Industrial Electrical Engineering and Automation

CODEN:LTH/(IEA-5462)/1-55/(2021)

Monte Carlo Simulations in Load Flow Calculations

-An Application on a Swedish 50 kV Network

Alice Jansson
Emma Åkerman

Division of Industrial Electrical Engineering and Automation Faculty of Engineering, Lund University

Abstract

The Swedish power system is changing due to the growing installations of non-dispatchable generation as well as the expected growth of variation of consumption levels. Mathematically, the power flows in the grid can to a larger extent be described by stochastic measures rather than deterministic. This suggests a need to introduce probabilistic methods for making load flow calculations as opposed to the deterministic methods used today. The aim of this thesis is to examine and illustrate the use of a probabilistic method for making load flow calculations. The chosen method is a Monte Carlo simulation multiple-scenario method. The method is first set up for a reference case and validated by comparison to real-life measurements. It is then applied on three cases which aim to illustrate the possible industrial applications of the method. The Monte Carlo simulation has generated accurate results in the reference case. The application of the method shows that it could be useful for generating more information on how an installation may affect the power flows in a selected study area. However, it's concluded that the method would need to be improved in order to apply it as a risk-based method.

Keywords - Probabilistic load flow, Power grid, Monte Carlo, Power line capacity

Preface

This thesis has been carried out at the division of Industrial Electrical Engineering and Automation (IEA) at the faculty of Engineering at Lund University in the spring semester of 2021. The thesis has been conducted in collaboration with E.ON Energy Distribution.

We would like to thank our supervisor Olof Samuelsson at IEA for good supervision and valuable comments throughout the work of this thesis. His expertise in the field has been both inspiring and helpful. We would also like to thank our supervisor Hampus Möller at E.ON Energy Distribution who has shown great support and enthusiasm for the work and always been available to answer all our questions. Hampus has also helped us with the network model and the simulation tools used in the thesis. Finally, we want to thank E.ON Energy Distribution for providing the topic of the thesis and the data for the study. The team at E.ON have shown great interest in our work and have been very supportive.

Due to the Covid-19 pandemic, the work of this thesis has been carried out in our homes. We would therefore like to thank Alice's flatmates (and her cat) for letting us occupy their kitchen for a whole semester. Special thanks also to Emma's partner Gustav for showing patience and for sometimes moving away from his workstation into the bedroom in order to give us space in the apartment. In spite of the difficult circumstances, we have enjoyed the work with this thesis and are thankful that it has worked out so well in every aspect.

List of Abbreviations

CHP Combined Heat and Power

DAE Differential Algebraic Equation

DC Data Center

MC Monte Carlo

PCC Pearson's Correlation Coefficient

PLF Probabilistic Load Flow

SLD Single Line Diagram

Contents

1 Introduction

This section will give a background to the subject of the thesis and provide a foundation for the work. The aim and research questions will be presented together with the limitations of the study. Furthermore, the previous work on which this thesis is built will be presented.

1.1 Background

The Swedish electricity system is undergoing major changes. The share of hydro- and nuclear power in the Swedish electricity mix has gone from constituting 96 % of all power generation in 1990 to around 80 % in 2018 [1]. The penetration of renewable and non-dispatchable electricity generation has been increasing each year in the 2010s, especially from the growing amount of installed wind power. Furthermore, Sweden has a goal to reach 100 % renewable electricity generation in 2040. The future of the Swedish electricity system is predicted to hold an increasing share of non-dispatchable electricity generation as well as a growing electricity consumption following the electrification of other energy sectors. The long term scenarios from the Swedish Energy Agency suggests that the variations in the total consumption of electricity is expected to grow, both in terms of energy and peak power [2].

The expected developments will change the basic characteristics of the electricity system. Grid capacity shortages have occurred in 2019, in the areas of Skåne and Stockholm. Forecasts predict an increasing risk of capacity shortages in the Swedish power grid in the future and the Swedish government has adopted a plan to analyze and prepare for these future scenarios [3]. High levels of new wind power installations and a fast growth of electricity consumption are expected to further add to the possible capacity shortages. One of the largest challenges is the time consuming process of reinforcements and construction of new transmission lines which take up to 10-15 years. This suggests that alternative solutions for avoiding overloading of the existing power lines are needed until the grid can be reinforced. One possible solution to this is working with flexibility in the power grid.

When investigating the need of flexibility, the loading of the power lines needs to be known. Thus load flow calculations are performed. The standard methods for calculating load flows in industry are deterministic and often based on a few worst case scenarios. This has worked well in the past when the variations in consumption levels have been slow and the largest part of electricity generation has been dispatchable. However, the introduction of more non-dispatchable and variable distributed power generation into the power grid means that the parameters in the power system are to a larger extent stochastic [4].

Considering the changes in the power system and how these changes are expected to continue, the European Commission recommends the usage of probabilistic methods for load flow calculations in the electricity system. The preferred methods according to this report are Monte Carlo (MC) methods [5]. The method which will be investigated in this report is a MC simulation multiple-scenario method. In short this method can be described as solving the power flow equations for a large number of different "scenarios". Each scenario consists of randomly drawn values for the load and generation parameters in the

study area. The resulting power flows in the branches are collectively treated as probability distributions, describing the probability of the power flows taking on different values. This method could be compared with the methods used today, which typically consider only a few worst-case scenarios for which the power flow equations are solved. The suggested method will be applied on a real-life case in the form of a network model of the Swedish power grid. This has been done with the use of historical measurement data provided by E.ON in combination with a defined MC simulation method for examining branch power flows in the selected study area.

1.2 Aim, Purpose and Research Questions

The aim of this study is to illustrate a probabilistic method for making load flow calculations, in order to approach a future where probabilistic calculations are made in the electrical grid.

The purpose is to:

- Identify a probabilistic method for making load flow calculations to examine the probability of thermal overload.
- Examine and illustrate the value of using probabilistic methods in load flow calculations.

The study is based on the following research questions:

- How can a Monte Carlo simulation method be applied in load flow calculations to study the risk of thermal overload?
- What input data is needed for the Monte Carlo simulation?
- What kind of results can be expected from the Monte Carlo simulation?
- How useful is the method in industrial applications?

1.3 Limitations

This thesis will create a proof of concept of how a probabilistic method could be used when making load flow calculations. A mathematical approach has been taken to a highly technical issue, which indicates some limitations. The use of deterministic methods in the modelling of the power grid is still standard in industry. Switching to probabilistic way of thinking will require some change of mindset as to how to handle risk levels and flexibility in the grid. This type of questions are out of scope for this thesis, which only aims at illustrating a possible tool for a future where probabilistic load flow calculations are to be used. Additional studies and methods for how to apply a probabilistic method on a real-life grid are required.

The thesis uses a selected study area of 10 nodes, which is a part of E.ON's network model in PSS/E. This enables an approach which is close to industry, but might not be applicable on any grid or sub-system. For more generalized results, a more general test grid could be used. For applications closer to the actual industrial use, larger parts of the network model could be included in the study. One aspect which followed the use of a small part of the network model was the limited computational effort needed. This thesis has not considered

how the computational effort scales with the number of nodes in the selected study area.

1.4 Related Work

The topic of probabilistic load flow (PLF) calculations has been discussed in research globally for several years. However, a deterministic approach is still typically used in industry. This thesis will build upon previous work that is done within the field and highlight the possibilities and limitations of PLF, as well as give examples of industrial applications of the method. The scientific articles covering the subject of different methods for PLF often present and compare both numerical and analytical approaches for PLF calculations. The numerical approach is typically a MC simulation. In [6], [7] and [8] this method is used as benchmark and is compared to the other techniques. This is due to the high accuracy of the MC simulation method.

The international council on large electric systems, CIGRE, published a comprehensive report in November 2020 covering topics such as how power system planning works today and current and future uncertainty factors. This report has been of great importance for laying the ground work of this thesis [4]. The previous mentioned sources that have been important for this thesis, ([6], [7] and [8]) all highlight the necessity of probabilistic methods for load flow calculations, especially in these times when the power system is under major development. The importance of correctly generating the input values and capture correlations between them is also discussed.

2 Theory

This section will present the Swedish power system and different methods for examining the risk of thermal overload. The statistical background of the MC simulation method will be presented. One of the cases that will be investigated in this thesis is the adding of a data center onto one of the stations in the investigated study area. For this purpose some background to data centers will be presented in the end of this section.

2.1 The Swedish Power System

The Swedish electricity grid is divided into three levels; the transmission network (400-200 kV), the regional sub-transmission networks (130-40 kV) and the local distribution networks (40-10 kV and 230/400 V). Figure 1 shows a sketch of the Swedish power grid. The transmission network is owned by Svenska Kraftnät and delivers electricity over long distances. Historically the transmission network has delivered electricity from the large hydro power plants in the north of Sweden to the areas in the south of Sweden with higher power consumption. The sub-transmission networks are mainly owned by the companies E.ON, Vattenfall and Ellevio. These networks deliver the electricity from the transmission network to the distribution networks. Some large industries and power plants are connected directly to the sub-transmission networks. The distribution networks are owned by a range of private companies. In total 160 Swedish companies own electric grids. The distribution network delivers electricity to the end consumers [9]. Historically no power generation is connected to the distribution network, but an increasing amount of local generation is becoming connected to the distribution network. This generation is typically in the form of residential solar or wind power generation.

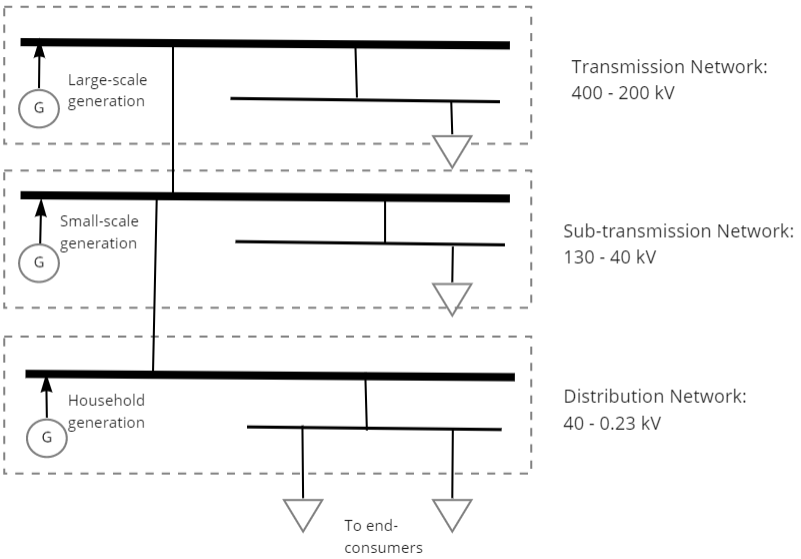


Figure 1: Sketch of the levels in the Swedish power grid.

The state agency Svenska Kraftnät has the over-all balance responsibility in the Swedish power grid. They are also responsible for providing a transmission network which has enough capacity to deliver the required amount of power at all times. For the sub-transmission and distribution networks, the grid owner is responsible for providing a network with sufficient capacity to deliver enough power. The grid owner is also responsible for making new connections to the grid and ensuring operational reliability [10]. The amount of connected wind power to the Swedish power grid is increasing, and connections of wind power parks are typically made directly onto the sub-transmission grid. The responsible party for wind power installations is therefore typically the sub-transmission grid owner. For wind power parks larger than 100 MW, permission can be applied for to connect directly onto the transmission network [11].

The sub-transmission and distribution grids can be divided into two categories; radial and meshed networks. This is illustrated in figure 2. The thick lines in this figure represent the buses/nodes in the network. In radial networks, the electricity has only one way to travel to the consumers. Meshed networks are defined by multiple ways for the electricity to travel to the different points in the grid [12]. Typically the lower-voltage distribution grids are radial networks. The sub-transmission grids are mainly radial but can also be meshed. When analyzing the electrical grid, calculating the power flows in these networks is essential.

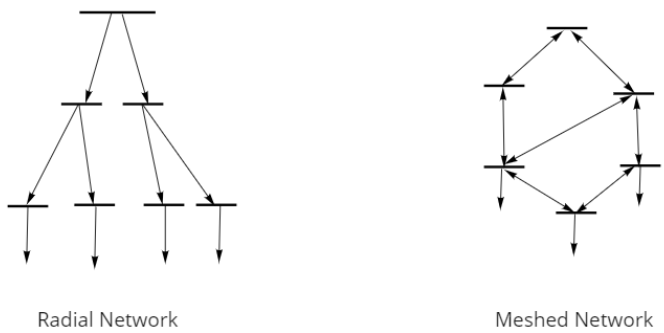


Figure 2: Radial and meshed networks.

2.2 The Power Flow Problem

Complex power can be defined as:

$$\begin{aligned} S = VI^* &= [V\angle\delta][I\angle\beta]^* = VI\angle(\delta-\beta) \\ &= \mathrm{VI}\cos(\delta-\beta) + j\mathrm{VI}\sin(\delta-\beta) \end{aligned} \tag{1}$$

where V$\angle\delta$ is the voltage across a circuit element and I$\angle\beta$ is the current through the element [13]. Alternatively the complex power can be expressed as

$$\mathbf{S} = \mathbf{P} + j\mathbf{Q} \tag{2}$$

where P is the real power and Q is the reactive power. The magnitude of the complex power is called the apparent power. Figure 3 shows the relationship between S, P, Q, V, I, β and δ.

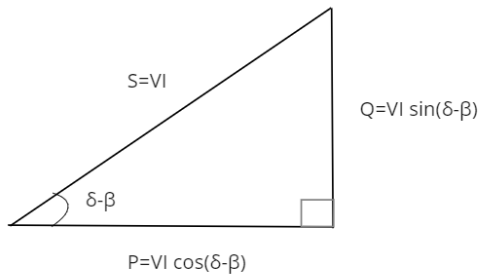


Figure 3: The power triangle.

In a network with N+1 buses, a nodal equation system can be formulated as

$$\begin{bmatrix} Y_{11} & Y_{12} & Y_{13} & \cdots & Y_{1N} \\ Y_{21} & Y_{22} & Y_{23} & \cdots & Y_{2N} \\ Y_{31} & Y_{32} & Y_{33} & \cdots & Y_{3N} \\ \vdots & \vdots & \vdots & & \vdots \\ Y_{N1} & Y_{N2} & Y_{N3} & \cdots & Y_{NN} \end{bmatrix} \begin{bmatrix} V_{10} \\ V_{20} \\ V_{30} \\ \vdots \\ V_{N0} \end{bmatrix} = \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ \vdots \\ I_N \end{bmatrix} \tag{3}$$

or written as

$$\mathbf{YV} = \mathbf{I} \tag{4}$$

where Y is the bus admittance matrix, V the bus voltages and I the bus currents [13]. The diagonal elements Y_{kk} are the sum of admittances at bus k. The off-diagonal elements Y_{kn} are the negative sum of admittances connected between buses k and n.

In a load flow problem, the expressions of complex power in (1) and (2) can be expressed as

$$S_k = \mathrm{P}_k + j\mathrm{Q}_k = V_k I_k^*. \tag{5}$$

Y_{kn} are taken from the admittance matrix (3). The voltage and admittance terms are defined as

$$
\begin{aligned}
V_k &= V_k e^{j\delta_k} \\
Y_{kn} &= Y_{kn} e^{j\theta_{kn}} .
\end{aligned}
\tag{6}
$$

The active power, P_k, and the reactive power, Q_k, can be set up as follows

$$
\begin{aligned}
P_k &= V_k \textstyle\sum_{n=1}^{N} Y_{kn} V_n \cos\left(\delta_k - \delta_n - \theta_{kn}\right) \\
Q_k &= V_k \textstyle\sum^{N} Y_{kn} V_n \sin\left(\delta_k - \delta_n - \theta_{kn}\right) .
\end{aligned}
\tag{7}
$$

Setting up these expressions for every node in the system creates a power flow problem which can be solved numerically using for example the Newton-Raphson method [13]. This is usually done using a commercial software like PSS/E.

2.3 Overhead Line Loading and Capacity

The Swedish power grid and its overhead-lines are regulated according to the managing authority Elsäkerhetsverket. A minimum distance between the high voltage overhead lines and the ground is set based on the area underneath the lines and the type and voltage level of the lines [14]. These regulations hold for the overhead transmission lines at all loads and weather conditions. The conductor sag of the transmission line predominantly depends on the loading of the line and the temperature, due to thermal elongation. Other factors that affect the conductor sag are wind and ice loading [15].

Typically in industry, seasonal overhead line ratings are defined based on temperature, wind, and solar radiation parameters. Each rating gives a set level for line loading which can not be exceeded due to the risk of the power lines hanging too close to the ground. This set level is in line with the regulations set by Elsäkerhetsverket.

2.4 Deterministic Power System Models

The methods of examining the risk of thermal overload in overhead lines today are typically based on deterministic reasoning. For the set of power flow equations presented in (7) that are set up for the system, a number of predefined scenarios are defined and solved for. This type of method is called a single-scenario model. The method is based on a network model, with static parameters at all buses in the system. The only parameters which are varied are some chosen loads and/or generation, often using scaling factors. These are defined as input values in the power flow equations. The system of power flow equations can then be solved for one scenario at the time, in which the input values are defined and sent into the solver [7].

The selected cases used in a deterministic method can be based on the real-life situations when the overall system load has been high or when the line ratings have been close to be exceeded. When new installations are considered, the system is set up in a power flow solver with one of the pre-defined scenarios as base. The new installation is then added to the corresponding bus in the model as an input value, whilst the rest of the model is kept constant. The system is solved for and the resulting branch power flows can be compared to the line

ratings. When this has been done for a number of different scenarios, a decision can be made as to whether the line loadings have enough margin to ensure that thermal overload will not occur. Figure 4 shows a flow chart of what such a deterministic method could look like.

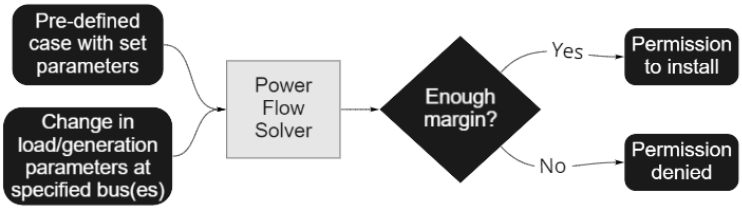


Figure 4: Flow chart of a deterministic power system model.

2.5 Probabilistic Power System Models

In power system planning, many factors are stochastic rather than deterministic. This suggests a need of introducing methods for PLF calculations [4], [8]. For European countries, the recommendation is to use probabilistic models in the future when making adequacy calculations for the power system. This is due to the stochastic behaviour of the uncertainties in the power system. These uncertainties are growing with the increasing installations of non-dispatchable power generation such as wind power. Specifically the use of Monte Carlo simulation methods has been recommended by the European Commission [5]. Similar methods could be used when examining power line capacities.

An alternative to the deterministic single scenario method described above are probabilistic multiple scenario methods. An example of a probabilistic multiple scenario method is the MC simulation. To use this, either historical or simulated data can be used as input values for the loads and generation in the power system. The input values are in the form of distributions, from which variables are picked at random. The system is solved a large number of times, each time for a different set of random variables. The results are then treated collectively as probabilistic distributions describing the probability of possible load flows [4]. For line capacity examinations, it is the resulting distributions which can be analyzed in order to decide if the margins are high enough.

Another way of using probabilistic methods is by applying them as risk-based methods. These accept a certain level of risk as a trade-off between security and economy. A risk-based model has a pre-defined level of accepted probability or risk of the undesired outputs. The model then allows for an optimization of the power system without exceeding the accepted level of risk. Today, scenario-based probabilistic methods are typically not used as risk-based models, but this is seen as a strong alternative in the future [4], [6]. Risk based models will probably be of value in combination with solutions such as flexibility.

2.6 Monte Carlo Simulation

MC simulation is a numerical method based on repetitive random sampling and statistical analysis used for complex and nonlinear models with several random variables. The simulation results in probability distributions of the desired random variables without requiring simplification of the nonlinear equations. The equation of MC estimation is presented in (8) [7].

$$\hat{E}(F) = \frac{1}{n_{\mathrm{S}}} \sum_{i=1}^{n_{\mathrm{S}}} F\left(S_i\right) \tag{8}$$

Here, $\hat{E}(F)$ is the estimated expectation of the desired random value, F is the desired random value and $S = [X_1, X_2, ..., X_n]$ is the system state where X_i is the ith input random variable and n_{S} is the number of iterations.

The MC simulation method is applicable to the nonlinear load flow equations in order to estimate a random variable such as power flow occurring in a line in a network. The use of MC simulations for estimating a probability of power flows consists in executing the deterministic load flow calculations a large number of times. With this technique the deterministic load flow equations can be used without simplification. However, due to the nonlinear nature of these equations and the iterative approach of the technique, MC sampling require in general larger computational efforts than deterministic methods [16].

2.6.1 Number of Iterations in a MC Simulation

The number of iterations needed in a MC simulation is independent of the system size. The stopping criteria for the MC simulation might be a fixed number of iteration or a set limit for the variance. It is challenging yet critical to find an n_{S} that is sufficient for the accuracy of the results yet not too high in order to save computational time. The main idea when deciding the number of iterations is to look for convergence of the MC simulations [7]. Here, convergence of the MC simulation means that a point is reached where more iterations does not change the result. That is, if you compare the outputs of the MC simulations and observe that more iterations makes no difference to the outcome, convergence is reached. This is what convergence means throughout this report. To do this one can either investigate when the variance of the solutions reach a desired limit, or one can visually compare the results from the MC sampling and observe when convergence of the results occur.

One metric to investigate similarity of two multidimensional vectors is to measure the distance between them in a multidimensional space. Here two such distance metrics, the Chebyshev distance and the City Block distance, will be introduced. These can be used to compare results from MC simulations of power flows. In order to determine the desired number of iterations, the results from runs with different number of iterations can be compared to each other using these metrics. A small distance between the results indicates that they are similar.

The Chebyshev distance is the L_∞ norm of the difference between the two vectors that are to be compared and is described by (9) [17].

$$D = \max_j \sqrt{(x_{sj} - y_{tj})^2} \tag{9}$$

City block distance (or Manhattan distance) is the $L1$-norm of the difference and is equivalent to the sum of the absolute difference. The expression is presented in (10) [17].

$$D = \sum_{j=1}^{n} |x_{sj} - y_{tj}| \tag{10}$$

These two distance metrics can be compared to the euclidian distance which is the shortest distance between two points in a given space. When comparing distances the euclidian distance is usually the go-to metric, when comparing histograms however, the Chebyshev or City Block distances are preferred [17].

2.7 Distributions of Modelled Load and Generation

The bus loads in PLF calculations are typically represented by normal distributions [18],[19]. The density function of the normal distribution is shown in (11), where σ^2 is the variance and μ is the mean value of the distribution and x is for example the bus load.

$$f(x) = \frac{1}{\sqrt{(2\pi)}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{11}$$

When modelling the wind power generation a normal distribution usually does not represent the data accurately enough [16]. This introduces a need to consider different approaches to power flow modelling from wind power generation. For some cases an exponential distribution might be a good fit. The density function of the exponential distribution is presented in (12).

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{12}$$

Here λ (> 0) is the so called rate parameter of the distribution. The mean value of the exponential distribution is defined as $\mu = 1/\lambda$.

2.7.1 Correlations

In order to be able to generate the input data in a representative way, correlations between the different distributions needs to be taken into consideration. This can be handled in several different ways [8].

If normal distributions are used, one method is to estimate the multivariate normal distribution that represents the data. This models the combined distribution of multiple variables, with correlations captured. It can be used to describe a set of correlated random variables, each of which is centered around a mean value. Basically, it is a generalisation of the one-dimensional normal

distribution to higher dimensions. Mathematically the multivariate normal distribution of a k-dimensional random vector $\mathbf{X} = [X_1, ..., X_k]^T$ can be written with the following notation

$$\mathbf{X} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{13}$$

where $\boldsymbol{\mu} = E[\mathbf{X}] = [E[X_1], ..., E[X_k]]^T$ is the k-dimensional mean vector and $\boldsymbol{\Sigma}$ is the kxk-dimensional covariance matrix such that

$$\Sigma_{i,j} = \mathrm{E}\left[(X_i - \mu_i)(X_j - \mu_j)\right] = \mathrm{Cov}\left[X_i, X_j\right]. \tag{14}$$

If one or more of the variables comes from another statistical distribution and no multivariate distribution can be found that explains their combined distribution, another approach is required. If the individual distributions of the variables and the correlation between them are known, copulas can be used. A brief description of the copula-technique will follow since a detailed explanation is beyond the scope of this thesis [20].

Copulas is a mechanism that allows us to separate the marginal distributions from the dependency structure of a multivariate distribution. It is possible to construct any multivariate distribution by specifying the marginal distribution and the copula. The copula is a function which generates a combination of values between 0 and 1 such that the relationship implied by the correlation between the variables is maintained [21].

Normally when calculating correlations, the Pearson's correlation coefficient (PCC) is used. This is a measure of linear correlation and is good for calculating a multivariate normal distribution. The PCC is the covariance of two variables divided by the product of their standard deviations. Essentially, the PCC is a normalized measure of the covariance. Equation (15) shows the equation for the PCC.

$$\rho_{X,Y} = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y} \tag{15}$$

In order to use copulas however, the correlations needs to be replaced with a rank correlation. Rank correlation is a statistic which measures the relationship between rankings of different variables or of different rankings of the same variable. Ranking is the assignment of ordering labels such as "first", "second" and so on to different observations of the variable. One measure of rank correlation is Kendall's Tau. This measure will be high if the rank of two variables are similar. For a set of observations of two joint random variables X and Y, any pair of observations (x_i, y_i) and (x_j, y_j) with $i < j$ is said to be concordent if the sort order of (x_i, x_j) and (y_i, y_j) is the same, otherwise they are discordant. Kendall's Tau coefficient is defined as described in (16) [21].

$$\tau = \frac{(\text{ number of concordant pairs }) - (\text{ number of discordant pairs })}{\binom{n}{2}} \tag{16}$$

2.8 Data Centres

As mentioned, one of the cases that will be investigated in this thesis is the adding of a data center (DC) onto one of the stations. This section will give some background to data centers.

DCs can be one or several buildings, or a dedicated space in a building where data is stored in form of different computer systems, storage systems, servers, processors and associated components. In Sweden there are 25 data centers, located from Malmö in the south to Luleå in the north. The power consumption of these DCs range from around 1 MW up to over 100 MW [22].

There are several types of DCs. One way to divide them is by investigating the size of their office space. So called mixed used DCs have large office spaces which consume about 20 % of the energy of the total DC. Mixed used DCs have a fluctuating load. This means that the load is higher during office hours and cold times of the year due to the heating of the office spaces. Mixed used DCs have a load profile more resembling that of commercial buildings [23].

DCs with small or negligible office space (around 5 % of total energy consumption) will here be called flat-load DCs. Flat load DCs have a more or less constant load. This means that the average load in summer and winter are almost identical, and the same goes for the daily and weekly loads. Characteristic for these DCs is that the energy use shows little correlation with weather [23].

3 Modelling and Presentation of Data

In order to investigate the use of a MC simulation for making load flow calculations, a network model of the Swedish transmission and sub-transmission systems have been used. This network model was provided by E.ON and contains in total 4 588 nodes. A study area of 10 nodes in a meshed 50 kV network has been selected for the study. Figure 5 shows a flow chart in which the input and output data in the simulation is defined. The input values are generated distributions, based on historical real-life data of the loads and generation in the selected study area. The output values are branch power flows in the selected study area. These can be compared to validation data of the actual power flows in the study area, which have been provided by E.ON.

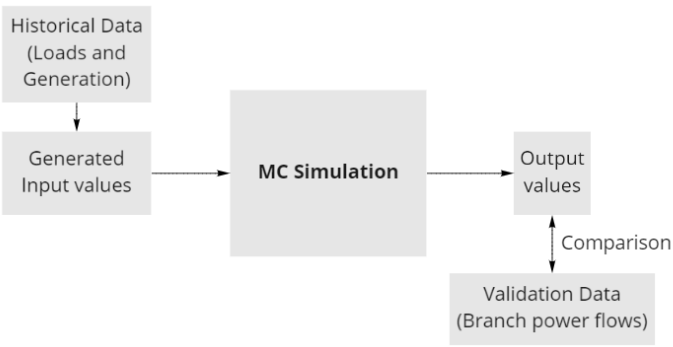


Figure 5: Flow chart presenting the different data sources in the MC simulation method.

3.1 Selected Study Area

The selected study area to be investigated in this thesis is a system of 10 stations in a meshed 50 kV network. The network is shown in figure 6. Stations A, B, C, D, E, F, H, I, and J have loads connected to their underlying buses. Stations A, C, D, and F have wind power generation and station G has a CHP plant connected to the buses. Stations C and I are connected to the sub-transmission network. The marked branches B-C, C-D, F-G, G-H and I-J have validation branch flow data available.

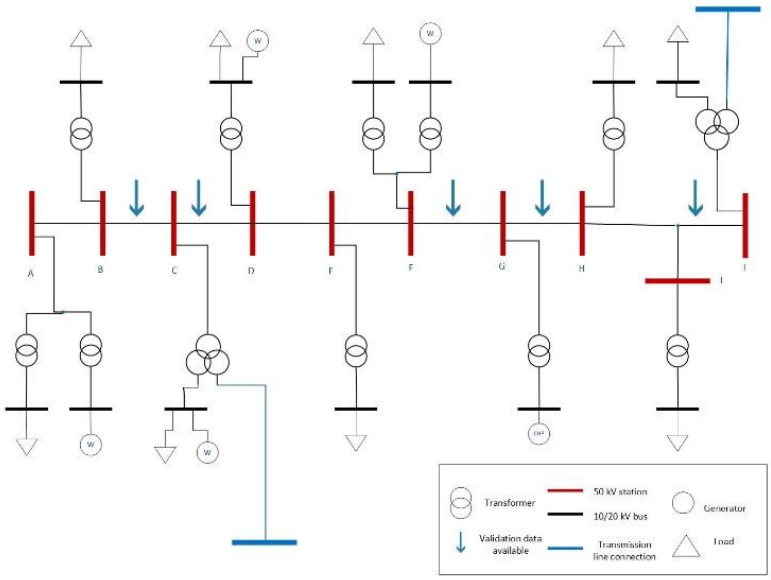
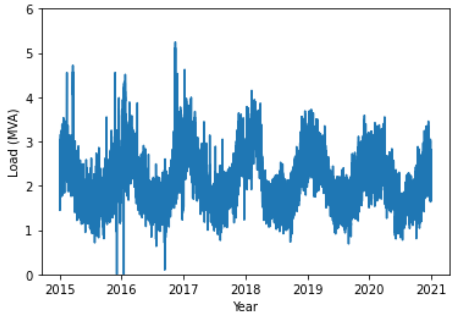


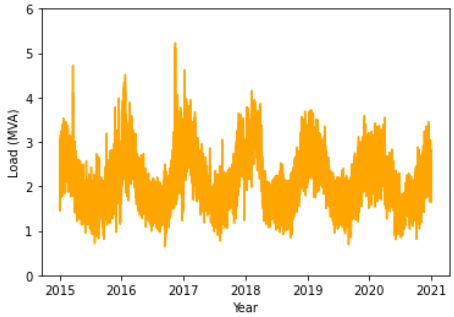
Figure 6: Single line diagram of the selected Swedish sub-system to be considered in the thesis.

3.2 Handling Outliers and Presenting the Data

The historical data consist of hourly load flow values through the transformers connecting the 50 kV and 10/20 kV buses. The data has been collected from a time span of 6 years. Furthermore, validation data in the form of branch power flows have been collected from the same time period. The time series for all six years from one of the loads is presented in figure 7 (a) as an example of what the collected data looks like. As can be observed in the figure, the data has some obvious outliers. To ensure the quality of the historical data, some of these outliers have been removed by an algorithm and replaced with interpolated values. Figure 7 (b) presents the same load as (a) but with the outliers removed.



(a)



(b)

Figure 7: The load in bus B over six years. The first picture (blue) presents the raw data, the second picture (orange) presents the data with outliers removed.

The algorithm used for outlier-removal uses a moving average and a moving standard deviation. From a window of 48 data points an average and a standard deviation is calculated. Values within this window that are located more than three standard deviations from the mean value is removed and replaced with an interpolated value. This window is moving forward one step at the time until all data points have been gone through. The algorithm is presented below.

```
1. select 48 data points
2. calculate mean and std of the 48 data points
3. upper limit = current mean + 3*std
4. lower limit = current mean - 3*std
5. if value outside limits:
       replace with interpolated value

   repeat
```

This procedure was done for all historical load data, all historical generation data and all validation branch power flow data. For the validation branch data however, the window was set to 500 data points instead of 48, due to the slower fluctuations observed in this data.

The number of removed and replaced outliers constitutes between 0.05 % and 2.2 % of the data points for the loads, between 1.7 % and 5.6 % of the wind

power generation data, almost 8 % of the CHP generation data and between 0.11 % and 2.42 % of the branch power flow data (the validation data). As seen in figure 7 (b) some values that might be considered outliers are left. This is since the limit for what an outlier is intentionally was set quite high in order to make sure that all possible extreme events would be taken into consideration.

In order to present this data in a consistent and clear manner, normalized histograms have been applied. In figure 8 two histograms are presented, on the left hand side without a normalized y-axis and on the right hand side with a normalized y-axis. The values have been normalized to a scale from 0 to 1 throughout the report. The main reason for the use of normalized histograms is to be able to compare the different histograms with each other in an easier manner. The absolute numbers on the y-axis in the histogram to the left is of no importance since they only represent the number of iterations chosen in the simulation, and is thus not relevant when comparing with other data sets.

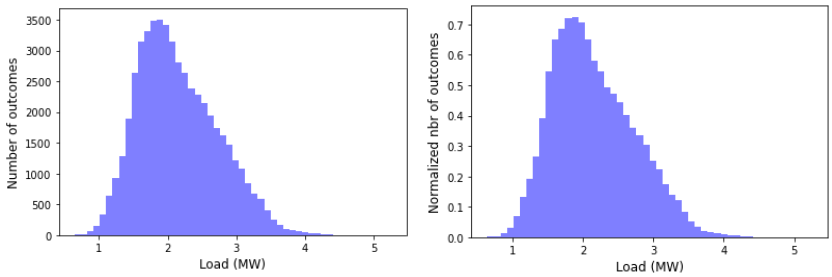


Figure 8: To the left is a histogram presenting the load in node B without normalized axis. To the right the same histogram is presented with normalized y-axis.

When working with statistical distributions, the important information received from a histogram is the shape of the distribution. In order to better present this, the use of individually scaled axes have been applied in this report. Figure 9 shows a comparison of the 9 load data sets presented in normalized histograms with equal axes to the left, and with individually scaled axes to the right. It becomes clear that data which is spread over a larger span, such as bus J, will have much lower peaks than data spread over smaller spans. The use of normalized histograms with individually scaled axes may result in the loss of some information in the form of the relative heights of the peaks, which the reader can keep in mind throughout the report.

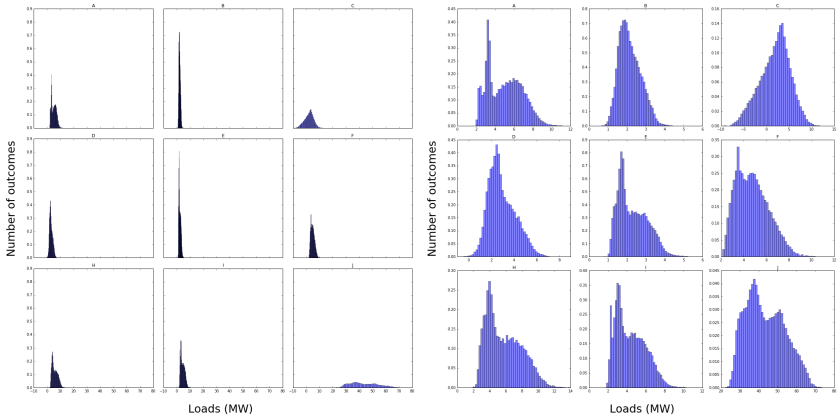


Figure 9: Normalized histograms of the historical load data sets presented with equal axes to the left and with individually scaled axes to the right.

3.3 Historical Load and Generation Data

The input values in the selected study area in figure 6 consisted of 9 loads, 4 wind power stations and 1 CHP plant. Input distributions for these were generated based on historical data, which was provided by E.ON. The data collected was the load flow through the transformers connecting the 10/20 kV buses to the 50 kV stations. The selected time span of collected data was from 2015-01-01 00:00:00 to 2020-12-31 23:00:00. The data was in the form of average hourly power flow through the transformer, which was the form of data which was provided by E.ON.

In some of the stations, as can be seen in figure 6, both load and wind power generation were connected. This was the case in stations A, C, D, and F. The historical data at these buses consisted of separate measurements of the total power flow and the power flow from the wind power plants. In order to separate the load from the wind power generation, the measured wind power production has been subtracted from the total power flow. The load data and the generation data have then been treated separately in the study due to their different characteristics. However, in some stations complete data was not available. This was the case in bus C, where not all wind power generation could be collected separately. This might have led to the data at this station being corrupted. Figures 10 - 12 show the historical data plotted in normalized histograms.

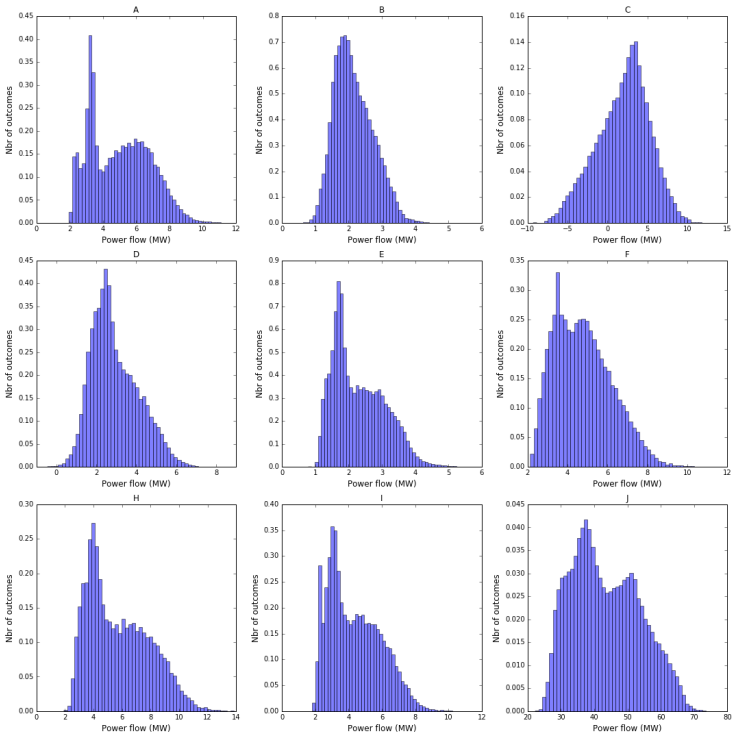


Figure 10: Historical data - Loads. Presented in normalized histograms. Note the use of different axes.

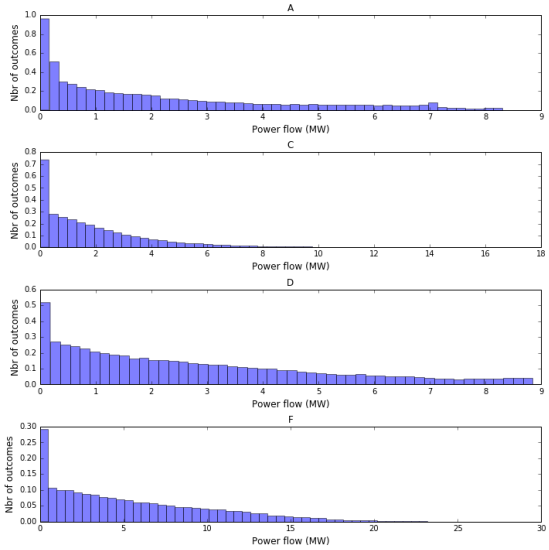


Figure 11: Historical data - Wind power generation. Presented in normalized histograms. Note the use of different axes.

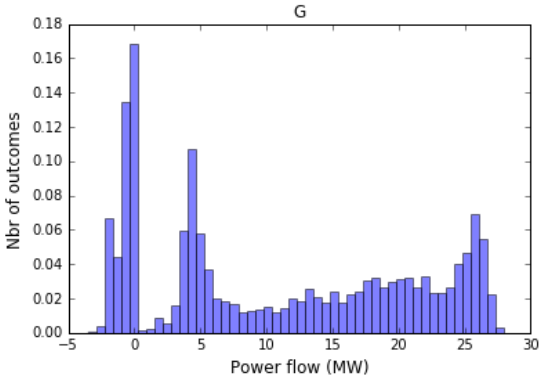


Figure 12: Historical data - CHP generation. Presented in normalized histogram.

When examining the historical wind power data, the average capacity factor was calculated to around 0.225. The capacity factor is a measurement of the ratio between the average output power to the maximum power rating of the wind power generation. A capacity factor of 0.225 in the selected sub-system can be compared with the usual span of capacity factors of wind power in Sweden which lies in the span of 0.2-0.4. The average capacity factor in Sweden is 0.26 [24].

The stations in the selected study area were all located in the same geographical region and were directly connected to each other in a chain-like formation as can be seen in figure 6. The load and CHP generation are both highly temperature dependent, as the loads increase with lower temperature and the CHP generation is dispatched when heat is needed. The wind power generation is dependent on wind speeds. This suggests a correlation between the historical data at the different stations, which has been calculated based on Pearson's correlation coefficient and is presented in figure 13. From this correlation matrix it can be seen that the loads and CHP generation are closely correlated, as can be expected based on their temperature dependency. The loads and CHP generation have therefore been grouped together in the rest of this report and treated collectively. The wind power generation variables are also highly correlated to each other, which can be explained by their wind speed dependence. The four wind power generation sets have thus been grouped together as well.

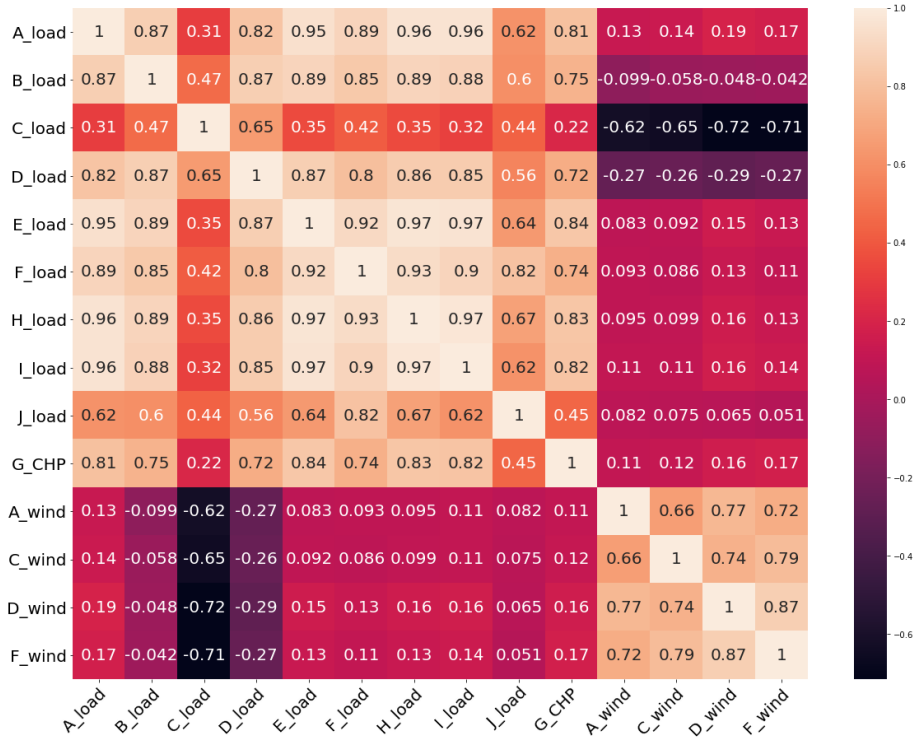


Figure 13: Correlation of historical load and generation data between the stations in the selected study area.

As can be seen in the matrix, the load at station C is not as correlated with the other loads as the other stations. As mentioned above, complete data was not available at this station. This becomes evident here, as the load at this stations is not as correlated to the other stations as expected. Furthermore, the load at station C is to a higher extent correlated to the other wind power stations than would be expected. This can also be seen in the matrix in figure 13 where high negative correlation is clearly visible. The reason for the correlation being negative is that the power at stations B, C and D flows in the opposite direction compared to the rest of the stations. This indicates that the load data at station C still has a significant amount of wind power generation included, which is a possible source of error.

3.4 Validation Branch Power Flow Data

The selected study area consists of 9 branches where output values were generated in the form of branch power flows. At 5 of these branches, validation data was available which could be used to verify the results. This data was taken from the same time span as the input data. The branches where validation data was available are B-C, C-D, F-G, G-H and I-J in the network in figure 6. The normalized histograms of the validation data are shown in figure 14.

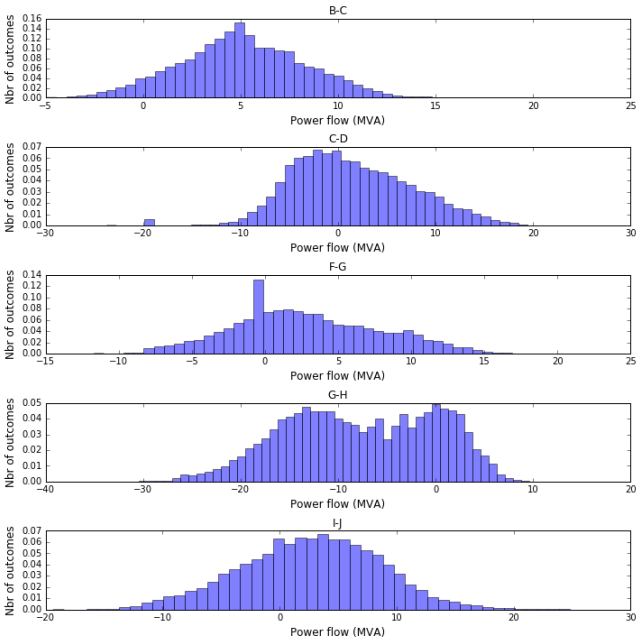


Figure 14: Historical data - Branch power flow. Note the different scales on the axes.

4 Suggested Probabilistic Method

This section describes the suggested PLF method, which is a MC simulation multiple-scenario method. The set-up of the MC simulation is presented, and two different trials on how to generate the method's input values are investigated and compared.

4.1 The Network Model in PSS/E

Throughout the study the Newton Raphson-based iterative power flow solver PSS/E was used to solve the power flow equations (7) for the defined input values. The network model used in PSS/E was a model of the Swedish transmission system provided by E.ON. This network model includes all of the transmission network and E.ON's sub-transmission network, as well as international connections. Everything down to 50 kV is included, and the lower voltage levels are modelled as aggregated loads or generation at the connection points of the transmission network and lower-voltage networks. These connection points are modelled as 20/10 kV buses which are connected to the 50 kV stations through transformers. The selected system is presented in figure 6.

For the network model in PSS/E, a pre-defined case was provided by E.ON. This case defined the fixed parameters in the model. The case was the maximum load case used when making deterministic calculations of the risk of thermal overload in the beginning of 2021. This case gives higher branch flows in general than the validation data and thus the case was scaled down to reach results closer to the historical measurements. The loads in the south of Sweden were scaled down 10 %. Furthermore, all nuclear power production was set to 8 % below maximal production. When this case had been defined in PSS/E, it was kept constant throughout all simulations, with the exception of the inputs at the 10 stations in the selected study area.

4.2 Suggested Probabilistic Method

The suggested probabilistic method is a MC simulation of the power flow system, resulting in a multiple-scenario based model. In the MC simulation, the varied input values were load and generation in the stations in the study area. These were the only parameters which were varied in the study, as it focused on how newly installed load and generation affects the power flow in the branches. If other parameters are chosen as input values, such as for example the topology of the network, different analyses could be made. For each iteration, a set of random input values were picked for the 9 loads, 4 wind power stations and 1 CHP unit. The system was solved by PSS/E and output data was generated. This process was then iterated a set number of times. After the MC simulation was run, the resulting values of each output variable were treated as probability distributions. As the output variables are branch power flows, the resulting distribution describes the probability of the branch power flow taking different values. A flow chart of the suggested MC simulation is shown in figure 15.

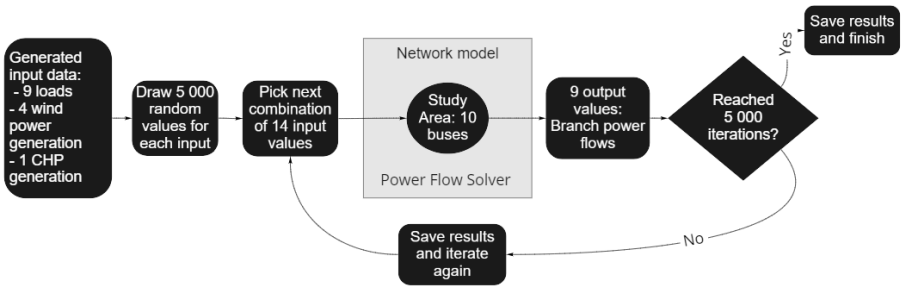


Figure 15: Flow chart of the suggested Monte Carlo simulation.

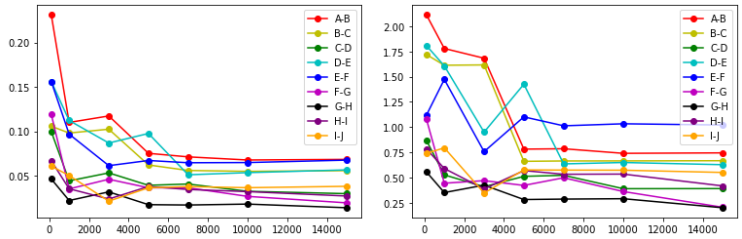
In order to get closer to a risk-based modelling approach, the suggested method could be used in combination with pre-defined overhead line ratings. If a level of accepted risk of thermal overload is decided, the MC simulation could be used to investigate how to optimize parts of the power system without exceeding this level.

4.2.1 Number of Iterations

In order to determine the number of iterations needed in the MC simulation, the number of iterations needed to reach convergence was decided on. The MC simulation was run with 100, 1 000, 3 000, 5 000, 7 000, 10 000, 15 000 and 20 000 iterations. The execution time was noted for each run.

The result when simulating with 20 000 iterations was used as benchmark when comparing the Chebyshev and City Block distances. The comparison was done by calculating the distance between the histogram from the simulation with 20 000 iterations with the histograms from the other simulations. A small distance is implying that the histograms are similar. When the result of the distance-tests were converging, a conclusion was drawn that the result would not change with higher number of iterations.

The results from these tests are presented in figure 16 where the two different distance metrics are plotted for all branches and for different number of iterations. It's clear that after a while both the Chebyshev distance and the City Block distance are converging, indicating that more iterations do not change the result. At 5 000 iterations all branches but one has converged. The execution time was around 200 s, which is significantly shorter than the execution time for 7 000 iteration which was 300 s. Even though one of the branches still had not converged at this point, the result from 5 000 iterations was considered to be good enough. Hence, 5 000 iterations were chosen as a sufficient number of iterations for the suggested MC simulation.



(a) Chebyshev distance (b) City block distance

Figure 16: The different distance metric presented for all branches. The x-axis present number of iterations, the y-axis is the distance between the histogram from the run with 20 000 iterations and the current histogram.

4.3 Generating Input Data: Trial 1

An issue when working with MC simulations is how to draw the random input values. In this study, generated statistical distributions were used from which random values were drawn. Two different types of input distributions have been examined, which are referred to as Trial 1 and Trial 2. In order to investigate which method of generating input values more accurately describes the actual loads and generations in the study area, the resulting power flows have been compared to the validation power flows.

The first trial of generating input data was to create uncorrelated statistical distributions. When observing the histograms of the historical data in figure 10, normal distributions seemed to be the best fit for the loads. The histogram showing the historical data of the CHP power plant in bus G as shown in figure 12 did not follow a normal distribution as clearly. However, theory suggests that both load and CHP generation are usually modelled by normal distributions, and therefore normal distributions where fitted to the historical loads and the historical CHP. These normal distributions were used for generating the input values of loads and CHP generation in Trial 1.

Figure 11 suggests that the historical data of the wind power generation may follow exponential distributions. Therefore exponential distributions were fitted to the historical data and used for generating the input data for the wind power generation in Trial 1. The different input data were generated independently. This means that the distributions in Trial 1 did not take into consideration any correlation between the input variables.

4.3.1 Generated Input Values

The generated normal distributions for the load input data are shown in figure 17, plotted in normalized histograms. The historical load data is shown in the same plot for comparison. The total number of samples for each load data set is 5 000.

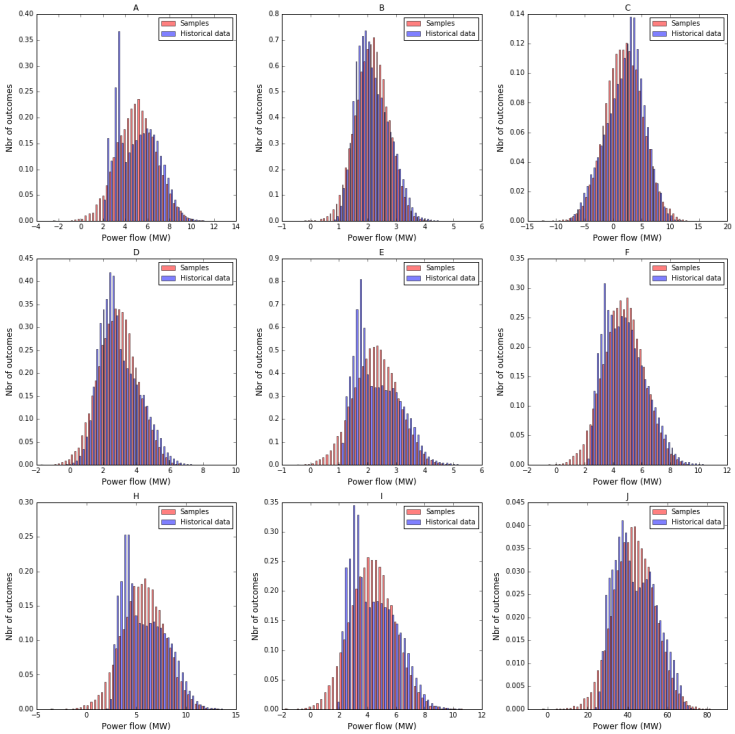


Figure 17: Generated normal distribution samples of loads (red) and historical load data (blue). Presented in normalized histograms.

Figure 18 shows histograms of the generated input data for wind power generation, based on exponential distributions. The historical data is shown in the same plot as a comparison.

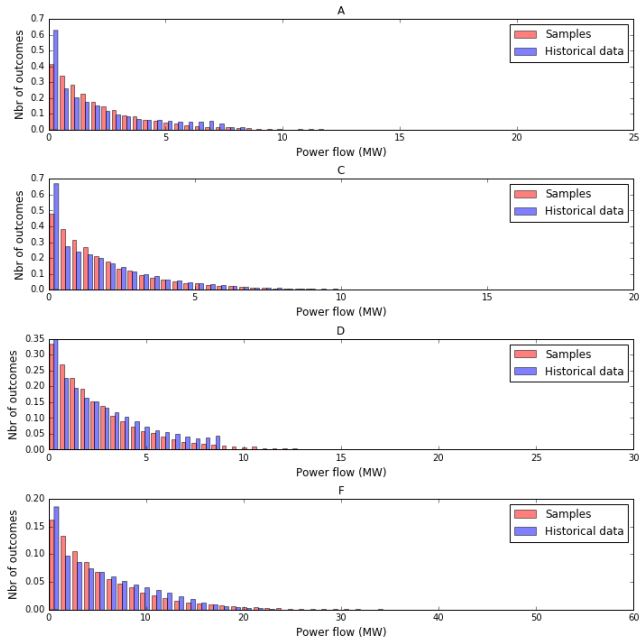


Figure 18: Generated exponential distribution samples of wind power generation (red) and historical wind power generation data (blue). Presented in normalized histograms.

Lastly, figure 19 shows the generated input data for the CHP generation, as a normal distribution. This is shown in a normalized histogram, together with the historical data for comparison.

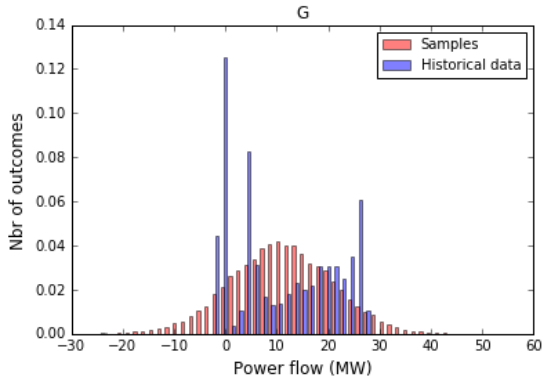


Figure 19: Generated normal distribution samples of CHP generation (red) and historical CHP generation data (blue). Presented in a normalized histogram.

4.3.2 Resulting Output Data

The output data from the model in form of power flows in the branches is presented in figure 20. The validation power flows are shown in the same plot for comparison.

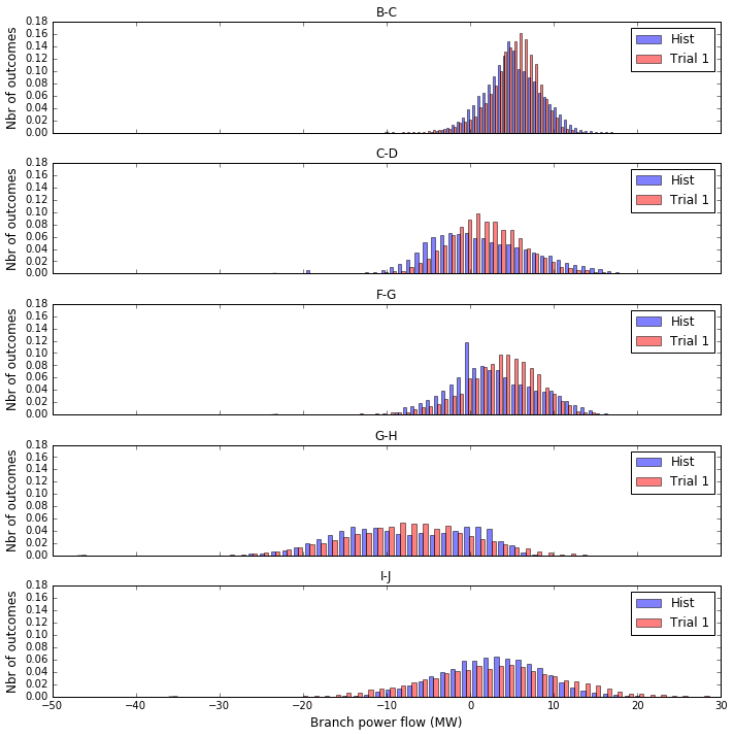


Figure 20: Resulting output power flows (red) plotted in normalized histograms. Validation data (blue) included as a comparison.

4.4 Generating Input Data: Trial 2

The second trial takes the correlation between the stations in the study area into consideration. As seen in figure 13, the loads are correlated with each other with correlation coefficients from 75 % up to 95 %. The wind power generations are correlated to each other with correlation coefficients from 55 % up to about 87 %. It can also be seen that the CHP power plant in station G is correlated with the loads from 45% to 84%, which is more than with the wind power generation. The approach has therefore been to generate the loads and CHP generation samples correlated to each other. Likewise the wind power generation samples have been generated correlated to the other wind power stations.

For the generated load and CHP input values, a multivariate normal distribution was used. The distribution was constructed using the mean values and covariances of the historical load and CHP generation data. The multivariate normal distribution describes a set of correlated random variables in the form of loads and generation, were each variable is clustered around a mean value. The mean values where given by the historical mean values for each of the loads and CHP plant respectively. The correlations between the buses were captured via the generated covariance matrix, as presented in table 1. 5 000 random variables were drawn from the distribution for each bus in the system. These random values where then set as the generated samples of the different loads and the generation from the CHP plant respectively.

Table 1: Covariance matrix for historical loads and CHP generation.

	A	B	C	D	E	F	H	I	J	G (CHP)
A	3.206	0.903	1.806	1.756	1.324	2.294	3.693	2.671	11.256	14.351
B	0.903	0.336	0.897	0.598	0.402	0.703	1.113	0.795	3.559	4.297
C	1.806	0.897	10.915	2.518	0.896	1.958	2.486	1.639	14.585	6.982
D	1.756	0.598	2.518	1.415	0.801	1.369	2.185	1.571	6.796	8.483
E	1.324	0.402	0.896	0.801	0.603	1.020	1.614	1.162	5.069	6.449
F	2.294	0.703	1.958	1.369	1.020	2.050	2.863	1.991	11.905	10.461
H	3.693	1.113	2.486	2.185	1.614	2.863	4.607	3.240	14.696	17.664
I	2.671	0.795	1.639	1.571	1.162	1.991	3.240	2.400	9.706	12.616
J	11.256	3.559	14.585	6.796	5.069	11.905	14.696	9.706	103.545	45.167
G (CHP)	14.351	4.297	6.982	8.483	6.449	10.461	17.664	12.616	45.167	98.155

In order to generate correlated random samples from exponential distributions, a method using copulas was used. This technique is somewhat more complicated than using multivariate normal distributions as described above. A correlation matrix between the wind power generations was calculated based on Kendall's Tau. This is presented in table 2. Exponential distributions were fitted to the data as in Trial 1. Copulas with the desired number of correlated entries were generated. Using the copulas in combination with the exponential distributions, 5 000 correlated and random input values were obtained.

Table 2: Correlation matrix generated for the historical wind power generation.

	A	C	D	F
A	1.000	0.632	0.481	0.642
C	0.632	1.000	0.556	0.721
D	0.481	0.556	1.000	0.518
F	0.642	0.721	0.518	1.000

4.4.1 Generated Input Values

Figures 21 - 22 show the generated input values obtained using the method presented above for loads and CHP generation.

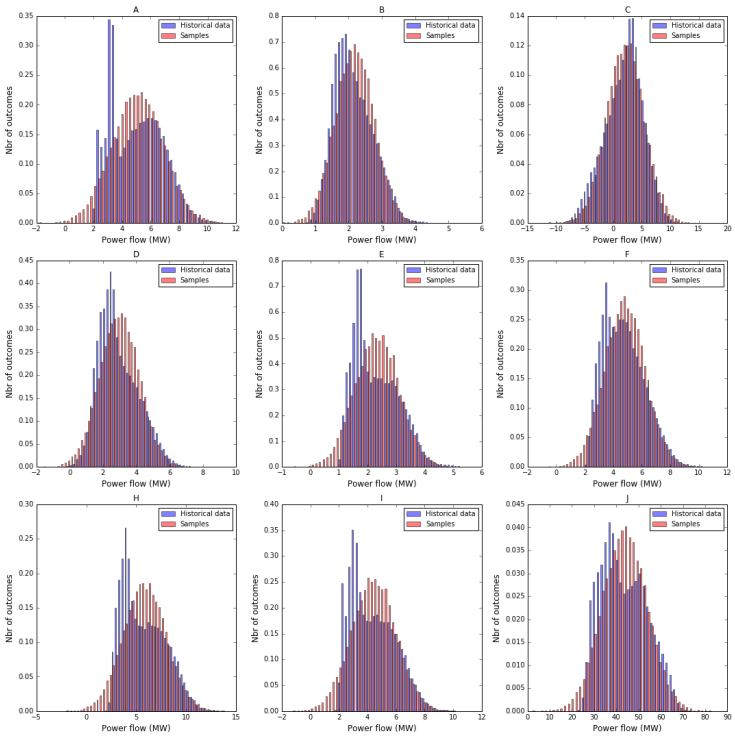


Figure 21: Generated load input values using multivariate distributions (red) and historical load data (blue). Presented in normalized histograms.

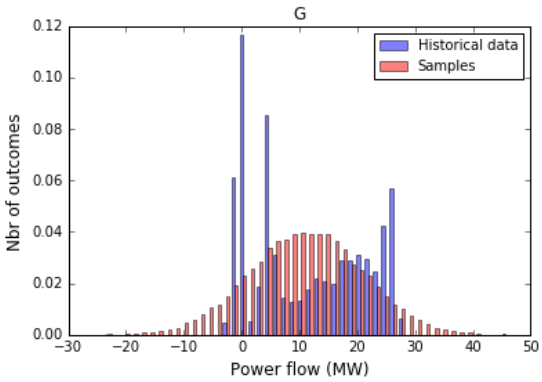


Figure 22: Generated CHP generation input values using multivariate distributions (red) and historical CHP generation data (blue). Presented in normalized histogram.

A correlations matrix was calculated to examine how well the samples were correlated. This correlation matrix is shown in figure 23. If compared with the historical correlations presented in figure 13, it can be seen that the correlations are similar.

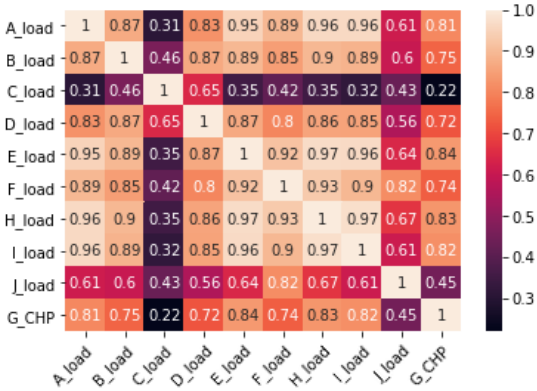


Figure 23: Correlation between generation input load and CHP generation values.

Figure 24 presents the generated input values of wind power generation obtained using the copula method. The historical wind power generation data is plotted in the same figure for comparison.

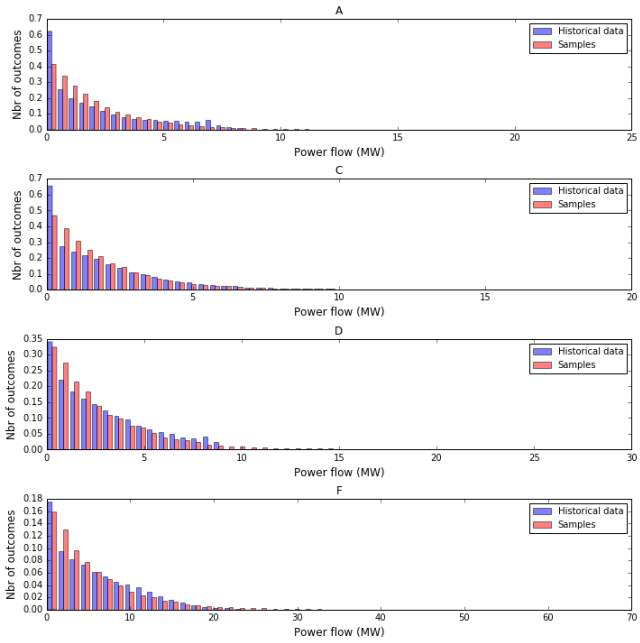


Figure 24: Generated correlated wind power generation values (red) and historical wind power generation data (blue). Presented in normalized histograms.

In order to check how well the samples were correlated, correlation matrices where calculated. Figure 25 presents the calculated correlation between the different input generations. Again, when compared to the historical correlations presented in figure 13 it can be seen that the correlations are similar.

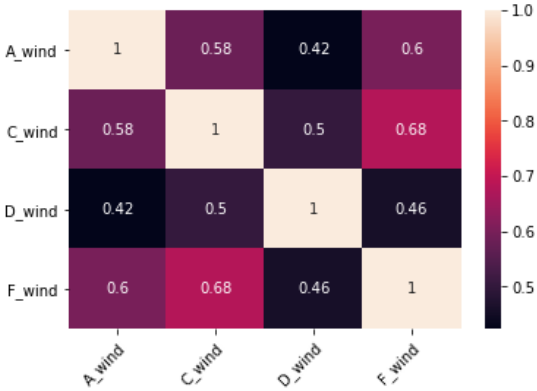


Figure 25: Correlation between generated wind power generation input values.

4.4.2 Resulting Output Values

The output power flows obtained using the method in trial 2 are presented in figure 26. The validation data for each branch is plotted together with the result for comparison. Even though the input variables are correlated, the correlation of the output values have not been evaluated since the results are generated in a random order.

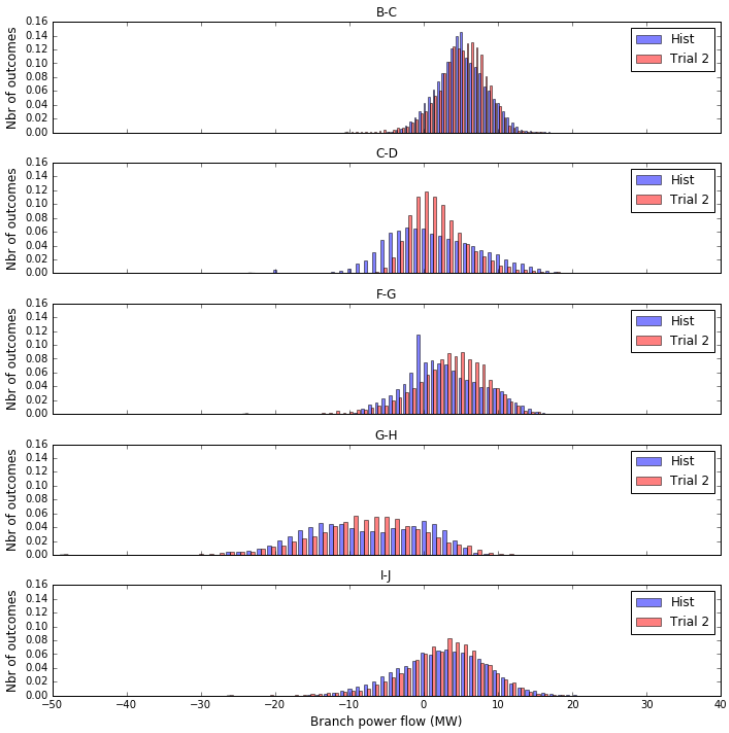


Figure 26: Output branch power flows (red) and validation data (blue) plotted in normalized histograms.

4.5 Comparison of Trials

The accuracy of the two trials were compared base on the mean values and standard deviations of the results. These were compared to the mean values and standard deviations of the validation data. These are all presented in figure 27.

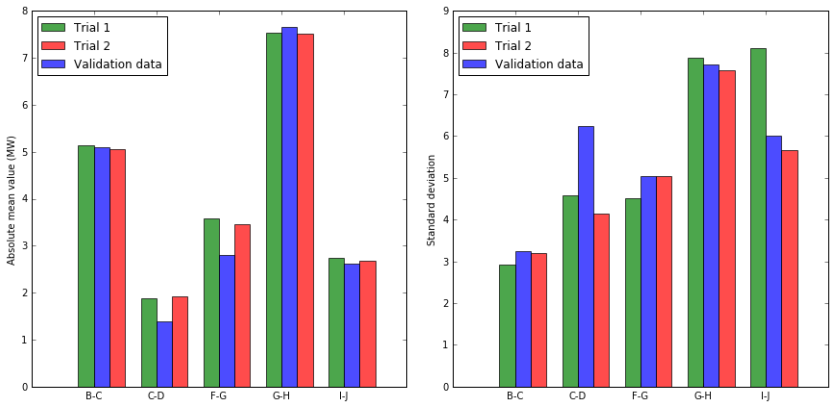


Figure 27: Comparison of resulting branch power flows in Trials 1 and 2. Left: Absolute mean values in simulated and validation branch power flows. Right: Standard deviations in simulated and validation branch power flows.

From the comparison of mean values and standard deviations, it can be concluded that Trial 2 gives more accurate standard deviations than Trial 1 in the branches F-G, G-H and I-J. The mean values do not differ much between the trials. It can be noted that the branch C-D gets a significantly higher difference in the standard deviation in Trial 2 than in Trial 1. However, the simulated power flows in this branch has a different shape than the validation data for both variations of input values, as can be seen in figures 20 and 26. This may be due to some error in the modelling of input values or in the MC simulation, but does not seem to be affected by the choice of input values.

As the method gives more accurate results when using the input values of Trial 2, the rest of the study will be based on this method for generating input values to the MC simulation.

5 Case Study

The suggested MC simulation method for investigating the risk of thermal overload as defined in Section 4 was applied on three different cases. The first case was an application of the method on a real-life installation of wind power generation which took place in the selected study area in 2014. This case aimed to further validate the method. The following case investigated a possible new installation in the study area in the form of a data centre. This case aimed to present how the method can be applied in a possible future scenario. The third case is an approach towards a risk-based method. It investigates how much wind power that can be installed in a specific bus without exceeding a set level of risk. The cases are all based on the input values of Trial 2, using 5 000 iterations.

5.1 Case 1: Added Wind Power Generation at Station D

The first case examined how a manually added wind power generation onto a station in the system compares with historical data before and after such an installation was made onto the grid. The real-life event was an installation of 4 MW wind power at station D in 2014, in addition to the previously installed 8 MW. For station D, generation data one year before and after the installation was available.

The same method for generating the input values as had been applied in Trial 2 was used as a base. Distributions with 5 000 random values for each input variable were created for the buses in the system, except for the generation input values at station D itself. In order to get input values for the wind power generation in station D before the installation in 2014, the input values for this bus were based on the historical data from the year before the installation. An exponential distribution was fitted to this data and used as the generation input data at station D. As the data in station D could not be used to calculate the new correlation with the other wind power input variables, the correlation was set to 0.5. The reason why the correlation could not be calculated was the different time periods of the data. As can be seen in table 2, the correlations using Kendall's Tau differed between 0.481-0.721 when calculated in Trial 2. The correlation of 0.5 is at the lower end of this measured correlation-span. A normalized histogram showing the input values before the new generation was added is shown in figure 28. The created samples are shown in red and the historical data from the 1-year period before the installation is plotted as a comparison in blue.

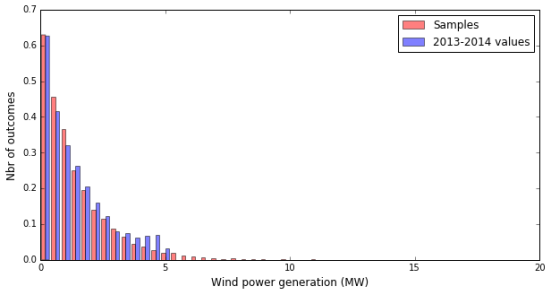


Figure 28: Input values for the wind power generation in bus D after 4 MW of installed capacity was added. Compares simulated values (red) with historical values (blue). Presented in a normalized histogram.

The installed generation was added manually to station D by setting up an exponential distribution with a mean value of $0.225 \cdot 4$ MW. This corresponds to the average capacity factor of the wind power generation as discussed in section 3.3. This distribution was defined to have a correlation to the other buses of 0.5, as discussed above. An exception was made for the correlation to station D which was set to 0.7. This was due to the likelihood of the newly installed wind power having a higher correlation to the station where it is installed than to the other stations. When the correlation matrix and distributions were set up, correlated input values were drawn for the four buses and for the new distribution. The generated values from the new generation was then added to the input values at station D. Figure 29 shows the created input values in red and the historical values for the year following the installation in 2014 in blue as a comparison.

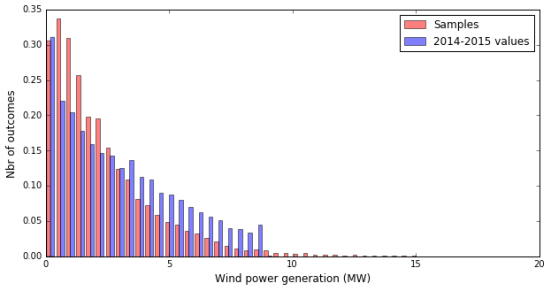


Figure 29: Input values for the wind power generation in bus D after 4 MW of installed capacity was added. Compares simulated values (red) with historical values (blue). Presented in a normalized histogram.

5.1.1 Results

The two different sample collections were sent into the MC simulation separately in order to compare how well the manual addition of wind power generation corresponds with the actual installation made in 2014. The simulated power flows before (red) and after (yellow) the installation are presented in normalized histograms in figure 30. The validation data from after the installation (blue) is plotted in the same figure for comparison.

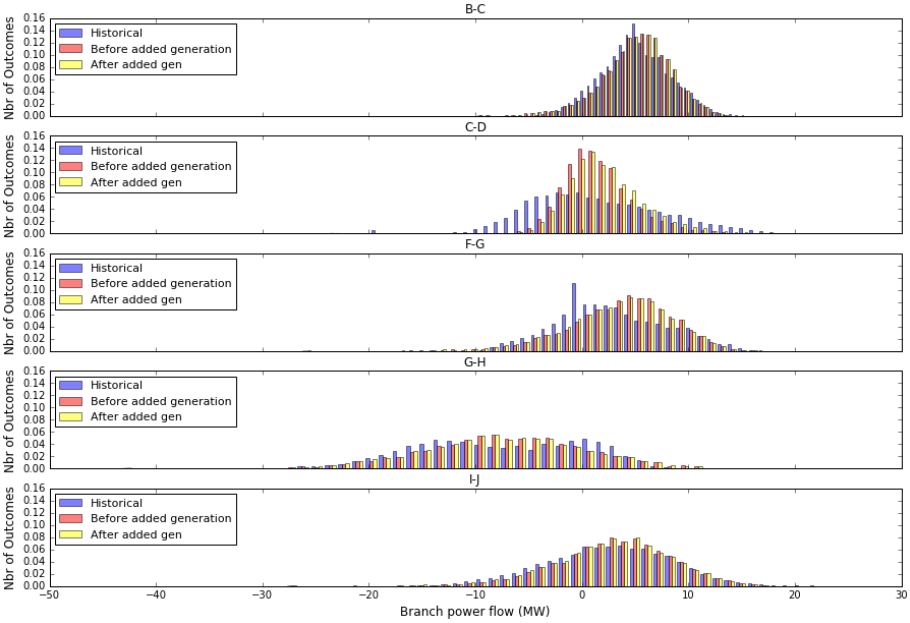


Figure 30: Resulting power flow in branches before adding generation (red) and after adding generation (yellow). Compares simulated output data with validation data (blue). Data presented in normalized histograms.

The aim with this case was to investigate if the manually added generation resulted in more accurate branch power flows than in the simulation without the added generation. As this is difficult to analyse in figure 30 due to the small changes, mean values and standard deviations were of the data were calculated before and after the installation. The mean values and standard deviations of the simulated output values are shown in figure 31. The mean values and standard deviations of the validation data are shown in the same plot as a comparison.

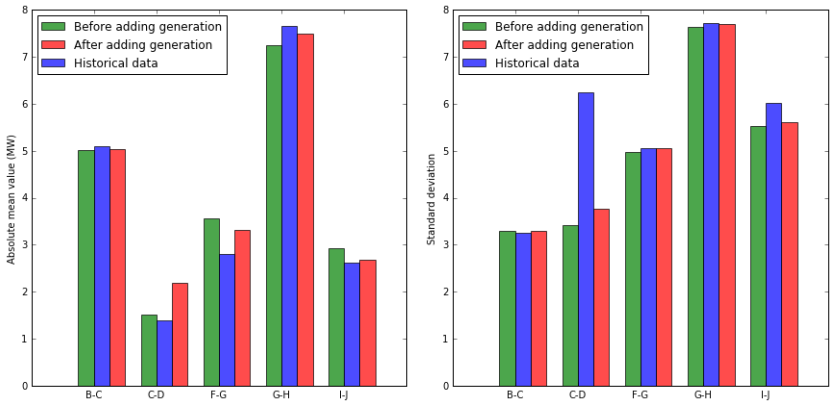


Figure 31: Comparison of simulated branch power flows before and after adding generation in Case 1. Left: Absolute mean values of branch flows in simulated and validation output data. Right: Standard deviations of branch flows in simulated and validation output data.

It can be concluded that the simulated mean values get closer to the validation mean values when the correct wind power is added to bus D. The exception is the branch C-D, where the opposite has happened. When looking at figure 32 however, it can be seen that the simulated power flow in this branch lies quite far from the validation power flow in both simulations. As has been discussed earlier, there seems to be an issue with the branch C-D, which affects the resulting output data. The standard deviations do not get much closer to the validation data, even though small changes are observed. These show that adding the correct generation have brought the simulated standard deviation closer to the standard deviation of the validation data. In branch C-D both of the simulated standard deviations are quite far off from the validation data.

This case suggests that the method for adding generation and load manually works well. However, due to the error in branch C-D, this conclusion may be optimistic. For situations when no historical data is available, similar techniques for generating input values may work well.

5.2 Case 2: Added Data Centre in Station H

The second case was an application of how the MC simulation could be used in a possible future scenario. The choice of adding a DC to a station in the selected study area was based on the fact that data centres have a relatively large energy consumption, and is likely to be installed in the area. Adding a DC was therefore thought to be a suitable way to examine the MC simulation. It is also realistic since there exists several data centres of different sizes in Sweden, and there are plans for even more data centres, both in the north and south parts of Sweden.

First of all, the input data as described in Trial 2 were used as a base. The results from the reference case were also used for comparison. A flat load data centre was chosen, as this is the simplest form of data centre in the sense of load types. This implies negligible office space and indicates that the load was

almost constant over the day and year. The load was chosen to be 10 MW which represents quite a large DC but still realistic to be placed anywhere in the country. The load of a flat load DC is centered around one value with small deviations from this value, hence the load was modelled as a normal distribution with mean value 10 MW and standard deviation 0.5.

Since these kind of DCs shows little correlation with weather, it was assumed that the correlation with the other loads would be small and could therefore be neglected. This new load was added to the already existing load in node H. The new load distributions were then used in the MC simulation with the same generation parameters as in Trial 2 and the simulation was iterated 5 000 times. The reference case was also run with 5 000 iterations and the results were compared.

5.2.1 Results

The resulting power flows in the nine branches before and after adding the DC to station H is plotted together as normalized histograms in figure 32. It can clearly be seen that several of the branches are affected by adding the DC, not only the branches directly connected to the station in question. However, it is obvious that the branches placed to the right of the DC-station are most affected.

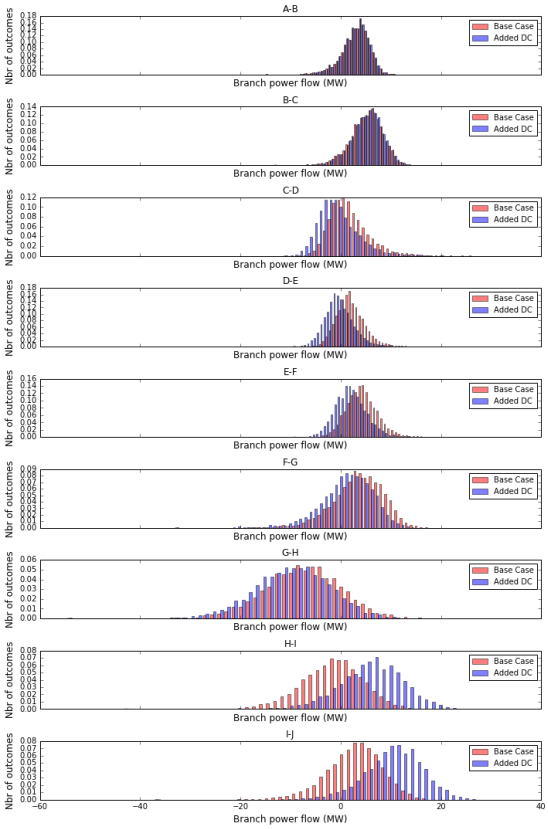


Figure 32: Power flow in the branches after adding load in form of a data centre in node H. For comparison the result from the reference case is plotted in the same figures.

From this case it becomes clear that the suggested MC simulation gives more information about future installations than a traditional deterministic approach. Instead of seeing how the power flows are affected in a few worst-case scenarios, a larger picture of how the installation affects the power flows is achieved. It therefore seems that the general knowledge produced by the multiple-scenario MC method is of use when examining the risk of thermal overload in a power grid.

5.3 Case 3: Maximum Added Wind Generation at Station F

The third case was another application of the MC simulation. In this case, the amount of wind power which could be installed at station F was investigated by using a risk-based approach. The aim was to examine if the MC simulation could be applied in such an investigation. In a risk-based model, the accepted level of risk for some unwanted event (such as thermal overload occurring) must be pre-defined. As it is not within the scope of this report to define any level of acceptable risk, this case has merely been used as an illustration of how risk-based modelling could be applied on the power system.

The investigation started with the MC simulation, using input data as defined in Trial 2. More wind power was then added to bus F using the same method as described in Case 1. The mean wind power was again added by using the capacity factor of 0.225. The new wind power was added gradually and the resulting branch power flows from the MC simulation were compared to a set rating of the branches.

The accepted risk level could be in the form of a percentage, a specified number of iterations or a set number of hours per year. In order to illustrate this, the case have been tested on a set number of hours per year (out of the total 8 760 hours in the year) in which the line rating is allowed to be exceeded. If the limit of the number of hours in a year was set to 5 hours, the number of allowed iterations with overload would be approximately 3 out of the 5 000 iterations. If the number instead was set to 10 hours per year, the number of iterations where thermal overload occurred would be allowed to be 6 out of the 5 000 iterations.

5.3.1 Results

A fixed line rating was set to 38 MW as an illustration. This level was chosen because it was sometimes exceeded and could therefore well illustrate how the risk-based method could be used. The only branch in which the power flow exceeded this rating was the branch G-H. Figure 33 shows the number of times out of the 5 000 iterations in which the resulting power flow of branch G-H exceeded 38 MW. The risk levels that have been chosen are 3 and 6 outcomes from the 5 000 iterations respectively. However, the results are still probabilistic, meaning that the amount of exceeding branch power flows are not increasing monotonously. It can be noticed that the limit of 3 outcomes is exceeded when 4 MW of wind power is added, but not when 5 MW of wind power is added. This makes it difficult to draw any conclusion from this case. The method needs to be improved in order to be able to apply it as a risk-based method. Furthermore, how a risk-based method is to be applied on a real life case needs to be examined further.

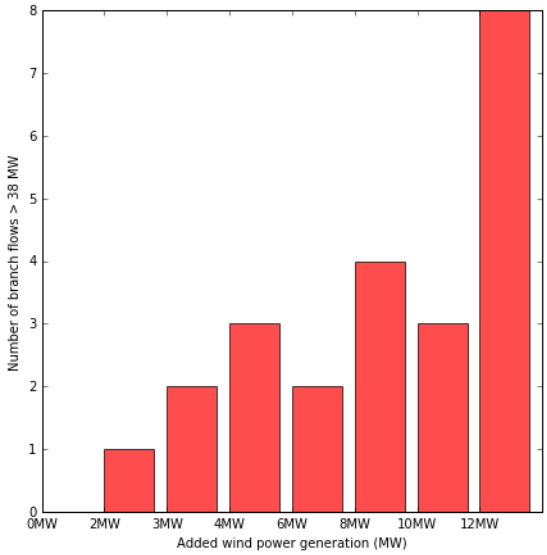


Figure 33: Number of hours where the power flow in branch G-H exceeds 38 MW.

An important issue with this case is the fact that the branch capacities are not constant, but depending on temperature, wind speed, etc. This makes it problematic to compare all resulting power flows to the same rating. In the suggested MC simulation of making load flow calculations, there is no information as to under which circumstances the resulting power flows occur. For that reason, it is not possible to compare the resulting flows with anything else than constant ratings or limits. As this case shows, the branch power flows quickly exceed the set rating of branch G-H as additional generation is added. In reality, these flows could occur at times when the rating is set higher, such as in winter time when the temperature is lower. In this method however, there is no way of proof-checking this. Therefore, this method can not be applied as a risk-based model without some improvements. Possible solutions to this issue would be to include temperature or time dependency in the method.

6 Discussion

In this section the choice of PLF method and its results will be discussed. The choices made when constructing the method, such as the use of historical data, the decided number of iterations, and the generation of input values will be considered. This thesis presents a small scale proof of concept and thus further studies to be done on PLF methods have been identified. These suggestions for future work are presented in the end of this section.

6.1 Choice of Probabilistic Method

The chosen probabilistic method is a MC simulation multiple-scenario method. A multiple-scenario method captures the stochastic behavior of the data, which results in a method which better represents a more generic image of reality than if only a few scenarios are considered. A probabilistic method is also more resilient to change in circumstances since reality is represented in a better way and not only dependent on a few scenarios. As has been mentioned, the European Commission has recommended all members in the EU to start using probabilistic methods when working with power system modelling. Since the multiple-scenario based MC simulation presented in this report is basically an extension of the currently used deterministic method, it is a good place to start. This technique could also be used in combination with for example machine learning predictions of the loads and generation in order to get even more accurate results of the actual branch loads for different future scenarios.

6.1.1 Historical Data

The historical data has been retrieved from measurements of the load flow through transformers connecting 10/20 kV buses and 50 kV stations. When working with real measurement data, it is important to be aware of possible measurement errors. The time series of the data were observed and obvious outliers were removed and replaced with interpolated data. Some deviating values were still left in the data. A decision was made not to be too strict in removing outliers, in order not to miss any extreme events. When examining the risk of thermal overload, it is typically these extreme events that are most important.

The data was collected over a time span of 6 years. This made it possible to collect variations over several years, something that conceivably reduced the impact of possible measurement errors and captured a wide range of measured extreme events. On the other hand, during the 6 years when the data was collected it is possible that some conditions have changed. This might be both in terms of consumer behaviour and in the generation levels. This has not been taken into consideration in this report. The historical data has simply been used as a reference, as if all historical data presented in one histogram is representative of the past 6 years in the investigated study area. Overall the historical data could however be assumed as representative of the study area on the level required for this analysis.

6.1.2 Comparison with Validation Data

When comparing the results from the reference case with the validation data in the five branches were historical data was available, it can be stated that for both Trial 1 and Trial 2 the estimated data matches the validation data quite well. The same can be said about the generated input data for both of the trials. When observing the histograms of the loads and wind power generation the estimated data matches the historical data in the general shape, extreme values and variances. The estimation of the CHP production however, does not show the same similarity. This is since it has been estimated as a normal distribution when the historical data clearly does not follow such a distribution. It might therefore be more accurate to model the CHP plant with a uniform distribution yet still correlate the production with the loads.

For both trials, the results in branch C-D shows a different shape than the validation data in the same branch. This difference continues throughout the study. The model has been tested through PSS/E in several different ways in order to exclude the risk of a programming error. Since no such source of error was found during any of the confirmation-tests, the issue was assumed to be located outside the investigated study area, perhaps in the overlying transmission network. This might suggest that the overlying network needs to be considered in the method, as well as the selected study area. One idea is that the power flows in the overlying network is correlated with the variables in the investigated power grid, this could have had an effect on the results. However, if the overlying transmission network makes such a large impact on branch C-D, the same level of error would be expected at the other branches in the meshed network as well. Another hypothesis for the source of error is that the measurements of the historical data in that particular branch could be corrupted or incorrect. As has been discussed, it is evident that the load data at station C has some irregularities. It might therefore be the failure of separating the load and generation input data in station C which has caused the deviation in branch C-D, or some other measurement error. Since the branch C-D has some obvious error which has not been located, this branch has not been paid as much attention when examining which use of input values to be chosen, nor in the following case studies.

The method has examined changes in 10 out of the 4 588 nodes in the network model. The rest of the parameters in the model were kept constant as defined by a pre-defined case. Initially the worst-case scenario provided by E.ON was used, but this gave branch power flows which were higher than the validation data. The case was thus scaled down to better represent a more average situation in the rest of the network model. This indicates that the behaviour in the overlying network model has some impact also on the results in the selected study area. This might be a reason to why the results in station C have a different shape than the validation data. A correlation in this station with some of the overlying network stations may need to be taken into account, instead of assumed as constant parameters.

6.1.3 Number of Iterations

When working with MC simulations, the number of iterations used is of importance in order to ensure enough accuracy but not use more computational power than necessary. As described in the theory section there are several different ways to do this, one standard approach is to investigate the variance of the solution, another way is to observe when the solution converges. Even though these tests can give a quantitative result, the issue of knowing which limit to set for the variance or when the solution is converging sufficiently remains. Ultimately, the human factor is difficult to avoid and some kind of judgment on what is "good enough" needs to be made.

Factors that need to be taken into consideration is how accurate the MC simulation should be in relation to the rest of the model uncertainties. In the presented MC simulation method for example, the modeled input variables are not identical with the historical data and the output will never be identical with the validation data. This indicates that the number of iterations does not need to be as high as if the general model uncertainties were smaller. Other limitations of accuracy of the result is the time required for the calculations. How much time that is worth spending on calculations in relation to the accuracy of the result needs to be investigated for every experiment using MC methods.

The choice of comparing convergence of the results that was made for this thesis (instead of investigating the variance) was made due to difficulties in finding a limit for the variance that would result in sufficiently accurate results. Furthermore, the use of convergence made it easier to compare and visualize the resulting number of iterations that was chosen. When iterating with the DAE solver PSSE/E, a lot of computational power was required in each iteration. As mentioned, the network model contained in total 4588 nodes, which resulted in severely increased time required for higher number of iterations. Due to the model uncertainties and the convergence of the results presented in figure 16, the number of iterations was set to 5 000. This was considered to be the most optimal trade off between convergence and time requirement for the MC method.

6.1.4 Input Values

When comparing the results from Trial 1 and Trial 2 a significant improvement in standard deviation (when comparing to validation data) can be observed for Trial 2. This is most likely due to the fact that the the correlation between the stations are taken into consideration. When picking correlated random values, the risk of accidentally picking a very warm and calm day in one station at the same time as a cold and windy day in another station is severely reduced. This results in a more correct variance in each random scenario and hence a more correct standard deviation. Since the computational effort is not higher in Trial 2 than in Trial 1, and the results seem to be more accurate in the form of standard deviations closer to the historical ones, Trial 2 was chosen as the method for generating input values for the PLF method.

It can be noted that there is some correlation between the wind power generation and the loads. In figure 13 correlations up 29 % can be observed. These correlations have not been taken into consideration. It would be interesting to

see if this correlation generated another result than that of Trial 2. However, when comparing Trial 1 with Trial 2 an improvement is made but the magnitude of the improvement is small. And if the improvement is due to correlations being considered, the result might improve but a risk is that the improvement would be of insignificant size. If correlations up to over 90 % results in a small improvement of the accuracy of the method, the plausibility of a correlation of 20 % leading to significant progress must be considered as low.

6.2 Application on Selected Cases

This report has used a case study of three cases with the aim to illustrate different uses of the suggested PLF method. In a real-life application of the suggested method for making load flow calculations, an analysis of how to interpret the results would have to be made. For example, questions like if there is a set level of risk which can be acceptable, or how much margin would be needed to the risk of thermal overload have to be considered. In actual applications, ratings are set by the grid owners, who also define their own methods when examining the risk of thermal overload. This type of questions have not been considered in this report, which has instead taken a mathematical approach.

6.2.1 Case 1: Manual Addition of Wind Power Generation

In the first case, the method was validated by making a manual addition of the same amount of wind power generation at station D as had actually been installed. The correlations between the wind power buses was calculated from historical data. As the data is not from the same time periods however, the correlations could not be calculated in this way. Instead the correlation between the wind power at bus D had to be approximated. This was made by picking a correlation value which was within the span of the calculated correlations between the historical data of wind power generation. These are shown in table 2. The choice of correlations of 0.5 with the other buses and 0.7 with bus D was an approximation and a possible source of error. The calculated average power factor of 0.225 could also be a source of error, but is however in the span of typical power factors of wind parks in Sweden.

When looking at mean values and standard deviations of the results, it can be seen that the addition of the wind power generation does bring the power flows closer to the validation data. However, validation data of the branch power flow was not available as far back in time as the installation was made for all the branches. This means that the validation data was the same as previously used, from the time period 2015-2020. It is possible that other changes have been made in these years which could have affected the power flows. However, as discussed earlier, the six-year data collection period has been concluded to represent typical years with some variation. The results therefore show that the method of adding new installations in the study area seems to work well, even though the results are hard to interpret due to the very small changes in the branch flow distributions when single installations are made. A similar method for generating input values could be used for stations where no historical data is available.

One exception to the accurate results in the branch power flows is the branch

C-D. The simulated mean value is further from the historical mean value after the new installation. However, looking at the results in figure 32, this branch power flow does not match the historical data in any of the two simulations. Furthermore, the standard deviations are shown to differ significantly in both simulations. This also corresponds with the earlier results of the MC simulation method, where the branch C-D has been identified to have some error. It could therefore be argued that the fact that the addition of the new generation brought the mean value further from the historical mean value is of low importance. This is due to the probability of there being some other issues with the modelling of this branch which are not related to the manual addition of wind power generation.

6.2.2 Case 2: Possible Future Application

The second case aimed at illustrating a possible future scenario and examine how well the suggested method could be applied. A possible installation onto the power grid would be a data centre. DCs typically have large loads and therefore need evaluation of the risk of thermal overload. This case was not something which could be compared and validated by looking at historical data. Following the study of Case 1 however, the method for adding a new load onto the grid in the study area seemed to work well. The chosen DC was of the flat load type, meaning that the correlation with temperature could be neglected. If other types of DCs are to be considered, the newly added load would have to be correlated with the loads at the other buses. This could be done in a similar manner as in Case 1.

The addition of a large load at station H has clearly affected the branch power flows. Especially the branches to the right of station H in the SLD are affected. Branches A-B and B-C do not seem to be very affected. This could be explained by them not being in the meshed part of the network, as can be seen in figure 6. The plots clearly show how the DC can be expected to affect the distributions of the power flows. This method gives more information about how an installation affects the power flows than the traditional deterministic method. Instead of a few scenarios and their resulting power flows, a more general picture of the distribution of the branch power flows is produced. Possible measures which can be compared to rates set by the grid owner would be the outer quantiles or maximal power flows. Exactly how such an analysis would be made by the grid owner or planner is out of scope for this report. This case shows however that the suggested method works well as a multiple-scenario based method for examining the risk of thermal overload.

6.2.3 Case 3: Application as a Risk-based Method

The third case was an attempt to apply the method as a risk-based method. Using line ratings as a basis, the amount of installed wind power without exceeding a set limit of risk was going to be optimized. The new installation was made larger until the risk level was exceeded. To illustrate how this could be done, the number of hours in a year in which a set rating was allowed to be exceeded was decided. This rating was set at a level which was sometimes exceeded in the case, in order to illustrate the method. However, such a strict line rating would in reality be applied under circumstances of high temperature, resulting

in high conductor sag in the power lines. This does not typically correspond to the times when the capacity in the lines is high, as this often occurs at low temperatures when the loads are high.

The suggested method gives no information of when the different branch power flows occur. This means that there is no information on which fixed power line rating should be applied. Large loads generally occur in winter time, but the strict line ratings are typically based on warm temperatures. The suggested method can therefore not be used directly as a risk-based model in combination with fixed line ratings. It is also true that fixed line ratings in themselves do not capture enough information as the accepted capacity in a branch is of a dynamic nature, depending on temperature and other factors.

Consequently, in order to be able to use the suggested probabilistic method with a risk-based modelling approach it needs to be improved. Dynamic line ratings and information of when the different power flows are expected to occur needs to be included in the method.

6.2.4 Possible Application in Industry

One of the aims of this report has been to investigate the possible use of MC simulations in PLF calculations. This has been investigated by the use of the above discussed cases. It has been made clear that the use of a multiple-scenario based probabilistic method will give more information about how new installations onto the grid affect the branch power flows than the deterministic single-scenario method. The resulting distributions show the probabilities of different branch power flows occurring in the selected study area.

For an application in industry however, the probabilistic model would need to include the whole model of the power grid. In this study, everything outside of the selected study area has been kept constant, and can therefore be a source of error. It is also of importance to know how the resulting branch power flows are affected by changes in the overlying network. This study therefore serves as a proof of concept but needs to be applied on a larger scale if used in industrial applications by the grid owners.

6.3 Future Work

This report has presented a proof of concept of how a probabilistic method for making load flow calculations and examining the risk of thermal overload could be designed. Further studies on this topic have been identified in line with the limitations discussed earlier in the report.

In order to work better with defined line ratings, a time series based modelling could be used for generating the input variables. For example, a simulation of the temperatures in a standard year could be used as input values. This would generate results which can be compared to dynamic line ratings based on temperature. Another option would be to include time or temperature in another way in the method, such as for example studying different seasons separately. This would allow the comparison with set line ratings in a more applicable manner. A risk-based modelling approach might then be possible. Today, risk-based methods are generally not used when examining the risk of thermal overload.

However, it has been mentioned in various sources as a potential tool in the future and could therefore be examined further.

This study has used aggregated loads as input data. This has worked well, but an alternative approach could be taken. The examination of the different aggregated loads show that they consist of a mixture of load types. A more accurate modelling approach would be to model the different underlying loads individually.

For an industrial application, the method would need to include a larger part of the network model. This would most likely increase the computational time needed. A study of how the computational time grows with larger study areas would therefore be of interest when considering future applications of the method. As this report shows, the results are already quite accurate when compared with historical data. It would thus be of interest to investigate how large the study area needs to be in order to examine how a new installation will affect the branch power flows. An investigation of the computing power and time needed in an industrial application would also be of use.

When comparing the results in this study, measurements of the mean value and the standard deviation of the distributions have been used. This is standard when working with probabilistic methods. However, in the application of this method when examining the risk of thermal overload, the extreme values might be of higher importance than the mean value. This is also the reason why the current deterministic methods have worked quite well until now. Different measures of the resulting distributions could therefore be developed and examined.

7 Conclusions

The aim of this thesis has been to illustrate the use of a probabilistic method when making load flow calculations in order to examine the risk of thermal overload. A Monte Carlo simulation multiple-scenario method has been selected. The study has been based on the following research questions:

- How can a Monte Carlo simulation be applied in load flow calculations to study the risk of thermal overload?
- What input data is needed for the Monte Carlo simulation?
- What kind of results can be expected from the Monte Carlo simulation?
- How useful is the method in industrial applications?

The MC simulation has been defined and tested on a reference case, which is a current real-life sub-transmission system located in Sweden. The results which have been compared with validation data show that the method gives accurate results with the exception for one of the branches. Some error in the collected data or in the network model may be the reason for this.

The generation of input values have been based on historical data, and two different methods have been compared. It was concluded that the use of statistical distributions which are correlated in line with historical correlations gave more accurate results. The input data may be improved by considering the aggregated loads and generation separately, or by including correlations between stations with low correlation as well. However, the comparison of the results to the validation data suggest that the input data is sufficient to generate accurate output data.

The method has been applied on three cases in order to show what type of results can be generated by the method. The results from the case with manually added load shows optimistic results, however in need of further investigation. The method generates results with more information than the deterministic single scenario method which is typically used in industry today. When adding a new installation, the whole expected range of probable branch power flows can be studied instead of a few cases. Lastly, the method would need to be improved if a risk-based modelling approach is to be taken. The method in combination with set line ratings do not give enough information to be applied in a reasonable manner.

Finally, this thesis has illustrated how a probabilistic method could be applied to power flow calculations. The method needs to be developed if applied in industry, but this thesis serves as a proof of concept for such an application.

References

[1] Statens Energimyndighet. Energiläget 2020. Statens Energimyndighet; 2020. (In Swedish).

[2] Statens Energimyndighet. Vägen till ett 100 procent förnybart elsystem. Statens Energimyndighet; 2018. (In Swedish).

[3] Sveriges Riksdag. Kapacitetsbrist i elnäten. Näringsutskottets betänkande (2020/21:NU5). Stockholm: Sveriges Riksdag; 2020. (In Swedish).

[4] CIGRE. TB 820 - Optimal power system planning under growing uncertainty. CIGRE; 2020. (Technical broschure)

[5] European Commission. Identification of Appropriate Generation and System Adequacy Standards for the Internal Electricity Market. Brussels: European Commission; 2015.

[6] Chen, P., Chen Z., Bak-Jensen, B. Probabilistic Load Flow: A Review. IEEE; DRPT2008: 1586-1591.

[7] Prusty, B., Jena, D. A critical review on probabilistic load flow studies in uncertainty constrained power systems with photovoltaic generation and a new approach. Renewable and Sustainable Energy Reviews. 2017; 69: 1286-1302.21

[8] Hanif Ramadhania, U., Shepero, M., Munkhammar, J., Widén, J., Etherden, N. Review of probabilistic load flow approaches for power distribution systems with photovoltaic generation and electric vehicle charging. International Journal of Electrical Power and Energy Systems,. 2020; 120.

[9] Svenska Kraftnät. Sveriges Elnät [Internet]. Svenska Kraftnät; 2021. (In Swedish). [Cited 2021-05-11]. Available from: https://www.svk.se/om-kraftsystemet/oversikt-av-kraftsystemet/sveriges-elnat/

[10] Svenska Kraftnät. Roller i kraftsystemet [Internet]. Svenska Kraftnät; 2021. (In Swedish). [Cited 2021-05-11]. Available from: https://www.svk.se/om-kraftsystemet/oversikt-av-kraftsystemet/roller-i-kraftsystemet/

[11] Svenska Kraftnät. Anslut till transmissionsnätet [Internet]. Svenska Kraftnät; 2021. (In Swedish). [Cited 2021-05-11]. Available from: https://www.svk.se/aktorsportalen/systemdrift-elmarknad/anslut-till-transmissionsnatet/

[12] Svenska Kraftnät, Elnät i fysisk planering [Internet]. Svenska Kraftnät; 2014. (In Swedish). [Cited 2021-05-11]. Available from: https://www.svk.se/siteassets/om-oss/rapporter/elnat-i-fysisk-planering-webb.pdf

[13] Glover, D., Sarma, M., Overbye, T. Power System Analysis and Design. 5 ed. Stamford: Cengage Learning; 2012.

[14] Elsäkerhetsverket. ELSÄK-FS 2008:1. Elsäkerhetsverket; 2021. (In Swedish).

[15] Lindberg, E. The Overhead Line Sag Dependence on Weather Parameters and Line Current [Degree Project]. Uppsala: Uppsala University; 2011.

[16] Reinders, J., Morren, J., Slootweg, H. Comparing Probabilistic Load Flow Methods in Dealing with Uncertainties at TSO/DSO Interface. 52nd International Universities Power Engineering Conference (UPEC); 2017.

[17] Shikhar, Y., Singh, V., Srivastava, R. Comparative Analysis of Distance Metrics for Designing an Effective Content-based Image Retrieval System Using Colour and Texture Features. International Journal of Image, Graphics and Signal Processing. 2017; 9: 58-65.

[18] Villanueva, D., Pazos, J., Feijoo, A. Probabilistic Load Flow Including Wind Power Generation. IEEE Transactions on Power Systems. 2011; 26(3): 1659-1667.

[19] Wang, XF., Song, Y., Irving, M. Modern Power Systems Analysis. New York: Springer Publishing; 2008.

[20] Karami, B. Correlated Variables in Monte Carlo Simulations [Internet]. Towards Data Science; 2021. [Cited 2021-05-11]. Available from: https://towardsdatascience.com/correlated-variables-in-monte-carlo-simulations-19266fb1cf29

[21] Haugh, M. An Introduction to Copulas [Internet]. Columia University; 2016. [Cited 2021-05-11]. Available from: http://www.columbia.edu/ mh2078/QRM/Copulas.pdf

[22] Baxtel Datacenter Resource [Internet]. Baxtel; 2015 - [Cited 2021-05-11]. Available from: https://baxtel.com/map

[23] Ghatikar, G., Piette, M.A., Fujita, S., McKane, A.,Han, J., Radspieler, A., Mares, K., Shroyer, D. Demand Response and Open Automated Demand Response Opportunities for Data Centers. California Energy Commission, PIER Program and Pacific Gas and Electric Company (PGE); 2010.

[24] Statens Energimyndighet. Vindkraftsstatistik 2017. Statens Energimyndighet; 2018.