Electricity Price Forecasting using Multivariate Price Time Series

Walid Demloj

Thesis submitted for the degree of
Master in Applied Computer and Information Technology - ACIT
(Applied Artificial Intelligence)
30 credits

Department of Computer Science
Faculty of Technology, Art and Design

Oslo Metropolitan University — OsloMet

Spring 2023

Electricity Price Forecasting using Multivariate Price Time Series

Walid Demloj

Electricity Price Forecasting using Multivariate Price Time Series

http://www.oslomet.no/

Printed: Oslo Metropolitan University — OsloMet

Contents

Acknowledgements **1**

Abstract **3**

1 Introduction and Overview **5**
1.1 Motivation and Contributions 7

2 Background and Related Work **9**
2.1 The Norwegian Electricity Market 9
2.1.1 The Wholesale Market 11
2.2 Main Concepts about Neural Networks 15
2.2.1 Neural Network Architectures 17
2.2.2 Long Short-Term Memory 18
2.3 Related Work 20
2.3.1 Electricity Price Forecasting using XAI 23

3 Data and Methodology **25**
3.1 Data Description 25
3.1.1 Challenges with the Dataset 26
3.1.2 Data Analysis 27
3.2 Data Processing 29
3.3 Methods 33
3.3.1 LSTM Model 33
3.3.2 Model Explanation 35
3.3.3 Evaluation Metrics 36

4 Results **39**
4.1 EPF with Multivariate Time Series LSTM 39
4.1.1 Principal Component Analysis 44
4.2 Interpreting Market Coupling using LIME 44

5 Discussion and Conclusion **49**
5.1 Discussion 49
5.1.1 Findings 49
5.1.2 Limitations 51
5.2 Conclusions and Future Work 52

A Correlation Heatmaps and European Bidding Zones **61**

List of Figures

2.1 Overview of the five BZNs in Norway. 11
2.2 The transmission grid of the Nordic countries. 12
2.3 A representation of the merit order curve. 14
2.4 Basic neural network architecture. 16
2.5 The architecture of an LSTM cell with its operators. 19

3.1 Electricity prices of the NO1 and NO4 Norwegian BZNs . . 26
3.2 Historical correlation of NO1 with NO2, NO4 and DE. . . . 27
3.3 Histograms of Pearson correlation coefficients. 28
3.4 Visualization of NaN values in the dataset. 30
3.5 Visualization of the day-ahead auction market. 34

4.1 Diebold-Mariano test between LSTM and DNN models. . . 41
4.2 Feature importance for NO1 using LIME. 45
4.3 Feature importances for selected European BZNs using LIME. 47

A.1 Covariance matrix between the bidding zones in 2019. 62
A.2 Covariance matrix between the bidding zones in 2020. 63
A.3 Covariance matrix between the bidding zones in 2021. 64
A.4 Covariance matrix between the bidding zones in 2022. 65

List of Tables

3.1 Optimal model architecture and hyperparameters. 34

4.1 Evaluation metrics for NO1. 40
4.2 Evaluation metrics for all bidding zones. 43

A.1 Bidding zone abbreviations and corresponding regions. . . . 66

Acknowledgements

I would like to express my deepest gratitude to my main supervisor Leonardo Rydin Gorjão for his exceptional support, guidance and dedication throughout this Master's thesis. His prowess in the field of power markets has truly been inspirational and without his constant commitment and valuable insights, this work would not have been possible. I am also grateful for the crucial knowledge shared by co-supervisor Pedro Lind. Together, their willingness to share expertise and feedback have been invaluable in shaping this thesis and elevating the quality of work. Lastly, many thanks to my family and friends for their unwavering support.

Abstract

Since the recent rise of European electricity prices, the field of electricity price forecasting (EPF) has gained increased popularity. With the renewable energy transition, the complexity of EPF has become more challenging due to the highly volatile nature of renewable energy sources. Additionally, the limited data covering the period with abnormally high prices, make EPF even more daunting. Accurate forecasts are therefore crucial in order to efficiently allocate energy resources. In this thesis, we introduce and investigate a novel approach to reduce the complexity of day-ahead EPF and better understand market coupling. Unlike univariate EPF, where features such as load, demand and weather forecasts are counted for, our approach is strictly based on multivariate electricity price time series from European electricity markets. We utilized an Long Short-Term Memory (LSTM) that successfully was capable of explaining electricity prices with varying accuracy throughout Europe. We found that the electricity market of Norway 1 (NO1) was the simplest to forecast, whereas the electricity markets of Denmark and Netherlands were the most difficult. Our LSTM yielded promising results and significantly outperformed a benchmark model using the same modeling approach. Nevertheless, multivariate price time series for EPF cannot be seen as the superior approach as its forecasts were lackluster in comparison to its counterpart. Moreover, using Local Interpretable Model Agnostic Explanations (LIME) we were able to quantify the importance of European electricity markets and analyze their interconnectivity. As expected, our results show that Germany and Great Britain are among the most influential. However, the electricity markets of Serbia and Croatia appear to have a strong connection with the high electricity prices.

Chapter 1

Introduction and Overview

Electricity has steadily since the 18^{th} century become one of the most important innovations of all time (Austin, 2021). The highly digital and technologically advanced world we know today, could not have been achieved without the contribution of electricity.

Since the first humans were alive 2.4 million years ago (Little, 2021) and up until the 18^{th} century, the world had steadily been changing, becoming more and more advanced as the millenniums went by. However, this constant advancement was nothing like what was to become after the discovery of electricity, which is of no coincidence. Once the true possibilities of electricity were discovered, the world quickly and drastically changed.

Whether you are switching your lights on in your apartment, starting your car or cooking, electricity is very likely to be the source of energy. We are constantly surrounded by electricity which is an integral part of our everyday lives and it allows us to more efficiently perform everything from basic, to more advanced tasks. Facilities such as schools and hospitals are entirely reliant on constant access to electricity in order to operate effectively. Electricity is without a doubt what keeps the world digital and innovative.

Given our heavy dependence on electricity, experiencing shortages or even more severe, electric power outages, can be detrimental to our societies and welfare. Modern day economies are reliant on access to cheap power in order to function properly. It is therefore crucial to be able to delegate generated electricity where the demand is high at any time during the day, and to do this both reliably and efficiently.

The European electricity markets experienced significant changes since their deregulation in the 1990s. Prior to the deregulation, the power sector was traditionally monopolistic and government controlled. Before the new regulations that liberalized the electricity sector took place, the electricity sector was suffering from an increasing dissatisfaction in performance. Most notably, the electric capacity was often exceeding the demand which resulted in an inefficient system (Bye & Hope, 2005). Electricity markets were therefore introduced to optimize efficiency. Additionally, a liberalized electricity market where different market participants can place bids and offers should lead to increased reliability and reduced costs (Jamasb &

Pollitt, 2005). Norway was one of the first countries in Europe to deregulate its electricity sector and the first to provide universal market access, with several countries following shortly after (Norwegian Ministry of Petroleum and Energy, 2021). In the context of energy science, universal market access refers to the ability of all market participants to access and trade in electricity markets, which is essential for well-functioning markets as it provides competition and transparency.

Electricity markets play a vital role in the balancing of power generation and demand. Given that there are no easy and efficient ways to store electricity for longer periods of time at this moment, the supply and demand of electricity must always be at an exact equilibrium[1] (Foroni et al., 2023). To ensure that power generation and consumption are always matched and in balance, electricity is traded in wholesale markets for a given time horizon ranging from short-term to long-term. The day-ahead market, regarded as the backbone of the European spot market and constitutes the largest market based on traded volume, takes place each day all year around. Up until a specific deadline, usually at 12:00 pm, electricity for each hour of the following day is traded.

Electricity is a complex commodity that presents several challenges due to its unique properties. The characteristics of always needing a balance between supply and demand, being non-storable and highly inelastic makes electricity subject to high volatility, sudden price spikes and seasonality. The volatility can partly be explained by the energy transition to renewable energy sources in today's power systems which are incredibly weather dependant. Additionally, the seasonality occurs based on the demand of electricity which often peaks during morning and midday hours. Moreover, households often consume more electricity during the winter season to heat their homes. In order to have a well grounded and predictable electricity market for those involved, modeling and forecasting electricity spot prices has gained increased research interest.

Electricity Price Forecasting (EPF) aims to provide reliable and accurate predictions of the electricity price. Most of the research is dedicated towards predicting the day-ahead market (Jedrzejewski et al., 2022), which requires models to forecast the 24 hourly prices for the following day, based on the same data. Accurate forecasts are essential for reaching the energy transition (Tschora et al., 2022). More recently, EPF has become increasingly challenging due to the current European energy crisis which has resulted in extraordinarily high electricity prices throughout Europe. Russia has historically been an integral gas supplier to continental Europe. However, after the invasion of Ukraine, its gas exports to Europe have significantly been reduced (Council of the EU, 2023). The Russian invasion of Ukraine further strengthened the severity of the energy crisis by disrupting the European energy markets. Electricity markets are intrinsically coupled with economical and political issues in Europe and generally around the world. Since the electricity markets became deregulated, security of supply

[1] Although electricity is a non-storable commodity, the EU funded stoRE research project aims to solve this issue.

has never been a more important and discussed issue in Europe.

Meanwhile, the popularity of machine learning (ML) is growing by the years as computational power and data storage have improved. ML has successfully been applied to several complex problems in fields ranging from finance (Rasekhschaffe & Jones, 2019) to healthcare (Davenport & Kalakota, 2019). Their ability to model non-linearity and extract meaningful information from complex domains have made them advantageous for EPF where statistical methods were superior at the start of the previous decade. However, as ML models are black boxes, their interpretability is limited. Understanding why a model makes its predictions is crucial for the field of EPF with these abnormal electricity prices.

1.1 Motivation and Contributions

Electricity prices throughout Europe are all coupled as countries exchange power between themselves. Consequentially, forecasting electricity prices quickly becomes a complex non-linear problem that is dependent on a country's generation and consumption, but similarly the wants and needs of its neighboring countries. Due to the nature of the underlying problem, ML techniques have successfully been applied to the field. Nevertheless, the recent surge in the complexity of EPF has resulted in a reduced level of ML model interpretability. Therefore, to reduce complexity and better understand market coupling, we are interested in the following research question: is it possible to accurately forecast day-ahead electricity prices for a given region based only on previous price time series of other regions?

The majority of the EPF literature has been focused on univariate time series analyses. These works comprise electricity prices for a single region and data such as generation by type, consumption and weather forecasts. To the best of our knowledge, no one has addressed multivariate time series analyses, where only price time series from different regions are used to make the final prediction. There are several reasons why our multivariate approach could be valuable. First of all, access to regions' historical data is not necessarily easy, but that is not the main reason for our work. As we are interested in how electricity markets are coupled, we have to look directly at the electricity prices in order to truly understand the price coupling between different regions. This could perhaps have been achieved using the univariate modeling approach, however the relationship between historical data and electricity prices might have long delays. For instance, today's water levels at the hydro reservoirs in Norway affect electricity prices for various weeks in the future. Thus, multivariate price time series modeling should better represent the simultaneous interactions between electricity markets. Fundamentally, all historical information of each region is already contained in the electricity prices.

This served as motivation for this thesis which is a two-part investigation. In this thesis, we first investigate the applicability of Long Short-Term Memory (LSTM) artificial neural networks in the field of EPF. LSTMs have shown promising results with sequential time series data and have successfully been applied to EPF (Zihan Chang & Chen, 2019). As the electricity power market is a complex system depending on many features, our main goal was to forecast day-ahead prices of European electricity markets based on the multivariate approach. Further on, most work in EPF is focused on Germany, however, our thesis is focused on Norway and specifically Norway bidding zone 1 (NO1), which serves the southern part of Norway. In the second part of our investigation, we wanted to analyze how the entire European electricity market was coupled, and whether there were any underlying contexts between the electricity prices from different regions. By training an ML model on price time series from different regions and applying that to forecast a region out of sample, we may be able to achieve more reliable forecasts and thus, better understand how the European electricity markets are coupled.

Moreover, understanding why a model makes its predictions is increasingly relevant, especially when working in a highly volatile market such as today's electricity market. Explainable artificial intelligence (XAI) will assist us in understanding what is being captured in a model. Adding to that, XAI allows us to gain insight into which features are more influential than others. In our case, XAI will serve as a tool to explain which electricity regions are the most influential in the European electricity exchange market. The goal is to disentangle the main drivers of electricity prices in European markets.

To summarize, our contributions are fourfold. Firstly, we present a novel approach for EPF based on multivariate time series analyses. Features strongly influencing electricity prices, such as gas prices, renewable energy generation and load are not needed to train the ML model, given that they are already weighted into the electricity price time series. Secondly, we introduce an LSTM capable of estimating electricity prices based on our novel approach. Thirdly, we provide the readers with insights into how the Norwegian power market works, how prices are determined and settled and how electricity markets are interconnected. Lastly, we quantify the importance of different European electricity markets such that we can disentangle drivers for European electricity prices. XAI will serve as a useful tool for the latter.

The rest of the thesis is structured as follows: In Chapter 2 the theoretical background, covering the Norwegian electricity market and the LSTM architecture, is presented. Additionally, an up-to-date literature review of the EPF field is included. Following, in Chapter 3 we describe our data and our entire modeling approach from data processing to model implementation. The chapter finishes with an explanation of how we will be evaluating our models. The main results of the ML models are presented and analyzed in Chapter 4 where we also quantify the importance of European electricity markets. Lastly, Chapter 5 concludes the thesis with a discussion of our main findings and our approach. We additionally provide a suggestion on what direction future works should be headed.

Chapter 2

Background and Related Work

In this chapter, the most relevant theoretical background will be covered. Starting with the Norwegian electricity market, we will walk through the process of how the market is structured, how prices are settled and what characterizes a well-functioning market. Following that is an overview of the main concepts linked to neural networks that are advantageous to be familiar with. Wrapping up the background section is a thorough explanation of how LSTMs are capable of learning from long-time series data. The chapter finishes with a section detailing the most relevant related works and current benchmark solutions in regard to EPF.

2.1 The Norwegian Electricity Market

As power markets are inherently different based on their location, regulations and sheer size, providing a general overview of power markets would be vague. We have therefore decided to limit the background of this thesis to the Norwegian power market which consists of unique characteristics while at the same time being representative of a typical European power market. Additionally, the Norwegian power market is an important market for the Nordic countries.

The Norwegian power market, as mentioned in Chapter 1, was liberalized with the introduction of the Energy Act of 1990 (Bye & Hope, 2005). A monopolistic operation of the power market, where power producers set the price for customers, was no longer optimal for the current time which lead to the power market becoming more accessible for several market participants. Competition between power suppliers meant that Norwegian consumers had the option to choose their power supplier based on the ones with the most competitive prices. Increasing competition is vital, regardless of sector, in order to achieve lower prices and higher quality of services (Council of Economic Advisers, 2016). With the deregulation of the power market, electrical power went from being directly decided by power producers to being determined by supply and demand.

Supply and demand are integral in order to understand how the power market works. In Norway and several other connecting countries, the power market is structured such that electricity will always be transported to areas with greater demand for electricity. In other words, power will flow to where its value is the greatest, which is enabled with the electricity grid and interconnectors (Ministry of Petroleum and Energy, 2016). For instance, if two regions in Norway are connected and one of them requires a lot more power than the other region, then the power demanding region will generally have higher power prices. Additionally, if the region with lower demand has a surplus of power, that power will flow to the other region as long as there is demand and the electricity grid is capable of transporting the power. This means that in regions with higher demand, prices will generally be higher, and regions with lower demand will have the opposite. However, this is strongly dependent on whether the regions have a surplus of power. If that is the case, then demand can still be high, but prices can also be low. We now enter an important principle in how prices are determined. As long as there is surplus and access to power, prices can be kept low. It is only when there is deficit of power that electricity prices will increase. In well-functioning markets, when a commodity's demand increases, its price will likewise increase.

The above-mentioned example of two regions in Norway is actually a realistic representation of how power is distributed in the country. As Norway is an elongated country, its nature is vastly varying from north to south. Since Norway's main power source comes from hydropower plants, that again are weather dependant, its power system also becomes weather dependant. The weather is often very different based on where you are located. As a consequence of the varying landscape, weather and transmission constraints, Norway is divided into five electricity price regions (NO1-NO5) where three of which are located in the south and NO1 is the main region which includes Oslo. In Figure 2.1 the five bidding zones (BZNs) are depicted. Since production and consumption happen at different places in the country, and the electricity grid has limited transmission capacity, areas with surpluses and deficits of electricity may arise (Statnett, 2021). More specifically, NO3 and NO4, the northernmost regions, are the BZNs in Norway with a surplus of electric power. Consequentially, these two BZNs have historically enjoyed lower electricity prices than the BZNs further south. This is especially true at the current time.

The reasoning for the lower electricity prices in NO3 and NO4 can be explained by several factors, many of which we already explained at the start of this chapter. First of all, and most importantly, is the low demand in comparison to the southbound regions. Secondly, many of the largest hydropower plants are located in the north, in order to fully take advantage of the steep landscape which is suitable for creating dams. The larger the height difference from the dam down to the turbine, the more electricity can be generated. Moreover, the limited transmission capacity in the electricity grid plays an important part. There are no continuous 420 kV transmission lines ranging across the country. 420 kV transmission lines are the largest and can carry the most amount of power in the entire electricity grid. They are capable of transferring large amounts of energy,

Figure 2.1: Overview of the five BZNs in Norway. Image taken from (Statnett, 2021).

but as can be seen in Figure 2.2, there is a break in the continuity near the NO5 and NO3 border. This works as a bottleneck given that there is no free flow of electricity from the northern regions to the south via the 420 kV transmission lines. However, this one missing line is not the sole reason for the price difference in the north and south. The total amount of transmission lines is not sufficient at the current moment. Compared to Sweden, Norway's transmission grid is clearly insufficient. If there was one BZN throughout the entire country, the price would have been the same everywhere, but this would require a massive upgrade of the electricity grid in order for electricity to freely flow across the country (Statnett, 2021).

It is worth mentioning that once power has been produced by power producers and supplied to the electricity grid, there is no way to continuously track that specific amount of electricity while it flows through the grid (Norwegian Ministry of Petroleum and Energy, 2021). In other words, electricity is indistinguishable. When industries and people consume electricity provided to them by their power supplier, they cannot know where that electricity was produced or how far it has traveled. Power producers are therefore paid for the amount they produce, while consumers are billed the amount they consume.

2.1.1 The Wholesale Market

The Nord Pool[1] power exchange was established in 1996 following the liberalization of the energy sector in the Nordic countries (Ministry of Petroleum and Energy, 2016). The power exchange, which was the first to open for cross-country power trade, is the leading power exchange in the Nordic region. Additionally, it is integrated with the European power

[1]https://www.nordpoolgroup.com/en/

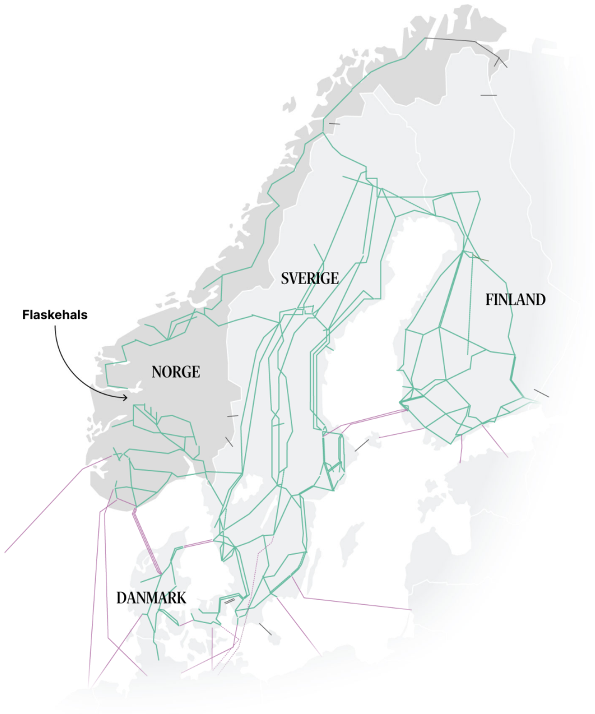


Figure 2.2: The transmission grid of the Nordic countries. The green lines represent 420 kV transmission lines. Both the black and pink lines are interconnectors. Image taken from (VG, 2023).

market via cross-border physical interconnectors and financial market integration. In Norway, there are interconnectors connecting the country to the Nordic countries and more recently, new interconnectors to Germany, United Kingdom and the Netherlands have been built and are operational.

As with most major decisions, there are both positives and negatives of trading power between countries. As previously highlighted, the Norwegian power system is greatly hydropower plant oriented, which makes it weather dependant. In Norway's neighboring countries, the reality is vastly different. Denmark, being a relatively flat country, is not suitable for hydropower plants, but benefits greatly from wind power. In Sweden and Finland, the usage of thermal power plants is popular. These differences enable countries coupled together to take advantage of several sources of power. When it is windy in Denmark and wind farms are producing electricity close to free of charge, it is often advantageous for neighboring countries to import excess electricity. Similarly, when the snow melting season begins in Norway, and the water level in the dams rise, it is beneficial to export excess electricity (The Norwegian Energy Regulatory Authority, 2023). The higher the water level in the dams, the higher the possible supply is. Besides, by importing cheap electricity, the power producers in Norway are able to save their water for later use, when its value may be greater. This principle is known as the water value and refers to the alternative value of the water when used at a later point.

To summarize, having interconnectors between the countries leads to power being consumed where it is most needed, provided the transmission capacity is sufficient, thus ensuring optimal use of resources and capacity. Additionally, it enables societies to benefit from security of supply at overall lower costs. However, the major downside is how prices in exporting regions can become higher, given that the supply is flowing elsewhere.

The wholesale electricity market is divided into two major parts that complement each other. First of all, the workhorse of the European power market, known as the day-ahead market, works as an auction between market participants. Since power production and consumption need to be in balance, the day-ahead market is available every day where participants can trade electricity for the following day. The day-ahead market is available in several exchanges and in Europe, Nord Pool and EPEX Spot[2] are the most popular. In the Nord Pool power exchange, electricity is traded every day between 08:00 and 12:00 CET. During this time window, power producers enter how much power they can produce for each hour of the next day, and at what price. On the other hand, power suppliers, that often trade electricity for everyday people, make bids by highlighting how much electricity they need every hour for the next day.

Once the auction ends, the electricity price for each hour of the following day is calculated using sophisticated algorithms that balance the available offers with current bids. All the European power exchanges are coupled together via the Single Day-Ahead Coupling (SDAC) to ensure that all the BZNs have a specific market clearing price (MCP) (Kühling et al., 2021).

The MCP is important to understand as it refers to the price where demand and supply are at an equilibrium. The 24 hourly prices obtained for the day-ahead market follow a standard supply-and-demand curve and the MCP is denoted as the intersection between these two curves. The cost of producing electricity greatly depends on the power source. The variable costs of producing electricity are lowest with renewable energy sources, while fossil fuels are the most expensive. The difference in variable costs means that power producers offer their electricity at different prices, in order to make a profit. Renewable power producers can offer their electricity cheaply and still make a profit, while fossil fuel producers need to put a higher price on their electricity to make a profit. While prices for the day-ahead market are being calculated, demand is matched with supply by always choosing the cheapest power source available until all demand is met. For instance, say we would like to buy 100 shares of Apple. Accidentally, we place our bid at double the price of a single share. Luckily, mechanisms would prevent us from paying double the price for our shares. In reality, our bids would be matched with the lowest asks until our order is fulfilled, or no shares are left to sell. In this example, the order book, a list of buy and sell orders, was "eaten up" using a bottom-up approach. This exact principle is used when prices are settled in the day-ahead market.

[2]https://www.epexspot.com/en

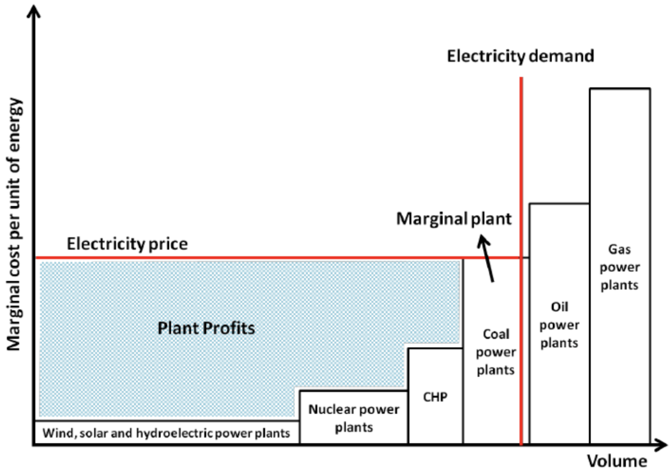


Figure 2.3: A representation of the merit order curve with demand (power volume required) along the x-axis and the marginal cost for producing power along the y-axis. Image taken from (Bahar & Sauvage, 2013).

The cheapest energy sources will always be used until demand is met. As we highlighted, renewable energy sources are offered at the cheapest price and will therefore always be taken into account at the start. Further on, if all demand is not met with the available renewable sources, other power sources, which are offered at higher prices, need to be taken into calculation. This continues until all demand is met and at that exact point, the MCP is determined. Therefore, the MCP can be referred to as the most expensive power source offered in order to serve demand. Every power producer whose offering price is below or equal to the MCP will supply electricity for that specific hour (Yan & Chowdhury, 2015). Figure 2.3 illustrates how the MCP is determined, which is also known as the merit-order principle (Trebbien et al., 2023). Additionally, the large profits for renewable power producers are further highlighted.

In order to ensure fairness in the markets, every power producer whose price was offered at a lower rate than the MCP will be paid the MCP (Yan & Chowdhury, 2015). Simply put, all power producers, regardless of the price they offer, will be paid the same if their power is used to serve demand. Renewable power producers are therefore percentage-wise the most profitable. As the day-ahead market allows you to trade power for each hour of the following day, the MCP is therefore determined each hour of a day.

To calculate how much electricity is needed for the following day's 24 hours can be cumbersome. Sudden increases in demand and likewise, sudden decreases in expected supply may occasionally arise. There are several situations in which supply may be inadequate. First of all, power plants may experience trouble producing the required amount of electricity for numerous reasons. There may be sudden changes in weather forecasts, or the transmission grid may become defective at places. Lastly, social events that require more electricity may also happen. In these types of situations, it is necessary to quickly have access to more electricity. The intraday market helps secure the balance between supply and demand by allowing participants to trade electricity close to delivery (Kühling et al., 2021). The intraday market is only available after the MCP has been calculated after the day-ahead auction closes. It offers physical electricity to be delivered at 15 minutes, 30 minutes or one hour after it is being traded. Flexibility is achieved with the intraday market, but the volumes traded are far lower than those of the day-ahead market. As a consequence, the prices in the intraday market are closely linked to the hourly prices calculated in the day-ahead market.

As previously mentioned, the Norwegian wholesale market is integrated with the Nordic market, which again is integrated with the greater European electricity market. Furthermore, all power exchanges in Europe are coupled together to create a single pan-European cross-zonal day-ahead electricity market (ENTSO-E, 2022). SDAC ensures that all power exchanges offer the same day-ahead market clearing prices for all the BZNs integrated into the European electricity market. Furthermore, an integrated day-ahead market leads to a more efficient and more competitive market, while also increasing overall liquidity. Moreover, SDAC utilizes different power sources across country borders more effectively (ENTSO-E, 2022).

The MCPs of the different BZNs are calculated with a common price coupling algorithm, EUPHEMIA (NEMO Committee, 2020). The sophisticated algorithm is the main contributor to achieving a balance between supply and demand and takes into account several key features related to price settlement. For instance, grid constraints and capacities set by different transmission system operators (TSOs) such as Statnett, and all the bids and offers during the day-ahead auction are used to calculate the joint MCP.

2.2 Main Concepts about Neural Networks

Intelligence, and more specifically human intelligence, allows us to think, learn and react to situations. It is what makes us capable of solving problems and learning from experience. Artificial intelligence (AI), one of the newer fields combining science and engineering, aims to quantify human behavior and thought processes by building intelligent entities. One of the main tools used to create these intelligent entities is artificial neural networks (ANNs).

The concepts of ANNs were introduced in the 1940s (McCulloch & Pitts, 1943), but the first practical implementation happened during the late 1950s when Rosenblatt introduced the well-known perceptron (Rosenblatt, 1958). Neural networks were introduced as tools to better understand and model complex data, but given the limited resources available at the time, such as data storage and processing power, the field of AI did not get off to a flying start. It was not until the mid-2000s that AI experienced a major boost in popularity across sectors, mostly because of the technological

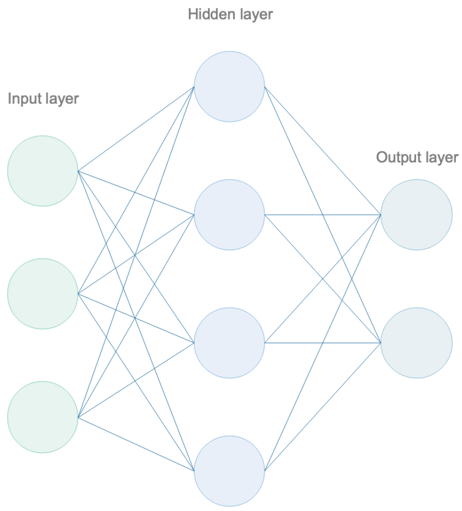


Figure 2.4: A typical architecture of a basic neural network consisting of one hidden layer.

advancements in computational power, data storage and the introduction of deep learning (LeCun et al., 2015).

ANNs replicate how the human brain is able to process information. In the same way as the human brain is built of neurons connected with each other via synapses, ANNs have a similar architecture. In Figure 2.4 the architecture of a basic neural network is depicted. It consists of different neurons connected together in different layers. Neural networks always start with an input layer and end with an output layer. All layers in between these two layers are called hidden layers and different transformations are often applied to these.

The neurons of a neural network communicate with each other via their connections. Firstly, each neuron is just a real number that can activate if the linear combination of its inputs is greater than a threshold value. If this is the case, the neuron can send information to nodes it is connected to, which is how information flows through neural networks (Russell & Norvig, 2009). Moreover, the strength of the connections, known as edges, is dependant on each edge's associated weight. The weights between the nodes either increase or decrease the strength of a signal, which plays a crucial part when neural networks are trained for specific tasks. Additionally, before a signal is sent to a connected node, the linear combination of a neuron's input is sent through an activation function. Activation functions are what differentiates ANNs from regular linear regression models. Without activation functions, each node in ANNs could be seen as a single linear regression model and thus, the neural networks would become giant linear regression models. Consequently, activation functions are what allow ANNs to become powerful tools for understanding complex data and if needed, forecasting it. Activation functions introduce the concept of non-linearity which is of the utmost importance given that most data cannot be modeled linearly. On top of this, most of the relationships in the real world, whether they are in finance, astronomy or biology, are non-linear. Since ANNs are able to capture these relationships, they become useful for modeling real-world problems.

As stated, AI and ANNs aim to mimic how the brain functions, but also how it is able to learn from experience. ANNs learn by repeating similar tasks just like humans would. In simple terms, input data is fed into ANNs. The data then flows through the entire network via the edges of the nodes with linear combinations and activation functions applied throughout the process. The output of the network is then measured against ground truth data which equals the network's error, also known as loss. This type of ML, where models learn by comparing predictions against actual output is known as supervised learning. Several methods for calculating the loss have been proposed, but the most popular ones are Mean Square Error (MSE) and Cross-Entropy loss (Christiansen et al., 2014). Once the loss has been calculated, it is time to tune the edges between nodes by either increasing or decreasing their weight. This process is known as backpropagation and eventually leads to lower loss values, which again leads to more accurate forecasts. In other words, ANNs are able to learn from experience and errors they make by simply changing how connections between nodes are structured. Loss functions are minimized using gradient descent optimization algorithms (Baldi, 1995). It is optimization algorithms such as gradient decent, that minimize the difference between the predicted output and the actual output. This is achieved by iteratively computing the gradient of the loss function with respect to the model's weights and biases. Simply put, in ML gradients refer to the partial derivatives of the loss function pointing in the direction of maximum increase for a given function. Since we want to minimize the difference between predicted values and actual values, the model's weight and biases have to be adjusted in the direction of the negative gradient. This process is repeated until the loss function converges, leading to the optimal weights and biases.

2.2.1 Neural Network Architectures

ANNs come in various architectures ranging from shallow to deep and comprehensive. However, even with all the changes in architecture, ANNs can primarily be divided into two main categories. Firstly, there are the feed-forward networks where the information only flows one way. Simply put, all edges point in the same direction which leaves us with a directed acyclic graph. No loops are therefore present in the network, and all nodes receive inputs from nodes in the preceding layers (Russell & Norvig, 2009). These types of networks are usually effective at processing non-time-related data because they do not have an internal state able to process previous outputs.

On the other hand, the second architecture, recurrent neural networks (RNNs), contain feedback loops throughout the network. This means that the flow of information not only flows forward, but it can flow to the same node several times. Since RNNs feed their outputs back into their inputs, and decisions are based on the current and previous inputs, these types of networks are able to capture different states. As a consequence, RNNs are able to exhibit temporal behaviors, unlike feed-forward networks. This makes them valuable for solving tasks where the data is sequential (Sherstinsky, 2020). RNNs are therefore widely applied in natural language processing tasks, such as next-word prediction and sentence generation, but also in time series problems. Generally, if features of a dataset are time dependant and follow each other in a particular order, RNNs can be useful for solving a vast number of problems. Given that RNNs have a temporal behavior, they inherently have a kind of memory, which again mimics the human brain.

Even though RNNs are able to make decisions based on previous inputs, they are highly prone to the vanishing and exploding gradient problem when time steps become too large (Sherstinsky, 2020). The more times an RNN is unfolded, the more unstable it becomes. During training, when the aim is to minimize the loss function by calculating the gradients of the parameters, the networks tend to become unstable because the gradients either become too small or too large. As RNNs leverage backpropagation through time and weights are shared within layers of the network, gradients will only become smaller or larger as time passes (Werbos, 1990). At each time step the states from previous steps will continuously become less significant with regard to the gradients. In simpler terms, RNNs suffer from short-term memory where information at the start of a sequence often is disregarded. There comes a point where the gradients propagating through the network become either too small or too large and the network is then unable to properly learn. In the case of vanishing gradients, gradients become so small that the weights between nodes are barely changed. Consequentially, the loss function converges very slowly which leads to slow model training. In the other case, gradients become very large and explode, resulting in weights that end up containing NaN-values (Not-a-Number). The loss function will most likely never converge to a global maximum because the steps taken to minimize the loss function become too large. These issues are more or less the sole reasons why RNNs are not as widely applied in the current day of time for complex sequential data analysis.

2.2.2 Long Short-Term Memory

Luckily, research aimed towards solving the issue of vanishing or exploding gradients in RNNs has been conducted with very promising findings. A different type of RNN was developed in 1997 to improve and minimize the problems linked with vanilla RNNs. The solution was a novel neural network named Long Short-Term Memory (LSTM) which is capable of storing long-term dependencies by the introduction of constant error carousels (Hochreiter & Schmidhuber, 1997). LSTMs were mainly introduced to solve the issue of short-term memory for vanilla RNNs and these issues were solved by internal mechanisms called gates that can regulate the flow of

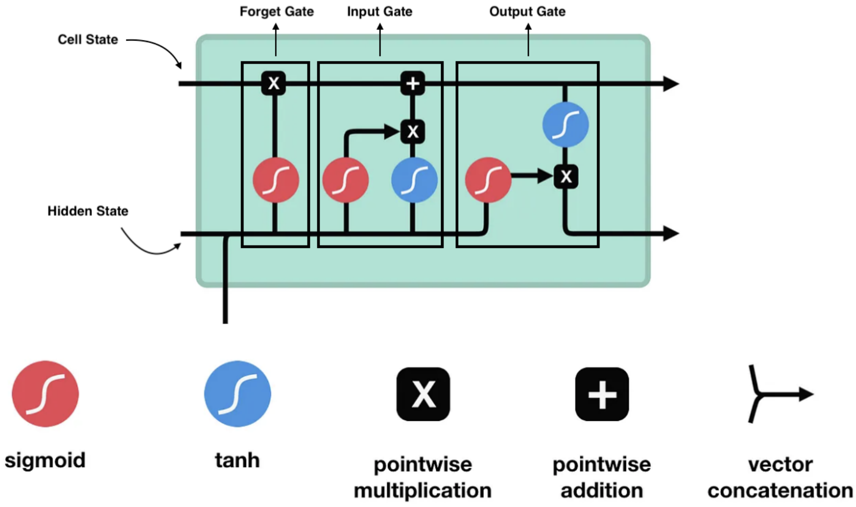


Figure 2.5: The architecture of an LSTM cell with its operators. Image taken from (Phi, 2018). Edits to original image were done.

information through a network. The gates enable LSTMs to process long sequences of data and decide what data is important to keep and what can be disregarded.

In Figure 2.5 an LSTM cell with all its internal operators is depicted. A thorough walkthrough of all its operators is key in order to grasp how LSTMs function and why they are reliable tools for time series prediction. First of all, as stated, the core concepts of LSTMs are their internal gates and cell states. The cell state, which is represented as the top line in Figure 2.5, works as the LSTM's memory. The cell state is able to keep relevant information from a sequence of data and can also be referred to as long-term memory. The short-term memory represents the lower line in the figure, which is known as the hidden state. Under ways during model training, new information is added or removed from the cell state via the other gates based on certain criteria. As can be seen from the figure, LSTM cells consist of three gates which are the forget gate, input gate and output gate. Each one of these gates is a different neural network that serves the purpose of deciding what information is allowed in the cell state (Phi, 2018).

Starting with the forget gate, its main purpose is to decide what information should be kept or discarded. The gate takes in two inputs which are information from the current input and the information from the previous hidden state (Staudemeyer & Morris, 2019). To simplify, the sum of the current input and the short-term memory is calculated and sent through a sigmoid activation function. The sigmoid function shrinks its inputs between the range of zero and one. After the activation function has been applied, the output is then multiplied with the current long-term memory. Since the sigmoid function shrinks its inputs between zero and one and is then multiplied with the cell state, one can think of the entire forget gate as the percentage of long-term memory to remember. As per

Equation 2.1, showcasing the sigmoid function, large positive values will be closer to one, whereas large negative values will be closer to zero. For instance, if the sigmoid function outputs the value of 0.60, 60% of the long-term memory will be kept, as 0.60 is multiplied with the current value of the long-term memory. On the opposite side, if the output of the activation function is zero, zero percent of the long-term memory will be kept.

$$S(x) = \frac{1}{1 + e^{-x}} \tag{2.1}$$

Once the percentage of long-term memory to remember has been calculated, it is time to move over to the input gate which is made up of two different neural networks. The rightmost network in Figure 2.5 calculates the potential long-term memory by combining the current input with the short-term memory which is sent through a tanh activation function which squishes the input between the values negative one and one. The other network also combines and multiplies the input and the short-term memory by their respective weights. Its main function is to decide what percentage of the potential memory to remember by the usage of the sigmoid activation function (Starmer, 2022). Once the potential long-term memory has been multiplied with the percentage of potential long-term memory to remember, its value is added to the cell state. Overall, the input gate determines how one should update the cell state or the long-term memory.

Finally, the last step of the LSTM updates the short-term memory. Just like the input gate, the output gate consists of two neural networks. However, this time the neural network with the tanh activation function calculates the potential short-term memory, while the other with the sigmoid activation function calculates the percentage of the potential short-term memory to remember (Starmer, 2022). The outputs of the two neural networks are multiplied together and represent the new short-term memory. Additionally, this is also the output of the entire LSTM. For longer sequences of data, the output of an LSTM cell will be the input for another cell which is known as the next time step.

To wrap it all up, LSTMs are great tools for capturing temporal dimensions which makes them suitable for data with strong memory. Furthermore, their ability to handle N-dimensional time series, makes them a valuable tool for multivariate time series forecasting.

2.3 Related Work

Electricity price forecasting (EPF) dates back to the early 2000s, a few years after the liberation of the European power market (Bunn, 2000; Nogales et al., 2002; Szkuta et al., 1999). During the 15-year period from 2000 to 2014, the field grew steadily with an increasing number of research papers being published each year (Weron, 2014). From 2014 to today's date, EPF has become an integral part of ensuring stability, efficiency and predictability for consumers, companies and power producers worldwide. A literature review of the entire EPF field is out of the scope of this thesis. The field of EPF contains vastly different methods ranging from short-term to long-term forecasting. Given that EPF has to be tailored towards specific markets and that this thesis's main purpose is to predict day-ahead prices, the literature review will be limited to forecasting day-ahead prices in Europe.

The field of EPF can generally be divided into three different categories with respect to the methodology used. These are statistical, deep learning and hybrid methods (Lago et al., 2021). Starting off with the statistical ones, these are mainly linear regression models that represent the output variable as a linear combination of its input features. Typically, the model architecture ranges from univariate to multivariate, where the latter is more applied and generally the one with the most accurate predictions (Lago et al., 2021). However, in (Ziel & Weron, 2018), the authors conduct an empirical study that tries to address whether the optimal structure from EPF is univariate or multivariate. 58 models were compared on 12 datasets across Europe. These were compared without the use of exogenous variables, such as weather conditions, fuel prices and renewable energy production. Exogenous variables simply refers to external factors that influence electricity prices. Although multivariate architectures ever so slightly averaged more accurate predictions, the authors concluded that multivariate architectures did not uniformly outperform its counterpart.

Auto-regressive linear regression models where only past data is used to predict the 24-hourly prices in the day-ahead market have additionally gained popularity among the statistical approaches. More specifically, the well-known ARIMA model has found its way to the field of EPF. Although not the first to implement ARIMA to forecast day-ahead prices, in (Jakaša et al., 2011), the auto-regressive model is applied to the German electricity market during the 2001–2011 period with satisfactory results.

According to (Lago et al., 2021), the introduction of linear regression models with a large number of features that utilize regularization techniques has been a great success in statistical EPF. If the number of features is large, promising results have been achieved by incorporating feature selection methods such as the least absolute shrinkage and selection operator (LASSO). The authors of (Ziel, 2016) also consider an auto-regressive model but include LASSO in order to capture intraday dependencies such as the time-varying cross-hour dependencies. Their model is additionally able to explain a great amount of variety in the data which again helps explain intraday behavior of electricity prices.

Moving over, the deep learning (DL) and machine learning (ML) approaches, which are more in line with what is to be expected of this thesis, have seen increased research interest since 2016. During the period 2017 to 2019, a total of 28 research papers where ML methods were applied, were published (Lago et al., 2021). Amongst these were (Wang et al., 2017) which was also the first paper published using deep learning. However, papers using shallow neural networks had already been published such as (Voronin & Partanen, 2013). In the first DL paper, the authors presented a novel deep learning architecture using stacked denoising autoencoders (SDA) to predict the day-ahead market with data from Nebraska, Arkansas, Louisiana, Texas, and Indiana hubs in the U.S. consisting of electricity prices, observed load and forecast load. Their dataset covers the period from January 2012 to November 2014, where the last three months were used to evaluate their model. Additionally, four other data-driven approaches such as shallow neural networks, support vector machines, multivariate adaptive regression splines and LASSO were used for comparison in order to validate their results. Furthermore, the performance of their approach is also validated on four metrics and the authors state that their model could be used to accurately predict electricity prices. However, further research aimed towards optimizing the structure of their model should be conducted.

In 2018, a comprehensive comparison of traditional algorithms was conducted in (Lago et al., 2018). It was the first paper to create a large-scale benchmark of new and existing models. The paper's main purpose was to provide a framework for EPF. The applicability of DL in EPF was still in its early days with limited literature. Therefore, to fill this gap, four DL models were proposed and compared to 27 common approaches consisting of both statistical and ML approaches. Their four models were a two-layered deep neural network (DNN), an LSTM, a Gated Recurrent Unit (GRU) and lastly, a convolutional neural network (CNN). These were all applied to the EPEX-Belgium market and the DNN, the LSTM and the GRU outperformed the other 27 approaches in a statistically significant matter. Additionally, the DNN outperformed the other DL approaches which are hypothesized to be because of the low amount of data. Furthermore, the consensus that deep learning approaches are more accurate than statistical ones, was first stated in the mentioned paper. In spite of that, the statement cannot be generalized given that the researchers only tested their models on one dataset.

Following the first benchmark paper was (Ugurlu et al., 2018), who performed EPF of the Turkish day-ahead market using data from January 2013 to December 2016. The authors also applied an RNN, however, it differs from (Lago et al., 2018) in the number of features utilized and that they proposed deep RNNs in contrast to shallow RNNs presented in (Lago et al., 2018). Their models were a novel multi-layered GRU and a multi-layered LSTM. Like previous papers, their models were tested and compared against both statistical and ML approaches. (Lago et al., 2018) stated that deep learning approaches generally were more accurate, and the results achieved in this paper are in line with that statement. Moreover, their GRU and LSTM both achieved lower errors than DNN which contradicts the findings of (Lago et al., 2018). An explanation of the contradiction may be linked to the market specifics, where each market has its own characteristics.

LSTMs used for EPF continued to gain interest in 2019 with (Zihan Chang & Chen, 2019) where a novel hybrid model based on wavelet transform and Adam-optimized LSTM (WT-Adam-LSTM) was proposed. Wavelet transform was used to decompose electricity prices series into a set of better-performing constitutive series. In other words, a more stable variance in the data was achieved by wavelet transform. Their model was trained and applied to the French electricity market, as well as the New South Wales market in Australia. Four cases, to validate their proposed model, were studied. Among these, the authors found that the Adam optimizer outperforms similar optimizing algorithms when it comes to EPF. Additionally, their hybrid model was compared to other hybrid statistical ARIMA models which were significantly outperformed on both datasets. It is concluded that their hybrid model exhibits better performance than that of existing models.

In 2021, the second major review paper of the EPF field (Lago et al., 2021), covering the period from the first review paper (Weron, 2014), was published. It is a comprehensive review of the current state-of-the-art algorithms, best practices and also includes an open-access benchmark. Their contributions are threefold. First, they analyzed the existing literature and selected two models which arguably could be considered state-of-the-art. These were the Lasso Estimated AutoRegressive (LEAR) model introduced in (Uniejewski et al., 2016), and the DNN from (Lago et al., 2018). These models were made available as part of an open-source python library, EPFTOOLBOX[3]. Second, five open-access benchmark datasets from markets across the globe, spanning six years each, were presented. These include markets such as Nord Pool, the French (EPEX-FR), Belgian (EPEX-BE), German (EPEX-DE) and the Pennsylvania–New Jersey–Maryland (PJM) market in the U.S. They consist of recent market data in order to include the effects of integrating renewable energy sources to power markets. Lastly, best practice guidelines for EPF were provided such that new research studies could become more reproducible and sound. A thorough discussion of which evaluation metrics work best and why, was also presented such that model evaluations in future works could become easier to perform. Moreover, the review paper highlights the importance of including ensembles in the context of EPF, further validating the findings of (Nowotarski et al., 2014), who stated that a combination of models trained on different calibration window lengths, led to more reliable and accurate forecasts.

2.3.1 Electricity Price Forecasting using XAI

During the last few years, increased efforts toward understanding ML models' outputs have been made (Tjoa & Guan, 2021). As ML approaches inherently are black-box models, scientific insights become limited. Understanding why ML models make the predictions they make has received significant attention in automated decision-making applications. In energy science, understanding which features drive electricity prices the most has always been a research question of interest. More recently, since the start of the European energy crisis, this research question has been further solidified. Thus far, ML has mainly been used to strictly forecast electricity prices, however, attempts at model explanations have gained increased research interest in EPF (Machlev et al., 2022).

EPF of the European electricity market was performed in (Tschora et al.,

[3]https://github.com/jeslago/epftoolbox

2022). More specifically, three datasets from the benchmark EPFTOOLBOX, created by (Lago et al., 2021) were used: France (EPEX-FR), Belgium (EPEX-BE) and Germany (EPEX-DE). Furthermore, they extended the benchmark by considering and incorporating unused predictive features such as price histories of neighboring countries and gas prices. Moreover, attempts at making their models more generalized were made by including recent data, from ENTSO-E Transparency Platform[4], such as the COVID-19 period and data covering the European energy crisis. Additionally, four ML approaches, support vector regressors (SVRs), random forest regressors (RFRs), DNNs and CNNs were applied to their datasets and compared against the two state-of-the-art models, LEAR and DNN from (Lago et al., 2018). Lastly, SHapley Additive exPlanations (SHAP) were used in their analysis to assess the importance of features in the prediction process. The paper marks the first that actively attempted to achieve accurate forecasts and provide explanations of model predictions. Based on their experiments and results, two key takeaways stand out. First, including more features in the datasets significantly improved model performance. However, based on SHAP values, feature contributions were strongly market specific, which makes perfect sense as the three considered markets are inherently different. Second, the authors conclude that SVRs and DNNs extract the most meaningful information which further verifies earlier statements.

In (Trebbien et al., 2023), an XAI ML model tailored to the German day-ahead market is presented. Just like (Tschora et al., 2022), SHAP values were used to disentangle the main drivers for the electricity prices. Simply put, SHAP values showed which features lead to higher or lower electricity prices. Their dataset was directly collected from the ENTSO-E Transparency Platform and includes data from the years 2017 to 2019. Additionally, power system features, such as day-ahead forecasts of load, solar generation, wind generation, the day-ahead total generation and imports and exports were collected. Furthermore, fuel prices such as oil and natural gas prices were included. The authors' model was a gradient-boosted tree (GBT) which achieved substantially more precise predictions of the electricity price than those of a benchmark model based on the merit order principle. Their model was able to explain 80% of the variability in the price time series, relative to the merit order approach which was able to explain 66%. To conclude, their analysis confirms that higher load leads to higher prices, whereas higher generation of power from renewable energy sources, such as wind or solar, leads to lower prices. Residual load was the main contributor to the electricity price according to their model which reflects reality. Lastly, export and import between neighboring countries was the fifth most important feature in regard to the price.

[4]https://transparency.entsoe.eu/

Chapter 3

Data and Methodology

This chapter covers all data-related topics ranging from dataset description to data processing. The main assumptions and techniques applied are highlighted and thoroughly justified. Thereafter, our methodology is presented which includes implementation details of our ML model. The chapter closes with an introduction to the evaluation metrics we will use.

3.1 Data Description

The dataset used for this thesis was directly collected from the ENTSO-E Transparency Platform[1], which makes pan-European electricity prices available at all time frames. The dataset consists of hourly electricity prices from day-ahead markets from BZNs spread throughout Europe. The price series included in the dataset dates back to the 12th of December 2014 and includes all dates up to the 1st of January 2023.

In total, the original dataset consists of 57 different BZNs, thus leading to 57 different numerical features. Furthermore, given that hourly entries are included, the size of the dataset quickly becomes comprehensive. The total amount of rows in the dataset is equal to 70633 which again amounts to 4096714 entries. Even though the dataset consists of 57 different features, they all represent the same entity, which of course is the electricity prices. In order to keep the dataset uncomplicated, all prices are continuous and given in €/MWh and not in each country's local currency, which is consistent with the unit of measurement used in EPF research. Given that the dataset covers the period between 2014 to 2023 and the surge in electricity prices was apparent in late 2021, there is a predominance of lower electricity prices which skews the distributions of the data. Figure 3.1 provides a clear visualization of how prices prior to late 2021 have been stably low.

The dataset is free of measurement uncertainties as it represents the agreed market prices of electricity in the European exchange markets. Nevertheless, there are still issues that need to be addressed and taken into account in order to achieve reliable ML model predictions.

[1] https://newtransparency.entsoe.eu/

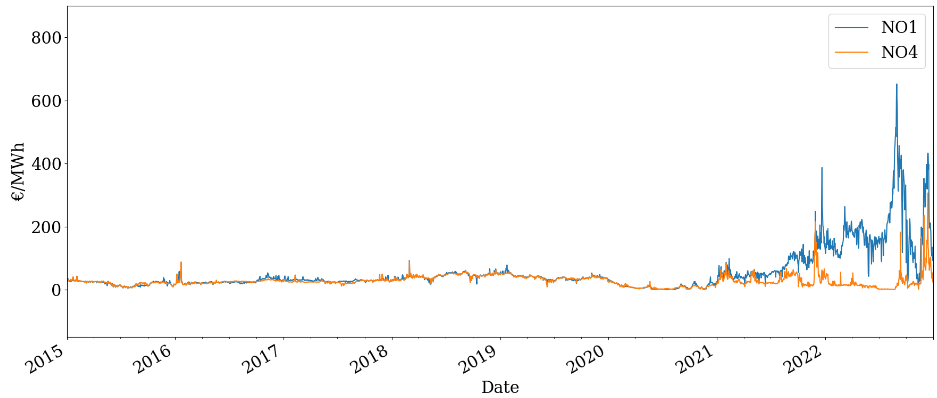


Figure 3.1: Electricity prices of the NO1 and NO4 Norwegian BZNs. See Figure 2.1 for their exact geographical locations.

3.1.1 Challenges with the Dataset

One of the main challenges with the dataset was the sheer number of missing (NaN) values spread across most of the features. Out of the 57 features, only one was perfectly represented, covering all hours from the 12th of December 2014 to the 1st of January 2023. Numerous tools and proposals for handling NaN values have already been researched and presented (Emmanuel et al., 2021). Finding the most optimal technique to handle NaN values in the dataset is a comprehensive research question in and of itself, which was not the aim of this thesis. In total, 20% of the data, or 849638 entries, were invalid NaN values, where the longest continuous sequence was 70609 rows long. The missing data is to be excepted as there are BZNs that have only been introduced recently. This naturally poses a complicated problem in designing ML models that rely on time-ordered data in equivalent time ranges. Most ML models, including LSTMs, are not capable of processing these values.

The NaN values were spread and occurred across the entire dataset as can be seen in Figure 3.4. The blue lines indicate NaN values, whereas the yellow color represents actual values. One can clearly see how most of the features contain NaN values, especially the Italian price regions. Moreover, a given time series may include several sequences of NaN values and not only one. The easiest option would have been to drop all dates consisting of at least one NaN value, however this would of course leave us with only one valid feature, as previously mentioned. Dropping these NaN values was therefore not an option.

Additionally, as the amount of high electricity prices is limited in comparison to ordinary prices in the dataset, ML models may struggle to produce reliable and consistent predictions. However, the price imbalance is necessary as we are trying to predict electricity prices in unprecedented times. Electricity prices have historically been in the 0.30 to 0.40 kroner per kWh range in Norway, which equates to 26.40 to 35.20 €/MWh using today's currency rate. These price levels have historically been considered relatively low. In order to reflect reality, we consider prices above 61.60 €/MWh (0.70 kr/kWh) to be extraordinarily high. This threshold is additionally consistent with the Norwegian electricity subsidy scheme that was introduced by the Norwegian government in order to compensate households and the agricultural sector against high electricity bills.

3.1.2 Data Analysis

As this thesis is invested in how electricity prices are coupled, especially during the current European energy crisis, a simple correlation plot provides useful information. In Figure 3.2 a historical plot from 2015 to the end of 2022, showing the four-month rolling Pearson correlation r of NO1 against NO2, NO4 and Germany (DE), is provided. A distinct difference can be spotted in the correlations from 2021 onwards. In the case of NO2 the difference is minimal which makes sense, as NO2 has always been closely coupled with NO1. On the other side, the relationship between NO1 and NO4 and DE from 2021 are opposite. Whereas NO4 historically has been relatively correlated with NO1, the relationship between the BZNs was significantly reduced around the same time the European energy crises began and prices quickly rose. The two northern BZNs of Norway have been relatively unaffected by the ongoing crisis throughout Europe, which explains the sudden decrease in correlation with NO1. The effect was the opposite for the German BZN. After the energy crisis, the German BZN became closely coupled with NO1, which is also a phenomenon with other BZNs affected by the energy crisis.

In Figure 3.3 the kernel density estimations of two time periods is presented. Each line represents the five-month rolling correlations of NO1

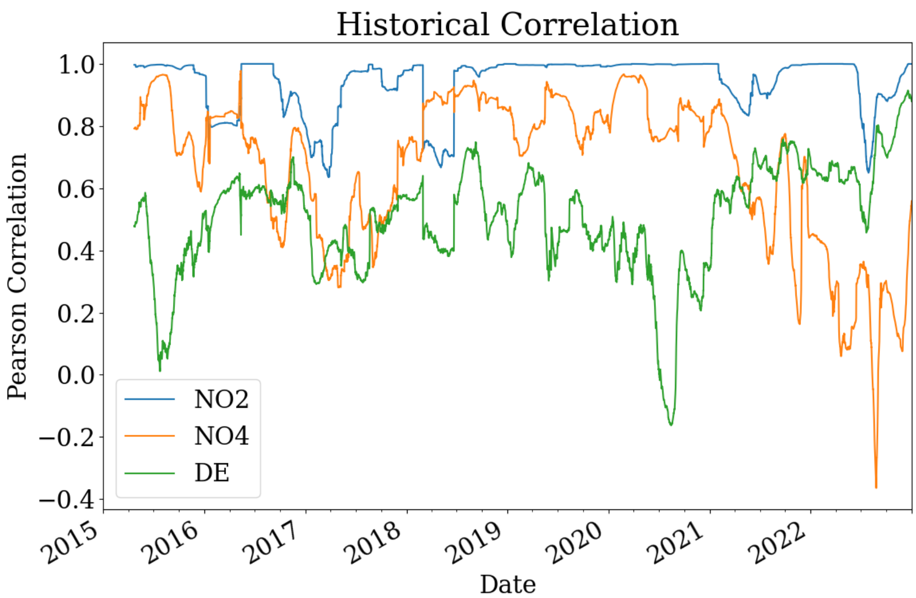


Figure 3.2: Historical correlation of NO1 with NO2, NO4 and DE.

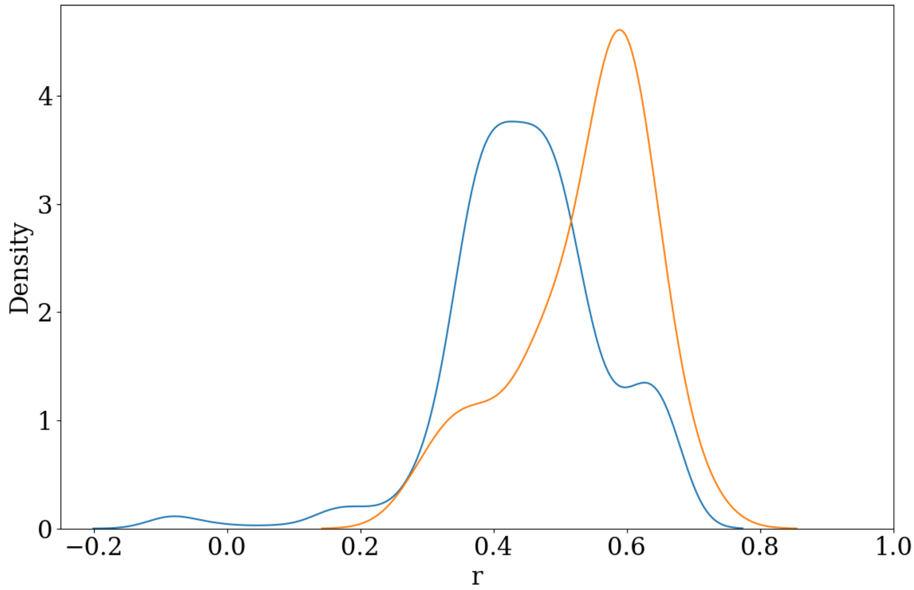


Figure 3.3: Kernel density estimations of the Pearson correlation coefficients r of all BZNs correlated by NO1 for the period before July 2021 (blue) and period after July 2021 (orange).

correlated by all other BZNs of the dataset during different time periods. The plots were achieved by calculating the mean of NO1 correlated by all other BZNs for each hour of the dataset. This resulted in a single correlation coefficient for all hours from 2015 to 2022. The blue, leftmost line represents the histogram of all correlation coefficients from 2015 to the end of June 2021, while the orange line consists of correlation coefficients from July 2021 to the end of 2022. It is interesting to see how the correlation distributions differ between the two periods. The blue line covers a wider range of correlation coefficients, whereas the orange line is more shallow with more concentrated coefficients. Additionally, the mode of the blue line is around the 0.40 correlation range, while the orange one is shifted further to the right, exceeding a value of 0.60. Consequentially, this tells us that the electricity markets in Europe generally became more correlated after the European energy crisis occurred.

To further analyze how the correlation between BZNs has developed throughout the years, we calculated the Pearson correlation between all BZNs for the years 2019 to 2022. The following Figures A.1, A.2, A.3 and A.4 in Appendix A all depict the covariance matrices for the years 2019, 2020, 2021 and 2022 respectively. Whereas most of the BZNs, with the exception of a few, seem to have a positive moderate correlation in 2019, we can notice the shift toward more correlated markets as the years went by. The electricity markets are highly efficient and prone to volatility. During 2020, which was heavily influenced by the COVID-19 pandemic, markets became slightly more correlated in comparison to 2019. However, there is a distinct difference in 2021 from the two previous years. With the exception of a handful of BZNs, NO3, NO4, SE1 and SE2, all other European BZNs became extremely correlated as the European energy crisis occurred. The remaining BZNs, which decorrelated with the rest of the markets, are to no surprise, zones that were not affected by the European energy crisis. In Figure 3.2, as previously mentioned, the northernmost BZN in Norway, NO4, became decorrelated with NO1 during 2021. Furthermore, SE1 and SE2, the two Swedish BZNs located in northern Sweden, likewise decorrelated with the southern BZNs SE3 and SE4, and thus the rest of Europe.

As the energy crisis continued throughout 2022, the closely coupled markets in 2021 were still a reality. However, even though record electricity prices were set across Europe, the closely coupled markets peaked in 2021. The following year was characterized by slightly smaller correlation coefficients between markets, albeit still highly correlated. In Figure A.4, the dark red shade, representing a high positive correlation, is marginally decreased in comparison to 2021.

Even though several markets were highly coupled with correlations close to one, we felt that it was not justifiable to drop a BZN based on this criterion alone. None of the BZNs had a correlation of one throughout the entire 2015–2022 period. Markets couple and decouple throughout time which was highlighted in Figure 3.2, where NO1 and NO2 experienced periods with decoupling, although still remaining highly correlated. We felt that valuable information might be lost if a BZN is dropped based on its correlation to other BZNs. Additionally, we wanted our data to reflect and preserve as much information about the European electricity markets. However, in order to minimize the effect of many highly correlated and similar BZNs we added a penalty term to the loss function of our LSTM that can shrink some of the coefficients whose importance is negligible. Furthermore, this constraint limits the potential for the LSTM predictions to solely duplicate the prices of correlated BZNs. Lasso regularization (L1) has successfully been applied to EPF in (Ziel, 2016) and simply eliminates unnecessary or redundant features from a model with aims at reducing the model's variance. High variance or overfitting occurs when a model is not capable of generalizing and instead captures random noise in the data. Overfitted models tend to memorize the training data instead of learning the underlying patterns, leading to poor performance on new unseen data. To validate our claim on whether valuable information was lost by fully dropping a feature, the statistical dimensionality reduction technique, Principal Component Analysis (PCA) was performed with its result showcased in Section 4.1.1.

3.2 Data Processing

As previously mentioned, the dataset was fairly clean with the exception of all the NaN values. As seen in Figure 3.4, it is evident that for all features the earlier periods are invalid NaN values. For this sole reason, the new dataset was set to start on the 1st of January 2015 and the entire month of

December 2014 was disregarded. Additionally, in order to keep the length of the dataset consistent, the two dates in 2023 were also disregarded. That is, the new dataset covered all hours between the first hour of 2015 and the last hour of 2022. Based on the same figure, the BZN of Montenegro (ME), only included invalid data, apart from a 24 hour recording. This feature was therefore dropped. Furthermore, a total of 18 features were dropped, including the Montenegro BZN, as their ratio of NaN values was too high. Hence, the dataset utilized for the training of our models comprised of 39 features and one target feature, giving rise to a total of 40 features.

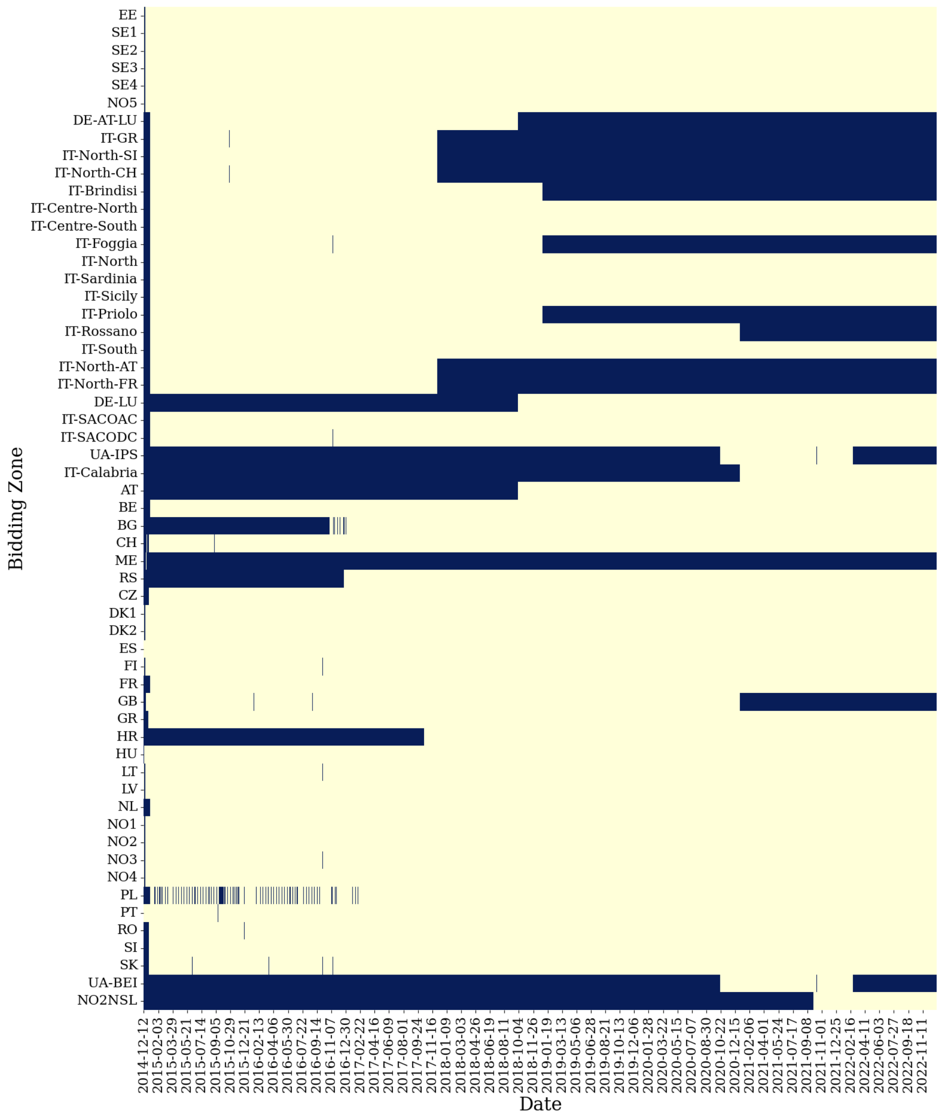


Figure 3.4: Visualization of the amount and location of all the NaN values in the dataset. The dark blue color represents NaN values, whereas the yellow color represent normal values. A full lookup table of the BZNs can be found in Table A.1.

Additionally, Germany has historically shared its BZN with both Austria and Luxembourg (Trebbien et al., 2023). This BZN was split in two on the 1st of October 2018, where one BZN covered Germany and Luxembourg and another Austria. Because of the BZN restructuring, the dataset includes three features which are the Germany, Austria and Luxembourg (DE-AT-LU) BZN, covering dates up until the 1st of October 2018, the new Germany and Luxembourg (DE-LU) BZN and the new Austria (AT) BZN from the 1st of October 2018 to the last hour of 2022. DE-AT-LU and DE-LU were therefore concatenated into a single feature covering all dates and named Germany. The same process was also applied to the Austrian BZN that now consisted of the DE-AT-LU and AT time series.

Several attractive methods for dealing with the NaN values were available, each with its own advantages and disadvantages. Firstly, there was a possibility to both do a backwards or forward fill, changing all NaN sequences with the last valid observation. However, as electricity prices are inherently seasonal, performing a backwards or forward fill would somewhat eliminate this seasonality for certain parts of the dataset, which was not a favorable scenario. Instead, we used linear interpolation to interject the NaN values between two valid observations. Forward interpolation was mainly used, however this technique is limited as series starting with NaN values will not be interjected. Therefore, we applied a combination of both forward interpolation and backward interpolation where it was needed. Nonetheless, if any time series either start or end with an invalid value, interpolation will default to forward or backward filling the first valid observation. Hence, certain BZNs may have longer sequences where the prices are constant with zero rate of increase or decrease. This is the case for the BZNs of Bulgaria (BG), Serbia (RS), Great Britain (GB) and Hungary (HR). Although our approach may have slightly reduced the stochastic nature of the non-stationary time series, we considered it justifiable due to the minor impact on the data.

As all features of the dataset were of the same type and scale, no general standardization or normalization techniques were applied. However, we winsorized the features in order to limit the most extreme prices in the dataset. It is worth noting that our goal was to be able to predict the recent high electricity prices and most of these would be considered outliers based on the earlier low electricity prices. Therefore, it was crucial to process these with care. Important information was contained in the high electricity prices, but a few values were considered extremely high. For instance, Bulgaria experienced a price of 6101.78 €/MWh on the 23rd of February 2017. Subsequently, approximately 2.70% of the data include prices passing the 1000 €/MWh range. Features were therefore winsorized only if their prices exceeded a set threshold value of 45 standard deviations. Values greater than the threshold were shrunken to the greatest value not exceeding the threshold. The value of 45 standard deviations was chosen as it was found to be the most effective at reducing the impact of the worst outliers while retaining as much of the original data as possible.

The data was further divided into train, validation and test sets based on the timestamps. The training set was composed of all timestamps up until the last hour of 2020. The validation set followed after, covering all hours of 2021 and lastly, the test set consisted of all the hours of 2022. Each of these three subsets was further divided into matrices containing the feature values ($\mathbf{X}$), and a single feature containing the output variable ($\mathbf{y}$), which in this case was the NO1 BZN. The data was divided sequentially for several reasons. First of all, it was necessary to have a sort of continuity of the price time series which is why we did not want to stochastically select dates from the entire range of dates. Secondly, our aim was to forecast the high electricity prices of 2022 based on historically lower prices, which is why our data had to be split according to years.

While analyzing the data we came across a few inconsistencies. We initially thought all prices were given in Euros, but this was not the case. Further investigating this issue lead to four features standing out and these were Bulgaria (BG), Romania (RO), Poland (PL) and Great Britain (GB). As the data was collected from the ENTSO-E Transparency Platform[2], we validated our findings by double checking if the mentioned regions were given in other currencies throughout the time period 2015 to the end of 2022. As this was the case and all regions were reported in their local currency, that is, Bulgarian lev, Romanian leu, Polish złoty and British pounds, they had to be converted to Euros. All inconsistent dates were converted to Euros by using the daily Euro exchange rate for each date. Specifically, the time series of Bulgaria required conversion for all dates prior to the 22nd of January 2022, while for the Romanian time series, dates prior to the 17th of June 2021 were converted. In addition, for the Polish time series, the dates between the 2nd of March 2017 and the 20th of November 2019 were converted from the Polish złoty to the Euro. Finally, the BZN of GB was reported in Euros from the 1st of January 2021. Thus, its previous dates were originally reported in pounds and had to be converted to the correct currency.

One last step was still required before the data could be fed as input to the LSTM model. Unlike ordinary ANNs, LSTMs require the shape of the input data to be of three dimensions with the following structure: batch size, timesteps and input dimensions. The timesteps explain how many previous steps the output should be dependent on, while the input dimensions refer to the number of features in the input data. The three subsets therefore had to be reshaped from a regular 2D shape to a 3D shape. This was achieved by applying a rolling window throughout the subsets which simultaneously processed batches of 24 hours. Subsequently, a lag of one day had to be applied to the input features in relation to the target value. This was necessary to mimic the behavior of the electricity markets, where the prediction for the following day is based on the information available from the previous day. We modeled our data such that all 24 hours of a day were dependant on the previous day's hours. Hence, the timesteps of our data were equal to 24. Consequentially, as the batches of 24 rows in the subsets are processed, a new dimension will be added. To demonstrate

[2]https://newtransparency.entsoe.eu/

the difference, the original training set had dimensions of $(52608, 39)$, where the first dimension indicates the total number of hours and the last dimension specifies the total number of features. The transformed three dimensional training set was equal to $(2191, 24, 39)$, where the first dimension now indicates the total number of days. The second dimension indicates the number of timesteps, comprising each hour, while the last dimension remains unchanged.

3.3 Methods

This thesis aims to explore and answer whether LSTMs are capable of accurately forecasting day-ahead electricity prices for a given European region based on multivariate price time series from other influential European BZNs. In this section, we first provide insight into our model architecture and finish the section with an explanation of how the model was evaluated.

3.3.1 LSTM Model

Once the data had been pre-processed and transformed into the correct format, it can be used to train our LSTM model to make predictions. The model architecture follows a straightforward vanilla LSTM approach with an LSTM layer as the input layer, followed by a single dense hidden layer (n_1), and lastly a single dense layer as the output layer. The model was implemented in PYTHON using the functions from the TENSORFLOW.KERAS[3] library. The input LSTM layer consists of ten units with a rectified linear unit (ReLU) as the activation function and L1 regularization set to 0.01. Moreover, the input layer expects data in a specific format, as earlier mentioned. Hence, the input shape of the layer is a three-dimensional shape reflecting the batch size, the number of timesteps to process and the number of features in the dataset. Following is the hidden layer consisting of 20 neurons with the ReLU activation function and lastly is the output layer with a single neuron representing a full day of 24 hours. Table 3.1 highlights the main model architecture that was chosen based on grid search hyperparameter tuning. Although random search, which allows for a wider search space by randomly sampling hyperparameters from a continuous or discrete distribution, is favored over grid search, we applied the latter. This was due to the fact that it is easier to compare the performance of different hyperparameters using grid search. Additionally, we already had a well-defined search space which allowed for a more exhaustive approach.

In the day-ahead auction, wholesale sellers and buyers submit bids for a given hour for the following day. The 24 hourly prices for the following day d are simultaneously calculated, hence the entire day and not only a single hour is calculated at once. In Figure 3.5 this process is presented. Therefore, in order to correctly model the day-ahead market our LSTM model has to

[3]https://www.tensorflow.org/api_docs/python/tf/keras

Table 3.1: Optimal hyperparameters for the LSTM model based on grid search.

Hyperparameter	Value
Activation function	ReLU
Dropout	No
Recurrent dropout	No
Regularization	L1 (0.01)
LSTM units	10
Dense n_1	20
Batch size	8
Optimizer	Adam
Learning rate	0.01

replicate this property, as this is a faithful replication of the market function as described in Section 2.1.1. We also considered the option of altering our model and data such that each hour of the following day d was dependent on the previous hour. Hence, we would simply be predicting a given hour based on previous hours regardless of the day. However, this approach was not applied as it directly contradicts how the day-ahead market operates. The electricity prices of each hour of a day d should all be disclosed at once, and not sequentially. Thus, in order to include market integration, we attempted to adhere to the structure of the electricity price markets to the greatest extent possible.

LSTM models are generally classified into four different architectures: one-to-one, one-to-many, many-to-one and many-to-many depending on the problem one wants to solve. Our task is a typical many-to-many sequential problem where we based on multivariate price time series are predicting the 24 next steps in a given sequence. In addition, given the characteristics of our LSTM model, it can be referred to as a multivariate multi-step LSTM model following a univariate framework for EPF.

Further, our model was optimized using the adaptive momentum estimation (Adam) (Kingma & Ba, 2014) optimizer with a learning rate set to 0.01. Moreover, the loss function was set as the mean absolute error (MAE) which is widely regarded as one of the better-suited loss functions when working with EPF (Lago et al., 2021). Generally, absolute errors that

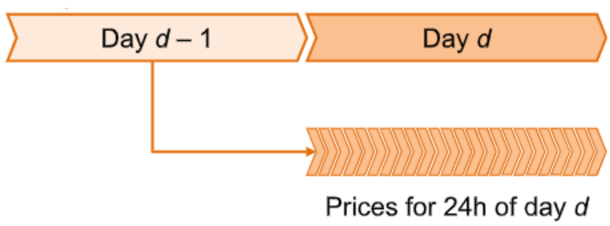


Figure 3.5: Visualization of the day-ahead auction. Bids for each hour of the following day d are made during the present day $d-1$. Image taken from (Lago et al., 2021).

predict the mean of a distribution are preferred over squared errors in EPF for a number of reasons. First, given that electricity price forecasts are used for decision-making purposes, such as bidding in the day-ahead auction or hedging strategies, the magnitude of the forecast errors is more informative. Second, squared errors are prone to be overly sensitive to large errors and less sensitive to small errors because they are prone to place more weight on large errors compared to smaller ones. This is an undesirable characteristic given that large financial implications may occur with suboptimal predictions. Lastly, MAE is more robust to outliers which our dataset is heavily influenced by. Ultimately, MAE is more capable of capturing the characteristics of electricity prices such as seasonality, high volatility and non-linearity. The smaller the value of the loss function, the closer the predicted values are to the actual values. Lastly, the cost of purchasing electricity is linear, absolute metrics are therefore the best to quantify the risk associated with forecasting errors (Lago et al., 2021).

The model was trained for 50 epochs with a batch size of eight on both a local Intel Iris Plus Graphics 640 1536 MB GPU with 8GB RAM and an external cluster, eX3[4] with efficient GPU capabilities provided by SimulaMet[5]. As a precaution against overfitting we applied early stopping which stops model training whenever the loss was not decreasing for a given number of epochs.

3.3.2 Model Explanation

As ANNs are black-box models, their predictions become difficult to explain. RNNs and LSTMs that include feedback loops in the architecture make these black-box solutions increasingly difficult to interpret. In order to efficiently provide trustworthy model explanations, our model was interpreted using Local Interpretable Model Agnostic Explanations (LIME) (Ribeiro et al., 2016) after training. LIME was applied because of its simplicity. It is model agnostic, meaning it can be applied by any type of ML model regardless of the architecture. Moreover, LIME can provide valuable insights into the decision-making process of complex models by identifying biases in model predictions.

LIME explains predictions of complex black-box ML models by creating simpler, more interpretable models that are trained to approximate the behavior of the original model in a local region of a given feature space. For instance, linear models can be used to approximate the behavior of the original black-box model in a local region around a given data point. The simple model approximates and explains each individual prediction of the test subset. In our case, the values returned by LIME show the local feature importance, which is the features that were the main contributors to a single prediction. Positive values indicate features positively influencing the price of a region, whereas negative values indicate the opposite. In other words, an increase in a positively contributing feature will likewise increase

[4]https://www.ex3.simula.no/
[5]https://www.simulamet.no/

the value of the target feature, meaning LIME also shows which features are positively correlated and negatively correlated. Similar to (Trebbien et al., 2023), we quantify a global feature importance from all the local LIME approximations by calculating the mean importance value for each feature during the 2022 period[6].

3.3.3 Evaluation Metrics

In order to evaluate our model in a clear and precise fashion, we followed the guidelines presented in (Lago et al., 2021). There are usually four metrics used to measure the accuracy of EPF and these are the MAE, the root mean squared error (RMSE), the mean absolute percentage error (MAPE) and lastly, the symmetric mean absolute percentage error (sMAPE):

$$\text{MAE} = \frac{1}{24\ N_d} \sum_{d=1}^{N_d} \sum_{h=1}^{24} \left| p_{d,h} - \hat{p}_{d,h} \right|, \tag{3.1}$$

$$\text{RMSE} = \sqrt{\frac{1}{24\ N_d} \sum_{d=1}^{N_d} \sum_{h=1}^{24} \left(p_{d,h} - \hat{p}_{d,h} \right)^2}, \tag{3.2}$$

$$\text{MAPE} = \frac{1}{24\ N_d} \sum_{d=1}^{N_d} \sum_{h=1}^{24} \frac{|p_{d,h} - \hat{p}_{d,h}|}{|p_{d,h}|}, \tag{3.3}$$

$$\text{sMAPE} = \frac{1}{24\ N_d} \sum_{d=1}^{N_d} \sum_{h=1}^{24} 2 \frac{|p_{d,h} - \hat{p}_{d,h}|}{|p_{d,h}| + |\hat{p}_{d,h}|}, \tag{3.4}$$

where $p_{d,h}$ and $\hat{p}_{d,h}$ denotes the ground truth and forecast electricity price on day d and hour h respectively. Lastly, N_d represents the number of days in the test dataset. Each of these evaluation metrics has their strengths and weaknesses.

The MAE, as previously touched upon, was one of the criteria used to evaluate our model. Again, the smaller the value of the metric, the more accurate the forecasts of the model are. Its equation is presented in Equation 3.1 and again, as the underlying risk associated with EPF is linearly dependent on the price and on the forecasting errors, absolute or linear evaluation metrics are the most informative. Quadratic evaluation metrics such as RMSE are however informative in the sense that it is easy to interpret the magnitude of the error. Moreover, as RMSE penalizes large errors more heavily, it becomes a useful and desirable metric in problems where large errors are more costly than small errors, such as in EPF. For instance, if a forecasting model consistently underestimates electricity prices during peak demand periods, RMSE may be a valuable evaluation metric. Furthermore, as RMSE is easy to interpret, it will be used in this thesis.

[6] Although similar, this paper calculated SHAP values and also normalized them by the highest value.

Although MAPE has consistently been used in the EPF literature, we have decided to look past this evaluation metric because of the findings listed in (Lago et al., 2021). According to (Lago et al., 2021) MAPE was unreliable as it completely disagreed with other evaluation metrics. Whereas the other metrics somewhat agreed on what the best models were, MAPE was far off, therefore it is disregarded in this thesis. Nevertheless, a variation of MAPE, sMAPE, has consistently been able to reliably evaluate EPF models and will be incorporated. sMAPE has the benefit that it is symmetrical and treats over-estimation and under-estimation errors equally, which is convenient in the context of EPF as both errors can have significant financial impacts.

Finally, drawing conclusions from non-stationary price time series can be cumbersome, especially between different BZNs and datasets. The field of EPF is dependent on having a general framework available that researchers can use to easily evaluate and compare their models. Hence, relative mean absolute error (rMAE) was introduced as an evaluation metric that is capable of providing relative measurements such that models can be validated on several datasets without losing context. The evaluation metric is defined in Equation 3.5 where the numerator, MAE, is divided by the MAE of a naive forecast. This is useful because the metric provides a measure of the accuracy of a given forecast relative to the magnitude of the actual price. Electricity prices vary over time, as we have seen in Figure 3.1. rMAE will therefore be used such that it easily can be compared against the results of other research papers.

$$\text{rMAE} = \frac{\frac{1}{24\ N_d}\sum_{d=1}^{N_d}\sum_{h=1}^{24}|p_{d,h} - \hat{p}_{d,h}|}{\frac{1}{24\ N_d}\sum_{d=1}^{N_d}\sum_{h=1}^{24}|p_{d,h} - \hat{p}_{d,h}^{naive}|}, \tag{3.5}$$

In addition to the four EPF evaluation metrics, we will evaluate our model with a baseline persistence model that simply assumes that today's electricity price will be equal to the mean of the previous day. In order to make the persistence model a bit more sophisticated the standard deviation of each previous day will randomly either be added or subtracted from today's day. In Equation 3.6, p_d denotes the price for all hours of day d, while $\overline{p}_{d-1}$ and σ_{d-1} denotes the mean price and the standard deviation of the previous 24 hours respectively. As electricity prices are non-stationary time series, meaning they include trends and seasonality, the persistence model will be a drastic simplification of the EPF problem.

$$p_d = \overline{p}_{d-1} \pm \sigma_{d-1} \tag{3.6}$$

Chapter 4

Results

This chapter presents the key findings of our research. We start off by covering the results gathered from our multivariate price time series LSTM model applied to the energy markets. Thereafter, a thorough comparison against another state-of-the-art EPF model will be conducted. Although the aim of the thesis is not to outperform the current state-of-the-art but rather to investigate an alternative approach, it will be useful to observe where this approach ranks and whether it holds any merit. Lastly, the main drivers of European electricity prices will be disentangled.

4.1 EPF with Multivariate Time Series LSTM

Our LSTM model was validated on the four evaluation metrics listed in Chapter 3. Combining forecasts in the field of ML has generally been beneficial in order to achieve more robust and trustworthy predictions. The case is no other in the field of EPF. According to (Lago et al., 2021), ensemble models, which are combinations of different models' forecasts, are valuable. Therefore, we adapted an ensemble model by simply taking the mean of the forecasts of five different runs.

Our developed ML model is capable of predicting NO1 day-ahead prices with an average performance of MAE = 21.17. That is, our model's forecasts are on average off by $\pm$ 22.42 €/MWh. For comparison the ML model developed in (Trebbien et al., 2023) achieved an MAE = 5.53, while the review paper of (Lago et al., 2021) achieved an MAE of 1.712 on their DNN model applied to the Nord Pool power market. Although our results look far off at first glance, it is worth reiterating a key characteristic of the electricity markets which is their inherent differences. As can be recalled from Chapter 2, each electricity market has its unique characteristics that ML models have to be designed and tailored towards. Therefore, it is not a given that a well-performing ML model in a certain electricity market will perform similarly in another. For instance, using an ML model tailored towards the German electricity market to forecast Turkish day-ahead prices will most likely yield unsatisfactory results given the market differences. Likewise, blindly comparing the performance of EPF evaluation metrics across datasets, electricity markets, research objectives and models, is a

Table 4.1: Summary of the evaluation metrics for the LSTM model, the benchmark DNN from (Lago et al., 2018) and the persistence model for NO1. The LSTM model outperforms both the DNN and persistence model on every metric and is highlighted by the gray shading.

	rMAE [€/MWh]	MAE [€/MWh]	RMSE [€/MWh]	sMAPE [%]
LSTM	1.04	21.17	25.78	12.64
DNN (Lago et al., 2018)	1.46	29.71	35.63	17.87
Persistence	1.93	40.05	64.78	25.70

naive and sub-optimal approach. Moreover, the timespans of price data covered in this thesis and the two aforementioned publications differ. Our inclusion of the most recent years, particularly after the start of the energy crisis, weaken the performance metrics of the LSTM output.

In order to achieve a fairer and significant comparison between state-of-the-art models we implemented the DNN from (Lago et al., 2018), which still is considered a benchmark EPF model. For the sake of achieving the fairest comparison between the two models, we placed great emphasis on using the exact same data and hyperparameters, with the exception of model architectures. Therefore, even though vanilla DNNs usually process 2D data, the same data processing steps listed for the LSTM in Chapter 3 were implemented. That is, the input data to the DNN were of three dimensions: batch size, timesteps and number of features.

Similarly, the results gathered from the DNN were a combination of five different model runs. All results for both the LSTM and the benchmark DNN are listed in Table 4.1. For all metrics, the lower the value the better the performance is. The gray color in each of the evaluation metric columns represent the best performance and one can clearly see how the LSTM outperforms the benchmark DNN on the multivariate price time series dataset. The results may not come as a surprise given that LSTMs generally are better suited tools for processing data with temporal dimensions. However, it is interesting to see how the performance of the benchmark DNN model significantly decreases from an MAE of 1.712 in (Lago et al., 2021) to 29.71 for our use case. On the contrary, the performance gain on both models from ordinary MAE to rMAE is vastly greater. The LSTM model explains the price with an average rMAE of 1.04 €/MWh which is considered promising.

Even though our proposed LSTM model outperformed the benchmark DNN for this particular use case, we further validated whether our model was superior in a statistically significant way. A Diebold-Mariano (DM) (Diebold & Mariano, 2002) test was conducted to assess the statistical significance of the differences in forecasting accuracy. The test is model-free, meaning it compares the forecasts of models and not the models themselves. They simply calculate z-scores, which represent the number of standard deviations an observed data point is above or below the mean of a distribution.

$$\mathrm{DM} = \sqrt{N}\,\frac{\hat{\mu}}{\hat{\sigma}} \tag{4.1}$$

Equation 4.1 represents how DM is calculated. N denotes the number of days in the out-of-sample test dataset, while $\hat{\mu}$ and $\hat{\sigma}$ denotes the sample mean and sample standard deviation respectively. Forecasts are compared with the null hypothesis that the difference between the forecasts is insignificant. In our case the acceptance level was set to 5%. That is, p-values greater than 0.05 meant that the null hypothesis could not be disregarded, and the difference between forecasts was not significant. On the other hand, forecasts of models with p-values lower than 0.05 were considered to be significantly better than their counterparts.

The DM test was jointly performed for all the 24 hours of each day (multivariate) in the test dataset instead of 24 independent tests for each hour (univariate). Thus, a single p-value was calculated for all days used for forecasting which makes plotting convenient. In our case, three of the best performing LSTM and DNN models, based on MAE, were used to perform a multivariate DM test. In Figure 4.1 the result is depicted. The closer the p-values are to zero, the more significant the difference between the forecasts of a model on the x-axis and a model on the y-axis is. The x-axis represents the better performing model, while the y-axis represent the worse. For instance, the forecasts of $LSTM_1$ are statistically better than all other models given by the five green squares. Likewise, the forecasts of DNN_1 are statistically better than the two other DNNs, but none of the LSTM models, marked by the black squares. The DM test shows how the forecasts of LSTMs are better than the DNNs for our EPF problem, which is consistent with the evaluation metrics.

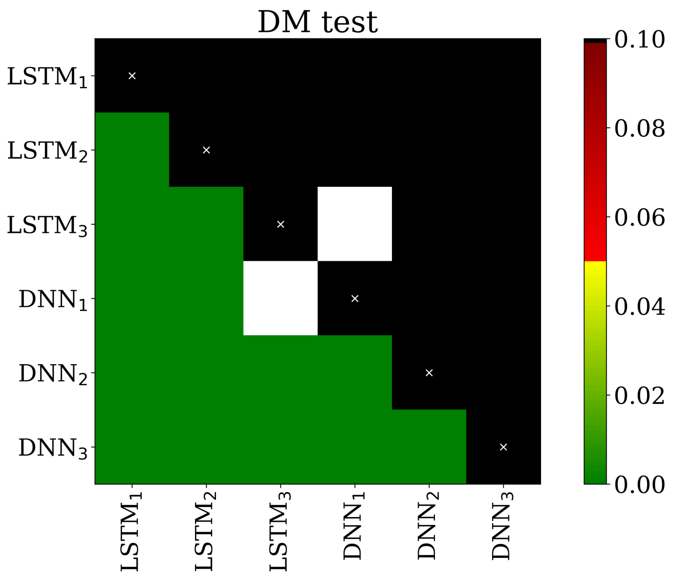


Figure 4.1: Diebold-Mariano test between the three best performing LSTM and DNN models with a 5% acceptance level. Low p-values (dark green) equals statistically better forecasts. Higher p-values equals worse forecasts. p-values greater than 0.10 (black) are disregarded.

This thesis' focus centers around NO1, however we were interested in how the LSTM performed when predicting other regions. Therefore, to get an overview of which regions are easier to predict than others, we forecast all BZNs in the dataset[1]. The same process presented in Chapter 3 was applied for each of the features and Table 4.2 presents all the results from the experiment. Although the LSTM is trained from scratch for each BZN, the results illustrate the generalizability of our implementation. The metrics highlighted by the gray color represent the best score achieved. Further on, the rightmost table represents how the baseline persistence model performed. The regions highlighted in boldface were predicted better than the LSTM model.

For the LSTM model several regions stand out. First off is the forecasts of IT-Centre-South which percentage-wise were the best with an average error of 9.47%. Likewise, the forecasts of Portugal (PT) outperformed the rest of the BZNs in regards to MAE and RMSE. With an average MAE and RMSE of 23.87 and 29.66 respectively, PT was one of the easier BZNs to forecast. However, in the case of sMAPE, PT was far from the best with slightly dissatisfying results. Moreover, Denmark 1 (DK1), Denmark 2 (DK2), Netherlands (NL) and the two northern Norwegian BZNs seem to be increasingly difficult for the LSTM to forecast accurately. It is especially interesting to compare the forecasts of NO3 and NO4 to the rest of the Norwegian BZNs. Whereas NO1 was the easiest to forecast of the five, NO3 and NO4, which decoupled from the rest of Norway and Europe, are two of the most difficult regions to forecast. However, the forecasts of NO2 and NO5, which are highly coupled with NO1, were significantly worse than that of NO1 and especially NO2. Regarding the Swedish BZNs, the results are the opposite of Norway. The two southern BZNs, SE4 and SE3, achieved the worst results, whereas the northern regions were easier to predict.

Moving over to the persistence model, there is a clear trend that the results are worse than the more sophisticated LSTM. Ideally, all BZNs should have been forecast worse than the LSTM, but that was not the case. Seven out of 39 BZNs were more accurately predicted using the simple persistence model and these are highlighted in boldface text. Given that a simple assumption such that tomorrow's price is equal to the mean $\pm$ the standard deviation of the current day, is capable of outperforming an LSTM on a few regions, speaks to the complexity of EPF, especially during the current European electricity crisis. Additionally, it provides good evidence that although markets may be coupled, they can be very different and exhibit vastly unique characteristics. Some of the worse predicted BZNs using the LSTM were on the contrary accurately predicted using persistence modeling. The difference in forecast accuracy for NO3 and NO4 between the LSTM and persistence model is remarkable. NO4 was also among the most accurately predicted regions with a MAE and RMSE of 12.09 and 37.26 respectively. It seems plausible that the persistence model is effective whenever prices are stagnant without extreme price

[1]With the exception of Great Britain as its entire validation and testing period consisted of interpolated data.

Table 4.2: MAE, RMSE, and sMAPE for all BZNs not dropped from the original dataset. The best values for each evaluation metric is highlighted by the gray color. The leftmost table represents the LSTM results, whereas the rightmost the persistence model. Instances where the persistence model was superior to its counterpart is highlighted in boldface. A full lookup table of the BZNs can be found in Table A.1.

LSTM

BZN	MAE	RMSE	sMAPE [%]
AT	49.97	63.58	21.75
BE	42.40	57.46	21.13
BG	69.99	86.55	31.23
CH	61.54	73.76	25.28
CZ	42.39	53.50	19.28
DE	60.78	75.56	31.34
DK1	95.38	121.99	56.79
DK2	83.88	113.53	51.79
EE	69.76	95.42	45.58
ES	26.69	39.23	19.56
FI	55.72	74.95	49.56
FR	56.74	88.25	21.49
GR	47.15	66.47	18.32
HR	62.13	78.36	25.19
HU	44.49	57.58	17.89
IT-Centre-North	40.73	58.13	12.65
IT-Centre-South	27.95	39.81	9.47
IT-North	73.74	100.42	24.51
IT-Sardinia	53.76	77.17	20.53
IT-Sicily	46.83	64.28	17.49
IT-South	88.85	115.62	32.49
LT	61.98	86.19	29.26
LV	37.55	55.38	17.63
NL	99.88	120.76	49.95
NO2	45.73	71.40	22.31
NO3	65.45	78.55	111.38
NO4	105.60	125.26	140.58
NO5	26.09	31.19	16.00
PL	46.73	61.60	32.31
PT	23.87	29.66	17.37
RO	41.11	53.03	17.57
RS	43.41	56.20	18.45
SE1	26.05	37.06	58.46
SE2	24.69	35.31	55.51
SE3	57.75	87.41	58.64
SE4	62.77	86.63	58.51
SI	32.78	44.53	13.42
SK	47.97	63.07	20.29

Persistence

BZN	MAE	RMSE	sMAPE [%]
AT	68.99	94.92	30.17
BE	79.38	108.27	38.44
BG	90.04	122.32	40.42
CH	**56.91**	78.02	**21.67**
CZ	74.88	100.76	35.36
DE	83.08	111.60	46.02
DK1	**76.73**	**103.06**	**46.79**
DK2	88.39	119.32	56.86
EE	96.21	141.30	56.85
ES	40.16	54.85	27.80
FI	91.71	132.77	74.74
FR	69.71	101.99	27.75
GR	80.31	113.79	29.73
HR	75.29	102.92	31.20
HU	78.76	107.38	33.61
IT-Centre-North	63.68	85.86	21.81
IT-Centre-South	64.89	87.80	23.02
IT-North	**63.31**	**86.86**	**21.26**
IT-Sardinia	74.61	113.26	28.59
IT-Sicily	68.37	91.95	26.80
IT-South	**62.76**	**84.55**	**22.80**
LT	94.36	143.40	48.61
LV	90.80	134.93	47.00
NL	**80.34**	**109.27**	**40.20**
NO2	46.96	72.62	26.37
NO3	**20.69**	**44.91**	**45.51**
NO4	**12.09**	**37.26**	**35.97**
NO5	37.87	61.60	24.08
PL	56.57	82.42	34.60
PT	37.99	51.01	27.15
RO	92.76	126.40	39.83
RS	71.65	98.51	28.94
SE1	30.90	53.93	61.38
SE2	35.52	63.89	64.62
SE3	87.83	125.92	84.25
SE4	95.36	134.68	82.79
SI	73.16	97.43	30.10
SK	83.47	112.84	36.27

spikes and fluctuations, which likely is the case for the good predictions of NO3 and NO4. Recalling back to Figure 3.1, we can observe how the prices of NO4 after 2021 do not fluctuate nearly as much as NO1. This assumption additionally holds weight as the BZN of Portugal (PT) was one of the more accurately predicted using the persistence model. Although the prices of PT fluctuated a lot more than NO3 and NO4, the volatility was far off what most of Europe experienced. Additionally, Sweden's two southern regions achieved some of the most underwhelming results. Analyzing the Swedish prices showed a clear trend of very fluctuating prices with large price spikes within a day, which most likely is the reason for the forecasting errors. On the other hand, the prices of Sweden 1 (SE1) and Sweden 2 (SE2) were significantly less volatile during the 2022 period than their southern counterparts. We highlight that the northern Swedish BZNs achieved far more accurate predictions than the southern ones.

4.1.1 Principal Component Analysis

PCA was applied and tested on our dataset. European electricity markets being coupled together adds to the complexity of our dataset as certain regions heavily influence others and vice versa. Seeing as several BZNs are highly correlated, reducing the overall complexity while maintaining as much of the data variability as possible may yield more accurate predictions of our LSTM model.

It is considered good practice to standardize the data before applying PCA in order to make sure the principal components reflect the true underlying structure of the data. Our dataset was therefore normalized between the range of zero and one. Thereafter, PCA with 95% variance, i.e. 95% of the variance in the original dataset is captured, was performed, resulting in 12 principal components. The principal components are linear combinations of the original features that capture the most variance. In other words, the first principal component is the linear combination that explains the most variance of the original data, whereas the second is a linear combination that explains the most variance of the remaining variables and so on.

PCA reduced the amount of features in the dataset by 69% as the original dataset used for training consisted of 39 features. The loadings of each principal component, which represent the contribution of each original feature to the principal component, were analyzed. No clear pattern could be observed from the loadings. There was not a single feature that contributed heavily to the overall variance of the dataset. All features more or less contributed evenly.

After applying PCA we trained our LSTM using the 12 principle components to predict NO1 and the results were very underwhelming. Whereas we originally achieved acceptable predictions, as can be seen in Table 4.1, the forecasts using PCA were far off and achieved an average MAE equal to 136.20 after five runs. Since the forecasts were worse and the loadings randomly distributed, PCA was disregarded as it was not sufficient for our use-case. The dimensionality reduction was not capable of capturing essential signals in the data. Thus, it was clear that the prices of the original BZNs contained far too valuable information and their structure was too complex to be captured by simple linear combinations.

4.2 Interpreting Market Coupling using LIME

One of the aims of the thesis was to disentangle the main drivers of the abnormal European electricity prices. XAI was therefore implemented using LIME to decipher model predictions. The feature importance approximated

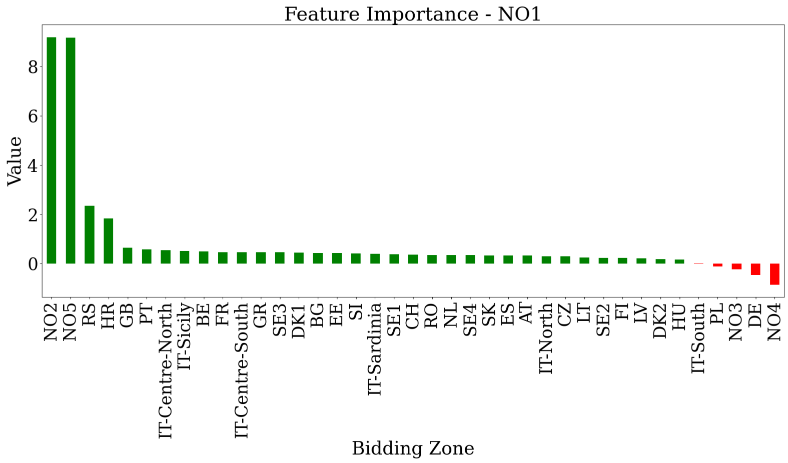


Figure 4.2: Feature importance for NO1 using LIME on the one year test subset. The most relevant BZNs are listed along the x-axis and their values along the y-axis. Green bars highlight BZNs positively contributing to predictions, whereas red highlight the opposite.

by LIME reveals which features have the strongest influence on the prices for a given region.

In Figure 4.2 the cumulative feature importance of NO1 in 2022 is depicted. Understandably, NO2 and NO5, the two most correlated BZNs are two of the main drivers for the electricity prices. Interesting is the moderate importance of the Serbian (RS) and Croatian (HR) features. Given that NO1 does not share any direct interconnectors to neither Serbia nor Croatia, this result is unexpected and will further be discussed in Chapter 5. On the other side, the features contributing most negatively to the prices of NO1 were NO3, DE and NO4. Again as expected, NO3 and NO4 which decoupled from NO1 during the middle of 2021, were the most negatively contributing features. However, it is quite unexpected seeing DE negatively influencing the electricity prices as Germany trades significant amounts of electricity with Norway. One would perhaps think that with the opening of NordLink[2], the subsea interconnector between southern Norway and Germany, DE should have a more important role in determining Norwegian electricity prices.

The feature importance of other important European electricity markets was additionally calculated and compared. More specifically, the results of Germany, Netherlands, Austria, France (FR), Sweden 1 and Denmark 2 are highlighted in Figure 4.3. From the six plots, we can see a clear difference in the number of BZNs that affected the forecasts in comparison to NO1 and SE1. Whereas all 39 features contributed to some degree to the prices of NO1 and SE1, a more concentrated group of BZNs contributed to the other European markets. Moreover, the feature importance values for NO1 and FR are significantly greater than those contributing to the prices in

[2]https://www.statnett.no/en/our-projects/interconnectors/nordlink/

the other BZNs. While most of the BZNs have feature importance values ranging between two and four, both NO1 and FR have values above seven, meaning these were more influential. In the case of all the regions, just like NO1, the Serbian and Croatian BZNs seem to play an important role as they are among the more important features.

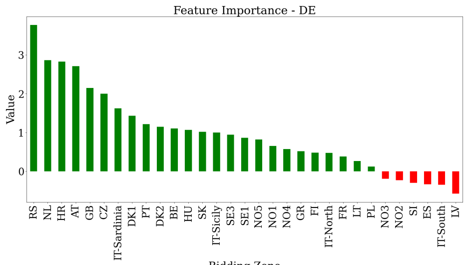
Including RS and HR, the German BZN seem to be highly coupled with NL, AT, GB and the Czech Republic (CZ). As Austria and Germany shared their bidding zone for several years, this result does not come as a surprise. Analyzing the Austrian BZN, we can see also see that its neighboring BZNs DE, IT-North and CZ are influential to the high prices. France is in addition coupled with its neighboring countries Switzerland (CH) and Belgium (BE).

For the two other Scandinavian regions we can also see a strong dependency on RS and HR, which has shown to be valuable features for the European electricity prices. Additionally, the Scandinavian regions including NO1 have an inverse relationship with DE. Whenever prices in DE decrease, they increase for the Scandinavian countries. We further see that several BZNs contribute negatively to the prices of SE1 and DK2 in comparison to the other plots. For SE1 there is additionally a strong relationship with the three southern Norwegian BZNs. On the other hand, the two northernmost Norwegian BZNs that geographically are closest to SE1, have a negative interaction. SE1 is additionally the BZN with the most balanced amount of positively and negatively influencing features.

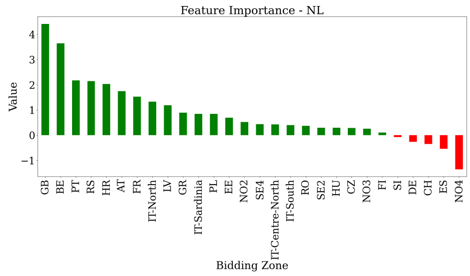
It is plausible that this interaction is given by how the two countries trade electricity as electricity prices and trades are closely related. Although generalizing electricity markets should be done with care given their complexity, in a well functioning electricity market, exporting electricity usually leads to higher prices for the exporting region, whereas importing leads to lower prices in the importing region due to a further source of electricity. Thus SE1 might find it attractive to export their power plant electricity to NO4 when there is demand for it, resulting in lower prices for NO4 and slightly higher for SE1.

Lastly, in DK2 its neighboring DK1 region is the most important, while the two Italian IT-North and IT-Sardinia are the most negative. Whereas NO5 and NO2 are significant contributors to the higher prices, NO1 is not. Although this behavior is counter intuitive as NO1, NO2 and NO5 are very similar, we have to remember that we are creating global feature importances from local explanations. From Figure 3.2, we recall that NO1 and NO2 decouple several times with the last time being in 2022. It might be the case that during certain dates in 2022, DK2 was heavily influenced by NO2 and NO5 and not by NO1.

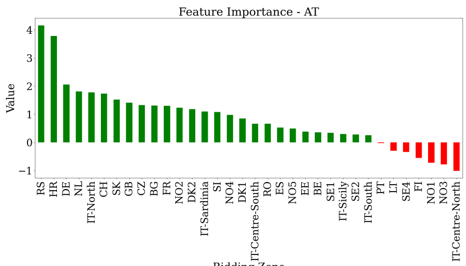
Summarizing, RS and HR seem to be key features with strong influence on several European electricity markets. Additionally, GB positively influences all BZNs showcased in the feature importance plots. Moreover, DE was also a contributing factor in all plots, however its contribution was both positive and negative. The influence GB and DE have over European electricity prices from the plots is not surprising. Great Britain and Germany are among the largest electricity markets in Europe, consequentially making them important in shaping the nature of the European energy market. Thus, our plots and LIME are capable of understanding the state of the energy markets. Furthermore, we hypothesize that the importance of GB and DE on European electricity prices also is contributed to the phase-out of nuclear power in Germany and the transition away from fossil fuels in Great Britain. The shift toward renewable energy sources, in addition to the European electricity crisis, has disrupted energy markets and perhaps strengthened the importance of GB and DE.



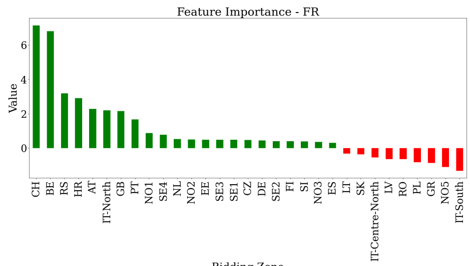
(a) Germany (DE)



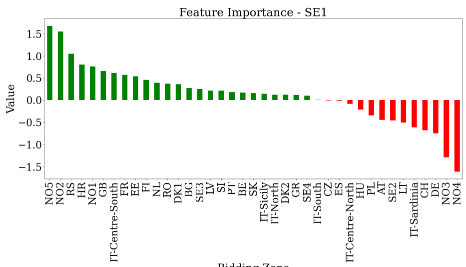
(b) Netherlands (NL)



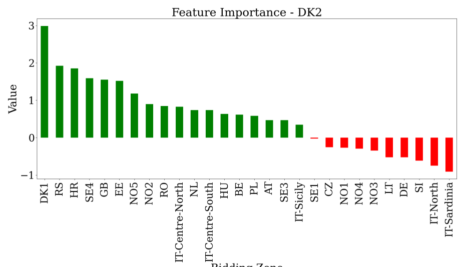
(c) Austria (AT)



(d) France (FR)



(e) Sweden 1 (SE1)



(f) Denmark 2 (DK2)

Figure 4.3: Feature importances for selected European BZNs using LIME on the one year test subset. The most relevant BZNs are listed along the x-axis and their values along the y-axis. Green bars highlight BZNs positively contributing predictions, whereas red highlight the opposite.

Chapter 5

Discussion and Conclusion

This final chapter starts off with a thorough discussion of the main findings of the thesis. Further on, a discussion of the strengths and crucial limitations of the thesis will be provided. We will close the thesis with a conclusion and suggestion on what future work and steps should be taken moving forward regarding multivariate EPF.

5.1 Discussion

To our knowledge, this is the first line of work aiming to forecast day-ahead electricity prices using multivariate price time series. Whereas several studies in EPF target single electricity markets, our thesis is based on multiple markets. We validate our approach on all BZNs in the open-access dataset to provide a well-balanced perspective regarding its validity. Furthermore, the modeling complexity using our approach is drastically reduced in comparison to ordinarily EPF. Our multivariate EPF approach with LIME allows us to analyze how electricity markets are interconnected and influence each other through the SDAC described in Section 2.1.1.

5.1.1 Findings

Our results show that NO1 was the most accurately predicted BZN across three out of four evaluation metrics with an average rMAE of 1.04, MAE of 21.17, RMSE of 25.78 and sMAPE of 12.64% using our novel modeling approach. While these results are lackluster in terms of other state-of-the-art methods, we reiterate the significant difference between our multivariate time series approach and other standard univariate time series approaches that take into account weather forecasts, load and demand. Accurate comparisons with other benchmark results are therefore not straightforward. Additionally, the European EPF literature is fairly concentrated around a handful of important electricity markets such as Germany and France. The literature on forecasting spot day-ahead prices for Norwegian BZNs is sparse, making comparisons even more difficult. That said, we found that our model, in the case of multivariate EPF, was capable of explaining electricity prices with greater precision than the benchmark DNN from (Lago et al., 2018). The DNN was outperformed when transforming the data to 3D to match that of the LSTM, but also without the transformation. The DM test in Figure 4.1 showed us that the forecasting differences were statistically significant and not due to randomness.

Our findings are in alignment with the findings of (Ugurlu et al., 2018). They came to the conclusion that models capable of processing data with temporal dimensions, such as LSTMs, are better suited tools for EPF as they yield more precise forecasts. However, it should be noted that their results regarding RNNs contradict the findings of (Lago et al., 2018) which achieved more accurate forecasts using DNNs. The two papers studied different electricity markets, making both findings hard to dismiss. However, as we know that electricity markets have unique characteristics, it could be the case that for some markets a given model is preferred. In our case, using LSTMs with multivariate price time series was the superior option.

We also found that the LSTM in most cases, which was expected, was capable of forecasting electricity prices with greater precision than the baseline persistence model. As we saw, BZNs such as DK1, DK2, NL, but also the isolated NO4 BZN, seem to be the most difficult for our LSTM to forecast. While no definite answer for why this is could be formulated, we have tried to justify our reasoning. First of all, various factors determine electricity prices. Since the European energy crisis, there has been a deficit of power and we saw in Figure 3.3 how markets became tightly coupled. The NO4 BZN was one of the exceptions where the market decoupled from the rest of Norway and Europe. Prior to the price increases in 2021, all five Norwegian BZNs were closely coupled with similar prices as illustrated in Figures A.1 and A.2. During model training, the LSTM is likely learning to replicate the prices of the four other Norwegian BZNs. Thus, because of the vastly different landscape in the test subset contrary to the train subset, the LSTM is not capable of generalizing and predicts NO4 to be similar to the southern BZNs.

In contrast, the below-average forecasts of DK1, DK2 and NL can only be hypothesized. By incorporating domain knowledge we can perhaps explain why this is. Denmark and the Netherlands are similar in several regards. First off, they are both flat low-lying countries located near the North Sea, making them highly suitable for both offshore and onshore wind power. Consequentially, a significant portion of the power generation in both countries stems from wind power. We conjecture that the poor forecasts are a result of the countries heavy reliance on highly volatile wind power, making the prices difficult to predict.

We further need to highlight the difficult landscape regarding EPF in recent years. Most research cited throughout this thesis has had the benefit of forecasting electricity prices in a stable market, mostly without persistent price spikes. Again, introducing uncertainty between model comparisons. Today's electricity situation in Europe is incomparable to what it was like during the 2010s. Thus, research prior to the European electricity crisis has usually trained and tested their models on similar market conditions, whereas the same cannot be said for this thesis. We trained our model on prices prior to 2021, which historically are considered low, and aimed at forecasting prices many fold larger than what the model was trained to do. This certainly affected our forecasting precision, making our model look more lackluster than it is. Additionally, since our model was not trained on 2021 and 2022 data, it has limited knowledge of the extraordinary market coupling that started during the European electricity crisis. Thus, the underlying relationships between BZNs in the test dataset were most likely not optimally captured. The most optimal approach in ML is of course to train and test models on similar conditions, but this would have required significantly more data with abnormal prices, which was not possible. We optimistically believe that the more data with higher prices become available, the easier it will be to more accurately forecast current market conditions using our approach.

Recalling back to the feature importance plots depicted in Figure 4.3 we mentioned the unusual importance of the RS and HR BZNs for most of the regions. We cannot exclude the fact that a significant portion of their price time series was linearly interpolated adding a certain bias to the data. In Figure 3.4, we see how the start of both time series are missing crucial data. It might be the case that the LSTM is sensitive to longer abrupt sequences of linear data, thus leading it to allocate greater emphasis on the interpolated data and skewing the importance of the feature. While we do believe the regions included underlying patterns about the European electricity markets that the model was capable of identifying, we are hesitant to fully conclude that these regions were superior contributors.

5.1.2 Limitations

It is important we acknowledge the limitations of this thesis, which may have affected the results acquired. First of all, the data preprocessing steps were crucial in order to most accurately present the reality of the European electricity markets. As we have shown, prices for several hours and dates were missing for many of the regions. To counteract this issue we interpolated the data using linear interpolation, meaning all missing dates in a given interval will have a constant change of rate. This is without a doubt not the most sophisticated way of interpolating time series data and moreover, it is not of stochastic nature, which electricity prices inherently are. Furthermore, the seasonal trends and volatility spikes that make electricity prices fluctuate will be lost in the interpolated range. We could perhaps have interpolated the missing dates in a stochastic manner, where uncertainty and variability of prices were taken into account, however, given the long interval length of some of the missing periods we felt that the benefits of stochastic interpolation would not outweigh the simplicity and efficiency of linear interpolation. When the missing dates exceed several days, the advantage of stochastic interpolation loses its value as the introduced variability more or less becomes random. However, it could perhaps be an interesting addition to implement Brownian Bridges where the missing values were limited to just a couple of days or even just a few hours to maintain the statistical properties of the time series. That said, the aim of the thesis was to investigate a novel approach for modeling EPF and not to find the most appropriate interpolation technique.

LIME, which was introduced to explain model predictions, also includes a mention-worthy limitation. The explanations of LIME are supposed to provide local explanations for a given data point. From this, we quantified a global feature importance by aggregating each local explanation and taking the mean of each observation. As LIME is not constructed for this use case, the results may not be completely accurate. That said, it does provide a clear overview and trend of which features are the main contributors.

Lastly, the generalizability of our LSTM model is not strong. Training the model to forecast a certain region and thereafter using the pre-trained model to forecast another out-of-sample BZN will not yield accurate results. However, given that all BZNs have their own unique characteristics we felt that optimizing the model towards generalizability would not be useful. In addition, we struggle to find a scenario where one would want to forecast a specific region where the model has been trained and tailored towards another region. Having a universal EPF model tailored towards the entire European electricity market would be out of scope for this thesis. Additionally, the signal-to-noise ratio for this specific task is very low, making it difficult to generalize. Generalizability was therefore not the main concern of the thesis, albeit regularization made the results more promising.

5.2 Conclusions and Future Work

To conclude, this thesis instigated the applicability of multivariate price time series to forecast European day-ahead markets. We introduced a novel approach using an LSTM that was successfully able of estimating spot day-ahead prices based only on multivariate price time series from different European BZNs. With varying accuracy and conviction, we have demonstrated the usefulness of said approach in regards to simplifying the complexity of EPF modeling and improving market coupling analyses. Our main LSTM tailored toward NO1 was capable of explaining electricity prices with an MAE of 21.17, meaning its forecast on average were 21.17 €/MWh off the actual prices. Although the results were promising for certain BZNs and different markets are characterized by unique properties, the errors achieved in comparison to univariate modeling are too high. We cannot therefore conclude that multivariate price time series perform better than ordinary univariate EPF. Thus, multivariate EPF should not be seen as a replacement for univariate EPF. Granted that features such as load and demand are weighted into the price time series, their standalone importance cannot be dismissed.

Moreover, we have analyzed how dependencies and dynamics between European electricity markets in the latter part of the 2010s have changed. Through the use of LIME, it was revealed that the contribution of Germany and Great Britain to electricity prices was of significant importance. Our thesis also confirmed that closely related BZNs, connected either via interconnectors or shared borders, have a greater influence in determining a given region's prices than others, as could be expected.

We believe this thesis is a solid starting ground for further research. Several open-ended questions still remain, one of which being the reasoning behind the lackluster DK1, DK2 and NL forecasts. While our conjecture might not be incorrect, this is definitely a line of study for further research. In addition, significant gains in performance might be achieved by introducing logarithmic returns. Electricity prices are inherently non-stationary and exhibit highly variable price movements. Hence, the high variance and volatility can be limited by modeling electricity prices using logarithmic returns. Finally, investigating the impact of including or excluding certain BZNs may result in deeper understanding of the complexity of electricity markets.

Bibliography

Austin, C. (2021). What is the greatest scientific discovery of all time? Retrieved January 19, 2023, from https://www.carbonscientific.co.uk/blogs/news/what-are-the-greatest-scientific-discoveries-of-all-time

Bahar, H., & Sauvage, J. (2013). *Cross-border trade in electricity and the development of renewables-based electric power: Lessons from europe* (OECD Trade and Environment Working Papers No. 2). OECD Publishing. https://doi.org/10.1787/5k4869cdwnzr-en

Baldi, P. (1995). Gradient descent learning algorithm overview: A general dynamical systems perspective. *IEEE Transactions on Neural Networks*, *6*(1), 182–195. https://doi.org/10.1109/72.363438

Bunn, D. (2000). Forecasting loads and prices in competitive power markets. *Proceedings of the IEEE*, *88*(2), 163–169. https://doi.org/10.1109/5.823996

Bye, T., & Hope, E. (2005). Deregulation of electricity markets: The norwegian experience. *Economic and Political Weekly*, *40*(50), 5269–5278. https://www.jstor.org/stable/4417519

Christiansen, N. H., Voie, P. E. T., Winther, O., & Høgsberg, J. (2014). Comparison of neural network error measures for simulation of slender marine structures. *Journal of Applied Mathematics*, *2014*, 1–11. https://doi.org/10.1155/2014/759834

Council of Economic Advisers. (2016). Benefits of competition and indicators of market power. Retrieved March 30, 2023, from https://obamawhitehouse.archives.gov/sites/default/files/page/files/20160502_competition_issue_brief_updated_cea.pdf

Council of the EU. (2023). Infographic - Where does the eu's gas come from? Retrieved May 13, 2023, from https://www.consilium.europa.eu/en/infographics/eu-gas-supply/

Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, *6:2*, 94–98. https://doi.org/10.7861/futurehosp.6-2-94

Diebold, F., & Mariano, R. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *20*, 134–44. https://doi.org/10.1080/07350015.1995.10524599

Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., & Tabona, O. (2021). A survey on missing data in machine learning. *Journal of Big Data*, *8*(1). https://doi.org/10.1186/s40537-021-00516-9

ENTSO-E. (2022). Single Day-ahead Coupling (SDAC). Retrieved January 29, 2023, from https://www.entsoe.eu/network_codes/cacm/implementation/sdac/

Foroni, C., Ravazzolo, F., & Rossini, L. (2023). Are low frequency macroeconomic variables important for high frequency electricity prices? *Economic Modelling*, *120*, 106160. https://doi.org/10.1016/j.econmod.2022.106160

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Jakaša, T., Andročec, I., & Sprčić, P. (2011). Electricity price forecasting – ARIMA model approach. *2011 8th International Conference on the European Energy Market (EEM)*, 222–225. https://doi.org/10.1109/EEM.2011.5953012

Jamasb, T., & Pollitt, M. (2005). Electricity market reform in the european union: Review of progress toward liberalization & integration. *The Energy Journal*, *26*, 11–41. http://www.jstor.org/stable/23297005

Jedrzejewski, A., Lago, J., Marcjasz, G., & Weron, R. (2022). Electricity price forecasting: The dawn of machine learning. *IEEE Power and Energy Magazine*, *20*(3), 24–31. https://doi.org/10.1109/mpe.2022.3150809

Kingma, D., & Ba, J. (2014). Adam: A method for stochastic optimization. *International Conference on Learning Representations*.

Kühling, J., Knapp, P., Kollmann, D., Nöcker, T., & Wamback, A. (2021). *Competition between electricity exchanges in germany* (Energy Sector Report of the Monopolies Commission No. 8). Monopolkommission. https://www.monopolkommission.de/images/PDF/SG/chapter-3-competition-between-electricity-exchanges.pdf

Lago, J., De Ridder, F., & De Schutter, B. (2018). Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy*, *221*, 386–405. https://doi.org/10.1016/j.apenergy.2018.02.069

Lago, J., Marcjasz, G., De Schutter, B., & Weron, R. (2021). Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy*, *293*, 116983. https://doi.org/10.1016/j.apenergy.2021.116983

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Little, B. (2021). How did humans evolve? Retrieved January 19, 2023, from https://www.history.com/news/humans-evolution-neanderthals-denisovans

Machlev, R., Heistrene, L., Perl, M., Levy, K., Belikov, J., Mannor, S., & Levron, Y. (2022). Explainable artificial intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. *Energy and AI*, *9*, 100169. https://doi.org/10.1016/j.egyai.2022.100169

McCulloch, W., & Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*(4), 115–133. https://doi.org/10.1007/BF02478259

Ministry of Petroleum and Energy. (2016). The power market and prices. Retrieved January 25, 2023, from https://www.regjeringen.no/en/topics/energy/the-electricity-grid/the-power-market-and-prices/id2076000/

NEMO Committee. (2020). Euphemia public description: Single price coupling algorithm. https://www.nemo-committee.eu/assets/files/euphemia-public-description.pdf

Nogales, F., Contreras, J., Conejo, A., & Espinola, R. (2002). Forecasting next-day electricity prices by time series models. *IEEE Transactions on Power Systems*, *17*(2), 342–348. https://doi.org/10.1109/TPWRS.2002.1007902

Norwegian Ministry of Petroleum and Energy. (2021). The power market. Retrieved January 20, 2023, from https://energifaktanorge.no/en/norsk-energiforsyning/kraftmarkedet/

Nowotarski, J., Raviv, E., Trück, S., & Weron, R. (2014). An empirical comparison of alternative schemes for combining electricity spot price forecasts. *Energy Economics*, *46*, 395–412. https://doi.org/10.1016/j.eneco.2014.07.014

Phi, M. (2018). Illustrated guide to LSTM's and GRU's: A step by step explanation. Retrieved February 10, 2023, from https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21

Rasekhschaffe, K. C., & Jones, R. C. (2019). Machine learning for stock selection. *Financial Analysts Journal*, *75:3*, 70–88. https://doi.org/10.1080/0015198X.2019.1596678

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "why should i trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. https://doi.org/10.1145/2939672.2939778

Rosenblatt, F. (1958). The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386–408. https://doi.org/10.1037/h0042519

Russell, S. J., & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach* (3rd ed.). Pearson.

Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena*, *404*, 132306. https://doi.org/10.1016/j.physd.2019.132306

Starmer, J. (2022). *Long Short-Term Memory (LSTM), Clearly Explained*. Youtube. Retrieved February 9, 2023, from https://www.youtube.com/watch?v=YCzL96nL7j0&ab_channel=StatQuestwithJoshStarmer

Statnett. (2021). Derfor har vi prisområder. Retrieved January 27, 2023, from https://www.statnett.no/om-statnett/bli-bedre-kjent-med-statnett/om-strompriser/fakta-om-prisomrader/

Staudemeyer, R. C., & Morris, E. R. (2019). Understanding LSTM – a tutorial into long short-term memory recurrent neural networks. *ArXiv*. https://doi.org/10.48550/arXiv.1909.09586

Szkuta, B., Sanabria, L., & Dillon, T. (1999). Electricity price short-term forecasting using artificial neural networks. *IEEE Transactions on Power Systems*, *14*(3), 851–857. https://doi.org/10.1109/59.780895

The Norwegian Energy Regulatory Authority. (2023). Norway and the european power market. Retrieved January 29, 2023, from https://www.nve.no/norwegian-energy-regulatory-authority/wholesale-market/norway-and-the-european-power-market/

Tjoa, E., & Guan, C. (2021). A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(11), 4793–4813. https://doi.org/10.1109/tnnls.2020.3027314

Trebbien, J., Rydin Gorjão, L., Praktiknjo, A., Schäfer, B., & Witthaut, D. (2023). Understanding electricity prices beyond the merit order principle using explainable AI. *Energy and AI*, *13*, 100250. https://doi.org/10.1016/j.egyai.2023.100250

Tschora, L., Pierre, E., Plantevit, M., & Robardet, C. (2022). Electricity price forecasting on the day-ahead market using machine learning. *Applied Energy*, *313*. https://doi.org/10.1016/j.apenergy.2022.118752

Ugurlu, U., Oksuz, I., & Taş, O. (2018). Electricity price forecasting using recurrent neural networks. *Energies*, *11*, 1255. https://doi.org/10.3390/en11051255

Uniejewski, B., Nowotarski, J., & Weron, R. (2016). Automated variable selection and shrinkage for day-ahead electricity price forecasting. *Energies*, *9*(8), 621. https://doi.org/10.3390/en9080621

VG. (2023). Strømprisen. Retrieved January 28, 2023, from https://www.vg.no/stromprisen/?utm_source=vgfront-personalized&utm_content=hovedlop_row6_pos1&utm_term=retargeting-stromprisen

Voronin, S., & Partanen, J. (2013). Forecasting electricity price and demand using a hybrid approach based on wavelet transform, ARIMA and neural networks. *International Journal of Energy Research*, *38*(5), 626–637. https://doi.org/10.1002/er.3067

Wang, L., Zhang, Z., & Chen, J. (2017). Short-term electricity price forecasting with stacked denoising autoencoders. *IEEE Transactions on Power Systems*, *32*(4), 2673–2681. https://doi.org/10.1109/TPWRS.2016.2628873

Werbos, P. (1990). Backpropagation through time: What it does and how to do it. *Proceedings of the IEEE*, *78*(10), 1550–1560. https://doi.org/10.1109/5.58337

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, *30*(4), 1030–1081. https://doi.org/10.1016/j.ijforecast.2014.08.008

Yan, X., & Chowdhury, N. A. (2015). Midterm electricity market clearing price forecasting using two-stage multiple support vector machine. *Journal of Energy*, 384528. https://doi.org/10.1155/2015/384528

Ziel, F. (2016). Forecasting electricity spot prices using lasso: On capturing the autoregressive intraday structure. *IEEE Transactions on Power*

Systems, *31*(6), 4977–4987. https://doi.org/10.1109/TPWRS.2016.2521545

Ziel, F., & Weron, R. (2018). Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Economics*, *70*, 396–420. https://doi.org/10.1016/j.eneco.2017.12.016

Zihan Chang, Y. Z., & Chen, W. (2019). Electricity price prediction based on hybrid model of adam optimized lstm neural network and wavelet transform. *Energy*, *187*, 115804. https://doi.org/10.1016/j.energy.2019.07.134

Appendix A

Correlation Heatmaps and European Bidding Zones

In this appendix we show correlation heatmaps of European electricity markets for the 2019 to 2022 period. Note that all BZNs are not represented in all four plots. During certain periods Pearson correlation coefficients may be undefined if the standard deviation between features is zero. This was the case for some of the BZNs which is why all are not represented in the plots. For all plots the BZNs of IT-GR, IT-North-SI, IT-North-CH, IT-Brindisi, IT-Foggia, IT-Priolo, IT-North-AT, IT-North-FR and ME were dropped. Moreover, in Figure A.1 UA-IPS, UA-BEI, IT-Calabria and NO2NSL were dropped, whereas IT-Calabria and NO2NSL were dropped in Figure A.2.

Finally, the appendix also includes a complete overview of all BZNs in the original dataset described in Section 3.1. Table A.1 provides information regarding the codes of all BZNs and their corresponding geographical location.

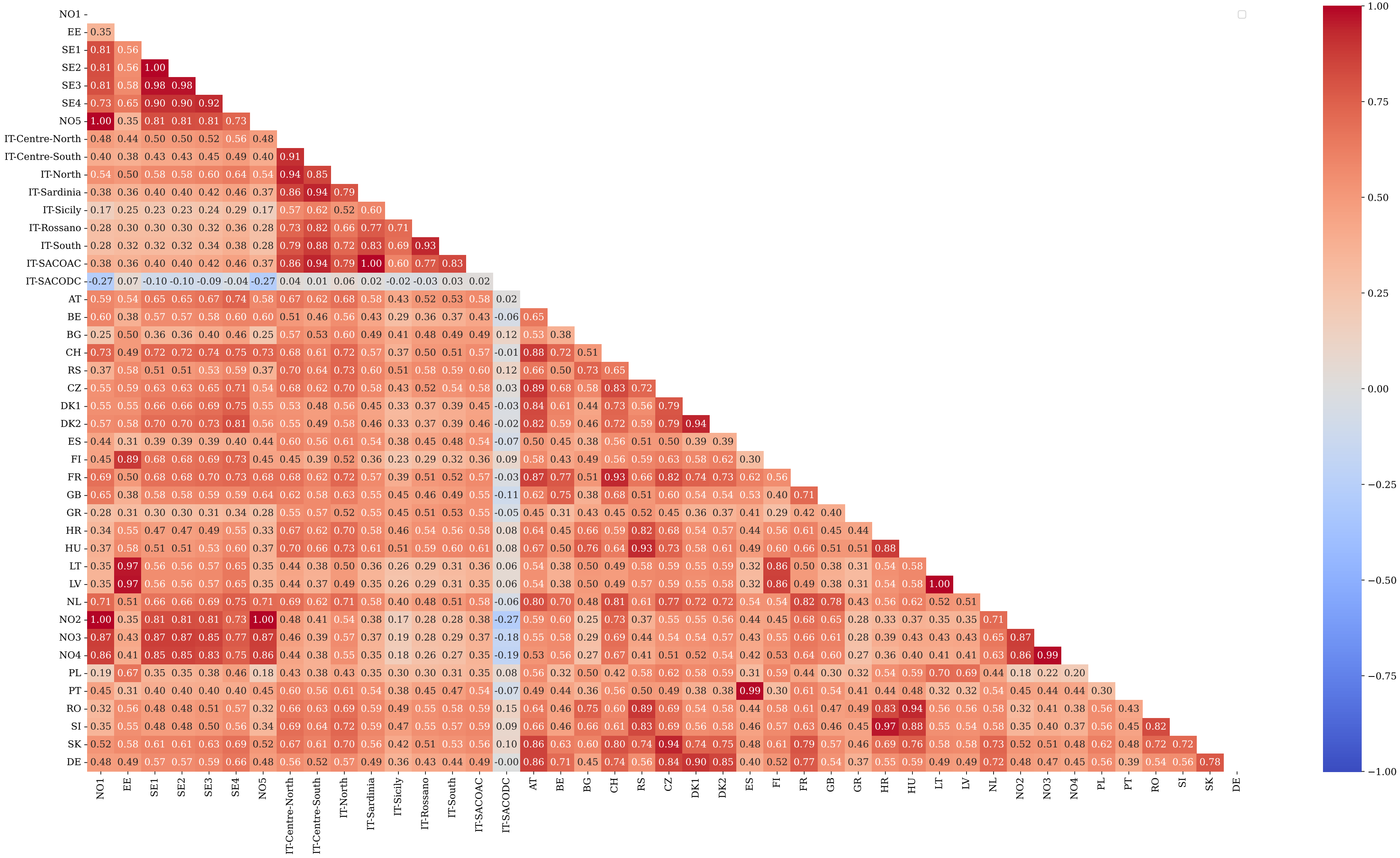


Figure A.1: Covariance matrix between the bidding zones in 2019.

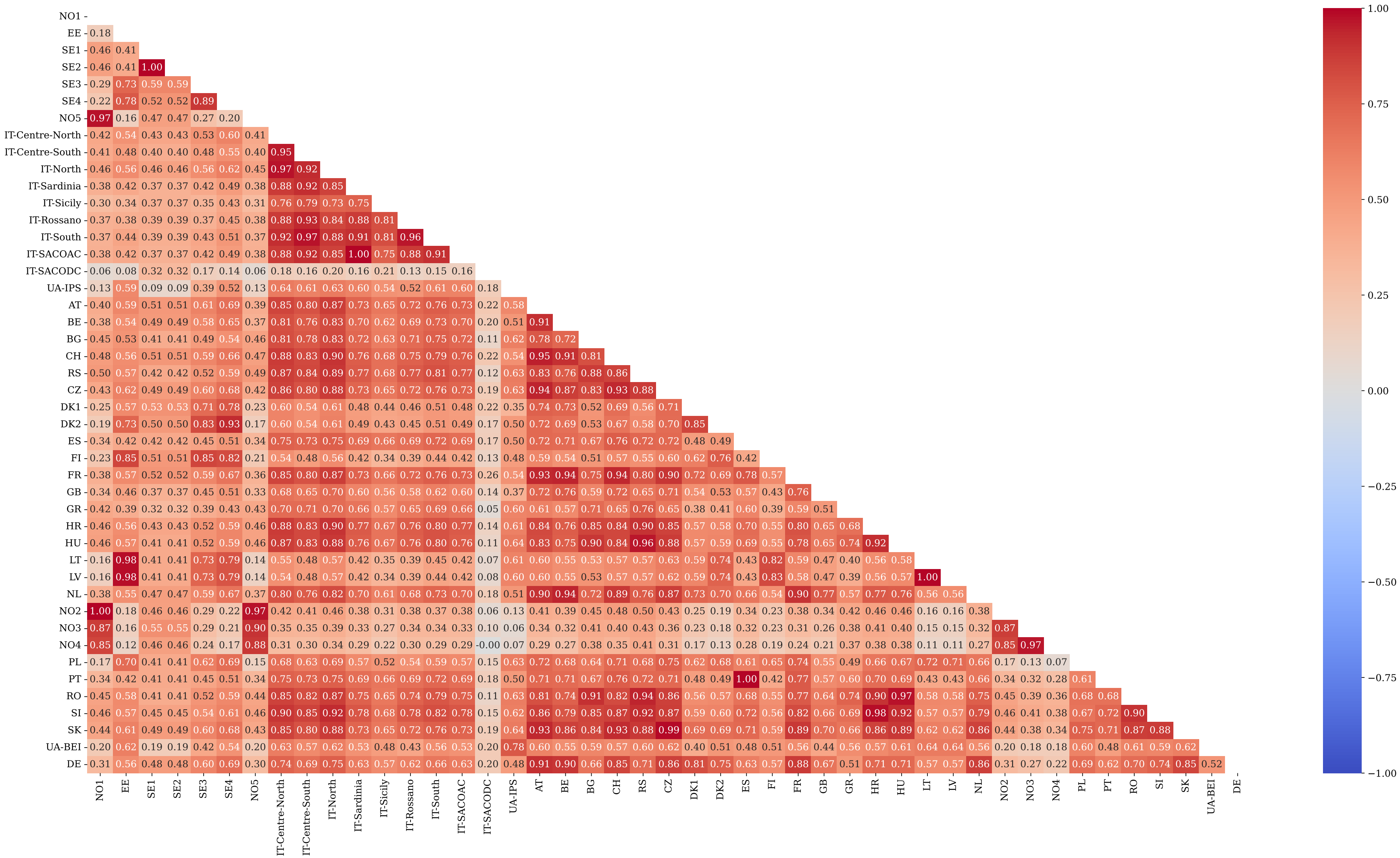


Figure A.2: Covariance matrix between the bidding zones in 2020.

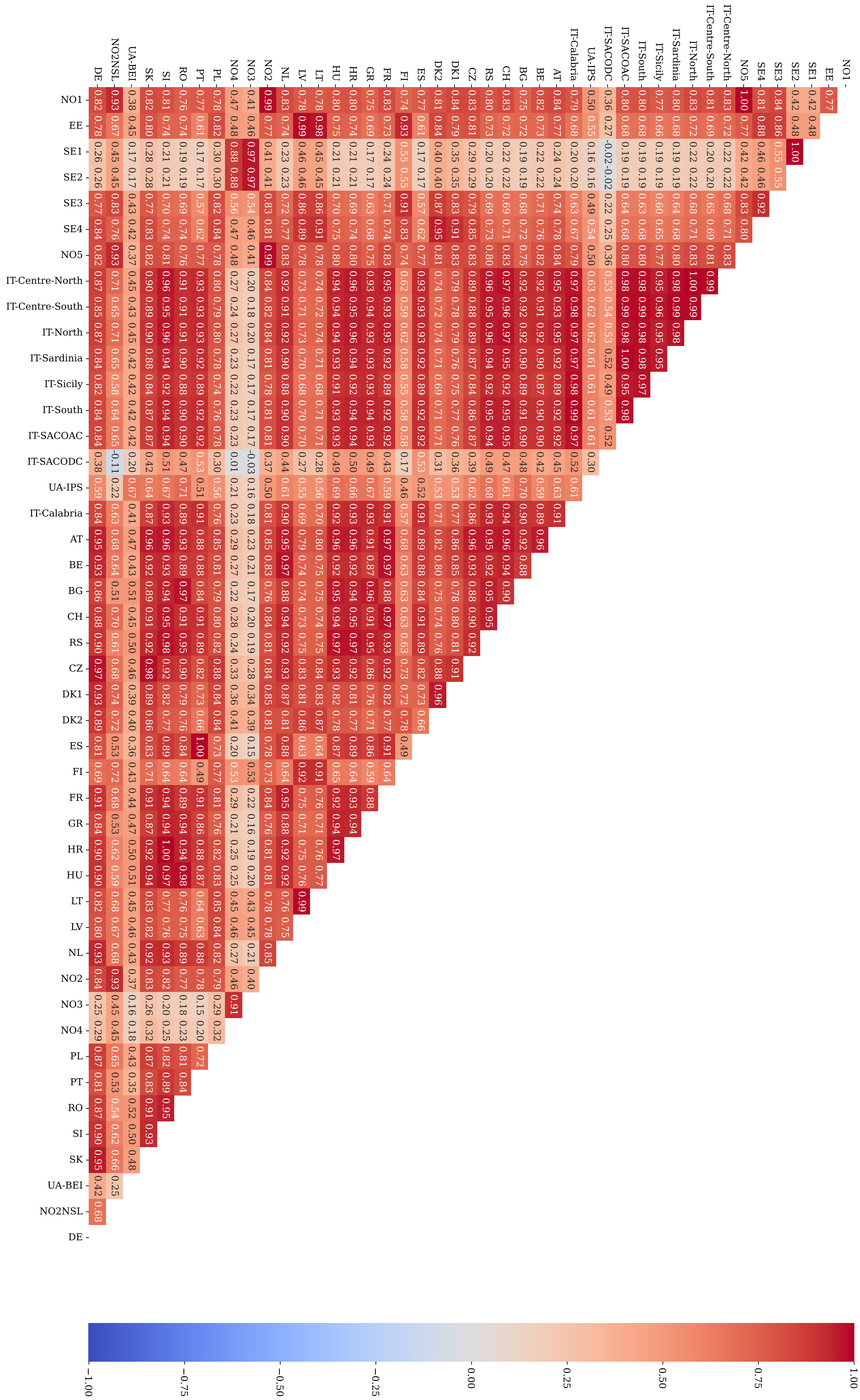


Figure A.3: Covariance matrix between the bidding zones in 2021.

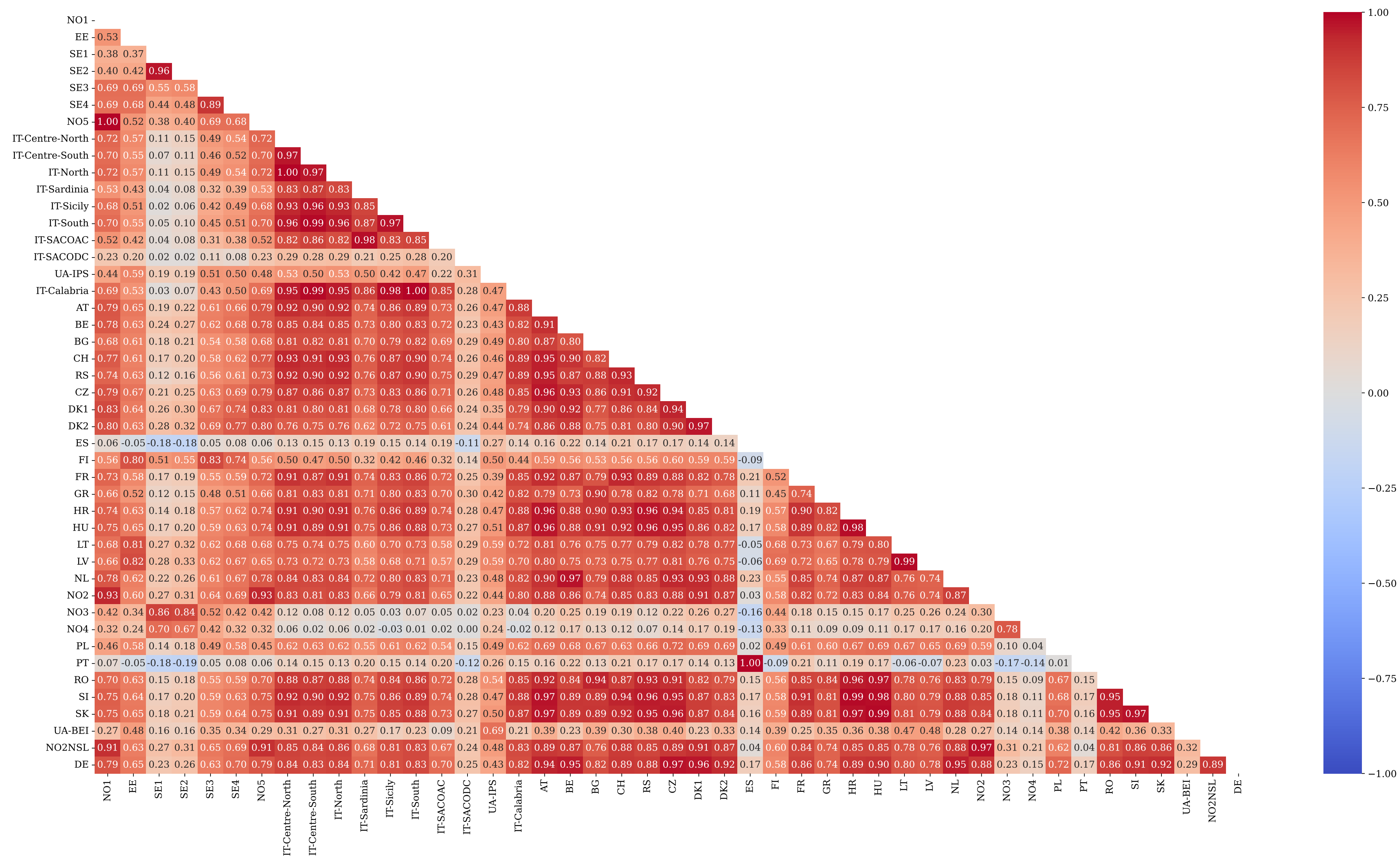


Figure A.4: Covariance matrix between the bidding zones in 2022.

Table A.1: Bidding zone abbreviations and corresponding regions.

Code	**Bidding Zone**
AT	Austria
BE	Belgium
BG	Bulgaria
CH	Switzerland
CZ	Czech Republic
DE-AT-LU	Germany + Austria + Luxembourg
DE-LU	Germany + Luxembourg
DK1	Denmark 1
DK2	Denmark 2
EE	Estonia
ES	Spain
FI	Finland
FR	France
GB	Great Britain
GR	Greece
HR	Croatia
HU	Hungary
IT-Calabria	Italy Calabria
IT-GR	Italy + Greece
IT-North	Italy North
IT-North-AT	Italy + Austria (North)
IT-North-FR	Italy + France (North)
IT-North-SI	Italy + Slovenia (North)
IT-North-CH	Italy + Switzerland (North)
IT-Centre-North	Italy Centre-North
IT-Centre-South	Italy Centre-South
IT-Sardinia	Italy Sardinia
IT-Sicily	Italy Sicily
IT-South	Italy South
IT-Brindisi	Italy Brindisi (Eliminated)
IT-Foggia	Italy Foggia (Eliminated)
IT-Priolo	Italy Priolo (Eliminated)
IT-Rossano	Italy Rossano (Eliminated)
IT-SACOAC	Italy
IT-SACODC	Italy
LT	Lithuania
LV	Latvia
ME	Montenegro
NL	Netherlands
NO1	Norway 1
NO2	Norway 2
NO2NSL	Norway 2 + North Sea Link to Great Britain
NO3	Norway 3
NO4	Norway 4
NO5	Norway 5
PL	Poland
PT	Portugal
RS	Serbia
RO	Romania
SE1	Sweden 1
SE2	Sweden 2
SE3	Sweden 3
SE4	Sweden 4
SI	Slovenia
SK	Slovakia
UA-BEI	Ukraine (Burshtyn Energy Island, synch ENTSO-E)
UA-IPS	Ukraine (Integrated Power System, synch Russia)