energies

MDPI

Article

Combined ANFIS–Wavelet Technique to Improve the Estimation Accuracy of the Power Output of Neighboring PV Systems during Cloud Events

Hasanain A. H. Al-Hilfi [1,2], Ahmed Abu-Siada [1,*] and Farhad Shahnia [3]

1 School of Electrical Engineering and Computing, Curtin University, Perth 6102, Australia; h.al-hilfi@postgrad.curtin.edu.au
2 Computer Center, Basrah University, Basrah 61028, Iraq
3 School of Engineering and Information Technology, Murdoch University, Murdoch 6150, Australia; F.Shahnia@murdoch.edu.au
* Correspondence: A.AbuSiada@curtin.edu.au

Received: 21 February 2020; Accepted: 26 March 2020; Published: 1 April 2020

Abstract: The short-term variability of photovoltaic (PV) system-generated power due to ambient conditions, such as passing clouds, represents a key challenge for network planners and operators. Such variability can be reduced using a geographical smoothing technique based on installing multiple PV systems over certain locations at distances of meters to kilometers. To accurately estimate the PV system's generated power during cloud events, a variability reduction index (*VRI*), which is a function of several parameters, should be calculated precisely. In this paper, the Wavelet Transform Technique (*WTT*) along with Adaptive Neuro Fuzzy Inference System (ANFIS) are used to develop new models to estimate the PV system's power output during cloud events. In this context, irradiance data collected from one PV system along with other parameters, including ambient conditions, were used to develop the proposed models. Ultimately, the models were validated through their application on a 0.7 km^2 PV plant with 16 rooftop PV systems in Brisbane, Australia.

Keywords: photovoltaic system; geographic smoothing; variability reduction index; ANFIS; wavelet transform

1. Introduction

Global warming due to greenhouse gas emissions produced by conventional fossil fuel-based power plants has urged all nations to invest in renewable energy, such as photovoltaic (PV) systems and wind energy [1]. According to [2], the generated electricity in Australia is divided into 79% fossil fuel-based generation whereas 21% is generated from renewable energy resources. Renewable energy continues to grow rapidly in Australia, which is planned to reach 50% of the overall domestic electricity production in 2024 and 100% in the year 2032 [3]. Although there are several advantages in using PV systems, including easy implementation in 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 it varies with short- (cloud movements) and long-term (irradiance daily and seasonal) variations. Long-term variability can be easily planned for based on the well-known daily and seasonal solar irradiance in various geographical areas. On the other hand, short-term irradiance fluctuation due to the unexpected passing of clouds of unknown size, direction and speed, may result in frequency instability due to the reduction in the generated power [4–6]. Estimating such power is one of the main challenges faced by network operators to prevent frequency instability due to such short-term power reduction, which is why many adopt a suitable backup energy storage system. According to utilities guidelines, the PV system's power output

reduction should not be more than 10% of the PV system's capacity over a minute [7]. Adopting an energy storage system to meet this requirement may be an expensive option for large PV plants [7,8]. Another cost-effective solution is through allocating the PV systems over a large geographical area. The robustness of this solution is shown in Figure 1, which reveals that the irradiance is getting smoother by increasing the number of PV systems over a geographical area with a separation distance of 20 km to 450 km [7].

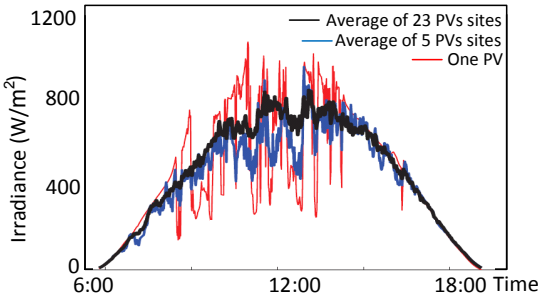

Figure 1. Average of the solar irradiance variability captured by 1, 5 and 23 photovoltaic (PV) system sites [7].

Geographic smoothing or the variability reduction index (*VRI*) is discussed in several studies in the literature. In such studies, different methods and metrics are presented to estimate and analyze the short-term variability in the PV system's power output. For instance, the correlation coefficient and standard deviation are employed in [9–14] while [15–17] utilized the variability score and variability index as mathematical metrics to calculate the power output variability of PV systems. Other studies introduced different metrics to investigate the power output variability, such as the Multifractal Detrended Fluctuation Analysis (MFDFA), which is a time series modeling method especially used for extreme ambient events [18]. Similarly, [19] adopted the Fourier analysis method to characterize the variability using power spectral density. In such studies, data used to calculate the PV system's power variability are obtained using generated synthetic PV system data based on real measurements as presented in [1,20]. Few studies have been introduced to estimate the power output variability using the Wavelet Transform Technique based on measured data of only one PV sensor and by considering the PV system's locations, cloud speed and irradiance time series [21–23]. Table 1 summarizes various methods presented in the literature along with the key finding of each study.

Although several models have been introduced to estimate the power output for a group of PV systems distributed within meters to kilometers, the estimation error calculated so far ranges between 8% to 20%, which reveals the relatively poor accuracy of such techniques [23]. As such, developing a new model with a higher estimation accuracy would be a significant contribution to this area of research. Such a high estimation accuracy model can be employed in power networks of high PV system penetration for better reliability, control and asset management, which is the main aim of this paper. In this paper, a combined *WTT* and Adaptive Neuro Fuzzy Inference System (ANFIS)-based technique with a higher estimation accuracy than existing models are presented. The proposed technique is employed to estimate the power output of a group of rooftop PV systems distributed over 0.7 km^2 area during variability events. The robustness of the proposed technique is validated through its application on real data collected from several PV systems distributed within a few hundred meters to 1 km distance.

In the remaining sections of this paper, brief reviews on the *VRI*, *WTT* and ANFIS are presented in Section 2. The proposed model methodology is presented in Section 3. Results and discussions are presented in Section 4. Section 5 presents a sensitivity analysis of the parameters used in the developed model followed by the main conclusions.

2. Variability and Correlation Coefficient

The *VRI* represents the ratio of the power variance of one PV system to the entire PV plant at a specific time scale [23] and is mathematically expressed as

$$VRI = \frac{{N_{\text{syst}}}^2}{(\rho_{n,m})^{\text{overall}}} \tag{1}$$

The correlation coefficient in Equation (1) is a measure of the relationship of two random variables [24]. A number of correlation models are presented in the literature [23,25–28] of which Pearson's is considered as one of the popular correlation models [25]. Reference [26] presented an empirical correlation coefficient ($\rho^{\text{Hf}}_{m,n}$, indicated by the Hoff model in the course of this study) that considers the timescale (T) and the distance ($d_{m,n}$) between two PV systems' location as below:

$$\rho^{\text{Hf}}_{m,n} = \frac{V_{clo} \times T}{d_{m,n} + V_{clo} \times T} \tag{2}$$

Perez [27] and Lave [23] ($\rho^{\text{Pz}}_{m,n}$ and $\rho^{\text{Lv}}_{m,n}$ refer to the Perez and Lave models in the rest of this paper) have also introduced empirical exponential models as in Equations (3) and (4), respectively.

$$\rho^{\text{Pz}}_{m,n} = \exp\left(-\frac{1.07 d_{m,n}}{V_{clo} \times T}\right) \tag{3}$$

$$\rho^{\text{Lv}}_{m,n} = \exp\left(\frac{-d_{m,n}}{0.5\ V_{clo} \times T}\right) \tag{4}$$

Reference [28] presented another mathematical model (referred to ACM model) by considering the cloud direction and size as in Equation (5).

$$\rho^{\text{ACM}}_{m,n} = \frac{2 \times 0.5^{(2-b_1)} - 0.5^{(2-b_2)} - 0.5^{(2-b_3)}}{2 \times \left(0.5 - 0.5^{(2-b_4)}\right)} \tag{5}$$

where b_1 to b_2 are normalized parameters considering V_{clo}, $d_{m,n}$ and cloud direction.

In addition, Reference [29] presented a *VRI* technique (referred to the *VRI*–GEP model in the rest of this study) by using the maximum values of the existing *VRI* models [7] and a genetic algorithm technique called Gene Expression Programming (GEP).

Table 1. Various estimation methods published in the literature.

Reference	Investigated System	Method	Findings
[9]	4 PV systems, Spain and Portugal with 200 to 500 km^2 and 1 and 5 min resolution.	Standard deviation, Hierarchical clustering method.	Classified the variability into low pressure, transient, local and high-pressure modes
[10]	24 PV systems, Canada, Germany and USA with 0.1–1 s resolution	Standard deviation	1-s resolution is a marking point. Less resolution is difficult to manage and control
[11]	99 PV systems, UK, 10 km distance and 1-s resolution	Average data and standard deviation	Distinguished three cases, clear, overcast and partly cloudy sky conditions.
[12]	16 PV systems Canada, few meters, 1500 km distance 1-h resolution	Correlation model, standard deviation and energy simulation program.	Increasing the area mitigates the power output variability.

Table 1. *Cont.*

Reference	Investigated System	Method	Findings
[13]	89 PV systems, UK, 2 to 50 km distance 1-h resolution.	Standard deviation to calculate the variability	Aggregated PV systems can mitigate the variability
[14]	20 PV systems in the USA, with 1-h time resolution.	Standard deviation and probability density function.	Model to calculate the variability using a clear-sky index.
[15]	9 PV systems, USA, 100 km^2 area, 1–30-s and 1-h resolution.	variability score and cumulative distribution function	New model to calculate the high frequencies of power variability
[16]	2 PV systems, Canada, with 41 sensors, distance of 400 km and 10-ms data time resolution.	Employed variability score and variability index.	Classified daily irradiance into 4 classes: clear, overcast, low and high frequencies solar variability.
[17]	10 PV systems in USA, with hundreds of km distance and 30-s time resolution	Variability score method	Variability score is a useful estimation metric. However, it is a poor predictor for PV tracking systems.
[18]	20 PV systems in Hawaii, 1 PV system in Spain, 1 PV system in Algeria, 1 PV system in Germany.	Correlation and the multifractal detrended fluctuation analysis. The study focused on temporal variability.	Variability is not stationary in which it was the highest in Hawaii and the lowest in Algeria.
[19]	50 PV systems, India, 400 km distance, 1-min resolution.	Fourier transform analysis and correlation coefficient.	Good smoothing in short-time events. Improvement for more added PV systems.
[1,20]	3 PV systems, Hong Kong, 14 to 18km distance and 1-min resolution.	Using system advisor model	An improved model to generate spatial synthetic PV data
[21]	99 PV systems, Germany, 10 to 12 km distance, 1-s. resolution.	Discrete wavelet transform model	Distinguished three conditions, clear, partly cloudy and overcast.
[22]	One PV plant, USA, 60 km. 1-s resolution	Discrete wavelet technique	Variability is higher at the end of the feeder.
[23]	550 PV systems, Japan and Nevada, 1–6 km^2 respectively, 1-s resolution.	Discrete wavelet transform model and variability reduction index.	Estimation of the power output with errors of 8% to 20%.

Thus, the *VRI* in Equation (1) can be calculated using any of the correlation coefficients presented in Equations (2) through (5) as below:

$$VRI_x = \frac{{N_{\text{syst}}}^2}{(\rho_{n,m})^{\text{overall}}} \tag{6}$$

where x refers to one of these four models.

3. Proposed Technique

The proposed technique is a combination of the *WTT* and the Adaptive Neuro-Fuzzy Inference System (ANFIS). Wavelet Transform is a signal processing technique that transforms periodic and

non-periodic time-series signals into the time-frequency domain [7]. Continuous Wavelet Transform (*CWT*) is used for a continuous time-series signals and is expressed as

$$CWT(a,f) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} \psi\left(\frac{t-f}{a}\right) \times y(t)\, dt \tag{7}$$

Because the irradiance time-series signal is discrete [23], Discrete Wavelet Transform (*DWT*) as in Equation (8) is more suitable for the analysis in this paper.

$$DWT^{\,mode-j}(j,c) = \frac{1}{\sqrt{T^{mode-j}}} \sum\nolimits_{k=1}^{\infty} \psi\left[\frac{k - c \times T^{mode-j}}{T^{mode-j}}\right] \times y[k] \tag{8}$$

where $y[k]$ is the discrete time-series signal of the PV power output, and a and c are scale and shift factors; respectively. The mother wavelet function could be one of Harr, Mexican Hat, Symlets, Coiflets, Biorthogonal, Meyer, Morlet and Daubechies [7]. The decomposed signals resulted from the *DWT* process can be divided into high-frequency (HF) and low-frequency (LF) modes. In this paper, a new technique based on ANFIS is proposed to capture the *DWT* features and model them [30].

Traditional mathematical modeling methods are inappropriate tools for dealing with stochastic and uncertain data, such as solar irradiance. An example of a more accurate modeling approach based on the integration of Artificial Neural Network (ANN) and Fuzzy Control System (FCS) is ANFIS, considered a powerful technique that can map the relationship between a set of input and output variables [31]. 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 organized structure with well-established fuzzy rules, reasoning and thinking [32]. 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 [33].

To build a model that aims at improving the estimation accuracy of the power output of several rooftop PV systems using ANFIS, it is important to identify the input parameters that influence the power output variability, such as ambient temperature (T_{em}), humidity (H_{u}), PV plant area (A^{pv}), the PV system site's power capacity (P^{cap}), cloud speed (V_{clo}) as well as the solar irradiance (I^{POA}). Some of these parameters are collected using a PV pyranometer located close to the PV system while other parameters are obtained based on the PV system's 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^{pv}), which is divided into 6 sub-areas with specific power capacity (P^{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 ($D_A^{\mathrm{cap}} = P^{\mathrm{cap}}/A^{\mathrm{pv}}$). 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 area layout 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_{clo} can be calculated using the method described in [7,24,34].
2. PV sensor data: This includes T_{em} and, H_{u} that can be directly collected from the pyranometer sensor installed in the PV system site.
3. *DWT* data: obtained using wavelet timescales (modes) analysis.

When the pyranometer is installed in the same tilt and orientation as the PV system, the output irradiance of the PV sensor will be a plane of the array (I^{POA}) [23]. Let $I^{\mathrm{POA}}[k]$ denote the output of the pyranometer, in which k is a time index.

To improve the ANFIS accuracy to estimate the power output, the time series (I^{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 of weather conditions over the day. For each sub-time series, the module power output can be calculated by [7]:

$$P_{\mathrm{POA}}^{\mathrm{mdu-in}}[k] = 0.12\, A^{\mathrm{pv}} I^{\mathrm{POA}}[k] \times \left[1 - 0.004\left(T_{\mathrm{em}}[k] + \frac{0.32\,\beta_{\mathrm{m}} \times I^{\mathrm{POA}}[k]}{8.91 + 2\,V_{\mathrm{clo}}} - 25\right)\right] \quad (9)$$

The average of the total power output of the PV plant ($P_{\mathrm{Eq}}^{\mathrm{mdu-out}}[k]$) is calculated by assuming each module is providing the same share of the overall power. Then, $P_{\mathrm{POA}}^{\mathrm{mdu-in}}[k]$ and $P_{\mathrm{Eq}}^{\mathrm{mdu-out}}[k]$ are normalized by dividing each by $P_{\mathrm{POA}}^{\mathrm{clr}}[k]$ (power output in a clear sky condition), which can be calculated by replacing $I^{\mathrm{POA}}[k]$ in Equation (9) by I_{clr}^{POA} [7],

$$I_{\mathrm{clr}}^{POA} = I_{\mathrm{clr}}^{\mathrm{d}} + I_{\mathrm{clr}}^{\mathrm{s}} + I_{\mathrm{clr}}^{\mathrm{g}} \quad (10)$$

The parameters in Equation (10) are calculated based on the data at the CAMS McClear website [35]. Then, Equations (11) and (12) can be derived:

$$P_{\mathrm{nr}}^{\mathrm{POA-in}}[k] = \frac{P_{\mathrm{POA}}^{\mathrm{mdu-in}}[k]}{P_{\mathrm{clr}}^{\mathrm{POA}}[k]} \quad (11)$$

$$P_{\mathrm{nr}}^{\mathrm{POA-out}}[k] = \frac{P_{\mathrm{Eq}}^{\mathrm{mdu-out}}[k]\ ]}{P_{\mathrm{clr}}^{\mathrm{POA}}[k]} \quad (12)$$

Using Equation (8), $P_{\mathrm{nr}}^{\mathrm{POA-in}}[k]$ and $P_{\mathrm{nr}}^{\mathrm{POA-out}}[k]$ can be decomposed into three timescale functions. Thus, *DWT* analysis of $P_{\mathrm{POA-in}}^{\mathrm{nrm}}[k]$ yields three high-frequency modes ($P_{\mathrm{mode}-j}^{\mathrm{HF-in}}[k]$) and three low-frequency modes ($P_{\mathrm{mode}-j}^{\mathrm{LF-in}}[k]$). In the same way, $P_{\mathrm{POA-out}}^{\mathrm{nrm}}[k]$ is analyzed into three high-frequency ($P_{\mathrm{mode}-j}^{\mathrm{HF-out}}[k]$) and low-frequency ($P_{\mathrm{mode}-j}^{\mathrm{LF-out}}[k]$) 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 Equation (13) below [36].

$$P_{\mathrm{nr-out}}^{\mathrm{Eq-ANFIS}} = inv(P_{\mathrm{mode}-1}^{\mathrm{HF-ANFIS}}[k]) + inv(P_{\mathrm{mode}-2}^{\mathrm{HF-ANFIS}}[k] + inv(P_{\mathrm{mode}-3}^{\mathrm{HF-ANFIS}}[k] + inv\left(P_{\mathrm{mode}-3}^{\mathrm{LF-ANFIS}}[k]\right) \quad (13)$$

$P_{\mathrm{mode}-j}^{\mathrm{HF-in}}[k]$ and $P_{\mathrm{mode}-j}^{\mathrm{LF-in}}[k]$ along with the parameters mentioned in points 1 and 2 above will be the ANFIS input parameters, whereas $P_{\mathrm{mode}-j}^{\mathrm{HF-out}}[k]$ and $P_{\mathrm{mode}-j}^{\mathrm{LF-out}}[k]$ are the outputs, as shown in Table 2. In Table 2, 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.

Table 2. Adaptive Neuro Fuzzy Inference System (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}), $V_{\mathrm{clo}} T_{\mathrm{em}}[k]$, $H_{\mathrm{u}}[k]$	$P_{\mathrm{mode}-1}^{\mathrm{HF-in}}[k]$	$P_{\mathrm{mode}-1}^{\mathrm{HF-out}}[k]$	$P_{\mathrm{mode}-1}^{\mathrm{HF-ANFIS}}[k]$
	$P_{\mathrm{mode}-2}^{\mathrm{HF-in}}[k]$	$P_{\mathrm{mode}-2}^{\mathrm{HF-out}}[k]$	$P_{\mathrm{mode}-2}^{\mathrm{HF-ANFIS}}[k]$
	$P_{\mathrm{mode}-3}^{\mathrm{HF-in}}[k]$	$P_{\mathrm{mode}-3}^{\mathrm{HF-out}}[k]$	$P_{\mathrm{mode}-3}^{\mathrm{HF-ANFIS}}[k]$
	$P_{\mathrm{mode}-3}^{\mathrm{LF-in}}[k]$	$P_{\mathrm{mode}-3}^{\mathrm{LF-out}}[k]$	$P_{\mathrm{mode}-3}^{\mathrm{LF-ANFIS}}[k]$

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

Then, by applying the inverse *DWT* as in Equation (13), the estimation of the equivalent normalized power output of several rooftop PV systems distributed within a certain area is evaluated.

By multiplying the results of Equation (13) by $P_{\mathrm{clr}}^{\mathrm{POA}}[k]$ and the number of PV modules (N^{mod}), the equivalent power output seen by the entire PV plant can be estimated as follows:

$$P_{\mathrm{out}}^{\mathrm{Eq-ANFIS}} = P_{\mathrm{nr-out}}^{\mathrm{Eq-ANFIS}} \times P_{\mathrm{clr}}^{\mathrm{POA}}[k] \times N^{\mathrm{mod}} \tag{14}$$

Figure 2 illustrates the flowchart of the main steps of the ANFIS–*DWT* technique proposed in this paper.

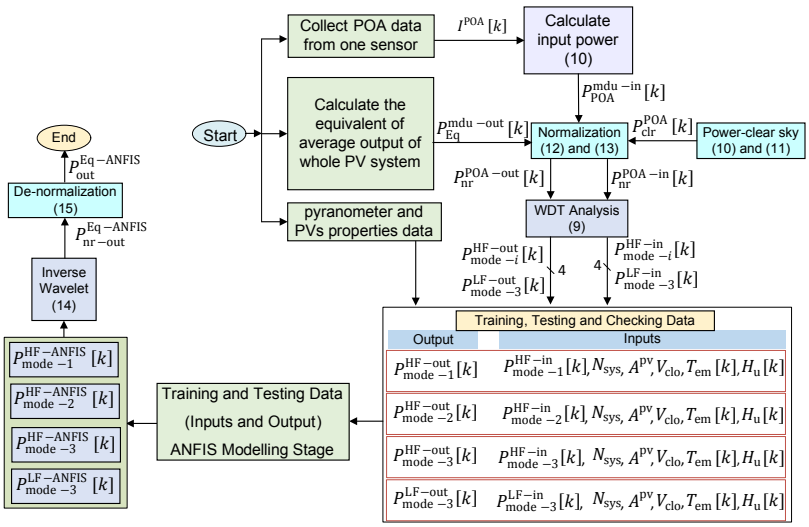

Figure 2. Proposed methodology for the proposed ANFIS–*DWT* technique.

Because variability is the main concern by power operators and planners, the variability power index (V_{pi}) has been introduced as a mathematical tool to show the amount of the power fluctuation at each *DWT* mode. First, $P_{\mathrm{nr-out}}^{\mathrm{Eq-ANFIS}}$ is analyzed into six timescales, namely 2, 4, 8, 16, 32 and 64 min, using the *DWT* method. To calculate the power at each timescale, the wavelet periodogram has been used as below [37]:

$$I_{\mathrm{w}}[k] = \frac{\left|P_{\mathrm{mode}-j}^{\mathrm{HF-ANFIS}}[k]\right|^2}{T^{\mathrm{mode}-j}} \tag{15}$$

then,

$$V_{pi}\left(T^{\mathrm{mode-j}}\right) = \frac{\int_0^{T_L} I_{\mathrm{w}}[k]dk}{T_{\mathrm{L}}} \tag{16}$$

According to [37], the variability reduction index (*VRI*) can be by calculated using Equation (17):

$$VRI = \frac{V_{\mathrm{pi}}\left(T^{\mathrm{mode}-j}\right)_{\mathrm{py}}}{V_{\mathrm{pi}}\left(T^{\mathrm{mode}-j}\right)_{\mathrm{ave}}} \tag{17}$$

4. Performance Evaluation of the Proposed Technique

Data of 445 days in the years 2016 and 2017 with different amounts of variabilities, where days with a completely clear sky were eliminated, were used to develop the proposed model. Data were collected from PV rooftop systems with a 2.14 MW power capacity that were distributed within 0.7 km^2 at the University of Queensland, Brisbane, Australia. The PV site is located at 153°29′44″ east longitude and 27°00′30″ south latitude. All PV systems faced the north at an inclination angle of 30° to produce

maximum possible power. Brisbane city weather is described as humid and warm all through 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. In summer, Brisbane's sky exhibits several passing clouds events, more than in winter [38].

The varied weather over the four seasons for Brisbane city and the PV systems being distributed around the University of Queensland makes it a suitable site for this study. PV sensor data were collected from the PV located at Site 7 in Figure 3.

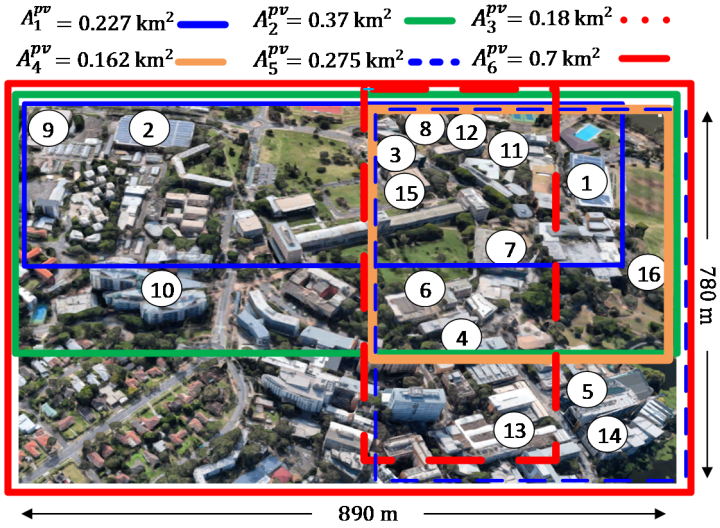

Figure 3. Layout of the six PV plants under study.

4.1. Simulation Results

The collected data of the 445 days comprise small, middle and large variability. For ANFIS training and testing data, 330 days were used and the rest of the data were used in the validation stage. To improve the ANFIS performance during the training stage, each day was divided into five sub-time series, with the first sub-time series assumed to start at 7:00 a.m. and ending at 8:59 a.m.; the second sub-time series was from 9:00 a.m. to 10:59 a.m.; the third sub-time series was from 11:00 a.m. to 12:59 p.m.; the fourth sub-time series was from 1:00 p.m. to 2:59 pm; and the fifth sub-time series was from 3:00 p.m. to 5:00 p.m.. The sunrise and sunset times have been taken into account in this division. The PV plant was also divided into six small areas, namely the whole PV plant (A_6^{pv}), $A_1^{\mathrm{pv}} = 0.227$ km^2 with 1.82 MW, $A_2^{\mathrm{pv}} = 0.37$ km^2 with 1.9 MW, $A_3^{\mathrm{pv}} = 0.17$ km^2 with 0.606 MW, $A_4^{\mathrm{pv}} = 0.162$ km^2 with 1 MW, $A_5^{\mathrm{pv}} = 0.275$ km^2 with 1.82 MW and $A_6^{\mathrm{pv}} = 0.7$ km^2 with 2.14 MW.

Following the methodology steps in Figure 2 and by using the input and output data in Table 2, the ANFIS model can be built and trained. Since the back-propagation algorithm is relatively slow [36], the hybrid-learning algorithm was adopted to train the proposed ANFIS model.

Figure 4 shows the error in estimating the power output when different MFs are employed. Results show that the gaussmf membership function is more accurate than other types after 100 epochs training modes and hence it was employed in the ANFIS model.

Figure 5 shows the Root Mean Square Error (*RMSE*) for the last epoch at each ANFIS model and at each sub-time series. Four ANFIS models were developed as shown in Table 2, third column.

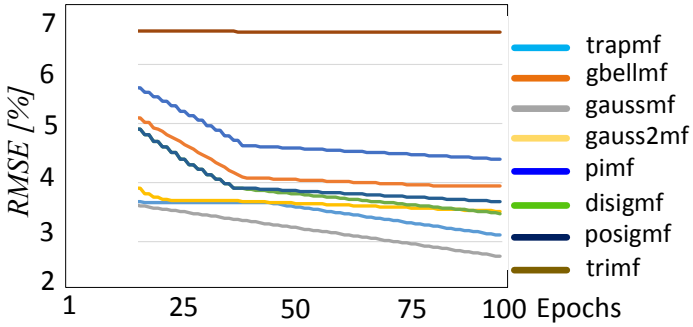

Figure 4. Resulted errors of different membership functions.

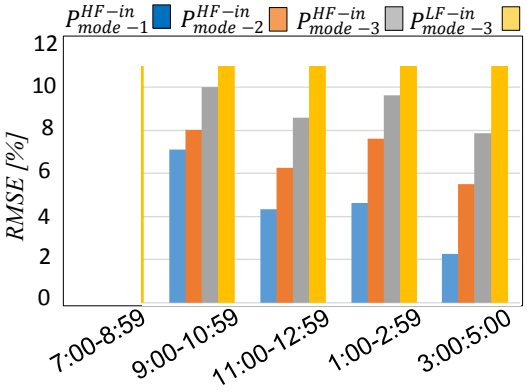

Figure 5. Errors after 100 ANFIS epochs for each sub-time series.

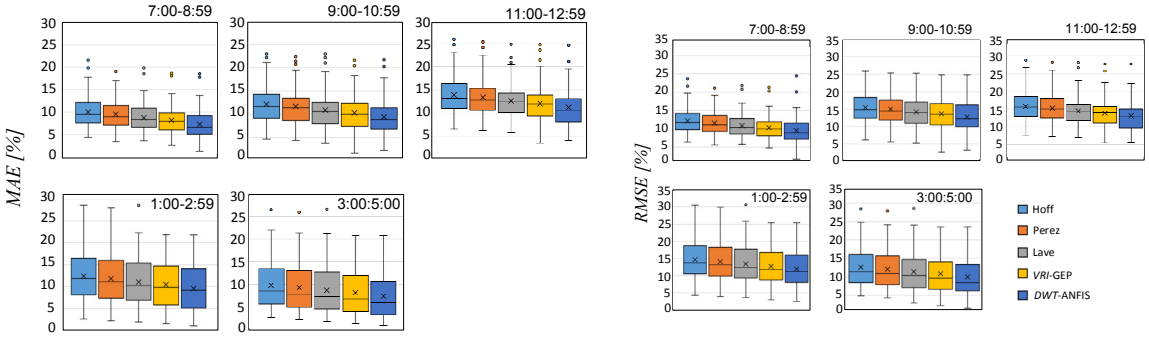

Figure 6. Comparison of the errors of the proposed ANFIS–*DWT* model and other models in the literature: (**a**) using *MAE* metric and (**b**) using *RMSE* metric

The model was validated using the remaining 115 days. Four signals ($P_{\text{mode}-1}^{\text{HF}-\text{ANFIS}}[k]$, $P_{\text{mode}-2}^{\text{HF}-\text{ANFIS}}[k]$, $P_{\text{mode}-3}^{\text{HF}-\text{ANFIS}}[k]$ and $P_{\text{mode}-3}^{\text{HF}-\text{ANFIS}}[k]$) represent the results and by applying Equations (13) and (14), the estimated PV system power power output can be obtained. Figure 6 depicts the error in the estimated power of the developed model for 115 days using Mean Absolute Error (*MAE*) and *RMSE* in comparison with other methods mentioned in Section 2, such as the Hoff, Perez, Lave and *VRI*–GEP models. Table 3 summaries the results of Figure 6 by using the average values of the *MEA* and *RMSE* for each method mentioned in the introduction section along with the proposed technique in this paper. This table also illustrates the minimum and maximum improvement values attained by the proposed technique when compared with the existing models in the literature. 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 11.44, 8.87, 7.46, 9.03 and 10.41, respectively, whereas the maximum improvements for the same

sub-time series are 36.89, 28.6, 24.06, 28.81 and 33.07, respectively. Similarly, results reveal a very similar tendency in the enhancement attained by the proposed technique when the *RMSE* metric is considered.

Table 3. MAE (%) and *RMSE* (%) average value comparison for different variability reduction index (*VRI*) models, including the proposed model in this paper.

Models	7:00–8:59 a.m.		9:00–10:59 a.m.		11:00–12:59 p.m.		1:00–2:59 p.m.		3:00–5:00 p.m.	
	MAE	***RMES***	**MAE**	***RMES***	**MAE**	***RMES***	**MAE**	***RMES***	**MAE**	***RMES***
Hoff	10.05	12.11	12.18	15.48	13.96	15.75	12.32	14.62	9.78	12.58
Perez	9.5	15.80	11.64	14.94	13.43	15.22	11.76	14.08	9.26	12.04
Lave	8.8	10.86	10.92	14.23	12.71	14.50	11.04	13.36	8.72	11.40
VRI-GEP	8.18	10.24	10.31	13.61	12.10	13.88	10.43	12.75	8.12	10.78
DWT-ANFIS	7.34	9.52	9.47	12.77	11.25	13.04	9.56	11.91	7.35	10.01
Minimum improvement (%)	11.44	7.641	8.87	6.58	7.46	6.44	9.03	7.05	10.41	7.84
Maximum improvemen (%)	36.89	27.26	28.6	21.20	24.06	20.76	28.81	22.74	33.07	25.83

To calculate the *VRI* parameter, the power output was analyzed from six wavelet modes using Equation (8), and then the wavelet periodograms factor in Equation (15) was employed at each wavelet mode. Using Equation (16), the amount of power involved in the variability $V_{pi}\left(T^{\text{mode}-i}\right)$ at each mode and each sub-time series can be evaluated. Figure 7 shows an example for one day, December 3, 2017, that was chosen to explain this step. The blue line in Figure 7 represents $V_{pi}\left(T^{\text{mode}-i}\right)$ for one PV system (No. 7), while the black and red lines represent $V_{pi}\left(T^{\text{mode}-i}\right)$ for the actual and predicted power output 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) characterizes the geographic smoothing or the *VRI* value. The *VRI* can be calculated using Equation (17) as shown in Figure 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 a.m., while it is 6.2, 7, 7, 7.5 at sub-time series 9:00–10:59 a.m., 11:00–12:59 p.m., 1:00–2:59 p.m. and 3:00–5:00 p.m., respectively. This is attributed to the change in the ambient parameters affecting the *VRI*, such as V_{clo}, T_{em}, and H_u during the daytime.

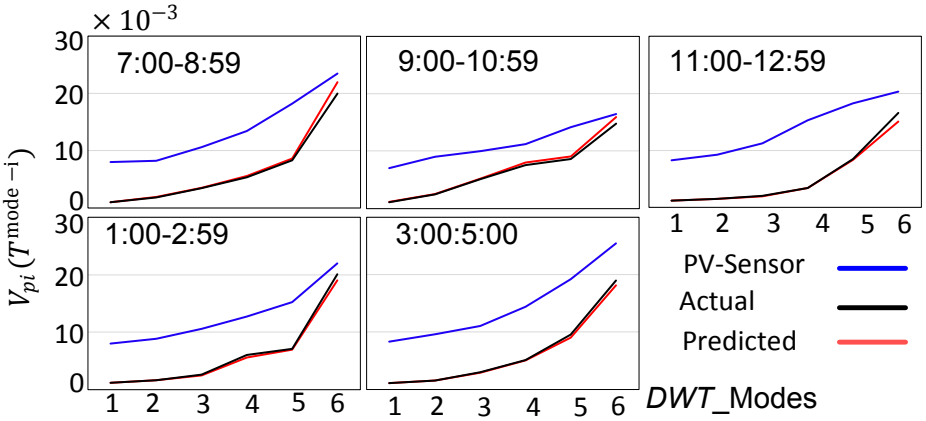

Figure 7. V_{pi} for individual PV system (blue), entire actual PV systems (black) and predicted (red).

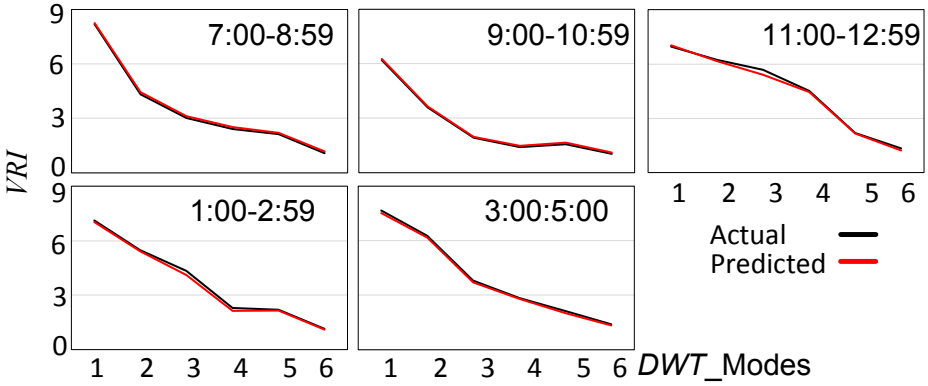

Figure 8. The actual (black) and predicted (red) *VRI* values, December 3, 2017.

4.2. Discussion

The estimation improvement strategy adopted in this study is based on dividing the time series data into several sub-time series (five sub-times series in this study) along with employing the Wavelet Transform Technique. 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 tempreture, zenith angle and power output would not 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 the input and output data.

Results of the proposed technique in this paper reveal its high accuracy when compared with other estimation models published in the literature. Results in Figure 6 and Table 3 also show a comparison of the model proposed in this paper and our recent *VRI*–GEP technique [29]. It can be observed that both the GEP and ANFIS techniques provide more accurate results than other techniques in the literature [39,40]. Reference [29] employed the GEP technique to develop a *VRI* model whereas in this paper, the ANFIS–*DWT* model is used to estimate the power output 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. Results show that the ANFIS model proposed in this paper attains a minimum and maximum improvement of 6.44% and 11.44% when compared with the *VRI*–GEP model by considering the MAE metric.

4.3. 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 was changed in an expected range as listed in Table 4 while maintaining all other parameters unchanged [41]. The effect of such a change on the power output was analyzed using the standard deviation (SD) metric. To standardize the model sensitivity analysis, SD is divided by the average value of the measured parameter. Figure 9 shows the impact of each parameter at different sub-time series. It can be seen that the input power parameter has a more significant impact than the ambient temperature while the humidity, cloud speed and the ratio of the PV plant capacity to the plant area have approximately the same impact. Table 4 illustrates the minimum and maximum values for each parameter that were considered during the sensitivity analysis.

Table 4. Minimum and maximum values for each model parameter.

Parameters	Minimum	Maximum
$P_{\mathrm{nr}}^{\mathrm{POA-in}}[k]$	1	0
$H_{\mathrm{u}}[k]$	10	95
V_{clo}	1	13
$T_{\mathrm{em}}[k]$	8	36
De	3	8

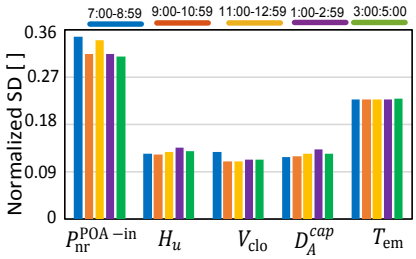

Figure 9. Sensitivity analysis for each model parameter at each sub-time series.

5. Conclusions

Using Wavelet-based and ANFIS modeling approaches, new PV system power output estimation models are presented in this paper. The developed models are employed to estimate the generated power of a group of PV systems distributed over 0.7 km^2 using irradiance data of one PV sensor located near the PV plant. Parameters such as PV area, capacity, cloud speed, temperature and humidity are considered as the inputs to the proposed model. Real data collected from a PV plant in Brisbane, Australia, were used to train the proposed model. Results reveal the high accuracy of the developed model when compared with the existing estimation models in the literature. Sensitivity analysis of each parameter reveals that the power output collected from the PV sensor has the largest impact on estimating the entire PV system's power output using the developed model. Ambient parameters, including humidity, cloud speed and the power capacity of the overall PV plant, have similar but less impactful effects on the estimated power.

Author Contributions: The draft paper is prepared by H.A.H.A.-H. as part of his PhD thesis. The paper is reviewed, edited and amended by A.A.-S. and F.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

H_u	humidity
$I_{\mathrm{clr}}^{\mathrm{d}}$	Incident direct irradiances on PV array (clear sky)
$I_{\mathrm{clr}}^{\mathrm{g}}$	Incident ground-reflected diffuse irradiances (clear sky)
$I_{\mathrm{clr}}^{\mathrm{s}}$	Incident sky diffuse irradiances (clear sky)
I_{w}	wavelet periodogram
I^{POA}	Plane of the array irradiance
j	Model-level
$inv(\ )$	Inverse *DWT* mode (level)
f	Shift factor
k	Time sample
N_{syst}	Number of the PV sites
N^{mod}	Number of PV modules
$()^{\mathrm{overall}}$	Entire PV plant

P^{cap}	PVs plant power capacity
$P^{\text{mdu-in}}_{\text{POA}}$	Pyranometer power equivalent to module power
$P^{\text{mdu-out}}_{\text{Eq}}$	Average output power of the overall PV plant
$P^{\text{POA}}_{\text{clr}}$	Clear sky power model
$P^{\text{POA-in}}_{\text{nr}}$	Normalized $P^{\text{mdu-in}}_{\text{POA}}$
$P^{\text{POA-out}}_{\text{nr}}$	Normalized $P^{\text{mdu-out}}_{\text{Eq}}$
P_{nr}	Could be $P^{\text{POA-in}}_{\text{nr}}$ or $P^{\text{POA-out}}_{\text{nr}}$
$P^{\text{HF-in}}_{\text{mode}-j}$	Wavelet high-frequency levels for $P^{\text{POA-in}}_{\text{nr}}$
$P^{\text{LF-in}}_{\text{mode}-j}$	Wavelet low-frequency levels for $P^{\text{POA-in}}_{\text{nr}}$
$P^{\text{HF-out}}_{\text{mode}-j}$	Wavelet high-frequency levels for $P^{\text{POA-out}}_{\text{nr}}$
$P^{\text{LF-out}}_{\text{mode}-j}$	Wavelet low-frequency levels for $P^{\text{POA-out}}_{\text{nr}}$
$P^{\text{HF-ANFIS}}_{\text{mode}-j}$	*DWT*-ANFIS high-frequency levels
$P^{\text{LF-ANFIS}}_{\text{mode-3}}$	*DWT*-ANFIS low-frequency levels
$P^{\text{Eq-ANFIS}}_{\text{nr-out}}$	Estimation of equivalent normalized output power (entire PVs)
$P^{\text{Eq-ANFIS}}_{\text{out}}$	Equivalent output power seen by the entire PV plant
T^{mode-j}	Timescale
T_{em}	Ambient temperature
T_{L}	Length of the time-series signal
x	can be Hoff, Perez, Lave or ACM correlation model
V_{clo}	cloud speed
VRI	Variability reduction index
V_{pi}	Variability power index
$V_{\text{pi}}\left(T^{\text{mode}-j}\right)_{\text{py}}$	V_{pi} for the pyranometer (or individual PV system)
$V_{\text{pi}}\left(T^{\text{mode}-j}\right)_{\text{ave}}$	V_{pi} for the overall PV plant
β_{m}	Mounting coefficient
ψ	Mother wavelet
ρ	Correlation coefficient

References

1. Emmanuel, M.; Rayudu, R.; Welch, I. Modelling impacts of utility-scale photovoltaic systems variability using the wavelet variability model for smart grid operations. *Sustain. Energy Technol. Assess.* **2019**, *31*, 292–305. [CrossRef]
2. The Clean Energy Council. *Clean Energy Australia Report*; The Clean Energy Council: Sydney, SY, Australia, 2019.
3. Blakers, A.; Stocks, M.; Lu, B. *Australia: The Renewable Energy Superstar*; Australian National University: Canberra, Australia, 2019.
4. Gari da Silva Fonseca Junior, J.; Nishitsuji, Y.; Udagawa, Y.; Urabe, C.T.; Saito, T.; Uno, F.; Ohtake, H.; Oozeki, T.; Ogimoto, K. Regional photovoltaic power fluctuations within frequency regulation control time frames: A study with high-resolution data. *Prog. Photovolt. Res. Appl.* **2018**, *26*, 402–413.
5. Hirooka, D.; Murata, N.; Fujimoto, Y.; Hayashi, Y. Temporal Interpolation of Gridded Solar Radiation Data for Evaluation of PV Fluctuations. *Energy Procedia* **2018**, *155*, 259–268. [CrossRef]
6. Júnior, C.F.M.; Waenga, A.F.C.; Pinto, D.A.F. Effects of the Photovoltaic Distributed Generation on Electricity Distribution System Voltage-Updated Review. *Braz. Arch. Biol. Technol.* **2018**, *61*. [CrossRef]
7. Al-Hilfi, H.; Shahnia, F.; Abu-Siada, A. An Improved Technique to Estimate the Total Generated Power by Neighboring Photovoltaic Systems Using Single-point Irradiance Measurement and Correlational Models. *IEEE Trans. Ind. Inform.* **2019**, *16*, 3905–3917. [CrossRef]
8. Al-Hilfi, H.A.; Shahnia, F.; Abu-Siada, A. Comparison of different correlational techniques in estimating the total generated power of neighboring photovoltaic systems. In Proceedings of the 2016 Australasian Universities Power Engineering Conference (AUPEC), Brisbane, QLD, Australia, 25–28 September 2016; pp. 1–6.

9. Rodríguez-Benítez, F.J.; Arbizu-Barrena, C.; Santos-Alamillos, F.J.; Tovar-Pescador, J.; Pozo-Vázquez, D. Analysis of the intra-day solar resource variability in the Iberian Peninsula. *Sol. Energy* **2018**, *171*, 374–387. [CrossRef]
10. Lohmann, G.M.; Monahan, A.H. Effects of temporal averaging on short-term irradiance variability under mixed sky conditions. *Atmos. Meas. Tech.* **2018**, *11*, 3131–3144. [CrossRef]
11. Lohmann, G.M.; Monahan, A.H.; Heinemann, D. Local short-term variability in solar irradiance. *Atmos. Chem. Phys.* **2016**, *16*, 6365–6379. [CrossRef]
12. Rowlands, I.H.; Kemery, B.P.; Beausoleil-Morrison, I. Managing solar-PV variability with geographical dispersion: An Ontario (Canada) case-study. *Renew. Energy* **2014**, *68*, 171–180. [CrossRef]
13. Palmer, D.; Koubli, E.; Cole, I.; Betts, T.; Gottschalg, R. Comparison of solar radiation and PV generation variability: System dispersion in the UK. *IET Renew. Power Gener.* **2017**, *11*, 550–557. [CrossRef]
14. Lauret, P.; Perez, R.; Aguiar, L.M.; Tapachès, E.; Diagne, H.M.; David, M. Characterization of the intraday variability regime of solar irradiation of climatically distinct locations. *Sol. Energy* **2016**, *125*, 99–110. [CrossRef]
15. Lave, M.; Broderick, R.J.; Reno, M.J. Solar variability zones: Satellite-derived zones that represent high-frequency ground variability. *Sol. Energy* **2017**, *151*, 119–128. [CrossRef]
16. Gagné, A.; Turcotte, D.; Goswamy, N.; Poissant, Y. High resolution characterisation of solar variability for two sites in Eastern Canada. *Sol. Energy* **2016**, *137*, 46–54. [CrossRef]
17. Lave, M.; Reno, M.J.; Broderick, R.J. Characterizing local high-frequency solar variability and its impact to distribution studies. *Sol. Energy* **2015**, *118*, 327–337. [CrossRef]
18. Madanchi, A.; Absalan, M.; Lohmann, G.; Anvari, M.; Tabar, M.R.R. Strong short-term non-linearity of solar irradiance fluctuations. *Sol. Energy* **2017**, *144*, 1–9. [CrossRef]
19. Klima, K.; Apt, J. Geographic smoothing of solar PV: Results from Gujarat. *Environ. Res. Lett.* **2015**, *10*, 104001. [CrossRef]
20. Tang, Y.; Cheng, J.W.; Duan, Q.; Lee, C.W.; Zhong, J. Evaluating the variability of photovoltaics: A new stochastic method to generate site-specific synthetic solar data and applications to system studies. *Renew. Energy* **2019**, *133*, 1099–1107. [CrossRef]
21. Madhavan, B.L.; Deneke, H.; Witthuhn, J.; Macke, A. Multiresolution analysis of the spatiotemporal variability in global radiation observed by a dense network of 99 pyranometers. *Atmos. Chem. Phys.* **2017**, *17*, 3317–3338. [CrossRef]
22. Emmanuel, M.; Rayudu, R.; Welch, I. Impacts of power factor control schemes in time series power flow analysis for centralized PV plants using wavelet variability model. *IEEE Trans. Ind. Inform.* **2017**, *13*, 3185–3194. [CrossRef]
23. Lave, M.; Kleissl, J.; Stein, J.S. A wavelet-based variability model (WVM) for solar PV power plants. *IEEE Trans. Sustain. Energy* **2012**, *4*, 501–509. [CrossRef]
24. Al-Hilfi, H.A.; Shahnia, F.; Abu-Siada, A. Enhancing the estimation of the overall produced power by several adjacent photovoltaic systems using existing correlational factors. In Proceedings of the 2018 IEEE 27th International Symposium on Industrial Electronics (ISIE), Cairns, QLD, Australia, 13–15 June 2018; pp. 43–47.
25. O'Rourke, N.; Hatcher, L. *A Step-by-Step Approach to Using SAS for Factor Analysis and Structural Equation Modeling*; Sas Institute: Cary, NC, USA, 2013.
26. Hoff, T.E.; Perez, R. Modeling PV fleet output variability. *Sol. Energy* **2012**, *86*, 2177–2189. [CrossRef]
27. Perez, R. *Wind Field and Solar Radiation Characterization and Forecasting: A Numerical Approach for Complex Terrain*; Springer: Berlin, Germany, 2018.
28. Arias-Castro, E.; Kleissl, J.; Lave, M. A Poisson model for anisotropic solar ramp rate correlations. *Sol. Energy* **2014**, *101*, 192–202. [CrossRef]
29. Hasanain, F.S.; Al-Hilfi, A.H.; Abu-Siada, A. Employing a Gene Expression-Based Technique to Improve the Accuracy of Estimating the Total Generated Power by Neighboring Photovoltaic Systems. In Proceedings of the ICPES 2019 Conference Perth, Perth, Australia, 10–12 December 2019.
30. Solgi, A.; Nourani, V.; Pourhaghi, A. Forecasting daily precipitation using hybrid model of wavelet-artificial neural network and comparison with adaptive neurofuzzy inference system (case study: Verayneh station, Nahavand). *Adv. Civ. Eng.* **2014**, *2014*, 279368. [CrossRef]
31. Alizdeh, M.J.; Joneyd, P.M.; Motahhari, M.; Ejlali, F.; Kiani, H. A wavelet-ANFIS model to estimate sedimentation in dam reservoir. *Int. J. Comput. Appl.* **2015**, *114*, 19–25.

32. *Fuzzy Logic Toolbox User's Guide*; The Mathworks Inc.: Massachusetts, MA, USA, 2018.
33. Abdulshahed, A.M.; Longstaff, A.P.; Fletcher, S. The application of ANFIS prediction models for thermal error compensation on CNC machine tools. *Appl. Soft Comput.* **2015**, *27*, 158–168. [CrossRef]
34. Lave, M.; Kleissl, J. Cloud speed impact on solar variability scaling–Application to the wavelet variability model. *Sol. Energy* **2013**, *91*, 11–21. [CrossRef]
35. CAMS McClear, Website. Retrieved 2018. Available online: http://www.soda-pro.com/web-services/radiation/cams-mcclear (accessed on 23 September 2018).
36. Seo, Y.; Kim, S.; Kisi, O.; Singh, V.P. Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. *J. Hydrol.* **2015**, *520*, 224–243. [CrossRef]
37. Lave, M.; Kleissl, J.; Arias-Castro, E. High-frequency irradiance fluctuations and geographic smoothing. *Sol. Energy* **2012**, *86*, 2190–2199. [CrossRef]
38. UQ Solar Photovoltaic Data. Website of the University of Queensland. Reterived 2018. Available online: http://solar.uq.edu.au/user/reportpower.php (accessed on 1 September 2018).
39. Moghaddasi, H.; Mahmoudi, I.; Sajadi, S. Comparing the Efficiency of Artificial Neural Network and Gene Expression Programming in Predicting Coronary Artery Disease. *J. Health Med. Informat.* **2017**, *8*. [CrossRef]
40. Sadrossadat, E.; Ghorbani, B.; Oskooei, R.; Kaboutari, M. Use of adaptive neuro-fuzzy inference system and gene expression programming methods for estimation of the bearing capacity of rock foundations. *Eng. Comput.* **2018**, *35*, 2078–2106. [CrossRef]
41. Wang, T.; Dyer, J.S.; Hahn, W.J. Sensitivity analysis of decision making under dependent uncertainties using copulas. *EURO J. Decis. Process.* **2017**, *5*, 117–139. [CrossRef]