Optimal Allocation of Norwegian Offshore Wind Power: A Copula Approach

How can a thoughtful placement of offshore wind parks reduce variability in production output?

Henrik Heltne Alfsvåg, Sander Sollie

Supervisor: Geir Drage Berentsen

Master thesis, Economics and Business Administration

Major: Business Analytics

NORWEGIAN SCHOOL OF ECONOMICS

Acknowledgements

Norwegian School of Economics

Bergen, June 2023

_____   _____

Henrik Heltne Alfsvåg   Sander Sollie

Abstract

This thesis investigates how to optimize stable wind production along the coast of Norway. The research is carried out by studying how well a compound dependency model, consisting of a time series and copula model, for simulation of wind power data performs compared to historical data when optimizing a portfolio for wind power production areas. The weights for the areas in the portfolio are computed so that the areas with the most stable joint power production are included. The findings of this research will contribute to the understanding of how effective different optimization approaches for offshore wind park placements are and provide insights into the selection of optimal areas for offshore wind power development in Norway.

The study's findings indicate that portfolio optimization performed on simulated data performs better than on historical data. Consequently, zero and low production values are reduced, and stability is increased for the portfolio made with simulated data. Moreover, Value at Risk (VaR) is argued to be a better performance measure for stable wind production than variance. The portfolio distribution when maximizing VaR is more left-skewed than the portfolio minimizing variance. Thus, maximizing VaR results in a higher variance, but less zero and low production values, and a higher average production which is argued to be more important.

The positive effect of dispersed wind parks regarding stable wind production is evident. Following the pattern of diminishing correlation as distance increases, the optimal combination of wind parks includes places throughout the Norwegian coast. All areas are included in the optimal solution, but the most influential areas which should be prioritized are Sørlige Nordsjø 2, South of Kristiansand, West of Tromsø, and North of Tanafjorden. When the criteria for stable wind production is extended to include a penalty factor for low average production, diversification is partly de-prioritized to include areas with high average production, among these, more southern areas are included.

Keywords – Offshore wind, Norwegian wind conditions, renewable energy production, copula, time series, portfolio approach, NORA3-WP.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation and Purpose

In recent years the Norwegian and global energy markets have gained increased attention in public discussions for a number of environmental, political, and economic reasons. The Norwegian energy-producing facilities' increased contribution of energy supply to the continent highlights the importance of generating a stable energy flow both within the country's borders and abroad. Further, increased public environmental awareness and political actions to ensure green industries and emissions reduction emphasize the need to use environmentally friendly energy sources. Among the political measures taken, the Paris Agreement from 2015 might be one of the most important (United Nations, 2016). The agreement states that the overarching goal is to hold the increase in the global average temperature to well below 2 degrees above pre-industrial levels and pursue efforts to limit the temperature to increase to 1.5 degrees above pre-industrial levels. It is estimated that the global greenhouse gas emission must decrease by 43% by 2030 to achieve the goal. However, the proposals and opportunities for reaching the goals are countless and diverse, and conflicting views complicate the matter further. For this reason, research and expert opinions could play an important role in facilitating decision-making and shedding light on the fundamental features of the challenges ahead.

In the case of energy production in Norway, the country is known for its high proportion of green energy production through its hydropower plants. To reduce the country's emission levels, the government incentivizes the use of environmentally friendly solutions (Regjeringen, 2022a). In later years, consumers and industries have shifted from fossil-based equipment to more use of electricity-based solutions, for instance, cars and heavy machinery. With increasing electricity export and a high consumption within own borders, the need for increased power production arises. In the search for a solution, the Norwegian government has started the work of granting licenses for the development of offshore wind power plants for several production sites, where the sites Utsira Nord and Sørlige Nordsjø II are the first areas that are approved for production, with more areas pending for clearance. On orders from the Norwegian Ministry of Petroleum and Energy, the Norwegian Water Resources and Energy Directorate (NVE) delivered a report including 20 areas for further investigation in April 2023 (NVE, 2023).

The potential of offshore wind as a source of energy is high, but a common critique of the production method is the intermittent character of wind (Cai and Bréon, 2021). As a result, it is difficult to anticipate the production of wind parks over short time horizons. It makes it challenging for producers to determine how much electricity will be provided to the market. In a market like the electricity market, where prices are set based on short-term supply and demand, this detail may prove crucial for both the electricity prices in the market but also the profitability of the producers. Therefore, this thesis focuses on investigating methods to reduce uncertainty for Norwegian offshore wind power production at a high temporal resolution.

1.2   Research Question

Through several reports in the later years, NVE has guided the government in selecting areas to grant licenses for production. The selection criteria are based on a compound qualitative and quantitative analysis of an extensive list of considerations. However, despite how detailed and comprehensive the reports are, it is without a doubt hard to capture the full complexity of a matter while at the same time keeping realistic features in the estimates. The analyses from NVE elaborate thoroughly upon features like the coastal area needed for production, capacity, and integration in the power grid and park types. Through this thesis, we intend to add to the work done by NVE and others to improve the selection method of areas up for evaluation, mainly based on the criterion of stabilizing power output.

To stabilize the production of wind power, two aspects should be considered. Firstly, each location's power generation varies over time, and we expect time dependency between observations for each site. Secondly, we expect a varying degree of dependence between locations due to weather conditions. Utilizing these two dependencies, the power output may be stabilized by combining wind power production from several sites that balance out one another. Therefore the thesis is based on the following research question:

When determining weights in the optimization for a portfolio with the most stable production for offshore wind park areas in Norway, how well does a compound dependency model for the simulation of wind power data perform compared to historical data, and which areas are included?

The thesis is structured as follows. An overview of the Norwegian offshore wind park situation and prior research on the diversification of wind power production is presented in Chapter 2. The data utilized for the analysis is described in Chapter 3, along with how the data's characteristics affect the choice of methods used. A presentation of the methods is included in Chapter 4. The method consists of a compound dependency model with time series and copula models, followed by portfolio optimization. The complete analysis of the compound dependency model and portfolio optimization is covered in Chapter 5, together with the findings. In Chapter 6, the results are analyzed and interpreted in light of the method's limitations. Lastly, the conclusion for the research question and recommendations for further research are provided in Chapter 7.

2 Background

2.1 The Electricity Market - Norway and Integration with Other Markets

The Norwegian electricity market is part of a joint Nordic market along with Sweden, Denmark, and Finland which again are enclosed in the European electricity market via transmission connections to bordering countries (Energifakta Norge, 2022). Further, Norway transmits electricity to Germany and the UK through the Nord Link and North Sea Link cables. The European market coupling covers countries that stand for close to 90 percent of the European power consumption. The interconnected character of the market may allow for efficiency gains in power production and consumption through proper power allocation and utilization of power grid capacity. It may also be beneficial in getting a higher proportion of consumption based on renewable energy sources.

The bidding for the wholesale market is organized through the day-ahead market, where typical agents are power producers, brokers, power suppliers, energy companies, and large industrial customers (Nord Pool, 2023). An hourly price for the next day is set based on the buy and sell bids so that the market is cleared. To account for the imbalance in power production (e.g., weather changes) between the closing and opening of the bidding in the day-ahead market, the intra-day market has continuous trading until one hour before the operation. This allows the agents to correct potential mismatches in their bid position and actual production or consumption. The market structure highlights the importance for agents to predict the production amount precisely. An estimate at high temporal resolution is therefore valuable to accurately set the bids and increase the efficiency of the market.

Due to transmission inefficiencies in the Norwegian power grid, the country is divided into five pricing areas (Energifakta Norge, 2022). This means that the power price may differ between locations. The area prices reflect the regional differences in the power situation that vary from hour to hour and season to season. These differences may cause power scarcity in one area and surplus in another. The price difference occurs when the transmission capacity is too low to transfer the power required to account for the

imbalance in power demand and supply. When the transmission capacity is sufficient, the prices between areas will turn out the same. However, the evaluation of power grid capacity and power balance is outside the scope of the thesis.

2.2   Wind Power Production in Norway

With currently seven floating windmills, Hywind Tampen is the only notable site in the current offshore wind power generation on the Norwegian continental shelf (Equinor, 2023). However, the Norwegian government aims to become a leading nation in offshore wind power production (Regjeringen, 2022b). The Norwegian government acknowledges that offshore wind power will be essential in meeting the demand for renewable energy in Norway and Europe in the coming years. Their goal is to assign areas for offshore wind park installation with a capacity of 30 GW within the year 2040.

For now, the two areas, Utsira Nord (UN) and Sørlige Nordsjø 2 (SN2), are the only areas approved for the installation of offshore wind parks. However, in April 2023, the Norwegian Water Resources and Energy Directorate (NVE) delivered a report including 20 areas for further investigation (NVE, 2023). Birkeland et al. (2023) further investigated the effects Sørlige Nordsjø 2 will have on the Norwegian power market. The implementation of energy grid solutions and it's impact on Norway and Europe was discussed. It is concluded that wind power will contribute more energy to the Norwegian power system, but to a small degree help the power balance. Further, a growing share of wind power with intermittent character will require that the rest of the power system has available flexibility (Energifakta Norge, 2021). The intention is to present a method to reduce the issue of intermittency by stabilizing wind power production. By combining power production from some or all of the 20 areas, wind power could considerably help the power balance. Our analysis was performed before the 20 areas were identified, meaning that our research was done using 13 other locations. The method is, however, transferable to other areas, so the difference does not notably impact the relevance of the thesis.

2.3   Previous Research on Diversification of Wind Power Production

Several studies have examined how the geographic dispersion of wind farms or the size of regions affects the smoothing effect of wind power. The first to examine these effects for arrays of various sizes was Kahn (1979). He discovered that the output reliability of wind power rose as the array size grew but with a declining marginal advantage. Additionally, more recent research demonstrates that interconnected parks help to reduce aggregate variability like Archer and Jacobson (2007). They found that scheduled and unscheduled outage times for interconnected parks could match the outage time for coal plants in the US. Therefore, the intermittency in stand-alone parks is moderated by interconnecting parks on the grid.

The sequence of methods used in this thesis is inspired by an article by Grothe and Schnieders (2011) on optimal wind power allocation in Germany. The authors present a selection method of weights for the wind park portfolio that improves stability in output by reallocating capacity for German wind parks. The main features of the modeling are executed similarly on the time series in this thesis. Nevertheless, the study differs in multiple ways. In the article by Grothe and Schnieders (2011) the data is on wind speed, not wind power which may lead to slight modeling differences. The geographical area in question differs, which may affect the results. Further, the article takes on a reallocation of resources for existing wind parks, while this is not the case in the current study. This means we do not have a real-life benchmark scenario for comparison, which must be accounted for in the analysis when evaluating the results.

The thesis draws further inspiration from a master thesis written by Osnes and Nesheim (2023) and uses the same data set they analyzed to compare the methods used to optimize the portfolio. The modeling approach in our thesis differs from Osnes and Nesheim in how the dependency between areas is modeled. The intention is to contribute to a more involved dependency calculation and compose a model that may surpass the results from using historical data alone. The optimization goal is altered in our thesis to avoid violating assumptions for the methodology used. Both theses draw on the research by Solbrekke and Sorteberg (2022) including data from NORA3-WP.

3   Data

In the following section, we present the data and the descriptive features of the variables used for the analysis. The data pre-processing is performed to more thoroughly understand the behavior of wind power generation along Norway's coastal line. This establishes the foundation for the analysis' methodology.

3.1   NORA3-WP

The thesis is based on a selection of feasible wind park areas derived from the work in "NORA3-WP: A high-resolution offshore wind power data set for the Baltic, North, Norwegian, and Barents Seas" (Solbrekke and Sorteberg, 2022). NORA3-WP is an open-access data set designed for use in research, government management, and for stakeholders to obtain pertinent information on wind resources and wind power throughout the development phase of new wind farm projects. The data set provides an overview of 25 wind resources and wind power-related characteristics for three chosen turbines every month. Additionally, the data set includes hourly wind speed and wind power generation for the three selected turbines. For this thesis, hourly wind power generation is used for the analysis.

In NORA3-WP, the wind resource and wind power variables are based on hourly wind speed data, air temperature, and air pressure from NORA3 (Haakenstad et al., 2021). The wind power variables are calculated using three turbines with different specifications. Table 3.1 lists the specifications for the turbines. The three turbines are the floating turbine called SWT-6.0-154 from Siemens, DTU-10.0-RWT, the known reference turbine from the Technical University of Denmark, and IEA-15-240-RWT the new reference turbine from the National Renewable Energy Laboratory. As a result of the different specifications, maximum production and responses to wind speed vary between the three turbines. Wind farm effects or other disturbances that can affect wind power production are not accounted for, meaning that an assumption of a stand-alone wind turbine is made (Solbrekke and Sorteberg, 2022). As this thesis focuses on the relationship between areas and not between turbines, it is argued that the choice of turbine will not affect the result as long as the same turbine is used for all areas. Therefore, the reference turbine from the National

Renewable Energy Laboratory, IEA-15-240-RWT, is used for the rest of the thesis. For convenience, the IEA-15-240-RWT turbine will be referenced as the turbine.

	SWT-6.0-154	DTU-10.0-RWT	IEA-15-240-RWT
Rated power, C_r (W)	6 000 000	10 000 000	15 000 000
Hub height (m)	101	119	150
Rotor diameter (m)	154	178.3	240
Specific rated power C_r/A (Wm^{-2})	161.1	200.3	165.8
cut-in (ms^{-1})	4.0	4.0	3.0
rated (ms^{-1})	13.0	11.4	10.59
cut-out (ms^{-1})	25.0	25.0	25.0

Table 3.1: Specifications Turbines (Solbrekke and Sorteberg, 2022).

The period included in the data set is from 1996 to 2019. The grid resolution is 3 x 3 km, and the data spans the eastern parts of the Norwegian Sea, the North Sea, the Baltic Sea, and parts of the Barents Sea. This makes for a data set with 652 grid points in the longitude direction and 1149 grid points in the latitude direction (Solbrekke and Sorteberg, 2022). To understand the behavior of wind power generation along the Norwegian coast and locate beneficial placements that stabilize one another, it is argued that a representation of locations along the Norwegian coastline is satisfactory. Therefore, 13 grid points representing different parts of the Norwegian coast, retrieved from Osnes and Nesheim (2023), are chosen to be a part of the report.

The 13 locations are presented in Figure 3.1. The locations are selected based on a qualitative analysis with factors like ocean depth, distance to an electricity grid, nature reserves, wind speed, and production output. They found that the area outside the southern part of Norway had the best wind conditions. A majority of locations were therefore chosen from this area. Further, they found that the correlation between the northern areas and the others was low, giving a benefit when combined. Areas that were too deep were excluded, and areas close to existing electricity grids were prioritized. Further, areas in danger of destroying nature reserves were excluded. With these characteristics as deciding factors, the 13 areas were chosen to represent viable areas along the entire coast of Norway. The Norwegian government has already made it possible to apply for a concession for wind power projects in Utsira Nord and Sørlige Nordsjø 2. These two areas are therefore included among the 13 locations.
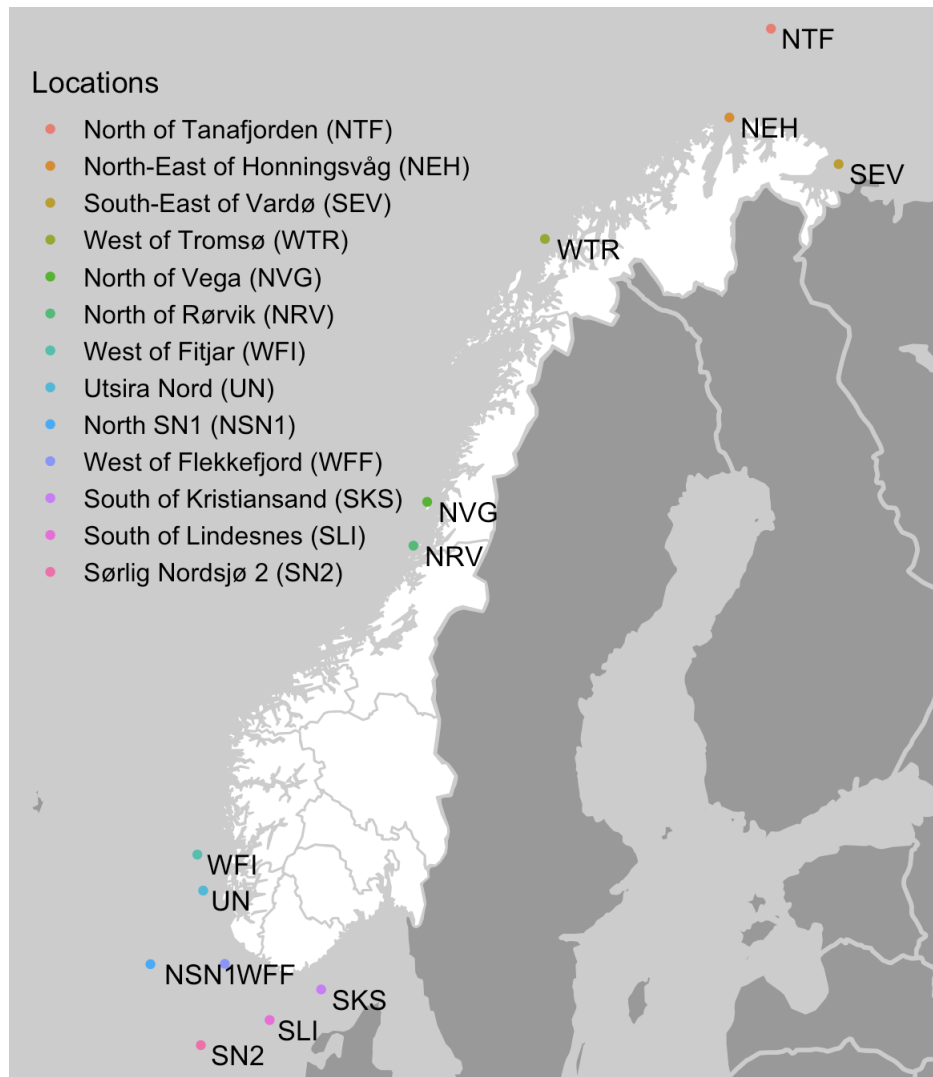
Figure 3.1: Map over the 13 locations used in the thesis.

In short, the data set analyzed in this thesis consists of hourly power production data from 13 areas along the coast of Norway between 1996 and 2019, estimated using the new reference turbine from the National Renewable Energy Laboratory, IEA-15-240-RWT.

3.2   Descriptive Statistics

To understand the behavior of wind power generation, we will investigate the descriptive statistics of the data in the following section. The data is based on the amount of power a turbine can produce over the time period specified. The turbines have limitations regarding maximum output and for which wind speeds power production is possible. The turbine specifications in Table 3.1 shows a cut-in value of 3.0 m/s, a cut-out value of 25.0 m/s, and a rated value of 10.59 m/s. This means that the turbine is not producing

power when the wind speed is lower than 3.0 m/s and when the wind speed is higher than 25 m/s. The rated value is a measure for which wind speed the turbine reaches maximum production: when the wind speed is higher than or equal to 10.59 m/s and less than 25.0 m/s, the maximum production level is achieved. This can be observed in Figure 3.2. Further, the figure shows how the power curve changes depending on the implementation of different storm control schemes through the lines for type 1 (SC1) or type 2 (SC2) storm control (Solbrekke and Sorteberg, 2022). The thesis does not account for the differences between various storm control schemes.
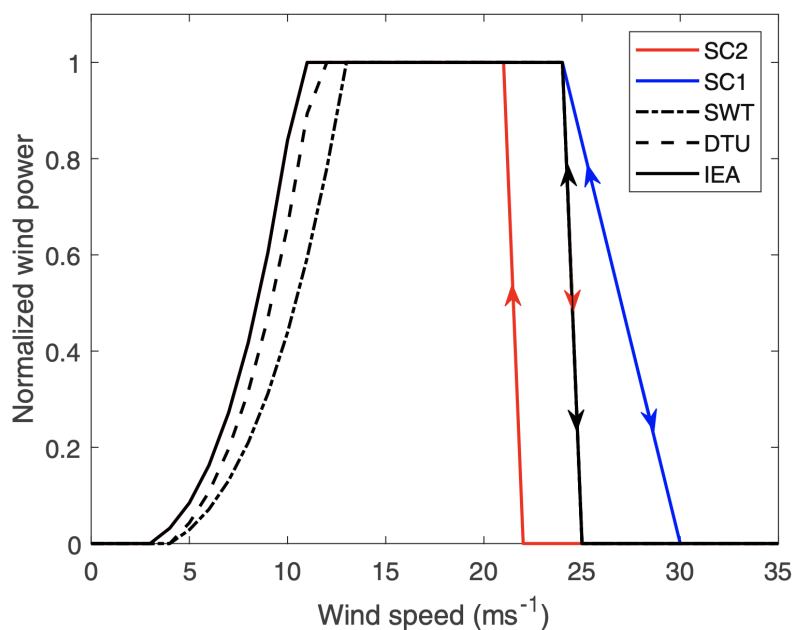
Figure 3.2: Power curves for the turbines (Solbrekke and Sorteberg, 2022).

As a result of the limitations in power production at different wind speeds, the distribution of the data is skewed. The nature of the data makes many values center around the maximum hourly output of 15 MW and 0 MW production, as observed in Figure 3.3. The limitations of the turbine set these upper and lower limits, leading to a non-normal distribution at the hourly level. Unlike data with a normal distribution, calculating the arithmetic mean and standard deviation will not guarantee representative values for the power production distribution. Therefore, a thorough descriptive analysis with adequate statistical features representing the data will be performed to understand the underlying characteristics of wind power production.
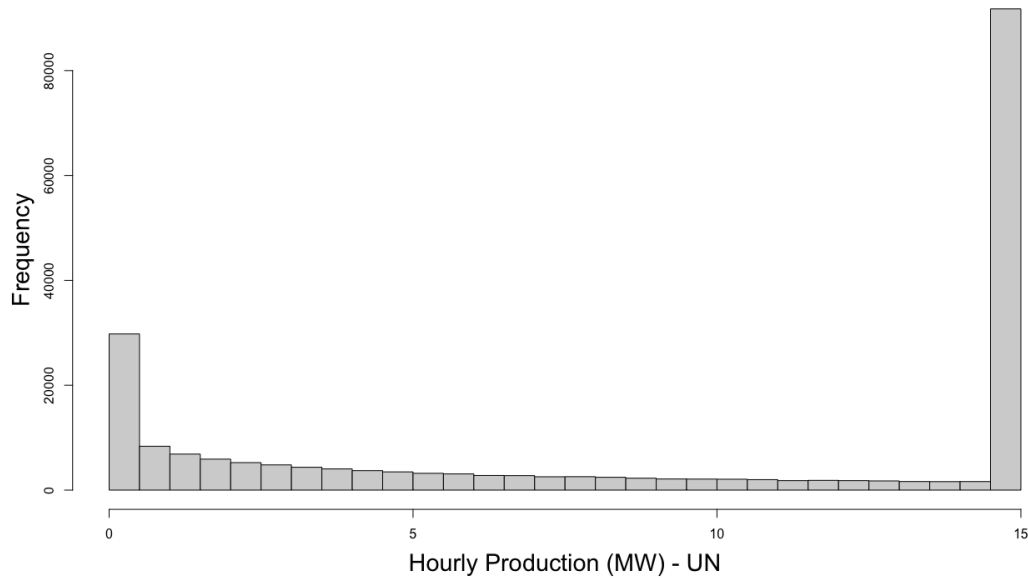
Figure 3.3: Hourly wind power production for Utsira Nord (UN).

The distribution of power production indicates large variability. This is further illustrated in the subsequent time series plots. The variability could potentially lead to high risk and problems in regard to balancing the demand and supply of the power market. To make decisions for the purpose of reducing the intermittency of power production, it is necessary to gain insight into how power production is behaving. Looking at how wind power behaves at different temporal resolutions assists in understanding the underlying characteristics of wind power production and further helps to reach the goal of reducing risk related to intermittency. In addition to the hourly power production, the data is aggregated, and the descriptive analysis is carried out at daily, weekly, and monthly time horizons. As observed in Figure 3.4, the distribution of power production changes when it is aggregated. On a daily level, the general shape remains with a slight increase in concentration around the middle values, while for the power production on a weekly level, the shape changes drastically, in closer resemblance to a normal distribution. This follows the Central Limit Theorem that says that the distribution of sample means is approximately normally distributed when the sample gets large (Ganti, 2023). Similarly, monthly values are also approximately normally distributed.
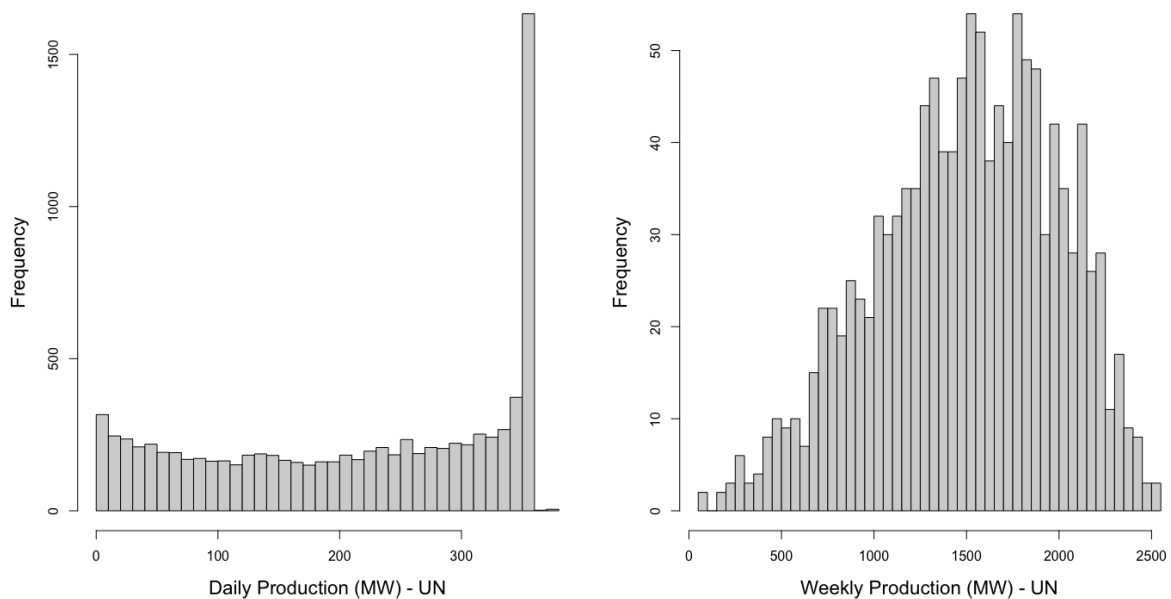
Figure 3.4: Daily and weekly wind power production for Utsira Nord (UN).

Table 3.2 shows descriptive statistics for all 13 locations on daily data. The mean of power production varies from approximately 180 to 239 MW in average power production, with the southern areas having the highest average. All areas have a negative kurtosis, indicating that the distributions should be light-tailed. Most areas also have a negative skewness, meaning a left-skewed distribution because of the concentration of values around the max production. For the areas with the lowest mean production, the skewness is positive because of a higher concentration of zero values. Corresponding observations are made for the hourly, weekly, and monthly data.

Areas	Mean	Sd	Kurtosis	Skewness
South of Lindesnes (SLI)	239.13 MW	113.46	-1.01	-0.57
Sørlige Nordsjø 2 (SN2)	234.85 MW	115.26	-1.07	-0.53
West of Flekkefjord (WFF)	233.06 MW	114.66	-1.08	-0.50
North of Sørlige Nordsjø 1 (NSN1)	229.73 MW	117.36	-1.15	-0.47
South of Kristiansand (SKS)	223.86 MW	116.50	-1.21	-0.38
Utsira Nord (UN)	214.00 MW	120.63	-1.32	-0.30
West of Fitjar (WFI)	209.86 MW	122.31	-1.36	-0.27
North of Tanafjorden (NTF)	207.08 MW	116.81	-1.32	-0.18
North-East of Honningsvåg (NEH)	206.86 MW	116.03	-1.31	-0.18
South-East of Vardø (SEV)	197.48 MW	117.56	-1.36	-0.06
North of Rørvik (NRV)	185.93 MW	119.15	-1.40	0.05
North of Vega (NVG)	182.13 MW	121.43	-1.42	0.08
West of Tromsø (WTR)	180.49 MW	118.14	-1.34	0.11

Table 3.2: Descriptive statistics of the locations based on daily data.

The average wind power production shows that the southern areas have the best wind conditions to produce the most power. However, high average production is not the only consideration. The irreversible character of investing in wind parks makes it important to have a balanced production. Once you build a wind park, you cannot move it, which requires long-term stability in power production to remain profitable. Short-term stability in production is required to ensure efficiency in the price and quantity provided to the electricity market. Therefore, we are interested in finding the dependency between the areas in an attempt to reduce the overall variability. Figure 3.5 show the correlation between the 13 areas. The locations are ordered with the southern areas on the left and the northern locations on the right. It is evident that locations close to one another correlate more strongly than locations that are far apart.
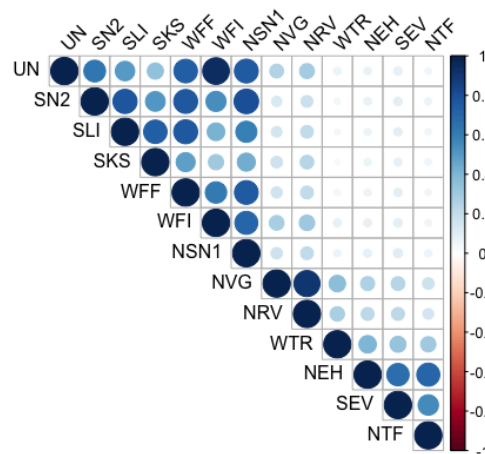
Figure 3.5: Correlation between the 13 areas.

Producing wind power in several areas with low correlation at the same time can be beneficial. This benefit results from the fact that there is a higher likelihood of other locations producing power while one place is not. To further investigate this effect, a regression is run where distance (km) is used as an explanatory variable to explain the correlation in power production between the locations. For all time horizons, the coefficient is significant and tells us that when distance increases, the correlation decreases. Figure 3.6 shows this relationship between distance and the correlation between the areas. The slope of the curve changes when distance increases and the effect is diminishing for longer distances. The diminishing effect means that there is a lower change in correlation for the longest distances than for short-distance areas. This difference in correlation between areas suggests that there are opportunities to diversify the risk (intermittency) of wind power production by including multiple locations in a joint portfolio of sites.
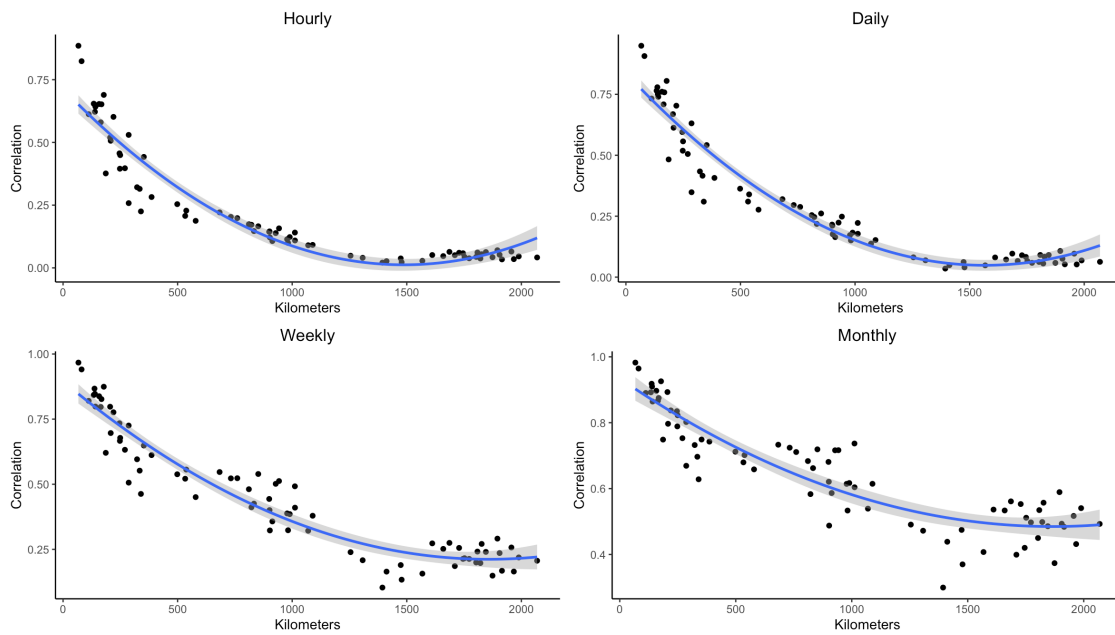
Figure 3.6: The relationship between distance and correlation across all 13 locations. Correlation in wind power production decreases when distance increases.
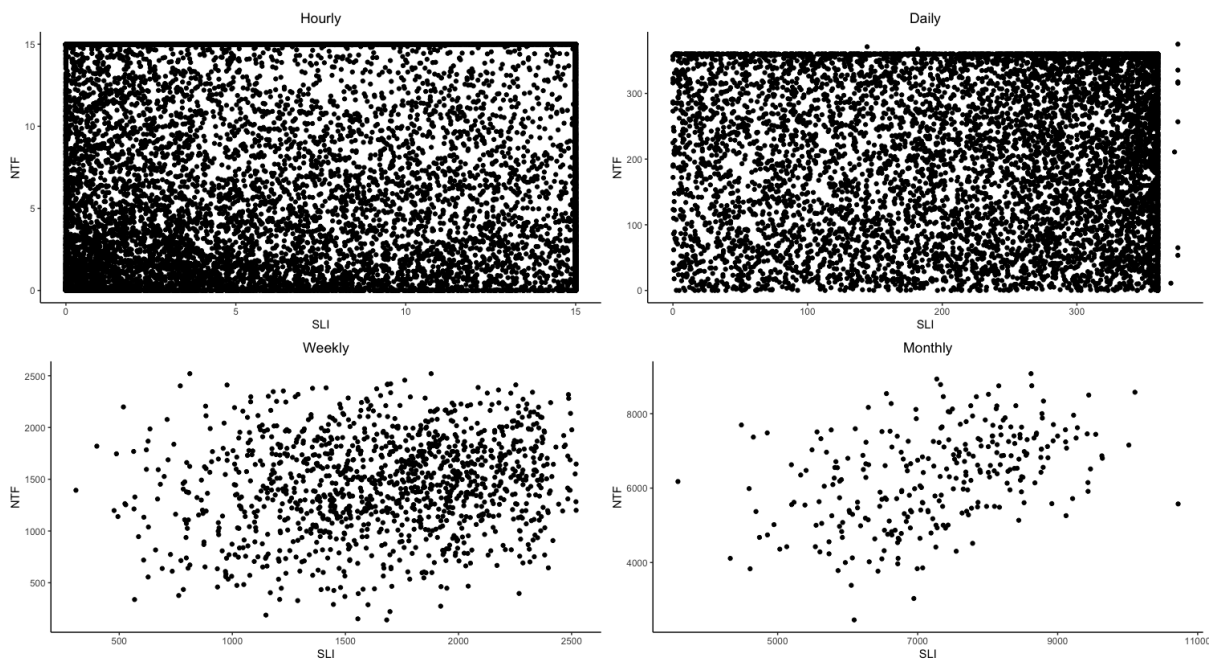
Figure 3.7: Scatterplot of the power production in South of Lindesnes (SLI) and North of Tanafjorden (NTF).
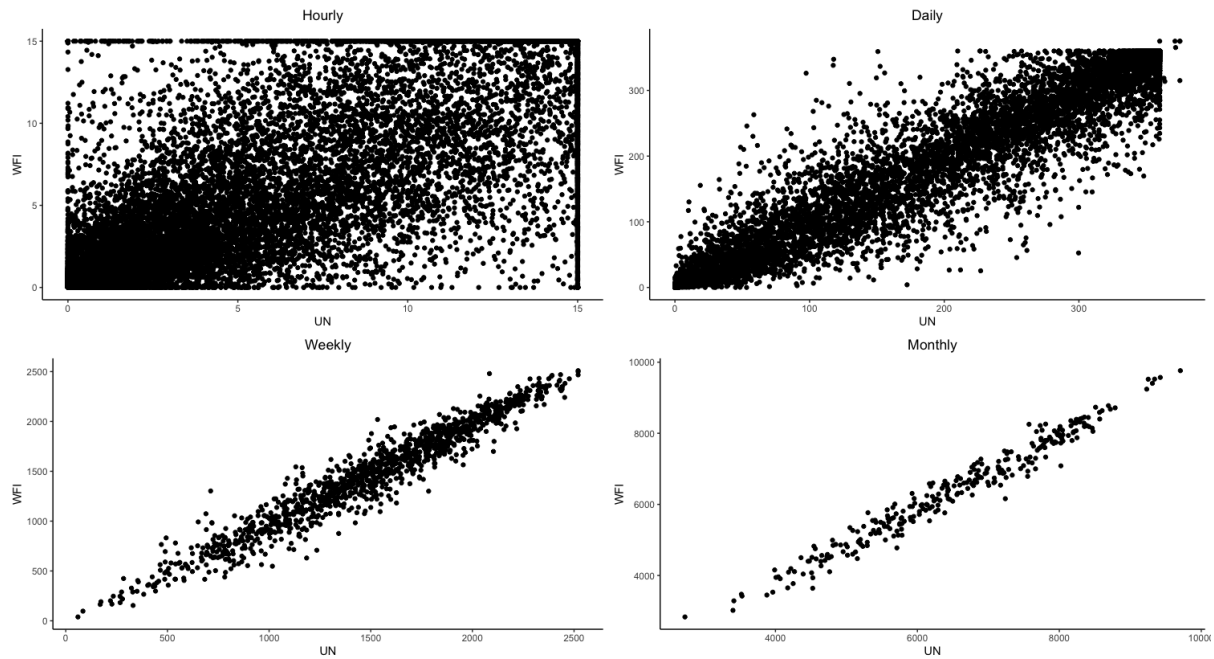
Figure 3.8: Scatter plot of the power production Utsira Nord (UN) and West of Fitjar (WFI).

As mentioned, the dependency between locations differs for the pairs. The scatter plots in Figure 3.7 and Figure 3.8 show the dependency between South of Lindesnes and North of Tanafjorden, and Utsira Nord and West of Fitjar, respectively. South of Lindesnes and North of Tanafjorden are the locations with the furthest distance apart and have a low correlation in wind power production. On the other hand, a clear correlation is evident between Utsira Nord and West of Fitjar, with high and low production values correlating. To be able to take advantage of these differences in dependencies, it is important to notice that there exists non-linear dependency in hourly and daily data. For the pair South of Lindesnes and North of Tanafjorden in Figure 3.7, the non-linear dependency is present in both hourly and daily production, while for the pair Utsira Nord and West of Fitjar in Figure 3.8, it is present in the hourly production. For all other pairs of locations, the same non-linear dependency is present in the hourly data and for many of the pairs in the daily data. The non-linear dependency can be observed as clustering in the scatter plots. There are clusters throughout the maximum production line, which is partially a function of the constraints of the turbines. Further, there are more concentrated observations for the lower values in the hourly data and more concentrated observations for higher values for select pairings of the daily data. These are clusterings that are not well explained by linear relationships. When modeling the dependency between locations, this must be taken into consideration.
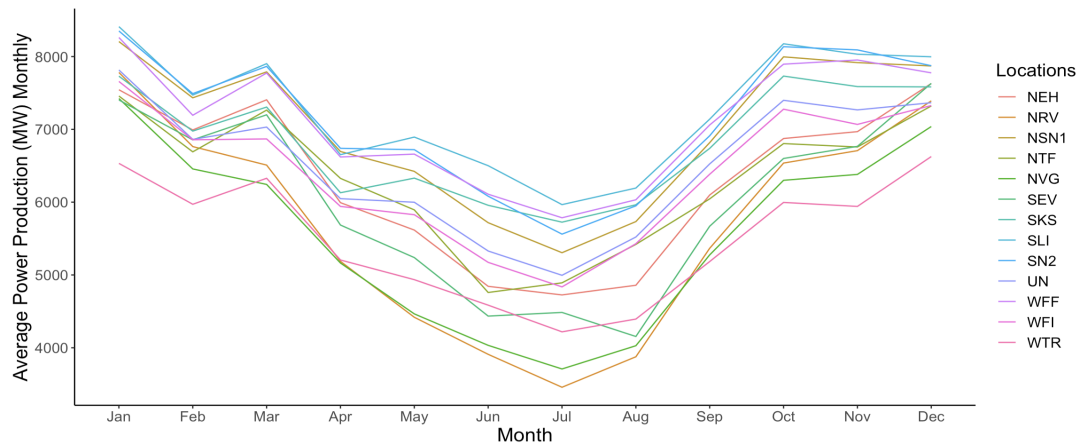
Figure 3.9: Average power production (MW) monthly. Shows seasonal elements in data.

The wind conditions for the areas are different across seasons, leading to a difference in average power production throughout the year. From Figure 3.9, it is apparent that power production peaks during the winter months and drop during the summer months. This indicates a seasonal pattern spanning over one year for all locations. The seasonal effect seems predictable as the variations occur for all locations, but to varying extents.

3.3   Time Series and Dependency Structure in the Data

The data demonstrate significant auto-correlation between lags, which makes the data non-stationary and dependent over time (Figure 3.10 and 3.11). In other words, an observation in time t in the series can be expressed as a function of observations from previous t's. To account for this in the analysis we attempt to fit models that can capture the time dependency in the observations.
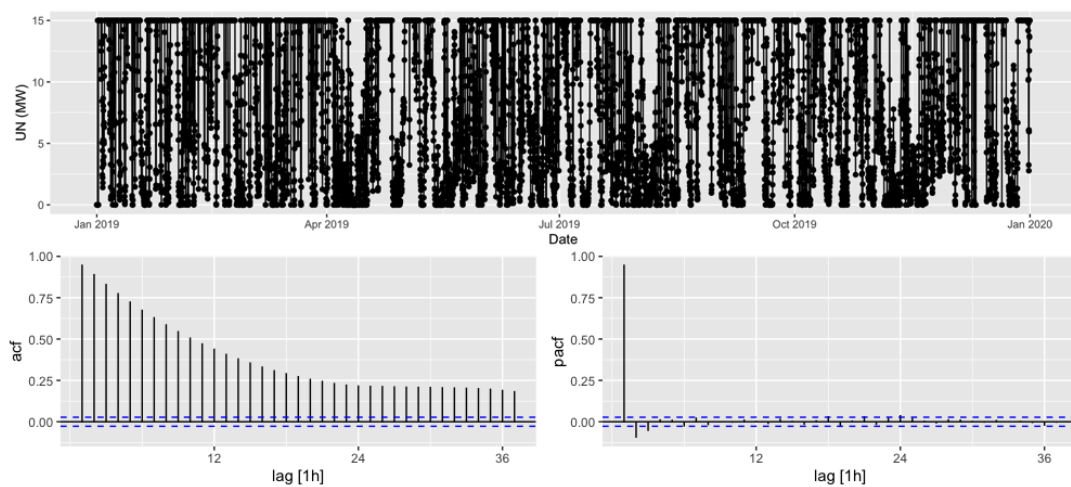
Figure 3.10: Autocorrelation plot Utsira Nord (UN) hourly.
Shows autocorrelation between lags. Extract of time series included (Jan 2019 - Dec 2019).
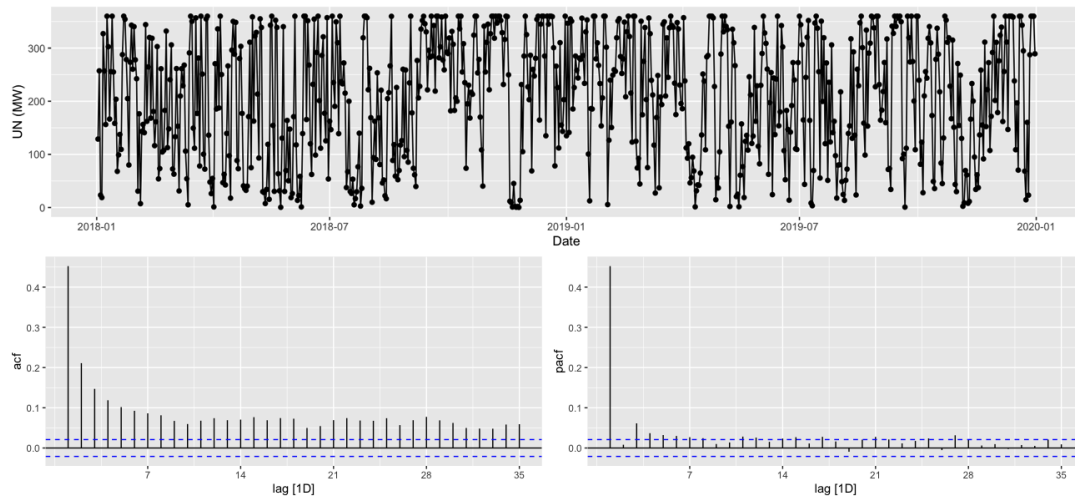
Figure 3.11: Autocorrelation plot Utsira Nord (UN) daily.
Shows autocorrelation between lags. Extract of time series included (year 2018 - 2019).

After presenting and investigating the data, it is apparent that we are dealing with non-normal data. The marginal distribution of wind power production is non-normal partly because of wind characteristics and partly because of the restrictions imposed by the wind turbine's upper and lower capacity. Additionally, we have found non-linear dependency structures between the locations. This should be accounted for to make a model that sufficiently explains how areas can be combined to stabilize wind production. The following section elaborates on time series and copula methods as ways to model these aspects of the data before it can be applied for portfolio optimization.

4   Methodology

The thesis investigates how different methods of modeling the dependency between wind power production sites can contribute to reduce variability in power production. The wind power data used is sequential and exhibit a serial correlation, meaning that there are time dependencies that can be modeled. We observe that the data also show non-linear dependency structures between the locations. The first part of the subsequent section elaborates on the methods used to model these dependencies through time series models and bi-variate copula models to make simulated wind power production. The last part of the methodology section elaborates on different optimization methods used to analyze to what extent production locations should be included in the optimal power production portfolio.

4.1   Time Series Modeling

4.1.1   STL Decomposition and Season NAIVE Model

A Seasonal and Trend decomposition using Loess (STL) is performed to get an impression of the components of the time series. The method was developed by Cleveland et al. (1990). It allows for a versatile decomposition method with all types of seasonality and allows for changes in the seasonal component over time. The components for the decomposition are an expression of the season S_t, trend T_t, and the remainder R_t for an observation. The components are then modeled with a seasonal NAIVE model assuming additive components. The model can be expressed as $y_t = S_t + A_t$, where A_t is the seasonally adjusted data consisting of the trend T_t and the remainder R_t.

4.1.2   Dynamic Harmonic Regression and ARMA Model

Dynamic Harmonic Regression is performed to model the dependence between lags in the wind power time series. The method uses Fourier terms to account for more extended period seasons that are not easily handled by methods such as the seasonal ARMA. The latter better handles shorter-term seasonal patterns (Hyndman and Athanasopoulos, 2018). The Fourier seasonal component of the series is complemented by ARMA errors to

model the short-term dynamics of the series. The resulting model should therefore be able to capture both the long and short-term patterns in the series. The models are fitted separately to each series, and the coefficients are selected for the model with the lowest Akaike Information Criteria (AIC). We use a step-wise algorithm proposed by Hyndman and Athanasopoulos (2021) to estimate the coefficients for the AR- and MA-processes in the series. The documentation of the algorithm can be found in the ARIMA function from the r-package *fable* (O'Hara-Wild et al., 2022).

The model has an intercept, a seasonal component S_t, and an error term. S_t is expressed as Fourier terms and uses sin and cos terms to smooth the seasonal element of the time series. If m is the seasonal period, the first few Fourier terms will be given by the following equations for time t (Hyndman and Athanasopoulos, 2021):

$$x_{1,t} = \sin\left(\tfrac{2\pi t}{m}\right), \quad x_{2,t} = \cos\left(\tfrac{2\pi t}{m}\right), \quad x_{3,t} = \sin\left(\tfrac{4\pi t}{m}\right), \tag{4.1}$$

$$x_{4,t} = \cos\left(\tfrac{4\pi t}{m}\right), \quad x_{5,t} = \sin\left(\tfrac{6\pi t}{m}\right), \quad x_{6,t} = \cos\left(\tfrac{6\pi t}{m}\right). \tag{4.2}$$

The maximum number of sin and cos pairs is $m/2$. The equivalent of using $m/2$ pairs is creating $m-1$ dummy variables for the periods. An advantage of using Fourier terms is that it often requires fewer variables than for the dummy approach, and the efficiency of the estimate increases. A downside is that the seasonality is constant. How many sin and cos terms K are used to model the seasonality for a given location is judged by the model's fit through the AIC. The entire model can be expressed in mathematical terms as:

$$y_t = bt + \sum_{j=1}^{K} \left[\alpha_j sin\left(\frac{2\pi jt}{m}\right) + \beta_j cos\left(\frac{2\pi jt}{m}\right) \right] + \eta_t \tag{4.3}$$

where η_t is an ARMA(p,q) process:

$$\eta_t = \sum_{i=1}^{p} \phi_i \eta_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t \tag{4.4}$$

The AR and MA coefficients ϕ and θ constitute the estimate's dependence on previous observations as well as previous errors ϵ. In the Dynamic Harmonic Regression, the AR coefficients are based on the part of the observation not already accounted for by the seasonal Fourier term.

In cases for which the seasonal component is less prominent in the data, judged by the selection criteria, an ARMA(p,q) process with a constant is selected:

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t \tag{4.5}$$

4.2   Vine Copula Theory

To model a high-dimensional dependence structure, a D-vine copula structure is used. The regular vine copulas (R-vine) method was initially introduced by Joe (1996). This method extends the use of bi-variate copulas to higher dimensions. The result is a more flexible dependence modeling between uni-variate distributions than the multivariate copula approach, which assumes a homogeneous dependence structure across the dimensions. The flexibility is added by fitting different copula families across dimensions to capture the dependence between variables.

A copula is a function that enables the joint distribution to be represented in terms of its marginals and the dependence structure among the marginals (Aas et al., 2009). Consider having a d-dimensional vector $X = (X_1, ..., X_d)$ with the joint distribution function $F = F(x_1, ..., x_d)$ with marginals $F_1 = F_1(x_1), ..., F_d(x_d)$. According to Sklar's theorem (1959), the copula associated with F is a d-dimensional distribution function C: $[0, 1]^d \to [0, 1]$ that satisfies $F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d); \beta)$. β is the parameter vector measuring the dependence between the marginals (Kim et al., 2013). The joint density function can be obtained from the previous equation as: $f(x_1, ..., x_d) = c_{1...d}(F_1, ..., F_d(x_d)) \cdot \sum_{k=1}^{d} f_k(x_k)$, where $c_{1...d}(\cdot)$ is a uniquely identified d-variate copula density. The density can be written as a product of pair copulas using conditional marginal densities (Aas et al., 2009).

Vine copulas allow for representation of the density for a d-dimensional multivariate distribution using $d(d-1)/2$ bi-variate copulas in a hierarchical manner. The first $d-1$ copulas have the dependence structures of ordinary bi-variate copulas, while the remaining copulas are in the form of conditional bivariate distributions. Researchers frequently use C- and D-vine copulas due to their restricting of possible vine structures, which serves well for computational purposes, interpretation, and to avoid issues with overfitting (Kim et al., 2013). In this thesis, we focus on the D-vine structure. The D-vine copula for d

dimensions uses $d - 1$ levels or trees to represent the full dependence structure. For the D-vine, the trees are structured so that each node has a maximum of 2 degrees - edges attached to a node. For a graphical representation of the tree structure of a D-vine copula, consult figure 5.8 on page 34.

The copula model is fitted to the residuals ϵ_t in equation 4.4 or 4.5, depending on the time series model that has been fitted for the areas. The objective is to model any dependency that cannot be represented as a function of prior observations; hence the residuals ϵ_t are used as input and transformed to the marginal interval using the empirical cumulative density function. Further, the resulting model will be used to create simulated wind power data. The simulated series consists of the time series estimates with sampled residuals for each area from the joint distribution to generate new observations. We use the simulated series to broaden the data basis for the optimization as an alternative approach to optimizing based on historical data. The optimization method is presented further in the next section. From the copula model fit, N samples are drawn for the variables in the joint distribution to form series for approximately 1000 years. The samples are expressed in the form of the marginal densities and are therefore transformed back using the quantile function (F_i^{-1} : *inverse of the marginal*). The resulting distributions are sampled residuals where the dependency between the variables is intact. Next, the 1000 years of wind power data are constructed using the model specifications from the time series analysis, with starting values from the original series and sampled residuals from the D-vine copula.

4.3   Portfolio Theory

Optimization of wind park portfolios is performed for the final part of our analysis. The optimization is performed using both historical data and simulated data, with the aim to analyze the impact of different inputs on the optimization. Further, a comparison of performance between different optimization methods is performed. The methods used are Markowitz minimum variance portfolio, unconstrained non-linear optimization of 5% VaR, and an unconstrained non-linear optimization of 5% VaR with penalized average return, and will be presented in the following section.

4.3.1   Markowitz Minimum Variance Portfolio

Markowitz (1952) introduced the Efficient Frontier (EF) as a financial tool to help investors select a desirable portfolio of assets. This is the foundation of the mean-variance analysis, one of the most frequently used portfolio optimization methods. The EF is based on a rule where the investor finds the expected return a desirable feature and the variance of the return undesirable. Following Markowitz, the mean-variance optimization problem will be presented in the following section (Würtz et al., 2015). To lower risk in the portfolio, the following objective function is used:

$$\min_{w} \ w^T \hat{\Sigma} w \tag{4.6}$$

$$\text{s.t.} \ w^T \hat{\mu} = \bar{r} \tag{4.7}$$

$$w^T 1 = 1 \tag{4.8}$$

The objective function 4.6 expresses that variance-covariance risk measure is to be minimized $\bar{\sigma}^2 = w^T \hat{\Sigma} w$, where the matrix $\hat{\Sigma}$ is an estimate of the covariance of the assets. Covariance is a statistical measurement used to establish how two random variables vary with respect to each other (Hayes, 2022). The first constraint (4.7) ensures that any return goals $\bar{r}$ are met, and it is expressed by multiplying the mean of the assets $\hat{\mu}$ with the respective weights invested. The vector w is the individual investments subject to the constraint in equation 4.8 that ensures that all available capital is invested. Based on this, the EF is calculated and shows the feasible region of portfolios. All portfolios on the EF are combinations of portfolios where no other portfolios give the same return with the same or lower risk. A variant of Markowitz's mean-variance method, minimum variance, is used for our optimization. The minimum variance portfolio is the portfolio with the lowest risk on the EF. The set of weights in the portfolio in the minimum variance portfolio can be expressed as: $w_* = \frac{\Sigma^{-1} 1}{1^T \Sigma^{-1} 1}$. The minimum variance portfolio is made with the R-package *fPortfolio* (Wuertz et al., 2020).

To use covariance to calculate the risk between assets in the portfolio, an assumption of bivariate normally distributed assets and linear relationships between the variables is made (Kent State University Libraries, 2023). As observed in the data section, this is not the case for the wind power data. To account for the non-normal data, an unconstrained

non-linear optimization method that maximizes the portfolio distribution's lower quantiles is introduced.

4.3.2   Unconstrained Non-linear Optimization of VaR

We use Value at Risk (VaR) as our performance measure to maximize the lower quantiles. Value at Risk is most commonly used in financial portfolios. VaR is a measure of the value of the observation at a particular quantile of a distribution. The interpretation of VaR in financial data is the probability of a potential loss (Kenton, 2023). Financial data have negative values in the lower tail, while wind power only have positive values. This changes the interpretation slightly: we want to stabilize wind power production, and by maximizing the 5% VaR, we make the values in the lower quantiles of the portfolio distribution as high as possible. Hence, zero and low wind power production events are minimized. Applying VaR also handles the issue of non-normality, as measuring the lower tail is not dependent on the distribution. Based on the previous, we argue that VaR seems like an appropriate performance measure for evaluating stability.

A non-linear unconstrained optimization method is used to maximize the VaR. A non-linear optimization problem is when either the objective or one of the constraints are non-linear functions of the decision variables (Bradley et al., 1977). Making the decision variables not restricted between any values makes the optimization unconstrained. We aim to maximize the 5 percent quantile of the portfolio distribution as shown in the objective function (4.9). The portfolio Y_t is made by summing the weighted wind production from each of the 13 areas i at time t as demonstrated by equation 4.10. X_{it} is the wind production from the k areas and w_i is the weights of each area in the portfolio. For the optimization, we are maximizing the lower quantiles for both the historical data and the simulated data made through time series and copula as explained in Chapter 4.2. Depending on which data is used, X_{it} represent simulated or historical wind power production. By finding optimal allocation for historical data and simulated data, we are able to compare the performance of the two methods when applying the weights to unseen test data.

$$\max_{v} \; Quantile(Y_t, 0.05) \tag{4.9}$$

$$Y_t = \sum_{i=1}^{k} X_{it} w_i \tag{4.10}$$

$$\text{w.r.t:} \qquad w_i = \frac{[1, exp(v_1), ..., exp(v_{k-1})]}{1 + \sum_{i=1}^{k-1} exp(v_i)} \tag{4.11}$$

$$\text{defined by} \qquad (v_1, ..., v_{k-1})$$

$$\text{which ensures} \qquad \sum_{i=1}^{k} w_i = 1 \;\; \text{and} \;\; w_i \in (0,1) \tag{4.12}$$

To make the optimization unconstrained, the weights w_i are calculated as shown with equation 4.11, defined by the decision variables in vector v. By having $k-1$ decision variables, but creating k weights in eq. 4.11, the calculation makes the weights implicitly meet the conditions in 4.12, causing the weights to sum to 1 and have values between 0 and 1. When the decision variables in vector v is used in equation 4.11 to optimize the objective function (4.9), the decision variables are not restricted between any values and are chosen freely. For the practical implementation of the optimization in R, the function *optim* from the package *stats* is used (R Core Team, 2022).

The BFGS algorithm is used for the optimization. The BFGS optimization algorithm belongs to the category of algorithms known as Quasi-Newton methods. The method leverages the second-order derivative of the objective function to locate optima and is one of the most used second-order methods for numerical optimization (AICorespot, 2021). The BFGS algorithm demands initial weights as an estimate of the optimal weights. For each step, the algorithm converges towards the objective function's optima. The initial weights can therefore influence the result as different optima can be located based on the initial weights. We find no information suggesting that unevenly distributed initial weights should be chosen for the wind park locations in question. Therefore, for the optimization problems in the thesis, evenly distributed weights summing to 1 are selected. By applying evenly distributed weights, all areas have equal possibilities for reaching high and low percentage involvement in the portfolio. Since the decision variables are transformed within the optimization, we must account for this to provide the correct input for initial weights. To find the initial weights v, the following equations are used:

$$w = (w_1, \ldots, w_k) = \left(\frac{1}{13}, \ldots, \frac{1}{13} \right) \tag{4.13}$$

$$v = \left(log \left(\frac{w_2}{w_1} \right), \ldots, log \left(\frac{w_k}{w_1} \right) \right) \tag{4.14}$$

4.3.3   Unconstrained Non-linear Optimization of VaR with Penalized Average Return

In the previous sub-chapter, a model improving stability based on the maximization of VaR was presented. As an extension of the previous model, a model including a penalty factor for average production below a threshold P is introduced. The method offers an approach to balance the importance of minimal zero and low production values, and a high average production. The maximization of 5% VaR is still included, but if a portfolio has a lower average production than the threshold, a negative value is added to the objective function. This makes the optimization more inclined to select portfolios producing close to or over the threshold. The alternative objective function is presented in equation 4.15. The rest of the optimization is the same as in equations 4.10 through 4.12. At the end of the objective function, a constant is added to scale the importance of the penalty factor. The penalty factor is multiplied by s to scale the impact of the penalty so that the objective function prefers portfolios with higher average production. The threshold P and scaling factor s are determined in the analysis.

$$\max_v \; Quantile(Y_t, 0.05) + min(0, (mean(Y_t) - P)) \cdot s \tag{4.15}$$

4.4 Method for Optimization on Simulated and Historical Wind Power Data

For clarity, the sequence of methods used in the thesis is presented in the following list. The list explains the process of making N observations of simulated wind power production through drawing residuals from the copula and substituting it into the time series models, and further optimizing portfolios for three different objective functions using the simulated data. Additionally, optimization is performed using historical data. For historical data, step 1-3 is omitted, and X_{it} = *historical wind power data (year 1996-2016)*. The optimal weights obtained from the optimizations are used on the test data (Wind power production, year 2017-2019) to compare the performance of the optimizations. This forms the basis for the portfolio results in Table 5.1 and the weights in Table 5.2 in the Analysis chapter.

Portfolios as a function of $w_1, \ldots, w_k$ is computed as follows:

1. Draw $(U_1, \ldots, U_k)$ from d-dimensional copula C

2. Compute sampled residuals $(\epsilon_1^s, \ldots, \epsilon_k^s)$ through $F_i^{-1}(U_i)$

   F_i^{-1} *is the inverse of the empirical cumulative density function used to back-transform the marginals from [0,1].*

3. Simulated power production X_{it} = *TS-estimate with ϵ_i^s from copula*

   (a) For DHR models: Following equation 4.3 and 4.4 substituting ϵ_t with ϵ_i^s

   (b) For ARMA models: Following equation 4.5 substituting ϵ_t with ϵ_i^s

   Any values breaching the limits of the turbine is adjusted to the minimum or maximum value.

   – **Repeat step 1-3 N times to obtain** $(X_{11}, \ldots, X_{k1}), \ldots, (X_{N1}, \ldots, X_{kN})$ –

4. Make Portfolio $Y = (Y_1, \ldots, Y_N)$ via $Y_t = \sum_{i=1}^{k} X_{it} w_i$

5. Optimize objective function using Y:

   (a) Compute empirical 5% VaR

   (b) Compute Markowitz minimum variance

   (c) Compute empirical 5% VaR with penalized avg return

   – **Repeat step 4-5 until convergence for optimal weights** (eq. 4.6 to 4.15) –

6. Apply optimal weights to test data and compare performance (table 5.1)

---

5   Analysis and Results

In this section, the implementation of the methodology on the wind power data is presented. The section is structured sequentially in the order the steps are performed: time series modeling, modeling of dependency between locations using time series residuals, simulation of data through the recomposed time series using sampled residuals, and lastly, a comparison of optimization methods and results.

5.1   Estimating Time Series with Dynamic Harmonic Regression Models

We estimate time series models for each location separately. The time series models use historical data to predict future values for the variables, and each series generally consist of three main components: trend, seasonality, and remainder (Hyndman and Athanasopoulos, 2021). In this section, we fit models to capture potential trend and seasonality components in the data.

We analyze time series with different temporal resolutions: mainly hourly and daily. The wind power data exhibit a noticeable difference in characteristics over the time horizons, mainly in terms of the noisiness of the data. Following the law of large numbers, the noise in the data is less influential in the aggregated data. For the hourly and daily data, we experience truncation to a larger extent than for weekly and monthly data caused by the maximum production capacity of the wind turbine. The occurrence of hourly and daily production at maximum capacity is not uncommon for any of the locations of interest. However, to claim that the production level is stable around the maximum capacity would be an exaggeration, as the volatility in production is high at the hourly and daily levels.

Since the time series are independently estimated for each location, the outcome of the model specifications and the necessity for pre-processing of the data differs for each series. From the initial descriptive analysis, we expect the series to have at least a yearly seasonal pattern with peaks for the winter months and low points for the summer months. Each time series is decomposed to get an impression of the trend and seasonality. Further, we fit Dynamic Harmonic Regression models, ARMA models, and a season NAIVE model based

on STL decomposition to each time series and compare the accuracy of the models. We use the season NAIVE with a simple yearly season as the benchmark model of comparison. The models seem capable of capturing the most prevailing information in the series.

We start by using hourly data because of the relevance of high temporal resolution to the mechanisms of the electricity market. The season NAIVE models give relatively poor in-sample fits, but the decomposition serves as a helpful tool to reveal that the data exhibit seasonal patterns. The data has no noteworthy trend component, and the most prevailing season is for the yearly period. The season NAIVE models suggest shorter period seasonality for daily, weekly, and monthly periods, but the subsequently presented models outperform the models' fit.

The accuracy varies among the model specifications for the Dynamic Harmonic Regression variations and ARMA models fitted to the data. For most cases, the seasonal element expressed as Fourier terms gives better in-sample fits than the seasonal ARMA models estimated through minimizing AIC. The exception is for SKS in daily data, where a non-seasonal ARMA model provides the best fit. Shorter period seasonality than yearly does not seem to provide a better fit for the Dynamic Harmonic Regression models. A list of the model components used for each series can be found in the appendix A1.
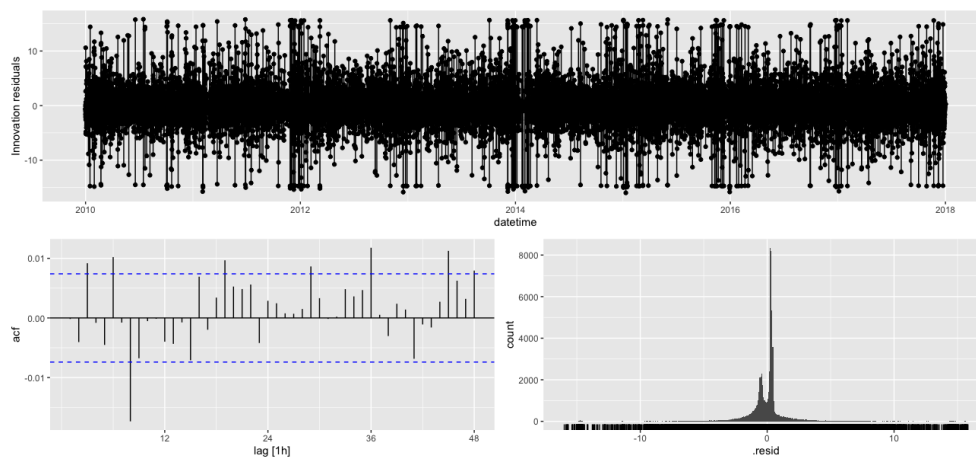
Figure 5.1: Residual plots for Dynamic Harmonic Regression model on hourly data for Utsira Nord (UN).

After fitting a model with Fourier terms and ARMA processes on the regression residuals η, we get time series residuals ϵ as shown in Figure 5.1. These results are consistent for all locations. The residuals show non-linearity and are not normally distributed. Further,

there is significant auto-correlation present in the residuals. Judged by the auto-correlation, the residuals are not resembling white noise in contrast to a well-fitted time series model. This is confirmed by the p-values from the Ljung-Box test, rejecting the null hypothesis of independently distributed residuals. The time series model is struggling to capture all information in the data, meaning that it is likely that there is remaining information in the residuals.
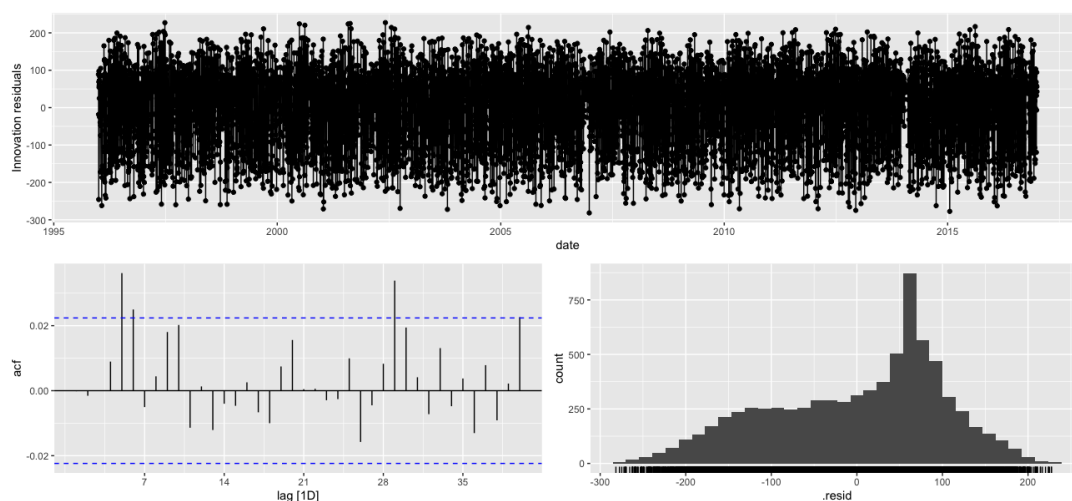
Figure 5.2: Residual plots for Dynamic Harmonic Regression model on daily data for South of Lindesnes (SLI).
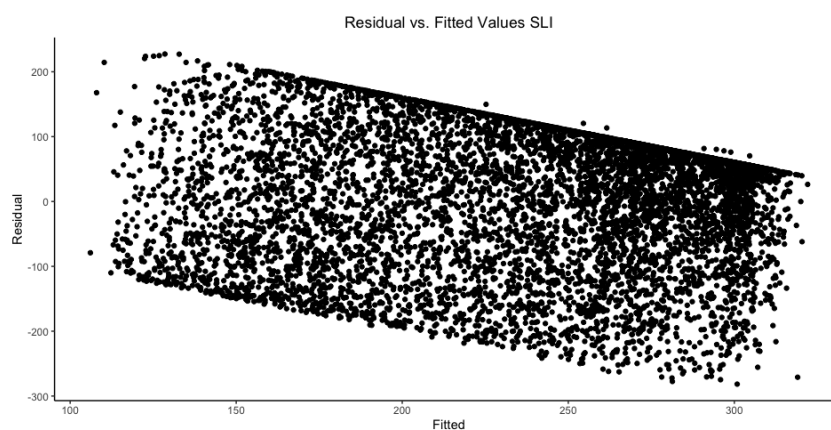
Figure 5.3: Residual vs. fitted values for Dynamic Harmonic Regression model on daily data for South of Lindesnes (SLI).

As with hourly data, it is clear from Figure 5.2 that the residuals of the time series model for daily data are not normally distributed and exhibit some autocorrelation between lags. This is an example of the models on a daily level and is similar for all locations with some nuances. The residual plots tell us that there may still be information in the residuals

that we have not managed to include in the model. The p-value from the Ljung-Box test confirms this suspicion for some of the locations, and the models may be be improved upon with more sophisticated modeling of the variance. Further, Figure 5.3 shows non-linear effects in the residuals. This is partly a result of the truncated power production data. The maximum and minimum values from the turbines create limitations for how large the residual can be on specific values, resulting in the clear-cut edges in the plot. Through testing for significant auto-correlation in residuals, we experience larger problems with dependency between lags for hourly models than for daily models. Although not perfect, the models for daily data do a better job of capturing the relevant information in the data.

Now that the time series models largely capture the time dependency, we aim to model the dependency between areas based on the dependency in the residuals of different locations at the same time t. As seen in Figure 5.2, the marginal distributions of the residuals are non-normal. Further, it can be observed in Figure 5.4 that there exists a non-linear dependency between locations, based on the clustering in the scatter plots. A copula model can help in modeling these features. Copula functions can model dependency between variables with different marginal distributions, including non-normal distributions. At the same time, copula functions can model the marginal distributions and the dependency structure separately, making it possible to model any type of dependency structure between variables (Bessis, 2010). The copula will be further addressed in the next sub-chapter 5.2.
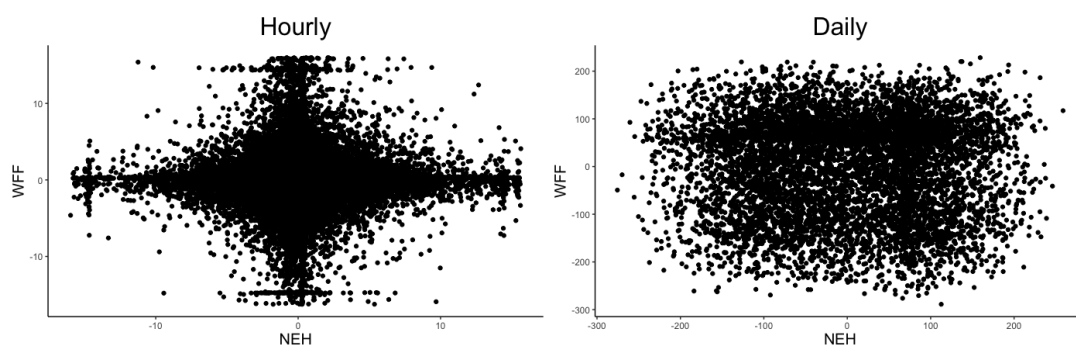
Figure 5.4: Dependency between residuals for West of Flekkefjord (WFF) and North-East of Honningsvåg (NEH).

For the copula method to be justified, the dependency of the residuals should be captured fully in lag 0 of the residuals, and the dependency should be constant over time (Grothe and Schnieders, 2011). This is important because if the dependency is captured in other

lags than 0, the copula model cannot sufficiently capture the dependency between the residuals. Similarly, the dependency between the residuals of the different locations must be somewhat stable over time so that we know the optimal placement of wind parks does not change over time. To ensure that all of the dependency is present in the concurrent residuals, pairwise cross-correlation is performed for all 13 locations. The desired result is to have no cross-correlation for lags not equal to 0. As observed in Figure 5.5, this is done on hourly and daily data. The shown example is of Utsira Nord and West of Fitjar, which we know from before are highly correlated locations. The lagged cross-correlation of the hourly data shows that multiple significant dependencies are located in other lags than lag zero. This result is consistent for all pairs of locations on hourly data. For the daily data, we observe that all dependency is located in lag 0.
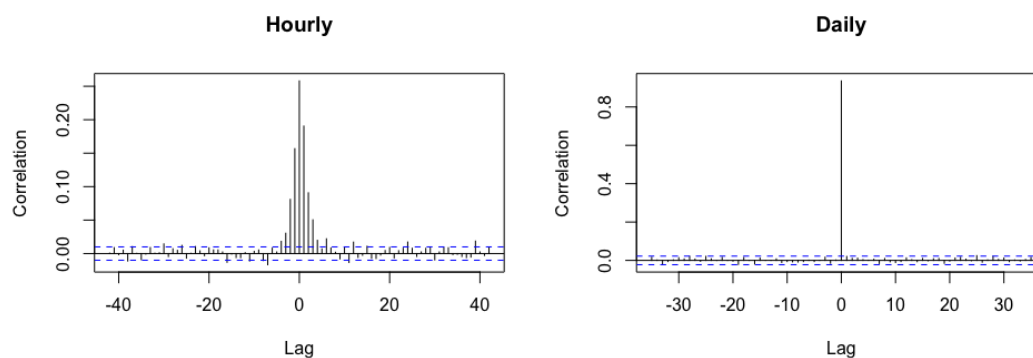
Figure 5.5: Cross-correlation of residuals for Utsira Nord (UN) and West of Fitjar (WFI) with hourly and daily data.

Daily data seems promising in regards to dependency being located in lag 0. Further analysis shows that there are some differences between pairs. Figure 5.6 shows two examples of pairwise cross-correlation. A high dependency is found in lag 0 for locations with high correlation. On the other hand, for locations with low correlation, low dependency is found in lag 0. Across most pairwise cross-correlations, other lags show significant dependency. This might be a result of the time series models' incapability to remove all auto-correlation in the residuals. However, no trend is found in these dependent lags, giving reason to believe that the information captured by other lags is negligible.
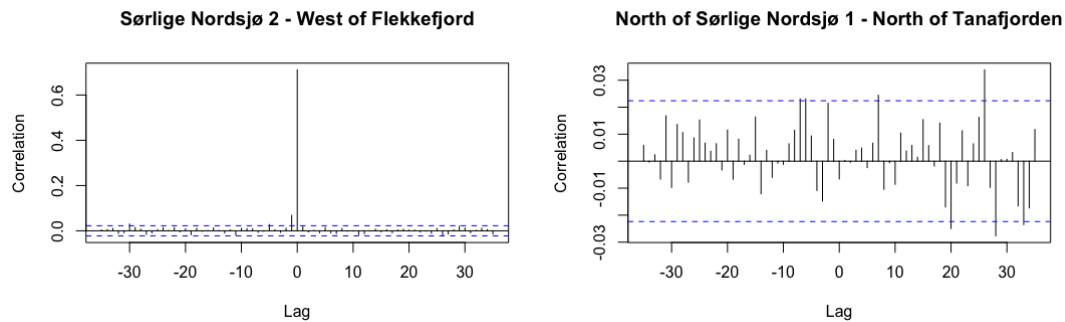
Figure 5.6: Cross-correlation between lags for pairs Sørlige Nordsjø 2 (SN2) - West of Flekkefjord (WFF) **&** North of Sørlige Nordsjø 1 (NSN1) - North of Tanafjorden (NTF).

Finally, the dependency between the locations should be stable over time. In Figure 5.7 below, it can be observed that the correlation of the residuals of the extracted pairs of locations is relatively stable over time. We can therefore use the data to decide wind park locations since the optimal location will not change over time.
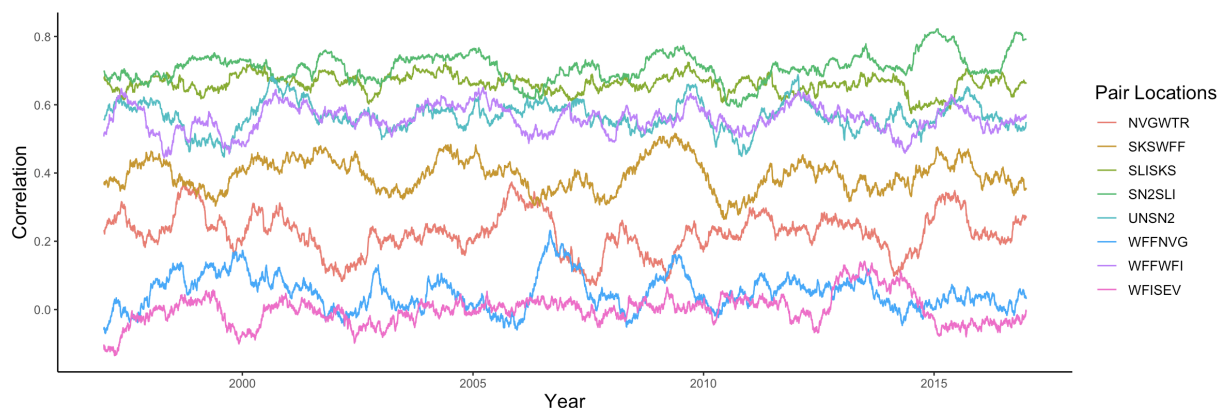
Figure 5.7: Correlation between locations over time for an extract of location pairs.

Based on the analysis done of the time series models and the resulting residuals, we have decided to only move forward with daily data. The time series models fitted on hourly data do not seem to adequately capture the behavior of the series, which reduce the reliance on the estimates. The dependency will not be able to sufficiently be modeled with copulas on the hourly data due to correlation in the residuals across lags. In the following section, copula models will be introduced based on the marginal distributions and the dependency structure in the daily residuals.

5.2   Finding Copula Structure and Simulating Data

So far, the models explain wind power at a location based on previous observations and error terms through the time series modeling. Further, we attempt to model the dependency between areas by modeling the dependence structure of the ARMA residuals (the error term ϵ_t of the time series models) for different locations through a d-dimensional distribution. In other words, the dependence between locations is modeled for effects not captured by the components of the time series estimates. By simulating errors while preserving the dependence between the locations, we can reconstruct the components of the data to make arbitrary long simulated sets of wind power data to use in the optimization.

A D-vine copula approach is performed here due to the distributions of the data: A pairwise assessment of the areas shows that the dependency structure differs from pair to pair. To model the dependencies between locations, we fit bi-variate and conditional bi-variate copulas to the residuals by using D-vine trees: The D-vine is a tree-like structure, with each node denoting a pair of variables, and each edge representing a bi-variate copula that models their dependence. The D-vine is built by repeatedly conditioning on pairs of variables until all pairings are linked. To find the tree structure for the 13 variables, we order the nodes using the shortest Hamiltonian path in terms of weights based on the absolute value of the empirical Kendall's Tau, as suggested in the R-documentation for the *VineCopula* package (Nagler et al., 2023). The first four levels of the resulting structure are presented in figure 5.8.

D-vine Tree Structure

T1:

(13)—(12)—(11)—(10)—(8)—(9)—(6)—(1)—(7)—(2)—(5)—(3)—(4)

13, 12   12, 11   11, 10   10, 8   8, 9   9, 6   6, 1   1, 7   7, 2   2, 5   5, 3   3, 4

T2:

(13,12)—(12,11)—(11,10)—(10,8)—(8,9)—(9,6)—(6,1)—(1,7)—(7,2)—(2,5)—(5,3)—(3,4)

13,11|12   12,10|11   11,8|10   9,10|8   8,6|9   9,1|6   6,7|1   1,2|7   7,5|2   2,3|5   5,4|3

T3:

(13,11|12)—(12,10|11)—(11,8|10)—(9,10|8)—(8,6|9)—(9,1|6)—(6,7|1)—(1,2|7)—(7,5|2)—(2,3|5)—(5,4|3)

13,10|12,11   12,8|11,10   11,9|10,8   10,6|8,9   8,1|9,6   9,7|6,1   6,2|1,7   1,5|7,2   7,3|2,5   2,4|5,3

T4:

(13,10|12,11)—(12,8|11,10)—(11,9|10,8)—(10,6|8,9)—(8,1|9,6)—(9,7|6,1)—(6,2|1,7)—(1,5|7,2)—(7,3|2,5)—(2,4|5,3)

13,8|12,11,10   12,9|11,10,8   11,6|10,8,9   10,1|8,9,6   8,7|9,6,1   9,2|6,1,7   6,5|1,7,2   1,3|7,2,5   7,4|2,5,3

Figure 5.8: First four levels in the D-vine tree for the model fitted to daily data.

We use copulas from the Gaussian and Archimedean families using maximum likelihood estimation and AIC as selection criteria for the model fit. The advantage of combining Gaussian and Archimedean families is that they allow for many different dependence structures to be fitted, where the dependency in the tails is modeled differently (Embrechts et al., 2001). We fit a multivariate Gaussian copula for reference to evaluate the model's fit. To estimate parameters for the model, the residuals need to be transformed to the unit interval, uniformly distributed [0,1]. The transformation is performed using the empirical cumulative distribution function (ECDF). For the multivariate copula, the estimation method requires a scaling factor of the ECDF of $n/(n+1)$ to manage the parameter estimation. The difference in transformation is negligible for the parameter estimation for the D-vine copula.

The resulting model has an AIC of -72354.57 and consists of a majority of non-Gaussian families: Gaussian 9, Student t 18, Clayton 3, Gumbel 1, Frank 25, Joe 3, rotated Clayton (180°) 4, rotated Gumbel (180°) 2, rotated (180°) Joe 2, rotated Clayton (90°) 4, rotated Clayton (270°) 5, rotated Gumbel (270°) 2. For a full summary of the model's fit, consult the appendix A2.1. The overweight of Archimedean copulas makes it reasonable to assume a non-linear dependence structure between the variables. The benchmark model confirms this: The multivariate Gaussian copula fit obtains an AIC of -66567.

We use the selected D-vine model to simulate 365 000 residual values (approximately 1000 years worth of values) for each of the variables. Since the simulated values are on the unit interval, we back-transform them using the inverse of the empirical cumulative distribution function. The next step is reconstructing the time series using the time series' model fit and the simulated residuals. The reconstruction is performed to obtain 1000 years of sequential daily wind power data for the locations to provide sufficient input for the optimization problem. The new series are assembled manually, modeling the data following equation 4.3 through 4.5 using the first few data points from the original data sets as starting points for the calculation. The size of the simulated series leads to some extreme values beyond the known limits of the turbine. Therefore, values below the minimum production and above the maximum daily capacity are rounded to the closest limit value of 0 and 360 MW. An arbitrarily selected three-year period from the simulated series is presented along with the original series in figure 5.9.
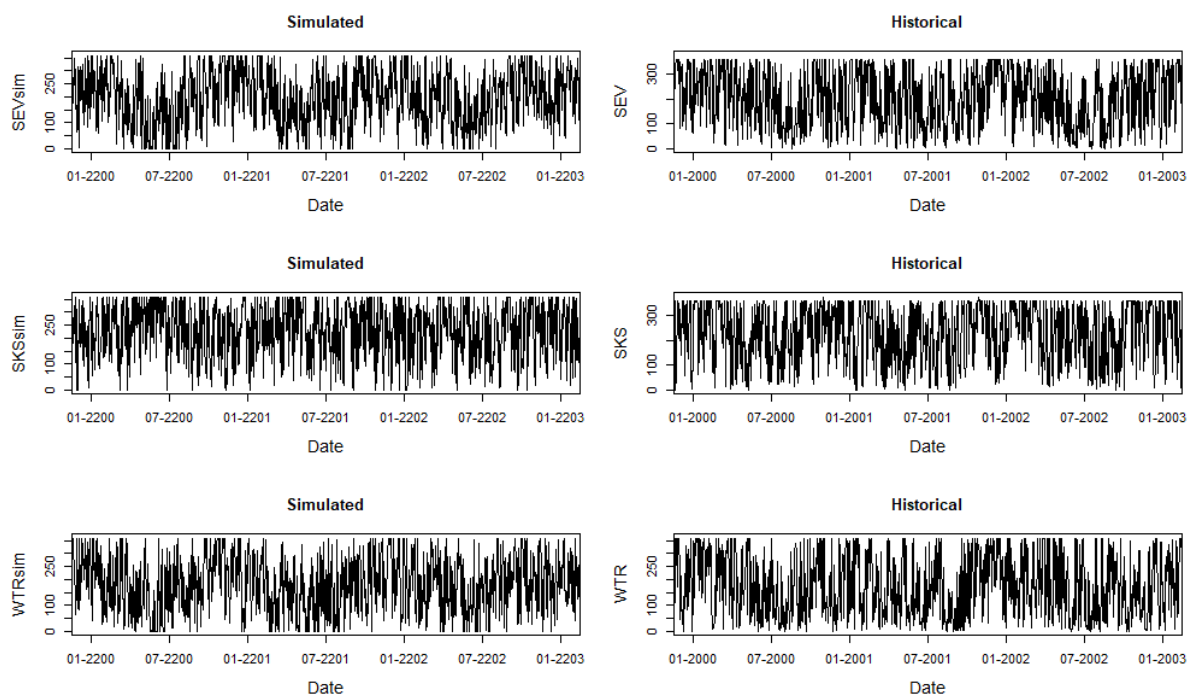
**Figure 5.9:** Simulated data vs. historical data of daily power production.

5.3   Optimization of Wind Park Portfolios and Results

After applying the time series and copula model to construct 1000 years of wind power data, the data will be used to find wind park locations that stabilize the output of wind power production on a daily basis. The thesis aims to investigate whether the optimization

of VaR on the simulated data can make a better combination of wind park sites than an approach using historical data directly. Further, different optimizing approaches are evaluated: Markowitz's minimum variance, optimization of VaR, and VaR with penalized average production. The optimization methods, VaR and Markowitz, are trained on both the simulated 1000 years of wind power data and the historical wind power data from 1996 to 2016, with a daily temporal resolution. The models aim to make well-performing portfolios for future wind production values. Therefore, the optimal weights from each method is tested on unseen test data, which none of the models are trained on. The test data is three years of historical daily wind power production from the year 2017 to 2019. Additionally, an analysis of the best combinations of areas is performed. The method with a penalty factor for portfolios with an average production below a threshold is included in this part to observe the difference in area combinations when high average production is considered important in addition to stable production.

Our analysis is based on the understanding that the best measure for stable wind power production is maximizing the lower quantiles of the portfolio distribution. This is founded on the basis that VaR is not dependent on normally distributed data and that by maximizing the lower quantiles, we minimize zero and low production periods. Zero and low production is an unfavorable characteristic of wind power that makes it difficult to provide stable production. Minimizing zero and low production periods will therefore lead to more stable wind production. The variability in wind power is hard to remove completely, but by maximizing the 5% quantile, we limit a higher portion of the variation to higher production values.

Since there is no baseline portfolio available, the analysis is built around a comparison of the methods. The performance measures of all methods can be observed in Table 5.1. The table includes performance measures (VaR, average production, and standard deviation) when the weights of each portfolio are applied to the unseen test data (daily power production year 2017-2019). The method for achieving the results in the table is summarized in the methodology sub-chapter 4.4. The values are in the form of a percentage value of installed capacity (maximum production). For instance, the 5% VaR of the *maxVaR* portfolio trained on simulated data thereby means that 95% of the time, more than 29.1% of the installed capacity is being produced.

Portfolio	VaR 0.05	Average	VaR 0.95	St.dev.
Simulated Data				
maxVaR	29.1 %	58.8 %	86.5 %	18.1 %
Markowitz	27.6 %	57.1 %	86.2 %	17.8 %
maxVaR+Penalty	26.7 %	62.9 %	92.3 %	20.8 %
Historical data				
maxVaR	27.0 %	59.7 %	88.7 %	18.6 %
Markowitz	27.8 %	57.3 %	85.7 %	17.7 %

Table 5.1: Performance measures for all portfolios. Weights from each portfolio tested on unseen test data (daily wind power production, the year 2017-2019). Values are in the form of a percentage of installed capacity.

5.3.1   Simulated and Historical Data

The results reveal that the non-linear unconstrained optimization of 5% VaR (*maxVaR*) on simulated data is performing best in regards to 5% VaR on the test data. Compared with the optimal weights retrieved from the same method using historical data, simulated data leads to an increase in 5% VaR of 7.8% (from 27.0% to 29.1%). This difference in VaR means that for 95% of the time the minimum proportion of installed capacity that produces wind power is 2.1% higher when using weights derived from the optimization of simulated data compared to historical data. Therefore, the modeling of dependence structures through the time series and copula shows clear improvements in the optimal wind park allocation when using VaR as the objective function. Judging by the standard deviation and return for the Markowitz portfolio, the impact of simulating data is not improving the output of the model.

5.3.2   VaR and Variance as Objective Function

When comparing the *maxVaR* portfolio with the *Markowitz* portfolio using simulated data, the 5% VaR is as expected worse as a consequence of the objective function. Similarly, the average production is lower using the Markowitz minimum variance method. This method, combined with wind power data with non-normal distributions, consistently creates portfolios with lower average production. An interesting observation, however, is that the Markowitz minimum variance method on historical data results in a higher 5% VaR than the approach of maximizing the lower quantiles on historical data. It seems reasonable to believe that because these two portfolios are made using only 21 years of historical data from the year 1996 to 2016, there are not enough observations to

make the weights stable for future values. Overall, the maximization of 5% VaR shows clear improvements compared to the Markowitz minimum variance portfolio when the dependency structure is modeled through time series and copula.

The output Y_t (see eq. 4.10), as a percentage of max capacity, of *maxVaR* and *Markowitz* made from the simulated data can be observed in Figure 5.10. The weights of the portfolios are applied to both the simulated data and the test data. The distribution of the *maxVaR* and *Markowitz* portfolio is visually very similar, except the *maxVar* portfolio being slightly more left-skewed. It has a small increase in variance, but it has less zero and low values and a higher average production which is argued to be more important. This supports the use of VaR for optimization compared to variance for optimal combinations of wind production sites.
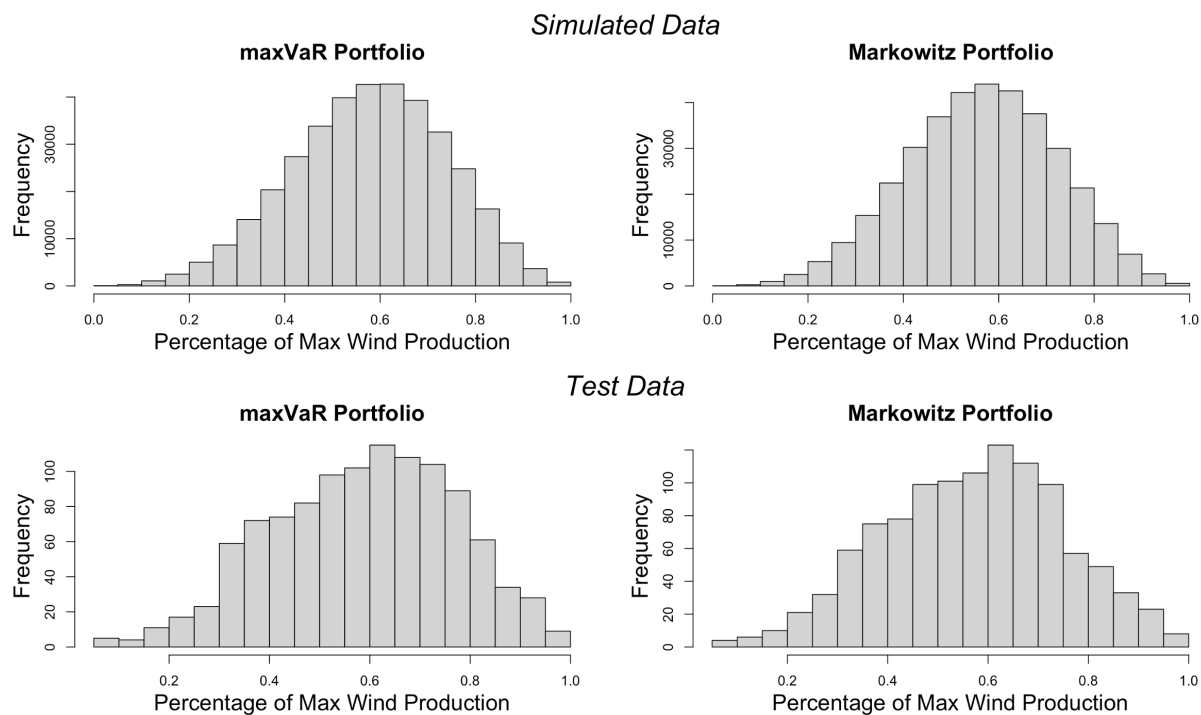
Figure 5.10: Histogram of the *maxVaR* and *Markowitz* portfolio made from simulated data.

5.3.3   The Areas Included in the Portfolios

The optimal weights used to achieve the portfolios with the most stable wind power production can be observed in Table 5.2. The weights can be interpreted as the proportion of available wind turbines that should be sited at each location. For the method ensuring the most stable wind power production, *maxVaR* with simulated data, it is clear that

diversification benefits the result since all areas are included in the portfolio. SKS and NTF have the largest shares and, consequently, are the most influential in terms of obtaining the best result. Further, SN2, WTR, and NEH are included with a high percentage share of the portfolio. These are representations of the whole coast of Norway, expressing the importance of distance between wind parks to achieve the most stable production. It is clear by comparing these weights to the weights of the other portfolios (excluding $maxVaR+Penalty$) that the regions SN2, SKS, WTR, and NTF have a high involvement rate. Therefore, it may be claimed that the combination of these areas yields the most promising outcomes in a portfolio for stable wind generation.

| | Simulated Data | | | Historical Data | |
Location	maxVaR	Markowitz	maxVaR+Penalty	maxVaR	Markowitz
Utsira Nord (UN)	4.05 %	0.00 %	2.08 %	5.59 %	0.00 %
Sørlig Nordsjø 2 (SN2)	10.90 %	6.65 %	8.83 %	9.52 %	7.57 %
South of Lindesnes (SLI)	4.98 %	0.00 %	24.04 %	11.81 %	1.41 %
South of Kristiansand (SKS)	19.69 %	19.17 %	13.97 %	10.70 %	15.95 %
West of Flekkefjord (WFF)	2.11 %	4.36 %	7.82 %	9.11 %	7.94 %
West of Fitjar (WFI)	0.38 %	9.07 %	0.54 %	2.95 %	8.96 %
North SN1 (NSN1)	6.41 %	4.46 %	13.00 %	4.19 %	0.71 %
North of Vega (NVG)	3.67 %	9.60 %	0.12 %	0.92 %	3.59 %
North of Rørvik (NRV)	6.54 %	4.00 %	0.10 %	1.10 %	8.34 %
West of Tromsø (WTR)	11.69 %	16.36 %	0.04 %	11.28 %	17.46 %
North-East of Honningsvåg (NEH)	10.12 %	4.52 %	20.10 %	2.72 %	2.20 %
South-East of Vardø (SEV)	5.04 %	8.49 %	0.13 %	13.35 %	9.54 %
North of Tanafjorden (NTF)	14.41 %	13.32 %	9.22 %	16.76 %	16.33 %

Table 5.2: Optimal Wind Park Portfolios.
The portfolios are made on both simulated data through time series and copula and historical data (1996-2016). In addition, the VaR optimization is compared with a Markowitz portfolio and a VaR optimization with a penalty factor for mean production under 62.5%.

For the optimization including a threshold of minimum desired average production, the threshold P (see eq. 4.15) is set to 62.5% of installed capacity ($\frac{225MW}{360MW}$). The threshold is set to 62.5% to get a substantial increase from the portfolios maximizing 5% VaR while still leaving enough room to allow for diversification. The maximum average production possible is 66.4% ($\frac{239.13MW}{360MW}$) when all the production is allocated to SLI (see table 3.2). The lower quantile value will, by nature, be substantially larger than the penalty factor. The scaling factor s is therefore set to 2.5 to make the impact of the penalty large enough to make the optimization prefer portfolios with higher average production. The optimal solution results in a 5% VaR of 26.7% and an average production of 62.9%. Comparing the $maxVaR$ and $maxVaR+Penalty$ portfolios, the latter is more left-skewed with a higher concentration of high production volumes, shown in Figure 5.11. To achieve a more dense

concentration of high production volumes, the portfolio has a fatter lower tail, meaning more zero and low production volumes.
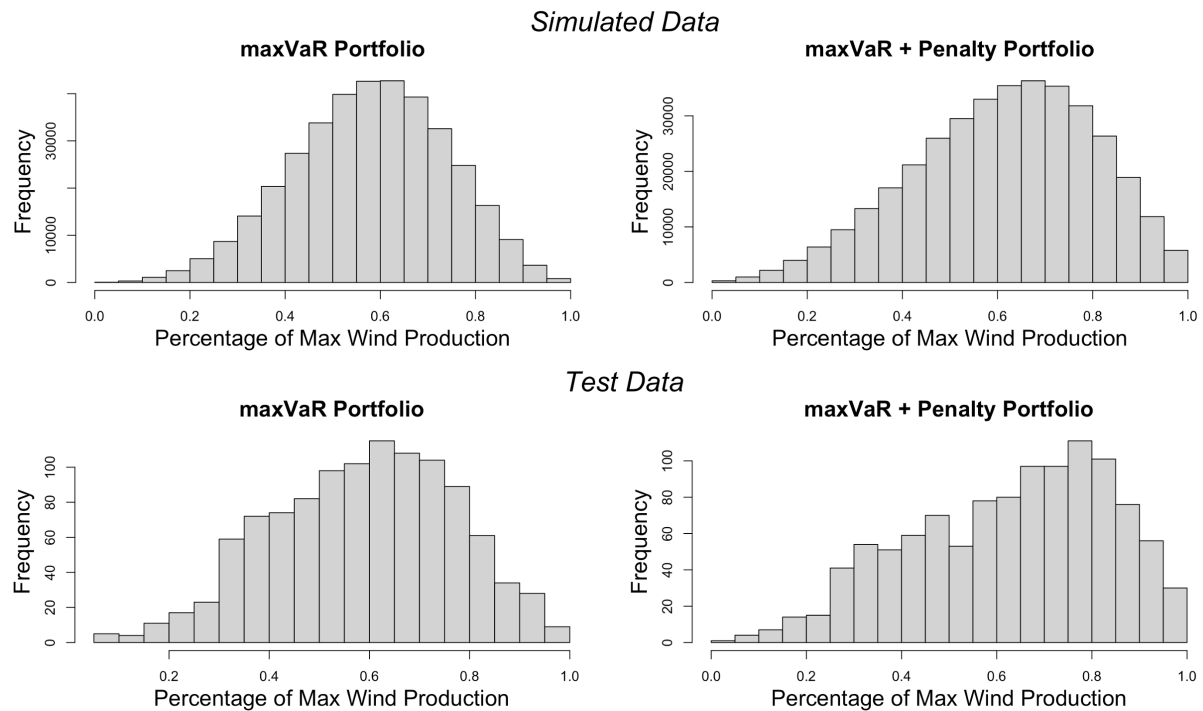
Figure 5.11: Histogram of the *maxVaR* and *maxVaR+Penalty* portfolio made from simulated data.

Since the *maxVaR+Penalty* model rewards higher production, a different combination of areas is optimal. The diversification is not as evident, leading to more areas having close to zero percent involvement. Consequently, the contribution of a few areas with high average production becomes more impactful in this portfolio. The seven locations with the highest average production are included: UN, SN2, SLI, SKS, WFF, and NSN1. To balance the output from the high average production sites, we observe that the portfolio assigns relatively high weights to the northern areas NEH and NTF. As established in the descriptive analysis, these are areas with low correlation to the southern high average production areas. The degree of importance given to a high average portfolio return has a large impact on the portfolio composition, compared with only optimizing the lower quantiles of the portfolio distribution. Thus, when deciding on locations to develop for wind power production, the decision of objective function is fundamental for finding the optimal allocation of resources.

From the analysis, it is clear that the unconstrained non-linear optimization of 5% VaR using simulated data results in the portfolio with the most stable production. The map in Figure 5.12 is a visual presentation of the capacity allocation for the optimal solution using the objective function from the *maxVaR* approach. If the objective was to cover the Norwegian government's goal of 30 GW installed capacity, for example, 2000 wind turbines with a capacity of 15 MW are needed (Regjeringen, 2022b). This would mean allocating 288 wind turbines to NTF assuming that the spatial requirement is fulfilled.
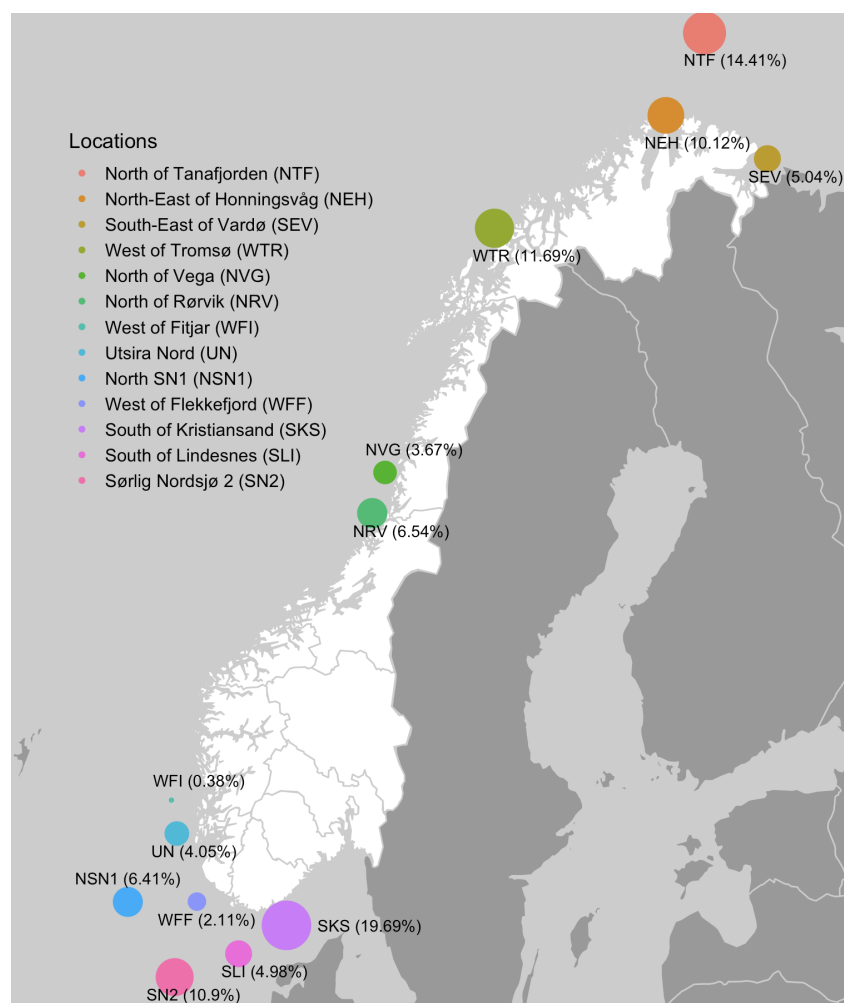
Figure 5.12: Map over the 13 locations with size of points corresponding with weights in the optimal portfolio for *maxVaR*.

The weights achieved from maximizing the model on daily data are also evaluated on hourly data. The results show a worse 5% VaR for our approach compared to Markowitz, but similarly to daily data, a higher average return is achieved compared to Markowitz. The weights obtained from *maxVaR* on 1000 years of simulated data are therefore not transferable to hourly data.

6 Discussion

The following section discusses the results and what insights they may provide. Additionally, the limitations of the study and its implications are discussed.

According to the analysis, our findings demonstrate that the method of simulating 1000 years of wind power production for each area based on the dependency structure through time series and copulas provides a stronger basis for determining the ideal weights for the combination of wind park locations. Further, the unconstrained non-linear optimization of VaR seems to be a better measurement than the variance for the purpose of reducing intermittency. For example, when comparing the *maxVaR* and *Markowitz* methods on simulated data, there is an increase of 5.4% in the 5% VaR and a 3.0% increase in average daily wind power production. This means that with the *maxVaR* method, the portfolio reduces zero and low production values, resulting in higher average production. Over several years, the benefit of choosing the *maxVaR* method could be valuable.

Through the optimization of portfolios, it is apparent that the application of VaR as performance measurement leads to many portfolios with very similar objective values. Changes in the initial values for the optimization result in different combinations of areas that maximizes the 5% VaR. The objective function arrives at approximately the same value, but the weights change. Despite arriving at different final weights for different initial weights, it is still possible to draw insightful conclusions from the results in terms of the placement of wind parks. Some areas are consistently present with a high percentage in the portfolios: SKS, WTR, NEH, and NTF. Many of these areas have a low correlation, making the combined production more stable. However, this attribute of the model makes it difficult to recommend specific weights for each site, as many combinations can lead to approximately the same result. Further, changing the objective function to maximize 1% or 10% VaR instead of 5% VaR will also slightly change the optimal portfolio's weights. However, the trend with SKS, WTR, NEH, and NTF being substantial participators in the portfolio remains.

The weights in the *maxVaR* optimal portfolio are distributed across all locations, emphasizing the positive effect of combining wind power production from dispersed wind parks. For example, if the optimal portfolio is compared to a portfolio including

only UN and SN2 since these are the only areas currently approved for offshore wind parks, the improvement in 5% VaR is major. The 5% VaR increases from 12.8% in the portfolio with UN and SN2 to 29.1% in the optimal portfolio, an increase of 127%. On the other hand, the average production of the $maxVaR$ portfolio only decreases by 7.1% (from 63.3% to 58.8%) from the UN and SN2 portfolios. The benefit achieved from the portfolio in terms of an increase in the 5% VaR can therefore be argued to be larger than the loss in average production.

A dilemma of what to prioritize arises when finding optimal combinations of wind production sites: minimal risk by maximizing 5% VaR or increasing risk to achieve higher average production. If the average output is fully prioritized, all production should be allocated to the South of Lindesnes because this location has the best wind conditions. The portfolio $maxVaR{+}Penalty$ is weighting this concern of minimizing intermittency through maximization of 5% VaR and keeping average production high. Diversification is in this portfolio partly de-prioritized for achieving higher average production, leaving more areas almost entirely out of the portfolio. Sites with good wind conditions (high average production) are in this portfolio represented with a higher percentage. The change in objective function leads to a decrease in 5% VaR of 8.2% (from 29.1% to 26.7%) and an increase in average production of 7.0% (from 58.8% to 62.9%). The maximum average output for one area, and consequently the highest average production that could be achieved for all portfolios, is 66.4% ($\frac{239.13MW}{360MW}$), as shown in Table 3.2 under the data section. This indicates that it is possible with this method to achieve a portfolio producing close to maximum average output while reducing the zero and low production values to make production more stable.

Based on our results, it is apparent that by diversification through the interconnection of wind parks along the coast of Norway, intermittency in wind power production can be reduced. However, there are some limitations to our method. Firstly, as discussed in our analysis of the Dynamic Harmonic Regression models, our time series model is not able to capture all the information that is available in the data. The time series model is an imperfect representation of the data, meaning there might be patterns in the data we cannot utilize fully. Additionally, the optimization method used does not guarantee that a global optimum is found. Depending on initial weights, different local optimums can be

found. The weights can therefore vary depending on initial weights, and it constitutes a risk of not reaching the optimal solution. Furthermore, because the research was done solely using data for wind power estimated for a single turbine, wake effects, maintenance, and other aspects were not considered. These are factors that can have a considerable impact on where wind parks are located. Consequently, the results in this thesis will only be able to offer one of many aspects regarding the optimal placement of wind parks. Finally, as of today, Norway is split into five pricing areas for power trading and does not have the infrastructure to take advantage of the interconnection of wind parks across the whole coast of Norway. As a result, the outcome is not directly applicable to the Norwegian case in its current state. However, the results of the method still show the positive effect of interconnected sites on the joint stability in power production. The method can be applied to other data, whether that is the currently suggested areas from NVE or different compositions of locations. An interesting approach would be to apply the method within pricing areas or between pricing areas where the interconnection of grid points may be feasible. With such an approach, the impact of the portfolio could be more closely linked to the price of electricity in the areas by altering the balance between supply and demand.

7 Conclusion

Due to global warming, emission goals, and increased energy consumption, renewable energy production has been on the agenda in Norway for several years. The subject of offshore wind power has surged as a candidate but has, among other things, been criticized for its intermittency. With Norway's ambitious production goals, reducing the intermittency in production should be of high interest to both the government and the power producers. The thesis has analyzed the interconnectedness of offshore wind parks along Norway's coast, focusing on stable conditions for wind power production as the main criteria. The analysis has taken on the following research question:

When determining weights in the optimization for a portfolio with the most stable production for offshore wind park areas in Norway, how well does a compound dependency model for the simulation of wind power data perform compared to historical data, and which areas are included?

To create simulated wind power data, different methods to represent the time dependency in the historical data were compared, and the dependency between locations was modeled through a D-vine copula on the time series' residuals. The time series were modeled on daily time resolution to sufficiently capture significant cross-correlation between the areas for the same time t. The sampling of residuals from the copula model constituted the new error component in the simulated series. Different optimization methods are then evaluated to find suitable combinations of locations with stable production as the decision criteria.

We find that maximizing Value at Risk (VaR) has favorable characteristics for what could be considered a stable production portfolio. Maximization of 5% VaR as opposed to minimizing variance leads to an improvement of 5.4% in the 5% VaR and a 3.0% increase in average production. The portfolio distributions of the optimization of 5% VaR and the Markowitz minimum variance are similar, except the portfolio distribution of the maximization of VaR is more left-skewed. Thus, the maximization of VaR has a slightly larger variance, but it has less zero and low values and a higher average production which is argued to be more important.

The study's findings show that portfolio optimization performed on simulated data performs better than on historical data. The 5% VaR increases by 7.8% when using simulated data instead of historical data. Higher 5% VaR reduces the occurrences of zero and low production values in the portfolio, making the production more stable. When computing the weights for the portfolio with the most stable production for offshore wind park areas in Norway, it is clear that using VaR as objective function and a compound dependency model for simulation of wind power data is preferable. Using this method, all areas are included in the portfolio, emphasizing the importance of diversification. However, the most influential areas which should be prioritized are Sørlige Nordsjø 2, South of Kristiansand, West of Tromsø, and North of Tanafjorden. When the objective function is altered to include a penalty factor for average production under 62.5%, diversification is partly de-prioritized to include areas with high average production. Thus, a higher portion is allocated to the southern areas. When prioritizing high average production, some areas have close to zero percent involvement, making the contribution of a few areas more influential.

For further research, making a better-performing time series model for daily and hourly data could be valuable. The design of hourly trading on the power market makes finding a suitable method for hourly data especially insightful. Investigating how this method can be combined with effects related to wind turbines and scaling of wind parks, like wake effect and maintenance, would further enrich the results. Lastly, using the method within the borders of pricing areas, future interconnected grids, or combining wind power with other energy sources could give results for a portfolio closer to a real-life scenario.

References

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and economics*, 44(2):182–198.

AICorespot (2021). An intro the bfgs optimisation algorithm. Last accessed from AICorespot May 12 2023: https://aicorespot.io/an-intro-the-bfgs-optimisation-algorithm/.

Archer, C. and Jacobson, M. (2007). Supplying baseload power and reducing transmission requirements by interconnecting wind farms. *Journal of Applied Meteorology and Climatology - J APPL METEOROL CLIMATOL*, 46:1701–1717.

Bessis, J. (2010). *Risk Management in Banking*, pages 427–449. John Wiley & Sons.

Birkeland, C., Buvik, M., Arnesen, F., de Brisis, J. M. H., Haukeli, I. E., Kirkerud, J. G., Mindeberg, S. K., Roos, A., Skaansar, E., Skulstad, H. T., Stavseng, A., and Wold, M. (2023). Vindkraft til havs i sørlige nordsjø 2. Last accessed from NVE May 23 2023: https://publikasjoner.nve.no/rapport/2023/rapport2023_04.pdf.

Bradley, S. P., Hax, A. C., and Magnanti, T. L. (1977). *Applied Mathematical Programming*. Addison-Wesley Publishing Company.

Cai, Y. and Bréon, F.-M. (2021). Wind power potential and intermittency issues in the context of climate change. *Energy Conversion and Management*, 240:114276.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I. (1990). Stl: A seasonal-trend decomposition procedure based on loess (with discussion). *Journal of Official Statistics*, 6:3–73.

Embrechts, P., Lindskog, F., and McNeil, A. (2001). *Modelling Dependence with Copulas and Applications to Risk Management*, pages 30–39. Eidgenössische Technische Hochschule Zürich.

Energifakta Norge (2021). Electricity production. Last accessed from Energifakta May 27 2023: https://energifaktanorge.no/en/norsk-energiforsyning/kraftproduksjon/.

Energifakta Norge (2022). The power market. Last accessed from Energifakta May 14 2023: https://energifaktanorge.no/en/norsk-energiforsyning/kraftmarkedet/.

Equinor (2023). The countdown has begun for norway's new era of offshore energy: floating wind. Last accessed from Equinor May 24 2023: https://www.equinor.com/magazine/offshore-wind-adventure.

Ganti, A. (2023). Central limit theorem (clt): Definition and key characteristics. Last accessed from Investopedia March 16 2023: https://www.investopedia.com/terms/c/central_limit_theorem.asp.

Grothe, O. and Schnieders, J. (2011). Spatial dependence in wind and optimal wind power allocation: A copula-based analysis. *Energy Policy*, 39(09):4742–4754.

Haakenstad, H., Breivik, O., Furevik, B. R., Reistad, M., Bohlinger, P., and Aarnes, O. J. (2021). Nora3: A nonhydrostatic high-resolution hindcast of the north sea, the norwegian sea, and the barents sea. *Journal of Applied Meteorology and Climatology*, 60(10):1443–1464.

Hayes, A. (2022). Covariance: Formula, definition, types, and examples. Last accessed from Investopedia February 16 2023: https://www.investopedia.com/terms/c/covariance.asp.

Hyndman, R. and Athanasopoulos, G. (2018). *Forecasting: principles and practice, 2nd edition.* OTexts: Melbourne, Australia. Last accessed from fpp2 May 11 2023: OTexts.com/fpp2.

Hyndman, R. and Athanasopoulos, G. (2021). *Forecasting: principles and practice, 3rd edition.* OTexts: Melbourne, Australia. Last accessed from fpp3 May 2 2023: OTexts.com/fpp3.

Joe, H. (1996). Families of m-variate distributions with given margins and m (m-1)/2 bivariate dependence parameters. *Lecture notes-monograph series*, pages 120–141.

Kahn, E. (1979). The reliability of distributed wind generators. *Electric Power Systems Research*, 2(1):1–14.

Kent State University Libraries (2023). Spss tutorials: Pearson correlation. Last accessed from Kent State University Libraries May 30 2023: https://libguides.library.kent.edu/SPSS/PearsonCorr.

Kenton, W. (2023). Understanding value at risk (var) and how it's computed. Last accessed from Investopedia May 12 2023: https://www.investopedia.com/terms/v/var.asp.

Kim, D., Kim, J.-M., Liao, S.-M., and Jung, Y.-S. (2013). Mixture of d-vine copulas for modeling dependence. *Computational Statistics & Data Analysis*, 64:1–19.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

Nagler, T., Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., and Erhardt, T. (2023). *VineCopula: Statistical Inference of Vine Copulas*. R package version 2.4.5: https://CRAN.R-project.org/package=VineCopula.

Nord Pool (2023). Trading. Last accessed from Nord Pool May 22 2023: https://www.nordpoolgroup.com/en/trading/.

NVE (2023). Identifisering av utredningsområder for havvind. Last accessed from NVE May 19 2023: https://veiledere.nve.no/havvind/identifisering-av-utredningsomrader-for-havvind/.

O'Hara-Wild, M., Hyndman, R., and Wang, E. (2022). *fable: Forecasting Models for Tidy Time Series*. R package version 0.3.2, https://CRAN.R-project.org/package=fable.

Osnes, A. and Nesheim, S. H. (2023). Assessing offshore wind farm placements in norway. Master's thesis, Norwegian School of Economics, Helleveien 30, 5045 Bergen.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Regjeringen (2022a). Grønt industriløft. Last accessed from Regjeringen May 30 2023: https://www.regjeringen.no/contentassets/1c3d3319e6a946f2b57633c0c5fcc25b/veikart-for-gront-industriloft.pdf.

Regjeringen (2022b). Havvind. Last accessed from Regjeringen May 12 2023: https://www.regjeringen.no/no/tema/naringsliv/gront-industriloft/havvind/id2920295/.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. In *Annales de l'ISUP*, volume 8, pages 229–231.

Solbrekke, I. M. and Sorteberg, A. (2022). Nora3-wp: A high-resolution offshore wind power dataset for the baltic, north, norwegian, and barents seas. *Scientific Data*, 9(1):362.

United Nations (2016). Paris agreement. Last accessed from European Union May 11 2023: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:22016A1019(01).

Wuertz, D., Setz, T., Chalabi, Y., and Chen, W. (2020). *fPortfolio: Rmetrics - Portfolio Selection and Optimization.* R package version 3042.83.1.

Würtz, D., Setz, T., Chalabi, Y., Chen, W., and Ellis, A. (2015). *Portfolio Optimization with R/Rmetrics.* Rmetrics Associaton & Finance Online Publishing. Last accessed from Rmetrics May 10 2023: https://www.rmetrics.org/downloads/9783906041018-fPortfolio.pdf.

Appendix

A1   Time Series Model Specifications

- **Time series - hourly**

UN:	Icpt. + Fourier season with 6K + ARIMA(3,0,3) error
SN2:	Icpt. + Fourier season with 3K + ARIMA(4,0,2) error
SLI:	Icpt. + Fourier season with 4K + ARIMA(2,0,3) error
SKS:	Icpt. + Fourier season with 3K + ARIMA(3,0,3) error
WFF:	Icpt. + Fourier season with 3K + ARIMA(2,0,3) error
WFI:	Icpt. + Fourier season with 3K + ARIMA(3,0,2) error
NSN1:	Icpt. + Fourier season with 3K + ARIMA(2,0,2) error
NVG:	Icpt. + Fourier season with 3K + ARIMA(1,0,5) error
NRV:	Icpt. + Fourier season with 3K + ARIMA(1,0,5) error
WTR:	Icpt. + Fourier season with 2K + ARIMA(2,0,4) error
NEH:	Icpt. + Fourier season with 2K + ARIMA(2,0,4) error
SEV:	Icpt. + Fourier season with 5K + ARIMA(3,0,3) error
NTF:	Icpt. + Fourier season with 2K + ARIMA(4,0,2) error

- **Time series - daily:**

UN:	Icpt. + Fourier season with 3K + ARIMA(2,0,2) error
SN2:	Icpt. + Fourier season with 3K + ARIMA(1,0,3) error
SLI:	Icpt. + Fourier season with 3K + ARIMA(1,0,1) error
SKS:	Icpt. + ARIMA(4,0,2)
WFF:	Icpt. + Fourier season with 6K + ARIMA(1,0,1) error
WFI:	Icpt. + Fourier season with 3K + ARIMA(2,0,2) error
NSN1:	Icpt. + Fourier season with 3K + ARIMA(2,0,2) error
NVG:	Icpt. + Fourier season with 1K + ARIMA(3,0,2) error
NRV:	Icpt. + Fourier season with 3K + ARIMA(2,0,2) error
WTR:	Icpt. + Fourier season with 1K + ARIMA(2,0,2) error
NEH:	Icpt. + Fourier season with 2K + ARIMA(3,0,0) error
SEV:	Icpt. + Fourier season with 5K + ARIMA(3,0,0) error
NTF:	Icpt. + Fourier season with 2K + ARIMA(2,0,0) error

A2   Copula Model Specifications

Node Pairs, Families, Parameters, Kendall's tau, Upper and Lower Tail Dependencies for the Fitted Copulas
t: t-copula, F: Frank, N: Gaussian, C: Clayton, G: Gumbel, J: Joe, S: survival 90, other degrees

tree	edge	family	cop	par	par2	tau	utd	ltd
1	4,3	2	t	0.67	4.86	0.47	0.33	0.33
1	3,5	2	t	0.73	4.32	0.52	0.40	0.40
1	5,2	2	t	0.73	4.45	0.52	0.39	0.39
1	2,7	2	t	0.75	5.18	0.54	0.39	0.39
1	7,1	2	t	0.73	4.51	0.52	0.39	0.39
1	1,6	2	t	0.94	3.75	0.78	0.71	0.71
1	6,9	5	F	1.04	0.00	0.11	0.00	0.00
1	9,8	2	t	0.88	4.65	0.68	0.57	0.57
1	8,10	5	F	1.44	0.00	0.16	0.00	0.00
1	10,11	5	F	1.69	0.00	0.18	0.00	0.00
1	11,12	5	F	4.18	0.00	0.40	0.00	0.00
1	12,13	5	F	3.34	0.00	0.34	0.00	0.00
2	4,5 | 3	1	N	-0.14	0.00	-0.09	0.00	0.00
2	3,2 | 5	2	t	0.41	4.74	0.27	0.18	0.18
2	5,7 | 2	2	t	0.32	5.61	0.21	0.11	0.11
2	2,1 | 7	2	t	0.10	11.05	0.06	0.01	0.01
2	7,6 | 1	2	t	-0.02	8.51	-0.01	0.01	0.01
2	1,9 | 6	2	t	-0.02	30.00	-0.01	0.00	0.00
2	6,8 | 9	33	C270	-0.05	0.00	-0.02	0.00	0.00
2	9,10 | 8	1	N	-0.10	0.00	-0.06	0.00	0.00
2	8,11 | 10	5	F	0.25	0.00	0.03	0.00	0.00
2	10,12 | 11	14	SG	1.02	0.00	0.02	0.00	0.02
2	11,13 | 12	2	t	0.54	6.24	0.36	0.18	0.18
3	4,2 | 3,5	2	t	-0.09	16.53	-0.06	0.00	0.00
3	3,7 | 5,2	2	t	-0.25	8.27	-0.16	0.00	0.00
3	5,1 | 2,7	2	t	0.44	7.25	0.29	0.11	0.11
3	2,6 | 7,1	2	t	-0.22	14.57	-0.14	0.00	0.00
3	7,9 | 1,6	5	F	-0.38	0.00	-0.04	0.00	0.00
3	1,8 | 6,9	1	N	-0.05	0.00	-0.03	0.00	0.00
3	6,10 | 9,8	33	C270	-0.03	0.00	-0.02	0.00	0.00
3	9,11 | 8,10	2	t	-0.04	30.00	-0.02	0.00	0.00
3	8,12 | 10,11	16	SJ	1.02	0.00	0.01	0.00	0.02
3	10,13 | 11,12	33	C270	-0.02	0.00	-0.01	0.00	0.00
4	4,7 | 3,5,2	3	C	0.14	0.00	0.07	0.00	0.01
4	3,1 | 5,2,7	1	N	-0.15	0.00	-0.09	0.00	0.00
4	5,6 | 2,7,1	5	F	-2.39	0.00	-0.25	0.00	0.00
4	2,9 | 7,1,6	6	J	1.01	0.00	0.01	0.01	0.00
4	7,8 | 1,6,9	5	F	-0.13	0.00	-0.01	0.00	0.00
4	1,10 | 6,9,8	6	J	1.01	0.00	0.01	0.02	0.00
4	6,11 | 9,8,10	1	N	-0.03	0.00	-0.02	0.00	0.00
4	9,12 | 8,10,11	23	C90	-0.01	0.00	0.00	0.00	0.00
4	8,13 | 10,11,12	23	C90	-0.02	0.00	-0.01	0.00	0.00

5	4,1 | 3,5,2,7	5	F	0.22	0.00	0.02	0.00	0.00
5	3,6 | 5,2,7,1	5	F	0.72	0.00	0.08	0.00	0.00
5	5,9 | 2,7,1,6	5	F	0.34	0.00	0.04	0.00	0.00
5	2,8 | 7,1,6,9	33	C270	-0.02	0.00	-0.01	0.00	0.00
5	7,10 | 1,6,9,8	1	N	0.00	0.00	0.00	0.00	0.00
5	1,11 | 6,9,8,10	6	J	1.00	0.00	0.00	0.01	0.00
5	6,12 | 9,8,10,11	5	F	0.11	0.00	0.01	0.00	0.00
5	9,13 | 8,10,11,12	13	SC	0.01	0.00	0.01	0.00	0.00
6	4,6 | 3,5,2,7,1	5	F	0.26	0.00	0.03	0.00	0.00
6	3,9 | 5,2,7,1,6	5	F	0.50	0.00	0.06	0.00	0.00
6	5,8 | 2,7,1,6,9	5	F	-0.20	0.00	-0.02	0.00	0.00
6	2,10 | 7,1,6,9,8	33	C270	-0.01	0.00	0.00	0.00	0.00
6	7,11 | 1,6,9,8,10	5	F	0.07	0.00	0.01	0.00	0.00
6	1,12 | 6,9,8,10,11	34	G270	-1.00	0.00	0.00	0.00	0.00
6	6,13 | 9,8,10,11,12	5	F	-0.07	0.00	-0.01	0.00	0.00
7	4,9 | 3,5,2,7,1,6	5	F	0.64	0.00	0.07	0.00	0.00
7	3,8 | 5,2,7,1,6,9	1	N	-0.04	0.00	-0.02	0.00	0.00
7	5,10 | 2,7,1,6,9,8	3	C	0.01	0.00	0.00	0.00	0.00
7	2,11 | 7,1,6,9,8,10	4	G	1.01	0.00	0.01	0.01	0.00
7	7,12 | 1,6,9,8,10,11	5	F	-0.04	0.00	0.00	0.00	0.00
7	1,13 | 6,9,8,10,11,12	5	F	0.11	0.00	0.01	0.00	0.00
8	4,8 | 3,5,2,7,1,6,9	1	N	-0.03	0.00	-0.02	0.00	0.00
8	3,10 | 5,2,7,1,6,9,8	34	G270	-1.00	0.00	0.00	0.00	0.00
8	5,11 | 2,7,1,6,9,8,10	13	SC	0.01	0.00	0.00	0.00	0.00
8	2,12 | 7,1,6,9,8,10,11	14	SG	1.01	0.00	0.01	0.00	0.01
8	7,13 | 1,6,9,8,10,11,12	23	C90	0.00	0.00	0.00	0.00	0.00
9	4,10 | 3,5,2,7,1,6,9,8	5	F	-0.04	0.00	0.00	0.00	0.00
9	3,11 | 5,2,7,1,6,9,8,10	23	C90	-0.01	0.00	0.00	0.00	0.00
9	5,12 | 2,7,1,6,9,8,10,11	13	SC	0.02	0.00	0.01	0.00	0.00
9	2,13 | 7,1,6,9,8,10,11,12	16	SJ	1.00	0.00	0.00	0.00	0.01
10	4,11 | 3,5,2,7,1,6,9,8,10	5	F	0.14	0.00	0.02	0.00	0.00
10	3,12 | 5,2,7,1,6,9,8,10,11	5	F	0.03	0.00	0.00	0.00	0.00
10	5,13 | 2,7,1,6,9,8,10,11,12	5	F	-0.07	0.00	-0.01	0.00	0.00
11	4,12 | 3,5,2,7,1,6,9,8,10,11	3	C	0.01	0.00	0.01	0.00	0.00
11	3,13 | 5,2,7,1,6,9,8,10,11,12	1	N	0.01	0.00	0.01	0.00	0.00
12	4,13 | 3,5,2,7,1,6,9,8,10,11,12	13	SC	0.01	0.00	0.01	0.00	0.00

Table A2.1: D-Vine Copula Tree