INCREASED PENETRATION OF DISTRIBUTED ROOF-TOP PHOTOVOLTAIC SYSTEMS IN SECONDARY LOW VOLTAGE NETWORKS: INTERCONNECTION IMPACT ANALYSIS AND MITIGATION

A Thesis Submitted to the

College of Graduate and Postdoctoral Studies

In Partial Fulfilment of the Requirements

For the Degree of Doctor of Philosophy

In the Department of Electrical and Computer Engineering

University of Saskatchewan

Saskatoon

By

Elisha Chukwuzemelu Asiri

PERMISSION TO USE

In presenting this thesis/dissertation in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis/dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis/dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

Requests for permission to copy or to make other uses of materials in this thesis/dissertation in whole or part should be addressed to:

Head of the Department of Electrical and Computer Engineering

57 Campus Drive

University of Saskatchewan

Saskatoon, Saskatchewan (S7N 5A9)

Canada

OR

Dean

College of Graduate and Postdoctoral Studies

University of Saskatchewan

116 Thorvaldson Building, 110 Science Place

Saskatoon, Saskatchewan (S7N 5C9)

Canada

ABSTRACT

The worsening global climatic condition has necessitated increased investments in renewable energy resources and in turn increased penetration of these resources in electricity grids worldwide. Distributed photovoltaic (PV) energy is one of the rapidly growing and viable forms of renewable energy. Distributed PV systems currently exist in two modes, a few number of large or utility scale systems and a plethora of small or residential scale roof-top systems which are rapidly growing in terms of number. Residential systems are alternatively called Behind the Meter (BTM) systems because they are not directly monitored by utility operators and are therefore invisible vis-à-vis their performances. While an individual Behind-The-Meter (BTM) system's size holds little significance in comparison to the inertia of the utility grid, the collective presence of numerous interconnected BTM systems within a single feeder has the potential to jeopardize the stability and security of utility operations. Conventional protective devices within distribution networks are designed to accommodate a unidirectional downstream power flow. However, as the integration of PV generators into utility grids intensifies, the prospect of reverse or upstream power flow becomes more probable. This development raises various apprehensions, including the potential for voltage level breaches and a notable reduction in the operational longevity of these devices.

BTM systems generally have a wide geographical coverage within a region and each system operates independently of others as well as the fact that their real-time performances are concealed in the net-load data relayed by electricity meters. Consequently, traditional forecasting methods have proved insufficient in predicting the outputs of PV systems on a regional level requiring the development of spatial aggregation approaches. Three basic sub-areas aimed at increasing the penetration of BTM PV generators in utility grids are the principal focus of this study. The sub-areas include performance analysis of BTM systems; day-ahead regional scale PV power forecasting model and a PV ramp events extraction model.

The first sub-area tries to address the challenges with small scale solar power performance data access on a regional basis. The performance analysis was aimed at evaluating the credibility and reliability of BTM data from public webpages and their representativeness for high profile

research. Consequently, this sub-area proposes and investigates the feasibility of the instrumentation of every invisible solar system for near real-time data monitoring. The investigation involved detailing the convergence between simulated and reported power outputs on a spectrum of orientation and tilt angles. Two simulation methods as well as two case studies public web repositories from which a subset of representative solar sites were adopted to provide a basis for the proposed approach. The results show that the proposed model is viable and feasible depending on the participation of certain key stakeholders in electricity market discourses.

Day-ahead forecasts are required by electricity market investors to make informed decisions on the trading floor. Whereas it is relatively easier to predict the performance of a few large-scale PV systems, a large number of small-scale PV systems with a wide geographical spread poses more challenges because they are not metered for real-time monitoring. This sub-area proposes an artificial neural network (ANN)-based model to achieve regional-scale day-ahead PV power forecasts from numerical weather predictions of weather variables excluding solar irradiance as inputs. The model was first implemented by dividing a region into clusters and selecting a representative site for each cluster using data dimension reduction algorithms. Solar irradiance forecasts were then generated for each representative PV system and the corresponding PV power was simulated. The cluster power output was obtained using a linear upscaling model and summed to produce regional-scale power forecasts. The model's accuracy is validated using power generation data of several distributed systems in California. Compared with available benchmark models with similar objectives, the proposed model performed significantly better.

Insufficient information on solar power ramp events is counterproductive to the operational flexibility and economics of electricity grids. Accurate solar ramp extraction and characterization in terms of ramp magnitude, rate and duration are useful to power system operators for system planning especially with regards to ensuring supply security and sizing ancillary services. The characterization of ramp events in historical databases is also useful for testing forecast models' accuracy in predicting significant solar ramp events that are of more concern to utility operators. A novel technique for solar power ramp events (SPREs) detection using the modified swinging door algorithm (MSDA) considering different time resolutions and weather profile is proposed in this sub-area. Firstly, the swinging door algorithm (SDA) is used to create ramp segments of the solar power data that are collected from different randomly selected systems. Afterwards, the

power generation variability patterns of these segments are studied. The SDA is then modified to merge adjacent segments according to the observations made by comparing the variability patterns. The solar power data simulated from irradiances measured with different time resolutions is utilized for performance validation and testing. The proposed technique shows much improved performance than existing detection algorithms with respect to the number of detected ramps, detection accuracy and in some cases, computation time.

ACKNOWLEDGEMENT

My sincere appreciation to Jesus Christ for the gift of life and health all through the period of this work.

I would also like to appreciate the immense support, understanding and guidance of my supervisors Prof. Tony C.Y. Chung and Prof. Liang Xiaodong. Many thanks also for the contributions of my advisory team members including Prof. Francis Bui, Prof. Khan Wahid and Prof Chris Zhang.

The SMARTGEN lab members created a supporting environment all through my research work, I am indeed grateful for this. Special mention goes to Osama, Nima, Bingzhi, Mohammed Mahgoub and Peiyan for their reliable and invaluable advice.

My family members, friends and most especially, my wife who reliably supported me emotionally, spiritually, and psychologically, I am grateful.

Lastly, my research work wouldn't have seen the light of day without the financial contribution from the Petroleum Technology Development Fund (PTDF), Nigeria.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ACF	Autocorrelation Function
ANN	Artificial Neural Network
AOI	Angle of Incidence
$ARENA$	Australian Renewable Energy Agency
$ARIMA$	Autoregressive Integrated Moving Average
$ARMAX$	Autoregressive Moving Average with Exogenous
BPI	Bird Performance Index
BTM	Behind the Meter
$CAISO$	California Independent System Operator
CC	Cloud Cover
CEC	Clean Energy Council
CMV	Cloud Motion Vector
CSI	Clear Sky Index
$CSIr$	Critical Success Index
DHI	Diffuse Horizontal Irradiance
DNI	Direct Normal Irradiance
$DOTY$	Day of the Year
$ECMWF$	European Centre for Medium-Range Weather Forecasts
ELM	Extreme Learning Machine
ENN	Elman Neural Network
FAR	False Alarm Ratio
FBS	Frequency Bias Score
$FFNN$	Feed Forward Neural Network
$FITs$	Feed-In-Tariffs

GDP	Gross Domestic Product
GHI	Global Horizontal Irradiance
GW	Gigawatt
HRA	Hour Angle
ISO	Independent System Operator
IEA	International Energy Agency
kW	Kilowatt
$L1-SW$	L1-Ramp Detect with Sliding Window
LF	Loss Factor
$LSTM$	Long Short-Term Memory
$MIDC$	Measurement and Instrumentation Data Center
ML	Machine Learning
$MSDA$	Modified Swinging Door Algorithm
MW	Megawatt
$NDFD$	National Digital Forecast Database
$NMAD$	Normalized Mean Absolute Difference
$NMAE$	Normalized Mean Absolute Error
$NMBD$	Normalized Mean Bias Difference
$NNCT$	No Negative Constraint Theory
$NOAA$	National Oceanic and Atmospheric Administration
$NREL$	National Renewable Energy Laboratory
$NRMSD$	Normalized Root Mean Square Difference
$NRMSE$	Normalized Root Mean Square Error
$NSRDB$	National Solar Radiation Database
NVI	Natural Variability Index
$NWPs$	Numerical Weather Predictions
$OpSDA$	Optimized Swinging Door Algorithm
PCA	Principal Component Analysis

POA	Plane of Array
POD	Probability of Detection
PR	Performance Ratio
$PRAA$	Parameter and Resolution Adaptive Algorithm
PV	Photovoltaic
$PVGIS$	Photovoltaic Geographic Information System
RH	Relative Humidity
$SAPM$	Sandia PV Array Performance Model
SDA	Swinging Door Algorithm
$sNeN$	start-NO-end-NO
$sNeY$	start-NO-end-YES
$SPREs$	Solar Power Ramp Events
SR	Success Ratio
STC	Standard Test Conditions
$sYeN$	start-YES-end-NO
$sYeY$	start-YES-end-YES
SZA	Solar Zenith Angle
T	Temperature
TW	Terrawatts
UTC	Coordinated Universal Time
$WMAPE$	Weighted Mean Absolute Percentage Error
$WPREs$	Wind Power Ramp Events
WS	Wind Speed

NOMENCLATURE

I_{POA}	Plane of Array Irradiance
I_{DH}	Diffuse Horizontal Irradiance
I_{GH}	Diffuse Horizontal Irradiance
I_{DN}	Direct Normal Irradiance
K_{DN}	Anisotropy index
K_{GH}	Clearness index
B	Regression coefficient
C	Regression coefficient
$I_{DN_{POA}}$	Plane of Array Direct Normal Irradiance
$I_{grd_{POA}}$	Plane of Array Ground reflected diffuse Irradiance
$I_{sky_{POA}}$	Plane of Array Sky diffuse Irradiance
I_{ext}	Extraterrestrial Irradiance
I_{ext_h}	Extraterrestrial irradiance at the horizontal plane
σ	Surface albedo
α	Array tilt angle
$\emptyset$	Latitude
δ_S	Solar azimuth angle
δ	Azimuth Angle of the array
$P_{mp_{DC}}$	Module power output at maximum power point

I_e	POA effective irradiance
P_{mpo}	Module power rating
ϑ	Module power temperature coefficient
$P_{mp_{AC}}$	AC power output
$P_{mp_{AC0}}$	Inverter optimum AC power rating at standard test condition
$P_{mp_{DCo}}$	Optimum DC power level at standard test condition
$P_{so_{DC}}$	Inverter starting DC power
C_o	Parameter for the AC – DC power relationship
C_1	Empirical coefficient for the linear variation of $P_{mp_{DCo}}$ with DC-voltage input
C_2	Empirical coefficient for the linear variation of $P_{so_{DC}}$ with DC-voltage input
C_3	Empirical coefficient for the linear variation of C_o with DC-voltage input
$V_{mp_{DC}}$	Input DC voltage at optimum condition
$V_{mp_{DCo}}$	Optimum DC voltage at standard test condition
$\hat{I}$	Normalized plane of array irradiance
$c_0 - c_5$	Coefficients of the quadratic model calibrated by measured data
$\hat{T}$	Module temperature
T_c	PV cell temperature
x_m	Modeled output values
x_r	Reported output values
n	number of values
θ_{SZA}	Solar zenith angle
σ_d	Declination angle

I_{GHcsk}	Clear sky irradiance at the horizontal plane
α_e	Elevation angle
f_{NN}	Neural network approximation function
m	Six meteorological predictors
t	Time
P_f	Projected power output
T_{amb}	Ambient Temperature
$a_1 - a_6$	PV power modelling equation coefficients
P_{region}	Regional-level PV system power outputs,
$P_{cluster}$	Cluster-level PV system power outputs,
P_{rep}	Representative PV system power outputs,
N	Total number of representative systems,
m_c	Total number of clusters,
j	$j-th$ representative system,
k	$k-th$ cluster, and
f	Mapping function representing the ratio between nominal capacities
$P_{parameter}$	Either the predictor or the target variable.
P_{scaled}	Scaled power value
I_{GH_f}	Forecasted irradiance
I_{GH_m}	Measured irradiance
$I_{GH_{m,peak}}$	Measured peak irradiance
ε	Door width (tolerance)
y	Signal magnitude

ub	Upper boundary of the current signal point
lb	Lower boundary of the current signal point
c	Current signal point
g	gate point
s	Start point of a ramp segment
e	End point of a ramp segment
D	Ramp definition
P_G	Installed capacity of the PV system
P_j	Instantaneous power at the beginning of a ramp segment
P_k	Instantaneous power at the end of a ramp segment
t_j	Ramp segment start time
t_k	Ramp segment stop time
μ_P	Arithmetic mean of PV power
ΔP_i	i-th data point of power step change
$\Delta \bar{P}$	Arithmetic mean of power step change and
k	Solar power data size
p_i	Solar power at point i
t_i	Equivalent time stamp at point i
p_k	Solar power of ramp segment k
S_k	Swinging door algorithm ramp segment
X_j	Modified swinging door algorithm ramp segment
Y	Solar power time series
R	Solar power time series temporal resolution
L_k	Length of line connecting two segments

1 INTRODUCTION

1.1 General

The implementation of policies and initiatives with the aim of increasing the proliferation of renewable energy schemes has had various levels of success on national and global scales. Worthy of mentioning are feed-in-tariffs, net-metering and power purchase agreements which are policies adopted to incentivize distributed renewable energy investments in different scales by all stakeholders especially commercial and residential investors. From the International Energy Agency (IEA), Figure 1.1 shows a global average of 87% increment in renewable energy capacity addition every five years from 2002 to 2022.

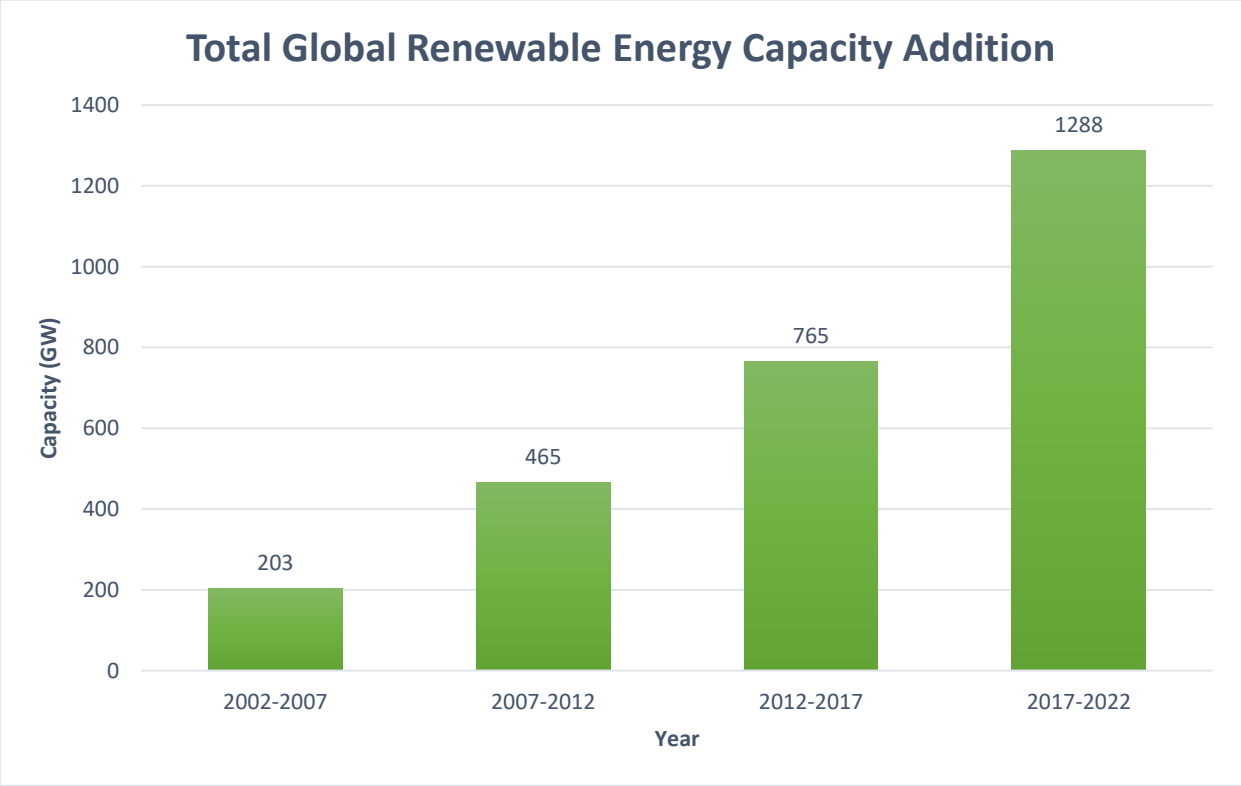

Figure 1.1. Total global renewable energy capacity additions from 2002 to 2022 in GW [1]

Power generation from PV technology has been found to be more prevalent and widespread especially roof-top solar than any other form of distributed generation. Many buildings grid-tied

to the distribution feeders have rooftop capacity that can accommodate varying PV system sizes from just a few kW to over 100kW. Global total PV installed capacity reached over 0.5TW in 2018 while almost 100GW were installed in 2018 alone with rooftop installations contributing about 30% of that figure and is predicted to steadily rise to 44GW on a low scenario and 76.5GW on a high scenario basis in 2023 [2]. At the close of 2019, California's residential solar PV installed capacity stood at 5128MW contributed by 923,405 solar projects [3].

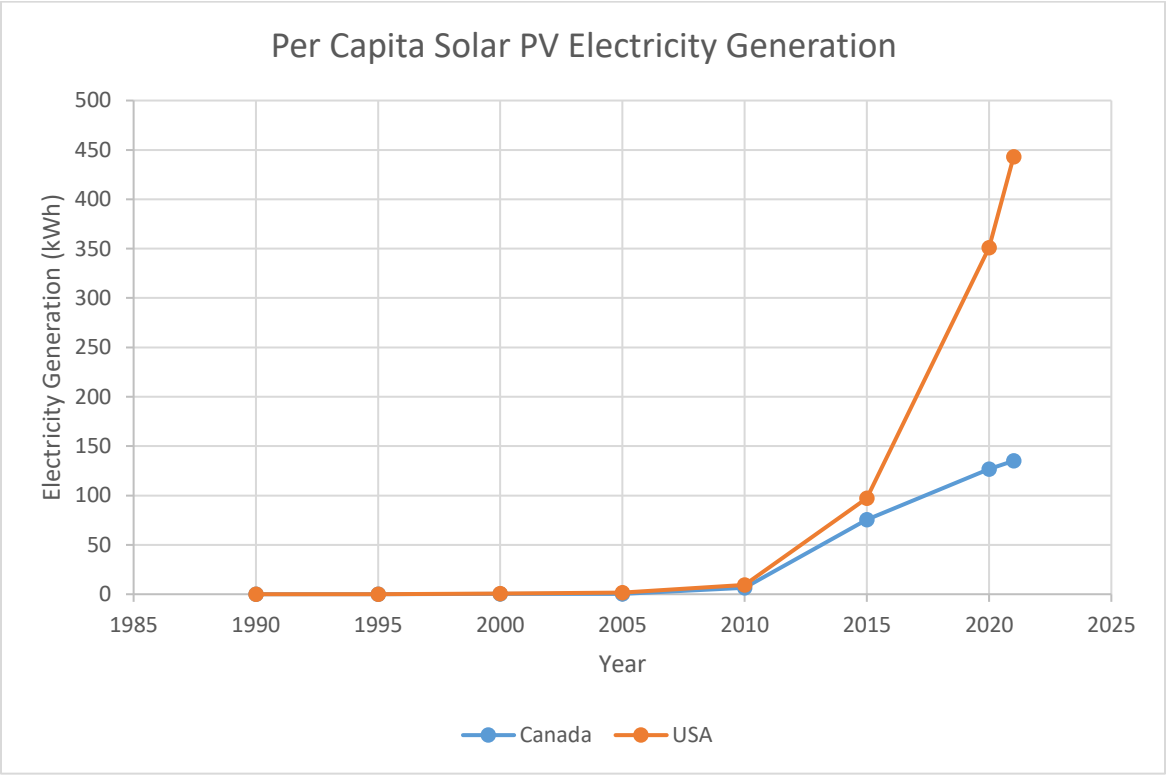

Figure 1.2. Per capita Solar PV electricity generation in the US and Canada. We obtained the total electricity generation per country data from the International Energy Agency (IEA) [4] and then normalized the data by the corresponding country's population to obtain the per capital data.

1.2 Renewable Energy Support Programs and Policies

One of the motivations for stakeholders to actively participate in rooftop PV investments includes the potential for financial returns accruable from energy bills savings. Notwithstanding the steep and continuous decline in the cost of PV materials, for investments to be economically viable and compete with large scale fossil fuel power generators counterparts, PV system owners

rely heavily on policies and support programs that offer sumptuous rebates on installation costs and incentives on energy sales to utility systems. Some of these programs include the Feed-in-Tariffs (FITs) and Net Metering. The Net Metering package rolled out by SaskPower in November 2018 and expected to run for 3 years or until a 16-megawatt cap is reached offered PV investors a rebate of $20,000 maximum ($0.61 per watt) to make up about 20% of equipment and installation cost [5]. However, this program was withdrawn and reviewed when the cap was reached just a year after its release because it couldn't be further sustained due to claims of heavy financial burden on taxpayers. Figure 1.2 shows the electricity generation from PV sources on a per capita basis in Canada and the US. A steep rise in generation beginning in 2010 was preceded by the introduction of the net metering and FITs policies in Ontario, Canada in 2006 and British Columbia, Canada in 2008. The FITs program was simultaneously implemented in Vermont, US in 2008. Table 1.1 shows some of the interventions by the Canadian government to boost investments in renewable energy projects.

Table 1.1. Some programs and policies implemented by the Canadian government to support renewable energy projects and cut greenhouse emissions .

No	Ref	Program	Year Introduced	Description
1	[6]	The Clean Energy and Electrification Program	2021	CAD $8 billion dollars to fund renewable energy-based projects for a 5-year period.
2	[7]	Clean Energy for Rural and Remote Communities Program	2016	Provides funding support to renewable energy projects in rural communities.
3	[8]	Smart Renewable and Electrification Pathways Program	2020	CAD $1.56 billion funding support for electricity grid modernization and renewable energy-based projects for an 8-year period
4	[9]	Sustainable Development Technology Canada (SDTC)	2001	A foundation that has provided CAD $2.5+ billion funding support for clean energy projects.

5	[10]	The Strategic innovation Fund	2017	Funding support for research and development of innovative technology
6	[11]	The Low carbon Economy Fund	2016	Over CAD $1.4 billion in Funding support for greenhouse gas emissions reduction projects.
7	[12]	The Pan-Canadian Framework on Clean Growth and Climate Change	2016	Provide multifaceted support to renewable energy and energy efficiency projects.

1.3 Problem Statement

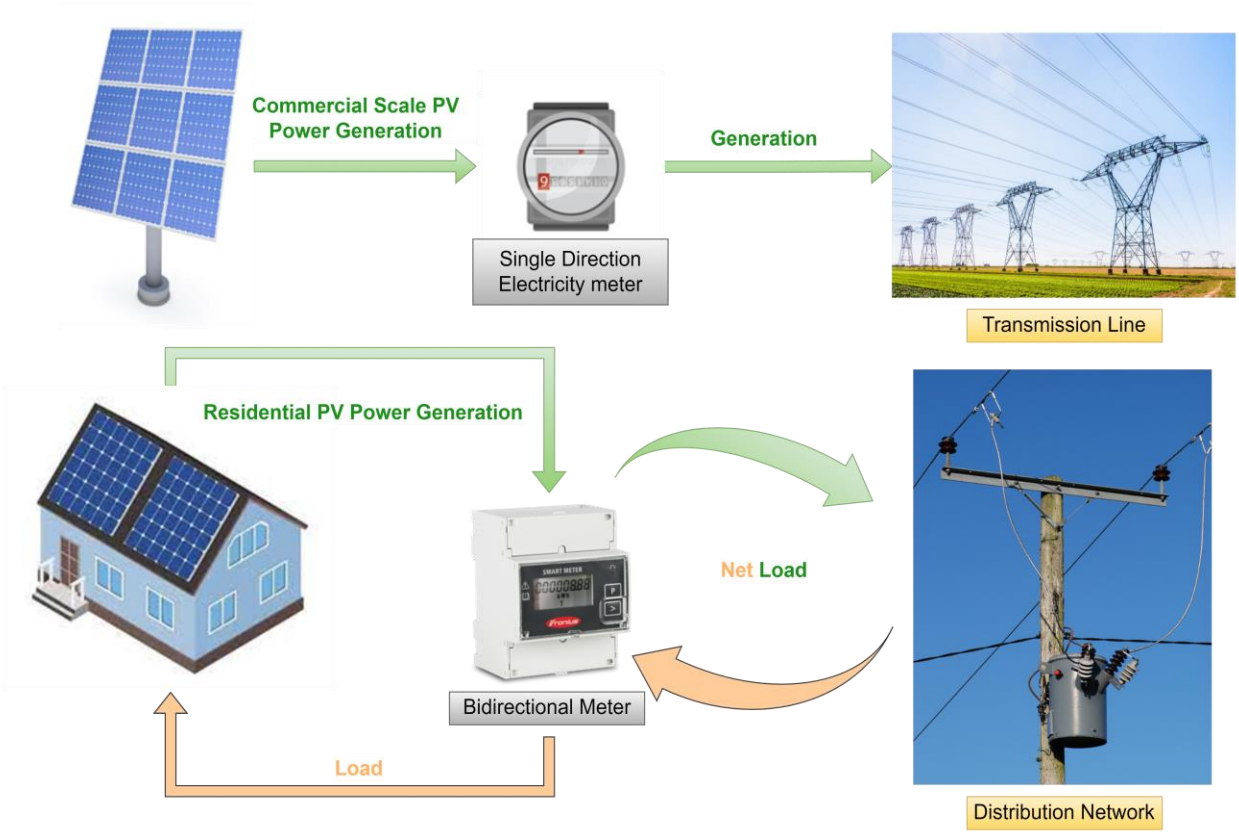

Figure 1.3. Schematical description of the basic setup of commercial and residential scale PV systems.

PV systems are either commercial or small-scale capacity as shown in Figure 1.3. Small scale installations are usually about 100kW and below and are often sited on the rooftops of residential or commercial buildings. In this document, small-scale systems are of interests and are synonymously called residential or rooftop or behind-the-meter PV systems. Granting that the impact of a single rooftop system could be negligible considering its capacity relative to the inertia of a distribution network, an aggregated output from a fleet of distributed systems on a single feeder network could significantly distort network stability. An underestimation of the impact of PV integration could significantly compromise the reliability, power quality, voltage fluctuation and frequency of the grid whereas an overestimation would limit the level of integration and in turn the economic, technical and environmental benefits accruable from renewables. There is therefore a need to accurately model the collective variability and uncertainty of grid-tied PV systems in order for utilities to fully understand and mitigate the challenges increased integration would pose to the grid.

1.4 Research Objectives and Outcomes

Basically, this research is aimed at supporting the increased penetration of distributed PV power in low voltage secondary networks. Because of the intermittency and variability of PV sources, increasing its penetration while maintaining power security requires two major actions. One involves designing sophisticated distribution grid planning tools that can effect highly detailed technical analysis to model real life systems considering different future scenarios of load and renewables integration and make recommendations for existing feeder lines upgrade requirements. The other action is to develop solar forecasting tools to provide various timescales forecasts with sufficient accuracies for generation scheduling and reserves operation. This thesis is further centered on the later aspect of forecast models development. Three broad concepts shown in Figure 1.4 form the basis for this thesis.

1. The first concept discussed in Chapter 2 and involving performance analysis of grid-tied residential PV systems was adopted to develop a proposition to securing a reliable and credible database of the performances of these systems. In this chapter, the problem of data unavailability stemming from the fact that behind-the-meter PV systems are not instrumented for direct monitoring is addressed and informed recommendations are made.

Figure 1.3 illustrates that only net load is visible to utility operators for residential PV systems as opposed to the actual generation in the case of commercial scale counterparts.

2. In chapter 3, a day-ahead prediction model of regional-scale generation from residential PV systems is presented. The training and test datasets used in this chapter are sourced from the recommendations in Chapter 2.
3. A solar power ramp events extraction algorithm is proposed in Chapter 4. The algorithm is developed as an alternative to traditional forecast models accuracy evaluation metrics. Ramp extraction is also relevant to electricity market stakeholders for determining ramping requirements of reserve and storage systems.

Chapter 5 provides a summary of and concludes this thesis.

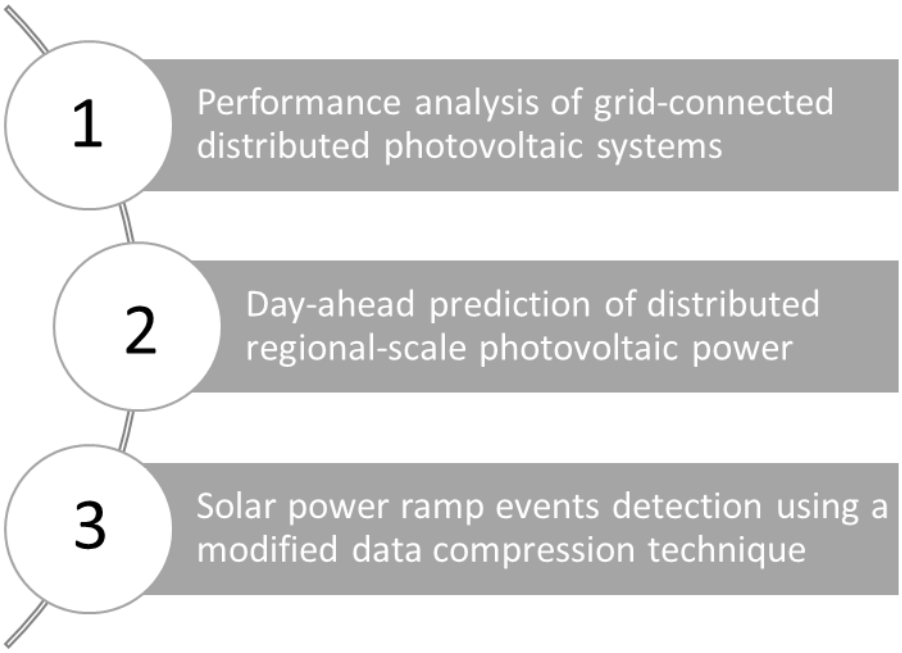

Figure 1.4. Schematic representation of the research sub-sections.

1.5 References

[1] "Total renewable capacity additions, 2002-2022 – Charts – Data & Statistics - IEA." https://www.iea.org/data-and-statistics/charts/total-renewable-capacity-additions-2002-2022 (accessed Apr. 11, 2023).

[2] Solar Power Europe, "Global Market Outlook: 2019-2023," *Glob. Mark. Outlook*, p. 92,

2019, [Online]. Available: http://www.solarpowereurope.org/wp-content/uploads/2019/05/SolarPower-Europe-Global-Market-Outlook-2019-2023.pdf

[3] "California Distributed; Generation Statistics." https://www.californiadgstats.ca.gov/charts/

[4] "Countries & Regions - IEA." https://www.iea.org/countries (accessed Apr. 11, 2023).

[5] A. White-Crummey, "Government to rush review of solar net metering program after outcry | Regina Leader Post." https://leaderpost.com/news/saskatchewan/sask-government-to-rush-review-of-rooftop-solar-program-in-wake-of-outcry.%0A

[6] Government of Canada, "Clean Energy and Electrification Program." https://www.nrcan.gc.ca/clean-growth-programs/clean-energy-electrification-program/23434.

[7] Government of Canada, "Clean Energy for Rural and Remote Communities Program," 2022, Accessed: Apr. 11, 2023. [Online]. Available: https://natural-resources.canada.ca/reducingdiesel

[8] "Smart Renewables and Electrification Pathways Program." https://natural-resources.canada.ca/climate-change/green-infrastructure-programs/sreps/23566

[9] "Home ~ Sustainable Development Technology Canada." https://www.sdtc.ca/en/ (accessed Apr. 11, 2023).

[10] "Strategic Innovation Fund." https://ised-isde.canada.ca/site/strategic-innovation-fund/en (accessed Apr. 11, 2023).

[11] "Low Carbon Economy Fund - Canada.ca." https://www.canada.ca/en/environment-climate-change/services/climate-change/low-carbon-economy-fund.html (accessed Apr. 11, 2023).

[12] Government of Canada, "Pan-Canadian Framework on Clean Growth and Climate Change first annual report," *Annex I: Status of all Pan-Canadian Framework actions*, 2017. https://www.canada.ca/en/services/environment/weather/climatechange/pan-canadian-framework.html (accessed Apr. 11, 2023).

2 PERFORMANCE ANALYSIS OF GRID-CONNECTED DISTRIBUTED PV SYSTEMS

2.1 Introduction

Roof-top PV systems are alternatively referred to as invisible solar or Behind the Meter (BTM) solar because their real online operating performances are not visible to utility system operators since they are located behind meters that relay only the net-load data. The performance of BTM solar systems alter the grid operating conditions by increasing the complexity of hourly net-load time series forecast which is a key input in the scheduling operation of generators [4][5]. Periods of low load coinciding with high renewables output increase the risk of overgeneration and consequently reverse power flow while the reverse is the case for high load and low PV output. On a regional scale, the collective impact of BTM solar on net-load could be ignored for a few systems but the proliferation of such systems on the distribution network underscores the need to innovate means to monitor how much is delivered to the grid in real time for research and forecasting.

The accuracy of PV prediction models are improved by the availability of actual distributed data for training of these models [4]. Models for regional level forecasts make used of a subset of representative PV systems that capture the main features of the whole set. These features include the distribution of systems orientations and inclinations, installed capacity, module types and spatial distribution. Another benefit for having the performance data is it provides information on the economic viability of BTM solar and support the detection of possible design defects and maintenance requirements especially for small scale PV investors.

Currently, various methods have been proposed to obtain the performance data. The U.S. Department of Energy offers funding support to researches exploring options to making BTM solar more visible: a combination of satellite weather data and detailed rooftop PV interconnection data is used to generate solar data sufficiently accurate for regional or utility scale aggregated output as input to California Independent Systems Operator (CAISO) automatic load forecasting system [6]. Another proposition with widespread acceptance in the PV community is to use the performance Ratio (PR) of the considered PV systems obtained over a long period of weather

patterns to approximate the output from the same systems receiving specific levels of insolation [7]. The PR of a PV system is simply the actual energy output from a system scaled by the rated array capacity at standard test conditions and the plane of array (POA) irradiance received during the period being considered. The problems with these methods lies in the fact that they rely on error free and continuous irradiance data acquisition from ground mounted pyranometers in the plane of array of the PV systems and imagery that requires additional processing. For a regional PV fleet spread over a wide geographical area, these methods would be financially cost intensive and instrumentation reliability issues could prove counterproductive to accuracy requirements.

Alternatively, we do not need to monitor every PV system continuously and directly or require weather data. Several other propositions to addressing the problem already discussed are data mining based options that rely on the assumption that the performances of neighboring PV systems exhibit sufficient levels of correlation to support the prediction of a system's output using the performance data of neighboring systems. [5] proposed a method that depends only on the power generation and location (Euclidean distance between PV sites) data of a small subset of representative PV sites to estimate the total output of a large set of PV sites. Upon analysis, this method was found to perform adequately well for hourly resolution during clear-sky periods but showed a high degree of mismatch between predicted and actual values during cloudy periods owning to factors including ignoring the impact of PV orientation on the generation output. [7] developed the Bird Performance Index (BPI) that rely only on geographical location, geometry and power output measurements of N-1 systems to predict for N systems, but this method produced results that were reasonably accurate only for weekly and longer terms power generation predictions.

In this subsection, our central focus and contribution to advancing knowledge involve the formulation of a proposition aimed at identifying a potentially robust and trustworthy data source, thus tackling the prevailing issue of data accessibility. We recommend the concept of a web-based monitoring and reporting model, designed to offer a viable solution.

Furthermore, this subsection is enriched by furnishing illustrative case studies. These cases showcase existing platforms that cater to owners of independent photovoltaic (PV) systems, enabling them to effectively oversee their energy generation. Some of these platforms even extend

their capabilities to include energy consumption monitoring, providing a comprehensive utility for system owners.

2.2 Web-Based Monitoring and Reporting

This approach is premised on the rationale that every PV system could be a data logger or smart meter. Modern PV systems are equipped with data logging and communication devices for automatic live monitoring and reporting of generation output on webpages in the public domain or with controlled access. "*It is very inexpensive to add data loggers to a PV system at installation time, and the communication cost are very low. Often homeowners themselves are interested in monitoring their PV system. As the power grid is a shared resource, regulators and the power companies need to apply more effort to collect the data necessary for making informed decisions*" [4].

The proposed approach requires utility operators to embark on the following:

1. Set up central repositories or data collection servers (webpages) for PV systems within their jurisdiction.
2. Create free accounts for PV system owners for automatic power generation data reporting and monitoring.
3. As part of the conditions in their legal and regulatory framework for power grid interconnection, require prospecting PV system owners in their offtake agreement to continuously monitor and display live power generation on the reporting platform for the period the PV system will be grid-tied.
4. Incentivize existing PV system owners to join the program. The incentivization could be providing free smart meters. Alternatively, utility operators can place the requirement for communicating output performance at the discretion of PV owners but tie certain benefits to compliance.
5. Sensitize and mandate licensed PV modules installers to link new systems to the web platform and voluntarily act as support administrators for the web repository.

A diagrammatic illustration of the proposed concept is show in Figure 2.1.

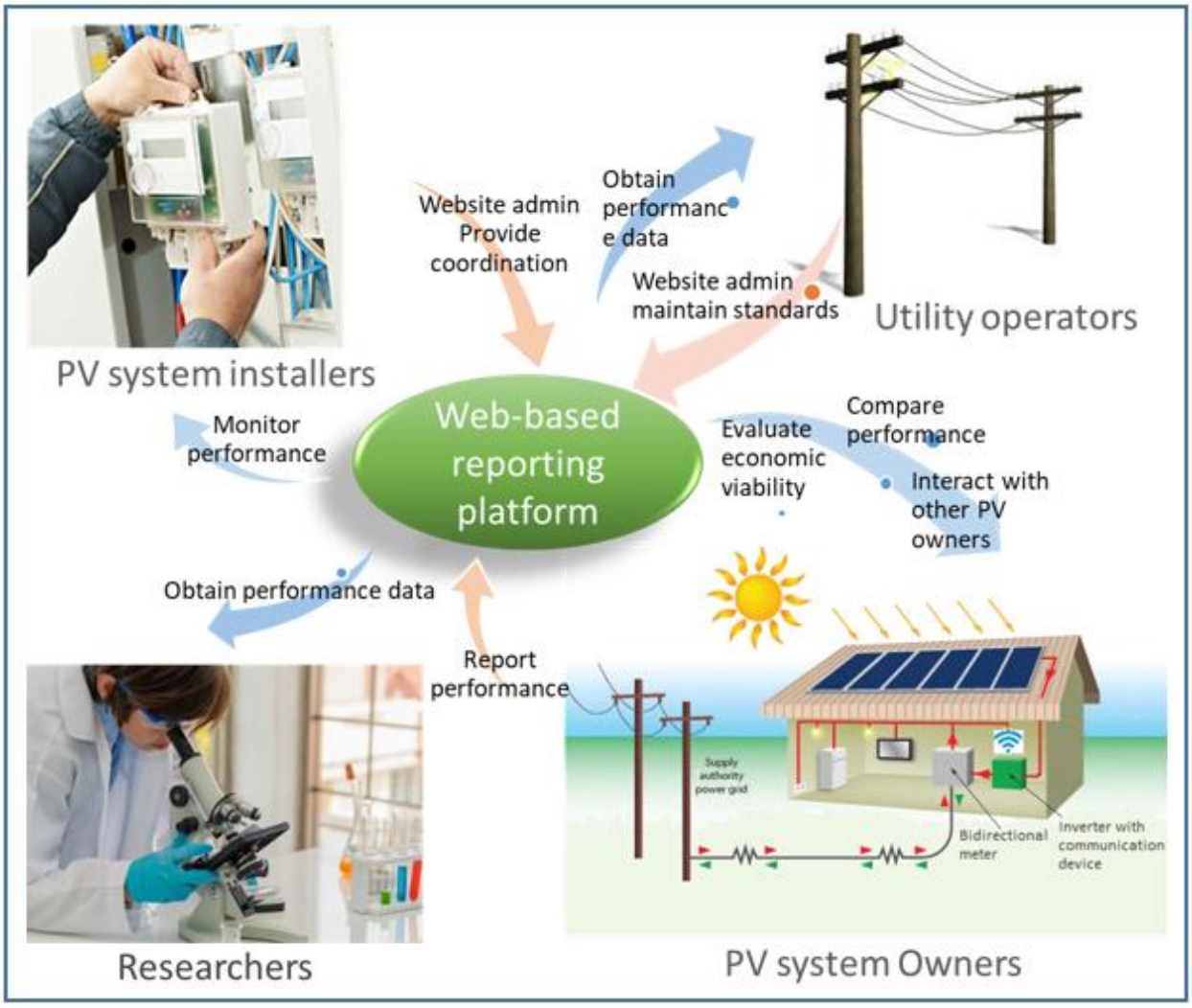

Figure 2.1. Diagrammatical illustration of the proposed web-based concept

At present, small scale PV systems are not required to communicate their generated power to utility operators in real-time. In SaskPower's Interconnection Agreements and Requirements document [8] which was revised in March 2005, it is clearly stated that the costs of future changes of generation facilities as required by SaskPower are the sole responsibility of generation investors. Also, as a metering requirement for interconnection of generators of capacity larger than 100kW, the brunt of providing facilities for communication and interfacing with and ensuring that generation output data is shared with SaskPower's metering unit rest on the shoulders of generator owners. It is presumed that this clause wasn't inserted for smaller systems because they constituted less risks as independent and sole systems and therefore their instrumentation for real-time monitoring was considered negligible. As earlier discussed, the trend in the proliferation of small-

scale distributed resources whose aggregated installed capacity on a regional level would be large enough to cause a significant disturbance in the grid is a compelling argument for the current standards to be reviewed and amended for enhanced PV data collection and communication.

Currently, there are online platforms that epitomize the model being discussed and although these platforms are not restricted on geographical regional basis, they offer various kinds of data and levels of access to the public domain. Table 2.1. contains a number of such existing webpages and synthetic PV data source.

Table 2.1. Examples of existing online PV Data sources with brief descriptions of their contents

Name of Webpage	**Description**	**Number of PV Systems**	**Predominant Country**
PVOutput[1]	Free online service	35,279 systems	Australia
Sunny Portal[2]	Private organization involved in PV systems design equipped with micro inverters for optional data sharing	Over 83,187 systems	Germany
National Renewable Energy Laboratory[3]	U.S. Department of Energy (2006 synthetic data)	6000 hypothetical solar plants	United States
Solrenview[4]	Subscription-based monitoring services provider	Over 4,942 systems	United States

Some potential benefits of creating the proposed platform where BTM solar output data are shared include:

[1] https://www.pvoutput.org/ladder.jsp
[2] https://www.sunnyportal.com/Templates/PublicPagesPlantList.aspx
[3] https://www.nrel.gov/electricity/transmission/solar_integration_%20methodology.html
[4] https://www.solrenview.com/cgi-bin/cgihandler.cgi?&sort=site_name&logo

1. It could be a primary source of numerous data containing excellent information of current real-world residential PV power generation performance for other methods of output estimation.
2. It will allow for comparisons between monitored systems.
3. It provides a basis for both technical, economic, and environmental impact assessments.
4. Researchers could leverage on the primary data from the webpage for technological innovations and advancements in the renewable energy world.

In SaskPower 2017 Stakeholder Engagement and Call for Evidence report [9] in collaboration with university researchers, stakeholders unanimously identified lack of PV education and community engagement as one of the barriers to growth in invisible PV investments and recommended enlightenment and support programs. The webpages could be made to have platforms for community discussion support forums where registered members talk about the health of their systems and share information relevant to the PV community – PVOutput offers this service. In addition, other estimation methods become outdated over time because of PV systems added to the grid and decommissioned or temporarily nonfunctioning systems either due to technical failures or dynamic site-specific issues like trees and high-rise buildings obscuring the path of the sun; snow covering and other objects casting shadows on the panels.

Notwithstanding the plausible benefits highlighted above, there are certain concerns raised about the feasibility of monitoring every grid-tied PV system. Without addressing these concerns, the proposed approach will be practically unfeasible and counterproductive at best. The bulk of the issues raised can be summarized into extra cost of instruments for measurement and data transfer; uncoordinated user-subjective data handling and archiving; inconsistency in uploading data due to instrumentation unreliability. The impact of erratic reporting situations caused by failure of communication devices could be reduced by archiving the data at the generation point during downtimes and then telemetered when communication is reestablished. In addition, residential PV system owners are concerned about the security of their connections to the internet for data transfer to prevent unauthorized access to their private information. As such, the data transfer to public network servers should be done through secured authentication means and should not require the opening of extra ports on homes owners firewall [4].

Unregulated or uncontrolled erroneous data (especially the site metadata) reporting can invalidate the benefits from accruable from continuous direct monitoring. The importance of correctly reporting the hardware data including the solar panel orientation, PV size rating and the site location cannot be overemphasized. Quality control techniques for detecting and correcting misreported PV data require accurate clear-sky models of PV output and rely on accurate PV site metadata to simulate the physical limits of the daily power measurements and in turn identify unexpected or outlying data [10]. The basic essence of the clear-sky model is to eliminate from the measured performance data the deterministic diurnal and seasonal solar variation signals caused by the relative motion between the earth and the sun leaving us with only the cloud-induced fluctuations.

2.3 Credibility of Data Reporting: PV Power Modelling

To recommend solutions for the concerns raised, it was needful to understand the impact of erroneously reported array orientation on simulated PV performances and the role utility operators can play in minimizing spurious reporting and increasing participation of more rooftop PV systems. Another objective is aimed at evaluating the credibility of the operational and performance data reported on existing public webpages by registered users whose PV systems are equipped with communication devices for live web-based monitoring and reporting.

The main idea here was to simulate the output of each PV system at multiple orientations and contrast the results with the measured power output reported on the websites to check for intolerable inconsistencies. A similar research was carried out by the Australian Renewable Energy Agency (ARENA) [4] who had data loggers installed at residential buildings reporting rooftop PV performance on PVOutput.org and compared the reported performance with the results from the data loggers.

Two simulation methods were adopted for this research. One involves combining the reported metadata with irradiance data from nearby ground weather station to model the hourly PV output whereas the other method relies on the European Commission PVGIS web application modeling tool.

Regardless of the simulation method adopted, the PV time series of each site was simulated using tilt angles within a tolerance of about 50% of the reported or expected value. To illustrate this, for a reported tilt of 20^0, a whole year's PV output will be simulated from 10^0 to 30^0 at least with 1^0 step until a reasonable pattern is observed from the results. This tilt angle bracket was adopted from careful observations of inclined roof-tops and array images reported by PV owners. Similarly, a range of values from 0^0 to 315^0 at intervals of 45^0 were considered for azimuth angle of every PV system. The angular measurements are usually from two reference directions (North and South) through a reference plane and towards the point of interest as shown in Figure 2.2. A 1^0 step was deemed unnecessary because of the minimal changes in the results obtained for the step size. If the 1^0 step was adopted, the azimuth angle would have 360 different possible outcomes and that coupled with the minimum of 20 outcomes for the tilt angle would yield a combined outcome of 7200 (360 times 20) for a single data point. For a whole year and assuming 9 daytime hours, a single PV system would have approximately 24million (365 times 9 times 7200) hourly data points.

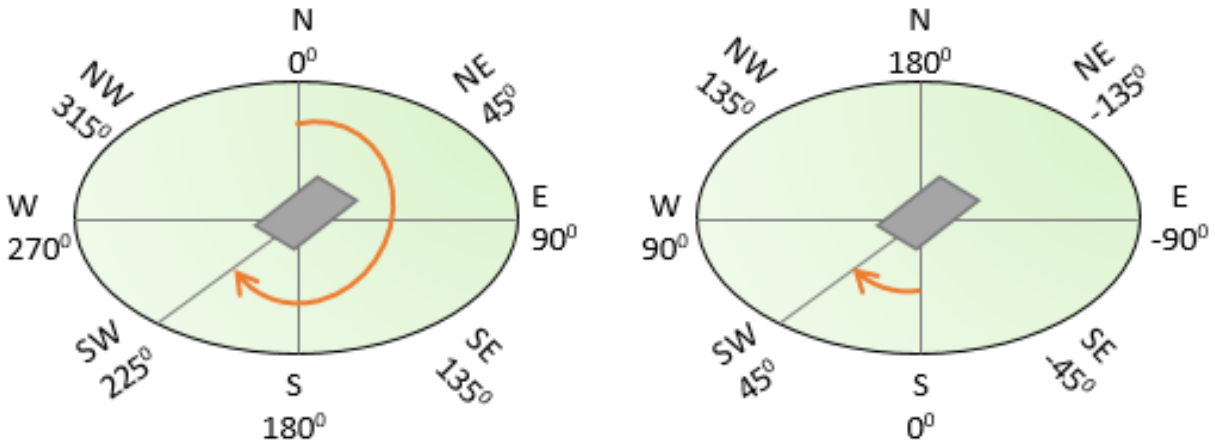

Figure 2.2. North and South reference directions for evaluating the azimuth angle. For the South (S 00), the measurement goes in either the clockwise or the anticlockwise direction.

2.4 PV Power Output Modeling

2.4.1 Combination Of Ground Station Data and PV Metadata

Historical one-minute averages of solar radiation dataset for a calendar year was obtained from the Australian Bureau of Meteorology ground-based weather station in Townsville. The solar

statistics included Solar Zenith Angle (SZA), Global Horizontal Irradiance (GHI), Direct Normal Irradiance (DNI) and Diffuse Horizontal Irradiance (DHI) obtained by point sensors at the horizontal plane. Measured power production data from the selected PV systems time-stamped and reported every five minutes on PVOutput server were also obtained for same period.

2.4.2 Data Quality Control

Two major issues were observed with the measured one-minute data: sensor deterioration-induced errors and in some cases a valuable percentage of missing measurements. Errors with the measured SZA were observed during periods of low solar altitude angles or when the sun was near the horizon which were likely due to the fact that the measurements were taken at a height of 4.4m above ground level. As such, if SZA >90^{o} , Cos(SZA) was set to 0. Errant irradiance values that were physically impossible or were quite larger than the clear-sky power output and negative values were also removed. There were periods of partial equipment fault resulting in only the GHI and SZA values recorded while total malfunction meant no data recorded. Periods of full outages lasting over one hour and days having up to three hours were filtered out. Consequently, the daily study period was limited to be from 9am to 5pm. For partial outages, the missing data points were either interpolated from other values or estimated using a decomposition model or obtained using the relation:

$$GHI = DHI + DNICos(SZA) \quad (Wm^{-2}) \tag{1}$$

The Plane of Array (POA) irradiance measurements is very critical to evaluating the performance of PV systems because most practical PV systems in order to capture more of the available solar radiation are usually installed with varying orientations relative to the horizontal. Since in this case, only the irradiance data measured by point sensors on the horizontal plane are available, there is need to convert this data to fixed axis POA irradiance values for each of the select PV systems. Generating estimates of the power generations required two basic steps involving estimating the Plane-Of-Array Irradiance (I_{POA}) and then used as radiation resource input to simulate the output power. Estimating I_{POA} from available measurements requires a combination of decomposition and transposition models.

2.4.3 Irradiance Decomposition and Transposition

Several empirical estimation models exists but for this analysis, the slightly modified Orgill and Hollands' radiation model [11] was the preferred decomposition model to estimate the DHI component from the GHI dataset because of its mathematical simplicity and minimum input requirements of clearness index (K_{GH}) and site latitude. More importantly, the model was also found to outperform other models in this analysis because its parameters were calibrated uniquely to locations in Australia. The diffused fraction was first obtained using expressions (2) – (10) then (1) was used to estimate the DNI. Zero subscripts indicate performance at STC.

$$I_{DH} = I_{GH} K_{DH} \tag{2}$$

$$K_{DH} = B - 0.3C \qquad if\ K_{GH} \leq 0.3 \tag{3}$$

$$K_{DH} = B - K_{GH} C \qquad if\ 0.3 \leq K_{GH} \leq 0.75 \tag{4}$$

$$K_{DH} = B - 0.75C \qquad if\ K_{GH} \geq 0.75 \tag{5}$$

$$B = 0.94 + 0.0118|\emptyset| \tag{6}$$

$$C = 1.185 + 0.0135|\emptyset| \tag{7}$$

The expression for the clearness index is given by (8)

$$K_{GH} = \frac{I_{GH}}{I_{ext_h}} \tag{8}$$

$$I_{ext_h} = I_{ext} Cos(SZA) \tag{9}$$

$$I_{ext} = 1367\left(1 + 0.033 \cdot Cos\left(\frac{360}{365}\right) \cdot DOTY\right) \tag{10}$$

where:

DOTY is the day of the year,

$\emptyset$ is the latitude,

I_{ext} is the extraterrestrial irradiance, and

I_{ext_h} represents the extraterrestrial irradiance at the horizontal plane.

Transposition models estimate the total POA irradiance obtained from the basic equations (11) – (17) by calculating and summing up the direct, ground reflected diffuse and the diffuse component from the sky hemisphere on the array plane.

$$I_{POA} = I_{DN_{POA}} + I_{grd_{POA}} + I_{sky_{POA}} \tag{11}$$

$$I_{DN_{POA}} = I_{DN} \cdot \cos AOI \tag{12}$$

$$I_{grd_{POA}} = I_{GH} \cdot \sigma \cdot \frac{1 - \cos\alpha}{2} \tag{13}$$

The POA sky diffuse radiation was calculated using Hay/Davies' model [12] since it was found to give the best results for this analysis:

$$I_{sky_{POA}} = I_{DH} \cdot \left[K_{DN} \cdot R_b + (1 - K_{DN}) \cdot \frac{1 + \cos\alpha}{2}\right] \tag{14}$$

$$R_b = \frac{\cos AOI}{\cos SZA} \tag{15}$$

$$K_{DN} = \frac{I_{DN}}{I_{ext}} \tag{16}$$

$$AOI = \cos^{-1}[\cos SZA_S \cdot \cos\alpha + \sin SZA_S \cdot \sin\alpha \cdot \cos(\delta_S - \delta)] \tag{17}$$

where:

α represent the array tilt angle,

σ is the surface albedo, and

AOI is the sun's angle of incidence.

It is worthy of note that since these models are purely empirical, their accuracies stated in the literature are subjective to the test data used in evaluating them. This means that a model could have lower accuracy levels when used on data from a different geographical location from the test data. Details of additional expressions for evaluating the model parameters are in [13].

2.4.4 Power Output Simulation

The horizontal solar radiation was transposed to different POA (tilt angles) and then the DC power output from the PV array was simulated using the PVWatts array performance model [14] provided by NREL because of its simple input parameters requirements and its average performance is similar to SAPM performance:

$$P_{mp_{DC}} = \frac{I_e}{1000} \cdot P_{mpo} \cdot [1 + \vartheta(T_c - T_{STC})] \quad for\ I_{POA} > 125W/m^2 \tag{18}$$

$$P_{mp_{DC}} = \frac{0.008 \cdot {I_e}^2}{1000} \cdot P_{mpo} \cdot [1 + \vartheta(T_c - T_{STC})] \quad for\ I_{POA} \leq 125W/m^2 \tag{19}$$

where:

$P_{mp_{DC}}$ represents the module power output at maximum power point,

P_{mpo} is the module power rating,

I_e is the POA effective irradiance,

ϑ is the module power temperature coefficient, and

T_c is the PV cell temperature respectively.

CEC and Sandia National Laboratories carried out detailed field and laboratory test on a plethora of commercial inverters and PV modules to generate a database[5][6] of inverter and modules parameters. The parameters for the inverters and modules for this analysis were obtained from this database. The empirically based Sandia Inverter model (19) – (23) [14] produced the best results for the AC power output $P_{mp_{AC}}$ from the input power $P_{mp_{DC}}$:

$$P_{mp_{AC}} = \left[\frac{P_{mp_{AC0}}}{X - Y} - Z(X - Y)\right] \cdot \left(P_{mp_{DC}} - Y\right) + Z\left(P_{mp_{DC}} - Y\right)^2 \tag{20}$$

$$X = P_{mp_{DCo}}\left[1 + C_1\left(V_{mp_{DC}} - V_{mp_{DCo}}\right)\right] \tag{21}$$

$$Y = P_{so_{DC}}\left[1 + C_2\left(V_{mp_{DC}} - V_{mp_{DCo}}\right)\right] \tag{22}$$

[5] https://pvpmc.sandia.gov/download/6840/

[6] https://pvpmc.sandia.gov/applications/pv_lib-toolbox/

$$Z = C_o\left[1 + C_3\left(V_{mp_{DC}} - V_{mp_{DCo}}\right)\right] \quad (23)$$

where:

$V_{mp_{DC}}$ is the input DC voltage at optimum condition,

$V_{mp_{DCo}}$ is the optimum DC voltage at STC,

$P_{mp_{AC0}}$ is the inverter optimum AC power rating at STC,

$P_{mp_{DCo}}$ is the optimum DC power level at STC, and

$P_{so_{DC}}$ is the inverter starting DC power respectively.

2.4.5 Power Output Simulation: PVGIS Modeling Tool

Admittedly, power simulations using irradiance data from ground-based weather monitoring stations would offer more accurate results for reasons being that measurements are taken using high-quality sensors at a similar elevation as the PV sites at close proximities to the weather stations. However, maintaining such ground-based stations is tedious and cost-intensive especially as it requires regular calibration and cleaning of the sensors to minimize systematic errors. Hence, a relatively few numbers of these stations exist and are usually situated remotely from one another leaving a gap which is being filled by solar radiation data from satellite images.

Consequently, since ground-based weather stations at locations considerably close to some of the selected PV sites for this research didn't exist, the PVGIS tool was used to model the power outputs of these sites. PVGIS was developed by the European Commission Joint Research Centre to aid PV performance studies. The online PVGIS 5 interface leveraged on a database of surface solar radiation estimates from satellite images of cloud cover to evaluate PV performances of hourly time resolution mapped to UTC time zone. The tool used a quadratic mathematical model (24) – (26) to approximate the impact of the input parameters on module efficiency and output [15]:

$$P_{mp_{AC}} = \hat{I} \cdot \left(P_{mpo} + c_0 \cdot \ln \hat{I} + c_1 \cdot \ln \hat{I}^2 + c_2 \cdot \hat{T} + c_3 \cdot \hat{T} \cdot \ln \hat{I} + c_4 \cdot \hat{T} \cdot \ln \hat{I}^2 + c_5 \cdot \hat{T}^2\right) \quad (24)$$

$$\hat{I} = \frac{I_{POA}}{1000} \quad (25)$$

$$\hat{T} = T_c - T_{STC} \quad (26)$$

where:

$c_0 - c_5$ are the coefficients of the quadratic model calibrated by measured data,

I' is the normalized plane of array irradiance, and

$\hat{T}$ is the module temperature.

Primary input data required include coordinates of the PV site; the array orientation (tilt and azimuth); the system loss factor; the period being considered; the PV nominal power and technology.

2.5 Case Study Public Webpages

Two notable websites from which a number of registered PV systems were selected after being subjected to certain criteria are PVOutput and Solrenview. These were selected for this analysis because they have contrasting profiles in terms of level of regulation and geographical region of dominance. While Solrenview is concentrated on the US, and the telemetering devices are linked to the subscription-based webpage by licensed PV or data acquisition service facilities installers, the reverse is the case for PVOutput which is a user-driven database and PV owners report their data unchecked. Both sites have a high number of registered systems and in addition, the PV sites metadata and output measurements are available in the public domain without strenuous requirements for accessibility. A select sample of PV systems could not be aggregately analyzed but were treated as independent entities because they had differing orientations and geographically distant from one another. For this reason and the fact that only a few PV systems meet the selection criteria, only a total of 9 PV sites were selected from both webpages.

2.5.1 PVOutput

The choice of PV systems handpicked from the databases fulfilled the following conditions: they were continuously monitored for at least one calendar year; there was adequate description of the PV systems metadata; inverter capacity was larger than or equal to the array rating to avoid issues resulting from curtailment, and they had zero shading levels. 6 PV systems were selected from PVOutput with profiles shown in Table 2.2. The PV power data for each of the

systems in the box plots in Figure 2.3 are normalized to their nominal capacities. From the plots, the maximum instantaneous power generated from any of the six systems is about 60% of the array rated capacity except for PV system 2 which contains outliers. The first 3 of these systems are located in Townsville and a ground-based weather station in active service was well within range so their 2018 PV datasets were collected and validated on individual basis with the 2018 PV output times series simulated using the modeling approach that relied on ground irradiance data. The 2016 datasets of the other 3 systems were similarly collected automatically using a web crawler and were compared with PV output time series simulated for same year using the PVGIS tool.

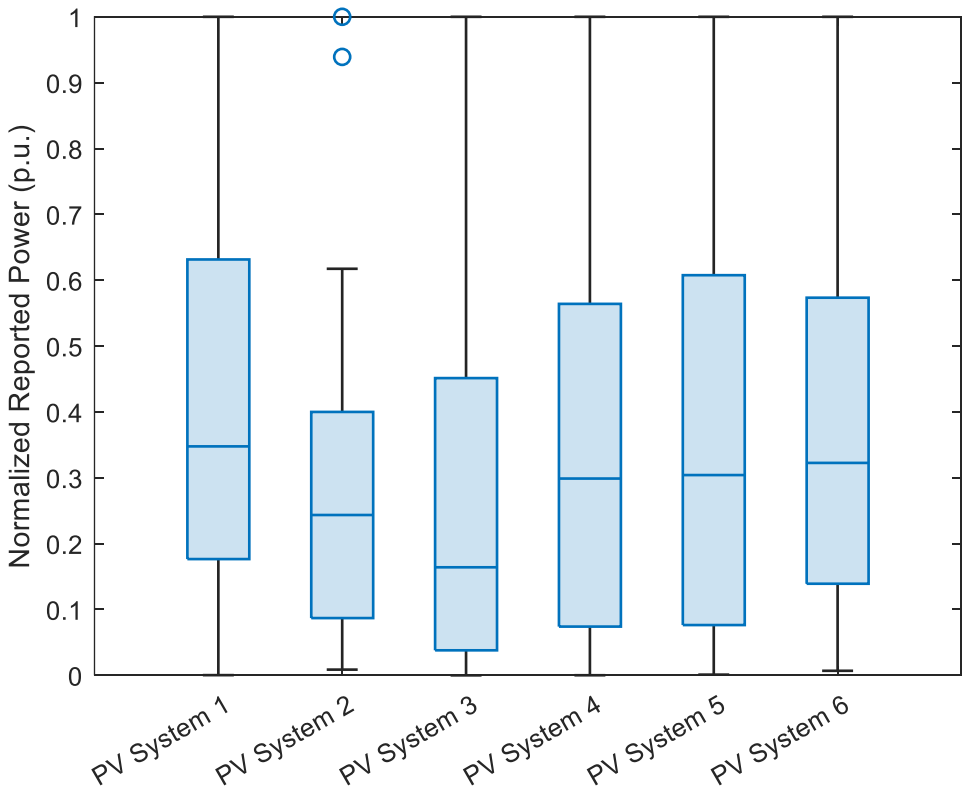

Figure 2.3: Box plots of the six PV systems from PVOutput. The plots are essentially used to detail the presence of outliers and the energy density of the systems.

All 6 PV systems have been reporting their AC power since 2012 in 5-minutes reporting time steps except for system 4 which started in 2015 and had about 9months records in 2016. Analysis on a one calendar year data when the systems were operational for most part of the year was deemed sufficient to capture seasonal trends in the accuracy of reported data and the

consistency in data reporting. In addition to the metadata shown in Table 2.2, all systems also reported their shading levels which was used to screen suitable systems.

Table 2.2: Reported profiles of Selected PV systems from PVOutput

Closest Weather Station: Townsville Australia (-19.2483, 146.7661)[7]			
Site No	**PV system 1**	**PV system 2**	**PV system 3**
Array size	4kW	3.04kW	4.5kW
Number of modules	16	16	18
Distance from weather station	5.63km	9.433km	5.67km
Coordinates (Lat/Long)	-19.2922/146.763	-19.3264, 146.731	-19.292/146.763
Module brand	CSUN250-60M	Linuo-190	SNM-P250(72)
Inverter brand	SMA SB 4000-TL	JSI-3000TL	Aurora - 5000
Tilt (^{0})	1	25	25
Azimuth (0^0N)	Northwest	West	North
Site No	**PV system 4**	**PV system 5**	**PV system 6**
Array size	1.92kW	2.66kW	4kW
Number of modules	8	14	16
Location	Germany	Denmark	UK
Coordinates (Lat/Long)	51.77/6.119	55.1407/15.0125	52.7675/1.166
Module brand	Diverse	China (1-crystal)	Sanyo HIT-H250
Inverter brand	4×Enecsys SMI-D480	Danfoss 3000	SMA SB 4000TL
Tilt (^{0})	20	45	30
Azimuth (0^0N)	Southeast	South	South

Since it was a primary goal of this study to uncover the issues with data from public PV monitoring webpages, data quality control algorithms to comprehensively identify measurement errors were intentionally not implemented. However, the daylight hours considered was restricted

[7] http://reg.bom.gov.au/climate/reg/oneminsolar/

to 9hours (9am to 5pm) local time. This decision was made to minimize the impact of distortions arising from systematic influences like shadows cast on PV modules during sunrise and sunset on the dataset. Also, by visual observation of the PV datasets from PVOutput, the inverters/recording equipment were programmed to cut off data recording either at a particular time or when the generated power fell below a threshold value (usually <3% array rating). For instance, system 1 consistently started/cut off reporting data from/below 48W while system 2 reporting was from 7:05am to 6pm each day.

2.5.2 Results and Analysis

The reported PV datasets were compared with the modeled PV profiles over the approximately one-year time period for hourly time intervals to detail the difference between them since the actual values are presumably not known. Normalized Root Mean Square Difference (NRMSD), Normalized Mean Absolute Difference (NMAD) and Normalized Mean Bias Difference (NMBD) statistical error metrics were used. While NRMSD measures the variation, NMAD reveals the mean absolute deviation of the simulated from the measured performance values and NMBD provides information on the systematic error or bias to account for over estimation or underestimation.

$$NMAD = \frac{\frac{1}{n}\sum_{i=1}^{n}\left|x_{m,i} - x_{r,i}\right|}{P_{mpo}} \times 100\% \tag{27}$$

$$NRMSD = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_{m,i} - x_{r,i}\right)^2}}{P_{mpo}} \times 100\% \tag{28}$$

$$NMBD = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_{m,i} - x_{r,i}\right)}{P_{mpo}} \times 100\% \tag{29}$$

where:

n is the number of values,

$x_{m,i}$ is the *i*th modeled output values,

$x_{r,i}$ is the *i*th reported outputs values.

Figures 2.4 – 2.9 display how the simulated power output changes relative to the measured counterpart while changing the inclination and azimuth angles as inputs to the simulation model obtained for the 6 PV systems. Since there is no standard threshold error value for determining the acceptable levels of disparity between reported and simulated values, the smallest difference (error) values obtained using the error metrics were regarded as best results and are shown in Table 2.3.

Table 2.3. Best results obtained for the PV systems and the corresponding tilt angles for the study period. Range of values indicates that the lowest NRMSD and NMAD values do not coincide on the same tilt angle.

PV system	**1**	**2**	**3**	**4**	**5**	**6**
NMBD (%)	1.85	-1.30	0.33	-1.38	1.50	1.90
NMAD (%)	4.46	6.20	9.35	6.47	9.60	9.10
NRMSD (%)	6.85	8.45	12.1	10.13	14.65	12.90
Reported tilt (^{0})	1	25	25	20	45	30
Best tilt (^{0})	1	26-27	15-20	3-5	34-35	12-14

Several observations can be made from Figures 2.4 – 2.9 based on the NRMSD and NMAD curves. Generally, in all six cases the best results were obtained at orientation angles that either matched with the reported data or can be considered to be close enough not to introduce a substantial deviation in the simulated from the reported datasets. While the first two systems show a near-perfect match, the third system has its best results revolving around the reported tilt angles with an exception for system 5 being the only case where the reported azimuth angle happens not to be the best case (Southwest), differing by 45^0. If there's anything to go by the steepness of the curves for systems 4 and 6 as they approach the reported tilt angles, the reported power output time series for these two systems are not reliable. A larger discrepancy between the reported tilt angles and the best fit is observed in both cases.

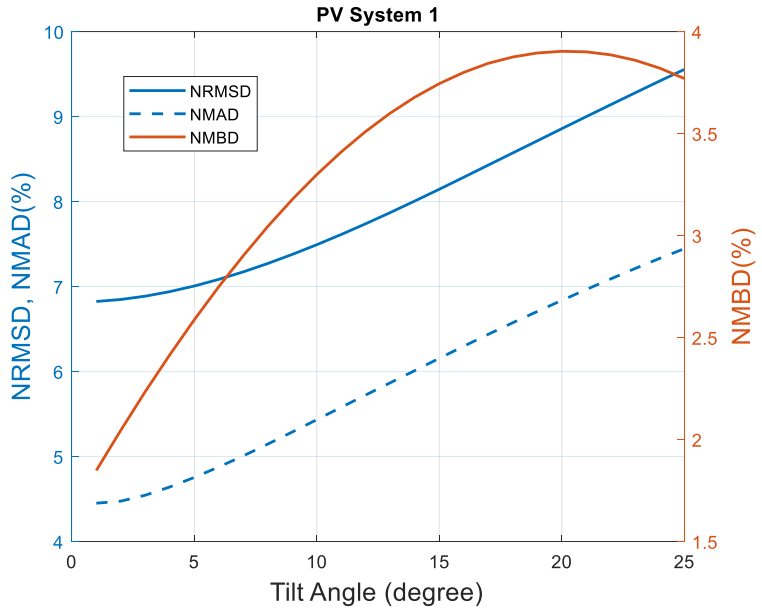

Figure 2.4. Variation of the error measures plotted against a steadily increasing angle of inclination of the PV array as input to the simulation model for PV systems from PVOutput webpage: PV System 1.

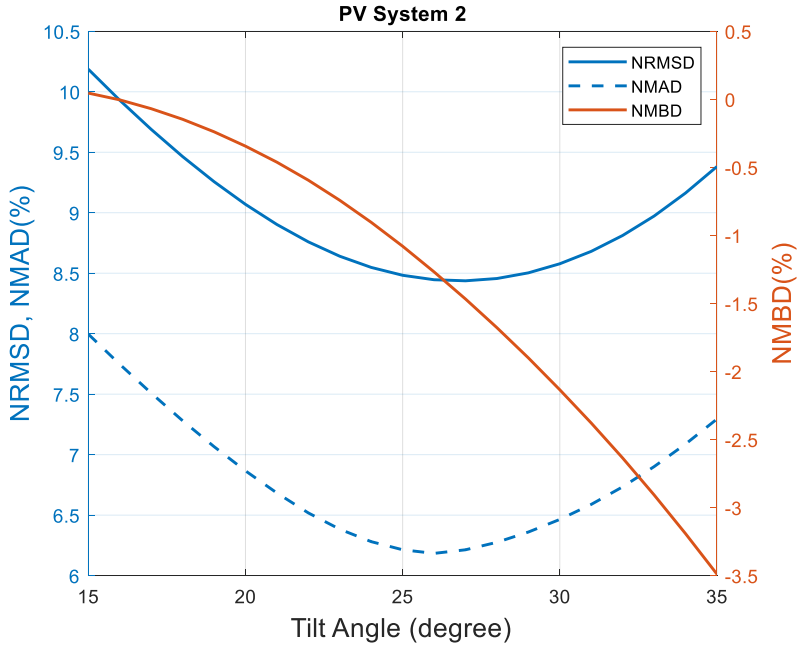

Figure 2.5. Variation of the error measures plotted against a steadily increasing angle of inclination of the PV array as input to the simulation model for PV systems from PVOutput webpage: PV System 2.

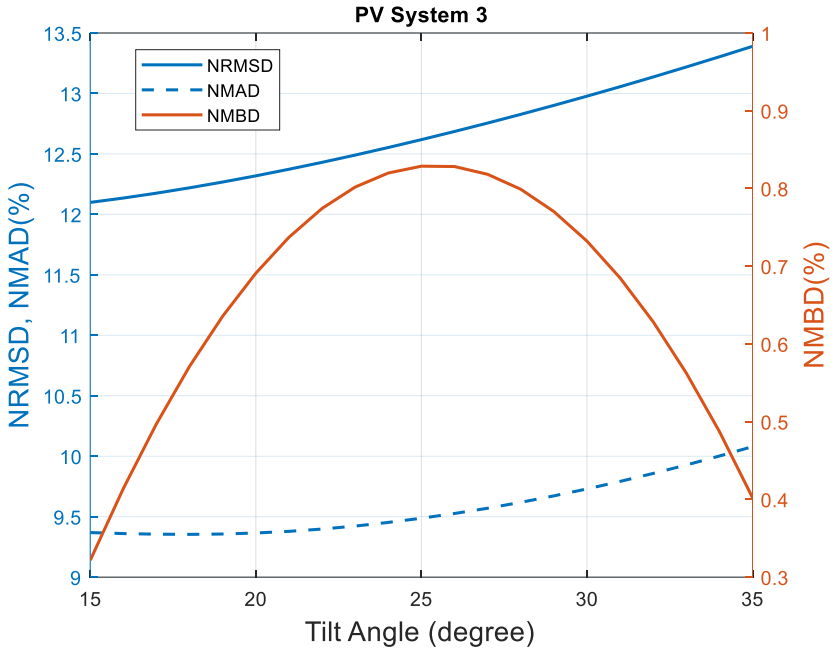

Figure 2.6. Variation of the error measures plotted against a steadily increasing angle of inclination of the PV array as input to the simulation model for PV systems from PVOutput webpage: PV System 3.

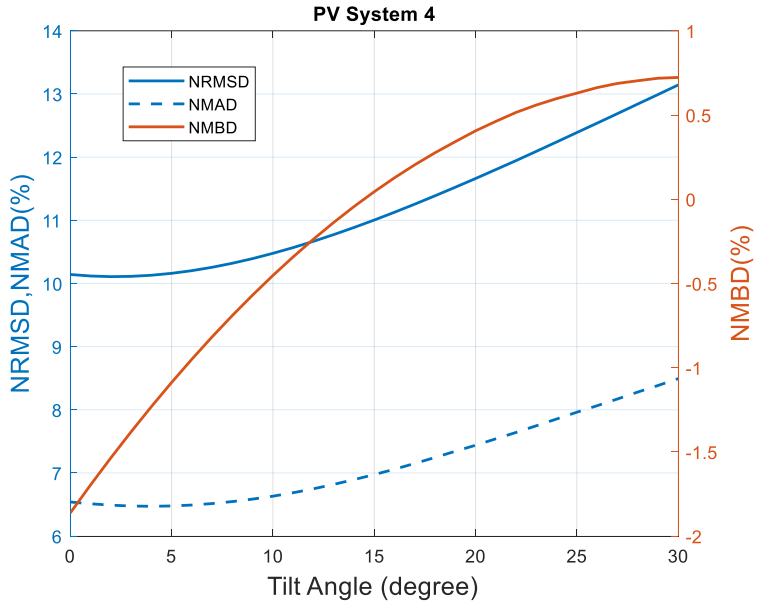

Figure 2.7. Variation of the error measures plotted against a steadily increasing angle of inclination of the PV array as input to the simulation model for PV systems from PVOutput webpage: PV System 4.

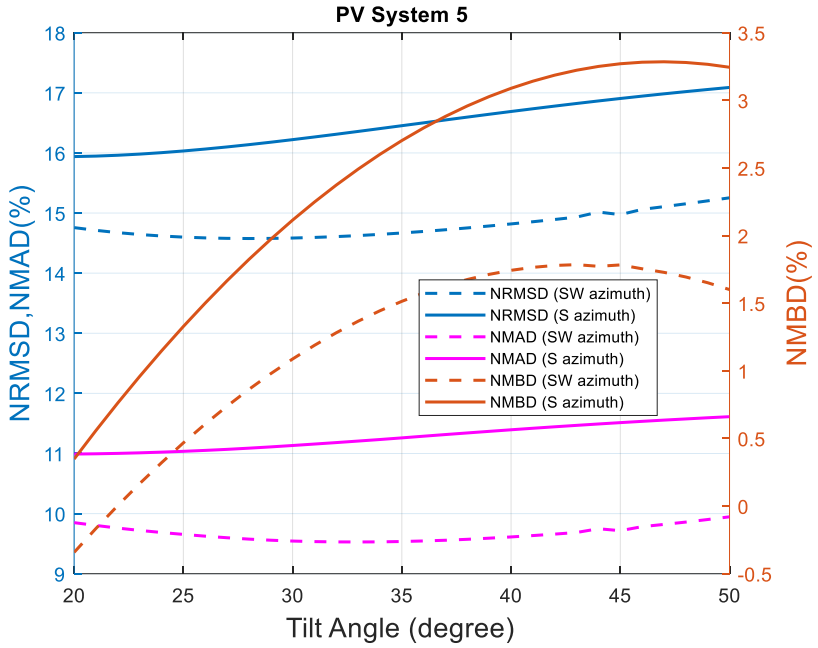

Figure 2.8. Variation of the error measures plotted against a steadily increasing angle of inclination of the PV array as input to the simulation model for PV systems from PVOutput webpage: PV System 5

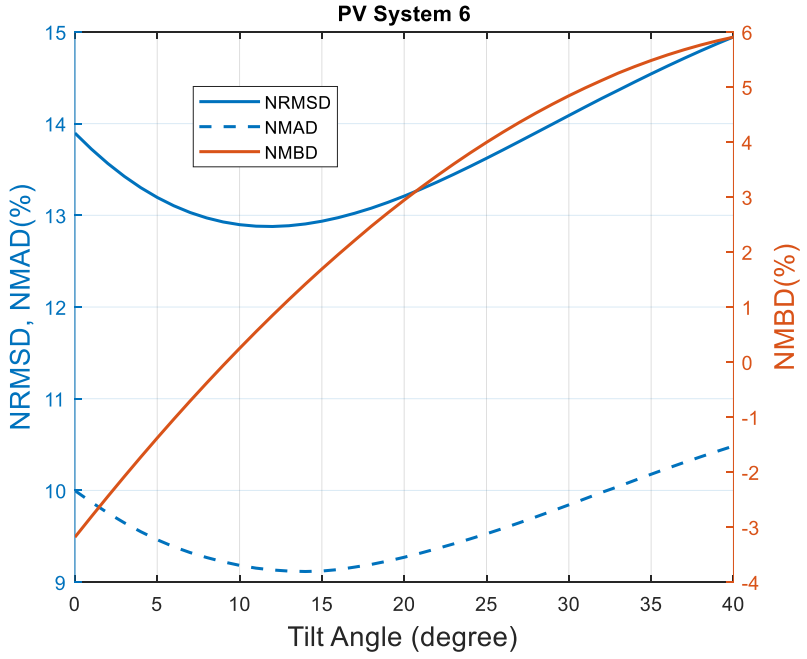

Figure 2.9. Variation of the error measures plotted against a steadily increasing angle of inclination of the PV array as input to the simulation model for PV systems from PVOutput webpage: PV System 6

The error measures for PV system 5 in Figure 2.8 when the reported azimuth angle of due South and best case of Southwest were used as inputs are shown. All other best results were gotten at the same azimuth angles reported by system owners.

Figures 2.10 – 2.15 are scatter plots of the simulated vs. reported hourly averaged power output relationship. They reflect the results obtained for the best inclination angles shown in Table 2.3. The plots were aimed at detailing the overall convergence of one power output dataset to the other while having neither as the standard or reference dataset. In other words, one can visualize the bias between both datasets from the scatter plots while the NMBD in Figures 2.10 – 2.15 show how this bias dynamically changes with the tilt angle however not in a linear manner. A perfect convergence would mean all the blue circles falling on the red line. Comprehensively, by visual observation there appears to be slightly positive bias indicating a power output overestimation. PV systems 1 and 2 show better alignment compared to the others although the best case did not necessarily mean the lowest NMBD.

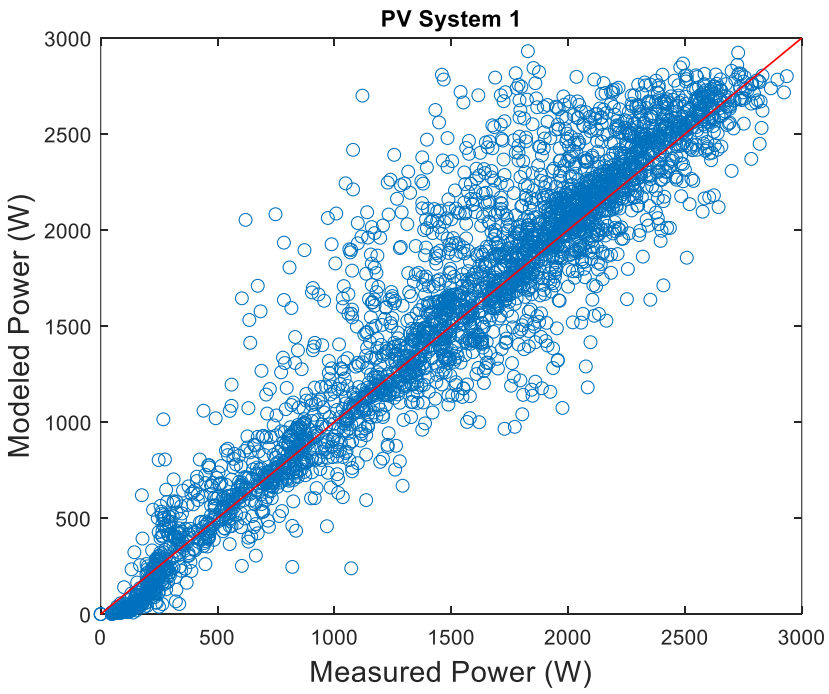

Figure 2.10. Scatter plots of simulated vs. measured or reported averaged hourly power output for the PV systems registered on PVOutput webpage for PV System 1.

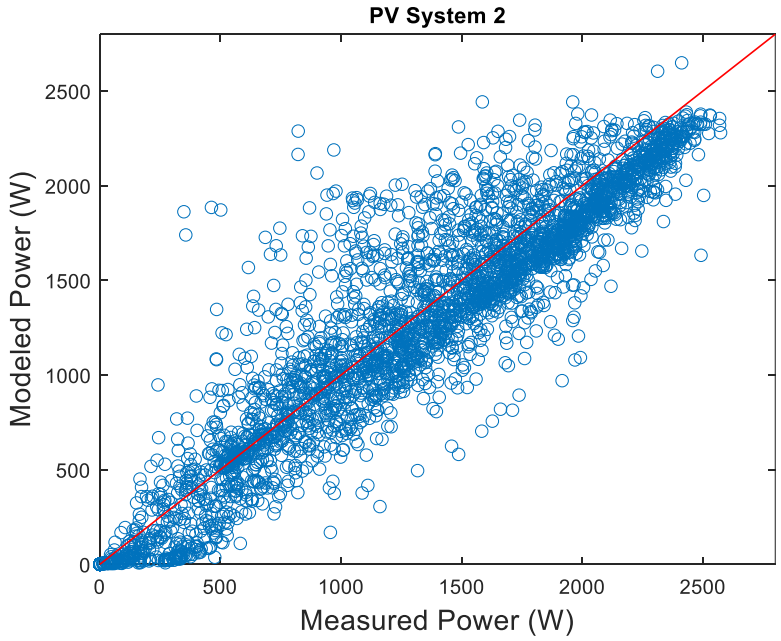

Figure 2.11. Scatter plots of simulated vs. measured or reported averaged hourly power output for the PV systems registered on PVOutput webpage for PV System 2.

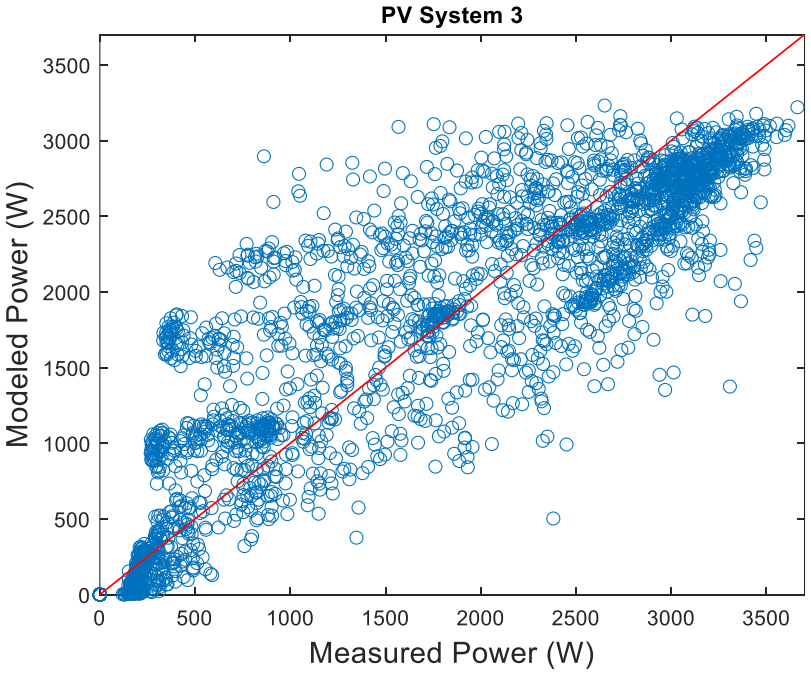

Figure 2.12. Scatter plots of simulated vs. measured or reported averaged hourly power output for the PV systems registered on PVOutput webpage for PV System 3.

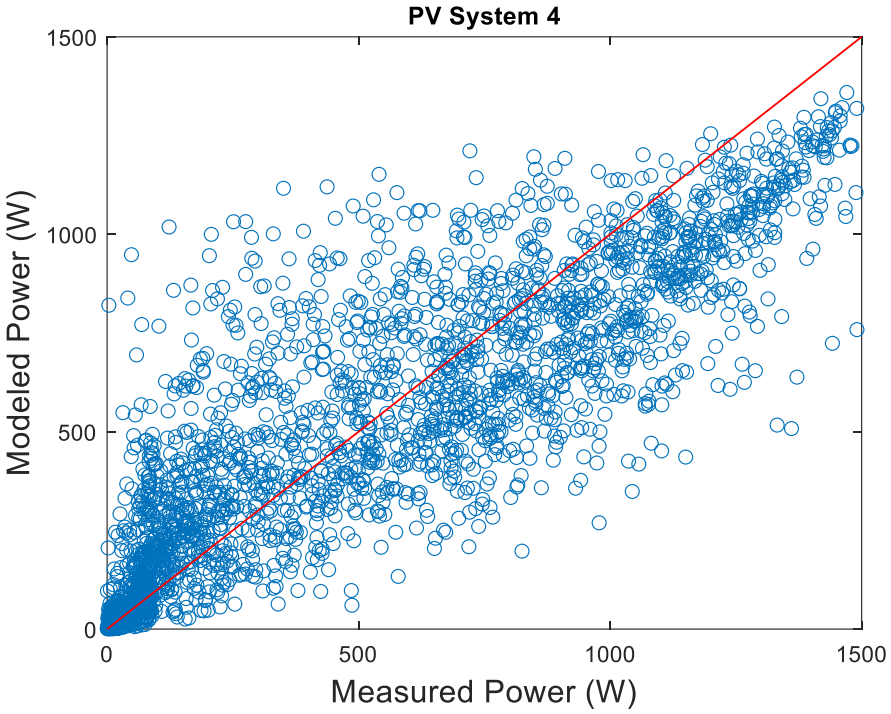

Figure 2.13. Scatter plots of simulated vs. measured or reported averaged hourly power output for the PV systems registered on PVOutput webpage for PV System 4

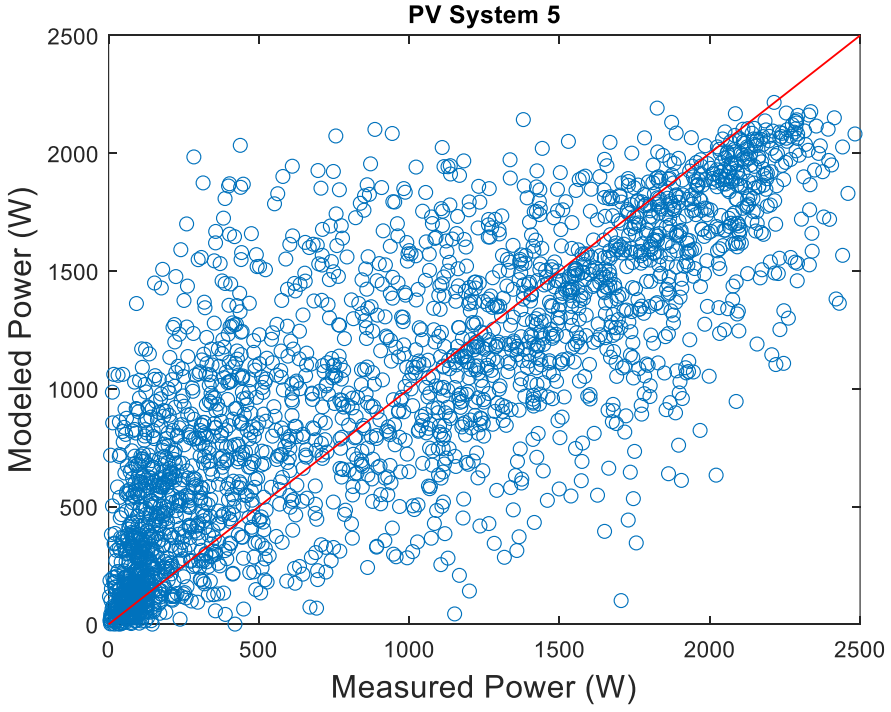

Figure 2.14. Scatter plots of simulated vs. measured or reported averaged hourly power output for the PV systems registered on PVOutput webpage for PV System 5.

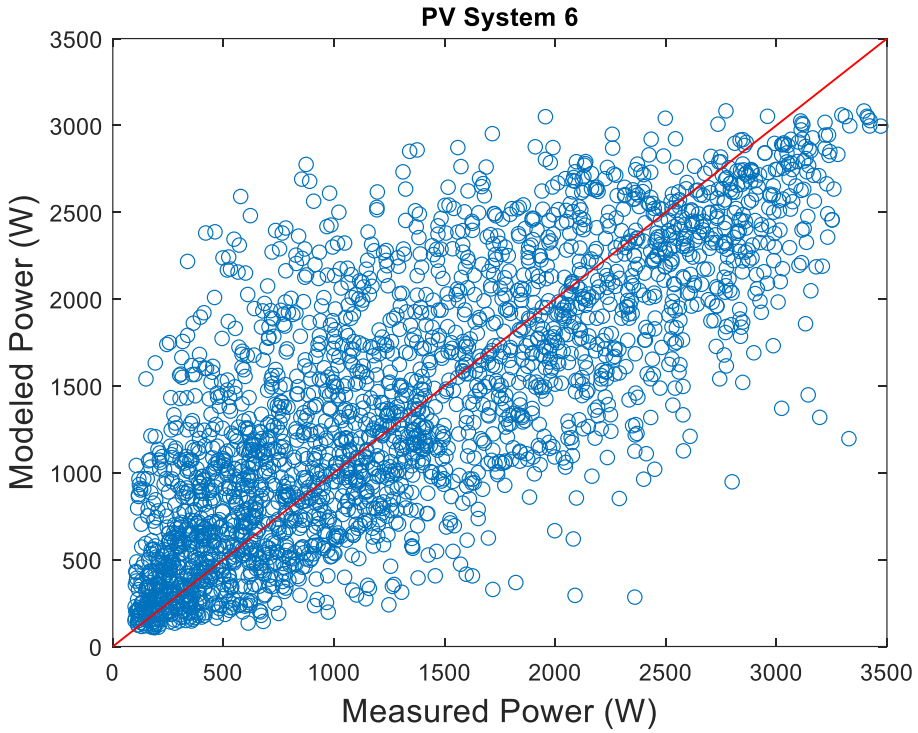

Figure 2.15. Scatter plots of simulated vs. measured or reported averaged hourly power output for the PV systems registered on PVOutput webpage for PV System 6.

Temporal aggregation of the results into monthly energy yield shown in Figures 2.16 – 2.21 uncover some unique features common among systems in the same geographical vicinity or that are expected to receive similar solar radiation. First, there is a better agreement between both datasets for each PV system which once again is more evident in the first three systems. Second, two clearly defined patterns of energy production can be observed. Systems 1-3 from Townsville Australia, have an almost regular energy generation all year round with the lowest output occurring precisely in the month of June while the maximum output was observed in the fourth quarter of the year. The first three months of system 3 were omitted from the analysis because a substantial lot of the dataset in this period was either missing or abysmally low due to faulty data logging equipment or downtime. This downtime couldn't be attributed to shading or other causes because of the prolonged period it occurred on a daily basis.

Systems 4-6 are situated in Europe which is in the northern hemisphere with Due South as the optimum azimuth angle. This second group also has a similar pattern with slight occasional breakaways and the energy outputs peaked towards the middle months of the year. The energy

yields in the winter months are significantly lower due to reduced number of sunshine hours per day and possibly snow cover. The results in Figures 2.16 – 2.21 are quite consistent with the annual weather reports of monthly average solar radiation received in these locations, the only anomaly being the mild deep in output witnessed in June in system 6. The October monthly output for System 4 was removed for same reason as in system 3.

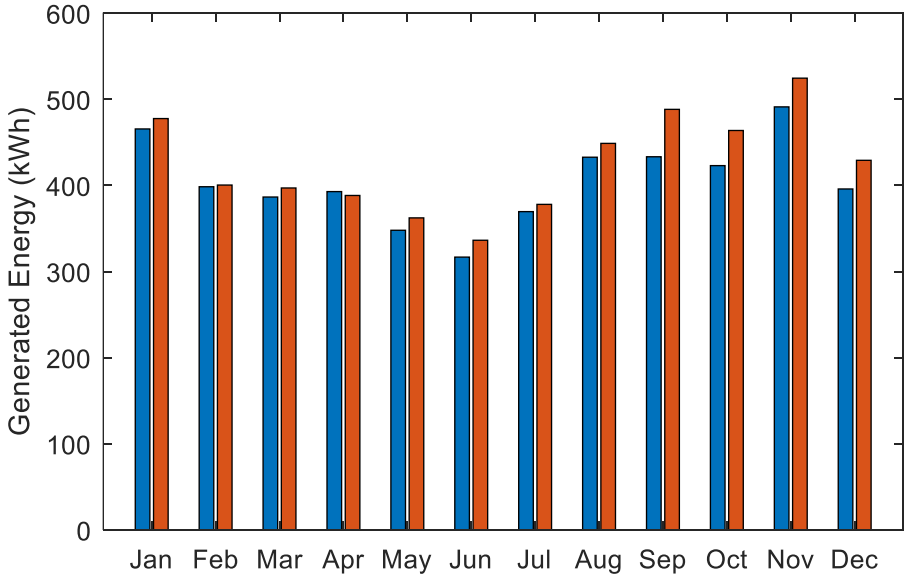

Figure 2.16. Monthly aggregated energy yield of the PV system 1 from PVOutput webpage.

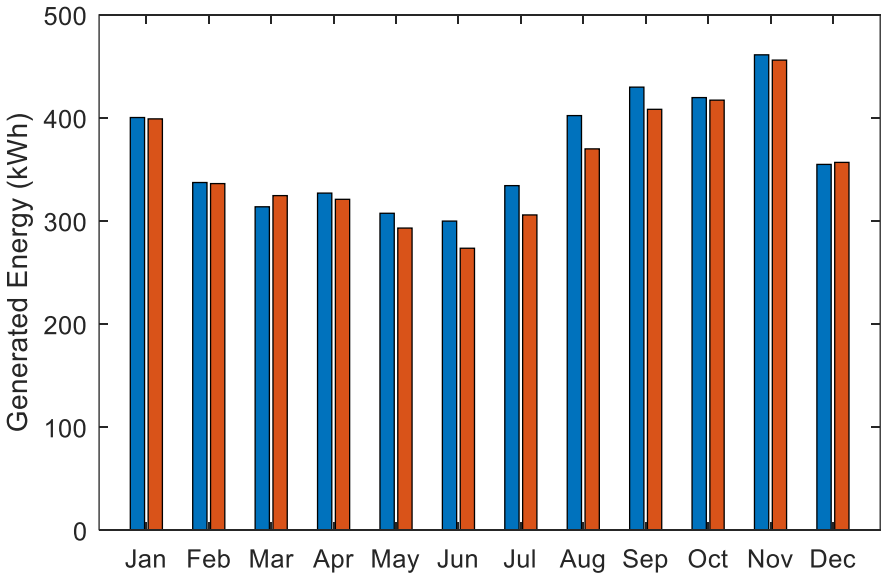

Figure 2.17. Monthly aggregated energy yield of the PV system 2 from PVOutput webpage.

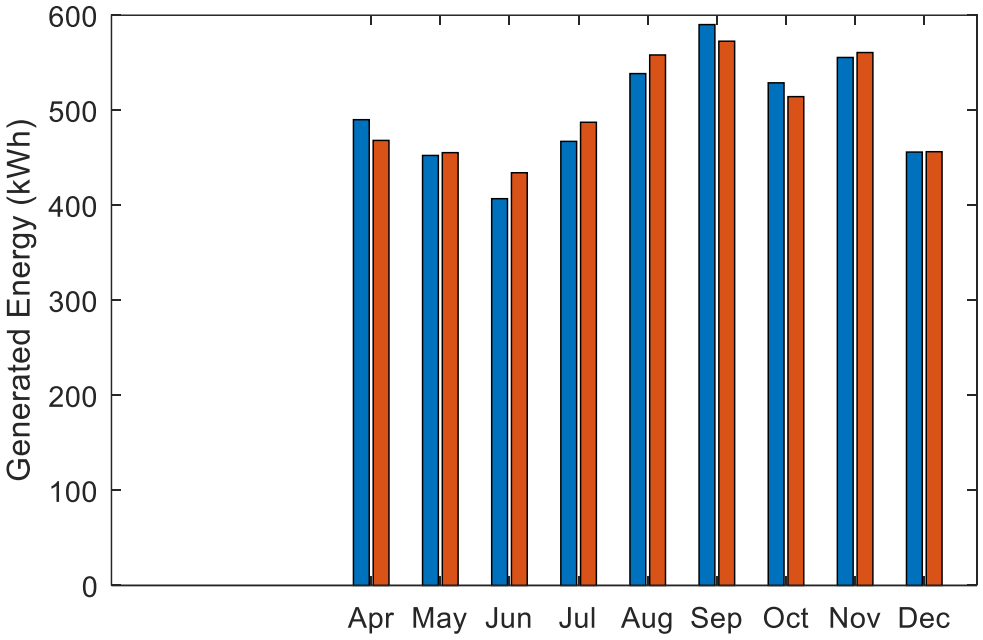

Figure 2.18. Monthly aggregated energy yield of PV system 3 from PVOutput webpage. Months with huge misleading or zero outputs were omitted.

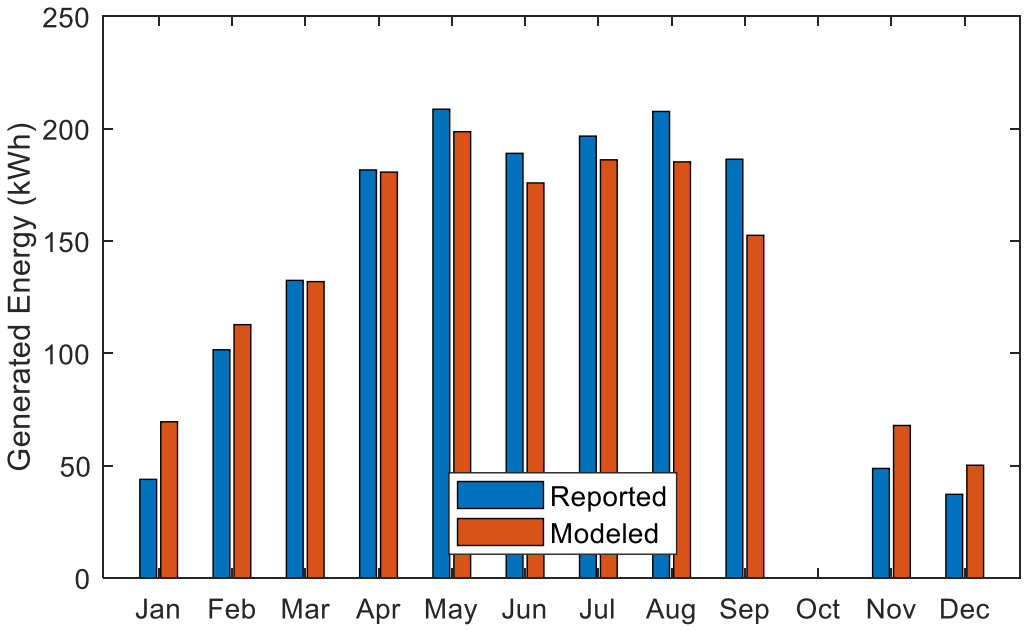

Figure 2.19. Monthly aggregated energy yield of PV system 4 from PVOutput webpage. Months with huge misleading or zero outputs were omitted.

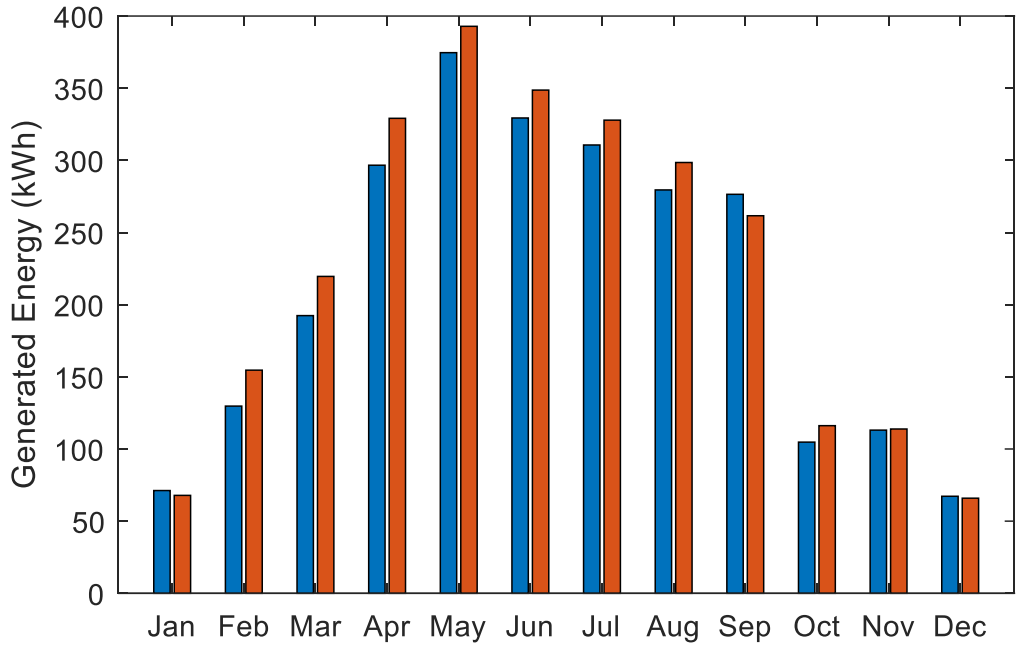

Figure 2.20. Monthly aggregated energy yield of PV system 5 from PVOutput webpage.

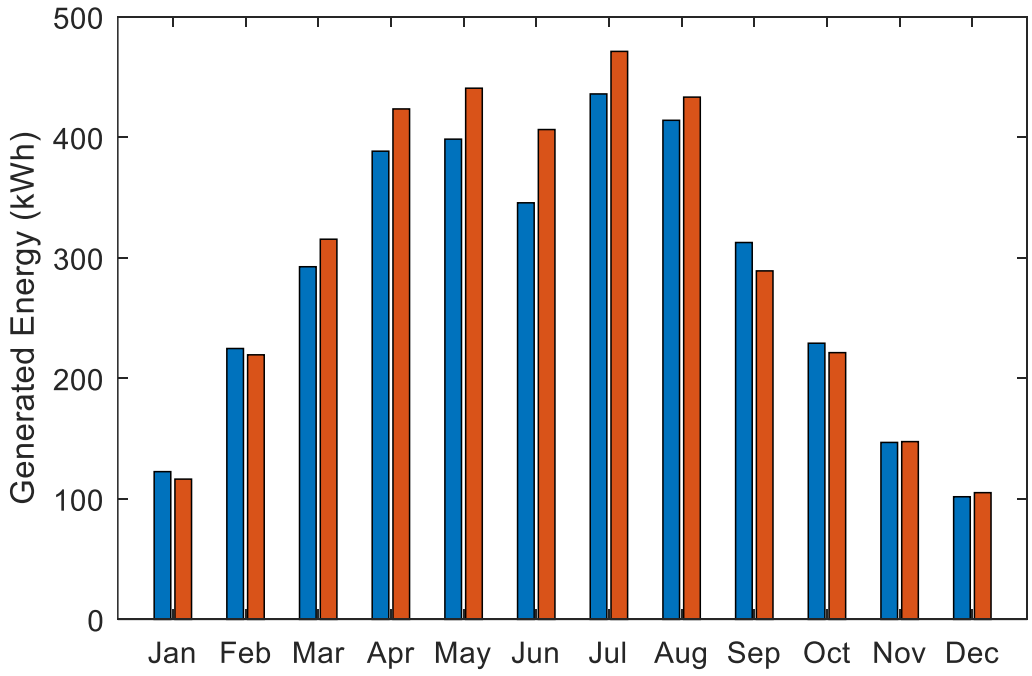

Figure 2.21. Monthly aggregated energy yield of PV system 6 from PVOutput webpage.

Basically, the results of the first three PV systems showed better match than the other half. The first reason that comes to mind for this mismatch is the simulation model since the PVGIS modeling tool is simply an approximation of the characteristics of PV panels at various irradiance levels.

2.5.3 SolrenView

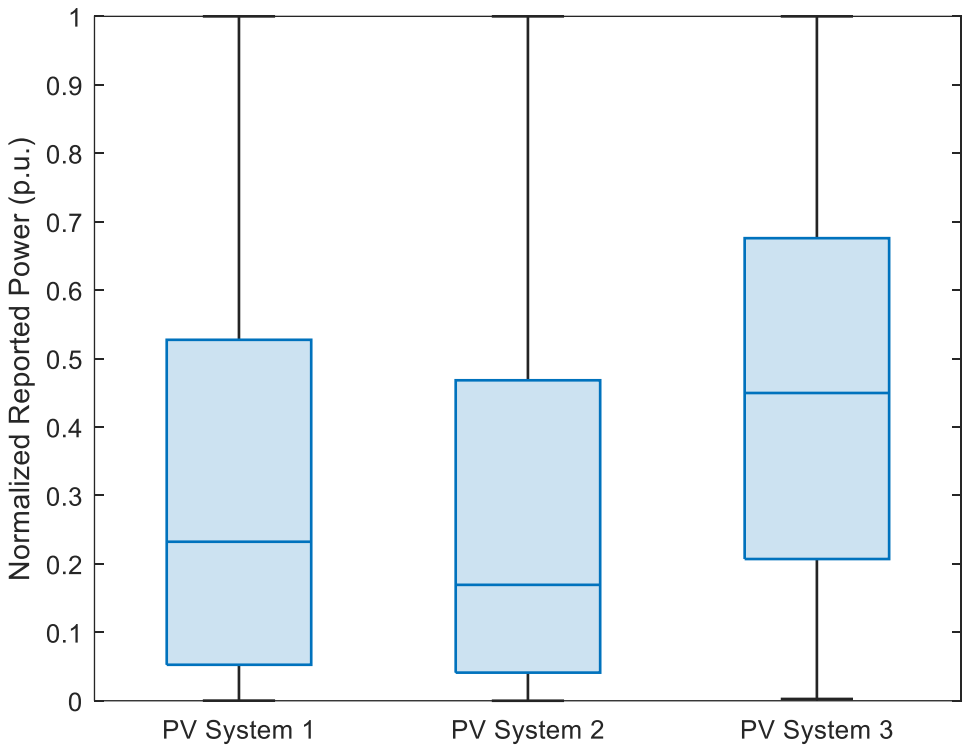

Figure 2.22. Box plots of the data of the three PV systems from SolrenView.

Adequate description of each PV system metadata wasn't reported as is the case in PVOutput and this creates more uncertainties in modelling the expected power output from measured irradiance with sufficient accuracy. The data provision requirement for grid integration or for systems (especially BTM) to receive feed-in tariff in many utility systems across the globe, includes only the address and nominal power. A primary input parameter for modelling is the array orientation and since this data wasn't provided, another method based on informed speculation although considered less effective was employed to obtain the tilt and azimuth angles. By visual inspection of the images reported by PV system owners and validating them with the

images from Google Earth of the reported addresses as shown in Figures 2.23 – 2.28, a range of values (a bracket) within which the angles were expected to fall was derived. The rationale was to obtain multiple simulations of the expected performance of each PV system using the orientation angles in the bracket and then using the best fitting results to identify the array orientation. Although there are no outliers in the box plot in Figure 2.22, a large percentage of power generation fell below 60% of the maximum generation.

The criteria for handpicking PV systems from this webpage were similar to the case for PVOutput. The only additional requirement was for the pictures of the systems to be reported by the owners and for images from Google Earth to show similar images or the presence of PV systems at the reported locations. Three PV systems were found to be across the board in meeting these criteria and were deemed sufficient as representative systems for this analysis. Their hourly output dataset in 2015 were extracted and compared with the simulated values using the metadata shown in Table 2.4.

Table 2.4. Reported profiles of selected PV systems from SolrenView.

PV system	1	2	3
Array size	121.275kW	12.2kW	49.704kW
Number of modules	385	40	152
Coordinates (Lat/Long)	41.637/-70.906	43.617/-72.96	32.851/-117.217
Module brand	SunPower SPR 315	Astroenergy 305	SunPower SPR 327
Inverter brand	Satcon PVS 100kW	4×Solectria PVI 6500	4×Solectria PVI 20TL
Expected tilt range (^{0})	0-5	30-45	5-15
Azimuth (0^0S)	South	South	south

A vast majority of registered PV systems on Solrenview including the selected systems had 100% reporting of hourly performance values round the clock for the spell of time they were active on the webpage. This would mean there were no periods of equipment failure however unprecedented it seems.

Figure 2.23. Pictures of site 1 reported by PV owners on SolrenView data monitoring webpage.

Figure 2.24. Aerial view of site 1 from Google Earth using the address reported on SolrenView.

Figure 2.25. Pictures of site 2 reported by PV owners on SolrenView data monitoring webpage.

Figure 2.26. Aerial view of site 2 from Google Earth using the address reported on SolrenView.

Figure 2.27. Pictures of site 3 reported by PV owners on SolrenView data monitoring webpage.

Figure 2.28. Aerial view of site 3 from Google Earth using the address reported on SolrenView.

2.5.4 Results and Analysis

Since irradiance data measured from weather stations at sufficiently close proximity to the PV systems weren't available, their outputs were estimated using the PVGIS Modeling Tool. The

modeled outputs were timestamped to UTC which was different from the local time at the locations of the PV systems. As such, there was need for time zone unification. From careful observations, the timestamping wasn't impacted by daylight saving, so the unification was simply shifting either of the recorded time-series several time steps back depending on the time difference.

Table 2.5. Best results obtained for the PV systems and the corresponding tilt angles for the study period. Range of values indicates that the lowest NRMSD and NMAD values do not coincide on the same tilt angle.

PV system	**1**	**2**	**3**
NMBD (%)	3.92	0.70	0.15
NMAD (%)	6.97	3.2	3.16
NRMSD (%)	10.01	8.01	6.30
Expected tilt range (^{0})	0-5	30-45	2-15
Best tilt (^{0})	0-3	20	13-15

The pictures reported by PV owners in Figures 2.23 – 2.28 give informed guesses of the tilt angles while the Google Earth images shows the modules facing due south. Similar to the case in PVOutput, Figures 2.29 – 2.31 show that the best results either matched with or revolve around the expected information. Regardless, the error levels in all cases are significant and calls for investigation.

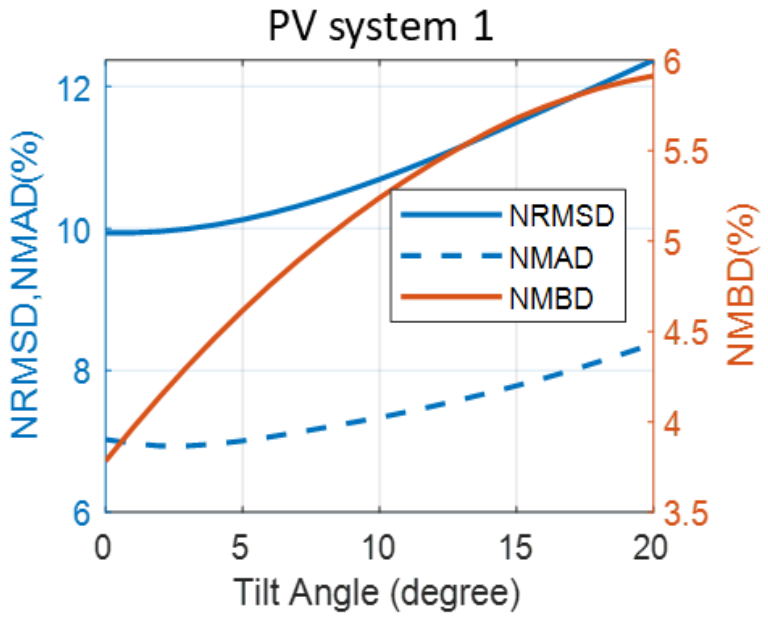

Figure 2.29. Variation of the error measures plotted against tilt angles of the PV array as input to the simulation model for PV systems registered on SolrenView webpage: PV system 1.

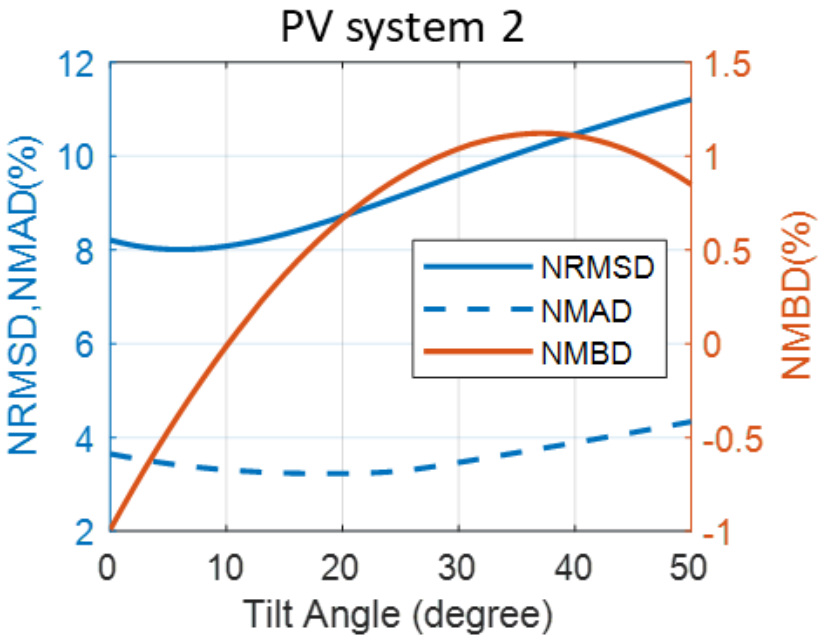

Figure 2.30. Variation of the error measures plotted against tilt angles of the PV array as input to the simulation model for PV systems registered on SolrenView webpage: PV system 2.

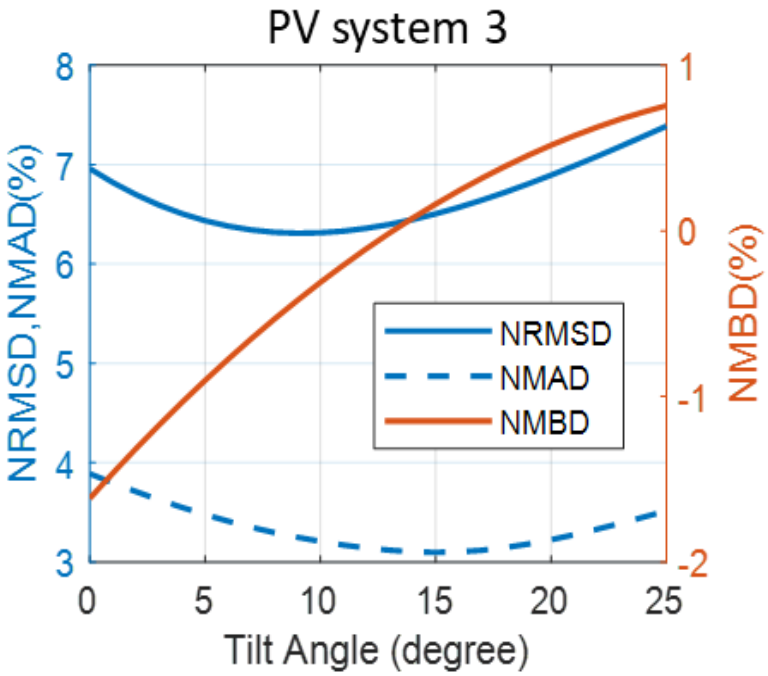

Figure 2.31. Variation of the error measures plotted against tilt angles of the PV array as input to the simulation model for PV systems registered on SolrenView webpage: PV system 3.

Figures 2.32 – 2.34 show positive biases in all the systems from SolrenView webpage with system 3 showing a better convergence. Although the first system is rooftop mounted, its array nominal capacity is well over its 100kW inverter which resulted in inverter clipping observed in the measured output. The PVGIS simulation model does not factor into consideration situations where the inverter is smaller than the array in capacity and consequently leading to output overestimation. PV system 1 was not deemed unfit for the analysis in order to see the effect of inverter clipping on the convergence of the outputs.

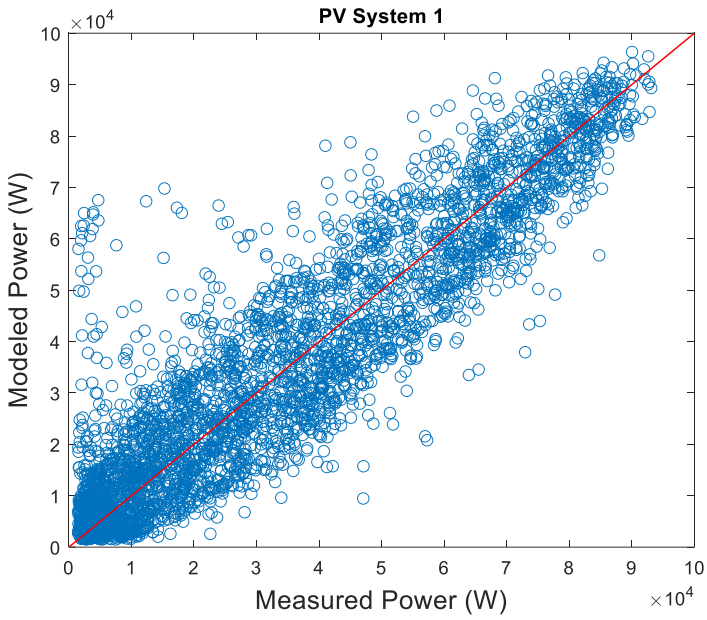

Figure 2.32. Simulated vs. measured or reported averaged hourly power output for SolrenView PV systems. These scatter plots reflect the results obtained for the best inclination angles shown in Table 2.5 for PV system 1.

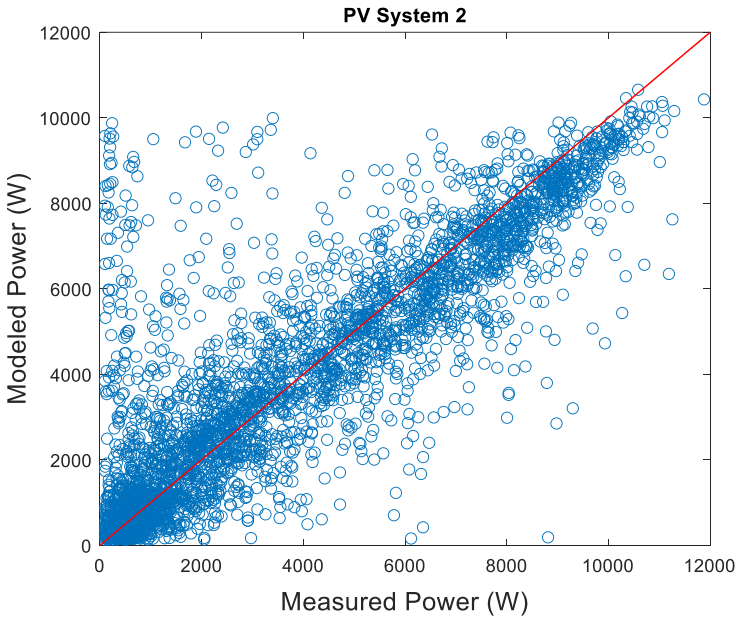

Figure 2.33. Simulated vs. measured or reported averaged hourly power output for SolrenView PV systems. These scatter plots reflect the results obtained for the best inclination angles shown in Table 2.5 for PV system 2.

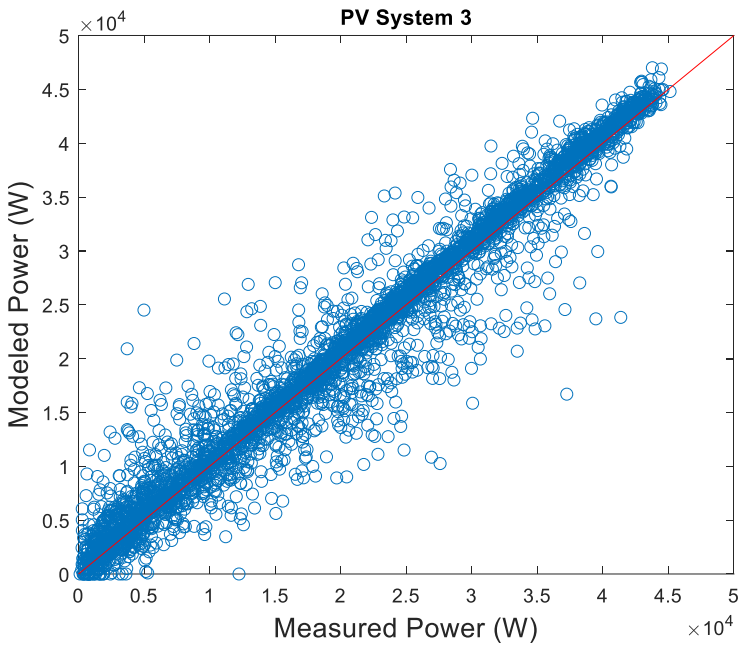

Figure 2.34. Simulated vs. measured or reported averaged hourly power output for SolrenView PV systems. These scatter plots reflect the results obtained for the best inclination angles shown in Table 2.5 for PV system 3.

The results displayed in Figure 2.35 – 2.40 support the narrative of better agreement between modeled and reported datasets with reduced temporal granularity. The GHI dataset for each location obtained from NREL's NSRDB datasets was modeled from satellite images using a physical model. On the other hand, comparing the systems with one another uncovers some information including the apparent case of system 1 showing less convergence possibly as a result of inverter clipping earlier discussed. Also, although all three systems are located in the US, there are two distinct energy generation profiles over the 12 months period howbeit the first two systems show a larger disagreement in February. Informed by reported array pictures, impact from shadows caused by nearby trees and high-rise buildings was ruled out as contributing to the obtained results.

Systems 1 and 2 are located in Fairhaven, Massachusetts and Rutland, Vermont respectively and generally have a similar climate as reported in annual US climate data by NOAA. Upon further investigations, there was a record-breaking blizzard across parts of the US for almost all through February and early March 2015. Cities like Fairhaven and Rutland witnessed prolonged heavy snowfalls and extreme winter storms which affected the PV generation as shown in the

hourly outputs in Figure 2.36, Figure 2.38, and Figure 2.40. There were cases of zero or near zero reported daily outputs and since this scenario was common to both systems, it was considered to be highly unlikely due to instrumentation failure. A solar panel fully occluded by snow cover can have no output and the climate of both cities show a possibility of as much as one inch snow cover on the ground from December to March for certain days per month.

The third system is located in San Diego, California and has a completely different climate compared to the other two systems. The results obtained for this system is corroborated by the fact that the lowest average daily sunshine hours in the city is just less than 8hours in December and there are no snow falls all year round.

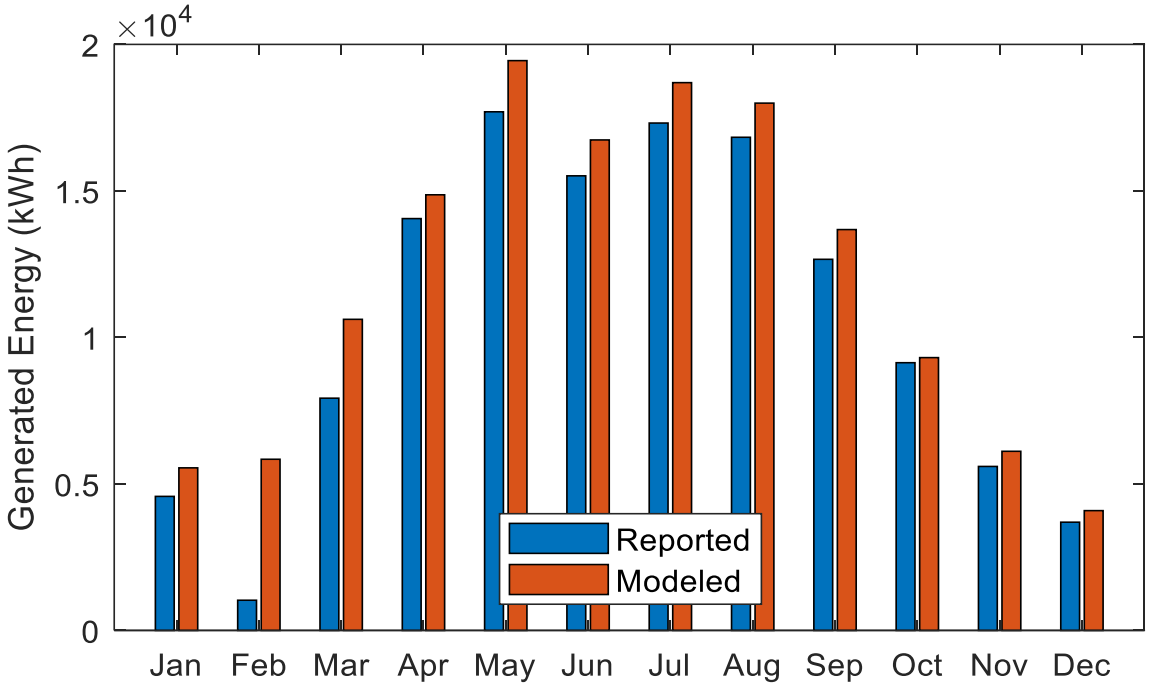

Figure 2.35. Monthly aggregated energy yield of PV system 1 from SolrenView webpage.

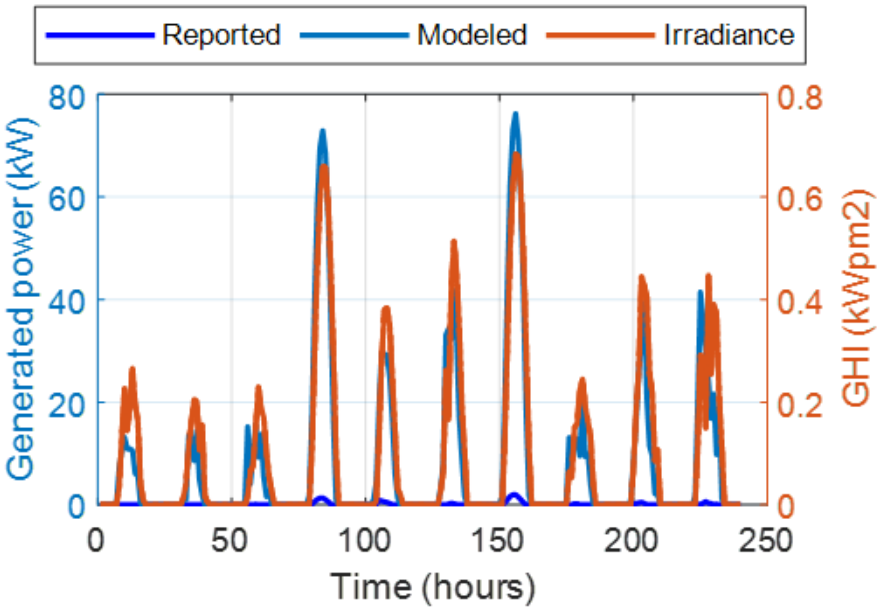

Figure 2.36. Hourly PV power generation and Global Horizontal Irradiance received from 9th to 18th February 2015 for PV System 1 from SolrenView.

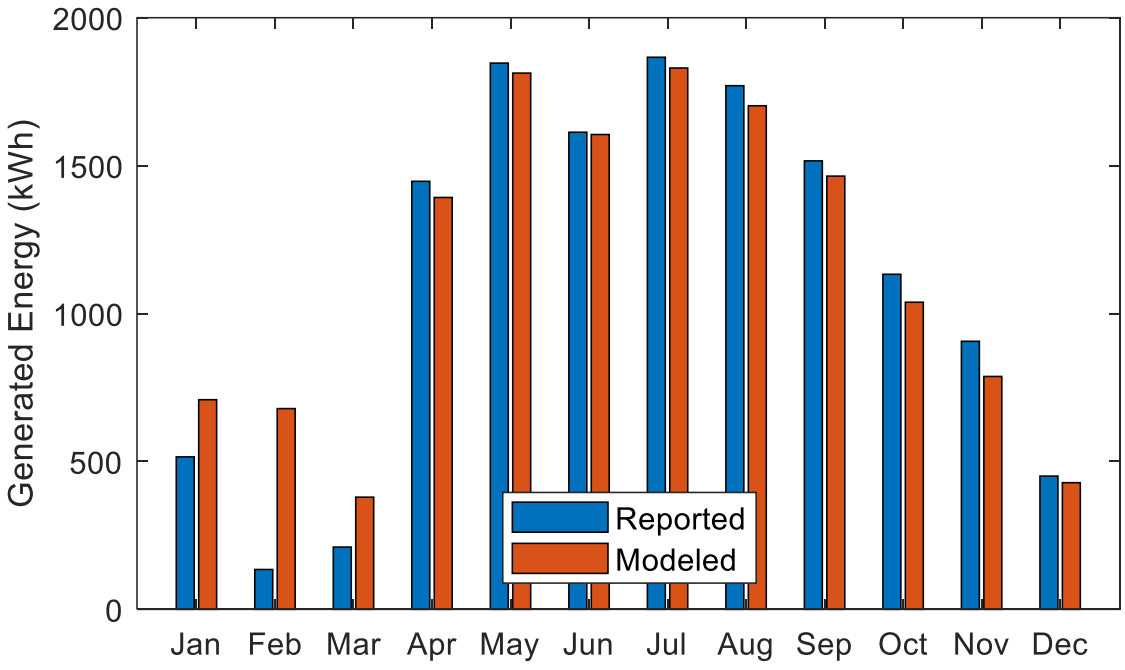

Figure 2.37. Monthly aggregated energy yield of PV system 2 from SolrenView webpage.

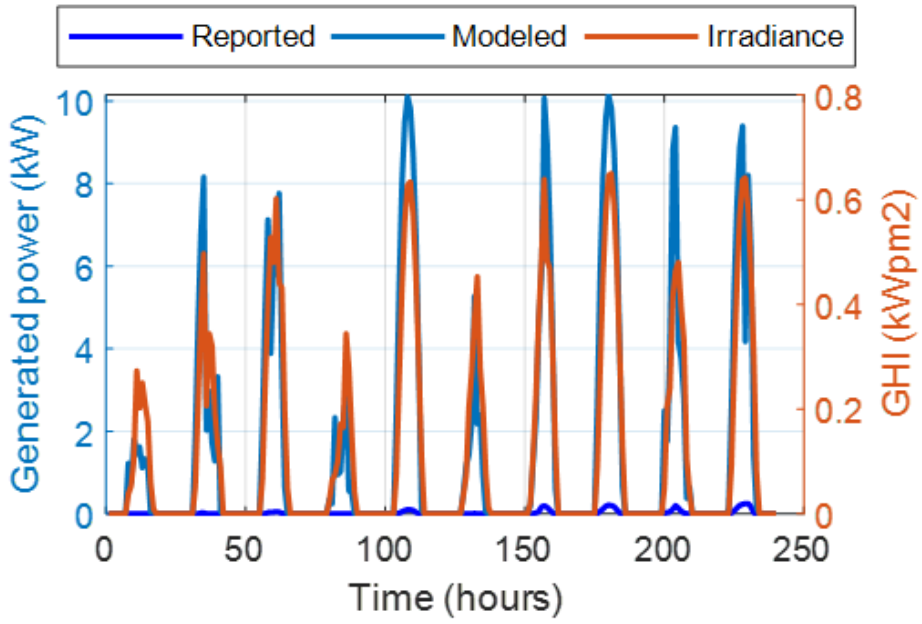

Figure 2.38. Hourly PV power generation and Global Horizontal Irradiance received from 9th to 18th February 2015 for PV System 2 from SolrenView.

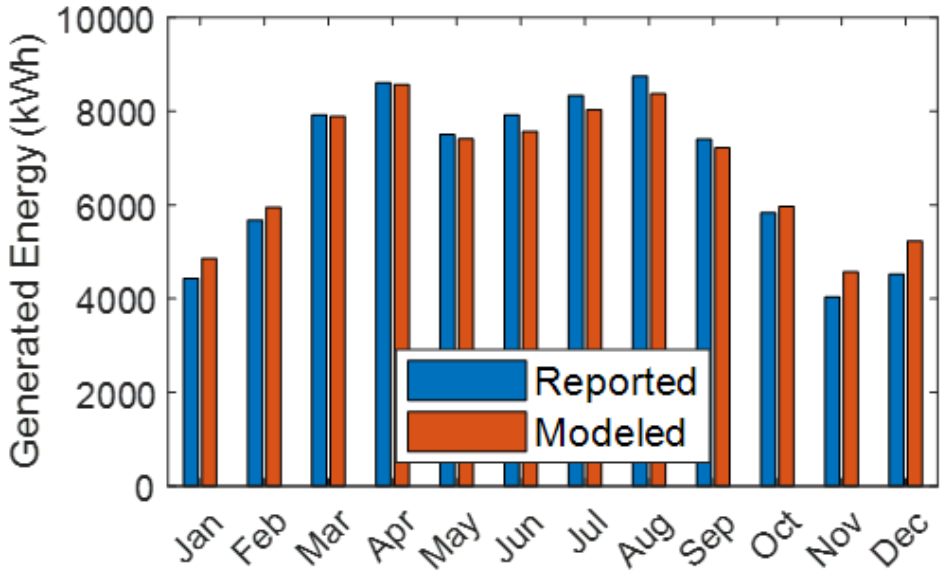

Figure 2.39. Monthly aggregated energy yield of PV system 3 from SolrenView webpage.

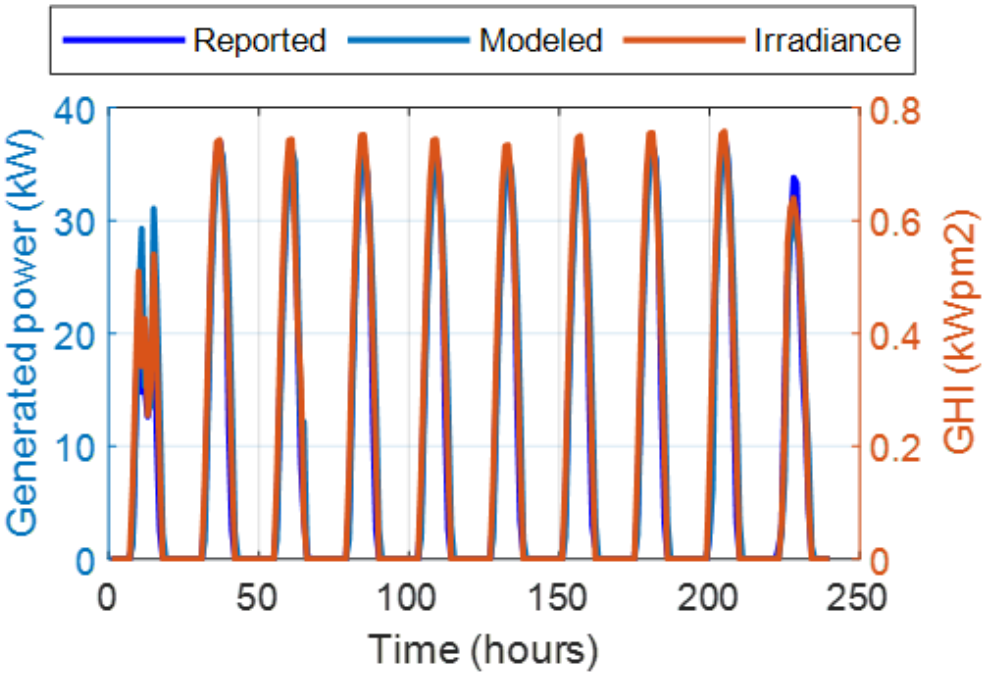

Figure 2.40. Hourly PV power generation and Global Horizontal Irradiance received from 9th to 18th February 2015 for PV System 3 from SolrenView.

2.6 Discussion

A discrete step size of one degree for the tilt angle does not adversely impact the error metrics since they are statistical averages of a large number of data points. In other words, depending on the system nominal capacity, the hourly generation data is not much different when simulated using a tilt angle differing by one degree per se. Consequently, the methodology applied wasn't focused on detecting the actual inclination and orientation angles with perfect accuracy but on revealing the possibility of the reported data being the actual case through the misalignment between reported and simulated power outputs.

Nine independent rooftop PV systems were analyzed. Divided equally into three groups, one group is located in Australia, the other two are located in the US and Europe respectively. These differing locations were preferred to highlight their unique climates expressed in total monthly sun hours received for the periods under study depicted in Figures 2.41 – 2.42. A stark difference in climate is seen between Australia and the other locations and there is a good match between the monthly PV outputs and the sun hours received. The statistical error metrics show considerable levels of errors. However, almost in every PV system considered, the lowest errors occurred at

array positioning angles coinciding with the reported information granting that an accuracy of not up to one degree is acceptable. The scatter plots also showed a good following between the data points with the monthly averages demonstrating better results although this doesn't reflect the large differences in the hourly data.

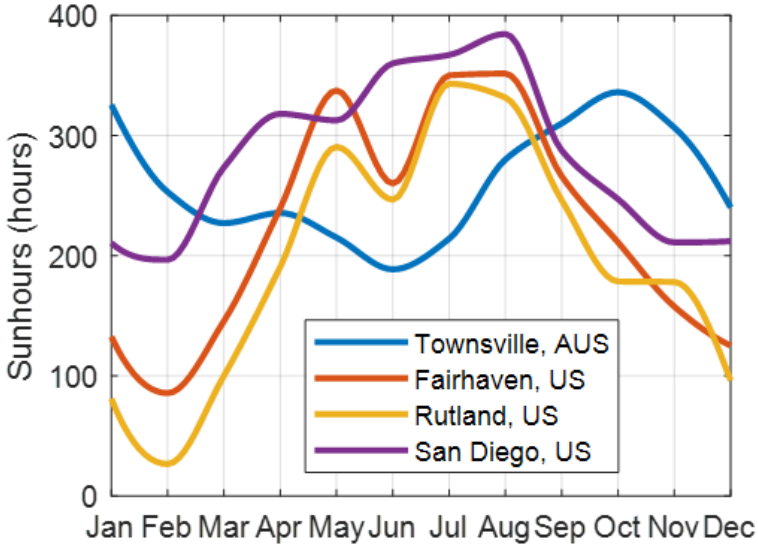

Figure 2.41. Total monthly sun hours for a calendar year in the seven locations of the analyzed PV systems [16]. 2018 calendar year for Australia (AUS) and 2015 for the US.

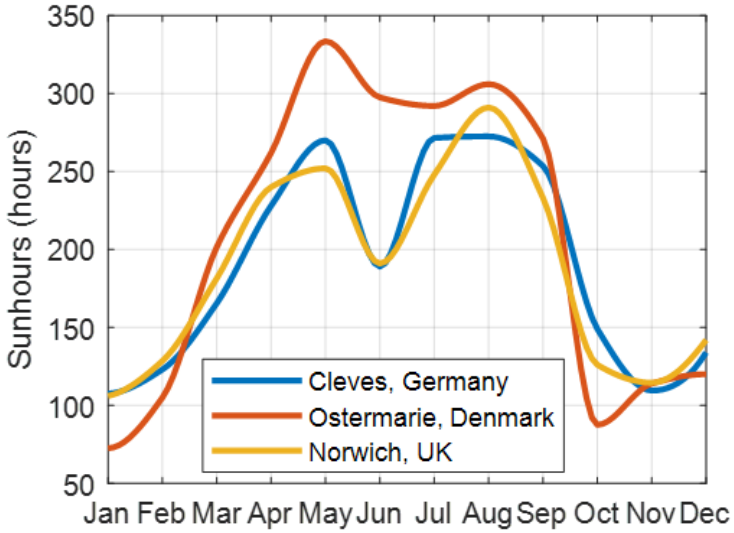

Figure 2.42. Total monthly sun hours for 2016 calendar year in three locations of the analyzed PV systems [16].

In contrast, the errors in the results are significant and are attributable to several sources including measuring equipment failure or malfunctioning. The ground based pyranometers showed erratic behaviours evidenced by missing or physically impossible measurements. Irradiance values had to be reconstructed to fix this problem which is also common for solar radiation data from satellite images. Furthermore, some PV systems have multiple inverters and as such, the failure of one becomes less detectable from the reported data when other inverters are in good operation.

The spread and systematic deviations displayed in the scatter plots bring to light a feature of the simulation methods used. A major difference is seen in the estimations at different irradiance levels. The PVGIS model performance drops at lower irradiances and that can be partly explained by the fact that irradiance modelling using cloud position and movement from satellite images is more prone to errors for images captured during periods of low sun elevation. Additionally, both PV output simulation models in this study are physical models that in simplifying the modelling process, make assumptions about certain parameters and physical processes which are not universally valid. As an example, a 14% system loss was assumed in the PVGIS model without a thorough analysis and the surface albedo for the transposition model was assumed to be constant at 0.2.

Temporal averaging could have added to the mismatch between reported and modeled data. The ground-based sensors had a 1-min time resolution while PVOutput recorded instantaneous power output every 5-mins and both were averaged to hourly resolution. Other dynamic error sources that could not be categorically accounted for are solar module degradation, shading from dusts and dirt and localized issues.

2.7 Conclusion

Regardless of the errors, the output data from distributed solar systems reported on public webpages is a reliable source and sufficiently paints a picture of the real-world performance under different conditions. In other words, the web-based monitoring and reporting concept is valid and feasible based on the results. To ensure a better data quality for high profile usages, data filtering may be recommended. Filtering spurious or errant data could be done by comparing neighboring systems whereas identifying physically impossible irradiance values by solar angle of incidence

or solar zenith angle-based filtering. A data quality control method with tremendous advantages and widespread acceptability involves normalizing the dataset with the clear-sky output simulated using the system parameters.

Contrasting the two case study webpages, it is needful to point out that there is need for upgrading the webpages. Although it is user-controlled, the installation of data logging equipment should be standardized or managed by utilities or licensed system installers. Adequate PV metadata description including the module orientation and inclination on the monitoring webpages is essential especially for research. Easy historical data access and acquisition should be ensured by impeccable cataloging. Hourly time resolution reduces the data storage requirement and reduced frequency of data logging could possibly minimize device failure.

The relevance of the proposed approach appreciates with growing investments in PV systems. Commercial scale PV systems can also use the proposed platform. To minimize social resistance to implementing the approach, we recommend that utility operators initiate a Call for Evidence to in order for PV power stakeholders and other interested persons or organizations to provide their submissions on the proposed approach.

Conclusions drawn may be questioned or considered misleading since the handpicked sites are only a minute fraction of the larger set that met the selection criteria and therefore not truly representative. Also, since every PV owner is acting independently of one another then generalizing the findings from a few systems to describe the whole lot as well as the selection process not being random could prove counterproductive at best. For the research scope and finitude, only a small number of systems could be analyzed. It is also a known fact that neighboring PV systems tend to have similar performances and since a consistency was found across the representative systems in this study, the conclusions reached are deemed valid.

2.8 References

[1] Solar Power Europe, “Global Market Outlook: 2019-2023,” *Glob. Mark. Outlook*, p. 92, 2019.

[2] “California Distributed; Generation Statistics.” [Online]. Available: https://www.californiadgstats.ca.gov/charts/.

[3] "No Title." [Online]. Available: https://leaderpost.com/news/saskatchewan/sask-government-to-rush-review-of-rooftop-solar-program-in-wake-of-outcry.%0A.

[4] X. Z. S Gould, A Blakers, J Domke, N Engerer, I Skryabin, F Mills, C Webers, B Williamson, "Short-term Machine-learning-based Forecasting of Distributed Solar Energy Production," *Intersolar*, no. July, 2013.

[5] H. Shaker, H. Zareipour, and D. Wood, "A data-driven approach for estimating the power generation of invisible solar sites," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2466–2476, 2016.

[6] "No Title." [Online]. Available: http://www.greentechmedia.com/articles/read/Turning-Rooftop-Solar-from-Invisible-Threat-to-Predictable-Resource.

[7] A. Golnas, J. Bryan, R. Wimbrow, C. Hansen, and S. Voss, "Performance assessment without pyranometers: Predicting energy output based on historical correlation," *Conf. Rec. IEEE Photovolt. Spec. Conf.*, pp. 002006–002010, 2011.

[8] SaskPower, "Generation Interconnection Requirements at Voltages 34 . 5 kV and Below," pp. 1–45, 2005.

[9] B. Dolter and M. J. Boucher, "'Let's Talk Solar' Final Report," 2017.

[10] S. Killinger, N. Engerer, and B. Müller, "QCPV: A quality control algorithm for distributed photovoltaic array power output," *Sol. Energy*, vol. 143, pp. 120–131, 2017.

[11] J. W. Spencer, "A comparison of methods for estimating hourly diffuse solar radiation from global solar radiation," *Sol. Energy*, vol. 29, no. 1, pp. 19–32, 1982.

[12] J. E. Hay, "Calculating solar radiation for inclined surfaces: Practical approaches," *Renew. Energy*, vol. 3, no. 4–5, pp. 373–380, 1993.

[13] "PV Performance Modeling Collaborative | An Industry and National Laboratory collaborative to improve Photovoltaic Performance Modeling." [Online]. Available: https://pvpmc.sandia.gov/. [Accessed: 02-Sep-2020].

[14] B. Marion, "Comparison of predictive models for photovoltaic module performance," *Conf. Rec. IEEE Photovolt. Spec. Conf.*, no. 1, 2008.

[15] T. Huld *et al.*, "A power-rating model for crystalline silicon PV modules," *Sol. Energy Mater. Sol. Cells*, vol. 95, no. 12, pp. 3359–3369, 2011.

[16] "World Weather Online." [Online]. Available: https://www.worldweatheronline.com/.

[17] A. Nespoli *et al.*, "Day-Ahead Photovoltaic Forecasting : A Comparison of the Most E ff

ective Techniques," pp. 1–15, 2019.

[18] J. Badosa, M. Haeffelin, N. Kalecinski, and G. Jumaux, "ScienceDirect Reliability of day-ahead solar irradiance forecasts on Reunion Island depending on synoptic wind and humidity conditions," vol. 115, pp. 306–321, 2015.

[19] P. E. Thornton, H. Hasenauer, and M. A. White, "Simultaneous estimation of daily solar radiation and humidity from observed temperature and precipitation : an application over complex terrain in Austria," vol. 104, pp. 255–271, 2000.

[20] K. U. N. Yang and T. Koike, "Estimating surface solar radiation from upper-air humidity," vol. 72, no. 2, pp. 177–186, 2002.

[21] R. Perez *et al.*, "PRELIMINARY EVALUATION OF AN APPROACH," no. September, 2004.

[22] R. Marquez and C. F. M. Coimbra, "Forecasting of global and direct solar irradiance using stochastic learning methods , ground experiments and the NWS database," *Sol. Energy*, vol. 85, no. 5, pp. 746–756, 2011.

[23] "PV education."

[24] J. Badosa and M. Haeffelin, "Scales of spatial and temporal variation of solar irradiance on Reunion tropical island," vol. 88, pp. 42–56, 2013.

[25] G. Hyndman, R.J., & Athanasopoulos, "Forecasting: principles and practice, 2nd edition," *OTexts: Melbourne, Australia*, 2018. [Online]. Available: OTexts.com/fpp2.

[26] 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," vol. 5, no. 2, pp. 597–606, 2015.

[27] A. Florita, B. Hodge, and K. Orwig, "Identifying Wind and Solar Ramping Events Preprint," no. January, 2013.

[28] Y. Qu, J. Xu, Y. Sun, C. Ching, S. Liao, and D. Ke, "Electrical Power and Energy Systems A parameter and resolution adaptive algorithm for rapid detection of ramp events in di ff erent timescale databases of the power system," *Electr. Power Energy Syst.*, vol. 112, no. May, pp. 393–403, 2019.

[29] D. M. Willy, T. L. Acker, D. Ph, E. Morgan, D. Ph, and R. K. Flood, "DEAD BAND METHOD FOR SOLAR IRRADIANCE AND POWER RAMP DETECTION ALGORITHMS," pp. 1–8.

[30] C. Hansen, J. Stein, and A. Ellis, “Statistical criteria for characterizing irradiance time series,” *Sandia Rep. Sand2010-7314*, no. October, pp. 1–51, 2010.

[31] L. Jinsong, Y. Dong, H. Yi, B. Xiaoxue, and G. An, “Application of Improved SDT Algorithm in CNC Machine Monitoring Field,” 2017.

[32] F. Zhang, S. Member, L. Cheng, X. Li, Y. Sun, and S. Member, “Application of a Real-Time Data Compression and Adapted Protocol Technique for WAMS,” pp. 1–10, 2014.

[33] M. Cui *et al.*, “An Optimized Swinging Door Algorithm for Identifying Wind Ramping Events,” *IEEE Trans. Sustain. Energy*, vol. 7, no. 1, pp. 150–162, 2016.

[34] M. Cui, J. Zhang, C. Feng, and A. R. Florita, “Characterizing and Analyzing Ramping Events in Wind Power , Solar Power , Load , and Netload.”

[35] M. Cui, J. Zhang, A. Florita, B. Hodge, D. Ke, and Y. Sun, “Solar Power Ramp Events Detection Using an Optimized Swinging Door Algorithm Preprint,” no. August, 2015.

3 DAY-AHEAD PV POWER PREDICTION SYSTEM FOR REGIONAL LEVEL FORECAST

3.1 Introduction

Electricity is an essential commodity to the world's GDP and as such the effective management of this precious item from generation to consumption is at the forefront of energy discourses. This brings on the concept of energy markets and the series of developments that have led to changes in the framework of how the market is managed. Innovations have exposed the inefficiency of a centrally controlled monopolized scheme necessitating the liberalization of the electricity market which gained global acceptance in the late nineties with California and some parts of the US adopting this market structure at the end of the century. Incorporating stochastic renewable energy power in electricity networks has certain impacts depending on its share in the generation mix. These stochastic means are sources of increased variability and unpredictability of the networks operation consequently impacting not only economic issues like the network operational cost and revenue [1],[2], but also the electricity market paradigms involving how biddings are tendered and accepted on the trading floors.

Wholesale electricity transactions or trading are done in different market pools – day-ahead, intra-day and balancing markets depending on the lead times when offers are made. Stakeholders in the forward or day-ahead market have to offer or accept offers at least 24hrs to the real time. The presence of renewables strongly impacts the market dynamics including the clearing price since they are categorized as potentially a free energy resource because of their associated zero fuel cost. As such, Independent System Operators (ISO) and renewable energy investors rely on day-ahead forecasts to function in this market floor. More accurate prediction of solar yields will tremendously improve decision-making and in turn result in increased integration of solar technologies in the grid [3].

Whereas predicting the output power is relatively easy for a single commercial-scale PV system, it is a challenging task for a large number of small-scale units spread within a regional utility system [4]–[6] due to the dynamism of installed capacity and performance inconsistencies that vary from site to site. Distributed energy resources, including rooftop PV systems, are currently increasing [7]. Energy yield from PV sources largely depends on two basic parameters: solar irradiance and outdoor

temperature (T) [8], [9]. Another important parameter is the efficiency of conversion from sunlight to electrical energy. Therefore, prediction models are concerned with evaluating solar resource levels and also the conversion rate of various PV technologies. The final step involves upscaling the model from a single or few systems to regional-scale capacity output. Three basic concepts considered in this study for regional scale PV power forecasting are:

1. Day-ahead irradiance forecasting model.
2. PV power simulation model.
3. Representative PV sites selection and PV power output upscaling model.

Figure 3.1 represents the modeling approach for a regional level PV power forecasting.

3.2 Day-Ahead Irradiance Forecasting Model for PV Power Prediction

There are various available techniques for forecasting PV power and several models exist to this end depending on the forecast horizon and its intended application. The forecasting process could be any of basically three methods viz. statistical, physical and a combination of both methods. Forecasting techniques like Artificial Neural Networks (ANNs), regression and persistence models that try to extract empirical relationships or information in observed data to predict future behavior fall into the family of statistical methods while physical methods use Cloud Motion Vectors (CMV) and mathematical expressions like partial differential equations to model the dynamics of the atmosphere [1]. Numerical Weather Prediction (NWP), Sky Imagery and Satellite Imaging are categorized as physical models. The rationale for using a hybrid of both methods is aimed at exploiting the unique strengths of each method. In combining both physical and statistical concepts to obtain an optimal forecast, the ratio of the blend is dependent on various determinants including the desired lead times, location globally and perhaps the forecaster's (forecast model designer) expertise level.

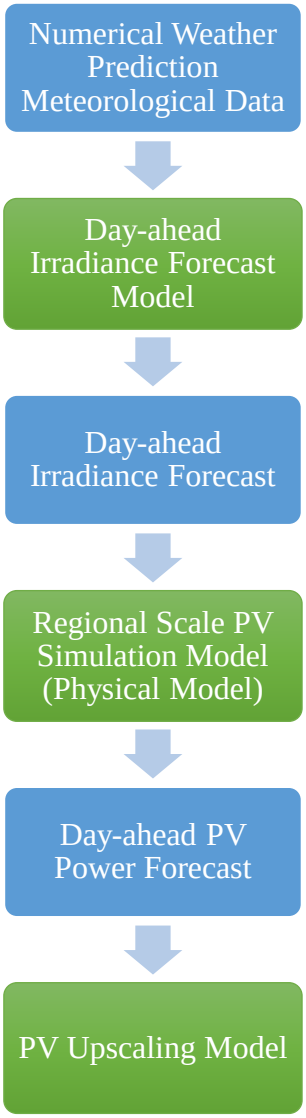

Figure 3.1. Overview of the proposed PV power prediction model

Very short-term forecasts which are usually provided within a couple of minutes to the real time are required by utility operators for balancing electricity supply with demand. Utility operators also make use of forecasts made available with a lead time of at least 24 hours with an hourly resolution for planning and scheduling operations of system infrastructure in the day-ahead market [10]. NWP of climatic and meteorological conditions at ground surface is significantly useful for this time range into the future because of the erratic and transient nature of clouds to deviate from previous situation. Irradiance observed during clear-sky days are less variable and

follow a diurnal pattern making it easier to predict than during overcast or cloudy periods where cloud movements modulate solar irradiance irregularly.

Weather variables are interdependent. Aerosols in the atmosphere contribute to irradiance scattering while gases like water vapor and ozone absorb irradiance [10],[11]. Surface irradiance attenuated and trapped in the atmosphere by the presence of these constituents in addition to irradiance reflection from ground surfaces could explain observed variable degrees of irradiance sensitivity to atmospheric temperature, relative humidity (RH), precipitation, wind speeds and other meteorological variables. Badosa et al. [12] concluded that daily formation of clouds are impacted by atmospheric humidity, local wind speed and direction. Consequently, cloud thickness and horizontal spread will vary between locations having different wind velocity and humidity profiles.

Another parameter to consider is the Solar Zenith Angle (θ_{SZA}) which describes in angular units the almost deterministic diurnal movement of the sun through the sky. At high SZAs, i.e. when the sun is closer to the horizon, sunlight has to navigate through a longer path through the atmosphere before reaching earth's surface resulting in more light being absorbed or scattered by aerosols and atmospheric gases. There is also an increased potential for obstruction of light by mountainous terrains during this period. Incorporating SZA in addition to temperature and precipitation as inputs to an irradiance forecast model will produce very good solar radiation estimates especially in locations with complex terrains and obstructed horizons [13].

Every forecasting effort is basically aimed at figuring out the sun's position and detailing the variability of atmospheric contents effect on the amount of sunlight reaching earth's surface. In other words, estimate the clear-sky irradiance (CSI) and then model how this irradiance is impacted by the presence of clouds. While simulating CSI is relatively easier, projecting into the future in excess of several hours cloud formation and dissipation to detail cloud cover distribution with good accuracy is a herculean task. Cloud cover is a non-prognostic variable hence we may have to rely on the relationships between other NWP outputs.

Based on correlations among weather variables [9], [14], Marquez et al. [15] developed a set of criteria to make inference on the most relevant predictors for solar irradiance forecasting. Gamma test and genetic algorithm were employed to obtain information on the relationships

between 11 meteorological and geo-temporal potential predictors and irradiance. The study arrived at a conclusion that sky (cloud) cover, (probability of) precipitation, (maximum and minimum) temperatures, and Cos(SZA) are overwhelmingly and critically important to enhance the forecasting accuracy. While RH, wind speed and direction were examined and not included in the best case list, the study did show that these variables could significantly improve forecast accuracy. Excluding sky cover from a list of commonly available weather data, Sangrody et al. [16] and Qing et al. [17] found that temperature and humidity had the strongest influence on accuracy of irradiance forecasts.

Simple models proposed in [18]–[20] trained irradiance forecast equations based on the empirical fit between sky cover data and the cloudless sky index. Yang et al. [21] developed a numerical model that used the upper air humidity as an input to estimate the hourly mean global horizontal irradiance (I_{GH}). The author proposed a method for estimating clear-sky index from RH profile in three atmospheric sublayers. The method simply involved simulating clear-sky radiation and then parameterizing cloud attenuation of this radiation using upper-air humidity classifications which represented different cloud types based on an assumption that a RH value above a threshold signified cloud existence. Humidity, temperature, and cloud amount data were fed into a fuzzy theory model for day-ahead insolation prediction in [22]; Si et al. [23] also included θ_{SZA}, whereas Moreno et al. [24] relied on only temperature and cloudiness forecast maps (cloud cover) as input data for a similar task. A common feature of these models is simulating clear-sky radiation and then parameterizing the attenuation of this radiation using different classifications of weather variables presumed to represent different cloud types.

The forecasted irradiances from these models show promising levels of correlation with ground-observed and satellite-derived benchmarks. Notwithstanding, the forecast performance status quo can be improved upon. In some cases, forecast models rely on only single predictor, such as cloud cover [19], which, if not reliably gathered, would severely impact the results. Yang's model in [21] depends on preestablished pressure and relative humidity (RH) constants, which are location sensitive and are not universally valid. Most of these models also require I_{GH} as a predictor and/or are applicable only to a single or small number of commercial-scale PV systems. Furthermore, based on our search, there isn't any preexisting literature on estimating day-ahead regional level distributed PV power without requiring solar irradiance as a primary input.

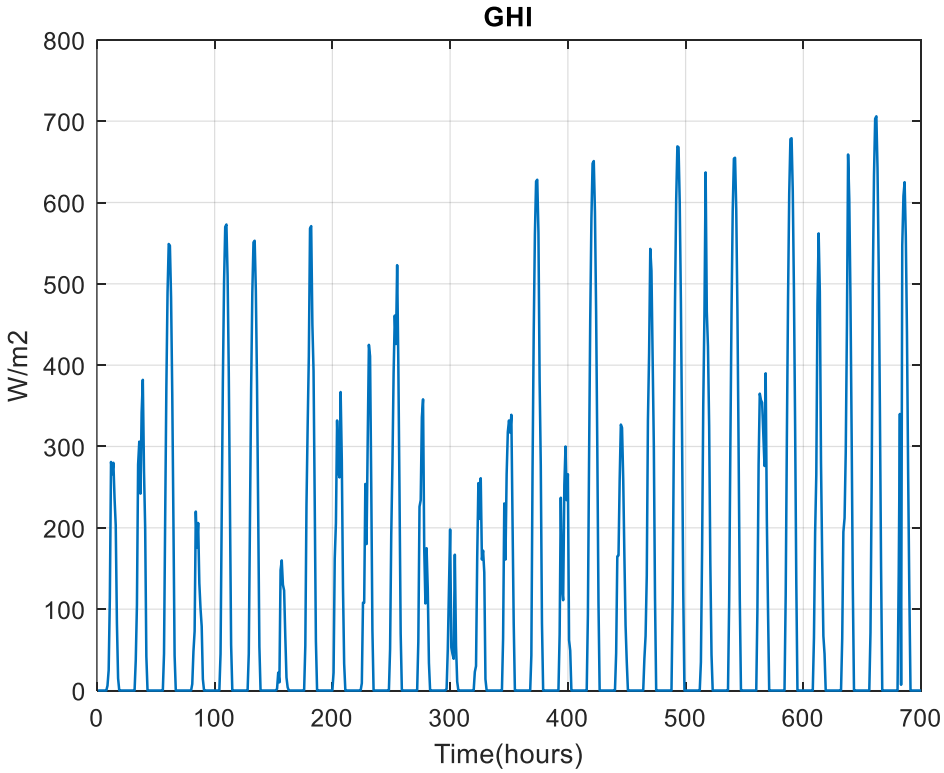

Figure 3.2. Time plot of the GHI variable.

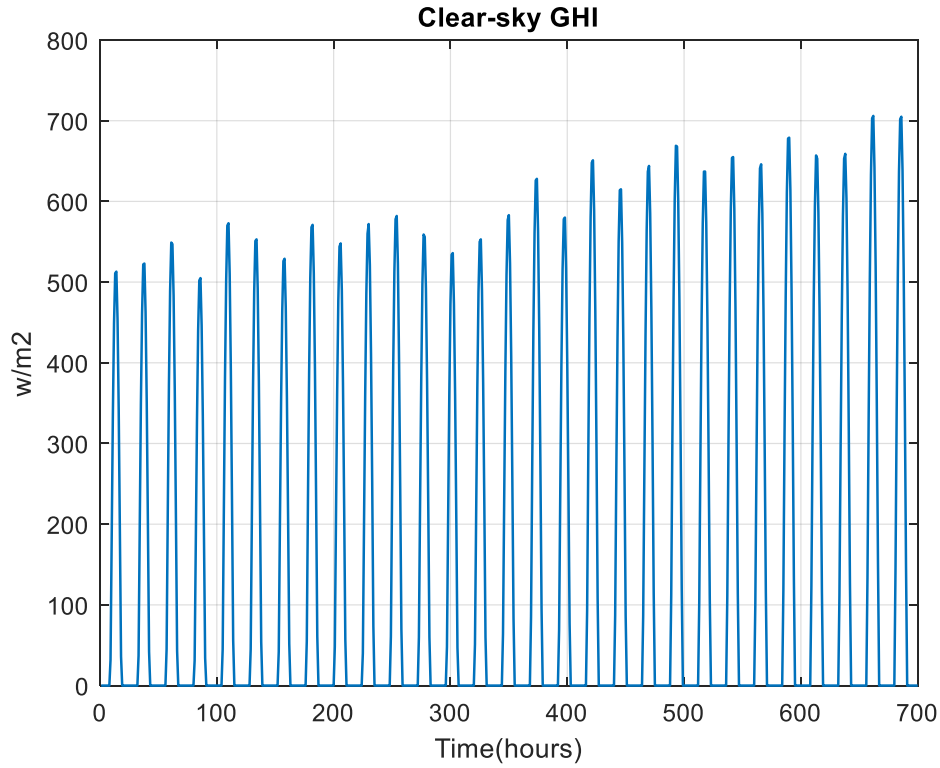

Figure 3.3. Time plot of the clear-sky GHI variable.

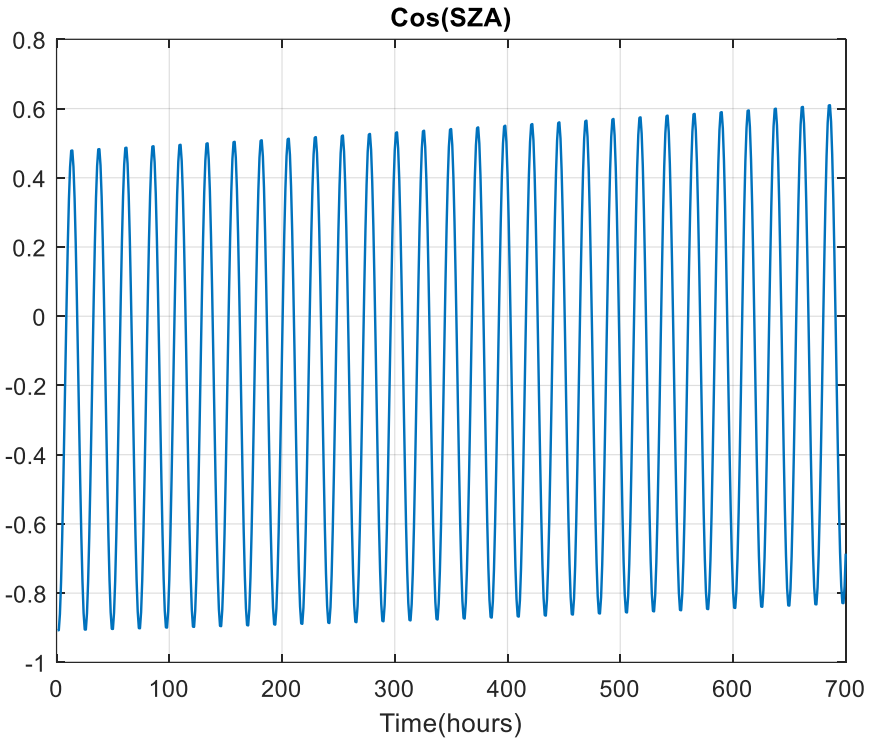

Figure 3.4. Time plot of the Cos(SZA) variable.

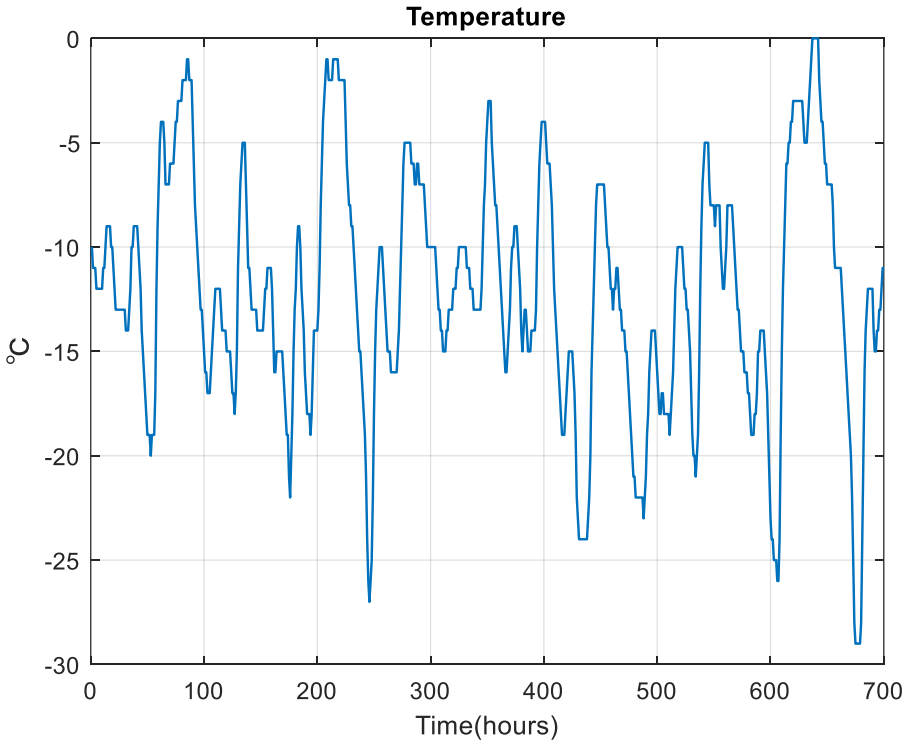

Figure 3.5. Time plot of the temperature variable.

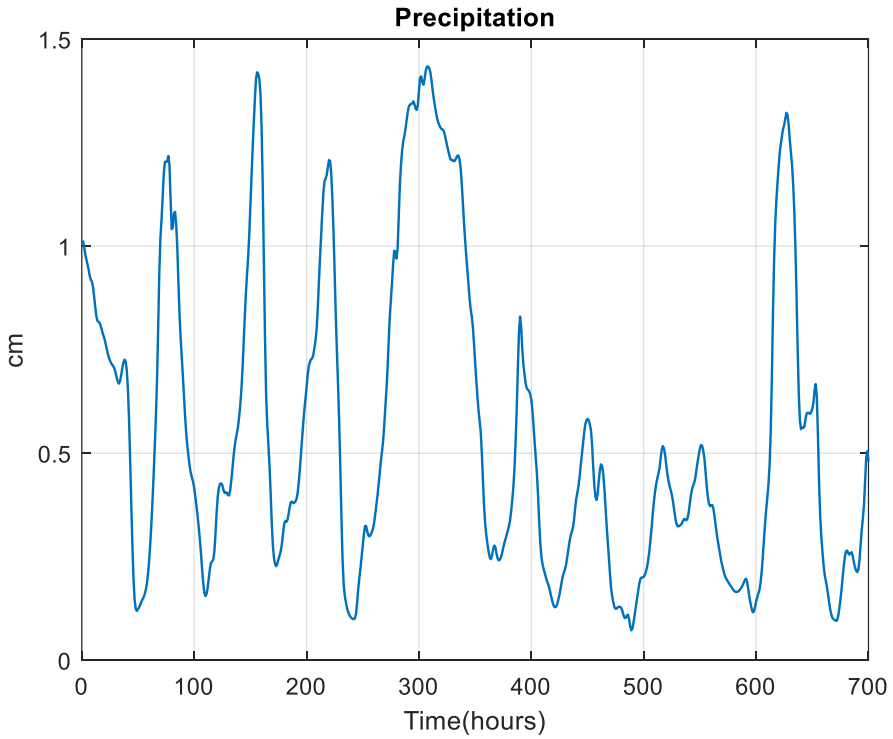

Figure 3.6. Time plot of the precipitation variable.

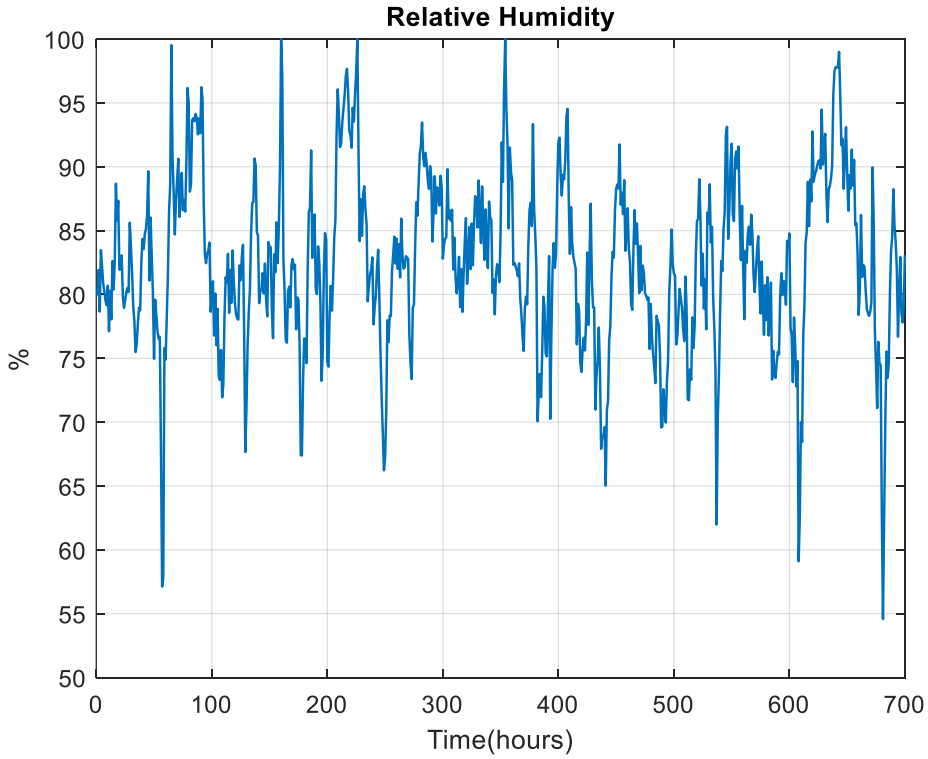

Figure 3.7.Time plot of the relative humidity variable.

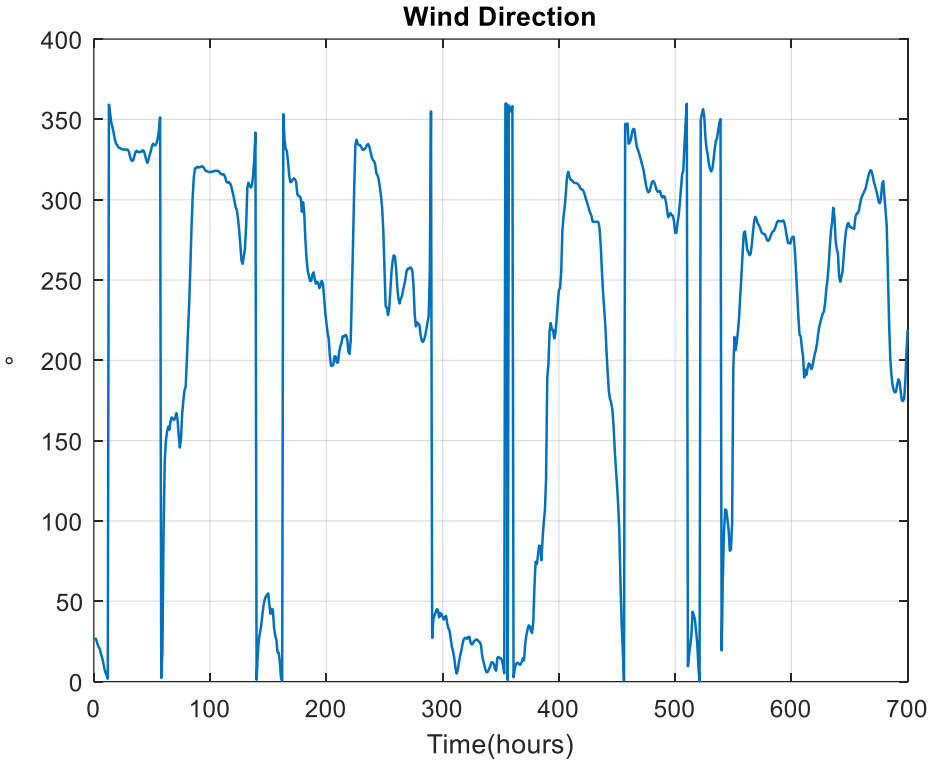

Figure 3.8. Time plot of the wind direction variable.

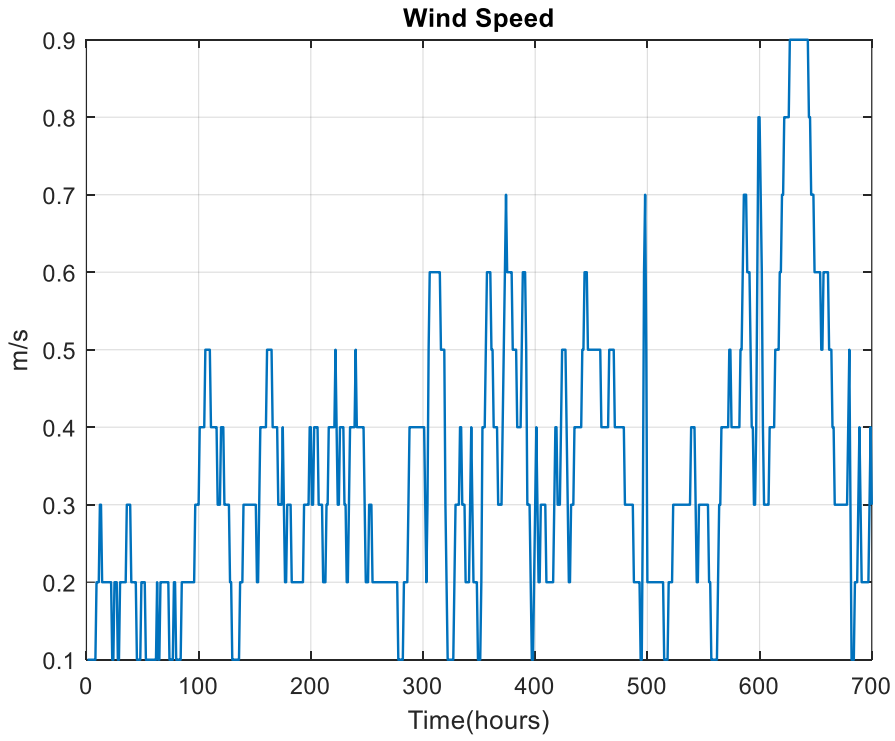

Figure 3.9. Time plot of the wind speed variable.

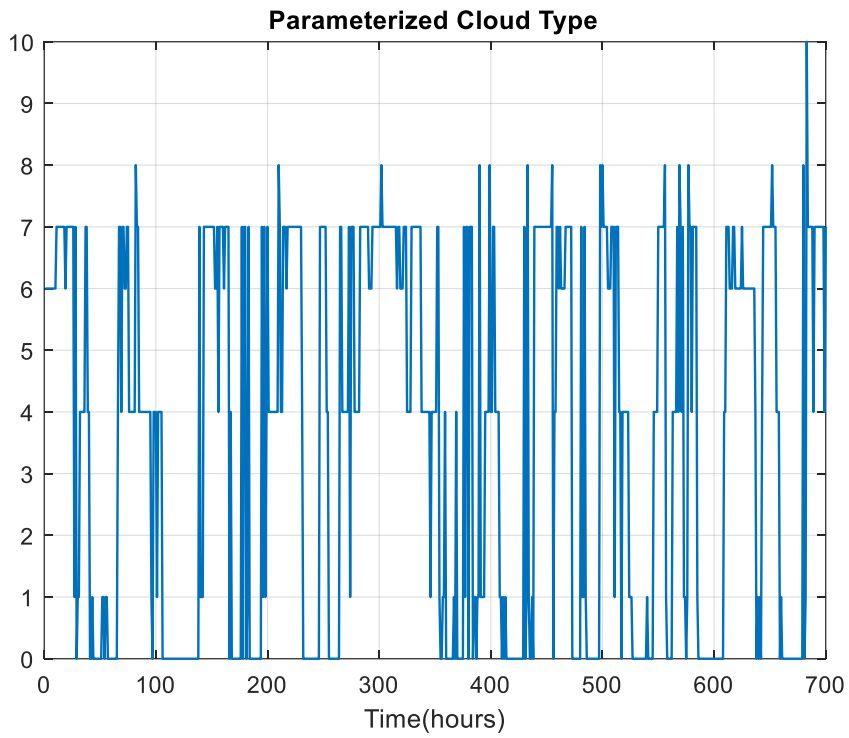

Figure 3.10. Time plot of the cloud type variable.

The data plotted in Figures 3.2 – 3.10 were obtained same period (first 30 days of January 2015) for each of the weather variables from National Renewable Energy Laboratory's (NREL) NSRDB datasets for a location with Latitude/Longitude 42.61/ -71.94.

These findings fueled our interests in exploring the possibility of developing irradiance forecast models featuring comparatively better accuracy consistent across test data from multiple sites. The model will also use multiple weather parameters as predictor variables to minimize the effect of single predictor errors on the results. In addition, global models of NWP in particular produce excessively coarse spatial and temporal resolution solar irradiance information that requires optimization in the form of bias correction to make it location-specific and sophisticated temporal interpolations for hourly applications [25]. For instance, the European Centre for Medium-Range Weather Forecasts (ECMWF) global model produces irradiance predictions with spatial and temporal resolutions of about 16 km and 3 hours, respectively [26].

Figures 3.2 – 3.4 involving Global Horizontal Irradiance (GHI), Clear-sky GHI and Cos (SZA) showed a strong daily seasonal pattern reminiscing the daily sun movement in the sky or rather the rotation of the earth about its axis while the others show a cyclic behaviour. The temperature plot in Figure 3.5 resembles a daily pattern of a steep climb and drop while precipitation showed an irregular cyclic pattern. Relative humidity mimics a white noise plot, windspeed has an increasing amplitude. Wind direction and cloud type do not show any notable or observable trend. In all, a finer resolution of the plots will uncover more slight changes and trends.

We use scatter plots to understand the relationships between the variables as shown in Figures 3.11 – 3.18. Data were obtained from the National Renewable Energy Laboratory's (NREL) NSRDB datasets [27] for a location with latitude/longitude 42.61/ -71.94. Clear-sky GHI is the clear-sky global horizontal irradiance. Clear and recognizable patterns can be observed between clear-sky irradiance, Cos(SZA), RH, temperature and GHI while the other 4 plots appear random except for the windspeed plot showing a weak negative relationship between windspeed and GHI. The Correlation coefficients measure the strength of only the linear relationships and therefore could be misleading because a weak linear correlation might mean a stronger non-linear relationship.

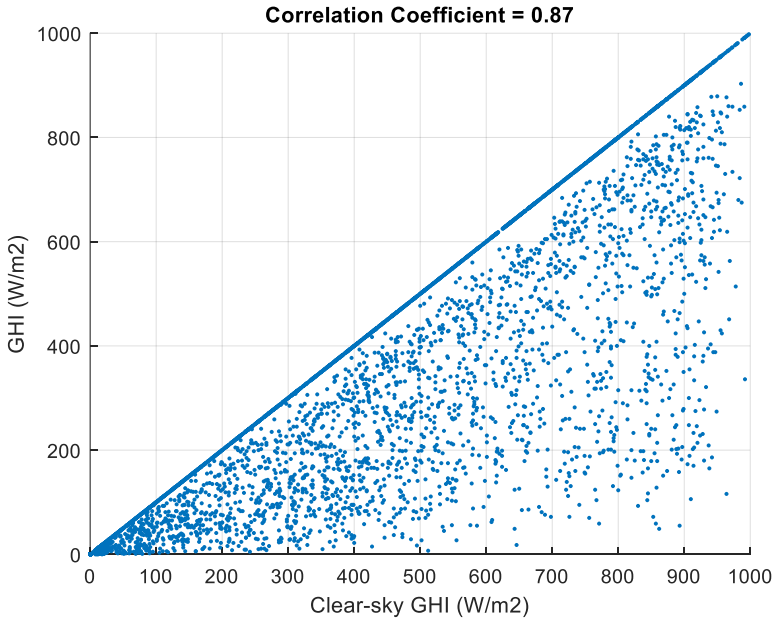

Figure 3.11. Solar irradiance plotted against clear-sky GHI variable with the corresponding correlation coefficient.

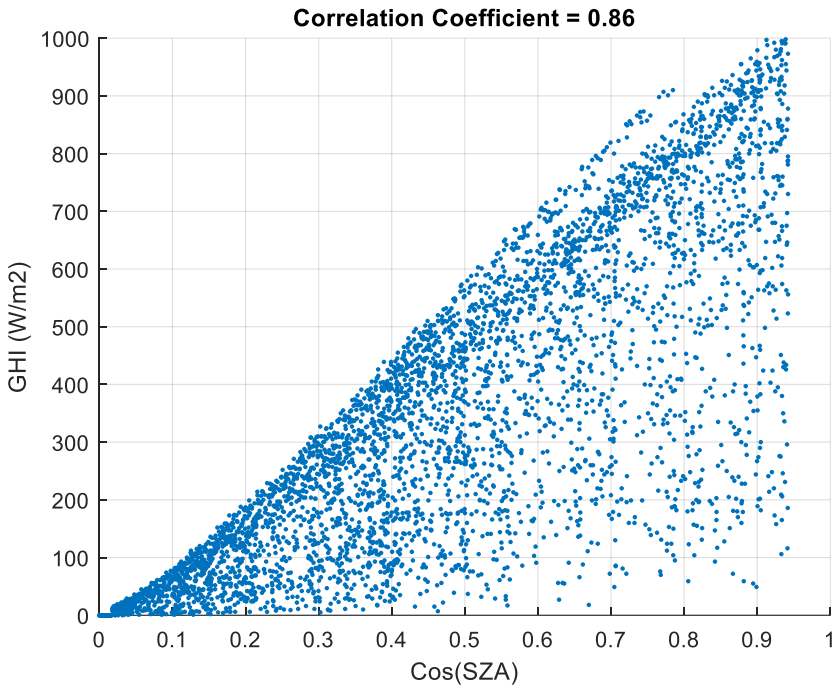

Figure 3.12. Solar irradiance plotted against Cos(SZA) variable with the corresponding correlation coefficient.

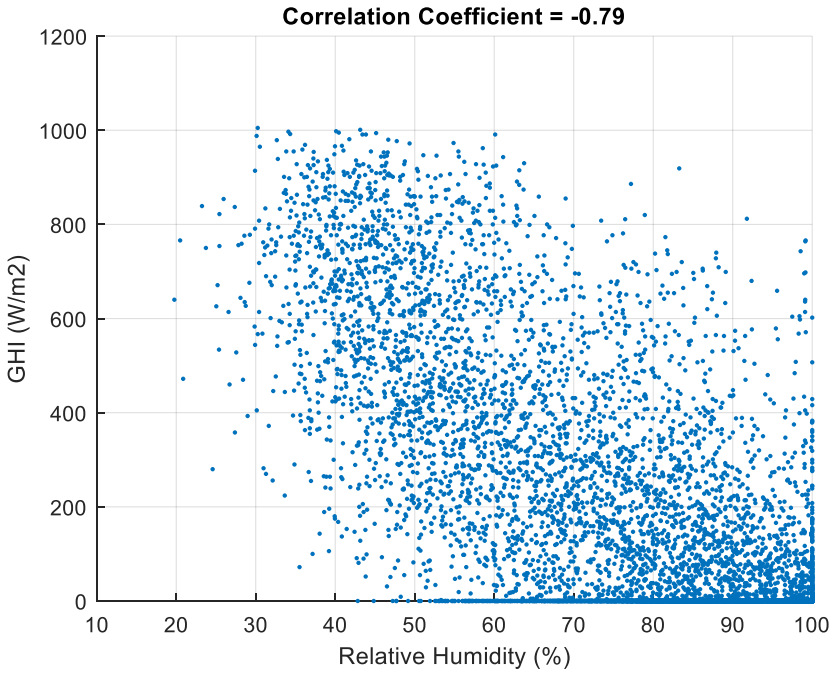

Figure 3.13. Solar irradiance plotted against relative humidity variable with the corresponding correlation coefficient.

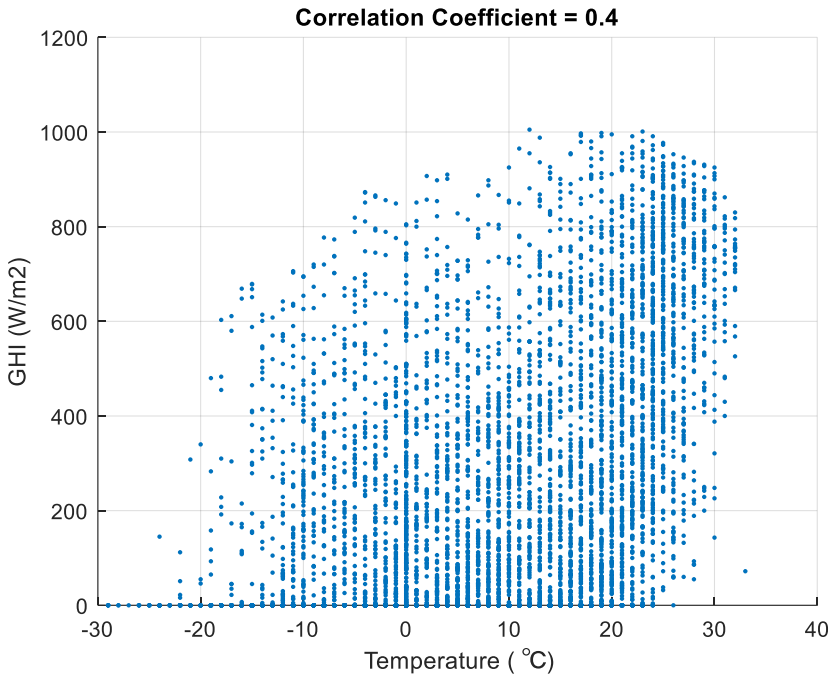

Figure 3.14. Solar irradiance plotted against temperature variable with the corresponding correlation coefficient.

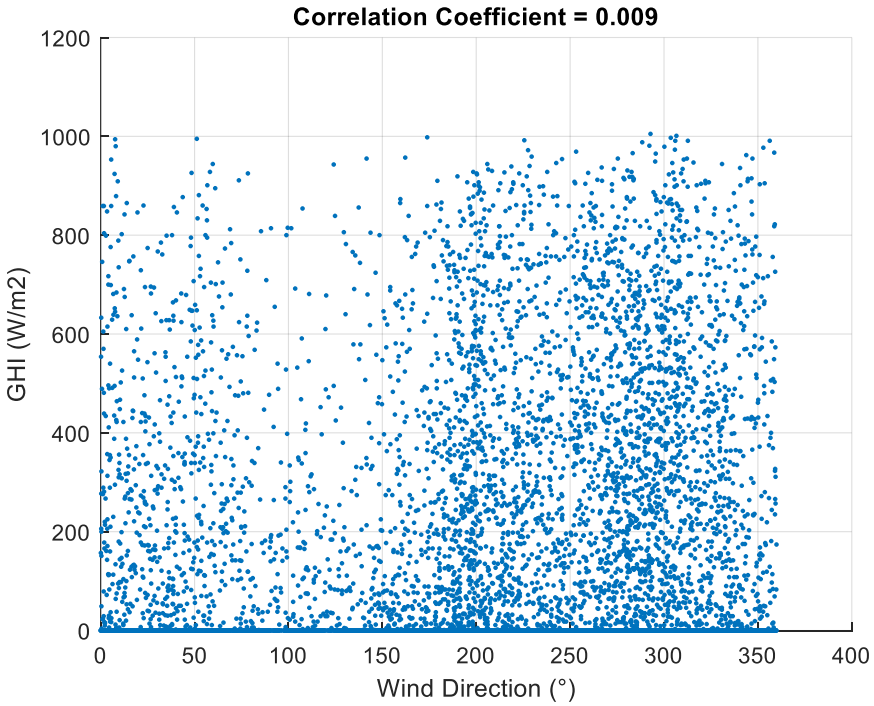

Figure 3.15. Solar irradiance plotted against wind direction variable with the corresponding correlation coefficient.

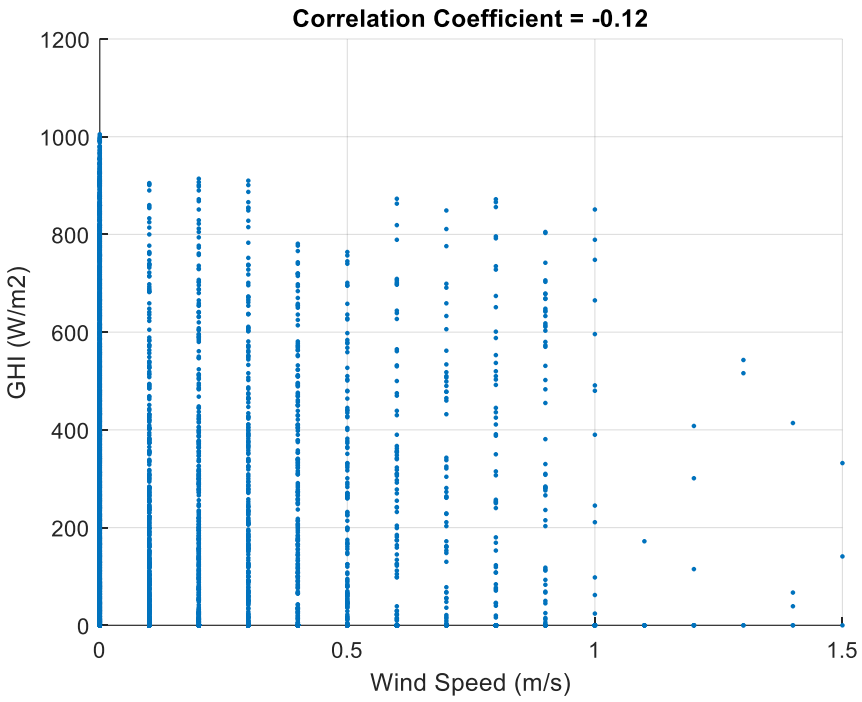

Figure 3.16. Solar irradiance plotted against wind speed variable with the corresponding correlation coefficient.

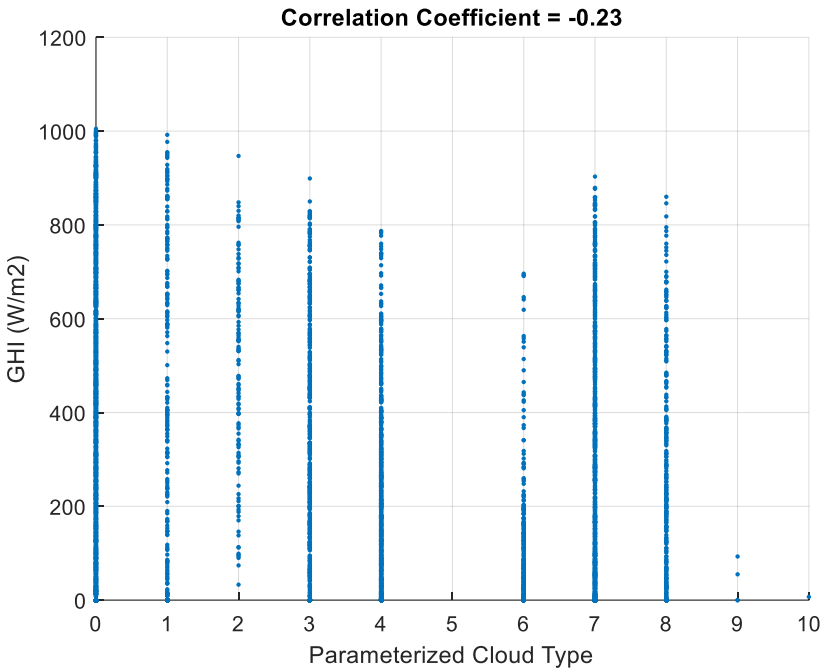

Figure 3.17. Solar irradiance plotted against parameterized cloud type variable with the corresponding correlation coefficient.

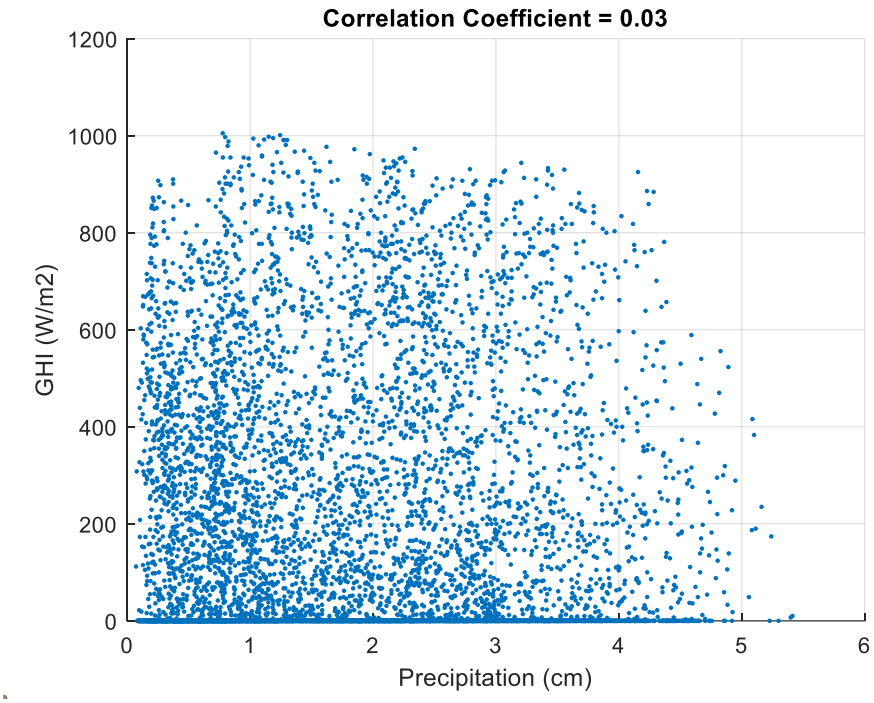

Figure 3.18. Solar irradiance plotted against precipitation variable with the corresponding correlation coefficient.

In this chapter, the main contributions include:

1. A model for estimating distributed regional-scale PV power generation based on weather variables from NWPs is proposed. The three basic concepts considered in this study are day-ahead irradiance forecasting, PV power simulation, and representative PV sites selection and power output upscaling. For day-ahead irradiance forecasts, A recalibrated clear-sky model is combined with an artificial neural network (ANN). Both physical and machine learning-based models are explored for PV power simulation. Data dimension reduction techniques including k-means clustering and principal component analysis (PCA) are employed for cluster creation and representative site selection. The regional output is estimated using linear upscaling expressions.
2. A basis for using autoregression of weather variables is provided to create input multiples for improved accuracy.
3. A perspective for redundant and relevant input weather variables for solar irradiance forecasts is provided.

Generally, deep neural networks are increasingly more popular and potentially superior to ANNs in terms of accuracy, relevance and applicability especially for forecasting applications. Most recent machine-learning-based forecast models show a transition from ANNs to deep neural networks [28]. Chen et al. [29] proposed a deep learning model for short-term predictions of windspeed. The proposed model consisted of a layer of three separate windspeed predictions using extreme learning machine (ELM), Elman neural network (ENN) and long short-term memory (LSTM) neural networks with the aim of exploiting their individual strengths. A second later of ELM was subsequently used for a non-linear combination of all three predictions for best results. Zhao et al. [30] also proposed using an ensemble of deep learning methods for vehicular traffic flow predictions instead. Similar to [29], a second layer of no negative constraint theory (NNCT) weight integration strategy was used to aggregate a cluster of LSTM-based predictions from the first stage into a final result.

However, the ANN is computationally simpler, requires less storage capacities and it is less time consuming than deep learning alternatives. The training data requirement for deep learning networks is huge [31] and not feasible for this study because of data availability constraints. Also, the performance of the proposed ANN-based model was comparatively sufficient and better given the prevailing realities of a small training data size and uncertain data quality.

3.2.1 Clear-Sky Irradiance Estimation

First, θ_{SZA} and I_{GHcsk} are calculated by simple empirical means. θ_{SZA} is the angle between the sun and the zenith and is related to the sun's elevation angle (α_e) by (1). There is no universally accepted or valid expression for estimating I_{GHcsk} due to sensitivity to local conditions and subjectivity to the training data. Several clear-sky models exist in the literature, ranging from basic models that require only the θ_{SZA} as input to complex models that are more difficult to implement due to unavailability of required input parameters. The choice of which model to use is found to be inconsequential, provided proper adjustment of the scaling is implemented to yield a sufficiently accurate representation of surface I_{GHcsk} [32].

With respect to the daily power generation profile, our initial trial of a randomly picked clear-sky model showed high magnitude forecast errors coinciding with midday when the sun is

expected to be at the zenith. On a seasonal basis, the bulk of the errors would occur between April and September when the expected irradiance and average number of daily sun hours are higher [33]. Consequently, this finding supported the notion that a good forecast system would require estimating the I_{GHcsk} component with good accuracy.

$$\theta_{SZA} = 90^0 - \alpha_e \quad (1)$$

$$\alpha_e = \sin^{-1}[\sin\sigma \sin\emptyset + \cos\sigma \cos\emptyset \cos(HRA)] \quad (2)$$

where:

σ_d is the declination angle,

$\emptyset$ is the latitude, and

HRA is the hour angle.

Further expressions to calculate σ and HRA can be obtained from [34]. The clear-sky model by [35] and that requires only θ_{SZA} as its input is adopted to calculate I_{GHcsk}. However, the model coefficients were recalibrated by regression analysis (ordinary lease squares) using data from the NREL NSRDB for the same period and location considered in this study. It is observed that a single expression cannot sufficiently model the relationship between θ_{SZA} and I_{GHcsk} at all irradiance levels, so we propose using (3) and (4) depending on a θ_{SZA} threshold obtained by optimization. This recalibration yields better results with error values that are below 3.5% of the reference data.

$$I_{GHcsk}(t) = 1100 \cdot \cos\big(\theta_{SZA}(t)\big)^{1.164} (w/m^2) \quad if\ \theta_{SZA} < 81^0 \quad (3)$$

$$I_{GHcsk}(t) = 1100 \cdot \cos\big(\theta_{SZA}(t)\big)^{1.25} (w/m^2) \quad if\ \theta_{SZA} \geq 81^0 \quad (4)$$

3.2.2 Artificial Neural Networks

To capture the complex and non-linear empirical relationships between meteorological variables and I_{GH} hourly values, a neural network is a preferred forecasting method [8], [36]. There are six variables as inputs, and ANNs show sufficiently better accuracy and adaptability than other statistical alternatives, such as the autoregressive moving average extrapolation (ARMAX) [37] and the autoregressive integrated moving average (ARIMA) models [38], especially for handling uncertain weather conditions. ANNs are considered one of the most effective statistical forecasting

methods and have widespread acceptability in energy forecasting discourses due to their ability to capture sudden changes in input-output relationships [39].

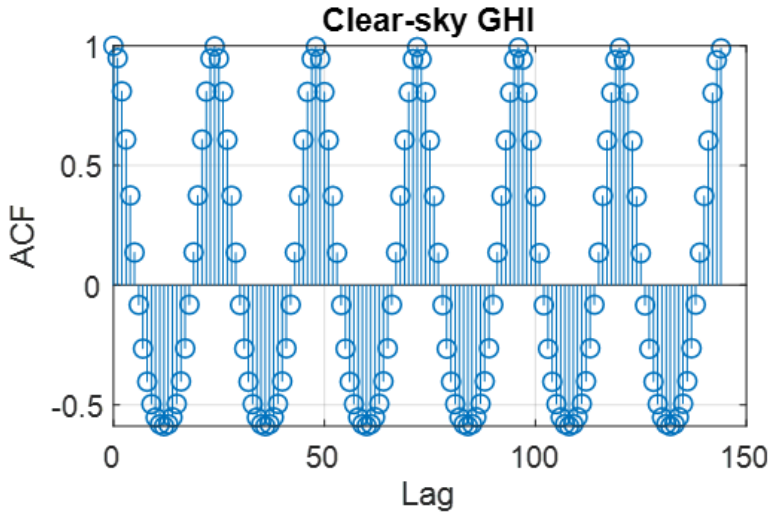

Figure 3.19. Autocorrelation function (ACF) of clear-sky GHI predictor. The peaks and troughs are 24 hours apart.

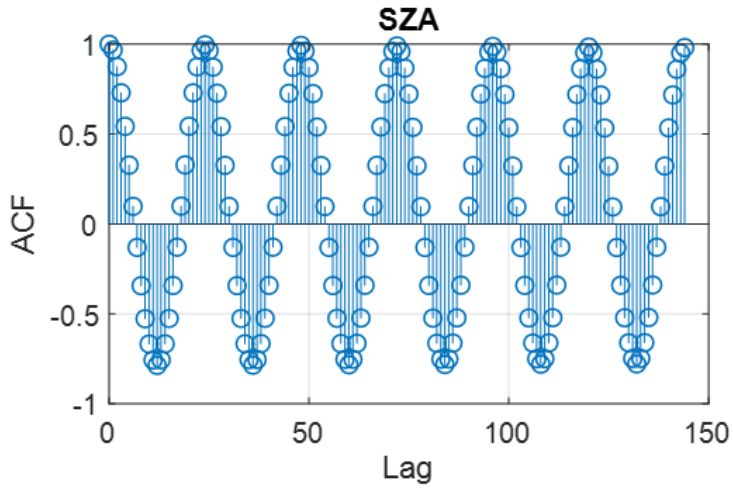

Figure 3.20. Autocorrelation function (ACF) of SZA predictor. The peaks and troughs are 24 hours apart. SZA is the solar zenith angle.

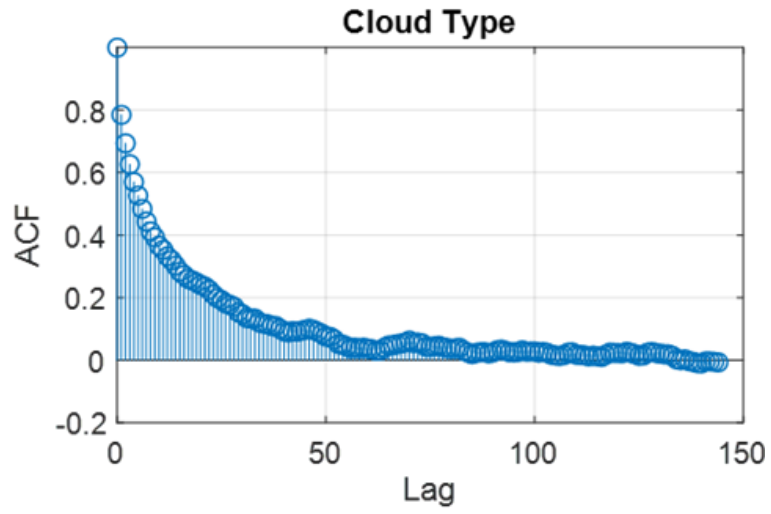

Figure 3.21. Autocorrelation function (ACF) of the cloud type predictor.

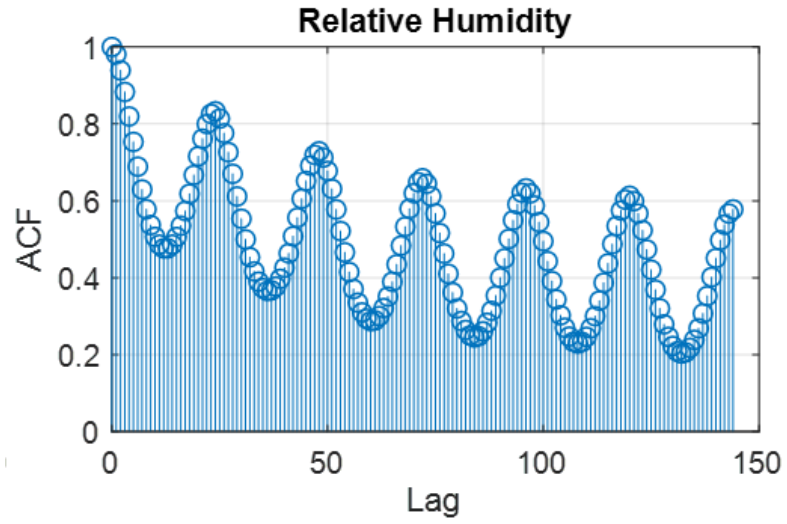

Figure 3.22. Autocorrelation function (ACF) of the relative humidity predictor. The peaks and troughs are 24 hours apart.

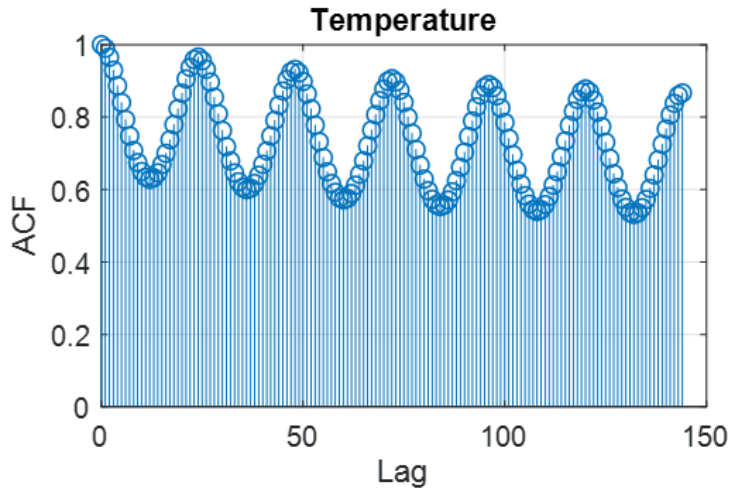

Figure 3.23. Autocorrelation function (ACF) of temperature predictor. The peaks and troughs are 24 hours apart.

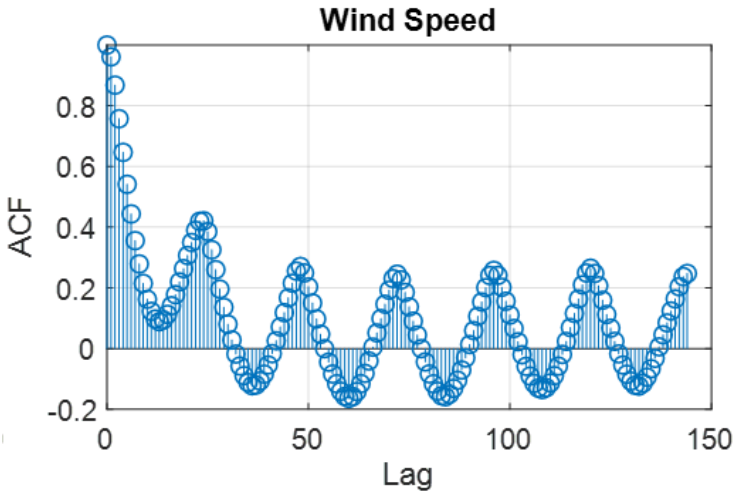

Figure 3.24. Autocorrelation function (ACF) of wind speed predictors. The peaks and troughs are 24 hours apart.

A feed-forward neural network (FFNN) [39], also known as multi-layer perceptron, is implemented because we find it to be the most popular and simple to implement architecture. The proposed neural network architecture comprises of the input, output (or target) and hidden layers. As already discussed, there are six weather variables as inputs and one target parameter. There is no standardized method for determining the optimal number of hidden layers and neurons, most proposed approaches are derived by trial and error. A common approach suggested in available

literature for choosing the number of hidden layer parameters involves cross-validating the performance of a model using different network configurations and adopting the network that minimizes the statistical error while avoiding overfitting problems [40]. Other approaches involve deciding the number of neurons by the ratio between the number of inputs to the target variables [41].

Consequently, the network architecture comprises of a single hidden layer with the number of hidden neurons arbitrarily (trial and error) chosen to be about two-third the number of input variables. Since algorithms in existing literature for estimating an optimal number of hidden layers and neurons did not meaningfully improve our results, these algorithms [40], [41] are not discussed in detail. The network is trained using the Levenberg-Marquardt back-propagation algorithm; this algorithm is relatively more efficient and consumes less time [42]. Two different scenarios of the proposed forecast model based on the autoregression of the predictors as model inputs are considered. Scenario 1, which generated two cases (Cases 1 and 2), aims to explore and exploit the dependency between historical values of the predictors by their combination to yield multiples of input values fed to the ANN. Case 1 involves 18 input neurons comprising three sets of the six predictors for each hour (t), the day before $(t-24)$, and two days before $(t-48)$. Case 2 uses hour (t), one hour prior $(t-1)$, and two hours prior $(t-2)$ to consider a shorter lag time. In Scenario 2, only one set of the six predictors constituting six input variables is fed to the ANN. This exploration is motivated by plots of the autocorrelation coefficients in Figures 3.19 – 3.24 with strong linear relationships between the lagged values. The autocorrelation function (ACF) plots show varying trends and seasonal behaviors. The cloud type plot shows a significant trend (positive ACF values), decaying slowly with time until the zero mark, while the others show a daily seasonality (scalloped shape) or a combination of both effects.

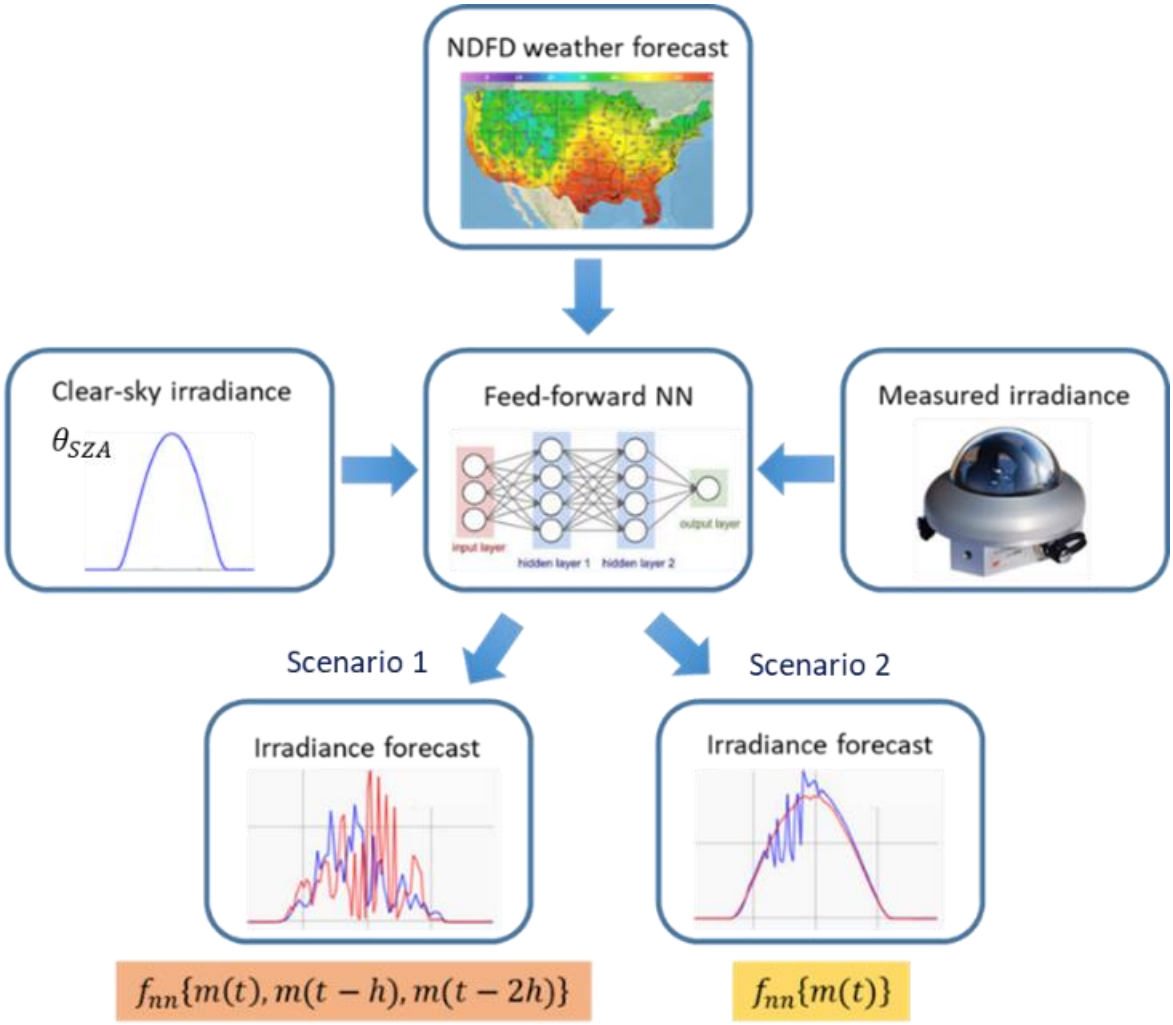

Figure 3.25. Block diagram of the proposed irradiance forecast model. Scenario 1 represents the exploratory Cases 1 and 2 with daily and hourly autoregressions of the predictor variables, respectively, whereas scenario 2 is without autoregressions.

Figure 3.25 represents a schematic structure of the proposed irradiance forecast model. Six predictor variables are fed as inputs to the ANN to predict the measured irradiance in a day-ahead time horizon. Temperature, humidity, cloud cover and windspeed predictor variables are obtained from NDFD whereas CSI and θ_{SZA} are simulated. The measured irradiance fed to the ANN as target for training and testing purposes was sourced from ground-based weather monitoring stations. The outputs in Figure 3.25 are mathematically described by (5) to (7).

Scenario 1:

$$Case\ 1: I_{GH}(t) = f_{NN}\{m(t), m(t-24), m(t-48)\} \quad (5)$$

$$Case\ 2: I_{GH}(t) = f_{NN}\{m(t), m(t-1), m(t-2)\} \quad (6)$$

Scenario 2: $I_{GH}(t) = f_{NN}\{m(t)\}$ (7)

where:

f_{NN} is the function approximated by the neural network,

m represents a set of the six meteorological predictors and

t is the time.

The input meteorological data (except $\cos(\theta_{SZA})$ and I_{GHcsk} which were simulated) used for training the networks are archived NDFD forecasts. Therefore, for Case 1, the observed or measured $m(t-24)$ and $m(t-48)$ are actual values that are available during the forecasting process and, as such, are preferred with respect to potentially increased model accuracy.

3.3 PV Power Simulation

After irradiance prediction, simulating power production from PV modules or an array can be done using either a physical model, a machine learning (ML) model, or a hybrid of both. ML models require historical PV power production and measured irradiance data for training a model to approximate the relationship between power output and irradiance forecasts. Physical approaches are comprised of several mathematical equations that try to capture the dynamics of different processes characterizing the conversion of sunlight to power. The hybrid option applies the physical model and then smooths the results using ML.

Although the physical model appears to be more cumbersome, it is often preferred as consistently monitored and reported historical power output data especially for behind-the-meter or residential PV systems are usually not available. Roof-top solar generators are alternatively called behind-the-meter systems because they are not equipped with the instrumentation for direct monitoring by utility operators. Every physical model requires I_{GH} to be decomposed into its direct normal (DNI) and diffused (DHI) components, and then fed to a transposition model that outputs the plane of array irradiance (I_{poa}). Several decomposition and transposition models exist [43] and can be applied depending on available input information and preferences informed by the location of the PV site. The accuracy of a model is location sensitive and changes accordingly

[43]. I_{poa} is converted to DC power and subsequently AC power output using a DC model and inverter model, respectively. The physical model illustrated in Figure 3.26 can be easily implemented using the PV_Lib Toolbox on MATLAB. The MATLAB scripts were developed by a collaboration of NREL and Sandia Laboratory for easier modelling of power output from PV arrays with fixed-tilt angles. Input variables include a description of the array metadata, the site location, irradiance (GHI) time series and the simulation period (date).

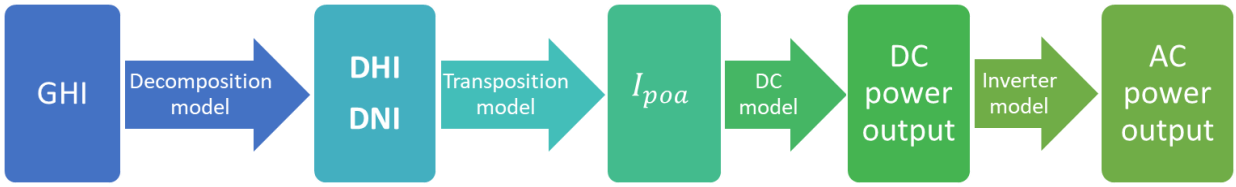

Figure 3.26. Flowchart showing the sequence for power conversion from global horizontal irradiance using the physical model of PV power simulation.

A major input parameter requirement for the physical model is the PV module orientation and inclination angles. The precision of decomposition and transposition models in estimating I_{poa} from horizontal irradiance components is heavily impacted by the module tilts. These parameters reported on public or user-controlled webpages are sometimes plagued by a variety of human oversight errors resulting from inaccurately estimated or reported values. At other times, these parameters are not even made available on these websites, consequently necessitating reliance on ML techniques or inference from mere physical evaluation of pictures provided by users. As such, an alternative ML-based simulation model is employed in this work in addition to the physical model. The quadratic power projection procedure in [44] is adopted and further modified in this chapter. Our modified model produces results with substantial power simulation accuracy. The modifications involve using a nonlinear least squares (NLS) function to derive the coefficients, $a_1 - a_6$, in (8) by fitting to measured PV power data:

$$P_f = a_1 + a_2 \cdot I_{poa} \cdot LF + a_3 \cdot T_{amb} + a_4 \cdot I_{poa} \cdot LF \cdot T_{amb} + a_5 \cdot \left(I_{poa} \cdot LF\right)^2 + a_6 \cdot T_{amb}^2 \quad (8)$$

where:

P_f is the projected power,

LF is the loss factor, and

T_{amb} is the ambient temperature.

3.4 Representative PV Site Selection and Power Output Upscaling Model

The number of distributed PV systems is on the rise and individual PV systems have differing unique features which make simulating a regional PV output as a single system unfeasible. Moreover, the basic idea of simulating every PV system in the region would require us to have a huge database of information as well as the capacity to handle a large number of computations. While regional utility operators maintain records of grid-tied PV systems, only the system capacity and location information are stored in all but metered large-scale systems.

Other data required for power output simulation are not readily available, including hardware configuration, make of panels and inverters, and dynamic characteristics, such as snow cover, shading levels and equipment failures. As a result, independent PV systems with the same capacity and available solar radiation can have different power profiles. Therefore, a model to navigate these hurdles and produce regional-scale PV power generation outputs with sufficient accuracy is required to fill in this technical gap.

Neighboring PV systems share similar power generation profiles, so a simplified model would be to simulate the PV power generation output of representative systems and then upscale the results to a regional level. This concept poses challenges with respect to how to optimally represent system parameters when selecting the representative lots as misrepresentation could introduce systematic errors. For a better description of the overall power production, basic properties including the distribution of system orientations, mix of module and inverter types, and an optimal spatial distribution of the representative set should be ensured.

The model proposed for this study involves dividing the region into geographical clusters based on the Euclidean distance between PV sites and select representative systems. Selection criteria within each cluster include proximity to the centroid of a cluster, nominal capacity, and prevalent tilt and orientation angles. Next, we simulate power outputs from each representative system and then use a simple mapping function to estimate the regional level based on any of the two modes described by (9) to (11).

$$P_{region}(t) = f\left(\sum_{i=1}^{N} P_{rep,i}(t)\right) \quad (9)$$

$$P_{cluster,j}(t) = f\left(P_{rep,j}(t)\right) \quad (10)$$

$$P_{region}(t) = \sum_{j=1}^{m} P_{cluster,k}(t) \quad (11)$$

where:

P_{region} is the regional-level PV system power outputs,

$P_{cluster}$ is the cluster-level PV system power outputs,

P_{rep} is the representative PV system power outputs,

N is the total number of representative systems,

m_c is the total number of clusters,

j is the $j-th$ representative system,

k is the $k-th$ cluster, and

f is the mapping function representing the ratio between nominal capacities.

3.5 The Proposed Model Execution

3.5.1 Data Collection from Case Study Solar Sites

To implement the proposed model, real-world data are used in this chapter. Three different datasets are collected, including measured solar irradiance from monitoring stations, weather data from NDFD and solar power measurements from a public webpage. Several webpages available in the public space with registered solar systems owners monitoring their outputs were considered for suitability for our study. Solrenview.com was preferred because the webpage offered round the clock data reporting with hourly time steps and minimal hurdles to automatic data extraction. However, individual PV systems had varying levels of inconsistencies and missing data entries. A

total of 25 independent systems within the state of California were selected to meet the research objectives. In addition to PV site metadata, power production data were collected for the period from March 2022 to September 2022 from the webpage.

3.5.2 Weather and Irradiance Datasets Description

The chosen 25 PV systems were divided into 4 clusters as shown in Table 3.1 and Figure 3.34 using k-means clustering technique [7]. All solar irradiance and weather data were collected for only the 4 representative systems for the same period as the power data. A 3-h temporal resolution NDFD day-ahead forecasts of the input predictors were collected and interpolated for hourly intervals. Raw irradiance files from available ground-based weather monitoring stations closest to the selected representative sites were also collected for every cluster for training and testing purposes. Notably, it was practically impossible to find the ideal or perfect situation for all of the systems, with the perfect case being consistent power generation reporting for at least one calendar year. In this study, 70% of the data in the dataset were used for extracting the irradiance-power relationship, and the remainder for out-of-sample testing.

Table 3.1. Selected solar site metadata categorized into clusters.

Clusters	**1**	**2**	**3**	**4**
Representative solar module capacity (kW)	48.1	114.9	40.3	10.6
Inverter capacity (kW)	42	100	42	10.8
Proximity to weather station (km)	30	16	0	31
Number of solar sites in a cluster	9	3	7	6
Total nominal cluster capacity (kW)	488	276.4	379.6	1093

3.5.3 Datasets Cleaning and Scaling

Two common characteristics of measured I_{GH} data are non-zero nighttime entries and missing entries caused by equipment failure. Physically impossible irradiance values were flagged using (12) [44] and were excluded from the analysis. Similarly, filtering was applied based on θ_{SZA}

to omit irradiances corresponding to θ_{SZA} values exceeding 95° ($\theta_{SZA} > 95°$) so that only daylight ($I_{GH} > 0$) hours were considered. I_{ext} is the extraterrestrial irradiance. We also scaled the proposed model input and output parameters using a min and max values of 0 and 1 respectively in (13).

$$I_{GH} < 1.5 \cdot I_{ext} \cos(\theta_{SZA})^{1.2} \tag{12}$$

$$P_{scaled} = \frac{P_{parameter} - P_{min}}{P_{max} - P_{min}} \tag{13}$$

where:

$P_{parameter}$ is either the predictor or the target variable.

For model validation and testing, the performance of the proposed irradiance forecast model and regional PV power forecast results are discussed by comparing with Yang's model in [21] and Perez's model in [19]. The two existing models have similar objectives to ours since they omitted irradiance data as a predictor variable. The proposed model is implemented in MATLAB R2022a on a PC with an Intel Core i7-6700 3.40GHz CPU and 16 GB RAM.

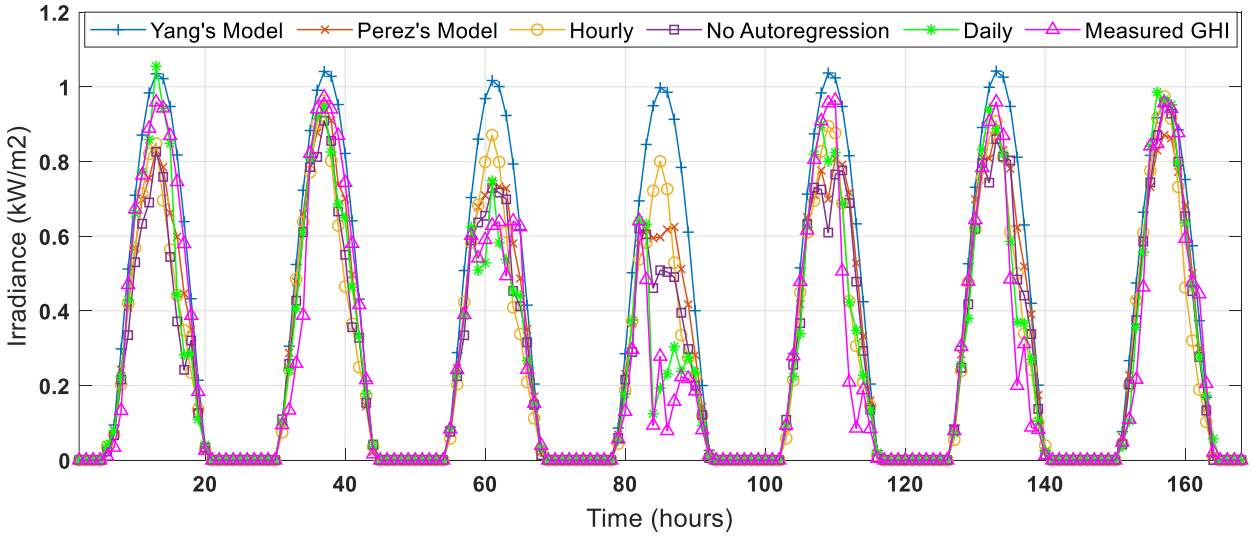

Figure 3.27. Irradiance time plots of the performance of the forecast models for the first 7 days of July 2022. Hourly and daily denote the autoregression lags.

3.5.4 Error Statistics

We use the following error metrics (14) – (16) to estimate how well the models under study

can replicate the measured or observed variables.

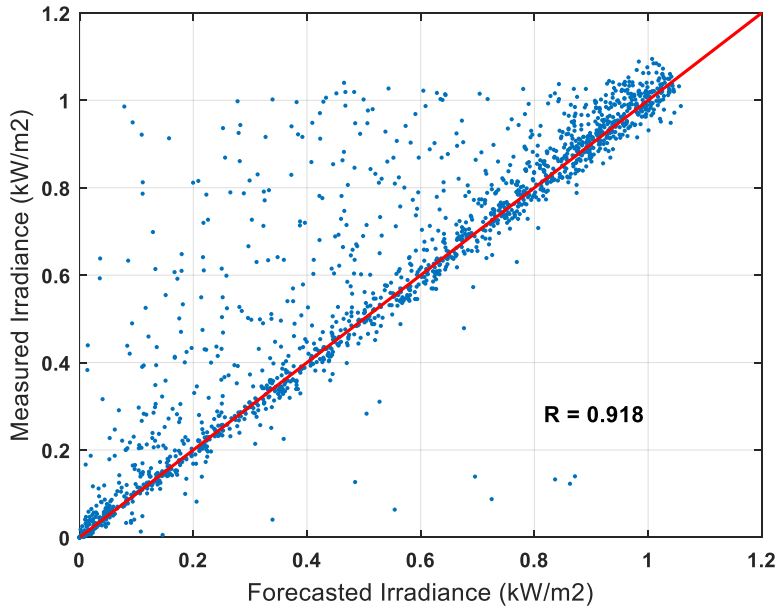

Figure 3.28. Scatter plot with correlation coefficient (R) of measured vs. forecasted irradiances for the entire test period using Yang's model.

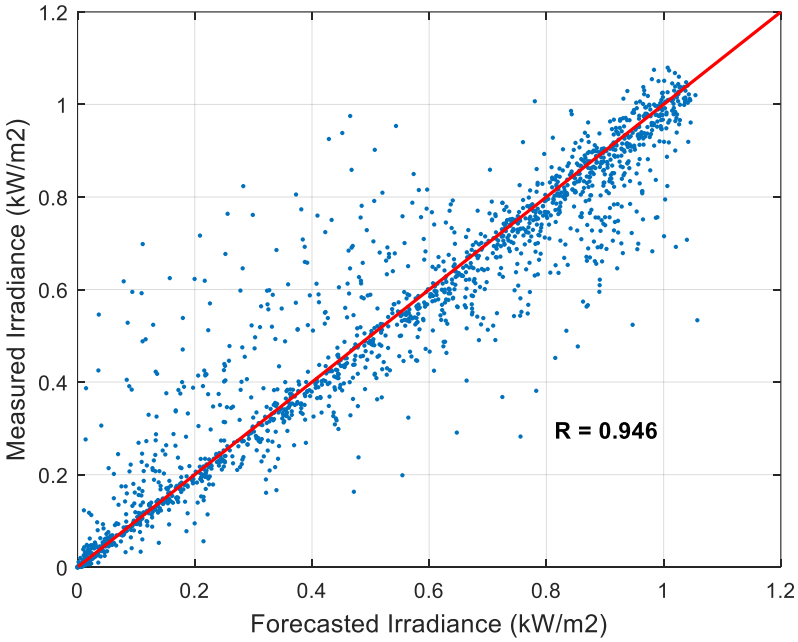

Figure 3.29. Scatter plot with correlation coefficient (R) of measured vs. forecasted irradiances for the entire test period using Perez's model.

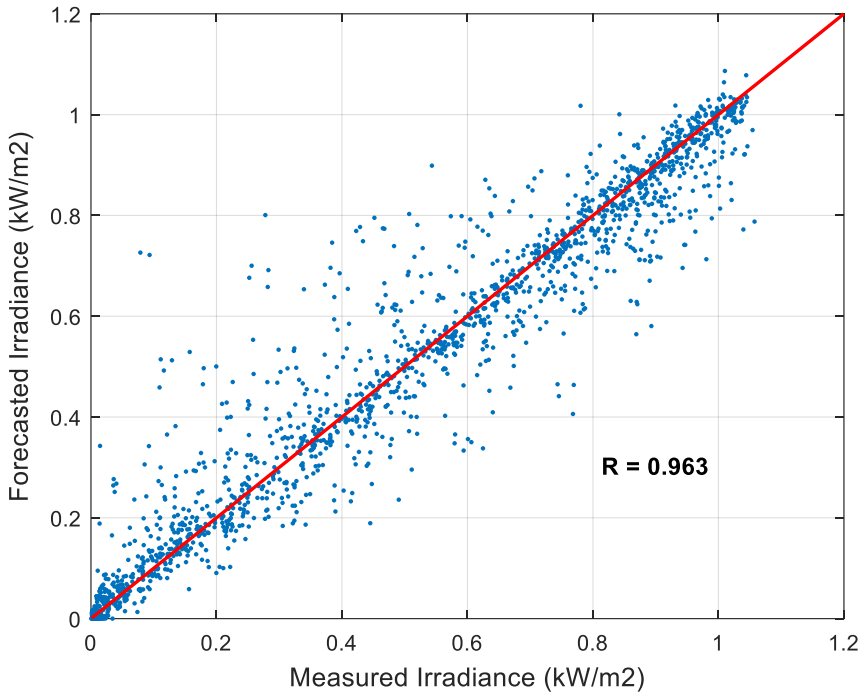

Figure 3.30. Scatter plot with correlation coefficient (R) of measured vs. forecasted irradiances for the entire test period using the hourly autoregression model.

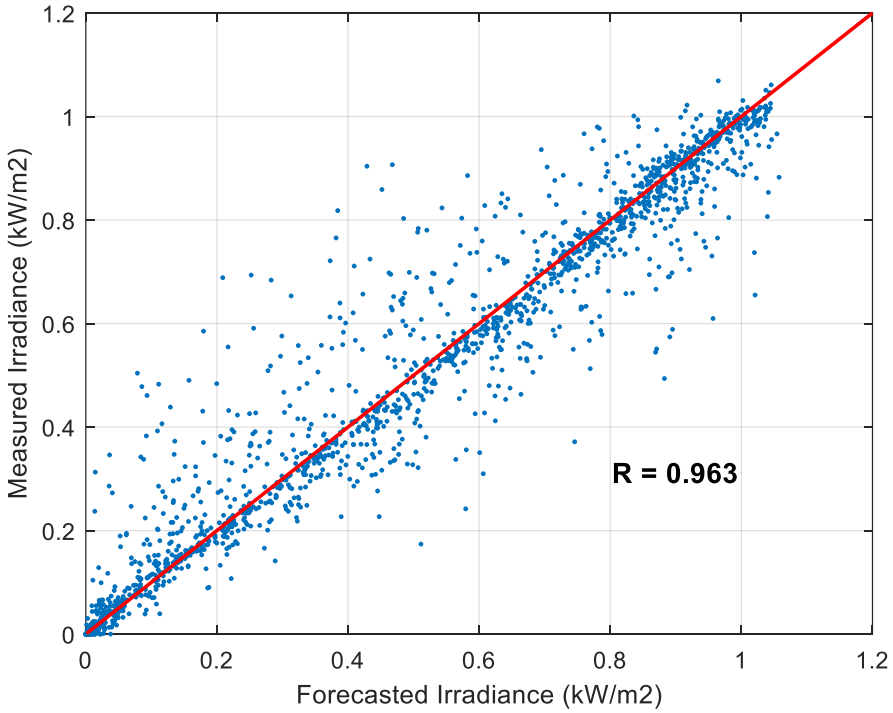

Figure 3.31. Scatter plot with correlation coefficient (R) of measured vs. forecasted irradiances for the entire test period using the model without autoregression.

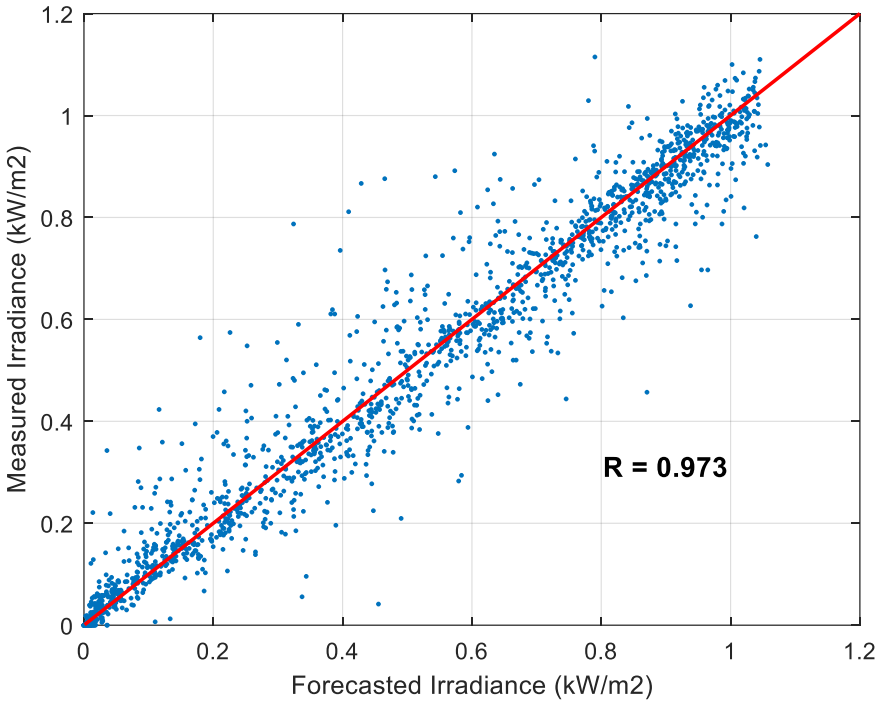

Figure 3.32. Scatter plot with correlation coefficient (R) of measured vs. forecasted irradiance for the entire test period using the daily autoregression model.

$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(I_{GH_{f,i}} - I_{GH_{m,i}}\right)^2}}{I_{GH_{m,peak}}} \times 100\% \quad (14)$$

$$NMAE = \frac{\frac{1}{n}\sum_{i=1}^{n}\left|I_{GH_{f,i}} - I_{GH_{m,i}}\right|}{I_{GH_{m,peak}}} \times 100\% \quad (15)$$

$$WMAPE = \frac{\sum_{i=1}^{n}\left(\left|I_{GH_{f,i}} - I_{GH_{m,i}}\right|\right)}{\sum_{i=1}^{n}\left|I_{GH_{m,i}}\right|} \times 100\% \quad (16)$$

3.5.5 Irradiance Forecast Model Performance

Using the first cluster as a case study, the results shown in Figures 3.27 – 3.32 are obtained after several trainings of the networks for both scenarios. Figure 3.32 primarily represents positives of the proposed model, whereas Figures 3.28 – 3.32 goes further to highlight the convergence levels. The proposed model's performance using some error metrics are shown in Table 3.2. The

Yang's model visibly overestimated the irradiance in the scatter plot. Normalized to $I_{GH\,m,peak}$, the daily autoregression model offered the best results, with an improved NMAE value of 4.58% compared to the 4.94% yield of the model without lags. This marginal improvement may question the justification of using autoregression because the results will not be significantly improved or affected irrespective of the combination of inputs to the network. Regardless, a slight change in accuracy metrics could mean a huge difference when the PV's nominal capacity is large. Thus, the daily autoregression model is preferred for PV power simulation in this study.

Furthermore, we employ the t-statistic ($t - stat$) metric detailed in [45] to evaluate the statistical significance of the models under study in Table 3.2. We use the metric to compare the forecasts with the target irradiance, smaller t-stat values represent better forecasts. A model is statistically significant or useful if its t-statistic value is below t-critical (t_c). All the models except Yang's model have the t-statistic below the critical threshold. It is worthy of note that the t-statistic metric is reliable only for small samples or where the population variance is unknown.

Table 3.2. Irradiance forecast accuracy using NRMSE and NMAE statistical metrics.

Forecast Method	**NRMSE (%)**	**NMAE (%)**	**t-stat ($t_c = 1.64$)**	**R**
Without Autoregression	8.44	4.94	0.194	0.96
Hourly Autoregression	8.42	5.07	0.106	0.96
Daily Autoregression	**6.58**	**4.58**	0.125	**0.97**
Perez Model	9.90	5.74	**0.07**	0.95
Yang Model	13.30	6.47	4.29	0.92

In the daily irradiance profile in Figure 3.27, only Day 3 and Day 4 from the left are either partly or almost fully cloudy. Analyzing the model forecast strength based on only clear-sky days would be counterproductive because the actual I_{GH} corresponds to I_{GHcsk} thereby concealing vital information about the model's strengths. With keen interest in Day 4, the daily autoregressive model had a significantly better following with the measured irradiance than the other models. Although only select portions can be feasibly sampled, visual inspection of the time plots shows reasonable correlations between measured and forecasted irradiance, especially for the daily

model. This observation is further confirmed by scatter plots in Figures 3.28 – 3.32.

Table 3.3. Pearson's coefficient of correlation (R) between the six meteorological variables

	CSI	**T**	**RH**	**CC**	**WS**	**SZA**
CSI	1.00	0.30	-0.22	0.01	0.26	**-0.99**
T	0.30	1.00	-0.25	0.03	0.06	-0.33
RH	-0.22	-0.25	1.00	**0.54**	-0.34	0.22
CC	0.01	0.03	**0.54**	1.00	-0.07	-0.01
WS	0.26	0.06	-0.34	-0.07	1.00	-0.27
SZA	**-0.99**	-0.33	0.22	-0.01	-0.27	1.00

3.5.6 Redundant Predictors Test

One of our objectives is to achieve the optimal forecast accuracy with the minimum number of input parameters. From the correlation coefficients in Table 3.3, there is a strong negative linear relationship between the CSI and SZA, making one of the variables redundant. Omar et al. [46] observed a similar correlation level and eliminated SZA at a penalty of only a marginal model error increase.

We carried out an analysis similar to the approach in [46] to identify, and if justified, exclude redundant variables as inputs. Based on strengths of the correlation coefficients in Figures 3.11 – 3.18 and Table 3.3, a limited number of input variable combinations are experimented with. The bar chart in Figure 3.33 illustrates the results of our analysis. As expected, the input combinations involving CSI and/or SZA can produce considerably more accurate results because the intrinsic characteristics of both variables share a resemblance to I_{GH} in the absence of clouds. Although RH shows a higher correlation than temperature, only a slight difference exists in their forecast accuracies. The optimal accuracy occurs when all six variables are included. The analysis indicates that every input variable contributes to accuracy improvement, with CSI and SZA being the most relevant. SZA is not removed as a predictor because even a marginal contribution to model error reduction is significant to this work considering a regional scale installed capacity.

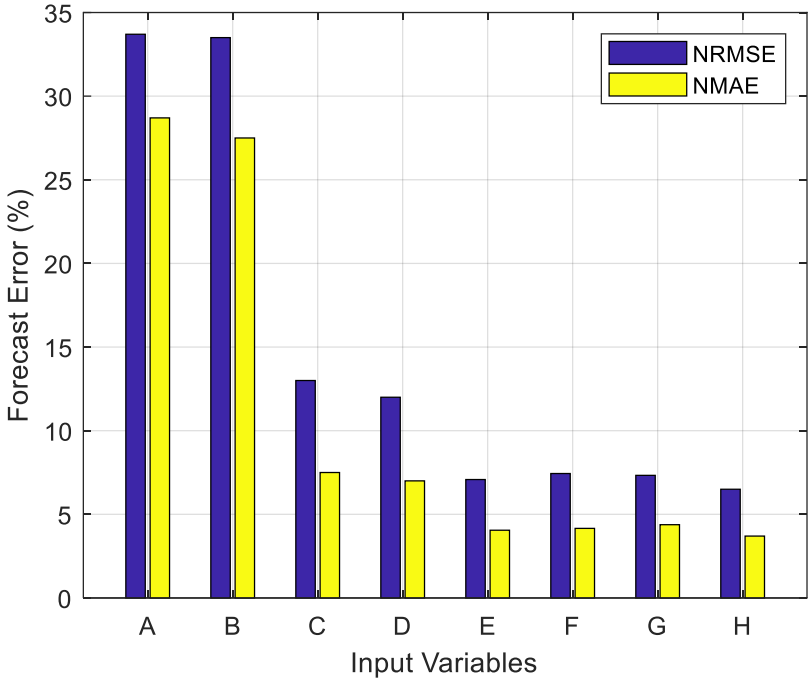

Figure 3.33. Average forecast error of eight meteorological variables mixes for redundancy check. A: T only; B: RH only; C: SZA only; D: CSI only; E: cloud cover, wind speed, temperature, RH and CSI; F: cloud cover, wind speed, temperature, RH, and SZA; G: wind speed, temperature, RH, CSI, and SZA.

3.5.7 Cluster Segmentation

The choice of representative PV systems is made by applying data dimension reduction techniques to the curated datasets. First, the k-means clustering algorithm is used to create four clusters of solar sites (in Table 3.1). The optimal number of clusters is determined by the geographical layout of the sites as shown in Figure 3.34. The objective of the algorithm is to group similar sites into clusters using only their longitude and latitude data while ensuring the dissimilarity of one cluster from another. The algorithm uses unsupervised learning methods to iteratively identify centroids and calculate the Euclidean distances between the locations of the sites. The sites are grouped into clusters using a (distance) minimization objective function. The solar site closest to the centroid of a cluster was considered a suitable candidate for modeling the whole cluster.

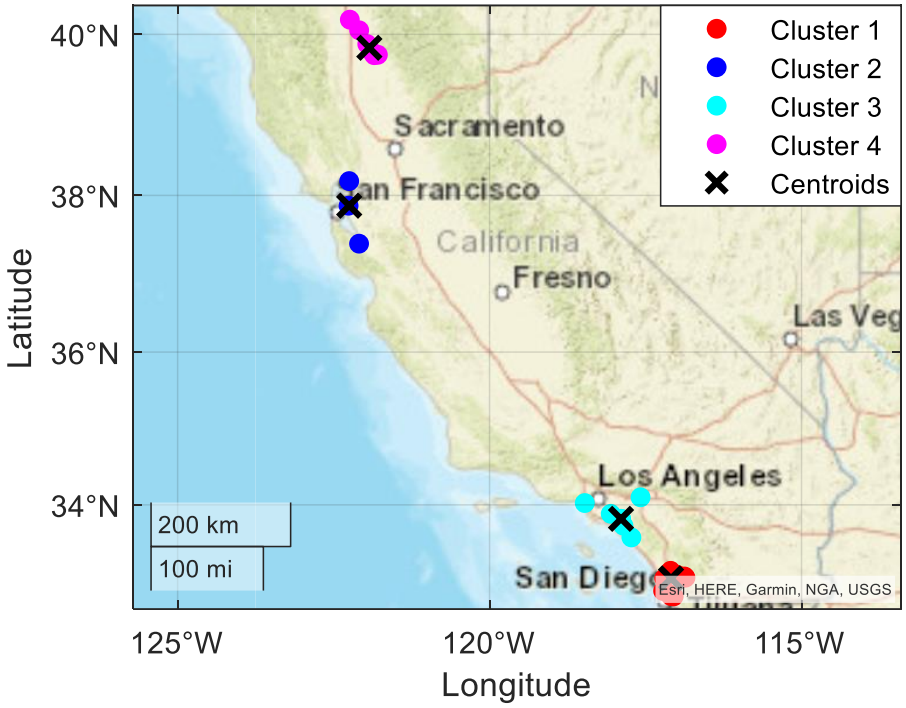

Figure 3.34. Clusters and centroids created using the k-means clustering algorithm. The Euclidean distance between the solar sites is the only determinants of the k-means clustering process.

To ensure a chosen solar site is either truly representative of the whole cluster or is the best available case, PCA (a linear dimension reduction technique) [7] is implemented in addition to k-means clustering on the normalized power generation time series data to rank the candidate sites in a cluster according to their principal components. The site containing more variant information is ranked the highest. Furthermore, the following conditions are taken into consideration in order of priority for selecting the representative sites: power production data and site metadata reporting; regular reporting without prolonged periods of missing data; reasonably close to a ground-based weather station providing irradiance measurement data with easy accessibility; and closest to the centroid of a cluster and the principal component using PCA. The sites with details in Table 3.1 and close to the centroids in Figure 3.34 were selected as the representative sites.

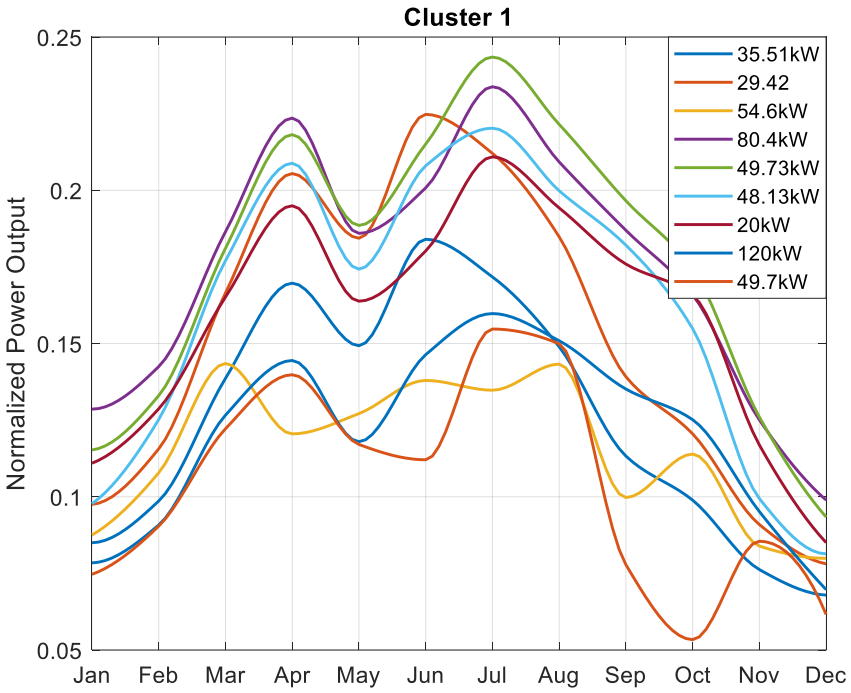

Figure 3.35. Monthly average power generation data in 2019 for each solar site in cluster 1 normalized to their reported respective nominal capacity in the clusters created using the k-means clustering technique.

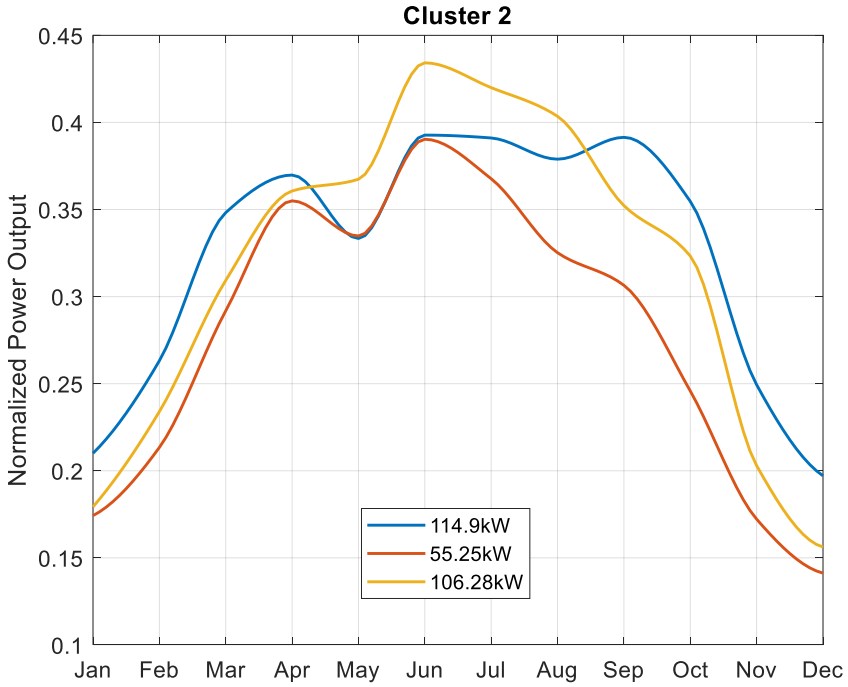

Figure 3.36. Monthly average power generation data in 2019 for each solar site in cluster 2 normalized to their reported respective nominal capacity in the clusters created using the k-means clustering technique.

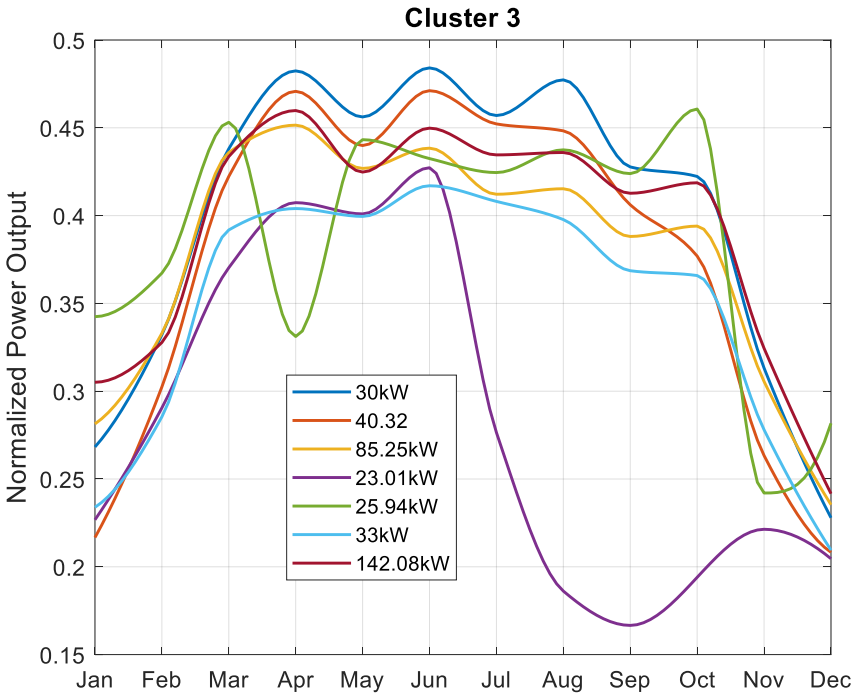

Figure 3.37. Monthly average power generation data in 2019 for each solar site in cluster 3 normalized to their reported respective nominal capacity in the clusters created using the k-means clustering technique.

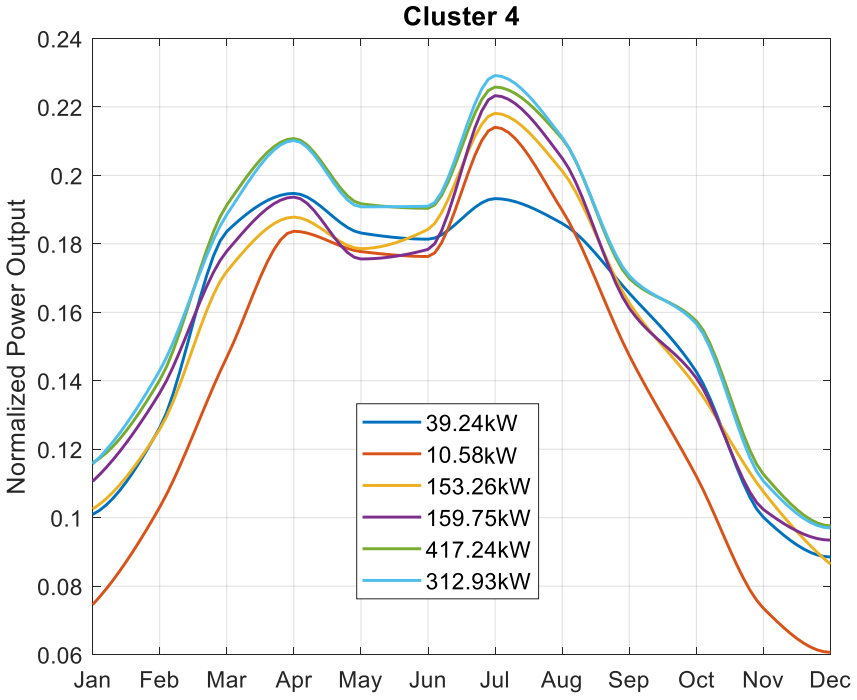

Figure 3.38. Monthly average power generation data in 2019 for each solar site in cluster 4 normalized to their reported respective nominal capacity in the clusters created using the k-means clustering technique.

Figures 3.35 – 3.38 demonstrate the rationale to use clusters of solar sites and representative systems. The line plots show the 25 PV systems categorized into four clusters. Their annual outputs are normalized to their individual nominal capacities and averaged on a monthly basis for easy visualization of power generation profiles. While the results show a similar power generation profile for sites in a cluster is realistically possible, solar sites still experience errant profiles resulting from several causes. For some sites with multiple inverters, a large error is observed when there are inverter downtimes, or some inverters stop reporting power output. The downtimes can be either due to electrical faults or connection breakdowns between communication devices and the webpage database. Cluster 1 has a significantly lower than average cluster power profile throughout the year for some sites relative to the predominant number of solar sites. A possible contributor to this mismatch is the incorrectly reported nominal installed capacity or situations where some PV modules may be decommissioned without due updates on the webpage. Other factors, including cover from trees/high rise buildings and snow, dirt/dust on solar panel surfaces, and cloud cover, that could potentially reduce power production, may not have substantially contributed to anomalies in Clusters 1 and 3.

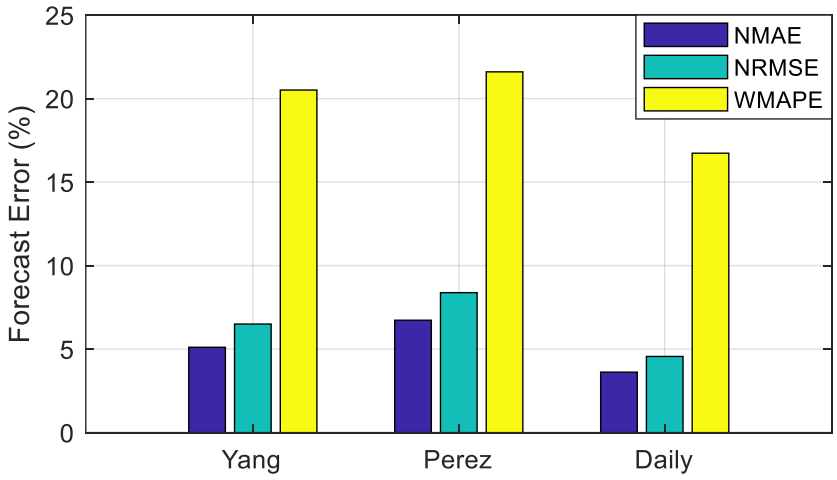

Figure 3.39. Benchmarking of the regional-level PV power forecast performances of the three models.

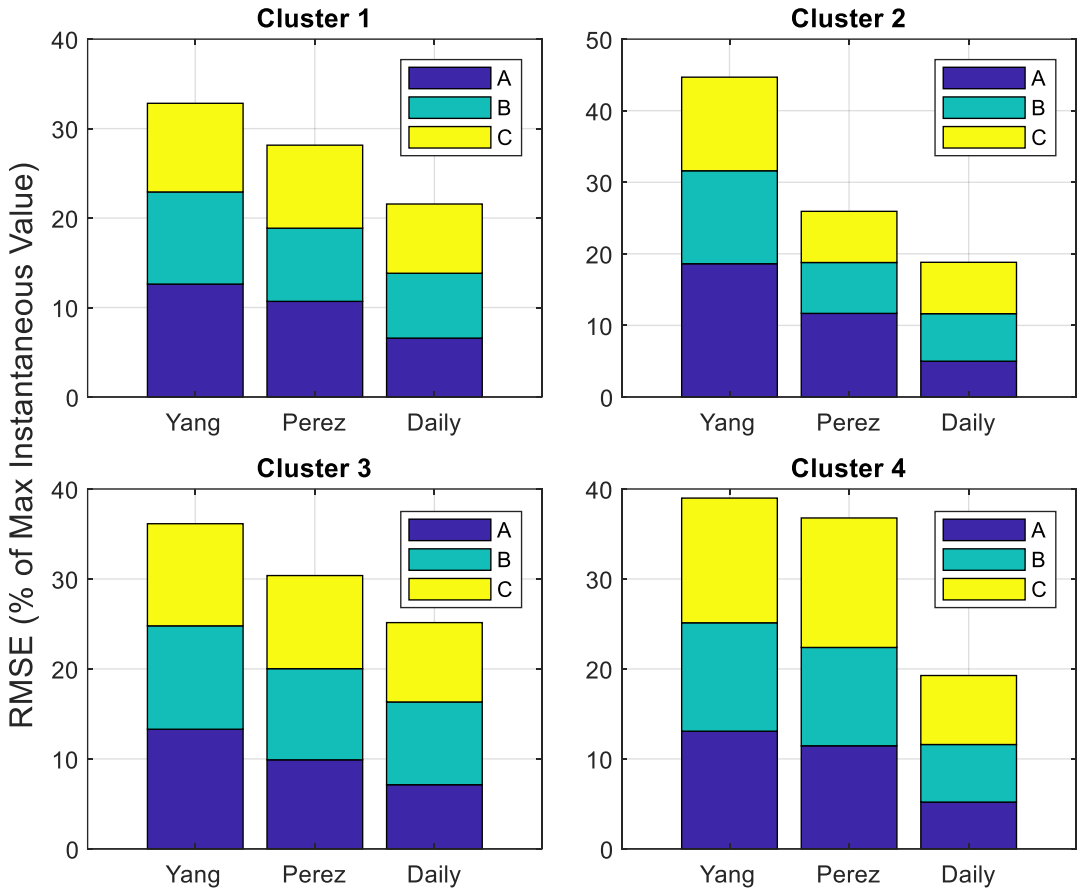

Figure 3.40. Bar chart of the non-cumulative cluster-level NRMSE as a percentage of maximum instantaneous value: (A) Irradiance forecast; (B) representative system power modelling; (C) cluster-level power modelling.

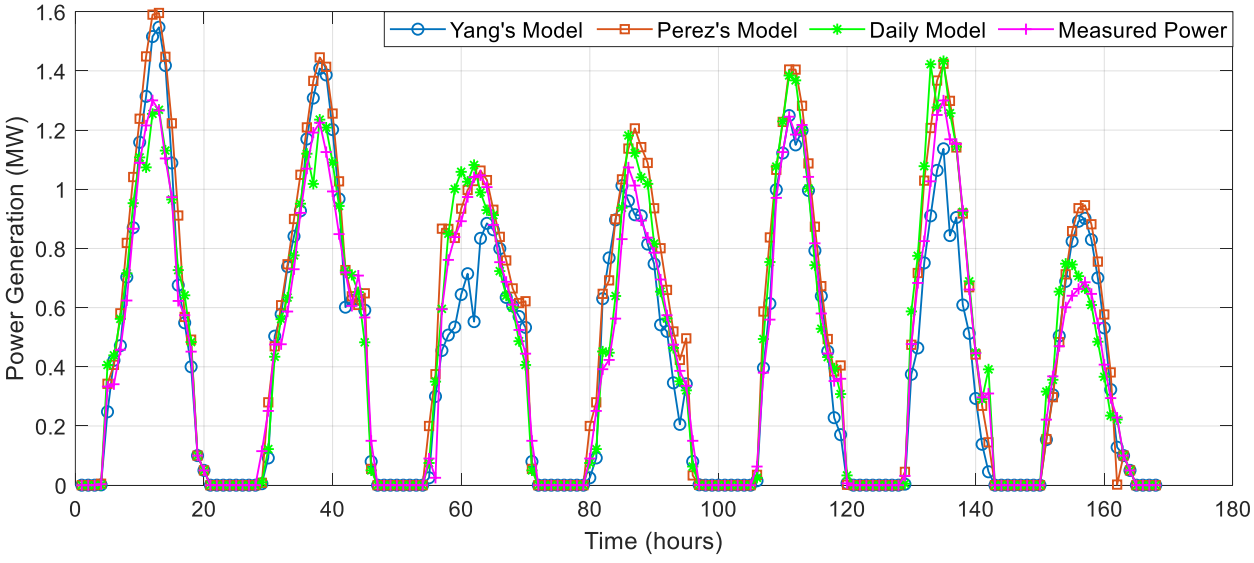

Figure 3.41. Time plot of a randomly sampled 7 days of the proposed model compared to Yang's and Perez' models for May 2022 regional-level power generation.

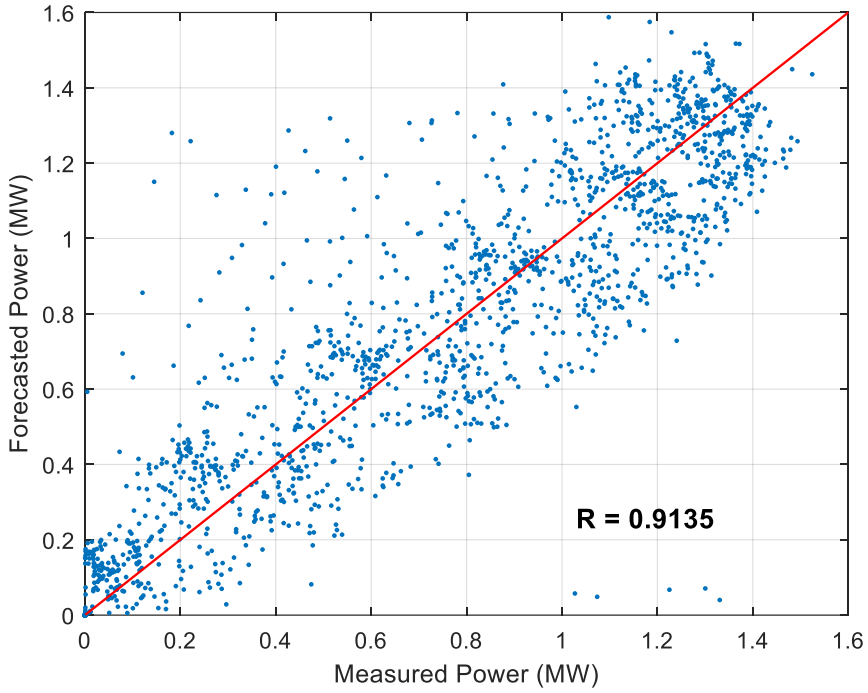

Figure 3.42. Scatter plot with the correlation coefficient of power generation at the regional level using Yang's model.

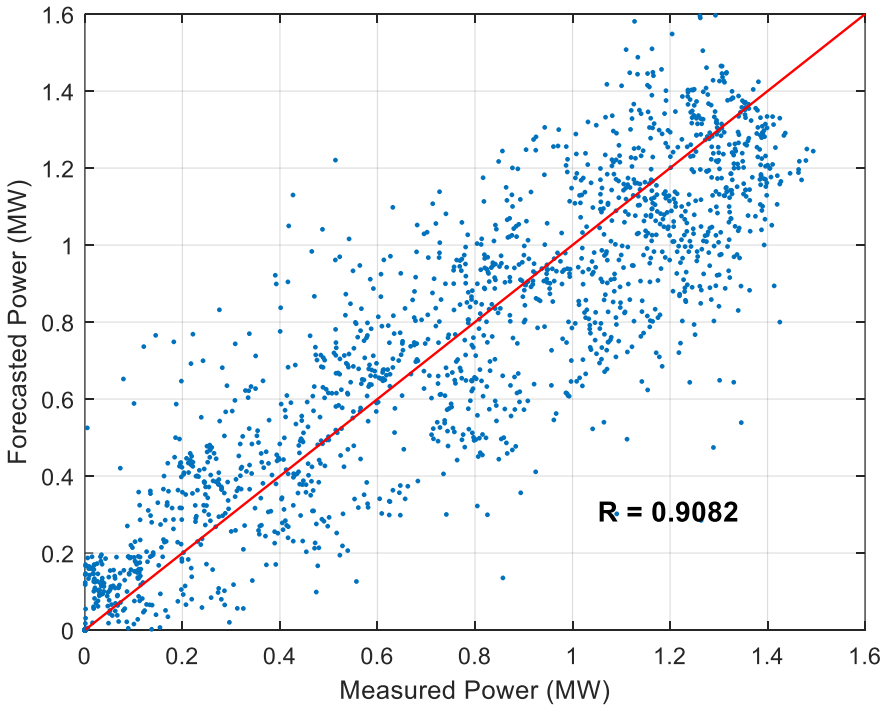

Figure 3.43. Scatter plot with the correlation coefficient of power generation at the regional level using Perez's model.

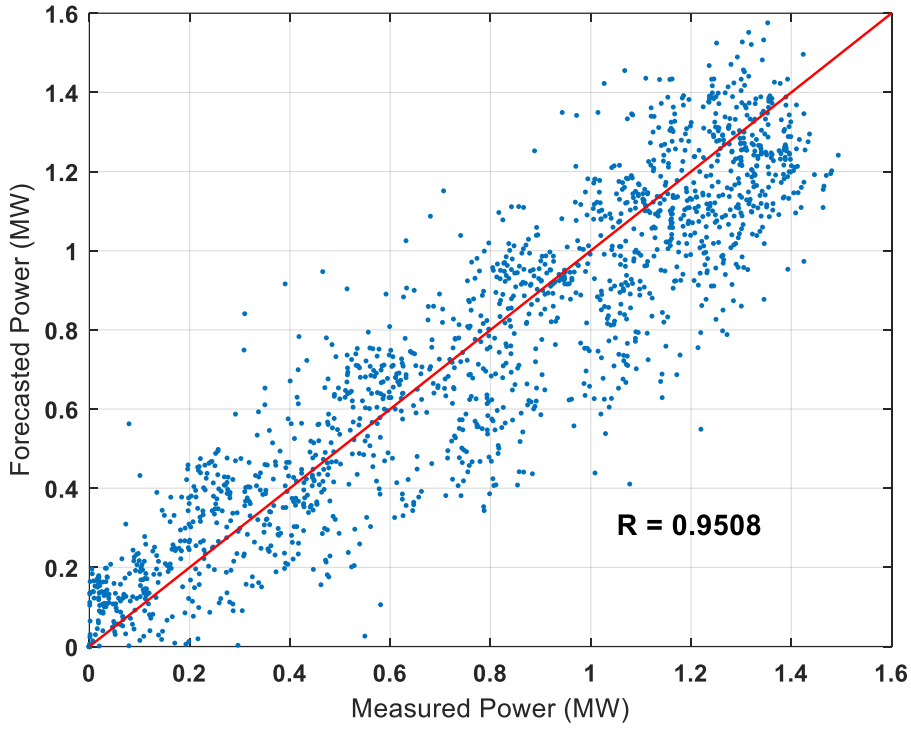

Figure 3.44. Scatter plot with the correlation coefficient of power generation at the regional level using the proposed Daily model.

3.5.8 Cluster And Regional Level Forecast Results

The daily autoregressive day-ahead forecast model formulated in (5) is used to obtain day-ahead irradiance predictions for the representative sites for the four clusters after conscientiously cleaning and normalizing the input data. A trial-and-error method informed by pictures of the solar sites is employed to estimate module inclinations for power simulation models. As expected, the ML option performs better relative to the physical model, primarily due to the latter's heavy reliance on accurate input inclination angles [47], which in this case are obtained by the unreliable trial-and-error approach. The physical model of PV power simulation is recommended when all input parameters are readily available, accessible, and gathered with acceptable reliability. The regional-scale power output is estimated by upscaling cluster outputs using mapping functions (10) and (11). With the nighttime values removed, results obtained for each stage from forecasting irradiance to estimating the regional PV power generation are detailed in Figure 3.39 and Figure 3.40 for easy visualization.

The proposed daily model yields result with NMAE and NRMSE errors of 3.62% and 4.56%, respectively, and visible in the time and scatter plots of Figures 3.41 – 3.44. However, significant improvements can be observed compared to Yang's and Perez' models in Figure 3.39. Whereas the irradiance forecast errors are consistently lower across the clusters for the proposed model, as shown in Figure 3.40, a larger percentage of the errors is introduced at the power modelling stage, which subsequently affects the upscaling model accuracy. A quite prominent example occurs in Cluster 3, which has NRMSE values of 9.2% and 8.82% for the representative sites and cluster level power simulations, respectively. This is in part a direct consequence of the PV power modelling method we adopted, which is still the best available case considering constraints related to unavailability of the required system metadata.

With regards to the WMAPE results, there is no distinctive or standout pattern observable in Figure 3.39 and Table 3.4 other than that the figures are considerably larger than the other error metrics and our proposed model is preferred. As the name suggests, the WMAPE in (15) is normalized by the sum of the absolutes of the measured variable. Consequently, the WMAPE is favorably biased towards positive errors ($I_{GH_m} > I_{GH_f}$) and puts a heavier penalty on overestimations [48]. It is also rational to say that poor test data quality also contributed to the inconsistencies in the WMAPE result.

Although the PV power datasets feature 24-hour reporting, careful analysis reveals that some PV systems are configured to cut off reporting when the generated power is below a threshold and instead report a zero integer in place of actual power output. Effects of geographical aggregation/averaging are evident in error reductions at regional levels for all models in Table 3.4. The degree of averaging as analyzed in several studies [49], [50] is directly proportional to the spatial distribution of systems as well as the number and size of systems in the region of interest.

Table 3.4. Evaluation of the proposed model performance in both the cluster and regional levels using NMAE and NRMSE metrics normalized to their maximum instantaneous power.

		Model	**Rep System Irradiance**	**Rep System Power**	**Cluster Power**
Cluster 1	NMAE (%)	Yang	9.50	7.23	7.53
		Perez	8.07	5.72	7.41
		Daily	**4.58**	**4.86**	**5.54**

	NRMSE (%)	Yang	12.62	10.30	9.91
		Perez	10.69	8.18	9.28
		Daily	**6.58**	**7.25**	**7.74**
	WMAPE (%)	Yang	21.44	18.86	21.8
		Perez	18.22	14.93	21.46
		Daily	**10.29**	**13.00**	**17.12**
Cluster 2	NMAE (%)	Yang	13.08	9.52	9.41
		Perez	7.33	4.90	5.34
		Daily	**3.18**	**3.06**	**4.71**
	NRMSE (%)	Yang	18.62	12.99	13.08
		Perez	11.68	7.11	**7.15**
		Daily	**5.00**	**6.64**	7.18
	WMAPE (%)	Yang	26.26	27.27	26.88
		Perez	14.73	14.02	14.85
		Daily	**6.55**	**8.64**	**13.17**
Cluster 3	NMAE (%)	Yang	6.47	7.55	8.08
		Perez	5.74	7.23	7.75
		Daily	**4.50**	**6.53**	**6.67**
	NRMSE (%)`	Yang	13.30	11.49	11.35
		Perez	9.90	10.13	10.35
		Daily	**4.50**	**9.20**	**8.82**
	WMAPE (%)	Yang	13.51	21.9	22.24
		Perez	11.98	20.93	21.29
		Daily	**9.40**	**18.91**	**18.28**
Cluster 4	NMAE (%)	Yang	10.04	8.41	8.77
		Perez	9.15	8.06	8.72
		Daily	**3.50**	**4.58**	**5.17**
	NRMSE (%)	Yang	13.08	12.05	13.86
		Perez	11.68	10.93	14.39
		Daily	**5.20**	**6.40**	**7.67**
	WMAPE (%)	Yang	23.97	26.4	39.53
		Perez	22.30	25.55	45.68
		Daily	**8.52**	**14.29**	**20.90**
Regional Power		Yang	Perez	Daily	
	NMAE (%)	5.11	6.73	**3.62**	
	NRMSE (%)	6.5	8.38	**4.56**	
	WMAPE (%)	20.51	21.60	**16.73**	
	R	0.91	0.91	**0.95**	

The time plot in Figure 3.41 aims to reveal the trends in forecast and recorded hourly values with particular attention paid to cloudy days in spring or summer when most of forecast errors

occur [33]. The rationale for this interest was fueled by an observation consistent across different forecast models, i.e., that cloud movement in a clear-sky day resulting in significant and sudden drops in power generation is more difficult to predict than a completely cloudy or clear-sky situation. Whereas a forecast model might predict the magnitude of a steep power rise or fall reasonably well, the ability to make these predictions with accurate timestamps is occasional. The proposed model shows slight overestimation and the noticeable timestamp mismatch.

Figure 3.45 shows convergence levels of the energy yield on a monthly basis. Apparently, all three models remarkably underestimated the energy yield for the first three consecutive months. In finding a plausible and coherent explanation for this observation, the energy yield with the monthly averaged irradiance is firstly compared in Figure 3.46 for power generation consistency. Although Figure 3.46 shares a resemblance with Figure 3.45 in terms of the overall output pattern, the former doesn't explain the heavy overestimation in the latter. Also, the fact that the underestimations did not occur throughout the study period, we ruled out the PV power modeler being the major error contributor. Probing further, the power upscaling model, which is heavily reliant on accurate nominal capacity values, is examined. Although a PV system metadata is generally considered static in nature, careful investigation revealed possible changes of these data within the study period. As an illustration, in the measured power generation data from SolrenView, there were instances of instantaneous power exceeding the reported installed capacity. These instances are physically impossible, and they suggest of a likely increase of the installed capacities sometime within the study period or an erroneously reported nominal size.

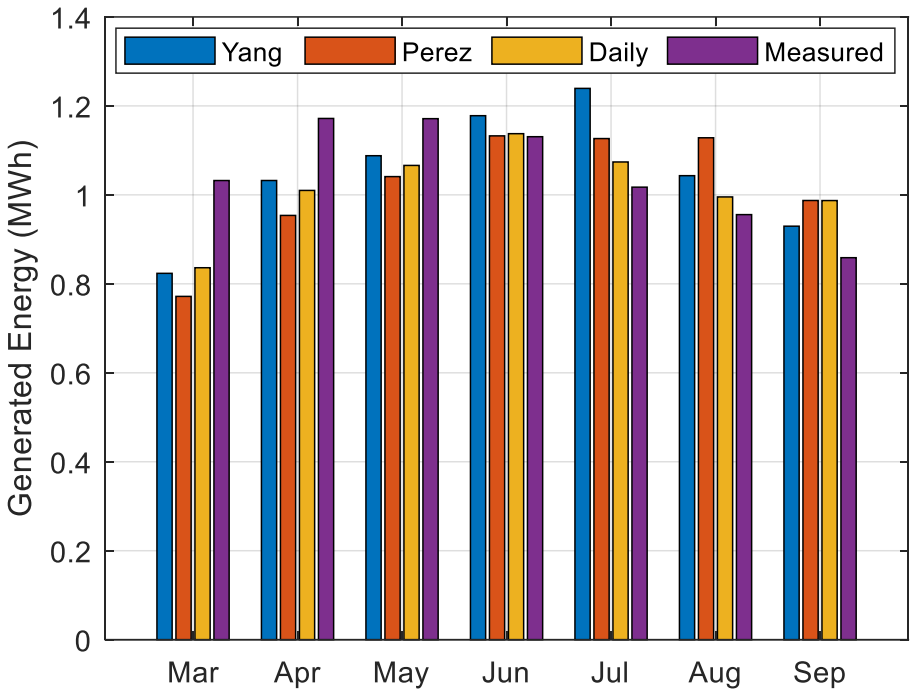

Figure 3.45. Regional level monthly averaged solar energy yield for the test period. Temporal aggregation ignores timestamp mismatch errors.

3.6 Conclusion

In this study, a model for day-ahead regional distributed PV power generation prediction is proposed in situations where numerical weather predictions of solar irradiance are unavailable. The model is comprised of three main components: irradiance forecast, PV power simulation, and cluster creation and power upscaling. The irradiance forecast model employs an ANN-based algorithm to predict day-ahead solar radiation from NWP meteorological variables excluding irradiance as inputs. The proposed model is first implemented by dividing the region into clusters and selecting a representative site for each cluster. The irradiance forecast is then obtained for every representative system and PV power simulated. The cluster power output is obtained using a linear upscaling model and summed to produce the regional scale generated power.

The results obtained justify the use of a small subset of representative solar sites to model the behavior of a larger set. The data dimension reduction techniques produce distinct clusters of solar systems with similar power profiles within each cluster. Spatial averaging in relation to a minimum distance between systems or region size produces a smoothing effect on the results. Day-

ahead irradiance forecasting from NWPs with Daily autoregression results in improved accuracy compared to approaches without regression.

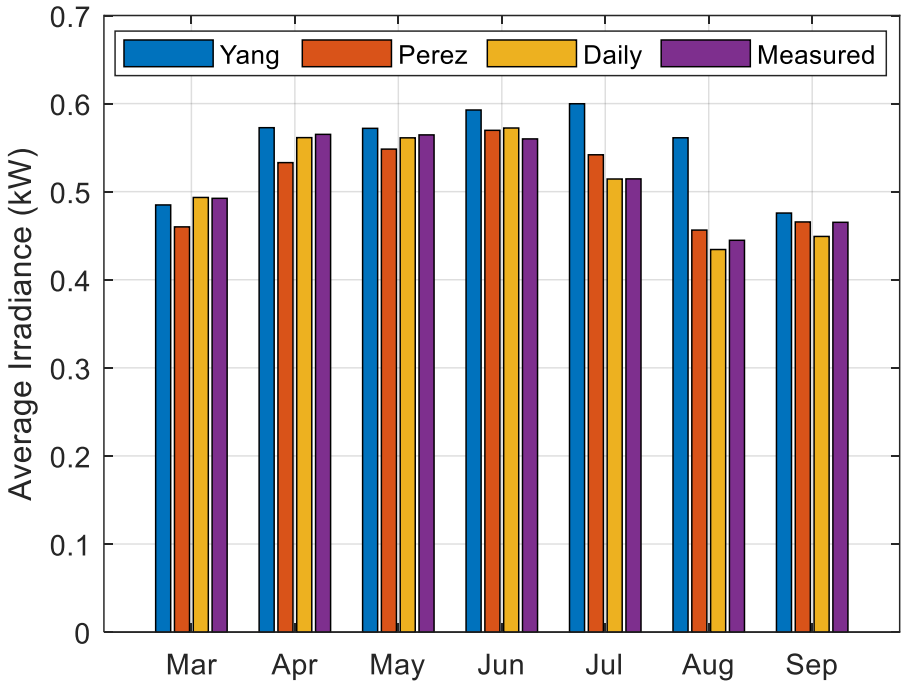

Figure 3.46. Cluster level monthly averaged irradiance forecast for the test period.

A simple linear upscaling model is found to be sufficient in this chapter. The physical model for power simulation is recommended when input data are available and reliable. Research efforts should be directed at making distributed solar system data readily available with improved reliability and easy access by adding data loggers and real-time communication devices during installation.

The output of the proposed model hold significance for the efficient operation of distribution networks and the optimization of dispatch scheduling. This model also carries potential utility for microgrids or independent electricity networks, encompassing residential units equipped with electric loads and photovoltaic (PV) generation capabilities, aiding in their planning endeavors. It is crucial to emphasize that the scope of the problem tackled within this subsection remains confined to situations where all pertinent PV systems are interconnected within the same distribution network. Without this shared network context, the issue and the solution become

inconsequential. The case study PV systems examined herein are geographically dispersed across California, as illustrated in Figure 3.34. Despite this geographical distribution implying connections to different distribution networks, the selection pool for case study systems was notably restricted due to the scarcity of BTM systems satisfying the stipulated criteria on public webpages. Consequently, our selection was dictated by this challenge in data availability. Moreover, it is worth noting that the extensive geographical dispersion of the systems primarily influences the model results through the mechanism of aggregation. However, research indicates that the impact of aggregation becomes minimal beyond a certain distance, thereby mitigating this effect's relevance in the context of the study. Information on regional scale PV power generation is also useful for the net-load forecasting since the net-load is the difference between the forecasted load and renewable energy generation.

3.7 References

[1] S. Leva, A. Nespoli, S. Pretto, M. Mussetta, and E. G. C. Ogliari, "PV plant power nowcasting: A real case comparative study with an open access dataset," *IEEE Access*, vol. 8, pp. 194428–194440, 2020, doi: 10.1109/ACCESS.2020.3031439.

[2] X. Zhang, Y. Li, S. Lu, H. F. Hamann, B. M. Hodge, and B. Lehman, "A Solar Time Based Analog Ensemble Method for Regional Solar Power Forecasting," *IEEE Trans. Sustain. Energy*, vol. 10, no. 1, pp. 268–279, 2019, doi: 10.1109/TSTE.2018.2832634.

[3] Y. Qu, J. Xu, Y. Sun, and D. Liu, "A temporal distributed hybrid deep learning model for day-ahead distributed PV power forecasting," *Appl. Energy*, vol. 304, no. September, p. 117704, 2021, doi: 10.1016/j.apenergy.2021.117704.

[4] E. Lorenz, J. Hurka, D. Heinemann, and H. G. Beyer, "Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 2, no. 1, pp. 2–10, 2009, doi: 10.1109/JSTARS.2009.2020300.

[5] H. Shaker, H. Zareipour, and D. Wood, "A data-driven approach for estimating the power generation of invisible solar sites," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2466–2476, 2016, doi: 10.1109/TSG.2015.2502140.

[6] I. Parvez, A. Sarwat, A. Debnath, T. Olowu, M. G. Dastgir, and H. Riggs, "Multi-layer

Perceptron based Photovoltaic Forecasting for Rooftop PV Applications in Smart Grid," *Conf. Proc. - IEEE SOUTHEASTCON*, vol. 2020-March, 2020, doi: 10.1109/SoutheastCon44009.2020.9249681.

[7] H. Shaker, H. Zareipour, and D. Wood, "Estimating Power Generation of Invisible Solar Sites Using Publicly Available Data," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2456–2465, 2016, doi: 10.1109/TSG.2016.2533164.

[8] S. Theocharides, G. Makrides, A. Livera, M. Theristis, P. Kaimakis, and G. E. Georghiou, "Day-ahead photovoltaic power production forecasting methodology based on machine learning and statistical post-processing," *Appl. Energy*, vol. 268, no. April, p. 115023, 2020, doi: 10.1016/j.apenergy.2020.115023.

[9] M. Sun, C. Feng, and J. Zhang, "Probabilistic solar power forecasting based on weather scenario generation," *Appl. Energy*, vol. 266, no. March, p. 114823, 2020, doi: 10.1016/j.apenergy.2020.114823.

[10] J. M. Morales, A. J. Conejo, H. Madsen, P. Pinson, and M. Zugno, *Integrating Renewables in Electricity Markets*. New York: Springer, 2014. doi: 10.1007/978-1-4614-9411-9.

[11] T. Stoffel, "Terms and Definitions," in *Solar Energy Forecasting and Resource Assessment*, Elsevier, 2013, pp. 1–19. doi: 10.1016/B978-0-12-397177-7.00001-2.

[12] J. Badosa, M. Haeffelin, N. Kalecinski, F. Bonnardot, and G. Jumaux, "Reliability of day-ahead solar irradiance forecasts on Reunion Island depending on synoptic wind and humidity conditions," *Sol. Energy*, vol. 115, pp. 306–321, 2015, doi: 10.1016/j.solener.2015.02.039.

[13] P. E. Thornton, H. Hasenauer, and M. A. White, "Simultaneous estimation of daily solar radiation and humidity from observed temperature and precipitation: An application over complex terrain in Austria," *Agric. For. Meteorol.*, vol. 104, no. 4, pp. 255–271, 2000, doi: 10.1016/S0168-1923(00)00170-2.

[14] W. Kong, Y. Jia, Z. Y. Dong, K. Meng, and S. Chai, "Hybrid approaches based on deep whole-sky-image learning to photovoltaic generation forecasting," *Appl. Energy*, vol. 280, no. September, p. 115875, 2020, doi: 10.1016/j.apenergy.2020.115875.

[15] R. Marquez and C. F. M. Coimbra, "Forecasting of global and direct solar irradiance using

stochastic learning methods , ground experiments and the NWS database," *Sol. Energy*, vol. 85, no. 5, pp. 746–756, 2011, doi: 10.1016/j.solener.2011.01.007.

[16] H. Sangrody, N. Zhou, and Z. Zhang, "Similarity-Based Models for Day-Ahead Solar PV Generation Forecasting," *IEEE Access*, vol. 8, pp. 104469–104478, 2020, doi: 10.1109/ACCESS.2020.2999903.

[17] X. Qing and Y. Niu, "Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM," *Energy*, vol. 148, pp. 461–468, 2018, doi: 10.1016/j.energy.2018.01.177.

[18] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker, D. Renné, and T. E. Hoff, "Validation of short and medium term operational solar radiation forecasts in the US," *Sol. Energy*, vol. 84, no. 12, pp. 2161–2172, Dec. 2010, doi: 10.1016/j.solener.2010.08.014.

[19] R. Perez, K. Moore, S. Wilcox, D. Renné, and A. Zelenka, "Forecasting solar radiation - Preliminary evaluation of an approach based upon the national forecast database," *Sol. Energy*, vol. 81, no. 6, pp. 809–812, Jun. 2007, doi: 10.1016/j.solener.2006.09.009.

[20] F. Kasten and G. Czeplak, "Solar and terrestrial radiation dependent on the amount and type of cloud," *Sol. Energy*, vol. 24, no. 2, pp. 177–189, 1980, doi: 10.1016/0038-092X(80)90391-6.

[21] K. Yang and T. Koike, "Estimating surface solar radiation from upper-air humidity," *Sol. Energy*, vol. 72, no. 2, pp. 177–186, 2002, doi: 10.1016/S0038-092X(01)00084-6.

[22] A. Yona, T. Senjyu, T. Funabashi, and C. H. Kim, "Determination method of insolation prediction with fuzzy and applying neural network for long-term ahead PV power output correction," *IEEE Trans. Sustain. Energy*, vol. 4, no. 2, pp. 527–533, 2013, doi: 10.1109/TSTE.2013.2246591.

[23] Z. Si, M. Yang, Y. Yu, and T. Ding, "Photovoltaic power forecast based on satellite images considering effects of solar position," *Appl. Energy*, vol. 302, no. May, p. 117514, 2021, doi: 10.1016/j.apenergy.2021.117514.

[24] G. Moreno, P. Martin, C. Santos, F. J. Rodríguez, and E. Santiso, "A day-ahead irradiance forecasting strategy for the integration of photovoltaic systems in virtual power plants," *IEEE Access*, vol. 8, pp. 204226–204240, 2020, doi: 10.1109/ACCESS.2020.3036140.

[25] E. Lorenz, T. Scheidsteger, J. Hurka, D. Heinemann, and C. Kurz, "Regional PV power prediction for improved grid integration," *Prog. Photovoltaics Res. Appl.*, vol. 19, no. 7, pp. 757–771, Nov. 2011, doi: https://doi.org/10.1002/pip.1033.

[26] V. E. Larson, "Forecasting Solar Irradiance with Numerical Weather Prediction Models," in *Solar Energy Forecasting and Resource Assessment*, Elsevier, 2013, pp. 299–318. doi: 10.1016/B978-0-12-397177-7.00012-7.

[27] "NSRDB." https://nsrdb.nrel.gov/data-viewer/query/query-tools (accessed Aug. 12, 2022).

[28] P. S. Kumar, V. P. Sakthivel, M. Raju, and P. D. Sathya, "A Comprehensive Review on Deep Learning Algorithms and its Applications," *Proc. 2nd Int. Conf. Electron. Sustain. Commun. Syst. ICESC 2021*, pp. 1378–1385, 2021, doi: 10.1109/ICESC51422.2021.9532767.

[29] M. R. Chen, G. Q. Zeng, K. Di Lu, and J. Weng, "A Two-Layer Nonlinear Combination Method for Short-Term Wind Speed Prediction Based on ELM, ENN, and LSTM," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6997–7010, 2019, doi: 10.1109/JIOT.2019.2913176.

[30] F. Zhao, G. Q. Zeng, and K. Di Lu, "EnLSTM-WPEO: Short-term traffic flow prediction by ensemble LSTM, NNCT weight integration, and population extremal optimization," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 101–113, 2020, doi: 10.1109/TVT.2019.2952605.

[31] N. K. Chauhan and K. Singh, "A review on conventional machine learning vs deep learning," *2018 Int. Conf. Comput. Power Commun. Technol. GUCON 2018*, pp. 347–352, 2019, doi: 10.1109/GUCON.2018.8675097.

[32] M. J. Reno and C. W. Hansen, "Identification of periods of clear sky irradiance in time series of GHI measurements," *Renew. Energy*, vol. 90, pp. 520–531, 2016, doi: 10.1016/j.renene.2015.12.031.

[33] L. Gigoni *et al.*, "Day-Ahead Hourly Forecasting of Power Generation from Photovoltaic Plants," *IEEE Trans. Sustain. Energy*, vol. 9, no. 2, pp. 831–842, 2018, doi: 10.1109/TSTE.2017.2762435.

[34] "PV education", [Online]. Available: https://www.pveducation.org/pvcdrom/properties-of-sunlight/elevation-angle

[35] J. Badosa, M. Haeffelin, and H. Chepfer, “Scales of spatial and temporal variation of solar irradiance on Reunion tropical island,” *Sol. Energy*, vol. 88, pp. 42–56, 2013, doi: 10.1016/j.solener.2012.11.007.

[36] S. Al-Dahidi, M. Louzazni, and N. Omran, “A Local Training Strategy-Based Artificial Neural Network for Predicting the Power Production of Solar Photovoltaic Systems,” *IEEE Access*, vol. 8, pp. 150262–150281, 2020, doi: 10.1109/ACCESS.2020.3016165.

[37] Y. Li, Y. Su, and L. Shu, “An ARMAX model for forecasting the power output of a grid connected photovoltaic system,” *Renew. Energy*, vol. 66, pp. 78–89, 2014, doi: 10.1016/j.renene.2013.11.067.

[38] S. Atique, S. Noureen, V. Roy, S. Bayne, and J. MacFie, “Time series forecasting of total daily solar energy generation: A comparative analysis between ARIMA and machine learning techniques,” *IEEE Green Technol. Conf.*, vol. 2020-April, pp. 175–180, 2020, doi: 10.1109/GreenTech46478.2020.9289796.

[39] A. Nespoli *et al.*, “Day-ahead photovoltaic forecasting: A comparison of the most effective techniques,” *Energies*, vol. 12, no. 9, pp. 1–15, 2019, doi: 10.3390/en12091621.

[40] D. Hunter, H. Yu, M. S. Pukish, J. Kolbusz, and B. M. Wilamowski, “Selection of proper neural network sizes and architectures-A comparative study,” *IEEE Trans. Ind. Informatics*, vol. 8, no. 2, pp. 228–240, 2012, doi: 10.1109/TII.2012.2187914.

[41] K. G. Sheela and S. N. Deepa, “Review on Methods to Fix Number of Hidden Neurons in Neural Networks,” *Math. Probl. Eng.*, vol. 2013, p. 425740, 2013, doi: 10.1155/2013/425740.

[42] S. Mishra, R. Prusty, and P. K. Hota, “Analysis of Levenberg-Marquardt and Scaled Conjugate gradient training algorithms for artificial neural network based LS and MMSE estimated channel equalizers,” *Proc. - 2015 Int. Conf. Man Mach. Interfacing, MAMI 2015*, no. Lm, 2016, doi: 10.1109/MAMI.2015.7456617.

[43] 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 J. Photovoltaics*, vol. 5, no. 2, pp. 597–606, 2015, doi: 10.1109/JPHOTOV.2015.2392938.

[44] S. Killinger, N. Engerer, and B. Müller, “QCPV: A quality control algorithm for distributed photovoltaic array power output,” *Sol. Energy*, vol. 143, pp. 120–131, 2017, doi: 10.1016/j.solener.2016.12.053.

[45] M. M. Khan, M. J. Ahmad, and B. Jamil, “Development of Models for the Estimation of Global Solar Radiation Over Selected Stations in India,” in *Energy, Transportation and Global Warming*, P. Grammelis, Ed. Cham: Springer International Publishing, 2016, pp. 149–160. doi: 10.1007/978-3-319-30127-3_13.

[46] N. Omar, H. Aly, and T. Little, “Optimized Feature Selection Based on a Least-Redundant and Highest-Relevant Framework for a Solar Irradiance Forecasting Model,” *IEEE Access*, vol. 10, pp. 48643–48659, 2022, doi: 10.1109/ACCESS.2022.3171230.

[47] M. J. Mayer, “Impact of the tilt angle, inverter sizing factor and row spacing on the photovoltaic power forecast accuracy,” *Appl. Energy*, vol. 323, p. 119598, Oct. 2022, doi: 10.1016/J.APENERGY.2022.119598.

[48] C. Tofallis, “A better measure of relative prediction accuracy for model selection and model estimation,” *J. Oper. Res. Soc.*, vol. 66, no. 8, pp. 1352–1362, 2015, doi: 10.1057/jors.2014.103.

[49] D. A. R. Barkhouse, O. Gunawan, T. Gokmen, T. K. Todorov, and D. B. Mitzi, “Yield predictions for photovoltaic power plants:empirical validation,recent advances and remaining uncertainties,” *Prog. Photovoltaics Res. Appl.*, vol. 20, no. 1, pp. 6–11, 2015, doi: 10.1002/pip.

[50] Y. Ue, R. Hara, H. Kita, Y. Saito, K. Takitani, and M. Saito, “Analysis of spatial smoothing effect of short-term fluctuation of global solar radiation based on data including quantization error,” *Electr. Eng. Japan (English Transl. Denki Gakkai Ronbunshi)*, vol. 180, no. 3, pp. 55–63, 2012, doi: 10.1002/eej.21274.

4 SOLAR POWER RAMP EVENTS DETECTION USING A MODIFIED DATA COMPRESSION TECHNIQUE

4.1 Introduction

One of the major challenges with the integration of solar photovoltaic (PV) energy sources into electric grids stems from the frequent occurrences of short-term, high magnitude surges (ramp up) and drops (ramp down) in their output [1]. In an electric grid with a high penetration of solar PV energy sources, the variability occurring within a short time horizon of a few seconds can cause voltage flickering while sudden changes that last for elongated time horizons of minutes can culminate into frequency problems for the grid [2],[3]. Sudden and high magnitude ramp events in the intervals of 5-25 minutes are of greater concerns to power system operators [4] as they would necessitate the availability of sufficient spinning reserves with high ramp rates to contain the power deficit or surplus [5],[6].

Analysis from an economic standpoint show that a substantial reduction in the levelized cost of energy can be obtained by combining optimized energy storage sizing methodologies and ramp events forecasting and characterization techniques to deal with the renewables intermittency problem [1],[7]. In compensating for the changes in power output from PV sources, energy storage capacity or spinning reserve requirements are decided based on the magnitude of the largest ramps [8],[9]. Furthermore, decisions on the required charging and discharging times are informed by metrics like the frequency distribution and duration of ramping events giving relevance to the accurate forecasting and quantification of these ramps [10],[11].

Ramp event detection can also provide training datasets of ramp events for forecasting applications [12]–[14]. Ramp characterization metrics including the count, magnitude and duration on a seasonal and daily basis from observed patterns in historical forecasts databases can be used to build sample sets of ramp events useful for post-processing forecasting methods [15], [16]. This rationale is based on the grounds that acquiring comprehensive information about the previous and current situations will make for better and more accurate forecast of future scenarios [17]. Also, conventional forecast methods accuracy indicators are either scale-dependent or the percentage error. These statistical measures only describe the overall systematic error since in simplistic terms, they are averages of the deviation of the forecasts from actual values. These measures do not

suffice as metrics for automatically detailing how well forecast models capture ramping events that are more problematic to utility operators [18].

In characterizing ramping events, a model should be driven towards evaluating the frequency of occurrence within varying time intervals as well as the duration of ramp events especially those whose magnitudes fall within the classification of critical ramp events. Another area of interest is the distribution of these ramp events including the seasonal and daily patterns of occurrences. In this chapter, we proposed a model that characterizes Solar Power Ramp Events (SPREs) based on the above metrics.

A diverse range of studies have been carried out on ramp detection methods. At present, there is limited literature on Wind Power Ramp Events (WPREs) [4], [15], [17], [19]–[24] and SPREs detection [4], [7], [19], [25]–[27]. Although any typical WPREs detection algorithms would similarly detect SPREs, these algorithms fall short in their performance since they do not take into account the peculiarity of the daily variability pattern of solar power. Consequently, without necessary adjustments, a detection algorithm solely designed for WPREs, if implemented on solar power data would inaccurately detect SPREs.

Sevlian and Rajagopal [22] developed the L1-SW algorithm that used a dynamic programming recursion method to merge adjacent segments for WPREs detection. Leveraging on the auxiliary ramp rules established in the L1-SW method, Lyners et al. [21] proposed a multi-parameter segmentation algorithm for WPRE extraction. Cui et al. [24] developed a generalized Gaussian Mixture Model to characterize WPREs probability distributions. Qu et al. [15] proposed a Parameter and Resolution Adaptive Algorithm (PRAA) to address issues of bad data and computation speed in WPREs detection. Cui et al. [17] combined the advantages of the sliding window (SW) from [22] with the Swinging Door Algorithm (SDA) for ramp characterization. Hossain et al. [8] and Madhab et al. [26] carried out statistical ramp analysis from which they proposed empirical models for estimating SPREs distribution. Willy et al. [27] used the dead band data compression method for SPREs detection. An Optimized Swinging Door Algorithm (OpSDA) was proposed by Cui et al. [19], [20], [25] for WPREs and SPREs detection. The algorithm which was proposed as an improvement of the SDA, involved ramp trends identification and adjacent segments combination using dynamic programing.

The following are some of the drawbacks in existing literature with respect to SPREs detection that need to be addressed:

1. In deciding an algorithm suitable for SPREs detection, we observed that most of the existing propositions share a common problem of computational complexity and as such, leave room for further improvements of the merging process. SPREs identification in an automated manner requires the process to be computationally simple and inexpensive to allow for online applications [4].
2. Other existing algorithms that could be considered simple to implement were tailored for WPREs and/or offered no definitive and consistent procedure for estimating the value of ε.
3. The dynamic programming in the OpSDA is a slight adjustment to the L1-SW method and the robustness of both methods are considerably impacted by the tunable or penalty parameter (tolerance). In both methods, the tolerance was decided by mere visual inspection or user experience which in other words, is simply trial and error. A suboptimal tolerance value will jeopardize their detection accuracy with respect to reproducibility and repeatability across different timeseries and databases with different time resolutions.

We therefore propose a relatively more efficient algorithm for SPREs detection in this chapter. Our objective is to develop an algorithm to characterize and analyze significant ramps events (SREs) in solar power measurements using parameters including ramp distribution, duration, and magnitude. The phrases ramp extraction, ramp characterization and ramp detection are synonymously used in this chapter.

The main contributions of this chapter include:

1. A novel SPREs detection algorithm with improved efficiency, accuracy and applicability regardless of the sampling time horizon is proposed.
2. Compared with the OpSDA, the proposed ramp detection with computational simplicity and lower execution time enables online applications. The proposed model performance is also less impacted by the SDA tolerance value. In existing literature, the OpSDA provides a more comprehensive account of SPREs detection and also a better performance than the

L1-SW or any other currently available SPREs detection algorithm, hence its choice as a benchmark in this study. The SDA is superior to the arch chord and dead band models in [27] and also other non-SDA-based models.

3. An expression for estimating an optimal value of the SDA tolerance for our model is proposed to ensure capturing the maximum possible ramp events with more accurate start- and stop-times.

The rest of this chapter takes the following structure: the development of the MSDA is laid out in Section 4.2; the experimental results compared with the OpSDA [25] and SDA are discussed in Section 4.3; and the conclusion is drawn in Section 4.5.

4.2 Modified Swinging Door Algorithm

4.2.1 Swinging Door Algorithm

Florita et al. first proposed using the SDA for WPREs and SPREs extraction in [4]. The SDA was formulated for data compression because of its structural simplicity, robustness and low computational requirements [4],[17]. The algorithm identifies ramps in a piecewise linear approximation manner by adopting a tunable threshold parameter (ε) called door width to create segments of the raw data as shown in Figure 4.1.

Only the initial and final data points of a segment are retained in Figure 4.1. The final point of a segment becomes the initial point for the next segment. In Segment 1, the doors ($\overline{3A}, \overline{7A}, \overline{3B}, \overline{7B}, \overline{3C}, \overline{3D}$) iteratively swing (lines drawn) from the pivots (points 3 and 7 on the vertical axis) to accommodate new data points. At the data point where the doors are parallel to each other or where the doors have over 180^0 interior angle, a new segment begins (at point D in Figure 4.1). Mathematically, this situation implies that $y_{lb} < y_c < y_{ub}$ is violated.
where:

y represents the signal magnitude,

t is the time,

the subscripts ub and lb are the upper and lower boundaries of the current signal point respectively,

c is the current signal point,

g is the gate point (point B in Figure 4.1) and

s is the start point of a segment.

$$y_{ub} = y_s + \left(\frac{y_g + \varepsilon - y_s}{t_g - t_s}\right)(t_c - t_s) \quad (1)$$

$$y_{lb} = y_s + \left(\frac{y_g - \varepsilon - y_s}{t_g - t_s}\right)(t_c - t_s) \quad (2)$$

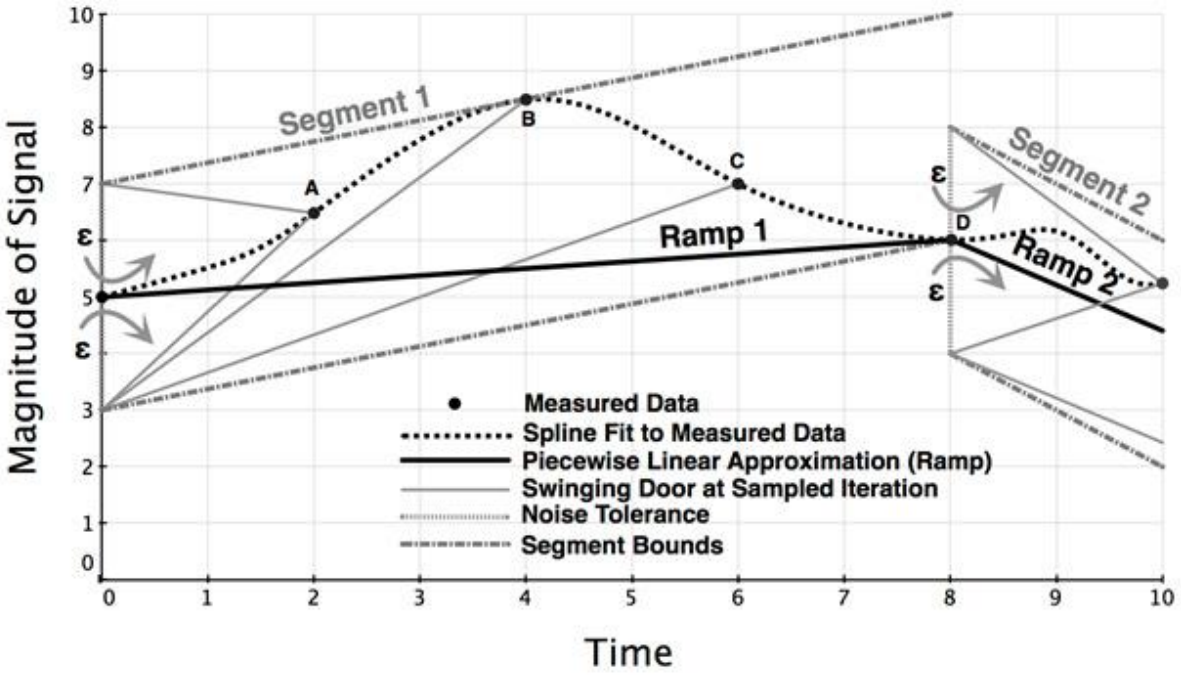

Figure 4.1. Schematic description of the swinging door algorithm [4]

The accuracy of the SDA is heavily influenced by the door width (ε) predefined by the user [17]. A static ε creates redundant segmentation points and inaccurate detection of ramp start- and stop-times and invariably ramp magnitude as shown in Figure 4.2. The ideal situation in Figure 4.2. beginning from point A should be a total of four significant SPREs ($\overline{AB}, \overline{BC}, \overline{CD}$ and $\overline{DE}$). Minimizing the problem of ramp segments redundancy by increasing the tolerance ε may mean ignoring significant inflection points and result in the identification of only a few large ramps when the SDA becomes under constrained [4]. The SDA therefore requires modifications to improve its suitability for SPREs detection and characterization.

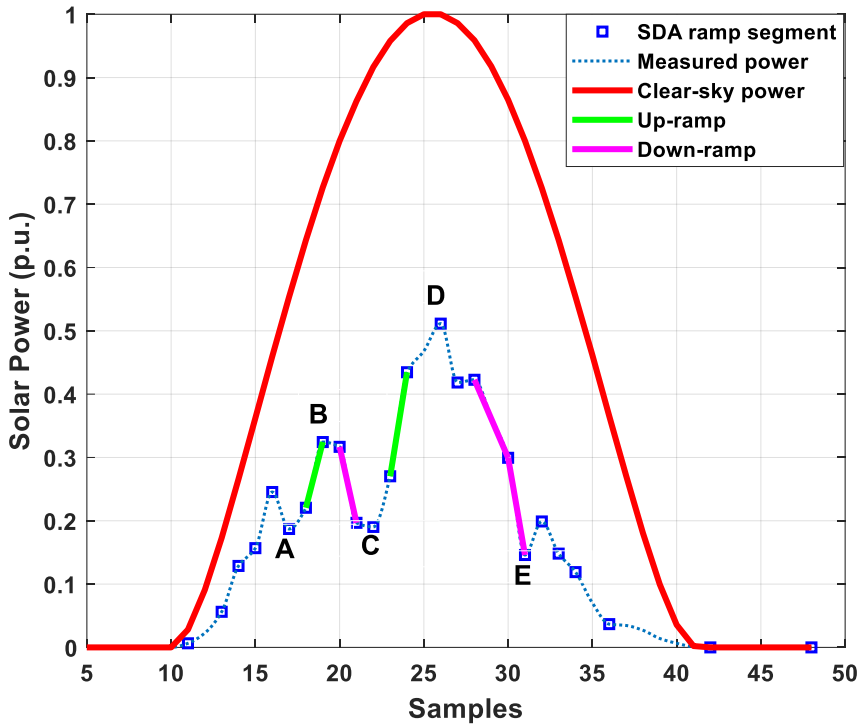

Figure 4.2. Application of the SDA for PV power ramps extraction on a 30-min resolution data.

4.2.2 Modified Swinging Door Algorithm Formulation

Our proposed algorithm simplifies the SPREs detection process by merging adjacent ramp segments created using the SDA. The basic principle of the MSDA is to merge redundant ramp segments created by the SDA into single ramp events with larger magnitudes and longer durations by checking for the occurrences and sizes of inflections in the SDA-segmented data. We carried out a preliminary study on a variety of measured PV power from geographically diverse systems and recording time horizons to understand the uniqueness of solar power variability patterns. The algorithm is then configured according to the observed patterns for SPREs detection using empirical and statistical tests. The algorithm is also adjusted to factor into consideration the impact of time resolution due to our observation that depending on cloud speed and orientation, the magnitude of individual solar ramps between adjacent time steps resonates inversely with the time resolution. A flowchart of the whole detection process is shown in Figure 4.3.

4.2.3 Significant Ramp Definitions

The algorithm is developed in this chapter to extract only significant SPREs by definition.

Since there is no unified or standard definition for significant SPREs in the literature, the three definitions (3) – (6) detailed in [25] are adopted in this study.

Definition 1: Solar power variation greater than 10% of P_G.

$$\left|P_j - P_k\right| > 0.1P_G \tag{3}$$

Definition 2: Solar power variation greater than 10% of P_G occurring within an hour or less.

$$\left|P_j - P_k\right| > 0.1P_G,\ t_j - t_k \leq 1\ hour \tag{4}$$

Definition 3(a): Solar power increment or up-ramp greater than 10% of P_G occurring within an hour or less.

$$P_k - P_j > 0.1P_G,\ t_j - t_k \leq 1\ hour \tag{5}$$

Definition 3(b): Solar power decrement or down-ramp greater than 8% of P_G occurring within an hour or less.

$$P_k - P_j < -0.08P_G,\ t_j - t_k \leq 1\ hour \tag{6}$$

where:

P_G is the installed capacity of the PV system,

P_j represents the instantaneous power at the beginning of a ramp segment,

P_k represents the instantaneous power at the end of a ramp segment,

t_j is the ramp segment start time,

t_k is the ramp segment stop time.

4.2.4 Natural Variability Index (NVI)

There are no generally accepted methods for optimally determining the tunable parameter of the SDA. In deciding a procedure for estimating the tolerance, our priority is to make the SDA door width small enough to optimally capture even small twists and turns in the input data. Although smaller tolerance values imply that the SDA becomes more constrained, this scenario presents us with the advantage of more accurate detection of ramp start- and stop-times after

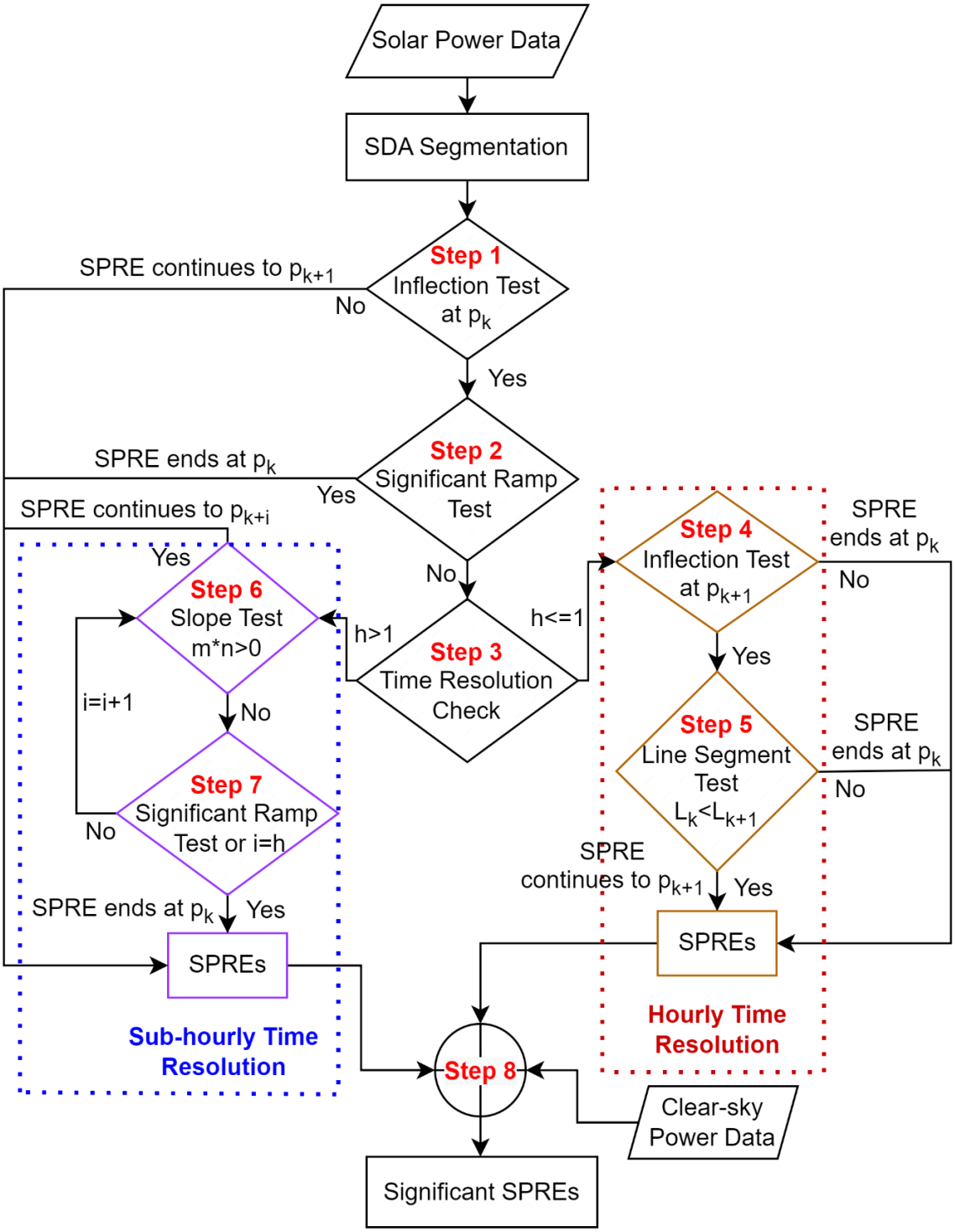

Figure 4.3. Flowchart of the overall solar power ramp events detection model proposed in this chapter.

implementing the merging of adjacent ramp segments using the MSDA. Upon sufficient verifications, we propose using a variable called the Natural Variability Index (NVI) [29] to

approximate an optimal value of ε for our model. Although primarily intended to be an irradiance variability pattern classification tool [30], it is found that the NVI is dynamic and consistent across datasets with different timescales and from different geographical locations in estimating ε.

$$NVI = \frac{\sigma_{\Delta P}}{\mu_P} \tag{7}$$

$$\sigma_{\Delta P} = \sqrt{\frac{1}{k-1}\sum_{i=1}^{k}(\Delta P_i - \Delta\bar{P})^2} \tag{8}$$

where:

μ_P is the arithmetic mean of PV power.

ΔP_i is the i-th data point of power step change.

$\Delta\bar{P}$ is the arithmetic mean of power step change and

k is the solar power data size.

4.2.5 MSDA Ramp Segments Merging Process

Consider a solar power time series $Y = \{(p_1, t_1), \dots, (p_i, t_i), \dots (p_N, t_N)\}$, $i \in N$ where p_i is the solar power and t_i is the equivalent time stamp. Implement the SDA to extract M number of data points of Y: $P = \{P_1, \dots, P_k, \dots, P_M\}$, $(k = 1, \dots, M)$, $M \leq N - 1, P_k = (p_k, t_k)$ as segments $S_k = \overline{p_k p_{k+1}}$. The segments are then merged based on changing ramp directions and magnitude of the differences between two data points: $S = \{X_1, \dots, X_j, \dots, X_T\}$, $(T \leq m)$, $X_j = (s_j, e_j)$, s_j and e_j representing the start- and end-points of the jth ramp event respectively. T represents the total number of detected SPREs. For the period $1 < k \leq M$, merge adjacent ramp segments using steps 1 – 7 after saving s_1 as the first data point. After the merging process, significant SPREs are sorted and extracted in Step 8.

Step 1. Check for an inflection at p_k using (9). If there's no inflection, i.e. $\frac{p_{k+1}-p_k}{p_k-p_{k-1}} \geq 0$, set $e_j = k + 1$. This means that the ramp event from $S_{k-1}$is assumed to continue into S_k thereby merging segments S_{k-1} with S_k. If $p_k = 0$, set $s_{j+1} = k$, a new ramp event begins at p_k. A

ramp event will always be terminated at any point power generation p_k drops to zero. If there is an inflection, $\frac{p_{k+1}-p_k}{p_k-p_{k-1}} < 0$ and $p_k \neq 0$, go to Step 2.

$$\frac{p_{k+1}-p_k}{p_k-p_{k-1}} <\geq 0 \tag{9}$$

Step 2. Compare the inflection size with the magnitude of the adopted significant ramp definition (D) using (10). If $|p_{k+1}-p_k| \geq D$, set $s_{j+1} = k$. The point of inflection p_k becomes the start point of a new ramp event. Otherwise, i.e. $|p_{k+1}-p_k| < D$, go to Step 3.

$$|p_{k+1}-p_k| <\geq D \tag{10}$$

Step 3. Calculate the value of h where $h = 60/R$. If $h \leq 1$, i.e., hourly resolution, go to the next step, otherwise (sub-hourly resolution), go to Step 6. R is the time resolution of Y in minutes.

Step 4. Check for a subsequent inflection at point p_{k+1}. If $\frac{p_{k+2}-p_{k+1}}{p_{k+1}-p_k} \geq 0$, set $s_{j+1} = k$ and save p_k to begin a new ramp event. If $\frac{p_{k+2}-p_{k+1}}{p_{k+1}-p_k} < 0$, go to Step 5.

Step 5. Calculate the lengths of lines L_k and L_{k+1} connecting the segments $\overline{p_k p_{k+1}}$ and $\overline{p_{k+1} p_{k+2}}$ respectively, using (11). If $L_k \geq L_{k+1}$, set $s_{j+1} = k$, a new ramp begins. If $L_k < L_{k+1}$, set $e_j = k+2$ signifying the ramp event from segment S_{k-1} continues to segment S_{k+1}. Go to Step 8.

$$L_k = \sqrt{(p_{k+1}-p_k)^2 + (t_{k+1}-t_k)^2} \tag{11}$$

Step 6. Calculate m_k and $n_{k+i|\ i=2}$ using (12) – (13). If $m \times n > 0$, set $e_j = k+i$, otherwise, go to Step 7.

$$m = \frac{p_k - p_{k-1}}{t_k - t_{k-1}} \tag{12}$$

$$n = \frac{p_{k+i} - p_k}{t_{k+i} - t_k} \tag{13}$$

Step 7. Check if the magnitude of the significant ramp criteria is triggered using (14). If (14) holds, set $s_{j+1} = k$, otherwise, repeat Step 6 with $i = i+1$ until a condition is triggered or $i = h$ $(i.e\ n_{k+i|\ \forall i=2,\dots,h})$. If no condition is triggered, set $s_{j+1} = k$. Go to Step 8.

$$|p_{k+i} - p_k| \geq D \tag{14}$$

Step 8. Since the proposed method was intended only to extract the significant ramps that are cloud-induced, ramping events simultaneously occurring both in the measured and clear-sky input datasets and that satisfy the threshold condition in (15) were omitted from the analysis. This is because solar variabilities caused by diurnal sun movement from sunrise to sunset are relatively deterministic and can be estimated with near perfect accuracies. Consequently, these variabilities pose no problems and are of no interest in this chapter.

$$Significant\ ramp\text{:}\ \left\{max\left[p_{s_j}, p_{e_j}\right] - min\left[p_{s_j}, p_{e_j}\right]\right\} \in D \tag{15}$$

In situations where (15) did not hold because of the time constraint in ramp Definitions 2 and 3, the ramp segments $S_{k=n}$ occurring within the data range $\left\{max\left[p_{s_j}, p_{e_j}\right] - min\left[p_{s_j}, p_{e_j}\right]\right\}$ were subjected to a significant ramp test using (14) and reducing the range one segment $(n-1)$ each time (16) is not achieved. n is the number of segments within the data range of interest.

$$\left\{max\left[p_{k,s_j}, p_{k,e_j}\right] - min\left[p_{k,s_j}, p_{k,e_j}\right]\right\} \in D \tag{16}$$

Basically, the SDA is employed to create ramp segments (blue boxes in Figure 4.2) from a solar power data with a window or tolerance value estimated using the NVI metric. The segments are then fed into the proposed algorithm which optimally merges adjacent segments that are in the same ramp direction to produce ramp events with more accurate characterizations. Thereafter, ramp events that reach pre-established thresholds of ramp definitions occurring only in the measured or actual power timeseries are labeled as significant solar ramp events. The next section details the implementation of the proposed method using a case study.

4.3 Experimental Results

4.3.1 Case Study Dataset Description and Preparation

Simulated PV power from raw data files of measured solar radiation with their timestamps were used for validating the proposed ramp detection method. The data files were obtained from the National Renewable Energy Laboratory (NREL)'s Measurement and Instrumentation Data Centre (MIDC) – University of Oregon (SRML) site [31]. The datasets comprise of an hourly

resolution irradiance data with 8,760 entries and a 1-min resolution with 525,600 entries spanning a full calendar year period from January 1, 2018, to December 31, 2018. Corresponding clear-sky data for the same location, the timeframe and resolutions were either simulated or collected from NREL National Solar Radiation Database (NSRDB). Other lower data resolutions were interpolated from the 1-min resolution dataset.

For data preprocessing, physically impossible data entries were filtered out – Negative power values occurring at lower sun elevation angles were replaced with zero. Nighttime power readings were also filtered to zero using a zenith-angle-based filter. Thereafter, the datasets were normalized relative to the installed capacity value. The proposed algorithm is implemented in Spyder (Python 3.8) on a PC with an Intel Xeon E5-2660v4 with two 2.00GHz CPU and 64 GB RAM.

4.4 MSDA SPRE Detection Test Results

The metrics proposed by [25] is used to evaluate and compare the accuracy of ramp detection by our proposed model with the OpSDA model. The comparisons are centered on computing time and statistical analysis of ramp features with respect to ramp rate and distribution. Basic visual inspection of portions of the detected ramping events are also considered valuable to detailing the improvement of the status quo by our proposed model.

Figures 4.4 – 4.12. show typical examples of SPREs detection over a single day period for three resolutions. Again, a ramping event in the figures was designated significant using Definition 1 and if not simultaneously occurring in the clear-sky power component. Table 4.1 displays the sizes (kW) of the extracted ramps depicted in Figures 4.4 - 4.12. The MSDA method produced larger magnitudes ramps. Figures 4.4 – 4.6. compares the SPREs extracted by the SDA, OpSDA and MSDA from an hourly resolution data having only 24 samples. Consistent across all three methods, only one up ramp and one down-ramp are extracted as significant. However, the MSDA showed a better performance in Figure 4.6 because it produces more accurate ramp start-time for the down-ramp and end-time for the up ramp.

Inaccurate detection of ramp start- and stop-times could be dire as information on the duration and size of the ramp would be erroneous. Therefore, optimally, the down ramp in Figures 4.4 – 4.6 should begin at the 11th hour mark whereas the up ramp terminates at the 15th hour.

Table 4.1. Detected ramps in kW using ramp definition 1 as shown in Figures 4.4 - 4.12.

1-min Resolution						
	SDA		OPSDA		MSDA	
	Up Ramp	Down Ramp	Up Ramp	Down Ramp	Up Ramp	Down Ramp
				115.03	122.13	120.32
				102.3	193.22	101.1
					192.4	190.9
Count	0	0	0	2	3	3
5-min Resolution						
	SDA		OPSDA		MSDA	
	Up Ramp	Down Ramp	Up Ramp	Down Ramp	Up Ramp	Down Ramp
	111.39	110.923	111.39	110.923	178.06	136.477
	126.093	128.569	185.88	128.569	239.693	138.42
				115.046	117.08	119.23
Count	2	2	2	3	3	3
1-hr Resolution						
	SDA		OPSDA		MSDA	
	Up Ramp	Down Ramp	Up Ramp	Down Ramp	Up Ramp	Down Ramp
	129.799	249.481	129.799	249.481	193.159	257.545
Count	1	1	1	1	1	1

Similar to the observation in Figures 4.4 – 4.6, the MSDA also comparatively produced better results with respect to the detected ramps magnitude and duration on the left side of Figure 4.7 – 4.12. Due to higher resolutions, there is more data samples and in turn more ramp segments but with smaller magnitudes. SPREs detection from the 1-min resolution data in Figures 4.10 – 4.12 makes apparent some remarkable observations.

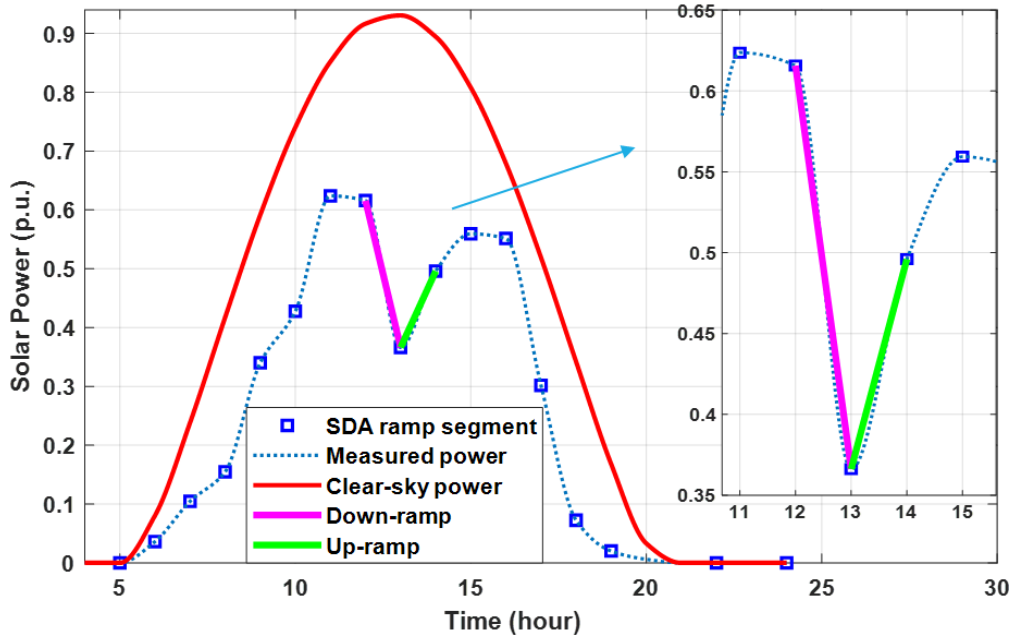

Figure 4.4. Visual comparison with insets of the SDA detection performance on a solar power data with hourly resolution and using the significant ramp Definition 1.

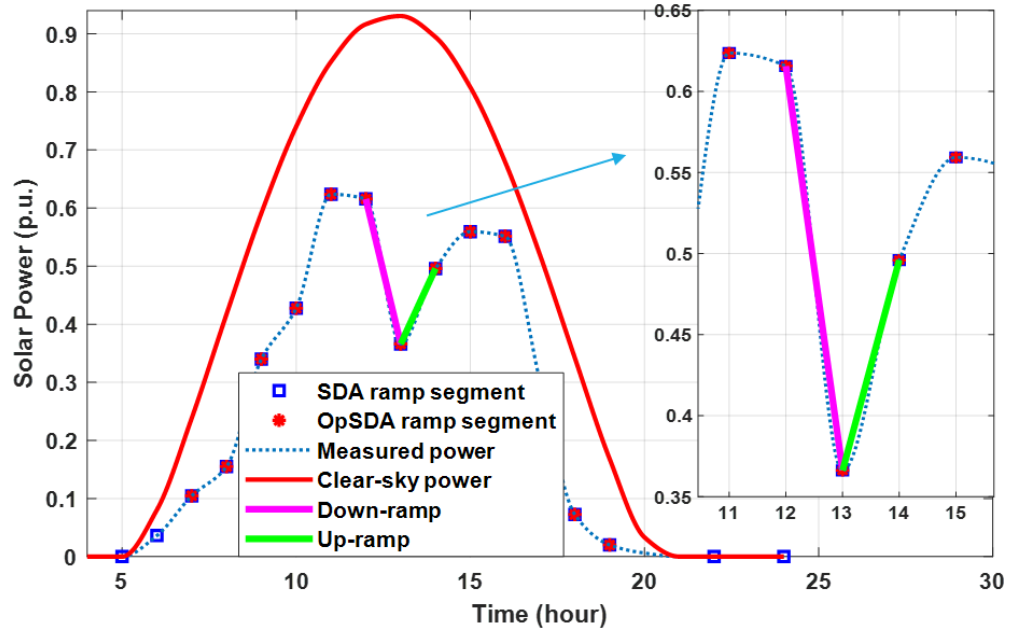

Figure 4.5. Visual comparison with insets of the OpSDA detection performance on a solar power data with hourly resolution and using the significant ramp Definition 1.

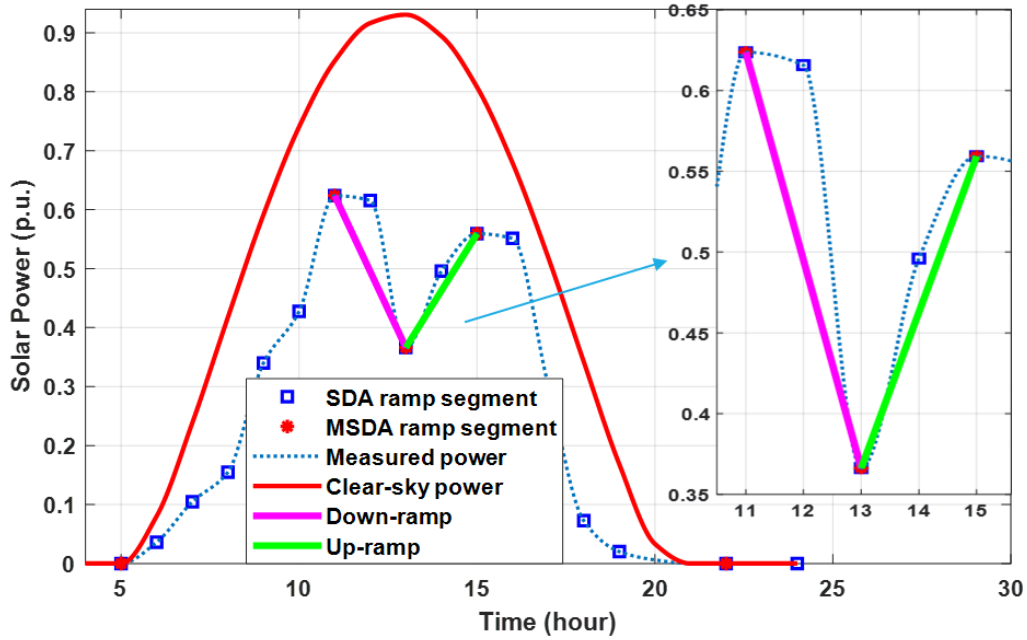

Figure 4.6. Visual comparison with insets of the MSDA detection performance on a solar power data with hourly resolution and using the significant ramp Definition 1.

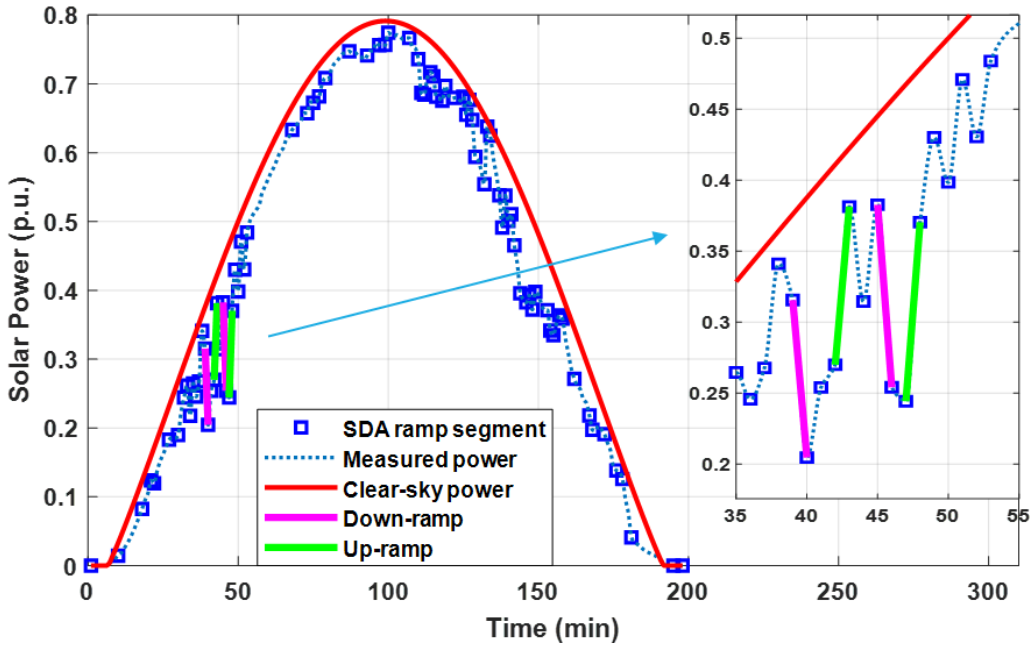

Figure 4.7. Visual comparison with insets of the SDA detection performance on a solar power data with 5-min resolution and using the significant ramp Definition 1.

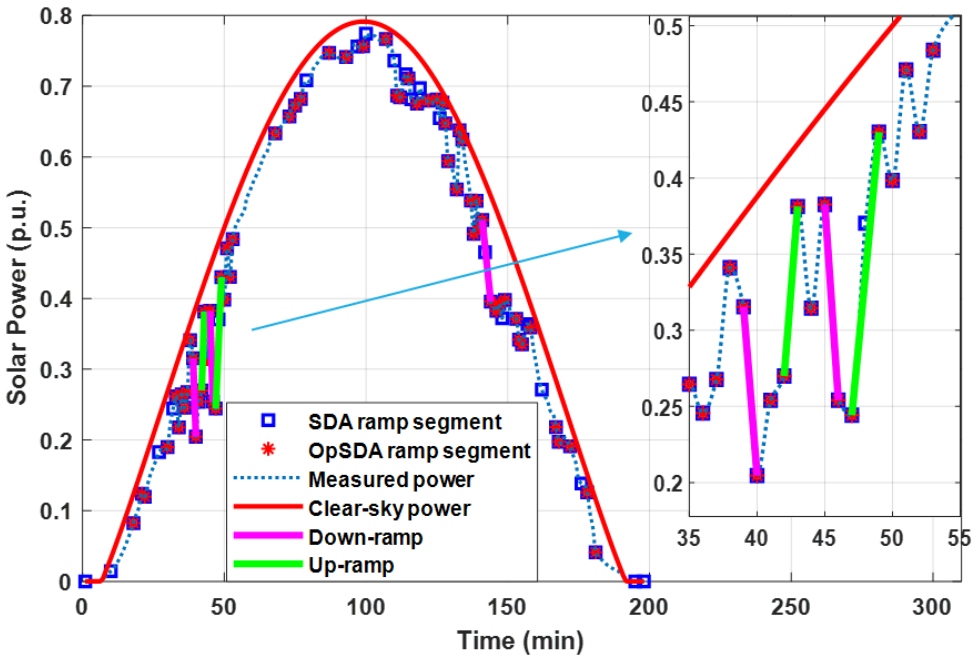

Figure 4.8. Visual comparison with insets of the OpSDA detection performance on a solar power data with 5-min resolution and using the significant ramp Definition 1.

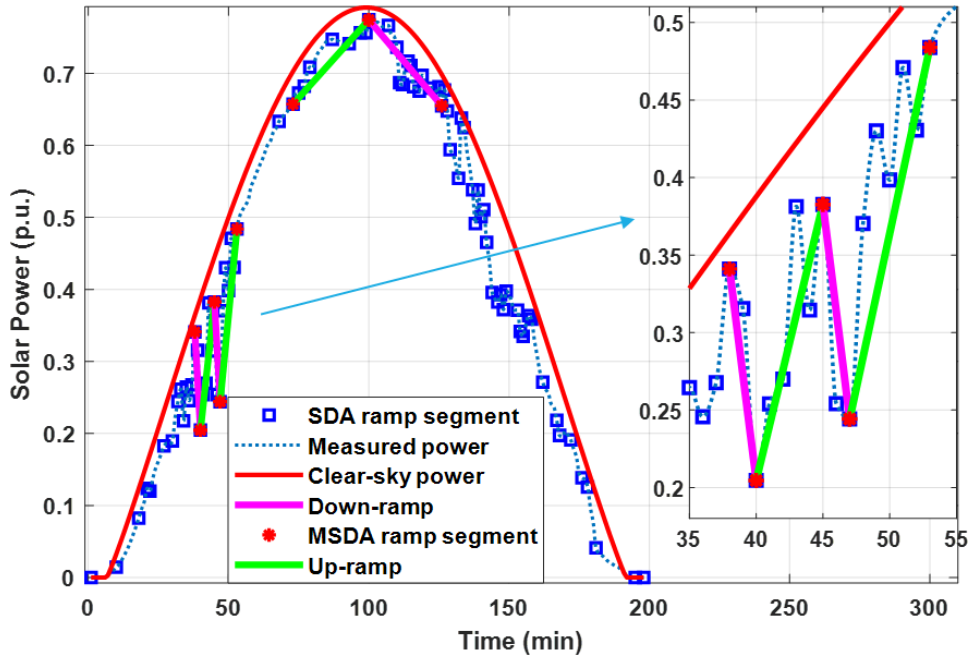

Figure 4.9. Visual comparison with insets of the MSDA detection performance on a solar power data with 5-min resolution and using the significant ramp Definition 1.

Although 523 SDA segments are created in Figure 4.7, the difference between any two adjacent segments is not large enough to be classified as a significant ramp event. Increasing the SDA tolerance value would be beneficial in reducing the number of segments and as a direct consequence, a possible detection of more SPREs. However, the SPREs would be sub-optimal – more detailed analysis on the sensitivity of ramp detection to a changing tolerance with respect to run time and optimality of results is given in [15], [19].

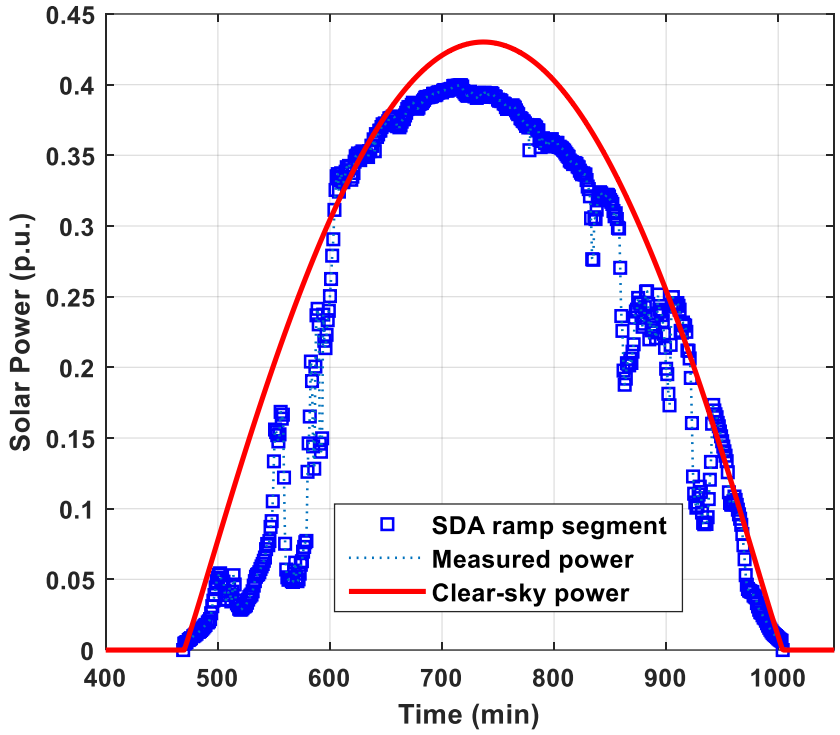

Figure 4.10. Visual comparison with insets of the SDA detection performance on a solar power data with 1-min resolution and using the significant ramp Definition 1.

Incidentally in Figure 4.12, the down-ramp which begins at about the 700th min and terminates at the 860th min appear to be inaccurate given its horizontal component. This is due to the significant ramp Definition 1 where the ramp magnitude is the only criterion for performance evaluation.

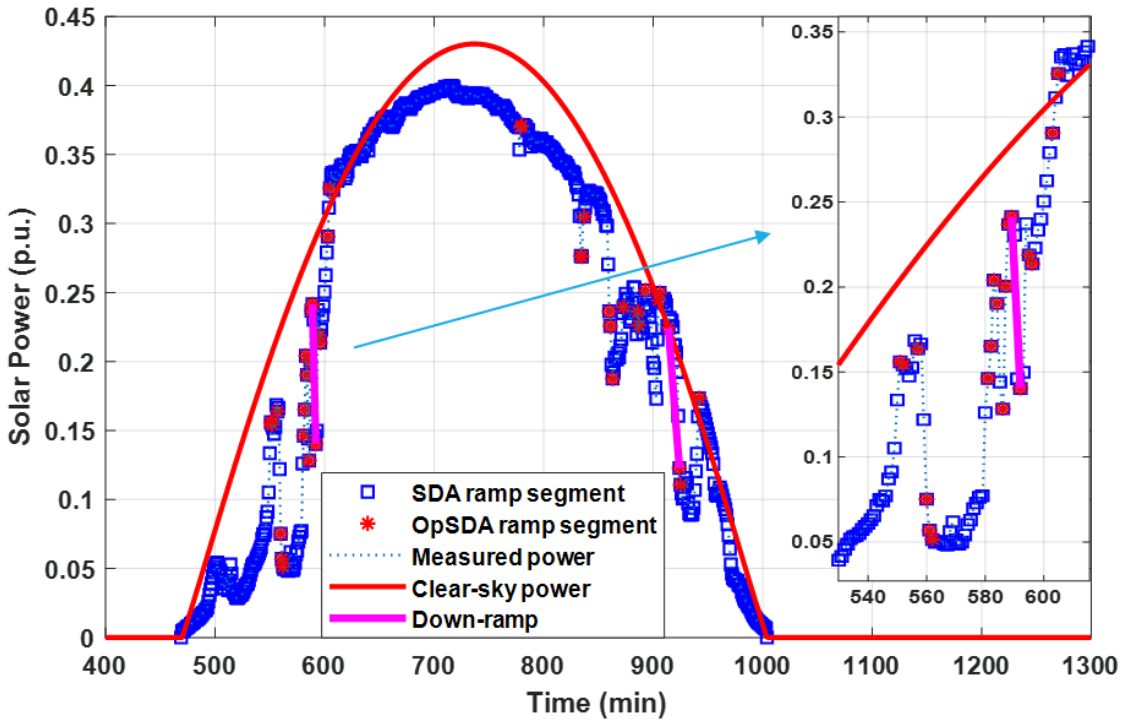

Figure 4.11. Visual comparison with insets of the OpSDA detection performances on a solar power data with 1-min resolution and using the significant ramp Definition 1.

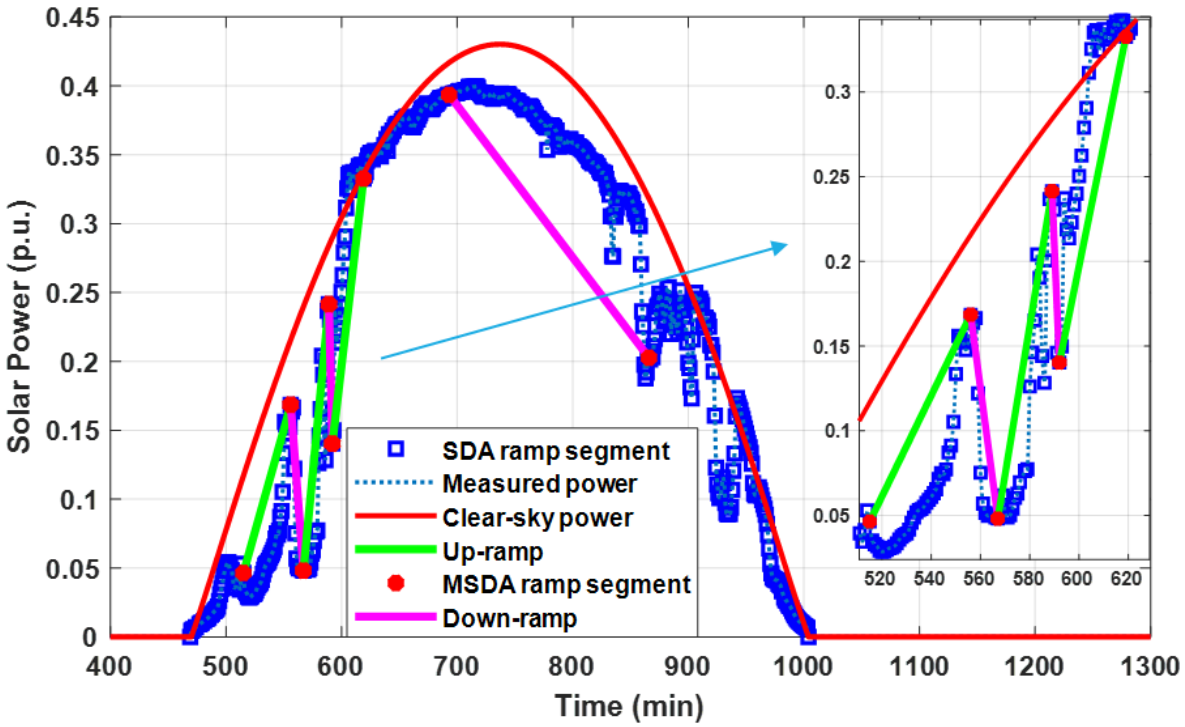

Figure 4.12. Visual comparison with insets of the MSDA detection performances on a solar power data with 1-min resolution and using the significant ramp Definition 1.

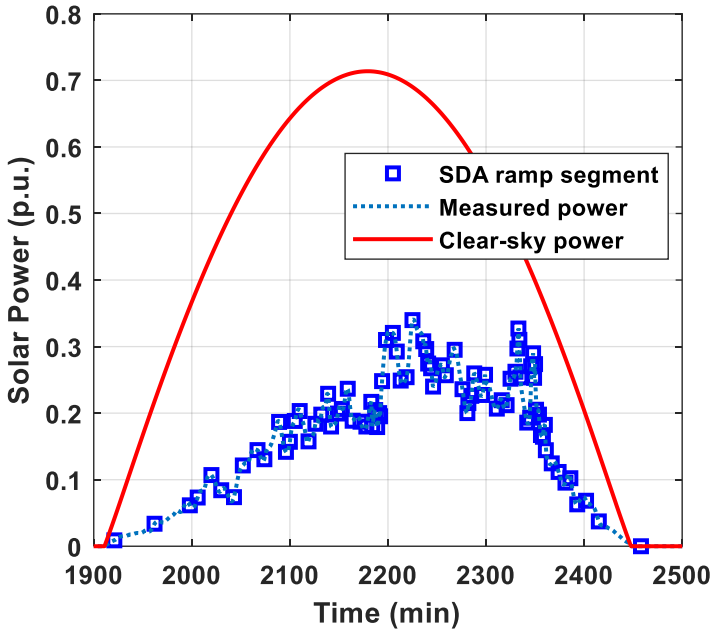

Figure 4.13. SDA detection performances on a solar power data with 1-min resolution and using an SDA tolerance value of 0.009.

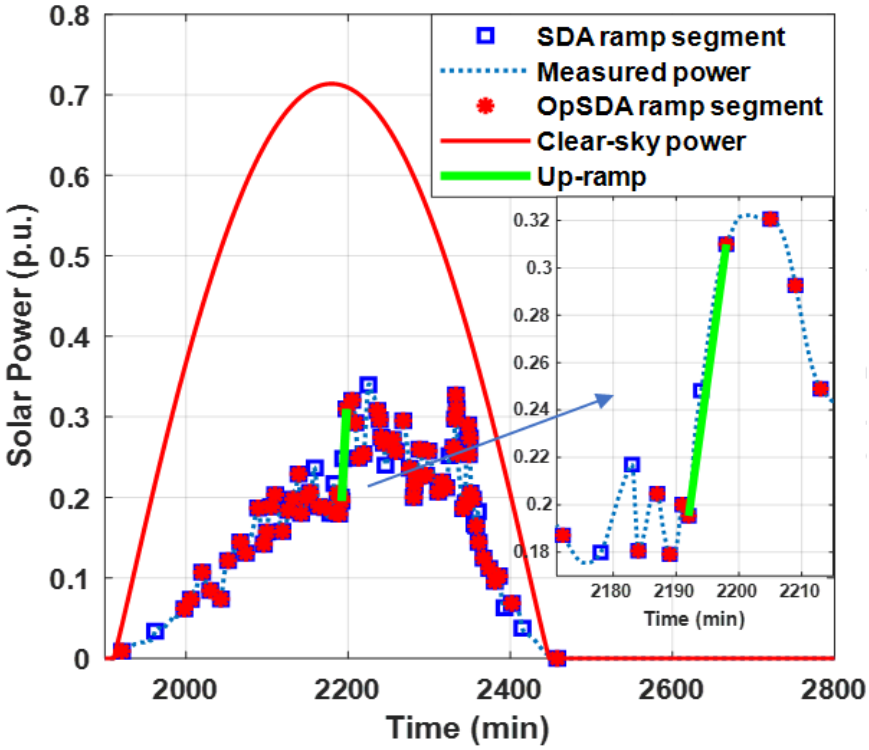

Figure 4.14. OpSDA detection performance on a solar power data with 1-min resolution and using an SDA tolerance value of 0.009.

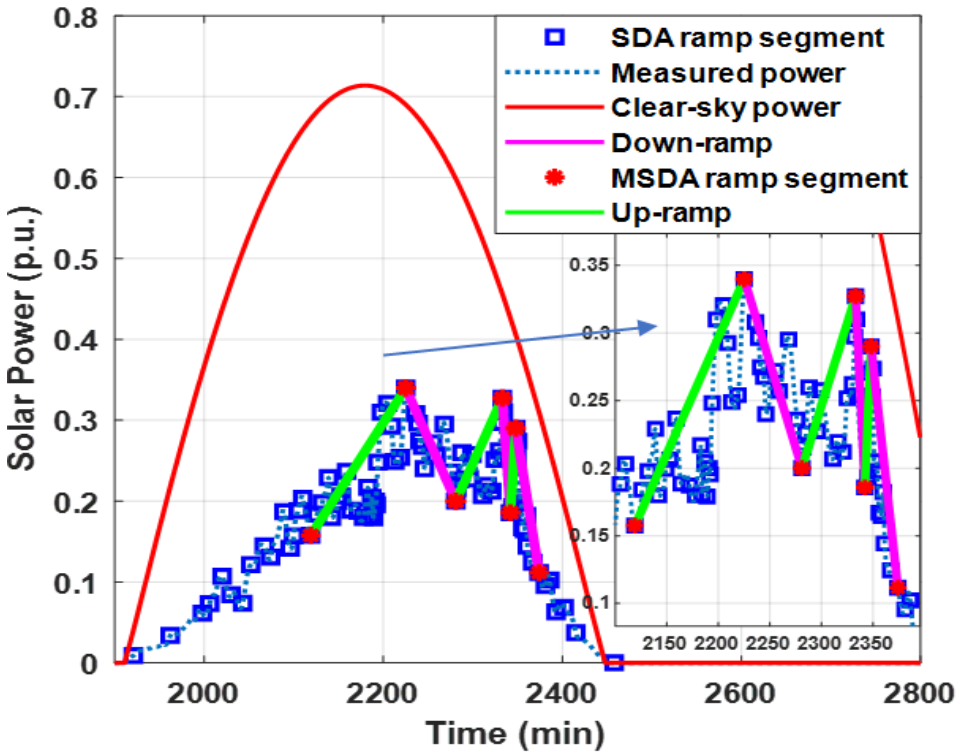

Figure 4.15. MSDA detection performance on a solar power data with 1-min resolution and using an SDA tolerance value of 0.009.

Similar to Figures 4.10 – 4.12, a 1-min resolution data is used to compare the models in Figures 4.13 – 4.15. To detail the proposed model performance in a different tolerance and climatic scenario, an ε value of 0.009 is used because it is the suggested optimal value in the OpSDA document. Also, the solar power data in Figur``es 4.13 – 4.15. was simulated using the irradiance data from NREL's University of Nevada – Las Vegas measuring station, which is over 1,800km away from the main test data location (NREL's University of Oregon). For easier visualization, a test day of relatively lower power generation and variability is selected. The maximum instantaneous power output is only about 35% of the nominal capacity as shown in Figures 4.13 – 4.15. The SDA detected no ramp events because the 1-min step changes are not large enough in magnitude to trigger any of the ramp definitions. The MSDA once more has detected more ramps and with more accurate start- and stop-times.

Tables 4.2., 4.3. and 4.4. show results of the number of detected ramps and the corresponding runtimes, the best results highlighted in bold. It is apparent that the basic SDA could be preferred to its modifications or optimizations because the former will always outperform the

latter if the performance evaluation criteria are based solely on computational simplicity and low execution time. However, if the number of detected ramps, ramp duration and magnitude detection accuracy are priority determinants, the MSDA will be best performer.

Table 4.2. Detected ramps using different definitions for 1-h resolution data.

Definitions	**Methods**	**Up-Ramps**	**Down-Ramps**	**Total**	**Runtime (s)**
Definition 1	SDA	185	155	340	**0.18**
	OpSDA	152	153	305	960.20
	MSDA	**194**	**169**	**363**	0.99
Definition 2	SDA	185	155	340	**0.18**
	OpSDA	148	155	303	960.17
	MSDA	**188**	**169**	**357**	0.92
Definition 3	SDA	185	164	349	**0.14**
	OpSDA	148	156	304	960.16
	MSDA	**188**	**177**	**365**	0.84

Table 4.3. Detected ramps using different definitions for 5-min resolution data.

Definitions	**Methods**	**Up-Ramps**	**Down-Ramps**	**Total**	**Runtime (s)**
Definition 1	SDA	1097	1106	2203	**5.12**
	OpSDA	1188	1171	2359	4328.93
	MSDA	**1324**	**1315**	**2639**	12.72
Definition 2	SDA	1097	1106	2203	**5.04**
	OpSDA	1183	1164	2347	4329.07
	MSDA	**1221**	**1295**	**2516**	12.63
Definition 3	SDA	1097	1136	2233	**5.58**
	OpSDA	1183	1191	2374	4328.86
	MSDA	**1221**	**1306**	**2527**	12.90

Table 4.4. Detected ramps using different definitions for 1-min resolution data.

Definitions	Methods	Up-Ramps	Down-Ramps	Total	Runtime (s)
Definition 1	SDA	2206	2216	4422	**55.05**
	OpSDA	2420	2380	4800	10443.45
	MSDA	**2747**	**2747**	**5494**	63.31
Definition 2	SDA	2208	2216	4424	**56.37**
	OpSDA	2420	2380	4800	10443.18
	MSDA	**2691**	**2744**	**5435**	62.95
Definition 3	SDA	2208	2306	4514	**57.42**
	OpSDA	2420	2452	4872	1044.27
	MSDA	**2691**	**2755**	**5446**	63.94

Some studies submitted that their proposed improvements of the SDA extracted more ramps than the SDA but they fail to also identify situations where the reverse might be the case or where the performances might be similar. It can be reasonably concluded that the number of ramps detected is influenced by some factors including the solar power data resolution and the magnitude of ramp definitions. Illustratively, a 1-h resolution might just be enough time for a majority of individual segments created by the SDA to be as large as 10% of the installed capacity. Whereas, several of these segments are merged into a single ramp by especially the OpSDA as evidenced in Table 4.2. Also worthy of mentioning is the measured power variability volatility or basically the weather profile. A highly variable climate is more likely to result in more SDA segments but fewer extracted ramps.

In Figures 4.16 – 4.24, the features of the detected SPRE using empirical statistics (magnitude, duration and rate) in different timescales are compared. Irrespective of timescale and ramp definition of interest, it is evident that all three detection algorithms share a similar probability density profile with respect to ramp magnitude. However, the MSDA had a wider empirical distribution of ramp size covering a significant area under the plots from about 0.1 p.u.

to 0.4 p.u. In terms of the average ramp magnitude variation across the timescales, there was no substantial observation worthy of mentioning.

As for the ramp rates and duration plots in Figures 4.16 – 4.24, considerable differences are observed. These differences could be attributed to a changing data resolution in the distributions. The ramp duration for 1-h resolution was longer and peaked around 60-min (the equivalent of a single step change) except for Figure 4.17 where the 1-h OpSDA showed a relatively horizontal distribution. Hence, the 1-min and 5-min timescales have higher ramp rates spread across the horizontal axis from 0 to 0.035p.u./min while the 1-h profile takes the resemblance of a normal distribution peaking around 0.003p.u./min. In all, the ramps detected by the MSDA have longer durations and larger magnitudes but with the penalty of lower ramp rates.

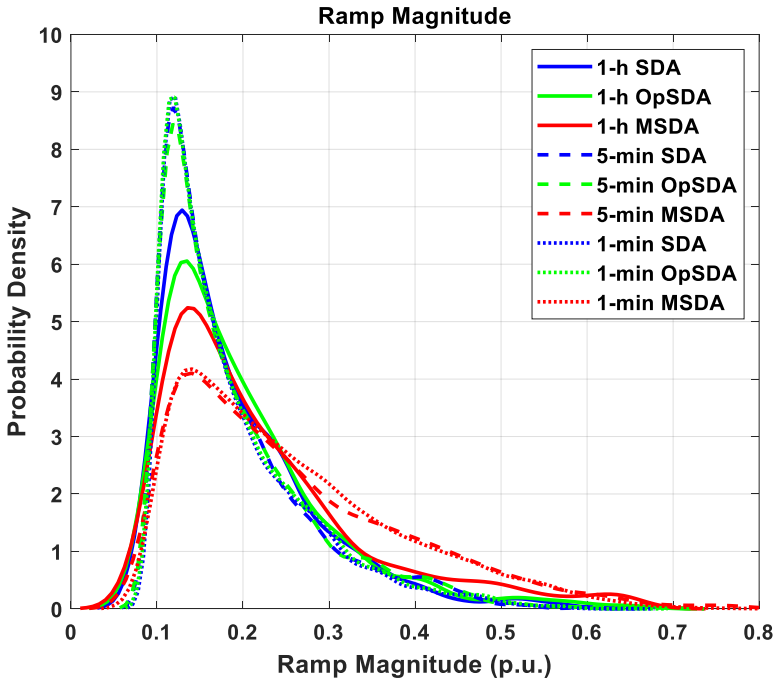

Figure 4.16. Probability density distributions of the magnitudes of the detected ramps by Definition 1.

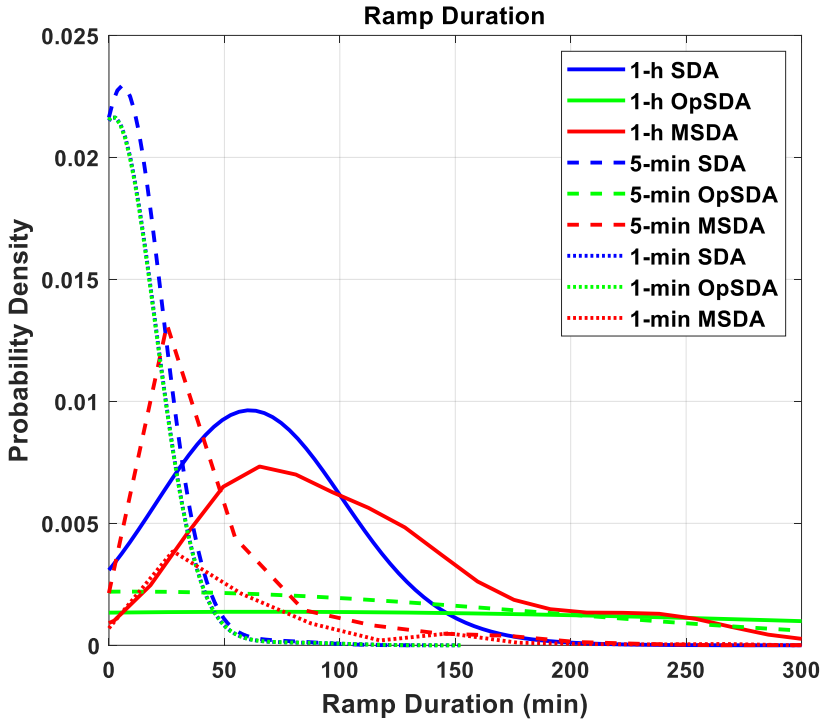

Figure 4.17. Probability density distributions of the durations of the detected ramps by Definition 1.

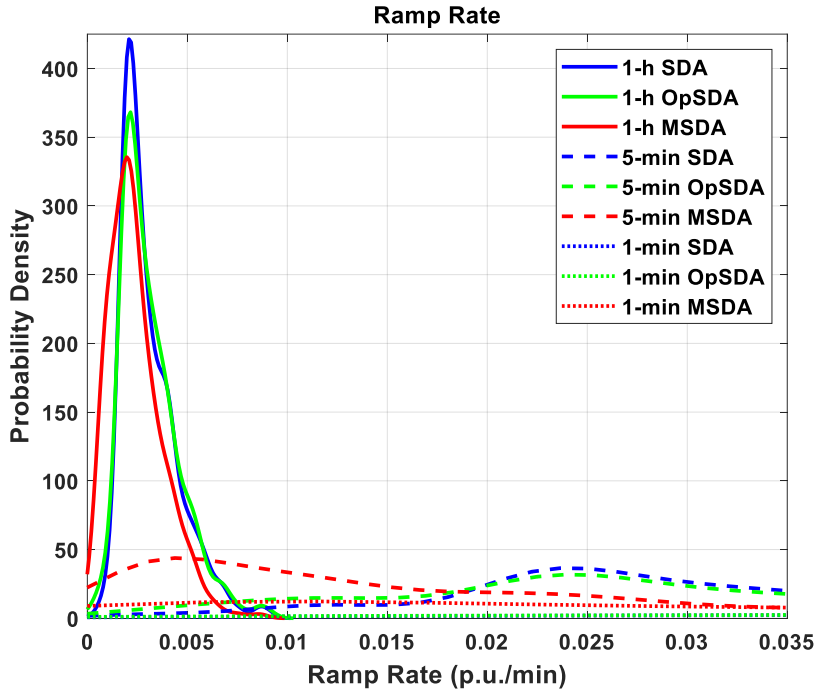

Figure 4.18. Probability density distributions of the rates of the detected ramps by Definition 1.

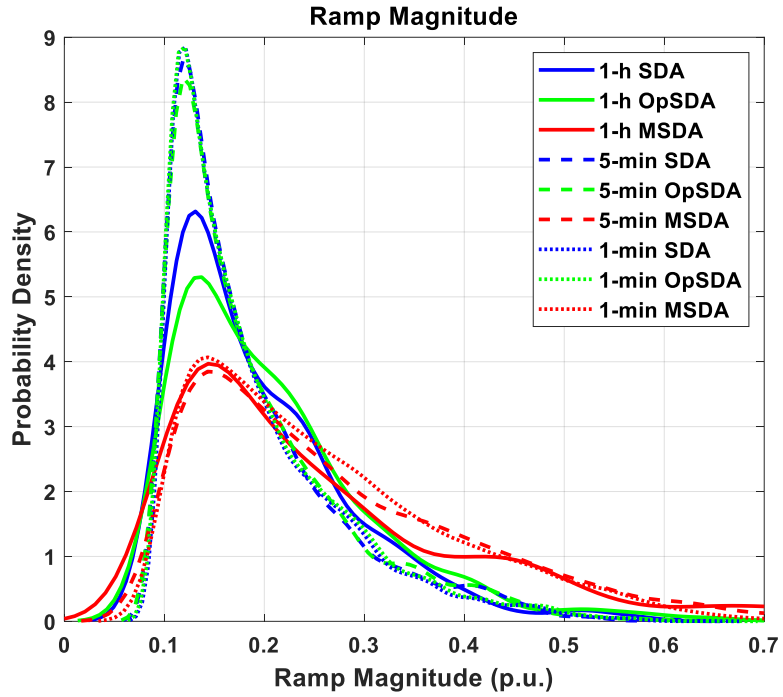

Figure 4.19. Probability density distributions of the magnitudes of the detected ramps by Definition 2.

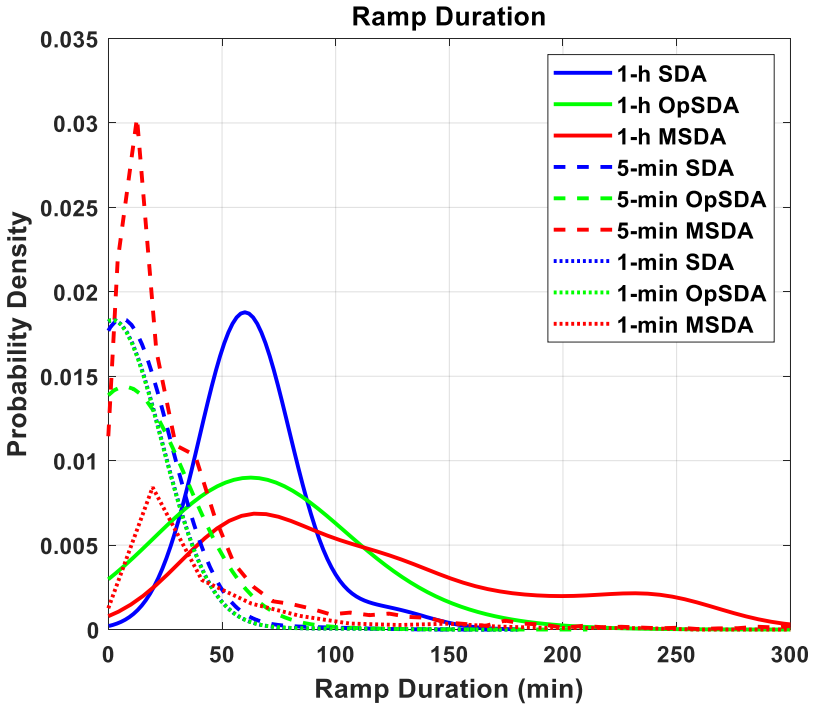

Figure 4.20. Probability density distributions of the durations of the detected ramps by Definition 2.

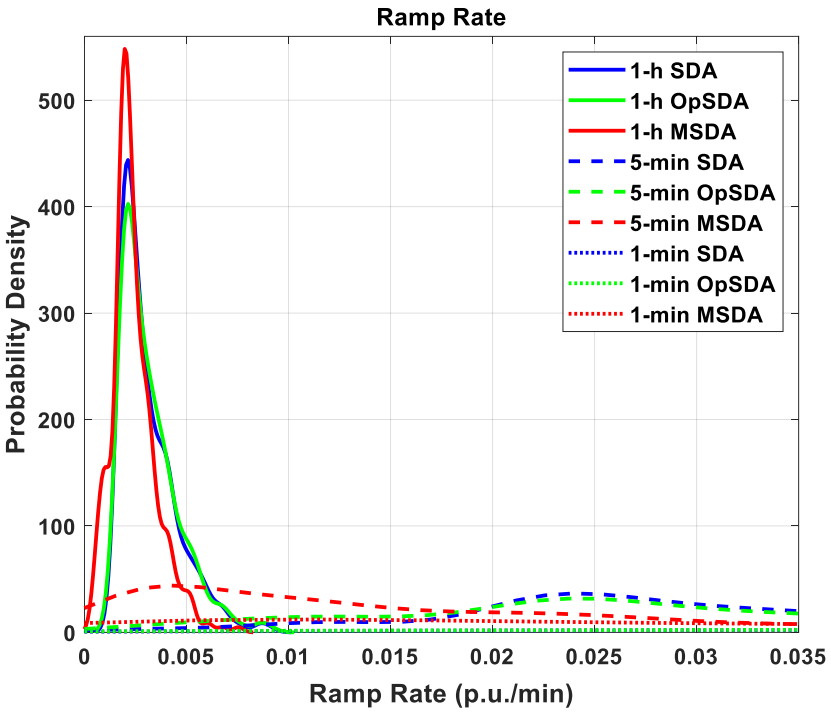

Figure 4.21. Probability density distributions of the rates of the detected ramps by Definition 2.

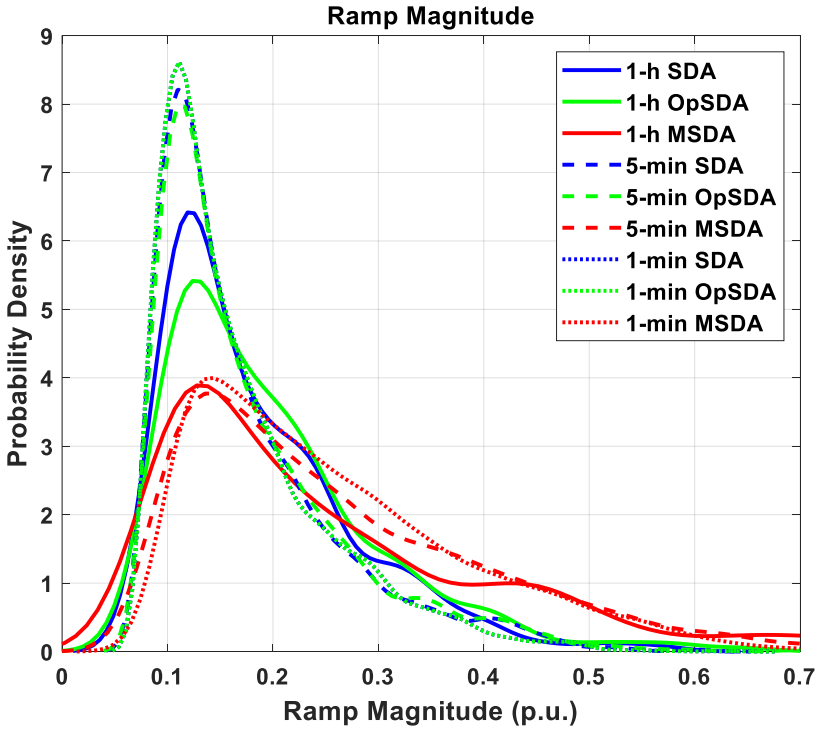

Figure 4.22. Probability density distributions of the magnitudes of the detected ramps by Definition 3.

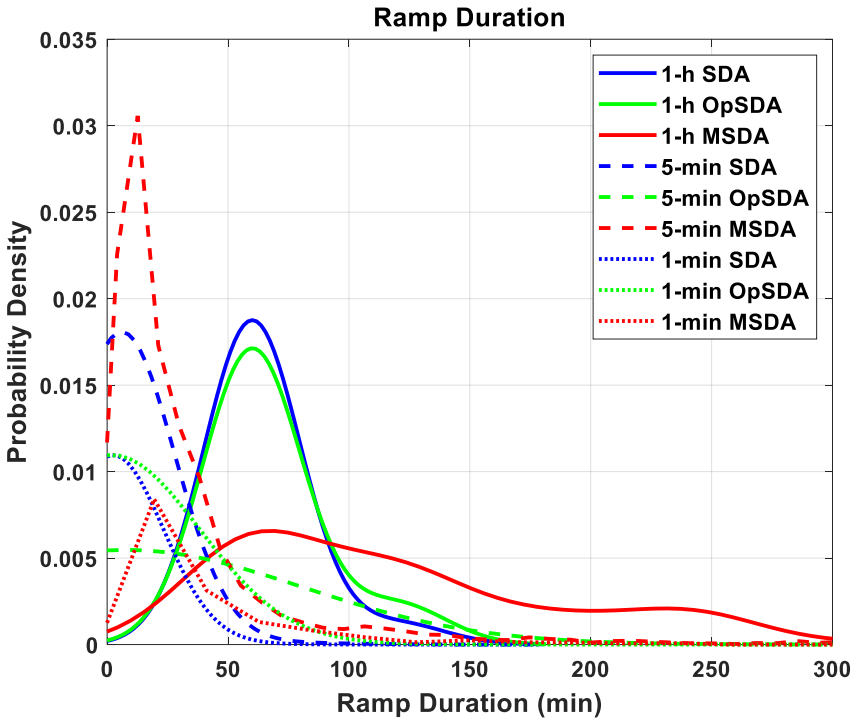

Figure 4.23. Probability density distributions of the durations of the detected ramps by Definition 3.

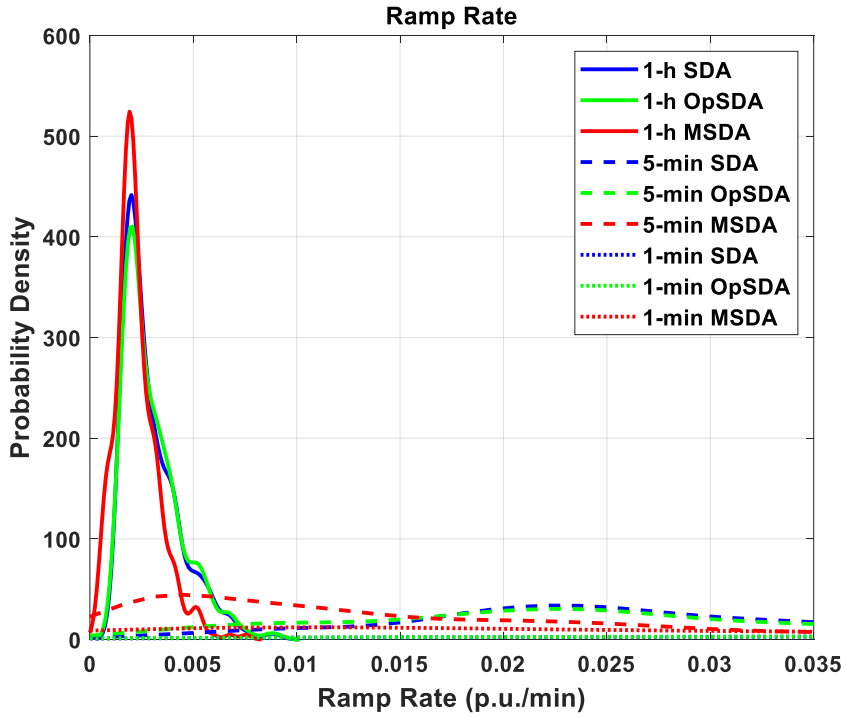

Figure 4.24. Probability density distributions of the rates of the detected ramps by Definition 3.

Table 4.5. Quantitative metrics for evaluating the performance of ramping events algorithms.

No	Metric	Representation	Description
1	POD	$\frac{sYeY}{sYeY + sNeY}$	Higher POD indicates better performance.
2	CSI	$\frac{sYeY}{sYeY + sNeY + sYeN}$	Ranges from 0 (worst case) to 1 (best case).
3	FBS	$\frac{sYeY + sYeN}{sYeY + sNeY}$	<1: more 'ends' detected accurately. >1: more 'starts' detected accurately
4	SR	$\frac{sYeY}{sYeY + sYeN}$	Higher SR means better performance.
5	FAR	$1 - SR$	Inverse proportionality to SR.

Additionally, the quantitative metrics detailed in [19] is employed for performance evaluation. The metrics are essentially focused on figuring out how well the ramp start (s) and end (e) times are captured by the ramp detection algorithms under analysis. The point at which an extracted ramp begins or ends is deemed accurate if that point is a local minima or maxima, i.e. if there's an inflection. If a ramping event start- and end-times are accurately detected, it is coded as sYeY (start-YES-end-YES) whereas it is coded sNeN (start-NO-end-NO) if the reverse is the case. Other combinations of detection include sYeN (start-YES-end-NO) and sNeY (start-NO-end-YES). The metrics include the Probability of Detection (POD), Critical Success Index (CSIr), Frequency Bias Score (FBS), Success Ratio (SR) and False Alarm Ratio (FAR) as shown in Table 4.5. This extensive analysis is motivated by the fact that accurate information on when a ramping event begins and terminates enables better informed scheduling decision making by utility operators [21].

For further performance evaluations, the results of the ramp start- and end-times detection accuracy in percentage of the total number of extracted ramps using Definition 1 are shown in Table 4.6. The best results (high SYEY and low SNEN) are highlighted in bold. The MSDA accurately detected an average of 69% and 3% of the total ramps SYEY and SNEN respectively for the three resolutions combined. The OpSDA and the SDA could only afford roughly 28%, 27% SYEY and 22%, 22% SNEN respectively.

Table 4.6. Results of the start- and end-points detection accuracy test fir 1-h, 5-min, and 1-min resolutions.

	SDA		OpSDA		MSDA	
1-h Resolution						
	End(Y)	End(N)	End(Y)	End(N)	End(Y)	End(N)
Start(Y)	109/32%	77/23%	98/32%	70/23%	**182/50%**	7/2%
Start(N)	85/25%	69/20%	70/23%	67/22%	156/43%	**18/5%**
Total	340		305		363	
5-min Resolution						
Start(Y)	595/27%	574/26%	637/27%	613/26%	**2111/80%**	158/6%
Start(N)	594/27%	440/20%	637/27%	472/20%	317/12%	**53/2%**
Total	2203		2359		2639	
1-min Resolution						
Start(Y)	973/22%	1149/26%	1104/23%	1248/26%	**4230/77%**	330/6%
Start(N)	1150/26%	1150/26%	1248/26%	1200/25%	769/14%	**165/3%**
Total	4422		4800		5494	

It is observed that the MSDA shows significant number of inaccurately detected ramp start-times (SNEY) especially for the 1-h resolution (43%). Upon investigation, the only possible and plausible cause is the daily pattern of sun movement from sunrise to sunset. Daily PV power values would rise from 0 to peak around midday and return to 0 at sunset, hence, would fail the accuracy test in Table 4.6 since these are not inflection points.

The results of the five metrics in Table 4.7 further corroborates the preference of the MSDA over its counterparts. Except for the dip in the POD in the 1-h resolution, the MSDA outperforms the other algorithms in almost all five performance indicators especially for 5-min and 1-min resolutions. This result is expected since the MSDA is designed to start and terminate a ramp only at large inflection and zero power points.

Table 4.7. Numerical results of the ramp detection performance evaluation metrics.

1-h Resolution					
	POD	CSI	FBS	SR	FAR
SDA	0.56	0.39	0.98	0.57	0.43
OpSDA	**0.58**	0.41	0.99	0.59	0.41
MSDA	0.54	**0.53**	0.55	**0.97**	**0.03**

5-min Resolution					
SDA	0.5	0.34	0.97	0.51	0.49
OpSDA	0.51	0.34	0.98	0.52	0.48
MSDA	**0.87**	**0.82**	0.94	**0.93**	**0.07**
1-min Resolution					
SDA	0.45	0.29	1.00	0.45	0.55
OpSDA	0.46	0.30	0.99	0.46	0.54
MSDA	**0.84**	**0.79**	0.90	**0.93**	**0.07**

4.5 Conclusion

In this chapter, a novel solar power ramp detection algorithm is proposed using a modified swinging door algorithm (MSDA). We programmed the MSDA for accurate detection of ramp start- and end-times while retaining the positives of the SDA. To demonstrate the robustness of the proposed model to produce good accuracy regardless of input data sampling rate, the results of the proposed model are compared with that of an existing optimization of the SDA (OpSDA) on the basis of the number of extracted ramps, ramp magnitude, duration, rate and other empirical parameters.

The results show that relative to the OpSDA, the MSDA offered significant performance improvements in every evaluation metric. In some instances, the OpSDA relative to the SDA extracted fewer ramps because the former combined several more ramp segments into a single ramp resulting in slightly longer ramp durations and larger magnitudes. Instances are also observed where the SDA without any modifications is preferred because of its lower execution time, robustness and computational simplicity.

Although the test data presented in this chapter is from residential (small scale) systems with small or inconsequential geographical spread, we affirm that the results are also valid for utility scale systems since the only difference would be reduced average variability due to geographical smoothing. The proposed model performance is not impacted by differing climates or geographical regions. We showed detailed results for different time resolutions representing varying variability rates or patterns. Also, since only solar variability profiles are considered for generating the model,

although very unlikely, the performance results might be different if the MSDA without any tweaking is applied on wind power data. In the future work, we will validate the MSDA for wind power ramp characterization and ramping events forecast.

4.6 References

[1] L. Polleux, T. Schuhler, G. Guerassimoff, J. P. Marmorat, J. Sandoval-Moreno, and S. Ghazouani, "On the relationship between battery power capacity sizing and solar variability scenarios for industrial off-grid power plants," *Appl. Energy*, vol. 302, no. July, p. 117553, 2021, doi: 10.1016/j.apenergy.2021.117553.

[2] K. M. Banjar-Nahor, L. Garbuio, V. Debusschere, N. Hadjsaid, T. T. H. Pham, and N. Sinisuka, "Study on Renewable Penetration Limits in a Typical Indonesian Islanded Microgrid Considering the Impact of Variable Renewables Integration and the Empowering Flexibility on Grid Stability," *Proc. - 2018 IEEE PES Innov. Smart Grid Technol. Conf. Eur. ISGT-Europe 2018*, pp. 1–6, 2018, doi: 10.1109/ISGTEurope.2018.8571673.

[3] A. Makibar, L. Narvarte, and E. Lorenzo, "On the relation between battery size and PV power ramp rate limitation," *Sol. Energy*, vol. 142, pp. 182–193, 2017, doi: 10.1016/j.solener.2016.11.039.

[4] A. Florita, B. M. Hodge, and K. Orwig, "Identifying wind and solar ramping events," *IEEE Green Technol. Conf.*, pp. 147–152, 2013, doi: 10.1109/GreenTech.2013.30.

[5] L. Exizidis, S. J. Kazempour, P. Pinson, Z. de Greve, and F. Vallée, "Sharing wind power forecasts in electricity markets: A numerical analysis," *Appl. Energy*, vol. 176, pp. 65–73, 2016, doi: 10.1016/j.apenergy.2016.05.052.

[6] R. Doherty and M. O'Malley, "A new approach to quantify reserve demand in systems with significant installed wind capacity," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 587–595, 2005, doi: 10.1109/TPWRS.2005.846206.

[7] M. J. Cui, D. P. Ke, Y. Z. Sun, D. Gan, J. Zhang, and B. M. Hodge, "A scenario generation method for wind power ramp events forecasting," *IEEE Power Energy Soc. Gen. Meet.*, vol. 2015-September, 2015, doi: 10.1109/PESGM.2015.7285818.

[8] M. K. Hossain and M. H. Ali, “Statistical analysis of ramp rates of solar Photovoltaic system connected to grid,” *2014 IEEE Energy Convers. Congr. Expo. ECCE 2014*, no. May, pp. 2524–2531, 2014, doi: 10.1109/ECCE.2014.6953737.

[9] B. Li *et al.*, “Sizing ramping reserve using probabilistic solar forecasts: A data-driven method,” *Appl. Energy*, vol. 313, p. 118812, May 2022, doi: 10.1016/J.APENERGY.2022.118812.

[10] 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,” *Electr. Eng. Japan (English Transl. Denki Gakkai Ronbunshi)*, vol. 166, no. 4, pp. 9–19, 2009, doi: 10.1002/eej.20723.

[11] C. Kamath, “Associating weather conditions with ramp events in wind power generation,” *2011 IEEE/PES Power Syst. Conf. Expo. PSCE 2011*, pp. 1–8, 2011, doi: 10.1109/PSCE.2011.5772527.

[12] M. Cui, D. Ke, Y. Sun, D. Gan, J. Zhang, and B. M. Hodge, “Wind Power Ramp Event Forecasting Using a Stochastic Scenario Generation Method,” *IEEE Trans. Sustain. Energy*, vol. 6, no. 2, pp. 422–433, Apr. 2015, doi: 10.1109/TSTE.2014.2386870.

[13] C. Kamath, “Understanding wind ramp events through analysis of historical data,” *2010 IEEE PES Transm. Distrib. Conf. Expo. Smart Solut. a Chang. World*, 2010, doi: 10.1109/TDC.2010.5484508.

[14] N. Cutler, M. Kay, K. Jacka, and T. S. Nielsen, “Detecting, categorizing and forecasting large ramps in wind farm power output using meteorological observations and WPPT,” *Wind Energy*, vol. 10, no. 5, pp. 453–470, Sep. 2007, doi: 10.1002/WE.235.

[15] Y. Qu, J. Xu, Y. Sun, C. Ching, S. Liao, and D. Ke, “Electrical Power and Energy Systems A parameter and resolution adaptive algorithm for rapid detection of ramp events in di ff erent timescale databases of the power system,” *Electr. Power Energy Syst.*, vol. 112, no. May, pp. 393–403, 2019, doi: 10.1016/j.ijepes.2019.05.033.

[16] J. Zhang, M. Cui, B. M. Hodge, A. Florita, and J. Freedman, “Ramp forecasting performance from improved short-term wind power forecasting over multiple spatial and temporal scales,”

Energy, vol. 122, pp. 528–541, 2017, doi: 10.1016/j.energy.2017.01.104.

[17] Y. Cui *et al.*, "Algorithm for identifying wind power ramp events via novel improved dynamic swinging door," *Renew. Energy*, vol. 171, pp. 542–556, 2021, doi: 10.1016/j.renene.2021.02.123.

[18] J. Zhang *et al.*, "A suite of metrics for assessing the performance of solar power forecasting," *Sol. Energy*, vol. 111, pp. 157–175, 2015, doi: 10.1016/j.solener.2014.10.016.

[19] M. Cui *et al.*, "An Optimized Swinging Door Algorithm for Identifying Wind Ramping Events," *IEEE Trans. Sustain. Energy*, vol. 7, no. 1, pp. 150–162, 2016, doi: 10.1109/TSTE.2015.2477244.

[20] M. Cui, J. Zhang, C. Feng, A. R. Florita, Y. Sun, and B. M. Hodge, "Characterizing and analyzing ramping events in wind power, solar power, load, and netload," *Renew. Energy*, vol. 111, pp. 227–244, 2017, doi: 10.1016/j.renene.2017.04.005.

[21] D. Lyners, H. Vermeulen, and M. Groch, "Wind power ramp event detection using a multi-parameter segmentation algorithm," *Energy Reports*, vol. 7, pp. 5536–5548, 2021, doi: 10.1016/j.egyr.2021.08.137.

[22] R. Sevlian and R. Rajagopal, "Detection and statistics of wind power ramps," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 3610–3620, 2013, doi: 10.1109/TPWRS.2013.2266378.

[23] H. Kuang, W. Jianhui, H. Zhang, H. Zhong, and G. Zhu, "A New Definition Method of Wind Power Ramp Sections," *IEEE Access*, vol. 8, pp. 202058–202066, 2020, doi: 10.1109/ACCESS.2020.3032547.

[24] M. Cui, C. Feng, Z. Wang, and J. Zhang, "Statistical representation of wind power ramps using a generalized Gaussian mixture model," *IEEE Trans. Sustain. Energy*, vol. 9, no. 1, pp. 261–272, Jan. 2018, doi: 10.1109/TSTE.2017.2727321.

[25] M. Cui, J. Zhang, A. Florita, B.-M. Hodge, D. Ke, and Y. Sun, "Solar Power Ramp Events Detection Using an Optimized Swinging Door Algorithm," *Proceedings of the ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Volume 2A: 41st Design Automation Conference*. Boston, Massachusetts, USA. August 2–5, 2015. V02AT03A027. ASME, doi: 10.1115/DETC2015-

46849.

[26] B. M. Mazumdar, M. Saquib, and A. K. Das, “ScienceDirect An empirical model for ramp analysis of utility-scale solar PV power,” *Sol. Energy*, vol. 107, pp. 44–49, 2014, doi: 10.1016/j.solener.2014.05.027.

[27] D. M. Willy, A. Dyreson, T. L. Acker, E. Morgan, and R. K. Flood, “Dead band method for solar irradiance and power ramp detection algorithms,” in *43rd ASES National Solar Conference 2014, SOLAR 2014, Including the 39th National Passive Solar Conference and the 2nd Meeting of Young and Emerging Professionals in Renewable Energy*, 2014, vol. 2, pp. 1204–1211.

[28] C. Hansen, J. Stein, and A. Ellis, “Statistical criteria for characterizing irradiance time series,” *Sandia Rep. Sand2010-7314*, no. October, pp. 1–51, 2010, [Online]. Available: http://energy.sandia.gov/wp/wp-content/gallery/uploads/107314.pdf

[29] J. S. Stein, C. W. Hansen, and M. J. Reno, “The variability index: A new and novel metric for quantifying irradiance and pv output variability,” *World Renew. Energy Forum, WREF 2012, Incl. World Renew. Energy Congr. XII Color. Renew. Energy Soc. Annu. Conf.*, vol. 4, pp. 2764–2770, 2012.

[30] Y. Cui, Y. Su, Y. Liu, Y. Liu, and D. Smith, “Study of variability metrics for solar irradiance and photovoltaic output,” *IEEE Power Energy Soc. Gen. Meet.*, vol. 2018-January, pp. 1–5, Jan. 2018, doi: 10.1109/PESGM.2017.8274193.

[31] “UOSMRL Daily Data.” https://midcdmz.nrel.gov/apps/daily.pl?site=UOSMRL&start=20160819&yr=2022&mo=6&dy=2 (accessed Jun. 14, 2022).

5 SUMMARY AND CONCLUSIONS

5.1 Summary

The highly volatile nature of power generation from PV systems and its associated unreliability is counterproductive to the transition from fossil fuels to more sustainable alternatives for electricity generation. However, as global population grows leading to increased power consumption and as the outcry for emissions reduction becomes increasingly formidable, it becomes inevitable that reliance on renewable energy sources like PV systems will increase. Consequently, energy discourses are centered on how to accommodate the attending challenges posed by these sources. Dealing with this problem, utility grids have to be smarter and forecasting technologies have to be sufficiently reliable. Smart grids and forecast models require data for performance improvements.

This thesis presented solutions to three different challenges in the renewable energy world with keen interests in small-scale PV systems. Although these systems, with respect to installed capacities seemingly do not post significant threats compared to commercial or large-scale counterparts, the proliferation of these small capacity systems concentrated in distributed networks poses an increasing risk of reversed power flow and complicates load predictions. The three challenges addressed in this thesis are discussed in chapters 2 to 4.

Chapter 2 presents an approach to solve the eventual problem of performance data unavailability for residential PV systems since this data is usually hidden from utility operators who only see net-load from residential buildings. The proposed approach involves a simple strategy of utility operators setting up web repositories for collecting these data in real- or near real-time. The logistics involved in creating such repositories are discussed. Resolutions to potential oppositions and challenges are presented alongside the benefits. Case studies of the implementation of similar approaches are also discussed to substantiate the claim that the proposed approach is practicable and feasible.

In chapter 3, a prediction model for day-ahead regional PV power forecasting is presented. The model is specifically relevant for areas where numerical weather predictions of solar irradiance are not available. An Artificial neural network that requires only meteorological

variables excluding solar irradiance as inputs is used for irradiance predictions. Both machine learning-based and physical models are considered for PV power modelling from solar irradiance. Case study small-scale PV systems are divided into clusters and representative systems selected using k-means clustering and principal component analysis algorithms. The regional power output is then modeled from the representative systems using simple linear relationships. The accuracy of the proposed significantly improves the status quo.

Chapter 4 presents a novel technique for solar power ramp events extraction and characterization. Only significant ramps by adopted ramp definitions are characterized based on ramp duration, magnitude and rate. Hence, ramps occurring in both clear-sky periods and the actual solar power data are ignored. The technique involves a modification of a data compression algorithm – the swinging door algorithm (SDA). Simple empirical and statistical tests are used to improve on the identified short comings of the SDA. The accuracy of the proposed modified swinging door algorithm (MSDA) was compared to the parent SDA and an already existing optimization of the SDA – optimized swinging door algorithm (OpSDA). The MSDA is found to accurately extract more ramps and computationally simple to execute.

5.2 Contributions

The main contributions of this thesis include:

1. Proposed an approach for residential PV power data collection to solve the problem of data unavailability.
2. Proposed a model for day-ahead regional scale PV power prediction in regions where solar irradiance data is unavailable or has intolerable errors.
3. Developed a solar power ramp characterization algorithm based on a modified variant of the SDA.
4. Proposed using the natural variability index for the automatic calibration of the tolerance of the SDA.

5.3 Conclusions

From the course of preparing this thesis, the following conclusions are reached:

1. Residential PV systems are increasingly more problematic as they grow in numbers.
2. The performance of small-scale PV systems is invisible to utility operators. The importance of developing a method for collecting the invisible data increases with their proliferation.
3. It is feasible to create web repository for collecting and documenting performance data in real- and near real-time.
4. Comparatively good accuracy forecasts can be obtained from models that do not require irradiance as inputs.
5. Physical models for modelling power outputs from irradiance data are more accurate than available alternatives.
6. Modifications of the SDA for solar power ramp extraction using basic empirical test could improve the status quo of ramp extraction.

5.4 Suggestions for Future Work

The following suggestions for future work are recommendations:

1. A Call for Evidence should be carried out to gather information on implementing the proposed web repository creation approach with the aim of reducing public and social resistance to implanting the approach.
2. If available and reliable, a larger sample set of real-world case studies of PV systems should be used to train the proposed model. Also, the physical model should be preferred for power output modelling.
3. The proposed ramp detection algorithm should be enhanced to increase its versatility for wind power ramp events detection.