

Faculty of Science and Engineering
Discipline of Electrical and Computer Engineering

**Short-Term Prediction of the Output Power of a Group of
Neighbouring PV Plants**

Hasanain A. H. Al-Hilfi

This thesis is presented for the Degree of
Doctor of Philosophy
in
Electric Engineering
of
Curtin University
Perth, Australia

July 2020

Declaration

I declare that this thesis is my own account of my research and contains as its main content work that has not previously been submitted for a degree at any tertiary education institution.

Hasanain A. H. Al-Hilfi

July 2020

Signature.....

Date: 22/07/2020

Dedication

The philosopher and the mentor of my life

MY MOTHER and MY FATHER'S SOUL

Abstract

The penetration of photovoltaic (PV) systems in existing power grids is increasing as they are considered attractive options for electricity generation in distribution networks. Due to the fact that they have a long productive life, operate silently, and are clean systems as well as the fact that they are becoming increasingly less expensive, they have become more competitive power generators compared to other generators (e.g. fossil fuel). However, a major operational challenge that exists for PV systems is the variability of their output power, which is mainly caused by the movements of clouds. This variability can be reduced through building several smaller PV systems distributed within an area, rather than a single, large PV system, often referred to as geographic smoothing. Geographic smoothing is defined as the smoothing that occurs in the output power variability of a particular PV plant during clouds events if the plant is distributed over a specific geographical area instead of being installed as a single block. This smoothing can be expressed mathematically using the variability reduction index (VRI), which denotes the reduced level of variations in the output power of a group of neighbouring PV systems. VRI considers the size of the PV systems, the distances between them, as well as the mathematical correlation coefficient that reflects the timescale, along with the speed, direction and density of the clouds passing over the PV systems. The VRI and solar irradiance, measured at one location, can be used to estimate the power generated by a group of neighbouring PV systems.

This thesis initially compares and assesses various VRI models through their application on a group of 16 small-scale neighbouring rooftop PV systems, distributed over a square kilometre area in Brisbane, Australia. Several approaches are then proposed to improve the estimation accuracy of the overall generated power by the group of neighbouring small-scale rooftop PV systems. This thesis reports on the improvement of several techniques for estimating the output power of an overall neighbourhood rooftop PV system distributed within an area of 0.7 km^2 using one pyranometer located close to the PV plant with a resolution of 1 minute.

First, considering the differences in the output results of existing techniques, a new VRI technique is designed by considering the maximum values of the different existing techniques at different timescales. Then, using the Wavelet Transform (WT) method and this new model, the output power of overall neighbouring rooftop PV systems has been estimated. The impact of the location of the pyranometer on estimation accuracy has also been evaluated.

Second, the Wavelet transform technique along with an Adaptive Neuro-Fuzzy Inference System are used to develop new models, denoted as the ANFIS-WT model in the course of this thesis, to estimate the PV system output power during cloud events. While the third and fourth models are developed by employing the Gene Expression Programming technique. The third model, referred to hereinafter as GEP-VRI, is used to correlate the distribution of the PV system and the irradiance measured by the pyranometer to calculate the VRI value and then estimate the total power generated by the PV systems using the GEP-VRI model and Wavelet Transform technique. The fourth model, referred to hereinafter as GEP-WT, uses the Gene Expression Programming technique directly without using the VRI model to estimate the output power of a group of neighbouring small-scale rooftop PV systems. The effective performance of each model proposed here is validated using real data collected by the Solar Project at the University of Queensland, Brisbane, Australia. The results reveal that the proposed technique using the Gene Expression Programming technique directly without using VRI is more accurate compared with existing approaches in the literature. In addition, a sensitivity analysis has been carried out to investigate the impact of the employed parameters in the new model.

Statement of Contributions to the publication

This thesis contains several papers published in several journals and international conferences as listed below.

- 1- “An improved technique to estimate the total generated power by neighbouring photovoltaic systems using single-point irradiance measurement and correlational models”

This paper has been published in the *IEEE Transactions on Industrial Informatics Journal*. Chapter three of the thesis is the result of this paper.

Key contribution from the candidate: creating the new idea for the publication, developing the theory and calculations, programming and formation, gathering the data, running the main programming, developing the case studies, analysis and demonstration of the results and formulating the manuscript.

Key contribution from co-authors: guiding, supporting and encouraging the idea, supporting with the investigation of the reasonableness of the idea, revising the results, studying and modifying the academic language and formatting.

- 2- “An ANFIS-Wavelet Technique to improve the Estimation Accuracy of the Output Power of Neighbouring PVs during Cloud Events”

This paper has been published in the *Energies Journal*. Chapter four of this thesis represents the outcome of this paper.

The candidate’s main contributions: creating the new idea for the publication, developing the theoretical basis for the simulation, performing the field tests, gathering the data, running the main programming, developing the case studies and the analysis and demonstrating the results and formulating the manuscript.

The key contribution from the co-authors: guiding, supporting and encouraging the idea, supporting the investigation of the reasonableness of the idea, revising the results, studying and modifying the academic language and formatting.

3- “A Gene Expression Technique-based Approach to Improve the Accuracy of Estimating the Total Generated Power by Neighbouring Photovoltaic Systems”

This manuscript has been submitted to IET Renewable Power Generation, Manuscript ID: RPG-2020-0279, 09/03/2020, round-2. The first part of the fifth chapter of this thesis presents the outcome of this article.

The candidate’s main contributions: creating the new idea for the publication, developing the theoretical basis for the simulation, performing the field tests, gathering the data, running the main programming, developing the case studies and analysis, demonstrating the results and formulating the manuscript.

Key contribution from the co-authors: guiding, supporting and encouraging the idea, supporting with the investigation of reasonableness of the idea, revising the results, studying and modifying the academic language and formatting.

4- “Estimating PVs Generated Power during Cloudy days using Gene Expression Programming”

This manuscript has been submitted to the IEEE Journal of Photovoltaics, Manuscript ID: JPV-2020-03-0112-R, 09/03/2020., round-2. The second part of the fifth chapter of this thesis presents the results of this article.

The candidate’s main contributions: creating the new idea for the publication, developing the theoretical basis for the simulation, performing the field tests, gathering the data, running the main programming, developing the case studies and analysis, demonstrating the results and formulating the manuscript.

The key contribution from the co-authors: guiding, supporting and encouraging the idea, supporting the investigation of the reasonableness of the idea, revising the results, studying and modifying the academic language and formatting.

Candidate Signature:  Supervisor Signature:

Date: 22/07/2020

Date:

Acknowledgments

To start with, I must thank Allah, the inspiring and true supporter in all areas of life. Then, I would like to express my deep thanks to Dr. Ahmed Abu-Siada for his perfect supervision and guidance during my PhD study. I found him by my side in every difficult period that I faced during the PhD. I have learned a lot from his personality and knowledge. He has always been helpful, friendly and ready to support me with his instructions and encouragement. I would also like to thank my Iraqi supervisor from my Master's study Abdul-Kareem Khamis Hassan for his great support even after receiving my Master's degree.

My great thanks must be introduced to the Iraqi ministry of higher education to award me a PhD scholarship and financial support (tuition fees, living cost, health insurance, and others).

I would like to offer my special thanks to everyone at the Electrical Engineering Department, who provided me with great support.

Finally, I would like to thank my wife, kids, brothers, sisters and friends for their continuous support and encouragement.

Table of Contents

Declaration	i
Dedication	ii
Abstract	iii
Acknowledgments.....	vii
Table of Contents	viii
List of Figures:	xi
List of Tables:	xiii
List of Symbols and Abbreviations.....	xv
Chapter 1 Introduction	1
1.1. Background.....	1
1.2. Research Problem and Motivation	2
1.3. Research Objectives	5
1.4. Outline of Research Contributions	6
1.5. Organisation of the Thesis.....	7
Chapter 2 Existing Methods for Estimating PV Output Power	8
2.1. Output Power Variability	8
2.2. Correlational Coefficients.....	18
2.3. Variability Reduction Index	21
2.4. Summary.....	22
Chapter 3 Proposed Correlational Model and Wavelet-Based Technique	24
3.1 Wavelet Transform	24
3.2 Proposed Methodology.....	26
3.3 Performance Evaluation	32

3.4	Summary.....	46
Chapter 4	ANFIS-Based Approach	48
4.1	ANFIS.....	48
4.2	Proposed Methodology.....	52
4.3	Performance Evaluation of the proposed technique	55
4.3.1	Sensitivity Analysis	63
4.4	Summary.....	64
Chapter 5	GEP-Based Approach	66
5.1.	GEP Technique.....	66
5.2.	GEP-VRI Model.....	73
5.2.1.	Proposed Methodology.....	73
5.2.2.	Performance Evaluation	76
5.2.3.	Sensitivity Analysis	80
5.3.	GEP-WT Model.....	82
5.3.1.	Proposed Methodology.....	82
5.3.2.	Performance Evaluation	85
5.3.3.	Sensitivity Analysis	106
5.4.	Comparison of the Proposed Models vs Existing Models in the Literature	
	108	
5.5.	Summary.....	111
Chapter 6	Conclusion and Recommendations	112
6.1.	Conclusions	112
6.2.	Recommendations for future work	114
References	116	

Publications Arising from this Thesis123

List of Figures:

Fig. 1.1 Solar irradiance exposed to PV array at the St Lucia campus of the University of Queensland on (a) a clear-sky day (01/01/2016).	3
Fig. 1.2 Solar irradiance exposed to a PV array at the St Lucia campus of the University of Queensland on a partly cloudy day (22/03/2016).	3
Fig. 1.3 Comparison of the variability in irradiance for one PV module, and a group of 5 and 23 PV systems at the Southern Great Plains network, USA	5
Fig. 3.1 b Sample illustration of the performance of a Wavelet transform	25
Fig. 3.2 Sample illustration of various Wavelet mother functions	25
Fig. 3.3 Flowchart of the proposed technique for estimating the overall power of a group of neighbouring PV systems	32
Fig. 3.4 The site of the study indicating the locations of the PV systems	34
Fig. 3.5 Approximated cloud speed for eight days in different seasons: a) calculated correlation coefficient on 04/07/2016, b) the fitted results for seven sample days	36
Fig. 3.6 The power generated by the PV module at PV site-10, as well as the average power generated by 7 and 16 rooftop systems across the studied site	38
Fig. 3.7 Average VRI calculated using the Hoff, Perez, Lave and ACM models for the considered site over 312 days at various cloud speeds and timescales.	40
Fig. 3.8 Comparison of the error between the power generated and estimated using different VRI models, calculated under (a) MAE, and (b) RMSE	42
Fig. 3.9 (a) VRI defined using the studied correlational models at each mode, (b) The difference between the anticipated overall power generated by each correlational model and the measured power	44
Fig. 3.10 Error and MAE of the calculated power using (11) and the power measured by the wattmeter for the PV array located next to the pyranometer over 6 sample days	45
Fig. 4.1 ANFIS architectonics	51
Fig. 4.2. The proposed methodology for the proposed ANFIS-WT technique	55
Fig. 4.3 Layout of the six PV plants under study	57
Fig. 4.4 Resulted in errors of different membership functions	57
Fig. 4.5 Errors after 100 ANFIS epochs for each sub-time series	58
Fig. 4.6 Comparison of the MAE and RMSE of the proposed ANFIS-WT model and other models in the literature	59
Fig. 4.7 V_{pi} for an individual PV system (blue), entire actual PV systems (black) and predicted (red).	62
Fig. 4.8 The actual (black) and predicted (red) VRI values, 3 Dec 2017	63
Fig. 4.9 . Sensitivity analysis for each model parameter for each sub-time series	64
Fig. 5.1 Flowchart of the GEP algorithm	67
Fig. 5.2. Flowchart for the calculation of $VRI_{mode} - iGEP$ using GEP and estimating the overall power of a small-scale PV plant	76
Fig. 5.3. MAE and RMSE for different techniques estimating the total power generated by the considered PV systems	78
Fig. 5.4. Evaluation of modelling stages of the $VRI_{mode} - iGEP$ model	80
Fig. 5.5. Sensitivity of the $VRI_{mode} - iGEP$ model for each model parameter	81
Fig. 5.6. GEP modelling methodology and estimating the overall output power of a group of rooftop PV systems distributed within 0.7 km ²	84
Fig. 5.7. The RMS errors for the estimated generated power during the GEP training process for various Wavelet modes	86
Fig. 5.8. Average MAE error for Table 5.14	92
Fig. 5.9. Average MAE error for Table 5.15	94
Fig. 5.10. Average MAE error for Table 5.16	95

Fig. 5.11. Average MAE error for Table 5.17.....	97
Fig. 5.12. Average MAE error for Table 5.18.....	98
Fig. 5.13. Average MAE error for Table 5.19.....	100
Fig. 5.14. Average MAE error for Table 5.20.....	101
Fig. 5.15. Average MAE error for Table 5.21.....	103
Fig. 5.16. Average MAE error for Table 5.22.....	104
Fig. 5.17. Error comparison between the estimated and the actual power using various models published in the literature and the model proposed in this chapter (GEP-WT)	105
Fig. 5.18. Sensitivity analysis of the estimated model parameters	108
Fig. 5.19. The maximum and minimum improvements of the GEP-WT model compared with other models using a) MAE and b) RMSE.....	110

List of Tables:

Table 2.1. VI and VS methods published in the literature	10
Table 2.2. Summary of the Discrete Fourier Transform methods used to estimate and analyse output variability	12
Table 2.3. Summary of Wavelet transform methods to estimate and analyse variability	15
Table 2.4. Summary of different methods to estimated and analyse output variability	17
Table 2.5 Comparison of the correlational coefficient models available in the literature	21
Table 3.1 The definition and method for calculating or retrieving the parameters, employed in (3.4) and (3.6).....	29
Table 3.2 Considered Wavelet modes and corresponding timescales in this chapter	30
Table 3.3 Considered distributed rooftop PV systems at the studied site	33
Table 3.4 Dominant wind direction and range of ambient temperature at the study site for 2016–2017.	35
Table 3.5 The correlational coefficient models that yield the largest average VRI for the considered site during the period of the study	41
Table 3.6 Comparison of the MAE and RMSE for differenet correlation models	43
Table 3.7 Percentage of days in which the proposed method has an accuracy of over 90% when the pyranometer is located at various PV sites.	46
Table 4.1 ANFIS input and output data and models after the training stage.....	54
Table 4.2 Comparison of MAE [%] average values for different VRI models including the proposed model in this chapter	60
Table 4.3 Comparison of RMSE [%] average values for different VRI models including the proposed model in this chapter	61
Table 4.4. Minimum and maximum values for each model parameter.....	64
Table 5.1. Creating an Initial Solution using a GEP algorithm	68
Table 5.2. Second stage: Selecting the best solutions.....	69
Table 5.3. Modification pattern: Replication, Mutation and Inversion operators and some work details and examples	70
Table 5.4 Modification Pattern: Transposition Operators	71
Table 5.5 Modification Pattern: Recombination Operators	72
Table 5.6 Data for the GEP model.....	77
Table 5.7 The improvement using the proposed technique over other models.....	79
Table 5.8. Various stages of the GEP model training.....	80
Table 5.9 GEP input and output variables as well as the resulting GEP models	85
Table 5.10 Developed <i>PHF – mode – 1GEPk</i> wavelet model.....	87
Table 5.11 Developed <i>PHF – mode – 2GEPk</i> wavelet model.....	88
Table 5.12 Developed <i>PHF – mode – 3GEPk</i> wavelet model.....	89
Table 5.13 Developed <i>PLF – mode – 3GEPk</i> Wavelet model	90
Table 5.14 Comparison of the proposed model and other published models using 10 samples for the first sub-time series ranging from 8:01 to 8:10, 1 January 2018.....	91
Table 5.15 Comparison of the proposed model and other published models using 10 samples for the second sub-time series ranging from 12:01 to 12:10, 1 January 2018.....	93
Table 5.16 Comparison of the proposed model and other published models using 10 samples for the third sub-time series ranging from 15:01 to 15:10, 1 January 2018	94
Table 5.17 Comparison of the proposed model and other published models using 10 samples for the first sub-time series ranging from 9:01 to 9:10, 17 Sept. 2017	95
Table 5.18 Comparison of the proposed model and other published models using 10 samples for the second sub-time series ranging from 12:01 to 12:10, 17 Sept. 2017	97
Table 5.19 Comparison of the proposed model and other published models using 10 samples for the third sub-time series ranging from 14:31 to 14:40, 17 Sept. 2017.....	99

Table 5.20 Comparison of the proposed model and other published models using 10 samples for the first sub-time series ranging from 10:01 to 10:10, 18 March 2017	100
Table 5.21 Comparison of the proposed model and other published models using 10 samples for the second sub-time series ranging from 13:51 to 14:00, 18 March 2017	102
Table 5.22 Comparison of the proposed model and other published models using 10 samples for the second sub-time series ranging from 16:01 to 16:10, 18 March 2017	103
Table 5.23 Comparison of the minimum and maximum errors for the main VRI models published in the literature and the proposed GEP-WT model	106
Table 5.24 Maximum and minimum values of the estimated model used in the sensitivity analysis	107
Table 5.25 Comparison of the minimum and maximum average errors calculated by each model (existing and proposed models) in this study	109

List of Symbols and Abbreviations

Symbol	Description
a	A function of T
a_1 to a_4	Constants defined by fitting the measured irradiance data in a specific location
avg	Average value
A_1 to A_4	Normalized values and are calculated considering the distance between the PV systems, as well as the cloud speed, size and direction
ACM	Anisotropic Correlation Model
A_m (m ²)	Area of the PV module
A_d^{pv}	PV plant area
ANFIS	Adaptive Network-Based Fuzzy Inference System
ANN	Artificial Neural Network
b	Constant
B_1 to B_9	Functions of distance and timescale, as well as the cloud speed and direction
c	Constant
$c_j, l_j,$ and m_j	MF parameters
$\cos(.)$	Cosine function
$covar ()$	The covariance functions
D_A^{cap}	P_d^{cap} / A_d^{pv}

$DHI_{clr}(W/m^2)$	Clear Sky Horizontal Diffuse Irradiance on Ground Level
DNI_{clr}	Clear Sky Direct Normal Irradiance
$exp(.)$	The Exponential Function
ET	Expression Tree
f	Shift Integer
g	Shift Factor
GA	Genetic Algorithm
GEP	Gene Expression Programming
$GHI(W/m^2)$	Global Horizontal Irradiance
GP	Genetic Programming
GR	Gene Recombination
$GHI_{clr}(W/m^2)$	Clear Sky Global Horizontal Irradiance
H_u	Humidity
i	An integer number that shows the modes of the discrete
$inv(.)$	Inverse Function
IO	Inversion Operator
IS	Insertion Sequence
I_w	Wavelet Periodogram
k (minute)	Discrete Time Index
$L_{m,n}$ (m)	Distance between m and n sites
$\ln()$	Natural Logarithmic Function
$m[k]$	The time-varying variables at location m
MAE	Mean Absolute Error

$\max(\cdot)$	Maximum Function
MF	Membership Function
M_j	Fuzzy set
MO	Mutation Operator
m-Si	Mono-Crystalline Silicon
$n(\cdot)$	Natural Logarithmic Function
$n[k]$	The time-varying variables at location n
N_{data}	Number of the power data set
N_j	Fuzzy set
N_{module}	Total number of PV modules
N_{system}	The number of PV systems
$O_{1,j}$	Outputs of the first layer of ANFIS
$O_{2,j}$	Outputs of the second layer of ANFIS
$O_{3,j}$	Outputs of the third layer of ANFIS
$O_{4,j}$	Outputs of the forth layer of ANFIS
$O_{5,j}$	Outputs of the fifth layer of ANFIS
OPR	One Point Recombination
P_{act}	Real (measured) total power generated by all PV systems
P_d^{cap}	PV system sites power capacity, $d = 1, 2, 3, \dots$
$P_{\text{Eq}}^{\text{mdu-out}}$	Average of the total output power of the PV plant
$P_{\text{HF-mode-}i}^{\text{GEP}}$	GEP-WT High-Frequency Modes
$P_{\text{LF-mode-}3}^{\text{GEP}}$	Low-Frequency Mode-3 of GEP-WT Model.

P_{mode-i}^{HF-in}	Wavelet High-Frequency Levels for P_{nrm}^{POA-in}
$P_{module}^{POA^{eq}}$ (W)	Approximate average output power of the PV modules
P_{module}^{POA} (W)	Output power of a PV module
$P_{mode-3}^{LF-ANFIS}$	DWT-ANFIS Low-Frequency Levels
$P_{mode-i}^{HF-ANFIS}$	DWT-ANFIS High-Frequency Levels
P_{mode-i}^{HF-out}	Wavelet High-Frequency Levels for $P_{nrm}^{POA-out}$
P_{mode-i}^{LF-in}	Wavelet Low-Frequency Levels for P_{nrm}^{POA-in}
P_{mode-i}^{LF-out}	Wavelet Low-Frequency Levels for $P_{nrm}^{POA-out}$
$P_{nr-out}^{Eq-ANFIS}$	Estimation of equivalent normalized output power (entire PVs)
$P_{nrm-out}^{Eq-GEP}$	Normalised GEP-WT model
P_{nrm}^{POA-in}	Normalised P_{POA}^{mdu-in}
$P_{nrm}^{POA-out}$	Normalised $P_{Eq}^{mdu-out}$
$POA(W/m^2)$	The Irradiance Incident on The Plane of The Array
$POA_{clr}(W/m^2)$	Clear Sky POA
$POA_{clr}^g(W/m^2)$	Ground-Reflected Diffuse Irradiances Incident on the PV Array
$POA_{clr}^s(W/m^2)$	Sky Diffuse Irradiances Incident on the PV Array
POA_{nrm}	Normalized POA
POA_{nrm}^{eq}	Equivalent Normalized POA
$POA_{clr}^d(W/m^2)$	Direct Irradiances Incident on the PV Array
POA^{eq} (W)	Equivalent POA

$P_{out}^{Eq-ANFIS}$	Equivalent Output Power seen by the entire PV plant
P_{out}^{Eq-GEP}	The Equivalent Generated Power of GEP-WT model
$P_{out}^{mdu-GEP}$	Module Output Power of GEP-WT model
P_{POA}^{mdu-in}	Pyranometer Power Equivalent to Module Power
P_{total}	Determined Power by GEP Technique
P_{POA}^{clr}	Output Power in a Clear Sky Condition
PSD	Power Spectral Density
p-Si	Polycrystalline Silicon
PV	Photovoltaic
R^2	R-square is a statistical quantity referred to how the data close to the fitted regression
RIS	Root IS
RMSE	Root Mean Square Error
RO	Replication Operator
SAM	System Advisor Model
SD	Standard Deviation
$\sin(\cdot)$	Sine Function
$\text{std}(\)$	Standard Deviation Functions
t (minute)	Time Function
T (minute)	Sampling Frequency of Solar Irradiances (timescale)
T_L	Length of the Time-Series Signal
Tmp (Celsius)	Ambient Temperature
T_{mode-i} (minute)	Time Scale

TPR	Two-Point Recombination
u	Scale Factor
V_c	Cloud Speed
VI	Variability Index
V_{pi}	Variability Power Index
$V_{pi}(T^{\text{mode}-j})_{\text{ave}}$	V_{pi} for the overall PV system
$V_{pi}(T^{\text{mode}-j})_{\text{py}}$	V_{pi} for the pyranometer (or individual PV system)
VRI	Variability Reduction Index
VS	Variability Score
V_w (m/s)	Wind Speed
W_j	Weight Vector
$\overline{W_j}$	Average of the rules firing strengths
WT	Wavelet Transform
WT_{cont}	The Wavelet Transform of a Continuous-Time Signal
WT_{disc}	Discrete WT
	Continuous-Time Signal
x_o	Input Variables
y_o	Input Variables
z	Could be Pearson, Hoff, Lave, ACM or Lonij model
$(\rho_{m,n}^z)^{\text{eq}}$	The equivalent of the overall PV systems correlation coefficient
θ_T	Tilt Angle of The PV Array
θ_Z	Zenith Angle of The PV Location

θ_A^{array}	Azimuth Angle of The PV Array
θ_A^{solar}	Solar Azimuth Angle
$\mu_{M_{j-2}}$ and μ_{N_j}	They can be any fuzzy MF such as Gaussian and bell-shaped
$\rho_{m,n}^{\text{Pearson}}$	Pearson's Correlation Coefficient
$\rho_{m,n}^{\text{ACM}}$	ACM model
$\rho_{m,n}^{\text{Akinobu}}$	Akinobu model
$\rho_{m,n}^{\text{Hoff}}$	Hoff model
$\rho_{m,n}^{\text{Lave}}$	Lave model
$\rho_{m,n}^{\text{Lonij}}$	Lonij model
$\rho_{m,n}^{\text{Mills}}$	Mills model
$\rho_{m,n}^{\text{Perez}}$	Perez model
π	Fuzzy Operator such as AND or OR Operators
β	Mounting Coefficient
$\psi(t)$	Wavelet Mother Function

Chapter 1 Introduction

This chapter discusses the background and motivation for the research presented in this thesis. The key research objectives and contributions of the thesis are summarised here. Finally, the organisation of the thesis and the key areas of each chapter are introduced.

1.1. Background

The Kyoto agreement raised the issue of global warming caused by greenhouse-gas emissions and urged individual nations to reduce energy generated using fossil fuels through relying more on renewable energy generation, such as PV systems and wind energy [1, 2]. According to [2], the electricity generated in Australia is divided into 79% fossil fuel-based generation, while 21% is generated from renewable energy resources. The use of renewable energy sources continues to grow rapidly in Australia, with plans for it to produce 50% of all national electricity production by 2024 and 100% by 2032 [3]. PV systems are currently very attractive power generation units within the electricity industry; hence, they are becoming more serious competitors for coal and natural gas-based generators [4]. The main source for these systems is solar irradiance. Solar irradiance (W/m^2) is a measure of the solar power (W) arriving within a certain area (m^2). Energy from the sun is exposed to several attenuating factors, such as aerosols, air molecules, water vapour and clouds, during its long journey to the earth's surface after passing through the atmosphere. Therefore, part of this energy is scattered and absorbed by these attenuating factors [5]. There are different kinds of solar irradiance mentioned in the literature. The solar energy that strikes the ground is called a total solar irradiance or global irradiance, which is composed of three kinds of irradiance – direct, diffuse and ground-reflected [6]. Direct irradiance, also called direct normal irradiance, is the beam of solar irradiance that comes directly from the sun and that hits the earth's surface at a perpendicular angle. Diffuse irradiance is the irradiance that is scattered by the attenuating factors, including the clouds. Ground-reflected irradiance is reflected from the ground and other elements like surface albedo material. Moreover, if the global irradiance incidents on a horizontal plane, the irradiance is called global horizontal irradiance. If a pyranometer, which is a PV

sensor that captures the irradiance and weather parameters, such as humidity and temperature, is tilted to the same angle as the PV array and has the same orientation, its output is the irradiance incident on the plane of the array, denoted by POA [7].

Although PV systems have several advantages including easy implementation at both residential and commercial sites along with their low cost and reliable operation, it still comprises some issues owing to its intermittent nature. Due to the ambient conditions, PV generated power is not constant and varies with short (cloud movements) and long-term (daily and seasonal irradiance) variations.

1.2. Research Problem and Motivation

The output power of PV systems is a direct result of the variability of meteorological conditions [8-10]. According to [11-15], at any locality, there is a significant correlation between the solar irradiance, experienced by the PV arrays, and the ambient conditions, such as cloud movements and wind speed. The impact of the size and speed of the cloud on solar irradiance is presented in [11, 12], while the influence of wind speed and direction is discussed in [13-15]. The aforementioned studies have reported extreme variability in solar irradiance on partly cloudy days.

Two types of output power variability are discussed for PV systems in the literature; long-term and short-term variability [16, 17]. The long-term variability refers to the movement of the sun during the daytime that results in the angle of the solar beams changing with respect to the PV panels [17, 18]. For example, the output power of a PV system in the mornings is different from midday, which is directly related to the irradiance exposed to the PV array throughout the day, represented by the parabola of Fig. 1.1 Therefore, it is easy to predict the long-term variability, and it is not a significant problem for electricity network operators [19]. On the other hand, the short-term variability in the output power of PV systems, currently a hot research topic for network operators [20], is related to the ambient wind speed, which changes the temperature of the PV modules [21-23], and the movement of the clouds passing over them. Fig. 1.2 shows typical short-term variability in the solar irradiance exposed to the PV arrays.

While the changes in the output of PV systems due to daily or seasonal variation in the ambient temperature can be easily predicted, fluctuations due to temporary cloud movements are less predictable [24, 25]. The short-term variability is a significant challenge for electricity network operators because of its low / unpredictability, for which providing preventive and improvement solutions is difficult or impossible [8, 19]. This challenge is more significant for large PV systems (especially when connected to weak power distribution networks [26]) than

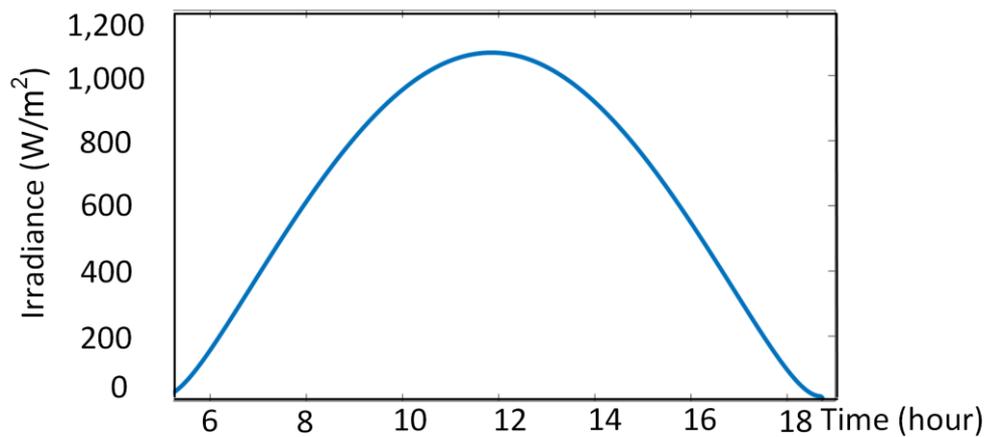


Fig. 1.1 Solar irradiance exposed to PV array at the St Lucia campus of the University of Queensland on a clear-sky day (01/01/2016).

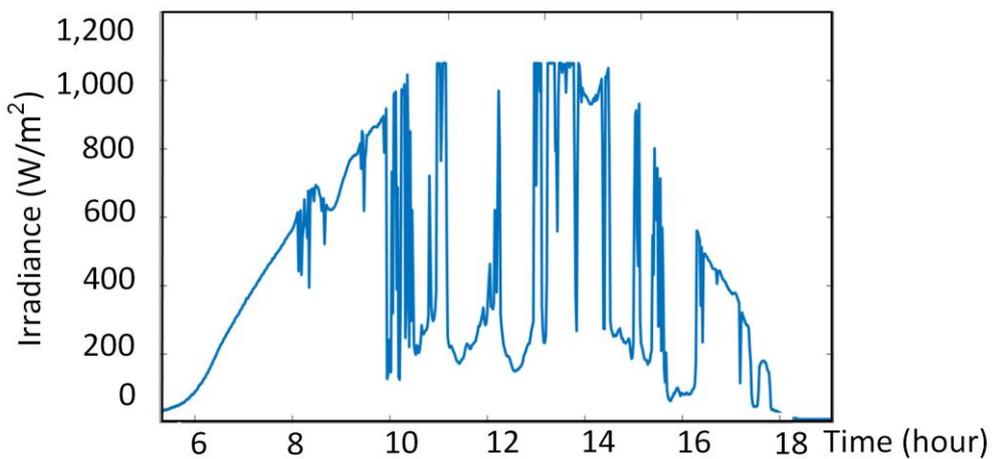


Fig. 1.2 Solar irradiance exposed to a PV array at the St Lucia campus of the University of Queensland on a partly cloudy day (22/03/2016).

small-scale rooftop ones, or when connected to a large interconnected stiff power grid. Such power fluctuations, if not properly regulated, may result in frequency instability [6]. Such utilities have set guidelines to limit the variability of the PV generated power to be less than 10% of the PV system capacity per minute [7]. The precise estimation of the power generated by PV systems during cloudy events facilitates the proper calculation of the required energy storage to mitigate the effect of power reduction and hence avoid further potential consequences. It will also facilitate proper planning for new utility-scale PV plants [8].

One widely accepted solution to overcome the short-term variability issue is employing energy storage systems, such as batteries and flywheels. Stored energy can smooth the output power of PV systems by compensating for the sharp variations in PV output power over a few minutes [17]. Another solution is through building multiple PV systems with smaller capacities spread over a reasonable distance from each other (e.g. a few hundred metres or kilometres), instead of a single large array. As all such systems will not simultaneously observe the same cloud movements, overall, they will have smoothed variability [16, 19, 28]. This mitigation technique is known as geographic smoothing. Geographic smoothing is defined as the smoothing that occurs in the output power variability of a particular PV plant during clouds events if the plant is distributed over a specific geographical area instead of being installed as a single block. As an example, Fig. 1.3 illustrates schematically the average irradiance seen by 5 and 23 PV systems, distributed 20–450km from each other at the Southern Great Plains network, USA [29], versus the irradiance seen by an individual system in the same area. As can be seen from this figure, the average irradiance seen by 5 PV systems has lower variability versus that of a single PV system; as such, the overall power generated by all PV systems (25 PV systems), will also be much smoother than that of a single PV.

From the above discussion, the current models, which presented to estimate the output power of the PV systems during clouds events, have been introduced with different precise estimations. The most accurate was introduced in [28] and displayed errors between 8% and 20%; however, this study only investigated two days. Another study presented a different mathematical method to estimate the PVs generated power with an estimation error range of 4% to 7%. While this model

results in higher estimation accuracy than the models presented in [28], this model is only valid for clouds passing at low speed (less than 0.5 m/s) and was validated using only 13 days [11]. Therefore, this thesis presents four models to estimate the generated power of PV systems during cloudy days by employing different methods that include machine learning techniques. One of the main advantage of machine learning is this technology's capacity to analyse large volumes of data and recognise the relationships and tendencies that might not be noticed by a human, such as ANN and ANFIS techniques. Another technique has also been used in this thesis, which is Gene Expression Programming (GEP), which is a powerful evolutionary algorithm that can be employed to build a mathematical correlation between two or more non-linear variables [20].

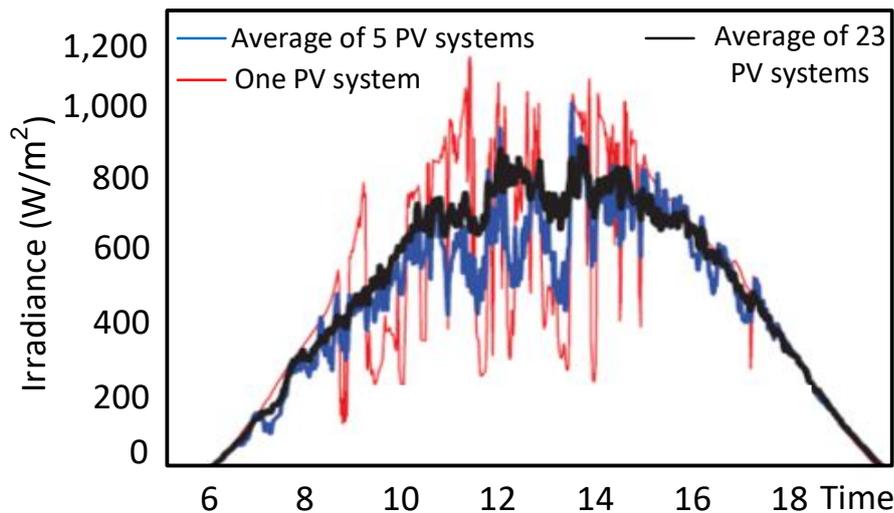


Fig. 1.3 Comparison of the variability in irradiance for one PV module, and a group of 5 and 23 PV systems at the Southern Great Plains network, USA.

1.3. Research Objectives

The main research objective is to develop an accurate technique for estimating the total power generated by a group of small-scale distributed rooftop PV systems when using only one pyranometer. Various approaches have been proposed and developed in this research that employ different types of machine learning tools. The major points can be summarised as

- Studying, analysing and comparing the accuracy of existing techniques in calculating and estimating the overall power generated by a group of neighbouring PV systems
- Considering the maximum value of existing models, a new model has been developed to estimate the total power generated by a group of neighbouring PV systems.
- Employing an artificial neural fuzzy inference system-based approach to develop and propose a new model to estimate the total power generated by a group of neighbouring PV systems.
- Using a Gene Expression Programming-based approach to develop and propose two models:
 - 1- A VRI model to estimate the total generated power of a group of neighbouring PV systems.
 - 2- A power model to estimate directly the total generated power of a group of neighbouring PV systems to further increase the accuracy of the estimation technique.

All the above-proposed models have been improved using one pyranometer and a Wavelet transform model.

1.4. Outline of Research Contributions

The main contributions of this thesis can be summarised as:

- Comparing the accuracy of existing techniques in estimating the total power generated by a group of neighbouring PV systems.
- Proposing and developing a simple but effective technique based on a Wavelet transform model to estimate the total power generated by a group of neighbouring PV systems.
- Proposing and developing an artificial neural fuzzy inference system-based approach to directly estimate the total generated power and increase the accuracy of the estimation technique.
- Proposing and developing a Gene Expression Programming-based approach to estimate the VRI to increase the accuracy of the estimation technique.
- Proposing and developing a Gene Expression Programming-based approach

to directly estimate the total generated power and further increase the accuracy of the estimation technique.

- Including more available weather parameters such as temperature and humidity to enhance the estimation accuracy and reliability of the proposed models.

1.5. Organisation of the Thesis

The remainder of the thesis is organised as follows:

Chapter 2 surveys and critically analyses the relevant literature to establish the research problems and some possible solutions.

Chapter 3 presents a comparison of the accuracy of existing techniques in estimating the total power generated by a group of neighbouring PV systems. As a result of this comparison, presenting a new model to estimate the output power of a group of PV systems using only one PV sensor, and using in-depth investigations to determine the best location for this sensor.

Chapter 4 introduces the developed Wavelet Transform and artificial neural fuzzy inference system-based approach to estimate the total power generated by a group of neighbouring PV systems.

Chapter 5 introduces the developed Wavelet Transform and Gene Expression Programming-based approach to estimate the total power generated by a group of neighbouring PV systems. At the end of this chapter, a comparison between existing models and the proposed models is also presented to illustrate the most accurate model.

Chapter 6 highlights the key research findings and summarises the significant contributions. This chapter also suggests some future directions in this research area.

Chapter 2 Existing Methods for Estimating PV Output Power

This chapter reviews the various existing techniques for estimating the power generated by neighbouring PV systems and highlights their key features and differences. It also provides examples of cases where existing techniques have been employed and tested. At the end of the chapter, research gaps will be highlighted and possible techniques for addressing those research problems suggested.

2.1. Output Power Variability

The main energy source for PV systems is the solar irradiance incident on these systems. The output power variability of PV systems is mainly caused by the variability of the solar irradiance. Many spatial and temporal factors influence solar irradiance variability, such as season, time of day, altitude, latitude, array tilt, and orientation as well as weather conditions. Although all of these factors impact the output power, only weather conditions, such as temperature, humidity, wind speed and size and shape of clouds, have a primary impact on short-term variability.

There are multiple approaches to analysing and estimating the VRI published in the literature, including the Variability Index, System Advisor Model, Power Spectral Density, Fast Fourier Transform, Standard Deviation, Wavelet Transform, and Correlational Models. These techniques are reviewed briefly here while the correlational models are introduced and discussed in detail in the next section. VRI or geographic smoothing is discussed in several studies in the literature; for example, reference [30] introduced a model to determine the variability amount by dividing the length of the global horizontal irradiance by the length of the clear sky irradiance, named the variability index (VI). If the day being tested has a clear sky, VI is equal to one whereas a larger VI value refers to greater variability. For instance, if the value of VI is in the range 2 to 7, this day is considered as a low variability day, whereas it is considered as a medium variability day if the VI's value is in the range 8 to 13. If this value is in the range 14 to 20, the day is considered as a high variability day while it is considered of extremal variability if the value of VI is more than 20 [30]. Employing the VI model, four variability areas

are investigated; clear, overcast, mixed, and high variability. The study considered four days with different amounts of variability at Oak Ridge National Laboratory in the USA within a time resolution ranging from 1 to 10 mins. Similarly, another mathematic model is introduced by [31], which investigates the maximum variability on the output power and is called the Variability Score (VS) by considering a model that adopts the maximum value of the ramp rate multiplied by the probability of the ramp rate and scaled by 100 to score the variability. The model was applied by using 10 PV systems in the USA, with hundreds of kilometres of distance and a 30 sec resolution. The result in [31] stated that VS can be a useful metric to estimate variability, and that a high variability score has a large impact on voltage variability, while a low variability score has less impact. However, VS is not a good predictor for a plane of array irradiance or for tracking PV systems. Both VI and VS techniques were used by [32] to analyse and estimate variability. This study used two PV system locations around 400 km apart in Canada using 41 PV sensors within a 10 ms time resolution. The results of this study were that daily irradiance can be divided into four classes; clear, overcast, low and high frequency solar variability. This study using the VS and VI models reported other results and considered good metrics for investigating and distinguishing variability; however, they are not suitable for investigating the variability with a time resolution greater than 1 sec. Another study used VS methods to classify the variability into zones based on variability frequencies [33]. This study employing two kinds of data, satellite-derived data for almost the entire area of the USA and nine ground measurements of irradiance with time resolution ranging from 1 sec to 30 sec. This study illustrated that the variability zones are divided into several zones in which the zone ranging from 25 to 75 is considered low variability, while the zone labelled 357 to 332 is considered very high variability. Table 2.1 summarizes the VS and VI methods presented above along with the key finding of each study.

Similarly, power spectral density (PSD) and wavelet transform techniques have been employed to analyse, calculate and estimate output power variability. By using the discrete Fourier Transform, [34] analyse the variability of the output power by collecting data from 6 PV systems in Spain with a time resolution of 1 sec. The study introduced an empirical model to estimate the variability in the output power

Table 2.1. VI and VS methods published in the literature

Ref.	Investigated system	Method	Findings
[30]	USA, 3 PV systems locations, temporal analysis study, 1 min to 10 mins	VI	-Introduced a new model called the Variability Index (VI) to determine the variability
[31]	USA, 10 PV systems, hundreds of km and 30 sec	VS	-Investigated the maximum variability on the output power -VS can be a useful method to estimate variability. -VS is not a good predictor for a plane of array irradiance or for tracking PV systems
[32]	Canada, 2 PV systems locations with 41 PV sensors 400 km, 10 mins	VI and VS	-The daily irradiance can be divided into four classes; clear, overcast, low and high frequency solar variability. -The VS and VI models considered good methods for investigating and distinguishing variability; however, they are not suitable for investigating the variability amount with time resolutions of more than 1-sec.
[33]	USA, 6 PV systems 6 to 60 km, 1 to 30 sec	VI and VS	-Using the power distributions function, an analytic model was proposed to analyse and calculate the variability

for a large-scale PV plant. Another study used power spectral density and fluctuation factor methods to analyse the output of a 9 PV system in Japan with 1 min resolution. This study stated that the irradiance on a cloudy day could be reduced to around 20 to 50% [35]. The PSD technique is used in [36] to analyse data with 10 sec and 10 mins time resolutions collected from two PV system sites located in Arizona, USA, which are distributed within 100 to 300 km from each other. The VRI was calculated using the solar irradiance data with resolutions of one and 10 mins. Similar work has been reported in [37], which defines the VRI between several PV systems, distributed within distances of between 20 and 150 km in Colorado, USA, with time resolutions of 1 to 5 mins. In [38], the authors used a signal decomposition hypothesis by introducing a low-pass filter approach to analyse the variability of 18 PV sensors distributed within a radius of 30 km in Nagoya using a Fast Fourier Transform analysing data sampled at 1 min resolution. The study demonstrated that the features of the average fluctuation can be evaluated using a low-pass filter method for different area sizes less than 20 km. In the same way, [39] analysed the PV system data using the PSD method and also employing the Fourier transform technique. In this work, the PV systems were distributed within 50 to 100 km at different locations in Gujarat, India, and using 1 min up to 1 hr time resolutions. The results showed that variability is reduced by 23% and 45% by using 1 hr and 6 hr resolutions respectively, in aggregate 20 PV plants. A new model to simulate the variability of the output power generated by a PV system using the PSD method considering a discrete Fourier model was also adopted in [40]. The model was created and investigated experimentally using 6 PV plants in Spain and 4 PV plants in the USA within an area 1100 km² and 2400 km² respectively with data collected at a 1 sec resolution. The model was built based on the number of PV systems and their size. The probability density function and power spectra analysis were also employed in [41], which used the irradiance collected from five PV sites around the world (Germany, USA, Algeria and Spain) and resolutions of 1 sec to 15 mins and distances of 0.25 to 0.75 km between the systems. They found that variability can be mitigated only when the spatial area is increased; however, the non-Gaussian behaviour in the cumulative power of the overall PV systems location was maintained even with distributed PV systems. Table 2.2 illustrates the summary of these methods and some findings.

Table 2.2. Summary of the Discrete Fourier Transform methods used to estimate and analyse output variability

Ref.	Investigated system	Method	Findings
[34]	-Spain, PV plants, 1 sec	Discrete Fourier Transform	-An empirical model to estimate variability
[35]	-Japan, 9 PV systems, 1 min	power spectral density and fluctuation factor	-The irradiance on a cloudy day could be reduced to around 20 to 50%
[36]	-USA, 2 PV system sites, 100 to 300 km, 10 sec and 10 mins	power spectral density	-Calculate and analyse VRI. The storage system is important to mitigate variability
[37]	-USA, 4 PV system sites, 20–150km, 1 to 5 minutes	power spectral density	-No correlation, significant mitigation at timescale less than 3-hour when the 4 PV systems site were averaged
[38]	-Nagoya, 18 PVs, 30 km, 1 min	Fast Fourier Transform analysing	-Variability can be evaluated using FFT
[39]	-India, 20 PVs sites, 50 to 100 km, 1 min to 1 hr	Fourier Transform analysing	-Variability reduced by 23% and 45% using 1 hr and 6 hr resolutions, respectively using 20 PV plants
[40]	-Spain, USA, 6 PV plants, 4 PV plants, 1100 km ² and 2400 km ² , 1 sec	discrete Fourier model	-A new model to simulate variability
[41]	-Germany, USA, Algeria and Spain, 5 PV sites, 0.25 to 0.75 km, 1 sec to 15 mins	Probability density function and power spectra analysis	-Variability can be mitigated only when the spatial area is increased; however, the non-Gaussian behaviour in the cumulative power of overall PV systems location was maintained even with distributed PV systems

Likewise, the Wavelet transform (WT) technique has been employed in many studies. In [42], the WT technique was used to analyse solar irradiance within 5 sec resolutions in Belgium. The study calculated the fluctuation in the output power using the fluctuation power index. This study focused on storage systems and its results stated that the variability increased in scattered cloud cover compared with clear and overcast sky conditions. The same method is presented in [43]; however, the data were collected from three different locations in Belgium and Australia with time resolutions of 1, 5, 8 sec. This study also focused on storage systems using the fluctuation power index, which is considered a significant tool for the necessary energy storage capacity. The non-stationary process of the output power variability was proved in [44]. The study approached the WT technique to analyse a sample of data with 1 sec time resolution collected from a plant located in Navarra, Spain of 800 PV systems. [56] Moreover, some studies have used a WT technique to investigate and distinguished variability characteristics. For instance, four conditions identified as clear, cirrus, partly cloudy and overcast were examined by implementing 99 PV sensors in Germany, distributed around a 10 to 12 km distance and a 1 sec resolution [46], whereas in [47], the study demonstrated global horizontal irradiance divided into two parts, direct and indirect irradiance, using 1 min data in Thailand. While [48] analysed the output power of one central utility-interactive PV plant in the USA by collecting data at 1 sec resolution to check the differences in the variability at two points; the closest point and the farthest point from the feeder source. The results showed that variability was higher at the end of the feeder source [48]. The WT technique was also used to study and compute geographic smoothing by considering 6 PV systems distributed within a 3 km distance in the USA using 1 sec resolution. The study stated that it detected an annual maximum ramp rate of around 60%, while the largest ramp rate was 40% in a 1 sec time series, whereas 5% ramps per second were never seen in the 10 MW PV plant [49]. Similar work was done by [50], using WT and averaging several PV sensor measurements distributed over a 2 km distance and a 1 sec resolution in the USA. The methods proved that it needs less than 25 PV sensors to obtain the same results obtained from WT methods to estimate PV variability. Similarly, a WT model used in [45] to estimate output power by calculating VRI at different timescales, then analysing the irradiance collected from one pyranometer located

close to the PV system in which the solar irradiance data are taken from two separate PV systems in Japan and the USA. Comparing the methods in [45, 48] with earlier ones, it can be seen that this method needs the solar irradiance data from one pyranometer, while the other techniques call for solar irradiance measurements at every PV system. A summary of the WT methods are presented in Table 2.3

Moving in a different direction, other studies have examined and analysed variability using different techniques and approaches [51-60]. By considering the whole area of Japan, [57] divided the area into eight locations with 1 hr data resolution. Using a set of regression formulae, the variability was estimated and evaluated. The study found that the aggregation of several PV systems led to reducing the PV output variability. Cross-correlation and standard deviation methods have been used in [58]. The study's data were collected from 100 PV systems distributed within 600×750 km in Germany using 5 mins resolution. The study revealed that considerable geographic smoothing has been achieved by aggregating several PV systems compared with observing PV systems separately. Considering the maximum and minimum observed variability values of the limited time windows of irradiance time series and using the power distribution function, an analytical model to analyse and calculate the variability was proposed in [51] by considering 6 PV systems within distances ranging from 6 to 60 km and 1 sec data time rate. A statistical-based approach called a Kriging tool was introduced in [52]. The Kriging tool was used with 1 sec irradiance time resolution and 45 PV sensors spread over 1.6 km in the USA. The achievements of this study indicate that the Kriging tool can give more accurate results using only four PV sensors within a short time ranged between 1 second and 1 min. Ref. [53] analysed irradiance data by employing a model called Multifractal Detrended Fluctuation Analysis, which is a time-series modelling method, especially for extreme events. The study achieved its results by considering several PV system sites around the world – in the USA using 17 PV sensors with 1 sec time resolution and distributed within an area of 0.56 km^2 , using 4 sensors in Spain and Algeria with 1 min time resolution, as well as in Germany with 1 min time resolution and a small PV plant. The study demonstrated that the irradiance time-series data have non-stationary variability. In

Table 2.3. Summary of Wavelet transform methods to estimate and analyse variability

Ref.	Investigated system	Method	Findings
[42]	-Belgium, 5 sec	WT technique	-Variability gets larger in the scattered cloud.
[43]	-Belgium, and Australia, 1, 5, 8 sec	WT technique	-A significant tool for the necessary energy storage system
[44]	-Spain, 800 PV systems, No distance, 1 sec	WT technique	-Variability is a non-stationary process. Classifying the irradiance into several levels
[45]	-Japan and Nevada, 550 PV systems, 1 km ² and 6 km ² , 1 sec	WT technique	-Estimates the output power of two different layout plants
[46]	-Germany, 99 PV systems, 10 to 12 km, 1 sec	WT technique	Four cases, clear, cirrus, partly cloudy and overcast
[47]	-Thailand, one PV sensor, one year data, 1 min	WT technique	-The global horizontal irradiance divided into two parts, direct and indirect irradiance
[48]	-USA, one central utility-interactive PV plant, 60 km, 1 sec.	WT technique	-Differences in the variability the closest point and the farthest point -The results were the variability was higher at the end of the feeder source
[49]	-USA, 6 PV systems, 3 km distance, 1 sec	WT technique	-60% of the ramp rates were detected, the largest ramp rate was 40%.
[50]	-USA, 45 PV systems, 2 km, 1 sec	WT technique	-Requires at less 25 PV sensors to obtain the same results obtained from the WT method.

the other study [54], PV system output power variability is analysed and evaluated using power intensity and ramp-rate methods. The study was conducted on 215 PV

house systems in Canada with 1 to 15 min time average and 1.5 km distance between each other for a centralised PV plant, and 83 km being the largest distance for the distributed PV plant. This study aimed to investigate the differences in the variability between distributed and centralised PV plants. The study confirmed that no significant difference exists in the variability within 15 minutes or greater timescales; however, the variability in the centralised PV system was found to be two times that of the distributed PV systems of equal capacity within timescales of 5 minutes and less. In [55], an investigation proved that the distribution of the short time variability is similar to a Laplacian distribution and is non-Gaussian. The study was implemented using two methods, namely governor free control and load frequency control with 321 PV sensors distributed over the total Japanese territory with a time scale of 1 sec. Because the data used to study and analyse the output power variability is considered a significant factor, ref. [56] introduced a new method to generate synthetic PV system data using real data and a simulation model called the System Advisor Model, which is a free modelling program used in the PV power industry. The study presented three PV system sites in Hong Kong with distances ranging from 14.4 to 17.7 km and 1 min time resolution. The findings of this study demonstrated that the variability in several PV systems that were spatially distributed is less than the variability in a single PV system. An empirical model is presented in [59]. The model is derived using statistical methods to analyse the variability events. In this study, the authors collected their data from a PV system located in India with 1 min resolution. The study stated that the proposed model can be employed to evaluate the frequency of PV ramp events for diverse time intervals. In the same way, [60] used a method called Daily Aggregate Ramp Rate to evaluate, categorise, and compare variability across several PV systems. The data was collected from 6 PV plants located in the USA and Canada with 1 min resolution. The study observed that the maximum ramp rates for 80, 48, 21 and 5 MW are 0.43, 0.53, 0.58 and 0.7 times the PV plants' capacities, respectively. Table 2.4 summarises these methods with some findings.

Table 2.4. Summary of different methods to estimated and analyse output variability

Ref.	Investigated system	Method	Findings
[57]	-Japan, 8 PV plants, 1 hr	-Regression formulae	-aggregation of several PV systems led to reducing the PV output variability
[58]	-Germany, 100 PV systems, 600 × 750 km, 5 mins	-Cross-correlation and standard deviation	-proved the smoothing geographic
[51]	-Spain, 6 PV plants 6 to 60 km, 1 sec	-Power distributions function	-An analytic model was proposed
[52]	-USA, 45 PV sensors 1.6 km, 1 sec	-Kriging tool	-Given more accurate results by using only four PV sensors.
[53]	-USA, Spain, Algeria and Germany, 21 PV, 1 sec and 1 min	-Multifractal Detrended Fluctuation Analysis	-The irradiance is non-stationary variability
[54]	-Canada, 215 PV, 1.5–83 km, 1 to 15 mins	-Power intensity and ramp-rate methods	-The variability in the centralised PV systems was found to be two times that of the distributed PV systems
[55]	-Japan, 321 PV systems, 1 sec	-Governor free control and load frequency control.	-Distribution variability is similar to a Laplacian distribution
[56]	-Hong Kong, 3 PV systems at sites 14.4 to 17.7 km, 1 min	-System Advisor Model (SAM)	-A new method was introduced to generate synthetic PV systems data by using real data
[59]	-India, 1 PV plant, 1 min	-Statistical methods	-The proposed model to evaluate the ramp rate
[60]	-USA and Canada, 6 PV plants, 1 min	-Daily Aggregate Ramp Rate	-The maximum ramp rates for 80, 48, 21,5 MW are 0.43, 0.53, 0.58, 0.7 times the PV plant's capacity, respectively

2.2. Correlational Coefficients

As briefly discussed above, different techniques are proposed in the literature to define different approaches and techniques to analyse and estimate VRI models. VRI can also be defined using a correlation model between the irradiance at two or more PV system sites. Several numerical values that represent the correlation between the solar irradiance at the PV system at different locations were introduced in the literature. For instance, Pearson's correlation coefficient model is a general mathematical technique to measure the linear relationship of the strength and direction of two random variables [61] and has been widely used to analyse statistic data in engineering, biology, finance, and market analyses [62]. It is formulated as

$$\rho_{m,n}^{\text{Pearson}} = \frac{\text{covar}(m[k] \times n[k])}{\text{std}(m[k]) \times \text{std}(n[k])} \quad (2.1)$$

in which $m[k]$ and $n[k]$ are the time-varying variables (i.e. the sampled solar irradiances in the context of this chapter) while $\text{covar}(\cdot)$ and $\text{std}(\cdot)$ are respectively the covariance and standard deviation functions. $\rho_{m,n}^{\text{Pearson}}$ ranges from -1 to $+1$ in which zero shows no correlation between the two solar irradiances while $+1$ and -1 respectively show a positive and negative linear correlation [63]. As seen from (2.1), the Pearson model does not consider the distance between the PV systems, while [64] introduces a correlation model (referred to hereinafter as the Akinobu model) that considers this distance (denoted by $L_{m,n}$), and is given by:

$$\rho_{m,n}^{\text{Akinobu}} = 1 - \exp\left(a - \frac{b \times T}{(L_{m,n})^c}\right) \quad (2.2)$$

in which the sampling frequency of solar irradiances is denoted by its corresponding interval T (in minutes) and $\exp(\cdot)$ denotes the exponential function. In (2.2), a is a function of T , while b and c are constants, independent of T . The value of solar irradiance is not used directly in the Akinobu model. Instead, the maximum variation of the irradiance and its standard deviation, over a moving evaluation window with a period of T , are employed in determining the parameters of (2.2). This method is validated for three sites with PV systems across Japan [64].

Ref. [65] uses the standard deviation to present a new correlation model (referred

to hereinafter as the Mills method), which is given by:

$$\rho_{m,n}^{\text{Mills}} = 0.5 \left[\exp\left(\frac{a_1 \times (L_{m,n})^{a_2}}{T}\right) - \exp\left(\frac{a_3 \times (L_{m,n})^{a_4}}{T}\right) \right] \quad (2.3)$$

where a_1 to a_4 are constants, defined by fitting the measured irradiance data in a specific location. This model is validated for 23 sites in Oklahoma and Kansas, USA, under 1-minute resolutions.

Although the correlational models (2.1) to (2.3) consider the presence of a cloud passing over the pyranometers (which is indirectly included in the measured solar irradiances), neither of them considers the cloud's speed. However, [66] introduces an empirical correlation model (referred to hereinafter as the Hoff model) that considers cloud speed (denoted by V_c), and is given by:

$$\rho_{m,n}^{\text{Hoff}} = \frac{V_c \times T}{L_{m,n} + V_c \times T} \quad (2.4)$$

This empirical model reflects the distance between two PV systems, as well as the corresponding interval, and the speed of the cloud passing over the pyranometers. By employing this model, the solar irradiance measured at three locations across the USA (i.e. Southwest, Southern Great Plains, and Hawaii) were analysed over a sample of 700,000 pairs for pyranometers, located at a distance of 10–250 km from each other, under 1, 2, 3, and 4-hour intervals.

Likewise, [9] also proposed another experimental method (referred to hereinafter as Perez's model), in the form:

$$\rho_{m,n}^{\text{Perez}} = \exp\left(\frac{L_{m,n} \times \ln(0.2)}{1.5 V_c \times T}\right) \quad (2.5)$$

in which $\ln(\cdot)$ is the natural logarithmic function. This model is used to analyse the solar irradiance with one-minute to hourly resolutions, for two locations across the USA while using satellite images to monitor cloud movements.

Similarly, [45] employs another model (referred to hereinafter as the Lave model) to find the correlation between two PV systems, which is given by:

$$\rho_{m,n}^{\text{Lave}} = \exp\left(\frac{-L_{m,n}}{0.5 V_c \times T}\right) \quad (2.6)$$

This model is used to analyse the solar irradiance data measured by 15 pyranometers distributed over 1–12 km around a PV system in Boulder, USA, under 1-second intervals. The Lave model suffers from two limitations: (a) A negative correlation when $L_{m,n} = V_c \times T$ while clouds are moving from one pyranometer to the other, and (b) Zero correlation when $L_{m,n} \gg V_c \times T$ while clouds are moving through two pyranometers without covering both simultaneously.

Neither of the above-mentioned models considers cloud direction, while the model suggested in [11] (referred to hereinafter as the Anisotropic Correlation Model (ACM) model) considers the direction of clouds, and is expressed by:

$$\rho_{m,n}^{\text{ACM}} = \frac{2 \times 0.5^{(2-A_1)} - 0.5^{(2-A_2)} - 0.5^{(2-A_3)}}{2 \times (0.5 - 0.5^{(2-A_4)})} \quad (2.7)$$

where A_1 to A_4 are normalized values and are calculated considering the distance between the PV systems, as well as cloud speed, size and direction. This model is used to analyse the solar irradiance data from 17 pyranometers, distributed within 1km² in Hawaii, USA, under 10 and 60-second resolutions. Similarly, [15] introduces another empirical correlation model (referred to hereinafter as the Lonij model) which considers cloud direction, and is formulated as:

$$\rho_{m,n}^{\text{Lonij}} = \frac{B_1 + B_2 - B_3}{\sqrt{(B_4 + B_5 - B_6) \times (B_7 + B_8 - B_9)}} \quad (2.8)$$

in which B_1 to B_9 are functions of distance and timescale, as well as cloud speed and direction. This model is used to study the correlation of residential rooftop PV systems, distributed over a 2,500 km² area in Tucson, USA, under an interval of 15 mins. Table 2.5 lists these methods, along with a brief summary of the information provided above.

Table 2.5 Comparison of the correlational coefficient models available in the literature

$\rho_{m,n}^z$	Considered Parameters					Model formula
	Cloud					
	$L_{m,n}$	V_c	direction	size	T	
Pearson [61]	✗	✗	✗	✗	✓	$\frac{\text{covar}(m[k] \times n[k])}{\text{std}(m[k]) \times \text{std}(n[k])}$
Akinobu [64]	✓	✗	✗	✗	✗	$1 - \exp\left(a - \frac{b \times T}{(L_{m,n})^c}\right)$
Millsy [65]	✓	✗	✗	✗	✓	$0.5 \left[\exp\left(\frac{a_1 \times (L_{m,n})^{a_2}}{T}\right) - \exp\left(\frac{a_3 \times (L_{m,n})^{a_4}}{T}\right) \right]$
Hoff [66]	✓	✓	✗	✗	✓	$\frac{V_c \times T}{L_{m,n} + V_c \times T}$
Perez [9]	✓	✓	✗	✗	✓	$\exp\left(\frac{L_{m,n} \times \ln(0.2)}{1.5 V_c \times T}\right)$
Lave [45]	✓	✓	✗	✗	✓	$\exp\left(\frac{-L_{m,n}}{0.5 V_c \times T}\right)$
ACM [11]	✓	✓	✓	✓	✓	$\frac{2 \times 0.5^{(2-A_1)} - 0.5^{(2-A_2)} - 0.5^{(2-A_3)}}{2 \times (0.5 - 0.5^{(2-A_4)})}$
Lonij [15]	✓	✓	✓	✗	✓	$\frac{B_1 + B_2 - B_3}{\sqrt{(B_4 + B_5 - B_6) \times (B_7 + B_8 - B_9)}}$

2.3. Variability Reduction Index

It is to be noted that the VRIs, calculated using various correlational models, can be different from each other because different models consider different factors, and some are only valid under specific conditions. Therefore, evaluating the

robustness of various correlation models is a research gap that is addressed in this thesis. Furthermore, the existing correlational models in the above studies only focus on large PV systems, except [45], which has discussed rooftop PV systems of a group of neighbouring households in a suburb or locality. As such, this thesis only focuses on distributed rooftop PV systems and compares the VRIs determined using different correlational models for such systems. This is an important research field as currently, evaluating the short-term variability of the output power of numerous rooftop PV systems within remote towns (which have weaker electrical power distribution networks, compared to those of large cities, and limited smart meters) is a challenge for electricity network operators around the world, including Australia [67].

Geographic smoothing can be expressed mathematically using a coefficient, referring to VRI that shows the variance of the sensor point versus the variance of the entire PV power systems, based upon the correlation of the PV systems with each other [39, 49, 68]. The VRI can be calculated using either of the above-mentioned correlational coefficient models. Given N_{system} neighbouring PV systems, first, the correlation coefficient between every two PV systems has to be defined. Then, a single number, representing the overall (equivalent) correlation between all PV systems (denoted by $(\rho_{m,n}^z)^{\text{eq}}$) can be determined as:

$$(\rho_{m,n}^z)^{\text{eq}} = \sum_{m=1}^{N_{\text{system}}} \sum_{n=1}^{N_{\text{system}}} \rho_{m,n}^z \quad (2.9)$$

using each correlational coefficient model (i.e., $z \in \text{Akinobu, Mills, Hoff, Perez, Lave, ACM, Lonij}$), in which the correlation of one PV system versus itself is unity (i.e., $\rho_{m,m}^z = +1$). Equation (9) is then used to determine the VRI of the output power of all PV systems using each model, as in [66].

$$VRI^z = \frac{N_{\text{system}}^2}{(\rho_{m,n}^z)^{\text{eq}}} \quad (2.10)$$

2.4. Summary

This chapter has discussed and reviewed various approaches and techniques proposed in the literature to analyse and estimate the variability of the output power

such as the Variability Index, Power Spectral Density, Fast Fourier Transform, Wavelet Transform, Standard Deviation, System Advisor Models and Correlational Models. It has considered and identified the most interesting study gaps in the field and suggested several approaches and techniques to address those study gaps. On the basis of this literature review, suitable approaches and techniques will now be developed and expressed in the next chapters, which can calculate and estimate the VRI parameter then estimate the output power of a PV plant of several neighbouring PV systems distributed across 0.7 km^2 and a 1 min resolution. The proposed approaches are more accurate than the models introduced in the literature.

Chapter 3 Proposed Correlational Model and Wavelet-Based Technique

This chapter analyses and compares the accuracy of different existing techniques mentioned in chapter two to determine the correlation factors and the VRI for a group of small-scale rooftop PV systems in a suburban area. Considering the differences in the VRI results, an enhanced VRI model is presented. Then, by using a Wavelet-based transform model, the generated output power of a group of neighbouring PV systems within 0.7 km² and irradiance data collected every one-minute by employing only one pyranometer, is calculated and estimated. The impact of the pyranometer location on the estimation accuracy is also evaluated. The results and discussion are presented in detail below.

3.1 Wavelet Transform

The WT is a powerful mathematical tool to analyse time-varying data. It can divide a time-varying signal into wavelets (i.e. wave-like oscillations with an amplitude that begins at zero, increases, and then decreases back to zero). Unlike the Fourier transform that only constructs frequency representations of a signal (see Fig. 3.1a a), the Wavelet transform can construct a signal's time-frequency representation when the signal possesses a good time and frequency localisation (see Fig. 3.1 b). The Wavelet transform is suitable for localising both the time and frequency domains, and can also discover several properties of the signal, such as the direction, breakdown point, discontinuities and self-similarity [69, 70]. It is also reversible, and furthermore, it can detect the frequency spectrum and their corresponding timescales for both periodic and non-periodic signals.

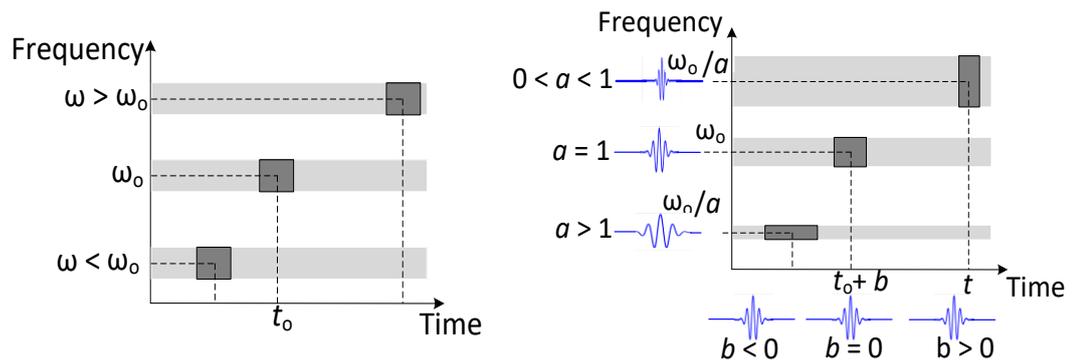


Fig. 3.1a Sample illustration of the performance of a Fourier transform

Fig. 3.1 b Sample illustration of the performance of a WT

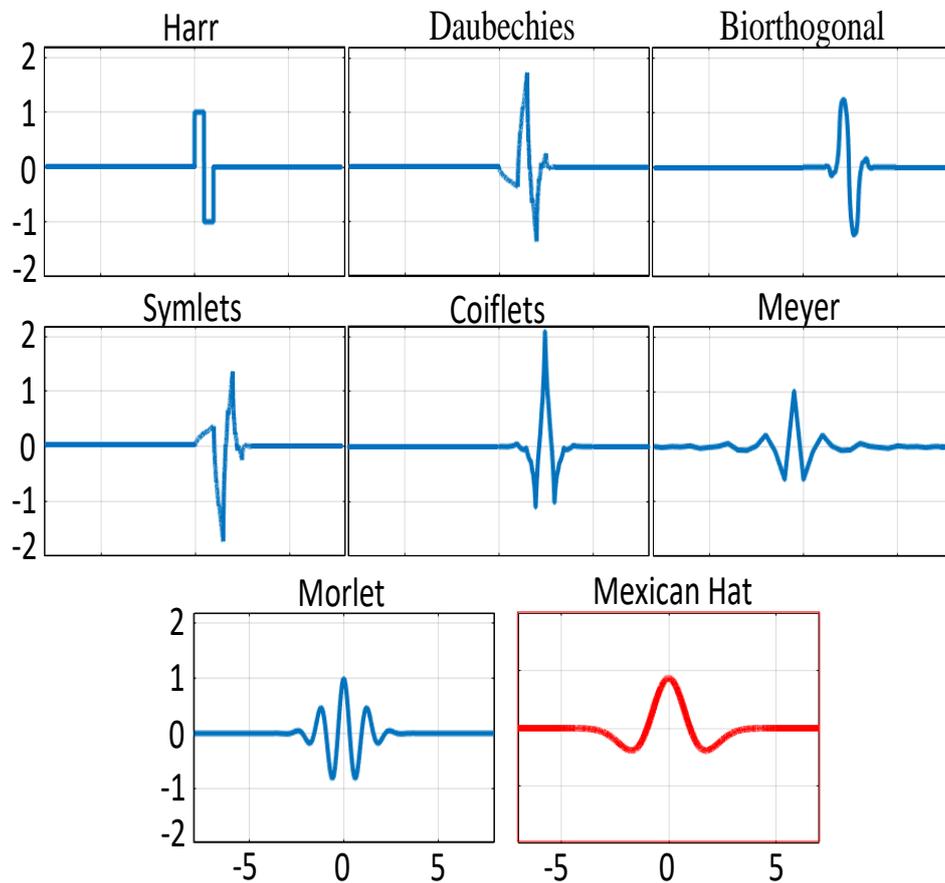


Fig. 3.2 Sample illustration of various Wavelet mother functions

The Wavelet transform for a continuous-time signal at a scale of $u > 0$ and a time shift factor of $g > 0$ is formulated as [71]:

$$WT_{\text{cont}}(u, g) = \frac{1}{\sqrt{|u|}} \int_{-\infty}^{\infty} \psi\left(\frac{t-g}{u}\right) \times x(t) dt \quad (3.1)$$

in which $\psi(t)$ is referred to as the Wavelet mother function and is a continuous function in both the time and frequency domains. The Wavelet mother function detects the signal changes and can explain it in a frequency-time plane. Various Wavelet mother functions are available such as Harr, Daubechies, Biorthogonal, Symlets, Coiflets, Meyer, Morlet and Mexican Hat [72-74] (see Fig. 3.2). Using (3.1), the translated and scaled versions of $\psi(t)$ are referred to as sub or daughter wavelets. As the POA data, captured by a sensor in this study, is discrete, (3.1) is expressed in the discrete domain as:

$$WT_{\text{disc}}(i, f) = \frac{1}{\sqrt{2^i}} \sum_{k=1}^{\infty} \psi\left[\frac{k-f \times 2^i}{2^i}\right] \times x[k] \quad (3.2)$$

in which $i > 0$ is an integer number that shows the modes of the discrete analysis, and f is an integer and discrete signal.

The WT model has been used in this thesis because the time-varying signal of the POA is discrete (due to the sampling intervals of a few seconds to a minute) and also non-periodic (due to the random variations caused by clouds passing). Thus, (3.2) is modified accordingly to derive the Wavelet model for the considered POA data, as depicted in (3.3). In addition, among various types of Wavelet mother functions, the Mexican Hat mother function has been used in this research, as its capability in yielding accurate results in the same field of study have been proven in [28, 49] versus other Wavelet mother functions.

$$WT_{\text{mode-}i}[k] = \frac{1}{\sqrt{T_{\text{mode-}i}}} \sum_{k=1}^{\infty} \psi\left[\frac{k-c \times T_{\text{mode-}i}}{T_{\text{mode-}i}}\right] \times POA[k] \quad (3.3)$$

where $WT_{\text{mode-}i}[k]$ refers to the resulting wavelet at mode- i ($1 \leq i \leq 6$) while $\psi[k]$ denotes the considered Wavelet mother function; $T_{\text{mode-}i}$ is the considered timescale for the Wavelet analysis.

3.2 Proposed Methodology

A technique is proposed in this chapter that aims at improving the accuracy of estimating the overall generated power by a group of neighbouring small-scale

rooftop PV systems. To this end, first, using a pyranometer, the solar irradiance is measured at a single point, across the neighbourhood PV systems, over a day. If the pyranometer is tilted at the same angle as the PV array and has the same orientation, its output is the irradiance incident on the plane of the array [7, 75]. Given the output of the sensor as $POA[k]$, the output power of a PV module ($P_{\text{module}}^{\text{POA}}[k]$) at that location can be calculated from [76]:

$$P_{\text{module}}^{\text{POA}}[k] = 0.12 A_m POA[k] \times \left[1 - 0.004 \left(Tmp[k] + \frac{0.32 \beta \times POA[k]}{8.91 + 2 V_w[k]} - 25 \right) \right] \quad (3.4)$$

in which A_m is the area of the PV module; $Tmp[k]$ and $V_w[k]$ are respectively the ambient temperature and wind speed while β is a mounting coefficient that equals 1.2 and 1.8 for a flat and sloped PV array installation, respectively [76]. In (3.4), k denotes the discrete time index. It is to be noted that if the pyranometer is flat mounted, its output will be the global horizontal irradiance (GHI). If this value is employed in (3.4), slight accuracy issues (e.g., in the range of 1-3% [77]) or even higher accuracy issues (e.g. up to 18.75% [78]) will be imposed, depending on the PV panels' tilt and azimuth. For such cases, the methods suggested in [7, 75] can be employed, which aim at finding more accurate methods to determine the POA from the GHI, as this projection is a highly non-linear transformation.

$POA[k]$ is then normalized to define $POA_{\text{norm}}[k]$ from:

$$POA_{\text{norm}}[k] = \frac{POA[k]}{POA_{\text{clr}}[k]} \quad (3.5)$$

using the clear sky POA of that day (denoted by $POA_{\text{clr}}[k]$), which is computed from [77]:

$$POA_{\text{clr}} = POA_{\text{clr}}^d + POA_{\text{clr}}^g + POA_{\text{clr}}^s \quad (3.6)$$

where POA_{clr}^d , POA_{clr}^g and POA_{clr}^s respectively denote the direct, ground-reflected diffuse and sky diffuse irradiances incident on the PV array, that are determined from the clear sky condition. Thus, both POA and POA_{clr} are the assumed input parameters for the proposed method. It is to be noted that some of the parameters

used in (3.4) and (3.6), such as the POA, the ambient temperature and wind speed, need to be measured by sensors; some, such as the size of the PV module, need to be checked from the PV catalogue; some, such as the tilt angle of the PV array, need to be defined by a site visit or by referring to the installation documents of PV arrays, while some, such as [79-81], need to be extracted from available resources on the internet for a locality's latitude and longitude, as well as the time of year. To avoid disturbing the flow of the text, all the relevant parameters employed in the equations are listed in Table 3.1. This table also discusses the factors affecting these parameters, as well as the method for calculating or extracting them.

$POA_{\text{nrm}}[k]$ is then analysed using a discrete Wavelet transform model with a Mexican Hat type Wavelet mother function [28, 82]. This technique has been extensively employed on non-periodic signals in power engineering applications (e.g. in [83-85], as well as for analysing solar irradiance data in [11, 12, 28, 48, 49, 86, 87]). Using the Wavelet transform, $POA_{\text{nrm}}[k]$ is decomposed into 6 modes using (3.3), where $WT_{\text{mode-}i}[k]$ refers to the resulting wavelet at mode- i ($1 \leq i \leq 6$) as listed in Table 3.2. As seen from this table, POA_{nrm} is decomposed into modes 1 to 6, which respectively show 2, 4, 8, 16, 32 and 64-minute timescales on which the study focuses.

At the same time, $(\rho_{m,n}^z)^{\text{eq}}$ is to be calculated from (2.9), for the Hoff, Perez, Lave, and ACM correlational models, introduced in chapter two, while replacing T with the above-mentioned figures for $T_{\text{mode-}i}$. This will express the overall correlational coefficient of the rooftop PV systems in the neighbourhood at the desired timescales. Then, $(\rho_{m,n}^z)^{\text{eq}}$ will be used in (2.10) to define the VRI of those PV systems, at each assumed timescale. This process yields six values for VRI, each representing one of the assumed timescales, for each of the four utilized correlational models (i.e. a total of $4 \times 6 = 24$ VRIs).

To improve the accuracy of the overall output power estimation, it is suggested to select the largest VRI, given by any of the models in (2.10) at each of the above-mentioned timescales (denoted by $VR_{\text{mode-}i}^{\text{max}}$); that is:

Table 3.1 The definition and method for calculating or retrieving the parameters, employed in (3.4) and (3.6)

Parameter	Definition	Dependence	Measured by or calculated from
A_m (m ²)	Size of a PV module		PV catalogue
POA_{clr}^d (W/m ²)	Direct irradiance incident on the PV array	-Direct-normal irradiance (DNI_{clr}) [*] -Tilt angle of the PV array (θ_T) [#] -Solar azimuth angle (θ_A^{solar}) [*] -Azimuth angle of the PV array (θ_A^{array}) [*] -Zenith angle of the PV location (θ_Z) [@]	$= DNI_{clr} \times (\cos(\theta_Z) \times \cos(\theta_T) + \sin(\theta_Z) \times \sin(\theta_T) \times \cos(\theta_A^{solar} - \theta_A^{array}))$ [77]
POA_{clr}^g (W/m ²)	Ground-reflected diffuse irradiance on the PV array	-Clear-sky horizontal irradiance (GHI_{clr}) [*] -Tilt angle of the PV array (θ_T) [#]	global $POA_{clr}^g = 0.1 GHI_{clr} \times (1 - \cos(\theta_T))$ [77]
POA_{clr}^s (W/m ²)	Sky diffuse irradiance on the PV array	-Clear-sky horizontal irradiance on ground level (DHI_{clr}) [*] -Tilt angle of the PV array (θ_T) [#]	$POA_{clr}^s = 0.5 DHI_{clr} \times (1 + \cos(\theta_T))$ [48]
Tmp (Celsius)	Ambient temperature	-Weather	-Weather monitoring system
V_w (m/s)	Ambient wind speed	-Weather conditions	
β	Mounting	-Mounting style of the	-Respectively 1.2 and

coefficient	PV array	1.8 for flat and sloped PV arrays
-------------	----------	--------------------------------------

* Retrieved for a PV location and day of the year from [81]; @ Retrieved for a PV location and day of the year from [79]; # retrieved from a PV site.

$$VRI_{mode-i}^{max} = \max(VRI_{mode-i}^Z) \quad (3.7)$$

where $\max(\cdot)$ is the maximum function. In such a scheme, the correlational coefficient model, which has the best representation of correlations among the PV systems at each timescale, will be employed. Therefore, the mismatch between the generated power by the PV systems and the estimated value, according to the proposed single-point POA, will be minimised.

Table 3.2 Considered Wavelet modes and corresponding timescales in this chapter

Mode- i	Timescale (T)
1	$2^1 = 2$ minutes
2	$2^2 = 4$ minutes
3	$2^3 = 8$ minutes
4	$2^4 = 16$ minutes
5	$2^5 = 32$ minutes
6	$2^6 = 64$ minutes

At this stage, the resulting wavelets from decomposing $POA_{nrm}[k]$ using (3.3) are divided by $\sqrt{VRI_{mode-i}^{max}}$ to derive the equivalent normalized POA, $(POA_{nrm}^{eq})_{mode-i}[k]$, seen by all PV systems in the neighbourhood over that timescale, as given by [28]

$$(POA_{nrm}^{eq})_{mode-i}[k] = \frac{WT_{mode-i}[k]}{\sqrt{VRI_{mode-i}^{max}}} \quad (3.8)$$

Now, the equivalent normalized POA over the considered day ($POA_{nrm}^{eq}[k]$) can be calculated by properly merging the values given by (3.8) with each other, using an inverse discrete Wavelet function, presented as:

$$POA_{nrm}^{eq}[k] = \sum_{i=1}^6 (POA_{nrm}^{eq})_{mode-i}[k] \quad (3.9)$$

At this stage, the equivalent POA seen by all PV systems in the neighbourhood can be determined from (3.10) and using the day's clear sky POA, as given by:

$$POA^{eq}[k] = POA_{nrm}^{eq}[k] \times POA_{clr}[k] \quad (3.10)$$

Replacing $POA[k]$ with $POA^{eq}[k]$ in (3.4) yields the approximate average output power of the PV modules in that neighbourhood over the considered day (denoted by $P_{module}^{POA^{eq}}$). Multiplying this figure with the total number of PV modules available in the neighbourhood (denoted by N_{module}) results in the estimated overall power of all rooftop PV systems, as given by:

$$P^{eq}[k] = N_{module} \times P_{module}^{POA^{eq}}[k] \quad (3.11)$$

Therefore, in summary, using the POA, measured at the location of one of the PV plants, and calculating the above-mentioned correlational factors, the aggregated overall power of all rooftop PV systems can be estimated, as detailed above. The above steps are schematically illustrated in the flowchart in Fig. 3.3.

It is to be emphasised again that this chapter has only focused on geographical smoothing because of the PV systems distributed in an area that sees various irradiances because of cloud movement. Therefore, the proposed method is not applicable for a spatially persistent model in which the irradiance is the same across all the PV systems which can occur on (a) a clear sky day without any clouding events, (b) a cloudy day in which a significantly large and uniform cloud is not causing variation in irradiances for the PV systems throughout the day. In both cases, the cloud speed estimation technique will not determine any numerical value for the cloud speed. Therefore, the proposed technique is inapplicable. It should be noted here that this is not a limitation for the proposed technique but is expected, as one of its inputs is unknown, and therefore the technique cannot yield any numerical outcome for those days.

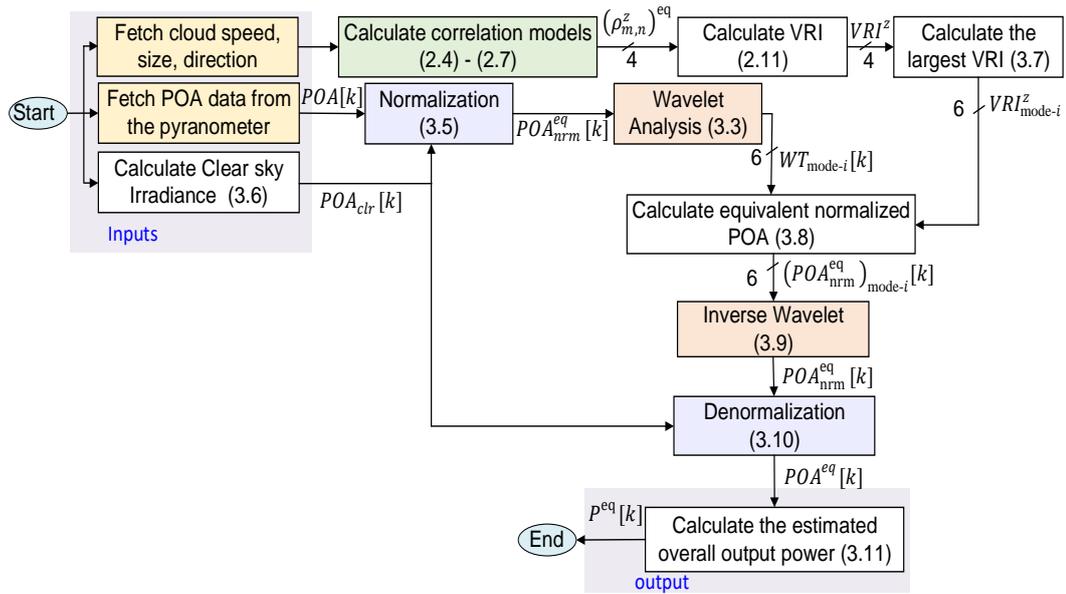


Fig. 3.3 Flowchart of the proposed technique for estimating the overall power of a group of neighbouring PV systems

3.3 Performance Evaluation

The St Lucia campus of the University of Queensland in Brisbane, Australia, is situated at a longitude of $153^{\circ}29'44''$ E and latitude of $27^{\circ}00'30''$ S. This campus has 16-rooftop PV systems with a total capacity of 2.14 MW distributed over 1 km^2 [27]. To obtain the highest output power, the PV modules face north and are tilted at 30° to be approximately equal to the site's latitude. Table 3.3 lists these PV systems, along with their dimensions, types and capacities, while Fig. 3.4 shows their locations on the map.

The city of Brisbane is characterised by humid subtropical weather and it experiences larger cloud coverage in summer than in winter. In winter, it is dry and middling warm and in summer, hot and humid.

The ambient temperature recorded at the site is $20.4\text{--}30.3^{\circ}\text{C}$ over the summer (December–February), $13.7\text{--}29^{\circ}\text{C}$ in autumn (March–May), $10.2\text{--}21.9^{\circ}\text{C}$ in winter (June–August), and $13.8\text{--}28.4^{\circ}\text{C}$ in spring (September–November) [27]. In addition, the wind at the site each month has a dominant direction, as indicated in Table 3.4.

Table 3.3 Considered distributed rooftop PV systems at the studied site

No	Location	Power rating (kWp)	Array area (m ²)	Module size (m)	N_{module}
1	UQ Center*	433.44	2,956	1.65 × 0.99	1,806
2	Multi-Level Carpark*	677.8	4,610	1.65 × 0.99	2,824
3	Sir Llew Edwards *	89.76	612	1.65 × 0.99	374
4	Prentice**	13.5	78	1.65 × 0.99	56
5	Advanced Engineering*	95.75	640	1.67 × 1	383
6	Learning and Innovation*	45.25	303	1.67 × 1	181
7	Global Change Institute*	138	941	1.65 × 0.99	575
8	General Purpose North*	66.5	432	1.64 × 0.99	266
9	Warehouse*	164	1,065	1.64 × 0.99	656
10	Queensland Bioscience Precinct**	34.66	173	1.56 × 1	106
11	Joyce Ackroyd*	54.3	342	1.65 × 0.99	209
12	Colin Clark *	51.5	324	1.65 × 0.99	198
13	Axon*	34.32	216	1.65 × 0.99	132
14	GP South *	72.56	457	1.65 × 0.99	279
15	Duhig North*	137.65	839	1.95 × 0.98	437
16	Pedestrian Link Bridge*	15.25	97	1.96 × 0.99	50

p-Si* = Polycrystalline Silicon,

m-Si** = Mono-crystalline Silicon

kWp represents the peak power in killo-watt generated by the PV system at mid-day (12:00 pm) where the irradiance is usually at its maximum level.

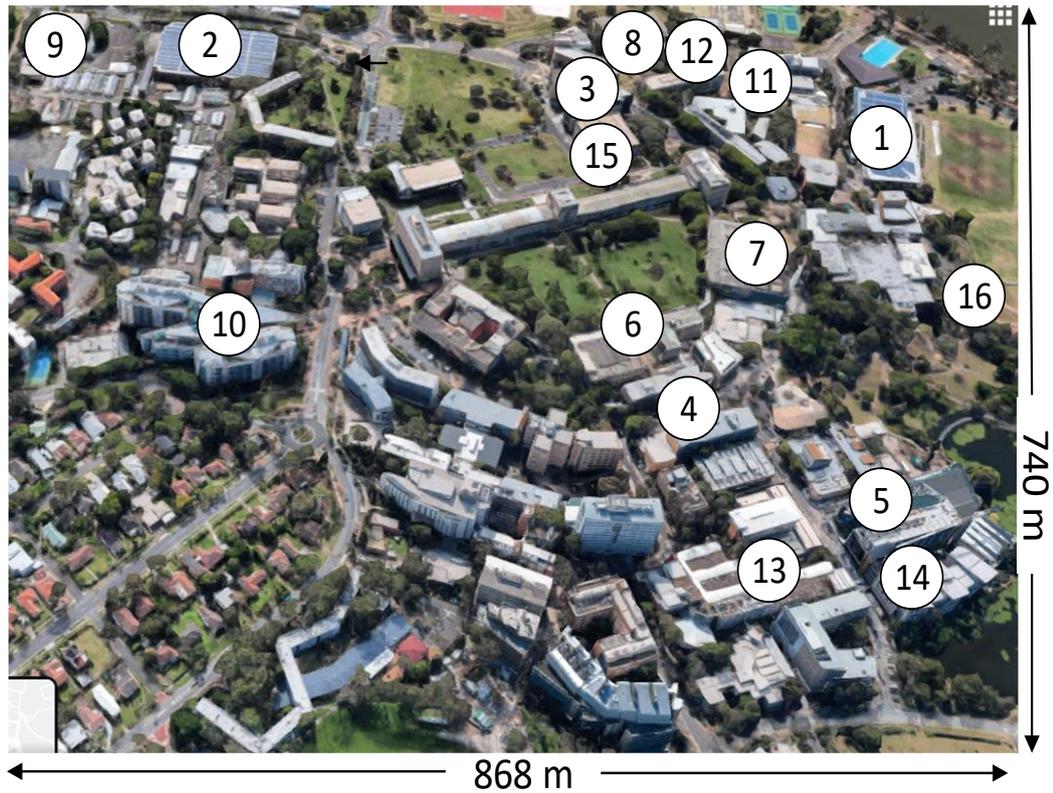


Fig. 3.4 The site of the study indicating the locations of the PV systems

The distributed PV systems at the study site suit the main concept of this study well because of the availability of the POA, seen by each rooftop PV system, and their individual generated power, along with the availability of the ambient temperature and wind speed, at one-minute resolution.

On the other hand, [27] and [77-81, 88] have been used to define the parameters employed in (3.4) and (3.6), considering the assumed locality and the focused time of the year. The cloud speed is also required in correlational coefficient models (2.4) to (2.7); however, because of the unavailability of this data, cloud speed is approximated in this study using the method proposed in [12], as detailed in section 3.2 below.

First, as mentioned above when calculating the correlational models (2.4)–(2.7), the cloud speed is required. This can be achieved by a commercially available product; however, this technology is now under development [89]. Therefore, many

Table 3.4 Dominant wind direction and range of ambient temperature at the study site for 2016–2017.

Season	Month	Wind direction	Ambient temperature (°C)
	March	93% West	21.2–25.2
Autumn	April	92% West	20.3–25.3
	May	96% West	16.3–25
	June	93% West	8.1–19.3
Winter	July	20% South, 19% Southwest	11–23
	August	65% West	12.6–20.9
	September	85% Southwest	12.6–19.7
Spring	October	88% west	17.4–22.9
	November	74% west	21.7–26
	December	50% west	22.1–27.7
Summer	January	93% West	21.1–27.5
	February	47% West	23.4–29.3

researchers use a technique to estimate the cloud speed. As an example, [12] and [13] estimate the cloud speed using two methods: The first method uses cross-correlation models for POA between each pair of sensors while the second uses three sensors distributed around a specific area and considers the time delay and distance between the points. Similarly, [12] uses a number of pyranometers on the ground to derive a scaling factor from which the cloud speed is calculated while the method in [90] collects satellite images and analyses them for this purpose. Most researchers prefer the ground method because data acquisition and interpretation from satellite images are complex and challenging [12, 14]. In the studies of this chapter, the cloud speed is derived using the method from [12]. Some pyranometers are considered across the neighbourhood. Using the Wavelet transform model of

(3.3), their data are decomposed into six modes, and the correlation among every two sensors is determined using (2.1) in each mode. The defined correlational model is assumed to be equal to that given by (2.6), from which the cloud speed is approximated on the basis of:

$$V_c[k] = \frac{-2}{T \times L_{m,n}} \ln(\rho_{m,n}^{\text{Pearson}}[k]) \quad (3.12)$$

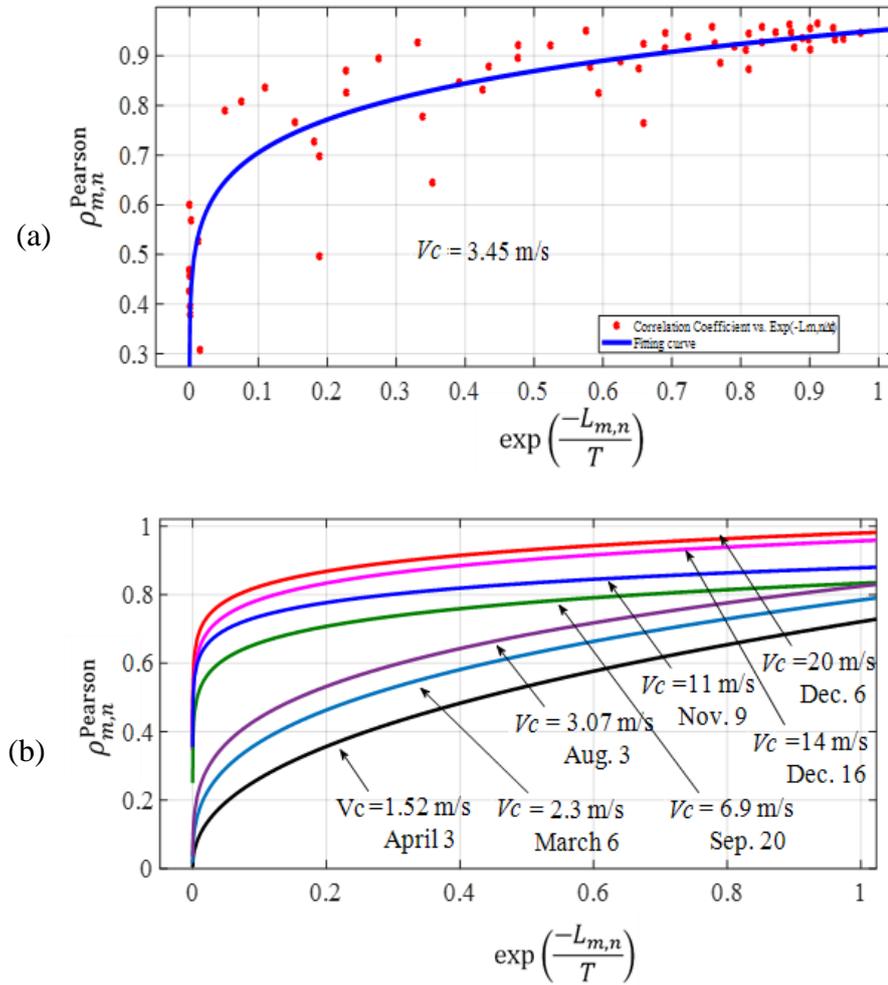


Fig. 3.5 Approximated cloud speed for eight days in different seasons: a) calculated correlation coefficient on 04/07/2016, b) the fitted results for seven sample days

The POA data from 5 rooftop PV modules are collected for each of 312 days. Then, considering the different timescales, the correlation of these PV systems

versus $\exp(-L_{m,n}/T)$ is plotted, as shown in Fig. 3.5 a. The results are fitted with an exponential curve to determine the approximate cloud speed. Using the same technique, Fig. 3.5 b illustrates different cloud speeds captured for some of the day's under study, varying between 1.52 and 20 m/s.

This information is used in correlational factors. It is to be noted that in this study, it is assumed that the speed, direction and shape of the cloud passing over one PV system is the same as that when it is passing over the other neighbouring PV systems during the same period [11, 20, 21]. This assumption is in line with other research in this field including research into cloud speed. The Wind Rose Rule is also used to define cloud direction based on which the direction of the wind is divided into eight directions of North (337.5–22.1°), Northeast (22.2–67.4°), East (67.5–112.4°), Southeast (112.5–157.4°), South (157.5–202.4°), Southwest (202.5–247.4°), West (247.5–292.4°), and Northwest (292.5–337.4°) [91].

The data from the pyranometer, installed next to PV site-7, is employed as the single POA measurement for the rest of this study. Fig. 3.6 shows the output power of the rooftop PV modules installed at the considered site on 6 March 2016. One of the curves in this figure corresponds to the generated power from one PV module (located at PV site-10), while the other two curves depict the average power, generated by a group of 7 and 16 PV modules across the studied site. As can be seen in Fig. 3.6, the power generated by the single PV module is larger than the average generated power by the group of 7 and 16 PV modules around 9:15 am.

This is because a small cloud passes over the neighbourhood which has not blocked the solar irradiance exposed to the single PV module, but is blocking it for the pyranometer. However, it can be seen that around 2:56 pm, a large cloud is passing over the neighbourhood and the measured power, generated by the considered single PV module, closely matches the average of the generated power by the group of 7 and 16 PV modules.

The proposed method with the flowchart in Fig. 3.3 is applied for this system over 312 days among 365 days in 2016 and 2017 (excluding clear sky days). The study considers days with large, intermediate and low variability to test the VRI during different ambient conditions.

Fig. 3.7 shows the average of the calculated VRIs from (2.10) for all considered rooftop PV modules for the days under study, using correlational coefficient models (2.4) to (2.7). The results are provided for each timescale (in Table 3.2) separately.

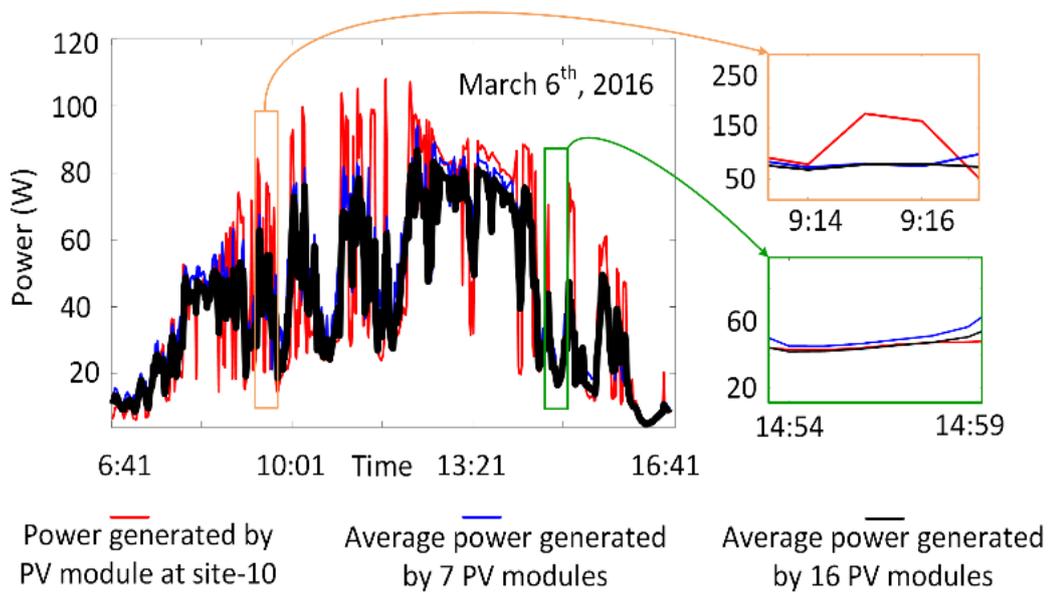
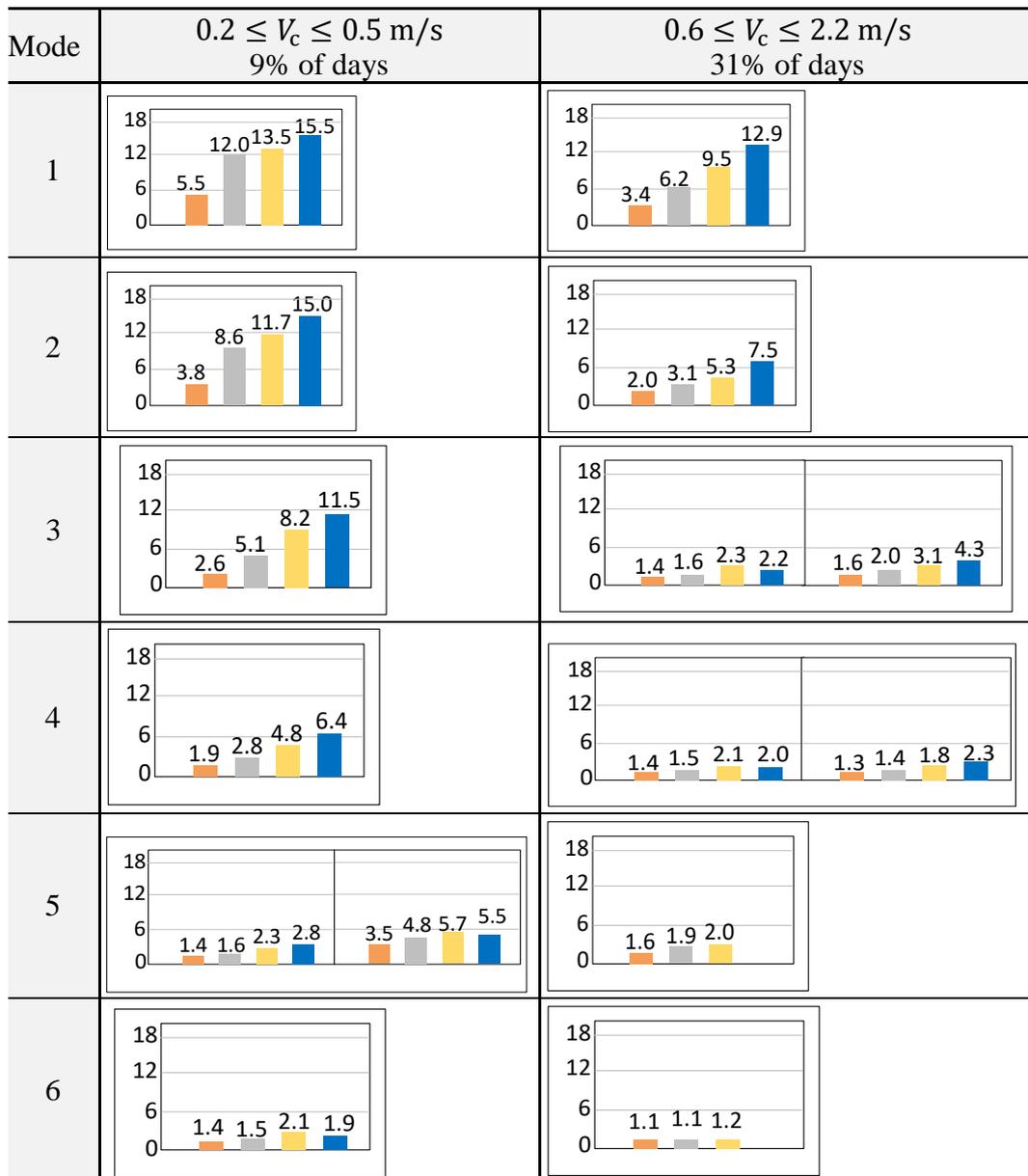
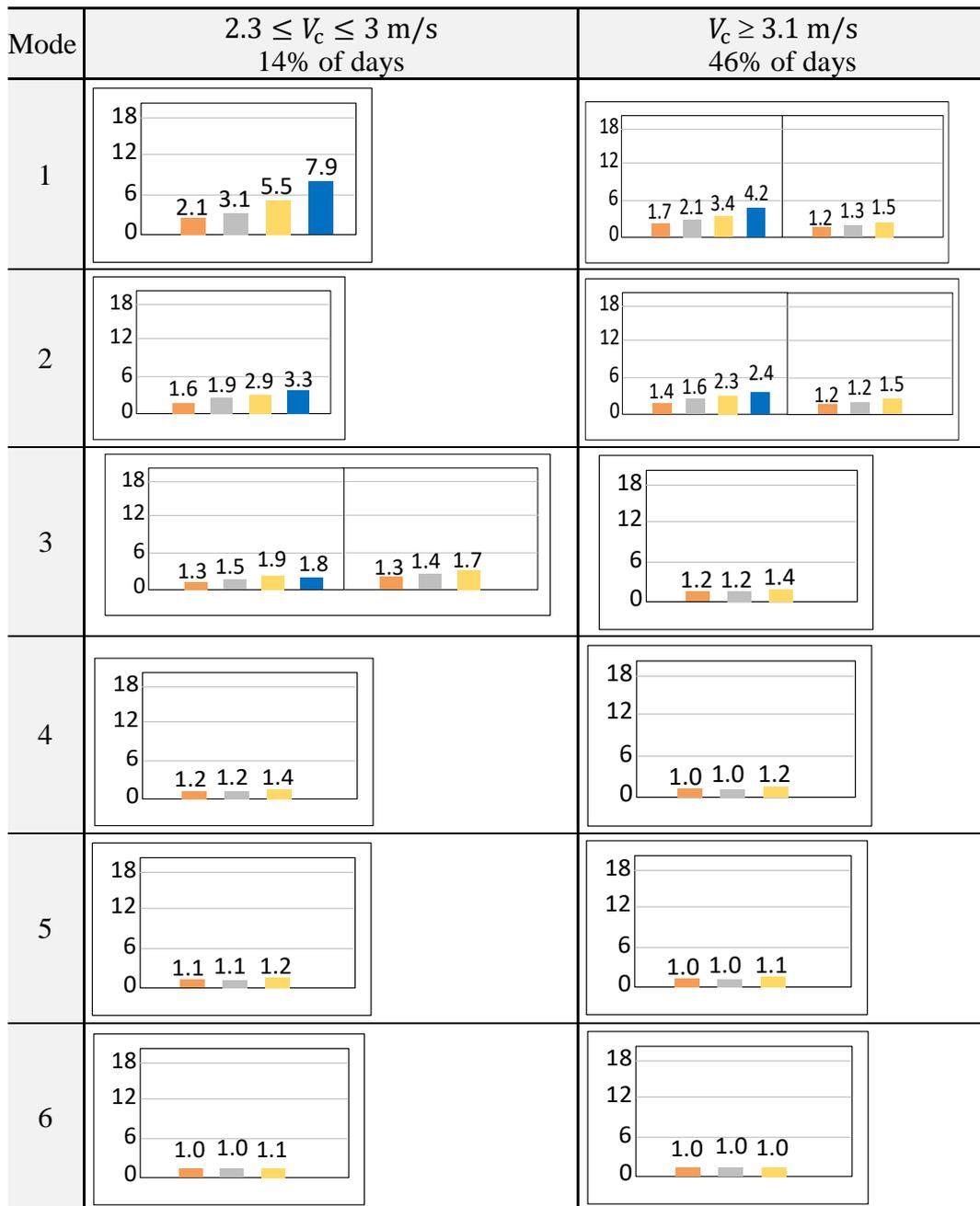


Fig. 3.6 The power generated by the PV module at PV site-10, as well as the average power generated by 7 and 16 rooftop systems across the studied site

Moreover, they are classified based on different ranges of the calculated cloud speed. Results show that for most days in the study, the trend in the correlational factors is the same for cloud speeds above 3.1m/s. Therefore, Fig. 3.7 has been categorised under cloud speeds of $0.2 \leq V_c \leq 0.5$, $0.6 \leq V_c \leq 2.2$, $2.3 \leq V_c \leq 3$ and $V_c \geq 3.1$ m/s. If for one range, two different types of trends are observed for the correlational factors, both of them are presented. From this figure, it can be seen that different correlational models yield unequal VRIs at various modes and cloud speeds. For example, for Fig. 3.7a, mode-3 at a cloud speed range of $0.2 \leq V_c \leq 0.5$ m/s, the ACM model provided the largest value for all days, whereas at the cloud speed range of $0.6 \leq V_c \leq 2.2$ m/s, the Lave model gave the largest value for some days (as shown in the first Figure) and the ACM model gave the largest value at the other days (as shown in the second Figure). Table 3.5 summarises the outcomes of Fig. 3.7 and shows that the ACM and Lave models yield the largest average VRI values for most of the modes.



(a)



(b)

■ Hoff ■ Perez ■ Lave ■ ACM

Fig. 3.7 Average VRI calculated using the Hoff, Perez, Lave and ACM models for the considered site over 312 days at various cloud speeds and timescales
a) for $0.2 \leq V_c \leq 2.2 \text{ m/s}$ b) for $2.3 \leq V_c \leq 3 \text{ m/s}$ and $V_c \geq 3.1 \text{ m/s}$

It should be noted that the power generated can experience rapid and large fluctuations, which are an important issue in the operation of power systems; however, the proposed technique in this chapter is more a measurement (estimation)

type technique. The common practice of demonstrating the performance of techniques such as the one proposed in this chapter illustrates the mean absolute error (MAE) or the root mean square error (RMSE) between the measured power generated by all PV modules at the considered site and the estimated power using

Table 3.5 The correlational coefficient models that yield the largest average VRI for the considered site during the period of the study

Cloud speed (m/s)	$0.2 \leq V_c \leq 0.5$	$0.6 \leq V_c \leq 2.2$	$2.3 \leq V_c \leq 3$	$V_c \geq 3.1$	
Day's ratio (%)	5.8	22	23	49.2	
Modes	1	ACM	ACM	ACM	ACM/Lave
	2	ACM	ACM	ACM	ACM/Lave
	3	ACM	ACM/ Lave	Lave	Lave
	4	ACM	ACM/ Lave	Lave	Lave
	5	ACM/Lave	Lave	Lave	Lave
	6	ACM	Lave	Lave	Lave

each of the correlational models [8, 11, 77, 92]. Fig. 3.8 illustrates these two quantities for the period of the study at the considered site. As mentioned in Table 2.5, the ACM model works for all modes if the cloud speed is below 0.5m/s. Therefore, the MAE and RMSE results in Fig. 3.8 for the ACM model are only shown for those days where the cloud speed is below 0.5m/s (and acceptable for the ACM model). From this figure, it can be seen that the proposed technique has the least MAE and RMSE and highest accuracy when compared with each of the existing correlational models. Table 3.6 summarises the findings from Fig. 3.8, and numerically compares the minimum and maximum observed MAE and RMSE for each correlational coefficient model versus the proposed method. From this table, it can be seen that the minimum observed MAE among the investigated correlational coefficient models varies between 2.6 and 3.7% (when the cloud speed is less than 0.5m/s) and between 1.2 and 1.4% (when the cloud speed is above 0.5m/s). However, these figures are respectively 2.5 and 1.1% for the proposed method. Likewise, the maximum observed MAE among the investigated

correlational coefficient models varies between 14.2 and 16.5% (when the cloud speed is less than 0.5m/s) and between 16.5 and 17.6% (when the cloud speed is above 0.5m/s); however, for the proposed method, these figures are 13.9 and 16.2% depending on the cloud speed.

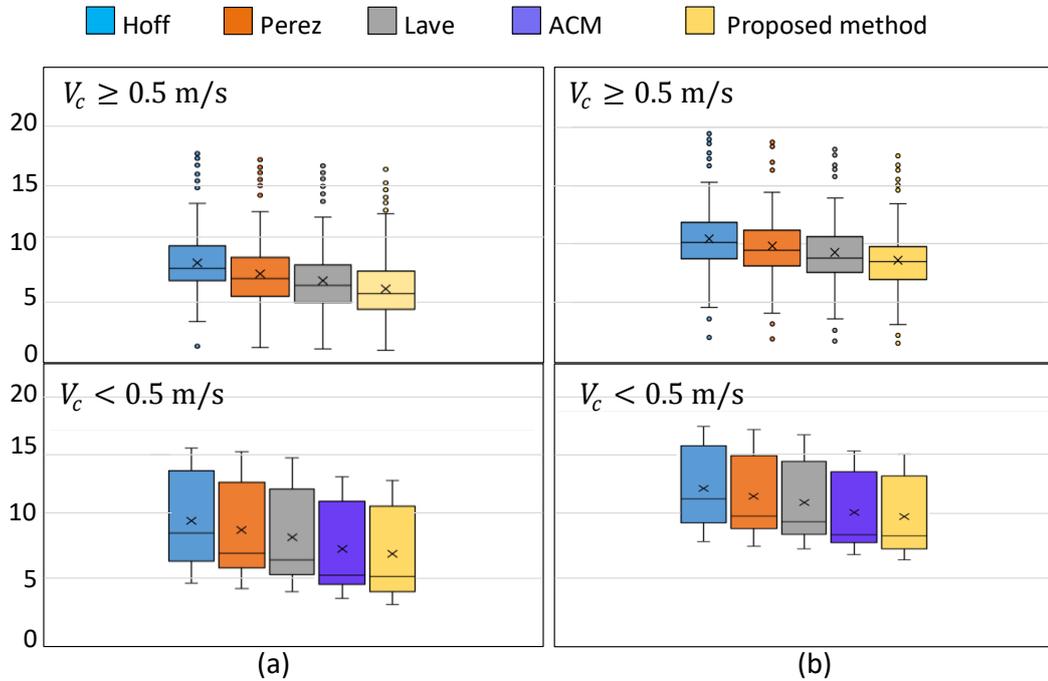


Fig. 3.8 Comparison of the error between the power generated and estimated using different VRI models, calculated under (a) MAE, and (b) RMSE

As such, the proposed method demonstrates minimum and maximum improvements respectively of 1.8 and 48% in the accuracy of the estimated total power, versus each considered correlational model based on the MAE. Likewise, this table shows a very similar trend in the improvement using the proposed method when the RMSE criterion is in focus. It can be seen that the proposed method presents minimum and maximum improvements of respectively 1.8 and 32% under the RMSE criterion. Moreover, Eq. (3.7) uses the largest VRI, given by the Hoff, Perez, Lave, and ACM correlational models, at each mode of the Wavelet analysis in the proposed technique. This improves the accuracy of the estimated power using the proposed technique.

To illustrate this, another study is conducted that demonstrates the direct relationship between the accuracy of the estimated power and VRI at each mode of the Wavelet analysis, as discussed below.

Let us consider a sample day. First, the overall power generated by all neighbouring PV systems at the considered study site is calculated and then converted into 6 modes, using the Wavelet transform from (3.3).

Table 3.6 Comparison of the MAE and RMSE for different correlation models

Cloud speed (m/s)	$V_c < 0.5$				$V_c \geq 0.5$			
Correlational coefficient model	MAE		RMSE		MAE		RMSE	
	Min	Max	Min	Max	Min	Max	Min	Max
Perez	3.6	16.2	6.6	17.4	1.3	17.0	1.8	18.9
Hoff	3.7	16.5	7.4	18.0	1.4	17.6	1.9	19.6
Lave	3.0	15.7	6.1	17.0	1.2	16.5	1.6	18.2
ACM	2.6	14.2	5.7	16.0	N/A		N/A	
Proposed method	2.5	13.9	5.6	15.0	1.1	16.2	1.5	17.7
Minimum improvement	4	2.2	1.8	6.6	9	1.8	6.6	2.8
Maximum improvement	48	18.7	32	20	27.2	8.6	26.6	24.6

Then, the overall power is estimated using the proposed technique and the available POA measurement at a single PV location and the VRI defined by each correlational model. To this end, (3.7) has been deactivated. Then, these estimates of the power generated are passed through the Wavelet transform to yield the corresponding values in 6 modes.

Fig. 3.9 a shows the VRI defined by each correlational model at each timescale, while the difference between the actual power in each mode versus that estimated by each correlational model is depicted in Fig. 3.9 b. As seen from this figure, in each mode, the correlational model that has the largest VRI delivers the greatest accuracy in the estimated power. Fig. 3.9 illustrates the results for one day only. Extending this study to all other days in the study, the analysis reveals the same finding. Therefore, a direct relationship is seen between the accuracy of the power estimation versus the VRI's value, which has been employed using (3.7) in the proposed technique, to improve the accuracy of power estimation. This is because

a larger VRI represents a smoother (better averaged) POA that is seen by all PV systems in that area. When this value is used in the proposed technique, the resulting estimated power is much smoother and thus closer to the overall actual power of the PV systems, yielding better estimation accuracy.

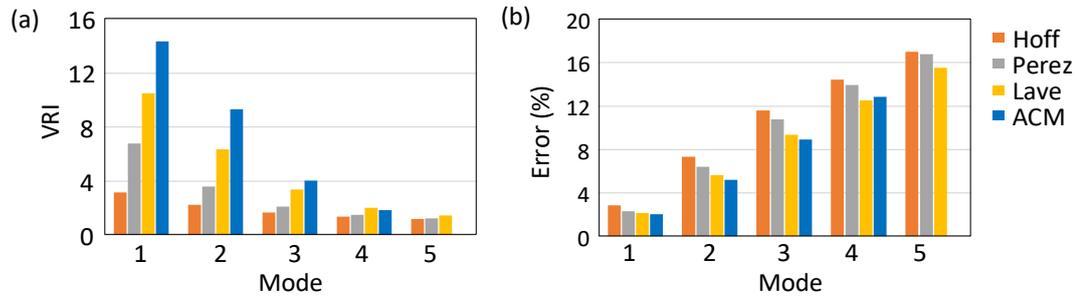


Fig. 3.9 (a) VRI defined using the studied correlational models at each mode, (b) The difference between the anticipated overall power generated by each correlational model and the measured power

It is also important to discuss the accuracy of the available POA data and provide a study to demonstrate the accuracy of the measured POA and ambient temperature by using them in (3.4), and comparing its outcome with the actual power, measured by the wattmeter, at the output of the PV array. The results of this comparison are provided in Fig. 3.10 for 6 sample days. This figure shows an MAE of 0.27–0.45% for the studied days. This finding is also valid for the other sample studied days at the considered site.

The above study demonstrates an MAE of less than 0.5% in the case of using one pyranometer per PV system. As seen in Table 3.6, the MAE varies between 1.2 and 17.6% depending on the employed correlational factor. It is to be noted that this increase in the MAE is mainly due to the use of a single pyranometer for the whole neighbourhood, which involves less capital expenditure. The proposed technique reduces the MAE to between 1.8 and 48%, depending on the cloud speed.

It is worth mentioning that the best calibration reported for the accuracy of pyranometers is 3% [90]. However, there is no officially recognised standard that is used to determine the net accuracy of a pyranometer measurement in the field for any given measurement condition. It is reported in [91] that the accuracy estimations published by pyranometer manufacturers are subjective because they do not conform to any standardized test method. Some manufacturers characterise

certain instrument offset characteristics in the lab; however, it is not always possible to transfer such test data from the laboratory environment to the real world

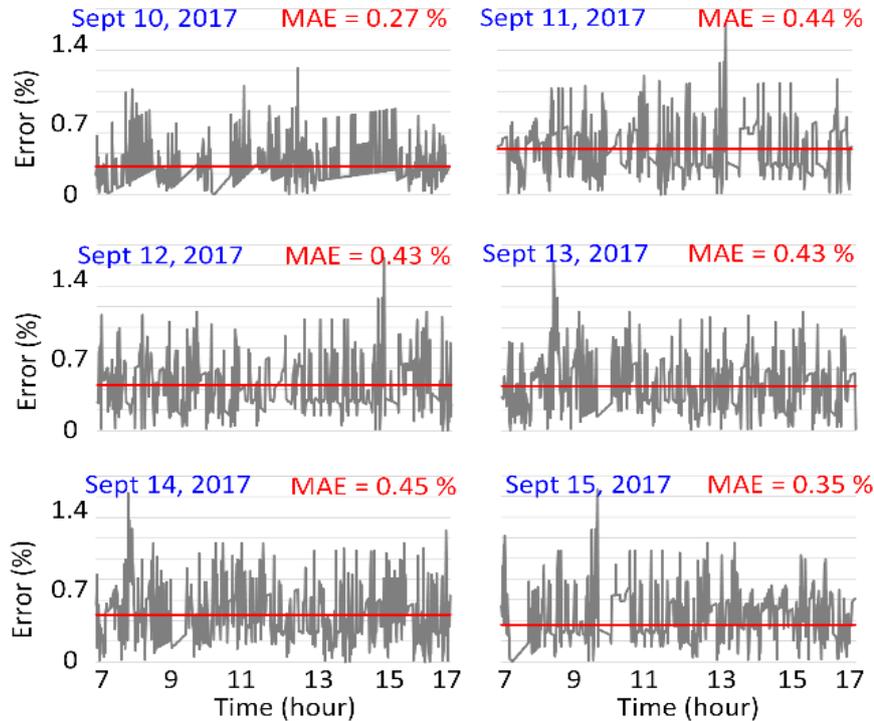


Fig. 3.10 Error and MAE of the calculated power using (11) and the power measured by the wattmeter for the PV array located next to the pyranometer over 6 sample days

Ultimately, to evaluate the impact of the location of the pyranometer on the accuracy of the estimated total output power, another study has been conducted. In general, the location of the most suitable point for installing the pyranometer, when a single POA measurement point is considered, highly depends on the ambient factors and needs to be evaluated for any considered site. In this chapter, the location of the pyranometer is assumed to be varied among 8 points across the considered site (i.e. next to PV sites 4, 7, 8, 9, 10, 11, 13, and 16 in Fig. 3.4) in an exhaustive search analysis. The remaining PV sites are not considered because they are very close to one of the above-mentioned selected sites. In each analysis, the above considered 312 days are re-analysed using the proposed method, and the percentage of days that the proposed method observes an accuracy above 90% in the estimated total output power is determined. This percentage is provided in Table 3.7 for each assumed location of the pyranometer. From this table, it can be seen

that when the pyranometer is located on either the northern or southern parts of the investigated site (e.g. PV sites 8, 11 and 13), the proposed method has an overall high accuracy during the investigated period. Among those when the pyranometer is located at PV site 8, the proposed method has an accuracy above 90% on 81% of the studied days. On the other hand, when the pyranometer is located on the western side of the studied site (e.g. PV sites 7 and 16), the proposed method has the least accuracy. As an example, for the pyranometer at PV site 16, the proposed method has an accuracy above 90% on only 20% of the studied days. Therefore, PV site 8 is found to be a very suitable location for the installation of the pyranometer for this case study.

Table 3.7 Percentage of days in which the proposed method has an accuracy of over 90% when the pyranometer is located at various PV sites.

Pyranometer located at site-	Percentage of days (%)
4	51
7	25
8	81
9	60
10	40
11	75
13	74
16	20

3.4 Summary

This chapter presented several models to express the correlation between two PV systems located in one area. These correlational models are used to determine the VRI of the power generated by these PV systems and estimate their overall generated power. This chapter focuses on defining the accuracy of various

correlational models in the literature when applied to a group of small-scale rooftop PV systems, distributed in a small area, versus large PV systems distributed over large areas, for which these models were developed and validated. Employing these models for the real system under consideration revealed that no single analysed correlational model always results in the most accurate estimation. The chapter then proposed a technique, which combines the outcomes of the existing correlational coefficient models, to improve the estimating accuracy. The Wavelet transform model with six modes has been used, which maximises the estimation accuracy by selecting a VRI from a model that has the highest value at each timescale. The developed method is validated by comparing the measured power generated by a group of 16 neighbouring small-scale rooftop PV systems, with the estimation results from the proposed method. The results show that the proposed method improves the accuracy by at least 1.8% and up to 32 or 48% (depending on whether the MAE or RMSE are employed) for 312 non-clear sky days of a single year. It must be noted that this technique uses a single pyranometer for the considered distributed PV systems. Hence, a study has also been conducted to investigate the location impact of the pyranometer on the results, and hence, the most appropriate location has been identified.

Chapter 4 ANFIS-Based Approach

This chapter presents a machine learning application by using the Adaptive Network-based Fuzzy Inference System (ANFIS) and Wavelet transform model to improve a model to estimate the output power of a group of neighbouring rooftop PV systems within 0.7 km². The model is derived by employing one pyranometer located close to the PV systems. The results of the model validation are also introduced. This chapter demonstrates that the accuracy of the proposed model is better than the existing models mentioned in the literature. A discussion and sensitivity analysis of the model parameters have been demonstrated and plotted.

4.1 ANFIS

Traditional mathematical modelling methods are inappropriate tools for dealing with stochastic and uncertain data, such as solar irradiance. More accurate modelling can be achieved based on the integration of the artificial neural network (ANN) and fuzzy control system (FCS) features to form ANFIS, which is a powerful technique that can map the relationship between a set of input and output variables [93]. ANFIS combines the features of ANN and FCS and can function in a noisy, uncertain and indefinite environment. In this technique, the ANN provides the FCS with learning capabilities, while the FCS provides ANN with an organised structure with well-established fuzzy rules reasoning and thinking [94]. The two learning algorithms adopted in the ANFIS technique are the back-propagation and hybrid approaches that act to reduce the error between real and estimated data [95].

FCS is a key system for making decisions in a fuzzy logic system as a primary function by using IF-THEN rules and by using OR or AND functions to map the essential decision function. Two famous models are presented in FCS – Mamdani and Sugeno's models. Sugeno's model is computationally more effective and more compact than Mamdani's models with sample data-based fuzzy functions, adaptive and optimization approaches [96]. The output of Sugeno's FCS is crisp and has mathematically intractable defuzzification. Therefore, Sugeno's FCS merged with the ANN technique can be used to model data that have a stochastic trend, such as solar irradiance. Sugeno's FCS uses a database that determines the membership

function (MF), which is used in the fuzzy rules. The membership function is a structural block of a fuzzy system employed for a particular problem since they affect a fuzzy inference system. They can have various shapes such as Gaussian, triangular and trapezoidal. It should be noted that the MF's values must vary between 0 and 1. Each MF takes only one rule, so the number of MFs should be equal to the number of FCS rules. To illustrate the ANFIS architectonics, the example below has been considered. In this example, a first-order Sugeno function and two IF-THEN rules are used [93]:

Rule-1

$$\text{IF } x_o \text{ is } N_1 \text{ AND } y_o \text{ is } M_1, \text{ THEN} \quad (4.1)$$

$$f_1 = a_1 x_o + b_1 y_o + r_1 \quad (4.2)$$

Rule-2

$$\text{IF } x_o \text{ is } N_2 \text{ AND } y_o \text{ is } M_2, \text{ THEN} \quad (4.3)$$

$$f_2 = a_2 x_o + b_2 y_o + r_2 \quad (4.4)$$

where x_o and y_o are the input variables, N_j and M_j are fuzzy sets, f_j the output variable, while a_j and b_j are parameters evaluated through the training mode. The ANFIS architectonics by using the Sugeno function, which is the main function used in the ANFIS modelling technique, can be depicted in Fig. 4.1 [95]. In this depiction, the square nodes refer to adaptive nodes, which are the parameters that have to learn, whereas the circular nodes are fixed nodes, which are the parameters that have to be fixed through the learning mode. ANFIS consists of five layers, and each one is explained below.

Layer-1: This is the first layer that obtains the input variables. The outputs of this layer are the fuzzy MF grade of the input variables given by:

$$O_{1,j} = \mu_{N_j}(x_o) \quad (4.5)$$

where $j = 1,2$

$$O_{1,j} = \mu_{M_{j-2}}(y_o) \quad (4.6)$$

where $j = 3,4$

Where x_o and y_o are the input variables to node j , while N_j and M_j are the linguistic labels (e.g high, low, etc.) linked with this node function and $\mu_{N_j}(x_o)$ and $\mu_{M_{j-2}}(y_o)$ can be any fuzzy MF, such as Gaussian and bell-shaped as stated below, respectively [95]:

$$\mu_{N_j}(x_o) = \frac{1}{1 + \left[\left(\frac{x_o - c_j}{l_j} \right)^2 \right] m_j} \quad (4.7)$$

$$\mu_{N_j}(x_o) = \exp \left[- \left(\frac{x_o - c_j}{l_j} \right)^2 \right] \quad (4.8)$$

where $j = 1,2$ and c_j , l_j , and m_j are the MF parameters.

Layer-2: This layer contains fixed nodes which involve a fuzzy operator such as AND or OR operators indicated by π . The output of this layer can be illustrated as:

$$O_{2,j} = W_j = \mu_{N_j}(x_o) * \mu_{M_j}(y) \quad (4.9)$$

These outputs of the two layers are called rules firing strengths.

Layer-3: This layer contains fixed nodes, as well as the indicator Nor referring to the normalisation stage, so this layer is called normalised firing strengths. The output of this layer can be given by:

$$O_{3,j} = \overline{W}_j = \frac{W_j}{W_1 + W_2} \quad (4.10)$$

Layer-4: The nodes in this layer are adaptive. The output of this layer is the product of the output of layer-3 (normalised firing strengths) and the first-order polynomial (for the first order of the Sugeno function) which can be given by:

$$O_{4,j} = \overline{W}_j f_j = \overline{W}_j (a_j x + b_j y + r_j) \quad (4.11)$$

Layer-5: This is the output of the ANFIS technique. It contains only one fixed node indicated by the Σ sample. This node explained the summation of all output signals illustrated by [93]

$$O_{5,j} = \sum_j \bar{W}_j f_j = \frac{\sum_j W_j f_j}{\sum_j W_j} \quad (4.12)$$

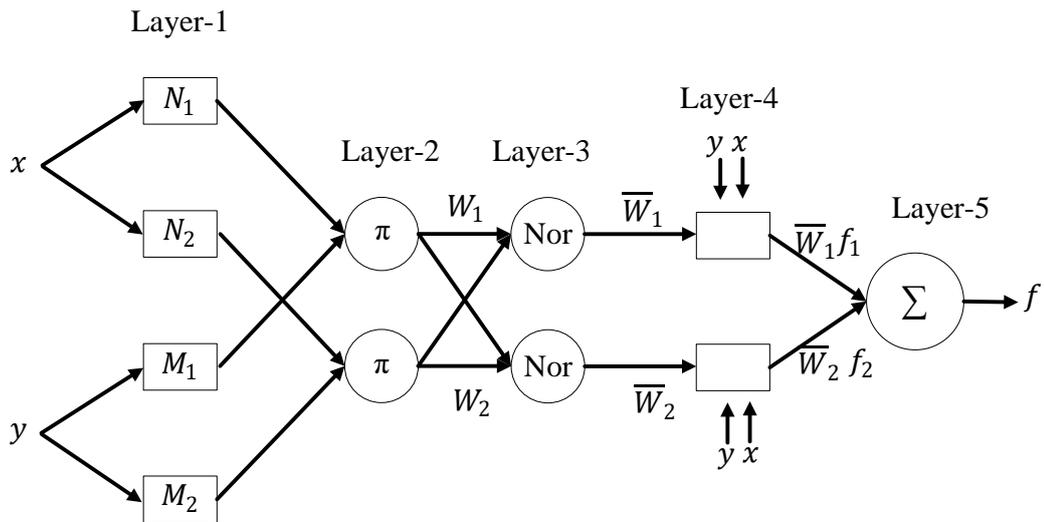


Fig. 4.1 ANFIS architectonics

The two learning algorithms adopted in the ANFIS technique are the back-propagation and hybrid approaches that act to reduce the error between real and estimated data [95]. Since the back-propagation algorithm is relatively slow [97], the hybrid-learning algorithm is adopted to train the proposed ANFIS model in this chapter.

In summary, there are several features [93, 94] that make the ANFIS technique eligible for estimating and predicting models involving uncertain and stochastic data; for example, ANFIS:

- does not need previous human skills;
- uses FCS rules to describe complex systems;
- entails fast and accurate learning;

- contains the desired data, several MFs to employ, and excellent illustration facilities using fuzzy rules;
- can be used for problem-solving with both numeric and linguistic information.

4.2 Proposed Methodology

The proposed technique in this chapter is a combination of a Wavelet transform and an ANFIS. To build a model that aims at improving the accuracy of estimations of the output power of several rooftop PV systems using ANFIS, it is important to identify the input parameters that influence the output power variability such as ambient temperature (Tmp), humidity (H_u), PV plant area (A_d^{pv}), PV system site power capacity (P_d^{cap}), cloud speed (V_c) as well as the solar irradiance. Some of these parameters are collected using a PV pyranometer located close to the PV systems, while other parameters are obtained based on the PV system properties. In the developed model, the input data are divided into three categories as below.

1. *PV system properties data:* The developed model considers the PV areas (A_d^{pv}) which is divided into 6 sub-areas with specific power capacity (P_d^{cap}) for each area $((A_1^{pv}, P_1^{cap}), (A_2^{pv}, P_2^{cap}), (A_3^{pv}, P_3^{cap}), (A_4^{pv}, P_4^{cap}), (A_5^{pv}, P_5^{cap}), (A_6^{pv}, P_6^{cap}))$. In which, (A_6^{pv}) represents the entire PV plant area while A^{pv} and P^{cap} are used to calculate the power density per unit area ($D_A^{cap} = P^{cap}/A^{pv}$). This division facilitates the application of the proposed model on various PV plants of different areas and capacities. The division is selected based on various parameters, such as the number of PV systems, their capacities and the layout of the area in which the PV systems are spread. The aim of this division is to improve the model's accuracy in estimating the generated power by each small area. The value of V_c can be calculated using the method described in [12]
2. *PV sensor data:* This includes Tmp and H_u that can be directly collected from the pyranometer sensor installed at the PV system site.
3. *WT data:* obtained using wavelet timescales (modes) analysis. When the pyranometer is installed with the same tilt and orientation as the PV systems, the

output irradiance of the PV sensor will be a plane of the array (*POA*) [45]. Let $POA[k]$ denote the output of the pyranometer, in which k is a time index.

To improve the ANFIS accuracy in estimating the output power, the time-series (*POA*) for each day is divided into five sub-time-series. This procedure enables the proposed technique to correlate the input and output variables at a specific period of the day more precisely due to the dynamic change in weather conditions over the day. For each sub-time series, the module output power can be calculated using (3.4) [76], in which the output of (3.4) is indicated by $P_{POA}^{mdu-in}[k]$.

The average of the total output power of the PV plant ($P_{Eq}^{mdu-out}[k]$) is calculated by assuming each module is providing the same share of the overall power. Then, $P_{POA}^{mdu-in}[k]$ and $P_{Eq}^{mdu-out}[k]$ are normalised by dividing each by $P_{POA}^{clr}[k]$ (output power under clear sky conditions), which can be calculated by replacing $POA[k]$ in (3.4) by POA_{clr} from (3.6).

Then, (4.13) and (4.14) can be derived as:

$$P_{nrm}^{POA-in}[k] = \frac{P_{POA}^{mdu-in}[k]}{P_{clr}^{POA}[k]} \quad (4.13)$$

$$P_{nrm}^{POA-out}[k] = \frac{P_{Eq}^{mdu-out}[k]}{P_{clr}^{POA}[k]} \quad (4.14)$$

Using (3.3), $P_{nrm}^{POA-in}[k]$ and $P_{nrm}^{POA-out}[k]$ can be decomposed into three timescale functions. Thus, WT_{mode-i} analysis of $P_{nrm}^{POA-in}[k]$ yields three high-frequency modes ($P_{mode-i}^{HF-in}[k]$) and three low-frequency modes ($P_{mode-i}^{LF-in}[k]$). In the same way, $P_{nrm}^{POA-out}[k]$ is analysed into three high-frequency ($P_{mode-i}^{HF-out}[k]$) and low-frequency ($P_{mode-i}^{LF-out}[k]$) modes. It is to be noted that from the inverse Wavelet transform, the first and second low-frequency modes can be ignored as they have no impact on the inverse wavelet process as explained in (4.15) below[97].

$$P_{nrm-out}^{Eq-ANFIS} = inv(P_{mode-1}^{HF-ANFIS}[k]) + inv(P_{mode-2}^{HF-ANFIS}[k]) + inv(P_{mode-3}^{HF-ANFIS}[k]) + inv(P_{mode-3}^{LF-ANFIS}[k]) \quad (4.15)$$

$P_{mode-i}^{HF-in}[k]$ and $P_{mode-i}^{LF-in}[k]$ along with the parameters mentioned in points 1 and 2 above will be the ANFIS input parameters, whereas $P_{mode-i}^{HF-out}[k]$ and $P_{mode-i}^{LF-out}[k]$

are the outputs, as shown in Table 4.1. In Table 4.1, the first and second columns represent the input and outputs of the ANFIS technique while the third column represents the ANFIS models after the training stages.

The results of the ANFIS modelling process comprise four ANFIS models, each one represents the ANFIS- WT model at a certain wavelet mode. For example, $P_{\text{mode-1}}^{\text{HF-ANFIS}}[k]$, $P_{\text{mode-2}}^{\text{HF-ANFIS}}[k]$ and $P_{\text{mode-3}}^{\text{HF-ANFIS}}[k]$ represent the equivalent wavelet analysis of mode-1, mode-2, and mode-3 for high-frequency signals, respectively, while $P_{\text{mode-3}}^{\text{LF-ANFIS}}[k]$ represents the equivalent wavelet analysis of mode-3 for a low-pass signal as shown in Table 4.1.

Then, by applying the inverse WT as in (4.15), the estimation of the equivalent normalised output power of a number of rooftop PV systems distributed within a certain area is evaluated.

Table 4.1 ANFIS input and output data and models after the training stage

Input parameters	Output	ANFIS models after training	
D_A^{cap} (could be for A_1^{pv} , A_2^{pv} , A_3^{pv} , A_4^{pv} , A_5^{pv} or A_6^{pv}), V_c , $Tmp[k]$, $H_u[k]$	$P_{\text{mode-1}}^{\text{HF-in}}[k]$	$P_{\text{mode-1}}^{\text{HF-out}}[k]$	$P_{\text{mode-1}}^{\text{HF-ANFIS}}[k]$
	$P_{\text{mode-2}}^{\text{HF-in}}[k]$	$P_{\text{mode-2}}^{\text{HF-out}}[k]$	$P_{\text{mode-2}}^{\text{HF-ANFIS}}[k]$
	$P_{\text{mode-3}}^{\text{HF-in}}[k]$	$P_{\text{mode-3}}^{\text{HF-out}}[k]$	$P_{\text{mode-3}}^{\text{HF-ANFIS}}[k]$
	$P_{\text{mode-3}}^{\text{LF-in}}[k]$	$P_{\text{mode-3}}^{\text{LF-out}}[k]$	$P_{\text{mode-3}}^{\text{LF-ANFIS}}[k]$

By multiplying the results of (4.15) by $P_{\text{clr}}^{\text{POA}}[k]$ and the number of PV modules (N_{module}), the equivalent output power seen by the entire PV plant can be estimated as follows:

$$P_{\text{out}}^{\text{Eq-ANFIS}} = P_{\text{nrm-out}}^{\text{Eq-ANFIS}} \times P_{\text{clr}}^{\text{POA}}[k] \times N_{\text{module}} \quad (4.16)$$

Fig. 4.2 illustrates the flowchart of the main steps of the ANFIS-WT technique proposed in this chapter.

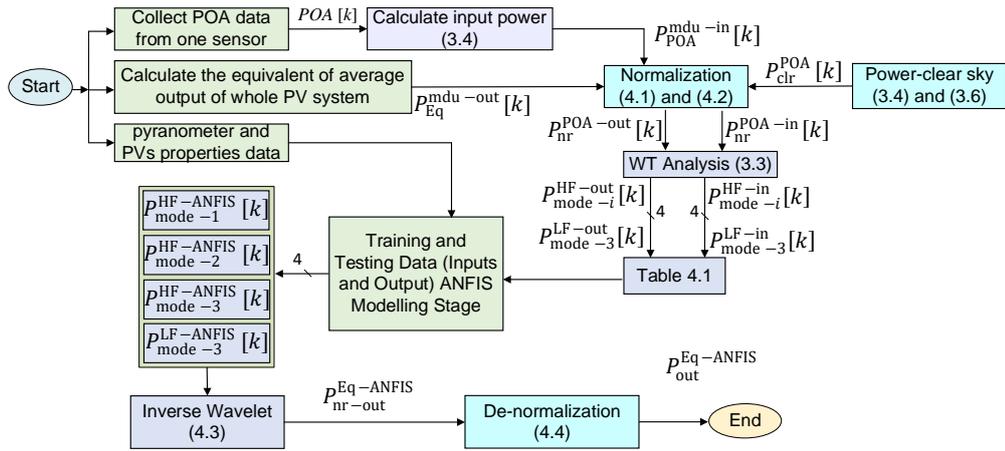


Fig. 4.2. The proposed methodology for the proposed ANFIS-WT technique

Because variability is the main concern of power operators and planners, the variability power index (V_{pi}) has been introduced as a mathematical tool to show the amount of power fluctuation at each WT mode. First, $P_{nr-out}^{Eq-ANFIS}$ is analysed into six timescales 2, 4, 8, 16, 32 and 64-min using the WT method. To calculate the power for each period, the wavelet periodogram has been used as below [49]:

$$I_w[k] = \frac{|P_{mode-i}^{HF-ANFIS}[k]|^2}{T^{mode-i}} \quad (4.17)$$

then,

$$V_{pi}(T^{mode-i}) = \frac{\int_0^{T_L} I_w[k] dk}{T_L} \quad (4.18)$$

where T_L is the length of the time data.

According to [49], VRI can be calculated using (4.19)

$$VRI = \frac{V_{pi}(T^{mode-i})_{py}}{V_{pi}(T^{mode-i})_{ave}} \quad (4.19)$$

4.3 Performance Evaluation of the proposed technique

Data from 445 days through 2016 and 2017 with different amounts of variabilities (days with a completely clear sky are eliminated) have been used to develop the proposed model. Data are collected from PV rooftop systems of 2.14 MW power capacity distributed within 0.7 km² at the University of Queensland in

Brisbane, Australia. The PV site is located at 153°29'44" east longitude and 27°00'30" south latitude. All PV systems face north at an inclination angle of 30° to produce maximum possible power. Brisbane city weather is described as humid, and warm all over the year. The temperature in summer and winter ranges between 21 to 35 °C, and 11 to 21 °C respectively, while it is 15 to 25 °C in autumn and spring. Brisbane's sky exhibits several more passing cloud events in summer than in winter [27].

The varied weather over the 4 seasons in Brisbane city and PV systems distributed around the University of Queensland makes it a suitable site for this study. PV sensor data have been collected from the PV located at site-7 in Fig. 4.3. The collected data from 445 days comprise small, medium and large levels of variability. For ANFIS training and testing data, 330 days have been used and the rest of the data are used in the validation stage. To improve the ANFIS performance during the training stage, each day is divided into five sub-time series, the first sub-time series is assumed to start at 7:00 am and ends at 8:59 am, the second sub-time series is from 9:00 am to 10:59, the third sub-time series is from 11:00 am to 12:59 pm, the fourth sub-time series is from 1:00 to 2:59 pm and the fifth sub-time series is from 3:00 pm to 5:00 pm. The times of sunrise and sunset have been taken into account in this division. Also, the PV plant was divided into six small areas, namely the whole PV plant (A_6^{pv}), $A_1^{pv} = 0.227 \text{ km}^2$ with 1.82MW, $A_2^{pv} = 0.37\text{km}^2$ with 1.9MW, $A_3^{pv} = 0.17\text{km}^2$ with 0.606 MW, $A_4^{pv} = 0.162\text{km}^2$ with 1 MW, $A_5^{pv} = 0.275\text{km}^2$ with 1.82 MW and $A_6^{pv} = 0.7\text{km}^2$ with 2.14 MW.

Following the methodology steps in Fig. 4.2 and using the input and output data in Table 4.1, the ANFIS model can be built and trained. Since the back-propagation algorithm is relatively slow [97], the hybrid-learning algorithm is adopted to train the proposed ANFIS model. Fig. 4.4 shows the error in estimating the output power when different MFs are employed. The results show that the gaussmf membership function is more accurate than other types after 100 epochs in training mode and hence it is employed in the ANFIS model.

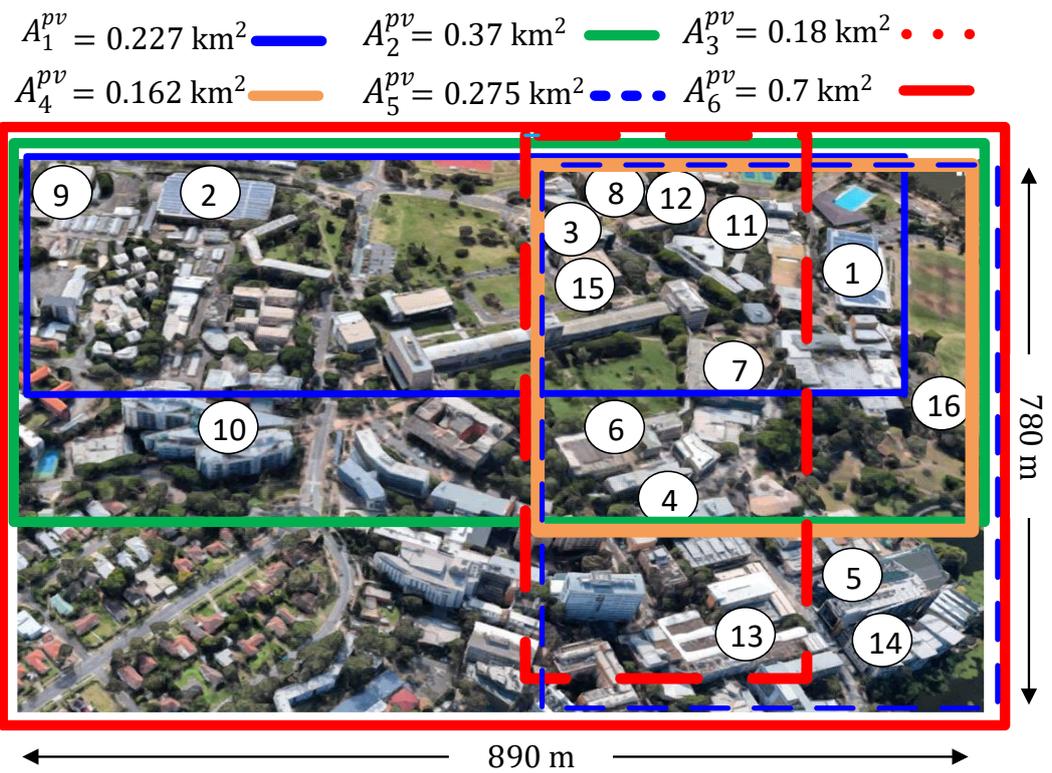


Fig. 4.3 Layout of the six PV plants under study

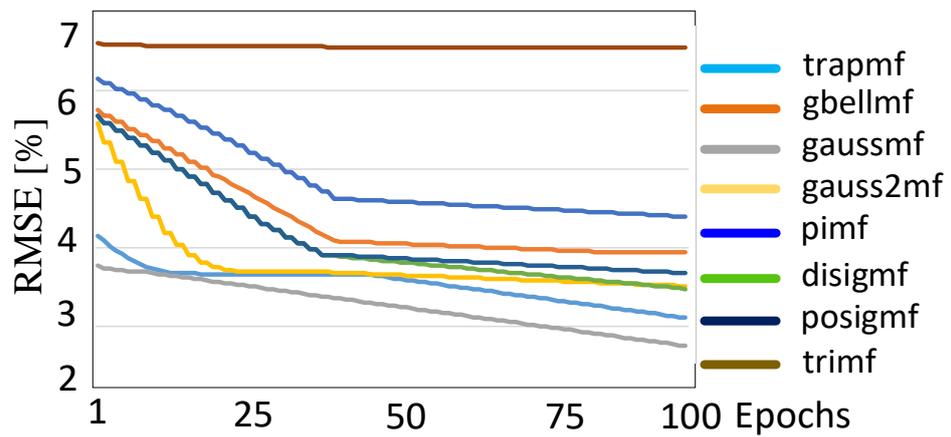


Fig. 4.4 Resulted in errors of different membership functions

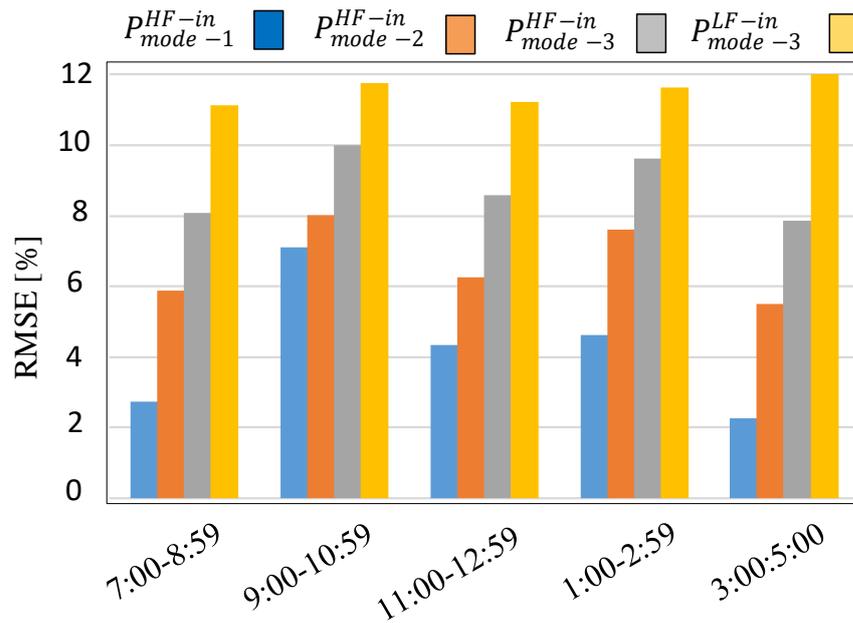


Fig. 4.5 Errors after 100 ANFIS epochs for each sub-time series

Fig. 4.5 shows the Root Mean Square Error (RMSE) for the last epoch at each ANFIS model and each sub-time series. Four ANFIS models have been developed as shown in Table 4.1, third column.

The model is validated using the remaining 115 days. Four signals ($P_{mode-1}^{HF-ANFIS}$ [k], $P_{mode-2}^{HF-ANFIS}$ [k], $P_{mode-3}^{HF-ANFIS}$ [k] and $P_{mode-3}^{LF-ANFIS}$ [k]) represent the results and by applying (4.15) and (4.16), the estimated PV systems output power can be obtained. Fig. 4.6 depicts the error in the estimated power of the developed model for 115 days using MAE and RMSE in comparison with other methods mentioned in chapter 2 such as Hoff, Perez, Lave as well as the GEP-VRI models from the next chapter (Chapter 5). Table 4.2 and Table 4.3 summarises the results of Fig. 4.6 using the average values of the MEA and RMSE, respectively, for each of the models along with the proposed technique in this paper. These tables also illustrate the minimum and maximum improvement values attained by the proposed technique when compared with existing models. It can be seen from these results that the proposed method is more accurate than other models published in the literature. Considering the MAE, the minimum improvements attained by the proposed technique for the five sub-time series are 11.4, 8.9, 7.5, 9.0 and 10.4; respectively whereas the maximum improvements for the same sub-time series are

36.9, 28.6, 24.1, 28.81 and 33.1; respectively. Similarly, the results reveal a very similar tendency in the enhancement attained by the proposed technique when the RMSE metric is considered.

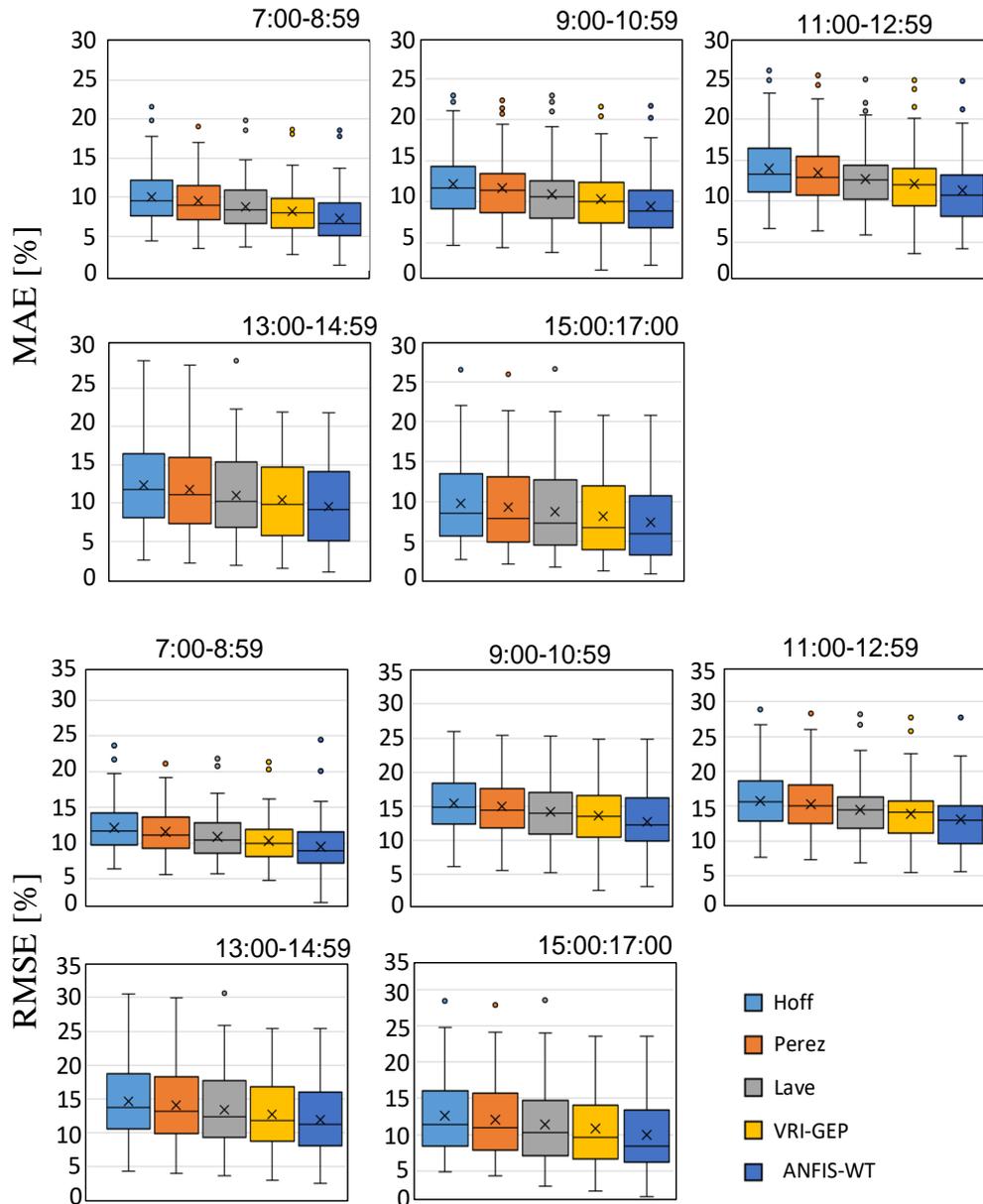


Fig. 4.6 Comparison of the MAE and RMSE of the proposed ANFIS-WT model and other models in the literature

The estimation improvement strategy adopted in this study is based on dividing the time-series data into several sub-time series (five sub-time series in this study) along with employing the wavelet transform method. ANFIS is employed to accurately

map the relationship between the input and output data set. Dividing the time-series data into five segments helps the ANFIS algorithm map the correlation between the input and output data rapidly and precisely due to the fact that, within each selected period, the environmental conditions such as temperature, zenith angle and output power won't change much. Likewise, the division conducted by the Wavelet transform method on each sub-time series into low and high-frequency bands reduces the ANFIS confusion during the training process to map the relationship between input and output data. The results of the proposed technique in this chapter reveal its high accuracy when compared with other estimation models published in the literature. The results in Fig. 4.6, Table 4.2 and Table 4.3 also show a comparison of the model proposed in this chapter and our recent GEP-VRI technique [98]. It can be observed that both GEP and ANFIS techniques provide more accurate results than other techniques in the literature [99, 100]. Ref [98] employed the GEP technique to develop a VRI model, whereas, in this chapter, an ANFIS-WT model is used to estimate the output power directly without calculating the VRI as per the strategy of all existing models in the literature. This reduces the calculation complexity and improves the estimation accuracy.

Table 4.2 Comparison of MAE [%] average values for different VRI models including the proposed model in this chapter

Model	7:00-8:59	9:00-10:59	11:00-12:59	13:00-14:59	15:00-17:00
	MAE	MAE	MAE	MAE	MAE
Hoff	10.1	12.2	13.9	12.3	9.8
Perez	9.5	11.6	13.4	11.8	9.3
Lave	8.8	10.9	12.7	11.0	8.7
GEP-VRI	8.2	10.3	12.1	10.4	8.1
ANFIS-WT	7.2	9.5	11.3	9.6	7.4
Minimum improvement	11.4	8.9	7.5	9.0	10.4
Maximum improvement	36.9	28.6	24.1	28.8	33.1

Table 4.3 Comparison of RMSE [%] average values for different VRI models including the proposed model in this chapter

Model	7:00-8:59	9:00-10:59	11:00-12:59	13:00-14:59	15:00-17:00
	RMSE	RMSE	RMSE	RMSE	RMSE
Hoff	12.1	15.5	15.8	14.6	12.6
Perez	15.8	14.9	15.2	14.1	12.0
Lave	10.9	14.2	14.5	13.4	11.4
GEP-VRI	10.2	13.6	13.9	12.8	10.8
ANFIS-WT	9.5	12.8	13.0	11.9	10.0
Minimum improvement	7.6	6.6	6.4	7.1	7.8
Maximum improvement	27.3	21.2	20.8	22.7	25.8

The results show that the ANFIS model proposed in this chapter attains a minimum and maximum improvement of 6.4% and 11.4% when compared with the GEP-VRI model by considering the MAE metric. The same results are achieved using the RMSE metric.

Ultimately, to calculate the VRI parameter, the output power has been analysed into six wavelet modes using (3.3) then the wavelet periodograms factor in (4.17) is employed at each wavelet mode. Using (4.18), the amount of power involved in the variability $V_{pi}(T^{\text{mode}-i})$ at each mode and each sub-time series can be evaluated. Fig. 4.7 shows an example for one day, 3 Dec 2017, that is chosen to explain this step. The blue line in Fig. 4.7 represents $V_{pi}(T^{\text{mode}-i})$ for one PV system (No.7), while the black and red lines represent $V_{pi}(T^{\text{mode}-i})$ for the actual and predicted output power for 16 PV systems, respectively. From these plots, the distance between the blue line (one PV system) and black and red lines (16 PV systems) characterises the geographic smoothing or the VRI value. The VRI can be calculated using (4.19) as shown in Fig. 4.8. These figures attest that VRI can be changed during the day. For example, the value of the VRI at mode-1 is 8.2 at sub-

time series 7:00–8:59 am, while it is 6.2, 7, 7, 7.5 at sub-time series 9:00–10:59 am, 11:00–12:59 pm, 1:00–2:59 pm and 3:00–5:00 pm; respectively. This is attributed to the change in the ambient parameters affecting the VRI such as V_c , Tmp , and H_u during the day.

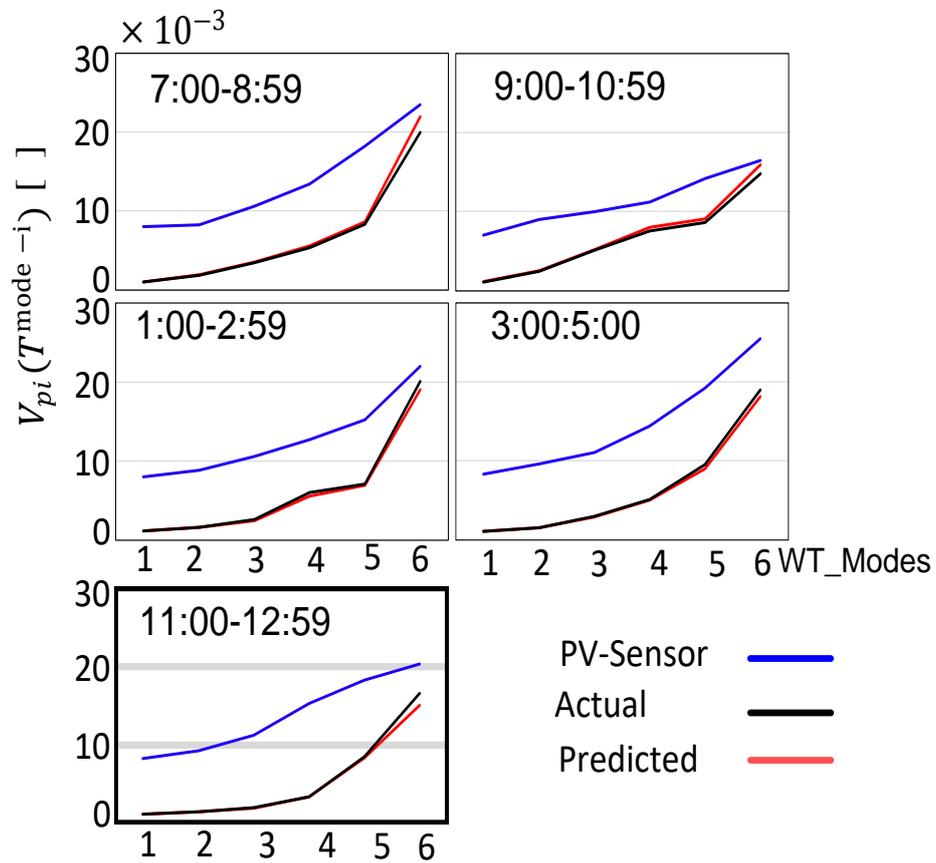


Fig. 4.7 V_{pi} for an individual PV system (blue), entire actual PV systems (black) and predicted (red)

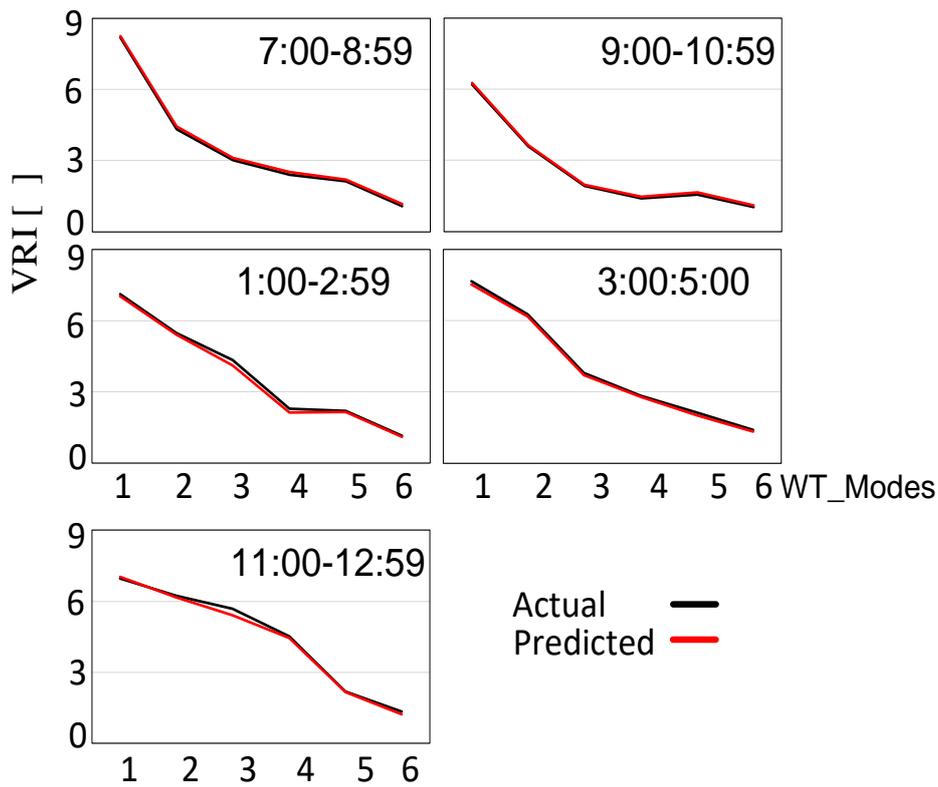


Fig. 4.8 The actual (black) and predicted (red) VRI values, 3 Dec 2017

4.3.1 Sensitivity Analysis

Following the validation of the proposed model’s accuracy in the above section, this section is aimed at investigating the sensitivity of the parameters used in the proposed model. In this regard, each parameter is changed within an expected range as listed in Table 4.4, while maintaining all other parameters unchanged [101]. The effect of such a change on the output power is analysed using the standard deviation (SD) metric. To standardise the model sensitivity analysis, SD is divided by the average value of the measured parameter. Fig. 4.9 shows the impact of each parameter for different sub-time series. It can be seen that the input power parameter has a more significant impact than the ambient temperature while humidity, cloud speed and the ratio of the PV plant capacity to plant area have approximately the same impact. Table 4.4 illustrates the minimum and maximum values for each parameter considered during the sensitivity analysis.

Table 4.4. Minimum and maximum values for each model parameter

Parameters	Minimum	Maximum
$P_{\text{norm}}^{\text{POA-in}} [k]$	1	0
$H_u [k]$	10	95
V_c	1	13
$Tmp [k]$	8	36
D_A^{cap}	3	8

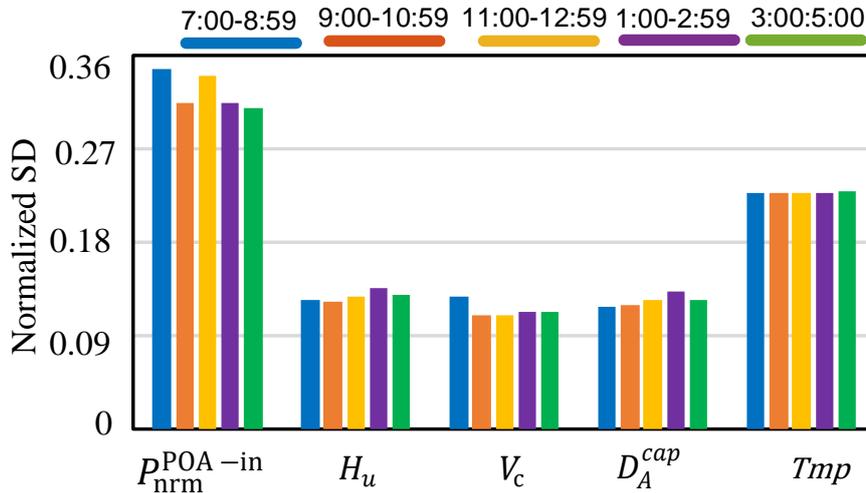


Fig. 4.9 . Sensitivity analysis for each model parameter for each sub-time series

4.4 Summary

This chapter has investigated a new method that used Wavelet-based and ANFIS modelling approaches to improve the new PV system output power estimation models. The developed models are employed to estimate the generated power of a group of PV systems distributed over 0.7km² using irradiance data from one PV sensor located near the PV system. Parameters such as PV area, capacity, cloud speed, temperature and humidity are considered as the inputs to the proposed model. Real data collected from the PV system in Brisbane, Australia are used to train the proposed model. The results reveal the high accuracy of the developed model when

compared with existing estimation models in the literature. The proposed model demonstrated 6.4 and 11.4% as minimum and maximum improvement levels when compared with the GEP-VRI model by employing the MAE approach. The sensitivity analysis of each parameter reveals that the output power collected from the PV sensor has the largest impact on estimating the output power of the entire PV system using the developed model. The impact of ambient parameters – humidity, cloud speed and the power capacity of the overall PV system – are similar and all have less impact on the estimated power.

Chapter 5 GEP-Based Approach

This chapter introduces an artificial intelligence technique, called GEP, to improve two models for estimating the total power generated by a group of neighbouring PV systems, spread over a distribution network using a single pyranometer for measuring the solar irradiance. The first model is employed to estimate a VRI value by correlating the distribution of the PV systems and the irradiance measured by the pyranometer. The model is then used to estimate the total power generated by the PV systems. While the second model is directly used to estimate the total generated power. Both proposed techniques consider the geographic variability reduction features and employ the WT technique to develop the proposed model and estimate the generated power. The effective performance of the proposed models are validated using real data collected by the Solar Project at the University of Queensland in Brisbane, Australia. The results reveal that the proposed techniques yield more accurate results when compared with other existing approaches in the literature. However, the model that estimated the total generated power directly is more accurate than the one that estimated the VRI value. Additionally, a sensitivity analysis has been carried out to investigate the impact of the employed parameters in the new models.

5.1. GEP Technique

This chapter aims to further enhance the accuracy of the output power estimations in the previous stage by using artificial intelligence or evolutionary programming techniques. Artificial intelligence techniques such as ANN and ANFIS, mentioned in chapter 4, are becoming significant machine tools to simulate, model and find correlations between non-linear data. Similarly, Genetic algorithm (GA), Genetic programming (GP), and GEP are powerful evolutionary computer programming techniques that can be used for the same purposes. Such techniques have been employed for various power systems applications [102-105].

The GEP technique is considered in this chapter, since it aims at developing two mathematic models: first, for VRI without needing the other correlation models (Hoff, Perez, Lave and, ACM), and second, directly for estimating the total

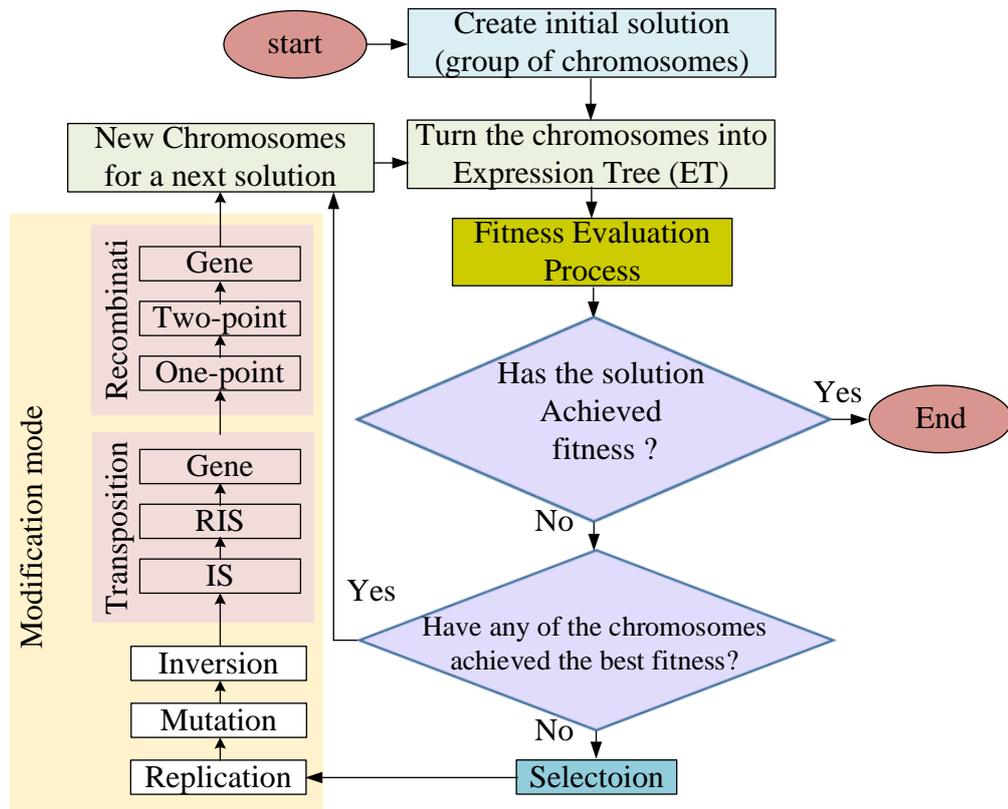


Fig. 5.1 Flowchart of the GEP algorithm

generated power.

The GEP technique is inspired by two early genetic techniques – GA and GP. Similar to these two techniques, the GEP technique selects solutions (individuals or chromosomes) based on several metrics and presents solutions at each stage based on a number of genetic operators [106]. However, while the solutions in the GA are of fixed length and linear strings and in the GP approach are of non-linear shapes and sizes, the solutions in the GEP are of fixed length, non-linear entities, different sizes and shapes, which combines the benefits of the GA and GP approaches [107, 108]. The main characteristic of the GEP technique is its ability to describe a mathematical model between dependent and independent parameters with high accuracy [109]. GEP is a powerful evolutionary technique that can be used to construct a mathematical model by training non-linear data to find the correlation between the considered input and output parameters [106, 110]. This technique can analyse, estimate, forecast and optimise the provided data set [111]. Fig. 5.1 shows a general flow chart for the GEP algorithm, which starts with generating random

solutions and evaluating their fitness. Each solution is modified using several genetic operators that change the form of the solution by changing the position of the codes (mutation operator), reversing the direction of the solution (inversion operator) or/and changing two segments between two different solutions (recombination operator). The process is continued until the best solution (a model with the least errors) is determined [108]. The outcomes of the GEP algorithm can be in three formats – data series, sub-expression trees (ETs) and mathematical formats, which are shown in the examples in Table 5.1.

Table 5.1. Creating an Initial Solution using a GEP algorithm

Function	Example												
Randomly created population or solution (a group of chromosomes) consists of a head and tail, contains some parameters of functions set (e.g. +, -, *, /, AND, OR etc.) and terminals set (variables and constants), while the tail contains only some terminals set.	<div style="text-align: center;"> <table border="0" style="margin: 0 auto;"> <tr> <td style="text-align: center;">Head</td> <td style="text-align: center;">Tail</td> <td></td> </tr> <tr> <td style="text-align: center;">0 1 2 3 4</td> <td style="text-align: center;">5 6 7</td> <td style="text-align: center;">0 1 2 3 4 5 6 7</td> </tr> <tr> <td style="text-align: center;">Q / + - a c a b</td> <td style="text-align: center;">Q * + + c b a a</td> <td></td> </tr> <tr> <td colspan="3" style="text-align: center;">Chromosome</td> </tr> </table> <p style="text-align: center;">data series format</p> <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;"> <p>sub-expression tree format</p> </div> <div style="text-align: center;"> <p>sub-expression tree format</p> </div> </div> $\sqrt{(a + c) \div (a - b) + \sqrt{(c + b) * (a + a)}}$ <p style="text-align: center;">Mathematical format</p> </div>	Head	Tail		0 1 2 3 4	5 6 7	0 1 2 3 4 5 6 7	Q / + - a c a b	Q * + + c b a a		Chromosome		
Head	Tail												
0 1 2 3 4	5 6 7	0 1 2 3 4 5 6 7											
Q / + - a c a b	Q * + + c b a a												
Chromosome													

The key stages of the GEP process are explained below:

First, an initial population (i.e. a group of chromosomes) is created randomly to represent the first suggested solution. Each chromosome consists of a head and a

tail in which the head comprises functions (e.g. +, -, ×, ÷, AND, OR) and terminals set (i.e. variables and constants), whereas the tail involves only terminals set. The solution is also expressed in sub-expression trees (see Table 5.1)

In the second stage, functions are used to measure how much the solution (the chromosomes) satisfies the fitness function (i.e. how closely the solution matches the actual output power measured from the PVs in the case of this study). GEP uses a few fitness function metrics, such as number of hits, precision and selection range, root mean square error (RMSE), and R-square. The fitness process is applied to each chromosome and rates them based on these metrics. The GEP selects the chromosomes according to a roulette-wheel rule in which each chromosome is given a slice based on the fitness process. Then, the roulette-wheel spins several times equal to the number of chromosomes. Finally, the largest slice is most likely to be selected, Table 5.2 illustrates this stage.

Table 5.2. Second stage: Selecting the best solutions

Function	Example																																				
Choose a number of chromosomes which succeed in the fitness test using the roulette-wheel rule; each chromosome is given a slice based on the fitness process then the roulette-wheel is spun several times equal to the number of chromosomes so the largest slice is more likely to be chosen.	<table border="0"> <tr> <td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td> <td>Fitness values</td> </tr> <tr> <td>Q/+-acdb</td><td>Q*++cbaa</td><td>-[1]=3</td> <td rowspan="6"> </td> </tr> <tr> <td>*/+dacab</td><td>*Q+ccbaa</td><td>-[2]=5</td> </tr> <tr> <td>*ad+/cab</td><td>*Q+ccbaa</td><td>-[3]=2</td> </tr> <tr> <td>Q/+-acab</td><td>Q*--cbba</td><td>-[4]=1</td> </tr> <tr> <td>+/Q/acab</td><td>Q*++cbca</td><td>-[5]=6</td> </tr> <tr> <td>Qc+-acdb</td><td>Q*++cdaa</td><td>-[6]=8</td> </tr> </table> <p>Chromosome No. 6 has the largest fitness value which equals 8</p>	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7	Fitness values	Q/+-acdb	Q*++cbaa	-[1]=3		*/+dacab	*Q+ccbaa	-[2]=5	*ad+/cab	*Q+ccbaa	-[3]=2	Q/+-acab	Q*--cbba	-[4]=1	+/Q/acab	Q*++cbca	-[5]=6	Qc+-acdb	Q*++cdaa	-[6]=8
0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7	Fitness values																					
Q/+-acdb	Q*++cbaa	-[1]=3																																			
*/+dacab	*Q+ccbaa	-[2]=5																																			
*ad+/cab	*Q+ccbaa	-[3]=2																																			
Q/+-acab	Q*--cbba	-[4]=1																																			
+/Q/acab	Q*++cbca	-[5]=6																																			
Qc+-acdb	Q*++cdaa	-[6]=8																																			

The selected chromosomes are then passed to the modification pattern. This pattern consists of several genetic operators to make appropriate changes to the selected chromosomes. The modification pattern consists of several operators that

work to make simple or essential changes depending on the type of operator to change the format of the chromosomes (solutions) to find an appropriate solution. The operators, which make minor changes, are replication, mutation and inversion operators. Minor changes include changing one or more codes and pasting them in another place in the same chromosome, more details are listed in Table 5.3.

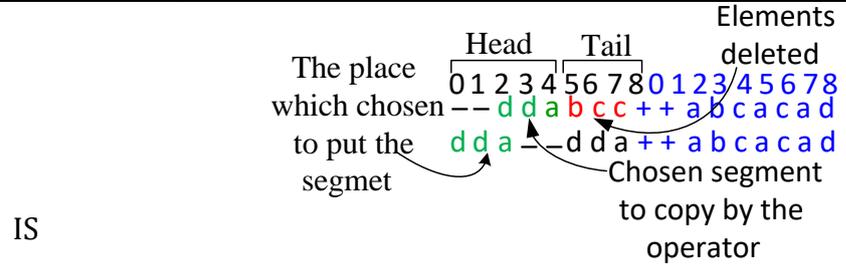
Table 5.3. Modification pattern: Replication, Mutation and Inversion operators and some work details and examples

Function	Example
Replication Operator: copies the selected chromosomes which are passed from the fitness step and passes them into modification mode	
Mutation Operator: change one or more than one code in the head or tail of the chromosome, randomly	<p>0123456701234567</p> <p>Q/+-acdbQ*++cbaa-[1]=3</p> <p>Qc+-acdbQ*++cdaa-[8]=5</p> <p>Represent chromosomes No. 1 and 8, code No. 5 (green colour (b)) in the second gene subjected to mutation operator to change into d code</p>
Inversion operator: work only in the head area. It works to select randomly part of the head of the gene then returns it to the same place after reversing all its parameters.	<p>0123456701234567</p> <p>*/+dacab*Q+ccbaa-[2]=5</p> <p>*ad+/cab*Q+ccbaa-[3]=2</p> <p>Represent chromosomes No. 1 and 8, code No. 5 (green colour (b)) in the second gene subjected to mutation operator to change into d code</p>

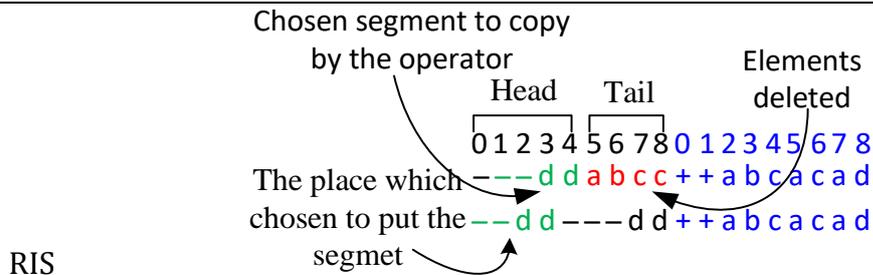
The other operators that cause major changes are Transposition Operators (IS, RIS and whole gene operators). These operators randomly select and copy a group of codes in the solution (serial codes) and paste them in the chromosome head (for more details see Table 5.4).

Table 5.4 Modification Pattern: Transposition Operators

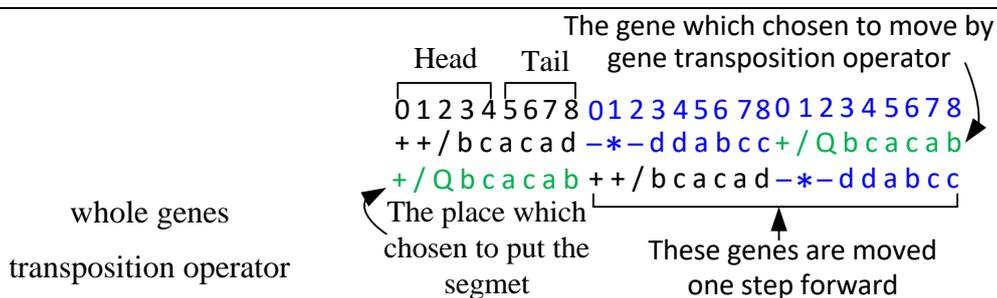
Function	Examples
----------	----------



Copy a small section from the chromosome randomly and put it in any location in the head of the gene except at the beginning of the chromosome (the root of sub-ET)



Copy a segment from a target gene that should start with a function element for another random position in the gene, then locate this segment in the first part of the gene so that the root of this gene is always a function element



Moves the whole gene to the beginning of the chromosome in which, in contrast to IS and RIS operators, the moved segment will be deleted from its original place

Table 5.5 Modification Pattern: Recombination Operators

Function	Examples
One-point Recombination	<p>Points are chosen by One point Recombination</p> <p>Parents</p> <p>Offsprings</p> <p>Change the position by the operator</p> <p>The operator chooses one point on each of the two chromosomes randomly and exchanges them to make two new chromosomes</p>
	<p>Two-point Recombination</p> <p>Parents</p> <p>Offspring</p> <p>The segment between two points chosen by two points operator</p> <p>Two points will be chosen by this operator on two chromosomes randomly. Then the segment between these points on the parent are exchanged with each other.</p>
Gene Recombination	<p>Gene Recombination</p> <p>Parents</p> <p>Offspring</p> <p>The genes which chosen by Gene Recombination operator to exchange between the parent</p> <p>The whole gene will be exchanged between two chromosomes (parent). The result will be two new chromosomes (offspring)</p>

The second type of this group of operators is the Recombination operators (OPR, TPR, and GR operators), which randomly select two solutions (parent) and exchange some codes between them to create two new solutions (offspring), see Table 5.5 for more details. All these operators randomly apply appropriate modifications to the chromosomes. At the end of the modification stage, new

chromosomes are introduced and they will be passed again to the fitness function to select the best solution.

5.2. GEP-VRI Model

This section presents a new VRI model by employing the GEP technique detailed below:

5.3. Proposed Methodology

Using the VRI calculated in (2.10), the total power generated by a group of distributed rooftop PV systems can be estimated based on the solar irradiance data of one pyranometer using only the WT function [45, 112]. This includes several stages, as summarised below:

Step 1

First, the solar irradiance data of the pyranometer are collected. The measured data include the irradiance incident on the plane of the array (POA), instead of the global horizontal irradiance (GHI) to further increase the accuracy of the estimation [45].

Step 2

The measured POA data are normalised ($POA_{nrm}[k]$) against the clear sky condition using (3.5)

Step 3

$POA_{nrm}[k]$ can then be decomposed into several modes (i.e. 6 modes in this study to denote timescales of 2, 4, 8, 16, 32 and 64 minutes) using the Wavelet transform in (3.3).

Step 4

Finally, as in (3.4), the total power generated by the number of PV modules, each with a cross-section of A_m , and experiencing an ambient temperature of $Tmp[k]$ is calculated from:

$$P_{\text{module}}^{\text{POA}}[k] = 0.12 A_m \times N_{\text{module}} \times POA^{\text{equ}}[k] \times \left[1 - 0.004 \left(Tmp[k] + \frac{0.32 \beta \times POA^{\text{equ}}[k]}{8.91 + 2 V_c[k]} - 25 \right) \right] \quad (5.1)$$

in which

$$POA^{equ}[k] = POA_{clr}[k] \times \sum_{i=1}^6 \frac{WT_{mode-i}[k]}{\sqrt{VRI_{mode-i}^{max}}} \quad (5.2)$$

The proposed technique consists of two stages of training the GEP model and testing it after being educated. These stages are discussed below:

The considered input parameters that impact of the VRI are the number of the PV systems (N_{system}), the area covered by these systems (A_d^{pv}), cloud speed (V_c) and timescale (T_{mode-i}). These parameters are commonly used in VRI calculations in the existing literature as discussed in Chapter 2. In addition, the ambient temperature (Tmp) and humidity (H_u) have also been considered in the input parameters.

To facilitate the application of the proposed model for PV systems in different areas and different sizes, the area under study should be divided into several sub-areas. This area division depends on various factors, such as the number of PV systems, their capacities and the area of land across which the PV systems are distributed. The aim of this division is to create a reasonable number of areas with various shapes and different numbers of PV systems and capacities to increase the ability of the model to estimate the generated power on the basis of the PV system with diversified PV numbers, capacities, areas, etc.

Having the input and output variables as discussed above, the system model can be formulated by setting and training the GEP technique. The settings of the GEP technique include:

- Choosing a fitness function. The fitness function should present the error of the power estimation in comparison with the real (measured) values. Various fitness functions can be employed such as the RMSE, MAE and R-square as below [2, 28]

$$RMSE = \sqrt{\frac{\sum_{j=1}^{N_{data}} (P_{total}[k])^j - P_{act}[k])^j)^2}{N_{data}}} \quad (5.3)$$

$$MAE = \frac{\sum_{j=1}^{N_{data}} |P_{total}[k])^j - P_{act}[k])^j|}{N_{data}} \quad (5.4)$$

$$R^2 = 1 - \frac{\sum_{j=1}^{N_{\text{data}}} (P_{\text{total}}[k]^j - P_{\text{act}}[k]^j)^2}{\sum_{j=1}^{N_{\text{data}}} (P_{\text{total}}[k]^j - P_{\text{act}}^{\text{avg}}[k]^j)^2} \quad (5.5)$$

where N_{data} is the number of the power data set in a day (between sunrise and sunset), $P_{\text{act}}[k]$ is the real (measured) total power generated by all PV systems during this period, while $P_{\text{act}}^{\text{avg}}[k]$ is its average value.

- Determining the terminal and function elements: This study has 6 independent variables (inputs) and one dependent output variable, as summarised in **Error! Reference source not found.**

- Selecting the GEP's chromosome details (i.e. the number of chromosomes, head length and the number of genes and the linking function of the genes) and the type and rate of the genetic operators. More details are provided in Table 5.6.

After determining the settings, the GEP technique can be trained by providing a sufficient amount of real input and output data. The training stage is deemed finalised when a negligible error is achieved between $P_{\text{act}}[k]$ and the determined power (denoted by $P_{\text{total}}[k]$) in all timescales.

After the training of the GEP model has been finalised, it can be tested by employing it to calculate the total power generated by the group of PV systems at the considered site. To this end, $VRI_{\text{mode}_i}^{\text{GEP}}$ will replace VRI in (5.2) to determine a more accurate $POA^{\text{equ}}[k]$.

The flowchart in Fig. 5.2 summarises all the above steps.

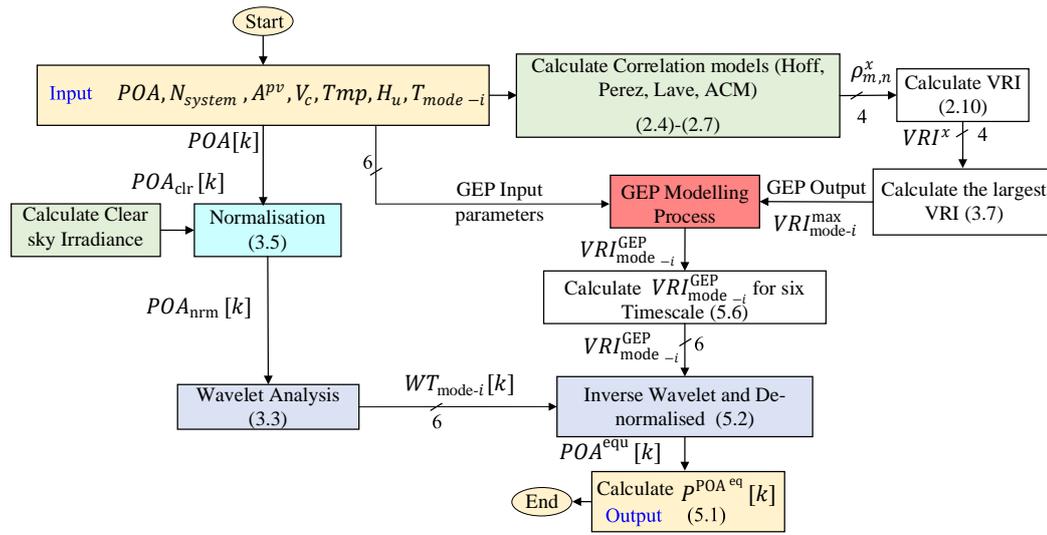


Fig. 5.2. Flowchart for the calculation of VRI_{mode-i}^{GEP} using GEP and estimating the overall power of a small-scale PV plant

5.3.1. Performance Evaluation

To evaluate the performance of the proposed technique, the 16 PV systems distributed across 0.7 km² at the St Lucia campus of the University of Queensland in Brisbane, Australia, as illustrated in Fig. 4.3 is considered in this study. Table 3.3 lists the specifications of these systems [27]. It is assumed that the pyranometer is installed next to PV system number 7.

The total area is divided into 6 sub-areas, each containing less than 9 PV systems of different capacities, as showing in Fig. 4.3. The developed GEP model is trained using the measured POA next to PV system-7 for 220 days in 2017 in a per minute range. The selected days covered various ambient conditions and seasons. The numbers of chromosomes, head length and the number of genes in the proposed GEP model are set at 30, 10 and 5, respectively.

The independent genes are linked through an addition function. The data of the developed GEP model is shown in Table 5.6. This table also shows the assumed types of genetic operators employed and their numerical values. The developed empirical-based model is tested by employing another set of data collected on different days. This testing stage did not result in any significant change between the predicted and measured power. The developed VRI equation on the basis of the

GEP model for the system under study is formulated in (5.6).

Table 5.6 Data for the GEP model

Number of chromosomes		30	
Head size		10	
Number of Genes		5	
Linking function		Addition	
Genetic Operator	rate	Genetic Operator	rate
Mutation	0.044	One-point recombination	0.3
Inversion	0.1	Two-point recombination	0.3
IS Transposition	0.1	Gene recombination	0.1
RIS Transposition	0.1	Gene transposition	0.1

It must be noted that there is no straight way to derive this equation and this is the fundamental reason that the GEP technique has been employed in this study, as this technique can derive such an empirical-based mathematical model based on the given input and output data.

The trained GEP is subsequently validated by employing it to determine the total power generation over 90 days of various ambient conditions during 2016. Fig. 5.3 illustrates the MAE and RMSE of the estimated POA using the developed GEP technique when compared with the real value over different days. In Fig. 5.3, the performance of the proposed technique is compared with the results that are revealed by the Hoff, Perez and Lave correctional coefficient models. Table 5.7 lists these errors as well as the minimum and maximum in the estimation accuracy improvement using the proposed technique. This table also shows the percentage of the studied days that each method has an accuracy of over 90%. As seen in Fig. 5.3 and Table 5.7, the proposed technique has the least errors compared with the other existing approaches.

$$\begin{aligned}
 \text{VRI}_{\text{mode-}i}^{\text{GEP}} = & \exp\left(-\frac{3 V_c^4}{2} - \left(1 + \frac{3 V_c}{2}\right) - A^{pv}\right) \\
 & + \exp\left(\sqrt{N_{\text{system}}} - \sqrt{(0.263 \log(0.67 T_{\text{mode-}i})^3 - 0.67 V_c)}\right) \\
 & + A^{pv} + \frac{T_{mp} \times N_{\text{system}}}{(1.11 T_{\text{mode-}i}) + T_{mp} + (1.5 V_c \times T_{mp})} \quad (5.6) \\
 & + ((11.12 - A^{pv}) - (4.46 \log(1.5 V_c))) \times \frac{N_{\text{system}}}{T_{\text{mode-}i} - 0.67 H_u} \\
 & + \left(\frac{(2 N_{\text{system}}^2 - 17.7 N_{\text{system}})}{\exp(N_{\text{system}})}\right)^{1/9}
 \end{aligned}$$

For example, considering the MAE metric, the maximum and minimum improvements using the proposed technique are respectively 43.7 and 8%, while the resulting improvements when using the RMSE metric are 51.4 and 8.4%, respectively. The proposed technique has also been able to estimate the total power for 82.3% of the studied days at an error rate of less than 10%, while this figure for the other approaches is between 57.6 and 67%. The reduction of the errors in the estimations using the proposed method is mainly because of the use of an artificial intelligence-based technique on top of existing empirical methods, which overtime tries to reduce the errors.

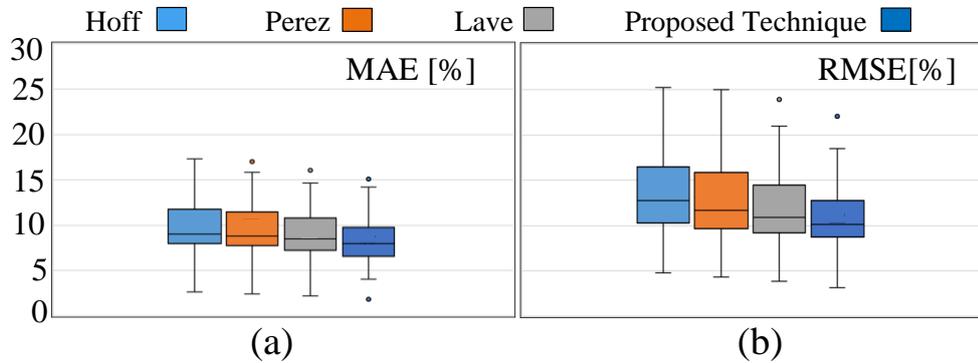


Fig. 5.3. MAE and RMSE for different techniques estimating the total power generated by the considered PV systems

Table 5.7 The improvement using the proposed technique over other models

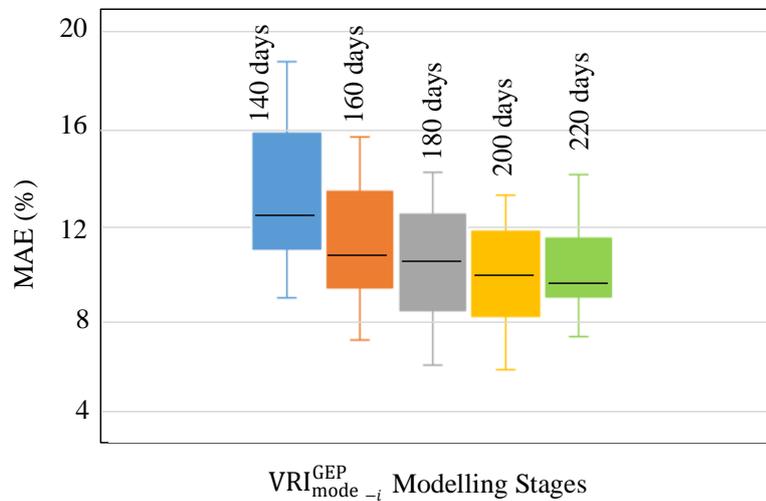
Models	MAE[%]		RMSE[%]		Standard deviation	Percentage of days that have accuracy more than 90%.
	Min	Max	Min	Max		
Hoff	2.6	17.3	4.8	25.2	3.2	57.6
Perez	2.4	17.0	4.4	25.0	3.1	62.3
Lave	2.2	16.3	3.9	23.9	2.9	67.0
VRI_{mode-i}^{GEP}	1.8	15.1	3.2	22.1	2.5	82.3
Maximum Improvement[%]	43.7	14.8	51.4	14.2		
Minimum Improvement[%]	19.1	8	21.6	8.4		

One important criterion for achieving highly accurate results using the proposed technique is ensuring proper training. Fig. 5.4 and Table 5.8 show the variation in the results of the proposed technique when a data set of 140 to 220 days is used in steps of 20. As seen from this figure and table, a data set of 220 has resulted in an acceptable level of accuracy and as such, is deemed to be suitable for finalising the training stage in the proposed technique.

The error in the method in this chapter is calculated as the difference between the powers recorded by the smart meters against that estimated by the proposed technique. In general, the smaller the error, the more accurate the proposed power estimation approach. For over 82% of the studied days, the estimation error of the proposed model was below 10%. It must be noted that all the existing and proposed methods are empirical, and as is expected from any empirical method, there will be errors if the employed assumptions are varied throughout the study period or if the assumptions are slightly different than the real values. Therefore, on those days that the error was above 10%, the error can be attributed to a change in the various assumptions used in the proposed approach, such as the change in cloud speed throughout the day (because in the considered study, cloud speed over a day is assumed to be constant).

Table 5.8. Various stages of the GEP model training

Modelling stage days	RMSE Average (%)	MAE Average (%)	R ²
140	14.2	12.55	0.80
160	12.02	10.62	0.83
180	11.21	10.5	0.84
200	9.8	10.1	0.87
220	9.5	9.8	0.90

Fig. 5.4. Evaluation of modelling stages of the VRI_{mode-i}^{GEP} model

Also, wind direction (cloud direction) can change in a day which may cause an error (again because in the proposed approach, it was assumed that wind/cloud direction remains the same throughout a day). These assumptions are in line with such assumptions used in the literature on power estimation

5.3.2. Sensitivity Analysis

In the above studies, the standard values for the input parameters of the developed empirical model are 25°C (ambient temperature), 50% (humidity), 0.5 km² (for a group of PV systems in an area), 16 (number of PV systems), 2-min (timescale), and 2 m/s (cloud speed). This section discusses the influence of each considered input parameter when determining VRI_{mode-i}^{GEP} using the proposed

technique.

The sensitivity of the developed model for each influencing parameter can be measured by varying the considered parameter within an acceptable range while all other input parameters are kept unchanged. For example, to evaluate the sensitivity of the proposed technique against variations in the ambient temperature, this parameter is changed between 5 to 36 °C. In the same way, humidity is varied from 5 to 95%, the area and number of all PV systems are varied respectively from 0.1 to 0.7 km² and from 9 to 16. Likewise, the timescale of the Wavelet model is varied from 2 to 64 minutes while cloud speed is varied from 0.5 to 10 m/s.

Fig. 5.5 illustrates the sensitivity analysis results for the parameters used to calculate VRI_{mode-i}^{GEP} . It can be seen from Fig. 5.5a that Tmp has a slightly linear impact on the model, resulting in an increase from 11 to 14.8 for VRI_{mode-i}^{GEP} when it increases from 5 to 36 °C. On the other hand, increasing the humidity factor results in an exponential increase from 12 to 19 when the humidity is changed from 5 to 95%, as seen in Fig. 5.5b. It should be highlighted that the Tmp and H_u parameters were neglected in the calculation of VRI models in the literature but

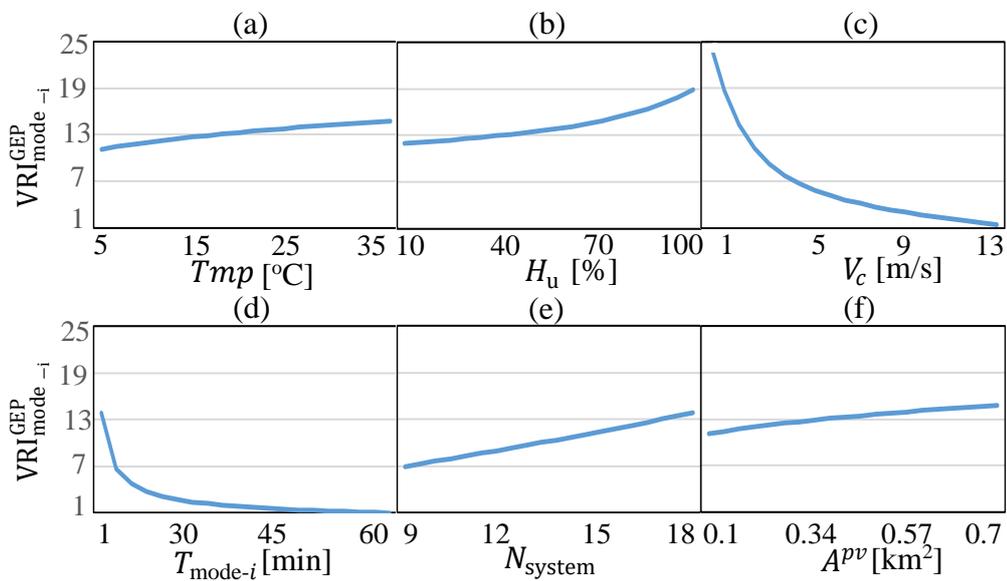


Fig. 5.5. Sensitivity of the VRI_{mode-i}^{GEP} model for each model parameter

have been considered in this work for the first time. This is one of the reasons for the improved accuracy of the power estimation by the proposed model.

The parameters which demonstrate the highest impact on the proposed calculation approach are cloud speed and Wavelet model timescale. Fig. 5.5c and Fig. 5.5d demonstrate a very sharp (exponential-type) variation in the output of the proposed technique after an increase in these two parameters. As such, more accurate results will be achieved by employing more accurate values for these parameters. This finding is in agreement with the findings of previous studies, such as [11, 66, 113]. Fig. 5.5e and f also demonstrate a linear increase in the output of the proposed model after increases in the number and area of the considered PV plants. This change is much sharper for the number of PV systems versus their area factor. As above, this finding is in agreement with the findings in [11, 66, 113].

5.4. GEP-WT Model

In this section, a new model considering the GEP and WT techniques is presented to estimate the generated output power of a group of PV systems directly, as detailed below.

5.4.1. Proposed Methodology

The GEP-WT methodology as shown in Fig. 5.6 is employed to improve the estimation of the output power of a group of PV systems distributed within 0.7 km². In this context, input and output variables along with other parameters influencing these models must be accurately identified as elaborated below.

In order to improve the accuracy of the GEP model, the input and output variables along with other parameters influencing the model must be thoroughly prepared and identified as elaborated below.

A- Data Preparation stage

- PV pyranometer parameters (such as T_{mp} and H_u) are collected from a PV pyranometer sensor located close to the PV plant.
- PV plant properties: These parameters can be identified from the PV plant specification such as A_d^{pv} and P_d^{cap} or through calculations such as V_c . In the

proposed model in this chapter, the PV area (A_d^{pv}) is divided into 6 sub-areas with a specific power capacity (P_d^{cap}) to improve the estimation accuracy. Several considerations are taken into account in dividing the area of the PV plant. This includes the geometrical shape of the site area, the number of PV systems and the overall PV capacity of each area, and the distance between each two PV systems. The purpose of this division is to facilitate the application of the proposed model to an area less than 1 km² to enhance the estimation accuracy. In this study, the considered PV plant is divided into 6 rectangular areas with specific capacities (A_d^{pv}, P_d^{cap} ; $d=1$ to 6) as shown in Fig. 4.3.

- WT parameters: The purpose of using WT analysis is to decompose the original time-series signal into several sub-signals with specific frequencies and time series to develop a precise correlation between the input and output variables. The parameters that are processed using WT include the output power of each plant. Since the PV pyranometer is installed in the same direction and tilt as the PV modules, its output represents the plane of the array (POA) [112]. Let $POA[k]$ denotes the output of the pyranometer, in which k is a time index. Then the equivalent module output power ($P_{POA}^{mdu-in}[k]$) can be given by (3.4) [112].

Subsequently, the average of the total output power of the PV plant ($P_{Eq}^{mdu-out}[k]$), which is equivalent to the output power of each PV module, can be calculated by assuming that each module is providing the same share of the overall power. Then, $P_{POA}^{mdu-in}[k]$ and $P_{Eq}^{mdu-out}[k]$ are normalized using (4.13) and (4.14) to yield $P_{nrm}^{POA-in}[k]$ and $P_{nrm}^{POA-out}[k]$, respectively.

At this end, using (3.3), $P_{nrm}^{POA-in}[k]$ and $P_{nrm}^{POA-out}[k]$ can be decomposed into three timescales. Therefore, WT analysis of $P_{nrm}^{POA-in}[k]$ yields three high-frequency modes ($P_{mode-i}^{HF-in}[k]$) and three low-frequency modes ($P_{mode-i}^{LF-in}[k]$). In the same way, $P_{nrm}^{POA-out}[k]$ is analysed into three high-frequency ($P_{mode-i}^{HF-out}[k]$) and low-frequency modes ($P_{mode-i}^{LF-out}[k]$). These low and high frequency modes are used as input and output data for the proposed GEP-WT model.

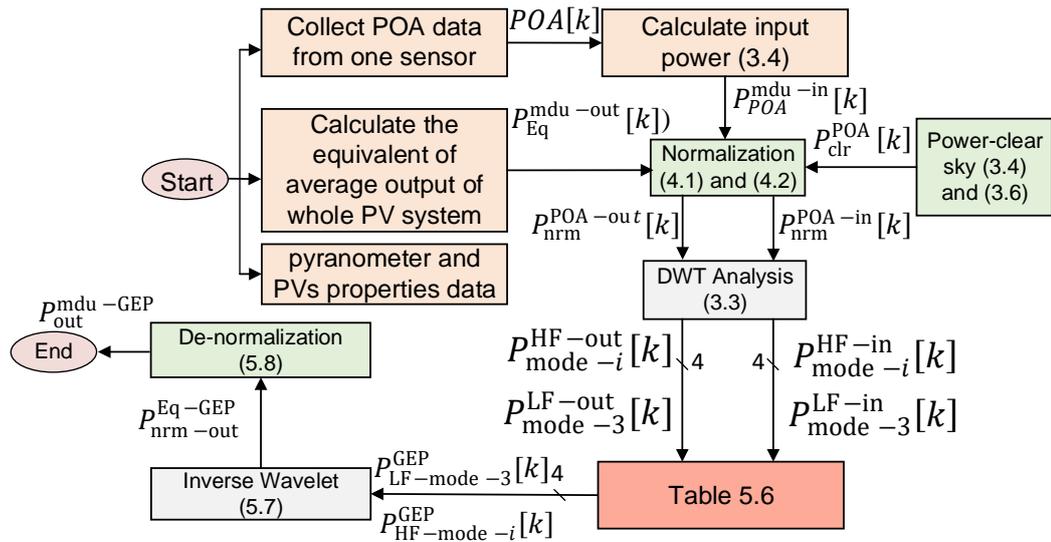


Fig. 5.6. GEP modelling methodology and estimating the overall output power of a group of rooftop PV systems distributed within 0.7 km²

Training and testing the data for the GEP technique results in four GEP models, each represents one of the Wavelet modes. Table 5.9 shows the input and output variables of the GEP training and testing stages as well as the GEP models after the modelling stages in which $P_{HF-mode-1}^{GEP}[k]$, $P_{HF-mode-2}^{GEP}[k]$, and $P_{HF-mode-3}^{GEP}[k]$ are the equivalent GEP-WT mode-1, mode-2, and mode-3 for the high-frequency modes, respectively while $P_{LF-mode-3}^{GEP}[k]$ is the equivalent GEP-WT of mode-3 for the low-frequency mode.

B- Application of the proposed GEP-Wavelet Model

To estimate the output power of the investigated PV plant, the inverse wavelet transform using three high frequency and one low-frequency modes is employed as below [114]:

$$P_{nrm-out}^{Eq-GEP} = inv(P_{HF-mode-1}^{GEP}[k]) + inv(P_{HF-mode-2}^{GEP}[k]) + inv(P_{HF-mode-3}^{GEP}[k]) + inv(P_{LF-mode-3}^{GEP}[k]) \quad (5.7)$$

By de-normalizing the result of (5.7) using $P_{clr}^{POA}[k]$, Eq. (5.8) can be derived

$$P_{out}^{mdu-GEP} = P_{nrm-out}^{Eq-GEP} \times P_{clr}^{POA}[k] \quad (5.8)$$

Then, the overall output power can be calculated by multiplying (5.8) with N_{module}

Table 5.9 GEP input and output variables as well as the resulting GEP models

Inputs	Output	GEP Models after Training and Testing running	
$Tmp[k], H_u[k], V_c, A_i^{pv}$ and P_i^{cap} as $(A_1^{pv} P_1^{cap}), (A_2^{pv} P_2^{cap}), (A_3^{pv} P_3^{cap}),$ $(A_4^{pv} P_4^{cap}), (A_5^{pv} P_5^{cap}), A_6^{pv} P_6^{cap}$	$P_{mode-1}^{HF-in}[k]$	$P_{mode-1}^{HF-out}[k]$	$P_{HF-mode-1}^{GEP}[k]$
	$P_{mode-2}^{HF-in}[k]$	$P_{mode-2}^{HF-out}[k]$	$P_{HF-mode-2}^{GEP}[k]$
	$P_{mode-3}^{HF-in}[k]$	$P_{mode-3}^{HF-out}[k]$	$P_{HF-mode-3}^{GEP}[k]$
	$P_{mode-3}^{LF-in}[k]$	$P_{mode-3}^{LF-out}[k]$	$P_{LF-mode-3}^{GEP}[k]$

$$P_{out}^{Eq-GEP} = P_{out}^{mdu-GEP} \times N_{module} \quad (5.9)$$

Fig. 5.6 shows a flowchart for the above process.

5.4.2. Performance Evaluation

The validation of the proposed technique is conducted on rooftop PV systems located at the University of Queensland in Brisbane, Australia. Data collected over 220 days with 1 min resolution are used for the GEP training and testing stages, while data collected over 90 days are used for the validation stage. Data are collected in the course of the years 2016 and 2017. Data have been collected from 16 PV systems locations distributed within 0.7 km² with an output power capacity of 2.14 MW. The days in which a clear sky or overcast sky was observed are not considered in developing the models. The University of Queensland site features are presented in the last chapter [27]. The PV sensor, which the data are collected from, is located at site-8 shown on Fig. 4.3. As mentioned in chapter 4, to enhance the estimation accuracy of the model, the PV plant is divided into six sub-PV plants with different areas and power capacities. In addition, for each individual day, the *POA* time-series signal is divided into three time segments in which the first time series is between 8:00 and 10:59, the second time series is between 11:00 to 13:59 and the third time series starts at 14:00 and ends at 17:00. The main variables of the GEP model are the terminal elements (inputs and output) and the function elements.

The 30 chromosomes with 3 genes each and 8 codes as head size and a maximum of 1000 generations are set for the proposed model. The genetic operator's impact rates are listed in Table 5.6. These features are fine-tuned several times to obtain the best possible results.

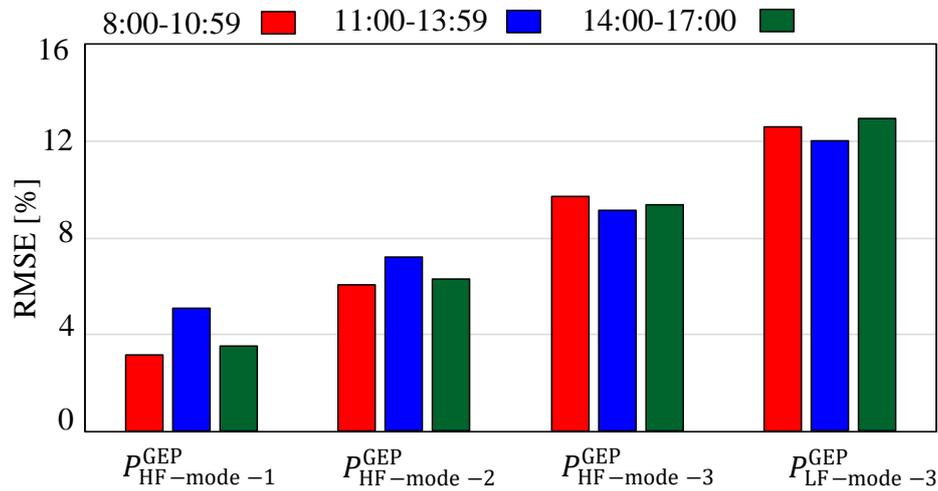


Fig. 5.7. The RMS errors for the estimated generated power during the GEP training process for various Wavelet modes

Fig. 5.7 shows the RMSE for the PV estimated power using the proposed GEP modelling of each Wavelet mode at each sub-time series. While Table 5.10, Table 5.11, Table 5.12 and Table 5.13 list the developed mathematical models resulting at the end of the GEP modelling stage.

Table 5.10 Developed $P_{HF-mode-1}^{GEP}[k]$ wavelet model

Time	Gene	Mathematic Modes
8:00-10:59	1	$\frac{P_{mode-1}^{HF-in}[k]}{3.5} + (4.7Tmp[k] \times (P_{mode-1}^{HF-in})^2)^2$
	2	$\frac{P_{mode-1}^{HF-in}[k]}{\sqrt{V_c + P_{mode-1}^{HF-in}[k] + \frac{Tmp[k] \times p^{cap^2}}{V_c}}}$
	3	$\frac{P_{mode-1}^{HF-in}[k] \times \sqrt{H_u[k] + 8.6}}{P_{mode-1}^{HF-in}[k] - 2A^{pv} + Tmp[k]}$
11:00-13:59	1	$e^{-10A^{pv}} \times \frac{(P_{mode-1}^{HF-in}[k])^3 \times Tmp[k]}{A^{pv}}$
	2	$e^{-12.7A^{pv}} \times (P_{mode-1}^{HF-in}[k] \times (P_{mode-1}^{HF-in}[k] + H_u[k] - A^{pv}))$
	3	$e^{Log(p^{cap}) - V_c} \times \frac{P_{mode-1}^{HF-in}[k]}{-15A^{pv}}$
14:00-17:00	1	$(\frac{H_u[k]}{27A^{pv}} \times Tmp[k] - 0.5V_c)^{0.5} \times P_{mode-1}^{HF-in}[k]$
	2	$\frac{P_{mode-1}^{HF-in}[k]}{C_{ap} + H_u[k] - 8}$
	3	$\frac{0.02H_u[k]}{Tmp[k]} \times P_{mode-1}^{HF-in}[k]$

Table 5.11 Developed $P_{HF-mode-2}^{GEP}[k]$ wavelet model

Time	Gene	Mathematic Modes
8:00-10:59	1	$\frac{P_{mode-2}^{HF-in}[k]}{(-5.7P_{mode-2}^{HF-in}[k] - 5.7Tmp[k])^2} + P_{mode-2}^{HF-in}[k]$
	2	$\frac{P_{mode-2}^{HF-in}[k]}{-3.6 + (Tmp[k])^{\frac{1}{3}}}$
	3	$\frac{P_{mode-2}^{HF-in}[k]}{e^{3p^{cap}}} + P_{mode-2}^{HF-in}[k]$
11:00-13:59	1	$\frac{P_{mode-2}^{HF-in}[k]}{(-5.7P_{mode-2}^{HF-in}[k] - 5.7Tmp[k])^2} + P_{mode-2}^{HF-in}[k]$
	2	$\frac{P_{mode-2}^{HF-in}[k]}{-3.6 + (Tmp[k])^{\frac{1}{3}}}$
	3	$\frac{P_{mode-2}^{HF-in}[k]}{e^{3p^{cap}}} + P_{mode-2}^{HF-in}[k]$
14:00-17:00	1	$\frac{P_{mode-2}^{HF-in}[k]}{(2V_c \times P_{mode-2}^{HF-in}[k]) - V_c - \sqrt{e^{V_c}}}$
	2	$\frac{P_{mode-2}^{HF-in}[k]}{p^{cap} + Log(p^{cap}) - 13}$
	3	$(P_{mode-2}^{HF-in}[k])^2$

Table 5.12 Developed $P_{HF-mode-3}^{GEP}[k]$ wavelet model

Time	Gene	Mathematic Modes
8:00-10:59	1	$P_{mode-3}^{HF-in}[k] + \frac{A^{pv}}{3}$
	2	$\sqrt{\text{Log}(\sqrt{Tmp[k]} + 30P_{mode-3}^{HF-in}[k]) + \frac{A^{pv}}{50}}$
	3	$\sqrt{\text{Log}((17A^{pv})^{\frac{2}{3}}) - \frac{H_u[k]}{0.5V_c - H_u[k]}}$
11:00-13:59	1	$-0.32 P_{mode-3}^{HF-in}[k] \times \text{Log}(V_c)$
	2	$(\frac{A^{pv}}{10} \times P_{mode-3}^{HF-in}[k])^4 \times (\frac{A^{pv}}{10} - Tmp[k] + \frac{1}{p^{cap}})$
	3	$(P_{mode-3}^{HF-in}[k] - \sqrt{0.2A^{pv}} \times P_{mode-3}^{HF-in}[k])$
14:00-17:00	1	$0.5(P_{mode-3}^{HF-in}[k] + A^{pv})$
	2	$(P_{mode-3}^{HF-in}[k])^2 \times (((-P_{mode-3}^{HF-in}[k]) \times V_c - A^{pv}) \times (-6)^{-4})$
	3	$\frac{-p^{cap}}{\frac{H_u[k]}{P_{mode-3}^{HF-in}[k]} + V_c}$

Table 5.13 Developed $P_{LF-mode-3}^{GEP}[k]$ Wavelet model

Time	Gene	Mathematic Modes
8:00-10:59	1	$\frac{Tmp[k]}{400} + \frac{A^{pv}}{40}$
	2	$\frac{(e^{\frac{V_c}{A^{pv} \times Tmp[k]} - p^{cap3}})}{14.6H_u[k] \times A^{pv} \times P_{mode-3}^{LF-in}[k]^{\frac{1}{3}}}$
	3	$P_{mode-3}^{LF-in}[k]$
11:00-13:59	1	$(\frac{Tmp[k]}{400} + \frac{0.5A^{pv}}{Tmp[k]})$
	2	$\frac{(e^{\frac{V_c}{14.6A^{pv}}} - p^{cap})}{14.6H_u[k] \times A^{pv} \times P_{mode-3}^{LF-in}[k]}$
	3	$P_{mode-3}^{LF-in}[k] - 0.54$
14:00-17:00	1	$P_{mode-3}^{LF-in}[k]$
	2	$0.5P_{mode-3}^{LF-in}[k] - 0.5e^{A^{pv} - \sqrt{((\frac{A^{pv} \times H_u[k]}{27} - A^{pv}) \times A^{pv}) + p^{cap}}}$
	3	$e^{A^{pv}} \times \frac{P_{mode-3}^{LF-in}[k]}{(Tmp[k] - (20p^{cap} \times A^{pv}))}$

The above tables show the mathematical models for the $P_{HF-mode-1}^{GEP}[k]$, $P_{HF-mode-2}^{GEP}[k]$, $P_{HF-mode-3}^{GEP}[k]$ and $P_{LF-mode-3}^{GEP}[k]$ wavelet models, respectively

based on 220 days of training and testing data. As an example to clarify the models in these tables, for the sub-time series 8:00–10:59, $P_{\text{HF-GEP}}^{\text{mode-1}}[k]$ is calculated from $P_{\text{HF-GEP}}^{\text{mode-1}}[k] = (\text{Gene}_1) + (\text{Gene}_2) + (\text{Gene}_3)$, which is the same for the other frequency modes. Then, the output power of this sub-time series can be estimated using (5.7). The same procedure is used to estimate the generated power for each sub-time series.

The validation of the GEP-WT model has been achieved using two different methods. First, employing 90 data samples not used in developing the proposed model in this chapter, which are collected during different time-series on three different days in different seasons in Australia. Numerical collected data along with the estimated power of various models as well as the MAE and its average for each model are listed in Table 5.14 to Table 5.22. The average MAE for each 10 samples of data, which represent sub-time series, is also plotted in Fig. 5.8 to Fig. 5.16. From these tables and figures, it can be seen that the average estimation errors (MAE) for the GEP-WT ranges from 3.2% to 8.1%, whereas the average estimation errors (MAE) for the other models are in the range of 6.5% to 14.9%. The results reveal the superiority of the proposed model (GEP-WT) over other models published in the literature in terms of PV power estimation accuracy.

Table 5.14 Comparison of the proposed model and other published models using 10 samples for the first sub-time series ranging from 8:01 to 8:10, 1 January 2018

1 January 2018, first sub-time series											
Model Parameters	Time	8:01	8:02	8:03	8:04	8:05	8:06	8:07	8:08	8:09	8:10
	T_{em}	27.2	27.2	27.2	27.1	27.1	27.1	27	27	26.9	26.9
	H_{u}	64.6	64.3	64.4	64.2	64.5	64.5	64.4	64.5	64.4	64.6
	V_{c}	1.34	1.34	1.34	1.34	1.34	1.34	1.34	1.34	1.34	1.34
	$P_{\text{mdu-in}}^{\text{POA}}$	41.83	37.75	35.59	34.41	32.79	30.1	27.5	25.06	24.49	25.78
	$P_{\text{mdu-out}}^{\text{Equ}}$	46.46	41.03	38.97	37.33	35.62	32.14	29.01	26.91	26.11	27.78

Estimated Power	Hoff	46.35	41.12	38.85	37.43	35.55	32.25	29.15	26.74	26.19	27.66	
	Perez	46.36	41.12	38.87	37.41	35.52	32.24	29.13	26.76	26.17	27.68	
	Lave	46.39	41.09	38.88	37.4	35.53	32.23	29.12	26.77	26.16	27.7	
	VRI_T^{\max}	46.4	41.09	38.89	37.39	35.53	32.22	29.11	26.77	26.16	27.7	
	GEP-WT	46.42	41.06	38.9	37.4	35.54	32.2	29.1	26.89	26.14	27.72	
MAE [%]	Hoff	10.57	8.47	11.61	10.53	7.12	11.06	13.54	16.66	8.32	11.76	
	Perez	9.61	8.84	9.76	8.44	9.79	9.17	11.32	15.28	6.33	10.05	
	Lave	7.29	5.53	8.8	6.61	9.22	8.17	10.32	13.78	5.23	8.1	
	VRI_T^{\max}	6.29	5.33	8.1	6.21	8.76	7.56	9.8	13.46	4.56	7.89	
	GEP-WT	3.79	2.53	6.8	7.11	8.22	5.67	8.82	1.78	2.93	5.9	
Average MAE [%]	Hoff	10.96										
	Perez	9.86										
	Lave	8.31										
	VRI_T^{\max}	7.8										
	GEP-WT	5.36										

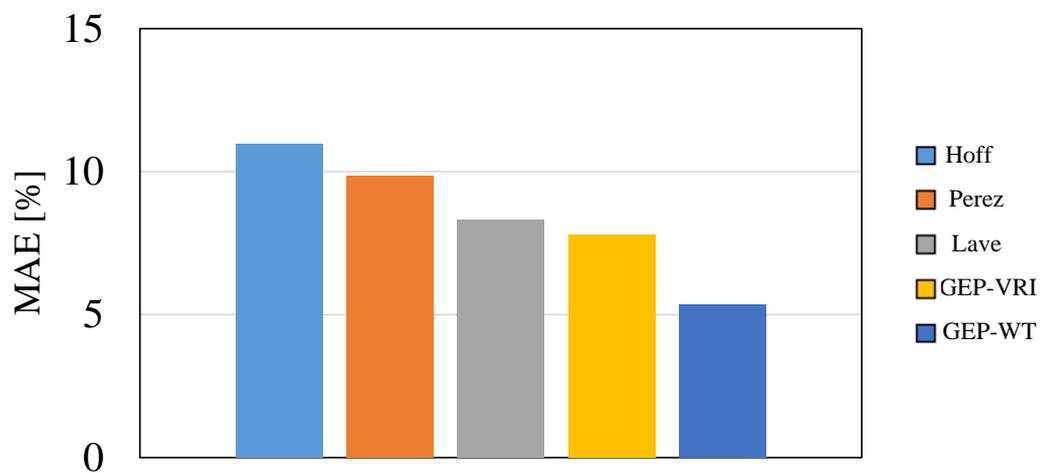


Fig. 5.8. Average MAE error for Table 5.14

Table 5.15 Comparison of the proposed model and other published models using 10 samples for the second sub-time series ranging from 12:01 to 12:10, 1 January 2018

1 January 2018, second sub-time series												
	Time	12:01	12:02	12:03	12:04	12:05	12:06	12:07	12:08	12:09	12:10	
Model Parameters	T_{em}	30.8	30.7	30.8	31	31.1	31.1	31.1	31.2	31.2	31.3	
	H_u	52.6	52.8	53	52.8	52.9	53.1	52.3	52.2	52.1	52.1	
	V_c	1.34	1.34	1.34	1.34	1.34	1.34	1.34	1.34	1.34	1.34	
	P_{mdu-in}^{POA}	124	131	152	139	112	98	110	210	207	201	
	$P_{mdu-out}^{Equ}$	133	149	160	140	127	141	182	188	196	192	
Estimated Power	Hoff	120	135	167	124	119	130	160	164	180	182	
	Perez	122	137	166	126	120	132	162	166	181	181	
	Lave	124	139	169	131	122	135	165	169	184	185	
	VRI_T^{max}	121	130	169	133	123	137	166	174	185	189	
	GEP-WT	130	143	164	135	132	150	175	192	190	190	
MAE [%]	Hoff	13	14	7	16	8	11	22	24	16	10	
	Perez	11	12	6	14	7	9	20	22	15	11	
	Lave	9	10	9	9	5	6	17	19	12	7	
	VRI_T^{max}	12	19	9	7	4	4	16	14	11	3	
	GEP-WT	3	6	4	5	5	9	7	4	6	2	
Average MAE [%]	Hoff	14.1										
	Perez	12.7										
	Lave	10.3										

	VRI_T^{\max}	9.9
	GEP-WT	5.1

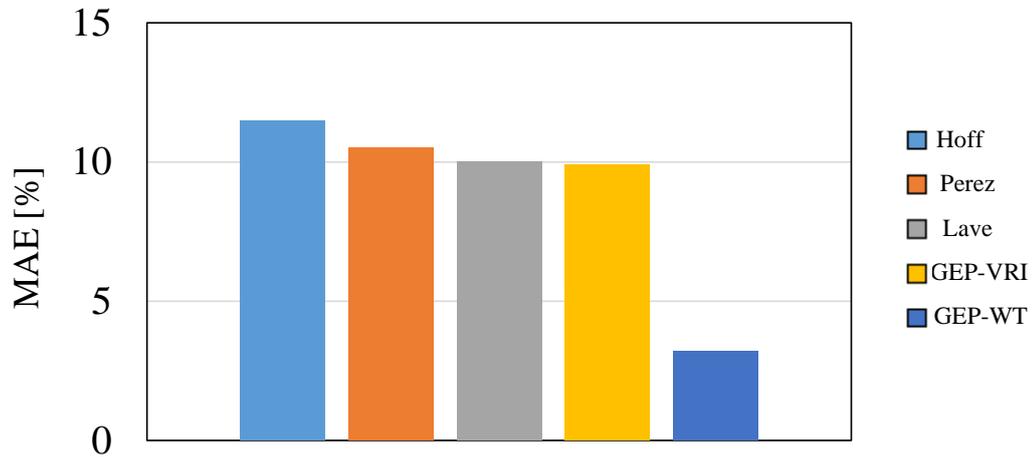


Fig. 5.9. Average MAE error for Table 5.15

Table 5.16 Comparison of the proposed model and other published models using 10 samples for the third sub-time series ranging from 15:01 to 15:10, 1 January 2018

1 January 2018, third sub-time series											
Model Parameters	Time	15:01	15:02	15:03	15:04	15:05	15:06	15:07	15:08	15:09	15:10
	T_{em}	31.3	31.4	31.3	31.2	31	30.8	30.8	30.8	30.9	30.9
	H_u	53.5	54	53.9	54	54.1	54.5	54.9	55	54.8	55
	V_c	1.34	1.34	1.34	1.34	1.34	1.34	1.34	1.34	1.34	1.34
	P_{mdu-in}^{POA}	98	96	98	57	53	73	54	68	80	73
	$P_{mdu-out}^{Equ}$	117	101	86	68	63	64	68	82	100	77
Estimated Power	Hoff	103	114	69	56	50	55	75	73	85	71
	Perez	102	113	72	58	54	54	77	75	87	71

	Lave	106	111	72	59	54	56	77	73	85	71
	VRI_T^{\max}	124	93	74	78	51	71	74	60	109	70
	GEP-WT	121	99	81	67	59	60	63	79	102	75
MAE [%]	Hoff	14	13	17	12	13	9	7	9	15	6
	Perez	15	12	14	10	9	10	9	7	13	6
	Lave	11	10	14	9	9	8	9	9	15	6
	VRI_T^{\max}	12	19	9	7	4	4	16	14	11	3
	GEP-WT	4	2	5	1	4	4	5	3	2	2
Average MAE [%]	Hoff	11.5									
	Perez	10.5									
	Lave	10									
	VRI_T^{\max}	9.9									
	GEP-WT	3.2									

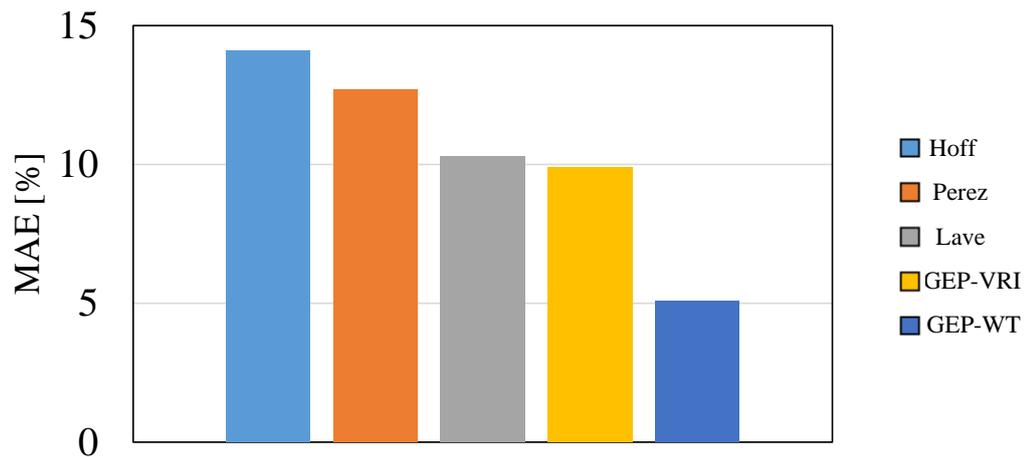


Fig. 5.10. Average MAE error for Table 5.16

Table 5.17 Comparison of the proposed model and other published models using

10 samples for the first sub-time series ranging from 9:01 to 9:10, 17 Sept. 2017

17 Sept. 2017, first sub-time series												
Model Parameters	Time	9:01	9:02	9:03	9:04	9:05	9:06	9:07	9:08	9:09	9:10	
	T_{em}	18.7	18.7	18.7	18.7	18.7	18.6	18.7	18.7	18.7	18.7	
	H_u	50.2	50.3	50.5	50.6	50.5	50.5	50.8	50.6	50.9	50.8	
	V_c	2.03	2.03	2.03	2.03	2.03	2.03	2.03	2.03	2.03	2.03	
	P_{mdu-in}^{POA}	70	70	66	65	69	66	69	69	64	65	
	$P_{mdu-out}^{Equ}$	75	75	73	71	74	73	75	75	71	71	
Estimated Power	Hoff	24	24	24	24	24	24	24	24	24	24	
	Perez	97	97	68	67	55	69	97	97	67	67	
	Lave	95	95	68	66	55	69	95	95	66	66	
	VRI_T^{\max}	89	89	68	63	59	69	89	89	63	63	
	GEP-WT	71	71	68	61	54	69	79	73	80	79	
MAE [%]	Hoff	99	99	64	67	52	63	99	99	67	67	
	Perez	22	22	5	4	19	4	22	22	4	4	
	Lave	20	20	5	5	19	4	20	20	5	5	
	VRI_T^{\max}	14	14	5	8	15	4	14	14	8	8	
	GEP-WT	4	4	5	10	20	4	4	2	9	8	
Average MAE [%]	Hoff	14.9										
	Perez	12.8										
	Lave	12.3										
	VRI_T^{\max}	10.4										
	GEP-WT	7										

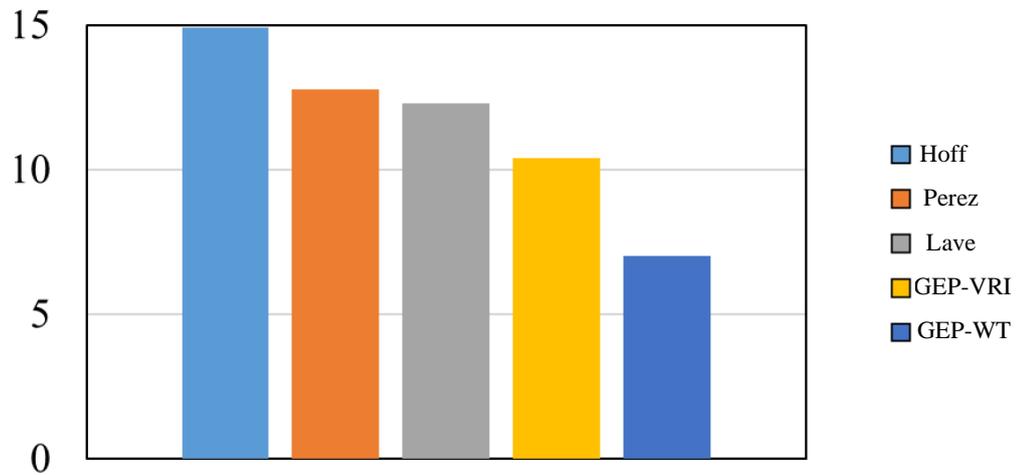


Fig. 5.11. Average MAE error for Table 5.17

Table 5.18 Comparison of the proposed model and other published models using 10 samples for the second sub-time series ranging from 12:01 to 12:10, 17 Sept. 2017

17 Sept. 2017, second sub-time series											
Model	Time	12:01	12:02	12:03	12:04	12:05	12:06	12:07	12:08	12:09	12:10
Parameter	T_{em}	22.6	22.6	22.5	22.5	22.5	22.5	22.4	22.5	22.5	22.5
	H_u	43.1	42.9	43	43.1	43.2	43	43	43.1	43.1	43.1
	V_c	2.03	2.03	2.03	2.03	2.03	2.03	2.03	2.03	2.03	2.03
	P_{mdu-in}^{POA}	168.55	168.95	166.28	142.68	165.6	160.26	167	162.56	163.35	164.3
	$P_{mdu-out}^{Equ}$	194.98	196.3	190.5	176.01	191.83	192.25	194.53	191.41	191.16	191.66
	Estimated Power	Hoff	195.12	196.45	190.62	176.1	191.99	192.38	194.64	191.28	191.26
Perez		195.11	196.44	190.62	176.09	191.97	192.37	194.63	191.3	191.25	191.59
Lave		195.11	196.43	190.6	176.08	191.95	192.36	194.63	191.31	191.24	191.59
VRI_T^{\max}		195.1	196.42	190.6	176.07	191.95	192.36	194.62	191.32	191.24	191.6

	GEP-WT	194.91	196.4	190.44	176.07	191.72	192.33	194.47	191.5	191.22	191.61
MAE [%]	Hoff	14.1	15.29	12.17	9.01	15.73	13.08	10.34	12.87	9.97	8.66
	Perez	13.08	14.3	11.99	8.43	13.26	12.12	10.14	11.48	9.31	7.65
	Lave	12.48	13.52	10.77	7.51	11.81	11.41	9.4	10.1	8.33	7.02
	VRI_T^{\max}	12.28	12.52	10.18	6.51	11.34	10.82	8.98	9.2	7.54	6.2
	GEP-WT	7.09	10.06	5.55	6.47	11.27	8.22	6.16	8.82	5.97	5.31
Average MAE [%]	Hoff	12.12									
	Perez	11.18									
	Lave	10.24									
	VRI_T^{\max}	9.56									
	GEP-WT	7.49									

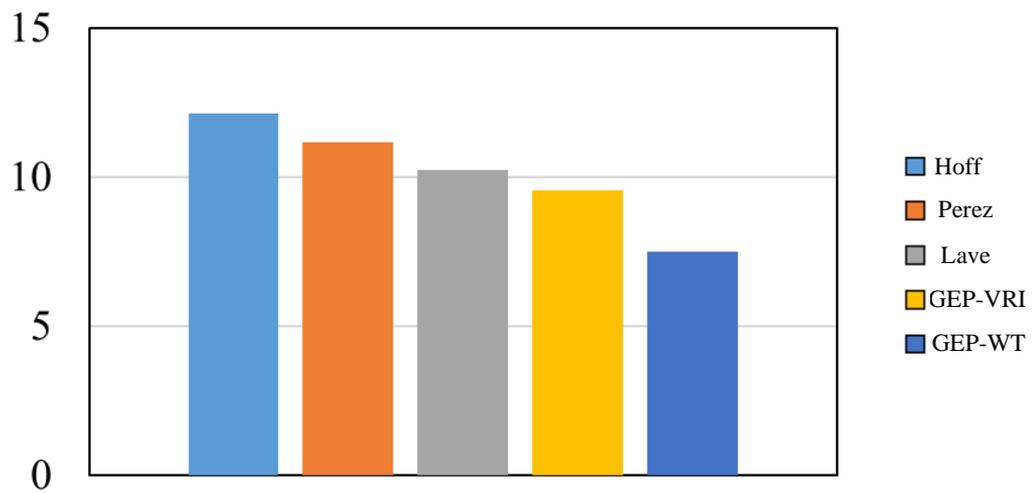


Fig. 5.12. Average MAE error for Table 5.18

Table 5.19 Comparison of the proposed model and other published models using 10 samples for the third sub-time series ranging from 14:31 to 14:40, 17 Sept. 2017

17 Sept. 2017, third sub-time series											
Model Parameters	Time	14:31	14:32	14:33	14:34	14:35	14:36	14:37	14:38	14:39	14:40
	T_{em}	22.6	22.9	22.9	22.9	22.8	22.7	22.6	22.6	22.5	22.4
	H_u	41.7	42.1	41.9	42.3	42.4	42.5	42.9	43	43.2	43.3
	V_c	2.03	2.03	2.03	2.03	2.03	2.03	2.03	2.03	2.03	2.03
	P_{mdu-in}^{POA}	111	109	110	110	109	73	57	36	35	38
	$P_{mdu-out}^{Equ}$	135	135	137	140	140	122	106	66	55	67
Estimated Power	Hoff	120	120	125	126	126	106	94	51	55	55
	Perez	124	124	127	125	125	108	95	52	51	53
	Lave	122	122	129	125	126	109	97	53	53	56
	VRI_T^{max}	124	124	130	127	129	110	99	55	45	51
	GEP-WT	131	131	132	131	132	114	102	69	50	59
MAE [%]	Hoff	15	15	12	14	14	16	12	15	0	12
	Perez	11	11	10	15	15	14	11	14	4	14
	Lave	13	13	8	15	14	13	9	13	2	11
	VRI_T^{max}	11	11	7	13	11	12	7	11	10	16
	GEP-WT	4	4	5	9	8	8	4	4	5	8
Average MAE [%]	Hoff	12.5									
	Perez	11.9									
	Lave	11.1									
	VRI_T^{max}	10.9									

	GEP-WT	6.8
--	--------	-----

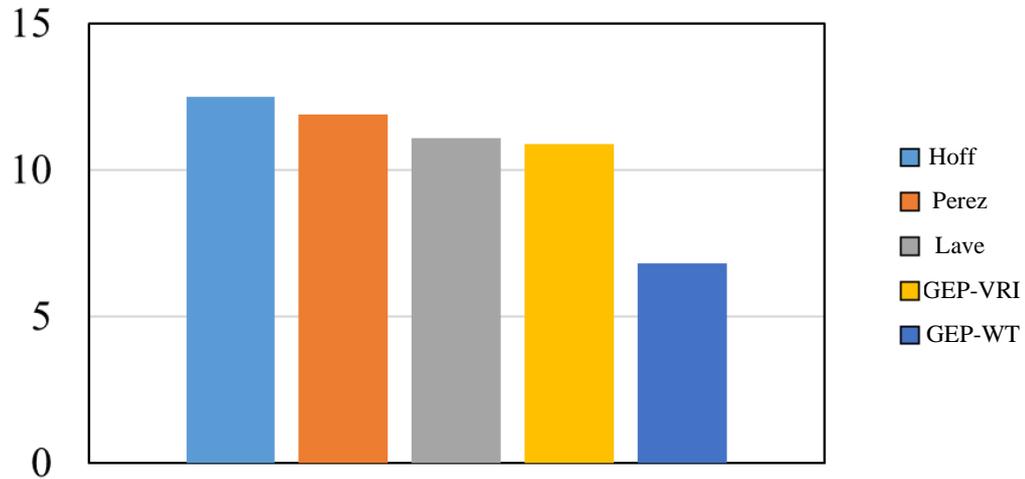


Fig. 5.13. Average MAE error for Table 5.19

Table 5.20 Comparison of the proposed model and other published models using 10 samples for the first sub-time series ranging from 10:01 to 10:10, 18 March 2017

18 March 2017 first sub-time series											
Model Parameters	Time	10:01	10:02	10:03	10:04	10:05	10:06	10:07	10:08	10:09	10:10
	T_{em}	28.3	28.2	28.2	28.2	28.2	28.4	28.5	28.4	28.5	28.5
	H_u	51	51	51.1	50.8	50.2	50.3	50	49.8	50	49.8
	V_{cloud}	4.12	4.12	4.12	4.12	4.12	4.12	4.12	4.12	4.12	4.12
	P_{mdu-in}^{POA}	148	149	151	153	155	159	154	152	154	153
	$P_{mdu-out}^{Equ}$	142	150	145	156	156	153	157	158	160	159
Estimated Power	Hoff	129	141	130	142	143	147	148	150	146	155
	Perez	134	144	133	142	145	149	149	151	144	152
	Lave	136	136	136	136	136	136	136	136	136	136
	VRI_T^{max}	137	148	134	145	146	153	155	155	145	153

	GEP-WT	139	139	139	139	139	139	139	139	139	139
MAE [%]	Hoff	13	9	15	14	13	6	9	8	14	4
	Perez	8	6	12	14	11	4	8	7	16	7
	Lave	6	3	10	12	8	2	4	3	16	5
	VRI_T^{\max}	5	2	11	11	10	0	2	3	15	6
	GEP-WT	3	5	6	5	8	1	2	7	5	4
Average MAE [%]	Hoff	10.5									
	Perez	9.3									
	Lave	6.9									
	VRI_T^{\max}	6.5									
	GEP-WT	4.6									

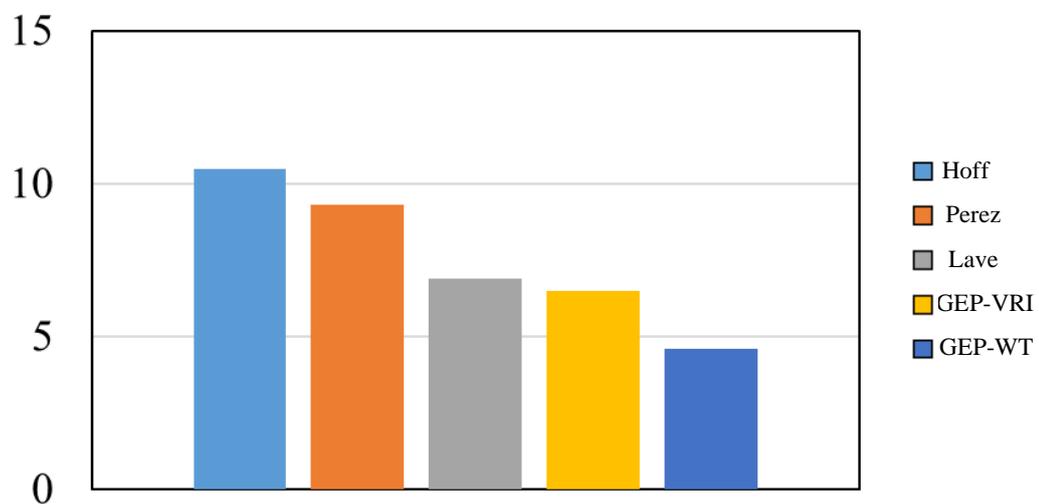


Fig. 5.14. Average MAE error for Table 5.20

Table 5.21 Comparison of the proposed model and other published models using 10 samples for the second sub-time series ranging from 13:51 to 14:00, 18 March 2017

18 March 2017, second sub-time series											
Model Parameters	Time	13:51	13:52	13:53	13:54	13:55	13:56	13:57	13:58	13:59	14:00
	T_{em}	30.2	30.2	30.2	30.3	30.4	30.4	30.3	30.3	30.3	30.2
	H_u	46.1	46.2	46.2	45.9	45.9	46	46.3	46.5	46.8	46.8
	V_c	4.12	4.12	4.12	4.12	4.12	4.12	4.12	4.12	4.12	4.12
	P_{mdu-in}^{POA}	50	65	192	76	62	59	54	52	53	60
	$P_{mdu-out}^{Equ}$	71	89	110	73	79	77	70	70	79	84
Estimated Power	Hoff	65	75	95	60	65	63	63	63	65	60
	Perez	68	79	99	67	69	62	63	63	62	61
	Lave	69	83	101	67	71	64	65	62	63	65
	VRI_T^{max}	69	84	103	68	73	65	66	63	63	64
	GEP-WT	74	93	116	79	84	83	80	80	90	101
MAE [%]	Hoff	6	14	15	13	14	14	7	7	14	24
	Perez	3	10	11	6	10	15	7	7	17	23
	Lave	2	6	9	6	8	13	5	8	16	19
	VRI_T^{max}	2	5	7	5	6	12	4	7	16	20
	GEP-WT	3	4	6	6	5	6	15	10	11	12
Average MAE [%]	Hoff	12.8									
	Perez	10.9									
	Lave	9.2									

	VRI_T^{\max}	8.4
	GEP-WT	8.1

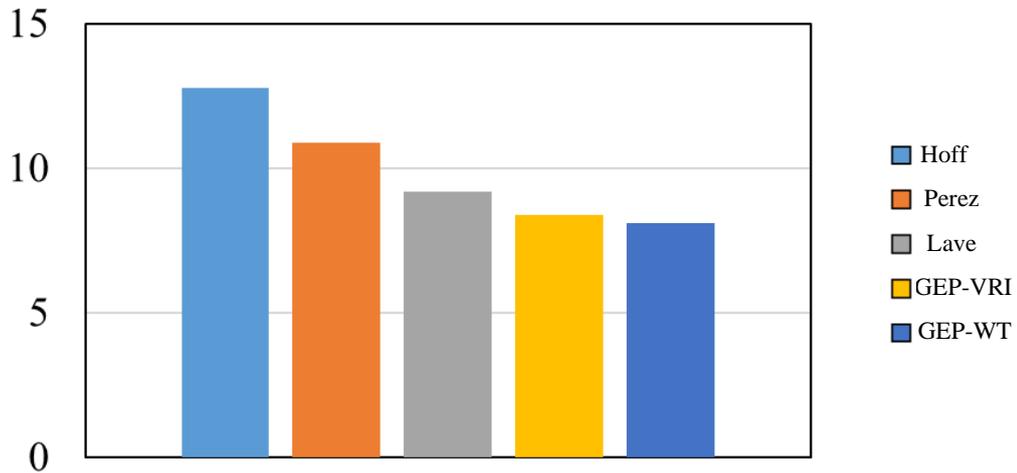


Fig. 5.15. Average MAE error for Table 5.21

Table 5.22 Comparison of the proposed model and other published models using 10 samples for the second sub-time series ranging from 16:01 to 16:10, 18 March 2017

18 March 2017, third sub-time series											
Model Parameters	Time	16:01	16:02	16:03	16:04	16:05	16:06	16:07	16:08	16:09	16:10
	T_{em}	28.7	28.8	28.7	28.7	28.7	28.6	28.6	28.6	28.6	28.6
	H_u	56.9	56.4	56.3	56.4	56.5	56.5	56.7	56.9	57.2	57.3
	V_c	4.12	4.12	4.12	4.12	4.12	4.12	4.12	4.12	4.12	4.12
	P_{mdu-in}^{POA}	29.16	20.4	18.54	15.68	13.78	64.89	13.63	11.35	10.49	53.41
	$P_{mdu-out}^{Equ}$	46.33	32.34	29.97	37.8	34.74	31.25	31.77	22.16	23.31	39.7
Estimated Power	Hoff	46.45	32.44	30.19	37.72	34.84	31.36	31.64	22.07	23.21	39.53
	Perez	46.44	32.43	30.16	37.73	34.83	31.36	31.65	22.08	23.22	39.54

	Lave	46.43	32.43	30.15	37.74	34.82	31.35	31.66	22.09	23.23	39.55
	VRI_T^{\max}	46.43	32.42	30.15	37.76	34.81	31.35	31.68	22.11	23.24	39.57
	GEP-WT	46.41	32.29	29.84	37.79	34.7	31.19	31.84	22.2	23.39	39.6
MAE [%]	Hoff	12.24	10.3	21.99	8.5	9.96	10.81	12.36	8.42	10.2	16.29
	Perez	10.9	9.25	19.31	7.47	8.45	10.44	11.84	8.09	9.12	15.57
	Lave	10.1	9.13	18.21	6.02	8.11	9.52	10.97	6.75	8.28	14.53
	VRI_T^{\max}	9.93	8.43	18.14	4.07	7.24	9.45	9.02	5.29	7.18	12.86
	GEP-WT	7.92	4.74	12.67	1.03	4.21	6.24	7.32	4.12	7.91	9.56
Average MAE [%]	Hoff	12.11									
	Perez	11.04									
	Lave	10.16									
	VRI_T^{\max}	9.16									
	GEP-WT	6.57									

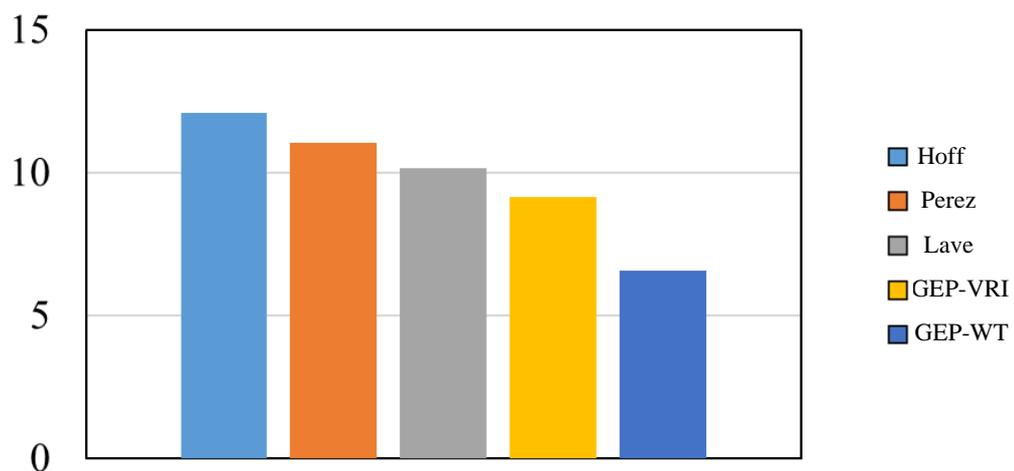


Fig. 5.16. Average MAE error for Table 5.22

The second way to illustrate the validation of the GEP-WT model is by employing 90 days using two error metrics – MAE and RMSE. Fig. 5.17 shows the errors of

the existing models (Hoff, Perez, Lave and GEP-VRI models) including the GEP-WT model. As shown in Fig. 5.17, the GEP-WT model results in the least errors when compared with other estimation models.

Table 5.23 lists the average numerical values for MAE and RMSE as well as the maximum and minimum enhancement due to the application of the GEP-WT approach compared with the other approaches. It can be seen that the maximum accuracy improvements of the GEP-WT approach using the MAE metric are 14.38, 12.46 and 6.62% at sub-time series of 8:00–10:59 am, 11:00–1:59 pm and 2:00–5:00 pm, respectively. While the maximum accuracy improvements of the GEP-WT approach are 47.61, 46.72, and 31.10% at sub-time series of 8:00–10:59 am, 11:00–1:59 pm and 2:00–5:00 pm, respectively. The same results can be illustrated using the RMSE metric.

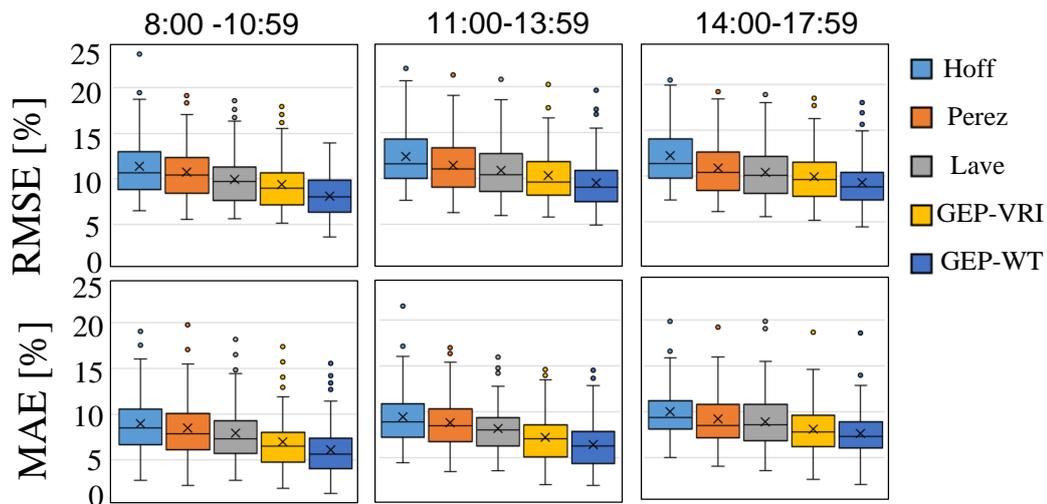


Fig. 5.17. Error comparison between the estimated and the actual power using various models published in the literature and the model proposed in this chapter (GEP-WT)

It must be noted that the enhancing strategy adopted to model the output power of a neighbourhood PV system is based on dividing the daily POA time-series data into three sub-time-series. Also, along with the WT technique, the relationship between the input and the output variables are quickly mapped. Both approaches, dividing the POA data and the WT, assist the GEP algorithm in drawing the correlation between the input and output variables promptly and precisely. As a

result of that, within each selected period, the environmental weather conditions, such as zenith angle, temperature, humidity and output power would not change much. Furthermore, the separation conducted by the WT approach on each sub-time series into high and low-frequency groups decreases the GEP confusion throughout the training progression to draw the relationship between the input and output variables.

Therefore, the proposed technique improved in this section demonstrated a high accuracy compared with the other models mentioned in Chapter 2 (Hoff, Perez and Lave) and Section 5.2 (GEP-VRI).

Table 5.23 Comparison of the minimum and maximum errors for the main VRI models published in the literature and the proposed GEP-WT model

Models	8:00–10:59 a.m.		11:00–1:59 p.m.		2:00–5:00 p.m.	
	MAE	RMES	MAE	RMES	MAE	RMES
Hoff	8.93	11.36	9.38	12.34	9.94	12.46
Perez	8.44	10.79	8.80	11.43	9.18	11.08
Lave	7.89	9.84	8.15	10.88	8.88	10.65
GEP-VRI	6.92	9.42	7.19	10.32	8.08	10.18
GEP-WT	6.05	8.09	6.39	9.49	7.58	9.50
Minimum						
improvement (%)	14.38	16.42	12.46	8.74	6.62	7.10
Maximum						
improvement (%)	47.61	40.38	46.72	30.07	31.10	31.19

5.4.3. Sensitivity Analysis

To investigate the sensitivity of the parameters used in the developed models in Table 5.24, each parameter in the model is changed within a range of maximum and

minimum limits while keeping all other parameters unchanged [101]. The effect of such a change on the output power is analysed using the standard deviation (SD). To standardise the sensitivity analysis of the model, the SD is divided by the average value of the measured parameter. Fig. 5.18 shows the sensitivity of each model parameter for each sub-time series. The results show that the most influential factor is the input power which represents the output power of the PV sensor followed by the area of the PV system and then the ambient temperature, whereas humidity, cloud speed and the power capacity of the overall PV system have less impact.

Table 5.24 Maximum and minimum values of the estimated model used in the sensitivity analysis

Parameters	Minimum	Maximum
$P_{\text{norm-out}}^{\text{Eq-GEP}}$	1	0
$H_u[\%]$	10	90
$V_c[\text{m/s}]$	0.5	13
$A_j^{\text{pv}} [\text{km}^2]$	0.1	0.7
$Tmp[^\circ\text{C}]$	10	36
$P_j^{\text{cap}} [\text{Mw}]$	0.15	2

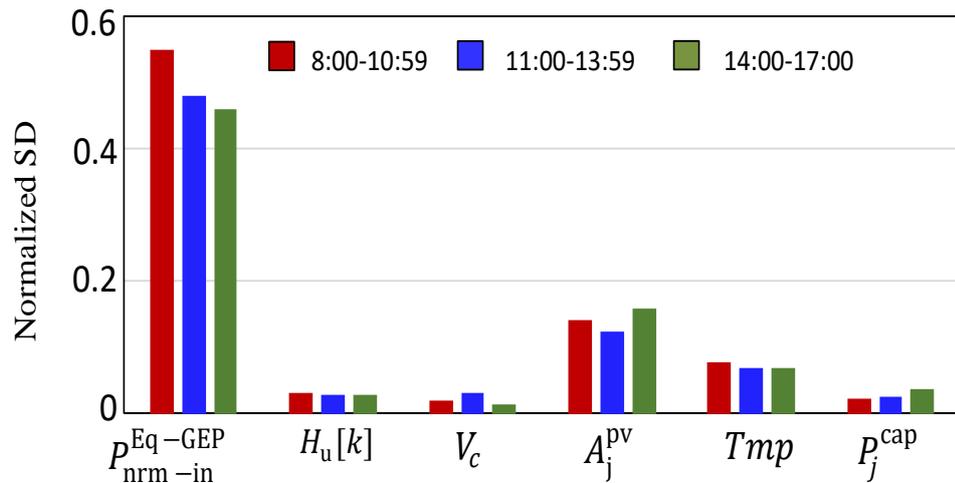


Fig. 5.18. Sensitivity analysis of the estimated model parameters

5.5. Comparison of the Proposed Models vs Existing Models in the Literature

Finally, this section presents a comparison of the existing models (Hoff, Perez and Lave) and the models improved in this thesis. Table 5.25 illustrates the final average errors using MAE and RMSE metrics calculated using the proposed models in this thesis (GEP-VRI, ANFIS-WT and GEP-WT) compared with the well-known existing models (Hoff, Perez and Lave). Fig. 5.19 illustrates graphically the average errors of each of the models investigated in this thesis as well as the maximum and minimum accuracy improvements of the GEP-WT model compared with the other models. From Table 5.25 and Fig. 5.19, it can be seen that by using MAE, the minimum improvements of the GEP-WT models range 8.65% (compared with the ANFIS-WT model) to 33.65% (compared with the Hoff model) whereas the maximum improvements of the GEP-WT models range from 15.2% (compared with the ANFIS-WT model) to 56.8% (compared with the Hoff model).

Table 5.25 Comparison of the minimum and maximum average errors calculated by each model (existing and proposed models) in this study

Models [%]	Minimum [%]		Maximum [%]	
	MAE	RMSE	MAE	RMSE
Hoff	9.8	12.1	13.9	15.8
Perez	9.3	12	13.4	15.8
Lave	8.7	10.9	12.7	14.5
GEP-VRI	8.1	10.3	12.8	13.9
ANFIS-WT	7.2	9.5	11.3	13
GEP-WT	6.25	8.52	10.4	12.23
GEP-WT model Improvement compare with Hoff	56.8	42.02	33.65	29.19
GEP-WT model Improvement compare with Perez	48.8	40.85	28.85	29.19
GEP-WT model Improvement compare with Lave	39.2	27.93	22.12	18.56
GEP-WT model Improvement compare with GEP-VRI	29.6	20.89	23.08	13.65
GEP-WT model Improvement compare with ANFIS-WT	15.2	11.50	8.65	6.30

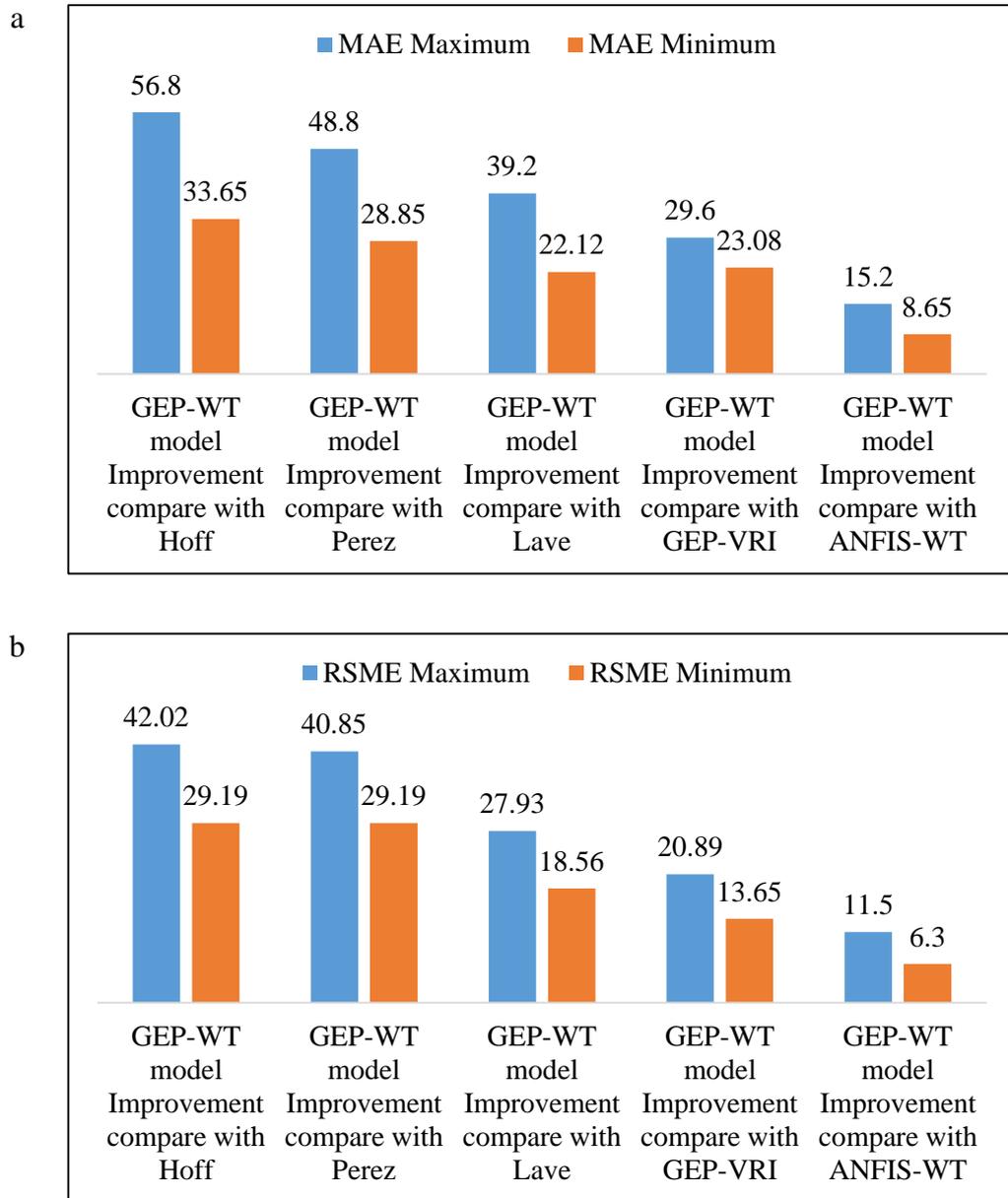


Fig. 5.19. The maximum and minimum improvements of the GEP-WT model compared with other models using a) MAE and b) RMSE

In summary, it can be concluded that the GEP algorithm could be more suited to dealing with the variability of the solar irradiance, especially with small-scale PV plants.

5.6. Summary

This chapter has presented two new models using the GEP technique. The first model is the GEP-VRI which is derived by employing the $VRI_{\text{mode-}i}^{\text{max}}$ model, which was introduced in Chapter 3, as an output variable, and the area of the PV plant, cloud speed, humidity, number of PV systems, ambient temperature and timescale as input variables. While the second model – GEP-WT – was simulated by employing the average output power of the overall PV plant analysed using WT as an output variable, and the output power of the pyranometer analysed by employing WT, the PV plant area, cloud speed, humidity and temperature as input variables. The validation results demonstrated the GEP-WT has greater accuracy than the GEP-VRI model because they have different training strategies. From the comparison results listed in Table 5.25, it can be seen that the minimum enhancements of the GEP-WT model, compared with the GEP-VRI model, are 15.2% and 8.65%. While the maximum enhancements of the GEP-WT model, compared with the ANFIS-WT model, are 29.6% and 23.07%. The accuracy of the GEP-WT model is due to the use of a strategy to provide accurate and specific data by dividing the time series for each day into a set of sub-time series in which some weather conditions are mostly the same during this sub-time series, such as temperature, humidity, zenith angle and solar irradiance. Employing the Wavelet transform technique to analyse the data and separate it (time-series data) into low and high frequencies, the confusion of the GEP algorithm will be reduced throughout the learning stage. Moreover, the strategy of dividing the PV plant into several sub-plants with different capacities and shapes helped produce a model capable of estimating the output power of PV plants of different sizes and capacities within 1 km²

Chapter 6 Conclusion and Recommendations

This chapter summarises the main findings of this thesis. Based on different findings and observations, some recommendations are also made for future research.

6.1. Conclusions

The study presented in this doctoral thesis focused on creating models using several approaches and several scientific strategies. The study was implemented by using several neighbouring rooftop PV systems distributed within approximately 1 km² and solar irradiance data with 1 minute resolution. The proposed models have also been created to estimate the generated power of the PV plant using only one PV pyranometer located close to the PV plant. The validation data was collected from the rooftop PV systems located in the University of Queensland in Brisbane, Australia.

The major scope of this study, the thesis structure, the study motivation and main subjects were explained in Chapter 1. In Chapter 2, several models and studies have been introduced. These studies and models from existing literature express correlations between two PV systems located in one area. These correlational models are used to determine the VRI of the power generated by these PV systems and estimate their overall generated power. This chapter also included two tables to summarise the approaches and the results obtained from these studies.

Chapter 3 focused on defining the accuracy of various correlational models in the literature when applied to a group of small-scale rooftop PV systems, distributed in a small area, versus large PV systems distributed over large areas, for which these models were developed and validated. Employing these models for the actual system under consideration revealed that no single correlational model always results in the most accurate estimation. Accordingly, a new model was introduced by adopting the maximum value of each of the previous models (Hoff, Perez, Lave and ACM) at each timescale, individually. Using the WT technique, POA data have been analysed into six modes and then compared with a maximum VRI model to estimate the overall output power of neighbouring PV systems. The developed method is validated by comparing the power generated by a group of 16

neighbouring small-scale rooftop PV systems, with the estimation results from the proposed method. It can be noted that the minimum error observed using MAE was 2.5 and 1.1 for the $VRI_{\text{mode-}i}^{\text{max}}$ model and between 2.6 and 1.2% for the other models. The results also show that the proposed method improves the accuracy by at least 1.8% and up to 32 or 48% (depending on whether the MAE or RMSE are employed) for 312 non-clear sky days of a year. Furthermore, since this technique uses a single pyranometer for the distributed PV systems in the study, the impact of the location of the pyranometer on the results has also been investigated and the most appropriate location has been identified.

In Chapter 4, using the Wavelet-based ANFIS modelling approaches, new PV system output power estimation models are presented. Parameters such as PV area, capacity, cloud speed, temperature and humidity are considered as inputs for the proposed model. Real data collected from the PV plant in Brisbane, Australia are used to train the proposed model. The results reveal the high accuracy of the developed model compared with the existing estimation models in the literature. For example, the minimum and maximum improvement using the ANFIS model are 11.4 and 36.9% compared with the other models (Hoff, Perez, Lave and GEP-VRI). Sensitivity analysis for each parameter reveals that the output power collected from the PV sensor has the greatest impact on estimating the output power of the entire PV system using the developed model. Ambient parameters including humidity, cloud speed and the power capacity of the overall PV plant all have less impact on the estimated power.

Chapter 5 aimed to propose two techniques with higher accuracy compared to similar existing techniques for estimating the total power generated by a group of neighbouring PV systems in a locality using the VRI. Existing techniques use mathematical approaches based on the correlation coefficient between every two PV systems. The two new techniques proposed here present a GEP-based artificial intelligence technique which merges the benefits of the GA and GP techniques to derive the correlation between the input parameters, such as the number and size of the PV panels, and the ambient conditions, such as the temperature and humidity and the POA, measured at one point only, and the total power generated across the neighbourhood. The first technique used the $VRI_{\text{mode-}i}^{\text{max}}$ parameter as an output

variable, and using the resulting model (GEP-VRI) and the WT approach, the overall generated power was estimated. The results have demonstrated that the proposed technique has a minimum and maximum improvement of 8 (or 8.4) and 43.7 (or 51)% using the MAE (or RMSE) metric. The proposed method has also demonstrated an accuracy of over 90% for 82% of the studied days which is far better than the accuracy of other techniques in the literature (i.e. 57.6 to 67%). In the second technique, the overall generated power has been modelled directly using the WT and GEP approaches. The result of this technique demonstrated a minimum and maximum improvement of 14.38 (or 16.42) and 47.61 (or 40.38)% using the MAE (or RMSE) metric. The sensitivity analysis of the second model shows that the output power collected from the PV sensor has the largest impact on the output power estimated by the GEP models. Ultimately, the thesis demonstrates that the GEP-WT model is the more accurate of the models for estimating the output power of a group of rooftop PV systems distributed within 1 km² using only one PV sensor and 1-minute solar irradiance data compared with existing models (Hoff, Perez and Lave models) and the other models suggested in this thesis (ANFIS-WT and GEP-VRI models).

6.2. Recommendations for future work

Based on the findings from this research, the points below can be suggested as potential future research subjects:

- Based on the results of this thesis, VRI could change during the daytime if cloud speed is not constant during that period. So it would be interesting to study the VRI parameter by dividing the data for each day into several sub-time series to calculate cloud speed at each sub-series time, the VRI for each sub-time series and then using the usual methods to estimate the generated power.
- This thesis focused on one-minute resolution; by using ANFIS and GEP techniques, it is possible to create other models using different time resolutions such as 2, 3, 4, 5 ...15-mins.
- This thesis also focused on a small-scale PV plant. It would also be interesting to use ANFIS and GEP techniques to model the generated power of medium and large-scale PV plants with different time resolutions.

- The Wavelet transform model is a good tool to analyse and evaluate uncertain and random data. More accurate results could be yielded when considering the more accurate Mother wavelet function. The author recommends using and comparing these different Mother functions to calculate and estimate the output power of PV systems.
- Because cloud speed plays an important role in estimating and analysing the output variability, the author recommends considering cloud speed at different times of the day. For example, cloud speed should be calculated from 7:00 to 10:00 and then also using the time series 10:01 to 13:00 and 13:01 to 16:00 and so on.
- Using the Machine Learning Technique, a model can be built to describe the size and speed of clouds by collecting a massive amount of data from different PV plants and locations. This will help the planner obtain more information when building new PV plants and appropriate storage systems.

References

- [1] F. Gaspari, "Aviation and Environmental Protection after the 2015 Paris Agreement: From Regulatory Unilateralism toward International Cooperation," *Issues Aviation L. & Pol'y*, vol. 15, pp. 347, 2015.
- [2] *CLEAN ENERGY AUSTRALIA REPORT*, 2019.
- [3] M. S. Andrew Blakers, Bin Lu, *Australia: the renewable energy superstar*, 2019.
- [4] J. H. P.W. Graham, J. Foster, et al., *GenCost 2018*, <https://www.csiro.au/~media/news-releases/2018/renewables-cheapest-new-power/gencost2018.pdf?la=en&hash=9a292e508f96eec337f1bfff4acdb54bc2da5cef>, 2018.
- [5] M. J. Reno, C. W. Hansen, and J. S. Stein, "Global horizontal irradiance clear sky models: Implementation and analysis," *SANDIA report SAND2012-2389*, 2012.
- [6] S. A. Mousavi Maleki, H. Hizam, and C. Gomes, "Estimation of hourly, daily and monthly global solar radiation on inclined surfaces: Models re-visited," *Energies*, vol. 10, no. 1, pp. 134, 2017.
- [7] S. Killinger, F. Braam, B. Müller, B. Wille-Hausmann, and R. McKenna, "Projection of power generation between differently-oriented PV systems," *Solar Energy*, vol. 136, pp. 153-165, 2016.
- [8] B. Elsinga, and W. van Sark, "Spatial power fluctuation correlations in urban rooftop photovoltaic systems," *Progress in Photovoltaics: Research and Applications*, vol. 23, no. 10, pp. 1390-1397, 2015.
- [9] R. Perez, *Wind Field and Solar Radiation Characterization and Forecasting: A Numerical Approach for Complex Terrain*: Springer, 2018.
- [10] M. David, F. H. R. Andriamasomanana, and O. Liandrat, "Spatial and temporal variability of PV output in an insular grid: Case of Reunion Island," *Energy Procedia*, vol. 57, pp. 1275-1282, 2014.
- [11] E. Arias-Castro, J. Kleissl, and M. Lave, "A Poisson model for anisotropic solar ramp rate correlations," *Solar Energy*, vol. 101, pp. 192-202, 2014.
- [12] M. Lave, and J. Kleissl, "Cloud speed impact on solar variability scaling—Application to the wavelet variability model," *Solar Energy*, vol. 91, pp. 11-21, 2013.
- [13] J. L. Bosch, Y. Zheng, and J. Kleissl, "Deriving cloud velocity from an array of solar radiation measurements," *Solar Energy*, vol. 87, pp. 196-203, 2013.
- [14] J. L. Bosch, and J. Kleissl, "Cloud motion vectors from a network of ground sensors in a solar power plant," *Solar Energy*, vol. 95, pp. 13-20, 2013.
- [15] V. P. Lonij, A. E. Brooks, A. D. Cronin, M. Leuthold, and K. Koch, "Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors," *Solar Energy*, vol. 97, pp. 58-66, 2013.
- [16] S. Pfenninger, and I. Staffell, "Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data," *Energy*, vol. 114, pp. 1251-1265, 2016.
- [17] H. A. H. Al-Hilfi, F. Shahnian and A. Abu-Siada, "Comparison of different correlational techniques in estimating the total generated power of neighboring photovoltaic systems," 2016 Australasian Universities Power Engineering Conference (AUPEC), Brisbane, QLD, Australia, 2016, pp. 1-6
- [18] H. A. Al-Hilfi, F. Shahnian, and A. Abu-Siada, "Enhancing the estimation of the overall produced power by several adjacent photovoltaic systems using existing correlational factors." 2018 IEEE 27th International Symposium on Industrial Electronics (ISIE) Cairns, QLD, Australia, 2018, pp 43-47.

- [19] C. Wan, J. Zhao, Y. Song, Z. Xu, J. Lin, and Z. Hu, "Photovoltaic and solar power forecasting for smart grid energy management," *CSEE Journal of Power and Energy Systems*, vol. 1, no. 4, pp. 38-46, 2015.
- [20] T. Kato, T. Inoue, and Y. Suzuoki, "Estimation of total power output fluctuation of high penetration photovoltaic power generation system." IEEE Power and Energy Society General Meeting, Detroit, MI, USA, 2011, pp. 1-7.
- [21] J. K. Kaldellis, M. Kapsali, and K. A. Kavadias, "Temperature and wind speed impact on the efficiency of PV installations. Experience obtained from outdoor measurements in Greece," *Renewable Energy*, vol. 66, pp. 612-624, 2014.
- [22] W. Charles Lawrence Kamuyu, J. Lim, C. Won, and H. Ahn, "Prediction model of photovoltaic module temperature for power performance of floating PVs," *Energies*, vol. 11, no. 2, pp. 447, 2018.
- [23] S. Amely Jumaat, F. Crocker, M. Helmy Abd Wahab, N. Hanis Mohammad Radzi, and M. Fakri Othman, "Prediction of Photovoltaic (PV) Output Using Artificial Neural Network (ANN) Based on Ambient Factors," *Journal of Physics: Conference Series*, vol. 1049, p. 012088, 2018
- [24] A. Hammer, J. Kühnert, K. Weinreich, and E. Lorenz, "Short-term forecasting of surface solar irradiance based on Meteosat-SEVIRI data using a nighttime cloud index," *Remote Sensing*, vol. 7, no. 7, pp. 9070-9090, 2015.
- [25] D. M. Willy, "Power output variability and metrics for utility-scale solar photovoltaics power plant," Northern Arizona University, 2012.
- [26] J. Susanto, F. Shahnia, and D. Ludwig, "A framework to technically evaluate integration of utility-scale photovoltaic plants to weak power distribution systems," *Applied energy*, vol. 231, pp. 207-221, 2018.
- [27] "UQ solar photovoltaic data, Website of the University of Queensland. Reterived 2018," <http://solar.uq.edu.au/user/reportpower.php>.
- [28] M. Lave, J. Kleissl, and J. S. Stein, "A wavelet-based variability model (WVM) for solar PV power plants," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 2, pp. 501-509, 2013.
- [29] A. D. Mills, and R. H. Wiser, "Implications of geographic diversity for short-term variability and predictability of solar power." pp. 1-9.
- [30] J. Stein, C. Hansen, and M. J. Reno, *The variability index: A new and novel metric for quantifying irradiance and PV output variability*, Sandia National Laboratories, 2012.
- [31] M. Lave, M. J. Reno, and R. J. Broderick, "Characterizing local high-frequency solar variability and its impact to distribution studies," *Solar Energy*, vol. 118, pp. 327-337, 2015.
- [32] A. Gagné, D. Turcotte, N. Goswamy, and Y. Poissant, "High resolution characterisation of solar variability for two sites in Eastern Canada," *Solar Energy*, vol. 137, pp. 46-54, 2016.
- [33] M. Lave, R. J. Broderick, and M. J. Reno, "Solar variability zones: Satellite-derived zones that represent high-frequency ground variability," *Solar Energy*, vol. 151, pp. 119-128, 2017.
- [34] J. Marcos, L. Marroyo, E. Lorenzo, D. Alvira, and E. Izco, "From irradiance to output power fluctuations: the PV plant as a low pass filter," *Progress in Photovoltaics: Research and Applications*, vol. 19, no. 5, pp. 505-510, 2011.
- [35] K. Otani, J. Minowa, and K. Kurokawa, "Study on areally-totalized PV systems," *Solar Energy Materials and Solar Cells*, vol. 47, no. 1-4, pp. 281-288, 1997.
- [36] A. E. Curtright, and J. Apt, "The character of power output from utility-scale photovoltaic systems," *Progress in Photovoltaics: Research and Applications*, vol. 16, no. 3, pp. 241-247, 2008.

- [37] M. Lave, and J. Kleissl, "Solar variability of four sites across the state of Colorado," *Renewable Energy*, vol. 35, no. 12, pp. 2867-2873, 2010.
- [38] T. Kato, S. Kumazawa, N. Honda, M. Koaizawa, S. Nishino, and Y. Suzuoki, "Development of estimation method of spatial average irradiance fluctuation characteristics considering smoothing effect around observation point," *Electronics and Communications in Japan*, vol. 97, no. 11, pp. 1-9, 2014.
- [39] K. Klima, and J. Apt, "Geographic smoothing of solar PV: results from Gujarat," *Environmental Research Letters*, vol. 10, no. 10, pp. 104001, 2015.
- [40] J. Marcos, Í. de la Parra, M. García, and L. Marroyo, "Simulating the variability of dispersed large PV plants," *Progress in Photovoltaics: Research and Applications*, vol. 24, no. 5, pp. 680-691, 2016.
- [41] M. Anvari, G. Lohmann, M. Wächter, P. Milan, E. Lorenz, D. Heinemann, M. R. R. Tabar, and J. Peinke, "Short term fluctuations of wind and solar power systems," *New Journal of Physics*, vol. 18, no. 6, pp. 063027, 2016.
- [42] A. Woyte, V. Van Thong, K. Purchala, R. Belmans, and J. Nijs, "Quantifying the occurrence and duration of power fluctuations introduced by photovoltaic systems." p. 7 pp. Vol. 3.
- [43] A. Woyte, R. Belmans, and J. Nijs, "Fluctuations in instantaneous clearness index: Analysis and statistics," *Solar Energy*, vol. 81, no. 2, pp. 195-206, 2007.
- [44] O. Perpiñán, and E. Lorenzo, "Analysis and synthesis of the variability of irradiance and PV power time series with the wavelet transform," *Solar Energy*, vol. 85, no. 1, pp. 188-197, 2011.
- [45] M. Lave, J. Kleissl, and J. S. Stein, "A wavelet-based variability model (WVM) for solar PV power plants," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 2, pp. 501-509, 2012.
- [46] B. L. Madhavan, H. Deneke, J. Witthuhn, and A. Macke, "Multiresolution analysis of the spatiotemporal variability in global radiation observed by a dense network of 99 pyranometers," *Atmospheric Chemistry and Physics*, vol. 17, no. 5, pp. 3317-3338, 2017.
- [47] N. Silsirivanich, "Fluctuation characteristics effect analysis of solar irradiance data by wavelet transform," *Energy Procedia*, vol. 138, pp. 301-306, 2017.
- [48] M. Emmanuel, R. Rayudu, and I. Welch, "Impacts of power factor control schemes in time series power flow analysis for centralized PV plants using wavelet variability model," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 3185-3194, 2017.
- [49] M. Lave, J. Kleissl, and E. Arias-Castro, "High-frequency irradiance fluctuations and geographic smoothing," *Solar Energy*, vol. 86, no. 8, pp. 2190-2199, 2012.
- [50] A. R. Dyreson, E. R. Morgan, S. H. Monger, and T. L. Acker, "Modeling solar irradiance smoothing for large PV power plants using a 45-sensor network and the Wavelet Variability Model," *Solar Energy*, vol. 110, pp. 482-495, 2014.
- [51] J. Marcos, L. Marroyo, E. Lorenzo, D. Alvira, and E. Izco, "Power output fluctuations in large scale PV plants: one year observations with one second resolution and a derived analytic model," *Progress in Photovoltaics: Research and Applications*, vol. 19, no. 2, pp. 218-227, 2011.
- [52] S. H. Monger, E. R. Morgan, A. R. Dyreson, and T. L. Acker, "Applying the kriging method to predicting irradiance variability at a potential PV power plant," *Renewable energy*, vol. 86, pp. 602-610, 2016.
- [53] A. Madanchi, M. Absalan, G. Lohmann, M. Anvari, and M. R. R. Tabar, "Strong short-term non-linearity of solar irradiance fluctuations," *Solar Energy*, vol. 144, pp. 1-9, 2017.
- [54] K. Adye, N. Pearre, and L. Swan, "Contrasting distributed and centralized photovoltaic system performance using regionally distributed pyranometers," *Solar Energy*, vol. 160, pp. 1-9, 2018.

- [55] J. Gari da Silva Fonseca Junior, Y. Nishitsuji, Y. Udagawa, C. T. Urabe, T. Saito, F. Uno, H. Ohtake, T. Oozeki, and K. Ogimoto, "Regional photovoltaic power fluctuations within frequency regulation control time frames: A study with high-resolution data," *Progress in Photovoltaics: Research and Applications*, vol. 26, no. 6, pp. 402-413, 2018.
- [56] Y. Tang, J. W. Cheng, Q. Duan, C. W. Lee, and J. Zhong, "Evaluating the variability of photovoltaics: A new stochastic method to generate site-specific synthetic solar data and applications to system studies," *Renewable energy*, vol. 133, pp. 1099-1107, 2019.
- [57] A. Murata, and K. Otani, "An analysis of time-dependent spatial distribution of output power from very many PV power systems installed on a nation-wide scale in Japan," *Solar energy materials and solar cells*, vol. 47, no. 1-4, pp. 197-202, 1997.
- [58] E. Wiemken, H. Beyer, W. Heydenreich, and K. Kiefer, "Power characteristics of PV ensembles: experiences from the combined power production of 100 grid connected PV systems distributed over the area of Germany," *Solar energy*, vol. 70, no. 6, pp. 513-518, 2001.
- [59] B. M. Mazumdar, M. Saquib, and A. K. Das, "An empirical model for ramp analysis of utility-scale solar PV power," *Solar Energy*, vol. 107, pp. 44-49, 2014.
- [60] R. Van Haaren, M. Morjaria, and V. Fthenakis, "Empirical assessment of short-term variability from utility-scale solar PV plants," *Progress in Photovoltaics: Research and Applications*, vol. 22, no. 5, pp. 548-559, 2014.
- [61] P. Schober, C. Boer, and L. A. Schwarte, "Correlation coefficients: appropriate use and interpretation," *Anesthesia & Analgesia*, vol. 126, no. 5, pp. 1763-1768, 2018.
- [62] V. DeGhett, "Effective use of Pearson's product-moment correlation coefficient: an additional point," *Animal Behaviour*, no. 98, pp. e1-e2, 2014.
- [63] N. O'Rourke, and L. Hatcher, *A step-by-step approach to using SAS for factor analysis and structural equation modeling*: Sas Institute, 2013.
- [64] A. Murata, H. Yamaguchi, and K. Otani, "A method of estimating the output fluctuation of many photovoltaic power generation systems dispersed in a wide area," *Electrical Engineering in Japan*, vol. 166, no. 4, pp. 9-19, 2009.
- [65] A. Mills, and R. Wiser, *Implications of wide-area geographic diversity for short-term variability of solar power*, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 2010.
- [66] T. E. Hoff, and R. Perez, "Modeling PV fleet output variability," *Solar Energy*, vol. 86, no. 8, pp. 2177-2189, 2012.
- [67] S. Sayeef et al., "Solar intermittency: Australia's clean energy challenge", CSIRO, Australia, 2012.
- [68] M. Lave, and J. Kleissl, "Testing a wavelet-based variability model (WVM) for solar PV power plants." pp. 1-6.
- [69] A. D. Darji, S. S. Kushwah, S. N. Merchant, and A. N. Chandorkar, "High-performance hardware architectures for multi-level lifting-based discrete wavelet transform," *EURASIP Journal on Image and Video Processing*, vol. 2014, no. 1, pp. 47, 2014.
- [70] L. Debnath, and F. A. Shah, *Wavelet transforms and their applications*: Springer, 2015.
- [71] R. Pathak, and A. Singh, "Distributional wavelet transform," *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, vol. 86, no. 2, pp. 273-277, 2016.
- [72] M. Mochizuki, and S. Moro, "Steady-state analysis using Haar wavelet transform in power electronics circuits including nonlinear elements." pp. 483-486.

- [73] M. S. Basha, and M. Ramakrishnan, "Color image contrast enhancement using daubechies d4 wavelet and luminance analysis," *International Journal of Computer Applications*, vol. 86, no. 6, 2014.
- [74] F. Wasilewski. "Wavelets- PyWavelets Documentation," 1/6/2018, 2018; <http://pywavelets.readthedocs.io/en/latest/ref/wavelets.html>
- [75] J. Freeman, D. Freestate, W. Hobbs, and C. Riley, "Using measured plane-of-array data directly in photovoltaic modeling: Methodology and validation." *IEEE 43rd Photovoltaic Specialists Conference (PVSC)*, Portland, OR, USA, 2016, pp. 2653-2656,
- [76] E. Skoplaki, A. Boudouvis, and J. Palyvos, "A simple correlation for the operating temperature of photovoltaic modules of arbitrary mounting," *Solar Energy Materials and Solar Cells*, vol. 92, no. 11, pp. 1393-1402, 2008.
- [77] M. Lave, W. Hayes, A. Pohl, and C. W. Hansen, "Evaluation of global horizontal irradiance to plane-of-array irradiance models at locations across the United States," *IEEE journal of Photovoltaics*, vol. 5, no. 2, pp. 597-606, 2015.
- [78] "PV System Performance: GHI to PoA, EcoSmart™ Solar," 2/5/2019; <https://ecosmartsun.com/pv-system-performance-3/pv-system-performance-ghi-to-poa>.
- [79] "NOAA Solar Position Calculator," 27/6/2017, 27/6/2017.
- [80] "MATLAB® PV Library," 27/6/2017, 27/6/2017.
- [81] "CAMS McClear," 27/6/2017, 27/6/2017.
- [82] C. Zeng, H. Lin, Q. Jiang, and M. Xu, "QRS Complex Detection Using Combination of Mexican-hat Wavelet and Complex Morlet Wavelet," *JCP*, vol. 8, no. 11, pp. 2951-2958, 2013.
- [83] A. Ashrafian, M. Mirsalim, and M. A. Masoum, "An adaptive recursive wavelet based algorithm for real-time measurement of power system variables during off-nominal frequency conditions," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 3, pp. 818-828, 2017.
- [84] F. F. El-Sousy, "Intelligent optimal recurrent wavelet Elman neural network control system for permanent-magnet synchronous motor servo drive," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 4, pp. 1986-2003, 2012.
- [85] T. Ji, Z. Lu, and Q. Wu, "Detection of power disturbances using morphological gradient wavelet," *Signal Processing*, vol. 88, no. 2, pp. 255-267, 2008.
- [86] M. Lave, J. Kleissl, A. Ellis, and F. Mejia, "Simulated PV power plant variability: Impact of utility-imposed ramp limitations in Puerto Rico." *IEEE 39th Photovoltaic Specialists Conference (PVSC)*, Tampa, FL, USA, 2013, pp. 1817-1821.
- [87] M. Lave, and J. Kleissl, "Testing a wavelet-based variability model (WVM) for solar PV power plants." *IEEE Power and Energy Society General Meeting*, San Diego, CA, USA, 2012, pp. 1-6.
- [88] P. Loutzenhiser, H. Manz, C. Felsmann, P. Strachan, T. Frank, and G. Maxwell, "Empirical validation of models to compute solar irradiance on inclined surfaces for building energy simulation," *Solar Energy*, vol. 81, no. 2, pp. 254-267, 2007.
- [89] T. Schmidt, "Short-term solar forecasting based on sky images for the integration of higher shares of PV generation in remote electricity networks," in *16th World renewable energy congress (WREC)*, Perth, 2017.
- [90] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker, and T. E. Hoff, "Short-term irradiance variability: Preliminary estimation of station pair correlation as a function of distance," *Solar Energy*, vol. 86, no. 8, pp. 2170-2176, 2012.
- [91] A. g. B. o. Meterology. "Wind Roses," 12/4/2017., 2017.; http://www.bom.gov.au/climate/averages/wind/wind_rose.shtml.

- [92] C. F. Coimbra, J. Kleissl, and R. Marquez, "Overview of solar-forecasting methods and a metric for accuracy evaluation," *Solar energy forecasting and resource assessment*, pp. 171-194, 2013.
- [93] M. J. Alizdeh, P. M. Joneyd, M. Motahhari, F. Ejlali, and H. Kiani, "A wavelet-ANFIS model to estimate sedimentation in dam reservoir," *International Journal of Computer Applications*, vol. 114, no. 9, pp. 19-25, 2015.
- [94] "Fuzzy logic toolbox user's guide," *The Mathworks Inc*, no. 2018, 2018.
- [95] A. M. Abdulshahed, A. P. Longstaff, and S. Fletcher, "The application of ANFIS prediction models for thermal error compensation on CNC machine tools," *Applied Soft Computing*, vol. 27, pp. 158-168, 2015.
- [96] F. Üneş, D. Joksimovic, and O. Kisi, "Plunging flow depth estimation in a stratified dam reservoir using neuro-fuzzy technique," *Water resources management*, vol. 29, no. 9, pp. 3055-3077, 2015.
- [97] Y. Seo, S. Kim, O. Kisi, and V. P. Singh, "Daily water level forecasting using wavelet decomposition and artificial intelligence techniques," *Journal of Hydrology*, vol. 520, pp. 224-243, 2015.
- [98] H. A. H. Al-Hilfi, F. Shahniah and A. Abu-Siada, "Employing a Gene Expression-based Technique to Improve the Accuracy of Estimating the Total Generated Power by Neighboring Photovoltaic Systems," 2019 9th International Conference on Power and Energy Systems (ICPES), Perth, WA, Australia, 2019, pp. 1-5
- [99] H. Moghaddasi, I. Mahmoudi, and S. Sajadi, "Comparing the Efficiency of Artificial Neural Network and Gene Expression Programming in Predicting Coronary Artery Disease," *J. Health Med. Inform*, vol. 8, no. 2, 2017.
- [100] E. Sadrossadat, B. Ghorbani, R. Oskooei, and M. Kaboutari, "Use of adaptive neuro-fuzzy inference system and gene expression programming methods for estimation of the bearing capacity of rock foundations," *Engineering Computations*, 2018.
- [101] T. Wang, J. S. Dyer, and W. J. Hahn, "Sensitivity analysis of decision making under dependent uncertainties using copulas," *EURO Journal on Decision Processes*, vol. 5, no. 1-4, pp. 117-139, 2017.
- [102] S. H. A. Kaboli, A. Fallahpour, N. Kazemi, J. Selvaraj, and N. A. Rahim, "An expression-driven approach for long-term electric power consumption forecasting," *American Journal of Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 16, 2017.
- [103] K. Karabacak, and N. Cetin, "Artificial neural networks for controlling wind-PV power systems: A review," *Renewable and Sustainable Energy Reviews*, vol. 29, pp. 804-827, 2014.
- [104] B. Alizadeh, and S. Jadid, "A dynamic model for coordination of generation and transmission expansion planning in power systems," *International Journal of Electrical Power & Energy Systems*, vol. 65, pp. 408-418, 2015.
- [105] S. S. S. Hosseini, and A. H. Gandomi, "Short-term load forecasting of power systems by gene expression programming," *Neural Computing and Applications*, vol. 21, no. 2, pp. 377-389, 2012.
- [106] A. Abu-Siada, "Improved Consistent Interpretation Approach of Fault Type within Power Transformers Using Dissolved Gas Analysis and Gene Expression Programming," *Energies*, vol. 12, no. 4, pp. 730, 2019.
- [107] A. Abu-Siada, "Correlation of furan concentration and spectral response of transformer oil-using expert systems," *IET science, measurement & technology*, vol. 5, no. 5, pp. 183-188, 2011.
- [108] C. Ferreira, *Gene expression programming: mathematical modeling by an artificial intelligence*: Springer, 2006.

- [109] A. Abu-Siada, L. S. Pin, and S. Islam, "A novel application of Gene Expression Programming in transformer diagnostics." Australasian Universities Power Engineering Conference, Sydney, NSW, Australia, 2008, pp. 1-5.
- [110] J. M. Ripalda, D. Chemisana, J. M. Llorens, and I. García, "Impact of spectral variations on photovoltaic energy production: A case study in the United States," *arXiv preprint arXiv:2003.08871*, 2020.
- [111] S. Hojjati, H.-Y. Jeong, and S. Jeon, "Prediction Model for Specific Cutting Energy of Pick Cutters Based on Gene Expression Programming and Particle Swarm Optimization," vol. 28, pp. 651-669, 12/01, 2018.
- [112] H. Al-Hilfi, F. Shahnia, and A. Abu-Siada, "An Improved Technique to Estimate the Total Generated Power by Neighboring Photovoltaic Systems Using Single-point Irradiance Measurement and Correlational Models," *IEEE Transactions on Industrial Informatics*, 2019.
- [113] G. M. Lohmann, "Irradiance variability quantification and small-scale averaging in space and time: A short review," *Atmosphere*, vol. 9, no. 7, pp. 264, 2018.
- [114] A. Solgi, V. Nourani, and A. Pourhaghi, "Forecasting daily precipitation using hybrid model of wavelet-artificial neural network and comparison with adaptive neurofuzzy inference system (case study: Verayneh station, Nahavand)," *Advances in Civil Engineering*, vol. 2014, 2014.

Publications Arising from this Thesis

Journal articles

- 1- **H Al-Hilfi**, F Shahniah, and A Abu-Siada (2019) An improved technique to estimate the total generated power by neighbouring photovoltaic systems using single-point irradiance measurement and correlational models, *IEEE Transactions on Industrial Informatics* doi:10.1109/TII.2019.2941218
- 2- **H Al-Hilfi**, A Abu-Siada, and F Shahniah, An ANFIS-Wavelet Technique to improve the Estimation Accuracy of the Output Power of Neighbouring PVs during Cloud Events, *Energies*, vol. 13, no. 7, pp. 1613, 2020.
- 3- **H Al-Hilfi**, A Abu-Siada, and F Shahniah, Estimating PVs Generated Power during Cloudy days using Gene Expression Programming, Submitted to the *IEEE Journal of Photovoltaics*, vol. 11, Issue: 1, pp 185 – 194, 2021. doi: 10.1109/JPHOTOV.2020.3029217
- 4- **H Al-Hilfi**, F Shahniah, and A Abu-Siada, A Gene Expression Technique-based Approach to Improve the Accuracy of Estimating the Total Generated Power by Neighbouring Photovoltaic Systems, Submitted to *IET renewable Power Generation*, Vol. 14, Issue 18, pp 3715 – 3723, 2020. doi: 10.1049/iet-rpg.2020.0279

Conference papers

- 5- **H Al-Hilfi**, F Shahniah and A Abu-Siada, “Employing a Gene Expression-based Technique to Improve the Accuracy of Estimating the Total Generated Power by Neighbouring Photovoltaic Systems,” 9th International Conference on Power and Energy Systems (ICPES), Perth, Dec. 2019.
- 6- **H Al-Hilfi**, F Shahniah and A Abu-Siada, “Enhancing the estimation of the overall produced power by several adjacent photovoltaic systems using existing correlational factors,” 27th IEEE International Symposium on Industrial Electronics (ISIE), pp.1-5, Cairns, Australia, June 2018.
- 7- **H Al-Hilfi**, A Abu-Siada and F Shahniah, “Comparison of different

correlational techniques in estimating the total generated power of neighbouring photovoltaic systems,” 26th Australasian Universities Power Engineering Conference (AUPEC), pp.1-5, Brisbane, Australia, Sept 2016.