Czech Technical University in Prague

Faculty of Electrical Engineering

Department of Economics, Management and Humanities

Pricing of Weather Derivatives

Master's Thesis

Study Program: Electrical Engineering, Power Engineering and Management

Field of Study: Economy and Management of Power Engineering

Scientific Adviser: doc. Ing. Július Bemš, PhD

BSc. Caner AYDIN

Prague 2019

MASTER'S THESIS ASSIGNMENT

I. Personal and study details

Student's name: **Aydin Caner** Personal ID number: **400259**
Faculty / Institute: **Faculty of Electrical Engineering**
Department / Institute: **Department of Economics, Management and Humanities**
Study program: **Electrical Engineering, Power Engineering and Management**
Branch of study: **Economy and Management of Power Engineering**

II. Master's thesis details

Master's thesis title in English:

Pricing of weather derivatives

Master's thesis title in Czech:

Pricing of weather derivatives

Guidelines:

Prepare general research on weather derivatives
Describe main weather indexes, contract types and hedging strategies
Analyze and simulate wind speed data
Evaluate wind speed derivatives based simulated data

Bibliography / sources:

Benth, F. E., Benth, J. Š.: Modeling and Pricing in Financial Markets for Weather Derivatives. Advanced Series on Statistical Science & Applied Probability: Volume 17. 2012
Yamada Y: Handbook of Wind Power Systems, Chapter 3-,Risk Management Tools for Wind Power Trades: Weather Derivatives on Forecast Error, 39-66

Name and workplace of master's thesis supervisor:

doc. Ing. Július Bemš, Ph.D., FEE CTU in Prague, K 13116

Name and workplace of second master's thesis supervisor or consultant:

Date of master's thesis assignment: **31.01.2019** Deadline for master's thesis submission: ________

Assignment valid until: **20.09.2020**

doc. Ing. Július Bemš, Ph.D.
Supervisor's signature

Head of department's signature

prof. Ing. Pavel Ripka, CSc.
Dean's signature

III. Assignment receipt

The student acknowledges that the master's thesis is an individual work. The student must produce his thesis without the assistance of others, with the exception of provided consultations. Within the master's thesis, the author must state the names of consultants and include a list of references.

Date of assignment receipt

Student's signature

Topic Registration Form

Student: Caner AYDIN

Study Program: Electrical Engineering, Power Engineering and Management

Topic: Pricing of Weather Derivatives

Details:

- Prepare general research on weather derivatives
- Describe main weather indexes, contract types and hedging strategies
- Analyze and simulate wind speed data
- Evaluate wind speed derivatives based simulated data

Literature for registration:

- Benth, F. E., Benth, J. Š.: Modeling and Pricing in Financial Markets for Weather Derivatives. Advanced Series on Statistical Science & Applied Probability: Volume 17. 2012
- Yamada Y: Handbook of Wind Power Systems, Chapter 3-, Risk Management Tools for Wind Power Trades: Weather Derivatives on Forecast Error, 39-66

“I hereby declare that this master’s thesis is the product of my own independent work and that I have clearly stated all information sources used in the thesis according to Methodological Instruction No. 1/2009 – “On maintaining ethical principles when working on a university final project, CTU in Prague “.

Date Signature

Abstract

Weather is one of the factors which has a crucial impact on the economy. Weather derivatives and weather insurances are two ways to hedge against unexpected weather conditions. Nowadays, many companies are hedging themselves against non-catastrophic situations with weather derivatives.

Since the world started to focus on renewable energy, wind power plants became popular at energy sector and since reliability is one of the most problematic issues of the renewables, hedging against this unreliability has become more important. Therefore, wind speed and wind power derivatives started to get into market.

The main aim of this dissertation to calculate the price of a wind power put option to hedge against the unreliability of the wind and performing sensitivity analysis to find out the main driver of the price. After providing the general information about weather derivatives such as definition of weather derivatives, differences between weather derivatives and insurances, potential risks which can be hedged against by weather derivatives during first chapter, different types of weather derivatives with different indexes and pay off calculations were demonstrated during the chapter 2. In addition to this, it is possible to find hedging strategies and examples within the same chapter. Statistical studies of the measured real wind speed data including seasonality, trend regression and residual analysis is provided during chapter 3. Modelling the variances of the residuals and transforming residuals into the standard residuals were briefly discussed too. According to the calculated residuals and standardized residuals, random number generation with MATLAB to follow the Monte Carlo Simulation approach was performed during the last chapter. Capacity factor difference between real-measured historical data and simulated data (in percentages), which is the one of the main parameters during calculation the price of the put option, was calculated. After transforming this difference into a parameter which is called as tick size and has monetary unit (according to market conditions), price of the option is calculated. During the last chapter, option prices with different contract durations were provided and evaluated with two different discounting approaches. At the end of the last chapter, a sensitivity analysis was run to find out main drivers of the price which are contract duration and capacity factor difference parameter with monetary unit.

Keywords: Weather Derivatives, Wind Power Derivative, Monte Carlo Simulation, Capacity Factor Difference, Pay-Off Function

Index

List of Tables

List of Figures

List of Graphs

List of Appendices

Acknowledgements

Firstly, I would like to state my sincere gratitude to my advisor, doc. Ing. Július Bemš, PhD for his clear, patient guidance, continuous support, motivation, knowledge and mentoring.

I also express my thanks to Mr. Jan Jandera for his time, support, guidance and understanding and Prof Fred Espen Benth for providing me the relevant data which is used during the statistical analysis.

Finally, I want to emphasize the importance of support of family during my whole master study and thank them with all my heart.

List of Abbreviations

ACF: Autocorrelation Function
ADF: Augmented Dickey-Fuller
AIC: Akaike's Information Criteria
AR: Autoregressive
ARCH: Autoregressive Conditionally Heteroskedastic Models
ARMA: Autoregressive Moving Average
ARIMA: Autoregressive Integrated Moving Average
BIC: Bayesian Information Criteria
CAT: Cumulative Average Temperature
CDD: Cooling Degree Days
CCDD: Cumulative Cooling Degree Days
CHDD: Cumulative Heating Degree Days
CME: Chicago Mercantile Exchange
EEX: European Energy Exchange
GARCH: Generalized Autoregressive Conditionally Heteroskedastic Models
GOF: Goodness of Fit
HDD: Heating Degree Days
KPSS Test: Kwiatkowski–Phillips–Schmidt–Shin Test
LIFFE: London International Financial Futures and Options Exchange
MA: Moving Average
MC: Monte Carlo Simulation
NOAA: National Oceanic and Atmospheric Administration
NY 1: New York Zone 1
NY 2: New York Zone 2
NY 3: New York Zone 3
OTC: Over the Counter
PACF: Partial Autocorrelation Function
pdf: Probability Density Function
Q-Q: Quantile - Quantile
SARIMA: Seasonal Autoregressive Integrated Moving Average
TX: Texas
USFE: U.S. Futures Exchange
WP: Wind Power in MWh
WS: Wind Speed in m/s

1. Introduction

Financial instruments, which are contracts between two or more parties and used as a hedging strategy against potential risks or used for earning returns with some accepted risk level, are called derivatives. One of the most important features of the derivatives is that their value depends on the fluctuation in its underlying asset. Another important point for the derivatives that they have the term "maturity" which means that they are real-time agreements for future possible risks or events [111,112].

Weather derivative is a type of derivative that its underlying assets are related to weather features or parameters and used against different types of weather risks. The importance and the trading volume of weather derivatives are increased significantly during last 20 years [3]. Companies, which are impacted by weather conditions in amount of sales or companies which are facing problems about operating daily activities, use these financial tools as a hedging strategy to prevent bigger losses. These companies are generally operating in tourism, construction, agricultural and energy sectors [4].

Weather derivatives are quite important for especially energy companies. There are different types of weather derivative contracts according to their indexes which are affecting the amount of electricity generated and consumed. Thus, it can be easily said that energy companies are the number one users of these derivatives. The percentages according to the different sectors for the years between 2005-2006, can be found in Figure 1 [18].

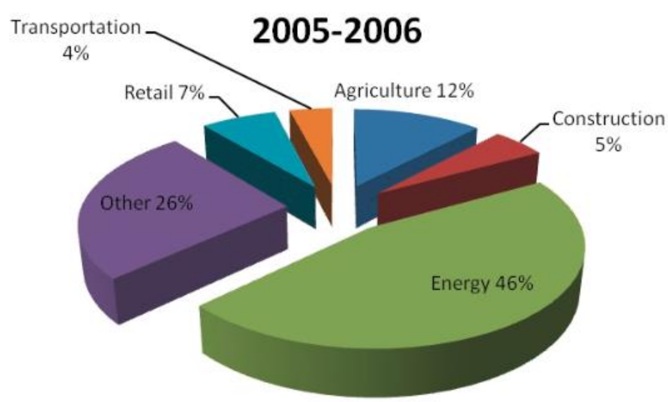


Figure 1. Usage of Weather Derivatives According to the Sectors

An ice-cream producer would face a decrease at their sales amount in a summer season which is cooler than average, or hotels especially providing services for winter sports would lose customers in a winter season which is warmer than an average winter season, or a wind farm would generate and sell less electricity in a season which has less wind speed than expected or in case of having unstable wind duration. Therefore, weather derivatives are crucial hedging opportunities for these kinds of enterprises. This hedging strategy provides stable profitability and there are some benefits for companies to have stable profit such as [10]

- Low volatility provides decreased interest rates for borrowed money
- Increases the value of the company in the stock market
- Reduces the bankruptcy risk for the company

In 1999 which is the year that CME (Chicago Mercantile Exchanges) launched its first weather derivative, one-seventh of the United States' economy was dependent on the weather [7]. In other words, the economy was weather sensitive, and this sensitivity increased during the last 20 years day by day and reached %20 percent of the USA economy [8]. In addition to CME and other exchange markets, these derivatives are also traded in OTC (over the counter) markets. The trades and contracts can be called less official and less transparent in pricing aspect at these types of markets. Currently, CME, which is the leader of the market, is providing weather derivative contracts for 24 US, 6 Canadian, 11 European, 3 Japanese and 3 Australian cities. (47 cities–list of the cities can be found in Appendix 1) [6]. With its first launch, there were 10 available cities. However, the transaction level increased rapidly. In two years, the market size for the weather derivatives reached 11,5 billion $[4]. There are and were other exchange markets which provides weather derivative contracts. These markets will be described in section 1.3.

Determined indexes are used for weather derivative contracts. These indexes are the main values (underlying assets of the contracts) that are used to compare and calculate the difference between what was estimated (or average) and what became real, occurred. CME uses indexes for temperature, frost, hurricanes, snowfall, and rainfall. Since it is easy to measure and find data for temperature, it is the most used index at the market which means that temperature based derivative contracts have the biggest portion at the weather derivative market [28].

Wind speed is another index which is not used by CME. Before its closure in 2008, USFE (US Future Exchange) had wind derivatives as a product within their portfolio [11]. Since USFE was closed and CME does not let wind speed trade, derivative contracts with wind related index could be found generally at OTC markets until EEX launched their wind power derivatives. Starting from September October 2016, EEX (European Energy Exchange) started to provide wind power futures as a weather derivative contract to hedge against the volatility in the wind which is used during generation energy from [86].

At this study wind-related derivatives, modeling wind data and pricing of wind power put option will be discussed. To determine the wind derivative parameters and dynamics, stochastic approach will be used and according to the generated autoregression model and estimated probability distribution, Monte- Carlo simulations will be performed during calculating and evaluating the price & value. Since there are error factors at AR (autoregressive), to reach the most possible result, simulations will be performed many times (50.000).

In summary, this study is going to explain the topics as follows. The first chapter includes a general introduction about weather and wind derivatives and explanation of their importance based on prepared general research on weather derivatives. In chapter two, indexes of weather derivative contracts, types of weather derivatives, payoff functions of weather derivatives, definition and information of Weather Derivative & OTC markets and hedging strategies with weather derivatives will be described. The following chapter provides studies and results for analyzing and modeling wind data and relevant time series' seasonality, autoregression, and residual analysis. In chapter four, analytical and graphical results from Monte Carlo (MC) simulation is demonstrated and evaluated to calculate the price of wind options. The overall summary about the results and overall assessment is described at the end in the conclusion part.

1.1. General Information for Weather Derivatives

With the growing economies, the effect of the undesirable weather on the companies' sales and profit amount increased year by year. Therefore, people started to think about a way which would help them to decrease the amount of risk in other words loss. The discovery of the weather derivatives realized as a result of this self-protection mechanism behavior. The first weather-related derivative contract was executed in USA, in 1997. The contract was between two power companies named Enron and Koch [5, 21]. After 1997, CME launched the first official weather futures market with two standard future contract types in 1999 [6].

1.1.1. What is a Weather Derivative?

Weather derivatives are young financial instruments [5] which provides protection against unexpected weather risks, in other words, they are being used to hedge against undesirable and unexpected weather conditions [12]. Even though they are new innovations in the finance sector, awareness of their importance is increasing day by day.

One of the most important points for understanding the weather derivatives is realizing its difference in the payoff from weather insurances. The payoff of the weather derivatives is done according to the difference between the index amount written on the signed contract and index which occurred.

In addition to the determined weather index which is the underlying asset of the weather derivative contracts, weather derivative financial instruments have other parameters which are also determined and written on the contracts. These parameters are [13, 14,15, 70];

a. Type of the contract: Type of the weather derivative contracts is not quite different from the standard derivatives. They can be call and put options, swaps, futures, etc.
b. Duration of contract (Contract Period): The agreements which are held between different two or more parties have always a start date and an end date. This end date is important since it generally means as the exercise date and it has an important impact while calculating the present value of the contract.
c. Agreement on the methodology of collecting data (location, source, etc.)
d. Determined weather index: According to sectors that companies are working in, they might be impacted by some certain weather outputs such as temperature, wind, snowfall. According to this difference, indexes are changing, and parties are making agreements with the index that they care about. HDD, CDD and daily wind speed are some examples of the weather indexes. (Unlike normal financial derivatives, weather derivatives are using indexes which are based on weather outcome.)
e. Premium: This is the amount of the money that the buyer of the contract pays (Long position pays, short position receives, and premiums are not valid for all types of the derivatives, such as futures)
f. Tick size and maximum payment amount (maximum payoff): Tick size is the amount of the money that will be received to the buyer of the contract according to the difference between the strike level and reality. The maximum price is like a protection for the seller. This is the maximum payoff amount that the buyer will receive.
g. Strike Level

1.1.2. Weather Risks

Since weather derivatives are hedging instruments against weather risks, understanding the meaning of and types of the risk becomes crucial. If it is needed to separate the weather risk in two main basic groups, they will be risks of catastrophic and non-catastrophic events. Catastrophic events occur with low probability. However, they might have bigger impacts. Unlike the catastrophic events, non-catastrophic events have a higher probability to occur with lower impacts. Windstorms, tornadoes, floods, and hurricanes can be called as catastrophic events. On the other hand, small volatilities and deviations from the standard, determined average of weather index, are called as non-catastrophic events [12]. Generally, the aim of the weather derivatives is hedging against these non-catastrophic events.

Potential risks according to the companies' sector are demonstrated in Table 1 which is taken from **[9]**.

Table 1. Economical Dependency of Different Sectors to the Weather

Risk Holder	Weather Type	Risk
Energy Industry	Temperature	Lower sales during warm winters or cold summers
Energy Consumers	Temperature	Higher heating/cooling costs during cold winters and hot summers
Beverage Producers	Temperature	Lower sales during cold summers
Building Material Companies	Temperature/Snowfall	Lower sales during severe winters (construction sites shut down)
Construction Companies	Temperature/Snowfall	Delays in meeting schedules during periods of poor weather
Ski Resort	Temperature	Lower revenue during winters with below-average snowfall
Agricultural Industry	Temperature/Snowfall	Significant crop losses due to extreme temperatures or rainfall
Municipal Governments	Temperature	Higher snow removal costs during winters with above-average snowfall
Road Salt Companies	Temperature	Lower revenues during low snowfall winters
Hydro-Electric Power Generation	Temperature	Lower revenues during periods of drought

1.1.3. Payoff Calculation of Weather Derivatives

As it is mentioned in section 1.1.1. weather derivatives have parameters which have impacts on the calculation of the payoff amount. According to the index and type of the weather derivative, the payoff function's formula is changing.

The most common weather derivatives contract are temperature-based ones [28]. They have common and popular indexes named HDD (Heating Degree Days) and CDD (Cooling Degree Days) which will be explained in the further sections. In addition to temperature-based derivatives, the payoff function of the wind derivatives is demonstrated in formulas 4 and 6. The payoff function showed at the formula 4 is the calculation according to wind speed index. In addition to wind speed indexed derivatives, it is possible to demonstrate the payoff function for wind power indexed derivatives which is described at formula 6.

At CME market, 18 °C is the base value (reference temperature) where the HDD or CDD indexes are calculated [6]. For temperature-based weather derivatives, HDD is generally used for winter season where CDD is used for summer seasons.

During the winter period, HDD is calculated based on the average temperature for each day. If a day is colder than 18 °C, then the difference between 18 °C and the relevant day's temperature is recorded as HDD of that day as seen at formula 1 [4].

$$\mathbf{HDD_i = max\ (18\ °C - T_i,\ 0)} \qquad \textbf{(1)}$$

On the other hand, during summer seasons, again CDD is calculated based on the average temperature and with the comparison between base temperature.

$$\mathbf{CDD_i = max\ (T_i - 18\ °C, 0)} \qquad \textbf{(2)}$$

Average temperature calculation is done with the sum of maximum and minimum values of the daily temperature and divided by two [4]. At formula 3; T_{min} and T_{max} represent the minimum and maximum value of the determined day's temperature respectively.

$$\textbf{Average Temperature} = \frac{\mathbf{Tmin+Tmax}}{\mathbf{2}} \qquad \textbf{(3)}$$

At formula 1 and 2, T_i represents the daily average temperature which is demonstrated in formula 3 and at formula 1 and 2, °C as is used as the unit. However, it is possible to find it with °F. The HDD and CDD are indexes which do not have a certain unit. They represent the temperature difference [17].

An example is demonstrated in Table 1 and Table 2 for calculation of HDD and CDD.

Table 2. Calculation of Daily HDD Values

Days	1	2	3	4	5	6	7	8	9	10	Sum
Average Daily Temperature (°C)	21	15	16	14	20	19	18	17	16	15	-
HDD	0	3	2	4	0	0	0	1	2	3	15

Table 3. Calculation of Daily CDD Values

Days	1	2	3	4	5	6	7	8	9	10	Sum
Average Daily Temperature (°C)	22	17	18	20	16	24	19	17	21	23	-
CDD	4	0	0	2	0	6	1	0	3	5	21

As it can be seen from Table 1 that the days which its temperature is greater than 18 °C has 0 as HDD values. On the other hand, the days demonstrated in Table 2 which temperature is less than 18 °C has 0 as CC value. the payoff function of these two contracts are done with multiplying the sum of the HDD or CDD index for determined period and multiplying it with the tick size (price index) which equals to 20$ in CME [17].

Giving information on payoff calculation related to the temperature-based derivatives is done since they are the most common one. However, since this study will focus on wind derivative pricing, the wind derivative index and payoff calculation are going to be explained too.

Since wind derivative contracts are generally at OTC markets, these contracts are less official than the contracts provided by CME. Therefore, it is possible to say that the base index is dependent on the agreement between the parties. Generally, historical data is used and an average for each day is calculated for the base value, reference point which equals to 18 °C for temperature derivatives [11]. However, as it is mentioned previously, EEX started to provide wind power futures in 2016 and the index description of EEX will be done at the last section.

Payoff function examples can be seen below for wind and temperature-based derivatives;

- Wind Speed Indexed Derivative Future Payoff Function [11]

$$\textbf{Payoff Amount (WS, T): min [} \alpha \textbf{ x (WS – K)]} \qquad \textbf{(4)}$$

where α, K, WS, T represent price index (determined monetary unit per index), determined base wind speed (m/s), daily average wind speed (m/s) and maturity time respectively. Payoff amount should be in the monetary unit. Since wind derivatives are generally traded at OTC markets, the value of the tick size is decided according to agreement between parties. CME is using 20$ as the tick price [17]. The price index for the EEX will also explained at last chapter while making the calculations like its index.

- Temperature-based HDD call option payoff function [4]

$$\textbf{Payoff Function (HDD, T): min [(} \alpha \textbf{ x (max (HDD – K, 0), H)]} \qquad \textbf{(5)}$$

where α, K, HDD, T, H represent price index (determined monetary unit per index), determined base temperature (°F or °C), daily average temperature (°F or °C) and maturity time respectively. Once again, the payoff amount should be in the monetary unit. H represents the maximum payoff amount which is called as the capped amount. This capped amount depends on the agreement which means that it is not valid for all contracts.

- Wind Power Indexed put option payoff function [11]

$$\textbf{Payoff Amount (WS, T): min [(} \alpha \textbf{ x (max (K – WP, 0), H)]} \qquad \textbf{(6)}$$

where α, K, WS, T, H represent price index (determined monetary unit per index), determined base wind power (in MW), daily average wind power (in MW) and maturity time respectively. Once again, the payoff amount should be in the monetary unit. H represents the maximum payoff amount which is called as the capped amount.

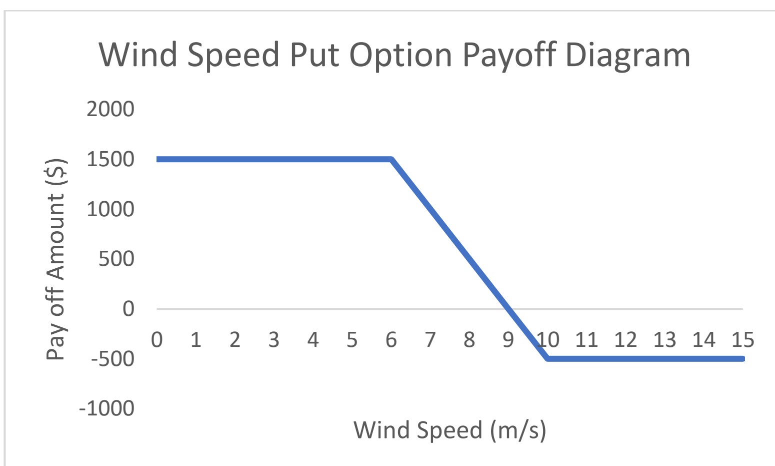
Wind Speed Call Option Payoff Example [4, 11, 19]:

A wind generator company named Karsiyaka is willing to hedge itself against the volatility of wind speed via OTC market with buying wind speed put option. Karsiyaka made an agreement with company Ceynir. The agreement between the companies is done according to the monthly average speed for January. For the base wind speed index (reference point) calculation of average daily wind speed is done from hourly average wind speeds. As the second step, the monthly average speed is set as the base index value (K). All data is taken from a determined and agreed source (web site or governmental offices) for a determined location.

The rule is set as if the average monthly average speed is lower than 6 m/s (like exercise price), Ceynir will pay the difference according to the index value to Karsiyaka. And the capped amount is set as 2000 $ for the payment amount.

Table 4. Payoff Amount Calculation for Different Wind Speeds

Input Parameters	K=10 m/s	α = 500 $	H = 2000 $	Pp = 500 $												
Monthly Average Wind Speed (m/s)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
K (base wind speed) - Monthly Average Wind Speed	10	9	8	7	6	5	4	3	2	1	0	0	0	0	0	0
Payoof According to Difference ($)	2.000	2.000	2.000	2.000	2.000	2.000	2.000	1.500	1.000	500	-	-	-	-	-	-
Pp ($)	500	500	500	500	500	500	500	500	500	500	500	500	500	500	500	500
Final Payoff Amount ($)	1.500	1.500	1.500	1.500	1.500	1.500	1.500	1.000	500	-	(500)	(500)	(500)	(500)	(500)	(500)



Graph 1. Example of Payoff Diagram for Wind Speed Put Option

Graph 1 demonstrates that the put option is providing hedging without any loss until 9 m/s. This value is less than the base value because of the option premium. If the monthly average wind speed is equal or higher than 10 m/s, total loss for the Karsiyaka will be 500 $ which can be covered with extra energy sales.

1.1.4. Categorization of Weather Derivatives

Even though the first weather derivative contract (it was a swap) was made according to temperature during the winter, it was obvious that other weather-indexed derivative types were needed. A weather derivative index becomes important for the buyer or the seller if it is related to its sector. The types of the derivatives according to the indexes are;

- Temperature
- Wind Speed & Wind Power
- Frost Days
- Hurricane
- Snowfall
- Rainfall
- Humidity

For agricultural companies, frost days and temperature indexes are important while the wind and hurricane is important for the energy sector's companies [7, 20].

1.2. Weather Derivatives' Differences

Weather derivatives took their places at the market after the launches of both standard derivatives and weather insurances. It can be said that weather derivatives took some of these financial instruments' features. Even though, there are some common features about the related sector or the payoff methodology, the differences between weather derivatives and weather insurances and standard derivatives cannot be neglected [10].

1.2.1. Differences Between Weather Derivatives and Weather Insurances

By the time weather derivatives started to be used as a risk preventing tool, weather insurances were being used to protect companies' assets against unexpected weather conditions. These conditions can be named as catastrophic conditions as mentioned at 1.1.2. Weather Risk. Even though, weather insurances had been used as the prior tool, they were able to cover losses against only catastrophic conditions. The need for the weather derivatives showed up when the market realized that they need to cover losses for the volume of demand and selling [8].

The difference can be seen from the payoff functions of derivative and insurance contracts. The payoff is calculated according to underlying assets for weather insurances. On the other hand, the payoff amount is calculated according to the index of the underlying asset for weather derivatives [2].

Table 5. Weather Insurance & Weather Derivative Risk Level and Probability

	Weather Derivatives	Insurance
Risk Level	Low	High
Probability	High	Low

Table 5 demonstrates that weather derivatives are covering the non-catastrophic losses where the weather insurances cover the catastrophic ones. Other important difference points are shown below [10, 13, 23, 70];

- Insurance contracts require proof of damage to pay out on the contrary of weather derivatives
- Weather insurances are generally more expensive than weather derivatives
- One of the main differences is that the weather derivatives can be used to make a profit in addition to hedging against risks even with speculations.

1.2.2. Differences Between Weather Derivatives and Standard Derivatives

Weather derivatives can be called as a combination of weather insurances and financial derivatives in terms of the features. However, weather derivatives are unique in terms of payoff and evaluation. At this chapter differences from the standard financial derivatives will be explained.

- Firstly, it is important to see that the underlying assets of weather derivatives, independently from its type, are meteorology dependent in other words, meteorological variables. These indexes are not directly tradable since they do not have monetary unit [24].
- It is known that; weather derivatives can be used for both hedging and making profits with speculations. However, if hedging is going to be focused on, it is not possible to mention a big difference. Weather derivatives are using as a quantity hedging where financial derivatives are being used for price hedging [10].
- It is possible to perform arbitrage pricing at financial derivatives. Since the weather derivatives' indexes are based on meteorological variables. Meteorological variables can be modeled with a probability function accurately. Thus, this prediction blocks arbitrage [24].

These three key points can be defined as where the weather derivatives differentiate from standard financial derivatives.

1.3. Weather Derivative and Over the Counter Markets

Derivatives are mainly being traded at two different types of markets which are official exchange derivative markets and over the counter markets. During the past years four of the exchange markets came to the fore;

- **CME:** The Chicago Mercantile Exchange Market is the leader of the weather derivative trade markets. CME describes themselves as "As the world's leading and most diverse derivatives marketplace, CME Group (www.cmegroup.com) is where the world comes to manage risk "[25]. Currently, they are offering monthly or seasonal derivative contracts according to four indexes; United States: HDD and CDD and Europe: HDD and CAT, snowfall and hurricane [26].
- **Liffe (London International Financial Futures and Options Exchange):** Until 2004, London International Financial Futures and Options Exchange was providing derivative contracts based on daily average temperatures in 3 main European cities; London, Berlin, and Paris.
- **USFE (United States Future Exchange Market):** United States Future Exchange Market was the first exchange market where it was possible to find wind-related derivatives. However, they stopped trade transactions in 2008 [11]. Before their closure, they were providing wind derivative trading opportunities in 7 locations. These 7 locations were chosen from Texas and New York (2 of them from New York and 5 of them from Texas) [11, 16]

- **EEX:** European Energy Exchange is a Germany origin exchange market platform which generally provides transactions for energy and energy-related products. It is possible to find wind power related futures at EEX [71, 87].

It is needed to agree on deciding a reliable weather data source. Collecting temperature-based data rather than other types is much easier. Therefore, weather derivatives except temperature indexed are generally being traded at OTC markets. As a result, it is important to understand the working rules of OTC markets.

OTC markets are trading markets where it is not possible to find a center or physical location. The deals and bargains are made between at least 2 parties according to the rules that they agree on. One important point of these deals is being less transparent. Since whole agreements are done between determined parties via e-mail, telephone, it is not possible track the price levels and rules at the relevant contracts [27].

Even though they are less official than the contract providing by official exchange markets, the amount of the transactions is quite considerable. According to [11], the value of the transactions in 2011 was approximately $36 million.

Even though weather derivatives except temperature indexed are generally are being traded at OTC markets, it is possible to find temperature-related derivatives at this market. The percentage of the temperature-indexed derivatives are taking the lead in OTC markets too. Figure 2 shows the percentage levels of each type according to PwC (Price Waterhouse Cooper) which is one of the biggest auditing company.

It can be seen from Figure 2 that CDD and HDD indexed contracts took the lead. Especially in 2005, it can be easily said that after temperature and rain indexed derivatives, wind derivative took the 3rd place according to the number of the transactions [28].

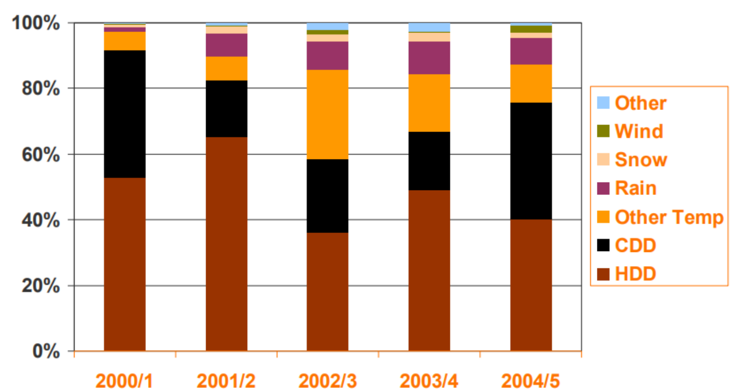


Figure 2. Distribution of Number of Contracts by Type

1.4. Wind, Wind Energy, Wind Derivatives Definition and Importance

General definitions and examples were given for weather derivatives until this chapter. As it is mentioned at the Introduction part, wind derivatives will be the main scope of this dissertation. Therefore, the definition of wind, energy generated by wind and importance of wind derivatives is explained at this chapter.

1.4.1. Definition and Importance of Wind

Wind is caused due to the air motion, it happens because of the air pressure happening in the different location of the earth. It can be said that the air moves from high pressure to low pressure. This air motion is called as the wind. Importance of the wind changes for the companies according to their sectors. Since the renewable energy growth rate is increasing, the main impact of the wind comes true for the energy sector. Since it is a cheap and unlimited energy source, the intention on the using wind to generate energy will continue to grow [29].

Horizontal directed wind is affecting the amount of the energy generated where the vertical part is affecting the temperature that is sensed. While mentioning the direction, the wind blows with different horizontal directions. For the energy sector, even though thanks to the latest technology the wind blades' angle can be arranged according to wind angle, estimation of the wind direction is important for less costed wind farms which do not use the latest technology to reduce the cost [2, 11].

1.4.2. Wind Derivatives and Importance of Their Usage

Energy is the first sector that comes to mind when the wind is mentioned. The amount and the price of the electricity generated depend on the power, speed and volume of the wind.

Predicting or having stable wind for future generation periods make companies stronger at the market in terms of the selling price. This opportunity lets wind farms to produce and sell at a stable environment with a higher price. However, real life is not like how it is wanted. Since the volatility is changing a lot for the wind, it makes difficult to make selling agreements between parties with higher prices which are the real value of the electricity generated [2].

Wind derivatives importance shows up in two main points;

a. Preventing opportunity loss or penalty
b. Preventing a decrease in the general amount of sale

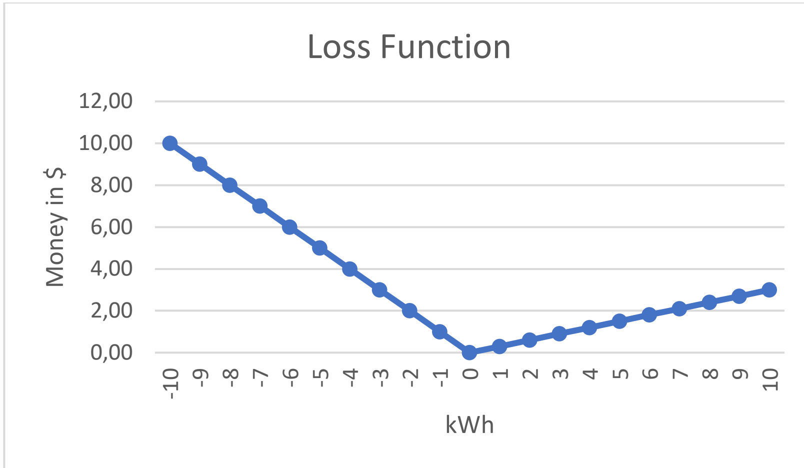
Apart from maintenance and operation downtimes, wind speed and duration are the main factors affecting the amount of electricity generated. Wind farms are making contracts between different parties to provide electricity. The amount and the price written on the contracts are important due to the volatility of wind which might cause opportunity loss or penalty [2].

Example: A contract between buyer and seller has the values of the electricity price per $0,20/ kWh and penalty amount as $1 for not providing per kWh. This contract puts the seller into risk in two ways; penalty and opportunity loss. Opportunity loss caused due to the normal average price which is $0,50/kWh. The function of both penalty and the opportunity loss can be seen at Graph 2.

Since the penalty amount is greater than the difference between the normal average electricity price and the contract price, the slope is higher on the left side.

Not only losses due to the penalty or price difference, but also the general sales amount to whole customers affect the revenue of the wind farms. This is the point where hedging is needed with wind derivatives.

A wind derivative can cover whole the losses mentioned above with a proper tick size and base value index.



Graph 2. Loss Function for the Example

2. Weather Derivatives Indexes, Contract Types, and Hedging Strategies

Information related to weather derivatives such as their indexes, contract types and hedging strategies will be explained in detail at this part.

2.1. Weather Derivative Indexes

Each derivative contract has an underlying asset. This underlying asset is one of the key points while pricing the contract and calculating the payoff function. Generally, at normal derivatives, the underlying assets are set with monetary units. On the other hand, for the weather derivatives, underlying assets are determined based on weather-related indexes. After determining the index, the transformation to the money can be done with price indexes [10,16].

2.1.1. Wind Derivative Indexes

Nowadays, wind-related derivatives are generally traded at OTC markets. Previously, many official exchange stock markets where it was possible to find wind derivative contracts were existing. Especially USFE was providing different types of contracts. However, it stopped its operations in 2008 as it is mentioned at chapter 1.3.

Currently, it is also possible to find wind derivative contracts at EEX (European Energy Exchange) and Nasdaq OMX which have only wind power indexes at their portfolio. However, it is known that in addition to wind power index, wind speed was and is also used as an index at the contracts [11,71,72,73].

2.1.1.1. Wind Speed Index

As it can be understood from the name of the index, this index is related to the wind speed which is measured at a determined location for a determined period.

The difference of this index from the HDD or CDD, the wind speed is not transformed into another parameter, but it is used as it is. For the temperature weather derivatives, the underlying asset of the contracts were generally called as HDD, CDD or CAT (Cumulative Average Temperature) whereas for the wind speed derivative contracts, the underlying asset is not called as something else but measured wind speed at a determined location for a determined period and the unit of the index is generally in m/s.

There are some important parameters and points which should be determined for the contracts such as;

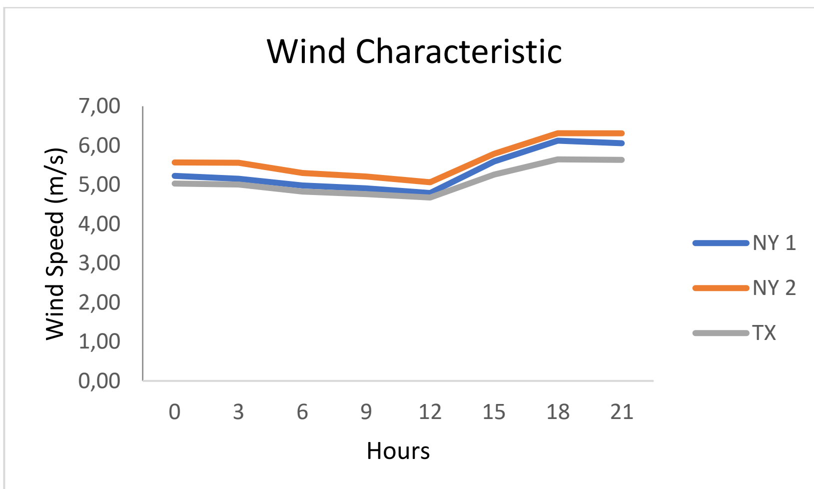
- Reference wind speed
- Maturity time
- The altitude of the wind speed measurement

Since it is known and mentioned previously that, USFE was the biggest wind derivative contract provider, its reference approach will be described here. The mentioned reference approach was also mentioned and used at many different studies. For determining the reference wind speed, which was called as a reference temperature, and which is also equal to 18 °C for temperature derivatives, USFE was using the daily average value of the last 20 years for the determined period [16].

The altitude of the measurement is another factor and parameter which affects the contracts since wind speed increases with height. The altitude measurement should be performed according to requirements. For example, for the wind farms or for a single wind turbine, the altitude of the measured wind speed should be close to the height of the turbine tower. If the measurement at the required altitude is not possible, it is possible to convert the measurements for different levels of altitudes with a determined formula. The relevant formula was described at chapter 3.1 with formula 14.

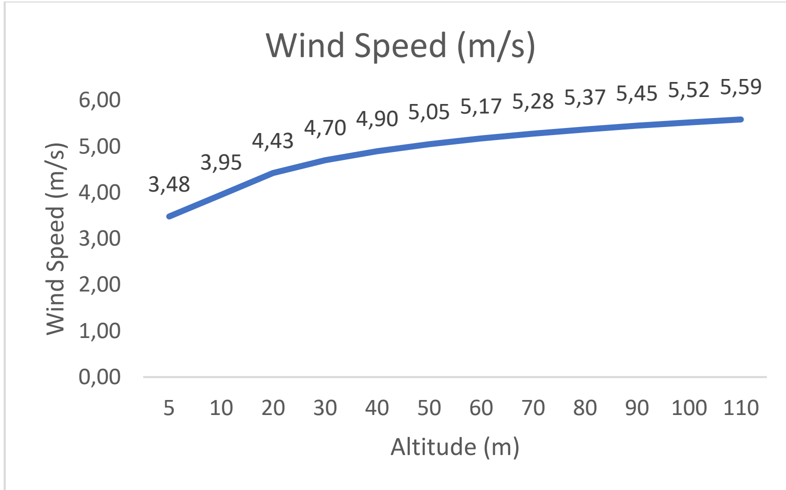
The maturity date is a common factor for each derivative contract and can be determined according to requirements. However, contracts or agreements are generally done weekly, monthly, quarterly or annually [4,11,16,74].

Graph 3 demonstrates the wind speed characteristic for the determined locations for this dissertation which are in New York and Texas. The characteristic of the graph demonstrates the diurnal cycle of the wind speed for the mentioned locations. Even though the measured wind speed values are different, the daily behavior of the wind is similar for the locations.



Graph 3. Wind Characteristic According to Different Locations - New York Zone 1, Zone 2 and Texas

Graph 4 demonstrates the wind speed change in line with the altitude. As mentioned before and at [11,74] the speed of the wind increases with an increase at altitude. During drawing Graph 4, the average value of the [30], which is the main data set of this dissertation, is taken as the reference wind speed and 80 meters is taken as the reference altitude and these values are used with Formula 14 and wind speeds in m/s at different altitude are calculated.



Graph 4. Wind Characteristic According to Different Height

2.1.1.2. Wind Power Index

Wind Power is the second type of index which is used for wind derivative contracts. This index is linked to the measured wind speed. The same measurement of the wind speed can be applied for this index calculation. However, one extra calculation step is needed to find out an answer for the following question; how much power can be generated from the measured wind speed.

This index is used at the current contracts which are provided by EEX and Nasdaq OMX. The power index at these contracts are represented in percentages as shown below;

$$\%\ \mathbf{wind\ power\ index} = \frac{\mathbf{Actual\ kWh\ produced}}{\mathbf{Maximum\ Energy\ that\ might\ have\ produced\ during\ the\ same\ period}} \qquad \mathbf{(7)}$$

The index price for the contracts is determined as 1,13 $ (1 €) per 1% capacity factor difference. The key factor while calculating this index that, wind turbines cannot extract whole energy coming from the wind. The power which can be extracted from the wind is calculated as shown below [11];

$$\mathbf{P} = \frac{\mathbf{A\rho}}{\mathbf{2}}\mathbf{v^3Cp} \qquad \mathbf{(8)}$$

Where P represents the amount of power extracted (watt), A represents the swept area of the wind blades in m^2, ρ represents the air density (kg/m3), v is the wind speed (m/s) hitting the blade and C_p is called as the power coefficient of the wind turbine. C_p is one of the most important parameters for the wind turbine. This parameter shows how much power can be extracted from the coming wind directly. The parameter values are provided and listed for each wind turbine by their suppliers. It can be said that this coefficient is more or less related to efficiency of the wind turbine and it cannot exceed a theoretical calculated limit.

According to Betz theory, the maximum value of the power coefficient equals to 0,59 which means it is not possible to extract more than 59% of the power from the coming wind [11, 75, 76].

2.1.2. Temperature Derivative Indexes

Currently, there are 3 main indexes are used at the exchange markets which are HDD, CDD, and CAT [17].

- HDD: Heating Degree Days
- CDD: Cooling Degree Days
- CAT: Cumulative Average Temperature

2.1.2.1. HDD (Heating Degree Days)

HDD is one of the indexes used at temperature-based weather derivatives and it represents the heating degree days. The value of the HDD has a meaning of the total number of degrees which are lower than the determined base temperature. The base temperature is calculated as 18 ° C for Europe and 65 ° F for the USA.

The calculation of the HDD index was shown at formula 1 previously which can be seen also below. An example was already provided with Table 2.

$$\mathbf{HDD_i = max\ (18\ °C - Ti,\ 0)} \qquad \textbf{(1)}$$

The idea behind the HDD is that it is not necessary to use energy for heating for the days which are warmer than this temperature [8, 17].

2.1.2.2. CDD (Cooling Degree Days)

CDD is the second most popular index like HDD for temperature-based weather derivatives. Once again, the calculation of the CDD index was shown at formula 2 previously. The idea behind the calculation is completely same with HDD.

$$\mathbf{CDDi = max\ (Ti - 18\ °C,0)} \qquad \textbf{(2)}$$

On the contrary to the HDD, CDD represents the total number of degrees which are higher than the base, reference temperature which is same with HDD.

CDD is mostly used at US-based contracts. It can be also used in Europe but according to the CME, European OTC markets and countries are using CAT. Therefore, CME provides contracts to the European countries with CAT instead of CDD [8,17].

2.1.2.3. CAT (Cumulative Average Temperature)

As mentioned at the previous section, this index is used by the European countries. CAT indexed contracts are provided by CME for the summer months. The index represents the value of the accumulated daily temperatures, the daily temperature is calculated as shown in formula 3 previously.

*This approach is valid for all temperature-based indexes described at this chapter. [8,17].

$$\textbf{Average Temperature} = \frac{\mathbf{Tmin+Tmax}}{\mathbf{2}} \qquad \textbf{(3)}$$

2.1.3. Snowfall

One of the other weather derivative contract types provided by CME is snowfall weather derivatives. It is provided for the US cities; Boston, New York, Chicago, Minneapolis, Detroit, Newark, Baltimore, Columbus, and Colorado.

The index for this contract type is the daily amount of the snowfall in inches and it is priced in dollars [77,78].

2.1.4. Rainfall

Rainfall indexed derivatives is another type of derivatives which is related to precipitation. According to [82, 83], it is possible to have two different rainfall related indexes. The index can be determined according to buyer's active working sector.

a. Total number of rainy days within the determined period
b. The total amount of rainfall within the determined period

CME used rainfall as an index at its contracts starting from 2010. The index that they are using at their contracts is the one which is written at the option b.

The exact description of the index is; "The CME Rainfall Index is equal to the sum of all actual rainfall recorded in the specified city during the contract period." [82, 83]

2.1.5. Hurricane

Hurricane contracts are also provided by CME. There is a determined contract index as an underlying asset. This index is called as CME Hurricane Index. The index is calculated according to possible damage that might be happened due to a hurricane. The index is calculated with a combination of two parameters; duration of the sustained wind speed and the radius of the hurricane.

There are some other indexes which classifies the hurricanes. The classification is done between 1-5. It represents the power of the storm. However, it does not represent the potential damage. Therefore, CME considers these two parameters to calculate the potential damage [84].

2.1.6. Humidity

Humidity index is not used as much as under indexes. In other words, it is used rarely. The index is called as critical humidity day (CHD). This index represents the total number of days which has humidity level more than the critical value [85].

2.2. Weather Derivative Types

Different types of weather-related derivatives are going to be explained at this chapter. At the end of the chapter, wind specific derivative examples will be demonstrated.

2.2.1. Weather Futures

As it can be understood from the name of "Future", these contracts are legal deals between two different parties, which will be related to buying or selling of an agreed underlying asset, happening not on today but in the future.

The terms of "futures" and "futures contract" have same denotations at the market. This contract type is counted as one of the weather derivative types. As it is mentioned at the beginning of the chapter since they are legal contracts, these contracts oblige the buyer & seller to buy & sell the relevant index which is the underlying asset of the contract on a determined future date.

The importance and popularity of this type of weather derivatives is also increasing as the other types. According to the determined index, the pay off- function is estimated according to the price index. Currently, CME is using 20$ as the price index. The price index can be differentiated according to the agreement between the buyer and seller. For these contracts, there is no need to pay the premium. The price calculation of a weather future contract with wind speed index can be found as an example below;

$$\textbf{FP (Index, T)} = \alpha \text{ x } \textbf{(max (WS – K)} \qquad \textbf{(9)}$$

where α represents tick size (monetary unit), K represent determined base value (wind speed in m/s) for the determined time point, T (maturity date) [8,11,22,79].

2.2.2. Weather Options

After the futures, options are one of the other types of weather derivatives which can be found at the markets very oftenly. It can be said that the trading approach of options is not so different from the future. There are two main difference between the options and the futures.

For the options, on the contrary to the futures, the owners of the option contract (long position) do not have to exercise the agreement, option. They have right about not exercising the option. In that case, the maximum loss for the option holder (long position) would be the premium amount of the option.

Now that option premium is mentioned, it can be said that the term of the premium is the second difference between the options and futures. Option premium means that the amount of the money which should be paid by the buyer (long position) to the seller (short position).

Weather options also have two different types as the options which have different indexes at the stock market; put and call which provides the buyer right to sell and right to buy respectively.

Even though the pay-off functions of these contracts are mentioned as an example at Chapter 1.1.3., it can be found once again below;

$$\textbf{Pc} = \alpha \text{ x } \textbf{(max (Determined Weather Index(observed) – Determined Base Weather Index, 0) (10)}$$

$$\textbf{Pp} = \alpha \text{ x } \textbf{(max (Determined Base Weather Index - Determined Weather Index(observed), 0) (11)}$$

where α represents once again the tick size in the monetary unit. The determined observed weather index and determined base weather index represents the stock and exercise indexes respectively [5,11,22].

2.2.3. Weather Swaps

Another type of weather derivative is weather swaps. These contracts are again signed between two parties. The aim of these contracts is changing risks occurring due to the weather volatilities to have a more stable cash flow for themselves.

According to the premium perspective, swaps are like the futures. Since it is common & mutual benefit or advantage, there is no need to pay a premium. In addition to this, swap contracts are generally signed as forwards and futures where there is and there is not a cap for the amount of payment respectively.

The payments of the swap depend on the agreement. However generally, one party pays a determined fixed price where the second party uncertain (variable) price [11,80].

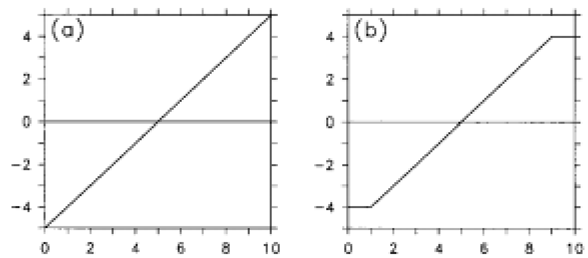


Figure 3. Swap Pay off Function with Uncapped and Capped Scenarios

As it can be seen from Figure 3 [81], the payoff amount can be capped (b) or uncapped (a) according to the agreement between the parties.

2.2.4. Wind-related Other Types of Contracts

Types of weather derivative contracts described at previous chapters are generally valid for every weather index. At this chapter, some extra types which are valid for wind-related contracts will be described. There are wind specific contracts named as [11];

- Wind Caps & Floors
- Wind Speed Cash Settlement Agreement
- Wind Farm Derivatives

2.2.4.1. Wind Caps & Floors and Wind Speed Cash Settlement Agreements

This agreement can be called as a combination of call options. For these contracts, there are again a buyer and a seller. The buyer will receive the money for each day where the wind speed exceeds the reference wind speed (exercise known as strike price too) at the end of each determined duration. The duration which is mentioned as period can differ as a day, week, month etc.

On the other hand, wind speed floors can be called as a combination of put options. The payoff of this contract is similar to the pull options. In this case, the buyer will receive the money if the wind speed cannot exceed the reference wind speed (strike price).

The pay-off functions of both types can be found below;

$$\textbf{Wind Cap} = \sum_{k=0}^{N} \frac{\max(WSk - K, 0)}{N} \qquad \textbf{(12)}$$

$$\textbf{Wind Floor} = \sum_{k=0}^{N} \frac{\max(K - WSk, 0)}{N} \qquad \textbf{(13)}$$

where N represents the total number of the periods (day-week-month-years etc.). As it can be seen from the payoff functions, the methodology and approach are quite similar to the options. WS_k is the average wind speed for day k and K is the reference wind speed. Both of these parameters are in m/s.

If the contract is based on daily durations, these contracts are called as Wind Speed Cash Settlement Agreements. The buyer receives a determined amount of cash flow for each day from the seller according to the type of contract (cap or floor) [11].

2.2.4.2. Wind Farm Derivatives

Many types of derivatives are explained in this chapter so far. The scope of the derivatives can be valid for a company, producer, power plant or a single wind turbine.

Since this dissertation is mainly focused on the wind type of the derivatives, it is better to mention another type of wind derivative contract called as wind farm derivatives.

For the wind turbine farm owners, it would be more efficient to hedge against risk not only for one wind turbine but for the whole wind farm.

The contract validation and the rules will be the same. However, there will one difference at the calculation part. Since it will be customer specific due to the different size of the wind farms, these contracts can be called as a tailored contract.

As it is mentioned at 2.1.2., for the wind power index, there is a coefficient which is provided by the producers called as power coefficient which has a direct impact on the efficiency. The contracts are calculated with an index which has an average efficiency coming from the whole wind turbines [11].

2.3. Weather Derivatives at Exchange Markets

The initial information for the exchange markets where weather derivatives were described previously during chapter 1.3. (Weather Derivative and Over the Counter Markets). As it is mentioned at that chapter weather derivatives are not only traded at the exchange market but also at over the counter markets. This rule is valid, especially for non-temperature indexed weather derivative contracts.

If it is needed to recall the famous exchange markets where you can find, or you could find weather derivatives now and, in the past, respectively, CME, USFE, Liffe, and EEX should be listed. Currently, CME and EEX markets are the most active ones. During this chapter, the contract conditions and parameters of the EEX and CME will be described.

Snowfall, temperature, hurricane index related weather derivatives can be found at CME (at its website). Options and futures are the contract types provided by the CME. As it is mentioned, the temperature-based weather derivatives have the highest amount of transactions. Therefore, the contract parameter explanation is done according to HDD option as shown in Table 6 [108].

Minimum Price Fluctuation	1 Index Point (= $20 per contract)
Pricing Unit	Dollars per index point
Trading Hours	MON-FRI: 8:30 a.m.-3:15 p.m.
Last Trade Date/Time	Second Exchange business day after the futures contract month, 9:00 a.m.
Contract Months	HDD: Nov, Dec, Jan, Feb, Mar plus Oct and Apr
Strike Price Intervals	HDD: 1 index point in a range of 1 to 3200 index points
Exercise Procedure	European Style
Product Code	Changes according to location
Underlying Contract	HDD Index

Table 6. CME - HDD Contract Parameters

Table 6 demonstrates the underlying asset of the contract, contract duration, and its index monetary unit. In addition, the basic conditions, it is possible to see the trading time interval and option type which can be European or American.

EEX contracts for wind power futures have the same rules, parameters, and conditions. Since they provide wind power futures, the underlying asset is based on wind power, (capacity factor of the plant) and the fluctuations are changing within two decimals. (in €/h with two decimals; 1 % ≙ 1 €/h) [76].

A risk which can be called as a disadvantage for exchange market derivatives is the location. It is called as location basis risk. Since it is not possible to find a contract which is valid for everywhere, it is possible for someone who is looking for hedging for a location that is not listed at CME or EEX will face with that risk. If someone who is in Serbia is imagined, this problem would occur because the USA provides contracts for the cities which are listed in Appendix 1 and EEX provides contracts for Germany and Austria. Nevertheless, the amount of transactions is getting increased day by day [109].

It is needed to mention the clearing houses which are one of the main differences for the exchange markets in respect to OTC markets. Clearing house is working to oblige the contract holders to meet the requirements of the contracts. According to the Bank of International Settlements, it is nice to have clearing office to diverse the risk happening due to the counterparties of the contracts [109].

As it can be seen from Table 6, it can be said that the weather derivatives which are provided by the exchange markets are very standardized and it really depends on the location base.

2.4. Hedging Strategies and Examples with Weather Derivatives

At this section, the hedging strategies against the unexpected weather conditions will be evaluated. The hedging strategies can be done via weather-related derivative contracts or without them. Both scenarios will be evaluated.

2.4.1. Hedging Against Weather Risks with Weather Derivatives

The hedging strategies and the position at the strategy depend on the company's sector. It is possible to be at short or long position according to sector and aim. One short and long positioned examples with payoffs is going to be given for different sectors and for different types of indexes.

Example 1: A natural gas provider would demand a colder winter instead of a warmer one. This is one of the hedging opportunities which is provided by CME. (Warm winter months: - Put HDD). The gas provider can hedge itself with buying a put option. This scenario is valid when it is expected to have a lower HDD than the average.

In that case, the company should buy an HDD option and determined the exercise price. Since it is expected that the exercise index would be higher than the expecting index, he can get some payoff to cover its losses due to the warm winter. The maximum loss will be limited with the option premium for having the contract which would happen when a colder winter exists. In this case, the cost of the option premium can be covered with the extra revenue which is coming from the extra selling. Graph 1, which is demonstrated in the first chapter, is an example of a payoff function for a put option

Example 2: This example would help to understand the hedging strategy in a different way. During the first example the company was at long position. At this case, it is assumed that a wind farm owner wants to hedge himself/herself against lower wind speed. The company provides wind option contract which means that the company owner will be at short position in this scenario. The company is providing wind call options. In this scenario, the wind farm owner will receive the premium money and it will be the owner's maximum profit. In this case, if the measured wind speed will be lower than the exercise point, the wind turbine owner would earn less revenue because of the less electricity generation and thanks to the premium the owner can cover some of these losses. If the contrary situation occurs, the owner needs to pay the payoff. However, thanks to higher wind speed the owner is able to generate and sell more which helps the owner to cover the payoff cost.

In addition to these examples, there are many companies from different sectors which can use weather derivatives to hedge against weather risks;

- Concert or Outdoor Activity Organizers ➔ Rainfall or temperature related weather derivatives
- Agricultural Activities ➔ Snowfall, frost, rainfall or temperature related weather derivatives
- Energy Companies (renewable) ➔ Wind, or sunny day related weather derivatives
- Hotels for Summer Activities ➔ Temperature and rainfall related weather derivatives
- Hotels for Winter Activities ➔ Temperature and snowfall related weather derivatives

2.4.2. Hedging Against Weather Risks without Weather Derivatives

Weather derivatives are not the only way to hedge against the weather risk. It is needed to mention one of the basic rules of reducing the risk; portfolio diversification [117].

For weather-related risks, it is possible to reduce the risk with contrary investment opportunities. As a second way to hedge weather-related risks to have different investments at different locations. For example, for a wind turbine owner, the risk can be reduced by planting the wind turbines into different locations rather than having all of them in one big plant [11].

3. Modeling Wind Data

As mentioned during section 2, there are two types of wind derivatives according to their indices;

1. Wind derivatives with wind speed index
2. Wind derivatives with wind power index

Wind speed index measurement is relatively easier than the measurement of wind power index. Therefore, derivatives with wind speed index are more common than derivatives with wind power index. Since it is more common to use wind speed index and since it can be calculated according to daily average wind speed, modeling the daily average wind speed data is more popular. However, if it is wanted to provide a study related to derivatives with wind power index, calculation of the amount of the energy generated via wind is becoming crucial. Therefore, at this point, modeling the data according to hourly average wind becomes necessary again [4].

3.1. Analytical Analysis of Data

The data used for the analysis at this dissertation is provided by Prof Fred Espen Benth in respect to his work [16] with Jūratė Šaltytė Benth. The communication, interview is done via e-mails [30]. According to [16], the data was taken from USFE official web site. However, since USFE does not provide any services or contracts anymore, it was needed to communicate with the author of [16] who is Prof Fred Espen Benth. The data of three locations from New York and one location from Texas was provided by him. (As mentioned before these are the four of seven locations which were listed as wind derivatives' locations by USFE.

It is an obvious fact that and also mentioned during previous sections, wind speed changes according to altitude. The relevant measurement should be performed according to the companies' needs or according to the height of the owned wind turbine. For example, since the average height of a wind turbine is 80 meters (280 feet) [32], a wind farm company should choose the index from wind speed data which is measured with approximately 80 meters height.

According to the conversation with Prof Fred Espen Benth, he does not recall the height of the measurement [30]. Furthermore, he did not mention the height of the data at [16] too. Since there is a lack of knowledge about the height of the data it is assumed that the data was measured at 80 meters height. Even though the certain measurement altitude is not known, and an assumption is made, it is also possible to check whether the assumption is close to reality or not. Therefore, the average wind speed is checked from a governmental web site [113] and found out that the wind speed values according to the assumption is close to reality. According to [113], the annual average measured 80-meter wind speed is close to 5,5 m/s. The average annual wind speed of data provided by [30] is approximately 5,37 m/s.

On the other hand, even though if a data is measured at different height then 80 meters, it is possible to convert the data as it is measured at 80 meters height or for any other altitude with an easy formula [11]. The relevant formula for estimation is shown below;

$$\mathbf{V(h)} = \mathbf{V(href)}\frac{\ln(\frac{h}{zo})}{\ln(\frac{href}{zo})} \qquad \textbf{(14)}$$

At formula 14, V(h) and V(h_{ref}) demonstrate the wind speed data (m/s) at determined height (m) and measurement (reference) height respectively. Z_o represents the roughness length (m) of the surface which equals to 0,0002 at sea and 0,03 at flat terrain with grass or very low vegetation, airport runway [33].

The data provided by the author of [16], was measured 8 times during a day with 3 hours intervals between each measurement. The average wind speed was calculated from these 8 samples which are valid for each day. In addition to altitude assumption, it is assumed that the calculated average daily wind speed will remain same during the day, in other words, it is assumed that wind speed does not have fluctuations over the day. The assumption will be valid while calculating the wind power according to both real and estimated data too.

The data also includes the 29th of February for leap years. Since the estimation will be done with Monte Carlo (MC) approach at MATLAB, to make the code easier and simplicity of the whole study, the data sets of those 29th of February were removed. There was no lack of data and the total amount of data is equal to 7.300 days which represents 20 years (1987-2007).

3.1.1. Box - Cox Transformation and Optimum Parameter Calculation

Having the data with normal distribution is making analysis easier. Box- Cox transformation which was invented by George Box and Sir David Cox is one of the techniques which helps not only to remove positive or negative skewness but also other complicated parameters, is used to make this transformation [34, 35].

Since the same data is used with [16], it is needed to check the parameters used at that article to make comparison. According to [16], Box – Cox parameter, λ, is chosen as 0,2 after many iterations. It is mentioned that this value is the optimum value for the data set to make it symmetric. The optimum value can be determined via software or iterations. It is known that the data set is positively skewed (longer right tail.) Choosing the related parameter with 0,25 step size between -3 and 3 is one of the common approaches to find out the optimum value [16,36].

$$y(\lambda)=\frac{(y^{\lambda})-1}{\lambda} \qquad \textbf{(15)}$$

For the λ=0 case, the transformation formula is calculated as y(λ)= ln (y). At both formulas, y(λ) demonstrates the transformed form where y demonstrates the raw data. Iteration methodology was used to determine the optimum λ value according to the skewness result. At formula 15, y represents the original data, y^{λ}, represents, box-cox transformed data and λ represent the box-cox coefficient.

The results at Table 7 are the results of iterations. The base, starting point was chosen according to the optimum result of [16], then small steps were taken to find out a better skewness and kurtosis value. According to [31, 116], Kurtosis value of normal distribution should be 3. Excel is used to calculate the iteration results and since Excel is built for business purposes, they use 0 as the Kurtosis value point for the normal distribution. Thus, the value of the Kurtosis shown in Table 7 is close to zero instead of three.

Table 7. Optimum Box – Cox Parameter Iteration Results

λ	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1	0,1011	0,102
Skewness	-0,003	0,083	0,168	0,254	0,341	0,428	0,517	0,606	0,697	0,790	-0,002	-0,001
Kurtosis	-0,292	-0,273	-0,231	-0,166	-0,077	0,037	0,177	0,345	0,541	0,769	-0,292	-0,292

Differently from [16], the value of the λ for the relevant data set is calculated as 0,102 instead of 0,2. The difference happened because at this dissertation the average wind speed of 3 locations (NY 1, 2, 3 – New York Zone 1,2,3) were taken as the raw data rather than one single location from New York (NY 1). This new λ value has less skewness value which is close to zero which is normal distribution's skewness value [31, 116].

According to [31], to get a proper histogram from the raw data, the number of class intervals and class widths should be calculated. First, number of classes should be calculated with Sturge's formula and then the class width should be determined.

Sturge's number of classes formula: 1+3,3log10(n) **(16)**

Where n represents the total amount of data.

Class Width: $\frac{\textbf{Largest observation}-\textbf{Smallest Observation}}{\textbf{Number of Classes}}$ **(17)**

For NY 1, 2, 3 average data n is taken as 7300. With 7300 daily average wind speed;

Number of classes = $1+3{,}3\log_{10}(7300) = 13{,}7489$ ➔ will be taken as 14.

Class Width (Bin Size at Histogram): $\frac{3{,}05-0{,}6}{14} = 0{,}175611324$

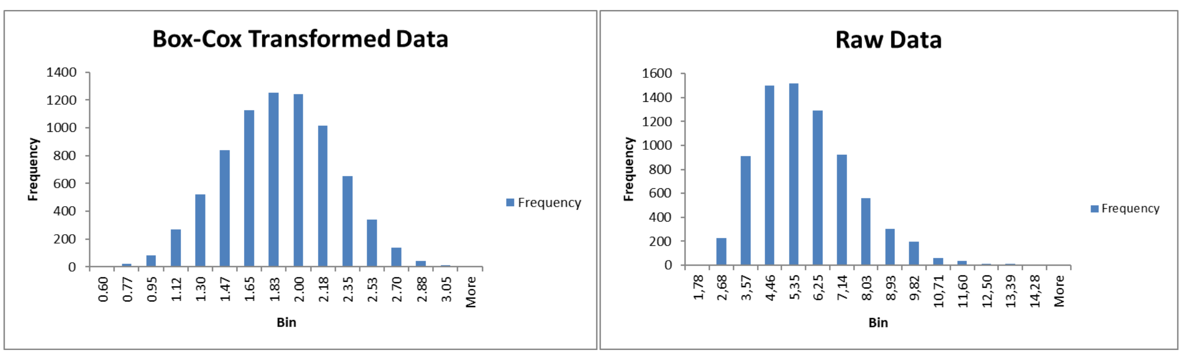


Figure 4. Histogram Demonstration of Box-Cox Transformed Data (Left) and Raw Data (Right)

It can be seen from Figure 4 and Table 7 that using 0,102 as the Box – Cox parameter, λ, reduced the skewness to the -0,001. (Normal distribution's skewness value is 0.) Bin frequency tables with graphs are demonstrated in Appendix 2. Since it is the first histogram of this dissertation, these number of classes and class width calculations were performed. It is possible to draw these histograms with MATLAB's histogram function too.

3.2. Seasonality Effect analysis

According to [43], “Seasonality is a characteristic of a time series in which the data experiences regular and predictable changes that recur every calendar year.” From the definition, it can be understood that fluctuations within a period represent seasonality. This period means within how much time the data shows the same characteristic feature. The period can be monthly, weekly or annual. A good example of seasonality can be understood from the change of ice-cream sales within a year which represents an increase in summer time and a decrease in winter time.

It is necessary to find out the seasonality period before analyzing it. Periodogram approach can be used to detect the seasonality period. The frequencies with the highest powers should give the periods. It can be called as also “power of frequency”.

To detect the seasonality period with Periodogram, R studio and R language were used as software and computing language respectively.

Before starting the explanation, it would be better to mention that the written code and the periodogram can be found in Appendix 3.

At the periodogram which shows the spectral density of the Box-Cox transformed wind speed data, it can be seen that the highest-powered frequencies are close to zero value. The results are shown in Table 8.

Table 8. Two Highest Power Frequencies of the Wind Speed Data

	Freq	Spec	Seasonality Period	
1st	0,002734	148,2791	365,71429	Annual
2nd	0,005469	11,71808	182,85714	Semi-Annual

It is known that it is possible to compute periods from frequency. For R language it is easy to write down the code shown below;

```
# convert the frequency to time periods
> period=1/frequency$f
> period
```

The results of the seasonality period are shown in Table 8. From the results it can be said that the average New York Box-Cox transformed wind speed data has annual and semi-annual seasonality [37].

The seasonality period shows how often the seasonal impacts are shown [16]. After detecting the period, these effects can be modeled with Fourier transform;

$$S(t)=a_0+a_1\cos(2\pi t/365)+a_2\sin(2\pi t/365)+ +a_3\cos(4\pi t/365)+a_4\cos(4\pi t/365) \quad (18)$$

Thanks to this Fourier transform, parameter estimation reduced significantly. It can be seen as a simplified way to evaluate. In addition to this benefit, Fourier transform is able to provide a smooth fitted pattern to the data [66]. At formula 18, a_i demonstrates the Fourier coefficients and t represents the day number in the data.

Trend analysis of the data is going to be explained during the further steps. However, a brief description is needed at the beginning. If the studies were done with temperature data, it can be said that an increasing trend would be expected because of some reasons such as global warming. Since wind speed data has been analyzed, it is needed to think about the reasons which might cause a trend. The most important reason which can cause a trend is the roughness change of the area. Since the data is taken from [30], the exact location of the measurements is not known. However, it is assumed that the measurements were done where it is possible to establish a wind speed farm where that a big roughness change cannot occur due to some new buildings. In summary, before analyzing the data, it can be said that a trend is not going to be expected [1].

The main harmonic at this seasonal model is ($2\pi/365$) which gives the annual part of the seasonality. For the semi-annual part, a combination of cos and sin function with the 2nd harmonic which equals ($4\pi t/365$) is added. In addition to these, t represents the number of the day from the data.

As next step, coefficients of the seasonality model are needed to be estimated with the help of MATLAB. For estimating the parameters 'Fourier**X**' function of MATLAB is used. The value of the X can be between 1 and 8. It shows the number of harmonics which will be used at the model [38].

Example:

General model Fourier8: $f(x) = a_0 + a_1*\cos(x*w) + b_1*\sin(x*w) + a_2*\cos(2*x*w) + b_2*\sin(2*x*w) + a_3*\cos(3*x*w) + b_3*\sin(3*x*w) + a_4*\cos(4*x*w) + b_4*\sin(4*x*w) + a_5*\cos(5*x*w) + b_5*\sin(5*x*w) + a_6*\cos(6*x*w) + b_6*\sin(6*x*w) + a_7*\cos(7*x*w) + b_7*\sin(7*x*w) + a_8*\cos(8*x*w) + b_8*\sin(8*x*w)$

Coefficients (with 95% confidence bounds).

Since the data has annual and semi-annual seasonality, the model should stop at the second step which means that during estimating the coefficients it is needed to use 'Fourier2'.

General model 'Fourier2': $f(x) = a_0 + a_1*\cos(x*w) + b_1*\sin(x*w) + a_2*\cos(2*x*w) + b_2*\sin(2*x*w)$

Estimated coefficient results are shown below with its graph;
Number of days ➔ 7.300 days (20 years)
Windspeeddata = wind speed symmetrized data(after Box-Cox transformation)
fouriermodel=fit(Numberofdays, Windspeeddata,'fourier2')
fouriermodel = General model Fourier2:
fouriermodel(x) = $a_0 + a_1*\cos(x*w) + b_1*\sin(x*w) + a_2*\cos(2*x*w) + b_2*\sin(2*x*w)$

Coefficients (with 95% confidence bounds):

a_0 = 1,776 (1,768, 1,784)
a_1 = 0,1898 (0,1774, 0,2022)
b_1 = 0,04138 (0,02034, 0,06242)
a_2 = -0,04005 (-0,05395, -0,02616)
b_2 = -0,03889 (-0,05273, -0,02504)
w = 0,01716 (0,01714, 0,01719)
>> plot(model)

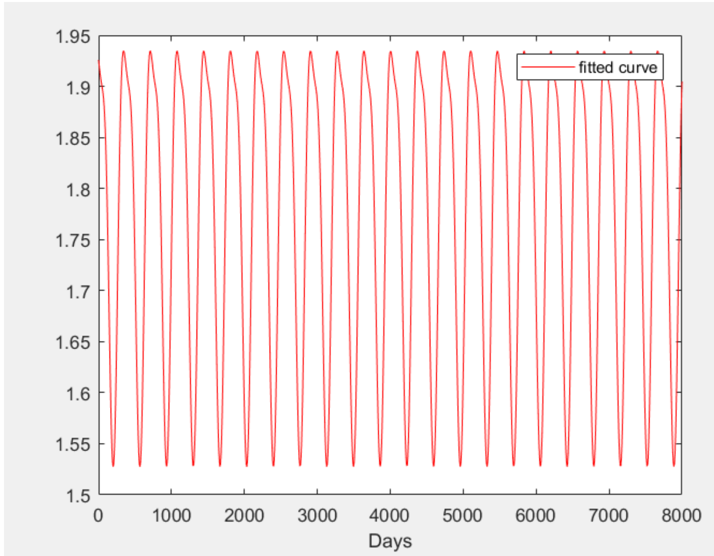


Figure 5. Seasonality Function Values According to Days

Figure 5 shows the change of the seasonality, S(t) in respect to days with annual and semi-annual seasonality. The coefficients which are shown above are used during plotting the graph. The average value of the box-cox transformed wind speed is a_0 = 1,776 which is the initial coefficient.

In addition to this, Figure 6 demonstrates the Box-Cox transformed daily average wind speed data. To make it more visible the graph shows only 4 years of the data. Once again MATLAB was used to demonstrate the graph. In addition to MATLAB, another software named Minitab could also be used to draw the graph. The graph of the full data generated by MATLAB and Minitab can be found in Appendix 4.

Seasonality impact can easily be seen from Figure 6. During the 20 years, the values of the wind speed data fluctuate over the years. According to the first visual analysis, it can be said that there is no going up or downtrend at the data as expected.

To continue with the further steps, firstly it is needed to remove seasonality from the data. After decomposition of the seasonality, the regression model should be specified.

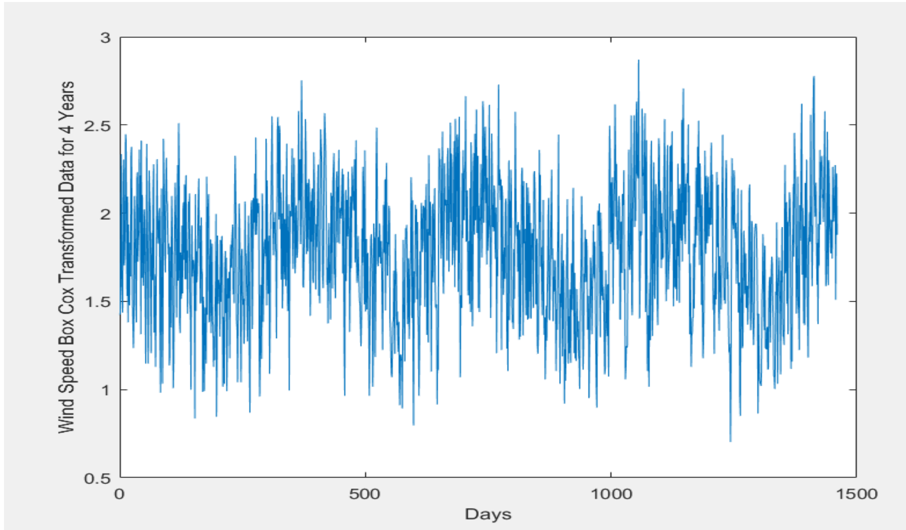


Figure 6. Box-Cox Transformed Daily Average Wind Speed Data for 20 Years

3.2.1. Trend Analysis

At the time series analysis, the trend represents the change of data during a long period of time. For good trend analysis, it is needed to have approximately at least 15-18 years. The change shows if the data is increasing or decreasing over the related period. It is important to detect the trend to set the regression model for the data. It will also help to understand if the data is stationary or not which will be important while choosing the regression model type. [41, 42].

As mentioned above, according to the first visual analysis it can be said that there is no going up or downtrend at the data. Minitab software is used to demonstrate the linear trend model and graph. To support this claim, moving average values are calculated and demonstrated with MATLAB too. The moving average values are following a nearly constant trend.

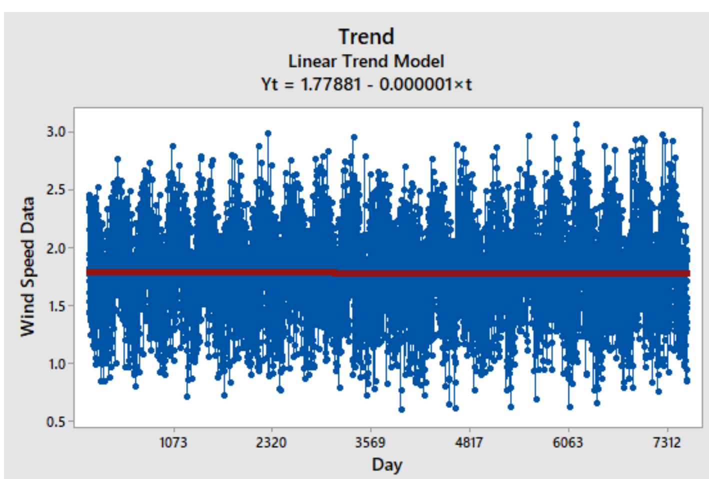


Figure 7. Trend Model of the Wind Speed Data

Since the main harmonic of the data is annual based, 365 days are used to calculate the moving average. The value of a day is calculated according to the previous 365 days. With this methodology, first-year data is lost. However, to prevent data loss, it is assumed that first-year data is equal to second-year data.

Figure 7 demonstrates that the wind speed data does not have a trend during the last 20 years. The trend formula which is calculated via Minitab, demonstrates that the change is on the 5th digit after comma. Thus, this part will be simplified, and it is assumed that there is no trend in data. In addition to that, Figure 8 shows moving term average of the data where the average value of the wind speed data stays approximately stable.

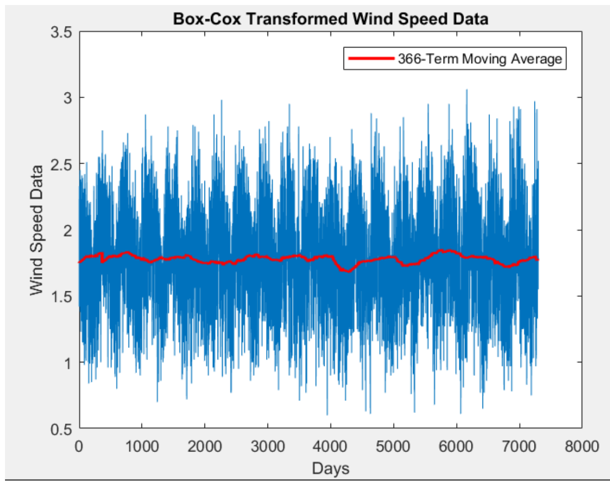


Figure 8. Demonstration of Moving Average of Wind Speed

3.2.2. Decomposition of Seasonality

It is necessary to remove seasonality from the data before creating a regression model. The regression model order is estimated according to partial autocorrelation analysis.

Time series can be in form with additive and multiplicative form in aspect of seasonality analysis. According to [46], if the fluctuations around the trend do not vary too much, the additive approach to decompose the seasonality from the raw data is more appropriate. Additive and multiplicative models can be seen below [40];

$$y_t=S_t+T_t+R_t \qquad (19)$$

$$y_t=S_t*T_t*R_t \qquad (20)$$

where y_t, S_t, T_t, R_t represents data, seasonal component, trend component, and residuals respectively.

For the relevant data, the additive form is chosen while decomposing the data. Since as it is demonstrated before, the data does not have an increasing or decreasing trend component and the fluctuations around the trend do not increase over the time of period. It can be said that fluctuations remain same during the time. If the seasonality can be removed from the data, it will be possible to set the related autoregressive model. (The relevant MATLAB code can be found in Appendix 5.)

There is a second approach to decompose seasonality from the data, seasonal filters. According to trend analysis, it is found that the trend is approximately stable. Thanks to this analysis, a stable seasonal filter could be used while decomposing too [39, 44].

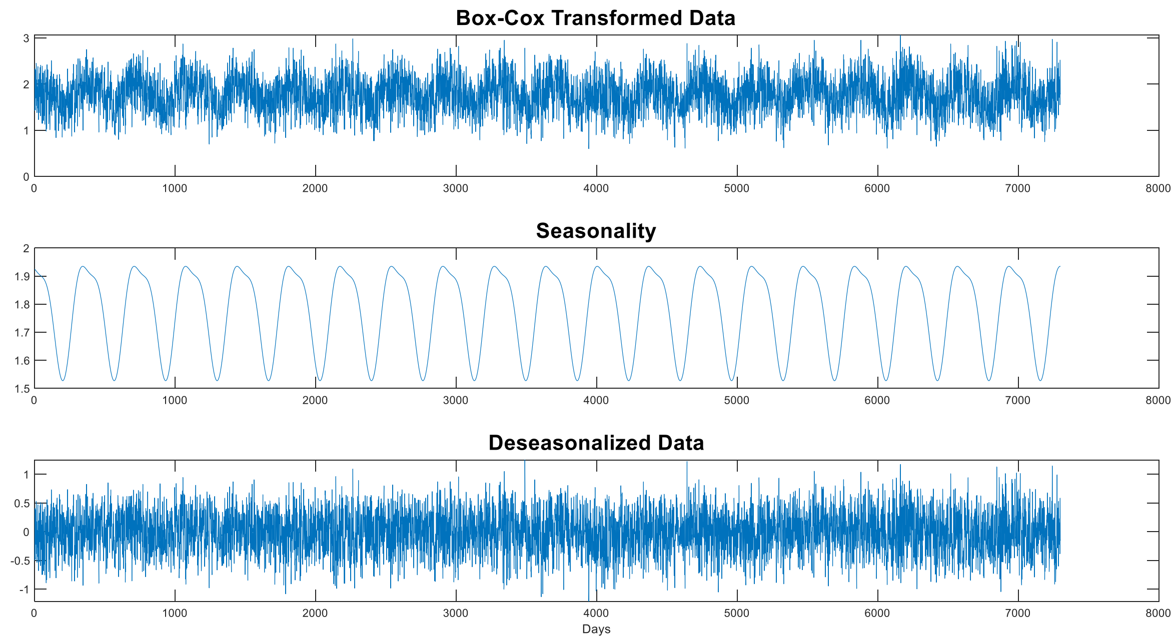


Figure 9. Comparison of Seasonal and Deseasonalized Data

Figure 9 demonstrates the comparison between raw data and deseasonalized data. It can be easily seen from Figure 8 that the seasonal fluctuations are decreased thanks to the decomposition.

3.2.3. Regression Model of the NY Data

During time series analysis, autocorrelation and partial autocorrelation are used frequently to find out the relation between data and its historical data.

The correlation coefficient summarizes and gives information about the strength of the relationship between two variables. There are two types of variables which are used during regression analysis, dependent and independent variables (predictors). The strength of the relationship between these variables is calculated with correlation coefficients. If the predictor variables are historical data of the dependent variable, autocorrelation is used to describe the relation with lagged data-predictors (y_t-s) and the dependent data yt. Partial autocorrelation also known as conditional correlation, is a quite similar approach to autocorrelation concept except for a difference with one aspect. For PACF, (Partial Autocorrelation Function) the condition is that the effects of the values between the dependent variable and the lagged variables are controlled. In other words, PACF analyses the relationship between the specific predictor and dependent variable with removing the impact of other predictor variables.

PACF is useful to find out the number of lags which are effective and should be used during the AR (Autoregressive) regression modeling (deciding the AR order) [41, 45].

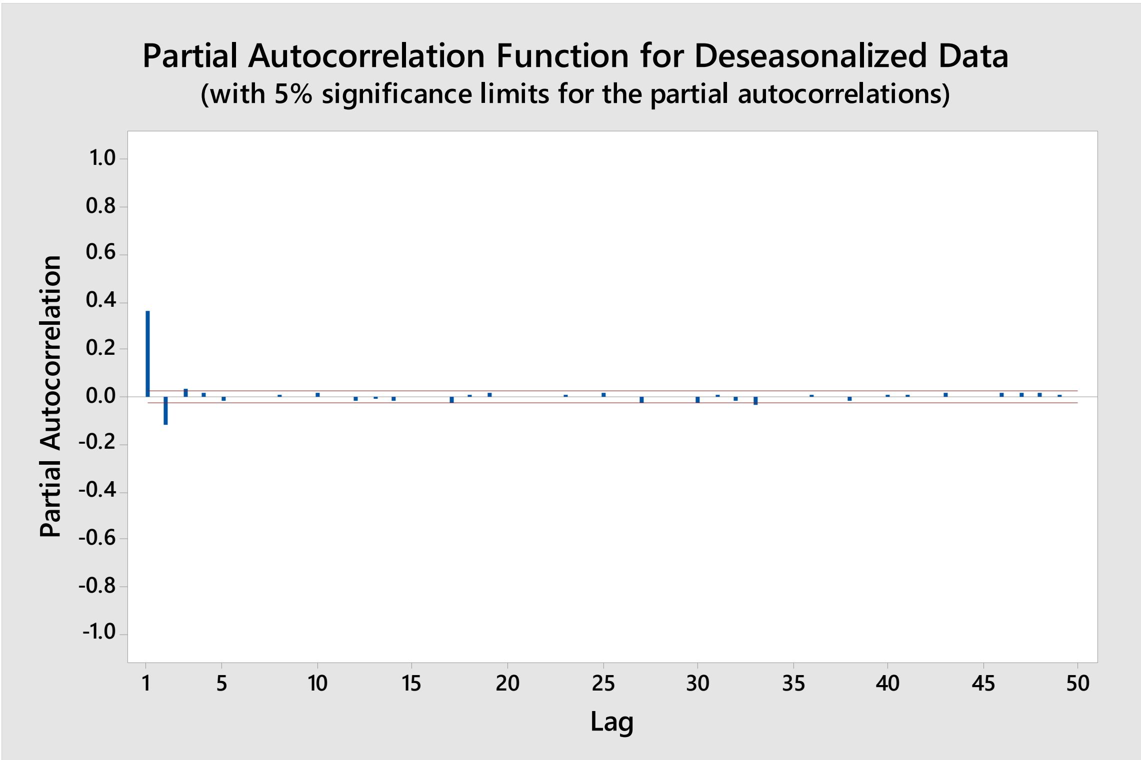


Figure 10. Partial Autocorrelation Function for Deseasonalized Wind Speed – 50 Lags

According to the data's PACF, Figure 10 demonstrates that the 3rd lag is the last lag which is effective to use at the regression model. After deciding the lag order, it should be analyzed that which type of regression model should be used. The analysis and comparison are going to be done between MA (Moving Average), AR and ARMA (Autoregressive Moving Average) models. The comparison between these model types is going to be done according to Akaike's Information Criteria methodology. To make it visible 50 lags are used for Figure 10. Since the period is 365 days, the PACF graph with 365 lags is demonstrated in Appendix 6.

The model's short description will be done in the next chapter. There is one more autoregression model type named ARIMA (Autoregressive Integrated Moving Average) which is used for non-stationary dataset. Since the data has no trend at the seasonality is removed from the data it is possible to say that the data is stationary. However, it is needed to prove that the data is stationary [53, 54].

Unit-root tests are used to check if the data is stationary or not. Two types of the unit-root tests are named as Augmented Dickey-Fuller and KPSS (Kwiatkowski–Phillips–Schmidt–Shin Test) are used to run the stationary test. These tests can be found at different softwares such as Gretl or RStudio. At this dissertation the Augmented Dickey-Fuller test is going to be used to check the stationary. According to ADF (Augmented Dickey-Fuller) test, the H_0 and the alternative hypothesis is shown below [55];

H_0 = Unit root is existing, the data series is not stationary

H_1= The data series is stationary

This test is measured as all other tests. Since the first hypothesis is not a good choice for us, it is needed to be rejected in other words it should not be accepted. For testing and evaluating the result, p-value is going to be used. If the test's p-value is smaller than the critical value, then H_o will be rejected. Gretl and RStudio is used for testing.

For RStudio testing, the number of lags is chosen according to RStudio suggestion. According to [56], "The default value of trunc((length(x)-1) ^ (1/3)) corresponds to the suggested upper bound on the rate at which the number of lags, k, should be made to grow with the sample size for the general ARMA(p,q) setup."

```
> adf.test(d, alternative = c("stationary", "explosive"),
+          k = trunc((length(x)-1)^(1/3)))

        Augmented Dickey-Fuller Test

data:  d
Dickey-Fuller = -18.999, Lag order = 19, p-value = 0.01
alternative hypothesis: stationary

Warning message:
Inadf.test(d, alternative = c("stationary", "explosive"), k = trunc(length(
d) -  :
  p-value smaller than printed p-value
```

Figure 11.RStudio Code and p-value result for ADF(Stationary) Test

As it can be seen from Figure 11, d represents the deseasonalized data, in other words seasonality removed, box-cox transformed data. As the alternative hypothesis, "stationary testing" is added and k represents the lag order which is important to test the hypothesis. This lag order is crucial since wrong choice of the number of lags might cause a correlation between residuals. Therefore, the suggested number of lags by R Studio is used [114].

The warning message is generated automatically by the system. Since the p-value is less than the critical value, H_o is not accepted. Different lag orders are tested and demonstrated in Appendix 7 with Gretl results. While choosing the lag order for ADF test, lag order 3 according to PACF and 365 (seasonality period) are also used. It is possible to find Gretl results of ADF test in Appendix 7.3.

These results demonstrate that data is stationary which means that it is not needed to use ARIMA (Autoregressive Integrated Moving Average) models during the regression analysis.

3.2.3.1. Regression Models and Akaike Comparison

There are many approaches, and methodologies to model a time series data. AR (Autoregressive), MA (Moving Average), ARMA (Autoregressive Moving Average) and ARIMA (Autoregressive Integrated Moving Average) models are some most common approaches for modeling.

AR Models: One of the simplest models of these autoregression related models is the AR model. The model demonstrates the linear regression between the dependent variable and its historical data.

$$X_t = \delta + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + A_t \qquad (21)$$

where X_t represents the current value of the time series data and A_t represents the white noise. δ and φi represent coefficient values. Formula 24 demonstrates the definition of δ. This constant parameter has a value when the mean of the series does not equal to zero [47, 48, 49, 50, 54].

<u>MA Models:</u> Moving average model uses a different approach. Instead of using the historical data of the dependent variable to make a forecast, the historical errors (white noise) are put into a regression model with the current value. The errors mentioned at this model is the White noise variables which can also be seen at formula 21.

$$\mathbf{X_t = \mu+\theta_1A_{t-1}+ \theta_2A_{t-2}+\cdots+ \theta_pA_{t-p} +A_t} \qquad \mathbf{(22)}$$

Where X_t represents the current value of the time series data and A_t represents the white noise values, μ represents the mean value of the series and θ_i represents coefficient values. [47, 48, 49, 50].

<u>ARMA Models:</u> This model is also known as the Box- Jenkins Model. This model is a regression model where it is established by the combination of AR and MA models. This model is used when stationary data is owned and used. If the time series data set is not stationary, it is suggested to make the data stationary with using differencing. This model is called as ARIMA. After differencing the data, ARMA model can be used to model the data [47, 48, 49, 50].

$$\mathbf{X_t= \delta +\varphi_1X_{t-1}+\varphi_2X_{t-2}+\cdots+\varphi p X_t-p+A_t +\theta_1A_{t-1}+ \theta_2A_{t-2}+\cdots+ \theta_pA_{t-p}} \qquad \mathbf{(23)}$$

At these formulas, δ is a constant value as shown below where μ represent the mean value of the process;

$$\boldsymbol{\delta = \left(1 - \sum_{i=0}^{p} \phi i\right) * \mu} \qquad \mathbf{(24)}$$

Since the seasonality is removed from the data it is expected to have the value of μ as zero which means that the value of the $\boldsymbol{\delta}$ will be zero too for the analyzed data set at this dissertation. Therefore, after deciding the regression model, the coefficients of the regressors will be determined via software without a constant value.

3.2.3.2. Akaike and Bayesian Information Criteria Methodology

AIC (Akaike's Information Criteria) and BIC (Bayesian Information Criteria) are two methodologies which are used to determine the best model for the related time series data. These are methods which only compare the variants.

These two criteria have two different goals where AIC and BIC serve goals 1 and 2 respectively [51].

Goal 1: Finding the model which generates the best predicted values ➔ Condition: There is no assumption that at least one of the models is true

Goal 2: Finding the true model ➔ Condition: It is assumed that there is one true model between variants.

One important difference between AIC and BIC is that both methodologies are penalizing the unused parameters, but BIC methodology does that more strongly. For both methodologies lower values are giving better results. These values do not mean anything, but they are used only for ranking the data or variants [52].

Table 9. AIC and BIC Values of Determined Possible Models

Model	Value of AIC	Value of BIC
AR (3)	4810,826	4838,408
AR (5)	4810,655	4852,029
ARMA (3,3)	4812,630	4860,899
ARMA (5,5)	4810,327	4886,179

According to Table 9, it can be easily seen that AR models for deseasonalized data are better than ARMA models. The AIC value of the AR (5) is a bit lower than the AIC value of AR (3). However, according to PACF 3 lags were determined for the order of the model. In addition to that BIC value of the AR (3) demonstrates better value. Thanks to these 2 facts, AR (3) model will be used as the regression model.

*AIC and BIC results are calculated by both MATLAB and Gretl software. The results of the Gretl can be found in Appendix 8.

3.2.3.3. Regression Model

After removing the seasonality from the original data and deciding the regression model type which is AR (3), the model is ready to be set as shown at formula 25.

$$\mathbf{Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3}} \qquad \textbf{(25)}$$

At formula 25, β represents the regression coefficients and Y_{t-i} represents the historical data.

Since the mean value of the data and the constant value of the seasonality function is equal (1,77 m/s) and also since the seasonality is subtracted from the box-cox transformed data, the constant value (β_0) does not exist at the regression model (It is expected to have it as zero). AR models generated by Minitab and Gretl prove that even though if a constant value is added during the generation of the AR model, the coefficients of the model remain same and the constant value is generated as more or less zero (0,0002). The results can be found in Appendix 9.

The coefficients of the AR model, MATLAB was used as the first choice. The results of the MATLAB are in line with other software tools, in other words, the coefficients obtained from Gretl and Minitab are similar with the MATLAB results.

MATLAB Results

>> armodel3=ar(deseasondata,3) armodel3 = Discrete-time AR model: A(z)*y(t) = e(t)

A(z) = 1 – 0,4112 z^-1 + 0,1314 z^-2 – 0,03475 z^-3

The coefficient results shown above is in format of lag operators. These operators are polynomial and used to define AR models. The opposite signed of the coefficients of the lag polynomial function represent the coefficient of the AR (3) model [57].

Table 10. Parameter Estimation for AR (3) Model

Parameters	β1	β2	β3
Estimations	0,4113	-0,1316	0,0347
p-value	1,30E-270	1,15E-25	0,003

H_0: βi=0, H_1: βi≠0 For:$\alpha = 10\%$

Table 10 shows the estimated values for the coefficients for the AR (3) model. The p-values are used to determine if any of the β coefficients are statically insignificant or not. Since all p-values are less than 0,1, it can be said that these β coefficients are statically significant. H_0 is not accepted, the result might be wrong with 10%.

MATLAB Results with Econometric Modeler

The parameter estimations shown above are done with AR function of MATLAB. MATLAB 2018 has a new application tool. The coefficient estimation and residual analysis graphs can be done via Econometric Modeler too.

There are regression model alternatives AR, MA, ARMA, and ARIMA at this tool. Since it is already decided to use AR (3) model, the relevant alternative is chosen as shown in Figure 12.

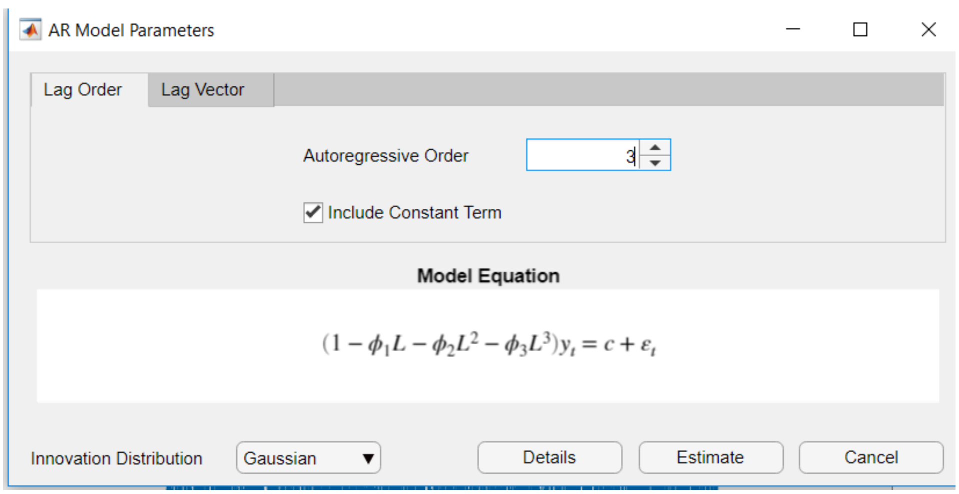
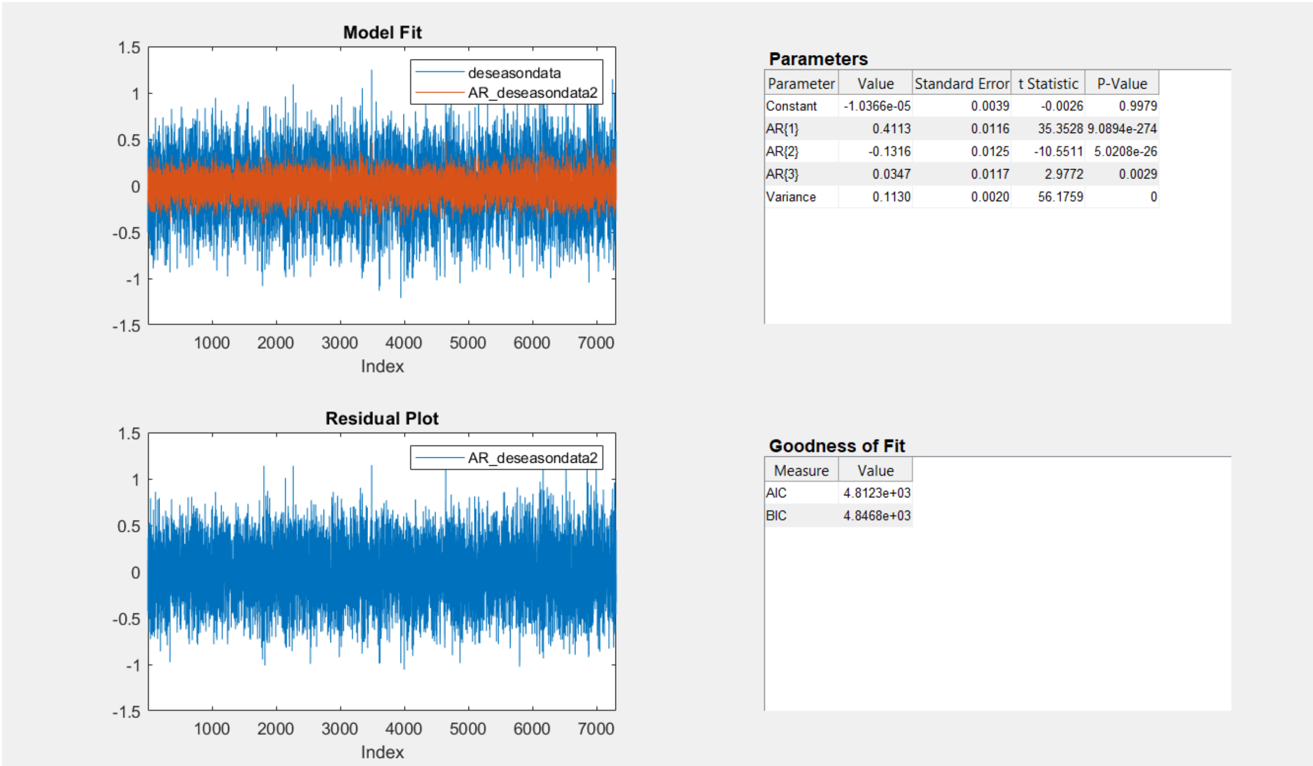


Figure 12. AR Model Order at Econometric Modeler

As it can be seen from Figure 12, this modeler also uses the lag operators, the basic principle behind estimating the coefficients is same. This modeler helps to analyze more user friendly. The estimated coefficients and AIC and BIC values are demonstrated in Figure 13.



Parameter	Value	Standard Error	t Statistic	P-Value
Constant	-1.0366e-05	0.0039	-0.0026	0.9979
AR{1}	0.4113	0.0116	35.3528	9.0894e-274
AR{2}	-0.1316	0.0125	-10.5511	5.0208e-26
AR{3}	0.0347	0.0117	2.9772	0.0029
Variance	0.1130	0.0020	56.1759	0

Measure	Value
AIC	4.8123e+03
BIC	4.8468e+03

Figure 13. Coefficients and AIC&BIC Values of AR (3) – Econometric Modeler

3.2.3.4. Residuals

After deciding and setting AR (3) model, the predicted values can be calculated and set. Predicted values play a significant role while calculating the residuals. Residuals demonstrate the difference between the predicted values according to the regression model and the real values. (Predicted values demonstrate the calculated values from the autoregression model)

$$E_t = R_t - P_t \qquad (26)$$

Where E_t , R_t and P_t demonstrate the residual value, real value and predicted value respectively. There are some features which can be called beneficial and useful for residuals.

1. The mean value of the residuals is expected to be zero
2. There should not be a correlation between the residual values.

These features can also be called as assumptions during setting the autoregression models. Therefore, after calculating the residuals, it is necessary to plot the residuals in different graphical methods and also demonstrate the ACF (Autocorrelation Function) of the residuals [59].

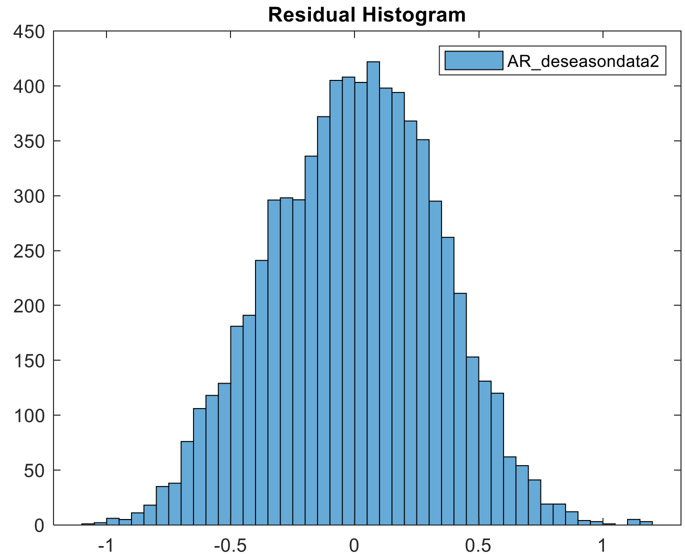


Figure 14. Residual Histogram of AR (3)

Figure 14 demonstrates the histogram of residuals coming from the autoregression model. The shape of the diagram looks like a normal distribution. Nevertheless, the values of the skewness and the kurtosis should be calculated to see if the values are close to 0 and 3 respectively [31, 116].

Table 11 demonstrates the results. The average value is really close to zero where the same can be said for the skewness (low level of negatively skewed distribution). The kurtosis level is a bit lower than 3 which means that the histogram pilot has a bit platy kurtotic behavior. However, platy kurtotic behavior is not so powerful.

Table 11. Skewness and Kurtosis Values of Residuals

Skewness	-0,05399177
Kurtosis	2,741049796
Average	5,92977E-06

Before evaluating the variance (volatility) of the residuals, general evaluation of the residuals coming from AR (3) model is going to be performed at this part. Figure 17 has four residual plots named as Normal Probability Plot, Versus Fits, Histogram, and Versus Order. These plots were generated by Minitab automatically. Since the evaluation of the histogram was done after Figure 13, it will not be repeated in this section.

The normal probability plot in Figure 15 demonstrates that the residuals are really close to normal distribution. The plot also demonstrates that the residuals' 50th percentile located on point zero at the x-axis. Since the normality probability plot does not have any curve at its shape, it can be said that the skewness does not exist and also the histogram does not have longer or shorter than the normal distribution [58] (Minitab Support demonstrations about the curves and tails can be found in Appendix 10).

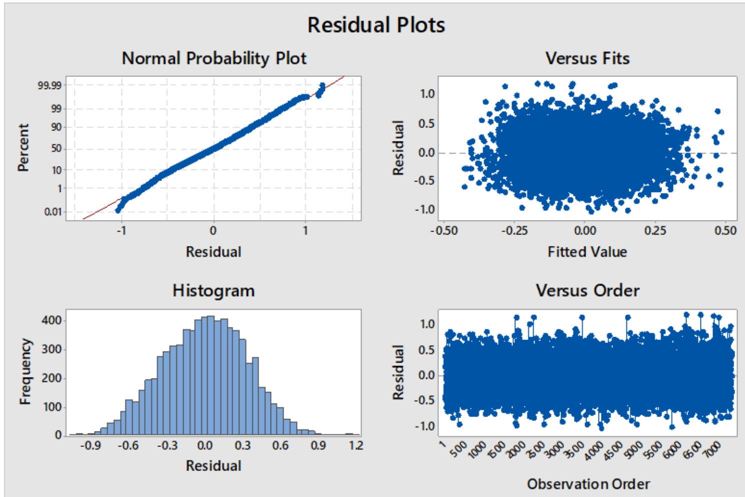


Figure 15. Residual Plots for the Wind Speed Data Before Removing Seasonality

Residual versus fits can also help to see if the residuals are normally distributed and have constant variances. The graph does not have any determined shape or direction, it is randomly scattered around zero. It is possible to check if the variance of the residuals are the same over the time. If the wideness of the y-axis changes over the x-axis, it means that the variance does not remain same. The further evaluation will be done at 3.2.3.4.1 [58,64].

The last graph named as Residuals vs Order demonstrates the relationship between each other, in other words, it is possible to check if the residuals are independent. As it can be seen from the graph that trend or shift does not exists which means that the residuals are not correlated [58].

After evaluating these graphs, it can be said that, the two general rules listed at the beginning of the chapter is satisfied. The kurtosis value of the data is not exactly same with the normal distribution's value. However, since it has a close value it is assumed that residuals are normally distributed.

3.2.3.4.1. Residuals' Variance Analysis

Even though the results demonstrated in Figure 14 imply that the residuals are close to the normal distribution (and it is assumed that the data is distributed with normal distribution) which is also supported also with values of skewness, kurtosis, and average, the residuals coming from the raw wind speed data are not iid (independent identically distributed). Since iid means that all variables have same mean value and variance and since a constant variance does not exist for each day of residuals, the volatility of the variance is needed to be modeled. This modeling could play a significant role according to prediction methodology. A further description is given at the end of this chapter.

For the residuals coming from the linear regression models, the volatility of the variance means that heteroscedasticity is existing. Heteroscedasticity checks the variance of the residuals for different x values to find out that if it is stable or not. If the variance is stable for different values of the x, it can be said that the homoscedasticity for the residuals is existing. Figure 16 is an example of heteroscedasticity existence. As it can be seen from Figure 16 [60] the variance is changing over for different x values at a regression model and Figure 16 is demonstrated as an example of Heteroscedasticity.

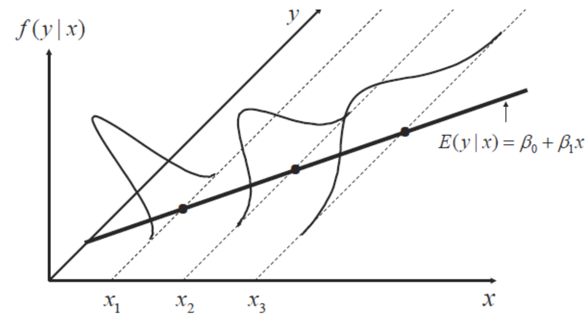


Figure 16. Heteroscedasticity Example

Since the analysis is performed with a time-series data set, the approach for analyzing the residuals' variance is going to be a bit different. Residual errors coming from time series analysis might be modeled. This model can also provide extra information to make the predictions better. Previous studies about modeling weather-related data mentioned that residuals coming from the AR or ARMA models have seasonal dependent changes [16,61,65,66]. The seasonal volatility of the data can be figured out from the ACF of the squared residuals. In addition to detecting seasonal volatility of the residuals, ACF of the squared residuals helps to determine if any ARCH (Autoregressive Conditionally Heteroskedastic Models) or GARCH (Generalized Autoregressive Conditionally Heteroskedastic Models) model is needed to be applied for the variance of the residuals. These models generally help to model the change in the variances [67].

Checking the necessity of adding ARCH or GARCH model into the variance model can be decided from the ACF plot of residuals and squared residuals. Since the regression model is an AR model, it is needed to check if ARCH model is needed or not (If ARMA model existed, it would be needed to check if it is needed to model GARCH effect).

ACF function of the residuals does not demonstrate any significant autocorrelation at any lag as it can be seen from Figure 17. PACF of the residuals can be found in Appendix 11. The PACF plot of the residuals do not have any significance at any lag as ACF plot's behavior. Since there is not any significant lag, it can be said that the residuals are uncorrelated which satisfies one of the 2 rules described in section 3.2.3.4.

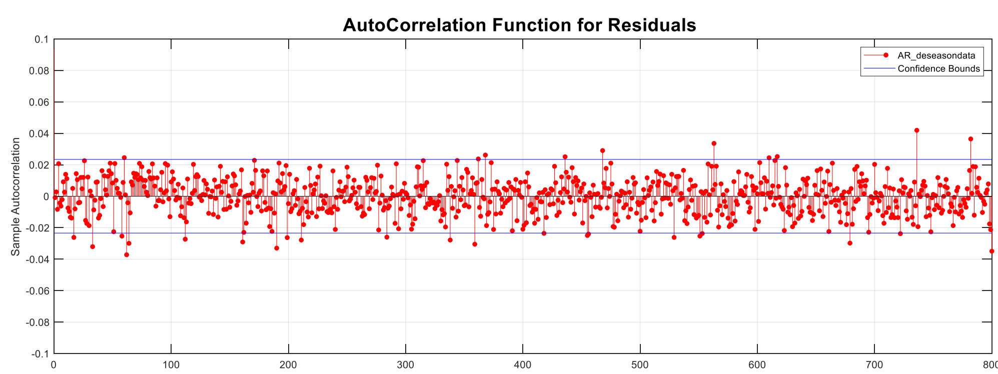


Figure 17. ACF for Residuals

Figure 18 demonstrates the ACF plot for the squared residuals. As it can be seen from the plot that small ARCH effect can be seen but since the significance of autocorrelation at different lags is not so high ARCH effect will not be modeled and put into the calculations.

Since it is assumed that the ARCH part does not exist at the residual parameter calculations, it is possible to write the residuals' model as shown below [16, 65, 67];

$$x(t) = w(t) * \sigma(t) \qquad (27)$$

where x(t), w(t) and σ(t) represent, residuals, white noise, and seasonal volatility respectively.

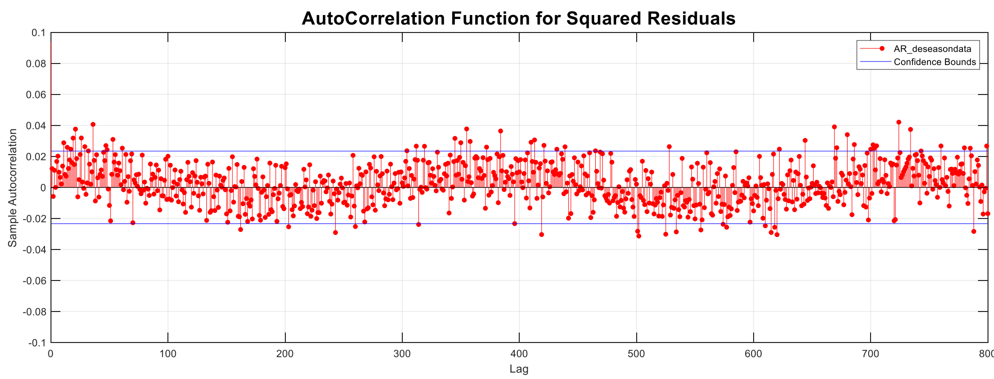


Figure 18. ACF for Squared Residuals

After calculating and modeling the seasonal volatility (variance) of the residuals, σ(t), the calculated residuals will be divided with σ(t) which is called also as the process of standardizing the residuals. Standard residuals are following normal distribution N (0, 1) [68, 69]. This step will be evaluated and performed after modeling the seasonal volatility of the residuals. After standardizing the residuals with dividing the residuals with seasonal volatility, random numbers will be generated with N (0, 1) distribution to reach the white noise part of the residuals and then seasonal volatility of each specific day will be added into the generated random number.

Modeling variance instead of standard deviation is one of the most convenient methods during calculation the volatility. Since variance eliminates the negative value, it prevents negative values to be canceled by the positive values. Modeling and working the variance instead of standard deviation is a common perspective for the previous weather and stock market-related studies and analysis as well [16, 61,65,66].

Since the source of the data is similar with [16], it is normal to expect to have a similar volatility characteristic in variance over the years. As it can be seen from Figure 19 during the beginning of March (approximately days between 50 and 70) and the last part of the autumn have higher variances. In summary the seasonal fluctuation is getting increase during the winter period and decreased during the summer period.

Once again as it can be seen from Figure 19, it follows trigonometric functions, sine, and cosine. Therefore, during the modeling of the seasonal volatility of the residuals, it is possible to use the Fourier transform once more as it was performed during the first seasonality analysis of the box-cox transformed wind speed data. The Fourier model which is going to be used, can be both a combination of sinus, and cosine or cosine only model.

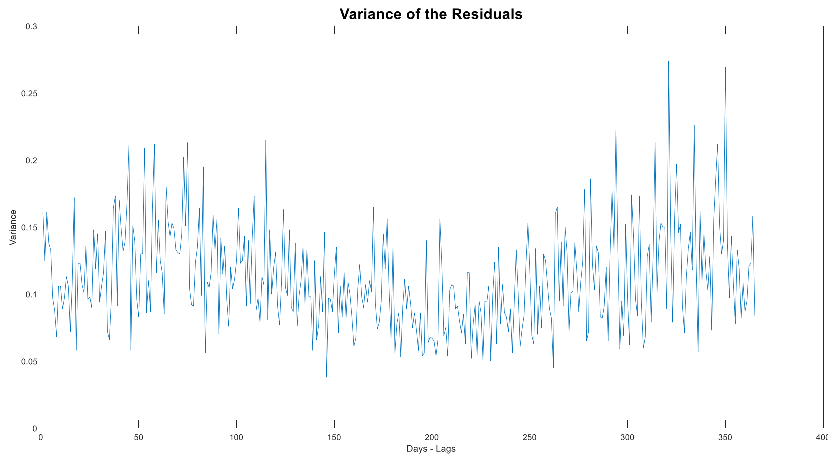


Figure 19. Variance of the Residuals

During the calculation of the variance for the residuals, the sample statistic approach should be followed. The pure raw data at the beginning was collected from 20 years which means that the variance value of each day can be calculated from 20 data. (20 values are existing for each day). Since the data set consists of 20 values, the population variance cannot be used. Due to this fact it is necessary to use the sample variance calculation approach. The formulas are shown below [62].

$$\sigma^2 = \sum_{i=1}^{N} (x_i - \mu)^2 / N \quad \textbf{(28)}$$

$$s^2 = \sum_{i=1}^{N} (x_i - \bar{x})^2 / (n-1) = \frac{n}{n-1} \; (\bar{x}^2 - (\bar{x})^2) \quad \textbf{(29)}$$

With this approach, it is important to understand that, the variance value will be the same for the same day of every year. The explanation can be found at formula 30. The periodicity of the variance is a year and the values variances of the 1st of January remain same for each year. (Variance of 01.01.1987 = 01.01.1988 = 01.01.2007) σ, σ^2, x, x̄ .and μ represent standard deviation, variance, data, sample mean, and population mean respectively.

$$\sigma^2 \, (t+365) = \sigma^2 \, (t) \quad \textbf{(30)}$$

where σ^2 (t+365) and σ^2 (t) demonstrates the variance of the residuals.

It is better to mention that during modeling, the seasonality of the box-cox transformed data, periodogram approach was used to understand and calculate the relevant periodicity to determine the amount of the coefficients and frequency at Fourier transform.

The situation is a bit different for the variance modeling. Since the periodogram cannot be drawn because of the fact shown in formula 30 (periodicity calculation cannot be performed- 1-year variance is existing), a different approach will be followed. During fitting the variance model, eyeball estimation will be followed. The estimation will be supported with the goodness of test results too.

3.2.3.4.2. Variance Model with Combination of Sin and Cos – Fourier Transform

For modeling the variance with Fourier Transform consisting of a combination of sin and cos functions, the same code and methodology will be used with seasonality modeling of the box-cox transformed data.

After trying to fit the many different levels of Fourier transforms starting from Fourier 2 and ending at Fourier 8, it is decided to use Fourier 4 since it fits the data most closely. The fitting is checked with eyeballing, checking the importance of coefficients and R square values. The R square values have a significant increase while changing from Fourier 3 to 4. After Fourier series with 4 terms, even though the numbers of terms were increased, the R squared values did not increase significantly (less than 1%). Thus, Fourier 4 is used to model the seasonal volatility at the residuals since simpler models are easier to analyze and evaluate. Once again, the decision on the number of harmonics (steps of the Fourier) was not calculated as it was done before. Since the periodicity is constructed as it was shown at formula 30, the periodogram approach was not followed.

General model Fourier 4: $f(x) = a_0 + a_1*\cos(x*w) + b_1*\sin(x*w) + a_2*\cos(2*x*w) + b_2*\sin(2*x*w) + a_3*\cos(3*x*w) + b_3*\sin(3*x*w) + a_4*\cos(4*x*w) + b_4*\sin(4*x*w)$

Coefficients (with % confidence bounds).

Figure 20 demonstrates the daily variance function with its fitted Fourier series. The fitted parameter estimation was performed by using 20 years of data. However, Figure 20 represents only 365 days from the whole 20 years.

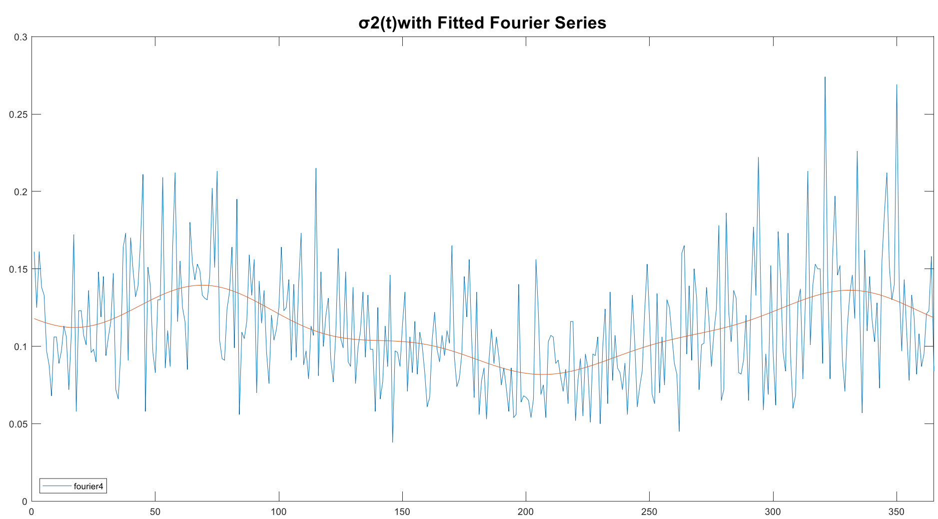


Figure 20. Daily Variance Function and Fitted Fourier Series

The coefficients of the fitted Fourier series demonstrated in Figure 20, can be shown below;

General model Fourier4:

varianceseasonality4(x) = $a_0 + a_1*\cos(x*w) + b_1*\sin(x*w) + a_2*\cos(2*x*w) + b_2*\sin(2*x*w) + a_3*\cos(3*x*w) + b_3*\sin(3*x*w) + a_4*\cos(4*x*w) + b_4*\sin(4*x*w)$

Coefficients (with 95% confidence bounds):

a_0 = 0,1126 (0,1118, 0,1133)

a_1 = 0,01932 (0,01818, 0,02047)

b_1 = 0,006428 (0,004991, 0,007865)

a_2 = -0,007358 (-0,008586, -0,00613)

b_2 = -0,005342 (-0,006639, -0,004046)

a_3 = -0,00512 (-0,006309, -0,003931)

b_3 = -0,003049 (-0,004366, -0,001733)

a_4 = -0,00089 (-0,00249, 0,0007095)

b_4 = -0,00623 (-0,007354, -0,005107)

w = 0,01721 (0,0172, 0,01723)

3.2.3.4.3. Variance Model with Cos Only Fourier Transform

Another approach for modeling the variance is using cos only Fourier series. Since MATLAB does not have direct modeling of this approach, it is needed to set the model **on my own**. For this approach, nlinfit function of the MATLAB is used.

Once again, it is needed to decide the level of the Fourier as in 3.2.3.4.1. At last, it is decided to use level 4 to set the model since it fits better. For setting the model the code shown below is used [63];

$$\sigma^2(t) = C_0 + \sum_{k=1}^{4} \mathrm{COSk}\left(\frac{2k\pi t}{365}\right) \qquad (31)$$

where $\sigma^2(t)$ represents the variance of the residuals and C_0 represents the initial coefficient of the cos only Fourier series.

```
Residualvariance=importdata('varianceofresiduals.txt');
fittedvariance=importdata('fittedfunction.txt');
figure
plot(Residualvariance)
numberofdata=transpose(1:365);
modelfun=@(c,x)(c(1)+c(2)*cos(2*pi*x/365)+c(3)*cos(4*pi*x/365)+c(4)*cos(6*pi
*x/365)+c(5)*cos(8*pi*x/365));
beta0=[1;1;1;1;1];
coefficients=nlinfit(numberofdata,Residualvariance,modelfun,beta0);
```

```
figure
plot(Residualvariance)
hold on
plot(numberofdata,fittedvariance)
hold off
```

The coefficients of the fitted Fourier series demonstrated in Figure 21, are demonstrated in Table 12.

Table 12. Estimated Coefficients for The Variance Function

C0	C1	C2	C3	C4
0,1126	0,0193	-0,0073	-0,0051	-0,0008

The article [16] which is used as the reference for the data, used cos only truncated Fourier series, which can be used for even functions, to model the seasonal volatility for residuals. Therefore, the model estimation trial was performed in this chapter. Even though the model is set properly with nlinfit function from MATLAB, according to the goodness of fit results, the previous model which is a combination of sine and cosine function can explain better. (5% difference is existing.) Thus, general model Fourier 4 will be used which is mentioned at chapter 3.2.3.4.2.

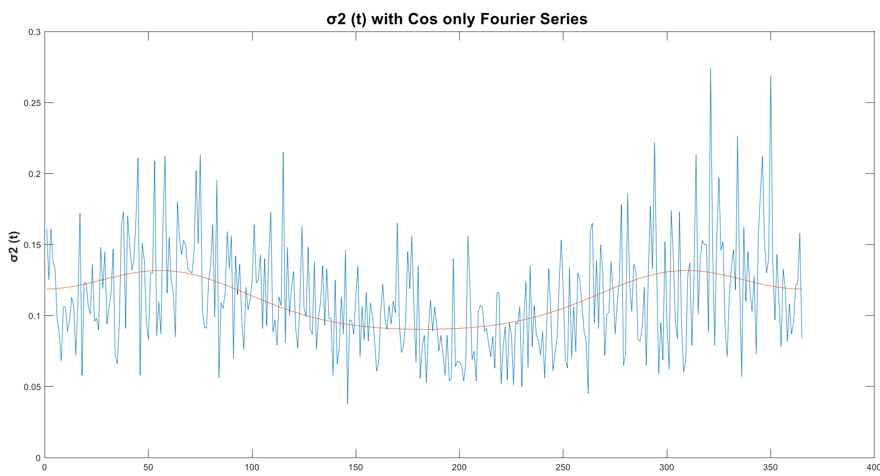


Figure 21. Daily Variance Function and Fitted Fourier Series – Cos Only

3.2.3.4.4. Removing Seasonal Volatility from Residuals

As it is mentioned at the previous section, the Fourier Series with 4th order will be used to model the seasonal volatility at the variances.

General model Fourier4:

$$\text{varianceseasonality4}(x) = a_0 + a_1*\cos(x*w) + b_1*\sin(x*w) + a_2*\cos(2*x*w) + b_2*\sin(2*x*w) + a_3*\cos(3*x*w) + b_3*\sin(3*x*w) + a_4*\cos(4*x*w) + b_4*\sin(4*x*w)$$

The coefficients which were calculated by MATLAB automatically are demonstrated in the previous section. After deciding the model and calculating the coefficients, to reach the standardized residuals or in other words removing seasonal volatility from the residuals should be performed. It is needed to re-call the formula 27 [16, 69];

$$\mathbf{x(t) = w(t)^* \sigma(t)} \qquad \mathbf{(27)}$$

where x(t), w(t) and σ(t) represent, residuals, white noise and seasonal volatility respectively.

To calculate each value for each day is done with the feval function of the MATLAB. After calculating the modelled values for each day, residuals are divided with these modeled seasonal values.

Figure 22 demonstrates that after removing the effect of σ(t), seasonal variation at the variances decreased significantly. In addition to this, even though previous evaluation for the residuals about being close to normal distribution was very satisfying, calculating the standardized residuals helps to generate random number with N (0, 1) distribution [69]. ACF function of the residuals after removing the seasonal volatility can be found at Appendix 12.

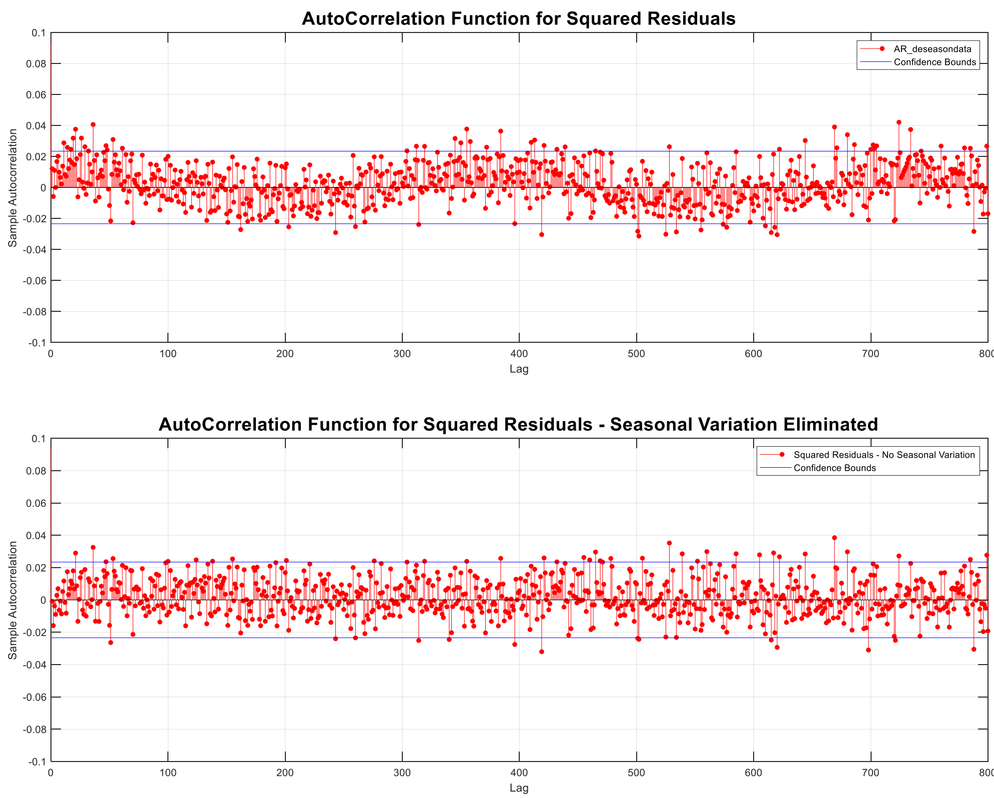


Figure 22. ACF Plot of Squared Res. before (up) and after (bottom) Removing Seasonal Volatility

During MC simulation, random numbers will be generated according to the normal distribution and this seasonal volatility will be added after the generation of the random number. The sum of these two values will be residual of the specific day. During the calculating the dependent variables with AR model, these residuals will be used.

4. Pricing Wind Options for New York with Monte Carlo Simulation Approach

Different methodologies, approaches are existing while using pricing the derivatives such as;

- Black – Scholes Approach
- Arbitrage-Free Pricing Approach
- Monte Carlo Simulation Approach
- Put & Call parity (In case one of them is known)
- Numerical Methods - when probability density function (pdf) is known
- Fang, Osterlee 's Cos (Fourier Method)

The first approach mentioned at this list cannot be used for weather-related derivatives. However, it is one of the most common methods which is used while calculating the price of the derivatives.

At this dissertation, Monte Carlo approach will be used during the calculation the price of the wind power put option. The general name of this approach is actuarial pricing method and Monte Carlo is one of the techniques to determine the price of the contract with this approach. The random number generation will be performed for the residuals and seasonal volatility of the residuals is going to be added into these generated numbers which will be put into the regression. As the last step, the price of the option will be calculated according to difference between predicted and real values [11, 16, 88, 89, 90]. In addition to this approach, in this section, pricing will be performed and calculated with a numerical method as well. The wind speed distribution will be estimated according to real measured data's distribution and the price will be calculated according to the randomly generated numbers which are generated with the real data's distribution. A comparison of these two methodologies will be evaluated too at the end of the chapter.

4.1. Wind Speed Estimation with Monte Carlo Simulation

Since it is the name of the accepted calculation approach at this dissertation, the name of the approach which is Monte Carlo will be used many times. As it is mentioned at the list of abbreviations, MC will be used to describe the Monte Carlo.

4.1.1. Monte Carlo Simulation Approach

MC is a computer-based technique which is used very often by people from different sectors. Companies & decision makers use this approach to determine or predict the potential risk they might face in the future. It can be called as an operational research technique which helps people to determine.

Nowadays, this technique which was used at the World War II (during the design of the atom bomb) first time, is used by the finance, project management, energy, engineering. Insurance, environment or transportation areas and sectors.

The technique is based on a big number of different tries and during each try, this methodology is using different generated random numbers. From this explanation, it is possible to say that, MC approach is not based on mathematical equations but tries. Since it is not a deterministic simulation model but a random model, the distribution of the outcomes can be determined. This is an advantage which shows that MC does not only show the possible result but also its frequency in other words probability [91,92,93].

4.1.2. Wind Speed Estimation with MC

In section 3, the wind speed data of the NY have been analyzed statistically. The steps of the analysis are described below;

- The measurement of the wind speed was performed eight times during a day, since existing and previous market conditions were dealing with the daily average wind speed, average daily wind speed of the 20 years were calculated. (7300 days between 1987 and 2007 & 29th of the Februaries were removed from the data)
- Box-Cox transformation was applied to fit the data into normal distribution to make the analysis easier.
- Trend and seasonality analysis were performed. It is realized that there was no trend. After detecting and modeling the seasonality, it was removed from the data.
- A stationary test is done to determine the type of regression model, and AR model selection is made and performed to find out the residuals.
- Residual analysis is performed, and seasonal volatility of the residuals was detected and removed from the residuals to reach the standardized residuals (N (0,1)) for random number generation.

It is known that MC is based on trials and the total number of the trails are generally in thousands. If total number of the trials is increased, the results would be closer to the expected results theoretically, in other words volatility at the results will be lower. For the MC simulation, MATLAB is used, and the total number of trials is set as 50.000. Total number of the trials is enough to have a result which is close to expected value theoretically and low enough to have the generated data within an acceptable time frame.

The seasonal volatility of the residuals listed as the last item at the above list is really important. This seasonal volatility and the residuals (generated randomly with (N (0,1))) are the main points (starting point) of the MC simulation approach. Since it is calculated that the seasonal volatility of each day is different, it is needed to add these seasonal fluctuations after generating the random numbers to reach the residual part of the regression model. To create an estimated MC wind speed, it is needed to follow the same steps listed above with one difference. It is needed to follow the steps from the latest one.

Firstly, it is needed to design the algorithm with a "for" loop in MATLAB. This loop helps to create a 365 x 50.000 sized matrix. It also helps to use the specific relevant residual for the determined day correctly. The 365 means the sum of the total days within a year. It can be increased or decreased according to the agreements. If the agreement period is shorter than 365 days, the relevant part of the simulated data can be taken and compared. If the agreement period is longer than 365 days, it is needed to increase the matrix size.

Secondly, it is needed to create random numbers for the white noise part of the regression model (please see formula 27). Since the seasonal volatility from the residuals is already removed to reach the standardized residuals which follow the normal distribution, it is needed to use normrnd function of the MATLAB to generate the random numbers. The MATLAB function demonstrates that the random generator will 1 random number from a normal distribution with 0 as the mean value and 1 as the standard deviation. The sigma which will be added to the regression model means the white noise part of the regression model which was represented at formula 27.

```
MATLAB Random Number Generation: sigma=normrnd (0,1);
```

For the next step, it is needed to recall the results shown in Figure 13. It is demonstrated again below in Figure 23 at next page. As it can be seen from the Model Fit and Parameters part, the initial points should be 0 for using the generated random numbers and put it into regression. Since there is no trend at data and since the seasonality is removed from the wind speed data, starting from 0 is an expected approach. One by one the estimation will be done for each day according to AR model shown below. In summary, after generating random numbers and adding the seasonal volatility of the residuals into it, the residual part of the regression model will be reached. Then, this calculated part is going to be added to the main regression model (AR (3)) to calculate the values for each day. The starting point of this regression model is zero since the mean value of the deseasonalized wind speed was zero which is supported by the initial coefficient of AR (3) model which is very close to zero and can be seen from Figure 23.

Wind Speed Estimation with AR Model at MATLAB:

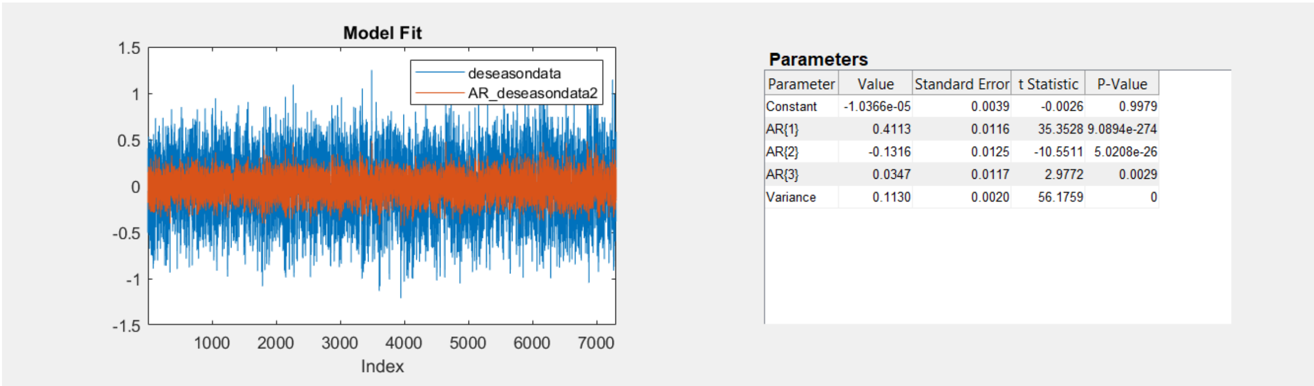
```
mc_noseason(i,j) = regressioncoefficients1 *mc_noseason(i,j) +
regressioncoefficients2*mc_noseason(i+1,j) +
regressioncoefficients3*mc_noseason(i+2,j) +
sigma*residualsvolatility(x);
```

The regression coefficients shown above, are the same coefficients that have been calculated before. They can be also seen again in Figure 23. The created loop which is mentioned as the first step helps to make this calculation for 50.000 times. In summary, firstly random number is generated and then multiplied by with the seasonal volatility and added it to the zero (expected value of the deseasonalized data). At the end, estimated values for deseasonalized data is calculated one by one with using AR model thanks to predecessor estimated data. Mc_noseason is a 365x50.000 sized matrix for the deseasonalized wind speed data.

After completing calculation this step for 365 x 50.000 matrix, it is needed to add seasonality to the data and re-transform it from Box-Cox. Since the seasonality is modeled before, it is just needed to calculate the seasonality values for 1 year and add it to the deseasonalized data.

It is needed to recall the formula 15, to make the transformation from Box-Cox transformed data to the real form of the data. After adding seasonality back to the estimated data, the reached values represent y(λ) at formula 15.

$$y(\lambda)= \frac{(y^{\lambda})-1}{\lambda} \qquad \textbf{(15)}$$



Parameter	Value	Standard Error	t Statistic	P-Value
Constant	-1.0366e-05	0.0039	-0.0026	0.9979
AR{1}	0.4113	0.0116	35.3528	9.0894e-274
AR{2}	-0.1316	0.0125	-10.5511	5.0208e-26
AR{3}	0.0347	0.0117	2.9772	0.0029
Variance	0.1130	0.0020	56.1759	0

Figure 23. Coefficients and AIC&BIC Values of AR (3) – Econometric Modeler

To have the MC data in the original form, formula 32 is used. This is the last step of estimating wind speed data via MATLAB. After this one, it is needed to make a comparison between the real 20- year data and the estimated data to find out if they are similar or not. The parameters used in formula 32 is the same with formula 15.

$$\mathbf{y = (y(\lambda) * \lambda +1)^{(1/\lambda)}} \qquad \textbf{(32)}$$

******Note:** During the residual analysis, seasonal volatility was detected and removed to reach the standardized residuals with N (0,1). After this step random number generation is performed from only N (0,1) distribution and seasonal volatility added to these generated numbers.

It is possible to follow a second approach. According to the second approach, it is not needed to detect the seasonal volatility of the variances. It is needed to calculate variance value for each day. After this calculation, the random number generation for each day can be done by 365 different distribution. The distribution will be again normal but this time the variance of the residuals will have different values for each day. Relevant MATLAB can be found below;

```
Residual value of day x = =normrnd (0, standarddeviation(x),1,1);
```

where standarddeviation(x) represents the standard deviation value of the specific x^{th} day. After generating the random number for each day, this generated value is going to be added to the regression model as shown below;

```
mc_noseason(i+3,j) = regressioncoefficients(3)*mc_noseason(i,j) +
regressioncoefficients(2)*mc_noseason(i+1,j) +
regressioncoefficients(1)*mc_noseason(i+2,j) + sigma;
```

where sigma represents the residual value of day x.

If the evaluation of the pricing is done with this approach, the results will not be so different from the results demonstrated in Table 15. The price difference will be only in ten dollars.

4.1.3. Real and Generated Data Comparison

To compare the real and generated data sets, it is needed to run goodness of fit test. One of the methodologies to make a comparison between two data sets to find out if they have similar distributions is Pearson's Chi-Square test. In addition to Pearson's test, the mean value and variance comparison will be performed too at this chapter.

Before running Pearson's Chi-Square test, the comparison and analysis of the mean value and the variance is going to be performed. This step is added to the dissertation since an assumption will be accepted according to the variance analysis results. The comparison is done between the average value of the real wind speed data and the average value of the MC simulated data. In addition to that, the comparison is done according to the daily and annual values. Since during the pricing, the daily difference will be used, it is meaningful to compare daily values.

The real measured data is coming from 20 years which is in total 7.300 days. On the other hand, there are 18.250.000 simulated data (50.000 data for each day). Even though 20 years data is real measured data and it should represent the population, the amount of data is not sufficient the call this as a population. Therefore, both data set is accepted as samples. Therefore, since it is not known about the standard deviation of the population, it is needed to use the t-test to run hypothesis testing. For the variance testing, it is needed to use F-test [94].

4.1.3.1. Mean Value Hypothesis Testing

H_0: $\mu_1 - \mu_2 = 0$ daily average wind speeds have the same mean value

H_1: $\mu_1 - \mu_2 \neq 0$ daily average wind speeds do not have the same mean value

The test results are evaluated according to p-value. The accepted confidence interval is accepted as 95%. Thus, if the value of the p is smaller than $\alpha/2$, H_o is not accepted, otherwise, H_o will be accepted (The result might be wrong with 5%). The mean comparison for daily results for the t-test is successful for 95% of the 365 days. H_o is not accepted only for 13 days. In addition to this, the comparison for annual data is also performed. The p-value of the annual comparison is 0,766639 which is higher than the $\alpha/2$ again. Thus, it can be said that the average annual wind speeds for both MC and real data have the same mean value (The result might be wrong with 5% probability).

4.1.3.2. Variance Testing

H_0: $\sigma^2{}_1 = \sigma^2{}_2$ daily average wind speeds have the same variance

H_1: $\sigma^2{}_1 \neq \sigma^2{}_2$ daily average wind speeds do not have the same variance

To compare the variances of the data sets, as it is mentioned F -test will be used and once again, the evaluation is performed according to p-value coming from the test. Excel is used to calculate the p-value coming from the F-test. H_o is not accepted only for 22 days which means that the variance of the two data sets is almost the same according to day-based comparison. (The result might be wrong with 5% probability.)

An annual variance comparison is also performed. Even though the daily based variance comparison gave a good p-value result which helps not to reject H_0, p-value of the annual value is less than $\alpha/2$ which forces not to accept H_0 (p-value for annual = 0,015). Not accepting H_o is an expected result. Since the number of data for MC simulated data is much higher than the real measured data, the variance, volatility is going to be less. This result is generally faced at MC simulation approach.

There are generally four probability distributions for a wind speed data which will be mentioned at 4.2. It is expected that both of the real and MC simulated data might follow one of these. The initial data has 7.300 values. On the other hand, the MC simulated data has more than 18 million values. If it is assumed that these 2 data sets are following the same probability density function, it is normal to see that the MC simulated data would have less variance due to the number of data. Therefore, this result is not a big surprise. On the other hand, since a daily comparison will be done and H_0 was not rejected, this MC simulated data while deciding the price level can be used.

4.1.3.3. Pearson's Chi-Square Test

Pearson's Chi-Square test helps to see the goodness of fit for the MC simulated data and real NY data. In other words, it compares the two data sets (generally one of them is coming from a theoretical distribution function) to find out how much they are similar. [95, 96, 97] are used as the reference for this chapter.

To run this test, it is needed to have different separated classes and class intervals. It is like determining the number of classes while drawing the histogram for a data set. There are two different types of the Chi-Square; goodness of fit test and test of independence. In this dissertation's case, it is needed to run the first of them. Therefore, it is needed to calculate the number of observations (frequencies) for each data set and it is needed to arrange the classes with at least 5 observations to make the comparison more precisely.

Even though during the process of calculating the derivative prices according to generated wind power from the wind speed via a wind turbine, a daily basis difference approach is going to be followed at the next chapters, weekly MC simulated data and real data will be used to compare the data sets while running the Chi-Square test. Since the real of the data is coming from only 20 years, the total number of the data would not be enough to arrange the classes for each observation for daily comparison. Therefore, weekly data comparison is performed to increase the number of data while making the comparison. Thanks to weekly comparison, it is possible to use the values of 140 days of real-measured data. (It is needed to have at least 5 observations for each class).

While deciding the number of classes, once again Sturge's formula (formula 16) is used and decided to have 7 classes which shown below in Table 13. Table 13 demonstrates the total number of classes and class intervals. The class intervals are decided the same for all weeks and it is decided according to data coming from each week. The lower limit is also meaningful because of the wind speed limit which will be described at 4.3.

Table 13. Number of Classes and Class Intervals

Class Number	1	2	3	4	5	6	7
Class Interval	<4	>=4 & <5	>=5 & <6	>=6 & <7	>=7 & <8	>=8 & <9	>=9

As it is mentioned above, even though during the pricing, daily data sets are going to be compared, weekly data sets are compared at this chapter to have a general idea whether the weekly data sets are similar at most cases. It is needed to calculate the degrees of freedom which and significance level. The values are;

- Degrees of freedom: df= number of classes -1 = 7-1=6
- Significance level: α = 0,05 (which is the same during the whole dissertation)

After deciding these two parameters, the critical value can be checked form the chi-square tables and the critical value for the determined parameters is approximately 12,6. The value for each week is calculated according to formula 33 [96] which is shown below;

$$\mathbf{X^2 = \frac{(O-E)^2}{E}} \qquad \textbf{(33)}$$

X^2, O, E represent the chi-square value, observed value, and expected value respectively. The Chi-Square distribution graph can be found in Appendix 13 with the Chi-Square values for the weekly data sets.

Normally, observed values and expected values should describe the MC simulated and real data sets respectively. However, since the total number of MC simulated data (even weekly) is much higher than the real data, it is assumed that the situation is vice versa. The comparison of mean value and variance is done daily bases was performed with the comparison of annual results. It was seen that even though the mean value remains the same, the variance of MC simulated data was lower than the real data. Thus, since the number of simulated data is high, the relevant calculated X^2 value will be high as well. If the number of runs, which was set as 50.000, is increased, the relevant X^2 value will increase as well which means that the X^2 will definitely be in the tail. Therefore, it is assumed that the MC simulated data represents the expected value where real data represents the observed value while running this test.

H_0: The observed frequencies' distribution is similar to expected frequencies' distribution

H_1: The observed frequencies' distribution is not similar to expected frequencies' distribution

After this assumption, the Chi-Square values are calculated based on the formula 33. The results can be seen in Appendix 13. It can be said that approximately X^2 value for 73% (38 weeks) of the weekly data sets (34 weeks in total) are not in the tail, and X^2 value for 13% (7 weeks) of the weekly data sets are close the critical value (it might not be in tail if the significance level is changed from 5% to 2,5%) which means that in total it can be said that H_0 might not be rejected for 86,5% of the weekly data set comparison.

As it can be seen both from the histograms of the MC Simulated and Real Data, the shape, and the mean value is similar. It is already mentioned that the variance if the annual data for MC simulated data is less than the real data's. The histograms can be found in Appendix 14. In addition to this, it should not be forgotten that MC simulated data is generated based on the real data and it can be called as a smooth version of the real data. Since the number of the simulations is much higher than the real data, it is normal to have a smoother version of the distribution. It is obvious, and it is also proven here that real data will always have more fluctuation, in other words, the variance that the MC simulated data will be less that the real data.

4.2. Distribution Fitting for Real and Simulated Data

According to different studies, the most popular distribution type for the wind speed data is Weibull distribution. However, this distribution type does not always give the perfect match. There is a possibility to have a different distribution which fits better than the Weibull distribution. Rayleigh, which is a special type of the Weibull distribution, Gamma or Log-normal distributions can fit to wind speed data better than the Weibull distribution. According to the location like onshore or offshore, these types might not fit as well. There are studies to find out the distribution for the specific wind speed of a specific location exists. The importance of finding the distribution might help in two ways; firstly, the distribution parameters for the simulated and real data can be compared and secondly it can be used to calculate the wind power generation's distribution to find out the most efficient wind turbine model for the location [98, 99, 100]. At this dissertation, the first benefit is focused mainly. However, the second aim will be used to calculate the estimated wind power from the generated wind speed data. In summary, the most popular distribution types which generally fit for the wind speed data are;

- Weibull
- Rayleigh
- Gamma
- Log–Normal

Therefore, a comparison between these distributions is going to be made to find out the best fit for the data. The comparison is made via one of the MATLAB Apps tools named as Distribution Fitter. Thanks to these apps it is possible to compare many distributions at the same time and decide which one is better. The comparison for choosing the best within these distributions are made according to their log-likelihood values.

Firstly, the comparison of the distribution is made for the real measured data. It is needed to recall that the data contains the average daily wind speed for 20 years and 7.300 data in total. Figure 24 demonstrates the histogram of the real measured data in addition to the probability density function of the four distribution types. It is obvious that Log–Normal and Gamma distributions fit better than the other two distributions.

In addition to Figure 24, Figure 25 can be used the see the fittest distribution for the real measured data. Figure 25 demonstrates the Quantile-Quantile (Q-Q) which compares the quantiles of each distribution with each other. Once again, Gamma and Log – Normal distributions step forward.

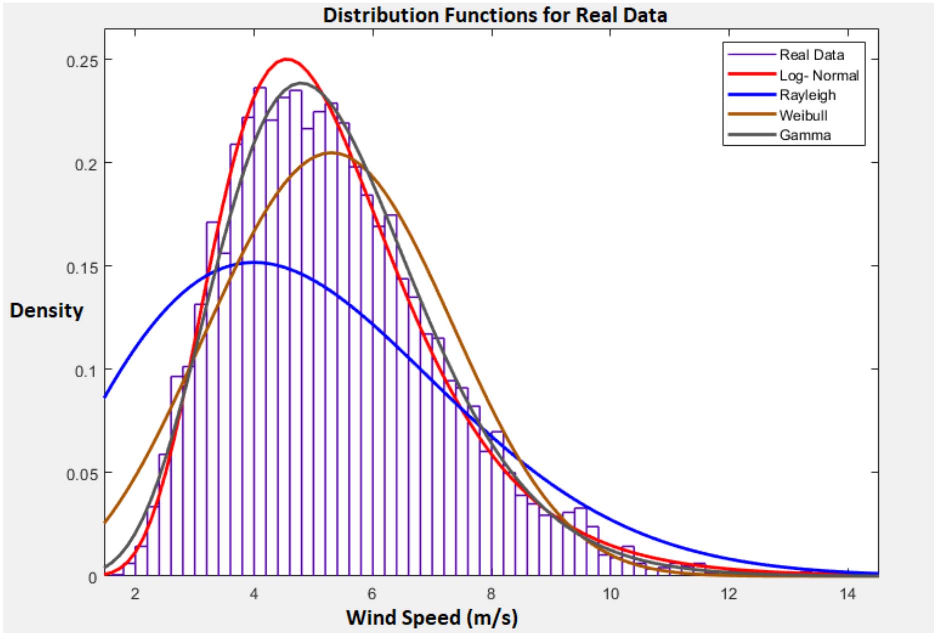


Figure 24. Distribution Functions (pdf) for Real Data

The comments about the best fitting distribution function and parameter comparison will be performed after evaluating the same graphs for MC simulated data.

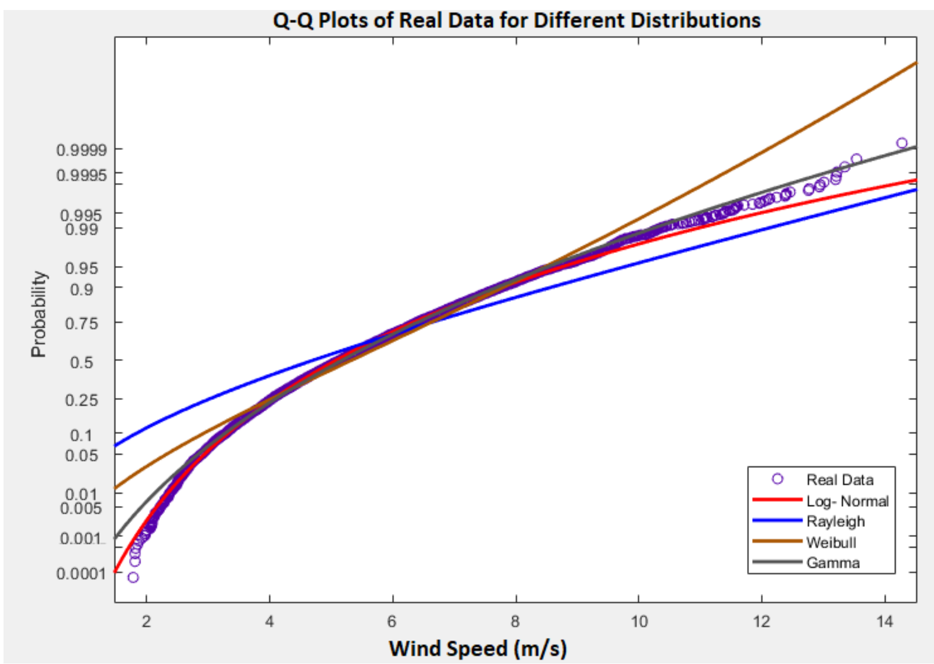


Figure 25. Q-Q Plots of Real Data for Different Distributions

For MC simulated data, firstly the comparison for pdf according to the histogram of MC simulated data is demonstrated. It is important to mention that the histogram is denser than the real measured data due to the number of data. (18 million vs 7.300) Since the MC simulated data is generated according to the N (0,1) and seasonal volatility of the residuals' variances, it is normal to expect to have the same distribution which fits better.

Figure 26 demonstrates that the Log–Normal distribution fits better than any other distributions. It can be said that Gamma can be called as the second-best fit distribution.

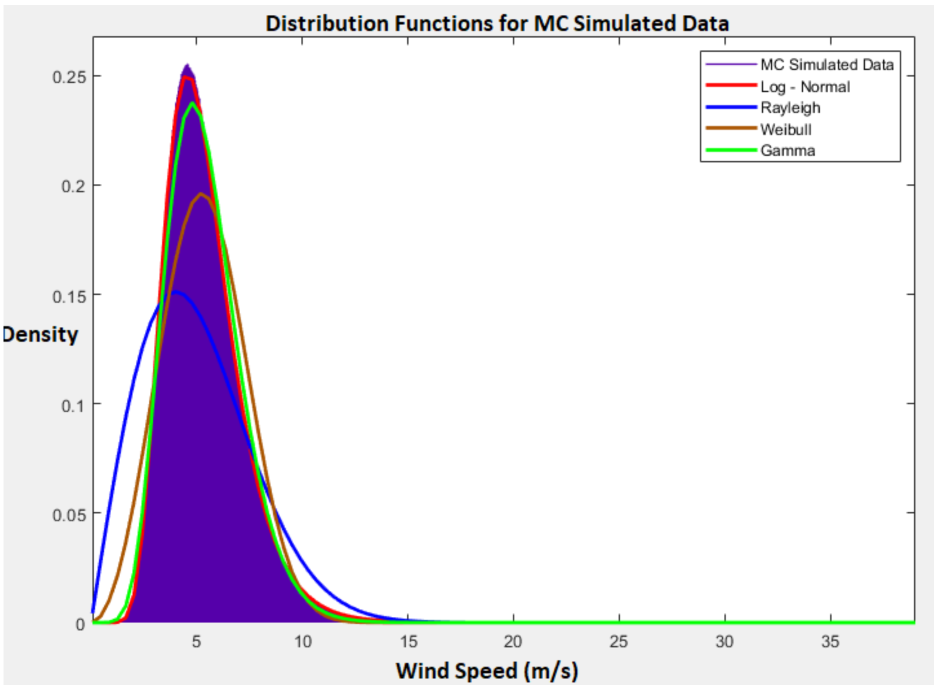


Figure 26. Distribution Functions of MC Simulated Data

In addition to Figure 26, Figure 27 demonstrates the Q – Q plot for MC simulated data. From the Q -Q plot, it can be seen that Log – Normal fits best and the comparison with others would be meaningless. However, the final evaluation will be done according to log-likelihood values as it is mentioned before.

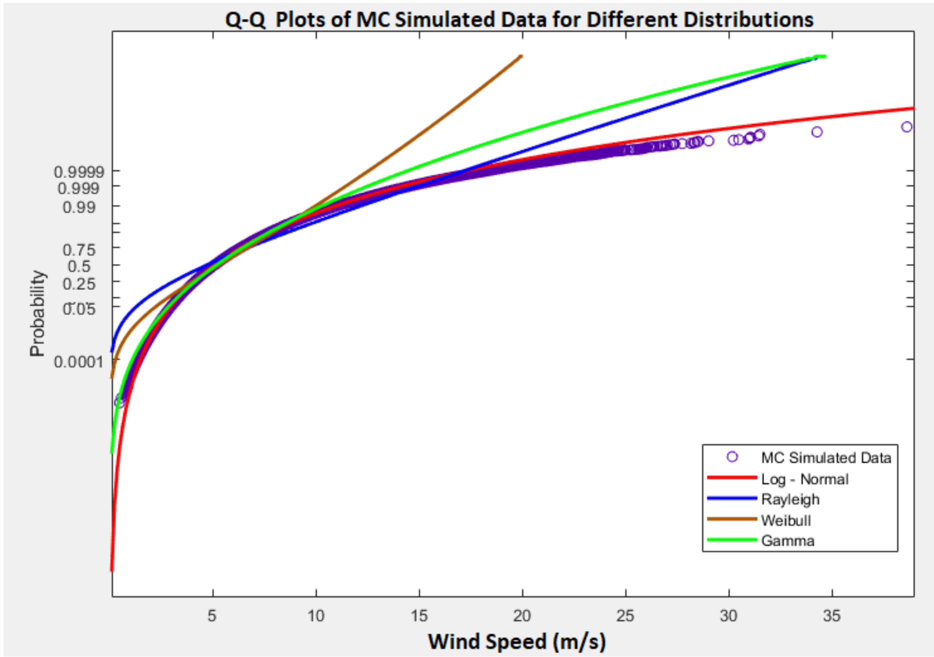


Figure 27. Q-Q Plots of MC Simulated Data for Different Distributions

Thanks to MATLAB distribution fitter, while managing the fits according to the histogram, it is possible to see the log-likelihood values. This log-likelihood value does not mean anything on its own. It is useful to make a comparison between distributions and find the best one within the choices. Since it is wanted to have a high log-likelihood value, higher values mean better fit in the relevant case [101].

According to the log-likelihood values of these four different distributions, the same ordering according to the fitting of the data sets for both real and MC simulated data have been reached. The order for the best fitting of the distributions to the data sets are;

1. Log-Normal
2. Gamma
3. Weibull
4. Rayleigh

These results can be seen before checking the log-likelihood values from the Figures 24,25,26 and 27. There might be only one concern for the real data comparison with Gamma and Log–Normal distributions but the likelihood values clear the doubts. The likelihood values of the distributions can be found in Appendix 15.

After deciding the distribution which fits the best to both data sets, the comparison of the coefficients is going to be made. Since it is decided Log- Normal decision fits better, the comparison of coefficients of this distribution is performed. Table 14 demonstrates the coefficients of Log–Normal distribution according to the two data sets.

Log–Normal distribution has two coefficients which are called as mean and scale parameters [99]. These parameters demonstrate the mean and standard deviation value of the log-transformed version of the original data. If the natural logarithm of both data sets is taken, the new log-transformed data sets will follow a normal distribution with the values of the mean(μ) and standard deviation (σ) which are shown at Table 14.

Table 14. Log – Normal Distribution Parameters

Log Normal Distribution		
Parameters	μ	σ
MC Simulated Data	1,62627	0,330195
Real	1,62535	0,331698
Interval (Real Data)	0,003882	0,002745

Table 14 demonstrates that the mean and standard deviation value for the MC simulated data is really close to the real data's values and they are within the confidence interval. It is clear that the aim of this chapter which is mentioned at the beginning is satisfied.

From these parameters, it is possible to calculate the mean and standard deviation values of the original distribution. The formulas for calculating the mentioned values can be found below [99];

$$\mathbf{m= exp\ (\mu+ \sigma^2/2)} \qquad \mathbf{(34)}$$

$$\mathbf{v=exp\ (2(\mu +\sigma^2))- exp\ (2\ \mu+ \sigma^2)} \qquad \mathbf{(35)}$$

where m and v represent the mean and variance value of the Log-Normal distributions. Generally, the wind speed's standard deviation is within 0-3 m/s [102]. If the standard deviation is calculated for both real and MC simulated data with using formula 35, it can be seen that the standard deviation values are 1,83 and 1,822 for real and MC simulated data respectively which is in line with the expectations.

4.3. Wind Power Calculation According to Wind Speed Data

In this section, wind power calculation according to measured and simulated wind speed data is going to be performed. This transformation (calculation) is needed because of the index which exists at the market (EEX). As it is mentioned at chapter 2.1.1.2. Wind Power Index. The power index is calculated according to installed capacity and used capacity, in other words, capacity factor. It would be better to recall the formula 7 as shown below;

$$\%\ \mathbf{wind\ power\ index} = \frac{\mathbf{Actual\ kWh\ produced}}{\mathbf{Maximum\ Energy\ that\ might\ have\ produced\ during\ the\ same\ period}} \qquad \mathbf{(7)}$$

There is a difference between the power which is held by the wind and the power which can be extracted from the wind by a wind turbine. The power which can be extracted from the wind by a wind turbine was explained at the same section again with formula 8;

$$\mathbf{P} = \frac{\mathbf{A\rho}}{\mathbf{2}}\mathbf{v^3Cp} \qquad \mathbf{(8)}$$

Since the parameters were explained before, the explanation is not made again. It is important to mention that the wind speed which hits the blades is not the same with the wind speed which is leaving from the wind turbine and wind speed which is leaving the wind turbine is not equal to zero. In that case generation, energy from the wind turbine would not be possible [11, 103].

The wind speed which is hitting the blades is called as the upstream wind speed and wind speed which leaving the wind turbine is called as the downstream wind speed. The v (m/s) showed at the formula 8 is the upstream wind speed. As it is mentioned here that, since the whole power from the wind cannot be extracted, in other words since the downstream wind speed is not equal to 0, using the upstream wind speed directly might seem wrong. However, that decrease is provided with C_p which is called the power coefficient. Generally, the value of this coefficient is provided by the producers [100].

Before explaining the calculation of the wind power and wind turbine selection part, the importance of the upstream value showed in formula 8 should be explained. There is one important thing that it is needed to take care of. As it was demonstrated in the formula 7, the option price depends on the energy that is actually generated and maximum energy that might have been produced. This explanation means that the capacity factor comparison should be valid for a certain period of time. At this point, the importance of the value of the wind speed which will be used at formula 8 comes to exist. Taking the average value for the whole contract duration might mislead and causes having wrong results. Therefore, daily comparison is performed while calculating the capacity factor difference. The example shown below would describe the difference better [103].

Example: This example aims to show the wind energy difference when the average wind speed of the whole time period is used against when real wind speeds are used. (It is assumed that A= 8000 m^2 and ρ = 1,2176)

Scenario 1: 5 m/s wind speed for 2 days

Generated Energy = $\frac{A\rho}{2}v^3 * 48\text{hour}$ = 0.5*8000*1,2176*$5^3$48 = 29,22 MWh

Scenario 2: 4 m/s 1 day and 6 m/s 1 day during the same period

Generated Energy = $\frac{A\rho}{2}v1^3.24\text{hour} + \frac{A\rho}{2}v2^3.24\text{hour}$ = 0,5*8000*1,2176*4^3*24 +0,5*8000*1,2176*6^3 *24 = 32,8 MWh

As it can be seen from the example, taking the average wind speed is causing to have wrong results. Therefore, during contract periods, the day-based comparison is performed instead of taking the average wind speed during the contract duration. This is the approach described at chapter 4.1. (Day-based comparison).

The second most important part of the calculation is calculating the capacity factors (wind power indexes) and the difference of this index between the real and MC simulated data. To calculate the power index (capacity factor), it is needed to calculate the power difference between the values of real data and MC simulated data. The difference is needed to be calculated between the cubic values of the wind speeds. This means that the cubic value of the wind speed of each real wind speed and MC Simulated wind speed difference is not calculated as shown below;

- $(WS_{real}-WS_{MC\ Simulated})^{3}$ ➔ Wrong
- $(WS_{real}^{3}-WS_{MC\ Simulated}^{3})$ ➔ Correct

This calculation will be explained in detail at the next chapter.

As it can be seen at formula 8, there are other parameters which are needed to be determined to calculate the real and simulated generation.

The first parameter is the air density (ρ) which has kg/m^3 as its unit. Air density is changing according to two main parameters; air pressure and temperature. The air density for the determined location is calculated according to the average annual temperature of New York [106] and 80 meters as the altitude (since it was assumed that, this was the tower length). The impact of the altitude becomes existing while calculating the air pressure. These parameters are calculated thanks to two online calculators [104, 105] and the final value of the air density parameter is reached as 1,2176 kg/m^3.

After the air density calculation, it is needed to choose the wind turbine. Web-pages of different producers were investigated, and the most convenient wind turbine is selected according to the capacity, height and cut in wind speed. General Electric GE 2,5-103 is chosen as the wind turbine which has 2,5 MW, 85 m and 3 m/s as the capacity, height and cut-in speed respectively [107]. Figure 28 [107] demonstrates the power curve of the wind turbine. This curve is important for reaching the C_p-values. Power coefficient values are changing according to upstream wind speed. The coefficients can be found in the Appendix 16.

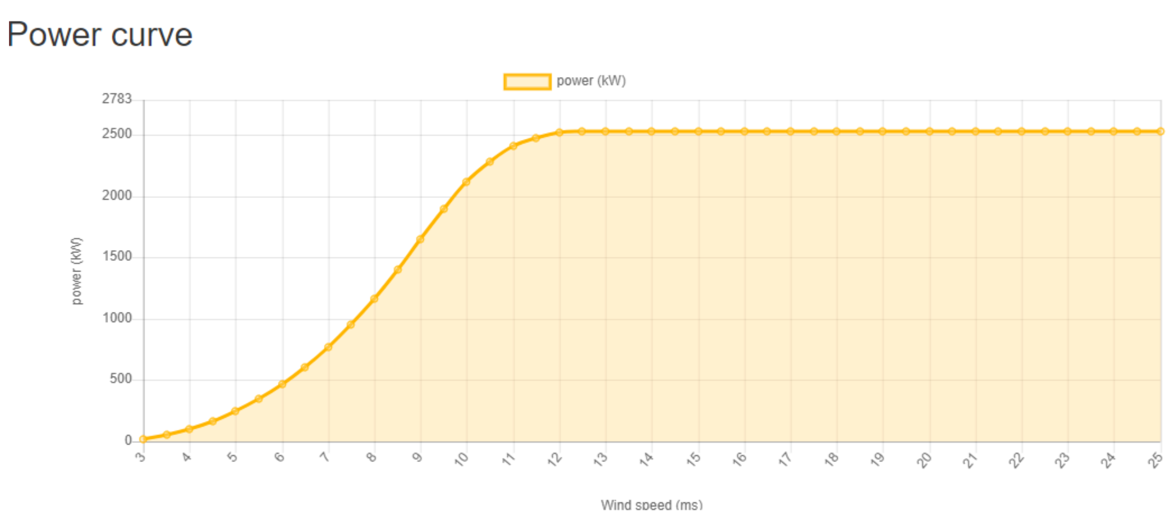


Figure 28. Power Curve of the Selected Wind Turbine

4.4. Price Decision for the Put Option

At this chapter, price calculations according to the different criteria given at this chapter will be performed. The contract duration will be changed while calculating the prices to see the seasonal price changes and its volatility.

In addition to that, since discounting is going to be used (to add the extrinsic value of the option into the calculation) during calculating the prices for the option contracts to calculate the present value of the contracts, choosing the discount rate (assumptions) will be discussed. Contract prices will be compared with the potential revenue lost to find out how it can be hedged against wind speed volatility risk wind this wind power put option.

At the end of the chapter, sensitivity analysis of the payoff function parameters which have impacts on the contract price will be demonstrated.

4.4.1. Discount Rate Choice and Discounting Method for Option Price Calculation

It is needed to remember the time value of money. Since the value of the money does not stay stable in time, it is needed to discount the contract prices with a meaningful discount rate which means that it is needed to put into the calculation of the time premium value. Time value can be described within the term of option extrinsic value which can be seen at formula 36.

Option Premium Price = Option Intrinsic Value + Option Extrinsic Value (36)

The intrinsic value of the options represents the payoff amount according to the contract which was demonstrated in formula 10 and 11. The extrinsic part of the option value can be defined as the difference between the option premium price and option's intrinsic value. There are two main factors which have impact on determining the extrinsic value of the option;

- Time Value
- Implied Volatility

Time is the first important factor which has an impact on the extrinsic value of the option. An increase in the duration of the contract would cause an increase in the extrinsic value. Contracts which have less duration would have lower prices. The duration of the contract has an impact to determine possibilities. In other words, an increase in time duration would cause a decrease for more accurate forecasting.

The behavior of the implied volatility is same with the time value. If the implied volatility is high for a determined derivative the premium value will be higher. The implied volatility is used to calculate the values of the Black – Scholes parameters. A higher volatility is going to cause higher prices. The impact of the implied volatility at this dissertation can be mentioned at seasonal volatility of the residuals. The contracts which are valid during the season where a higher seasonal volatility exists, have higher prices [120, 121, 122, 123].

For the case which is analyzed within this dissertation, the extrinsic value can be determined & calculated when the option is going to be exercised as shown in formula 36.

Apart from the extrinsic value of the option, it is needed to put the time value of the money into calculation. To perform this, it is needed to perform discounting and a determined and reasonable discount rate should be used.

It is assumed that the discount rate would be equal to the risk-free rate. The risk-free rate can be called as the required return rate without any risk. This definition is valid in theory and since it exists in theory, but not in practice, any specific determined value for the risk-free rate does not exist. Therefore, people at the market are generally choosing the governmental bond's rate as the risk-free rate. At this dissertation, it is assumed that this general rule is valid. Therefore, treasury yield rates will be used as a base for the discount rate.

The wind speed data set that is valid for NY. Therefore, US-based treasury yield rate is going to be used. Since the contract durations are changing as monthly, quarterly, annual, treasury yield rates according to these periods will be used. Since the minimum period of US treasury yield is 3 months and since the rate for 3 months, 6 months and 12 months are so close (from 2,42% to 2,44%) one yield are will be used and will be set as the main rate. The accepted discount rate is 2,43%. It is also assumed that the discount rate will be stable and will not change during the analysis period [110].

According to previous studies and researches, while calculating the option price in respect to the time value of the money, continuously compounding interest approach is used [5, 11, 16]. At this dissertation, the discounting type will be continuously compounding too which is in line with previous studies. Since the generation and selling prices are continuous, this approach is accepted.

$$\mathbf{PV = FVe^{-rT}} \qquad \textbf{(37) [118]}$$

Formula 37 demonstrated the continuously compounding discounting calculation where r and T represent the discount rate and the total duration of the relative contract or case.

According to these assumptions the present value of the option price will be calculated with formula 38 shown below;

$$\mathbf{PV(Option\ Premium\ Price) = FV(Option\ Premium\ Price) * e^{-rT}} \qquad \textbf{(38)}$$

Where r and T represent the risk-free rate and contract duration of the option. The monetary unit for the option price will be in dollars at this dissertation.

4.4.2. Prices of Wind Put Option Contract with Different Contract Periods

As it is mentioned before at the introduction part, EEX is currently the only exchange market where it is possible to find wind-related weather contracts. They are providing wind power futures and the payoff of the contracts depends on the capacity factor difference between the expected and realized. 1% capacity difference represents 1 € and payoff is calculated with multiplying this difference with duration of the contract (in hours). The value in percentage demonstrates the money with the €/h unit. Therefore, to have the payoff in Euros, it is needed to multiply the value which is in €/h with the contract period in hours [76]. Since Dollars are used as the main monetary unit at this dissertation, currency change is going to be performed. 1,13 will be the exchange rate for € to $ conversion which means that 1% difference represents 1,13 $. Since the prediction is performed for the future values, it is not possible to determine the exact value for the currency. Therefore, 1,13 will be accepted as the exchange rate and it is assumed that this rate remains same during the whole calculation period.

An option contract can be priced with different durations. In addition to contract period there is one more main driver which has an impact on the option price; capacity factor difference. Formula 39 demonstrates the pricing method of option contracts for EEX.

Option Premium Value = (Expected Capacity Factor – Observed Capacity Factor) *Contract Duration (39)

where option premium value, (Expected Capacity Factor – Observed Capacity), Contract Duration parameters have $, $/h, and h as their units respectively.

This payoff calculation approach is used at this dissertation. During the analysis, expected capacity factor and observed capacity factor represent the real-measured historical data and MC simulated data respectively.

Since contract duration is one of the main drivers during calculating the option price, different durations are used during the calculations. The duration can be in days, weeks or it can be longer. However, at this dissertation contract period is set as monthly, quarterly and annually.

Table 15 demonstrates the put option prices which have different contract durations. If formula 39 is recalled, it will be seen that there are two main parameters; difference and contract duration. For each contract, the number of days within the contract duration is calculated and multiplied with 24 to have the duration in hours.

Example: Contract Duration for January: 24*31= 744 hours

(Expected Capacity Factor – Observed Capacity Factor) is the second parameter which has an impact to determine the option price. The expected capacity factor represents the 20 years' data measured in NY. From those 20 years, an average is calculated for each day. The observed capacity represents the MC simulated data. MC simulation was performed 50.000 times and each of these 50.000 runs is compared with the calculated average values of the real historical data. During the calculation, the capacity factor difference is summed when the expected capacity factor is less than the observed capacity factor. Since this is a hedging strategy against less production, this approach is accepted at this dissertation. Related MATLAB code can be seen below;

```
WindEnergyDifference =(max(ExpectedCapacityFactor- ObservedCapacityFactor,0))*113
```

where expected and observed capacity factors are in percentage. Therefore, 113 is used to make the energy difference in $/h. (113 = 100*1,13 to calculate it from € to $)

The comparison of capacity factor is performed for each day. In the end, the sum value of the difference for each day is calculated to find out the total difference occurred during the contract period. Before calculating the sum value of the capacity factor difference, it is needed to discount these values to add the time premium value of the contract.

The discounting is done daily-basis. After calculating the capacity difference, it is multiplied with 113 to have it in a monetary unit. And as it is mentioned at 4.4.1., continuously discounting is performed to prevent overlook time value. It is important to emphasize that discounting is performed x times where x represents the total days in the contract period. In other words, the discounting is not performed after adding all values to reach the total difference. As it is mentioned the total day-based discounted capacity differences are summed to reach the total value. This process is done for each of the 50.000 runs.

The mean value of this 50.000 runs' discounted capacity difference is calculated with a standard deviation. As of the last step these values are multiplied with contract durations which are in hours to reach the option prices.

Table 15. Option Prices with Different Durations and Time Periods (Put Option)

Contract Duration	Option's Parameters		
	Mean Value & Option Price ($)	Standard Deviation ($)	Capacity Difference According to Wind Speed ($/h)
Jan	5.704,6	1.385,0	6,84%
Feb	4.250,2	1.161,7	5,61%
Mar	4.822,4	1.221,2	5,76%
Apr	4.329,8	1.063,5	5,35%
May	2.898,2	731,3	3,48%
Jun	2.390,3	573,0	2,97%
Jul	1.924,5	449,6	2,32%
Aug	1.133,7	339,0	1,37%
Sep	2.184,1	616,4	2,73%
Oct	3.456,8	922,0	4,19%
Nov	4.729,8	1.239,1	5,94%
Dec	4.477,7	1.215,2	5,45%
1st Quarter	14.825,0	2.186,1	6,09%
2nd Quarter	9.616,0	1.419,6	3,39%
3rd Quarter	5.036,7	816,2	2,06%
4th Quarter	12.616,0	1.978,5	5,13%
Annual	42.312,0	3.385,4	4,32%

The first column at Table 15 demonstrates the option prices ($), the second column represents the standard deviation and the last column represents the average capacity difference according to 50.000 MC runs.

It is assumed that these prices do not include any clearing or transaction costs. Normally, these costs should be added to the contract too.

4.4.2.1. Annual One Time Discounting

As it is mentioned at the previous chapter, putting discounting into the calculation is important for putting time value part into consideration. At this chapter discounting approach will be differentiated. The discounting approach can be determined within the parties at the OTC market or official exchange markets can determine their own approach.

Calculated values demonstrated in Table 15 were performed according to day-based discounting. On the other hand, in this chapter, a different approach will be used and a comparison between the two approaches will be discussed.

The different point of the discounting calculation will be the discounting times. At the previous approach, it is mentioned the total discounted day-based capacity differences are summed to reach the total value. This time, the sum of the day-based capacity difference will be calculated and one discounting is performed at the end of the contract. In other words, the discounting is performed after adding all values to reach the total difference.

Table 16. Option Price Difference According to Discounting Approach

Contract Duration	Option's Parameters		
	Mean Value & Option Price ($)	Standard Deviation ($)	Capacity Difference According to Wind Speed ($/h)
Annual - Day Based Discounting	42.312	3.385,4	4,32%
Annual - One Time Discounting	41.853	3.363,7	4,32%

Table 16 demonstrates that day-based discounting gives higher prices for the contracts. Since the discounting period for each day except the last day is longer, it is expected to reach these results.

4.4.2.2. Put Option Price Calculation According to Log-Normal distribution

During this chapter, another approach will be used to determine the option price. With this approach it is also possible to determine option prices with any contract duration. However, only the option price which has a year as its contract duration will be calculated to make the general evaluation. The option prices demonstrated in Table 15 were calculated with steps shown below;

- Box-Cox Transformation of raw wind speed data (to make the data normal distributed to make the analysis easier)
- Seasonality and Trend Analysis of the transformed data

- Removing Seasonality from the data (Since the trend is so low, the trend was not removed from the data)
- Choosing the autoregression model and calculating the AR parameters for the lags
- Calculating residuals according to chosen AR model
- Calculating variances (seasonal volatilities) of the residuals for each day and removing it from the residuals
- Generating (normal distributed) random numbers for the white noise part of the residuals and adding the seasonal volatility to reach the residuals
- Putting these residuals into regression to calculate whole wind speed data
- Adding back the seasonality which was removed before
- Re-transform Box-Cox data into the original version of the wind speed data

As the last step, these steps were performed 50.000 times and compared with the real-measured historical data and mean value of the difference is accepted as the capacity factor difference. (after transforming wind speed to energy)

The approach which will be performed during this chapter also includes MC strategy. As it was demonstrated in 4.2. Distribution Fitting for Real and Simulated Data chapter, log-normal distribution fits better to the real-measured historical wind speed data. The mean parameter and standard deviation of the neutral log transformed data was demonstrated in Table 14 and is demonstrated again in Table 17.

Table 17. Log – Normal Distribution Parameters for Real Measured Data

Log Normal Distribution		
Parameters	μ	σ
Real	1,62535	0,331698

The main difference at this approach will be the random number generation part. The random number generation will be done according to Log-Normal distribution. For this random number generation, MATLAB will be used once again. MATLAB's logrnd function can be used to generate random numbers according to the Log – Normal distribution. The relevant function can be found below;

```
lognormalrandomnuumbergeneration(i,j)=lognrnd(1,62535,0,331698);
```

As it can be seen from the function, it is needed to put the log-transformed parameters to generate the random numbers. These parameters can be calculated by MATLAB automatically thanks to the distribution fitter application.

During this approach, steps which are listed below are followed;

- Calculating the main parameters of raw wind speed data (Log – Normal distribution Parameters)
- Random Number Generation with lognrnd function of MATLAB
- Compared the 50.000 run results with raw wind speed data
- Calculate capacity factor difference after transforming wind speed to energy)

Table 18 demonstrates the option prices which have one year as the contract duration for two approaches. The second approach has a higher price. Even though the same parameters of the Log-Normal distribution according to real wind speed data is used, it ended with higher expected pay-off, option price.

Table 18. Option Prices Calculated with Two Different Approaches

Contract Duration	Option's Parameters		
	Mean Value & Option Price ($)	Standard Deviation ($)	Capacity Difference According to Wind Speed ($/h)
Annual – AR & Seasonality Included Approach	42.070	3.374,6	4,30%
Annual – Log-Normal Distribution Random Number Generation Approach	49.756	2.726	5,08%

Annual – AR & Seasonality Included Approach is going to be called as 1st approach where Annual – Log- Normal Distribution Random Number Generation Approach is going to be called as the 2nd one. The reasons caused these results are listed below;

- Residuals GOF (goodness of fit) to normal distribution is better than wind speed data's GOF to Log-Normal distribution
- The first approach estimates the MC simulated wind speed values according to AR, but second approach does not include the relation between the values
- The first approach also deals with seasonality and puts it into the calculation where the second approach does not.

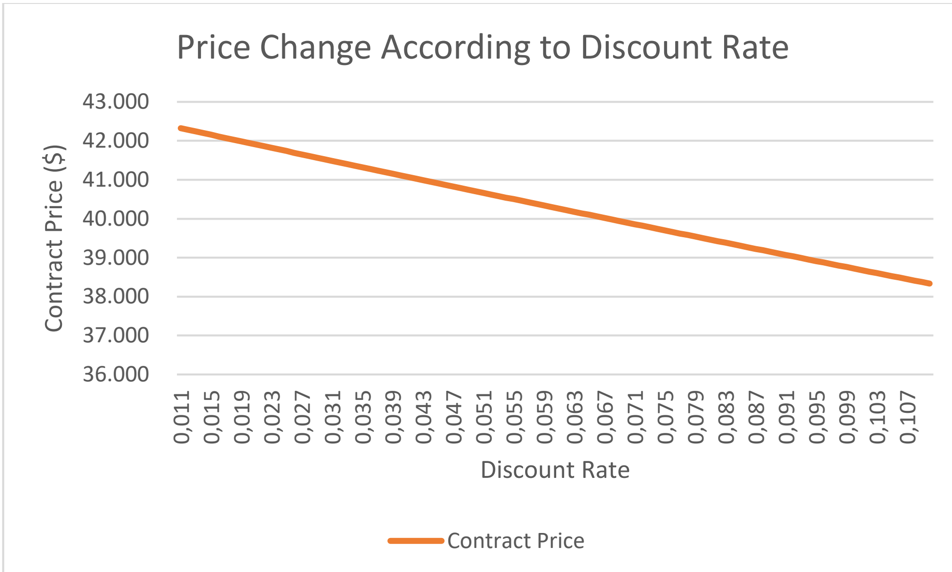
Both approaches can be used to determine the option price. It depends on the agreements between the parties at the OTC market or it depends on the official contract providers (exchange markets).

4.4.3. Option Price - Sensitivity Analysis

At this chapter, sensitivity analysis which can show the importance of the parameters and can also find out the most important main driver is performed to see the impacts of the parameters. During calculation of the put option price, capacity factor difference, contract durations were set as the main parameters. In addition to these two parameters, the discount rate has also impact on the option price due to the time value of the money. In summary sensitivity analysis for option price will be performed for;

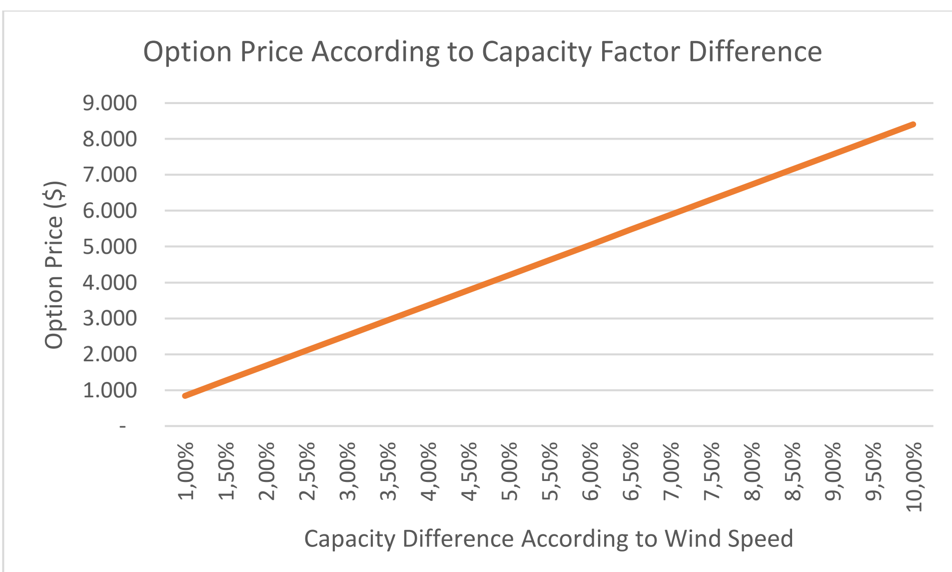
- Contract Duration
- Discount Rate
- Capacity Factor Difference

Graph 5 demonstrates the less important parameter which has an impact on the option price. It is known that the discount rate has an inverse proportion with the option price because when it is increased, the present value of the money will be less. Therefore, Graph 5 demonstrates an expected result. When the discount rate increases, the current price of the option is getting decreased.



Graph 5. Option Price Change According to Discount Rate

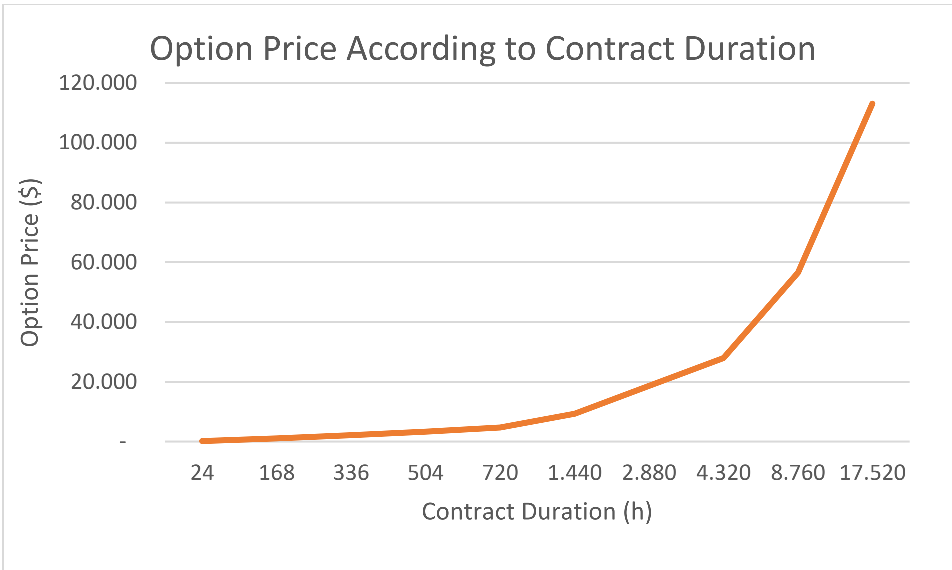
The impact of another parameter can be seen in Graph 6. It is important to remember that the allowed minimum changes were on the 4th digit after the decimal. Graph 6 demonstrates the change in the capacity factor difference with 0,5% steps. If the capacity factor difference changes from 1% to 2% the price will be double. This parameter demonstrates a linear impact as it can be seen from the Graph 6.



Graph 6. Option Price Change According to The Capacity Factor Difference

The impact of the last parameter can be seen in Graph 7. It is needed to determine the minimum contract allowance. If the minimum allowance is set as 1 hour, the impact of this parameter will be the same with the capacity factor difference. However, the minimum contract duration can be a week at EEX.

Therefore, changing the contract from minimum allowance (1 week) to two weeks would increase the price 168 times more. In summary, it is possible to say that, contract duration is the most important parameter during calculation the option prices according to the current market conditions. However, the minimum contract duration might be different at OTC markets. If it is allowed to change the contract duration in hours, the impact of this parameter would be linear like capacity difference parameter.



Graph 7. Option Price Change According to Contract Duration

4.4.4. Comparison of Option Price and Revenue Loss

In this section evaluation of option price against the revenue loss will be evaluated. It is important to understand how it can be hedged against wind speed volatility risk wind this wind power put option.

Before starting analyzing the situation, it should be better to mention one assumption. During the comparison analysis, it is assumed that a wind turbine owner is selling directly the electricity that he/she produce and since the owner is a renewable energy producer it is assumed that the owner has the guarantee of selling the whole electricity to the grid without any problem.

There is one more assumption which was accepted. Even though the relevant wind speed data is valid for NY, since the contract provider EEX provides contracts only for Germany and Austria, the electricity selling prices will be taken according to Germany legislation.

It is known that the general payoff function of a put option is [90];

$$\mathbf{Vp = max\ (E - S) - P_p} \qquad \mathbf{(38)}$$

where E, S, Pp represent exercise price, stock price and option premium respectively. As a put option buyer (long position), it is important to have E-S > Pp to prevent loss and gain revenue. During the comparison one-year period will be accepted as the contract duration.

The expected E-S value for this case is approximately 4,32%. As the option owner, the profit generation will start with 4,32% and bigger differences.

If the payoff function of the EEX is recalled, for each 1% difference, 1,13$ will be paid. With the wind turbine used during the analysis 1% capacity difference represents 25 kWh (in one hour) and 219 MWh in a year.

The electricity feed-in tariff prices differentiate according to two types [115];

- **Onshore:** € cent 4,66 – 8,38 per kWh (§ 46 EEG 2017) minus € cent 0,4 per kWh (§ 53 no. 2 EEG 2017
- **Offshore:** € cent 3,9 – 1,4 per kWh (§ 47EEG 2017) minus € cent 0,4 per kWh (§ 53 no. 2 EEG 2017).

Since the price conditions are changing according to applicants, it is assumed that the owner would receive the average price of these feed in tariffs which are;

- **Onshore:** (6,52 – 0,4) = 6,12 € cent per kWh = 6,915 $ cent/kWh
- **Offshore:** (6,52 – 0,4) = 2,1 € cent per kWh = 2,373 $ cent/kWh

In case of 25kWh of loss, the potential loss will be;

- **Onshore:** 6,915 $ cent/kWh * 25 kWh = 1,73 $
- **Offshore:** 2,373 $ cent/kWh * 25 kWh = 0,60 $

Depending on the wind turbine type, the benefit of the contract is changing. If the wind turbine is an offshore one, after the exercise point (4,32%), it is possible to cover the revenue loss with extra payments. On the other hand, if the wind turbine is an onshore one, after the exercise point (4,32%), the benefit of the option would be a decrease in the revenue loss. The revenue loss will increase however, the impact would be less.

Conclusion

The technology improves every day. Thanks to this improved technology, people can take actions including protecting themselves against potential risks by changing or controlling them. Even though mankind has very good level of technology, there is one thing that people cannot control; the events and incidents caused by the weather. Since mother nature will always continue to surprise us, it is needed to be ready against it.

It is possible to predict the weather-related events and parameters. Thanks to this opportunity people can prepare themselves against the possible outcomes of nature. Weather derivatives are one of the outputs of this preparation to hedge against weather-related situations. This output is protecting people against the nature and weather in economic aspects. Many sectors which are currently active and popular in the world are generally weather sensitive. Therefore, the importance and popularity of the weather derivatives have been increasing day by day. The weather's impact on USA's economy worth billions. From 1997 at OTC market and from 1999 at CME, weather derivatives started to be traded. During the first two years after the first launch of the weather derivative at CME, the transactions reached 11.2 billion dollars which can be seen as a good evidence for its importance. Thanks to this success, different types of weather derivatives such as option, future, swap, etc. with different weather indexes such as temperature-related, wind-related, snow & rain-related, became widespread at both OTC and official exchange markets.

The exchange markets where weather derivatives can be found, has been changed during the last twenty years. Currently, CME in USA and EEX in Europe is leading the sector. However, it was possible to find weather derivatives in Liffe (London) and USFE (USA) exchange markets. OTC market should not be underrated within the scope of transaction amount.

Before weather derivatives it was not possible to hedge against unexpected non-catastrophic weather conditions. Hedging against catastrophic conditions has been provided with weather insurances which covers generally the physical damages and expenses. The gap for covering the losses caused by situations such as, decrease in sales or delay in a construction is fulfilled by weather derivatives which differentiate itself with its underlying asset from the standard derivatives. Apart from its underlying asset, other parameters of weather derivatives are very similar to the parameters of standard derivatives (cap amount, premium and etc.).

Most frequent indexes which are used for weather derivatives are; HDD and CDD for temperature, wind speed and wind power for wind, snow and rainfall indexes for snow and rain related derivatives. For temperature and wind related derivatives, payoff function is quite similar. It is needed to determine a base, reference point according to historical data and pay-off amount is calculated according to the difference between this reference point and the real measured data. There is one important step which should be calculated for the wind power derivatives. It is needed to know and calculate the wind power according to measured wind speed, and determined swept area of the wind turbine, air density, power coefficient of the wind turbine correctly.

Thanks to these many types of weather derivatives (options, futures, swaps etc.) with different types indexes (temperature, wind, snow, rain, hurricane) it is possible to hedge in many situations. The hedging strategy can be in short or long position which is same with standard derivatives. The strategy is generally decided according to sector and purpose.

Even though, temperature related derivatives are the most common derivatives, the importance and transactions of the wind-related weather derivatives are increasing, and this increase occurs in line with the share of the renewable energies within all energy production and installed capacity. Since there are obligations about increasing the renewable share within the whole energy production or installed capacity, new power plants which uses renewables to produce energy have been installed more frequently. Therefore, **the main aims of this dissertation were calculating price of wind power put option for a determined location with two different approaches and make a sensitivity analysis to find out the main driver of the price.** The underlying asset of the wind power option described at this dissertation was set as the capacity factor difference of the selected (assumed) wind turbine & power plant.

The price for the wind power put option can be described as the discounted expected pay off according to difference between the historical measured wind speed's capacity factor and the estimated capacity factor according to historical measured wind speed. Therefore, a deep statistical analysis was needed to meet this requirement. Under the light of this analysis, the estimation, in other words prediction of the possible future wind speed was performed with MC simulation approach.

The measured wind speed data used within this dissertation is valid for New York and it was provided from USFE database by Prof. Benth. It was assumed that the altitude of the measurement was 80 meters. The assumption was validated with annual average wind speed which is provided by US government.

As mentioned before, MC approach was used for estimating the wind speed. MC approach is based on random number generation. There are two alternatives for generating the random numbers.

1. Random number generation for residuals coming from autoregression model
2. Random number generation according to the distribution of real measured wind speed data

Both approaches were performed at this dissertation and comparison of the results was provided. For the first approach, it is needed to perform some steps which is a general approach at time series analysis; seasonality and trend detection, decomposition of seasonality, setting autoregression model and residual analysis.

It is known that; statistical analysis is easier when the relevant data is normal distributed. It is possible transform non-normal distributed data into the normal distributed one with Box-Cox transformation. After performing the transformation, seasonality period was needed to be detected to set the seasonality model with Fourier series. Periodogram was used for this aim and two seasonal periods, which are semi-annual and annual, were detected. For the trend analysis, it is needed to think the main factor which might cause a decreasing or increasing trend. Construction of the new buildings is the main reason which might cause this situation. It was assumed that the exact location for the measurement of the wind speed is available to build and install a wind turbine and generally these areas are located far away from the residential areas. Therefore, since it was not expected to have a new construction at the determined area, a trend within the data set was not expected too. This expectation was validated with the trend analysis which was performed via Minitab software. Moving average of the wind speed data also demonstrated the same result.

After modeling the seasonality with Fourier series, this modeled seasonality was removed from the main data. Since trend was not detected, additive approach was used for the decomposition. Seasonality detection and decomposition is one of the most important steps of time series analysis.

Setting a regression model is the key point of the prediction. An accurate regression model would help to determine and calculate the dependent variable according to its predictors. If it is needed to set a regression model for a time-series data, autoregression or moving average approach should be used. There is a chance to use also combination of autoregression and moving average, ARMA. These regression models are valid for stationary data. Therefore, a stationary test was run (ADF Test) to verify stationarity of the data.

As next step, comparison of different models was performed with different lags. According to Akaike Bayesian and Criteria's values and also p-values of the regression coefficients, AR (3) model was chosen to set the regression model. According to AR (3), it was needed to calculate the residuals which describes the difference between the real and predicted value.

These residuals are the main point for random number generation with the first approached described above. During the residual analysis, two different approaches were followed again.

1. Detecting the seasonal volatility of the residuals and decompose it to reach standardized residuals which is distributed with N (0,1) and generate number with this normal distribution for all days
2. Calculating the variance of each day's residual and generate numbers for each day with different distributions which are normally distributed but with different variances

Within the scope of first approach, seasonal volatility was modeled for the residuals. Once again, Fourier series was used to model this seasonal. For the second approach, it was accepted that the variance is periodically repeating itself over a year. (σ^2 (t+365) = σ^2 (t))

These two approaches provided very similar results while calculating the price of the option. After generating the random numbers with MATLAB (normrnd function) with one of these approaches to have the residuals, same steps starting from the end were performed to reach the wind speed data with measured real wind speed data's format.

The residuals were taken and put into the regression model. Since the expected value of the residuals is zero and since there are 3 lag operators for the AR model, the first three values of the residuals were set as zero at the beginning of the analysis. After the calculation of each dependent variable, seasonality was added back into the data and re-transformed from Box-Cox into the original form (This form of the data is called as MC simulated data from now on).

To make comparison of similarity between the real measured and MC simulated data, Chi-Square tests, mean value and variance comparison was performed. All these comparison tests gave satisfied results. In addition to these tests, one more control was performed. It was found that the most fitted distribution for the real measured and MC simulated data is Log-Normal distribution. The calculated parameters of Log-Normal distribution for MC simulated data was within the confidence level of calculated parameters of Log-Normal distribution for real-measured data. The difference between these two data sets was the main point of the expected pay-off, in other words, price of the option.

Wind power derivatives provided by EEX were used to determine the tick size which transform this expected difference into a parameter with monetary unit. EEX is using 1,13$/h for 1% capacity factor difference occurred because of wind speed change. Since the hedging is wanted to set against the lower energy production (for one wind turbine), **1,13* max (capacity factor according to real measured data – capacity factor of MC simulated data, 0) *contract duration (in hours)** was used as the pay-off function. To calculate the capacity factor, potential power production was calculated and divided by the installed capacity of the wind turbine.

The prices were calculated for 12 different months, 4 different quarters and for one year. As it was mentioned, the prices were increasing in line with the contract duration. In addition to this, it was found that the prices are increasing for the months which has higher seasonal volatility for the residuals.

Another approach which was mentioned above for the random number generation, was generating random numbers according to the most fitted function to the real-measured data; Log-Normal (lognrnd function of MATLAB was used). The calculated option prices are higher than the previous approach. Thanks to this fact, it is possible to say that residuals GOF to normal distribution is better than wind speed data's GOF to Log-Normal distribution. In addition to this, the first approach estimates the MC values according to AR, but second approach does not include the relation between the values of the real data. In other words, generating random numbers for residuals and putting them into the regression gives closer estimated data to the real data. During the calculation of option price with these two approaches, risk free rate was used as the discount rate to calculate the present value of the option price.

To find out the main driver of the option price, sensitivity analysis was performed. The sensitivity analysis was run for three different parameters; discount rate, capacity factor difference and contract duration. Even though, the all pay off function depends on the capacity factor difference, the main driver which has the biggest impact on the price is the contract duration. This result is faced because of the current exchange market conditions. According to market conditions the minimum change in contract duration is one week. Therefore, changing the contract from minimum allowance (1 week) to two weeks would increase the price 168 times more. On the other hand, If the minimum allowance was set as 1 hour, the impact of this parameter would be approximately same with the capacity factor difference.

In summary, calculating the prices for the option contracts with random number generation for residual and putting them into regression gives more accurate results. The prices are higher for the months which have higher seasonal volatilities.

The incentives support for the wind energy is changing according to the wind turbine type; offshore and onshore. Therefore, according to the wind turbine type (offshore or onshore), the benefits of this option contract is changing too. For offshore wind turbines this option contract provides a cover against the revenue loss with extra payments. On the other hand, for an onshore wind turbine, this option contract will provide a decrease in revenue loss, in other words this option would not cover the whole revenue loss but will decrease it. This situation is valid for the current situation, a new analysis might be needed to perform to see the situation in the future.

Bibliography and References

[1] Benth, F. E., Benth, J. Š.: Modeling and Pricing in Financial Markets for Weather Derivatives. Advanced Series on Statistical Science & Applied Probability: Volume 17. 2012

[2] Yamada Y: Handbook of Wind Power Systems, Chapter 3-, Risk Management Tools for Wind Power Trades: Weather Derivatives on Forecast Error, 39-66

[3] AKKAYA, G Cenk: HAVA DURUMU TÜREVLERİ VE İKLİM RİSKLERİNE KARŞI POTANSİYEL FAYDALARI ÜZERİNE BİR İNCELEME, 1997.

[4] CHAMBERS, Nurgül assoc. prof. dr. İklim Türevleri ve Fiyatlandırılması (Weather Derivatives and Their Pricing). P. 55–66.

[5] DUMITRAS, DANIELA and KUVAITSEVA, REGINA. Pricing Temperature Weather Derivatives. MSc. Lund University, 2012.

[6] CME GROUP, 2011. The Weather Derivatives Markets at CME Group: A Brief History. [online]. 2011. Available at:

Available at:

https://www.cmegroup.com/education/files/Weather_Derivatives_Markets_at_CME_Group.pdf

[7] BELLINI, FRANCESCA. The weather derivatives market: Modeling and pricing temperature. Ph.D. University of Lugano, 2005.

Available at: https://doc.rero.ch/record/5176/files/1_2005ECO001.pdf

[8] CARABELLO, FELIX. Introduction to Weather Derivatives. Investopedia [online]. 2018.

Available at:

https://www.investopedia.com/trading/market-futures-introduction-to-weather-derivatives/

[9] JANDA, KAREL and VYLEZIK, TOMAS. Financial Management of Weather Risk with Energy Derivatives. 2011. Munich Personal RePEc Archive,Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague

[10] TASTAN, BIRHAN and HAYFAVI, ASSOC. PROF. DR. AZIZE. MODELING TEMPERATURE AND PRICING WEATHER DERIVATIVES BASED ON TEMPERATURE. Ph.D. Middle East Technical University, 2016.

[11] HOYER, STEVEN and OOSTERLEE, PROF.DR. C. W. Wind derivatives: hedging wind risk. MSc. Delft University of Technology, 2013.

[12] STULEC, Ivana, BAKOVIC, Tomislav and HRUSKA, Domagoj, 2012. Weather Risk Management in Energy Sector. Annals & Proceedings of DAAAM for International 2012 &. 2012. Vol. 23, no. 1, p. 89–94.

[13] GEYSER, J M. WEATHER DERIVATIVES: CONCEPT AND APPLICATION FOR THEIR USE IN SOUTH AFRICA. Agrekon. 2004. Vol. 43, no. 4, p. 444-464. DOI 10.1080/03031853.2004.9523660. Informa UK Limited

[14] PRABAKARAN, S. DR. TEMPERATURE STOCHASTIC MODELING FOR PRICING OF WEATHER DERIVATIVE MARKET. . Presentation.

Available at :

https://cea.javeriana.edu.co/documents/153049/9215453/Temperature+stochastic+modeling+for+pricing+of+weather+derivate+market-Prabakaran.pdf/a2a7eed8-b9bc-4144-9d8f-d16776a55796

[15] Weather Market Overview. Climetrix [online].

Available at: https://www.climetrix.com/WeatherMarket/MarketOverview/

[16] BENTH, FRED ESPEN and BENTH, JŪRATĖ ŠALTYTĖ. Dynamic pricing of wind futures. In: Energy Economics. 31. 2009. p. 16-24.

[17] Weather Options Overview – CME Group. Cmegroup.com [online]. 2016.

Available at:

https://www.cmegroup.com/education/articles-and-reports/weather-options-overview.html

[18] Weather Derivatives: Answers to 10 most popular questions – Part B | Dr. Antonis Alexandridis' Research Webpage. Blogs.kent.ac.uk [online]. 2015.

Available at:

https://blogs.kent.ac.uk/antonis/2015/07/31/weather-derivatives-answers-to-10-most-popular-questions-part-b/

[19] SARASWAT, NEHA. WEATHER DERIVATIVES, An Emerging Market in Indian Context. MA. Gokhale Institute of Politics and Economics, 2012.

[20] What are weather derivatives? Jbwere.com.au [online].

Available at: https://www.jbwere.com.au/blog/2016/12/what_are_weatherder

[21] Innovative weather bonds get cool reception from investors. Bizjournals.com [online]. 1999.

Available at: https://www.bizjournals.com/houston/stories/1999/11/22/story7.html

[22] KORDI, KONSTANTINA. Pricing Weather Derivatives. MSc. University of Piraeus, 2012.

[23] CUTHBERTSON, PROF. KEITH. ENERGY AND WEATHER DERIVATIVES. KEITH CUTHBERTSON FINANCE MAN [online].

Available at:

http://keithcuthbertson.weebly.com/uploads/3/8/9/1/38918267/ch-27-energy-weather-final.pdf

[24] CAO, MELANIE, LI, ANLONG and WEI, JASON. Weather Derivatives: A New Class of Financial Instruments. SSRN Electronic Journal [online]. 2003. P. 1-30. DOI 10.2139/ssrn.1789810.

Available at:
https://poseidon01.ssrn.com/delivery.php?ID=697001127126073123125118075124082074103024036044086003100074100125124081023125089029110032053022109049003123000070100100065113116083094022086123071107114127005054040019020023020095122120111024001066095020018091100008015088004104100080076115065120&EXT=pdfElsevier BV

[25] CME Group Overview. Cmegroup.com [online].

Available at: https://www.cmegroup.com/company/history/

[26] Weather Products Homepage - CME Group. Cmegroup.com [online].

Available at: https://www.cmegroup.com/trading/weather/

[27] INVESTOPEDIA. Over-The-Counter Market. Investopedia [online].

Available at: https://www.investopedia.com/terms/o/over-the-countermarket.asp

[28] STELL, JOHN. Results of the 2005 PwC Survey, Weather Risk Management Association. Presentation. 2005.

[29] ELİBÜYÜK, UFUK, YAKUT, ALI KEMAL and İBRAHIM ÜÇGÜL, İBRAHIM. Süleyman Demirel Üniversitesi Rüzgâr Enerjisi Santrali Projesi. Dergipark.gov.tr [online]. 2016.

Available at: http://dergipark.gov.tr/download/article-file/327653

[30] ESPEN BENTH, FRED. Wind Speed Data Request of [16] via e-mail. [email]. 2018.

E-mail address: Fred Espen Benth <fredb@math.uio.no>

[31] TASHPULATOV, MGR. SHERZOD, M.A., PH.D. Lecture 1: Statistical fundamental concepts. Descriptive Statistics. Moments. Lecture. 2017.

[32] Wind turbine heights and capacities have increased over the past decade. Eia.gov [online]. 2017.

Available at: https://www.eia.gov/todayinenergy/detail.php?id=33912

[33] GEERTS, B. and LINACRE, E. Roughness length. Www-das.uwyo.edu [online]. 1999

Available at: http://www-das.uwyo.edu/~geerts/cwx/notes/chap14/roughness.html

[34] SCOTT, DAVID. Box-Cox Transformations. Onlinestatbook.com [online].

Available at: http://onlinestatbook.com/2/transformations/box-cox.html

[35] What is a Box Cox Transformation Statistics How To [online]. 2015.

Available at: https://www.statisticshowto.datasciencecentral.com/box-cox-transformation/

[36] OLIVIER, JAKE and NORBERG, MELISSA M. Positively Skewed Data: Revisiting the Box-Cox Power Transformation. International Journal of Psychological Research [online]. 2010. Available at:

https://www.researchgate.net/publication/50946373_Positively_Skewed_Data_Revisiting_the_Box-Cox_Power_TransformationResearchGate

[37] Detecting Seasonality Using Fourier Transforms in R - Anomaly. Anomaly [online]. 2015.

Available at: https://anomaly.io/detect-seasonality-using-fourier-transform-r/

[38] Fourier Series- MATLAB & Simulink. Mathworks.com [online].

Available at: https://www.mathworks.com/help/curvefit/fourier.html

[39] Seasonal Adjustment Using a Stable Seasonal Filter- MATLAB & Simulink. Mathworks.com [online].

Available at: https://www.mathworks.com/help/econ/seasonal-adjustment.html

[40] Methods and formulas for Decomposition - Minitab. Support.minitab.com [online].

Available at: https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/time-series/how-to/decomposition/methods-and-formulas/methods-and-formulas/#additive

[41] EMEC, HAMDI. ZAMAN SERI ANALIZINDE TEMEL KAVRAMLAR. . Presentation.

Available at:

http://kisi.deu.edu.tr//hamdi.emec/UygEko/ZAMAN_SERI_ANALIZINDE_TEMEL_KAVRAMLAR.pdf

[42] HANEDAR, ONDER. DURAĞANLIK ANALİZİ, BİRİM KÖK TESTLERİ VE TREND. Lecture.

Available at: http://debis.deu.edu.tr/userweb//onder.hanedar/dosyalar/Metin.pdf

[43] STAFF, INVESTOPEDIA. Seasonality. Investopedia [online].

Available at: https://www.investopedia.com/terms/s/seasonality.asp

[44] Seasonal Filters- MATLAB & Simulink. Mathworks.com [online]. 2018.

Available at: https://www.mathworks.com/help/econ/seasonal-filters.html

[45] GUNES, MERVE. ZAMAN SERİLERİNDE MÜDAHALE ANALİZ TEKNİKLERİ ÜZERİNE BİR ÇALIŞMA. M.Sc. NAMIK KEMAL ÜNİVERSİTESİ, 2010.

[46] Hyndman, R. and Athana-sopou-los, G. (n.d.). Forecasting: Principles and Practice. [ebook] OTexts, p.Chapter 6 Time series decomposition. Available at: https://otexts.org/fpp2/decomposition.html.

[47] Alonso, A. and Martos, C. (2012). Time Series Analysis Autoregressive, MA and ARMA processes.

Available at: http://www.etsii.upm.es/ingor/estadistica/Carol/TSAtema4petten.pdf

[48] Onlinecourses.science.psu.edu. (2018). Lesson 1: Time Series Basics | STAT 510. [online]

Available at: https://onlinecourses.science.psu.edu/stat510/node/41/.

[49] Itl.nist.gov. (n.d.). 6.4.4.4. Common Approaches to Univariate Time Series. [online]

Available at: https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc444.htm#MA.

[50] YALTA, A. (2011). Zaman Serileri Ekonometrisine Giris - Box-Jenkins Yöntemi.

[51] Wasserman, L. (n.d.). Lecture Notes 16 Model Selection. [online] Stat.cmu.edu.

Available at: http://www.stat.cmu.edu/~larry/=stat705/Lecture16.pdf.

[52] Snipes, M. and Taylor, C. (2014). Model selection and Akaike Information Criteria: An example from wine ratings and prices. Science Direct. [online]

Available at: https://www.sciencedirect.com/science/article/pii/S2212977414000064.

[53] Investopedia. (n.d.). Autoregressive Integrated Moving Average - ARIMA. [online]

Available at: https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp.

[54] People.duke.edu. (n.d.). Introduction to ARIMA models. [online]

Available at: https://people.duke.edu/~rnau/411arim.htm.

[55] Ugurlu, E. (2015). Durağanlık, Birim Kök Sınamaları- Stationarity, Unit Root Tests. ResearchGate.

Available at: https://www.researchgate.net/publication/ 281647245_ Duraganlik Birim Kok Sinamalari-Stationarity_Unit_Root_Tests

[56] Augmented Dickey–Fuller Test. (n.d.). Rstudio - Help.

[57] Uh.edu. (n.d.). Short Introduction to Time Series. [online]

Available at: http://www.uh.edu/~bsorense/arma-intro2012.pdf.

[58] Support.minitab.com. (n.d.). Residual plots for Fit Regression Model - Minitab. [online]

Available at : https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics /regression/how-to/fit-regression-model/interpret-the-results/all-statistics-and-graphs/ residual -plots /.

[59] Otexts.org. (2019). Evaluating the regression model - Chapter 5.3. [online]

Available at: https://otexts.org/fpp2/regression-evaluation.html.

[60] TASHPULATOV, MGR. SHERZOD, M.A., PH.D. Lecture 8: Pairwise regression analysis. (2018), Czech Technical University in Prague

[61] BENTH, F. and BENTH, J. (2006). The volatility of temperature and pricing of weather derivatives. Quantitative Finance, Vol. 7, No. 5, October 2007, 553–561, pp.553-561.

[62] Investopedia. (n.d.). Understanding Variance. [online]

Available at: https://www.investopedia.com/terms/v/variance.asp.

[63] Mathworks.com. (n.d.). Nonlinear regression - MATLAB nlinfit. [online]

Available at: https://www.mathworks.com/help/stats/nlinfit.html.

[64] Newonlinecourses.science.psu.edu. (2019). Residuals vs. Fits Plot | STAT 501. [online]

Available at: https://newonlinecourses.science.psu.edu/stat501/node/277/

[65] Benth, J. and Benth, F. (2012). Energy Economics - A critical view on temperature modeling for application in weather derivatives markets. 34th ed. Science Direct, pp.592 - 602.

[66] Campbell, S. and Diebold, F. (2000). Weather Forecasting for Weather Derivatives, The Wharton Financial Institutions Center.

[67] Auquan, T. (2017). Time Series Analysis for Financial Data VI— GARCH model and predicting SPX returns. [online] Medium.

Available at: https://medium.com/auquan/time-series-analysis-for-finance-arch-garch-models-822f87f1d755.

[68] University of Washington, Lecture 7 Linear Regression Diagnostics. (2004).

Available at: https://courses.washington.edu/b515/l7.pdf

[69] IIST, Massey University. (n.d.). Standardised residuals. [online]

Available at: http://www-ist.massey.ac.nz/dstirlin/CAST/CAST/Hdiagnostics/diagnostics4.html

[70] Alaton, P., Djehiche, B. and Stillberger, D. (n.d.). On Modeling and Pricing Weather Derivatives. (citeseerx.ist.psu.edu)

Available at:
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.198.6547&rep=rep1&type=pdf

[71] Eex.com. (2019). EEX Homepage. [online] Available at: https://www.eex.com/en#/en.

[72] Contract Specifications. (2019). Nasdaq Oslo ASA and Nasdaq Clearing AB.

Available at: https://business.nasdaq.com/media/181001%20Joint%20Appendix%202%20-%20Contract%20Specifications%20mark%20up_tcm5044-65879.pdf

[73] Gersema, G. and Wozabal, D. (2017). An equilibrium pricing model for wind power futures. Energy Economics, pp.64-74.

[74] Theweatherprediction.com. (2019). WIND SPEED INCREASING WITH HEIGHT. [online] Available at: http://www.theweatherprediction.com/habyhints3/749/.

[75] Watson, D. (2015). Wind Turbine Power Coefficient (Cp). [online] Ftexploring.com.

Available at: http://www.ftexploring.com/wind-energy/wind-power-coefficient.htm

[76] EEX Wind Power Future. (2017). eex group. Document is requested via e-mail. [email]. 2018.

E-mail address: Aarti Dalal (Sales Manager - Sales Power & Environmental Markets - London Office < Aarti.Dalal@eex.com>

[77] Cmegroup.com. (2019). Snowfall Futures and Options. [online]

Available at: https://www.cmegroup.com/trading/weather/files/snowfall-fact-card.pdf.

[78] Luo, L. (2010). Snowfall derivative pricing: Index and daily modeling for the snowfall futures. Master. Concordia University.

[79] CHEN, J. (2018). Weather Future. [online] Investopedia. Available at: https://www.investopedia.com/terms/w/weatherfuture.asp.

[80] Thornes, J. (2003). An introduction to weather and climate derivatives. Weather, 58(5), pp.193-196.

[81] Jewson, S. and Caballero, R. (2003). The use of weather forecasts in the pricing of weather derivatives. Meteorological Applications, 10(4), pp.377-389.

[82] Cmegroup.com. (n.d.). Seasonal Rainfall Index Futures Final Settlement Procedure. [online] Available at: https://www.cmegroup.com/trading/weather/files/Seasonal-Rainfall-Index-Futures-Final-Settlement-Procedure.pdf.

[83] Turner, Barry PhD. Weather Derivatives and Short-Period Rainfall Indices. Dept. of Atmospheric and Oceanic Sciences, McGill University

[84] Cmegroup.com. (n.d.). CME Hurricane Index Futures and Options. [online]

Available at: https://www.cmegroup.com/trading/weather/files/WT106_NEWHurricaneFC.pdf.

[85] Bartkowiak, M. (2009). Weather derivatives. [online] Research Gate.

Available at: https://www.researchgate.net/publication/275532497_Weather_derivatives.

[86] Eex.com. (2016). EEX: Trading in Wind Power Futures to begin in early October. [online]

Available at: https://www.eex.com/en/about/newsroom/news-detail/eex--trading-in-wind-power-futures-to-begin-in-early-october/56354.

[87] Eex.com. (n.d.). EEX Group. [online] Available at: https://www.eex.com/en/about/eex/eex-ag.

[88] Magris, M. (n.d.). COS method (Fang, Osterlee). [online] Mathworks.

Available at: https://www.mathworks.com/matlabcentral/fileexchange/67944-cos-method-fang-osterlee.

[89] Hamisultane, H. (2008). Which Method for Pricing Weather Derivatives ?. [online] Halshs.archives-ouvertes.fr. Available at: https://halshs.archives-ouvertes.fr/halshs-00355856.

[90] doc. Ing. Julius Bems, Ph.D.,Financial Management Lecture Notes. (2018).

[91] The Economic Times. (2016). Monte Carlo Simulation. [online]

Available at: https://economictimes.indiatimes.com/m/mcc/articleshow/52187686.cms

[92] Monte Carlo Simulation: What Is It and How Does It Work?. (n.d.). What is Monte Carlo Simulation?. [online]

Available at: https://www.palisade.com/risk/monte_carlo_simulation.asp.

[93] DEMIROGEN, O. (1998). Talep Tahmininde Monte Carlo Simulasyon Tekniginin Kullanilmasi. Ataturk Universitesi IIBF Dergisi, 12(1-2), pp.229-240.

[94] Statistics How To. (2019). F Test: Simple Definition, Step by Step Examples -- Run by Hand / Excel. [online]

Available at: https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/hypothesis-testing/f-test/.

[95] University of Pennsylvania Ling.upenn.edu. (2008). Tutorial: Pearson's Chi-square Test for Independence. [online] Available at: https://www.ling.upenn.edu/~clight/chisquared.htm.

[96] Statistics Solutions. (2019). Chi-Square Goodness of Fit Test - Statistics Solutions. [online] Available at: https://www.statisticssolutions.com/chi-square-goodness-of-fit-test/.

[97] Stattrek.com. (n.d.). Chi-Square Goodness of Fit Test. [online] Available at: https://stattrek.com/chi-square-test/goodness-of-fit.aspx.

[98] Akgül, F., Şenoğlu, B. and Arslan, T. (2016). An alternative distribution to Weibull for modeling the wind speed data: Inverse Weibull distribution. Energy Conversion and Management, 114, pp.234-240.

[99] AKYUZ, H. (2017). Statistical Analysis of Wind Speed Data with Weibull, Lognormal and Gamma Distributions. Cumhuriyet Science Journal, 38-4, pp.68-76.

[100] Igissenov, A. (2017). Designing parallel working wind power station in South Kazakhstan. MSc. Czech Technical University in Prague.

[101] Support.minitab.com. (n.d.). What is log-likelihood? - Minitab. [online]

Available at: https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling- statistics /regression /supporting-topics/regression-models/what-is-log-likelihood/.

[102] KURBAN, M., HOCAOĞLU, F. and MERT KANTAR, Y. (n.d.). WEIBULL DAĞILIMI KULLANILARAK RÜZGAR HIZ VE GÜÇ YOĞUNLUKLARININ İSTATİSTİKSEL ANALİZİ. Afyon Kocatepe University, JOURNAL OF SCIENCE, [online] pp.205-218.

[103] Masters, G. (n.d.). Renewable and Efficient Electric Power Systems. Stanford University: A JOHN WILEY & SONS, INC., PUBLICATION, pp.312-352.

Available at: https://fenbildergi.aku.edu.tr/pdf/0702/7-2(213-226).pdf.

[104] Gribble, S. (n.d.). Air density calculator (useful for estimating rho). [online] Gribble.org.

Available at: https://www.gribble.org/cycling/air_density.html.

[105] Mide Technology. (n.d.). Air Pressure at Altitude Calculator. [online]

Available at: https://www.mide.com/pages/air-pressure-at-altitude-calculator.

[106] Data, U. (n.d.). Climate New York - New York and Weather averages New York. [online] Usclimatedata.com.

Available at: https://www.usclimatedata.com/climate/new-york/new-york/united-states/usny0996.

[107] En.wind-turbine-models.com. (2019). General Electric GE 2.5-103 - 2,50 MW - Wind turbine. [online] Available at: https://en.wind-turbine-models.com/turbines/1293-general-electric-ge-2.5-103.

[108] Cmegroup.com. (2019). US Monthly Weather Heating Degree Day (HDD) Contract Specs - CME Group. [online] Available at: https://www.cmegroup.com/trading/weather/temperature/us-monthly-weather-heating_contractSpecs_options.html.

[109] Hoornaert, A. (2018). Hedging Weather Risk Using Weather Derivatives. Master. Ghent University.

[110] Bloomberg.com. (2019). Bloomberg - Markets United States Rates & Bonds . [online]

Available at: https://www.bloomberg.com/markets/rates-bonds/government-bonds/us [Accessed 12 Apr. 2019].

[111] Investopedia. (2019). Derivative. [online]

Available at: https://www.investopedia.com/terms/d/derivative.asp.

[112] ADIGÜZEL Hümeyra PhD., A. and YILMAZ Ayşe Altıok PhD, A. (2015). ACCOUNTING FOR DERIVATIVES AS AN HEDGING INSTRUMENT UNDER INTERNATIONAL ACCOUNTING STANDARDS. MALI COZUM, pp.15-31.

[113] Windexchange.energy.gov. (2019). WINDExchange: New York 80-Meter Wind Resource Map. [online] Available at: https://windexchange.energy.gov/maps-data/95.

[114] Statistics How To. (2016). ADF - Augmented Dickey Fuller Test - Statistics How To. [online] Available at: https://www.statisticshowto.datasciencecentral.com/adf-augmented-dickey-fuller-test/.

[115] GmbH, e. (2019). Renewable energy policy database and support: single. [online] Res-legal.eu. Available at: http://www.res-legal.eu/search-by-country/germany/single/s/res-e/t/promotion/aid/feed-in-tariff-eeg-feed-in-tariff/lastp/135/.

[116] Itl.nist.gov. (2019). 1.3.5.11. Measures of Skewness and Kurtosis. [online]

Available at: https://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm.

[117] Investopedia. (2019). Risk and Diversification: Diversifying Your Portfolio. [online]

Available at: https://www.investopedia.com/university/risk/risk4.asp.

[118] Investopedia. (2019). Continuous Compounding. [online]

Available at: https://www.investopedia.com/terms/c/continuouscompounding.asp.

[119] Itl.nist.gov. (n.d.). 1.3.6.7.4. Critical Values of the Chi-Square Distribution. [online] Available at: https://www.itl.nist.gov/div898/handbook/eda/section3/eda3674.htm.

[120] GANTI, A. (2019). Implied Volatility (IV). [online] Investopedia.

Available at: https://www.investopedia.com/terms/i/iv.asp.

[121] CHEN, J. (2019). Extrinsic Value Definition. [online] Investopedia.

Available at: https://www.investopedia.com/terms/e/extrinsicvalue.asp.

[122] Optionstrading.org. (2019). Price of Options - Extrinsic & Intrinsic Value. [online] Available at: http://www.optionstrading.org/introduction/how-options-really-work/price.

[123] SEGAL, T. (2019). Understanding Time Value. [online] Investopedia. Available at: https://www.investopedia.com/terms/t/timevalue.asp.

Appendices

Appendix 1: Current Available Cities for Trading CME Weather Derivatives

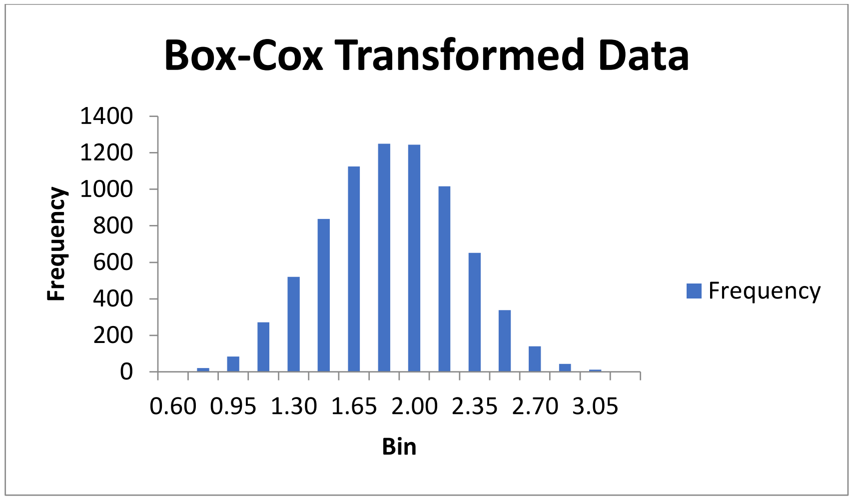
(based on data from [6])

Countries&Cities	US	Canada	Europe	Japan	Australia
1	Atlanta	Calgary	London	Tokyo	Bankstown, Sydney
2	Chicago	Edmonton	Paris	Osaka	Brisbane Aero
3	Cincinnati	Montreal	Amsterdam	Hiroshima	Melbourne Regional
4	New York	Toronto	Berlin		
5	Dallas	Vancouver	Essen		
6	Philadelphia	Winnipeg	Stockholm		
7	Portland		Barcelona		
8	Tucson		Rome		
9	Des Moines		Madrid		
10	Las Vegas		Oslo-Blindern		
11	Detroit		Prague		
12	Minneapolis				
13	Houston				
14	Sacramento				
15	Salt Lake City				
16	Baltimore				
17	Boston				
18	Kansas City				
19	Colorado Springs				
20	Jacksonville				
21	Little Rock				
22	Los Angeles				
23	Raleigh Durham				
24	Washington D.C				

Appendix 2: Bin – Frequency Tables and Histogram Graphs of NY's Wind Speed Data

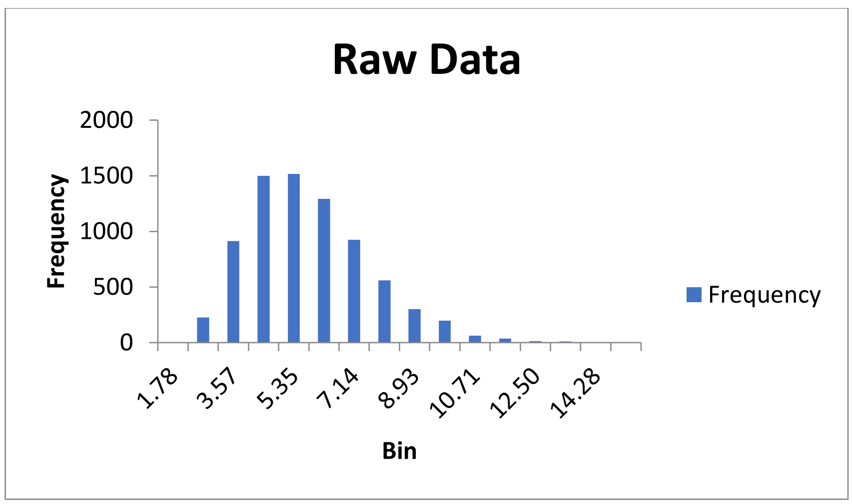
Appendix 2.1. Bin – Frequency Tables and Histogram Graphs of NY's Box-Cox Transformed Wind Speed Data

Bin	Frequency
0.60	1
0.77	20
0.95	84
1.12	271
1.30	520
1.47	837
1.65	1125
1.83	1250
2.00	1244
2.18	1017
2.35	652
2.53	338
2.70	140
2.88	44
3.05	11
More	1



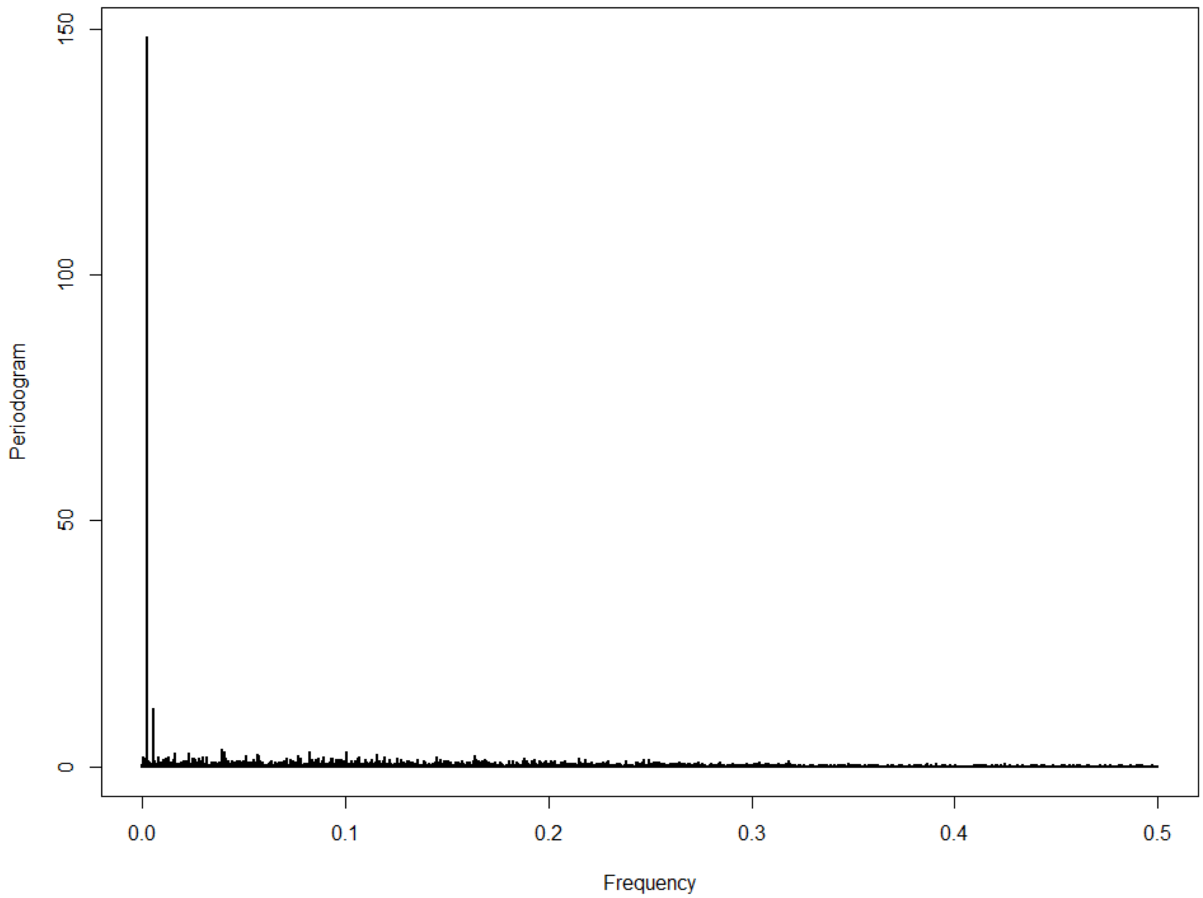
Appendix 2.2. Bin – Frequency Tables and Histogram Graphs of NY's Raw Wind Speed Data

Bin	Frequency
1.78	1
2.68	226
3.57	912
4.46	1499
5.35	1516
6.25	1292
7.14	924
8.03	559
8.93	302
9.82	198
10.71	63
11.60	37
12.50	14
13.39	10
14.28	2
More	0



Appendix 3: Periodogram of Box-Cox Transformed Wind Speed Data and R Language Code for Calculating Seasonality Period

Appendix 3.1. Periodogram of Box-Cox Transformed Wind Speed Data

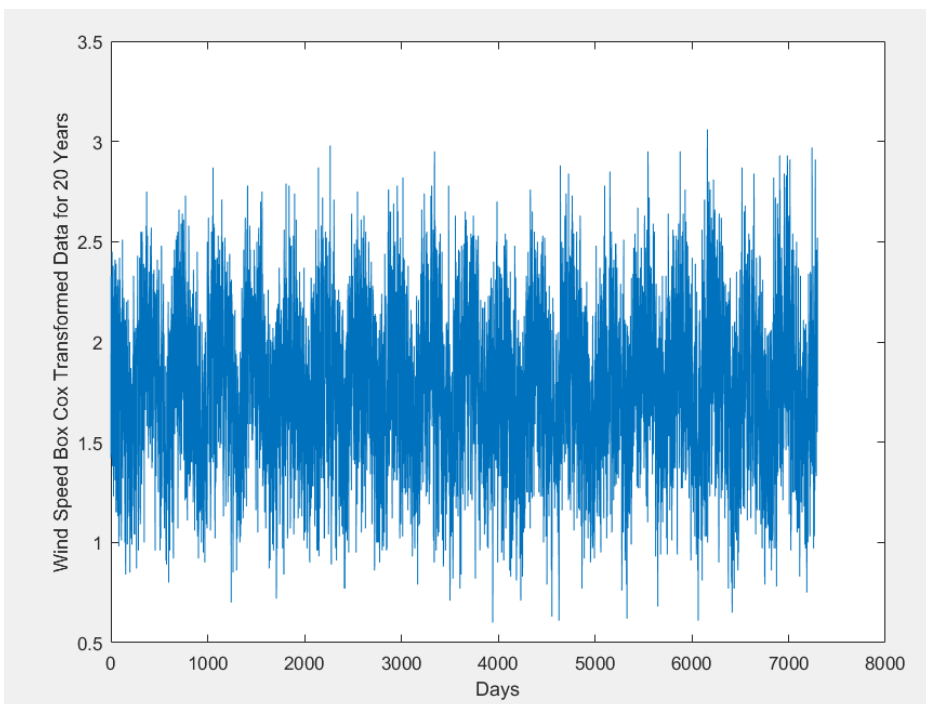


Appendix 3.2. R Language Code for Calculating Seasonality Period

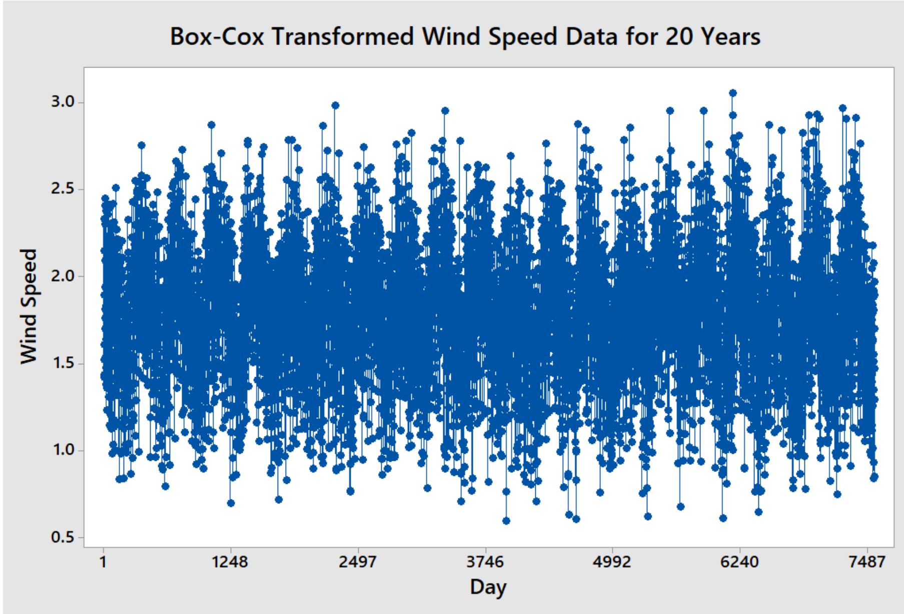
```
> WindSpeedData <- read_excel("C:/Users/I355824/OneDrive - SAP SE/Desktop/cv
ut master/3rd semester/Individual Project/tez programlari/rstudio sonuclariu
/WindSpeedData.xlsx")
> View(WindSpeedData)
> install.packages("TSA")
> library (TSA)
> seasonality=periodogram(WindSpeedData$`Box-Coxed Transformed Wind Speed`)
> seasonalits=data.frame(freq=seasonality$freq,spec=seasonality$spec)
> order=seasonalits[order(-seasonalits$spec),]
> frequency=head(order,3)
> frequency
           freq       spec
21  0,002734375 148,279080
42  0,005468750  11,718080
302 0,039322917   3,610572
> # display the 3 highest "power" frequencies
> # convert frequency to time periods
> period=1/frequency$f
> period
[1] 365,71429 182,85714 25,43046
```

Appendix 4: Box-Cox Transformed Daily Average Wind Speed Data for 20 Years

Appendix 4.1. Box-Cox Transformed Daily Average Wind Speed Data for 20 Years(MATLAB)



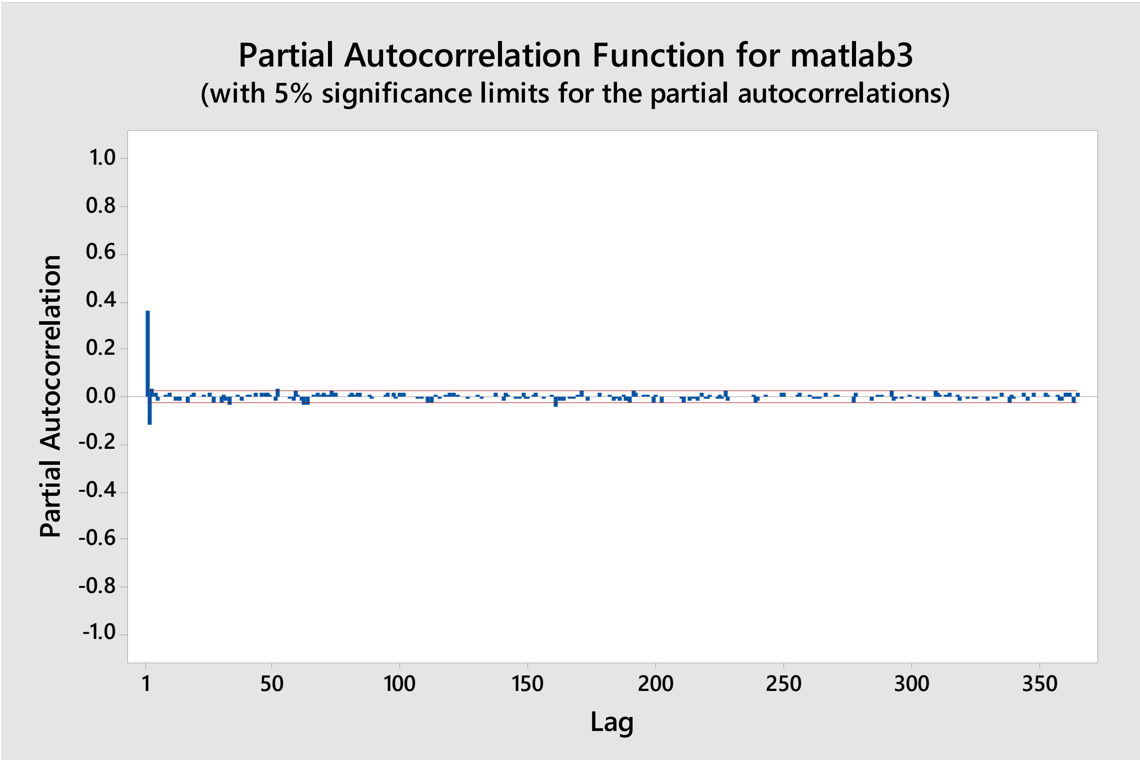
Appendix 4.2. Box-Cox Transformed Daily Average Wind Speed Data for 20 Years(Minitab)



Appendix 5: Decomposing Seasonality with MATLAB

```
%seasonality formula and removing it
>> originaldata=importdata('boxcoxdata3.txt');
>> llenght=transpose (1:7300);
>> seasonality=fit (length, originaldata,'fourier2');
>> mevsin=feval (seasonality,llength);
>> deseasondata=originaldata-mevsim;
```

Appendix 6: Partial Autocorrelation Function for Deseasonalized Wind Speed – 365 Lags



Appendix 7: RStudio Code and p-value result for ADF(Stationary) Test and Gretl Results

Appendix 7.1. RStudio Code and p-value result for ADF(Stationary) Test – Lag

adf.test(d, alternative = c("stationary", "explosive"),k = 365)

Augmented Dickey-Fuller Test

Dickey-Fuller = -4,2623, Lag order = 365, p-value = 0,01

alternative hypothesis: stationary

Warning message:

In adf.test(d, alternative = c("stationary", "explosive"), k = 365) : **p-value smaller than printed p-value**

Appendix 7.2. RStudio Code and p-value result for ADF(Stationary) Test – Lag Order=3

adf.test(d, alternative = c("stationary", "explosive"),k = 3)

Augmented Dickey-Fuller Test

Dickey-Fuller = -38,499, Lag order = 3, p-value = 0,01

alternative hypothesis: stationary

Warning message:

In adf.test(d, alternative = c("stationary", "explosive"), k = 3) : **p-value smaller than printed p-value**

Appendix 7.3. Gretl Results of ADF(Stationary) Test – Lag Order=3

```
Augmented Dickey-Fuller test for v1
testing down from 19 lags, criterion AIC
sample size 7300
unit-root null hypothesis: a = 1

  test without constant
  including 3 lags of (1-L)v1
  model: (1-L)y = (a-1)*y(-1) + ... + e
  estimated value of (a - 1): -0,672471
  test statistic: tau_nc(1) = -38,5032
  asymptotic p-value 2,326e-013s
  1st-order autocorrelation coeff. for e: 0,000
  lagged differences: F(3, 7292) = 37,708 [0,0000]
```

Appendix 8: Akaike and Bayesian Information Criteria Results from Gretl

Appendix 8.1. AR (3) Model – Including Constant Value

Model 5: ARMA, using observations 1960-01-01:1979-12-26 (T = 7300)
Dependent variable: v1

Mean dependent var	−1,18e-15	S.D. dependent var	0,363788
Mean of innovations	0,000017	S.D. of innovations	0,336220
Log-likelihood	−2.401,413	Akaike criterion	4.812,826
Schwarz criterion	4.847,304	Hannan-Quinn	4.824,681

Appendix 8.2. ARMA (3,3) Model

Model 3: ARMA, using observations 1960-01-01:1979-12-26 (T = 7300)
Dependent variable: v1

Mean dependent var	−1,18e-15	S.D. dependent var	0,363788
Mean of innovations	−3,31e-06	S.D. of innovations	0336123
Log-likelihood	−2.399,315	Akaike criterion	4.812,630
Schwarz criterion	4.860,899	Hannan-Quinn	4.829,228

Appendix 8.3. AR (5) Model

Model 11: ARMA, using observations 1987-01-01:2006-12-26 (T = 7300)
Dependent variable: Regmodel
Standard errors based on Hessian

Mean dependent var	−1,18e-15	S.D. dependent var	0,363788
Mean of innovations	−3,32e-06	S.D. of innovations	0,336124
Log-likelihood	−2.399,327	Akaike criterion	4.810,655
Schwarz criterion	4.852,029	Hannan-Quinn	4.824,882

Appendix 8.4. ARMA (5, 5) Model

Model 8: ARMA, using observations 1960-01-01:1979-12-26 (T = 7300)
Dependent variable: v1

Mean dependent var	−1,18e-15	S.D. dependent var	0,363788
Mean of innovations	−5,45e-06	S.D. of innovations	0,335883
Log-likelihood	−2.394,163	Akaike criterion	4810,327
Schwarz criterion	4.886,179	Hannan-Quinn	4836,409

Appendix 9: AR Models Generated by Minitab and Gretl

Appendix 9.1. Minitab AR Model without Constant

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.4112	0.0117	35.15	0.000
AR 2	-0.1315	0.0126	-10.47	0.000
AR 3	0.0348	0.0117	2.97	0.003

Number of observations: 7300

Appendix 9.2. Minitab AR Model with Constant

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.4112	0.0117	35.14	0.000
AR 2	-0.1315	0.0126	-10.47	0.000
AR 3	0.0348	0.0117	2.97	0.003
Constant	-0.00002	0.00394	-0.01	0.996
Mean	-0.00003	0.00574		

Since the mean value of the data and the constant value of the seasonality function is equal and also since the seasonality is subtracted, it is normal to have the mean value (constant) for the AR model as zero. Adding or removing the constant value while generating the AR model do not affect the coefficients of the AR model.

Appendix 9.3. Gretl AR Model

Model 4: ARMA, using observations 1987-01-01:2006-12-26 (T = 7300)
Dependent variable: v1
Standard errors based on Hessian

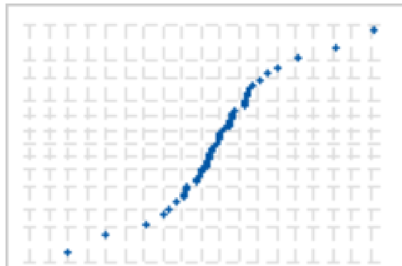
	Coefficient	*Std. Error*	*z*	*p-value*	
phi_1	0,411139	0,0116977	35,15	<0,0001	***
phi_2	−0,131494	0,0125552	−10,47	<0,0001	***
phi_3	0,0347493	0,0117002	2,970	0,0030	***

Mean dependent var	−1,18e-15	S.D. dependent var	0,363788
Mean of innovations	−4,37e-06	S.D. of innovations	0,336220
Log-likelihood	−2.401,413	Akaike criterion	4.810,826
Schwarz criterion	4.838,408	Hannan-Quinn	4.820,310

Appendix 10: Interpretation of Normal Probability Plot for Residuals

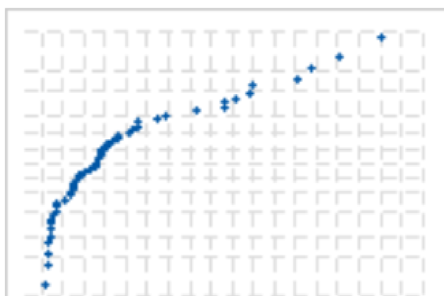
Appendix 10 demonstrates different normal probability plots for different situations [58].

Appendix 10.1. Distribution with Long Tails



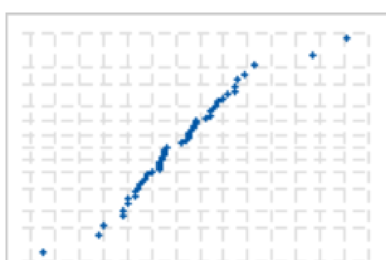
S- Curve demonstrates that the distribution has long tails.

Appendix 10.2. Distribution with Right Skewness



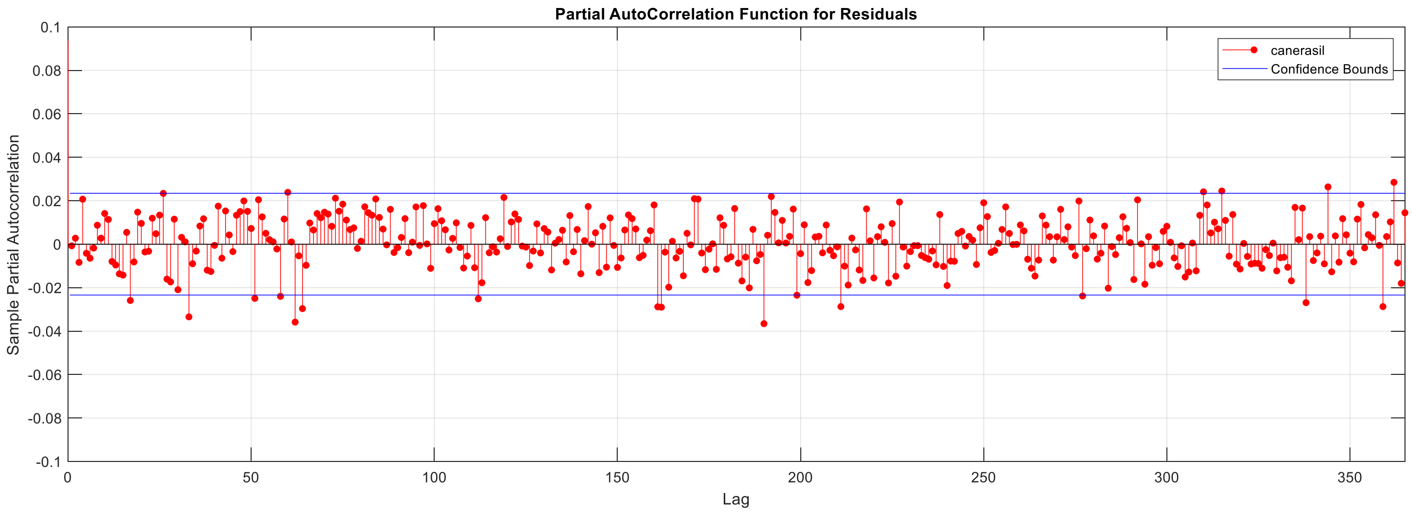
Downward Curve demonstrates that the distribution is a right skewed distribution.

Appendix 10.3. Distribution with Outliers



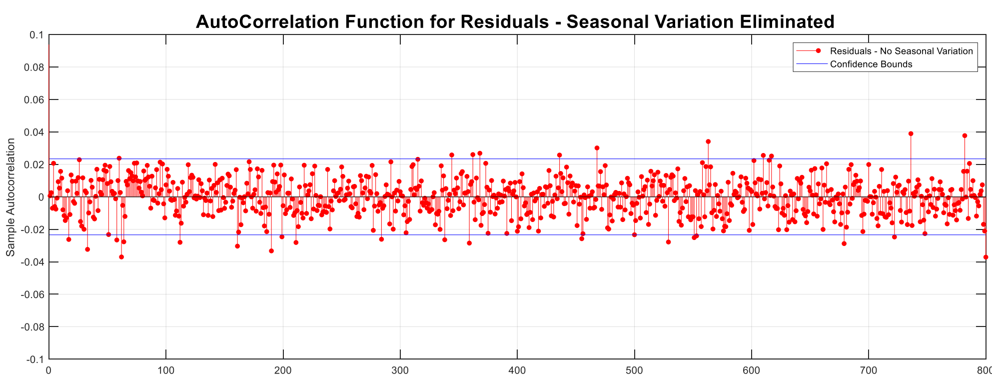
The points standing away from the intensive part of the line demonstrates that there are some outliers at the relevant distribution.

Appendix 11: PACF for Residuals of AR (3)



Since it was already seen from ACF of the residuals that there is no significant autocorrelation between residuals, it is normal to have same result for PACF of the residuals. At any lag, there is no autocorrelation between the residuals.

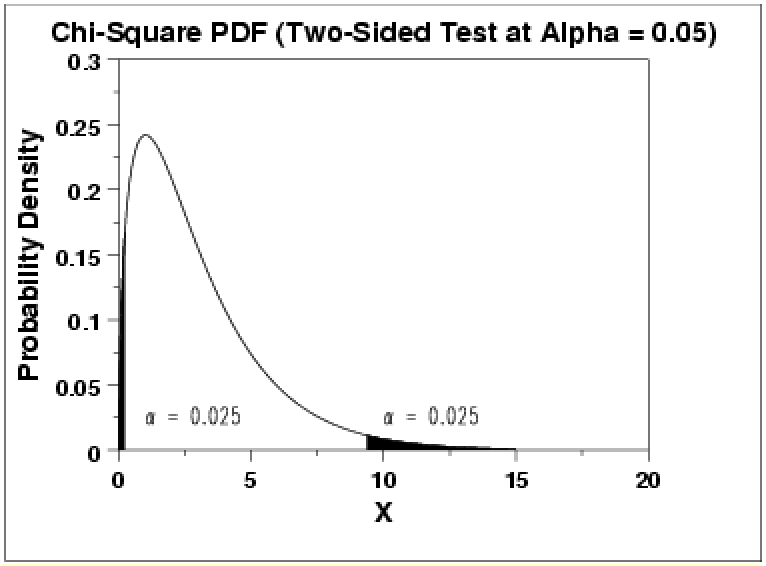
Appendix 12: ACF Plot for Residuals after Removing the Seasonal Variation



After removing the seasonal variation from the residuals, there is still not significant autocorrelation between any lags of the residuals.

Appendix 13: Chi-Square Distribution Probability Density Function and Chi-Square Values for Weekly Data Comparison (MC Simulated and Real Data)

Appendix 13.1. Chi-Square Distribution Probability Density Function [119]

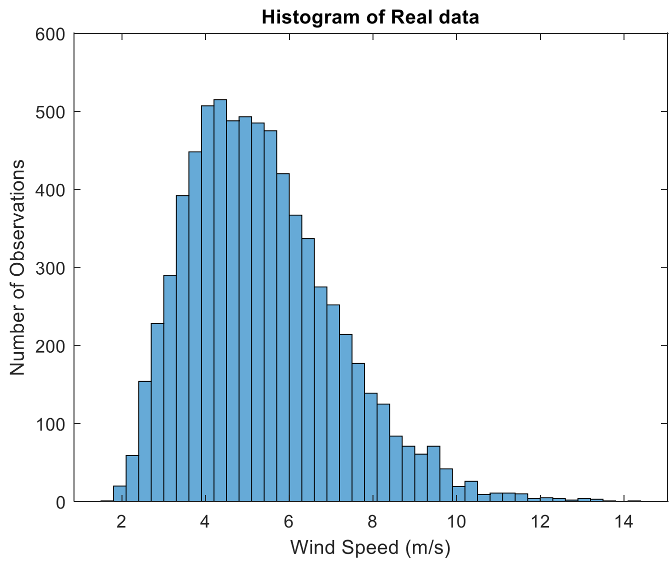


Appendix 13.2. Chi-Square Values for Weekly Data Comparison (MC Simulated and Real Data)

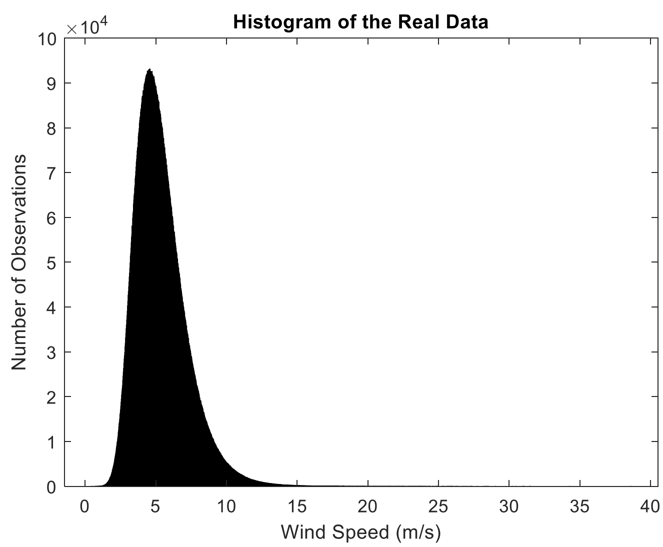
Week No	1	2	3	4	5	6
Chi-Square Value	8,7266	2,1226	8,5112	15,9248	2,7315	3,2487
Week No	7	8	9	10	11	12
Chi-Square Value	9,8194	4,5413	12,1014	16,5134	11,4005	7,8174
Week No	13	14	15	16	17	18
Chi-Square Value	11,9426	4,2478	5,5058	20,3476	10,8656	15,3140
Week No	19	20	21	22	23	24
Chi-Square Value	5,8092	8,6840	1,1033	9,5029	17,4241	13,9109
Week No	25	26	27	28	29	30
Chi-Square Value	3,0381	10,7922	13,9894	24,1910	5,4680	18,1677
Week No	31	32	33	34	35	36
Chi-Square Value	3,3188	6,0876	11,1567	7,3608	5,8047	22,7572
Week No	37	38	39	40	41	42
Chi-Square Value	13,3481	5,8030	11,1575	17,3406	14,0091	7,0853
Week No	43	44	45	46	47	48
Chi-Square Value	6,0862	10,9338	7,2461	8,5973	6,5394	11,4469
Week No	49	50	51	52		
Chi-Square Value	3,6000	31,7490	7,9670	4,1910		

Appendix 14: Histogram of MC Simulated and Real Data

Appendix 14.1. Histogram of Real Data



Appendix 14.2. Histogram of MC Simulated Data



As it can be seen both from the histograms of the MC Simulated and Real Data, the shape and the mean value is similar. It is already mentioned that the variance if the annual data for MC simulated data is less than the real data's.

Appendix 15: Log-Likelihood Values for Distribution Fits with Real and MC Simulated Data

	Real	MC
Rayleigh	-1,56640E+04	-3,92593E+07
Weibull	-1,45680E+04	-3,68434E+07
Log-Normal	-1,41670E+04	-3,53512E+07
Gamma	-1,41850E+04	-3,55193E+07

Appendix 16: Power Coefficient (Cp) for Selected Wind Turbine Based on Upstream Wind Speed

Wind Speed (m/s)	Cp
3,5	0,25
4,5	0,36
5,5	0,41
6,5	0,43
7,5	0,44
8,5	0,45
9,5	0,43
10,5	0,39
11,5	0,32
12,5	0,25
13,5	0,2

As it can be seen from the table, this wind turbine has its one of the highest Cp-value at NY's average wind speed; 5,37 m/s