

**PERFORMANCE ANALYSIS OF SPECIFIC YIELD FOR SOLAR
PV BASED ON STATISTICAL TECHNIQUE**

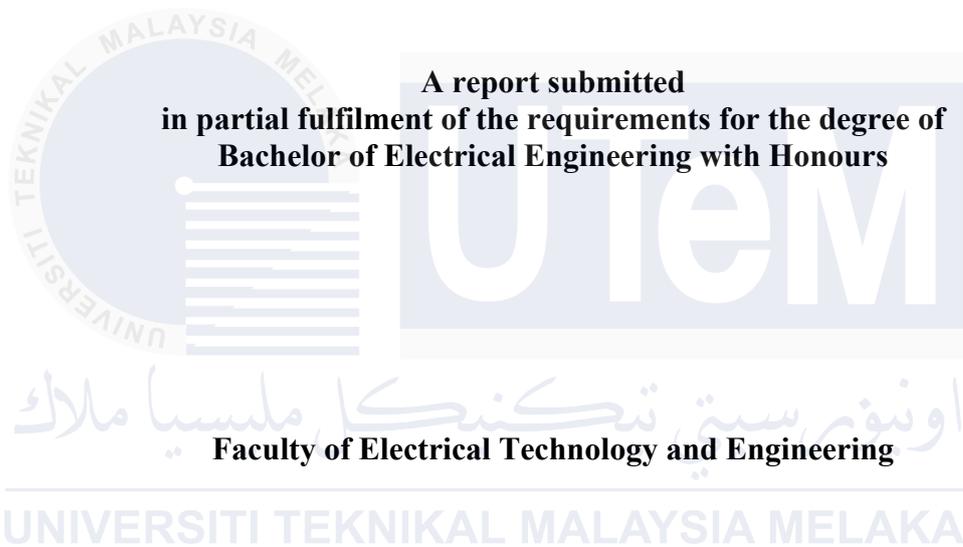


**BACHELOR OF ELECTRICAL ENGINEERING WITH HONOURS
UNIVERSITI TEKNIKAL MALAYSIA MELAKA**

2024

**PERFORMANCE ANALYSIS OF SPECIFIC YIELD FOR SOLAR PV BASED ON
STATISTICAL TECHNIQUE**

MEGAT ZHAFRANN AMYR BIN MOHAMAD FAKHRI



UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2024

DECLARATION

I declare that this thesis entitled "PERFORMANCE ANALYSIS OF SPECIFIC YIELD FOR SOLAR PV BASED ON STATISTICAL TECHNIQUE" is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in the candidature of any other degree.

Signature :



Name :

Megat Zhafrann Amyr Bin Mohamad Fakhri

Date :

24 June 2024

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

APPROVAL

I hereby declare that I have checked this report entitled "title of the project", and in my opinion, this thesis fulfils the partial requirement to be awarded the degree of Bachelor of Mechatronics Engineering with Honours

Signature

:



Supervisor Name

:

Dr. Arfah Binti Ahmad

Date

:

4th July 2024

اونيورسيتي تيكنيكل مليسيا ملاك

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

DEDICATIONS

To my beloved mother Rozita binti Bachik and father Mohamad Fakhri bin Abd Rashid



ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Arfah binti Ahmad, for her unwavering support, invaluable guidance, and insightful feedback throughout the course of this research. Her encouragement and dedication have been instrumental in the completion of this study.

I would also like to extend my sincere thanks to my panel, Dr. Aine Izzati binti Tarmizi and Dr. Farhan bin Hanaffi, for their constructive criticism, expert advice, and thoughtful suggestions, which greatly enhanced the quality of this research.

Additionally, I am grateful to Universiti Teknikal Malaysia Melaka (UTeM) for providing the necessary resources and a conducive environment for my research.

Finally, I would like to thank my family and friends for their continuous support and encouragement.



ABSTRACT

This study develops a time series model to forecast the specific yield of solar panels in the Faculty of Electrical Technology and Engineering (FTKE) area. The study uses data from four types of solar panels around the (FTKE) area: Thin Film (TF) solar panels, Heterojunction (HIT) solar panels, monocrystalline solar panels and polycrystalline solar panels. Initially, the analysis begins with data pre-processing to calculate descriptive statistics, processing missing values and data merging. Then, the descriptive statistics are calculated, revealing that TF solar panels have the maximum specific yield. The study is then focus on forecasting the specific yield of TF solar panels. Subsequently, the time series models ARMA and ARIMA are developed using Minitab software to analyze the processed data. The forecast model's accuracy will be evaluated through Mean Absolute Error (MAE) and Mean Squared Error (MSE) to determine the best forecast model. The developed model is used to forecast the specific yield at FTKE, UTeM in the future.

ABSTRAK

Kajian ini membentuk model siri masa untuk meramalkan hasil spesifik panel solar di kawasan Fakulti Teknologi Kejuruteraan Elektrik (FTKE). Kajian ini menggunakan data daripada empat jenis panel solar di sekitar kawasan FTKE: panel solar Thin Film (TF), panel solar Heterojunction (HIT), panel solar monocrystalline dan panel solar polycrystalline. Pada mulanya, analisis dimulakan dengan prapemprosesan data untuk mengira statistik deskriptif, memproses nilai yang hilang dan penggabungan data. Kemudian, statistik deskriptif dikira, menunjukkan bahawa panel solar TF mempunyai hasil spesifik maksimum. Kajian ini kemudian memfokuskan pada meramalkan hasil spesifik panel solar TF. Seterusnya, model siri masa ARMA dan ARIMA dibangunkan menggunakan perisian Minitab untuk menganalisis data yang diproses. Ketepatan model ramalan akan dinilai melalui Ralat Mutlak Min (MAE) dan Ralat Kuasa Dua Min (MSE) untuk menentukan model ramalan terbaik. Model yang dibangunkan akan digunakan untuk meramalkan hasil spesifik di FTKE, UTeM pada masa hadapan.

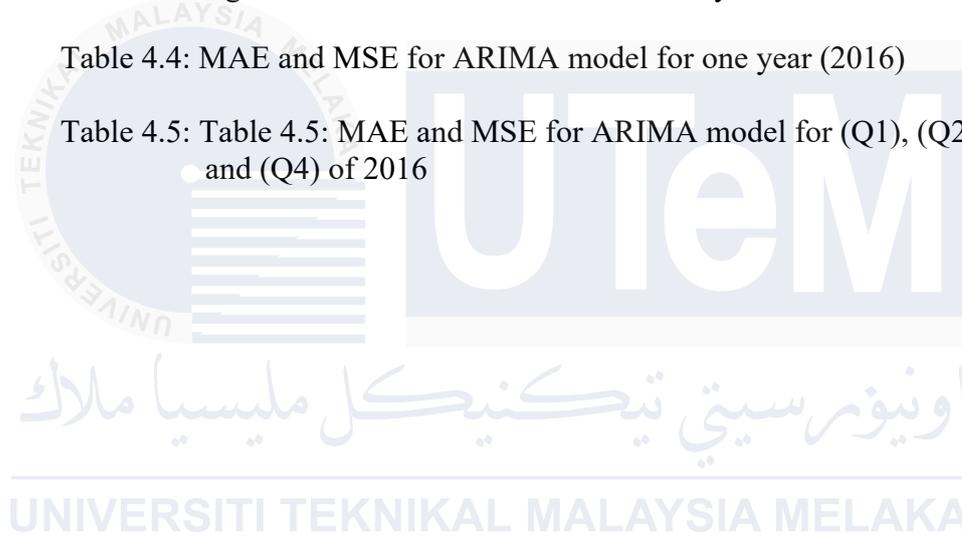
TABLE OF CONTENTS

	PAGE
DECLARATION	
APPROVAL	
DEDICATIONS	
ACKNOWLEDGEMENTS	2
ABSTRACT	3
ABSTRAK	4
TABLE OF CONTENTS	5
LIST OF TABLES	7
LIST OF FIGURES	8
LIST OF SYMBOLS AND ABBREVIATIONS	11
CHAPTER 1 INTRODUCTION	12
1.1 Background	12
1.2 Problem Statement	16
1.3 Objective	16
1.4 Scope	17
1.5 Project Motivation	17
CHAPTER 2 LITERATURE REVIEW	18
2.1 Types of Forecasting Method for Solar Radiation	18
2.1.1 Probability Distribution	18
2.1.2 Artificial Neural Network	19
2.1.3 Hybrid Model	20
2.1.4 Comparative Analysis of Past Case Studies	21
2.2 Time Series Analysis	21
2.2.1 Type of Time Series Analysis	22
2.3 Factors that Contribute to The Time Series Analysis	23
2.4 Related Case Studies on Time Series Analysis	25
2.4.1 A Guide to Solar Power Forecasting using ARMA Models	25
2.4.2 ARMA Model for Short-term Forecasting of Solar Potential: Dakar Site	27
2.4.3 Forecasting of Total Daily Solar Energy Generation using ARIMA28	
CHAPTER 3 METHODOLOGY	29
3.1 Introduction	29
3.2 Harvesting Raw Data	30
3.3 Data Pre-Processing	31
3.3.1 Data Merging	31

3.3.2	Missing Values	33
3.3.3	Descriptive Statistics	34
3.4	Time Series Model	34
3.4.1	ARMA Model	35
3.4.2	ARIMA Model	35
3.4.3	Model Parameter Selection	36
3.4.3.1	Stationary Data	36
3.4.3.2	Autocorrelation Function (ACF)	37
3.4.3.3	Partial Autocorrelation Function (PACF)	38
3.5	Error Measurement	38
3.5.1	Mean Absolute Error (MAE)	38
3.5.2	Mean Squared Error (MSE)	39
CHAPTER 4 RESULTS AND DISCUSSIONS		40
4.1	Preliminary Results	40
4.1.1	Monthly Data Trend	40
4.1.2	Daily Data Trend	47
4.2	Descriptive Statistics	49
4.3	Stationarity	50
4.3.1	Probability Plot	50
4.3.2	Box-Cox Transformation	51
4.3.3	Autocorrelation Factor (ACF)	52
4.3.4	Partial Autocorrelation Factor (PACF)	53
4.4	Time Series Model Development	54
4.5	Error Measurement	56
4.6	Forecasting Future Values by using ARIMA Model	58
CHAPTER 5 CONCLUSION AND RECOMMENDATIONS		61
5.1	Conclusion	61
5.2	Future Works	61
REFERENCES		62

LIST OF TABLES

Table 1.1: Characteristics of solar panel	14
Table 2.1: Summary of two types of ANN model [15]	18
Table 3.1: Series number of type of solar panel	31
Table 4.1: Descriptive statistics of solar panels	48
Table 4.2: Non-significant ARIMA model for case study data	54
Table 4.3: Significant ARIMA model for case study data	55
Table 4.4: MAE and MSE for ARIMA model for one year (2016)	56
Table 4.5: Table 4.5: MAE and MSE for ARIMA model for (Q1), (Q2), (Q3) and (Q4) of 2016	57



LIST OF FIGURES

Figure 1.1: Conversion of solar to electrical energy [3]	13
Figure 1.2: Monocrystalline solar panel at the FTKE administration rooftop [8]	14
Figure 1.3: TF solar panel at the FTKE laboratory rooftop [8]	15
Figure 1.4: HIT solar panel at the FTKE administration rooftop [8]	15
Figure 1.5: Polycrystalline solar panel at the FTKE entrance [8].	15
Figure 2.1: Plot of independent uncorrelated random variable [20]	21
Figure 2.2: Typical weather factor correlation curve [26].	24
Figure 2.3: Meteorological factors affecting solar power forecasting by month [28]	25
Figure 2.4: Estimated p and q values for ARMA model for 14 hours [18]	26
Figure 3.1: Flowchart of the project	29
Figure 3.2: Raw data from SSG Laboratory	30
Figure 3.3: Imported data in Microsoft Excel file format.	31
Figure 3.4: The summary results of data merging in February 2016 for (a) polycrystalline (b) HIT (c) Thn-Film (d) monocrystalline solar panels from each type of solar panel	32
Figure 3.5: Datasheet with missing values	33
Figure 3.6: List of missing values in February 2016	33
Figure 3.7: Time series plot of stationary and non-stationary data [34]	37
Figure 3.8: PACF plot of time series after differencing [36]	38
Figure 4.1: Total energy yield and specific yield in January 2016	40
Figure 4.2: Total energy yield and specific yield in February 2016	41
Figure 4.3: Total energy yield and specific yield in March 2016	41

Figure 4.4:	Total energy yield and specific yield in April 2016	42
Figure 4.5:	Total energy yield and specific yield in May 2016	42
Figure 4.6:	Total energy yield and specific yield in June 2016	43
Figure 4.7:	Total energy yield and specific yield in July 2016	43
Figure 4.8:	Total energy yield and specific yield in August 2016	44
Figure 4.9:	Total energy yield and specific yield in September 2016	44
Figure 4.10:	Total energy yield and specific yield in October 2016	45
Figure 4.11:	Total energy yield and specific yield November 2016	45
Figure 4.12:	Total energy yield and specific yield in December 2016	46
Figure 4.13:	Daily trend for Polycrystalline solar panel	47
Figure 4.14:	Daily trend for Monocrystalline solar panel	47
Figure 4.15:	Daily trend for Thin Film solar panel	48
Figure 4.16:	Daily trend for HIT solar panel.	48
Figure 4.17:	Probability plot of TF specific yield in (a) one year, (b) first quarter, (c) second quarter, (d) third quarter and (e) fourth quarter of 2016	50
Figure 4.18:	Probability plot after normalization of TF specific yield in (a) one year, (b) first quarter, (c) second quarter and (d) third quarter of 2016	51
Figure 4.19:	ACF plot for TF specific yield in (a) one year, (b) first quarter, (c) second quarter, (d) third quarter and (e) fourth quarter of 2016	52
Figure 4.20:	PACF plot for TF specific yield in (a) one year, (b) first quarter, (c) second quarter, (d) third quarter and (e) fourth quarter of 2016	53
Figure 4.21:	Forecast data trend for one whole year of 2016 for TF solar panel	58
Figure 4.22:	Forecast data trend for the first quarter of 2016 for TF solar panel	55

Figure 4.23: Forecast data trend for the second quarter of 2016 for TF solar panel	55
Figure 4.24: Forecast data trend for the third quarter of 2016 for TF solar panel	56
Figure 4.25: Forecast data trend for the fourth quarter of 2016 for TF solar panel	56



LIST OF SYMBOLS AND ABBREVIATIONS

FTKE	-	Faculty of Electrical Technology and Engineering
UTeM	-	Universiti Teknikal Malaysia Melaka
SSG	-	Smart Grid Research Laboratory
PV	-	Photovoltaic
TF	-	Thin-Film
HIT	-	Heterojunction
ANN	-	Artificial Neural Network
AR	-	Autoregressive
MA	-	Moving Average
ARMA	-	Autoregressive Moving Average
ARIMA	-	Autoregressive Integrating Moving Average
MAE	-	Mean Absolute Error
MSE	-	Mean Squared Error
MAPE	-	Mean Absolute Percentage Error
RMSE	-	Root Mean Squared Error
ACF	-	Autocorrelation Function
PACF	-	Partial Autocorrelation Function

CHAPTER 1

INTRODUCTION

1.1 Background

Solar power is a form of renewable energy that harnesses the sun's rays and transforms the thermal energy to generate electricity. Solar power can be produced in two ways; photovoltaic (PV) and concentrated solar power (CSP). PV harnesses light and converts it into electricity using the photovoltaic effect and CSP uses mirrors to focus sunlight to generate electricity [1].

The level of solar radiation will determine the energy produced by a solar panel. An accurate solar forecast helps to boost the penetration level of a PV system, improve system reliability, and reduce power uncertainty on the grid. Different techniques have been developed to forecast solar radiation such as probability distribution, artificial neural network, hybrid model and time series analysis. Time series analysis methods are commonly used for forecasting, as they utilize historical solar radiation data to analyse the behaviour of solar radiation and provide accurate forecasts [2].

Solar energy is converted into electricity through the photovoltaic effect. When photons strike a semiconductor material like silicon, they release electrons from their atoms, leaving behind a vacant space. The stray electrons move around randomly, looking for another hole to fill. There are two types of semiconductors in solar cells, which are p-type and n-type. These two are joined together to form a p-n junction. An electric field is formed in the junction region. The electron starts moving to the negative side, the n-side. The electric field causes the charged particles to move in one direction and positively charged particles to move in the other direction [3].

The sunlight comprises photons, also known as the small bundles of electromagnetic radiation energy. When photons hit the surface of a photovoltaic cell, their energy is transferred to the electrons in the cell. As a result, the electron gets excited and starts jumping to a higher energy state known as the conduction band.

Once the electrons jump to the conduction band, they leave holes in the valence band. This electron movement creates energy, making two charge carriers and an electron-hole pair. The motion of electrons when they move in the excited state causes electricity conversion by solar cells [4]. Figure 1.1 illustrates how solar cells work.

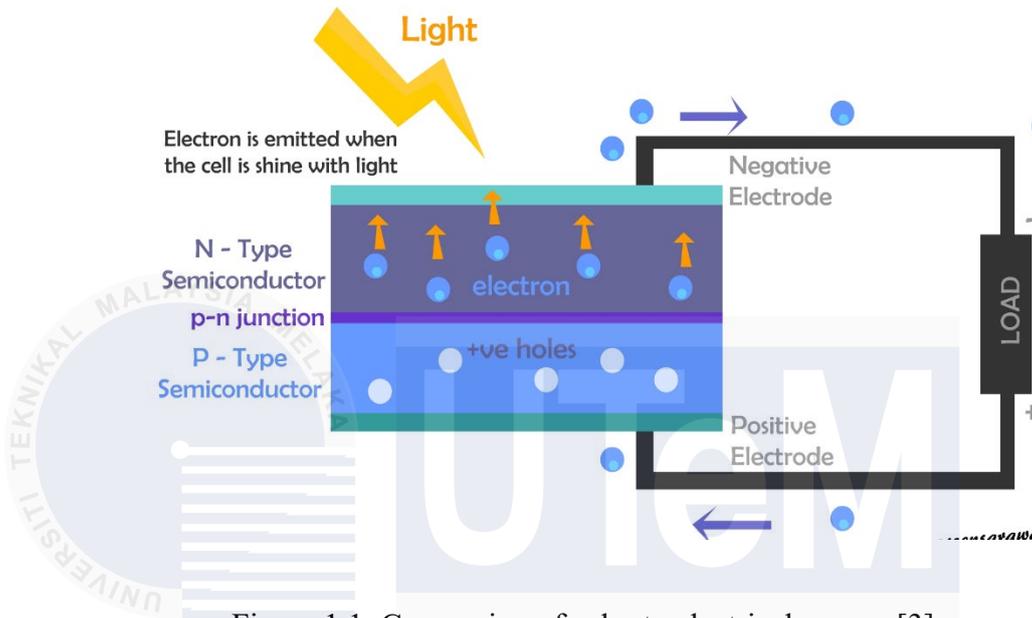


Figure 1.1: Conversion of solar to electrical energy [3].

Faculty of Electrical Technology and Engineering (FTKE) installed several types of solar panels; Monocrystalline, Polycrystalline, Thin-Film (TF) and Heterojunction with Intrinsic Thin layer (HIT). Figure 1.2 to Figure 1.5 show solar panel in FTKE building. The solar panel system used for research activities such as partial shading analysis, PV forecasting, PV plant design and PV system performance.

Based on Table 1.1, each solar panel has its own material, such as Silicon fragments, Crystalline Silicon, Amorphous Silicon, Indium Tin Oxide and Cadmium Telluride. According to the table, TF has the shortest lifespan (15 to 20 years), followed by Polycrystalline (25 years) Monocrystalline (25 years) and HIT (25-30 years). HIT has the highest efficiency (20-22%) while Polycrystalline has the lowest efficiency (13-18%). Monocrystalline and TF have an efficiency of (15-22%) and (19-20.4%) respectively. Temperature coefficient is the percentage decrease in output for every 1°C increase in temperature from 25°C. Hence, TF solar panel has the most minor effect (-0.2) on temperature, followed by HIT (-0.3), Monocrystalline (-0.40) and Polycrystalline (-0.43) [5-7].

Table 1.1: Characteristics of solar panel [5-7]

	Polycrystalline	Monocrystalline	HIT	TF
Material	Multiple silicon fragments	Single crystal of silicon	Crystalline Silicon, Amorphous Silicon, Indium Tin Oxide	Cadmium Telluride
Lifespan	25 years	25years	25-30 years	15 – 20 years
Efficiency	13-18%	15-22%	20-22%	19-20.4%
Average temperature coefficient	-0.43%/°C	-0.40%/°C	-0.3%/°C	-0.2%

In this study, the method used to forecast the solar radiation is time series analysis. The data collected are form solar panel installed by Smart Grid Research Laboratory (SSG) in FTKE. The solar radiation data are from 1st January 2016 until 31st December 2016. The time series model is used as a forecast model for the solar radiation which will benefit the solar panel at FTKE building.



Figure 1.2: Monocrystalline solar panel at the FTKE administration rooftop [8]



Figure 1.3: TF solar panel at the FTKE laboratory rooftop [8]

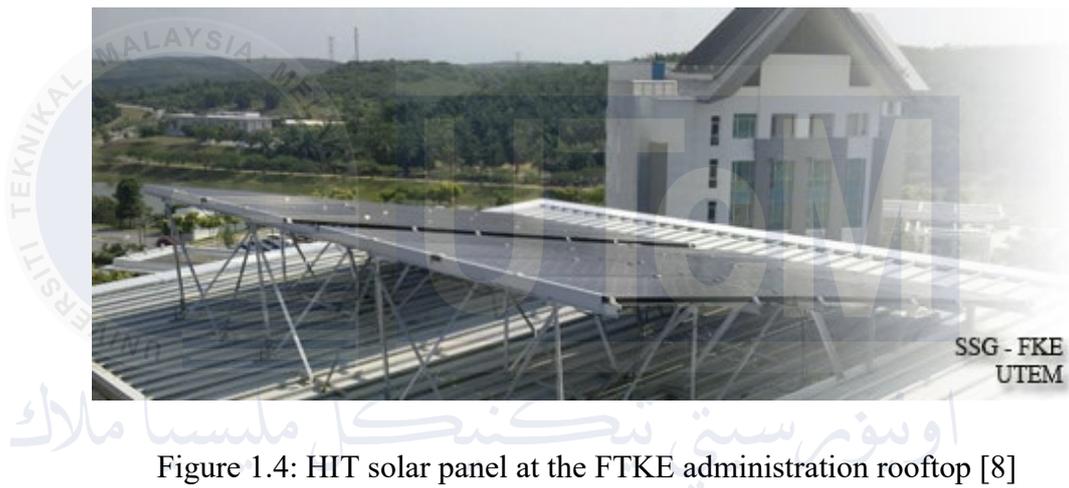


Figure 1.4: HIT solar panel at the FTKE administration rooftop [8]



Figure 1.5: Polycrystalline solar panel at the FTKE entrance [8]

1.2 Problem Statement

For the past years, humans heavily rely on oil, fossil fuels and coal to generate power for various purposes. These sources constitute the primary drivers of global climate change, responsible for more than 90% of total carbon dioxide emissions [9].

Solar energy is considered as one of the main renewable sources in Malaysia due to its abundant sunlight which provide a reliable consistent source of power. Malaysia's tropical climate and strategic location close to the equator results in high solar irradiance [10]. A series of government initiatives have been introduced throughout the years to boost the application of renewable energy to achieve 70% renewable energy by 2050. Net Energy Metering (NEM), Feed-in Tariff (FiT), Large Scale Solar (LSS), and Green Electricity Tariff (GET) are the initiatives by the government to increase the usage of solar panel in Malaysia [11]. However, the amount of energy produced by solar panels can be inconsistent due to weather conditions like cloudy or rainy days. The weather inconsistency makes grid integration a challenging task. Thus, solar forecasting becomes vital to ensure grid stability, reliability, and efficient operation of the power system [2]. Therefore, this study will forecast specific yield data in FTKE to ensure the best time series analysis for accurate forecast model of specific yield.

1.3 Objective

1. To pre-process raw data of specific yield produced by solar panel at FTKE through data cleaning.
2. To analyse the clean data of specific yield from solar panel at FTKE by using time series analysis.
3. To measure the accuracy of the time series model based on error measurement.

1.4 Scope

This study has a few potential limitations. One of the limits are outdated data, as the data used for this study from 2016. This study does not use recent year data because of malfunction equipment at SSG laboratory. Secondly, this study only use data collected from solar panel around FTKE area. Thus, the forecast model is not suitable to use at other location as it will result with inaccurate forecast. Next, the limitation is imperfect data, as missing values are spotted at random timestamp. Subsequently, the limitation is the study used time series analysis method only. Although, there are several forecasting methods such as probability distribution, artificial neural network, and hybrid model. Lastly, the study is limited in the error measurement method, as the study only uses mean average error (MAE) and mean squared error (MSE). Other methods, such as mean absolute percentage error (MAPE) and root mean squared error (RMSE), can also be used for further error measurement analysis.

1.5 Project Motivation

The motivation of this study is to reduce electricity bills. Malaysia now relies on non-renewable energy sources such as fossil fuels and coal. With time, the fossil fuel will run out which could affect the increment of electricity bills in the future. The FTKE building uses a lot of electricity as the building need electricity to supply power for high voltage laboratory equipment, high horsepower air conditioning and electrical equipment. Therefore, to take a step forward a solar forecast model is needed to manage battery energy storage system for electricity usage management and associate cost.

CHAPTER 2

LITERATURE REVIEW

2.1 Types of Forecasting Method for Solar Radiation

The following topic discussed about the methods used to analyse the solar radiation data.

2.1.1 Probability Distribution

A probability distribution is a statistical approach that indicates the likelihood of various outcomes in a random experiment or event. It assigns probabilities to different potential results. Studies in [12] and [13] used probability distribution to predict solar irradiance using gathered data. The study used four probability functions that are Normal, Rayleigh, Weibull, and Log normal to determine the most suitable fit for the data. The equation for normal distribution is shown in equation (2.1).

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (2.1)$$

where σ is standard deviation and μ is mean. The equation for Rayleigh distribution is as shown in equation (2.2)

$$f(x; \sigma) = 1 - e^{-\frac{x^2}{2\sigma^2}} \quad (2.2)$$

where σ is the shape parameter, e is Euler's number and x is random variable. The equation for Weibull distribution is as shown in equation (2.3).

$$f(x; \lambda, k) = \frac{\gamma}{\alpha} \left(\frac{x - \mu}{\alpha}\right)^{\gamma-1} \exp^{-\left(\frac{x-\mu}{\alpha}\right)^\gamma} \quad x \geq \mu; \gamma, \alpha > 0 \quad (2.3)$$

where γ is the shape parameter, α is the scale parameter, and μ is the location parameter. The equation for Log Normal distribution as shown in equation (2.4).

$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} e^{\left[-\frac{\ln(x) - \mu^2}{2\sigma^2} \right]} \quad (2.4)$$

where μ is the shape parameter, σ is the scale parameter of the distribution.

The MSE concept was applied to calculate the error and to get the best-fitted probability distribution. The standard distribution with the smallest MSE with respect to the data is considered as the best candidate distribution. Results in [13] show that normal distribution is the most suitable distribution function for summer and rainy seasons, and Rayleigh distribution is the most suitable for winter.

Based on these studies, it has been demonstrated that forecasting of solar irradiance, as well as its accuracy for sampled data, can be accommodated using various probability distribution functions [12,13].

2.1.2 Artificial Neural Network

An artificial neural network (ANN) is a computational model inspired by the structure and functioning of the human brain. It is a type of machine learning algorithm designed to recognize patterns, learn from data, and make predictions or decisions [14]. The architecture of the ANN comprises three layers. Typically, there is an input layer that receives gathered data, an output layer that generates computed information, and one or more hidden layers that facilitate the connection between the input and output layers through processing unit neurons [15].

A study in [15] uses nine parameters that are latitude, longitude, altitude, year, month, mean ambient air temperature, mean station level pressure, mean wind speed, and mean relative humidity as the inputs of ANN model. The sole output is the prediction of the monthly average global solar radiation. The author collected data from Bangalore, Chennai, Kolkata, New Delhi, and Mumbai to conduct data training for the ANN model that is run on the MATLAB software. Two models of ANN were created where in model 1, data from Chennai, Kolkata, New Delhi, and Mumbai are used only for training, and testing will be done for data in Bangalore only. In another case, model 2 uses all five places' data for training and testing. The performance of the two ANN models was evaluated using MAE and RMSE error measurement. From the result of error measurement, model 2 is a slightly more accurate solar forecast model

as it has a lower error measurement value compared to model 1. Table 2.1 is the summarise between the two ANN models used in [15].

Table 2.1: Summary of two types of ANN model [15]

	Training		Testing		Error Measurement	
	No of city	City	No of city	City	MAE	RMSE
Model 1	4	Chennai Kolkata New Delhi Mumbai	1	Bangalore	0.83	1.08
Model 2	5	Chennai Kolkata New Delhi Mumbai Bangalore	5	Chennai Kolkata New Delhi Mumbai Bangalore	0.78	0.95

2.1.3 Hybrid Model

Hybrid model analysis is the combination between linear and nonlinear models. This method is used to improve forecast accuracy. A study in [16] used hybrid model that harness the optimal performance of individual models across various temporal horizons, combining both machine learning and deep learning models. The study combines hybrid model with models based on satellite imagery and numerical weather prediction for improving intra-day solar radiation forecast and proved it is the optimal choice for short term hourly intra-day solar forecasting [16].

The hybrid model is designed to begin by independently computing the hourly weighs for each individual machine learning and data learning model. The rankings are assigned based on the cumulative weight of each model. The top six performing models out of thirteen will be selected. The selected model is based on the process of decision tree hybrid model where the data will go through training to improve the accuracy of the selected models [16]. Later, the data will go through error measurement analysis. The error measurement value is compared between hybrid model and non-hybrid model. The comparison summarises that, hybrid model able to increase efficiency of the selected models around 5% to 10% [16].

2.1.4 Comparative Analysis of Past Case Studies

Based on sub-section (2.1.1) to (2.1.3), it can be concluded that there are several methods used to forecast data gain from solar panels, such as ANN, probability distribution, and hybrid models. Some of the forecast models are used to forecast solar radiation or solar irradiance, while this study focuses on forecasting specific yields. All past studies and this study use historical data to generate the forecast model. There are several techniques are used to measure the accuracy of the forecast models, such as MAE, RMSE, and MAPE. However, for this study, the methods used will be MAE and MSE to measure the accuracy.

2.2 Time Series Analysis

Time Series pertains to the sequence of observations collected in constant time intervals, be it daily, monthly, quarterly, or yearly. Time series analysis involves developing models used to describe the observed time series and understand the "why" behind its dataset. This involves creating assumptions and interpretations about a given data. Time series forecasting makes use of the best-fitting model essential to predicting future observations based on the complex processing of current and previous data [19]. Figure 2.1 shows a sample for independent uncorrelated variables.

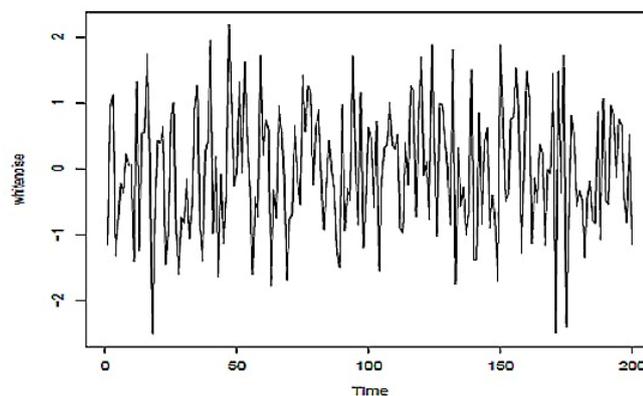


Figure 2.1 Plot of independent uncorrelated random variable [20]

From Figure 2.1, it is evident that the plot shows no clear pattern or trend, highlighting the uncorrelated nature of the random variables. Each point is independent and does not provide predictive information about any other point. Consequently, the dataset is unsuitable for forecasting because it lacks any underlying pattern, trend, or correlation that could be used to predict future values [20].

2.2.1 Type of Time Series Analysis

Time series analysis involves the study and interpretation of data collected over successive points in time. The choice of analysis depends on the specific characteristics of the time series data and the goals of the analysis, such as forecasting, anomaly detection, or understanding underlying patterns [21]. The following topic will explain type of time series model.

(i) Box Jenkins model

The Box-Jenkins Model is a mathematical framework developed to predict data ranges by utilizing inputs from a designated time series. This model is versatile and capable of analysing various types of time series data for forecasting purposes. Originating from the collaborative efforts of mathematicians George Box and Gwilym Jenkins, the foundational concepts of this model were outlined in their 1970 publication titled "Time Series Analysis: Forecasting and Control." Specifically tailored for short to medium-term predictions, the Box-Jenkins Model is most effective when forecasting within time frames of 18 months or less [22].

The Box-Jenkins Model forecasts data using three principles: autoregression, differencing, and moving average. These three principles are known as p , d , and q , respectively. The autoregression (p) process tests the data for its level of stationarity. If the data being used is stationary, it can simplify the forecasting process. If the data being used is non-stationary it will need to be differenced (d). The data is also tested for its moving average fit (which is done in part q of the analysis process). Overall, initial analysis of the data prepares it for forecasting by determining the parameters (p , d , and q), which are then applied to develop a forecast.

(ii) ARMA model

The time series ARMA (Autoregressive Moving Average) model is a statistical approach used for analysing and forecasting time series data. The ARMA model combines two key components: Autoregressive (AR) and Moving Average (MA) [23].

The model is defined by two parameters, p and q , which represent the number of autoregressive and moving average terms in the model, respectively. The ARMA model is used to forecast future values of a time series based on its past values. The process of forecasting using an ARMA model involves estimating the model parameters, fitting the model to the data, and then using the model to make predictions [24].

(iii) ARIMA model

An autoregressive integrated moving average, abbreviated as ARIMA, is a statistical model employed for analysing time series data. Its purpose is to enhance comprehension of the dataset or forecast future trends. The ARIMA model is a form of regression analysis that gauges the strength of one dependent variable relative to other changing variables. The model's goal is to predict future values by examining the differences between values in the series instead of through actual values [20].

The ARIMA model is an enhancement of the ARMA model. The ARIMA model extends the ARMA model by including an "integration" component, which allows it to handle non-stationary data. Each component in ARIMA functions as a parameter with a standard notation. For ARIMA models, a standard notation would be ARIMA with p , d , and q , where integer values substitute for the parameters to indicate the type of ARIMA model used. In the ARIMA model, differencing is applied to the data to achieve stationarity. Stationarity indicates a consistent pattern in the data over time. The aim of differencing is to eliminate any trends or seasonal patterns within the data.

2.3 Factors that Contribute to The Time Series Analysis

The large-scale deployment of photovoltaics (PV) for generating electricity plays an important role to mitigate global warming. However, the fluctuation in PV output power presents difficulties in managing the power grid. Generally, PV output power relies on historical solar irradiance data, associations among meteorological variables, and the impact of weather conditions in spatially neighbouring areas [25].

A study in [26], the forecast utilizes data from a PV power station in the Ashland region of the United States. Numerous weather conditions impact PV output as well as environmental factors. Figure 2.2 depicts a graph illustrating the correlation of PV power generation in typical weather.

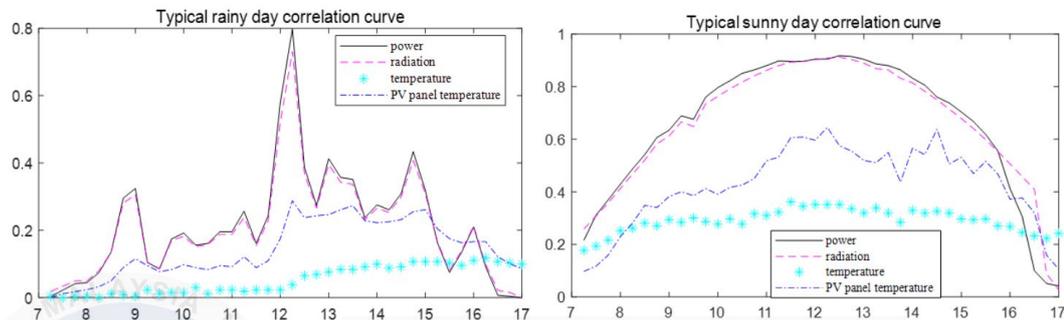


Figure 2.2: Typical weather factor correlation curve [26].

Figure 2.2 compare PV power generation on a typical rainy day (left) and a typical sunny day (right). On rainy days, power generation and solar radiation fluctuate significantly, indicating inconsistent and less efficient PV performance due to cloud cover. In contrast, sunny days show a smooth, consistent increase in power generation and solar radiation peaking around midday, with higher and more stable values, indicating optimal conditions for PV efficiency [26].

A study in [27] conducted an analysis on the behaviour of different types of PV modules under a range of temperature of 20-60°C. The outcome from the analysis shows CIGS PV modules seem to be better choice in high temperature conditions due to low temperature coefficients. With temperature increasing, the reverse saturation current increases rapidly which cause major changes in voltage rate.

In [28], the study analysed six meteorological factors as shown in Figure 2.3 with the goal of accurately predicting solar power. The six meteorological factors are proven affect the solar power forecast in the following descending order: solar radiation, sunlight, wind speed, temperature, cloud cover, and humidity. Solar radiation has the greatest influence on solar power forecasting, while humidity has effectively no influence on the solar power forecast.

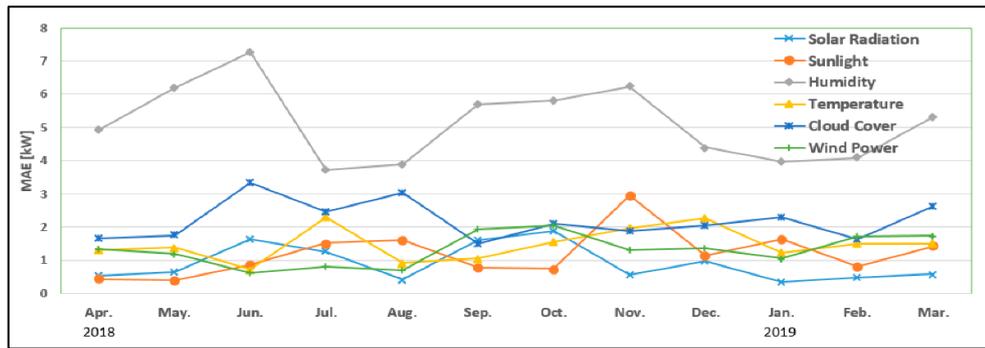


Figure 2.3: Meteorological factors affecting solar power forecasting by month [28]

In conclusion, the efficiency and performance of solar power systems are influenced by a multitude of factors such as cloud cover, humidity, temperature, and wind power. Each element plays a crucial role in determining the overall effectiveness of solar panels. As the world navigates the transition towards cleaner and sustainable energy sources, a comprehensive understanding of these influencing factors is essential for the effective design, installation, and solar forecast.

2.4 Related Case Studies on Time Series Analysis

Case studies are essential tools that offer real-world applications for theoretical concepts, enabling in-depth analyses of specific situations or events. They contribute to problem-solving skills, allowing individuals to learn from both successes and failures. Therefore, case studies on time series analysis are a valuable resource that help to improve and make decisions in analysing time series data.

2.4.1 A Guide to Solar Power Forecasting using ARMA Models

A study in [30], stated that solar power forecasting techniques can be generally categorized into two groups: (i) physics-based models, which predict solar power using numerical weather predictions and solar irradiation data, and (ii) statistical models, which forecast solar power directly based on historical data.

The data used in [30] is based in Australia. The sample comprises around nine months of data, with observations for 14 hours per day, spanning from 6:00 am to 7:00 pm. The author constructs an ARMA (p, q) model and employs Monte Carlo sampling based on the optimal ARMA model to generate hourly scenarios for each hour. Figure 2.4 shows estimated p and q values for the 14 hours of the day.

Hour	6:00	7:00	8:00	9:00	10:00	11:00	12:00
p	1	1	1	1	2	1	1
q	1	1	1	1	1	1	2
	13:00	14:00	15:00	16:00	17:00	18:00	19:00
p	1	1	1	1	1	2	1
q	1	1	1	1	1	1	2

Figure 2.4: Estimated p and q values for ARMA model for 14 hours [18]

The process involved formulating the ARMA model equation, stationarity checking, selecting parameters, and making predictions with the model. Equation (2.5) is the general equation for ARMA model.

$$ARMA(p, q) = x_t = \sum_{i=1}^p \phi_i x_{t-i} + \epsilon_t + C + \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (2.5)$$

From the equation, p is the order of the autoregressive polynomial, q is the order of the moving average polynomial, ϕ is the autoregressive model's parameter, θ is the moving average model's parameter, C is constant and ϵ_t is error terms (white noise). The output variable depends linearly on the current and various past values.

The accuracy of the time series model is measured by calculating MAE and RMSE. From the findings, the actual and predicted data is 3.3% and 5.1% of the maximum actual value respectively. In conclusion, this paper outlines a straightforward and concise process for fitting an ARMA model to historical solar radiation data.

2.4.2 ARMA Model for Short-term Forecasting of Solar Potential: Dakar Site

A study in [29] introduces a model for short-term solar potential forecasting using the ARMA model. The study involves analysing data collected by assessing energy production from PV solar sources in the Sahelian zone.

The ARMA model is used to predict global solar potential over the next 24 hours. The ARMA (p, q) model is utilized to identify the optimal p and q parameters for a more accurate fit to the variable under the consideration. Model calibration is conducted using data collected at the Dakar station, covering hourly records from October 2016 to September 2017. The selection of this model is based on its reliability and applicability across diverse global scales. The RStudio software is utilized for simulation purposes. Equation (2.6) is the ARMA models used in [29].

$$x_t = \sum_{i=1}^p \phi_i x_{t-i} + \epsilon_t + C + \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (2.6)$$

The parameter ϕ_i represents the autoregressive process, θ_i represents the moving average process, and ϵ_t is the residual of the model. To check the stationarity of a series, Dickey-Fuller test is used. The author notes that classical tests, such as the augmented Dickey-Fuller and Phillips Fuller tests, can also be employed. Additionally, for forecasting purposes, there are several methods used to verify the performance and reliability of the model. This paper specifically focuses on a few validation methods, using error measurements such as RMSE, MAE, and the coefficient of determination (R^2).

To summarize, the research in [29] demonstrates the reliability of the ARMA model for solar forecasting. This model can serve as a dependable decision support tool for the planning and management of PV solar power plants.

2.4.3 Forecasting of Total Daily Solar Energy Generation using ARIMA

A study in [30] use ARIMA model to predict the daily total solar energy output of a 10kW solar panel, accounting for both seasonal and nonseasonal variations. The solar panels were installed on the roof of the Group Nire building at the Reese Research Centre in Lubbock, Texas. The data spanned a full year, commencing from September 6, 2017, to November 5, 2018.

The application of the ARIMA model revealed the necessity of having stationary data. The author emphasized that non-stationary time series required transformation operations, such as differencing, logging, and deflating, to achieve stationarity in the time series. The ARIMA model equation is presented in (2.7), with the model's effectiveness depending on the values of autoregressive (p), moving average (q), and model differential (d).

$$ARIMA = \hat{y}_t = \mu + \phi_1 y_t + \dots + \phi_p \epsilon_{t-p} - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q} \quad (2.7)$$

\hat{y}_t is the d^{th} different of non-stationary time series of the model. The autoregressive lags term is denoted by p , while the differencing and moving average lag terms are represented by d and q , respectively. Parameters ϕ_i and θ_1 correspond to the AR and MA terms, respectively.

This study demonstrated the use of ARIMA statistical modelling for predicting the overall daily solar energy. The analysis involved determining the model parameters and assessing their validity through various criteria. Verification methods included the AIC as outlined in (2.8) where k is independently adjusted number of parameters and the sum of squared residuals (SSE) as specified in (2.9) where n is the data point, \hat{x} is the predicted value and x_i is the deviation of observe value. Essential error analysis was used to evaluate the overall performance of the model.

$$AIC = -2 \log(\text{maximumlikelihood}) + 2k \quad (2.8)$$

$$SSE = \sum_{i=1}^n (x_i - \hat{x})^2 \quad (2.9)$$

The outcome shows that the model's accuracy was evaluated using R software. For the data gathered for this model, the MAE was 17.70%.

CHAPTER 3

METHODOLOGY

This section explains about the flowchart, data preparation, time series model and error measurement analysis in this project.

3.1 Introduction

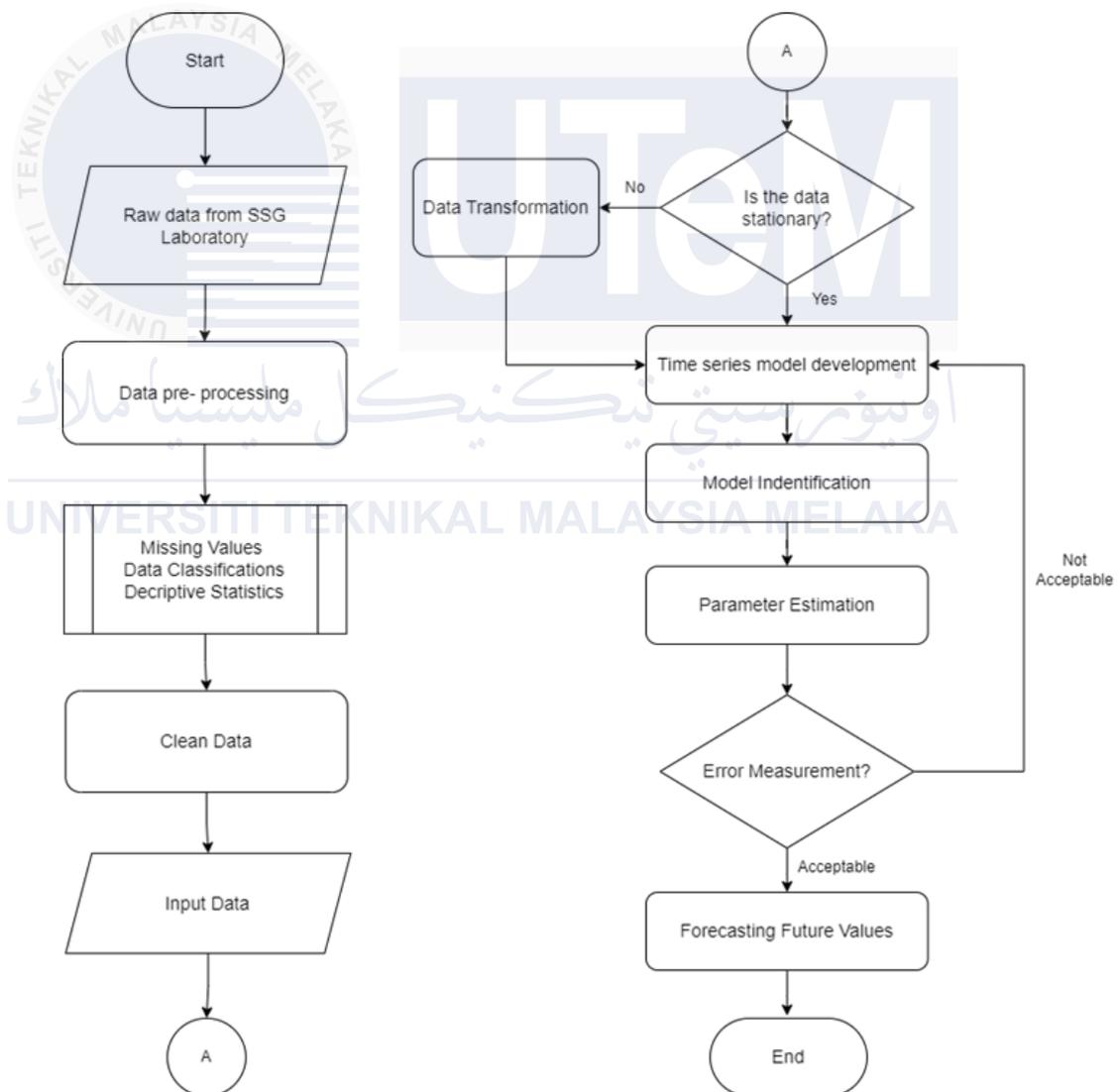


Figure 3.1 Flowchart of the project

Figure 3.1 shows the overall implementation process of this project. The process started with raw data acquired from the SSG laboratory. Next is data pre-processing to calculate descriptive statistics, processing missing values and data merging. Afterwards is data cleaning, where the necessary data will be arranged in a table format. As a result, the input data is ready to be used in the time series analysis.

The input data will be examined to identify if the data is stationary or not. If the data is non-stationary, it will go through data transformation to achieve stationary data. Once the data is stationary, the data will go through time series model development. The model identification process is to identify which models to use between ARMA and ARIMA models. If the data is initially stationary, the data will use the ARMA model. Otherwise, if the data has gone through data transformation, the data will be used for the ARIMA model.

Later, the process continues with parameter estimation. The estimation parameters are p , q and d . Consequently, the error measurement of the time series model will be calculated. If the error measurement value is high, the process will be repeated starting from time series model development. If the values of error measurement are acceptable, the time series model is ready to forecast future values of specific yield.

3.2 Harvesting Raw Data

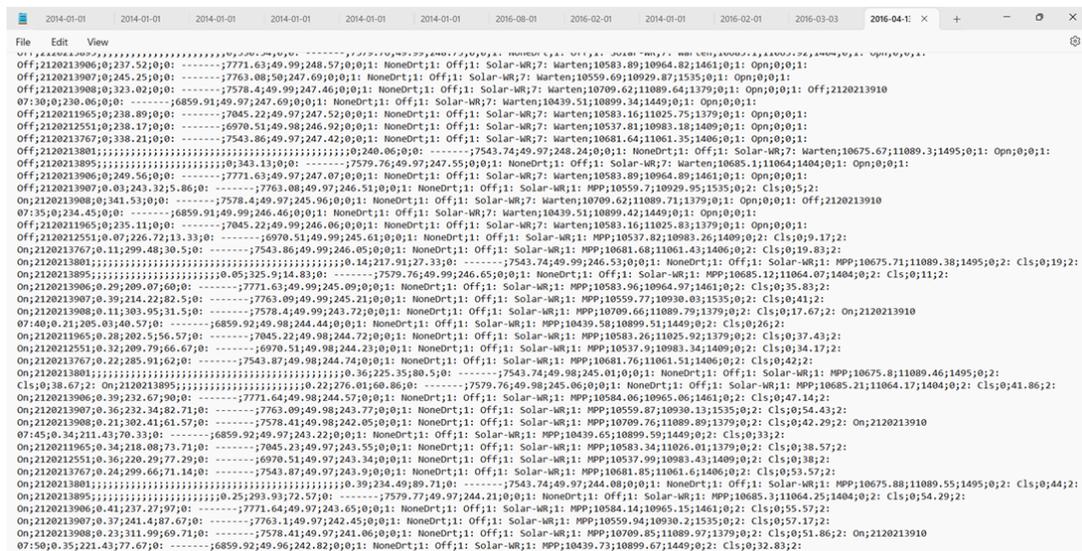


Figure 3.2: Raw data from SSG Laboratory

The raw solar radiation data was obtained through solar panels in FTKE. The data were gathered daily at intervals of five minutes for one year, from 1st January 2016 until 31st December 2016. Figure 3.2 shows the raw data in CSV file format collected from the SSG laboratory in FTKE. The collected data is transferred into Microsoft Excel, as shown in Figure 3.3. The data consists of timestamp, current, voltage, power, and frequency.

Figure 3.3: Imported data in Microsoft Excel file format.

3.3 Data Pre-Processing

3.3.1 Data Merging

The data were categorized based on the series number in the raw file. The raw data consists of solar radiation from polycrystalline, monocrystalline, TF and HIT solar panels in the FTKE area. The data is categorized based on series number. Table 3.1 shows the type of solar panel with its respective series number.

Table 3.1: Series number of type of solar panel

Type of solar panel	Series number
Polycrystalline	21202111965, 2120212551, 2120213767
Monocrystalline	2120213895, 2120213907, 2120213908
TF solar panel	2120213855, 2120213894, 2120213897
HIT solar panel	2120213801, 2120213906, 2120213910

Following this, the data is used for the calculation of specific yield. Specific yield in a solar plant refers to the amount of electricity generated by a photovoltaic system per unit of installed capacity [31]. Equation (3.1) is the formula for specific yield. Figure 3.4 is an example of the summary result of four types of solar panel merging data in February 2016.

$$\text{Specific Yield} = \frac{\text{System's annual average energy yield (kWh)}}{\text{System's installed capacity (kWp)}} \quad (3.1)$$

DATE	POLY-I	POLY-II	POLY-III	E-TOTAL	S-YIELD
1	10.12	10.29	10.13	30.54	5.194
2	9.27	9.47	9.29	28.03	4.767
3	7.19	7.31	7.21	21.71	3.692
4	8.23	8.35	8.22	24.8	4.218
5	4.63	4.74	4.68	14.05	2.389
6	6.72	6.89	6.78	20.39	3.468
7	6.07	6.2	6.11	18.38	3.126
8	8.02	8.19	8.07	24.28	4.129
9	12.91	13.22	13.03	39.16	6.660
10	4.73	4.85	4.79	14.37	2.444
11	7.28	7.42	7.32	22.02	3.745
12	9.75	10.02	9.81	29.58	5.031
13	8.84	9.01	8.86	26.71	4.543
14	7.55	7.67	7.58	22.8	3.878
15	9.67	9.8	9.67	29.14	4.956
16	10.51	10.83	10.63	31.97	5.437
17	9.73	10.02	9.88	29.63	5.039
18	7.06	7.21	7.12	21.39	3.638
19	3.62	3.73	3.69	11.04	1.878
20	7.85	8.02	7.91	23.78	4.044
21	9.86	10.19	10.03	30.08	5.116
22	10.5	10.83	10.67	32	5.442
23	10.3	10.56	10.43	31.29	5.321
24	10.13	10.36	10.22	30.71	5.223
25	10.32	10.67	10.54	31.53	5.362
26	10.03	10.4	10.31	30.74	5.228
27	9.65	10	9.9	29.55	5.026
28	9.98	10.3	10.21	30.49	5.185
29	9.95	10.01	9.89	29.85	5.077

(a) Polycrystalline solar panel

DATE	HIT-I	HIT-II	HIT-III	E-TOTAL	S-YIELD
1	10.72	10.73	10.71	32.16	5.702
2	9.77	9.8	9.78	29.35	5.204
3	7.52	7.56	7.54	22.62	4.011
4	8.54	8.56	8.56	25.66	4.550
5	4.71	4.74	4.75	14.2	2.518
6	6.96	7.01	6.99	20.96	3.716
7	6.22	6.25	6.24	18.71	3.317
8	8.26	8.3	8.3	24.86	4.408
9	13.35	13.44	13.43	40.213	7.130
10	4.83	4.87	4.86	14.56	2.582
11	7.47	7.52	7.51	22.5	3.989
12	10.35	10.39	10.38	31.12	5.518
13	9.21	9.23	9.21	27.65	4.902
14	7.82	7.85	7.85	23.52	4.170
15	10.15	10.18	10.15	30.48	5.404
16	11.26	11.29	11.24	33.79	5.991
17	10.31	10.35	10.32	30.98	5.493
18	7.37	7.4	7.38	22.15	3.927
19	3.64	3.69	3.68	11.01	1.952
20	8.12	8.15	8.14	24.41	4.328
21	10.46	10.5	10.48	31.44	5.574
22	11.31	11.33	11.29	33.93	6.016
23	11.04	11.07	11.03	33.14	5.876
24	10.69	10.71	10.68	32.08	5.688
25	11.03	11.07	11.03	33.13	5.874
26	10.75	10.81	10.76	32.32	5.730
27	10.38	10.42	10.38	31.18	5.528
28	10.68	10.72	10.67	32.07	5.686
29	10.37	10.4	10.34	31.11	5.516

(b) HIT solar panel

DATE	TF-I	TF-II	TF-III	E-TOTAL	S-YIELD
1	12.99	13.1	13.16	39.25	6.290
2	11.46	11.63	11.65	34.74	5.567
3	9.01	9.11	9.14	27.26	4.369
4	9.98	10.13	10.13	30.24	4.846
5	5.55	5.67	5.65	16.87	2.704
6	8.03	8.17	8.16	24.36	3.904
7	7.52	7.6	7.63	22.75	3.646
8	9.49	9.66	9.67	28.82	4.619
9	15.72	15.95	15.99	47.66	7.638
10	5.82	5.89	5.92	17.63	2.825
11	8.63	8.77	8.79	26.19	4.197
12	12.1	12.31	12.31	36.72	5.885
13	10.9	11.05	11.09	33.04	5.295
14	9.1	9.27	9.28	27.65	4.431
15	12.21	12.29	12.35	36.85	5.905
16	13.35	13.52	13.58	40.45	6.482
17	12.14	12.35	12.38	36.87	5.909
18	8.89	8.96	9	26.85	4.303
19	4.3	4.41	4.43	13.14	2.106
20	9.37	9.53	9.54	28.44	4.558
21	12.05	12.28	12.29	36.62	5.869
22	13.46	13.61	13.68	40.75	6.530
23	13.17	13.35	13.41	39.93	6.399
24	12.48	12.69	12.71	37.88	6.071
25	12.99	13.18	13.22	39.39	6.312
26	12.43	12.64	12.65	37.72	6.045
27	12.42	12.57	12.6	37.59	6.024
28	12.72	12.9	12.93	38.55	6.178
29	12.57	12.62	12.69	37.88	6.071

(c) Thin-Film solar panel

DATE	MONO-I	MONO-II	MONO-III	E-TOTAL	S-YIELD
1	10.55	11.36	11.42	33.33	5.446
2	9.43	10.38	10.55	30.36	4.961
3	7.78	8.02	8.18	23.98	3.918
4	8.36	9.14	9.3	26.8	4.379
5	5.02	5.21	5.32	15.55	2.541
6	6.97	7.51	7.65	22.13	3.616
7	6.63	6.79	6.88	20.3	3.317
8	8.18	8.94	9.06	26.18	4.278
9	13.86	14.45	14.67	42.977	7.022
10	5.21	5.34	5.41	15.96	2.608
11	7.85	8.09	8.2	24.14	3.944
12	9.95	10.85	11.07	31.87	5.208
13	9.17	9.77	9.94	28.88	4.719
14	7.99	8.39	8.53	24.91	4.070
15	10.44	10.71	10.6	31.75	5.188
16	10.98	11.73	11.74	34.45	5.629
17	10.14	10.92	11.06	32.12	5.248
18	7.66	7.95	8.05	23.66	3.866
19	3.99	4.09	4.15	12.23	1.998
20	8.36	8.77	8.89	26.02	4.252
21	10.25	11.12	11.33	32.7	5.343
22	10.99	11.68	11.76	34.43	5.626
23	10.88	11.46	11.54	33.88	5.536
24	10.75	11.28	11.45	33.48	5.471
25	10.87	11.54	11.63	34.04	5.562
26	10.54	11.33	11.47	33.34	5.448
27	10.17	10.76	10.94	31.87	5.208
28	10.55	11.08	11.19	32.82	5.363
29	10.74	10.74	10.86	32.34	5.284

(d) Monocrystalline solar panel

Figure 3.4: The summary results of data merging in February 2016 for (a) polycrystalline (b) HIT (c) Thin-Film (d) monocrystalline solar panels

3.3.2 Missing Values

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		POLYCRYSTALLINE				HIT			TF		MONOCRYSTALLINE		
2		2120211965	2120212551	2120213767	2120213801	2120213906	2120213910	2120213855	2120213894	2120213897	2120213895	2120213907	2120213908
3	TimeStamp	E-Total	E-Total	E-Total	E-Total	E-Total	E-Total	E-Total	E-Total	E-Total	E-Total	E-Total	E-Total
4	hh:mm	kWh	kWh	kWh	kWh	kWh	kWh	kWh	kWh	kWh	kWh	kWh	kWh
5	07:05	6859.91	7045.22	6970.51							7543.74	7771.63	7763.08
6	07:10	6859.91	7045.22	6970.51	7543.86	7579.76	7578.4				7543.74	7771.63	7763.08
7	07:15	6859.91	7045.22	6970.51	7543.86	7579.76	7578.4				7543.74	7771.63	7763.08
8	07:20	6859.91	7045.22	6970.51	7543.86	7579.76	7578.4				7543.74	7771.63	7763.08
9	07:25	6859.91	7045.22	6970.51	7543.86	7579.76	7578.4				7543.74	7771.63	7763.08
10	07:30	6859.91	7045.22	6970.51	7543.86	7579.76	7578.4				7543.74	7771.63	7763.08
11	07:35	6859.91	7045.22	6970.51	7543.86	7579.76	7578.4				7543.74	7771.63	7763.09
12	07:40	6859.92	7045.22	6970.51	7543.87	7579.76	7578.41				7543.74	7771.64	7763.09
13	07:45	6859.92	7045.23	6970.51	7543.87	7579.77	7578.41				7543.74	7771.64	7763.1
14	07:50	6859.92	7045.23	6970.51	7543.88	7579.77	7578.41				7543.75	7771.65	7763.1
15	07:55	6859.93	7045.24	6970.52	7543.88	7579.78	7578.42				7543.75	7771.65	7763.11
16	08:00	6859.93	7045.24	6970.53	7543.89	7579.79	7578.43				7543.76	7771.67	7763.12
17	08:05	6859.94	7045.25	6970.54	7543.91	7579.8	7578.45				7543.77	7771.68	7763.13
18	08:10	6859.96	7045.27	6970.56	7543.94	7579.83	7578.48				7543.79	7771.71	7763.17
19	08:15	6859.97	7045.29	6970.58	7543.97	7579.87	7578.51				7543.81	7771.76	7763.21
20	08:20	6859.99	7045.31	6970.61	7544	7579.9	7578.55	8533.2	8653.53	8669.45	7543.83	7771.79	7763.25
21	08:25	6860.01	7045.33	6970.63	7544.04	7579.94	7578.58	8533.22	8653.54	8669.48	7543.85	7771.83	7763.29
22	08:30	6860.04	7045.36	6970.67	7544.07	7579.98	7578.63				7543.87	7771.88	7763.34
23	08:35	6860.06	7045.38	6970.7	7544.12	7580.03	7578.67	8533.24	8653.56	8669.5	7543.89	7771.93	7763.38
24	08:40	6860.07	7045.41	6970.73	7544.16	7580.07	7578.71	8533.27		8669.53	7543.92	7771.97	7763.42
25	08:45	6860.09	7045.43	6970.76	7544.17	7580.07	7578.71	8533.3		8669.56	7543.95	7772.01	7763.46
26	08:50	6860.12	7045.47	6970.8				8533.34		8669.59	7543.99	7772.05	7763.5

Figure 3.5: Datasheet with missing values

From the data file, missing values were spotted in the spreadsheet at random timestamps. Figure 3.5 shows an example of a datasheet with missing values. To overcome this problem, the average of the past three consecutive readings is used to estimate the missing values. Figure 3.6 shows the list of missing values in February 2016.

Date	Solar	Time	Date	Solar	Time	Date	Solar	Time
12/2/2016	POLYCRYSTALLINE	19:20	11/2/2016	POLYCRYSTALLINE		21/2/2016	POLYCRYSTALLINE	
	HIT	7:25, 19:20		HIT	7:25		HIT	7:25
	TF	7:25, 19:20		TF	7:25		TF	7:25
	MONOCRYSTALLINE	19:20		MONOCRYSTALLINE	19:25		MONOCRYSTALLINE	
22/2/2016	POLYCRYSTALLINE	7:20	12/2/2016	POLYCRYSTALLINE	7:20	22/2/2016	POLYCRYSTALLINE	7:20
	HIT	7:20		HIT	7:20		HIT	7:20
	TF	7:20		TF	7:20		TF	7:20
	MONOCRYSTALLINE	7:20		MONOCRYSTALLINE	7:20		MONOCRYSTALLINE	7:20
32/2/2016	POLYCRYSTALLINE		13/2/2016	POLYCRYSTALLINE		23/2/2016	POLYCRYSTALLINE	7:20, 19:25
	HIT	7:25		HIT	7:25, 19:55		HIT	7:20, 19:25
	TF	7:25		TF	7:25, 19:55		TF	7:20, 19:25
	MONOCRYSTALLINE			MONOCRYSTALLINE	7:25		MONOCRYSTALLINE	7:20, 19:25
4/2/2016	POLYCRYSTALLINE	19:05	14/2/2016	POLYCRYSTALLINE	7:25, 19:15	24/2/2016	POLYCRYSTALLINE	7:15
	HIT	7:25, 19:00, 19:05		HIT	7:25, 19:15		HIT	7:15
	TF	7:25, 19:00, 19:05		TF	7:25, 19:15		TF	7:15
	MONOCRYSTALLINE			MONOCRYSTALLINE	19:15		MONOCRYSTALLINE	7:15
5/2/2016	POLYCRYSTALLINE	7:25	15/2/2016	POLYCRYSTALLINE		25/2/2016	POLYCRYSTALLINE	7:20
	HIT	7:25, 19:25		HIT	7:25, 19:15		HIT	7:20
	TF	7:25, 19:25		TF	7:25, 19:15		TF	7:20
	MONOCRYSTALLINE	7:25		MONOCRYSTALLINE			MONOCRYSTALLINE	
6/2/2016	POLYCRYSTALLINE	7:25	16/2/2016	POLYCRYSTALLINE	7:20	26/2/2016	POLYCRYSTALLINE	19:25
	HIT	7:25, 19:20		HIT	7:20, 19:20, 19:25		HIT	7:20, 19:25
	TF	7:25, 19:20		TF	7:20, 19:25		TF	
	MONOCRYSTALLINE	7:25, 19:20		MONOCRYSTALLINE	7:20, 19:25		MONOCRYSTALLINE	
7/2/2016	POLYCRYSTALLINE		17/2/2016	POLYCRYSTALLINE	7:15, 7:20	27/2/2016	POLYCRYSTALLINE	19:25
	HIT	7:30		HIT	7:15, 7:20		HIT	7:20, 19:25
	TF	7:30		TF	7:15, 7:20		TF	7:20, 19:25
	MONOCRYSTALLINE			MONOCRYSTALLINE	7:15, 7:20		MONOCRYSTALLINE	19:25
8/2/2016	POLYCRYSTALLINE		18/2/2016	POLYCRYSTALLINE	19:25	28/2/2016	POLYCRYSTALLINE	7:15
	HIT	7:30		HIT	7:25, 19:20, 19:25		HIT	7:15
	TF	7:30		TF	7:25, (19:15 until 19:25)		TF	7:15
	MONOCRYSTALLINE			MONOCRYSTALLINE	19:25		MONOCRYSTALLINE	
9/2/2016	POLYCRYSTALLINE		19/2/2016	POLYCRYSTALLINE	19:25	29/2/2016	POLYCRYSTALLINE	19:25
	HIT			HIT	7:25, 19:25		HIT	19:25
	TF			TF	7:25, 19:25		TF	19:25
	MONOCRYSTALLINE			MONOCRYSTALLINE	19:25		MONOCRYSTALLINE	19:25
10/2/2016	POLYCRYSTALLINE	7:30	20/2/2016	POLYCRYSTALLINE	7:25, 19:15			
	HIT	7:30, 19:20		HIT	7:25, 19:20			
	TF	7:30, 19:20		TF	7:25, 19:20			
	MONOCRYSTALLINE			MONOCRYSTALLINE	7:25			

Figure 3.6: List of missing values in February 2016

3.3.3 Descriptive Statistics

Descriptive statistics are used to describe the essential characteristics of a data set in an analysis, which can represent the entire population or a sample of the population [32]. The descriptive statistics for this study include minimum value, maximum value, mean and standard deviation.

The maximum value is the most significant or highest value observed in a data set, and the minimum value is the smallest value observed in the data set. The mean is the average value of a time series over a given period. Mean is calculated by summing up all the values in the data set and dividing it by the total amount of data. The mean formula is shown in equation (3.2). The standard deviation measures the dispersion or variability of the data points within a specific time series data set [33]. It indicates how spread out the variables are from the mean [30]. The equation of standard deviation is shown in equation (3.3).

$$\text{Mean} = \frac{\sum_{i=1}^n x_i}{n} \quad (3.2)$$

where $\sum_{i=1}^n x_i$ is the summing of total variable and n is the number of variables.

$$\text{Standard Deviation} = \sqrt{\frac{\sum_{i=1}^n (X - x_i)^2}{n - 1}} \quad (3.3)$$

where x is the value of observation data, x_i is the mean value and n is the amount of data.

3.4 Time Series Model

Time series data analysis is the way to predict time series based on past behaviour. Prediction is made by analysing underlying patterns in the time-series data [14].

3.4.1 ARMA Model

An ARMA model is type of statistical model that are used in time series analysis for stationary data. ARMA model is a combination of an AR and MA in the model. The model is a model that predicts the current value based on the past values from the same time series. Equation (3.4) shows the equation for AR model.

$$AR(p) = x_t = \sum_{i=1}^p \phi_i x_{t-i} + \varepsilon_t + C \quad (3.4)$$

where x_t is the level of current observation; ϕ_i parameter of autoregressive process; p is the term of lagged observation; ε_t is an additional white noise error term and C is constant.

The MA model create current values based on the errors from the past forecasts instead of using the past values like AR. Equation (3.5) shows the equation for MA model.

$$MA(q) = x_t = \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (3.5)$$

where x_t is the level of current observation; q is the term of moving average value; θ_i is parameter of moving average process; ε_t is a residual of the model. The model development combine (3.4) and (3.5) to form the general formula of ARMA model as shown in equation (3.6). This equation observes the output variables depends linearly on the current and past value.

$$ARMA(p, q) = x_t = \sum_{i=1}^p \phi_i x_{t-i} + \varepsilon_t + C + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (3.6)$$

3.4.2 ARIMA Model

ARIMA model is a type of model used in time series analysis for analysing and forecasting time series data. The model can be used for analysing data with trends and seasonality.

The ARIMA model has three components which are (AR), (I) and (MA). AR refers to the dependence of current value on previous values in time series. The number of past values (p) is called the lag. Integrated (I) represents the differencing of raw observation to allow the time series to become stationary. The number of times series go through differencing (d) is called degree of differencing. Moving average (MA), refers to the dependency between an observation and a residual error from a moving average applied to lagged observation. The number of past errors used (q) is called the order of the moving average. The equation for ARIMA model is as shown in equation (3.7).

$$ARIMA(p, q, d) = \mu + \phi_1 y_t + \dots + \phi_p y_{t-p} - \theta_1 e_1 - \dots - \theta_q e_{t-q} \quad (3.7)$$

where ϕ_1 autoregression parameter; θ is moving average parameters; μ is constant; $\theta_1 e_1$ degree of differential; $\phi_p y_{t-p}$ is d^{th} a stationary ARIMA model. The parameters p , d and q can be determined using statistical techniques such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) [28].

3.4.3 Model Parameter Selection

Model parameter selection determine the performance and accuracy of the models used for forecasting time series analysis.

3.4.3.1 Stationary Data

A time series is said to be stationary if its statistical properties are constant and it does not exhibit seasonality. A time series is considered stationary if the value of mean or standard deviation is constant, and there is no trend or seasonality in the data. A non-stationary time series has the statistical properties change over time and there is a trend and seasonality component [34]. Figure 3.7 shows the difference data trend between stationary data and non-stationary data.

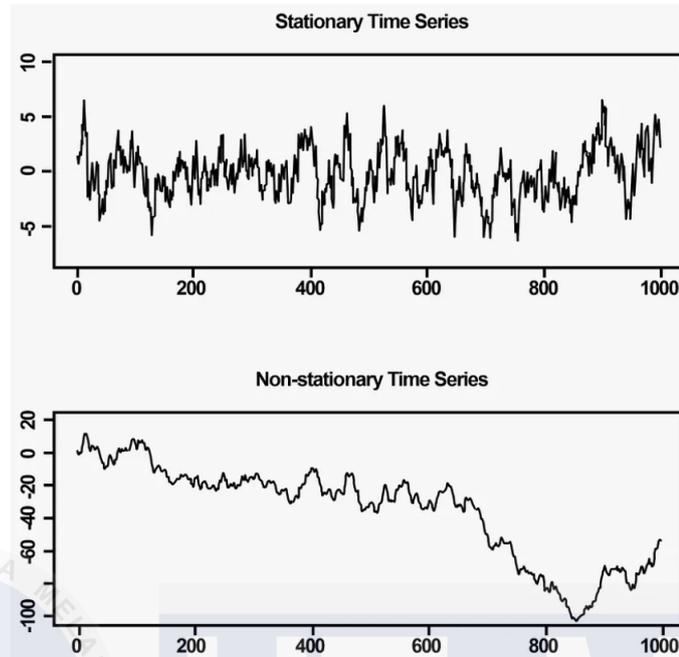


Figure 3.7: Time series plot of stationary and non-stationary data [34]

3.4.3.2 Autocorrelation Function (ACF)

The autocorrelation function (ACF) is a measure of correlation coefficient between the series and past values. ACF offers a way to examine the current value of the time series data and past value to identify trends and patterns. For the function to work, the complete data of time series need to be analysed with one or more lagged version of the data for comparison. As a result, the strength between variables can be evaluated [35].

The fundamental of the correlation at lag k is the correlation between the original data series, x_t and the same series moved forward one period represented as x_{t-1} . The autocovariance at lag k is defined in equation (3.8).

$$\text{cov}(x_t, x_{t-1}) = E(x_t - \mu)(x_{t-1} - \mu) \quad (3.8)$$

where E is the estimation; μ is the mean of the data observation. The autocorrelation at lag k is defined as in equation (3.9).

$$\text{ACF}, \rho_k = \frac{E(x_t - \mu)(x_{t-1} - \mu)}{\sigma_x^2} \quad (3.9)$$

where σ_x^2 is the variance of stochastic process; E is the estimation; μ is the mean of the data observation; k is specific number of periods. The ACF used in time series analysis to identify the appropriate order MA terms in a series model.

3.4.3.3 Partial Autocorrelation Function (PACF)

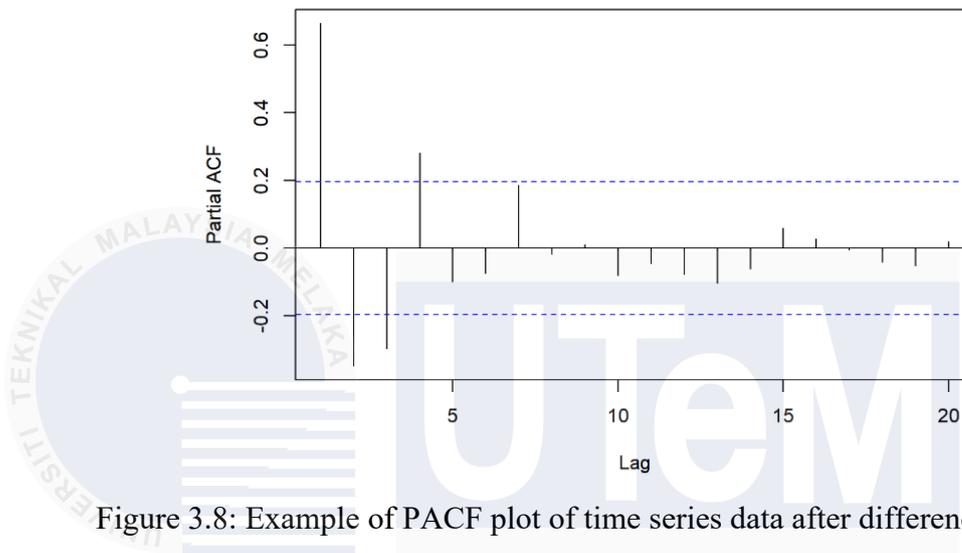


Figure 3.8: Example of PACF plot of time series data after differencing [36]

Figure 3.8 shows PACF plot after differencing. In time series analysis, the PACF gives the partial correlation of a stationary time series with its own lagged values, regressed the values of the time series at all shorter lags [37].

3.5 Error Measurement

Error measurement focus on measuring the accuracy and magnitude of error in the forecasted values when compared to the actual values. They emphasize the magnitude of errors rather than the specific direction and provide insights into the overall performance and precision of the forecasting model [38].

3.5.1 Mean Absolute Error (MAE)

MAE is a measure of the average size of the mistakes in a collection of predictions, without taking their direction into account. It is measured as the average absolute difference between the predicted values and the actual values and is used to assess the effectiveness of a regression model [39]. The MAE formula is shown in equation (3.10).

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - \hat{P}_i| \quad (3.10)$$

where n is number of observations; P_i is observed data; \hat{P}_i is predicted data, $|P_i - \hat{P}_i|$ is absolute value of the difference of $P_i - \hat{P}_i$.

3.5.2 Mean Squared Error (MSE)

The MSE indicates how close a regression line is to a set of points. It does this by taking the distances from the points to the regression line (these distances are the errors) and squaring them. The squaring is necessary to remove any negative signs. It also gives more weight to larger differences [40]. The MSE equation is as shown in (3.11).

$$MSE = \frac{1}{n} \sum_{i=1}^n (P_i - \hat{P}_i)^2 \quad (3.11)$$

where n is the number of observations of error; P_i is observed data; \hat{P}_i is predicted data, $(P_i - \hat{P}_i)^2$ is squared different absolute error.

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Preliminary Results

Total energy yield and specific yield were calculated. The data were obtained for each solar panel from 1 January 2016 until 31 December 2016. The data were collected between 7:00 am and 6:00 pm. There were 366 energy yield data (kWh) and specific yield (kWh/kWp). The preliminary results were analysed to detect the daily and monthly trends for each type of solar panel. The data is taken from four types of solar panels: Polycrystalline, Monocrystalline, HIT solar panel and TF solar panel.

4.1.1 Monthly Data Trend

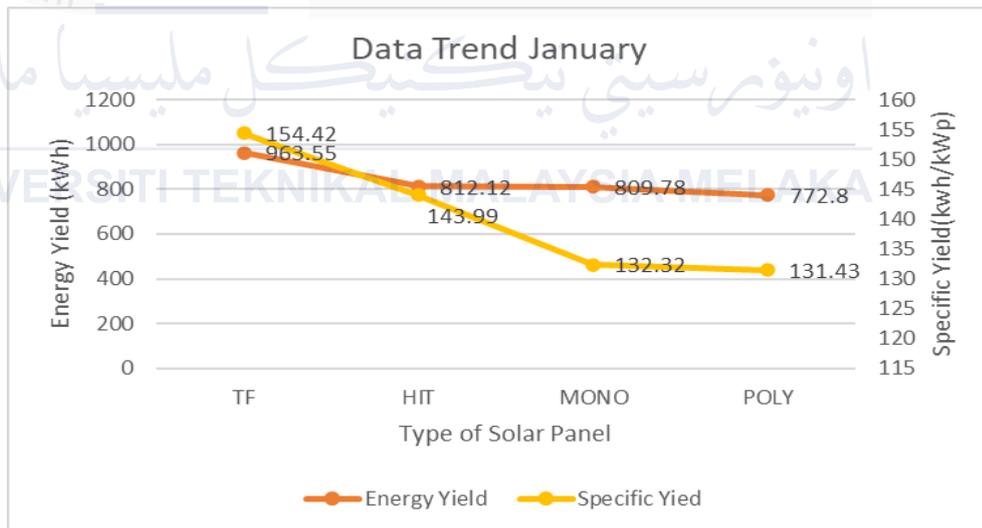


Figure 4.1: Total energy yield and specific yield in January 2016

Based on Figure 4.1, the highest energy yield generated by the TF solar panel is 963.55kWh, followed by HIT and monocrystalline. Polycrystalline (772.8kWh) generates the least energy yield. Data trend for specific yield in January shows that the highest specific yield is TF solar panel followed by HIT, monocrystalline and polycrystalline solar panel.

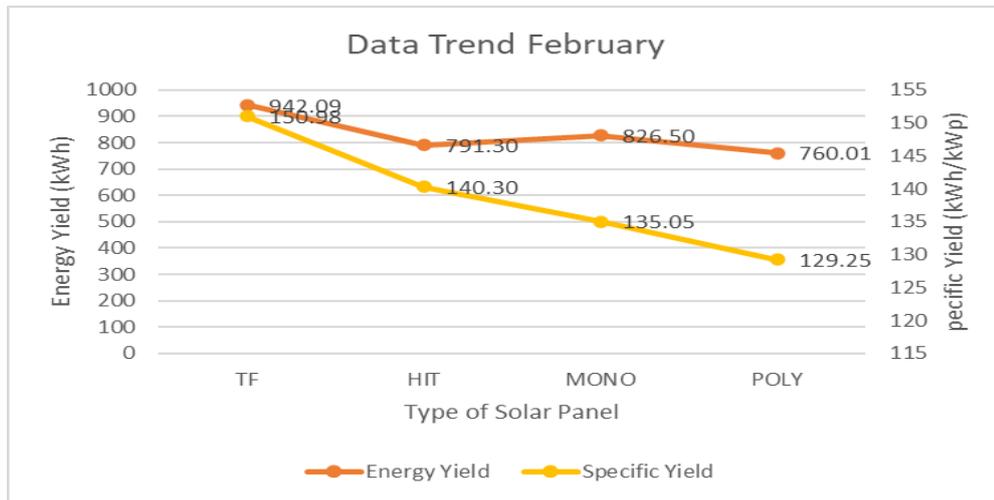


Figure 4.2: Total energy yield and specific yield in February 2016

Based on Figure 4.2, the energy yield decreased from the previous month except for the monocrystalline solar panel. The increased energy yield of the monocrystalline solar panels might be because of a decrease in humidity by 3% in February [41]. However, the reduction of the energy yield of three other solar panels might be affected by other factors such as shading, inverter efficiency, and dust.

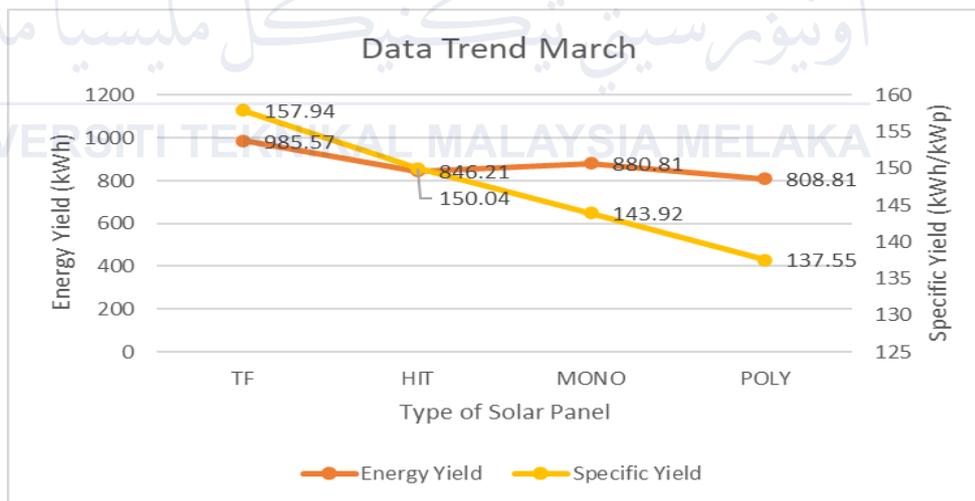


Figure 4.3: Total energy yield and specific yield in March 2016

In March 2016, the highest energy yield was by the TF solar panel, which was 985.57kWh, slightly higher than the month before. As shown in Figure 4.3, the increment of all types of solar panels may be affected due to the higher amount of solar radiation in March. Data trend for specific yield in March 2016 shows that TF solar panel has the highest value (157.94kWh/kWp) followed by HIT, monocrystalline and polycrystalline.

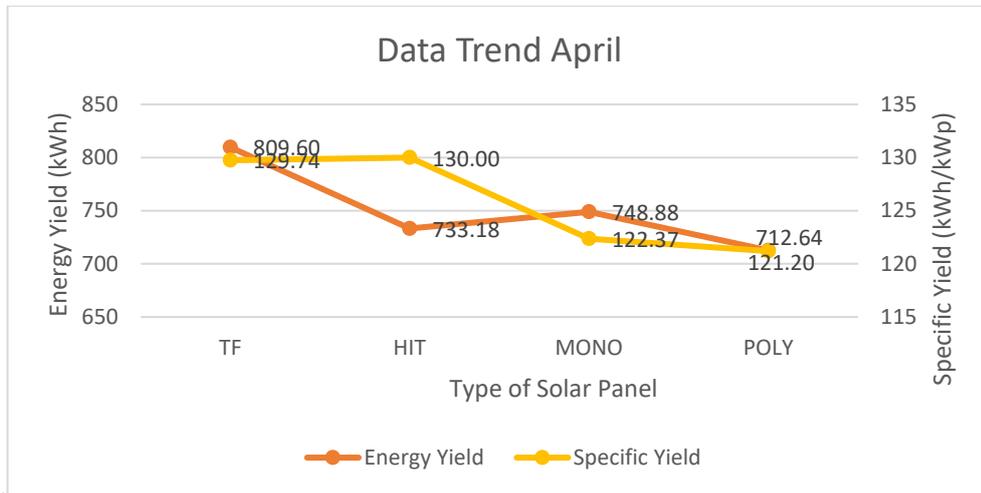


Figure 4.4: Total energy yield and specific yield in April 2016

In Figure 4.4, the TF solar panel generates the highest energy yield; the value is 809.60kWh lower than the previous month. Each solar panel's energy yield decreases: the decrement might be related to cloudy weather as humidity increased by 4% in April [41]. Data trend for specific yield in April shows that TF produced the highest specific yield; meanwhile, HIT has the lowest specific yield value.

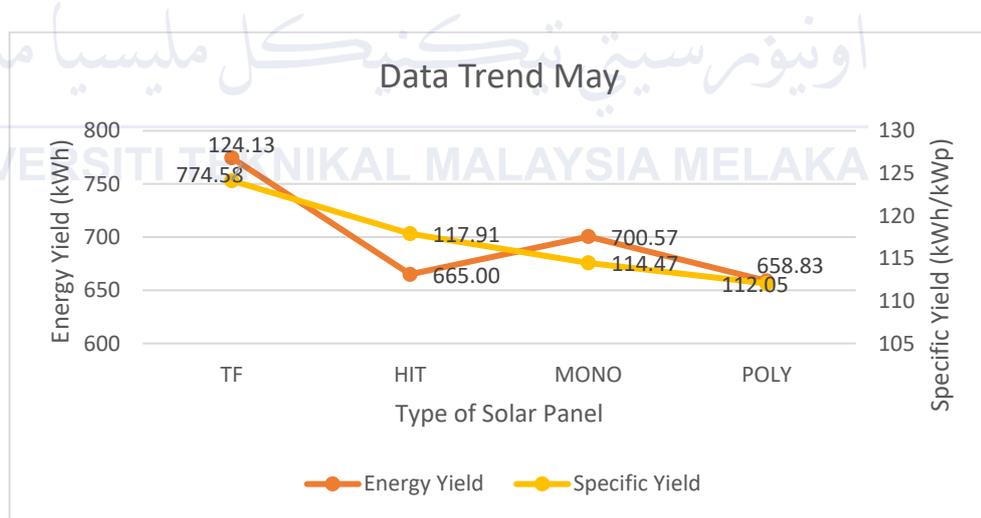


Figure 4.5 Total energy yield and specific yield in May 2016

In Figure 4.5, the values of energy yield and specific yield continue to decrease as the highest energy yield is TF with a value of 774.58kWh followed by monocrystalline (700.57kWh), polycrystalline (658.83kWh) and HIT solar panel (665kWh). The decrease in energy yield in May might be affected by the 2% increase in humidity [41]. Data trends for specific yield indicate that the highest is TF, followed by monocrystalline, HIT, and polycrystalline.

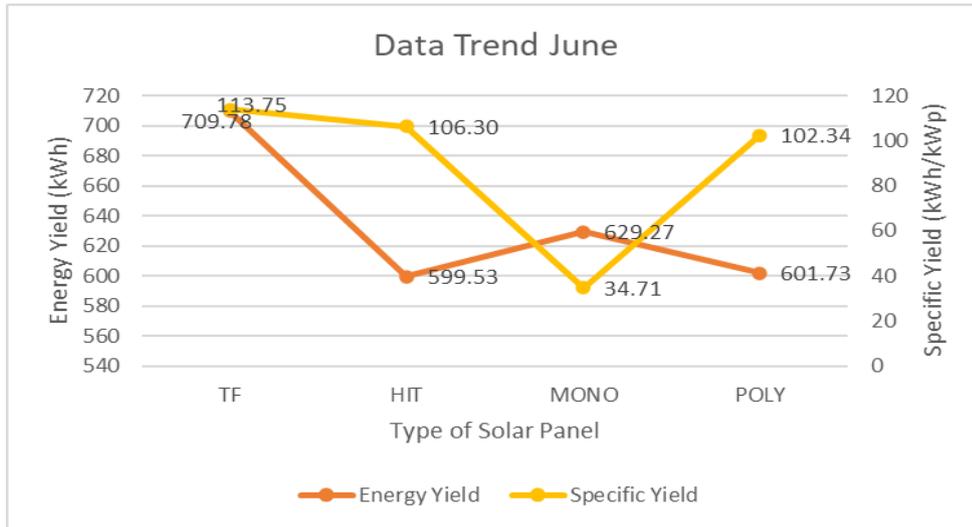


Figure 4.6: Total energy yield and specific yield in June 2016

Figure 4.6 shows that the energy yield and specific yield for June show decreases for all type of solar panels. However, the daily data trend shows that the specific yield value for monocrystalline solar decreased drastically by about 30% from the previous month.

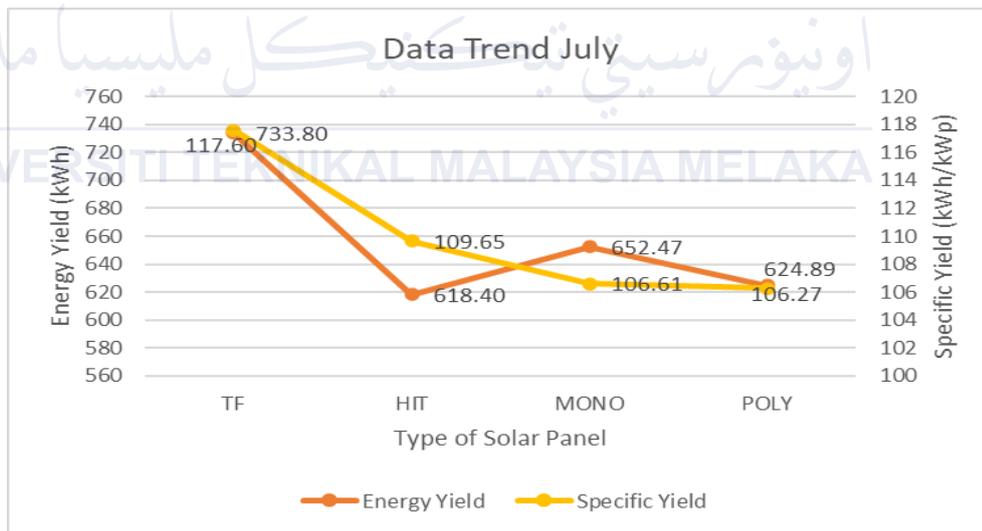


Figure 4.7: Total energy yield and specific yield in July 2016

In July, energy yield by all types of solar panels increases after five months of consecutive decrement. TF solar panels still produce the highest energy yield, followed by monocrystalline, polycrystalline and HIT. The factors affecting the increment could be the increase of solar radiation or the solar being cleaned. The data trend in Figure 4.7 shows that the TF solar panel has the highest specific yield (117kWh/kWp), and polycrystalline has the lowest value (106.27kWh/kWp).

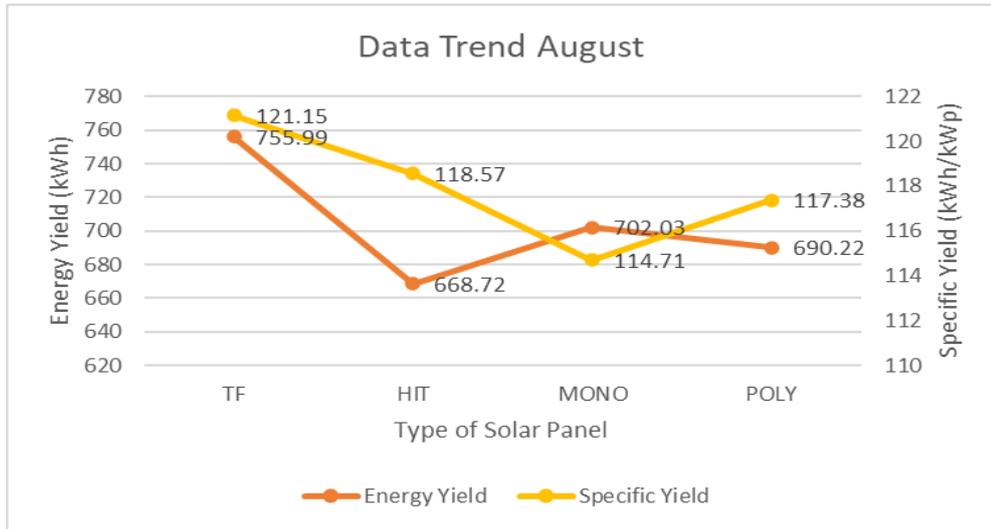


Figure 4.8: Total energy yield and specific yield in August 2016

From Figure 4.8, the value of energy yield and specific yield continues to increase. The humidity in Melaka decreased by 2% in August [41], which might correlate with the increment of energy yield in August. TF solar panels still dominate with the highest value of energy yield (756 kWh) and specific yield (121.15 kWh/kWp). The lowest value for energy yield is HIT solar panel (668.72kWh), while monocrystalline has the lowest specific yield value (114.71kWh/kWp).

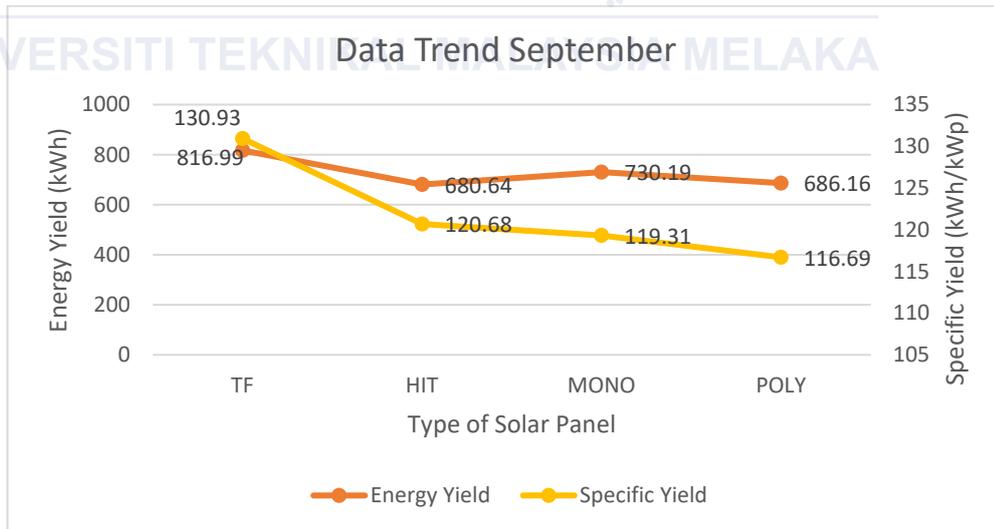


Figure 4.9: Total energy yield and specific yield in September 2016

Referring to Figure 4.9, all three types of solar panels have an increment of energy yield and specific yield. The solar panels mentioned are TF, HIT and monocrystalline. However, polycrystalline solar panels have slightly decreased from the previous month. This situation might be affected by shading, inverter efficiency,

or dust, which impacts the output power. Under those circumstances, polycrystalline has the second lowest value of energy yield (686.16kWh) and the lowest specific yield value (116.69kWh/kWp). TF has the highest value of energy yield (817kWh) and the highest value of specific yield (130.93kWh/kWp). HIT solar panels have the lowest energy yield of 680.64kWh.

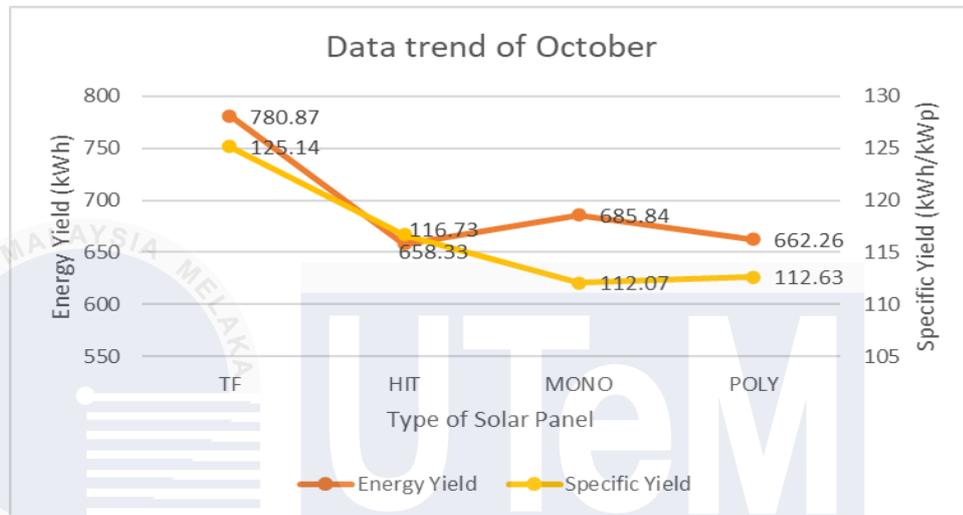


Figure 4.10: Total energy yield and specific yield in October 2016

Figure 4.10 shows that all four solar panels started to plummet again in October — as Melaka's humidity increased by 2% [41]. The highest energy yield produced by solar panels is the TF solar panel (780.87kWh), and the lowest is the HIT solar panel (658.33kWh). TF remains the highest for the specific yield, while monocrystalline has the lowest value.

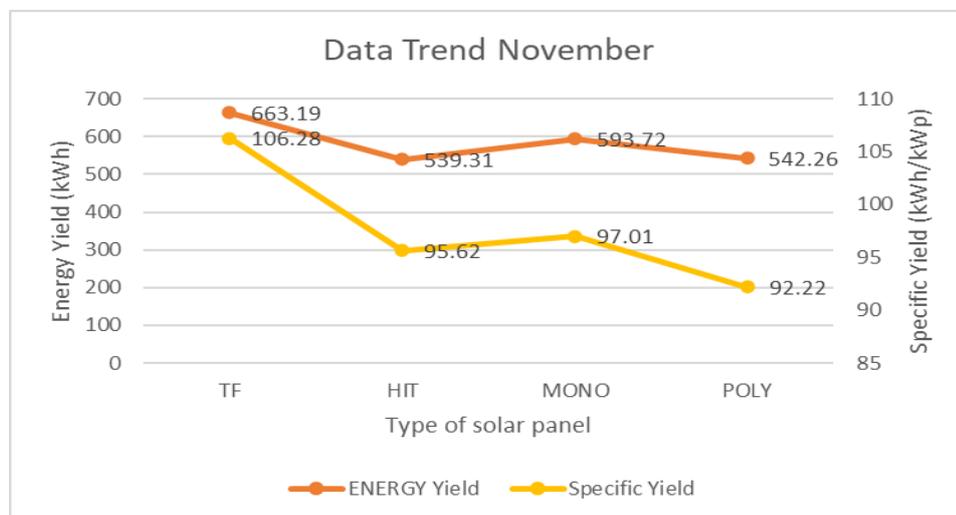


Figure 4.11: Total energy yield and specific yield November 2016

Based on Figure 4.11, November has the lowest energy and specific yield produced in 2016. The resulting output may be affected by humidity, as Melaka's highest percentage of humidity in 2016 was in November (86%) [41]. TF solar panels have the highest energy yield (663.19kWh) and specific yield (118.18kWh/kWp). Meanwhile, HIT solar panels have the lowest value of energy yield (594.41kWh) and polycrystalline has the lowest value of specific yield (92.22kWh/kWp).

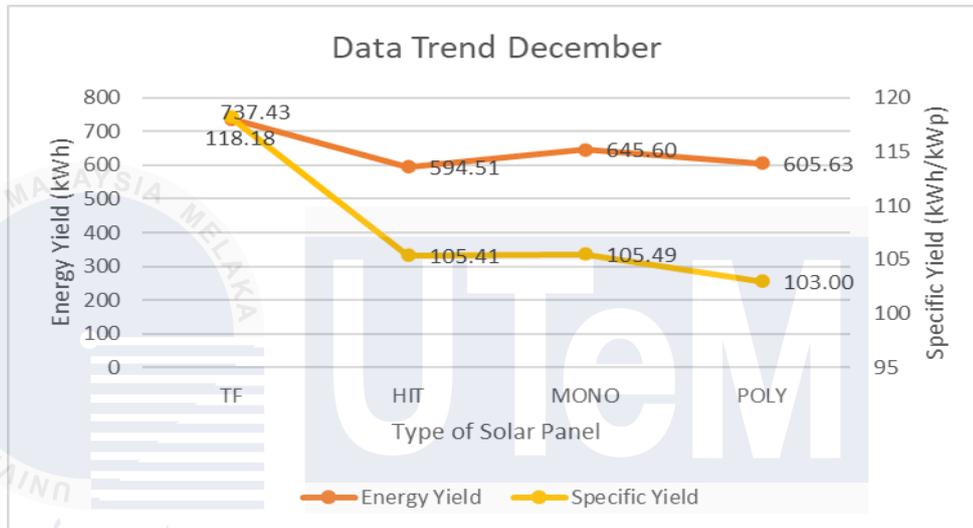


Figure 4.12: Total energy yield and specific yield in December 2016

Referring to Figure 5.12, in December 2016, the values started to increase from the previous month. TF solar panel continues its streak of highest value in energy yield (737.43kWh) and specific yield (118.18kWh/kWp). Monocrystalline has the second highest value of energy yield (645.6kWh) and specific yield (105.49kWh/kWp). The lowest energy yield is from HIT solar panels (594.51kWh), and polycrystalline solar panels have the lowest specific yield (103kWh/kWp).

In conclusion, TF solar panels consistently showed the highest energy and specific yield over the months. Polycrystalline has the lowest value for energy yield from January until May. However, from June until December, HIT solar panels have the lowest energy yield. For specific yield, polycrystalline has dominated the lowest value for nine months, and the remaining three months are from monocrystalline solar panels. The data indicated fluctuations in energy generation influenced by various factors such as dust, shading and meteorological factors.

4.1.2 Daily Data Trend

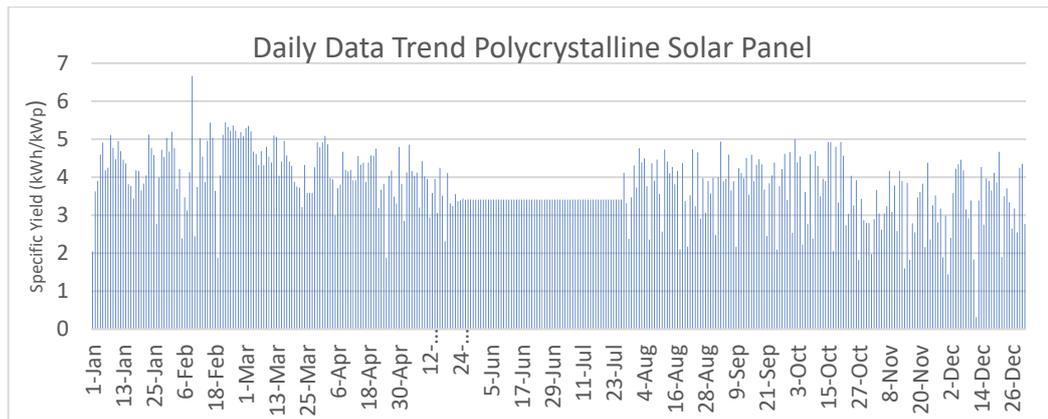


Figure 4.13: Daily trend for Polycrystalline solar panel

Figure 4.13 shows the trend in daily specific yield for polycrystalline solar panels in 2016. The specific yield of the solar system was calculated using equation (3.1), the energy yield values were acquired from the datasheet, and the kilowatt peak (kWp) data were obtained from the SSG laboratory website. Based on Figure 4.13, the highest specific yield that polycrystalline solar panels can produce is on 9th February 2016, with approximately 6.66kWh/kWp. The least specific yield was 0.32kWh/kWp on 12th December 2016. The constant value of the specific yield from 20th May until 26th July is due to the estimation of missing values in that period.

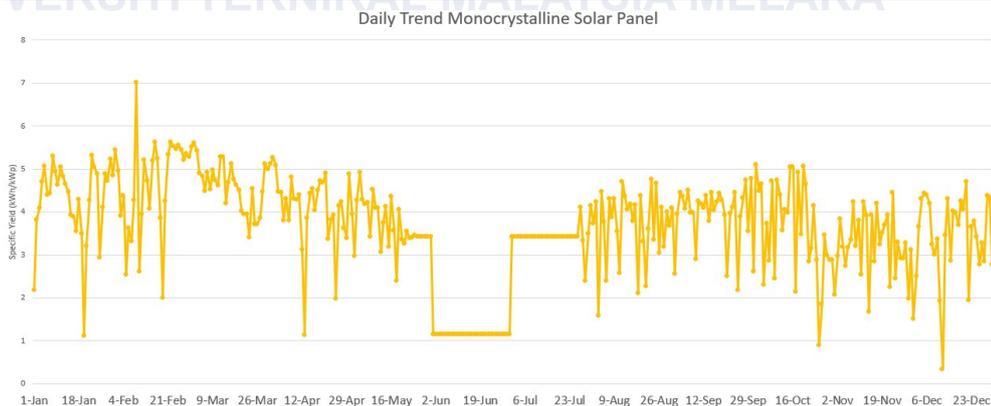


Figure 4.14 Daily trend for Monocrystalline solar panel

Figure 4.14 shows the trend in daily specific yield for monocrystalline solar panels in 2016. The highest specific yield that polycrystalline solar panels can produce is on 9th February 2016, with approximately 7.02kWh/kWp. The least specific yield was 0.34kWh/kWp on 12th December 2016. The static value of the specific yield from 20th May until 26th July is due to the estimation of missing values in that period.

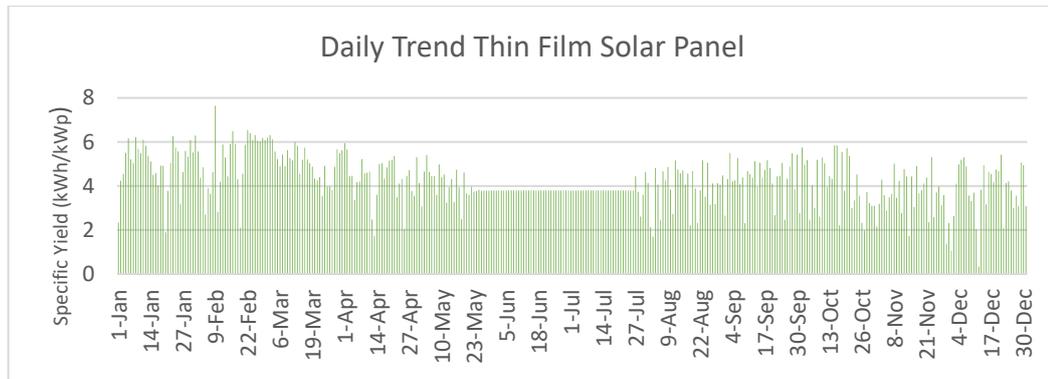


Figure 4.15: Daily trend for Thin Film solar panel

Figure 4.15 shows the trend in daily specific yield for monocrystalline solar panels in 2016. The highest specific yield that polycrystalline solar panels can produce is on 9th February 2016, with approximately 7.64kWh/kWp. The least specific yield is 0.34kWh/kWp on 12th December 2016. The constant value of the specific yield from 20th May until 26th July is due to the estimation of missing values in that period.

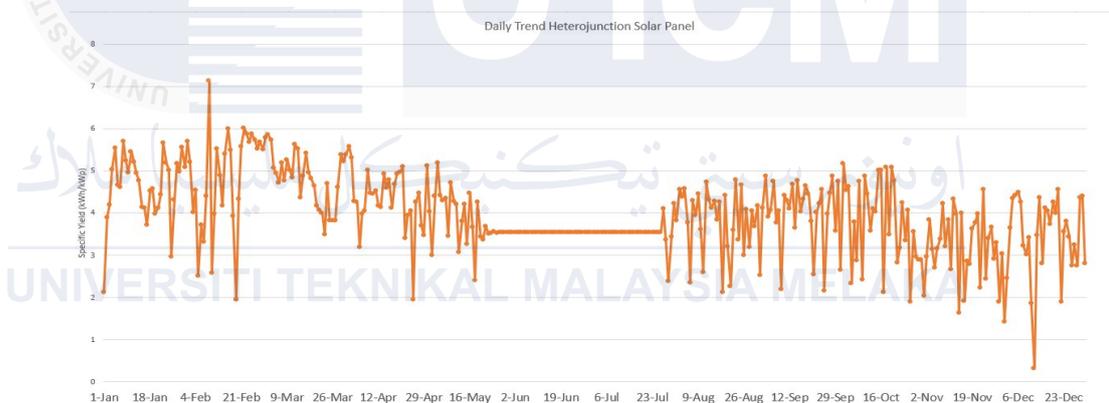


Figure 4.16: Daily trend for HIT solar panel.

Figure 4.16 shows the trend in daily specific yield for HIT solar panels in 2016. The highest specific yield that polycrystalline solar panels can produce is on 9th February 2016, with approximately 7.13kWh/kWp. The least specific yield was 0.33kWh/kWp on 12th December 2016. The static value of the specific yield from 20th May until 26th July is due to the estimation of missing values in that period.

In conclusion, the graph of daily data trends highlights the varying specific yields for different types of solar panels over the 12 months. The highest specific yield was around 7kWh/kWp, and the lowest was around 0.33kWh/kWp. Moreover, there has been a trend of decrement values of specific yield from the end of October 2016 until December 2016 for all type of solar panels in FTKE.

4.2 Descriptive Statistics

Descriptive statistics analysis provides insights into features and patterns of data. Various measures and methods are used to interpret data. Table 4.1 summarises four types of solar panels with descriptive statistics.

Table 4.1 Descriptive statistics of solar panels

Parameter	Polycrystalline solar panel	Monocrystalline solar panel	HIT solar panel	TF solar panel
Maximum Value	6.66	7.02	7.13	7.64
Minimum Value	0.32	0.34	0.33	0.34
Mean	3.78	3.66	3.98	4.24
Standard Deviation	0.83	1.16	0.93	1.07

Referring to Table 4.1, TF solar panel had the highest maximum value of specific yield in 2016 (7.64 kWh/kWp), followed by HIT, monocrystalline and polycrystalline. Hence, TF solar panels produced the highest energy yield for every peak power of the PV system. Meanwhile, polycrystalline had the lowest minimum value of specific yield in 2016 (0.32 kWh/kWp), followed by HIT, monocrystalline and TF solar panels. Therefore, polycrystalline solar panels produced the lowest energy yield for every peak power of the PV system. The result shows that TF solar panels are more efficient in hot and humid climates than polycrystalline solar panels.

Secondly, TF solar panels have the highest mean values of specific yield followed by HIT, polycrystalline and monocrystalline, which means that the TF solar panels are the most efficient for generating power in the FTKE area. Next, polycrystalline has the lowest standard deviation, followed by HIT, monocrystalline and TF solar panels. The low standard deviation value shows that the specific yield data clustered tightly around the mean value.

In conclusion, TF solar panels produce the highest specific yield compared to other types of solar panels. Therefore, for the purpose of developing a forecasting model, the next chapter will focus on the ARIMA model for TF solar panels.

4.3 Stationarity

The probability plot compares the data quantiles to the expected quantiles from a theoretical distribution. If the data points align closely with a straight line, it indicates that the dataset roughly follows the normal distribution. Otherwise, techniques such as differencing can be applied in time series analysis to make the data stationary. Differencing involves calculating the differences between consecutive observations in a time series. This method helps to eliminate trends or seasonality in the data, making it stationary.

4.3.1 Probability Plot

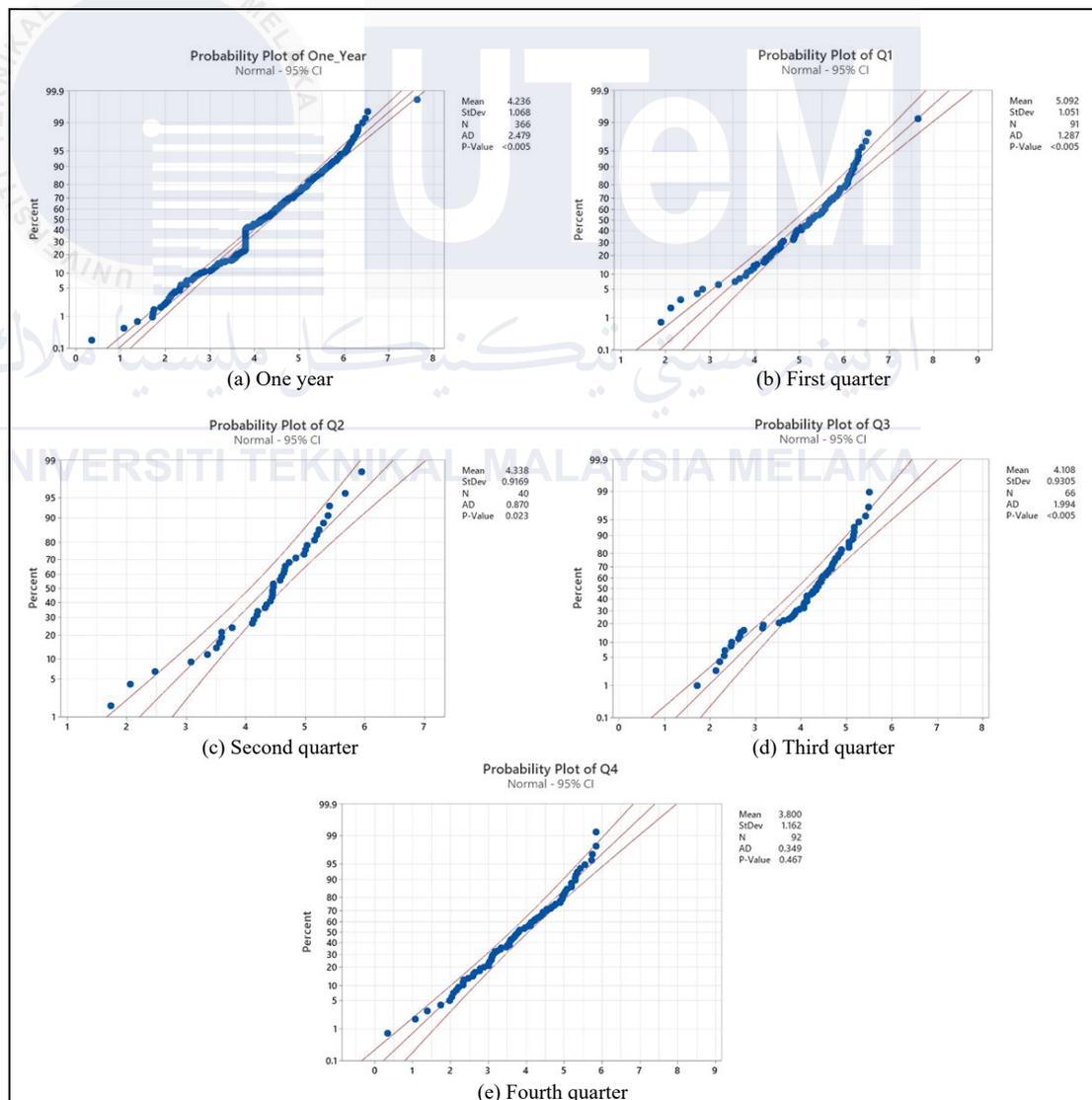


Figure 4.17: Probability plot of TF specific yield in (a) one year, (b) first quarter, (c) second quarter, (d) third quarter and (e) fourth quarter of 2016

Figure 4.17 (a) to Figure 4.17 (e) presents the probability plot for the specific yield of TF solar panel. The results indicate that the p-value for the fourth quarter of 2016 is 0.467, which is higher than the significance level ($\alpha = 0.05$), suggesting that the data follows a normal distribution. However, the p-values for the one year, first quarter, second quarter, and third quarter of 2016 are below the significance level, indicating that the data does not follow a normal distribution. Therefore, normalization or transformation methods should be applied to make the data follows normal distributed.

4.3.2 Box-Cox Transformation

The Box-Cox transformation transforms the data so that its distribution is as close to a normal distribution as possible. It is useful when dealing with data that violates the assumption of normality and constant variance. These diagnostic plots help assess the effectiveness of the transformation and verify if the transformed data meet the assumption of normality and constant variance.

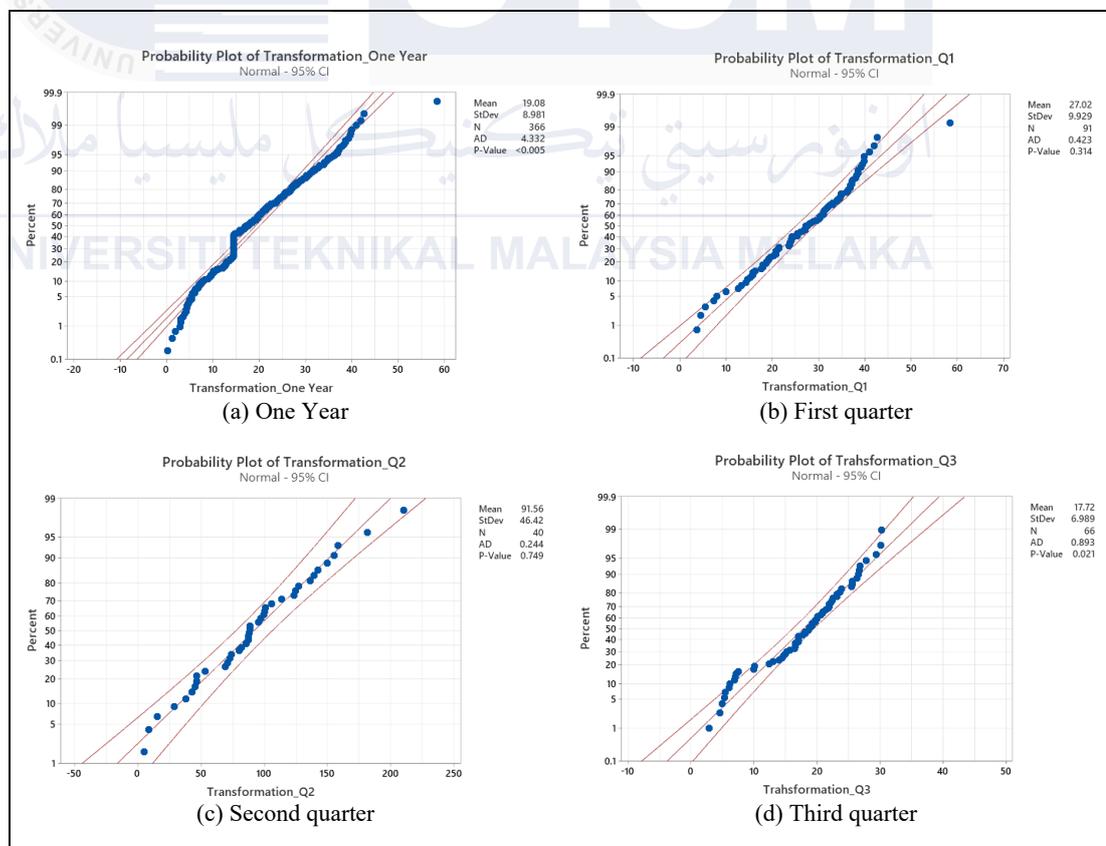


Figure 4.18: Probability plot after normalization of TF specific yield in (a) one year, (b) first quarter, (c) second quarter and (d) third quarter of 2016

Figure 4.18 (a) to Figure 4.18 (d) shows the probability plot after normalization using Box-Cox transformation. The results indicate that the p -values for the first and second quarters are above the significance level ($\alpha = 0.05$), suggesting that the Box-Cox transformation method successfully made this data normally distributed. However, the p -values for the one year and third quarter remain below the significance level, indicating that the transformation fail to make the data normally distributed.

4.3.3 Autocorrelation Factor (ACF)

Auto-correlation refers to the correlation between a time series and its own past values at different points in time. It quantifies how current values are related to previous values within the time series.

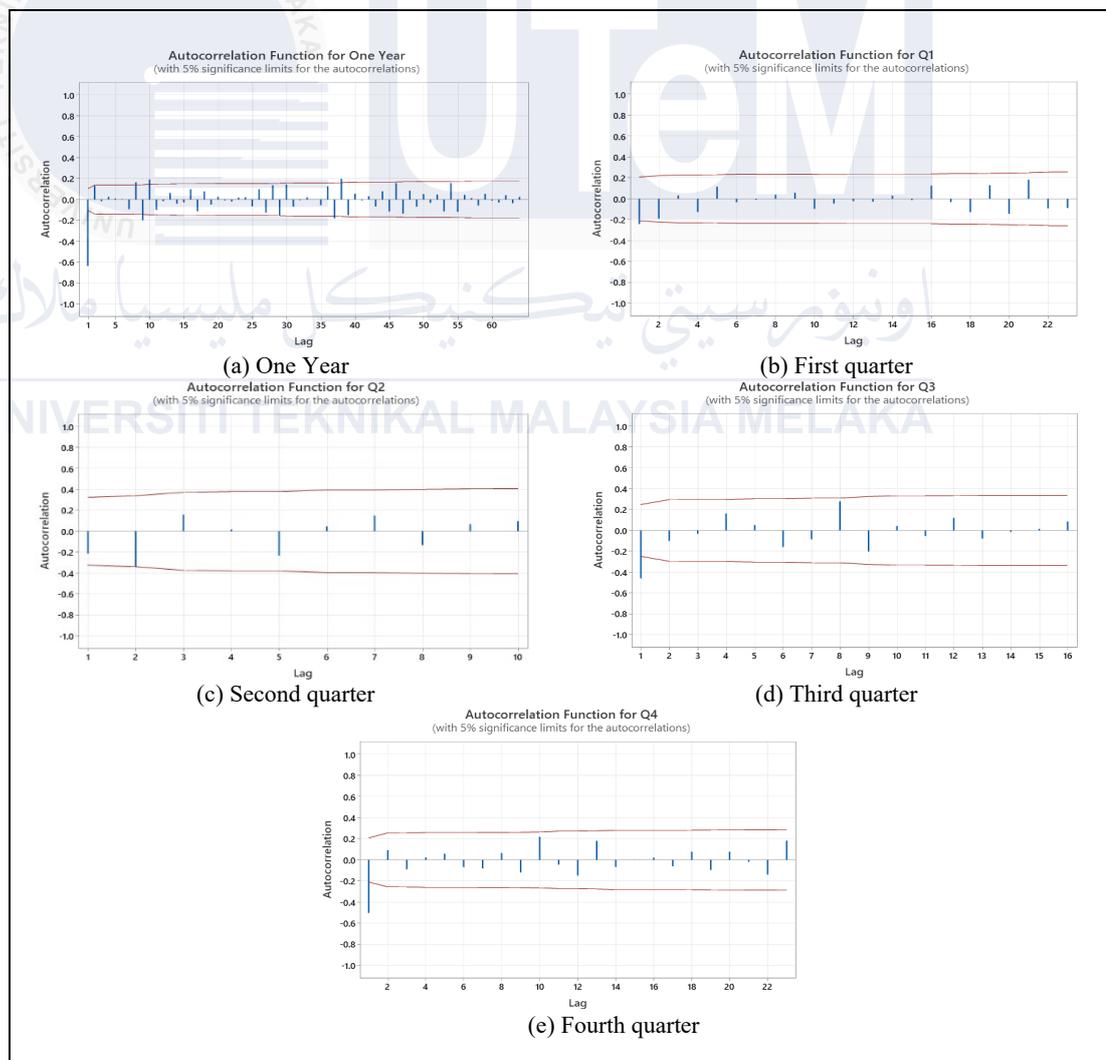


Figure 4.19: ACF plot for TF specific yield in (a) one year, (b) first quarter, (c) second quarter, (d) third quarter and (e) fourth quarter of 2016

Figure 4.19 (a) to Figure 4.19 (e) displays the ACF graph for the case study data, showing the correlation between the current observation and its lagged values. The red horizontal line marks the significance threshold. ACF values that exceed this line are considered non-stationary. In Figure 4.19, the second quarter of 2016 has all ACF values within the red horizontal line, illustrating the stationarity of the data. Meanwhile, for the one year, first, third, and fourth quarters of 2016, there is one ACF value that exceeds the red horizontal line.

4.3.4 Partial Autocorrelation Factor (PACF)

A PACF captures a direct correlation between time series and a lagged series of data.

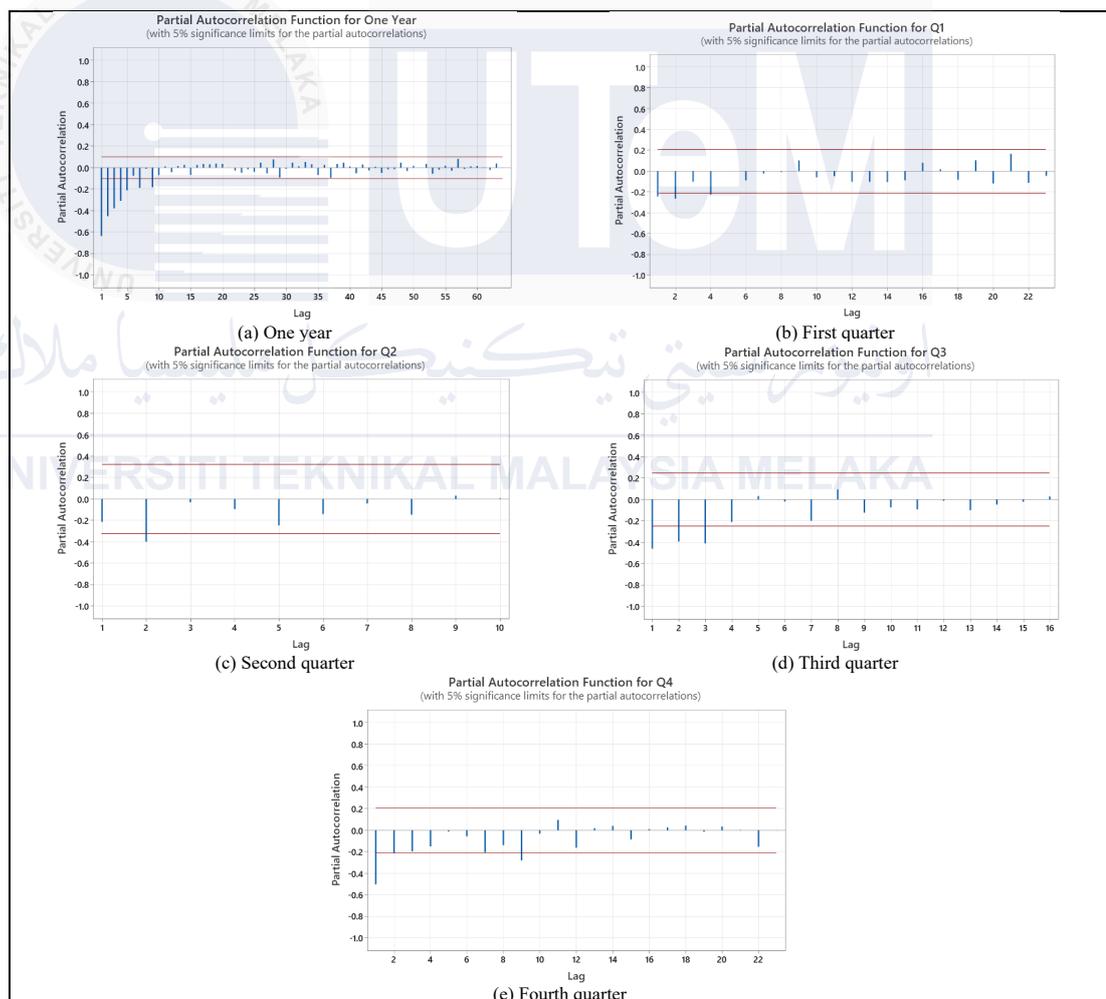


Figure 4.20: PACF plot for TF specific yield in (a) one year, (b) first quarter, (c) second quarter, (d) third quarter and (e) fourth quarter of 2016

Figure 4.20 (a) to Figure 4.20 (e) depicts the PACF plot, with a red horizontal line representing the threshold value of the PACF where any value above this line is deemed significant. From Figure 4.20, it is observed that the PACF plot for the one-year and the third quarter of 2016 exhibits multiple non-stationary values, suggesting potential non-stationarity in the data. Conversely, for the first, second, and fourth quarters, all PACF values fall within or have only one exceeding the significance threshold, indicating that the data can be considered as stationary.

4.4 Time Series Model Development

To determine whether the association between the response and each term in the model is statistically significant, a comparison of the p -value and the significance level was done to assess the null hypothesis. The null hypothesis is that the term is not significantly different from 0, which indicates that no association exists between the term and the response. Usually, the significance level used is 0.05.

The data of one year, the first quarter, second quarter, third quarter and the fourth quarter of 2016 were tested with 26 ARIMA models. The p -values of these models were observed to determine their significance.

For instance, when testing an ARIMA model (1,0,1) on the first quarter of 2016, the results reveal that the p -value for the AR 1 and MA 1 terms are 0.261 and 0.432, respectively. Both p -values exceed the significance level, suggesting that the model is not significant. Table 4.2 shows analysis of all non-significant ARIMA models tested for one year, first quarter, second quarter, third quarter, and fourth quarter of 2016.

Table 4.2 Non-significant ARIMA model for specific yield of TF solar panel

ARIMA model (p,d,q) for one year							
(2,0,2)	(2,2,1)	(1,2,1)	(1,1,2)	(0,2,2)	(2,1,1)	(2,1,2)	(2,2,0)
ARIMA model (p,d,q) for first quarter							
(0,2,2)	(1,0,1)	(2,1,2)	(1,1,2)	(1,2,2)	(2,0,0)	(2,0,2)	(2,2,2)
ARIMA model (p,d,q) for second quarter							
(0,0,2)	(2,1,1)	(0,2,2)	(1,0,0)	(1,0,1)	(1,0,2)	(1,1,0)	(1,1,1)
(2,1,2)	(1,1,2)	(1,2,1)	(1,2,2)	(2,0,0)	(2,0,2)	(2,2,1)	(1,2,2)
ARIMA model (p,d,q) for third quarter							
(0,0,1)	(0,0,2)	(0,1,2)	(2,1,1)	(1,0,0)	(1,0,1)	(1,0,2)	(1,1,1)
(2,1,2)	(1,1,2)	(1,2,2)	(2,0,0)	(2,0,1)	(2,0,2)		
ARIMA model (p,d,q) for fourth quarter							
(0,0,1)	(0,0,2)	(0,1,2)	(1,0,0)	(1,0,1)	(1,0,2)	(1,1,1)	(2,1,2)
(1,2,2)	(2,0,0)	(2,0,1)	(2,0,2)	(2,2,1)	(1,1,2)		

Conversely, when testing an ARIMA model (0,1,1) on the first quarter of 2016, the analysis reveals that the p-value for the MA 1 is 0 which is lower than the significance level, indicating the ARIMA model is significant. Table 4.3 shows analysis of all significant ARIMA models tested for one year, first quarter, second quarter, third quarter, and fourth quarter of 2016.

Table 4.3 Significant ARIMA model for case study data

ARIMA model (p,d,q) for one year							
(0,0,1)	(0,0,2)	(0,1,1)	(0,1,2)	(2,1,0)	(2,0,0)	(1,2,2)	(1,0,0)
(1,0,1)	(1,0,2)	(1,1,0)	(1,1,1)	(2,0,1)	(2,2,2)		
ARIMA model (p,d,q) for first quarter							
(0,1,1)	(2,2,0)	(0,0,1)	(2,1,1)	(1,0,2)	(1,1,1)	(2,0,1)	(2,1,0)
(1,2,0)	(0,1,2)	(1,0,0)	(1,1,0)	(1,2,1)	(2,2,1)		
ARIMA model (p,d,q) for second quarter							
(0,1,1)	(2,1,0)	(2,2,0)	(1,2,0)	(0,0,1)	(0,1,2)	(2,0,1)	
ARIMA model (p,d,q) for third quarter							
(0,1,1)	(2,1,0)	(2,2,0)	(1,2,0)	(0,2,2)	(1,1,0)	(1,2,1)	(2,2,1)
(2,2,2)							
ARIMA model (p,d,q) for fourth quarter							
(0,1,1)	(2,1,0)	(2,2,0)	(1,2,0)	(2,1,1)	(0,2,2)	(1,1,0)	(1,2,1)
(2,2,2)							

4.5 Error Measurement

Error measurement is the difference between a measured quantity and its actual value. It helps to understand the accuracy of a prediction or estimation. For this analysis, only significant models as mentioned in Table 4.3 are used for forecasting purposes. The MAE and MSE methods are used to measure the accuracy of the predicted models.

Table 4.4: MAE and MSE for ARIMA model for one year (2016)

Model (p,d,q)	MAE	MSE
(0,0,1)	0.825	0.680
(0,0,2)	0.797	0.636
(0,1,1)	0.866	0.750
(0,1,2)	0.905	0.819
(2,1,0)	0.792	0.628
(2,0,0)	0.758	0.574
(1,2,2)	0.882	0.778
(1,0,0)	0.805	0.647
(1,0,1)	0.795	0.632
(1,0,2)	0.853	0.727
(1,1,0)	0.884	0.781
(1,1,1)	0.912	0.831
(2,0,1)	0.845	0.714
(2,2,2)	0.806	0.649

The best ARIMA model for specific yield in 2016 is (2,0,0). This model has the lowest error compared to the other 13 significant ARIMA models. This analysis is supported by the lowest values of MSE and MAE as shown in Table 4.4. However, the MAE and MSE values are still considered high, above 0.5. This is because the data used for one entire year of 2016 are not stationary.

Table 4.5: MAE and MSE for ARIMA model for (Q1), (Q2), (Q3) and (Q4) of 2016

No	Model (p,d,q)	Q1		Q2		Q3		Q4	
		MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
1	(0,1,1)	0.381	0.145	0.530	0.281	1.119	1.253	0.316	0.100
2	(2,1,0)	0.735	0.540	0.659	0.434	1.186	1.406	0.452	0.204
3	(2,2,0)	1.063	1.129	0.790	0.625	1.260	1.588	0.554	0.306
4	(1,2,0)	0.634	0.402	0.886	0.784	1.100	1.210	2.586	6.690
5	(1,1,0)	0.721	0.520	-	-	0.894	0.799	0.205	0.042
6	(1,2,1)	0.422	0.178	-	-	0.870	0.756	0.195	0.038
7	(0,2,2)	-	-	-	-	1.077	1.160	0.322	0.104
8	(2,2,2)	-	-	-	-	1.246	1.553	0.276	0.076
9	(2,2,1)	0.295	0.087	-	-	1.831	3.354	-	-
10	(2,1,1)	0.715	0.511	-	-	-	-	0.377	0.142
11	(0,0,1)	0.210	0.044	0.626	0.392	-	-	-	-
12	(2,0,1)	0.265	0.070	0.419	0.176	-	-	-	-
13	(0,1,2)	0.385	0.148	0.505	0.255	-	-	-	-
14	(1,0,2)	0.255	0.065	-	-	-	-	-	-
15	(1,1,1)	0.473	0.223	-	-	-	-	-	-
16	(1,0,0)	0.256	0.066	-	-	-	-	-	-

The ARIMA model (0,0,1) is the most accurate model for forecasting the specific yield of the TF solar panel in FTKE for the first quarter of 2016. This model exhibits the lowest error among the 13 other significant models, as indicated by the values of MAE and MSE presented in Table 4.5. For the second quarter of 2016, the best ARIMA model for forecasting specific yield of TF solar panel in FTKE is (2,0,1). This analysis is supported by the lowest values of MAE and MSE as shown in Table 4.5. The most precise ARIMA model for forecasting the specific yield in the third quarter of 2016 is (1,2,1). This model stands out among the other eight significant ARIMA models tested due to its lowest MAE and MSE value, as shown in Table 4.5. However, the MAE and MSE values are still considered high, above 0.5. This is because the data used for the year 2016 are not stationary. The fourth quarter of 2016

for ARIMA (1,2,1) stands out as the most precise choice for forecasting specific yield of the TF solar panel in FTKE. Among the other 8 significant models, the smallest margin of error as shown the value of MAE and MSE as presented in Table 4.5.

In conclusion, the best ARIMA models are:

1 year data: ARIMA (2,0,0)

1st quarter: ARIMA (0,0,1)

2nd quarter: ARIMA (2,0,1)

3rd quarter: ARIMA (1,2,1)

4th quarter: ARIMA (1,2,1)

4.6 Forecasting Future Values by using ARIMA Model

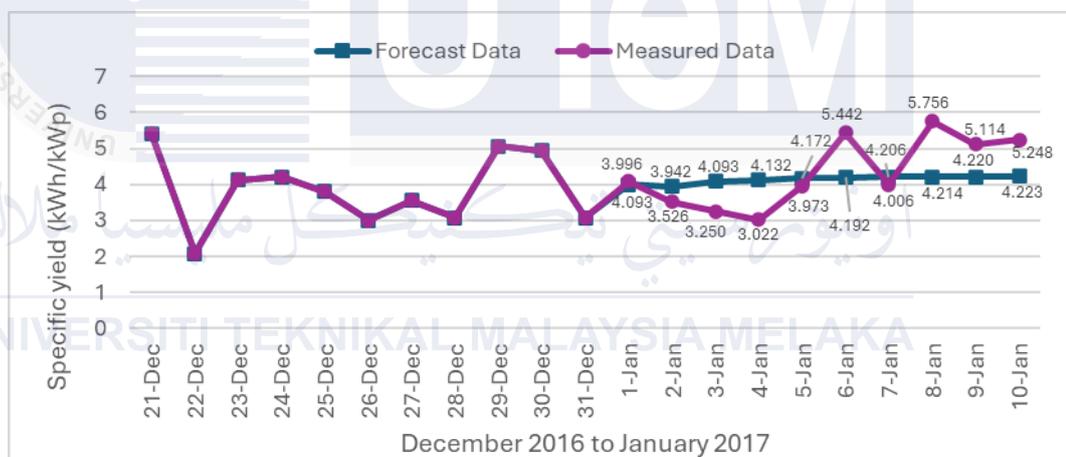


Figure 4.21: Forecast trend for one whole year of 2016 for TF solar panel

Figure 4.21 displays the data trend of the forecasted specific yield using the ARIMA model (2,0,0). The model was developed using data from 1st January 2016 to 31st December 2016. With this timeframe, a specific yield forecast can be made from 1st January 2017 to 10th January 2017. The forecasted data trend indicates that the specific yield slightly increases from 4.093 kWh/kWp on the 1st January to 4.223 kWh/kWp on the 10th January 2017. This differs from the measured data which shows significant fluctuations during that timeframe. The forecast data does not follow the trendline of the measured data because the data used to develop the forecast model is not stationary.

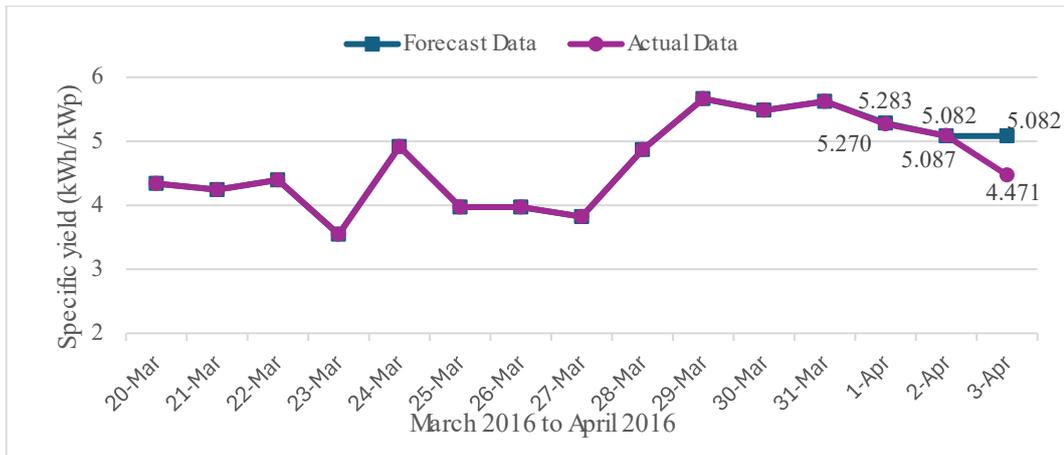


Figure 4.22: Forecast trend for the first quarter of 2016 for TF solar panel

Figure 4.22 exhibits the data trend of forecast specific yield of the ARIMA (0,0,1). The model is created based on data spanning from 1st January 2016 to 31st March 2016. This timeframe enables the forecast of specific yield from 1st April to 3rd April 2016. The forecast follows the trendline of the measured data as there is only a slight difference between the forecast and measured data.

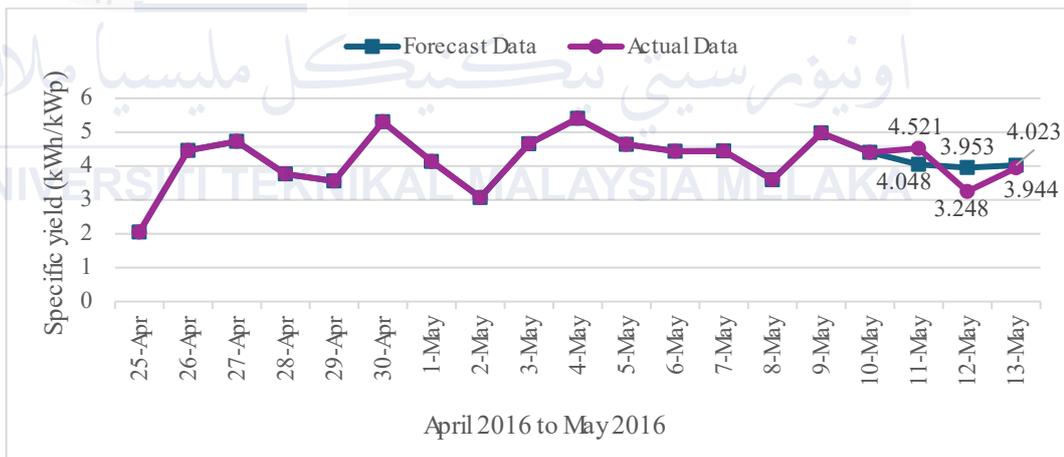


Figure 4.23: Forecast trend for the second quarter of 2016 for TF solar panel

Figure 4.23 illustrates the data trend in forecast specific yield data produced by the ARIMA (2,0,1). The model was developed using data ranging from 1st April to 10th May 2016. This time period allows for the prediction of specific yield values from 11th May 2016 to 13th May 2016. The forecast data trend indicates a decrease in value from 11th May (4.048 kWh/kWp) to 12th May (3.953 kWh/kWp) and then increase on 13th May (4.023 kWh/kWp).

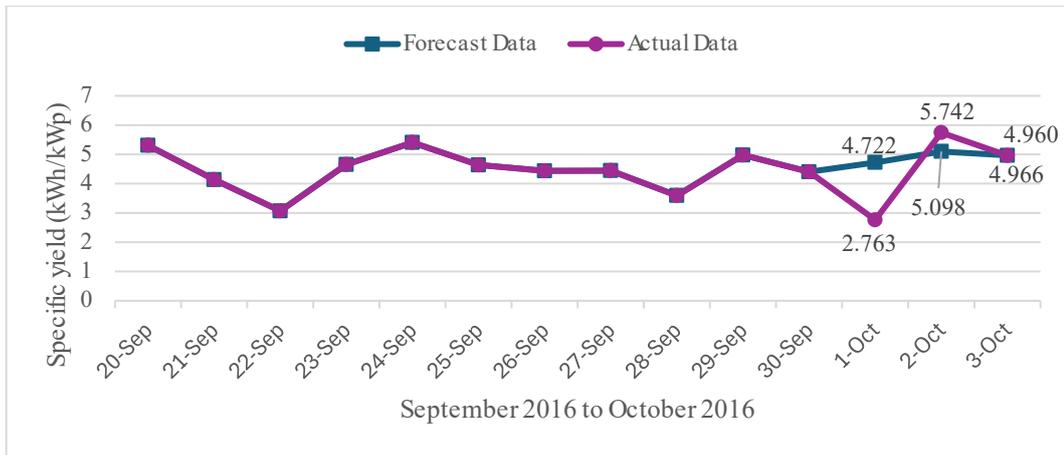


Figure 4.24: Forecast trend for the third quarter of 2016 for TF solar panel

Figure 4.24 illustrates the trend of forecasted specific yield using the ARIMA (1,2,1). This model was built based on data collected from 1st 2016 to 30th September 2016. Utilizing this timeframe, the model can be forecast from 1st October to 3rd October 2016. The forecast data shows a major difference between actual and forecast data on the 1st October 2016 to 2nd October 2016. This is because the data used to build the ARIMA model is not stationary.

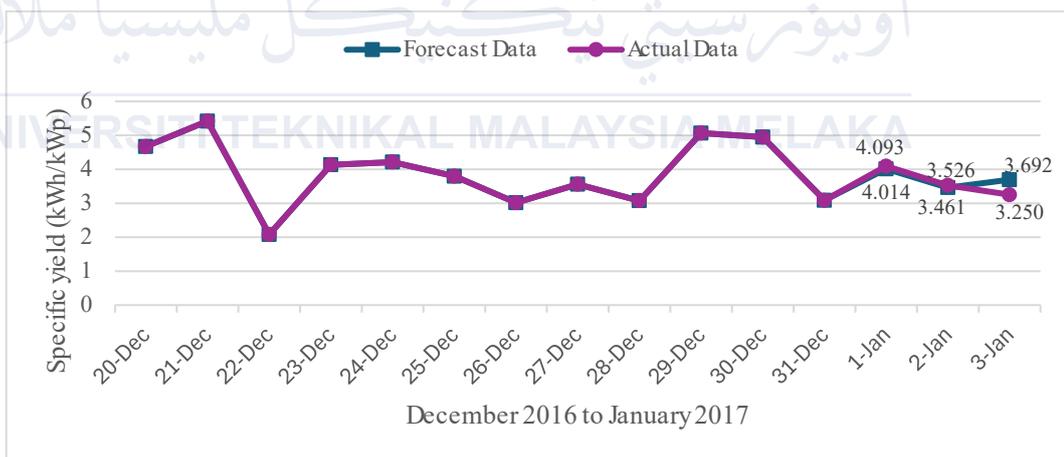


Figure 4.25: Forecast data for the fourth quarter of 2016 for TF solar panel

Figure 4.25 displays the data trend of forecast specific yield using the ARIMA model (1,2,1). The model was developed using data from 1st October to 31st December 2016. With this timeframe, a specific yield forecast can be made from 1st January 2017 to 3rd January 2017. The forecast follows the trendline of the measured data as there is only a slight difference between the forecast and measured data.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

Based on the time series analysis, it has been determined that the Thin-Film solar panel exhibits the highest specific yield for solar radiation, suggesting it as the ideal solar panel for power generation in the FTKE area. Subsequently, a forecasting model for specific yield utilizing ARIMA models was developed by using the Minitab software. Five ARIMA models were developed: (2,0,0), (0,0,1), (2,0,1), (1,2,1), and (1,2,1), each applied to five distinct timeframes; the entire year of 2016, the first, second, third, and fourth quarters of 2016. The accuracy of these models was evaluated, revealing highly precise results with MAE values are 0.758, 0.210, 0.419, 0.870, and 0.195, along with MSE values are 0.574, 0.044, 0.176, 0.756, and 0.038 for the same respective timeframes. Finally, this analysis offers valuable insights into the trend analysis of specific yield, enhancing comprehension of its patterns and trends in the area of FTKE, UTeM area. The ARIMA model and its error measurement provides crucial information for decision-making and planning concerning the utilization and management of solar energy at the facility.

5.2 Future Works

In future studies, researchers can enhance specific yield forecasting models at FTKE, UTeM, by considering meteorological factors like temperature, humidity, and cloud cover. By including these factors, the models can better predict how solar radiation changes based on weather conditions. Also, extending the forecasting period to months or years can help assess long-term solar energy planning and decision-making at FTKE. This would give useful information for improving how resources are used and managed. By making these improvements, researchers can help increase the accuracy of forecasting specific yield in FTKE, UTeM.

REFERENCES

- [1] R. Alles, “Concentrated Solar Power (CSP) Vs Photovoltaic (PV): An In-depth Comparison,” *SF Magazine*, Oct. 05, 2021. <https://www.solarfeeds.com/mag/csp-and-pv-differences-comparison/>
- [2] K. J. Iheanetu, “Solar Photovoltaic Power Forecasting: A Review,” *Sustainability*, vol. 14, no. 24, p. 17005, Dec. 2022, doi: <https://doi.org/10.3390/su142417005>
- [3] “Photovoltaic Effect,” *VEDANTU*. <https://www.vedantu.com/physics/photovoltaic-effect>
- [4] “Photovoltaic - Harvesting the Power of the Sun,” *Green Sarawak*, Nov. 10, 2017. <https://greensarawak.com/photovoltaic-harvesting-the-power-of-the-sun/>
- [5] A. Bakhtiari, “What is Heterojunction Solar Panels - AE Solar,” *AESOLAR*, Nov. 24, 2021. <https://old.ae-solar.com/heterojunction-modules/> (accessed Jan. 10, 2024).
- [6] “Solar panel temperature coefficients | Greenwood,” *www.greenwoodsolutions.com.au*. <https://www.greenwoodsolutions.com.au/news-posts/solar-panel-temperature-coefficients>
- [7] “Solar Panel Components (List And Functions) - 2023,” *Sol Voltaics*, May 08, 2022. <https://solvoltaics.com/solar-panel-components/>
- [8] F. K. E. UTeM. (FKE), “Solar PV & Smart Grid Research Laboratory (SSG).” <https://fke.utem.edu.my/eps/index.php/facilities/solar-pv-smart-grid-ssg#solar612>
- [9] Environmental and Energy Study Institute, “Fossil Fuels,” *www.eesi.org*, Jul. 22, 2021. <https://www.eesi.org/topics/fossilfuels/description#:~:text=Fossil%20fuels%E2%80%94including%20coal%2C%20oil>
- [10] “Is Going Solar a Good Idea in Malaysia?,” *www.buysolar.my*. <https://www.buysolar.my/resources/articles/is-going-solar-a-good-idea-in-malaysia#:~:text=Thanks%20to%20Malaysia>
- [11] “Renewable Energy Malaysia – Renewable Energy Malaysia,” *SEDA Malaysia*, 2023. <https://www.seda.gov.my/reportal/>
- [12] Y. D. Arthur, “Probability Distributional Analysis of Hourly Solar Irradiation in Kumasi-Ghana,” *Int. J. Bus. Soc. Res.*, no. 1964, pp. 63–75, 1988, doi: 10.18533/ijbsr.v3i3.57.

- [13] S. Dukkipati, V. Sankar, and P. S. Varma, "Forecasting of solar irradiance using probability distributions for a PV System: A case study," *Int. J. Renew. Energy Res.*, vol. 9, no. 2, pp. 741–748, 2019, doi: 10.20508/ijrer.v9i2.9216.g7643
- [14] M. Rouse, "What is an Artificial Neural Network (ANN)? - Definition from Techopedia," *Techopedia.com*, 2019.
<https://www.techopedia.com/definition/5967/artificial-neural-network-ann>
- [15] N. Premalatha and A. Valan Arasu, "Prediction of solar radiation for solar systems by using ANN models with different back propagation algorithms," *Journal of Applied Research and Technology*, vol. 14, no. 3, pp. 206–214, Jun. 2016, doi: <https://doi.org/10.1016/j.jart.2016.05.001>.
- [16] J. Thaker, R. Höller, and M. Kapasi, "Short-Term Solar Irradiance Prediction with a Hybrid Ensemble Model using EUMETSAT Satellite Images," *Short-Term Solar Irradiance Prediction with a Hybrid Ensemble Model Using EUMETSAT Satellite Images*, Nov. 2023, doi: <https://doi.org/10.20944/preprints202311.1538.v1>
- [17] Tableau, "Time Series Analysis: Definition, Types, Techniques, and When It's Used," *Tableau*, 2022. <https://www.tableau.com/learn/articles/time-series-analysis>
- [18] B. Singh and D. Pozo, "A Guide to Solar Power Forecasting using ARMA Models," *Proc. 2019 IEEE PES Innov. Smart Grid Technol. Eur. ISGT-Europe 2019*, pp. 0–5, 2019, doi: 10.1109/ISGTEurope.2019.8905430
- [19] O. Gerasymov, "Using Machine Learning for Time Series Forecasting Project," *CodeIT*, Jan. 04, 2023. <https://codeit.us/blog/machine-learning-time-series-forecasting>
- [20] M. Pourahmadi, "A Course in Time Series Analysis," *Am. Stat.*, vol. 56, no. 1, pp. 77–77, 2002, doi: 10.1198/tas.2002.s131.
- [21] A. Ramesh, "Time Series Analysis: Definition, How it Works, Purpose & Uses," *www.strike.money*. <https://www.strike.money/technical-analysis/time-series-analysis> (accessed Dec. 12, 2023).
- [22] G. Scott, "Box-Jenkins Model," *Investopedia*, 2019.
<https://www.investopedia.com/terms/b/box-jenkins-model.asp>

- [23] A. Dotis-Georgiou, "Time series forecasting with ARMA and InfluxDB," InfoWorld, Dec. 15, 2022. <https://www.infoworld.com/article/3681894/time-series-forecasting-with-arma-and-influxdb.html#:~:text=An%20ARMA%20or%20autoregressive%20moving> (accessed Dec. 12, 2023).
- [24] Jason Brownlee, "How to Create an ARIMA Model for Time Series Forecasting in Python," Machine Learning Mastery, Jan. 08, 2017. <https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/>
- [25] L. Cheng, H. Zang, and G. Sun, Multi-Meteorological-Factor-Based Graph Modeling for Photovoltaic Power Forecasting, vol. 12, no. 3, pp. 1593–1603, Feb. 2021, doi: <https://doi.org/10.1109/tste.2021.3057521>.
- [26] K. Zhang and G. Zou, "Photovoltaic Output Prediction Method Based on Weather Forecast and Machine Learning," Journal of Physics: Conference Series, vol. 2320, no. 1, p. 012032, Aug. 2022, doi: <https://doi.org/10.1088/1742-6596/2320/1/012032>.
- [27] A. Al-Khazzar, Behavior of Four Solar PV Modules with Temperature Variation, vol. 6, pp. 1091–1099, Nov. 2016.
- [28] N. Son and M. Jung, "Analysis of Meteorological Factor Multivariate Models for Medium- and Long-Term Photovoltaic Solar Power Forecasting Using Long Short-Term Memory," Applied Sciences, vol. 11, no. 1, p. 316, Dec. 2020, doi: <https://doi.org/10.3390/app11010316>.
- [29] A. Mbaye et al., "ARMA model for short-term forecasting of solar potential: application to a horizontal surface of Dakar site," OAJ Mater. Devices, vol. 4, no. 1, 2019, doi: 10.23647/ca.md20191103
- [30] S. Atique, S. Noureen, V. Roy, V. Subburaj, S. Bayne, and J. MacFie, "Forecasting of total daily solar energy generation using ARIMA: A case study," 2019 IEEE 9th Annu. Comput. Commun. Work. Conf. CCWC 2019, no. January, pp. 114–119, 2019, doi: 10.1109/CCWC.2019.8666481.

- [31] “What is a specific yield in a solar plant?,” *Quora*, 2019. <https://www.quora.com/What-is-a-specific-yield-in-a-solar-plant#:~:text=Specific%20yield%20in%20a%20solar%20plant%20refers%20to%20the%20amount>
- [32] A. Hayes, “Descriptive Statistics: Definition, Overview, Types, Example,” *Investopedia*, Mar. 21, 2023. https://www.investopedia.com/terms/d/descriptive_statistics.asp
- [33] National Library of Medicine, “Finding and Using Health Statistics,” *www.nlm.nih.gov*, 2012. [https://www.nlm.nih.gov/oet/ed/stats/02-900.html#:~:text=A%20standard%20deviation%20\(or%20%CF%83](https://www.nlm.nih.gov/oet/ed/stats/02-900.html#:~:text=A%20standard%20deviation%20(or%20%CF%83)
- [34] R. Santra, “Stationarity in Time Series,” *Medium*, May 12, 2023. <https://medium.com/@ritusantra/stationarity-in-time-series-887eb42f62a9>
- [35] Indeed, “What is the autocorrelation function and how do you use it?,” May 25, 2023. <https://uk.indeed.com/career-advice/career-development/autocorrelation-function>
- [36] “RPubs - ACF PACF,” *rpubs.com*. <https://rpubs.com/yqzrp09/acf-pacf> (accessed Jan. 10, 2024).
- [37] “About: Partial autocorrelation function,” *dbpedia.org*. https://dbpedia.org/page/Partial_autocorrelation_function (accessed Jan. 10, 2024).
- [38] D. Andrés, “Error Metrics for Time Series Forecasting - ML Pills,” *Machine Learning Pills*, Jun. 22, 2023. [https://mlpills.dev/time-series/error-metrics-for-time-series-forecasting/#:~:text=Mean%20Squared%20Error%20\(MSE\)%20is](https://mlpills.dev/time-series/error-metrics-for-time-series-forecasting/#:~:text=Mean%20Squared%20Error%20(MSE)%20is) (accessed Jan. 10, 2024).
- [39] “What is Mean Absolute Error,” *Deepchecks*. <https://deepchecks.com/glossary/mean-absolute-error> (accessed Jan. 10, 2024).
- [40] S. Glen, “Mean Squared Error: Definition and Example,” *Statistics How To*. <https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/mean-squared-error/>
- [41] Steffen Thorsen, “Weather in December 2016 in Malacca City, Malacca, Malaysia,” *www.timeanddate.com*, 2016. <https://www.timeanddate.com/weather/malaysia/malacca-city/historic?month=12&year=2016> (accessed Jan. 18, 2024).