SOLAR IRRADIANCE FORECASTING USING GLOBAL FORECAST SYSTEM (GFS)

HO MUN BOCK

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

SOLAR IRRADIANCE FORECASTING USING GLOBAL FORECAST SYSTEM (GFS)

HO MUN BOCK

THIS REPORT IS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF BACHELOR OF ELECTRONIC ENGINEERING WITH HONOURS

FACULTY OF ELECTRONIC AND COMPUTER ENGINEERING
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2019/2020

DECLARATION

I declare that this report entitled "*Solar Irradiance Forecasting Using Global Forecast System (GFS)*" is the result of my own work except for quotes as cited in the references.

Signature :

Author : HO MUN BOCK

Date : 17 June 2020

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Bachelor of Electronic Engineering with Honours.

Signature :

Supervisor Name : Dr. HO YIH HWA

Date : 25 JUN 2020 ....

DEDICATION

First of all, I would like to express my greatest gratitude to my family, my father, Ho Seng Tack, my mother, Lee Pooi Khuan, my brother and sisters for supporting me in completing the thesis whenever I am tried and I feel emotional. My family that really helps me to reduce a lot of my stress. Thanks a lot for my family members that gave me their strength and believe to me.

Moreover, I want to thank to my fellow friends. Thank you for having time with me and give mental support that I really need. Thank to my friends that always give me a hand when I need help. Without you guys, I have no idea how I can go through all this way.

ABSTRACT

Smart grid is the next generation of electrical grid that can be integrated with renewable source to produce electric power. Then, there is a challenge to integrate renewable energy such as solar into smart grid. This is because the output of solar energy is related to the solar irradiance which is lack of stability due to weather variation. Therefore, solar irradiance forecasting is the solution to solve this problem. Then, multiple regression (MR) and neural network (NN) models were built. For the model validation, neural network model has achieved correlation coefficient (R) of 0.9173 and root mean square error (RMSE) of 114.1820 which better than MR model. With weather forecast produced by seasonal ARIMA, one day ahead (inter-day) and an hour (intraday) solar irradiance were forecasted by MR and NN models. In addition, there is Global Forecast System (GFS) was applied then blended model formed by blending MR, NN and GFS models together. As the result, for inter-day forecasting, weighted mean absolute percentage error (WMAPE) of 11.79% achieved by NN model on sunny day. Then, blended model has the lowest WMAPE of 30.50% on cloudy day. On the other hand, NN model has outperformed compared to other models for intraday forecasting.

ABSTRAK

Smart grid adalah grid elektrik generasi seterusnya yang dapat disatukan dengan sumber yang boleh diperbaharui untuk menghasilkan tenaga elektrik. Kemudian, ada cabaran untuk menggabungkan tenaga boleh diperbaharui seperti solar ke grid pintar. Ini kerana output tenaga suria berkaitan dengan pancaran matahari yang kurang stabil kerana perubahan cuaca. Oleh itu, ramalan penyinaran solar adalah penyelesaian untuk menyelesaikan masalah ini.Kemudian, model regresi berganda (MR) dan rangkaian saraf (NN) dibina. Untuk pengesahan model, model rangkaian neural telah mencapai koefisien korelasi (R) 0,9173 dan min kesilapan persegi (RMSE) 114,1820 yang lebih baik daripada model MR. Dengan ramalan cuaca yang dihasilkan oleh ARIMA bermusim, satu hari lebih awal (antara hari) dan satu jam (dalam sehari) cahaya matahari diramalkan oleh model MR dan NN. Di samping itu, ada Sistem Ramalan Global (GFS) yang diaplikasikan kemudian dicampur model yang dibentuk dengan menggabungkan model MR, NN dan GFS bersama-sama. Hasilnya, untuk ramalan antara hari, ralat peratusan mutlak berwajaran (WMAPE) 11.79% dicapai oleh model NN pada hari yang cerah. Kemudian, model campuran mempunyai WMAPE terendah 30.50% pada hari mendung. Akhir sekali, model NN mempunyai prestasi yang lebih baik berbanding model lain untuk ramalan dalam sehari.

ACKNOWLEDGEMENTS

I would like to thank my faculty, Faculty of Electronic and Computer Engineering, FKEKK that gives me the opportunity to conduct my final year project. Also, thank you for my faculty that provides a good environment and complete facility to allow me to do my work more efficiency.

Besides that, I want to thank my final year supervisor, Dr. Ho Yih Hwa for giving me guidance and advice throughout the entire project. Furthermore, thank you Dr. Ho that spent your time to discuss and share your knowledge. Thank you for reviewing my work, progress and thesis. Thank you to give me guideline that helps me to improve my work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

Temp	:	Temperature
RH	:	Relative humidity
WS	:	Wind speed
P	:	Pressure
MR	:	Multiple regression
NN	:	Neural network
ARIMA	:	Autoregressive integrated moving average
VIF	:	Variance inflation factor
R	:	Correlation coefficient
RMSE	:	Root mean square error
WMAPE	:	Weighted mean absolute error

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background

Over the past decades, coal and natural gas are the natural resources used as conventional power generation in Malaysia [1]. Then, the burning process on coal and natural gas produced huge amount of carbon dioxide which causes global warming. So, the way to reduce and replace the use of those primary resources is to apply renewable energy such as solar energy.

Solar energy is kind of clean energy which energy can be generated based on sunlight. Then, solar panel is the main component in solar energy. The working principle of solar panel is the photovoltaic cells consisted in solar panel that used to convert direct sunlight into electricity. Also, our country has the benefit to expand solar energy due the geography of Malaysia is located at the equatorial region which is hot climate throughout the year. In order to apply solar energy to the user, there is smart grid technology that can be dealt with the solar energy.

Smart grid is known as next generation of electrical grid based on digital system to provide energy to customer via full duplex communication. The operation of this system is to control and monitor the information from energy supply to improve efficiency, optimize the reliability and transparency also reduce the cost of energy supply [2]. Smart grid covers the evolution for power generation, transmission and distribution. For example, long distance transmission with minimize loses and digital control of distribution. Also, smart grid can adjust power flow and recover power transfer service when there is incident like transformer was down happened [3]. A model set of smart grid includes power generation, transmission, distribution and residential consumption based on renewable energy shown as Figure 1.1.

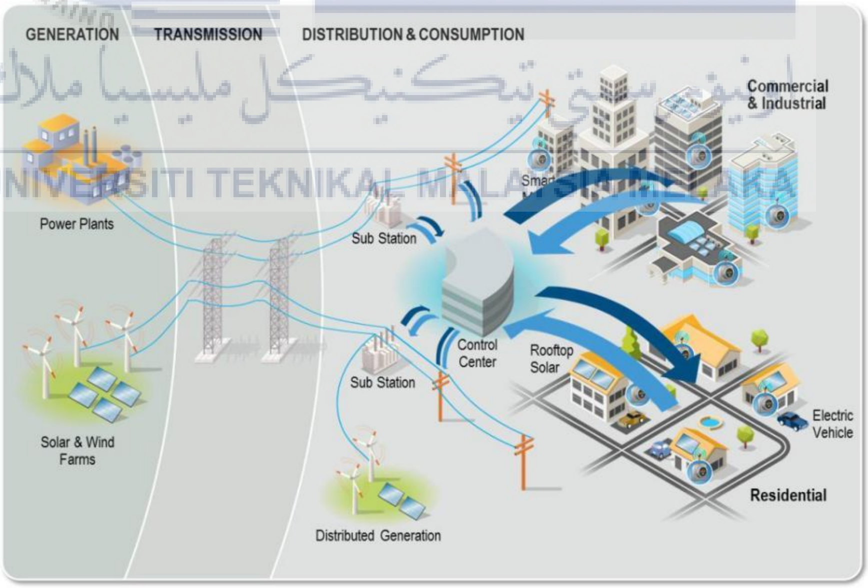

Figure 1.1: A model set of Smart Grid [4]**.**

Besides that, there are a lot of benefits by applying smart grid technology. One of the benefits is smart grid has the function of demand side management which allows customer to save their power consumption during peak hour. Furthermore, smart grid enables the grid from power distribution connected with solar panel as well as renewable resources to reduce carbon dioxide emission. Then, power storage involved for load balancing is also another benefit of smart grid. Other than that, smart sensor has inserted into smart grid to eliminate electricity failure and increase reliability [5]. Moreover, there is advanced metering infrastructure (AMI) as part of smart grid by connecting smart meter which can measure and analyze the energy usage. On the other words, AMI provides management capability that allows user to know their real-time energy usage [6].

Moreover, smart grid that operates in small scale is known as smart microgrid. There are AC, DC and hybrid AC/DC microgrids [7]. AC microgrid is utilized existing AC grid while DC microgrid is based on distributed renewable sources such as solar energy. Then, the hybrid AC/DC microgrid is enabled to change in between gird-connected and island modes. For the island mode, it is referred to the renewable energy management with two forms of operation. One form is without the connection of electric grid. Another form is the connection with electric grid in parallel in case there is run out the storage of solar energy. Since the output of solar power is directly related to solar irradiance, then solar irradiance forecasting is essential to implement together with smart grid technology. This is because solar irradiance can be affected by the weather variations.

In the term of sustainability, there is consideration of solar energy in the fields of environment, economic and social. The traditional power generator based on fossil fuel such as coal, petroleum and natural gas. Fossil fuel is also known as non-renewable resource. The depletion of fuel fossil can cause serious problem which is energy crisis [8]. This is why the way to expand in renewable energy such as solar energy. Then, solar irradiance forecast technique helps to stable the production of electric based on solar energy. So, the renewable energy can replace the use of fossil fuel and also reduce carbon dioxide emission.

Due to development technologies, the cost of solar panel is getting cheaper. There is low cost based on solar energy to generate electric power [9]. However, solar irradiance forecast system is needed to know how much solar energy will be produced for the next day in order to sustain and stable the energy produced by solar energy. When the electric energy is generated based on solar energy, the cost of power generator based on fossil fuel can be reduced.

With the smart grid technology, resident can able to integrate their own solar panel on the roof top. The benefit of integrating solar panel is to generate electric based on solar energy and reduce the electric bill from the power supplier. However, the solar energy is required the support of storage battery [10]. Then, solar irradiance forecast takes part to know how much solar energy produced to maintain the battery life as well as protect the investment of solar panel.

In conclusion, this project is going to build models that used to forecast solar irradiance. Therefore, the solution of forecasting is the way to support the field of solar energy and solve the problem of integration solar energy into smart grid.

1.2 Problem statement

Nowadays, smart grid is defined as smart electrical network that combines with electrical network and digital communication which enables to monitor and control the system. Besides that, smart grid has capable of providing electrical energy from renewable resources [11]. The increase in contribution of renewable energy source into the grid is part of smart grid initiative. However, the intermittency of renewable resource due to weather variation is a challenge for grid operators when integrating renewables such as solar energy into electrical network. Thus, forecasting is a solution for system grid operators in order to manage the efficiency of solar photovoltaic (PV) and satisfy the demand of energy consumers. Then, Global Forecast System (GFS) model is weather forecast model that run by National Oceanic and Atmospheric Administration (NOAA) which is produced by National Centers for Environmental Prediction (NCEP). GFS dataset have temperature, winds, soil moisture, atmospheric ozone and irradiance that can be used to forecast the weather for 16 days in the future. GHI (Global Horizontal Irradiance) is one of the measurements from GFS model which can used to forecast the solar irradiance. Even though GFS model can forecast the solar irradiance but there is a gap of 3-hour interval which may happen a lot of weather changes. Multiple regression (MR) and neural network (NN) models based on ARIMA are proposed to forecast solar irradiance. The forecast result for blended model is determined by blending with MR, NN and GFS models together.

1.3 Objectives

The objectives of my project are provided in order to overcome the problem, which are:

i. To forecast solar irradiance by using multiple regression and neural network models based on ARIMA.
ii. To analysis the accuracy and performance of solar irradiance forecasting.
iii. To determine solar irradiance forecasting by blending with global forecast system (GFS).

1.4 Scope of work

This research will focus on solar irradiance forecasting using multiple regression and feed-forward neural network based on ARIMA models. In modelling process, 4 months data (from 24 Oct 2019 to 24 Feb 2020) of temperature, humidity, wind speed, pressure and time (in decimal) used as independent variables while solar irradiance as dependent variable. Multiple regression (MR) and neural network (NN) models were built. Both MR and NN models had been validated using correlation coefficient (R) and root mean square error (RMSE). Only the model with best validation was chosen. Then, ARIMA models were generated for the purpose of weather forecast. Also, blended model had been formed by blending MR, NN, GFS models together to determine solar irradiance forecast. Moreover, weighted mean absolute percentage error (WMAPE) as performance metric was used to evaluate each model in different testing days.

In the terms of modern tools, MATLAB version 2016a is the main computing software used in my project. Besides that, Excel is used to manage all the data. Also, NOAA climate and weather toolkit used to read the GFS file.

1.5 Organization of thesis

This thesis is divided into five chapters. Chapter 1 is introduction of project which consists of background, problem statement, objective and scope of work. Chapter 2 is literature review that describes related works had been done by other researchers. Then, identify gap from other works and define my own method.

Furthermore, chapter 3 is methodology of project. This chapter included overall flow chart, sub flow charts, weather forecast (inter-day and intraday), forecast blending and performance metric. All of the content had described.

Besides that, chapter 4 is result and discussion. In this chapter, data analysis and correlation between independent variables and dependent variable had been recorded. The models of multiple regression, neural network and ARIMA were built. The best model is identified based on model validation. Blended model is formed by blending with MR, NN and GFS models. Then, forecast performance for each model were evaluated in different testing days.

Finally, chapter 5 is conclusion and recommendation for the future works. This chapter describes the achievement of my works and the methods to improve my project in the future.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Nowadays, the world is in the generation of information and technology. Technology is changing frequently to fulfill the human requirements. Forecasting technique is also one of the growing technologies due to the evolution of generation. Then, there are many different kinds of forecasting techniques generated. For instance, regression as statistical approach while artificial neural network (ANN) as machine learning approach and autoregression integrated moving average (ARIMA) as time series approach. Also, Global Forecast System (GFS) model is one of existing weather forecast models that can be applied. Therefore, various research papers on solar irradiance forecasting can be found. In addition, those were related to my proposed methods that can be used as references in order to achieve the objectives of my project.

2.2 Statistical approach

Regression is a process of modelling between a dependent variable and one or more independent variables. [12], [13]. Regression with more than one independent variable called multiple regression. The kind of regression can be defined as linear and quadratic. Linear regression is linear form while quadratic regression is non-linear form.

However, the regression model which has two or more independent variables can cause the issue of multicollinearity [14]. The effect of multicollinearity that makes the coefficients of regression insignificant when there are many similar independent variables. So, in order to avoid multicollinearity, then variance inflation factor (VIF) can be used to detect whether the independent variables are correlated to each other or not. Also, the range of VIF as shown in Table 2.1. For instance, if the independent variables with the VIF is above 5, then one of them should be removed.

Table 2-1: The range of VIF.

VIF	Correlation between independent variables
1	Not correlated
Between 1 and 5	Moderately correlated
Greater than 5	Highly correlated

The goal of regression is to find the best fitted line (can be linear or quadratic) by the method of least-squares fit then estimate the coefficients [15]. As stated in [16], the general equation for multiple linear and quadratic regression are shown in Eq. (2.1) and (2.2) where β_0 = intercept, β_1, β_2 = linear coefficients, β_{11}, β_{22} = quadratic coefficients, β_{12} = interaction coefficient, X_1, X_2 = parameters and ε = random error that follows normal distribution with mean 0.

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \quad (2.1)$$

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} {X_1}^2 + \beta_{22} {X_2}^2 + \beta_{12} X_1 X_2 + \varepsilon \quad (2.2)$$

Moreover, there are papers had been proposed for the statistical method to forecast solar irradiance. *Massidda et al.* [17] had reported a research on use of multilinear adaptive regression splines and numerical weather prediction to forecast the power output of a PV plant in Borkum, Germany. Multivariate Adaptive Regression Splines (MARS) is the method applied to form the model then predict the one day ahead photovoltaic power from power production. The input data used are history power output data and forecast weather data from Global Forecasting System (GFS). There are various performance measures such as coefficient of determination (R-squared), root mean square error (RMSE), mean bias error (MBE) and mean average error (MAE) used in this study. As the result, RMSE of 1106.1 kWh and MAE of 862.5 kWh as the error of daily energy production which equal to 11.0% and 8.6% of its maximum. In addition, the model can be further improved by putting huge number of input variables as well as extended data to build complex regression to increase the accuracy.

A paper of multivariate regression for prediction of solar irradiance had c by *Nalina U et al.* [18]. Multivariate regression model developed by solar irradiance (dependent variable) and air temperature and relative humidity (independent variable) for 4 days ahead solar irradiance forecast. The result showed that the standard error of 2.242% for the model.

A research on multiple linear regression equation for estimation of daily averages solar radiation in Chonburi, Thailand had completed by *Mekparyup et al.* [19]. Multiple linear regression (MLR) equation generated based on the input data of solar radiation, maximum and minimum temperature, humidity, hours of bright sunshine, water vapor pressure, sea level pressure collected from Meteorological Observation Bureau, Thai Meteorological Department since 2005 to 2009. Then, MLR model is evaluated by adjusted coefficient of determined (adjusted R-squared) and standard error of estimation (S). The best results of MLR model achieved are R-squared of 0.923 and S of 0.101.

2.3 Machine learning approach

Artificial neural network (ANN) uses machine learning algorithm that allows the neurons learn like a human brain [20]. In neural network, it consists of input, hidden and output layers. For each layer, there is neuron (also called as node) which connected in between multi-layer networks [20]. In mathematically, the output of neuron can be calculated by Eq. (2.3) where x = input, w = weight, b = bias, f = activation function and y = output. Furthermore, different implementation of activation function like sigmoid or hyperbolic tangent sigmoid which can affect the output of neuron. Then, the structure of neuron as shown in Figure 2.1.

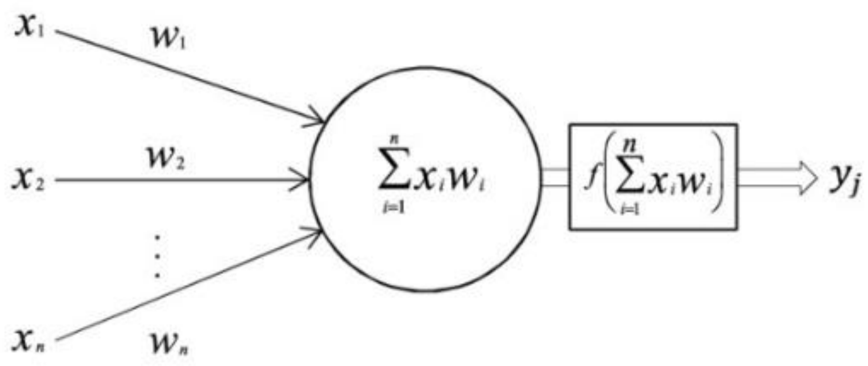

Figure 2.1: The structure of a neuron [21]**.**

$$y_j = f\left(\sum_{i=1}^{n} x_i \,.\, w_i + b\right) \tag{2.3}$$

Feed forward neural network (FFNN) is kind of neural networks with the application of forecasting. FFNN is based on backpropagation learning algorithm which is known as supervised learning. The function of backpropagation is to update weights to minimize error. [12]. Then, with the adjusted weights, the desired output can only achieve as actual data.

Furthermore, there are a lot of research papers had been done for on solar irradiance forecast using artificial neural network (ANN). *Khatib et al.* [22] had reported solar energy prediction method using artificial neural networks (ANNs). The developed network based on ANN is used to predict clearness index. Latitude, longitude, day number, and sunshine are the inputs while the output is clearness index. Then, the result of MAPE for predicted global solar irradiance is 5.92% which diffused solar radiation is 9.8%.

Marquez et al. [23] had developed a medium-term solar irradiance forecasting model by applying 11 predicted meteorological variables from the US National Weather Service's (NWS) forecasting database as inputs to an Artificial Neural

Network (ANN) model. The metrics included are root mean square error (RMSE) and correlation coefficient (R-squared). Then, the model has achieved R-squared of 0.945 and improved RMSE for GHI by 10 – 15%. However, the built model has less accuracy by increasing forecast horizon.

A research on solar insolation forecast using Artificial Neural Network for Malaysian weather had investigated by *Chua et al.* [24]. The paper presented 12 hourly solar insolation forecasting using Artificial Neural Network (ANN). Prediction for the next day 12 hours solar insolation based on multi-level perceptron (MLP) with back propagation technique model. The best performance MLP forecast model with least error of MSE and high R-squared. There are Com 3 and Com 4 investigated under sunny and rainy condition. As the result, the performance of Com 3 for sunny and rainy with MSE of 0.2% and 0.09% while Com 4 are 0.2% and 0.05%.

According to *Kumar et al.* [25], the title of Artificial Neural Network model for precise estimation of global solar radiation with the method of feed-forward neural network had been reported. The inputs are mean value of year, duration (year), month, latitude, longitude, sunshine, temperature, humidity, wind speed and rainfall while the output is solar radiation. As the result, mean percentage error, root mean square error and mean bias error mean percentage error between measured and estimated solar radiation are the range of -4.16 to 4.82, 0.02 to 0.26 and - 0.30 to 0.08.

Abuella et al. [26] had presented solar power forecasting using Artificial Neural Networks. 12 independent variables from the European Centre for Medium-Range Weather Forecasts (ECMWF) used to forecast solar irradiance. Root mean square error (RMSE) and correlation coefficient (R) to evaluate the performance of models. There is also multiple linear regression (MLR) model conducted then compared to feed - forward curve fitting model (ANN) model. As the result, ANN model has achieved RMSE of 0.0554 and R-squared of 0.97 which is better than MLR model on the test case of May 2014.

A paper about day-ahead solar irradiance forecasting for microgrids using a long short-term memory recurrent neural network: a deep learning approach has been done by *Munir Husein et al.* [27]. The data included are historical solar irradiance, dry-bulb temperature, dew-point temperature and relative humidity to develop long short term recurrent neural network (LST-RNN) model that forecasts hourly day ahead solar irradiance. Also, LST-RNN model is compared to feedforward neural network (FFNN) model. In the result, LST-RNN model has achieved RMSE of 60.31 W/m^2 which is better than FFNN model.

Kim et al. [28] had published a two-step approach to solar power generation prediction based on weather data using machine learning. There are base model and two-stage approach used in this research. Base model is only based on weather forecast while two-stage approach model combined of weather forecast and measured weather to predict power generation. The weather forecast included rain type, sky type, temperature, humidity, wind speed, wind direction and elevation while measured weather data included radiation atmospheric pressure, vapor pressure and surface temperature. There are various algorithms such as support vector regression (SVR), classification and regression tree (CART), k-nearest neighbors (k-NN), adaptive boosting (AdaBoost), random forest regression (RFR) and artificial neural network (ANN) applied to models. In the result, R-squared of 70.5% as the best result achieved by random forest regression algorithm.

2.4 Time series approach

Time series forecasting is a model that forecast future value based on past value [29]. There are a lot of kind of time series models such as AR, MA, ARMA, ARIMA stated in [30]. ARIMA is a time series model that combined autoregression (AR), integration (I) and moving average (MA). ARIMA model that deal with seasonality is known as seasonal ARIMA. Usually, ARIMA model presented as $ARIMA\ (p, d, q)$ while seasonal ARIMA as $SARIMA\ (p, d, q)(P, D, Q)s$ presented in [31].

Then, the general equation of seasonal ARIMA without constant as shown in Eq. (2) where p = number of autoregression, q = number of moving average, P = number of autoregression in seasonal, Q = number of moving average in seasonal, s = seasonality and ε_t = Gaussian random variable with mean (0) and variance, σ. Furthermore, the representation of components in Eq. (2.4) as shown in Table 2.2.

$$\emptyset_p(B)\varphi_P(B^s)(1-B^s)^D(1-B)^d = \theta_q(B)\vartheta_Q(B^s)\varepsilon_t \tag{2.4}$$

Table 2-2: The representation of components in seasonal ARIMA model [31].

Component	Representation
Regular AR(p)	$\emptyset_p(B) = 1 - \emptyset_1 B^1 - \cdots - \emptyset_p B^p$
Seasonal AR(P)	$\emptyset_P(B^s) = 1 - \varphi_1 B^s - \cdots - \varphi_P B^{Ps}$
Regular MA(q)	$\theta_q(B) = 1 - \theta_q B^1 - \cdots - \theta_q B^{qs}$
Seasonal MA(Q)	$\theta_Q(B^s) = 1 - \theta_Q B^{Qs} - \cdots - \theta_Q B^{Qs}$
Differencing, d	$(1-B)^d$
Seasonal differencing, D	$(1-B^s)^D$
Backshift operator, B	$By_t = y_{t-1}$

Moreover, autocorrelation function (ACF) and partial autocorrelation function (PACF) used to estimate the elements of p and q from ARIMA model [29]. However, the input is suggested to form in stationary with the process of differencing before ACF and PACF are applied.

Furthermore, time series model is also widely used on solar forecasting. A study of localized solar power prediction based on weather data from local history and global forecasts had been done by *Poolla et al*. [32]. This study aimed to forecast 18 hour ahead solar power output based on measured weather data and 18 hour ahead global forecast weather data from High Resolution Rapid Refresh (HRRR). ARX model developed by history weather measurements and exogenous weather forecasts. Then, ARX model compared to AR model that without the exogenous forecast. Root mean square error (RMSE), mean absolute error MAE and root mean square (RMS) used to verify the performance of models. As the result, the ARX model using both local history and global forecast has achieved in higher mean RMS of 80.07% while AR model using local history is 73.42%. This result indicated that the ARX model has higher accuracy than AR model.

A paper of time series ARIMA model for prediction of daily and monthly average global solar radiation: The case study of Seoul, South Korea had reported by *Mohammed H. Alsharif et al.* [33]. In this paper, seasonal ARIMA models were developed to predict monthly and daily solar radiation based on the dataset of solar radiation over the past 37 years. The structure of seasonal ARIMA was determined by ACF and PACF. The result showed that RMSE = 33.18 and R-squared = 79% for monthly solar radiation model while RMSE = 104.26 and R-squared = 68% for daily solar radiation model.

2.5 Hybrid approach

Hybrid model is a mix of two or more models together to achieve better forecast result. A study on forecast horizon and solar variability influences on the performances of multiscale hybrid forecast model had completed by *Monjoly et al.* [34]. In this study, multiscale hybrid forecast model (MHFM) model was developed based on a hybrid autoregressive (AR) and neural network (NN) model combined with multiscale decomposition methods for 1 hour ahead solar irradiance forecast. For the input, solar global radiation is used in different time interval. Then, the best performance of MHFM model achieved is rRMSE of 2.91% on clear sky days while rRMSE of 6.73% on cloudy sky days as the worst. Therefore, it is a challenge to forecast solar irradiance due to microclimates.

Based on the research with the title of SARIMA-SVM hybrid model for the prediction of daily global solar radiation time series, the author *Sabrina Belaid Boualit et al.* had been proposed a hybrid model based on seasonal autoregression integrated moving average (SARIMA) and support vector machine (SVM) to predict daily solar radiation [31]. As the result, normalized root mean square error (NRMSE) and correlation coefficient (R) achieved by the hybrid model are 14.529% and 0.874. which improved 0.384% and 0.007 from SARIMA model.

2.6 Global Forecast System (GFS)

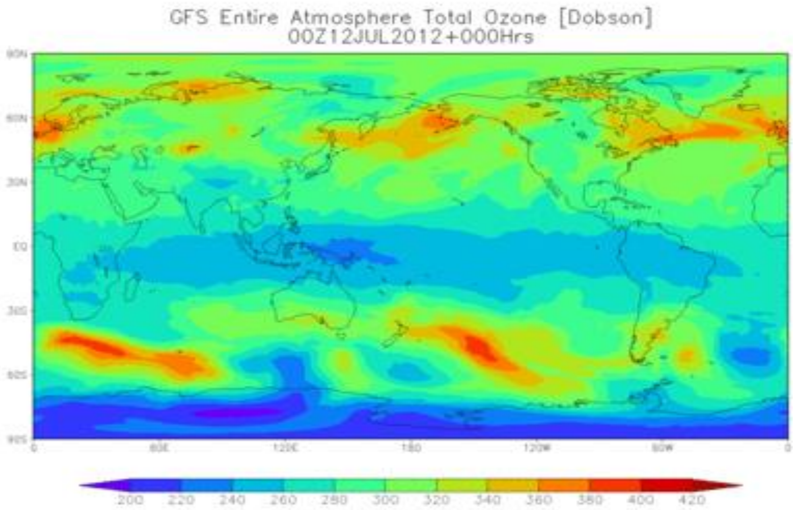

Figure 2.2: Global Forecast System (GFS) model [35]**.**

Global Forecast System (GFS) shown in Figure 2.2 is a weather forecast model produced by the National Centers for Environmental Prediction (NCEP). There are various of forecast data include of solar irradiance provided by GFS model. Also, GFS model had been used for research purposes.

According to *L. Martin* [36], the paper of global and direct normal irradiance forecast using GFS and statistical techniques had been studied. The aim of this research is to forecast hourly global solar irradiance (GHI) in different stations based on statistical downscaling with neural network (NN) and Global Forecast System (GFS). Then, the dataset of hourly solar irradiance obtained from the stations of Meteorological and Climatological Service of Navarra. Furthermore, relative mean bias deviation (rMBD) and relative root mean squared deviation (rRMSD) to measure the error of model. GFS model has showed lower rRMSD for the most of stations. However, there is huge improvement based on neural network model compared to GFS model for some stations.

A research on solar irradiance forecasting in the tropics using Numerical Weather Prediction and statistical learning had reported by *Verboi et al*. [37]. Weather Research and Forecasting (WRF) model with statistical techniques to generate day-ahead hourly solar irradiance. Dimensionality reduction, model output statistics (MOS) and removal of yearly and daily cycles are post-processing steps in the system. Smart persistence, climatological and Global Forecasting System (GFS) used to compare with WRF in the terms of RMSE and MAE. The result showed WRF-solar-PCA model has RMSE and MAE that are 1% and 5% lower than GFS, 23% and 22% lower than smart persistence, 10% and 15% lower than climatological.

2.7 Summary

In short, all the research paper reviewed the study on different forecast approaches. For example, statistical, machine learning, time series, hybrid and GFS model. The advantage and disadvantage for each forecast approach was summarized in Table 2.3. Therefore, the forecast techniques of multiple regression as statistics approach and neural network as machine learning will be chosen to build forecast models. Then, ARIMA as time series forecasting is suitable on weather forecast. With the weather forecast, then only the forecast models are able to forecast day ahead or an hour ahead. Then, from the hybrid, blended model will be formed by combining two or more models. Furthermore, the metrics such as RMSE and R are considered to validate forecast model while metric of WMAPE is used to evaluate forecast performance for all the models in different testing days.

Table 2-3: Advantage and disadvantage of different forecast approaches.

Forecast approach	Advantage	Disadvantage
Statistical	Simple to implement and transparent	Cannot apply too much inputs
Machine learning	Deal with various of data but opaque	Strong computation required
Time series	Only dependent variable required	Without the support of independent variables
Hybrid	Overcome limitation for a single model	Increased complexity of model
Global Forecast System (GFS)	Available and costless	Fixed forecast interval (3h)

CHAPTER 3

METHODOLOGY

3.1 Introduction

As the chapter of methodology, there is content of my work noted as following: data collection, overall flow chart, sub-flow chart for multiple regression, neural network and ARIMA models, principle of ACF and PACF, principle of weather forecasting for inter-day and intraday, also forecast blending. Then, each part of work involved will be explained in this chapter.

3.2 Data collection

There are two data sources collected and used in my project. The first data source is from the weather station located at faculty of electronic and computer engineering (FKEKK) which provides temperature, humidity, wind speed, pressure and solar irradiance.

Another data source is from the NOAA global forecast system (GFS) model which provides forecasting solar irradiance with 3-hour interval. GFS data is available on website https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs. Then, NOAA climate and weather toolkit (WCT) is used to read the GFS file due to the dataset of GFS is in form of grb2 file.

4 months of data (from 24 Oct 2019 to 24 Feb 2020) used to train models. New input data were chosen based on different forecast days. Four forecast days used which are 29 Feb, 5 Mar, 10 Mar, 12 Mar 2020. Lastly, each forecast day will be evaluated with measured value on that day.

3.3 Overall flow chart

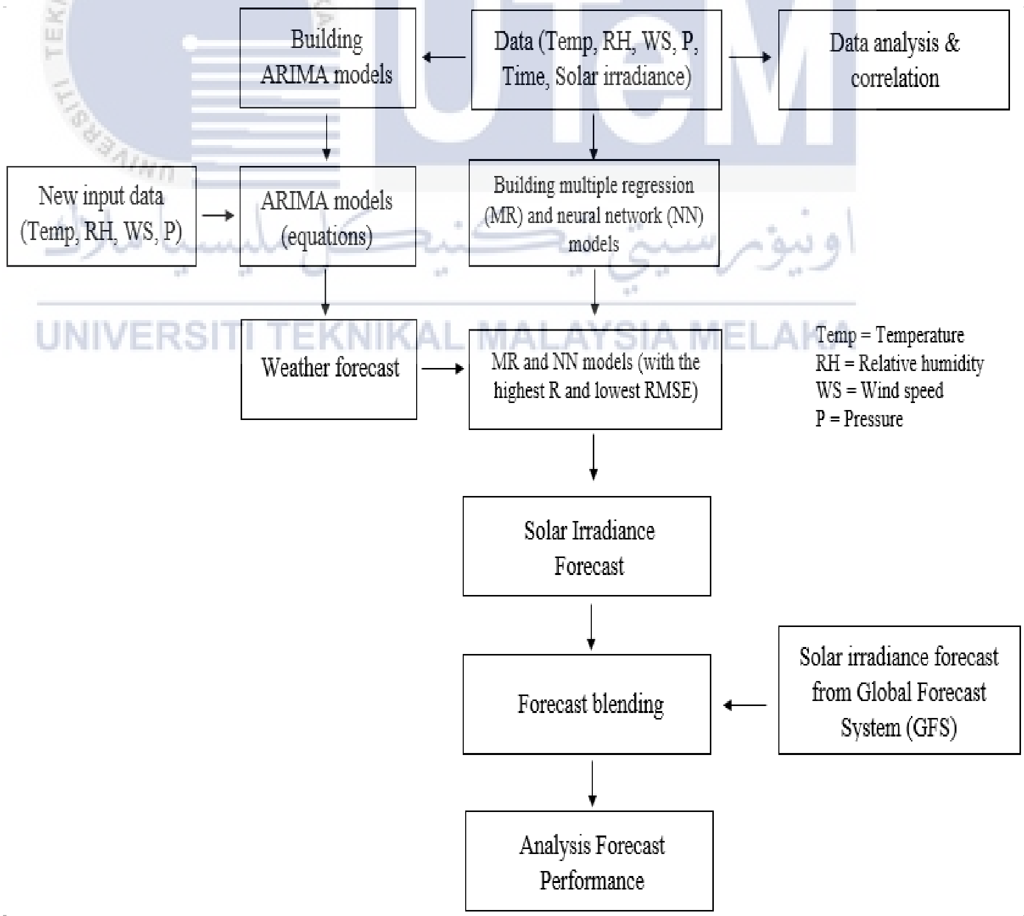

Figure 3.1: The overall flow chart.

In the Figure 3.1, the flow chart is start with the data which consists of temperature, humidity, wind speed, pressure, time (in decimal) and solar irradiance. Then, the data will flow into the parts of data analysis and correlation, building multiple regression (MR) and neural network (NN) models also ARIMA models. After building MR and NN models, only the best models will be chosen to forecast solar irradiance. After that, the purpose of ARIMA models is to produce forecast weather for inter-day and intraday. So, the weather forecast will then be imported into MR and NN models for one day ahead or an hour ahead forecast solar irradiance. Furthermore, the solar irradiance forecast from MR and NN models will be blended with the solar irradiance forecast from global forecast system (GFS) to form blended model. Also, the forecast performance for each model will be evaluated.

3.4 Data analysis and correlation

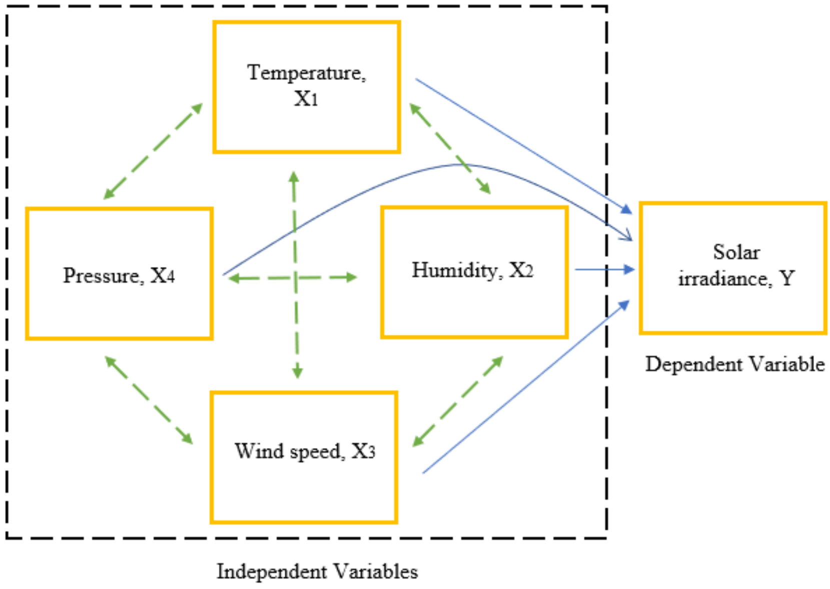

Figure 3.2: Data relationship between independent variables and dependent variable.

Based on Figure 3.2, there is step of data analysis and correlation will be carried out to understand the data. This is concerned on how the relationship between independent variables and dependent variable. Then, the purpose of analysis is to avoid the issue of multicollinearity. Multicollinearity occurs when two independent variables are similar and corelated to each other. When multicollinearity is present in models, the models may not be estimated precisely. Also, independent variable that has strong relationship with dependent variable which is useful for the forecast model.

3.5 Multiple regression model

In multiple regression model, this is process of modelling with the relationship of independent variables and dependent variable stated in literature review. Then, the flow chart of multiple regression model is shown as Figure 3.3.

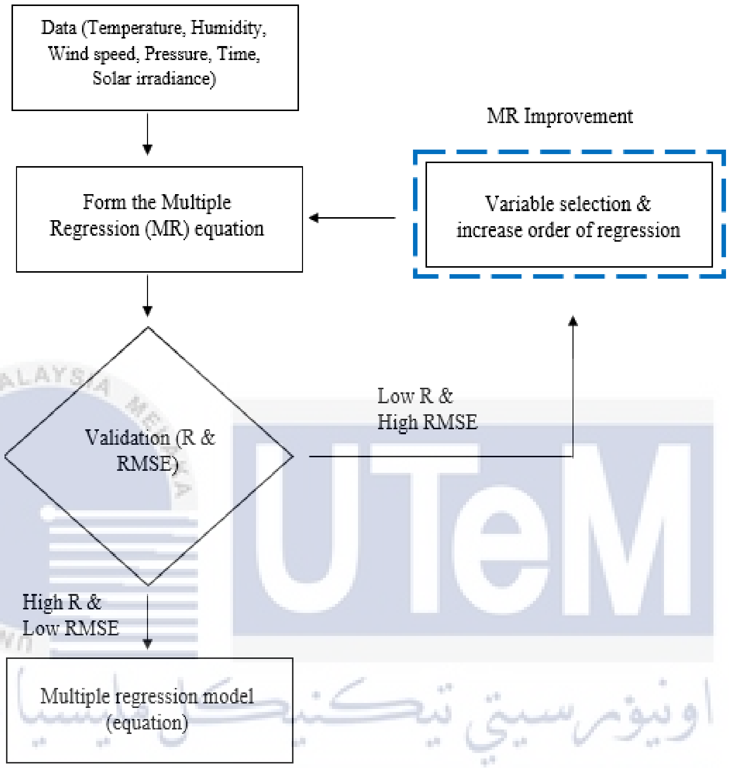

Figure 3.3: Flow chart (multiple regression model).

Figure 3.3 presents flow chart of multiple regression (MR). In the beginning, the data of temperature, humidity, wind speed, pressure and time will be imported to build the MR model (equation). Then, the next step is to validation model using R and RMSE. After that, the loop will keep repeating if there are low R and high RMSE achieved. Inside the loop, there is improvement by increasing the order of regression. Also, select the different variables to generate the different multiple regression models until the model is achieved its best validation result.

3.6 Neural network model

Feed-forward neural network is used as the structure of neural network model. Levenberg-Marquardt algorithm is used as backpropagation learning algorithm. After that, 80% of input data for training and rest of 20% for validation. Then, the flow chart neural network is shown in Figure 3.4.

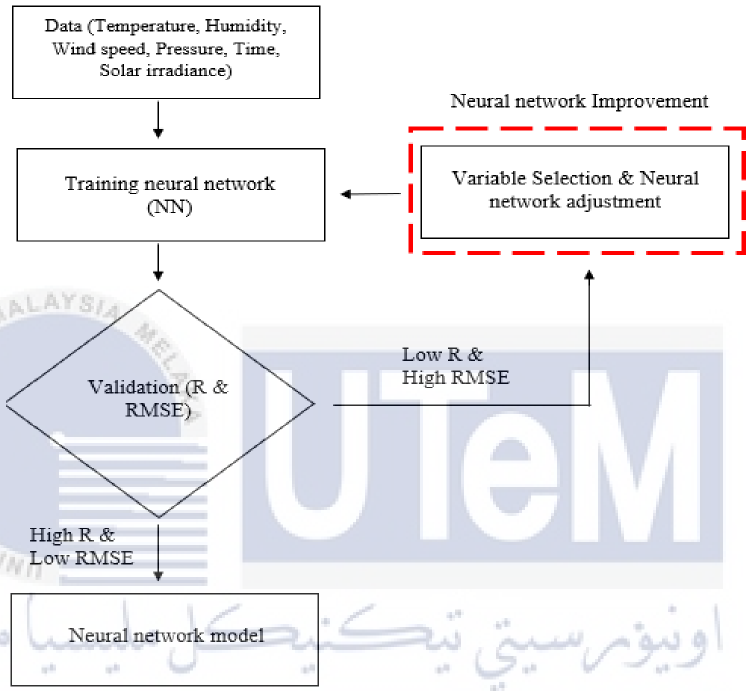

Figure 3.4: Flow chart (neural network model).

The flow chart of neural network model shows in Figure 3.4 is definitely similar to the Figure 3.3. The only difference is the neural network improvement although the part of variable selection is the same. Then, besides of variable selection, neural network will be adjusted with the number of hidden layers in order to achieve the highest R and lowest RMSE as usual. Hence, only the neural network model with the best validation is used to forecast.

3.7 ARIMA model

In this part, there is flow chart of ARIMA in Figure 3.5. Then, following by the principle of PACF and ACF. Also, the method to weather forecast for inter-day and intraday.

3.7.1 Flow chart of ARIMA model

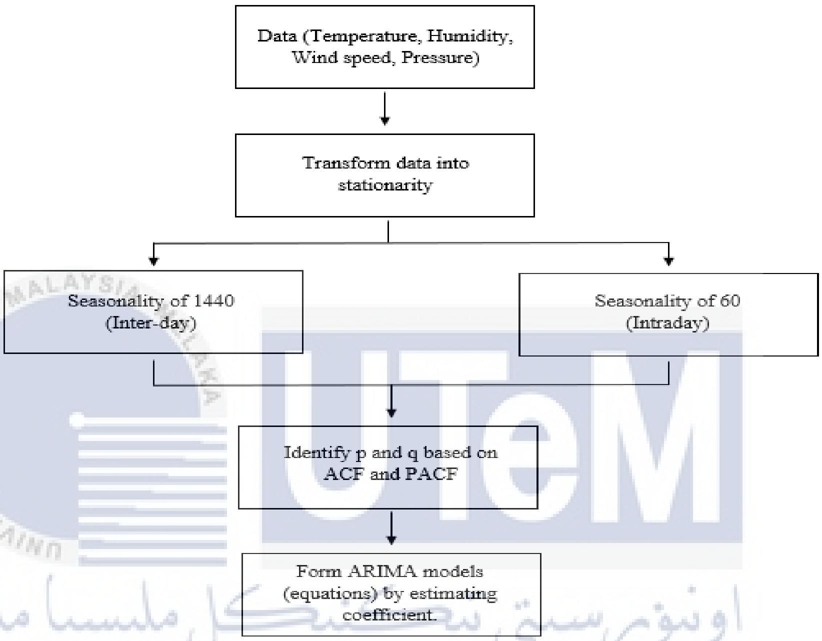

Figure 3.5: Flow chart of seasonal ARIMA.

Figure 3.5 shows the flow chart of seasonal ARIMA model. Then, the first step is to transform data into stationarity by using differencing. Differencing is to remove trend and seasonal from the data. After that, seasonality of model stated as 1440 for inter-day while 60 for intraday. Then, the graphs of ACF and PACF will be plotted to identify the terms of p and q according to the principle of ACF and PACF in Table 3.1. After that, the coefficients of ARIMA equation will be estimated based on their corresponding parameters of temperature, humidity, wind speed and pressure. Finally, 8 seasonal ARIMA equations are expected to be formed which 4 equations for both inter-day and intraday.

3.7.2 Principle of ACF and PACF

Table 3-1: Principle of ACF and PACF [29]**.**

Model order	Autocorrelation (ACF)	Partial autocorrelation (PACF)
AR (p)	Decays	Cuts off after lag p
MA (q)	Cuts off after lag q	Decays
ARMA (p, q)	Decays	Decays

According to the principle of ACF and PACF shown in Table 3.1, the terms of p and q for ARIMA model is defined based on the graphs of ACF and PACF. Also, the terms of P and Q (seasonal) followed as same as the terms of p and q.

3.7.3 Method to weather forecast (Inter-day)

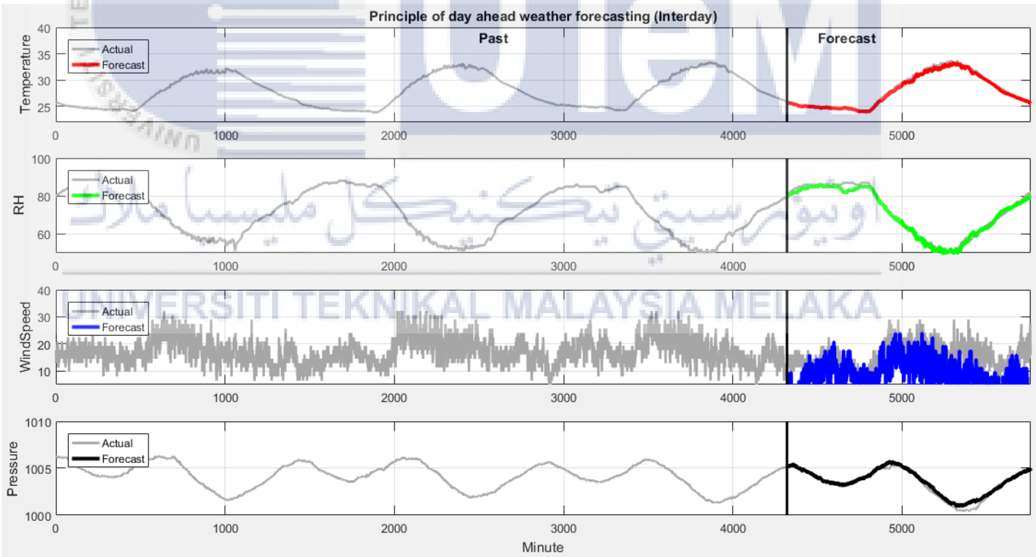

Figure 3.6: Weather forecast (Inter-day).

Figure 3.6 shows that the weather forecast for inter-day. There are left and right parts shown in the Figure 3.6. The left part stated as past value while the right part stated as forecast value. Three previous days used as part data to forecast the weather of the next day. For instance, if the forecast day is 29 Feb 2020, then three previous days used are from 26 to 28 Feb 2020.

3.7.4 Method to weather forecast (Intraday)

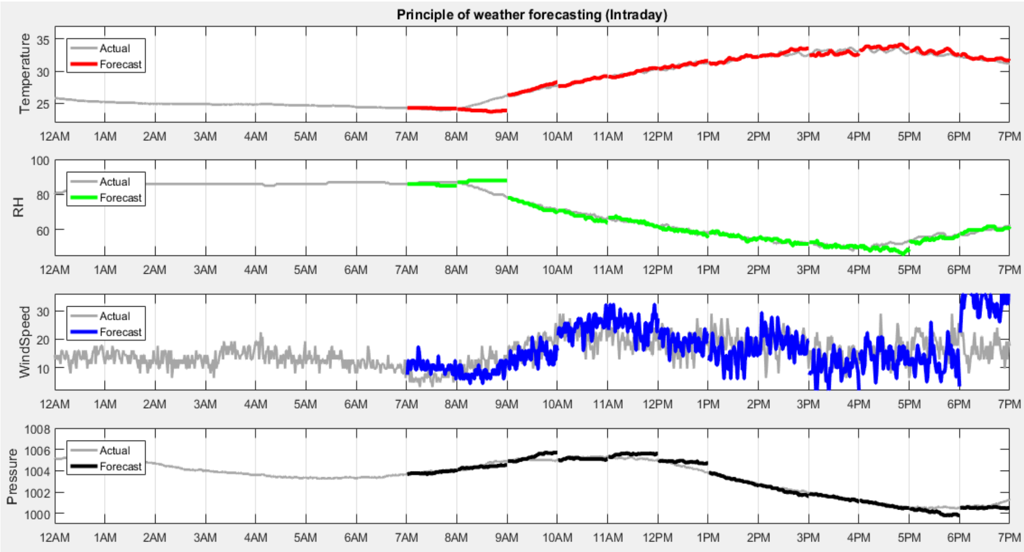

Figure 3.7: Weather forecast (Intraday).

Figure 3.7 shows weather forecast for intraday. The input data from 12AM to 7AM used to forecast first hour which is 7 – 8 AM. After go through an hour, the actual data from 7 – 8 AM is collected. To forecast the following hour (8 – 9AM), the input data will be updated to 12 – 8AM by adding cumulatively new actual data. Then, the input data will update repeatedly hour by hour to forecast every hour ahead until 7PM.

3.8 Forecast blending

Since the data of forecasting solar irradiance from GFS model has gap of 3-hour interval. So, in order to make the GFS data blended with other models, GFS data is required to convert to the amount as same as other models. Then, the method of interpolation implemented on GFS data. Interpolation is the process to fulfill the missing points within two known points. By applying interpolation, GFS data with 3-hour interval is converted to data with 1-min interval. After that, it can only proceed to the next step when the GFS data is with the 1-min interval. The next step, there is addition for each value of multiple regression (MR), neural network (NN) and GFS models. Therefore, each average value is plotted to form as blended model.

3.9 Performance metric

There are some performance metrics used in my project for the purpose of evaluating data analysis, model validation and forecast accuracy. Then, the first metric is variance inflation factor (**VIF**) which estimates multicollinearity between independent variables [14], [38]. The formula of VIF as shown in Eq. (3.1) where R = correlation coefficient.

$$VIF = \frac{1}{1 - R^2} \tag{3.1}$$

In order to compute VIF, then correlation coefficient should be calculated first. So, the next metric is correlation coefficient (**R**) which measures the correlation between two variables. R is not only applied between independent variables but also dependent variable. The higher R indicates the stronger correlation. Moreover, R is also used in model validation. Then, the R equation is presented as Eq. (3.2) where x = input variable and y = targeted variable.

$$R = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \tag{3.2}$$

Besides that, there is also root mean square error (**RMSE**) used to validate model. For the better model, the RMSE is desired as low as possible. The formula of RMSE is shown in Eq. (3.3) where y_i = actual value and $\hat{y}_i$ = forecast value.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{3.3}$$

After that, weighted mean absolute percentage error (**WMAPE**) as stated in [39] is the metric used to measure the forecast performance. The denominator of WMAPE is based on summation of actual value which is able to handle zero value. The formula of WMAPE is shown as Eq. (3.4) where y_i = actual value and $\hat{y}_i$ = forecast value.

$$WMAPE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{\sum_{i=1}^{n} y_i} \tag{3.4}$$

3.10 Summary

As conclusion, there are several steps to conduct in order to complete this project. So, the first step is data analysis is a way to understand data then determine the variable. Then, the next step is model validation to find out the best model. After that, seasonal ARIMA used to produce weather forecast. Also, the process of forecast blending is described. Then, the last step is the performance metrics are defined.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction

This chapter will focus the result of data analysis, validation for multiple regression and neural network and ARIMA models also forecast performance. Then, model validation is measured by R and RMSE. Next, WMAPE is evaluated the forecast performance for all models with different testing days (inter-day and intraday). To under the forecast performance, then performance metrics used to evaluate the results.

4.2 Results for data analysis and correlation

There are two parts conducted for data analysis and correlation. The first part is data analysis and correlation between independent variables. Then, second part is the data analysis and correlation between independent variables and dependent variable. Independent variables are temperature, humidity, wind speed and pressure while dependent variable is solar irradiance.

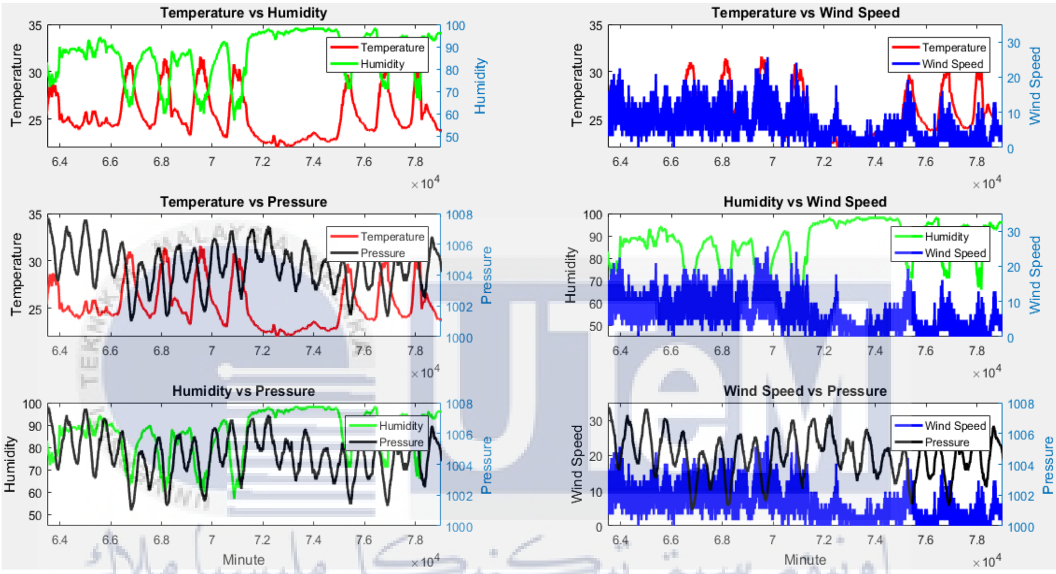

Figure 4.1: The data analysis between independent variables (part of data).

Figure 4.1 shows that the data analysis between independent variables with only part of data. The reason of the result presented as part of data due to the better view for the graph analysis. Based on the observation, the relationship between temperature and humidity is the most significant result which is inversely related to each other. Besides of that, it is difficult to recognize the relationship for the rest of the independent variables.

Table 4-1: R and VIF between independent variables.

Parameter	R	VIF
Temperature & Humidity	-0.8713	4.1519
Temperature & Wind Speed	0.2452	1.0639
Temperature & Pressure	-0.2544	1.0692
Humidity & Wind Speed	-0.3773	1.1659
Humidity & Pressure	0.1994	1.0414
Wind Speed & Pressure	0.1071	1.0116

Table 4.1 shows R and VIF between independent variables. Based on the range of VIF stated in Table 2.1, the relationship between temperature and humidity has the highest VIF of 4.1519 which means moderately correlated. For the rest of independent variables, the values of VIF are close to 1 which means not correlated. Fortunately, all the parameters are not exceeded VIF of 5 therefore all can be used for modelling.

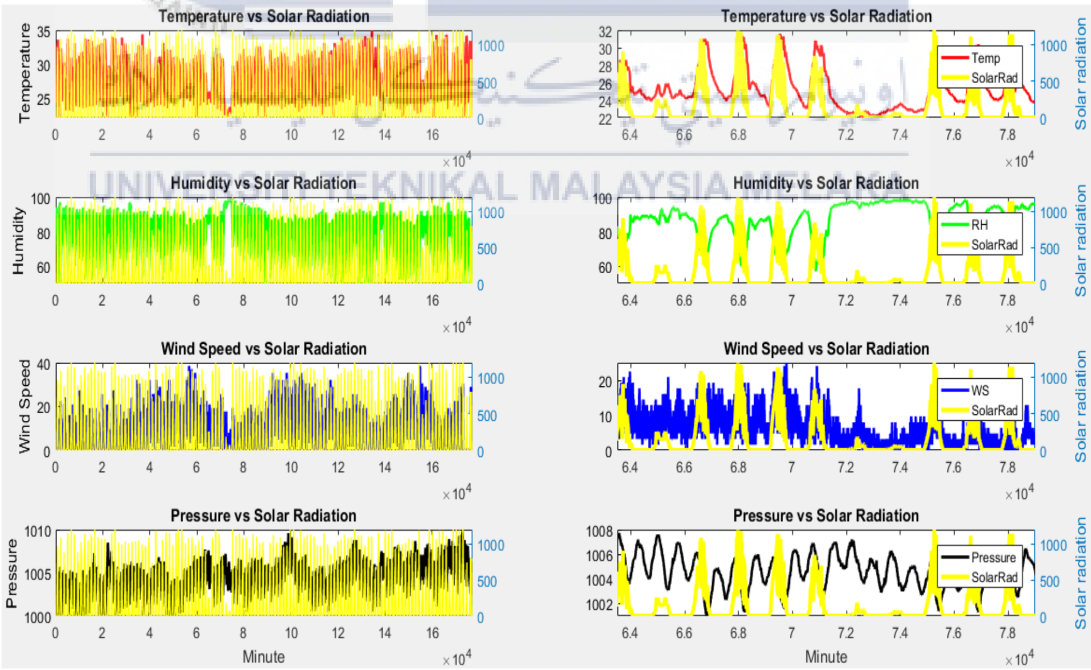

Figure 4.2: The data analysis between independent variables and dependent variable. Left column: Overall data. Right column: Part of data.

Figure 4.2 shows that the data analysis between dependent variable and independent variables. There are two columns which left column shows overall data view while the right column shows zoom in view from overall data. The reason that the additional part of zoom in view due to better view for data analysis. So, the relationship result can be observed and explained. As the result, there is high solar irradiance when temperature is high. Besides that, humidity is inverse of temperature therefore solar irradiance will drop as humidity rise. Then, wind speed is not steady related to solar irradiance. Sometime high wind speed come with high solar irradiance but sometime not. For the pressure, it can be observed that the solar irradiance rises at the moment when pressure drop from its peak.

Table 4-2: R and RMSE for independent variables and dependent variable.

Parameter	R	RMSE
Temperature & Solar Irradiance	0.6884	327.2544
Humidity & Solar Irradiance	-0.6878	313.7555
Wind Speed & Solar Irradiance	0.3552	336.6058
Pressure & Solar Irradiance	8.01e-04	865.0954

In the Table 4.2, the relationship between temperature and solar irradiance has the highest R of 0.6884. This is because strong solar light which causes higher temperature. So, temperature is mostly related to solar irradiance. Other than that, the lowest R of 8.01e-04 and the highest RMSE of 865.9054 is the relationship of pressure and solar irradiance.

4.3 Model validation

R and RMSE used to validate the multiple regression and neural network model. The validation of models with different combination of parameters will be recorded. The model with the best validation result is chosen as the final model. In the end, the form of equation for multiple regression model and the final structure for neural network model will be shown.

4.3.1 Result of validation for multiple regression model

Table 4-3: Validation for multiple regression model with different parameters.

Parameter	1st order (Linear)		2nd order (Quadratic)	
	R	RMSE	R	RMSE
Temp	0.6885	208	0.7148	200
Temp, RH	0.7113	201	0.7197	199
Temp, RH, WS	0.7246	197	0.7589	187
Temp, RH, WS, Pressure	0.7403	193	0.8000	172
Temp, RH, WS, Pressure, Time	0.7642	185	**0.8815**	**135**

Based on the Table 4.3, it shows the result for the model validation of multiple regression. With the increase of parameters, then R is increased while RMSE is decreased. There is also improvement from 1st order (linear) to 2nd order (quadratic) regression. As the result, R is increased from 0.7642 to 0.8815 and RMSE is decreased from 185 to 135 with all the parameters.

4.3.1.1 Result of multiple regression model (equation)

$$\begin{aligned} y = {} & 4005X_1 + 1185X_2 - 1005.1X_3 - 186.6X_4 - 45860X_5 - 0.43917X_1X_2 \\ & + 2.8699X_1X_3 - 3.9952X_1X_4 - 104.36X_1X_5 \\ & + 0.0072604X_2X_3 - 1.1856X_2X_4 + 23.461X_2X_5 \\ & + 0.93005X_3X_4 - 1.2566X_3X_5 + 46.991X_4X_5 + 2.7276{X_1}^2 \\ & + 0.027146{X_2}^2 - 0.12369{X_3}^2 + 0.18551{X_4}^2 - 855.27{X_5}^2 \end{aligned} \tag{4.1}$$

With the estimated of coefficients, the equation of multiple regression model as shown in Eq. (4.1) where X_1 = temperature, X_2 = humidity, X_3 = wind speed, X_4 = pressure, X_5 = time and y = solar irradiance.

4.3.2 Result of validation for neural network model

Table 4-4: Validation for neural network model with different parameters.

Parameter	R	RMSE
Temp	0.7278	195.3103
Temp, RH	0.7298	194.9912
Temp, RH, WS	0.7709	181.4396
Temp, RH, WS, Pressure	0.8308	159.8094
Temp, RH, WS, Pressure, Time	0.9067	121.0106
With only 1 hidden layer that consists of 5 neurons.		

From the Table 4.4, this is the validation result of neural network model. For this validation, the neural network is set to only 1 hidden layer that consists of 5 neurons. As the result, the model is getting better as the increase of parameters. With only the parameter of temperature, R of 0.7278 and RMSE of 195.3103 achieved by neural network. Then, the neural network with all parameters has the highest R of 0.9067 and the lowest RMSE of 121.0106.

Table 4-5: Model validation with different number of hidden layers.

Parameter	No. of hidden layer	No. of neuron	R	RMSE
Temp, RH, WS, Pressure, Time	1	5 (for each hidden layer)	0.9067	121.0106
	2		0.9153	116.3766
	3		**0.9173**	**114.1820**

Table 4.5 shows the validation of model with all the parameters and different number of hidden layers. Each hidden layer has same amount of neuron which is 5 neurons. From the table previous, the neural network is validated with 1 hidden layer that consists of 5 neurons. Now, neural network is validated by increasing of hidden layer. In the result, that the neural network has achieved the highest R of 0.9173 and the lowest RMSE of 114.1820 with 3 hidden layers. As the more hidden layer, the deeper neural network, then the validation result is better. By the comparison of multiple regression and neural network models, the validation result for neural network is higher than multiple regression. This may explain that the neural network based on the backpropagation algorithm while multiple regression is based the best fit line that regressed by all the parameters.

4.3.2.1 The final structure of neural network model

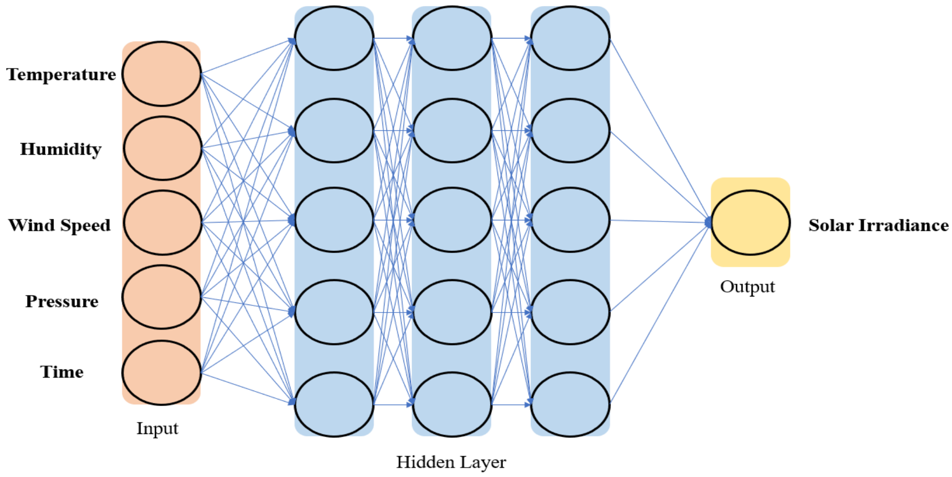

Figure 4.3: The final structure of neural network.

Figure 4.3 shows that the final structure of neural network model. It consists 3 hidden layers and each hidden layer has 5 neurons. Total of 15 neurons in the neural network. The structure of neural network as shown in Figure has the highest R and lowest RMSE so it is chosen as the best neural network model used to forecast solar irradiance.

4.4 Results of ACF and PACF (Inter-day)

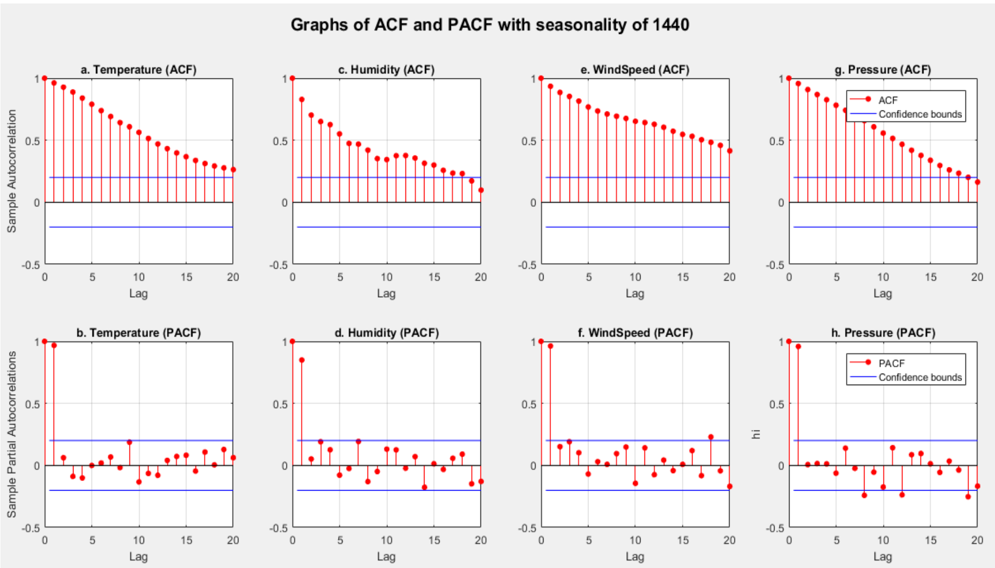

Figure 4.4: ACF and PACF with seasonality of 1440 (Inter-day).

Figure 4.4 shows the ACF and PACF plots for weather parameters with seasonality of 1440. For the results, ACF plots show that those lags are towards zero. Also, PACF plots with the only significant point on lag (1). As conclude, AR (1) is chosen to use in ARIMA models for seasonality of 1440.

4.5 Results of ACF and PACF (Intraday)

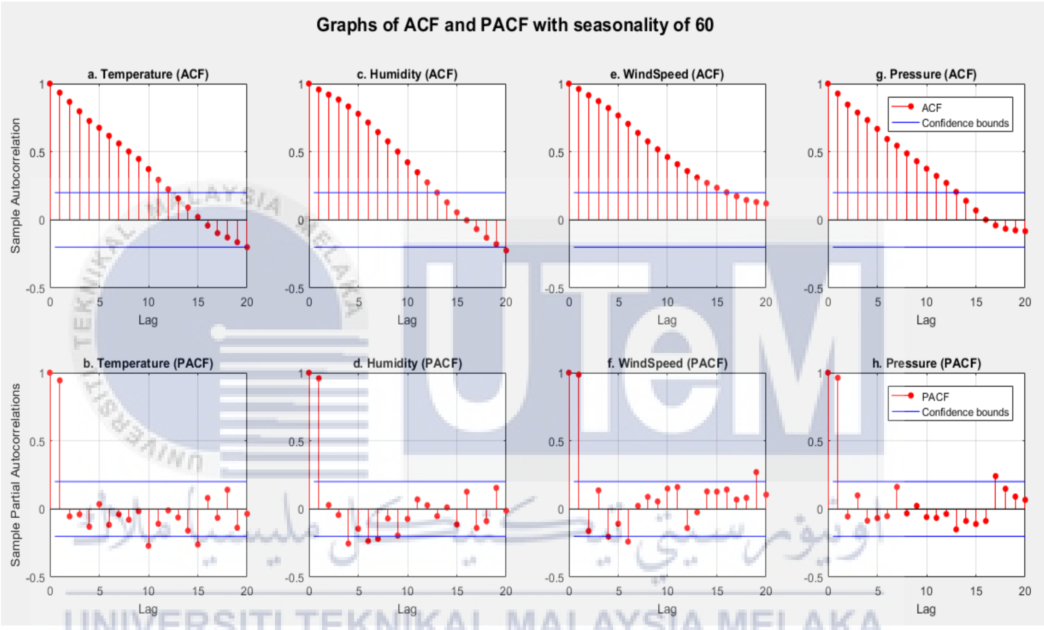

Figure 4.5: ACF and PACF with seasonality of 60 (Intraday).

Figure 4.5 shows the result of ACF and PACF plots for each weather parameters with seasonality of 60. The upper row is all about ACF plots while lower row is PACF plots. ACF plots have the behavior towards zero even there are some negative lags. Then, PACF plots show the significant points on some lags. However, the lag (1) is still the most significant value. So, it can be concluded that AR (1) is suitable to use in ARIMA.

4.6 ARIMA equation

Based on the differencing process and the result of ACP and PACF in Figure 4.4 and 4.5, the final ARIMA models for inter-day and intraday are known as $SARIMA\ (1,1,0)(1,1,0)_{1440}$ and $SARIMA\ (1,1,0)(1,1,0)_{60}$. Then, the equations of SARIMA model with their specific coefficients are shown in below.

4.6.1 ARIMA equation (Inter-day)

$$(1-(-0.1755)B)(1-\ (-0.1755)B^{1440})(1-B)^{1}(1-B^{1440})^{1}y_t = 0.3040 \tag{4.2}$$

$$(1-(-0.0462)B)(1-\ (-0.0462)B^{1440})(1-B)^{1}(1-B^{1440})^{1}y_t = 0.7999 \tag{4.3}$$

$$(1-(-0.1882)B)(1-\ (-0.1882)B^{1440})(1-B)^{1}(1-B^{1440})^{1}y_t = 12.1441 \tag{4.4}$$

$$(1-(-0.2463)B)(1-\ (-0.2463)B^{1440})(1-B)^{1}(1-B^{1440})^{1}y_t = 0.0062 \tag{4.5}$$

For inter-day weather forecasting, $SARIMA\ (1,1,0)(1,1,0)_{1440}$ for temperature, humidity, wind speed and pressure which are represented by Eq. (4.2), (4.3), (4.4) and (4.5).

4.6.2 ARIMA equation (Intraday)

$$(1-(-0.1738)B)(1-\ (-0.1738)B^{60})(1-B)^{1}(1-B^{60})^{1}y_t = 0.3254 \tag{4.6}$$

$$(1-(-0.0434)B)(1-\ (-0.0434)B^{60})(1-B)^{1}(1-B^{60})^{1}y_t = 0.8616 \tag{4.7}$$

$$(1-(-0.1882)B)(1-\ (-0.1882)B^{60})(1-B)^{1}(1-B^{60})^{1}y_t = 12.2030 \tag{4.8}$$

$$(1-(-0.2433)B)(1-\ (-0.2433)B^{60})(1-B)^{1}(1-B^{60})^{1}y_t = 0.063 \tag{4.9}$$

For intraday weather forecasting, $SARIMA\ (1,1,0)(1,1,0)_{60}$ for temperature, humidity, wind speed and pressure which are represented by Eq. (4.6), (4.7), (4.8) and (4.9).

4.7 Results of solar irradiance forecasting (Inter-day)

Four different testing days which are 29 Feb 2020, 5 Mar 2020, 10 Mar 2020, 12 Mar 2020 chosen to evaluate the forecast performance for each model based each model based on the metric of weighted mean absolute percentage error (WMAPE).

4.7.1 Solar irradiance forecasting on 29 Feb 2020 (Inter-day)

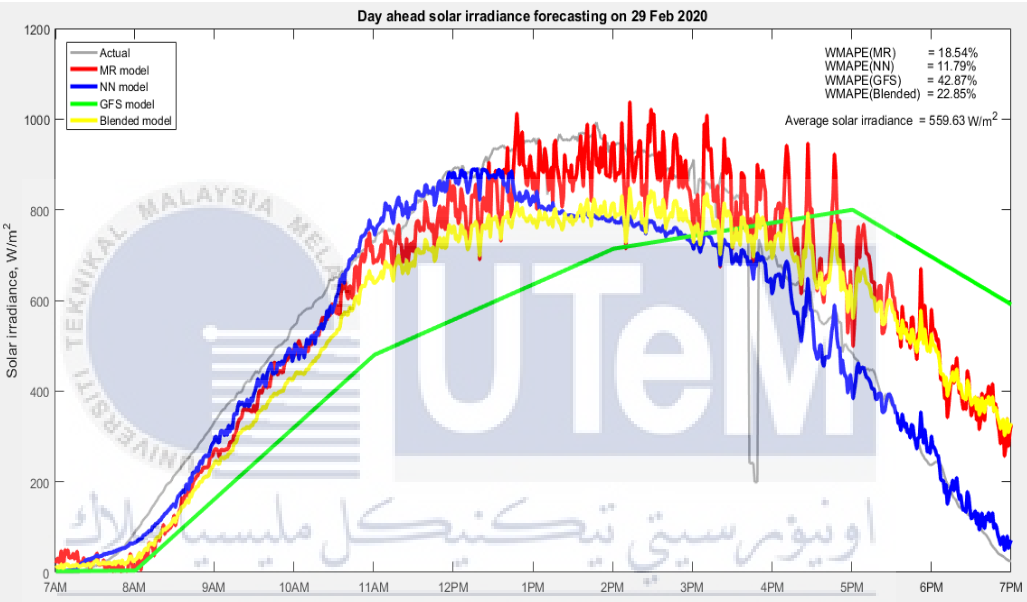

Figure 4.6: Day ahead solar irradiance forecasting on 29 Feb 2020.

In the Figure 4.6, WMAPE for neural network (NN) model is 11.79% which 6.75% lower than multiple regression (MR) model. Besides that, WMAPE for GFS and blended models which are 42.87% and 22.85% respectively. The pattern of solar irradiance acts as normal on this day. Therefore, neural network model is the best model used to forecast on sunny day with normal solar irradiance.

4.7.2 Solar irradiance forecasting on 5 Mar 2020 (Inter-day)

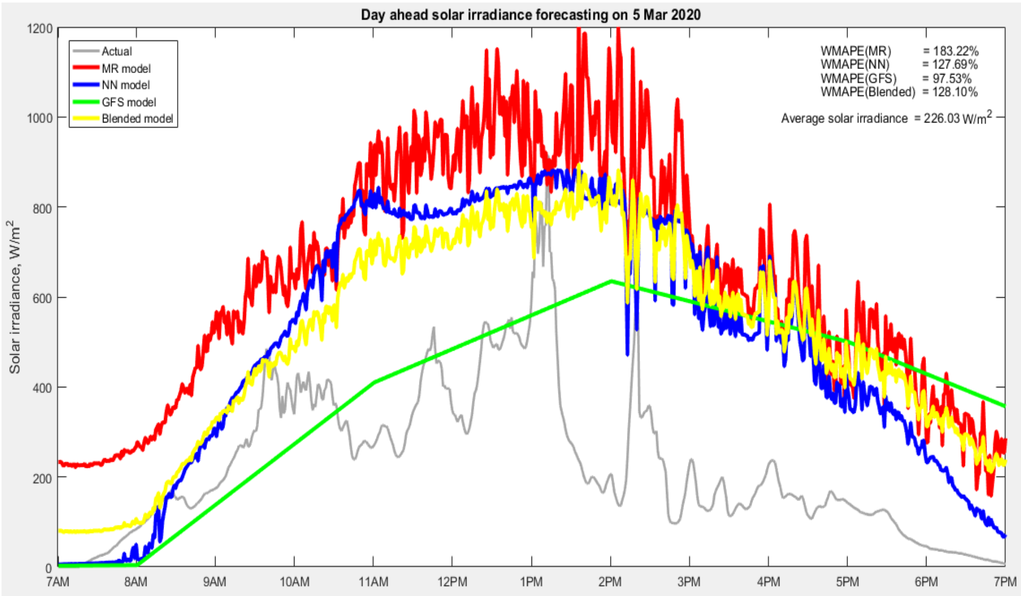

Figure 4.7: Day ahead solar irradiance forecasting on 5 Mar 2020.

From the Figure 4.7, none of the models has performed a good forecasting on this day. Even the lowest WMAPE for GFS model is 97.53% which is close to 100%, then WMPAE for the rest of models are exceeded 100%. This is because of unexpected low solar irradiance due to heavy rainy day or overcast day. So, this result indicates that the models is not able to forecast on the extreme day with low solar irradiance.

4.7.3 Solar irradiance forecasting on 10 Mar 2020 (Inter-day)

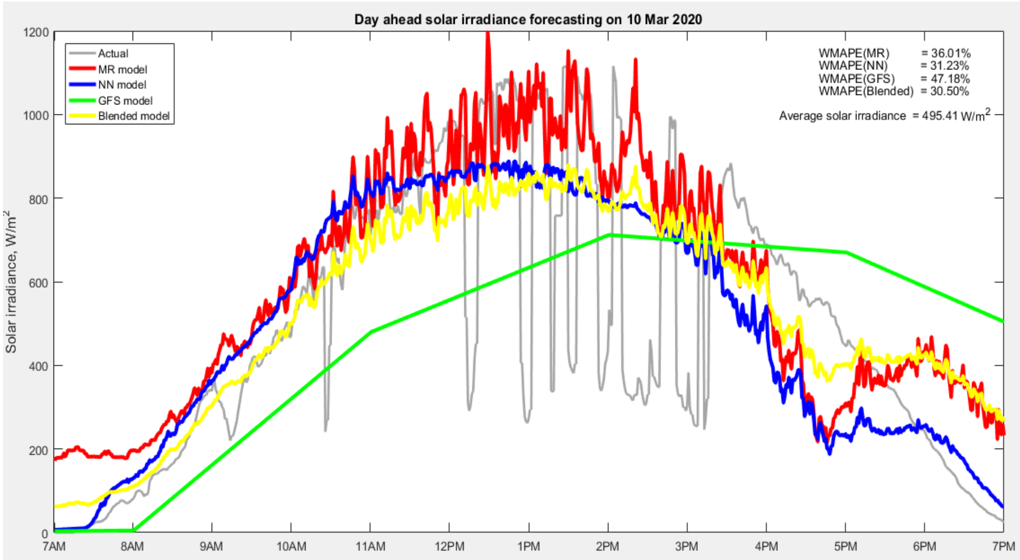

Figure 4.8: Day ahead solar irradiance forecasting on 10 Mar 2020.

Based on the Figure 4.8, WMAPE for neural network (NN) model is 31.23% which 4.78% lower than multiple regression (MR) model. Even though neural network (NN) model is better compared to multiple regression (MR) model, but the blended model has the best performance which WMAPE is 30.50%. There is the fluctuation of solar irradiance on this day which means the average solar irradiance value is lower than normal. Then, blended model has the benefit on moderate solar irradiance since it is the combination of neural network (NN), multiple regression (MR) and GFS models. Overall, blended model has showed a good forecast performance on cloudy day with medium solar irradiance value.

4.7.4 Solar irradiance forecasting on 12 Mar 2020 (Inter-day)

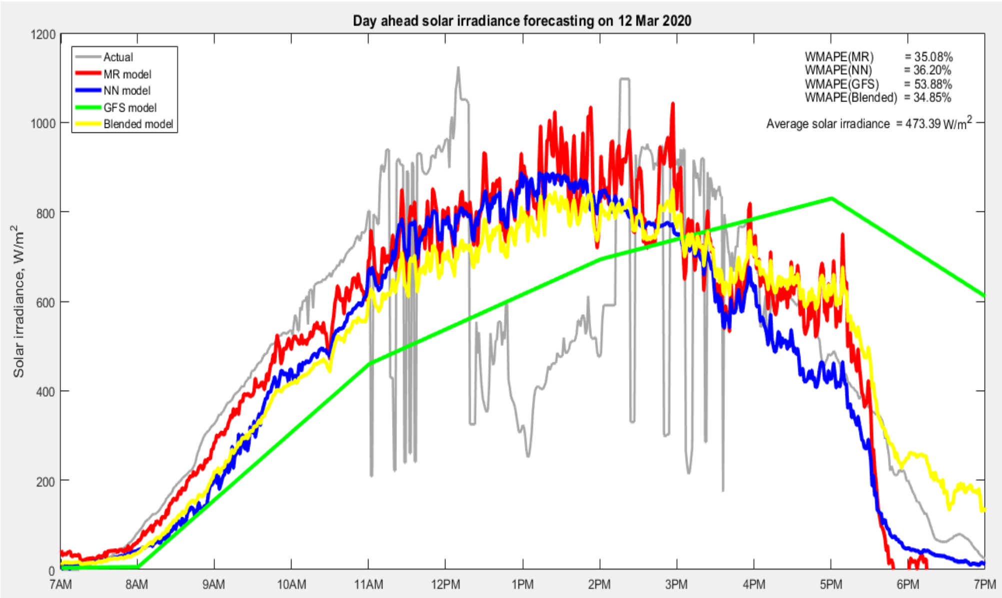

Figure 4.9: Day ahead solar irradiance forecasting on 12 Mar 2020.

In the Figure 4.9, the forecast performance for multiple regression (MR) model is slightly better than neural network (NN) model. WMAPE for multiple regression (MR) model is 35.08% while 36.20% for neural network (NN) model. However, blended model has the lowest WMAPE which is 34.85%. The pattern of solar irradiance on this day is similar to the Figure. The only difference is that there is sudden drop of solar irradiance at peak of the day. So, it can be defined as the day with medium low solar irradiance. As the result, blended model has better forecast performance compared to other models on the day with medium low solar irradiance.

4.8 Results of solar irradiance forecasting (Intraday)

Since there is extreme day that difficult to forecast accurately for inter-day. Then, this scenario of intraday for the purpose of solving the inaccurate forecast on the extreme day with low solar irradiance. The testing days are followed as the forecast days used for inter-day.

4.8.1 Solar irradiance forecasting on 29 Feb 2020 (Intraday)

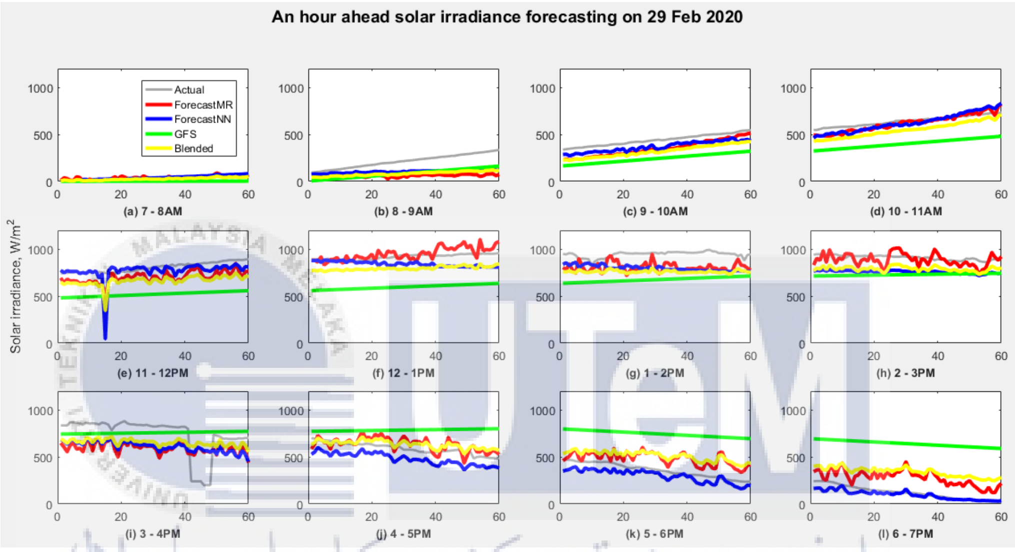

Figure 4.10: An hour ahead solar irradiance forecasting on 29 Feb 2020.

Table 4-6: Forecast performance for each model in different hour (29 Feb 2020).

Time	WMAPE (%)			
	MR model	NN model	GFS model	Blended model
7-8 AM	86.24	65.02	93.30	**52.84**
8-9 AM	72.39	**51.72**	60.54	61.55
9-10 AM	20.73	**16.04**	44.58	27.11
10-11 AM	**5.88**	7.11	37.64	14.80
11-12 PM	12.91	**6.30**	35.51	17.58
12-1 PM	**5.26**	8.42	34.87	13.27
1-2 PM	**14.62**	15.67	28.45	19.56
2-3 PM	**5.91**	17.05	20.00	13.21
3-4 PM	30.21	28.62	**18.36**	24.55
4-5 PM	10.98	18.75	32.98	**8.58**
5-6 PM	34.62	**16.86**	107.87	41.97
6-7 PM	132.57	**23.44**	433.62	182.20

Figure 4.10 presents an hour ahead solar irradiance forecasting on 29 Feb 2020. In each forecast hour, the forecast performance for each model is summarized in Table 4.6. There are five times of the lowest WMAPE for neural network (NN) model which is varied between 6.30% to 51.72%. Other than that, the times of the lowest WMAPE achieved for multiple regression (MR), GFS and blended models which are four, one and two respectively. This could be explained that the neural network (NN) model is not only perform well on whole day as showed in Figure but also hourly forecast within the day. In short, neural network (NN) model has the best forecast performance on 29 Feb 2020.

4.8.2 Solar irradiance forecasting on 5 Mar 2020 (Intraday)

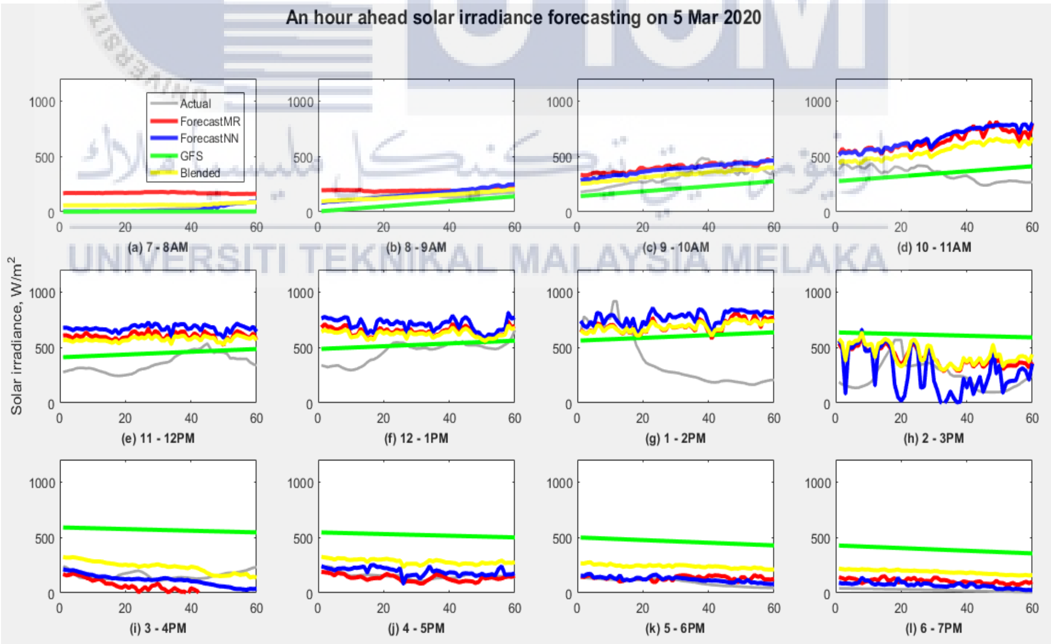

Figure 4.11: An hour ahead solar irradiance forecasting on 5 Mar 2020.

Table 4-7: Forecast performance for each model in different hour (5 Mar 2020).

Time	WMAPE (%)			
	MR model	NN model	GFS model	Blended model
7-8 AM	473.44	**34.81**	93.85	120.60
8-9 AM	37.73	21.94	48.41	**13.17**
9-10 AM	27.81	23.35	34.89	**15.38**
10-11 AM	88.45	94.91	**26.41**	61.61
11-12 PM	67.52	87.05	**26.89**	59.79
12-1 PM	38.50	53.02	**15.55**	35.01
1-2 PM	92.02	104.59	**80.16**	91.21
2-3 PM	**75.52**	79.83	151.10	86.67
3-4 PM	90.10	**36.11**	258.88	56.68
4-5 PM	**16.63**	27.61	237.69	83.00
5-6 PM	42.91	**27.23**	380.24	148.63
6-7 PM	328.65	**178.25**	1472.43	659.78

Figure 4.11 shows an hour ahead solar irradiance forecasting on 5 Mar 2020. The forecast performance for each model calculated and stated in Table 4.7. As the result, both neural network (NN) and GFS models have performed four times of the lowest WMAPE. Besides that, multiple regression (MR) and blended models have achieved only two times of the lowest WMPAE. This outcome indicates that the neural network is still not able to forecast low solar irradiance from 10AM to 1PM (peak of the day). Fortunately, GFS models has provided better performance to overcome this kind of situation. Overall, it seems like neural network (NN) and GFS models can be applied together for getting a good forecast on 5 Mar 2020.

4.8.3 Solar irradiance forecasting on 10 Mar 2020 (Intraday)

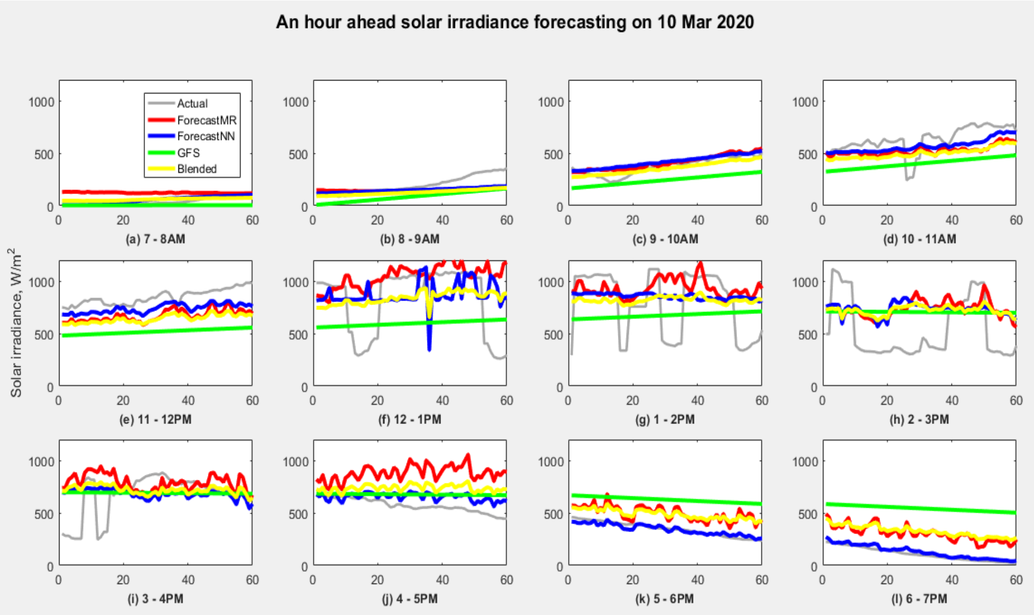

Figure 4.12: An hour ahead solar irradiance forecasting on 10 Mar 2020.

Table 4-8: Forecast performance for each model in different hour (10 Mar 2020).

Time	WMAPE (%)			
	MR model	NN model	GFS model	Blended model
7-8 AM	297.10	**70.35**	92.21	97.31
8-9 AM	35.59	**28.71**	55.96	35.72
9-10 AM	**9.70**	10.28	35.99	10.98
10-11 AM	21.12	**15.47**	38.41	24.80
11-12 PM	21.15	**12.84**	38.67	24.18
12-1 PM	**30.33**	32.90	45.40	32.59
1-2 PM	**32.91**	33.35	40.04	34.15
2-3 PM	59.36	59.92	**58.78**	59.29
3-4 PM	24.64	23.85	21.31	**21.14**
4-5 PM	53.85	**15.45**	18.06	28.68
5-6 PM	42.21	**6.32**	80.33	40.49
6-7 PM	181.97	**19.92**	402.33	201.37

Figure 4.12 shows an hour ahead solar irradiance forecasting on 10 Mar 2020. The forecast performance for each model on 10 Mar 2020 noted in Table 4.8. Neural network (NN) model has seven times of the lowest WMAPE which is varied between 6.32% to 70.35% while multiple regression (MR) model is three times. Besides that, both GFS and blended models have only one time of the lowest WMAPE over this day. By comparing the times of the WMAPE, it can be concluded that the neural network (NN) model has better forecast performance compared to other models on 10 Mar 2020.

4.8.4 Solar irradiance forecasting on 12 Mar 2020 (Intraday)

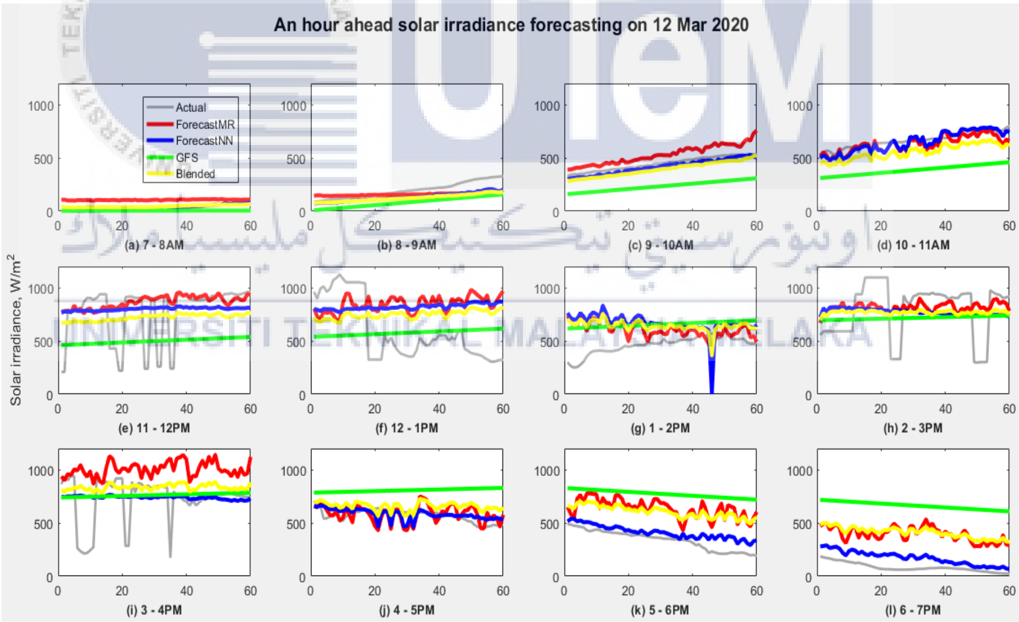

Figure 4.13: An hour ahead solar irradiance forecasting on 12 Mar 2020.

Table 4-9: Forecast performance for each model in different hour (12 Mar 2020).

Time	WMAPE (%)			
	MR model	NN model	GFS model	Blended model
7-8 AM	311.50	**25.52**	88.22	89.32
8-9 AM	**28.44**	33.31	57.80	36.65
9-10 AM	18.93	**6.55**	47.41	11.68
10-11 AM	7.43	**6.86**	42.13	16.69
11-12 PM	**21.02**	24.91	46.68	30.26
12-1 PM	63.99	59.98	**42.66**	55.54
1-2 PM	**33.44**	48.11	41.90	40.50
2-3 PM	**25.52**	29.45	30.46	28.48
3-4 PM	38.72	20.20	**19.79**	21.29
4-5 PM	12.17	**8.02**	40.85	15.99
5-6 PM	87.06	**21.08**	127.66	78.60
6-7 PM	376.58	**105.96**	698.10	393.55

Figure 4.13 shows an hour ahead solar irradiance forecasting on 12 Mar 2020. The forecast performance for each model on 12 Mar 2020 recorded in Table 4.9. Six times of the lowest WMAPE varied between 8.02% to 105.96% achieved by neural network (NN) model. The times of the lowest WMAPE for multiple regression (MR), GFS and blended models are four, two and zero. None of one time presented the lowest WMAPE for blended model. As the result, neural network (NN) model is defined as the best forecast model on 12 Mar 2020.

4.9 Discussion

In discussion, each part of result will be explained further. It also enables to state the points that never be mentioned and provides a good overview for the entire work.

First of foremost, data analysis is the first part in my project. Data analysis reflects on how the data affect to each other. As the result of the relationship between independent variables, there is no strong relationship between independent variables which means multicollinearity does not exist. On the other word, there is no duplicate and redundant independent variable. Also, all independent variables are suitable to put into forecast model. Besides that, temperature is the most suitable to apply because it is highly related to solar irradiance.

Second, the results of model validation showed that the neural network (NN) model has better R and RMSE compared to multiple regression (MR) model. However, the model validation is definitely difference than forecast testing. Even the result of model validation is great but it does not mean the model is great at forecasting. So, it is essential to test those models with at least few different testing days before concluding the performance of the models. The coefficient of seasonal ARIMA for inter-day and intraday are quite similar therefore the only difference is the seasonality.

For inter-day forecasting, there is expected result as neural network has the best performance on sunny day while blended model with better forecast on cloudy day. Since the most of data that refer to sunny day, then the neural network model with powerful algorithm which makes the forecast on sunny day more accurate. Besides that, the solar irradiance forecast obtained from GFS model is usually lower than the actual solar value. Therefore, blended model that combined MR, NN and GFS to provide moderate solar irradiance forecast which close to the solar value on cloudy day. Furthermore, all the models are also expected that not able to forecast on the extreme day. This is why intraday forecast takes part to overcome the issue.

Other than that, for intraday forecasting, although neural network model has outperformed than other models due to its deep neural network property, but still neural network has not the best performance for each forecast hour. Moreover, unlikely multiple regression as a form of equation, neural network is an opaque model which computes result based on neurons. Overall, it can be concluded that never forecast solar irradiance that only based on a single model but apply multiple models to get better forecast result.

4.10 Summary

In short, the results of all the works have discussed. Independent variables of temperature, humidity, wind speed and pressure are suitable to use. From the results of model validation, neural network model is better than multiple regression model which is within expectation. However, neural network model is not the model that forecast well for inter-day forecasting. Neural network can forecast well on sunny day. Then, blended model achieved the lowest WMAPE on cloudy day. Unfortunately, none of the models is able to forecast on the extreme day with low solar irradiance. So, intraday forecasting takes part to forecast on the extreme day. On the other hand, neural network model has shown better forecast performance compared to other models for intraday forecasting. Furthermore, GFS model has showed the potential to forecast low solar irradiance at peak of the extreme day for intraday forecasting.

CHAPTER 5

CONCLUSION AND FUTURE WORKS

5.1 Introduction

This chapter states the works that have been done and discusses the recommendation of future works. Overall, all the objectives have been achieved. All the models have successfully been built. Then, forecast performance for each model was evaluated in different testing days.

5.2 Conclusion

Solar irradiance forecasting is an essential technique that required to apply on the field of solar energy especially there is growing on green energy nowadays. In this project, multiple regression (MR) and neural network (NN) models have been built to forecast solar irradiance. With the weather forecast generated by ARIMA models, then only MR and NN models can forecast one day ahead (inter-day) and an hour ahead (intraday) solar irradiance. After that, blended model was formed by blending MR, NN and GFS models together. Neural network has the best result of model validation with R of 0.9173 and RMSE of 114.1820. As the result of inter-day forecast, WMAPE of 11.79% obtained for neural network model on sunny day. For the cloudy days, blended model has achieved the lowest WMAPE of 30.50% and 34.45%. In addition, none of models is able to perform well on the extreme day with low solar irradiance. On the other hand, neural network model is superior to other models on intraday forecast. Even though GFS model has not performed well on most of testing days but it is useful to apply on forecast the peak of day that towards low solar irradiance.

5.3 Future works

Based the obtained result, there is still a long way to go to achieve a perfect forecast system. In the future, there are a lot of possible ways can be applied to improve the result of solar irradiance forecasting. Then, one of the ways is to increase the quantity of the data. Since there is only 4 months of data used in my project, then it is possible to get better forecast model if there are more few years of data used to train the model.

Besides that, at this moment, the model is only based on the input parameters of temperature, humidity, wind speed, pressure and time (in decimal point). So, various of data can be further implemented. For example, the data of precipitation and clouds from satellite and the dataset from Numerical Weather Prediction (NWP) and European Centre for Medium-Range Weather Forecasts (ECMWF). As a reminder, the data cannot simply put into model to train, data analysis and preparation are very important steps to filter out unnecessary data to ignore overfitting problem. Therefore, more time required to deal with the huge number of various data. Also, more computation power required in order to complete model training.

Moreover, another factor that affected on forecast model which is forecast technique. Different forecast techniques can be applied in the future. For instance, develop a hybrid model based on the technique of LSTM-RNN or even create own custom forecast algorithm.

REFERENCES

[1] W. Regan *et al.*, “Solar Farm Fact Sheet,” *Researchgate.Net*, vol. 12, no. 12, pp. 528–533, 2019, doi: 10.3390/en12122437.

[2] K. M. Ravi Eswar, “Smart Grid-Future for Electrical Systems,” *Int. J. Electr. Electron. Res.*, vol. 3, no. June, pp. 603–612, 2015, [Online]. Available: www.researchpublish.com.

[3] H. H. S. Davood Mohammadi Souran, “Smart Grid Technology in Power Systems,” *Adv. Intell. Syst. Comput.*, vol. 357, no. July 2014, pp. 1097–1108, 2016, doi: 10.1007/978-3-319-18416-6.

[4] K. Q. Fung, A. Zicklin-kanter, and C. Hardwick, “Smart Cities and,” no. April, 2015.

[5] T. Vijayapriya and D. P. Kothari, “Smart Grid: An Overview,” *Smart Grid Renew. Energy*, vol. 02, no. 04, pp. 305–311, 2011, doi: 10.4236/sgre.2011.24035.

[6] G. Rahman, M. F. Bin, R. Chowdhury, A. Al Mamun, R. Hasan, and S. Mahfuz, “Summary of Smart Grid : Benefits and Issues,” *Int. J. Sci. Eng. Res.*, vol. 4,

no. 3, pp. 1–7, 2013.

[7] T. Jain and E. S. N. P. Raju, “Hybrid AC / DC Micro Grid : An Overview,” *Fifth Int. Conf. Power Energy Syst. Kathmandu*, no. October 2014, pp. 1–7, 2013, doi: 10.13140/2.1.1847.8081.

[8] V. Manieniyan, M. Thambidurai, and R. Selvakumar, “Study on Energy Crisis and the Future of Fossil,” *Shee*, no. October, pp. 7–12, 2009, doi: 10.13140/2.1.2234.3689.

[9] L. M. Fraas, *Low-cost solar electric power*, vol. 9783319075. 2014.

[10] E. O. Ogunniyi and H. C. V. Z. Pienaar, “Overview of battery energy storage system advancement for renewable (photovoltaic) energy applications,” *Proc. 25th Conf. Domest. Use Energy, DUE 2017*, no. April, pp. 233–239, 2017, doi: 10.23919/DUE.2017.7931849.

[11] S. A. . N. and T. N. M. S. Zafar, K. Nawaz, “Integration of Renewable Energy Sources in Smart Grid : a Review,” *Int. J. Sci. Res.*, vol. 50, no. 4, pp. 311–327, 2013, [Online]. Available: http://www.ijsr.net/conf/NCKITE2015/112.pdf.

[12] A. S. Y. Ahmad Zia Ul-Saufie, Hazrul abdul hamid, “Comparison Between Multiple Linear Regression And Feed forward Back propagation Neural Network Models For Predicting PM 10 Concentration Level Based On Gaseous And Meteorological Parameters Ahmad Zia Ul-Saufie (Corresponding author) Clean Air Research G,” vol. 1, no. 4, pp. 42–49, 2011.

[13] E. Ostertagová, “Modelling using polynomial regression,” *Procedia Eng.*, vol. 48, no. December 2012, pp. 500–506, 2012, doi:

10.1016/j.proeng.2012.09.545.

[14] M. O. Akinwande, H. G. Dikko, and A. Samson, “Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis,” *Open J. Stat.*, vol. 05, no. 07, pp. 754–767, 2015, doi: 10.4236/ojs.2015.57075.

[15] S. J. Miller, “The method of least squares,” *Handb. Res. Informatics Healthc. Biomed.*, pp. 181–184, 2006, doi: 10.4018/978-1-59140-982-3.ch023.

[16] Priyanka Sinha, “Multivariate Polynomial Regression in Data Mining: Methodology, Problems and Solutions,” *Int. J. Sci. Eng. Res.*, vol. 4, no. 12, pp. 962–965, 2013, [Online]. Available: https://www.ijser.org/paper/Multivariate-Polynomial-Regression-in-Data-Mining-Methodology.html.

[17] L. Massidda and M. Marrocu, “Use of Multilinear Adaptive Regression Splines and numerical weather prediction to forecast the power output of a PV plant in Borkum, Germany,” *Sol. Energy*, vol. 146, pp. 141–149, 2017, doi: 10.1016/j.solener.2017.02.007.

[18] U. Nalina, V. Prema, K. Smitha, and U. R. K, “Multivariate Regression for Prediction of Solar Irradiance,” no. July, 2015, doi: 10.1109/ICDSE.2014.6974633.

[19] J. Mekparyup, K. Saithanu, and J. Dujjanutat, “Multiple linear regression equation for estimation of daily averages solar radiationin Chonburi, Thailand,” *Appl. Math. Sci.*, vol. 7, no. 73–76, pp. 3629–3639, 2013, doi: 10.12988/ams.2013.35253.

[20] E. Grossi and M. Buscema, "Introduction to artificial neural networks," *Eur. J. Gastroenterol. Hepatol.*, vol. 19, no. 12, pp. 1046–1054, 2007, doi: 10.1097/MEG.0b013e3282f198a0.

[21] S. Vieira, W. H. L. Pinaya, and A. Mechelli, "Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications," *Neurosci. Biobehav. Rev.*, vol. 74, no. January, pp. 58–75, 2017, doi: 10.1016/j.neubiorev.2017.01.002.

[22] T. Khatib, A. Mohamed, K. Sopian, and M. Mahmoud, "Solar energy prediction for Malaysia using artificial neural networks," *Int. J. Photoenergy*, vol. 2012, 2012, doi: 10.1155/2012/419504.

[23] R. Marquez and C. F. M. Coimbra, "Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database," *Sol. Energy*, vol. 85, no. 5, pp. 746–756, 2011, doi: 10.1016/j.solener.2011.01.007.

[24] H. G. Chua, B. C. Kok, and H. H. Goh, "Solar Insolation Forecast Using Artificial Neural Network for Malaysian Weather," no. April, pp. 4–9, 2012.

[25] R. Kumar and S. Pathania, "Artificial Neural Network Model for Precise Estimation of," no. May, 2016.

[26] M. Abuella and B. Chowdhury, "Solar power forecasting using artificial neural networks," *2015 North Am. Power Symp. NAPS 2015*, no. November, 2015, doi: 10.1109/NAPS.2015.7335176.

[27] M. Husein and I. Y. Chung, "Day-ahead solar irradiance forecasting for

microgrids using a long short-term memory recurrent neural network: A deep learning approach," *Energies*, vol. 12, no. 10, 2019, doi: 10.3390/en12101856.

[28] S. G. Kim, J. Y. Jung, and M. K. Sim, "A two-step approach to solar power generation prediction based on weather data using machine learning," *Sustain.*, vol. 11, no. 5, 2019, doi: 10.3390/SU11051501.

[29] M. K. Douglas C. Montgomery, Cheryl L. Jennings, *Introduction to Time Series Analysis and Forecasting*. John Wiley & Sons. Inc .. Hoboken. New Jersey, 2008.

[30] G. M. L. George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, *Time Series Analysis: Forecasting and Control*, Fifth. John Wiley & Sons, Inc., Hoboken, New Jersey, 2016.

[31] S. B. Boualit and A. Mellit, "SARIMA-SVM hybrid model for the prediction of daily global solar radiation time series," *Proc. 2016 Int. Renew. Sustain. Energy Conf. IRSEC 2016*, no. November, pp. 712–717, 2017, doi: 10.1109/IRSEC.2016.7983867.

[32] C. Poolla and A. K. Ishihara, "Localized solar power prediction based on weather data from local history and global forecasts," *2018 IEEE 7th World Conf. Photovolt. Energy Conversion, WCPEC 2018 - A Jt. Conf. 45th IEEE PVSC, 28th PVSEC 34th EU PVSEC*, no. June, pp. 2341–2345, 2018, doi: 10.1109/PVSC.2018.8548173.

[33] M. H. Alsharif, M. K. Younes, and J. Kim, "Time series ARIMA model for prediction of daily and monthly average global solar radiation: The case study

of Seoul, South Korea," *Symmetry (Basel).*, vol. 11, no. 2, pp. 1–17, 2019, doi: 10.3390/sym11020240.

[34] S. Monjoly, M. André, R. Calif, and T. Soubdhan, "Forecast horizon and solar variability influences on the performances of multiscale hybrid forecast model," *Energies*, vol. 12, no. 11, 2019, doi: 10.3390/en12122264.

[35] H. Shao, "Introduction to Data Assimilation & Gridpoint Statistical Interpolation (GSI) System," 2018, [Online]. Available: https://dtcenter.org/HurrWRF/users/tutorial/2018_NCWCP_tutorial/lectures/11-HWRFtutJan2018_DAGSIoverview_Shao.pdf.

[36] L. Martin, "Global and direct normal irradiance forecast using GFS and statistical techniques," no. March 2011, 2014.

[37] H. Verbois, R. Huva, A. Rusydi, and W. Walsh, "Solar irradiance forecasting in the tropics using numerical weather prediction and statistical learning," *Sol. Energy*, vol. 162, no. February, pp. 265–277, 2018, doi: 10.1016/j.solener.2018.01.007.

[38] R. M. O'Brien, "A caution regarding rules of thumb for variance inflation factors," *Qual. Quant.*, vol. 41, no. 5, pp. 673–690, 2007, doi: 10.1007/s11135-006-9018-6.

[39] S. Eroshenko and A. Khalyasmaa, "Weather data errors analysis in solar power stations generation forecasting," *E3S Web Conf.*, vol. 51, pp. 1–6, 2018, doi: 10.1051/e3sconf/20185102002.

APPENDICES

```
filename_Read_1 = 'Data (Latest).xlsx';

Temp = xlsread(filename_Read_1, 'Sheet1',
'C2:C177196');
RH = xlsread(filename_Read_1, 'Sheet1',
'D2:D177196');
WindSpeed = xlsread(filename_Read_1, 'Sheet1',
'E2:E177196');
Pressure = xlsread(filename_Read_1, 'Sheet1',
'F2:F177196');
Time = xlsread(filename_Read_1, 'Sheet1',
'G2:G177196');
SolarRad = xlsread(filename_Read_1, 'Sheet1',
'H2:H177196');

C = {Temp RH WindSpeed Pressure Time};
Data = cell2mat(C);

Data_New = Data';          %Temp, RH, WS,
Pressure, Time (Daytime)
SolarRad_New = SolarRad'; %SolarRad (Daytime)

x = Data_New;
t = SolarRad_New;

trainFcn = 'trainlm';  % Levenberg-Marquardt
backpropagation.

% Create a Fitting Network
hiddenLayerSize = [5 5 5];
net = fitnet(hiddenLayerSize,trainFcn);

net.trainParam.epochs = 800;
net.trainParam.max_fail = 10;      %Validation
checks (stop training if there is consecutive 6
validation error increase)
%net.performParam.regularization = 0.5;

% Choose Input and Output Pre/Post-Processing
Functions
net.input.processFcns =
{'mapminmax'};           %Set the matrix row
between -1 and 1.
net.output.processFcns = {'mapminmax'};

net.divideFcn = 'dividerand';  % Divide data
randomly
net.divideMode = 'sample';  % Divide up every
sample
net.divideParam.trainRatio = 80/100;
net.divideParam.valRatio = 20/100;
net.divideParam.testRatio = 0/100;

net.performFcn = 'mse';  % Mean Squared Error

net.plotFcns =
{'plotperform','plottrainstate','ploterrhist','pl
otregression', 'plotfit'};

% Train the Network
% net = init(net);
[net,tr] =
train(net,x,t);                          %tr =
training record

% Test the Network
y = net(x);
e = gsubtract(t,y);
performance = perform(net,t,y);

% Calculate Training, Validation and Test
Performance
trainTargets = t .*
tr.trainMask{1};                     %t = solar as
target (output)
valTargets = t .* tr.valMask{1};
testTargets = t .* tr.testMask{1};
trainPerformance =
perform(net,trainTargets,y);          %y = net(x)
while x = data for each parameters
valPerformance = perform(net,valTargets,y);
testPerformance = perform(net,testTargets,y);

%Train_RMSE = sqrt(trainPerformance)
Val_RMSE = sqrt(valPerformance)
%Test_RMSE = sqrt(testPerformance)

%figure, plotperform(tr)
%figure, plotregression(t,y)    % R-squared for
overall
```

```
load('C:\Users\Asus\Desktop\Workspace\MR_Coefficient (Latest).mat')
load('C:\Users\Asus\Desktop\Workspace\NN (Latest).mat')

%Read data
filename_Read_1 = 'Data_Read/Data (9 Mar - 12 Mar 2020).xlsx';

Temp = xlsread(filename_Read_1, 'Sheet1', 'C2:C5761');
RH = xlsread(filename_Read_1, 'Sheet1', 'D2:D5761');
WindSpeed = xlsread(filename_Read_1, 'Sheet1', 'E2:E5761');
Pressure = xlsread(filename_Read_1, 'Sheet1', 'F2:F5761');
Time = xlsread(filename_Read_1, 'Sheet1', 'G2:G5761');
SolarRad = xlsread(filename_Read_1, 'Sheet1', 'H2:H5761');
One = xlsread(filename_Read_1, 'Sheet1', 'I2:I1441');

filename_Read_2 = 'Data_GFS/GFS (12 Mar 2020).xlsx';
Orig_SolarRad = xlsread(filename_Read_2, 'Sheet1', 'D2:D11');

y1 = Temp(1:4320);          % 3 days
y2 = RH(1:4320);
y3 = WindSpeed(1:4320);
y4 = Pressure(1:4320);

ActTemp1 = Temp(4321:5760);
ActRH1 = RH(4321:5760);
ActWS1 = WindSpeed(4321:5760);
ActP1 = Pressure(4321:5760);

% ARIMA model for weather forecasting
model_1 = arima('AR', {-0.1816}, 'ARLags',1, 'D', 1, 'SAR', {-0.1816}, 'SARLags', 1, 'Seasonality', 1440, 'constant', 0, 'variance',
0.2948);      %Temp
model_2 = arima('AR', {-0.0499}, 'ARLags',1, 'D', 1, 'SAR', {-0.0499}, 'SARLags', 1, 'Seasonality', 1440, 'constant', 0, 'variance',
0.7431);      %RH
model_3 = arima('AR', {-0.1883}, 'ARLags',1, 'D', 1, 'SAR', {-0.1883}, 'SARLags', 1, 'Seasonality', 1440, 'constant', 0, 'variance',
12.1809);     %WS
model_4 = arima('AR', {-0.2463}, 'ARLags',1, 'D', 1, 'SAR', {-0.2462}, 'SARLags', 1, 'Seasonality', 1440, 'constant', 0, 'variance',
0.0062);      %Pressure

Temp1 = forecast(model_1, 1440, 'Y0', y1);        % Forecast next day
RH1 = forecast(model_2, 1440, 'Y0', y2);
WindSpeed1 = forecast(model_3, 1440, 'Y0', y3);
Pressure1 = forecast(model_4, 1440, 'Y0', y4);

%Multiple regression
Time1 = Time(1:1440);     % 7-8   AM

TempRH1 = Temp1 .* RH1;
TempWS1 = Temp1 .* WindSpeed1;
TempP1 = Temp1 .* Pressure1;
TempT1 = Temp1 .* Time1;
RHWS1 = RH1 .* WindSpeed1;
RHP1 = RH1 .* Pressure1;
RHT1 = RH1 .* Time1;
WSP1 = WindSpeed1 .* Pressure1;
WST1 = WindSpeed1 .* Time1;
PT1 = Pressure1 .* Time1;
Temp21 = Temp1 .* Temp1;
RH21 = RH1 .* RH1;
WindSpeed21 = WindSpeed1 .* WindSpeed1;
Pressure21 = Pressure1 .* Pressure1;
Time21 = Time1 .* Time1;

One1 = One;

X1 = [One1 Temp1 RH1 WindSpeed1 Pressure1 Time1 TempRH1 TempWS1 TempP1 TempT1 RHWS1 RHP1 RHT1 WSP1 WST1 PT1 Temp21 RH21 WindSpeed21 Pressure21
Time21];

ForecastMR1 = X1 * b1;

%Neural network
C1 = {Temp1 RH1 WindSpeed1 Pressure1 Time1};

Data1 = cell2mat(C1);
Data_New1 = Data1';

simOut1 = sim(net, Data_New1);
ForecastNN1 = simOut1';

% GFS Interpolation
Interp_Time = 1:(1/179.95):10;     %For 1 minute interval  (Latest) For 1620 from 11PM to 2AM
Interp_SolarRad = interp1(Orig_SolarRad,Interp_Time);

Interp_SolarRad_New = Interp_SolarRad';

GFS1 = Interp_SolarRad_New(61:1500);

%Forecast blending
Blended1 = (ForecastMR1 + ForecastNN1 + GFS1) ./ 3;

Act1 = SolarRad(4741:5460);
MeanAct1 = mean(Act1);

ErrorMR1 = Act1 - ForecastMR1(421:1140);
ErrorNN1 = Act1 - ForecastNN1(421:1140);
ErrorGFS1 = Act1 - GFS1(421:1140);
ErrorBlended1 = Act1 - Blended1(421:1140);

RMSE_MR1 = sqrt(mean((ErrorMR1).^2));
RMSE_NN1 = sqrt(mean((ErrorNN1).^2));
RMSE_GFS1 = sqrt(mean((ErrorGFS1).^2));
RMSE_Blended1 = sqrt(mean((ErrorBlended1).^2));

WMAPE_MR1 = (sum(abs(ErrorMR1))./sum(Act1))*100;
WMAPE_NN1 = (sum(abs(ErrorNN1))./sum(Act1))*100;
WMAPE_GFS1 = (sum(abs(ErrorGFS1))./sum(Act1))*100;
WMAPE_Blended1 = (sum(abs(ErrorBlended1))./sum(Act1))*100;

ForecastMR1_New = ForecastMR1(421:1140);
ForecastNN1_New = ForecastNN1(421:1140);
GFS1_New = GFS1(421:1140);
Blended1_New = Blended1(421:1140);

% Display result
figure
plot(SolarRad(4741:5460), 'color', [0.65 0.65 0.65], 'linewidth',2);hold on
plot(ForecastMR1_New, 'r', 'linewidth',3);hold on
plot(ForecastNN1_New, 'b', 'linewidth',3);hold on
plot(GFS1_New, 'g', 'linewidth',3); hold on
plot(Blended1_New, 'y', 'linewidth',3)
xlim([0 720])
ylim([0 1200])
% xlabel('Minute')
set(gca,'XTick', [1 60 120 180 240 300 360 420 480 540 600 660 720]);
set(gca,'XTickLabel',{'7AM','8AM','9AM','10AM','11AM','12PM','1PM','2PM','3PM','4PM','5PM','6PM','7PM'});
ylabel('Solar irradiance, W/m^2')
title('Day ahead solar irradiance forecasting on 12 Mar 2020')
legend('Actual', 'MR model', 'NN model', 'GFS model', 'Blended model', 'location', 'northwest')

text(580, 1150, sprintf('WMAPE(MR)          = %0.2f%%', WMAPE_MR1))
text(580, 1120, sprintf('WMAPE(NN)          = %0.2f%%', WMAPE_NN1))
text(580, 1090, sprintf('WMAPE(GFS)        = %0.2f%%', WMAPE_GFS1))
text(580, 1060, sprintf('WMAPE(Blended)  = %0.2f%%', WMAPE_Blended1))
text(550, 1000, sprintf('Average solar irradiance  = %0.2f%', MeanAct1))
text(687, 1002, 'W/m^2')
```