Corvinus University of Budapest
Doctoral School of Economics, Business and Informatics

# Artificial Intelligence Forecasting Techniques For Reducing Uncertainties In Renewable Energy Applications

Ph.D. Dissertation

Business Informatics

Supervisors: Prof. Csaba Csaki, Ph.D.

Mutaz AlShafeey

Budapest, 2022

Mutaz Abdelrazzaq Saleh AlShafeey

Corvinus University of Budapest
Department of Business Informatics

Supervisors: Prof. Csaba Csaki, Ph.D.

## CONTENTS

## TABLE OF FIGURES

## LIST OF TABLES

## DATA SHARING POLICY

All renewable energy production datasets used in this thesis were obtained from renewable energy farms, 3Comm Hungary (3comm.hu/) and E.on Hungary (eon.hu/), while weather data was collected from Solcast (http://solcast.com) which specialized in weather data modeling. As well, the data collection was conducted throughout the lifetime of the project. The pre-analysis data should be requested directly from the mentioned sources.

For clarifying the types of data, appendices containing all of the variable lists that were used in this thesis are provided. All data specifications including the size, time frame, format, number of files, data dictionary, and the codebook are documented and can be provided upon agreed request. Additionally, the newly created variables from the models and analyses are updated to the data specification.

The post-analysis data may be useful for researchers who plan to conduct studies in topics related to energy forecasting. However, normally new users are required to gain permission for data use from the data sources.

The pre-analyses data is stored in CSV format. The post-analysis data is stored in CSV and Matlab (M file) format. If requested, other data formats, including CSV, Excel, R, and SPSS can be transformed.

Data sharing will require two steps of permission. 1) data use agreement from the data sources mentioned above for pre-analysis data use, and 2) data use agreement from the department of business informatics (in Corvinus University of Budapest) for post-analysis data use. Preparation for data sharing will begin after the completion of the thesis.

## PERSONAL DATA AND GENDER DISCLAIMER

This thesis deals with artificial intelligence forecasting systems. None of the required datasets include any human-related data like gender, age, ethnicity, or any other personal data. As a result, the biological variable of sex is not applicable in this research. Even though the data itself doesn't have personal or gender information, the artificial intelligence biased is considered while designing the proposed system. The gender aspects were considered while analyzing, interpreting, and disseminating the findings.

As developing and designing artificial intelligence forecasting systems is one of the main outcomes of this research, the focus on gender bias-free systems is crucial. In this research the bias-free system is achieved through:

1. Utilizing weather and past energy data (no human or gender indication data was used).
2. Consider and report gender awareness whenever it is necessary such as in energy policy to ensure gender-equitable power.

# I. INTRODUCTION

## I.1. General overview

One of the most distinguishing features of this era is the rapid technological development and the changes accompanying it, including the increasing demand for natural resources. These changes have led to a steady increase in demand for electric power as well. As a result, a rapid continuous increase in energy demand was observed in the past 50 years. Also, this demand is expected to grow further in the next 50 years. The International Energy Agency (IEA) World Energy Outlook 2010 and 2021 state that the average annual increase of the total primary energy demand lies between 1 and 2% (Mandil, 2004) (Agency, 2021) as can be seen in Table 1.

The need to increase energy generating capacity is a critical economic issue due to the financial burdens associated with increasing energy demand. With the high increase in energy demand expected in the near and far future, looking for available resources to fulfill the future demand is crucial, especially for the electricity generation sector.

**Table 1 World Total Energy Supply in exajoule (EJ)**[1]

| Energy Source/Type | Stated policies scenario | | | | | | Compound average annual growth rate to 2020 (%) | |
|---|---|---|---|---|---|---|---|---|
| | 2010 | 2019 | 2020 | 2030 | 2040 | 2050 | 2030 | 2050 |
| Renewables | 47.7 | 65.8 | 68.5 | 109.0 | 153.0 | 192.5 | 4.8 | 3.5 |
| Oil | 172.1 | 187.9 | 171.4 | 198.5 | 199.6 | 198.3 | 1.5 | 0.5 |
| Natural Gas | 115.1 | 141.4 | 138.7 | 155.9 | 168.0 | 174.0 | 1.2 | 0.8 |
| Nuclear | 30.1 | 30.5 | 29.4 | 34.0 | 38.4 | 40.5 | 1.5 | 1.1 |
| Coal | 153.0 | 162.2 | 155.8 | 150.2 | 132.9 | 116.8 | -0.4 | -1.0 |
| Total | 544.7 | 613.0 | 589.1 | 671.0 | 714.8 | 743.9 | 1.3 | 0.8 |

Traditional energy generation methods which rely on fossil fuels (coal, oil, natural gas) and their derivatives produce air, noise, and environmental pollution (Ashi et al., 2014). Such pollution has negative impacts on humans and nature such as climate change, greenhouse effect, and deforestation (Nelson & Starcher, 2015). Moreover, the depletion of fossil fuels and inequality in the distribution of energy consumption and reserves causes serious problems for current energy systems (Perea-Moreno et al., 2018). Thus far, among all the available resources to produce power, oil is the most widely used method, followed by coal and natural gas as can be seen in Table 2 (Goswami & Kreith, 2015) (Agency, 2021). Therefore, generating energy from green sustainable resources becomes an imperative necessity on the long

[1] Based on (Mandil, 2004) and (Agency, 2021)

timescale to supply the exponentially growing demand and reduce greenhouse gas emissions (Nelson & Starcher, 2015).

**Table 2 Fuel Shares in World Total Primary Energy Supply [2]**

| Source | Source Share (%) per year | | |
|---|---|---|---|
| | 2020 | 2030 | 2050 |
| Oil | 29 | 30 | 27 |
| Coal | 26 | 22 | 16 |
| Natural gas | 24 | 23 | 23 |
| Renewables | 12 | 16 | 26 |
| Nuclear | 5 | 5 | 5 |
| Others | 4 | 4 | 3 |

One of the promising green resources is renewable energy (Alshafeey & Csáki, 2019). Renewable energy is a very wide term that includes a broad spectrum of different resources. The main common characteristic between all renewable resources is being "self-renewing" (Bull, 2001).

Renewable resources such as solar, wind, tidal wave, or biomass can offer a reliable solution for the energy demand problem (Almutairi et al., 2021). Green technologies such as solar and wind are among the main sustainable technologies that may offer competitive advantages and their use has been lately accelerated (Prasad et al., 2021). Since photovoltaic (PV) equipment can be easily installed almost everywhere and operates efficiently in different geographical regions with low maintenance required, solar energy is considered to be an effective environmentally friendly technology for energy production (Alshafeey & Csáki, 2019). Another

[2] Based on (Goswami & Kreith, 2015) and (Agency, 2021)

growing trend in renewable energy generation is wind resources. Wind technologies offer reliable, eco-friendly, simple, and low-maintenance methods for energy generation (Grigsby, 2018).

The trend of generating energy from renewable sources, especially the energy generated from solar and wind resources has received wide approval due to its advantages such as the ease of generation and its availability in most geographical locations. Furthermore, renewable resources can complement each other, taken together can contribute to energy security by reducing foreign energy dependency (Bull, 2001).

Irrespective of all the advantages of utilizing renewable technology for energy production, there are some hindrances limiting growth and wider utilization. One of the crucial drawbacks of renewable solutions is low energy converting efficiency or the density problem (Huang et al., 2013) (Nelson & Starcher, 2015). Even though sources like sun and wind have the potential to supply the whole earth's energy demand (Bull, 2001), however, the current energy conversion efficiency is relatively low and energy harvesting technologies need more improvements.

Renewable technologies have also some environmental issues to be considered. While conventional fossil fuel power plants produce different types of pollution (air, noise, waste products,…etc.) perceived issues like odor from biomass, avian mortality from wind turbines, brine from geothermal energy, and visual pollution from solar panels must be considered when applying renewable solutions (Nelson & Starcher, 2015). So the common problem between renewable, nuclear, and conventional power plants is the “not in my backyard” issue.

Another major drawback of utilizing renewable resources is variability. Variability problems in some renewable resources come from the reliance of some renewable energy resources on the weather variables for producing energy. For instance, solar cells rely mainly on solar radiation to produce energy, while wind turbines rely mainly on wind speed. The nature of weather variables -which unfortunately highly fluctuates over time- leads to generation uncertainty (Alshafeey

& Csáki, 2019). Fluctuation and uncertainty in energy production lead to uncertainty in economic benefits. Calculating economic indicators such as energy pricing, rate of return, and payback period is challenging under generation uncertainty. Additionally, uncertainty can affect grid stability in case of grid-connected PV farms (Alshafeey & Csáki, 2019).

As can be concluded, the energy produced by renewable resources like solar and wind depends on many factors. Among the most vital factors are solar radiation, temperature, wind speed, humidity, and the conversion efficiency of technology. Thus, to control the amount of the potential amount of renewable energy, these factors must be studied and optimized.

For large applications such as grid-connected renewable energy farms, any small fluctuation in any variable might highly affect the amount of generated energy. Fluctuating energy production has serious consequences. These consequences might be either economic, such as the inability to calculate energy pricing, rate of return, and other economic elements; or technical, such as under or over generation of energy, and more importantly, fluctuation may lead to grid instability.

To achieve renewable energy generation stability there are many solutions that have been used to overcome the above problems. For instance, energy storage units can be used to stabilize power. Storage units act as a buffer by supplying load when there is a shortage and storing energy when there is excess generation. Yet, using storage units is impractical for large applications. In addition, storage units like batteries have limited usage cycles and it has to be replaced after a certain time adding extra costs (Koohi-Fayegh & Rosen, 2020) (Wang et al., 2012).

Other possible solutions include using hybrid systems like solar-diesel, wind-diesel, or solar-wind hybrid systems. Diesel hybrid systems are a well-known solution, especially in remote locations. The diesel generator can provide energy whenever there is a supply shortage from renewable resources. But this solution is not always economical and may be impractical for inter-cities applications. Moreover, hybrid Diesel systems show a bad performance for diesel generators (Yamegueu et al., 2011) (Cavalcante et al., 2021).

One promising solution for renewable energy generation stability is to enhance renewable energy forecasting (Singh, 2013) (Devaraj et al., 2021). If the potential renewable energy can be accurately predicted with lower uncertainty, renewable energy systems can be better designed and optimized helping grid operators in managing power supply and demand (Pazikadin et al., 2020). Accurate forecasts would improve grid stability as well (Rodríguez et al., 2018). Additionally, accurate energy forecasting does not only provide value through reduced imbalance penalties (incurred due to the difference between the scheduled and actually delivered energy) but also leads to increased competitiveness by providing advanced knowledge in real-time energy market trading.

Improving renewable energy forecasting and creating accurate forecasting models are among the most important aspects of renewable energy production and are considered to be one of the 'hottest' topics in the renewable energy research field.

Renewable energy forecasting models are software solutions that can be used to forecast the future values of renewable energy generation. Like any system that predicts the future, the forecasted value of energy would have a degree of uncertainty and errors. A good forecasting model can predict future values with minimum errors and uncertainties (Cammarano et al., 2012).

Many approaches can be used to forecast renewable energy generation such as physical modeling, statistical modeling, artificial intelligence techniques, and their hybrids which have frequently been employed (Wang et al., 2018) (H. Wang et al., 2019). Each method has its own pros and cons. As the physical method is mainly concerned with generating forecasts based on atmospheric variables, this method is often considered computationally expensive (Sweeney et al., 2020). Statistical modeling approaches on the other hand aim to reveal the mathematical relationship between time-series data of renewable energy. With the development of computing techniques and hardware, artificial intelligence-based forecasting models can now provide promising forecasting performance compared to physical or statistical approaches due to their potential abilities in data mining and feature-extracting (Daut et al., 2017).

Renewable energy forecasting is a sophisticated process, many factors affect forecasting accuracy. Yet, forecasting horizon and resolution, forecasting model inputs, and forecasting methods and techniques are the main factors (Ahmed et al., 2020)

## I.2. Problem statement

Renewable energy forecasting is an essential part of decision-making for future energy development – globally, regionally, and nationally (Debnath et al., 2018). While the attention towards energy forecasting started back in the 1960s (Nguyen, 2005), the oil crisis in the 1970s had boosted the interest in energy forecasting by emphasizing traditional fuel sources' dependency effects on the economies. More specifically, the role of exogenic political issues on the oil markets (Barsky & Kilian, 2004). Interest in renewable energy forecasting has heightened over the years due to an overwhelming consensus on its importance for reliable integration with existing power grids as penetration of these resources is increasing.

Most renewable energy forecasting emphases are placed on wind and solar energy owing to their variability, limited predictability, and instantaneous response to weather phenomena (Sweeney et al., 2020). The problem of solar and wind energy forecasting is multi-dimensional. As discussed earlier there are three main factors that affect forecasting accuracy, and each of these factors can be considered as one dimension of the forecasting problem.

The first dimension of the forecasting problem considers the forecasting horizon and resolution. As different horizons and resolutions lead to different forecasting accuracy (Das et al., 2018) (Wu & Hong, 2007), optimizing the best horizon and resolution to maximize the accuracy is the first challenge. However, in most cases, horizon and resolution are set by the regulatory authorities or grid operators, and thus there might be no control over this dimension. For instance, many grid operators in the European Union (EU) are required to report 24-hours-ahead with 15 minutes resolution for each grid-connected PV farm (Orasch, 2009) (Zsiborács et al., 2019). Yet, the intraday (15 minutes) resolution is still an appealing factor to study as the 15 minutes resolution might not be the proper choice for all renewable energy resources.

The second dimension is formed by forecasting model inputs. Many data sets can be utilized by the forecasting model as inputs including weather, meteorological, past energy, and geographical datasets. Furthermore, each data set might contain several variables. As there are too many variables that can be used for designing and building renewable energy forecasting systems, selecting input variables imprudently increases cost, computational complexity, and forecasting errors (Raza et al., 2016). Choosing the right variables is one of the challenges in designing renewable energy forecasting models. Another important factor regarding the input data is availability. Historical data should be collected over several years or even more. Unfortunately, some renewable sites don't have full datasets covering all the required variables for an extended period of time.

The third dimension of the forecasting problem is the question of forecasting technique. This means identifying and tailoring one (or more) suitable techniques to forecast renewable energy for the required horizon in the required resolution while utilizing the available data. Many available techniques can be used to design forecasting solutions, yet not all techniques are suitable to achieve the forecasting objectives within the available data. Some techniques might perform better under certain conditions. For example, Artificial Neural Networks (ANN) can be used for complex nonlinear, nonanalytical, and nonstationary stochastic forecasting problems (Inman et al., 2013), but certain ANN structures may also be used to provide quick predictions with a high accuracy ratio even with smaller input data (Khishe et al., 2018) (Akkaya & Çolakoğlu, 2019).

It can be concluded that designing a renewable energy forecasting solution is a very complex task. Different dimensions must be considered before and during the design. Additionally, the dimensions of the forecasting problem are interconnected. Any change in any dimension might affect the accuracy of the overall forecasting system.

## I.3. Aims, objectives, and research questions

The main aim of this research is to employ artificial intelligence technologies in renewable energy forecasting. This will help renewable energy farms and operators in providing better energy forecasting accuracy based on the available data, which will also reflect on grid stability and enhance renewable energy integration with electricity grids. Another aim is to provide researchers, energy practitioners, grid operators, and decision-makers with a comprehensive guide for forecasting methods based on the available data. Hence, one of the secondary targets is comparing the current widely utilized methods of forecasting (such as multiple linear regression) with machine learning methods. Part of creating a comprehensive guide is to study the state of the art and research status by performing tech mining analysis. Tech mining helps in identifying the most active authors, countries, affiliations, as well as the evolution and most recent trends in the field. Therefore, the objectives of this research consist of four main parts.

First, to collect, study and analyze the documents published in the field of renewable energy forecasting using artificial intelligence technologies. This step is targeted at providing a comprehensive overview of the field by allocating the main bibliometrics indicators, which is also helping in finding the research gaps.

Second, to study and analyze the meteorological weather and past generation data variables to enhance the selection of input data that will be used for designing, training, and building renewable energy forecasting models. This step is targeted at reducing costs, reducing complexity, and improving the accuracy of the suggested system.

Third, to further analyze forecasting horizons and resolutions. This step is targeted at analyzing the effects of utilizing different input data resolutions in forecasting accuracy for different forecasting horizons.

Fourth, to study and analyze different algorithms and techniques utilizing different input data. This step is targeted at finding the best algorithms, techniques, and hybrid combinations to assure the highest forecasting accuracy.

Since many dimensions are interconnected and must be considered while designing a forecasting system, this research has four main research questions, each question deals with one (or more) aspects of the design problem and the objectives of this thesis.

The first research question deals with data and data availability problems. There are too many variables that can be used for designing and building renewable energy forecasting systems. So the first research question is:

- Which variables should be used to design, train, and build renewable energy forecasting models to improve forecasting accuracy while reducing costs and computational complexity?

The second research question deals with forecasting models and techniques. Different forecasting methods lead to different forecasting accuracy. Based on the application and data availability, the forecasting model selection criteria can be tailored. Also, various algorithms and techniques can be used (together) in building hybrid renewable energy forecasting systems. Consequently, the second research question is:

- What are the algorithms and techniques to design, train, and build renewable energy forecasting models that can improve forecasting accuracy based on the available data?

The third research question deals with forecasting horizon and resolution. Different forecasting horizons and resolutions lead to different forecasting accuracy. Consequently, the third research question is:

- What are the resolutions that can be utilized to design, train, and build renewable energy forecasting systems to assure the highest forecasting accuracy?

As mentioned earlier, most grid operators in the EU require a 15 minutes resolution forecast, thus the fourth research question regarding resolution is:

- Does the regulatory 15-minutes forecasting resolution provide similar accuracy when forecasting wind and solar?

The uniqueness of this research does not only derive from providing accurate forecasting methods while reducing costs and computational complexities, but also from the ability to enhance renewable energy integration for grids even for locations where limited data variables are available. While the main innovative part is to provide a comparative analysis between different artificial intelligence forecasting methods like ANN, SVM, and KNN based on different input data methods. Moreover, a hybrid wind energy forecasting model will be designed, trained, and tested to forecast wind energy for both long and short-term forecasting.

# II. RENEWABLE ENERGY AND THE CHALLENGE OF ELECTRICITY PRODUCTION FORECAST

As described in Chapter 1, the main research focus of this study is to employ artificial intelligence technologies in renewable energy forecasting. Renewable energy is very general and contains many possible sources such as wind, solar, and hydro. Moreover, each of these resources might be used in different ways, for example, solar energy can be used for water heating or generating electricity using photovoltaic (PV) cells. So first section II.1 provides an overview of the energy and electricity market. Then, the main renewable sources that can be used for electricity generation were identified in section II.2. In sections II.3 andII.4, the growth of renewable energy resources was discussed, as well as the challenges which limit wider utilization of these resources, and some possible solutions. Finally, the forecasting solution was discussed in sections II.5, II.6, and II.7.

## II.1. Electricity market

Energy is considered to be one of the most important resources for any community. With the enormous technological development and the entry of modern means of communication into every detail of daily life, energy has become a necessity for ensuring the continuity and development of societies.

With the current pattern of economic growth, the World Energy Council predicts that global energy demand will grow by 45% - 60% by the year 2030 compared to what it was in 2010. At the same time, in the European Union (EU) this growth is expected to reach 15% to 20% (Tvaronavičienė et al., 2020). Electrical power demand is one of the important needs as well. In fact, electricity is the fastest-growing energy demand (Conti et al., 2016). For electricity demand, the World Energy Council expects that the demand will be doubled by 2060. The study explained this huge increase to the growing urban lifestyles and rising incomes. Additionally, other studies show a trend of continuous growth in electricity demand even though world gross domestic product (GDP) growth slowed in the past two decades (Council, 2019) (Conti et al., 2016).

The electricity market has changed significantly in the past decade. Aiming to improve the economic benefit of this massive market, changes regarding deregulation and competition impacted the appearance of the wholesale electricity market (Ventosa et al., 2005) (Wolak, 2021). This new shift allows decentralized electricity generating units to supply the grid instead of relying on central state – or utility-based – units. Decentralized electricity generation units compete to provide electricity at a price set by the market. Thus, cheaper and greener electricity generation technologies are crucial in today's market.

The appearance of decentralized units has also encouraged firms to invest in renewable resources. Renewable resources have great advantages for decentralized units such as ease of application and generation, low maintenance cost, and the ability to operate without the need for huge investment in infrastructure. Those factors increase the investment in renewable electricity generation as can be seen in Figure 1. It can be noticed that solar and wind resources are the most attractive for investors. These renewable alternatives are further discussed in the next section.

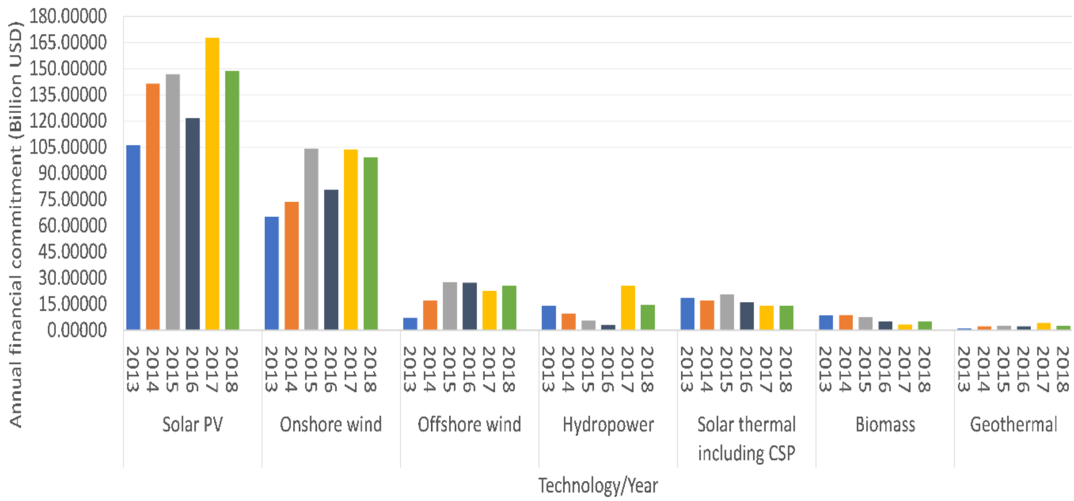


**Figure 1 Annual financial commitment to renewable energy**[3]

## II.2. Renewable alternatives of electricity production

Electricity is not freely available in nature, thus producing electricity requires transforming energy from other forms into electrical power. This transformation or generation is normally carried out in power plants. Power plants might be driven by heat from burning fossil fuel or nuclear fission, kinetic energy from floating water or wind, or by other means such as Photovoltaic (PV) or geothermal. This section focuses on renewable sources for electricity generation like hydro, biomass, solar, and wind.

### *II.2.1. Hydro energy*

[3] Source: IRENA and CPI (2020) Global Landscape of Renewable Energy Finance 2020, International Renewable Energy Agency, Abu Dhabi.

Electricity produced from hydropower ("hydroelectricity") takes up almost 20% of the world's total electricity. Despite the relatively low annual financial commitment for hydro resources shown in Figure 1, 70% of total electricity generated by renewable resources is produced by hydropower, which makes hydropower the most widely used renewable resource (Infield & Freris, 2020). This might be explained by the large construction wave of the hydropower fleets back in the 1960s to the 1980s ((IEA), 2020).

Hydroelectric power is an essential part of the electrical grid. The output power of the hydro can be controlled to augment variable generation sources like solar and wind, preventing any power shortages in lean hours. Likewise, being efficient at low generation levels allows the hydroelectricity plants to follow predicted and unpredicted changes in power demand, knowing that hydro generators can respond within minutes to changes in power demand (Price, 2014).

While the price of electricity generated by hydro is relatively low, the construction of hydroelectric infrastructure can be complex and cause some environmental impacts. Building such infrastructure might lead to loss of arable land and population displacement, disrupt the natural ecology of the surroundings, and affect habitats. Also, dams have a risk of having dam bursts, which would have catastrophic effects (Brown, 2021).

### *II.2.2. Biomass energy*

A promising alternative energy source is biomass. Various organic materials such as wood, agricultural residues, and animal or human waste can be utilized as biomass. Those materials can be converted into energy by several methods including:

- Direct combustion to produce heat. Burning organic materials like wood or animal products is a common method for converting biomass to energy. The energy produced from burning organic materials can be used directly for heating or for generating electricity in steam turbines.

- Thermochemical conversion to produce solid, gaseous, and liquid fuels: The produced fuel (e.g. biodiesel) can be used as a source of energy by either direct use for generating heat or by machines (green transportation).
- Chemical conversion to produce liquid fuels. A chemical process known as transesterification can be utilized to transform animal fat, vegetable oils, and greases into usable fuel.
- Biological conversion to produce liquid and gaseous fuels. Ethanol and renewable natural gasses can be produced by anaerobic fermentation of biomass products. Ethanol and renewable natural gasses (known as biogases) can be utilized as fuel for vehicles or to produce heat and electricity.

Even though biomass seems to be a very promising source of energy, many challenges limit its applicability. Processes of biomass transformation require energy and good infrastructure, adding extra costs to the capital investment. Moreover, utilizing biomass may affect the land and water used by societies (Evans et al., 2010). Therefore, in regions that already suffer a great lack of food this could potentially lead to rising prices and increased social dilemmas.

### *II.2.3. Nuclear energy*

Nuclear energy is the use of nuclear reactions to generate heat. The heat is then (most commonly) used to produce electricity. Nuclear decay, nuclear fusion, and nuclear fission reactions are used to obtain nuclear power. With 434 nuclear reactors operating in 32 different countries, producing 3-400 GW of electricity, nuclear power can be considered a “mature technology” (Sims et al., 2003). Furthermore, nuclear power plants now have improved safety tools, increased plant performance, and lifetime extension thus becoming competitive compared to other electricity generation methods (Comsan, 2010). Since nuclear energy is a low-carbon technology, it is seen as a major opportunity for the decarbonization of global economies (Prăvălie & Bandoc, 2018).

Despite being a reliable source of energy, still, some serious disadvantages must be considered before utilizing the nuclear solution. Radioactive waste is one of the major nuclear issues. Also, nuclear power plants have relatively low thermal efficiencies and a non-negligible risk of accidents (Pioro & Duffey, 2015).

### *II.2.4. Solar energy*

One of the most important sources of green energy today is solar power. The star of our solar system radiates a near-infinite amount of energy toward our planet. Harvesting and converting solar energy efficiently can contribute to reducing the energy bill (Al Shafeey & Harb, 2018).

There are various ways to utilize the power of the Sun. Its heat may be used directly as a heating appliance, but one of the greatest potential lies in photovoltaic (PV) technology. Although PV is not a new concept, technology itself is the center of attention for innovation and development. The basic concept behind this method of conversion is that absorbing light causes the excitation of an electron or other charge carrier to a higher-energy state, whilst differences in the electrochemical potentials and the ejection of electrons force the built-in electrical field to move, thus creating electricity (Price, 2014).

The drop in PV modules cost, ease of application without the need for huge infrastructure, and the subsidization for commercial PV systems by state governments have increased the growth of PV system utilization (Price, 2014) (Madsen et al., 2019) as can be seen in Figure 2.

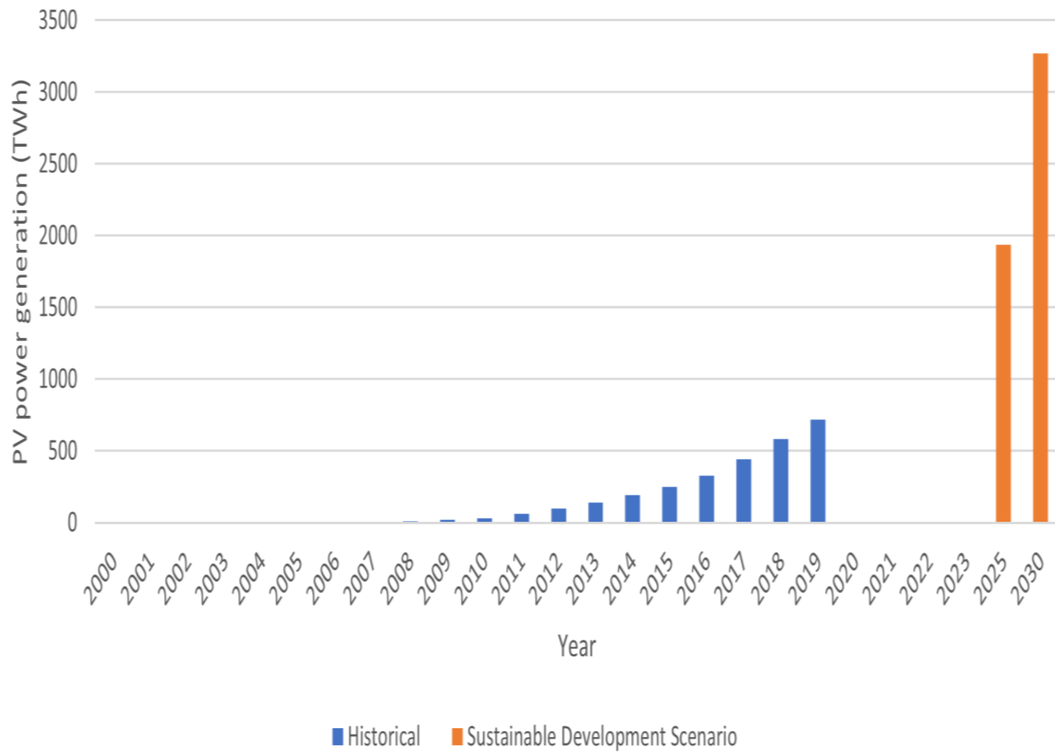


**Figure 2 Solar PV power generation in the Sustainable Development Scenario, 2000-2030 [4]**

As PV systems rely directly on sunlight (solar radiation) to generate electricity, the reliability of this source is the biggest challenge. Solar radiation fluctuates during seasons, months, days, and even minutes. This means the electricity generated by PV systems also fluctuates with time, which is unfavorable. Thus, a backup energy source or storage unit(s) might be needed to cover the power shortages in lean hours.

In case of grid-connected solar systems or farms, PV modules are connected to the grid via inverters, thus the energy generated is fed to the national grid. The potential fluctuations in grid-connected PV power might affect grid stability. Therefore, forecasting the potential amount of PV energy is of great importance,

[4] Source: IEA, Solar PV power generation in the Sustainable Development Scenario, 2000-2030, IEA, Paris

especially if PV is a stand-alone system where there are no other means of energy balance control (like storage or hybrid systems).

### *II.2.5. Wind energy*

Wind is one of the oldest major sources of power for mankind. Wind was – and still is being- used for grinding grains, pumping water, as well as long-distance transportation (i.e. sailing) (Grigsby, 2018). Another key usage of wind power is generating electricity. The kinetic energy of floating wind can be transformed into electricity via wind turbines. The huge potential of generating electricity by wind has encouraged many countries to utilize this resource. Thus till 2010 wind turbines were installed in over 70 countries (Grigsby, 2018).

The advantages of utilizing wind for electricity generation are similar to other renewable resources as it is available in most geographical locations including offshore. Unlike nuclear and fossil fuel generators, wind does not require water to generate electricity. The disadvantage which might be challenging for better utilization of wind power is fluctuations. Similarly to PV power, wind is a highly fluctuating variable. Consequently, energy generated by wind is not stable and changes over time. Hence, forecasting wind power is very important for grid-connected wind turbines in order to stabilize the grid.

It can be concluded from the previous sections that electricity production from renewable resources like nuclear, hydro, and biomass is relatively easier to control. The amount of electricity generation from these resources can be calculated and monitored based on the technical specifications of the power plants. Meanwhile, the amount of generated electricity from sources like wind and solar is hard to control as it depends on external atmospheric variables. Hence, the focus of this thesis will be directed toward unstable renewable sources like solar and wind.

## II.3. Renewable energy growth

The continuous growth in energy demand attracts many scholars to develop and improve energy technologies, where energy production can be easier, cheaper, and most importantly clean.

The advantages of utilizing clean sources for energy-generating have accelerated the spread of renewable energy usage. As a result, continuous huge growth in the energy produced by renewable resources has been observed. With 2.9% annual growth, renewables are (and predicted to be) the fastest-growing source of electricity generation for the years 2012 to 2040 (Conti et al., 2016). According to the Renewables Global Status Report (REN21), about 70% of the net addition in the global power generation came from renewable resources in 2017 (Network, 2018).

Solar energy generation is of great interest for many reasons. The main reason is the huge potential amount of solar energy, the sum of solar energy that strikes the earth in one hour is enough to supply the energy demand of the whole planet in a year (Lewis, 2007). Solar power is theoretically capable of supplying the whole world's total energy demand (Görig & Breyer, 2016). Other factors that attracted the attention to solar energy in general and photovoltaic (PV) in particular are the ease and the possibility of generating energy in various geographical regions, even inside cities at the local residential or commercial levels without the need to change the entire infrastructure, unlike other renewable energy resources (Ashi et al., 2014). The above-mentioned factors had eased the process of organizing and creating energy policies like tax incentives, feed-in tariffs, and market share quotas. The effective policies and the ease of installation encouraged the construction of solar power stations for generating electricity.

Out of the total global growth in renewables, solar energy alone has shown a 50% annual growth rate in the last decade (Victoria et al., 2021). Solar energy resources would be a potential solution for many current era problems. Traditional power plants are responsible for 25% of anthropogenic emissions (Jerez et al., 2015). Greenhouse gas (GHG) emissions can be reduced using solar resources for producing energy as the GHG emissions associated with solar energy generation are less than the traditional oil and gas ones (Şen, 2004). The range of CO2 emissions produced by solar resources is estimated at around 0.03-0.09 Kg per kilowatt-hour, while the CO2 emissions produced by coal and natural gas are estimated at around 0.64-1.63 and 0.27-0.91 Kg per kilowatt-hour, respectively (Kabir et al., 2018). The reduction of toxic gases by using solar resources will not only reflect on nature but it has also direct effects on mankind. A study done by Machol (Machol & Rizk, 2013) stated that using renewable energy resources instead of fossil fuels would minimize premature mortality rates, decrease the loss in workdays, and improve the overall healthcare economic benefits.

Moreover, solar technologies are more labor-intensive compared to fossil fuels. More jobs can be created per unit of energy generated by solar technology, which reflects positively on the energy labor market (Kabir et al., 2018).

The efficiency of solar technologies has shown a solid increase, and additionally, a steady decrease in costs was also observed, especially in photovoltaic (PV) technologies. For instance, the total costs for a PV module were 1.3 USD in 2011, decreasing to 0.5 USD in 2014, which is almost 60% cost decreasing in four years and it is expected to decrease furthermore (Parkinson, 2015). Despite the improvement in the efficiency and the reduction in the costs of the PV technologies, new developments are still raising, specifically, the development of new methods to enhance the total efficiency of the PV module and improve the economic benefits (Jäger-Waldau, 2006) (Parida et al., 2011) (Razykov et al., 2011).

Wind is also among the most utilized renewable source. Wind resources have performed well recently and have also improved environmental, climate, visibility, and noise pollution impact. Despite the fact that it produces less energy output compared to fossil fuels, wind can still be an efficient energy source with a high potential to meet energy needs. Further, the use of wind energy has only a partial impact on the environment (Suryakiran et al., 2020). These factors have encouraged many countries to increase the installation of wind farms to produce electricity. Scotland for example had supplied 100% of its electricity demand from wind resources in November 2018 (Suryakiran et al., 2020). Some other countries have reached relatively high levels of wind energy penetration, such as 39% of electricity production in Denmark, and 14% in Ireland (Suryakiran et al., 2020).

In OECD countries, particularly OECD countries in Europe, most of the renewable energy growth comes from solar and wind energy generation resources (Conti et al., 2016).

## II.4. Renewable energy challenges and some solutions

Irrespective of all the advantages of utilizing renewable technology for energy production, there are some hindrances limiting growth and wider utilization. One of the crucial drawbacks of renewable solutions is low energy converting efficiency or the density problem (Huang et al., 2013) (Nelson & Starcher, 2015). Even though sources like sun and wind have the potential to supply the whole earth's energy demand (Bull, 2001), the current energy conversion efficiency is relatively low and energy harvesting technologies need more improvements.

Depending on the concentration ratio, the maximum theoretical limit of efficiency for solar cells crystalline silicon (c-Si) with bandgap energy (Eg) of 1.1 eV is approximately 31% or 41% as specified by the Shockley Queisser Efficiency Limit (SQ Limit) (Huang et al., 2013). While the efficiency of wind turbines can reach a maximum of 59.26% as stated by the Betz limit, still factors like blade number losses, whirlpool losses, end losses, and airfoil profile losses prevent wind turbines from reaching the theoretical Betz limit (Blackwood, 2016).

Beside the low conversion efficiency, renewable energy resources might have some environmental issues. Biomass energy might not be cost-effective, and burning biomass can result in air pollution. The startup and maintenance costs of geothermal energy resources can be expensive. Hydropower plants (dams) can cause flooding and have significant ecological impacts on local hydrology. While the main problem of solar and wind is the reliance on weather variables like solar radiation and wind speed (Ellabban et al., 2014).

The reliance on weather variables for wind and solar resources causes the variability problem. As discussed earlier, solar cells rely mainly on solar radiation to produce energy, while wind turbines rely mainly on wind speed. The nature of weather variables -which unfortunately highly fluctuates over time- leads to generation uncertainty (Alshafeey & Csáki, 2019).

As can be concluded, the energy produced by renewable resources like solar and wind depends mainly on weather factors. Among the most vital factors are solar radiation, temperature, wind speed, and humidity. Therefore, to control the potential amount of renewable energy, these factors must be studied and optimized.

In the production of power with large applications like grid-connected renewable energy farms, the fluctuations in energy supply can cause instability in the grids. Variation in the produced energy has serious consequences. Besides the economic issues with fluctuating power sources, some technical issues such as frequency and voltage anomalies, overloading of existing transmission lines, and demand/supply mismatch might affect the grid.

To achieve renewable energy generation stability, there are many solutions that have been used to overcome the above problems. For instance, energy storage units can be used to stabilize power. Storage units act as a buffer by supplying load when there is a shortage and storing energy when there is excess generation. Yet, using storage units is impractical for large applications. In addition, storage units like batteries have limited usage cycles and it has to be replaced after a certain time adding extra costs (Koohi-Fayegh & Rosen, 2020) (Wang et al., 2012).

Other possible solutions include using hybrid systems like solar (or wind)-diesel hybrid systems or solar-wind hybrid systems. Diesel hybrid systems are a well-known solution, especially in remote locations. The diesel generator can provide energy whenever there is a supply shortage from renewable resources. However, this solution is not always economical and may be impractical for inter-cities applications. Moreover, hybrid Diesel systems show a bad performance for diesel generators (Yamegueu et al., 2011) (Cavalcante et al., 2021).

One promising solution for renewable energy generation stability is to enhance renewable energy forecasting (Singh, 2013) (Devaraj et al., 2021). If the potential renewable energy can be accurately predicted with lower uncertainty, renewable energy systems can be better designed and optimized helping grid operators in managing power supply and demand (Pazikadin et al., 2020). Accurate forecasts would improve grid stability as well (Rodríguez et al., 2018). Additionally, accurate energy forecasting does not only provide value through reduced imbalance penalties (incurred due to the difference between the scheduled and actually delivered energy) but also leads to increased competitiveness by providing advanced knowledge in real-time energy market trading.

Improving renewable energy forecasting and creating accurate forecasting models are among the most important aspects of renewable energy production and are considered to be one of the 'hottest' topics in the renewable energy research field as will be seen in chapter 3. The renewable energy forecasting methods will be discussed in the next section.

## II.5. Renewable energy forecasting methods

One of the simplest approaches to forecast renewable energy is using the average values of historical renewable energy and weather records (Abunima et al., 2019). However, the average method is not suitable because averages do not represent the full range of values, which will be reflected by having some considerable errors in the forecasted values. Furthermore, those errors and the inherent uncertainty of weather variables will be aggravated, leading to additional uncertainty (Linguet et al., 2016).

Renewable energy forecasting models may be categorized as either deterministic forecasting or uncertainty analysis (Liu et al., 2019), as can be seen in Figure 3. Depending on the renewable energy forecasting model input data, deterministic renewable energy forecasting models can be divided into three main groups (Liu et al., 2019) (Foley et al., 2012): physical, statistical, and intelligent models. Also, a fourth hybrid category can be added. Although the definition of hybrid modeling is quite vague, still, hybrid forecasting model refers to the combination of two or more different algorithms or methods.

In the physical approach, explanatory variables from Numerical Weather Prediction (NWP) are used to forecast renewable energy. The Explanatory variables (mainly hourly mean of weather variables) are derived from a meteorological model of the weather dynamics, then they are used to predict renewable energy for a given number of steps ahead.

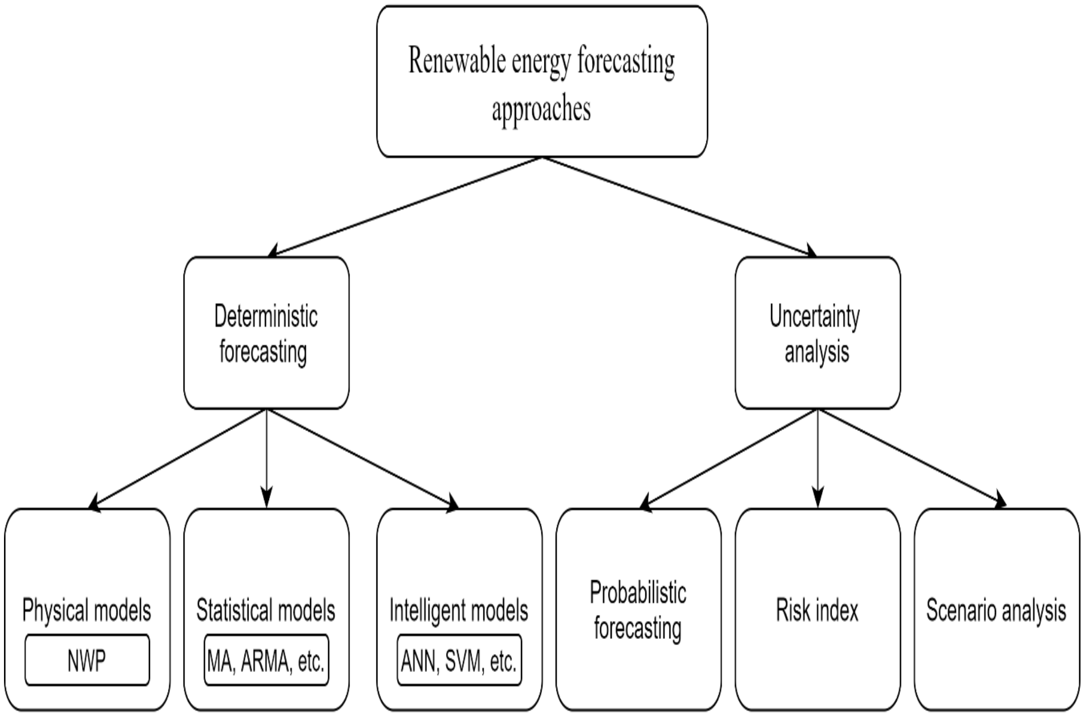


**Figure 3 Renewable energy forecasting classification based on adopted approaches**[5]

Statistical methods use analysis of historical time series of weather variables (Foley et al., 2012). In this group, statistical approaches are used to forecast weather variables or to directly forecast renewable energy production. Some widely used statistical renewable energy forecasting techniques include moving average (MA), autoregressive (AR), autoregressive moving average (ARMA), and, autoregressive integrated moving average (ARIMA)

Intelligent methods – with being the most recent trend – are based on artificial intelligence or more specifically, machine learning (Liu et al., 2019). Machine Learning is called the learning approach because the models are able to learn from

[5] Based on Liu, H., Chen, C., Lv, X., Wu, X., & Liu, M. (2019). Deterministic wind energy forecasting: A review of intelligent predictors and auxiliary methods. *Energy Conversion and Management*, *195*, 328-345. and Foley, A. M., Leahy, P. G., Marvuglia, A., & McKeogh, E. J. (2012). Current methods and advances in forecasting of wind power generation. *Renewable Energy*, *37*(1), 1-8.

the relationship between the predicted and forecasted renewable energy using historical time series data. The main machine learning approaches are artificial neural networks (ANN), support vector machines (SVM), and k-nearest neighbors (KNN).

The other main category (uncertainty analysis) focuses on the representation of uncertainty in renewable energy forecasting context and includes probabilistic forecasting (Gneiting, 2011) as well as solutions based on risk index (Pinson & Kariniotakis, 2004) or generation of scenarios (Pinson et al., 2009). Probabilistic forecasting may be parametric or non-parametric and uses a probabilistic measure to estimate the uncertainty in the future of power generation (Zhang & Wang, 2016).

To obtain the goals of renewable energy forecasting, first relevant data should be collected and processed properly. The data needed for building wind and solar forecasting models are mainly the historical ground weather data and energy consumption/generation data.

The historical ground weather data can be collected from ground weather station for a specific location, normally for a long period (like twenty or twenty-five years), those types of data consist of different weather parameters such as global horizontal irradiance (GHI), direct normal irradiance (DNI), diffuse horizontal irradiance (DHI), as well as ambient temperature, and wind speed.

Other weather parameters like humidity and sunshine hours would be useful and helpful for building more accurate models. Those weather parameters are used for creating typical meteorological year (TMY) data. The TMY data is used to calculate the potential energy that could be generated by wind and solar systems (Rodríguez-Gallegos et al., 2018). The time-series energy consumption/generation data is also important. Processing this time-series data is crucial not only for indicating the amount of energy needed at a certain time but also for analyzing and forecasting energy usage patterns and linking them with new applications like smart grids and the internet of things (IoT).

Analyzing the consumption patterns is the key to success and the first step towards “smart energy”. With such analysis, the energy-saving programs can be assisted and a good demand-supply balance can be achieved. Since the supply by renewable sources is uncertain, creating as accurate as possible wind and solar energy prediction models becomes valuable (Singh & Yassine, 2018).

As mentioned earlier, creating accurate renewable energy prediction models depend mainly on weather variables. In other words, predicting the weather variables accurately will lead to accurate energy prediction models. Despite the historical weather data can be used directly to build prediction models, initiating decisions for renewable energy projects based on historical weather data without any statistical processing should be avoided for two main reasons. The first reason is the lack of some specific weather variables data like solar radiation records in certain locations for the required long period. Secondly, even if a good historical record exists, analyzing this historical record inherently assumes that the future and the historical solar radiation profiles are exactly the same which is unrealistic (Brook & Finney, 1987).

There are some simple solutions to this problem. The simplest solution is using the average values of weather variables from historical records (Abunima et al., 2019). This solution is not appropriate because the average method does not represent the full range of values which leads to overestimating and underestimating the predicted amount of energy. Furthermore, this error will be aggravated due to the inherent uncertainty of the weather nature, creating more uncertainty (Linguet et al., 2016). Thus, a robust renewable energy prediction model must retain the statistical properties of the weather variables, while allowing the development of a certain level of stochastic behavior and eliminating a certain degree of uncertainty.

So, renewable energy prediction model has to create future data for a specific location. The future data demonstrates the same behavior as the original historical data but with improvements and most importantly with the ability to accurately forecast the uncertain future reading. Accordingly, the output of the prediction model is useful in improving the reliability of wind and solar generation.

To build a renewable energy prediction model that can achieve the goals of prediction, reduce uncertainty, and improve prediction accuracy, new methods have to be used. Building such models will not be easy as the time-series weather variables data has a highly non-stationary nature. Further, the model must accurately predict the future potential amount of renewable energy that can be generated at a specific time in a specific location, to overcome the uncertainty problem and improve the grid stability.

Wind and solar prediction models output is a future reading of the potential amount of energy, which means it has a time horizon. Depending on the prediction horizon and forecasting period there are two main methodologies to forecast wind and solar energy. The first methodology includes satellite sky imagery and Numerical Weather Prediction (NWP) while the second methodology includes machine learning models (Voyant et al., 2017).

For very short-term prediction which was defined by Gordon Reikard to range from 5 minutes to 6 hours (Reikard, 2009), statistical models including the artificial neural networks (ANNs) can be applied (Diagne et al., 2013). NWP models typically perform better than satellite sky imagery in forecasting longer horizons, or 4 to 6 hours onward predictions (Perez et al., 2010), thus each of the two methodologies has its use. Yet, machine learning methods can be used for both long and short-term predictions (Voyant et al., 2017).

One important fact to know about solar prediction is that solar energy production is mainly affected by solar radiation. Solar radiation is composed of infrared and ultraviolet energy waves. The range of its wavelength is between 300 to 3000 nanometers. Additionally, solar radiation has three components (Akter & Shoeb, 2015):

1. Global horizontal irradiance (GHI)
2. Diffuse horizontal irradiance (DHI)
3. Direct normal irradiance (DNI).

The GHI is the total incoming amount of solar radiation on a horizontal surface and is composed of the DNI and DHI (Akter & Shoeb, 2015) (Perez-Astudillo & Bachour, 2015).

GHI is a vital factor to be measured for planning and constructing any projects or applications of photovoltaic (PV) systems (Perez-Astudillo & Bachour, 2015).

## II.6. Major factors affecting renewable energy forecasting

Renewable energy forecasting is a sophisticated process, many factors affect forecasting accuracy. Yet, forecast horizon and resolution, forecasting model inputs, and forecasting methods and techniques are the main factors (Ahmed et al., 2020).

### *II.6.1. Forecast horizon and resolution*

The forecasting horizon can be defined as the time between the present and the effective time of the predictions, while forecasting resolution is the frequency of the predictions (Antonanzas et al., 2016). Forecasting horizon is one of the major factors that affect forecasting accuracy (Das et al., 2018). Most studies have categorized renewable energy forecasting horizon into four major classes: very short-term (few seconds to 30 minutes), short-term (30 minutes to 6 hours), medium-term (6 hours to one day), and long-term forecasting (days, weeks,.., etc) (Raza & Khosravi, 2015) (de Marcos et al., 2019). Although it should be noted that the time-scale classification of forecasting models in the literature is relatively vague (Liu & Chen, 2019). Thus far there is not any international classification criterion (Nespoli et al., 2019) (Sobri et al., 2018).

The relationship between the forecasting horizon and accuracy is reverse: forecasting accuracy decreases significantly for longer horizons (Ahmed et al., 2020). Day-ahead renewable energy forecasting is of the utmost importance in decision-making processes (Cococcioni et al., 2011). Moreover, certain grid operators in the European Union (EU) are required to report a 24-hours-ahead with 15 minutes resolution forecast from each grid-connected PV farm (Orasch, 2009) (Zsiborács et al., 2019).

Resolution is the frequency of the input variables per time unit. Similar to the horizon, the input data resolution is also affecting the accuracy of the forecast (Wu & Hong, 2007). Generally, past energy production data are collected with high sampling resolution, such as 10 minutes (Hao & Tian, 2019), 15 minutes (Ding et al., 2019), and one hour. (Li et al., 2018b). When a longer prediction horizon is required, the original high-resolution data are usually averaged to build up low-resolution data (Gallego et al., 2011). Unfortunately, the process of averaging will lead to a lot of information losses as the rapid fluctuations in original high-resolution data will be neglected (Wang et al., 2017).

### *II.6.2. Forecasting model inputs*

Most renewable energy output is highly correlated with weather variables (F. Wang et al., 2019). Yet, not all weather variables have the same significance for renewable energy forecasting. Choosing among potential variables depends on the availability of related data for the required location and time period, as some locations might not have full datasets covering all the required parameters, especially for an extended period of time. Depending on the parameters used – i.e. the so-called explanatory variables utilized – prediction models can be built using three different approaches (Aggarwal & Saini, 2014) (Bacher et al., 2009) (Jafarzadeh et al., 2012) (Khatib et al., 2012):

- structural methods that only utilize the geographical and meteorological parameters as inputs;
- time-series methods that only utilize the historical data of renewable energy as inputs;
- hybrid methods that utilize renewable energy historical data as well as other variables like geographical and meteorological parameters as inputs.

It should be noted, that there are two basic approaches to time-series forecasting: direct forecasting and multi-step rolling forecasting. While in the direct approach only actual historical data is utilized (i.e. always being one-time horizon behind), in the rolling approach the predictions of the previous values are used like they were actual values when predicting the next value (being one resolution step behind). Although there are some claims that multi-step rolling forecasting is slightly better than the direct option for certain tasks (see for example (Lan et al., 2019) for frequency-based solar irradiation forecasting), for renewable energy output this method has been found to be problematic. This is because the error generated in each step is propagated to the subsequent steps (Sahoo et al., 2020). Thus, it is found to be less accurate due to the accumulation of the error along the prediction horizon (Galicia et al., 2019).

### *II.6.3. Forecasting models and techniques*

As discussed earlier, forecasting can be performed using several methods, including physical, statistical, or intelligent methods. Moreover, each method has different modeling techniques. Utilizing different methods and techniques leads to different forecasting accuracy as any forecasting technique has its own pros and cons (Aslam et al., 2021). Based on the application and data availability, the forecasting model selection criteria can be tailored (Ineichen, 2006). For example, physical methods are not suitable for short-term renewable energy forecasting (Aslam et al., 2021). Likewise, most of the current statistical renewable energy forecasting models are designed as linear models, where their abilities to solve complex forecasting problems are limited to longer forecasting time horizons (Aslam et al., 2021). Choosing the right forecasting technique is not an easy or direct task. As intelligent forecasting models are among the most recent and widely utilized, this thesis is focusing on artificial intelligence methods, more precisely, machine learning models which will be discussed in the next section.

## II.7. Artificial Intelligence Machine Learning prediction models for renewable energy forecasting

### *II.7.1. Machine learning forecasting models*

Machine Learning (ML) is the part of Artificial Intelligence (AI) that studies artifacts. More specifically ML studies algorithms that improve their performance with experience (Harrington, 2012). Thus any ML model aims to get better and more accurate output with more data fed to the model. ML models are able to find the complex relations between the model's inputs and outputs even when the representation is impossible. This ability allows ML to be used for many purposes like classification problems, pattern recognition, spam detection, and most noteworthy in this study; data mining and forecasting problems (Voyant et al., 2017).

It was mentioned previously that machine learning methodology is part of artificial intelligence methods. The intelligent system can learn from datasets, giving the system good abilities to learn and improve the outputs without explicit programming.

Considering the huge amount of historical data needed to build the renewable energy prediction model and the data would still be fed to the system day by day (as wind and solar live data would still be collected and fed to the prediction model), using ML methodology is one of the most suitable choices. As ML forecasting models work with huge datasets, preprocessing and data preparation steps have to be undertaken before the model can be trained for the actual forecast. After data preparation, the ML models can be trained for forecasting tasks. When machine learning is used for predictions, the ML system consists of 'output' or 'response' variables and one or a set of 'input' or 'explanatory' variables. Using training samples of known values, an estimation or approximate values of the function can be found (Voyant et al., 2017). The error between the real and predicted values can be represented by a loss function. The loss function is mathematically represented by several methods like squared error and absolute error (Voyant et al., 2017).

There are different machine learning techniques used for renewable energy forecasting, supervised learning, unsupervised learning, and meta-learning algorithms are the main technique (Lantz, 2019). The choice of the appropriate ML technique is a very important step. Before choosing the ML technique, the main machine learning tasks must be determined first. The four main machine learning tasks are (Lantz, 2019):

- Classification
- Numeric prediction
- Pattern detection
- Clustering.

Thus, the task will drive the choice of the machine learning technique. Table 3 below shows the different machine learning techniques and their main tasks (Lantz, 2019).

**Table 3 Machine Learning Models and Learning Tasks**[6]

| Model | Learning Task |
|---|---|
| **Supervised learning algorithms** | |
| K-nearest neighbors | Classification |
| Naive Bayes | Classification |
| Decision Trees | Classification |
| Classification rule learners | Classification |
| Linear regression | Numeric prediction |
| Regression trees | Numeric prediction |
| Model trees | Numeric prediction |
| Neural networks | Multi-use |
| Support vector machines | Multi-use |
| **Unsupervised learning algorithms** | |
| Association rules | Pattern detection |
| K-means clustering | Clustering |
| **Meta-learning algorithms** | |
| Bagging | Multi-use |
| Boosting | Multi-use |
| Random forests | Multi-use |

[6] Based on Lantz, B. (2019). Machine learning with R: expert techniques for predictive modeling. Packt Publishing Ltd.

A considerable amount of research discusses the utilization of machine learning in forecasting, especially Neural Networks (NNs). NN can improve time-series predictions and enhance the accuracy of the forecast (Makridakis et al., 2018). Yet, the superiority of the AI and machine learning forecasting models have some limitations (Makridakis et al., 2018):

- their conclusions are based on a few, or even a single time series, raising questions about the statistical significance of the results and their generalization;
- the methods are evaluated for short-term forecasting horizons, often one-step-ahead, not considering medium and long-term ones;
- no benchmarks are used to compare the accuracy of ML methods versus alternative ones.

The objective evaluation of AI and machine learning forecasting algorithms is highly important as these models are computationally demanding. Using ML models in some cases might be a waste of resources. Moreover, many ML techniques can be utilized for forecasting. Different techniques lead to different forecasting accuracies and require different resources (data, computational resources, time,..etc.). Choosing the right technique for the required accuracy is one of the main challenges for ML models.

Among all the mentioned machine learning techniques, the Artificial Neural Networks (ANN) has many attractive advantages in forecasting renewable energy. ANN has the ability to solve complex nonlinear, nonanalytical, and nonstationary stochastic problems without the need for complex computer programming (Inman et al., 2013). The abilities and advantages of ANN forecasting models have influenced many scholars and practitioners to use ANN in solar forecasting – including irradiation and energy production prediction. As a result, a rising number of research reports and ANN-based forecasting applications have been observed since 1990 (Garud et al., 2021).

The use of ANN techniques for solar energy forecasting was utilized to forecast day-ahead solar energy with 1-hour forecasting resolutions utilizing historical weather data as well as time-series power data (Chen et al., 2011). But ANN has also been applied as a base for short-term solar power prediction models (Almonacid et al., 2014) that showed a good performance for 1-hour power forecasting. Another short-term prediction model using ANN utilized both temperature and solar irradiance data (Oudjana et al., 2013) and achieved good accuracy of forecasting power. What is worth noting is that ANN was the most applied technique for solar power forecasting over the last ten years especially for short-term prediction, as 48% of related articles published between 2009-2019 were using ANN (Mellit et al., 2020) and 97% of those articles were forecasting power for short or very short term horizon (and only 3% were forecasting medium and long term).

According to (Maldonado-Correa et al., 2019), ANN is among the most frequent machine learning models for wind power forecasting. ANN is used also in hybrid models, where more than one technique or different ANN algorithms were applied for the forecast. Table 4 below summarizes some machine learning facts in wind power forecasting.

**Table 4 Some facts about wind power forecasting using machine learning[7]**

| | |
|---|---|
| Frequent model used | ANN, Hybrid ANN models |
| Frequent Data source | Wind farm datasets |
| Frequent locations of the wind farms<br>(from where the data was obtained) | Europe and China |
| Frequent software used in the forecast | Matlab |

[7] Base on (Maldonado-Correa et al., 2019

Even though it is difficult to achieve accurate wind prediction using a single prediction forecasting method, ANN is still one of the most reliable methods (Li et al., 2020).  Forecasting wind power with ANNs generally includes four main steps (Hossain et al., 2021):

- Step 1: Data pre-processing
- Step 1: Develop the forecasting model
- Step 3: Training the model
- Step 4: Forecasting and measuring performance.

Training of ANN is performed in a supervised manner. It is assumed that a training set is available, given as historical data and containing some inputs (weather or past wind power values in case of time-series data) and the corresponding desired outputs, which are all presented to the network. The adequate selection of inputs for ANN training is highly influential to the success of the training. The most popular ANN learning algorithm for wind forecasting is the backpropagation algorithm, where the input is passed through the layers until the final output is calculated, then it is compared to the real output to find the error. The error is then propagated back to the input adjusting the weights and biases in each layer. The standard backpropagation learning algorithm is the steepest descent algorithm that minimizes the sum of square errors. However, the standard backpropagation learning algorithm tends to converge slowly.

### *II.7.2. The use of machine learning in PV energy forecasting*

The abilities and advantages of using machine learning (especially ANN) forecasting models have influenced many researchers to use these models in solar forecasting – including irradiation and energy production prediction (H. Wang et al., 2019). For example, three different short-term prediction models using ANN were built in (Oudjana et al., 2013). The first model utilized temperature data to forecast power and showed huge errors; the second model utilized solar irradiance data which resulted in better forecasting accuracy; while the third model showed the best accuracy and it utilized both temperature and solar irradiance to forecast power.

ANN PV power forecasting model based on a self-organizing feature map (SOFM) was proposed in (Yousif et al., 2017), where the suggested model uses solar irradiance and ambient temperature to forecast PV power. The results show that using ANN based on SOFM improves prediction accuracy.

Real-time solar irradiance was used to make two-hour-ahead solar irradiance levels forecasting in (Vanderstar et al., 2018): the proposed method uses ANN to forecast the irradiance and genetic algorithm to optimize array size and position in order to obtain the most accurate prediction. The suggested method shows adequate forecasting capabilities, yet, it has some limitations as this method only works for non-zero solar Global Horizontal Irradiance (GHI) values.

In (Notton et al., 2019) ANN models were proposed to forecast different solar irradiance components for 1 to 6-hour horizons. The results show that ANN is a very promising method to forecast solar radiation. Also, several ANN forecasting models were proposed to predict hourly solar irradiance in six different locations in Nigeria (Bamisile et al., 2020). The results show that all of the proposed ANN models performed well and can be used for PV performance calculation.

Multiple weather variables such as temperature, precipitation, wind speed, and solar irradiation were used to build a multi-channel convolutional neural network (CNN) prediction model in (Heo et al., 2021). The suggested model extracts meteorological as well as geographical features of PV sites from raster image datasets. The results show high forecasting capabilities, however, to avoid any biased prediction, sufficient data should be included.

ANN was the most applied technique for solar power forecasting over the last ten years especially for short-term prediction as 48% of related articles published between 2009-2019 were using ANN (Mellit et al., 2020).

Unlike ML methods which formulate solar energy prediction problems as a black box, statistical methods reveal the mathematical relationship between the input variables and the output (H. Wang et al., 2020). Such statistical methods include Autoregressive Moving Average (ARMA), Auto-Regressive Integrated Moving Average (ARIMA), exponential smoothing, and regression (F. Wang et al., 2020) (Das et al., 2018).

Multiple Linear Regression (MLR) is also popular in PV solar power forecasting (e.g. (De Giorgi et al., 2014), (Oudjana et al., 2012), (Pitalúa-Díaz et al., 2019)). Regression methods establish a relationship between the explanatory (meteorological and geographical) variables and dependent variables (the forecasted PV power) (Das et al., 2018).

Table 5 provides a brief chronological overview of the main PV forecasting methods applied for different horizons using various resolutions and input variables.

**Table 5 Forecasting methods, horizons, resolutions, and variables (source: author)**

| Reference | Forecast horizon | Forecast resolution | Methods | Variables |
|---|---|---|---|---|
| (Oudjana et al., 2012) | 7 days | 24 hours | Linear regression, MR, neural network | Global irradiance, temperature |
| (Al-Messabi et al., 2012) | 10 and 60 min. | 10 and 60 min | Dynamic ANN | Actual and past values of power |
| (Ogliari et al., 2013) | 24 hours- | 1 h | ANN hybrid approach | Weather variables |
| (De Giorgi et al., 2014) | 1–24 hours | 1–24 hours | Statistical methods based on MR analysis; ANN | PV power, module temperature, ambient temperature, solar irradiance |
| (Chu et al., 2015) | 5–15 min. | 5 min. | Many methods including cloud tracking, k-NN, ANN | Power past values and sky images |
| (Leva et al., 2017) | 24 hours- | 1 h | ANN | Power and solar radiation past values, Numerical Weather Prediction variables |
| (Pitalúa-Díaz et al., 2019) | 30 days | 5 min | MR, Gradient Descent Optimization (GDO) and Adaptive Neuro-Fuzzy Inference System (ANFIS) | Solar radiation, ambient temperature, wind speed, daylight hour, and PV power |

### *II.7.3. The use of machine learning in wind energy forecasting*

Many research articles focus on deterministic wind energy forecasting. Jung et al. (Jung & Broadwater, 2014) presented an overview of different wind speed and wind power forecasting models. Physical, statistical, spatial correlation, and regional forecasting models were reviewed. It was found that choosing the best wind forecasting model is a hard task as various models will perform differently in different situations, yet using a combination of numerous methods and/or techniques (hybrid) strives in leveraging the strength of different models and improving the accuracy. Qian et al. presented a comprehensive review of different decomposition-based hybrid models for wind energy forecasting. The authors discussed decomposition methods, the challenges of these methods and finally provided a comparative analysis of various decomposition-based models (Qian et al., 2019). Zendehboudi et al. presented a review of the development and application of SVM in wind and solar energy forecasting. The authors found out that for both wind and solar energy forecasting, hybrid SVM models perform better than other models (Zendehboudi et al., 2018). Hybrid wind energy forecasting models were reviewed based on weighted-based, preprocessing, optimization, and residual error modeling by Ren et al.in (Ren et al., 2015), Xiao et al. in (Xiao et al., 2015), and Tascikaraoglu et al. in (Tascikaraoglu & Uzunoglu, 2014).

Recent trend shows a growing number of applications that are based on AI technologies, this applies to wind forecasting as well. A vast number of researchers are improving intelligent technologies to accurately predict wind speed and power. Sideratos and Hatziargyriou used machine learning methods for short-term wind power forecasting, they used a combination of fuzzy logic and neural network techniques to forecast wind farm power output. The authors stated that the results can be used effectively for operational planning in 1–48 h ahead wind farm (Sideratos & Hatziargyriou, 2007). Rahmani et al. proposed a hybrid system that consists of two meta-heuristic techniques under the category of swarm intelligence to forecast the energy output of a wind farm. The empirical results indicate that the proposed technique can estimate the output wind power based on the wind speed and ambient temperature with acceptable accuracy (Rahmani et al., 2013). Zameer et al. proposed a wind power prediction system that uses a combination of machine learning techniques for feature selection and regression. The authors stated that the proposed model performs better than the existing prediction models in terms of performance measures, and can be used as an effective wind power prediction model (Zameer et al., 2015). Chi et al. studied the performance of direct and iterative methods for multi-step ahead wind speed forecasting, three machine learning methods including linear regression, multi-layer perceptron, and support vector machine was developed. The results show that neither direct nor iterative forecasting can always outperform each other in terms of all the error measures (Chi et al., 2015).

Some other researches mainly focus on ML technologies for wind forecasting, Barbounis et al. used three local recurrent neural networks to provide 72 time-steps ahead wind speed and power forecasts. The results show that the suggested model has outperformed the static rivals in terms of forecast errors (Barbounis et al., 2006). Barbosa de Alencar et al. proposed different models like neural network, ARIMA, and hybrid (ARIMA and ANN) for short, medium, and long-term wind power prediction. The hybrid model shows the smallest errors for all forecasting horizons (Barbosa de Alencar et al., 2017). Wang et al. proposed hybrid models utilizing various ML techniques such as Support Vector Regression (SVR) with seasonal index adjustment (SIA) and Elman recurrent neural network (ERNN). The hybrid models were applied in three different sites in China and predicted the behaviors of daily wind in a reasonable way (Wang et al., 2015). Yao Zhang and Jianxue Wang developed a combination of the k-nearest neighbor algorithm (k-NN) and the kernel density estimator (KDE) method for probabilistic wind power forecasting. The suggested approach showed a good forecasting performance (Zhang & Wang, 2016).

A review of solar and wind energy forecasting research published during the last five years was discussed by (Alkhayat & Mehmood, 2021), it was found that most of the studies included in the review proposed models to forecast the next 24 hours or less. Moreover, very few researchers studied and developed models to predict wind energy for different horizons i.e long and short-term. Table 6 summarizes some research involving forecasting wind power.

**Table 6 Literature concerning wind power forecasting (source: author)**

| Authors | Input data | Forecast horizon | Method | Techniques used |
|---|---|---|---|---|
| (Jursa & Rohrig, 2008) | Wind power, NWP time series | Short-term | Intelligent models | ANN and KNN |
| (Ghadi et al., 2014) | NWP, SCADA | Short-term | Intelligent models | combination of imperialistic competitive algorithm (ICA) and ANN |
| (Zameer et al., 2015) | Wind speed, Relative Humidity, Temperature | Medium-term | Hybrid | ANN, SVR |
| (Eseye et al., 2017) | NWP | Medium-term | Intelligent models | Genetic algorithm/ANN (GA-ANN) |
| (Li et al., 2018a) | NWP | Short-term | Intelligent models | SVM |
| (Zhang et al., 2019) | Wind speed and power | Short-term | Hybrid | long short-term memory network |

| | | | | |
|---|---|---|---|---|
| | | | | (LSTM) algorithm and Gaussian mixture model (GMM) |
| (Xiang et al., 2020) | Wind speed, and power | Short-term | Hybrid | Secondary decomposition (SD) and bidirectional gated recurrent unit (BiGRU) |
| (Nam et al., 2020) | Past 24 h renewable electricity supply | Short-term | Hybrid | Empirical mode decomposition, LSTM, gated recurrent unit |
| (Wang et al., 2021) | Wind speed, and power | Short-term | Intelligent models | Deep learning network stacked by independent recurrent autoencoder (IRAE) |
| (Yildiz et al., 2021) | Meteorological wind speed, direction, and power | Short-term | Hybrid | Variational mode decomposition (VMD), Convolutional neural network |

To augment the above literature review a further analysis has been conducted by using tech mining analysis in the next chapter.

# III. TECH MINING ANALYSIS

One of the aims of this study is to provide researchers, industry, businesses, and decision-makers with a comprehensive guide on renewable energy forecasting using artificial intelligence technologies. Part of creating this guide involves studying the state of art for this field of science and summarizing the most recent research outputs, most active authors, institutions, and most trendy topics.

In addition, conducting tech mining will help in better understanding the topic of renewable energy forecasting using artificial intelligence technologies and its subtopics. The outcomes of tech mining analysis will be also used side by side with the literature review to conceptualize the research gap. This chapter also seeks to address the status of a set of scientific productions in the world which is indexed in the Scopus database using scientometrics indicators.

## III.1. Tech mining overview

Scientific and technical documents databases contain important research results that are valuable to the researchers, industry, business, and decision-making communities. Analyzing those documents can be useful in showing the trends and relations of the analyzed topic (Bortoluzzi et al., 2021). To analyze such large unstructured data, methods to handle unstructured data sets must be used. One of the methods is bibliometric (tech mining) analysis (Xie et al., 2019). The cross science between quantitative analysis and statistical methods is normally referred as bibliometric analysis (Alshafeey et al., 2018). Using data mining techniques to perform bibliometric analysis on technology fields is known as tech mining (Ziegler, 2009).

Tech mining allows researchers to investigate scientific and technical documents, as well as extract valuable information, summarize the latest research outcomes, and understand the recent research directions and the evolution of a certain topic. As part of tech analysis, statistical tools are used. The process starts with a collection of bibliographic scientific documents and publications. This collection is then broken down into lists that focus on several publishing patterns (Ellegaard & Wallin, 2015). Such patterns include authors' production, national bibliographies, subject bibliographies, geographical and institutional aspects (Ellegaard & Wallin, 2015).

Since one of the major problems of utilizing renewable energy is the uncertainty in production (AlShafeey & Csáki, 2021), the purpose of this chapter is to investigate the status and the evolution of scientific studies in the field of renewable energy forecasting using artificial intelligence methods. Applying tech mining analysis to the mentioned domain would help in summarizing the most recent research results to researchers, industry, and decision-makers.

## III.2. Materials and methods

In order to achieve the objectives of this study, a combination of systematic, objective, and quantitative literature review methods along with content analysis methods were applied. The starting point was a systematic search for the related literature in the selected database.

Scopus database was chosen; the ease of discovery of peer-reviewed research and the vast range of energy research work were the main reasons for database selection. The extraction of the articles was based on specific searching criteria to ensure the relativity of the selected articles. The following keywords and Boolean searching criteria were used: TITLE-ABS-KEY ( power OR energy ) AND forecasting AND ( ( artificial AND intelligence ) OR ( machine AND learning ) ) AND ( LIMIT-TO ( LANGUAGE , "English" ) ).

The documents were extracted and analyzed early in 2020, then later in 2021, the search has updated to cover the latest results, hence, the date of the search was 5 October 2021. The search was limited to the English language with available full text. In total, over 25000 related articles were identified and extracted from the Scopus database into a tabular format. The search results were downloaded in Comma Separated format (CSV). These articles were checked and cleaned in Python data analysis environment. After Python check, over 18000 articles were extracted. The collected data were then analyzed using R software packages. The Bibliometrix package and supporting packages including dplyr, Matrix, and ggplot2 were used. The convert2df function was used in order to covert the data into a bibliographic data frame that matches the tags used in Scopus.

### III.3. A general summary of the extracted articles

The main information about the extracted articles is summarized in Table 7. In total, 18107 articles were extracted for analysis. These articles were published in 187 different sources like journals, books, conference proceedings, etc. The articles were published between 1991 and 2022, this indicates that the science of renewable energy forecasting using artificial intelligence is relatively new. It was also found that on average it takes almost 2.2 years for each article to get its first citation. The 18107 extracted documents were written by 25133 authors. Only 477 authors published documents without any coauthors (single-authored documents), while the remaining 24656 authors were collaborating. The high level of collaboration between authors reflects on the published documents as only 4% of the documents (738 documents) were published by a single author. This shows the high level of collaboration.

Another indicator that shows the high level of collaboration is the average number of documents by authors which is found to be 0.72.

**Table 7 Main information about the extracted articles (source: author)**

| **Description** | **Results** |
|---|---|
| Main information about data | |
| Timespan | 1991:2022 |
| Sources (Journals, Books, etc) | 187 |
| Documents | 18107 |
| Average years from publication | 2.2 |
| Authors | |
| Authors | 25133 |
| Authors of single-authored documents | 477 |
| Authors of multi-authored documents | 24656 |
| Authors collaboration | |
| Single-authored documents | 738 |
| Documents per Author | 0.72 |
| Authors per Document | 1.39 |

## III.4. Annual scientific production

An increasing number of published documents that discuss renewable energy forecasting by AI can be observed over the years. As mentioned earlier, this topic is relatively new, as can be clearly observed in Figure 4 which shows annual scientific production. Although articles were started to get published in 1991, there was not any significant increase until 2009. After 2009, a steady increase in the number of published documents may be observed. This increase then became exponential after 2017, which shows great recent attention to this topic.

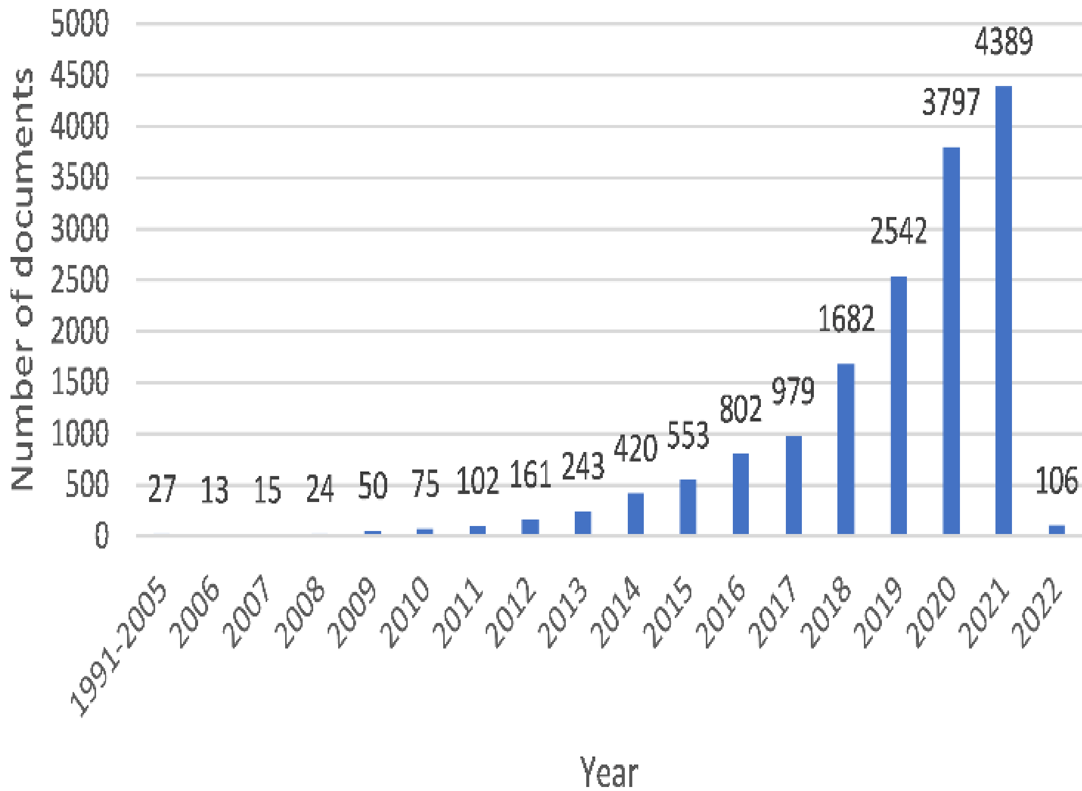


**Figure 4 Annual scientific production (source: author)**

## III.5. Most relevant authors

One of the important aspects of studying bibliometrics is identifying the relevant authors in the field. Getting familiar with the most relevant authors of a research field would not only help in recognizing the established, prolific, and emerging researchers who contribute to the scholarly but also implies responsibility and accountability of the published documents.

Therefore, the most productive authors were identified. Wang J. was the most productive author with 375 published documents, followed by Wang Y., with 315 published documents. The top authors and the number of their publications can be found in Figure 5.

It can be observed that the top 10 authors published almost 12.5% of the total published documents.

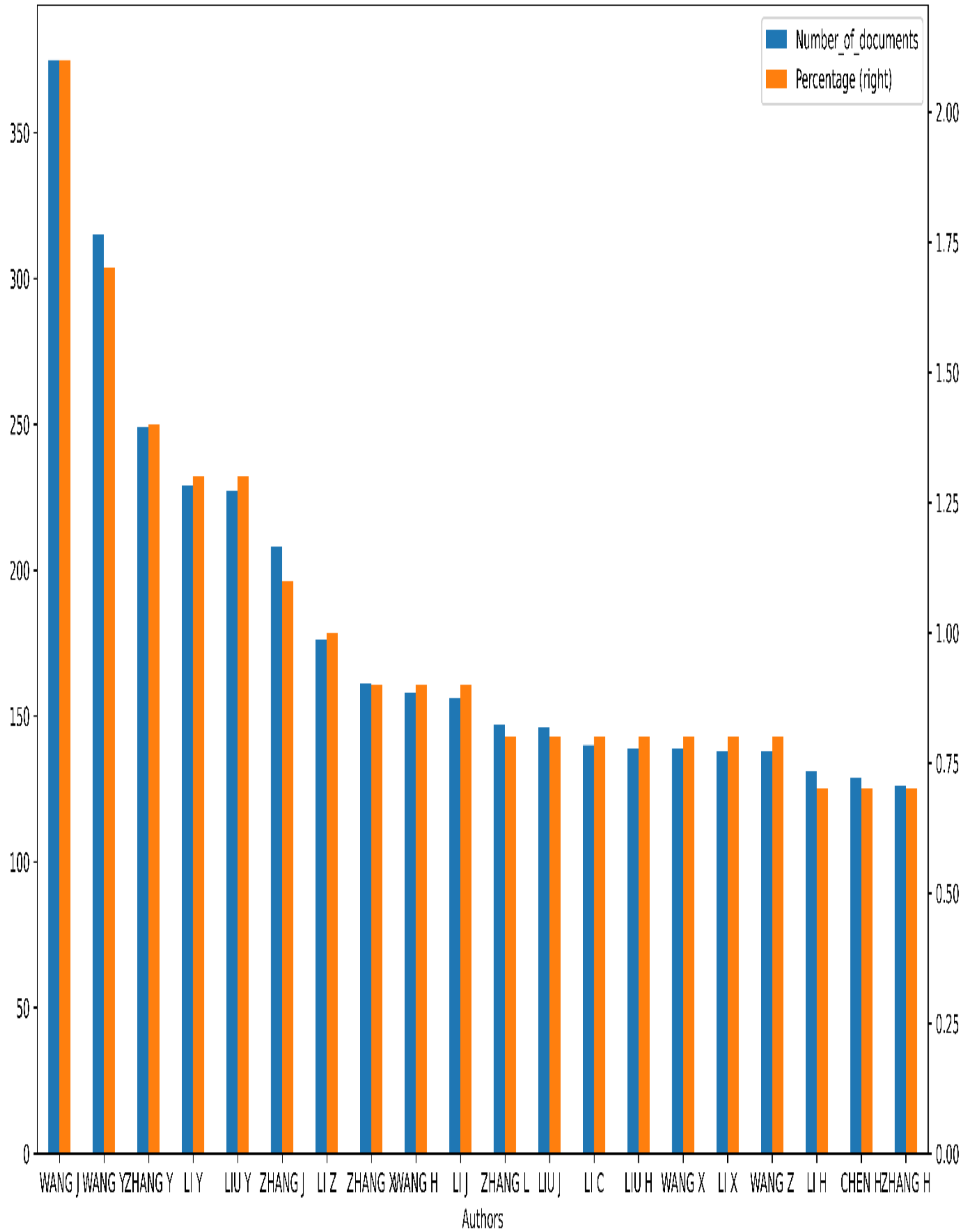


**Figure 5 Most relevant authors (source: author)**

Figure 6 below shows the top authors' production over time (N.Articles), and the number of total citations per year (TC per Year). Figure 6 confirms that the studied topic here is relatively new as most of the top authors have started publishing after 2011. Moreover, the recently increasing number of published articles and citations confirms the raise of attention for this topic.

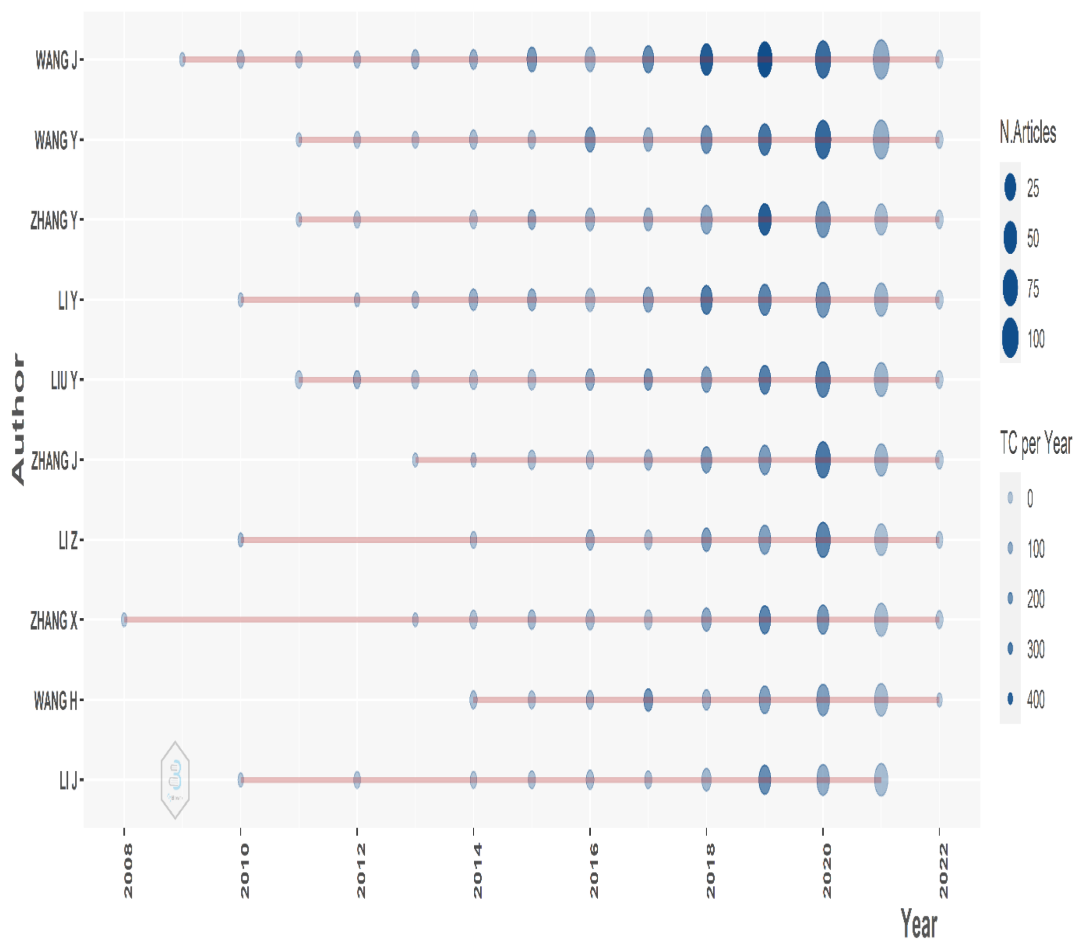


**Figure 6 Authors' production over time (source: author)**

### III.6. Most productive countries, affiliations, and collaboration

It is also worth analyzing the countries contributions to publications. So, most productive countries are pointed out as can be seen in Figure 7. Note that some documents were published in collaboration between two or more countries. So in Figure 7, SCP refers to Single Country Publication, while MCP refers to Multiple Countries Publications. It was found that China is the most active country. With over 5000 (22%) published documents in the field. China is leading for both SCP and MCP, which shows high level of collaboration between authors from China and the international communities. It is also interesting that China alone contributes more than the following top 5 countries. Another interesting fact is the large number of documents published by Asian authors, Asia had contributed to the field more than any other place. This shows the huge interest of Asian institutions in general and the Chinese institutions in particular for this field of research.

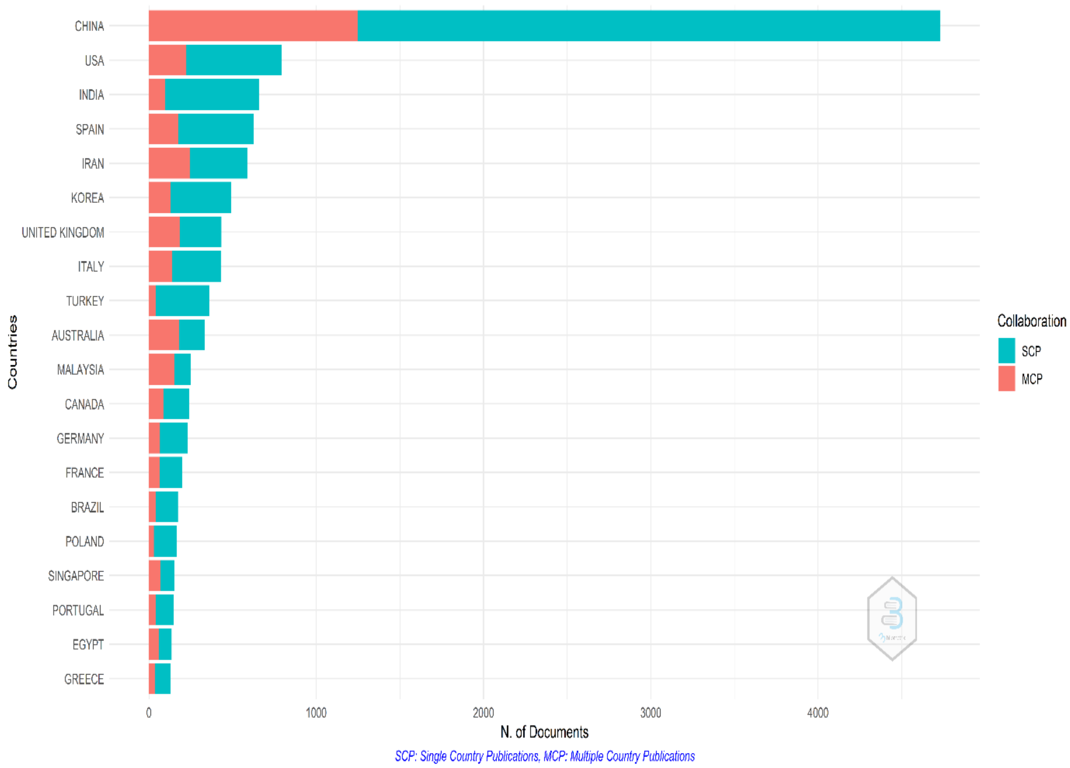


**Figure 7 Most productive countries (source: author)**

Besides analyzing the countries contributions, citations per country and average article citations are also worth exploring, as it gives an indication of the leading countries for research on this field. Average article citations can also give an indication of the research quality and popularity. Figure 8 shows the top countries by citations and the average article citations. The analysis shows that China is the most cited country, followed by the USA, Iran, and Spain. Once again, China has more citations than the following 5 countries taken together. Yet, China has an enormous number of published documents, and thus comparing the absolute number of citations might not be a fair comparison. Hence, the average article citations indicator was also used to compare the quality of the published documents.

The analysis shows that China is the most cited country as well, followed by the USA, Iran, and Spain. Once again, China has more citations than the following 5 countries taken together. Yet, China has an enormous number of published documents, and thus comparing the absolute number of citations might not be a fair comparison. Hence, the average article citations indicator was also used to compare the quality of the published documents. By looking at the average article citations in Figure 8, it can be seen China has on average 18.18 citations per document which is not a leading number.  While even though countries like Singapore, Malaysia, and Hong Kong have not published many articles in the field, yet, they have the highest citations per article. This shows the high level of quality and the high impact of academic work on the field. It can be also concluded that the high number of publications and citations for documents published in Chinese institutions does not necessarily mean the most accepted research quality.

Another interesting fact is the dominance of the Asian documents for both, total citations and average article citations. Here again, it can be clearly seen that documents published in Asia like China, Iran, and India have the highest number of citations. While countries such as Singapore, Malaysia, and Hong Kong have the highest citations per article. Even Pakistan, which is not on the top productive countries list, has a significant number of citations and high average article citations, higher than Spain for example which is one of the most productive and cited countries. This is another confirmation of the huge interest of Asian institutions in this field of research.  This also shows that the Asian market may be the hub and the producer of renewable energy forecasting technologies in the following years.

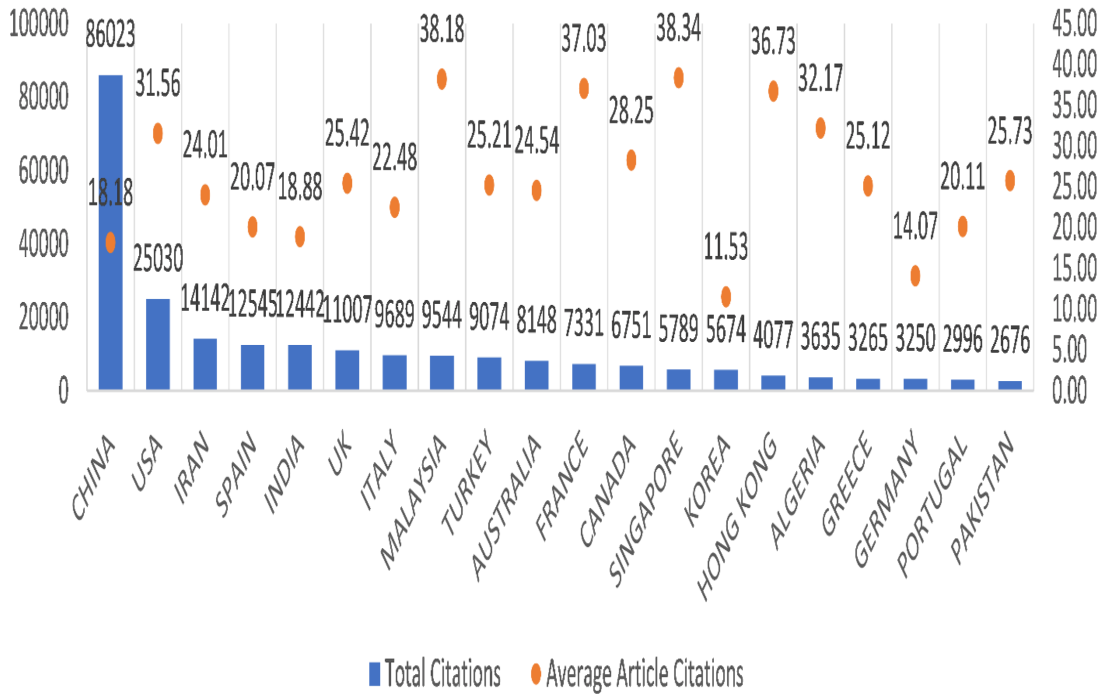


**Figure 8 Citations per Country (source: author)**

As it was found in this and previous sections, there is a high level of collaboration between authors. To study and examine the movement of researchers from one country to another the affiliation of the authors was studied as well.

Table 8 shows the most relevant affiliations and the number of documents published by each institution.

**Table 8 Most relevant affiliations (source: author)**

| Affiliation | Documents |
|---|---|
| NORTH CHINA ELECTRIC POWER UNIVERSITY | 1279 |
| HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY | 775 |
| TIANJIN UNIVERSITY | 480 |
| ISLAMIC AZAD UNIVERSITY | 435 |
| TSINGHUA UNIVERSITY | 435 |
| DONGBEI UNIVERSITY OF FINANCE AND ECONOMICS | 434 |
| SOUTHEAST UNIVERSITY | 344 |
| SHANGHAI JIAO TONG UNIVERSITY | 342 |
| ZHEJIANG UNIVERSITY | 340 |
| CENTRAL SOUTH UNIVERSITY | 326 |

As expected, it was found that the most relevant affiliations are related to Chinese institutions like north China electric power university, Hua Zhong University of Science and Technology, and Shanghai Jiao Tong University. Also, other than China, the most relevant affiliations list mainly includes Asian institutions like Islamic Azad University, the University of Tehran, and the University of Malaya.

To further study the level of collaboration between different institutions in different countries, country collaboration map was provided as can be seen in Figure 9.


**Figure 9 Country collaboration map (source: author)**

The map shows that Chinese institutions have the highest level of collaboration, where they collaborate with institutions in the UK, North America, Australia, and East Asia. It can be also seen that Europe (except the UK) has a relatively low level of national and international level of collaboration, compared to other places with a similar level of prestigious research and industrial institutions. South America and Africa have the lowest national and international levels of collaboration.

Another factor studied here is the sponsors of the extracted documents. It was found that most fund comes from Chinese institutions like National Natural Science Foundation of China, which funded over 5000 documents as can be seen in Figure

10. It was noticed that the fund contribution from African and South American institutions is very low compared to other continents.

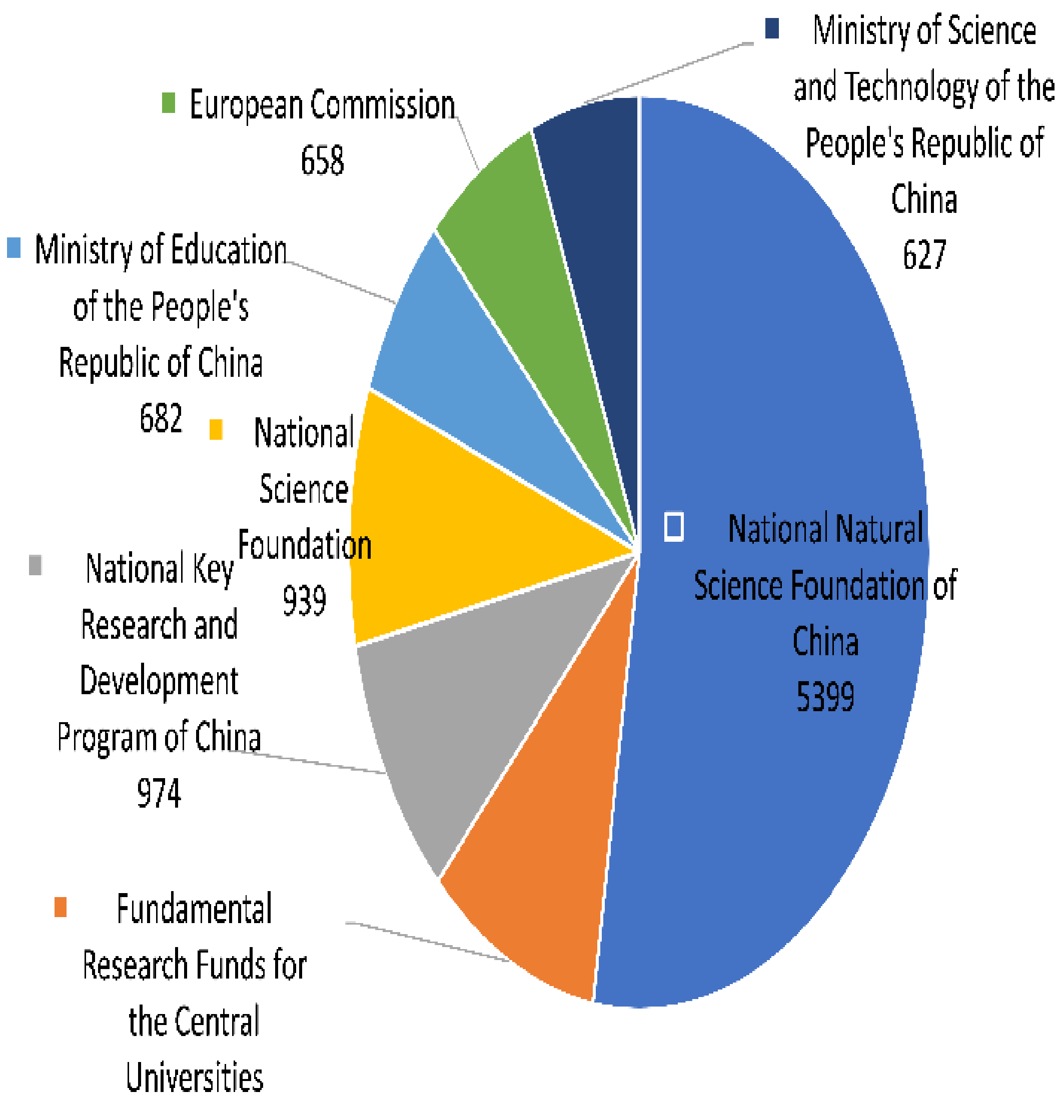


**Figure 10 Top sponsor institutions (source: author)**

## III.7. Thematic map

Another important indicator to analyze while studying bibliometric analysis is the thematic map. Conducting the thematic map analysis for forecasting renewable energy using artificial intelligence methods helps energy researchers, policymakers, and practitioners in discovering the field's current and future insights, providing the potential direction of research development, as well as delineating the conceptual structure of the field.

Thematic map is basically clusters of published documents' keywords and the connection between these keywords, it is based on co-word network analysis and clustering. The clusters' interconnection maps form themes. In themes, the keywords and their interconnected clusters draw a network map, known as thematic networks. The main characteristics of such themes are density and centrality. The map vertical axis represents the density which measures the cohesiveness among the nodes, while centrality is represented in the horizontal axis. Centrality measures the degree of correlation among different topics (Esfahani et al., 2019).

The degree of development and importance is measured in thematic maps by analyzing density and centrality. Greater relations a certain node has with other nodes in a thematic map are represented by higher centrality. Therefore, centrality measures the correction degree among different topics. Likewise, higher cohesiveness and internal correlation among nodes are represented by higher density. In other words, a density of a research field signifies its capability to sustain and develop itself. Thematic maps are intuitive graphs, themes can be analyzed according to the quadrant in which they are placed. The lower-right quadrant is basic themes, the lower-left quadrant is emerging or disappearing themes, the upper-right quadrant is motor themes, and the upper-left quadrant is specialized niche themes. The thematic map for renewable energy forecasting using artificial intelligence methods is shown in Figure 11.

Themes in the upper-right quadrant (motor theme) of Figure 11 are well developed and important for forming the research field such as “smart grid” and “optimization”. Themes in the upper-left quadrant (specialized niche themes) are well-developed internal ties but unimportant external ties and thus have marginal importance for the field such as “artificial neural network” and “prediction”. Note that neural network is one of the most studied machine learning techniques for forecasting renewable energy thus it was clustered as well-developed. Themes in the lower-left quadrant (emerging or disappearing themes) are weakly developed and marginal, mainly representing either emerging or disappearing themes such as “forecasting”. Themes in the lower-right quadrant (basic themes) are important for a research field but are not developed, so this quadrant groups transversal and general, basic themes such as “machine learning”, “deep learning”, and “artificial intelligence”.

By analyzing the thematic map in Figure 11, it can be concluded that the field needs to merge research focuses on important but not well-developed topics. For example, merging smart grid and deep learning research topics, or studying renewable energy forecasting and artificial intelligence, which is the goal of this thesis.

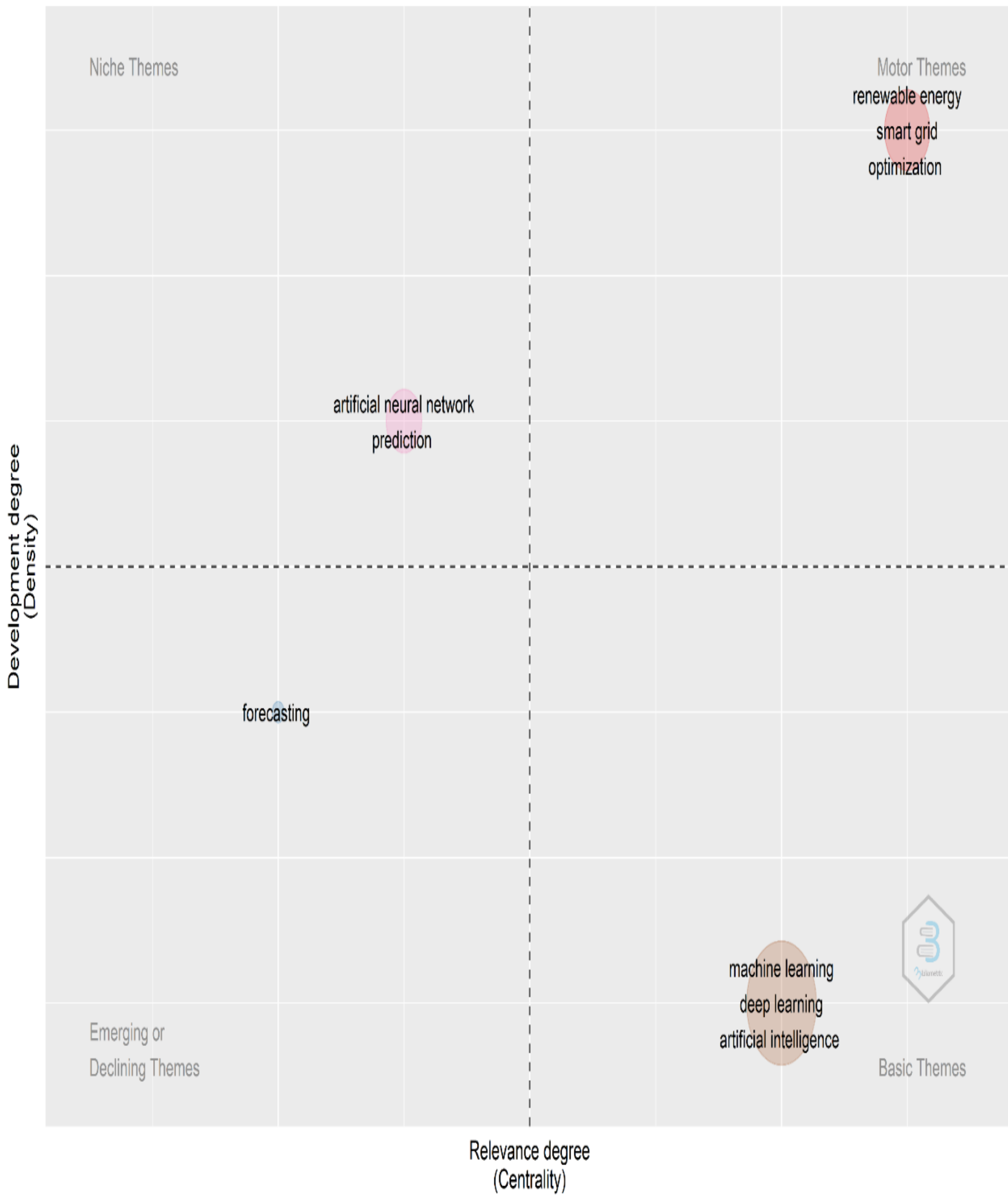


**Figure 11 Thematic map for renewable energy forecasting using artificial intelligence methods (source: author)**

.

## III.8. Trend topics

The exponential growth in the number and heterogeneity of research papers increases the difficulty to obtain a synthetic illustration of the research topics being investigated. As each field of research covers a large dynamic number of subfields and contains a huge variety of information. Heterogeneity between fields makes any specific subjects analysis a complex problem. As a result, it is difficult to obtain a wide perspective illustration of which exact topics of research are new, active, collapsing, or have been ephemeral. Exploring the status of research topics gives a good insight regarding the level of interest for a certain topic in the field. To shed light on more specific topics within the field and the evolution of topics within renewable energy forecasting using artificial intelligence, the trending topic analysis was conducted and the results are mentioned in Table 9. The analysis was performed based on the frequency of the keywords in the published documents. Hence, certain topics are considered trendy in a certain year if the frequency of their appearance as a keyword is higher than other topics for that year. For instance, Support Vector Regression (SVR) (see Table 9) used to be a trendy topic for the years 2014-2018, as that term was among the most frequent keywords in each year between 2014-2018 with an average frequency of 48 times per year.

It is noticeable from Table 9 that trendy topics are changing with time. Topics that used to be trendy in 2015-2020 are no longer considered trendy. Some topics like artificial intelligence, machine learning, and deep learning are gaining more and more attention, and have been trendy in the past few years. Recently, new emerging topics are rising. Some topics like long short-term memory (lstm), transfer learning, and climate change are currently gaining huge interest. Covid-19 is also one of the trendy topics which are being frequently studied in the context of forecasting renewable energy.

**Table 9 Trending topics within renewable energy forecasting using artificial intelligence field (source: author)**

| Topic | Trend time frame | Frequency |
|---|---|---|
| support vector regression (svr) | 2014-2018 | 48 |
| support vector machines | 2015-2020 | 61 |
| neural networks | 2016-2020 | 1218 |
| smart grid | 2017-2020 | 383 |
| genetic algorithm | 2017-2021 | 215 |
| artificial intelligence | 2018-2021 | 404 |
| machine learning | 2019-2021 | 1191 |
| deep learning | 2019-2021 | 671 |
| climate change | 2019-2021 | 80 |
| lstm | 2020-2021 | 122 |
| transfer learning | 2020-2021 | 63 |
| long short-term memory (lstm) | 2020-2021 | 52 |
| covid-19 | 2020-2021 | 49 |

## III.9. Conceptual Model of Research Gap

Various conceptualization methods can be used in the identification of research gaps. Methods like citation analysis, content analysis, meta-analysis, systematic reviews, future research, and limitations have been used (Farooq, 2017). However, there is not any definite process to identify the research gap defined in the literature (Farooq, 2017).

In this work, the tech mining analysis alongside systematic literature reviews was used to identify the research gaps. The conceptual model for identifying research gaps can be shown in Figure 12.

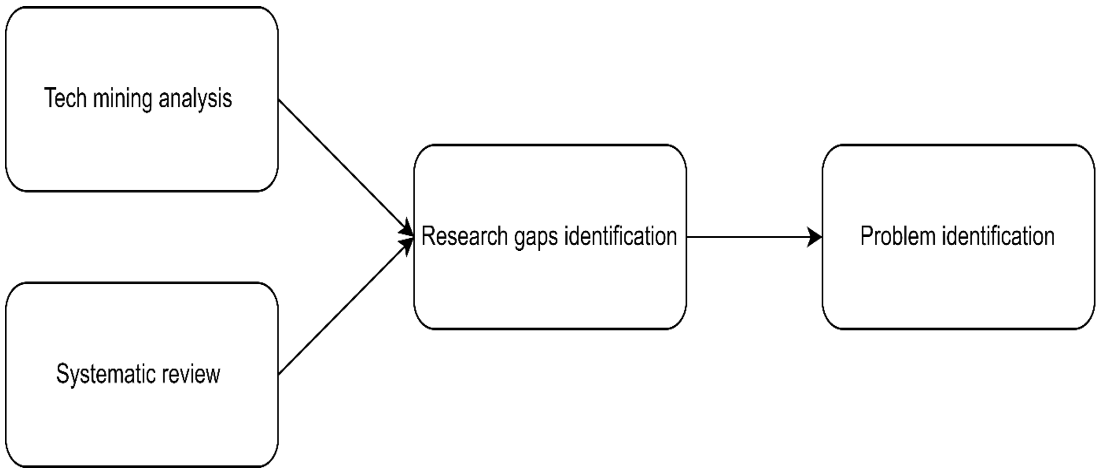


**Figure 12 The conceptual model for identifying research gaps (source: author)**

The research questions were further investigated to specify the research gaps. This investigation was done in chapter II (systematic review) and chapter III (tech mining analysis). The conclusions of these two chapters were then used to identify the research gaps and later to better understand the research problem.

Four main conclusions and corresponding research gaps can be drawn from chapters II and III:

1. While ANN and regression are heavily studied techniques for PV energy forecasting, most of the literature focus either on testing one of these techniques utilizing different input data methods (see for example (Inman et al., 2013) and (Oudjana et al., 2013)) or on testing these two techniques utilizing the same input data methods (such as (Ogliari et al., 2013), (Chu et al., 2015), and (De Giorgi et al., 2014)). Hence, the first research gap is the lack of comparison and analysis of the performance

for the ANN and regression modeling techniques using structural, time-series, and hybrid input data methods.

2. Most of the studies regarding wind energy forecasting proposed models to forecast the next 24 hours or less (Alkhayat & Mehmood, 2021). Moreover, very few researchers study and develop models to predict wind energy for different horizons i.e long and short-term. Hence, the second research gap is the lack of wind energy forecasting models that can predict both long and short-term energy production with decent accuracy.
3. While it has been established that different forecasting horizons lead to different accuracies, the impact of input data resolution could bear some clarification. Most studies have not clarified how the input data resolution might affect the forecasting performance. Hence, the third research gap is the lack of investigation of how different input data resolution affects prediction performance for both PV and wind energy.
4. By analyzing the thematic map in section III.7, it can be concluded that the field needs to merge research focuses on important but not well-developed topics. So, studying renewable energy forecasting and artificial intelligence is highly required. Hence, the fourth research gap is the need for investigating interconnected important but not well-developed topics such as renewable energy forecasting and artificial intelligence.

Thus, while addressing the above-mentioned research gaps, this work will focus on achieving the main objective mentioned in section I.3 which is employing artificial intelligence technologies in renewable energy forecasting. As well as, helping renewable energy farms and operators to provide better forecasting accuracy based on the available data, which will also reflect on grid stability and enhance renewable energy integration with electricity grids.

# IV. METHODOLOGY

## IV.1. Goals of the study and research questions

One of the key challenges facing renewable energy forecasting models is the task of choosing the right input variables to be utilized by the right ML techniques. Each of the ML techniques performs differently depending on weather variables, geographical location, or other complicated factors. Consequently, the methodology is carefully designed to solve this issue and achieve the aims of the thesis.

As proposed in Chapter I, the aim of this study is to employ artificial intelligence technologies in renewable energy forecasting and provide the researchers, energy practitioners, grid operators, and decision-makers with a comprehensive guide for forecasting methods based on the available data. Part of the comprehensive guide is to study the state of the art and research status by performing tech mining analysis which was provided in chapter III, while the second part will be done to compare the performance of different forecasting models based on the input data and to suggest a hybrid method for wind energy forecasting.

To achieve the aims of this study, four main objectives were set (see section I.3). Each objective deals with one (or more) aspect of the renewable energy forecasting problem. These aspects were specified in four research questions. The goal of this study will be achieved by answering the research questions and completing the objectives.

The main aim of this research is to employ artificial intelligence technologies in renewable energy forecasting. This will help renewable energy farms and operators in providing better energy forecasting accuracy based on the available data, which will also reflect on grid stability and enhance renewable energy integration with electricity grids. Another aim is to provide researchers, energy practitioners, grid operators, and decision-makers with a comprehensive guide for forecasting methods based on the available data. Hence, one of the secondary targets is comparing the current widely utilized methods of forecasting (such as multiple linear regression) with machine learning methods. Part of creating a comprehensive guide is to study the state of the art and research status by performing tech mining analysis. Tech mining helps in identifying the most active authors, countries, affiliations, as well as the evolution and most recent trends in the field. Therefore, the objectives of this research consist of four main parts.

First, to collect, study and analyze the documents published in the field of renewable energy forecasting using artificial intelligence technologies. This step is targeted at providing a comprehensive overview of the field by allocating the main bibliometrics indicators, which is also helping in finding the research gaps.

Second, to study and analyze the meteorological weather and past generation data variables to enhance the selection of input data that will be used for designing, training, and building the renewable energy forecasting models. This step is targeted at reducing costs, reducing complexity, and improving the accuracy of the suggested system.

Third, to further analyze forecasting horizons and resolutions. This step is targeted at analyzing the effects of utilizing different input data resolutions in forecasting accuracy for different forecasting horizons.

Fourth, to study and analyze different algorithms and techniques utilizing different input data. This step is targeted at finding the best algorithms, techniques, and hybrid combinations to assure the highest forecasting accuracy.

Since many dimensions are interconnected and must be considered while designing a forecasting system, this research has four main research questions, each question deals with one (or more) aspects of the design problem and the objectives of this thesis.

The first research question deals with data and data availability problems. There are too many variables that can be used for designing and building renewable energy forecasting systems. So the first research question is:

- Which variables should be used to design, train, and build renewable energy forecasting models to improve the forecasting accuracy while reducing costs and computational complexity?

The second research question deals with forecasting horizon and resolution. Different forecasting horizons and resolutions lead to different forecasting accuracy. Consequently, the second research question is:

- What are the resolutions that can be utilized to design, train, and build renewable energy forecasting systems to assure the highest forecasting accuracy?

As mentioned earlier, most grid operators in the EU require a 15 minutes resolution forecast, thus a secondary research question regarding resolution is:

- Does the regulatory 15-minutes forecasting resolution provide similar accuracy when forecasting wind and solar?

The fourth research question deals with forecasting models and techniques. Different forecasting methods lead to different forecasting accuracy. Based on the application and data availability, the forecasting model selection criteria can be tailored. Also, various algorithms and techniques can be used (together) in building hybrid renewable energy forecasting systems. Consequently, the fourth research question is:

What are the algorithms and techniques to design, train, and build renewable energy forecasting models that can improve the forecasting accuracy based on the available data?

## IV.2. Approaches

The method consists of four main stages as can be seen in Figure 13. Each stage deals with one or more research questions. The first stage is where geographical, meteorological, and past power data are collected. Therefore, in this stage, all the collected variables will be used to build a variable selection model. The function of this model is to select the most suitable variables to perform the forecast while reducing costs and computational complexities. The variable selection model consists of sub-models, where statistical methods like correlation and regression will be used. The input of this model is all the collected variables while the output is the variables that should be used.

In the second stage, the output variables of the first stage will be utilized to build ANN and multiple regression models. Based on the input data method, i.e structural, time-series, or hybrid, six models will be built and tested to forecast solar energy. The input of this model is the output variables of stage one while the output is the forecasted solar energy.

In the third stage, different machine learning techniques will be designed, trained, and tested to forecast wind power using only past generation time-series data. Three main techniques will be tested including ANN, KNN, SVM. Ultimately, a hybrid machine learning model will be designed and tested. Note that wind energy data sets include only past generation data, and hence, only past generation time series data will only be used to build the mentioned models.

In the fourth stage, different input data resolutions will be used to test its effect on output accuracy. In this stage, solar and wind forecasting models will be built, trained, and tested using the ANN forecasting technique. Note that wind power data sets include only past generation data, and hence, time series past generation data of solar and wind will be used in this stage.

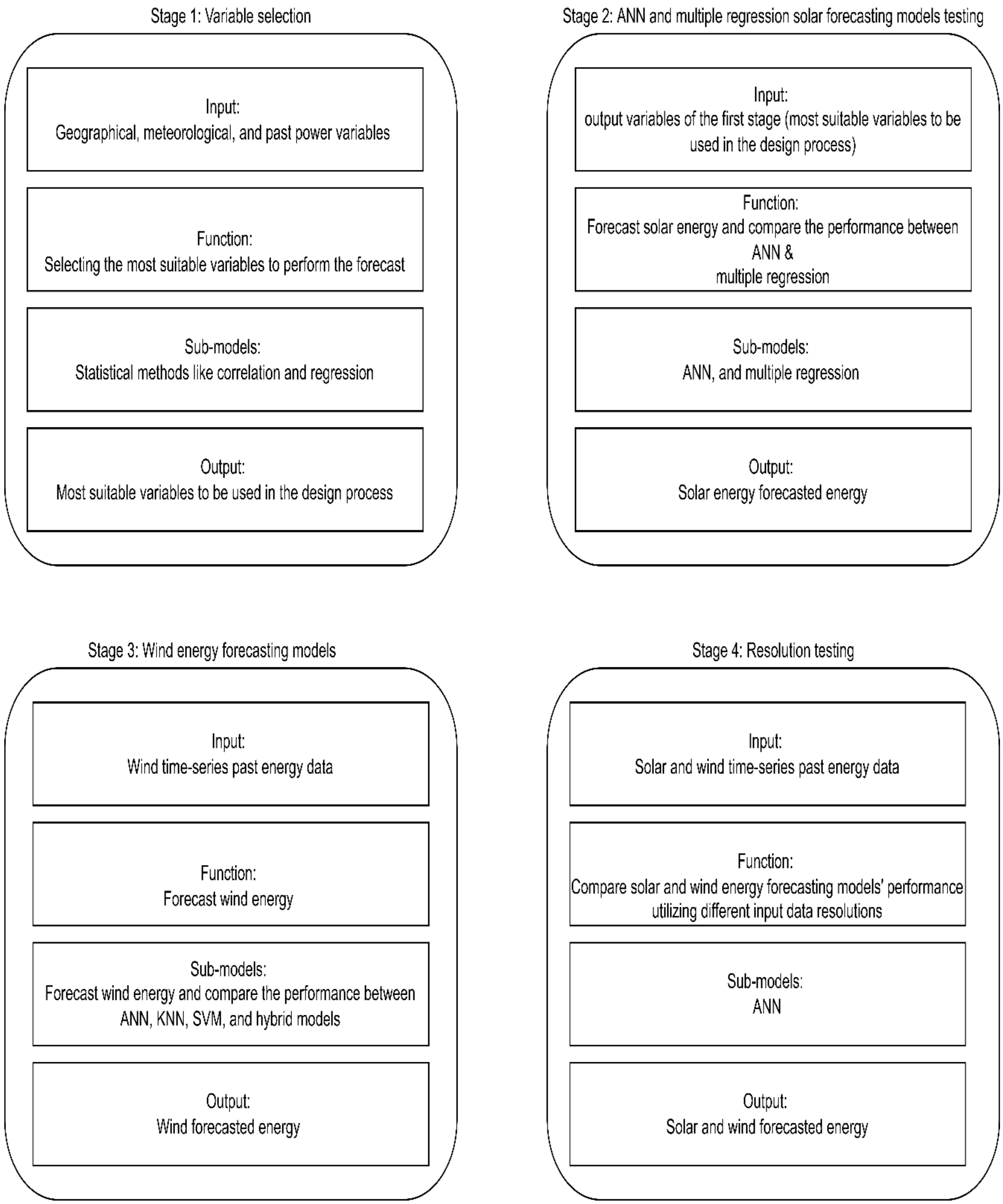


**Figure 13 Methodology stages (source: author)**

## IV.3. Data

To fulfill the goal of this research renewable energy production data, as well as geographical and meteorological parameters that affect energy production are needed. Moreover, the data should be collected over a long time horizon.

The focus of this thesis is grid-connected solar and wind farms, thus the data collection centers around real data provided by such farms and specialized weather modeling firms that provide data for solar and wind farms.

### *IV.3.1. Data collection*

The huge amount of data needed for this research requires collaborations between different data and renewable energy organizations. Thus, different prestigious international firms are collaborating with the authors to provide the necessary data.

The data was collected from three main sources. The first source is grid-connected solar farms where the actual solar energy generation data and some weather parameters were collected. The second source is a grid-connected wind turbine, where the actual wind energy generation data were collected. While the third source is a specialized weather and energy data modeling firm, where the historical solar radiation, geographical and meteorological data were obtained.

Solar energy production data is provided by 3Comm-Hungary[8] through 'SolarEdge Technologies[9], Inc.' platform. By inventing better ways to manage and collect energy produced by solar systems, SolarEdge is one of the world leaders in

---
[8] http://www.3comm.hu/
[9] https://www.solaredge.com/

the solar energy industry with 1.38 million mentored systems and 2400 employees in 130 countries around the globe (solaredge, 2020).

3Comm provided the authors with full access to three PV grid-connected solar farms. The three solar farms are located in Hungary, more specifically, in Szeged, Mindszent, and Kiskunhalas cities, with a peak power of 546 kWp , 547.8 kWp and 546.15 kWp, respectively.

The access was used to collect data needed for training, building, and testing the solar forecasting models. The access allows getting the amount of produced energy from the three sites as well as some weather parameters, so a programming code was written to get the produced energy's data every 15 minutes.

For the site in Szeged, the data was collected since it was established on April 13th, 2017, and the live data is still collected (for future research beyond this thesis), yet the data used in this work is till Aril 18,2020. The same applies to the Mindszent site, which was established on September 10th, 2018, and Kiskunhalas site which was established on January 15th,2019.

Table 10 shows detailed information about the three grid-connected solar sites including location, peak power, established date, data obtained, and other technical information. Note that in this thesis, only data from the Szeged site was used as it is the oldest one, which means more data is available. The data from the other two sites are still collected to be used in future research.

The second source of data is E.ON[10]. E.ON is an international group that focuses on new energy resources, energy networks, renewable energy, and energy solutions. E.ON operates in 13 European countries and plays a key role in the energy market and innovative energy solutions. E.ON provided the authors with energy generation data from a 2 MW wind turbine located in Csetény, Hungary. Table 11 shows some information about the wind turbine like location, peak power, data obtained, and other technical details.

[10] https://www.eon.com/en.html

**Table 10 Detailed information about the three grid-connected solar sites (source: author)**

| Site | Szeged | Mindszent | Kiskunhalas |
|---|---|---|---|
| **Location** | Sándorfalvi Út 10, Szeged,Csongrad Megye,Hungary | Szabadság Utca 92, Mindszent,Csongrad Megye,Hungary | Pirtó Hrsz: 074/22, Kiskunhalas, Hungary |
| **Date of the establishment** | 13/04/2017 | 10/09/2018 | 21/01/2019 |
| **Used data period** | 13/04/2017 - 18/04/2020 | 10/09/2018 - 18/04/2020 | 21/01/2019 - 18/04/2020 |
| **Peak power kWp** | 546 | 547.8 | 546.2 |
| **Data obtained** | Energy generated (every 15 minutes)<br>Humidity, temperature, wind speed (every hour) | Energy generated (every 15 minutes)<br>Humidity, temperature, wind speed (every hour) | Energy generated (every 15 minutes)<br>Humidity, temperature, wind speed (every hour) |
| **PV cells' model** | ND-RJ260 | ND-RB275 | ND-RB275 |
| **Number of modules** | 2100 | 1992 | 1986 |
| **Number of inverters** | 18 | 16 | 18 |

**Table 11 Some information about the wind turbine (source: author)**

| Description | Details |
|---|---|
| Location | Csetény, Hungary |
| Peak power MWp | 2 |
| Data obtained | Energy generated (every 15 minutes) |
| Data collection period | 1/5/2019 till 31/5/2020 |
| Turbine's model | VESTAS V90 |
| Rotor diameter | 90 m |
| Swept area | 6,362 $m^2$ |
| Carbon Footprint[11] | 9.7 g CO2e/kWh |
| Return on energy break-even[11] | 9 months |
| Lifetime return on energy[11] | 26 times |
| Recyclability rate[11] | 82% |

The third source of data is Solcast. Solcast[12] is an international organization specialized in developing data and tools needed for planning, constructing, operating, and managing renewable energy systems. Solcast provided over 3000 clients with data for 1 million locations in Europe, Asia, and North America (Solcast, 2020).

Solcast provided the authors with historical data records for the three sites mentioned earlier in Table 10. The provided datasets include air temperature, cloud

[11] Based on Vestas v90 datasheet, more details: https://www.vestas.com/en/products/2-mw-platform/V90-2-0-MW

[12] https://solcast.com/

opacity, dewpoint temperature, Diffuse Horizontal Irradiance (DHI), Direct Normal Irradiance (DNI), Direct Beam Horizontal Irradiance (EBH), Global Horizontal Irradiance (GHI), precipitable water, relative humidity, snow depth, and wind speed. All the parameters were collected every 15 minutes as will be seen in the next section.

### *IV.3.2. Variables (predictors)*

The data were collected from the three previously-mentioned sources. The collected data consists of energy production data, as well as geographical and meteorological variables.

Since solar cell temperature was not provided by the data sources, it needed to be calculated (more details in the following section).

Table 12 shows the variables used in this study, the table also includes the resolution, type of the variable, unit, and the original source. All meteorological datasets were collected with 15 minutes resolutions between April 13, 2017, and April 18, 2020.

**Table 12 Variables used in the study (source: author)**

| Name | Type | Unit | Source |
|---|---|---|---|
| Air temperature | Historical weather | Celsius | Solcast |
| Cell temperature | Calculated | Celsius | Calculations |
| Wind speed | Historical weather | m/s | Solcast |
| Cloud opacity | Historical weather | Percentage (%) | Solcast |
| Dewpoint temperature | Historical weather | Celsius | Solcast |
| DHI, EBH, DNI, GHI, and GTI | Historical weather | W/m2 | Solcast |
| Precipitable water | Historical weather | Centimeters | Solcast |
| Relative humidity | Historical weather | Percentage (%) | Solcast |
| Snow depth | Historical weather | Centimeters | Solcast |
| PV energy generation | Historical power | Wh | SolarEdge |
| Wind energy generation | Historical power | Wh | E.ON |
| Forecasted PV energy | Predicted variable | Wh | Prediction model output |
| Forecasted wind energy | Predicted variable | Wh | Prediction model output |

It should be noted that the available collected data from the sources have different time-frames, i.e solar energy datasets contain over four years of generation data, while only 13 months of wind generation data. Likewise, weather variables are not available for the wind turbine location. Hence, only the time-series input data method can be performed for wind forecasting as the other two input methods (structural and hybrid) requires additional data (see section II.6.2).

#### *IV.3.2.1. Cell temperature calculation*

It was mentioned in the previous section that the solar cell temperature was not provided by the data sources. Since this variable is critical in forecasting solar energy, it was calculated using the following equation (Mattei et al., 2006) (Trinuruk et al., 2009):

$$T_c = T_a + (T_s - 20) * \frac{solar\ irradiance}{800}$$

**Equation 1**

Where $T_c, T_a$, and $T_s$ are the cell, ambient, and the Standard Test Conditions (STC) temperature in Celsius. The $T_s$ for the PV models used for this study is 25°.

### *IV.3.3. Descriptive Analysis*

Ranging from diagnostic to predictive, different types of data analysis can be conducted. Descriptive data analysis is one of the most straightforward analyses to describe or summarize past and present data. Additionally, descriptive data analysis might be helpful in creating accessible data insights. So, in this section, a descriptive analysis of the collected data is provided.

#### *IV.3.3.1. PV energy data descriptive analysis*

It was mentioned in the previous section that PV past generation data was collected from several grid-connected PV farms. Yet, only one of the farm sites will be used in this study. The technical and other details can be found in Table 10 and Table 12. The past PV energy generation data from the Szeged site was investigated, and the basic descriptive analysis can be found in Table 13. Note that the data analyses were performed using Python, While modeling and forecasting were performed using Matlab. The results from the descriptive analysis clearly show the stochastic nature of PV energy. For instance, the standard deviation is high which indicates data are more spread out from the mean. Another expected fact is that 50% of collected PV energy is less than 1 Wh. This is due to zero energy generation at night and cloudy periods. Another important descriptive measure for PV energy is the most and least frequent value(s). As can be in

Table 14, zero is the most frequent value. This high frequency of zero is expected for the same reason mentioned above. High values (higher than 100 KWh) are infrequent and rarely observed.

**Table 13 Basic descriptive analysis for PV energy data (source: author)**

| Measure | Value |
|---|---|
| Mean | 21957.12 |
| Standard deviation | 34604.22 |
| Median | 1.00 |

| | |
|---|---|
| Mininmum value | 0.00 |
| 25th Percentile | 0.00 |
| 50th Percentile | 1.00 |
| 75th Percentile | 33292.25 |
| Maximum value | 196683.00 |

**Table 14 Most and least frequent values of the generated PV energy(source: author)**

| Generated PV energy (Wh) | Frequency |
|---|---|
| Most frequent | |
| 0 | 55581 |
| 1 | 66 |
| 2 | 48 |
| 3 | 32 |
| 4 | 23 |
| Least frequent | |
| 79096 | 1 |
| 7755 | 1 |
| 103660 | 1 |
| 102318 | 1 |

To better understand the stochastic nature of the PV data, the seasonality distribution might be important, especially for forecasting purposes. Hence, Figure 14, Figure 15, and Figure 16 show the yearly, monthly, and daily average PV energy generation respectively.

The yearly PV energy generation in Figure 14 shows that the PV farm generated between 20000 and 25000 Wh in the past years, and hence it would be expected that the yearly forecasted energy will be within or near that range.

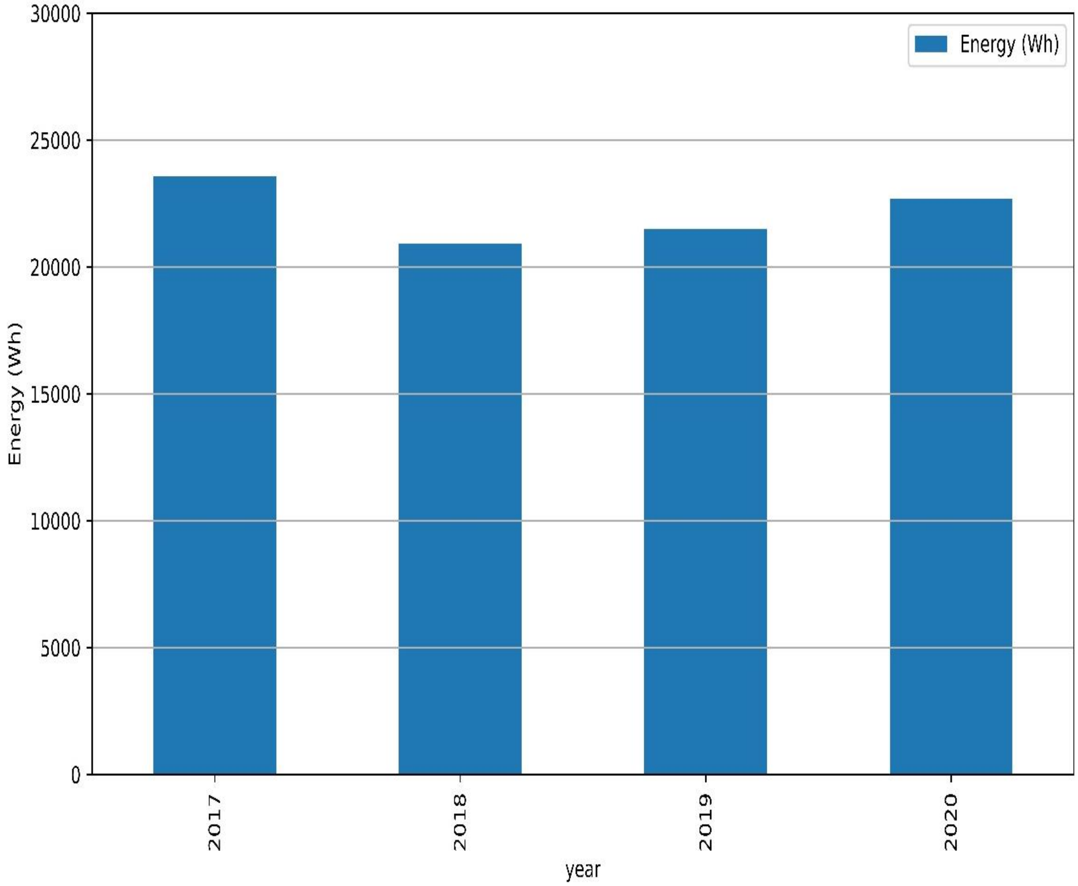


**Figure 14 Yearly mean PV energy generation (source: author)**

The monthly PV energy generation is shown in Figure 15. It can be seen that the highest average generation occurs in July. While the lowest occurs in December and January. For the studied location in Szeged, it can be seen that there are eight months (March till October) where the PV energy generation is around or higher than the yearly average generation (22000 Wh). This indicates that summer, fall, and spring months produce most of the yearly energy.

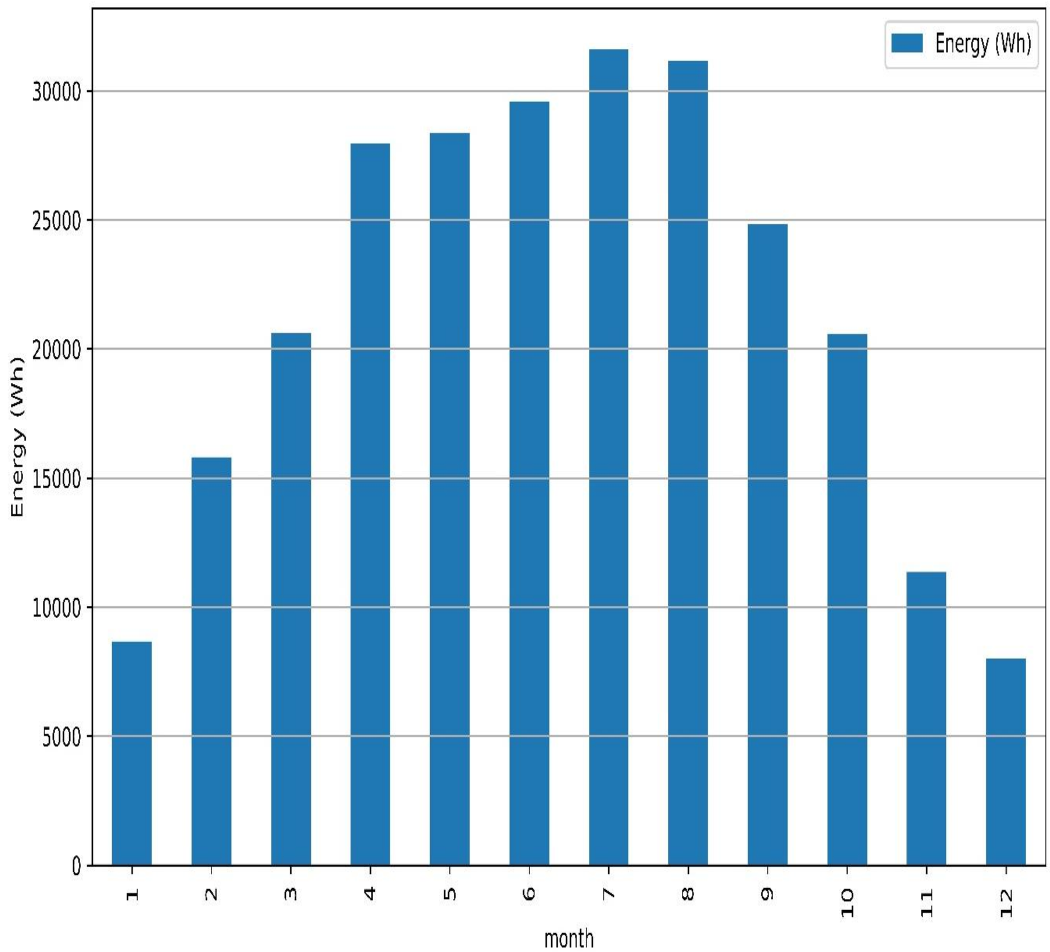


**Figure 15 Monthly PV energy generation (source: author)**

The hourly PV energy generation is shown in Figure 16. It can be seen that the graph follows the typical average hourly PV energy generation where midday hours generate way higher than the average. While in the lean hours' generation is almost zero. As expected, the highest average generation occurs at 11. While the zero generation occurs -normally- between 20:00 and 4:00.

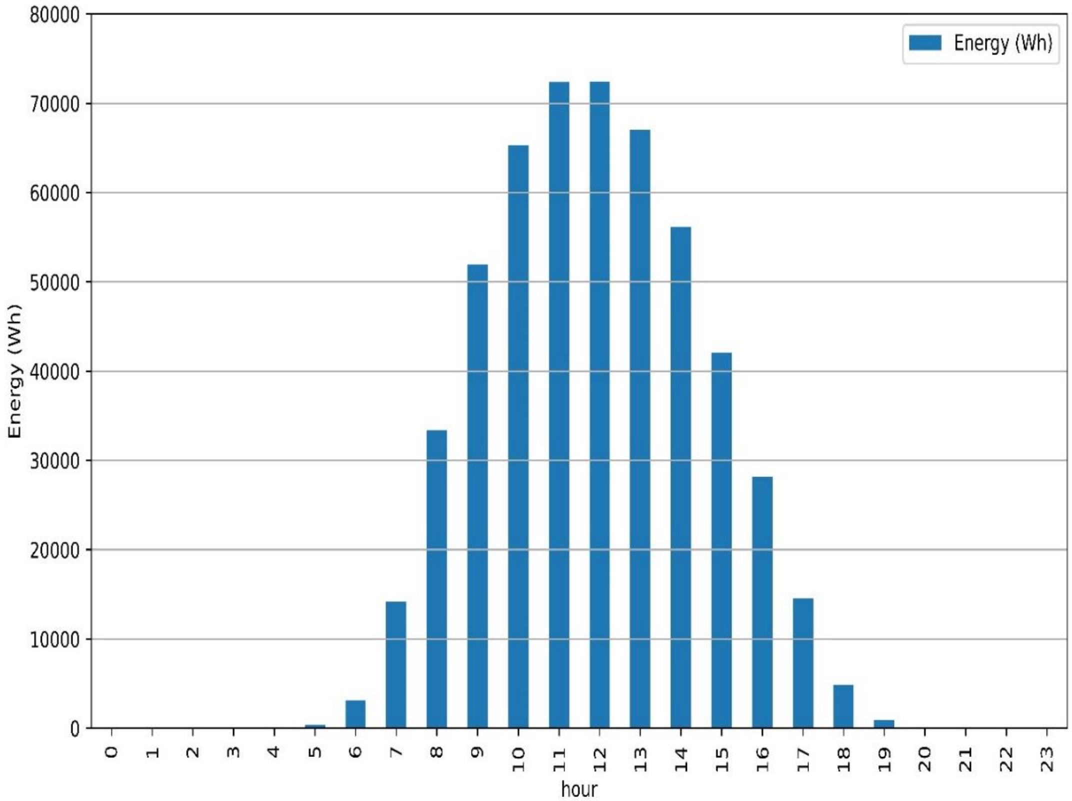


**Figure 16 Hourly PV energy generation (source: author)**

### *IV.3.3.2. Meteorological data descriptive analysis*

The same descriptives applied in section IV.3.3.1 for PV past energy data were applied for the meteorological data. A summary of the basic descriptive analysis for the meteorological data can be seen in Table 15. Note that all the units and details of these variables were mentioned earlier in Table 12.

**Table 15 A summary of the basic descriptive analysis for the meteorological data (source: author)**

| Variable | Mean | Standard deviation | Median | Min. value | 25th | 50th | 75th | Max. value |
|---|---|---|---|---|---|---|---|---|
| Air temperature | 13.3 | 9.1 | 13.6 | -17.7 | 5.9 | 13.6 | 20.3 | 37.7 |
| Cloud Opacity | 25.8 | 31.1 | 4.9 | 0.0 | 0.0 | 4.9 | 49.9 | 100.0 |
| Dewpoint temperature | 6.9 | 7.0 | 7.4 | -20.1 | 1.4 | 7.4 | 12.5 | 24.2 |
| Dhi | 72.9 | 108.9 | 3.0 | 0.0 | 0.0 | 3.0 | 118.0 | 605.0 |
| Dni | 152.8 | 277.2 | 0.0 | 0.0 | 0.0 | 0.0 | 162.0 | 1005.0 |
| Ebh | 87.0 | 180.9 | 0.0 | 0.0 | 0.0 | 0.0 | 42.0 | 840.0 |
| Ghi | 160.0 | 243.1 | 3.0 | 0.0 | 0.0 | 3.0 | 254.0 | 946.0 |
| Gti Fixed Tilt | 181.3 | 285.5 | 2.0 | 0.0 | 0.0 | 2.0 | 262.0 | 1097.0 |
| Gti Tracking | 198.9 | 300.6 | 3.0 | 0.0 | 0.0 | 3.0 | 328.0 | 988.0 |
| Precipitable Water | 19.2 | 9.0 | 18.2 | 2.1 | 11.7 | 18.2 | 25.9 | 48.20 |
| Relative Humidity | 68.4 | 17.9 | 70.0 | 16.6 | 55.2 | 70.0 | 83.5 | 100.0 |
| Snow Depth | 0.1 | 0.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 4.0 |
| Wind Speed | 3.1 | 1.5 | 2.8 | 0.0 | 2.10 | 2.8 | 3.9 | 14.8 |

The results from the descriptive analysis for the meteorological show different nature of each variable in terms of having stochastic nature. For instance, the standard deviation is high for all solar radiation variables, meanwhile, it is lower for some other variables like air temperature. Generally, Table 15 shows that the PV site has a mild climate. The mean temperature is 13.3 C° and only 25% of the times temperature fall below 5.9. All other variables like humidity, wind speed, and snow depth do not show any signs of extreme weather conditions for long periods.

It should be noted that although all variables in Table 12 and Table 15 are worth being further analyzed, the aim of this study is not to deeply study the meteorological data. Rather, studying meteorological variables aims at helping to understand weather patterns that might affect renewable energy generation. To that end, only key weather variables such as air temperature, gti fixed tilt, and wind speed will be further analyzed, as these variables can summarize weather patterns that affect PV energy generation in the given PV site.

Hence, for the three above-mentioned variables, the most and least frequent value(s) were found. As can be in Table 16, 20.4 is the most frequent air temperature. Other values such as 16.9, 16.5,16.8, and 20.2 are among the most frequent. This shows that the PV farm site has a mild climate where no hot or cold temperatures are recorded often. This claim can be confirmed by looking at the least frequent temperatures. Extreme values such as -17.2, -15.1, -14.9, -13.4, and -16.9 are rarely occurred (with each recorded only once).

The Gti fixed-tilt shows a very similar distribution as the PV energy one. This similar behavior is expected knowing that Gti fixed-tilt is the most important variable for PV energy generation. Zero value is the most frequent value, while high values (higher than 1000 w/m2) are infrequent and rarely observed.

Wind speed distribution also confirms that the site has a mild climate. 2.6 m/s is the most frequent speed. All the top five most frequent speeds are around 2.5 m/s. Extreme wind speeds above 13 are so rare yet it was recorded several times.

**Table 16 Most and least frequent values of some meteorological variables (source: author)**

| Air temperature (C°) | Frequency | Gti fixed tilt (W/m2) | Frequency | wind speed (m/s) | Frequency |
|---|---|---|---|---|---|
| Most frequent | | | | | |
| 20.4 | 469 | 0 | 54920 | 2.6 | 4150 |
| 16.9 | 465 | 1 | 359 | 2.4 | 4031 |
| 16.5 | 462 | 2 | 332 | 2.3 | 4011 |
| 20.2 | 462 | 3 | 314 | 2.5 | 4000 |
| 16.8 | 461 | 4 | 287 | 2.7 | 3863 |
| Least frequent | | | | | |
| -17.2 | 1 | 1081 | 1 | 12.0 | 2 |
| -15.1 | 1 | 1084 | 1 | 14.3 | 1 |
| -14.9 | 1 | 1097 | 1 | 11.8 | 1 |
| -13.4 | 1 | 1094 | 1 | 13.2 | 1 |
| -16.9 | 1 | 1087 | 1 | 13.3 | 1 |

To better understand the nature of the three meteorological variables, the seasonality distribution might also be helpful, especially for forecasting purposes. Hence, figures 18-26 show the yearly, monthly, and daily mean values for temperature, gti fixed tilt, and wind speed.

The yearly mean air temperature in Figure 17 shows that the average temperature ranges between 16 and 13 C° in the past years. The slightly lower temperature in 2020 is because the data was collected till April 18, 2020, and thus the summer of this year is not included in the average.

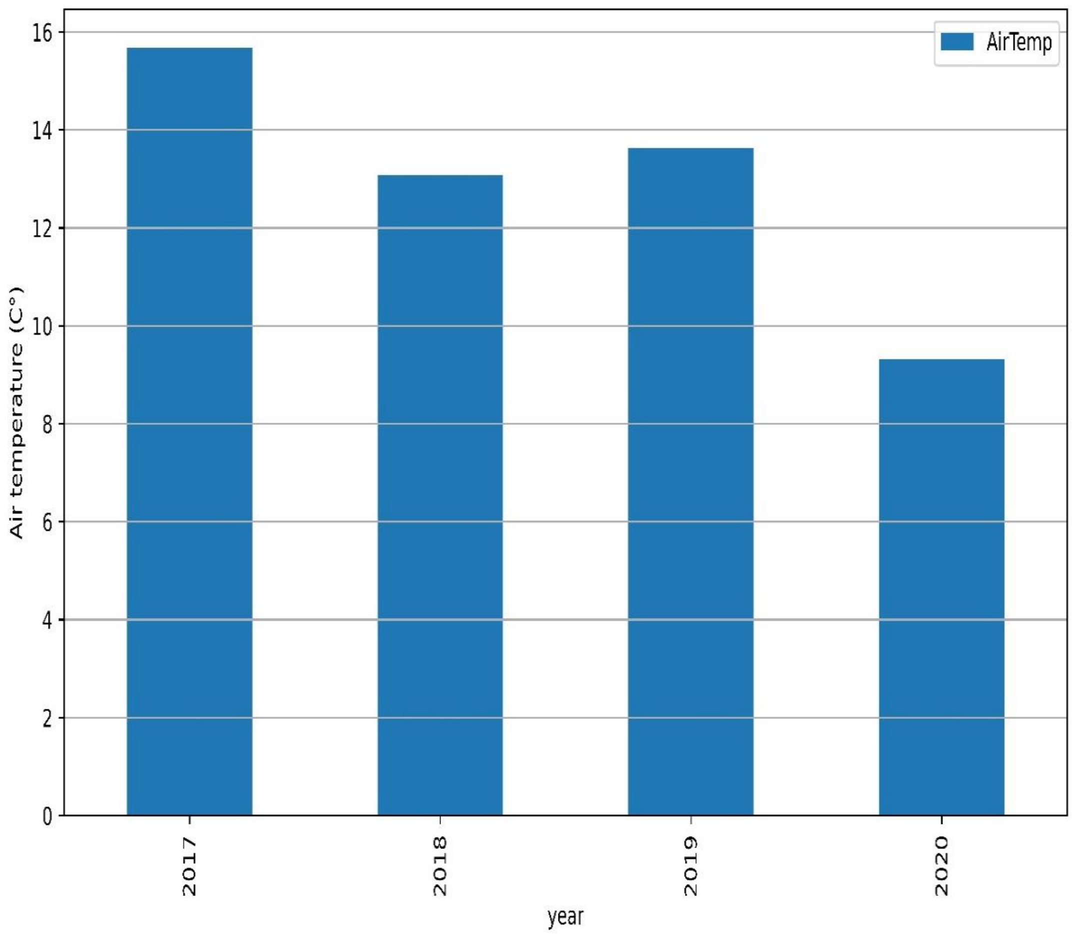


**Figure 17 Yearly mean temperature (source: author)**

The monthly mean air temperature is shown in  Figure 18. It can be seen that the highest mean temperature occurs in August.  While the lowest occurs in January. For the studied location in Szeged, it can be seen that there are seven months (April till October) where the average temperature is higher than 10 and nine months higher than 5. This indicates that the site has a mild climate.

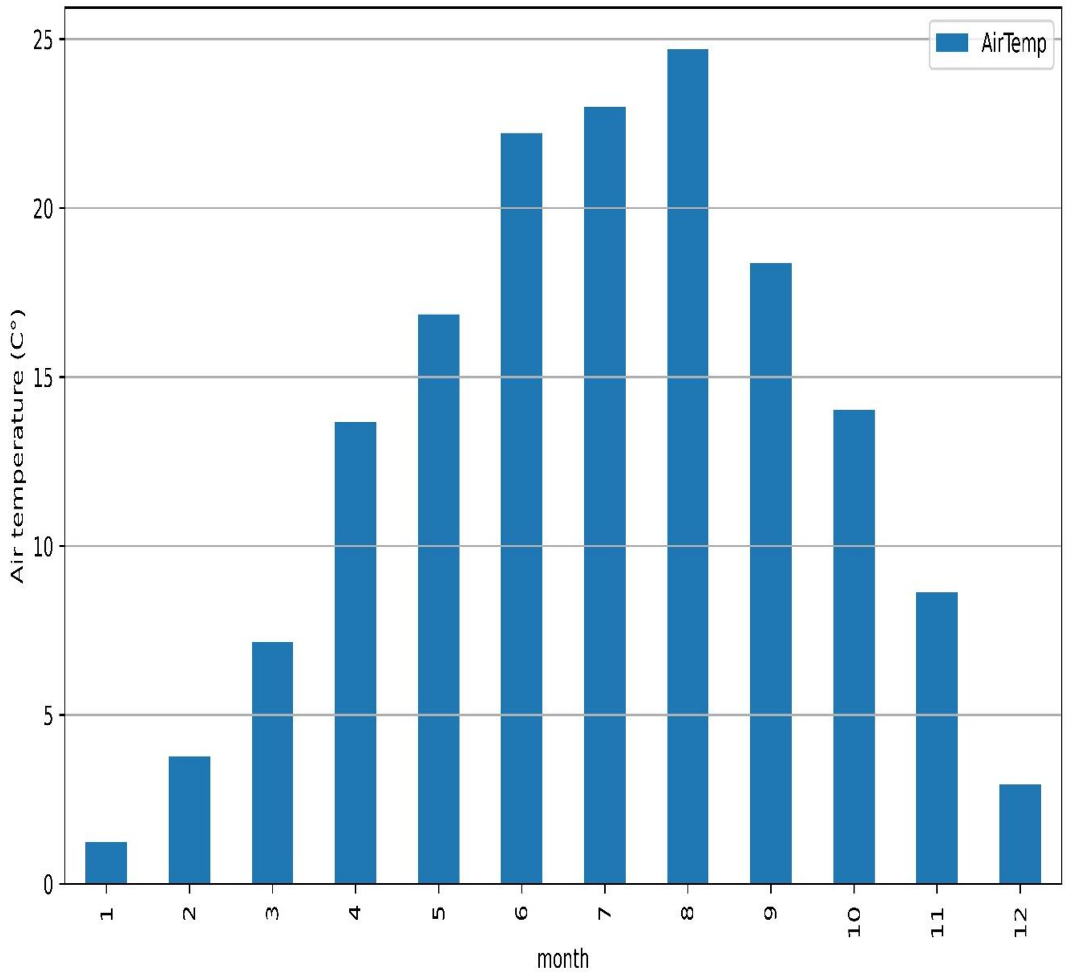


**Figure 18 Monthly mean temperature (source: author)**

The hourly air temperature is shown in Figure 19. It can be seen that the highest average temperatures occur between 13:00 and 15:00. The lowest average temperatures occur between 3:00 and 5:00.

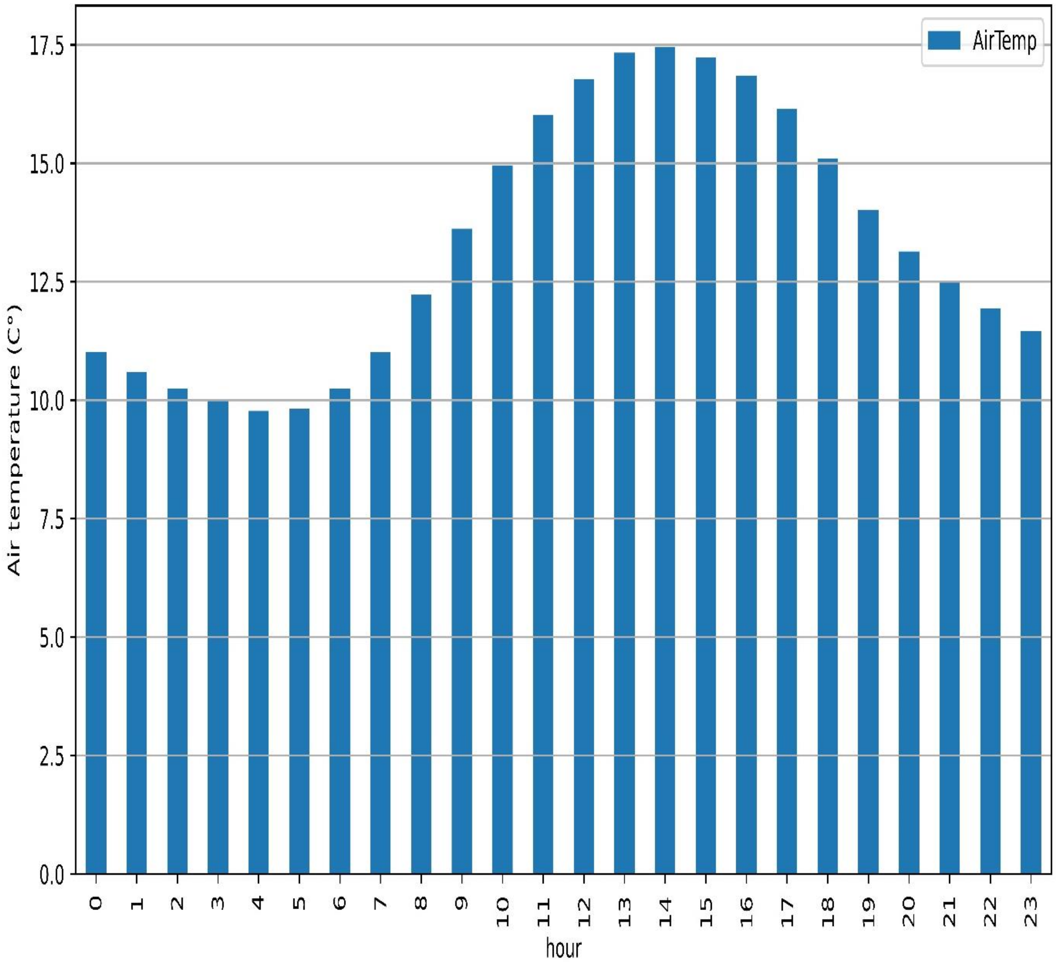


**Figure 19 Hourly mean temperature (source: author)**

The yearly mean Gti fixed tilt in Figure 20 shows that the average Gti fixed-tilt values range between 170 and 200 w/m2 in the past years.

The monthly Gti fixed-tilt values are shown in  Figure 21. . It can be seen that the highest average Gti fixed-tilt occurs in August.  While the lowest occurs in December and January. It can be also seen that there are six months (April till September) where the Gti fixed-tilt value is around or higher than the yearly average generation (181 w/m2).

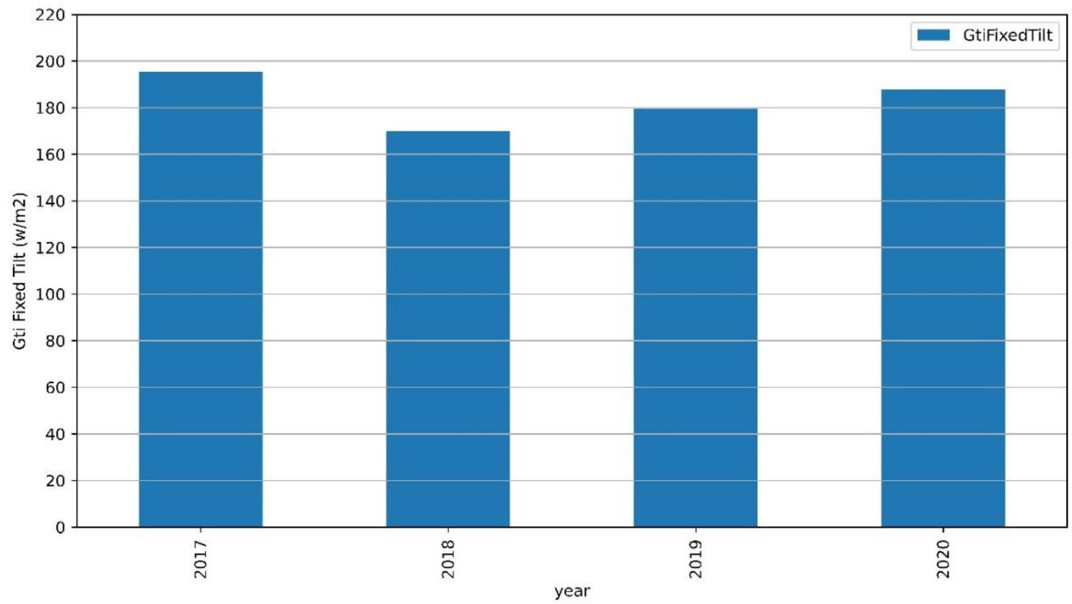


**Figure 20 Yearly mean Gti Fixed Tilt (source: author)**

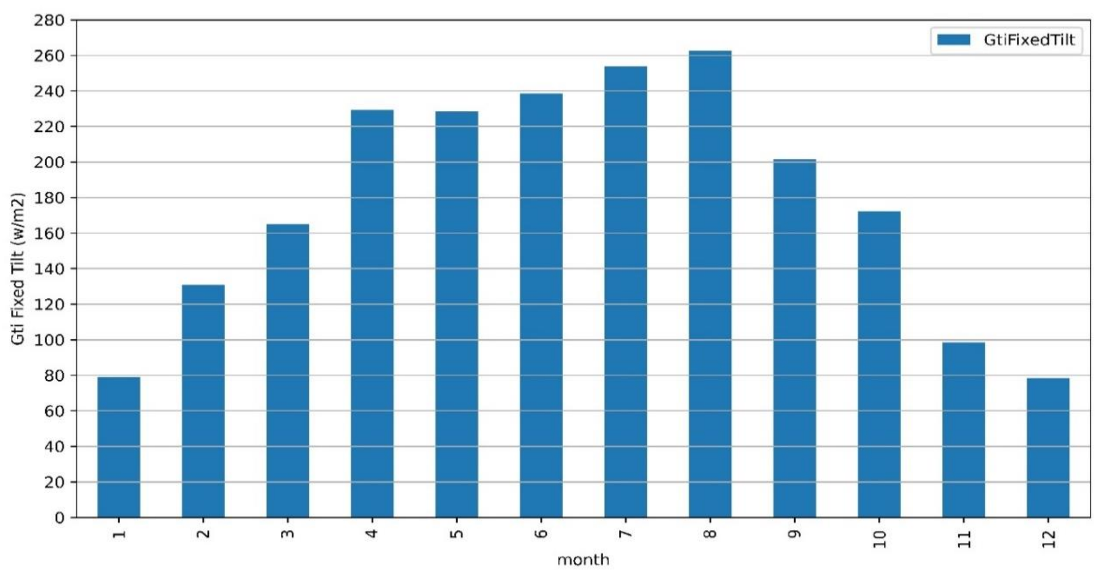


**Figure 21 Monthly mean Gti Fixed Tilt (source: author)**

The mean hourly Gti fixed-tilt values are shown in  Figure 22. It can be seen that the graph follows the typical average hourly Gti where midday hours have way higher values than the average.  While in the lean hours' generation is almost zero. As expected, the highest average Gti fixed-tilt value occurs at 11. While the zero generation occurs -normally- between 20:00 and 4:00.

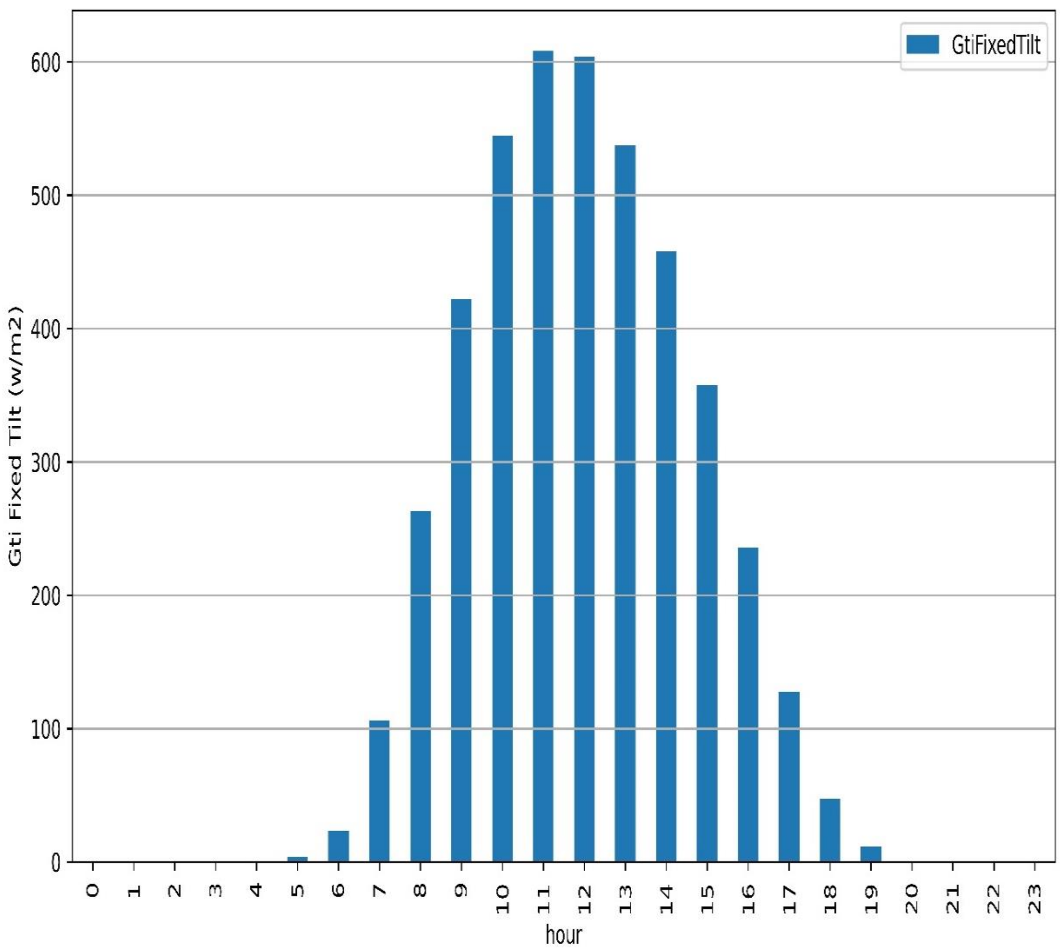


**Figure 22 Hourly mean Gti Fixed Tilt (source: author)**

The yearly mean wind speed in Figure 23 shows that the average wind speed ranged between 3 and 3.5 m/s in the past years.

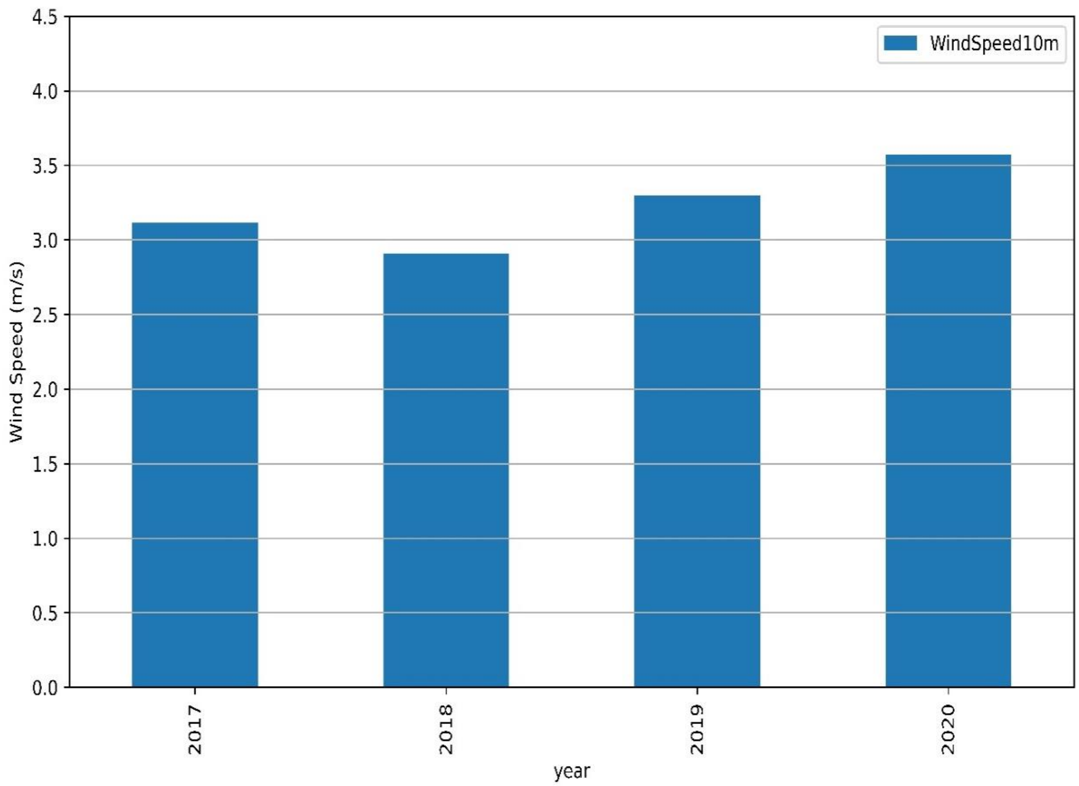


**Figure 23 Yearly mean wind speed (source: author)**

The monthly mean wind speed is shown in Figure 24. It can be seen that the highest mean wind speed occurs in February. While the lowest occurs in June and August. Generally, the graph shows that summer months are less windy than other seasons. Nevertheless, the site has very mild wind speeds and most months have monthly wind speeds close to the mean average wind speed (3.17 m/s).

The hourly wind speed is shown in Figure 25. It can be seen that the daytime is relatively windier than the nighttime. Wind speed tends to have higher values from 8:00 till 15:00.

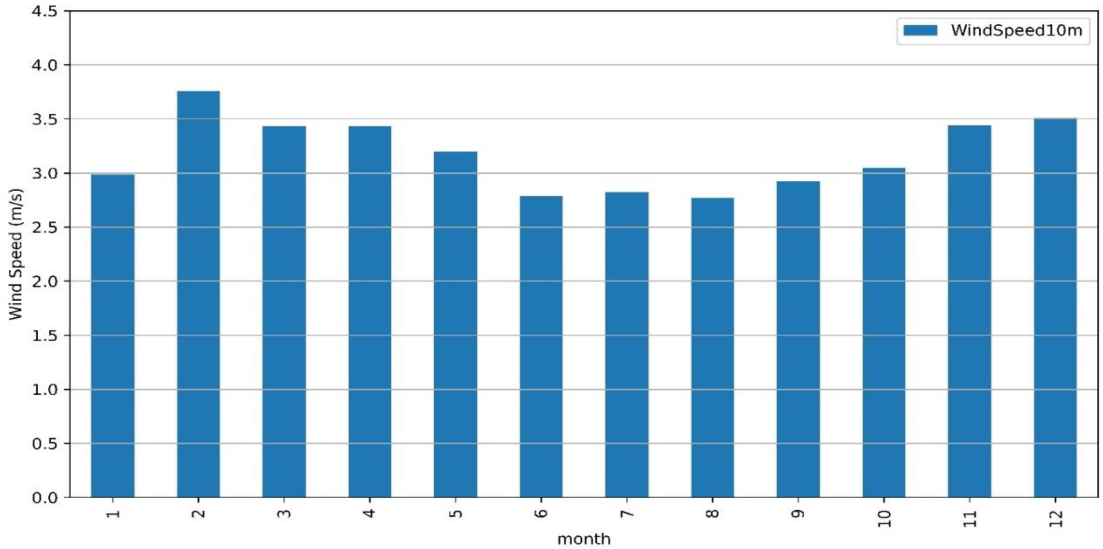


**Figure 24 Monthly mean wind speed (source: author)**

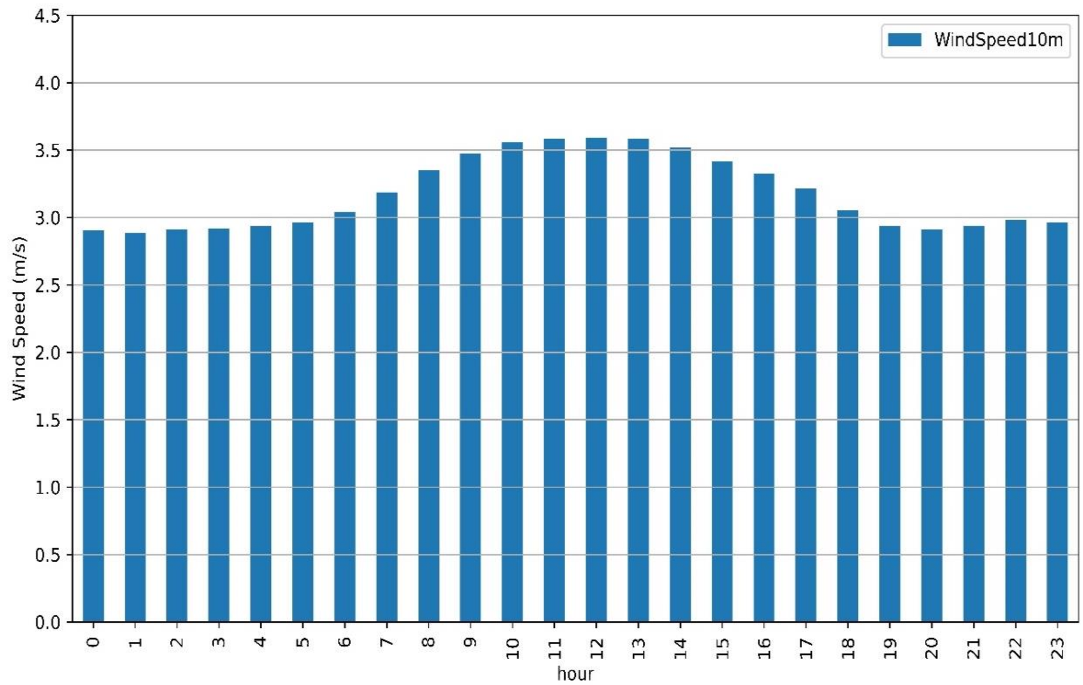


**Figure 25 Hourly mean wind speed (source: author)**

#### *IV.3.3.3. Wind data descriptive analysis*

The past wind energy generation data from the Csetény site was investigated, and the basic descriptive analysis can be found in Table 17. The results from the descriptive analysis clearly show the stochastic nature of wind energy. For instance, the standard deviation is high which indicates data are more spread out from the mean. Another fact is that 50% of collected wind energy is less than 60 KWh.

**Table 17 Basic descriptive analysis for wind data (source: author)**

| **Measure** | **Value** |
|---|---|
| Mean | 126.07 |
| Standard deviation | 149.52 |
| Median | 60.1 |
| Minimum value | 0.00 |
| 25th Percentile | 3.80 |
| 50th Percentile | 60.10 |
| 75th Percentile | 203.30 |
| Maximum value | 582.80 |

The most and least frequent wind energy generation value(s) were also found. As can be in Table 18, zero is the most frequent value. Yet, high wind energy generation values (higher than 400KWh) are among the most and least frequent list. This indicates that wind energy has an even more stochastic nature than PV energy, and high random values might occur often which makes it harder to forecast compared to PV energy.

**Table 18 The most and least frequent wind energy generation (source: author)**

| **Generated wind energy (KWh)** | **Frequency** |
|---|---|
| Most frequent | |
| 0 | 8934 |
| 488.4 | 78 |
| 488.1 | 61 |
| 1.0 | 56 |
| 1.1 | 49 |
| Least frequent | |
| 332.8 | 1 |
| 319.6 | 1 |
| 493.7 | 1 |
| 419.1 | 1 |
| 490.3 | 1 |

To better understand the stochastic nature of the wind generation data, the seasonality distribution might be important, especially for forecasting purposes. Hence, Figure 26, Figure 27, and Figure 28 show the yearly, monthly, and daily average wind energy generation respectively.

The yearly wind energy generation in Figure 26 shows that the wind farm generated between 100 and 150 KWh in the past years. Yet the data was collected from May 2019 till June 2020, thus for both years, many monthly data are not provided. More monthly and yearly data is needed to make a firm conclusion regarding the yearly mean data.

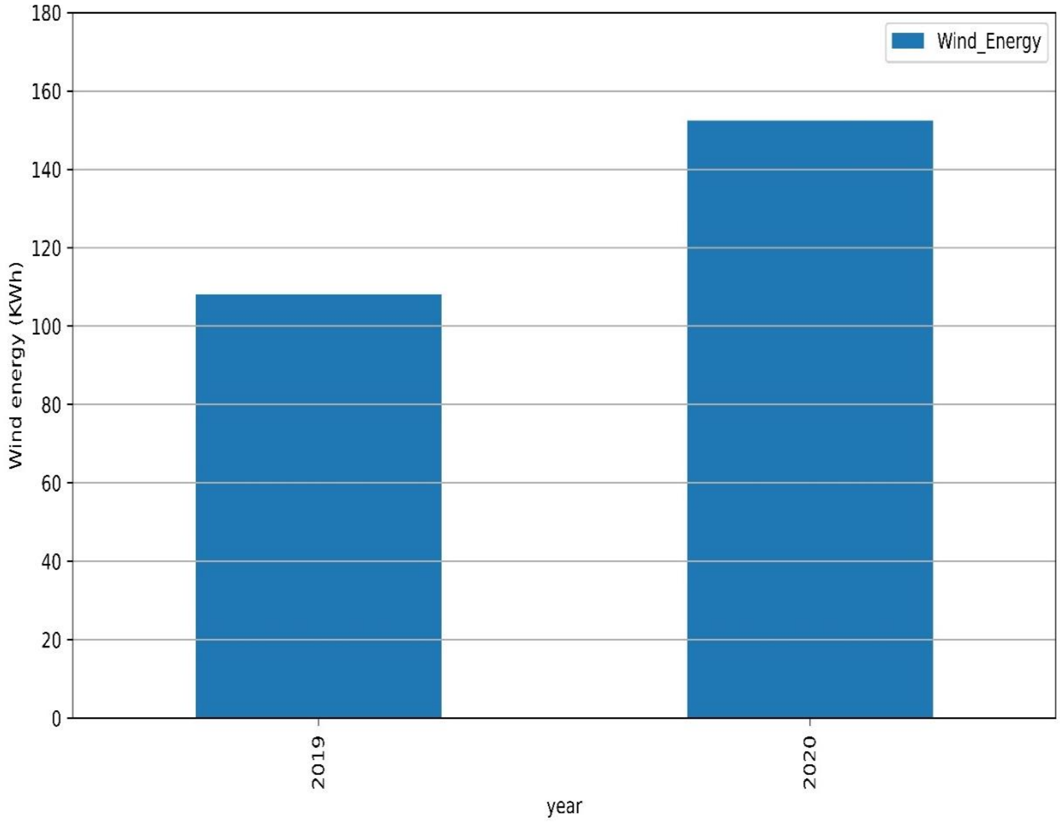


**Figure 26 Yearly mean wind energy generation (source: author)**

The monthly mean wind energy generation is shown in Figure 27. It can be seen that the highest average generation occurs in February. While the lowest occurs in June and August. For the studied location in Csetény, it can be seen that there are four months (February, March, May, and December) where the wind energy generation is around or higher than the yearly average generation (126 KWh). It can be also noticed that these months belong to different seasons.

The hourly mean wind energy generation is shown in Figure 28. It can be seen that the average hourly wind energy generation does not have a firm peak hour(s), yet it tends to have lower values during the mornings between 6:00 and 11:00.

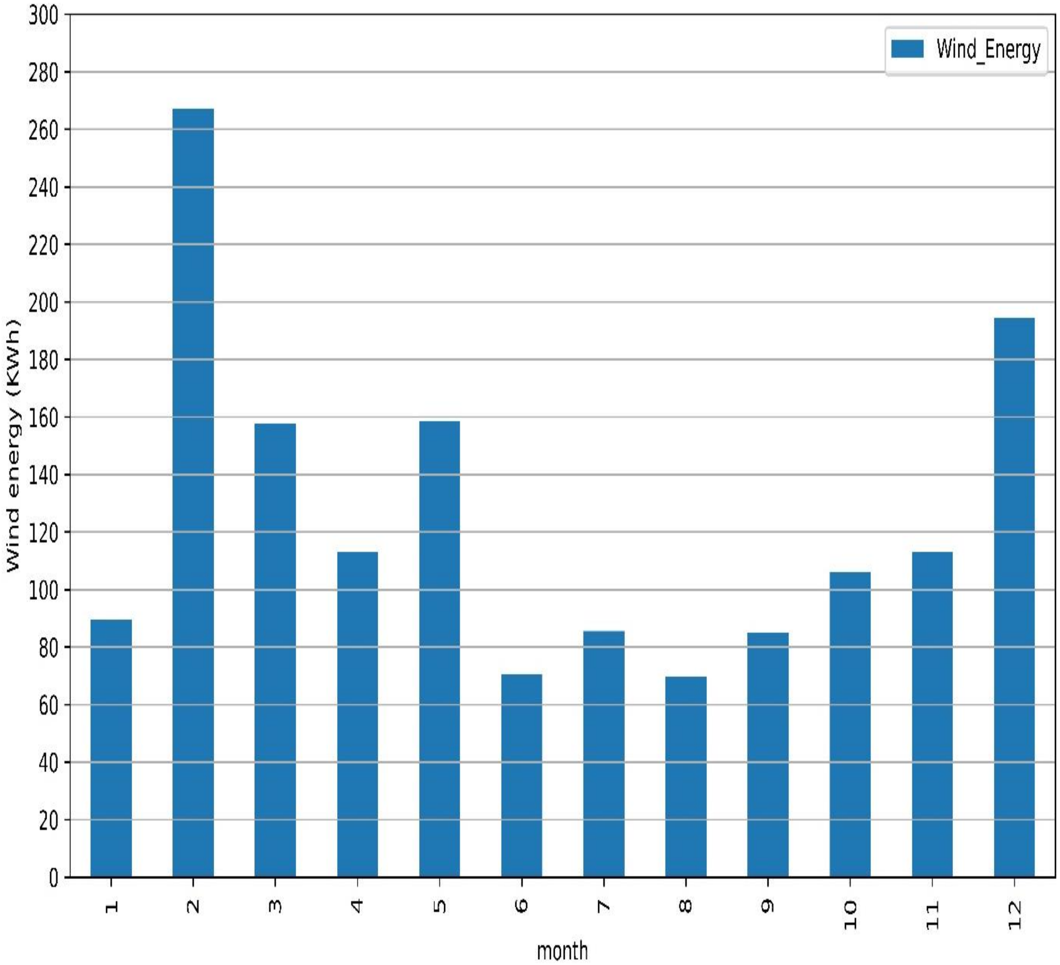


**Figure 27 Monthly mean wind energy generation (source: author)**

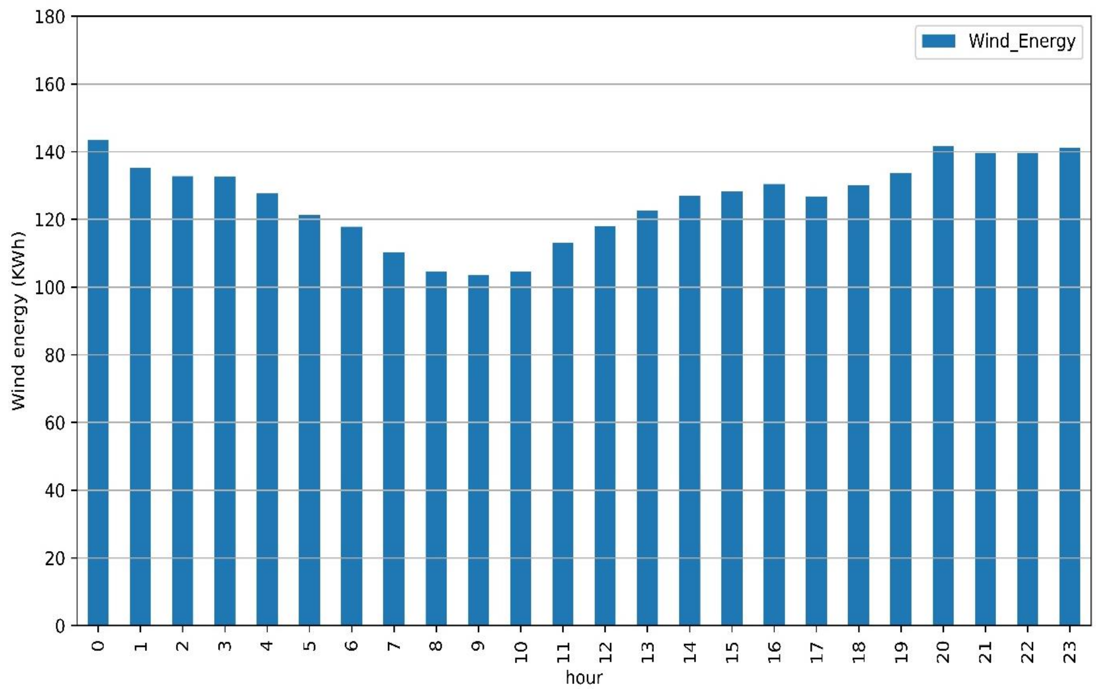


**Figure 28 Hourly mean wind energy generation (source: author)**

### IV.4. Evaluation methods

To evaluate the performance of the forecasting model, an evaluation method is needed. The accuracy of the forecasting models can be evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), Coefficient of Determination (COD), and Error (ε). Equations 2-5 summarize the evaluation methods used in this study (Ahmed et al., 2020) (Elsheikh et al., 2019), where n is the number of observations, $xt$ and $yt$ are the observed (real) and the forecasted output power values at time $t$, respectively, and $\overline{x}$ is the average of the observed values.

$$\text{Mean Absolute Error (MAE)} = \frac{1}{n} * \sum_{i=1}^{n} |xt - yt|$$

**Equation 2**

$$\text{Mean Squared Error (MSE)} = \frac{1}{n} * \sum_{i=1}^{n} (xt - yt)^2$$

**Equation 3**

$$\text{Coefficient of Determination } (R^2) = \frac{\sum_{i=1}^{n} (xt - yt)^2}{\sum_{i=1}^{n} (xt - \bar{x})^2}$$

**Equation 4**

$$\text{Error } (\varepsilon) = xt - yt$$

**Equation 5**

MAE is a quantity that is used in order to measure the closeness of the predicted values to the measured values. MSE measures the average of the squares of the errors – and thus embodies not only how widely the estimates are spread from the real sample but also how far off the average estimated value is from the true value. COD has been used to show how close prediction model results are to the actually measured data line as a fitted regression line (also known as R-squared ($R^2$) score and is generally used for testing hypotheses). Error is the actual (not absolute) value of the difference between the estimation and the corresponding actual value. RMSE is calculated to measure the prediction of a given approach thus showing the so-called scattering level produced by the model. For higher modeling accuracy, MAE, MSE, ε, and RMSE indices should be closer to zero but the COD value should be closer to 1.

As each of these measures has a different scale, and different high and low, it is hard to compare the model's performance under different scales. Different scales with different possible highs and lows for each performance measure might be confusing.

Evaluating forecasting models based on different scaled measures might lead to biased conclusions. One of the methods used to avoid bias is to normalize the performance measures (Poli & Cirillo, 1993) (Haghverdi et al., 2018). Thus, all of the mentioned measures were normalized to have a value between 0 and 1, where 0 represents the lowest errors (best possible performance) and 1 the highest errors (worst possible performance) as shown in Equation 6:

$$\text{M normalized} = \frac{(M - M\ minimum)}{(M\ maximum - M\ minimum)}$$

**Equation 6**

Where M represents the values of each performance measure. For all measures, higher values mean worse performance, but for COD it is the opposite. Therefore, in order to make COD on the same scale where 0 is the best possible performance, Equation 7 was applied.

$$\text{Scaled COD} = |1 - COD|$$

**Equation 7**

Where Scaled COD is the adjusted COD. In Equation 7, the lowest possible value is 0 which represents the best possible performance. 1 is the highest possible value and represents the worst possible performance.

To evaluate each Ml technique, all of the 4 performance measures will be used. The 4 measures will be used to create a method score (MS). So, for each method, the average value of the 4 normalized measures will be computed and used as an indicator for the method’s performance, as can be found in Equation 8. The MS will be used by the suggested hybrid algorithm to evaluate each method tested for wind energy forecasting.

$$MS = \frac{(MAE\ normalized + MSE\ normalized + \varepsilon\ normalized + Scaled\ COD)}{4}$$

**Equation 8**

## IV.5. Multiple regression models

A general multiple linear regression model can be denoted as follows:

$$y = \beta_0 + \beta_1 v_1 + \beta_2 v_2 + \beta_3 v_3 + \beta_4 v_4 + \cdots + \beta_n v_n + \varepsilon$$

**Equation 9**

Where $v_1$, $v_2$, …, $v_n$ are the input variables (1 to n). The coefficient $\beta_0$ is the intercept, while values of $\beta_1$, …, $\beta_n$ denote the slope coefficient of each input (explanatory) variable, and $\varepsilon$ is the error (the amount by which the predicted value is different from the actual value). The regression model estimates the best values of $\beta_1$, …, $\beta_n$ leading to least error $\varepsilon$.

Given all or part of the explanatory variables (Table 12), the multiple regression model can be used to forecast PV output power ($y$) for a given time $t$. To forecast future values of $y$ within the forecast horizon (i.e. $y_{(t+1)}, y_{(t+2)}, \cdots, y_{(t+h)}$, where $h$ is the number of time periods between the present and the effective time of predictions), past values of $v$ are required. Since the resolution is 15 minutes, to forecast one day ahead $h$ is set to 96 time periods. Consequently, for a given time $t$, $y$ is predicted using $v$ values from time $(t - h)$. Notice, forecasting should start on the second day of data collection to avoid negative times (see section IV.7).

In structural method, the input variables fed to the multiple regression model are only the meteorological and geographical variables – with low correlation variables left out as will be seen in section V.1. Past values of selected weather variables were utilized according to Equation 10 which denotes the Structural Multiple Regression model (SMR):

$$y_{(t)} = \beta_0 + \beta_1 v_{1(t-h)} + \beta_2 v_{2(t-h)} + \beta_3 v_{3(t-h)} + \cdots + \beta_n v_{n(t-h)} + \varepsilon$$

**Equation 10**

PV output power can be forecasted by knowing its past values (Cococcioni et al., 2011). Therefore, for the time-series method only PV output power values were used as input such that actual past PV output power values were used to predict future values ($y(t)$) of the output power according to Equation 11 which denotes the Time-series Multiple Regression model (TMR):

$$y(t) = \hat{\beta}_0 + \hat{\beta}_1 y_{(t-h-0)} + \hat{\beta}_2 y_{(t-h-1)} + \cdots + \hat{\beta}_h y_{(t-h-95)}) + \varepsilon$$

**Equation 11**

Where $\hat{\beta}_1$, …, $\hat{\beta}_h$ denote the slope coefficient of each input (explanatory) variable, the hat symbol used here just to indicate that each regression model has its unique beta values. $y_{(t-h-0)}, y_{(t-h-1)}, \ldots, y_{(t-h-95)}$ are the past PV power values starting from 24 hours before the forecasting takes place (i.e. 24 hours before the 1st prediction) and goes until 48 hours before the 1st prediction. In other words, $y_{(t-h-0)}$ is the past PV power value 24 hours (96 time period) before forecasting $y_{(t)}$, $y_{(t-h-1)}$ represents the past PV power value 97 time period before forecasting $y_{(t)}$, and $y_{(t-h-95)}$ represents the past PV power value 191 time period before forecasting $y_{(t)}$. So, for predicting $y_{(t)}$, 96 past PV power values are utilized by the model. Note that to forecast $y_{(t)}$, the past PV power values 48 hours prior to the forecast are required. This ensures that there is no overlap between the prediction horizon period (of 24 hours) and the data representing the actual power generated, which becomes available at the end of each day.

Finally, in the hybrid method past values of all the variables, including actual past values of the power variable as well as past values of the weather variables were fed to the model to predict future PV output power values as shown in Equation 12 representing the Hybrid Multiple Regression model (HMR):

$$y(t) = \tilde{\beta}_0 + \beta_1 v_{1(t-h)} + \beta_2 v_{2(t-h)} + \beta_3 v_{3(t-h)} + \cdots + \beta_n v_{n(t-h)} + \hat{\beta}_1\, y_{(t-h-0)} + \hat{\beta}_2\, y_{(t-h-1)} + \cdots + \hat{\beta}_h\, y_{(t-h-95)}) + \varepsilon$$

**Equation 12**

Where $\tilde{\beta}_0$ is the intercept for the HMR model.

## IV.6. Machine learning models

Several machine learning forecasting models will be used in this work. ANN, SVM, KNN forecasting models will be used. Note that some models like ANN will be used in both PV and wind energy forecasting. SVM and KNN models will be used only for wind energy forecasting. The reason behind utilizing different models for PV and wind is mainly the data availability. As mentioned in section IV.3, PV data sets include many meteorological and past energy variables, while the wind dataset includes only past energy data. Plus, the available datasets have different timeframes where PV past generation data was collected over a longer horizon.

### *IV.6.1. ANN forecasting model*

ANN is a network of “neurons” that are arranged in a layered structure. Figure 29 shows a simple diagram of ANN where input variables arrive from the bottom, while the forecasted variable(s) (output) appears at the top layer. An ANN also includes one or more hidden layers and hidden neurons. Such an ANN structure where the information flow is directed in one direction only (from the bottom to the top layer) is called Multi-Layer Feed-Forward Neural Network (MLFFNN). Note that in this work MLFFNN networks are used. In order to evaluate the performance of the training algorithm performance, the error (difference between the MLFFNN output and the real measured output) is determined using MSE. To minimize MSE values between the observed real output and the forecasted output back-propagation algorithm is used where MSE values are utilized to update the weights and the biases of the network.

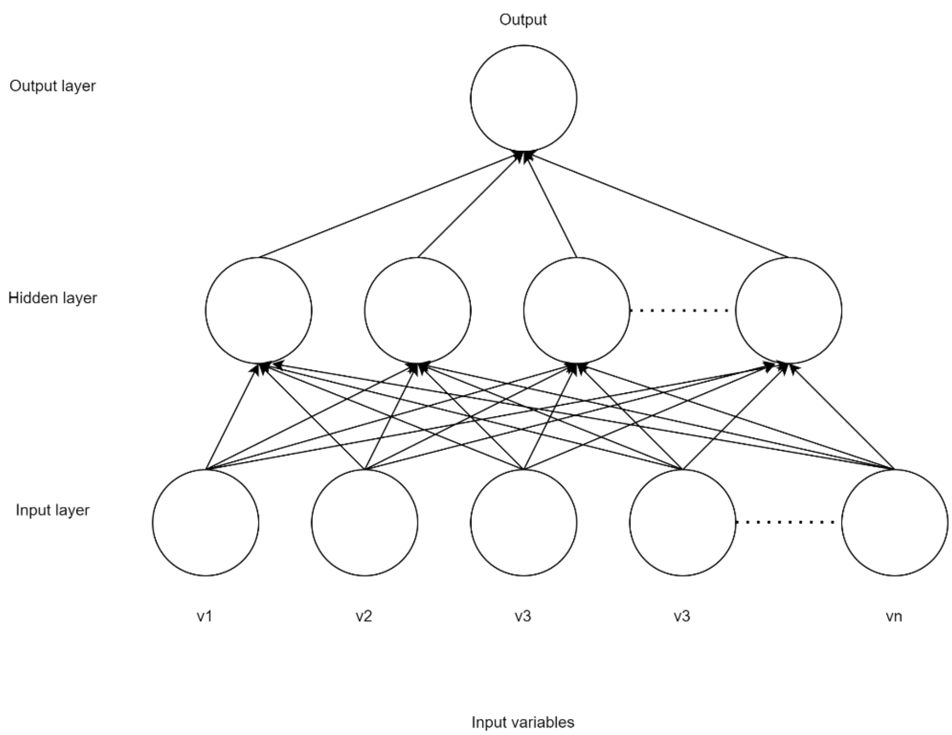


**Figure 29 A neural network with n inputs and one output (source: author)**

In MLFFNN, each layer receives its inputs from the previous layer (except for the input layer), so the outputs of a certain layer are the inputs to the next one. Inputs to each neuron in a given (hidden or output) layer are combined using a weighted linear combination. Then a nonlinear activation (transfer) function φ modifies the results before it is ready to be output. This network structure has many advantages for this forecasting context as this structure works well with big data and provides quick predictions after training. Moreover, it can be applied to solve complex non-linear problems and same accuracy ratio can be achieved even with even smaller data (Khishe et al., 2018) (Akkaya & Çolakoğlu, 2019). Figure 30 shows the flow of information in one artificial neuron, where $w_1$, $w_2$,…, $w_n$ are the weights corresponding to input data $v_1$, $v_2$,..., $v_n$ respectively (Kim, 2017).

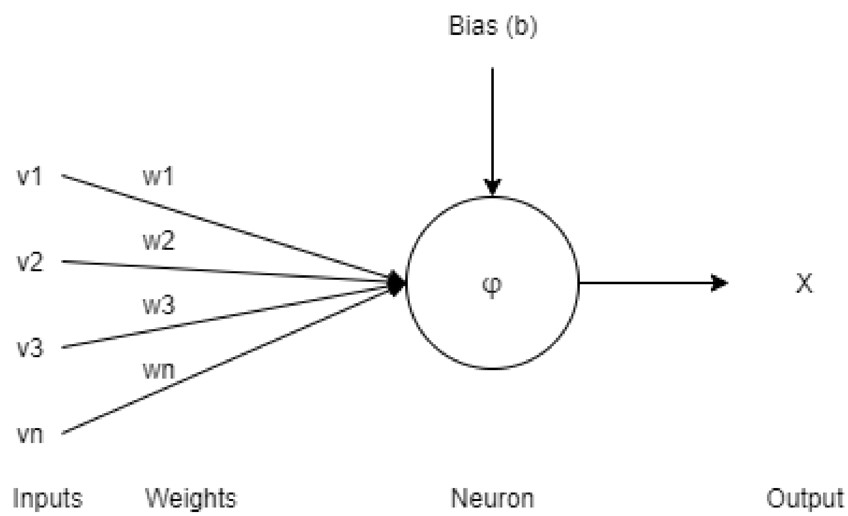


**Figure 30 Flow of information in an artificial neuron (source: author)**

Before reaching the neuron, the input signal of each neuron from the previous layer is multiplied by its dedicated weight. Once the weighted signals are collected, they are added to create the weighted sum ($w_s$ ) as denoted by Equation 13, where $b$ is the bias for the neuron:

$$w_s = w_1 \times v_1 + w_2 \times v_2 + \cdots + w_n \times v_n + b$$

**Equation 13**

The weighted sum equation can be written with matrices as in Equation 14 (Kim, 2017):

$$w_s = wv + b$$

**Equation 14**

Where $w$ and $v$ are defined as:

$$w = [w_1 \ \ w_2 \ \ w_3 \ldots \ldots w_n] \text{ and } v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \cdot \\ \cdot \\ \cdot \\ v_n \end{bmatrix}$$

**Equation 15**

Then finally, the neuron enters the weighted sum into the activation function and yields its output as shown in Equation 16:

$$X = \varphi(w_s)$$

**Equation 16**

One of the most used activation functions is the sigmoid function (Ghritlahre & Prasad, 2018) as given by Equation 17:

$$\varphi(w_s) = \frac{1}{1 + e^{-w_s}}$$

**Equation 17**

Note that the sigmoid function is replaced by the Tansig transfer function whenever negative values are found in the input or output layers (Ghritlahre & Prasad, 2018) as follows:

$$\varphi(w_s) = \frac{1 - e^{-2w_s}}{1 + e^{-2w_s}}$$

**Equation 18**

The above equations represent one neuron but each individual neuron has its own specific set of weights and bias on the inputs. The weights of neurons are initially set to random values. Training data is fed to the bottom (input) layer and it passes through the succeeding layers, getting multiplied and added together as described in the equations, until it finally arrives, drastically transformed, at the output layer. Information is stored in form of weights. Those weights have to be changed to train the ANN with new information. In this work back-propagation algorithm is applied to minimize the MSE between the real observed output and forecasted output from the MLFFNN, the weights are adjusted in proportion to the input value(s) and the output error (MSE) as can be seen in Equation 19 (Talaat et al., 2020).

$$\text{Min}\ (MSE) = \ \min\left(\frac{1}{n} * \sum_{i=1}^{n} \left(y_t - p_t\right)^2\right)$$

**Equation 19**

The change in weights and biased are calculated as in Equation 20 and Equation 21 respectively (Talaat et al., 2020) (Leema et al., 2016):

$$\Delta w_n = \gamma\ (y_n - \ p_n)$$

**Equation 20**

$$\Delta b_n = \gamma\ (y_n - \ p_n)$$

**Equation 21**

Where $\Delta w_n$ is the change of weight for the nth neuron, $\Delta b_n$ is the change of the bias for the nth neuron, and $\gamma$ is the learning rate. Subsequently, the adjusted weights

($w_{adjusted}$) and biases ($b_{adjusted}$) are donated as in Equation 22 and Equation 23 respectively.

$$w_{adjusted} = w + \Delta w$$

**Equation 22**

$$b_{adjusted} = b + \Delta b$$

**Equation 23**

One unit of this process (when training data is passed forward through the neural network and then the weights of each neuron are adjusted based on the error) is called an epoch. During training, the weights are repeatedly adjusted each epoch. This loop will continue until either a specific number of epochs is reached or when the value of MSE reaches the lowest possible limit (typically when MSE does not change for several epochs).

This research utilizes a fully connected MLFFNN as described above with one hidden layer. For each of the three methods used, the ANN was fed with the same set of input variables as were the corresponding MR models. Notice, this implies a differing number of input neurons for each method used. Also, ANN will be used to forecast wind and PV energy, hence the full network details will be provided in the following chapters.

### *IV.6.2. SVM forecasting model*

Since SVM was first introduced by Vladimir Vapnik in 1992, this method became a popular machine learning tool for classification and regression (Vapnik, 2013). SVM regression is a nonparametric supervised learning technique as it relies on kernel functions. Unlike other methods and techniques where regression models try to minimize the errors (squared error normally) between the real and predicted values, Support Vector Regression (SVR) tries to minimize the coefficients (W) as can be seen in Equation 24 (Awad & Khanna, 2015) – or, more precisely, try to minimize the norm of the coefficient vector. Thus, in SVR the error term is represented by constraints where the absolute error is set less than or equal to a specified margin called the maximum error ($\epsilon$), as can be seen in Equation 25 (Awad & Khanna, 2015). In other words, SVR tries to find the best fit line, such that this fit is the hyperplane that has the maximum number of points, as can be seen in Figure 31.

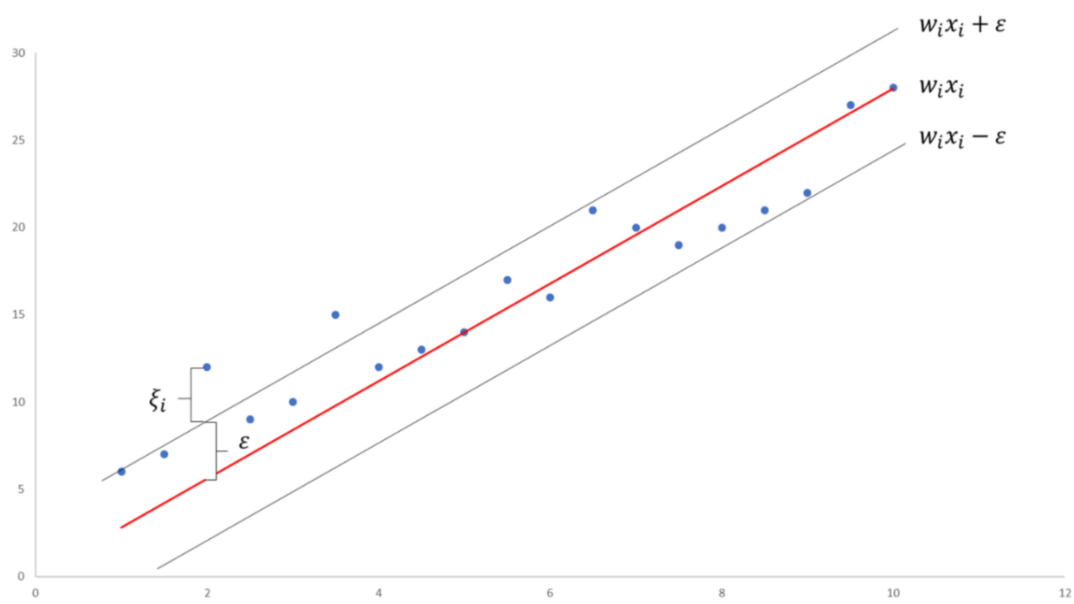


**Figure 31 Simple illustrative example of SVR**[13]

[13] Source: https://towardsdatascience.com/an-introduction-to-support-vector-regression-svr-a3ebc1672c2

$$Minimize: \frac{1}{2} \|w\|$$

**Equation 24**

$$Constraints: |y_i - w_i x_i| \leq \epsilon$$

**Equation 25**

Sometimes, some values might fall outside the specified margin ($\epsilon$), these values are called slack variables and donated by $\xi$. Slack variables have the potential to exist, so it has to be minimized. Thus, Equation 24 and Equation 25 become as in Equation 26 and Equation 27 (Awad & Khanna, 2015). Note that C in Equation 26 is an additional hyperparameter, as C increases, the tolerance for points outside of $\epsilon$ increases. As C move towards 0, the tolerance approaches 0 and the equation collapses into the simplified (infeasible) one.

$$Minimize: \frac{1}{2} \|w\| + C \sum_{i=1}^{n} |\xi_i|$$

**Equation 26**

$$Constraints: |y_i - w_i x_i| \leq \epsilon + |\xi_i|$$

**Equation 27**

In non-linear SVR, kernel functions (K) transform the data into a higher dimensional feature space to make it possible to perform the linear separation as shown in Equation 28, where b is the bias and $a$ and $a^*$ are the Lagrange multipliers (or dual variables) and are nonnegative real numbers (Awad & Khanna, 2015).

$$y = \sum_{i=1}^{n} (a_i - a_i^*) . K\langle x_i, x\rangle + b$$

**Equation 28**

In this work, the Gaussian radial basis kernel function was used as can be seen in Equation 29, where $\|X - X'\|^2$ is the squared Euclidean distance between the two feature vectors $X$ and $X'$. $\sigma$ is a free hyperparameter.

$$K(X, X') = \exp\left(-\frac{\|X - X'\|^2}{2\sigma^2}\right)$$

**Equation 29**

### *IV.6.3. KNN forecasting model*

K-nearest Neighbours Regression or KNN is also a non-parametric method for prediction. KNN uses feature similarity to predict the values of any new data points. This means that the new point is assigned based on how closely it resembles the points in the training set. The main steps of forecasting wind power using KNN technique can be boiled down to the following three steps:

1. Calculating the distance between the new point and each training point.
2. Based on the closest distance (which is calculated in step 1), the closest k data points are selected.
3. The average of the k data points is the prediction for the new point.

Many methods can be used to calculate the distance between new and training points, yet, Euclidean distance, Manhattan distance, and Mahalanobis distance are the most commonly used (Zhang & Wang, 2016). In this work, the Euclidean distance was utilized as can be seen in Equation 30.

$$\text{Euclidean distance} = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$$

**Equation 30**

### *IV.6.4. Hybrid forecasting model*

The fourth model used in this work is a hybrid model. In the suggested hybrid model, ANN, SVM, and KNN techniques are used. The hybrid model has three main tasks, first task is to detect the most accurate technique among ANN, SVR, and KNN by analyzing its performance in the past 24 hours. This task is done by comparing MS values for the past 24 hours. The model with the least MS value is considered the most accurate one. The model with the most accurate prediction will be selected to do the forecast for the upcoming 24 hours. Selecting the model to make the next 24 hours forecast is the second task. Finally, in the third task, the prediction data from the selected model is further modified, and an error correction process is conducted. Figure 32 shows the hybrid model workflow.

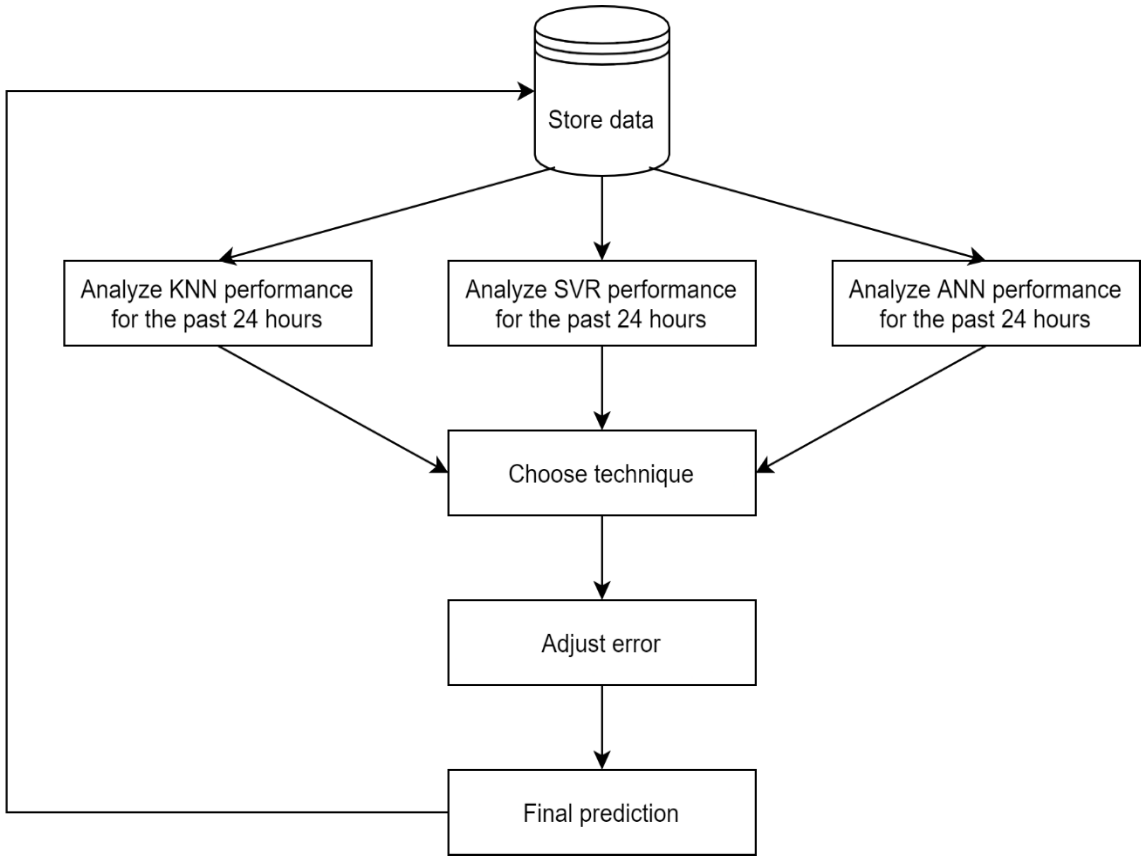


**Figure 32 Hybrid model workflow (source: author)**

As the model is chosen to do a future forecast based on its past performance, some errors are expected. However, these potential errors can be reduced. Thus, for the selected model (the model which is selected to do the next 24 hours forecast) the average past errors for each time step are calculated as can be seen in Equation 31. The term $(t - 24 * n)$ represents the time steps 24, 48, 72,……n hours before the prediction starts.

$$\text{Average error for each time step } (\mathrm{ae}_t) = \frac{1}{n} * \sum_{i=1}^{n} (\mathrm{y}_{(t-24*n)} - \mathrm{x}_{(t-24*n)})$$

**Equation 31**

Then the final prediction will be the sum of the original prediction done by the selected model and the ae as can be seen in Equation 32.

$$\text{Final prediction}_{t} = \text{Original prediction}_{t} + \text{ae}_{t}$$

**Equation 32**

## IV.7. The use of forecasting models for PV energy forecasting

Key challenges facing solar energy forecasting models include the task of choosing the right method and the need to select appropriate inputs to achieve the most accurate prediction. Consequently, one aim of this thesis is to investigate two of the main techniques for building prediction models to accurately forecast PV output production: multiple regression (MR) and artificial neural network (ANN). To that end, structural, time-series, and hybrid data input methods mentioned in the previous section will be used to build different forecasting models and experiment with different input (predictor) settings.

Figure 33 depicts a general overview of the steps forming the development process. Building the forecasting models starts by feeding the historical weather and PV power data to the models. Structural models are fed with historical weather data, time-series models are fed with historical PV power data, while hybrid models are fed with both weather and PV power historical data. Each model is forecasting the PV output power for the selected horizon with a given resolution set. The forecasted PV output power values are then stored. When the real values become available as a fact data from the PV farm, this data is used to calculate the performance of each model (i.e. in comparison to the stored prediction) as well as to update the historical data records (which means this real-time data is later applied to update the model).

As part of the research, a large amount of historical data was collected to build the prediction models. The data collected covered the period April 13, 2017 to April 18, 2020 (3 years). The data used to train the models for prediction is measured data. Past data is used as an input to forecast the next day, albeit differently depending on the input method. For example, the measurement of April 13 had been used to predict expected output for April 14 in case of the structural model, and for April 15th in cases of the time-series and hybrid models. In other words, for the structural model weather data from exactly 24 hours earlier is used to forecast for a given point in time – e.g. any timeslot of April 14 may be predicted using data from April 13. However, for the time series data prediction (and, therefore, for the hybrid model as well) a full past day data of generated power is needed as input for the model to forecast the next day – e.g. a prediction done on April 14 to predict the same timeslot on April 15 (one-day-ahead) uses data covering a full 24 hours going back (thus including data from April 13), consequently, no prediction is possible for April 14 if data is not available from April 12. This was repeated until April 18, 2020. Thus, all models were continuously trained and tested over the data covering a 3-year period. For all ANN models trained here, the data was split into three segments: 70% training 15% validation, and 15% test set. Since this implies tens of thousands of values of each variable, it would be hard to visualize the forecasted versus the real power values for the whole period. Therefore, the last day of testing (18th of April 2020) was used to visualize and compare the performance of the different models. This day appeared to be a good test day as it had a few dips during the day due to weather changes during the day (as opposed to an average stochastic PV power curve).

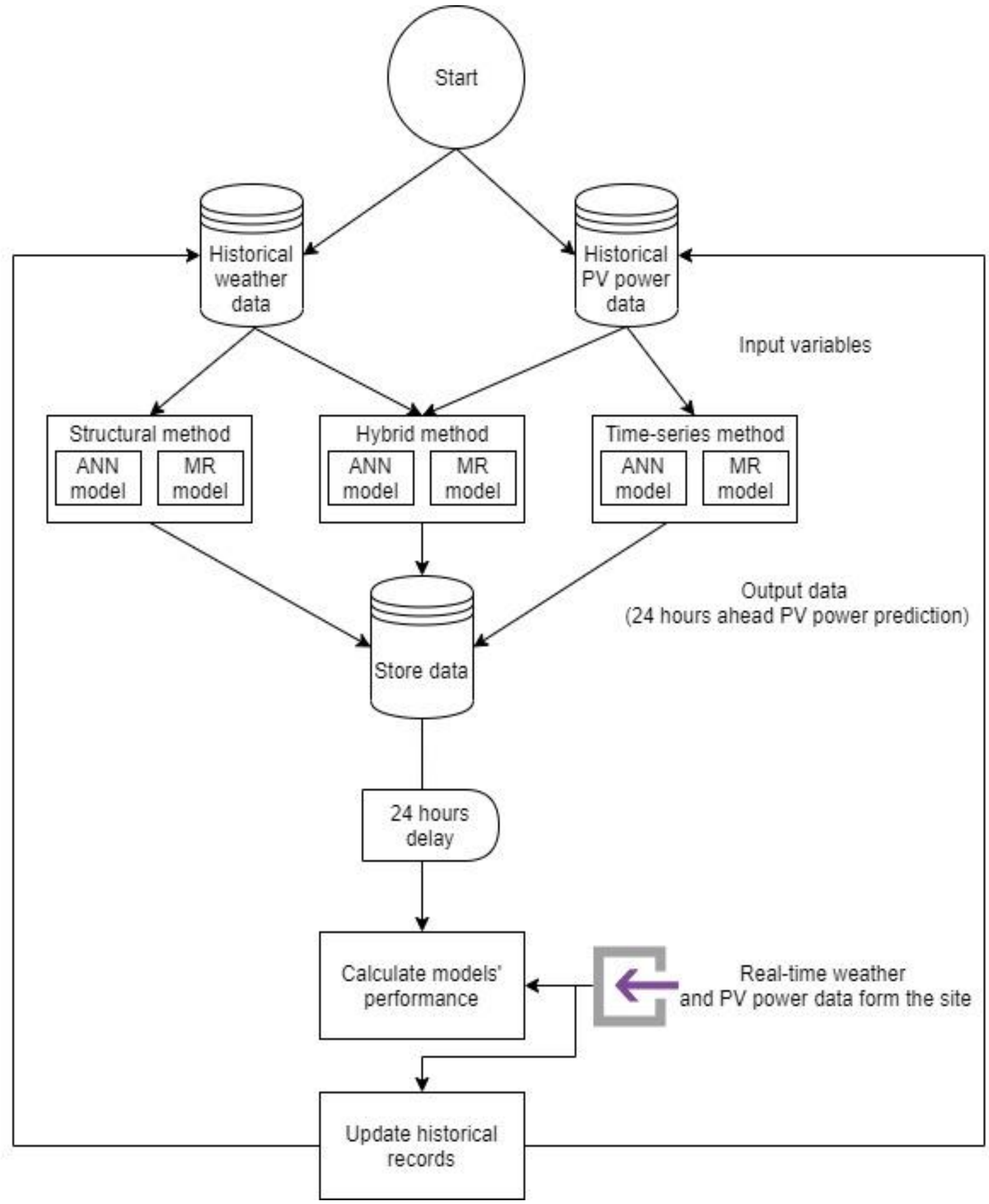


**Figure 33 General overview of the PV forecasting flowchart (Source: Authors)**

## IV.8. The use of forecasting models for wind energy forecasting

Another aim of this thesis is to build ML forecasting models, then utilize different ML forecasting techniques to develop a hybrid method for wind power forecasting. The suggested system is able to utilize Artificial Neural Networks (ANN), Support Vector Machines (SVM), and k-nearest neighbors (KNN) techniques and choose the most accurate prediction based on the past performance of all models. Moreover, a comparative analysis is provided, comparing the performance of each ANN, SVM, KNN model with the suggested hybrid one. Each model is forecasting the wind output power for 24 hours horizon with 15 minutes resolution. Figure 34 below shows the general overview of the methodology. To achieve the aims of this study, a large amount of data is required to build and test the proposed prediction models. Thus, wind power data was collected from a 2 MW wind turbine for the period May 1, 2019, till June 13, 2020. The data was collected in 15 minutes resolutions.

Even though ML wind power forecasting models show robust abilities and good performance, yet, the superiority of the AI and machine learning forecasting models have some limitations (Makridakis et al., 2018) as described in section II.7.

Hence, to avoid AI and machine learning forecasting models limitations, 13 months of time-series data was used to forecast day ahead (96 steps ahead) wind power. Also, three machine learning wind power forecasting models were built, utilizing different ML techniques. Thus, ANN, SVM, and KNN wind power forecasting models were built. Then, the three above-mentioned ML techniques were used to build a hybrid model. Finally, the performance of this hybrid model was benchmarked, analyzed, and compared with ANN, SVM, and KNN models for short and long-term horizons.

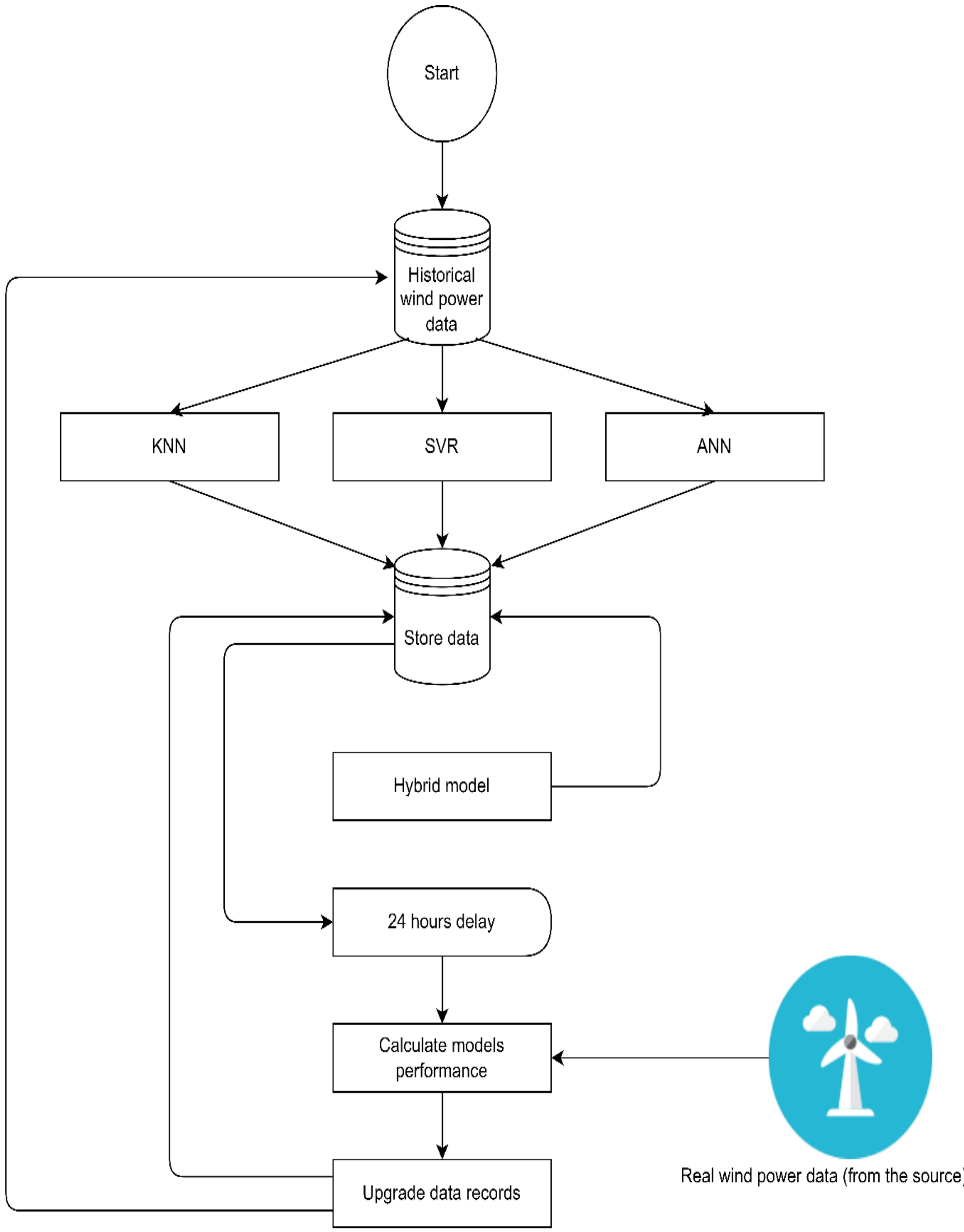


**Figure 34 General overview of the wind forecasting flowchart (Source: Authors)**

### IV.9. Resolution and horizon modeling

To test the effect of utilizing input data with different resolutions, two sets of ANN time series forecasting models were designed, built (i.e. trained), and tested to forecast wind and PV out power for 24 hours ahead. Each set utilizes input data with resolutions of 15, 30, and 60 minutes. Once the six models were trained, their accuracy was then calculated. Subsequently, a comparative analysis was conducted to determine the best settings leading to the best performance.

The input to the time-series ANN models are past energy values, therefore, to train the ANN models, both actual PV and wind past energy values were collected covering a bit more than 13 months. Data collection started on May 1, 2019, and lasted till June 13, 2020. PV and wind past generation time-series data were collected from the sources mentioned earlier in Table 12. All data were collected in 15, 30, and 60 minutes resolutions.

Figure 35 shows the overview of the methodology. The process starts by collecting the past generation data for the PV farm and wind turbine. The data is used in its original resolution as collected, thus, data was not averaged to build up lower resolutions. Then six ANN forecasting models were designed and trained: ANN models were built to forecast PV and wind energy both with 15, 30, and 60-minute resolutions.

The target horizon of the forecast is 24 hours ahead. The outputs (forecasted values of PV and wind energy) were then stored. After 24 hours delay, when the real generation values have become available (as the real production values are always lagging 24 hours behind the forecasted ones), the performance of each model was calculated. The output data is used to update the historical records and then to continue the training of the models.

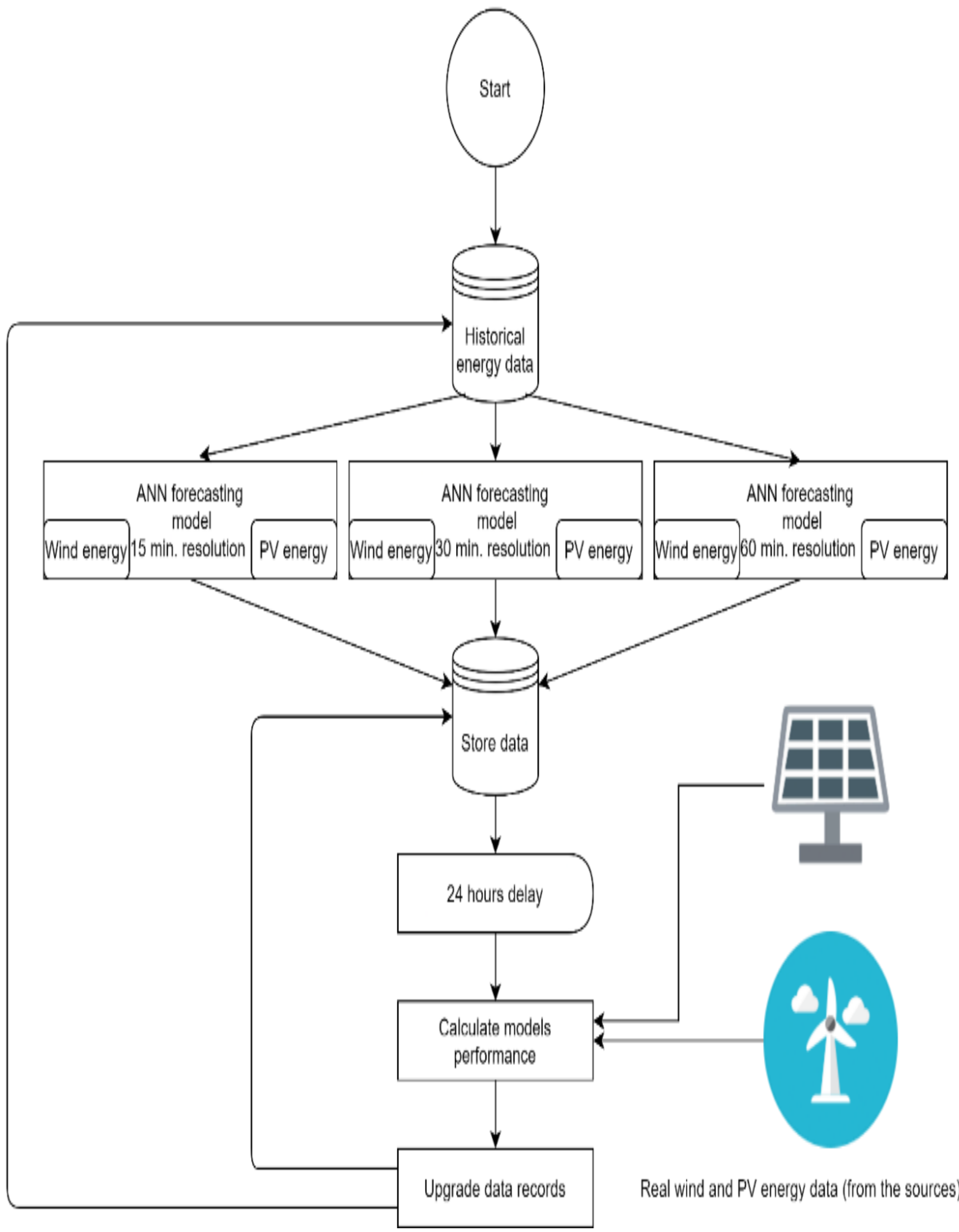


**Figure 35 Flowchart providing a general overview of the resolution testing (source: author)**

# V. EVALUATING NEURAL NETWORK AND LINEAR REGRESSION PHOTOVOLTAIC POWER FORECASTING MODELS BASED ON DIFFERENT INPUT METHODS

The studied PV energy forecasting system is built and tested for the Szeged site (Table 10  on page 86). The other two sites will be tested in future research when more data is collected as they are relatively new sites and more data is needed.

## V.1. Data preparation and variables selection

PV power output is highly correlated with weather variables (F. Wang et al., 2019). Yet, not all weather variables have the same significance for PV power forecasting. A correlation analysis was done to determine the significance of each collected variable shown in Table 12 before it was used in the modeling: see Table 19 for results.

As can be seen in Table 19, some meteorological factors have higher significance than others. Solar irradiance components have the highest significant factors, especially GTI fixed-tilt. Generally, it can be concluded that all variables collected here can be used in the modeling, yet variables that have low correlations with the output power could be excluded: in this study the threshold is set to 0.1, therefore wind speed, snow depth, and precipitable water are excluded.

**Table 19 The correlation between PV output power and meteorological variables (source: author)**

| Input variables | Correlation with PV output power |
|---|---|
| Air Temperature | 0.42 |
| Cloud Opacity | -0.26 |
| Dewpoint Temperature | 0.18 |
| DHI | 0.68 |
| DNI | 0.84 |
| EBH | 0.87 |
| GHI | 0.95 |
| GTI Fixed Tilt | 0.96 |
| GTI Tracking | 0.90 |
| Precipitable Water | 0.09 |
| Relative Humidity | -0.53 |
| Snow Depth | -0.09 |
| Wind Speed | 0.09 |
| Cell Temperature | 0.52 |

## V.2. PV power generation prediction model

After developing the suggested models as mentioned in the methodology (see more on chapter IV), a series of experiments were constructed to measure the performance of each model variant and to compare their prediction abilities using the evaluation measures described in section IV.4. The overall (average) performance of each variant was computed at the end of training and testing (i.e. over 3 years). Then, as an additional demonstration that enables some representative visualization, the performance for forecasting the output power for the 18th of April 2020 was also computed and compared to the overall performance.

### *V.2.1. Multiple regression models*

Initially, the multiple regression model was developed utilizing only meteorological and geographical variables, then another MR model was built to utilize only time-series data of PV solar power, and finally, a third model was developed utilizing both PV power historical data as well as geographical and meteorological parameters as inputs.

#### *V.2.1.1. Structural Multiple Regression model (SMR) performance*

Over the training and testing period, this model shows a good performance with a 0.94 COD, 14.84 MAE, 1054.74 MSE, and, 32.47 RMSE. Figure 36 shows the frequency distribution of the error. It can be seen that the most frequent (which were recorded more than 1000 times) errors recorded are small errors ranging between -20 and +20 kW. However, some fairly large errors can also be observed.

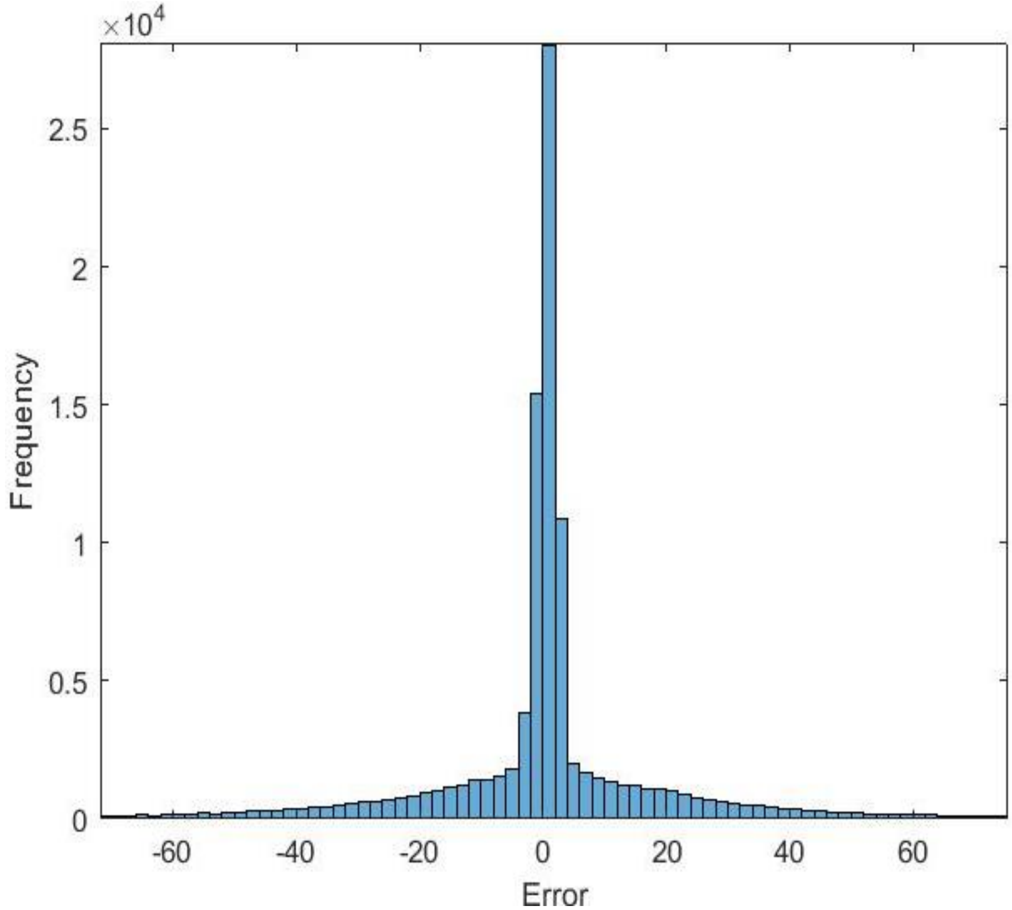


**Figure 36 Frequency distribution of the error in SMR (source: author)**

The SMR model was able to forecast the energy for the 18th of April with a 0.92 COD, 22.44 MAE, 1815.52 MSE, and, 42.60 RMSE performance measures. Which is a bit less than the overall performance. Figure 37 shows the forecasted vs. observed power for the mentioned day.

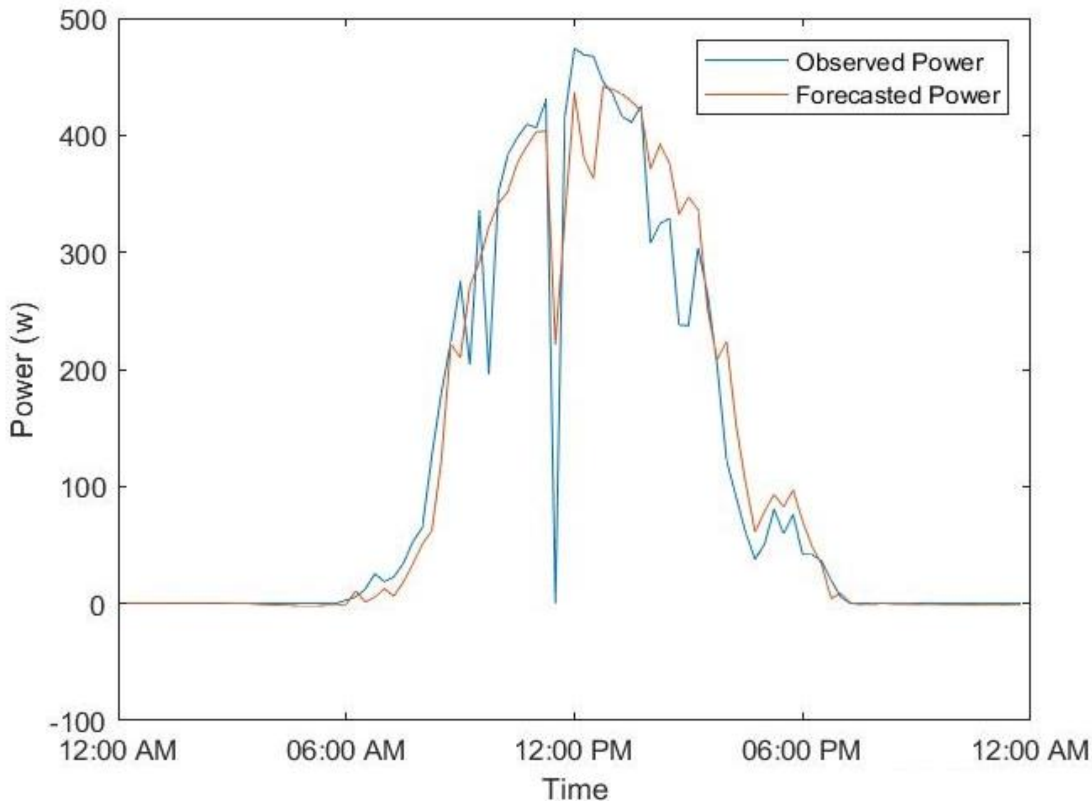


**Figure 37 Forecasted vs. observed power for the SMR (source: author)**

Note that the model utilizes real values, that is they have 100% accuracy. This explains the very good prediction performance. In case meteorological parameters can only be provided with some degree of uncertainty, the SMR might have less accurate performance. Figure 38 shows a sensitivity analysis for the SMR model where the effect of uncertainty in the input variables can be observed on the forecasted power.

The sensitivity analysis shows the normalized percent changes in the forecasted PV output power with the normalized percent of input variables uncertainty. The point (1,1) on the graph represent 100% accurate inputs, therefore, there is no change in forecasted power. An uncertainty between 0 and +/ – 40% in any of the input variables (i.e. 0.6 and 1.4 in the x-axis) leads to huge changes in the model's output, thus affecting the performance.

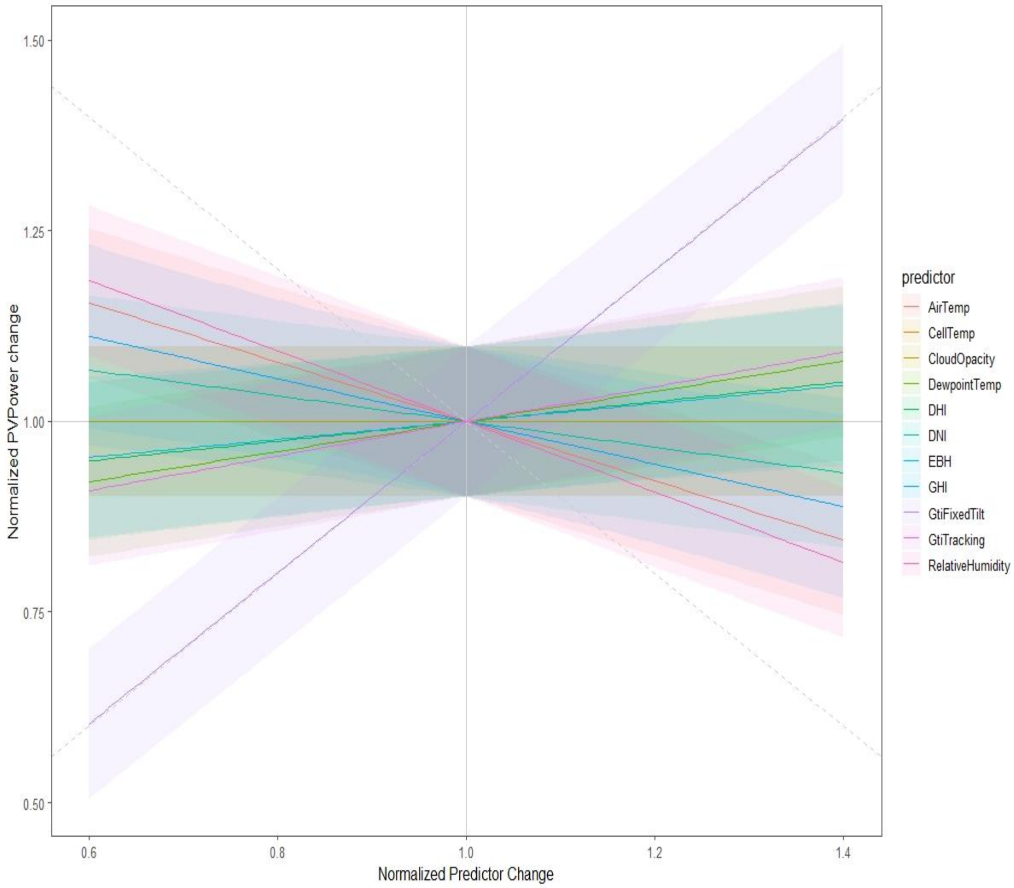


**Figure 38 Sensitivity analysis for the SMR model (source: author)**

### *V.2.1.2. Time-series Multiple regression model (TMR) performance*

Over the training and testing period, this model shows poor performance, much worse than the SMR with a 0.68 COD, 45.83 MAE, 5584.5 MSE, and, 74.72 RMSE. Figure 39 shows the frequency distribution of the error. Some huge errors were recorded, additionally, the errors are not distributed around zero (most frequent errors do not equal zero).

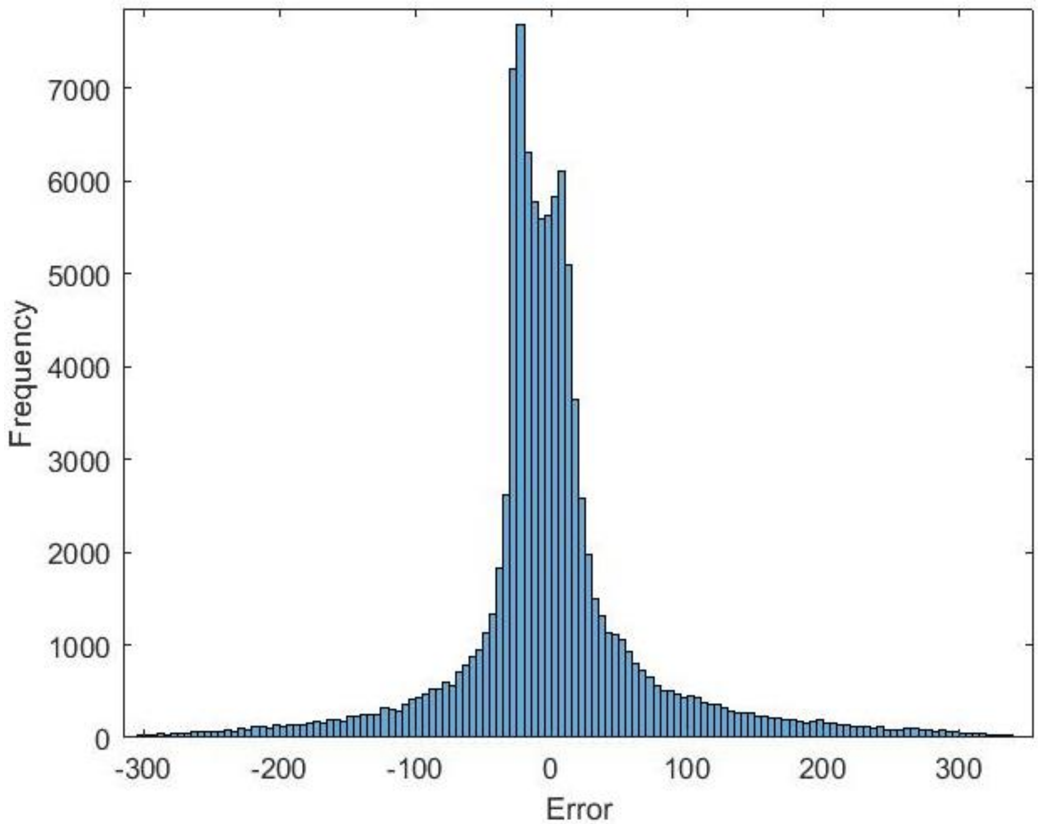


**Figure 39 Frequency distribution of the error in TMR (source: author)**

Figure 40 shows the forecasted vs. observed power for the 18th of April 2020. It can be noticed from error measures and Figure 40 that the TMR performs worse than the SMR. The TMR could not predict the sudden drop in the output PV power just before noon, while this drop was better predicted by the SMR.

The SMR model was able to forecast the energy for the 18th of April with a 0.88 COD, 32.97 MAE, 3091.27 MSE, and, 55.59 RMSE performance measures. Which is above the overall performance of this model. This can be explained as the 18th of

April does not have huge weather variations, thus the TMR performs better than other days.

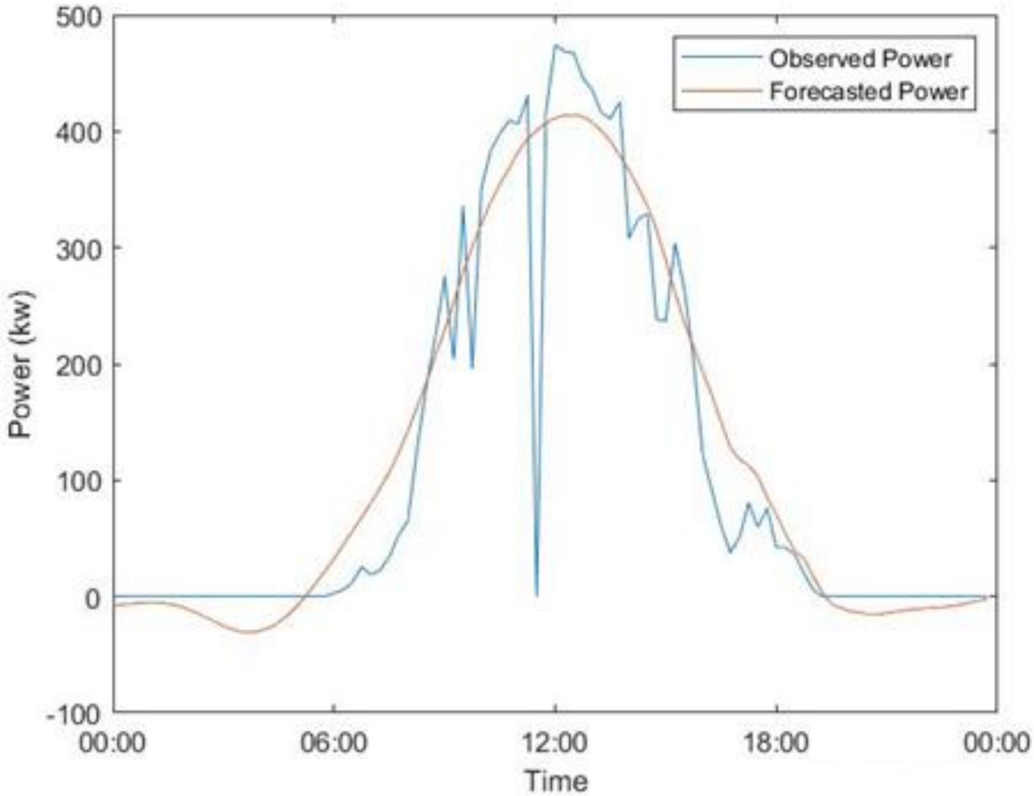


**Figure 40 Forecasted vs. observed power for the TMR (source: author)**

#### *V.2.1.3. Hybrid Multiple regression model (HMR) performance*

Over the training and testing period, this HMR model shows the best overall performance compared to TMR and SMR, with a 0.95 COD, 16.05 MAE, 835.68 MSE, and, 28.90 RMSE. Figure 41 shows the frequency distribution of the error. Even though the HMR shows better performance, it shows more errors between -50 and 50 KW with over 1000 frequency.

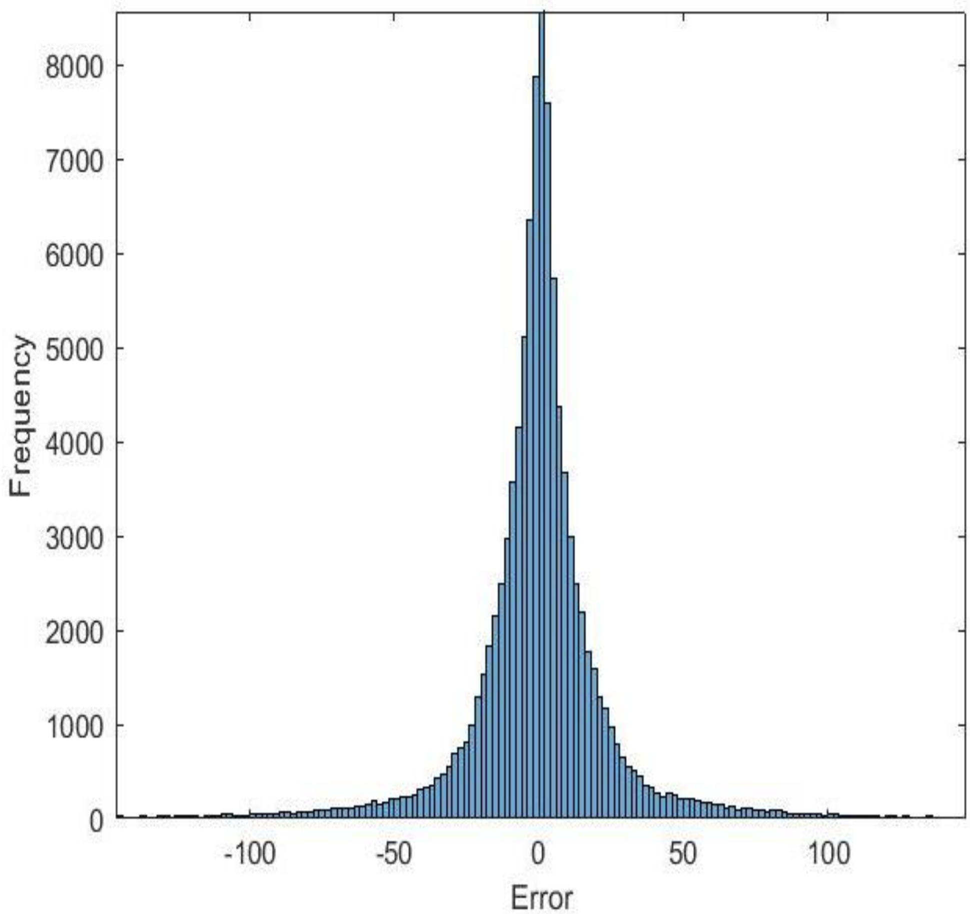


**Figure 41 Frequency distribution of the error in HMR (source: author)**

The HMR model was able to forecast the energy for the 18th of April with 0.93 COD, 22.26 MAE, 1796.84 MSE, and, 42.38 RMSE performance measures. Which is a bit less than the overall performance. Figure 42 shows the forecasted vs. observed power for that day.

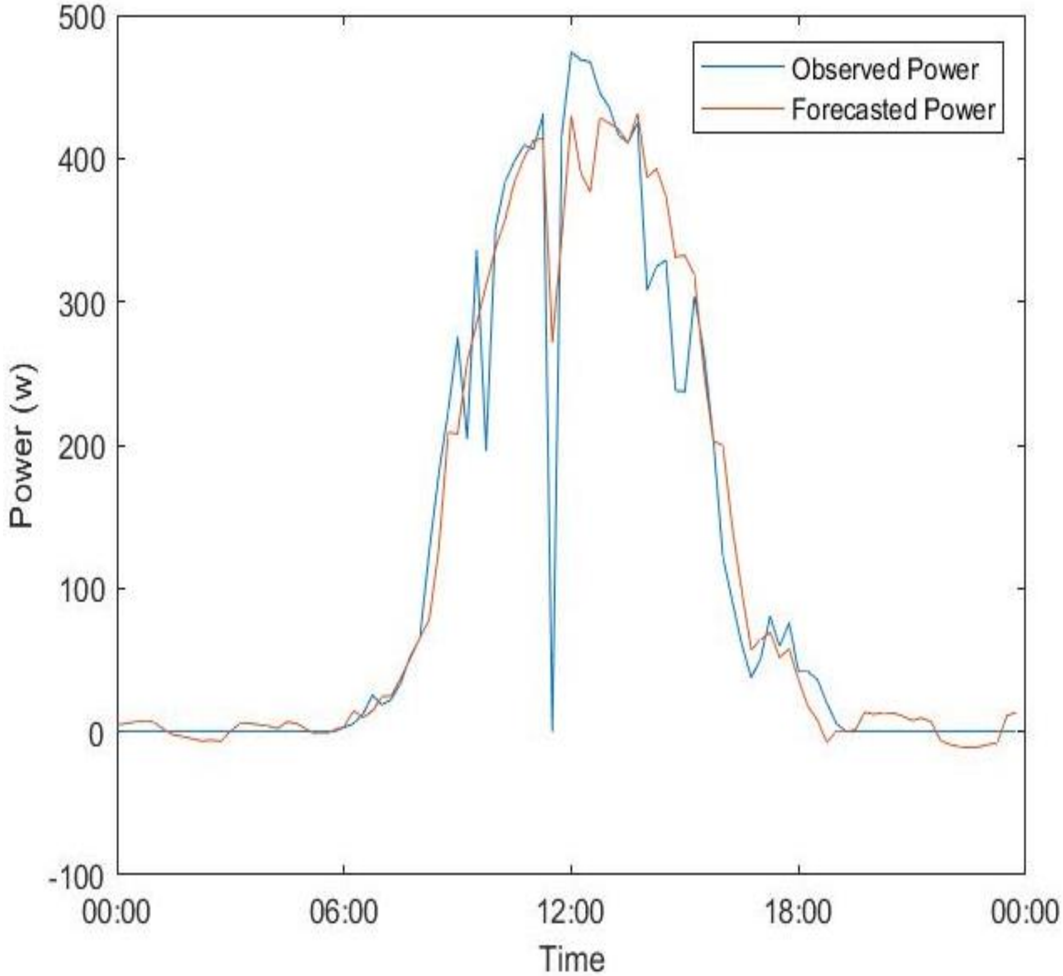


**Figure 42 Forecasted vs. observed power for the HMR (source: author)**

### *V.2.2. Artificial Neural Network models*

The ANN models were built using three input data methods, the same way as were the MR models as described in section IV.6.1. Initially, the ANN model was developed utilizing only selected meteorological and geographical variables, then another ANN model was developed to utilize only time-series data of PV power, and finally, a third ANN model was developed to utilize PV power historical data as well as geographical and meteorological parameters as inputs.

As explained in the methodology, this research utilizes a fully connected MLFFNN with one hidden layer. For each of the three methods used, the ANN was fed with the same set of input variables as were the corresponding MR models. Notice, this implies a differing number of input neurons for each method used.

The number of input neurons, therefore, are 11, 96, and 107 for the structural, time series, and hybrid input methods respectively. The number of hidden neurons is an important parameter for ANN. With few hidden neurons the ANN might not be able to generate a function that indicates the underlying problem while having more hidden neurons than required may result in over-fitting of the training set and reducing the ability to generalization the out-of-sample data (Setyawati, 2005). Therefore the number of hidden neurons was set to be 33% (one-third) of the number of inputs. Table 20 shows the settings of the ANN parameters. Although training time had not been limited, the actual running time for the set number of epochs to be trained was ranging from seconds to a couple of hours, while forecasting times were, of course, very short (fraction of a second).

**Table 20 ANN parameters for PV energy forecasting (source: author)**

| Parameter | Description | Value for each method | | |
|---|---|---|---|---|
| | | Structural | Time series | Hybrid |
| Number of inputs | Number of input data variables | 11 | 96 | 107 |
| Number of outputs | Number of output forecasted variables | 1 | 1 | 1 |
| Number of hidden neurons | Number of hidden neurons | 4 | 32 | 35 |
| Maximum Epochs | Maximum number of training iterations before training is stopped | 1000 | 1000 | 1000 |
| Maximum Training Time | Maximum time in seconds before training is stopped | ∞ | ∞ | ∞ |
| Performance Goal | The minimum target value of MSE | 0 | 0 | 0 |

#### *V.2.2.1. Structural Artificial Neural Network model (SANN) performance*

Over the training and testing period, this model shows a good performance with a 0.95 COD, 13.13 MAE, 943.53 MSE, and, 30.26 RMSE. The SANN reached the best performance (least MSE) after 140 epochs (iterations) as shown in Figure 43. The error distribution is in Figure 44 which shows the total frequency of errors as well as the error frequency in the training, validation, and test sets.

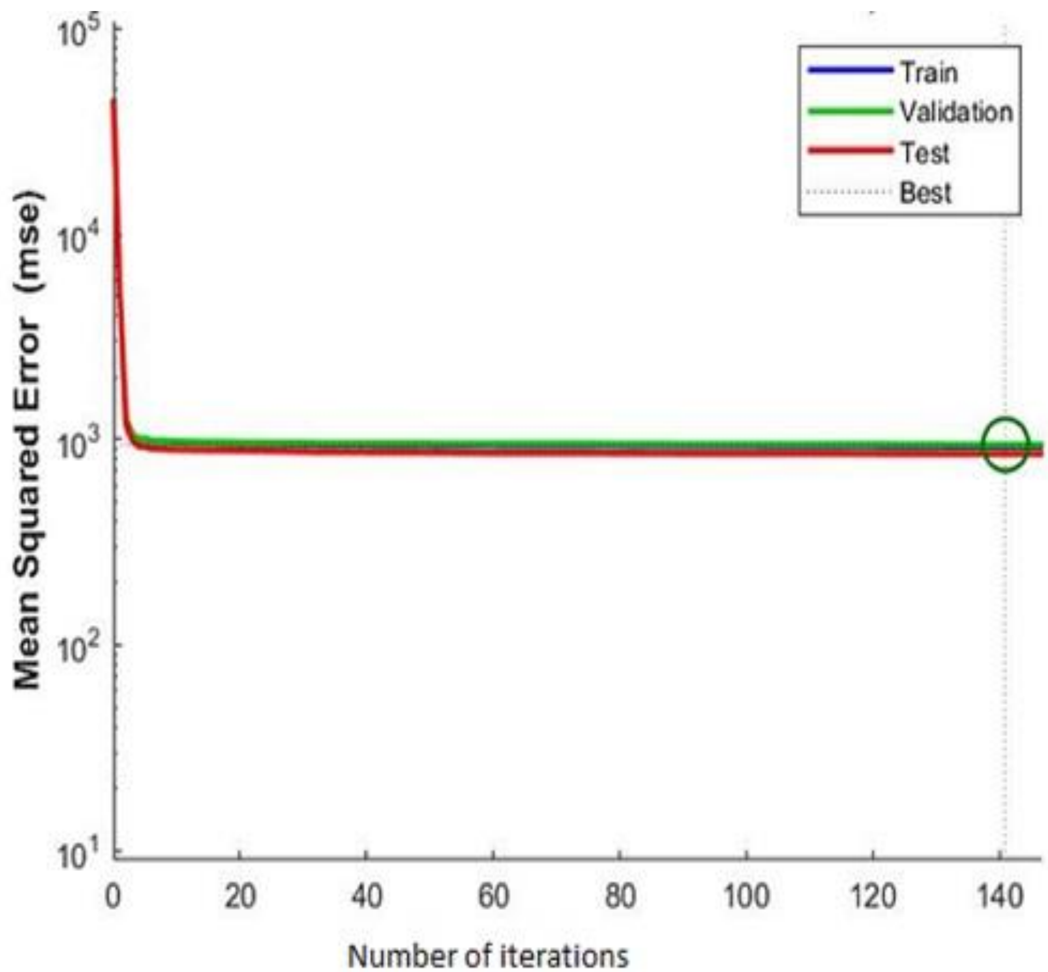


**Figure 43 SANN performance (source: author)**

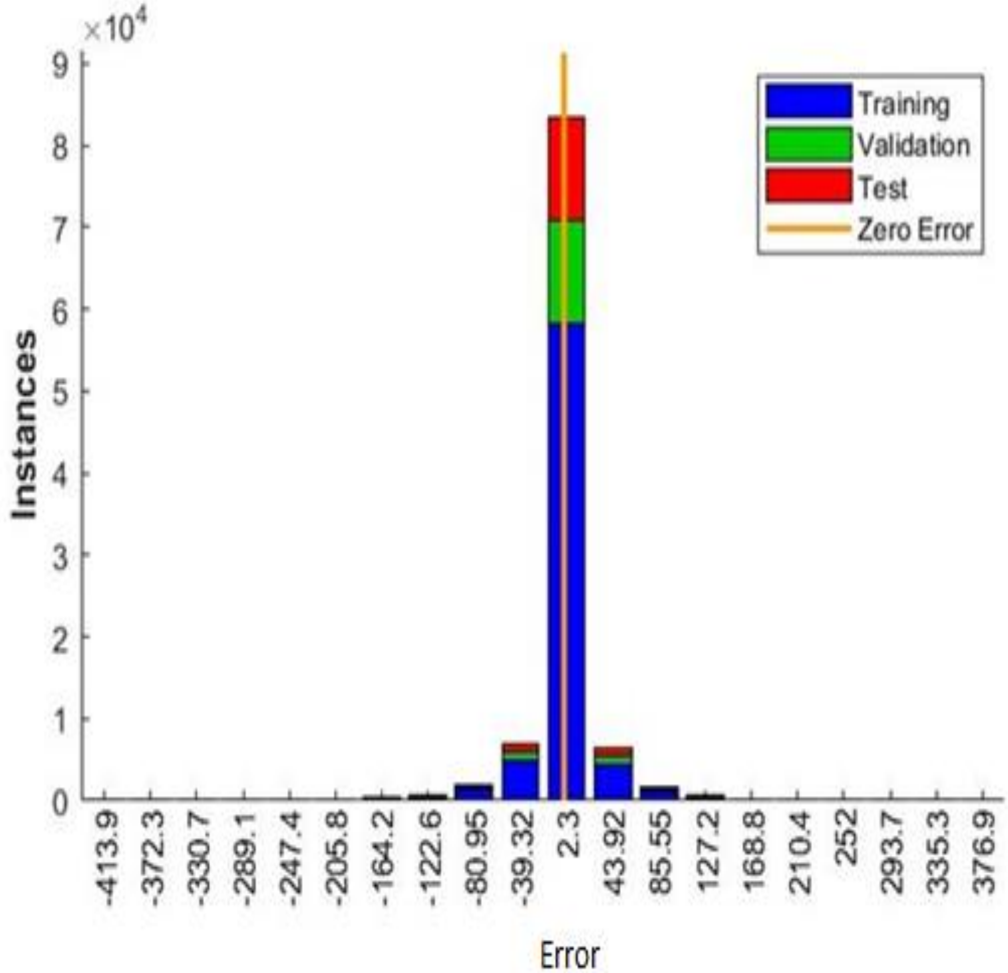


**Figure 44 Frequency distribution of the error in SANN (source: author)**

The TMR model was able to forecast the energy for the 18th of April with 0.93 COD, 20.96 MAE, 1752.54 MSE, and, 41.86 RMSE performance measures. Which is a bit less than the overall performance. Figure 45 shows the forecasted vs. observed power for that day.

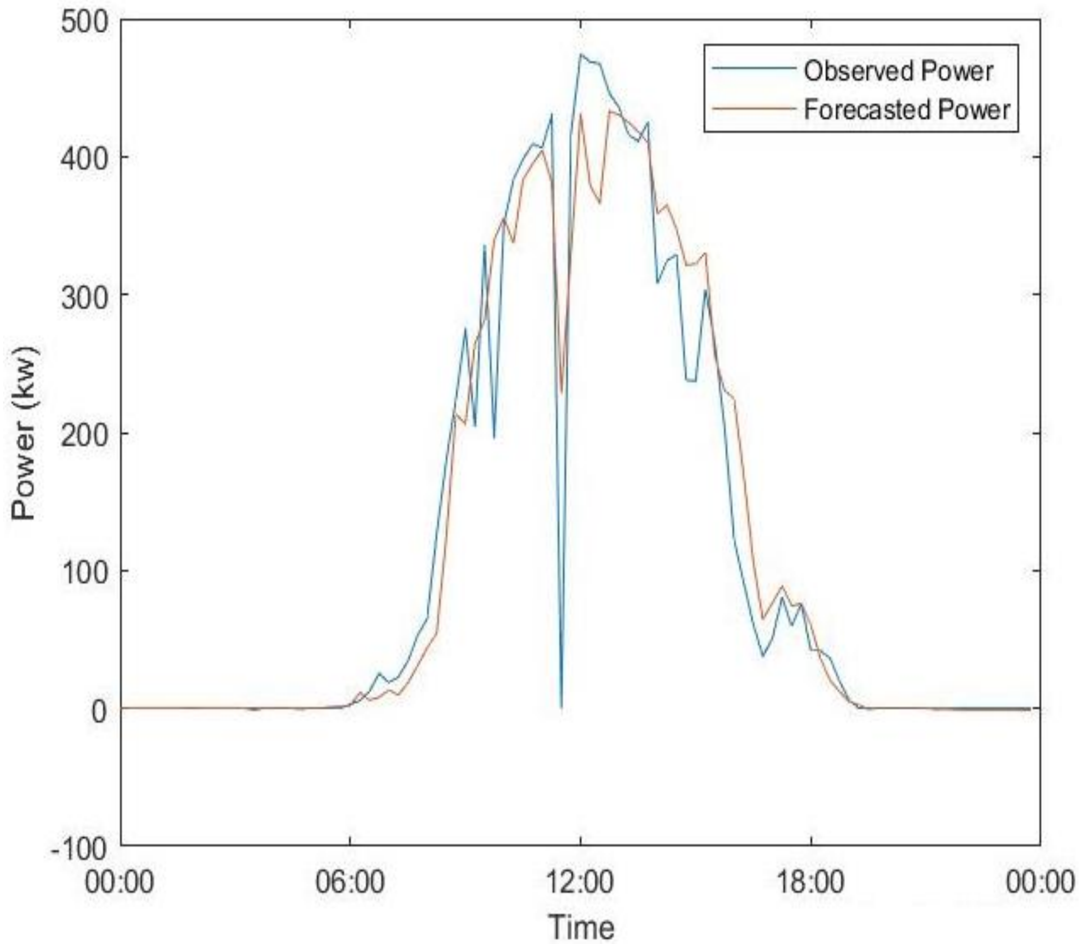


**Figure 45 Forecasted vs. observed power for the SANN (source: author)**

### *V.2.2.2. Time-series Artificial Neural Network model (TANN) performance*

Over the training and testing period, this model shows a fair performance, slightly better than the TMR with a 0.75 COD, 36.38 MAE, 4329.87 MSE, and, 36.38 RMSE. The TANN reached the best performance (least MSE) after 241 epochs (Figure 46). The error distribution can be found in Figure 47.

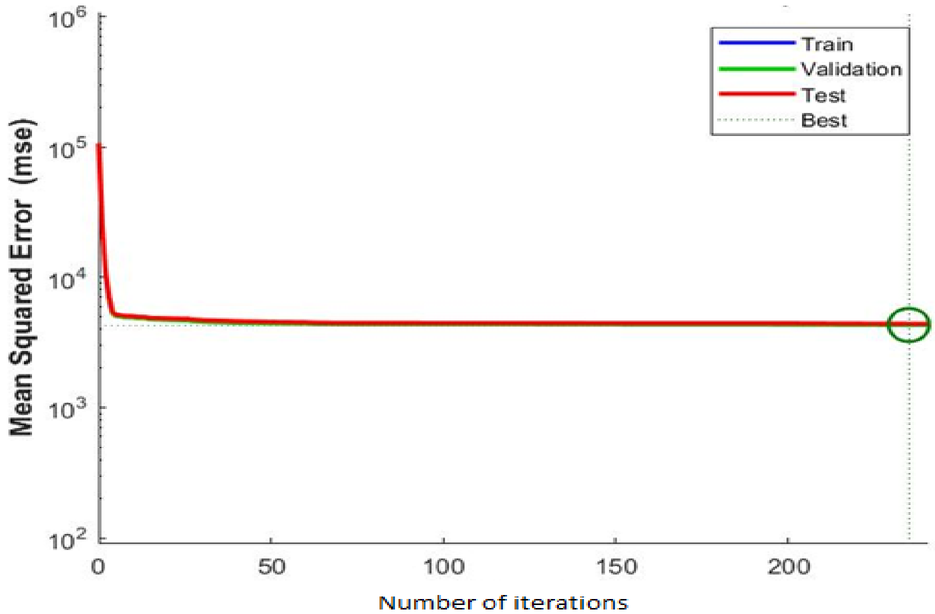


**Figure 46 TANN performance (source: author)**

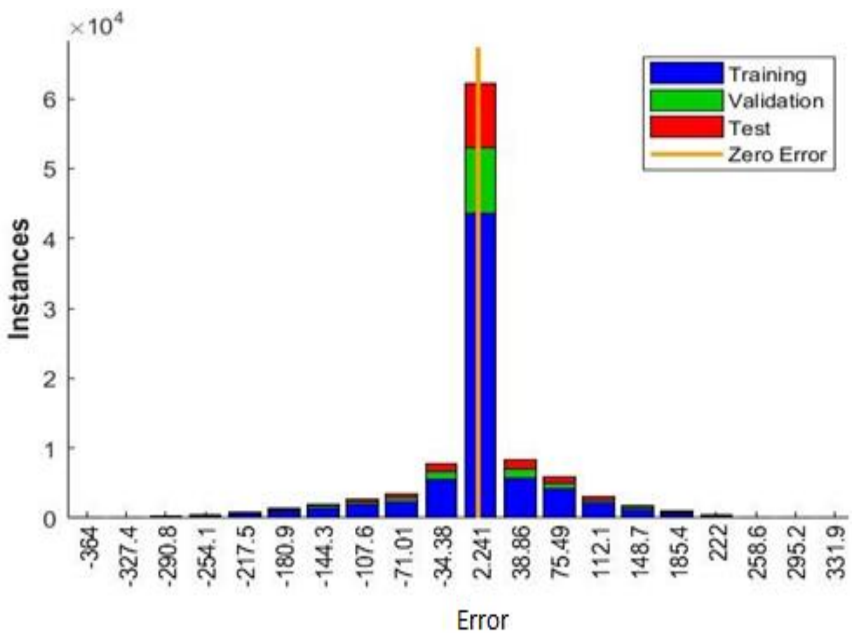


**Figure 47 Frequency distribution of the error in TANN (source: author)**

The TANN model was able to forecast the energy for the 18th of April with 0.87 COD, 32.57 MAE, 3135.32 MSE, and, 55.99 RMSE performance measures. This is better than the overall performance but no dip is predicted, for the same reason mentioned in V.2.1.2. Figure 48 shows the forecasted vs. observed power for the mentioned day.

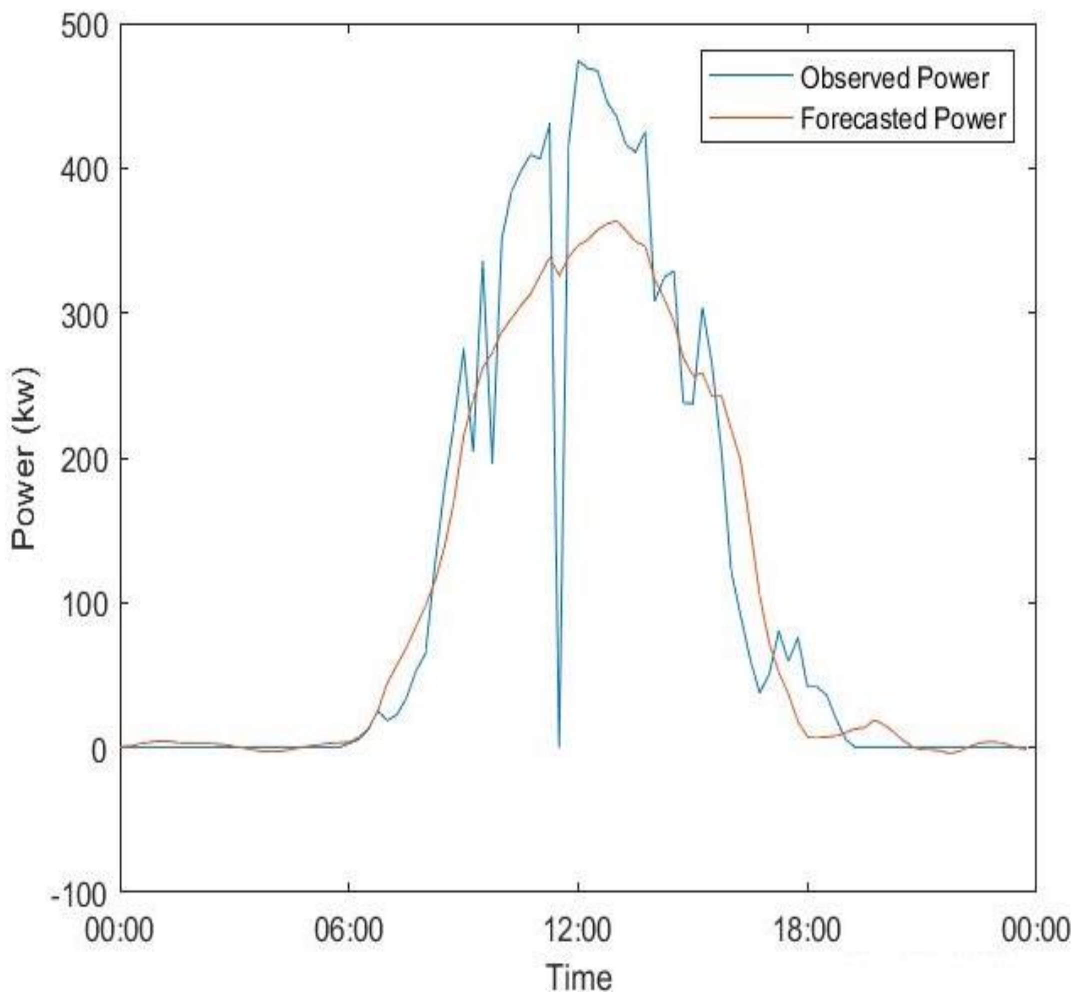


**Figure 48 Forecasted vs. observed power for the TANN (source: author)**

#### *V.2.2.3. Hybrid Artificial Neural Network model (HANN) performance*

Over the training and testing period, the model shows a good performance, way better than the TANN with a 0.96 COD, 13.52 MAE, 914.10 MSE, and, 30.23 RMSE. The HANN reached the best performance (least MSE) after 29 epochs as shown in Figure 49. The error distribution can be found in Figure 50.

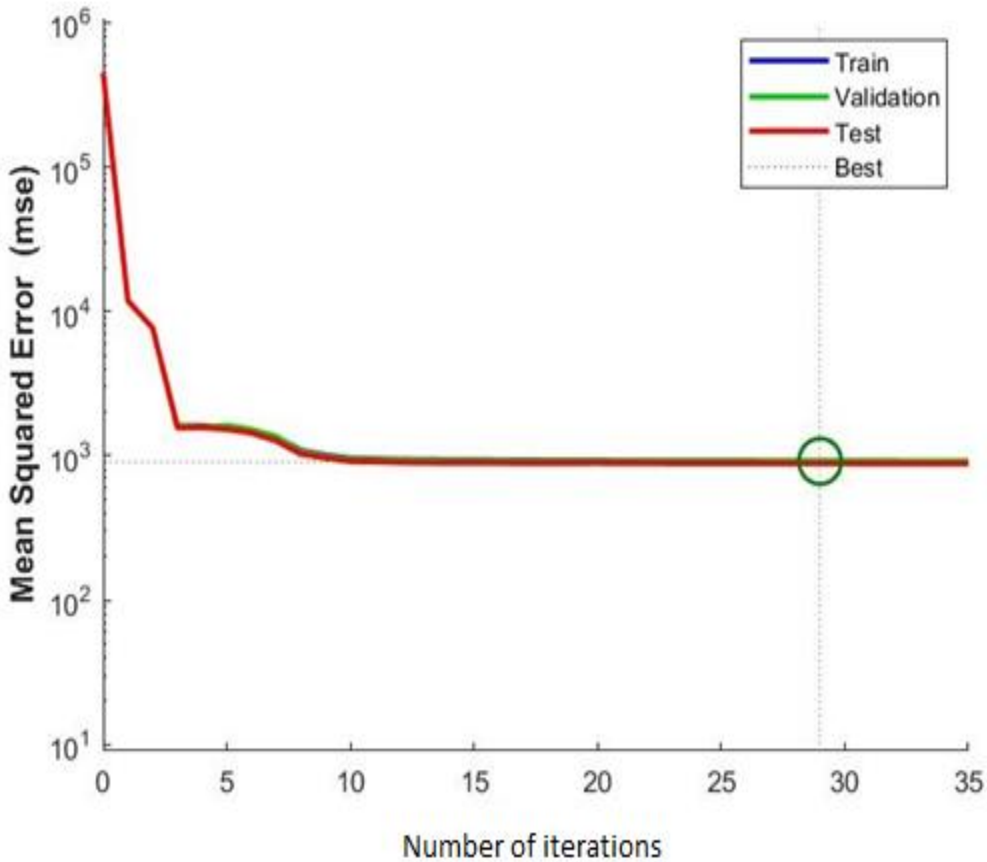


**Figure 49 HANN performance (source: author)**

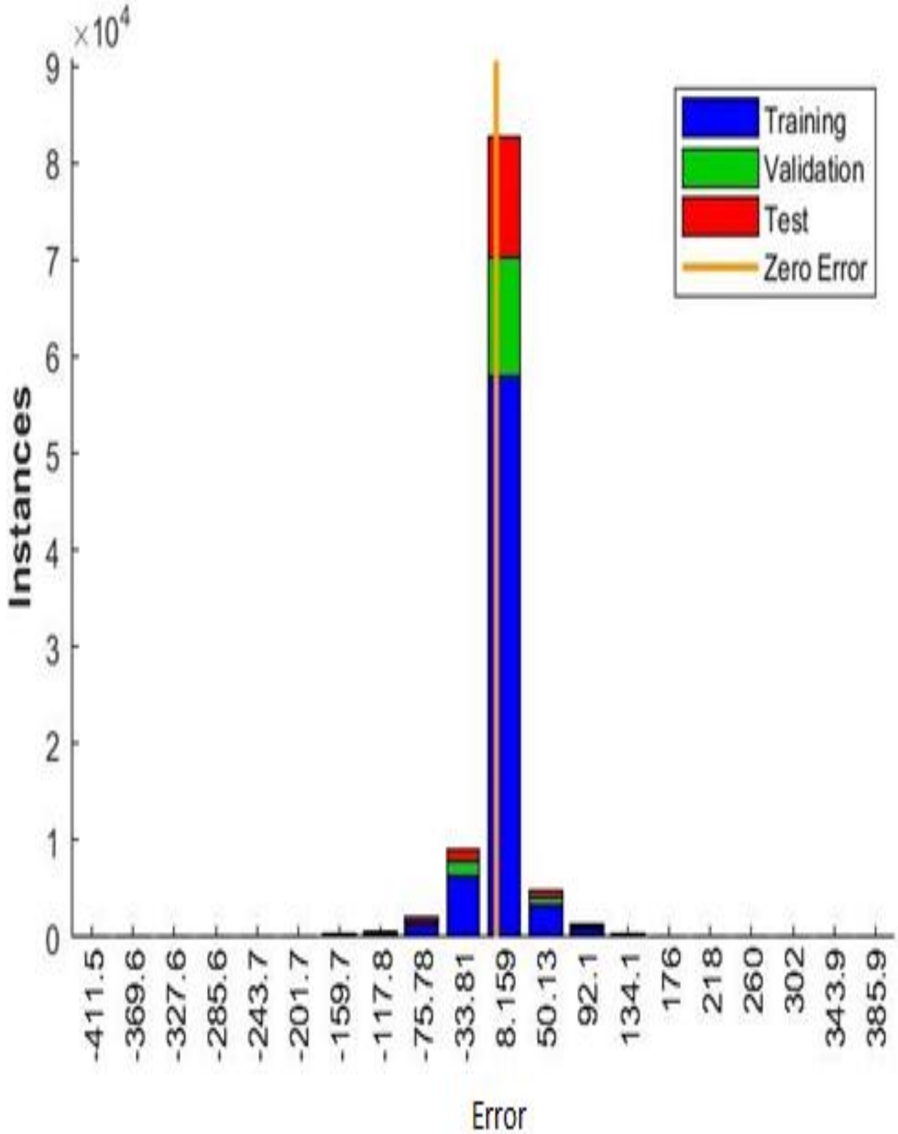


**Figure 50 Frequency distribution of the error in HANN (source: author)**

The HANN model was able to forecast the energy for the 18th of April with 0.94 COD, 19.0 MAE, 1626.35 MSE, and, 40.32 RMSE performance measures. Which is almost the same as the expected performance. Figure 51 shows the forecasted vs. observed power for the mentioned day with the daily dip predicted.

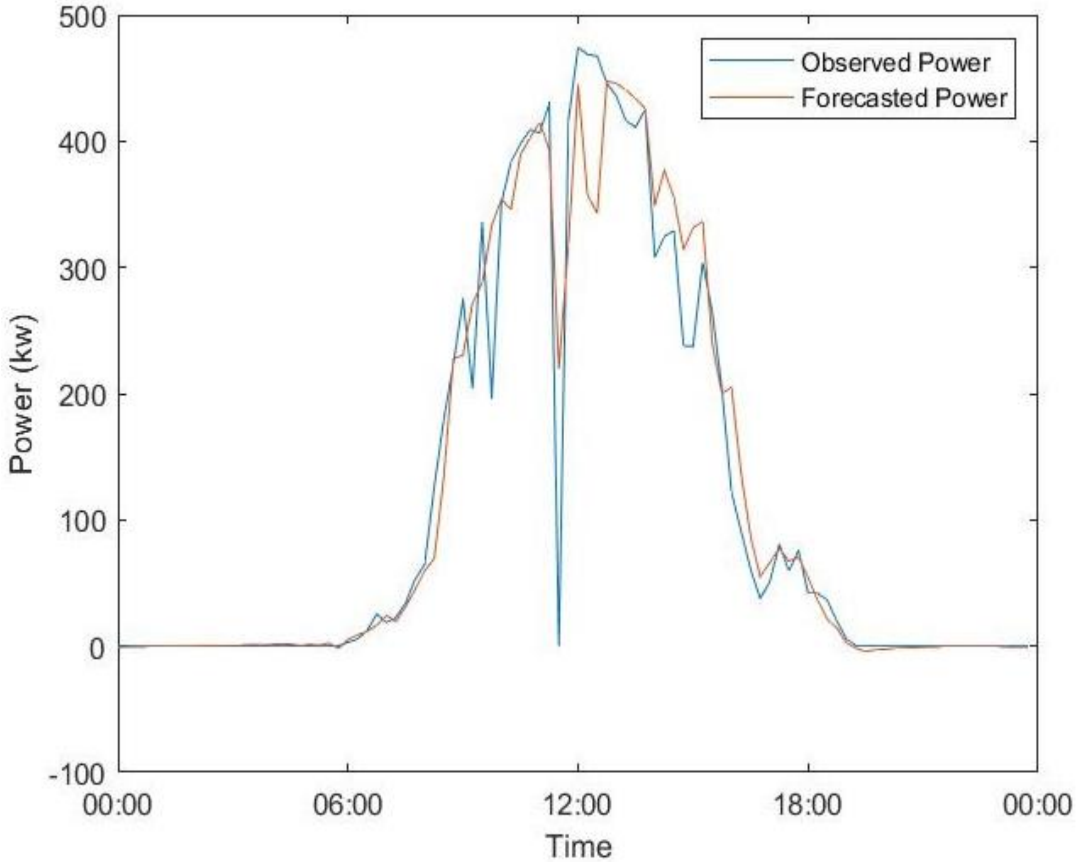


**Figure 51 Forecasted vs. observed power for the HANN (source: author)**

### *V.2.3. Performance comparison*

In this section, a performance comparison for all the models that were designed and tested in the previous subsections is provided. Table 21 summarizes the overall performance data of the different models.

**Table 21 Performance measures comparison (source: author)**

| Model | | Performance measures | | | |
|---|---|---|---|---|---|
| | | **COD ($R^2$)** | **MAE** | **MSE** | **RMSE** |
| MR | SMR | 0.94 | 14.84 | 1054.74 | 32.47 |
| | TMR | 0.68 | 45.83 | 5584.5 | 74.72 |
| | HMR | 0.95 | 16.05 | 835.68 | 28.90 |
| | **Average** | **0.86** | **25.57** | **2491.64** | **45.36** |
| ANN | SANN | 0.95 | 13.13 | 943.53 | 30.26 |
| | TANN | 0.75 | 36.38 | 4329.87 | 36.38 |
| | HANN | 0.96 | 13.52 | 914.10 | 30.23 |
| | **Average** | **0.89** | **21.01** | **2062.5** | **32.29** |

As Table 21 shows, the difference between MR and ANN is very clear in the time-series data, where TANN performance is highly superior compared to TMR. Moreover, even though SMR and SANN show comparable performances, the SMR is sensitive to the uncertainty in the input variables as discussed in the sensitivity analysis (section V.2.1.1). HANN has the highest COD, the lowest MSE, and RMSE, thus the HANN has the best overall performance in all the used measures except MAE where SANN has the lowest value. It can be noticed that ANN is generally overperformed MR models – as can also be observed from the diagrams in Figure 52 which show average overall performance for the MR and ANN.

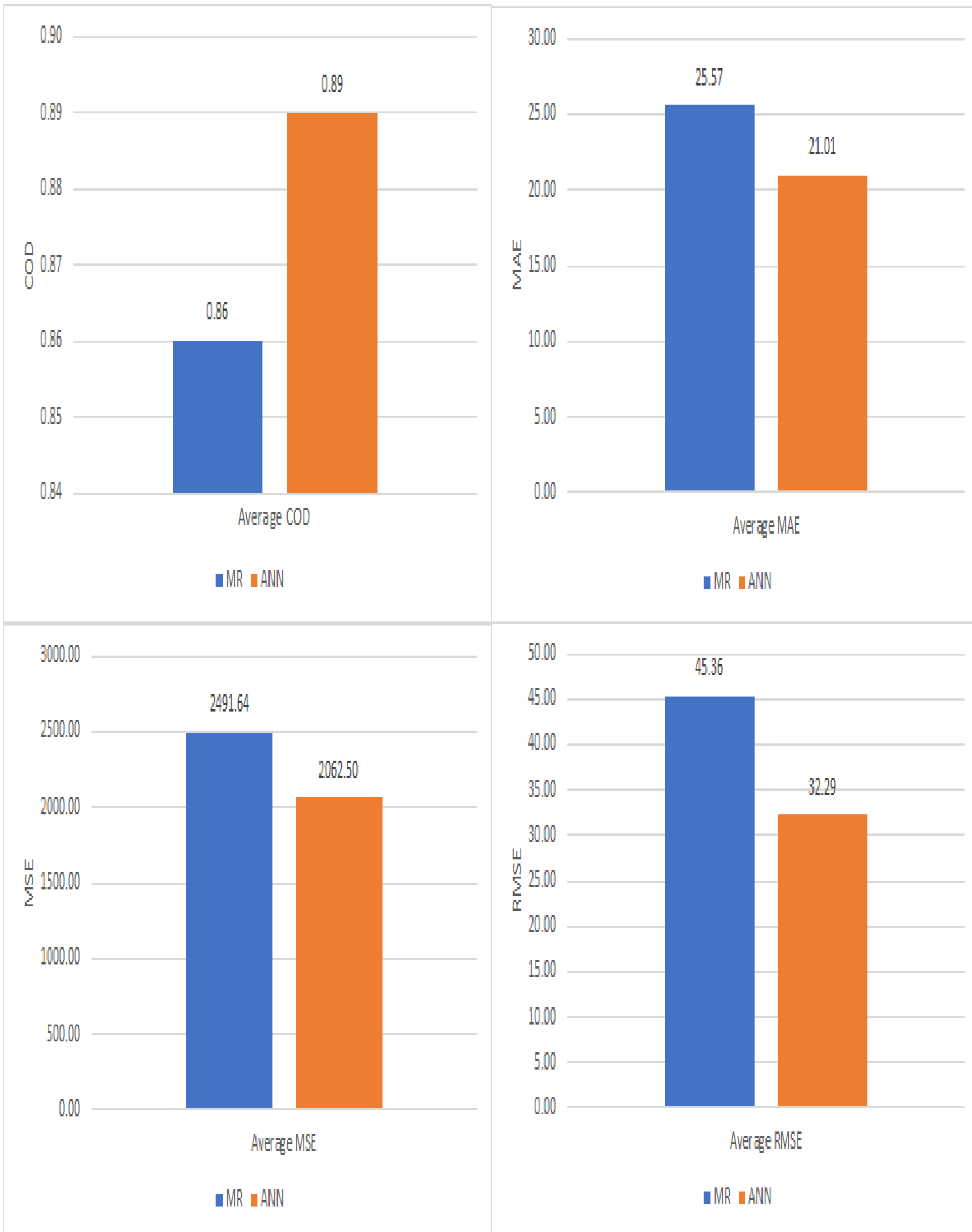


**Figure 52 Average performance measures comparison between MR and ANN (source: author)**

Figure 53 shows average performance measures comparison between structural, time-series, and hybrid methods. Time-series models have the worst performances. Although hybrid and structural models have close performance values, hybrid models perform vaguely better in the MSE and RMSE measures, while structural models overperform the hybrid ones in the MAE measure.

When comparing the performance of the models for forecasting PV output power on April 18th, 2020 similar conclusions can be made, as expected: ANN models were slightly overperforming MR ones (see Table 22).

Reiteratively, time-series models show unfavorable performance compared to the structural or hybrid models. The hybrid models show the highest COD and least MAE, MSE, and RMSE.

So, the outcomes of this chapter answer the first and the fourth research questions while focusing on the research gaps mentioned in III.9. Regarding the first research question which is “Which variables should be used to design, train, and build renewable energy forecasting models to improve the forecasting accuracy while reducing costs and computational complexity?”, it was proved that some variables might not be useful in the forecasting process, thus variables such as wind speed, snow depth, and precipitable water were excluded.

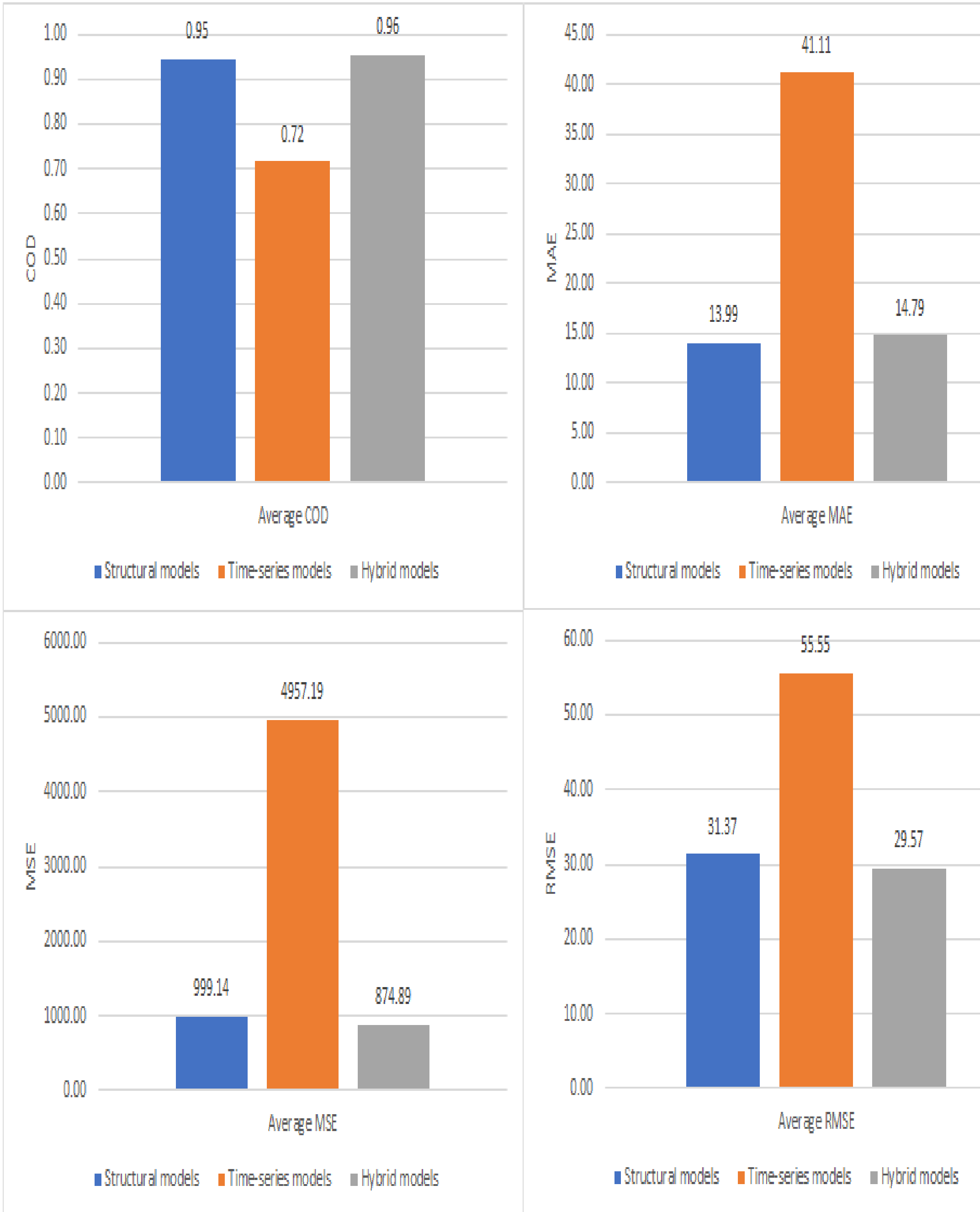


**Figure 53 Average performance measures comparison between structural, time-series, and hybrid models (source: author)**

**Table 22 Performance measures comparison for PV output power for the 18th of April 2020 (source: author)**

| Model | | Performance measures | | | |
|---|---|---|---|---|---|
| | | COD ($R^2$) | MAE | MSE | RMSE |
| MR | SMR | 0.92 | 22.44 | 1815.52 | 42.60 |
| | TMR | 0.88 | 32.97 | 3091.27 | 55.59 |
| | HMR | 0.93 | 22.26 | 1796.84 | 42.38 |
| | **Average** | **0.91** | **25.89** | **2234.54** | **46.86** |
| ANN | SANN | 0.93 | 20.96 | 1752.54 | 41.86 |
| | TANN | 0.87 | 32.57 | 3135.32 | 55.99 |
| | HANN | 0.94 | 19.0 | 1626.35 | 40.32 |
| | **Average** | **0.91** | **24.18** | **2171.40** | **46.06** |

The first research question deals with data and data availability problems. There are too many variables that can be used for designing and building renewable energy forecasting systems. Based on this chapter, it was found that some meteorological variables have a very low correlation with the PV power and hence can be eliminated (not included in the modeling process)

The third research question deals with forecasting models and techniques. Different forecasting methods lead to different forecasting accuracy. Based on the application and data availability, the forecasting model selection criteria can be tailored. Consequently, the fourth research question is “What are the algorithms and techniques to design, train, and build renewable energy forecasting models that can improve the forecasting accuracy based on the available data?”, the outcomes of this chapter shows that ANN models perform better regardless of input method, while hybrid input method is better in accuracy for both MR and ANN.

In the next chapter, the wind forecasting models are discussed.

# VI. WIND POWER FORECASTING DISCUSSION AND ANALYSIS

After building the models and the evaluation methods based on the mathematical background provided earlier in sections IV.6.1, IV.6.2, IV.6.3, and IV.8, the models were utilized to make a 24 hours wind energy forecasting. All models were trained and tested for 406 days, starting on the 5th of May, 2019, till the 13th of June, 2020. The training period lasts for 399 days (5th of May, 2019, till 6th of June, 2020), while the testing period lasts for 7 days (7th till 13th of June, 2020).

## VI.1. ANN forecasting model performance

ANN ML technique was utilized to build the ANN forecasting model as was discussed earlier. Over the training period, ANN shows poor performance in predicting 24 hours ahead wind energy with a MAE of 117.8, MSE of 21200.9, and COD of 0.06. After training, the testing period has taken place. As expected, ANN shows a very poor performance with 80.0 MAE, 8024.41 MSE, and 0.05 COD. Figure 54 below shows the observed (real) wind energy vs. the ANN forecasted output energy for the testing period.

Although ANN shows very poor abilities in 24 hours energy forecasting, this model shows very good abilities in long-term forecasting. ANN analysis shows that the ANN model was able to predict the total amount of energy generated in the train and test period with a 0.43% error, as will be shown later.

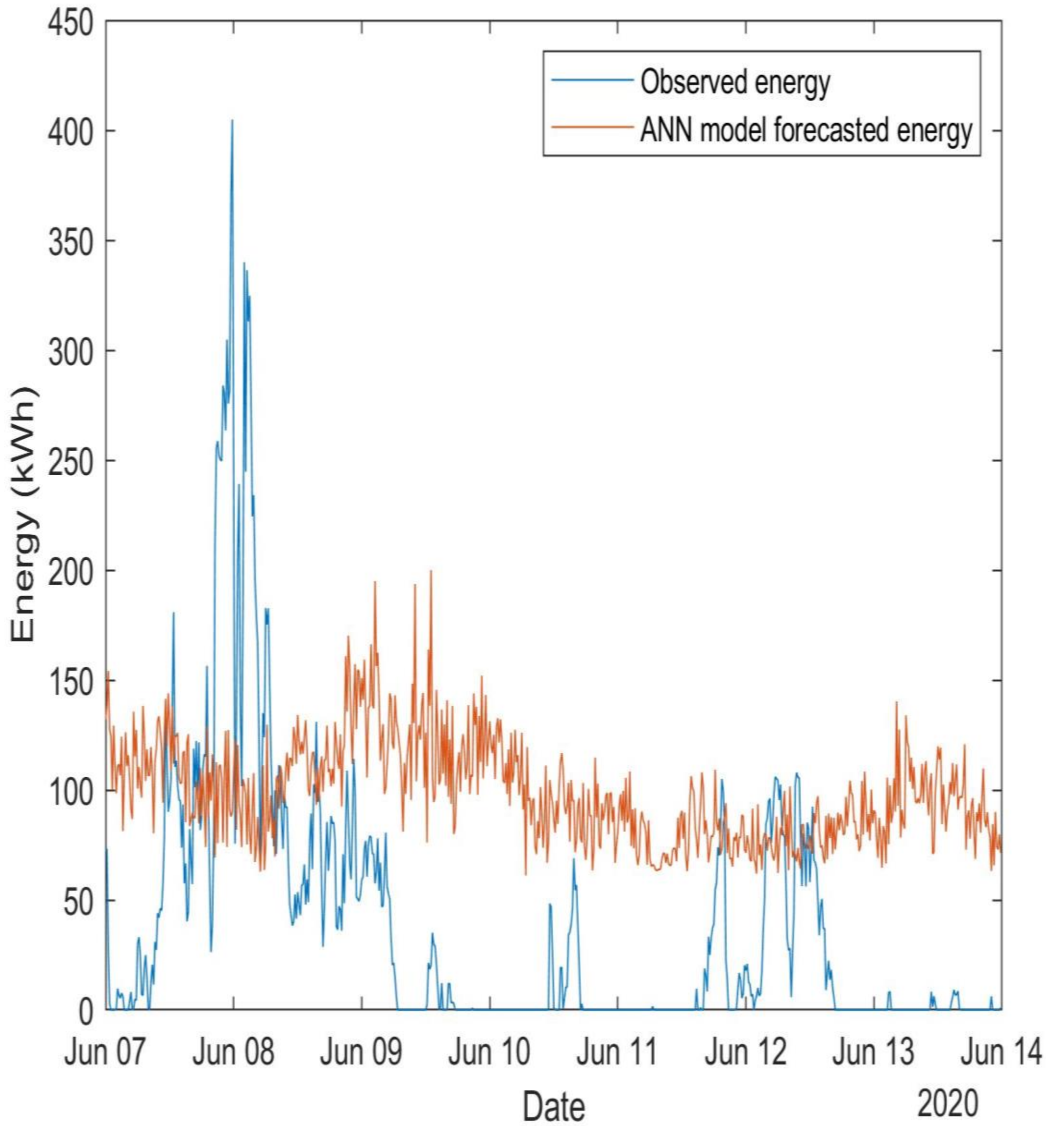


**Figure 54 Observed Vs. forecasted wind energy for the ANN model throughout the test period (source: author)**

## VI.2. SVR forecasting model performance

SVR ML technique was utilized to build the SVR forecasting model as was discussed earlier. Over the training period, SVM shows fair performance in predicting 24 hours ahead wind energy with a MAE of 68.22, MSE of 12995.2, and COD of 0.42. After training, the testing period has taken place. SVM shows poor performance with 42.68 MAE, 3874.623 MSE, and 0.01 COD. Figure 55 below shows the observed (real) wind energy vs. the SVM forecasted output energy for the testing period.

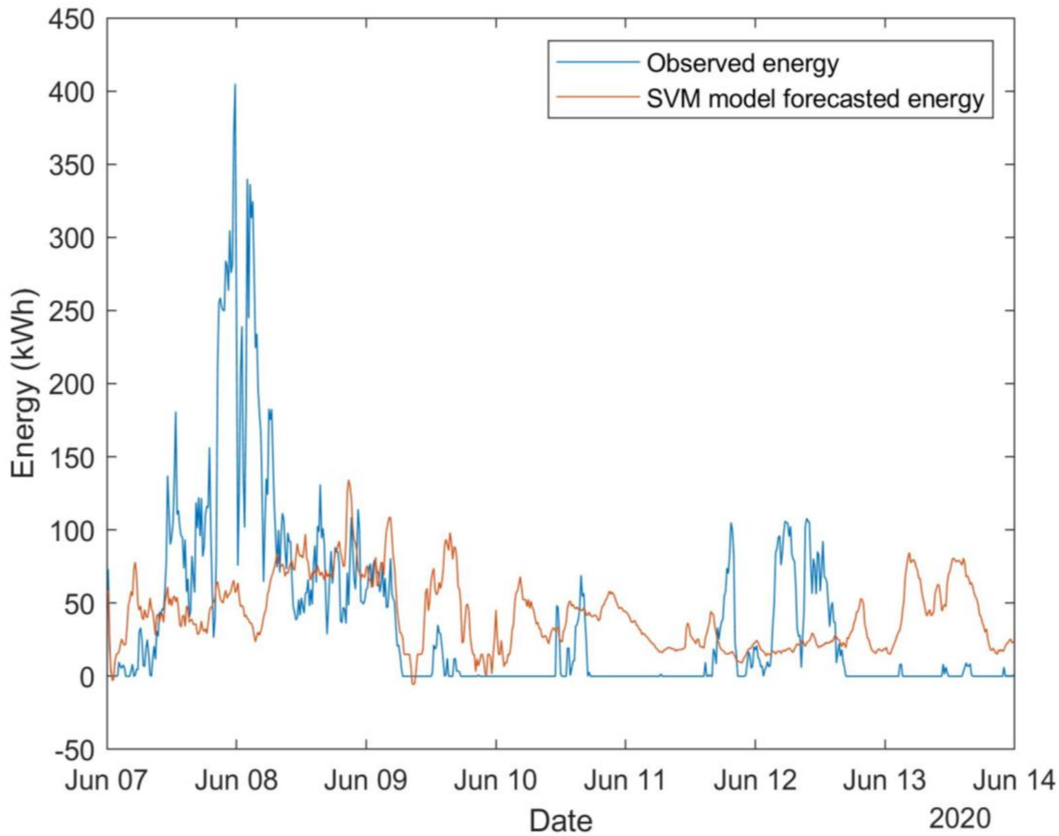


**Figure 55 Observed Vs. forecasted wind energy for the SVM model throughout the test period (source: author)**

As SVM shows fair abilities in 24 hours energy forecasting, it was expected that this model will show fair abilities in long-term forecasting as well. SVM analysis shows that the SVM model was able to predict the total amount of energy generated in the train and test period with a 28.08% error.

### VI.3. KNN forecasting model performance

KNN ML technique was utilized to build the KNN forecasting model as was discussed earlier. Over the training period, KNN shows a very good performance in predicting 24 hours ahead of wind energy with a MAE of 39.18, MSE of 3983.48, and COD of 0.82. After training, the testing period has taken place. KNN shows a very good performance with 19.07 MAE, 1088 MSE, and 0.72 COD. Figure 56 below shows the observed (real) wind energy vs. the KNN forecasted output power for the testing period.

Although KNN shows very good abilities in 24 hours energy forecasting, yet this model shows fair abilities in long-term forecasting (worse than ANN technique for example). KNN analysis shows that the KNN model was able to predict the total amount of energy generated in the train and test period with a 4.23% error.

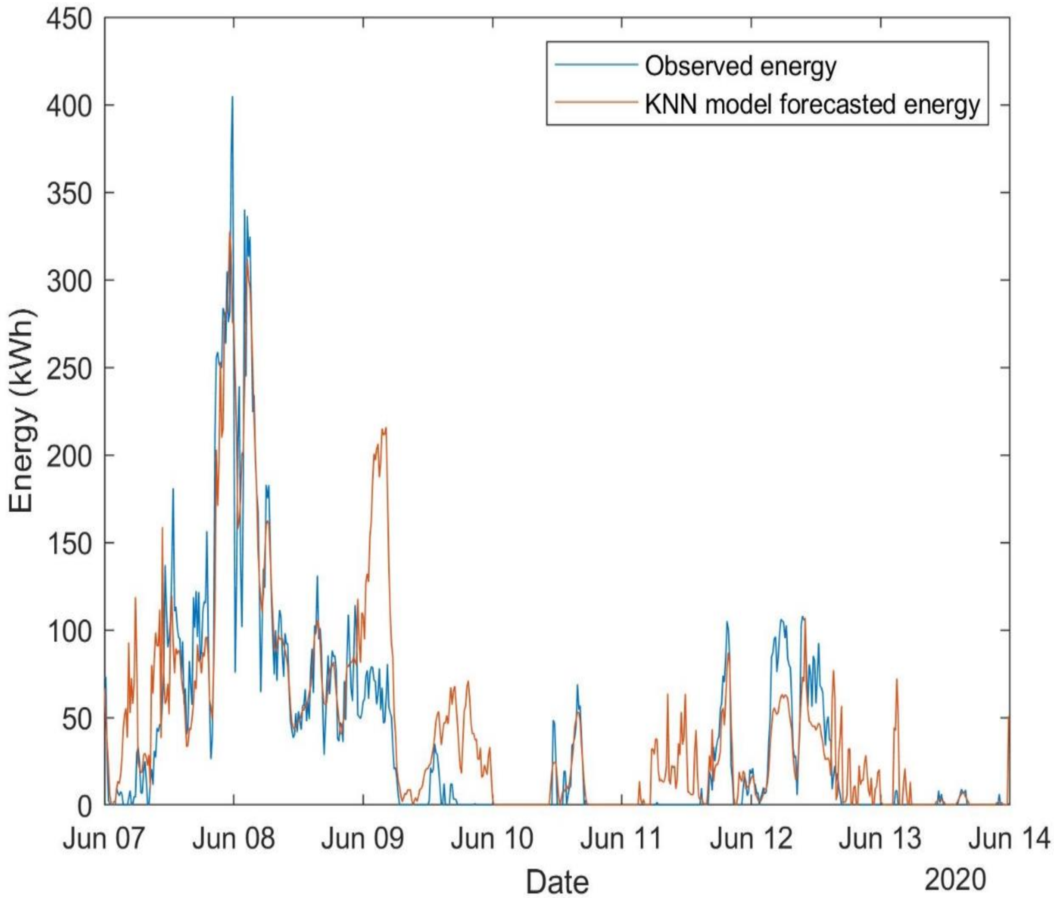


**Figure 56 Observed Vs. forecasted wind energy for the KNN model throughout the test period (source: author)**

### VI.4. Hybrid forecasting model performance

ANN, SVM, and KNN ML techniques were utilized to build the hybrid forecasting model. Over the training period, the hybrid model shows good performance in predicting 24 hours ahead wind energy with a MAE of 63.70, MSE of 9167.50, and COD of 0.6. After training, the testing period has taken place. The hybrid model also shows a good performance with 59.13 MAE, 5356.57 MSE, and 0.37 COD. Figure 57 below shows the observed (real) wind energy vs. the ANN forecasted output energy for the testing period.

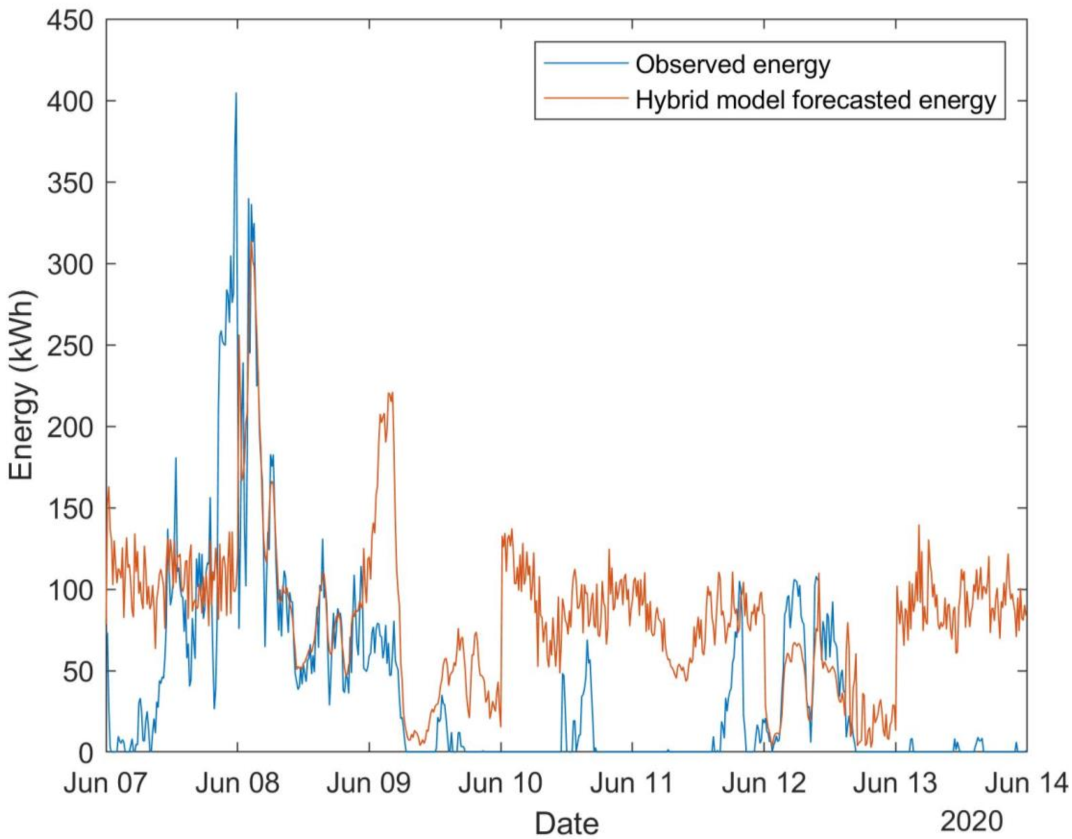


**Figure 57 Observed Vs. forecasted wind energy for the hybrid model throughout the test period (source: author)**

Although the hybrid model shows a bit lower forecasting abilities than KNN in 24 hours energy forecasting, the hybrid model shows very good abilities in long-term forecasting, better than KNN and SVM for example (more details in the following sections). The hybrid model analysis shows that this model was able to predict the total amount of energy generated in the train and test period with a 1.27% error.

What is worth adding regarding the hybrid method is the selected models for performing the next 24 hours prediction. As discussed earlier, one of the hybrid model's tasks is to choose a model to perform the prediction for the next 24 hours based on the past 24 performances. During the analysis, it was noticed that from the 406 days of training and testing ANN was never good enough to be selected by the algorithm to perform the next 24 hours forecast. On the other hand, KNN was selected 259 times, while SVM was selected 147 times as can be seen in Figure 58.

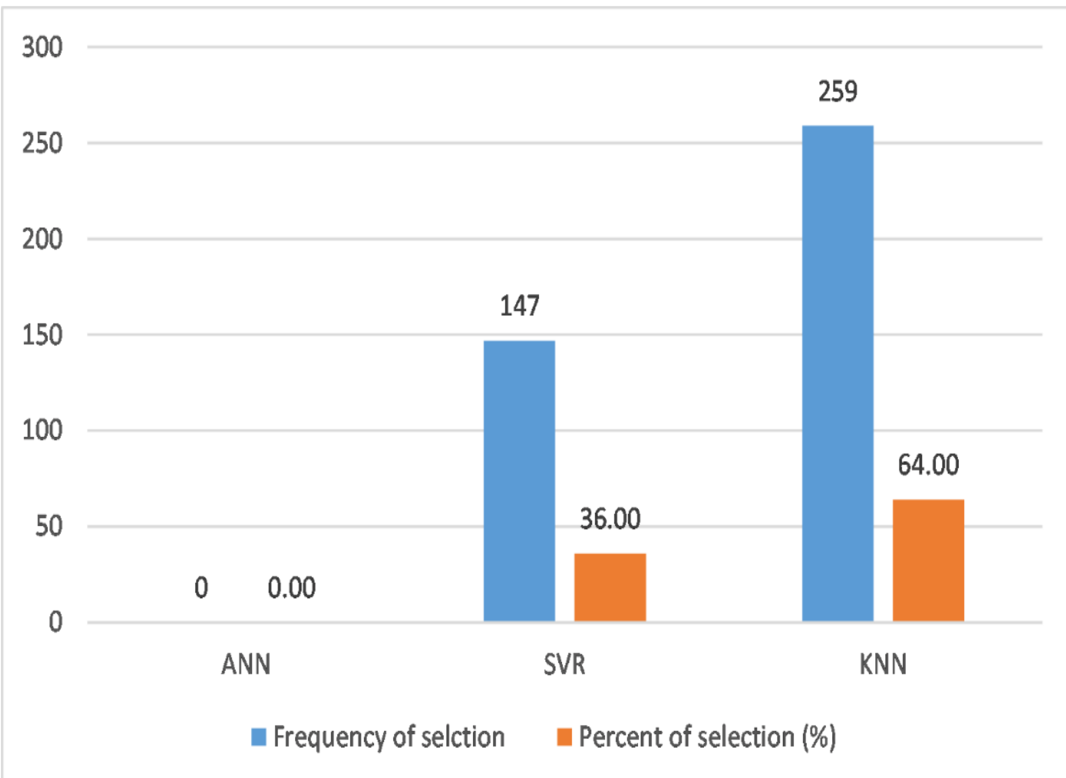


**Figure 58 The frequency and percent of techniques selected by the algorithms to perform the next 24 hours forecast (source: author)**

## VI.5. Performance comparison

This section shows and compares all the analysis results for each model. The prediction abilities for forecasting day-ahead wind power are analyzed and benchmarked, moreover, the long-term forecasting abilities are also provided. Table 23 shows the performance data of the different models during the training period.

**Table 23 Performance measures comparison during the training period (source: author)**

| Model | Performance measure | | |
|---|---|---|---|
| | MAE | MSE | COD |
| ANN | 117.83 | 21200 | 0.061 |
| SVR | 68.22 | 12995.24 | 0.42 |
| KNN | 39.18 | 3983.48 | 0.82 |
| Hybrid | 63.70 | 9167.50 | 0.60 |

During the training period, ANN shows the worst performance among the models built in this study. ANN COD is very low, additionally, MAE and MSE are almost 2 times more compared to SVR. SVR shows fair performance, better than ANN but way worse than hybrid or KNN. KNN shows the best performance for 24 hours ahead of wind energy forecasting. KNN has the least MAE, MSE, and highest COD. The hybrid model shows very good performance as well, its performance is close to KNN.

Comparing the performance by method score (MS) results in a similar conclusion where the ANN technique has the worst performance among the 4 models in every single measure, thus its score is 1, while KNN has the best performance in every single measure, and thus its score is 0 as can be seen in Figure 59.

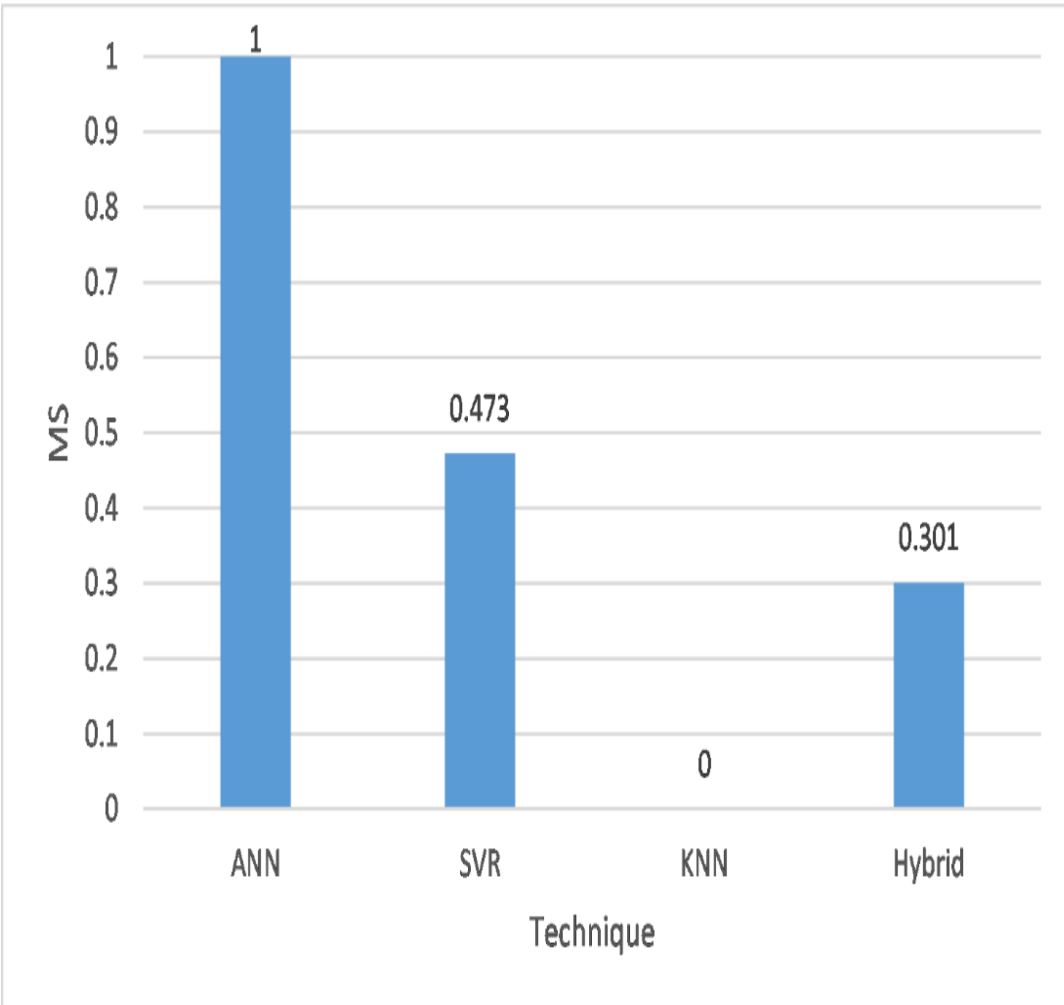


**Figure 59 MS for each technique during the training period (source: author)**

As discussed earlier, after the training, all of the models were tested for 7 days. Table 24 shows the performance data of the different models during the testing period.

**Table 24 Performance measures comparison during the testing period (source: author)**

| Model | Performance measure | | |
|---|---|---|---|
| | MAE | MSE | COD |
| ANN | 80.00 | 8024.41 | 0.03 |
| SVR | 42.68 | 3874.62 | 0.01 |
| KNN | 19.07 | 1088.05 | 0.72 |
| Hybrid | 59.13 | 5356.57 | 0.36 |

During the testing period, ANN shows the worst performance among the models built in this study. Here again, utilizing ANN technique results in very low COD and high MAE and MSE values. SVR shows fair performance, better than ANN but way worse KNN. KNN shows the best performance as it has the least MAE, MSE, and highest COD. The hybrid model shows good performance as well, yet the MAE and MSE values are higher than expected from the training period. Figure 60 shows the MS score for each technique during the testing period.

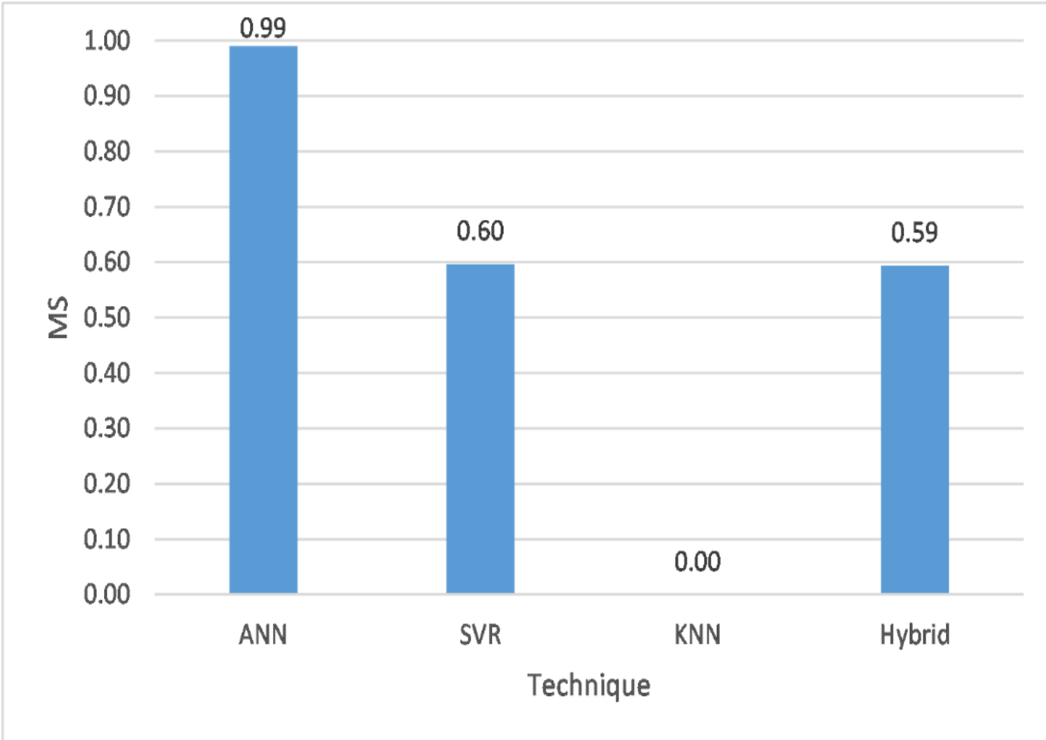


**Figure 60 MS for each technique during the testing period (source: author)**

Another important comparison to add is the long-term prediction performance during the training and testing period. Throughout the 406 days where the models are trained and tested, the total amount of energy generated by the wind turbine was 4899197.3 kWh. The sum of the daily forecasted energy for each model was calculated for 406 days, where the best long-term prediction performance can be conducted and compared to the total amount of energy generated by the wind turbine. Surprisingly, even though ANN shows the worst daily performance, yet, this model shows the best long-term prediction performance with only 0.43% error as can be seen in Figure 61. SVR shows the worst long-term prediction abilities with 28% errors. KNN shows good long-term prediction abilities with a 4.24% error rate, yet the hybrid method overperforms it with only 1.28% error.

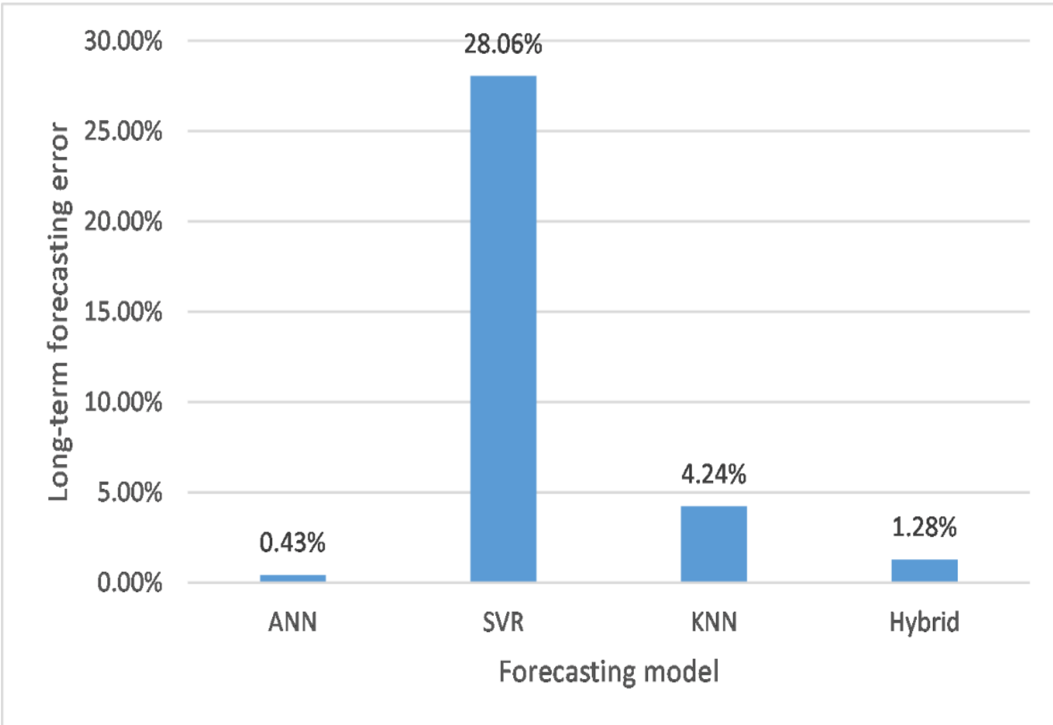


**Figure 61 Long-term forecasting error for each model (source: author)**

The outcomes of this chapter answer the third research question regarding forecasting models and techniques, while focusing on long and short-term wind forecasting (research gaps). All in all, even KNN shows very good day-ahead forecasting abilities, yet its long-term performance can be improved. The opposite case was observed in ANN, where the short-term prediction is poor, while the long-term is very accurate. The hybrid model inherited the good prediction abilities from each model, thus the suggested hybrid model shows very good abilities in both long and short-term prediction, as a result, it can be utilized in long and short wind energy forecasting while maintaining good forecasting accuracy.

# VII. THE IMPACT OF INPUT DATA RESOLUTION ON NEURAL NETWORK FORECASTING MODELS FOR WIND AND PHOTOVOLTAIC ENERGY GENERATION

While it has been established that different forecasting horizons lead to different accuracies, the impact of input data resolution could bear some clarification. Therefore, the objective of this chapter is to investigate how input data resolution affects prediction performance with the goal of helping wind and solar energy producers to improve their scheduling and to inform regulators on potential scheduling policy and planning issues.

To address the above objective, two sets of ANN time series forecasting models were designed, built (i.e. trained), and tested to forecast wind and PV out power for 24 hours ahead, each set utilizing input data with resolutions of 15, 30, and 60 minutes. Once the six models were trained, their accuracy was then calculated. Subsequently, a comparative analysis was conducted to determine the best settings leading to the best performance.

## VII.1. ANN time series forecasting models preparation

In this chapter PV and wind forecasting models using fully connected MLFFNNs were built and tested with three different resolutions as described in sections IV.6.1 and IV.9. This implies a differing number of input neurons for each resolution tested.

The number of input neurons, therefore, are 96, 48, and 24 for the 15, 30, and 60 input resolutions respectively for both PV and wind. Another important parameter for ANN is the number of hidden neurons.

Few hidden neurons might affect the ability of ANN to generate a proper function that solves the forecasting problem, while in the contrast, adding more hidden neurons might result in over-fitting of the training set and, therefore, lowering the ability of generalization (Setyawati, 2005). Hence, the number of hidden neurons was set to be 33% (one-third) of the number of inputs. In addition, instead of a time limit, the number of Epochs was set to a limit to control running time. Table 25 shows all settings of ANN parameters depending on input data resolution. For all ANN models trained here, the data was split into three segments: 70% for the training set, 15% for the validation set, and 15% for the test set. During each epoch, the training set is used to train the models and update the network weights and biases. While the validation set is used to monitor the errors during the training process. The training error normally decreases during each epoch, and this applies to the validation set error as well. However, when the network begins to overfit the data, the error on the validation set typically begins to increase. The network weights and biases are saved at the minimum value of the validation set error to ensure that no overfitting has occurred. The test set error is not used during training, but it is used to compare the performance of each epoch. In this work, the performance of each tested model was calculated during the training and testing period, i.e. using 85% of the original dataset.

**Table 25 ANN parameters for resolution testing (source: author)**

| Parameter | Description | Value for each resolution | | |
|---|---|---|---|---|
| | | 15 | 30 | 60 |
| Number of inputs | Number of input data variables | 96 | 48 | 24 |
| Number of outputs | Number of output forecasted variables | 1 | 1 | 1 |
| Number of hidden neurons | Number of hidden neurons | 32 | 16 | 8 |
| Maximum Epochs | Max. number of training iterations before training is stopped | 1000 | 1000 | 1000 |
| Maximum training time | Max. time before training is stopped | ∞ | ∞ | ∞ |
| Performance Goal | The min. target value of MSE | 0 | 0 | 0 |

## VII.2. Results and discussion

The performance of each model was calculated during the training and testing period (one full year). Additionally, as it is difficult to visualize the performance of each tested model for one year,  the performance of each forecasting model was visualized for the last week of the testing which covers the period 7th to 13th of June 2020 as can be seen in Figure 62 (solar) and Figure 63 (wind).

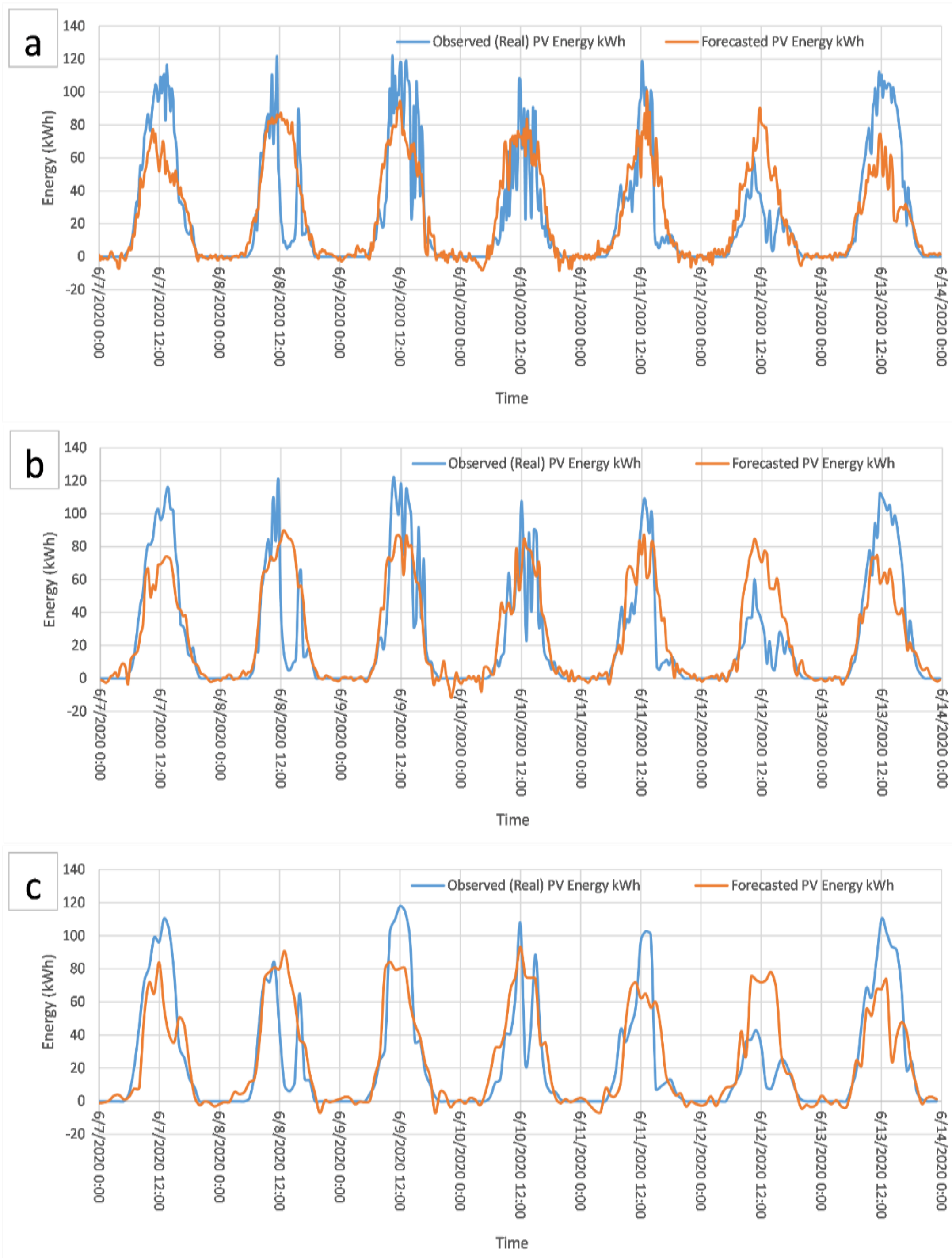


**Figure 62 PV energy forecasting model performance utilizing (a) 15; (b) 30; and (c) 60 minutes input data resolution (source: author)**

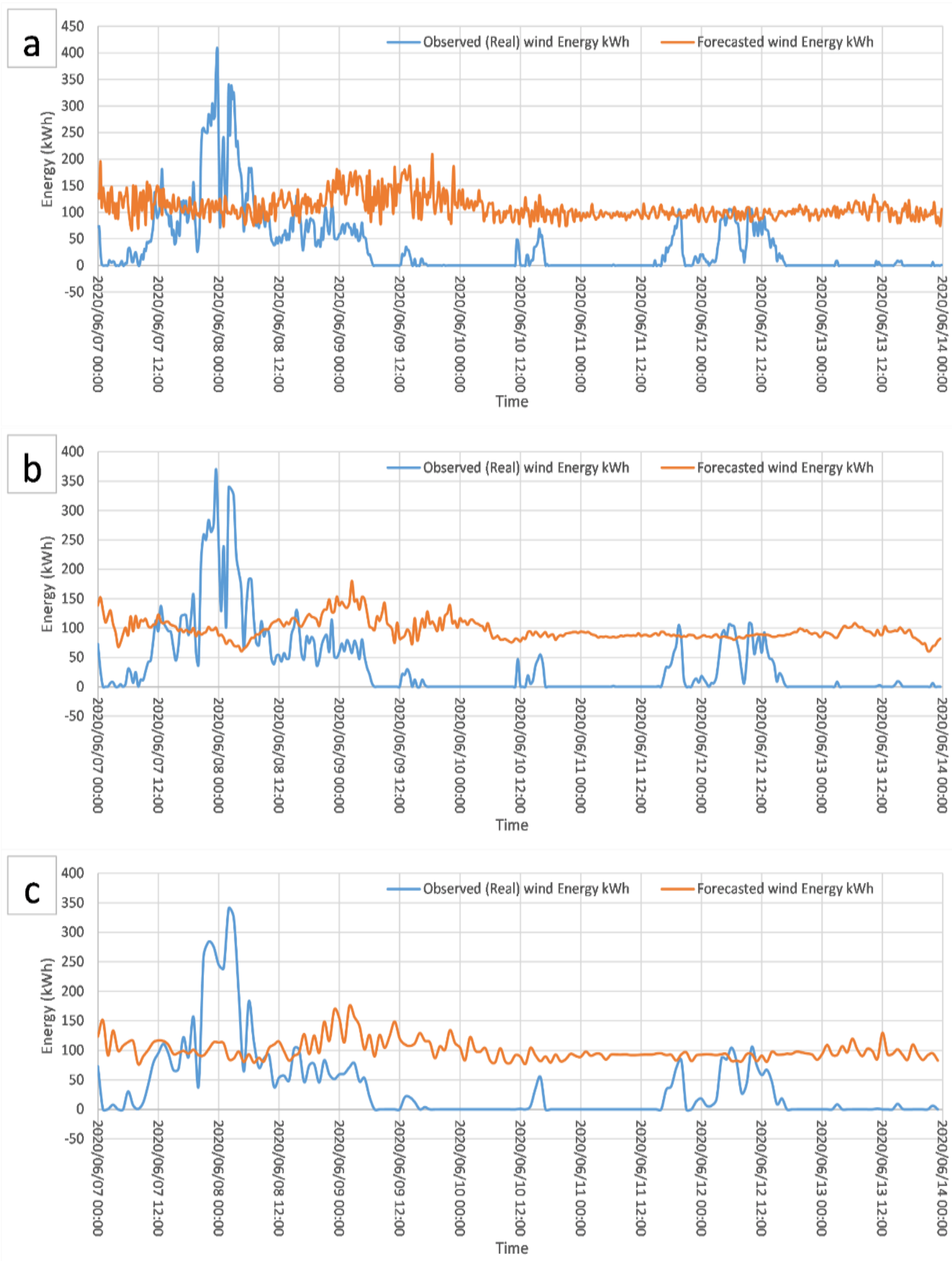


**Figure 63 Wind energy forecasting model performance utilizing (a) 15; (b) 30; and (c) 60 minutes input data resolution (source: author)**

In Figure 62 (a) it can be seen that the PV forecasting model utilizing 15 minutes of input data resolution has good prediction abilities, yet some errors can be observed. Specifically, larger error may be observed on the 8th of June where the model failed to predict the sudden dip that happened in the afternoon. As the input variables and the forecasted values have high resolution, the small fluctuations in the real energy production can be detected. For instance, the 9th and 10th of June show fast fluctuations in the produced energy. The forecasting model was partially able to predict these sudden fast fluctuations, yet it could not accurately predict steep movements.

Figure 62 (b) shows the performance of the PV forecasting model utilizing 30 minutes of input data. This model also shows good prediction abilities. Yet again, some errors can be observed especially on the 8th of June. It can be noticed here that the sudden production fluctuations can still be detected but smaller fluctuations could not be detected as frequently as in the previous model of higher input data resolution.

Figure 62 (c) shows the performance of the PV forecasting model utilizing a 60-minute resolution of input data. It can be seen that this model has higher forecasting errors for the 12th of June 2020. Also, as the input and forecasted data have a lower resolution than the two previous models, production fluctuations appear more smoothly.

Generally, it can be observed from Figure 62 that using ANN forecasting models utilizing only past energy data (time-series past generation data) leads to good forecasting performance. Still, the sudden fluctuations in energy production could not be accurately predicted. Moreover, different input data and forecasts resolutions show different behavior in detecting and forecasting these fluctuations (as will be discussed later in this section in comparison to wind forecasting).

Figure 63 shows the performance of the wind energy forecasting models utilizing different input data resolutions. Generally, it can be noticed that the ANN time-series forecasting model is not good enough in predicting wind energy.

Figure 63 (a) shows the performance utilizing input data of 15 minutes resolutions. This wind model was not able to predict the energy accurately especially in the last few days of testing (9th till 13th of June) when the actual produced energy was zero most of the time. Utilizing input data of 30 and 60 minutes didn't improve the forecasting performance much as can be seen in Figure 63 (b) and (c).

In the 15-minute resolution forecasting model, the forecasted values were fluctuating in a steeper manner than some of the real generation values. The steep fluctuations have still existed when utilizing 30, and 60-minute resolutions but in a less frequent manner.

As the designed ANN forecasted model only relies on the past generation time-series data, the output forecasted values were greatly deviating from the real values and huge errors were marked. This indicates that catching seasonality and patterns of wind energy generation by ANN forecasting models requires additional input variables compared to PV energy forecasting.

The results discussed above show a big variance in PV and wind forecasting performance as represented in Table 26. ANN time-series method was efficient in predicting the PV energy output with average COD of 0.75. Also, the average MAE and MSE are 9.95 and 303.55 respectively. Same method with the same data utilization approach shows very poor abilities in forecasting wind energy with a 0.064 COD, 117.59 MAE, and 20870.79 MSE.

Table 26 also shows that ANN time-series method has in general similar abilities in forecasting the PV output energy regardless of the input data resolutions. Although all performance measures are very close and comparable, the 60 minutes resolution shows higher values of MAE and MSE, yet slightly lower COD. This indicates higher input data resolutions lead to slightly better accuracy – and, interestingly, 30 minutes performs slightly better than 15 minutes in some performance measures like COD and MSE. However, the MAE value decreases for higher resolutions indicating slightly better prediction abilities.

**Table 26 Performance measures comparison for different resolutions (source: author)**

| Model | | Performance measures | | |
|---|---|---|---|---|
| | | COD | MAE | MSE |
| PV forecasting | 15 min. resolution | 0.75 | 9.72 | 300.13 |
| | 30 min. resolution | 0.76 | 10.00 | 297.81 |
| | 60 min. resolution | 0.74 | 10.13 | 312.73 |
| | **Average** | **0.75** | **9.95** | **303.55** |
| Wind forecasting | 15 min. resolution | 0.05 | 119.69 | 21040.5 |
| | 30 min. resolution | 0.07 | 116.29 | 20607.7 |
| | 60 min. resolution | 0.05 | 116.80 | 20964.0 |
| | **Average** | **0.06** | **117.59** | **20870.7** |

With respect to the effect of different input data resolutions on the forecasting model accuracy, it was found that performance measures are similar to the ones found in the literature. However, this study took an integrated view. For example, the MAE values for day ahead forecasting horizon varies between 7-12 depending on the input data and the technique utilized(Rajagukguk et al., 2020). Similarly, it was also confirmed from the literature that wind forecasting models tend to have higher MAE values. Moreover, for some models, the values of MAE do vary greatly between 1 (or even less) up to even a few hundred(H. Wang et al., 2019). Our results are more specific, however, as most other studies only present percentage difference.

COD values are close for all tested resolutions, while some significant differences can be seen between the different utilized resolutions, especially between the 60-minute and 15-minute resolutions in MAE and MSE. As can be seen in Figure 64 Performance measures comparison of PV energy forecasting utilizing different input data resolutions, utilizing data with 30 minutes resolutions instead of 60-minute resolutions improved (decreased) MAE by 1.33% and MSE by 4.77%. While utilizing data with 15-minute resolutions instead of 60-minute resolutions improved MAE by 4.10% and MSE by 4.03%. Utilizing a 15-minute resolution instead of 30 does not show any significant improvement. Actually, COD and MSE measures show a deteriorated improvement of 0.26% and 0.78%. Yet MAE shows a 2.81% improvement.

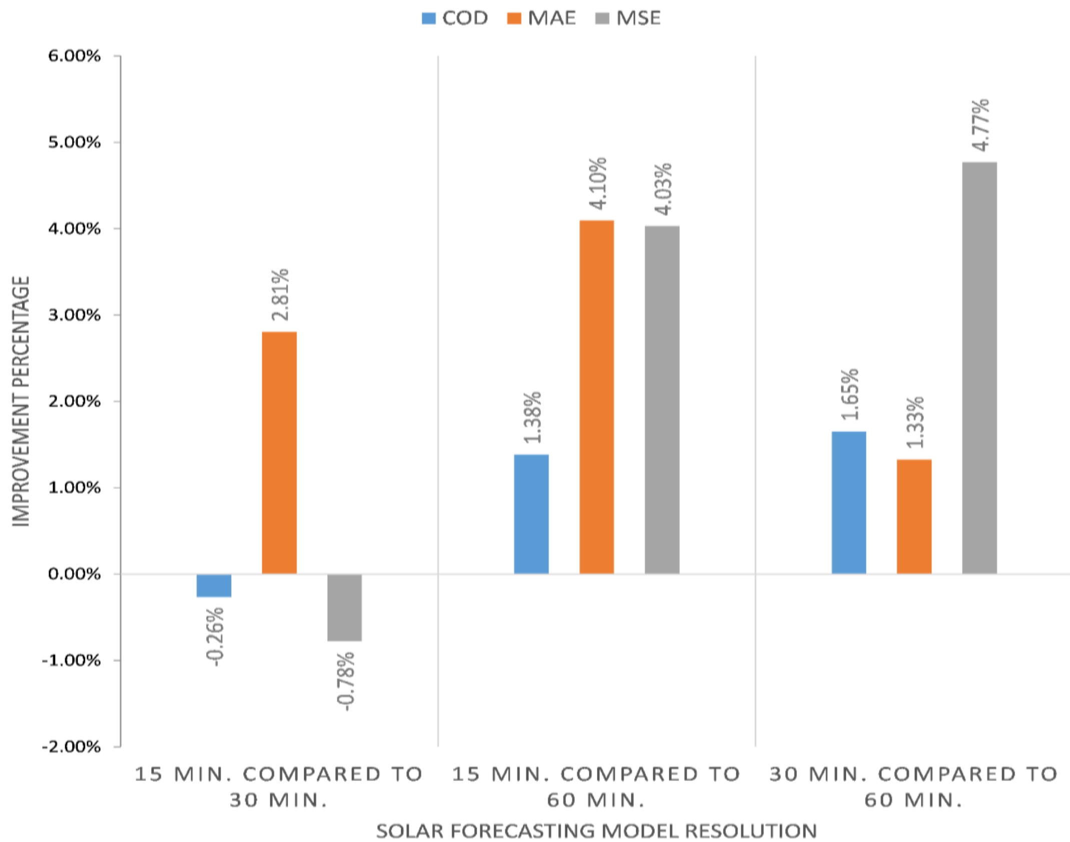


**Figure 64 Performance measures comparison of PV energy forecasting utilizing different input data resolutions**

For wind energy forecasting, different input data resolutions show some effects on the forecasting performance. The 30 mins resolutions show the lowest MAE and MSE. While higher values of MAE and MSE were observed utilizing 15 mins of input data resolution. Interestingly, here 60 minutes perform better, than 15 mins. But every resolution leads to weak performance in general.

As can be seen from Figure 65 Performance measures comparison of wind energy forecasting utilizing different input data resolutions, utilizing data with 30 minutes resolutions instead of 60-minute resolutions improved COD by 31.68%. Note that even after this huge improvement, COD values for both 30 and 60-minute resolutions are still low. Utilizing data with 15-minute resolutions instead of 60-minute resolutions improved COD by 1.18%, but MAE and MSE did not improve. Utilizing a 15-minute resolution instead of 30 does not show any improvement. on the contrary, all the performance measures show deteriorated values.

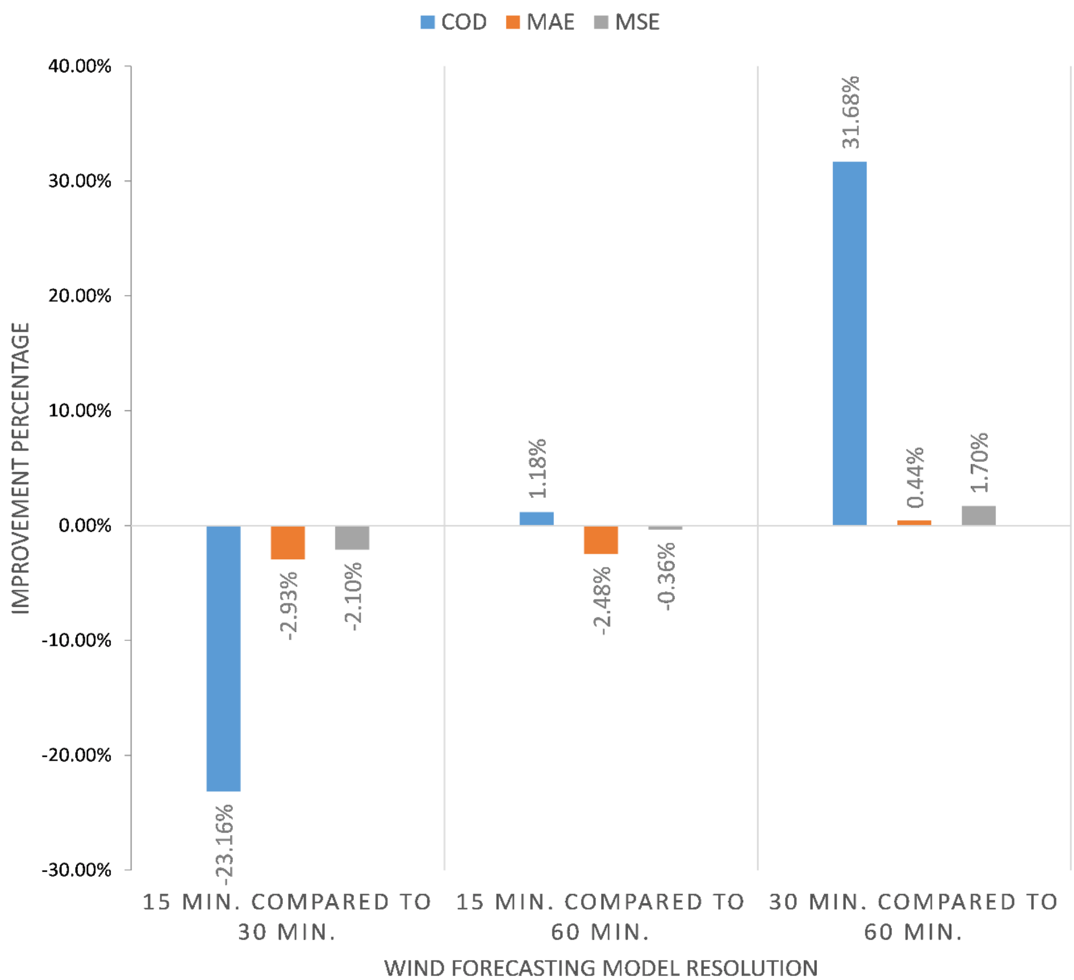


**Figure 65 Performance measures comparison of wind energy forecasting utilizing different input data resolutions**

Setting and formulating renewable energy policy is not a straightforward task as it can also be concluded from the practical forecasting provided in this research. There might be many applicable methods not only for forecasting renewable energy but also for evaluating these models. Thus far, there are no international standardized criteria according to which such models can be designed and evaluated, nor any international classification criteria.

So, the outcomes of this chapter answer the third research question while focusing on the research gaps mentioned in III.9. Regarding the second research question which is “What are the resolutions that can be utilized to design, train, and build renewable energy forecasting systems to assure the highest forecasting accuracy?”, and the fourth research question which is “Does the regulatory 15-minutes forecasting resolution provide similar accuracy when forecasting wind and solar?”, the results indicate that predicting wind power utilizing the time-series data alongside with ANN method might not lead to good forecasting accuracy at all. The results also indicate that for intraday renewable energy forecasting, using a 15-minute resolution might not lead to the best accuracy for all forecasting purposes. Input data resolutions have only a small effect on the accuracy of the ANN time-series PV forecasting model as forecasting measures are fairly close when utilizing 15, 30, or 60 minutes input data resolution. Yet, the 15 minutes resolution shows slightly better forecasting performance. Hence it can be concluded that different renewable energy predictions might require different models, methods, and input data settings. A powerful forecasting method for one renewable energy resource does not necessarily mean that this method is also powerful for forecasting other renewable sources.

# VIII. CONCLUSIONS

This thesis has demonstrated the use of artificial intelligence forecasting techniques in wind and solar energy prediction utilizing different input data methods. To that end, the work was done in four major parts. In the first part, a systematic literature review was conducted. In the second part, neural network and linear regression photovoltaic power forecasting models based on different input methods were evaluated. The third part studies wind energy forecasting. Finally, the fourth part discusses the impact of input data resolution on neural network forecasting models for wind and photovoltaic energy generation

In the first part, tech mining analysis in the field of renewable energy forecasting using AI technologies was discussed. To that end, documents from Scopus database were extracted after applying appropriate search methods. The extracted documents were then transformed into a special R data frame. Full-tech mining analysis was performed. The most active authors and countries were found. Collaboration between countries and the citations were also analyzed. Advanced analysis methods like thematic maps analysis were performed as well.

The results show that there is huge growth in the number of published documents in recent years. Asian countries like China, India, and Iran have the highest number of documents, the highest number of citations, and high number of citations per document. The results also show that Chinese institutions are very active not only in publishing but also in funding research in this field. Also, by analyzing thematic maps it was found that merging some research topics like smart grid and deep learning research or merging renewable energy forecasting and artificial intelligence, is highly needed. Finally, trend topics and subtopics in the field were identified.  The trendy topics have been changing over time. For instance, support vector regression and extreme learning were among the trendy topics back in 2014-2015, while in 2020-

2021 topics like climate change, long short term memory (lstm), and covid-19 are among the trendy topics list.

In the second part, two different techniques of PV energy prediction modeling, namely ANN and MR were analyzed. Depending on the input variables utilized, forecasting models were built using three different approaches: structural, time-series, and hybrid. The six models were built to predict the PV solar power for a 546 kWp grid-connected solar farm located in Hungary. This part is targeted to help PV farms improving their power prediction, therefore, the horizon and the resolution of the forecasts were set based on the forecasting regulations affecting certain grid operators in the European Union. Hence, all forecasting models were built and designed to forecast PV output power for a 24-hour ahead horizon with 15 minutes resolution. So, a historical data set including 3-years of geographical and meteorological variables was collected for the site of this specific PV farm along with actual PV power values. This data was used to build, train, and test the models.

The results indicate that ANN forecasting models have higher COD and lower MAE, MSE, RMSE values compared to the MR, regardless of the method used for building the forecasting models. It was also found that using the hybrid method to build prediction models results in better prediction accuracy for both MR and ANN while using the time-series method results in the least accurate forecasting models.

After analysing the results of this work using real farm data, it was confirmed that ANN technique performs better than the MR. This is true regardless of the input method used to build the models. It was also found that using the hybrid method of input data to build the forecasting models leads to better forecasting accuracy regardless of the technique used. The results of sensitivity analysis show that input variables and corresponding data quality have huge effects on the models' output when utilizing the structural technique. Consequently, in case of poor data quality or inaccurate weather data it is recommended to avoid using the structural method, especially when using the structural method to build MR forecasting models. To summarize, farm operators may have better results using ANN-based models with hybrid input approach.

Finally, some improvements might be done to expand this work. One possibility is to compare the performance of the tested forecasting models for different horizons and resolutions. In theory, the performance of the forecasting models is decreasing for longer forecasting horizons. It would be worthwhile to study the performance of the tested models for medium- and long-term forecasting horizons.

In the third part, four machine learning wind forecasting models were built and tested. First, three machine learning techniques namely ANN, SVR, and KNN were utilized to build three forecasting models to predict wind energy for the next 24 hours. Then, a hybrid model based on the three tested techniques was designed, trained, and tested. Moreover, a performance analysis was provided, comparing and benchmarking the different models. The data used in this study was collected from a 2 MW wind turbine located in Hungary. The data collection, model training, and model testing last for 406 days between 2019-2020. The results show that ANN has very poor performance in short-term (24 hours ahead) prediction, while, KNN shows a very good performance in predicting wind energy for the next 24 hours. The hybrid model shows a good performance, way better than ANN and SVR, but slightly lower than KNN.

On the contrary of ANN's poor performance on sort-term forecasting, it shows excellent long-term forecasting abilities. ANN was able to forecast the total wind energy generated by the wind turbine with a 0.43% error. KNN's performance for the long-term was worse than its performance for short-term forecasting with a 4.24% error in long-term forecasting. The hybrid method shows a very good long-term forecasting performance with a 1.28% error. It can be observed from the analysis that the hybrid model maintains good forecasting abilities in both long and short-term forecasting, thus it can be used in both cases while performing pretty well.

In the fourth part, different input data resolutions were used to build and test PV and wind energy forecasting. ANN modeling technique was used utilizing past energy values as time-series input data. A key characteristic of this part is that it used real site data for both PV and wind energy forecasting covering a full year, while most similar reports in the literature used either a 30-day moving window or a few months of historical production data. Energy values were collected from a 546 kWh grid-connected solar farm and a 2 MW wind turbine, both located in Hungary. All energy values were collected with three different resolutions of 15, 30, and 60 minutes – hence, energy for lower resolutions were not calculated. Then the forecasting models were trained and tested to predict the output energy of both PV and wind farms for a 24-hour ahead horizon, utilizing the above input data resolutions.

It was found that ANN time-series model was efficient in predicting the PV energy regardless of the input data resolution. In fact, input data resolutions have only a small effect on the accuracy of the ANN time-series PV forecasting model as forecasting measures are fairly close when utilizing 15, 30, or 60 minutes input data resolution. Yet, the 15 minutes resolution shows slightly better forecasting performance.

The same model approach shows poor performance in predicting wind energy. ANN time-series wind forecasting model has huge errors in forecasting wind energy regardless of the input data resolution. Yet, the 30 minutes input data resolution shows a slightly better performance. These results show that forecasting energy production in a 15-minute resolution might not assure high prediction accuracy for all renewable resources. Different renewable energy resources might need different input data resolutions to attain better forecasting accuracy.

In summary, it can be concluded that ANN time-series forecasting models are suitable for predicting PV output energy, while these models (at least in this form) might not be the best choice for predicting wind energy. Furthermore, utilizing different input data resolutions might not help in improving wind energy forecasting accuracy.

This part was limited to the ANN forecasting method, future research should investigate ANN model variants and other data input methods (such as structural and hybrid input data) for better wind energy prediction performance. One other clear limitation is that the work was only concerned with day-ahead forecasting, but in the future, the plan is to look at both shorter and longer-term predictions as well.

# IX. REFERENCES


(IEA), I. E. A. (2020). Hydropower Special Market Report Analysis and forecast to 2030. (IEA). https://iea.blob.core.windows.net/assets/4d2d4365-08c6-4171-9ea2-8549fabd1c8d/HydropowerSpecialMarketReport_corr.pdf

Abunima, H., Teh, J., & Jabir, H. J. (2019). A new solar radiation model for a power system reliability study. IEEE Access, 7, 64758-64766.

Agency, I. E. (2021). World Energy Outlook 2021 (World Energy Outlook, Issue. IEA. https://iea.blob.core.windows.net/assets/4ed140c1-c3f3-4fd9-acae-789a4e14a23c/WorldEnergyOutlook2021.pdf

Aggarwal, S., & Saini, L. (2014). Solar energy prediction using linear and non-linear regularization models: A study on AMS (American Meteorological Society) 2013–14 Solar Energy Prediction Contest. Energy, 78, 247-256.

Ahmed, R., Sreeram, V., Mishra, Y., & Arif, M. (2020). A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. Renewable and Sustainable Energy Reviews, 124, 109792.

Akkaya, B., & Çolakoğlu, N. (2019, 2019). Comparison of Multi-class Classification Algorithms on Early Diagnosis of Heart Diseases y-BIS Conference 2019: Recent Advances in Data Science and Business Analytics, İstanbul, Turkey.

Akter, M. S., & Shoeb, M. A. (2015). A Novel Model to Calculate Global Tilted Irradiation (GTI) from Solar Variables Using Netcdf and Rstudio. International Journal of Scientific & Engineering Research, 6(2).

Al-Messabi, N., Li, Y., El-Amin, I., & Goh, C. (2012). Forecasting of photovoltaic power yield using dynamic neural networks. The 2012 International Joint Conference on Neural Networks (IJCNN), Al Shafeey, M., & Harb, A. M. (2018). Photovoltaic as a promising solution for peak demands and energy cost reduction in Jordan. 2018 9th International Renewable Energy Congress (IREC),

Alkhayat, G., & Mehmood, R. (2021). A review and taxonomy of wind and solar energy forecasting methods based on deep learning. Energy and AI, 100060.

Almonacid, F., Pérez-Higueras, P., Fernández, E. F., & Hontoria, L. (2014). A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a PV generator. Energy Conversion and Management, 85, 389-398.

Almutairi, A., Sayed, K., Albagami, N., Abo-Khalil, A. G., & Saleeb, H. (2021). Multi-Port PWM DC-DC Power Converter for Renewable Energy Applications. Energies, 14(12), 3490. https://www.mdpi.com/1996-1073/14/12/3490

Alshafeey, M., Asemi, A., & Rashdan, O. (2018). Industrial revolution 4.0, renewable energy: A content analysis. Proceedings of FIKUSZ Symposium for Young Researchers,

Alshafeey, M., & Csáki, C. (2019). A Case Study of Grid-Connected Solar Farm Control Using Artificial Intelligence Genetic Algorithm to Accommodate Peak Demand. Journal of Physics: Conference Series,

AlShafeey, M., & Csáki, C. (2021). Evaluating neural network and linear regression photovoltaic power forecasting models based on different input methods. Energy Reports, 7, 7601-7614.

Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de-Pison, F. J., & Antonanzas-Torres, F. (2016). Review of photovoltaic power forecasting. Solar Energy, 136, 78-111.

Ashi, A., Joudeh, A. A., Shafeey, M., Sababha, B. H., & Istehkam, S. N. (2014). A PV solar tracking system: Design, implementation and algorithm evaluation. Information and Communication Systems (ICICS), 2014 5th International Conference on, Aslam, S., Herodotou, H., Mohsin, S. M., Javaid, N., Ashraf, N., Aslam, S. J. R., & Reviews, S. E. (2021). A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids. 144, 110992.

Awad, M., & Khanna, R. (2015). Support vector regression. In Efficient learning machines (pp. 67-80). Springer.

Bacher, P., Madsen, H., & Nielsen, H. A. (2009). Online short-term solar power forecasting. Solar energy, 83(10), 1772-1783.

Bamisile, O., Oluwasanmi, A., Obiora, S., Osei-Mensah, E., Asoronye, G., & Huang, Q. (2020). Application of deep learning for solar irradiance and solar photovoltaic multi-parameter forecast. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 1-21.

Barbosa de Alencar, D., de Mattos Affonso, C., Limão de Oliveira, R. C., Moya Rodriguez, J. L., Leite, J. C., & Reston Filho, J. C. (2017). Different models for forecasting wind power generation: Case study. Energies, 10(12), 1976.

Barbounis, T. G., Theocharis, J. B., Alexiadis, M. C., & Dokopoulos, P. S. (2006). Long-term wind speed and power forecasting using local recurrent neural network models. IEEE Transactions on Energy Conversion, 21(1), 273-284.

Barsky, R. B., & Kilian, L. J. J. o. E. P. (2004). Oil and the macroeconomy since the 1970s. 18(4), 115-134.

Blackwood, M. J. U. J. o. M. M. O. T. (2016). Maximum efficiency of a wind turbine. 6(2), 2.

Bortoluzzi, M., de Souza, C. C., Furlan, M. J. R., & Reviews, S. E. (2021). Bibliometric analysis of renewable energy types using key performance indicators and multicriteria decision models. 143, 110958.

Brook, M. J., & Finney, B. A. (1987). Generation of bivariate solar radiation and temperature time series. Solar Energy, 39(6), 533-540.

Brown, T. (2021). Hydropower in the Context of Sustainable Energy Supply: A Review of Technologies and Challenges.

Bull, S. R. (2001). Renewable energy today and tomorrow. Proceedings of the IEEE, 89(8), 1216-1226.

Cammarano, A., Petrioli, C., & Spenza, D. (2012). Pro-Energy: A novel energy prediction model for solar and wind energy-harvesting wireless sensor networks. 2012 IEEE 9th International Conference on Mobile Ad-Hoc and Sensor Systems (MASS 2012),

Cavalcante, R. L., Costa, T. O., Almeida, M. P., Williamson, S., Galhardo, M. A. B., & Macêdo, W. N. (2021). Photovoltaic penetration in isolated thermoelectric power plants in Brazil: Power regulation and specific consumption aspects. International Journal of Electrical Power & Energy Systems, 129, 106648.

Chen, C., Duan, S., Cai, T., & Liu, B. (2011). Online 24-h solar power forecasting based on weather type classification using artificial neural network. Solar energy, 85(11), 2856-2870.

Chi, Z., Haikun, W., Tingting, Z., Kanjian, Z., & Tianhong, L. (2015). Comparison of two multi-step ahead forecasting mechanisms for wind speed based on machine learning models. 2015 34th Chinese Control Conference (CCC),

Chu, Y., Urquhart, B., Gohari, S. M., Pedro, H. T., Kleissl, J., & Coimbra, C. F. (2015). Short-term reforecasting of power output from a 48 MWe solar PV plant. Solar Energy, 112, 68-77.

Cococcioni, M., D'Andrea, E., & Lazzerini, B. (2011). 24-hour-ahead forecasting of energy production in solar PV systems. 2011 11th International Conference on Intelligent Systems Design and Applications,

Comsan, M. (2010). Nuclear electricity for sustainable development: Egypt a case study. Energy Conversion and Management, 51(9), 1813-1817.

Conti, J., Holtberg, P., Diefenderfer, J., LaRose, A., Turnure, J. T., & Westfall, L. (2016). International energy outlook 2016 with projections to 2040.

Council, W. E. (2019). World Energy Insights Brief 2019 (World Energy Insights Brief, Issue. W. E. Council.

Das, U. K., Tey, K. S., Seyedmahmoudian, M., Mekhilef, S., Idris, M. Y. I., Van Deventer, W., . . . Stojcevski, A. (2018). Forecasting of photovoltaic power generation and model optimization: A review. Renewable and Sustainable Energy Reviews, 81, 912-928.

Daut, M. A. M., Hassan, M. Y., Abdullah, H., Rahman, H. A., Abdullah, M. P., Hussin, F. J. R., & Reviews, S. E. (2017). Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review. 70, 1108-1118.

De Giorgi, M. G., Congedo, P. M., & Malvoni, M. (2014). Photovoltaic power forecasting using statistical methods: impact of weather data. IET Science, Measurement & Technology, 8(3), 90-97.

de Marcos, R. A., Bello, A., & Reneses, J. (2019). Electricity price forecasting in the short term hybridising fundamental and econometric modelling. Electric Power Systems Research, 167, 240-251.

Debnath, K. B., Mourshed, M. J. R., & Reviews, S. E. (2018). Forecasting methods in energy planning models. 88, 297-325.

Devaraj, J., Madurai Elavarasan, R., Shafiullah, G., Jamal, T., & Khan, I. (2021). A holistic review on energy forecasting using big data and deep learning models. International Journal of Energy Research.

Diagne, M., David, M., Lauret, P., Boland, J., & Schmutz, N. (2013). Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. Renewable and Sustainable Energy Reviews, 27, 65-76.

Ding, M., Zhou, H., Xie, H., Wu, M., Nakanishi, Y., & Yokoyama, R. J. N. (2019). A gated recurrent unit neural networks based wind speed error correction model for short-term wind power forecasting. 365, 54-61.

Ellabban, O., Abu-Rub, H., Blaabjerg, F. J. R., & reviews, s. e. (2014). Renewable energy resources: Current status, future prospects and their enabling technology. 39, 748-764.

Ellegaard, O., & Wallin, J. A. J. S. (2015). The bibliometric analysis of scholarly production: How great is the impact? , 105(3), 1809-1831.

Elsheikh, A. H., Sharshir, S. W., Abd Elaziz, M., Kabeel, A., Guilan, W., & Haiou, Z. (2019). Modeling of solar energy systems using artificial neural network: A comprehensive review. Solar Energy, 180, 622-639.

Eseye, A. T., Zhang, J., Zheng, D., Ma, H., & Jingfu, G. (2017). Short-term wind power forecasting using a double-stage hierarchical hybrid GA-ANN approach. 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA),

Esfahani, H., Tavasoli, K., Jabbarzadeh, A. J. I. J. o. D., & Science, N. (2019). Big data and social media: A scientometrics analysis. 3(3), 145-164.

Evans, A., Strezov, V., & Evans, T. J. (2010). Sustainability considerations for electricity generation from biomass. Renewable and sustainable energy reviews, 14(5), 1419-1427.

Farooq, R. J. I. J. o. M. R. (2017). A framework for identifying research gap in social sciences: Evidence from the past. 16(4), 66-75.

Foley, A. M., Leahy, P. G., Marvuglia, A., & McKeogh, E. J. (2012). Current methods and advances in forecasting of wind power generation. Renewable Energy, 37(1), 1-8.

Galicia, A., Talavera-Llames, R., Troncoso, A., Koprinska, I., & Martínez-Álvarez, F. (2019). Multi-step forecasting for big data time series based on ensemble learning. Knowledge-Based Systems, 163, 830-841.

Gallego, C., Pinson, P., Madsen, H., Costa, A., & Cuerva, A. J. A. E. (2011). Influence of local wind speed and direction on wind power dynamics–Application to offshore very short-term forecasting. 88(11), 4087-4096.

Garud, K. S., Jayaraj, S., & Lee, M. Y. J. I. J. o. E. R. (2021). A review on modeling of solar photovoltaic systems using artificial neural networks, fuzzy logic, genetic algorithm and hybrid models. 45(1), 6-35.

Ghadi, M. J., Gilani, S. H., Afrakhte, H., & Baghramian, A. (2014). A novel heuristic method for wind farm power prediction: A case study. International Journal of Electrical Power & Energy Systems, 63, 962-970.

Ghritlahre, H. K., & Prasad, R. K. (2018). Application of ANN technique to predict the performance of solar collector systems-A review. Renewable and Sustainable Energy Reviews, 84, 75-88.

Gneiting, T. (2011). Quantiles as optimal point forecasts. International Journal of forecasting, 27(2), 197-207.

Görig, M., & Breyer, C. (2016). Energy learning curves of PV systems. Environmental Progress & Sustainable Energy, 35(3), 914-923.

Goswami, D. Y., & Kreith, F. (2015). Energy efficiency and renewable energy handbook. CRC Press.

Grigsby, L. L. (2018). Electric power generation, transmission, and distribution. CRC press.

Haghverdi, A., Washington-Allen, R. A., Leib, B. G. J. C., & Agriculture, E. i. (2018). Prediction of cotton lint yield from phenology of crop indices using artificial neural networks. 152, 186-197.

Hao, Y., & Tian, C. J. A. e. (2019). A novel two-stage forecasting model based on error factor and ensemble method for multi-step wind power forecasting. 238, 368-383.

Harrington, P. (2012). Machine learning in action. Manning Publications Co.

Heo, J., Song, K., Han, S., & Lee, D.-E. (2021). Multi-channel convolutional neural network for integration of meteorological and geographical features in solar power forecasting. Applied Energy, 295, 117083.

Hossain, M. A., Chakrabortty, R. K., Elsawah, S., & Ryan, M. J. J. J. o. C. P. (2021). Very short-term forecasting of wind power generation using hybrid deep learning model. 296, 126564.

Huang, X., Han, S., Huang, W., & Liu, X. (2013). Enhancing solar cell efficiency: the search for luminescent materials as spectral converters. Chemical Society Reviews, 42(1), 173-201.

Ineichen, P. J. S. E. (2006). Comparison of eight clear sky broadband models against 16 independent data banks. 80(4), 468-478.

Infield, D., & Freris, L. (2020). Renewable energy in power systems. John Wiley & Sons.

Inman, R. H., Pedro, H. T., & Coimbra, C. F. (2013). Solar forecasting methods for renewable energy integration. Progress in energy and combustion science, 39(6), 535-576.

Jafarzadeh, S., Fadali, M. S., & Evrenosoglu, C. Y. (2012). Solar power prediction using interval type-2 TSK modeling. IEEE Transactions on Sustainable Energy, 4(2), 333-339.

Jäger-Waldau, A. (2006). European Photovoltaics in world wide comparison. Journal of non-crystalline solids, 352(9-20), 1922-1927.

Jerez, S., Tobin, I., Vautard, R., Montávez, J. P., López-Romero, J. M., Thais, F., . . . Déqué, M. (2015). The impact of climate change on photovoltaic power generation in Europe. Nature communications, 6, 10014.

Jung, J., & Broadwater, R. P. (2014). Current status and future advances for wind speed and power forecasting. Renewable and Sustainable Energy Reviews, 31, 762-777.

Jursa, R., & Rohrig, K. (2008). Short-term wind power forecasting using evolutionary algorithms for the automated specification of artificial intelligence models. International Journal of Forecasting, 24(4), 694-709.

Kabir, E., Kumar, P., Kumar, S., Adelodun, A. A., & Kim, K.-H. (2018). Solar energy: Potential and future prospects. Renewable and Sustainable Energy Reviews, 82, 894-900.

Khatib, T., Mohamed, A., & Sopian, K. (2012). A review of solar energy modeling techniques. Renewable and Sustainable Energy Reviews, 16(5), 2864-2869.

Khishe, M., Mosavi, M., & Moridi, A. (2018). Chaotic fractal walk trainer for sonar data set classification using multi-layer perceptron neural network and its hardware implementation. Applied Acoustics, 137, 121-139.

Kim, P. (2017). Matlab deep learning. With Machine Learning, Neural Networks and Artificial Intelligence, 130, 21.

Koohi-Fayegh, S., & Rosen, M. A. J. J. o. E. S. (2020). A review of energy storage types, applications and recent developments. 27, 101047.

Lan, H., Zhang, C., Hong, Y.-Y., He, Y., & Wen, S. (2019). Day-ahead spatiotemporal solar irradiation forecasting using frequency-based hybrid principal component analysis and neural network. Applied Energy, 247, 389-402.

Lantz, B. (2019). Machine learning with R: expert techniques for predictive modeling. Packt Publishing Ltd.

Leema, N., Nehemiah, H. K., & Kannan, A. (2016). Neural network classifier optimization using differential evolution with global information and back propagation algorithm for clinical datasets. Applied Soft Computing, 49, 834-844.

Leva, S., Dolara, A., Grimaccia, F., Mussetta, M., & Ogliari, E. (2017). Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power. Mathematics and computers in simulation, 131, 88-100.

Lewis, N. S. (2007). Toward cost-effective solar energy use. science, 315(5813), 798-801.

Li, C., Lin, S., Xu, F., Liu, D., & Liu, J. (2018a). Short-term wind power prediction based on data mining technology and improved support vector machine method: A case study in Northwest China. Journal of Cleaner Production, 205, 909-922.

Li, C., Lin, S., Xu, F., Liu, D., & Liu, J. J. J. o. C. P. (2018b). Short-term wind power prediction based on data mining technology and improved support vector machine method: A case study in Northwest China. 205, 909-922.

Li, L.-L., Zhao, X., Tseng, M.-L., & Tan, R. R. (2020). Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm. Journal of Cleaner Production, 242, 118447.

Linguet, L., Pousset, Y., & Olivier, C. (2016). Identifying statistical properties of solar radiation models by using information criteria. Solar Energy, 132, 236-246.

Liu, H., Chen, C., Lv, X., Wu, X., & Liu, M. (2019). Deterministic wind energy forecasting: A review of intelligent predictors and auxiliary methods. Energy Conversion and Management, 195, 328-345.

Liu, H., & Chen, C. J. A. E. (2019). Data processing strategies in wind energy forecasting models and applications: A comprehensive review. 249, 392-408.

Machol, B., & Rizk, S. (2013). Economic value of US fossil fuel electricity health impacts. Environment international, 52, 75-80.

Madsen, D. N., Hansen, J. P. J. R., & Reviews, S. E. (2019). Outlook of solar energy in Europe based on economic growth characteristics. 114, 109306.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. PloS one, 13(3), e0194889.

Maldonado-Correa, J., Solano, J., & Rojas-Moncayo, M. (2019). Wind power forecasting: A systematic literature review. Wind Engineering, 0309524X19891672.

Mandil, C. (2004). World energy outlook 2004. World, 2004.

Mattei, M., Notton, G., Cristofari, C., Muselli, M., & Poggi, P. (2006). Calculation of the polycrystalline PV module temperature using a simple method of energy balance. Renewable energy, 31(4), 553-567.

Mellit, A., Massi Pavan, A., Ogliari, E., Leva, S., & Lughi, V. (2020). Advanced Methods for Photovoltaic Output Power Forecasting: A Review. Applied Sciences, 10(2), 487.

Nam, K., Hwangbo, S., & Yoo, C. (2020). A deep learning-based forecasting model for renewable energy scenarios to guide sustainable energy policy: A case study of Korea. Renewable and Sustainable Energy Reviews, 122, 109725.

Nelson, V. C., & Starcher, K. L. (2015). Introduction to renewable energy. CRC press.

Nespoli, A., Ogliari, E., Leva, S., Massi Pavan, A., Mellit, A., Lughi, V., & Dolara, A. (2019). Day-ahead photovoltaic forecasting: A comparison of the most effective techniques. Energies, 12(9), 1621.

Network, R. E. P. (2018). Renewable Energy Policy Network for the 21st Century (Ren21). http://www.ren21.net/gsr-2018/pages/imprint/imprint/

Nguyen, Q. K. (2005). Long term optimization of energy supply and demand in Vietnam with special reference to the potential of renewable energy Universität Oldenburg].

Notton, G., Voyant, C., Fouilloy, A., Duchaud, J. L., & Nivet, M. L. (2019). Some applications of ANN to solar radiation estimation and forecasting for energy applications. Applied Sciences, 9(1), 209.

Ogliari, E., Grimaccia, F., Leva, S., & Mussetta, M. (2013). Hybrid predictive models for accurate forecasting in PV systems. Energies, 6(4), 1918-1929.

Orasch, W. (2009). Regulatory framework for RES-E system integration in Europe-Description and analysis of different European practices. Appendix to Deliverable, 8. https://ec.europa.eu/energy/intelligent/projects/sites/iee-projects/files/projects/documents/greennet-incentives_greennet_incentives_grid_regulation_practices_en.pdf

Oudjana, S., Hellal, A., & Mahammed, I. H. (2013). Power forecasting of photovoltaic generation. International Journal of Electrical and Computer Engineering, 7(6), 627-631.

Oudjana, S. H., Hellal, A., & Mahamed, I. H. (2012). Short term photovoltaic power generation forecasting using neural network. 2012 11th International Conference on Environment and Electrical Engineering, Parida, B., Iniyan, S., & Goic, R. (2011). A review of solar photovoltaic technologies. Renewable and sustainable energy reviews, 15(3), 1625-1636.

Parkinson, G. (2015). Why solar costs will fall another 40% in just two years. RE New Economy, 20.

Pazikadin, A. R., Rifai, D., Ali, K., Malik, M. Z., Abdalla, A. N., & Faraj, M. A. (2020). Solar irradiance measurement instrumentation and power solar generation forecasting based on Artificial Neural Networks (ANN): A review of five years research trend. Science of The Total Environment, 715, 136848.

Perea-Moreno, M.-A., Hernandez-Escobedo, Q., & Perea-Moreno, A.-J. (2018). Renewable energy in urban areas: Worldwide research trends. Energies, 11(3), 577.

Perez-Astudillo, D., & Bachour, D. (2015). Variability of measured global horizontal irradiation throughout Qatar. Solar Energy, 119, 169-178.

Perez, R., Kivalov, S., Schlemmer, J., Hemker Jr, K., Renné, D., & Hoff, T. E. (2010). Validation of short and medium term operational solar radiation forecasts in the US. Solar Energy, 84(12), 2161-2172.

Pinson, P., & Kariniotakis, G. (2004). On-line assessment of prediction risk for wind power production forecasts. Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology, 7(2), 119-132.

Pinson, P., Madsen, H., Nielsen, H. A., Papaefthymiou, G., & Klöckl, B. (2009). From probabilistic forecasts to statistical scenarios of short-term wind power production. Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology, 12(1), 51-62.

Pioro, I., & Duffey, R. (2015). Nuclear power as a basis for future electricity generation. Journal of Nuclear Engineering and Radiation Science, 1(1).

Pitalúa-Díaz, N., Arellano-Valmaña, F., Ruz-Hernandez, J. A., Matsumoto, Y., Alazki, H., Herrera-López, E. J., . . . Velázquez-Contreras, E. F. (2019). An

ANFIS-Based Modeling Comparison Study for Photovoltaic Power at Different Geographical Places in Mexico. Energies, 12(14), 2662.

Poli, A. A., & Cirillo, M. C. J. A. E. P. A. G. T. (1993). On the use of the normalized mean square error in evaluating dispersion model performance. 27(15), 2427-2434.

Prasad, S., Venkatramanan, V., & Singh, A. (2021). Renewable energy for a low-carbon future: policy perspectives. In Sustainable Bioeconomy (pp. 267-284). Springer.

Prăvălie, R., & Bandoc, G. J. J. o. e. m. (2018). Nuclear energy: Between global electricity demand, worldwide decarbonisation imperativeness, and planetary environmental implications. 209, 81-92.

Price, G. D. (2014). Power Systems and Renewable Energy: Design, Operation, and Systems Analysis. Momentum Press.

Qian, Z., Pei, Y., Zareipour, H., & Chen, N. (2019). A review and discussion of decomposition-based hybrid models for wind energy forecasting applications. Applied energy, 235, 939-953.

Rahmani, R., Yusof, R., Seyedmahmoudian, M., & Mekhilef, S. (2013). Hybrid technique of ant colony and particle swarm optimization for short term wind energy forecasting. Journal of Wind Engineering and Industrial Aerodynamics, 123, 163-170.

Rajagukguk, R. A., Ramadhan, R. A., & Lee, H.-J. J. E. (2020). A review on deep learning models for forecasting time series data of solar irradiance and photovoltaic power. 13(24), 6623.

Raza, M. Q., & Khosravi, A. (2015). A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. Renewable and Sustainable Energy Reviews, 50, 1352-1372.

Raza, M. Q., Nadarajah, M., & Ekanayake, C. (2016). On recent advances in PV output power forecast. Solar Energy, 136, 125-144.

Razykov, T. M., Ferekides, C. S., Morel, D., Stefanakos, E., Ullal, H. S., & Upadhyaya, H. M. (2011). Solar photovoltaic electricity: Current status and future prospects. Solar Energy, 85(8), 1580-1608.

Reikard, G. (2009). Predicting solar radiation at high resolutions: A comparison of time series forecasts. Solar Energy, 83(3), 342-349.

Ren, Y., Suganthan, P., & Srikanth, N. (2015). Ensemble methods for wind and solar power forecasting—A state-of-the-art review. Renewable and Sustainable Energy Reviews, 50, 82-91.

Rodríguez-Gallegos, C. D., Bieri, M., Gandhi, O., Singh, J. P., Reindl, T., & Panda, S. (2018). Monofacial vs bifacial Si-based PV modules: Which one is more cost-effective? Solar Energy, 176, 412-438.

Rodríguez, F., Fleetwood, A., Galarza, A., & Fontán, L. (2018). Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control. Renewable energy, 126, 855-864.

Sahoo, D., Sood, N., Rani, U., Abraham, G., Dutt, V., & Dileep, A. (2020). Comparative analysis of multi-step time-series forecasting for network load dataset. 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT),

Şen, Z. (2004). Solar energy in progress and future research trends. Progress in energy and combustion science, 30(4), 367-416.

Setyawati, B. R. (2005). Multi-layer feed forward neural networks for foreign exchange time series forecasting West Virginia University]. Morgantown, West Virginia

Sideratos, G., & Hatziargyriou, N. D. (2007). An advanced statistical method for wind power forecasting. IEEE Transactions on power systems, 22(1), 258-265.

Sims, R. E., Rogner, H.-H., & Gregory, K. (2003). Carbon emission and mitigation cost comparisons between fossil fuel, nuclear and renewable energy resources for electricity generation. Energy policy, 31(13), 1315-1326.

Singh, G. K. (2013). Solar power generation by PV (photovoltaic) technology: A review. Energy, 53, 1-13.

Singh, S., & Yassine, A. (2018). Big data mining of energy time series for behavioral analytics and energy consumption forecasting. Energies, 11(2), 452.

Sobri, S., Koohi-Kamali, S., & Rahim, N. A. (2018). Solar photovoltaic generation forecasting methods: A review. Energy Conversion and Management, 156, 459-497.

solaredge. (2020). about-us. Retrieved 15th of April from https://www.solaredge.com/corporate/about-us

Solcast. (2020). Data and tools to build the solar powered future. Retrieved April 17 from https://solcast.com/?_ga=2.194580525.16390729.1587033345-878486021.1579729414

Suryakiran, M. N. S., Begum, W., Sudhakar, R., & Tiwari, S. K. J. A. i. S. G. T. (2020). Development of wind energy technologies and their impact on environment: A review. 51-62.

Sweeney, C., Bessa, R. J., Browell, J., Pinson, P. J. W. I. R. E., & Environment. (2020). The future of forecasting for renewable energy. 9(2), e365.

Talaat, M., Farahat, M., Mansour, N., & Hatata, A. J. E. (2020). Load forecasting based on grasshopper optimization and a multilayer feed-forward neural network using regressive approach. 196, 117087.

Tascikaraoglu, A., & Uzunoglu, M. (2014). A review of combined approaches for prediction of short-term wind speed and power. Renewable and Sustainable Energy Reviews, 34, 243-254.

Trinuruk, P., Sorapipatana, C., & Chenvidhya, D. (2009). Estimating operating cell temperature of BIPV modules in Thailand. Renewable energy, 34(11), 2515-2523.

Tvaronavičienė, M., Baublys, J., Raudeliūnienė, J., & Jatautaitė, D. (2020). Global energy consumption peculiarities and energy sources: Role of renewables. In Energy Transformation Towards Sustainability (pp. 1-49). Elsevier.

Vanderstar, G., Musilek, P., & Nassif, A. (2018). Solar forecasting using remote solar monitoring stations and artificial neural networks. 2018 IEEE Canadian conference on electrical & computer engineering (CCECE),

Vapnik, V. (2013). The nature of statistical learning theory. Springer science & business media.

Ventosa, M., Baıllo, A., Ramos, A., & Rivier, M. (2005). Electricity market modeling trends. Energy policy, 33(7), 897-913.

Victoria, M., Haegel, N., Peters, I. M., Sinton, R., Jäger-Waldau, A., del Cañizo, C., . . . Kaizuka, I. J. J. (2021). Solar photovoltaics is ready to power a sustainable future. 5(5), 1041-1056.

Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F., & Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: A review. Renewable Energy, 105, 569-582.

Wang, F., Xuan, Z., Zhen, Z., Li, K., Wang, T., & Shi, M. (2020). A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework. Energy Conversion and Management, 212, 112766.

Wang, F., Zhang, Z., Liu, C., Yu, Y., Pang, S., Duić, N., . . . Catalão, J. P. (2019). Generative adversarial networks and convolutional neural networks based weather classification model for day ahead short-term photovoltaic power forecasting. Energy conversion and management, 181, 443-462.

Wang, H., Lei, Z., Zhang, X., Zhou, B., & Peng, J. (2019). A review of deep learning for renewable energy forecasting. Energy Conversion and Management, 198, 111799.

Wang, H., Liu, Y., Zhou, B., Li, C., Cao, G., Voropai, N., & Barakhtenko, E. (2020). Taxonomy research of artificial intelligence for deterministic solar power forecasting. Energy Conversion and Management, 214, 112909.

Wang, J., Qin, S., Zhou, Q., & Jiang, H. (2015). Medium-term wind speeds forecasting utilizing hybrid models for three different sites in Xinjiang, China. Renewable Energy, 76, 91-101.

Wang, L., Tao, R., Hu, H., & Zeng, Y.-R. (2021). Effective wind power prediction using novel deep learning network: Stacked independently recurrent autoencoder. Renewable energy, 164, 642-655.

Wang, X., Adelmann, P., & Reindl, T. (2012). Use of LiFePO4 batteries in stand-alone solar system. Energy Procedia, 25, 135-140.

Wang, Y., Hu, Q., Meng, D., & Zhu, P. J. A. e. (2017). Deterministic and probabilistic wind power forecasting using a variational Bayesian-based adaptive robust multi-kernel regression model. 208, 1097-1112.

Wang, Y., Hu, Q., Srinivasan, D., & Wang, Z. J. I. T. o. S. E. (2018). Wind power curve modeling and wind power forecasting with inconsistent data. 10(1), 16-25.

Wolak, F. A. (2021). Wholesale electricity market design. In Handbook on Electricity Markets. Edward Elgar Publishing.

Wu, Y.-K., & Hong, J.-S. (2007). A literature review of wind forecasting technology in the world. 2007 IEEE Lausanne Power Tech,

Xiang, L., Li, J., Hu, A., & Zhang, Y. (2020). Deterministic and probabilistic multi-step forecasting for short-term wind speed based on secondary decomposition and a deep learning method. Energy Conversion and Management, 220, 113098.

Xiao, L., Wang, J., Dong, Y., & Wu, J. (2015). Combined forecasting models for wind energy forecasting: A case study in China. Renewable and Sustainable Energy Reviews, 44, 271-288.

Xie, X., Fu, Y., Jin, H., Zhao, Y., & Cao, W. (2019). A novel text mining approach for scholar information extraction from web content in Chinese. Future Generation Computer Systems.

Yamegueu, D., Azoumah, Y., Py, X., & Zongo, N. (2011). Experimental study of electricity generation by Solar PV/diesel hybrid systems without battery storage for off-grid areas. Renewable energy, 36(6), 1780-1787.

Yildiz, C., Acikgoz, H., Korkmaz, D., & Budak, U. (2021). An improved residual-based convolutional neural network for very short-term wind power forecasting. Energy Conversion and Management, 228, 113731.

Yousif, J. H., Kazem, H. A., & Boland, J. (2017). Predictive models for photovoltaic electricity production in hot weather conditions. Energies, 10(7), 971.

Zameer, A., Khan, A., & Javed, S. G. (2015). Machine Learning based short term wind power prediction using a hybrid learning model. Computers & Electrical Engineering, 45, 122-133.

Zendehboudi, A., Baseer, M., & Saidur, R. (2018). Application of support vector machine models for forecasting solar and wind energy resources: A review. Journal of cleaner production, 199, 272-285.

Zhang, J., Yan, J., Infield, D., Liu, Y., & Lien, F.-s. (2019). Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model. Applied energy, 241, 229-244.

Zhang, Y., & Wang, J. (2016). K-nearest neighbors and a kernel density estimator for GEFCom2014 probabilistic wind power forecasting. International Journal of forecasting, 32(3), 1074-1080.

Ziegler, B. E. (2009). Methods for bibliometric analysis of research: renewable energy case study Massachusetts Institute of Technology].

Zsiborács, H., Hegedűsné Baranyai, N., Csányi, S., Vincze, A., & Pintér, G. (2019). Economic analysis of grid-connected PV system regulations: A hungarian case study. Electronics, 8(2), 149.