



*Degree project in Electric Power Systems*

Second cycle 30 credits

**KTH ROYAL INSTITUTE  
OF TECHNOLOGY**

# **Towards a Stochastic Operation of Switzerland's Power Grid**

Master thesis

**ALBAN MAURY**

---

*This thesis is my final work for the master's program in Sustainable Energy Engineering at KTH, Stockholm. It was performed in collaboration with KTH Royal Institute of Technology in Stockholm and Swissgrid, Switzerland's Transmission System Operator.*

*I would like to thank the people at Swissgrid who supported me during this work, especially Matthias Bucher for broadening my perspectives and providing relevant insights, and Marc Hohmann for the technical support and the relevant ideas. I would also like to thank my supervisor and examiner Mohammad Reza Hesamzadeh for his availability and his support throughout the thesis. His inputs were always relevant and allowed me to have more academic insight into the thesis.*

*I would also like to thank my team and more generally all the people at Swissgrid for their kindness and readiness to share their knowledge and talk about their ongoing projects. I have had a great time at Swissgrid and will keep a good memory of my time here. Thank you all.*

---

## Abstract

As Europe's power production becomes increasingly reliant on intermittent renewable energy sources, uncertainties are likely to arise in power generation plans. Similarly, with the growing prevalence of electric vehicles, electric demand is also becoming more uncertain. These uncertainties in both production and demand can lead to challenges for European power systems. This thesis proposes the use of Monte-Carlo simulations to translate uncertainties in power generation and demand into uncertainties in the power grid.

To integrate stochasticity in the forecasts, this thesis separates the multivariate probabilistic forecasting problem by first forecasting the marginal loads individually and probabilistically. Copula theory is then used to integrate spatial correlations and create realistic scenarios. These scenarios serve as inputs for Monte-Carlo simulations to estimate uncertainties in the power system. The methodology is tested using power injection data and the power system model of Switzerland.

The results demonstrate that integrating stochasticity in forecasts improves the reliability of the power system. The proposed approach effectively models the uncertainty in both production and demand and provides valuable information for decision-making.

**Keywords:** *Monte-Carlo simulations, probabilistic forecasting, copula theory, power flow analysis.*

## Sammanfattning

I takt med att Europas elproduktion blir alltmer beroende av intermittenta förnybara energikällor kommer det sannolikt att uppstå osäkerheter i planerna för elproduktion. På samma sätt blir efterfrågan på elektricitet mer osäker i takt med att elfordon blir allt vanligare. Dessa osäkerheter i både produktion och efterfrågan kan leda till utmaningar för de europeiska kraftsystemen. I denna avhandling föreslås att Monte-Carlo-simuleringar används för att omvandla osäkerheter i elproduktion och efterfrågan till osäkerheter i elnätet.

För att integrera stokasticitet i prognoserna separerar denna avhandling det multivariata probabilistiska prognosproblemet genom att först individuellt och probabilistiskt prognostisera belastningar. Kopulateori används sedan för att integrera rumsliga korrelationer och skapa realistiska scenarier. Dessa scenarier tjänar som indata för Monte-Carlo-simuleringar för att uppskatta osäkerheterna i kraftsystemet. Metodiken testas med hjälp av data om inmatning av el och med hjälp av Schweiz kraftsystem.

Resultaten visar att integrering av stokasticitet i prognoser förbättrar kraftsystemets tillförlitlighet. Den föreslagna metoden modellerar effektivt osäkerheten i både produktion och efterfrågan och ger värdefull information för beslutsfattandet.

**Nyckelord :** *Monte Carlo-simuleringar, probabilistiska prognoser, kopulateori, analys av energiflöden..*

# Contents

<b>Nomenclature</b>	<b>8</b>
<b>1 Introduction</b>	<b>10</b>
1.1 Research objective and question . . . . .	10
1.2 Scope and limitations . . . . .	10
1.3 Sustainability aspects . . . . .	11
1.4 Thesis structure and tools . . . . .	11
<b>2 Related work</b>	<b>13</b>
2.1 Probabilistic power flows methods . . . . .	13
2.2 Probabilistic load forecasting . . . . .	14
<b>3 Methodology</b>	<b>16</b>
3.1 Generating scenarios for probabilistic power flows . . . . .	17
3.2 Non-parametric probabilistic forecasting of the loads . . . . .	18
3.2.1 Temporal analysis . . . . .	20
3.2.2 Comparison of forecasting models . . . . .	22
3.2.3 Key Performance Indicators definition . . . . .	25
3.3 Sampling random variables for Monte-Carlo simulations . . . . .	26
3.3.1 Correlation analysis . . . . .	26
3.3.2 Copula theory . . . . .	28
3.3.3 Copula methodology applied to scenario generation . . . . .	28
3.4 Power flows calculations . . . . .	31
3.4.1 Applied topology . . . . .	31
3.4.2 AC power flows . . . . .	31
3.4.3 Contingency analysis . . . . .	32
<b>4 Results and discussion</b>	<b>33</b>
4.1 Non-parametric forecasting of the marginal loads . . . . .	33
4.1.1 Comparison of the different forecasting models . . . . .	33
4.1.2 Performance analysis . . . . .	34
4.1.3 Example of forecasts . . . . .	35
4.2 Integrating spatial correlations using copula theory . . . . .	36
4.2.1 Choice of the copula family . . . . .	37
4.2.2 Scenarios for PPF . . . . .	38
4.3 PPF: an example . . . . .	40
4.3.1 Power flows . . . . .	40
4.3.2 Contingency analysis . . . . .	41
<b>5 Conclusion</b>	<b>43</b>
<b>References</b>	<b>44</b>

## List of Figures

1	Monte-Carlo simulations for probabilistic power flows. . . . .	16
2	Density estimation for a specific node . . . . .	18
3	Day-Ahead load forecasting . . . . .	19
4	Autocorrelation to past values and Fourier transformation of the total load. . .	21
5	Machine learning models . . . . .	23
6	Architecture of a neural network . . . . .	24
7	Correlation heatmaps . . . . .	27
8	Generating random correlated vectors for MCS using copula theory . . . . .	30
9	Mapping random correlated vectors . . . . .	30
10	Performance of the forecasting model . . . . .	34
11	Example of non-parametric forecasts for a sample day . . . . .	36
12	Example of scenarios for a sample day . . . . .	38
13	Correlation graphs for different nodes . . . . .	39
14	Power flows examples . . . . .	41
15	N-1 violations examples . . . . .	42

## List of Tables

1	PICP and PINAW at 90% for the three tested models, in %. . . . .	34
2	RMSE comparison between different copula families . . . . .	37

## Abbreviations

<b>AC</b>	Alternative current
<b>ANN</b>	Artificial Neural Network
<b>ARIMA</b>	Autoregressive integrated moving average
<b>CNN</b>	Convolutional Neural Network
<b>CDF</b>	Cumulative Density Function
<b>DACF</b>	Day-Ahead Congestion Forecast
<b>DC</b>	Direct current
<b>DFT</b>	Discrete Fourier transform
<b>GAN</b>	Generative Adversarial Network
<b>IDCF</b>	Intraday Congestion Forecast
<b>KPI</b>	Key Performance Indicator
<b>LSTM</b>	Long Short-Term Memory
<b>MCS</b>	Monte-Carlo Simulation
<b>PDF</b>	Probability density function
<b>PLF</b>	Probabilistic Load Forecasting
<b>PPF</b>	Probabilistic power flow
<b>PSS/E</b>	Power System Simulator for Engineering
<b>RNN</b>	Recurrent Neural Network
<b>RTSN</b>	Real-Time Snapshot
<b>STLF</b>	Short-term load forecasting
<b>TSO</b>	Transmission System Operator
<b>VAE</b>	Variational Autoencoder

# Nomenclature

$(\mathbf{u}^i)_{1 \leq i \leq k}$	Random correlated vectors from $[0, 1]^d$
$C$	Copula function
$F_{X_i}$	Cumulative density of $X_i$
$R$	Correlation matrix
$U_i$	Random variable following a uniform distribution on $[0, 1]$
$X_i$	Random variable
$PICP$	Prediction Interval Coverage Probability
$PINAW$	Prediction Interval Normalized Average Width
$RMSE$	Root Mean Squared Error
$\sigma$	Sigmoid function
$b_i, b_f, b_o$	Respective bias matrices
$C_t$	Cell value
$h_t$	Hidden state for $t$
$i_t, f_t, o_t$	Input, forget and output gate
$\tanh$	Hyperbolic tangent function
$W_i, W_f, W_o$	Respective weight matrices
$x_t$	Input vector
$d = NT$	Size of the vector to forecast
$N$	Number of nodes in Switzerland and its belt
$T$	Time horizon used for forecasting, here always equal to 24 hours
$\mathbf{F} = (X_{F_1}, \dots, X_{F_m})$	Categorical temporal features
$\mathbf{X} = (X_1, X_2, \dots, X_p)$	Input vector for the forecasting model

$\mathbf{Y} = (Y_1, Y_2, \dots, Y_d)$	Output vector to forecast
$L_q(\mathbf{Y}, \hat{\mathbf{Y}})$	Mean Pinball Loss for quantile $q$
MW	Megawatt
MWh	Megawatt hour

# 1. Introduction

The increasing share of intermittent renewables and demand flexibility in the European power grid has led to new challenges. Power production and demand are becoming more uncertain, making them harder to predict. These uncertainties in the production and demand raise new challenges for power systems and more specifically for Transmission System Operators (TSOs) as they translate to uncertainties in the power systems, making this operation and planning of the grid more complex. For this reason, traditional point models are being challenged, and new methods are necessary to address these uncertainties. This work aims at translating the uncertainties in power production and demand into uncertainties in the power systems to increase the reliability and security of their operations.

## 1.1. Research objective and question

This thesis aims to explore the potential methods for incorporating uncertainty estimation in Day-Ahead congestion forecasts for the high-voltage power grid in Switzerland. Currently, the forecast models are point forecasts, providing net injected power estimates for each node in the grid and the hour of the following day, which are then used to estimate the power flows for each hour. However, this approach offers limited insight into potential uncertainties in power injections and their impacts on the grid's reliability.

Therefore, this thesis proposes an efficient method to estimate the Day-Ahead load distribution for each grid element. To do so, Monte-Carlo simulations are used: the main idea is to generate many possible injection scenarios for each hour to determine power flows, resulting in an estimation of the load distribution for each hour and each grid element. In addition, to ensure compliance with the N-1 security criterion, which requires the system to remain secure in the event of any single element's failure, the impact of N-1 contingencies is also studied.

The central research question of this thesis is to investigate the creation of realistic scenarios for probabilistic power flows. This approach aims to improve the reliability of the grid and provide better insight to operators in the control room by offering more comprehensive and accurate estimations of potential load distributions, eventually leading to increased security and reliability of the power grid.

## 1.2. Scope and limitations

The main focus of this thesis is to provide a day-ahead probabilistic power flow forecast for the high-voltage grid in Switzerland. Therefore, this thesis focuses only on Day-Ahead forecasts, with one calculation per hour of the following day. Swissgrid also creates throughout the day an Intraday Congestion Forecasts (IDCF) that provides the operators useful insight closer to real-time data, based on newer data, but shorter periods are here not considered. Integrating stochasticity over longer periods can also provide fruitful information, for example, to study the impact on the grid of the development of renewables, but this is not considered here.

This thesis focuses only on the high-voltage grid (220 and 380 kV) within Switzerland and not the whole of Europe as Swissgrid is only responsible for the operation of this part of the European grid. However, Switzerland is interconnected to France, Germany, Italy, and

Austria, so the power flows are calculated by considering certain nodes around Switzerland to provide results that are closer to reality. Congestions and N-1 contingencies abroad are not investigated here as Swissgrid is not responsible for them.

Remedial actions, i.e., possible actions such as topological changes or redispatch that can be enforced by the operator to avoid congestions and N-1 contingencies, are not investigated here. A probabilistic remedial action optimizer that studies the impact of possible remedial actions probabilistically, could represent a significant step forward to helping the operator in decision-making but will be considered here out of scope.

### 1.3. Sustainability aspects

With the growing part of intermittent renewable electricity generation such as wind and solar PV, uncertainties in the power production schedules have increased over the past years. Despite increased accuracy of the forecast of renewable power generation, some uncertainties can remain (Taşçıkaraoğlu 2017). The uncertainties on the demand side are also expected to grow in the coming years with the democratization of electric vehicles (D’Ettorre et al. 2022).

Both injection and demand uncertainties result can result in power flows variations in real-time compared to the forecasts, which might eventually lead to unforeseen congestion and contingencies. Probabilistic power flows (PPFs) can thus account for the uncertainties on the loads and thus on the power flows, leading to estimations of the probability of congestions and associated risks for the grid. In the future, with the growing part of intermittent renewables, the decreasing share of thermal power generation, and the increasing demand flexibility, PPF will become more and more useful for grid operation and decision-making.

PPFs might also contribute to the reduction of greenhouse gas (GHG) emissions. Indeed, PPFs could mitigate the curtailment of renewable energy production by integrating scenarios with higher renewable production than expected. Moreover, PPFs might also help reduce redispatch<sup>1</sup> at a European level. For instance, if the wind power production in the north of Germany must be curtailed and some thermal power plants in the southern parts of the country must increase their production as part of the redispatch in the country will result in higher GHG emissions. Therefore, PPF could help achieve Sustainable Development Goal 7 and help better integrate and optimize the use of renewable production (United Nations 2022).

### 1.4. Thesis structure and tools

This thesis is organized as such: section 2 provides an overview of the different PPF and probabilistic load forecasting methods, whereas section 3 presents the proposed methodology in detail. The results are discussed in section 4, and section 5 summarizes this work and proposes possible improvements. Moreover, for privacy reasons, all the results shown here are standardized and anonymized.

The main tool used here is Python and several Python libraries, such as:

- seaborn and matplotlib for plotting,
- pandas for data analysis,

---

<sup>1</sup>Redispatch corresponds to changes in the power production schedule of a power plant due to congestions on the power grid. For example, controllable production can be lowered on a specific node and increased elsewhere to mitigate possible congestions. This service is paid for by the TSO to the owner of the power plant.

- TensorFlow and Scikit-learn for deep learning and machine learning models,
- the PSS/E python API for calculating the power flows,
- statsmodels for copula theory.

## 2. Related work

### 2.1. Probabilistic power flows methods

With the increase of weather-dependent power production, of demand flexibility and traded volumes between the market zones, uncertainties in the power systems are increasing and playing a critical role in the operations and planning of electrical systems. Uncertainties in power systems come from the demand, the power generation, or the network itself. With the increasing share of electric vehicles, the demand, which is stochastic by nature, is expected to become more and more difficult to forecast, resulting in higher uncertainties. On the generation side, solar PV or wind power have uncertainties related to the difficulty of forecasting the weather accurately. On the network side, uncertainties can come from network element failures or temperature-dependent transmission line ratings. These factors result in higher unpredictability for the electrical grid, thus requiring changing the traditional point models to probabilistic ones to incorporate uncertainties in power flow forecasts to improve operations and planning of power systems (Tuinema et al. 2020).

Probabilistic power flows (PPF) have first been proposed in the literature in 1974 by considering the uncertainty in the node data (Borkowska 1974). There are three main methods for PPF: analytical methods, approximation methods and Monte-Carlo simulations (MCS) methods (Bin et al. 2018):

1. **Analytical methods:** Analytical methods consist in simplifying the power flow equations by linearizing them. The main analytical method is the cumulant method, which consists in representing the outputs' cumulants as a linear combination of the inputs' cumulants<sup>2</sup>. Different reconstruction techniques can then be applied to the cumulants to retrieve the probability density function (PDF), such as Laplace Transform or Gram-Charlier expansion (Carr, Mc Grath, and Ní Fhloinn 2020). These methods reduce the computation time considerably, but some errors can occur due to the linearization of the power flows equations (C. Wang et al. 2020).
2. **Approximation methods:** Approximation methods are effective methods to keep a reasonable computational burden whilst improving the accuracy of PPF. The most common method is the point-estimate method (PEM) which consists in approximating the moments of the outputs' random variables by calculating the power flows deterministically for a few representative points. The advantages of PEMs are the relatively low computational burden, and the possibility to use non-linear power flow equations, but conventional PEMs do not capture spatial correlations between nodes (Delgado and Domínguez-Navarro 2014).
3. **Monte-Carlo simulations:** MCS consist in computing the power flows for many scenarios, which allows thus to estimate the power flow distributions. Compared to analytical methods and PEMs, MCS have shown better accuracy but require many iterations and thus good computational capacity. Appropriate sampling techniques can also incorporate spatial correlations between nodes. (Hajian, Rosehart, and Zareipour 2013). However, MCS have a heavy computational burden, so scenario reduction can be used

---

<sup>2</sup>In probability theory, cumulants correspond to the coefficients of the power series expansion of the natural logarithm of the moment generating function.

to select only relevant scenarios and thus reduce the computational time. Certain sampling techniques such as Latin Hypercube Sampling (LHS) have been used to consider only non-overlapping scenarios, nonetheless, LHS does not incorporate possible correlations between loads and generation at different buses (Xu and Yan 2015a). More advanced sampling techniques have been proposed to tackle this issue by adapting the LHS, such as Latin Supercube Sampling with linear regression (Xu and Yan 2015b) or Latin Hypercube important Sampling (Q. Li, X. Wang, and Rong 2018).

For grid operations, the computational burden is not really a limiting factor, and the highest accuracy possible is required. For these reasons, MCS are usually preferred over other methods.

## 2.2. Probabilistic load forecasting

Several probabilistic load forecasting methods (PLF) have been proposed in the literature. PLF can be either parametric or non-parametric. The parametric approach consists in assuming the shape of the distribution (e.g., Gaussian) and forecasting its conditional parameters (e.g., mean vector and covariance matrix) using for example the maximum likelihood estimation (Anderson and Olkin 1985). The most common non-parametric approach consists in forecasting the different conditional quantile levels by defining a quantile loss. The main advantage of the non-parametric approach is that it requires fewer assumptions on the inputs but comes at a higher computational cost (Steinwart and Christmann 2011). More recently, generative models that are traditionally applied to image or text generation have been applied to generate loads scenarios, for example, using Generative Adversarial Networks (GANs) or Variational Auto-Encoders (VAEs). Such models have proven their overall good performance. However, GANs are difficult to properly condition and stabilize, whereas VAEs show too wide distributions (Liao et al. 2022).

Some works have also been done to compare the parametric and the non-parametric approaches. Both can show similar performance, but the former requires assuming the distribution shape. On the contrary, the latter does not require any assumption on the distribution shape but does not allow directly integrate correlations between the features (Gürses-Tran, Flamme, and Monti 2020).

Assuming the distribution shape depends on the injection type. For loads, the Gaussian assumption is usually reasonable. However, for generation, the assumption depends on the power production technology. For example, for wind power, the most common assumption is to apply a Weibull distribution to the wind speed, from which the power production with its uncertainties can be derived. The uncertainties in solar PV production can also be derived from the uncertainties in the solar irradiance that can be modeled via a Beta distribution (Ebeed and Aleem 2021).

To face this issue, Konstantinou, Savvopoulos, and Hatziargyriou 2021 use Gaussian Mixtures Models (GMMs) to be able to model more complex distribution shapes. Autoencoders are used to project the demand profiles in a latent space to eliminate redundant information. In this space, a GMM is fitted. The scenarios are then sampled from the latent space. Probabilistic metrics are defined to account for the sharpness and accuracy of the predictions. However, the use of the latent space does not allow a clear understanding of the model.

Making realistic assumptions on the distribution shape requires case-specific assumptions.

Toubeau et al. 2019 shows that the non-parametric approach shows better overall performance for probabilistic forecasting of load, wind production, and Day-Ahead prices, whilst the performance is similar on PV production. In the second step, a copula-based sampling method is used to integrate spatiotemporal correlations and create scenarios for a probabilistic optimization of day-ahead electricity market bids, which shows a significant increase in the profit of the considered portfolio. Simard and Remillard 2013 also use copula theory to capture the spatial dependencies for a multivariate time series forecasting problem applied to financial data.

Classic machine learning techniques have been applied to load forecasting problems, such as linear regressions, support vector machine regressions, Random Forests or Auto-Regressive Integrated Moving Average models. Gradient Boosting Regressors (GBRs), which combine several weak decision trees into a stronger model by minimizing the residual loss at each step, has also shown great performance for time series forecasting (Nie et al. 2021). Such models have proven a relatively good performance for load forecasting but fail to capture correlations as they can only produce one output per input (Czapaj, Kamiński, and Sołtysik 2022). To model uncertainties in the forecasts, other approaches have been proposed, such as bootstrap methods. These methods consist of resampling the data to train the same model on different but similar datasets, which allows for multiple results and thus the estimation of a confidence interval of the prediction. (Xiao, M. Li, and Zhang 2022).

More recently, Artificial neural networks (ANNs) have been widely used for time series forecasting because they can extract nonlinear dependencies in the input data, such as temporal and spatial dependencies. Moreover, such models allow more flexibility, such as multidimensional data processing and multi-step ahead prediction (Lim and Zohren 2021). Simple neural network architectures, such as dense ANNs, have shown good performance in predicting time series. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) can achieve better performance but require a fine-tuning of their hyperparameters, which can be complicated (Qin 2019).

CNNs have historically been used for image datasets but have also shown good performance by extracting relationships that are invariant across spatial dimensions. RNNs have traditionally been used to generate sequences, especially in natural language processing, but its architecture is also suitable for time series forecasting as it can retain past information that is recursively updated at each time step. More specifically, Long-Short Term Memory (LSTM) neural networks have shown outstanding performance in a variety of domains, such as speech recognition and time-series forecasting, as they can process an entire sequence of data given their specific architecture (Lindemann et al. 2021). Gated Recurrent Units (GRUs) are another type of RNN that has shown good performance in time-series forecasting, especially with financial data (Shen et al. 2018).

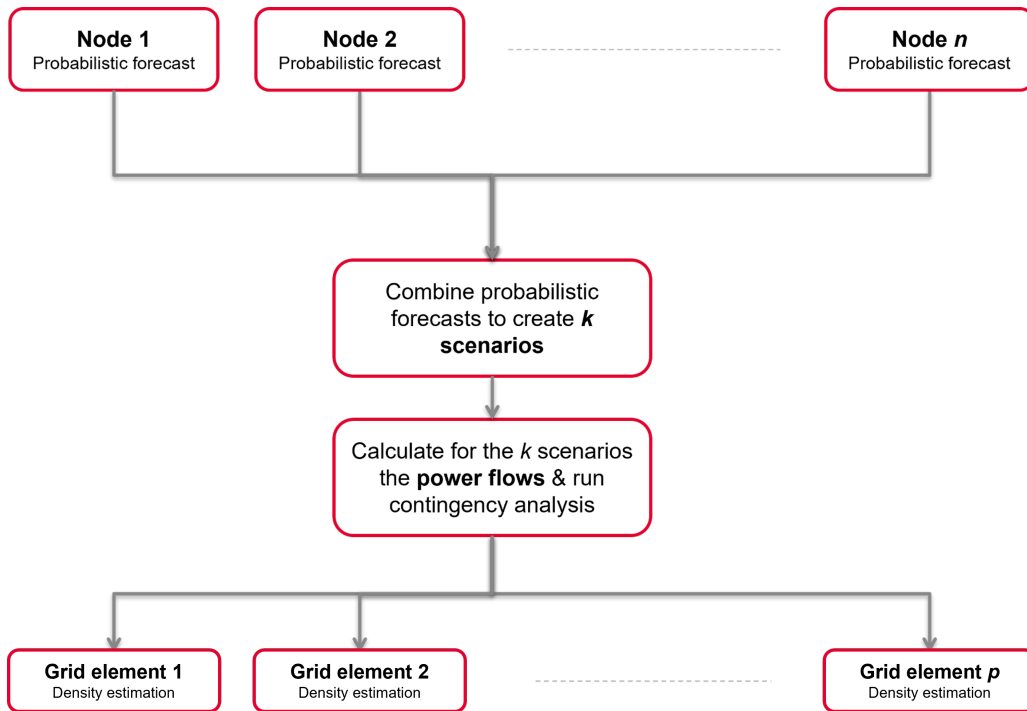
Overall, there are several methods to integrate stochasticity in load forecasting. The following section justifies the chosen methodology.

### 3. Methodology

Because analytical and approximation methods require too many simplifications (see Section 2), it was decided that both techniques are not suitable for the intended study and a numerical approach based on Monte-Carlo simulations is thus preferred. This approach consists in feeding to a deterministic load flow model many inputs with different combinations of nodal power values. The non-linear power flow equations can thus be calculated with no approximations, resulting in high accuracy. The main drawback is the computational burden, which is why it was decided to select only a few relevant scenarios. Monte-Carlo simulation was therefore decided to be the most appropriate method for Day-Ahead PPF.

Figure 1 presents the methodology used for Monte-Carlo simulations for PPF. First, a probabilistic forecast is made for the load at each node. The term "load" refers to the net load, which can be either positive or negative as it includes both production and demand in the lower grid levels. By convention, if the net load is positive, the node is a consumption node and is retrieving power from the system, and vice-versa.

These probabilistic forecasts are then combined to create  $k$  realistic scenarios for which the power flows are then computed. This leads to an estimation of the density of each of the grid elements  $p$ .



**Figure 1:** Monte-Carlo simulations for probabilistic power flows.

This section presents the detailed methodology that is used for creating these scenarios.

#### The historical loads

For the thesis, as the focus is put on day-ahead forecasts, the Day-Ahead congestion forecast data (DACF) and Real-Time Snapshots (RTSN) are retrieved. The RTSN data contains the real-time information about injection power (active and reactive power) at each node for

each hour for each day for the nodes within Switzerland and its close belt<sup>3</sup>. The topology of the grid is also stored within the data, meaning that outages or scheduled maintenance are integrated within the data. On the other hand, the DACF data contains the same information but is based on Swissgrid's forecast for the following day. As of today, the DACF is constructed by retrieving non-binding power production schedules from the power plants. The loads are estimated by taking the load from the same day same hour a week ago, and the loads are then scaled to match Switzerland's net position, defined as the difference between exports and imports.

From this historical data, the net load for each node is retrieved and stored in a .csv file. The net load corresponds to the net sum of load and generation in the lower grid levels. By convention, the load is positive and the generation negative, so the net load can be either negative or positive, depending on if the node retrieves power from the grid or injects power to it.

The rows correspond to the timestamps and the columns to the nodes. As retrieving data is quite a fastidious process, the loads have only been retrieved for the past four last years, from 2019 to 2022 included. For the sake of this thesis, this amount of data is reasonable and sufficient, but retrieving more historical data could further improve the performance of the forecasting models.

### 3.1. Generating scenarios for probabilistic power flows

Creating these scenarios is a difficult task because of the large dimension of the problem: there are over 200 nodes in the high-voltage grid in Switzerland and its belt that need to be forecasted probabilistically for 24 hours. Moreover, the loads are expected to be spatially correlated: indeed, for example, production and consumption are expected to be somewhat positively correlated since the production is higher when the demand is higher. Generation can be also to a certain extent correlated: PV facilities located close by are expected to have similar behavior. Multivariate probabilistic forecasting, therefore, needs to consider these correlations. To integrate them, the two main techniques are the following (see section 2):

1. **Parametric approach:** the distribution shape is assumed (e.g., Gaussian) and its parameters are forecasted (e.g., mean vector and correlation matrix). This allows to directly forecast the distribution considering the spatiotemporal correlations between the nodes. However, it requires assuming the distribution shape, which can be unreasonable.
2. **Non-parametric approach:** the non-parametric approach consists in forecasting the different quantile levels for each node separately and combining the quantile forecasts using copula theory to integrate the spatial correlations and create realistic scenarios. The advantage of this approach is that it does not require any assumption on the distribution shape, but the spatial dependencies must be modeled independently, for example using copula theory.

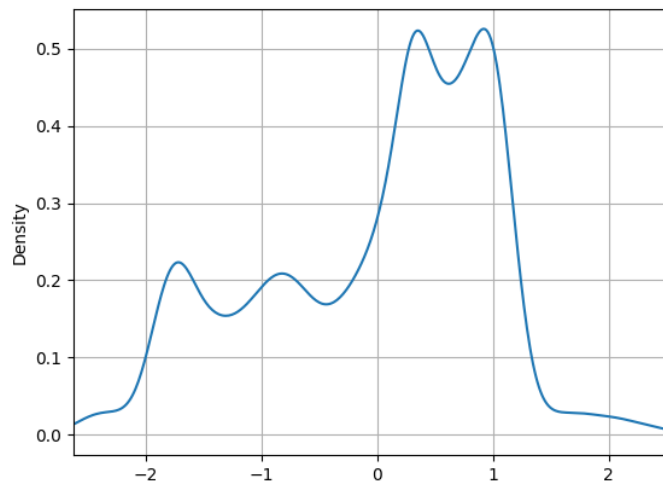
To choose between the parametric and the non-parametric approaches, the distribution shape of the loads is here investigated.

Switzerland's power production mostly comes from nuclear and hydropower plants. Switzer-

---

<sup>3</sup>The belt corresponds to all the nodes close to the Swiss border that have an influence on the Swiss power system.

land also has a total installed capacity of 2.5GW of pumped hydropower (BFE 2022). Such facilities are used as batteries: they can either turbine the water to generate electricity or pump it to store it in a lake at a higher altitude to produce electricity later. They have a strong bimodal behavior as their power production schedule is based on the market prices and cannot be assumed to be Gaussian. Other nodes, such as some which contain production and consumption in the lower grid levels, can also not be assumed Gaussian, as Figure 2 shows. More complex assumptions on the distribution shape should thus be used, such as Gaussian Mixture Models, which consist of a weighted average of Gaussians, but they require a fine-tuning of the distribution shape model. For this reason, the non-parametric approach is preferred over the non-parametric one.



**Figure 2:** Density estimation for a node with production and consumption in the lower grid level. The  $x$ -axis is in normalized units of power.

### 3.2. Non-parametric probabilistic forecasting of the loads

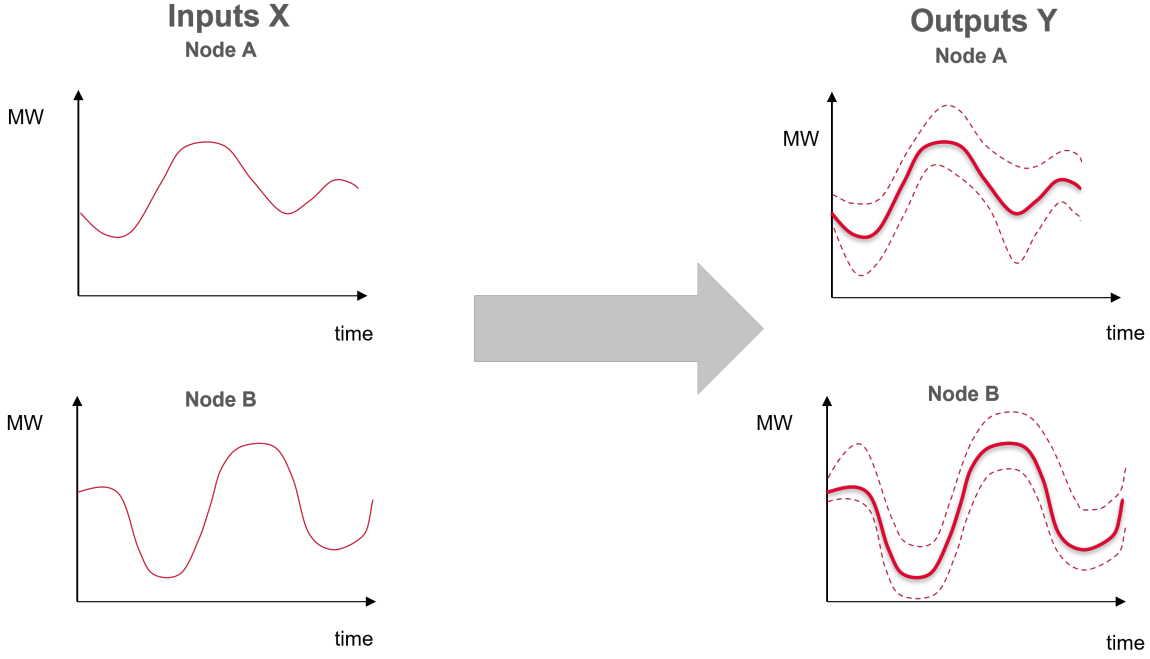
Non-parametric probabilistic forecasting consists in forecasting the conditional quantiles for the following day. It is non-parametric because it does not require any assumption on the distribution shape. In statistics and probability, a quantile is a value that divides a dataset into several subgroups of equal sizes. Specifically, for  $0 < q < 1$ , the  $q$ -th quantile is a value in a dataset such that  $q$  percent of the data is less than or equal to that value, and  $(1 - q)$  percent of the data is greater than or equal to that value. For example, the median is the 50<sup>th</sup> percentile or the 0.5 quantile, as it divides the dataset into two equal halves.

Quantile regression is a statistical technique used to model the relationship between a response variable and one or more predictor variables. Unlike traditional linear regression, which focuses on modeling the conditional mean of the response variable given the predictor variables, quantile regression models the conditional quantiles of the response variable.

In simple mathematical terms, the goal of quantile regression is to model the conditional relationship between an output vector  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_d)$  and an input vector  $\mathbf{X} =$

---

<sup>3</sup>Normalized here means that the data was transformed by subtracting the average and dividing by the standard deviation.



**Figure 3:** Day-Ahead load forecasting

$(X_1, X_2, \dots, X_p)$ . Let  $q$  be a quantile of interest, such as the 50th percentile (median) or the 95th percentile. The goal of quantile regression is to estimate the conditional  $q^{th}$  quantile of  $y_i$  given  $\mathbf{X}$ , denoted by  $q(y_i|\mathbf{X})$ . Figure 3 summarizes the problem in the bivariate case for two hypothetical nodes, nodes A and B. The goal here is to forecast the conditional quantiles for the next day for both nodes, based on the values of these nodes the previous day.

To estimate the quantity  $q(y_i|\mathbf{X})$ , a loss function that penalizes the difference between the actual quantile and the predicted quantile is used, such as the Pinball Loss. The Pinball loss or tilted absolute loss is given by:

$$l_q(y_i, \hat{y}_i) = \begin{cases} (y_i - \hat{y}_i)q & \text{if } y_i \geq \hat{y}_i \\ (\hat{y}_i - y_i)(1 - q) & \text{otherwise} \end{cases} \quad (1)$$

with  $q$  the target quantile,  $y_i$  the real value and  $\hat{y}_i$  the quantile forecast of  $Y_i$ . For the multivariate case, the mean pinball loss is used, defined as the average over the nodes:

$$L_q(\mathbf{Y}, \hat{\mathbf{Y}}) = \frac{1}{N} \sum_{k=1}^N l_q(y_i, \hat{y}_i) \quad (2)$$

The tilted absolute value function places more weight on differences in the  $q^{th}$  quantile than on differences in other quantiles.

Machine learning and deep learning models are therefore perfectly suited for quantile regression as they can use the quantile loss to forecast the conditional quantiles. These models then iteratively minimize this loss across all the observations in the dataset and obtain coefficients corresponding to the minimal loss. Machine learning models can only fit one quantile  $q$ , but deep learning models allow forecasting of several quantile levels at the same time, for example by calculating the Pinball Loss as average loss over the desired quantiles. For the sake of simplicity, it was decided here that the deep learning models would forecast all

the quantiles simultaneously. Once a model is fitted, it can be used to make new predictions about  $Y_i$  for new values of  $\mathbf{X}$ . However, fitting one model per quantile could further increase overall performance.

Therefore, quantile regression allows modeling of each conditional distribution  $Y_i$  given  $\mathbf{X}$ , rather than just the conditional mean. In this thesis,  $\mathbf{Y}$  corresponds to loads of all the nodes for the following day and is thus of size  $d = NT$ ,  $N$  being the number of nodes and  $T = 24h$ , the hours of the following day to forecast.

To decide what should be given as input to forecast  $\mathbf{Y}$ , the autocorrelations of the dataset are analyzed. Moreover, external temporal features are analyzed to determine relevant time parameters that can be used as external features, such as the time of the day or the season, as the behavior of the consumers and thus the behavior of the production is expected to be strongly dependent on them.

### 3.2.1. Temporal analysis

The historical loads show strong temporal patterns. This section focuses on the autocorrelation, i.e., to past lagged values, and on the analysis of the main frequencies of the signal to look for temporal patterns and behavioral changes in the production and consumption over time, as shown in Figure 4.

#### Autocorrelation analysis

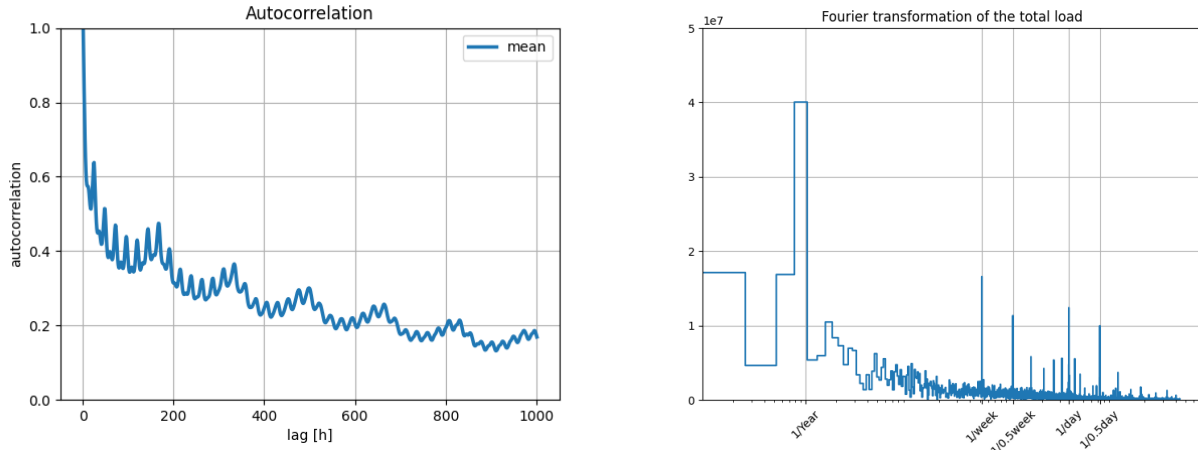
Autocorrelation, i.e., correlation to past lagged values, provides valuable information and understanding of the loads. On average, the loads show a great autocorrelation, i.e., a strong positive correlation to past lagged values. Figure 4a shows the autocorrelation averaged over the nodes for lagged values in hours. This figure shows that the loads are on average positively correlated to past recent values  $t - 1h$  and  $t - 2h$ , to  $t - 24h$  and  $t - 48h$ , as well as to  $t - 168h$ , which corresponds to the same hour the same day a week ago, and further decreases over time. The autocorrelation analysis shows that on average the net loads are positively correlated to past recent values, peaking around  $t - 24h$ . Forecasting models should thus be based on the values for the loads at  $t - 24h$ . As the other values for the net loads of the previous day show a relatively high autocorrelation, adding all the hours of the previous day could help improve the forecasting models by providing additional valuable information.  $t - 1h$  cannot be used for forecasting as the focus is put on day-ahead forecasting so the newest data is not available yet.

#### Discrete Fourier transformation

To analyze the time dependence of the loads throughout the year, the discrete Fourier transformation (DFT) can be used. Intuitively, the DFT decomposes a signal into a sum of sinusoidal waves with different frequencies, amplitudes, and phases. The magnitude of each frequency component in the DFT gives information about the corresponding frequency components in the original signal. The DFT is a mathematical operation that can be applied to any signal to analyze its main frequencies, and it extends the Fourier decomposition of a periodic signal to non-periodic signals. Mathematically, given a sequence of  $N$  complex numbers  $x_n$  for  $n = 0, 1, \dots, N - 1$ <sup>4</sup>, the DFT  $X_k$  is a sequence of  $N$  complex numbers defined

---

<sup>4</sup>In this work,  $x_n \in \mathbf{R}$  as they correspond to net load values.



(a) Autocorrelation averaged over the nodes. (b) Fourier coefficients of the total load (log scale).

**Figure 4:** Autocorrelation to past values and Fourier transformation of the total load.

by:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-2\pi i k n / N} \quad (3)$$

for  $k = 0, 1, \dots, N - 1$ ,  $i$  being the complex number such that  $i^2 = -1$ . The Euclidean norm  $\|X_k\|_2$  of each complex number  $X_k$  can then be computed to analyze the main frequencies of the signal.

Figure 4b shows the norm of the coefficients of the DFT for the total net load in Switzerland. Each peak corresponds to a main frequency of the total net load, and therefore as evolutions of consumption and production behaviors throughout the year. The peak at a frequency of  $1/year$  for example shows that the total load follows a yearly pattern, meaning that it changes throughout the year: in this case, the total load is higher in the winter because the demand and therefore the production is higher in the winter due to lower temperatures. It decreases in spring and summer when temperatures become milder before increasing again in autumn. The peak at a frequency of  $1/week$  shows that the total net load also follows a weekly pattern, less important than the yearly pattern as the peak is not as important. Indeed, this can be explained by the different behaviors during the weekdays and the weekends. The peak at  $1/day$  corresponds to behavior changes during the day: the net load is usually lower at night than during the day because human activity is lower. The intraday peak at  $1/0.5day$  corresponds to peak consumption hours that happen twice a day, most often in the morning when people wake up, and in the evening when people get home after work. The other peaks are harder to interpret and will therefore not be considered.

### Temporal features for forecasting models

Day-ahead forecasting models should thus be based on the past 24 hours to forecast the following 24 hours. The main frequencies of the loads should be integrated, either as categorical features for machine learning models or as cosine and sine for deep learning models. Indeed, integrating them as frequencies allows to locate similar periods or months nearby on a circle, which is reasonable as they are likely to have similar behavior. For example, January (1) and December (12) are both winter months, so production and consumption

behaviors are expected to be similar. They should thus have similar categorical features and be located close by on the circle of their frequency.

Forecasting models should thus integrate these frequencies as inputs. Traditional machine learning models can only consider one-dimensional features, and these features are thus embedded as categorical features. For example, a time column is added to the inputs as  $0, 1, \dots, 23$  and months are added as  $1, 2, \dots, 12$ . However, this representation puts 23:00 and 00:00 very far from each other, even though they are supposed to be quite similar and should therefore be close in terms of mathematical distance. Deep learning models can handle two-dimensional features and the frequencies can thus be embedded as frequencies, i.e., as sine and cosine with the frequencies listed above. For each frequency above  $f$  corresponding to a peak, the temporal features can be embedded as frequencies. For example, the daily peak with a frequency  $f = 1/24h$  can be embedded as  $\cos(2\pi ft)$  and  $\sin(2\pi ft)$  where  $t$  is the time of the day. For deep learning models, these features are embedded as a column vector  $\mathbf{F} = (X_{F_1}, \dots, X_{F_m})$  which summarizes the main characteristics of the net load.

### 3.2.2. Comparison of forecasting models

Several forecasting models are tested and tuned to find the most suitable one for this problem. Several models with growing complexities are tested and compared.

The dataset, i.e., the loads for the years 2019 to 2022 in Switzerland and its belt, is then divided into three subsets: one for training, one for validation and one for testing. The model tries to minimize the loss on the training dataset, whereas the validation set is used to tune the hyperparameters<sup>5</sup> of the model and prevent it from overfitting. The performance of the models is then evaluated on the test dataset. The performance of several models is then compared using the following two probabilistic Key Performance Indicators (KPIs) to estimate the sharpness and the accuracy of the results (Konstantinou, Savvopoulos, and Hatziaargyriou 2021). The KPIs are detailed in section 3.2.3. Figure 5 represents a simple overview of the machine learning models<sup>6</sup>, whereas figure 6 explains the architecture of neural networks in this case. The focus is put on the inputs and the outputs of each model as well as their basic architectures.

### Quantile linear regression

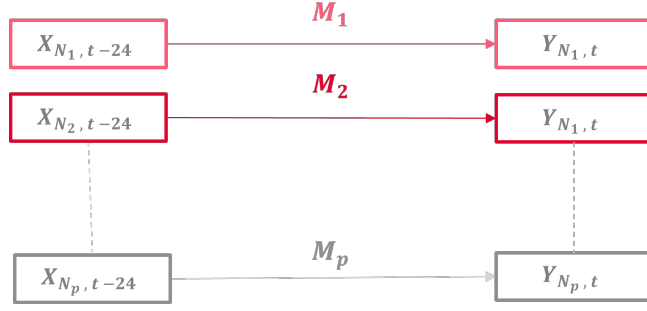
The simplest model for quantile regression is to use a quantile linear regression. This linear regression minimizes the Pinball loss by assuming that the output  $Y_{i,t}$  is a linear function of the input  $X_{i,t-24}$  for each quantile and each node (see Figure 5):

$$Y_{i,t} = a_i X_{i,t-24} + b_i \quad (4)$$

with  $a_i$  and  $b_i$  the coefficients of the linear regression for each node  $i$  and each quantile. These coefficients are time independent as linear regression cannot integrate any external feature. Because it is the simplest model, quantile regression will be used as a benchmark.

<sup>5</sup>A hyperparameter in machine learning is a parameter that is set prior to the training of a model and is not learned from the data. Hyperparameters control aspects of the model such as the complexity of the model, the learning rate, or the number of hidden layers in a neural network (Wu et al. 2019).

<sup>6</sup>Quantile regression and Gradient Boosting regression



**Figure 5:** Both quantile regression and HistGBR consider the nodes individually. For each node and each quantile, one model  $M_i$  is fitted based on the value 24 hours ago.

### Gradient Boosting Regression

The most suitable machine learning for quantile regression in large dimensions is Scikit’s learn Hist Gradient Boosting Regressor (HistGBR). This model combines several weak decision trees into a stronger model by minimizing the residual loss at each step. This model can only consider one-dimensional inputs: each node is then considered independently and one model is fitted for each considered quantile and each node so that the forecasted quantity becomes  $q(Y_{i,t_0+h} | X_{i,t_0+h-24}, F_1, \dots, F_m)$  for  $1 \leq h \leq 24$ . The categorical features are here embedded as the time of the day (0 to 23), the day of the week (1 to 7), the month of the year (1 to 12), and the year (0 to 3).

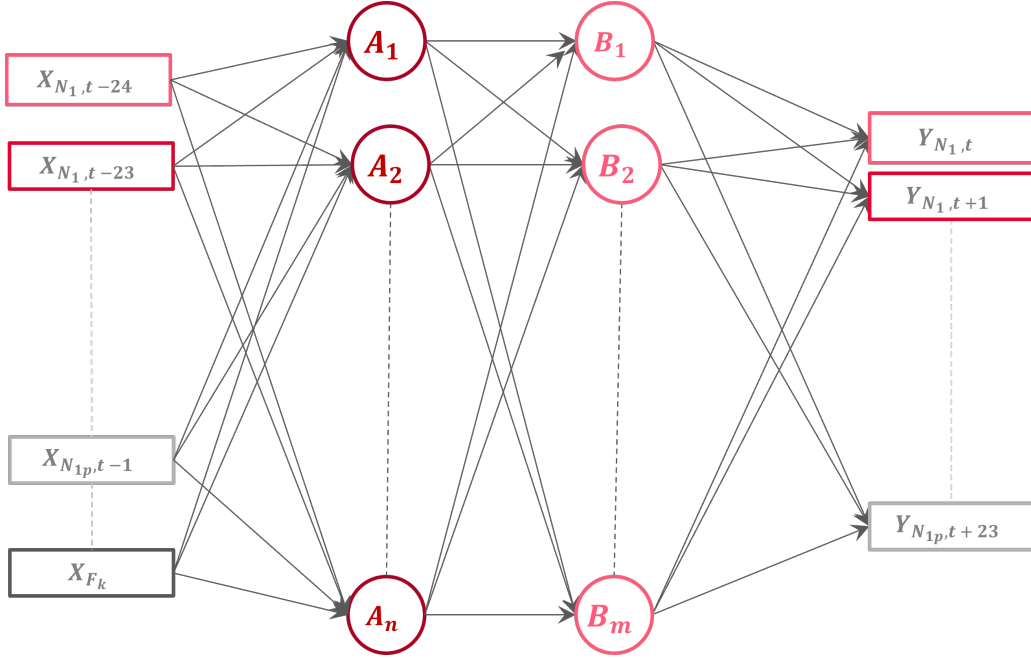
### Dense neural network

As GBR can only handle one-dimensional inputs, neural networks were tested for their ability to handle complex relationships between inputs and produce several outputs at once.

The rough idea behind neural networks is that they are modeled after the structure of the human brain. Like the human brain, a neural network is composed of many interconnected neurons that work together to process information and make decisions. The basic building block of a neural network is a neuron, which takes in one or more inputs, performs a mathematical calculation on those inputs, and produces an output. The calculation that a neuron performs is determined by a set of weights and biases that are assigned to it during the training process. During training, a neural network is presented with a set of input data and corresponding targets that represent the desired output for that input. The network adjusts its weights and biases to minimize the difference between its output and the target output. Once the network has been trained, it can be used to make predictions on new input data by passing the data through the network and computing the output.

The pinball loss defined in section 3 is calculated as the average over several quantile levels, so that all the desired quantiles for every node can be forecasted with the same model. Moreover, the features are here embedded using the cosine and the sine of the main frequencies of the signal, as described in 3.2.1.

More specifically, a dense neural network (DNN) was tested here. A DNN contains several layers which are deeply connected with their preceding layers. This means the neurons of the layer are connected to every neuron of its preceding layer. The input given is therefore now the vector  $\mathbf{X} = (X_{1,t_0}, X_{1,t_0-1}, \dots, X_{d,t_0-23}, X_{F_1}, \dots, X_{F_m})$  (see Figure 6).



**Figure 6:** Example of a dense neural network with two hidden layers A and B with respectively  $n$  and  $m$  neurons. The layers are deeply connected: each neuron is connected to all neurons of the neighboring layers. The gray arrows represent the connections between the inputs, the neurons, and the outputs. Each arrow corresponds to a weight. During the training, the neural network tries to find the weights that minimize the loss.

### Long Short-Term Memory neural network

A Long Short-Term Memory (LSTM) neural network is a type of artificial neural network that is suitable for time-series forecasting as it can remember information for a long time (see section 2). This section presents the general idea and the main equations of LSTM neural networks (Sak, Senior, and Beaufays 2014).

More specifically, the memory cell has three main components: an input gate, an output gate, and a forget gate. The input gate controls which information is allowed into the cell, the forget gate decides which information to forget, and the output gate controls which information is outputted from the cell. When a new piece of data is inputted into the neural network, it first passes through the input gate, which determines which parts of the data are important and should be stored in the memory cell. Then, the forget gate decides which information should be discarded from the memory cell based on its relevance. Next, the memory cell updates itself by combining the new input data with the information that was previously stored. Finally, the output gate decides which information should be outputted from the memory cell as the final prediction or result.

Mathematically, an LSTM can be represented as a set of equations that update its internal state at each time step. First, the LSTM receives an input vector  $x_t$  and a previous hidden state  $h_{t-1}$ . It then computes three gates that determine how much information should be let through:

- The input gate  $i_t$ , which decides how much new information should be added to the memory cell. This is computed as:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (5)$$

where  $W_i$  and  $b_i$  are weight matrix and bias vector respectively.<sup>7</sup>

- The forget gate  $f_t$ , which decides how much of the previous memory cell  $C_{t-1}$  should be retained. This is computed as:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (6)$$

where  $W_f$  is a weight matrix and  $b_f$  is a bias vector.

- The output gate  $o_t$ , which decides how much of the current memory cell should be output as the hidden state. This is computed as:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (7)$$

where  $W_o$  and  $b_o$  are the weight matrix and bias vector respectively.

Next, the LSTM updates its memory cell by combining the previous cell value  $C_{t-1}$  with the new input information:

$$C_t = f_t C_{t-1} + i_t \tanh(W_C[h_{t-1}, x_t] + b_C) \quad (8)$$

where  $W_C$  and  $b_C$  are the weight matrix and bias vector respectively. Finally, the LSTM updates its hidden state  $h_t$  based on the updated memory cell  $C_t$  using the tanh function to squash the cell values between -1 and 1:

$$h_t = o_t \tanh(C_t) \quad (9)$$

In this case, the input data is reshaped so that the LSTM network can analyze the temporal patterns. The input vector  $\mathbf{X}$  is here reshaped to a matrix with  $N$  rows and  $T$  columns, corresponding to all the load values of the previous 24 hours. Since LSTM neural network can handle multidimensional data, the temporal features are embedded as sine and cosine and are added as rows to the input matrix.

### 3.2.3. Key Performance Indicators definition

To compare the performance of the models detailed above, this section defines two Key Performance Indicators (KPIs) that are traditionally used to assess the performance of probabilistic forecasts, namely the Prediction Interval Normalized Average Width (PINAW) and the Prediction Interval Coverage Probability (PICP). Both are defined at a certain confidence level and account respectively for the sharpness and the accuracy of probabilistic forecasts. These metrics are defined independently from the loss as they are easier to interpret.

#### PINAW

The PINAW accounts for the sharpness of the predictions and is defined for a certain level of confidence  $0 < \alpha < 0.5$  as the average of the prediction interval sizes, normalized with the load range to compare the nodes to each other. Mathematically:

$$PINAW = \frac{1}{NT} \sum_{k=1}^N \frac{1}{LR(k)} \sum_{i=1}^T |Y_{k,t,0.5+\alpha/2} - Y_{k,t,0.5-\alpha/2}| \quad (10)$$

---

<sup>7</sup>The sigmoid function  $\sigma$  is defined as  $\sigma(x) = (1 + \exp(-x))^{-1}$  for  $x \in \mathbf{R}$

with  $N$  the number of nodes,  $T$  the number of periods in the test dataset,  $Y_{k,t,0.5+\alpha/2}$  the prediction corresponding to the quantile  $q = 0.5 + \alpha/2$  for the node  $k$  at time  $t$ ,  $LR(k) = \max(X_k) - \min(X_k)$  the load range of the node  $k$  over the whole dataset. The PINAW is unitless and is thus expressed in %. A low PINAW means that the predictions are sharp, whereas a high PINAW means that the predictions are wide. The PINAW should thus be as small as possible.

### PICP

To account for the accuracy of the predictions, the Prediction Interval Coverage Probability (PICP) is used. It corresponds to the percentage of points that fall within the confidence interval for a certain level  $\alpha$ . It is again calculated as the average over the nodes and the considered period. Mathematically,

$$PICP = \frac{1}{NT} \sum_{k=1}^N \sum_{i=1}^T I_{k,t} \quad (11)$$

where  $I_{k,t} = 1$  if the true value  $x_{k,t} \in [Y_{k,t,0.5-\alpha/2}, Y_{k,t,0.5+\alpha/2}]$  and 0 elsewhere. The PICP is unitless too and will be expressed in % as well. A PICP close to 1 means that almost all the values fall within the prediction interval so that the accuracy of the model is good, and vice versa.

Three different forecasting models are tested here, and their performance is evaluated based on the metrics defined above, calculated over the test dataset. More specifically, a Gradient Boosting Regressor (GBR), a dense neural network, and an LSTM neural network are investigated here as they have proven to be efficient for time series forecasting (see section 2).

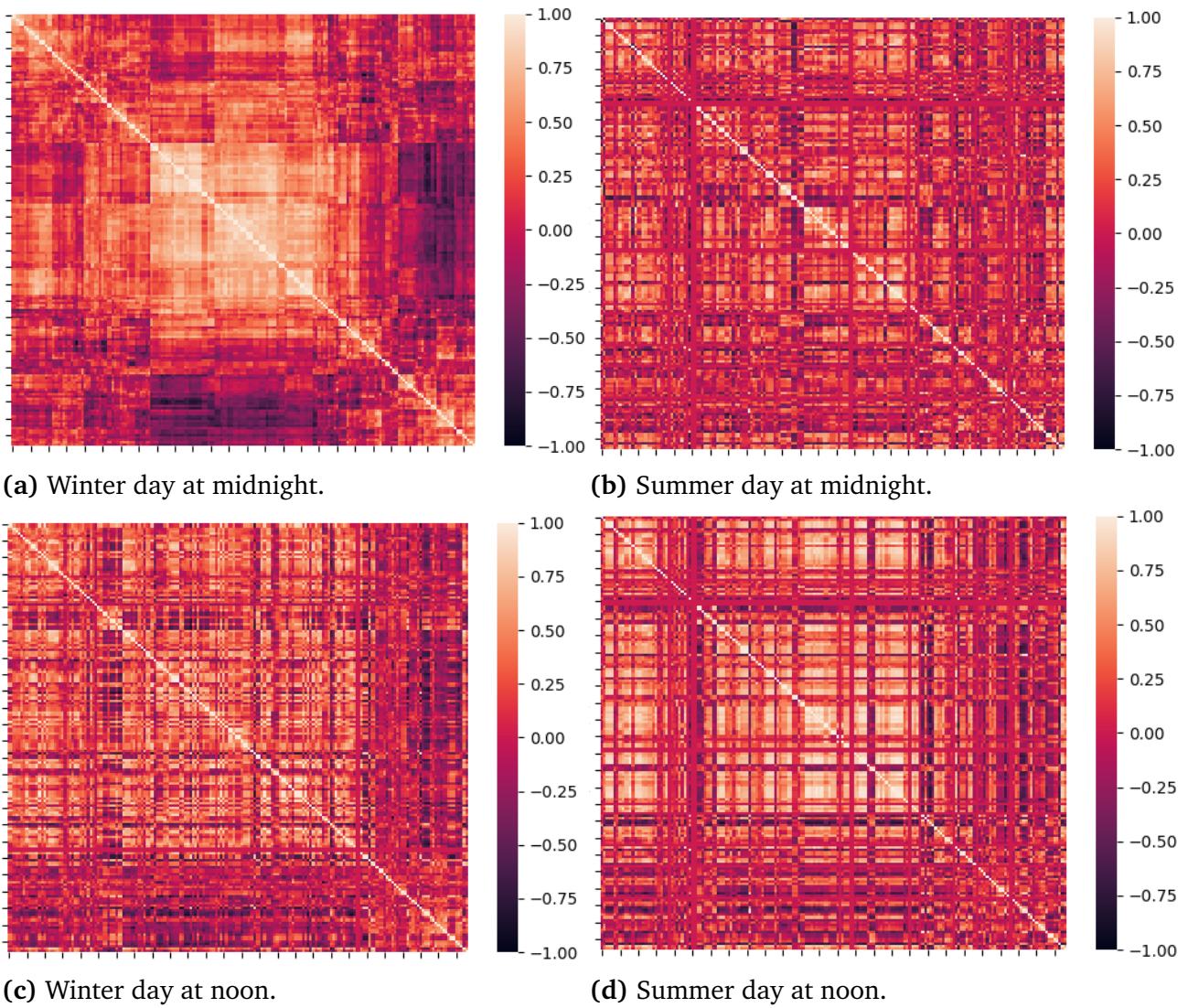
## 3.3. Sampling random variables for Monte-Carlo simulations

As the production increases with the consumption, the nodes are expected to be spatially correlated. The methodology developed by Toubreau et al. 2019 based on copula theory is here adapted to integrate dynamically the spatial correlations in the scenarios. Section 3.3.1 justifies the need to integrate dynamically the spatial correlations. Section 3.3.2 presents the copula theory, whereas section 3.3.3 presents the algorithm to create scenarios for PPF that integrate spatial correlations. The output of this sampling algorithm can then be used as input for PPF.

### 3.3.1. Correlation analysis

The correlation matrix of the loads shows that the nodes cannot be considered independently. Figure 7 shows the correlation matrix for all the nodes for a winter and a summer day calculated based on the eight past weeks for a specific hour, here midnight. More precisely, the correlation matrix was here calculated by considering only the values from the eight previous weeks at midnight.

Figure 7 shows that some nodes are positively correlated, some are negatively correlated, and some are rather uncorrelated. For example, if the sun is shining, nodes close by with PV capacity will be positively correlated. Production nodes are also somewhat correlated: indeed, the production needs to increase with the demand to meet it. Negative correlations



**Figure 7:** Correlation heatmaps for the nodes for a winter and a summer day based on the previous 8 weeks at midnight and at noon. The  $x$  and  $y$  axis correspond to the nodes in Switzerland and its belt, in the same order for the two considered days.

correspond to the correlations between production and consumption nodes: indeed, production is increasing with the demand, which leads to negative correlations as they are opposite in sign. This can also correspond to production nodes and nodes with installed pumped hydro capacity as such nodes can store power.

Moreover, correlations change over time: the correlation heatmap for a summer day and a winter day changes significantly (see Figure 7a vs 7b). This can be explained by the changes in the behavior of the production and consumption over the year: in the summer for example, when the snow in the mountains is melting, the hydro production increases, which leads to different production sites and a different correlation heatmap. The correlation heatmap also changes throughout the day: for both days the correlation heatmaps have differences between noon and midnight (see Figure 7a vs 7d), but the differences at noon for the summer day and the winter day are smaller (see Figure 7c vs 7d).

The analysis of the correlation matrix shows that the nodes cannot be considered independently as some are positively or negatively correlated. Moreover, the correlations change

throughout the year and throughout the day. Correlations must therefore be considered dynamically.

### 3.3.2. Copula theory

A classic method to integrate correlations is to model the multivariate distribution using copula theory and sample from it (see section 2). This section presents an overview of the copula theory and how to incorporate it for Monte-Carlo simulations for PPF.

Copula theory is used to construct multivariate distributions to model the dependency structure. The approach to constructing a multivariate distribution using a copula is based on the idea that each marginal distribution can be transformed using the cumulative density function (CDF) so that each marginal distribution has a uniform distribution. The dependence structure can then be expressed as a multivariate distribution of the obtained uniforms in a so-called copula. Copulas were first introduced in the late 1950s by Sklar. Sklar himself described copula as a method to "join together one-dimensional distribution functions to form multivariate distribution functions" (Plischke and Borgonovo 2019). Sklar's theorem is the basis of the copula theory and is described below for the bivariate case, and it can of course be extended to higher dimensions.

**Theorem 1 (Sklar's theorem)** *Let  $X_1$  and  $X_2$  be two random variables and  $F_{X_1}$  and  $F_{X_2}$  their respective cumulative density functions. Their joint distribution  $F_{X_1 X_2}$  can be written as a copula  $C$ :*

$$F_{X_1 X_2}(x_1, x_2) = C(F_{X_1}(x_1), F_{X_2}(x_2)) \quad (12)$$

*If  $F_{X_1}$  and  $F_{X_2}$  are continuous, then  $C$  is unique; otherwise, the copula  $C$  is unique on the range of values of the marginal distributions.*

**Definition 3.1 (Copula)**  *$C: [0, 1]^2 \rightarrow [0, 1]$  is a copula if  $C$  is a joint cumulative distribution function of a two-dimensional random vector on the unit square  $[0, 1]^2$  with uniform margins. For two random variables  $U_1$  and  $U_2$  following a uniform distribution on  $[0, 1]$ , it is defined as:*

$$C: [0, 1]^2 \rightarrow [0, 1] \quad (13)$$

$$(u_1, u_2) \mapsto P(U_1 \leq u_1, U_2 \leq u_2) \quad (14)$$

There are different families of copulas, for example, Gaussian, Elliptical, or Archimedean copulas. Two families of copula are tested here, namely Gaussian and Student copulas, and their capacity to accurately capture the spatial correlations between the nodes is tested. Moreover, the effect of the number of samples is investigated to determine the number of samples that can achieve reasonable performance.

### 3.3.3. Copula methodology applied to scenario generation

Copula theory can therefore model the spatial dependency structure of the nodes. The method developed here is adapted from Toubreau et al. 2019 to create scenarios by combining the non-parametric forecasts and the copula theory. Following these steps, it can be applied to create realistic scenarios for PPF for a specific day:

1. A non-parametric forecasting of the marginal loads is obtained following section 3.2,

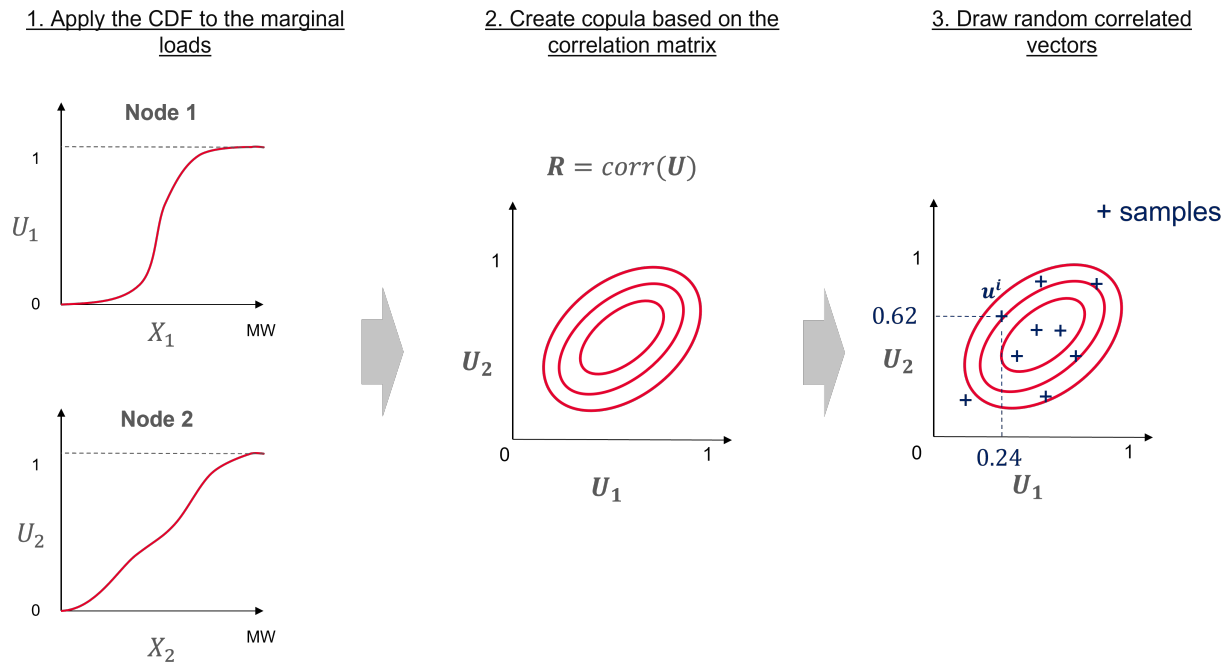
2. Copula theory: to integrate the dynamic dependency structure, the following calculations are done based on the previous 8 weeks:
  - (a) The marginal loads are uniformed using the CDF,
  - (b) The correlation matrix  $R$  of size  $d = NT$ ,  $N$  being the number of nodes and  $T = 24h$ , is computed. This makes it possible to integrate spatial and temporal dependencies structure between the nodes and to model the evolution of the dependency structure throughout the day,
  - (c) A copula is created based on the correlation matrix  $R$ ,
  - (d)  $k$  random correlated vectors  $(\mathbf{u}^i)_{1 \leq i \leq k} \in [0, 1]^d$  are drawn from the copula.
3. The  $(\mathbf{u}^i)_{1 \leq i \leq k}$  are mapped back to  $MW$  based on the non-parametric forecasts of the marginal loads: for each coordinate of each sample  $\mathbf{u}^i$ , the corresponding quantile is associated. For example, if  $u_1^0 = 0.6$ , then for the first coordinate, the quantile 0.6 is chosen and the corresponding value in  $MW$  is used, and so on and so forth. The values between the forecasted quantiles are linearly interpolated for this process.

Steps 1. and 2. are independent. Step 3. combines forecasts 1. and the correlated vectors 2. to create the scenarios by mapping the correlated vectors using the forecasted quantiles for each node.

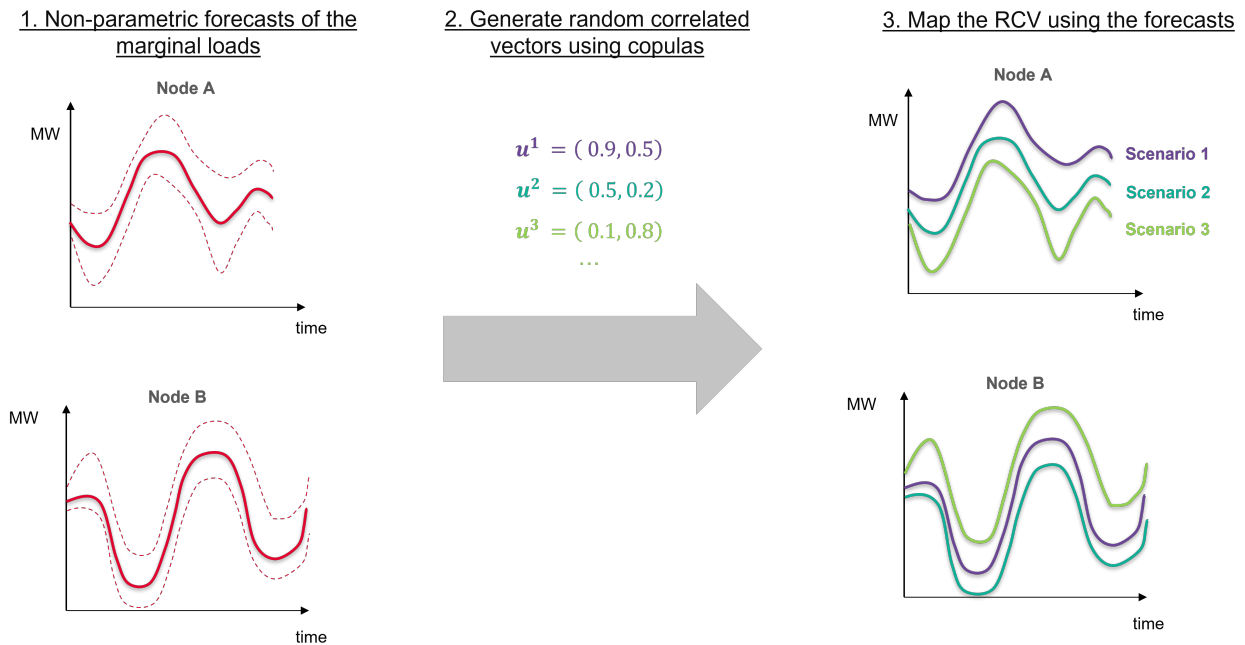
Figure 8 summarizes step 2 of the above methodology in the bivariate case for two hypothetical nodes, nodes 1 and 2, that are assumed to be positively correlated. First, the marginal net loads  $X_i$  are transformed into uniform distributions  $U_i$  using their respective CDFs. Then, the correlation matrix  $R$  of the vector  $\mathbf{U} = (U_1, U_2)$  is computed and the copula is created. The marginals of the copula are therefore uniformly distributed. Random correlated vectors  $(\mathbf{u}^i)_{1 \leq i \leq k} \in [0, 1]^2$  are then sampled from the copula. These vectors thus integrate the spatial dependency structure of the nodes. These vectors take values in  $[0, 1]$  and are mapped back to  $MW$  using the non-parametric forecasts. For example, for a hypothetical vector  $\mathbf{u}^i = (0.24, 0.62)$  sampled from the copula, a vector  $x^i = (q_{24}, q_{62})$  is created by taking for the first coordinate the quantile 24% of the first coordinate and for the second coordinate the quantile 62% of the second coordinate. The vector  $x_i$  thus integrates the dependency structure between nodes 1 and 2.

Figure 9 summarizes how the non-parametric forecasts and the random correlated vectors sampled from the copula are used to create the scenarios. The values are hypothetical. Each of the random correlated vectors generated by the copula are mapped using the non-parametric forecasts: the corresponding quantile is associated to each coordinate of each random correlated vector. For example, for  $\mathbf{u}^1 = (0.9, 0.5)$ , the value associated with the quantile 90% is taken for node A, and the value associated with the quantile 50% is taken for node B. The random correlated vectors are mapped for each hour for each node, but this is not shown on the figure.

Overall, this method allows the modeling of any kind of marginal distribution whilst dynamically integrating the spatiotemporal dependency structure of the nodes. Using copula theory, random correlated vectors are created that are then used to create scenarios for PPF using the non-parametric forecasts.



**Figure 8:** Generating random correlated vectors for MCS using copula theory - example for the bivariate case.



**Figure 9:** Example of mapping random correlated vectors. For  $u^1 = (0.9, 0.5)$ , the value associated with the quantile 90% is taken for node A, and the value associated with the quantile 50% is taken for node B.

### 3.4. Power flows calculations

For each of the  $k$  scenarios, the power flows are calculated deterministically with the PSS/E python API. This section presents the applied network topology as well as how the solution of the AC power flows equations are estimated. Eventually, it presents the N-1 criterion and how contingency analysis is used to ensure that it is respected.

#### 3.4.1. Applied topology

The topology of the grid has a great impact on the power flows. Indeed, outages or maintenance can directly influence the power flows in the high-voltage network. In this thesis, the focus is put on retrieving the topology from the DACF and changing the loads within Switzerland and its belt in the DACF, so that planned outages and maintenance are integrated into the power flow calculations. To compare the scenarios to the real-time values, the power flows are also calculated with the real-time power injections based on the DACF topology. Indeed, the RTSN topology also integrates possible remedial actions that might have been enforced by the operators, such as topological changes or production redispatch. These remedial actions should not be considered in the calculations of the power flows as they correspond to real-time operations and rely on human decisions which thus cannot be forecasted. The topology used for the computation of the power flows for the scenarios is the DACF topology, so planned outages and maintenance are integrated into their calculations.

#### 3.4.2. AC power flows

AC power flow is a method used in power systems analysis to determine the steady-state behavior of an electrical power system. It consists in calculating the voltage, current, and power flow through each component of the system, such as generators, transformers, and transmission lines, under normal operating conditions. The AC power flow equations can be written as a set of nonlinear equations, where the unknowns are the voltage magnitude and phase angle at each node in the system. The equations are derived from Kirchhoff's laws (Monticelli 1999).

To solve these non-linear equations, the Newton-Raphson method is commonly used. This method starts with an initial guess for the voltage magnitude and phase angle at each node in the power system, and then iteratively updates those values until a convergence criterion is met. More precisely, this method consists in linearizing the nonlinear equations at each iteration using a first-order Taylor series approximation. This linearized system is then solved using matrix algebra to obtain an updated estimate for the unknown variables. The updated estimates are then used to linearize the equations again, and the process is repeated until the convergence criterion is met. The convergence criterion is typically based on the difference between the estimated and actual values of the variables, and it is typically set to a small tolerance value. This criterion ensures that the iterative process stops when the solution is close enough to the actual solution (Tinney and Hart 1967).

The Newton-Raphson method is widely used in power systems analysis because it is efficient and can handle large and complex systems with non-linear behavior. However, it can be sensitive to the choice of initial guesses, and it may not converge or converge slowly in some cases. As a result, various modifications and improvements have been proposed to improve the convergence and robustness of the method, such as the use of acceleration techniques and

pre-conditioning methods, but such improvements will not be considered here (Tostado-Véliz et al. 2021).

### **3.4.3. Contingency analysis**

In power systems analysis, N-1 security is a criterion used to ensure the reliability of a power system. The term N-1 refers to the concept of analyzing the system with the assumption that any one component (such as a generator, transmission line, or transformer) can fail or be taken out of service without causing a system-wide failure.

Contingency analysis ensures that the N-1 criterion is respected. The purpose of contingency analysis is to identify potential points of failure within the power system by analyzing the system's response to a variety of different failure scenarios, such as the loss of a generator or transmission line and evaluating the impact on the overall system. Using the PSS/E Python API, the response to various failure scenarios can be estimated for each monitored grid element. In this thesis, the focus is put on the worst-case failure for each grid element, i.e., the failure with the greatest impact on the system. Indeed, the operator in the control room uses this information to take a decision and possibly enforce remedial actions if necessary.

Contingency analysis is therefore an important tool for ensuring the reliability and stability of large, interconnected power systems. Indeed, they are subject to a range of external factors, such as weather events, equipment failures, and other disruptions that can result in grid failures and eventually power outages. By identifying potential points of failure and developing remedial actions to mitigate those risks, utilities such as SwissGrid can help ensure that their customers have reliable access to power.

In this thesis, only the failures with the greatest impact on the grid will be studied. Indeed, failures with a medium to low impact on the grid will not be considered here as they have little influence on it.

## 4. Results and discussion

This section presents the main results of the thesis and provides some insight into the performance of the final model. This section is divided according to the methodology developed in section 3. It focuses on the non-parametric of the marginal loads, on the generation of random correlated vectors, and on the scenario creation. The final scenarios are used as input for PPF and tested for a sample day. A few examples are shown.

### The historical loads

The first challenge was to retrieve historical data. The data is stored in RAW files, which can be read by PSS/E<sup>8</sup> and contain a collection of unprocessed data that specifies a Bus/Branch network model for the establishment of a power flow working case. A RAW file has multiple groups of records (data blocks), with each group containing a particular type of data needed in power flow (Lee et al. 2022). These RAW files are read via the PSS/E python API. From these RAW files, the active power at each node is retrieved. This power corresponds to the net load, i.e., the sum of generation (negative) and consumption (positive) in the lower grid levels at each node. If the net load is positive, the node is retrieving power from the grid. If the net load is negative, the node is injecting power into the grid.

### 4.1. Non-parametric forecasting of the marginal loads

This section focuses on finding the best model for non-parametric forecasting of the marginal loads. As a reminder, the goal is to forecast for each node  $i$  its conditional quantiles for the following day  $q(Y_{i,t_0+1,\dots,t_0+24} | \mathbf{X}_{t_0-1,\dots,t_0-24}, F_1, \dots, F_m)$ , based on the values from the past 24 hours of all the nodes  $\mathbf{X}_{t_0-1,\dots,t_0-24}$ , to which some temporal features  $F_1, \dots, F_m$  are added (see section 3.2). For the sake of simplicity, the model always forecasts from 00:00 to 23:00 for a day  $D$ , based on 00:00 to 23:00 for  $D - 1$ , with one value per hour of the day.

The quantiles 1%, 10%, 20%, 50%, 80%, 90% and 99% are forecasted for each model. The granularity is higher on the edges due to higher variations in the forecasts on the edges.

As mentioned in section 3, the data is split into three datasets, namely train, validation, and test dataset, corresponding to respectively 70%, 20%, and 10% of the data. The performance is evaluated for the different models by calculating the PICP and PINAW of the forecasts for each node, averaged over the considered period. The sections below compare the performance of the models presented in section 3.2.2.

#### 4.1.1. Comparison of the different forecasting models

For the GBR model, an exhaustive search over specified parameter values was performed, such as the maximum depth of each tree, the learning rate, or  $l^2$  regularization. The results shown here are obtained with the optimal hyperparameters for this model. For the DNN and the LSTM, the hyperparameters search was done manually by trial and error. The results shown here are for the best-found hyperparameters for each model. Whilst the performance of the GBR is probably close to its best performance thanks to the exhaustive search, the performance of the other models might be further increased by a more refined tuning of their

---

<sup>8</sup>Power System Simulator for Engineering (PSS/E) is a software developed by Siemens used by power system engineers to simulate electrical power transmission networks.

hyperparameters. The KPIs are calculated only for the interval 90% centered around the median.

KPI	Linear QR	GBR	DNN	LSTM
PICP (%)	75	67	75	72
PINAW (%)	66	24	45	40

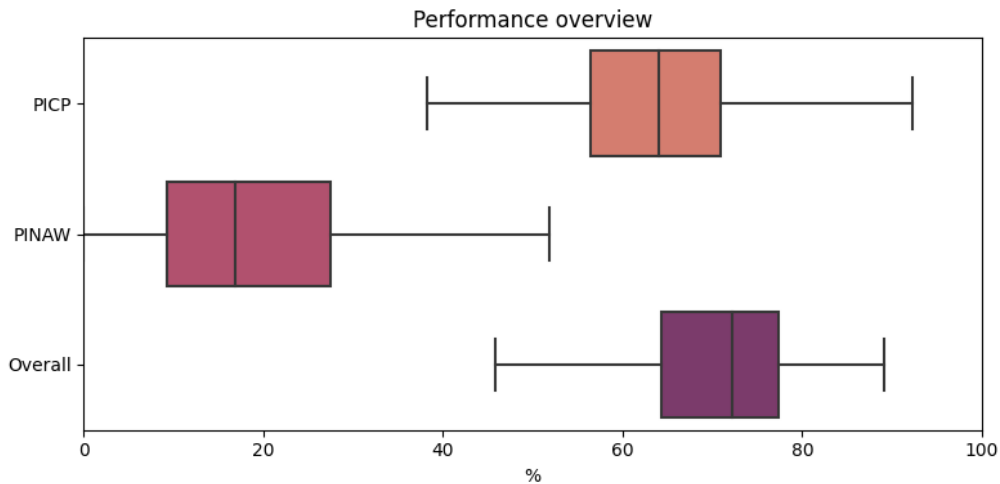
**Table 1:** PICP and PINAW at 90% for the three tested models, in %.

Table 1 summarizes the performance of the three tested models on the test dataset. As expected, the linear quantile regression shows a relatively low performance compared to the other more advanced models. The forecasts are quite broad (relatively high PINAW), most likely because this model cannot integrate any external features such as temporal features, which is why the PICP is relatively high. This model is quite simple as  $Y_{i,t}$  is just a linear function of  $X_{i,t-24}$ , which explains why the forecast is too broad.

The GBR model shows an overall relatively good performance as it has a quite high PICP and the lowest PINAW. Moreover, there is a trade-off between accuracy and sharpness: indeed, if the forecasts are very broad (high PINAW), the accuracy is likely to be close to 100%. Reciprocally, tighter forecasts (low PINAW) are likely to lead to a lower PICP. For this reason, the two tested neural networks have a higher PICP and PINAW: the higher accuracy comes at the cost of more broad predictions. The LSTM reaches a better performance than a DNN, but its performance is still lower than the GBR.

Given these performances, it was decided to choose the GBR model. Indeed, it was decided that the forecasts should rather be sharper and a bit less accurate than the opposite. However, further tuning of the neural networks' hyperparameters, as well as more advanced models, could reach higher performance. More specifically, an auto-regressive LSTM, which forecasts every hour sequentially and feeds each forecast to the next one, could be tested and might reach higher performance.

#### 4.1.2. Performance analysis



**Figure 10:** Average PICP and PINAW distribution for the nodes in Switzerland and its belt.

This section presents a more detailed analysis of the performance of the GBR under the

defined KPIs. The results are shown for each node in Switzerland. Figure 10 shows the PICP and the PINAW for the nodes in Switzerland and its belt. The PICP and PINAW are calculated as averaged values over the test dataset so that the distribution represents the differences in performance across the nodes. The overall KPI is a linear combination of the PICP and PINAW to account for their trade-off. It is calculated for each node as  $overall = \frac{1}{2}(PICP + (1 - PINAW))$  and represents the overall performance. If it is close to 100% for a node, it means that the PICP is high and the PINAW is low at the same time, which shows the overall good performance of the model on this specific node.

Figure 10 shows a relatively great disparity between the nodes. Indeed, some nodes have a very good overall performance, meaning that they have a high PICP and a low PINAW. The performance was analyzed under the specificities of each node to try to look for patterns in the performance, for example looking at whether production or consumption nodes had the best performance, but no distinct pattern was found. Some nodes have a surprisingly low PICP or high PINAW due to a lack of data and are here not represented.

#### 4.1.3. Example of forecasts

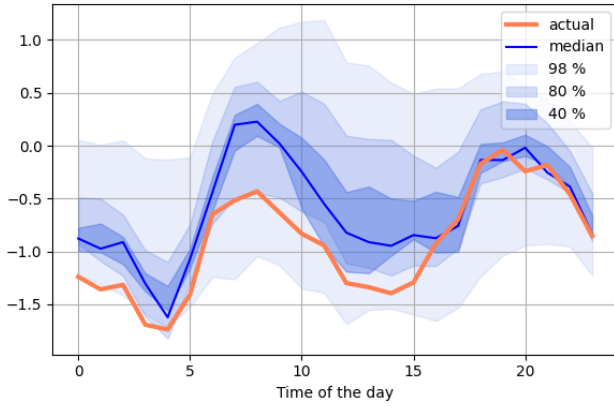
This section presents the forecasted confidence interval for four different nodes in Switzerland, nodes A, B, C, and D for a spring day from the test dataset (Figure 11).

Nodes A and B are both consumption nodes and nodes C and D both have consumption and production in the lower grid levels. The unit on the  $y$ -axis is in normalized units of power. The  $x$ -axis represents the time of the day. The blue zones correspond to the confidence intervals at different confidence levels, centered around the median.

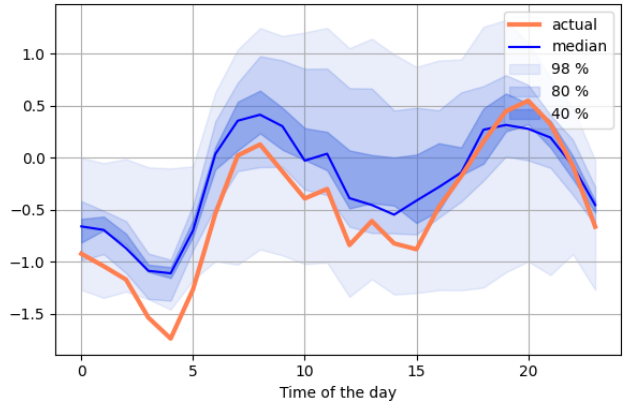
For nodes A and B, the load increases sharply in the early morning before decreasing again during the day. In the evening, the load peaks again before starting to decrease. This is a classic shape for load: the load increases when people wake up in the morning and come home in the evening. The forecasting model is here capable of capturing the evolutions of the loads throughout the day.

Node C has several hydropower plants (among them pumped hydropower plants) in the lower grid levels, whereas node D only has one pumped hydro facility. For node C, the load is negative at the two bottoms of the curve, meaning that the node injects power into the grid, whereas it is positive in the morning and in the afternoon, meaning that the consumption is higher than the production for this node. This depends on the power production plans of the power plants and their decisions are based on market prices, which is why the behavior of power plants can be hard to predict. Node D has also several hydro facilities, including pumped hydropower plants that are used to store power. The production pattern is here quite clear as the power plants produce at almost constant power in the morning and in the evening, leading to a strong bimodal behavior in this case. Again, the model captures quite accurately the evolutions of the net load throughout the day. The square shape of node D is less well captured by the model, maybe due to the strong changes during the day as the model seems to smooth the evolution of this net load. Nodes C and D are harder to predict since they depend on the decisions of the power plants to produce or not, and therefore by extension depend on the market. Here, integrating market signals such as Day-Ahead spot prices could allow for increasing the performance of the model.

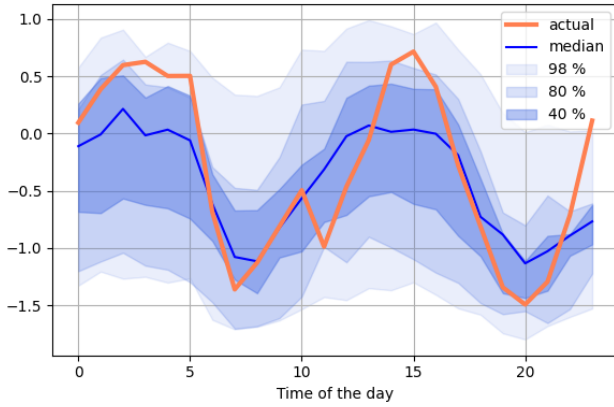
Overall, the model quite accurately captures the evolutions of the net loads throughout



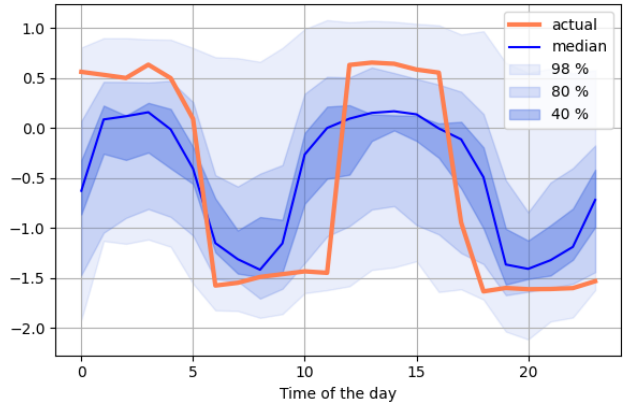
(a) Node A - consumption.



(b) Node B - consumption.



(c) Node C - production and consumption.



(d) Node D - production and consumption.

**Figure 11:** Example of forecasts for a day in spring for four different nodes in Switzerland. The orange line represents the actual values. The blue line represents the median forecast, whereas the blue zones correspond to different confidence intervals centered around the median, namely 40%, 80%, and 98%. The  $y$ -axis is in normalized units of power.

the day. Predictions can perhaps be considered a bit too broad as the quantiles 1% and 99% are probably a bit too conservative, leading to very broad forecasts for these four nodes. Moreover, a few points are outside of the 98% confidence interval, such as at 4:00 for node B. For these reasons, before implementation in Swissgrid's system, more advanced forecasting models should be tested to improve its overall performance.

## 4.2. Integrating spatial correlations using copula theory

As the non-parametric forecasts of the marginal do not integrate any kind of spatial dependencies between the net loads, copula theory is then used to integrate a spatial dependency structure and create scenarios for PPF. This section presents the main results from the copula theory. Two families of copulas are compared based on their ability to capture the spatial dependencies between the nodes. The effect of the number of samples is also studied. Scenarios for PPF are then created by associating the correlated vectors generated from the copula to the different quantile forecasts from section 3.2.

### 4.2.1. Choice of the copula family

There are different kinds of parametric copulas. Two of them are investigated here, namely Gaussian and Student, to determine which can capture the best spatial correlations between the nodes. They both take as input the correlation matrix. To assert the performance of these families:

1. The marginals are transformed to a uniform distribution using the CDF and the copula is created based on the correlation matrix  $\mathbf{R}$
2. Samples are drawn from the copula. To investigate the effect of the number of samples, 20, 50, 100, and 1000 samples are drawn,
3. The samples are transformed back to  $MW$  using the inverse CDF,
4. The correlation matrix  $\mathbf{R}'$  of the samples is computed for each number of samples,
5. The Root Mean Squared Error (RMSE) between  $\mathbf{R}$  and  $\mathbf{R}'$  is computed for each number of samples. The best copula is chosen as the one which has the lowest overall RMSE.

Table 2 summarizes the RMSE for the tested copulas and different numbers of samples drawn. The RMSE is unitless as it has the same unit as the Pearson correlation coefficients and is thus shown in %. This table shows that the Gaussian copula is the most appropriate to capture the correlations as it reaches the lowest RMSE for each number of samples. However, the differences between the copula families become tinier when the number of samples drawn increases. This table shows that family choice has rather little influence on capturing the dependence structure. From now on, Gaussian copulas will be used to model the dependency structure between the nodes because it can achieve better performance for fewer samples.

Number of samples	RMSE Gaussian	RMSE Student
10	28 %	32 %
20	19 %	24 %
50	13 %	15 %
100	9 %	10 %
1000	4 %	5 %

**Table 2:** RMSE comparison between different copula families. The index corresponds to the number of samples drawn from the copula, and the values to the RMSE for each number of samples and copula family, in %.

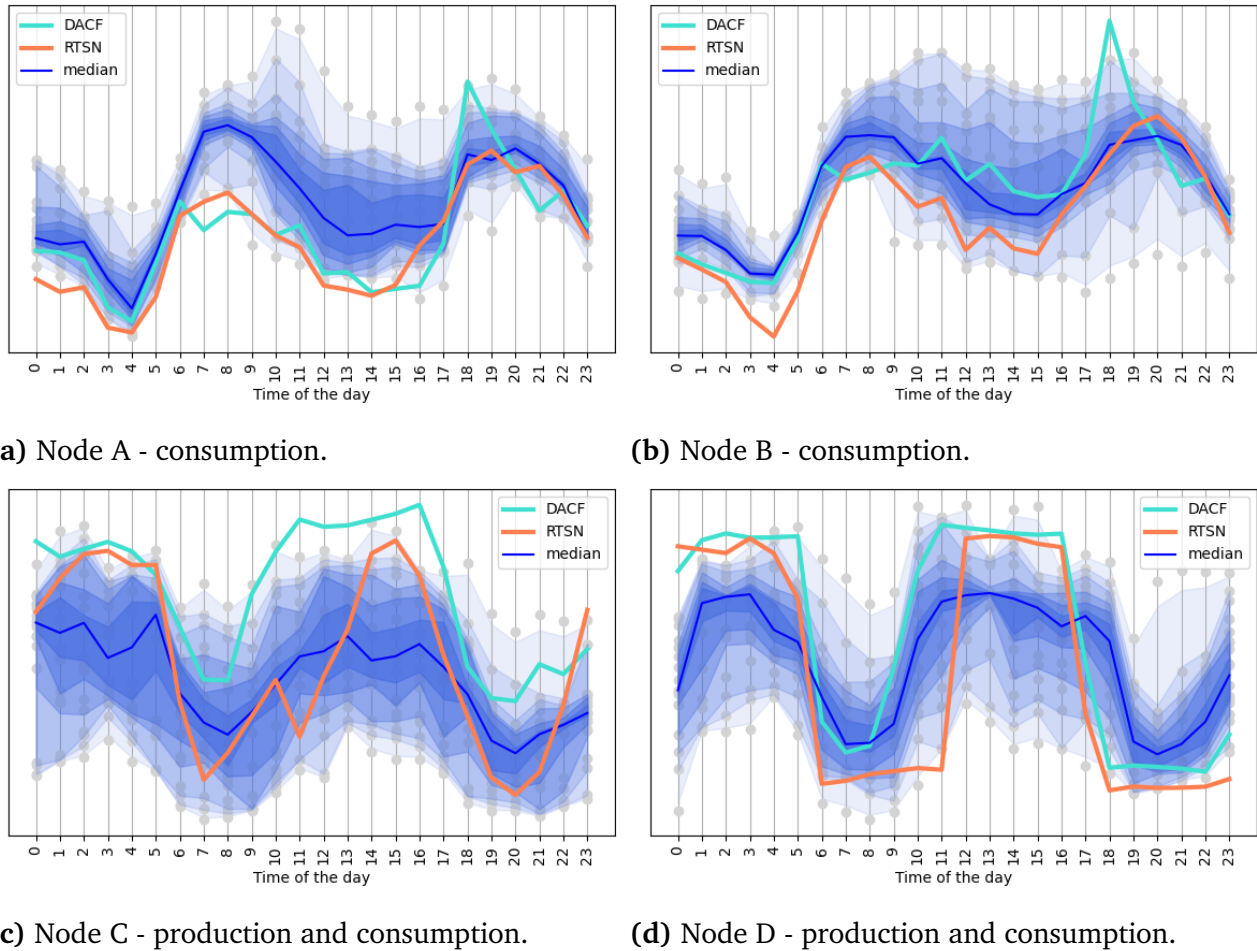
Moreover, the number of samples influences the RMSE: the more samples, the lower the RMSE. This can be explained by Borel's law of large numbers: "If an experiment is repeated a large number of times independently under identical conditions, then the proportion of times that any specified event occurs approximately equals the probability of the event's occurrence on any particular trial". In simple words: the larger the number of repetitions, the better the approximation.

Therefore, drawing a larger number of samples allows a more accurate representation of the spatial correlations. However, there is here a trade-off between accuracy and computational time, and the gains in accuracy between 20, 50, 100, and 1000 samples might be low compared to multiplying the computational time by 2.5, 5, or 50. For this reason, calculations will be made with 20 samples, but higher accuracy at a higher computational cost might be achieved using more samples.

### 4.2.2. Scenarios for PPF

This section presents a few examples of scenarios that can be used for PPF. The non-parametric forecasts from section 4.1 are combined with random correlated vectors using copula theory. More specifically, as it is not realistic to forecast every percentile for every node, the missing quantiles are linearly interpolated for each node.

Figure 12 shows the net load evolution throughout the day for four different nodes, the same nodes as in section 4.1.3. Nodes A and B are consumption nodes, whereas nodes C and D have both production and consumption in the lower grid levels.



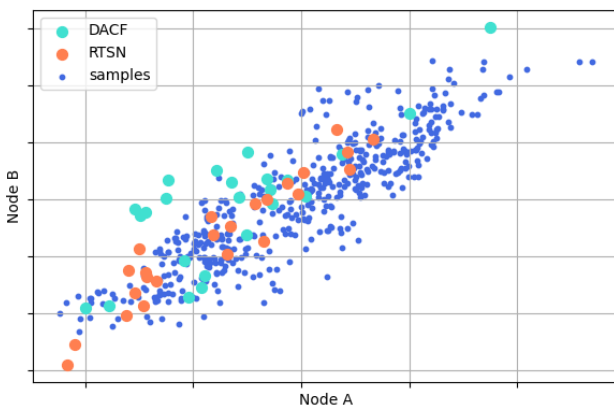
**Figure 12:** Example of scenarios for a day in spring for four different nodes in Switzerland. The orange line represents the actual values, and the turquoise line the DACF forecast. The blue line represents the median line of all the scenarios for each hour, whereas the blue zones correspond to different confidence intervals centered around the median, namely 20%, 40%, 60%, 80%, 90%, and 98%. The  $y$ -axis is in normalized units of power.

Nodes C and D both have consumption and production capacity. For node D, the DACF, Swissgrid’s current model, seems to perform better than the scenarios. Indeed, the DACF integrates power production schedules (PPS) that are retrieved from the power plants. These PPS are non-binding, meaning that the power plants can decide to change their production plan throughout the day depending on the market. For nodes A and B, the load is a bit overestimated in the first half of the day, and the actual load is outside of the prediction range of the model. This is probably due to an overestimation in the non-parametric forecast.

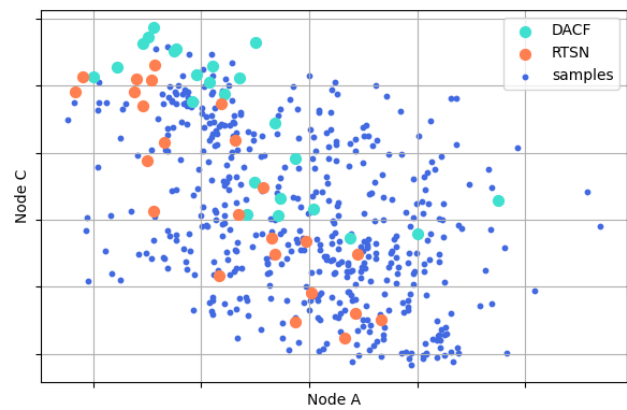
The net load for node A is indeed overestimated in the non-parametric forecast (see Figure 11a).

The scenarios differ slightly from the non-parametric forecasts (Figure 11 vs 12). For the four nodes, the confidence intervals of the non-parametric forecasts and of the scenarios are globally the same. However, for these four nodes, the confidence intervals of the scenarios are smaller than the confidence intervals of the non-parametric forecasts, especially for nodes A and B. This leads to an actual value of the net load outside of the 98% confidence interval of the scenarios for nodes A and B. This might be specific to this date and these nodes, however here copula theory might also exclude unlikely scenarios, such as low loads scenarios for nodes A and B in the early morning.

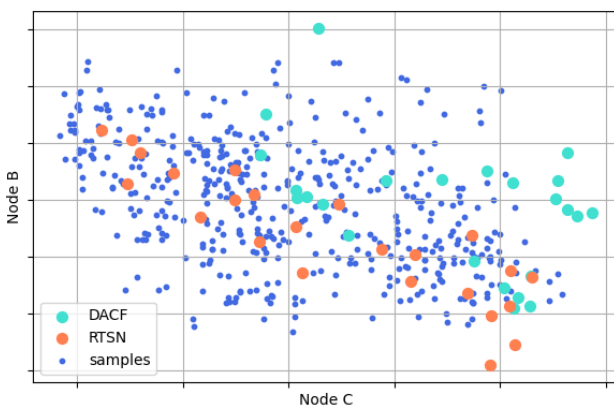
Figure 13 shows the correlations between the same nodes. Figure 13a shows the correlation between nodes A and B, which are both consumption nodes located in the same area, around Zürich. Nodes A and B are for this day strongly positively correlated. The samples and the DACF accurately represent the correlations between the two nodes. Figure 13b shows the correlation between the consumption node A and the production node C. Both nodes are geographically quite far away, but the negative correlation between the two nodes is still vis-



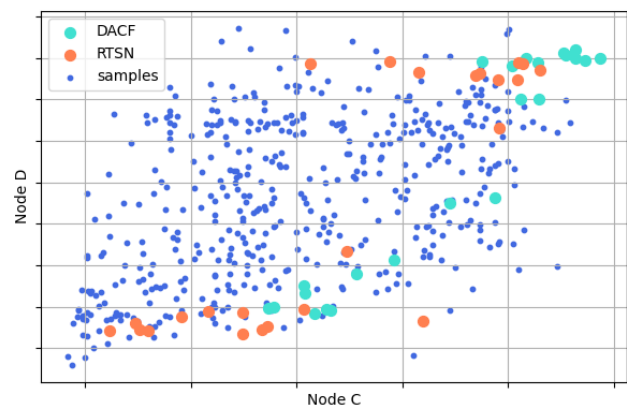
(a) Node A versus node B.



(b) Node A versus node C.



(c) Node C versus node B.



(d) Node C versus node D.

**Figure 13:** Correlation graphs for the same day for different combinations of nodes A, B, C, and D. The blue dots represent the samples, the turquoise points represent the forecasts of Swissgrid's current forecast, the DACF, and the orange points represents the RTSN correlations.

ible. Again, the samples represent quite accurately the correlation, even though they are too many samples in the upper right zone.

Figure 13c shows that nodes C and B are rather negatively correlated, and the samples again capture quite accurately the correlations between these two nodes. Figure 13d shows the correlation between nodes C and D, which are both mostly production nodes. Due to the bimodal behavior of node D in real-time (Figure 12d), the samples here fail to represent the correlations accurately. Indeed, this is due to the choice of the copula: a Gaussian copula cannot represent bimodal modes, so the samples are scattered between the two modes of node D. The choice of the copula could thus be further investigated to face such challenges. However, this configuration is rather rare in the dataset. Such cases should not lead to high errors in the power flows.

This method thus allows the creation of probabilistic scenarios that manage to integrate relatively well the spatial correlations between the nodes.

### 4.3. PPF: an example

The scenarios above can then be used as inputs for PPF. This section focuses on the results of the Monte-Carlo simulation for the same spring day as shown in the scenarios above. The results correspond to the power flows and the N-1 violations for each hour of the day.

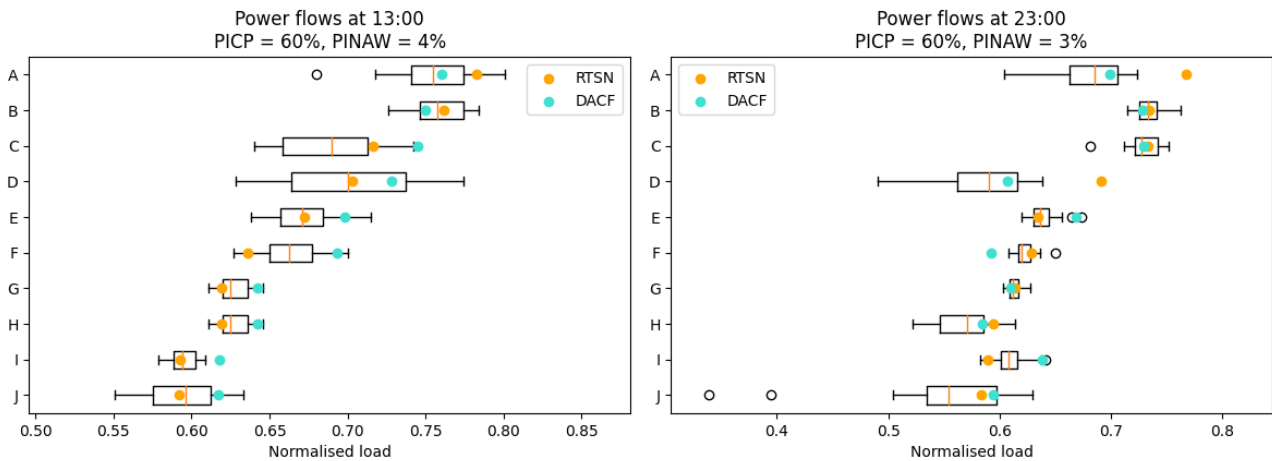
As the Swiss grid has many grid elements, only the grid elements with the highest relative load will be studied here. The relative load is defined as the ratio of the actual load, calculated as the norm of active and reactive power, and the maximal load, which is defined as the maximum capacity of a grid element in normal operating conditions. The values shown on the  $x$ -axis are thus between 0 and 1 and will be interpreted as percentages. The  $y$ -axis corresponds to the most loaded grid elements, denoted alphabetically. They do not correspond to the nodes mentioned in the previous section: they here correspond to grid elements, mainly transmission lines and transformers. The KPIs defined in section 3 are reused here. The confidence level chosen is 80%, meaning that the PICP and PINAW are calculated for the 80% confidence interval. The PINAW is here calculated as the average PINAW, that are here already normalized.

#### 4.3.1. Power flows

Figure 14 shows the power flows computed for the same spring day in the early afternoon and in the late evening. The grid elements are denoted alphabetically and do **not** correspond in the two examples, meaning that A in Figure 14a and in Figure 14b are not necessarily the same grid elements. The boxplots summarize the distribution of the power flows for each grid element based on the scenarios. The orange line corresponds to the median value of the scenarios, the box itself to the interquartile range ( $IQR$ )<sup>9</sup> and the whiskers correspond to  $q_1 - 1.5IQR$  and  $q_3 + 1.5IQR$ . The outliers are represented as single dots outside of the whiskers.

Figure 14a shows that the scenarios provide additional relevant information compared to the DACF. Indeed, for almost all the grid elements, the median of the scenarios is closer to the real-time value than the DACF (except for A). The accuracy of the scenarios is also relatively

<sup>9</sup>The interquartile range is defined as the difference between third quartile  $q_3$  and the first quartile  $q_1$ , corresponding to the 75th and 25th percentiles of the data.



(a) Power flows at 13:00.

(b) Power flows at 23:00.

**Figure 14:** Power flows at 13:00 and 23:00 for the considered spring day. Only the ten most loaded elements are shown for each hour, denoted alphabetically. The index does **not** represent the same grid elements for the two examples. The orange dots represent the actual value, the turquoise point Swissgrid’s current point forecast (DACF), and the boxplot summarizes the power flows estimated from the scenarios.

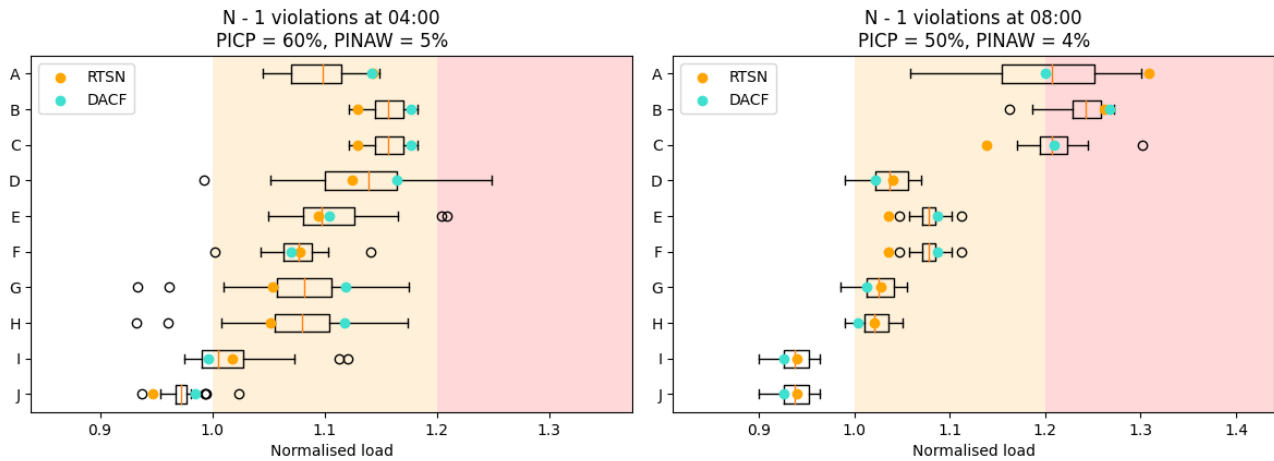
good as the PICP reaches 60%, but the points that are outside the confidence interval fall just outside of it. The PINAW is also quite low: on average, the confidence interval on the prediction is equal to 4% on the element’s capacity.

Figure 14b shows the power flows at 23:00 for the same day. Even though the PICP and PINAW are similar to 13:00, the situation is here not the same. The scenarios perform quite well for most grid elements, but for some of them the error is quite big, especially for A and D. A and D are in this case two transmission lines that are connected to each other through a node, which explains that the error on A is also present on D. However, on the nodes on the extremities of A and D, the scenarios cover the real-time values. The error for these two elements thus comes from other nodes.

#### 4.3.2. Contingency analysis

Contingency analysis is a tool used to ensure the N-1 criterion is respected. For each grid element, the impact of a failure of every other grid element is studied. This section presents the greatest impact of a potential failure. N-1 violations correspond to an overloading component after a grid failure. N-1 violations are a concern in power system planning and operation, and measures are taken to prevent them from occurring or to mitigate their impact if they do occur. N-1 violations are even more important than power flows calculations. Indeed, the grid is designed so that contingencies barely happen. The operators need to make sure that the power grid is N-1 secure at any time, meaning that in case of any grid element failure, all the other grid components should not be overloaded. An N-1 violation corresponds to a violation of the N-1 criterion, meaning that a grid element would be overloaded in case of a grid element failure. N-1 violations can occur quite often, and one of the roles of the operators in the control room is to prevent them and mitigate them. In this section, the graphs show the ten largest impacts of the failures on the system, which correspond to N-1 violations if the normalized load is over 1, meaning that the grid element would be

overloaded. In simple words, it answers the question: what can happen in the worst case for example for element A if one element falls out?



(a) CNE at 04:00.

(b) CNE at 08:00.

**Figure 15:** N-1 violations at 04:00 and 08:00 for the same day. The legend is the same as for Figure 14.

Figure 15 shows the N-1 violations for the same day, but at 04:00 and 08:00. On this specific date, the system faces a high risk: there are many grid elements that face an N-1 violation. However, the DACF overestimated the risk in most cases, which shows that a scenario-based approach can further increase reliability in the system by providing additional information. At 04:00, the scenarios show that the DACF is a high-load scenario for elements B, C, D, G, and H, whereas the scenarios show that the real-time value is expected to be lower, which was here the case. At 08:00, the scenarios are centered around the DACF for element A, and the real-time value is a bit outside the range of the scenarios, but again the scenarios provide useful additional information as it shows the value in the worst case. For element C at 08:00, the DACF and the scenarios both overestimated the risk, but the element is still facing an N-1 violation.

## 5. Conclusion

The increasing share of intermittent renewables and demand flexibility in the European power grid has led to new challenges, making it harder to predict power production and demand. Traditional deterministic models are being challenged, and new methods are necessary to address these uncertainties. This work aims to increase the system's security and reliability by integrating stochasticity in the demand and production.

The method developed here is based on Monte-Carlo simulations, which estimate uncertainty on each network element for the following day by calculating power flows many times for different realistic scenarios. Creating these scenarios is a real challenge due to the spatio-temporal correlations of the multivariate probabilistic problem. For this reason, it was chosen to separate the problem into two parts. First, a non-parametric probabilistic forecast of the marginal loads was used based on the temporal correlations as it can model any distribution shape. Secondly, the spatial correlations between the nodes were integrated by using copula theory to generate random correlated vectors. The scenarios are created by combining the non-parametric forecasts of the marginal loads and the random correlated vectors. These scenarios can then be used as inputs for Monte-Carlo simulations and estimate of the density of the different network elements, as well as the possible N-1 violations.

This work has shown promising results; however, further improvements are required before implementation in real-time operations. There are two ways to increase the accuracy of the methodology. On the one hand, the forecasting model is likely to be improved by more advanced models, such as auto-regressive LSTM, which predicts each hour sequentially. Other external features could also be added to increase the accuracy, such as market prices, weather data and forecasts, and holidays, which have here not been considered. On the other hand, other copula configurations could also be tested, such as defining for example a copula for each month of the year based on the data for the same months of the previous years, so that weekdays and weekends could be distinguished. Moreover, a risk metric and a probabilistic remedial action optimizer could be implemented to suggest possible remedial actions in case of N-1 violations in most scenarios for example.

Overall, this work presents a detailed methodology for creating scenarios for probabilistic power flows. The results are promising, but further tuning and improvements are required before implementation in Swissgrid's system.

## References

- Anderson, T.W. and I. Olkin (1985). "Maximum-likelihood estimation of the parameters of a multivariate normal distribution". In: *Linear Algebra and its Applications* 70, pp. 147–171. ISSN: 0024-3795. DOI: [https://doi.org/10.1016/0024-3795\(85\)90049-7](https://doi.org/10.1016/0024-3795(85)90049-7). URL: <https://www.sciencedirect.com/science/article/pii/0024379585900497>.
- BFE (2022). "Large-scale hydropower". URL: <https://www.bfe.admin.ch/bfe/en/home/supply/renewable-energy/hydropower/large-scale-hydropower.html>.
- Bin, L. et al. (June 2018). "Probabilistic Load Flow Analysis of Power System Network Considering Uncertainty with Generation and Correlated Loads". In: *International Journal of Simulation: Systems, Science and Technology* 19, pp. 6.1–6.7. DOI: 10.5013/IJSSST.a.19.03.06.
- Borkowska, Barbara (1974). "Probabilistic Load Flow". In: *IEEE Transactions on Power Apparatus and Systems* PAS-93.3, pp. 752–759. DOI: 10.1109/TPAS.1974.293973.
- Carr, Michael, Mark Mc Grath, and Eabhnat Ní Fhloinn (2020). "14 - Laplace transforms: Engineering applications of Laplace transforms". In: *Calculus for Engineering Students*. Ed. by Jesús Martín-Vaquero et al. Mathematics in Science and Engineering. Academic Press, pp. 305–326. DOI: <https://doi.org/10.1016/B978-0-12-817210-0.00021-7>. URL: <https://www.sciencedirect.com/science/article/pii/B9780128172100000217>.
- Czapaj, Rafał, Jacek Kamiński, and Maciej Sołtysik (2022). "A Review of Auto-Regressive Methods Applications to Short-Term Demand Forecasting in Power Systems". In: *Energies* 15.18. ISSN: 1996-1073. DOI: 10.3390/en15186729. URL: <https://www.mdpi.com/1996-1073/15/18/6729>.
- D’Ettorre, F. et al. (2022). "Exploiting demand-side flexibility: State-of-the-art, open issues and social perspective". In: *Renewable and Sustainable Energy Reviews* 165, p. 112605. ISSN: 1364-0321. DOI: <https://doi.org/10.1016/j.rser.2022.112605>. URL: <https://www.sciencedirect.com/science/article/pii/S1364032122005007>.
- Delgado, C. and J.A. Domínguez-Navarro (2014). "Point estimate method for probabilistic load flow of an unbalanced power distribution system with correlated wind and solar sources". In: *International Journal of Electrical Power Energy Systems* 61, pp. 267–278. ISSN: 0142-0615. DOI: <https://doi.org/10.1016/j.ijepes.2014.03.055>. URL: <https://www.sciencedirect.com/science/article/pii/S0142061514001665>.
- Ebeed, Mohamed and Shady H. E. Abdel Aleem (2021). "Chapter 1 - Overview of uncertainties in modern power systems: uncertainty models and methods". In: *Uncertainties in Modern Power Systems*. Ed. by Ahmed F. Zobaa and Shady H.E. Abdel Aleem. Academic Press, pp. 1–34. ISBN: 978-0-12-820491-7. DOI: <https://doi.org/10.1016/B978-0-12-820491-7.00001-3>. URL: <https://www.sciencedirect.com/science/article/pii/B9780128204917000013>.
- Gürses-Tran, Gonca, Hendrik Flamme, and Antonello Monti (2020). "Probabilistic Load Forecasting for Day-Ahead Congestion Mitigation". In: *2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pp. 1–6. DOI: 10.1109/PMAPS47429.2020.9183670.

- Hajian, Mahdi, William D. Rosehart, and Hamidreza Zareipour (2013). “Probabilistic Power Flow by Monte Carlo Simulation With Latin Supercube Sampling”. In: *IEEE Transactions on Power Systems* 28.2, pp. 1550–1559. DOI: 10.1109/TPWRS.2012.2214447.
- Konstantinou, Theodoros, Nikolaos Savvopoulos, and Nikos Hatziargyriou (2021). “Scenario Based Probabilistic Energy Demand Forecasting using Autoencoders and Gaussian Mixture Models”. In: *2021 International Conference on Smart Energy Systems and Technologies (SEST)*, pp. 1–6. DOI: 10.1109/SEST50973.2021.9543133.
- Lee, Ryungyeong et al. (2022). “Advanced Probabilistic Power Flow Method Using Vine Copulas for Wind Power Capacity Expansion”. In: *IEEE Access* 10, pp. 114929–114941. DOI: 10.1109/ACCESS.2022.3218644.
- Li, Quan, Xin Wang, and Shuaiang Rong (Nov. 2018). “Probabilistic Load Flow Method Based on Modified Latin Hypercube-Important Sampling”. In: *Energies* 11, p. 3171. DOI: 10.3390/en11113171.
- Liao, Wenlong et al. (2022). “Scenario prediction for power loads using a pixel convolutional neural network and an optimization strategy”. In: *Energy Reports* 8, pp. 6659–6671. ISSN: 2352-4847. DOI: <https://doi.org/10.1016/j.egyr.2022.05.028>. URL: <https://www.sciencedirect.com/science/article/pii/S2352484722008745>.
- Lim, Bryan and Stefan Zohren (Feb. 2021). “Time-series forecasting with deep learning: a survey”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379.2194, p. 20200209. DOI: 10.1098/rsta.2020.0209. URL: <https://doi.org/10.1098/rsta.2020.0209>.
- Lindemann, Benjamin et al. (2021). “A survey on long short-term memory networks for time series prediction”. In: *Procedia CIRP* 99. 14th CIRP Conference on Intelligent Computation in Manufacturing Engineering, 15-17 July 2020, pp. 650–655. ISSN: 2212-8271. DOI: <https://doi.org/10.1016/j.procir.2021.03.088>. URL: <https://www.sciencedirect.com/science/article/pii/S2212827121003796>.
- Monticelli, A. (1999). “Power Flow Equations”. In: *State Estimation in Electric Power Systems: A Generalized Approach*. Boston, MA: Springer US, pp. 63–102. ISBN: 978-1-4615-4999-4. DOI: 10.1007/978-1-4615-4999-4\_4. URL: [https://doi.org/10.1007/978-1-4615-4999-4\\_4](https://doi.org/10.1007/978-1-4615-4999-4_4).
- Nie, Peng et al. (2021). “Prediction of home energy consumption based on gradient boosting regression tree”. In: *Energy Reports* 7, pp. 1246–1255. ISSN: 2352-4847. DOI: <https://doi.org/10.1016/j.egyr.2021.02.006>. URL: <https://www.sciencedirect.com/science/article/pii/S2352484721001049>.
- Plischke, Elmar and Emanuele Borgonovo (2019). “Copula theory and probabilistic sensitivity analysis: Is there a connection?” In: *European Journal of Operational Research* 277.3, pp. 1046–1059. ISSN: 0377-2217. DOI: <https://doi.org/10.1016/j.ejor.2019.03.034>. URL: <https://www.sciencedirect.com/science/article/pii/S0377221719302929>.
- Qin, Hongqian (2019). *Comparison of Deep learning models on time series forecasting : a case study of Dissolved Oxygen Prediction*. arXiv: 1911.08414 [eess.SP].
- Sak, Haşim, Andrew Senior, and Françoise Beaufays (2014). *Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition*. arXiv: 1402.1128 [cs.NE].

- Shen, Guizhu et al. (2018). “Deep Learning with Gated Recurrent Unit Networks for Financial Sequence Predictions”. In: *Procedia Computer Science* 131. Recent Advancement in Information and Communication Technology: pp. 895–903. ISSN: 1877-0509. DOI: <https://doi.org/10.1016/j.procs.2018.04.298>. URL: <https://www.sciencedirect.com/science/article/pii/S1877050918306781>.
- Simard, Clarence and Bruno Remillard (Jan. 2013). “Forecasting Time Series with Multivariate Copulas”. In: *SSRN Electronic Journal* 3. DOI: 10.2139/ssrn.2295120.
- Steinwart, Ingo and Andreas Christmann (2011). “Estimating conditional quantiles with the help of the pinball loss”. In: *Bernoulli* 17.1, pp. 211–225. DOI: 10.3150/10-BEJ267. URL: <https://doi.org/10.3150/10-BEJ267>.
- Taşcıkaraoğlu, Akın (2017). “Chapter 5 - Impacts of Accurate Renewable Power Forecasting on Optimum Operation of Power System”. In: *Optimization in Renewable Energy Systems*. Ed. by Ozan Erdiñç. Boston: Butterworth-Heinemann, pp. 159–175. ISBN: 978-0-08-101041-9. DOI: <https://doi.org/10.1016/B978-0-08-101041-9.00005-3>. URL: <https://www.sciencedirect.com/science/article/pii/B9780081010419000053>.
- Tinney, William. F. and Clifford E. Hart (1967). “Power Flow Solution by Newton’s Method”. In: *IEEE Transactions on Power Apparatus and Systems* PAS-86.11, pp. 1449–1460. DOI: 10.1109/TPAS.1967.291823.
- Tostado-Véliz, Marcos et al. (2021). “An improved version of the Continuous Newton’s method for efficiently solving the Power-Flow in Ill-conditioned systems”. In: *International Journal of Electrical Power Energy Systems* 124, p. 106389. ISSN: 0142-0615. DOI: <https://doi.org/10.1016/j.ijepes.2020.106389>. URL: <https://www.sciencedirect.com/science/article/pii/S0142061520301812>.
- Toubeau, Jean-François et al. (2019). “Deep Learning-Based Multivariate Probabilistic Forecasting for Short-Term Scheduling in Power Markets”. In: *IEEE Transactions on Power Systems* 34.2, pp. 1203–1215. DOI: 10.1109/TPWRS.2018.2870041.
- Tuinema, Bart W. et al. (2020). “Probabilistic Power Flow Analysis”. In: *Probabilistic Reliability Analysis of Power Systems: A Student’s Introduction*. Cham: Springer International Publishing, pp. 179–208. ISBN: 978-3-030-43498-4. DOI: 10.1007/978-3-030-43498-4\_6. URL: [https://doi.org/10.1007/978-3-030-43498-4\\_6](https://doi.org/10.1007/978-3-030-43498-4_6).
- United Nations (2022). “The 17 goals”. URL: <https://sdgs.un.org/goals>.
- Wang, Chenxu et al. (2020). “A scenario-based analytical method for probabilistic load flow analysis”. In: *Electric Power Systems Research* 181, p. 106193. ISSN: 0378-7796. DOI: <https://doi.org/10.1016/j.epsr.2019.106193>. URL: <https://www.sciencedirect.com/science/article/pii/S0378779619305127>.
- Wu, Jia et al. (2019). “Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization”. In: *Journal of Electronic Science and Technology* 17.1, pp. 26–40. ISSN: 1674-862X. DOI: <https://doi.org/10.11989/JEST.1674-862X.80904120>. URL: <https://www.sciencedirect.com/science/article/pii/S1674862X19300047>.
- Xiao, Ling, Miaotong Li, and Shenghui Zhang (2022). “Short-term power load interval forecasting based on nonparametric Bootstrap errors sampling”. In: *Energy Reports* 8, pp. 6672–6686. ISSN: 2352-4847. DOI: <https://doi.org/10.1016/j.egyrs.2022.05.016>. URL: <https://www.sciencedirect.com/science/article/pii/S2352484722008629>.

- Xu, Xiaoyuan and Zheng Yan (2015a). “Probabilistic load flow evaluation with hybrid Latin Hypercube Sampling and multiple linear regression”. In: *2015 IEEE Power Energy Society General Meeting*, pp. 1–5. DOI: 10.1109/PESGM.2015.7285724.
- (2015b). “Probabilistic load flow evaluation with hybrid Latin Hypercube Sampling and multiple linear regression”. In: *2015 IEEE Power Energy Society General Meeting*, pp. 1–5. DOI: 10.1109/PESGM.2015.7285724.

