Lappeenranta-Lahti University of Technology LUT
School of Engineering Science
Computational Engineering and Technical Physics
Computer Vision and Pattern Recognition

Suzan Yemane

DEEP FORECASTING OF RENEWABLE ENERGY PRODUCTION WITH NUMERICAL WEATHER PREDICTIONS

Master's Thesis

Examiners: Professor Lasse Lensu
Professor Galina Malykhina

Supervisors: D.Sc. (Tech.) Toni Kuronen
D.Sc. (Tech.) Aleksei Romanenko
Associate Professor Arto Kaarna
Professor Lasse Lensu

ABSTRACT

Lappeenranta-Lahti University of Technology LUT
School of Engineering Science
Computational Engineering and Technical Physics
Computer Vision and Pattern Recognition

Suzan Yemane

Deep forecasting of renewable energy production with numerical weather predictions

Master's Thesis

2021

52 pages, 29 figures, 14 tables.

Examiners: Professor Lasse Lensu
Professor Galina Malykhina

Keywords: renewable energy, photovoltaic production, machine learning, probabilistic forecasting

Renewable energy is used in many countries with numerous health, environmental and economic benefits. Forecasting renewable energy production helps to predict future energy produced in order to achieve demand and supply balance. This thesis focuses on a deep learning method for forecasting renewable energy production from photovoltaic panels using weather data to improve the predictions. The main focus of this thesis is forecasting photovoltaic production and investigating if the prediction performance could be improved by using numerical weather prediction over a 36-hour prediction horizon using a state-of-the-art probabilistic forecasting model, probabilistic forecasting with autoregressive recurrent networks. The forecasting result of photovoltaic production with weather data prediction shows a slight improvement. The forecasting performance of the selected model is also compared with benchmark models i.e., naïve predictor, seasonal naive predictor, and constant predictor.

PREFACE

First of all, I would like to thank to my supervisors, D.Sc. (Tech.) Toni Kuronen, D.Sc. (Tech.) Aleksei Romanenko, Associate Professor Arto Kaarna, Professor Lasse Lensu for their great support and guidance throughout my thesis. Their patient advice and availability of much needed supervision has been of immense help.

Secondly, I would like to thank Lappeenranta-Lahti University of Technology (LUT) for providing me the opportunity to study at their esteemed university. I would like to thank Professor Anders Lindfors from Finnish Meteorological Institute (FMI), for providing the data that I used in my thesis.

Words cannot express my deepest gratitude to my parents, Tesfai and Mihret, for their prayers, unrelenting support, love and inspiration in all aspects of my life. I am grateful to my sister and brother, Ruth, Ebenezer and Philemon, for always being there for me and their encouragement. Finally, I thank to all my friends and fellow classmates who helping me during this difficult time of my studies.

Lappeenranta, June 1, 2021

Suzan Yemane

CONTENTS

1 INTRODUCTION **7**
1.1 Background . . . 7
1.2 Objectives and delimitations . . . 8
1.3 Structure of the thesis . . . 9

2 FORECASTING RENEWABLE ENERGY **10**
2.1 Renewable energy production . . . 10
2.2 Weather impacts on renewable energy production . . . 11
2.3 Time series data and prediction problem . . . 12
2.3.1 Statistical time-series forecasting . . . 12
2.3.2 Machine learning based time-series forecasting . . . 14
2.3.3 Deep learning based time series forecasting . . . 16
2.4 Summary . . . 21

3 DEEP FORECASTING OF RENEWABLE ENERGY PRODUCTION **22**
3.1 Description of the probabilistic forecasting with autoregressive recurrent networks model . . . 22
3.1.1 Model architecture . . . 22
3.1.2 Likelihood . . . 23
3.1.3 Hyperparameters of the model . . . 24
3.2 Description of the baseline models . . . 25

4 EXPERIMENTS AND RESULTS **26**
4.1 Data Description . . . 26
4.1.1 Photovoltaic production data . . . 26
4.1.2 Weather data . . . 27
4.2 Data preparation . . . 29
4.3 Evaluation criteria . . . 31
4.4 Forecasting photovoltaic production . . . 32
4.4.1 Description of the training . . . 32
4.4.2 Testing hyperparameters . . . 33
4.4.3 Relevance of input variables . . . 35
4.4.4 Forecasting of photovoltaic production . . . 36
4.4.5 Result of benchmark models . . . 40

5 DISCUSSION **46**
5.1 Current study . . . 46
5.2 Future work . . . 47

6 CONCLUSION **48**

REFERENCES **49**

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARIMAX	ARIMA with Inclusion of eXogenous Variables
BNN	Bayesian Neural Network
CART	Classification and Regression Tree
CNN	Convolutional Neural Network
CS	Clear Sky
DeepAR	Probabilistic Forecasting with Autoregressive Recurrent Networks
DeepTCN	Probabilistic Forecasting with Temporal Convolutional Neural Network
DHI	Diffuse Horizontal Irradiance
DL	Deep Learning
DNI	Global Normal Irradiance
DT	Decision Trees
EU	European Union
FMI	Finnish Meteorological Institute
GHI	Global Horizontal Irradiance
GRU	Gated Recurrent Unit
KNN	K-Nearest Neighbor
LSTM	Long Short-Term Memory
MAAPE	Mean Arctangent Absolute Percentage Error
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
NN	Neural Network
NWP	Numerical Weather Prediction
PV	Photovoltaic
RE	Renewable Energy
RES	Renewable Energy Resources
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
S2S	Sequence to Sequence
SARIMA	Seasonal ARIMA
SMAPE	Symmetric Mean Absolute Percentage Error
SVM	Support Vector Machine

1 INTRODUCTION

1.1 Background

The energy demand is increasing day by day and the world has almost doubled its energy consumption on the last decades [1]. Fossil fuels are still the main source of energy. Fossil fuels are hydrocarbons, such as natural gas, coal, and petroleum. Fossil fuels take more years to be produced, and existing supplies are consumed considerably quicker than new ones are produced. Fossil fuels also affect the environment and are one of the main reasons for the global warming effect because they emit greenhouse gases. Renewable Energy (RE) is another alternative source of energy generated from renewable resources, which can be replenished in nature such as sunlight, rain, wind, tides, waves, biomass and geothermal heat. Using RE is important in countries looking for a clean and environment-friendly source of energy. Prediction of RE has become more important to the operation of electrical power systems.

The consumption of RE by the European Union (EU) has been increasing from past two decades [2]. This is possible due to the adoption of policy measurements and regulations that resulted in massive growth of RE consumption across the region. From 2005 to 2015 the growth almost doubled from 9% to 16.7%.

Forecasting RE production provides valuable information about the energy to be generated in the near future and its changes that helps in the operation of electrical grid and energy market planning. However, forecasting accurate RE production is a challenging task due to the irregular and complex nature of RE data and several factors affecting the generation. In addition, the accuracy of the forecasting is also affected by weather data and weather forecasts, and their related uncertainties [3]. To forecast the RE production several methods have been developed, including statistical methods, Machine Learning (ML) based approaches and their hybrids [1].

To improve the prediction, along with the historical data the Numerical Weather Prediction (NWP) can be used [4]. The data can be collected from a specific location i.e., meteorological station, and a Photovoltaic (PV) panel or a wind turbine. When the weather information are collected from the same location where the prediction are performed, the forecasting performance can be improved. For instance, as shown in Figure 1, due to the increase in wind speed and its direction at wind turbine 1, after certain hours its speed profile will get seen near the place of wind turbine 3. Similarly, the value of solar irradi-

ance might get lower within some time when the clouds move from region one to region two or region three.

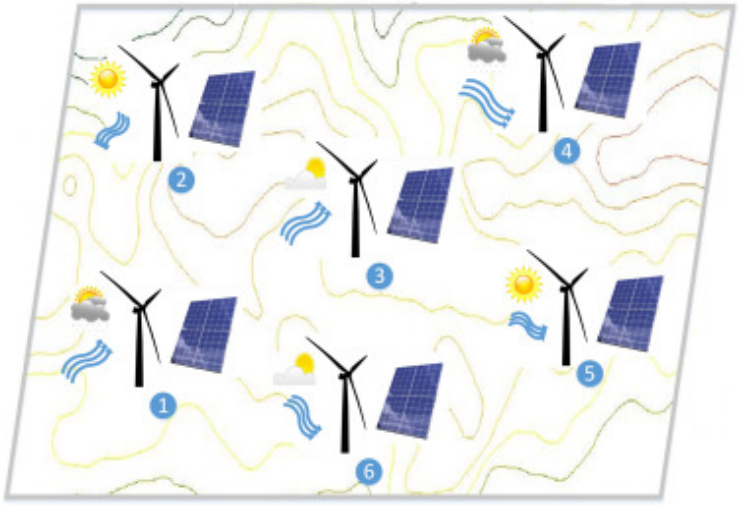


Figure 1. A map with distributed wind and solar power plants with different weather conditions [4].

In recent years, Deep Learning (DL) has become an essential and promising part of time series forecasting due to its capability for learning the characteristics of the nonlinear features and structures in data [5, 6]. This thesis focuses on the DL forecasting method for PV production using numerical weather prediction data. The inputs of the model used are time series of historical PV production data and weather data. The data used in this thesis have been collected from the Finnish Meteorological Institute (FMI) station located in Helsinki, Finland.

1.2 Objectives and delimitations

The main objective of this thesis is to provide probabilistic forecasting for PV production. The weather data is used in addition to the historical data of PV production to improve the forecasting performance. The goal is to perform quantitative evaluation of the forecasting performance over the prediction results and to compare these results with other benchmark models.

The objectives are formulated as follows:

- Study and develop deep learning method for forecasting PV production based on PV generation history data and weather information for 36 hours and one hour prediction horizon.
- Study the relevance of input variables.
- Quantitatively evaluate the model performance against a benchmark.

1.3 Structure of the thesis

The rest of the thesis is organized as follows: Chapter 2 presents literature review related to the forecasting of RE production. The proposed method for forecasting PV production is discussed in Chapter 3. Chapter 4 includes the experiments and the results. Chapter 5 presents the discussion of the research work made and ideas for future work. The thesis work is briefly concluded in Chapter 6.

2 FORECASTING RENEWABLE ENERGY

2.1 Renewable energy production

RE refers to the energy produced from natural resources like sunlight, rain, tides, waves, wind, etc. RE has more advantages when it is compared to fossil fuels, such as it is infinite supply, clean and less carbon which is useful in the protection of the environment [1]. It also helps in reducing air pollution, soil pollution, greenhouse effect, and release of waste gas and water liquid. Because of these advantages, RE one of the energy sources that shows rapid growth in recent years. Among different RE sources, the use of wind energy is clean, indispensable and inexpensive.

Although RE is considered to be another alternative to fossil fuels because it is environmental friendly and protecting natural resources, it has also uncertainty, which influence the stability and reliability of the electrical grid system [1]. At the same time, to use RE a vast number of power electronics are required, which in turn reduces the stability of the system [2].

Finland is among the leading EU countries using Renewable Energy Resources (RES). According to Motiva [7], due to the long cold winter months in Finland, there is high energy consumption for lighting and heating. Thus, reliable and uninterrupted sources of an electricity and heat production are essential. RE is used as a source of energy to reduce the dependency on the imported energy and due to its positive impact on the regional and national economy. Some of RES are forest chips, wind power, heat pumps and solar energy.

Among all RES, wind energy and solar energy are common resources of energy resource because they are inexhaustible and environmental-friendly. However, they are considered as a non-guaranteed supply because of their frequent variation. Thus, such RESs are more challenging in both providing a stable and safe energy supply and in integrating into an electricity grids [8]. The inherent character of these RESs can affect the prediction accuracy. RE production forecasting is important for considering related uncertainties of the data, to the operation, planning, and management of electrical grid and power energy systems [8]. Accurate prediction of RE production ensures better safety of the electrical grid, optimal operation of power system and balancing the demand and supply of energy production. However, forecasting RE production remains challenging due its intermittent and stochastic nature.

2.2 Weather impacts on renewable energy production

Weather is the change in atmospheric states in a short period of time. It is usually presented in terms of temperature, cloudiness, humidity, precipitation, brightness, wind, visibility and atmospheric pressure. RES are dependent on weather conditions which adds uncertainty and makes them harder to predict [4]. Weather parameters such as wind speed, solar irradiance, humidity, are crucial to the renewable energy production.

According to [9], in PV power production, the most influential weather parameter is solar resource. The reliability of solar resource measurement requires various tools and techniques and this should be characterized to the best in order to increase the PV power output. The solar irradiance that reaches to the top of atmosphere is called the total sky irradiance. There are different components of solar resources:

- **Global Normal Irradiance (DNI)**: the amount of radiation reaches on the flat surface. Its orientation is always normal (perpendicular) to the ray.
- **Diffuse Horizontal Irradiance (DHI)**: the solar radiation can be partially scattered back to the surface due to the presence of clouds. Thus, total radiation comes from the combination of the partial direct normal irradiance, the amount of radiation scattered by air molecules and particles is called the diffuse radiation. Measuring the diffuse radiation on flat surface with horizontal orientation presents the diffuse horizontal irradiance.
- **Global Horizontal Irradiance (GHI)**: the total amount of radiation on horizontal surface of the earth.

The relationship between these different components can be expressed as:

$$GHI = DNI \times \cos(Z) + DHI \tag{1}$$

where Z is the angle between the current position of the ray and vertical direction to the surface. The solar energy resources are affected by various factors of atmospheric composition but mainly their are affected by the cloud coverage [10].

To obtain the solar resource data, there are mainly two approaches: (1) To measure directly using various tools and techniques (2) estimating or modeling the solar irradiance using data like images of the cloud, characteristics detected from weather satellites, NWP. The instrument type used to measure the solar resource data used in this thesis are known as Pyranometer [11]. The Solis-Heliosat model is used to forecast GHI for southern Finland based on the Clear Sky (CS) model Pvlib Solis and the weather satellite data-Heliosat

(SARAH-2). The satellite-based data for solar radiation data can be retrieved using different models [12].

Generally, the weather variables used for predicting solar power generation are temperature, relative humidity, dew point, sky cover and wind. Amount of sunlight and air temperature are most influential variables that could affect PV power outputs [11, 13].

2.3 Time series data and prediction problem

Time series data is an ordered sequence of observations recorded through time. The intervals among these data points are not always the same, they might be different. Time series data contain arbitrary information and can be uni-variate or multivariate. The concept of time series forecasting is to estimate how far the sequence of observations will continue. The main tasks of time series forecasting are to understand the structure and underlying features of the observed data.

Time series forecasting is a prediction of future observation based on past observations [14]. Let $Y = \{y_1, ..., y_n\}$, be a time-series. The predicted future data points of $Y = y_n + h$, where h denotes the forecasting length i.e., the prediction horizon (see Figure 2). The forecasting horizon could be single step ahead or multiple-step. However, increasing forecasting horizon might increase the uncertainty of the forecasting results.

In time-series forecasting, various factors affect the prediction and many time series data are complex. Solving these problems is not straightforward using statistical and physical models. To solve time series forecasting problem several approaches are introduced: statistical models [15], ML based methods [16–20] and DL based methods [21–31].

2.3.1 Statistical time-series forecasting

The conventional statistical forecasting models impose the use of statistics based on historical data to predict the output. Their aim is to uncover the relation between a one or more variables. It is easy to estimate a predictive value and very effective for short-term prediction [1].

Autoregressive (AR) model is one of the baseline technique used for forecasting time series [15]. The prediction is made based on the analysis of statistical properties of time-

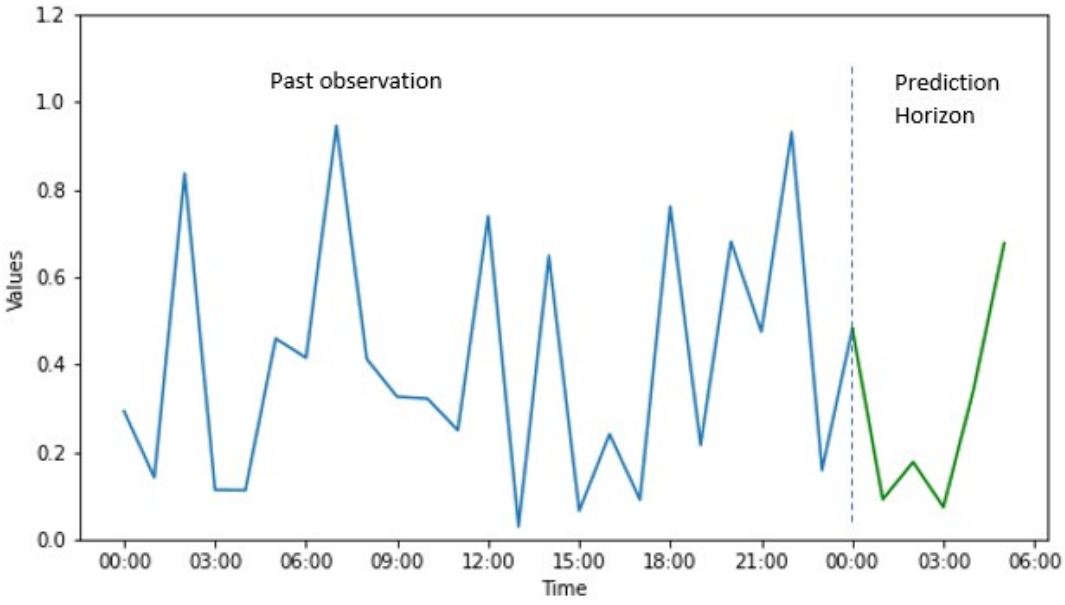


Figure 2. Illustration of time series forecasting with past observation and prediction horizon.

series. The training data points needs to be stationary which means the observations at one point should be similar to the same observation of the recent pasts. Basic AR models predict the future value based on the past observation.

Autoregressive Integrated Moving Average (ARIMA) model is similar to that of AR model. The idea of the model is to make time-series stationary by differencing raw observations [32]. The ARIMA model uses the dependent relationship between observations and the residual errors. A standard notation used for $ARIMA(p, q, d)$, where p represents order of AR, d the number of differencing, and q order of moving average. Results have shown that the ARIMA model is suitable for forecasting. The result of ARIMA is a linear equation which can be expressed as:

$$\hat{d} = C + \sum_{i=1}^{p} \phi X_{t-i} - \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \tag{2}$$

where $\hat{Y}$ is the predicted result, t time step, X the input data. ϕ is the coefficient of the AR model and θ is the coefficient of moving average, C is a constant value and ϵ is the forecasting error.

Seasonal ARIMA (SARIMA) is similar to ARIMA model except an additional condition

is added for regular events and trend [15]. They are denoted by $SARIMA(p,d,q)(P,D,Q)_s$ where p, d and q are trend autoregression, trend difference and trend moving average, which are related to the trend of original data. The P, D, Q are the seasonal autoregressive, differencing and moving average order and s is the number of time steps at one seasonal period.

ARIMA with Inclusion of eXogenous Variables (ARIMAX) considers the influence of external parameters other than the past history data on the prediction [15]. The past b, which corresponds to the time series is added as a linear combination to the standard ARIMA models. ARIMAX model is denoted by ARIMAX(p,d,q,b). Both SARIMA and ARIMA are combined to make SARIMAX.

However, the statistical approaches developed for forecasting RE are expressed as a linear regression models. These models have limitations on solving prediction problems with long term forecasting. [1].

2.3.2 Machine learning based time-series forecasting

Artificial Neural Network (ANN) is a network of artificial neurons designed to simulate the way the human brain works. The neurons are organized in layers [19]. The ANN consists of the input layer (predictors), and the output layers (forecasts) (see Fig 3). ANN can have one layer or more hidden layers. In ANN with more than one hidden layers i.e., multilayer feed-forward network, the outputs of the neurons in one layer serve as an input to the next layer. The output of z_i, of neuron i in the hidden layer can be represented by:

$$z_j = b_j + \sum_{i=1}^{N} w_{i,j} x_i \tag{3}$$

where N is the number of input neurons, $w_{i,j}$ it the weight of connection between neurons, x_i is the input neuron, and b_j is the bias to the neuron. The output z_i is modified by the activation function. The activation function aims to impose non-linearity on the neural network which allows learning more complex patterns. The most commonly used activation function is sigmoid (see Eq. 4).

$$s(z_i) = \frac{1}{1 + e^{-z_i}} \tag{4}$$

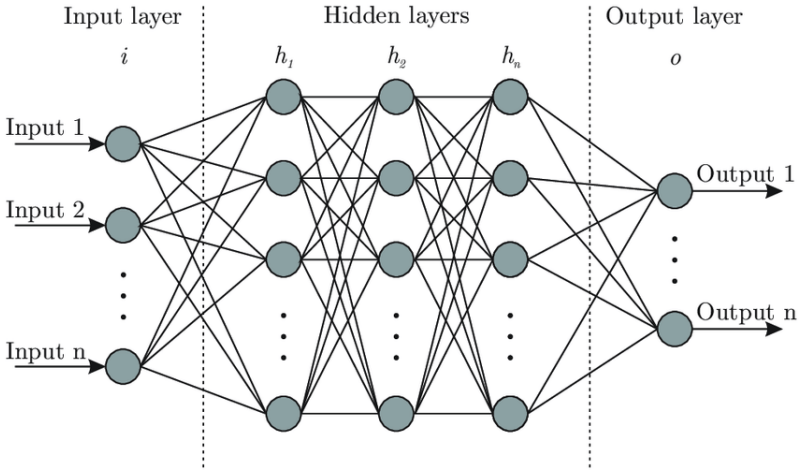


Figure 3. ANN Architecture [20]

Bayesian Neural Network (BNN) is a Neural Network (NN), developed based on a Bayesian probabilistic formulation [16]. The aim of BNN is to relate the strength of the NN with the concept of stochastic modeling and probabilistic models. The main concept of BNN is to enforce the network weights to follow a prior distribution. During the prediction time, the statistical model generates the posterior distribution of the weights produces probabilistic guarantees on the predictions.

K-Nearest Neighbor (KNN) is a method most commonly used for classification and regression-based. The prediction is based on the K-nearest neighbors training points of the point to be predicted [16]. To compute the distance between that given query point and other training points, different distance metrics can be used. Euclidean distance is usually preferred when there is low dimensionality in the data. Let $\jmath(x)$ be the set of K-nearest neighbors of the query point x. KNN prediction $\hat{y}$ is defined as:

$$\hat{y} = \frac{1}{K} \sum_{m \in \jmath(x)} y_m \tag{5}$$

where y_m is the predicted output based on the training data point x_m. The value K is the key parameter in this method. A value of K should be large enough to get a smooth fit and lower variance. The number of similar input data resulted in similar output and the accuracy of the forecasting depends on it.

Decision Trees (DT) is a ML based method that can be applied for time series forecasting problems [15]. The structure of DT starts with a root node and keep splitting to non-leaf nodes(decision nodes). A sequence of tests based on specific conditions along the decision nodes is performed starting from the root node until it reaches a leaf node. The

predicted output is made at the leaf node according to the local model associated with it. There are different variants of DTs methods. The most common method used for RE forecasting is Classification and Regression Tree (CART). The main advantage of CART is in eliminating the variables that do not contribute to the prediction and it is not complex [16].

Support Vector Machine (SVM) is a model, developed to solve classification and regression problems but it can also be applied for time series prediction [18]. The support vector regression method is used for time series prediction as it is autoregression in time [33]. The main idea of support vector regression is transformation of the data into a space of high dimensional feature using a nonlinear kernel function. Comparing to ANN, SVM performs in low computational complexity and achieves good convergence results. However, the main challenge of SVM is the accuracy depends on the on the value of the kernel parameters [17].

2.3.3 Deep learning based time series forecasting

The main draw-back of statistical-based method and basic ML approaches are the dependencies on predefined non-linear parameters and long-term forecasting. In the past recent years, DL, the branch of ML, shows more promising outcomes due to three major features i.e., extracting features from complex patterns, unsupervised learning, and training on large dataset compared to other models [1]. The use of DL for time series forecasting overcomes the limitations of statistical and traditional ML models. DL based models they show better performance in long term forecasting.

Convolutional Neural Network (CNN) model it is most commonly developed for classification problems such as object recognition, speech recognition, and natural language processing [22]. A typical CNN consists of an input layer, convolution layer, fully connected layer, and output layer (see Fig. 4). The convolution layer applies kernels or filters to extract important features from the input. This structure allows the filters to be learned that are able to discover specific features in the input data. The activation function is applied to present non-linearity into the CNN model. The most popular activation functions are Rectified Linear Unit (Relu), softmax, etc. [5]

Recurrent Neural Network (RNN), is a variant of DL model designed for modelling sequential data i.e., time-series data. It can be used for speech recognition, language translation, and in predictions problems [6]. RNNs use the neurons internal state to process the sequences of inputs. RNN provides interdependencies among the data of input sam-

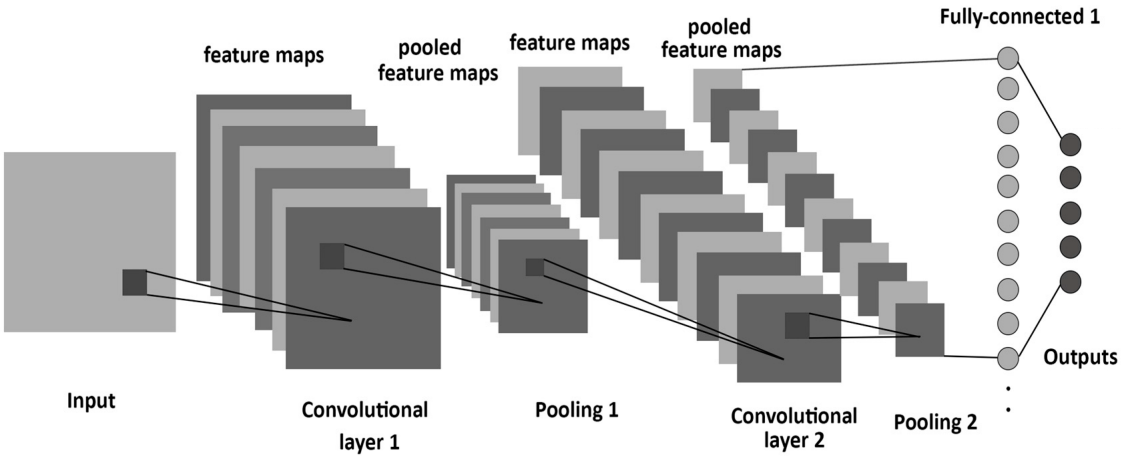


Figure 4. The structure of a CNN [5].

ples. The output produced from the previous step affects the parameters of the current input. However, RNNs suffer from vanishing gradient problems, which means, they are incapable of handling long term dependencies [23].

In [24], Hochreiter and Schmidhuber developed a method called Long Short-Term Memory (LSTM) which is type of RNN, in order to address the problem of short-term memory. By using mechanisms called gates, the flow of information is regulated. The cell state, input, output, forget gates are the main concepts of LSTMs. It allows the information to be stored within LSTM cell throughout the process. The information that goes to the cell state is added or removed by the forget gate and adjusted by the input gate. The structure of LSTM cell is presented in Figure 5. The forget gate operation can be defined as:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{6}$$

where t, is the time step, f_t is forget gate at t, $W_f$is the weight between forget gate and input gate, x_t is the input, h_{t-1} is previous state, b_f is a bias vector, σ is a sigmoid activation function. The input gate updates the cell status and can be defined as:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{7}$$

$$\hat{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{8}$$

where i_t is the input gate at time t, b_i is a bias vector, $\tanh$ is the hyperbolic tangent activation function and $\hat{C}_t$ is the vector generated by $\tanh$. The network gets information

from the forget gate and input gate and saves the cell state information in C_t.

$$C_t = f_t \times C_{t-1} + i_t \times \hat{C}_t \tag{9}$$

The value of the next hidden state is determined by the output gate. It can be defined as:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{10}$$

$$h_t = o_t \times \tanh(C_t) \tag{11}$$

where o_t is the output gate at t ,b_o is a bias vector and h_t is LSTM output.

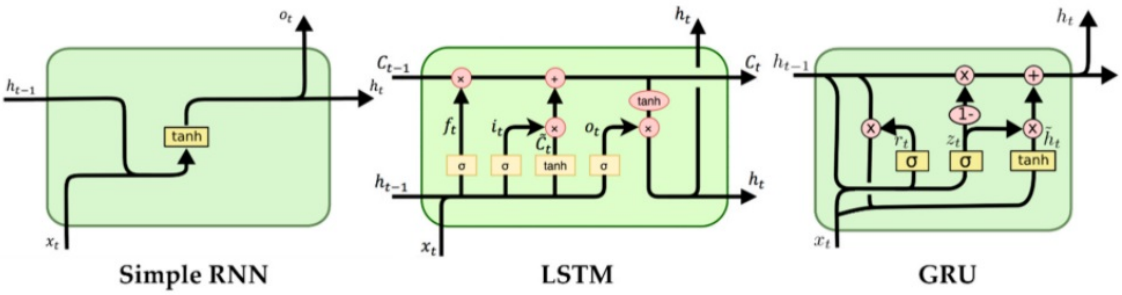


Figure 5. The different inner structures of RNN cells [34].

Gated Recurrent Unit (GRU) is another evolution of RNNs which is similar to LSTM [25]. To address the problem of vanishing gradient, GRU uses the update and reset gate. The update gate is the same to the forget and input gate of the LSTM. It decides on keeping or forgetting information. Both LSTM and GRU perform better and ensures that important features will not be vanished during long-term transmissions by using various gates. However, GRU performed well only on small datasets [26]. Figure 5 presents the inner structure of RNN, LSTM and GRU.

In [27], LSTM method is proposed for PV power generation forecasting. The historical data of solar irradiance is combined with the sky forecast, to produce synthetic weather forecasts using K-means algorithm. The study shows that the accuracy is improved by 33% using the synthetic forecast comparing with the hourly sky forecast. The forecast accuracy is also improved up to 44.6% with daily sky forecast. The proposed method is evaluated using RNN, and extreme learning machine and a generalized regression neural network. However, LSTM shows higher forecasting accuracy.

The LSTM RNNs are often used as Sequence to Sequence (S2S) models [28]. This model is widely used in language translation. It is used in a situation when the length of the input and the output is different. Kharlova et al. [35] proposed a S2S model with attention mechanism to predict PV power generation. The proposed model consists of an encoder and decoder LSTM blocks. The encoder produces fixed dimensional vector from the input sequence to encapsulate valuable information and transfers that vector to the decoder to produce the output sequence see Figure 6. This helps the decoder to obtain only important features.

However, the decoder can access only the earlier state of the encoder. The attention mechanism is introduced to address this longer-range dependency problem. The aim of attention mechanism is to impose unequal weights to the features after their importance are evaluated. Then the important features get more attention from the network. In [35], attention mechanism is used to extract the information on the contribution of the inputs to the model in order to increase the forecasting performance.

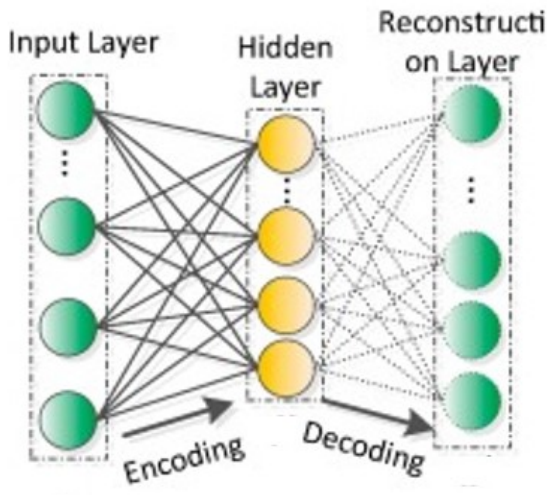


Figure 6. The basic unit of an auto-encoder [1].

The existing DL methods for time series forecasting can be categorized under deterministic and probabilistic methods [1]. The RE data, due to the irregular and complex nature, contain more or less uncertainty. The deterministic forecasts are providing less information regarding uncertainty. Probabilistic forecasting provides additional information on the level of uncertainty related to predictions.

The main aim of a probabilistic forecast is to make a probability distribution over the

predictive results [18]. There are two approaches to build a Probability Density Function forecasting (PDF), parametric assume underlying statistical distributions in the data and nonparametric approach do not rely on any distribution.

Chen et al. [29], presented a model called Probabilistic Forecasting with Temporal Convolutional Neural Network (DeepTCN) for a probabilistic forecasting based on CNN. The proposed model can work under the parametric and nonparametric approach to evaluate probability density. The model has designed stacked residual block, which has the ability to learn complex features throughout time series data from both past observation and exogenous variables.

Salinas et al. [30] developed a DL based approach Probabilistic Forecasting with Autoregressive Recurrent Networks (DeepAR) to produce accurate probabilistic forecasts. The method is based on training an autoregressive RNN model on a multiple number of time series. The output produced from the network it not the predictive results but the parameters for the probabilistic function. Different types of probabilistic functions can be applied. However, two different models are considered i.e., Gaussian likelihood and negative-binomial likelihood models. The samples generated from the likelihood models fed back as an input to the next point until the prediction ranges ends. Repeating sampling multiple times produces probability distribution. This method can forecast time series with little or no history available. This method is implemented on different real-world datasets and generates accurate probabilistic forecasts [30].

Mashlakov et al. [31] studied the implementation of new global DL based models for deterministic and probabilistic multivariate forecasting and whether these models can be applied for energy forecasting. Different DL models which show good performance forecasting multivariate time series such as DeepAR [30], DeepTCN [29], Long and Short Term Temporal Patterns with Deep Neural Networks (LSTNet) [36], and Self-Attention Network for Multivariate Time Series Forecasting (DSANet) [37], were selected for the empirical comparison. It also shows that by using these global DL models, it is possible to achieve satisfactory levels of forecast accuracy, uncertainty risk, and the impact of exogenous covariates in the forecasting of energy systems for short-term forecasts. However, to improve the forecast accuracy of the intermittent wind data additional exogenous information is required.

2.4 Summary

In this section the literature review made on Chapter2 is summarized. Several approaches have been discussed for forecasting time series data. Statistical methods consist of AR models such as ARIMA, SARIMA, have been discussed as the basis for time series forecasting. However, the main drawbacks of these approaches are their limitation for short-term forecasting. In many recent studies, the DL approaches have been considered as a solution to the drawbacks of statistical methods. In addition, LSTM RNN is the most commonly method used for forecasting renewable energy production. The combination models of the global DL are also one of the main research interests in order to increase the model performance. Moreover, the DL models are also used together with probability models for probabilistic forecasting of time series e.g., DeepAR.

3 DEEP FORECASTING OF RENEWABLE ENERGY PRODUCTION

This section describes the model implemented in this thesis. Based on the literature review in Section 2.3.3 DeepAR, found to be the appropriate model for deep forecasting of renewable energy production. DeepAR has shown good performance in probabilistic forecasting problems.

3.1 Description of the probabilistic forecasting with autoregressive recurrent networks model

DeepAR is a method for generating accurate probabilistic predictions based on training autoregressive recursive network models on multiple data sets [30]. It contains RNN using either LSTM or GRU cells, which integrates a likelihood function like Gaussian/negative binomial, to make the probabilistic forecasting.

DeepAR learns the correlation characteristics of multiple time series by combining similar time series. Apart from input data to the model, the covariates time series i.e., hour of the day, week of the year, etc. are also generated automatically. These covariates time series allow the DeepAR model to learn seasonal pattern and plays a role in the improvement of the forecasting performance.

In DeepAR, the results obtained from the network is not the actual predicted value, rather the parameters which are used for the probability model. The predicted value is sampled from the probability model see Figure 7.

3.1.1 Model architecture

The value of the time series i at time t is denoted by $z_{i,t}$. The goal is it to model the conditional distribution $P(z_{i,to:T}|z_{i,1:to-1}, x_{i,1:T})$ of the feature values $[z_{i,t0}, z_{i,t0+1}, ..., z_{i,T}] := z_{i,t0:T}$ using a past value $[z_{i,1}, ..., z_{i,to-2}, z_{i,to-1}] := z_{i,1:to-1}$, where the covariates $x_{i,1:T}$ are assumed to be known. The time ranges $[1 : t0 - 1]$ and $[t0 : T]$ are respectively the conditioning window range (context length) and the prediction window range (prediction length). The summary of the model is presented in Figure 7.

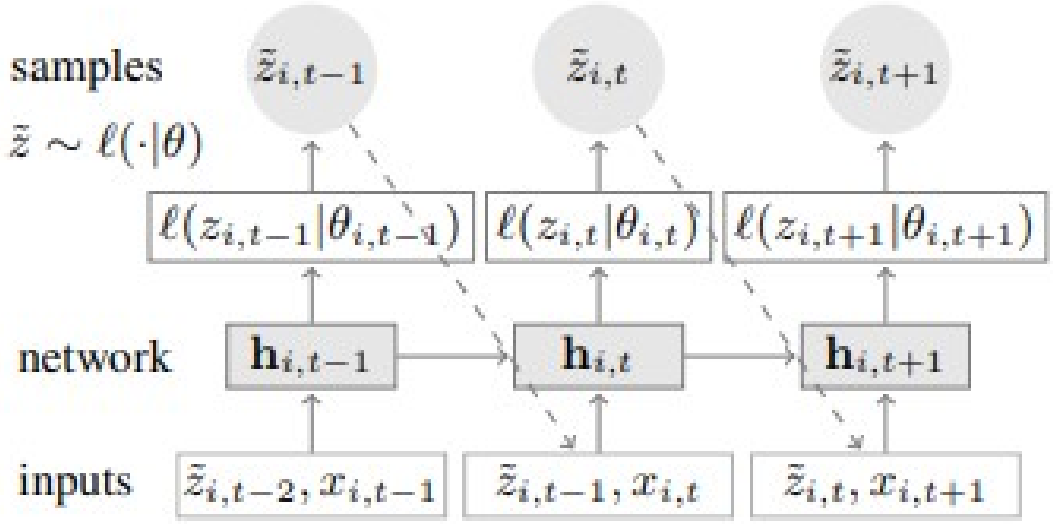


Figure 7. Summary of the model training: At time t, the inputs to the network are the covariates $x_{i,t}$, value at the previous time $z_{i,t-1}$, and the previous network output $\mathbf{h}_{i,t-1}$. The network output $\mathbf{h}_{i,t} = h(\mathbf{h}_{i,t-1}, z_{i,t-1}, x_{i,t}, \Theta)$ is then used to compute the parameters $\theta_{i,t} = \theta(h_{i,t}, \Theta)$ of the likelihood $\ell(z \mid \theta)$, which is used for training the model parameters [30].

The model distribution $Q_\Theta(z_{i,t0:T}|z_{i,1:t0-1}, x_{i,1:T})$ is assumed to contain a product of likelihood factors:

$$Q_\Theta(z_{i,t0:T}|z_{i,1:t0-1}, x_{i,1:T}) = \prod_{t=to}^{T} Q_\Theta(z_{i,t}|z_{i,1:t-1}, x_{i,1:T}) = \prod_{T}^{t=to} l(z_{i,t}|\theta(h_{i,t}, \Theta)) \quad (12)$$

with the output of an auto regressive recurrent network $\mathbf{h}_{i,t}$, $\mathbf{h}_{i,t} = h(\mathbf{h}_{i,t-1}, z_{i,t-1}, x_{i,t}, \Theta)$. Since the models is an auto regressive this $\mathbf{h}_{i,t}$ output will be an input for the next step $\mathbf{h}_{i,t+1}$, where h is a LSTM cell, and the likelihood $\ell(z_{i,t} \mid \theta(h_{i,t}))$ is a fixed distribution parametrized by a function $\theta(h_{i,t}, \Theta)$.

3.1.2 Likelihood

The likelihood $\ell(z \mid \theta)$ should reflect the statistical properties of the data on its best. In this thesis, the negative-binomial likelihood is chosen to represent the positive data. For the next input the network predicts the parameters for θ.

$$\ell_{\mathrm{NB}}(z \mid \mu, \alpha) = \frac{\Gamma(z + \frac{1}{\alpha})}{\Gamma(z+1)\Gamma(\frac{1}{\alpha})}(\frac{1}{1+\alpha\mu})^{\frac{1}{\alpha}}(\frac{\alpha\mu}{1+\alpha\mu})^{z} \tag{13}$$

$$\mu(\mathbf{h}_{i,t}) = \log(1 + \exp(\mathbf{W}_{\mu}^{T}\mathbf{h}_{i,t} + b_{\mu})) \tag{14}$$

$$\alpha(\mathbf{h}_{i,t}) = \log(1 + \exp(\mathbf{W}_{\mu}^{T}\mathbf{h}_{i,t} + b_{\alpha})) \tag{15}$$

To ensure the data is positive, both the parameters (α and the μ) are attained from the network output with softplus activation. The shape parameter α and the μ are used to parameterize the negative binomial distribution, where the variance is scaled relative to the mean, i.e., $Var[z] = \mu + \mu^2\alpha$ by the shape of the parameter α.

3.1.3 Hyperparameters of the model

The accuracy of time series forecasting depends on the optimal value of the hyperparameters of the model. In DeepAR model two hyperparameters are crucial, which are fixed hyperparameters of the problem:

- **Context length**: this indicates the number of time points considered as input size for the model.
- **Prediction length**: the number of time points considered for the model to predict.

Since the model generates a probabilistic forecasting there is another important parameter called likelihood which is described in Section 3.1.2. The type of cell, the number of the hidden layers of the RNN and the number of cells used in each hidden layer is also considered as hyperparameter to be tuned. To determine the values pf these parameters several tests are performed see Section 4.4.2.

In addition to the above mentioned hyperparameters, DeepAR has also learning rate, number of epochs, batch size, early stop patience parameters. The learning rate controls how quickly the model is learning. Epochs is the number of times that passes over the training data. Patience is used for learning rate scheduler which reduces the learning rate if there is no improvement in the training. The batch size is the number of training instances used in one iteration.

The DeepAR model has also parameters associated with covariates. These covariates can be used to provide additional information to the input of the model to improve the forecasting performance. In this thesis two additional covariates are used one is a time

feature i.e., minute of hour, hour of day, day of month, day of week and the second one is a dynamic feature (weather data).

3.2 Description of the baseline models

To evaluate the forecasting performance obtained from the DeepAR model, some baseline models are used. These models are as benchmarks in this thesis. These baseline models used are as follows:

- Naïve forecasting.
- Seasonal Naive Predictor.
- Constant Predictor.

Naïve model predicts the value based on the past history data. Each forecast values are set to the previous value observed (see Eq. 16). The parameters used in this model are the past history data, prediction horizon h and frequency of the data.

$$\hat{y}(t+h|t) = yt \tag{16}$$

Seasonal Naive predictor is similar to naïve model which considers the seasonality of the data during prediction. In this case, each of the forecast value is equal to the last observed data of the same season. In Seasonal Naive predictor, for each time series y, it produces a forecast:

$$\hat{y}(T+k) = y(T+k-h) \tag{17}$$

where T is the prediction horizon, $k = 0, ..., T-1$, and h is season length.

A constant predictor is a predictor that always produces constant value for all forecast values. The model takes the constant value C and the prediction length as its parameters, then for each time series y, it produces a forecast of same constant value C.

$$\hat{y}(t+h|t) = C. \tag{18}$$

4 EXPERIMENTS AND RESULTS

4.1 Data Description

The data used in this thesis are PV production data and weather data. The target time series is the PV production and to observe whether using weather data can improve the prediction result the weather data is also used. The detailed description of the data is presented below.

4.1.1 Photovoltaic production data

The PV production data have been collected from the FMI site located in Helsinki. The data is provided as one dataset,it ranges from $2015/08/26 - 2020/12/31$ year. The PV production data consists of four historical PV power output time series. The frequency of the time series is one minute. The values at every time step corresponds to the PV in W [38].

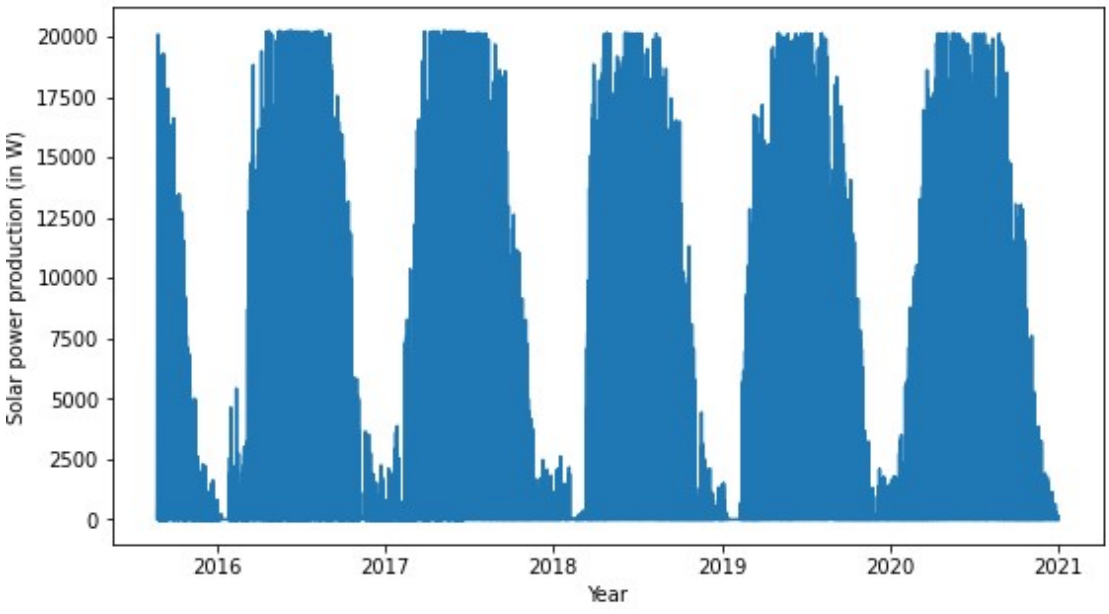


Figure 8. The whole PV production data.

At the beginning of the data there are some missing value. The used data is set to start from $2015/12/21$ in order to remove the range that has missing values. During dark period of winter, the values of all the PV production data are almost zeros which indicates no power is generated. However, during summer time, the PV production data reaches to its maximum value. The PV production data of the whole data, different season and with different frequency are illustrated in Figure. 8-14.

There are a set of 23 PV panels located in Southern Finland [38]. The sites are indexed by ID number. The PV production data used in this thesis is as site ID 2 data used in [38]. The data are sampled 3600 per hour. The range of the PV is 5 and 20 kWh which provide PV power output data measured by inverter.

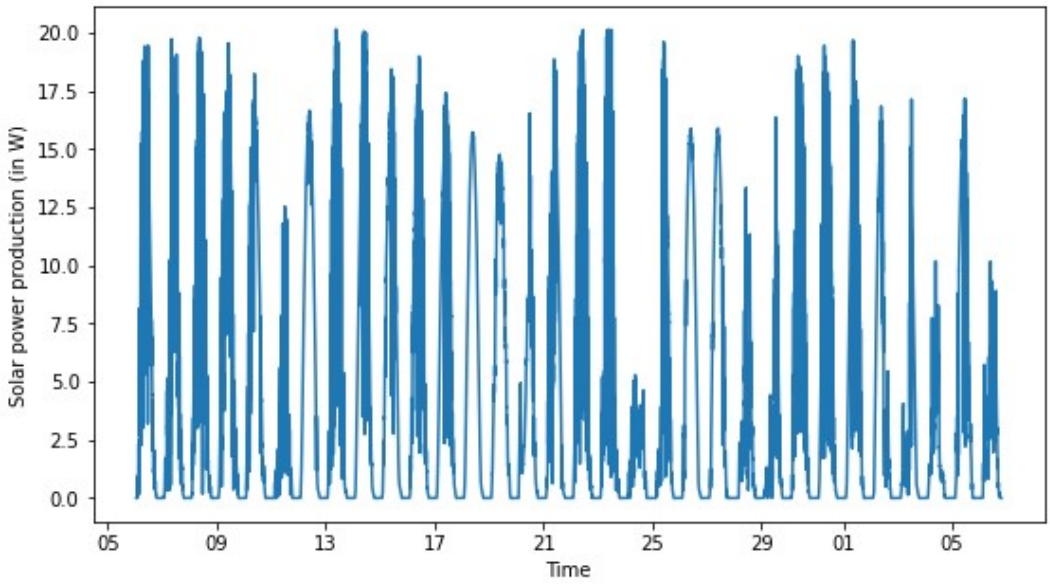


Figure 9. The PV production data of one month.

4.1.2 Weather data

The NWP data are collected from FMI station Helsinki, Finland. The data consists of 10 sub-time series with one minute and ten-minute resolution. The list of the the sub-time series is listed below Table 2.

These variables are used to order to improve the forecasting accuracy of PV production. The variables used for forecasting PV production is depending on its importance which

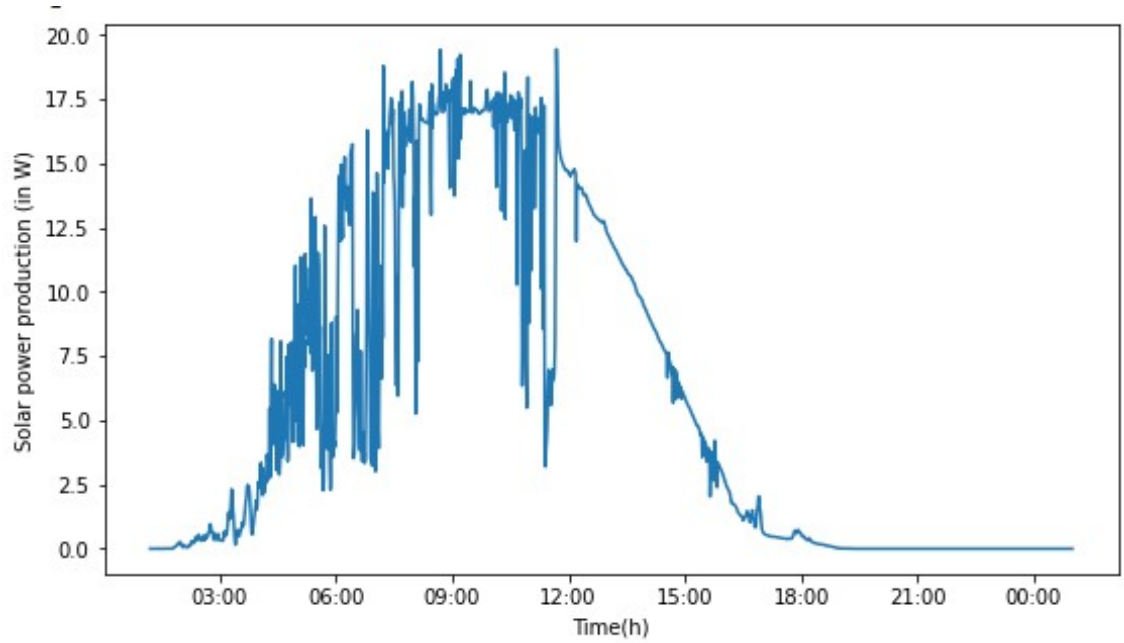


Figure 10. The PV production data of one day with frequency of one minute.

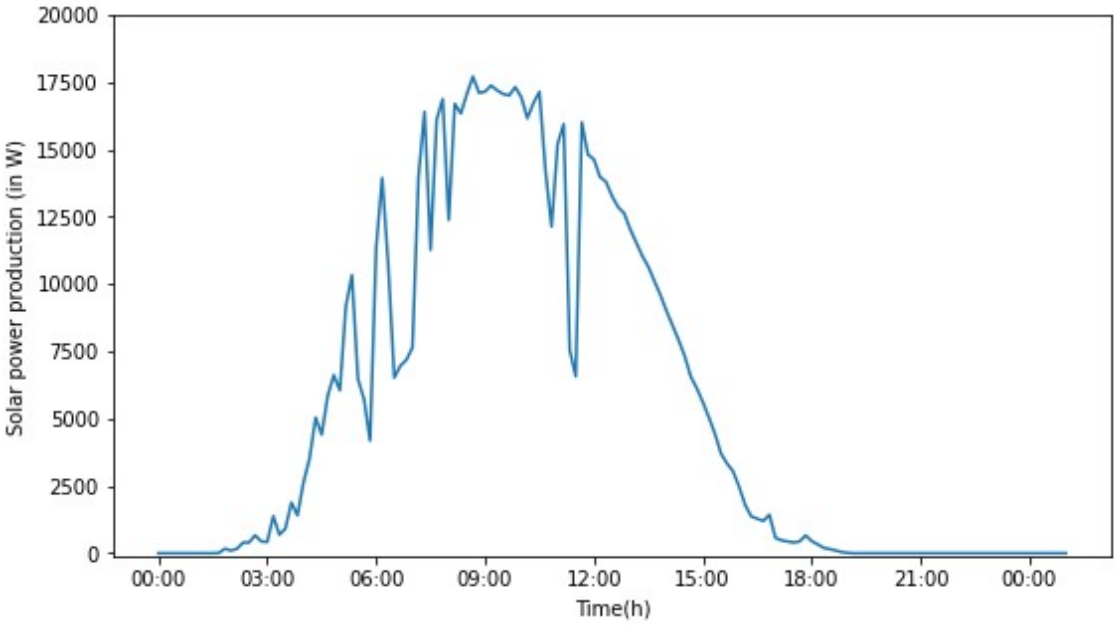


Figure 11. The PV production data of one day with frequency of ten minutes.

is studied in Section 4.4.3 and used the combination of these variables as an additional feature to the input of the model.

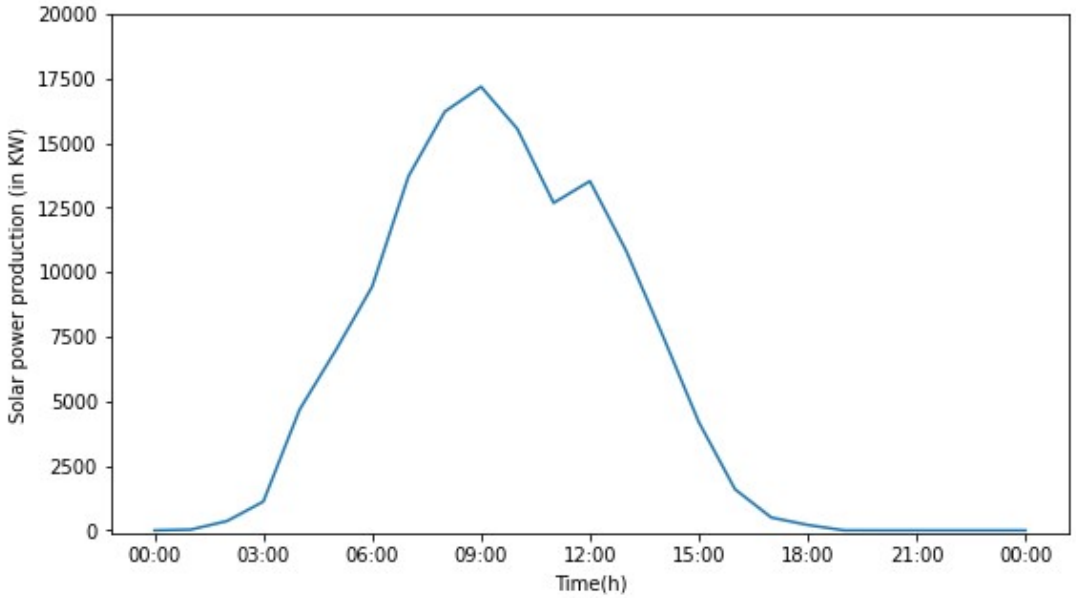


Figure 12. The PV production data of one day with frequency of one hour.

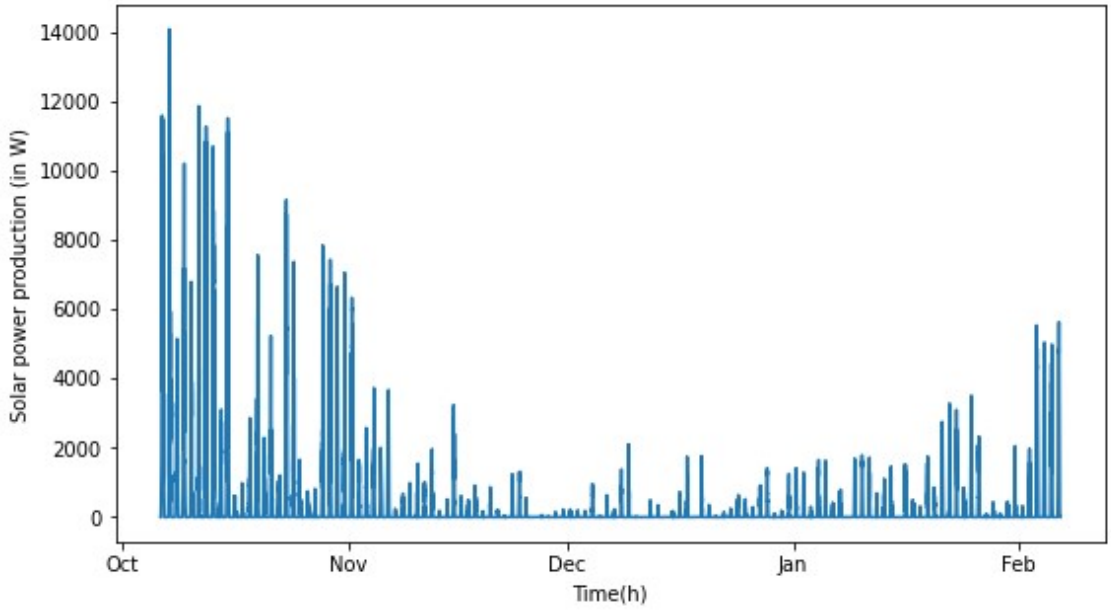


Figure 13. The PV production data during winter time.

4.2 Data preparation

The PV output time series is used as a target from the PV production data and the weather prediction time series are used as additional features depending on its importance. The inputs used for training the model for both the experiments with historical data only and when weather data is added are illustrated in Table 3.

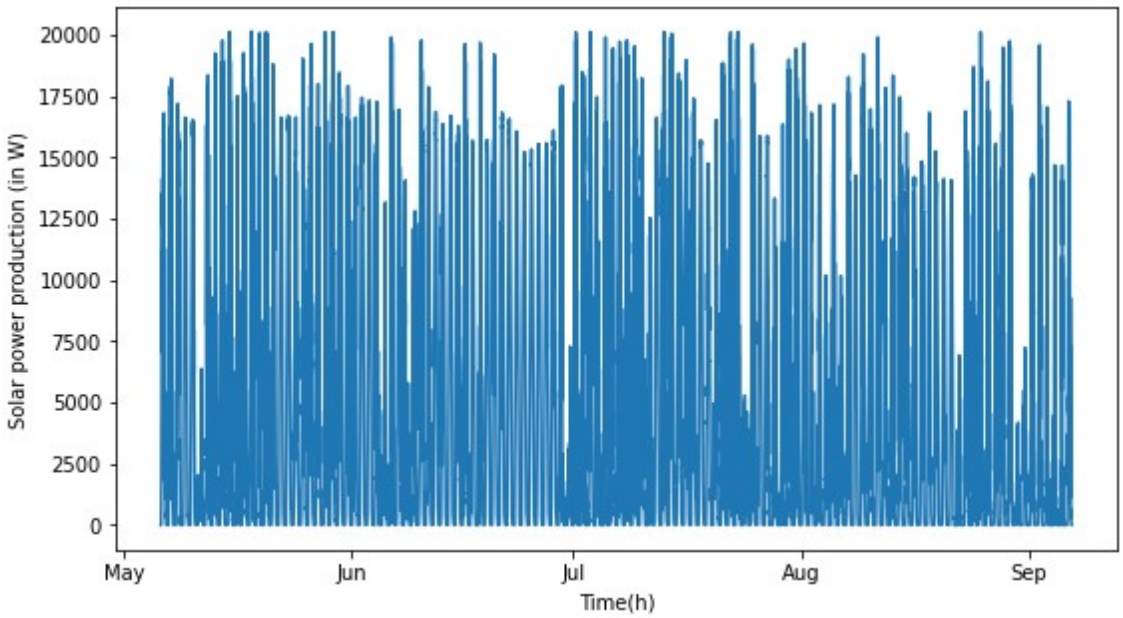


Figure 14. The PV production data during summer time.

Table 2. Weather data variables.

Variables	Values
Global radiation on horizontal surface	1 min
Diffuse radiation on horizontal surface	1 min
Direct normal irradiance (perpendicular surface)	1 min
Global radiation on inclined PV plane-of-array	1 min
Air temperature on roof, vicinity of PV arrays	1 min
PV module temperature (North Eastern corner)	1 min
PV module temperature, (South Western corner)	1 min
Air temperature at close by weather station	10 min
Wind speed at close by weather station	10 min
Wind speed at close by weather station	10 min

Table 3. Input data of the model.

Input records	Target: PV	Time feature	NWP
Historical data only	Yes	Yes	No
With weather data	Yes	Yes	Yes

The original frequency of the data is ranging from one minute to ten minutes. The frequency of the time series data is re-sampled to one-hour frequency before feeding to the model training, to avoid unnecessary noise from the data. For the weather the data the two minutes and ten minutes sub-time (see Table 2) series are first interpolated before re-

sampled to one hour frequency. Then the data is split into two sub-sets: for training model the first 80% of the data is used and the last 20% of the data is used for testing the model accuracy. The whole data is used to create the training and testing data chronologically.

4.3 Evaluation criteria

The performance of the predictive model is evaluated using different accuracy metrics. The evaluation metrics evaluate the trained model by measuring the difference between the forecast value and the observed value by using different algorithms. The Root Mean Square Error (RMSE) is calculated as follows :

$$RMSE = \sqrt{\frac{1}{nT}\sum_{i,t}(\hat{y}_{i,t} - y_{i,t})^2} \tag{19}$$

where $y_{i,t}$ is the actual value of the data i at time t, $\hat{y}_{i,t}$ is the mean of the predicted data, n is the number of time series and T is the prediction horizon [39].

The Mean Absolute Error (MAE) measures the average values of the errors in a set of predictions. It is defined as:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \tag{20}$$

where n is a time steps, y_t is the actual value and $\hat{y}_t$ is the forecast value [40].

The Mean Absolute Percentage Error (MAPE) is another often used metric for measuring forecasting accuracy. However, it has one significant disadvantage, if the time series contains zero or close-to-zero values it generates infinite values. To address this problem another metrics is proposed in [41] called Mean Arctangent Absolute Percentage Error (MAAPE), which inherently follows the MAPE philosophy. The MAAPE is defined as follows:

$$MAAPE = \frac{1}{n}\sum_{t=1}^{n}\arctan(|\frac{y_t - \hat{y}_t}{y_t}|) \tag{21}$$

Symmetric Mean Absolute Percentage Error (SMAPE) measures the accuracy based on the percentage of the errors [42]. It is usually defined as follows:

$$SMAPE = \frac{100\%}{n} \sum_{t=1}^{n} \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2} \tag{22}$$

where A_t is the real value and F_t is the predicted value.

The forecast distribution accuracy is evaluated using weighted quantile loss. The quantile are in the range of $[0, 1]$, The weighted quantile loss is defined as follows:

$$wQuantileLoss[\tau] = 2\frac{\sum_{i,t} Q_{i,t}^{(\tau)}}{\sum_{i,t} |y_{i,t}|}, \text{ with } Q_{i,t}^{(\tau)} = \begin{cases} (1-\tau)|q_{i,t}^{(\tau)} - y_{i,t}|, & \text{if } q_{i,t}^{(\tau)} > y_{i,t} \\ \tau|q_{i,t}^{(\tau)} - y_{i,t}|, & \text{otherwise} \end{cases} \tag{23}$$

$q_{i,t}^{(\tau)}$ is the τ-quantile of the forecast distribution which the model predicts [39].

4.4 Forecasting photovoltaic production

In this section the experiment done in the forecasting of PV production is discussed. This experiment includes the following points:

- Training of the model.
- Analyze the forecasting results.
- Evaluating the relevance of input variables.
- Compare the model with other benchmark models.

4.4.1 Description of the training

The total window length T is sum of prediction range and conditioning range, which is constant in each experiment. Multiple training instances are created during training by choosing windows with various starting points. The conditioning range and prediction range are associated with the starting point of each training sample.

The training is performed by selecting only one starting point for each experiment. The window length T is illustrated in Figure 15 in association with the time ranges $[1, t_{0-1}]$ and $[t_0, T]$, which refer to the conditioning range (past) and prediction range (future) for time series of i at time t, $Z_{i,t}$ respectively.

To evaluate the most suitable window length, for the prediction horizon of 36 hours, different values of the context length are tested. The focus of this thesis is 36 hours

prediction horizon.

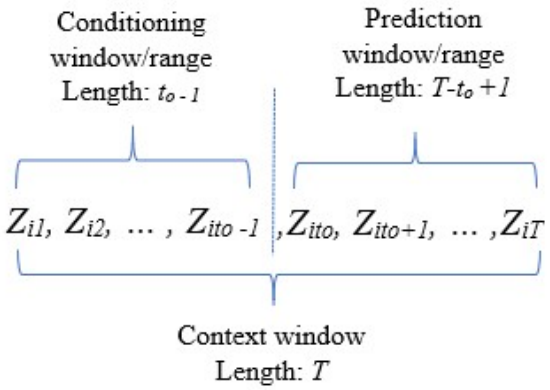


Figure 15. Illustration of the ranges during training.

The different combination of the parameters for prediction length and context length are tested, for the prediction horizon of 36 hours the context length 168 is selected and for one hour prediction horizon, context length 24 is selected based on the performance evaluation 5.

4.4.2 Testing hyperparameters

The optimal value for the hyperparameters are selected based on the results obtained from the evaluation criteria. Different values within the given range are selected for testing as described in Table 4. The parameters selected for testing are as follows:

- The context length.
- The learning rate of the model training.
- The number of epochs of the training.
- The number of cells.
- The number of layers

For testing each parameter, the values of the other parameters are set to their default value [43]. The results obtained are presented in 16, 17. The performance in this experiment is evaluated using RMSE and MAAPE evaluation metrics. Then the optimal value of each parameter is deduced based on the minimal value of the evaluation metrics.

Table 4. Hyperparameters to optimize.

Hyperparameters	Values
Context length	[24 48 72 168]
Number of layers	[2 3 4]
Number of cells	[40 50 80 100]
Learning rate	[1e-3 1e-2 1e-1]
Number of epochs	[20 50 80 100]

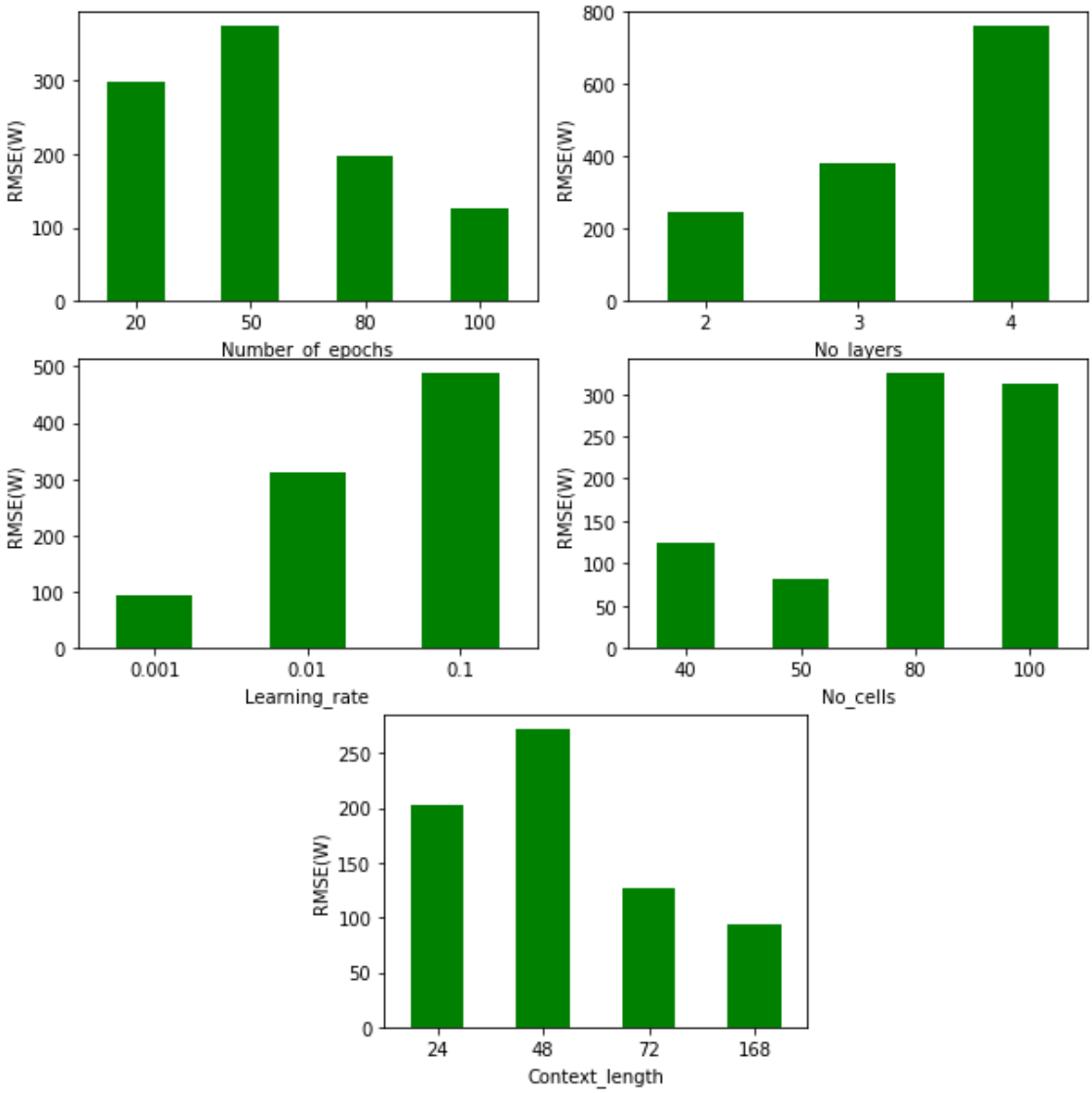


Figure 16. Comparison of hyperparameters with different values RMSE.

The forecasting performance is influenced by the parameter values. For the parameters number of epochs, learning rate value, number of layers, and number of cells, the optimal value corresponds with the default value of the model. However, the context length for

the prediction horizon of 36 hours used is 24*7=168, which is equivalent to one week observation. For the prediction horizon of one hour the context length is considered to be 24. The context length is associated with the seasonality of the time series which varies on the resolution of the data.

Table 5. Hyperparameters fixed for 36-hour prediction horizon.

Parameters	Value
Prediction length	36
Context length	168
Number of layers	2
Number of cells	50
Learning rate	0.001
Maximal number of epochs	100

4.4.3 Relevance of input variables

The relevance of the input time series for forecasting PV production is analyzed using feature importance score, which is presented in Figure 18. There are 10 input time series used as weather features. The relevance of the weather variables are investigated against the target time series PV production output. The goal is to select weather variables that can be used as weather feature in forecasting PV production. From the weather data the most relevant variables are global radiation, diffuse radiation and direct irradiance and air temperature.

The feature importance is evaluated based on chi-square statistical testing. The chi-square is a statistical test of independence, which measures the degree of dependency of two variables. In this thesis, the variables are target time series value (pv_inv_out) and the other feature variables (all weather variables). Then by performing chi-square statistics between each feature variable and the target variable, the relationship between the them is observed.

According [11], in order to produce the power output of the PV system, first the global radiation and the PV module temperature should be estimated. The used radiation can be determined using at least two of the following weather variables: DNI, DHI or GHI see

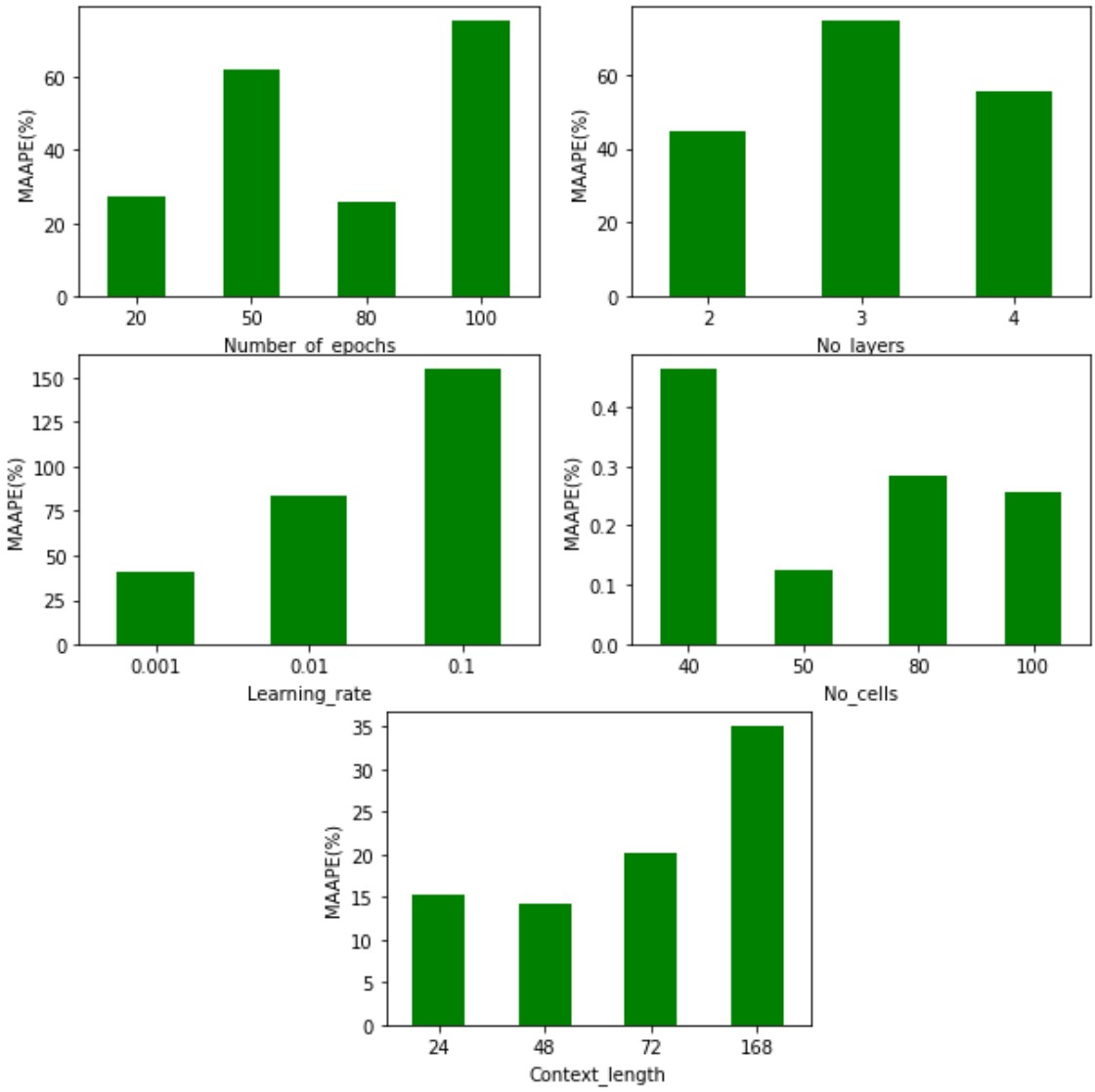
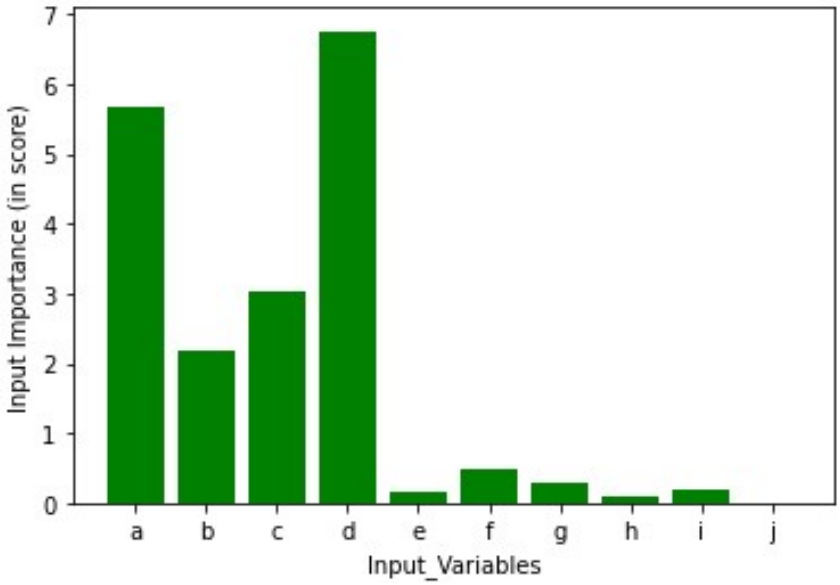


Figure 17. Comparison of hyperparameters with different values (MAAPE).

Section 2.2. According to the estimated score returned from the feature importance score function, we can observe that the specified parameters have high score.

4.4.4 Forecasting of photovoltaic production

The forecasting of PV production for 36-hour and one-hour prediction horizon is done based on the selected parameters see Table 5. The target time series is the PV production output and additional feature used is time feature and weather data. The weather variables used for the experiment are the GHI, DNI and air temperature. The variables are first selected based on their feature importance which is studied in Section 4.4.3.



a	global radiation on horizontal surface	f	PV module temperature, NE corner
b	diffuse radiation on horizontal surface	g	PV module temperature, SW corner
c	direct normal irradiance	h	air temperature
d	global radiation on inclined PV plane-of-array	i	wind speed
e	air temperature on roof, vicinity of PV arrays	j	wind direction

Figure 18. Relevance of the input variables.

The plots of the forecasting results are presented with the predicted distribution represented by its quantiles. The results are plotted all with same quantile values quantile (0.8), quantile (0.9) and median Figure 19- 21.

Table 6. Results of the forecast prediction for 36-hour with PV data only.

	RMSE (in W)	MAAPE (in %)
Best	18	6.01
Average	1499	64.26
Worst	4424	93.26

The PV forecasting results are evaluating based on RMSE and MAAPE evaluation metrics. The result obtained using weather data with the PV production historical data for both prediction horizon of 36-hour and one hour shows slight improvement, which is beyond expectation. The forecasting results are illustrated as shown in Table 6 and 9.

Table 7. Results of the forecast prediction for 36-hour with NWP.

	RMSE (in W)	MAAPE (in %)
Best	11	0
Average	1400	37
Worst	4424	98

Table 8. Results of the forecast prediction for one-hour with PV data only.

	RMSE (in W)	MAAPE (in %)
Best	8	0
Average	7.14	0
Worst	174	93.26

Table 9. Results of the forecast prediction for one hour with NWP.

	RMSE (in W)	MAAPE (in %)
Best	0	0
Average	5	0
Worst	169	78

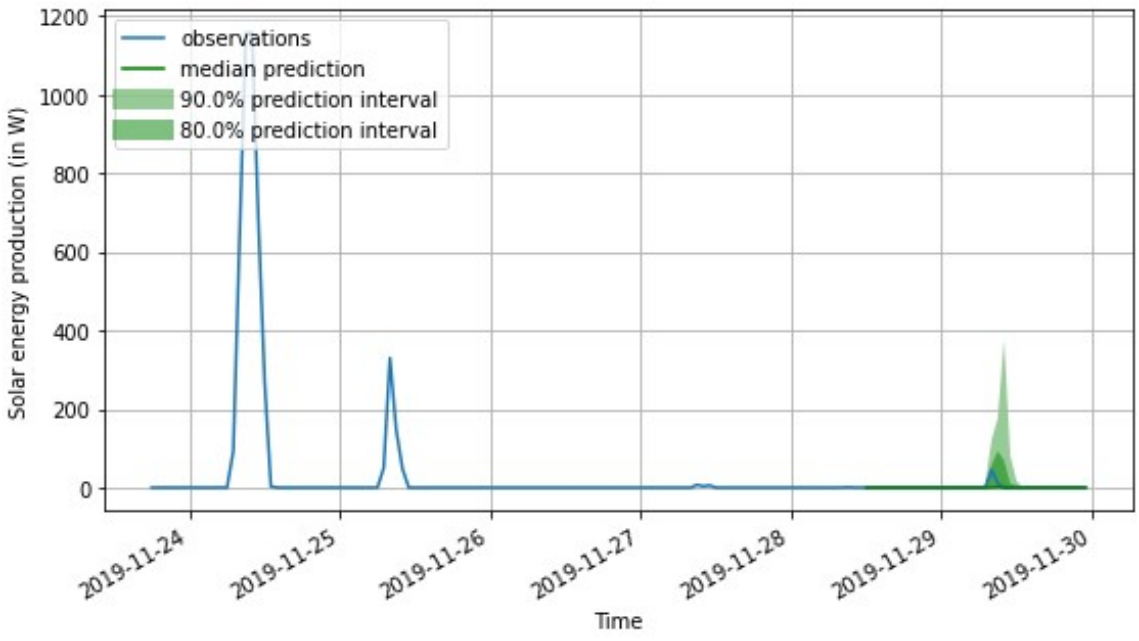


Figure 19. Best forecast result for 36-hour prediction horizon with NWP.

The best, average and worst forecast samples of the PV forecasting are presented in Figure 19-21. Most of the best forecast samples are predicted during winter time. During these time, there is no PV power is produced. That is why, the prediction becomes easier.

The best forecasting sample is same with its corresponding true value. Hence, the forecast is more accurate for these samples.

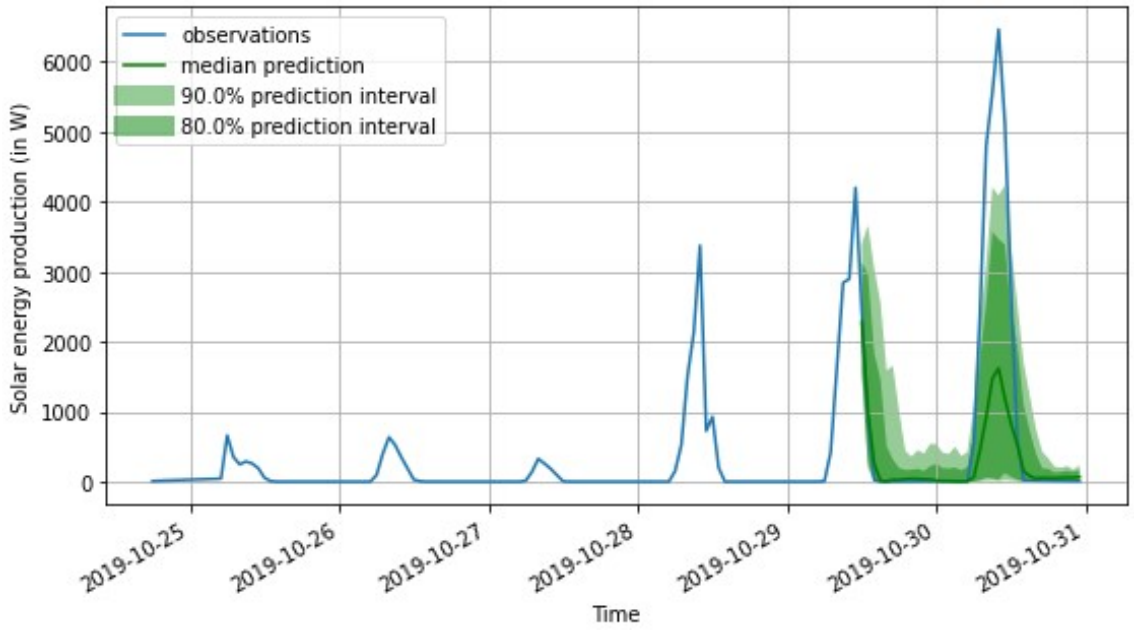


Figure 20. Average forecast result for 36-hour prediction horizon with NWP.

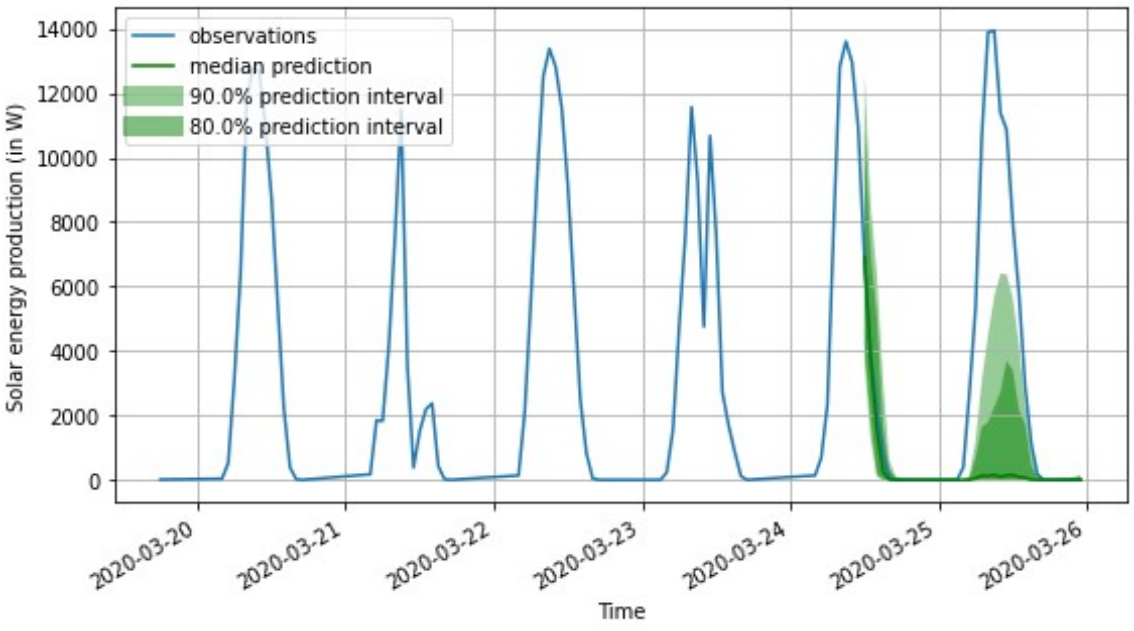


Figure 21. Worst forecast result of 36-hour prediction horizon with NWP.

4.4.5 Result of benchmark models

The prediction result achieved from the DeepAR is compared with other baseline models (see Section 3.2) in order to evaluate its forecasting performance. The forecasting performance is evaluated for the prediction horizon of 36-hour and one hour. The forecasting performance of each model is evaluated using RMSE and MAAPE evaluation metrics and compared with the result obtained from the DeepAR model as shown in Table 13.

Naïve forecasting model is used to forecast the PV production for 36-hours and one-hour prediction horizon. The results obtained for the best, average and worst are presented in Table 10, Figure 22- 24. For one hour prediction for the best and average, the forecast samples are from the dark periods.

Table 10. Results of the forecast prediction for 36-hour using Naive model .

	RMSE (in W)	MAAPE (in %)
Best	0	0
Average	4289	64.7
Worst	9989	62.33

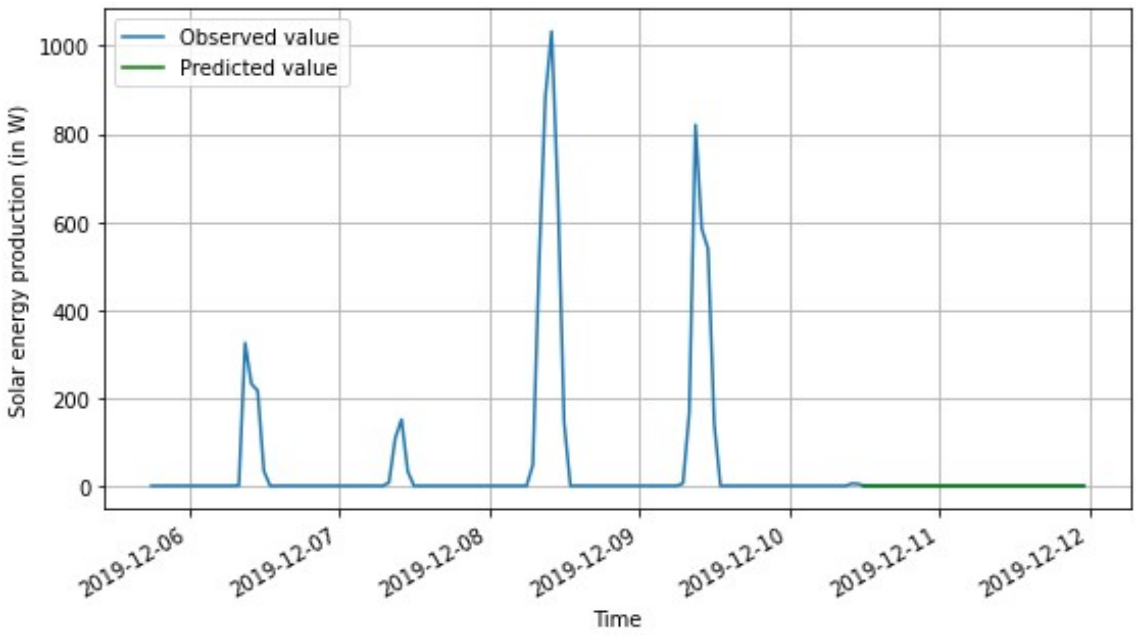


Figure 22. Best forecast result of 36-hour prediction horizon using Naive model.

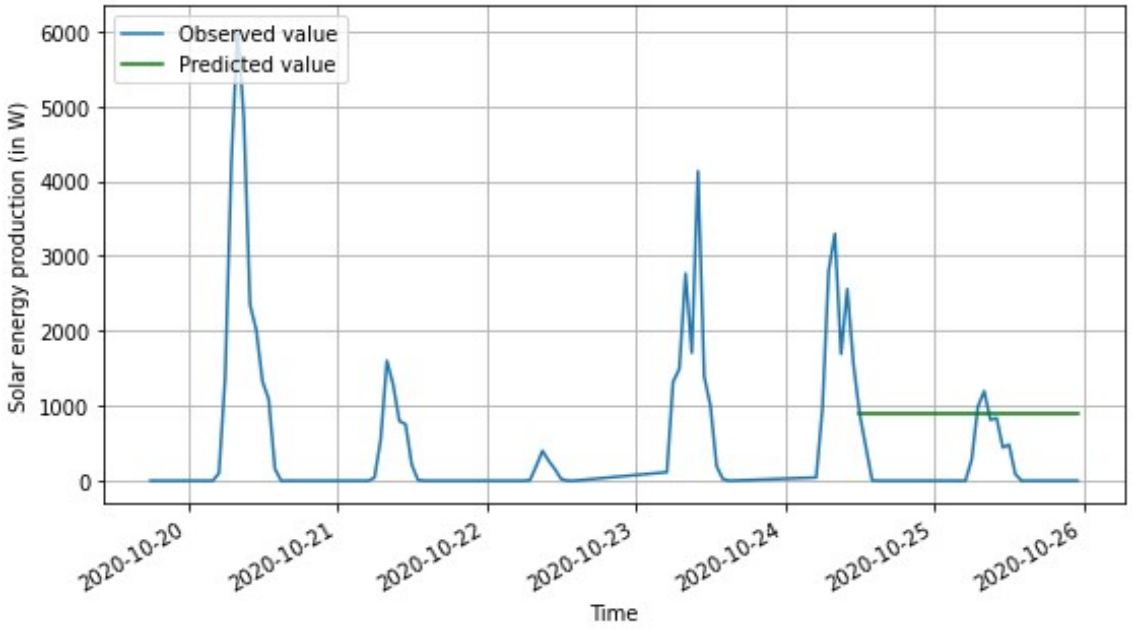


Figure 23. Average forecast result of 36-hour prediction horizon using Naive model.

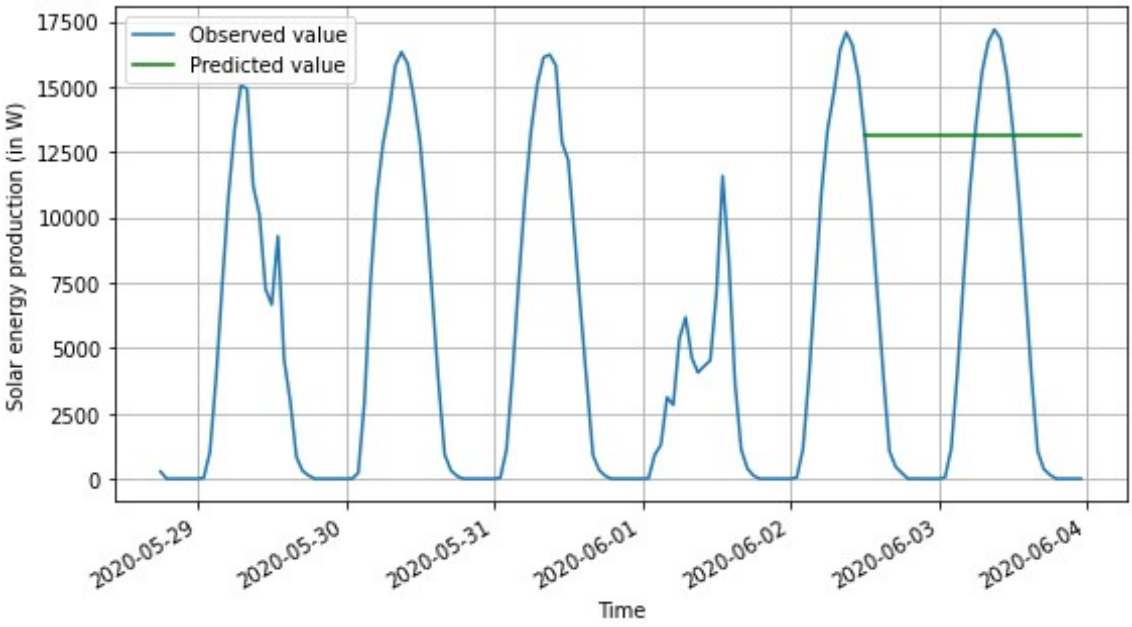


Figure 24. Worst forecast result of 36-hour prediction horizon using Naive model.

The seasonal naive predictor, predicts the value based on the last observed data from the same season. To predict the PV production for 36-hour and one hour, the results are obtained using season length of 24. The season length varies based on the seasonality of the data used for prediction. The results obtained are for the best, average and worst are presented in Table 11, Figure 25- 27.

Table 11. Results of the forecast prediction for 36-hour using Seasonal naive predictor .

	RMSE (in W)	MAAPE (in %)
Best	15	10.12
Average	1119	40.27
Worst	5716	59.27

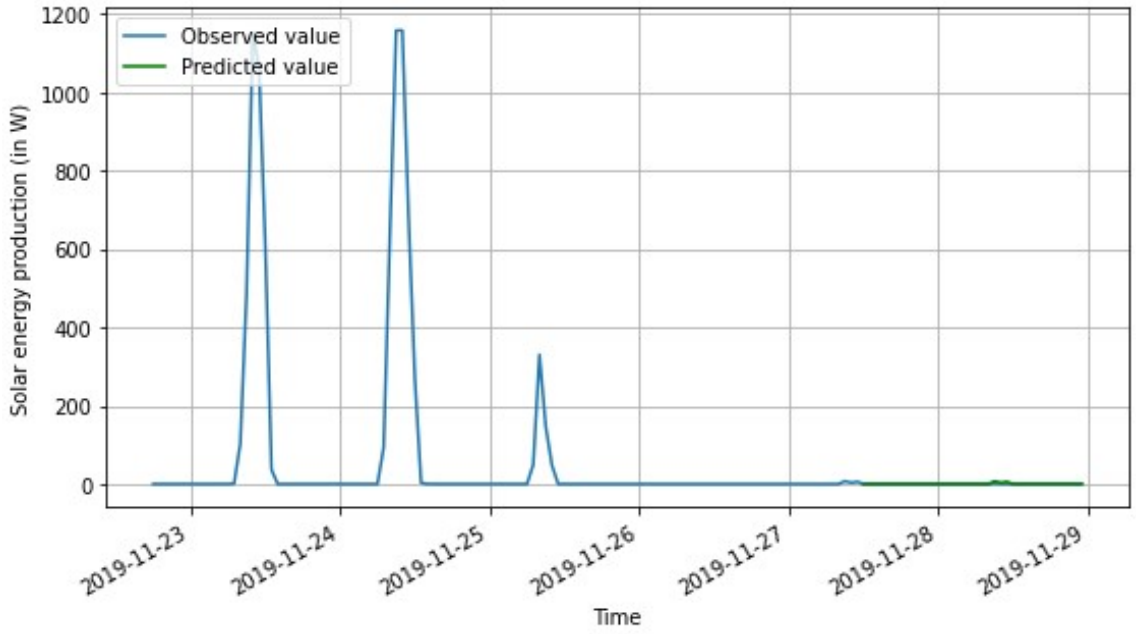


Figure 25. Best forecast result of 36-hour prediction horizon using seasonal naive predictor.

A constant predictor produces same forecast value for all prediction horizon. This model is suitable when the forecast values needs to be constant. Since the data is not constant in this case, the default value zero is predicted for the forecast values. The results obtained are for best forecast sample is similar to naive predictor Figure 22, average and worst are presented in Table 12, Figure 28, 29.

Table 12. Results of the forecast prediction for 36-hour using constant predictor .

	RMSE (in W)	MAAPE (in %)
Best	0	0
Average	3078	47
Worst	8216	58.9

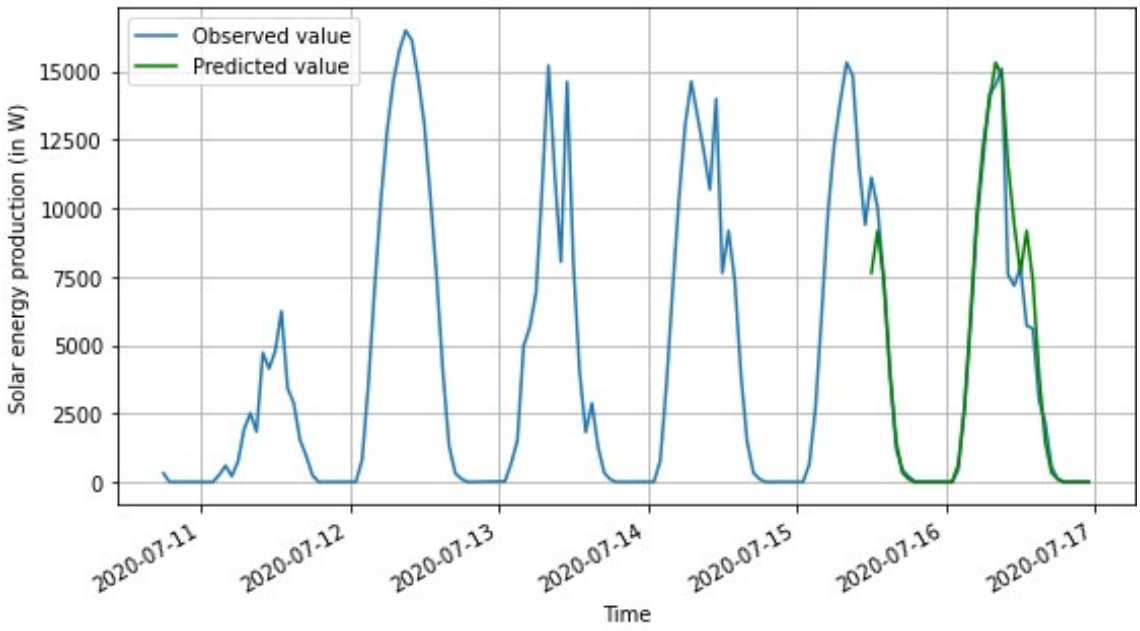


Figure 26. Average forecast result of 36-hour prediction horizon using seasonal naive predictor.

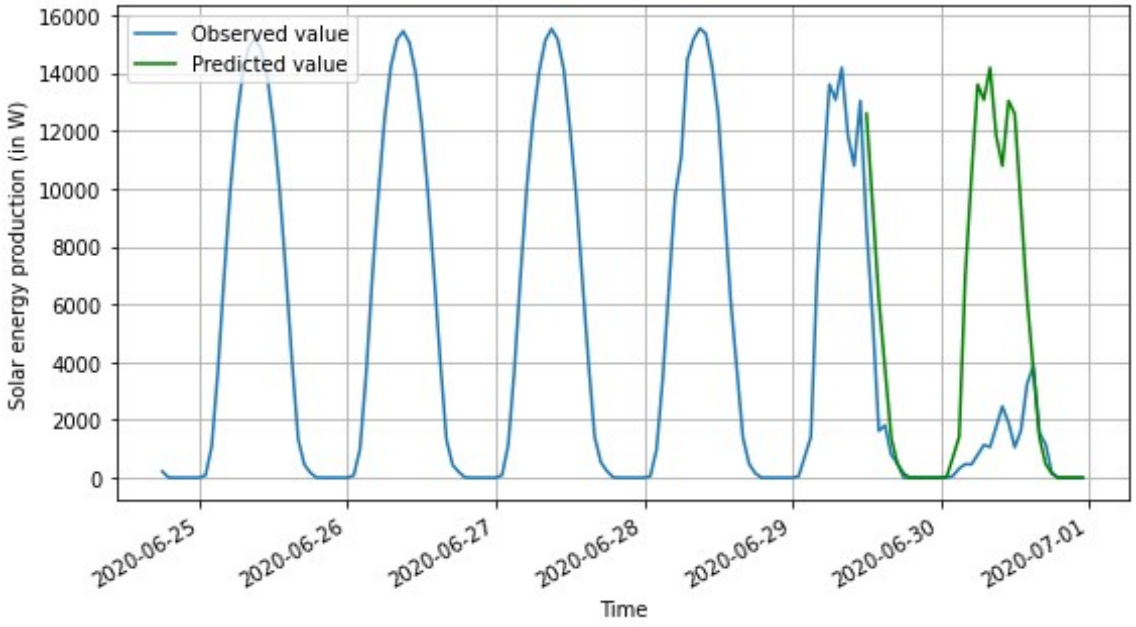


Figure 27. Worst forecast result of 36-hour prediction horizon using seasonal naive predictor.

The results obtained from the baseline models are compared to DeepAR model and the DeepAR model outperforms for 36-hour prediction horizon. The result is presented in Table 13-14. However, for one hour prediction the results of the benchmark models for best and average forecast are more accurate due to forecast samples are taken during winter time.

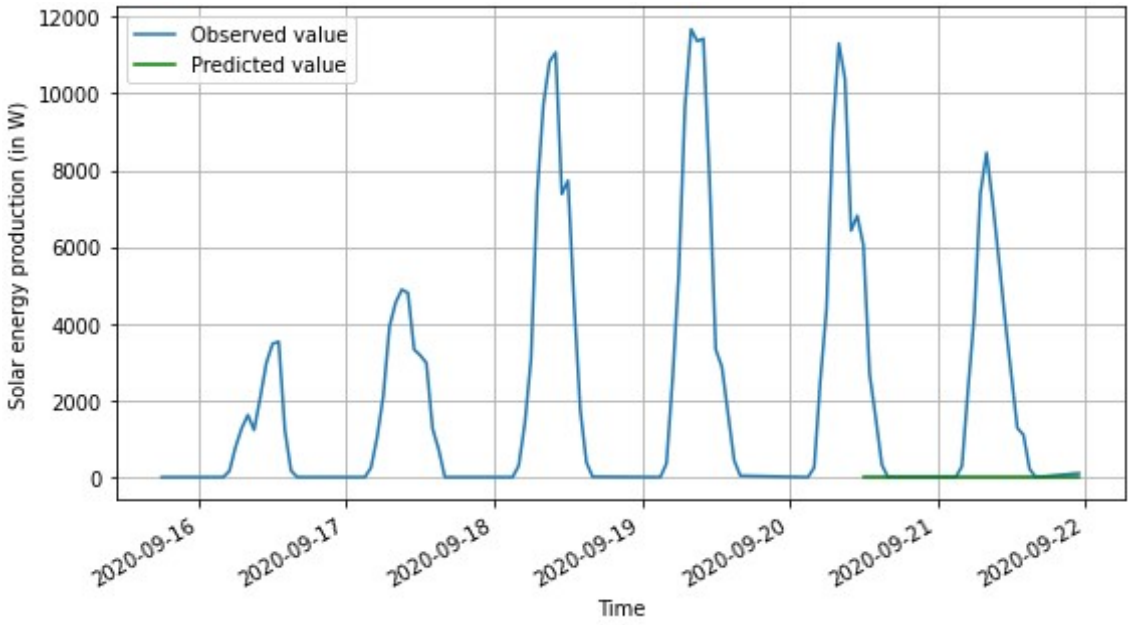


Figure 28. Average forecast result of 36-hour prediction horizon using constant predictor

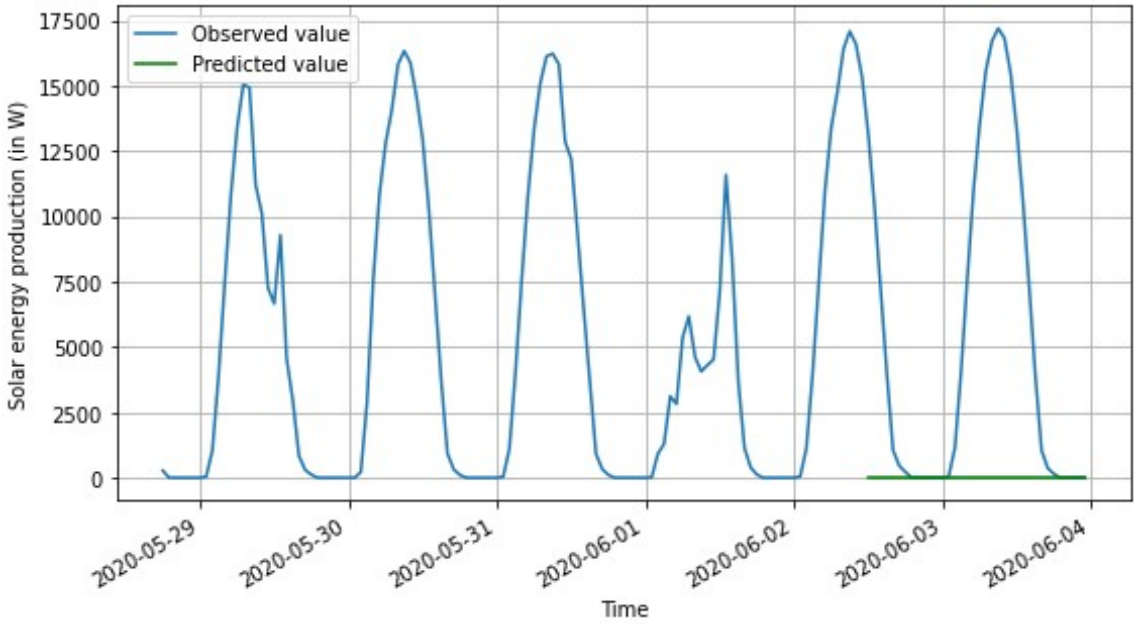


Figure 29. Worst forecast result of 36-hour prediction horizon using constant predictor

The results achieved from DeepAR and seasonal naive predictor, the difference is less comparing to naïve and constant predictor. The reason for that is both these models involve the seasonality of the data during prediction. But, the naïve and constant predictor performs worst as forecast values are constant for the whole prediction length. However, apart from the evaluation metrics used for comparison the DeepAR model provides additional metrics (see Section 4.3) for evaluating the forecast distribution.

Table 13. Results of the forecast 36-hour prediction horizon RMSE (in W).

	DeepAR	Naïve	Seasonal Naive Predictor	Constant Predictor
Best sample	11	0	15	0
Average sample	1400	4289	1119	3078
Worst sample	4424	9989	5716	8216
Average error	1778	3781	2148	4091

Table 14. Results of the forecast one-hour prediction horizon RMSE (in W).

	DeepAR	Naïve	Seasonal Naive Predictor	Constant Predictor
Best sample	0	0	0	0
Average sample	5	0	0	0
Worst sample	169	194	163	194
Average error	22	19	20	19

5 DISCUSSION

5.1 Current study

DL techniques are used in many time series forecasting problem and show promising results. The DL models can be integrated with probability model to make probabilistic forecasting. RNN based model become the most popular model for predicting time series forecasting due to its ability to extract features from a given data. In this study the DeepAR model is used for the probabilistic forecasting of the PV production.

The forecasting of RE production on this thesis is focused only on the PV production forecasting. Before predicting the PV production the hyperparameters of the model are tested with different possible values. Then, their impact on the prediction result are evaluated using the evaluation metrics. However, the prediction results are more affected by the context length than by other hyperparameters. One possible reason could be the dependency of the context length in the seasonality of the input data to the model.

Two cases were considered when evaluating the model performance: one case is using historical PV data only and the second case is using weather data along the with historical PV production data. The weather variables used in the model are investigated based on its feature importance score since all the weather variables have not equal contribution to the prediction result. The weather variables used are GHI, DHI and air temperature. Using the selected weather variables, the prediction result for 36-hour and one-hour prediction horizon is slightly improved. The reason could be due to limited set of weather parameters used. The weather data used in this thesis does not include the weather parameters like cloud coverage, amount of rain intensity, dew point, relative humidity and other parameters see Section 2.2.

The prediction result obtained is also compared with other benchmark models i.e naïve predictor, seasonal naive predictor and constant predictor. The DeepAR model outperforms in the prediction result compared to these models for the prediction horizon of 36-hour. The seasonal naive predictor is suitable when there exist seasonality in the given time series. The best forecasting sample is obtained during winter time because during these periods there is no PV power generated, which is easier to predict.

5.2 Future work

According to the literature review presented in Section 2.3.3, CNN model has the ability to learn complex patterns and extract features. DeepTCN model is capable of learning intricate patterns throughout the time series from both history data and past and exogenous data. Therefore using DeepTCN model for probabilistic forecasting can be a good idea for predicting RE production. The results achieved from the probabilistic models can be also compared to physical prediction models, in order to evaluate the performance of the probabilistic prediction model.

Moreover, using additional weather variables i.e., cloud cover, relative humidity, dew point as input to the model can improve the prediction performance. In addition, this thesis is only focused on PV production forecasting. However, other forms of RE production i.e., wind energy can be also predicted using probabilistic forecasting models.

6 CONCLUSION

In this thesis, the DeepAR model is used for predicting PV production data. The DeepAR model combines the RNN deep learning model with probabilistic model to bring probabilistic forecasting.

The data used in this thesis is collected from the FMI station located in Helsinki, Finland. The model is used to forecast 36-hour and one hour prediction horizons of PV production data. The prediction result is evaluated using PV historical data only and with weather data. Since all components of the weather are not contributing to the enhancement of the forecasting, the weather data is first investigated its importance before using for the model. The weather variables selected for the model are the GHI, DHI and air temperature. However, the obtained results slightly improved with weather data for forecasting PV production.

The DeepAR model is also compared with other baseline models (naïve, seasonal naive predictor and constant predictor) to compare its forecasting performance. As the result the DeepAR shows better result comparing to these benchmark models for the prediction of 36-hour horizon.

REFERENCES

[1] Huaizhi Wang, Zhenxing Lei, Xian Zhang, Bin Zhou, and Jianchun Peng. A review of deep learning for renewable energy forecasting. *Energy Conversion and Management*, 198:111799, 10 2019.

[2] Renewable Energy Prospects for the European Union. https://www.irena.org/publications/2018/Feb/Renewable-energy-prospects-for-the-EU, 2018. [Online; accessed January, 11, 2021].

[3] J. R. Andrade and R. J. Bessa. Improving Renewable Energy Forecasting With a Grid of Numerical Weather Predictions. *IEEE Transactions on Sustainable Energy*, 8(4):1571–1580, 10 2017.

[4] Akın Taşcıkaraoğlu. Chapter 5 - Impacts of Accurate Renewable Power Forecasting on Optimum Operation of Power System. In Ozan Erdinç, editor, *Optimization in Renewable Energy Systems*, pages 159–175. Butterworth-Heinemann, Boston, 01 2017.

[5] Saleh Albelwi and Ausif Mahmood. A framework for designing the architectures of deep convolutional neural networks. *Entropy*, 19(6):242, 06 2017.

[6] Jatin Bedi and Durga Toshniwal. Deep learning framework to forecast electricity demand. *Applied Energy*, 238:1312–1326, 03 2019.

[7] Renewable Energy Forecasting in Finland. `https://www.motiva.fi/files/8542/Renewable_Energy_in_Finland_2014.pdf`, 2014. [Online; accessed January,30 2021].

[8] Gilles Notton, Marie-Laure Nivet, Cyril Voyant, Christophe Paoli, Christophe Darras, Fabrice Motte, and Alexis Fouilloy. Intermittent and stochastic character of renewable energy sources: Consequences, cost of intermittence and benefit of forecasting. *Renewable and Sustainable Energy Reviews*, 87:96–105, 05 2018.

[9] David S. Renné. Effective solar resource methodologies for sustainable PV applications. In Ali Sayigh, editor, *Photovoltaics for Sustainable Electricity and Buildings*, pages 25–53. Springer International Publishing, 2017.

[10] Viivi Kallio-Myers, Aku Riihelä, Panu Lahtinen, and Anders Lindfors. Global horizontal irradiance forecast for finland based on geostationary weather satellite data. *Solar Energy*, 198:68–80, 03 2020.

[11] Herman Böök, Antti Poikonen, Antti Aarva, Tero Mielonen, Mikko R. A. Pitkänen, and Anders V. Lindfors. Photovoltaic system modeling: A validation study at high latitudes with implementation of a novel DNI quality control method. *Solar Energy*, 204:316–329, 07 2020.

[12] Richard Müller, Uwe Pfeifroth, Christine Träger-Chatterjee, Jörg Trentmann, and Roswitha Cremer. Digging the METEOSAT treasure—3 decades of solar surface radiation. *Remote Sensing*, 7(6):8067–8101, 06 2015.

[13] H. Sangrody, M. Sarailoo, N. Zhou, A. Shokrollahi, and E. Foruzan. On the performance of forecasting models in the presence of input uncertainty. In *North American Power Symposium (NAPS)*, pages 1–6. IEEE, 09 2017.

[14] Chatfield Chris. *Time-series Forecasting*. Chapman and Hall/CRC., 1st edition, 2001.

[15] Mathieu Bourdeau, Xiao qiang Zhai, Elyes Nefzaoui, Xiaofeng Guo, and Patrice Chatellier. Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustainable Cities and Society*, 48(9):2210–6707, 07 2019.

[16] Nesreen Ahmed, Amir Atiya, Neamat Gayar, and Hisham El-Shishiny. An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29:594–621, 08 2010.

[17] J. Shi, W. Lee, Y. Liu, Y. Yang, and P. Wang. Forecasting Power Output of Photovoltaic Systems Based on Weather Classification and Support Vector Machines. *IEEE Transactions on Industry Applications*, 48(3), 05 2012.

[18] D. W. van der Meer, J. Widén, and J. Munkhammar. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. *Renewable and Sustainable Energy Reviews*, 81:1484–1512, 01 2018.

[19] Athanasopoulos G. Hyndman R.J. *11.3 Neural network models | Forecasting: Principles and Practice*. OTexts: Melbourne, Australia, 2nd edition, 2018.

[20] Facundo Bre, Juan M. Gimenez, and Víctor D. Fachinotti. Prediction of wind pressure coefficients on building surfaces using artificial neural networks. *Energy and Buildings*, 158:1429–1441, 2018.

[21] Anastasia Borovykh, Sander Bohte, and Cornelis W. Oosterlee. Conditional Time Series Forecasting with Convolutional Neural Networks. *arXiv:1703.04691*, 09 2018.

[22] I. Koprinska, D. Wu, and Z. Wang. Convolutional Neural Networks for Energy Time Series Forecasting. In *International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 07 2018.

[23] Y. Bengio, P. Simard, and P. Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2):157–166, 03 1994.

[24] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[25] Kyunghyun Cho, Bart van Merrienboer, Dzmitry Bahdanau, and Yoshua Bengio. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. *arXiv:1409.1259*, 10 2014.

[26] S. Yang, X. Yu, and Y. Zhou. LSTM and GRU neural network performance comparison study: Taking yelp review dataset as an example. In *International Workshop on Electronic Communication and Artificial Intelligence (IWECAI)*, pages 98–101. IEEE, 06 2020.

[27] M. S. Hossain and H. Mahmood. Short-Term Photovoltaic Power Forecasting Using an LSTM Neural Network and Synthetic Weather Forecast. *IEEE Access*, 8:172524–172533, 2020.

[28] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to Sequence Learning with Neural Networks. *arXiv:1409.3215*, 12 2014.

[29] Yitian Chen, Yanfei Kang, Yixiong Chen, and Zizhuo Wang. Probabilistic forecasting with temporal convolutional neural network. *Neurocomputing*, 399:491–501, 2020.

[30] David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3):1181 – 1191, 2020.

[31] Aleksei Mashlakov, Toni Kuronen, Lasse Lensu, Arto Kaarna, and Samuli Honkapuro. Assessing the performance of deep learning models for multivariate probabilistic energy forecasting. *Applied Energy*, 285:116405, 2021.

[32] Chirag Deb, Fan Zhang, Junjing Yang, Siew Eang Lee, and Kwok Wei Shah. A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, 74:902–924, 02 2017.

[33] U Thissen, R van Brakel, A. P de Weijer, W. J Melssen, and L. M. C Buydens. Using support vector machines for time series prediction. *Chemometrics and Intelligent Laboratory Systems*, 69(1):35–49, 11 2003.

[34] Simple RNN vs GRU vs LSTM :- difference lies in more flexible control. https://www.pluralsight.com/guides/introduction-to-lstm-units-in-rnn, 09 2020. [Online; accessed February, 11, 2021].

[35] E. Kharlova, D. May, and P. Musilek. Forecasting photovoltaic power production using a deep learning sequence to sequence model with attention. In *International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE, 07 2020.

[36] Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long- and short-term temporal patterns with deep neural networks. *The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval*, page 95–104, 2018.

[37] Siteng Huang, D. Wang, Xuehan Wu, and A. Tang. Dsanet: Dual self-attention network for multivariate time series forecasting. *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, page 2129–2132, 2019.

[38] Herman Böök and Anders V. Lindfors. Site-specific adjustment of a NWP-based photovoltaic production forecast. *Solar Energy*, 211:779–788, 11 2020.

[39] DeepAR Forecasting Algorithm,AmazonSageMaker. http://docs.aws.amazon.com/sagemaker/latest/dg/deepar.html. [Online; accessed April, 28, 2021].

[40] Mean absolute error,Wikipedia. http://en.wikipedia.org/wiki/Mean_absolute_error. [Online; accessed May, 16, 2021].

[41] Sungil Kim and Heeyoung Kim. A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3):669–679, 2016.

[42] Symmetric mean absolute percentage error,Wikipedia. http://en.wikipedia.org/wiki/Symmetric_Mean_absolute_error. [Online; accessed May, 16, 2021].

[43] GluonTS documentation. https://ts.gluon.ai/api/gluonts/gluonts.trainer.html. [Online; accessed May,17 2021].