INFERENCE METHOD FOR INVERTER SIZING RATIO FOR PV PLANT USING SATELLITE-DERIVED SOLAR IRRADIANCE DATA.

TAN HUI LI

A project report submitted in partial fulfilment of the requirements for the award of Bachelor of Engineering (Hons.) Electrical and Electronic Engineering

Lee Kong Chian Faculty of Engineering and Science
Universiti Tunku Abdul Rahman

April 2022

DECLARATION

I hereby declare that this project report is based on my original work except for citations and quotations which have been duly acknowledged. I also declare that it has not been previously and concurrently submitted for any other degree or award at UTAR or other institutions.

Signature :

Name : Tan Hui Li

ID No. : 980623-01-6292

Date : 10/5/2022

APPROVAL FOR SUBMISSION

I certify that this project report entitled **"INFERENCE METHOD FOR INVERTER SIZING RATIO FOR PV PLANT USING SATELLITE-DERIVED SOLAR IRRADIANCE DATA"** was prepared by **TAN HUI LI** has met the required standard for submission in partial fulfilment of the requirements for the award of Bachelor of Engineering (Hons.) Electrical and Electronic at Universiti Tunku Abdul Rahman.

Approved by,

Signature : LIM B. H.

Supervisor : Ir Dr. Lim Boon Han

Date : 12 May 2022

Signature :

Co-Supervisor : Dr. Lai An Chow

Date : 12 May 2022

ABSTRACT

Inverter Sizing Ratio (ISR) is the ratio of the installed DC capacity of a photovoltaic (PV) power plant to the AC output power rating of an inverter. It is important to choose an optimal ISR to design an economical design of a utility-scale PV power plant. The ground-based measurement solar irradiance datasets can have higher temporal resolution, which can be made up to 1-min intervals, whereas the satellite-derived datasets are only available for a lower temporal resolution such as 15 minutes to 1-hour intervals. Therefore, an optimal ISR computed by satellite-derived irradiance datasets might be less reliable as compared to the optimal ISR computed by ground-based datasets.

In executing the first objective of this project, ISR of a 10MW PV plant was computer-simulated, and the optimal ISR value was chosen based on the lowest Levelized Cost of Energy (LCOE) for a project lifespan. Both ground-based measurement and satellite-derived irradiance datasets were used to simulate the optimal ISR. The optimal ISR values were compared and to the relationship was investigate the relationship between them. It is found that different types of solar irradiance sensors can cause different solar irradiance distribution profile and thus impact on the optimal ISR value. The result shows that the irradiance sensor with low accuracy will contribute to the overestimation of optimal ISR.

The second objective is to infer the optimal ISR to the other sites using satellite-derived datasets. The inference was done by using an Artificial Neural Network (ANN) model since the relationship between optimal ISR simulated by ground-based and satellite-derived datasets is complicated to be derived. The percentage of difference and errors between optimal ISR simulated by ground-based and satellite-derived datasets is reduced after the inferencing of optimal ISR using the ANN model developed. The percentage of difference is reduced to 5.8%, Mean Squared Error is 0.0128 and Root Mean Squared Error is 0.1132.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS / ABBREVIATIONS

AC	Alternating current
ANN	Artificial Neural Network
D	The period of time for the discrete value of the output power, hours
DC	Direct current
$E_{AC\_per\ day,act}$	AC electricity yield per day of a PV system with consideration of clipping power during event of high irradiance, kWh
E_{AC_sum}	Total electricity generated by PV system within a specific time frame
GHI	Global Horizontal Irradiance
GTI	Global Tilted Irradiance
ISR	Inverter sizing ratio
LCOE	Levelized cost of electricity
MLP	Multi-layer Perceptron
MSE	Mean squared error
O & M	Operation and maintenance
P_{AC_exp}	The expected instantaneous AC output power without consider clipping of power, MW
P_{AC_max}	The maximum AC power rating of inverter, MW
P_{AC_rated}	The AC power rating of inverter, MW
$P_{actual}(t)$	The actual instantaneous AC power output of inverter at that time interval, MW
PR_{fix}	Fixed performance ratio
PV	Photovoltaic
PVGIS	Photovoltaic Geographical Information System
PR	Performance ratio
t	interval of time, hours

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 General Introduction

Inverter is the significant element which help to convert the direct current (DC) electricity produced from a photovoltaic (PV) plant to alternating current (AC) electricity which can be delivered to the grid. When designing the PV plant system, it is a must to take extra consideration on how much DC power can be produced from the PV plant and how much AC output power can be achieved, which lead to the topic discussed in this study, the Inverter Sizing Ratio (ISR) or can be called as DC-AC Ratio.

ISR is the ratio of installed DC capacity of PV plant to the AC output power rating of an inverter. Oversizing of solar PV array where the ISR is greater than 1 is a common practice in this industry. It is common that the ISR of inverter is in between 1.1 to 1.2 for high solar irradiation area and 1.3 to 1.4 for low solar irradiation area. (Zidane, 2020) Oversizing of PV array is because usually the PV panels do not perform at their peak output due to environmental influences. For example, a 5 kW solar panel array, the average amount of generated power could be on average much closer to 4 kW only, 1 kW less than what the inverter can actually handle. Thus, it is common that the ISR of inverter is normally set to be greater than 1. So, the average efficiency of inverter can be increased. (Solar Power Co., 2019) For average output power, the PV output power could not achieve the inverter capacity but for instantaneous power, during event of high solar irradiance, the output can exceed the inverter rated capacity if the inverter is downsized, thus clipping of output power will happen.

In Figure 1.1, the clipping of output power due to inverter rating is described. The oversize of an inverter can reduce the generation of electricity being clipped, but the investors have to invest higher initial cost to purchase a higher rating inverter which might reduce their profit in a long run. Contrary, if the inverter is downsized, it will clip the production of electricity power during the event of high solar irradiance and result the same profit loss circumstance happened to the investor. This impact will be more vivid if applying a low power rating inverter at the site which have potential to harvest more solar energy yield.

The cost of an inverter is usually derived as levelized cost of electricity (LCOE), and the LCOE stated is proportional to the cost to build the PV power plant which is usually expressed in dollars per watt. Thus, choosing an optimal ISR is important to balance between the reduction of the capital cost of the PV plant and the loss of profit due to the clipping of electricity generation in the event of high solar irradiance. However, the analysis of yield of a PV plant needs at least past 10 years of solar irradiance datasets to investigate the climate around the selected PV plant area. Ground-based datasets have higher accuracy to estimate the inverter sizing ratio due to its high-resolution irradiance data, but it must be equipped with a physical weather station which must be around the area that is selected to construct the PV plant.

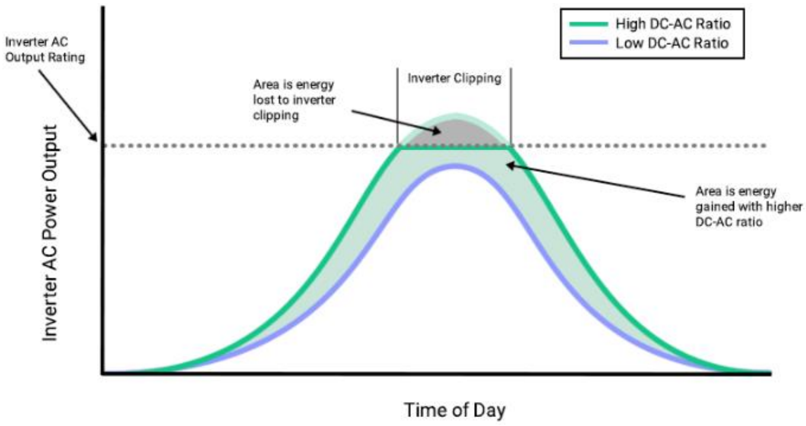

Figure 1.1 The inverter AC output over 24 hours of a day for systems with low and high DC-to-AC ratio. (David, 2021)

Another important point to be mentioned is that the tropical climate has more variation on the weather events around the tropic areas. For example, it might have two different types of weather in one city. Thus, the build of weather station must be reasonable near to the selected area in order to get the correct weather information. However, this is impossible and worthless to intentionally build and maintain that much of weather station around the country just to find the most suitable location to build the PV plant in order to get the maximised performance and profit. Thus, the advantage of satellite-derived datasets is highlighted due to this reason, it can obtain the data for any location over large range of geographical areas but in lower time resolution as compared to the ground-based measurement datasets. The coverage area from different online source of satellite-derived solar irradiance datasets is shown in Figure 1.2. This characteristic of satellite-derived datasets has overcome the limitation of coverage of geographical area in ground-based datasets, which help to useful information to estimate the yield of PV plant for the solar investor, the decision maker to more understand about the project they invested. In result, increase in number of solar investors help to promote the adoption of solar energy in tropical area to slow down the progress of climate change around the mother earth.

However, the accuracy of the value of optimal ISR simulated from the satellite-derived datasets might not that high as compared to the optimal ISR value calculated by the ground-based solar irradiance datasets. Based on the research done by Väisänen, et al. (2019), the paper states that using solar irradiance with higher temporal resolution is beneficial to finding the optimal ISR. Low temporal resolution datasets can cause some solar irradiance data that is highly fluctuated unseen. For example, a 1-hour interval satellite-derived solar irradiance datasets might miss the peak irradiance value during that time interval and affect the optimal ISR outcome. Thus, this project is carried out to overcome this disadvantage of satellite-derived solar irradiance datasets and make the optimal ISR simulated by satellite-derived datasets become closer to the optimal ISR obtained from the ground-based datasets.

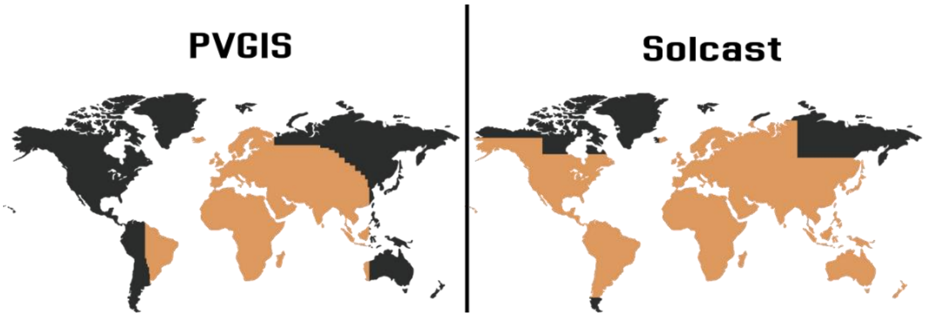

Figure 1.2 PVGIS and Solcast Solar Irradiance Coverage. (Solcast, n.d.)

1.2 Problem Statement

The ground-based measurement can have higher temporal resolution of solar irradiance datasets which can be made up to 1-minute interval. However, for the satellite-derived datasets, they are in lower time resolution such as 15 minutes to 1-hour interval, and the value are averaged within that particular time interval or sampled at every given time interval, these cause some irradiance data that is highly fluctuated unseen. In these recent years, there are few studies about the comparison of the solar irradiance collect from ground-based measurement and satellite-derived database has been carried out but lack of study about relationship between optimal ISR simulated by ground-based and satellite-derived solar irradiance datasets. This is because the relationship between optimal ISR simulated by ground-based and satellite-derived datasets is complex and the effect of parameters used in calculation of optimal ISR is unpredictable yet. Thus, in this project, an ANN model will be developed to infer the optimal ISR using satellite-derived datasets.

In addition, the loss of solar irradiance peak values in satellite-derived datasets need to be further investigated and studied to look for solution to overcome this disadvantage since it is not confirmed yet if the change of resolution of solar irradiance with different time interval could affect the optimal ISR significantly.

1.3 Importance and Contribution of the Study

The final expected outcome of this study is to infer the optimal ISR using satellite-derived datasets to the interested sites which is not equipped weather station or any solar irradiance sensor. This inference allowed advantage of satellite-derived datasets which have large coverage of geographical area being utilized.

Moreover, the analysis of relationship between optimal ISR simulated from ground-based and satellite-derived datasets is not common in tropical region, particularly in Malaysia. This is very important topic to help the development of PV system in tropics which can provide a reference range of optimal ISR value of a PV plant in the tropics, where the weather station with long-term collection of irradiance data is not yet available or ready in sufficient amount in most of the potential country in tropics. If the analysis which can help to compensate the lack of compliances of optimal ISR from satellite-derived datasets in this study, it will help to overcome the insufficient weather station problem and give a boost for development of PV industry in tropics.

Tropical region is the region which is suitable for large scale solar farm due to large geographical flat plain area and nearly constant daytime period along the year. Choosing a right size of inverter is important to achieve a cost-efficient PV system design, which help the investors to have shorter payback period. Thus, more investors are willing to participate in large-scale solar farm and this can help the government to promote the renewable energy indirectly. In addition, it is predicted that fossil fuels from mother earth will run out in 2060 if people keep burning it at current rate. Thus, this study helps to promote the adoption of solar energy which able to slow down this burning rate.

1.4 Aims and Objectives

The objectives in this study are:

(i) To investigate the relationship of optimal inverter sizing ratio simulated by using ground-based solar irradiance meter and satellite-derived data at co-located sites.

(ii) To infer the optimal inverter sizing ratio to other sites using the satellite-derived data.

1.5 Scope and Limitation of the Study

The most significant limitation in this study is that the sites to collect ground-based datasets are all from Malaysia area, which mean the outcome of study might be only suitable for Malaysia climate and not applicable for other tropical regions such as Indonesia, Philippines, and Thailand.

Moreover, in optimal ISR calculation, there are some terms such as Ross Coefficient, the degradation rate of PV panel etc. are constant values. These terms take a significant role which will affect the LCOE value, where the LCOE value can decide the value of optimal ISR. However, these terms might be not a constant value in real situation during operation of PV system. For example, the degradation rate of PV might increase due to high ambient temperature and humidity whereas the Ross Coefficient can be different for different type of PV panel model and installation method. All these terms might affect the value of optimal ISR for PV plant located at different sites, especially if inference is applied to obtain the optimal ISR for other sites which might have different climate with the sites having ground-based measurement in this study.

On the other hand, in the process of analysing of optimal ISR, capital of system installation and operation and maintenance (O&M) fee in this project are fixed value, which mean interest of loan, incentive, inflation rate and any other risk payment are not included in the calculation, which might cause the LCOE reduction and affect the estimated optimal ISR value.

1.6 Outline of the report

Chapter 1: Introduction

The content in this chapter will briefly introduce the background of the topic, issues arisen in Photovoltaic industry, the importance and contribution of study and the objectives of the study.

Chapter 2: Literature Review

This chapter discusses the current research done by other researcher and the technology which is related to the topic discussed.

Chapter 3: Methodology and Work Plan

The flow of the project and the methods applied will be presented in this chapter.

Chapter 4: Results and Discussion

The results are showed in form of graphs and tables in this chapter. Then the results will be analysed and discussed.

Chapter 5: Conclusion and Recommendations.

A summary of overall project will be discussed and some recommendations related to this topic can be conducted in future are presented.

CHAPTER 2

LITERATURE REVIEW

2.1 Solar Irradiance in Tropical Area

In Figure 2.1, it can be observed that the Global Horizontal Irradiance (GHI) value in tropical region is fall within the range of 1534 kWh/m^2 to 1826 kWh/m^2, which is able to harvest a great yield of renewable solar energy. Figure 2.2 has further show the potential Photovoltaic power output in Malaysia which is also showing the potential capability of harvesting high yield of solar energy in Malaysia. Not forgot to mention that for the countries fall on or near to the equator, the nearly constant daylight period along the year is actually more suitable for the operation of On-grid PV plant.

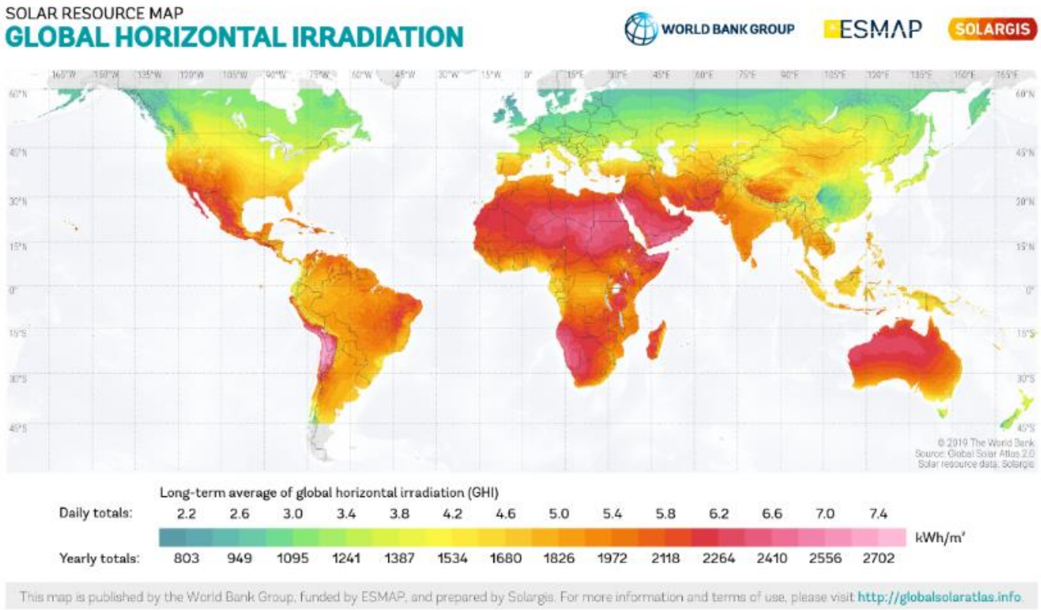

Figure 2.1 World solar resource map with GHI value. (Solargis, 2021)

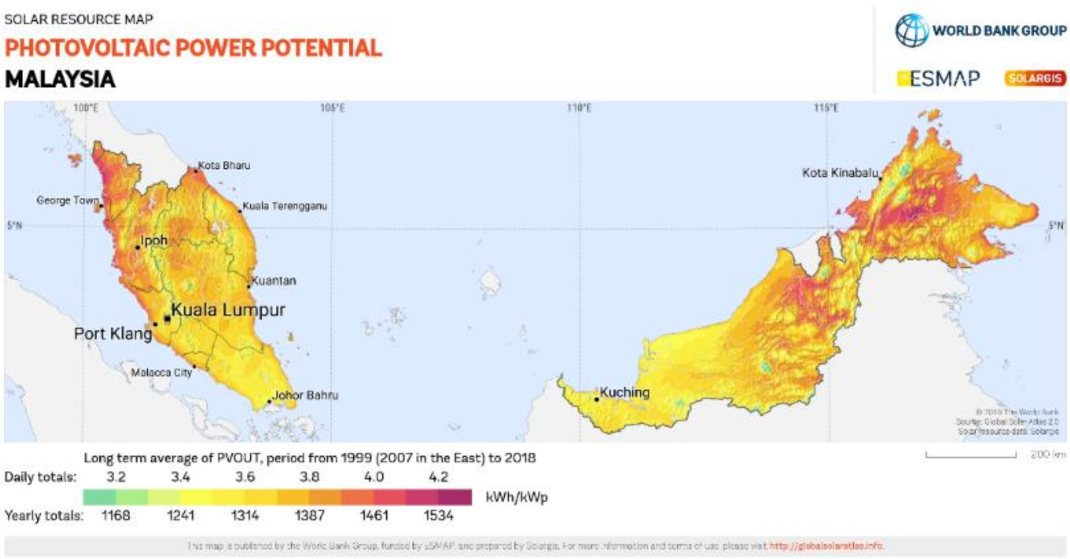

Figure 2.2 Malaysia solar resource map with potential PV power output. (Solargis, 2021)

Figure 2.3 shows the solar irradiance distribution profile with different irradiance levels. This distribution includes 8 sites of the in Malaysia which are distantly distributed and have different annual solar irradiation and the irradiance data are retrieved from PVGIS website. The solar irradiance distribution pattern in Malaysia can be observed with this file which will be helpful during inference of optimal ISR to other sites.

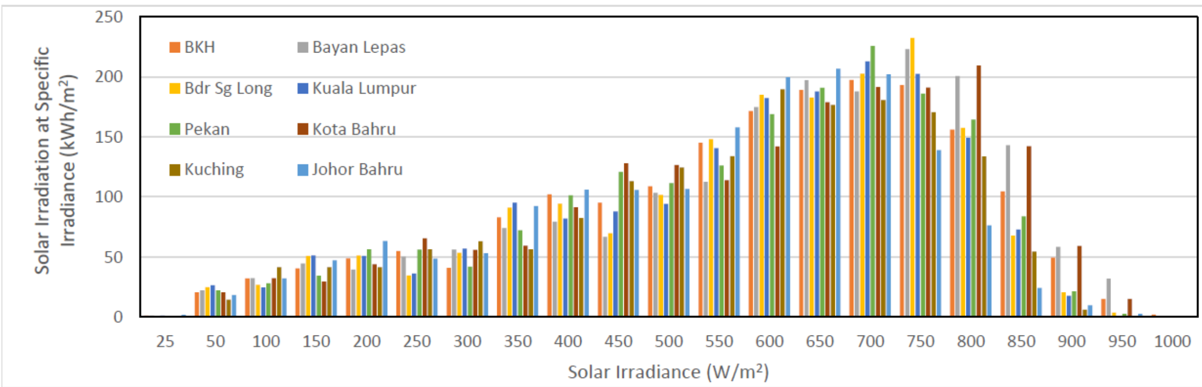

Figure 2.3 Solar irradiance distribution profile in Malaysia. (Lai and Lim, 2019)

Next, to further investigate the PV system performance to determine the optimal ISR, Global Tilted Irradiance (GTI) is used instead of Global Horizontal Irradiance (GHI). From the research work of Lim (2021), the percentage difference of the annual solar irradiance distribution profile of GTI and GHI from the same sites are 1.85% only, which is a negligible value. However, this small change can affect the value of optimal ISR of the PV system. The value of optimal ISR obtained from the GHI is 0.09 higher than the GTI. In addition, in Malaysia region, based on the outdoor experiment carried out by Mamun, Islam and Hasanuzzaman (2021), the optimal tilt angle is 15°. Thus, using GTI able to improve the optimal ISR value accuracy.

2.2 Inverter Sizing Ratio (ISR)

The major components of a grid-connected PV system are included: PV modules, mounting or tracking system, inverters, transformer and the grid connection cabling. In this project, the ISR value discussed here is for designation of the central inverters, which are mostly used for utility-scale PV project. The advantages of central inverters are high reliability and ease of installation. Figure 2.4 showed an overview of a megawatt scale On-grid PV system. (Miller and Lumby, 2012)

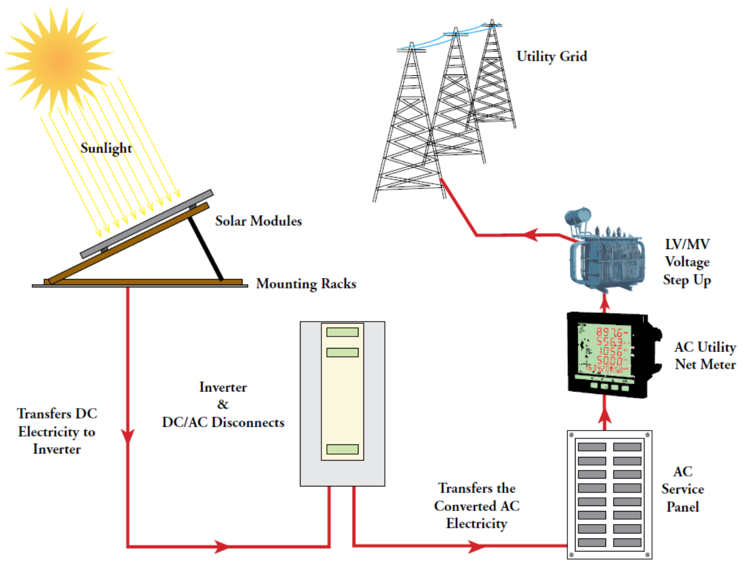

Figure 2.4 On-grid PV plant overview. (Miller and Lumby, 2012)

ISR also known as DC-AC ratio of inverter, the general equation of ISR is given by below:

$$ISR = \frac{P_{PV}}{P_{AC}} \quad (2.1)$$

Where

ISR = Inverter sizing ratio

P_{PV} = DC input power from the PV array, W

P_{AC} = AC output power of inverter, W

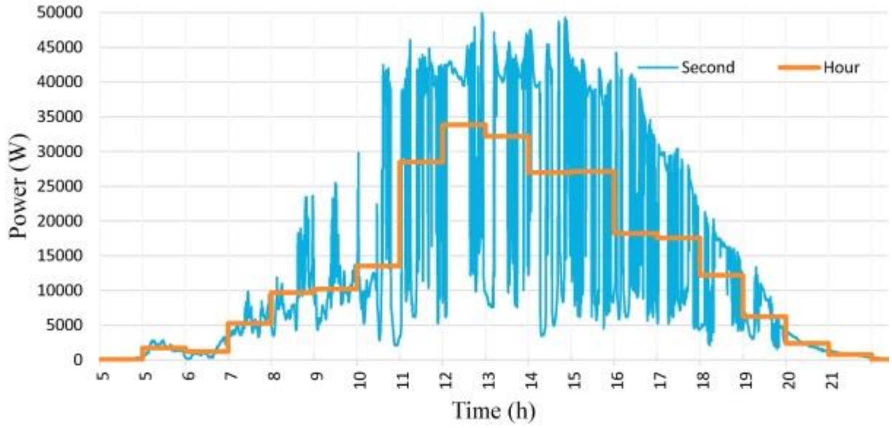

Figure 2.5: Production of PV plant with 1-sec and 1-hour resolutions during cloudy day. (Väisänen, et al., 2019)

A research paper done by Väisänen, et al (2019) was investigating the impact of time resolution of solar irradiance to the optimal sizing ratio. Based on the Figure 2.5, it can be observed that using 1-sec resolution data can avoid loss of irradiance. The researchers conclude that using of 1-hour resolution irradiance data to determine ISR can potentially skewed the result and cause more significant undersizing of inverter. Thus, the research paper states that using solar irradiance with smaller time resolution is beneficial to find the optimal ISR.

Therefore, the finding of this research paper can be utilized as one of the factors to investigate the relationship between ground-based measurement and satellite-derived irradiance data, where the ground-based measurement irradiance data is usually smaller than satellite-derived data.

2.3 Levelized Cost of Electricity (LCOE)

LCOE is defined as the estimation of the revenue needed to construct and operate a generator over a specified cost recovery period. (U.S. Energy Information Administration, 2021)

Usually, this fundamental calculation will be conducted in the preliminary phase of an energy generation project. The value of LCOE is helpful to decide whether the project is worth to proceed as compare with other electricity generation project. It helps to estimate if a project will break even or become profitable in future. (CFI, n.d.)

The general formula to compute the LCOE is presented as:

$$LCOE = \frac{Present\ Value\ of\ System\ Total\ Cost\ over\ Lifetime}{Electricity\ Generated\ by\ System\ over\ Lifetime} \qquad (2.2)$$

The present value of system total cost is the sum of initial capital cost and the predicted maintenance cost of the system over the lifetime of system whereas the electricity generated is AC energy generated at the end of inverter. The LCOE value is usually expressed in dollar per watt unit. (Lai and Lim, 2019)

2.4 Optimal Inverter Sizing Ratio

In the research paper done by Lai and Lim (2019), a method to compute optimal ISR which can minimize the LCOE of the PV system is introduced. The optimal ISR can be determined by choosing on the ISR value which give lowest LCOE. The method used in the paper divides the components of performance ratio of PV plant into system-dependent and non-system-dependent components. In this optimal ISR computation, it considers the inverter overloading ratio and the degradation of PV module and inverter. This assumption can increase the reliability of optimal ISR computed to apply in real situation. Thus, this method is adapted and applied in this project.

2.5 Artificial Neural Network (ANN)

ANN is one of the main tools used in machine learning. It relies on training data by providing enough sets of input data to learn and improve their accuracy overtime. The idea of ANN is inspired by the biological neural network in the human brain. By referring to the Figure 2.6 and Figure 2.7, the relationship between artificial and biological neural network are shown in the Table 2.1.

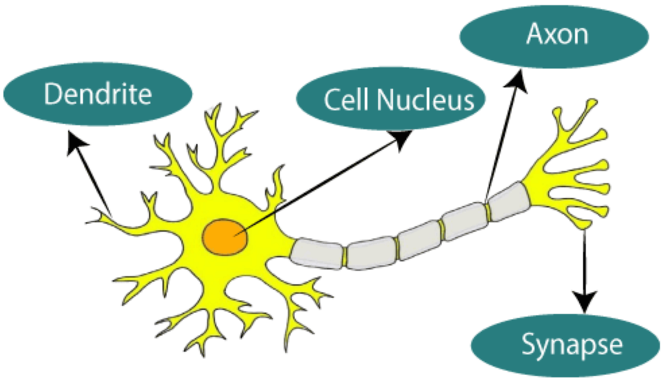

Figure 2.6 Biological neuron structure. (JavaTpoint, n.d.)

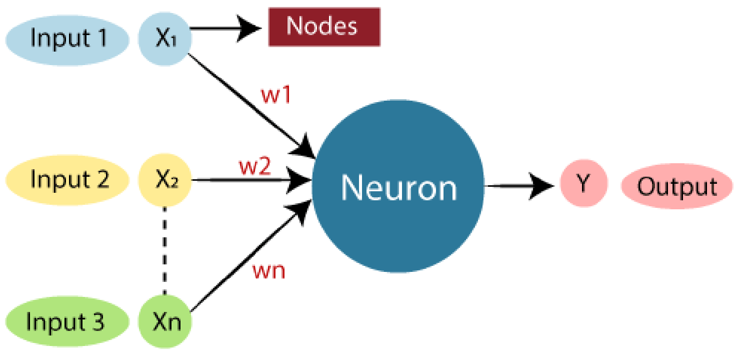

Figure 2.7 Artificial neuron structure. (JavaTpoint, n.d.)

Table 2.1: The functions of components in ANN corresponding to the biological neural network.

Biological Neural Network	Artificial Neural Network
Dendrites	Inputs
Cell Nucleus	Nodes
Synapse	Weight
Axon	Output

The neural networks here include 3 types of layers which are input layer, hidden layers, and output layer. The input layer will receive the input data with several formats and parameters which will be provided by the programmer. Then the data will be analysed and activate the corresponding nodes in hidden layers. The hidden layers will perform all the calculation to find out the hidden feature and pattern and past it to next layers. The next layer can either next hidden layer or output layer. After the input goes through a series of transformations and analysis in the hidden layers, the final output results will be conveyed to the output layers. (JavaTpoint, n.d.)

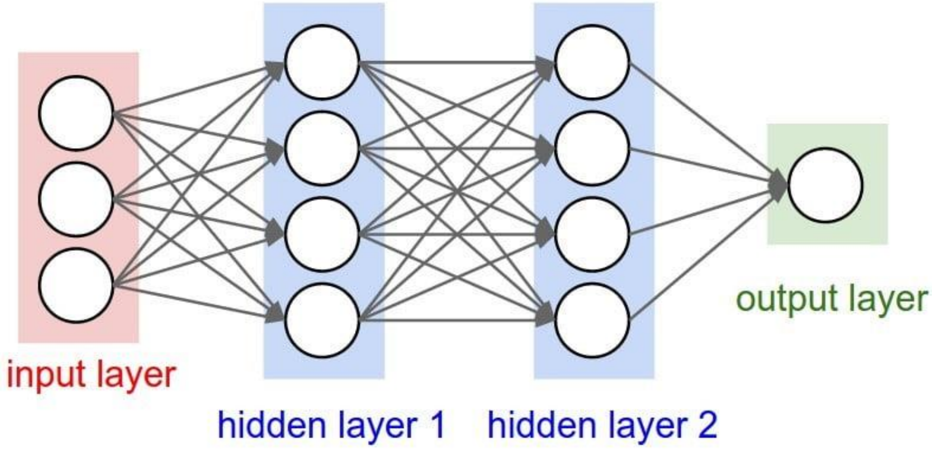

Figure 2.8 Types of layers available in an artificial neural network. (Dormehl, 2019)

The propagation of information can be various forms of forward and backward among the hidden layers. The change of flows of information can become multiple types of ANN, each of it comes with different specific use cases and complexity levels. The most basic type is the Feed-forward Neural Network which its information flows in forward direction only (input to output). Next, the network which is great at learning handwriting and language recognition are the Recurrent Neural Network. This network allows the data to flow in multiple directions which help it to possess better learning ability. (Dormehl, 2019) Other than these, there are some other ANN models popular in nowadays such as Convolutional or Deconvolutional Neural Network, and the Modular Neural Network. These high-level ANN models allow it to process more complex information such as signal processing and image classification. (Burns and Burke, 2021)

The basic unit of ANN is called perceptron, which is the artificial neuron that comprise with ANN. Each individual node has their own linear regression model, which is composed by input data, weight, a bias (or threshold), and an output. The general formulas below showed the relationship between them and how it works for decision making.

To simplify the equation, the summation of each perception is first change to dot product first. (Nielsen, 2019)

$$w \cdot x = \sum_{i=1}^{m} w_i x_i \tag{2.3}$$

Where

m = number of perceptron stacked

w_i = the weight of the specific perceptron

x_i = the input data

$$output = \begin{cases} 0 & if\ w \cdot x + bias \ < 0 \\ 1 & if\ w \cdot x + bias \ > 0 \end{cases} \tag{2.4}$$

Where

bias = the threshold value determines the activation of output

Other than the basic neurons, sigmoid neurons are introduced to make the training of the network easier. Sigmoid neurons are similar to basic perceptron but it able to reduce the change of output if small change of weight and bias is occurred. The sigmoid function is defined as:

$$\sigma(z) \equiv \frac{1}{1+e^{-z}} \tag{2.5}$$

Combining the equation (2.19) and (2.20), the output of the sigmoid neuron is:

$$Output = \frac{1}{1+\exp\left(-\sum_j w_i x_i - bias\right)} \tag{2.6}$$

In Figure 2.9, the yellow hexagons represent the transfer function applied, and the blue box is the intermediate results from each layer or the final result at output layer. (Mesquita, 2021)

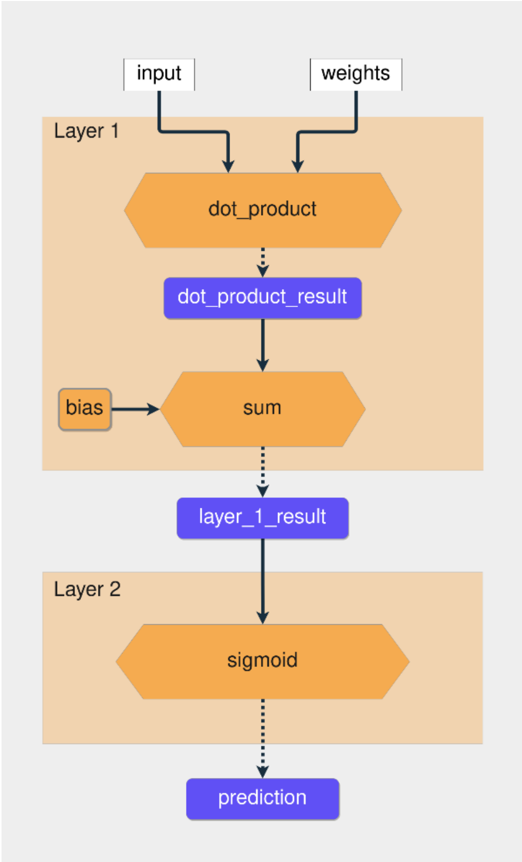

Figure 2.9 The flow of computation inside the ANN

However, this flow of computation does not have the mechanism to detect the error and training. A feedback loop needs to be applied to make the prediction closer to the targeted value and reduce error in output. Gradient descent and backpropagation algorithms can be used to adjust the weight and bias. Before the correction of prediction started, mean squared error (MSE) is computed to understand the magnitude of the error. The formula of the error here is the squared difference between prediction and target value. It can be written in the following equation (2.22): (Mesquita, 2021)

$$error = (prediction - target)^2 \tag{2.7}$$

After getting the MSE error, if the (prediction – target) term is treated as a single variable X, it can be found that the error can be derived as quadratic function, where error = $(X)^2$. The error is plotted as y axis. As the graph showed in Figure 2.10, if the current error point is located at A, to reduce the error down to 0 value, gradient descent algorithm is now used to find the direction and rate of descent to update the weights in the ANN. The overview on how to partial derivatives inside the ANN is shown in Figure 2.11. If more than one hidden layer is applied, chain rule can be executed to get error in each stage and update weight along the flow of data. (Mesquita, 2021)

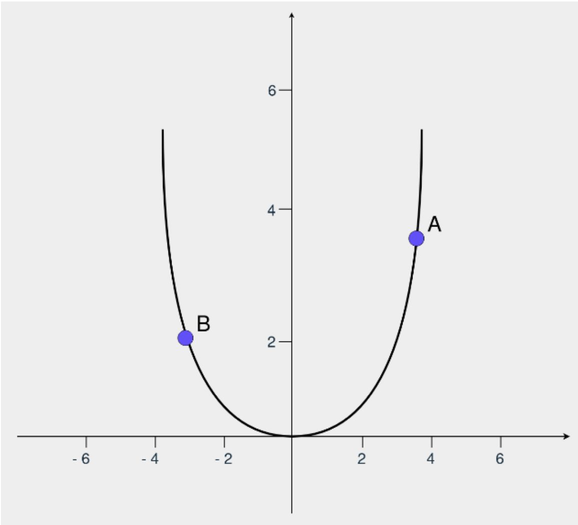

Figure 2.10 An example of quadratic function.

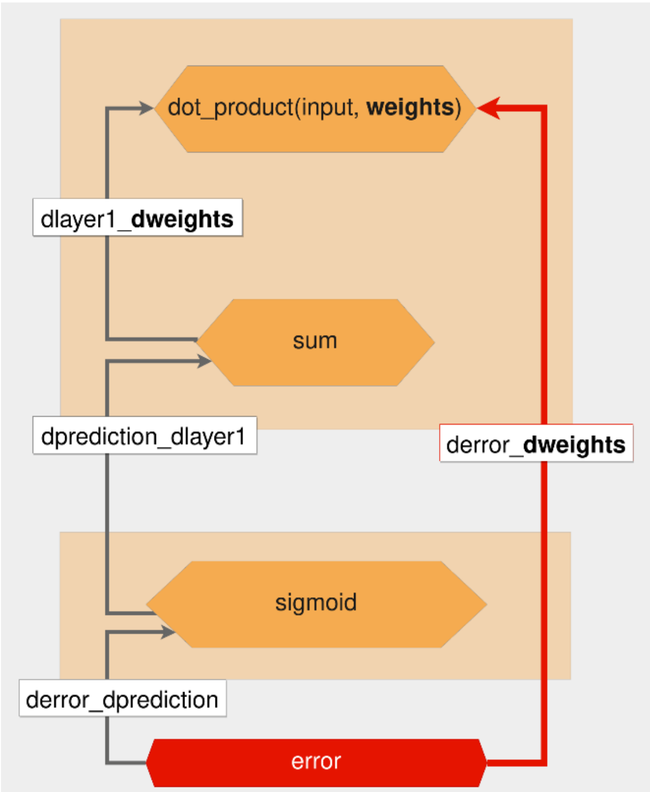

Figure 2.11 The partial derivatives to adjust weight in ANN.

Next, with the backpropagation algorithm, the bias of the ANN can be updated by taking the derivative of the error function with respect to the bias using the same logic as the partial derivative in adjusting the weight. Figure 2.12 shows the flow chart of partial derivative to adjust the bias value. (Mesquita, 2021)

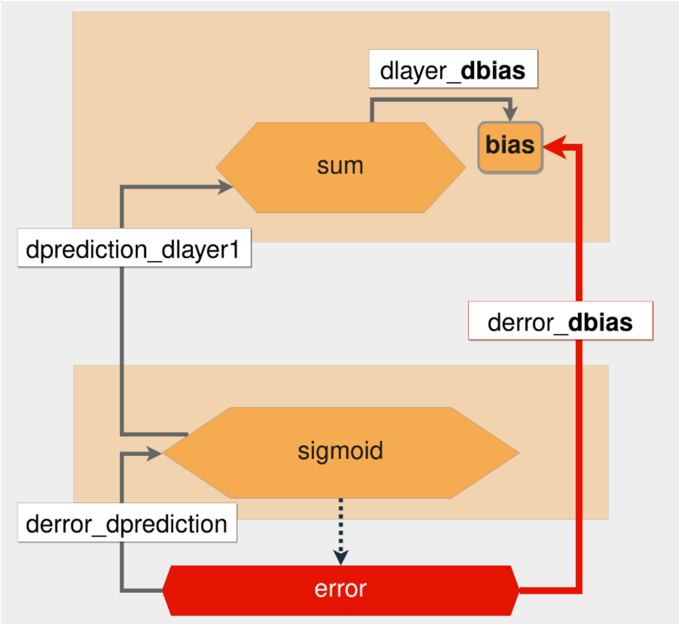

Figure 2.12 The partial derivatives to compute the bias gradient.

2.6 Method of Estimation (Inference)

Statistical inference is a process to estimate or make conclusion about one or more characteristics in a statistical population. The estimation can be done by analysis of sample data drawn from the statistical population. It allows people to do estimation of the value needed when the parameters are varied by assessing the relationship between the dependent and independent variables. Inference provides a probable ranges of values for the characteristics in a statistical population. (Camm *et al.*, 2015)

CHAPTER 3

METHODOLOGY AND WORK PLAN

3.1 Introduction

The relationship between optimal ISR simulated by ground-based and satellite-derived datasets is complex. Thus, in this project, an ANN model will be developed to infer the optimal ISR using satellite-derived datasets.

Figure 3.1 below shows the overview of the flow of methodology. The flow of methodology can be divided into 5 stages. The first stage is mainly dealing with the solar irradiance data from ground-based and satellite-based datasets. If the datasets have missing data, interpolation needs to be carried out. The second stage is to calculate the monthly optimal ISR value and monthly irradiation which will be used in the training and the testing of ANN model. The third stage is to construct and train the ANN model. The fourth stage in the methodology flow is to optimize the ANN model by using the training datasets until lowest loss and error is obtained. Lastly, the fifth stage is to infer the monthly optimal ISR of others site that do not have ground-based measurement solar irradiance After inferencing of optimal ISR, the results of inference will be analysed, and the performance of ANN model will be evaluated.

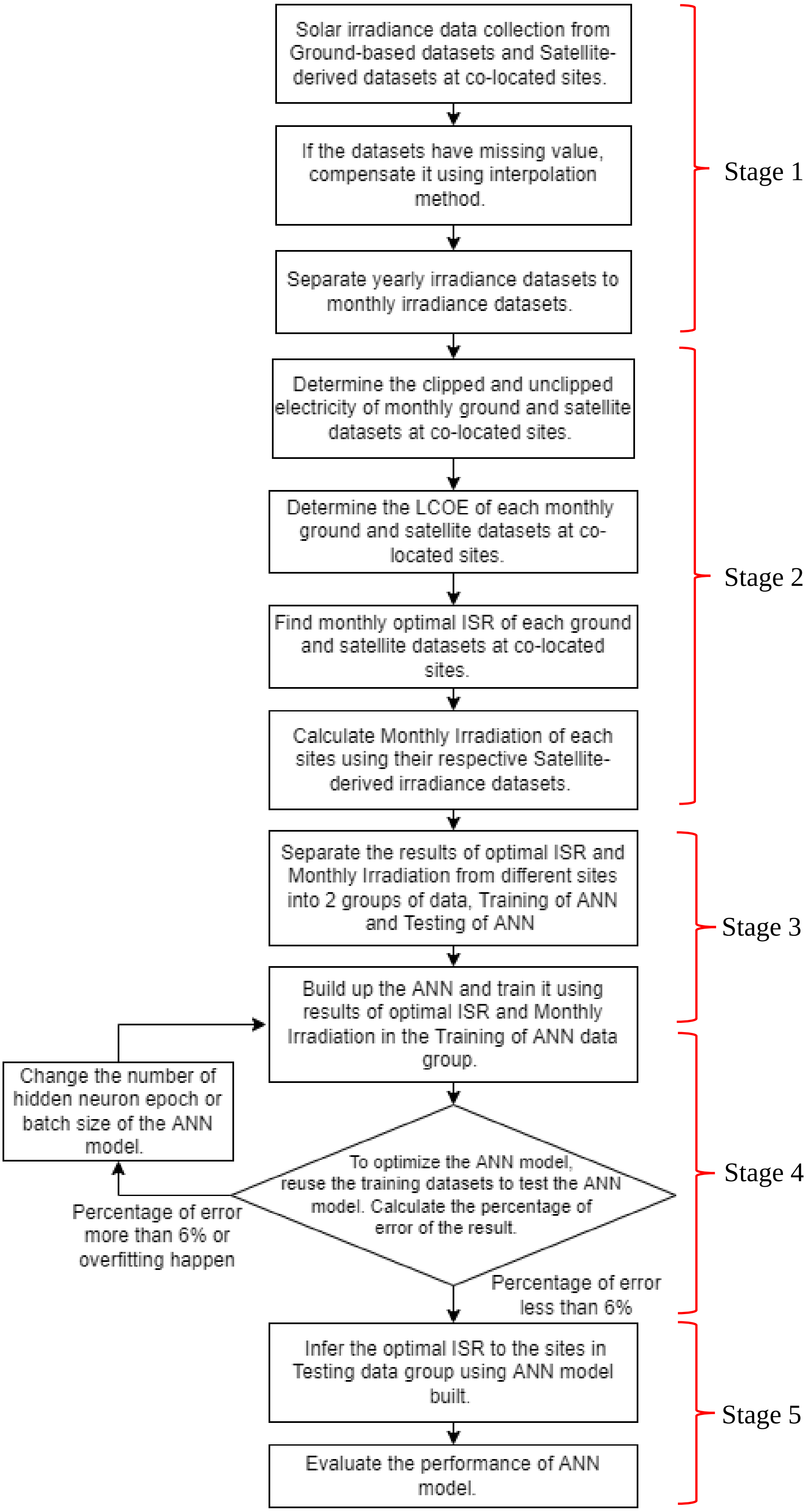

Figure 3.1: General Flowchart of the methodology.

3.2 Retrieving solar irradiance datasets

There are 2 types of sources for the solar irradiance data, the ground-based measurement datasets, and the satellite-derived datasets. The solar irradiance datasets must include the date and the time of the data collected, GHI and ambient temperature. The solar irradiance datasets from ground-based measurement and satellite-derived must be co-located, so that the relationship of optimal ISR from both ground-based measurements and satellite-derived datasets can be investigated. In this project, the optimal ISRs from 10 sites in Malaysia are investigated. The irradiance datasets in all 10 sites will be available in both ground-based measurement and satellite-derived irradiance datasets. However, some of the ground-based measurement datasets have a problem of missing data due to some technical problem in the weather station. The process to compensate the missing data problem will be explained in detail in Section 3.4.

The Table 3.1 below showed the 10 sites which will be used in this project.

Table 3.1: Summary of available sites in this project.

No.	Name	Location	Number of Months Available in Year 2019 Solar Irradiance Datasets
1	UTAR	Selangor	12
2	Bukit Kayu Hitam (BKH)	Kedah	12
3	AyerKeroh	Melaka	6
4	Permatang	Penang	10
5	AxisBC	Selangor	12
6	Balakong	Selangor	11
7	LFLogistic	Selangor	10
8	StarkenACC	Selangor	9
9	TelePower	Selangor	7
10	SRTech	Selangor	10

3.3 Software and programming language used

Microsoft Excel is used to store the solar irradiance datasets and LCOE results from different ISR. The function of "Conditional Formatting" in Excel is used to select the ISR value which have the lowest LCOE value as optimal ISR. The Pivot Table in Excel is helpful to check the missing day or time in the solar irradiance datasets.

Python language is chosen to write the program for interpolation, calculation of optimal ISR, training and testing of the ANN model, and evaluation for the performance and prediction of the ANN model built. Moreover, it is open-source and free to use, and easy to write and read since Python is a high-level programming language which has English-like syntax.

3.4 Interpolation on missing data

Sometimes the irradiance data from ground-based measurement might have data loss problem. This might be due to the hardware problem such as malfunction of the equipment to store and upload the date or the sensor used in weather station. Thus, before the calculation of optimal ISR started, the total amount of data needs to be checked to see if any data loss occurred. If there are missing of data found in the source of irradiance and ambient temperature data, it can be compensated using interpolation method before proceeding to next stage. The searching of missing datetime is done by using the Pivot Table function in the Microsoft Excel software.

The interpolation method used in this project will be done by using the interpolation function from the Pandas library in python. The available functions are linear, nearest, cubic, spline etc. These interpolation function can only interpolate the series of data which has continuously same time interval. A sample python program using the stated interpolation function was done before, the code is shown in Appendix A and the results was compared with the original data which has no missing value to compare the error of each interpolation functions. The methodology of the interpolation python program refers the study done by Mohamad, Lai and Lim (2022). 10% to 50% of data are randomly removed from the irradiance data. Then different interpolation functions are applied to the irradiance data and the error of the results are calculated and analyzed.

At the end of the testing on interpolation function, linear interpolation function has the lowest error among the tested functions. The comparison of the sum of error among the tested interpolation function are shown in Figure 3.2. Thus, the linear interpolation function from the Pandas library is used to interpolate the missing data in the raw solar irradiance data. However, there are some when the missing data is too much (more than 10 consecutive time interval) or the missing data is located at the beginning or end of the day (where starting point and ending point cannot be estimated), that particular day with missing data will not be interpolated.

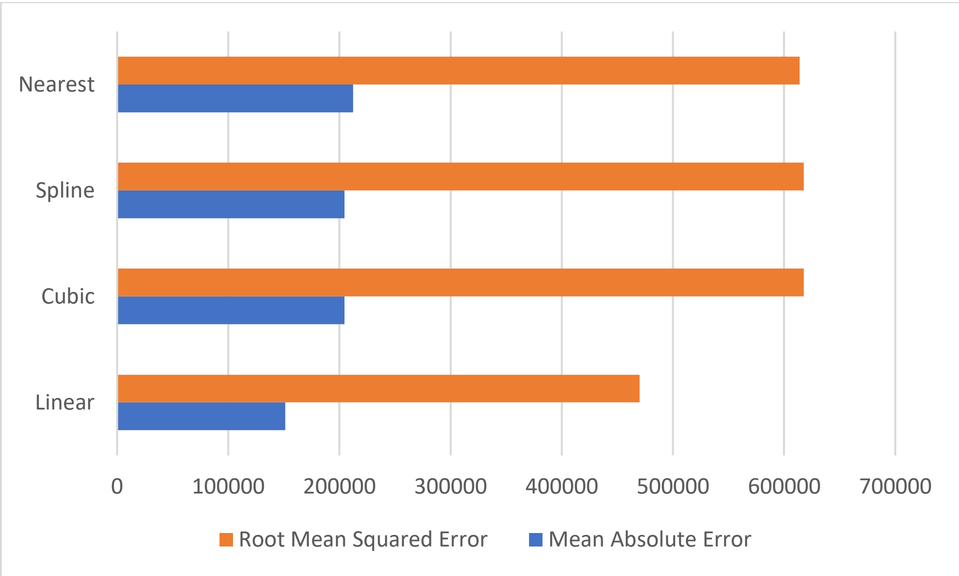

Figure 3.2 Error comparison among the tested interpolation function.

3.5 Calculate LCOE to determine optimal ISR

In this project, the optimal ISR will be determined based on the LCOE value. An iterative method will be used to calculate the LCOE value of PV system with different ISR given. The value of ISR will be iterated from 1.2 to 1.94 with a 0.01 step value. The LCOE of PV system with different ISR will be recorded. The ISR value with lowest LCOE value will be selected as the optimal ISR, which means the inverter is designed to produce the electricity at the lowest cost and should be the most profitable in long run. The computation of LCOE with different ISR values will be programmed with Python Language and the results being stored in a Microsoft Excel file.

There are total of 10 sites that provide the ground-based measurement solar irradiance. However, the irradiance data collected from these sites is only referring to one year, which is year 2019. Since the number of solar irradiance datasets is limited, the computation of LCOE and Optimal ISR is modified from yearly value to monthly value. The yearly irradiance data and ambient temperature in both the ground-based measurement datasets and satellite-derived datasets from different sites will be divided into monthly data. This can be done by checking the month element in the date-time data of the irradiance datasets.

Table 3.2: The fixed parameters and their nominal values used in LCOE computation.

Parameters	Value	Unit
DC Capacity	10.00	MW
Degradation Rate	0.50	%/ year
Fixed Performance Ratio	0.92	-
Temperature Coefficient	-0.0038	/°C
Ross Coefficient	0.0234	°C/ (W/m^2)
Inverter Overpower Ratio	1.1	-
O & M Cost	200 000.00	RM/ year
Nominal Tariff Rate	0.28	RM/ kWh
PV Module Price	1.28	RM/ W
Inverter Price	0.52	RM/ W
Other Cost - Installation Cost - Miscellaneous Fees - Other Balance-of-system Cost	1.68	RM/ W

The PV system size is set at 10 MW in the computing of LCOE value with different site irradiance data. There are some parameters need to be fixed with their nominal value before the computation of LCOE started. The nominal values can be fixed because they are less subjected to variation at least in this region and will not yield a significant change in calculating the optimal ISR. The values are provided by my supervisor, Ir Dr. Lim Boon Han through his experience in the field. Some sensitivity analyses were conducted by previous student (Winston, 2021). The parameters and their nominal values are shown in the Table 3.2. (Lim, 2021)

First of all, the AC electricity yield per day of the PV system is calculated based on the formula below: (Lai and Lim, 2019)

$$E_{AC\_per\ day} = \sum_{t=1}^{t=288} [D \times (P_{PV} \times PR(t) \times GHI(t))] \tag{3.1}$$

Where

$E_{AC\_per\ day}$ = AC electricity yield per day of a PV system, kWh

P_{PV} = the capacity of PV system to be installed, MW

$PR(t)$ = performance ratio (PR) during interval of time

$GHI(t)$ = Global Horizontal Irradiance to the PV array during interval of time, W/m^2

D = the period of time for the discrete value of the output power, hours

t = interval of time, hours

Based on the data used, the D and t can be modified, if the format of data is in 5-minutes interval, D will be 5 minutes and the t will be started from 0 to 288 if the range of time of data sampled is 24 hours. 288 is the total amount of data per day, which can be understand as equation below:

$$\frac{60\ min}{5\ min} \times 24\ hours\ per\ day = \ 288\ sampled\ data \tag{3.2}$$

Next, the PR terms need to be further processed before passing the value to calculate the AC electricity yield. The performance ratio of a PV system is contributed by 2 groups of losses, fixed loss, and unfixed loss, specifically proposed by Lai and Lim (2019) for the performance of PV system located in the tropics, where the performance ratio has less than 3% variation across different months of a year. The small variation of performance ratio allows us to convert the yearly data to monthly data without causing significant change on optimal ISR value. Fixed loss will only affect the PR value when clipping of energy in small scale, it usually can be controlled with periodic maintenance, and good design. However, the unfixed losses are usually easily affected by environmental factor around the PV array. For example, the PV array performance will drop if the ambient and module temperature is increased, and the varied loading factor of inverter contributes to the inverter conversion loss. Table 3.3 has summarised and listed different type of loss in the system. (Lai and Lim, 2019)

Table 3.3: Type of Losses.

Name	Fixed or Not Fixed Loss
Soiling Loss	Fixed
Near Shading Loss	Fixed
Low Irradiance Loss	Fixed
Module Quality Loss	Fixed
Module Array Mismatch Loss	Fixed
Ohmic Loss in Wire	Fixed
Incident Angle Modifier Factor Loss	Fixed
PV Loss due to Temperature	Not fixed
Inverter Conversion Loss	Not fixed

Therefore, the performance ratio, PR(t) can be derived as the formulas (3.3) to (3.5) below: (Lai and Lim, 2019)

$$PR(t) = PR_{fix} \times \eta_{inv}(l) \times f_{temp}(t) \tag{3.3}$$

Where

PR_{fix} = the sum of fixed loss in PR as listed in Table 3.3.

$\eta_{inv}(l)$ = the inverter conversion efficiency depends on the loading factor of the inverter used

$f_{temp}(t)$ = the derating factor of PV array caused by change of temperature

In the equation (3.5), the $\eta_{inv}(l)$ stated is chosen a series of number sampled from efficiency curve of inverter in the PVsyst software. The inverter used in this project simulation is SG2500 model from Sungrow company. The Figure 3.3 showed the efficiency curve of the selected inverter model.

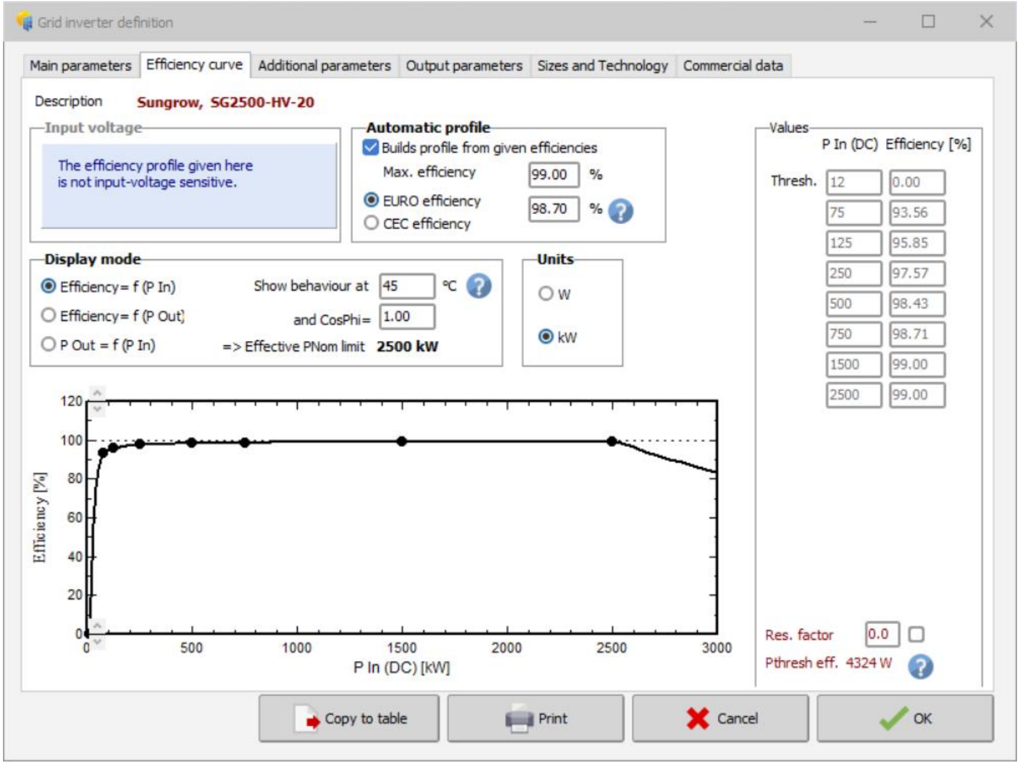

Figure 3.3: Screenshot of inverter efficiency curve from PVsyst software.

$$f_{temp}(t) = 1 + \gamma(T(t) - T_{STC}) \tag{3.4}$$

Where

γ = the temperature coefficient for power of PV module

$T(t)$ = the instantaneous PV module temperature, °C

T_{STC} = the reference temperature stated in Standard Test Conditions (STC), °C

$$T(t) = T_{amb}(t) + (GHI(t) \times C_{ROSS}) \tag{3.5}$$

Where

$T_{amb}(t)$ = the ambient temperature of PV module, °C

C_{ROSS}= Ross Coefficient slope, °C / (W/m^2)

After combining the equation (3.3) and (3.5), a summarised equation to get the expected instantaneous AC output power at that time interval can be derived, without consider any clipping power due to inverter rating. (Lai and Lim, 2019)

$$P_{AC\_exp}(t) = P_{PV} \times GHI(t) \times PR_{fix} \times \eta_{inv}(l) \times \{1 + \gamma[(T_{amb}(t) + GTI(t) \times C_{ROSS}) - T_{STC}]\} \quad (3.6)$$

Where

$P_{AC\_exp}(t)$ = the expected instantaneous AC output power without consider clipping of power, MW

However, during the event of high solar irradiance happened, clipping of AC power will occur due to the maximum AC power rating of inverter, P_{AC_max}. The P_{AC_max} of an inverter is usually designed as 1.10 times greater than the rated power of inverter, P_{AC_rated} value. The actual AC output power is no longer same as the P_{AC_exp} but change to the P_{AC_max}. Thus, the actual AC power output of inverter, $P_{actual}(t)$ is derived in the equations below: (Lai and Lim, 2019)

$$P_{AC\_max} = P_{AC\_rated} \times 1.10 \quad (3.7)$$

Where

P_{AC_max} = the maximum AC power rating of inverter, MW

P_{AC_rated} = the AC power rating of inverter, MW

$$P_{AC\_rated} = P_{PV}/ISR \quad (3.8)$$

Where

ISR = inverter sizing ratio

The ISR value used in the equation (3.8) is the iterated values in the range from 1.2 to 1.95.

$$P_{actual}(t) = \begin{cases} P_{AC\_exp} & if\ \ P_{AC\_exp} < P_{AC\_max} \\ P_{AC\_max} & if\ \ P_{AC\_exp} > P_{AC\_max} \end{cases} \tag{3.9}$$

Where

$P_{actual}(t)$ = the actual instantaneous AC power output of inverter at that time interval, MW

Next, the actual instantaneous AC power, $P_{actual}(t)$ of each time interval in a day needed to be summing up, if the solar irradiance datasets applied is 5-minutes resolution, the actual value of AC electricity yield per day of a PV system, $E_{AC\_per\ day,act}$ is showed by below: (Lai and Lim, 2019)

$$E_{AC\_per\ day,act} = \sum_{t=1}^{t=288}[D \times P_{actual}(t)] \tag{3.10}$$

Where

$E_{AC\_per\ day,act}$ = AC electricity yield per day of a PV system with consideration of clipping power during event of high irradiance, kWh

Then, the $E_{AC\_per\ day,act}$ is summed up to get the total electricity generation in particular month. The degradation rate of PV modules will be increased by year, usually is with the rate of +5% per year. In conclusion, the electricity yield, for a specific month can be obtained by using the equations below: (Lai and Lim, 2019)

$$E_{AC\_per\ month,act} = (1 - N_y d) \sum_{N=1}^{N=30} E_{AC\_per\ day,act} \tag{3.11}$$

Where

N = Number of days in that specific month

d = the degradation rate of the PV module

N_y = the number of years PV module used

There are 3 number of years PV module will be used in the LCOE computation, which are 15 years, 21 years, and 25 years. 15 years is selected as one of the number of years PV module used is because the life span of inverter is usually 10-15 years. Moreover, 21 years is chosen because most of the Solar Power Purchase Agreement (SPPA) chose 21 years as period of agreement. Lastly, 25 years is chosen is because the life span and warranty of PV module is usually up to 25 years only.

Next, the total generated electricity yield by the PV system, E_{AC_sum} within a specific time frame, L can be predicted. (Lai and Lim, 2019)

$$E_{AC\_sum} = \sum_{N_y=1}^{N_y=L} E_{AC\_per\ month,act} \tag{3.12}$$

Where

E_{AC_sum} = total electricity generated by PV system within a specific time frame.

Furthermore, the monthly LCOE of the system is derived in the following equations: (Lai and Lim, 2019)

$$LCOE = \frac{(capital + O\&M\ cost) \times \left({}^{N}\!/_{365}\right)}{E_{AC\_sum}} \tag{3.13}$$

Where

capital = the capital cost of the system, RM

O&M cost = the operating and maintenance fee of the system within a specific time frame, RM

To calculate the capital cost of the PV system, the equation from (3.14) to (3.16) is computed. The expected saving cost of a downsized inverter is given by the following equation: (Lai and Lim, 2019)

$$\$_{inv,exp} = P_{PV}\left(1 - \frac{1}{ISR}\right) \times Price_{inv} \tag{3.14}$$

Where

$\$_{inv,exp}$ = the cost saved by downsized inverter which included the clipped electricity, RM

$Price_{inv}$ = the market price of the inverter with chosen capacity, RM/W

The net saved cost of downsized inverter can be obtained by the following equation: (Lai and Lim, 2019)

$$\$_{inv,net} = \$_{inv,exp} - (P_{AC\_exp} - P_{actual}) \times \$_{TARRIF} \qquad (3.15)$$

Where

$\$_{inv,net}$ = The net saved cost by downsizing the inverter using optimal ISR, RM

$\$_{TARRIF}$ = The market price to sell electricity for a specific plant capacity, RM

The market price of the PV system installed can obtained by the following equation:

$$Price_{PV\ system} = Price_{PV\ module} + Price_{inv} + Price_{\ Other} \qquad (3.16)$$

Where

$Price_{PV\ system}$ = the market price of the PV system, RM/W

$Price_{PV\ module}$ = the market price of the PV module, RM/W

$Price_{Other}$ = the market price of the other cost in PV system, RM/W

The interest rate of loan, incentives and cost incurred to manage risk are not considered in the net present value of the future cost. The capital cost of the system is the cost after obtaining the saved cost of downsized inverter. The formula to calculate the capital cost are shown below: (Lai and Lim, 2019)

$$capital = P_{PV}\left(Price_{PV\ system}\right) - \$_{inv,exp} \qquad (3.17)$$

The calculated capital cost is substituted into equation (3.13) to calculate monthly LCOE. This computation of LCOE will be repeated for different ISR value which is in range of 1.2 to 1.95 with 0.01 step. The LCOE and respective ISR value is recorded and exported from python software to Microsoft Excel as shown in Figure 3.4 below.

A1

	A	B	C	D	E
1		**15 years**	**21 years**	**25 years**	
2	**for ISR 1.2**	0.14088	0.1073	0.09397	
3	**for ISR 1.21**	0.14064	0.10713	0.09382	
4	**for ISR 1.22**	0.14044	0.10698	0.09369	
5	**for ISR 1.23**	0.14023	0.10683	0.09357	
6	**for ISR 1.24**	0.14003	0.10668	0.09344	
7	**for ISR 1.25**	0.13983	0.10654	0.09332	
8	**for ISR 1.26**	0.13964	0.1064	0.0932	
9	**for ISR 1.27**	0.13944	0.10626	0.09308	
10	**for ISR 1.28**	0.13925	0.10611	0.09296	
11	**for ISR 1.29**	0.13905	0.10597	0.09284	
12	**for ISR 1.3**	0.13887	0.10584	0.09272	
13	**for ISR 1.31**	0.13869	0.10571	0.09261	
14	**for ISR 1.32**	0.13851	0.10558	0.09251	
15	**for ISR 1.33**	0.13834	0.10545	0.0924	
16	**for ISR 1.34**	0.13817	0.10533	0.09229	
17	**for ISR 1.35**	0.138	0.10521	0.09219	
18	**for ISR 1.36**	0.13783	0.10508	0.09208	

Figure 3.4: Example result of LCOE for different ISR value.

Next, with the aid of “Conditional Formatting” function is Microsoft Excel, the ISR value with lowest monthly LCOE is selected as monthly optimal ISR value for specific sites and month. Repeat the computation of LCOE and monthly optimal ISR with the satellite-derived irradiance datasets which is co-located with the ground-based measurement solar irradiance datasets. Then, repeat the whole process to another sites.

3.6 Separate available datasets into training and testing data group

There are 3 types of data need to be generated, before modelling of ANN. Firstly, two types of the data will be the input data of the ANN model, which are the monthly solar irradiation and the optimal ISR computed from satellite-derived datasets. On the other hand, the remaining one type of data is optimal ISR computed from ground-based measurement datasets, which will be the target value in the training process and also will be compared with the optimal ISR that are predicted by the ANN model to evaluate the ANN model performance.

By referring to the Table 3.1, the number of datasets is determined by the number of months available in year 2019 solar irradiance datasets. The total number of datasets available is 99.

The datasets are divided into two groups. 80% of the available datasets (about 80 datasets) will be chosen to be the first group, which help to train the ANN and investigate the relationship between optimal ISR compute from ground-based datasets and satellite-derived datasets. The remaining 20% of datasets (about 19 datasets) will be chosen to be the second group to test and evaluate the accuracy of the ANN by only providing the optimal ISR computed from satellite-derived datasets. Evaluation of ANN model will be done by comparing the output result of ANN with the original value of optimal ISR computed from ground-based datasets. Thus, the second objective as stated in Section 1.4 can be fulfilled.

3.7 Construction of Artificial Neural Network (ANN)

The basic working principle of ANN is described and explained in the Section 2.4. However, it is time consuming and difficult to create an ANN from scratch. Therefore, with the help of external open-source library available in Python, the training process could be easier and more efficient. The library imported is called Keras which is from TensorFlow platform. The deep learning algorithm in library able to learn a non-linear function approximator for either regression or classification.

In training stage of ANN model, in order to achieve the first objective as stated in Section 1.4, the monthly optimal ISR computed by satellite-derived datasets and monthly irradiation of the specific site are set to be input for ANN whereas the monthly optimal ISR simulated by ground-based datasets at co-located site is set to be the final target value in the output layer of ANN. The schematic diagram of ANN model which designed in this project is shown in Figure 3.5.

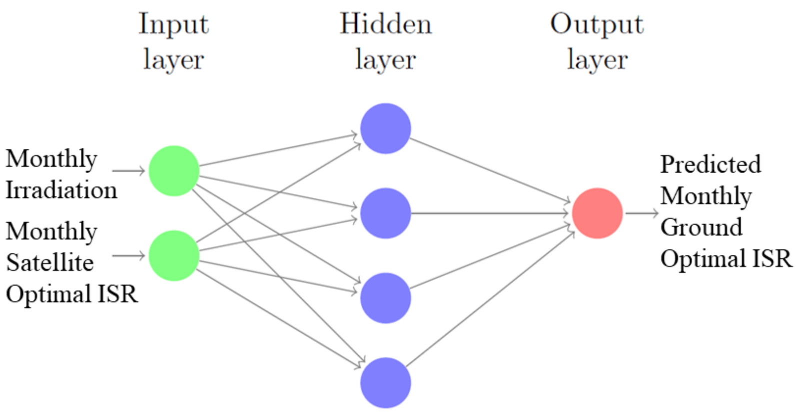

Figure 3.5: Schematic diagram of the ANN model.

Table 3.4 below showed the list of the parameters of ANN model used in this project.

Table 3.4: Parameters of the ANN model.

Parameters	Description or value
Activation Function (Hidden Layer)	ReLU
Activation Function (Output Layer)	Linear
Optimizer	Adam
Loss Function	Mean Absolute Error
Metrics for evaluation of model	Mean Squared Error
Number of Hidden Layer	1
Number of Hidden Neuron	4 (range from 3 to 5)
Epochs	1300 (range from 500 to 1500)
Batch Size	2 (range from 2 to 20)

In a neural network, the activation function is responsible to transform the summed weighted input from the node into the activation of the node. The activation function used is ReLU and Linear. ReLU is a piecewise linear function. When the input is positive value, it will be transfer directly to the output of function without changing its value, otherwise is zero. By comparing to Sigmoid and Hyperbolic tangent activations function, ReLU overcomes the problem of vanishing gradient and make the models learn faster with better performance.

The selected optimizer is “Adam”, which is known as a combination of RMSprop and Stochastic Gradient Descent (SDG) with momentum. It uses the squared gradients to scale the learning rate like RMSprop and it takes the advantage of momentum by using moving average of the gradient instead of gradient itself like SGD with momentum, which make the modelling process require less memory and quicker to converge toward the minima (which give lowest error between predicted and targeted value).

Moreover, the loss function and metrics are used to optimize and evaluate the performance of model. The loss function will work together with optimizer by sending back the difference to adjust the model weights through the back propagation stage as stated in Section 2.3. On the other hand, the metrics function output will gradually converge to zero which will show the progress of optimization of ANN model, where the predicted value is approaching to the targeted value as the iteration of epochs from 0 to 1400. In addition, to be highlighted that during the training stage of ANN model, 20% of training datasets is extract out from training datasets for validation purpose at the end of each epoch. Then, the history data of ANN model is recorded and a graph of both loss and validation error vs epoch is plotted. Both loss and validation error should gradually decrease as epoch increase and approach to zero at the end of training process. The Figure 3.6 below showed an example of error graph after the training of the ANN model done. The graph also helped in defining the epoch required by the training of model to avoid underfitting or overfitting.

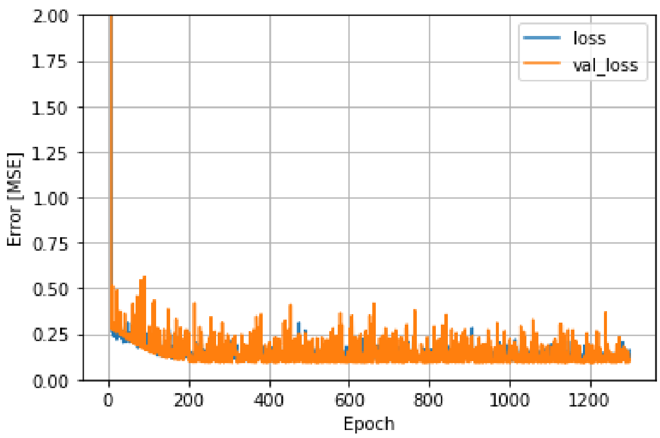

Figure 3.6: Graph of error vs epoch of the ANN model.

Furthermore, a graph showed the comparison between predicted value and targeted value of testing datasets and training datasets is also recorded and plotted using Python Matplot library as shown in Figure 3.7 below. Besides, the percentage of different between predicted and actual value for both training datasets and testing datasets is calculated.

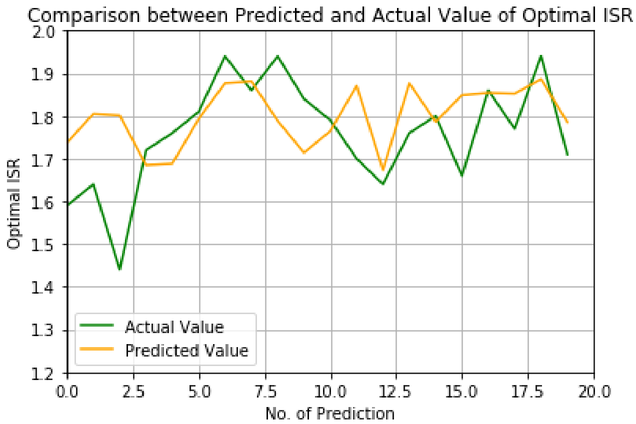

Figure 3.7: Comparison between predicted and actual value of optimal ISR.

3.8 Optimization of ANN model

Repeat the training process with different combination of number of hidden neurons, number of epochs and the batch size. When the percentage of difference between actual value and predicted value of testing datasets which are simulated from different ANN model is less than 6%, the value of parameters used is recorded and the ANN model is saved in .h5 file format to further used for prediction of testing datasets. The reason that 6% of difference is set as threshold value is because the average percentage of difference of optimal ISR simulated by ground-based measurement and satellite-derived irradiance datasets of all available sites is 6%.

3.9 Infer the optimal ISR to other sites

After the training and optimization of the ANN is done, the ANN model with best performance will be used for inference of monthly optimal ISR of other sites. The inference of optimal ISR will use the data available in testing datasets. By referring to Table 3.1, AxisBC and AyerKeroh sites are selected as testing datasets, these two sites will give 18 datasets, which is about 18% of the total available datasets.

Firstly, the optimized ANN model will be loaded from the saved files. This ANN model will be used to infer the optimal ISR value. The input of the ANN model is the monthly irradiation of the specific site and their respective monthly optimal ISR value simulated by satellite-derived datasets. Then, the percentage difference of predicted monthly optimal ISR value and actual monthly optimal ISR simulated with the ground-based datasets which is at co-located site is calculated. The mean square error of the inferred value and actual value is calculated.

3.10 Evaluation of ANN model performance

The percentage of difference, Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) is calculated in stage 4 and 5. These errors are recorded and analysed to evaluate the performance of ANN model used for inference of optimal ISR. The equations below showed the formula of errors:

$$Percentage\ of\ Difference = \frac{|Actual\ ISR - Predicted\ ISR|}{Actual\ ISR} \times 100\% \quad (3.18)$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(Actual\ ISR - Predicted\ ISR)^2 \quad (3.19)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Actual\ ISR - Predicted\ ISR)^2} \quad (3.20)$$

3.11 Milestone

In Figure *3.8*, the milestones of the project across two semesters are listed out.

Milestone	Activity Involved	Date of Completion
Preliminary investigation for problem statement	1.1	13/6/2021
Identify the Objectives and Scope of Project	1.2, 1.3	27/6/2021
Literature Review on current method to compute optimal ISR	2.1, 2.2	11/6/2021
Process the irradiance data	3.1, 3.2	13/7/2021
Build the program to compute LCOE, Optimal ISR and Monthly Irradiance	3.3	27/9/2021
Literature Review to build a non-linear regression ANN model	2.1, 2.2	23/12/2021
Build and Optimize the ANN Model	3.4	30/2/2022
Inference of Optimal ISR with ANN Model	3.5	28/3/2022
Evaluate the performance of ANN Model	3.5, 4	28/3/2022
Writing of Final Report	2.1, 2.2, 4	24/4/2022

Figure 3.8: Project Milestones

3.12 Gantt Chart

In Figure 3.9, the Gantt Chart for two semesters of this Final Year Project is shown.

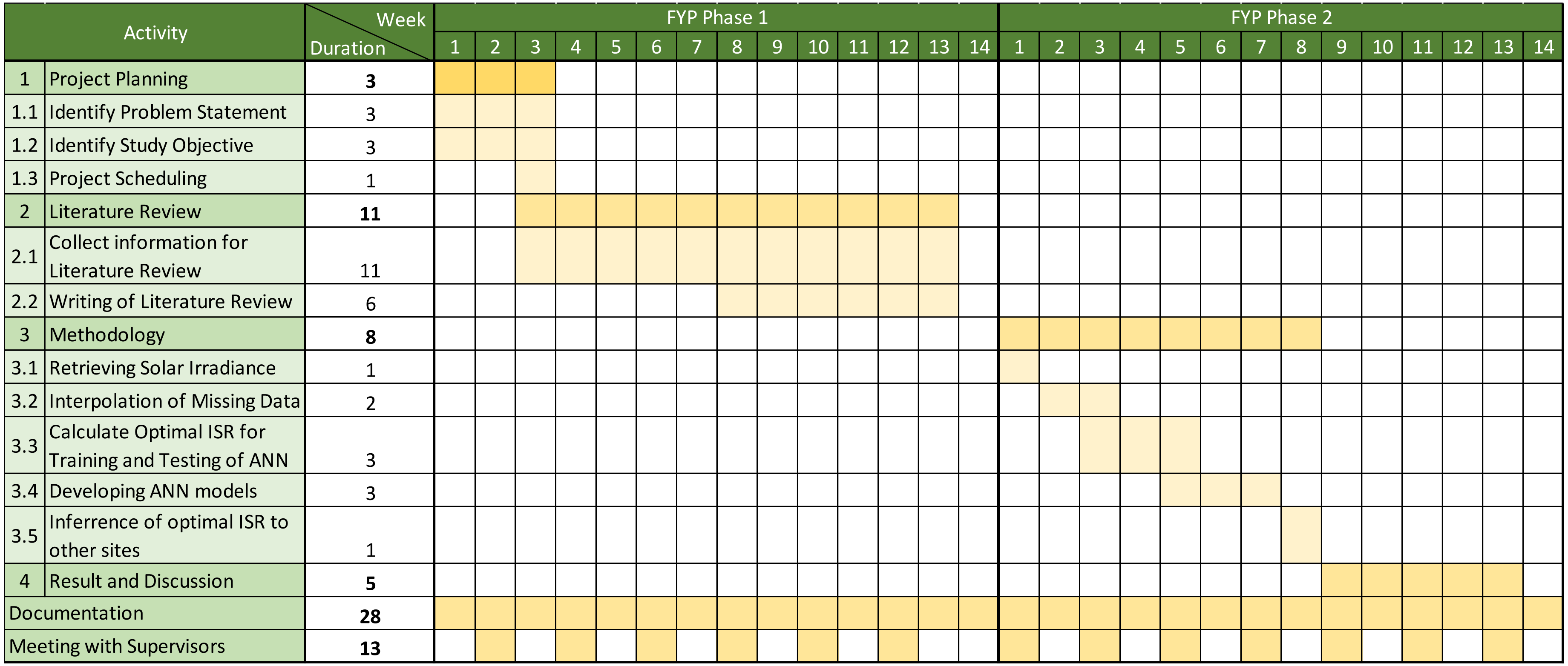

Figure 3.9: Gantt chart of the project.

CHAPTER 4

RESULTS AND DISCCUSSIONS

4.1 Introduction

In this chapter, the ISR values which are computed by ground-based measurement and satellite-derived datasets will be discussed to investigate the relationship of optimal ISR computed by ground-based measurement and satellite-derived datasets. Next, the solar irradiance distribution of the sites was plotted to investigate how the changing of solar irradiance distribution profile will impact the optimal ISR value. Then, the final ANN model built, and the results of inference are discussed. The performance of ANN model is evaluated with MSE, RMSE, and percentage of difference between the actual and predicted value.

4.2 Summary of irradiance datasets from available sites

Table *4.1* showed the summary of solar irradiance datasets from different sites. Most of the sites are located in Selangor area. Two of the sites are selected as testing datasets, which are AxisBC site and AyerKeroh site. The remaining sites will be categorized as training datasets.

To be highlighted that, the minimum number of days per month is 27 days. Some of the sites have problem of missing data in their monthly solar irradiance data due to malfunction of weather station equipment. The monthly datasets with less than 27 days per month will be removed. Lack of irradiance data due to missing day will cause the increase of optimal ISR value. The detailed explanation will be further elaborated in Section 4.3.

In addition, the type of the irradiance sensor which is recorded in Table *4.1* is to check how the type of sensor can affect the ground-based measurement solar irradiance data. This information will be useful in the analysis of result in Section 4.3.

Table 4.1: Summary information of the sites which provide the irradiance datasets to use in project. (The highlighted row is the sites used in inference of optimal ISR)

No.	Name	Location	Number of Month Available (>27 days)	Type of Irradiance Sensor
1	UTAR	Selangor	12	Pyranometer
2	Bukit Kayu Hitam (BKH)	Kedah	12	Pyranometer
3	AyerKeroh	Melaka	6	Pyranometer
4	Permatang	Penang	10	Pyranometer
5	AxisBC	Selangor	12	Photodiode Sensor
6	Balakong	Selangor	11	Photodiode Sensor
7	LFLogistic	Selangor	10	Photodiode Sensor
8	StarkenACC	Selangor	9	Photodiode Sensor
9	TelePower	Selangor	7	Photodiode Sensor
10	SRTech	Selangor	10	Photodiode Sensor

4.3 Comparison between optimal ISR computed from ground-based measurement and satellite-derived irradiance datasets.

Figure 4.1 to Figure 4.10 has shown the monthly optimal ISR computed by using ground-based measurement and satellite-derived irradiance datasets of each sites. The first 3 sites, which are UTAR, BKH and AyerKeroh, have overall lower monthly optimal ISR computed with ground-based measurement (Ground ISR) as compared to their monthly optimal ISR computed with satellite-derived irradiance datasets (Satellite ISR). In contrast, the rest of the 7 sites have higher Ground ISR as compared to Satellite ISR.

ISR usually have smaller value when the irradiance datasets have more irradiation with high solar irradiance (higher than 1000 W/m^2). This is because, the clipping of power generation during event of high irradiance will result to design a smaller size of DC capacity in a PV plant, which lead to smaller ISR value. However, normally the ground-based measurement which have higher time resolution irradiance data will have more solar irradiation in solar irradiance range as compared to satellite-derived irradiance data. This is because satellite-derived irradiance datasets have lower time resolution which can cause some missing of the peak irradiance value during that time interval. Thus, the Satellite ISR should be higher than Ground ISR in most of the case. By referring to the graph of monthly optimal ISR computed by using ground-based measurement and satellite-derived irradiance datasets, 7 out of 10 sites do not obey the statement. The 7 sites are Permatang, AxisBC, Balakong, LFLogistic, StarkenACC, TelePower and SRTech. Therefore, the solar irradiance distribution of the site is investigated in the next paragraph.

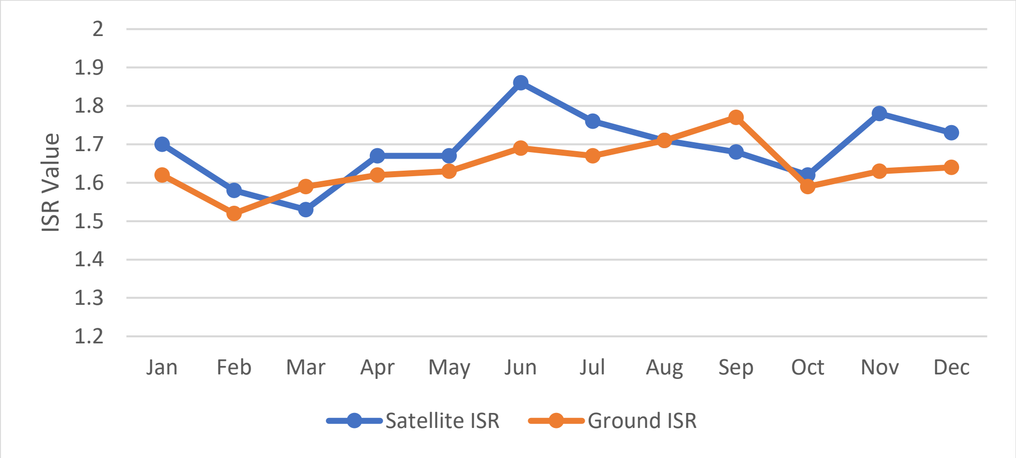

Figure 4.1: UTAR monthly optimal ISR computed by using ground-based measurement and satellite-derived irradiance datasets.

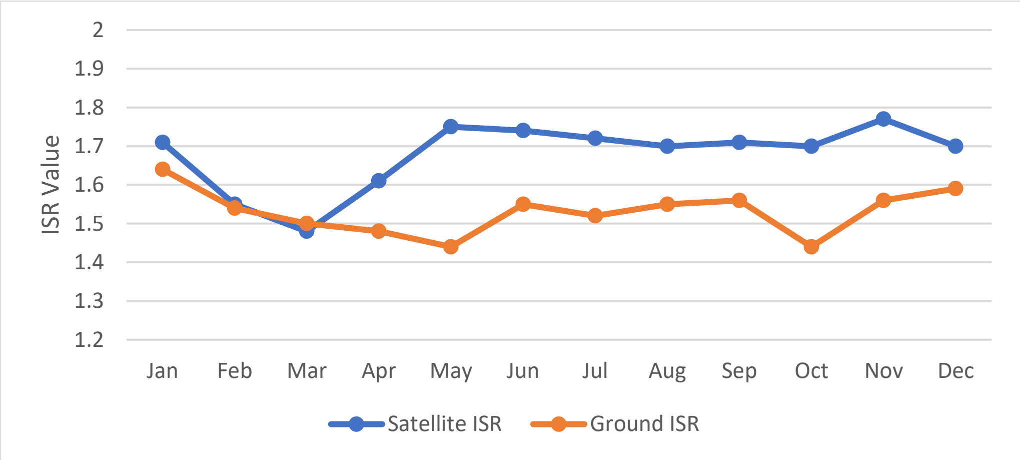

Figure 4.2: BKH monthly optimal ISR computed by using ground-based measurement and satellite-derived irradiance datasets.

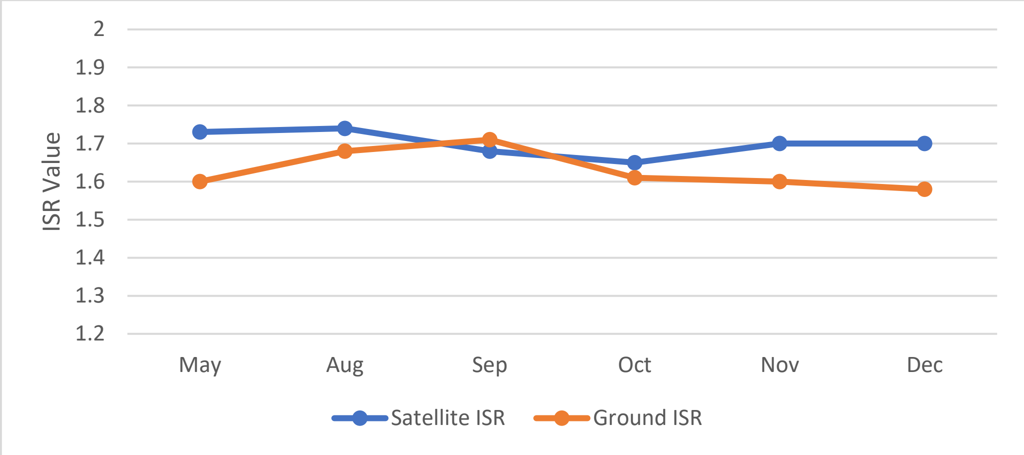

Figure 4.3: AyerKeroh monthly optimal ISR computed by using ground-based measurement and satellite-derived irradiance datasets.

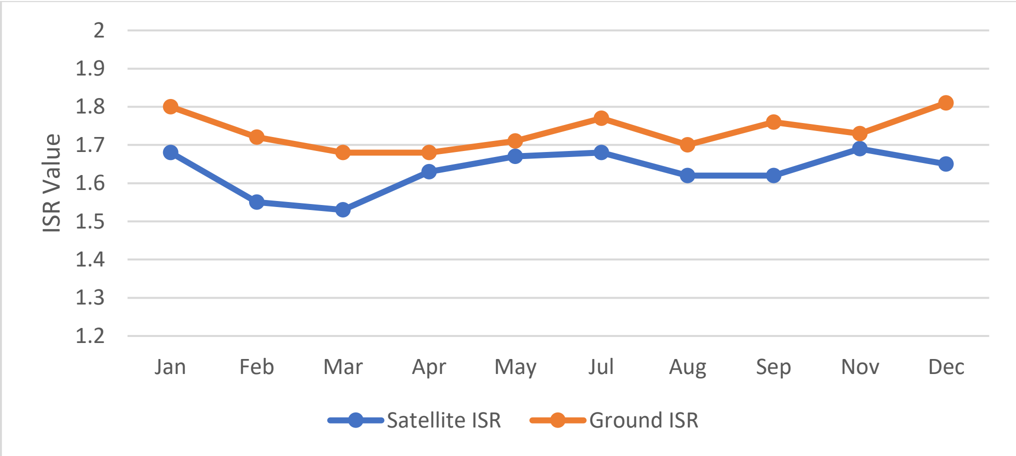

Figure 4.4: Permatang monthly optimal ISR computed by using ground-based measurement and satellite-derived irradiance datasets.

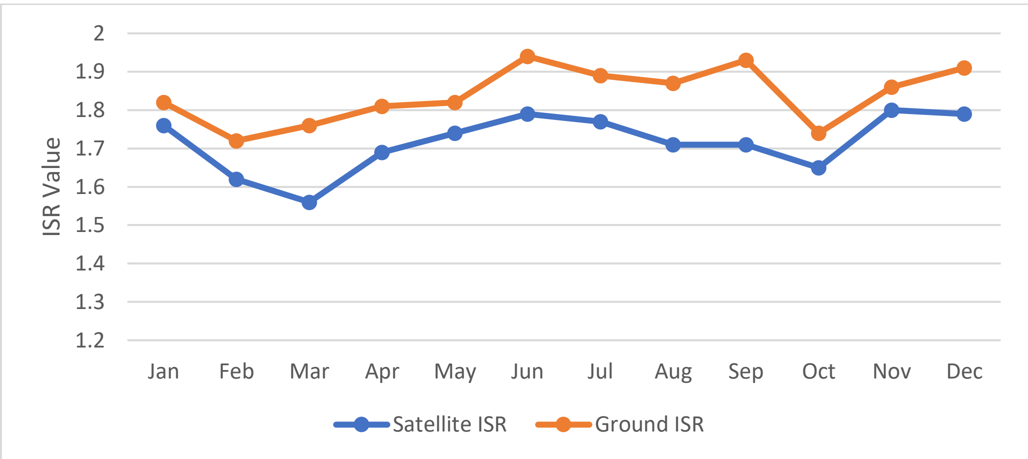

Figure 4.5: AxisBC monthly optimal ISR computed by using ground-based measurement and satellite-derived irradiance datasets.

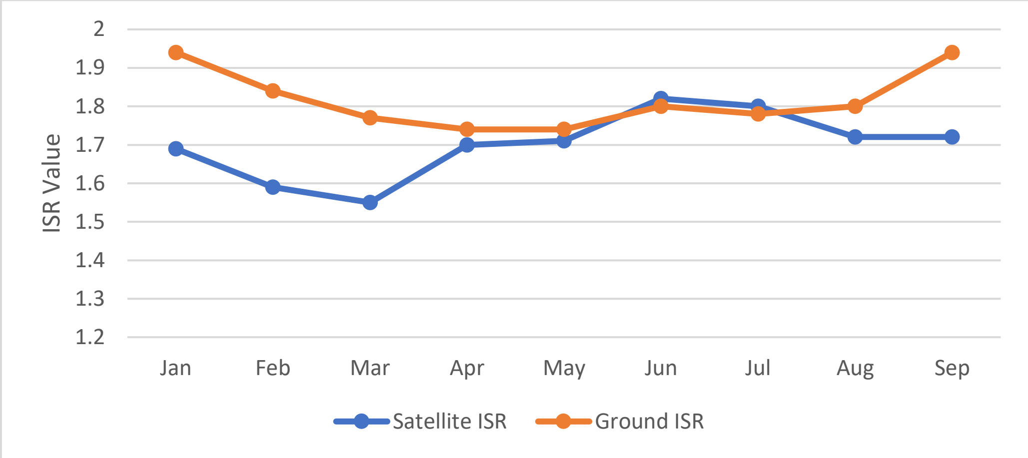

Figure 4.6: Balakong monthly optimal ISR computed by using ground-based measurement and satellite-derived irradiance datasets.

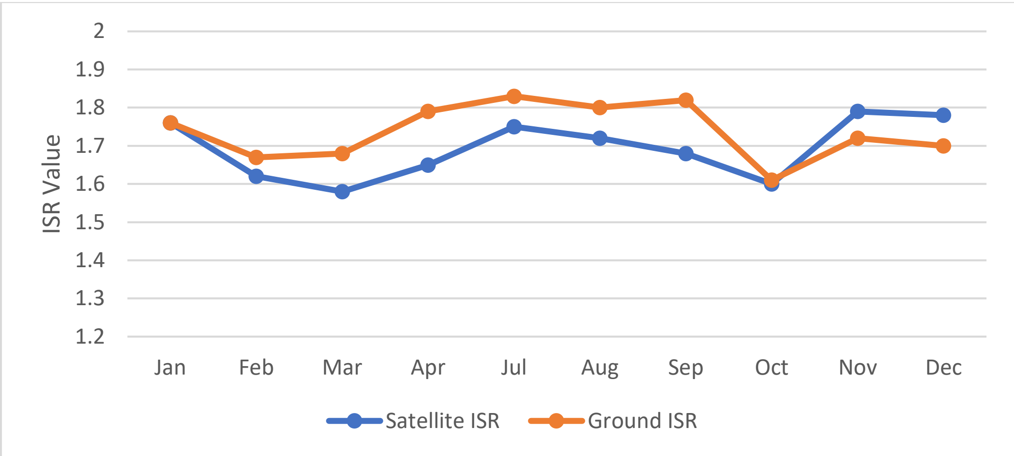

Figure 4.7: LFLogistic monthly optimal ISR computed by using ground-based measurement and satellite-derived irradiance datasets.

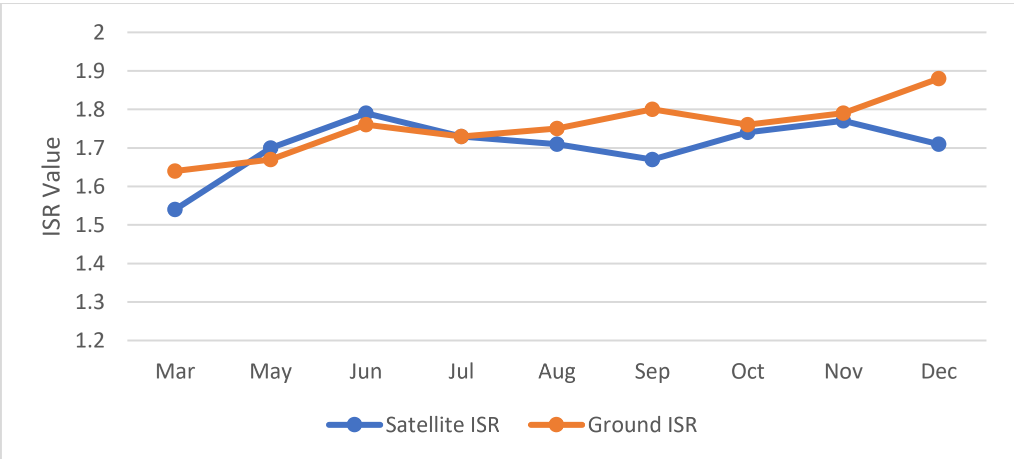

Figure 4.8: StarkenACC monthly optimal ISR computed by using ground-based measurement and satellite-derived irradiance datasets.

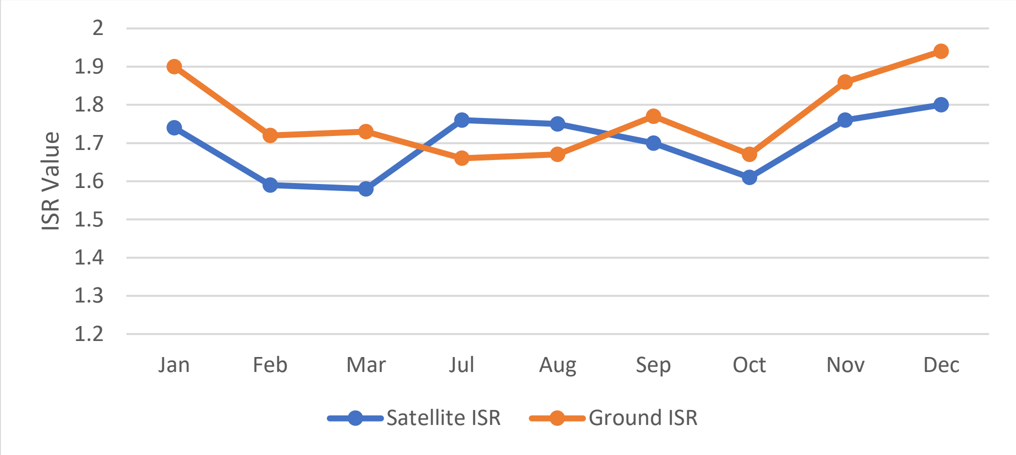

Figure 4.9: TelePower monthly optimal ISR computed by using ground-based measurement and satellite-derived irradiance datasets.

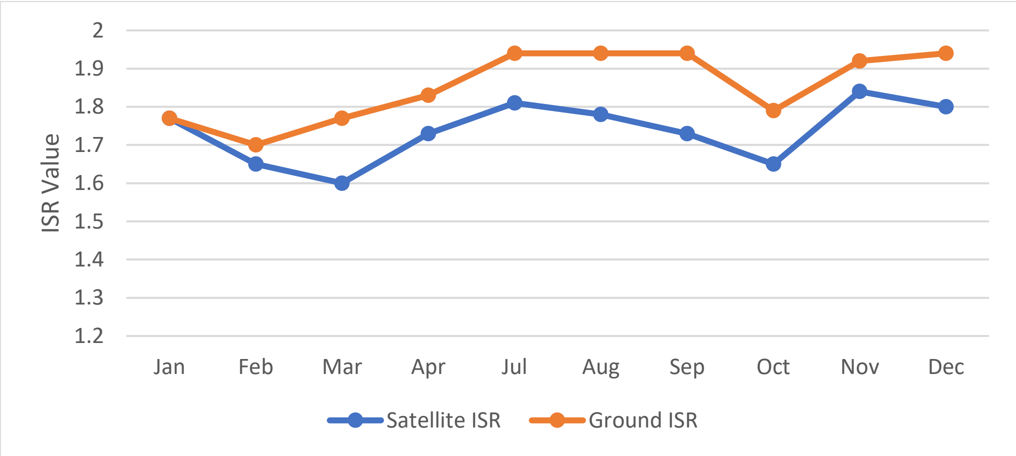

Figure 4.10: SRTech monthly optimal ISR computed by using ground-based measurement and satellite-derived irradiance datasets.

The figures below show the solar irradiance distribution profile in 2019 of some selected sites. Figure 4.11 and Figure 4.12 are UTAR and BKH sites, where their Satellite ISR is higher than Ground ISR and vice versa for next two figures, Figure 4.13 and Figure 4.14 which are from AxisBC and Balakong sites. By taking UTAR and BKH sites as reference of common case where the Satellite ISR is higher than Ground ISR, it can be observed that the irradiation in high irradiance range of ground-based measurement datasets is lesser than the satellite-derived datasets in the solar irradiance distribution of AxisBC and Balakong sites. This situation leads to the situation where the Ground ISR in AxisBC and Balakong sites are higher than their Satellite ISR occurred.

By referring to the solar irradiance distribution graph of AxisBC and Balakong, it can be observed that the irradiation of high solar irradiance in ground-based measurement datasets have value near to zero. At the same time, the solar irradiance sensor used in AxisBC and Balakong sites are both Photodiode sensor, which is widely used in agriculture field. The Photodiode type of irradiance sensor have lower accuracy and might have different working range of irradiance as compared to the Silicon Solar Radiation type of sensor. The actual value of Ground ISR among the sites with Photodiode irradiance sensor should be lower than the current computed value.

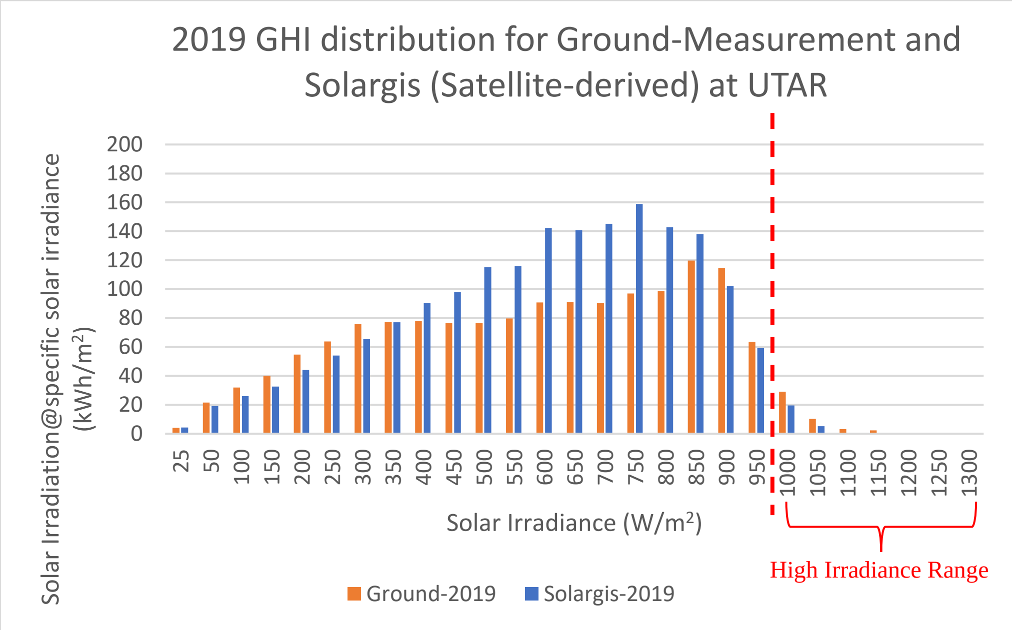

Figure 4.11 Solar Irradiance distribution for ground-based measurement and satellite-derived irradiance datasets at UTAR site.

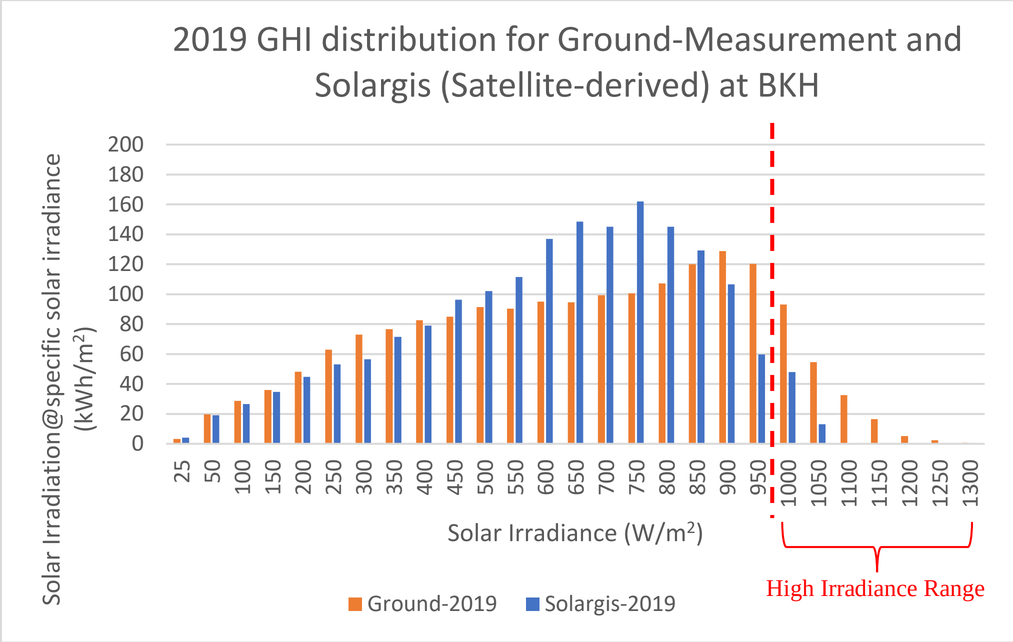

Figure 4.12: Solar Irradiance distribution for ground-based measurement and satellite-derived irradiance datasets at BKH site.

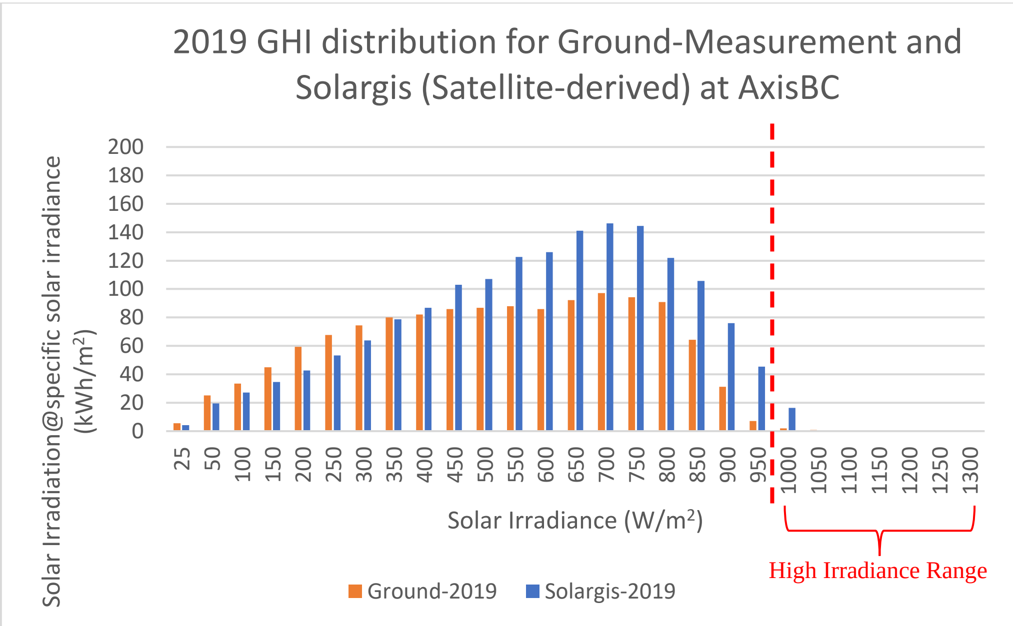

Figure 4.13: Solar Irradiance distribution for ground-based measurement and satellite-derived irradiance datasets at AxisBC site.

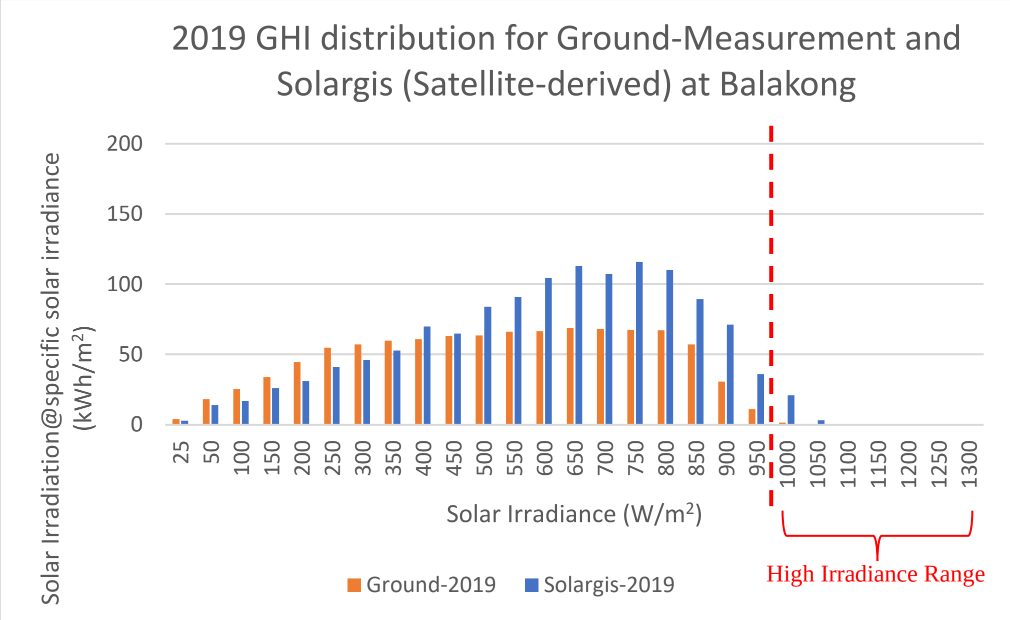

Figure 4.14: Solar Irradiance distribution for ground-based measurement and satellite-derived irradiance datasets at Balakong site.

4.4 Final ANN model after the optimization is done.

By referring to the Table 3.4, the parameters are changed within the specified range while the percentage of difference between predicted and actual value of Ground ISR by using the training datasets is recorded. Only the parameter combinations which have percentage difference less than 6% and without overfitting will be considered as parameters used in final ANN model. At the end of optimization, the parameter combination with lowest percentage of difference will be selected as final ANN model, which will be used to infer the Ground ISR of other sites which are under testing datasets category. Table 4.2 showed the final parameters of optimized ANN model which give smallest percentage of difference and have step loss and validation loss converging to zero.

Table 4.2: The final parameters of ANN model selected as optimized model.

Parameters of ANN	Values after optimization of ANN model
Number of Hidden Layer	1
Number of Hidden Neuron	4
Epochs	1300
Batch Size	2

Figure 4.15 shows the change of step loss and validation loss when the epoch of training increase from 0 to 1300. The step loss and validation loss are converged to a small constant value after 200 epoch and keep fluctuating until the pre-set 1300 epoch. However, it can be observed that during the 1300 epoch, the losses decrease. The Table 4.3 is the final losses and error of the ANN model when the epoch is 1300.

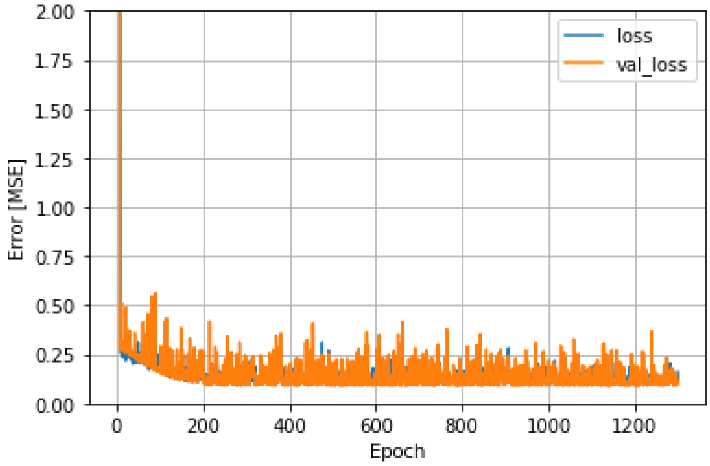

Figure 4.15: The graph of step loss and validation loss along the pre-set epochs.

Table 4.3: The final losses and error of the ANN model developed when the epoch is 1300.

Step loss	0.1608
Mean squared error	0.0421
Validation loss	0.1059
Validation mean squared error	0.0156

4.5 Comparison between inferred optimal ISR and the actual optimal ISR value.

In this section, the result of prediction by using both training datasets and testing datasets will be showed and each of their predicted and actual value will be compared. The percentage of difference and the mean square error of predicted and actual value will be calculated.

Figure 4.16 and the table in Appendix C record the predicted and actual value of Ground ISR using training datasets. The reason that training datasets has been reused to predict the Ground ISR is to verify the function and reliability of the ANN model before proceeding to do the inference of Ground ISR to other unknown sites. An ANN model which can function well should be able to give a good prediction with the datasets which is used to trained it. From the table in Appendix C, the average percentage of difference between prediction and actual Ground ISR value is 5.6%, which is less than 6%. The 6% of difference is the average percentage of difference between original value of Satellite ISR (one of the ANN inputs) and Ground ISR (the value that ANN designed to predict). If the average percentage of difference between prediction and actual Ground ISR value is less than 6%, it means that the trained ANN model is functioning and perform well to predict Ground ISR when the Satellite ISR is used as ANN input data. In other words, the difference between Satellite ISR and its corresponding Ground ISR is reduced.

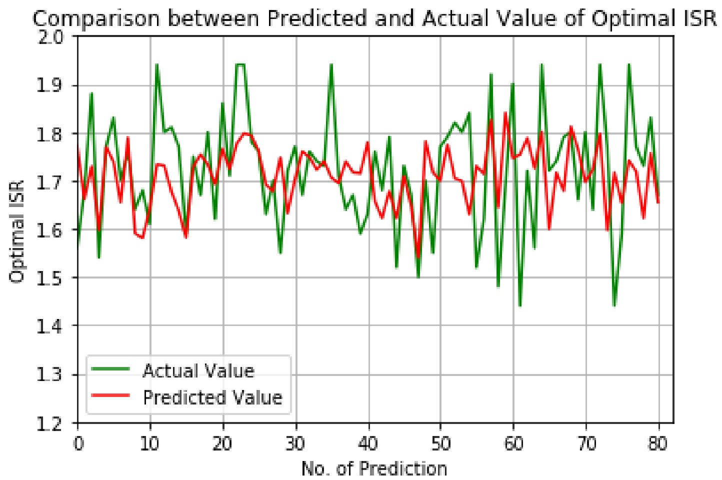

Figure 4.16: The comparison between predicted and actual value of Ground ISR by using the training datasets after the ANN training was done.

Figure 4.17, Figure 4.18 and Table 4.4 show the graphs and results of the prediction of the ANN model. From the Table 4.4, the average percentage of difference between prediction and actual Ground ISR value by using testing datasets is 5.8%, which is less than 6%. It means that the trained ANN model is functioning and perform well to predict Ground ISR and the difference between Satellite ISR and its corresponding Ground ISR is reduced. Furthermore, by referring to the table in Appendix B, the percentage of different of actual Satellite ISR and Ground ISR is varied from range of 21.5% to 0% whereas the percentage of difference of actual Ground ISR and Predicted Ground ISR using testing datasets is reduced to range of 10.6% to 0.5%. The range of percentage of difference is reduced with the application of ANN model developed to predict the Ground ISR of a specific site.

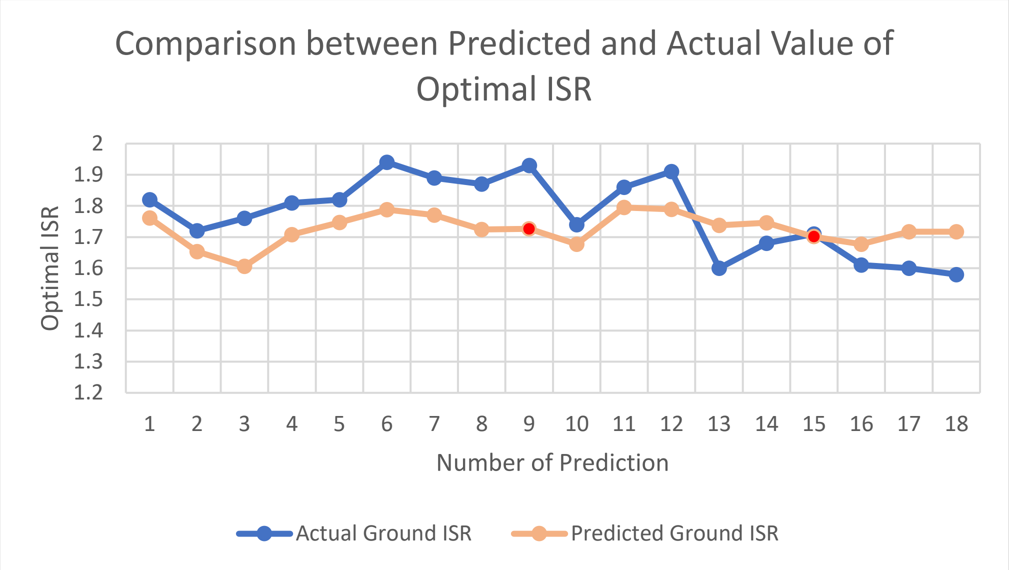

Figure 4.17: The comparison between predicted and actual value of Ground ISR by using the testing datasets. (The highlighted red spot showed the optimal ISR with maximum and minimum percentage of difference.)

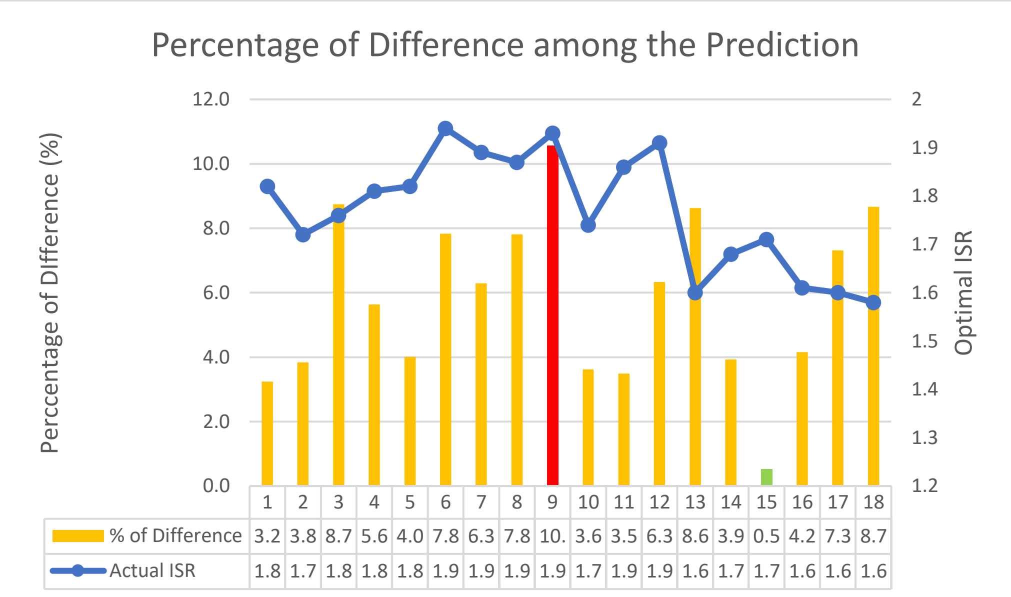

Figure 4.18: Percentage of difference between predicted and actual Ground ISR among the prediction. (The red bar is the maximum value whereas the green bar is the minimum value of percentage of difference.)

Figure 4.17, Figure 4.18 and Table 4.4 show the graphs and results of the prediction of the ANN model. From the Table 4.4, the average percentage of difference between prediction and actual Ground ISR value by using testing datasets is 5.8%, which is less than 6%. It means that the trained ANN model is functioning and perform well to predict Ground ISR and the difference between Satellite ISR and its corresponding Ground ISR is reduced. Furthermore, by referring to the table in Appendix B, the percentage of different of actual Satellite ISR and Ground ISR is varied from range of 21.5% to 0% whereas the percentage of difference of actual Ground ISR and Predicted Ground ISR using testing datasets is reduced to range of 10.6% to 0.5%. The range of percentage of difference is reduced with the application of optimized ANN model to predict the Ground ISR of a specific site.

Table 4.4: Results of ANN model predicting Ground ISR and comparison with actual value. (The highlighted row showed the results with maximum and minimum percentage of difference.)

Name	No. of Prediction	Monthly Irradiance (W/m^2)	Actual ISR	Predicted ISR	Percentage of Difference (%)
AxisBC (Selangor)	1	147.560	1.82	1.761	3.2
	2	150.531	1.72	1.654	3.8
	3	169.055	1.76	1.606	8.7
	4	145.265	1.81	1.708	5.6
	5	138.095	1.82	1.747	4.0
	6	118.724	1.94	1.788	7.8
	7	132.487	1.89	1.771	6.3
	8	143.865	1.87	1.724	7.8
	9	128.192	1.93	1.726	10.6
	10	149.169	1.74	1.677	3.6
	11	126.747	1.86	1.795	3.5
	12	118.148	1.91	1.789	6.3
AyerKeroh	13	149.346	1.6	1.738	8.6

(Melaka)	14	146.112	1.68	1.746	3.9
	15	145.854	1.71	1.701	0.5
	16	149.274	1.61	1.677	4.2
	17	140.061	1.6	1.717	7.3
	18	137.219	1.58	1.717	8.7
		(Average)			5.8
		Mean Squared Error			0.0128
		Root Mean Squared Error			0.1132

Moreover, in the Table 4.5, the Mean Square Error (MSE), Root Mean Square Error (RMSE) and average percentage of difference are listed out to evaluate the ANN model developed. It can be observed that the MSE and RMSE of the actual and predicted Ground ISR which are computed by training datasets and testing datasets has reduced as compared to the errors of actual Satellite ISR and Ground ISR. Thus, it can be concluded that, the ANN model developed is able to use Satellite ISR with monthly irradiance of the sites to infer Ground ISR where the sites do not have ground-based measurement solar irradiance datasets.

Table 4.5: Summary of errors before and after prediction of Ground ISR using ANN model developed.

	Original Satellite ISR and Ground ISR	Actual Ground ISR and Predicted Ground ISR using training datasets	Actual Ground ISR and Inferred Ground ISR using testing datasets
MSE	0.0152	0.0136	0.0128
RMSE	0.1234	0.1165	0.1132
Average Percentage of Difference (%)	6	5.6	5.8

4.6 Suggestions to improve the performance of ANN model based on the analysis done.

By referring to the Table 4.5, the errors reduce where the inferred Ground ISR value is closer to the actual Ground ISR as compared to original Satellite ISR value. However, the decrement of errors is small, and the performance of ANN model is not great enough to support the accuracy of the inferred optimal ISR.

In Section 4.3, it can be concluded that type of sensor used to measure solar irradiance will affect the optimal ISR value. Photodiode sensor which has lower accuracy will clip the receiving of irradiation of high irradiance (>1000 Wh/m^2). This can cause the overestimation of optimal ISR. Thus, it is suggested that the weather station should adapt the irradiance sensor with high accuracy such as pyranometer and silicon solar radiation sensor. To be highlighted that, the sensor must be installed without any self-shading or shaded by other building around it. The Figure 4.19 below showed the wrong installation of weather station. The type of sensor used in the Figure 4.19 is photodiode sensor which has low accuracy. The yellow circle showed the shaded region of PV panel cause by weather station. Moreover, it can be observed that the anemometer at the top of weather station caused shading to the irradiance sensor.

Figure 4.19: An example of wrong installation of weather station.

Moreover, by referring to the Table 4.1, there are 7 out of 10 sites are located at Selangor region. To increase the reliability of ANN model, the geographical location of sites to measure and collect solar irradiance datasets should be distantly distributed in Malaysia. Distantly distributed sites can collect different pattern of annual or monthly solar irradiation. Thus, the ANN model built can be trained with more pattern of solar irradiance distribution and improve its performance when it is used to infer optimal ISR of other sites.

Lastly, the total number of datasets should be increased. A total number of 99 datasets is insufficient for ANN model to draw a conclusion about the relationship between Ground ISR and Satellite ISR value. By referring to the ANN model working principle and explanation of inference in Section 2.5 and 2.6, greater number of datasets could enhance the accuracy to infer the optimal ISR value.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

Both objectives stated in the study were achieved. The relationship between optimal ISR computed by ground-based measurement and satellite-derived irradiance datasets was investigated with the type of irradiance sensor used and the irradiance distribution profile of the 10 sites. In short, the decrement of the time resolution of irradiance data could cause the overestimation of optimal ISR due to the clipped of energy during event of high irradiance. Moreover, the accuracy of the irradiance sensor also could clip the high irradiance and caused the optimal ISR to be overestimated.

Next, the second objective is to infer the optimal ISR of other sites with satellite-derived irradiance datasets. This objective was accomplished by using ANN model to infer optimal ISR to other sites. The inferred optimal ISR have lower errors as compared to the optimal ISR simulated by satellite-derived irradiance datasets. However, being an exploratory study, the result of inference is not satisfied. One of the issues was the insufficiency on decrement of error could not ensure the reliability and accuracy of ANN model in real life. By referring to the Table *4.5*, the percentage of difference is reduced from 6% to 5.8%, MSE is reduced by 0.0024 (-15.8%) and the RMSE is reduced by 0.0102 (-8.3%) only. The reduction of error is too small. Thus, it is suggested to increase the total number of datasets for training and testing ANN model. Therefore, the performance of ANN model will improve. Furthermore, it was suggested that the geographical location of the sites should be distributed evenly in Malaysia region instead of focusing on a specific location and infer the optimal ISR of different region.

5.2 Recommendation

It is recommended that the future research should attempt to consider other parameters which are related to the ground-based measurement sites weather information such as monthly irradiance data, humidity, and other related information. In addition, the future work needs to understand and identify the mechanism involved in the satellite-derived datasets. Moreover, the impacts of missing data and interpolated datasets to the optimal ISR simulation should be verify in the future research.

REFERENCES

Burns, E. and Burke, J., 2021. *What is a neural network? Explanation and examples.* [online] Available at: <https://searchenterpriseai.techtarget.com/definition/neural-network> [Accessed 29 August 2021].

Camm, J. et al., 2015. *Essentials of Business Analytics*. Stamford: Cengage Learning.

CFI, n.d. *Levelized Cost of Energy (LCOE)*. [online] Available at <https://corporatefinanceinstitute.com/resources/knowledge/finance/levelized-cost-of-energy-lcoe/> [Accessed 31 August 2021]

David, B., 2021. *Choosing the Right Size Solar Inverter*. [online] Available at: <https://www.aurorasolar.com/blog/choosing-the-right-size-inverter-for-your-solar-design-a-primer-on-inverter-clipping/> [Accessed 23 August 2021].

Dormehl, L., 2019. *What is an artificial neural network? Here's everything you need to know*. [online] Available at: <https://www.digitaltrends.com/cool-tech/what-is-an-artificial-neural-network/> [Accessed 28 August 2021].

European Commission, 2021. *Photovoltaic Geographical Information System (PVGIS)*. [online] Available at: < https://ec.europa.eu/jrc/en/pvgis> [Accessed 23 August 2021].

Fisherdong, 2011. *Positive and Negative Correlation*. [online] Available at: < https://fisherdong.wordpress.com/2011/03/25/positive-and-negative-correlation/ > [Accessed 3 September 2021]

JavaTpoint, n.d. *Artificial Neural Network Tutorial.* [online] Available at: <https://www.javatpoint.com/artificial-neural-network> [Accessed 29 August 2021].

Lai, K.Y. and Lim, B.H., 2019. *Optimal Inverter Sizing Ratio for Photovoltaic Power Plants in Malaysia*. In: Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, International Conference on Applied Energy 2019. Västerås, Sweden, 12-15 August 2019. Sweden: Energy Proceedings.

Lai, K.Y. and Lim, B.H., 2019. *Comparative Study for Time-specific Ross Coefficient and Overall Ross Coefficient for Estimation of Photovoltaic Module Temperature*. In: Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, 2019 IEEE Conference on Sustainable Utilization and Development in Engineering and Technologies (CSUDET). Penang, Malaysia, 7-9 November 2019. Malaysia: IEEE.

Lim, W. J. L., 2021. *Inverter Sizing Ratio for PV plant in the Tropics*. Degree. Universiti Tunku Abdul Rahman. Available at: < http://eprints.utar.edu.my/4062/> [Accessed 29 August 2021]

Mamum, M. A. A., Islam, M. M. and Hasanuzzaman, M., 2021. Effect of tilt angle on the performance and electrical parameters of a PV module: Comparative indoor and outdoor experimental investigation. *Energy and Built Environment*, [online] Available at: < https://www.sciencedirect.com/science/article/pii/S2666123321000179> [Accessed 29 August 2021].

Mesquita, D., 2021. *Python AI: How to Build a Neural Network & Make Predictions*, [online] Available at: < https://realpython.com/python-ai-neural-network/#computing-the-prediction-error> [Accessed 29 August 2021]

Miller, A. and Lumby, B., 2012. *Utility Scale Solar Power Plants. A guide for developers and investors.* [e-book] New Delhi: International Finance Corporation (IFC) World Bank Group. Available at: UNECE Wiki < https://wiki.unece.org/download/attachments/25267247/SOLAR%2BGUIDE%2BBOOK.pdf?api=v2> [Accessed 29 August 2021].

Mohamad, N. B., Lai, A. C. and Lim, B. H. (2022) 'A case study in the tropical region to evaluate univariate imputation methods for solar irradiance data with different weather types', *Sustainable Energy Technologies and Assessments*, 50(October 2021), p. 101764. doi: 10.1016/j.seta.2021.101764.

Nielsen, M., 2019. *Chapter 1: Using neural nets to recognize handwritten digits* [online] Available at: < http://neuralnetworksanddeeplearning.com/chap1.html > [Accessed 29 August 2021]

Scikit-learn, n.d. *1.17. Neural network models (supervised)* [online] Available at: < https://scikit-learn.org/stable/modules/neural_networks_supervised.html > [Accessed 30 August 2021]

Solar Power Co., 2019. *What does Oversizing a Solar Power System Mean?* [online] Available at: <https://solarpowerco.com.au/what-does-oversizing-a-solar-power-system-mean/> [Accessed 28 August 2021].

Solargis, 2021. *Download solar resource maps and GIS data for 200+ countries and regions*. [online] Available at: <https://solargis.com/maps-and-gis-data/download/world> [Accessed 24 August 2021].

Solcast, n.d. *Compare Solcast historical and TMY products to PVGIS*. [online] Available at: < https://solcast.com/historical-and-tmy/pvgis/> [Accessed 3 September 2021]

U.S. Energy Information Administration, 2021. *Levelized Costs of New Generation Resources in the Annual Energy Outlook 2021*, [online] Available at: < https://www.eia.gov/outlooks/aeo/pdf/electricity_generation.pdf> [Accessed 31 August 2021]

Zidane, T., Zali, S. and Adzman, M., 2020. *PV array and inverter optimum sizing for grid-connected photovoltaic power plants using optimization design*, [online] Available at: <https://iopscience.iop.org/article/10.1088/1742-6596/1878/1/012015#references> [Accessed 24 August 2021].

APPENDICES

Appendix A: Python code for interpolation

```
import pandas as pd
import numpy as np
import glob

# get the filename in the data folder provided----------------
            ----------------
case =glob.glob('BZ_Full/*.csv')

# declare variable--------------------------------------------
            ----------------
sum_MAE_linear, sum_RMSE_linear, sum_MAE_cubic, sum_RMSE_cubic
            = 0,0,0,0
sum_MAE_spline, sum_RMSE_spline, sum_MAE_nearest,
            sum_RMSE_nearest = 0,0,0,0


# read file---------------------------------------------------
            ----------------
for j in range (len(case)):
    df = pd.read_csv(case[j])
    df['test_data'] = np.nan

# to extract the data set with missing part and fill it with
            NaN---------------
    for i in range(len(df)):
        if df['Missing'][i] == 1:
            df['test_data'][i] = np.nan
        else:
              df['test_data'][i] = df['Actual'][i]

#Interpolation part-------------------------------------------
            ----------------
    linear = df['test_data'].interpolate(method='linear')
    cubic = df['test_data'].interpolate(method='cubic')
    spline = df['test_data'].interpolate(method='spline',
            order=3, s=0.)
    nearest = df['test_data'].interpolate(method='nearest')
    result = pd.concat([linear,cubic, spline,nearest], axis=1)
    result.columns = ['linear', 'cubic', 'spline','nearest']
    final = pd.concat([df, result],axis = 1)
```

```
# check error --------------------------------------------------
            -----------------
    diff_linear = abs(final['Actual'] - final['linear'])
    mae_linear = diff_linear.mean()
    sum_MAE_linear += mae_linear
    rmse_linear = np.sqrt((diff_linear ** 2).mean())
    sum_RMSE_linear += rmse_linear

    diff_cubic = abs(final['Actual'] - final['cubic'])
    mae_cubic = diff_cubic.mean()
    sum_MAE_cubic += mae_cubic
    rmse_cubic = np.sqrt((diff_cubic ** 2).mean())
    sum_RMSE_cubic += rmse_cubic

    diff_spline = abs(final['Actual'] - final['spline'])
    mae_spline = diff_spline.mean()
    sum_MAE_spline += mae_spline
    rmse_spline = np.sqrt((diff_spline ** 2).mean())
    sum_RMSE_spline += rmse_spline

    diff_nearest = abs(final['Actual'] - final['nearest'])
    mae_nearest = diff_nearest.mean()
    sum_MAE_nearest += mae_nearest
    rmse_nearest = np.sqrt((diff_nearest ** 2).mean())
    sum_RMSE_nearest += rmse_nearest

# export the output file---------------------------------------
            -----------------
    s = case[j]
    start = s.find('\\') + 1
    end = s.find('.csv', start)
    dest = 'various type of interpolation result/' +
            s[start:end] +'.csv'
    final.to_csv(dest,index = False)

# print out the sum of error for comparison-------------------
            -----------------
print ('sum of MAE for linear interpolation is: ' +
            str(sum_MAE_linear))
print ('sum of MAE for cubic interpolation is: ' +
            str(sum_MAE_cubic))
print ('sum of MAE for spline interpolation is: ' +
            str(sum_MAE_spline))
print ('sum of MAE for nearest interpolation is: ' +
            str(sum_MAE_nearest))
```

```
print ('sum of RMSE for linear interpolation is: ' +
           str(sum_RMSE_linear))
print ('sum of RMSE for cubic interpolation is: ' +
           str(sum_RMSE_cubic))
print ('sum of RMSE for spline interpolation is: ' +
           str(sum_RMSE_spline))
print ('sum of RMSE for nearest interpolation is: ' +
           str(sum_RMSE_nearest))
```

Appendix B: Monthly information of each site with the percentage of difference and errors between Satellite ISR and Ground ISR.

Name	Month	Monthly Irradiance (W/m^2)	Satellite ISR	Ground ISR	Percentage of Difference (%)
UTAR	Jan	159.455	1.7	1.62	4.9
	Feb	163.289	1.58	1.52	3.9
	Mar	179.636	1.53	1.59	3.8
	Apr	148.524	1.67	1.62	3.1
	May	157.729	1.67	1.63	2.5
	Jun	124.646	1.86	1.69	10.1
	Jul	150.777	1.76	1.67	5.4
	Aug	142.825	1.71	1.71	0.0
	Sep	141.406	1.68	1.77	5.1
	Oct	156.250	1.62	1.59	1.9
	Nov	131.473	1.78	1.63	9.2
	Dec	139.928	1.73	1.64	5.5
BKH	Jan	153.090	1.71	1.64	4.3
	Feb	171.546	1.55	1.54	0.6
	Mar	200.463	1.48	1.5	1.3
	Apr	166.580	1.61	1.48	8.8
	May	145.065	1.75	1.44	21.5
	Jun	137.105	1.74	1.55	12.3
	Jul	152.566	1.72	1.52	13.2
	Aug	140.744	1.7	1.55	9.7
	Sep	136.345	1.71	1.56	9.6
	Oct	142.986	1.7	1.44	18.1

	Nov	129.726	1.77	1.56	13.5
	Dec	151.112	1.7	1.59	6.9
Permatang	Jan	154.350	1.68	1.8	6.7
	Feb	158.608	1.55	1.72	9.9
	Mar	186.405	1.53	1.68	8.9
	Apr	155.176	1.63	1.68	3.0
	May	137.063	1.67	1.71	2.3
	Jul	152.100	1.68	1.77	5.1
	Aug	152.072	1.62	1.7	4.7
	Sep	136.219	1.62	1.76	8.0
	Nov	140.880	1.69	1.73	2.3
	Dec	153.225	1.65	1.81	8.8
AxisBC	Jan	147.560	1.76	1.82	3.3
	Feb	150.531	1.62	1.72	5.8
	Mar	169.055	1.56	1.76	11.4
	Apr	145.265	1.69	1.81	6.6
	May	138.095	1.74	1.82	4.4
	Jun	118.724	1.79	1.94	7.7
	Jul	132.487	1.77	1.89	6.3
	Aug	143.865	1.71	1.87	8.6
	Sep	128.192	1.71	1.93	11.4
	Oct	149.169	1.65	1.74	5.2
	Nov	126.747	1.8	1.86	3.2
	Dec	118.148	1.79	1.91	6.3
Balakong	Jan	157.605	1.69	1.94	12.9
	Feb	159.980	1.59	1.84	13.6
	Mar	176.322	1.55	1.77	12.4
	Apr	146.649	1.7	1.74	2.3
	May	154.117	1.71	1.74	1.7
	Jun	115.236	1.82	1.8	1.1
	Jul	138.811	1.8	1.78	1.1
	Aug	133.948	1.72	1.8	4.4
	Sep	129.913	1.72	1.94	11.3

LFLogistic	Jan	144.289	1.76	1.76	0.0
	Feb	150.218	1.62	1.67	3.0
	Mar	164.812	1.58	1.68	6.0
	Apr	142.026	1.65	1.79	7.8
	Jul	132.052	1.75	1.83	4.4
	Aug	146.864	1.72	1.8	4.4
	Sep	123.502	1.68	1.82	7.7
	Oct	143.602	1.6	1.61	0.6
	Nov	125.044	1.79	1.72	4.1
	Dec	114.788	1.78	1.7	4.7
Starken ACC	Mar	170.379	1.54	1.64	6.1
	May	139.829	1.7	1.67	1.8
	Jun	119.143	1.79	1.76	1.7
	Jul	139.028	1.73	1.73	0.0
	Aug	117.369	1.71	1.75	2.3
	Sep	123.878	1.67	1.8	7.2
	Oct	133.330	1.74	1.76	1.1
	Nov	115.197	1.77	1.79	1.1
	Dec	112.716	1.71	1.88	9.0
TelePower	Jan	142.651	1.74	1.9	8.4
	Feb	145.196	1.59	1.72	7.6
	Mar	166.053	1.58	1.73	8.7
	Jul	134.585	1.76	1.66	6.0
	Aug	145.693	1.75	1.67	4.8
	Sep	130.538	1.7	1.77	4.0
	Oct	151.287	1.61	1.67	3.6
	Nov	120.031	1.76	1.86	5.4
	Dec	116.238	1.8	1.94	7.2
SRTech	Jan	145.288	1.77	1.77	0.0
	Feb	146.307	1.65	1.7	2.9
	Mar	165.094	1.6	1.77	9.6
	Apr	142.826	1.73	1.83	5.5

	Jul	133.294	1.81	1.94	6.7
	Aug	141.345	1.78	1.94	8.2
	Sep	128.119	1.73	1.94	10.8
	Oct	147.594	1.65	1.79	7.8
	Nov	124.391	1.84	1.92	4.2
	Dec	110.793	1.8	1.94	7.2
AyerKeroh	May	149.346	1.73	1.6	8.1
	Aug	146.112	1.74	1.68	3.6
	Sep	145.854	1.68	1.71	1.8
	Oct	149.274	1.65	1.61	2.5
	Nov	140.061	1.7	1.6	6.2
	Dec	137.219	1.7	1.58	7.6
				(Average)	6.0
				MSE	0.0152
				RMSE	0.1234

Appendix C: Results of ANN model predicting Ground ISR using training datasets and comparison with actual value.

Actual ISR	Predicted ISR (using training datasets)	Percentage of Difference (%)
1.56	1.771	13.5
1.68	1.661	1.1
1.88	1.729	8.0
1.54	1.598	3.8
1.77	1.769	0.1
1.83	1.739	5.0
1.7	1.654	2.7
1.76	1.788	1.6
1.64	1.59	3.0
1.68	1.58	6.0
1.61	1.64	1.9
1.94	1.734	10.6
1.8	1.731	3.8

1.81	1.677	7.3
1.77	1.637	7.5
1.59	1.581	0.6
1.75	1.728	1.3
1.67	1.754	5.0
1.8	1.733	3.7
1.62	1.693	4.5
1.86	1.766	5.1
1.71	1.724	0.8
1.94	1.777	8.4
1.94	1.797	7.4
1.78	1.793	0.7
1.76	1.761	0.1
1.63	1.691	3.7
1.7	1.678	1.3
1.55	1.748	12.8
1.72	1.632	5.1
1.77	1.701	3.9
1.67	1.76	5.4
1.76	1.748	0.7
1.74	1.722	1.0
1.73	1.74	0.6
1.94	1.706	12.1
1.71	1.694	0.9
1.64	1.739	6.0
1.67	1.717	2.8
1.59	1.715	7.9
1.63	1.779	9.1
1.76	1.657	5.9
1.68	1.622	3.5
1.79	1.678	6.3
1.52	1.622	6.7
1.73	1.709	1.2

1.67	1.647	1.4
1.5	1.54	2.7
1.7	1.782	4.8
1.55	1.717	10.8
1.77	1.7	4.0
1.79	1.774	0.9
1.82	1.704	6.4
1.8	1.699	5.6
1.84	1.63	11.4
1.52	1.73	13.8
1.62	1.713	5.7
1.92	1.825	4.9
1.48	1.644	11.1
1.69	1.841	8.9
1.9	1.747	8.1
1.44	1.754	21.8
1.72	1.787	3.9
1.56	1.725	10.6
1.94	1.801	7.2
1.72	1.6	7.0
1.74	1.716	1.4
1.79	1.678	6.3
1.8	1.812	0.7
1.66	1.763	6.2
1.8	1.697	5.7
1.64	1.722	5.0
1.94	1.796	7.4
1.77	1.597	9.8
1.44	1.716	19.2
1.59	1.653	4.0
1.94	1.741	10.3
1.77	1.718	2.9
1.73	1.621	6.3

1.83	1.756	4.0
1.67	1.654	1.0
	(Average)	5.6
	Mean Squared Error	0.0136
	Root Mean Squared Error	0.1165

Appendix D: Python Program to compute LCOE of a site using monthly irradiance datasets.

```
import pandas as pd
import numpy as np
import glob
import re

# import irr data
for name_file in glob.glob('AyerKeroh_Satellite\*'):
    irr_df = pd.read_excel(name_file)

    ## calculate the day per month based on number of data per
day
    day_per_month = len(irr_df)/48

    # Data needed
    DC_cap = 10000000
    Ross_C_slope = 0.0234
    Ross_C_intercept = 0.187
    Temp_C = -0.0038
    Temp_STC = 25
    Fix_PV_loss = 0.92
    Inv_overpwr_ratio = 1.1

    ## change the time resolution here
    PtoE_Coeff = 30/60
    tariff = 0.28
    price_af_MAAQ = 0.09
```

```
    MAAQ = 16000000 *(day_per_month/365)
    Period = [15,21,25]
    Inv_curve_eff = {0:0, 0.5:0, 3:0.9233, 5:0.9534, 10:0.9756,
15:0.981,
                      20:0.986, 25:0.988, 30:0.9889, 35:0.9892,
40:0.9895,
                      45:0.9896, 50:0.9898, 55:0.9899, 60:0.99,
65:0.9894,
                      70:0.9892,     75:0.9889,     80:0.9886,
85:0.9883, 90:0.9882,
                      95:0.9881,     100:0.9879,     105:0.984,
110:0.98}

    Maintenance_per_yr = 200000 *(day_per_month/365)
    Inverter = 0.52
    Other_cost = 1.68

    # define range and step size of ISR
    final_df_name  =  ['for  ISR  '+  str(i/100)  for  i  in
range(120,190,+1)]
    final_df  =  pd.DataFrame(np.nan,  index  =  np.arange(3),
columns = final_df_name)


    for ratio in range(120, 190, +1):
        ISR = ratio/100

        Rated_inv_cap = DC_cap / ISR
        Overpwr_inv_cap = DC_cap / ISR * Inv_overpwr_ratio
        Tot_cost_per_W = Inverter + Other_cost
        System_cost = DC_cap * Tot_cost_per_W
        Cost_saved_by_undersized_inv = (DC_cap - Rated_inv_cap)
* Inverter
        Final_system_cost      =      System_cost      -
Cost_saved_by_undersized_inv
```

```
        ##Calculate instantaneous power generated
        Theoretical_Power = irr_df ['GHI'] /1000 * DC_cap
        Estimated_Module_Temp         =         irr_df['Temp']+
Ross_C_intercept +irr_df ['GHI'] * Ross_C_slope
        Temp_Derating_Factor       =       1       +       Temp_C
*(Estimated_Module_Temp - Temp_STC)
        Theor_Power_bf_Inv      =      Theoretical_Power      *
Temp_Derating_Factor * Fix_PV_loss
        Inv_loading_percent      =      Theor_Power_bf_Inv      /
Rated_inv_cap * 100

        result_df        =        pd.concat([Theoretical_Power,
Estimated_Module_Temp, Temp_Derating_Factor,
                                Theor_Power_bf_Inv,
Inv_loading_percent],axis=1)
        result_df.columns       =       ['Theoretical_Power','
Estimated_Module_Temp', 'Temp_Derating_Factor',
                                'Theor_Power_bf_Inv',
'Inv_loading_percent']

        result_df['Inv_eff_from_loading'] = np.nan

        for i in range(len(result_df)):

            search_key = result_df['Inv_loading_percent'][i]
            res      =      Inv_curve_eff.get(search_key)      or
Inv_curve_eff[
            min(Inv_curve_eff.keys(), key = lambda key: abs(key-
search_key))]

            result_df ['Inv_eff_from_loading'] [i]= res

        ## Calculate clipped power
        Theor_Power_af_Inv = result_df ['Theor_Power_bf_Inv'] *
result_df ['Inv_eff_from_loading']
```

```
        result_df            =            pd.concat([result_df,
Theor_Power_af_Inv.rename('Theor_Power_af_Inv')],axis=1)

        result_df['Theor_Power_af_Clipping'] = np.nan

        result_df.loc[result_df['Theor_Power_af_Inv']         >
Overpwr_inv_cap, 'Theor_Power_af_Clipping'] = Overpwr_inv_cap
        result_df['Theor_Power_af_Clipping']                  =
result_df['Theor_Power_af_Clipping'].fillna(result_df['Theor_P
ower_af_Inv'])

        generation_df = result_df[['Theor_Power_af_Inv']].copy()
        compiled_df = pd.DataFrame(np.nan, index = np.arange(25),
columns = ['Generation(kWh)','Clipped Generation(kWh)', 'Energy
Loss(kWh)', 'Income with MAAQ limit(RM)'])
        degrad_rate = 0

        ## run for 25 year to get the energy of degraded PV
system
        for i in range(25):
            name = 'year '+ str(i+1)
            generation_df['generation   in   '+   name   ]   =
generation_df ['Theor_Power_af_Inv']*(1-degrad_rate)
            generation_df['generation_af_clipping in '+ name ]
= np.nan
            generation_df.loc[generation_df['generation  in  '+
name ] > Overpwr_inv_cap, 'generation_af_clipping in '+ name ]
= Overpwr_inv_cap
            generation_df['generation_af_clipping in '+ name ]
=       generation_df['generation_af_clipping       in       '+
name ].fillna(generation_df['generation in '+ name ])
            compiled_df        ['Generation(kWh)'][i]         =
generation_df['generation in '+ name ].sum() /1000 * PtoE_Coeff
            compiled_df   ['Clipped   Generation(kWh)'][i]    =
generation_df['generation_af_clipping in '+ name ].sum() /1000
* PtoE_Coeff
```

```
            degrad_rate += 0.005

        compiled_df['Energy            Loss(kWh)']            =
compiled_df['Generation(kWh)']      -      compiled_df['Clipped
Generation(kWh)']
        compiled_df.loc[compiled_df['Clipped Generation(kWh)'] >
MAAQ,      'Income      with      MAAQ      limit(RM)']      =
(MAAQ*tariff)+(compiled_df['Clipped    Generation(kWh)']-MAAQ)*
price_af_MAAQ
        compiled_df['Income    with    MAAQ    limit(RM)']    =
compiled_df['Income                  with                  MAAQ
limit(RM)'].fillna(compiled_df['Clipped
Generation(kWh)']*tariff)

        LCOE = []
        for year in Period:
            name = 'Clipped Energy for ' + str(year) +' years'
            compiled_df [name] = np.nan
            compiled_df   [name][0]   =   compiled_df['Clipped
Generation(kWh)'].iloc[0: year].sum()

            LCOE    =    LCOE    +    [(Final_system_cost    +
(Maintenance_per_yr * year)) / compiled_df [name][0]]

        for k in range(len(LCOE)):
            final_df['for ISR ' + str(ISR) ][k] = LCOE[k]

    final_transposed = final_df.round(decimals = 5).transpose()

    temp = re.findall(r'\d+', name_file)
    res = list(map(int, temp))
    final_transposed.to_excel('AyerKeroh_Satellite  result/  A_S
output '+str(res)+'.xlsx')
    ## export to output file
```

Appendix E: Python Program to train ANN model.

```
from tensorflow import keras
import pandas as pd
import matplotlib.pyplot as plt
#from sklearn.utils import shuffle
# create some data
# input
datasets = pd.read_excel('all input.xlsx')
#datasets = shuffle(datasets)
#datasets.reset_index(inplace=True, drop=True)

train_dataset = datasets.sample(frac=0.8, random_state=0)
test_dataset = datasets.drop(train_dataset.index)

xs = train_dataset[['Monthly Irr','Satellite ISR']].to_numpy()
ys = train_dataset[['Ground ISR']].to_numpy()

xp = test_dataset[['Monthly Irr','Satellite ISR']].to_numpy()
yp = test_dataset[['Ground ISR']].to_numpy()


# Build a simple model
model = keras.Sequential()
model.add(keras.layers.Dense(4,input_shape=(2,),  activation  =
'relu')) # hidden layer
model.add(keras.layers.Dense(1, activation = 'linear')) #output
layer

#model.summary()
#to train model
model.compile(optimizer='adam',     loss='mean_absolute_error',
metrics = ['mean_squared_error'])
result  =  model.fit(xs,ys,epochs  =  1600,  batch_size  =  2,
validation_split = 0.2)
```

```
##plot error vs epoch graph
def plot_loss(history):
  plt.plot(history.history['loss'], label='loss')
  plt.plot(history.history['val_loss'], label='val_loss')
  plt.ylim([0, 2])
  plt.xlabel('Epoch')
  plt.ylabel('Error [MSE]')
  plt.legend()
  plt.grid(True)

plot_loss(result)

test = model.predict(xs)
prediction = model.predict(xp)

## plot graph to compare actual and predicted value
fig,ax = plt.subplots()
plt.xlabel('No. of Prediction')
plt.ylabel('Optimal ISR')
plt.title('Comparison  between  Predicted  and  Actual  Value  of
Optimal ISR')
plt.axis([0, 20, 1.2, 2])
ax.plot(yp, color='green', label = 'Actual Value')
ax.plot(prediction, color='orange', label = 'Predicted Value')
ax.legend(loc = 'lower left')
plt.grid(True)
plt.show()

Actual_ISR=pd.DataFrame(yp,columns=['Actual ISR'])
Predicted_ISR   =   pd.DataFrame(prediction,columns=['Predicted
ISR'])
df = pd.concat([Actual_ISR,Predicted_ISR],axis =1)
final_df = df.round(decimals = 3)


fig2,ax2 = plt.subplots()
```

```
plt.xlabel('No. of Prediction')
plt.ylabel('Optimal ISR')
plt.title('Comparison between Predicted and Actual Value of
Optimal ISR')
plt.axis([0, 80, 1.2, 2])
ax2.plot(ys, color='green', label = 'Actual Value')
ax2.plot(test, color='red', label = 'Predicted Value')
ax2.legend(loc = 'lower left')
plt.grid(True)
plt.show()

Actual_ISR2=pd.DataFrame(ys,columns=['Actual ISR'])
Predicted_ISR2 = pd.DataFrame(test,columns=['Predicted ISR
(using training datasets)'])
df = pd.concat([Actual_ISR2,Predicted_ISR2],axis =1)
final_df2 = df.round(decimals = 3)

model.save('Monthly ISR ANN model/HL4_EP1300_BS2_Trial3.h5')
```