SOLAR POWER FORECASTING USING WAVELET TRANSFORM AND MACHINE LEARNING APPROACHES

NOR AZLIANA BINTI ABDULLAH

INSTITUTE FOR ADVANCED STUDIES
UNIVERSITY OF MALAYA
KUALA LUMPUR

2020

SOLAR POWER FORECASTING USING WAVELET TRANSFORM AND MACHINE LEARNING APPROACHES

NOR AZLIANA BINTI ABDULLAH

DISSERTATION SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF PHILOSOPHY

INSTITUTE FOR ADVANCED STUDIES
UNIVERSITY OF MALAYA
KUALA LUMPUR

2020

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Nor Azliana Binti Abdullah

Matric No: HGF160002

Name of Degree: Master of Philosophy

Title of Dissertation: Solar Power Forecasting using Wavelet Transform and Machine Learning Approaches

Field of Study: Power Energy

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

[SOLAR POWER FORECASTING USING WAVELET TRANSFORM AND MACHINE LEARNING APPROACHES]

ABSTRACT

Generation of photovoltaic (PV) power is intermittent in nature and integration of PV system into the grid system causes an imbalanced power production and power demand. One of the efforts to reduce this problem is to forecast the generation of solar power in the PV system. Solar power forecasting requires the collection of solar power and meteorological data. Hence, this work collected solar power data and various meteorological data (global radiation, tilted radiation, temperature surrounding, humidity surrounding, PV module/ PV panel temperature and wind speed) from Universiti Teknikal Malaysia Melaka (UTeM). A pre-processing process is carried out to ensure that solar power data and meteorological data can be simplified. The proposed work of this study is divided into four phases of works. The work in Phase 1 presents the solar power data and meteorological data into three forecasting models such as Multi-Layer Perceptron (MLP), Radial Basis Function Neural Network (RBFNN) and Adaptive Neuro-Fuzzy Inference System (ANFIS). The performance of every forecasting model is estimated. The work in Phase 2 proposes a Wavelet Transform (WT) technique to remove noise in solar power data and meteorological data. The existence of noise in data is due to the presence of dirt on the sensor of measurement. The denoised solar power and meteorological data are then presented to MLP, RBFNN and ANFIS to conduct the forecasting process. The performance of MLP, RBFNN and ANFIS in Phase 1 and Phase 2 are compared. The comparison result is presented in Phase 3 to estimate the efficiency usage of WT to eliminate noise. The result in Phase 3 depicts an improved performance of MLP, RBFFN and ANFIS when employing WT technique. This can be proven when the values of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) for MLP, RBFNN and ANFIS in Phase 2 are smaller than the values of MAE and RMSE in

Phase 1. Apart from that, the Correlation of Coefficient (R) values for MLP=0.9793, RBFNN=0.9788 and ANFIS=0.9799 in Phase 2 are greater than the R-values of MLP=0.9709, RBFNN=0.9722 and ANFIS=0.9674 in Phase 1. The work in Phase 3 also selects the most accurate forecasting model based on the values of MAE, RMSE and R depicted by MLP, RBFNN and ANFIS in Phase 1 and Phase 2. The result of this work proves that the integration of WT with the ANFIS (WT-ANFIS) surpasses the performance of other forecasting models by providing the lowest MAE value of 0.0278 and lowest RMSE value of 0.0385. The work in the final phase which is Phase 4 includes the integration of Hybrid Firefly and Particle Swarm Optimisation (HFPSO) to optimise the premise parameters of WT-ANFIS. It is observed from the result of WT-ANFIS-HFPSO that the Mean Square Error (MSE) value of 0.0012175, RMSE value of 0.034892 and MAE value of 0.025361 are the lowest compared to the integration of WT-ANFIS with single Firefly (WT-ANFIS-FF) and single Particle Swarm Optimisation (WT-ANFIS-PSO). Furthermore, the WT-ANFIS-HFPSO presents the R-value of 0.98220 which indicates the capability of the model to follow the data pattern efficiently. From the comparative analysis, WT-ANFIS-HFPSO has confirmed its reliability as a forecaster of solar power.

Keywords: Adaptive Neuro-Fuzzy Inference System; Firefly; Hybrid Firefly and Particle Swarm Optimisation; Particle Swarm Optimisation; Wavelet Transform.

[RAMALAN TENAGA SURIA MENGGUNAKAN JELMAAN GELOMBANG DAN KAEDAH-KAEDAH PEMBELAJARAN MESIN]

ABSTRAK

Penjanaan tenaga suria fotovoltaik (PV) adalah tidak sejanjar. Oleh itu, penggunaaan sistem PV di sistem grid telah menyebabkan penjanaan dan permintaan kuasa elektrik menjadi tidak seimbang. Salah satu usaha untuk menstabilkan penggunaan tenaga suria dalam sistem grid ialah melalui ramalan tenaga suria di sistem PV. Dalam merealisasikan usaha untuk membuat ramalan tenaga suria, data tenaga suria dan data kaji cuaca perlu dikumpul. Kajian ini telah mengumpul data tenaga suria beserta beberapa data kaji cuaca (radiasi global, radiasi condong, suhu sekitar, kelembapan sekitar, suhu panel/modul dan kelajuan angin) dari Universiti Teknikal Malaysia Melaka (UTeM). Data tersebut akan diproses terlebih dahulu untuk menghasilkan data tenaga suria dan data kaji cuaca yang tidak kompleks. Penyelidikan ini mencadangkan empat fasa penting dalam menghasilkan dapatan kajian. Fasa 1 telah memperkenalkan data tenaga suria dan data kaji cuaca kepada tiga model ramalan iaitu Perseptron Lapisan Berbilang (MLP), Rangkaian Neural Fungsi Asas Jejari (RBFNN) dan Sistem Inference Neural Kabur Ubah Suai (ANFIS). Prestasi setiap model ramalan itu akan dianggarkan. Fasa 2 mencadangkan teknik Jelmaan Gelombang (WT) untuk membuang hingar yang hadir dalam data. Kehadiran hingar tersebut disebabkan oleh habuk yang melekat di sensor pengukur. Data-data yang tidak mengandungi hingar akan diperkenalkan kepada model MLP, RBFNN and ANFIS dalam melakukan proses ramalan tenaga suria. Fasa 3 kajian ini telah membandingkan prestasi model MLP, RBFNN dan ANFIS di Fasa1 dan Fasa 2 untuk membuktikan kecekapan teknik WT untuk membuang hingar. Hasil kajian menunjukkan bahawa penggunaan teknik WT untuk membuang hingar telah berjaya meningkatkan prestasi model MLP, RBFNN dan ANFIS. Hal ini dapat dibuktikan apabila nilai Min Ralat Mutlak (MAE) dan nilai Punca Min Kuasa Dua (RMSE) bagi MLP, RBFNN dan ANFIS di Fasa 2 adalah

lebih kecil berbanding nilai MAE dan nilai RMSE di Fasa 1. Selain itu, nilai Pekali Korelasi (R) untuk MLP= 0.9793, RBFNN= 0.9788 dan ANFIS= 0.9799 dalam Fasa 2 adalah lebih tinggi berbanding nilai MLP=0.9709, RBFNN=0.9722 dan ANFIS=0.9674 dalam Fasa 1. Fasa 3 dalam kajian ini juga turut menghasilkan sebuah model ramalan yang paling tepat berdasarkan nilai-nilai MAE, RMSE dan R yang diberikan oleh model MLP, RBFNN dan ANFIS dalam Fasa 1 dan Fasa 2. Kajian ini membuktikan bahawa teknik WT bersama dengan model ANFIS (WT-ANFIS) berjaya memberi prestasi yang paling baik dengan memberikan nilai MAE bersamaan dengan 0.0278 dan nilai RMSE bersamaan dengan 0.0385 yang paling rendah berbanding model-model ramalan yang lain. Kajian di fasa yang terakhir iaitu Fasa 4 telah membangunkan kaedah Hibrid Algoritma Kelip-Kelip dan Pengoptimuman Kerumunan Zarah (HFPSO) dalam mengoptimumkan parameter premis WT-ANFIS. Hasil kajian ini membuktikan bahawa model WT-ANFIS-HFPSO berjaya mencapai nilai Min Kuasa Dua (MSE) sebanyak 0.0012175, nilai RMSE sebanyak 0.034892 dan nilai MAE sebanyak 0.025361. Nilai-nilai tersebut adalah yang paling rendah berbanding integrasi WT-ANFIS dengan Algoritma Kelip-Kelip (WT-ANFIS-FF) dan integrasi WT-ANFIS dengan Pengoptimuman Kerumunan Zarah (WT-ANFIS-PSO). Tambahan lagi, nilai R yang diberikan oleh model WT-ANFIS-HFPSO adalah bersamaan dengan 0.98220 yang menunjukkan keupayaan model tersebut untuk mengikuti corak data dengan lebih cekap. Melalui analisa perbandingan, WT-ANFIS-HFPSO telah membuktikan keupayaannya sebagai peramal tenaga suria yang bagus.

Keywords: Algoritma Kelip-Kelip; Hibrid Algoritma Kelip-Kelip dan Pengoptimuman Kerumunan Zarah; Jelmaan Gelombang; Pengoptimuman Kerumunan Zarah; Sistem Inference Neural Kabur Ubah Suai

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Allah S.W.T for giving me the opportunity, determination and strength to complete my Master's Degree in Power Energy.

I would like to express my deepest appreciation to my supervisor, Prof. Ir. Dr Nasrudin Bin Abd Rahim, who continually gives his constant guidance, valuable expert suggestion and encouragement throughout the research work. Without his guidance and persistence help, this dissertation would not have been possible.

I am deeply indebted to my dear friends and UMPEDAC staff for their invaluable help and moral support throughout this course of research works.

I would like to give a special thanks to my beloved family for encouraging me and inspiring me to follow my dreams. I am especially grateful to my husband and my parents, who supported me emotionally.

Last but not least, I humbly extend my thanks to all concerned persons who have co-operated with me in this regard.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

ABC	:	Artificial Bee Colony
ACF	:	Autocorrelation
ACO	:	Ant Colony Optimisation
AIC	:	Akaike Information Criterion
AIFNN	:	Evolving Fuzzy Neural Network and Simulated Annealing
ANFIS	:	Adaptive Neuro-Fuzzy Inference System
ANN	:	Artificial Neural Network
AR	:	Autoregressive Moving Average
ARIMA	:	Autoregressive Integrated Moving Average
ARMA	:	Autoregressive Moving Average
BFG	:	Broyden-Fletcher-Goldfarb-Shanno Quasi-Newton
BMD	:	Bangladesh Meteorological Department
BP	:	Backpropagation
BPNN	:	Backpropagation Neural Network
CABC	:	Chaotic Artificial Bee Colony
CGB	:	Conjugate Gradient Backpropagation with Powell-Beale
CGF	:	Conjugate Gradient Backpropagation with Fletcher-Reeves
CGP	:	Conjugate Gradient Backpropagation with Polak-Ribiere
CVRMSE	:	Coefficient of Variance based on Root Mean Square Error
CWT	:	Continuous Wavelet Transform
DBSCAN	:	Density-Based Spatial Clustering of Applications with Noise
DNI	:	Direct Normal Irradiance

DHI	:	Diffuse Horizontal Irradiance
DWT	:	Discrete Wavelet Transform
ECMWF	:	European Centre for Medium-Range Weather Forecast
ELM	:	Extreme Learning Machine
ESS	:	Energy Storage System
ESSS	:	Exponential Smoothing State Space
EP	:	Evolutionary Programming
FCM	:	Fuzzy C-Means
FF	:	Firefly
FFT	:	Fast Fourier Transform
FHRCNN	:	Hyper-Rectangular Composite Neural Network
FIS	:	Fuzzy Inference System
FT	:	Fourier Transform
GA	:	Genetic Algorithm
gaussmf	:	Gaussian curve membership function
GEM	:	Environment Canada's Global Environmental Multiscale
GFS	:	Global Forecast System
GGA	:	Grouping Genetic Algorithm
GHI	:	Global Horizontal Irradiance
GP	:	Genetic Programming
GRBF	:	Gaussian Radial Basis Function
GRNN	:	General Regression Neural Network
GSI	:	Global Solar Irradiance
GSR	:	Global Solar Radiation
HAPE	:	Hybrid of Ant Colony Optimisation and Particle Swarm Optimisation

IACC	:	Improved Ant Colony Clustering
IS	:	Input Parameter Selection
HFPSO	:	Hybrid Firefly and Particle Swarm Optimisation
HSV	:	Hue Saturation Value
k-NN	:	k-Nearest Neighbour
LM	:	Lavenberg Marquardt
LOESS	:	Local Polynomial Regression Fitting Smoothing
logsig	:	Logistic sigmoid activation function
LS-SVM	:	Least Square Support Vector Machine
MA	:	Moving Average
MAE	:	Mean Absolute Error
MAPE	:	Mean Absolute Percentage Error
MBE	:	Mean Bias Error
MEF	:	Mean Error Function
MLP	:	Multi-Layer Perceptron
MLR	:	Multiple Layer Regression
MPE	:	Mean Percentage Error
MRE	:	Mean Relative Error
NAM	:	North American Model
NASA	:	National Aeronautics and Space Administration
NRMSE	:	Normal Root Mean Square Error
NSE	:	Nash-Sutcliffe Equation
NSGA II	:	Non-dominated Sorting Genetic Algorithm II
NWP	:	Numerical Weather Prediction
PACF	:	Partial Correlation
PSO	:	Particle Swarm Optimisation

purelin	:	Linear activation function
PV	:	Photovoltaic
R	:	Correlation of Coefficient
R^2	:	Coefficient of Determination
RAM	:	Random Access Memory
RBFNN	:	Radial Basis Function Neural Network
RGB	:	Red, Green, Blue
RMSE	:	Root Mean Square Error
RNN	:	Recurrent Neural Network
RP	:	Resilient Backpropagation
RVM	:	Relevance Vector Machine
SA	:	Simulated Annealing
SAT	:	Satellite
SCG	:	Scaled Conjugate Gradient
SOM	:	Self-Organising Maps
SVM	:	Support Vector Machine
SVR	:	Support Vector Regression
SRSVRCABC	:	Seasonal Recurrent Support Vector Regression Model with Chaotic Artificial Bee Colony
tansig	:	Hyperbolic tangent sigmoid activation function
TDNN	:	Time Delay Neural Network
TSI	:	Total Sky Imager
USI	:	University of California, San Diego Sky Imager
UTeM	:	Universiti Teknikal Malaysia Melaka
WRF	:	Weather Meso-scale
WT	:	Wavelet Transform

$x(n)$	:	Data sequence in the time domain
$X(k)$	:	Data sequence in the frequency representation
N	:	Lengths of the data sequence
y_i	:	Observations of the noisy signal
$g(x_i)$	:	Cubic polynomial
$g''(x_i)$	:	Second derivative of the cubic polynomial
p	:	Weighting parameter
a	:	Scaling parameter
b	:	Translation parameter
ψ	:	Mother wavelet
$W(a,b)$	:	Signal in the frequency domain
K	:	Dimension of the signal
t	:	Time sampling index of a function
β_p	:	Constant value for the AR model
w_t	:	White Gaussian noise with the zero means
θ_q	:	Constant value for the MA model
V	:	Loss function
N	:	Number of values in the estimation data set
T	:	Transpose of the matrix
θ_N	:	Estimated parameters
y_t	:	Current data
y_{t-1}	:	First data before the current data
y'_t	:	Data after the first differencing

$y_t{}^*$	:	Data after the second differencing
y_{t-2}	:	Second data before the current data.
d	:	Number of differences needed for stationarity
β_0	:	Intercept of a plane
β_1	:	Mean changes in the response variable that corresponds to the unit of x_1 when x_2 is kept constant
$\beta 2$	:	Mean changes of the response variable due to a unit change of x_2 when x_1 is kept constant
ϵ	:	Random error
w	:	Weight of network
b	:	Bias vector
α	:	Learning rate
e	:	Error rate
$\phi(x)$	:	Non-linear mapping function
ε	:	Tube size
ζ_i and $\zeta_i{}^*$	:	Slack parameters
σ	:	Gauss parameter
P_{best}	:	Personal best
G_{best}	:	Global best
V_{max}	:	Maximum number of allowable velocity
V_{min}		Minimum number of allowable velocity
ω	:	Inertia weight
v_i (t-1)	:	Previous velocity of a particle
v_i (t+1)		New velocity of a particle
C_1	:	Cognitive parameter

C_2	:	Social parameter
r1 and r2	:	Uniformly distributed number in the range of 0 to 1
$x_i(t-1)$	:	Previous position of a particle
$x_i(t+1)$	:	New position of a particle
ω_{max}	:	Final inertia weights
ω_{min}	:	Initial inertia weights
Itr_{max}	:	Maximum number of iteration
Itr	:	Current iteration number
$x_{60-\min}$	:	60-minutes interval value
$\sum_{t=1}^{60} x_{t-\min}$	:	1-minute interval until 60- minutes interval
x_{norm}	:	Normalised data value
$y_{\max}$	:	Maximum value of the target vector
$y_{\min}$	:	Minimum value of the target vector
$x_{\max}$	:	Maximum value of input vectors
$x_{\min}$	:	Minimum value of input vectors
W_n	:	Weight
φ_n	:	Activation function
σ	:	Spread value
A_i and B_i	:	Fuzzy set variables of i^{th} rule
p_i, q_i and r_i	:	Consequent parameters
μ_{A_i}	:	Membership function of i^{th} rule A_i
c_j	:	Centre of Gaussian membership function
σ_j	:	Width of Gaussian membership function

w_j	:	Firing strength
$\overline{w_j}$	:	Normalisation of the firing strength
I_f	:	Forecasted value
I_m	:	Actual value
$\overline{I_m}$	:	Mean of actual values
$\overline{I_f}$	:	Mean of forecasted values
n	:	Number of observations

CHAPTER 1: INTRODUCTION

1.1 Background and Motivation

The production of electricity from renewable energy resources has increased from year to year due to its wide-ranging benefits of controlling the level of pollution as well as reducing the amount of carbon dioxide emission to atmosphere (Behera et al., 2018). Solar energy becomes the most promising renewable energy resources and it offers significant advantages of being a clean, cost-free and abundant source (Semero et al., 2018).

The market shares of solar energy in the production of electricity is continuing to increase. As a result, the need for reliable electricity production in meeting the demand is increasing as well. The generation of electricity from solar energy is variable as it is influenced by various environmental factors such as solar radiation, wind speed, temperature surrounding, and humidity surrounding. This variability problem raises concerns among the system operators as they have to face certain challenges to operate and to dispatch the generated power to a transmission system (Shi et al., 2012).

Several methods have been developed to optimise the production of solar energy. One of them is to forecast the demands on the loads to improve the operation stability of the power system (Rodríguez et al., 2018). The other method, which is addressed in this work, is to forecast the power generation from renewable energy (solar energy) resources. An accurate solar power forecasting has increased the capability of energy trading companies and dispatching centre of the power network to make accurate decisions on certain important issues such as scheduling arrangement and operation control of the power system (Eseye et al., 2018). On top of that, the reliability and power quality of the overall power system can be improved as well (Shi et al., 2012).

An accurate solar energy forecasting will maintain the security of a grid and this process can only be achieved from a model that forecasts solar energy with the greatest accuracy. Several physical and statistical models are developed recently to forecast solar power with different time horizons. The illustration of this forecasting model and forecasting horizon can be shown in Figure 1.1. A physical model predicts certain factors (solar radiation or temperature) which directly influence the production of solar power. Later, those predicted factors (solar radiation or temperature) are used as inputs of a model to forecast the future values of solar power (Wang, J. et al., 2018). It should be noted that the physical model is difficult to be installed and the maintenance cost is also expensive which make it less desirable in the forecasting area.

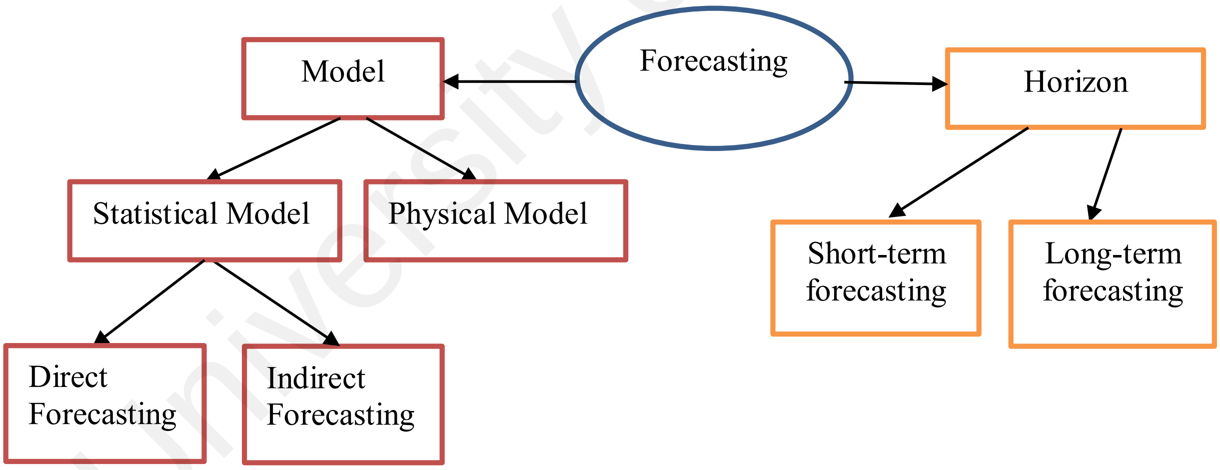


Figure 1.1: Forecasting Model and Forecasting Horizon

The other model which is a statistical model uses mathematical models or builds machine learning algorithms to directly forecast solar power without using any physical model (Wang, J. et al., 2018). Previous researchers prefer to utilise statistical model instead of a physical model because it is much easier and more efficient to be used in the area of solar power forecasting (Leva et al., 2017). Furthermore, it can be classified into direct and indirect forecasting. Direct forecasting refers to a statistical model that directly forecast solar power as the model output. On the contrary, indirect forecasting model forecasts the values of solar radiation in the first place and those forecasted values of solar radiation are used in the photovoltaic (PV) performance model to obtain the forecasted values of solar power (Huang, C. et al., 2018).

Forecasting of solar power can be divided into short-term forecasting and long-term forecasting. Short-term forecasting refers to a forecast that is made from minutes to hours and up for several days. Apart from that, long-term forecasting is used to forecast solar power for a longer period which is up to months and years (Malik, 2016). Certain models are suitable to be utilised for short-term forecasting and certain other models are fitted for long-term forecasting. The model suitability for forecasting is measured based on the capability of forecasting models to generalise the data that has been introduced to them. This means that an accurate forecasting model manages to produce forecasted data that is closest to actual data.

Other than a good selection of the forecasting model, the optimal selection of data is one of the requirements that can increase the forecasting accuracy. The data which is collected for solar power forecasting must be validated to ensure the highest model accuracy. Therefore, any poorly behaved data, for instance, noisy data must be filtered instantly. The noise in time series data has caused the data to become non-stationary that leads to the wrong model coefficient. This will significantly cause the forecasting process to become less accurate.

This work comprises of four main phases. The first phase focuses on the performance of forecasting models, namely, Multi-Layer Perceptron (MLP), Radial Basis Function Neural Network (RBFNN) and Adaptive Neuro-Fuzzy Inference System (ANFIS) when using the noisy data to forecast future values of solar power. The second phase employs the Wavelet Transform (WT) technique to eliminate the presence of noise in the solar power and meteorological data. The denoised data are then introduced to the MLP, RBFNN and ANFIS to conduct the forecasting process. The third phase compares the performance of MLP, RBFNN and ANFIS in Phase 1 and Phase 2 in showing the effectiveness of WT as a noise elimination technique. Furthermore, the third phase selects the most accurate forecasting model based on a model that offers supreme performance. The fourth phase of this work uses the Hybrid Firefly-Particle Swarm Optimisation (HFPSO) approach to optimise the parameter of the most accurate forecasting model. The utilisation of HFPSO is estimated to achieve a greater forecasting accuracy.

1.2 Problem Statement

The demand for pollution-free electricity has driven an approach of integrating renewable energy resources with non-renewable energy resources in the grid system. The adding of renewable energy resources has caused a new technological challenge to the grid as it needs to face the variability production of renewable energy resources, particularly solar energy. The variability of solar energy is depending on weather characteristics such as passing cloud, humidity surrounding, solar radiation, and temperature surrounding. This variability introduces uncertainty in power generation at the grid as an exact amount of solar power variation is unpredictable.

The challenge of variability can be reduced through several methods. The first method is to store the output power in large-scale energy storage such as the pumped hydroelectric or batteries. The second approach is to balance regional deficits or excesses by using a long-distance transmission (Crabtree et al., 2011). The last method to be employed is to forecast load demand or power generation so that conventional generation capacity can be switched on or switched out instantly.

Out of three methods to reduce the variability problem, this study focuses on short-term forecasting of solar power. The forecasting process is implemented by using several statistical models that require the collection of historical meteorological data as well as historical solar power data to predict future solar power data. Note that the existence of noise in data will affect the accuracy of forecasting models. To address this problem, a noise elimination technique which is known as WT is proposed to remove noise in data.

This study forecasts solar power data by using three statistical models that are identified as MLP, RBFNN and ANFIS models. The relevance to developing three forecasting models is to find a model that has the greatest accuracy to forecast solar power data. Each forecasting model works differently and it produces the results based on the nature of the data that has been collected. Therefore, the performance of every forecasting model has been compared to find the most accurate model for solar power forecasting.

The statistical forecasting models work through an adjustment process of parameters. Some of the parameters are determined by users through a trial and error process and some of them are estimated according to algorithms existed in the respective forecasting model. Nevertheless, both methods do not give ideal parameter values for forecasting. As a result, the forecasting model will not give its maximum accuracy during the forecasting process.

To optimise the value of parameters, several optimisation techniques have been utilised by previous works. (Fei & He, 2015) utilised Artificial Bee Colony (ABC) algorithm to select the optimum kernel parameters for the Relevance Vector Machine (RVM) model and (Awan et al., 2014) used ABC algorithm to optimise set of neuron connection weights for the Artificial Neural Network (ANN). The outcomes of both works show the supreme performance of ABC optimisation algorithm. (Aybar-Ruiz et al., 2016) presented a novel approach where Genetic Algorithm (GA) was used to find an optimal feature of Extreme Learning Machine (ELM) model and GA was also proposed to find the optimum parameters of the Support Vector Machine (SVM) in the work of (Liu, D. et al., 2014). It is obvious from the results of both works that utilisation of GA has improved the performance of the forecasting model. The Particle Swarm Optimisation (PSO) algorithm was also used by (Bahrami et al., 2014) as it was combined with the grey model for load forecasting. On the other hand, (Ibrahim & Khatib, 2017) proposed a hybrid model of random forest and Firefly (FF) algorithm to predict hourly Global Solar Radiation (GSR) where FF algorithm was employed to optimise the number of trees and leaves per tree in the random forest approach. It is realised from the results of both works that the utilisation of PSO and FF algorithms provide a better performance of the forecasting model.

PSO is preferable to be used as an optimisation algorithm because it is a simple algorithm and it can be used to solve any continuous problem efficiently (Adrian et al., 2015; Niknam et al., 2013). Nonetheless, it is not easy to acquire a solution from PSO algorithm because it has a shortcoming of premature convergence (Premalatha, K. & Natarajan, 2009). The FF algorithm is likely to be used in the optimisation area because it tends to not experience premature convergence (Aydilek, 2018). Though it has some demerits of being trapped in local optima as well as giving a poor performance in the high dimensional problem (Ali, 2014) In this work, an HFPSO approach is proposed where it is a hybridisation of PSO and FF algorithms. It is used to update and to optimise the parameter of the most accurate forecasting model which has been estimated beforehand. It is expected that the usage of HFPSO will increase the performance of the most accurate forecasting model.

1.3 Research Objectives

The objectives of this study are:

1. To reduce noisy data from on-site raw data collection using Wavelet Transform (WT) in MATLAB software.
2. To simulate Multi-Layer Perceptron (MLP), Radial Basis Function Neural Network (RBFNN) and Adaptive Neuro Fuzzy Inference System (ANFIS) as the forecasting models of solar power.
3. To compare the forecasting model accuracy performance of MLP, RBFNN and ANFIS.
4. To optimise the parameter of the most accurate forecasting model using Hybrid Firefly-Particle Swarm Optimisation (HFPSO) approach.

1.4 Scope of the Study

Solar power forecasting process of this work is carried out by collecting solar power data and meteorological data from PV systems installed on the rooftop of laboratory and administration building at the Faculty of Electrical Engineering in Universiti Teknikal Malaysia Melaka (UTeM), Malaysia. The PV system is located at a longitude of 102.3° E, a latitude of 2.3° N and an altitude of 70 meters above sea level. The collection of solar power (W) data is carried out from a solar monitoring system in one-minute time step resolution that provides 241200 measured power data. All solar power data is collected from 8 A.M. to 7 P.M.

In the meantime, the meteorological data of this work is obtained from the weather station which is installed at the same location as the PV system. The meteorological variables which are global radiation (W/m^2), titled radiation (W/m^2), temperature surrounding (°C), PV module/ PV panel temperature (°C), wind speed (m/s) and humidity surrounding (%) are collected for every minute. Similar to solar power data, all of the meteorological variables are collected from 8. A.M. until 7 P.M.

Solar power data and meteorological data might be exposed to noise due to several factors. In this case, WT is used to remove the noisy data from on-site raw data collection. The relevance of choosing WT instead of other noise elimination methods is due to its capability to remove noise in data as well as to maintain important information of signals at the same time.

Aforementioned, forecasting models can be categorised as the physical model and the statistical model. For this work, three statistical models, namely, MLP, RBFNN, and ANFIS are used to forecast future values of solar power data. The MLP is one type of ANN method that is widely used for forecasting due to its benefit of making the decision function directly from the training process. The other forecasting model which is RBFNN is practically employed for time series forecasting and function approximation. The last forecasting model, namely, the ANFIS model is a combination method of ANN and Fuzzy Inference System (FIS). The utilisation of ANFIS combines the benefits of both ANN and FIS models which enables the knowledge to be represented in an interpretable way while possessing a training algorithm to adjust parameters of knowledge. The performance of every forecasting model is estimated. After that, their performance is compared to find a model that provides the greatest accuracy.

This work also suggests using an optimisation algorithm which is known as HFPSO to update and optimise the parameters of the most accurate forecasting model. It is estimated that the usage of HFPSO will enhance the performance of the most accurate forecasting model.

Proposed works of using a noise elimination technique, the forecasting models and an optimisation approach are simulated and implemented in MATLAB software version Release 2015b (R2015b). The implementation of WT for noise elimination process is done in the MATLAB toolbox where it has a specialised function that enables the user to use it more efficiently. Apart from that, the MLP, RBFNN and ANFIS models are simulated by developing the MATLAB codes. Similar to forecasting models, the HFPSO approach is implemented from the MATLAB code development process. All of the simulations are done inside the ASUS laptop, model A550C which features 4 gigabytes Random Access Memory (RAM), Intel Core i3 processor, 5 gigabytes memory size, 2 number of core processors and 1.8 gigahertz speed.

1.5 Thesis Outline

The thesis consists of five chapters and the outline of every chapter is summarised as follows:

Chapter 2 presents the literature reviews that summarise the theoretical framework associated with noise elimination techniques, followed by various forecasting models including the physical and statistical model. The final part of this chapter discusses several optimisation algorithms that have been used previously in the forecasting area.

Chapter 3 highlights the methodology of the research. The first part of this chapter covers the data collection and data pre-processing procedures in the forecasting process. This chapter also discusses an application of WT to remove noise in solar power data and meteorological data. Apart from that, the application of MLP, RBFNN, and ANFIS to forecast solar power data have been explained in detail in this chapter. The last part of the chapter presents the utilisation of HFPSO to update and to optimise the parameter of the most accurate forecasting model.

Chapter 4 focuses on the results of solar power forecasting followed by discussions of the findings. This chapter has been divided into four phases. The first phase (Phase 1) presents the results of MLP, RBFNN and ANFIS models where the data presented to those models are not been denoised by any noise elimination technique. The second phase (Phase 2) elaborates the performance of all forecasting models when utilising the denoised data from WT. The third phase (Phase 3) makes a performance comparison for MLP, RBFNN, and ANFIS from the previous phases. Later, the most accurate forecasting model has been selected. The final phase which is the fourth phase (Phase 4) discusses the performance of the most accurate forecasting model when utilising the HFPSO approach to update and to optimise its parameters.

Chapter 5 states the conclusion of this study. This final chapter also covers the recommendations for further work.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

In this chapter, the concept of solar power forecasting is discussed in details. It is then followed by discussions of several noise elimination techniques that are used to remove the noisy solar power and meteorological data. This chapter also focuses on different types of solar power forecasting methods which are further categorised into physical models and statistical models. The final part of this chapter investigates about various optimisation algorithms which are used to update and to optimise the parameter values of the forecasting model.

2.2 Solar Power Forecasting

The security of the power system is maintained by supplying adequate amounts of electricity to users in meeting the demand at the loads. This supply and demand equilibrium can only be achieved from schedules of dispatching department of a power system that manages the amount of power delivered to users. The utilisation of the PV system to produce the electricity in the grid has imposed a challenge to dispatching management to manage the exact amount of power that can be delivered to users. This is due to the variability factor of solar power which makes it difficult to generate constant solar powers. Thus, several measures have been undertaken to secure the integration of solar energy into the grid system. One measure that is addressed in this work is known as solar power forecasting.

Many previous works such as (Bouzerdoum et al., 2013; Larson et al., 2016; Persson et al., 2017; Sperati et al., 2016; Tang et al., 2018; Wang, J. et al., 2018; Wang, J. et al., 2017; Wang, F. et al., 2015) carried out the forecasting process of solar power. Note that solar power forecasting is not an easy process as it is highly depending on the various environmental factors that are fluctuating over time. To mitigate this shortcoming, several forecasting models have been developed recently which are further categorised into physical models and statistical models. In this chapter, several physical models are being discussed, namely, sky imager, Numerical Weather Prediction (NWP) and satellite cloud motion vector models. This chapter also elaborates about various statistical models of solar power forecasting which are known as Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), Multiple Layer Regression (MLR), ANN, FIS and SVM.

Both categories of forecasting models are dealing with certain data variables to forecast solar power as the model output. It is important to note that various data variables are exposed to noises. The existence of noises corrupts the data signal (Lyu et al., 2014). As a result, the forecasting model cannot forecast solar power data accurately. Therefore, previous researchers have resorted to the use of several denoising techniques such as Fourier Transform (FT), moving average filter, smoothing cubic spline and WT to enhance the performance of forecasting models.

Furthermore, statistical models are preferred to be used due to its less complexity and shorter computation time during the forecasting process (Wang, F. et al., 2012). They require an adjustment on their parameter values to improve the model performance. Usually, those parameters are being estimated according to trial and error methods that cannot provide optimum parameter values for the forecasting model. However, the parameters of the statistical model can be ideally specified by using certain optimisation algorithms such as ABC, Ant Colony Optimisation (ACO), GA, FF and PSO as discussed in this chapter.

Above-mentioned types of denoising techniques, types of forecasting models and types of optimisation algorithms can be summarised in the forecasting strategy as shown in Figure 2.1.

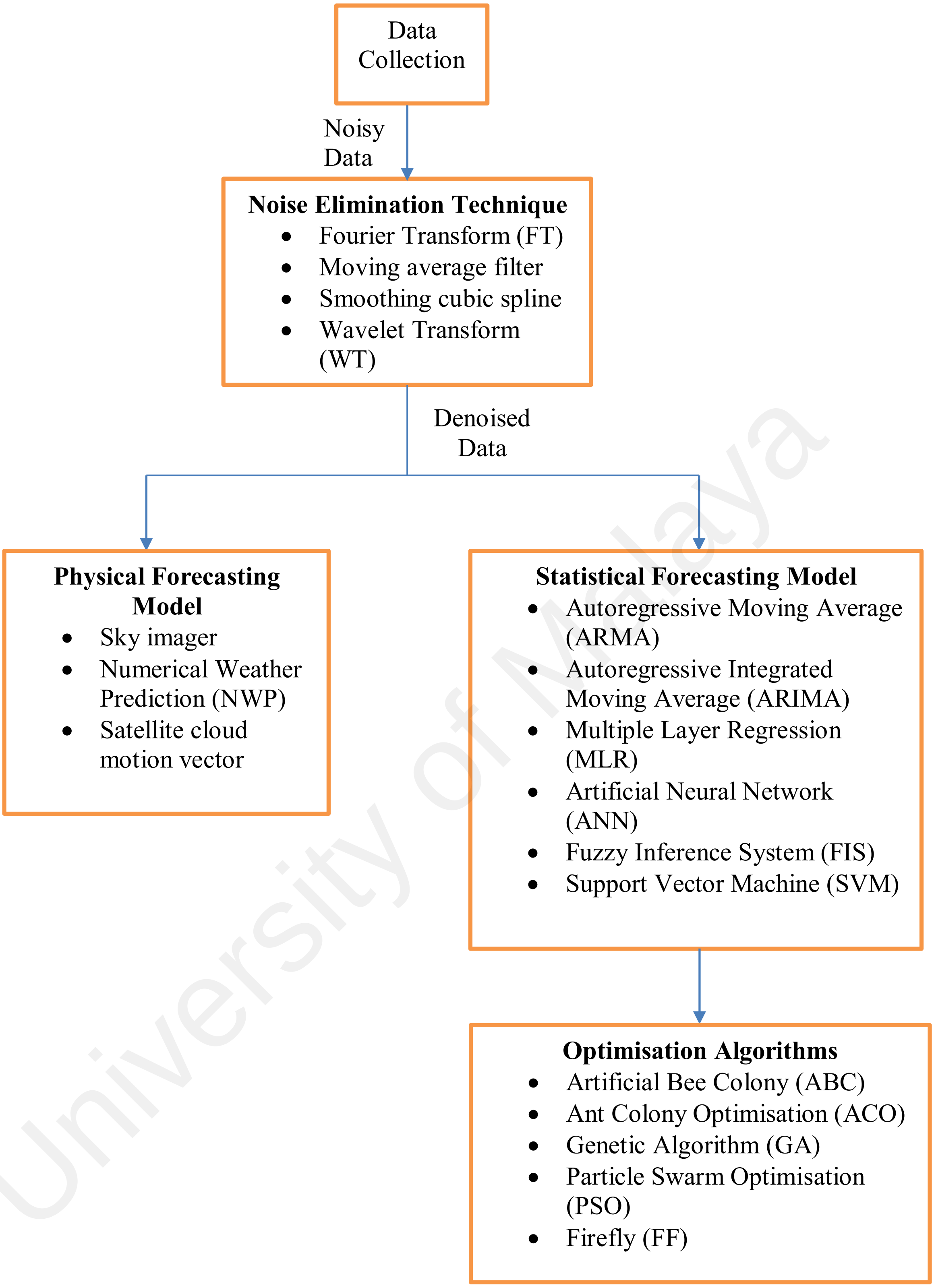


Figure 2.1: Forecasting Strategy

2.3 Noise Elimination Techniques

The data that will be presented to the forecasting model is exposed to noises that gradually degrading the performance of the forecasting model. There are numerous methods which are employed to remove the noise in time series data. Further discussion of those methods can be obtained in the following subsection.

2.3.1 Fourier Transform (FT)

In 1807, Joseph Fourier, a famous French mathematician had made a discovery that enables a periodic function to be represented in the sum of complex exponentials which later had been extended to a non-periodic function (Ismail, Mohd Tahir et al., 2014). A process of obtaining a spectrum of frequencies from time-dependent data is known as Fourier Analysis. FT is a technique used in Fourier Analysis and it can be employed in various applications. One of those applications is to remove the existence of noises in time series data. An example process of removing noisy data in modem application by using the FT can be illustrated in Figure 2.2.

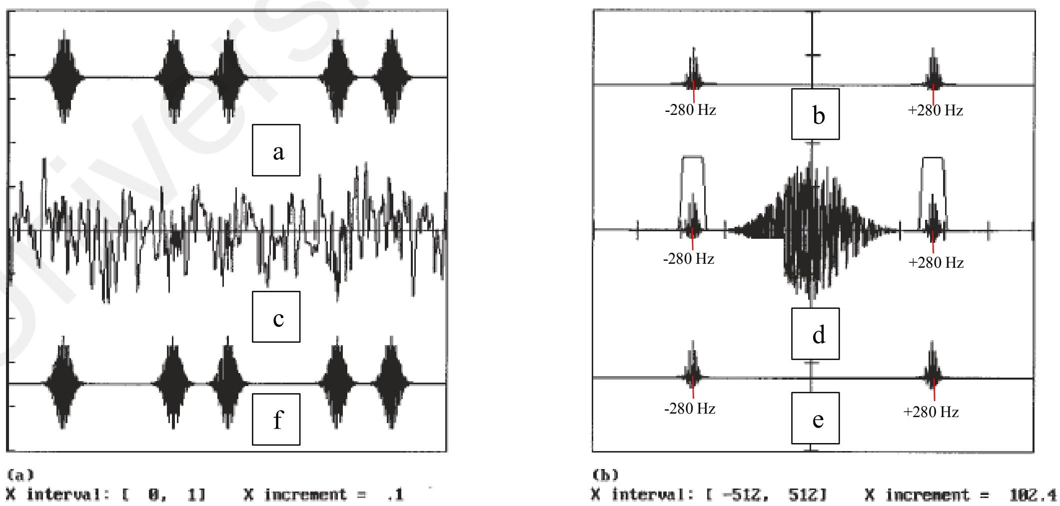


Figure 2.2: (a)Original Signal; (b)Fourier Coefficients of Original Signal; (c) Original Signal after Adding Noise; (d) Fourier Coefficients of Noisy Signal and Filter Function (e) Fourier Coefficients after Multiplication of Filter Function. (f) Denoised Signal (Walker, 1997)

The original signal without the interference of noise can be represented in Figure 2.2 (a). In Figure 2.2 (b), the original signal in Figure 2.2 (a) can be represented in a form of Fourier coefficients and the highest magnitude coefficients are located at the frequencies of ±280 Hz.

As can be seen in Figure 2.2 (c), the original signal has been inflicted by noises. Figure 2.2 (d) displays the Fourier coefficient representation of the original signal that is separated with the noisy signal. The highest Fourier coefficients of the original signal in Figure 2.2 (d) are clustered around frequencies of ±280. Meanwhile, the Fourier coefficients of the noisy signal lie around the origin and subsequently undergone a magnitude reduction to zero value when it is approaching frequencies of ±280.

The noise elimination process by FT is done through the multiplication process of the original signal and noisy signal with a filter function. In this case, the coefficients of the original signal are multiplied by the factor of 1 and noisy signal are multiplied by the factor of 0. Hence, the coefficients of the original signal are recovered as portrayed in Figure 2.2 (e). Later, the frequency coefficients in Figure 2.2 (e) are translated into time coefficients as shown in Figure 2.2 (f).

Classical FT faces problems when analysing a transient signal. Thus, this problem is mitigated by introducing a Fast Fourier Transform (FFT) algorithm that represents the signal in a discrete form. The equation of the FFT algorithm is shown in Equation 2.1 (Ismail, Mohd Tahir et al., 2014).

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi/N}, \quad 0 \le k \le N-1 \tag{2.1}$$

From Equation 2.1, $x(n)$ signifies data sequence in the time domain, $X(k)$ denotes the data sequence in the frequency representation and N indicates the lengths of the data sequence. The data that has been transformed into frequency coefficients will undergo a noise elimination process. Later, those coefficients will be converted back into the time domain according to the inverse FFT algorithm as shown in Equation 2.2.

$$x(n) = \frac{1}{N}\sum_{n=0}^{N-1} X(k)e^{j2\pi/N}, \; n = 0,1,...., N/2 - 1 \tag{2.2}$$

2.3.2 Moving Average Filter

Moving average filter is functioning as a smoothing filter that eliminates the noises in sampled signals. It gives a good representation in the time domain and poor representation in the frequency domain (Mathuranathan, 2010). The application of this method is to average several input samples to produce one single output at one time. The length of the filter is a key factor that determines the smoothness of the output signal. An increasing number of the filter length produces a smooth output signal while decreasing number of filter length gives a less smooth output signal (Mathuranathan, 2010).

Equation 2.3 shows the equation used by moving average filter technique (Alessio et al., 2002). As can be seen in Equation 2.3, *n* represents the number of filters while *x* signifies the input vector.

$$y[i] = \frac{1}{n}\sum_{k=0}^{n-1} x[i-k] \tag{2.3}$$

For a better understanding of the above-mentioned equation, an expression of moving average filter that utilises five number of filters is shown in Equation 2.4. As can be seen in Equation 2.4, 5-Moving Average Filter has averaged the current input value and four previous input values to obtain a new output value.

$$y[i] = \frac{1}{5}(x[i] + x[i-1] + x[i-2] + x[i-3] + x[i-4]) \tag{2.4}$$

2.3.3 Smoothing Cubic Spline

Smoothing cubic spline method is a combination of natural cubic spline and curvature minimisation that can practically be used to eliminate noises in time series data (Shirkey, 2019). Generally, the natural cubic spline technique represents the noisy data as a piecewise function and it tends to visit every noisy point in a signal as shown in Figure 2.3. The result of this technique is unsatisfying because it produces a rough estimation.

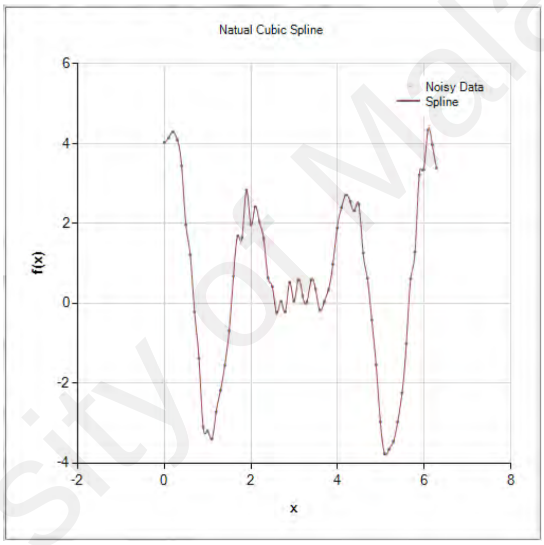


Figure 2.3: Utilisation of Natural Cubic Spline in Fitting the Noisy Data (Shirkey, 2019)

Due to this matter, a new smoothing cubic spline has been introduced where it manages to capture the noisy points in the signal without mimics those noisy points. The implementation of smoothing cubic spline will produce a smooth estimation as illustrated in Figure 2.4.

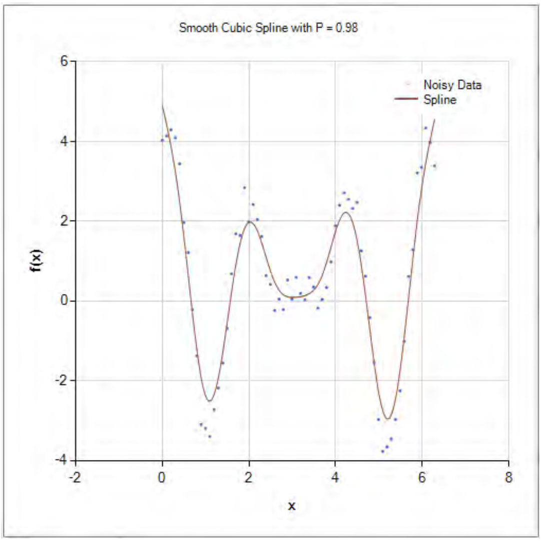


Figure 2.4: Utilisation of Smoothing Cubic Spline in Fitting of Noisy Data (Shirkey, 2019)

Above-mentioned, smoothing cubic spline is a combination of natural cubic spline and curvature minimisation techniques. The utilisation of natural cubic spline will minimise square errors which attempt to attach a spline line closer to the noisy line (Shirkey, 2019). On the other hand, the curvature minimisation technique tries to free the spline from the noisy line to minimise the curvature. It is seen that both techniques are functioning in opposition to each other.

The functions of the natural cubic spline technique and curvature minimisation technique can be shown in Equation 2.5 and Equation 2.6, respectively. Notation y_i in Equation 2.5 signifies the observations of the noisy signal while $g(x_i)$ denotes a cubic polynomial. In Equation 2.6, $g''(x_i)$ indicates the second derivative of the cubic polynomial.

$$\sum_{i=1}^{N}(y_i - g(x_i))^2 \tag{2.5}$$

$$\int dx(g''(x_i))^2 \tag{2.6}$$

For smoothing cubic spline technique, these two equations are being joined together by weighting parameter (p) that is functioning to balance the functions of both techniques as well as to produce a smooth result. The merging equation can be shown in Equation 2.7 where the value of p lies in the range of 0 to 1.

$$p\sum_{i=1}^{N}(y_i - g(x_i))^2 + (1-p)\int dx(g''(x_i))^2 \tag{2.7}$$

The value of p =1 will result in interpolation spline as the curvature constraint has been invalidated. Hence, the spline line tends to visit all noisy points in a signal which means that the smoothing cubic spline has been reverted to a natural cubic spline. Apart from that, the utilisation of p =0 will invalidate the natural cubic spline constraint. Hence, the p value must be ideally selected when removing the noisy data. An appropriate choice of p value will improve the performance of smoothing cubic spline when eliminating noises in time series data.

2.3.4 Wavelet Transform (WT)

The noise elimination application of FT, moving average and smoothing cubic spline techniques tend to suppress noisy data whilst erase important information of the original signal (Lyu et al., 2014). Due to this matter, WT technique has gained many interests among signal processing community due to its benefits. Firstly, the WT manages to decompose time series data into time and frequency representation simultaneously (Lyu et al., 2014). Besides, it can analyse non-linear and non-stationary signal which make it easier to be implemented in the area of signal denoising (Sharma et al., 2016).

The concept of WT technique is quite similar to FT in term of using a basic function when transforming the time-series signal. For FT case, the basic functions used are known as cosine and sine signals. On the contrary, WT employs basic functions in the form of the mother wavelet. WT technique can be divided into two categories, namely, Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT).

2.3.4.1 Continuous Wavelet Transform (CWT)

A detailed analysis of a signal can be obtained from CWT because it provides every single information about the strength of frequency at each timestamp (Sharma et al., 2016). The expression for CWT is shown in Equation 2.8 where it represents three important parameters, namely, scaling parameter, translation parameter and mother wavelet.

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(x) \psi\left(\frac{x-b}{a}\right) dx \tag{2.8}$$

According to Equation 2.8, a scaling parameter is represented by notation a and it is used to control the spread of wavelet when expanding or compressing an original signal. A low value of scaling factor will result in a detailed graph as the signals are being tapered. Apart from that, a high value of scaling factor will produce a less detailed graph due to signal stretching. The other control parameter which is the translation parameter is signified by notation b in Equation 2.8. The translation parameter is responsible to determine the centre of wavelet. The last parameter is known as mother wavelet and it is denoted by notation ψ while $W(a,b)$ indicates the signal in the frequency domain.

2.3.4.2 Discrete Wavelet Transform (DWT)

As has been mentioned in Subsection 2.3.4.1, CWT manages to transform a time series signal into a more detailed signal. Suitable set value of scaling and translation parameters are adequate rather than the full-range value of parameters. This is because a suitable set value of parameters is merely sufficient to preserve the important information of a signal. Hence, a DWT algorithm has been introduced which provides a discrete value of scaling parameter as well as translation parameter. Previous researchers prefer to utilise DWT instead of CWT due to its simple implementation and it produces a better result (Lyu et al., 2014).

$$W(a,b)=\frac{1}{\sqrt{2^{a}}}\sum_{k=0}^{K-1}f(t)\psi\left(\frac{t-b2^{a}}{2^{a}}\right) \tag{2.9}$$

Equation 2.9 shows the algorithm of DWT where the scaling parameter is denoted by a notation of a and the translation parameter is represented by notation of b. Besides, K signifies the dimension of a signal and t characterises the time sampling index of a function.

2.3.5 Comparison of Noise Elimination Techniques

The techniques that have been mentioned above have their advantages and disadvantages during the elimination process of noise in time series data. These advantages and disadvantages are summarised in Table 2.1 (Agayev, 2015; Chen, M.-Y. & Chen, 2014; Ismail, M. T. et al., 2014; Mohammadi et al., 2015; Singh & Mohapatra, 2019; Walker, 1997).

Table 2.1: Summarisation of Noise Elimination Technique

Method	Advantage	Disadvantage
Fourier Transform (FT)	• Decompose time series data into a frequency representation • Good in maintaining the information of amplitude, phase and harmonics during transformation	• Fail to give time information of the signal • Require too many information to reconstruct signal locally • Can only represent the signal in cosine and sine function
Moving Average Filter	• Good to produce the smallest amount of high-frequency noise • Conceptually simple to implement	• Lose some data at the end or beginning of time series data • Take some time to estimate the ideal window size to produce a denoised signal • Usage of large window size will induce a large latency in any signal passing through a filter which is not suitable in real applications
Smoothing Cubic Spline	• Capture noisy points without mimics those points • Produce a smooth estimation	• Consume some time to estimate the weighting parameter of every data point. Yet, those values of the weighting parameter only give a low effect to signal.
Wavelet Transform (WT)	• Decompose time series data in time and frequency representation simultaneously • Able to analyse a non-linear and non-stationary signal • Represent signal in various type of mother wavelets	• Take some time to estimate the ideal type of mother wavelet. However, the same type of mother wavelet is used in every data point.

According to Table 2.1, the technique of WT is the best alternative to be used in the noise elimination process because it possesses some valuable benefits rather than FT, moving average filter and smoothing cubic spline techniques. Hence, this work applies the WT technique to eliminate noise occurrence in time series data.

2.4 Forecasting Models

As has been mentioned in Chapter 1, the forecasting models can be categorised into physical models and statistical models. In this chapter, the basic principle of three physical models which are sky imager, NWP and satellite cloud motion vector model is discussed in detail. Besides, this chapter analyses the basic principle of various statistical models which are known as ARMA, ARIMA, MLR, ANN, FIS and SVM.

2.4.1 Physical Models

Several types of physical models are widely used for forecasting. They are known as sky imager, NWP and satellite cloud motion vector models (Pelland, Remund, et al., 2013).

2.4.1.1 Sky Imager

Sky imager is a small device that is equipped with a digital camera to obtain various images of the sky. The images are analysed to determine cloud motion determination, cloud height measurement as well as solar energy availability (Sobri et al., 2018). It is comprised of three elements, namely, digital camera, arm and digital video recorder as illustrated in Figure 2.5. The digital camera is positioned in an upward or downward over the centre of the dome and it is used to capture the direct or reflected images of the sky. Meanwhile, an arm is functioning to block any direct sunlight on the device whilst a digital video recorder is integrated with image processing software to analyse those images (Inc, 2014).

There are several types of sky imagers existed and the efficiency of every type is different. For instance, one type of sky imager is known as Total Sky Imager (TSI) and it can be illustrated in Figure 2.5. This model is not preferable in the area of solar forecasting due to its limitation in image resolution as well as contains a shadow band that obstruct the full view of the sky (Inman et al., 2013).

Figure 2.5: Total Sky Imager Model TSI-880 (Inc, 2014)

Solar forecasting from sky imager can be made from 10 minutes to 30 minutes ahead and this forecasting process can be done according to several steps. At first, a sky imager uses its digital camera to capture the images of the sky in a wide horizon. Later, an analysis of the sky images is made to estimate the cloud motion vectors and the cloud heights above the ground. The cloud motion vectors are estimated by mapping the original images on a flat space and those images are pre-processed to obtain the successive images. The mapping process is implemented to ensure a uniform size of cloud motion vectors. Lastly, all of the information such as cloud locations and cloud motion vectors are utilised to estimate the cloud cover, solar irradiance and solar power data (Pelland, Remund, et al., 2013).

The existence of multiple cloud layers in the sky has caused the lower level of clouds to cover the upper level of clouds and this phenomenon has caused the changing of cloud geometry (Pelland, Remund, et al., 2013). As a result, there are some variations in the cloud motion vectors that limit the spatial space defined by the field of view of sky imager. Hence, this sky imager model is not preferred to be used in the forecasting area.

Gohari et al. forecasted solar power output to compare the accuracy of the TSI model with the University of California, San Diego Sky Imager (USI) model. The forecasting horizon used in this work was in 15-minutes ahead of forecasting with the time step of the 30-second interval. The result of this work indicated the supremacy performance of the USI as the trend correlation result shown by USI was more consistent than TSI. The result obtained is logical as the TSI has several missing information due to the existence of a shadow band. As a result, the image compression will occur and the resolution of images is lowered (Gohari et al., 2014).

Peng et al. proposed a TSI model to track the cloud in three-dimensional space and the image features of the clouds were used to forecast solar irradiance. This work designed some algorithms to determine the based heights and the wind fields of multiple cloud layers from the multiple TSI. The information obtained from multiple TSI was used to stitch the images to generate a larger cloud and thus increasing the forecasting horizon of the TSI model. Besides, the stitched images were used to capture the fluctuation of solar irradiance. Compared with the persistence model, the proposed work of using multiple TSI had achieved about 26% improvement than the persistence model (Peng et al., 2015)

Alonso-Montesinos et al. developed a novel approach of emerging sky camera technology to forecast beam solar radiation, global solar radiation and diffuse solar radiation. The proposed model converted the digital image into solar irradiance data by using a technology of Red, Green, Blue (RGB) and Hue Saturation Value (HSV) colour space. The maximum cross-correlation method is then applied to estimate future solar irradiance values. The result of this work showed high reliability of the proposed model as it provided the average Normal Root Mean Square Error (NRMSE) values of 25.44%.

11.60% and 11.17% for beam irradiance, diffuse irradiance and global irradiance, respectively (Alonso-Montesinos et al., 2015).

2.4.1.2 Numerical Weather Prediction (NWP)

NWP utilises mathematical models that require observation of the current weathers to forecast the future state of the weather. Those current conditions include wind velocity, wind direction, temperature, moisture, and surface pressure. They are served as the inputs of a numerical computer model through data assimilation process to forecast the future values of weather conditions. Usually, the NWP model is used for the forecasting horizon of 6 hours to 2 weeks.

The NWP model consists of two types, namely, the global model and the mesoscale model (Sobri et al., 2018). The global model simulates the evolution of the weather at the worldwide scales. The initial conditions of global model are derived from the satellite, radar and ground station measurements that will be further processed and interpolated into the 3D grid. It is a computationally and an intensive model and only 14 models of this global type are currently in operation at the worldwide (Pelland, Remund, et al., 2013). To reduce the computation requirement, the relatively coarse resolution of the global model uses the grid spacing of 40 km to 90 km. On the other hand, the mesoscale model estimates the feature of the weather at the limited space such as the countries and regions. It is important to note that the utilisation of a limited geographical area provides a higher resolution and more details images than the global model.

The solar forecasting process by the NWP method consists of several steps. Firstly, the earth surface is divided into grids. The distance between two grid points is called resolution. Low resolution signifies a larger distance between the grids and it requires less computation. On the other hand, high resolution denotes a smaller distance between the grids and it needs more computation. The second step of solar forecasting by the NWP model applies the mathematical equations in the computer to estimate the condition of the weather. Lastly, the weather evolution is used to estimate the future values of solar irradiance as well as future power values.

Fernandez-Jimenez et al. described the development of short-term forecasting process of energy production by using a coupled of three modules that were known as global NWP, a mesoscale NWP and an energy production forecast model. The global NWP forecasted the weather conditions for a given set of positions that covered the entire world. The outputs from global NWP became the inputs to the mesoscale NWP that produced the forecasted values of weather variables nearer to the location of the PV plant. The third module covered the usage of several types of statistical models which were known as ARIMA, k-Nearest Neighbour (k-NN), ANN and ANFIS models. The result of this work portrayed the supreme performance of ANN when being coupled with the global NWP and mesoscale NWP (Fernandez-Jimenez et al., 2012)

Mathiesen & Kleissl analysed five primary techniques for solar irradiance forecasting, namely, Ineichen clear sky model, persistence model, clear sky model interpolated North American Model (NAM), Global Forecast System (GFS) model and European Centre for Medium-Range Weather Forecast (ECMWF) model. The result of this work portrayed the best performance of the ECMWF model followed by the GFS model and the NAM model. Meanwhile, the significant problems were found in Ineichen clear sky model and persistence model which showed the inaccuracy of those models to forecast solar irradiance (Mathiesen & Kleissl, 2011)

Pellan et al. utilised the global NWP, namely, Environment Canada's Global Environmental Multiscale (GEM) model for solar irradiance forecasting. A comparison was made between forecasted outputs of this work, solar irradiance data from the 10 North-American ground stations and alternating current power data from three Canadian

PV systems. Two post-processing techniques that were known as spatial averaging and bias removal by using the Kalman filter were applied to the forecasted data of solar irradiance. The utilisation of the post-processing techniques showed a better performance of the proposed model compared to the persistence model and GEM model that did not employ the post-processing techniques during forecasting (Pelland, Galanis, et al., 2013).

2.4.1.3 Satellite Cloud Motion Vector

The satellite cloud motion vector model is conceptually similar to the sky imager model as both models recognise the cloud patterns according to visible or infrared images from a satellite-based sensor that flies overhead (Pelland, Remund, et al., 2013). However, the satellite cloud motion vector model is effective to be used for one minute until five hours ahead of solar irradiance forecasting. The solar forecasting process for satellite cloud motion vector models requires several procedures. The first procedure is to forecast the conditions of the clear sky by using the existing input conditions such as aerosol content, water vapour and elevation. Next, the consecutive images from a satellite are combined to identify the cloud motion vector field which is used to identify the locations of the cloud. Lastly, the forecast values of solar irradiance are derived by combining the clear sky model with the satellite images (Sobri et al., 2018).

This model offers some improvements than the classical satellite method. In classical satellite method, visible spectrum channels are used which reduce the model accuracy for solar irradiance forecasting. On the contrary, the satellite cloud motion vector model integrates the infrared channels with the visible spectrum channels to give more time history to the model and increase the model accuracy for forecasting (Pelland, Remund, et al., 2013).

This approach can provide a large spatial scale when detecting the cloud motion vector. Nonetheless, the large spatial resolution will not directly detect the clouds due to the large convection of the cloud. Therefore, this approach will become less efficient during rapid cloud forming or cloud dissipation. Additionally, the time to download and to process the images is much slower than the sky imager model. As a result, the forecasting process of this method will not be updated frequently.

Aguiar et al. forecasted Global Horizontal Irradiance (GHI) by combining ground measurement data and exogenous input data provided by satellite (SAT) and ECMWF model. The data combination was presented to the ANN model to conduct the forecasting process. The performance of ANN with different input combinations was compared with Smart-Persistence model and Climatological model in term of Root Mean Square Error (RMSE). It was shown in the result that the combination of the ground data with the exogenous data (SAT and ECMWF) had improved the intra-day forecast. Both at Pozo Izquierdo station and Las Palmas station, the forecast model that combined ANN with SAT and ECMWF (ANN+SAT+ECMWF) gave the best result than other forecasting strategies (Aguiar et al., 2016).

Dong et al. applied satellite image analysis and utilised hybrid model which comprised of Exponential Smoothing State Space (ESSS) and ANN for the hourly forecasting of solar irradiance. The cloud cover index from satellite image was derived and analysed by the Self-Organising Maps (SOM). Then, the ESSS model was employed to forecast the future cloud cover index. Lastly, the ANN was applied to obtain the forecasted solar irradiance from the forecasted value of cloud cover index which was acquired from the ESSS. The performance of the proposed forecasting models was compared with linear exponential smoothing, simple exponential smoothing, random walk and ARIMA techniques. The simulation results had portrayed the 6% accuracy improvement of the proposed hybrid model than other statistical time series models (Dong et al., 2014).

Marquez et al. employed the input information from the satellite images and introduced the information to the ANN model to forecast the values of GHI. The satellite images analysis which comprised of velocimetry and cloud index were used as input variables of the ANN model. The performance of the proposed ANN was compared with the persistence model. The simulation result had proven the accuracy improvement of the proposed ANN model than other persistence models which were used by other studies. This suggested that the hybrid model that utilised a combination of stochastic learning, image processing as well as ground telemetry offered more substantial accuracy and robustness for the development of forecasting model (Marquez et al., 2013).

2.4.2 Statistical Models

Statistical models give higher accuracy and they are much easier to be set up. It can be divided into two types which are statistical non-learning models and statistical learning models.

2.4.2.1 Statistical Non-Learning Models

Statistical non-learning models employ the mathematical knowledge to find the relationship as well as to analyse the data pattern when conducting the forecasting process. The types of statistical non-learning models which are mostly employed are known ARMA, ARIMA and MLR.

(a) ***Autoregressive Moving Average (ARMA)***

ARMA is a great model to understand and to forecast the future values of time series data. Many researchers preferred to utilise ARMA in the forecasting area because of its ability to extract the statistical properties of data based on the Box-Jenkins method. The ARMA model can only be used for stationary data (Ji & Chee, 2011). Hence, any non-stationary data must be transformed into stationary data before it is introduced to the ARMA model. The key idea of this model is to predict the current value of the time-series data, x_t based on the j past values of the series: $x_{t-1}, x_{t-2}, \ldots x_{t-j}$ where j signified the number of past values that are required to predict the current value of data.

The ARMA defines a stationary data in two polynomials terms, namely, Autoregressive (AR) and Moving Average (MA). It is referred to as ARMA (p, q) where p indicates the order of AR and q represents the order of MA. The AR model is employed to measure a correlation between dependent and independent variables of the data. The AR model with the order of p is abbreviated as AR (p) where the expression is shown in Equation 2.10. x_t is stationary data, β_p is a constant value for the AR model (p≠0) and w_t is the white Gaussian noise with the zero mean.

$$x_t = \beta_1 x_{t-1} + \beta_2 x_{t-2} + \ldots.. + \beta_p x_{t-p} + w_t \tag{2.10}$$

The MA part is abbreviated as MA (q) and it is shown in Equation 2.11 where θ_q denotes the constant value for the MA model (q≠0).

$$x_t = w_t + \theta_1 x_{t-1} + \theta_2 x_{t-2} + \ldots.. + \theta_q x_{t-q} \tag{2.11}$$

The combination of AR (p) and MA (q) will form an ARMA (p, q) and it can be shown in Equation 2.12 (Ji & Chee, 2011).

$$x_t = \beta_1 z_{t1} + ... + \beta_p z_{tp} + w_t + \theta_1 w_{t-1} + ... + \theta_q w_{t-q} \tag{2.12}$$

Values of p and q in Equation 2.12 characterise the AR order and MA order, respectively. Furthermore, $\{w_t; t=0, \pm 1, \pm 2, ...\}$ denotes the sequences of white Gaussian noises. The order of ARMA is estimated by using the Partial Correlation (PACF) or Autocorrelation (ACF) techniques. The example plots of PACF and ACF which were obtained from the work of (Ji & Chee, 2011) are illustrated in Figure 2.6 and Figure 2.7. From Figure 2.6 and Figure 2.7, the residual data is reduced abruptly after one lag which means that the values of p and q are equivalent to one for both techniques

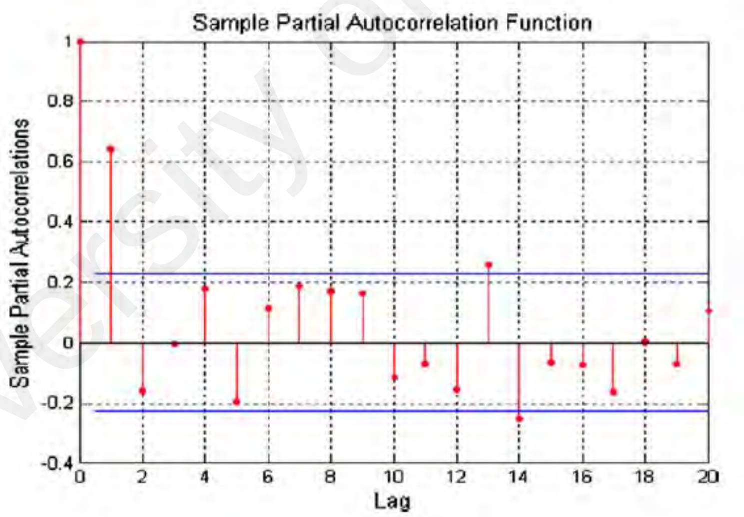


Figure 2.6: Partial Correlation Plot of Residual Data (Ji & Chee, 2011)

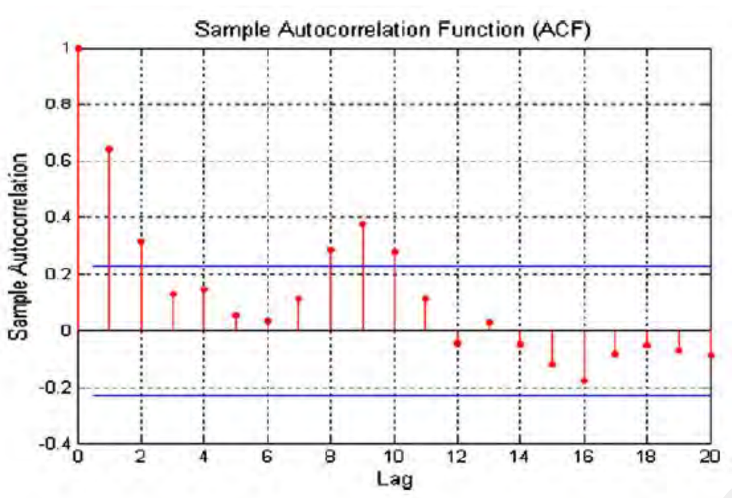


Figure 2.7: Autocorrelation Plot of Residual Data (Ji & Chee, 2011)

Torres et al. forecasted hourly wind speed by using the ARMA model. The daily evolution of wind speed existed in non-stationary nature. Thus, the wind speed data was transformed and standardised before it was used by the ARMA model. The performance of the ARMA model was compared with the persistence model. The result showed that the transformation and standardisation of the original time series data had significantly improved the performance of the ARMA model to forecast the hourly wind speed. This could be proven when the RMSE values of the ARMA model were smaller than the persistence model (Torres et al., 2005).

Ji & Chee developed the ARMA model and Time Delay Neural Network (TDNN) to predict the hourly solar radiation. The data of solar radiation is non-stationary due to the existence of the trend. Therefore, this work employed a detrending approach in producing stationary data. The ACF and PACF were utilised to estimate the order of AR and MA. Later, the ARMA was employed to forecast solar radiation. Meanwhile, the TDNN model was used to forecast values of solar radiation and the result showed that that TDNN model was more sensitive than the ARMA model. To capture the benefit of both models, the ARMA and TDNN were combined. In the combination of ARMA and TDNN models, the ARMA model was used to forecast the linear component of solar radiation while the TDNN model was employed to forecast the non-linear component of solar radiation. The numerical result of this study had shown that stability and accuracy improvement of the combined ARMA and TDNN models for solar radiation forecasting (Ji & Chee, 2011).

Huang and Shih proposed the ARMA model for the short term load forecasting by considering the non-Gaussian process. The cumulant and bispectrum concept were employed to justify the Gaussianity characteristic of the load data. This approach was tested on the practical system and the results were compared with the other forecasting strategies. The result of this work portrayed the supremacy of the proposed ARMA model for short term load forecasting as it gave the smallest error percentage than other forecasting strategies (Huang, S.-J. & Shih, 2003).

(b) ***Autoregressive Integrated Moving Average (ARIMA)***

ARIMA is the generalisation of the ARMA model and it is applied to cases that owning stationary or non-stationary data. For the case that has non-stationary data, the differencing technique which corresponds to 'integrated' part of the ARIMA model is applied to non-stationary data several times to produce the stationary data.

As have been discussed in Section 2.4.2.1 (a), the ARMA model only tolerates with stationary data and a specified method has to be used to transform the non-stationary data into the stationary data. In the case of the ARIMA model, the differencing technique is applied to the non-stationary data and it works by eliminating the trend and seasonality of the data as well as stabilising the mean of the data (Hassan, 2014).

The example of the differencing technique applied on two consecutive data can be shown in Equation 2.13 where y_t denotes the current data, y_{t-1} signifies the first data before the current data and y'_t refers to the data after the first differencing process.

$$y'_t = y_t - y_{t-1} \tag{2.13}$$

If the data is not stationary after the first differencing, the data can be further differenced for the second time. The second time of the differencing process is known as second-order differencing where the expression of this process is shown in Equation 2.14. y_t* represents the data after the second differencing and y_{t-2} denotes the second data before the current data.

$$y_t* = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \tag{2.14}$$

The obtained ARIMA model is constructed as ARIMA (p,d,q) where the term *d* refers to some differences needed for stationarity. After the differencing process, the ACF and PACF are applied to the stationary time series data to obtain the order of *p* and *q* for the AR and MA models, respectively.

Colak et al. created the multi-period prediction of solar radiation by using the ARMA model, the ARIMA model and the persistence model. By comparing all of the multi-period prediction, the ARIMA (2,2,2) provided the outstanding performance to forecast solar radiation. It was then followed by the ARMA (1,2). The persistence method revealed the worst prediction accuracy for all type of forecasting horizons. The simulation result of this work had proven the accuracy of the ARIMA (2,2,2) model over ARMA and persistence models for the forecasting process of solar radiation. (Colak et al., 2015)

Hassan predicted total solar radiation and diffuse solar radiation by using the regression model and the ARIMA model. Statistical analysis of this work had revealed the supremacy of the regression approach as well as ARIMA (2,1,1) for the forecasting of total solar radiation and diffuse solar radiation. The regression approach depicted the lowest RMSE and highest Nash-Sutcliffe Equation (NSE) values. Meanwhile, the ARIMA (2,1,1) revealed the lowest Mean Percentage Error (MPE) and Mean Bias Error (MBE) values. It could be concluded that both models managed to forecast the total solar radiation and diffuse solar radiation efficiently (Hassan, 2014).

Yang et. al applied the ARIMA model to forecast the next hour solar irradiance by utilised Global Solar Irradiance (GSI), Direct Normal Irradiance (DNI), Diffuse Horizontal Irradiance (DHI) and cloud cover as the input parameters. Seasonality and trend in solar irradiance were removed by Local Polynomial Regression Fitting Smoothing (LOESS) technique while irregularity in the solar irradiance was eliminated by decomposition technique. Three forecasting models were proposed with the different combination of input variables. Model 1 utilised GHI to forecast the next hour of GHI. Model 2 forecasted DHI and DNI separately and then combined the results of both forecasts to predict GHI. The last model which was Model 3 considered the cloud cover effect as the input variable. It was found from the result that Model 3 outperformed other models during solar irradiance forecasting (Yang et al., 2012).

(c) ***Multiple Layer Regression (MLR)***

Regression is a process to find a correlation between the response (dependent) variable and predictor (independent) variable. For the MLR model, a relationship between the predictor variables with a response variable is obtained by applying a linear equation. (Sobri et al., 2018). The simplest form of MLR comprises of two predictor variables and one response variable and the expression is shown in Equation 2.15

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \quad (2.15)$$

where Y denotes the response variable and β_0 is identified as the intercept of a plane. Besides, β_1 and β_2 are known as partial regression coefficients. β_1 indicates mean changes in the response variable that corresponds to the unit of x_1 when x_2 is kept constant. On the other hand, β_2 represents the mean changes of the response variable due to a unit change of x_2 when x_1 is being fixed. The notation of ϵ refers to a random error in the forecasting process and it is known as residuals (Amral et al., 2007).

The MLR model is developed according to the following assumptions (Kenton, 2019). The first assumption is made by finding a linear relationship between the response and the predictor variables. Besides, the predictor variables must not overly correlate with each other. Furthermore, the predictor variables are selected independently and randomly from the data population. The last assumption made in this model is to ensure that residuals are normally distributed and has a zero mean value.

There are three major usages of the MLR model (Solutions, 2013). The first usage is to estimate the strength effect of the predictor variables on the response variable. Secondly, this MLR is employed to forecast the effects and impacts of change that gives an understanding of how much the response variable changes when there is some changes occurred on the predictor variables. Thirdly, the MLR model is utilised to forecast the future trend and future values of certain data.

The Coefficient of Determination (R^2) is a typical statistical metric used to measure the variation in the response variable. The increasing number of predictor variables in the MLR model will cause the R^2 to be increased even though the predictor variables are not highly correlated to the response variable. R^2 lies in the value of 0 to 1. The value closer to 1 denotes a better estimation for the regression function to fit the data (Amral et al., 2007).

Kicsiny forecasted the temperature of solar collectors by using MLR-based models that represented a direct relationship between the input and output variables of solar collectors. The performance of the MLR model was compared with a physically-based model. The investigation of this work had given the mean accuracy value of 4.6% for the MLR model and 7.8% for the physically-based model. This showed an outstanding performance of the MLR model to forecast the temperature of the solar collector. To validate the results of the MLR model, the input variables of four different days had been used to forecast the temperature of solar collector and the results had signified the superior performance of the MLR model during the forecasting process (Kicsiny, 2014).

Hong et al. proposed the naïve MLR approach for short term load forecasting. This work included the engineering concept of load forecasting which covered hourly load forecasting, energy forecasting, peak load forecasting, valley load forecasting, peak hour load forecasting and valley hour load forecasting. The average Mean Absolute Percentage Error (MAPE) results for hourly load, daily energy, daily peak load, daily valley load, daily peak hour load and daily valley hour load were 5.01%, 3.52%, 3.97%, 4.98%, 4.29% and 5.48%, respectively. All of the numerical results obtained from the naïve MLR model showed the outstanding performance of the investigated model for the short term load forecasting (Hong, T., Wang, P., & Willis, H. L., 2011, July).

Abuella utilised the MLR model to forecast solar energy. A sensitivity analysis had been conducted to find the most correlated input variables. The result from the sensitivity analysis had shown that the surface solar radiation down, the surface thermal radiation down and the top net solar radiation were the most correlated input variables with solar energy. The simulation results of this work were obtained from three zones and the accuracy of the proposed model in every zone was estimated according to the RMSE values. The RMSE values obtained from Zone 1, Zone 2 and Zone 3 were 0.0725, 0.0741, and 0.0742, respectively, and the average RMSE value for all three zones was 0.0736. This had proven that the data pre-processing technique such as sensitivity analysis of variables had improved the performance of the MLR model for solar energy forecasting (Abuella, 2015, April).

2.4.2.2 Statistical Learning Models

Statistical learning models use machine learning algorithms to understand and to predict the future event of the data. It consists of supervised learning and unsupervised learning. Supervised learning is used to map input data on output data according to the input-output pair (Russell & Norvig, 2010). It has the training data that consists of the input values which is typically a vector and desired output values. It also uses its algorithm to analyse the training data and then produce the model outputs. The most widely used supervised learning models are known as ANN, FIS and SVM.

(a) ***Artificial Neural Network (ANN)***

ANN is an imitation of the cerebral cortex of the human brain and it consists of neurons that are functioning to emit the signals after they receive a strong signal from another neuron (Al-Shamisi et al., 2013).

Despite carrying the same function as the human neurons, the structure of neurons in the ANN is not as complicated as the human neurons. They depend on interconnected layers which are known as an input layer, a hidden layer and an output layer as shown in Figure 2.8. Apart from that, certain types of ANN permit more than one hidden layer.

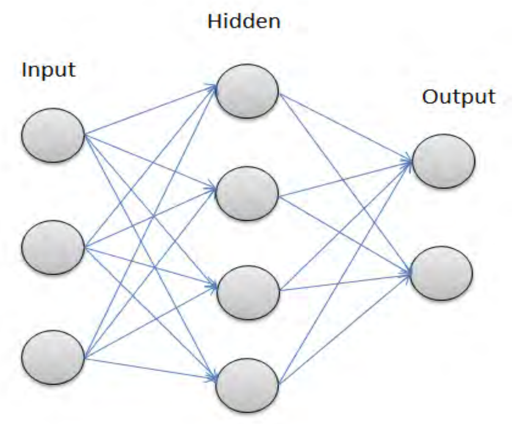


Figure 2.8: General Structure of Neural Network (Tutorials Point, 2017)

Weights and biases are two parameters that assist the interconnection of the neurons and they are updated to the new values until the network achieves the minimum error. The functions that are used to update the weights and biases are shown in Equation 2.16 and Equation 2.17

$$w(k+1) = w(k) + 2\alpha e(k)P(k) \tag{2.16}$$

$$b(k+1) = b(k) + 2\alpha e(k) \tag{2.17}$$

where w represents the weight of network and b denotes the bias vector existed in the network. Besides, α represents the learning rate, e depicts the error rate and P characterises the input vector.

In ANN, two vital processes, namely, the training process and the testing process are conducted. The training process is a process to adjust the values of weights and biases of input neurons until the desired output values are obtained (Ghanbarzadeh et al., 2009; Sulaiman et al., 2012). At the initialisation of the training process, the weights and biases are randomly initialised (Routh et al., 2012). In the case of supervised learning, the actual output values are supplied to the networks. Due to this matter, ANN will tune the values of weights and biases until the generated output values are close to the desired actual output values (Routh et al., 2012). The difference between the generated output values of ANN and desired actual output values is called an error.

The training process also requires a parameter specification where the number of hidden layers and the number of hidden neurons are selected wisely to increase the accuracy of ANN. A proper selection of hidden neurons enables the ANN to easily detect the data pattern as well as to ease the execution of non-linear mapping process on the input and output data (Routh et al., 2012).

The second process of the ANN is known as the testing process where it is a recall process that validates the generalisation capability of the trained network over new and unseen data(Rumbayan & Nagasaka, 2012). After the testing process of ANN, the tested output values are compared with the desired actual values according to regression analysis or other validation error metrics. This comparison will estimate the accuracy performance of the ANN model.

The common problem occurred in the ANN model is overfitting issue where the network has learnt the data well in the training process but it faces the problem of learning the new data (Illias et al., 2015). This means that the errors produced during the training process are small but the insertion of the new data has caused the errors to become bigger in the testing process.

This problem can be prevented by using several useful techniques. One of them is called as early stopping where the training process should be stopped before all weights are been utilised (Fadare, 2009). Another technique to reduce the overfitting problem is to implement the regularisation approach(Doan & Liong, 2004). The idea of this technique is to penalise the ANN model during the training process according to the magnitude of weights. This encourages the network to map the input data on the outputs of training data in such a way to keep a small value of weights and biases.

Rabbi et al. presented the ANN model to predict monthly solar energy. The data from the six cities in Bangladesh was used as the training data while the data from the other two cities in Bangladesh was applied as testing data. The result of this work indicated the suitability of ANN model to be used in the forecasting process of solar energy as the MAPE values obtained by all of the cities in the training and testing phases were less than 7% (Rabbi et al., 2016).

Nandi et al. initiated research that applied the ANN model in solar radiation forecasting. The training data was collected from the website of the National Aeronautics and Space Administration (NASA) and testing data were obtained from the database of the Bangladesh Meteorological Department (BMD). The result of this study depicted the MAPE value of 3.023%, RMSE value of 0.00138 and regression value of 0.99 which indicated a good performance of the ANN model to forecast solar radiation. Apart from that, the authors made a performance comparison between the ANN model and the empirical model. The result showed the supreme performance of the ANN model due to the lower values of RMSE and MAPE. (Nandi et al., 2016)

Celik et al. developed the ANN model to forecast solar radiation. The performance of this model was being evaluated according to types of input parameters and types of training algorithms. Six input parameters were used and they were combined in the different ways before being used by ANN. Other than that, the performance of two types of training algorithms which were Lavenberg-Marquardt algorithm and Scaled Conjugate Gradient algorithm was estimated. The result of this study indicated that monthly mean sunshine duration, monthly mean temperature, altitude and month were the most optimum types of input parameters in the ANN model. Apart from that, the Lavenberg-Marquardt algorithm was superior to the Scaled Conjugate Gradient algorithm because it gave a smaller value of MAPE and higher value of R^2 (Çelik et al., 2016)

Awad & Qasrawi developed the ANN model to predict solar energy cells production. The proposed model of this study used the K-clustering algorithm to determine the centre of ANN. The authors made a performance comparison between the proposed model and the traditional model. The traditional model was a model that utilised the trial and error method to estimate the centre of the ANN. It could be found in the result that the proposed model outperformed the traditional model in the training process and testing process of solar energy forecasting (Awad & Qasrawi).

(b) ***Fuzzy Inference System (FIS)***

In 1975, Zadeh introduced a fuzzy set theory that represents a piece of data information into a degree of membership function that mostly taken in real values of 0 until 1 (Zadeh, 1975). The membership function is a curve that maps every data point to a degree of the membership function. Usually, those data points are mapped between 0 to 1 and the example of an illustration of the fuzzy set theory is portrayed in Figure 2.9. According to Figure 2.9, $\mu_A(x)$ represents the membership function that is assigned to every data point to detect whether a person *(x)* is in the tall *(A)* category or not.

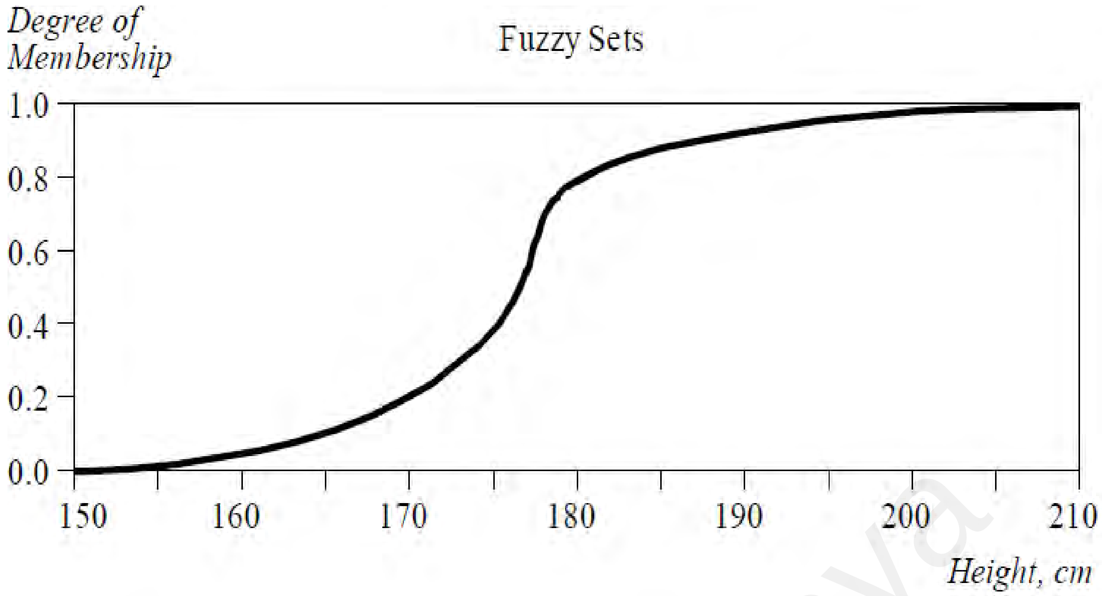


Figure 2.9: Fuzzy Set Mapping of Tall Men (Oentaryo, 2005)

The value of the membership function in Figure 2.9 can be obtained as follows:

$\mu_A(x)=1$ IF x belongs in A category;

$\mu_A(x)$ =0 IF x does not belong in A category;

0< $\mu_A(x)$<1 IF x partly belongs in A category;

where A=210cm

There are various types of membership functions existed and some of them are depicted in Figure 2.10. The selection of an ideal membership function for a fuzzy approach is prepared by experts where the suitability of each membership function is chosen according to the factors of simplicity, speed, convenience and efficiency (Jang & Mizutani, 1996).

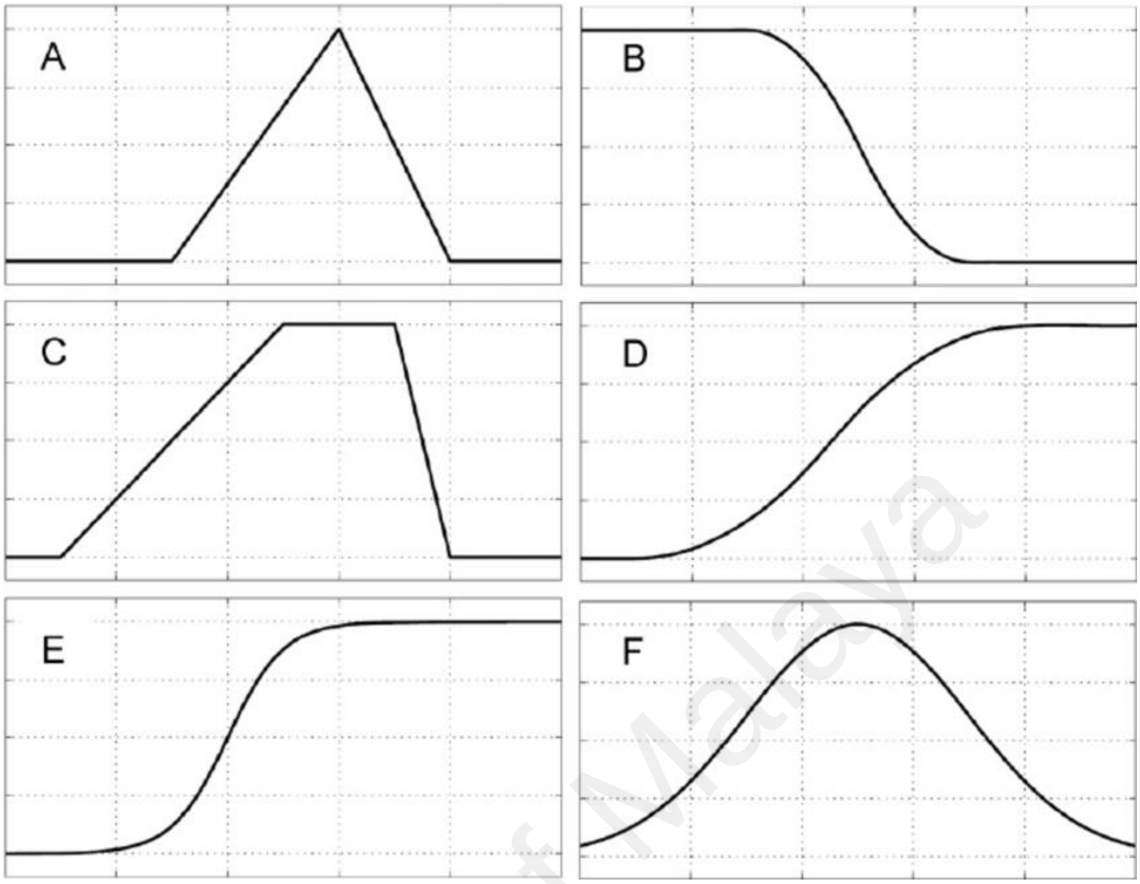


Figure 2.10: Type of Membership Functions: (A) Triangular; (B) Z-shape; (C) Trapezoidal; (D) S-shape; (E) Sigmoid; (F) Gaussian (Rajabi et al., 2010)

The fuzzy set is connected to form the fuzzy rules according to an IF-THEN statement that can be shown as follows:

IF antecedent(s) **THEN** consequent (s)

The antecedent in IF-THEN rule is estimated according to the fuzzification process where a crisp input value is transformed into a fuzzy set value. The antecedent might be more than one in a single fuzzy rule and those antecedents are connected according to various fuzzy operators such as fuzzy intersection (AND), fuzzy union (OR) and fuzzy complement (NOT). The result of the fuzzification process is applied to the consequent part in the process of fuzzy reasoning (Negnevitsky & Intelligence, 2005). The consequent part might be set into a single number. Thus, that single number experienced the defuzzification process to obtain a crisp output from a fuzzy representative value.

In the forecasting area, FIS is a model that utilises the fuzzy set theory to map all input points into an output value. It comprises of two types which are Mamdani and Sugeno method. For the Mamdani type of FIS, the consequent part consists of the fuzzy set and a crisp output is obtained through the defuzzification process of the overall fuzzy set. On the other hand, the Sugeno type of FIS contains the real numbers either linear or constant at the consequent part of the fuzzy rule and it can be shown as follows:

IF x is A and y is B **THEN** z=ax+bx+c

A and B are located in the antecedent part and they are known as fuzzy sets. Meanwhile, ax+bx+c lies at the consequent part and it is a mathematical function that exists in the linear form. The crisp output of this Sugeno type is acquired by averaging the output weight produced at every rule.

The FIS model involves three main processes, namely, fuzzification process, fuzzy reasoning process and defuzzification process (Krishnamoorthy et al., 2012). For the fuzzification process, a suitable membership function is selected to transform the crisp input points into membership grades. Later, the appropriate fuzzy operators are applied in the fuzzy reasoning process to obtain a fuzzy set at the output point. The defuzzification process is then being applied to the fuzzy output point to extract the crisp output value.

Mamlook et al. proposed the hourly ahead of short-term load forecasting by using a fuzzy logic controller to decrease the forecasted error and the processing time. The proposed model utilised weather, time, historical data and random disturbance as the input variables of the fuzzy model. These input parameters were chosen according to their importance to the load forecasting. The forecasted values from the fuzzy model were compared with the conventional forecasting model and the result showed the superior performance of the fuzzy model as it provided more accuracy and better outcome (Mamlook et al., 2009).

Liao & Tsao employed the Evolving Fuzzy Neural Network and Simulated Annealing (AIFNN) in the load forecasting where it was a combination of the Hyper-Rectangular Composite Neural Network (FHRCNN), Evolutionary Programming (EP) and Simulated Annealing (SA). The first part of this work employed the FHRCNN to forecast the load. Later, the EP and SA were merged and used to find the optimal solutions for parameters of FHRCNN such as weights, membership function and sensitivity factor. The merging of EP and SA had given a good opportunity for FHRCNN to search in the globe optimal value and local optimal value. It had been proven from the result that the proposed AIFNN model managed to conduct load forecasting with the greatest accuracy (Liao & Tsao, 2004).

(c) ***Support Vector Machine (SVM)***

SVM is one type of supervised machine learning approach that is used in pattern recognition, regression and forecasting (Chen, J.-L. et al., 2013). It has a learning principle which is called structural risk minimisation and it is used to minimise the generalisation error rather than minimise the training error only (Vapnik, 1999). The fundamental concept of SVM is to map the input vector of x_i into higher-dimensional feature space (hyperplane) according to the non-linear mapping process. After that, the linear solutions that correspond to the non-linear solutions are performed in hyperplane order and it can be formulated as in Equation 2.18 where $\phi(x)$ is the non-linear mapping function, w is the unit normal vector to the hyperplane and it is also known as weight vector. Meanwhile, b refers to the distance from the origin to the hyperplane

$$f(x) = w.\phi(x) + b \tag{2.18}$$

The structure of the SVM can be depicted in Figure 2.11 where input vector x_n is mapped through the non-linear mapping function of $\phi(x)$. The output from the $\phi(x)$ is weighted and applied with the bias to produce a linear output of *f(x)*.

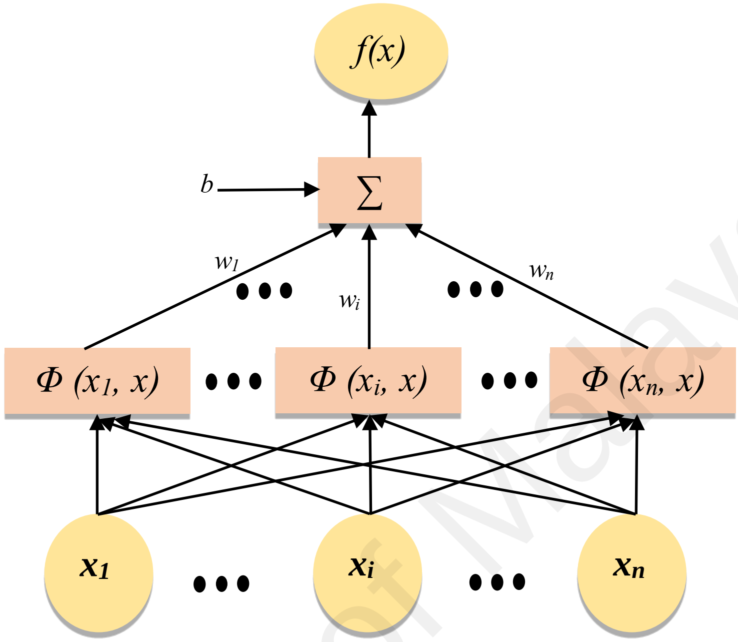


Figure 2.11: Architecture of Support Vector Machine (Eseye et al., 2018)

The optimal values of w and b are estimated by minimising the regularised risk function as shown in Equation 2.19

$$R_{SVM_S}(C) = C\frac{1}{n}\sum_{i=1}^{n} L\varepsilon(d_i, y_i) + \frac{1}{2}\|w\|^2 \tag{2.19}$$

where C signifies error penalty parameter that is used to control the data fitting in the training and generalisation process. n denotes the data size. $L\varepsilon(d_i, y_i)$ signifies the empirical error and it is measured by the function of $L\varepsilon$ that is shown in Equation 2.20 where d_i is denoted as desired value, y_i represents the output value of SVM and ε is known as tube size.

$$L\varepsilon(d_i - y_i) = \begin{Bmatrix} |d_i - y_i| - \varepsilon |d_i - y_i| \geq \varepsilon \\ 0 \leq \varepsilon \end{Bmatrix} \tag{2.20}$$

Meanwhile, the term $\frac{1}{2}\|w\|^2$ in Equation 2.19 refers to the regularisation term. (Zendehboudi et al., 2018).

The value of w is further transformed into a primal function given by Equation 2.21, with the constraint in Equation 2.22 that introduce the positive slack parameters ζ_i and ζ_i*.

$$Minimize R_{SVMS}(w,\zeta_i^{(*)}) = C\sum_{i=1}^{n}(\zeta_i + \zeta_i*) + \frac{1}{2}\|w\|^2 \tag{2.21}$$

$$subjected \begin{Bmatrix} d_i - w\phi(x_i) - b_i + \leq \varepsilon + \zeta_i * \\ w\phi(x_i) + b_i - d_i \leq \varepsilon + \zeta_i * \\ \zeta_i*, \zeta_i \geq 0 \end{Bmatrix} \tag{2.22}$$

The Lagrange method has then been introduced to replace the term of *w* with the Lagrange multiplier (α_i, α_i*) to exploit the optimality constraint. The new function is shown in Equation 2.23

$$f(x,\alpha_i,\alpha_i*) = \sum_{i=1}^{n}(\alpha_i - \alpha_i*)K(x_i,x_j) + b \tag{2.23}$$

$K(x_i,x_j)$ is identified as the kernel function and the value of the kernel function is obtained from the inner product of two vectors x_i and x_j, $K(x_i,x_j) = \phi(x_i)\ \phi(x_j)$. There are four types of basic kernel functions existed and they are identified as linear, sigmoid, polynomial and radial basis functions. Among them, the radial basis function is preferable to be used due to its computational efficiency, reliability and superior adaptation to optimise other adaptive techniques (Sobri et al., 2018). It only requires a solution in term of the linear equation instead of demanding a computationally quadratic programming problem (Shamshirband et al., 2014). The radial basic function of the kernel is defined in Equation 2.24 where σ represents the Gauss parameter that equivalent to the width of the radial basis function. It is used to determine the dominance region of the support vectors in the data space.

$$K(x_i, x_j) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{\sigma^2}\right) \tag{2.24}$$

Bae et al. forecasted solar irradiance by using weather classification as well as SVM regression techniques. In this work, the k-clustering algorithm was employed to determine the number of clusters according to the Silhouette values. Three clusters were chosen and they were identified as sunny days, partially cloudy days and rainy days. The forecasting process was accomplished by the SVM with the eight input variables. Numerical results of this work portrayed the supremacy of the SVM approach for solar irradiance forecast. Besides, this work had shown that only a small amount of Energy Storage System (ESS) installation capacity was needed to absorb the prediction errors. This will decrease the installation cost for a large-scale PV system (Bae et al., 2016).

Piri et al. proposed the SVM model to estimate solar radiation at two different climates that were Bojnoord and Zahedan. Two types of kernel functions were applied to the SVM approach which were known as the polynomial and radial basis function. The numerical result had proven that the polynomial basis function was more qualified to be used at Bojnoord station while the radial basic function was competent to be used at the Zahedan station. All nominated results had strongly proved the capability of the SVM technique for solar radiation forecasting (Piri et al., 2015).

J.-L. Chen et al. proposed the SVM to predict daily solar radiation. Seven SVM models with different input combinations as well as five empirical models were developed for the solar forecast. The simulation results portrayed that all of the SVM models were outperforming the empirical models. Between those newly developed SVM models, the SVM1 model, which utilised the sunshine ratio as the input variable, was preferred to be used due to its greatest accuracy (Chen, J.-L. et al., 2013).

Ekici carried out a study to forecast solar insolation by using the Least Square Support Vector Machine (LS-SVM). The performance of the developed LS-SVM model was estimated by several statistical metrics. From this study, the value of RMSE=0.0043841%, R^2=99.29%, Mean Relative Error (MRE)=9.9617%, Mean Error Function (MEF)=3.3188 and Coefficient of Variance based on Root Mean Square Error (CVRMSE)=0.094611 had shown the capability of the LS-SVM to estimate solar insolation accurately (Ekici, 2014).

2.4.3 Comparison of Forecasting Models

Section 2.4.1 and Section 2.4.2 discuss on various physical and statistical forecasting models. Both physical and statistical models have their advantages and disadvantages that will be summarised in Table 2.2 (Basu & Halder, 2017; Liu, H. et al., 2010; Wang, F. et al., 2012).

Table 2.2: Summarisation of Physical Models and Statistical Models

Model	Advantage	Disadvantage
Physical	• Provides detailed information on the data • Beneficial for long term prediction	• Low forecasting accuracy • Complex processes that need strong computing power • Require appropriate and frequent calibration • Expensive model • Require a longer time to produce a result
Statistical	• High forecasting accuracy • Simple construction and application • Inexpensive model • Shorter computation time for the forecasting process	• Need a large amount of data to estimate model parameters • Suitable for short term prediction

As can be seen in Table 2.2, statistical models offer more advantages than physical models. In the aspect of disadvantages, the statistical models possess some shortcomings in terms of data availability and the forecasting horizon. This work uses eleven months of data which is merely sufficient for statistical models to carry out the forecasting process. Apart from that, this work focuses on the hour ahead of forecasting which is in the category of short term forecasting. Hence, the statistical models are practically suitable to be used by this work for conducting the forecasting process of solar power.

Section 2.4.2 also elaborates about two types of statistical models which are known as statistical non-learning models and statistical learning models. This work considers the benefits and limitations of both models in selecting a model that can provide better accuracy. Those benefits and limitations of statistical learning and statistical non-learning models are summarised in Table 2.3 (Cruz & Wishart, 2006; Wuest et al., 2016).

Table 2.3: Summarisation of Statistical Learning Models and Statistical Non-Learning Models

Model	Advantage	Disadvantage
Statistical Non-Learning	• Easy to implement • Low computational requirement	• Some methods are not applicable for non-stationary data • Not suitable to be used for noisy data
Statistical Learning	• Able to handle high-dimensional and multivariate data • Comfortable adjustment of the parameter to increase the model performance • Suitable for stationary or non-stationary data • Some of the models can tolerate with noisy data • Provide better accuracy	• Difficult to understand the structure of the algorithm • Irrelevant and redundant data will affect the model performance

According to Table 2.3, statistical learning models provide better alternatives for the forecasting process. Though, they have some limitations that can be overcome. Firstly, it is undeniable that algorithms of statistical learning models are more complex than statistical non-learning models. Yet, the accuracy shown by statistical learning models is better than statistical non-learning models. This is the reason why these statistical learning models are preferable in the forecasting area even their algorithms are complicated to be understood. The other shortcoming is related to irrelevant and redundant data that affect model performance. To overcome this matter, this work has utilised several pre-processing techniques to provide reliable data for solar power forecasting.

2.5 Optimisation Techniques

The optimisation process executes a problem and obtains various solutions. Those solutions are then compared to acquire an optimum solution. The employment of an optimisation process will maximise the efficiency of a certain production. In the forecasting area, there are numerous optimisation algorithms are employed to enhance the accuracy of the forecasting models. The only optimisation algorithms that are going to be discussed in this chapter are known as ABC, ACO, GA, PSO and FF.

2.5.1 Artificial Bee Colony (ABC)

ABC algorithm is firstly introduced by Karaboga (2005) that inspired from the foraging behaviour of the honey bees. In nature, foraging is an important process for honey bees to maintain the continuity of life. The foraging behaviours include searching for the most profitable source around the hive, recruiting the other bees to the most profitable source, abandoning the exhausted source as well as finding the new potential rich source (Ozkan et al., 2011). There are three types of foraging bees, namely, employed bees, onlooker bees and scout bees where each type of them is assigned with a specific task. The employed bee is associated with the tasks of finding the most profitable food source and sharing that information with the onlooker bees. The onlooker bees are given a task to choose the best food source and the scout bees are responsible to make a random search to discover a new food source.

- Employed Bees

The employed bees search every food source in the neighbourhood and calculate the important properties of the food source such as the distance of the food source with the hive, the taste of the nectar in the food source, the richness of energy and the difficulty to extract the energy in the food source (Karaboga & Akay, 2009). After the food exploitation, the employed bees return to the hive and exchange the information to the onlooker bees by doing a waggle dance (Hong, W.-C., 2011).

- Onlooker Bees

The onlooker bees watch waggle dances from numerous employed bees and choose the most profitable food source. The selection process of the most profitable food source is depending on the probability value. This means that the onlooker bees choose the best food source that has a probability which is proportional to the quality and quantity of the food source (Awan et al., 2014). After the food selection process, the onlooker bees start to search for the best food source according to the information given by employed bees.

- Scout Bee

The scout bees are associated with the abandoned food source and the random replacement of food source with the new one. To carry out the replacement process, the employed bees have turned into scout bees (Hong, W.-C., 2011). The position of the food source that is not updated for a certain number of cycles is considered as the abandoned food source and scout bees will replace it with a new one.

Fei & He proposed a hybrid model of wavelet decomposition and ABC algorithm-based relevance vector machine to forecast the wind speed. The first part of the work had decomposed the wind speed signal by using wavelet decomposition process. After that, the decomposed wind speed signal was used by the RVM to forecast the future wind speed and ABC was used to select the optimum kernel parameters for the RVM model. The experimental result of the work showed the feasible performance of the proposed model when it was compared with other forecasting strategies (Fei & He, 2015).

Awan et al. conducted a short-term load forecasting by hybridising the ANN model with the ABC algorithm. In this work, the ABC was used as an alternative learning scheme to obtain the optimised set of neuron connection weights for the ANN. The combination of ANN with ABC had portrayed the improved convergence rate of ANN This signified that ANN was not getting trapped in the local minimum. Besides, the performance of ABC was compared with other optimisation algorithms and the result showed the superior performance of ABC for wind speed forecasting (Awan et al., 2014).

Gürbüz et al. forecasted the energy consumption by using three forecasting models which were linear model, quadratic model and ANN model. In the linear and quadratic models, the ABC algorithm was used to find the ideal values of energy consumption parameters while the ANN had employed ABC to find the optimum weight values. All nominated results had strongly proved the capability of the ABC algorithm to optimise the parameters of linear, quadratic and ANN models (Gürbüz et al., 2013).

W.-C. Hong Hong presented a hybrid model that combined the Recurrent Neural Network (RNN), Support Vector Regression (SVR) and Chaotic Artificial Bee Colony (CABC) to examine their potentiality for electric load forecasting. The proposed hybrid model was identified as the Seasonal Recurrent Support Vector Regression Model with Chaotic Artificial Bee Colony (SRSVRCABC). The utilisation of CABC in

SRSVRCABC will overcome the premature local optimum problem of the forecasting model. Apart from that, the performance of two alternatives models, namely, the ARIMA model and TF-ε-SVR-SA model were compared with the SRSVRCABC model and the result from this work had indicated the significant superiority of the proposed SRSVRCABC among other alternatives in terms of forecasting accuracy (Hong, W.-C., 2011).

2.5.2 Ant Colony Optimisation (ACO)

ACO is introduced by Dorigo and Gambardella that mimics the foraging behaviour of the ant colonies (Dorigo et al., 1996). Communication between the ants is carried out by leaving the chemical trails called pheromone on their way of finding the food source. When an isolated ant found the pheromone, there is a huge probability for that isolated ant to follow it, and thus marks another trail with its pheromone. This foraging behaviour of the ants is known as the autocatalytic behaviour and the increasing number of ants that follows the trail will cause the trail to become more attractive (Dorigo et al., 1996).

The autocatalytic behaviour of the ants is further explained by the experimental setting that is illustrated in Figure 2.12. As can be seen in Figure 2.12 (a), the ants move along the path from food source A to nest E and vice versa. Suddenly, as in Figure 2.12 (b), an obstacle appears and blocks the pathway. Therefore, the ants that move from food source A to nest E will stop at point B and make a decision either to turn left or right. Same goes to the ants that move from nest E to food source A, where they stop at point D and they must decide either to turn left or right. Their choice to turn left or right is highly influenced by the intensity of pheromone left by the previous ants. Figure 2.12 (c) has clearly illustrated that the majority of the ants are turning right. It is due to the high level of pheromone that gives a stronger stimulus to the ants to move to the right (Toksarı, 2007).

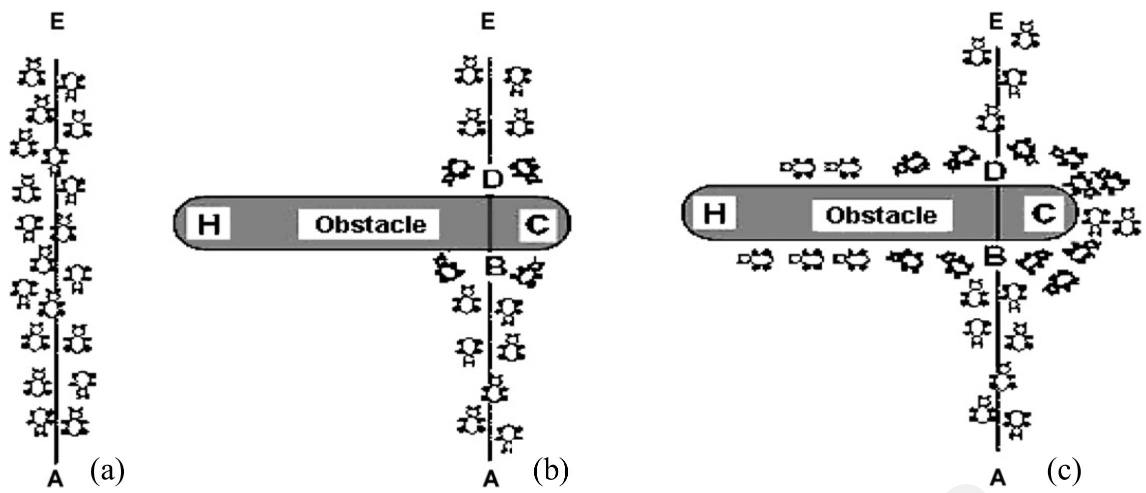


Figure 2.12: Ant Behaviour: (a) Ants move in the path between point A and E; (b) Obstacle is interposed along the path; (c) Ants choose the shorter path as more pheromone is laid on the ground (Toksarı, 2007)

Note that the first ant that reaches point B or point D in Figure 2.12 (b) is called as an isolated ant and it can make a choice either to turn left or right as there is no trail of pheromone left by the preceding ant. The first ant that chooses path BCD will reach point D faster than the ant that chooses the path BHD. This is due to the reason that the path BCD is shorter than the path BHD. The ants that move from the nest of point E to point D find out that there is a stronger pheromone trail on the path of DCB. Thus, there is a high probability for those ants to choose the path DCB instead of DHB. As a result, the number of ants that follow the path BCD or DCB is higher than the number of ants that follow the path BHD or DHB. Thus, the quantity of pheromone left on the shorter path is more than the longer path (Toksarı, 2007).

Rahmani et al. hybridised the ACO with PSO in the forecasting process of energy at the wind farm. To forecast the wind energy, a mathematical model that consisted of the S-curve and the parabola functions had been utilised. The hybrid of ACO and PSO (HAPE) was employed in a six-dimensional solution space in that mathematical model. The utilisation of HAPE was estimated to increase the quality of the results with faster convergence profile. The result obtained from HAPE was compared with single PSO and single ACO which were applied to the same mathematical model. Result of this work indicated that the HAPE model could forecast the wind energy better than single PSO and single ACO algorithms (Rahmani et al., 2013).

Similar to Rahmani et al., the authors in Kiran et al. proposed a hybrid algorithm of ACO and PSO (HAPE) to forecast the energy demand in Turkey. In this work, the hybrid algorithm had been applied to the linear model and the quadratic model. The results obtained from HAPE provided better relative error estimation for both linear and quadratic models (Kıran et al., 2012).

Li & Han had employed an Improved Ant Colony Clustering (IACC) for short-term load forecasting. In the IACC model, each load data was represented by an ant. After that, the parallel optimisation characteristics of ACO and the volatile quotient method were used to change the amount of information. The utilisation of IACC had improved pheromone concentration on every path and enhanced the heuristic function to accelerate the searching process. It had been proven from the result that IACC managed to forecast the short-term load efficiently (Li & Han, 2008).

2.5.3 Genetic Algorithm (GA)

GA is firstly introduced by John Holland in 1960 where it is based on the evolutionary natural-based algorithm that inspired from natural selection, the evolution of chromosomes and survival of the fittest in the biological world (Jadidi et al., 2018; Platt et al., 2010). It is originally a binarily coded algorithm that uses the possible solutions in terms of ones and zeros when modelling the search problem. The bit-string of the binary code is a representation of parameters in the search problem and GA is employed to maintain the population of the knowledge structure by introducing the suitable candidate solutions. The population of set rules are then evolved for several generations to improve the performance of the population.

In the optimisation area, GA starts to optimise a search problem by randomly generating numbers of chromosomes where they are called population. Each chromosome in the population represents an individual solution to the search problem. At this initial stage, the fitness value of every chromosome is evaluated. Later, the parent chromosomes are selected according to the several techniques such as a roulette-wheel selection, a tournament selection as well as an elitist selection (Jadidi et al., 2018).

The selected parent chromosomes are being applied with the crossover process and mutation process to produce the new chromosomes which are known as offsprings. The crossover is a process to divide and exchange the parent chromosomes and several crossover techniques can be used in the GA algorithm. Some of them are identified as a uniform crossover, single-point crossover, two-point crossover and linear crossover (Adewuya, 1996). On the other hand, the mutation is a random process that changes a part of the parent chromosomes based on the defined mutation rate. The result of the mutation process will cause a random change for solution exploration. Various mutation approaches are used where they are known as a uniform mutation, boundary mutation, Gaussian mutation and non-uniform mutation. Note that the non-existence of the mutation process will cause the GA algorithm to converge rapidly at the local optimum and causing the model inaccuracy.

Jadidi et al. employed a combination of different machine learning techniques to forecast GHI. In the work of Jadidi et al., Density-Based Spatial Clustering of Applications with Noise (DBSCAN) was used to detect and to remove anomalies. Non-dominated Sorting Genetic Algorithm II (NSGA II) was then utilised in the process of feature selection. A forecasting model known as MLP was employed to forecast GHI and two optimisation algorithms which were PSO and GA were used to tune the parameters of MLP. The result when utilising PSO and GA was compared and it was realised that GA provided better performance than PSO (Jadidi et al., 2018).

Aybar-Ruiz et al. presented a novel approach that merged the Grouping Genetic Algorithm (GGA) and ELM algorithm to forecast solar radiation. In the proposed work, GGA made the optimal feature selection and ELM was applied as the forecasting model. The proposed hybrid model (GGA-ELM) was used in different cases. For the first case, the GGA-ELM used the output from Numerical Weather Meso-scale model (WRF). The second case employed GGA-ELM to predict solar radiation at different time tags but using the predictive variables from WRF. The last case made a complete prediction by including the previous values of solar radiation and the outputs of WRF. The result of the proposed work had shown excellent performance for all cases according to different error metrics (Aybar-Ruiz et al., 2016).

Liu et al. developed a hybrid algorithm for short-term wind speed forecasting. In the first part of work, WT was used to remove random fluctuation of the wind speed and the signal obtained by WT was introduced to the SVM model to establish the forecasting process. Apart from that, GA was proposed to find the optimum parameters of the SVM model. The proposed work also included the Granger causality test to choose proper lags temperature data. Furthermore, ACF and PACF were employed to select proper lags of historical wind speed. The performance of the hybrid model was compared with the persistence model and SVM-GA model without WT. The result of this work indicated the suitability of the proposed model to be used in the forecasting process of wind speed when it was compared to other forecasting strategies (Liu, D. et al., 2014).

2.5.4 Particle Swarm Optimisation (PSO)

PSO is an evolutionary type algorithm that was introduced by Kennedy and Eberhart in 1995. This algorithm is inspired by the social behaviour of fish schooling or bird flocking (Rahmani et al., 2013). Typically, animals in a flock move randomly and tend to follow the member that has the nearest position to the food sources. The information about the position of the food source is obtained through communication with the member that has the best position with the food source. The process of exchanging information repeats until the food source has been discovered.

The PSO algorithm uses the nature work of this animal society in finding the best position (optimal value) of an optimisation problem. In PSO, an individual that exists in a population is called a particle. Each of them conducts two important processes which are the exploration process and exploitation process to find the best position (food source). The exploration is the ability of a particle to explore several search space to find the best position. On the other hand, exploitation is the capability possessed by a particle to concentrate on the allocated position to refine the candidate solution. With these two processes, a particle manages to fly to the search space area and to produce two reasoning abilities that are known as Personal best (P_{best}) solution and Global best (G_{best}) solution. The determination of a solution for PSO is influenced by various other parameters which are explained as follows:

- **Personal best (P_{best}):** P_{best} value in the PSO algorithm is defined as the best position or the best fitness value that has been attained by every particle in a population (Bashir & El-Hawary, 2009). In every iteration, a particle updates the P_{best} value by comparing the previous P_{best} and the current P_{best}. If the current P_{best} is better than the previous one, the PSO algorithm will update the P_{best} value according to the current value and the previous value will be omitted.

- **Global best (G_{best}):** The G_{best} is the solution reached by the best position among entire particles in a population (Catalão et al., 2011). A particle moves randomly in the search space and follows another member that has a position nearest to the food source. Thus, the member that has the best position (position near to the food source) acts as an attractor that entices other particles to move to that best position. Eventually, all of the particles converge to that best solution which is known as G_{best} solution. In the PSO algorithm, G_{best} value needs to be updated regularly to reduce the premature convergence of a population.
- **Velocity Clamping:** Velocity clamping parameter is functioning to control the global exploration process of a particle (Rini et al., 2011). It is important to note that there will be a maximum number of allowable velocity (V_{max}) for a particle to travel. Hence, any particle travels with a velocity beyond the limit set by V_{max} will be set with the V_{max} value (Rini et al., 2011). The equation of velocity clamping is shown in Equation 2.25 where ω denotes the inertia weight, v_i (t-1) represents the previous velocity of a particle, C_1 and C_2 signify the cognitive parameter and social parameter, respectively. r_1 and r_2 indicate the uniformly distributed number in the range of 0 to 1 and x_i(t-1) represents the previous position of a particle.

$$v_i(t) = \omega v_i(t-1) + C_1 r_1 (P_{best} - x_i(t-1)) + C_2 r_2 (G_{best} - x_i(t-1)) \qquad (2.25)$$

- **Cognitive Parameter (C_1) and Social Parameter (C_2):** C_1 and C_2 are the acceleration weights that attract a particle towards the values of P_{best} and G_{best} (Mohandes, 2012). The small values of both parameters will cause a particle to roam away from the P_{best} and G_{best} while the large values cause the abrupt movement of a particle towards the target position. Thus, the values of C_1 and C_2 must be ideally selected.

- **Inertia Weight (ω):** The inertia weight will control the exploration and exploitation processes of a particle by dynamically adjusting its velocity. The adjustment is made by controlling the effect produced by the previous velocity on the current velocity. A large value of inertia weight facilitates global exploration while a small value smooths the local exploration of the population (Ren et al., 2014). An ideal value of inertia weight provides a balance between global and local exploration (Bashir & El-Hawary, 2009). Due to this matter, the initial value of inertia weight is set to a larger value to enable a particle to make global exploration. Later, the value of inertia weight is decreasing gradually to enable local exploration to occur (Mohandes, 2012). The value of inertia weight is decreasing linearly according to the expression shown in Equation 2.26

$$\omega = \frac{\omega_{\max} - \omega_{\min}}{Itr_{\max}} Itr \tag{2.26}$$

where $\omega_{\max}$ represents the final inertia weights and $\omega_{\min}$ denotes the initial inertia weights. $Itr_{\max}$ signifies the maximum number of iteration and Itr indicates the current iteration number.

Ren et al. proposed an IS-PSO-BP model that combined a Backpropagation (BP) neural network, PSO and input parameter selection (IS) for the wind speed forecasting. The authors had utilised two methods of IS which were lateral data selection and longitudinal data selection and the performance of each method was compared. The comparison outcome revealed the supremacy of longitudinal selection method over the lateral selection method. After that, the chosen input variables were presented to the PSO-BP for the wind power forecasting. The final result showed an outstanding performance of IS-PSO-BP to forecast wind power data better than ARIMA and basic BP models(Ren et al., 2014).

Bahrami et al. presented the short term load forecasting by the application of WT, grey model and PSO approach. The WT was used to eliminate the high-frequency components of the previous day, the grey model was utilised as a forecasting model and PSO was employed as an optimisation algorithm of the grey model. To verify the efficiency of the work, the performance of the proposed model was compared with previous methods studied by the authors. Simulation results had shown a favourable performance of the proposed model that combined WT, grey model and PSO approach for the load forecasting (Bahrami et al., 2014).

Mohandes carried out a study to estimate the monthly mean of GSR by using a combination of PSO and ANN. The performance of the PSO-ANN was compared with the empirical model. The numerical result of this work revealed the superiority of PSO-ANN as it managed to forecast the monthly mean of GSR with minimum MAPE value of 3.4% compared to the empirical model that gave the minimum MAPE value of 5.7% (Mohandes, 2012).

2.5.5 Firefly Algorithm (FF)

FF is a stochastic metaheuristic search algorithm which was designed by Yang in 2007 (Jallad et al., 2018). It mimics the characteristic of fireflies that emit the short flashes to communicate as well as to hunt (Aydilek, 2018). Typically, all of the fireflies are unisex and they tend to follow another firefly that is more attractive and has greater brightness regardless of the sex. However, they tend to move randomly in a population if there is no brighter and more attractive firefly existed (Arunachalam et al., 2014).

For the implementation of FF algorithm in optimisation area, each firefly is called a particle in the swarm. Every firefly has its brightness which represents a candidate solution in the dimensional search problem. The position of the brightest particle is assumed to have a better solution. Due to this, the FF algorithm will assist other fireflies to find the brightest position of a firefly in the swarm.

The light intensity *(I)* of a firefly is proportional to its brightness and inversely proportional to the distance between a firefly with a light source *(r)*. The relationship between *I* and *r* can be represented in Equation 2.27 where I_S denotes the intensity of the light source. According to Equation 2.27, an increasing distance of a firefly with the light source will cause *I* to decrease. Hence, the brightness degree of firefly will be decreased as well.

$$I(r) = \frac{I_s}{r} \tag{2.27}$$

The attractiveness of firefly can be determined according to its brightness. This means that the most attractive firefly is the one that has the highest degree of brightness and it is powerful enough to attract other fireflies to its position. Apart from the brightness factor, other parameters influence the attractiveness of a firefly and those parameters can be shown in Equation 2.28.

$$\beta(r) = \beta_0 e^{-\gamma r_{ij}^2} \tag{2.28}$$

The brightness of a firefly at $r = 0$ is represented by notation β_0. The other parameter which is a coefficient of the light absorption is signified by notation γ. The Cartesian distance between two fireflies is expressed in Equation 2.29 and it is indicated by notation r_{ij}. In this work, two types of fireflies are named as firefly x_i and firefly x_j. As can be seen in Equation 2.29, $x_{i,k}$ is the k^{th} component of the position vector for firefly x_i while $x_{j,k}$ is the k^{th} component of the position vector for firefly x_j. The notation *d* in Equation 2.29 denotes the dimensionality of the problem in the search space.

$$r_{ij} = \|x_i - x_j\| = \sqrt{\sum_{k=1}^{d} (x_{i,k} - x_{j,k})^2} \tag{2.29}$$

By assuming that firefly x_j is brighter than firefly x_i, the movement equation of firefly x_i towards the position of firefly x_j can be shown in Equation 2.30 where the first term of equation portrays the previous position of the firefly x_i. The second term in Equation 2.30 represents the attractiveness equation as in Equation 2.28. The third term in Equation 2.30 governs the randomisation parameter (α) and a random number (rand) that spread in the range value of 0 to 1 (Kora & Krishna, 2016).

$$x_{i,k}(t+1) = x_{i,k}(t) + \beta_0 e^{-\gamma r_{ij}^2}(x_{j,k}(t) - x_{i,k}(t)) + \alpha(rand - 0.5) \quad (2.30)$$

Ibrahim & Khatib proposed a hybrid model of random forest and FF algorithm to predict hourly GSR. The FF algorithm was employed to optimise the number of trees and leaves per tree in the random forest approach. The experimental result of the work was compared with three other methods of forecasting, namely, optimised artificial neural network model by firefly algorithm, conventional artificial neural network and conventional random forest model. It was realised from the result that the proposed model provided better performance than other forecasting models in terms of the prediction speed as well as prediction accuracy (Ibrahim & Khatib, 2017).

Olatomiwa et al. hybridised the SVM with FF in the forecasting process of monthly mean horizontal GSR. To forecast the horizontal GSR, three meteorological variables which were sunshine duration, maximum temperature and minimum temperature were used as the inputs of the SVM model. The result obtained from hybrid SVM-FF was compared with ANN and Genetic Programming (GP) models. Result of this work indicated that the hybrid SVM-FF provided an accurate prediction than other aforementioned models which signified the efficiency of hybrid SVM-FF model as a forecaster of horizontal GSR (Olatomiwa et al., 2015).

Haque et al. presented the short term forecasting of PV power by the application of hybrid WT technique, fuzzy ARTMAP model (FA) and FF algorithm. The WT was used to eliminate the ill-behaved feature of PV power data, the FA network was utilised to capture the fluctuation of the non-linear PV power signal and FF was employed as an optimisation algorithm of the FA model. To verify the efficiency of the work, the performance of the proposed model was compared with several forecasting methods which were RBFNN, Backpropagation Neural Network (BPNN), General Regression Neural Network (GRNN), hybrid WT and BPNN model, hybrid WT and RBFNN model, hybrid WT and GRNN model and hybrid WT and FA model. Simulation results had shown a favourable performance of the proposed model that combined WT, FA and FF compared to other forecasting alternatives (Haque et al., 2013).

2.5.6 Comparison of Optimisation Techniques

The optimisation techniques that are discussed in Subsection 2.5.1 until Subsection 2.5.5 have their benefits and limitations when optimising a problem. The benefits and limitations of each optimisation technique can be summarised in Table 2.4 (Adrian et al., 2015; Aydilek, 2018; Fu et al., 2015; Gerhardt & Gomes, 2012; Gupta & Gupta, 2014; Pei et al., 2019; Yan & Li, 2011).

Table 2.4: Summarisation of Optimisation Technique

Method	Advantage	Disadvantage
Artificial Bee Colony	• Fast convergence • Has a good ability for local search • Fewer setting parameter	• An accurate solution cannot be discovered rapidly • Require high computational effort • Requires high evaluation of the objective function • Requires new fitness test for the new algorithm parameters • Tends to face the premature convergence
Ant Colony Optimisation	• Has a memory • Efficient to solve discrete problems • A good solution can be discovered rapidly	• Tends to face the premature convergence • Has a weak ability for local search • Not effective to solve the continuous problem • Probability distribution tends to change at each iteration
Genetic Algorithm	• Information can be exchanged via mutation or crossover • Can be used efficiently to solve continuous problems	• Does not has any memory • Tends to face the premature convergence • Has a weak ability for local search • Require high computational effort
Particle Swarm Optimisation	• Has a memory • Easy to be implemented as it uses simple operators • Efficient to solve continuous problems	• Tends to face the premature convergence • Has a weak ability for local search
Firefly Optimisation Algorithm	• Short computational time • Fewer setting parameter • Less likely to face premature convergence	• Sometimes it is trapped in local optima • No memorising capability • Gives poor performance in high dimensional problem

As can be seen in Table 2.4, all of the optimisation techniques tend to have their benefits and limitations. The utilisation of a single algorithm in the optimisation area is restrictive on the non-linear and high dimensional problems (Ali, 2014). Due to this matter, a hybridisation of optimisation algorithms is highly suggested to mitigate the above-mentioned problem (Abdullah et al., 2012). In this work, the PSO and FF algorithms have been hybridised to optimise the parameters of the forecasting model. The utilisation of hybrid PSO and FF is called HFPSO algorithm and it is estimated to increase the performance of the forecasting model better than utilisation of single PSO and single FF algorithms in the optimisation problem.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Introduction

This study aims to forecast hourly solar power data by using various statistical learning models. Solar power forecasting is carried out by collecting solar power data and several meteorological data. The former is used as the corresponding output of statistical learning models while the latter is utilised as input variables of statistical learning models. The amount of data collected may consist of noise due to changes of instruments as well as the presence of dirt on the sensor of measurement (Renani et al., 2016). The noisy data will degrade the performance of forecasting models. Thus, it is very important to remove the noisy data before further analysis can be carried out. In this work, a noise elimination technique which is known as WT is introduced and the usage of WT is deemed to enhance the performance of forecasting models.

The denoised data (solar power and meteorological variables) from WT is introduced to several statistical learning models which are MLP, RBFNN and ANFIS. The historical data of solar power and meteorological variables are presented to the above-mentioned forecasting models during the training process. The training process of every forecasting model is done iteratively until the network is converged. The testing process is then conducted where the future data of the meteorological variables are inserted to the trained forecasting models to forecast future data of solar power. Note that this work compares the performance of every forecasting model to select a model that offers the highest accuracy.

Commonly, the parameters of the forecasting model are determined by the users through the trial and error method which requires more time and efforts. To overcome this problem, an optimisation algorithm can be utilised to select the optimum values of the respective parameters. This work proposes an optimisation technique which is known as HFPSO to optimise the selected parameters of the most accurate forecasting model which is chosen beforehand.

3.2 Proposed Forecasting Strategy

The strategy employed by this study is depicted in Figure 3.1.

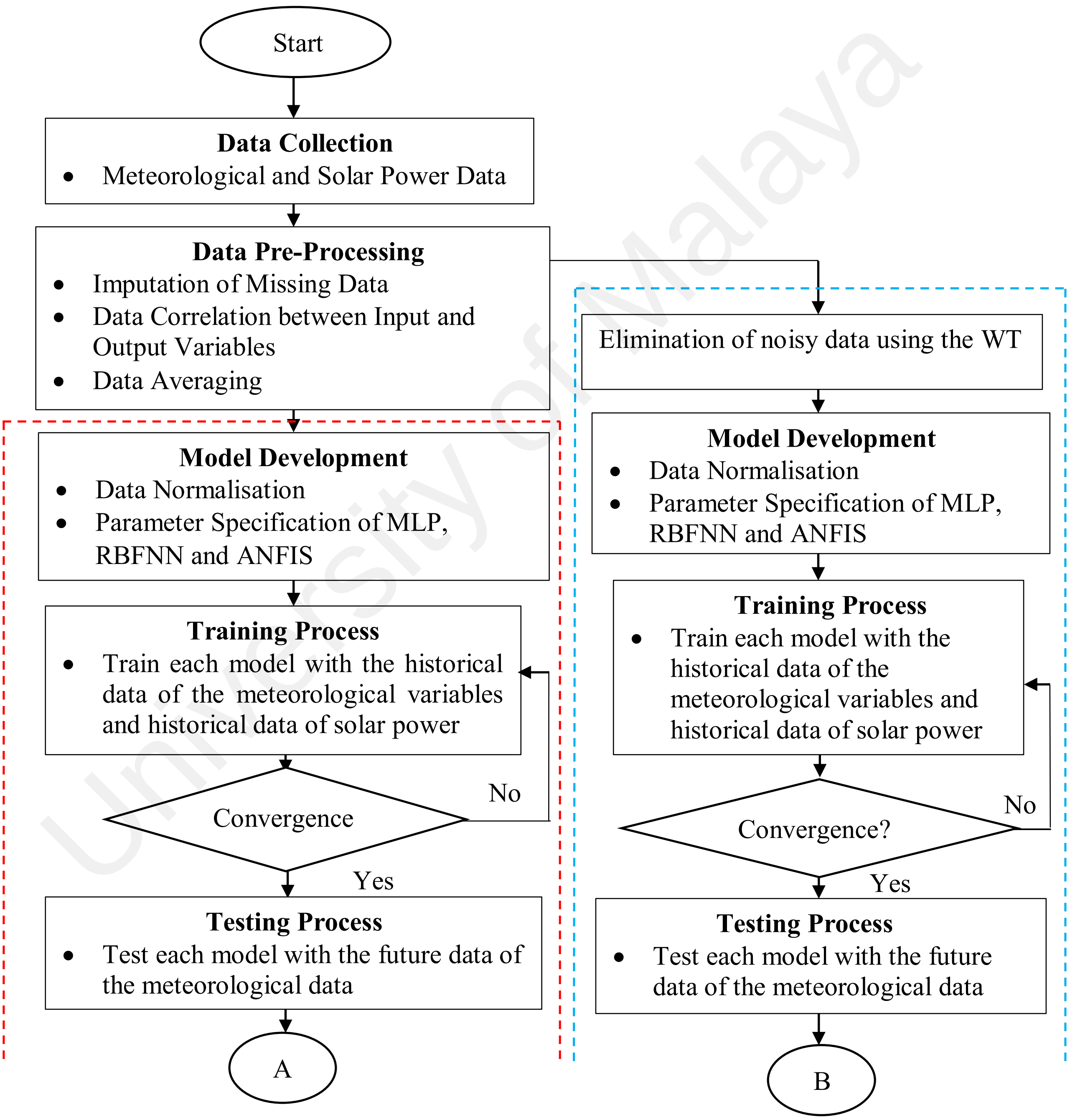


Figure 3.1: Proposed Forecasting Strategy

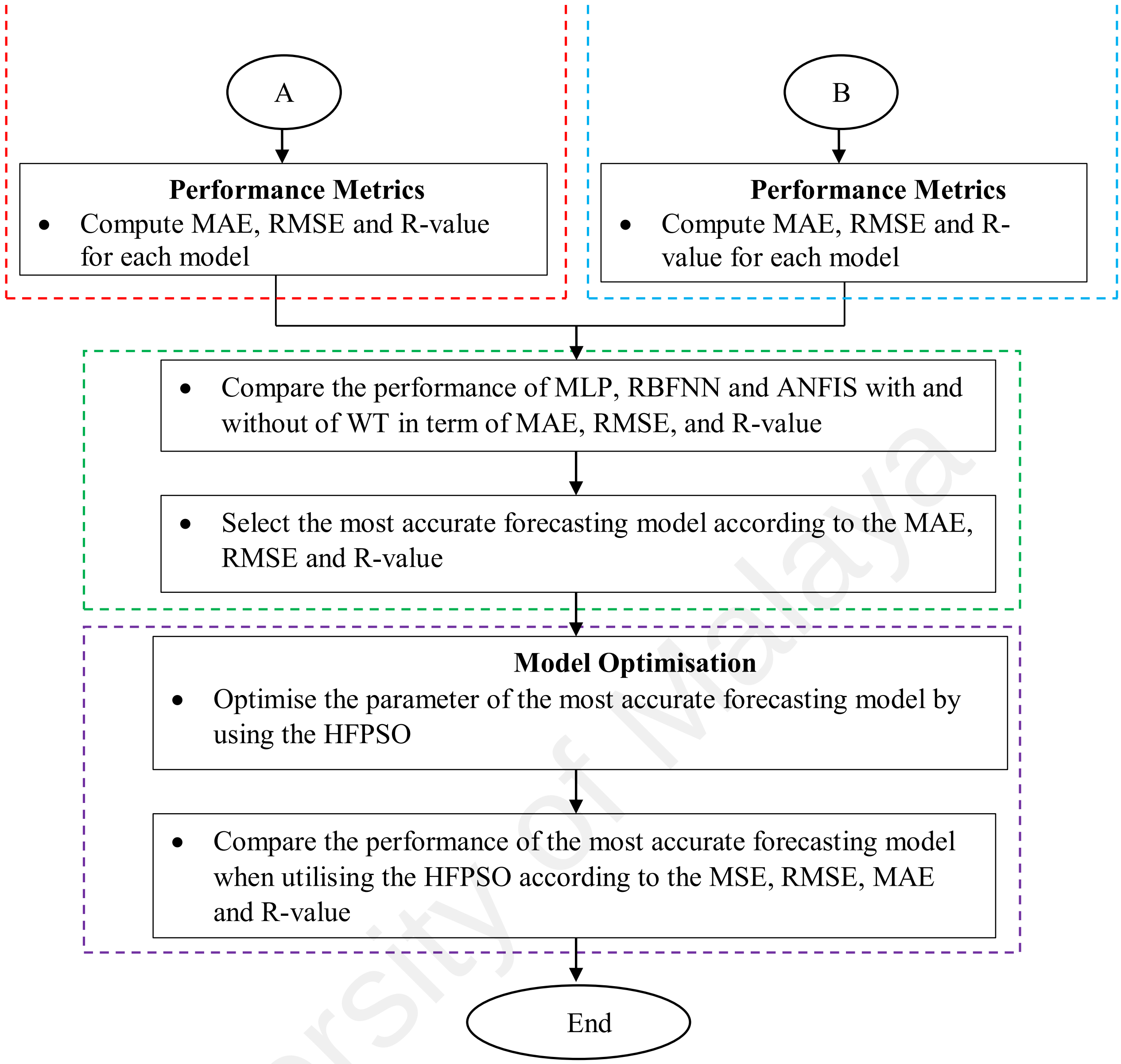


Figure 3.1, continued

The methodology of this work covers several phases. For Phase 1 (red dashed line), the MLP, RBFNN and ANFIS models receive the noisy data when conducting the forecasting process. On the other hand, Phase 2 (blue dashed line) provides the performance of MLP, RBFNN and ANFIS when utilising the data that has been denoised by WT. Aforementioned, the utilisation of WT to remove the noisy data is deemed to increase the performance of the forecasting model. To prove this matter, the performance of MLP, RBFNN and ANFIS in Phase 1 and Phase 2 is compared and the comparison result is shown in Phase 3 (green dashed line). Other than that, Phase 3 covers the selection process of the most accurate forecasting model by comparing the Mean

Absolute Error (MAE), RMSE and Correlation of Coefficient (R) of every forecasting model in Phase 1 and Phase 2. Lastly, Phase 4 (purple dashed line) optimises the parameter of the most accurate forecasting model by using the HFPSO approach. Phase 4 also includes the comparison result of the most accurate forecasting model when utilising an optimisation algorithm of HFPSO, single PSO and single FF during solar power forecasting.

3.3 Data Collection

The data of this work is collected from the PV system installed on the rooftop of admin and laboratory building at the Faculty of Electrical Engineering in UTeM. The PV system is located at a latitude of 2.3°N and a longitude of 102.3° E. The installation setup of the PV system can be shown in Figure 3.2 where the tilting angle of every module is at 10°.

Figure 3.2: PV System Installation Setup on the Rooftop of Admin and Laboratory Building at the Faculty of Electrical Engineering in UTeM

The data variable of this study consists of solar power data which is obtained from the solar monitoring system. Apart from that, several meteorological variables such as humidity surrounding, wind speed, PV module/ PV panel temperature, global radiation, temperature surrounding and tilted radiation are acquired from the weather stations as shown in Figure 3.3. Solar power and meteorological variables are associated with historical data for eleven months starting from the 1st January 2017 until 30th November 2017 with a time step of 1-minute resolution. For every variable, the time interval of the data is obtained from 8 A.M. until 7.00 P.M.

Figure 3.3: Weather Station Installed on the Rooftop of Admin and Laboratory Building at the Faculty of Electrical Engineering in UTeM

3.4 Data Pre-Processing

The data pre-processing is a technique to transform raw data into a simplified version of data. In this work, the data pre-processing technique can be categorised into three stages. The first stage relates to a process of imputing the missing values, the second stage is to find a correlation between every meteorological variable with solar power and the next stage is the data averaging process. All of the above-mentioned stages can be further described in the following subsection 3.4.1, subsection 3.4.2 and subsection 3.4.3

3.4.1 Missing Data Imputation

The missing data is data of a certain variable that cannot be stored in an observation. One of the common occurrences of missing data is due to the sensor malfunction. The phenomenon of missing data gives a significant effect on data reliability and contributes to several problems. This phenomenon will introduce bias in the data which causes the data handling and data analysing to become more complicated. Other than that, the missing data causes the data insufficiencies for forecasting as certain forecasting models automatically remove the missing data case. Some forecasting models perform the analysis of the missing data case but provide insignificant result due to a small amount of data usage.

Imputation of missing data is one of the solutions to overcome these problems. It is a process to replace the missing values of observation with the estimation values of the available information. Some of the previous works imputed the missing data by averaging the previous values and future values of missing data (Akarslan & Hocaoglu, 2016; Hocaoglu & Serttas, 2017). This future-previous imputation attenuates the correlation between imputed values and measured values which makes it less preferable to be utilised.

Due to this matter, this work imputes the missing data according to an interpolation process. Among several types of interpolations processes, the linear interpolation is a simplest and the quickest method for missing data imputation. It is a popular method that is utilised by previous works such as (Cornaro et al., 2015; David et al., 2016; Elsinga & van Sark, 2017; Gutierrez-Corea et al., 2016) to impute the missing values. Linear interpolation creates a new data point by fitting a linear line between a discrete set of known data points. The equation of linear interpolation is shown in Equation 3.1

$$\frac{y-y_1}{x-x_1}=\frac{y_2-y_1}{x_2-x_1}; \quad (3.1)$$

where (x_2, y_2) and (x_1, y_1) are two coordinates of known data points while (x , y) signifies the coordinate of missing value

3.4.2 Correlation between Meteorological Variables and Solar Power

Association of two variables can be measured according to a degree of correlation that varies from the strongest until the weakest as shown in Figure 3.4. The correlation value of -1 between two variables signifies a perfect negative correlation while the correlation value of 1 indicates a perfect positive correlation. The negative correlation shows the opposite trend between investigated variables. This means that whenever the values of one variable increases, the values of other variable decreases and vice versa. On the other hand, the positive correlation denotes the same trend of data for both variables. As the value of one variable is increasing, the value of another variable is increasing as well.

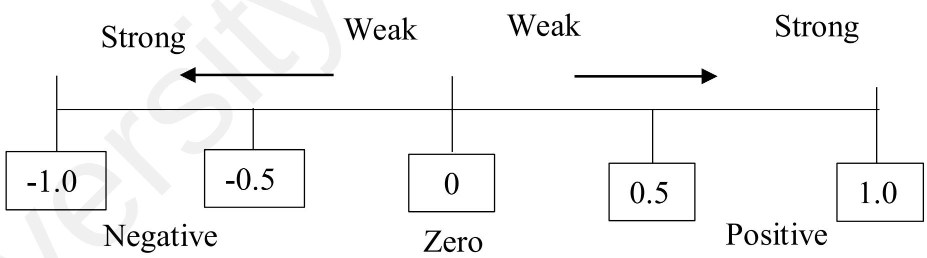


Figure 3.4: Pearson Correlation Scale

A good selection of input variables will improve the accuracy of forecasting models. In this work, the ideal input parameters of forecasting models are selected by determining the Pearson Correlation between solar power with every meteorological variable (global radiation, tilted radiation, temperature surrounding, humidity surrounding, PV module/ PV panel temperature and wind speed). The result of the correlation analysis is depicted in Table 3.1.

Table 3.1: Correlation Value between Solar Power and Meteorological Variable

Meteorological Variables	Correlation Value
Global Radiation	0.84
Tilted Radiation	0.86
Temperature Surrounding	0.48
Humidity Surrounding	-0.49
PV module/PV panel temperature	0.60
Wind Speed	0.41

According to Table 3.1, tilted radiation and global radiation depict the highest positive correlation with solar power. This represents a strong relationship between each variable with solar power. Other than that, the correlation value obtained from the PV module/ PV panel temperature is 0.60 which signifies a moderately strong relationship between PV module/ PV panel temperature with solar power. Besides, temperature surrounding and wind speed acquire the positive correlation value of 0.48 and 0.41, respectively, whilst humidity surrounding obtains the negative correlation value of -0.49. The temperature surrounding, wind speed and humidity surrounding highlight the weak relationships with solar power. Due to this matter, this work uses three highest correlated parameters, namely, PV module/ PV panel temperature, tilted radiation and global radiation as the input parameters of the forecasting models.

3.4.3 Data Averaging

This work collects the data in 1-minute resolution from 1st January 2017 until 30th November 2017 with a starting time of 8.00 A.M. until 7.00 P.M. The time horizon for this study is one hour ahead of forecasting. Thus, 1-minute of data resolution is averaged into 60-minutes of data resolution to match the forecast horizon and the equation that is used to average the data is shown in Equation 3.2. The averaging process results in 4020 data samples for every meteorological variable as well as for solar power data.

$$x_{60-\min} = \frac{\sum_{t=1}^{60} x_{t-\min}}{60} \tag{3.2}$$

$x_{60-\min}$ signifies an hourly (60-minutes) interval value and $\sum_{t=1}^{60} x_{t-\min}$ represents the summation of the 1-minute interval until hourly (60- minutes) interval.

3.5 Noisy Data Elimination using Wavelet Transform (WT)

Noise is defined as high variations data which is combined with real data (Han et al., 2007). It exists due to several reasons such as changes of instruments or the presence of dirt on the sensor of measurement (Renani et al., 2016). The noise signal needs to be isolated from the true signal to prevent the performance degradation of the forecasting model. This work utilises WT to remove the existence of noise in solar power and meteorological data. The application of noise removal technique by the WT is carried out according to three important processes which are decomposition, thresholding and reconstruction that are shown in Figure 3.5.

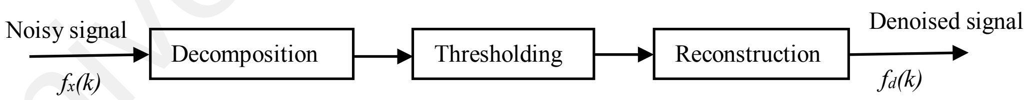


Figure 3.5: Noise Removing Technique

In the decomposition process, a signal that contains noise is decomposed repeatedly to yield high-frequency and low-frequency signals. The former is recognised as detailed signals whereas the latter is denoted as approximation signals. The level of decomposition is one of the important parameters to be determined in this process and this work chooses five levels of decomposition. This means that solar power data and three types of meteorological variables that have been choosing in Subsection 3.4.2 (tilted radiation, PV module/ PV panel temperature and global radiation) are decomposed five times to acquire the finest detailed signals and approximation signals. Figure 3.6 illustrates the five levels of the decomposition process of a noisy signal. Notation *s* signifies the original signal (global radiation, tilted radiation, PV module/ PV panel temperature or solar power) that contains noises. Notation *d* and *a* in Figure 3.6 represents the detailed signal and approximation signal, respectively.

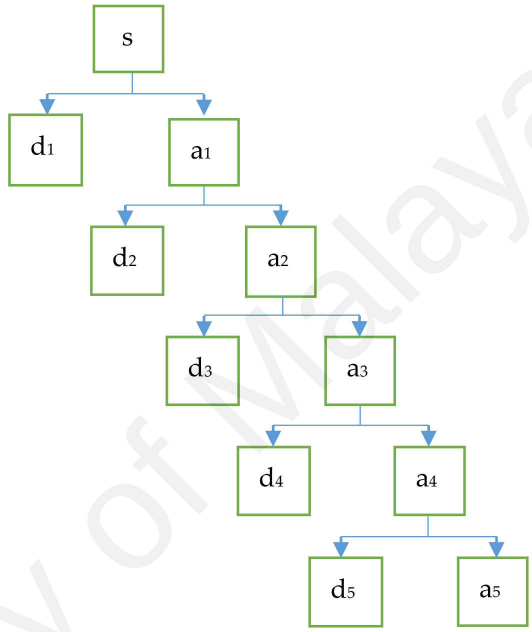
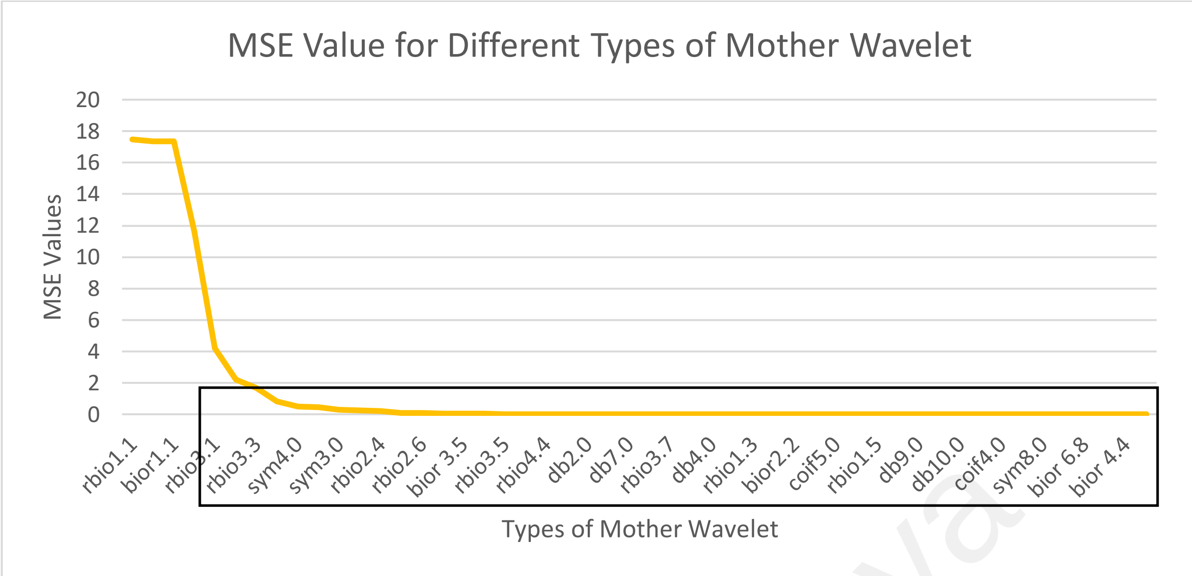
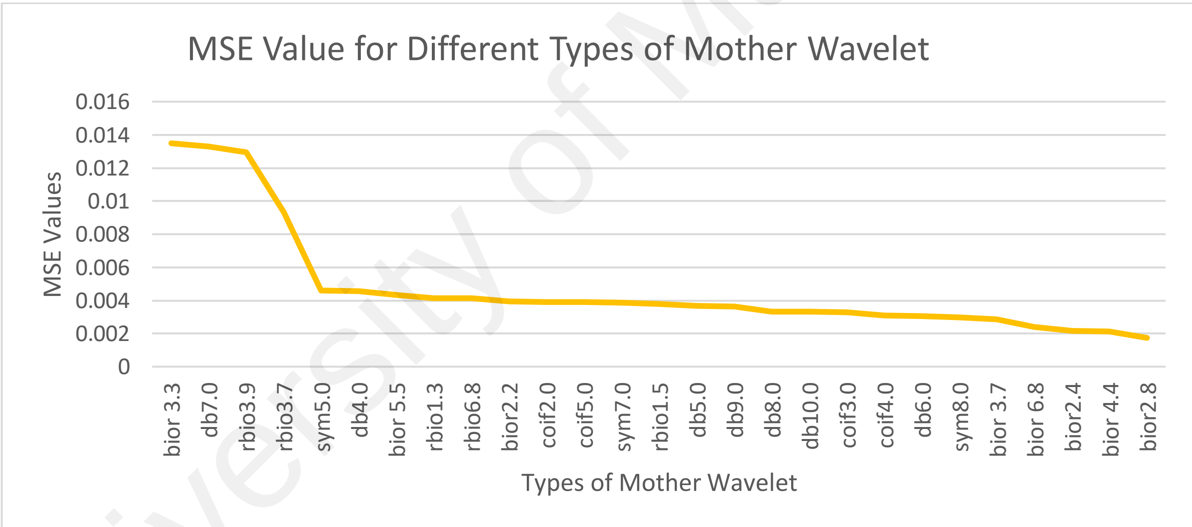


Figure 3.6: Five Levels of Wavelet Decomposition Process

The other control parameter of the decomposition process is the type of mother wavelet (ψ). An appropriate selection of the type of ψ enables a signal to be easily separated with noise components in the thresholding process (Sharie et al., 2019). Besides, a higher similarity between a mother wavelet with the output signal allows the components of the signals to be better decomposed into the wavelet coefficients (Sharie et al., 2019). Authors in the work of (Ngui et al., 2013) mentioned that there is no standard technique to be employed when selecting an ideal mother wavelet. Due to this, this work selects the ideal type of mother wavelet according to the trial and error approach. An ideal type of mother wavelet is the one that can give the lowest MSE value.



(a)



(b)

Figure 3.7: MSE Values for Different Types of Mother Wavelet: (a) All Types of Mother Wavelet; (b) Several Types of Mother Wavelet that has MSE values in range of 0.001754-0.013488

Figure 3.7 (a) illustrates the MSE values for all types of mother wavelet that has been used in this study. As can be seen in Figure 3.7 (a), there is a huge difference of MSE value between each type of mother wavelet until it can barely see which type of mother wavelet that can offer the lowest MSE value. Hence, another graph has been re-illustrated in Figure 3.7 (b) where the MSE values plotted in this figure is in the range of 0.001754 until 0.013488. The graph illustration of Figure 3.7 (b) gives a clear visual of which type of mother wavelet that can provide the lowest MSE value. As can be seen in Figure 3.7 (b), the lowest MSE value can be obtained by Biorthogonal 2.8 with the MSE value of 0.001754. Hence, Biorthogonal 2.8 has been elected as an ideal type of mother wavelet for all types of signals which are tilted radiation, global radiation, PV module/ PV panel temperature and solar power signal.

Biorthogonal type of mother wavelet consists of several members which are shown in Figure 3.8. Every member in the Biorthogonal uses two wavelets instead of single wavelet. One of the wavelets (on the left side) is utilised during the decomposition process while the other wavelet (on the right side) is employed during the reconstruction process (MathWorks, 2019).

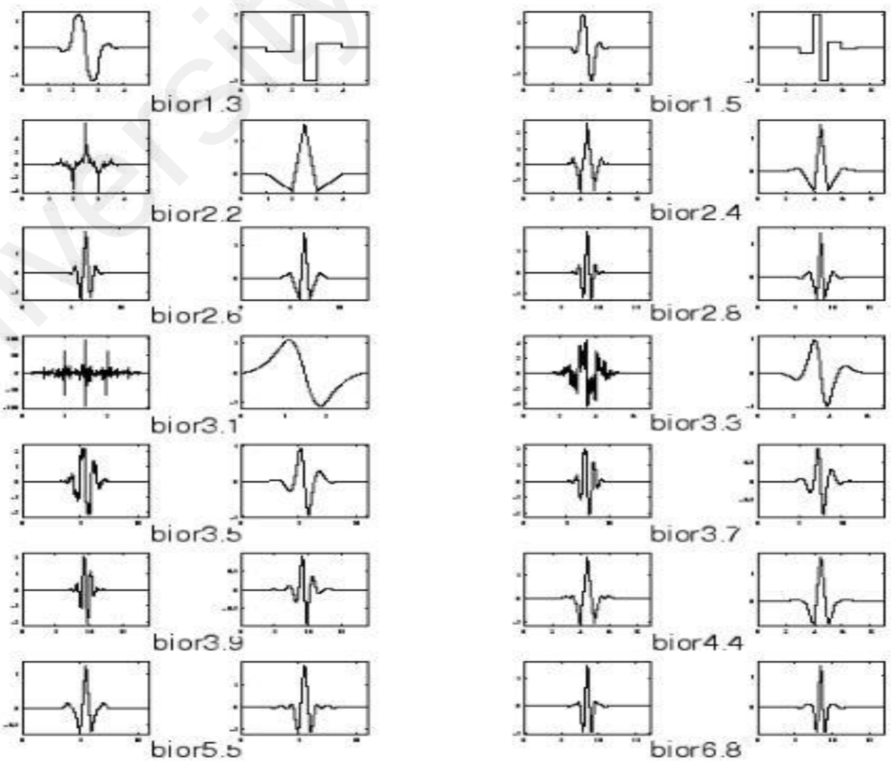


Figure 3.8: Members of Biorthogonal Wavelet (MathWorks, 2019)

The wavelets used in the decomposition process as well as in reconstruction process can have two different vanishing moments. In the wavelet system, the vanishing moments are ideally selected to maintain a smooth function of wavelet (Tian & Wells, 1998). A greater number of vanishing moments will create a sparse representation of the wavelet function. A greater number of vanishing moment is used in the decomposition process while a small number of vanishing moment is used in the reconstruction process. In this work, the 2 and 8 vanishing moments are used in the reconstruction process and decomposition process, respectively. That is why the type of mother wavelet used by this work is called Biorthogonal 2.8.

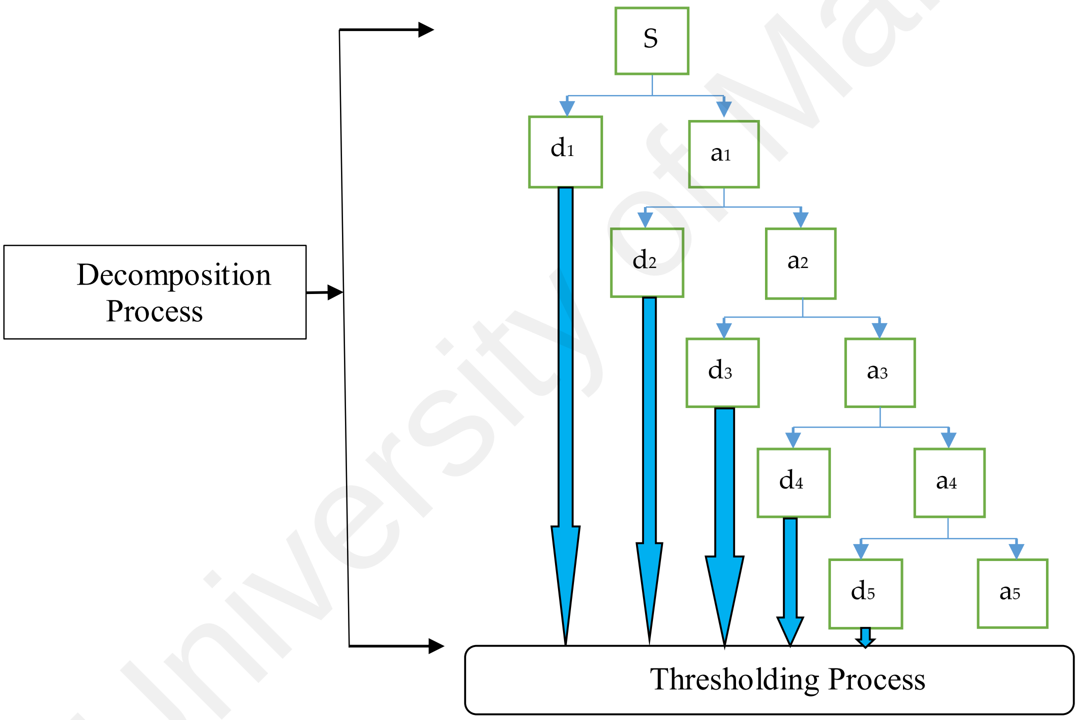


Figure 3.9: Thresholding Process of Detailed Signal

In the analysis of WT technique, important information of a signal lies in the approximation signal and it is hardly affected by any noise. On the other hand, the detailed signal signifies noisy data points of a signal (Cohen, 2012). Therefore, the thresholding process as shown in Figure 3.9, will select a threshold value for each detailed signal in every level of decomposition. After that, any coefficients of a detailed signal which are below that threshold value will be set to the zero values and the coefficients that are above the threshold value will experience the shrinking process by subtracting the threshold value from the coefficient value. There are several types of threshold estimation methods used to select the threshold value. This work selects the ideal type of threshold estimation method according to the trial and error approach and the result depicts an outstanding performance of fixed type threshold estimation method.

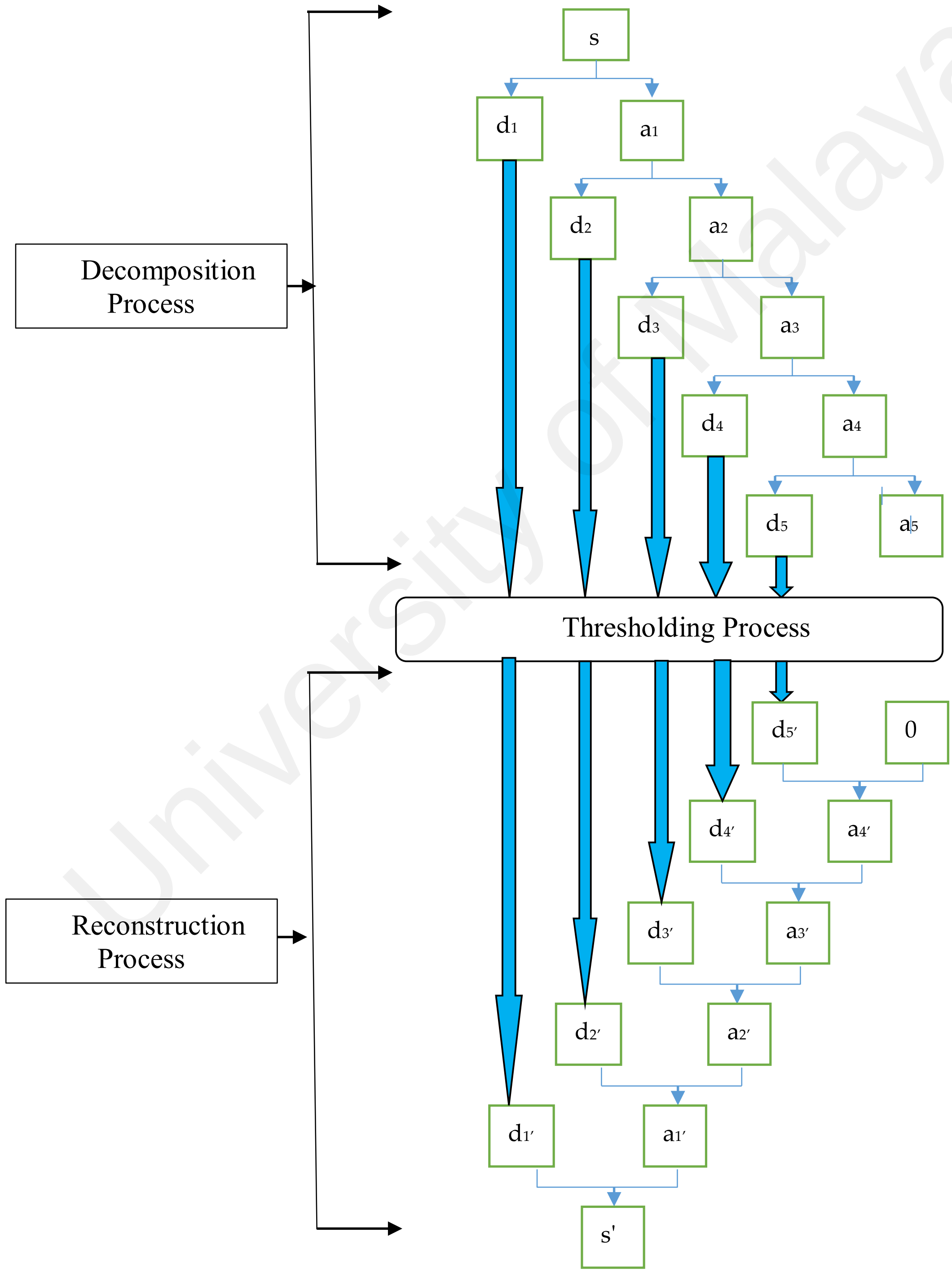


Figure 3.10: Reconstruction Process of Denoised Signal

The last process is known as the reconstruction process. At this stage, the new detailed coefficients, which are obtained from the thresholding process, are reconstructed. The structure of the reconstruction process is portrayed in Figure 3.10.

3.6 Model Development

The development of forecasting models requires two vital processes which are data normalisation and parameter specification of the forecasting model. The processes above-mentioned are explained in the following subsection.

3.6.1 Data Normalisation

Data normalisation is mostly used to scale the attributes of data in the range of [-1 1] and [0 1]. It is a practical solution to prevent the mixing of large and small input values for every forecasting model. Furthermore, the input values need to be in the same order of magnitude to prevent the bias of each input variable. This study utilises the min-max method to normalise the solar power data and meteorological data in the range of [0 1]. The min-max equation used by this work is shown in Equation 3.3

$$x_{norm} = y_{\min} + \left[(y_{\max} - y_{\min}) \left(\frac{x - x_{\min}}{x_{\max} - x_{\min}} \right) \right] \tag{3.3}$$

where x_{norm} represents the normalised data value, y_{max} denotes the maximum value of the target vector and y_{min} signifies the minimum value of the target vector. The notation x indicates the actual data value to be normalised, x_{max} characterises the maximum value of input vectors and x_{min} represents the minimum value of input vectors. For the application of the min-max method of this work, y_{max} and y_{min} are set to values of 1 and 0, respectively.

3.6.2 Parameters Specification of Forecasting Model

Three types of forecasting models developed by this work are known as MLP, RBFNN and ANFIS. Every model has its parameters that will be set to the ideal values to improve the model accuracy. There is no specific method that can be used to find the ideal parameter values of each forecasting model. Due to this matter, this work employs a trial and error method to select the parameter value for every forecasting model. Later, the performance of every forecasting model is compared and the parameters of the most accurate forecasting model will be estimated according to HFPSO approach.

3.6.2.1 Multi-Layer Perceptron (MLP)

MLP is one type of ANN and it is particularly used to learn a linear or non-linear relationship between input data and output data. The architecture of the MLP is graphically depicted in Figure 3.11. The notations of x_1: x_3 represent the input neurons of the MLP model. In this work, three types of input neurons are being used which are known as PV module/ PV panel temperature, tilted radiation and global radiation. Hence, the variable of PV module/PV panel temperature is denoted by x_1, variable of tilted radiation is signified by x_2 and variable of global radiation is represented by x_3. The input neurons of x_1: x_3 is responsible to transmit the signals to the neurons in the hidden layer.

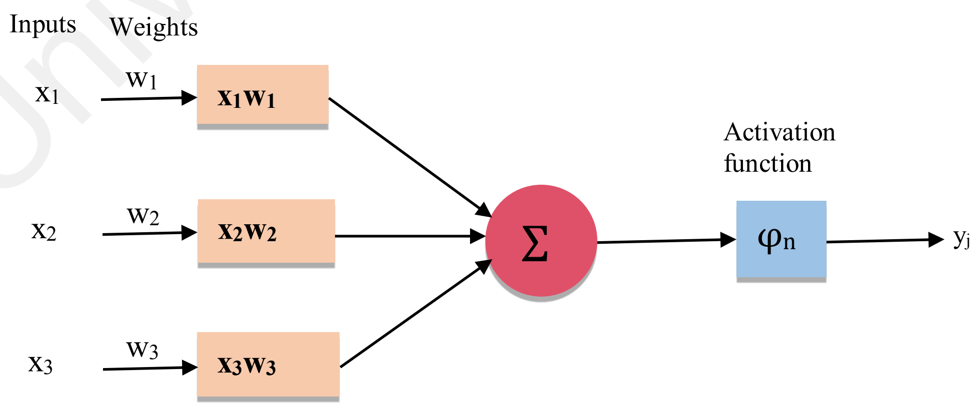


Figure 3.11: Architecture of MLP

Before the arrival of the signals at the hidden layer, the input neurons of x_1: x_3 are applied by a strong connection called as weight (w_1: w_3) that later forming the weighted signals (x_1w_1: x_3w_3). The weighted signals are added together and become the input of an activation function (φ_n) at the hidden layer where *n* represents the number of hidden layers. Note that the MLP model permits the utilisation of more than one hidden layer. In this work, one hidden layer is used because it already gives a good performance of the MLP model during the forecasting process.

The φ_n is a mathematical function which introduces linear or non-linear characteristics of ANN. In the hidden layer of MLP, the non-linear activation function is used and the most desirable type of functions are known as logistic sigmoid (*logsig*) and hyperbolic tangent sigmoid (*tansig*). This means that the summation value of x_1w_1: x_3w_3 is applied with the non-linear activation function.

After that, the output of the φ_n becomes an input to the output layer. The signals in the output layer travel and being weighted in a similar way of what input signals have experienced. However, the φ_n at the output layer can be either linear (*purelin*) or non-linear activation function (*tansig or logsig*). At the last stage, the output neurons produce the output signal known as $y_{j.}$ In this regard, the notation of y_j denotes the forecasted value of solar power.

Generally, the MLP is a supervised learning model that utilises a training algorithm to adjust the weights and biases of the MLP. The employment of an ideal training algorithm will increase the performance of the MLP due to the production of output values that have small differences with the actual values. There are many types of training algorithms that are widely used in the MLP model. Each algorithm has its function to tune the weights and biases of the MLP model. This work has utilised several types of backpropagation algorithm and the superior one is chosen according to the better performance offered during the training process of solar power forecasting.

The number of hidden layers, the number of hidden neurons, the type of activation function and the type of training algorithm are found iteratively in the trial and error process. The ideal values of those parameters are chosen according to the lowest errors existed between actual and forecasted values of solar power.

3.6.2.2 Radial Basis Function Neural Network (RBFNN)

The RBFNN is one type of ANN and it is composed of three core layers which are an input layer, a hidden layer and an output layer. The architecture of the RBFNN is represented in Figure 3.12. The neurons in each layer are responsible to receive and to transmit signals from an input layer to a hidden layer and finally to an output layer. Unlike MLP, signals assigned to input neurons x_1: x_3 in RBFNN are transmitted directly to a hidden layer without being weighted. For forecasting purpose, this work assigns the variables of PV module/PV panel temperature, tilted radiation and global radiation with the notations of x_1, x_2 and x_3, respectively.

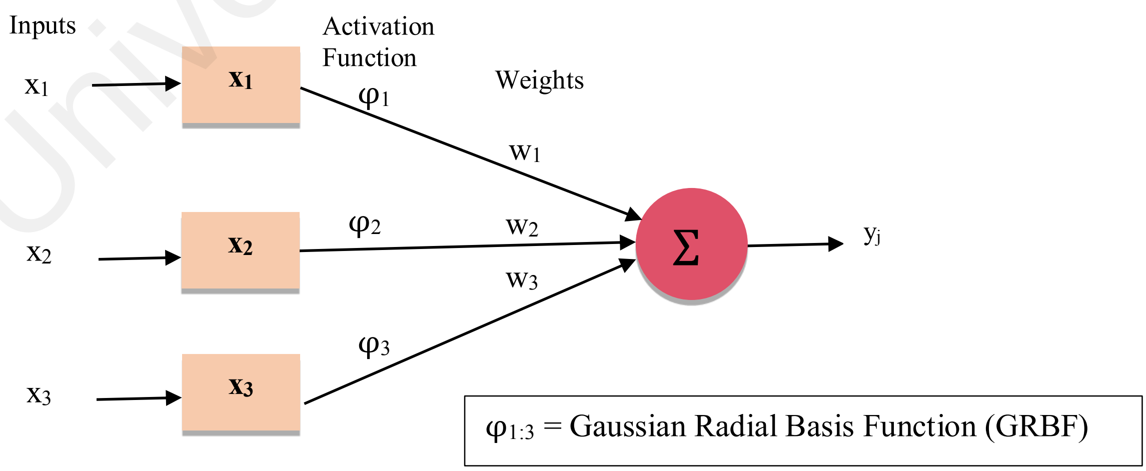


Figure 3.12: Architecture of RBFNN

Upon receiving the input signal, every hidden neuron is applied with an activation function (φ_1: φ_3). The type of φ_1: φ_3 used in the RBFNN is called a Gaussian Radial Basis Function (GRBF) where it can be shown in Equation 3.4. Later, each hidden neuron that has passed through the activation function will be weighted with weights (w_1: w_3) and added together before being sent to the output layer to produce the output signal (y_j). In this regard, the output y_j signifies the forecasted values of solar power

$$\varphi_n(x) = e^{\left[-\frac{1}{2\sigma_j^2}\left\|x-x_j\right\|^2\right]} \quad (3.4)$$

As can be seen in Figure 3.12, GRBF becomes one of the parameters that control the output production of RBFNN. By referring to the equation of GRBF in Equation 3.4, two important parameters control the GRBF value. The first parameter is known as the spread value (σ) that is functioning to control the rate of decreasing function. Large spread value produces a slow rate of decreasing function and vice versa. Besides, the distance value between the centre of the activation function (x_j) and the input neuron is one of the control parameters of GRBF. A small distance value produces the highest output value of GRBF while a large distance value provides the smallest output value of GRBF.

Other than GRBF function, the maximum number of hidden neurons is also a parameter that controls the output of RBFNN. Clearly, for the RBFNN model, the centre of GRBF, the spread of GRBF and the maximum number of hidden neurons are the parameters that need to be specified ideally to produce an accurate performance of RBFNN for forecasting purpose.

3.6.2.3 Adaptive Neuro-Fuzzy Inference System (ANFIS)

An ideal FIS model depends on the determination of a suitable membership function as well as choosing a representable rule of function that can convey the behaviour of the model. These two aspects of FIS are determined according to expert knowledge which becomes a shortcoming for the FIS model. Other than that, FIS has another shortcoming of not possessing any learning capabilities. To minimise these shortcomings, the FIS model is combined with an ANN model. This combination produces a fuzzy-neuro model that represents the knowledge in a linguistic and numerical form as well as having a training algorithm that enables the knowledge to learn and to improve from the experience.

The fuzzy-neuro model can be combined in three ways which are the cooperative neuro-fuzzy system, concurrent neuro-fuzzy system and hybrid neuro-fuzzy system (Nauck et al., 1997). For the cooperative neuro-fuzzy system, the neural network model is utilised to determine the fuzzy rule or membership of the FIS while the execution process is fully implemented by the FIS (Abraham, 2001).

On the other hand, the concurrent neuro-fuzzy system employs the neural network to provide the input data as well as to change the output of the FIS. This means that FIS and ANN are working together in the entire process. The FIS is used to pre-process the input data and the ANN is employed to post-process the output data or vice versa (Vieira et al., 2004)

Lastly, the hybrid neuro-fuzzy system uses the training algorithm of a neural network to decide the parameters of the FIS model. Later, the execution process is done by the FIS model.

In this work, a hybrid neuro-fuzzy model which is known as ANFIS model has been chosen as one of the forecasting models. It is chosen because of the learning ability of the neural network to adjust the parameters of the membership function. Apart from that, linguistic and numerical knowledge can be easily integrated (Vieira et al., 2004).

Two types of parameters which are premise parameters and consequent parameters are the control parameters of the ANFIS model. The former is updated according to gradient descent backpropagation algorithm whilst the latter is determined according to the least square estimation algorithm.

The structure of the ANFIS can be described through a model assumption that has three inputs x_1, x_2, and x_3. The IF-THEN rules of these inputs can be expressed as follows:

Rule 1: IF x_1 is A_1, x_2 is B_1 and x_3 is C_1 THEN $f_1=p_1+q_1+r_1$

Rule 2: IF x_1 is A_2, x_2 is B_2 and x_3 is C_2 THEN $f_2=p_2+q_2+r_2$

The A_i, B_i and C_i denote the fuzzy set variables of i^{th} rule, x_1, x_2 and x_3 signify the input neurons, fi represents the output of the ANFIS model with the estimation of consequent parameters known as p_i, q_i and r_i. Note that the division of x_1: x_3 into the i^{th} rule of A_i, B_i and C_i is achieved according to three approaches.

The first approach is known as grid partitioning where the number of membership function of the FIS model, the type of membership function for input data and the type of membership function for output data of the FIS model is determined optimally by the user. This approach is suitable to be used on a small number of input data which makes it less preferable in the forecasting area (Vlădăreanu et al., 2018). This is due to a reason that large input numbers increase the number of rules which leads to a complex computation.

The second approach is the subtractive clustering approach that produces the scattering partition on the input data of x_1: x_3. Its user-determined parameter is known as the influence radius which specifies the range of influence for the cluster centre in the data dimension.

The third approach uses the Fuzzy C-Means (FCM) clustering to extract a suitable set of rules to model the behaviour of the data. For the FCM clustering approach, the Gaussian curve membership function (*gaussmf*) is utilised as membership function of the x_1: x_3 while linear membership function (*linear*) is employed as the membership function for the output data. The important parameter that needs to be specified by this approach is the number of clusters.

In this work, the FCM clustering approach has been chosen to divide the input data of x_1: x_3 into a certain number of cluster. The notations of x_1, x_2 and x_3 represent the input variables of PV module/PV panel temperature, tilted radiation and global radiation, respectively. After the division of data, each input data in the x_1: x_3 will pass through several processes as shown in Figure 3.13. Every process involved is explained as follows:

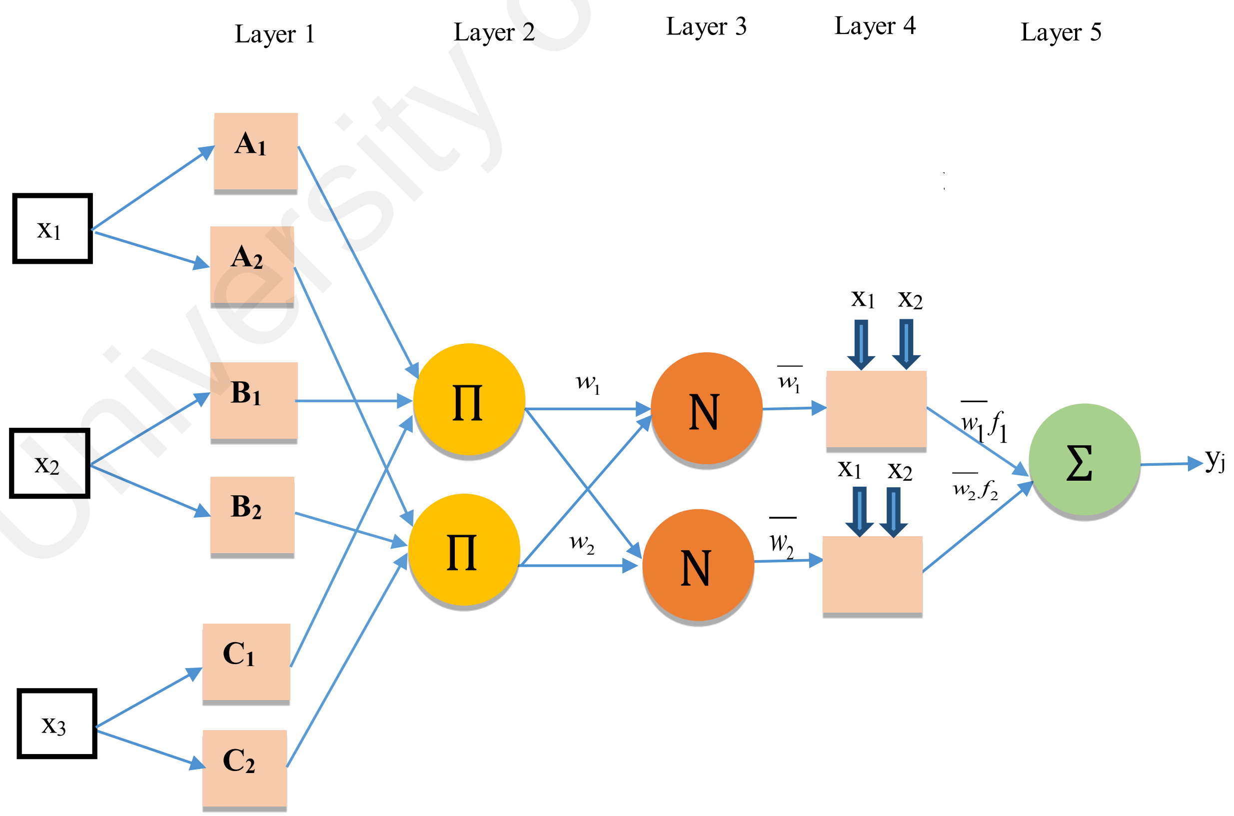


Figure 3.13: Architecture of ANFIS

Layer 1: The first layer of the ANFIS model conducts the fuzzification process. According to Figure 3.13, the input data of x_1: x_3 has been divided into two clusters of A_i, B_i and C_i where *i* signifies the number of clusters. Abovementioned, the FCM clustering approach utilises *gaussmf* as the membership function of input data x_1: x_3. Hence, the fuzzification process will assign the function of *gaussmf* to each data in the x_1: x_3. The expression of *gaussmf* is shown in Equation 3.5

$$O_{1,j} = \mu_{A_i}(x) = \exp(-\frac{1}{2}(\frac{x_1 - c_i}{\sigma_i})^2) \tag{3.5}$$

where $O_{1,j}$ represents the output of the j^{th} node at the first layer while μ_{A_i} denotes the membership function of i^{th} rule A_i. Meanwhile, the notations of c_i and σ_i signifies centre and width of the *gaussmf*, respectively, which is identified as premise parameters.

Layer 2: The second layer in the ANFIS model multiplies the membership grades of i^{th} rules A_i, B_i and C_i. The multiplication of membership grades produces a j^{th} node output that is known as firing strength (w_j) and it can be expressed as in Equation 3.6.

$$O_{2,j} = w_j = \mu_{A_i}(x_1)\mu_{B_i}(x_2)\mu_{C_i}(x_3) \tag{3.6}$$

Layer 3: The output of j^{th} node at Layer 3 is recognised as the normalisation of the firing strength ($\overline{w_j}$) where the w_j of each rule is divided with the total w_j of all rules as shown in Equation 3.7.

$$O_{3,j} = \overline{w_j} = \frac{w_1}{w_1 + w_2} \tag{3.7}$$

Layer 4: In this Layer 4, the j^th node calculates the contribution of the i^th rule to the overall output response as shown in Equation 3.8. The notations of p_i, q_i and r_i are denoted as the consequent parameters of the ANFIS model.

$$O_{4,j} = \overline{w_j} f_i = \frac{w_1}{w_1 + w_2}(p_i x + q_i x + r_i x) \tag{3.8}$$

Layer 5: The process in the fifth layer is recognised as defuzzification process where each node produces an output according to the contribution of every rule as shown in Equation 3.9. The output is identified as forecasted values of solar power and it is notated by letter y_j in Figure 3.13.

$$O_{5,j} = \sum \overline{w_j} f_j \tag{3.9}$$

For the training process of the ANFIS model, two types of training algorithms which are gradient descent BP algorithm and hybrid algorithm are employed. Note that the hybrid algorithm is a combination of gradient descent BP with the least square algorithm where the former is used to determine the values of premise parameters and the latter is utilised to regulate the values of consequent parameters. In this work, the hybrid algorithm has been used instead of gradient descent BP algorithm because it gives a better result during the forecasting process of solar power.

3.6.3 Training and Testing Processes of Forecasting Model

After the parameter selection for forecasting model (MLP, RBFNN and ANFIS), the historical values of normalised solar power, normalised global radiation, normalised PV module/ PV panel temperature, and normalised tilted radiation are divided into training data and testing data to perform the training and testing processes. Aforementioned in Subsection 3.4.3, the averaging process results in 4020 samples of data for every meteorological data and solar power data. To the best of author's knowledge, there is no specific method that estimates an ideal size of the training and testing dataset. Due to this matter, this work divides the samples into 80% of training data and 20% of testing data, which was according to the previous works of (Adeoti & Osanaiye, 2012; Benali et al., 2019; Halabi et al., 2018; Premalatha, N. & Valan Arasu, 2016). Hence, the data division in this work is equivalent to 3216 samples for the training process and 804 samples for the testing process.

One important challenge arises whether the usage of 3216 samples of data is sufficient to train the forecasting models. To address this challenge, this study uses the rule of thumb, namely, the rule of 10 in finding the right amount of data for forecasting models (Haldan, 2015). According to this rule, the lowest amount of training data needed to achieve a good performance of machine learning methods is equivalent to 10 times the parameters in the investigated model. In this study, four parameters are used which are tilted radiation, global radiation, PV panel/PV module temperature and solar power and the minimum number of training data needed in this study is 40 samples of data only. Hence, the usage of 3216 samples is merely sufficient to train the forecasting models of MLP, RBFNN and ANFIS.

Later, the training data is presented to every forecasting model in the training process to familiarise each model with the pattern of data. The training process of MLP involves the input parameters and output parameters. The training process of MLP is initiated when each input signal, namely, tilted radiation, global radiation and PV panel /PV module temperature is being multiplied with a strong connection called as weight at the input layer. Later, the weighted signal from each input signal is being added together and becoming the input to the non-linear activation function at the hidden layer. Then, the outcome of the activation function is sent to the output layer. It will be multiplied with the weight and is applied with linear activation function at the output layer before producing the final result which is known as the forecasted value of solar power. The forecasted value of solar power is compared with the actual value of solar power obtained from on-site raw data collection. If there is a huge difference between both values, the forecasted values of solar power will be backpropagated from the output layer to the hidden layer and finally at the input layer. In this backpropagation stage, the weight at the input and output layer will be tuned and adjusted by using training algorithm to produce the forecasted value of solar power that has very small difference with the actual value of solar power.

The training process of RBFNN involves the transmission of the input signals in the input layer directly to the hidden layer. In the hidden layer, each signal is applied with an activation function which is called as GRBF. Later, the outcome of the activation function is multiplied with the weight and they are added together before being sent to the output layer. The signal in the output layer signifies the forecasted values of solar power.

For the ANFIS model, the training process is initiated by dividing each input signal into two clusters. After that, every cluster in an input signal is being assigned with the function of *gaussmf* which later forming the membership grade. The membership grade for each input signal is then being multiplied according to the specified cluster and it is known as firing strength. Later, the firing strength of each cluster is being divided with the total firing strengths of all clusters. At this stage, the firing strength is known as normalisation of firing strength. Lastly, the normalisation of firing strength for each cluster is multiplied with the premise parameters before forming the output signal which is known as forecasted values of solar power.

The training process is executed repeatedly until the convergence of the forecasting model is met. In this study, the training data in all forecasting models have been trained for 15 times. This is due to the reason that 15 times for training the data is sufficient enough for all forecasting models to achieve the lowest generalisation error.

After that, the testing process introduces the input testing data (PV module/ PV panel temperature, global radiation, and tilted radiation) to the trained model to generate solar power values. The forecasted solar power data is then compared with the actual solar power data to estimate the performance of the forecasting model.

3.6.4 Performance Metrics Evaluation of Forecasting Model

The accuracy of every forecasting model is evaluated according to performance metrics analyses. This work employs four types of performance metrics, namely, R, RMSE, Mean Square Error (MSE) and MAE where the equations are shown in Equation 3.10 until Equation 3.13

$$R = \frac{\sum(I_m - \overline{I_m})(I_f - \overline{I_f})}{\sqrt{\sum(I_m - \overline{I_m})^2 \sum(I_f - \overline{I_f})^2}} \tag{3.10}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(I_f - I_m\right)^2} \tag{3.11}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(I_f - I_m\right)^2 \tag{3.12}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|I_f - I_m\right| \tag{3.13}$$

where I_f represents the forecasted value and I_m signifies the actual value. Furthermore, $\overline{I_m}$ indicates the mean of actual values, $\overline{I_f}$ designates the mean of forecasted values and n is for the number of observations. The estimation of RMSE and MSE gives the measurement variability for the forecasted data and actual data (Yona et al., 2007). On the other hand, the average distance between the actual and forecasted can be determined according to the parameter of MAE (Rana et al., 2016).

Small values of RMSE, MSE and MAE indicate an outstanding performance of the forecasting model when performing the forecasting process of solar power. On the other hand, R-value specifies a degree of correlation between actual values and forecasted values of solar power. Any forecasting model that can depict an R-value that is close to 1 is considered an excellent model for forecasting.

After the performance metrics estimation for each forecasting model, their performance metrics are then compared to select a model that can offer the highest accuracy for forecasting. In this case, the most accurate forecasting model is chosen according to the production of the smallest values of RMSE, MAE, as well as the highest R-value during solar power forecasting.

3.7 Optimisation of the Most Accurate Forecasting Model by using Hybrid Firefly and Particle Swarm Optimisation (HFPSO)

The hybrid optimisation technique is a combination of two optimisation algorithms with the aims to improve the quality of the result, to decrease the running time during the optimisation process and to show greater flexibility during a complex problem (Aydilek, 2018; Raidl & Puchinger, 2008). Apart from that, the hybridisation of two algorithms will extract the powerful benefits of each algorithm while eliminating the weaknesses (Thangaraj et al., 2011). In this work, two optimisation algorithms, namely, PSO and FF are hybridised to become HFPSO. The PSO is used in the global search process because it manages to converge rapidly during exploration while FF is generally employed in the local search process to provide better exploitation (Aydilek, 2018).

Generally, the HFPSO algorithm employs five important steps to perform the optimisation process. The steps above-mentioned are identified as initialisation stage, distance updating stage, position updating stage, memory updating stage and termination stage.

Step 1 Initialisation: Initialisation of HFPSO is a process to initialise all of the input parameters in the algorithms of PSO and FF. Table 3.2 depicts the type of input parameters for FF and PSO algorithms as well as the values that have been specified to each of them. All of the values are obtained from the previous works of (Bahrami et al., 2014; Bashir & El-Hawary, 2009; Ren et al., 2014; Thirupathaiah, 2018).

Table 3.2: Input Parameters for HFPSO Algorithm

Type of Input Parameters	**Symbol**	**Value**
Maximum Search Range Limit	X_{max}	1
Minimum Search Range Limit	X_{min}	0
Number of Particles	nPop	25
Number of Maximum Iteration	Itr_{max}	1000
Inertia Weight	ω	1
Cognitive Parameter	C_1	1
Social Parameter	C_2	2
Attractiveness of Fireflies at Zero Distance	β_0	2
Randomisation Parameter	α	0.2
Coefficient of the Light Absorption	γ	1

Apart from that, the initialisation step initialises the position of the particle to be at random. This step also includes the fitness evaluation of a particle by referring to the objective function allocated to every particle. In this work, the objective function assigned to the HFPSO is in term of RMSE. Therefore, a particle can estimate its initial fitness according to RMSE value that has been assigned to them. Lastly, the particle stores the initial position as the P_{best} and the best initial position achieved by all particles is kept as G_{best}.

Step 2 Distance Updating: In every iteration, the distance between particle x_j with the P_{best} and G_{best} is updated. The distance between P_{best} with the particle x_j is denoted as r_{px}. Meanwhile, r_{gx} signifies the distance between the G_{best} with particle x_j. The expressions for r_{px} and r_{gx} can be shown in Equation 3.14 and Equation 3.15, respectively, where notation *d* in both expressions had signified the dimensionality parameter of the search problem whilst component of spatial coordinate is denoted by notation *k* (Umbarkar et al., 2017).

$$r_{px} = \sqrt{\sum_{k=1}^{d} (P_{best\,i,k} - x_{j,k})^2} \tag{3.14}$$

$$r_{gx} = \sqrt{\sum_{k=1}^{d} (G_{best\,i,k} - x_{j,k})^2} \tag{3.15}$$

Step 3 Position Updating: After the distance between particle x_j with the P_{best} and G_{best} is updated, the fitness values of particle x_j and particle x_i are compared. If the fitness value of particle x_j is bigger than the fitness value of particle x_i, the position updating process is handled by the FF algorithm by updating the new position based on Equation 3.16. At this stage, the local search process is started. If the fitness value of particle x_j is smaller than the fitness value of particle x_i, the global search process will occur. At this stage, the position updating process is handled by the PSO algorithm and the new position is updated according to Equation 3.17.

$$x_{i,k}(t+1) = x_{i,k}(t) + \beta_0 e^{-\gamma r_{ij}^2}(x_{j,k}(t) - x_{i,k}(t)) + \alpha(rand - 0.5) \tag{3.16}$$

$$x_{i,k}(t+1) = \omega x_{i,k}(t) + C_1 e^{-r_{px}^2}(P_{best\ i,k}(t) - x_{i,k}(t)) + C_2 e^{-r_{gx}^2}(G_{best\ i,k}(t) - x_{i,k}(t)) + \alpha(\gamma - 0.5) \tag{3.17}$$

$x_{i,k}(t+1)$ is the new position of particle x_i, $x_{i,k}(t)$ denotes the previous position of particle x_i and $x_{j,k}(t)$ denotes the previous position of particle x_j.

Step 4 Memory Updating: The fitness of a particle is evaluated according to the current objective function attained. For the case of P_{best}, if the current position has a smaller value than the previous position, the P_{best} is updated according to the current position. Otherwise, the P_{best} value retains the previous position value. The process of updating the P_{best} value is portrayed in Equation 3.18 where *f(x)* represents the objective function attained by the particles.

$$P_{best_{i,k}}^{(t+1)}) \begin{cases} x_{i,k}^{(t)} & \text{if} \quad f(x_{i,k}^{(t+1)}) \geq f(x_{i,k}^{(t)}) \\ x_{i,k}^{(t+1)} & \text{if} \quad f(x_{i,k}^{(t+1)}) < f(x_{i,k}^{(t)}) \end{cases} \tag{3.18}$$

After that, the fitness of all particles in the swarm is evaluated to obtain the ideal value of G_{best}. Similar to the case of P_{best}, the new value of G_{best} is updated to the current value if it is smaller than the previous G_{best}.

Step 5 Termination Stage: The processes in Step 2 until Step 4 repeats until the maximum iteration is achieved or the convergence is reached. Later, the sorting process of the particle according to individual fitness value is implemented. The above-mentioned steps when utilising HFPSO is vividly described in the flowchart as shown in Figure 3.14.

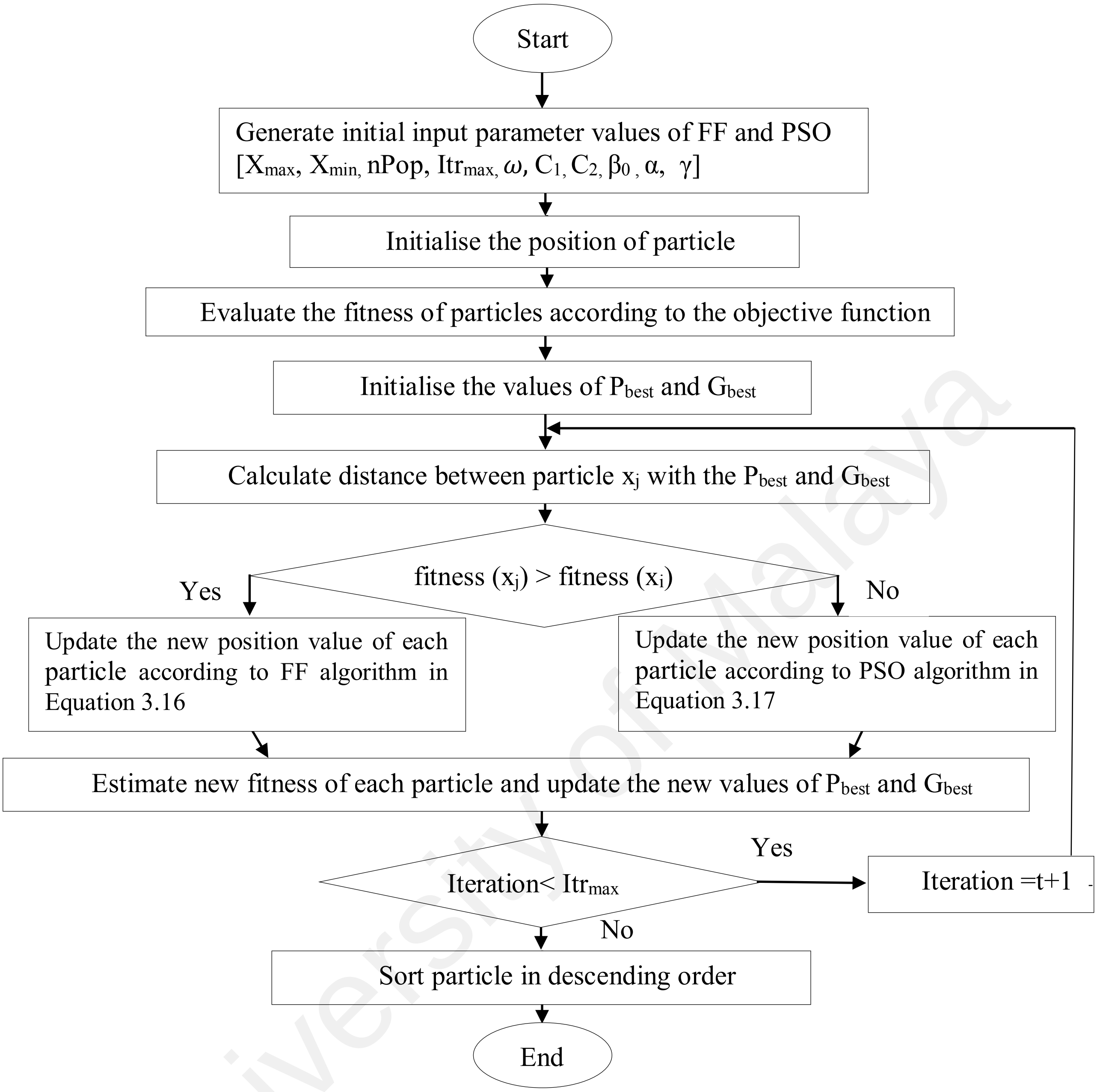


Figure 3.14: Flowchart of HFPSO Algorithm

CHAPTER 4: SIMULATION RESULTS

4.1 Introduction

This chapter discusses the simulation results of this study during solar power forecasting and it is further divided into four phases of works. For the Phase 1 and Phase 2, the result of MLP is assessed according to a different type of training algorithm. Furthermore, the result of RBFNN is measured based on the variation number of spread while the result of ANFIS is evaluated according to the variation of cluster numbers.

The third phase (Phase 3) compares the performance of MLP, RBFNN and ANFIS and selects the most accurate forecasting model according to the values of MAE, RMSE and R depicted by each forecasting model in Phase 1 and Phase 2. A forecasting model that gives the lowest RMSE and MAE values as well as provides the highest R-value is chosen as the most accurate forecasting model. The final phase compares the result of HFPSO with single PSO and single FF according to the values of RMSE, MSE, MAE and R.

4.2 Phase 1: Performance of Forecasting Models without Utilisation of WT

This section discusses the results that are obtained from MLP, RBFNN and ANFIS when utilising the noisy solar power and noisy meteorological data.

4.2.1 MLP Model

Three layers of MLP which comprise of one input layer, one hidden layer and one output layer are used. The hidden layer utilises *logsig* activation function while the output layer employs *purelin* activation function because both activation functions enhance the accuracy of the MLP model. Apart from that, the best performance of the MLP model is obtained with the usage of 12 hidden neurons. This subsection discusses the result of MLP when uses nine training algorithms, namely, Lavenberg Marquardt (LM) (*trainlm*), Broyden-Fletcher-Goldfarb-Shanno Quasi-Newton (BFG) (*trainbfg*), Resilient Backpropagation (RP) (*trainrp),* Scaled Conjugate Gradient (SCG) (*trainscg*), Conjugate Gradient Backpropagation with Powell-Beale (CGB) (*traincgb*), Conjugate Gradient Backpropagation with Fletcher-Reeves (CGF) (*traincgf*), Conjugate Gradient Backpropagation with Polak-Ribiere (CGP) (*traincgp*), One Step Secant Backpropagation (*trainoss*) and Gradient Descent with Momentum and Adaptive Learning Rate Backpropagation (*traingdx*). The result of every training algorithm is displayed in Table 4.1 and the regression plots for every training algorithm are illustrated in Figure 4.1.

Table 4.1: Performance Comparison of Training Algorithm for MLP Model

Algorithm	**MAE**	**RMSE**	**R**
trainlm	0.0471	0.0619	0.9700
trainrp	***0.0471***	***0.0617***	***0.9709***
trainbfg	0.0497	0.0636	0.9693
trainscg	0.0499	0.0637	0.9700
traincgb	0.0497	0.0637	0.9684
traincgf	0.0478	0.0621	0.9702
traincgp	0.0488	0.0619	0.9705
traingdx	0.0499	0.0663	0.9681
trainoss	0.0606	0.0745	0.9583

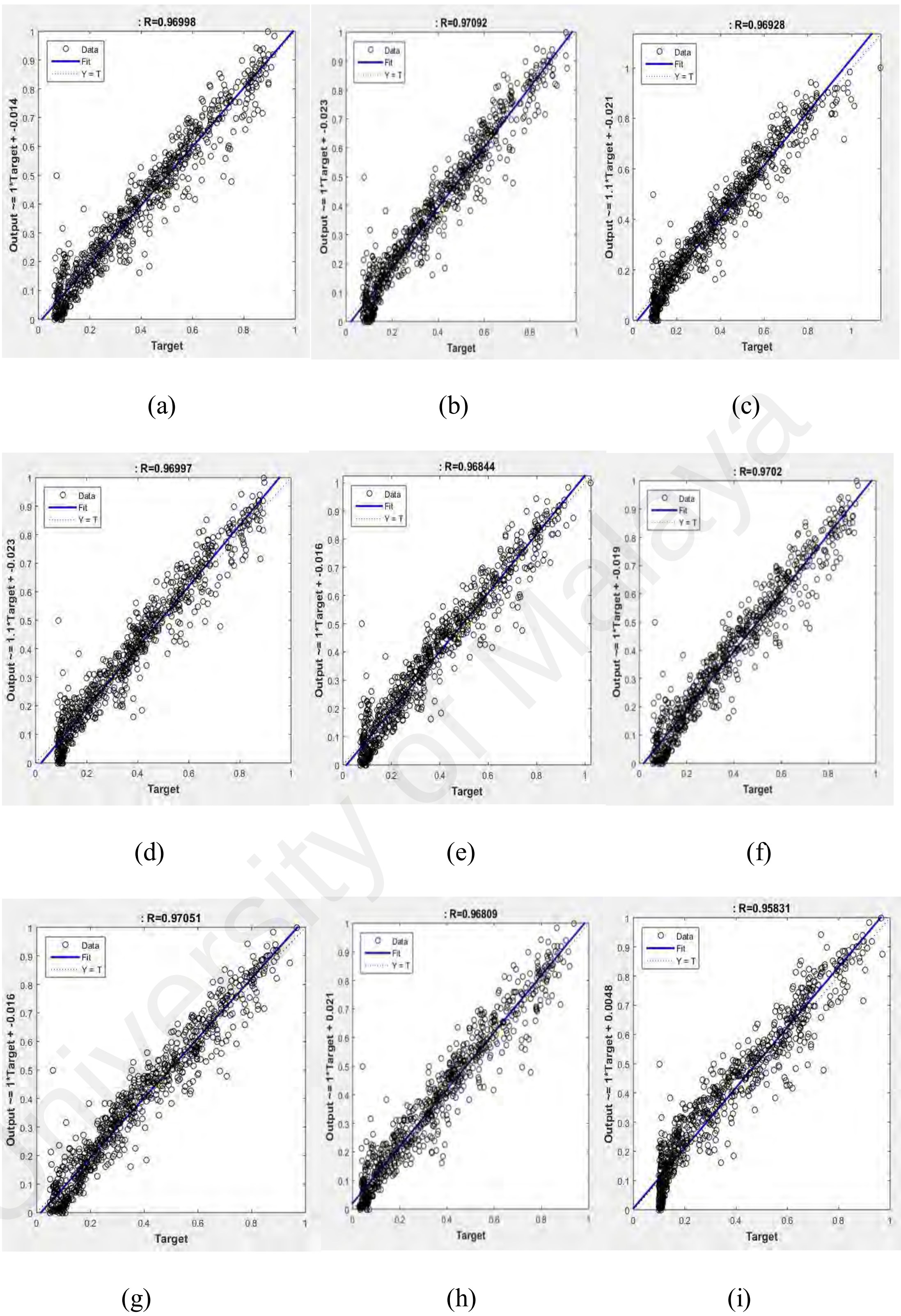


Figure 4.1: Regression Plots of Training Algorithm for MLP model: (a)trainlm; (b)trainrp; (c)trainbfg; (d)trainscg; (e)traincgb; (f)traincgf; (g)traincgp; (h)traingdx; (i)trainoss

As can be seen in Table 4.1, the lowest MAE is depicted by *trainrp* algorithm with a value of 0.0471 while the lowest RMSE value of 0.0617 is also obtained from the *trainrp* algorithm. In the meantime, the highest R-value of 0.9709 in Table 4.1 and Figure 4.1 is attained from the t*rainrp* algorithm. These results show that the majority of forecast data that is tested with the *trainrp* algorithm is closer to the actual data. Due to this matter, the *trainrp* algorithm is chosen as an ideal training algorithm for the MLP model.

4.2.2 RBFNN Model

The main parameters to be determined in RBFNN are the maximum number of hidden neurons, the centre of activation function and the spread value of activation function. Generally, the RBFNN adds the hidden neuron one at a time until the lowest MSE is achieved. In this work, 40 number of hidden neurons are chosen because they give the lowest MSE value. Another parameter which is the centre of the activation function is selected according to the forward search strategy that chooses the value based on the input that has the highest error. The spread value of activation function is varied to estimate the performance of the RBFNN in solar power forecasting and the result is tabulated in Table 4.2.

Table 4.2: Performance Comparison of Spread Value for RBFNN Model

Spread	MAE	RMSE	R
1	0.0480	0.0635	0.9684
2	0.0487	0.0629	0.9698
3	0.0469	0.0606	0.9721
4	0.0469	0.0610	0.9717
5	***0.0467***	***0.0604***	***0.9722***
6	0.0469	0.0605	0.9721
7	0.0469	0.0606	0.9721
8	0.0479	0.0616	0.9710
9	0.0497	0.0638	0.9683
10	0.0500	0.0640	0.9681

In this investigation, the utilisation of five spread values has increased the accuracy of RBFNN. This is proven when the values of MAE=0.0467, RMSE=0.0604 and R=0.9722 are outperforming other numbers of spreads.

4.2.3 ANFIS Model

Aforementioned, the division of input and output data can be achieved by using grid partitioning, subtractive clustering and FCM clustering approaches. This study neglects the usage of grid partitioning approach because the greater number of input leads to an increasing number of rules. As a result, the mapping between the input data and output data cannot be performed effectively. Previous works such as (Benmouiza & Cheknane, 2019; Mirrashid, 2014; Mollaiy-Berneti, 2016) compared the performance of FCM clustering and subtractive clustering and the outcomes had proved the supremacy of the FCM clustering approach in giving an accurate result. Thus, this study employs the FCM clustering approach to estimate the number and the type of membership functions for input data and output data.

In the FCM clustering approach, the number of clusters is an important parameter to be set up. The optimum number of clusters is found by testing the various number of clusters and the result is shown in Table 4.3. This work found that six number of clusters are outperforming the results of other numbers of clusters. This is proven due to the lowest MAE value of 0.0504 and the lowest RMSE value of 0.0647 obtained during the forecasting process.

The outstanding performance of six number of clusters is further proven according to the R-value of 0.9674 which is shown in Table 4.3 and Figure 4.2. The value of 0.9674 indicates that forecasted solar power data is perfectly correlated with actual solar power data. These results specify that six number of clusters in the ANFIS model has depicted a reliable result for solar power forecasting.

Table 4.3: Performance Comparison of Number of Clusters for ANFIS Model

Number of Clusters	MAE	RMSE	R
2	0.0504	0.0653	0.9663
3	0.0509	0.0658	0.9658
4	0.0513	0.0654	0.9663
5	0.0512	0.0653	0.9665
6	***0.0504***	***0.0647***	***0.9674***
7	0.0510	0.0653	0.9665
8	0.0504	0.0660	0.9661
9	0.0524	0.0685	0.9630
10	0.0536	0.0688	0.9631
11	0.0525	0.0669	0.9650
12	0.0537	0.0695	0.9618
13	0.0544	0.0704	0.9609
14	0.0538	0.0721	0.9587
15	0.0546	0.0710	0.9600

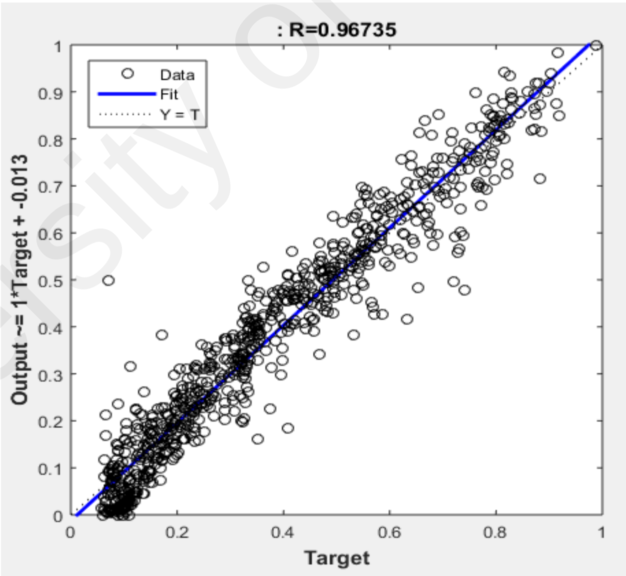
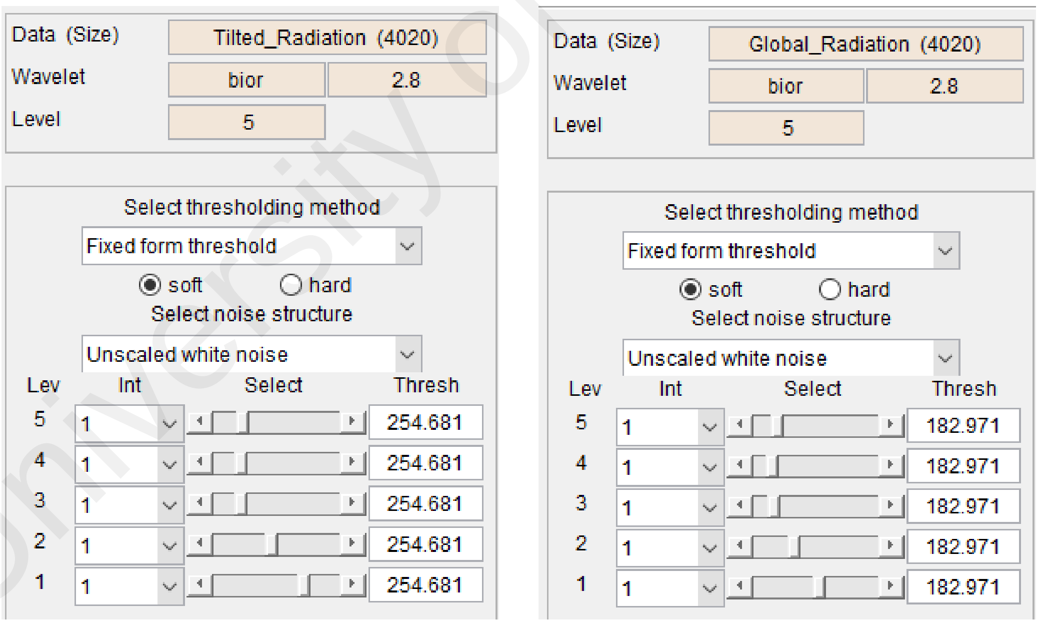


Figure 4.2: Regression Plot for ANFIS Model

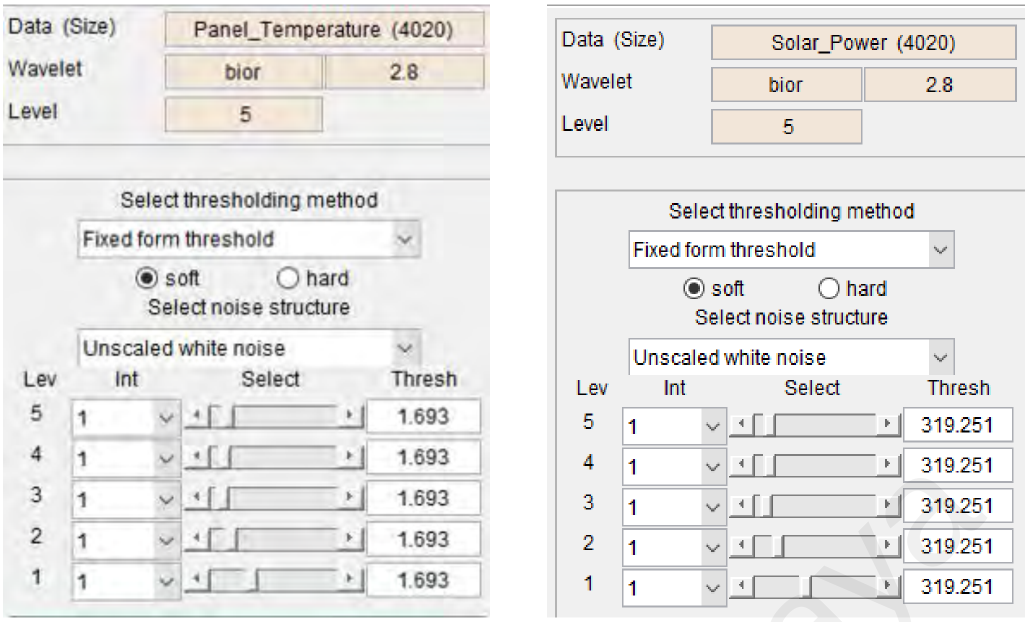
4.3 Phase 2: Performance of Forecasting Model with Utilisation of WT

This section investigates the performance of MLP, RBFNN and ANFIS when using the WT to remove the noise in the tilted radiation, global radiation, PV module/ PV panel temperature and solar power data. The WT removes the noise by performing the decomposition process, the thresholding process and the reconstruction process. In the decomposition process, the level of decomposition and the type of mother wavelet need to be determined by the user. In Phase 2 of work, five levels of decomposition and mother wavelet of Biorthogonal 2.8 are selected because they depict the lowest MSE values. Apart from that, the threshold coefficient in the thresholding process is chosen according to the fixed thresholding technique. The threshold values for tilted radiation, global radiation, PV panel/ PV module temperature and solar panel can be acquired in Figure 4.3.



(a) (b)

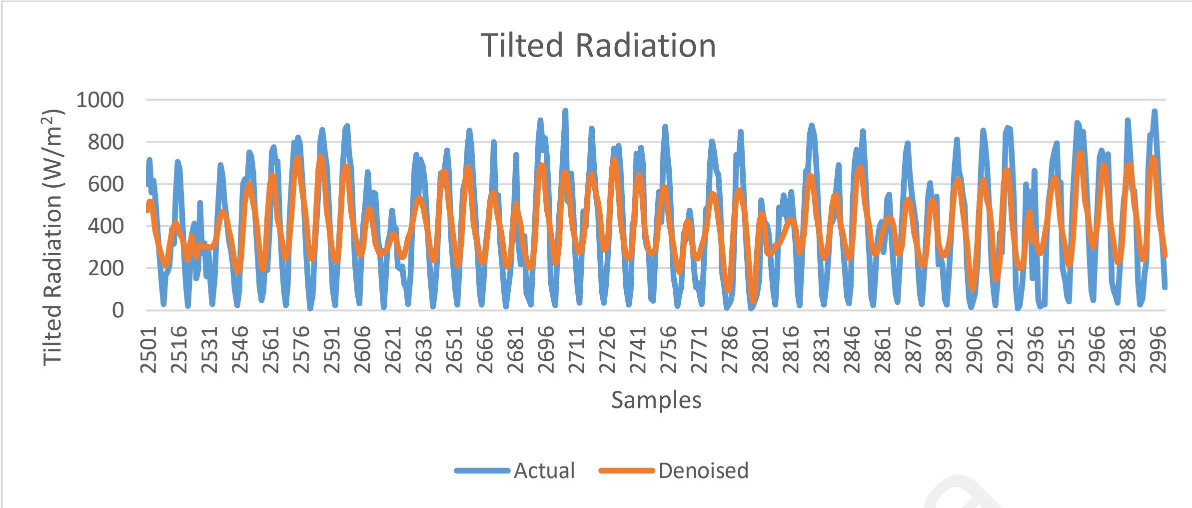
Figure 4.3: Threshold Value for Every Level of Decomposition (a) Tilted Radiation; (b) Global Radiation; (c) PV Panel/PV Module Temperature; (d) Solar Power



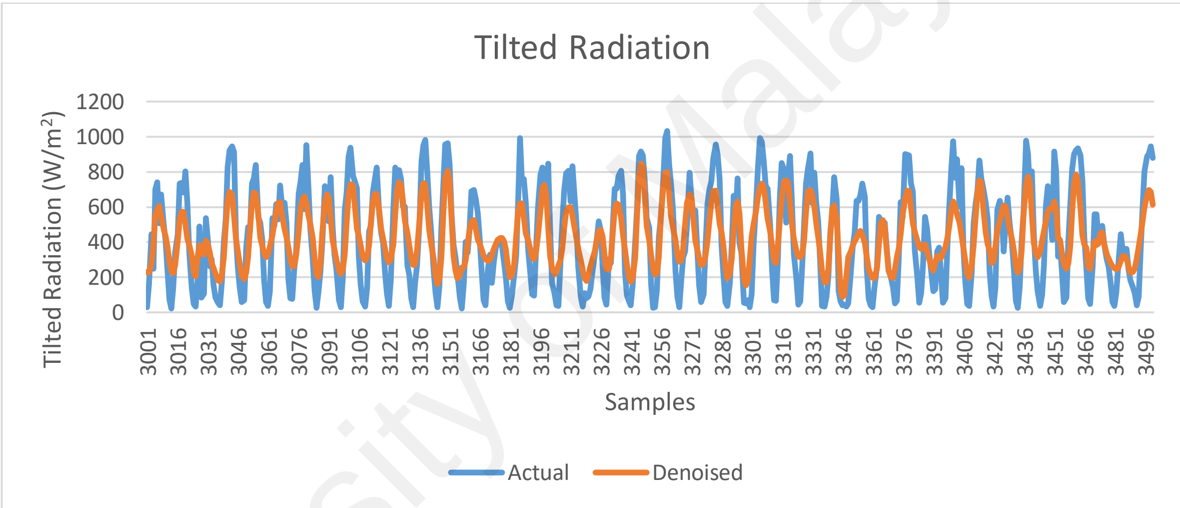
(c) (d)

Figure 4.3, continued

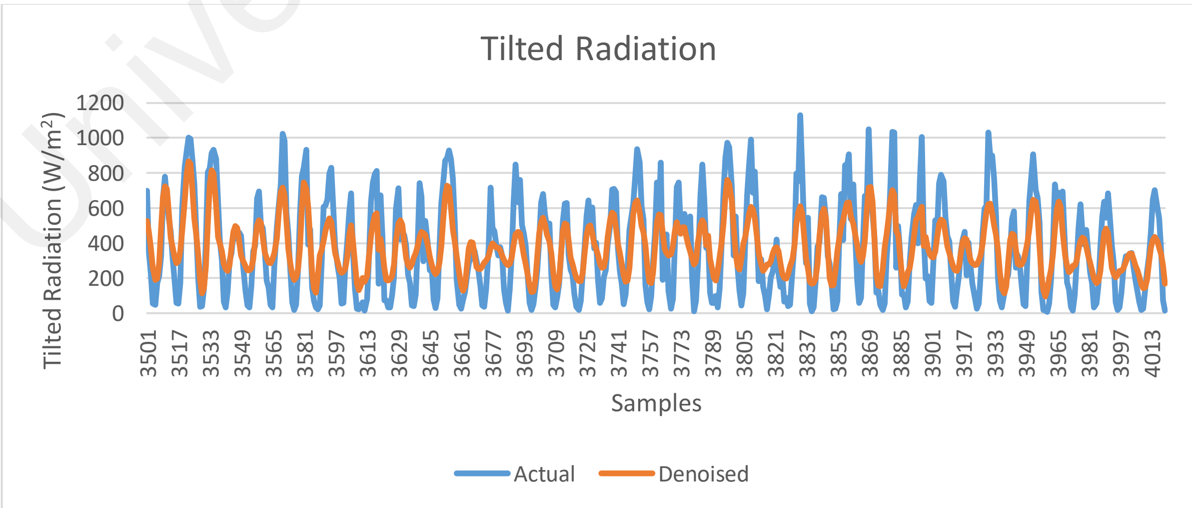
The behaviour of the tilted radiation, global radiation, PV module/ PV panel temperature and solar power signals when utilising the WT approach can be portrayed in Figure 4.4 until Figure 4.7. Aforementioned, there are 4020 samples of tested data used in this study. However, this study manages to illustrate the denoised data of WT for the 2501th until 4020th samples only. It is because it will be difficult to see the difference between the actual signal and denoised WT signal if all 4020 samples are plotted.



(a)

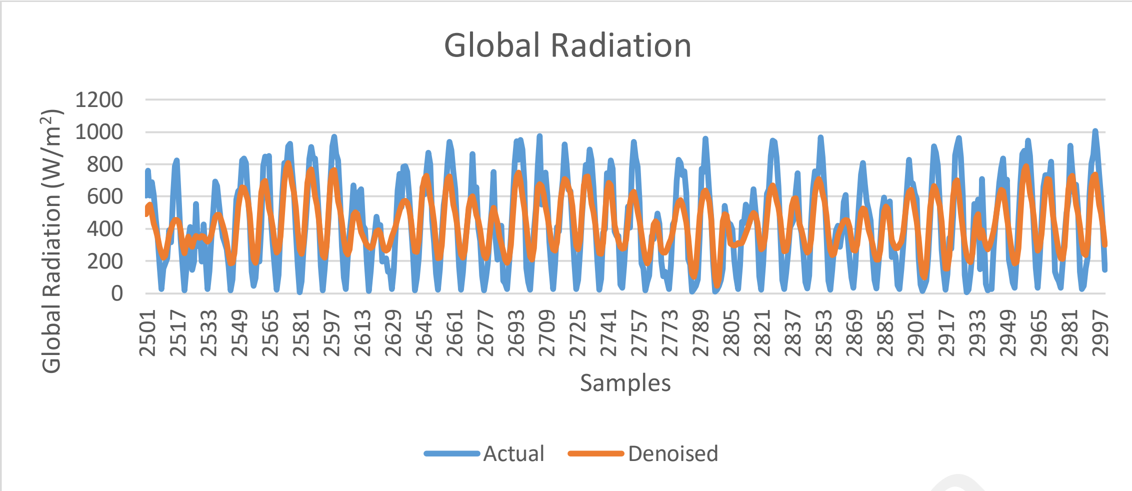


(b)

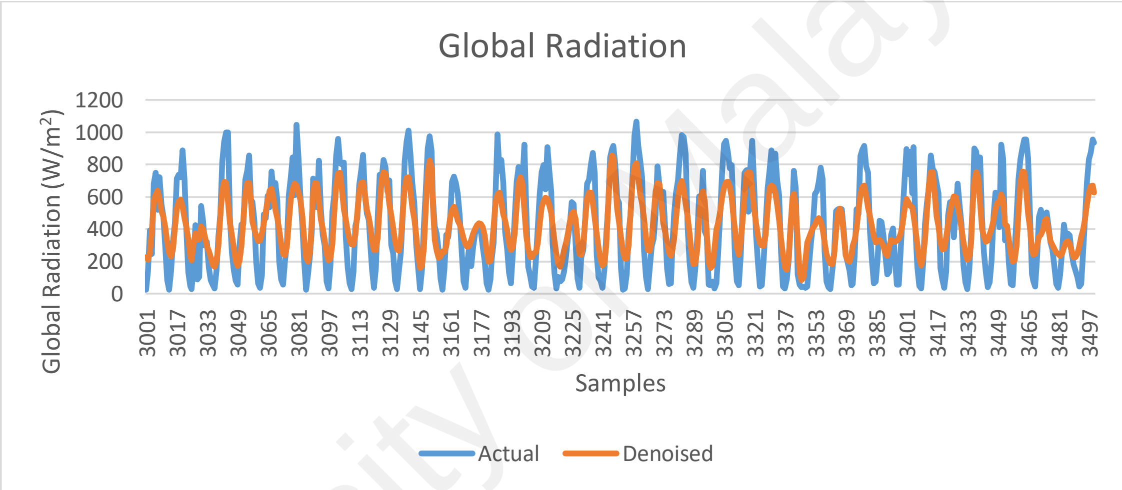


(c)

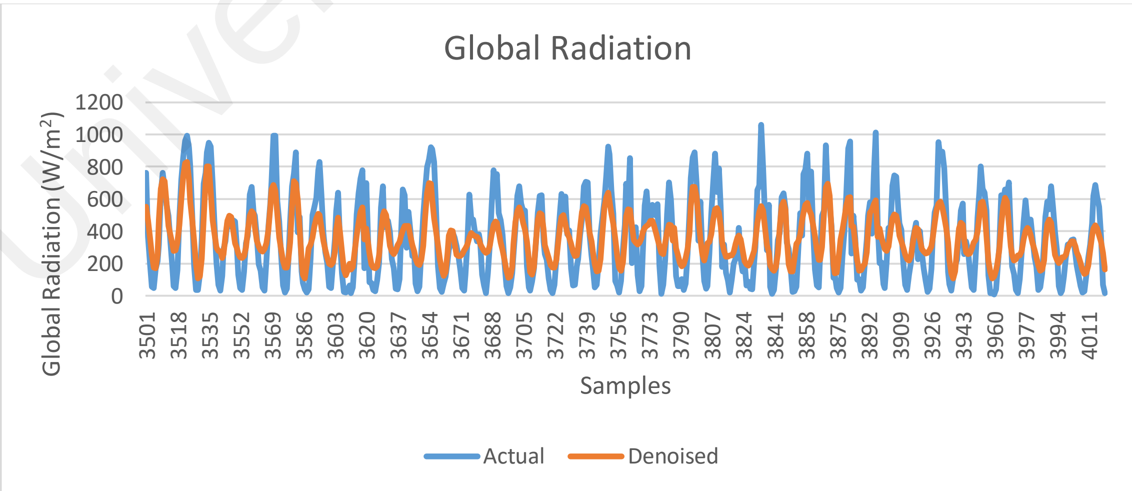
Figure 4.4: Actual and Denoised WT of Tilted Radiation (a) 2501th-3000th of Tested Data; (b) 3001th-3500^{h} of Tested Data; and (c) 3501th-4020th of Tested Data



(a)

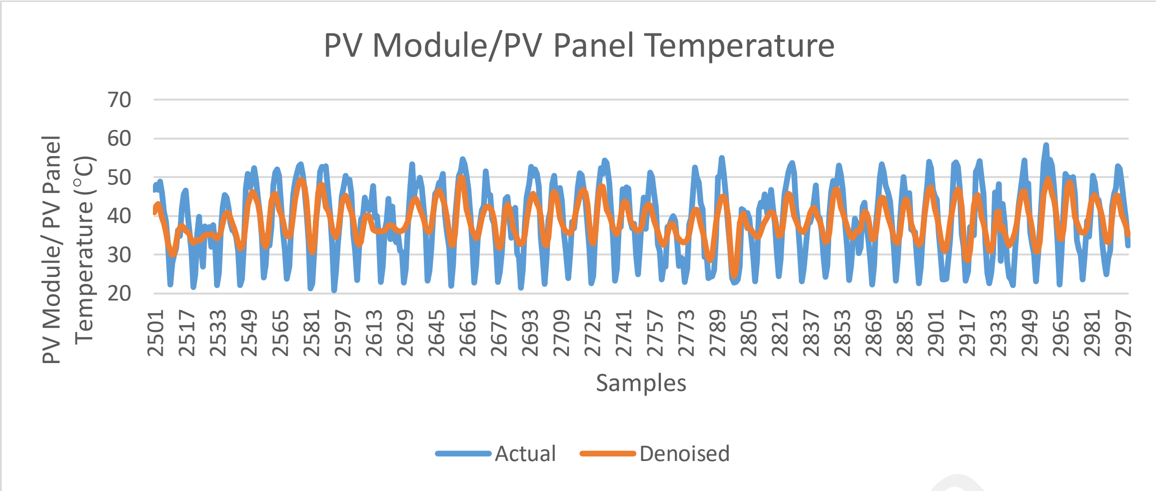


(b)

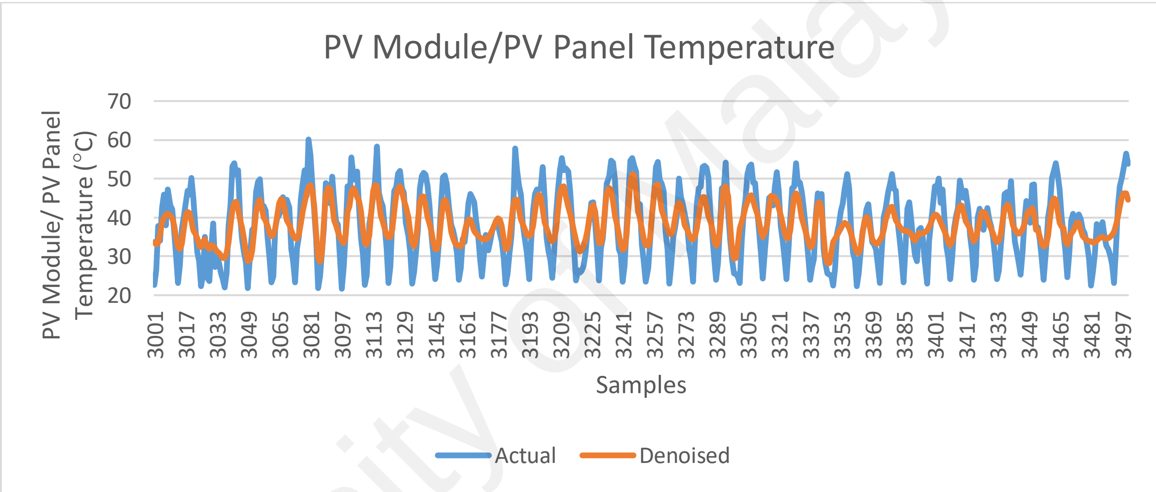


(c)

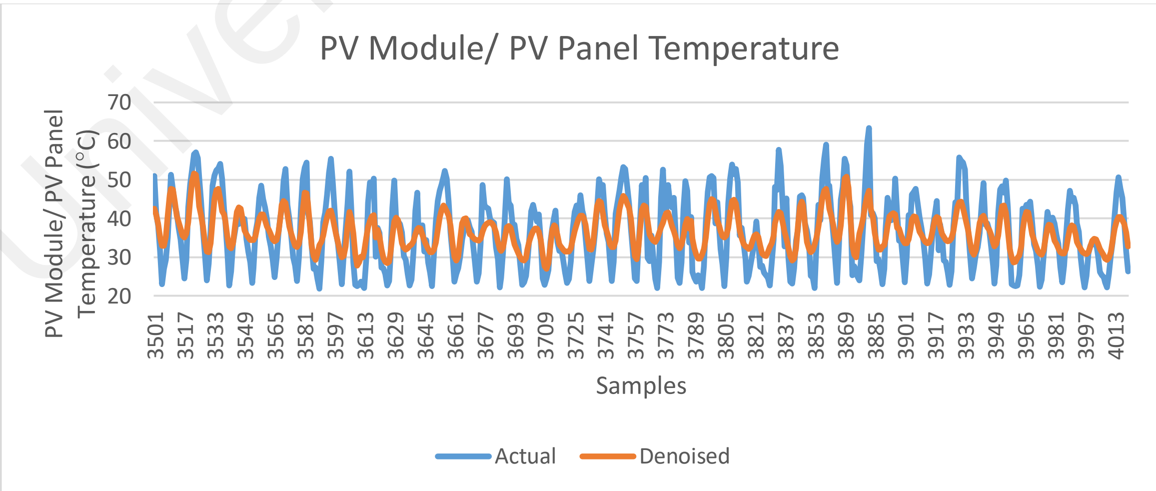
Figure 4.5: Actual and Denoised WT of Global Radiation (a) 2501th-3000th of Tested Data; (b) 3001th-3500h of Tested Data; and (c) 3501th-4020th of Tested Data



(a)

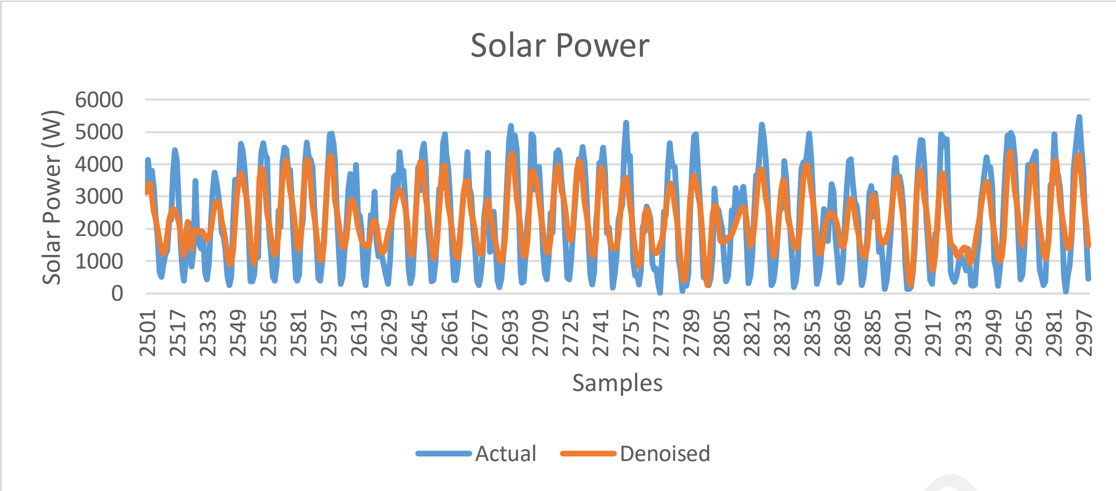


(b)

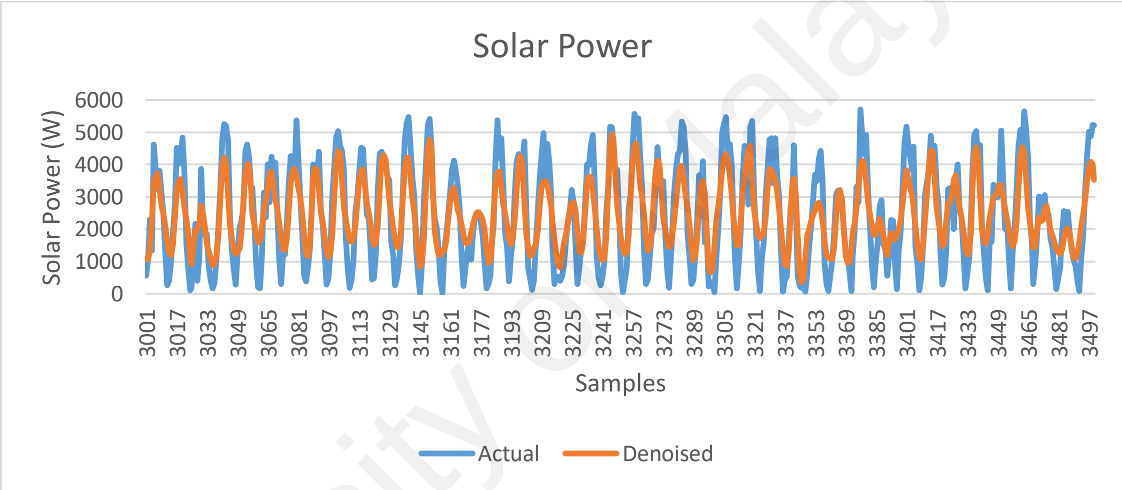


(c)

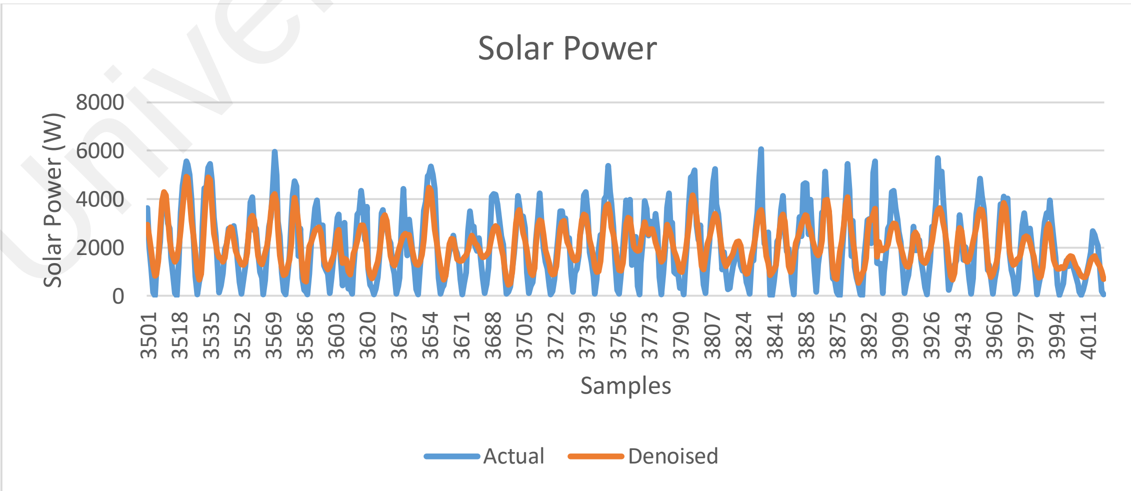
Figure 4.5: Actual and Denoised WT of PV Module/ PV Panel Temperature (a) 2501th-3000th of Tested Data; (b) 3001th-3500^{h} of Tested Data; and (c) 3501th-4020th of Tested Data



(a)



(b)



(c)

Figure 4.6: Actual and Denoised WT of Solar Power (a) 2501th-3000th of Tested Data; (b) 3001th-3500^{h} of Tested Data; and (c) 3501th-4020th of Tested Data

As can be seen in Figure 4.4 until Figure 4.7, the application of WT has eliminated the noise exists in the signals of tilted radiation, global radiation, PV module/PV panel temperature and solar power. The feasibility of WT can be further evaluated by inserting the denoised values into the forecasting models of MLP, RBFNN and ANFIS. The results of using the denoised WT signals in MLP, RBFNN and ANFIS can be obtained in subsection 4.3.1. subsection 4.3.2 and subsection 4.3.3.

4.3.1 WT-MLP Model

For the integration of WT with the MLP model (WT-MLP), the type of activation functions, the number of the hidden layer and the number of hidden neurons that are used in this study are described in Table 4.4.

Table 4.4: Important Parameters for WT-MLP

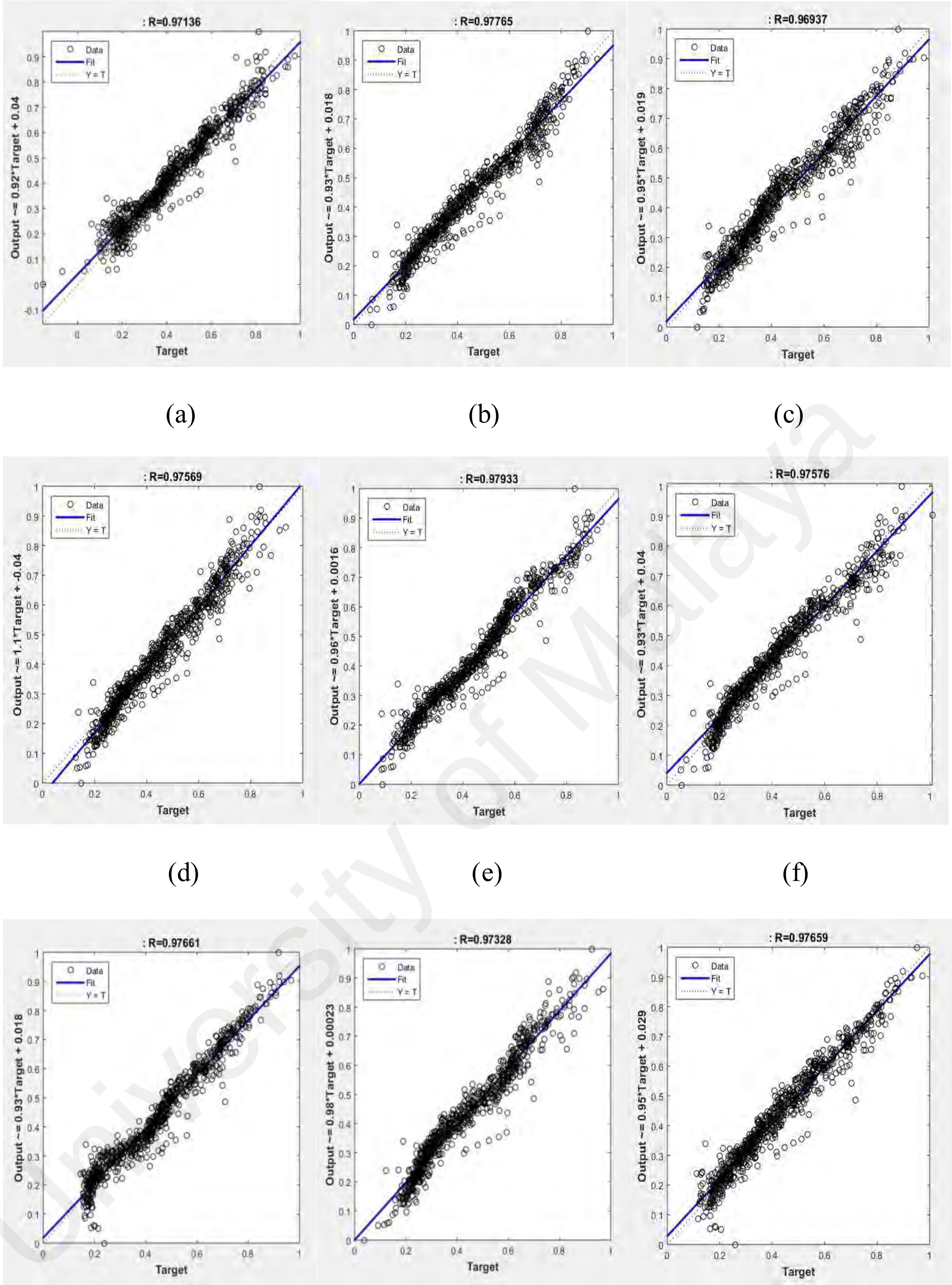
Network Type	**WT-MLP Model**
Number of Hidden Layer	1
Number of Hidden Neurons	18
Activation Function in Hidden Layer	*tansig*
Activation Function in Output Layer	*purelin*

The value of these parameters is chosen because of its supreme performance during solar power forecasting. On the other hand, different types of training algorithms are varied to find an algorithm that provides the greatest accuracy during forecasting.

Table 4.5 indicates the result of various training algorithms in the WT-MLP model and Figure 4.8 displays the regression plots of the tested training algorithms. From Table 4.5, the lowest MAE is depicted by the *traincgb* algorithm which gives a value of 0.0294 and the lowest RMSE value of 0.0410 is also obtained from the *traincgb* algorithm. The R-value metric defines a correlation between the actual and forecasted solar power data. According to R-values in Table 4.5 as well as regression plots in Figure 4.8, the highest R-value is attained from *traincgb* algorithm that provides the greatest value of 0.9793. The result proves that *traincgb* has outperformed other training algorithms by providing reliable results during solar power forecasting.

Table 4.5: Performance Comparison of Training Algorithm for WT-MLP Model

Algorithm	**MAE**	**RMSE**	**R**
trainlm	0.0343	0.0471	0.9714
trainrp	0.0302	0.0429	0.9777
trainbfg	0.0351	0.0468	0.9694
trainscg	0.0338	0.0449	0.9757
traincgb	***0.0294***	***0.0410***	***0.9793***
traincgf	0.0347	0.0441	0.9758
traincgp	0.0312	0.0430	0.9766
traingdx	0.0321	0.0431	0.9733
trainoss	0.0304	0.0415	0.9766



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 4.7: Regression Plots of Training Algorithms for WT-MLP: (a)trainlm; (b)trainrp; (c)trainbfg; (d)trainscg; (e)traincgb; (f)traincgf; (g)traincgp; (h)traingdx; (i)trainoss

4.3.2 WT- RBFNN Model

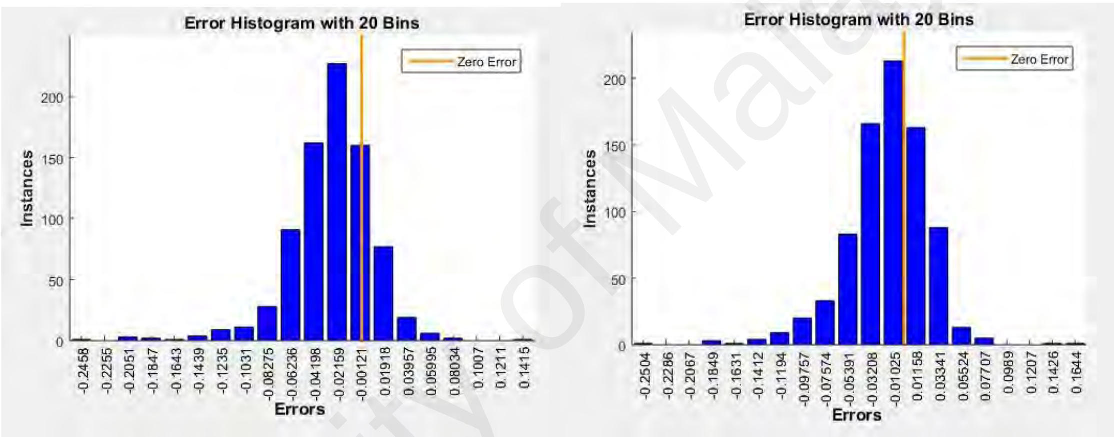
The incorporation of WT and RBFNN (WT-RBFNN) requires the measurement of several parameters which are the maximum number of hidden neurons, the centre of activation function and the spread of activation function. The trial and error approach is employed to select the maximum number of hidden neurons and the result displayed 40 hidden neurons as an ideal number. Similar to the RBFNN model in Subsection 4.2.2, the centre of the activation function is chosen according to the forward search strategy.

Table 4.6: Performance Comparison of Spread Value for WT-RBFNN Model

Spread	**MAE**	**RMSE**	**R**
1	0.0358	0.0576	0.9584
2	0.0397	0.0559	0.9689
3	0.0363	0.0498	0.9776
4	0.0356	0.0482	0.9794
5	0.0334	0.0450	***0.9820***
6	0.0335	0.0452	0.9817
7	0.0332	0.0451	0.9814
8	0.0330	0.0449	0.9815
9	***0.0303***	***0.0419***	0.9788
10	0.0305	0.0424	0.9769

The spread value of the activation function is varied to find a value that offers the highest accuracy. The result to vary the spread value of the activation function is tabulated in Table 4.6. It demonstrates that nine spread value outperforms other numbers of spread concerning MAE and RMSE values. It provides the MAE and RMSE values of 0.0303 and 0.0419, respectively, which are the lowest than other numbers of spread. On the other hand, the highest R-value of 0.9820 in Table 4.6 is obtained from five spread value instead of nine spread value. Due to this matter, an analysis of error histogram plots is performed and shown in Figure 4.9 to estimate which number of spread should be used in WT-RBFNN for solar power forecasting.

The result in Figure 4.9 signifies that the majority of errors in five spread values is located inside the value of 0.02159 which lies next to the zero error range values. On the contrary, most of the errors in nine spread values are inside the value of 0.01025 which are positioned at the zero error range values. Thus, this work chooses nine spread values as an ideal spread value for WT-RBFNN because the majority of errors in nine spread values are located in zero error range values.



(a) (b)

Figure 4.8: Error Histogram Bars of Tested Data for WT- RBFNN: (a) 5 Spread Value; (b) 9 Spread Value

4.3.3 WT-ANFIS Model

For the WT and ANFIS (WT-ANFIS) combination, the FCM clustering method is employed to divide the input and output data. Aforementioned in Subsection 4.2.3, *gaussmf* activation function and *linear* activation function is utilised as the membership function for the input neurons and the output neuron, respectively. Another control parameter which is the number of the cluster has been analysed and the result is tabulated in Table 4.7.

It can be seen from Table 4.7 that the performance of two clusters has achieved considerably higher accuracy than other numbers of the cluster. It depicts the MAE and RMSE values of 0.0278 and 0.0385, respectively, which are the lowest than the other number of clusters. Besides, the highest R-value of 0.9799 is obtained from two clusters which are surpassing the results of the other number of clusters.

Table 4.7: Performance Comparison of Number of Clusters for WT-ANFIS Model

Number of Clusters	**MAE**	**RMSE**	**R**
2	***0.0278***	***0.0385***	***0.9799***
3	0.0320	0.0413	0.9769
4	0.0298	0.0401	0.9778
5	0.0289	0.0396	0.9785
6	0.0300	0.0401	0.9770
7	0.0285	0.0396	0.9772
8	0.0297	0.0410	0.9777
9	0.0303	0.0416	0.9773
10	0.0320	0.0432	0.9774
11	0.0311	0.0422	0.9752
12	0.0325	0.0448	0.9715
13	0.0328	0.0440	0.9758
14	0.0307	0.0424	0.9734
15	0.0359	0.0479	0.9660

To verify the supremacy usage of two clusters in WT-ANFIS, the R-value in Table 4.7 is plotted according to the regression plot shown in Figure 4.10. According to Figure 4.10, majority of the forecasted solar power data is highly correlated with the actual solar power data. This signifies the reliability usage of two clusters in WT-ANFIS during solar power forecasting.

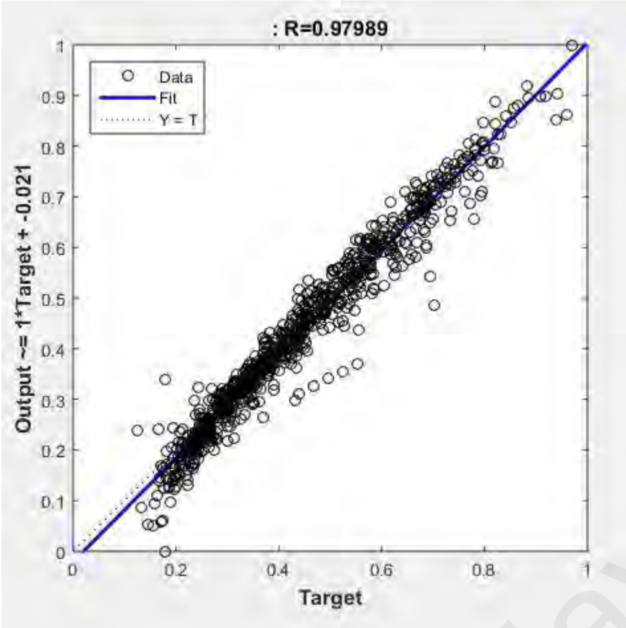


Figure 4.9: Regression Plot for WT-ANFIS Model

4.4 Phase 3: Evaluation of WT Performance for Noise Elimination Process and Selection of the Most Accurate Forecasting Model

The results in Phase 3 are divided into two categories. The first category compares the performance of MLP, RBFNN and ANFIS in Phase 1 and Phase 2. The comparison is made to prove the superiority of WT technique to remove noisy data. The second category estimates the values of RMSE, MAE and R for every forecasting model in Phase 1 and Phase 2. A model that can provide the lowest values of RMSE and MAE, as well as the highest value of R, is nominated as the most accurate forecasting model.

4.4.1 Evaluation of WT Performance for Noise Elimination Process

The effectiveness of using the WT approach is estimated by comparing the results of the forecasting models in Phase 1 and Phase 2. In this regard, three forecasting models are being employed which are identified as MLP, RBFNN and ANFIS. The results which are previously shown in Phase 1 and Phase 2 is vividly described in the following Table 4.8.

Table 4.8: Performance Comparison of Forecasting Models with and without Utilisation of WT

Model	Performance Metrics					
	MAE		RMSE		R	
	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2
MLP	0.0471	0.0294	0.0617	0.0410	0.9709	0.9793
RBFNN	0.0467	0.0303	0.0604	0.0419	0.9722	0.9788
ANFIS	0.0504	0.0278	0.0647	0.0385	0.9674	0.9799

The MAE and RMSE values in Table 4.8 is further described in the histogram plots in Figure 4.11 and Figure 4.12. As can be seen from both figures, the employment of WT to remove noise in the data has improved the performance of MLP, RBFNN and ANFIS. This can be proven when the values of MAE and RMSE for MLP, RBFNN and ANFIS are lowered from Phase 1 to Phase 2.

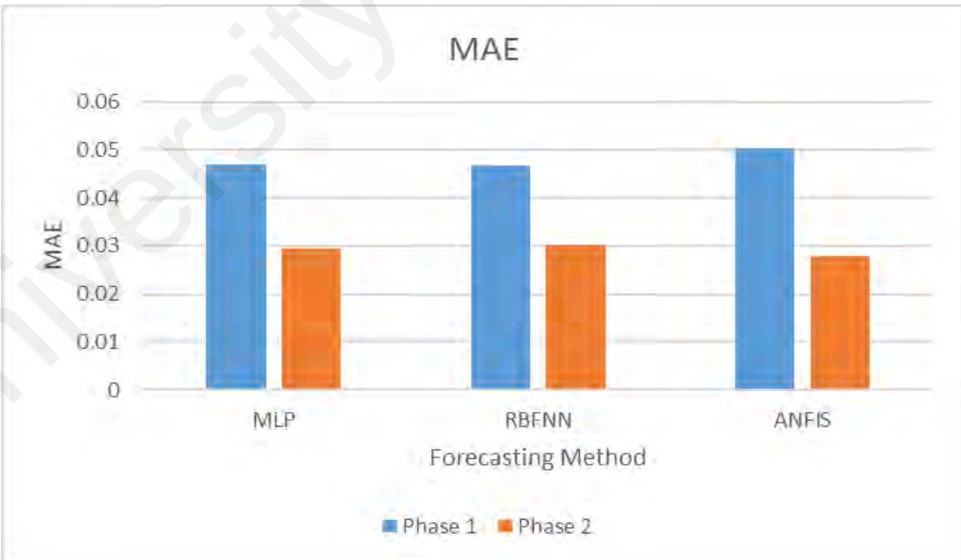


Figure 4.10: MAE Comparison for Forecasting Models in Phase 1 and Phase 2

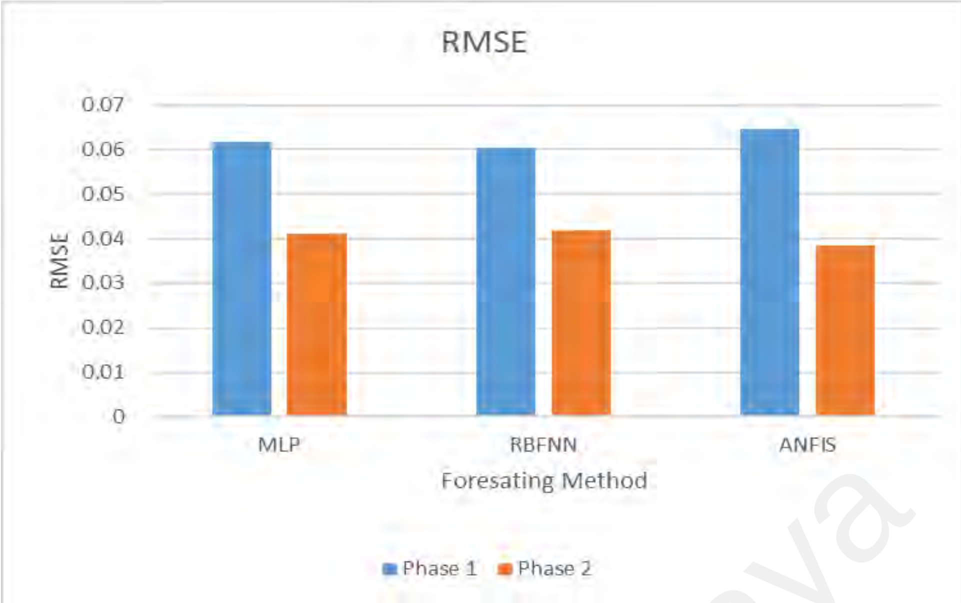


Figure 4.11: RMSE Comparison for Forecasting Models in Phase 1 and Phase 2

Apart from that, the R-values of MLP, RBFNN and ANFIS have been improved from 0.9709, 0.9722 and 0.9674 in Phase 1 to 0.9793, 0.9788 and 0.9799 in Phase 2, respectively. The high value of R indicates a high correlation between the actual data and forecasted data. Hence, the results had proven the applicability of WT to remove the noisy data because it manages to improve the performance of MLP, RBFNN and ANFIS during solar power forecasting.

4.4.2 Selection of the Most Accurate Forecasting Model

It is highly essential to regulate the performance metrics of every forecasting model to propose a model that can highlight the most accurate results among other forecasting models. For this reason, the performance metrics of RMSE and MAE for every model is portrayed in Figure 4.13. The results depict that the combination of WT and ANFIS method (WT-ANFIS) is surpassing the other five strategies by providing the lowest MAE value of 0.0278 and lowest RMSE value of 0.0385.

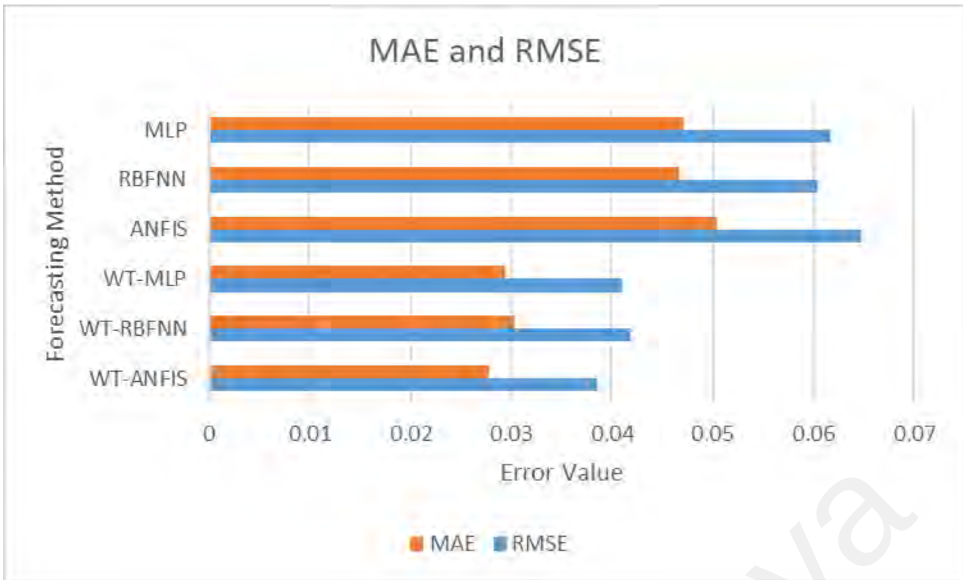


Figure 4.12: Comparison of MAE and RMSE Values for Every Forecasting Model

Apart from the MAE and RMSE values, another metric that needs to be considered in choosing the most accurate forecasting model is according to R-value. The information of R-values which are previously shown in Tables 4.8 is vividly described in the following Figure 4.14. According to Figure 4.14, the highest prediction accuracy has been achieved by the combination of WT and ANFIS method (WT-ANFIS) by providing the R-value of 0.97989. Hence, the WT-ANFIS technique has been elected as the most accurate model for solar power forecasting.

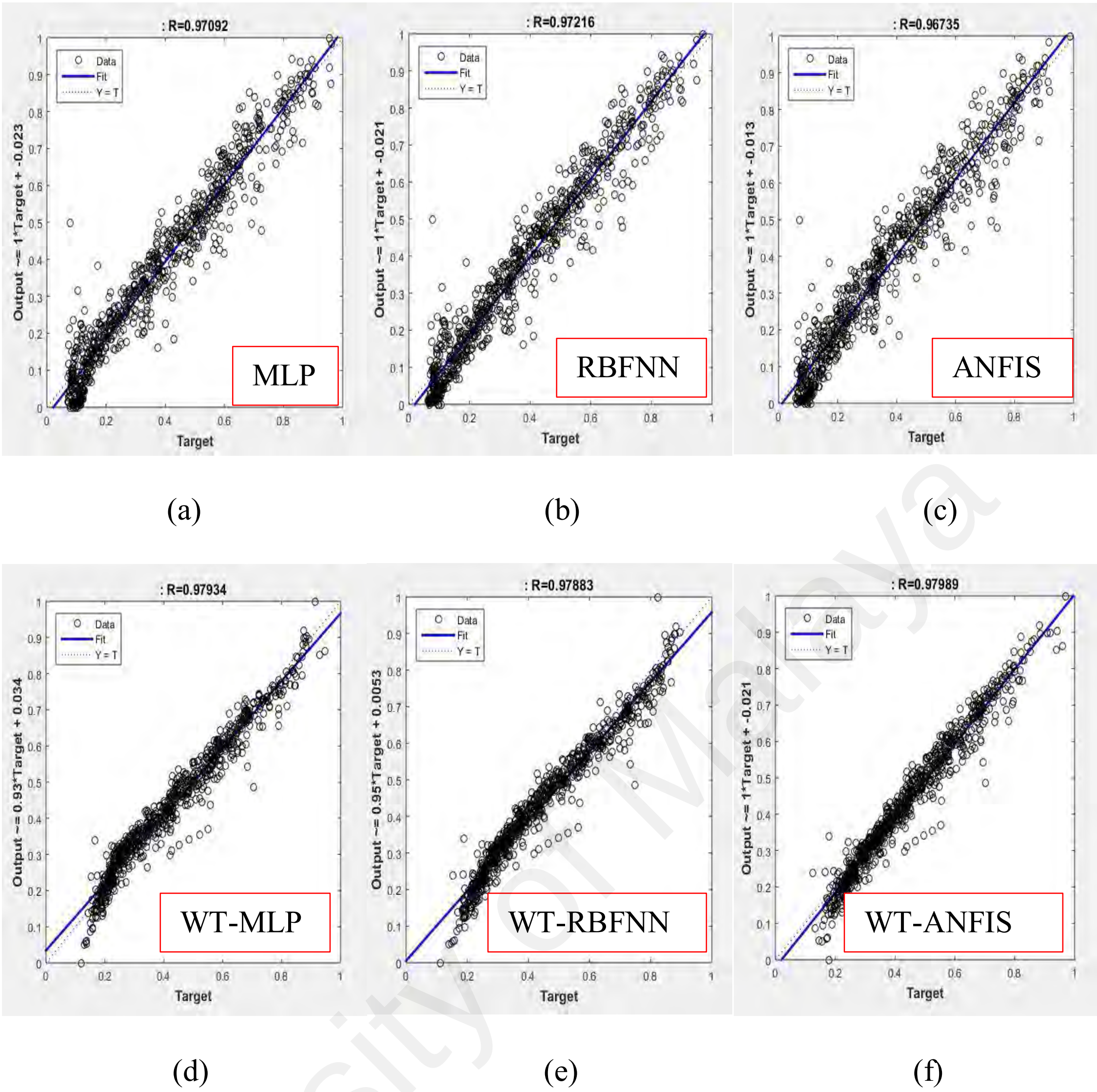
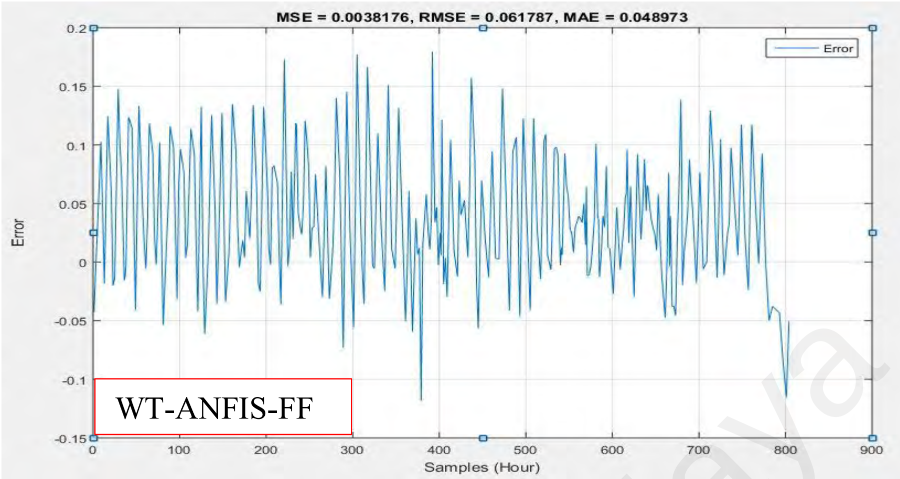


Figure 4.13: Comparison of R-Value for Every Forecasting Model: (a)MLP; (b)RBFNN; (c)ANFIS; (d)WT-MLP; (e)WT-RBFNN; (f)WT-ANFIS

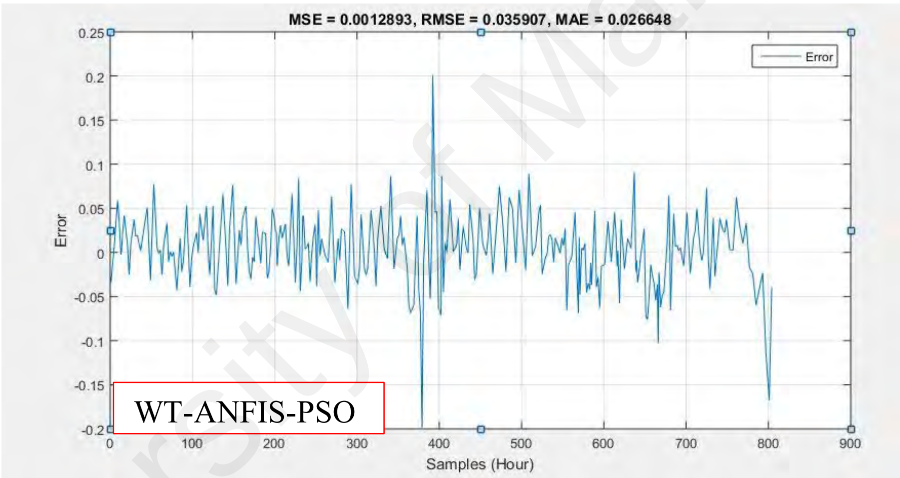
4.5 Phase 4: Model Optimisation by using HFPSO

An optimisation algorithm will optimise a set of parameters to minimise the objective function of a forecasting model. In this work, the HFPSO algorithm is used to optimise the parameters of the most accurate forecasting model. As can be seen in Subsection 4.4.2, the model comparison has defined WT-ANFIS as the most accurate model during solar power forecasting. Hence, the HFPSO algorithm is used to optimise the premise parameters of the WT-ANFIS model. The results from the proposed WT-ANFIS-HFPSO is compared with the results of WT-ANFIS combined with single FF

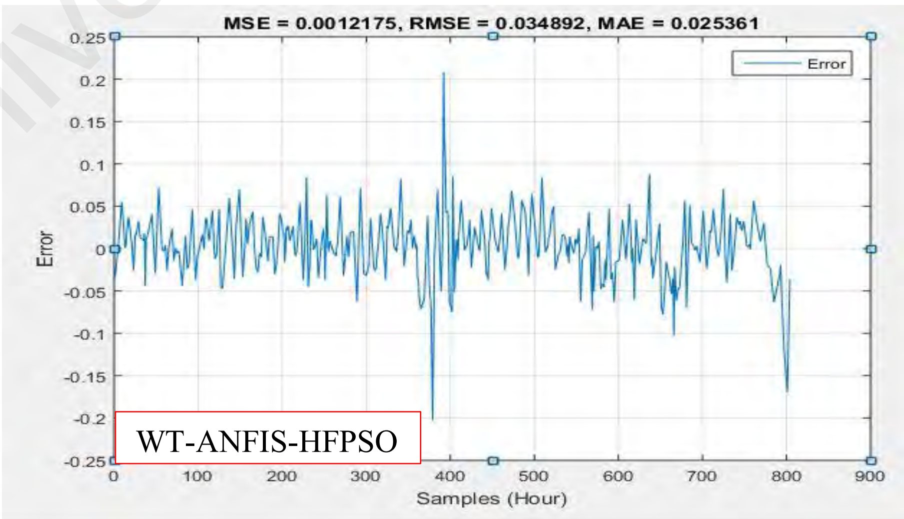
(WT-ANFIS-FF) and WT-ANFIS combined with single PSO (WT-ANFIS-PSO) algorithm.



(a)



(b)

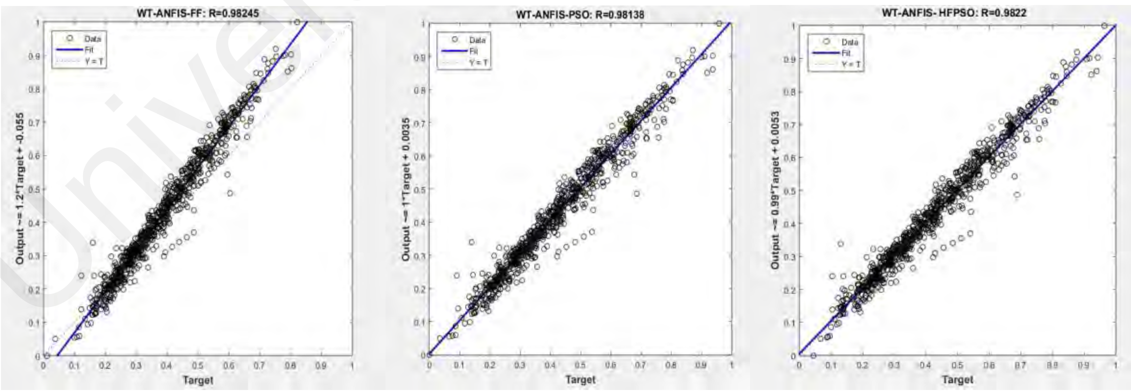


(c)

Figure 4.14: MSE, RMSE and MAE Comparison: (a) WT-ANFIS-FF; (b) WT-ANFIS-PSO; (c) WT-ANFIS-HFPSO

The comparison results are analysed according to several performance metrics. Figure 4.15 visuals the comparison of the MSE, RMSE and MAE values when using the WT-ANFIS-FF, WT-ANFIS-PSO and WT-ANFIS-HFPSO. It is observed that the values of MSE=0.0012175, RMSE=0.034892 and MAE=0.025361 which are obtained from the WT-ANFIS-HFPSO are the lowest compared to the WT-ANFIS-FF and WT-ANFIS-PSO algorithms. The hybrid model of HFPSO has increased the performance of WT-ANFIS when it is compared to the single FF and single PSO algorithms.

Meanwhile, Figure 4.16 portrays the regression plots for WT-ANFIS-FF, WT-ANFIS-PSO and WT-ANFIS-HFPSO. It can be observed that the method of WT-ANFIS-FF presents the highest R-value of 0.98245 compared to the models of WT-ANFIS-HFPSO and WT-ANFIS-PSO. However, WT-ANFIS-FF cannot be considered as the supreme model for solar power forecasting because there are deviations of some estimated areas with the actual values as shown in Figure 4.16(a). Hence, WT-ANFIS-HFPSO model is selected as the most supreme forecaster of solar power because it gives a higher R-value than WT-ANFIS-PSO and there are very small deviations existed between the estimated areas from the actual values.

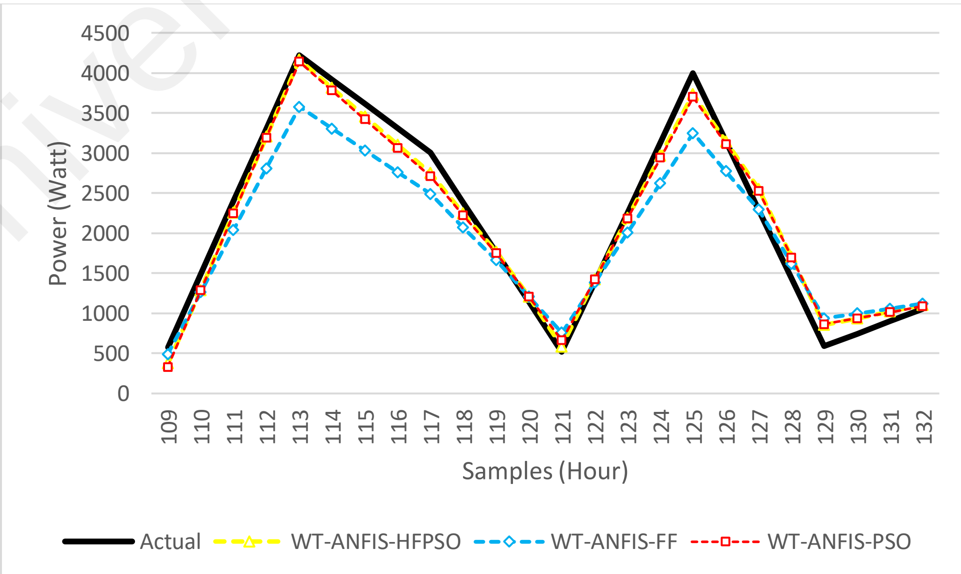


(a) (b) (c)

Figure 4.15: Regression Plots Comparison: (a) WT-ANFIS-FF: R=0.98245; (b) WT-ANFIS-PSO: R=0.98138 ; (c) WT-ANFIS-HFPSO: R=0.98220

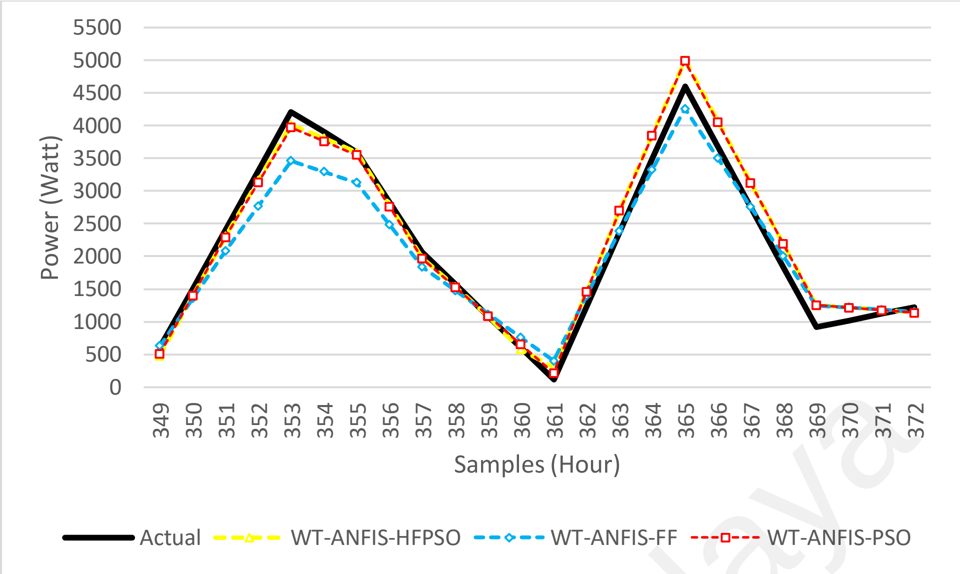
According to Figure 4.15 and Figure 4.16, a single FF algorithm and a single PSO algorithm do not give a good performance for solar power forecasting. This might due to the condition of FF algorithm that is getting trapped in the local minima as well as due to the premature convergence problem faced by the PSO algorithm. Hence, a hybridisation of both algorithms will mitigate the above-mentioned problems and give a better precision for solar power forecasting.

To have a better view of result for the WT-ANFIS-HFPSO model, Figure 4.17 illustrates the comparisons of the actual and forecasted power profile for WT-ANFIS-HFPSO, WT-ANFIS-FF and WT-ANFIS-PSO simultaneously. According to Figure 4.17, the forecasted solar power values that are attained from WT-ANFIS-PSO and WT-ANFIS-FF are close to the actual values of solar power. Though, the WT-ANFIS that is optimised by the HFPSO algorithm provides better results as the majority of the forecasted solar power data are very close to the actual solar power data. In the end, it is concluded that the simulation results in Phase 4 have shown the potential benefit of WT-ANFIS-HFPSO as an accurate forecaster of solar power.

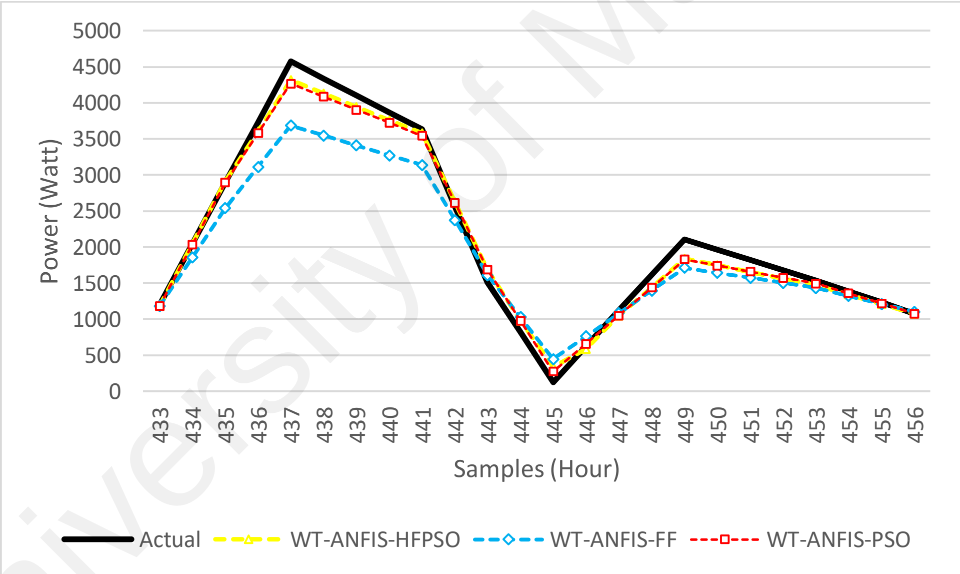


(a)

Figure 4.16: Actual and Forecasted Values of Tested Data: (a) 109th-132th of Tested Data; (b) 349th-372th of Tested Data; and (c) 433th-456th of Tested Data



(b)



(c)

Figure 4.17, continued

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

This chapter concludes the overall work for solar power forecasting by using a noise elimination technique(WT), several forecasting models (MLP, RBFNN and ANFIS) as well as an optimisation algorithm (HFPSO). This chapter is further divided into two subsections which are conclusions and recommendations for further work.

5.1 Conclusions

The first phase (Phase 1) of the simulation results discusses the performance of the MLP, RBFNN and ANFIS when employing noisy data for solar power forecasting. For the MLP model, one hidden layer with 12 hidden neurons are chosen in this study. Meanwhile, *logsig* and *purelin* are selected as the activation functions for the hidden layer and output layer, respectively. This work has tested the efficiency of the MLP model according to several training algorithms and *trainrp* has been chosen as an ideal one. Apart from that, RBFNN employs 40 hidden neurons and the centre of the activation function is selected according to the forward search strategy. The spread value of RBFNN is varied and the result has shown the improved result of five spread value compared to the other number of spreads. In the case of ANFIS model, FCM clustering approach is used to divide the input and output data into the certain number of clusters and this work found that six number of clusters are outperforming the results of the other number of clusters.

For the second phase of result (Phase 2), the WT is integrated with the MLP, RBFNN and ANFIS model. The WT-MLP model selects one hidden layer with 18 hidden neurons during the forecasting process of solar power. Moreover, *tansig* is chosen as the activation function at the hidden layer while *purelin* is denoted as the activation function at the output layer. Various training algorithms are used in the forecasting process of WT-MLP and *traincgb* is chosen as an ideal algorithm. In the WT-RBFNN forecasting, 40 hidden neurons are elected and forward search strategy is used to estimate the centre of the activation function. Similar to the RBFNN in Phase 1, the spread value is varied and the result demonstrates that nine spread value has outperformed the other number of spreads. The WT-ANFIS applies the FCM clustering method for the division of input and output data into the number of clusters. The result shows that the performance of two clusters has achieved considerably higher accuracy than the other number of the clusters.

To show the effectiveness of the WT as a noise elimination tool, the results in the third phase (Phase 3) has compared the forecasting accuracy of MLP, RBFNN and ANFIS in both phases (Phase 1 and Phase 2). Any forecasting model that utilises the denoised data from the WT achieves a higher forecasting accuracy. This is specifically proven when MAE and RMSE values for MLP, RBFNN and ANFIS are lowered from Phase 1 to Phase 2. Apart from that, the R-values of MLP, RBFNN and ANFIS have been improved from 0.9709, 0.9722 and 0.9674 in Phase 1 to 0.9793, 0.9788 and 0.9799 in Phase 2, respectively.

Phase 3 result of this work also covers the selection process of the most accurate forecasting model. It is worth mentioning that the integration of WT with the ANFIS has been selected as the most accurate forecasting model. The WT-ANFIS has given the MAE and RMSE values of 0.0278 and 0.0385, respectively, outperforming the other forecasting strategies.

This work also proposes the HFPSO algorithm to find optimised premise parameters for the WT-ANFIS model. It appears from the result that the WT-ANFIS-HFPSO forecasts solar power with better accuracy than WT-ANFIS-FF and WT-ANFIS-PSO. This is proven when WT-ANFIS-HFPSO provides the values of MSE=0.0012175, RMSE=0.034892 and MAE=0.025361 which are the lowest compared to WT-ANFIS-FF and WT-ANFIS-PSO. Other than that, the WT-ANFIS-HFPSO model provides an R-value=0.98220 that indicates the capability of the model to follow the pattern of the data efficiently. Furthermore, the actual and forecasted power profiles of WT-ANFIS-HFPSO have shown that majority of the forecasted solar power data are very close to the actual solar power data. In conclusion, all of the results in Phase 4 have verified the efficiency of WT-ANFIS-HFPSO as the superior forecasting model of solar power.

5.2 Further Works

Three aspects, namely, noise elimination aspect, forecasting aspect and optimisation aspect have been considered in this study. The proposed works are effective to be used in the forecasting area, particularly in solar power forecasting. However, there is still room of improvements that can be suggested in the future. Those recommendations are listed as follows:

- The data from other locations can be introduced for solar power forecasting. It is important to enable the generalisation of the proposed works to different locations.
- The forecast horizon of every forecasting model can be varied to days, weeks or months. The variation of forecast-horizon will estimate the practicality of the proposed works on the short-term and long-term forecasting horizon
- The performance of other noise elimination techniques can be estimated. The outcomes of other noise elimination techniques can be compared with the outcome of WT. It is essential to prove the superiority of WT as a noise elimination technique.
- The performance of other hybrid optimisation algorithms can be included. The results of other hybrid optimisation algorithms can be compared with the result of HFPSO to identify a hybrid optimisation algorithm that can provide a better performance

REFERENCES

Abdullah, Deris, Mohamad, & Hashim. (2012). A new hybrid firefly algorithm for complex and nonlinear problem *Distributed Computing and Artificial Intelligence* (pp. 673-680): Springer.

Abraham. (2001). *Neuro fuzzy systems: State-of-the-art modeling techniques.* Paper presented at the International Work-Conference on Artificial Neural Networks.

Abuella. (2015, April). *Solar power probabilistic forecasting by using multiple linear regression analysis*. Paper presented at the SoutheastCon 2015.

Adeoti, & Osanaiye. (2012). Performance analysis of ANN on dataset allocations for pattern recognition of bivariate process. *Math. Theory Model, 2*, 53-63.

Adewuya. (1996). *New methods in genetic search with real-valued chromosomes.* Massachusetts Institute of Technology.

Adrian, Utamima, & Wang. (2015). A comparative study of GA, PSO and ACO for solving construction site layout optimization. *KSCE Journal of Civil Engineering, 19*(3), 520-527.

Agayev. (2015). Solar Radiation Data Analysis In Baku By Using Daubechies Wavelets. *IJISM, 3*(3), 163-167.

Aguiar, Pereira, Lauret, Díaz, & David. (2016). Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting. *Renewable Energy, 97*, 599-610.

Akarslan, & Hocaoglu. (2016). A novel adaptive approach for hourly solar radiation forecasting. *Renewable Energy, 87*, 628-633.

Al-Shamisi, Assi, & Hejase. (2013). Artificial neural networks for predicting global solar radiation in al ain city-uae. *International Journal of Green Energy, 10*(5), 443-456.

Alessio, Carbone, Castelli, & Frappietro. (2002). Second-order moving average and scaling of stochastic time series. *The European Physical Journal B-Condensed Matter and Complex Systems, 27*(2), 197-200.

Ali. (2014). A review of firefly algorithm. *ARPN Journal of Engineering and Applied Sciences, 9*(10), 1732-1736.

Alonso-Montesinos, Batlles, & Portillo. (2015). Solar irradiance forecasting at one-minute intervals for different sky conditions using sky camera images. *Energy Conversion and Management, 105*, 1166-1177.

Amral, Ozveren, & King. (2007). *Short term load forecasting using multiple linear regression.* Paper presented at the 2007 42nd International universities power engineering conference.

Arunachalam, AgnesBhomila, & Babu. (2014). *Hybrid particle swarm optimization algorithm and firefly algorithm based combined economic and emission dispatch including valve point effect.* Paper presented at the International Conference on Swarm, Evolutionary, and Memetic Computing.

Awad, & Qasrawi. Enhanced RBF neural network model for time series prediction of solar cells panel depending on climate conditions (temperature and irradiance). *Neural Computing and Applications*, 1-12.

Awan, Aslam, Khan, & Saeed. (2014). An efficient model based on artificial bee colony optimization algorithm with Neural Networks for electric load forecasting. *Neural Computing and Applications, 25*(7-8), 1967-1978.

Aybar-Ruiz, Jiménez-Fernández, Cornejo-Bueno, Casanova-Mateo, Sanz-Justo, Salvador-González, & Salcedo-Sanz. (2016). A novel grouping genetic algorithm–extreme learning machine approach for global solar radiation prediction from numerical weather models inputs. *Solar Energy, 132*, 129-142.

Aydilek. (2018). A hybrid firefly and particle swarm optimization algorithm for computationally expensive numerical problems. *Applied Soft Computing, 66*, 232-249.

Bae, Jang, & Sung. (2016). Hourly solar irradiance prediction based on support vector machine and its error analysis. *IEEE Transactions on power systems, 32*(2), 935-945.

Bahrami, Hooshmand, & Parastegari. (2014). Short term electric load forecasting by wavelet transform and grey model improved by PSO (particle swarm optimization) algorithm. *Energy, 72*, 434-442.

Bashir, & El-Hawary. (2009). Applying wavelets to short-term load forecasting using PSO-based neural networks. *IEEE transactions on power systems, 24*(1), 20-27.

Basu, & Halder. (2017). Importance of Numerical Weather Prediction in Variable Renewable Energy Forecast.

Behera, Majumder, & Nayak. (2018). Solar photovoltaic power forecasting using optimized modified extreme learning machine technique. *Engineering Science and Technology, an International Journal, 21*(3), 428-438.

Benali, Notton, Fouilloy, Voyant, & Dizene. (2019). Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components. *Renewable energy, 132*, 871-884.

Benmouiza, & Cheknane. (2019). Clustered ANFIS network using fuzzy c-means, subtractive clustering, and grid partitioning for hourly solar radiation forecasting. *Theoretical and applied climatology, 137*(1-2), 31-43.

Bouzerdoum, Mellit, & Pavan. (2013). A hybrid model (SARIMA–SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. *Solar Energy, 98*, 226-235.

Catalão, Pousinho, & Mendes. (2011). Short-term wind power forecasting in Portugal by neural networks and wavelet transform. *Renewable Energy, 36*(4), 1245-1251.

Çelik, Teke, & Yıldırım. (2016). The optimized artificial neural network model with Levenberg–Marquardt algorithm for global solar radiation estimation in Eastern Mediterranean Region of Turkey. *Journal of cleaner production, 116*, 1-12.

Chen, & Chen. (2014). Online fuzzy time series analysis based on entropy discretization and a Fast Fourier Transform. *Applied Soft Computing, 14*, 156-166.

Chen, Li, & Wu. (2013). Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration. *Energy Conversion and Management, 75*, 311-318.

Cohen. (2012). Signal denoising using wavelets. *Project Report, Department of Electrical Engineering Technion, Israel Institute of Technology, Haifa*.

Colak, Yesilbudak, Genc, & Bayindir. (2015). *Multi-period prediction of solar radiation using ARMA and ARIMA models.* Paper presented at the 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA).

Cornaro, Pierro, & Bucci. (2015). Master optimization process based on neural networks ensemble for 24-h solar irradiance forecast. *Solar Energy, 111*, 297-312.

Crabtree, Misewich, Ambrosio, Clay, DeMartini, James, . . . Sauer. (2011). *Integrating renewable electricity on the grid.* Paper presented at the AIP Conference proceedings.

Cruz, & Wishart. (2006). Applications of machine learning in cancer prediction and prognosis. *Cancer informatics, 2*, 117693510600200030.

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

Doan, & Liong. (2004). *Generalization for multilayer neural network bayesian regularization or early stopping.* Paper presented at the Proceedings of Asia Pacific Association of Hydrology and Water Resources 2nd Conference.

Dong, Yang, Reindl, & Walsh. (2014). Satellite image analysis and a hybrid ESSS/ANN model to forecast solar irradiance in the tropics. *Energy Conversion and Management, 79*, 66-73.

Dorigo, Maniezzo, & Colorni. (1996). Ant system: optimization by a colony of cooperating agents. *IEEE Transactions on Systems, man, and cybernetics, Part B: Cybernetics, 26*(1), 29-41.

Ekici. (2014). A least squares support vector machine model for prediction of the next day solar insolation for effective use of PV systems. *Measurement, 50*, 255-262.

Elsinga, & van Sark. (2017). Short-term peer-to-peer solar forecasting in a network of photovoltaic systems. *Applied Energy, 206*, 1464-1483.

Eseye, Zhang, & Zheng. (2018). Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and Meteorological information. *Renewable Energy, 118*, 357-367.

Fadare. (2009). Modelling of solar energy potential in Nigeria using an artificial neural network model. *Applied Energy, 86*(9), 1410-1422.

Fei, & He. (2015). Wind speed prediction using the hybrid model of wavelet decomposition and artificial bee colony algorithm-based relevance vector machine. *International Journal of Electrical Power & Energy Systems, 73*, 625-631.

Fernandez-Jimenez, Muñoz-Jimenez, Falces, Mendoza-Villena, Garcia-Garrido, Lara-Santillan, . . . Zorzano-Santamaria. (2012). Short-term power forecasting system for photovoltaic plants. *Renewable Energy, 44*, 311-317.

Fu, Liu, Tong, Wang, & Zhao. (2015). *A novel firefly algorithm based on improved learning mechanism.* Paper presented at the International conference on logistics engineering, management and computer science (LEMCS 2015).

Gerhardt, & Gomes. (2012). *Artificial bee colony (ABC) algorithm for engineering optimization problems.* Paper presented at the International Conference on Engineering Optimization.

Ghanbarzadeh, Noghrehabadi, Assareh, & Behrang. (2009). *Solar radiation forecasting based on meteorological data using artificial neural networks.* Paper presented at the 2009 7th IEEE International Conference on Industrial Informatics.

Gohari, Urquhart, Yang, Kurtz, Nguyen, Chow, . . . Kleissl. (2014). Comparison of solar power output forecasting performance of the Total Sky Imager and the University of California, San Diego Sky Imager. *Energy Procedia, 49*, 2340-2350.

Gupta, & Gupta. (2014). Evaluation of a New Modified Firefly Algorithm. *2014 Recent Advances and Innovations in Engineering (Icraie)*.

Gürbüz, Öztürk, & Pardalos. (2013). Prediction of electricity energy consumption of Turkey via artificial bee colony: a case study. *Energy Systems, 4*(3), 289-300.

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

Halabi, Mekhilef, & Hossain. (2018). Performance evaluation of hybrid adaptive neuro-fuzzy inference system models for predicting monthly global solar radiation. *Applied Energy, 213*, 247-261.

Haldan. (2015). How Much Training Data Do You Need? , from https://medium.com/@malay.haldar/how-much-training-data-do-you-need-da8ec091e956

Han, Cheng, Xin, & Yan. (2007). Frequent pattern mining: current status and future directions. *Data mining and knowledge discovery, 15*(1), 55-86.

Haque, Nehrir, & Mandal. (2013). *Solar PV power generation forecast using a hybrid intelligent approach.* Paper presented at the 2013 IEEE Power & Energy Society General Meeting.

Hassan. (2014). ARIMA and regression models for prediction of daily and monthly clearness index. *Renewable Energy, 68*, 421-427.

Hocaoglu, & Serttas. (2017). A novel hybrid (Mycielski-Markov) model for hourly solar radiation forecasting. *Renewable Energy, 108*, 635-643.

Hong. (2011). Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm. *Energy, 36*(9), 5568-5578.

Hong. (2011, July). *A naïve multiple linear regression benchmark for short term load forecasting*. Paper presented at the 2011 IEEE Power and Energy Society General Meeting.

Huang, Cao, Peng, Li, Zhang, Wang, . . . Wang. (2018). Day-Ahead Forecasting of Hourly Photovoltaic Power Based on Robust Multilayer Perception. *Sustainability, 10*(12), 4863.

Huang, & Shih. (2003). Short-term load forecasting via ARMA model identification including non-Gaussian process considerations. *IEEE Transactions on power systems, 18*(2), 673-679.

Ibrahim, & Khatib. (2017). A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm. *Energy Conversion and Management, 138*, 413-425.

Illias, Chai, Abu Bakar, & Mokhlis. (2015). Transformer incipient fault prediction using combined artificial neural network and various particle swarm optimisation techniques. *PLOS one, 10*(6), e0129363.

Inc. (2014). TSI-880 Automatic Total Sky Imager. from http://www.yesinc.com/products/data/tsi880/

Inman, Pedro, & Coimbra. (2013). Solar forecasting methods for renewable energy integration. *Progress in energy and combustion science, 39*(6), 535-576.

Ismail, Mamat, Hamzah, & Karim. (2014). *Forecasting performance of denoising signal by Wavelet and Fourier Transforms using SARIMA model.* Paper presented at the AIP Conference Proceedings.

Ismail, Mamat, Hamzah, & Karim. (2014). Forecasting Performance of Denoising Signal by Wavelet and Fourier Transforms using SARIMA Model. *Proceedings of the 21st National Symposium on Mathematical Sciences (Sksm21): Germination of Mathematical Sciences Education and Research Towards Global Sustainability, 1605*, 961-966. doi: 10.1063/1.4887720

Jadidi, Menezes, de Souza, & de Castro Lima. (2018). A hybrid ga–mlpnn model for one-hour-ahead forecasting of the global horizontal irradiance in Elizabeth city, North Carolina. *Energies, 11*(10), 2641.

Jallad, Mekhilef, Mokhlis, Laghari, & Badran. (2018). Application of hybrid meta-heuristic Techniques for optimal load shedding planning and operation in an islanded distribution network integrated with distributed generation. *Energies, 11*(5), 1134.

Jang, & Mizutani. (1996). *Levenberg-Marquardt method for ANFIS learning.* Paper presented at the Proceedings of North American Fuzzy Information Processing.

Ji, & Chee. (2011). Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN. *Solar Energy, 85*(5), 808-817.

Karaboga, & Akay. (2009). A comparative study of artificial bee colony algorithm. *Applied mathematics and computation, 214*(1), 108-132.

Kenton. (2019). Multiple Linear Regression – MLR Definition. from https://www.investopedia.com/terms/m/mlr.asp

Kicsiny. (2014). Multiple linear regression based model for solar collectors. *Solar Energy, 110*, 496-506.

Kıran, Özceylan, Gündüz, & Paksoy. (2012). A novel hybrid approach based on particle swarm optimization and ant colony algorithm to forecast energy demand of Turkey. *Energy Conversion and Management, 53*(1), 75-83.

Kora, & Krishna. (2016). Hybrid firefly and particle swarm optimization algorithm for the detection of bundle branch block. *International Journal of the Cardiovascular Academy, 2*(1), 44-48.

Krishnamoorthy, Boopathy, Palanikumar, & Davim. (2012). Application of grey fuzzy logic for the optimization of drilling parameters for CFRP composites with multiple performance characteristics. *Measurement, 45*(5), 1286-1296.

Larson, Nonnenmacher, & Coimbra. (2016). Day-ahead forecasting of solar power output from photovoltaic plants in the American Southwest. *Renewable Energy, 91*, 11-20.

Leva, Dolara, Grimaccia, Mussetta, & Ogliari. (2017). Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power. *Mathematics and Computers in Simulation, 131*, 88-100.

Li, & Han. (2008). *Short-term power load forecasting using improved ant colony clustering.* Paper presented at the First International Workshop on Knowledge Discovery and Data Mining (WKDD 2008).

Liao, & Tsao. (2004). Application of fuzzy neural networks and artificial intelligence for load forecasting. *Electric Power Systems Research, 70*(3), 237-244.

Liu, Niu, Wang, & Fan. (2014). Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. *Renewable Energy, 62*, 592-597.

Liu, Tian, Chen, & Li. (2010). A hybrid statistical method to predict wind speed and wind power. *Renewable Energy, 35*(8), 1857-1861.

Lyu, Kantardzic, & Arabmakki. (2014). *Solar irradiance forecasting by using wavelet based denoising.* Paper presented at the Computational Intelligence for Engineering Solutions (CIES), 2014 IEEE Symposium on.

Malik. (2016). *Application of artificial neural network for long term wind speed prediction.* Paper presented at the Advances in Signal Processing (CASP), Conference on.

Mamlook, Badran, & Abdulhadi. (2009). A fuzzy inference model for short-term load forecasting. *Energy Policy, 37*(4), 1239-1248. doi: 10.1016/j.enpol.2008.10.051

Marquez, Pedro, & Coimbra. (2013). Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs. *Solar Energy, 92*, 176-188.

Mathiesen, & Kleissl. (2011). Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States. *Solar Energy, 85*(5), 967-977.

Mathuranathan. (2010). Moving Average Filter (MA filter). from https://www.gaussianwaves.com/2010/11/moving-average-filter-ma-filter-2/

MathWorks. (2019). Introduction to Wavelet Families. from https://www.mathworks.com/help/wavelet/gs/introduction-to-the-wavelet-families.html

Mirrashid. (2014). Earthquake magnitude prediction by adaptive neuro-fuzzy inference system (ANFIS) based on fuzzy C-means algorithm. *Natural hazards, 74*(3), 1577-1593.

Mohammadi, Shamshirband, Tong, Arif, Petković, & Ch. (2015). A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation. *Energy Conversion and Management, 92*, 162-171.

Mohandes. (2012). Modeling global solar radiation using Particle Swarm Optimization (PSO). *Solar Energy, 86*(11), 3137-3145.

Mollaiy-Berneti. (2016). Optimal design of adaptive neuro-fuzzy inference system using genetic algorithm for electricity demand forecasting in Iranian industry. *Soft Computing, 20*(12), 4897-4906.

Nandi, Rahman, & Riadh. (2016). *A comparative study on ANN techniques in predicting solar radiation for various meteorological locations of Bangladesh.* Paper presented at the Informatics, Electronics and Vision (ICIEV), 2016 5th International Conference on.

Nauck, Klawonn, & Kruse. (1997). *Foundations of neuro-fuzzy systems*: John Wiley & Sons, Inc.

Negnevitsky, & Intelligence. (2005). A guide to intelligent systems. *Artificial Intelligence, 2nd edition, pearson Education*.

Ngui, Leong, Hee, & Abdelrhman. (2013). *Wavelet analysis: mother wavelet selection methods.* Paper presented at the Applied mechanics and materials.

Niknam, Narimani, & Jabbari. (2013). Dynamic optimal power flow using hybrid particle swarm optimization and simulated annealing. *International Transactions on Electrical Energy Systems, 23*(7), 975-1001.

Oentaryo. (2005). Automated driving based on self-organizing GenSoYager neuro-fuzzy system. *Centre for Computat Intelligence, School of Computer Engineering, Nanyang University, Singapore, Technical Report C2i-TR-002/05*, 533-536.

Olatomiwa, Mekhilef, Shamshirband, Mohammadi, Petković, & Sudheer. (2015). A support vector machine–firefly algorithm-based model for global solar radiation prediction. *Solar Energy, 115*, 632-644.

Ozkan, Kisi, & Akay. (2011). Neural networks with artificial bee colony algorithm for modeling daily reference evapotranspiration. *Irrigation Science, 29*(6), 431-441.

Pei, Huayu, Zheqi, & Meibo. (2019). *A Novel Hybrid Firefly Algorithm for Global Optimization.* Paper presented at the 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS).

Pelland, Galanis, & Kallos. (2013). Solar and photovoltaic forecasting through post-processing of the Global Environmental Multiscale numerical weather prediction model. *Progress in Photovoltaics: Research and Applications, 21*(3), 284-296.

Pelland, Remund, Kleissl, Oozeki, & De Brabandere. (2013). Photovoltaic and solar forecasting: state of the art. *IEA PVPS, Task, 14*, 1-36.

Peng, Yu, Huang, Heiser, Yoo, & Kalb. (2015). 3D cloud detection and tracking system for solar forecast using multiple sky imagers. *Solar Energy, 118*, 496-519.

Persson, Bacher, Shiga, & Madsen. (2017). Multi-site solar power forecasting using gradient boosted regression trees. *Solar Energy, 150*, 423-436.

Piri, Shamshirband, Petković, Tong, & ur Rehman. (2015). Prediction of the solar radiation on the Earth using support vector regression technique. *Infrared Physics & Technology, 68*, 179-185.

Platt, Li, Li, Poulton, James, & Wall. (2010). Adaptive HVAC zone modeling for sustainable buildings. *Energy and Buildings, 42*(4), 412-421.

Premalatha, & Natarajan. (2009). Hybrid PSO and GA for global maximization. *Int. J. Open Problems Compt. Math, 2*(4), 597-608.

Premalatha, & Valan Arasu. (2016). Prediction of solar radiation for solar systems by using ANN models with different back propagation algorithms. *Journal of applied research and technology, 14*(3), 206-214.

Rabbi, Nandi, Saleh, Faisal, & Mojumder. (2016). *Prediction of solar irradiation in Bangladesh using artificial neural network (ANN) and data mapping using GIS technology.* Paper presented at the Development in the in Renewable Energy Technology (ICDRET), 2016 4th International Conference on the.

Rahmani, Yusof, Seyedmahmoudian, & Mekhilef. (2013). Hybrid technique of ant colony and particle swarm optimization for short term wind energy forecasting. *Journal of Wind Engineering and Industrial Aerodynamics, 123*, 163-170.

Raidl, & Puchinger. (2008). Combining (integer) linear programming techniques and metaheuristics for combinatorial optimization *Hybrid metaheuristics* (pp. 31-62): Springer.

Rajabi, Bohloli, & Ahangar. (2010). Intelligent approaches for prediction of compressional, shear and Stoneley wave velocities from conventional well log data: A case study from the Sarvak carbonate reservoir in the Abadan Plain (Southwestern Iran). *Computers & Geosciences, 36*(5), 647-664.

Rana, Chandra, & Agelidis. (2016). *Cooperative neuro-evolutionary recurrent neural networks for solar power prediction.* Paper presented at the Evolutionary Computation (CEC), 2016 IEEE Congress on.

Ren, An, Wang, Li, Hu, & Shang. (2014). Optimal parameters selection for BP neural network based on particle swarm optimization: A case study of wind speed forecasting. *Knowledge-based systems, 56*, 226-239.

Renani, Elias, & Rahim. (2016). Using data-driven approach for wind power prediction: A comparative study. *Energy Conversion and Management, 118*, 193-203.

Rini, Shamsuddin, & Yuhaniz. (2011). Particle swarm optimization: technique, system and challenges. *International journal of computer applications, 14*(1), 19-26.

Rodríguez, Fleetwood, Galarza, & Fontán. (2018). Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control. *Renewable Energy, 126*, 855-864.

Routh, Yousuf, Hossain, Asasduzzaman, Hossain, Husnaeen, & Mubarak. (2012). *Artificial neural network based temperature prediction and its impact on solar cell.* Paper presented at the 2012 International conference on informatics, electronics & vision (ICIEV).

Rumbayan, & Nagasaka. (2012). Solar irradiation estimation with neural network method using meteorological data in Indonesia. *International journal of technology, 2*, 110-120.

Russell, & Norvig. (2010). Artificial Intelligence-A Modern Approach (3rd internat. edn.): Pearson Education.

Semero, Zhang, & Zheng. (2018). PV power forecasting using an integrated GA-PSO-ANFIS approach and Gaussian process regression based feature selection strategy. *CSEE Journal of Power and Energy Systems, 4*(2), 210-218.

Shamshirband, Petković, Saboohi, Anuar, Inayat, Akib, . . . Gani. (2014). RETRACTED: Wind turbine power coefficient estimation by soft computing methodologies: Comparative study: Elsevier.

Sharie, Mosavi, & Rahemi. (2019). Determination of an appropriate mother wavelet for de-noising of weak GPS correlation signals based on similarity measurements. *Engineering Science and Technology, an International Journal*.

Sharma, Yang, Walsh, & Reindl. (2016). Short term solar irradiance forecasting using a mixed wavelet neural network. *Renewable Energy, 90*, 481-492.

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

Shirkey. (2019). Smoothing Cubic Splines. from https://www.centerspace.net/smoothing-cubic-splines

Singh, & Mohapatra. (2019). Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting. *Renewable Energy, 136*, 758-768.

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

Solutions. (2013). What is Linear Regression? , from https://www.statisticssolutions.com/what-is-linear-regression/

Sperati, Alessandrini, & Delle Monache. (2016). An application of the ECMWF Ensemble Prediction System for short-term solar power forecasting. *Solar Energy, 133*, 437-450.

Sulaiman, Rahman, Musirin, & Shaari. (2012). *Artificial neural network versus linear regression for predicting Grid-Connected Photovoltaic system output.* Paper presented at the 2012 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER).

Tang, Mao, Wang, & Nelms. (2018). Solar power generation forecasting with a LASSO-based approach. *IEEE Internet of Things Journal, 5*(2), 1090-1099.

Thangaraj, Pant, Abraham, & Bouvry. (2011). Particle swarm optimization: hybridization perspectives and experimental illustrations. *Applied mathematics and computation, 217*(12), 5208-5226.

Thirupathaiah. (2018). Enhancement of power quality in wind power distribution system by using hybrid PSO-firefly based DSTATCOM. *International Journal of Renewable Energy Research (IJRER), 8*(2), 1138-1154.

Tian, & Wells. (1998). *Vanishing moments and biorthogonal wavelet systems.* Paper presented at the INSTITUTE OF MATHEMATICS AND ITS APPLICATIONS CONFERENCE SERIES.

Toksarı. (2007). Ant colony optimization approach to estimate energy demand of Turkey. *Energy Policy, 35*(8), 3984-3990.

Torres, Garcia, De Blas, & De Francisco. (2005). Forecast of hourly average wind speed with ARMA models in Navarre (Spain). *Solar Energy, 79*(1), 65-77.

Tutorials Point. (2017). Artificial Intelligence – Neural Networks. from https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_neural_networks.htm

Umbarkar, Balande, & Seth. (2017). *Performance evaluation of firefly algorithm with variation in sorting for non-linear benchmark problems.* Paper presented at the AIP Conference Proceedings.

Vapnik. (1999). An overview of statistical learning theory. *IEEE transactions on neural networks, 10*(5), 988-999.

Vieira, Dias, & Mota. (2004). *Neuro-fuzzy systems: a survey.* Paper presented at the 5th WSEAS NNA international conference on neural networks and applications, Udine, Italia.

Vlădăreanu, Căpitanu, & Vlădăreanu. (2018). Neuro-Fuzzy Modelling of the Metallic Surface Characterization on Linear Dry Contact between Plastic Material Reinforced with SGF and Alloyed Steel. *Materials, 11*(7), 1181.

Walker. (1997). Fourier analysis and wavelet analysis. *Notices of the AMS, 44*(6), 658-670.

Wang, Li, Ran, Che, & Zhou. (2018). A short-term photovoltaic power prediction model based on the Gradient Boost Decision Tree. *Applied Sciences, 8*(5), 689.

Wang, Mi, Su, & Zhao. (2012). Short-term solar irradiance forecasting model based on artificial neural network using statistical feature parameters. *Energies, 5*(5), 1355-1370.

Wang, Ran, & Zhou. (2017). A Short-Term Photovoltaic Power Prediction Model Based on an FOS-ELM Algorithm. *Applied Sciences, 7*(4), 423.

Wang, Zhen, Mi, Sun, Su, & Yang. (2015). Solar irradiance feature extraction and support vector machines based weather status pattern recognition model for short-term photovoltaic power forecasting. *Energy and Buildings, 86*, 427-438.

Wuest, Weimer, Irgens, & Thoben. (2016). Machine learning in manufacturing: advantages, challenges, and applications. *Production & Manufacturing Research, 4*(1), 23-45.
Yan, & Li. (2011). An effective refinement artificial bee colony optimization algorithm based on chaotic search and application for pid control tuning. *Journal of Computational Information Systems, 7*(9), 3309-3316.
Yang, Jirutitijaroen, & Walsh. (2012). Hourly solar irradiance time series forecasting using cloud cover index. *Solar Energy, 86*(12), 3531-3543.
Yona, Senjyu, Saber, Funabashi, Sekine, & Kim. (2007). *Application of neural network to one-day-ahead 24 hours generating power forecasting for photovoltaic system.* Paper presented at the Intelligent Systems Applications to Power Systems, 2007. ISAP 2007. International Conference on.
Zadeh. (1975). The concept of a linguistic variable and its application to approximate reasoning—II. *Information sciences, 8*(4), 301-357.
Zendehboudi, Baseer, & Saidur. (2018). Application of support vector machine models for forecasting solar and wind energy resources: A review. *Journal of cleaner production, 199*, 272-285.

LIST OF PUBLICATIONS

ISI Publication (Accepted and Published)

1. Abdullah, N.A., Abd Rahim, N., Gan, C.K., & Nor Adzman, N. (2019). Forecasting Solar Power using Hybrid Firefly and Particle Swarm Optimisation (HFPSO) for Optimising the Parameters in Wavelet Transform-Adaptive Neuro Fuzzy Inference System (WT-ANFIS). *Applied Sciences*, 9(16), 3214.

Conference

2. Abdullah, N.A., Koohi-Kamali, S., & Rahim, N.A. (2018). Forecasting of solar radiation in Malaysia using the Artificial Neural Network and Wavelet Transform. 5th IET Clean Energy and Technology Conference 2018 (CEAT 2018), 5-6 Sept. 2018, Pullman Hotel, Kuala Lumpur