SCHOOL OF BUSINESS ADMINISTRATION
DEPARTMENT OF ACCOUNTING AND FINANCE
MASTER OF SCIENCE IN ACCOUNTING AND FINANCE

Master’s Thesis

FORECASTING SOLAR POWER GENERATION IN ENERGY MARKETS

by

MALATARAS KONSTANTINOS
Supervising Professor: Alexandridis Antonis

Dissertation submitted in partial fulfillment of requirements for the degree of MSc in Accounting & Finance, at the University of Macedonia.

October 2022

Dedication

This thesis is dedicated to my parents, whose perseverance during challenging times has been a tremendous source of inspiration. Their continuous moral, spiritual, emotional and financial support has made this and every other achievement in my life feasible.

To my grandfather who always insisted that his grandchildren should improve their lives through education and has been a strong advocate of hard work and sacrifice. I hope this achievement fulfils, at least to some extent, the course he envisioned for me.

To my uncle who I always looked up to and admired and to whom I am very grateful for his contribution in the start of my career.

Acknowledgments

I would like to express my deepest appreciation to my supervisor Dr. Antonis Alexandridis, who even thought our conflicting schedules, found the time to provide me with valuable feedback and guidance. His expertise was a valuable factor that greatly contributed to the realization of this thesis.

Furthermore, I would like to thank my friend Andreas for his valuable insight, and without whom, the data collection phase of this thesis would be almost impossible.

Περίληψη

Ο στόχος της παρούσας διπλωματικής εργασίας είναι να προτείνει ένα απλό αλλά ισχυρό μοντέλο το οποίο θα μπορεί να χρησιμοποιηθεί για την πρόβλεψη της παραγόμενης ηλιακής ενέργειας στα πλαίσια των αγορών ενέργειας. Χρησιμοποιώντας μετεωρολογικά δεδομένα και ιστορικά δεδομένα ηλιακής ενέργειας από 1/1/2018 έως και 31/12/2021, ένα Νευρωνικό Δίκτυο εκπαιδεύτηκε για να δημιουργεί επαναλαμβανόμενες ημερήσιες προβλέψεις ενός βήματος προς το μέλλον σχετικά με την ηλιακή ενέργεια που θα παραχθεί από ένα φωτοβολταϊκό πάρκο. Για να αξιολογηθεί το μοντέλο, η ακρίβεια των προβλέψεων του Νευρωνικού Δικτύου συγκρίνεται με προβλέψεις που προκύπτουν από ένα μοντέλο τυχαίου περιπάτου, ένα μοντέλου κινητού μέσου όρου και ένα μοντέλο πολλαπλής γραμμικής παλινδρόμησης. Το Νευρωνικό Δίκτυο παρουσιάζει μεγαλύτερη ακρίβεια σε σχέση με το μοντέλο τυχαίου περιπάτου και το μοντέλο κινητού μέσου όρου και παρόμοια ακρίβεια σε σχέση με το μοντέλο γραμμικής παλινδρόμησης. Έπειτα, η ακρίβεια των μοντέλων αξιολογείται σε διαφορετικά πλαίσια αγορών ενέργειας. Τα μοντέλα αξιολογούνται ανάλογα με το εισόδημα που δημιουργούν για τους παραγωγούς ενέργειας βασιζόμενα στην ενέργεια η οποία δεσμεύεται στην Αγορά Επόμενης Μέρας και στα κόστη Εξισορρόπησης της Αγοράς Εξισορρόπησης. Παρατηρείται πως σε ένα πλαίσιο αγοράς με μία Τιμή Εξισορρόπησης, όπως η ελληνική αγορά ενέργειας, ένα λιγότερο ακριβές μοντέλο μπορεί να οδηγήσει σε μεγαλύτερα εισοδήματα για τους παραγωγούς, υπό συγκεκριμένες συνθήκες. Ωστόσο, όταν τα μοντέλα εξετάζονται σε ένα πλαίσιο αγοράς με δύο Τιμές Εξισορρόπησης, είναι εμφανές πως τα πιο ακριβή προβλεπτικά μοντέλα (στην προκειμένη περίπτωση το Νευρωνικό Δίκτυο και η γραμμική παλινδρόμηση) εξασφαλίζουν υψηλότερα κέρδη (ή χαμηλότερα κόστη εξισορρόπησης) για τους παραγωγούς ηλιακής ενέργειας.

Λέξεις κλειδιά: Πρόβλεψη Ηλιακής Ενέργειας, Φωτοβολταϊκά, Αγορές Ενέργειας, Τιμή Εξισορρόπησης, Ανάλυση Σφάλματος.

Abstract

The aim of this thesis is to propose a simple but robust forecasting model that can be used for solar power forecasting in the context of energy markets. Utilizing meteorological data and historical values of PV (photovoltaic) power spanning from 1/1/2018 to 31/12/2021, an ANN model was trained to generate recursive one step ahead daily predictions, on an expanding window, of the PV power that a solar power plan will generate. To evaluate the model, the forecasting accuracy of the ANN is then benchmarked against the Persistent Model, a three-day Moving Average and a Multiple Linear Regression model. The ANN is found to outperform the Persistent Model and the Moving Average and to perform similarly to the Linear Regression, as far as statistical evaluations are concerned. After that, the accuracy of these models is evaluated in different frameworks of energy markets. The models are evaluated based on the income they generate for the producers depending on the energy pledged on the Day Ahead Market and the Imbalance fees of the Imbalance Market. It is found that in a Single Imbalance Price energy market, like the Greek one, a less accurate forecasting model can lead to higher profits for the energy producers under certain circumstances. On the other hand, when the models are evaluated in a Dual Imbalance Price energy market framework, it is found that the most accurate forecasting models (in this case the ANN model and the Regression model) generate higher profits (or lower imbalance fees) for the energy producers.

Keywords: Solar Power Forecasting, PV, Electricity Markets, Imbalance Price, Error Analysis.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Global warming and energy crises in recent decades have led to the rapid development of clean renewable energy sources (RES). It is now evident that if the usage of fossil fuels is maintained at the current rate, these sources of energy will be quickly depleted, and their continued usage will lead to even greater environmental deterioration. Moreover, fossil fuel sources are located in specific parts of the world with few countries controlling a significant portion of their production. Even though energy markets (mostly oil) were relatively stable for the so called "golden era" (Smith, 2009), the years from 1874 to 1974, recent geopolitical tension and conflicts have led to unstable prices with grave consequences for the countries heavily dependent on importing fossil fuels. Even more recently, with the Russian invasion of Ukraine, the issue of energy security for many fossil fuel importing countries has been raised (Prisecaru, 2022). All these factors have led to an increased interest in the research and development of renewable energy sources, with solar power being one of the most promising, mainly due to its potential and availability.

Solar energy is the radiant energy of the sun. Sun generates enormous amounts of electromagnetic energy from the fusion of hydrogen gas that takes place in its surface. Solar power plants utilize this energy to produce electricity. Currently there are two types of solar power plants: solar thermal systems and photovoltaic plants. Solar thermal technology utilizes sunlight for heating water which produces steam. Then, the steam is used to rotate turbines in power plants. This technology has been implemented mostly in Spain and USA (Gonzalo et al., 2019). The great disadvantage of this technology is that it requires large facilities in order to be effective. On the other hand, photovoltaics (PV) utilize sunlight to activate free electrons in semiconductors embedded on panels in order to create electric charge. These panels can be installed in open spaces of all sizes, they can even be integrated into buildings, cars or create huge solar power plants. Over the last decades PVs have been widely adopted and their demand has increased substantially. One of the main reasons that has led to the significant growth of PV power in recent years, is the constant technological improvements. The state-of-the-art PV panels are cheaper and more efficient than the older models, while at the same time being more modular and versatile in their installation, requiring less maintenance, and having a longer service life (Raza et al, 2016).

Due to the constantly increasing popularity of PVs, PV generated power is reaching higher levels of penetration in the power grid. However, the generation of PV power depends on unpredictable meteorological factors such as solar irradiance, temperature, humidity etc. Furthermore, the output of a PV system constantly and dynamically changes due to the variability of climate factors. Therefore, the accurate forecast of the energy production of a PV system is inherently challenging. The unpredicted flow of PV power (either a surplus or a deficit) into the power grid can adversely affect the stability and the scheduling of the power grid and cause significant economic losses (Ahmed & Khalid, 2019). On the other hand, an accurate forecast of the PV power generated will reduce the uncertainty and volatility that is introduced to the system, improve system stability, and power quality and possibly lead to greater adoption of PV systems. The above have led to an increased interest from researchers, in the past decade, in addressing this challenging task.

Two other factors have recently intensified the need for accurate PV energy generation forecasting. One of them is the political commitments that have been made by different countries, according to global agreements, to address climate change. The most important of these agreements is the Paris Agreement, signed in 2015, where the signing parties pledged to a slower rate of temperature increase and a carbon neutral world by the second half of the century. One very efficient way to eliminate the carbon emissions of fossils fuels, leading to the slower increase of global temperature and carbon neutrality, would be to replace them with renewable energy sources, such as solar. However, for solar power to achieve greater penetration and adoption, it has to become more reliable and predictable, leading to an increased need for enhanced forecasting.

The other factor that has led to a growing demand for accurate PV power forecasting is the liberalization of electricity markets. In modern electricity markets, energy producers are pledging their estimated energy production for the next day, in the Day-Ahead market, effectively locking in a price for the specified volume. If the prediction is significantly flawed the system operator has to utilize another energy source to immediately compensate for the deviation from the declared energy, resulting in increased cost for the system. In these cases, a penalty fee is often required from the producer to cover the increased costs of the system. This is a significant issue for solar energy producers since their estimated output is difficult to determine. This is one significant factor that has led to substantive research in optimizing PV power forecasting, with many researchers supporting that increases in forecasting accuracy can have significant economic effects on the revenue of solar power producers (Zhang et al., 2015; De Giorgi et al., 2015).

The scope of this thesis is twofold. Firstly, to propose a simple but accurate forecasting model which will attempt to predict the next day's PV energy so that a PV power producer can utilize this prediction on their participation in energy markets. Secondly, to evaluate this predicting model in the context of energy markets and especially regarding the income of the PV producer that will be generated by the use of this model and the imbalance fees that the inaccuracy of this model will impose on the PV power producer's income.

The rest of this thesis is organized as follows. In Chapter 2 a review of the literature regarding the way the PV power forecasting has been developed is presented. This section deals with the different factors that affect the forecasts as well as with the different models that have been proposed. Afterwards, Chapter 3 presents and analyzes the data utilized in this research and the different data sources from which the data were derived. In Chapter 4, the methodology followed throughout this research is presented outlining the ANN model proposed, the benchmarking process and the evaluation of the forecasting errors in the context of energy markets. In Chapter 5 the results of the forecasting effort are presented and discussed. Finally, in Chapter 6 some general remarks are presented along with the limitations that hindered this research.

Chapter 2

Literature Review

As it is presented in the previous chapter, there is a growing interest in accurately forecasting PV generated power, which has led to rich literature. In this chapter, the most commonly used methods for PV power forecasting will be presented. The first section of this chapter follows the steps of the researchers creating forecasting PV power models. Firstly, the different input variables that have been used in literature are discussed. Secondly, the different forecasting horizons that researchers utilize in their forecasts are presented. Lastly, the ways they evaluate their models are introduced. In the second section of the chapter the different categories of models that have been used in literature are presented and the individual models are discussed.

2.1. Input selection for the PV power forecasting models

The data used as input to a forecasting model has a direct influence on its prediction accuracy. Improper input selection can lead to forecast errors, delays, costs, and computational complexity. Under these circumstances, low accuracy rates can appear even for highly capable forecasting models (Kilkenny & Robinson, 2018).

As it is previously presented, solar energy is generated from the sun in the form of solar irradiance. The PV panels are comprised of semiconductors that convert solar radiance to electricity. Thus, the power that will be generated from the PV system is greatly dependent on solar radiance. Moreover, other meteorological observations, such as atmospheric temperature, solar panel temperature, humidity, wind speed and direction have been used as inputs for forecasting the PV power output.

The combination of meteorological parameters that are used depends on the geographical location and the local climate, since the same meteorological observation can have different impact on different geographical locations. As a result, it can be safely assumed that the correlation of the different weather observations and the PV power output will vary across different locations. However, the correlation of the inputs and output is crucial for the performance of a forecasting model. In this case, the correlation of each input is calculated, and only the strongly (positively or negatively) correlated input variables are used as input to the forecasting models, while the weakly correlated variables should be disregarded.

Literature (De Giorgi et al., 2014; Chen et al., 2011) clearly supports the existence of a remarkably strong correlation between PV power output and solar radiance, compared to other meteorological variables. As far as the temperature is concerned, research findings are mixed. Some researchers (Chen et al., 2011) suggest that a strong correlation exists between ambient temperature and PV power output, while others (Das et al., 2018) indicate a weak correlation between them.

Other meteorological inputs such as wind speed, have also been studied regarding their correlation the PV power output. The linkage between wind speed and PV output may not be so obvious. However, wind speed is important since it can mitigate the heat and reduce the PV cells temperatures. It is important to note that the efficiency of PV panels depends on the temperature of the panel (the correlation between them is negative) which rises during operation, due to the absorption of radiation. Thus, both the wind speed and the cell temperatures should be reviewed. Researchers (Raza et al., 2016; Ahmed et al., 2020) have studied the correlation between wind speed and PV power output and found it to be positive but weak. Regarding the temperature of PV panels Schwingshackl et al., (2013) and Ting-Chung & Hsiao-Tse (2011) have found it to be strongly correlated with PV power generation.

During the design of a forecasting model of PV power generation, one could be tempted to use a large number of input vectors, in order to increase its accuracy and performance. However, at the same time, the computational cost and the model complexity will also be increased. Therefore, the utilization of correlation during the design of a forecasting model, in order to achieve an optimal number of inputs, is crucial, as it has been pointed out in literature (Ahmed et al., 2020).

2.2. Forecasting horizon

The forecasting horizon can be defined as the time period between the actual time of prediction and the time that the prediction will take effect. Researchers have created three categories of the forecasting horizon for PV power forecasting: short-term, medium-term, long-term (Das et al., 2018).

i) Short term horizon

This is the most popular and most researched forecasting horizon, since it is quite useful in different environments, such as the electricity markets (economic load dispatch and power system operation-balancing) and renewable energy integrated power systems management. Typically, the short-term horizon is between 30 and 360 minutes (Ren et al., 2015), however this range has been extended by some, to include ranges of one to several hours, one day or even up to one week as short-term forecast horizon (Das et al., 2018).

ii) Medium term horizon

Medium term horizon usually covers the span between 6 and 24 hours (Ren et al., 2015). However, a few researchers consider one day, one week and up to a month as part of this category. This horizon is used in the planning and maintenance scheduling of power systems.

iii) Long term horizon

The mark that distinguishes the long-term horizon from the medium-term one is the one day (Ren et al., 2015). Forecasts that predict outcomes more than 24 hours in advance are categorized as long-term. Similar to the cases presented above, different definitions of this horizon have been given, with some researchers categorizing periods from one month to one year as long-term (Das et al., 2018). These horizons are utilized in the planning of long-term power generation, transmission, and distribution, as well as for researching seasonal trends (Han et al., 2019). However, these forecasting models have reduced accuracy, since weather fluctuations, in such long time periods into the future, cannot be accurately predicted.

There are a great number of studies that research the impact of the forecasting horizon on the models' accuracy. Lipperheide et al. (2015) studied the performance of a PV power forecasting model by keeping the other parameters constant and altering the forecasting horizon. The RMSE of the model fell to 3.2 % at 20 seconds forecasting, from 15.5% at 180 seconds. Lonji et al. (2013) also concluded that the accuracy of a PV forecasting model varied with the changes in the forecasting horizon. Therefore, the forecasting horizon must be taken into consideration before designing the appropriate forecasting model.

2.3. Forecasting model performance evaluation

In general, performance estimation is essential for evaluating a model's forecasting ability. Specifically, since a great amount of power produced from PV power systems has been introduced to the grid, the balance and stability of the grid system is dependent on the ability of the PV producers to correctly forecast their output and commit it to the grid. The grid will face imbalances if the provided energy is either inadequate or more than the predicted amount. Consequently, the accuracy of PV power forecasting models is essential for ensuring grid stability and balance as well as the further adoption of PV power generation.

The evaluation of the models' accuracy presented in literature is based on various evaluation metrics. These standardized performance measures include Mean Square Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) etc. and are presented in the following paragraphs.

Mean square error (MSE)

Mean square error is a measurement of the average of the squared errors of a forecasting procedure. Due to the errors being squared MSE is always greater or equal to zero, with the closer being to zero the better the accuracy of the predictive model. The MSE is calculated by:

$$MSE = \sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n} \tag{2.1}$$

where $\hat{y}_i$ the predicted solar energy on day *i* and y_i the actual energy generated.

Root Mean Square Error (RMSE)

RMSE calculates the mean value of the error by considering the square root of the difference between the forecasted variable and the actual observation. RMSE is nothing more than the square root of MSE that was presented in the above paragraph. Same as MSE, RMSE is always non-negative, and the lower its value the better the accuracy of the model. However, RMSE can be used to evaluate models utilizing the same dataset. It cannot be used to compare models that use different datasets as inputs because RMSE is scale dependent. RMSE is calculated by:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \tag{2.2}$$

Normalized Root Mean Square Error

In order to address the barrier of RMSE mentioned above, different methods of normalization are proposed in literature. The most popular are:

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}} \tag{2.3}$$

where y_{max} and y_{min} represent the range of the dataset.

$$NRMSE = \frac{RMSE}{\bar{y}} \tag{2.4}$$

where $\bar{y}$ is the mean of the variable *y*.

Mean Absolute Percentage Error (MAPE)

MAPE is used to evaluate the predicting models' accuracy by calculating the deviation of the predicted value from the actual value as a percentage. MAPE can be calculated from the formula below:

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right| \tag{2.5}$$

Even though these metrics have been widely used in literature some researchers have suggested that the exclusive use of these measures for the performance evaluation of models is not sufficient (Ahmed & Khalid, 2019). They propose that each model's accuracy should be evaluated based on the domain for which the forecast has been made. These domains include economic dispatch, optimal energy storage, profit maximization for energy market stakeholders and optimal reserve size. Since in these domains the forecasting error can have significant economic effects on stakeholders a simple statistical evaluation of a model is not enough, and other processes have to be used. In these cases, Monte Carlo simulations are often utilized in order to evaluate the models' performance (Doga et al., 2016; Haessig et al., 2015; Ortega-Vasquez & Kirschen, 2008). In other cases, the forecasting errors are incorporated in new models that lead to improved economic results for stakeholders (Kaur et al., 2016; Howlader et al., 2015).

2.4. Classification of Models

Several modeling approaches have been utilized for PV power generation forecasts. These methods can be grouped and categorized into the five major categories that are presented below.

2.4.1 Persistence model

The persistence model is an elementary forecasting model, mainly used as a benchmark for other forecasting models. In this model, the forecasted output of the next day is assumed to remain the same as the day before. The forecast output for the next day can be described by the equation below:

$$\hat{P}_t = P_{t-1} \tag{2.6}$$

where $\hat{P}_t$ is the forecasted generated power, and P_{t-1} is the power generated on the previous day. This model is usually used for short-term forecasting, especially in the one-hour time horizon. The accuracy of this model depends greatly on the stability of the weather conditions. If the weather is stable enough, the output of the previous day is a reliable indicator of the power that will be generated on the next day. However, as the time horizon of the forecast increases, the unpredictability of the weather conditions increases as well, leading to a significant decrease of this model's accuracy (Perez et al., 2018).

2.4.2 Physical models

Physical models involve meteorological observations of the lower atmosphere, creating a numerical weather prediction (NWP) model consisting of "*discretized conservation equations of* *mass, momentum, energy and other fundamental principles of physics"* (Zhao et al., 2016). These models are usually used for long-term forecasting, in which case they usually outperform pure statistical forecasts (Cassola & Burlando, 2012). These models are further distributed into two subcategories based on their scale: mesoscale models and global models. Mesoscale models can process atmospheric observations for limited geographical areas such as regions, continents, or countries while global models can provide forecasts on a global scale. For the creation of such models, weather databases are needed. There are currently about 15 weather information providers that are active in data collection, which are usually managed by state organizations (US NOAA, ESMWF, etc.) (Ahmed at al., 2020). NWP models can accurately forecast weather conditions for more than 15 days ahead (Lorenz et al., 2012) using equations to capture the physical state and dynamic nature of the atmosphere. However, as is the case with the persistence model, NWP models produce better results when the weather conditions remain relatively stable (Soman et al., 2010). When the weather conditions change abruptly, less accurate forecasts are more likely to occur.

2.4.3 Statistical models

Statistical forecasting models utilize historical and real time generated data. They require fewer inputs than Deep Learning models and show better accuracy in short-term forecasting compared to NWP models (Ahmed et al., 2020). These models utilize mathematical equations to interpret patterns and extract correlations from the historical data provided. Usually, the algorithms consist of curve fitting, moving averages and autoregressive models (Firat et al., 2010). The way these models minimize error is by estimating the difference between the actual observed past value and the predicted value of the forecasting model and trying to reduce it. Statistical models proposed in literature are exponential smoothing, regression method, autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA).

Exponential smoothing

In the exponential smoothing method or exponentially weighted moving average (EWMA) weights are allocated to the historical data, which then are exponentially reduced from the most recent observation to the oldest. In this method greater importance is given to the most recent data compared to the older ones. This method was first proposed by Brown (1957) and was later modified by Holt in 1957 and Winter in 1960. Therefore, it is now called the Holt-Winter's method

(Tratar & Strmčnik, 2016). The most used equation to describe this model is the one presented below:

$$Y_{t+1} = aY_t + (1-a)\hat{Y}_t = \hat{Y}_t + a(Y_t - \hat{Y}_t) \quad (2.7)$$

where Y_t is the current observation, $\hat{Y}_t$ is the predicted value and a is the smoothing constant which remains between 0 and 1. Therefore, the EWMA indicates that the forecasted output equals to the sum of the last forecasted value and the error adjusting factor. Similar to the simple moving average method, the EWMA method is not appropriate for time-series with trends, since if a trend exists the EWMA forecast would lag behind the trend. Gumus and Kilic (2018) used an EWMA process to accurately predict the solar radiance and sunlight duration of the coming years for a specific region in Turkey.

Regression method

The regression method is a statistical model used to estimate the relationship between explanatory and dependent variables. In this model the dependent variable is forecasted by a function that is calculated from the known explanatory variables. In the case of PV power forecasting the forecasted power is considered the dependent variable while the meteorological variables are considered the explanatory variables. Verna et al., (2016) forecasted the PV power generation, including three different regression models: linear, logarithmic, and polynomial regressions. The study found linear relationships between temperature, cloud coverage, elevation angle and azimuthal angle and the PV power output while for the wind speed and humidity no linear relationship was found. Moreover, the only variable which did not seem to have a logarithmic relationship with the generated power was cloud coverage. As presented in the study above, a complex non-linear mathematical model and a great number of explanatory variables are required to design a regression-based forecasting model, which constitutes the weakness of this method.

Autoregressive moving average (ARMA)

ARMA is a statistical model frequently used in forecasting. The model is a combination of the AR (autoregression) and MA (moving average) models, and it has been examined by many researchers (David et al., 2016; Huang et al., 2012; Mora-Lopez & Sidrach-de Cardona, 1998) in solar power forecasting and it has consistently presented a high accuracy. The mathematical expression that is used to describe the model is as follows:

$$X(t) = \sum_{i=1}^{p} \alpha_i X(t-i) + \sum_{j=1}^{q} \beta_j e(t-j) \tag{2.8}$$

where the predicted power generation is depicted as the *X(t)* function which is a combination of the AR and MA models. Therefore, *p* and *q* indicate the order, α_i and β_j are the coefficients of the AR and MA models respectively and *e(t)* is the randomly generated white noise which is not corelated with the model's predictions. ARMA models are quite flexible and can be used to describe several different time series by adjusting the orders (*p* and *q*). The main reason ARMA models are so frequently used is their ability to extract the statistical properties of the data. However, this approach requires the time-series data to be static, thus creating a significant disadvantage. Huang et al. (2012) forecasted future solar power generation in California, based on solar radiance data, utilizing the ARMA model, which significantly outperformed the persistence model.

Autoregressive integrated moving average (ARIMA)

The ARIMA model is also known as Box-Jenkins model and was developed by George Box and Gwilym Jenkins in 1976 (Box & Jenkins, 1976). The ARIMA model is an extension of ARMA, and it is a popular and accurate model for forecasting in a short-term horizon. This model comes as an answer to the disadvantage of ARMA that is mentioned above. The ARIMA has the ability to remove any non-stationarity from the data. Its components are the same as ARMA with the addition of an integrated part. The general form of the ARIMA model is as follows:

$$\begin{gathered} \Phi_p(B)\Delta^d X_t = \Theta_q(B) a_t \\ \Phi_p(B) = 1 - \varphi_1 B - \varphi_2 B^2 \dots \varphi_p B^p \\ \Theta_q = 1 - \theta_1 B - \theta_2 B^2 \dots \theta_q B^q \end{gathered} \tag{2.9}$$

where *B* is the backward shift operator, $\Delta = 1\text{-}B$ and $BX_t = X_{y-1}$ is the backward difference Φ_p and Θ_q are polynomial numbers of order *p* and *q*, respectively. Consequently, the ARIMA *(p,q,d)* model is a summarization of an autoregressive part (*p*), an integrating part (*d*) and a moving average part (*q*). Das (2021) used an ARIMA model to forecast the PV power output in a short-term horizon, comparing it with an analytical model, and found the ARIMA to be more accurate across all the different time horizons studied in the research (15 min, 30 min, 45 min and 75 min). Similarly, Cadenas et al. (2016) compared the ARIMA model to a multivariate neutral network for forecasting one step ahead wind speed, finding the accuracies of the two models to be remarkably close. In another approach, Haiges et al. (2017) used different ARIMA models to forecast the electricity generation capacity in Malaysia, with the ARIMA (0,2) proving to be the most accurate.

2.4.4 Machine learning models

These models are utilizing the advances of machine learning, which relies on the ability of artificial intelligence (AI) to extract patterns and correlations from historical data and improve its predictive abilities by a repetition of training runs. In recent years, energy research has been flooded with great volumes of data due to the extensive use of smart sensors both in energy production and consumption (Torabi et al., 2018). This increased volume of data led to an increase in machine learning models' popularity, since they are the ones which can effectively manage big data and easily extract dependencies and correlations from these data (Mosavi et al., 2019). Machine learning approaches that are frequently used in PV power generation forecasting are: Artificial neural networks (ANNs), Extreme learning machine (ELA) and Support vector machine (SVM).

Artificial neural networks

ANNs have increased dramatically in popularity for forecasting purposes since they were first proposed. These methods have been used extensively in literature for forecasting PV power output with a very high level of accuracy (Gutierrez Corea, et al., 2016; Leva et al., 2015; Ding et al., 2011; Pedro & Coimbra, 2012; Paoli et al., 2011). ANNs have been so popular because of the non-linear complex relations that describe the meteorological data. ANNs are more suitable, compared to statistical methods, in processing non-linear and complex data.

ANNs' structure is based on the biological neurons of the human brain. In Figure 1 a simple ANN architecture is presented. ANNs consist of several interconnected cells (neurons), which can be divided into the three main components of an ANN: input, hidden and output layers. The input layers consist of neurons that receive the input information. The hidden layer, which can be composed of a single or several layers, processes the information from the input layer. While the output layer receives the processed information and provides the output. Each of these layers is comprised of several neurons connected through certain weights to the other nodes in the next layer. Each neuron cell consists of two parts, as can be seen in Figure 2. The first part is the *combination function* which sums the weighted inputs. The second part is the *activation function* which transforms the output of the combination function into the output of the node (or the network). The most used activation functions are presented in Table 1.

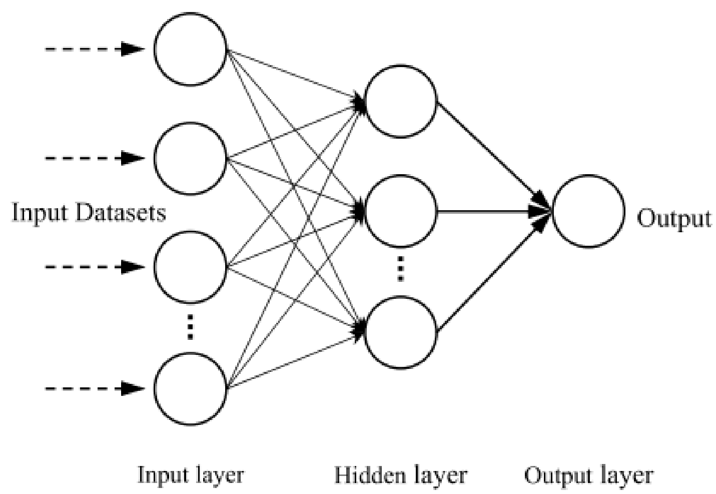


Figure 1. Basic ANN structure (Sobri et al., 2018).

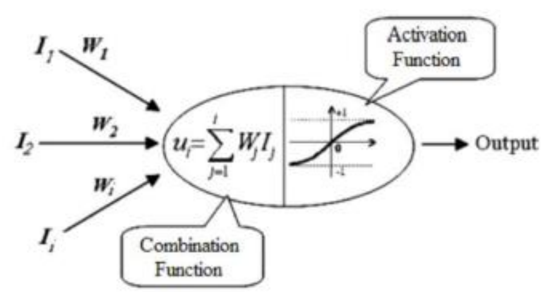


Figure 2. Mathematical model of an ANN neuron (Aminzadeh & De Groot, 2006).

In the literature the following models of neural networks have been used for PV output forecasting:

- Recurrent neural networks (Qing & Niu, 2018; De Giorgi et al., 2014; Chupong & Plangklang, 2011)
- Back propagation neural networks (B.M. Shah et al., 2015; Liu et al. 2015; Notton et al, 2013)
- Radial basis function neural networks (Lu & Chang, 2018; Mori & Takahashi, 2012)
- Self-organizing neural networks (Chen et al., 2011; Yang et al., 2014)

Table 1. Commonly used activation functions of ANNs (Raza et al., 2016).

Function	**Formula**
Sigmoid	$f(u) = \frac{1}{1+e^{-u}}$
Hyperbolic tangent sigmoid (tanh-sig)	$f(u) = tanh(u) = \frac{e^{u}-e^{-u}}{e^{u}+e^{-u}}$
Gaussian radial basis	$f(u) = exp\left(\frac{-\lvert u-m\rvert^{2}}{2\sigma^{2}}\right)$
Linear	$f(u) = u$
Unipolar step function	$f(u) = \begin{cases} 0, if\ u < 0 \\ 1, if\ u > 0 \end{cases}$
Bipolar step function	$f(u) = \begin{cases} -1, if\ u < 0 \\ 0, if\ u = 0 \\ 1, if\ u > 0 \end{cases}$
Unipolar linear function	$f(u) = \begin{cases} 1, if\ u < 0 \\ y, if\ u < 1 \\ 1, if\ u > 1 \end{cases}$
Bipolar linear function	$f(u) = \begin{cases} -1, if\ u < -1 \\ y, if\ \lvert u\rvert < 1 \\ 1, if\ u > 1 \end{cases}$

Extreme learning machine (ELM)

The ELM model is an advanced approach for a single layer feed forward network proposed by Huang et al. in 2006. In the ELM model the weights of the single hidden layer are randomly selected and never updated, while the output weights are extracted analytically through a least-squares algorithm. The main advantage of this method is that, since the hidden layer weights will not be updated, the output layer weights are calculated in a single step, thus reducing the computational time needed dramatically. Therefore, the ELM can learn faster compared to standard ANNs and it also demonstrates an easier implementation since no decision about the architecture of the network is to be made. Consequently, it is preferred by researchers as a simple and effective algorithm with smaller training input requirements, and thus it has seen extensive use in PV power forecasting (Al-Dahidi, et al., 2018; Li et al., 2018; Behera & Nayak, 2019; Hossain, et al., 2017).

Support vector machine (SVM)

SVM is a supervised learning algorithm developed by Vapnik in 1995 as a classification algorithm. However, SVM has recently been used for regression problems too. The algorithm is based on statistical learning theory for structural risk minimization. The application of SVM in time series regression is called support vector regression (SVR). The SVR algorithm is a non-linear regression algorithm, where the input data is mapped into a high-dimensional space by a nonlinear mapping function. After that, a linear regression is performed in that space. The main advantage of SVM is that it produces one optimal solution, as opposed to ANNs which can converge to local minima and produce multiple solutions. In literature different approaches to SVM and SVR have been proposed (Alfadda et al., 2017; Yang et al., 2016; Shi et al., 2012; Wolff et al., 2016)

2.4.5 Hybrid models

Each single model approach to the PV power forecasting comes with the limitations of the standalone technique used. To address these limitations a hybrid model, the combination of two or more techniques, is proposed. These models can achieve greater accuracy by taking advantage of the strengths of each single approach and combining them. Hybrid models have performed better than the individual techniques in PV power forecasting. Specifically, Yona et al. (2013) combined a fuzzy interference model with a RNN for forecasting PV power. In this approach the fuzzy interreference model was used to smoothen the meteorological that were used by the RNN to forecast the power generation. In another approach, Li et al. (2020) proposed a combination of wavelet packet decomposition (WPD) and long short memory network (LSTM) for short-term PV power forecasting. WPD decomposes the PV output time series data and then LSTM is used to predict the PV output using each subclass of the decomposed data and meteorological variables as inputs. This hybrid model achieved higher accuracy when compared to other single RNN methods. It should be noted that in literature, the combination of wavelet transform with neural networks is quite common (Zhang et al., 2020; Heydari et l., 2019; Zhu et al., 2015). In these cases, wavelet transformation is used for filtering and denoising the input data. The data are then passed as inputs to the neural network, which forecasts the PV power output with great accuracy.

Even though there are advantages, hybrid models come with some disadvantages, too. They present high computational complexity, and their performance is greatly dependent on the performance of each individual model (Ramsami & Oree, 2015). Their accuracy is affected by the selection of each underlaying strategy (if one component performs poorly the whole model will too).

Chapter 3

Data

The data utilized for the creation of the predictive model are derived from three different sources. For the needs of this thesis, the dataset created consists of 1461 observations of 17 variables in a daily frequency, spanning the period between the 1st of January of 2018 and the 31st of December of 2021. This dataset consists of meteorological observations and one variable describing the generated power of a PV plant used as the dependent variable. The meteorological variables include 7 weather variables measured at a surface level and 9 atmospheric variables provided by satellite.

The meteorological data were collected from two different meteorological agencies:

- The Hellenic National Meteorological Service (HNMS) kindly provided the author with historical meteorological observations recorded from a meteorological station located in Tatoi, Attica, Greece at longitude 23.78, latitude 38.10 and at an altitude of 225 m. The list of variables provided by HNMS is presented on Table 2.
- Copernicus Atmospheric Monitoring Service, a European atmospheric satellite surveillance system, provided historical, highly accurate, atmospheric observations regarding different measures of solar radiation. The collected atmospheric data are presented on Table 3.

Table 2. Meteorological variables provided by HNMS.

Variable	**Scale**	**Description**
Max Temperature	℃	The maximum temperature recorded in each day.
Min Temperature	℃	The minimum temperature recorded in each day.
Average Temperature	℃	The average temperature of each day as a weighted average of the temperatures recorded each hour of the day.
Cloud Coverage	8^{ths} of covered sky	Average daily cloud coverage.

Variable	Scale	Description
Low Cloud Coverage	8^{ths} of covered sky	Average daily cloud coverage in the lower atmosphere.
Wind Speed	Knots	Average daily wind speed, derived from the weighted average of the hourly wind speed.
Rainfall	mms	Average daily rainfall, derived from the weighted average of the hourly rainfall.

Table 3. Meteorological provided by Copernicus Atmospheric Monitoring Service.

Variable	Scale	Description
TOA	Wh/m^2	Irradiation on horizontal plane at the top of the atmosphere.
Clear sky GHI	Wh/m^2	Clear sky global irradiation on horizontal plane at ground level.
Clear sky BHI	Wh/m^2	Clear sky beam irradiation on horizontal plane at ground level.
Clear sky DHI	Wh/m^2	Clear sky defuse irradiation on horizontal plane at ground level.
Clear sky BNI	Wh/m^2	Clear sky beam irradiation on mobile plane following the sun at normal incidence.
GHI	Wh/m^2	Global irradiation on horizontal plane at ground level.
BHI	Wh/m^2	Beam irradiation on horizontal plane at ground level.
DHI	Wh/m^2	Diffuse irradiation on horizontal plane at ground level.
BNI	Wh/m^2	Beam irradiation on mobile plane following the sun at normal incidence.

By utilizing weather variables measured both at a surface and an atmospheric level, the model is expected to perform better. Even though atmospheric data regarding solar irradiance are sufficient for training forecasting models which perform with satisfactory accuracy (Perez et al., 2021), the combination of these data with weather data provided by surface weather stations has been proven to lead to forecasting models with enhanced accuracy (Agoua et al., 2021). This is mainly because variables measured at surface level (such as temperature, wind speed etc.) can impact the PV power generation, as already mentioned in the previous chapter.

For the collection of PV generated power data, usually, the installation of special equipment – energy analyzers- is required in the PV parks' substations. Energy analyzers are electric measuring devices that record and collect information regarding electric parameters such as current, voltage, frequency, power and quality of power (Yavor, 2009). However, due to not having access to a PV plant or to an electric analyzer, the data were collected using an alternative. The required data were aggregated from the website *www.pvouput.com*, where PV plant operators / owners upload data from their respective PV plants regarding power generated, efficiency etc. The only variable that will be used in the scope of this thesis will be the daily power generated. The data collected refers to a PV plant with a 10-kW capacity located in the area of Oropos in East Attica, Greece. The said plant consists of 40 PV panels, each one with a power of 250 Watts, installed with a 35^{o} tilt and a South – Southeast orientation. Table 4 depicts the descriptive statistics of the meteorological and energy variables.

Table 4. Descriptive statistics.

Variables	Mean	Median	Minimum	Maximum	Std. Dev.	C.V.[1]	Skewness	Kurtosis
Max Temp	22.15	21.60	-0.30	42.60	7.90	0.36	-0.01	2.08
Min Temp	12.07	11.60	-8.90	28.40	6.86	0.57	0.07	2.18
Avg Temp	17.66	17.10	-2.50	37.45	7.66	0.43	0.07	1.94
Cloud Coverage	2.88	2.60	0.00	8.00	2.14	0.74	0.37	2.06
Low Cloud	1.80	1.38	0.00	7.38	1.61	0.89	0.87	3.00
Windspeed	5.43	4.25	0.00	25.50	4.12	0.76	1.34	4.97
Rainfall	1.37	0.00	0.00	84.70	5.47	3.99	6.48	62.94
TOA	8069.1	8314.7	4091.77	11580.96	2654.82	0.33	-0.14	1.52
Clear-sky GHI	5770.70	5954.49	2793.36	8989.88	1996.20	0.35	-0.11	1.53
Clear-sky BHI	4379.50	4381.08	1820.87	7847.02	1644.70	0.38	0.10	1.67

[1] Coefficient of Variance

Variables	Mean	Median	Minimum	Maximum	Std. Dev.	C.V.[1]	Skewness	Kurtosis
Clear-sky DHI	1417.35	1324.85	366.26	3811.10	620.31	0.44	0.99	3.92
Clear-sky BNI	7564.83	7511.42	4376.68	11540.08	1569.88	0.21	0.05	2.12
GHI	4827.38	4816.58	979.90	8781.33	2247.85	0.47	0.01	1.67
BHI	3264.70	3098.40	16.88	7758.63	2144.82	0.66	0.17	1.80
DHI	1569.02	1368.93	594.36	3778.25	695.76	0.44	0.88	2.98
BNI	5363.83	5725.35	71.54	11215.38	2968.18	0.55	-0.20	1.87
PV power	48.75	52.32	3.66	87.27	24.99	0.51	-0.29	1.72

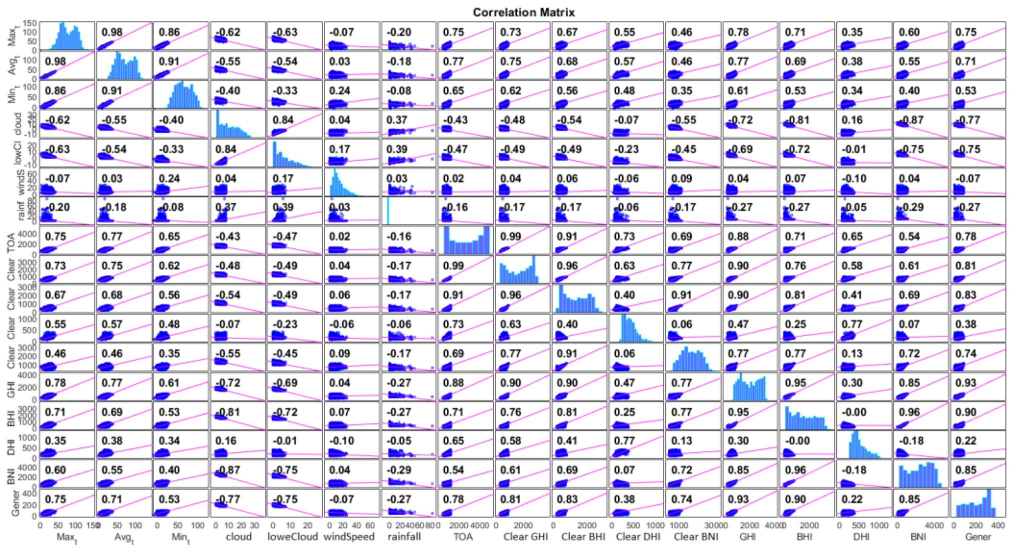
For all the variables the mean and median present small differences which means that the data tend to be symmetrically distributed. Exceptions to the above statement constitute the cases of rainfall and windspeed. This is corroborated from the skewness of each variable. For most of the variables, skewness is very close to 0, presenting a symmetrical distribution. The variables with the higher skewness are rainfall, with the highest skewness among all variables, and lower cloud coverage, windspeed, clear sky DHI and DHI all with a skewness close to 1. As far as the dispersion of each variable is concerned, windspeed, rainfall and clear sky DHI seem to have a leptokurtic distribution (more dispersed distribution with greater outliers) while the rest of the variables present a platykurtic distribution. Exemptions are the low cloud coverage and DHI with kurtosis similar to the normal distribution. It should be noted that the very high kurtosis of rainfall can be explained by many of the observations being equal to zero, due to the limited raining days in Greece.

In Table 5. the Pearson correlation (Eq. 3.1) matrix of the variables is presented. Pearson correlation measures the strength of the linear relationship (either positive or negative) between two variables.

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{3.1}$$

where $cov(X,Y)$, the covariance of variables X and Y and σ_X and σ_Y their standard deviation.

Table 5. Pearson correlation matrix of the variables.



As expected, the different temperature measurements are highly correlated with each other. Furthermore, the cloud measurements have a negative correlation with all the solar radiance measurements except DHI. Windspeed presents very weak, almost non-existent correlation with all the variables. Similarly, rainfall presents weak correlation with the cloud measurements and the solar radiance variables. The solar radiance variables are strongly correlated with each other, with the only exception being the DHI measurements which present lower correlations with the other solar variables. This is because DHI represents solar radiation that is not transmitted directly by the sun but radiation that has been scattered and diffused by clouds or particles in the atmosphere and comes from different directions (Perez-Astudillo & Bachour, 2014).

The variable with the weakest correlation regarding the power generated seems to be windspeed. This finding is in line with other studies (Liu et al., 2018), which also found that wind speed and direction have the lowest correlation with the produced energy among the variables examined. On the other hand, solar radiance variables show a very strong positive correlation with the energy produced, as expected. The variable with the highest correlation between the solar radiance measurements is GHI, this is reasonable since GHI represents the total amount of shortwave irradiation absorbed from a surface which is parallel to the ground (Lopes et al., 2018). On the other hand, DHI has the lowest correlation among the solar variables, as mentioned above this is because DHI represents the diffused solar radiance.

Furthermore, in order to examine whether historical observations of PV power generated have a predictive ability, the autocorrelation of the PV power was examined. The property of long memory is quite evident as presented in the correlograms below where the autocorrelation appears to be persistently strong. Regarding the partial autocorrelation, the first lag is the one that presents the most significant one even if the partial autocorrelation persists, although decreasing, until lag 15.

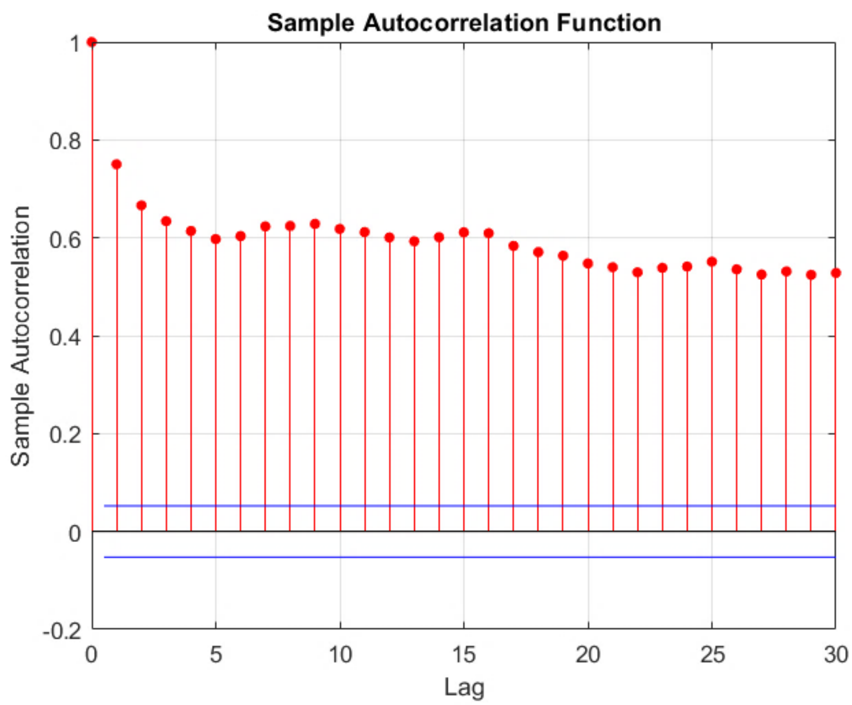


Figure 3. Autocorrelation of PV power generated.

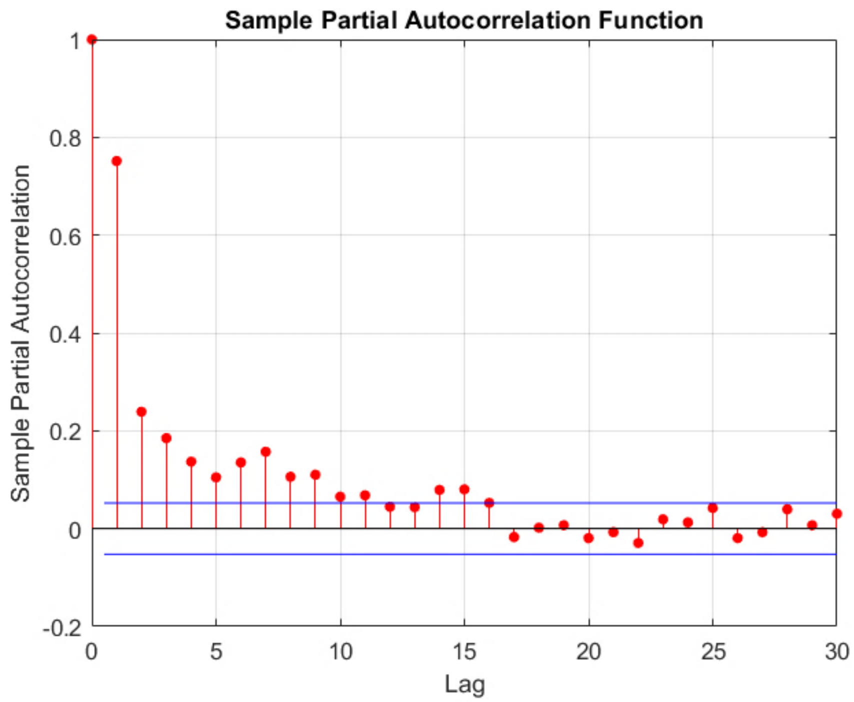


Figure 4. Partial autocorrelation of PV power generated.

Chapter 4
Methodology

As previously stated, this research is structured on two pillars. The first one is the creation of a simple but robust forecasting model to accurately forecast one day ahead PV generated power. The second pillar is to evaluate the financial benefits that the utilization of this model can create for PV power producers who participate in the competitive electricity market. In this Chapter the methodology for the creation and evaluation of the forecasting model will be presented. The forecasting model proposed is an ANN trained with historical meteorological and power data. The model will, then, be benchmarked against a multiple linear regression and two persistent models to evaluate its performance. Lastly, it will be evaluated in an energy market framework based on the imbalance fees generated from the model's inaccuracy.

4.1 Data pre-processing

After the meteorological and energy output data were collected, the process of data cleansing was initiated. Since the collected data were derived from three different providers, each with its own templates and standards, the data needed to be aggregated into a single dataset. The different inconsistencies in the uniformity of the dataset were addressed and then the data were properly timestamped. After that, the missing values of the dataset were filled in using linear interpolation (Eq. 4.1).

$$y' = y_1 + (x' - x_1)\frac{(y_2 - y_1)}{(x_2 - x_1)} \tag{4.1}$$

4.2 Feature selection

As stated before, the selection of input variables that significantly affect the target-variable is one of the most important steps in the design of a forecasting model. The selection of the features that comprise the input vector greatly influences the performance of the forecasting model. A high number of input features, called forecasting parameters, makes the model complex and computationally expensive and especially in the case of neural networks, it can lead to poor generalization. On the other hand, too few forecasting parameters lead to an incomplete and suboptimal forecasting model. Therefore, it is essential to find an adequate choice of input variables leading to the highest possible model performance, with the lowest complexity and costs. In this research this issue is approached by using Stepwise Regression to assess the influence of each meteorological variable to the power generated from the PV plant before introducing them to the model. Stepwise Regression is a method which iterates regression models, each with a combination of different explanatory variables. The model starts with no predictive variables and in each iteration, variables are added or removed based on their statistical significance (the thresholds for addition and removal are based on their p-values). In this way it is ensured that only the variables that significantly affect the PV output are selected as inputs and thus, the training of the model will require less time and computational power. The stepwise regression algorithm led to the selection of 6 meteorological variables to be used as input to the ANN presented to the Table below (Table 6).

It should be noted that feature selection with the use of stepwise regression is an automated process. Automated processes, in model building never come without disadvantages. Quite a few authors (Thompson, 1989; Harrell, 2001; Flom & Cassell, 2007) have suggested that statistical tests of stepwise regression may lead to too low p values which can, in their turn, lead to overfitting thus creating a false confidence in the final model. However, stepwise regression remains quite popular and has been used in feature selection of PV forecasting problems (Ramsami & Oree, 2015).

Table 6. ANN input meteorological variables produced from Stepwise Regression.

Input variables
Cloud coverage
Low cloud coverage
Rainfall
Clear-sky GHI
GHI
BNI

4.3 Training of ANN

The ANN used in this study consists of three layers: an input layer, a single hidden layer and an output layer. The input layer receives two input vectors: the vector of the 6 meteorological parameters on day *t-1*, and the power generated on day *t-1*. This lagged term is included in the network so as to capitalize on the predictive property of the past observations of the generated power which were presented above. The output layer produces the forecasted PV power on day *t*. Therefore, the problem is to describe the relationship between the inputs and the output based on the data included in the database described before. In other words, the problem can be stated as: *Is it possible to find a formula which can be used to estimate the total PV power that will be generated one day ahead, based on the meteorological data observed and the power generated today?*

The dataset of 1461 daily observations were split into two subsets of 1200 (from 1/1/2018 to 14/4/2021) and 261 observations (from 15/4/2021 to 31/12/2021). The first subset was used for the training of the ANN (80% was used for training and 20% for validation). This dataset is comprised of daily observations for the meteorological variables: cloud coverage, low cloud coverage, rainfall, Clear – sky GHI, GHI and BNI, as well as the PV power generated for the days from 1/1/2018 to 14/4/2021. The second subset which contained the observations for the same variables from 15/4/2021 to 31/12/2021, was used for evaluating the forecasting ability of the model. An expanding window approach was used for the recursive generation of one-step ahead forecasts, a method which introduces an ever-increasing training sample to the ANN and is expected to yield more accurate results.

A set of input and output data is used to train the network during the training step. After the introduction of the input data the network calculates its output, which is then compared to the actual target. The difference between these two values produces an error (e). The performance function, which is usually used for training feed forward networks, and also used in this case, is the mean squared error with regularization performance function $\overline{E}$ as:

$$\bar{E} = \frac{1}{n}\left(\beta \sum_{i=1}^{n} e_i^2 + \alpha \sum_{j=1}^{n} w_j^2\right) \tag{4.2}$$

where α, β are regularization parameters and w_j the weight of each neuron.

After a number of iterations (training epochs) the mean square error between the target and the network generated outputs settles to a minimum value.

The learning algorithm of the ANN used in this study is the Bayesian Regularization algorithm. This approach is followed to avoid overfitting. Bayesian Regularization is used to automatically select regularization parameters during the training of the ANN, in order to produce a generalized network. The goal of ANN training is to produce a network with the lowest possible error, which is able to respond well to new data, different from the ones it was trained with. When this is the case, the network is said to be well generalized. Bayesian Regularization improves the generalization process by limiting the weights of the network. If the weights of the network are smaller, the network will respond more smoothly to new data (Kusuma et al., 2021). Equation (4.2) is used to update the weights of the ANN in order to reach a network with the lowest error. However, conventional training algorithms, face difficulties in determining the size of the regularization parameters α, β. Bayesian regularization optimizes these parameters as follows:

$$\begin{cases} \alpha = \dfrac{\gamma}{2E_w} \\ \beta = \dfrac{n-\gamma}{2E_e} \end{cases} \tag{4.6}$$

where $\gamma = n - 2atr(H)^{-1}$, H the Hessian matrix of function $\bar{E}$, $E_w = \frac{1}{n}\sum_{j=1}^{n} w_j^2$ and $E_e = \frac{1}{n}\sum_{i=1}^{n} e_i^2$.

For the network, the linear function (Eq. 4.3) is used as the transfer function of the output layer, while the sigmoid function (Eq. 4.4) is used as an activation function for the neurons of the hidden layer.

$$f(u) = u \tag{4.3}$$

$$f(u) = \frac{1}{1+e^{-u}} \tag{4.4}$$

Since the value of the sigmoid function is limited in the range of 0 to 1, the training data should be rescaled before being introduced to the ANN. The rescaling of the meteorological data and the PV output data is performed by the function below:

$$x'_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{4.5}$$

Where x_i is the i^{th} component of the original data, x_{min} and x_{max} are the minimum and maximum values of the input data respectively.

The number of neurons that constitute the hidden layer was based on a simple “trial and error” process. It was concluded that a single hidden layer with 16 neurons was the optimal architecture of the network for this study. After repeatedly testing and evaluating the validation performance of the network with different combinations of numbers of hidden layers and neurons, this specific architecture was the one that presented the lowest error during these tests.

It should be noted that for the training of the ANN the Neural Net Time Series tool of MatLab® (Ver. R2022a) was utilized.

4.4 Benchmarking models

After the training of the ANN three simple models were created, in order to compare the performance of the more complex and sophisticated ANN to more basic but robust predictive models that have been used in literature.

Persistence model

The model that will be used as a benchmark will be the persistence model (Eq. 4.7), which will be used as a floor in the benchmarking process. If the performance of any complex forecasting model falls below the performance of the naïve model, it should be instantly disregarded since it offers a worse performance at a much higher computational cost.

$$\hat{P}_t = P_{t-1} \tag{4.7}$$

where $\hat{P}_t$ is the prediction of the next day’s power and P_{t-1} is the power generated on the previous day.

Moving Average

A simple moving average will be used as the second model. In this case the power that will be produced is assumed to be equal to the average of the three previous days:

$$\hat{P}_t = \frac{(P_{t-1} + P_{t-2} + P_{t-3})}{3} \tag{4.8}$$

Linear Regression

A multiple linear regression predictive model is constructed as an upper limit for the benchmarking process, due to its high accuracy presented in literature (Abuella & Chowdhury, 2015). The linear regression was trained with the same data used to train the ANN and the forecasting approach will also be recursive one step ahead forecasts on an expanding window. The predictive function can be presented as:

$$\hat{P}_t = \beta_0 + \beta_1 X_{clouds} + \beta_2 X_{lowclouds} + \beta_3 X_{rainfall} + \beta_4 X_{ClearCHI} + \beta_5 X_{GHI} + \beta_6 X_{BNI} + \beta_7 P_{t-1} \tag{4.9}$$

where β_0 to β_7 the regression coefficients and X_{clouds} to X_{BNI} the respective input variables and P_{t-1} the power generated on day *t-1*.

4.5 Evaluation of the forecasting model

The ANN predictive model will be evaluated on two different levels. The model will be benchmarked against the models presented above to compare their performance and accuracy. For benchmarking, the evaluation metrics widely used in literature will be utilized. These measures describe the difference between the observed values of the dependent variable and their predicted values. Obviously, lower values of these metrics mean that the predictive model returns values for the dependent variable that is quite close to the actual, observed value and therefore the model presents higher predictive accuracy and lower predictive error. The evaluation metrics that will be used for benchmarking are presented in the table below:

Table 7. Error evaluation metrics.

Metric	Formula
RMSE	$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$
NRMSE	$NRMSE = \frac{RMSE}{y_{max} - y_{min}}$
MAE	$MAE = \frac{1}{n} \sum_{i=1}^{n} \lvert \hat{y}_i - y_i \rvert$

where, $\hat{y}_i$ and y_i respectively the predicted and actual values of the PV output generated and y_{max} and y_{min} the maximum and minimum recorded values of y.

4.6 Economic analysis

Lastly, an analysis will be carried out to evaluate the economic impact of the accuracy of the forecasting method, focusing the analysis on the PV energy that is introduced to the electric grid through the electricity markets.

Greece is in a special place where not only its electricity market structure changes to a harmonized EU target model, but it is happening while there is a great penetration of RES to the market (Kontochristopoulos et al., 2021). Generally, in liberalized energy markets, it is assumed that electricity generation sources are utilized based on their marginal cost. Therefore, in most cases, the most expensive source of electricity in operation sets the electricity price for the whole market, which is the marginal price of the electricity market. This is a very logical pricing scheme for conventional energy sources (fossil fuels etc.). However, due to the massive adoption of RES, this assumption can become obsolete, since the most common RES technologies, such as PVs, have almost zero marginal costs (Kraan et al., 2018). The minimal marginal costs combined with special, risk-free RES subsidization schemes that act as incentives for the green energy transition, has led to distorted market prices and wrong economic signals to market actors (Ntomaris et al., 2022). To address these issues market policies have changed to ensure a level-playing field for all generating sources. According to the European Commission's State Aid Guidelines (2014–2020), RES producers will be compensated based on market mechanisms and will be subject to imbalance costs.

To evaluate the effects of these imbalance costs on PV power producers a simple model was built based on two markets of the Greek wholesale electricity market:

Day-Ahead Market (DAM): In this market a Day Ahead Scheduling (DAS) is performed by the electricity market operator (EnEx) to optimize the allocation of all the available energy resources. Market participants who represent generating resources, have to place a bid for their available generation capacity for the next day. Traders and energy suppliers also place their offers, demanding the available energy (EnEx, 2021). Taking the bids and the different system limitations into consideration, EnEx clears the electricity market, aiming to meet the forecasted demand, while at the same time minimizing the overall system costs (Dagoumas & Polemis, 2017). The output of the DAM is the timetable indicating where all the available electricity generating sources are committed, along with the system's marginal price for the next day.

Balancing Market: The Balancing Market deals with system imbalances in real-time which are settled by the Transmission System Operator (TSO). The imbalance price is the price of DAM plus a markup which is cost reflective of the expenses incurred by the System Operator to provide real time balancing energy to the system (Ntomaris, 2022). For example, if the system is undersupplied, parties with a shortage must pay the markup while parties with an energy surplus receive this markup in return for their contribution of energy to the system. On the other hand, if the system is oversupplied, parties with a surplus will have to accept a discounted price for their uncommitted generated energy.

With respect to the above, a simple function is created to evaluate the economic effect of the predictive error of the ANN forecasting model. This function is intended for a solar power producer who commits all their available predicted next-day power output $\hat{P}$ on the Day-Ahead Market at a spot price *S*. The profit of the producer can be calculated by the product of $\hat{P} * S$. The next day, if the actual PV output (P) is different from the one pledged on the DAM, the difference of the output, which is equal to the predictive error ($e = P - \hat{P}$), is either bought from or sold to the System Operator settled by the imbalance price (I). Thus, the profit function of the producer can be:

$$\Pi = \hat{P} * S + e * I \tag{4.10}$$

However, when considering the state of the system (either oversupplied or undersupplied) the concept of opportunity cost arises. For example, let's consider a producer with a surplus of the energy pledged on DAM, if the system is oversupplied the imbalance price will probably be lower than the DAM price, thus there is an opportunity cost for the producer (they could have pledged more energy at the higher DAM price, if they had forecasted accurately). Therefore, the profit function of the producer, incorporating opportunity costs, can be described as follows:

$$\Pi_{opportunity} = P * S + e * (I - S) \tag{4.11}$$

It should be stressed that this model is not meant to realistically describe the imbalance clearing process in the Greek electricity market at the time of writing, but just to be used as a simple profit function for the scope of this research.

Chapter 5
Results

5.1 Statistical evaluation

The ANN was trained to obtain a relationship between the input variables and the energy that will be produced in the next day by the PV system. The regression plot for each model is presented in Figures 5 to 8. These figures depict the comparison between the predicted output energy of the PV generated by each model (vertical axis) and the actual measured values (horizontal axis) for the period ranging from 15 April 2021 to 31 December 2021. Table 8 summarizes the quantitative evaluation of the predictive model compared to the linear regression, moving average and persistence model in terms of the statistical parameters RMSE, NRMSE, MAE, R, and standard deviation.

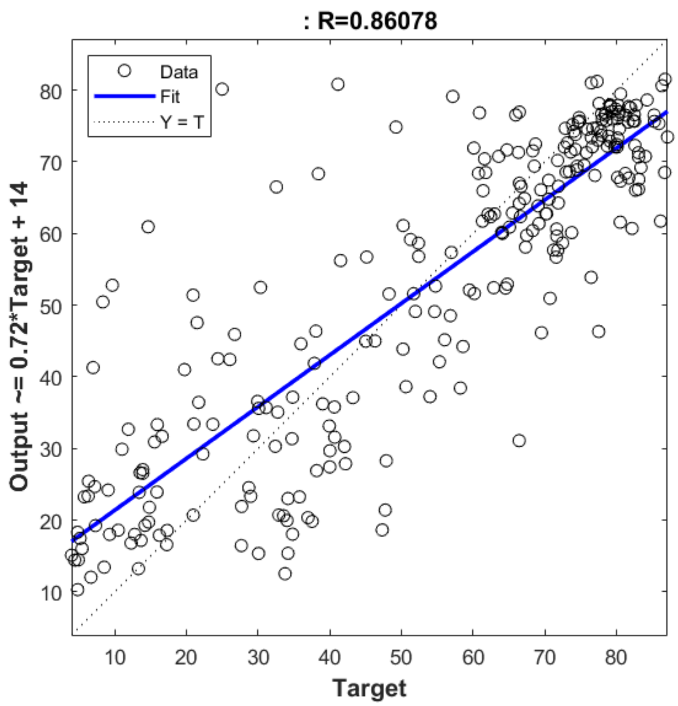


Figure 5. ANN regression results.

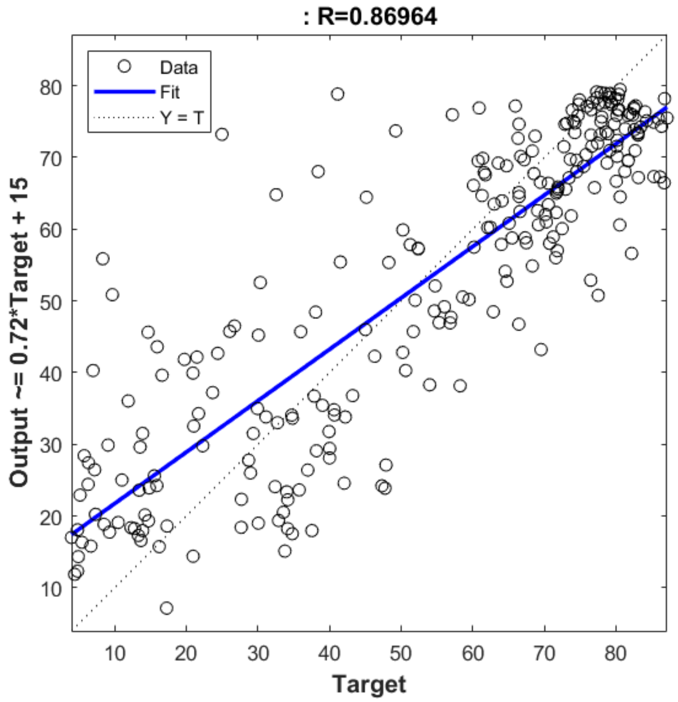


Figure 6. Linear regression results.

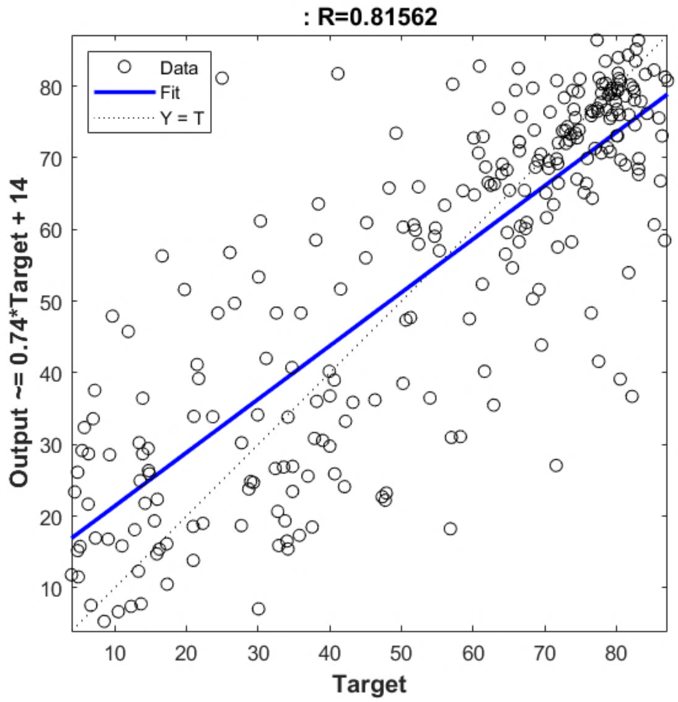


Figure 7. Moving Average regression results.

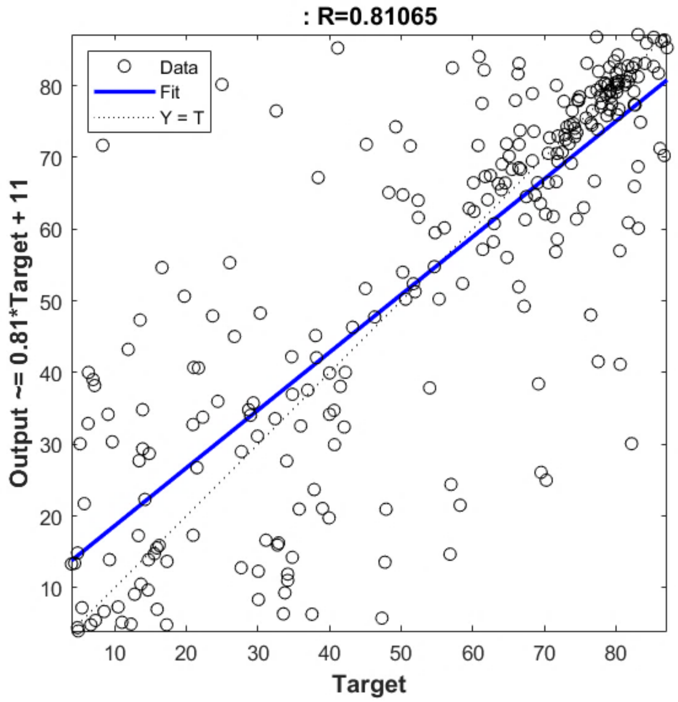


Figure 8. Persistence Model regression results.

Table 8. Evaluation of Models.

Model	RMSE	NRMSE	MAE	R	σ
Regression	**12.7265**	**15.31%**	9.8256	**0.8696**	**12.729**
ANN	13.0743	15.72%	**9.7394**	0.8608	13.076
Moving Average	15.0330	18.08%	10.6407	0.8156	15.056
Persistence	15.7189	18.90%	10.2998	0.8106	15.748

Results are consistent with the literature that indicates that ANN outperforms the persistence and moving average models, even at short term horizons in which they are at the peak of their performance (Zhang et al., 2015). On the other hand, the linear regression outperforms the ANN regarding all evaluation metrics except MAE. RMSE is more penalizing for greater errors and outliers, while the MAE equally weights the errors. Therefore, according to the Table above, the ANN produces some greater outlier errors compared to the regression model, but it produces a lower average deviation from the target value. These findings are similar to other presented in literature (Sharma et al., 2018), where a simple ANN is found to outperform persistent models while being outperformed (by a small margin) by linear and polynomial regressions.

Another way that is worth examining the errors of each model, is the magnitude of the error that each model produces based on each season. The data were split into 3 seasons: data from 15/4/21 - 30/6/21, were assigned to the season *spring*, while the data from 1/7/21 - 30/9/2021 were assigned to *summer* and the data from 1/10/2021 - 31/12/2021 were assigned to *winter*. The RMSE of errors of each model in each specific season are presented at the Table below, where the lowest RMSE per season is in bold. Furthermore, the standard deviation of the energy produced in each season is also presented in the Table below.

Table 9. Model RMSE based on the season.

Model	**Spring**	**Summer**	**Winter**
Regression	**14.425**	**10.390**	13.336
ANN	14.673	11.423	**13.200**
Moving Average	18.137	11.928	14.960
Persistence	18.746	11.288	16.716
Variable	**Standard Deviation**		
Energy	17.219	14.062	15.609

The regression model presents the lowest error in spring and summer, while the ANN produces the lowest error in winter, although the errors of these two models in spring and winter are quite close. The ANN model produces a lower RMSE, compared to the persistence model and the moving average, during spring and winter, seasons that have a lot more volatile meteorological conditions compared to summer, as it is suggested by the standard deviation of the produced energy. On the other hand, the persistence model generates a lower RMSE compared to the ANN in the summer. This seems logical, since in the summer, where there are only small deviations in the daily weather (compared to the other two seasons), usually in the temperature, the past days can be sufficient predictors of the future ones. The results presented above agree with the results presented in Kardakos et al. (2012) where the persistent model also outperformed an ANN model in one day ahead forecasting during the summer. The authors of the abovementioned study conclude that *“the persistence model (…) lead(s) to superior or comparable forecasts only during the summer months and September, where the favorable weather conditions in Greece eliminate the need to use a more sophisticated short-term forecasting model of PV power generation.”*

5.2 Economic evaluation

In order to get a tangible overview of the predictive performance of the ANN model an analysis has been carried out to evaluate the economic impact of the models. As previously presented, producers of renewable energy resources can participate in the wholesale electricity market through programmed transactions. In the Day Ahead Market producers declare the amount of energy they will provide to the energy grid the next day. However, imbalance charges are applied when there is a discrepancy between the volume declared and the volume actually provided. For this analysis the profit function that was presented in the previous chapter is used. Utilizing this function, the daily profits of the producer, adjusted for opportunity costs, will be calculated. The actual data of the DAM price and imbalance price for the span of 15/4/2021 to 31/12/2021 will be used for this calculation. The data were acquired from HEnEx and IPTO respectively. It should be noted that the original data were provided in 15-minute iterations, so a weighted average was used to calculate the daily average prices. In Figure 9 the daily average DAM and imbalance prices are presented.

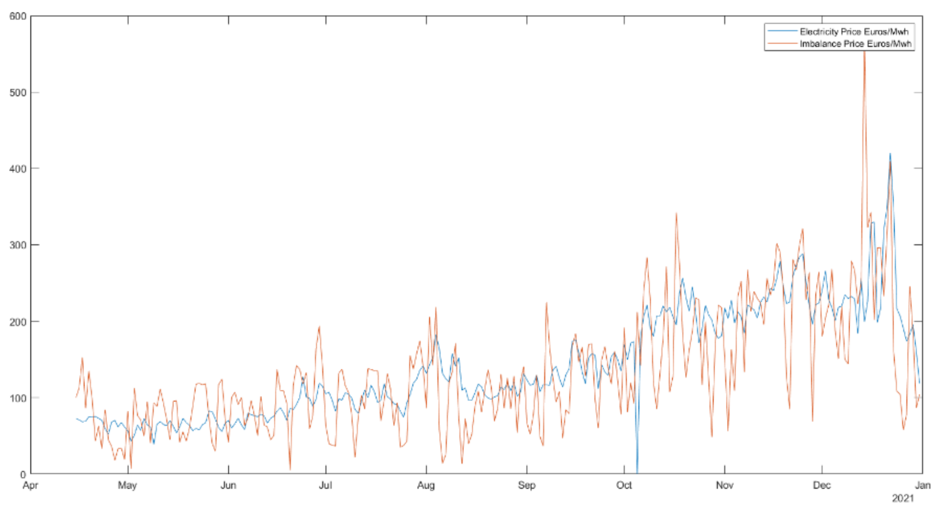


Figure 9. DAM and Imbalance prices for the Greek electricity market.

Finally, in Table 10 the total income of the producer, adjusted for opportunity costs (Eq. 4.11), that are generated with each predictive model are presented.

Table 10. Producer incomes in the Greek electricity market.

Model	**Income (€)**
Regression	1,593.2
ANN	1,592.9
Moving Average	1,591.6
Persistence	1,599.7

Surprisingly, the persistence model, the least accurate one with respect to the statistical measurements, yields the most income for the producer. This is because the Single Price Imbalance system that is used in the Greek energy market allows the Balance Responsible Parties (BRPs), the producer in this case, to profit from being on the opposite direction of the grid (Ntomaris et al., 2022). In this case, for the period that is examined, the mean daily difference between the Imbalance price and DAM price is –7.5751 €/MWh, as well as the Imbalance price was lower than the DAM price in 139 days of the 261 studied (53.26%), also the difference between the Imbalance price and DAM presents a skewness of 0.8739. The distribution of the daily difference is presented in Figure 10.

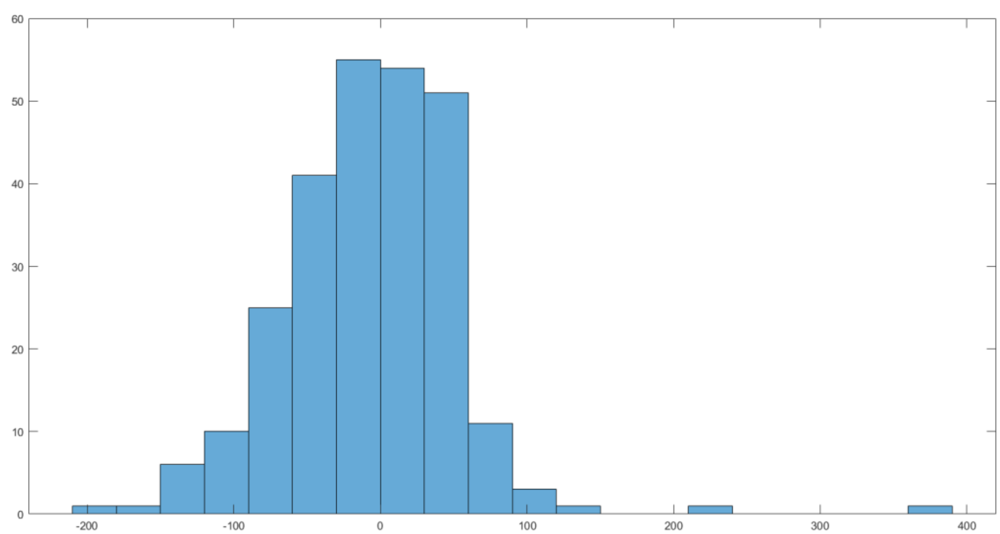


Figure 10. Distribution of the daily difference between Imbalance and DAM price.

This means that the grid was usually oversupplied with energy and so the producers who had a surplus of energy should accept a discounted price for it and thus, leading to an Imbalance price lower than the DAM price. In this situation the producers who produce less than the energy pledged at DAM can acquire the energy they lack in the discounted Imbalance price leading to higher profits. For reference, a producer who generates a constant forecasting error of $e = P - \hat{P} = -30$ kWh, will generate an income of 1,697.0 € for the period examined, an income far greater than the ones generated from the predictive models above. This result is in line with the literature where it is pointed out that the Single Price Imbalance system *"... may produce weak incentives to balance and motivate the BRPs to be imbalanced in the opposite direction from the system deviation."* (Ntomaris et al, 2022).

However, if the models are evaluated in another framework, like the one used in De Giorgi et al. (2015) or in Matsumoto and Yamada (2018), the results can be different. In these studies, the economic impact of the forecasting errors is evaluated based on a Dual Price Imbalance market. If the energy that the producer injects to the grid surpasses the scheduled one, then the producer is compensated for the excess energy at a price *DAM – c*. On the other hand, if the producer generates less energy than the scheduled one, they must procure the missing energy at a price *DAM + l*. In this framework, no matter the system's condition (oversupplied or undersupplied), deviations of any direction are penalized, thus, providing high incentives for the producers to be balanced at all times.

Therefore, the profit functions below are generated:

$$\Pi_{opportunity} = \begin{cases} P * S - c * e, when\ e > 0 \\ P * S + l * e, when\ e < 0 \end{cases} \tag{5.1}$$

The first factor of the function represents the income that the producer will receive when there is no prediction error (scheduled energy equals the energy generated), the second factor of the function which is always negative, represents the opportunity cost compared to the perfect prediction. On the Table 11, below, the incomes generated from each predictive model are presented as a sum of the daily incomes divided by the maximum income the producer can achieve in the period examined (income when there are no forecasting errors). For this calculation, *c* and *l* were set to 50% of the DAM price.

Table 11. Normalized incomes for a Dual Price Imbalance framework.

Model	**Normalized Income (%)**
Regression	88.84
ANN	**88.86**
Moving Average	87.96
Persistence	87.93

In this market framework, the ANN model yields the greater income among the four models studied, although very close to the income generated by the regression model. For reference, the same producer used before as an example (stable forecasting error of -30 kWh) would achieve a normalized income of 71.57% in this case. These results are in line with both the statistical error measurements presented above as well as with the rationale of the Dual Price Imbalance Market. The predictive models with the lowest errors, both in terms of MAE and RMSE, are the ones that receive the least penalization from the market framework.

Chapter 6

Conclusion

In summary, this research attempts to create a simple and robust forecasting model which can be used by solar power producers in the context of electricity markets. The previous day's weather conditions and PV output are provided as input and the model recursively, and utilizing an expanding window, forecasts the PV power that will be generated the next day. Using a dataset comprising of daily data of weather variables and PV power outputs of a specific solar park from 1/1/2018 until 31/12/2021 an ANN was trained. Specifically, the data from 1/1/2018 to 14/4/2021 were used to train the network while the rest of the data was used for its evaluation. The trained model was benchmarked against the persistent model, a three-day moving average and a multiple linear regression. The results of each model were compared via the use of standard evaluation metrics and in the context of two different energy market frameworks. The results of this evaluation indicate that the ANN model can predict the next day's generated PV energy with greater accuracy than the persistent model and the moving average. Compared to the regression model the ANN presents a lower MAE (9.74 compared to 9.82) but a higher RMSE (13.07 compared to 12.73). Evaluating the models in a Single Price Imbalance market was not conclusive regarding their accuracy. On the other hand, as expected, the two more accurate predictive models lead to higher incomes for producers and lower imbalance fees in a Dual Price Imbalance market. These findings seem to agree with the existing literature which suggests that ANNs outperform simpler models even in short term forecasting horizons, in which the latter are most accurate, and perform similarly to regression models (e.g., Sharma et al., 2018; Raza et al., 2018).

Although promising, it should be noted that this research and by extension its findings are hindered by the limitations imposed by data availability and quality issues. Thus, restricting the quality of the results that could have been produced from conducting forecasts with hourly data for a time horizon of 24 hours ahead. Most studies presented in literature regarding 24 hours ahead forecasts usually construct an ANN which receives as inputs weather forecasts of the next day and generates the forecasted PV energy, with the data being mostly in hourly iterations (Chupong & Plangklang, 2011; De Giorgi et al., 2014; Leva et al., 2015; Liu et al., 2015). On the other hand, most of the few studies that have approached the forecasting effort from a time series point of view simply utilize only historical PV data, again in hourly iterations (e.g., Kardakos et al., 2012), while the ones that use exogenous inputs also use the weather forecast of the next day as the Network's input (e.g., Sharma et al., 2018). Another significant point is that most of these studies utilize PV data from PV panels set up specifically for research purposes which are accompanied by weather stations that provide accurate data and measurements. That being said, due to the lack of access to hourly PV data or next day hourly forecasts for weather variables, this research used, as an alternative, the methodology described in Chapter 4, which is based on daily observations and historical meteorological data. This is why the results of this study cannot be directly compared to other studies presented in literature. However, in Chapter 5 the results have been compared to some comparable results of specific studies. As it can be deduced from the above, access to specialized PV and weather data could lead to a more complete study with more accurate results comparable to the ones presented in literature.

However, this must not take away from the purpose of this master's thesis which is also to act as an advocate for the need of accurate PV power forecasting in the context of energy markets. Now, more than ever, this need is evident since energy markets are moving away from special subsidization regimes for RES and towards a competitive framework where renewable energy producers will have to deal with imbalance costs and fees just as the conventional energy producers. Therefore, improvements in RES power forecasting are needed in order to not only expedite the transition away from polluting fossil fuels but also lead to more stable electric grids and lower electricity prices.

References

Abuella, M., & Chowdhury, B. (2015, April). Solar power probabilistic forecasting by using multiple linear regression analysis. In *SoutheastCon 2015* (pp. 1-5). IEEE.

Agoua, X. G., Girard, R., & Kariniotakis, G. (2021). Photovoltaic power forecasting: Assessment of the impact of multiple sources of spatio-temporal data on forecast accuracy. *Energies*, 14(5), 1432.

Agreement, P. (2015, December). Paris agreement. In Report of the Conference of the Parties to the United Nations Framework Convention on Climate Change (21st Session, 2015: Paris).

Ahmed, A., & Khalid, M. (2019). A review on the selected applications of forecasting models in renewable power systems. *Renewable and Sustainable Energy Reviews*, 100, 9-21.

Ahmed, R., Sreeram, V., Mishra, Y., & Arif, M. D. (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.

Al-Dahidi, S., Ayadi, O., Adeeb, J., Alrbai, M., & Qawasmeh, B. R. (2018). Extreme learning machines for solar photovoltaic power predictions. *Energies*, *11*(10), 2725.

Alfadda, A., Adhikari, R., Kuzlu, M., & Rahman, S. (2017, April). Hour-ahead solar PV power forecasting using SVR based approach. In 2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT) (pp. 1-5). IEEE.

AlKandari, M., & Ahmad, I. (2020). Solar power generation forecasting using ensemble approach based on deep learning and statistical methods. Applied Computing and Informatics.

Aminzadeh, F., & De Groot, P. (2006). Neural networks and other soft computing techniques with applications in the oil industry. Eage Publications.

Behera, M. K., & Nayak, N. (2020). A comparative study on short-term PV power forecasting using decomposition based optimized extreme learning machine algorithm. *Engineering Science and Technology, an International Journal*, *23*(1), 156-167.

Bjarghov, S., Löschenbrand, M., Saif, A. I., Pedrero, R. A., Pfeiffer, C., Khadem, S. K., ... & Farahmand, H. (2021). Developments and challenges in local electricity markets: A comprehensive review. *IEEE Access*, 9, 58910-58943.

Box, G. E., & Jenkins, G. M. (1976). Time series analysis: Forecasting and control San Francisco. *Calif: Holden-Day*.

Brown, R. G. (1957, January). Exponential smoothing for predicting demand. In *Operations Research* (Vol. 5, No. 1, pp. 145-145). 901 ELKRIDGE LANDING RD, STE 400, LINTHICUM HTS, MD 21090-2909: Inst Operations Research Management Sciences.

Cadenas, E., Rivera, W., Campos-Amezcua, R., & Heard, C. (2016). Wind speed prediction using a univariate ARIMA model and a multivariate NARX model. *Energies*, 9(2), 109.

Cassola, F., & Burlando, M. (2012). Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output. *Applied energy*, 99, 154-166.

Chatziagorakis, P., Ziogou, C., Elmasides, C., Sirakoulis, G. C., Karafyllidis, I., Andreadis, I., ... & Voutetakis, S. (2016). Enhancement of hybrid renewable energy systems control with neural networks applied to weather forecasting: the case of Olvio. *Neural Computing and Applications*, *27*(5), 1093-1118.

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.

Chupong, C., & Plangklang, B. (2011). Forecasting power output of PV grid connected system in Thailand without using solar radiation measurement. *Energy Procedia*, 9, 230-237.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273-297.

Dagoumas, A. S., & Polemis, M. L. (2017). An integrated model for assessing electricity retailer's profitability with demand response. *Applied Energy*, 198, 49-64.

Das, S. (2021). Short term forecasting of solar radiation and power output of 89.6 kWp solar PV power plant. *Materials Today: Proceedings*, 39, 1959-1969.

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.

David, M., Ramahatana, F., Trombe, P. J., & Lauret, P. (2016). Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models. *Solar Energy, 133,* 55-72.

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 Giorgi, M. G., Congedo, P. M., Malvoni, M., & Laforgia, D. (2015). Error analysis of hybrid photovoltaic power forecasting models: A case study of mediterranean climate. *Energy conversion and management*, 100, 117-130.

Dogan, G., Labeau, P. E., Maun, J. C., Sprooten, J., Galvez, M., & Sleurs, K. (2016, April). Monte Carlo sampling vs. discrete forecast error scenarios in grid reliability assessment for short-term planning. In *2016 IEEE International Energy Conference (ENERGYCON)* (pp. 1-6). IEEE.

EnEx, (2021). HEnEx SPOT TRADING RULEBOOK, Day-Ahead & Intra-Day Markets Trading *Rulebook Version 1.5*

European Commission, Guidelines on State aid for environmental protection and energy 2014-2020 (2014).

Tratar, L. F., & Strmčnik, E. (2016). The comparison of Holt–Winters method and Multiple regression method: A case study. *Energy*, *109*, 266-276.

Firat, U., Engin, S. N., Saraclar, M., & Ertuzun, A. B. (2010, December). Wind speed forecasting based on second order blind identification and autoregressive model. In *2010 Ninth International Conference on Machine Learning and Applications* (pp. 686-691). IEEE.

Flom, P. L., & Cassell, D. L. (2007, November). Stopping stepwise: Why stepwise and similar selection methods are bad, and what you should use. In *NorthEast SAS Users Group Inc 20th Annual Conference* (Vol. 11).

Gonzalo, A. P., Marugán, A. P., & Márquez, F. P. G. (2019). A review of the application performances of concentrated solar power systems. *Applied Energy*, 255, 113893.

Gumus, B., & Kilic, H. (2018). Time dependent prediction of monthly global solar radiation and sunshine duration using exponentially weighted moving average in southeastern of Turkey. *Thermal Science*, 22(2), 943-951.

Gutierrez-Corea, F. V., Manso-Callejo, M. A., Moreno-Regidor, M. P., & Manrique-Sancho, M. T. (2016). Forecasting short-term solar irradiance based on artificial neural networks and data from neighbouring meteorological stations. *Solar Energy*, 134, 119-131.

Haessig, P., Multon, B., Ahmed, H. B., Lascaud, S., & Bondon, P. (2015). Energy storage sizing for wind power: impact of the autocorrelation of day-ahead forecast errors. *Wind Energy*, 18(1), 43-57.

Haiges, R., Wang, Y. D., Ghoshray, A., & Roskilly, A. P. (2017). Forecasting electricity generation capacity in Malaysia: an auto regressive integrated moving average approach. *Energy Procedia*, 105, 3471-3478.

Han, Y., Wang, N., Ma, M., Zhou, H., Dai, S., & Zhu, H. (2019). A PV power interval forecasting based on seasonal model and nonparametric estimation algorithm. *Solar Energy*, 184, 515-526.

Harrell, F. E. (2001). *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis* (Vol. 608). New York: springer.

Hossain, M., Mekhilef, S., Danesh, M., Olatomiwa, L., & Shamshirband, S. (2017). Application of extreme learning machine for short term output power forecasting of three grid-connected PV systems. *journal of Cleaner Production*, *167*, 395-405.

Howlader, H. O. R., Matayoshi, H., & Senjyu, T. (2015). Distributed generation incorporated with the thermal generation for optimum operation of a smart grid considering forecast error. *Energy Conversion and Management*, 96, 303-314.

Huang, R., Huang, T., Gadh, R., & Li, N. (2012, November). Solar generation prediction using the ARMA model in a laboratory-level micro-grid. In *2012 IEEE third international conference on smart grid communications (SmartGridComm)* (pp. 528-533). IEEE.

Simoglou, C. K., Vagropoulos, S. I., Bakirtzis, E. A., Kardakos, E. G., Chatzigiannis, D. I., Biskas, P. N., & Bakirtzis, A. G. (2015, September). Large-scale res integration in electricity markets: Challenges and potential solutions. In *2015 50th International Universities Power Engineering Conference (UPEC)* (pp. 1-6). IEEE.

Kaur, A., Nonnenmacher, L., Pedro, H. T., & Coimbra, C. F. (2016). Benefits of solar forecasting for energy imbalance markets. *Renewable energy*, 86, 819-830.

Kazem, H. A., & Yousif, J. H. (2017). Comparison of prediction methods of photovoltaic power system production using a measured dataset. *Energy conversion and management*, 148, 1070-1081.

Kontochristopoulos, Y., Michas, S., Kleanthis, N., & Flamos, A. (2021). Investigating the market effects of increased RES penetration with BSAM: A wholesale electricity market simulator. *Energy Reports*, 7, 4905-4929.

Kraan, O., Kramer, G. J., & Nikolic, I. (2018). Investment in the future electricity system-An agent-based modelling approach. *Energy*, 151, 569-580.

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.

Li, Q., Wu, Z., & Xia, X. (2018). Estimate and characterize PV power at demand-side hybrid system. *Applied Energy*, *218*, 66-77.

Lipperheide, M., Bosch, J. L., & Kleissl, J. (2015). Embedded nowcasting method using cloud speed persistence for a photovoltaic power plant. *Solar Energy*, *112*, 232-238.

Liu, J., Fang, W., Zhang, X., & Yang, C. (2015). An improved photovoltaic power forecasting model with the assistance of aerosol index data. *IEEE Transactions on Sustainable Energy*, 6(2), 434-442.

Mora-Lopez, L. L., & Sidrach-de-Cardona, M. (1998). Multiplicative ARMA models to generate hourly series of global irradiation. *Solar Energy*, *63*(5), 283-291.

Lonij, V. P., Brooks, A. E., Cronin, A. D., Leuthold, M., & Koch, K. (2013). Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors. *Solar energy*, *97*, 58-66.

Lopes, F. M., Silva, H. G., Salgado, R., Cavaco, A., Canhoto, P., & Collares-Pereira, M. (2018). Short-term forecasts of GHI and DNI for solar energy systems operation: assessment of the ECMWF integrated forecasting system in southern Portugal. *Solar Energy*, 170, 14-30.

Lorenz, E., & Heinemann, D. (2012). Prediction of solar irradiance and photovoltaic power.

Mayer, K., & Trück, S. (2018). Electricity markets around the world. *Journal of Commodity Markets*, 9, 77-100.

Mellit, A., Sağlam, S., & Kalogirou, S. A. (2013). Artificial neural network-based model for estimating the produced power of a photovoltaic module. *Renewable Energy*, 60, 71-78.

Mosavi, A., Salimi, M., Faizollahzadeh Ardabili, S., Rabczuk, T., Shamshirband, S., & Varkonyi-Koczy, A. R. (2019). State of the art of machine learning models in energy systems, a systematic review. *Energies*, 12(7), 1301.

Ntomaris, A. V., Marneris, I. G., Biskas, P. N., & Bakirtzis, A. G. (2022). Optimal participation of RES aggregators in electricity markets under main imbalance pricing schemes: Price taker and price maker approach. *Electric Power Systems Research*, 206, 107786.

Ortega-Vazquez, M. A., & Kirschen, D. S. (2008). Estimating the spinning reserve requirements in systems with significant wind power generation penetration. *IEEE Transactions on power systems*, 24(1), 114-124.

Orwig, K. D., Ahlstrom, M. L., Banunarayanan, V., Sharp, J., Wilczak, J. M., Freedman, J., ... & Marquis, M. (2014). Recent trends in variable generation forecasting and its value to the power system. *IEEE Transactions on Sustainable Energy*, 6(3), 924-933.

Pedro, H. T., & Coimbra, C. F. (2012). Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy*, 86(7), 2017-2028.

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.

Pérez, E., Pérez, J., Segarra-Tamarit, J., & Beltran, H. (2021). A deep learning model for intra-day forecasting of solar irradiance using satellite-based estimations in the vicinity of a PV power plant. *Solar Energy*, 218, 652-660.

Ramsami, P., & Oree, V. (2015). A hybrid method for forecasting the energy output of photovoltaic systems. *Energy Conversion and Management*, 95, 406-413.

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

Ren, Y., Suganthan, P. N., & 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.

Schwingshackl, C., Petitta, M., Wagner, J. E., Belluardo, G., Moser, D., Castelli, M., ... & Tetzlaff, A. (2013). Wind effect on PV module temperature: Analysis of different techniques for an accurate estimation. *Energy Procedia*, *40*, 77-86.

Sharma, V., Cali, U., Hagenmeyer, V., Mikut, R., & Ordiano, J. Á. G. (2018, June). Numerical weather prediction data free solar power forecasting with neural networks. *In Proceedings of the Ninth International Conference on Future Energy Systems* (pp. 604-609).

Shi, J., Lee, W. J., Liu, Y., Yang, Y., & Wang, P. (2012). Forecasting power output of photovoltaic systems based on weather classification and support vector machines. *IEEE Transactions on Industry Applications*, *48*(3), 1064-1069.

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

Soman, S. S., Zareipour, H., Malik, O., & Mandal, P. (2010, September). A review of wind power and wind speed forecasting methods with different time horizons. In *North American power symposium 2010* (pp. 1-8). IEEE.

Thompson, B. (1989). Why won't stepwise methods die? *Measurement and Evaluation in Counseling and Development*, *21*(4), 146-148.

Yu, T. C., & Chang, H. T. (2011, April). The forecast of the electrical energy generated by photovoltaic systems using neural network method. In *2011 International Conference on Electric Information and Control Engineering* (pp. 2758-2761). IEEE.

Torabi, M., Hashemi, S., Saybani, M. R., Shamshirband, S., & Mosavi, A. (2019). A Hybrid clustering and classification technique for forecasting short-term energy consumption. *Environmental progress & sustainable energy, 38*(1), 66-76.

Ugurlu, U., Oksuz, I., & Tas, O. (2018). Electricity price forecasting using recurrent neural networks. *Energies, 11(5)*, 1255.

Verma, T., Tiwana, A. P. S., Reddy, C. C., Arora, V., & Devanand, P. (2016, January). Data analysis to generate models based on neural network and regression for solar power generation forecasting. *In 2016 7th international conference on intelligent systems, modelling and simulation (ISMS)* (pp. 97-100). IEEE.

Wan, C., Zhao, J., Song, Y., Xu, Z., Lin, J., & Hu, Z. (2015). Photovoltaic and solar power forecasting for smart grid energy management. *CSEE Journal of Power and Energy Systems, 1(4)*, 38-46.

Wang, K., Qi, X., & Liu, H. (2019). A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. *Applied Energy, 251*, 113315.

Wolff, B., Kühnert, J., Lorenz, E., Kramer, O., & Heinemann, D. (2016). Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data. *Solar Energy, 135*, 197-208.

Yang, X., Ren, J., & Yue, H. (2016, August). Photovoltaic power forecasting with a rough set combination method. In *2016 UKACC 11th International Conference on Control (CONTROL)* (pp. 1-6). IEEE.

Yavor, M. (2009). Electrostatic energy analyzers. *Advances in Imaging and Electron Physics, 157*, 213-258.

Yona, A., Senjyu, T., Funabashi, T., & Kim, C. H. (2013). Determination method of insolation prediction with fuzzy and applying neural network for long-term ahead PV power output correction. *IEEE Transactions on Sustainable Energy*, *4*(2), 527-533.

Zhang, J., Hodge, B. M., Lu, S., Hamann, H. F., Lehman, B., Simmons, J., ... & Tedesco, J. (2015). Baseline and target values for regional and point PV power forecasts: Toward improved solar forecasting. *Solar Energy, 122*, 804-819.

Zhang, T., Lv, C., Ma, F., Zhao, K., Wang, H., & O'Hare, G. M. (2020). A photovoltaic power forecasting model based on dendritic neuron networks with the aid of wavelet transform. *Neurocomputing, 397*, 438-446.

Zhao, J., Guo, Z. H., Su, Z. Y., Zhao, Z. Y., Xiao, X., & Liu, F. (2016). An improved multi-step forecasting model based on WRF ensembles and creative fuzzy systems for wind speed. *Applied Energy, 162*, 808-826.

Zhu, H., Li, X., Sun, Q., Nie, L., Yao, J., & Zhao, G. (2015). A power prediction method for photovoltaic power plant based on wavelet decomposition and artificial neural networks. *Energies, 9(1)*, 11.