# Machine learning of proxies for power systems reliability management

**Auteur :** Duchesne, Laurine
**Promoteur(s) :** Wehenkel, Louis
**Faculté :** Faculté des Sciences appliquées
**Diplôme :** Master en ingénieur civil électricien, à finalité approfondie
**Année académique :** 2015-2016
**URI/URL :** http://hdl.handle.net/2268.2/1374

*University of Liège*
*Faculty of Applied Sciences*
*Department of Electrical Engineering and Computer Science*


# Machine learning of proxies for power systems reliability assessment

Master thesis submitted for the degree of Msc in Electrical Engineering
by Laurine Duchesne

Supervised by Prof. Louis Wehenkel

*Academic Year 2015-2016*

*University of Liège*
*Faculty of Applied Sciences*
*Department of Electrical Engineering and Computer Science*


# Machine learning of proxies for power systems reliability assessment

Master thesis submitted for the degree of Msc in Electrical Engineering
by Laurine Duchesne

Supervised by Prof. Louis Wehenkel

*Academic Year 2015-2016*

# Machine learning of proxies for power systems reliability assessment

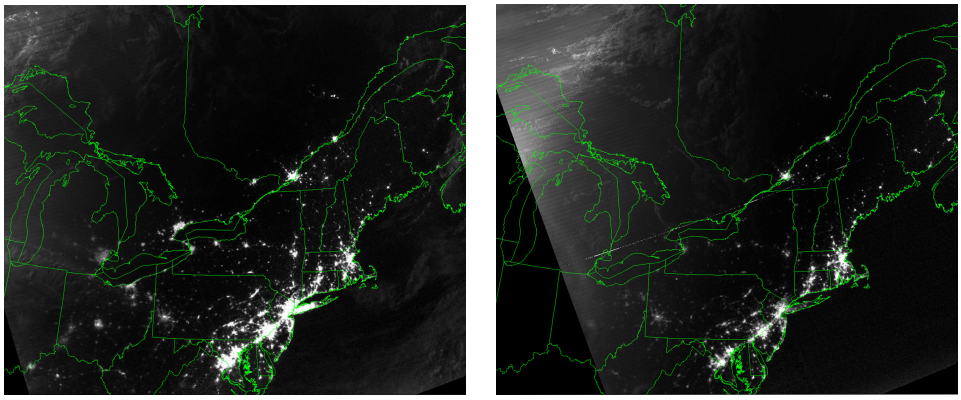
Laurine Duchesne, supervised by Prof. Louis Wehenkel

Electrical Engineering
University of Liège

*Academic year 2015-2016*

## Abstract

Nowadays, electrical power is essential for the functioning of our society. Power system reliability management intends to prevent service interruptions while minimising the socio-economic costs. In order to meet this purpose, it takes decisions based on a reliability criterion which is currently a deterministic criterion. However, several factors such as the ageing of the power systems' components, the increasing share of renewable energies and the development of the electricity market complicate the work of power system's operators. All of this calls for a new probabilistic reliability management approach and criterion, that is able to take into account the stochastic nature of a power system. This should allow operators to design and operate their system more economically while ensuring a desired level of reliability. In this context, the European project GARPUR developed a new reliability management approach and criterion. This master's thesis intends first to study the behaviour of this new reliability management approach and criterion when its use in real-time operation is anticipated in the day-ahead operation planning context, and to build simplified models (called proxies) with machine learning so as to mimic the optimisation tool implementing the criterion. In order to reach these objectives, a database is generated using Monte-Carlo simulations and the optimisation tool developed in the GARPUR project. The dataset is first analysed by using classical exploratory statistical methods and then by computing feature importances derived from tree-based ensemble methods to assess the link between input parameters and target quantities computed by optimising real-time decisions based on the reliability criterion. Finally, several machine learnt proxies are tested to assess their accuracy to predict several target outputs of the decision making program. The analysis demonstrated the importance of the weather on the results. A comparison of the proposed probabilistic reliability criterion with a "pseudo" deterministic criterion based on a predefined list of contingencies has shown that the deterministic criterion is more costly for an identical reliability level. Furthermore, considering the fixed N-1 set of contingencies is typically conservative, but may sometimes be insufficient. The building of proxies also revealed that tree-based ensemble methods seem to be the most accurate estimators. In particular the models based on extremely randomized trees were found to be the most accurate models to predict most target outputs.



(a) About 20h before the blackout (b) About 7h after the blackout

Figure 1: Representation of the blackout that occurs on 2003/08/14 in the Northeast of the United States and in Ontario (Canada). Figures (a) and (b) are satellite images (Images from [1]). It can be seen that there is almost no light in cities like Toronto and Ottawa seven hours after the blackout occurrence. The blackout is the worst consequence of a failure in power system reliability management. These images emphasise the importance of power system reliability management.

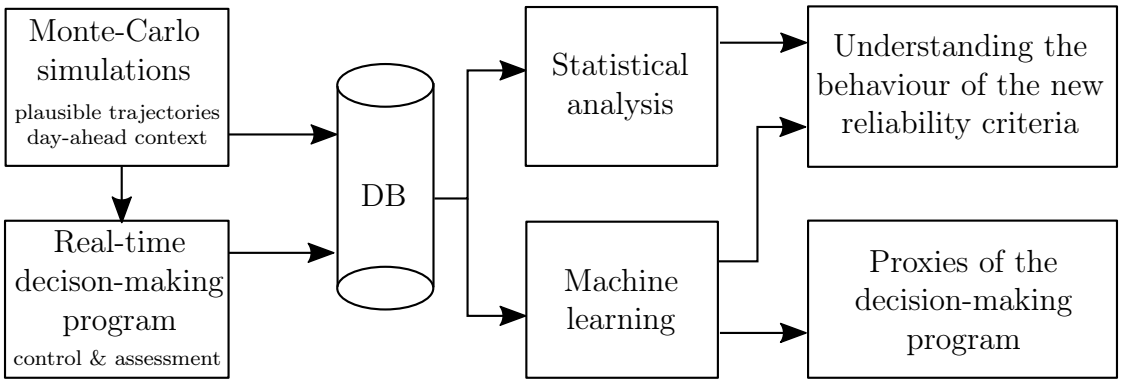



Figure 2: Block diagram representing the main steps of this master's thesis. First a database (DB) is generated in the following way. Monte-Carlo simulations are used to generate different plausible trajectories for the system in a day-ahead context(input data). Then, all these trajectories are applied to the real-time decision-making program based on the Reliability Management Approach and Criterion (RMAC) developed by GARPUR (output data). After that the database is analysed using classical statistical methods and with the help of machine learning in order to understand the behaviour of the RMAC. Finally, the last part consists in building proxies of the decision-making program.

# Acknowledgements

First of all, I would like to express my gratitude to my supervisor Professor Louis Wehenkel for the opportunity he gave me to work on this project and thanks to whom I discovered a fascinating field of study. He guided me throughout the year in my work and took the time to answer to my questions. I also would like to emphasize his great availability, his valuable comments and support throughout this year and his proofreading of my report.

Then, I am grateful to Efthymios Karangelos for all his help, suggestions and explanations as well as his implementation of the decision-making program on which this work is based.

I would also like to thank my family and friends for their support, patience and encouragements throughout all these years of study.

Finally, I would like to acknowledge Elodie Burtin and Hubert Woszczyk for their attentive proofreading.

Laurine Duchesne

# Contents

# Chapter 1

# Introduction

Electric power is essential in our modern society. Power system reliability management is therefore important to guarantee a continuous supply. Its purpose is to take optimal decisions in planning and operation, so as to maintain the reliability of the power system at a desired level while minimising the socio-economic costs.

Nowadays, most transmission system operators base their reliability management work on a deterministic criterion, named the N-1 criterion. Under this criterion, the system is considered as sufficiently reliable, if it can continue to operate without any service interruption after the loss of any single one of its components (lines, transformers,...). The uncertainty in power system is increasing due to several factors. Some of these factors are the ageing of power system components, the rise of renewable energies and the development of electricity markets. In order to take into account the stochastic behaviour of the system, it is desirable to switch to a new probabilistic reliability criterion. The European project GARPUR is aimed to "design, develop, assess and evaluate such new reliability criteria" [2]. This master's thesis is in continuity with the GARPUR project.

The purpose of this master's thesis is twofold. The first objective is to study the behaviour of the new probabilistic reliability criterion developed by the GARPUR project in the context of day-ahead operation planning. The second objective is to explore how machine learning could be exploited to build simplified models of a decision-making program, called proxies. The advantage in using proxies instead of the full program is to gain computational speed. However, it is at the expense of a loss in precision. Figure 1.1 illustrates the main steps of this master's thesis.

The first step is to create a database. Monte-Carlo simulations are used to generate the input variables which are then applied to the real-time decision-making program developed by Karangelos *et al.* in [3] to produce output variables. On one hand, these output variables are studied both statistically and with the help of machine learning algorithms and in particular with feature selection methods. On the other hand, supervised learning algorithms are applied to the database in order to build proxies in order to predict the target outputs.

The report is organised as follows. Chapter 2 is a reminder of the basics of power

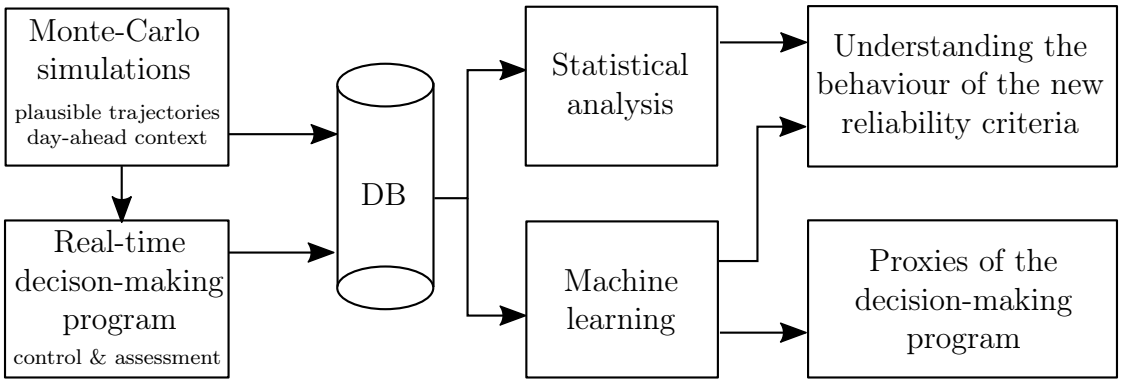



Figure 1.1: Main steps of the master's thesis.

system operations and describes the learning algorithms that are used throughout the report.

Chapter 3 describes the modelling of the real-time decision making problem, on which is based this work.

Chapter 4 explains in detail how the database is generated and in particular the different hypotheses used.

Chapter 5 is an analysis of the database generated according to the procedure developed in Chapter 4. The purpose is twofold: first verify that the database was correctly generated by checking if the constraints are met and secondly understand the influence of several parameters on the results of the decision making problem.

Chapter 6 determines the feature ranking based on their information about each target output with the help of random trees ensembles. This allows to identify those features that most strongly influence the value of each target output variable.

Finally, Chapter 7 studies the accuracy of proxies built by several machine learning algorithms when they are used to predict the target outputs of the problem.

The report ends by a short conclusion and the discussion of some future work in Chapter 8.

# Chapter 2

# Background

This chapter is dedicated to the theoretical background needed for this work. The first section presents some notions in power system operation, with a focus on reliability management. The second section introduces briefly the main learning algorithms used in this project. The machine learning library used in this work is Scikit-learn [4].

## 2.1 Power systems

Power systems are infrastructures that supply and deliver electricity to end-users. The delivering system is split into two distinct parts: a transmission system that delivers electrical power from generators to load centres and a distribution system that conveys this power from the load centres to consumers. Some end-users are directly connected to the transmission network. The voltage level is different for both systems. In Belgium, the transmission system is a very high-voltage network (70kV, 150kV, and 380kV) whereas the distribution systems concern the lower voltage levels [5].

This work focuses on power system reliability at the level of the transmission system and in particular on power system operations. Both terms are detailed in the two next subsections. Finally, this section ends with a comparison between power system operation in practice and in research along with the presentation of the GARPUR project.

### 2.1.1 Power system reliability

The reliability of a power system refers to the ability of this system to perform its intended tasks under given conditions. Power system reliability management intends to ensure that the system is reliable by taking a certain number of decisions under uncertainty in order to meet a reliability criterion while minimising the socio-economic costs. This reliability criterion is a basis to assess whether or not a system is reliable [6].

In other words, power system reliability management consists in ensuring the continuous operation of the system, despite the various disturbances and threats

that can act on the power system, and while at the same time maximising the social-welfare.

A power system is submitted to various threats and disturbances such as a component's outage and unexpected variations in demand and generation from renewable energies. Large disturbances such as the unexpected loss of a transmission line or a power plant are called contingencies. The term contingency also covers simultaneous outages of several components.

System's operators work to ensure that the system is always secure, that is the system is stable and within its operating limits, even after the occurrence of a contingency. Their work can be divided in three main tasks: grid development, asset management and power system operation.

Grid development consists in evaluating the possible long-term scenarios (over ten to twenty years) for the evolution of the level and location of the demand and the generation so as to adapt the structure of the transmission system in consequence. It should take into account the future level of reliability as well as an economic constraint. Asset management mainly consists in inspecting, repairing, and replacing the system's components. In particular, it consists in scheduling the preventive maintenance of components over a mid-term horizon of a few months to a few years, and repairing them in case of failure.

On the other hand, power system operation ensures that the system is always secure. In real-time, it consists in monitoring the system state and taking actions if needed to bring the system back to a secure state. Ahead in time, it consists in forecasting the possible future states of the system, over horizons of several hours to a few weeks, in order to plan those actions that need to be taken in advance to enable secure operation in real-time. Some actions are intra-day rescheduling, definition of the day-ahead reserves and maintenance rescheduling.

This work focuses on power system operation and therefore this is described in more detail in the following subsection.

### 2.1.2 Power system operation

This subsection first introduces the different operating states of a power system. Then the two aspects of reliability management are explained. They are reliability assessment and control. The first one consists in evaluating current reliability of the system and comparing it against pre-determined reliability criteria. The second evaluates the actions to be taken by the operator in order to maintain the reliability. Finally, this subsection ends with the description of some mathematical tools used by operators.

#### Operating states

Five operating states can be distinguished [7]:

1. The system is in a *normal* state if it is stable and within its operating limits. Furthermore, it is able to withstand contingencies.

2. If the system is stable and within its operating limits but is not able to withstand some contingencies, it is in an *alert* state. To come back to a normal state, some preventive actions should be taken by the operator.

3. If a contingency occurs in an alert state, it could lead to an *emergency* state, where some operational limits are violated. To come back to an alert state, emergency actions should be taken.

4. The system is in an *in extremis* state if some operating limits are violated and the system is unstable. In that case, some further components may be lost leading to service interruptions and possibly to a blackout, and the operator must take heroic actions in order to limit the loss.

5. Finally, the system is in a *restorative* state when it stabilises after an in extremis state. The operator tries to restore the lost components, load and generation, in order to go back to a normal or alert state as quickly as possible.

To put back the system in a normal or alert state, preventive and corrective (or emergency) actions can be taken. Preventive actions are taken before the occurrence of a disturbance and can be the starting-up or shutting-down of generating units or generation re-dispatch for example. Corrective actions are taken after the occurrence of a contingency. Possible corrective actions are load shedding, generation re-dispatch or setting of the phase-shifting transformers.

### Reliability assessment and control

Power system reliability management can be divided into two sub-problems: reliability assessment and reliability control.

Reliability assessment intends to evaluate the level of risk of the current operating state of the system (or, more generally, over a certain number of trajectories over a certain time-horizon), in order to verify whether it is compliant with the reliability criterion used by the system operator.

Reliability control consists in defining the preventive actions to apply in order for the system to be in a normal state and the corrective actions that will put the system back in a normal or alert state after occurrence of a contingency [8]. These actions can, for example, be determined with the help of a security-constrained optimal power flow (SCOPF) introduced hereunder.

### ED, UC, PF, OPF and SCOPF

Various optimisation problems solved by operators in order to operate the system and ensure the reliability while maximising the social-welfare of end-users are described hereafter.

One of of the simplest optimisation problem is the *economic dispatch* (ED). It is typically used in real-time to determine the electrical power to be generated by the individual generating units in operation, say during the next hour. It minimises the generation costs under the constraints that the total electrical power generated must be equal to the demand to be served and that the power produced by a generating unit must be within the production range of that unit. This dispatching neglects the transmission loss and the operating limits, such as the maximum power that can flow through a transmission line [9]. In order to take into account the output power from renewable energies, they can be considered here as a negative demand.

The *unit commitment* (UC) problem determines which generating units should be on or off at any moment during a certain future time-interval (typically of several hours to a few days) in order to minimise the generation costs. It is computed in advance and based on a forecast demand over the considered period. It takes into account the time during which the units must stay up once started up and also the time the units should stay down once shut down. Thus, for each time-step and for a certain number of subsets of the set $\mathcal{G}$ containing all the generating units, it computes the economic dispatch and it eventually selects the configuration that minimises the costs. It is a more complex problem than the economic dispatch, first because it embeds several such economic dispatch problems but especially because it also involves integer variables [9].

The state of an electrical system is described with an *AC power flow analysis* (PF). Given the active and reactive powers (or the voltage set-points) injected by the generating units or withdrawn by the loads at each bus, it is possible to compute the amount of active and reactive powers transiting in each transmission line and the voltage magnitude at each node. From that information, it is easy to compute the currents in each line and the voltage phases. A bus can be characterised by four variables: the active and reactive powers and the magnitude and phase of the bus voltage. Two variables per bus are enough to determine the whole system [10].

The AC power flow is computationally expensive. The DC power flow is an approximation of the AC power flow analysis that allows to have linear equations instead of quadratic ones (such as $P = RI^2$). The approximations listed hereafter are used.

- The relation between the active powers and the voltage phases is linear.
- The active losses in the transmission lines are neglected.
- The voltage magnitudes are expressed in per unit and are all equal to 1.
- The reactive power transit is neglected.

With a power flow analysis, the system's operators are able to verify that no network constraint is violated for a particular generation dispatch. One step further, it is possible to determine the best dispatching according to specific criteria with an *optimal power flow program* (OPF). The OPF solves multiple power flow problems in order to find this optimal dispatching while satisfying the network constraints.

The best solution can be different depending on the objectives. These ones could be to minimise the generation costs, to minimise the transmission line losses or to maximise the security for instance.

A *security-constrained optimal power flow* (SCOPF) intends to find an operating point such that all the system variables stay under limits, even in case of a contingency. Different security constraints can be defined and they depend on the reliability criterion.

### 2.1.3 Operation planning: practice vs. research

One important security criterion used by system operators is a deterministic criterion called N-1 criterion. A system operated according to this criterion is able to withstand the loss of any single one of its elements [11]. It could be a generator, a transmission line or a transformer for example. Other approaches are the N-$k$ criteria, where $k$ is an integer. $k$ is rarely greater than 2 because of the high computational cost it would require for larger values.

Current researches in reliability management tend to develop techniques to plan and operate the system using probabilistic criteria (*e.g.* [8][12][13][14] ). The probabilistic approach takes into account the probabilistic nature of a power system. Indeed, load and renewable energy generation as well as unplanned component outages are probabilistic events. This approach would allow the operators to evaluate the risk of a contingency and determine if it would be profitable to take actions. It would also allow for a better choice between preventive and corrective actions. With a probabilistic approach, it would be possible to design a system able to withstand the contingencies leading to a risk above a defined threshold and only these contingencies. It should result in a more economical system [15].

Nowadays, most transmission system operators (TSOs) in Europe are still using a deterministic criterion for reliability management [16]. Therefore there exists a gap between the research and the practice. In this context, the European project GARPUR intends to design a new reliability criterion to improve reliability management [2]. It would allow to reach a better balance between reliability and socio-economic costs.

In GARPUR, the new criterion is developed for three decision contexts (power system operations, asset management and grid development) in order to have a coherence at the different time-scales. It is intended to be deployed at the Pan European level. This criterion defines not only a criterion for assessment but also an optimisation problem for control and is called Reliability Management Approach and Criteria (RMAC). This RMAC in the real-time decision making context is described in the next chapter.

As already said in the introduction, this master's thesis is in continuity with the GARPUR project. Indeed, the purpose is to study the behaviour of the new reliability criteria in day-ahead operation planning context and then assess the use

of learning algorithms to build proxies of the decision-making programs that are based on these reliability criteria.

Before moving to the machine learning part of this background, it is noteworthy to mention that our work is certainly not the first work on the use of machine learning in the context of power system reliability management. Amongst several researches, one can cite the European project iTesla (Innovative Tools for Electrical System Security within Large Areas) [17]. The project aims to develop a new tool for real-time dynamic security assessment. In this project, machine learning is used in the context of operation planning so as to build security rules in the form of decision trees and these latter rules may then be used in real-time to predict the security-status of the system if a certain contingency occurs, based on the state of the system before occurrence of that contingency.

## 2.2 Machine learning

The main supervised learning algorithms used in this master's thesis are briefly introduced in the following subsections. In this work, the target outputs have numerical values and thus only regression models are used. We therefore limit the discussion to the regression counterpart of each learning algorithm. The first section introduces some general considerations and notations used in the sequel. The second section introduces two algorithms based on randomized ensembles of regression trees. Then the third and fourth sections describe linear regression algorithms and their kernelized versions: respectively support vector regression and ridge regression.

The last subsection already introduces the cross-validation method used in Chapters 6 and 7 to find the optimal meta-parameters for each learning algorithm described in this section. It also provides a reminder of all the meta-parameters that need to be tuned when applying these algorithms to a particular problem.

### 2.2.1 General notions for supervised regression

In supervised learning, the goal is to learn a mapping from an input space $\mathcal{X}$ to an output space $\mathcal{Y}$, based on a sample $S = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$ of input-output pairs. In our context, inputs are vectors in $\mathbb{R}^p$ and outputs are real numbers. $N$ is the number of observations in the dataset and $p$ the number of input features. The notation $\mathbf{x} \in \mathbb{R}^p$ is used to represent the vector of input feature values of one observation (or object), and $y \in \mathbb{R}$ denotes one output value.

In order to define the notion of a "good mapping" (or model, or predictor), one considers a loss function $\ell(y, y') \mapsto \mathbb{R}^+$, and a target probability distribution $\mathbb{P}_{\mathcal{X},\mathcal{Y}}$ over $\mathcal{X} \times \mathcal{Y}$, and for any given predictor $f \in \mathcal{Y}^{\mathcal{X}}$ defines its expected loss by

$$L(f, \mathbb{P}_{\mathcal{X},\mathcal{Y}}) = \mathbb{E}_{\mathbb{P}_{\mathcal{X},\mathcal{Y}}}\{\ell(f(\mathbf{x}), y)\}, \tag{2.1}$$

where $\mathbb{E}_{\mathbb{P}_{\mathcal{X},\mathcal{Y}}}$ denotes the mathematical expectation along the target distribution $\mathbb{P}_{\mathcal{X},\mathcal{Y}}$. The goal of supervised learning is then to exploit the available sample of

input-output pairs in order to find a predictor that minimises the expected loss. In our context of regression, we use the *square loss*, i.e. $\ell(y, y') = (y - y')^2$.

To analyse the properties of a given supervised learning algorithm, one needs to make further assumptions about the mechanism that generates the learning samples. In the standard setting of supervised learning, the target distribution is unknown, but one assumes that the learning sample $S$ is obtained by sampling identically and independently (*i.i.d.*) $N$ times from this target distribution $\mathbb{P}_{\mathcal{X},\mathcal{Y}}$. Thus the expected (square) loss of any $f$ can be estimated by its *empirical (square) loss*

$$\hat{L}(f, S) = \frac{1}{N} \sum_{i=1}^{N} \{(f(\mathbf{x}_i) - y_i)^2\}. \tag{2.2}$$

Therefore, many supervised learning algorithms explicitly consider the minimisation of the empirical loss as (part of) the objective function of an optimisation problem solved to find the best predictor in a certain space $\mathcal{F}$ of candidate input-output mappings. Below we use the notation $\hat{f}_S$ to highlight the dependence on the used sample $S$ of the predictor produced by a supervised learning algorithm.

The model $\hat{f}_S$ produced by a certain supervised learning algorithm and hence its expected loss (also called generalisation error or testing error) depends on the random nature of the learning sample $S$. Therefore, an expected generalisation error is defined as the expectation of the expected loss over all the learning samples, *i.e.* $\mathbb{E}_S\{\mathbb{E}_{\mathbb{P}_{\mathcal{X},\mathcal{Y}}}\{(\hat{f}_S(\mathbf{x}) - y)^2\}\}$. It can be decomposed into three terms. But let's first introduce the notion of bayes model. In general, the bayes model $f_B(\mathbf{x})$ is defined as the "ideal" predictor, i.e. the function $f \in \mathcal{Y}^{\mathcal{X}}$ minimising the expected loss; in the case of the square loss, the solution is $f_B(\mathbf{x}) = \mathbb{E}_{\mathbb{P}_{\mathcal{Y}|\mathbf{x}}}\{y\}$, and one shows moreover that the expected loss over all learning samples $S$ can be decomposed as:

$$\begin{aligned} \mathbb{E}_S\{\mathbb{E}_{\mathbb{P}_{\mathcal{X},\mathcal{Y}}}\{(\hat{f}_S(\mathbf{x}) - y)^2\}\} \;=\; & \mathbb{E}_{\mathbb{P}_{\mathcal{X},\mathcal{Y}}}\{(f_B(\mathbf{x}) - y)^2\} \\ & + \; \mathbb{E}_{\mathbb{P}_{\mathcal{X}}}\{(f_B(\mathbf{x}) - \mathbb{E}_S\{\hat{f}_S(\mathbf{x})\})^2\} \\ & + \; \mathbb{E}_{\mathbb{P}_{\mathcal{X}}}\{\mathbb{E}_S\{(\hat{f}_S(\mathbf{x}) - \mathbb{E}_S\{\hat{f}_S(\mathbf{x})\})^2\}\}. \end{aligned} \tag{2.3}$$

The first term is the residual error and quantifies the variation of $y$ from the bayes model $f_B(\mathbf{x})$. It cannot be avoided. The second term represents the bias, which is the error between the averaged model over all learning sets and the bayes model or true (conditional) expectation of the target variable. Finally, the third term represents the variance which indicates the variations of $\hat{f}_S(\mathbf{x})$ in function of the learning set $S$.

In the context of the design of machine learning algorithms, one major issue is to find the proper bias-variance tradeoff. Indeed, increasing the size of the hypothesis space $\mathcal{F}$ of the algorithm typically allows to reduce its bias; however this generally comes with an increase of its variance. This compromise is directly related to two other important concepts used in machine learning, overfitting and regularisation. A model overfits the data when it predicts well the learning set but has poor performance with the test set; it is said that it generalises poorly. Algorithms producing such models have a low bias but a too high variance. Regularisation tries to avoid overfitting by reducing the size of the hypothesis space, for instance by adding to the

loss function a penalty term taking into account the model complexity. It reduces the variance but may increase the bias.

We refer the reader to [18] for an in depth theoretical motivation of these ideas and the analysis of a large class of methods proposed in the machine learning community.

To conclude this section of general considerations, we would like to stress here the fact that once a sample $S$ has been used by a supervised learning algorithm in order to find a predictor, say $\hat{f}_S$, this sample can not be used a second time in order to obtain an unbiased estimate of the expected loss of this predictor $\hat{f}_S$. In practice, we therefore need to obtain a second sample $S'$, independent from $S$ and of sufficient size, in order to carry out the validation of the resulting predictors. In order to clearly distinguish these two samples whenever appropriate, we will use the term *learning sample* (or learning set) to denote the former, and *testing sample* (or test set) to denote the latter.

### 2.2.2 Ensembles of regression trees

In order to reduce the variance and/or the bias of a machine learning algorithm, it is possible to combine the predictions of several models. The methods performing this are called ensemble methods. There are two main families of ensemble methods: the averaging methods where the predictions of each model are averaged and the boosting methods that combine several weak learners to form a strong one.

The forest-based learning algorithms aggregate decision tree predictors. For regression models, the predictions obtained by each are averaged to give the final result.

Before describing the two methods used in this work, let's recall briefly how a regression tree is built [18]. Starting from the full learning set, one cuts it in two regions $R_1$ and $R_2$ according to a splitting criterion. This splitting criterion is defined by a splitting variable $v_j$ and a split point $s$ and thus $R_1$ and $R_2$ are functions of $(j, s)$. The prediction for both parts of the dataset is their mean output value, that is $\hat{f}(\mathbf{x}) = c_1 = E\{\mathbf{y}|\mathbf{x} \in R_1\}$ if $\mathbf{x} \in R_1$ and $\hat{f}(\mathbf{x}) = c_2 = E\{\mathbf{y}|\mathbf{x} \in R_2\}$ if $\mathbf{x} \in R_2$. The splitting criterion is chosen in such a way that it minimises the mean square error:

$$\min_{j,s} \left\{ \min_{c_1} \sum_{\mathbf{x} \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{\mathbf{x} \in R_2(j,s)} (y_i - c_2)^2 \right\} \tag{2.4}$$

Once the dataset is divided into two distinct parts, all the search for a splitting variable and a cut-point is done independently for both parts. The same method is used until the tree is completely built. It is possible to build incomplete trees and it is often done because a full tree is sensitive to overfitting. This is called pruning and it can be done before the tree is completely built (pre-pruning) or after (post-pruning). Pre-pruning can be done for instance by specifying a minimum number

of samples needed to split a node. This minimum number is called $n_{min}$. A tree can be post-pruned by removing some subtrees until finding the minimum average prediction error over an independent validation sample.

### Random forests

Breiman defined random forests in [19] as follows:

*"A random forest is a classifier consisting of a collection of tree-structured classifiers* $\{h(\mathbf{x}, \Theta_k), k = 1, ..., M\}$ *where the* $\{\Theta_k\}$ *are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input* $\mathbf{x}$*."*

$\Theta_k$ is a random vector, independent of $\Theta_{1,...,k-1}$ but drawn according to the same distribution, and $\mathbf{x}$ is the input vector. $h(\mathbf{x}, \Theta_k)$ is a classifier but the definition could be applied to regressors. In that case $h(\mathbf{x}, \Theta_k)$ is a regressor and the final prediction is the average value of each tree's result.

Different types of random forests are reported in the literature. As an example, bagging (**b**ootstrap **agg**regat**ing**) was introduced by Breiman in [20] and consists in using bootstrap replicates of the learning set to build each tree and then aggregating the predictions. A bootstrap replicate is a a subset of the learning set. In this subset, all the samples have been drawn randomly with replacement, meaning that the same object can be found several times in a particular subset. Another example of random forest was introduced by Dietterich in [21]. In that case, the splitting criterion is chosen uniformly randomly amongst the $K$ best splits.

In this work, the random forest algorithm is the one implemented by Scikit-learn and used by Breiman in [19] for his experiments. It consists in using bagging associated to a random selection of features. When splitting a particular node, instead of considering all the features and then selecting the best one to split the data, a number $K$ of features are randomly selected and the best split is chosen among those $K$ attributes. One can modify this parameter $K$ to introduce more or less randomness. When $K = 1$, the splitting variable is chosen completely at random whereas if $K$ is equal to the number of features, only bagging brings randomness.

Note that there is no post-pruning of the grown trees in [19].

To summarise, this algorithm has three *meta-parameters*, namely $M \geq 1$ (the number of trees in the ensemble), $K \in \{1, 2, \ldots, p\}$ (the smaller the $K$, the stronger the randomization), and $n_{min} \geq 2$ (the larger the $n_{min}$, the shorter the individual trees). In terms of accuracy, the larger $M$ the better, while the optimal values of $K$ and $n_{min}$ are problem dependent. In practice, all three parameters need to be tuned to the problem at hand.

### Extremely randomized trees

The extremely randomized trees algorithm [22] is also an ensemble method based on trees but it goes one step further in the randomness. As for the random forest,

it selects randomly $K$ features but the cut-point for each of them is also chosen at random. Then the best split is selected.

There are three parameters for this method. $K$ was already introduced and is the number of features selected at random. $M$ is the number of trees and $n_{min}$ is the minimum number of samples required to split a node.

Contrarily to other random forest algorithms, the whole learning set is used to learn the trees. This intends to minimise the bias.

This algorithm thus has exactly the same meta-parameters $M, k$, and $n_{min}$ as random forests. However, for a given problem, their suitable values may differ from those of the latter algorithm.

### 2.2.3 Support vector regression

Support vector machines are an important family of methods in machine learning. They usually perform well and their computational complexity is independent from the number of features.

Since only regression is used in this work, this section focuses on support vector regression (SVR). The idea behind SVR is to learn a linear function in the feature space $\mathcal{X}$ that deviates from the learning outputs by at most $\epsilon_{SVR}$ and that is as flat as possible. The problem is written as an optimisation problem that minimises the norm of the coefficient vector $\mathbf{w} \in \mathcal{X}$ under the constraints that all the samples should lie within a tube of radius $\epsilon_{SVR}$ around the predicted function. This function has the form: $\hat{f}(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$, with $b \in \mathbb{R}$.

However, it may happen that it is impossible to find such a function. In that case, it is possible to introduce two slack variables $\xi_i$ and $\xi_i^*$ in the optimisation problem so as to allow for deviations larger than $\epsilon_{SVR}$. These larger deviations are penalised via the $\epsilon$-insensitive loss function proposed by Vapnik and that is equal to 0 if $|y_i - \hat{f}(\mathbf{x}_i)| < \epsilon_{SVR},\ i = 1, ..., N$ and is equal to $|y_i - \hat{f}(\mathbf{x}_i)| - \epsilon_{SVR}$ otherwise.

Formally, the SVR problem is a convex optimisation problem that can be written such as [23]:

$$\begin{aligned} \min_{b,\mathbf{w},\xi,\xi^*} \quad & \frac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \\ \text{s.t.} \quad & \begin{cases} y_i - \mathbf{w}^T\mathbf{x_i} - b \leq \epsilon_{SVR} + \xi_i & \forall i = 1, 2, ..., N \\ -(y_i - \mathbf{w}^T\mathbf{x_i} - b) \leq \epsilon_{SVR} + \xi_i^* & \forall i = 1, 2, ..., N \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \tag{2.5}$$

The constant $C > 0$ defines a compromise between the model's smoothness ($\|\mathbf{w}\|_2$ denotes the Euclidean norm of the weight vector, whose components are proportional to the partial derivatives of the model predictions with respect to the individual input features) and the degree of tolerance for deviations greater than $\epsilon_{SVR}$. The

larger the $C$, the less deviations are tolerated but at a price of an increase in the model complexity.

In its dual form, this problem is stated as follows [18]:

$$\begin{aligned} \min \quad & \epsilon_{SVR} \sum_{i=1}^{N} (\alpha_i^* + \alpha_i) - \sum_{i=1}^{N} y_i(\alpha_i^* - \alpha_i) + \frac{1}{2} \sum_{i,i'=1}^{N} (\alpha_i^* - \alpha_i)(\alpha_{i'}^* - \alpha_{i'}) \mathbf{x}_i^T \mathbf{x}_{i'} \\ \text{s.t.} \quad & \begin{cases} 0 \leq \alpha_i, \alpha_i^* < C \\ \sum_{i=1}^{N} (\alpha_i^* - \alpha_i) = 0 \end{cases} \end{aligned} \tag{2.6}$$

and its solution is

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^{N} (\alpha_i^* - \alpha_i) \mathbf{x_i}^T \mathbf{x} + b. \tag{2.7}$$

The term $(\alpha_i^* - \alpha_i)$ is equal to 0 for all points lying in the tube defined by $\epsilon_{SVR}$. We skip the derivations leading to the calculation of $b$ at the optimum.

If a linear function is not suitable in the input space, it is possible to learn a linear function in a space induced by a positive kernel. Let's call it the kernel space. A kernel $K$ is a symmetric function such that $K(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$, where $\mathcal{X}$ is the input space. The function $K$ is a positive kernel if for any finite choice of inputs $\mathbf{x}_1, \ldots, \mathbf{x}_m$ the $m \times m$ Gram matrix $[K(\mathbf{x}_i, \mathbf{x}_j)]$ is positive semi-definite. Mercer's theorem states that if the kernel is positive, it is possible to express it as $K(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^T \phi(\mathbf{x}')$, where $\phi(\cdot)$ is a (typically non-linear) mapping from the input space to the kernel space. With this relation, instead of computing $\phi(\mathbf{x}_i)$ and $\phi(\mathbf{x})$ and then replacing $\mathbf{x}_i^T \mathbf{x}$ by $\phi(\mathbf{x}_i)^T \phi(\mathbf{x})$ in equations (2.6) and (2.7), it is possible to replace the dot product in the kernel space by the function $K(\mathbf{x}_i, \mathbf{x})$. This is computationally more efficient [24].

This is called the kernel trick and it can be used for many linear algorithms. Using the kernel trick, the SVM predictor becomes:

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^{N} (\alpha_i^* - \alpha_i) K(\mathbf{x}_i, \mathbf{x}) + b. \tag{2.8}$$

It is worth noticing that, contrarily to tree-based estimators, SVR are not scale-invariant nor shift-invariant. If the ranges of the features are too different, it is generally preferable to normalise them before applying the algorithm.

### 2.2.4 Ridge regression and kernel ridge regression

Ridge regression is a least-square regression with a $\mathbb{L}2$-norm normalisation. The vector of regression coefficients $\mathbf{w}$ are shrunk towards 0 to reduce the variance. The regression model is expressed as: $\hat{f}(\mathbf{x}) = \mathbf{w}^T \mathbf{x}$ and is linear in its parameters.

The coefficients can be obtained with the following formula:

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} \left( y_i - \mathbf{w}^T \mathbf{x}_i \right)^2 + \lambda \left\| \mathbf{w} \right\|_2^2 \right\}, \tag{2.9}$$

where $\lambda$ is the regularisation meta-parameter. We will further denote by $X \in \mathbb{R}^{N \times p}$ the input data matrix and by $y \in \mathbb{R}^N$ the vector of target outputs, in order to write this expression in a matrix form. This gives:

$$\begin{aligned} \mathbf{w}^* &= \underset{\mathbf{w}}{\operatorname{argmin}} \left\{ \left\| (\mathbf{y} - \mathbf{X}\mathbf{w}) \right\|_2^2 + \lambda \left\| \mathbf{w} \right\|_2^2 \right\} \\ &= \underset{\mathbf{w}}{\operatorname{argmin}} \left\{ (\mathbf{y} - \mathbf{X}\mathbf{w})^T (\mathbf{y} - \mathbf{X}\mathbf{w}) + \lambda \mathbf{w}^T \mathbf{w} \right\} \end{aligned} \tag{2.10}$$

The greater the value of $\lambda$, the greater the shrinkage. When $\lambda = 0$, the ridge regression is equivalent to a least-square regression. Compared to a least-square regression, the ridge regression has a larger bias but a smaller variance that could result in a smaller prediction error [18].

The kernel ridge regression is a ridge regression with a kernel [24]. There is no clear dot product in that case but it is possible to make it appear. The solution of equation (2.10) is:

$$\mathbf{w}^* = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I}_p)^{-1} \mathbf{X}^T \mathbf{y}, \tag{2.11}$$

where $\mathbf{I}_p \in \mathbb{R}^{p \times p}$ is the identity matrix. If we apply the matrix inversion lemma, we can rewrite this as:

$$\mathbf{w}^* = \mathbf{X}^T (\mathbf{X}\mathbf{X}^T + \lambda \mathbf{I}_N)^{-1} \mathbf{y}, \tag{2.12}$$

and define the vector of dual variables $\alpha = (\mathbf{X}\mathbf{X}^T + \lambda \mathbf{I}_N)^{-1} \mathbf{y}$. In order to work in a kernel space, the term $\mathbf{X}\mathbf{X}^T$ can then be replaced by the Gram matrix $[K(\mathbf{x}_i, \mathbf{x}_j)]$ derived from the learning sample in order to compute the dual variables. Their values are then injected into equation (2.12):

$$\mathbf{w}^* = \mathbf{X}^T \alpha = \sum_{i=1}^{N} \alpha_i \mathbf{x}_i, \tag{2.13}$$

and the predictor $\hat{f}(\mathbf{x})$ is then expressed by:

$$\hat{f}(\mathbf{x}) = (\mathbf{w}^*)^T \mathbf{x} = \sum_{i=1}^{N} \alpha_i \mathbf{x}_i^T \mathbf{x} = \sum_{i=1}^{N} \alpha_i K(\mathbf{x}_i, \mathbf{x}), \tag{2.14}$$

where the kernel function $K$ can be used to make predictions.

It is worth noticing that we have used in our derivations a linear model (without intercept term $b$), and not an affine model as in the SVR. Therefore, the data is assumed to be centred prior to the application of this method. As for the SVR, ridge regression is not scale-invariant.

### 2.2.5 Meta-parameter tuning with cross-validation

Each of the aforesaid models depend on meta-parameters that influence the computational complexity, smoothness, and most notably the accuracy of the learned models. One way to determine suitable problem specific values of these parameters is to use a K-fold cross-validation (CV). The principle of K-fold cross-validation is to divide the learning set into K folds, learn a model with K-1 folds and compute the prediction error with the remaining fold. If we repeat this procedure K times with a different fold left out, we obtain K prediction errors that can be averaged for assessing the generalisation error [18]. This is called the cross-validation score in the sequel.

In order to find the best meta-parameters for each estimator, we define a range of values for each meta-parameter and then determine the cross-validation score for each combination. The optimal meta-parameters are the ones that maximise the CV score.

It is interesting at this point to remind the meta-parameters that will be tuned in Chapters 6 and 7. The random forest and extremely randomized trees have three meta-parameters: $K$ is the number of randomly selected features at each node, $M$ is the number of trees and $n_{min}$ the minimum number of samples to split a node.

The ridge regression and kernel ridge regression algorithms depend on $\lambda$, a coefficient that controls the amount of shrinkage. On the other hand, two parameters influence the complexity of the SVR models: $\epsilon_{SVR}$ and $C$. $\epsilon_{SVR}$ determines the maximal error $y - \hat{f}(\mathbf{x})$ that is ignored and $C$ is a regularisation parameter.

Finally, some parameters are kernel-dependent. The three kernels used in this work are the linear kernel, the polynomial kernel and the Gaussian radial basis function (rbf) kernel. In Scikit-learn, they are defined as:

- linear kernel: $\mathbf{x}^T\mathbf{x}$,
- polynomial kernel: $(\gamma\mathbf{x}^T\mathbf{x} + 1)^d$,
- Gaussian rbf: $e^{-\gamma\|\mathbf{x}-\mathbf{x}'\|_2^2}$.

One kernel meta-parameter is $\gamma$ for both the polynomial and the rbf kernel. The polynomial kernel has an extra parameter: the degree $d$ of the polynomial. The parameter $\gamma$ influences the range of actions of the kernel. It is sometimes written as $\frac{1}{2\sigma^2}$ and since the rbf can be seen as a low-pass filter, $\frac{1}{2\gamma}$ can be seen as the bandwidth of the kernel [24].

# Chapter 3

# Modelling of the real-time decision making problem

This chapter describes the real-time decision-making problem developed by Karangelos *et al* in [3] and [25]. The model is based on a probabilistic reliability management approach and criterion (RMAC).

Three principles define the RMAC: a reliability target, a socio-economic objective and a discarding principle. These 3 notions are developed in the first section. The second section describes how the decision-making problem is implemented. Finally, the last section introduces the IEEE-Reliability Test System 1996 (IEEE-RTS96)[26]. This is the system used to apply the decision-making program.

The content of this chapter is based on the article [3] and the supplementary material [25]. Most notations are borrowed from these papers.

## 3.1 General principle

The model is based on a probabilistic Reliability Management Approach and Criterion (RMAC). This approach is characterised by 3 elements set forth hereafter.

A *reliability target* ensures that the probability of unacceptable states is below a certain threshold $\epsilon$. An unacceptable state is a state where the interruption of services to end-users has a severe socio-economic impact. It may occur in case of occurrence of a neglected contingency or a failure of the corrective actions. Mathematically, the reliability target is expressed as: $P\{(x_0, x_c, x_c^b) \notin X_a | (c, b) \in \mathcal{C} \times \mathcal{B}\} \leq \epsilon$, where $X_a$ is the set of acceptable states and $(x_0, x_c, x_c^b)$ describes the sequence of states of the system resulting from the application of a contingency $c$ and a corrective control behaviour $b$ to the pre-contingency state $x_0$; $x_c$ is the post-contingency state before the application of corrective control and $x_c^b$ the post-contingency state after the application of corrective decisions. $\mathcal{C}$ is the set of considered contingencies and $\mathcal{B}$ is the set of considered possible behaviours of the corrective actions. Notice that set $\mathcal{B}$ always contains one so-called normal behaviour, corresponding to perfect operation of the planned corrective control, as well as one or several abnormal behaviours, corresponding to the various failure modes of corrective control.

The objective function is a *socio-economic objective*. The optimisation program finds a set of preventive and corrective actions that minimises the costs of preventive and corrective control and the risk which is defined as the expected cost of service interruptions. To compute it, a severity function measures the socio-economic impact of service interruptions. Mathematically, the risk $R$ is defined as: $R = \sum_{c \in \mathcal{C}} \pi_c(w_0) \cdot \sum_{b \in \mathcal{B}} \pi_b(w_0) \cdot S(x_c^b, w_0)$ where $w_0$ represents the weather conditions, $\pi_c(w_0)$ is the probability of occurrence of contingency $c \in \mathcal{C}$, $\pi_b(w_0)$ is the probability of the realisation of behaviour $b \in \mathcal{B}$ and $S(x_c^b, w_0)$ is the severity function depending on the post-contingency state of the system $x_c^b$ and the weather conditions $w_0$. Note that $\pi_b$ depends on the weather conditions but also on the post-contingency state $x_c$ and the corrective decisions $u_c$.

Given the large number of possible contingencies, a *discarding principle* is defined to reduce the set of contingencies. Indeed, many contingencies have a severity function close to 0 or a small probability to occur and it is thus not necessary to consider the full set. The subset of discarded contingencies is such that the total risk associated with them is below a constant threshold $\Delta E$, *i.e.* $R_{\mathcal{C} \setminus \mathcal{C}_c} = \sum_{c \in \mathcal{C} \setminus \mathcal{C}_c} \pi_c(w_0) \sum_{b \in \mathcal{B}} \pi_b(w_0) \cdot S(x_c^b, w_0) \leq \Delta E$, where $\mathcal{C}_c \subseteq \mathcal{C}$ contains only the non-discarded contingencies.

In order to state the problem formally, several elements still need to be defined:

- $\mathcal{U}_0(x_0)$ is the set of all possible preventive decisions for a pre-contingency state $x_0$.
- $u_0 \in \mathcal{U}_0(x_0)$ is a preventive control decision.
- $\mathcal{U}_c(x_0, u_0, c)$ is the set of all possible corrective actions for the contingency $c$, state $x_0$ and preventive control decision $u_0$.
- $u_c \in \mathcal{U}_c(x_0, u_0, c)$ is a corrective control decision for a state $x_0$, a preventive decision $u_0$ and if contingency $c \in \mathcal{C}_c$ occurs.
- $\mathcal{U}(x_0)$ is the set of all possible combinations of preventive and corrective decisions for a pre-contingency state $x_0$.
- $u \in \mathcal{U}(x_0)$ is a combination of preventive and corrective decisions for a pre-contingency state $x_0$.
- $w_0$ represents the weather conditions and is considered constant regarding the operation horizon.
- $CP(x_0, u_0, w_0)$ is the preventive control cost.
- $CC(x_c, u_c, w_0)$ is the corrective control cost.

The problem is stated as follows:

$$\min_{u \in \mathcal{U}(x_0)} \left[ CP(x_0, u_0, w_0) + \sum_{c \in \mathcal{C}_c} \pi_c(w_0) \cdot \left( CC(x_c, u_c, w_0) + \sum_{b \in \mathcal{B}} \pi_b(w_0) \cdot S(x_c^b(u_0, u_c), w_0) \right) \right] \tag{3.1}$$

$$\text{s.t.} \quad P\{(x_0, x_c, x_c^b) \notin X_a | (c, b) \in \mathcal{C}_c \times \mathcal{B}\} \leq \epsilon \tag{3.2}$$

while

$$R_{\mathcal{C} \setminus \mathcal{C}_c} \leq \Delta E \tag{3.3}$$

This problem can be divided into two parts: a control part based on equations (3.1) and (3.2) and an assessment part checking condition (3.3). The control part aims at finding a set of optimal preventive and corrective decisions and the assessment part computes the severity associated with each contingency given the control decisions and determines the residual risk $R_{\mathcal{C} \setminus \mathcal{C}_c}$.

The optimisation problem can be solved iteratively by first defining a subset of contingencies $\mathcal{C}_c$, then computing the optimal preventive and corrective decisions with the control part and finally evaluating if the residual risk $R_{\mathcal{C} \setminus \mathcal{C}_c}$ is less than the meta-parameter $\Delta E$ (3.3) using the assessment part. If not, the assessment part, based on the control decisions previously computed, is used to return a new subset of contingencies and the control part is applied with this new subset. This iterative process ends when the residual risk is below the defined threshold.

## 3.2 Implementation

This section describes in detail the implementation of the control part and the assessment part of the program.

The control part consists in a Security Constrained Optimal Power Flow (SCOPF). Unlike usual SCOPF programs, the set of considered contingencies is not based on a deterministic approach such as the N-1 (or N-k) criterion but on a probabilistic criterion as defined by the RMAC (equation (3.3)). This optimisation program is implemented with GAMS (General Algebraic Modelling System), which is a modelling system for optimisation programs. The SCOPF implementation is described in the first subsection.

The assessment part evaluates the risk for the system with a cascade failure simulator. It is implemented with MATLAB and is described in the second subsection.

### 3.2.1 SCOPF

The implemented SCOPF is a mixed integer linear programming (MILP) problem. The problem finds the control decisions that minimise the cost of power system reliability management. In order to have a linear program, the DC power flow approximation is used.

As defined by the RMAC, the objective function consists in the minimisation of operation costs, ie. the minimisation of the sum of the preventive action costs, expected corrective action costs and risk.

Three types of states are considered in this problem: a pre-contingency state before any contingency occurs, an intermediate short-term post-contingency state before applications of corrective actions and a final post-contingency state after applications of corrective actions. For each such state, the power at each node should be balanced and the power flowing through transmission lines should be below the thermal ratings (permanent ratings are used for the pre-contingency and final states, while short-term ratings are used for the intermediate state). In case the limits are violated, it is considered that all the load is lost.

The basis for preventive and corrective actions is the market clearing outcome. The market clearing is a simple optimisation problem that finds the dispatch of generating units' production that matches the demand while minimising the cost. The line thermal ratings are not taken into account.

A preventive decision $u_0$ consists in the re-dispatching of the power output of generating units. The preventive action cost is thus the preventive re-dispatching cost compared to the market clearing dispatch.

A corrective decision $u_c$ is composed of a set of elementary corrective actions. An elementary corrective action can be either the re-dispatch of a generating unit or the adjustment of a phase-shifting transformer. Each elementary action can fail with a given probability and the corrective decision $u_c$ is considered as failed if at least one of the elementary operations (phase shift or re-dispatch) has failed. The probability of failure of corrective control $\pi_f(u_c)$ is thus the sum of the probability of failure of each elementary operation composing the correction decision. This is expressed as:

$$\pi_f(u_c) = \sum_{g \in \mathcal{G}} \pi_f^g \cdot \gamma_g^c + \sum_{l_p \in \mathcal{L}_p} \pi_f^{l_p} \cdot \lambda_{l_p}^c,$$

where $\mathcal{G}$ is the set of generators, $\pi_f^g$ is the probability of failure of the re-dispatch of generator $g \in \mathcal{G}$ and $\gamma_g^c$ is a binary variable which is equal to 1 if the re-dispatch of generator $g$ is one corrective elementary operation of $u_c$ and 0 otherwise. Similarly, $\mathcal{L}_p$ is the set of phase-shifting transformers, $\pi_f^{l_p}$ the probability of failure of the adjustment of transformer $l_p \in \mathcal{L}_p$ and $\lambda_{l_p}^c$ is a binary variable with value 1 if the setting of transformer $l_p$ is one corrective elementary action applied in case of occurrence of contingency $c$ and with value 0 otherwise.

The corrective control cost for a contingency $c$ is equal to the sum of the costs of the elementary corrective actions corresponding to $c$. The expected corrective control cost is the sum of the corrective costs for each contingency in $\mathcal{C}_c$ multiplied by the probability of occurrence of this contingency. The cost of phase-shift transformers is considered to be zero. Note that only the ramp-up costs are considered in the objective function. The corrective control cost is always positive.

In the implementation, two scenarios could lead to an unacceptable state. The first one happens when there is a failure of the corrective control after the occurrence of the corresponding contingency. Indeed, it is assumed that if the corrective operation was not necessary to put the system back under its long-term operational limits, it would not be chosen in a cost-minimisation program. Thus a failure of one corrective elementary operation would lead to the loss of the system's load. The second scenario corresponds to situations where it is impossible to apply corrective action after the occurrence of a contingency $c$ because the short-term operational limits are violated before. In these cases, two possibilities exist for the program. The first one solves the problem with preventive actions that will prevent the system to be above the short-term limits after occurrence of $c$. The other possibility is to relax the post-contingency constraints for $c$. In that case, the occurrence of $c$ will lead to an unacceptable trajectory and $c$ is a relaxed contingency. This is possible because of the reliability target, that allows the probability of unacceptable states to be greater than 0.

The severity is the cost in case of unacceptable state and the risk is the expected severity. In this implementation, the cost is computed as the product of the value of lost load and the amount of lost load, i.e. the amount of active load:

$$S_{max} = \sum_{i \in \mathcal{D}} v_i(w_0) \cdot p_{d,i},$$

where $\mathcal{D}$ is the set of demand nodes, $v_i(w_0)$ is the value of lost load for node $i$ (possibly dependent on weather conditions) and $p_{d,i}$ is the load active power of demand node $i$. The assumption that all the load is lost in case of unacceptable behaviour is there to compensate any under-estimation that could be caused by other assumptions. The probability of being in an unacceptable state is:

$$P_{unacceptable} = \sum_{c \in \mathcal{C}_r} \pi_c + \sum_{c \in \mathcal{C}_c \setminus \mathcal{C}_r} \pi_c \pi_f(u_c),$$

where $\mathcal{C}_r \subseteq \mathcal{C}_c$ is the subset of relaxed contingencies, and $\pi_f(u_c)$ denotes the probability of failure of the corrective control actions planned for contingency $c$. From the reliability target (3.2), we have that: $P_{unacceptable} < \epsilon$. Finally, the risk $R_{\mathcal{C}_c}$ for the set of contingency $\mathcal{C}_c$ is defined as:

$$R_{\mathcal{C}_c} = \sum_{c \in \mathcal{C}_r} \pi_c \cdot S_{max} + \sum_{c \in \mathcal{C}_c \setminus \mathcal{C}_r} \pi_c \pi_f(u_c) \cdot S_{max}.$$

It is worth noting that relaxing a contingency increases the risk and thus is accounted in the objective function.

### 3.2.2 Assessment

The conservative assumption used in the SCOPF formulation that all the load is lost in case of unacceptable state is not used for the computation of the risk $R_{\mathcal{C}}$, corresponding to the risk considering all contingencies. Instead, a cascade failure

simulator simulates what would happen in the system if a contingency occurs. It is therefore possible to determine the amount of load shedding and generator tripping for each contingency.

The simulator considers each contingency separately. For each one of them, it first updates the topology of the system and searches for isolated nodes and islands. An island is a zone of the network separated from the rest of the system. The program checks if the nodes and the islands are balanced, meaning that the production is equal to the demand.

If they are balanced, a DC power flow problem is solved to determine the power flow in each line. If an island is not balanced at this step, the power flowing through the lines belonging to that island are temporarily set to 0. Finally the line power flows are analysed to determine if there will be an overload on a line and the time before the next overload.

When an island or an isolated node is unbalanced, the program studies the possibility to balance it.

If the load is greater than the production, the program determines if it is possible to increase the production before the next overload, considering the ramping up limits. If all the generators are at maximum capacity, some loads are shed.

Similarly, if the demand is less than the production, the possibility of decreasing the production is evaluated according to the ramping down rates and the time available before the next overload. If all generators are at minimum capacity, some generators are tripped. It may happen in that case that some load shedding are necessary if the power output of the generator tripped was greater than the imbalance.

If the time available before the next overload is not sufficient to balance the system, the power output of each generator is updated as much as the available time allows it.

Time is then advanced until reaching the next overload. The system topology is updated in consequence and everything is recomputed: search for isolated nodes or islands, balance checking, DC power flow, search for next overload and balancing. If the system is not balanced but there is no next overload, time is advanced by steps of 0.5 minute.

The program stops for a given contingency when all the load is lost or if all the islands and isolated nodes are balanced with no overload coming.

The assessment program is schematically illustrated in Figure 3.1.

Once the amount of load shedding per contingency is computed, it is possible to assess the criticality of a contingency. It is defined as the sum of lost load multiplied by the value of lost load. Thus for contingency $c$, the criticality $crit_c$ is computed

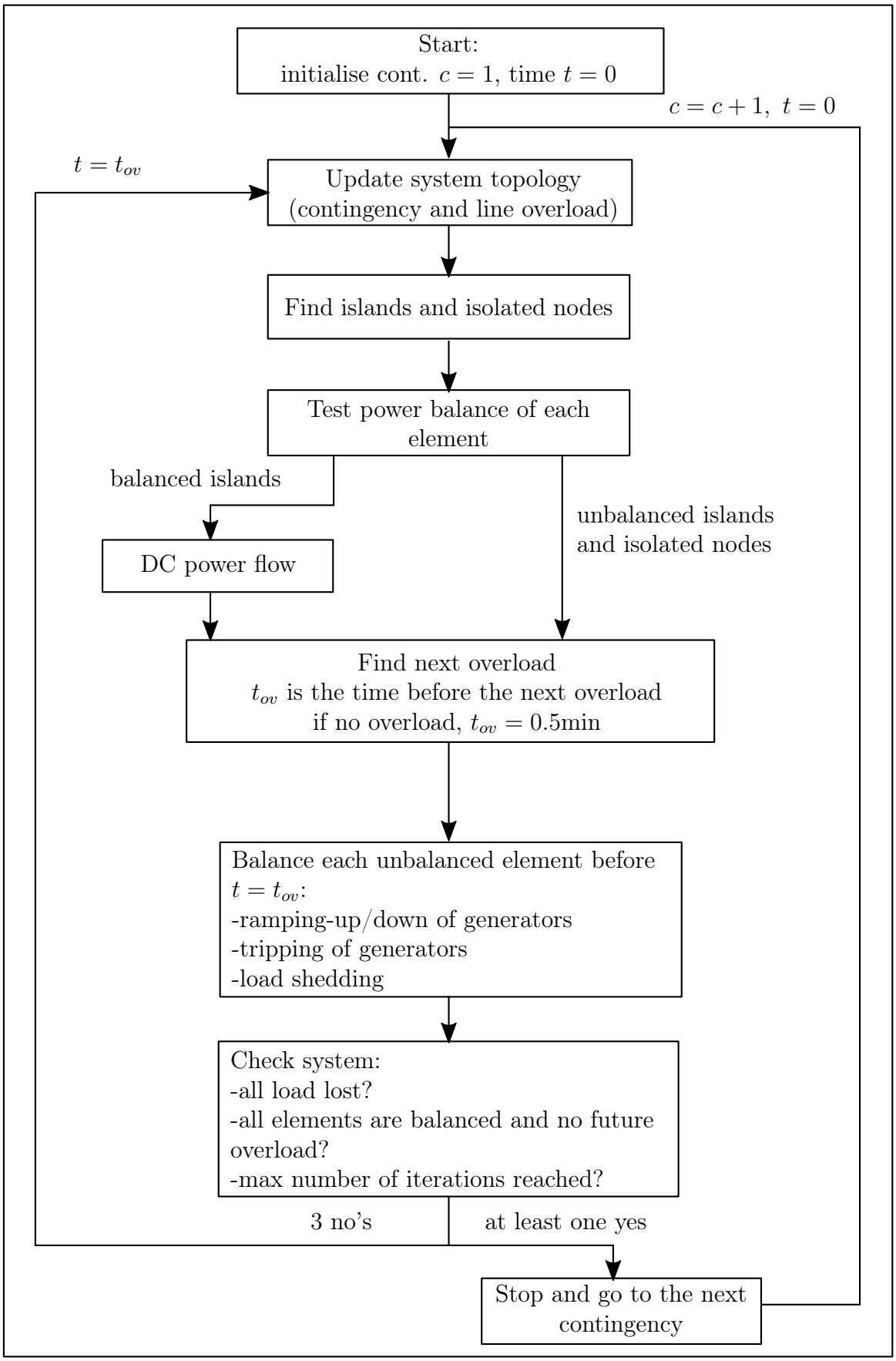



Figure 3.1: Main steps of the assessment program for contingency $c$.

as $crit_c = \sum_{i \in \mathcal{D}} v_i(w_0) \cdot ll_{i,c}$, where $v_i(w_0)$ is the value of lost load for bus load $i \in \mathcal{D}$ and $ll_{i,c}$ is the amount of lost load for bus load $i$ and contingency $c$. The expected criticality $ec_c$ of a contingency $c$ is the criticality multiplied by the probability that this load shedding happens. If no corrective decision was pre-computed for that contingency, this is simply the probability for $c$ to occur. If a corrective decision was pre-computed for that $c$, the probability of load shedding is the probability for $c$ to occur multiplied by the probability that the application of the corrective decision fails.

It gives:

$$ec_c = \sum_{c \in \mathcal{C} \setminus \mathcal{C}_{corr}} \pi_c \cdot \sum_{i \in \mathcal{D}} v_i(w_0) \cdot ll_{i,c} + \sum_{c \in \mathcal{C}_{corr}} \pi_c \pi_f(u_c) \cdot \sum_{i \in \mathcal{D}} v_i(w_0) \cdot ll_{i,c},$$

where $\mathcal{C}_{corr} \in \mathcal{C}_c$ is the set of contingencies for which corrective control is planned, and $\pi_f(u_c)$ represents the probability of failure of a precomputed corrective control decision $u_c$.

Eventually, the risk $R_{\mathcal{C}}$ of the system is the sum for all contingencies of the expected criticality: $R_{\mathcal{C}} = \sum_{c \in \mathcal{C}} ec_c$. The residual risk is the sum of the expected criticality for the set of contingencies neglected in the SCOPF: $R_{\mathcal{C} \setminus \mathcal{C}_c} = \sum_{c \in \mathcal{C} \setminus \mathcal{C}_c} ec_c$.

## 3.3 Introduction to the IEEE-RTS96 system

A modified version of the single area IEEE Reliability Test System-1996 [26] is used in our simulations in order to study the model of Karangelos *et al.* It is a 24-bus system developed in 1996. The system topology is represented in Figure 3.2. This schematic is not to scale. Compared to the standard IEEE-RTS96 network, 6 wind farms with a capacity of 200 MW have been added in order to analyse a system with high wind penetration. The number and locations of these wind farms are consistent with what is advised in [27].

Each generating unit is characterized by a unit type, from which depend the minimum and maximum power this generating unit is able to provide and also the ramping rate. In the system, there are 33 generating units that are described in Appendix A. The maximum generation capacity of the system is 3405MW considering only the thermal units and the 6 hydropower units. The cost of generation is non-linear and in this system is approximated by a piecewise linear function with four steps. The redispatch is considered constant whatever the value of the change.

The demand (or load) is defined as well in [26]. There are 17 bus loads. These ones are characterized by a load in MW and a voll (value of lost load) which allows to estimate the cost of service interruptions.

There are 33 transmission lines and 5 phase-shifting transformers (PST). They are characterised by a reactance, a continuous rating in MVA and a short-term

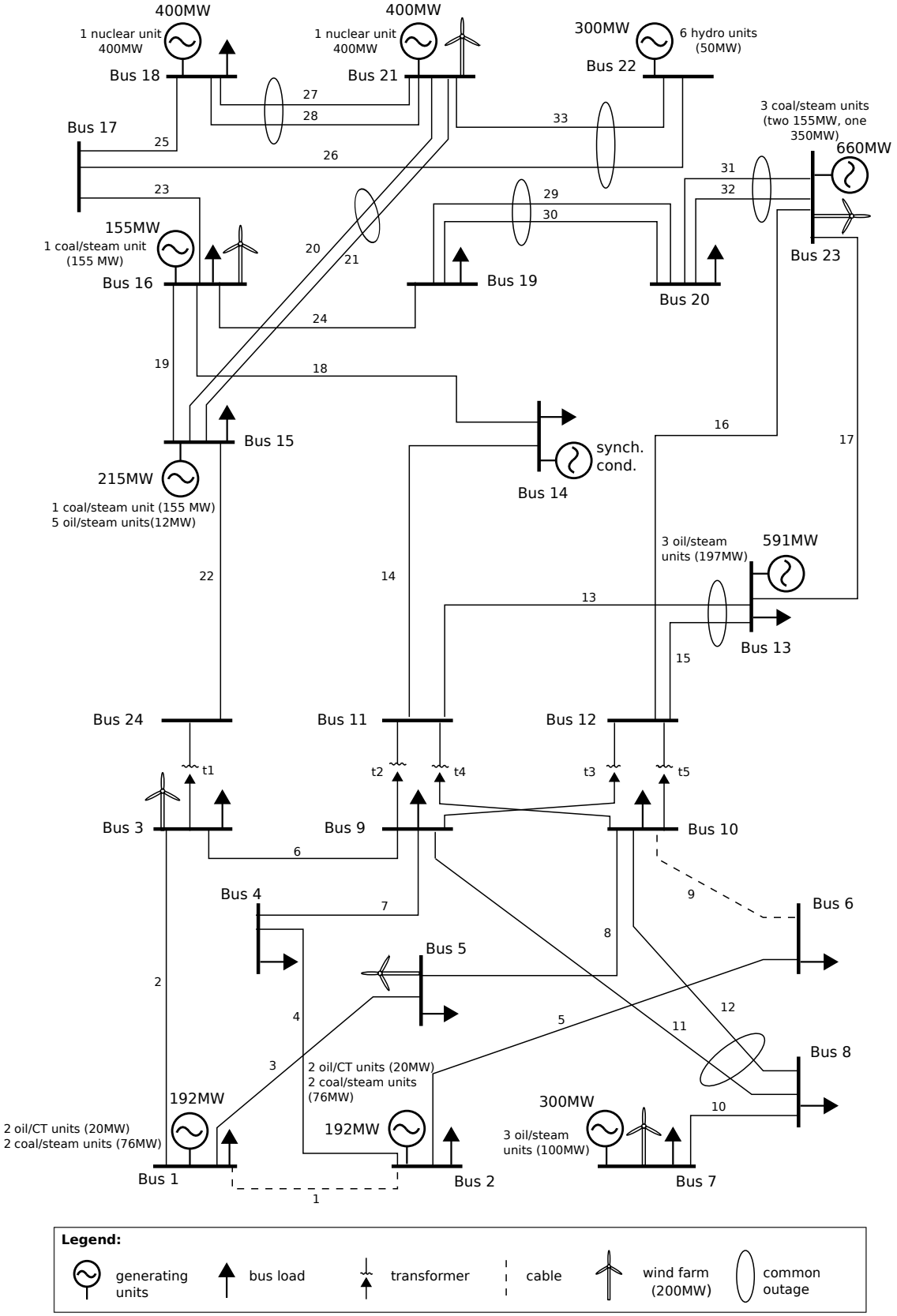



Figure 3.2: Topology of the IEEE-RTS96 system.

emergency rating also in MVA. The two last terms represent the maximum loading that the line can carry respectively continuously and during 15 minutes. The PST are also defined by a minimum and a maximum phase shift.

In this model, the contingencies consist in the outage of transmission lines. There are 46 contingencies considered. One is the no outage case, 33 correspond to the outage of a single line and 5 correspond to the outage of a single PST. Finally the 7 last contingencies are common mode double outage contingencies. They are circled in Figure 3.2. For the sake of simplicity, the approximation used in this work is that the set $\mathcal{C}$ is composed of only 46 contingencies. In order to compute the probability of occurrence of these contingencies, the number of forced outages per year given in [26] for each line and transformer is used. In case of common outages, the recommendation is to take 7.5% of the single outage rate as the common outage rate. The method used to compute these probabilities is explained in Chapter 4.

# Chapter 4

# Database generation

This chapter describes how the database is generated in a day-ahead context. In a day-ahead operation planning context, demand and wind production are predicted the day before for the next day. In practice, there is always a forecast error and some unforeseen events that could happen such as the outage of one transmission line or of a generating unit. Using Monte-Carlo simulations, it is possible to generate a large number of plausible trajectories based on the forecast errors and the probability of outage of components.

The generation of this database serves several purposes. The first objective is to study the behaviour of the RMAC described in Chapter 3 from different perspectives and in particular to compare it with the N-1 criterion. The second one is to create a dataset in order to apply learning algorithms and build proxies.

The first section describes the different steps in the generation of the database. The second section explain how Monte-Carlo simulations are used to create a plausible dataset and in particular the probability distributions of each parameters. Finally, the last section describes how the meta-parameters $\Delta E$ and $\epsilon$ are chosen.

## 4.1 Overall principle

The overall principle of calculation is schematically illustrated in Figure 4.1. It proceeds as follows.

We consider that at noon the day before, the demand and the wind production for the 24 hours of next day are predicted. A market clearing for the next day is then computed, based on the assumption that all the lines and all the generators will be available tomorrow. This step will define a unit commitment (UC), an economic dispatch (ED) and the reserve requirements for the next day. The market clearing does not consider any security criterion, *i.e.* the thermal limits of transmission lines are neglected.

When this is done, a state (corresponding to 1 hour) is randomly generated. An economic dispatch, based on the market clearing computed the day before, is calculated to balance the production and the demand. Only the N-security crite-

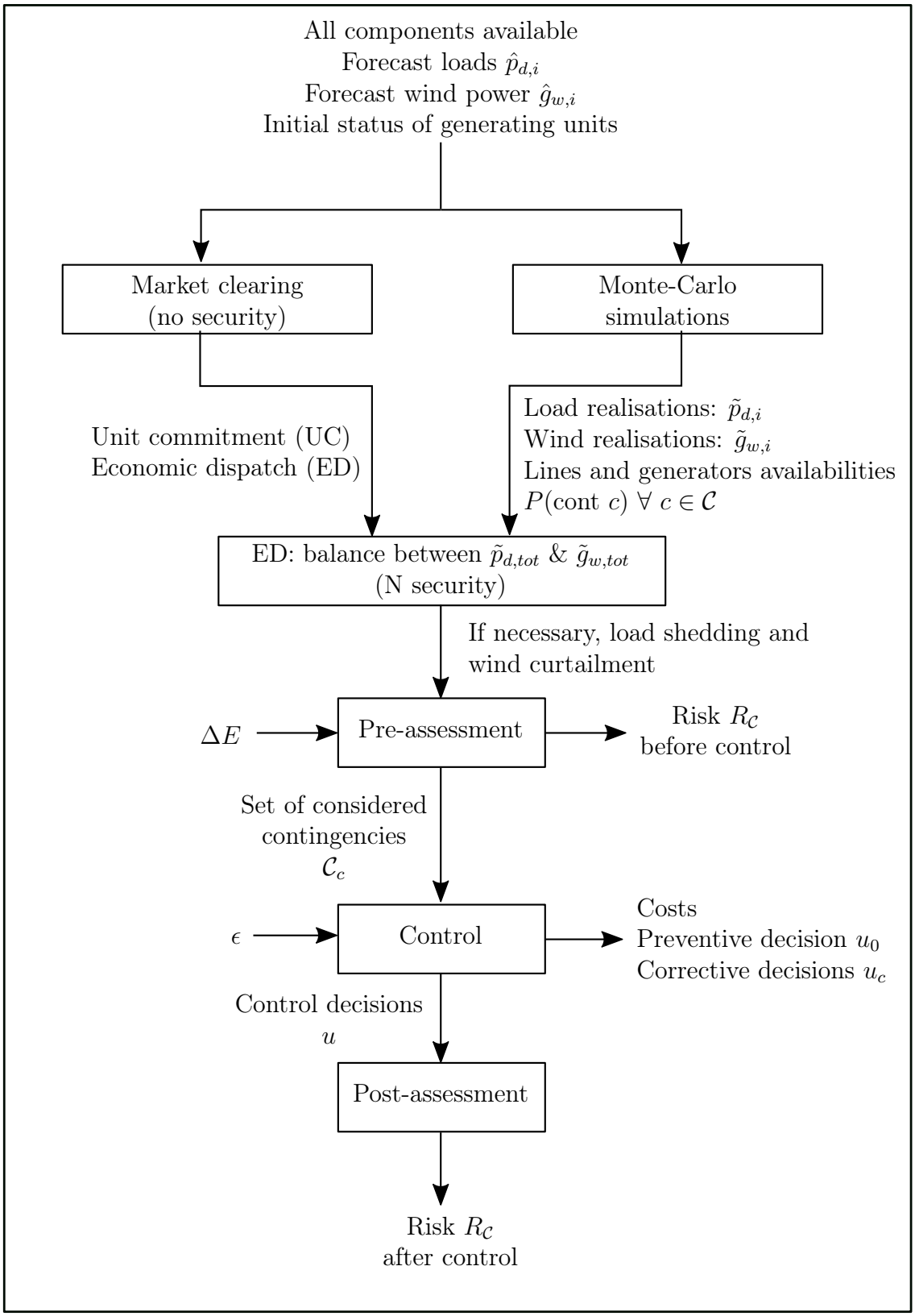



Figure 4.1: Schematic representation of the main steps of the database generation.

rion is taken into account. It means that the modification to the initial dispatch is done to balance the production and the load but also to prevent the violation of the

transmission lines' permanent thermal limits. It may happen that it is impossible to balance the production and the consumption given the availability of the transmission lines and generating units. In that case, load shedding and wind curtailment are performed. Whenever it is possible, wind curtailment is preferred.

From this new dispatch that takes into account the transmission line limits and the realisation of the variables, the pre-assessment block first calculates the risk before any further security control decisions are decided, and by using the cascade simulator for each contingency in the full set $\mathcal{C}$. In addition to the "Risk before control", it outputs a set of contingencies $\mathcal{C}_c$ that will be used by the control RMAC in order to define some preventive and corrective actions. This set of contingencies is chosen in such a way that the residual risk $R_{\mathcal{C}\setminus\mathcal{C}_c}$ (before control) is smaller than the specified value of $\Delta E$. It is worth noticing that since the market outcome is computed with a forecast load and not a realisation, the basis for preventive and corrective actions is not the market clearing outcome as was explained in Chapter 3 but the economic dispatch computed to balance the production and the realised load.

Finally, the post-assessment block uses the preventive and corrective control decisions computed by the RMAC control block in order to re-calculate the current risk of the system, again by using the cascade simulator and by considering all the contingencies and not only the ones in $\mathcal{C}_c$, and by also taking into account the possible failures of the computed corrective actions. The residual risk $R_{\mathcal{C}\setminus\mathcal{C}_c}$ is also re-computed to take into account the effects of preventive actions. It is possible that the residual risk does not satisfy anymore the discarding principle $R_{\mathcal{C}\setminus\mathcal{C}_c} \leq \Delta E$. In that case, an iterative approach should be used in order to find the set $\mathcal{C}_c$ and the corresponding preventive and corrective control actions satisfying the principle. The iterative approach is not used in this work in order to facilitate the interpretation of the results.

## 4.2 Day-ahead forecast data

For the generation of the database, the predictions for only one day are considered and the samples consist of different realisations for that particular day. The day chosen is a Tuesday during Winter.

The forecast demand per bus is given by the IEEE-RTS 96 data [26]. Figure 4.2 displays the load profile for the day in question.

The wind farms power generation forecast data per hour are taken from the website [28]. The system wind output power forecast for the day considered in this work is represented in Figure 4.3.

A more detailed description of the data used for market clearing can be found in Appendix A.

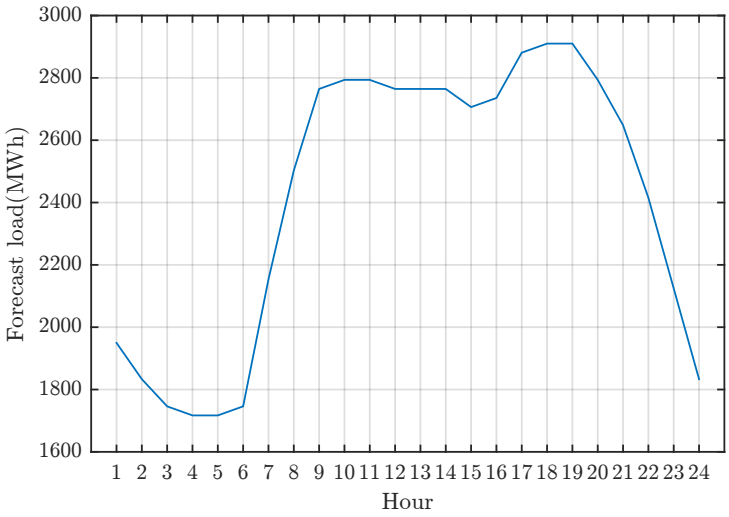



Figure 4.2: System forecast load per hour for Tuesday of week 50 (Winter). Data comes from [26].

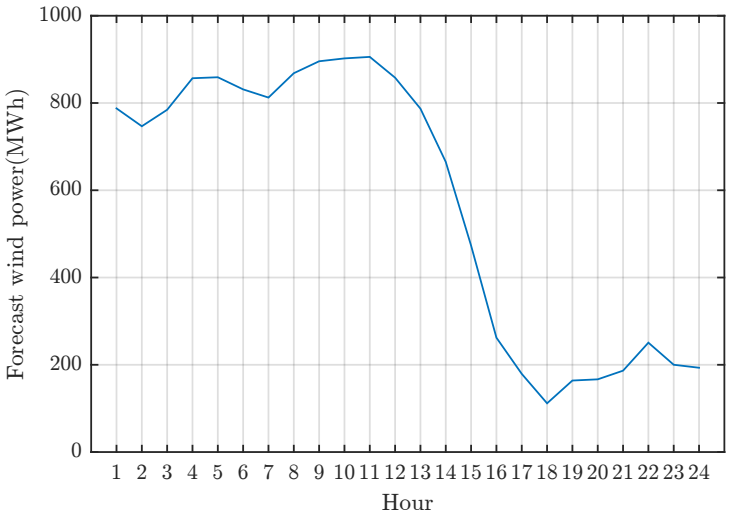



Figure 4.3: System forecast wind power per hour. Data comes from [28].

## 4.3 Monte-Carlo simulations

Each input variable could be modified to generate a database. However, in the day-ahead context of this project, the variables that should change are the ones that vary from one day to another, such as the load, the wind farms generation and the topology of the network for instance.

Some parameters are chosen to be constant, either because we do not think they would change significantly from one day to another or because there are not enough data to describe the uncertainties on the predictions. Those parameters are the following:

- the production cost,
- the value of lost load,
- the minimum and maximum phase shifts of the transformers,
- the reactance of the transmission lines,
- the continuous and emergency rating of the lines,
- the probability of failure of corrective actions.

The following subsections describe in detail how the other parameters are generated. The pseudo-code describing the Monte-Carlo simulations can be found in Appendix C.2.

### 4.3.1 Weather

The weather has an influence on the real-time reliability management of the system. For the sake of simplicity, only two states are considered: normal and adverse. The weather is adverse when there is lightning, thunderstorm, ice or snow [29]. It is assumed that the whole system has the same weather.

In this work, the weather has an impact on the transmission lines. Indeed, when the weather is adverse, their probability of outage increases. This affects the availability of transmission lines as well as the probabilities of contingencies.

In order to be able to analyse the weather influence, each state is generated twice: once with a normal weather and the other with an adverse weather.

### 4.3.2 Generating units

To assess which generators will be available the next day, a Poisson distribution is used to determine the probability of unavailability [30]. The failure rate per hour ($\lambda$) is the inverse of the Mean Time To Failure (MTTF), given in [26] for each type of generating unit. The probability of having $x$ occurrences of failure in $t$ hours is:

$$P_t(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}. \tag{4.1}$$

We consider that the generating units will not be repaired in one day and thus the probability of unavailability of a generator is given by 1 minus the probability that there is no failure ($x = 0$) in $t$ hours:

$$P_{outage}(t) = 1 - P(x = 0) = 1 - e^{-\lambda t}. \tag{4.2}$$

The probabilities of outage for each unit group can be found in Table 4.1. For simplicity, the generating units are assumed to be in only two possible states, up or down.

In our study, the parameter $t$ ranges from 12 to 35h.

Table 4.1: Number of outages per hour $\lambda$ and probability of unavailability for $t = 12$, 24 and 36h for each type of generating unit.

| Unit group | $\lambda$ [outage/h] | $P_{outage}(t = 12h)$ | $P_{outage}(t = 24h)$ | $P_{outage}(t = 36h)$ |
|---|---|---|---|---|
| U12 | $3.4\ 10^{-4}$ | 0.0041 | 0.0081 | 0.012 |
| U20 | $2.2\ 10^{-3}$ | 0.026 | 0.052 | 0.077 |
| U50 | $5.1\ 10^{-4}$ | 0.006 | 0.012 | 0.018 |
| U76 | $5.1\ 10^{-4}$ | 0.0061 | 0.012 | 0.018 |
| U100 | $8.3\ 10^{-4}$ | 0.01 | 0.0198 | 0.0296 |
| U155 | $1\ \ 10^{-3}$ | 0.012 | 0.025 | 0.037 |
| U197 | $1.1\ \ 10^{-3}$ | 0.013 | 0.025 | 0.037 |
| U350 | $8.7\ 10^{-4}$ | 0.010 | 0.021 | 0.031 |
| U400 | $9.1\ 10^{-4}$ | 0.011 | 0.027 | 0.032 |

### 4.3.3 Transmission lines and transformers

The same reasoning as for the generating units is used to determine the probability that there is a forced outage on a line or a transformer. However, the weather has a strong impact on the probability of failure of transmission lines.

The failure rate per hour is significantly increased for adverse weather. Some examples given in [31] show that the number of outages per year can be multiplied by a factor 30. In this work, the outage rates per year will be multiplied by 30 in case of adverse weather for all the lines. The outage rates per year for the cables (branches 1 and 9) and the transformers are not impacted by the weather.

Two examples of the probabilities of unavailability can be found in Table 4.2. The first example (line 1) is a cable and therefore the probability of unavailability is independent of the weather. The second example (line 22) is a typical line and both probabilities of unavailability are shown. The transmission lines outage probabilities for all the lines and both weather conditions can be found in Appendix C.1.

Table 4.2: Number of outages per hour $\lambda$ and probability of unavailability for $t = 12$, 24 and 36h for the cable 1 and the transmission line 22. $\lambda$ for cable 1 is identical for both weather status.

| Line | $\lambda$ [outage/h] | $P_{outage}(t = 12h)$ | $P_{outage}(t = 24h)$ | $P_{outage}(t = 36h)$ |
|---|---|---|---|---|
| 1 (both) | $2.74\ 10^{-5}$ | $3.29\ 10^{-4}$ | $6.57\ 10^{-4}$ | $9.86\ 10^{-4}$ |
| 22 (normal) | $4.68\ 10^{-5}$ | $5.62\ 10^{-4}$ | $1.12\ 10^{-3}$ | $1.68\ 10^{-3}$ |
| 22 (adverse) | 0.00140 | 0.0167 | 0.0331 | 0.0493 |

The outage probabilities of the transformers for different values of $t$ can be observed in Table 4.3. The number of outages per hour is identical for the 5 phase-shifting transformers.

Table 4.3: Number of outages per hour $\lambda$ and probability of unavailability for $t = 12$, 24 and 36h for a phase-shifting transformer.

| $\lambda$ [outage/h] | $P_{outage}(t = 12h)$ | $P_{outage}(t = 24h)$ | $P_{outage}(t = 36h)$ |
|---|---|---|---|
| $2.28\ 10^{-6}$ | $2.74\ 10^{-5}$ | $5.48\ 10^{-5}$ | $8.22\ 10^{-5}$ |

### 4.3.4 Load realisation

The error on the load for day-ahead forecast is assumed to follow a normal distribution [29] with a mean equal to 0. In order to take into account both the correlation in the fluctuation of the different bus loads and their independence, two error terms are added to the prediction value. The first one is common to all the loads and the second is independent. In other words, for the bus load $i$, the realisation load $\tilde{p}_i$ is given by:

$$\tilde{p}_i = (1 + \epsilon_{\alpha_d} + \epsilon_{\beta_{d,i}}) \cdot \hat{p}_i, \tag{4.3}$$

where $\hat{p}_i$ is the predicted load of bus $i$, $\epsilon_{\alpha_d}$ is constant for all buses and drawn randomly according to a normal law $\mathcal{N}(0, \alpha_d)$ and $\epsilon_{\beta_{d,i}}$ is different for each bus and drawn according to a law $\mathcal{N}(0, \beta_d)$.

**Determination of parameters $\alpha_d$ and $\beta_d$**
In order for the realisations to be close to what could be encountered in practice, the parameters $\alpha_d$ and $\beta_d$ are determined according to values found in the literature. In a study realised by the NREL (National Renewable Energy Laboratory) in 2012, a normalised standard deviation $\sigma_{d,global}$ of about 0.03 was found for the total day-ahead load forecast error [32]. The normalised standard deviation is defined as the standard deviation divided by the yearly average load. The error on a single bus load is assumed to have a greater standard deviation. It is arbitrarily defined as $\sigma_{d,local} = 0.1$.

The variance of the total load realisation can be computed:

$$\begin{aligned}
&\sum_{i=1}^{17} \tilde{p}_i = \sum_{i=1}^{17} \hat{p}_i(1 + \epsilon_{\alpha_d} + \epsilon_{\beta_{d,i}}) \\
\Leftrightarrow\ &\text{var}\left(\sum_{i=1}^{17} \tilde{p}_i\right) = \text{var}\left(\sum_{i=1}^{17} \hat{p}_i(1 + \epsilon_{\alpha_d} + \epsilon_{\beta_{d,i}})\right) \\
\Leftrightarrow\ &\text{var}\left(\sum_{i=1}^{17} \tilde{p}_i\right) = \text{var}\left(\epsilon_{\alpha_d} \sum_{i=1}^{17} \hat{p}_i\right) + \text{var}\left(\sum_{i=1}^{17} \hat{p}_i \epsilon_{\beta_{d,i}}\right) \\
\Leftrightarrow\ &\text{var}\left(\sum_{i=1}^{17} \tilde{p}_i\right) = \left(\sum_{i=1}^{17} \hat{p}_i\right)^2 \alpha_d^2 + \text{var}\left(\sum_{i=1}^{17} \hat{p}_i \epsilon_{\beta_{d,i}}\right).
\end{aligned}$$

Considering that $\hat{p}_i = l_i \hat{p}_{tot}$ for the bus load $i$ with $l_i \geq 0$ and $\sum_{i=1}^{17} l_i = 1$ and $\hat{p}_{tot}$ being the total predicted load, it is possible to simplify this expression. Indeed, the

variables $l_i$ are constant for each hour (as defined in [26]). One can write:

$$\begin{aligned}
\Leftrightarrow \quad & \text{var}\left(\sum_{i=1}^{17}\tilde{p}_i\right) = \left(\sum_{i=1}^{17}\hat{p}_i\right)^2 \alpha_d^2 + \text{var}\left(\sum_{i=1}^{17} l_i \hat{p}_{tot}\epsilon_{\beta_{d,i}}\right) \\
\Leftrightarrow \quad & \text{var}\left(\sum_{i=1}^{17}\tilde{p}_i\right) = \left(\hat{p}_{tot}\right)^2 \alpha_d^2 + \left(\hat{p}_{tot}\right)^2 \sum_{i=1}^{17} l_i^2 \beta_d^2 \\
\Leftrightarrow \quad & \left(p_{avg}\right)^2 \sigma_{d,global}^2 = \left(\hat{p}_{tot}\right)^2 \alpha_d^2 + \left(\hat{p}_{tot}\right)^2 \sum_{i=1}^{17} l_i^2 \beta_d^2 \\
\Leftrightarrow \quad & \left(\hat{p}_{tot}\right)^2 \sigma_{d,global}^2 \simeq \left(\hat{p}_{tot}\right)^2 \alpha_d^2 + \left(\hat{p}_{tot}\right)^2 \sum_{i=1}^{17} l_i^2 \beta_d^2 \\
\Leftrightarrow \quad & \sigma_{d,global}^2 \simeq \alpha_d^2 + \sum_{i=1}^{17} l_i^2 \beta_d^2.
\end{aligned}$$

$p_{avg}$ is defined as the average load over a year. The approximation that $\hat{p}_{tot} \simeq p_{avg}$ is used to simplify the calculations.

The variance of the load on bus load $i$ can be computed as:

$$\begin{aligned}
& \tilde{p}_i = \hat{p}_i(1 + \epsilon_{\alpha_d} + \epsilon_{\beta_{d,i}}) \\
\Leftrightarrow \quad & \text{var}\left(\tilde{p}_i\right) = \text{var}\left(\hat{p}_i(1 + \epsilon_{\alpha_d} + \epsilon_{\beta_{d,i}})\right) \\
\Leftrightarrow \quad & \sigma_{d,local}^2 (\hat{p}_i)^2 = (\hat{p}_i)^2 \alpha_d^2 + (\hat{p}_i)^2 \beta_d^2 \\
\Leftrightarrow \quad & \sigma_{d,local}^2 = \alpha_d^2 + \beta_d^2
\end{aligned}$$

Finally, one obtains a system of two equations with two unknowns:

$$\begin{cases}
\sigma_{d,global}^2 &= \alpha_d^2 + \sum_{i=1}^{17} l_i^2 \beta_d^2 \\
\sigma_{d,local}^2 &= \alpha_d^2 + \beta_d^2
\end{cases}$$

Injecting the values of $\sigma_{d,global}$, $\sigma_{d,local}$ and $l_i$, we have $\alpha_d = 0.0144$ and $\beta_d = 0.09896$. This is approximated by $\alpha_d = 0.015$ and $\beta_d = 0.1$.

In order to check the results, 24000 samples are generated with $\alpha_d = 0.015$ and $\beta_d = 0.1$, thus 1000 samples per hour. Note that this set of samples is only used to verify the computation of the values of $\alpha_d$ and $\beta_d$. The relative forecast error $E_{d,i}^h$ per bus load and per hour is computed as:

$$E_{d,i}^h = \frac{\hat{p}_{d,i}^h - \tilde{p}_{d,i}^h}{\hat{p}_{d,i}^h},$$

where $i = 1, ..., 17$ represents the bus load index and $h$ the hour. The exponent $h$ is just a notation. When analysing the standard deviation of the relative forecast error, we obtain an average standard deviation of 0.1012 and a maximum and minimum values of respectively 0.0938 and 0.1076. It corresponds to what is expected.

To find the relative standard deviation of the system load forecast error for each hour, one computes the relative standard deviation as the standard deviation of $\tilde{p}_{d,tot}^h$

divided by the system average load over one year. For the IEEE-RTS 96 system, $p_{avg} = 2194MWh$. The mean relative standard deviation is 0.0309 and the maximum value is 0.0380 while the minimum value is 0.0219. It is close to the relative standard deviation observed by the NREL. Therefore, the assumption $\hat{p}_{tot} \simeq p_{avg}$ is valid.

Three histograms of the total demand for 2 different hours of the day can be seen in Figure 4.4 and the histograms representing the distribution of samples at 2 bus loads for the same hour are represented in Figure 4.5.

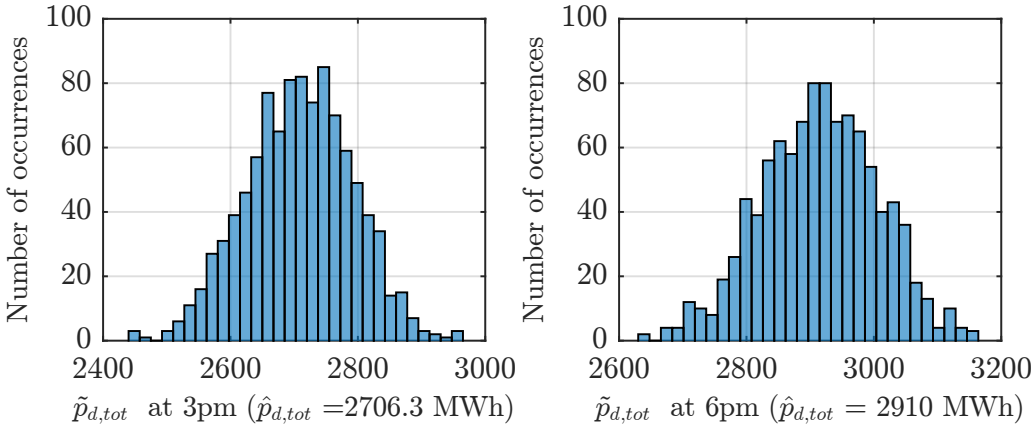



Figure 4.4: Histograms of the system load realisations at respectively 3pm and 6pm. The normalised standard deviations are respectively 0.0344 and 0.0380.

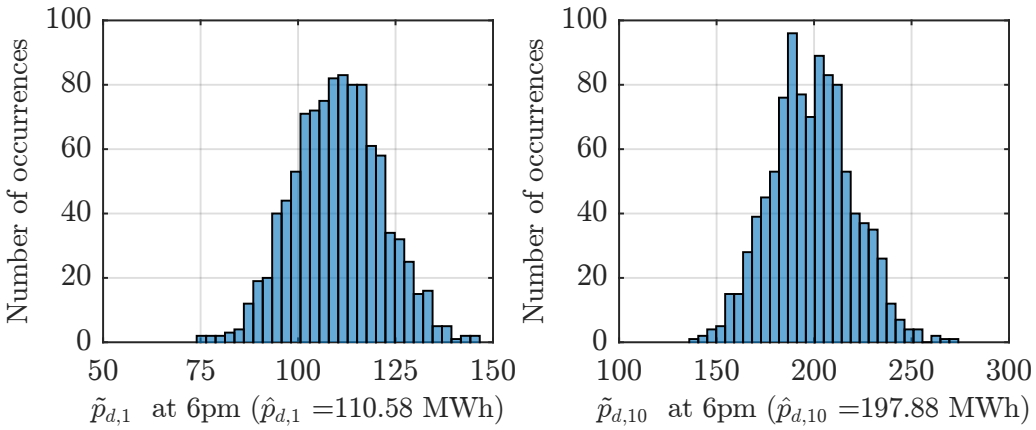



Figure 4.5: Histograms of the load realisations at 6pm respectively for bus load 1 and 10. The normalised standard deviations are respectively 0.1030 and 0.1063.

To identify the relationship between the load forecast error of several nodes, some scatter plots of the relative forecast error are displayed in Figure 4.6. It can be seen that there is few correlations between the errors at the different bus loads.

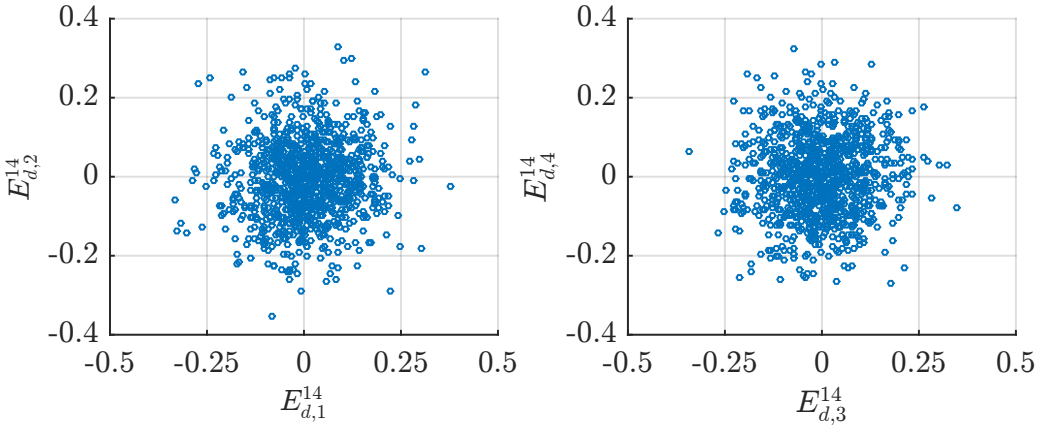



Figure 4.6: Scatter plots of the relative load forecast error $E^h_{d,i}$ for different bus loads at 2pm.

There is no change in the demand in function of the state of the weather. The weather influence is included in the normal distribution of forecast error.

### 4.3.5 Wind generation realisation

The forecasting of wind farms production is one of the main issues of operation planning. The usual approximation is to consider that the forecast error follows a normal distribution as for the load. However several recent studies showed that this is not the case. In [33], wind forecast errors from several countries were analysed and it was concluded that a hyperbolic distribution could better fit the error than a normal distribution. Another proposal was to use a Beta pdf [34].

Given the difficulty to obtain some real parameters for these two laws, a normal distribution is chosen and is defined in the same way as for the load forecast error. The realised generation $\tilde{g}_i$ for the wind farm $i$ is thus given by:

$$\tilde{g}_{w,i} = (1 + \epsilon_{\alpha_w} + \epsilon_{\beta_{w,i}}) \cdot \hat{g}_{w,i},$$

where $\hat{g}_{w,i}$ is the predicted production of wind farm $i$, $\epsilon_{\alpha_w}$ is constant for all wind farms and drawn randomly according to a normal law $\mathcal{N}(0, \alpha_w)$ and $\epsilon_{\beta_{w,i}}$ is different for each wind farm and drawn according to a law $\mathcal{N}(0, \beta_w)$.

**Determination of parameters $\alpha_w$ and $\beta_w$**
According to a study of the NREL, a normalised standard deviation $\sigma_{w,global} = 0.13$ was obtained for the total day-ahead wind power forecasting errors of the California Independent System Operator (CAISO) region [32]. The normalised standard deviation was obtained by calculating the standard deviation and dividing it by the system maximum capacity. Another study indicates a variance of 60.4MW$^2$ for a typical 30 MW-capacity wind farm from the western wind and solar integration study data set [35]. This leads to a normalised standard deviation $\sigma_{w,local} = \dfrac{\sqrt{60.4}}{30} = 0.259$.

The same approach as for the load is used to estimate the parameters $\alpha_w$ and $\beta_w$ defining the distribution of wind power forecast errors. Nevertheless, the repartition of production among the wind farms is different from one hour to the next. Therefore all the wind farms are considered to produce the same proportion of power for the sake of simplicity.

The variance of the total forecast generation error can be computed as:

$$\begin{aligned}
& \sum_{i=1}^{6} \tilde{g}_{w,i} = \sum_{i=1}^{6} \hat{g}_{w,i}(1 + \epsilon_{\alpha_w} + \epsilon_{\beta_{w,i}}) \\
\Leftrightarrow \quad & \text{var}\left(\sum_{i=1}^{6} \tilde{g}_{w,i}\right) = \text{var}\left(\sum_{i=1}^{6} \hat{g}_{w,i}(\epsilon_{\alpha_w} + \epsilon_{\beta_{w,i}})\right) \\
\Leftrightarrow \quad & \text{var}\left(\sum_{i=1}^{6} \tilde{g}_{w,i}\right) = \left(\sum_{i=1}^{6} \hat{g}_{w,i}\right)^2 \alpha_w^2 + \text{var}\left(\sum_{i=1}^{6} \hat{g}_{w,i}\epsilon_{\beta_{w,i}}\right) \\
\Leftrightarrow \quad & (C_{w,tot})^2\, \sigma_{w,global}^2 = (\hat{g}_{w,tot})^2\, \alpha_w^2 + (\hat{g}_{w,tot})^2 \left(\frac{1}{6}\right)^2 \beta_w^2 \\
\Leftrightarrow \quad & (\hat{g}_{w,tot})^2\, \sigma_{w,global}^2 \simeq (\hat{g}_{w,tot})^2\, \alpha_w^2 + (\hat{g}_{w,tot})^2 \left(\frac{1}{6}\right)^2 \beta_w^2 \\
\Leftrightarrow \quad & \sigma_{w,global}^2 \simeq \alpha_w^2 + \frac{1}{36}\beta_w^2
\end{aligned}$$

Note that $\sigma_{w,global}$ is defined as the standard deviation divided by the capacity of the system and not the forecast production. In this computation, it is considered that $\hat{g}_{w,tot} \simeq C_{w,tot}$ where $C_{w,tot}$ is the maximum wind power generation. The reason is that $\hat{g}_{w,tot}$ is different for each hour and it is a way to have a solution independent of the predicted value.

The variance of the forecast generation error for one wind farm is:

$$\sigma_{w,local}^2 = \alpha_w^2 + \beta_w^2.$$

It is the same reasoning as for the load and the approximation $\hat{g}_{w,i} = C_{w,i}$ where $C_{w,i}$ is the capacity of wind farm $i$ is used.

The resolution of the system :

$$\begin{cases} \sigma_{w,global}^2 &= \alpha_w^2 + \dfrac{1}{36}\beta_w^2 \\ \sigma_{w,local}^2 &= \alpha_w^2 + \beta_w^2 \end{cases}$$

gives $\alpha_w = 0.124$ and $\beta_w = 0.227$. This is approximated by $\alpha_w = 0.12$ and $\beta_w = 0.23$.

However, it appeared that the approximations performed were not valid. Given $\hat{g}_{w,tot}$ for each hour and still considering that $\hat{g}_{w,i} = \hat{g}_{w,tot}/6$, a less approximated system of equations is:

$$\begin{cases} \sigma_{w,global}^2 &= \dfrac{\hat{g}_{w,tot}}{C_{w,tot}}\left(\alpha_w^2 + \dfrac{1}{36}\beta_w^2\right) \\ \sigma_{w,local}^2 &= \dfrac{\hat{g}_{w,tot}/6}{C_{w,tot}/6}\left(\alpha_w^2 + \beta_w^2\right) = \dfrac{\hat{g}_{w,tot}}{C_{w,tot}}\left(\alpha_w^2 + \beta_w^2\right) \end{cases} \tag{4.4}$$

Solving this system leads to different values of $\alpha_w$ and $\beta_w$ per hour. It gives better results in term of variance but the solution is not general and depends on the predicted wind power for each hour. The values of $\alpha_w$ and $\beta_w$ per hour for the particular day chosen and dependent on the total predicted wind power can be seen in Table 4.4.

Table 4.4: Solutions $\alpha_w$ and $\beta_w$ for the system of equations (4.4) in function of the total forecast wind power per hour.

| Hour | $\hat{g}_{w,tot}$(MWh) | $\alpha_w$ | $\beta_w$ | Hour | $\hat{g}_{w,tot}$(MWh) | $\alpha_w$ | $\beta_w$ |
|---|---|---|---|---|---|---|---|
| 1 | 787.66 | 0.189 | 0.346 | 13 | 787.05 | 0.19 | 0.346 |
| 2 | 746.73 | 0.2 | 0.365 | 14 | 665.01 | 0.224 | 0.4099 |
| 3 | 785.51 | 0.19 | 0.348 | 15 | 473.12 | 0.315 | 0.576 |
| 4 | 856.76 | 0.174 | 0.318 | 16 | 262.14 | 0.569 | 1.04 |
| 5 | 859.02 | 0.174 | 0.317 | 17 | 179.22 | 0.83 | 1.52 |
| 6 | 831.22 | 0.18 | 0.328 | 18 | 111.70 | 1.336 | 2.44 |
| 7 | 812.46 | 0.184 | 0.336 | 19 | 163.91 | 0.9105 | 1.66 |
| 8 | 868.26 | 0.172 | 0.314 | 20 | 166.65 | 0.896 | 1.636 |
| 9 | 895.66 | 0.167 | 0.304 | 21 | 186.54 | 0.8 | 1.461 |
| 10 | 902.24 | 0.165 | 0.302 | 22 | 250.69 | 0.595 | 1.088 |
| 11 | 905.68 | 0.165 | 0.301 | 23 | 200.07 | 0.746 | 1.363 |
| 12 | 858.00 | 0.174 | 0.318 | 24 | 193.22 | 0.77 | 1.411 |

Using this method to generate the samples sometimes leads to negative values or values greater than the maximum capacity. In order to create realistic samples, the wind output power is set to 0 if the value is negative and is equal to the maximum capacity ($C_{w,i}$) if the value is greater than $C_{w,i}$. However this has the effect of decreasing the variance and modifying the mean. The mean value of the realised standard deviation per hour for a set of 24000 samples (1000 per hour) is 0.1256 with a maximum value of 0.1494 and a minimum value of 0.0726. This is close to the value observed in the literature. Note that this set is not the one used in the generation of the database. It is simply a test set.

Concerning the normalised standard deviation for a single wind farm, the maximum value is 0.3900 and the minimum value is 0.0044. The targeted relative standard deviation is 0.259 and the mean value is 0.2095. The mean value is acceptable and thus there is no change to $\alpha_w$ and $\beta_w$ to adapt to the truncation of values.

It is worth noting that the large difference in relative standard deviation for different hours is due not only to the truncation of values but also to the approximation that each wind farm produces the same amount of power. Depending on the hour, this assumption is more or less correct.

The histogram of the total wind production for 1000 realisations can be seen in Figure 4.7 for 2 different hours. It can be observed that the variance is effectively lower in case of a low production. Figure 4.8 shows that a lot of values are set to 0 in case of a low prediction of wind power and similarly, many samples are set to the

maximum capacity in case of high wind power output forecast.

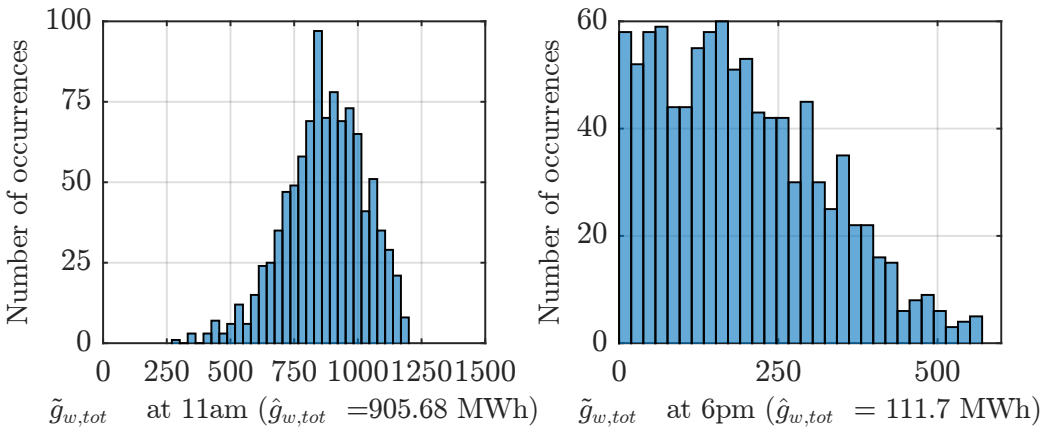



Figure 4.7: Histogram of the total wind power realisations at 11am and 6pm. The relative standard deviations are respectively equal to 0.1323 and 0.1074.

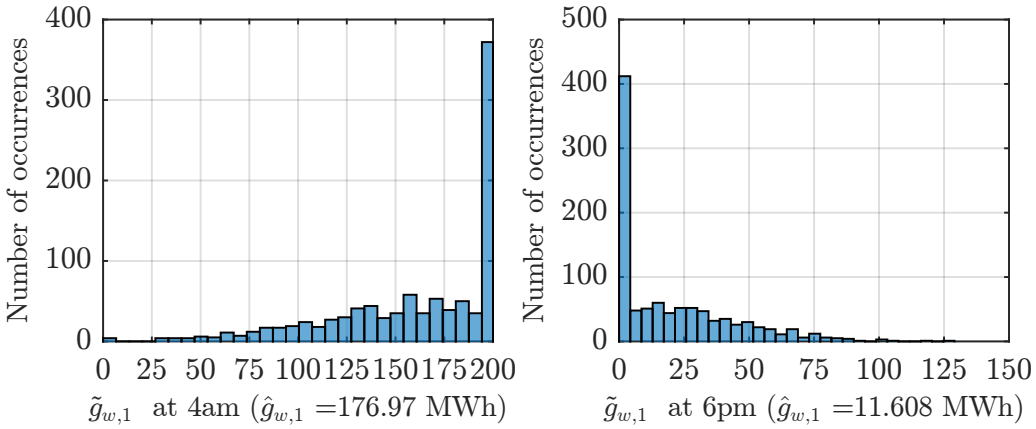



Figure 4.8: Histogram of the wind power realisations at wind farm 1 at respectively 4am and 6pm. The normalised standard deviations are respectively 0.2153 and 0.1087.

Figure 4.9 represents some scatter plots of the relative wind generation forecast error for different wind farms. This error $E^h_{w,i}$is computed as:

$$E^h_{w,i} = \frac{\hat{g}^h_{w,i} - \tilde{g}^h_{w,i}}{\hat{g}^h_{w,i}},$$

where $i = 1, ..., 6$ represents the wind farm index and $h$ the hour. It can be seen in the graphs that the correlation between different wind farms for the same time of the day is low.

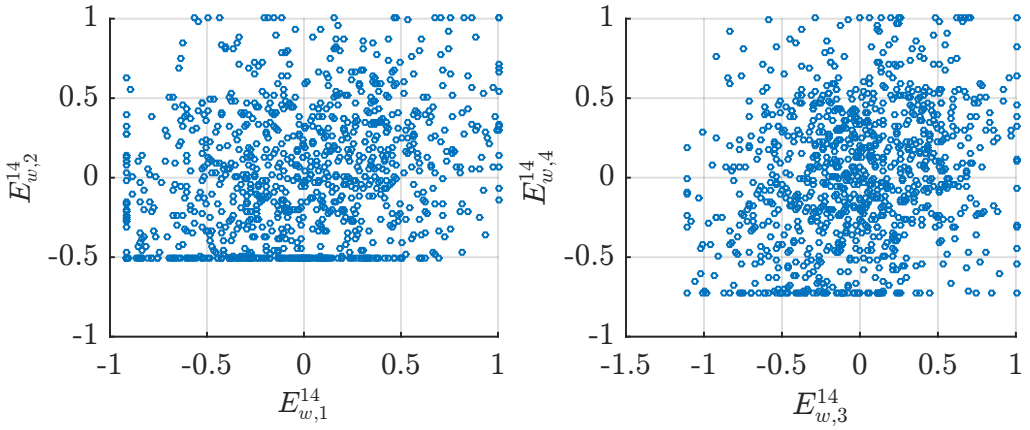



Figure 4.9: Scatter plots of relative forecast error per wind farm at 2pm for 1000 samples.

### 4.3.6 Probabilities of contingencies

The considered set of contingencies is composed of 7 (common mode) double outages and 38 single outages. The contingency $c = 1$ is the no outage case, therefore there are 46 contingencies. The 7 common outages correspond to the lines circled in Figure 3.2.

Given the mean time to failure (MTTF) per line and per common outage, one can compute the probability of occurrence of only one contingency. A MTTF is also known for common outage contingencies.

The probability of having at least one occurrence of contingency $c$ in 1 hour given the MTTF is:

$$P(\text{failure}_c) = 1 - e^{-\lambda_c \Delta t},$$

with $\Delta t = 1\text{h}$ and $\lambda_c = \dfrac{1}{MTTF_c}$ for $c = 2, ..., 46$ for the 38 single outages and the 7 common outages. The Poisson law is again chosen, as for the determination of generating units and transmission lines availability.

The probability of no failure is thus the product of the probabilities for each line to be in service and for each common outages to not occur:

$$P(\text{no failure}) = \prod_{c=2}^{46} (1 - P(\text{failure}_c)). \tag{4.5}$$

We need to compute the probability that only contingency $c$ occurs. It is given by the probability that the failure described by contingency $c$ occurs multiplied by the probability that all the other possible failures do not occur. This gives:

$$P(\text{cont}_c) = P(\text{failure}_c) \cdot \prod_{j=2, j\neq c}^{46} (1 - P(\text{failure}_j)). \tag{4.6}$$

It is worth mentioning that by considering a set of 46 contingencies, a lot of contingencies are neglected, such as the combination of 3 lines out-of-service, 4 lines out-of-service and so on. Two possibilities are considered. Either their criticality is neglected and they are incorporated in the no outage case. Or they are all considered as a unique contingency where their associated criticality could be the worse criticality computed among the 46 studied contingencies and their probability is 1 minus the sum of the probabilities of the 46 studied contingencies. In this work, the first possibility is chosen for simplicity and thus the risk is underestimated. Therefore, the probability of having no outage is not equal to $P(\text{no failure})$. It is slightly increased to have $\sum_{c\in\mathcal{C}} \pi_c = 1$, where $\mathcal{C}$ contains the 46 contingencies previously introduced:

$$P(\text{cont}_1) = 1 - \sum_{c\in\mathcal{C}, c\neq 1} P(\text{cont}_c).$$

The probabilities of contingencies are modified for two reasons: the weather and the availability of the lines.

Concerning the weather, the number of failures per hour $\lambda$ is multiplied by 30 in case of adverse weather, except for the cables (lines 1 and 9) and the 5 transformers.

In case of unavailability of particular lines, the probabilities of outage of these lines are equal to 1. If the unavailable line was part of a double common outage contingency, this has no influence on the probability of the double common outage. This probability is equal to 1 only when both lines are unavailable.

If at least one line is unavailable, the definition of the no outage case should change. In fact, the probability of no outage is 0 since some factors in equation (4.5) are equal to 0. Thus a redefinition of the no failure probability is needed: $P(\text{no failure})$ is the probability to have no failure of the available lines or available common outages. If $\mathcal{C}_{available}$ is the set of contingencies that represent the available common outages and the failure of the available lines and transformers , $P(\text{no failure})$ is computed as:

$$P(\text{no failure}) = \prod_{c\in\mathcal{C}_{available}} (1 - P(\text{failure}_c)).$$

Note that the contingency 1, corresponding to the no outage case, is not included in $\mathcal{C}_{available}$. For each $c \in \mathcal{C}_{available}$, the probability for only contingency $c$ to occur $P(\text{cont}_c)$ is also slightly modified compared to equation (4.6):

$$P(\text{cont}_c) = P(\text{failure}_c) \cdot \prod_{j\in\mathcal{C}_{available}, j\neq c} (1 - P(\text{failure}_j)).$$

The contingencies that are not included in the set $\mathcal{C}_{available}$ describe the failure of the same lines, thus the same system topology. In that case, to avoid taking into account the same expected criticality several times, the probabilities of the redundant contingencies are set to 0. Thus,

$$P(\text{cont}_c) = 0, \forall c \notin \mathcal{C}_{available} \;\&\; c \neq 1.$$

For instance, if the line 1 is unavailable, contingency 1 (no outage case) and 2 (fault on line 1) describe the same system topology, that is a system without line 1. The probability of contingency 2 is thus set to 0.

Finally, the probability of the no outage case, defined here as no failure of available lines and no available common outages is

$$P(\text{cont}_1) = 1 - \sum_{c \in \mathcal{C}_{available}} P(\text{cont}_c).$$

## 4.4 Meta-parameters

In the RMAC, there are three meta-parameters: $\epsilon$, $\Delta E$ and $X_a$. $\epsilon$ is the threshold for the reliability target, $\Delta E$ is a threshold for the residual risk and $X_a$ defines the set of acceptable trajectories. In this work, we only modify $\epsilon$ and $\Delta E$.

This section describes how these two parameters are determined for each sample. We first explain how $\epsilon$ is sampled and then how $\Delta E$ is computed based on the risk of a "pseudo N-1" criterion. Next we discuss the notion of infeasible cases and explain how these infeasible cases are handled in our simulations. Finally, we wrap up by presenting the complete procedure incorporating all these steps.

### 4.4.1 Sampling the value of $\epsilon$

In order to study the impact of $\epsilon$ on the outputs, $\epsilon$ is defined randomly for each state. The distribution is chosen such as the values of $\epsilon$ are uniformly distributed on a logarithmic scale between $[10^{a_\epsilon}, 10^{b_\epsilon}]$. The value of $a_\epsilon$ is determined by assigning a fixed value to $\epsilon$ and checking the number of infeasible states (see Sections 4.4.2 and 4.4.3). The first value for which there are not too many infeasible samples will be used to determine the minimum value of $\epsilon$. The value of $b_\epsilon$ is determined by setting $\epsilon = 1$ for a certain number of states and looking for the maximum value of the probability of unacceptable states. Indeed, when $\epsilon = 1$, there is no constraint on $P_{unacceptable}$.

In order to count the number of infeasible states for a given $\epsilon$, the set of considered contingencies is fixed and contains only the N-1 contingencies. The results for various values of $\epsilon$ can be seen in Table 4.5. Despite the high number of infeasible samples for $\epsilon = 10^{-5}$ and $\epsilon = 10^{-6}$, the minimum value is set to $\epsilon = 10^{-5}$. This is a compromise between a low number of infeasible states and the study of results with a smaller value of probability of unacceptable behaviours.

Considering a set of samples of size 1200 with $\epsilon = 1$ and a set of considered contingencies containing only the single line outages, the value 0.0119 was found to be the largest value of the probability of unacceptable states. Thus the variable $b_\epsilon$ is chosen to be equal to -1.

The value of $\epsilon$ is interesting to study on a logarithmic scale (base 10), it is why it was chosen to select a uniform distribution on a logarithmic scale. The formula

Table 4.5: Number of feasible and infeasible samples for different values of $\epsilon$. $Nb_{\text{feasible}}$ is the number of samples that are feasible, $Nb_{\text{infeasible}}$ is the number of infeasible samples and $Nb_{\text{total}}$ is the total number of samples for a particular value of $\epsilon$. The states are generated randomly and are all different.

| $\epsilon$ | $Nb_{\text{feasible}}$ | $Nb_{\text{infeasible}}$ | $\frac{Nb_{\text{infeasible}}}{Nb_{\text{total}}}$ |
|---|---|---|---|
| $10^{-6}$ | 19 | 221 | 92.1% |
| $10^{-5}$ | 24 | 216 | 90% |
| $10^{-4}$ | 86 | 154 | 64.2% |
| $10^{-3}$ | 152 | 88 | 36.7% |

is:

$$\epsilon = 10^{a_\epsilon + (b_\epsilon - a_\epsilon)\cdot \text{rand}(0,1)}$$

where $a_\epsilon = -5$, $b_\epsilon = -1$ and $\text{rand}(0,1)$ is a function that outputs a pseudo-random number uniformly distributed between 0 and 1.

### 4.4.2 Computing the value of $\Delta E$

The value of $\Delta E$ is more difficult to fix. In order to avoid choosing it completely arbitrarily, one possibility is to apply the N-1 criterion when computing the control actions and then calculating the residual risk associated with this criterion.

The deterministic N-1 criterion corresponds to a version of the RMAC using a fixed set of considered contingencies $\mathcal{C}_{N-1}$ that contains only the single outage contingencies, an $\epsilon$ equal to 0 and a probability of failure of corrective actions also equal to 0 [3].

However, when we used this deterministic N-1 criterion with our samples, 82% of them led to an infeasible optimisation problem at this stage. Rather than rejecting such infeasible cases from the database, we thus decided to implement a modified version of the N-1 criterion (to distinguish it we will use the term "pseudo N-1"). This "pseudo N-1" SCOPF problem merely consists in relaxing the acceptability constraint in the N-1 calculation by using the randomly generated value of $\epsilon$, instead of $\epsilon = 0$ and also considering the probability of failure of corrective actions. This allows us to study the impact of considering a fixed set of contingencies instead of choosing a set of contingencies dependent on the state of the system when computing $\Delta E$.

Therefore $\Delta E$ is equal to the residual risk $R_{\mathcal{C}\setminus\mathcal{C}_{N-1}}$ computed from the control decisions taken according to the "pseudo N-1" criterion. $R_{\mathcal{C}\setminus\mathcal{C}_{N-1}}$ is thus the sum of the expected criticalities of common outage contingencies.

### 4.4.3 Handling the remaining infeasible cases

Even with the above modified version of the N-1 criterion, the resulting optimisation problem may still be infeasible. Therefore, in case a sample leads to an infeasible

optimisation problem at the previous stage, the problem will be further relaxed in order to enable the calculation of the value of $\Delta E$.

Rather than trying to determine the minimal increase of $\epsilon$ to make the problem feasible, we decided to merely set its value to 1 in order to make sure at once that the "pseudo N-1" criterion will become feasible. Moreover, if at this stage, when calculating the value of $\Delta E$ with the "pseudo" N-1 criterion, $\epsilon$ is set to 1, the same value of 1 is also used later on with the RMAC, so as to ease the comparison of the two approaches.

Finally, while the values of $\epsilon$ thus used in the RMAC always correspond to a feasible "pseudo N-1" criterion, there is no guarantee that the RMAC itself will remain feasible. It turns out that indeed occasionally the RMAC problem is infeasible. We decided to exclude these samples from our final database.

In the dataset, 36% of the samples are infeasible with the "pseudo" N-1 criterion. Considering only the feasible cases, 0.3% of them are RMAC infeasible.

### 4.4.4 Complete procedure

The complete structure of the program incorporating the calculation of $\Delta E$ and the possible relaxation of $\epsilon$ via the "pseudo" N-1 approach is represented in Figure 4.10.

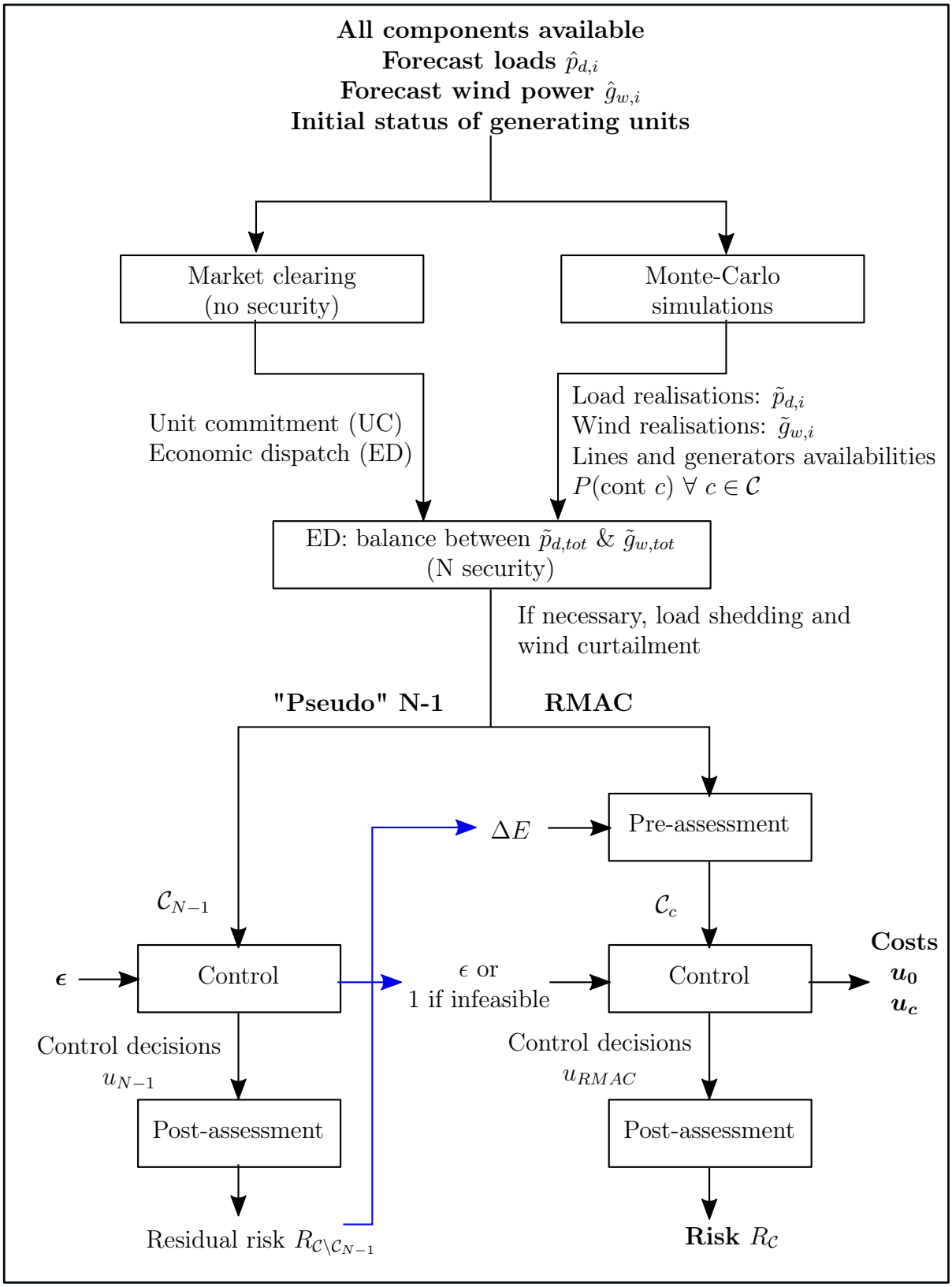



Figure 4.10: Schematic representation of the generation of samples for both the "pseudo N-1" criterion and the RMAC. The residual risk of the "pseudo N-1" criterion is used to define $\Delta E$ for the RMAC. If the sample was infeasible with the "pseudo N-1" criterion, $\epsilon_{N-1}$ is set to 1 and thus $\epsilon_{RMAC}$ is also set to 1.

# Chapter 5

# Database analysis

The generated database contains 4623 samples. This strange number is due to the discarding of some samples for various reasons. First, we discarded the samples that were feasible with the "pseudo" N-1 criterion but not with the RMAC. Other samples were infeasible at the N-security rescheduling step. Indeed if a bus connected only to generating units is isolated, the only way to balance the node is to trip the generators. This is not implemented in the rescheduling step and thus the samples had to be discarded. Finally, this non-logical number of samples is also due to some computer problems.

This chapter contains a description of the database and the analyses of some results. The first section describes the different variables of the problem. The next sections contain analyses of the results: Section 2 describes the topology of the system; Section 3 the distribution of the meta-parameters $\Delta E$ and $\epsilon$ and the proportion of infeasible values of $\epsilon$; Sections 4 and 5 show respectively the impact of the control actions and of the weather on the risk and on the residual risk; Section 6 compares the "pseudo N-1" criterion with the RMAC. Section 7 describes the macroscopic variables introduced in order to study more generally the preventive and corrective control actions. Finally, the last section recalls the output variables on which learning algorithms will be applied.

## 5.1 Variable description

This section enumerates and describes the different variables used in the subsequent chapters when applying supervised learning methods. The first subsection describes the input features and the second subsection the target output variables.

### 5.1.1 Input features

There are 184 features, which are enumerated hereafter (notice that the names we will use in our analyses to refer to these variable are written within parentheses):

- the hour of the day (*hour*),
- the 17 bus loads (*load i*),

- the 6 wind farm power outputs (*wind i*),
- the 33 generating units' availability (*generator i status*),
- the 33 transmission lines' availability (*line i status*),
- the 5 transformers' availability ($transformer\ i\ status$),
- the 46 probabilities of contingencies ($probability\ of\ contingency\ i$),
- the value of $\epsilon$ (*epsilon*),
- the value of $\Delta E$ (*Delta E*),
- the weather status (normal or adverse) (*weather status*),
- the system load (*total load*),
- the total wind power output (*total wind power*),
- the system load minus the total wind production ($load - wind\ production$),
- the number of unavailable transmission lines ($number\ of\ lines\ unavailable$),
- the number of unavailable transformers ($number\ of\ transformers\ unavailable$),
- the number of unavailable generators ($number\ of\ generators\ unavailable$),
- the market power output of each generator ($market\ power\ for\ gen\ i$),
- the minimum production of the system without considering wind power (*min production wo wind*),
- the maximum production of the system without considering wind power (*max production wo wind*).

### 5.1.2 Target output variables

Some interesting outputs of the control program are the total cost, the preventive control cost, the expected corrective control cost, the probability of being in an unacceptable state and the risk associated to the relaxed contingencies and the possible failure of corrective actions. The total cost is the sum of the preventive control cost, the expected corrective control cost and the risk. This risk will be called expected criticality in the rest of the report in order to avoid confusing it with the risk $R_{\mathcal{C}}$ computed with the assessment program. They are two differences in the evaluation of both risks. The first difference is that to estimate the expected criticality, it is assumed that if there is a violation of the thermal limits, all the load is lost. On the contrary, the amount of lost load is estimated through cascade failure simulations for the risk $R_{\mathcal{C}}$. The second difference is the set of contingencies used. The expected criticality and the risk are assessed using respectively the set $\mathcal{C}_c$ and the set $\mathcal{C}$.

Other outputs of the SCOPF program are the preventive re-dispatch of generating units, the corrective re-dispatch and the corrective adjustments of phase-shifting

transformers.

Finally, the risk $R_{\mathcal{C}}$ output from the assessment program allows the evaluation of the reliability of the system after control.

Since there are 33 generating units and 5 phase-shifting transformers in the system, the total number of outputs is 1749. As this number is large, we will try to determine more macroscopic variables accounting for the preventive and corrective actions. This is discussed in Section 5.8.

## 5.2 Load and wind power

This section shows the distribution of the total load, the total wind power generation and the net load that is defined as the total load minus the total wind generation. These distributions can be seen respectively in Figures 5.1, 5.2 and 5.3.

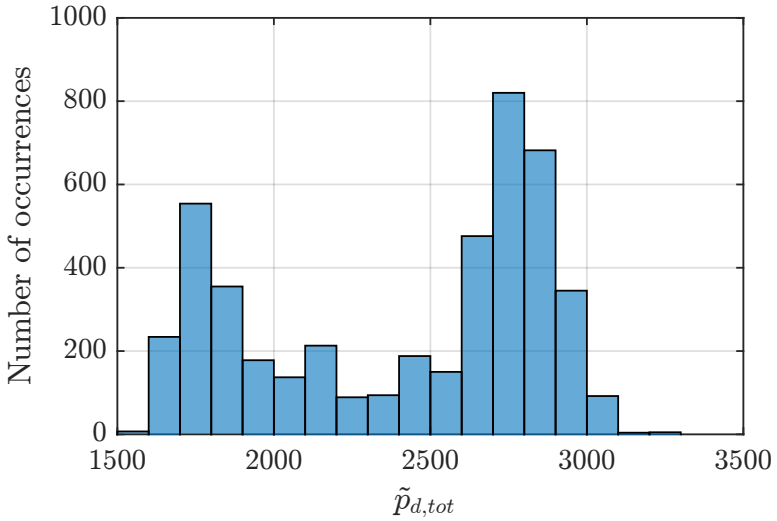



Figure 5.1: Histogram of total system load.

In each histogram, one can distinguish a bimodal distribution. This is more pronounced in the load distribution than in the wind generation distribution. This is due to the load and wind forecasts. Indeed, in Figures 4.2 and 4.3, two levels of load and wind generation can be observed. The high level of wind generation is about 800MWh whereas the low level is around 200MWh. These values correspond to both peaks in Figure 5.2. Similarly, the high level of total load is approximately 2800MWh and the low level of total load is around 1800MWh and these two values correspond to the peaks in Figure 5.1.

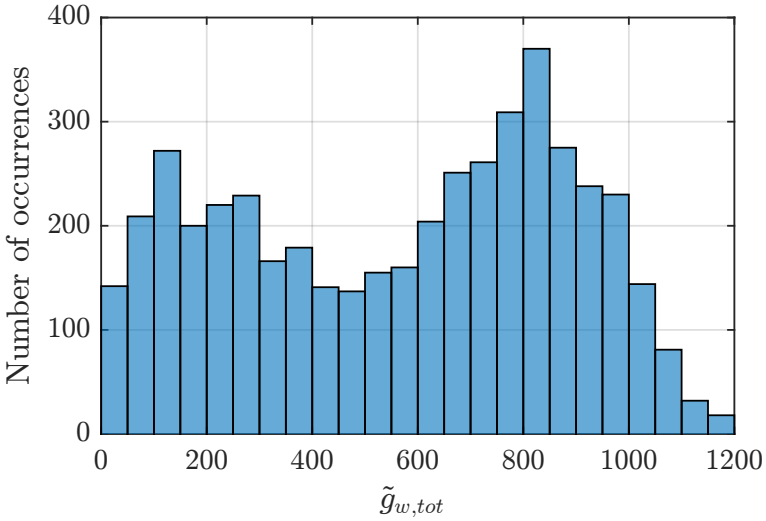



Figure 5.2: Histogram of total wind generation.

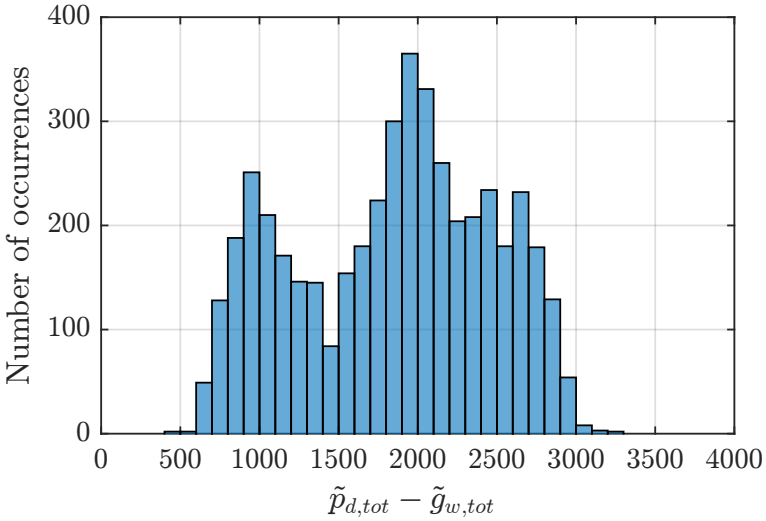



Figure 5.3: Histogram of the system net load.

## 5.3 Topology of the system

### 5.3.1 Generating units

The probability for a generating unit to be unavailable is in the order of $10^{-2}$. Thus one can expect to have many samples with at least one generator out-of-service. In Figure 5.4, the histogram of the number of generators unavailable per sample is represented. One can notice that there are at most 5 generators out-of-service at the same time. Furthermore, almost half the samples have at least one generating unit unavailable.

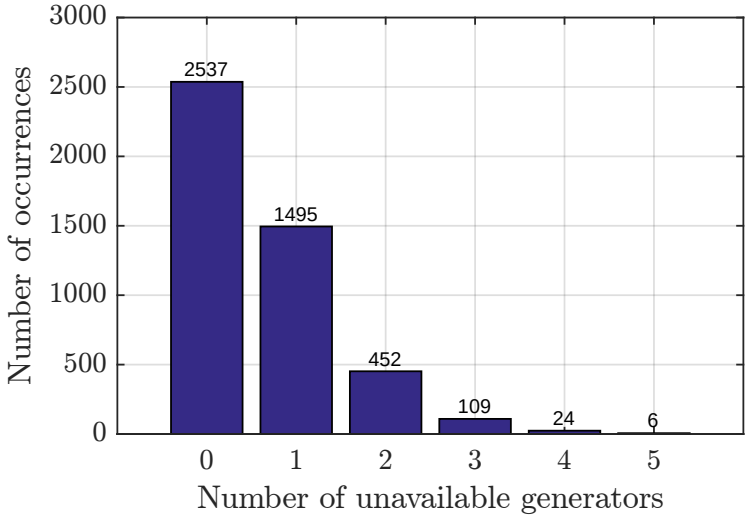



Figure 5.4: Histogram of the number of unavailable generating units per observation.

### 5.3.2 Transmission lines

The probability of unavailability of a transmission line depends on the weather. It is thus interesting to study the availability of lines in the samples as a function of the weather status. In Figure 5.5, a bar graph of the number of unavailable lines per sample as a function of the weather status is represented.

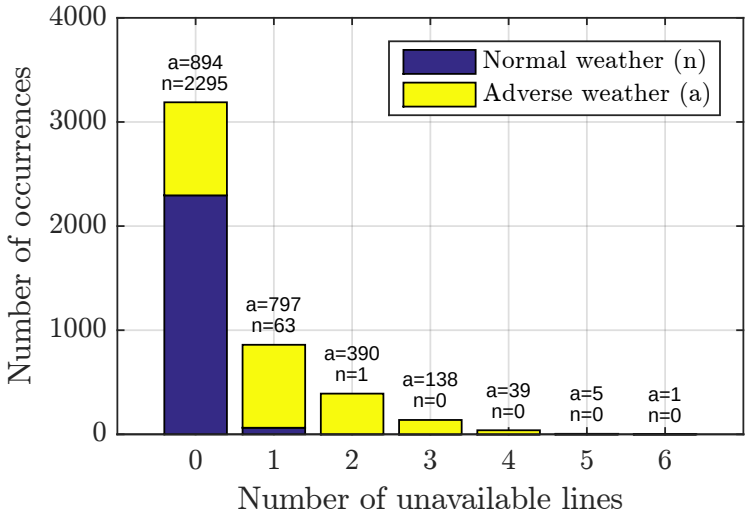



Figure 5.5: Stacked bar graph of the number of out-of-service lines per sample in case of normal and adverse weather.

For a normal weather, the probability of unavailability of a line is small (about $10^{-4}$) and thus the maximum number of out-of-service lines is two and it is the case for only one sample. Most samples correspond to the complete topology with all the

transmission lines available.

Concerning the adverse weather state, more samples have at least one line unavailable and there are a maximum of 6 transmission lines out-of-service at the same time. It is due to the fact that the probability of unavailability of a line is approximately one hundred times greater in case of adverse weather compared to a normal weather.

Note that there are about 51% of samples with a normal weather in the database.

### 5.3.3 Phase-shifting transformers

In the dataset, only one sample has a transformer unavailable. Indeed, the probability of outage of a transformer is small (around $10^{-5}$) and is independent of the weather. There are also a fewer number of transformers.

## 5.4 Meta-parameters $\epsilon$ and $\Delta E$

### 5.4.1 $\epsilon$

The value of $\epsilon$ is randomly drawn according to a uniform distribution on a logarithmic scale. When it has a value 1, it means that with the drawn $\epsilon$ the case was infeasible. Figure 5.6 represents the stacked histogram of $\log_{10}\epsilon$ for feasible and infeasible samples. It can be seen that about one third of the samples were infeasible with the "pseudo" N-1 criterion. The distribution of the values of $\epsilon$ for these infeasible samples, before this parameter was set to 1, is represented in yellow in the graph. One can verify that the drawn values of $\epsilon$ are uniformly distributed on a logarithmic scale between $10^{-5}$ and $10^{-1}$.

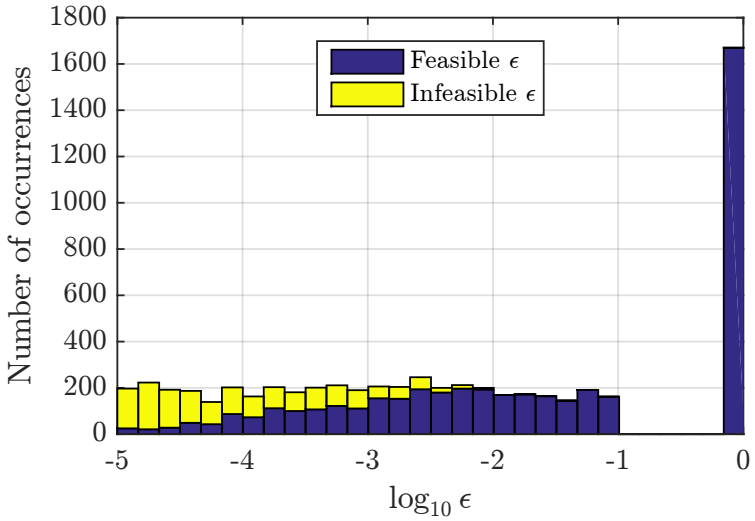



Figure 5.6: Stacked histogram of values of $\log_{10}\epsilon$ in the dataset.

For information, Figure 5.7 shows a scatter plot of the probability of being in an unacceptable state vs. the value of $\epsilon$. The logarithmic scale is used to better show the values. However, the 92 samples with a probability of unacceptable states equal to 0 have been discarded. It can be seen that $P_{unacceptable} \leq \epsilon$, thus the reliability target is met. In this graph, two markers are used to distinguish the weather status of the samples. One can observe that the probability of unacceptable states is in average larger for adverse weather.

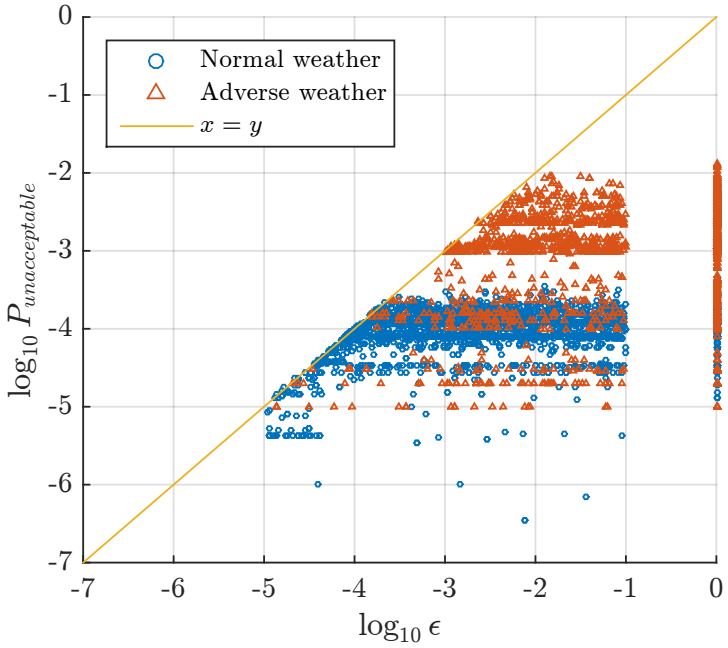



Figure 5.7: Values of $\log_{10} P_{unacceptable}$ vs. values of $\log_{10} \epsilon$. The 92 samples with a zero-value for the probability of unacceptable states have been discarded.

### 5.4.2 $\Delta E$

As said previously, the value of $\Delta E$ is obtained by finding the residual risk using the N-1 approach for each sample. The histogram of this parameter can be seen in Figure 5.8. The values of $\Delta E$ are comprised between 0 and 7579.5$. In order to check that the set of considered contingencies $\mathcal{C}_c$ meets the constraint $R_{\mathcal{C}\setminus\mathcal{C}_c} \leq \Delta E$, a scatter plot of the value of $R_{\mathcal{C}\setminus\mathcal{C}_c}$ before control vs. the value of $\Delta E$ is shown in Figure 5.9. The discarding principle is effectively verified before control.

### 5.4.3 Infeasible cases

The proportion of the original values of $\epsilon$ that led to infeasibility of the "pseudo N-1" criterion is quite high in the dataset. $\epsilon$ had indeed to be set to 1 for about 36% of the samples. In Figure 5.6, if we look at the distribution of the infeasible $\epsilon$ values, it can be seen that the smaller $\epsilon$, the larger the number of infeasible cases.

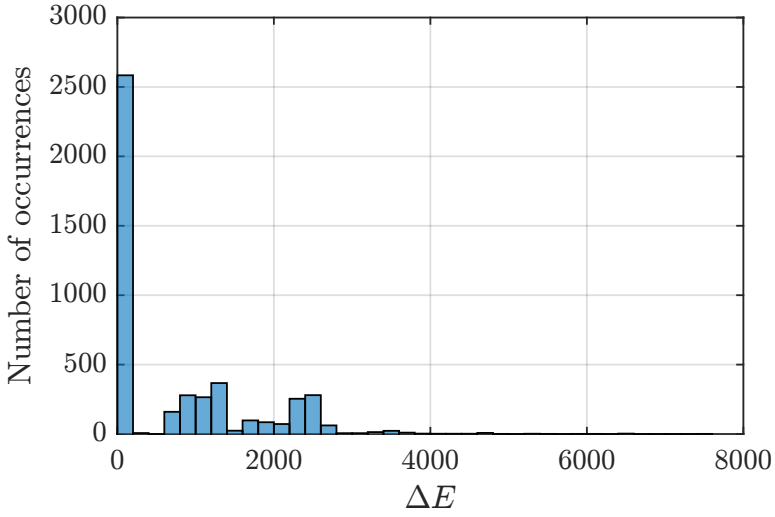



Figure 5.8: Histogram of values of $\Delta E$ in the dataset.

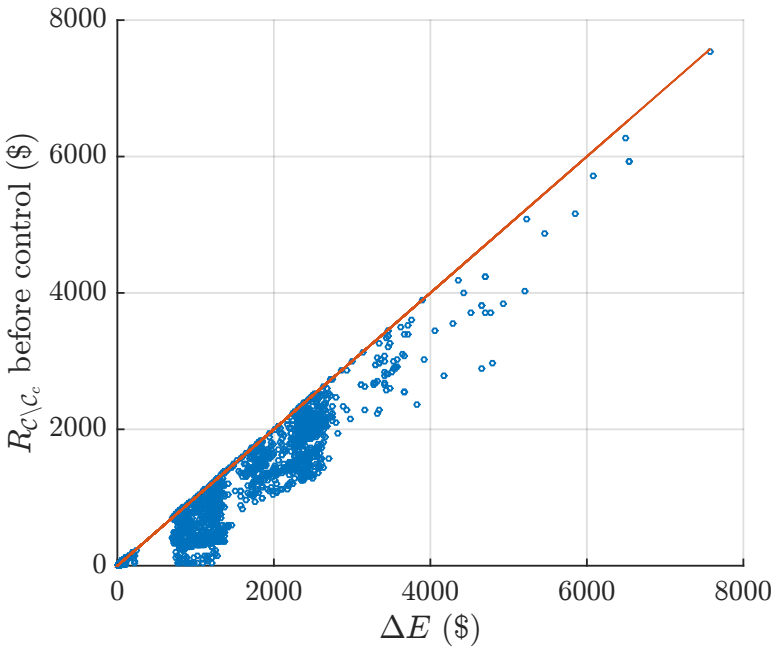



Figure 5.9: Values of $R_{\mathcal{C}\setminus\mathcal{C}_c}$ before control vs. values of $\Delta E$.

Another cause of infeasibility can be the increase of the probabilities of contingencies. It can be seen in Figure 5.10 that the weather status of most infeasible samples is adverse.

Given the way the database is generated, most infeasible samples were infeasible with the "pseudo" N-1 criterion. However 9 samples were feasible with the N-1 criterion and $\epsilon < 1$ but infeasible with the RMAC. Even if they have been discarded, it is interesting to understand why they were infeasible.

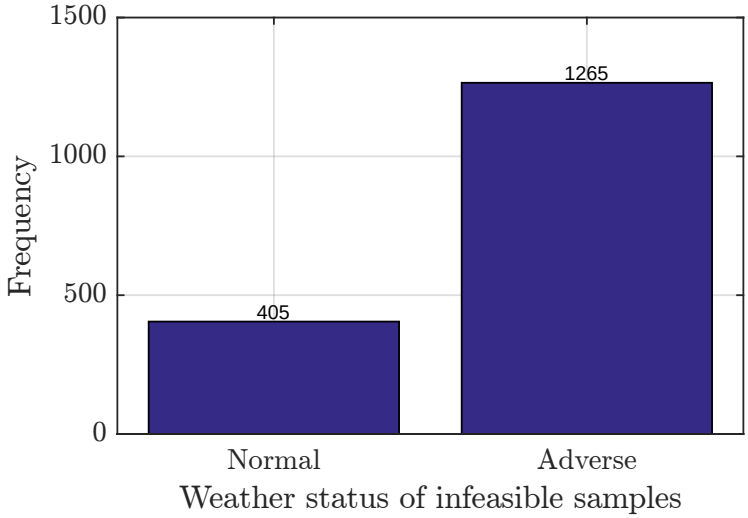



Figure 5.10: Histogram of the weather status of infeasible samples.

In Figure 5.11, two particular system's topologies leading to an RMAC infeasible case are represented on the same schematic. In the first topology, line 17 is unavailable and in the second, line 10 is unavailable.

If line 17 is unavailable and contingency 41 (in blue in Figure 5.11) is a considered contingency, the bus 13 is isolated. In one RMAC infeasible sample of the database, the load on that bus is smaller than the minimum generation. Therefore the only way to make the problem feasible is to relax contingency 41. However, in that case, relaxing contingency 41 increases the probability of being in an unacceptable state beyond the chosen value of $\epsilon$. As a consequence, the sample is RMAC infeasible. As a reminder, a relaxed contingency is a considered contingency for which no preventive or corrective decisions are taken.

Another "RMAC infeasible" case corresponds to the topology where line 10 is unavailable. For this sample, contingency 40 (in green in Figure 5.11) is considered and it is impossible to balance bus 8 since it is isolated and only connected to a load centre. Thus contingency 40 should be relaxed for the problem to be feasible, again at the price of increasing the probability of unacceptable states.

In both cases, it is impossible to use preventive and corrective actions to handle double outage contingencies, leading to an infeasible RMAC.

## 5.5 Assessment before and after control

This section studies the impact of control on the risk. The first subsection compares the total risk before and after control while the second subsection looks at the residual risk and analyses if it changes after control.

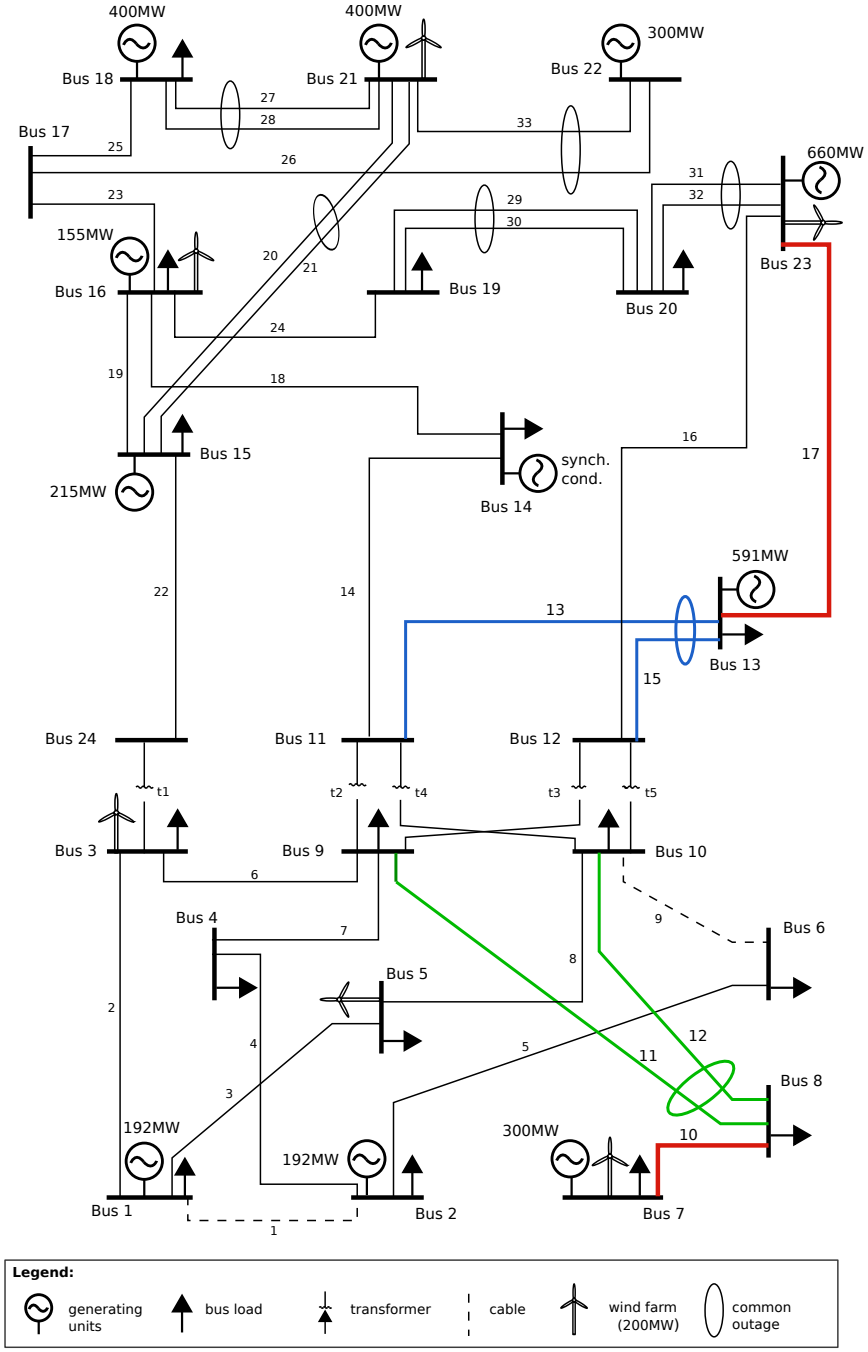



Figure 5.11: Representation of the modified IEEE-RTS96 system where two configurations of the system leading to an RMAC infeasible case are highlighted. In the first configuration, line 17 is unavailable while in the second configuration, line 10 is unavailable. The unavailable lines are represented in red. The double outages contingency that must be relaxed is represented in blue for the first configuration and in green for the second configuration. They correspond respectively to the common mode double outage of lines 13 and 15 and the common mode double outage of lines 11 and 12.

### 5.5.1 Risk

Figure 5.12 shows the histograms of the risk before and the risk after control as well as their differences. If everything works well the risk before control actions should be greater than the risk after control actions. It can be seen in the lower graph that it is the case for almost all samples. The difference between the risk before and the risk after is mostly positive and the range of values of the risk before control is greater. However, in the database, 11 samples have a risk that is increased with the control actions. It represents 0.2% of the cases. In these cases, the risk associated with the considered contingencies decreases as expected. However, the risk associated with the contingencies $c \in \mathcal{C} \backslash \mathcal{C}_c$ rises, causing the total risk to be greater after control than before.

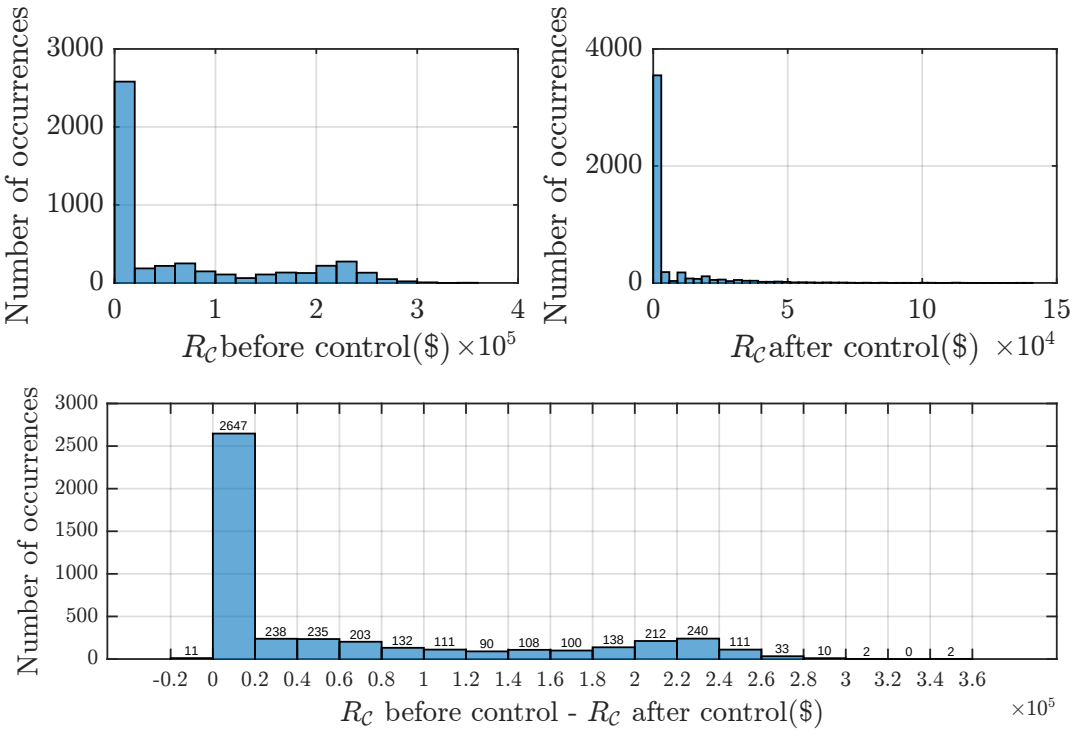



Figure 5.12: Histograms of values of risk before control and risk after control in the dataset. The graph in the bottom represents the difference between the risk before control and the risk after control.

On average, the risk decreases with control by an amount of $6.05 \cdot 10^4$\$. If we put in relation the decrease in risk $R_{\mathcal{C}}$ due to control and the total cost for each sample (Figure 5.13), it can be seen that few correlations exist between both.

### 5.5.2 Residual risk

About half the samples have a residual risk that is identical before and after control. This means that the preventive re-dispatch does not affect the criticality of discarded contingencies in these cases.

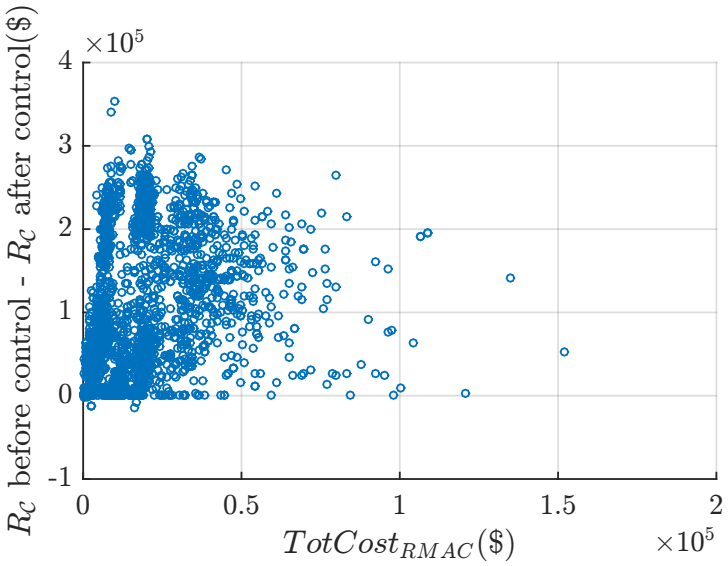



Figure 5.13: Decrease in risk $R_\mathcal{C}$ due to application of control actions vs. total cost.

For the samples where the residual risk changes, it is interesting to check if the residual risk increases or decreases after control. For 38% of the samples, it decreases by an average of 230.16 $. For the remaining 13%, it rises by an average of 4331$. The small amount of samples with an increasing residual risk indicates that for most of the samples, no iterative approach would be needed to satisfy the discarding principle, *i.e.* $R_{\mathcal{C}\setminus\mathcal{C}_c} \leq \Delta E$.

The distributions of the residual risks before and after control as well as their differences can be seen in Figure 5.14.

## 5.6 Weather influence

In this section, the influence of the weather on some results is studied. We recall that each realisation is normally present twice in the dataset, once for a normal weather and once for an adverse weather. However, due to some computers problems and the discarded samples, only 1513 realisations of load and wind generation are present twice. The analysis of the weather influence is performed with these 3026 samples.

The only differences between a normal weather sample and its adverse weather counterpart are the probabilities of contingencies and the availability of the transmission lines. Indeed, the availability of a line is randomly drawn based on the MTTF and this variable is weather-dependent.

### 5.6.1 Control costs in function of the weather

Figure 5.15 shows four histograms. Each one represents the distribution of the differences between the adverse weather costs and the normal weather costs. As one could expect, the control cost is usually greater in case of adverse weather. The greatest differences can be found with the expected criticality. Indeed, this variable

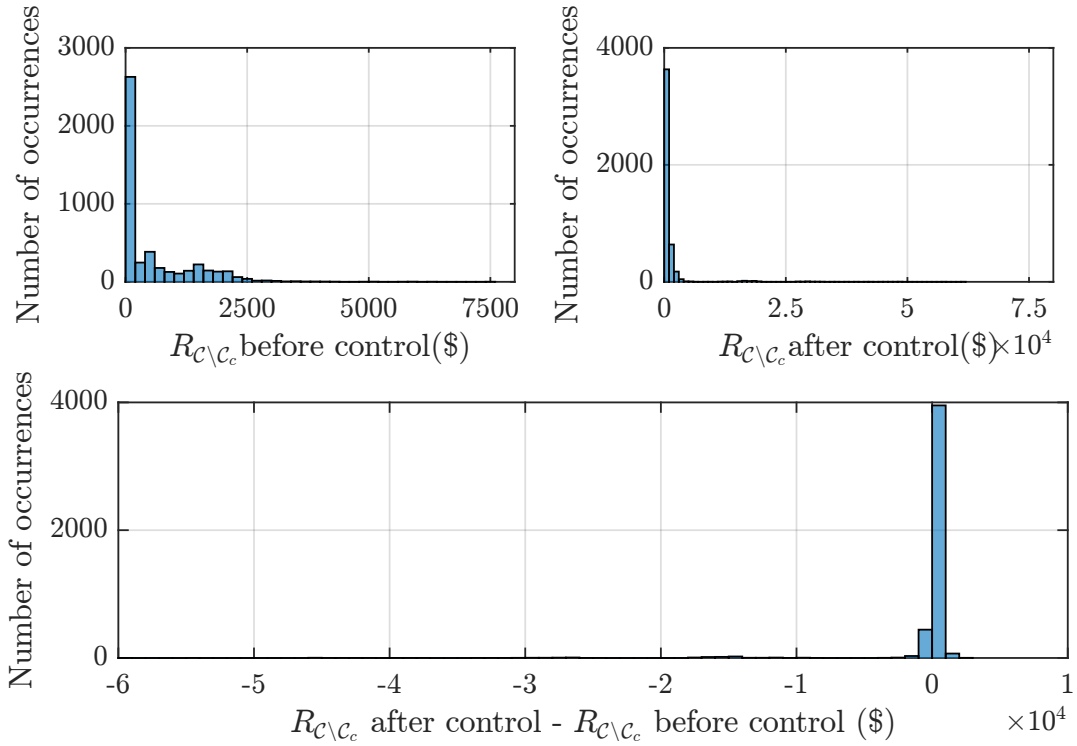



Figure 5.14: Histograms of values of residual risk before control and residual risk after control in the dataset. The graph in the bottom represents the difference between the residual risk before control and the residual risk after control.

is directly impacted by the weather since the severity is weighted with the contingencies' probabilities.

It is also probable that the increase in contingencies' probabilities leads to more preventive actions. Indeed, the cost of corrective actions and the probability of being in an unacceptable state are proportional to the probabilities of contingencies. Therefore, in case of adverse weather, their cost rises up, possibly making preventive control more advantageous.

### 5.6.2 Risks in function of the weather

In order to compare the risk before control and the risk after control for a normal and for an adverse weather, the distributions of $R_{\mathcal{C},adv} - R_{\mathcal{C},nor}$ for both risks are plotted in Figure 5.16. $R_{\mathcal{C},adv}$ and $R_{\mathcal{C},nor}$ are respectively the risk in case of adverse weather and the risk in case of normal weather. As expected, both risks are greater when the weather is adverse than when the weather is normal for most samples. The average differences between adverse and normal weather for the risks before control and after control are respectively $1.26 \cdot 10^5$\$ and 9978.3\$.

In about 0.4% of the cases, the risk before control is higher for a normal weather than for an adverse weather. One possible reason is that load shedding may be performed at the rebalance step in case of outage of lines due to adverse weather.

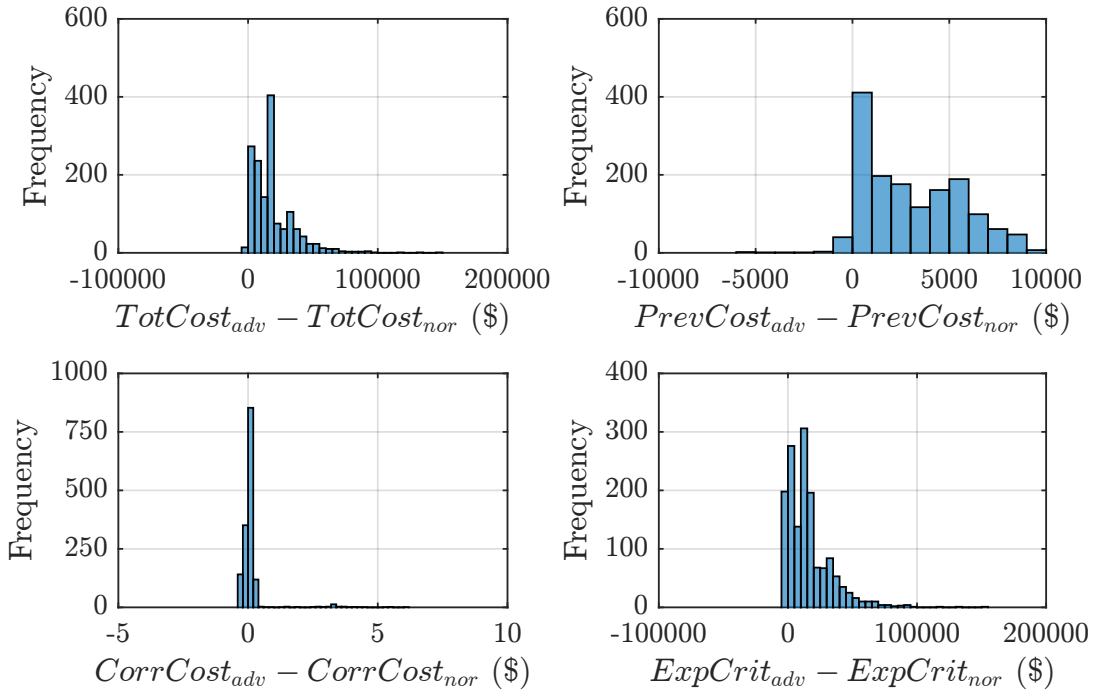



Figure 5.15: Histograms of the differences between control costs during an adverse weather and control costs during a normal weather. They represent respectively, from left to right and top to bottom, the difference in total cost, preventive cost, corrective cost and expected criticality.

This would decrease the system load and lead to a more robust system. The load lost at the rebalance step is not considered when computing the risk.

About 6% of the samples have a risk after control larger for a normal weather than for an adverse weather. Among these samples, less than 10% had a risk before control larger for the normal weather. Thus there is another reason for the risk with a normal weather to be higher. In some cases, the increase in contingencies' probabilities and the loss of one or more lines force the use of more preventive actions. This results in a more reliable system at a higher cost.

## 5.7 Comparison between the "pseudo" N-1 criterion and the RMAC

For each sample, the control program was applied twice, once according to the "pseudo" N-1 criterion and once according to the RMAC. The only difference between both criteria is the set of considered contingencies. This section studies the impact of the RMAC or N-1 criterion on the outputs of the control and assessment blocks.

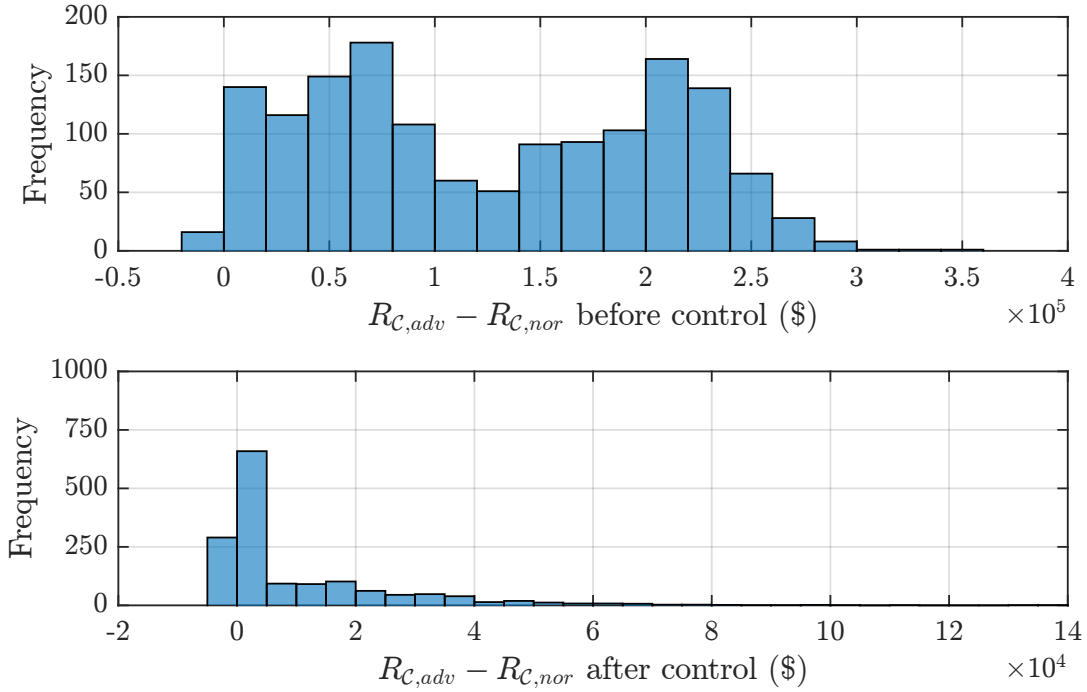



Figure 5.16: The histogram on the top represents the difference between the risk before control computed for an adverse weather and the risk for a normal weather and the histogram on the bottom represents the difference in risk after control for a normal and an adverse weather.

### 5.7.1 Control

#### Total cost

As already mentioned, the total cost is the sum of the preventive control cost, the expected corrective control cost and the expected criticality. In the dataset, 77% of the samples have a total cost larger with the N-1 approach than with the RMAC.

0.15% of the samples have a total cost equal to 0 in case of RMAC and none sample have a zero cost with the N-1 approach. A cost of 0 may happen if the value of $\Delta E$ is such that the set $\mathcal{C}_c$ contains only the contingency 1, corresponding to the no outage case.

Over 4623 samples, 68 have an identical total cost with both approaches.

In Figure 5.17, the histogram of the differences between the N-1 total cost and the RMAC control cost is represented.

#### Preventive control cost

The preventive control cost is greater with the N-1 criterion for 48% of the states. A quarter of the samples have a preventive cost equal to 0 with the RMAC criterion and

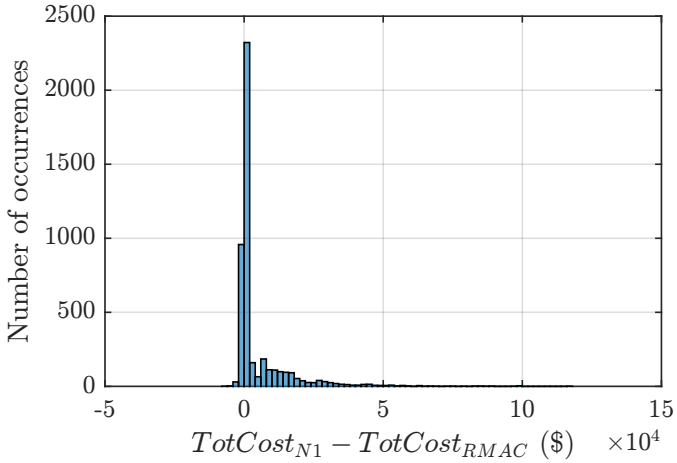



Figure 5.17: Histogram representing the difference between the total cost obtained with the N-1 criterion and the one obtained with the RMAC.

only 14% of the samples have no preventive actions with the "pseudo" N-1 criterion. The histogram of the differences between preventive control costs computed with both approaches can be seen in Figure 5.18.

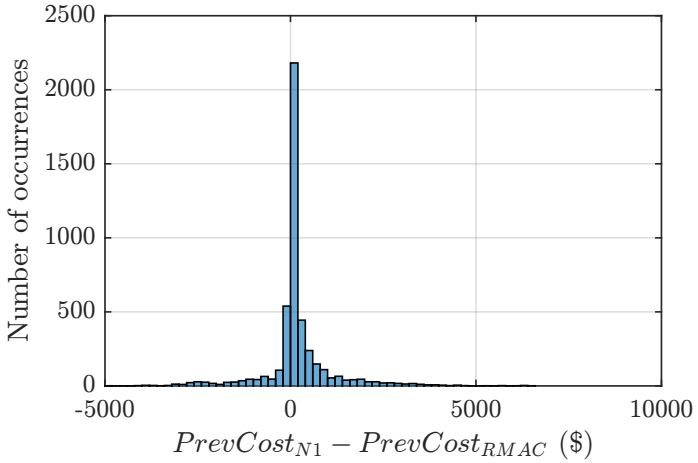



Figure 5.18: Histogram of the difference between the preventive control cost obtained with the "pseudo" N-1 criterion and the one obtained with the RMAC.

### Expected corrective control cost

The expected corrective cost is equal to 0 in 66% of the samples for both criteria. As can be seen in Figure 5.19, the impact of the criterion on the corrective cost is weak. In 53% of the cases, the expected corrective cost is the same.

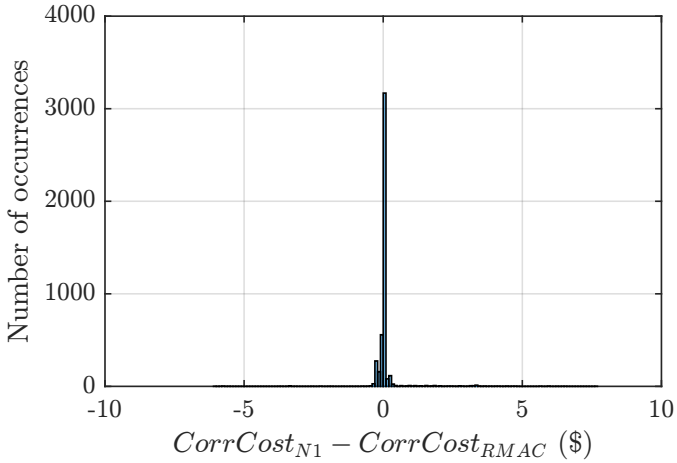



Figure 5.19: Histograms of the difference between the expected corrective cost obtained with the N-1 criterion and the one obtained with the RMAC.

### Expected criticality

The distribution of the differences in expected criticality for both approaches is represented in Figure 5.20. 72% of the samples have a larger expected criticality with the "pseudo" N-1 criterion.

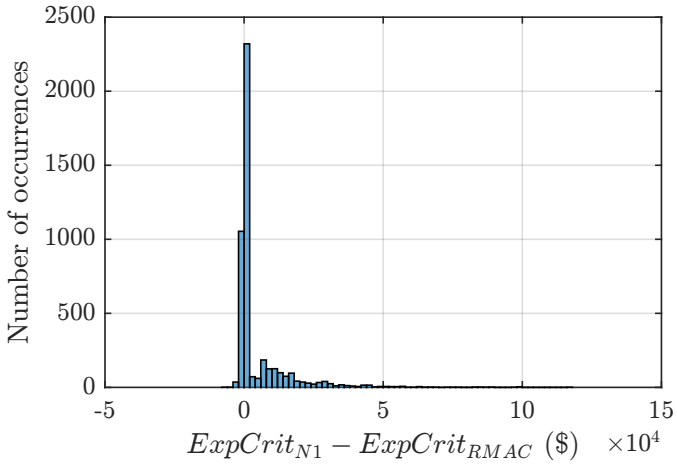



Figure 5.20: Histograms of the difference between the expected criticality obtained with the N-1 criterion and the one obtained with the RMAC.

However the set of considered contingencies is different for both criteria. Consequently, this analysis may not be meaningful. It is more interesting to compare the risk after control considering the same set of contingencies. This is done in Subsection 5.7.2.

### Probability of unacceptable states

The distribution of difference in probabilities of being in an unacceptable state for the "pseudo" N-1 criterion and the RMAC can be seen in Figure 5.21. 69% of the samples have a larger probability to be in an unacceptable state with the "pseudo" N-1 criterion.

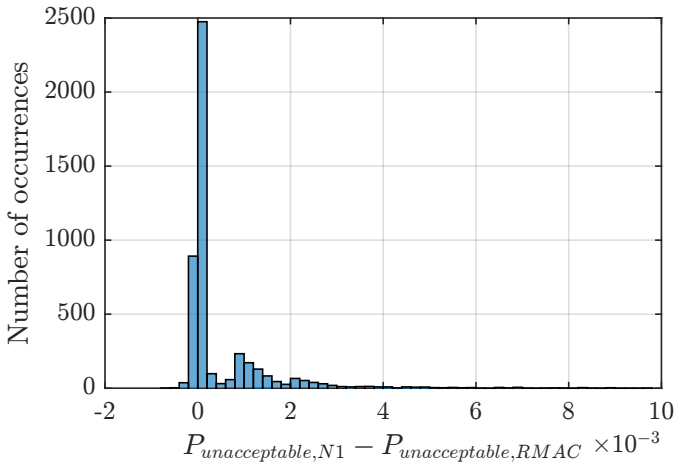



Figure 5.21: Histograms of the difference between the probability of unacceptable states obtained with the N-1 criterion and the one obtained with the RMAC.

## 5.7.2 Assessment

In the comparison between the "pseudo" N-1 criterion and the RMAC, the assessment part is the most interesting. Here we will analyse the impact of choosing a fixed set of contingencies on the risk computed while considering all the contingencies. Furthermore, another interesting analysis is to analyse the set of considered contingencies $\mathcal{C}_c$ to observe if it is close in average to $\mathcal{C}_{N-1}$.

The analysis of the subset of contingencies considered for the RMAC allows to determine which are the lines that are at risk. Indeed, if a line is often considered, that means that if we do nothing to prevent the impact of an outage, the risk is high. For some other contingencies, almost never selected, the impact of a failure is smaller or zero and therefore it is not necessary to take actions.

### Risk

The risk before control is the same for both approaches. Let's look at the risk after control actions. Figure 5.22 represents the difference between the risk after control computed with the "pseudo" N-1 approach and the one after control computed with the RMAC. It is not obvious on the graph but most samples have a difference in risk comprised between -2000 and 4000 $. For 60% of the samples, the difference is positive, meaning that the RMAC approach leads to a smaller risk. The average decrease is equal to 3495.4$.

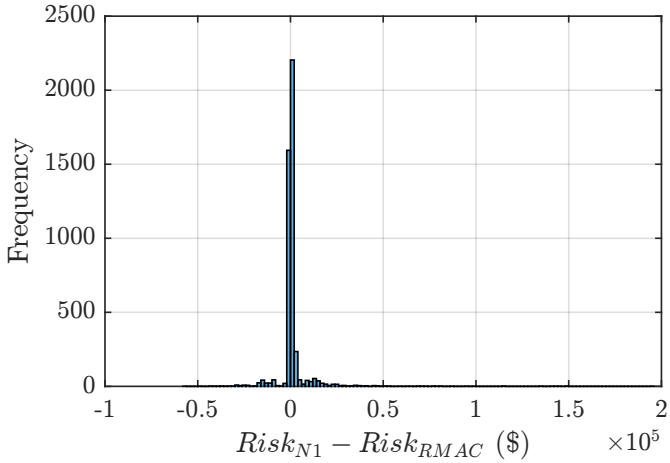



Figure 5.22: Histogram of the difference between the risks after control using the "pseudo" N-1 approach and the RMAC.

40% of the samples have a risk larger with RMAC than with the "pseudo" N-1 approach. The average decrease in risk is 2010.1$. In order to understand why it is the case, let's have a closer look at a sample chosen randomly among the ones where the N-1 approach seems more efficient in the point of view of the risk. For this sample, the risk obtained with the N-1 approach is equal to 111.77$ while the risk after RMAC control is 441.66$. Only 6 contingencies belong to $\mathcal{C}_c$: the no outage case, 3 from $\mathcal{C}_{N-1}$ and two double outages contingencies. $\mathcal{C}_c$ is much smaller than $\mathcal{C}_{N-1}$ that contains 39 contingencies. If we compare the total cost of the "pseudo" N-1 approach with the one of the RMAC approach, the results are respectively 3796$ and 386.4$. In fact, only corrective actions corresponding to the adjustment of phase-shifting transformers are used with the RMAC. Thus the cost corresponds to the expected criticality. In contrast, preventive actions and costly corrective actions are used with the N-1 approach. Thus the higher risk has to be opposed to a smaller cost of control actions.

Another case to look at is the one where the difference between the N-1 risk and the RMAC risk is the greatest in favour of the N-1 approach. For this particular case, the risk after control is greater than before control for the RMAC. Indeed, with the new dispatch, some contingencies with a zero criticality before control are associated to a non-zero criticality after control. Additionally some contingencies with a criticality different from 0 but discarded have a greater criticality after application of control actions. This is one of the cases for which an iterative approach would be necessary to finally have a residual risk smaller than $\Delta E$. In this sample, 3 single outage contingencies have a criticality different from 0 after the RMAC control and one single contingency that was discarded thanks to the threshold $\Delta E$ has a greater associated criticality. Since all these contingencies are considered for the "pseudo" N-1 approach, their criticality does not increase and thus the risk is reduced compared to the risk assessed before applying control actions (the common outage contingencies are not affected by the control action in this case).

For many samples (1303), the difference in risk is smaller than 1$. If we compare the control cost of both approaches for these samples, we observe that the cost is most of the time higher with the fixed set of contingencies $\mathcal{C}_{N-1}$. This can be seen in Figure 5.23. Concerning the preventive cost, it is most often similar. However, when the preventive cost is not equal, it appears that it is larger with the $N-1$ approach. For the corrective cost, it is also most often equal and there is no strong tendency for the cost to be larger with the "pseudo" N-1 criterion than with the RMAC. Finally, the expected criticality is most often greater with the N-1 criterion.

In general, one can conclude that for identical risk, choosing a fixed set of contingencies independently of the current state of the network is more costly.

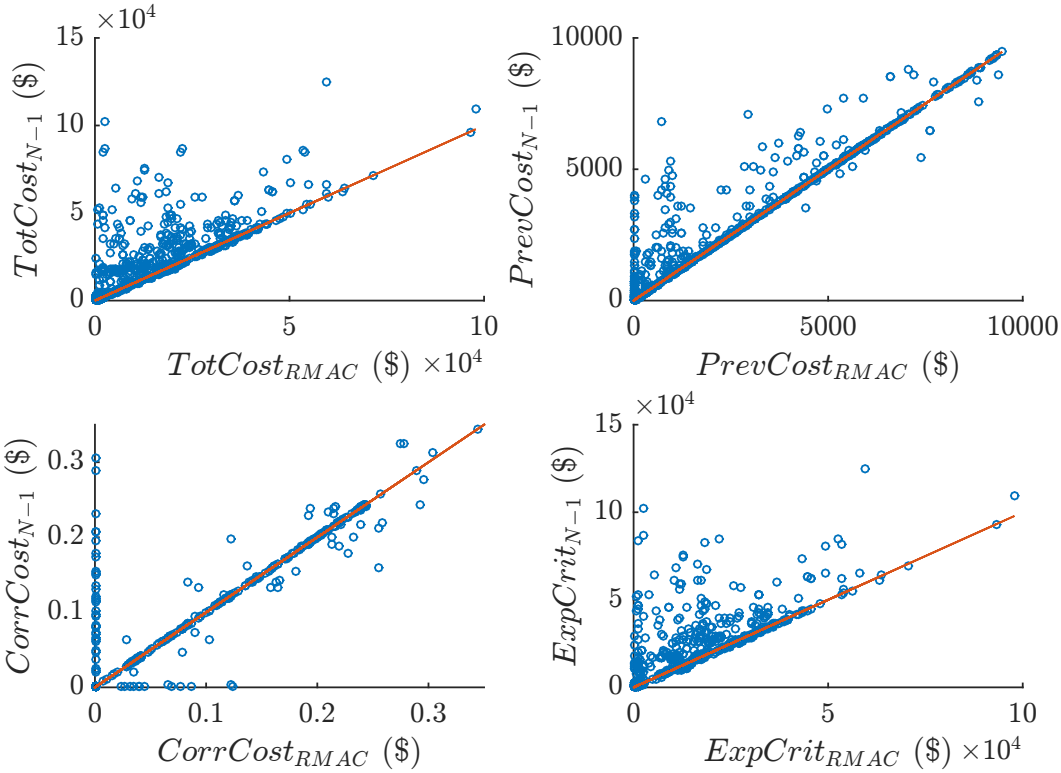



Figure 5.23: Scatter plots of the costs for "pseudo" N-1 approach and RMAC. All the samples have a difference in risk less than 1$ for both approaches. Note that the corrective cost graph has been zoomed in in order to show most samples and facilitate the analysis. Less than 20 samples have been discarded because of the zoom.

### Set of contingencies $\mathcal{C}_c$

In order to facilitate the analysis of the sets of considered contingencies, Figure 5.24 depicts a representation of the IEEE-RTS96 system where the numbers associated to each transmission line, transformer and double outage are the contingency numbers.

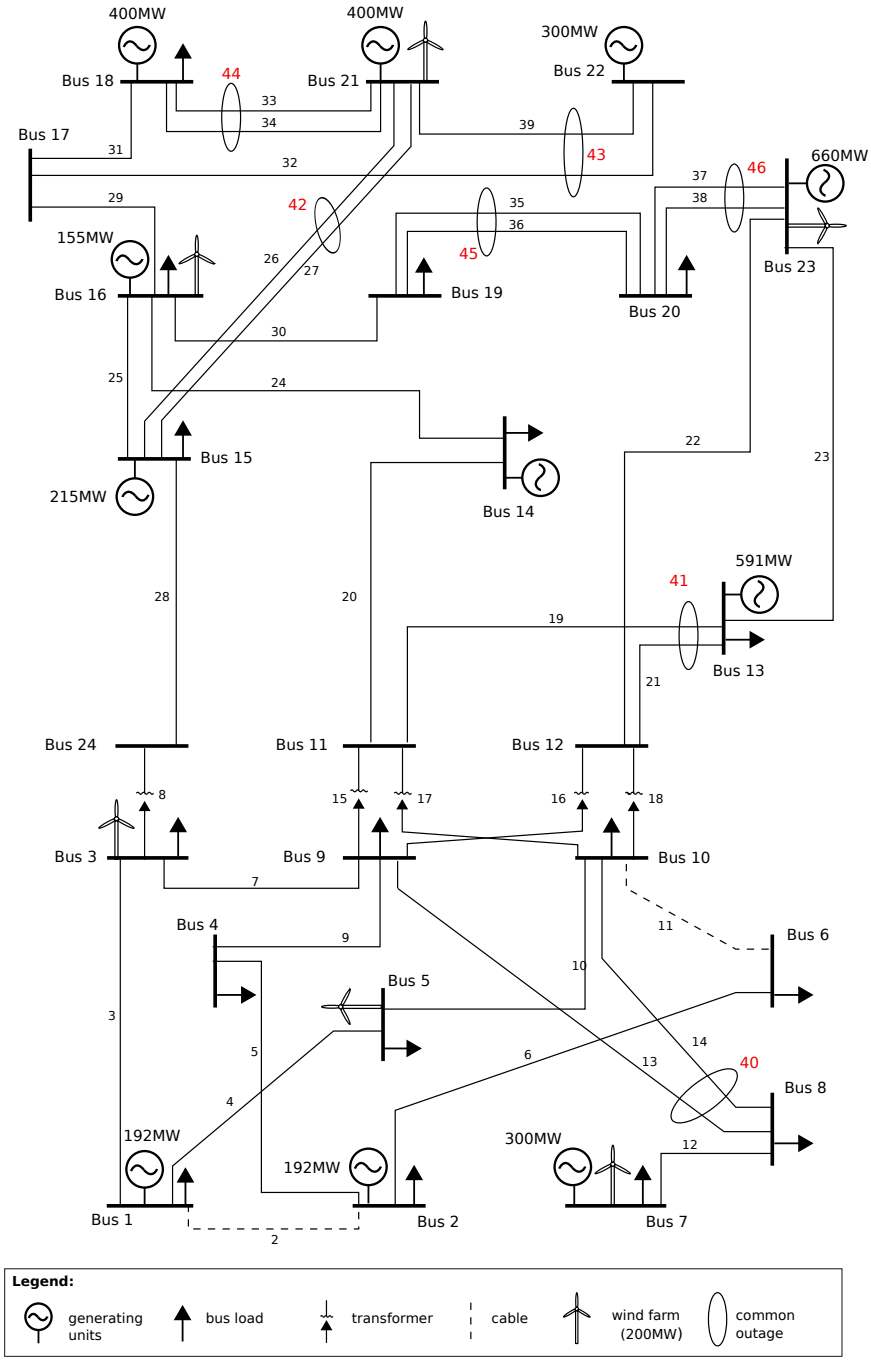



Figure 5.24: Representation of the modified IEEE-RTS96 system. The lines' numbering indicates which line is in outage with which contingency. For example, contingency 2 corresponds to the outage of the line connecting bus 1 and bus 2. Similarly, the number linked to a common mode double outage indicates to which contingency it corresponds.

The first interesting analysis is to observe which contingencies are considered more often. Figure 5.25 shows the number of occurrences of each contingency among all the considered sets $\mathcal{C}_c$.

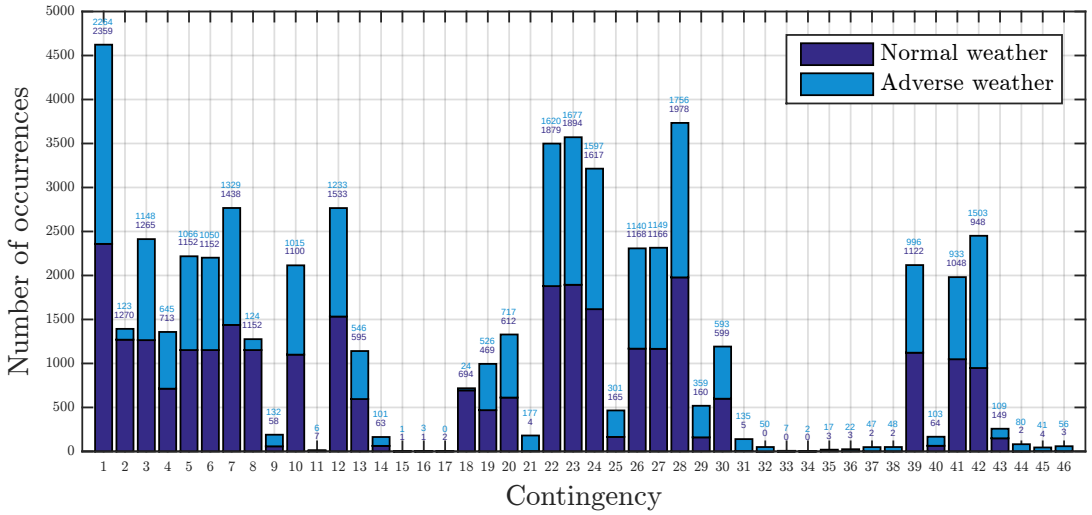



Figure 5.25: Number of occurrences of each contingency in all the sets $\mathcal{C}_c$.

Contingency 1 (no outage case) is of course selected each time. The contingency that is the most considered is the 28th. It corresponds to the outage of the transmission line joining bus 24 and bus 15. It indicates that this connection is important in the system.

Contingencies 22, 23 and 24 are considered in more than 69% of the samples. In contrast, several contingencies do not seem to be considered often. Contingencies 15, 16 and 17 correspond to the outage of transformers and have a small probability to occur (in the order of $10^{-6}$). Contingency 11 has probably a power flowing that is small and contingencies 33 to 38 define the outage of lines that are redundant. This may explain why these contingencies are less frequently selected.

The proportion of single outage contingencies considered in average is 26% (10/38) and the proportion of considered double outages is 15.6% (1/7). The numbers of single and double outage contingencies considered per sample are represented respectively in Figures 5.26 and 5.27. It can be observed the the complete set $\mathcal{C}_{N-1}$ is never selected. The common mode double outage contingencies are discarded in 28% of the samples.

Concerning the single outage contingencies, the distribution is bimodal. In general, the weather has not a strong impact on the number of considered contingencies. The distributions of the realisations of the total load and total wind power generation are also bimodal. However, as can be seen in Figure 5.28, this does not explain the bimodal distribution of the number of considered single outage contingencies.

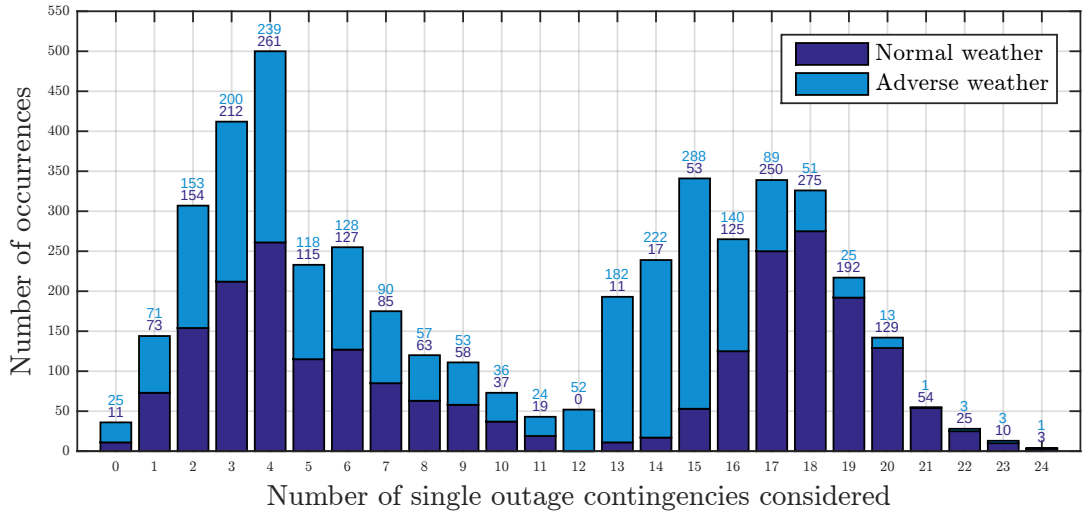



Figure 5.26: Histogram of the number of single contingencies in $\mathcal{C}_c$. The weather status of the samples is highlighted.

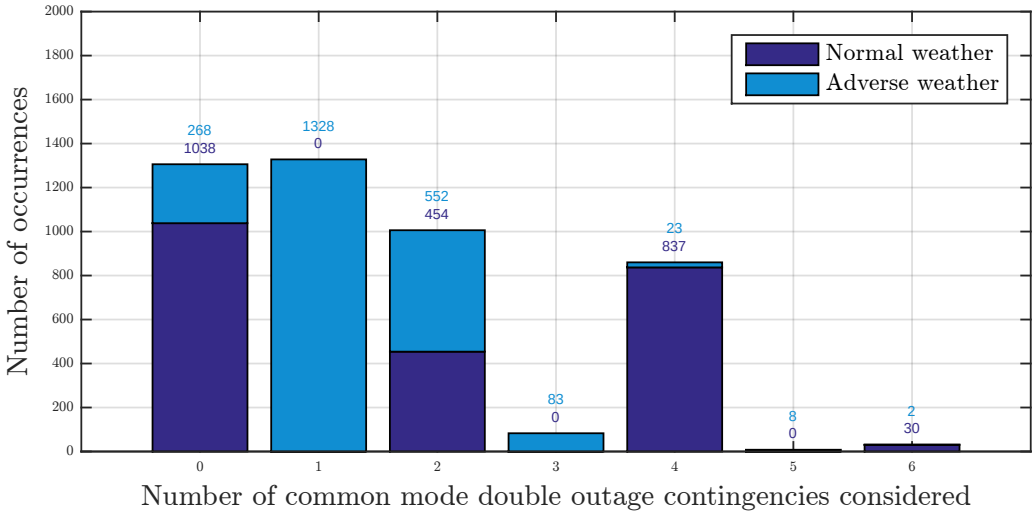



Figure 5.27: Histogram of the number of common mode double outage contingencies in $\mathcal{C}_c$. The weather status of the samples is highlighted.

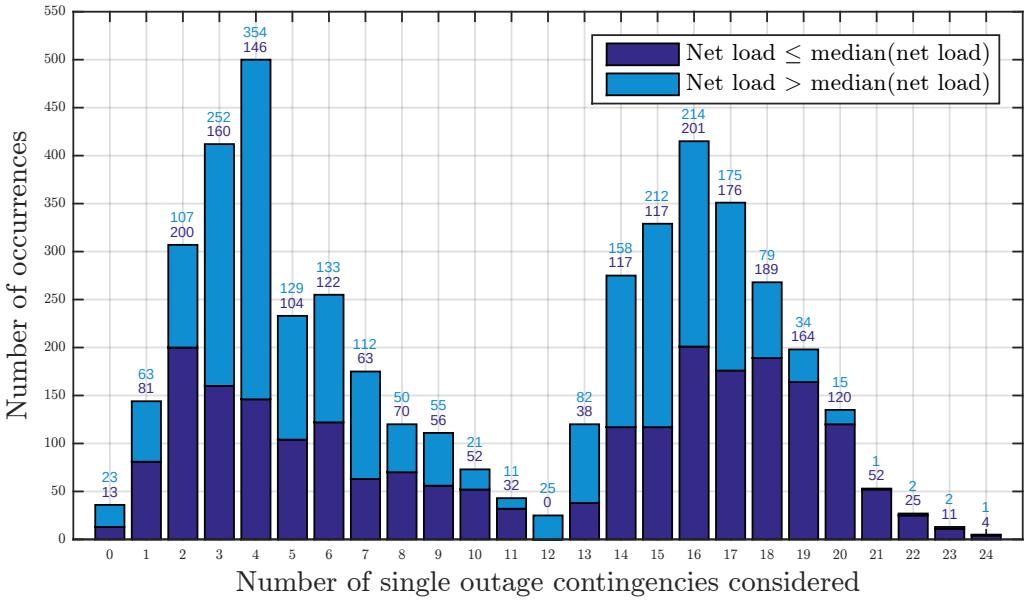



Figure 5.28: Histogram of the number of single contingencies in $\mathcal{C}_c$. The level of net load (total demand minus wind generation) of each sample is highlighted, depending on its value compared to the median.

## 5.8 Macroscopic variables representing the preventive and corrective control actions

Instead of analysing each corrective control action separately, we analyse them with a more macroscopic point of view. Concerning the preventive actions, the preventive ramp-up ($\Delta P^+_{0,g}$) or ramp-down ($\Delta P^-_{0,g}$) is given for each generator. One possible macroscopic variable is the sum of the ramps-up. In fact, $\sum_{g\in\mathcal{G}}(\Delta P^+_{0,g} + \Delta P^-_{0,g}) = 0$, thus it is better to study either $\sum_{g\in\mathcal{G}}(\Delta P^+_{0,g})$ or $\sum_{g\in\mathcal{G}}(\Delta P^-_{0,g})$ to assess the amount of re-dispatch. We chose to consider the sum of the ramps-up as the preventive macroscopic variable. The histogram of this variable is represented in Figure 5.29.

Regarding the corrective ramps-up and ramps-down of generating units, it is possible to create similar macroscopic variables for each contingency. For the sake of simplicity, only one contingency will be studied. In order to choose one, we represented the number of times corrective re-dispatch is planned per contingency. This can be seen in Figure 5.30. It was computed by counting the number of times $\sum_{g\in\mathcal{G}}(\Delta P^+_{c,g})$ is greater than 0, where $\Delta P^+_{c,g}$ is the corrective ramp-up of generator $g$ for contingency $c$. One can see in the figure that the contingency for which corrective re-dispatch is planned the largest number of times is contingency 28. Therefore, this is the contingency for which corrective re-dispatch will be studied. The histogram

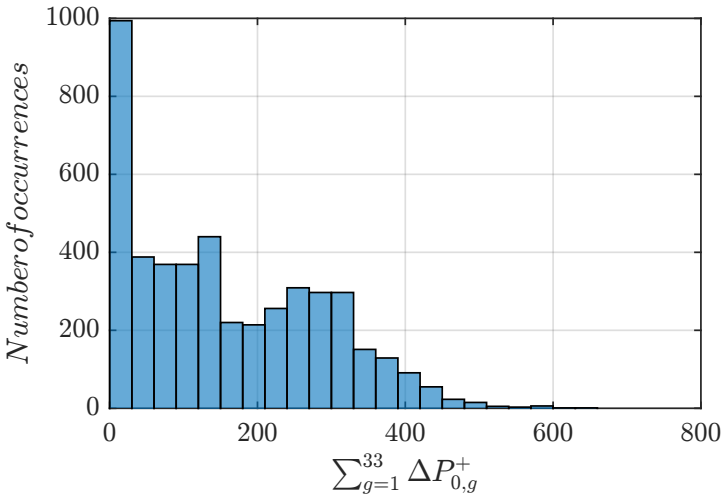



Figure 5.29: Histogram of the preventive macroscopic variable $\sum_{g \in \mathcal{G}} \Delta P_{0,g}^{+}$.

of the corrective re-dispatch macroscopic variable can be seen in Figure 5.31.

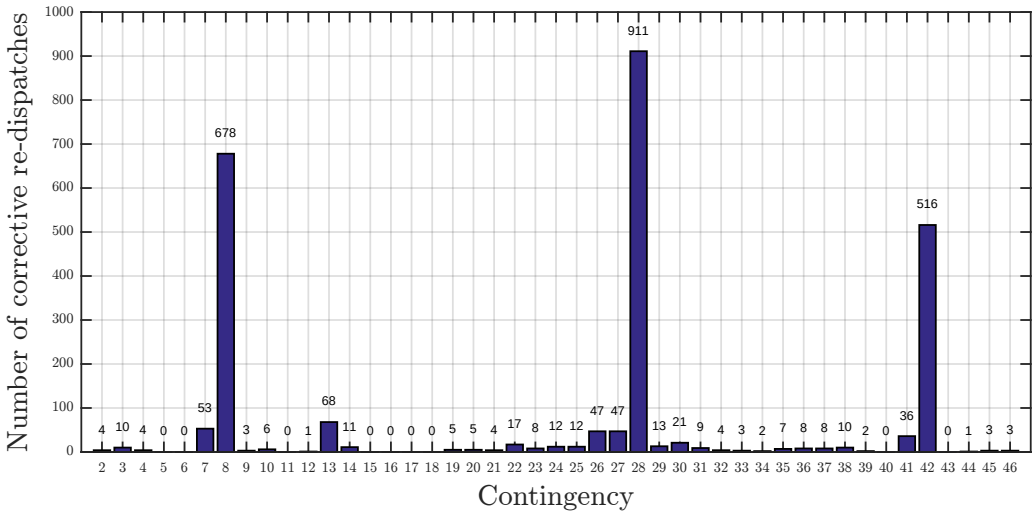



Figure 5.30: Number of times corrective re-dispatch is planned per contingency.

Finally, the phase shifts of transformers do not necessarily cancel each others. Therefore, we do not need to sum only the positive phase shifts. The macroscopic variable is thus the sum of the absolute value of phase shifts per contingency, *i.e.* $\sum_{l \in \mathcal{L}} |\Delta\theta_{c,l}|$. If we plot a graph representing the number of times adjustment of phase-shifting transformers is used as a corrective action per contingency, the contingency 23 is the contingency for which this type of corrective actions is most often planned. Thus this is the contingency for which phase shifts of transformers will be studied more in detail. The histogram of the corrective phase shifts of transformers macroscopic variable can be seen in Figure 5.33.

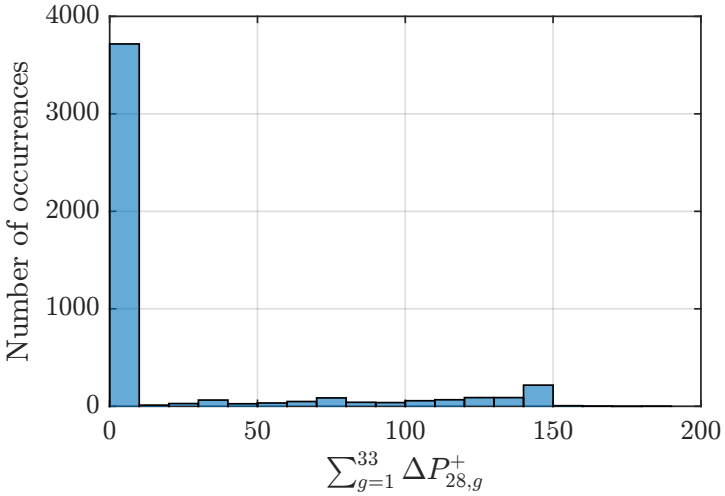



Figure 5.31: Histogram of the corrective re-dispatch macroscopic variable $\sum_{g \in \mathcal{G}} \Delta P_{28,g}^{+}$.

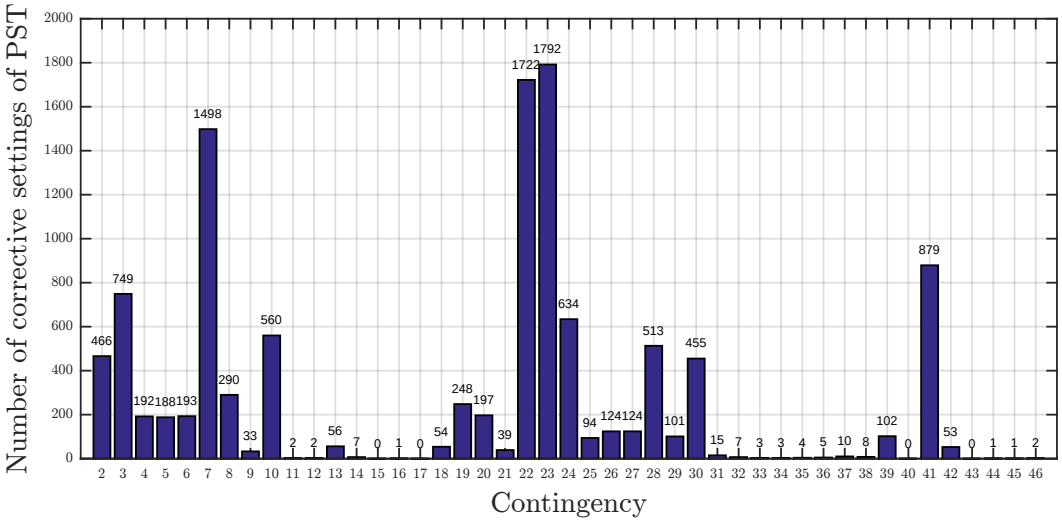



Figure 5.32: Number of times corrective phase shifts of transformers are planned per contingency.

Note that adjustment of phase-shifting transformers is used more often than corrective re-dispatch, which is logical since in the implementation, these actions are considered as having no cost.

For information, the maximum value of $\sum_{l \in \mathcal{L}} |\Delta\theta_{c,l}|$ among all contingencies is 2.08° and the maximum value of $\sum_{g \in \mathcal{G}} \Delta P_{c,g}^{+}$ is 360MW.

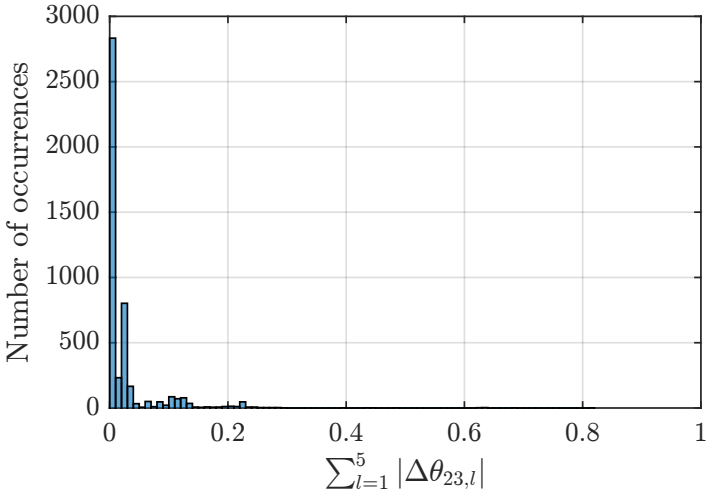



Figure 5.33: Histogram of the corrective PST macroscopic variable $\sum_{l\in\mathcal{L}} |\Delta\theta_{23,l}|$.

## 5.9 Outputs summary

This section is a summary of the outputs that will be used for the machine learning part of this work.

The cost variables are the following:

- the total cost,
- the preventive control cost,
- the expected corrective control cost,
- the expected criticality.

The other variables computed from the control program are:

- the probability of being in an unacceptable state,
- the sum of preventive control ramps-up $\sum_{g\in\mathcal{G}} \Delta P^{+}_{0,g}$,
- the sum of corrective control ramps-up $\sum_{g\in\mathcal{G}} \Delta P^{+}_{c,g}$ for contingency 28,
- the sum of corrective control phase shifts $\sum_{l\in\mathcal{L}} |\Delta\theta_{c,l}|$ for contingency 23.

As a reminder, contingencies 28 and 23 are the contingencies for which respectively corrective control re-dispatch and setting of phase-shifting transformers are used the largest number of times. Eventually, the last output is the risk $R_{\mathcal{C}}$, computed with the post-assessment program.

# Chapter 6

# Analysis of the relevance of input features

The aim of this chapter is to study the strength of the impact of the different input variables defined in the previous chapter on the control decisions and the risk computed with the model developed by Karangelos *et al.* in [3].

In Chapter 2, some tree ensemble methods were briefly introduced. In the present chapter, random forests and extremely randomized trees are used as a tool to rank the features as a function of their importance. This ranking will also be used in the next chapter to select the features that are more relevant for predictions. The ranking objective is to understand which variables have more effect on the outputs. On the other hand, the selection of a subset of features may avoid overfitting and make the prediction computationally more efficient.

In the following two sections, we first present the algorithms and the methodology used to compute the importances of the different input features in terms of their contribution to the prediction of a certain target output variable. The core of the chapter then reports the results we have obtained when analysing the different target variables of interest, and discusses them from a physical point of view.

## 6.1 Background

This section introduces briefly how feature importances are determined with random forests and extra-trees. The metrics used to select a model and assess its performances are also defined.

### 6.1.1 Feature importances derived from random tree ensembles

In this work, we exploit the so-called 'feature importances' that can be computed as a by-product of training models in the form of random forest or extremely randomized trees. The reason is that they provide a non-parametric multivariate ranking and assessment of the amount of information provided by each input feature in conjunction with the other ones, without any particular assumption about linearity,

gaussianity or independence among features.

Two measures to evaluate feature importances with tree-ensemble methods were proposed by Breiman in [19][36]. The first one is the Mean Decrease Accuracy (MDA) and the second one is the Mean Decrease Impurity (MDI) [37].

The MDA consists in randomly permuting the values of a feature and observe the mean square error of the resulting trees with out-of-bag samples. If the mean square error does not change, it means that this feature has little influence on the prediction. On the other hand, if the MSE increases, the feature is important for the result. This measure is also called permutation importance.

The MDI is based on the reduction of impurity performed at each split. The decrease of an impurity measure is what is minimised at each split. For regression trees, the variance is used. For a single tree $T$, the decrease in impurity is weighted by the proportion of samples that will reach the node. This proportion is approximated by the number of training samples reaching a particular node and thus if $\Delta i(s_t, t)$ is the decrease of impurity at node $t$ and for split $s_t$, the importance of a variable $v_j$ is defined as:

$$FI_T(v_j) = \sum_{v(s_t)=v_j} p(t)\Delta i(s_t, t), \tag{6.1}$$

where $p(t)$ is the proportion of training samples reaching node $t$ and $v(s_t)$ is the splitting variable at node $t$. For a tree-ensemble method, $FI_T(v_j)$ is computed for all trees and the results are then averaged to give :

$$FI_{forest}(v_j) = \frac{1}{M}\sum_T FI_T(v_j) = \frac{1}{M}\sum_T \sum_{t\in T: v(s_t)=v_j} p(t)\Delta i(s_t, t), \tag{6.2}$$

where $M$ is the number of trees in the forest.

We refer the interested reader to reference [37], for a study of the theoretical properties of these MDI measures.

### 6.1.2 Metric used for model selection and assessment

As already explained, in our study we will use both random forests and extremely randomized trees as the core methods in order to calculate the feature importances for different target output variables. However, these methods may lead to different performances in terms of their capability to predict the different target output variables, and these performances will typically also depend on the chosen values for the meta-parameters of these methods. In order to show meaningful results it is therefore necessary to assess the predictive accuracy of different models, and base the analysis on those models and meta-parameter settings which are most suitable for each target output variable.

In the context of regression, several metrics may be used to compare different models and choose the best settings of their meta-parameters. In our study, we decided to use the so-called $R^2$-score.

$R^2$ is also called the coefficient of determination. It is computed on the basis of a sample of size $N$ by [38]:

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}, \tag{6.3}$$

where $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ is the mean of the true values, and $\hat{y}_i$ denotes the model prediction.

The coefficient of determination can be (loosely) interpreted as the fraction of the variance of the target output variable that is explained by the model. The second term in the definition can be seen as a relative mean square error and should be as close as possible to 0. The best possible value for the $R^2$-score is 1, and corresponds to a model that perfectly predicts all the target output values of the sample used to estimate its value. On the other hand, a model that would predict the constant value $\bar{y}$ (the sample mean of the output target values) would obtain an $R^2$-score of 0. The $R^2$-score is a normalised measure of accuracy, which interpretation is independent on the scaling of the target output variable.

## 6.2 Methodology

The first step in the determination of the feature ranking is to find the best estimator and the associated parameters. Therefore, for each output, the two models (random forest and extra trees) are tested and a 5-fold cross validation (CV) is used to evaluate the best setting for their meta-parameters.

The procedure is the following: first of all, the dataset is divided into 2 subsets, one containing 80% of the samples and called the learning set and the other containing 20% of the samples and called the test set. The division is carried out at random. Then, a cross-validation test is used together with the learning sample in order to determine the best meta-parameter values for a given estimator. All the possible combinations are tested and the chosen one is the one with the maximum CV score. The CV score used is the $R^2$-score. After that, in order to determine the score of the estimator on the test set, the complete learning set is used to learn a model with the corresponding best parameters. Finally, this model is used to predict the values of the test set and then the predicted values are compared to the true values to compute the $R^2$-score. This latter score is called in the sequel the test score to distinguish it from the CV score.

The tested values for the meta-parameters are enumerated hereafter:

- $M$: $500, 1000$;
- $K$: $p, \sqrt{p}, 1, \frac{p}{2}, \frac{p}{3}$;
- $n_{\min}$: $2, 3, 4, 10$.

Table 6.1 summarises the results obtained. It details for each output the best meta-parameter values per estimator as well as the resulting CV and test scores.

Table 6.1: Best parameters per output for each estimator, corresponding CV score and test score. Both are $R^2$-score. The CV-score is computed with the learning set (80% of the dataset) and the test score with the test set (20% of the dataset). To determine the test score, the model is learnt with the learning set and then used to predict the outputs of the samples from the test set. The results are then compared to the true values.

| Output | Model | Parameters | | | CV score | Test score |
|---|---|---|---|---|---|---|
| | | $M$ | $K$ | $n_{min}$ | | |
| Total cost | ET | 1000 | $p/2$ | 2 | 0.7653 | **0.7731** |
| | RF | 1000 | $p/3$ | 2 | 0.7271 | 0.7534 |
| Preventive control cost | ET | 500 | $p$ | 4 | 0.7724 | **0.7980** |
| | RF | 1000 | $p/2$ | 2 | 0.7916 | 0.7776 |
| Expected corrective control cost | ET | 500 | 1 | 2 | 0.2705 | 0.3472 |
| | RF | 500 | $\sqrt{p}$ | 2 | 0.2673 | **0.4081** |
| Expected criticality | ET | 500 | $p/2$ | 5 | 0.7443 | **0.7445** |
| | RF | 500 | $p/3$ | 3 | 0.7118 | 0.7301 |
| $P_{unacceptable}$ | ET | 500 | $p/2$ | 3 | 0.7847 | **0.7954** |
| | RF | 500 | $p/3$ | 3 | 0.7590 | 0.7842 |
| Risk after control | ET | 500 | $p/2$ | 4 | 0.6807 | **0.6868** |
| | RF | 1000 | $p/3$ | 4 | 0.6432 | 0.6292 |
| Macroscopic preventive dispatch | ET | 1000 | $p$ | 2 | 0.6667 | 0.7140 |
| | RF | 500 | $p/2$ | 2 | 0.7001 | **0.7373** |
| Macroscopic corrective dispatch for contingency 28 | ET | 500 | $p$ | 4 | 0.5950 | 0.5368 |
| | RF | 500 | $p/2$ | 2 | 0.6520 | **0.5987** |
| Macroscopic corrective PST for contingency 23 | ET | 500 | $\sqrt{p}$ | 2 | 0.4156 | **0.4106** |
| | RF | 1000 | $\sqrt{p}$ | 2 | 0.4048 | 0.3877 |

As can be observed, random forests and extremely randomized trees obtain similar results in terms of accuracy, even if the latter are slightly better in 6 out of the 9 cases. We also observe that the test scores and the cross-validation scores are close to each other, even if in most cases the former is slightly better. It has a value of approximately 70%, except for the three variables representing the corrective actions and cost were the performances of the model drop significantly compared to the other outputs. The random forest estimator performs better when predicting the expected corrective control cost, the macroscopic preventive control variable and the macroscopic corrective re-dispatch variable. The macroscopic corrective re-dispatch variable is better predicted than the expected corrective control cost and the macroscopic corrective setting of phase-shifting transformers (PST) variable. In principle, the larger the $M$ the better the expected performance; the fact that sometimes $M = 500$ outperforms $M = 1000$ indicates that further increasing the value of $M$ would not significantly improve the scores.

In the rest of this chapter, we will consider the model (and its "optimal" meta-parameters) shown in bold in the table to carry out the analysis of input features

importances for each target output variable.

## 6.3 Feature ranking and discussion

Now that the best model along with suitable values of the meta-parameters have been determined for each target output, it will be used to assess feature importances. The procedure is similar as the one used to determine the best model. Namely, for each target output variable, we first re-sample a learning set, containing 80% of the samples, in order to build the corresponding model. The importances and the corresponding ranking of the input features are derived from this model. In Scikit-learn, the importance measure is the Mean Decrease Impurity and therefore, this is the one used in this work. Note that the feature importances in Scikit-learn are normalised, in such a way that the sum of importances of all features is 1. Finally, the performances of the model are again tested using the remaining 20% of the dataset. Notice that in our simulations we actually carried out the analysis for each target output several times (with different randomly selected learning and test sets), in order to assess whether the resulting ranking and importances were stable or not. Although the precise values of the feature importances are indeed influenced by the random selection of the learning sample, we did not notice any counter-intuitive changes from one run to the other, so that in the rest of this chapter we will report only the values for a single such split for each target output.

This section begins with a description of the feature relevance analysis for the different terms of the objective function used in the RMAC control part, then it continues with the risk computed at the post-assessment stage and finally it presents the results for the three macroscopic control variables. Given the large number of features, only the 40 most important features are shown for each output.

### 6.3.1 Costs computed by the RMAC control part

#### Total cost

First of all, the histogram of the total cost used as objective function by the control part of the RMAC is displayed in Figure 6.1. The mean equals 10,841$ and the standard deviation is 15,048$. It can be seen that the weather has a strong influence on the value of the total cost. Indeed, the maximum value of the total cost for a normal weather is 8273$. Most samples with an adverse weather have a total cost greater than this value. The mean value of the total cost for normal weather is 1470.2$ and the standard deviation equals 1028.8$. On the other hand, the mean and standard deviation of the total cost for samples with an adverse weather are respectively 20,666$ and 16,567$.

The feature importances of the 40 most important input features used when they are used to predict the total cost are displayed in Figure 6.2.

It can be seen that in the 10 most important features, six are probabilities of contingencies. Furthermore, except for “contingency 1” and “contingency 28”, all these

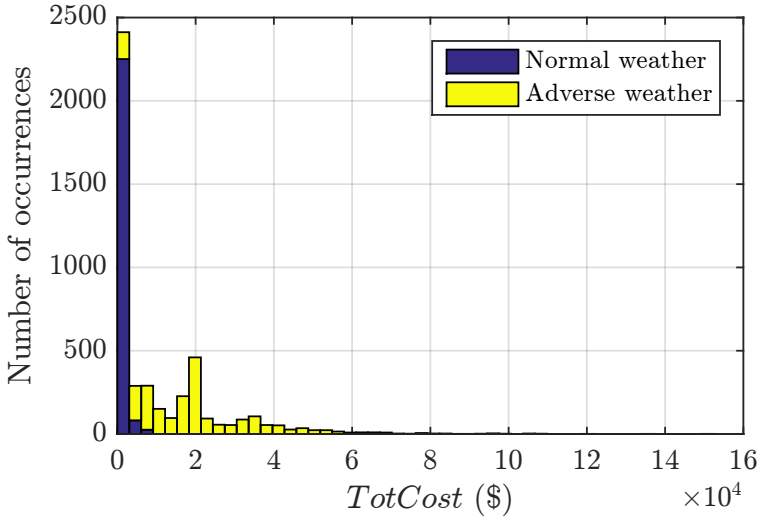



Figure 6.1: Histogram of the total cost. The weather status of the samples are highlighted.

contingencies are common mode double outage contingencies. The other important features are the discarding threshold $\Delta E$, the weather status, the market clearing dispatch of a generator and the net load (difference between the total demand and the total wind generation). These 10 features account for a little bit more than 50% of the feature importances. Note that *market power of gen i* is an unfortunate naming for the market clearing dispatch of generator $i$. It has nothing to do with the usual definition of market power.

The probability of contingency 46 is the most important feature and its histogram is displayed in Figure 6.3. Looking at the graph on the top, we distinguish two bars. They indicate the weather status of the sample. The smallest probabilities correspond to a normal weather and the largest to an adverse weather. If we zoom in, we see that there are not only two possible values. Indeed, the probability of a contingency is equal to the probability of failure of the corresponding lines multiplied by the probability of having no failure of other lines (see Section 4.3.6). If a line is unavailable, it modifies $P(\text{no failure})$ and thus $P(\text{cont}_c)$. As a result, the probabilities of contingencies contain the information that other lines are unavailable. The probability can also have a zero-value, indicating that the two lines corresponding to the common mode double outage are both unavailable. All the double outage contingencies have a similar histogram. If the feature importances are computed several times with different learning and test sets (but with the same proportion of samples in both), the top ten variables are identical but the order might change a little bit. However, the most important variable is always a double outage probability. This means that the information brought by a double outage probability is, in general, important in the determination of the total cost.

As was expected when looking at Figure 6.1, the weather seems to be important in the determination of the total cost. Indeed, this information is brought by each

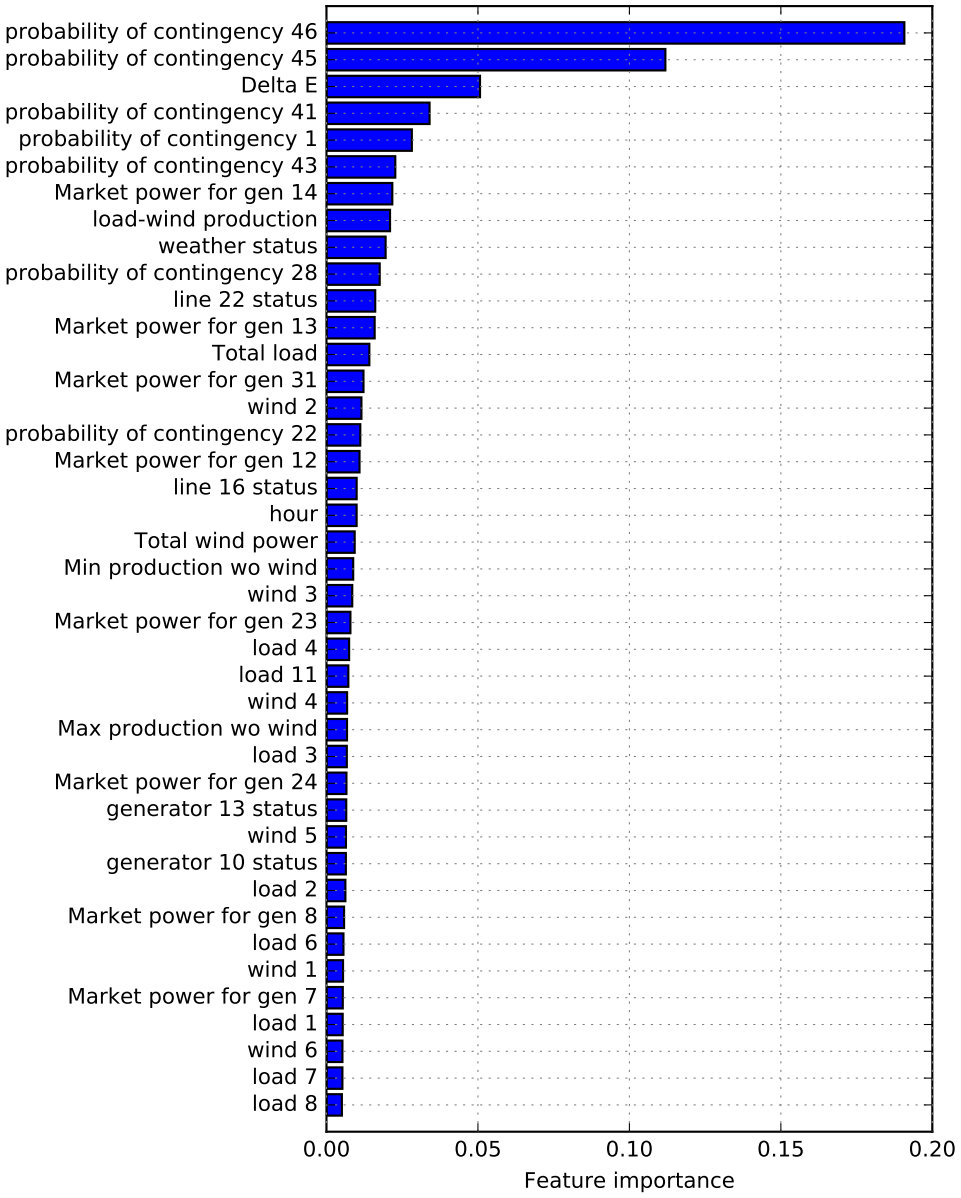



Figure 6.2: Feature importances of the 40 most important variables to predict the total cost. The $R^2$ score is equal to 0.764 and the cumulative feature importance of these 40 features is 0.770.

probability of contingency and by the variable *weather status*.

Other important variables are the discarding threshold, the difference between the load and the wind generation and the market clearing output of generator 14. $\Delta E$ has an impact on the set of considered contingencies and therefore it is logical that it is an important variable. Generator 14 is on the bus 13. Figure A.1 is a representation of the IEEE-RTS 96 system where the locations of each generating unit, load and wind farm is indicated. The unit can produce up to 197MWh and its re-dispatching cost is the highest, considering only the units that are on. The bus 13 is one of the buses between the bottom of the network where more loads

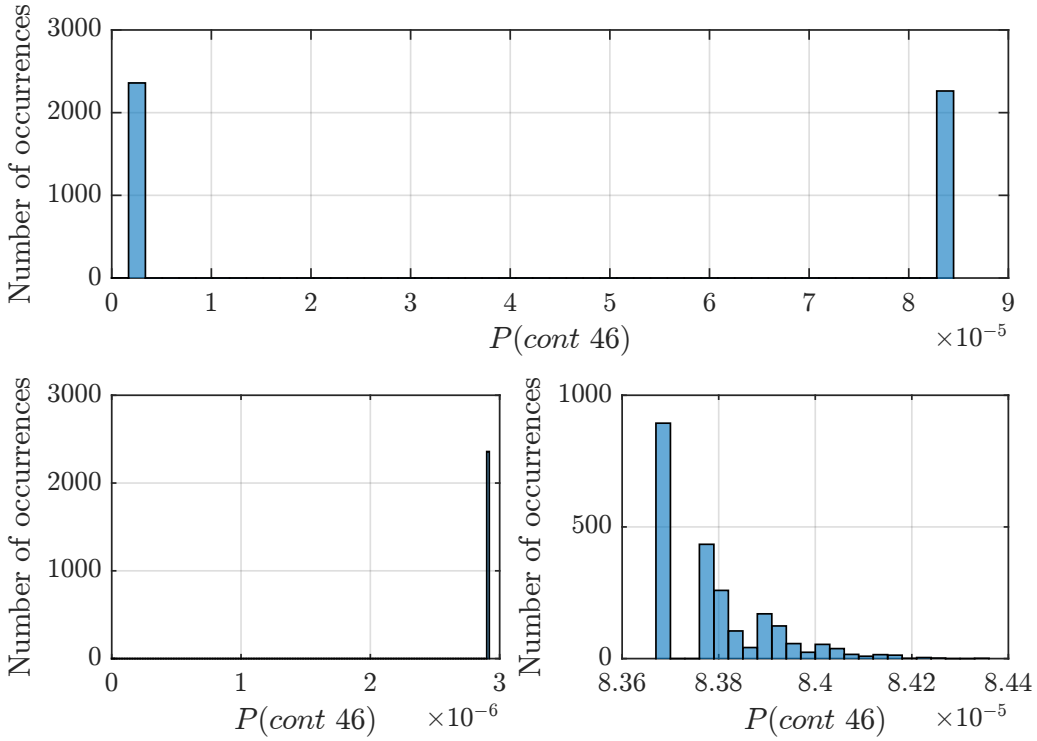



Figure 6.3: The graph at the top is the histogram of the probability of contingency 46. The two histograms at the bottom are also histograms of the probability of contingency 46 but the one on the left displays only the values less than $4 \cdot 10^{-5}$ and the graph on the right displays only the values greater than $4 \cdot 10^{-5}$.

are present and the top of the network where the large generating units dominate. The strategic location of generator 14 and the high re-dispatching cost may explain why the market clearing output of this generator is important. The market clearing output per hour and the re-dispatching price are detailed in Table B.2. Figure 6.4 shows the sum over all the samples of the absolute value of the ramps-up and ramps-down per generator. One can observe that generators 23 and 24 are the most used. They are nuclear units with the lowest re-dispatching price and a large capacity of production, thus it is logical. Generating units 12, 13 and 14 are all on the same bus and are often used despite their larger cost. This is why their market clearing outputs influence the total cost. While this reasoning seems plausible, we have not found any explanation to understand why gen 14 is more important than gen 12, even if the difference is small.

It is interesting to notice that the status of line 22 and the probability of contingency 28 are also important features. The contingency 28 is the failure of line 22 and this line is important for the network. This line is indeed a connection between the part of the network where most loads are located and the part when the large generating units are. An outage of this line can have a significant impact on the system.

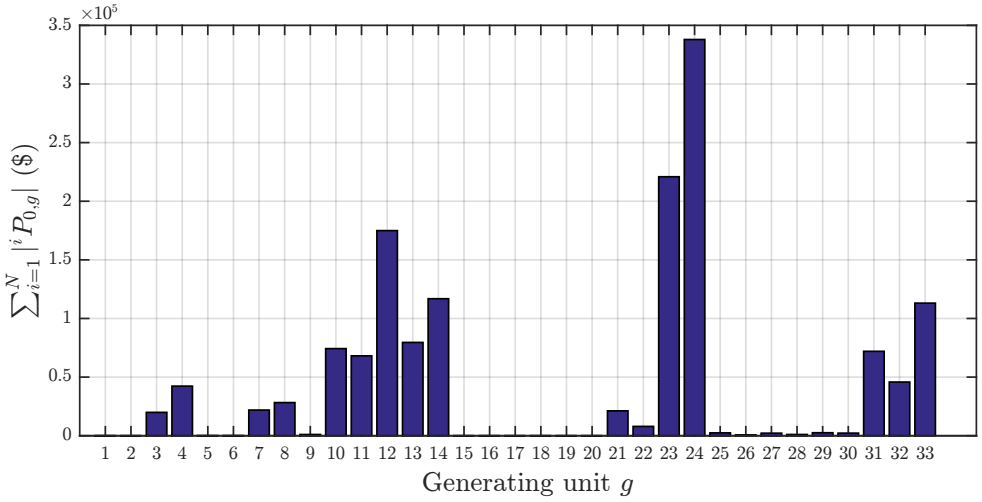



Figure 6.4: Sum over each sample of the absolute value of ramp-up or ramp-down per generator, $\sum_{i=1}^{N} |^i P_{0,g}|$, where $N$ is the number of samples and $^i P_{0,g}$ is the preventive ramp-up/down of generator $g$ in the sample $i$.

### Preventive control cost

Figure 6.5 shows the histogram of the preventive control cost. The mean and the standard deviation respectively equal 1812.5$ and 2477.1$. As for the total cost, the weather influences the preventive control cost. Indeed, the means for a normal and an adverse weather are respectively 392.73$ and 3291.9$. The standard deviations are equal to 1041.9$ and 2666.5$.

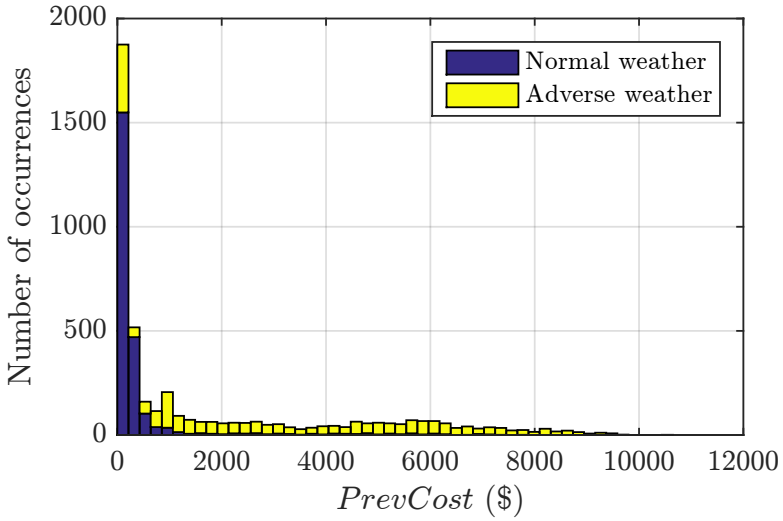



Figure 6.5: Histogram of the preventive control cost. The weather status of the samples are highlighted.

If we study the important features for the preventive control cost (Figure 6.6),

we see that the most important feature is the probability of contingency 24. The other important variables are related to difference between the load and the wind production and the market clearing outputs of some generators. Given that the possible preventive actions are the re-dispatch of the generating units, it seems logical that they are important in the determination of the preventive control cost. $\epsilon$ is also an important variable. This can be explained by the fact that for a small $\epsilon$, more preventive actions need to be taken to meet the reliability criterion (equation (3.2)).

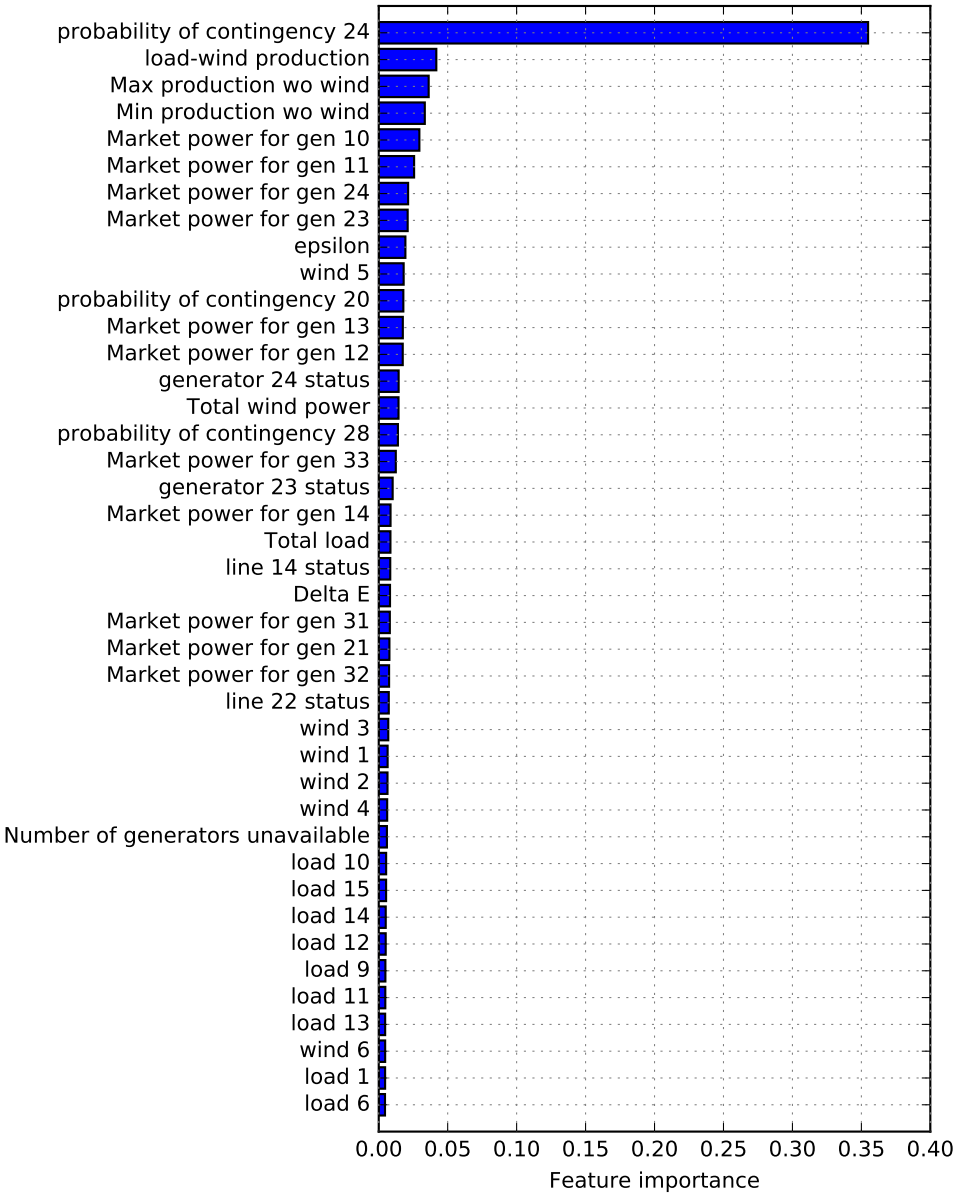



Figure 6.6: Feature importances of the 40 most important variables to predict the preventive control cost. The $R^2$ score is 0.805 and the 40 variables represent 86% of the cumulated importance.

The variable *probability of contingency 24* accounts for 35% of the sum of feature importances. This contingency is considered in 70% of the samples. Therefore, the

corresponding line is an important line for the system. If we remove that feature and determine the new most important features, it can be seen in Figure 6.7 that probability of contingency 24 is replaced by two probabilities of contingencies. The sum of their importance is approximately 0.35. Furthermore, the test score is similar. As a conclusion, the information contained in the variable *probability of contingency 24* can be brought by a combination of two other variables. The other variables are mostly similar but the difference between load and wind generation as well as $\epsilon$ do not appear anymore. The variable *line 18 status* is in this case an important variable. The contingency corresponding to the failure of this line is the contingency 24. Therefore, it can be concluded that line 18 is an important line for preventive control. The fact that probability of contingency 24 is replaced by contingency 20 seems logical. Both corresponding lines are connected to the bus 14. Concerning contingency 27, in Figure 6.4, it can be seen that the generating unit 24, on bus 21, is the generating unit for which the total amount of ramping-up/down is the highest. It is therefore an important unit for the preventive control. Line 20 (contingency 27) is a line that connects bus 21 to the bus 15. It is one of the most direct connection from this nuclear unit to the loaded part of the system. This can explain why the probability of contingency 27 is important.

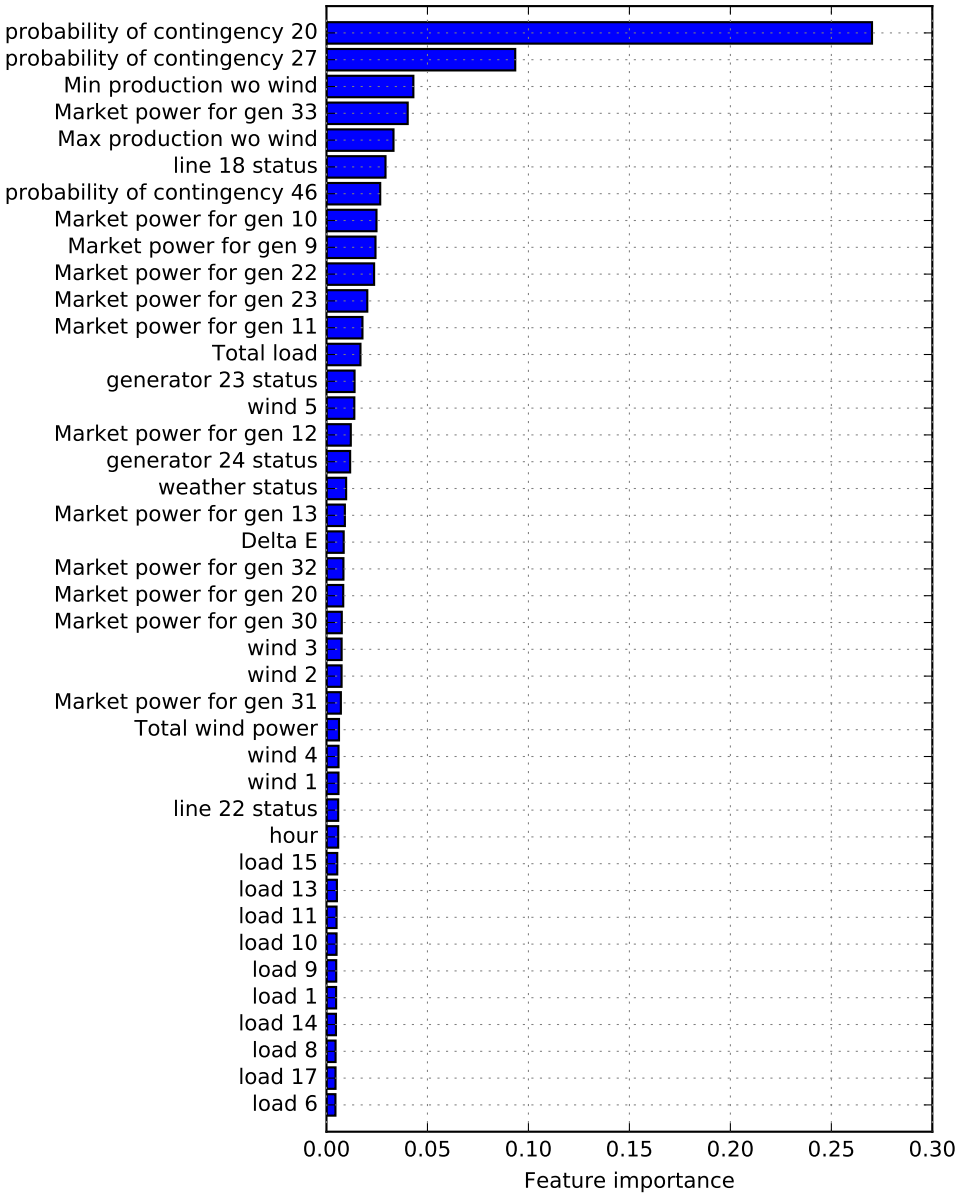



Figure 6.7: Feature importances of the 40 most important variables to predict the preventive control cost when feature *probability of contingency 24* is removed. The $R^2$ score is 0.833 and the 40 variables represent 86% of the cumulated importance.

### Expected corrective control cost

The histogram of the expected corrective control cost is represented in Figure 6.8. The weather does not have an impact as strong as for the other outputs. The mean value and standard deviation for a normal weather are respectively 0.0616$ and 0.09$ and for an adverse weather there are 0.1520$ and 0.6215$. There are 3034 zero-valued samples: 1344 with a normal weather and 1690 with an adverse weather.

The random forest and extra-trees algorithms do not give good results. Indeed, the test score is equal to 41%. Nevertheless, the feature importances of the 40 most important attributes are displayed in Figure 6.9. One can notice that the most

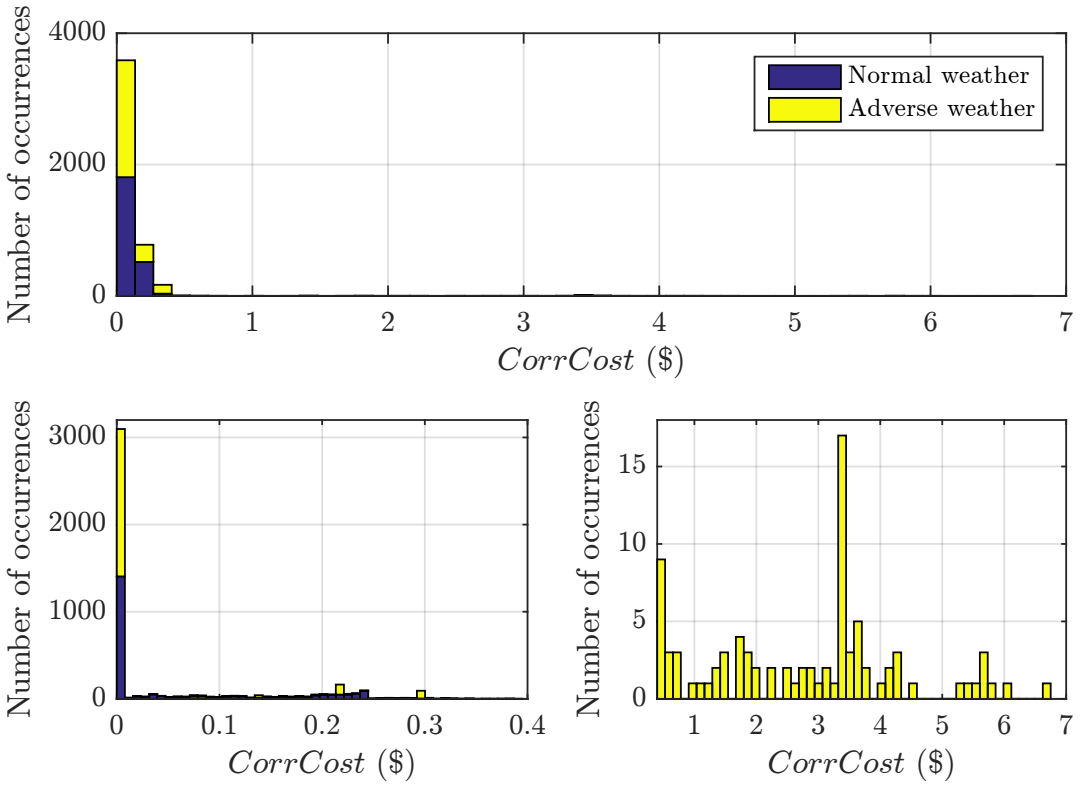



Figure 6.8: Histogram of the expected corrective control cost. The weather status of the samples are highlighted. The two graphs at the bottom represents also the histogram of the expected corrective cost but the one on the left only displays samples smaller than or equal to 0.4 $ and the one on the right shows the samples with a cost greater than 0.4$. There is no normal weather samples in the graph on the right.

important variables seem to be $\Delta E$ and the realisations of load and wind generation.

The weather status does not appear in the 40 most important variables. That was expected given the distribution of the expected corrective cost.

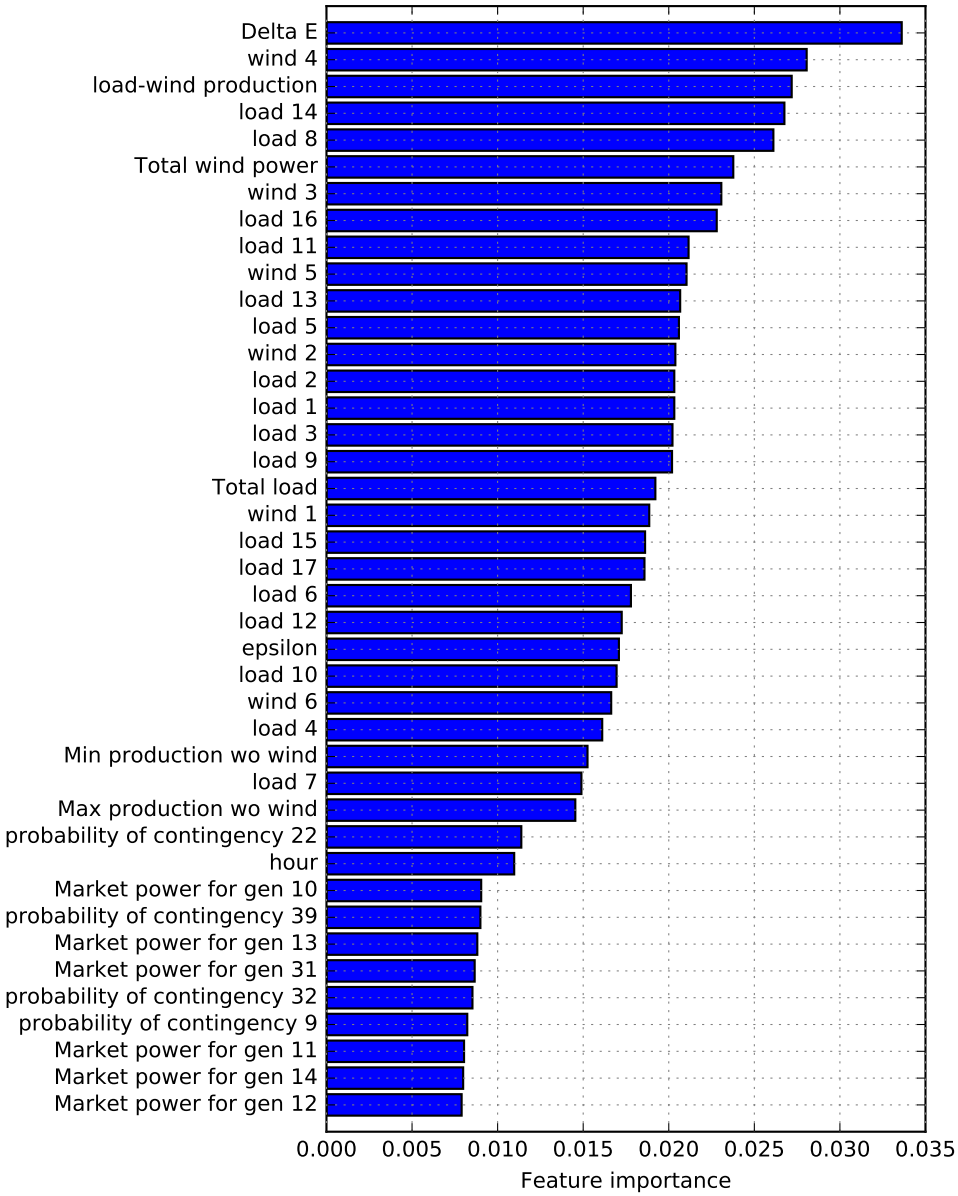



Figure 6.9: Feature importances of the 40 most important variables to predict the expected corrective control cost. The $R^2$ score equals 0.42 and the cumulative feature importance of those 40 attributes is 70.87%.

### Expected criticality

The histogram of the expected criticality can be seen in Figure 6.10. The mean and standard deviations are respectively 9028.7$ and 14567$. As for the preventive control cost, the weather has a strong impact on the value of the variable. The mean considering only the adverse weather samples is ten times larger than the mean considering only the normal weather samples. The standard deviation is also much greater: 524$ for the normal weather and 17278$ for the adverse weather.

The feature importances of the expected criticality can be seen in Figure 6.11. The results are similar to the ones of the total cost. Indeed, the most important fea-

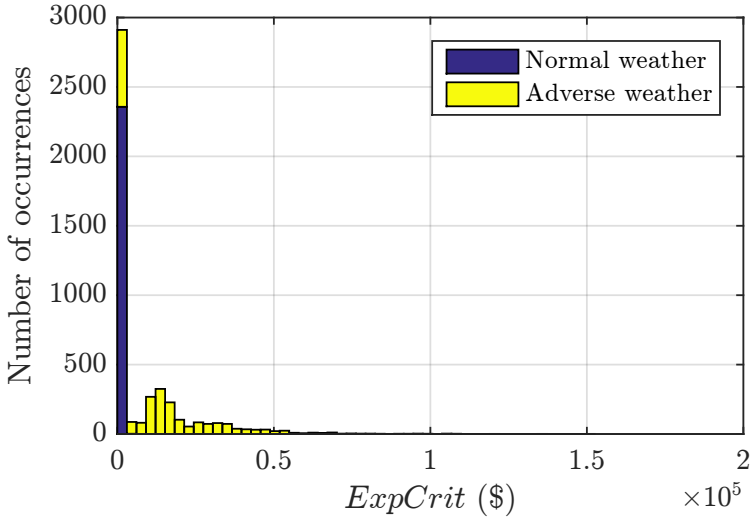



Figure 6.10: Histogram of the expected criticality. The weather status of the samples are highlighted.

tures are $\Delta E$, the weather status, the probabilities of common mode double outage contingencies and the market power output of generator 13 or 14. These two generating units are identical and on the same bus (bus 13). It is interesting to notice that the features related to the load have a relative importance of less than 0.02, despite the fact that the maximum severity is proportional to a linear combination of the loads.

The similarities between the feature importances related to the total cost and the expected criticality indicate that a large part of the total cost is due to the expected criticality. Indeed, in average the expected criticality constitutes 77% of the total cost. In order to analyse the impact of the weather on the repartition between the expected criticality and the preventive control cost, two pie charts are represented in Figure 6.12. The chart on the left corresponds to the normal weather samples and the chart on the right to the adverse weather samples. The average proportion of the preventive cost in the total cost is larger for an adverse weather. It can be explained by the fact that the probabilities of contingencies are larger and thus relaxing a contingency or using corrective control increases the probability of being in an unacceptable state more than if the weather was normal. Therefore, to meet the reliability target, more preventive control must be used.

Amongst the 40 most important features, a lot of variables are probabilities of contingencies. This is expected since the expected criticality is proportional to them. However, the first probabilities are not contingencies that are neither considered often or relaxed often. For instance, contingency 42 is selected 2451 times and relaxed 1446. However, its relative importance is only 0.01. Figure 6.13 shows the number of times each contingency is relaxed among the samples.

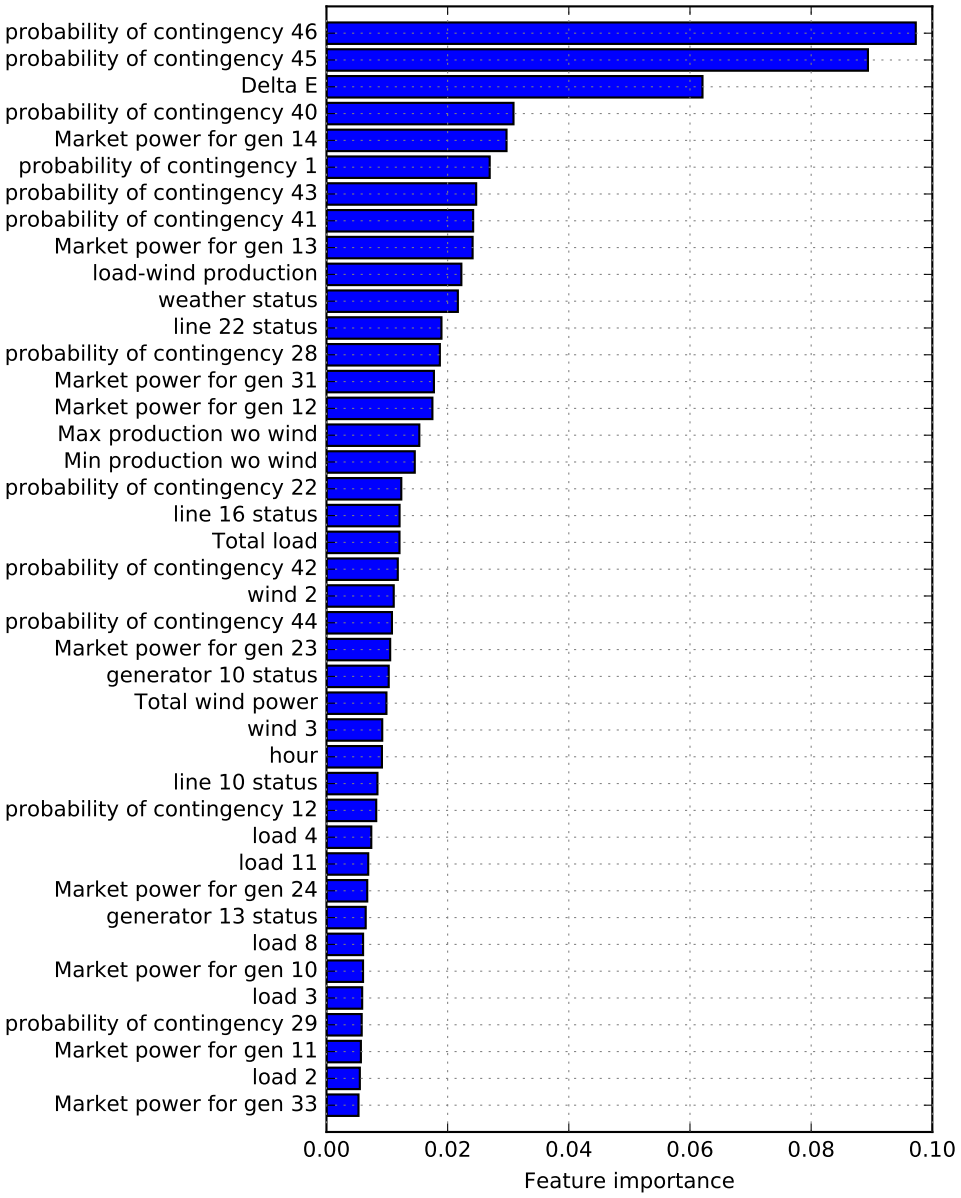



Figure 6.11: Feature importances of the 40 most important variables to predict the expected criticality. The $R^2$ score is 0.725 and the sum of the 40 importances equals 0.7563.

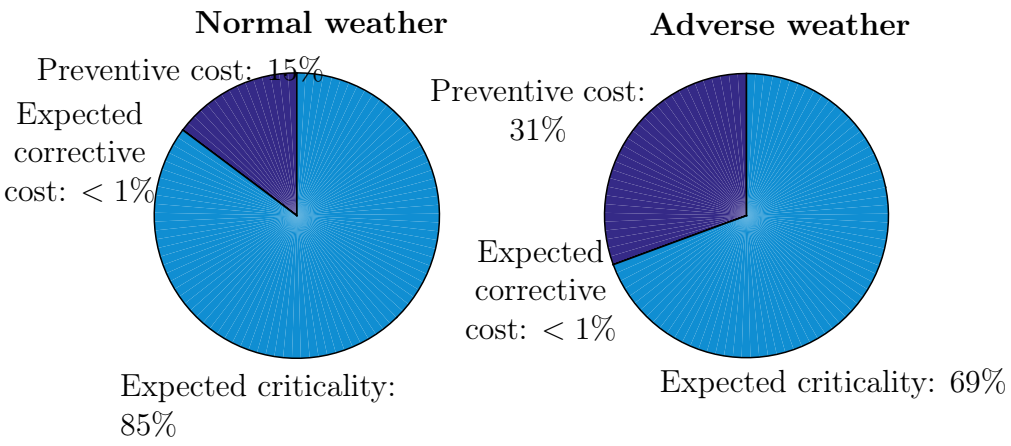



Figure 6.12: Average proportion of the different variables composing the total cost.

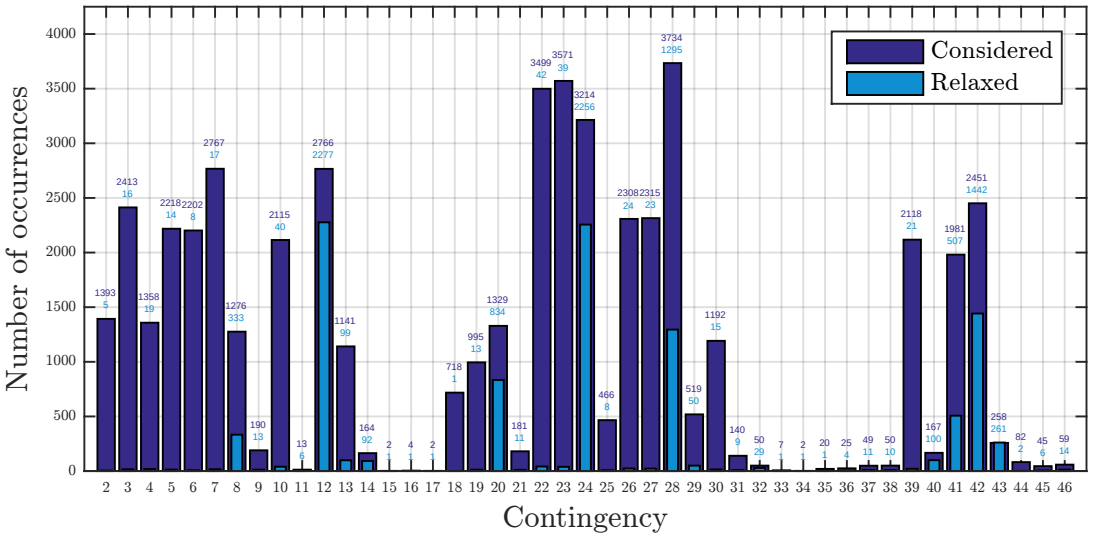



Figure 6.13: Number of times a contingency is considered and relaxed in the dataset.

### Probability of unacceptable states

The histogram of the probability of unacceptable states is represented in Figure 6.14.

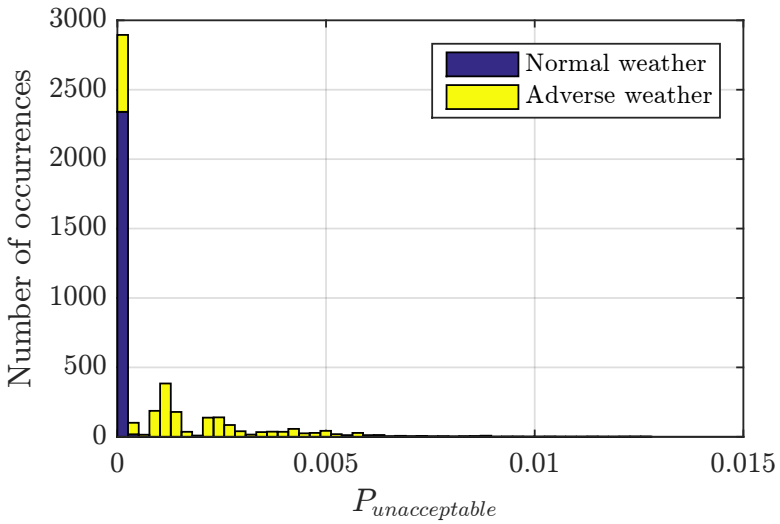



Figure 6.14: Histogram of the probability of unacceptable states. The weather status of the samples are highlighted.

The feature importances are similar to the ones of the expected criticality. It was expected given the relation between the two variables. Surprisingly, $\epsilon$ does not appear in the list of the 40 most important variables. In fact, in Figure 5.7, it can be seen that the reliability target constraint is not often active. Furthermore, $\epsilon$ is equal to 1 in one third of the samples and in that case there is no constraint on $P_{unacceptable}$. It may explain why $\epsilon$ is not an important variable to determine $P_{unacceptable}$.

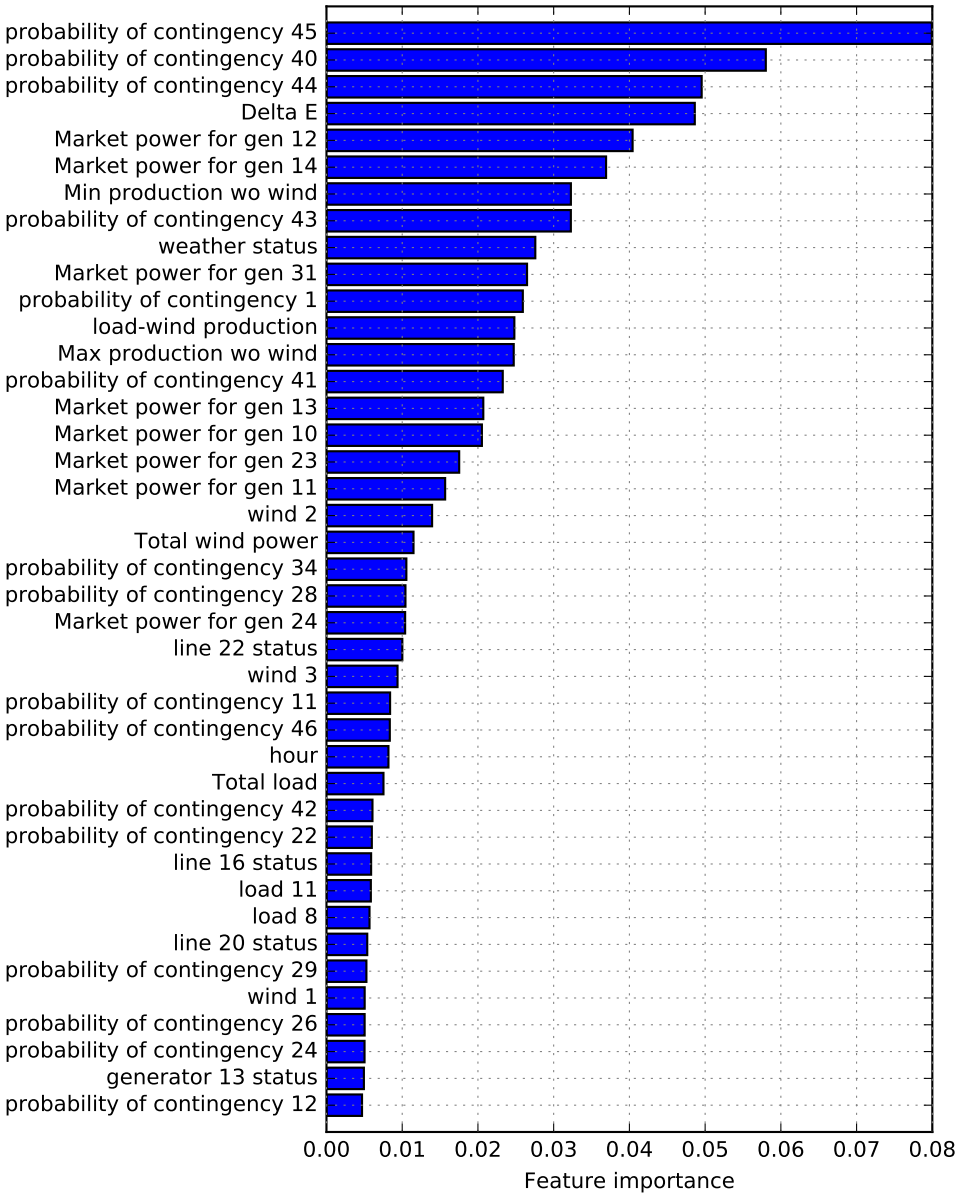



Figure 6.15: Feature importances of the 40 most important variables to predict the probability of unacceptable states. The test score is equal to 0.797 and the sum of the 40 importances is 0.7741.

### 6.3.2 Risk

The histogram of the risk after control can be seen in Figure 6.16. The shape of the distribution is similar to the one of the other output variables. The means for respectively a normal and an adverse weather are 1031.9 and 11313$. The standard deviations are 579.42$ and 16773$.

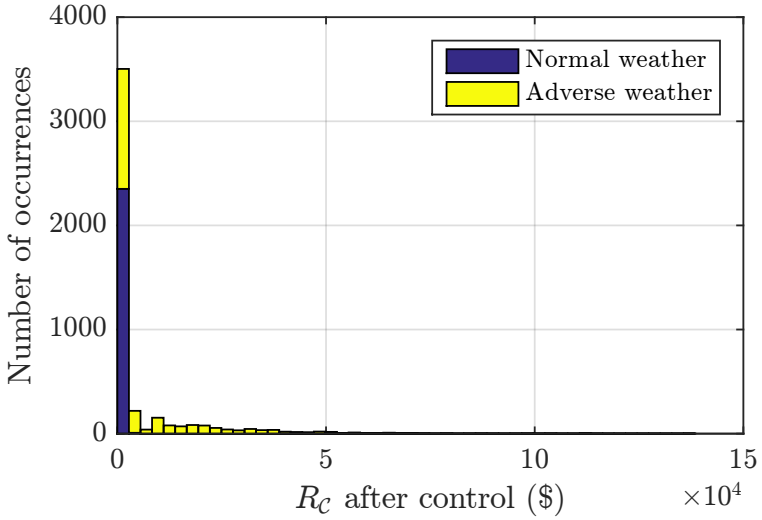



Figure 6.16: Histogram of the risk $R_C$ after control. The weather status of the samples are highlighted.

In Figure 6.17, it can be seen that the most important feature to predict the risk after control is the discarding threshold $\Delta E$. It accounts for almost one eighth of the total feature importance. The probability of contingency 44 is the second important attribute. The risk after control depends also on the difference between the system load and the wind production. The following variables are market power outputs and probabilities of contingencies. The weather does not appear in the list. It is probably because control actions tend to attenuate the impact of weather on the risk of the system.

It is interesting to notice that contingency 22 is the failure of line 16 and contingency 28 is the failure of line 22. Both line status and both probabilities of contingencies are important attributes. The lines connect the two parts of the system and therefore are important in the determination of the risk.

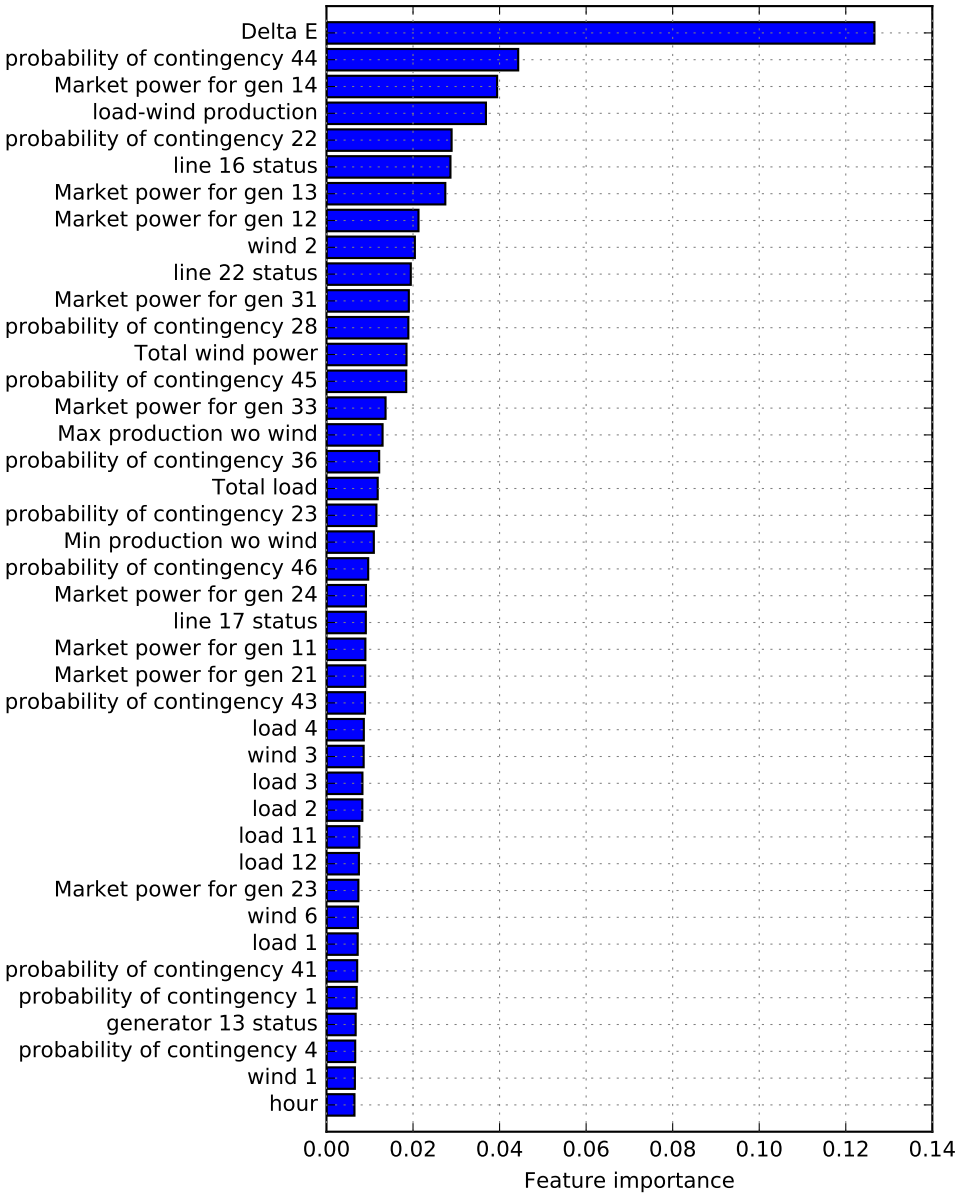



Figure 6.17: Feature importances of the 40 most important variables to predict the risk after control actions. The sum of the 40 feature importances is equal to 0.7008 and the $R^2$ score is 0.664.

### 6.3.3 Preventive control actions

The histogram of the macroscopic variable representing the preventive actions can be seen in Figure 5.29.

The 40 most important features are displayed in Figure 6.18. The first feature is the probability of contingency 24, as for the preventive control cost.

The results are similar to the ones of the preventive control cost. However, there are more probabilities of contingencies and less market clearing outputs.

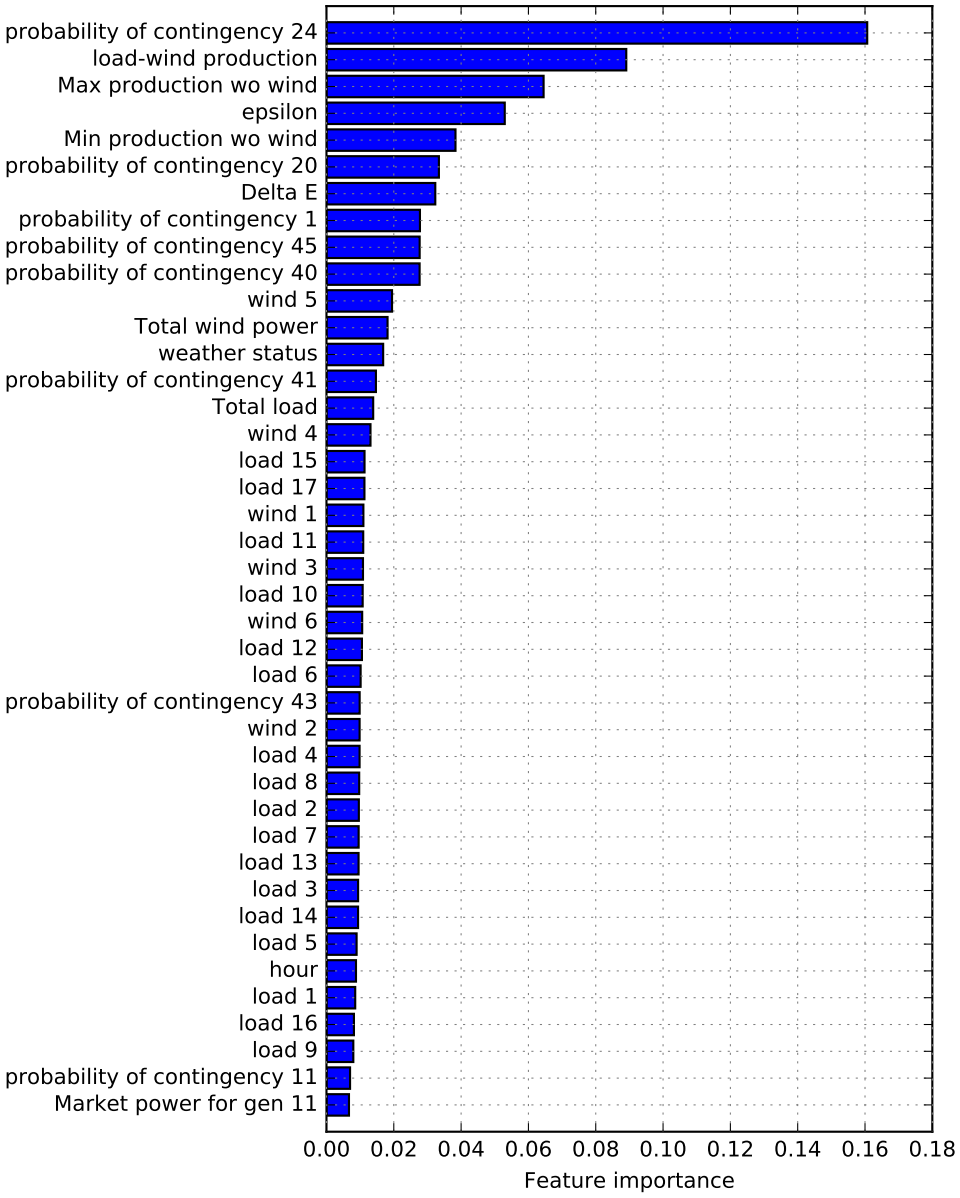



Figure 6.18: Feature importances of the 40 most important variables to predict the macroscopic preventive variable. The $R^2$ score equals 0.732 and the sum of the 40 feature importances is 0.8840.

### 6.3.4 Corrective control actions

The histograms of the macroscopic variables representing the corrective actions can be seen in Figure 5.31 for the re-dispatch and in Figure 5.33 for the adjustment of phase-shifting transformers.

The feature importances of the corrective re-dispatch variable can be seen in Figure 6.19. The production of the wind farm 1 on bus 3 is the most important variable. This means that when contingency 28 occurs, the production of this wind farm is important to determine the corrective re-dispatching. The load 3, also located on the bus 3 is the first important load. The meta-parameters $\Delta E$ and $\epsilon$

are also important to predict the amount of re-dispatching due to contingency 28. Contingency 28 is the most considered variables and is often relaxed. The fact that this contingency is considered depends on $\Delta E$ and $\epsilon$ determines the relaxation.

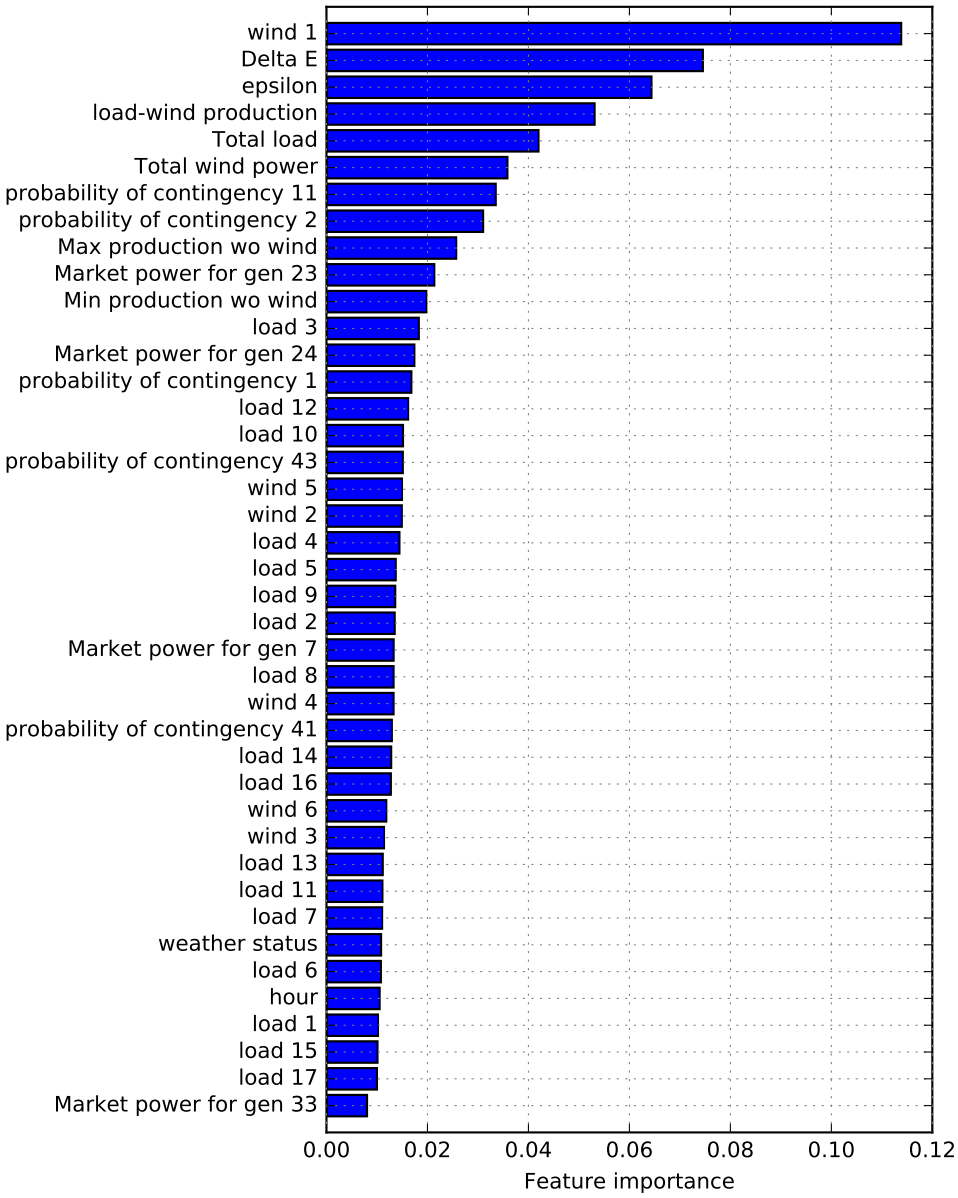



Figure 6.19: Feature importances of the 40 most important variables to predict the macroscopic corrective re-dispatch variable (for contingency 28). The sum of the 40 feature importances is 0.8965 and the $R^2$ score is 0.651.

The feature importances of the corrective adjustment of phase-shifting transformers variable can be seen in Figure 6.20. The test score is low and, as a consequence, the importances are more uniformly distributed. The important features are mostly related to load and wind production. There are also some market clearing outputs.

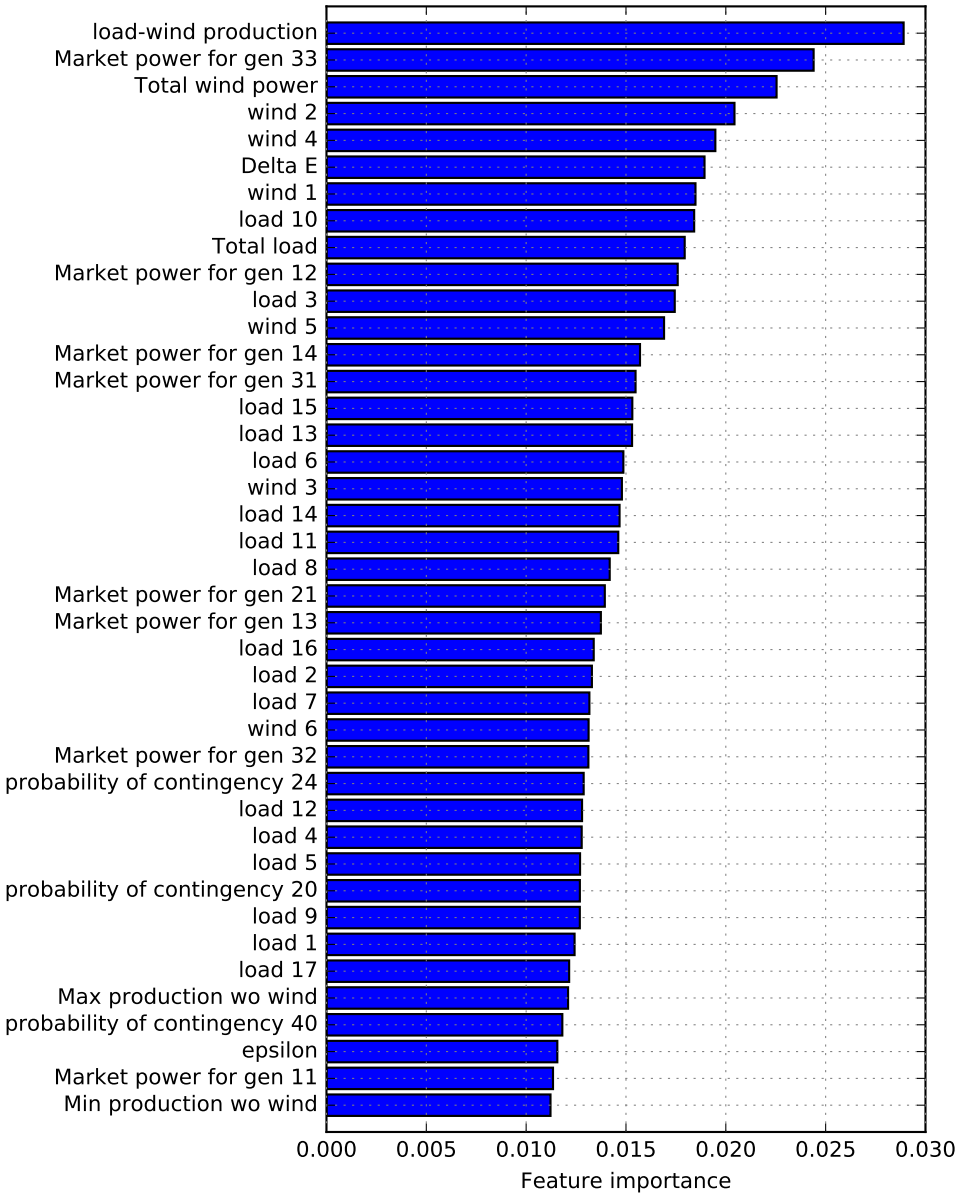



Figure 6.20: Feature importances of the 40 most important variables to predict the macroscopic corrective PST variable (for contingency 23). The sum of the 40 feature importances is 0.6181, the test score equals 0.395.

# Chapter 7

# Predictive model

This chapter is dedicated to the evaluation of the performances of different machine-learnt predictors of the target output variables described in Section 5.9.

The first section enumerates the machine learning algorithms we have used and presents the methodology applied to assess their performances. The second section presents the results obtained for each target variable and details some analyses performed for the *total cost* target variable. In particular, the influence of each meta-parameter on the predictors' performances is studied, as well as the impact of the size of the learning set. Feature selection is also tested to assess if it could improve the accuracy of the prediction. This chapter ends with a discussion about the obtained results.

## 7.1 Methodology

As in Chapter 6, the dataset is divided into two subsets: a learning set containing 80% of the samples and a test set containing the remaining 20%. The learning set and the test set are exactly the same as the ones used to find the meta-parameters for the random forest and extremely randomized trees estimators in the previous chapter.

In order to determine the best estimator to predict each target output, the following supervised learning settings are investigated:

- linear ridge regression (RR),
- kernel ridge regression with a polynomial and a gaussian radial basis function (rbf) kernel (resp. KRR poly, KRR rbf),
- support vector regression with a linear and a gaussian radial basis function (rbf) kernel (resp. SVR linear, SVR rbf),
- random forests (RF),
- extremely randomized trees (ET).

A normalisation of the features is preferable in our case for the ridge and kernel ridge regressions and the SVR models. To do so, we apply the following formula to all features before applying these algorithms:

$$x_{i,v}^{st} = \frac{x_{i,v} - \mu_v}{\sigma_v}, \tag{7.1}$$

where $x_{i,v}$ is the value of feature $v$ for the sample $i$, $\mu_v$ is the mean value of the feature $v$, considering only the learning set, and $\sigma_v$ its standard deviation in the learning set. Notice that these means and standard deviations computed on the learning set are also used to normalise the test set.

It is worth noticing that the target output is also normalised in the case of the ridge regression. It is due to the fact that the target outputs' means are generally far from zero while no intercept term is used in the formulation (2.9) to compensate. Standardising the target outputs solves the problem. This situation was not encountered with the other models and therefore, the target outputs are centred only with the ridge regression.

When possible, different combinations of meta-parameters are tested using a 5-fold cross-validation. As in chapter 6, a set of relevant combinations are tested and the selected one is the one with the highest cross-validation score (which is an $R^2$-score). In order to evaluate the best method, the learning sample is used to learn a model for each algorithm with the corresponding optimal meta-parameters. Then these models are used to predict the outputs of the test set and the best estimator is the one leading to the highest $R^2$-score on the test set.

Subsection 2.2.5 discusses the different meta-parameters that need to be tuned for each algorithm. The tested values for these meta-parameters are:

- $M$: 500, 1000 (for tree-ensemble methods),
- $n_{min}$ : 2, 3, 4, 10 (for tree ensemble methods),
- $K$ : $p$, $p/2$, $p/3$, $\sqrt{p}$, 1 (for tree ensemble methods),
- $\lambda$: $10^{-5}$, $10^{-4}$,...,$10^{7}$ (for ridge and kernel ridge regression),
- $\gamma$: $10^{-9}$, $10^{-8}$,...,$10^{4}$ (for polynomial and rbf kernel),
- $\epsilon_{SVR}$: adapted for each target output in function of their range of values:
  - for total cost, preventive control cost, expected criticality and risk: $10^{-2}$, $10^{-1}$,...,$10^{5}$,
  - for expected corrective control cost, $P_{unacceptable}$, and the macroscopic variable representing the corrective adjustment of PST: $10^{-8}$, $10^{-7}$,...,$10^{1}$,
  - for both macroscopic variables representing preventive and corrective redispatch: $10^{-2}$, $10^{-1}$,...,$10^{4}$,
- $C$: $10^{-4}$, $10^{-3}$,...,$10^{8}$ (for SVR),
- degree $d$: 2, 3, 4, 5 (for polynomial kernels).

## 7.2 Results

This section develops the results obtained for the different target outputs. The results obtained for the total cost are developed in detail. For the other target outputs, only the final results are presented.

### 7.2.1 Total cost target output

Regarding the total cost, the optimal meta-parameters for each estimator and the corresponding scores are first presented. Then the influence of each tuning parameter on the learning and CV scores is studied. After that, considering the large number of features and the low relevance of some of them, the next step is to evaluate each of these models with subsets of the feature set. We hope that it will increase the accuracy of the prediction while decreasing the computation cost. Finally, the influence of the size of the learning set on the score is presented.

#### Optimal meta-parameters

Table 7.1 presents the optimal meta-parameters for each estimator and the corresponding scores. The learning and test sets used in this study are exactly the same for all estimators. It can be seen that the best estimators are the random forest and the extremely randomized trees. A scatter plot of the predicted values obtained with the best estimator vs. the true values of the test set is shown in Figure 7.1.

Table 7.1: Best meta-parameters and corresponding scores for each estimator used to predict the total cost. The learning and test sets used in this study are exactly the same for all estimators.

| Estimator | Parameters | Train score | CV score | Test score |
|---|---|---|---|---|
| RF | $M = 1000$, $K = p/3$, $n_{min} = 2$ | 0.9638 | 0.7271 | 0.7534 |
| ET | $M = 1000$, $K = p/2$, $n_{min} = 2$ | 1 | 0.7653 | 0.7731 |
| RR | $\lambda = 100$ | 0.5235 | 0.5299 | 0.5361 |
| RBF KRR | $\lambda = 0.01$ $\gamma = 10^{-3}$ | 0.9434 | 0.7185 | 0.7288 |
| poly KRR | $\lambda = 10$, $d = 4$, $\gamma = 0.01$ | 0.9613 | 0.7447 | 0.7462 |
| linear SVR | $\epsilon_{SVR} = 10000$, $C = 1000$ | 0.5343 | 0.4903 | 0.5268 |
| RBF SVR | $\epsilon_{SVR} = 1000$, $C = 10^6$, $\gamma = 10^{-3}$ | 0.9434 | 0.7017 | 0.7288 |

To observe the impact of each meta-parameter on the score, we analysed the evolution of the score when modifying one meta-parameter. The values of the other meta-parameters are the optimal ones that can be found in Table 7.1. We limit the analysis to the train scores and the cross-validation scores.

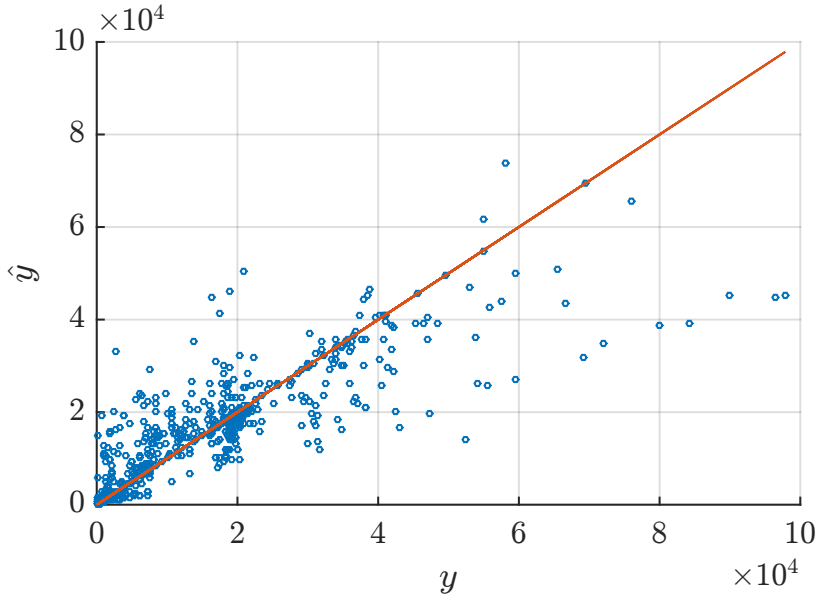



Figure 7.1: True value of the total cost ($y$) against predicted value ($\hat{y}$) with the extremely randomized trees. The samples belong to the test set.

The first meta-parameter studied is $K$, the maximum number of features selected randomly at each node in the forest algorithms. It can be seen in Figure 7.2 that the score is relatively stable for $K = p, \frac{p}{2}, \frac{p}{3}$ and decreases for smaller $K$.

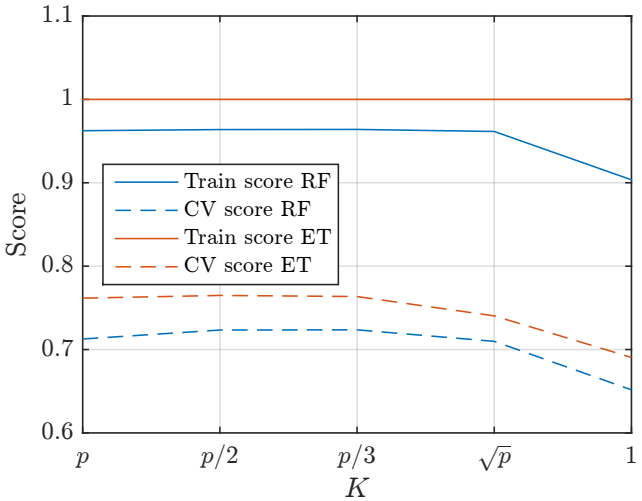



Figure 7.2: Impact of the meta-parameter $K$ on the training and CV scores to predict the total cost.

Figure 7.3 shows the impact of the number of trees $M$ on the performances of the model. It can be seen that from $M = 100$, the performances of the model do not increase significantly.

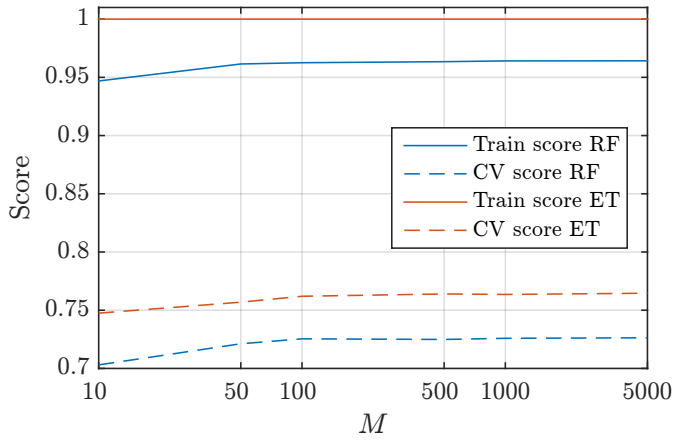



Figure 7.3: Impact of the meta-parameter $M$ on the training and CV scores to predict the total cost.

About the minimum number of samples to split a node, Figure 7.4 shows that the CV score decreases when $n_{min}$ is greater than 5. Therefore, it is possible to decrease the complexity of the model by setting $n_{min} = 5$ while keeping an acceptable level of performances.

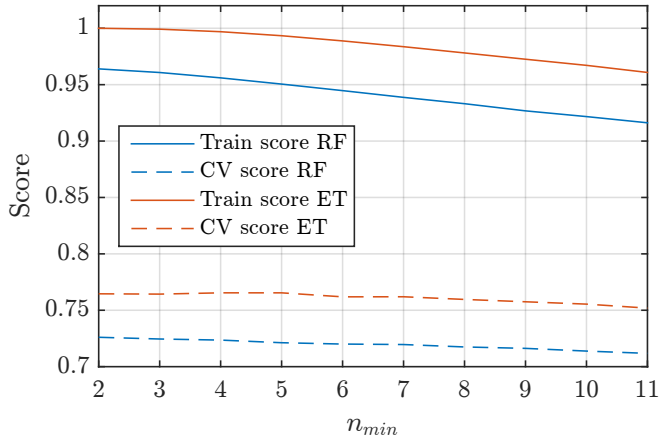



Figure 7.4: Impact of the meta-parameter $n_{min}$ on the training and CV scores to predict the total cost.

Depending on the kernel, the meta-parameter $\lambda$ has a different impact on the scores. It can be seen in Figure 7.5 that for a ridge regression, when $\lambda$ is smaller than 0.01, the CV score is negative. It is almost constant in the range 0.1-1000 and decreases for larger $\lambda$. For the rbf and the polynomial kernel, there is a small peak at respectively $\lambda = 0.01$ and $\lambda = 10$ and no negative scores in the range of values studied.

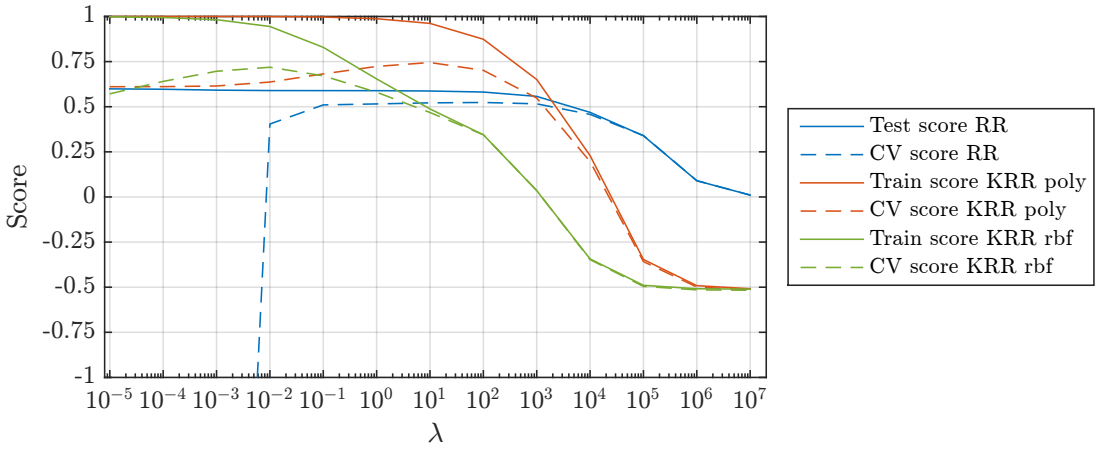



Figure 7.5: Impact of the meta-parameter $\lambda$ on the training and CV scores to predict the total cost.

If we observe the impact of $\epsilon_{SVR}$ on the SVR estimators, we see that it has a negligible effect, except when it becomes too large (greater than $10^{-4}$). For all values, the rbf SVR outperforms the linear SVR.

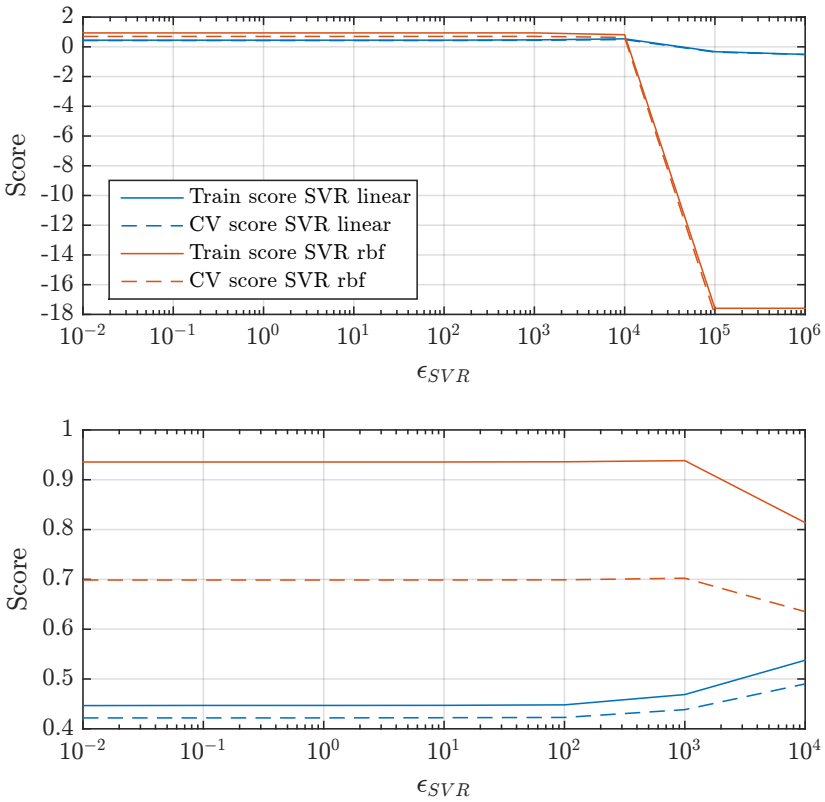



Figure 7.6: Impact of the meta-parameter $\epsilon_{SVR}$ on the training and CV scores to predict the total cost. The graph at the bottom is a zoom of the graph at the top.

Figure 7.7 shows the impact of the meta-parameter $C$ on the SVR estimators. For $C$ smaller than 10, the scores are negative. The rbf SVR is better than linear SVR when C is sufficiently large.

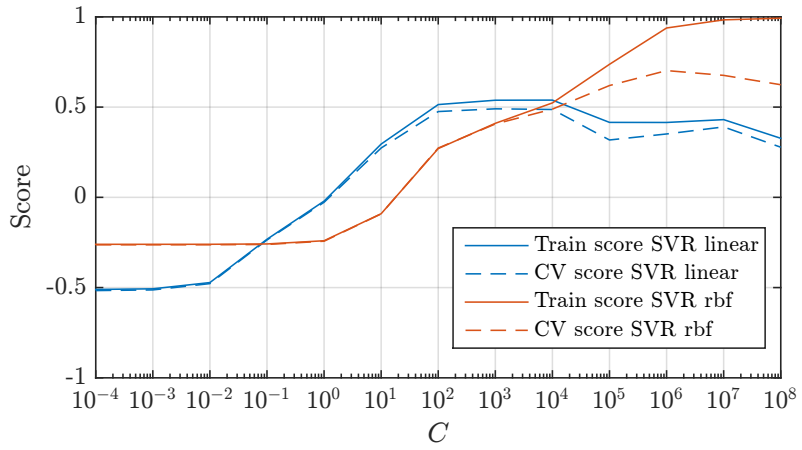



Figure 7.7: Impact of the meta-parameter $C$ on the training and CV scores to predict the total cost.

$\gamma$ is a kernel characteristic. As can be seen in Figure 7.8, the impact of $\gamma$ on the performances of the different estimators concerned is great. For all estimators, the best values are between $10^{-3}$ and $10^{-2}$.

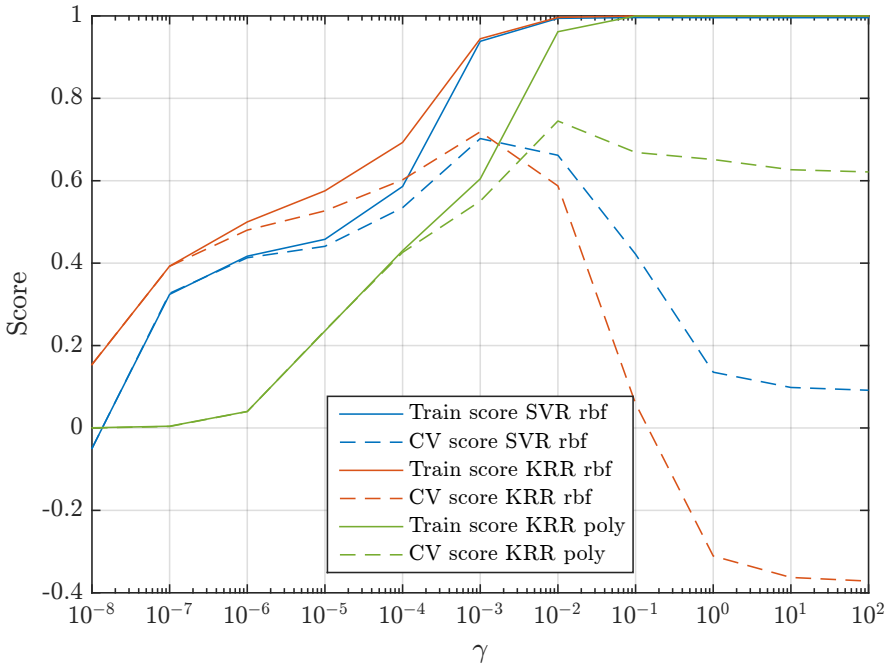



Figure 7.8: Impact of the meta-parameter $\gamma$ on the training and CV scores to predict the total cost.

The impact of the degree $d$ of the polynomial kernel can be seen in Figure 7.9. The optimal degree is 4.

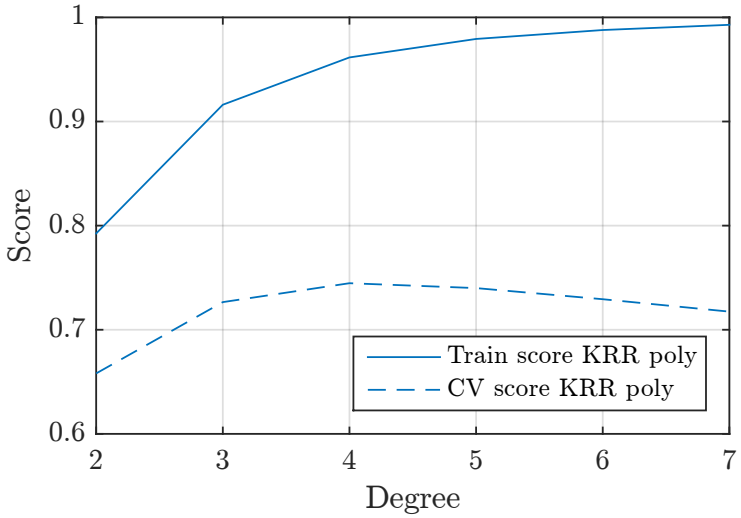



Figure 7.9: Impact of the degree of the polynomial kernel on the training and CV scores to predict the total cost.

### Subsets of features

In this subsection, we study the impact of selecting a subset of features for learning. The principle is the following. First of all, the feature importances determined in chapter 6 are used to establish a feature ranking. Then, a subset of the best features is determined, in such a way that the cumulated importance exceeds a given threshold, called importance threshold. The subset is used to learn the different models with the meta-parameters determined in Table 7.1. The results for different importance threshold are displayed in Figure 7.10.

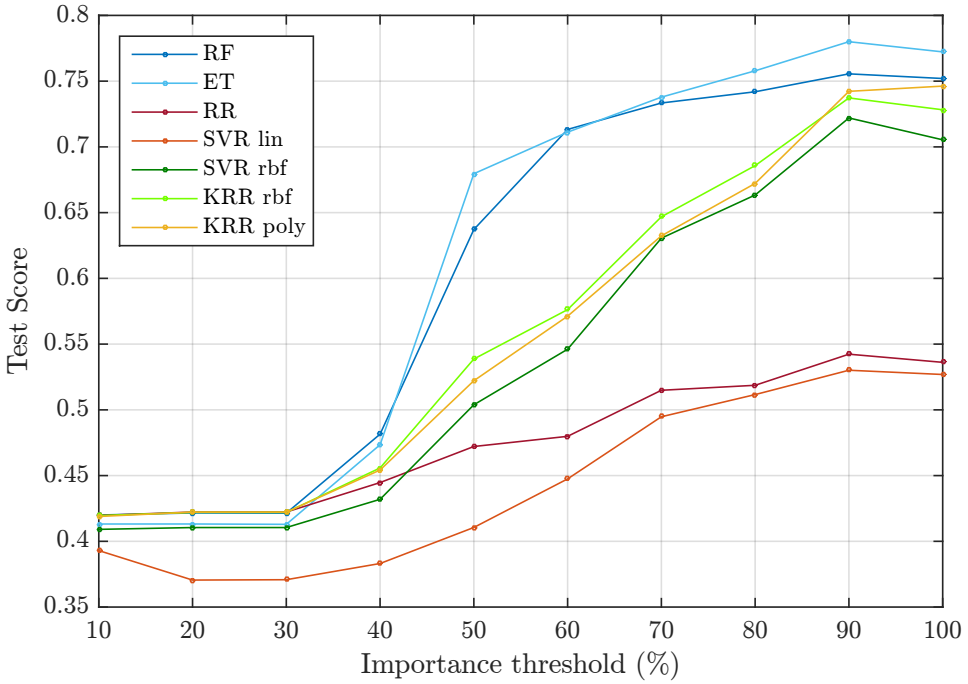



Figure 7.10: Results of feature selection for the target output total cost. The number of selected features per threshold of cumulated importance are [1, 2, 2, 5, 9, 17, 29, 47, 74, 163].

It can be seen that choosing a smaller number of features does not improve significantly the test score. However, the performances of several models are increased with a subset of 74 features (corresponding to an importance threshold of 90%). The extremely randomized trees model is still the best model. On can notice that tree ensemble methods are clearly better with a smaller subset of features.

Regarding the computation cost, we compared the time needed to learn two extremely randomized trees models. The first one is built with the feature subset corresponding to a cumulative importance of 90% and the second one with the complete set of features. We obtained respectively a learning time of 26.88s and 47.42s. Therefore, using feature selection with an importance threshold of 90% is advantageous, both for efficiency and accuracy. Note that we did not notice any significant decrease in the learning time if the 163 features corresponding to an importance threshold of 100% were used instead of the complete set.

### Influence of the size of the learning set

The CV score and train score for different estimators and sizes of the learning set can be seen in Figure 7.11. For all estimators, when the size increases, the CV score increases. Thus, in order to improve the results, the size of the database should be increased.

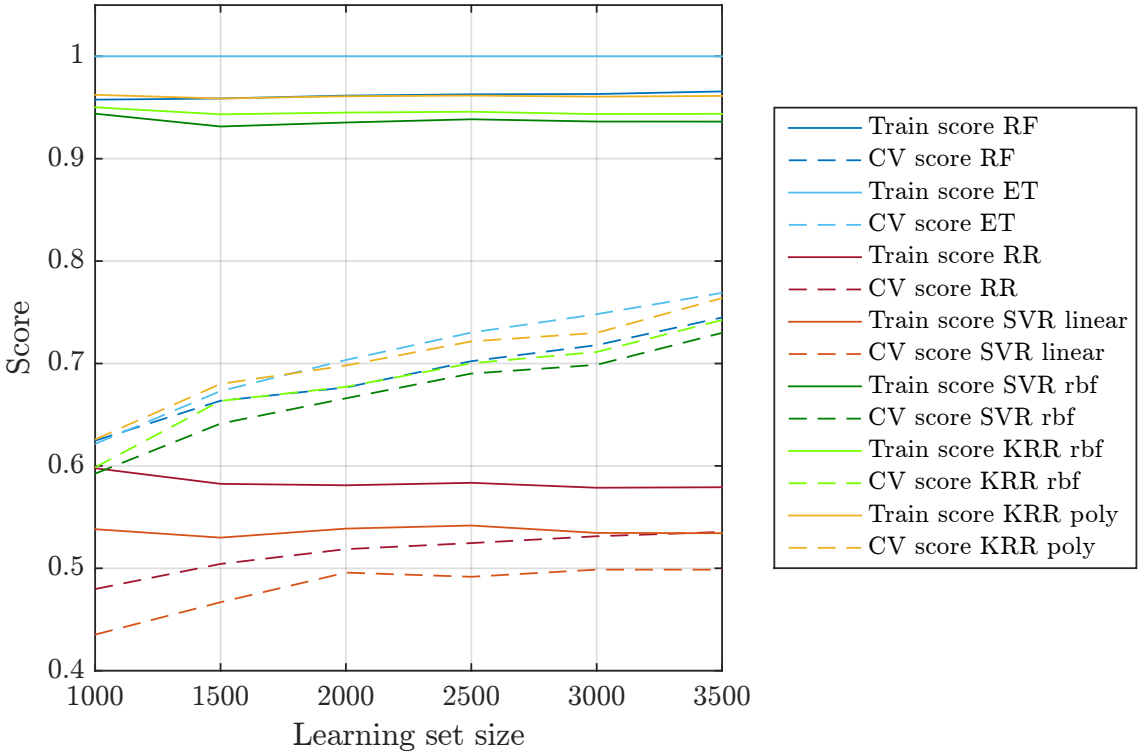



Figure 7.11: Influence of the size of the learning set on training and CV scores for the target output total cost.

### 7.2.2 Preventive control cost

The optimal meta-parameters and the corresponding scores for each estimator are presented in Table 7.2. In terms of scores, the results are similar to the one obtained for the total cost. The forests are the best estimators with a test score of more than 75%.

Table 7.2: Best meta-parameters and corresponding scores for each estimator used to predict the preventive control cost.

| Estimator | Parameters | Train score | CV score | Test score |
|---|---|---|---|---|
| RF | $M = 1000$, $K = p/2$, $n_{min} = 2$ | 0.9773 | 0.7916 | 0.7776 |
| ET | $M = 500$, $K = p$, $n_{min} = 4$ | 0.9988 | 0.7724 | 0.7980 |
| RR | $\lambda = 10$ | 0.6241 | 0.5882 | 0.5973 |
| RBF KRR | $\lambda = 10^{-2}$ $\gamma = 10^{-3}$ | 0.8846 | 0.7259 | 0.7557 |
| poly KRR | $\lambda = 10$, $d = 3$, $\gamma = 0.01$ | 0.8527 | 0.7315 | 0.7543 |
| linear SVR | $C = 100$, $\epsilon_{SVR} = 1000$ | 0.5891 | 0.5754 | 0.6105 |
| RBF SVR | $\epsilon_{SVR} = 10^3$, $C = 10^5$, $\gamma = 10^{-3}$ | 0.8283 | 0.7107 | 0.7373 |

The scatter plot comparing the true preventive control cost and the predicted preventive control cost can be seen in Figure 7.12. The set is the test set and the predictive model is the extremely randomized trees.

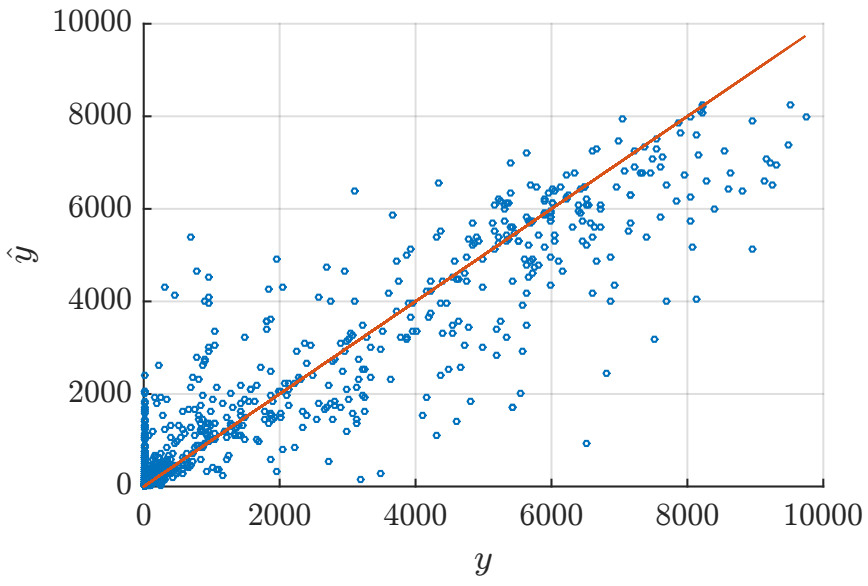



Figure 7.12: True value of the preventive control cost ($y$) against predicted value ($\hat{y}$) with the extremely randomized trees. Test score: 0.7980.

### 7.2.3 Expected corrective control cost

Table 7.3 presents the optimal tuning parameters and the corresponding scores for each estimator. This target output is not well predicted. Nevertheless, the best model is the random forest one and a scatter plot opposing the true expected cost to the predicted one for the test set is shown in Figure 7.13. The graph of the left is identical to the graph on the right, only the axis limits change to show how smaller values of the expected corrective cost are predicted.

Table 7.3: Best meta-parameters and corresponding scores for each estimator used to predict the expected corrective control cost.

| Estimator | Parameters | Train score | CV score | Test score |
|---|---|---|---|---|
| RF | $M = 500$, $K = 1$, $n_{min} = 2$ | 0.8295 | 0.2673 | 0.4081 |
| ET | $M = 500$, $K = \sqrt{p}$, $n_{min} = 2$ | 1.0000 | 0.2405 | 0.3472 |
| RR | $\lambda = 1000$ | 0.0987 | 0.0337 | 0.0522 |
| RBF KRR | $\lambda = 10^{-5}$ $\gamma = 0.1$ | 1.0000 | 0.2153 | 0.2672 |
| poly KRR | $\lambda = 1000$, $d = 5$, $\gamma = 0.01$ | 0.3858 | 0.0998 | 0.1649 |
| linear SVR | $C = 0.001$, $\epsilon_{SVR} = 0.1$ | 0.0066 | 0.0071 | 0.0055 |
| RBF SVR | $\epsilon_{SVR} = 10^{-7}$, $C = 100$, $\gamma = 0.1$ | 1.0000 | 0.2425 | 0.2905 |

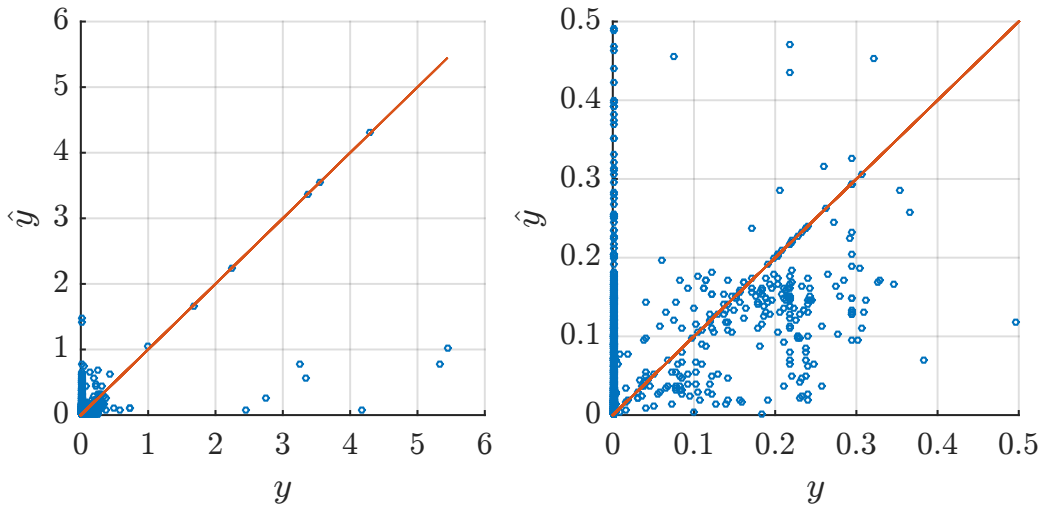



Figure 7.13: True value of the expected corrective control cost ($y$) against predicted value ($\hat{y}$) with the random forest model. Test score: 0.4081.

### 7.2.4 Expected criticality

The scores for each estimator with the optimal meta-parameters are shown in Table 7.4. Figure 7.14 is a scatter plot opposing the predicted value of the expected criticality with the extremely randomized trees estimator to the true value. The samples come from the test set.

Table 7.4: Best meta-parameters and corresponding scores for each estimator used to predict the expected criticality.

| Estimator | Parameters | Train score | CV score | Test score |
|---|---|---|---|---|
| RF | $M = 500$, $K = p/3$, $n_{min} = 3$ | 0.9620 | 0.7118 | 0.7301 |
| ET | $M = 500$, $K = p/2$, $n_{min} = 5$ | 0.9520 | 0.7443 | 0.7445 |
| RR | $\lambda = 100$ | 0.5202 | 0.4596 | 0.4663 |
| RBF KRR | $\lambda = 0.01$, $\gamma = 0.001$ | 0.9375 | 0.6876 | 0.6990 |
| poly KRR | $\lambda = 10$, $d = 4$, $\gamma = 0.01$ | 0.9571 | 0.7164 | 0.7136 |
| linear SVR | $\epsilon_{SVR} = 10000$, $C = 1000$ | 0.4559 | 0.4111 | 0.4326 |
| RBF SVR | $\epsilon_{SVR} = 1000$, $C = 10^6$, $\gamma = 10^{-3}$ | 0.9311 | 0.6679 | 0.6716 |

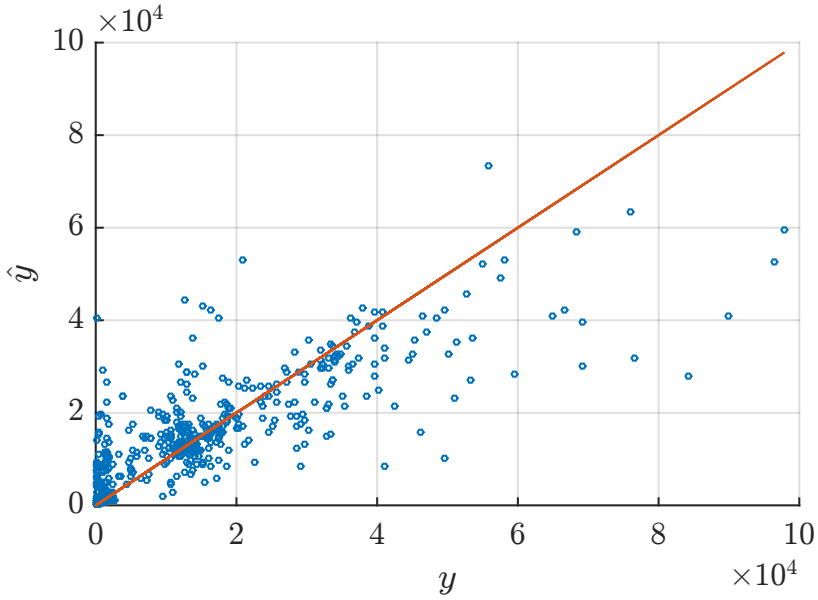



Figure 7.14: True value of the expected criticality ($y$) against predicted value ($\hat{y}$) with the extremely randomized trees. Test score: 0.7445.

### 7.2.5 Probability of unacceptable states

As can be seen in Table 7.5, the best estimator is the extremely randomized trees but it does not outperform significantly the random forest estimator. A scatter plot showing the values of probability of unacceptable states in the test set and the corresponding predicted probabilities with the best estimator is shown in Figure 7.15.

Table 7.5: Best meta-parameters and corresponding scores for each estimator used to predict the probability of unacceptable states.

| Estimator | Parameters | Train score | CV score | Test score |
|---|---|---|---|---|
| RF | $M = 500$, $K = p/3$, $n_{min} = 3$ | 0.9623 | 0.7590 | 0.7842 |
| ET | $M = 500$, $K = p/2$, $n_{min} = 3$ | 0.9963 | 0.7847 | 0.7954 |
| RR | $\lambda = 100$ | 0.5258 | 0.4753 | 0.4861 |
| RBF KRR | $\lambda = 0.01$, $\gamma = 0.001$ | 0.9369 | 0.7332 | 0.7531 |
| poly KRR | $\lambda = 10$, $d = 4$, $\gamma = 0.01$ | 0.9558 | 0.7587 | 0.7647 |
| linear SVR | $\epsilon_{SVR} = 0.001$, $C = 0.001$ | 0.4559 | 0.4336 | 0.4326 |
| RBF SVR | $\epsilon_{SVR} = 10^{-5}$, $C = 0.1$, $\gamma = 0.001$ | 0.9208 | 0.7112 | 0.7306 |

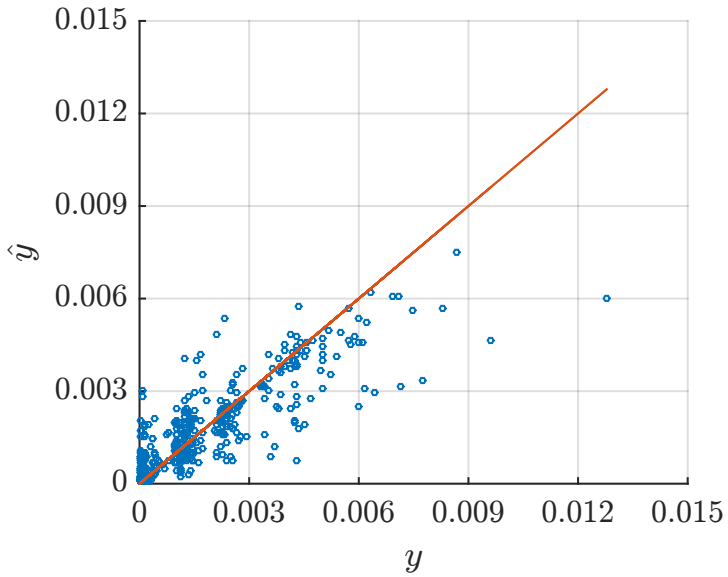



Figure 7.15: True value of the probability of unacceptable states ($y$) against predicted value ($\hat{y}$) with the extremely randomized trees. Test score:0.7954.

### 7.2.6 Risk after control

Table 7.6 includes the scores obtained for each estimator with the optimal tuning parameters. The best estimator is again a tree-based ensemble method: the extremely randomized trees. A scatter plot of the predicted values of the risk with this estimator compared to the true values can be seen in Figure 7.16.

Table 7.6: Best meta-parameters and corresponding scores for each estimator used to predict risk $R_{\mathcal{C}}$ after control.

| Estimator | Parameters | Train score | CV score | Test score |
|---|---|---|---|---|
| RF | $M = 1000$, $K = p/3$, $n_{min} = 4$ | 0.9460 | 0.6432 | 0.6292 |
| ET | $M = 500$, $K = p/2$, $n_{min} = 4$ | 0.9965 | 0.6807 | 0.6868 |
| RR | $\lambda = 100$ | 0.4878 | 0.4164 | 0.4014 |
| RBF KRR | $\lambda = 0.01$, $\gamma = 0.001$ | 0.9070 | 0.6095 | 0.5884 |
| poly KRR | $\lambda = 10$, $d = 4$, $\gamma = 0.01$ | 0.9311 | 0.6277 | 0.6059 |
| linear SVR | $\epsilon_{SVR} = 10000$, $C = 100$ | 0.4292 | 0.3840 | 0.3978 |
| RBF SVR | $\epsilon_{SVR} = 1000$, $C = 10^6$, $\gamma = 0.001$ | 0.8873 | 0.5823 | 0.5609 |

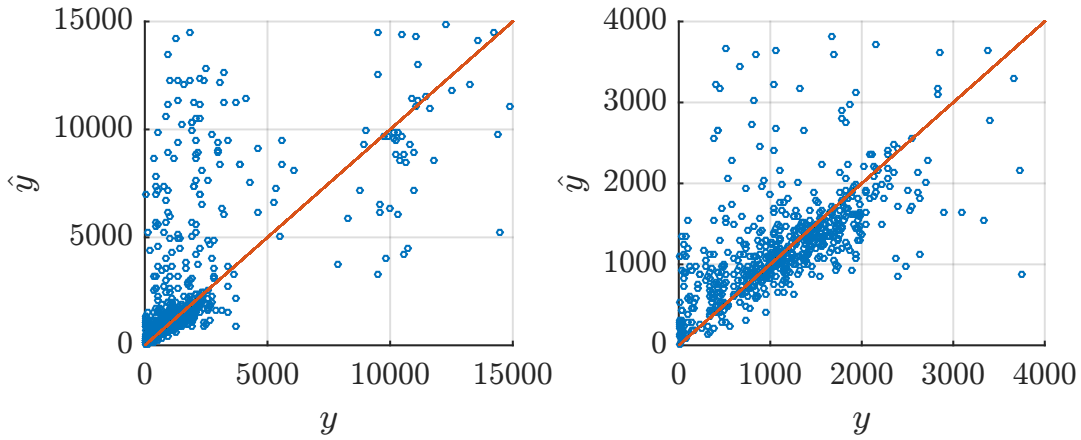



Figure 7.16: True value of the probability of the risk $R_{\mathcal{C}}$ after control ($y$) against predicted value ($\hat{y}$) with the extremely randomized trees. Test score: 0.6868. The graph on the right is the same as the one on the left, except for the axis limits.

### 7.2.7 Macroscopic preventive dispatch variable

Table 7.7 contains the optimal meta-parameters for each estimator as well as the corresponding train, CV and test scores. The random forest regressor is the best model. In that case, the tree-based ensemble methods clearly outperform the other estimators. In Figure 7.17, one can observe the predicted values of the macroscopic preventive variables opposed to the true values from the test set.

Table 7.7: Best meta-parameters and corresponding scores for each estimator used to predict the macroscopic preventive variable.

| Estimator | Parameters | Train score | CV score | Test score |
|---|---|---|---|---|
| RF | $M = 500$, $K = p/2$, $n_{min} = 2$ | 0.9610 | 0.7001 | 0.7373 |
| ET | $M = 1000$, $K = p$, $n_{min} = 2$ | 0.9999 | 0.6667 | 0.7140 |
| RR | $\lambda = 100$ | 0.4900 | 0.4533 | 0.4695 |
| RBF KRR | $\lambda = 0.01$, $\gamma = 0.001$ | 0.7907 | 0.5719 | 0.5918 |
| poly KRR | $\lambda = 10$, $d = 3$, $\gamma = 0.01$ | 0.7464 | 0.5785 | 0.5902 |
| linear SVR | $\epsilon_{SVR} = 10$, $C = 100$ | 0.4766 | 0.4326 | 0.4379 |
| RBF SVR | $\epsilon_{SVR} = 10$, $C = 100$, $\gamma = 0.01$ | 0.7361 | 0.5233 | 0.5451 |

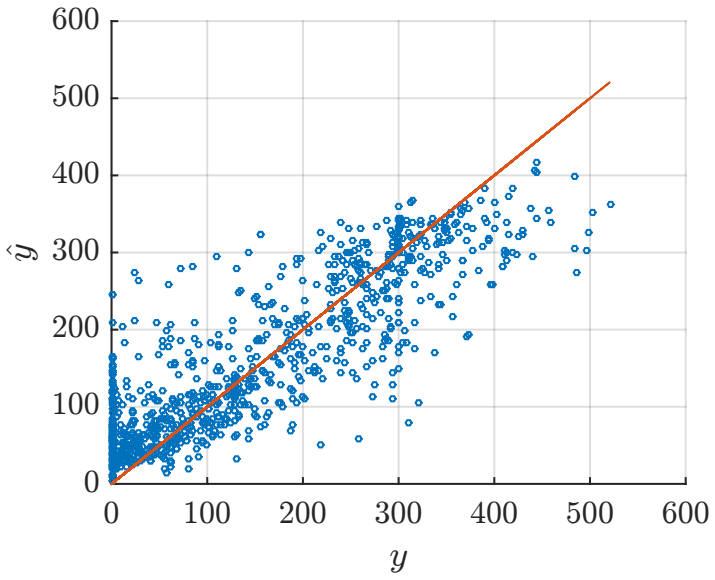



Figure 7.17: True value of the probability of the macroscopic preventive dispatch variable ($y$) against predicted value ($\hat{y}$) with the random forest model. Test score: 0.7373.

### 7.2.8 Macroscopic corrective dispatch variable

Table 7.8 presents the optimal tuning parameters and the scores of each estimator. Tree-based ensemble methods are the best estimators and in particular the extremely randomized trees. In Figure 7.18, a scatter plot comparing the predicted values of the macroscopic corrective dispatch variable for contingency 28 to the true values can be seen. The samples are from the test set and are predicted with the extremely randomized trees model.

Table 7.8: Best meta-parameters and corresponding scores for each estimator used to predict the macroscopic corrective dispatch variable.

| Estimator | Parameters | Train score | CV score | Test score |
|---|---|---|---|---|
| RF | $M = 500$, $K = p/2$, $n_{min} = 2$ | 0.9541 | 0.6520 | 0.5987 |
| ET | $M = 500$, $K = p$, $n_{min} = 4$ | 0.9946 | 0.5950 | 0.5368 |
| RR | $\lambda = 100$ | 0.3258 | 0.2896 | 0.2882 |
| RBF KRR | $\lambda = 0.1$, $\gamma = 0.01$ | 0.7884 | 0.4602 | 0.4351 |
| poly KRR | $\lambda = 100$, $d = 5$, $\gamma = 0.01$ | 0.5941 | 0.4347 | 0.4437 |
| linear SVR | $\epsilon_{SVR} = 10$, $C = 100$ | 0.1278 | 0.1289 | 0.1452 |
| RBF SVR | $\epsilon_{SVR} = 10$, $C = 100$, $\gamma = 0.01$ | 0.6210 | 0.4221 | 0.4318 |

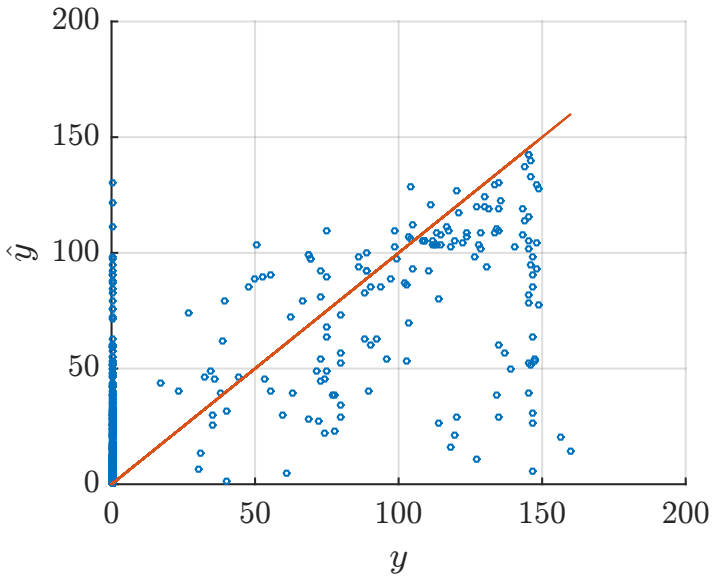



Figure 7.18: True value of the probability of the macroscopic preventive dispatch variable ($y$) against predicted value ($\hat{y}$) with the extremely randomized trees. Test score: 0.5987.

### 7.2.9 Macroscopic corrective PST variable

Finally, Table 7.9 summarises the results obtained for the macroscopic variable representing the corrective adjustment of PST for contingency 23. As for the expected corrective control cost, this variable is difficult to predict. The best estimator is the extremely randomized trees model and a scatter plot of the true values *vs.* the predicted values of the test set is shown in Figure 7.19.

Table 7.9: Best meta-parameters and corresponding scores for each estimator used to predict the macroscopic corrective PST variable.

| Estimator | Parameters | Train score | CV score | Test score |
|---|---|---|---|---|
| RF | $M = 1000$, $K = sqrtp$, $n_{min} = 2$ | 0.9196 | 0.4048 | 0.3877 |
| ET | $M = 500$, $K = \sqrt{p}$, $n_{min} = 2$ | 1.0000 | 0.4156 | 0.4106 |
| RR | $\lambda = 100$ | 0.2753 | 0.2204 | 0.2212 |
| RBF KRR | $\lambda = 1$, $\gamma = 0.01$ | 0.6540 | 0.3417 | 0.3458 |
| poly KRR | $\lambda = 100$, $d = 4$, $\gamma = 0.01$ | 0.6101 | 0.3353 | 0.3357 |
| linear SVR | $\epsilon_{SVR} = 0.01$, $C = 0.1$ | 0.1550 | 0.1457 | 0.1350 |
| RBF SVR | $\epsilon_{SVR} = 0.01$, $C = 0.1$, $\gamma = 0.01$ | 0.5399 | 0.3290 | 0.3180 |

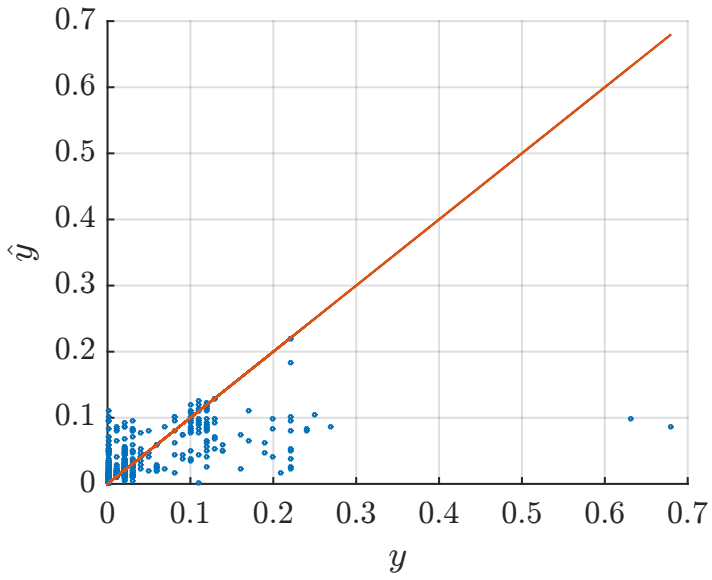



Figure 7.19: True value of the probability of the macroscopic preventive dispatch variable ($y$) against predicted value ($\hat{y}$) with the extremely randomized trees. Test score:0.4106.

## 7.3 Discussion

From the previous section, several conclusions can be drawn. First of all, tree ensemble methods outperform the other methods tested and most of the time, the best estimator is the extremely randomized trees model. Furthermore, linear models lead to the worst results for all the target variables. Ridge regression often outperforms SVR for similar kernels. For the target outputs with a test score greater than 70%, the kernel ridge regression with a polynomial kernel is often the best alternative to the forest algorithms.

The maximum $R^2$ score was obtained for the preventive control cost and was 0.7980. From the different scatter plots, one can observe that most samples are not correctly predicted. However, the cloud of points seem to follow the "$\hat{y} = y$" line. Three target variables in particular are not well predicted with any considered estimator. These three variables are related to the corrective control actions. The macroscopic variable accounting for the corrective re-dispatch has a slightly better score than the two others.

From the analysis of the tuning meta-parameters realised for the total cost, it can be concluded that a number of trees greater than 100 does not improve significantly the score. Furthermore, the minimum number of samples to split a node ($n_{min}$) could also be increased until 5 without losing accuracy. All of this could reduce the complexity of the model without notably modifying the performances of the estimator.

Figure 7.10 indicates that feature selection does not improve significantly the score for the total cost. However, it is possible that the tuning parameters are not optimal for the subsets of features. In a second approach, it could be interesting to verify if the scores of the rbf and polynomial kernel methods do not improve with other values of meta-parameters.

Finally, the performances of all the models increase when the size of the learning set rises (Fig 7.11). Therefore, in order to improve the accuracy of the prediction, more samples should be generated.

# Chapter 8

# Conclusions and future work

In this master's thesis, we have studied the probabilistic reliability criterion developed for real-time operation in the framework of the GARPUR project, from the viewpoint of day-ahead operation planning.

With the statistical analysis performed in Chapter 5, we compared a variant of the currently used deterministic criterion ("pseudo" N-1 criterion) to the RMAC proposed by GARPUR. Two observations can be drawn. First of all, for an equivalent risk after control (representing the reliability level), the total cost is higher when using the "pseudo" N-1 criterion. Indeed, both preventive control cost and expected criticality are larger with the "pseudo" N-1 criterion. Thus, adapting the set of covered contingencies to the state of the system, as implied by the GARPUR RMAC, results in a more economical solution for an equivalent risk. Secondly, when looking at the set of covered contingencies by the RMAC, it was noticed that not all single contingencies were selected at the same time. Moreover, at least one common mode double outage contingency was considered in 72% of the cases. This tends to show that the deterministic set $\mathcal{C}_{N-1}$ can be overly conservative or insufficient. Finally, the analysis of the number of times a contingency is selected emphasises the importance of some transmission lines, in particular the ones connecting the area where most of the load is present to the one containing the large generating units.

In Chapter 6, an analysis of the relevance of some features was performed. It highlights the importance of probabilities of contingencies in the determination of the different target variables. Other relevant variables are the discarding threshold $\Delta E$ and the net load. The weather is also essential for the prediction of most target output variables. This was confirmed by the statistical analysis. The weather information is also present in the probabilities of contingencies, emphasising even more the impact of the weather status on the results of the decision-making program.

Finally, in Chapter 7, we studied the predictive power of different estimators for this application and concluded that extremely randomized trees are for most outputs of the decision-making program the best algorithm in terms of accuracy. The best test scores ($R^2$ scores) reach 80%, meaning that we manage to explain in the best case only 80% of the variance of the target variable. Another remark is that linear models are not convenient for this application. However, kernel ridge

regression with a polynomial or a gaussian kernel had results close to the tree-based ensemble methods.

We must keep in mind that the stated conclusions depend on the assumptions made throughout this report. One possible future work is to modify some of these assumptions and verify if the conclusion are still valid. For instance, it would be relevant to use a criterion similar to the deterministic N-1 approach currently used by the European TSOs, instead of the "pseudo" N-1 criterion to compare with the proposed RMAC.

Other possible improvements would be to integrate an iterative approach in the decision-making program in order to guarantee that the discarding principle is met with the optimal control decisions, or to modify the value of lost load as a function of the weather to be closer to what happens in practice. Regarding all the infeasible cases, instead of setting $\epsilon$ to 1, we could increase the value of $\Delta E$ such that the problem is feasible and the new discarding threshold is $\Delta E + \lambda$ and consider as an output of the problem the increment $\lambda$.

Currently the same physical approximations are used both in the optimisation program (the control part) and the simulation program (pre- and post-assessment parts). Although the assumptions are necessary to simplify and speed up the optimisation problem, it would be interesting to design the assessment program with less restrictive hypotheses (for example, by considering an AC power flow instead of a DC one) in order to verify that the hypotheses made in the control part are indeed valid.

The predictive performances of the different estimators have been studied for several reasons. First of all, the decision-making program is slow and therefore using machine learning to predict the preventive and corrective control decisions and/or the resulting optimal objective function values could considerably speed-up the look-ahead reliability assessment process. It would allow to generate thousands of plausible trajectories in a look-ahead mode with Monte-Carlo simulations and then verify with proxies the control cost and the risk after control for all these trajectories.

In a second step, these proxies could be imported towards other time-horizons decision-making programs in order to take into account how operation is performed in real-time. For instance, in the context of day-ahead operation planning, proxies of real-time reliability management would allow to take into consideration the impact of day-ahead decisions on the real-time operations. For this objective in particular, tree-ensemble methods are not ideal and would require more research to include them in optimisation problems. Therefore, given that for the target outputs that are better predicted, the polynomial KRR models give results close to the tree-based ensemble methods, it would be interesting to develop this model in particular and try to improve the performances, either by increasing the size of the learning set or by refining the tuning of the meta-parameters.

# Appendix A

# Data for the IEEE-RTS96 network

This chapter contains the data used in the control and assessment programs. The first section contains a representation of the IEEE-RTS 96 system [26]. The second section provides the data for the generating units while the third and fourth sections includes the day-ahead forecast values of respectively load and wind production.

## A.1 IEEE-network

First of all, the IEEE network is represented in Figure A.1. The lines, loads, wind farms and generating units' numbering is indicated in the figure.

## A.2 Generating units data

Table A.1 gives information about the generating units of the IEEE-RTS96 system, such as the maximum and minimum production when the unit is up, the ramping rate and the unit type. These data come from [26].

## A.3 Load

In order to study the impact of prediction errors both for the load and wind generation, all the states are generated according to the predictions for one day. This day is the Tuesday of the $50^{th}$ week. It is a winter day with a high averaged demand. Table A.2 and A.3 give the predicted load per hour for respectively the $9^{th}$ first buses and the $8^{th}$ last bus loads.

## A.4 Wind power

In order to have a system with wind farms, the IEEE-RTS96 system is modified. [27] advices to add 6 wind farms with a maximum capacity of 200MW. The location of these power plants can be seen in Figure 3.2.

The day-ahead data for wind power output come from [28] that generated 100 scenarios for the 15 regions of Denmark. The data was used in [39][40]. For this

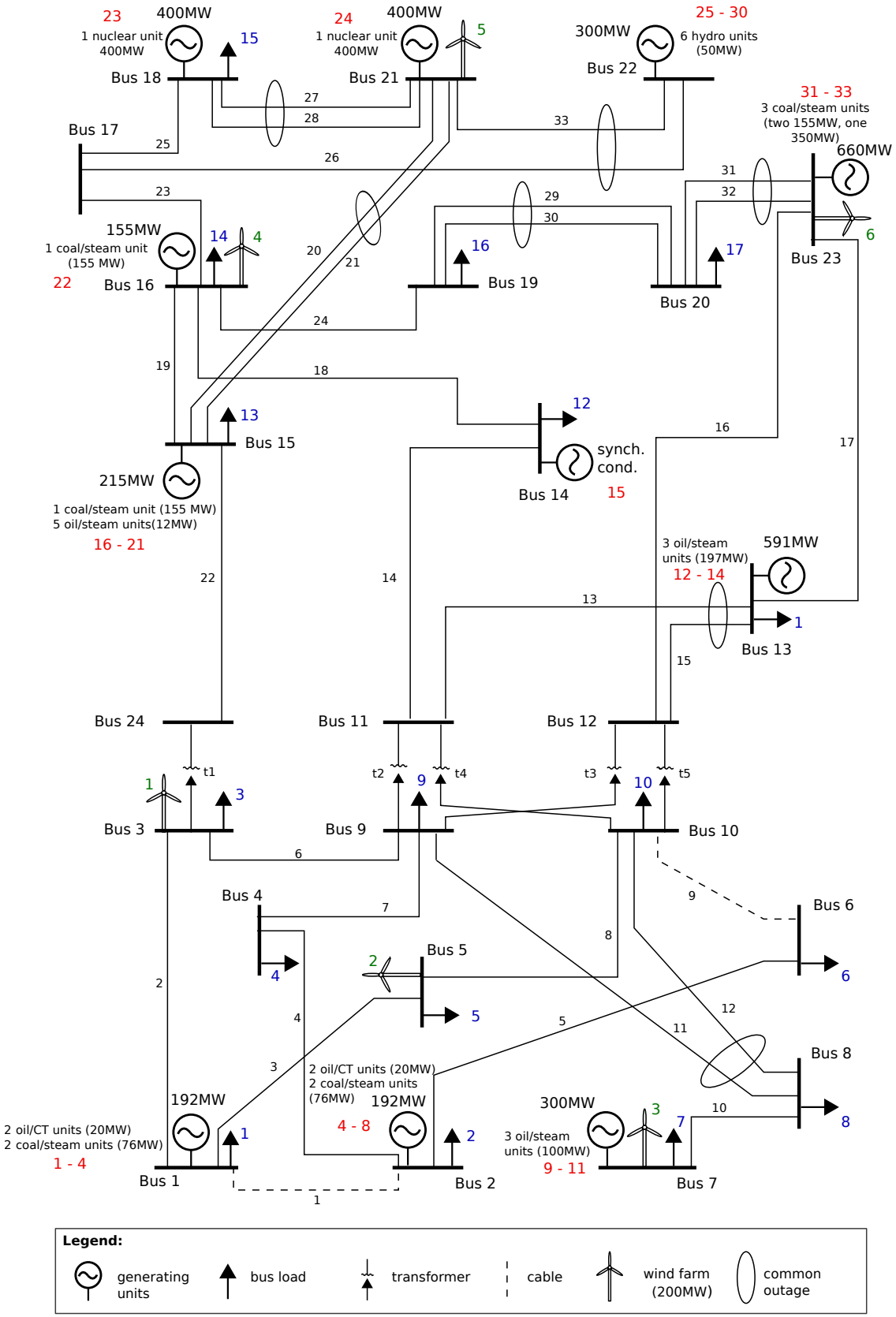



Figure A.1: IEEE RTS-96 network. The lines' numbering is indicated in black, the loads' in blue, the wind farms' in green and the generating units' in red.

Table A.1: Data for each generating unit of the IEEE-RTS96 system.

| G | Name | Unit type | Bus ID | $P_{min}$ (MW) | $P_{max}$ (MW) | Ramping rate (MW/min) |
|---|---|---|---|---|---|---|
| 1 | U20 | Oil/CT | 1 | 16 | 20 | 3 |
| 2 | U20 | Oil/CT | 1 | 16 | 20 | 3 |
| 3 | U76 | Coal/Steam | 1 | 15.2 | 76 | 2 |
| 4 | U76 | Coal/Steam | 1 | 15.2 | 76 | 2 |
| 5 | U20 | Oil/CT | 2 | 16 | 20 | 3 |
| 6 | U20 | Oil/CT | 2 | 16 | 20 | 3 |
| 7 | U76 | Coal/Steam | 2 | 15.2 | 76 | 2 |
| 8 | U76 | Coal/Steam | 2 | 15.2 | 76 | 2 |
| 9 | U100 | Oil/Steam | 7 | 25 | 100 | 7 |
| 10 | U100 | Oil/Steam | 7 | 25 | 100 | 7 |
| 11 | U100 | Oil/Steam | 7 | 25 | 100 | 7 |
| 12 | U197 | Oil/Steam | 13 | 69 | 197 | 3 |
| 13 | U197 | Oil/Steam | 13 | 69 | 197 | 3 |
| 14 | U197 | Oil/Steam | 13 | 69 | 197 | 3 |
| 15 | SynCond | | 14 | 0 | 0 | 0 |
| 16 | U12 | Oil/Steam | 15 | 2.4 | 12 | 1 |
| 17 | U12 | Oil/Steam | 15 | 2.4 | 12 | 1 |
| 18 | U12 | Oil/Steam | 15 | 2.4 | 12 | 1 |
| 19 | U12 | Oil/Steam | 15 | 2.4 | 12 | 1 |
| 20 | U12 | Oil/Steam | 15 | 2.4 | 12 | 1 |
| 21 | U155 | Coal/Steam | 15 | 54.3 | 155 | 3 |
| 22 | U155 | Coal/Steam | 16 | 54.3 | 155 | 3 |
| 23 | U400 | Nuclear | 18 | 100 | 400 | 20 |
| 24 | U400 | Nuclear | 21 | 100 | 400 | 20 |
| 25 | U50 | Hydro | 22 | 10 | 50 | 50 |
| 26 | U50 | Hydro | 22 | 10 | 50 | 50 |
| 27 | U50 | Hydro | 22 | 10 | 50 | 50 |
| 28 | U50 | Hydro | 22 | 10 | 50 | 50 |
| 29 | U50 | Hydro | 22 | 10 | 50 | 50 |
| 30 | U50 | Hydro | 22 | 10 | 50 | 50 |
| 31 | U155 | Coal/Steam | 23 | 54.3 | 155 | 3 |
| 32 | U155 | Coal/Steam | 23 | 54.3 | 155 | 3 |
| 33 | U350 | Coal/Steam | 23 | 140 | 350 | 4 |

work, the data corresponds to the fourth scenario from hour 12 to hour 35. The power output of the wind farms corresponds respectively to the power output of regions 2 to 7.

Table A.2: Forecast load per hour for the $9^{th}$ first load buses on the $50^{th}$ Tuesday of the year (winter). The data comes from [26].

| Hour | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 74.09 | 66.29 | 122.83 | 50.69 | 48.74 | 93.59 | 85.79 | 116.98 | 118.93 |
| 2 | 69.67 | 62.33 | 115.50 | 47.67 | 45.83 | 88.00 | 80.67 | 110.00 | 111.83 |
| 3 | 66.35 | 59.36 | 110.00 | 45.40 | 43.65 | 83.81 | 76.82 | 104.76 | 106.51 |
| 4 | 65.24 | 58.37 | 108.16 | 44.64 | 42.92 | 82.41 | 75.54 | 103.01 | 104.73 |
| 5 | 65.24 | 58.37 | 108.16 | 44.64 | 42.92 | 82.41 | 75.54 | 103.01 | 104.73 |
| 6 | 66.35 | 59.36 | 110.00 | 45.40 | 43.65 | 83.81 | 76.82 | 104.76 | 106.51 |
| 7 | 81.83 | 73.22 | 135.66 | 55.99 | 53.84 | 103.36 | 94.75 | 129.20 | 131.36 |
| 8 | 95.10 | 85.09 | 157.66 | 65.07 | 62.56 | 120.12 | 110.11 | 150.16 | 152.66 |
| 9 | 105.05 | 93.99 | 174.16 | 71.88 | 69.11 | 132.70 | 121.64 | 165.87 | 168.63 |
| 10 | 106.16 | 94.98 | 176.00 | 72.63 | 69.84 | 134.09 | 122.92 | 167.62 | 170.41 |
| 11 | 106.16 | 94.98 | 176.00 | 72.63 | 69.84 | 134.09 | 122.92 | 167.62 | 170.41 |
| 12 | 105.05 | 93.99 | 174.16 | 71.88 | 69.11 | 132.70 | 121.64 | 165.87 | 168.63 |
| 13 | 105.05 | 93.99 | 174.16 | 71.88 | 69.11 | 132.70 | 121.64 | 165.87 | 168.63 |
| 14 | 105.05 | 93.99 | 174.16 | 71.88 | 69.11 | 132.70 | 121.64 | 165.87 | 168.63 |
| 15 | 102.84 | 92.01 | 170.50 | 70.36 | 67.66 | 129.90 | 119.08 | 162.38 | 165.08 |
| 16 | 103.95 | 93.00 | 172.33 | 71.12 | 68.38 | 131.30 | 120.36 | 164.12 | 166.86 |
| 17 | 109.47 | 97.95 | 181.50 | 74.90 | 72.02 | 138.28 | 126.76 | 172.85 | 175.73 |
| 18 | 110.58 | 98.94 | 183.33 | 75.66 | 72.75 | 139.68 | 128.04 | 174.60 | 177.51 |
| 19 | 110.58 | 98.94 | 183.33 | 75.66 | 72.75 | 139.68 | 128.04 | 174.60 | 177.51 |
| 20 | 106.16 | 94.98 | 176.00 | 72.63 | 69.84 | 134.09 | 122.92 | 167.62 | 170.41 |
| 21 | 100.63 | 90.04 | 166.83 | 68.85 | 66.20 | 127.11 | 116.52 | 158.89 | 161.53 |
| 22 | 91.78 | 82.12 | 152.16 | 62.80 | 60.38 | 115.93 | 106.27 | 144.92 | 147.33 |
| 23 | 80.72 | 72.23 | 133.83 | 55.23 | 53.11 | 101.97 | 93.47 | 127.46 | 129.58 |
| 24 | 69.67 | 62.33 | 115.50 | 47.67 | 45.83 | 88.00 | 80.67 | 110.00 | 111.83 |

Table A.3: Forecast load per hour for the $8^{th}$ last load buses on the $50^{th}$ Tuesday of the year (winter). The data comes from [26].

| Hour | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|
| 1 | 132.58 | 181.32 | 132.58 | 216.42 | 68.24 | 228.11 | 124.78 | 87.74 |
| 2 | 124.66 | 170.50 | 124.66 | 203.50 | 64.17 | 214.50 | 117.33 | 82.50 |
| 3 | 118.73 | 162.38 | 118.73 | 193.81 | 61.11 | 204.28 | 111.74 | 78.57 |
| 4 | 116.75 | 159.67 | 116.75 | 190.58 | 60.09 | 200.88 | 109.88 | 77.26 |
| 5 | 116.75 | 159.67 | 116.75 | 190.58 | 60.09 | 200.88 | 109.88 | 77.26 |
| 6 | 118.73 | 162.38 | 118.73 | 193.81 | 61.11 | 204.28 | 111.74 | 78.57 |
| 7 | 146.43 | 200.27 | 146.43 | 239.03 | 75.37 | 251.95 | 137.82 | 96.90 |
| 8 | 170.18 | 232.74 | 170.18 | 277.79 | 87.59 | 292.80 | 160.17 | 112.62 |
| 9 | 187.99 | 257.10 | 187.99 | 306.86 | 96.76 | 323.45 | 176.93 | 124.40 |
| 10 | 189.96 | 259.80 | 189.96 | 310.09 | 97.78 | 326.85 | 178.79 | 125.71 |
| 11 | 189.96 | 259.80 | 189.96 | 310.09 | 97.78 | 326.85 | 178.79 | 125.71 |
| 12 | 187.99 | 257.10 | 187.99 | 306.86 | 96.76 | 323.45 | 176.93 | 124.40 |
| 13 | 187.99 | 257.10 | 187.99 | 306.86 | 96.76 | 323.45 | 176.93 | 124.40 |
| 14 | 187.99 | 257.10 | 187.99 | 306.86 | 96.76 | 323.45 | 176.93 | 124.40 |
| 15 | 184.03 | 251.69 | 184.03 | 300.40 | 94.72 | 316.64 | 173.20 | 121.78 |
| 16 | 186.01 | 254.39 | 186.01 | 303.63 | 95.74 | 320.04 | 175.07 | 123.09 |
| 17 | 195.90 | 267.92 | 195.90 | 319.78 | 100.83 | 337.07 | 184.38 | 129.64 |
| 18 | 197.88 | 270.63 | 197.88 | 323.01 | 101.85 | 340.47 | 186.24 | 130.95 |
| 19 | 197.88 | 270.63 | 197.88 | 323.01 | 101.85 | 340.47 | 186.24 | 130.95 |
| 20 | 189.96 | 259.80 | 189.96 | 310.09 | 97.78 | 326.85 | 178.79 | 125.71 |
| 21 | 180.07 | 246.27 | 180.07 | 293.94 | 92.68 | 309.83 | 169.48 | 119.16 |
| 22 | 164.24 | 224.62 | 164.24 | 268.10 | 84.54 | 282.59 | 154.58 | 108.69 |
| 23 | 144.45 | 197.56 | 144.45 | 235.80 | 74.35 | 248.54 | 135.96 | 95.59 |
| 24 | 124.66 | 170.50 | 124.66 | 203.50 | 64.17 | 214.50 | 117.33 | 82.50 |

Table A.4: Forecast wind power per hour on the $50^{th}$ Tuesday of the year (winter).

| Hour \ Wind farm | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 151.52 | 142.56 | 117.33 | 132.22 | 134.96 | 109.07 |
| 2 | 150.87 | 149.16 | 103.83 | 129.05 | 129.31 | 84.505 |
| 3 | 155.51 | 153.07 | 96.596 | 136.91 | 138.2 | 104.21 |
| 4 | 176.97 | 163.9 | 101.78 | 147.6 | 150.87 | 115.64 |
| 5 | 158.18 | 167.56 | 97.149 | 155.09 | 161.79 | 119.26 |
| 6 | 164.23 | 158.88 | 105.93 | 146.22 | 149.12 | 106.84 |
| 7 | 161.29 | 151.36 | 96.221 | 145.51 | 147.62 | 110.46 |
| 8 | 156.65 | 154.2 | 112.01 | 150.42 | 154.83 | 140.15 |
| 9 | 155.47 | 152.9 | 123.85 | 155.15 | 158.92 | 149.37 |
| 10 | 149.18 | 155.16 | 140.79 | 144.96 | 156.84 | 155.31 |
| 11 | 150.2 | 161.4 | 144.81 | 147.68 | 146.74 | 154.85 |
| 12 | 147.27 | 160.23 | 139.4 | 139.74 | 134.43 | 136.93 |
| 13 | 136.73 | 154.01 | 124.26 | 131.82 | 117.86 | 122.37 |
| 14 | 104.59 | 132.58 | 94.677 | 115.74 | 100.36 | 117.07 |
| 15 | 59.13 | 75.826 | 71.724 | 92.459 | 72.583 | 101.4 |
| 16 | 23.055 | 32.629 | 38.868 | 59.43 | 42.602 | 65.554 |
| 17 | 9.2203 | 16.098 | 20.975 | 57.519 | 23.298 | 52.108 |
| 18 | 11.608 | 6.6926 | 5.3617 | 44.342 | 12.52 | 31.178 |
| 19 | 11.389 | 22.09 | 6.7552 | 68.326 | 20.512 | 34.834 |
| 20 | 19.344 | 7.7529 | 5.8236 | 74.911 | 25.246 | 33.576 |
| 21 | 38.189 | 36.299 | 5.9281 | 55.177 | 24.064 | 26.883 |
| 22 | 47.956 | 36.151 | 35.641 | 40.953 | 44.531 | 45.461 |
| 23 | 46.935 | 28.948 | 14.972 | 39.654 | 26.077 | 43.484 |
| 24 | 46.429 | 31.755 | 20.77 | 34.747 | 24.212 | 35.305 |

# Appendix B

# Market clearing results

This chapter describes the results of the market clearing dispatch. Concerning the day-ahead unit commitment, Table B.1 represents the status of the generating units at the beginning of the day. They were chosen arbitrarily. It is considered, for the sake of simplicity, that the start-up and shut-down times are null.

Tables B.2 and B.3 summarises the results of the market clearing step and indicates for each unit the re-dispatching cost in $/MWh.

Table B.1: Initial status of each generating unit at the beginning of the $50^{th}$ Tuesday of the year.

| Generating unit | Maximum capacity | Unit type | Initial status |
|---|---|---|---|
| 1 | U20 | Oil/CT | down |
| 2 | U20 | Oil/CT | down |
| 3 | U76 | Coal/Steam | down |
| 4 | U76 | Coal/Steam | down |
| 5 | U20 | Oil/CT | down |
| 6 | U20 | Oil/CT | down |
| 7 | U76 | Coal/Steam | down |
| 8 | U76 | Coal/Steam | down |
| 9 | U100 | Oil/Steam | down |
| 10 | U100 | Oil/Steam | down |
| 11 | U100 | Oil/Steam | down |
| 12 | U197 | Oil/Steam | down |
| 13 | U197 | Oil/Steam | down |
| 14 | U197 | Oil/Steam | down |
| 15 | 0 | SynCond | up |
| 16 | U12 | Oil/Steam | down |
| 17 | U12 | Oil/Steam | down |
| 18 | U12 | Oil/Steam | down |
| 19 | U12 | Oil/Steam | down |
| 20 | U12 | Oil/Steam | down |
| 21 | U155 | Coal/Steam | down |
| 22 | U155 | Coal/Steam | down |
| 23 | U400 | Nuclear | up |
| 24 | U400 | Nuclear | up |
| 25 | U50 | Hydro | up |
| 26 | U50 | Hydro | up |
| 27 | U50 | Hydro | up |
| 28 | U50 | Hydro | up |
| 29 | U50 | Hydro | up |
| 30 | U50 | Hydro | up |
| 31 | U155 | Coal/Steam | down |
| 32 | U155 | Coal/Steam | down |
| 33 | U350 | Coal/Steam | up |

Table B.2: Market clearing output per hour and re-dispatching price for generating units 1 to 16. The re-dispatching is in $/MWh. The results are rounded to the second decimal for readability.

| Hour \ Gen. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 38 | 41.43 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 15.2 | 15.2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 15.2 | 15.2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 15.2 | 15.2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 15.2 | 15.2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 15.2 | 15.2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 0 | 0 | 66.10 | 60.8 | 0 | 0 | 55.2 | 55.2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | 0 | 0 | 76 | 76 | 0 | 0 | 76 | 76 | 0 | 25 | 37.45 | 69 | 0 | 0 | 0 | 0 |
| 9 | 0 | 0 | 76 | 76 | 0 | 0 | 76 | 76 | 0 | 25 | 39.79 | 69 | 69 | 69 | 0 | 0 |
| 10 | 0 | 0 | 76 | 76 | 0 | 0 | 76 | 76 | 0 | 37.13 | 25 | 69 | 69 | 69 | 0 | 0 |
| 11 | 0 | 0 | 76 | 76 | 0 | 0 | 76 | 76 | 0 | 25 | 26.70 | 69 | 69 | 69 | 0 | 0 |
| 12 | 0 | 0 | 76 | 64.6 | 0 | 0 | 76 | 76 | 0 | 25 | 25 | 69 | 69 | 69 | 0 | 0 |
| 13 | 0 | 0 | 76 | 76 | 0 | 0 | 76 | 76 | 0 | 25 | 36.82 | 69 | 69 | 69 | 0 | 0 |
| 14 | 0 | 0 | 76 | 76 | 0 | 0 | 76 | 76 | 0 | 50 | 50 | 69 | 87.94 | 69 | 0 | 0 |
| 15 | 0 | 0 | 76 | 76 | 0 | 0 | 76 | 76 | 0 | 50 | 50 | 118.2 | 71.96 | 84.40 | 0 | 0 |
| 16 | 0 | 0 | 76 | 76 | 0 | 0 | 76 | 76 | 0 | 50 | 50 | 118.2 | 131.97 | 118.2 | 0 | 0 |
| 17 | 0 | 0 | 76 | 76 | 0 | 0 | 76 | 76 | 0 | 67.9915 | 50 | 157.6 | 157.6 | 157.6 | 0 | 0 |
| 18 | 16 | 0 | 76 | 76 | 16 | 16 | 76 | 76 | 0 | 50 | 67.13 | 157.6 | 157.6 | 157.6 | 0 | 0 |
| 19 | 0 | 0 | 76 | 76 | 0 | 0 | 76 | 76 | 0 | 80 | 63.42 | 157.6 | 157.6 | 157.6 | 0 | 0 |
| 20 | 0 | 0 | 76 | 76 | 0 | 0 | 76 | 76 | 0 | 50 | 50 | 149.76 | 157.6 | 157.6 | 0 | 0 |
| 21 | 0 | 0 | 76 | 76 | 0 | 0 | 76 | 76 | 0 | 50 | 50 | 118.2 | 113.00 | 118.2 | 0 | 0 |
| 22 | 0 | 0 | 76 | 76 | 0 | 0 | 76 | 76 | 0 | 50 | 28.76 | 69 | 69 | 69 | 0 | 0 |
| 23 | 0 | 0 | 38 | 38 | 0 | 0 | 59.21 | 38 | 0 | 25 | 25 | 69 | 69 | 69 | 0 | 0 |
| 24 | 0 | 0 | 15.2 | 15.2 | 0 | 0 | 15.2 | 15.2 | 0 | 25 | 25 | 69 | 69 | 69 | 0 | 0 |
| Re-dispatching price | 58.91 | 58.91 | 21.48 | 21.48 | 58.91 | 58.91 | 21.48 | 21.48 | 33.47 | 33.47 | 33.47 | 33.41 | 33.41 | 33.41 | 0 | 42.10 |

Table B.3: Market clearing output per hour and re-dispatching price for generating units 17 to 33. The re-dispatching price is in $/MWh. The results are rounded to the second decimal for readability.

| Hour \ Gen. | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 320 | 322.61 | 50 | 50 | 50 | 50 | 50 | 50 | 0 | 0 | 140 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 296.17 | 320 | 50 | 50 | 50 | 50 | 50 | 50 | 0 | 0 | 140 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 200 | 291.09 | 50 | 50 | 50 | 50 | 50 | 50 | 0 | 0 | 140 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 189.74 | 200 | 50 | 50 | 50 | 50 | 50 | 50 | 0 | 0 | 140 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 200 | 187.48 | 50 | 50 | 50 | 50 | 50 | 50 | 0 | 0 | 140 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 244.38 | 200 | 50 | 50 | 50 | 50 | 50 | 50 | 0 | 0 | 140 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 320 | 343.64 | 50 | 50 | 50 | 50 | 50 | 50 | 0 | 0 | 140 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 358.88 | 400 | 50 | 50 | 50 | 50 | 50 | 50 | 0 | 0 | 140 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 398.75 | 400 | 50 | 50 | 50 | 50 | 50 | 50 | 54.3 | 0 | 140 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 400 | 400 | 50 | 50 | 50 | 50 | 50 | 50 | 78.23 | 0 | 140 |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 400 | 400 | 50 | 50 | 50 | 50 | 50 | 50 | 85.22 | 0 | 140 |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 400 | 400 | 50 | 50 | 50 | 50 | 50 | 50 | 54.3 | 62.60 | 140 |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 400 | 400 | 50 | 50 | 50 | 50 | 50 | 50 | 71.63 | 93 | 140 |
| 14 | 0 | 0 | 0 | 0 | 54.3 | 0 | 400 | 400 | 50 | 50 | 50 | 50 | 50 | 50 | 93 | 82.24 | 140 |
| 15 | 0 | 0 | 0 | 0 | 54.3 | 0 | 400 | 400 | 50 | 50 | 50 | 50 | 50 | 50 | 93 | 93 | 214.31 |
| 16 | 0 | 0 | 0 | 0 | 93 | 0 | 400 | 400 | 50 | 50 | 50 | 50 | 50 | 50 | 124 | 124 | 259.90 |
| 17 | 0 | 0 | 0 | 0 | 93 | 54.3 | 400 | 400 | 50 | 50 | 50 | 50 | 50 | 50 | 124.59 | 155 | 280 |
| 18 | 0 | 0 | 0 | 0 | 112.07 | 54.3 | 400 | 400 | 50 | 50 | 50 | 50 | 50 | 50 | 155 | 155 | 280 |
| 19 | 0 | 0 | 0 | 0 | 93 | 54.3 | 400 | 400 | 50 | 50 | 50 | 50 | 50 | 50 | 143.58 | 155 | 280 |
| 20 | 0 | 0 | 0 | 0 | 93 | 54.3 | 400 | 400 | 50 | 50 | 50 | 50 | 50 | 50 | 124 | 124 | 262.69 |
| 21 | 0 | 0 | 0 | 0 | 78.36 | 54.3 | 400 | 400 | 50 | 50 | 50 | 50 | 50 | 50 | 124 | 124 | 227.5 |
| 22 | 0 | 0 | 0 | 0 | 54.3 | 54.3 | 400 | 400 | 50 | 50 | 50 | 50 | 50 | 50 | 93 | 93 | 180.25 |
| 23 | 0 | 0 | 0 | 0 | 54.3 | 54.3 | 400 | 400 | 50 | 50 | 50 | 50 | 50 | 50 | 54.3 | 91.19 | 140 |
| 24 | 0 | 0 | 0 | 0 | 54.3 | 54.3 | 345.08 | 320 | 50 | 50 | 50 | 50 | 50 | 50 | 54.3 | 54.3 | 140 |
| Re-dispatching price | 42.10 | 42.10 | 42.10 | 42.10 | 16.98 | 16.98 | 8.82 | 8.82 | 0 | 0 | 0 | 0 | 0 | 0 | 16.98 | 16.98 | 17.57 |

# Appendix C

# Monte-Carlo simulations

This chapter contains some information about the Monte-Carlo simulations. In particular, the first section presents the probability of unavailability of transmission lines for a normal and an adverse normal and section 2 includes a pseudo-code describing the input data generation.

## C.1 Probability of unavailability of transmission lines

The transmission lines outage probabilities for different values of $t$ can be observed in Table C.1 for a normal weather and in Table C.2 for an adverse weather.

## C.2 Pseudo-code for the Monte-Carlo simulations

The pseudo-code used to generate the different states of the system can be found in Algorithms 1 and 2. Note that all the samples are independent from each others. There are the same number of samples with a normal weather as with an adverse weather and the 24 hours are uniformly distributed in the dataset.

Table C.1: Number of outages per hour and probability of unavailability for $t = 12$, 24 and 36h per transmission line for a normal weather.

| Line | Outages per hour | $P_{outage}(t = 12h)$ | $P_{outage}(t = 24h)$ | $P_{outage}(t = 36h)$ |
|---|---|---|---|---|
| 1 | $2.74\ 10^{-5}$ | $3.29\ 10^{-4}$ | $6.57\ 10^{-4}$ | $9.86\ 10^{-4}$ |
| 2 | $5.82\ 10^{-5}$ | $6.98\ 10^{-4}$ | $1.40\ 10^{-4}$ | $2.09\ 10^{-3}$ |
| 3 | $3.77\ 10^{-5}$ | $4.52\ 10^{-4}$ | $9.04\ 10^{-4}$ | $1.36\ 10^{-3}$ |
| 4 | $4.45\ 10^{-5}$ | $5.34\ 10^{-4}$ | $1.07\ 10^{-3}$ | $1.60\ 10^{-3}$ |
| 5 | $5.48\ 10^{-5}$ | $6.57\ 10^{-4}$ | $1.31\ 10^{-3}$ | $1.97\ 10^{-3}$ |
| 6 | $4.34\ 10^{-5}$ | $5.20\ 10^{-4}$ | $1.04\ 10^{-3}$ | $1.56\ 10^{-3}$ |
| 7 | $4.11\ 10^{-5}$ | $4.93\ 10^{-4}$ | $9.86\ 10^{-4}$ | $1.48\ 10^{-3}$ |
| 8 | $3.88\ 10^{-5}$ | $4.66\ 10^{-4}$ | $9.31\ 10^{-4}$ | $1.40\ 10^{-3}$ |
| 9 | $3.77\ 10^{-5}$ | $4.52\ 10^{-4}$ | $9.04\ 10^{-4}$ | $1.36\ 10^{-3}$ |
| 10 | $3.43\ 10^{-5}$ | $4.11\ 10^{-4}$ | $8.22\ 10^{-4}$ | $1.23\ 10^{-3}$ |
| 11 | $5.02\ 10^{-5}$ | $6.03\ 10^{-4}$ | $1.21\ 10^{-3}$ | $1.81\ 10^{-3}$ |
| 12 | $5.02\ 10^{-5}$ | $6.03\ 10^{-4}$ | $1.21\ 10^{-3}$ | $1.81\ 10^{-3}$ |
| 13 | $4.57\ 10^{-5}$ | $5.48\ 10^{-4}$ | $1.10\ 10^{-3}$ | $1.64\ 10^{-3}$ |
| 14 | $4.45\ 10^{-5}$ | $5.34\ 10^{-4}$ | $1.07\ 10^{-3}$ | $1.60\ 10^{-3}$ |
| 15 | $4.57\ 10^{-5}$ | $5.48\ 10^{-4}$ | $1.20\ 10^{-3}$ | $1.64\ 10^{-3}$ |
| 16 | $5.94\ 10^{-5}$ | $7.12\ 10^{-4}$ | $1.42\ 10^{-3}$ | $2.14\ 10^{-3}$ |
| 17 | $5.60\ 10^{-5}$ | $6.71\ 10^{-4}$ | $1.34\ 10^{-3}$ | $2.01\ 10^{-3}$ |
| 18 | $4.34\ 10^{-5}$ | $5.20\ 10^{-4}$ | $1.04\ 10^{-3}$ | $1.56\ 10^{-3}$ |
| 19 | $3.77\ 10^{-5}$ | $4.52\ 10^{-4}$ | $9.04\ 10^{-4}$ | $1.36\ 10^{-3}$ |
| 20 | $4.68\ 10^{-5}$ | $5.62\ 10^{-4}$ | $1.12\ 10^{-3}$ | $1.68\ 10^{-3}$ |
| 21 | $4.68\ 10^{-5}$ | $5.62\ 10^{-4}$ | $1.12\ 10^{-3}$ | $1.68\ 10^{-3}$ |
| 22 | $4.68\ 10^{-5}$ | $5.62\ 10^{-4}$ | $1.12\ 10^{-3}$ | $1.68\ 10^{-3}$ |
| 23 | $3.995\ 10^{-5}$ | $4.79\ 10^{-4}$ | $9.58\ 10^{-4}$ | $1.44\ 10^{-3}$ |
| 24 | $3.88\ 10^{-5}$ | $4.66\ 10^{-4}$ | $9.31\ 10^{-4}$ | $1.40\ 10^{-3}$ |
| 25 | $3.65\ 10^{-5}$ | $4.38\ 10^{-4}$ | $8.76\ 10^{-4}$ | $1.31\ 10^{-3}$ |
| 26 | $6.16\ 10^{-5}$ | $7.40\ 10^{-4}$ | $1.48\ 10^{-3}$ | $2.22\ 10^{-3}$ |
| 27 | $3.995\ 10^{-5}$ | $4.79\ 10^{-4}$ | $9.58\ 10^{-4}$ | $1.44\ 10^{-3}$ |
| 28 | $3.995\ 10^{-5}$ | $4.79\ 10^{-4}$ | $9.58\ 10^{-4}$ | $1.44\ 10^{-3}$ |
| 29 | $4.34\ 10^{-5}$ | $5.20\ 10^{-4}$ | $1.04\ 10^{-3}$ | $1.56\ 10^{-3}$ |
| 30 | $4.338\ 10^{-5}$ | $5.20\ 10^{-4}$ | $1.04\ 10^{-3}$ | $1.56\ 10^{-3}$ |
| 31 | $3.88\ 10^{-5}$ | $4.66\ 10^{-4}$ | $9.31\ 10^{-4}$ | $1.40\ 10^{-3}$ |
| 32 | $3.88\ 10^{-5}$ | $4.66\ 10^{-4}$ | $9.31\ 10^{-4}$ | $1.40\ 10^{-3}$ |
| 33 | $5.14\ 10^{-5}$ | $6.16\ 10^{-4}$ | $1.23\ 10^{-3}$ | $1.85\ 10^{-3}$ |

Table C.2: Number of outages per hour and probability of unavailability for $t = 12$, 24 and 36h per transmission line for an adverse weather.

| Line | Outages per hour | $P_{outage}(t = 12h)$ | $P_{outage}(t = 24h)$ | $P_{outage}(t = 36h)$ |
|---|---|---|---|---|
| 1 | $2.74\ 10^{-5}$ | $3.29\ 10^{-4}$ | $6.57\ 10^{-4}$ | $9.86\ 10^{-4}$ |
| 2 | 0.00175 | 0.0207 | 0.0411 | 0.0609 |
| 3 | 0.00113 | 0.0135 | 0.0268 | 0.0399 |
| 4 | 0.00134 | 0.0159 | 0.0316 | 0.0469 |
| 5 | 0.00164 | 0.0195 | 0.0387 | 0.0575 |
| 6 | 0.00130 | 0.0155 | 0.0308 | 0.0458 |
| 7 | 0.00123 | 0.0147 | 0.0292 | 0.0434 |
| 8 | 0.00116 | 0.0139 | 0.0276 | 0.0411 |
| 9 | $3.77\ 10^{-5}$ | $4.52\ 10^{-4}$ | $9.04\ 10^{-4}$ | $1.36\ 10^{-3}$ |
| 10 | 0.00103 | 0.0123 | 0.0244 | 0.0363 |
| 11 | 0.00151 | 0.0179 | 0.0355 | 0.0528 |
| 12 | 0.00151 | 0.0179 | 0.0355 | 0.0528 |
| 13 | 0.00137 | 0.0163 | 0.0323 | 0.0481 |
| 14 | 0.00134 | 0.0159 | 0.0316 | 0.0469 |
| 15 | 0.00137 | 0.0163 | 0.0323 | 0.0481 |
| 16 | 0.00178 | 0.0211 | 0.0418 | 0.0621 |
| 17 | 0.00168 | 0.0199 | 0.0395 | 0.0586 |
| 18 | 0.00130 | 0.0155 | 0.0308 | 0.0458 |
| 19 | 0.00113 | 0.0135 | 0.0268 | 0.0399 |
| 20 | 0.00140 | 0.0167 | 0.0331 | 0.0493 |
| 21 | 0.00140 | 0.0167 | 0.0331 | 0.0493 |
| 22 | 0.00140 | 0.0167 | 0.0331 | 0.0493 |
| 23 | 0.00120 | 0.0143 | 0.0284 | 0.0422 |
| 24 | 0.00116 | 0.0139 | 0.0276 | 0.0411 |
| 25 | 0.00110 | 0.0131 | 0.0260 | 0.0387 |
| 26 | 0.00185 | 0.0220 | 0.0434 | 0.0644 |
| 27 | 0.00120 | 0.0143 | 0.0284 | 0.0422 |
| 28 | 0.00120 | 0.0143 | 0.0284 | 0.0422 |
| 29 | 0.00130 | 0.0155 | 0.0308 | 0.0458 |
| 30 | 0.00130 | 0.0155 | 0.0308 | 0.0458 |
| 31 | 0.00116 | 0.0139 | 0.0276 | 0.0411 |
| 32 | 0.00116 | 0.0139 | 0.0276 | 0.0411 |
| 33 | 0.00154 | 0.0183 | 0.0363 | 0.0540 |

```
state_realisation
Input: wind forecast(hour), load forecast(hour), weather, hour
Output: wind realisation, load realisation, generators outages, lines outages,
        transformers outages, probabilities of contingencies, epsilon
if weather = adverse then
    for line in list of lines do
        if line is not a cable then
            outage rate per hour is multiplied by 30;
        end
    end
end
t=hour+12;
// day-ahead predictions are done at noon
for generator in list of generators do
    failureProbability= fail_prob(generator, t);
    randomly generate a number randomNb uniformly distributed in [0, 1];
    if randomNb<failureProbability then
        generator is out-of-service;
    end
end
for line in list of lines do
    failureProbability= fail_prob(line, t);
    randomly generate a number randomNb uniformly distributed in [0, 1];
    if randomNb<failureProbability then
        line is out-of-service;
    end
end
for pst in list of transformers do
    failureProbability= fail_prob(pst, t);
    randomly generate a number randomNb uniformly distributed in [0, 1];
    if randomNb<failureProbability then
        transformer is out-of-service;
    end
end
common_error_factor_d is a random number generated with law N(0, α_d);
for busLoad in busLoad list do
    ind_error_factor_d is a random number generated with law N(0, β_d);
    loadRealisation=load prediction*(1+common_error_factor_d +
    ind_error_factor_d);
end
common_error_factor_w is a random number generated with law N(0, α_w);
for windFarm in windFarm list do
    ind_error_factor_w is a random number generated with law N(0, β_w);
    windRealisation=wind prediction*(1+common_error_factor_w +
    ind_error_factor_w);
end
```

**Algorithm 1:** Pseudo-code for Monte-Carlo simulations of system states (first part).

```
// Compute probabilities of contingency
cont_prob(updated topology, updated lambda);


cont_prob
Input: updated topology, lambda
Output: probabilities of contingencies
find the set C_available of contingencies where the line (or the 2 lines) in outage
is (are) available;
// Compute the no outage probability and the probability of
   failure of each line and double outage
initialise prob_no_failure = 1 ;
for c ∈ C_available do
  | failure_prob(c)=fail_prob(c, t=1h);
  | prob_no_failure = prob_no_failure * (1-failure_prob(c)) ;
end
// Compute the probability of contingency for each available
   contingencies
for c ∈ C_available do
  | contingency_prob(c)=prob_no_failure * failure_prob(c) /
  | (1-failure_prob(c));
end
// Compute the probability of contingency for each unavailable
   contingency and the no outage case
for c ∉ C_available do
  | contingency_prob(c)=0 ;
end
contingency_prob(1) = 1-sum(contingency_prob(c)) for all c;

fail_prob
Input: element, t
Output: probability of failure
lambda=outage_rate(element);
prob=1-exp(-lambda*t);
```

**Algorithm 2:** Pseudo-code for Monte-Carlo simulations of system states (second part).

# Bibliography


[1] DMSP satellite images representing the blackout that occurred on 2003-08-14, in the United States and Canada. http://earthobservatory.nasa.gov/IOTD/view.php?id=3719, accessed on 2016-06-07.

[2] GARPUR: Generally Accepted Reliability Principle with Uncertainty modelling and through probabilistic Risk assessment. Collaborative R&D project co-funded by the European Commission(7th Framework Programme). http://www.garpur-project.eu/, accessed on 2016-05-31.

[3] E. Karangelos and L. Wehenkel, "Probabilistic reliability management approach and criteria for power system real-time operation," in *Submitted for presentation at Power Systems Computation Conference (PSCC), Genoa, Italy*, June 2016.

[4] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[5] Elia, "Legal framework." http://www.elia.be/en/about-elia/legal-framework, accessed on 2016-05-20.

[6] GARPUR, "D1.1 State of the art on reliability assessment in power systems," April 2014. Available at http://www.garpur-project.eu/deliverables.

[7] L. H. Fink and K. Carlsen, "Operating under stress and strain," *IEEE Spectrum;(United States)*, vol. 15, no. 3, 1978.

[8] C. Hamon, *Probabilistic security management for power system operations with large amounts of wind power*. PhD thesis, KTH, Electric Power Systems, 2015. QC 20150508.

[9] A. Wood and B. Wollenberg, *Power Generation, Operation and Control*. John Wiley & Sons, 2nd ed., 1996.

[10] A. Von Meier, *Electric power systems: a conceptual introduction*. John Wiley & Sons, 2006.

[11] ENTSO-E, "Network code on operational security," September 2013. Available at /http://networkcodes.entsoe.eu/operational-codes/operational-security/.

[12] D. Kirschen and D. Jayaweera, "Comparison of risk-based and deterministic security assessments," *Generation, Transmission & Distribution, IET*, vol. 1, no. 4, pp. 527–533, 2007.

[13] E. Karangelos, P. Panciatici, and L. Wehenkel, "Whither probabilistic security management for real-time operation of power systems?," in *Bulk Power System Dynamics and Control-IX Optimization, Security and Control of the Emerging Power Grid (IREP), 2013 IREP Symposium*, pp. 1–17, IEEE, 2013.

[14] F. Capitanescu, S. Fliscounakis, P. Panciatici, and L. Wehenkel, "Day-ahead security assessment under uncertainty relying on the combination of preventive and corrective controls to face worst-case scenarios," *PSCC proceedings Stockholm (Sweden) 2011*, 2011.

[15] R. Marceau, J. Endrenyi, R. Allan, F. Alvarado, G. Bloemhof, T. Carlsen, G. Couto, E. Dialynas, N. Hatziargyriou, D. Holmberg, *et al.*, "Power system security assessment: A position paper," *Electra*, vol. 175, December 1997.

[16] GARPUR, "D1.2 Current practices, drivers and barriers for new reliability standards," September 2014. Available at http://www.garpur-project.eu/deliverables.

[17] iTesla: Innovative Tools for Electrical System Security within Large Areas.Collaborative R&D project co-funded by the European Commission (7th Framework Programme). http://www.itesla-project.eu/, accessed on 2016-06-06.

[18] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics, 2nd ed., 2009.

[19] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[20] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[21] T. G. Dietterich, "An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization," *Machine Learning*, vol. 40, no. 2, pp. 139–157, 2000.

[22] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006.

[23] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.

[24] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[25] E. Karangelos and L. Wehenkel, "Probabilistic reliability management approach and criteria for power system real-time operation - supplementary material," Oct. 2015. [Online], Available at http://www.montefiore.ulg.ac.be/~karangelos/.

[26] P. Wong, P. Albrecht, R. Allan, R. Billinton, Q. Chen, C. Fong, S. Haddad, W. Li, R. Mukerji, D. Patton, *et al.*, "The ieee reliability test system-1996. a report prepared by the reliability test system task force of the application of probability methods subcommittee," *Power Systems, IEEE Transactions on*, vol. 14, no. 3, pp. 1010–1020, 1999.

[27] C. Ordoudis, P. Pinson, J. Morales González, and M. Zugno, *An Updated Version of the IEEE RTS 24-Bus System for Electricity Market and Power System Operation Studies.* Technical University of Denmark, 2016.

[28] W. Bukhhsh, "Data for stochastic multiperiod optimal power flow problem," March 2015. [Online], Available at https://sites.google.com/site/datasmopf/home.

[29] R. Billinton and R. N. Allan, *Reliability evaluation of power systems.* Springer, 2nd ed., 1996.

[30] R. Billinton, *Power System Reliability Evaluation.* Gordon and Breach, 1970.

[31] R. Billinton and L. Wenyuan, "A novel method for incorporating weather effects in composite system adequacy evaluation," *Power Systems, IEEE Transactions on*, vol. 6, no. 3, pp. 1154–1160, 1991.

[32] B.-M. Hodge, A. Florita, K. Orwig, D. Lew, and M. Milligan, "A comparison of wind power and load forecasting error distributions," in *Proc. 2012 World Renewable Energy Forum*, 2012.

[33] B.-M. Hodge, D. Lew, M. Milligan, H. Holttinen, S. Sillanpää, E. Gómez-Lázaro, R. Scharff, L. Söder, X. G. Larsén, G. Giebel, *et al.*, "Wind power forecasting error distributions: An international comparison," in *11th Annual International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants Conference*, 2012.

[34] H. Bludszuweit, J. A. Domínguez-Navarro, and A. Llombart, "Statistical analysis of wind power forecast error," *Power Systems, IEEE Transactions on*, vol. 23, no. 3, pp. 983–991, 2008.

[35] A. Florita, B.-M. Hodge, and M. Milligan, "Wind power forecasting error frequency analyses for operational power system studies," in *11th International Workshop on Large-Scale Integration of Wind Power in Power Systems Proceedings, Lisbon, Portugal*, 2012.

[36] L. Breiman, "Manual on setting up, using, and understanding random forests v3.1," *Statistics Department University of California Berkeley, CA, USA*, 2002.

[37] G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts, "Understanding variable importances in forests of randomized trees," in *Advances in Neural Information Processing Systems*, pp. 431–439, 2013.

[38] Scikit-learn developers, "Model evaluation: quantifying the quality of predictions." [Online], Available at http://scikit-learn.org/stable/modules/model_evaluation.html, accessed on 2016-05-11.

[39] P. Pinson *et al.*, "Wind energy: Forecasting challenges for its operational management," *Statistical Science*, vol. 28, no. 4, pp. 564–585, 2013.

[40] W. A. Bukhsh, C. Zhang, and P. Pinson, "A multiperiod opf model under renewable generation uncertainty and demand side flexibility," July 2014. [Online], Available at http://arxiv.org/abs/1407.2744.