8-2023

Digital Twins and Artificial Intelligence for Applications in Electric Power Distribution Systems

Deborah George
dgeorg2@clemson.edu

DIGITAL TWINS AND ARTIFICIAL INTELLIGENCE FOR APPLICATIONS IN ELECTRIC POWER DISTRIBUTION SYSTEMS

A Thesis

Presented to

the Graduate School of

Clemson University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in

Electrical Engineering

---

by

Deborah George

August 2023

Accepted by:

Dr. Ganesh Kumar Venayagamoorthy, Committee Chair

Dr. Rajendra Singh, Committee Member

Dr. Yingjie Lao, Committee Member

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

List of Figures (Continued)

ABSTRACT

As modern electric power distribution systems (MEPDS) continue to grow in complexity, largely due to the ever-increasing penetration of Distributed Energy Resources (DERs), particularly solar photovoltaics (PVs) at the distribution level, there is a need to facilitate advanced operational and management tasks in the system driven by this complexity, especially in systems with high renewable penetration dependent on complex weather phenomena.

Digital twins (DTs) or virtual replicas of the system and its assets, enhanced with AI paradigms can add enormous value to tasks performed by regulators, distribution system operators and energy market analysts, thereby providing cognition to the system. DTs of MEPDS assets and the system can be utilized for real-time and faster-than-real-time operational and management task support, planning studies, scenario analysis, data analytics and other distribution system studies.

This study leverages DT and AI to enhance DER integration in an MEPDS as well as operational and management (O&M) tasks and distribution system studies based on a system with high PV penetration. DTs have been used to both estimate and predict the behavior of an existing 1 MW plant in Clemson University by developing asset digital twins of the physical system. Solar irradiance, temperature and wind-speed variations in the area have been modeled using physical weather stations located in and around the Clemson region to develop ten virtual weather stations. Finally, DTs of the system along with virtual and physical weather stations are used to both estimate and predict, in short time intervals, the real-time behavior of potential PV plant installations over the region. Ten virtual PV plants and three hybrid PV plants are studied, for enhanced cognition in the system. These physical, hybrid and virtual PV sources enable situational awareness and situational intelligence of real-time PV production in a distribution system.

DEDICATION

To my parents, my family and my friends.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. G. Kumar Venayagamoorthy. Additionally, I would like to thank Dr. Rajendra Singh and Dr. Yingjie Lao for serving as my advisory committee and their advice and support.

I would like to acknowledge the Real-Time Power and Intelligent Systems (RTPIS) Laboratory (http://rtpis.org/) for providing world-class lab facilities for conducting my research. I would also like to acknowledge the National Science Foundation and the Duke Energy Distinguished Professor Endowment Fund for directly or indirectly funding this MS Thesis research.

This work is supported by the National Science Foundation under grant CNS 2131070 as well as the Duke Energy Distinguished Professor Endowment Fund. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding authorities.

CHAPTER ONE

INTRODUCTION

With the continuous modernization and changes in the Electric Distribution system comes new difficulties, increased complexities and needs for supporting infrastructure and technology. Historically, the distribution system used to require comparatively less comprehensive design, and include less operational and maintenance complexity, especially when compared to the transmission or generation side of the grid. However, with the advent of distributed energy resource penetration (DER) at the distribution level, new operational and management challenges began to quickly arise. DERs have led to significant hardware upgrades and other infrastructural changes, opened new opportunities for distributed generation, microgrids and the like and quickly led to changing attitudes with respect to the distribution system.

Prior to 2018, DERs were to be quickly disconnected during abnormal voltage conditions as per IEEE 1547-2003 standard. However, new standards have been defined as per IEEE 1547-2018 [1] and 1547a-2020 [2] defining DER interconnection and integration standards. FERC 2222 [3] additionally bolstered the change, allowing the participation of aggregated DERs in wholesale markets. A large portion of DERs consist of distributed photovoltaics (DPV), additionally supported by the ever decreasing prices of PV panel production. PV is however inherently dependent on climatic phenomena, while being a non-dispatchable, grid-following source when connected to the grid [4].

DER integration and by extension DPV, have become therefore, an area where extensive research, system modeling at various time-scales and real-time simulations are to be performed, enabling distribution system operator support for volt-var control and optimization. New products in the distribution automation space such as ADMS and DERMS systems, STATCOMS and smart inverters have become commonly employed. DSOs need short and long-term PV power forecasts to enable intraday and next-day energy market trading, and enhanced volt-var control (VVC) and volt-var-optimization (VVO) for day-ahead planning and system management, needed to mitigate power and

voltage instability common in PV sources [5]. These studies are also required by regulatory and utility support for planning, reliability and resilience studies analyzing the impact of utility scale PV installations on the grid, hosting capacity of the system and the setting of VVC control points and sophisticated control strategies for DPV O&M optimization [6].

Essentially situational awareness, or real-time awareness of the system's behavior and situational intelligence, and predictive awareness or data-based, informed behavioral analysis of how the system will continue to behave over time is needed. This means a highly complex, dynamic, non-stochastic system's behavior is to be modeled, often at varying time scales from real-time to faster than real-time over long periods.

Digital Twin (DT) technology is therefore an emerging avenue of research in this area [9]. DTs, or a dynamic real-time virtual model of a process, system or asset, However, given the complexity, cost and modeling hardware, software and other resource requirements involved in DT development, the employment of physics-based DT creates added challenges, especially when a physics-driven DT may be highly area and installation specific. Data-driven approaches to DT modeling, however, can be a solution to this issue. Data-driven DTs and hybrid DTs that are physics-based with artificial intelligence paradigms are scalable, easily optimizable, can be applied to new and varying installation types and locations, and can be comparatively cost-effective to design and deploy.

Objectives of the Thesis

The objectives of this thesis were to primarily mitigate the complexity and costs involved with PV power behavior modeling and develop a scalable, cost-effective means to generate realistic PV plant data streams using available measurement sites. Additionally, a means to enable estimation and predictive analysis of PV plant behavior in real-time and PV power production over a region was a focus, to further produce situational awareness and situational awareness of PV power production with the eventual aim of easier integration of DPVs in an MEPDS.

Objectives of Chapter 2

An explorative review of the emerging MEPDS and needs in the system is needed to identify gaps, potential use-cases and typical applications currently utilized for leveraging DTs and AI enabled support in MEPDS

Objectives of Chapter 3

PV power prediction and estimation is an area of constant research. DSOs need support for operational and management needs, especially for PV power forecasting, both over short and long-term. For instance, cloud-cover drops with over 50% of peak power over a minute can generate fluctuations in PV power generation, leading to power quality issues and other reliability challenges. Voltage regulation at the distribution level is additionally increasingly challenging if there are DPVs. Estimation and short-term prediction of PV power over an entire area is necessary for optimized DPV and PV plant operation and employment of effective VVC strategies.

Objectives of Chapter 4

Solar irradiance and solar irradiance variations based on climate defines PV power generation, while effective decision making and DT modeling needs high fidelity data at small timesteps in real time at various regions

Objectives of Chapter 5

Prediction and estimation of PV power generation in the area, in various placement locations is needed for enhanced observability and monitoring of the system.

Contributions of the Thesis

- A careful review of the major challenges with the changing MEPDS and the use-case for DT technology for common O&M challenges, especially as pertains to DPV and PV power production is provided. AI is identified as an enabler of cost-effective, scalable and optimizable DT design, particularly suited to the behavior modeling of PV plants and further enabling advanced applications in the MEPDS. In particular, the use-cases of data-driven DT models for DPV optimization and integration are presented.

- A DT approach was utilized for scalable, low cost and reliable PV plant behavior estimation and prediction in real-time, accurately estimating and predicting PV power production over various cloud cover patterns utilizing multi-layer perceptrons and echo state networks to model its behavior and was tested over a period of 3 months of solar PV data. The DTs developed can be utilized for traditional DT use cases. For instance, PV power estimation and prediction can be used for fault detection, condition monitoring and predictive analysis in a PV site with available measurements and for missed data imputation in a PV site with missing or erroneous PMU measurements.
- Intelligent mutation consisting of mathematical based models such as inverse distance weighting as well as data-driven inferences of solar irradiance, temperature and wind speed variations between three existing weather stations were used to identify trends and predict weather at user-selected locations over the region. Using this information, 10 new virtual weather stations (VWS) were developed, providing estimated solar irradiance, temperature and associated wind speed and direction in those locations in real-time.
- A multi-DT methodology for estimating and forecasting PV power over the selected region in real-time is developed supporting various scenarios and purposes, enabling situational awareness and situational intelligence of PV power generation in a distribution system. PV plants are generated are of three types and are categorized as Case 1: Traditional PV sources, where a physical weather station and prediction or estimation DT of an existing PV plant is developed, Case 2: Hybrid PV sources where an existing PV plant and these scenarios can be further validated and optimized as needed for PV power prediction or estimation in locations where measurement data is available, and in locations with non-existing measurements where data imputation is necessary.

Publications

- A paper titled “Digital Twins and AI for Photovoltaic Plant Power Estimation and Prediction in Electric Power Distribution Systems”, consisting of chapters 2 and 3 of the thesis have been submitted to the 33rd Australasian Universities Power Engineering Conference (AUPEC2023).

- A paper titled “Situational Awareness and situational intelligence for distributed PV plants”, consisting of parts of chapters 3,4 and 5 has been submitted to the 2023 IEEE Symposium Series on Computational Intelligence.

Organization of the Thesis

The rest of the thesis is outlined as follows:

- Chapter 2: A comprehensive review of DT technology, the beneficial inclusion of data-driven, here AI, paradigms in its design particularly as a means to mitigate complexity and support O&M challenges in the MEPDS and DPV integration is presented.
- Chapter 3: ESNs and MLP based methodologies for data-driven DT design is employed for PV power estimation and prediction acting as a DT of a physical system, Clemson’s R06 site. The DT performance is validated and tested against real-time behavior of the site over the spring.
- Chapter 4: Virtual weather stations that generate real-time input streams of solar irradiance and temperature based on spatial variations over the site and three existing physical weather stations are then developed.
- Chapter 5: Finally, all weather station sources available at the site, now including new virtual weather stations are fed into the tested, designed DT model, enabling estimation and prediction of PV power and real-time PV plant behavior, resulting in “Virtual Real-Time Photovoltaic Plants” or V-RT-PVPs.
- Chapter 6: The conclusions drawn by the work, summarized results and future work to be done in this area is presented.

CHAPTER TWO
DIGITAL TWINS FOR SOLAR PHOTOVOLTAIC PLANTS

Introduction

In recent years, the modern electric power distribution system (MEPDS) has undergone significant changes due to the increased penetration of distributed energy resources (DERs) and the growing presence of electric vehicles (EVs) with grid charging capabilities. These developments have brought about planning and design challenges to regulators and energy market changes as well as new operational and management challenges for and distribution system operators. Among the many complexities that arise from DER inclusion, renewable energy sources, particularly photovoltaic (PV) such as Distributed PV (DPV) systems, from rooftop solar to utility scale installations have played a crucial role [8].

By harnessing data-driven models and artificial intelligence-based approaches, DTs, or high-fidelity virtual replicas of the system or system assets can offer a comprehensive, intelligent support infrastructure for PV plant integration in the MEPDS. MEPDS DTs can facilitate real-time system observability, condition monitoring, fault recognition, and predictive analysis [9]. DTs enhanced with AI capabilities can leverage DT models as platforms for AI based DT services supporting MEPDS operational and management tasks, and further enable cognition in the MEPDS. The potential of Digital Twins in addressing operational and management challenges accompanying PV integration in the MEPDS is further explored.

MEPDS O&M Challenges

The traditional distribution system is in a transitional stage, with complexity and its associated needs growing exponentially alongside the incorporation of DERs in the system. New infrastructures such as enhanced sensing, decision making, intelligent control and optimized communication protocols thereby becomes integral in an MEPDS. These distribution system changes lead to added operational and management complexity explored below.

Transformation of the MEPDS

There is a progressive shift from traditional power distribution systems with unidirectional power flows to a modern electric power distribution system (MEPDS) with bidirectional power flows. MEPDS are characterized by their increased decentralization of power generation due to the penetration of DERs which further enables active distribution networks (ADNs) and microgrids. Supporting cyber physical system infrastructure is needed to handle this transition, which can be considered as 3C or Computational, Communication and Control frameworks. This includes intelligent information and communication technology (ICT) to support advanced metering infrastructure (AMI), micro-PMUs and PDCs, and accompanying optimal decision-making support such as DERMS and DRMS. With the inclusion of new technologies to support various MEPDS complexities, the system can support or integrate advanced functionality, enabling efficient, secure, reliable and resilient operation with improved power quality [10]. A diagrammatic representation of the MEPDS can be seen in Figure 2.1, adapted from [11].

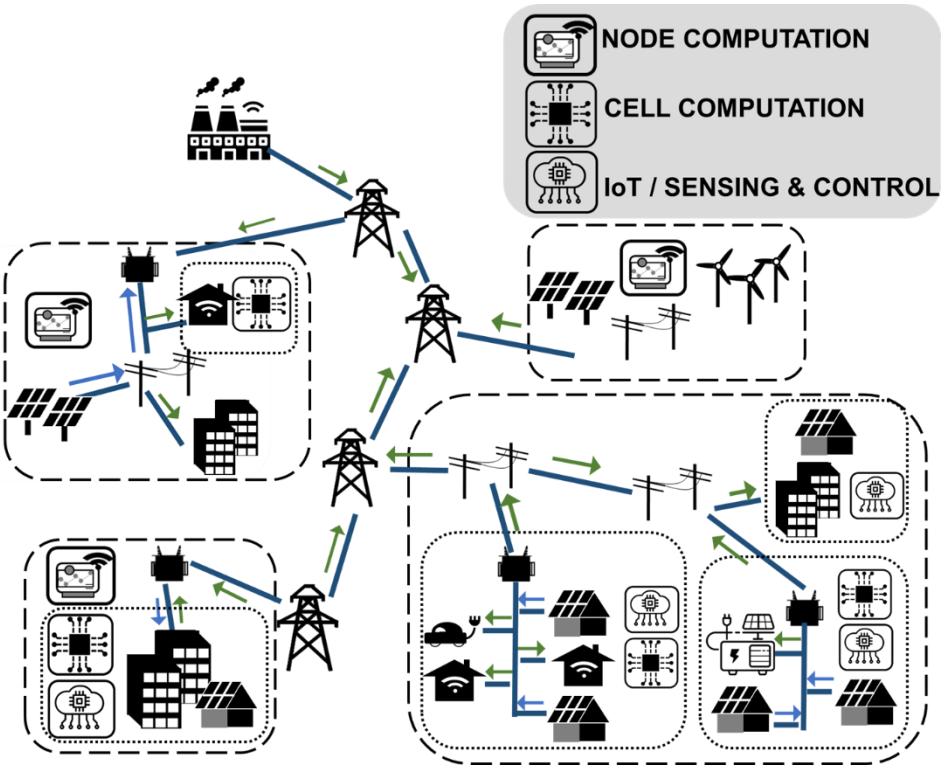


Figure 2.1 3C framework for MEPDS design, adapted from [11]

A large portion of MEPDS challenges thereby lie in DER penetration, of which distributed photovoltaics (DPV) is a large aspect. Photovoltaic (PV) plants present new challenges, particularly in the context of distributed PV integration, surrounding system monitoring, state estimation, control and overall operation and management. Typical challenges involve lack of system visibility, non-dispatchable nature and potential back-flow into transmission necessitating volt-var optimization and control, and distribution hardware upgrades driven by hosting capacity studies driven by understanding PV plant temporospatial dynamics [12].

Achieving real-time system monitoring and dynamic state estimation is challenging due to the incorporation of DPV sources. DPVs can introduce bi-directional power flow and can hinder visibility of on-site network parameters. Significant cloud cover leading to power generation drops over 70% of peak expected power, can lead to sudden losses of PV generation, causing voltage fluctuations further leading to power quality issues [16], while DPV installations can additionally be behind the meter causing low visibility, i.e., impact of cloud cover and partial shading losses are bundled with home-owners' power usage. Additionally, there is a need to compensate for missed and unavailable data measurements while visualizing PV plant performance in real-time for performance analysis and potential degradation or faults in the system. DPV estimation and forecasting is needed to ensure accurate monitoring and estimation using advanced sensing, monitoring, and communication protocols and new technologies.

Optimization and decision-making frameworks and decision-making for Distribution System Operators (DSOs) are crucial to effectively support interconnected DER systems with high renewable penetration. This involves intelligent volt-var optimization strategies based on neural network driven methodologies [39], optimal placement, utilization and control of smart inverters and other VVO equipment such as on-load tap changers (OLTC), switchable capacitor banks, static var compensator (SVC), and monitoring equipment, changes in equipment hosting capacity as necessary, and usage of supporting infrastructure such as centralized or decentralized Distributed Energy Resource Management

Systems (DERMS) [40] to manage and optimize reactive voltage injection and prevent network congestion.

Long-term O&M challenges accompanying PV integration include reliability estimation, optimal corrective and preventative actions, maintenance planning, and integration of legacy systems with newer equipment, line or feeder upgrades and voltage regulation and monitoring equipment such as smart inverters, D-FACTs, SVCs and other devices to support utility-scale PV installations at the distribution side [13]. Resilience evaluation of the MEPDS, particularly of feeders and feeder configuration, considering PV production in scenarios such as light load and peak PV production, voltage regulation during severe cloud cover impact, and asset management constraints such as power transfer management and microgrid stability once PV is integrated are crucial considerations [14].

Finally, in a system where DERs, microgrids, and Active Distribution Networks (ADNs) coexist with high penetration of PVs, short and day-ahead forecasts of PV power are needed to support intra-day and day ahead trading in energy markets [15]. The MEPDS needs to accommodate multiple stakeholders and prosumers, ensure enhanced system security through access controls, and establish optimal regulatory requirements to balance revenue generation, power trading, and operational costs while maintaining optimal power dispatch.

Digital Twins – Definition and Architecture

The DT was originally devised as a parallel virtual space and a simulation tool in the field of product development and testing throughout its lifecycle. They have now evolved far beyond its original use case in product development, and are increasingly adopted in various industries such as the aerospace industry [18], automobile manufacturing and in the power industry. A Digital Twin is defined as a virtual model or representation of a physical system, asset or process enhanced with data connections enabling the transfer of data insights and process data from the virtual representation in real-time (RT) back to the physical system. The DT has three major components: the physical system, the virtual model or “twin” and a constant real-time informational or control flow exchange, enabling “twinning” of the digital model

with the changes in the physical system. It thus acts as a self-evolving, multi-scale, multi-physical and hierarchical digitized simulation of a system or processes in real-world, dealing with both real-time data as well as historical data [19].

DT models can be as simplistic as a 3D system representation to complex, high-fidelity twins of operational vehicles, for instance. The virtual model developed can be self-evolving, multi-scale, multi-physical and hierarchical digitized simulation of a system, asset or processes in the real-world, dealing with both real-time data as well as historical data. DTs therefore can be classified on various levels and aspects with various user-defined nomenclatures on aspects ranging from the level of synchronization and fidelity with the physical system, time of creation of the DT, application type, and level of maturity[20].

DTs can be built using physics or mathematical based models, as well as using data-driven approaches such as Artificial Intelligence (AI) paradigms. The utilization of artificial intelligence (AI) paradigms for DT development, known as a data-driven DT [21] can significantly enhance the value-add of DTs while providing adaptivity, efficiency, reliability, scalability, memory, reasoning and perceptiveness of the linked physical system. These DTs enhanced with AI functionality are also known as Cognitive Digital Twins (CDTs). CDTs thereby interact with, learn and adapt to the physical system, as well as enable separate DT model entities to be linked together for enhanced O&M functionality [22].

In the context of an MEPDS, two types of DTs can be built, when categorized based on the physical system:

- *System DTs* or Digital Twins replicating the overall MEPDS behavior at various time-scales, levels of fidelity such as DTs of distribution network topology [25], network parameters such as current, voltage, frequency and active and reactive power [27] and so on
- *Asset DTs* or Digital Twins replicating MEPDS assets, such as DTs of inverter dynamics [28], PV plants [29],[33], HVAC load models of a building, energy storage system capacity modeling and degradation, etc.

Digital Twins and the MEPDS

A cohesive, constantly expanding digital entity or entities that reflects MEPDS complexity and leverages expanding, real-time flow of data in the system to effectively enable or improve grid long-term and short-term O&M operations is a valuable proposition. DTs that can mimic complex real-world systems, its operation, processes and behavior offer a promising means for an overarching intelligent support infrastructure, simulating system behavior, and a platform for facilitating computationally intelligent tasks and grid edge solutions in the MEPDS [23]. The availability of a digital model, or DT of distribution system entities and processes enables real-time system monitoring at required time-scales, data maintenance and data synchronization of archived data and processes, condition monitoring, fault recognition, alert generation, granularized views and enhanced visualization and predictive analysis of the MEPDS system and its assets [24].

Furthermore, utilization of a data-driven DT or AI-based simulated virtual replicas of a power system and associated processes can provide a cohesive platform or environment for further building of advanced AI based applications supporting MEPDS operation and management. Traditional DT services include:

- Real-time condition and health monitoring of system and system assets
- Predictive analysis
- Data collection and archiving
- System and equipment fault recognition, alert generation
- Dynamic state estimation and advanced observability

DT services can be deployed onto traditional DTs to utilize collected data and develop new inferences. DT services be developed using AI paradigms, creating a Cognitive Digital Twin (CDT). AI-based DT services built onto an MEPDS can include:

- Volt-var optimization and control

- Asset planning and allocation
- Optimal decision support and parallel scenario testing
- Distribution network or equipment fault detection
- System topology monitoring
- Optimizable load and source forecasting
- Parallel scenario-testing
- Creation of synthetic data and missed-data imputation
- Distribution network fault identification, prediction and classification

Exports of entity and process behavior from CDT and DT models provide valuable analyses, for instance, this data can be useful for business intelligence and reporting, energy markets, prosumers and planning operations. The DT thereby works in real-time to support and facilitate operation and management tasks in what will consequently evolve into a highly controllable, scalable, resilient and adaptable cyber physical MEPDS.

Need for data-driven Digital Twins

Designing DTs for an evolving, multi-stakeholder MEPDS system with high renewable penetration, accompanying complex behavior and process modeling requirements becomes highly computationally inefficient and resource expensive.

In MEPDS systems with high penetrations of DPVs, physics-based or mathematical-based DT modeling will need to account for stochastic or non-linear behavior dependent on environment processes ranging from cloud cover variation, wind speed changes, shading losses, bundled behind-the-meter readings where load usage is to be predicted, seasonal variation, humidity, as well as PV panel degradation and aging, and other accompanying variables. Additionally, given the small-scale of rooftop PV installations for instance, intensive modeling and setting up accompanying measurement equipment and management of data streams may not be economically viable.

Therefore, the accompanying costs and resources required to develop DTs for multi-dimensional and multi-stakeholder systems that are to be scalable, highly computationally efficient, multi-faceted, and cyber-secure, while serving the needs of multiple stakeholders, as in the case of MEPDS DTs, are exponential high. The incorporation of DT services can be another unwieldy piece in this puzzle, where facilitation of these services requires additional management and monitoring of service functionality by the respective stake-holder.

Complexity mitigation and efficiency improvement of the DT design process can be accomplished through the utilization of data-driven DT models. DT services and applications can similarly be enhanced through the incorporation of AI based paradigms. AI paradigms enable function approximation and reasonably accurate estimation of non-linear and stochastic processes, require little maintenance and are easily adaptable and scalable compared to mathematical or physics-based DTs which may be system specific.

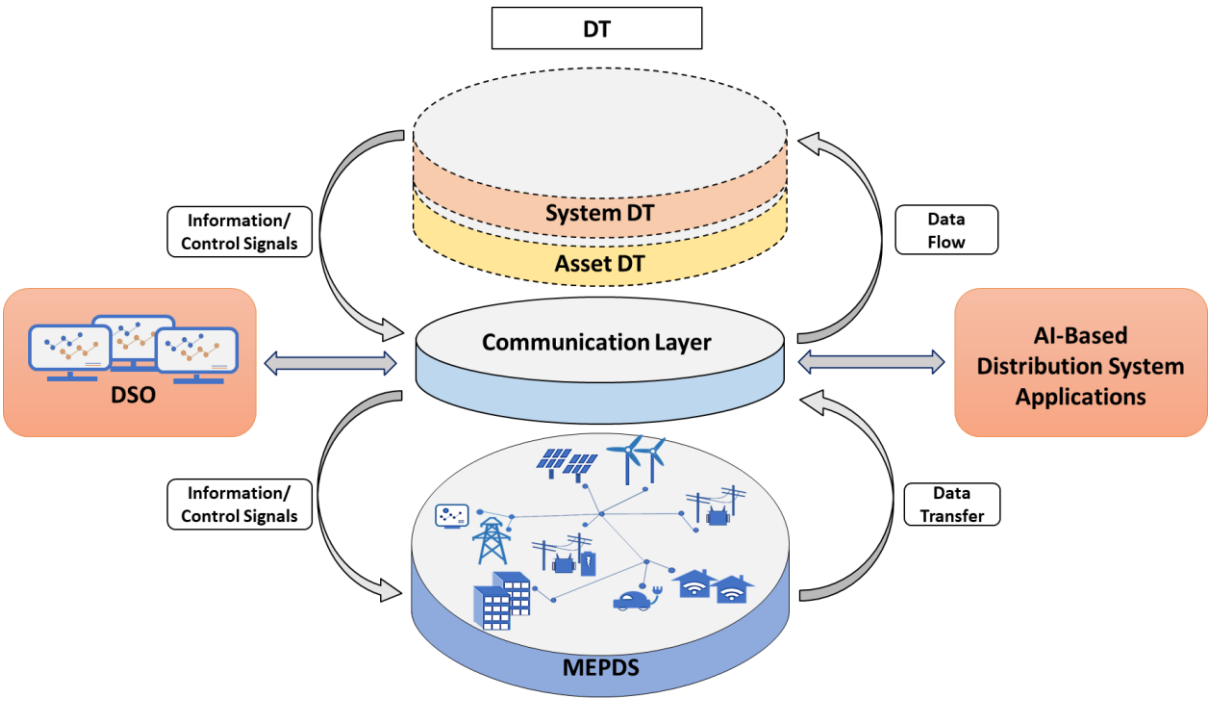


Figure 2.2 Digital twin and AI collaboration for enhanced MEPDS operations and management

Additionally, there is a unique, synergistic effect to utilizing data-driven DTs as a platform to deploy AI-based DT services in an MEPDS. Real-time data streams are available in data-driven DT

environments, that are typically pre-processed, synchronized and in an appropriate format, and are additionally archivable as needed, in comparison to physics-based DTs which are typically built onto separate modeling environments with their own proprietary tools such as MATLAB, RSCAD (RTDS) and the like. Applications built supporting MEPDS O&M tasks such as DERMS, DRMS or BI applications can easily utilize this data for greater insight of the system. Inferences made using these DT can also guide DSO or regulator decision making in real-time, or help with distribution system studies. The utilization of data-driven DTs and CDTs therefore facilitates high-quality data and an effective platform to develop plug-and-play tools or enhanced MEPDS support applications

DTs for Solar Photovoltaic plants

Much of current changes around the MEPDS can be directly attributed to the increased penetration of DERs at the distribution level along with additional impact of electric vehicles (EVs) and grid charging, leading to new O&M challenges. A large portion of this DER complexity can be attributed to renewable inclusion, especially photovoltaic (PV) and Distributed PV (DPV) penetration in the MEPDS. The integration of PV plants presents various operational and management challenges due to the non-dispatchable nature and associated operational and management issues arising from renewable penetration. This necessitates volt-var control and the inclusion of distribution automation technologies and distribution system upgrades to improve power quality, system reliability and resilience, enable effective monitoring, fault recognition, and predictive analysis. New standards have been developed to deal with DER interconnection, such as FERC 2222 and IEEE 1547.9-2022, permanently changing the distribution system landscape in the United States.

A cohesive, constantly expanding digital entity or entities that reflects, in real-time, PV plant complexity and leverages expanding, real-time flow of data in the system to effectively support or optimize PV plant operational and control challenges and other long-term operations from planning and design to asset management and maintenance becomes a valuable proposition. DTs of distributed PV

sources and PV assets enables real-time system observability and monitoring at required time-scales, fault recognition and analysis, and advanced distribution system studies and analysis.

The operational and management challenges faced by photovoltaic plants in the MEPDS can be effectively addressed through the utilization of DTs, with a particular focus on data-driven models and AI-based approaches. By enabling real-time system observability, fault recognition, and analysis, as well as supporting advanced distribution system studies, DTs offer significant potential for optimizing the performance of PV plants and enhancing the integration of renewable energy sources within the modern electric power distribution system.

Digital Twin Applications

Digital models of the system, system assets and other entities itself provides support for DPV integration, operation and management in the MEPDS. Advanced DT services, or applications supporting various O&M tasks can then be built or supported by information from DTs. This includes specific operations of PV plants such as real time monitoring and control, dynamic state estimation, volt-var optimization and decision support, long and short-term forecasting, resilience and reliability testing, power quality improvement among others, that are usually modeled or optimized using AI paradigms. The DT therefore behaves as an optimized, dynamic, scalable and flexible environment for services or applications, testing and implementation of various AI enabled operation and management tasks.

Many implementations of DTs [28]-[52] and prototypal-DTs [39] currently exist in power distribution systems with AI-based applications. They support specific distribution system O&M needs surrounding PV integration and DPV or PV plant management at the distribution level. These are categorized in Table 2.1 and further explored below.

DTs can be used for enhanced state estimation, or modeling system and system asset dynamics. In [29] AI and Digital Twin models were used to model asset DTs of PV panels, DC-DC converter and the final system, using IoT data to create a PV CDT capable of monitoring PV plant dynamics, utilized for condition monitoring of the system. DTs of regional multiple energy systems (RMES) on CloudPSS are

developed and tested on a PV system, where power flow and electromagnetic transient (EMT) models are utilized to monitor distribution system parameters and health in [30]. In [31], genetic algorithms and digital twins were used to identify PV dynamics of neighboring stations with weather data, using simple parameters such as tilt, albedo, solar irradiance and the like, creating a self-learning DT model. In [32] a DT of the maximum power point tracking (MPPT) algorithm was developed using mathematical approaches to model shading losses at each PV string, optimized using an Artificial Neural Network (ANN).

In [28], a DT of a distribution power transformer was used for state estimation of medium and low voltages in distribution feeders using waveform monitoring. In [41], A DT consisting of OpenDSS based time-series feeder measurements and particle swarm optimization (PSO) was utilized for dynamic state estimation and setting of optimized DER reactive power setpoints in the system.

In [40] a novel DERMS solution is proposed by adopting the real-time optimal power flow (OPF) for coordinated control of the distributed PV inverters in a real-time manner, creating a prototypal DT. In [45], a real-time digital simulator (RTDS) was used to create a DT of distribution power flow, with data imported with GridVis.

A large portion of DT applications deals with volt-var optimization. For instance, a DNN approach has been used for volt-var optimization in [36]. In [37], a DT was utilized to create an Automatic Voltage Regulation (AVR) hierarchical coordinated control strategy for PV inverters to keep voltages in low-voltage (LV) distribution grids within specified limits during cloud cover and other scenarios affecting PV power generation. In [39] cellular computational networks (CCN) was used to create a data-driven DT for hierarchical DER optimization, at the individual node, zone level and at the distribution system level.

DTs have been used for PV forecasting for a distribution network in [17]. In [53], a DT based day-ahead integrated energy system scheduling under load and renewable energy uncertainties using deep neural networks. and decision support [20] using IoTs, edge computing and cloud architectures.

An online analysis digital twin (OADT) approach is seen in [34], developed to support the realization of a new power grid online analysis solution architecture for DERs, and includes the usage of DTs to aid DSOs by automating distribution rules.

In [50] fog computing and IoT networks were used to develop a data-driven DT for probabilistic solar and load forecasting to improve system reliability.

A DT based Power-Hardware-in-the-Loop (PHiL) was utilized for distribution network topology assessment and fault detection in DERs in [35]. In [44] and [46] DTs of DPVs were used for fault identification and condition monitoring of the system.

Table 2. 1 DT APPLICATIONS IN MEPDS O&M TASKS

O&M Task(s)	***DT Types***	***Application(s)***
Modeling	Asset And System DTs	Situational Awareness [29], PV asset modeling [30], PV generation imputation [31], MPPT optimization [32]
Monitoring And Estimation	Asset DT	Distribution network [28], Dynamic state estimation and DER setpoint optimization[41]
Power Flow Analysis	System DT	Power flow calculation for PV inverter control [40], Power Flow Calculations [45]
DERMS Support	System DT	Distribution network VVO [36], [39], Automatic voltage regulation for PV inverters [37]
DRMS Optimization	System DT	Load and Energy Management [34]
Forecasting	Asset DT	Demand [48], Network Reliability [50]
Fault Diagnosis	Asset DT	Distribution network topology assessment[35], DPV fault diagnosis [44],[46]

DTs thus enable artificial intelligence paradigms that allow for increased efficiency and optimized control, reduction in process redundancy and costs involved with incorporation of automation and algorithmic improvements and standardized data processing across systems or management levels, while significantly reducing risk in planning and optimization tasks. The dual inclusion of digital twin (DT) technology paired with AI frameworks for DT modeling in MEPDS and DPVs is thereby highly advantageous

Summary

DTs are highly advantageous as supporting infrastructure for various MEPDS and DPV O&M tasks. Additionally, data-driven digital twins, or DTs developed using AI paradigms are highly suitable to employ in modeling systems dependent on non-linear or stochastic environmental processes such as PVs and therefore in an MEPDS with high PV penetration. AI methodologies further enable DT modelers to optimally model the virtual space, reduce time-to-market, and create computationally efficient virtual models.

This shows that through the usage of AI algorithms in both the design phase of an MEPDS DT and in the development of advanced functionalities during the operational and working phase of the DT, DT and AI become enabling technologies for each other, facilitating grid-edge solutions and stake-holder needs in an MEPDS, and further enabling cognition in the system.

CHAPTER THREE

PV PLANT DIGITAL TWIN

Introduction

As detailed in chapter 2, data-driven digital twin models can be utilized to deploy advanced MEPDS operations. Towards this purpose, computationally efficient, scalable neural network architectures such as Multi-Layer Perceptrons (MLP) and Echo State Networks (ESN) have been utilized to develop asset digital twins for estimation and prediction of PV plant power outputs. A 1 MW PV plant at Clemson University is utilized for these studies. The modeling, optimization and testing results of the PV DTs are further explored in this section.

Digital Twins for PV Plants

A DT of power distribution system assets such as generators, loads, transformers and DER sources can enable generation of realistic datasets of a system over a wide range of scenarios and further act as a valuable aid for estimating and predicting the behavior of a complex system for power system studies. For instance, volt-var optimization, energy management, planning studies, demand response studies and further forecasting, analysis and optimization of the physical system can be facilitated via DTs of the system integrated into testbeds or RTDS systems.

Traditionally, physics-based [59], mathematical approaches [60], data-driven [61] or hybrid combinations [62], [63] of those methods can be used for creating DTs of DERs, in particular, of PV plants. Additionally, different PV plant entities, various levels of fidelity and varying time-scales can be modeled using these DTs based on the use-case requirements.

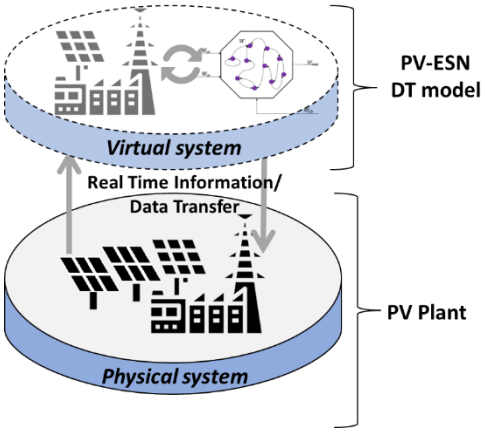


Figure 3.1. Digital Twin components for a PV plant

Some distribution system assets, in particular DERs such as PV plants or wind turbines are complex in nature to model using physics-based approaches, especially in real-time, as is the case with traditional DTs. As PV penetration or in particular distributed PV (DPV) continues to rise in shares with both count of installations and by capacity now connected to the grid [6], PV power estimation and prediction is necessary.

However, modeling the realistic behavior of PV DTs for various operational conditions and time-scales requires the utilization of extensive resources. Additionally, once implemented, physics-based models can be difficult to maintain, operate and update, especially when advanced functionality or MEPDS application support is necessary from the model. For instance, typical operations performed in distribution systems studies include combinations different configurations, placements and scales of PV plants, while additionally accounting for climatic or external factors that are a result of these changes.

A purely mathematical or physical model is therefore typically insufficient to account for growing system complexity and is time-consuming to model, while the current and future state and needs of the PV plant and MEPDS system as well as time dependent PV performance metrics such as panel ageing and degradation, shading losses and the like can be better captured by AI-paradigms, as detailed in chapter 2. Data-driven PV plant DTs can be used to support MEPDS planning operations or support a Distributed Energy Resource Management System (DERMS). PV DTs integrated into a distribution testbed can simulate system behavior due to addition, removal or scaling of new PV plants

in surrounding geographical areas, further allow estimation hosting capacity and network congestion, support reliability and resiliency testing and so on, all in real-time, faster-than-real-time or as time series, supporting quasi-static and time-series analyses. MEPDS O&M challenges such as volt-var optimization and scenario testing as well as simplistic operations load-relief and sizing analysis and the like can be simulated using PV DTs.

PV Plant System

A PV plant DT has been employed here to model PV power generation of an existing real-world PV-plant at a physical site, shown in Figure 3.1 and Figure 3.2. The physical site consists of a weather station which is a CR300 weather data logger that monitors and logs solar irradiance, temperature, wind speed and wind direction data for a second interval. This is placed accordingly to measure the input solar irradiance to a 1 MW PV array at Clemson University with corresponding solar power generation data streamed from micro-PMUs (μPMUs). This data is archived and timestamped in openHistorian, a back-office No-SQL database used for streaming, archiving and integrating process control and synchro phasor data for further analysis or real-time querying of power generation flows.

Figure 3.1 1 MW PV plant at Clemson University's R-06 carpark

Figure 3.2 CR300 Campbell Scientific Measurement and Control Datalogger

System environment for DT development

- The data-driven DT was developed on Python (3.9) and Jupyter Notebook.
- A high performance computing (HPC) environment was used, run on the Palmetto Cluster, consisting of 8 CPUs with 2 resource chunks and 30gb memory per chunk. 2 GPUs, V100, were used per CPU.
- Data collection of weather data was done using CR300 web API over Python and pre-processed for correct sample counts using Python's pandas library. Python's openHistorian library was used for reading data points archived on OpenHistorian's noSQL database.
- Real-time streaming was done over TCP/IP connections of weather station databases and openHistorian's localhost instance.
- Python packages used include numpy, pandas, pyESN, scikit-learn, openHistorian, Dash and Plotly.
- Data from PMUs and CR300 was collected from February 27rd to May 16th , for the Clemson-Anderson-Pendleton region over spring, pre-processed to be a second interval. Unusable measurements and days with long periods of missing data were excluded from the training dataset.

DT architecture for PV power estimation using MLPs

DTs of DPVs, and further, of system assets are useful tools in power distribution system studies. The system behavior and actionable metrics related to a particular PV plant's behavior such as condition monitoring, fault recognition, predictive maintenance and other traditional DT tasks can be enabled through DTs of a system. This enables enhanced

monitoring and visibility of the asset, which helps in front-of-the-meter PV installations and of the distribution system as a whole.

Modeling PV plant behavior has its own challenges. A characteristic feature of DERs and particularly PV plants, is the system's non-linear behavior. The equation defining typical PV power generated at time instant *t* is seen in equation 3.1, where G_{std} and T_{std} are the standard test conditions for solar radiation and cell temperature, respectively, and α_T is the manufacturer temperature coefficient of a PV module. PV generation at a site therefore is dependent on temperature and solar irradiance received at the site over time, and panel material which defines panel degradation over time as well as PV panel installation size, tilt and type. Solar irradiance falling on the panel is additionally impacted by external variables affecting shading losses such as geospatial and atmospheric conditions, which further impacts PV plant performance.

$$PV(t) = P_{Peak} \frac{G(t)}{Gstd} - \alpha_T[Tc(t) - T_{std}] \qquad (3.1)$$

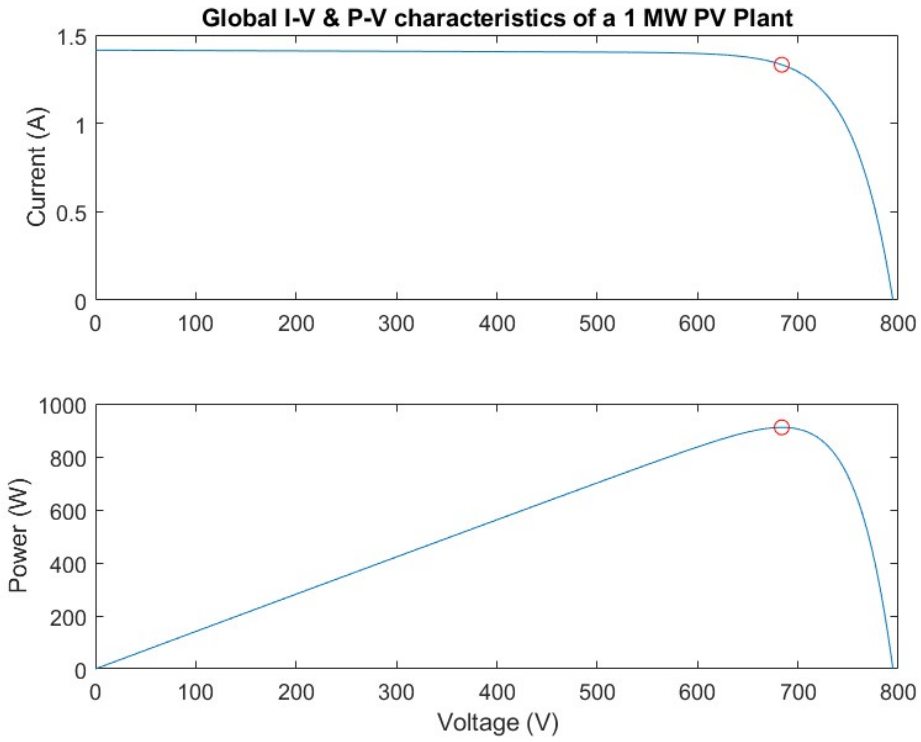


Figure 3. 3 Ideal I-V and P-V characteristic curves at 480 W/m2 irradiance

An MLP is used to model system behavior and estimate PV plant performance in a manner that reflects changing plant performance in near real-time. The MLP model is fed real-time data and is retrained appropriately to reflect changing PV plant conditions in real time.

MLPs are feedforward neural networks that typically enable informational flow in one direction. MLPs have input layers, one or more hidden layers containing neurons with trainable and optimizable “weights”, and output layers. From the name, MLPs consist of neurons, or “perceptrons”. These perceptrons apply weighted sums to input samples that are passed through to an activation function to finally produce an output. These outputs can be then fed as inputs to the neurons in the subsequent layer, if more than one hidden layer is present. The architecture of the MLP used for data-driven DT modelling of PV power estimation is seen in Figure 3. 4. The equation used to train the MLP is in equation 3.2, where

$y(t)$ is the output, obtained through tuning output weights W_{out}, reference weights W_{ref} against an activation function $h(t)$ and transfer function $r(t)$

$$y(t)=f(W_{out}h(t)+W_{ref}r(t)). \quad (3.2)$$

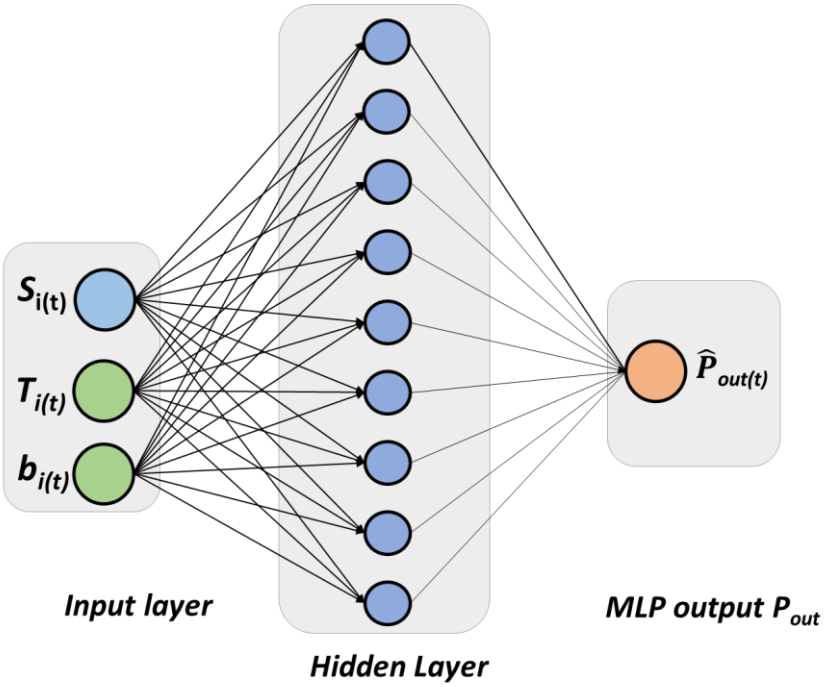


Figure 3. 4 MLP Model for PV power estimation at current time interval

The weights and biases of the MLP are further adjusted iteratively using a process called backpropagation, enabling learning. Backpropagation calculates the gradient of the loss function with respect to the weights and biases and updates them in a way that minimizes the error between the predicted output and the true output. This process is typically performed using optimization techniques such as stochastic gradient descent (SGD) or its variants.

Python's scikit-learn module is used to define the MLP architecture, here "MLPRegressor". A single layer MLP, with a logistic or "sigmoid" activation function, with neurons ranging from 10 to 25, learning rates of 0.04 to 0.001, bias of 1 and quasi-newton

solver Limited-memory BFGS was used to estimate the output at the 1 MW PV plant, R06. The training function is set such that up to 5000 epochs can be used, with early stopping and validation practiced when value has converged to <0.0001 tolerance. The parameters were chosen for optimal computational efficiency with acceptable accuracy, enabling near-real-time DT estimations of PV power. The results of a grid search for optimal hyperparameters can be seen in Table 3.1.

Table 3.1 Tuned Hyperparameters for best DT performance

Model Category	***Day Category***	***Neurons(n)***	***Activation function***	***Solver***	***Learning rate (α)***
Estimation	Sunny	10	Logistic	LM-BFGS	0.04
Estimation	Moderately cloudy	20	Logistic	LM-BFGS	0.002
Estimation	Cloudy	20	Logistic	LM-BFGS	0.001

Three MLPs are separately trained offline on historical data, categorized by daily cumulative solar irradiance received at R06 into “cloudy”, “moderately cloudy” and “sunny” days. Input data used as features for the network are solar irradiance, temperature and bias, with a measurement interval of one second. These MLPs are used for near-real-time estimation of PV plant performance and are therefore PV plant estimation DTs that can be optimized for increased system complexity and knowledge depth, as well as granularity required.

Online training and predictions are then done to utilize the PV DT for estimation as in Figure 3.5. Real-time data obtained from the weather data monitor at the PV plant is used as inputs to each MLP model, pre-trained on historical data, estimating PV power output at the current time interval in real-time for each day category. The MSE of MLP predictions from each day category is compared and monitored at every minute to ensure the selected MLP model category's estimation accuracy or MSE does not fall below 98%. If it does, the MLP model is switched to the model with the lowest MSE amongst the three. The predictions with the lowest MSE is visualized, archived and finally averaged to identify the type of cloud coverage over the day.

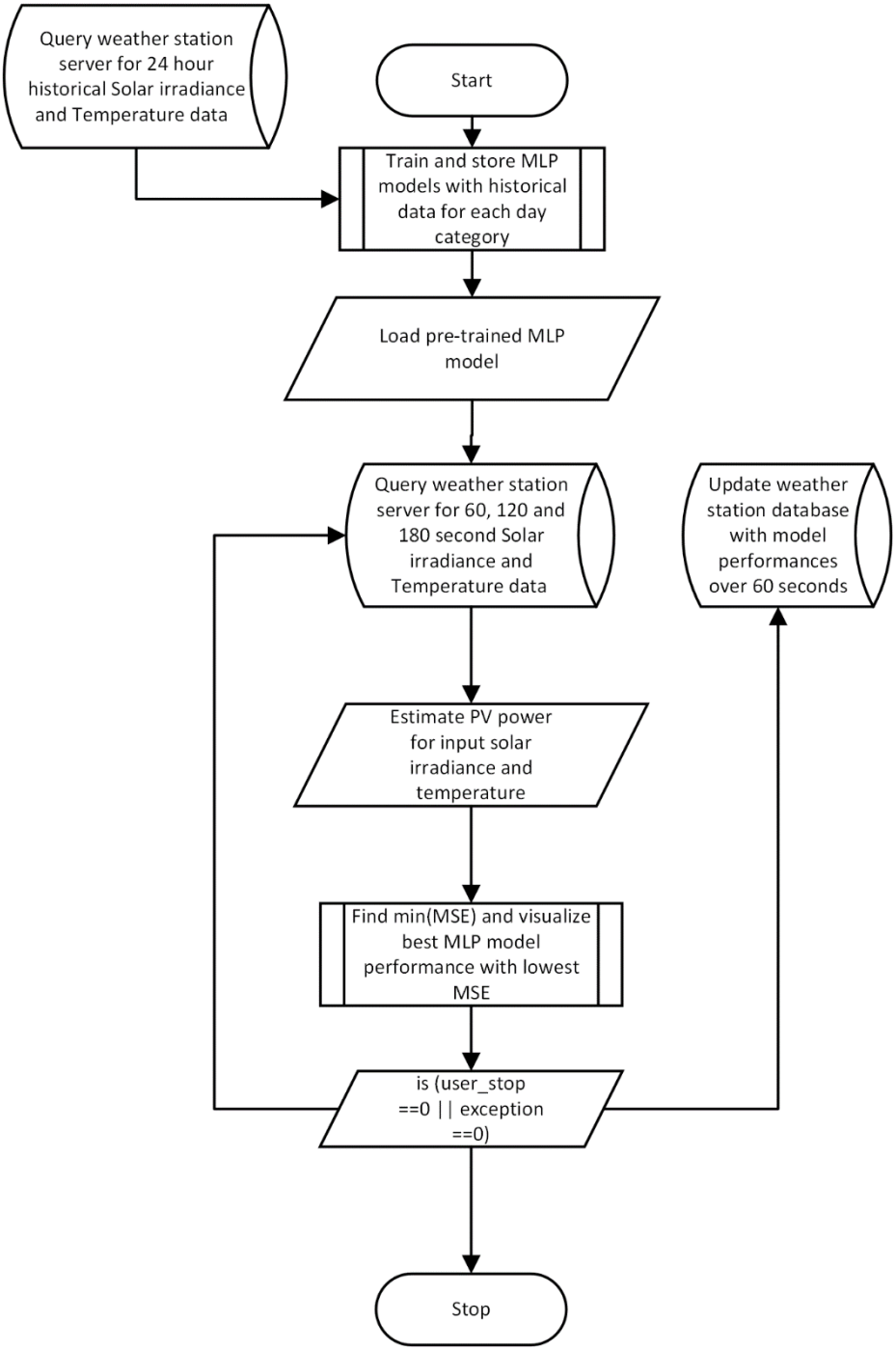


Figure 3. 5 MLP Real-Time estimation flowchart

DT architecture for PV power prediction using ESNs

Short term PV power forecasts can be useful for DSO needs and applications such as cloud cover impact identification, inverter voltage regulation and volt-var optimization and

fault prediction. To effectively model plant performance in real-time and faster-than-real-time, ESNs were used.

An ESN is a form of a recurrent neural network (RNN), created by Herbert Jaeger as a paradigm falling under the category of reservoir computing. The ESN's architecture allows for quick solutions of non-linear, complex function approximations, along with the advantage of simpler training processes and lower computational requirements compared to a traditional RNN. Since ESNs are adapted from RNN concepts, ESNs have a type of memory or reservoir that retains an internal state that holds the non-linear transformation of previous input states. ESN architectures enable modeling non-linear or stochastic system behavior by mapping low dimension input values into high dimensional spaces, capturing the temporal patterns of a system [64].

The architecture of the ESNs, shown in Figure 3. 6, consists of two main features, the reservoir weights and the readout weights. The ESN reservoir consists of randomly initialized sparsely connected synaptic neurons, connected to other weights and input weights in its hidden layer. Its weights are randomly assigned and fixed. The second ESN feature is the readout layer. The connection between the reservoir neurons and the outputs are called readout weights and are adapted during the training of the network to give the ESN outputs the best approximation of the measured target values. The read-out layer thereby decodes the reservoir activations, and furthermore can be trained using linear regression without backpropagation as in a typical feed-forward neural network, with activations that can be fed back to the reservoir, as in the case of a feedback connection.

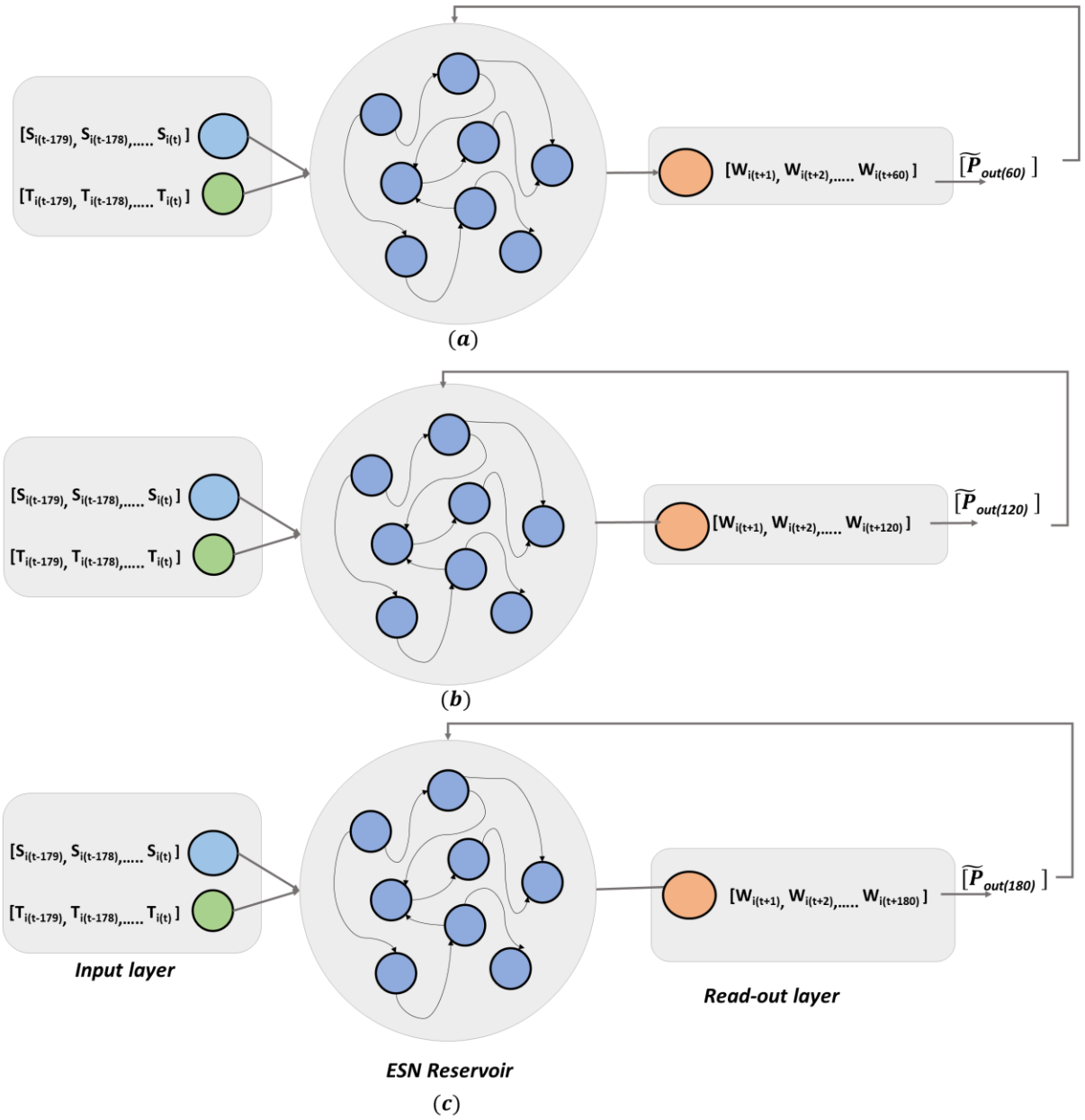


Figure 3. 6 ESN Model Architecture for PV power prediction for 60 (a), 120 (b) and 180 (c) seconds ahead

All these factors allow ESN to be frequently employed in AI models for time series data prediction with accompanying inherent uncertainty such as PV power generation [65]. For PV power prediction, two parameters are needed to accurately model the PV array's power generation or approximate the equation in 3.3, the solar irradiance and cell temperature.

$$x(n+1)=f(W_x(n)+W_{in}u(n+1)+W_{fb}y(n)) \qquad (3.3)$$

The equation for ESNs can be represented in Equation 3.2 where $x(n)$ is the n-dimensional reservoir state, f is an activation function, W is the (n x n) reservoir weight matrix, W_{in} is the (n x k) input weight matrix, $u(n)$ is the k-dimensional input signal, W_{fb} is the (n x m) output feedback matrix, and $y(n)$ is the m-dimensional output signal.

For PV power prediction, three parameters are needed to accurately model the PV array's power generation; the solar irradiance, temperature, and the output power of the PV array. A training data set consisting of historical solar irradiance and temperature for 30 days over various operational conditions is collected from the physical system, categorized and utilized to train and test the accuracy of ESN operation. Training data is categorized into 3 main categories based on solar irradiance waveforms from cloud coverage; sunny, moderately cloudy and extremely cloudy. Each training data set category is used to fit a different ESN model, each with three different read-out weight sets allowing for the system to iteratively predict 60, 120 and 180 seconds ahead.

Three ESNs are trained separately with cloudy, moderately cloudy and sunny days datasets and are further utilized in the DT model. Each ESN structure consists of one input, 350 to 550 neurons in the reservoir and one output. Hyperparameters such as optimal leaking rates, or sparsity, and spectral radius were adaptively selected and optimized based on differing solar irradiance profiles, and weights for the ESN were generated randomly, using a random seed to prevent variation over trials. For prediction, 350, 500 and 550 neurons were identified to have the highest accuracy for all sunny, moderately cloudy and cloudy days. There is no introduction of noise to this model for prediction, however, a feedback was introduced.

Table 3.2 Tuned Hyperparameters for best DT performance

Model Category	***Day Category***	***Reservoir neurons (n)***	***Sparsity***	***Spectral Radius***	***Noise***	***Feedback***
Prediction	Sunny	350	0.2	0.3	0	True
Prediction	Moderately cloudy	500	0.1	0.25	0	True
Prediction	Cloudy	550	0.1	0.4	0	True

Online training and predictions are then done to realize the PV DT as in Figure 3.7. Real-time data obtained from the weather data monitor at the PV plant is used as inputs to each ESN model category, predicting PV power output for 60, 120 and 180 seconds ahead. The MSE of ESN predictions from each model category is compared and constantly monitored to make sure the selected ESN model category doesn't let prediction accuracy or MSE fall below 98% before being switched. The predictions with the lowest MSE, is visualized, archived and finally averaged to identify the type of cloud coverage over the day.

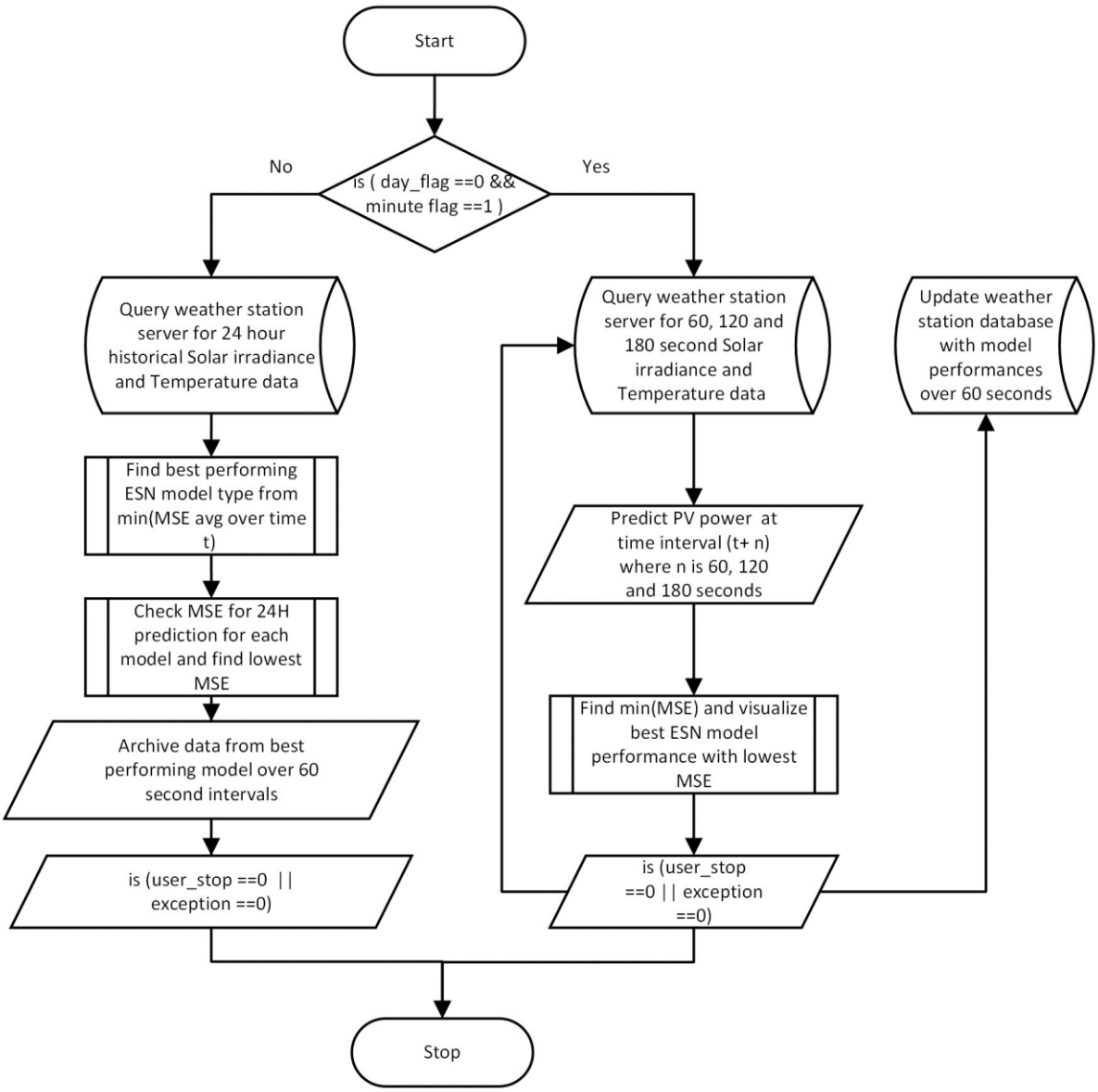


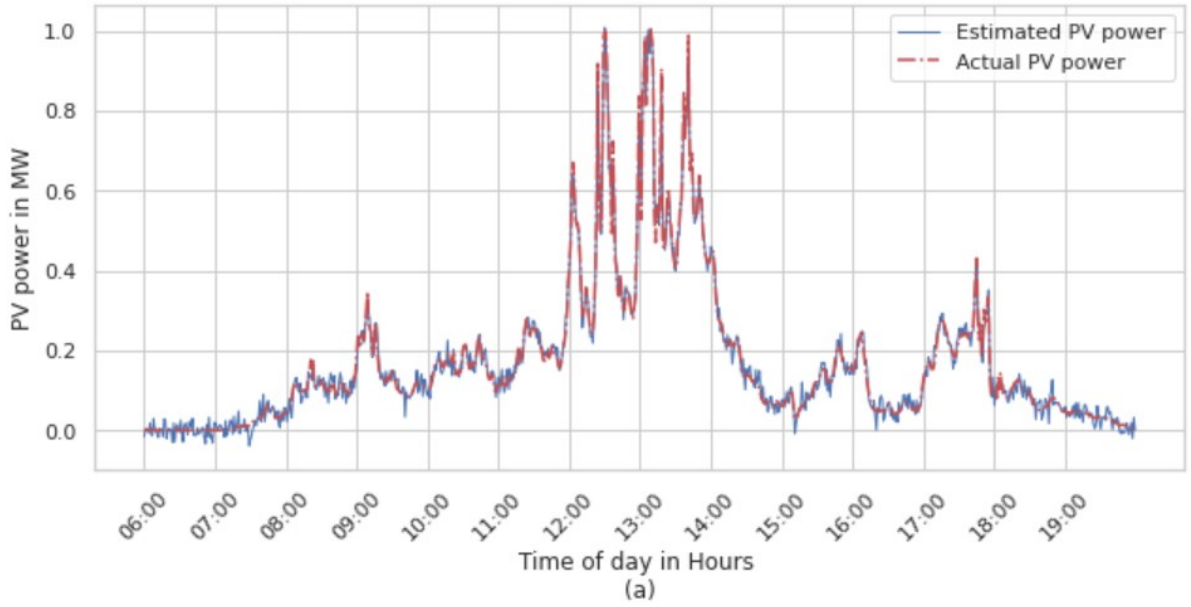
Figure 3. 7 ESN Real-Time prediction flowchart

Results and Discussion

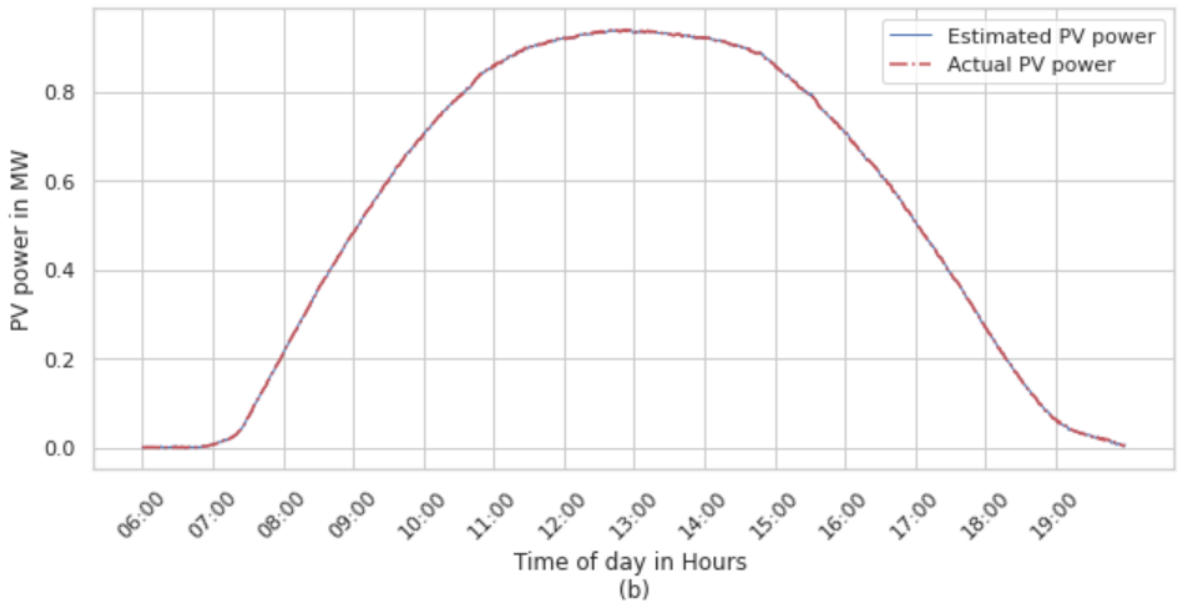
DT performance for PV power estimation

The MLP-DT for estimation is implemented on Python 3.9 and Jupyter Notebook using the sklearn library. A multi-layer perceptron regressor (MLP) model was used. The neurons, learning rate, solver and activation function has been selected to obtain maximum accuracy while maintaining near real-time estimation

The estimation PV DT is used to estimate or “twin” the real time PV performance at the 1 MW physical PV plant. The DT is run for online estimation over moderately cloudy, cloudy and sunny conditions. Results of the MLP model for three day categories with training and testing estimations are depicted in Figures 3.8 and 3.9.



(a)



(b)

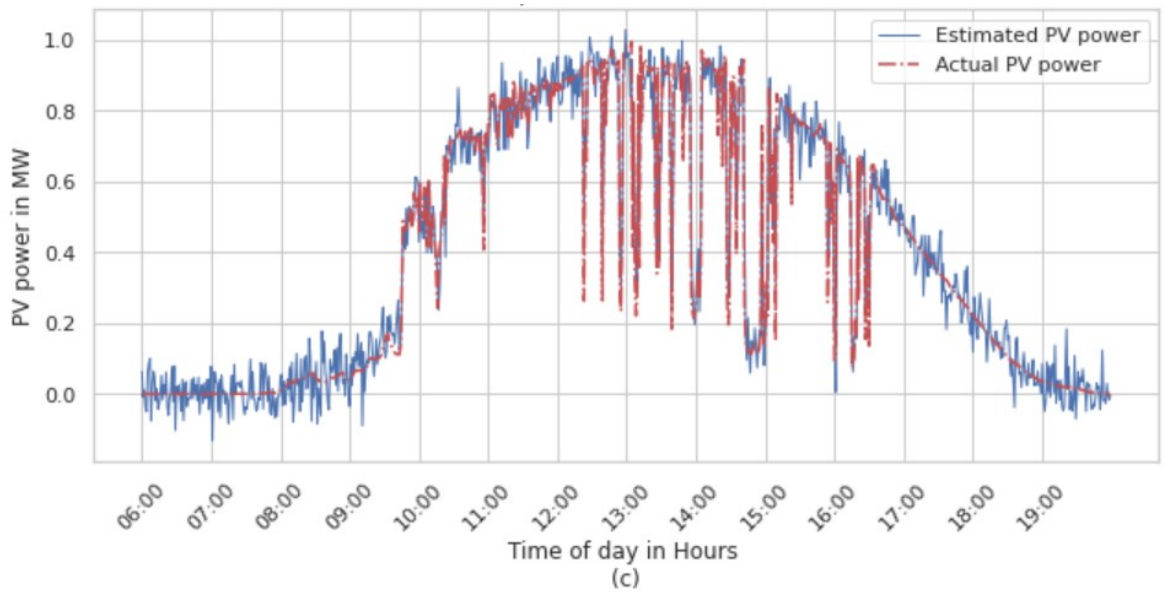
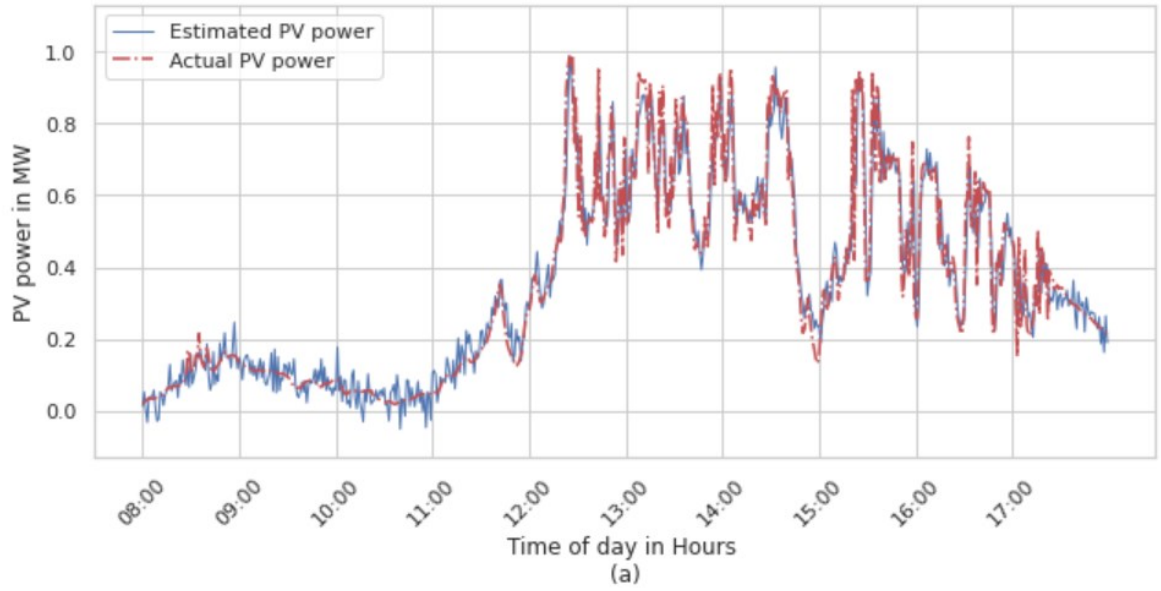
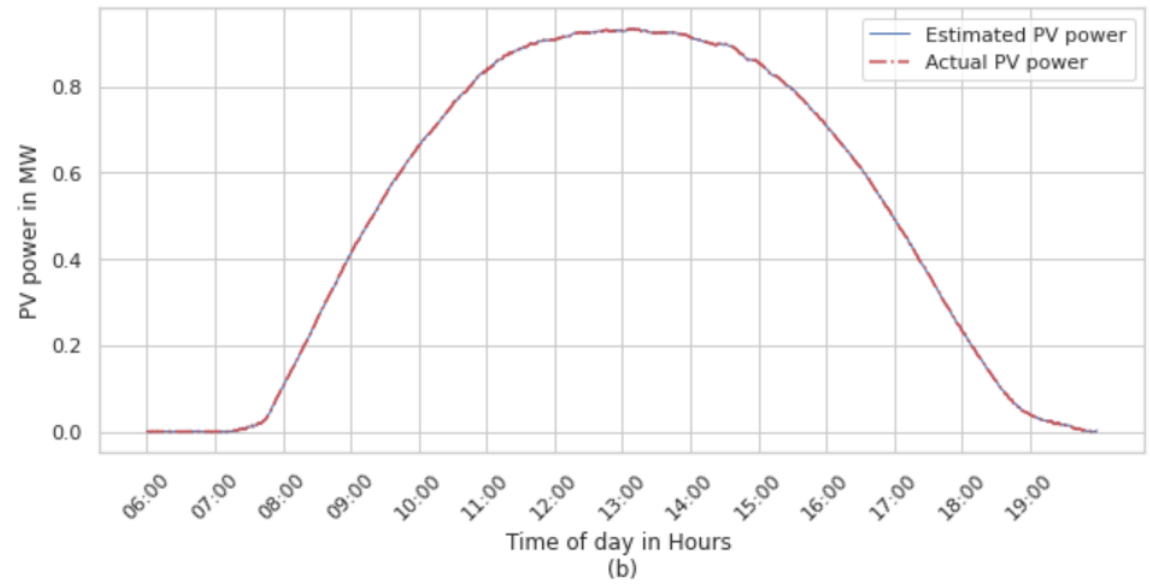


Figure 3. 8 DT estimation for cloudy (a), sunny (b) and moderately cloudy (c) days plotted for testing dataset





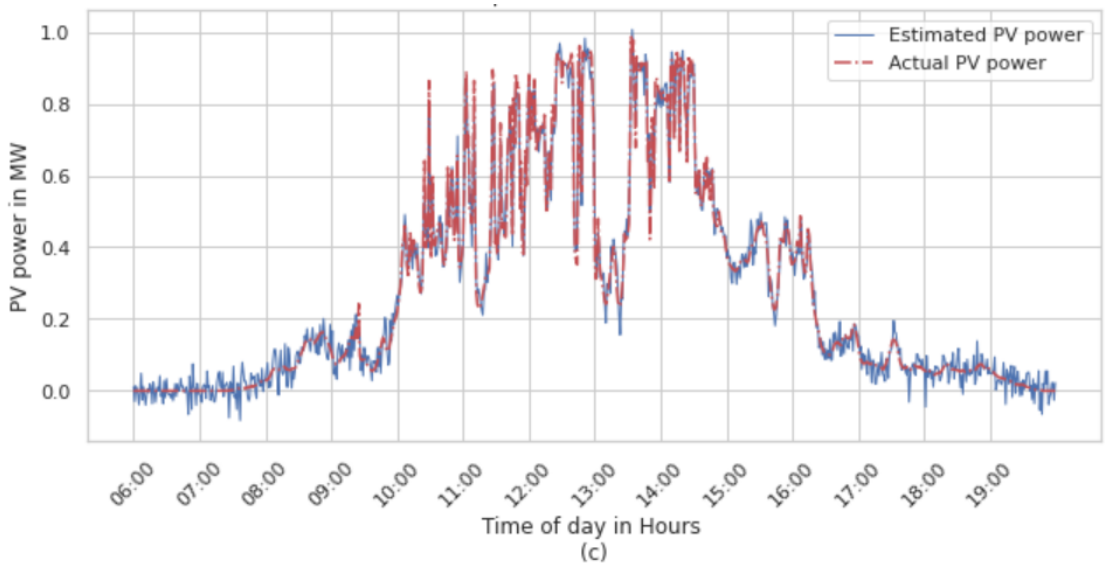


Figure 3. 9 DT estimation for cloudy (a), sunny (b) and moderately cloudy (c) days plotted for training dataset

$$MSE = \frac{1}{n}\sum_{i}^{n}(W_{actual(i)} - \widetilde{W}_{pred(i)})^2 \quad (3.4)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i}^{n}(W_{actual(i)} - \widetilde{W}_{pred(i)})^2} \quad (3.5)$$

$$MAE = \frac{1}{n}\sum_{i}^{n}\left|W_{actual(i)} - \widetilde{W}_{pred(i)}\right| \quad (3.6)$$

$$MAPE = \frac{1}{n}\sum_{i}^{n}\left|\frac{W_{actual(i)} - \widetilde{W}_{pred(i)}}{W_{actual(i)}}\right| \quad (3.7)$$

25 trials were performed to ensure DT consistency for each day category. Each model's output performance with accompanying neuron counts can be seen in Table 3.3, where mean squared error (*MSE*) (3.4) and root mean squared error (*RMSE*) (3.5), mean absolute error (*MAE*) (3.6), and mean absolute percentage error (*MAPE*) (3.7) of the predicted vs actual values during testing the model are calculated along with average prediction interval time.

Table 3.3 DT performance for PV power estimation

Day Category	***Dataset type***	***Neurons***	***MAE***	***MSE***	***RMSE***	***MAPE***
Sunny	Training	10	0.0002	0.0003	0.0173	0.0049
Moderately Cloudy	Training	20	0.0087	0.009	0.0949	0.0089
Cloudy	Training	20	0.0092	0.0098	0.0990	0.0135
Sunny	Testing	10	0.0003	0.0002	0.0141	0.0064
Moderately Cloudy	Testing	20	0.0075	0.0087	0.0933	0.0112
Cloudy	Testing	20	0.009	0.0095	0.0975	0.0142

DT performance for PV power prediction

The prediction PV DT is used for short term forecasts of the real-time PV power generation at the 1 MW physical PV plant, allowing PV DT monitoring in real-time or faster-than-real-time. The DT is tested for online prediction over moderately cloudy, cloudy and sunny conditions. Results of the ESN model for 60, 120 and 180 seconds ahead estimation for training dataset for each cloud category is depicted in Figures. 3.10, 3.11 and 3.12

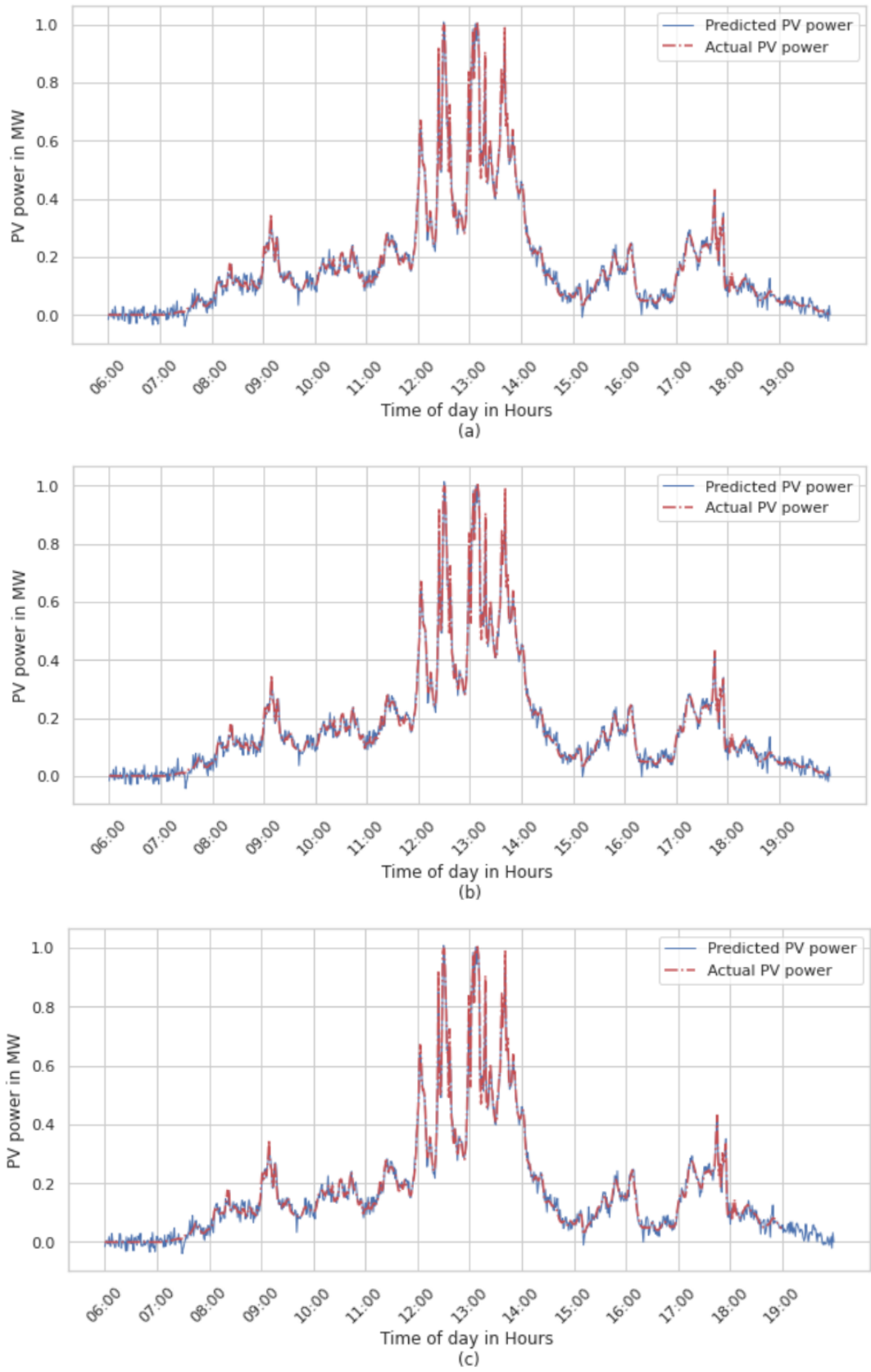


Figure 3. 10 DT prediction for 60 (a), 120 (b) and 180 (c) seconds ahead plotted for training dataset for a cloudy day

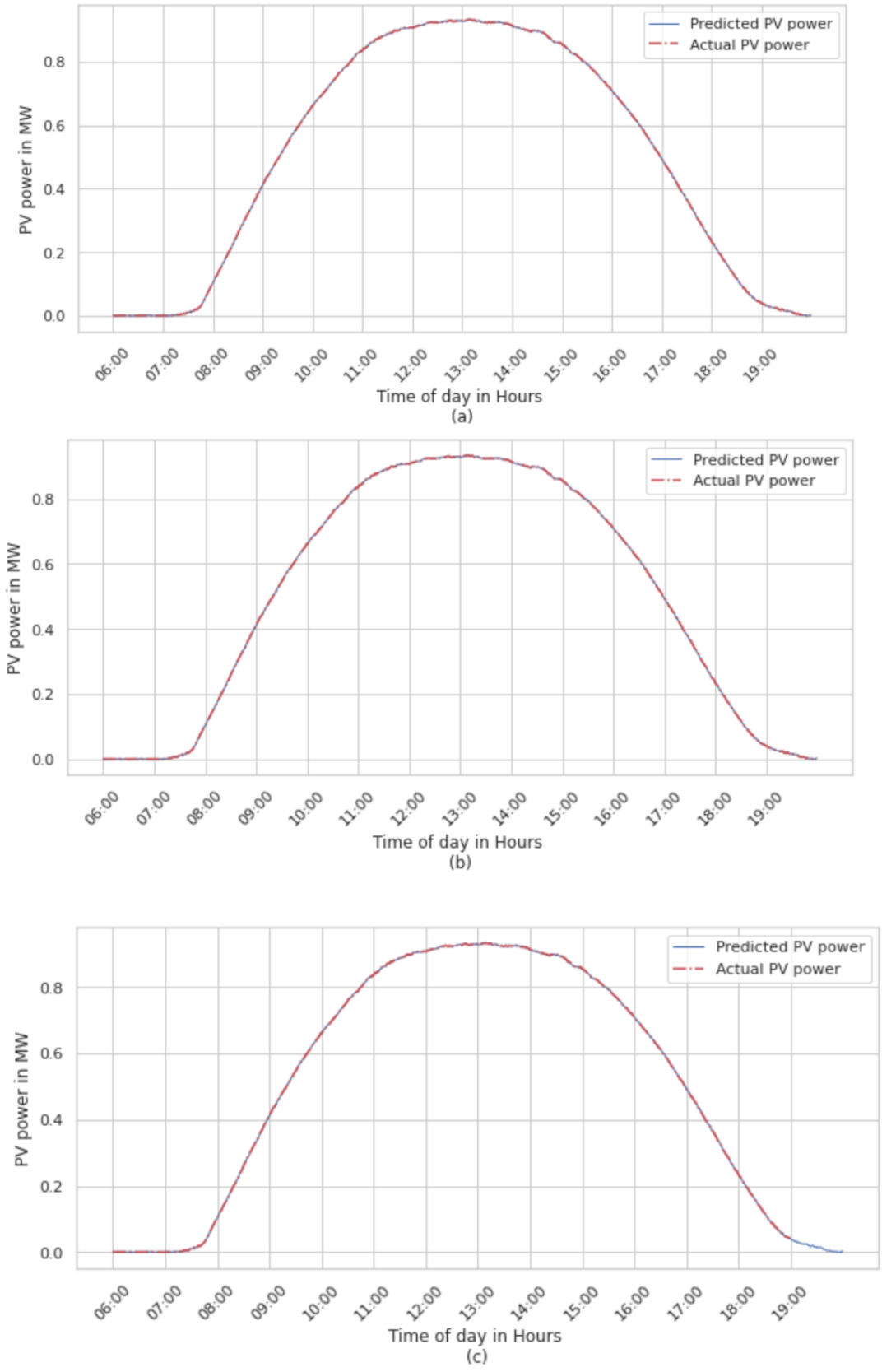
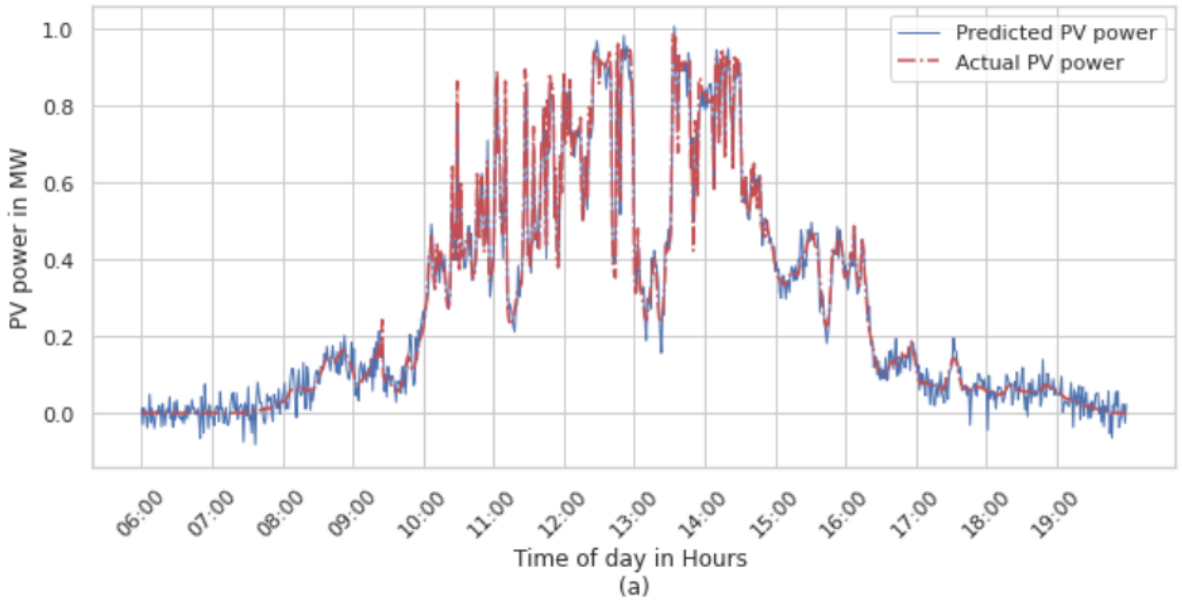
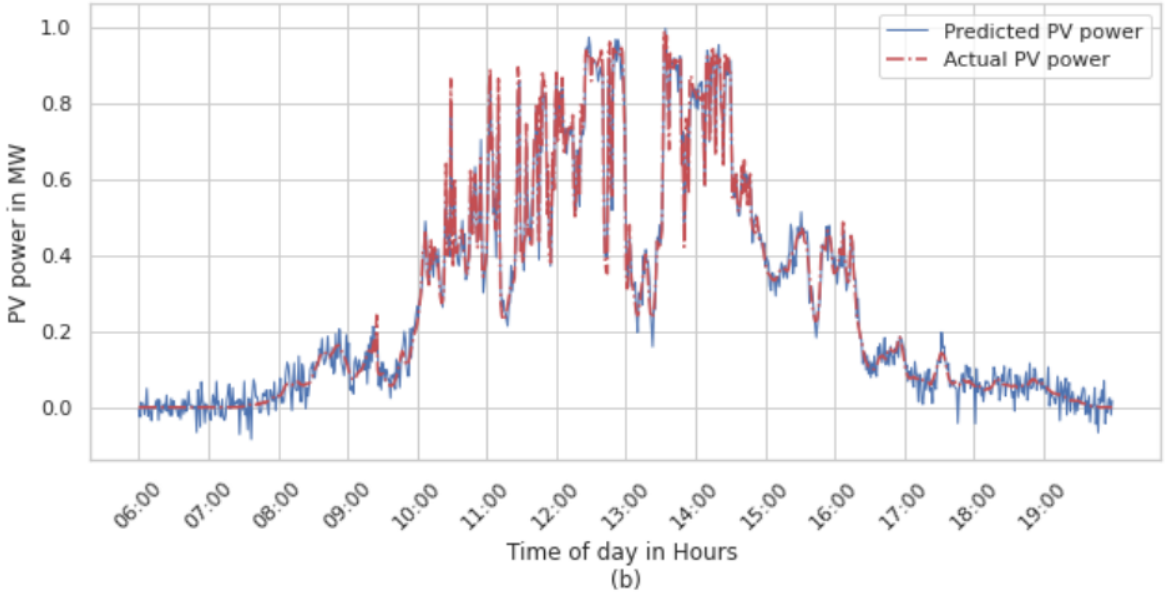


Figure 3. 11 DT prediction for 60 (a), 120 (b) and 180 (c) seconds ahead plotted for training dataset for a sunny day


Predicted PV power
Actual PV power
PV power in MW
1.0
0.8
0.6
0.4
0.2
0.0
06:00
07:00
08:00
09:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00
19:00
Time of day in Hours
(a)


Predicted PV power
Actual PV power
PV power in MW
1.0
0.8
0.6
0.4
0.2
0.0
06:00
07:00
08:00
09:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00
19:00
Time of day in Hours
(b)

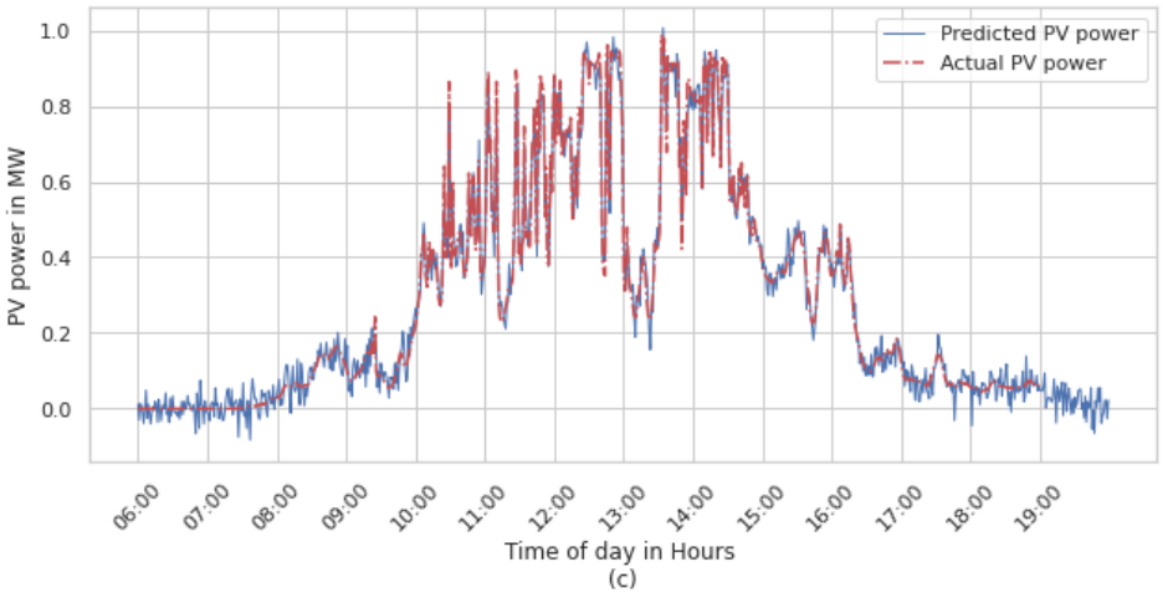


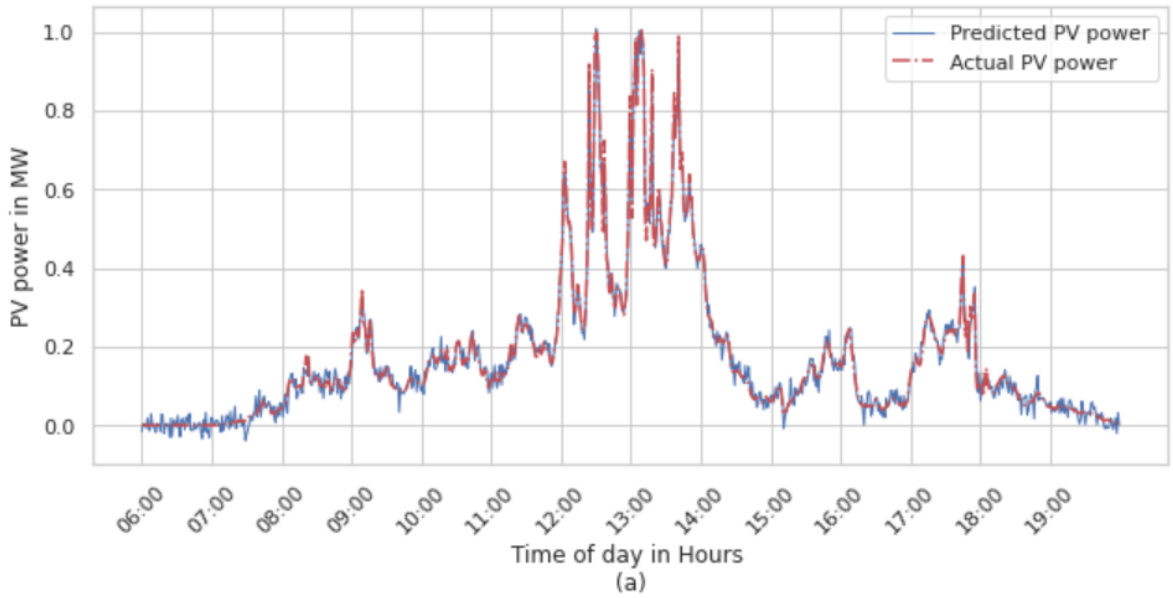
Figure 3. 12 DT estimation for 60 (a), 120 (b) and 180 (c) seconds ahead plotted for training dataset for a moderately cloudy day

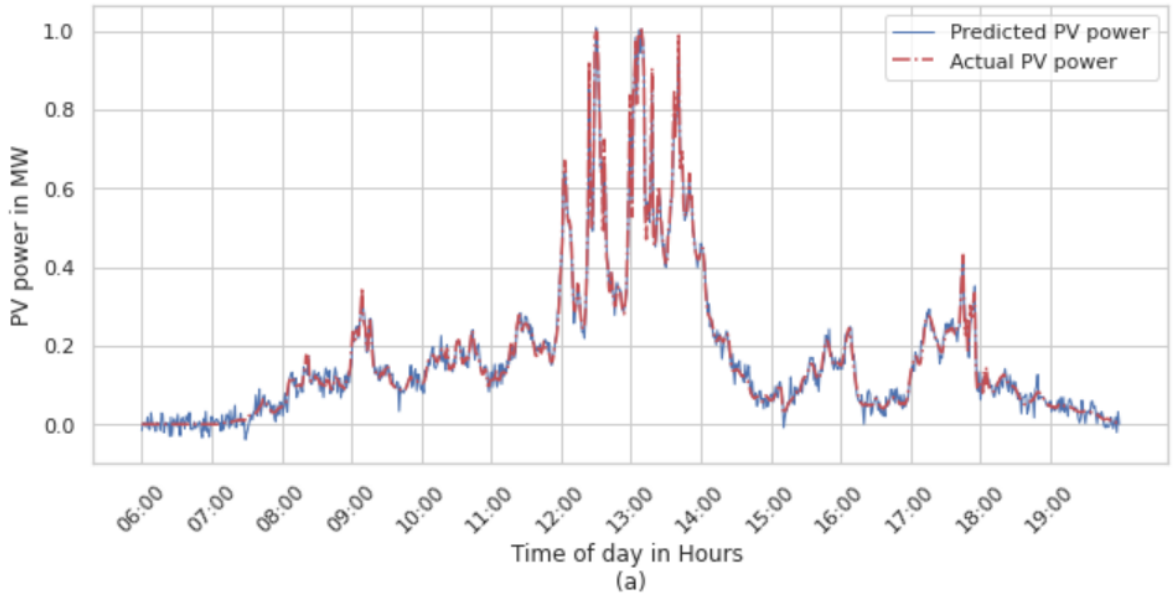
Results of the average ESN performance for training data over 25 trials can be seen in Table 3.4. Bolded results show the best results of prediction interval accuracy. The DT predicts best 60 seconds ahead, with 120 seconds and 180 seconds showing similar results. The highest accuracy is for a sunny day, as expected, due to the comparatively lower complexity of solar irradiance waveform over the day. Moderately cloudy and cloudy days have similar prediction accuracies.

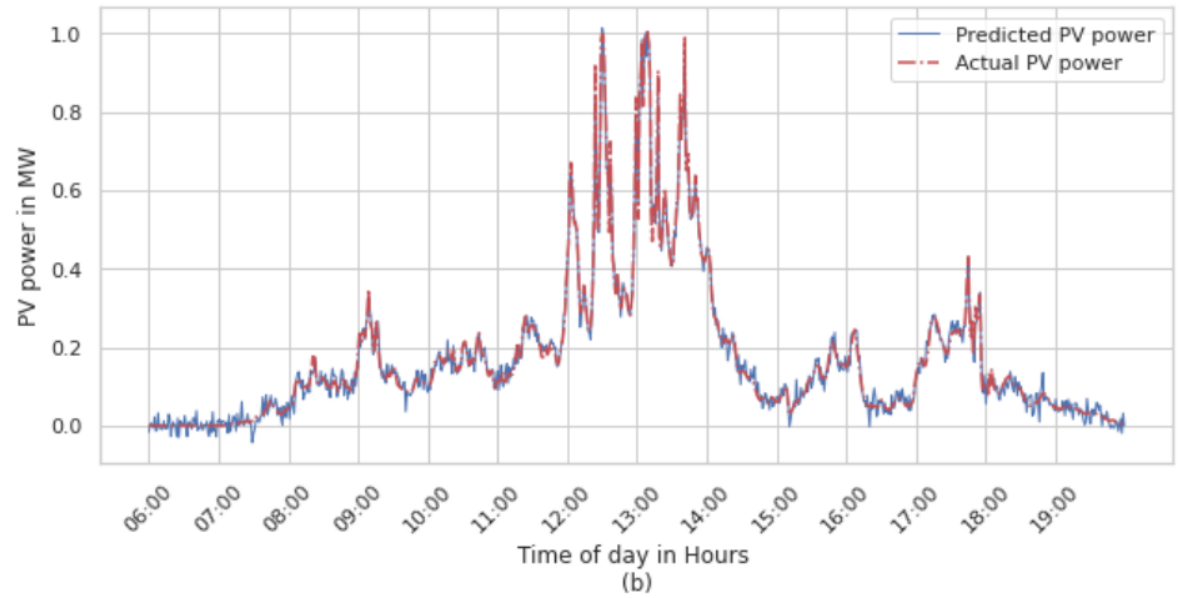
Table 3.4 ESN model performance for Training data

Day Category	*Reservoir neurons (n)*	*Metric*	*60 seconds*	*120 seconds*	*180 seconds*
Sunny	200	MAE	**0.0006**	0.0008	0.0010
	200	MAPE	**0.0250**	0.0336	0.0404
	200	MSE	**0.0007**	0.0009	0.0012
	200	RMSE	**0.0265**	0.0300	0.0346
Moderately Cloudy	550	MAE	**0.0069**	0.0082	0.0100
	550	MAPE	**0.2466**	0.3997	0.3803
	550	MSE	**0.0089**	0.0108	0.0119
	550	RMSE	**0.0943**	0.1039	0.1091
Cloudy	450	MAE	**0.0158**	0.0168	0.0176
	450	MAPE	**0.2342**	0.2482	0.2211
	450	MSE	**0.0183**	0.0197	0.0190
	450	RMSE	**0.1353**	0.1404	0.1378

Results of the ESN model for 60, 120 and 180 seconds ahead estimation for testing dataset for each cloud category is depicted in Figures. 3.13, 3.14 and 3.15.


Predicted PV power
Actual PV power
PV power in MW
1.0
0.8
0.6
0.4
0.2
0.0
06:00
07:00
08:00
09:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00
19:00
Time of day in Hours
(a)


Predicted PV power
Actual PV power
PV power in MW
1.0
0.8
0.6
0.4
0.2
0.0
06:00
07:00
08:00
09:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00
19:00
Time of day in Hours
(a)



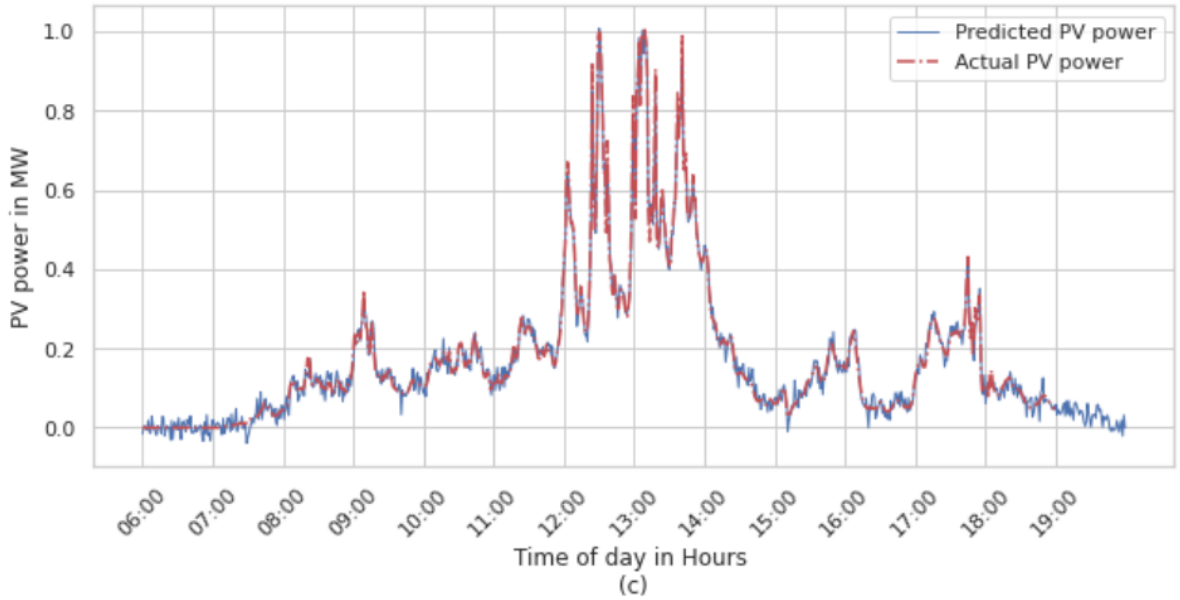
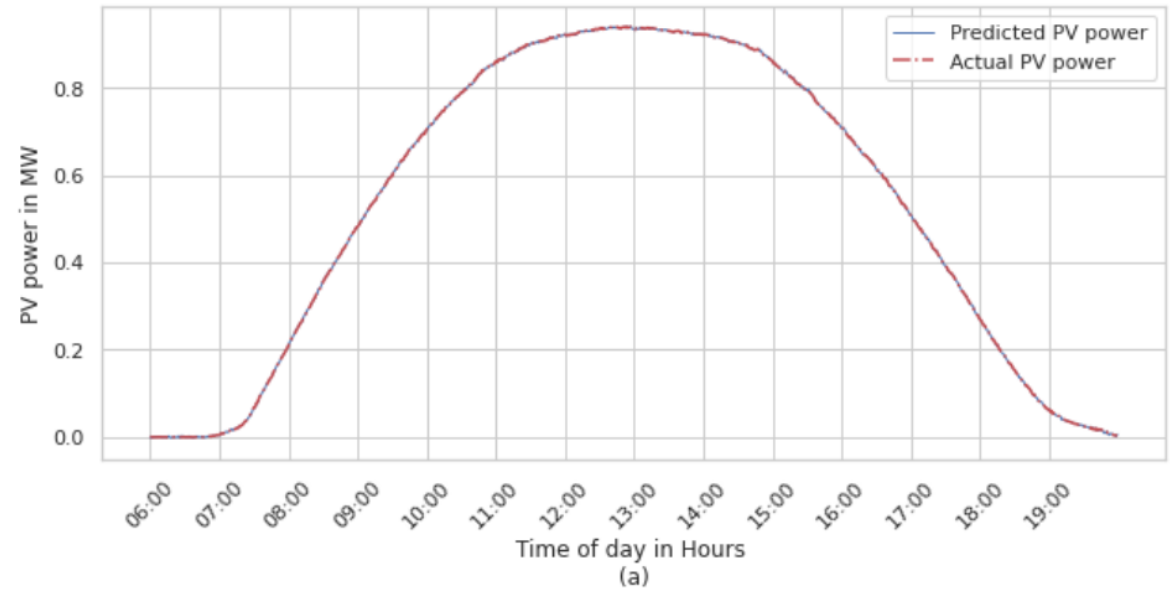
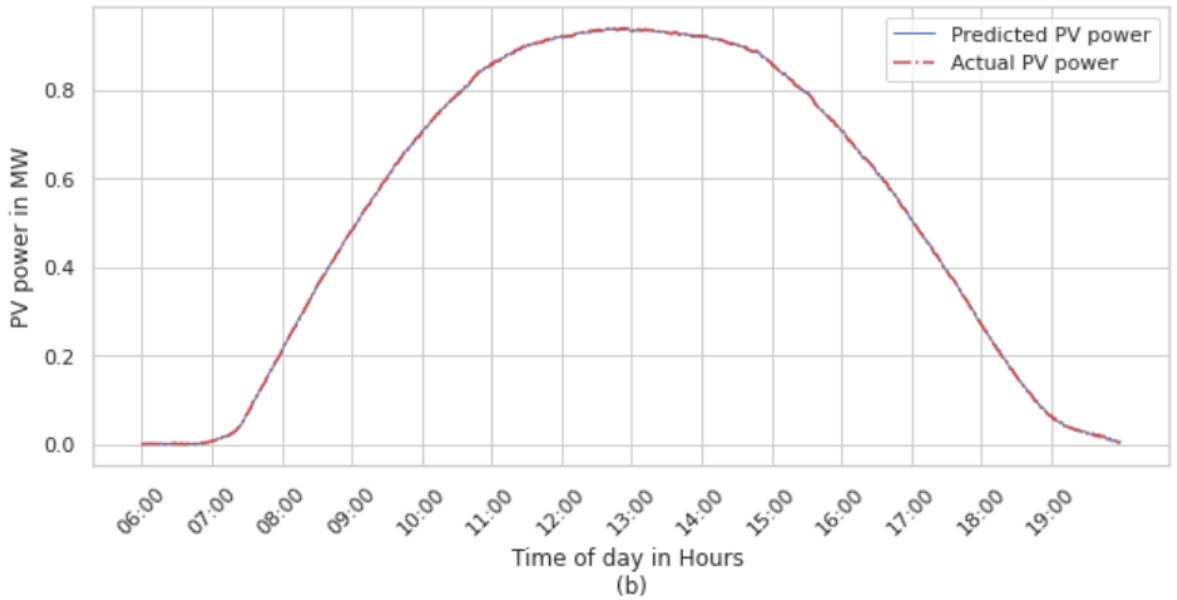


Figure 3.13 DT predicted for 60 (a), 120 (b) and 180 (c) seconds ahead plotted for testing dataset for a cloudy day


Predicted PV power
Actual PV power
PV power in MW
0.0
0.2
0.4
0.6
0.8
06:00
07:00
08:00
09:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00
19:00
Time of day in Hours
(a)


Predicted PV power
Actual PV power
PV power in MW
0.0
0.2
0.4
0.6
0.8
06:00
07:00
08:00
09:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00
19:00
Time of day in Hours
(b)

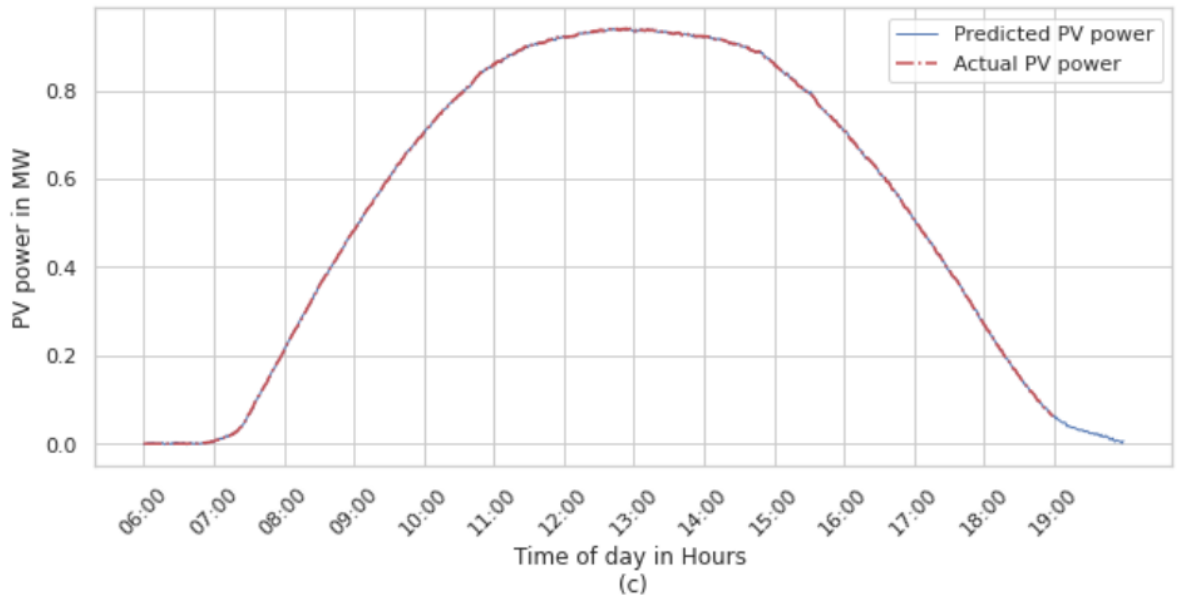
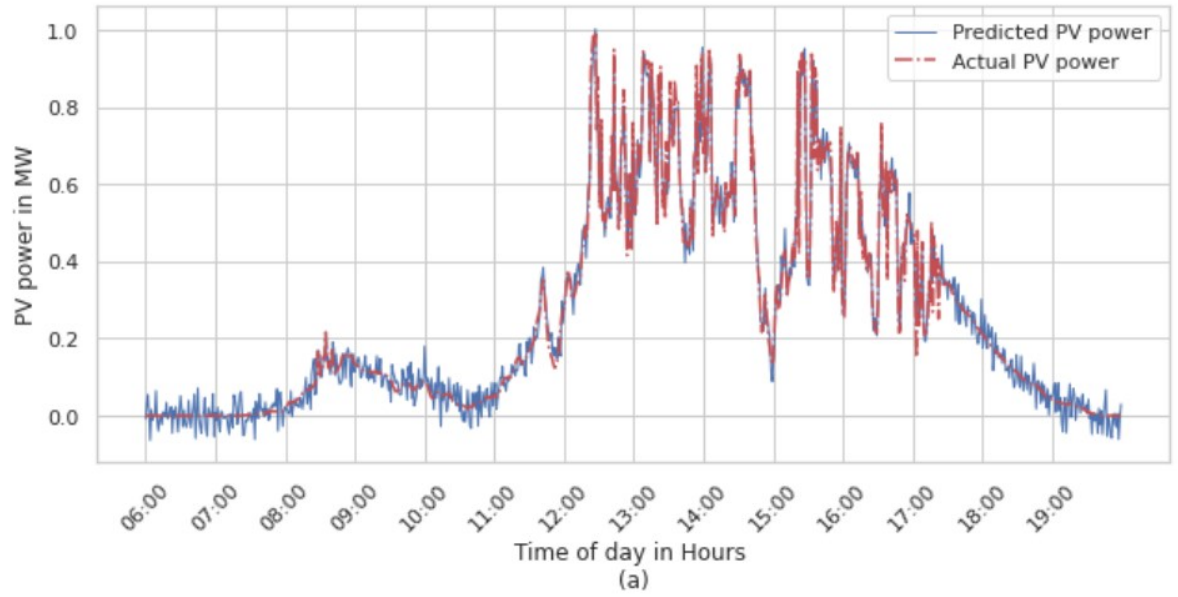
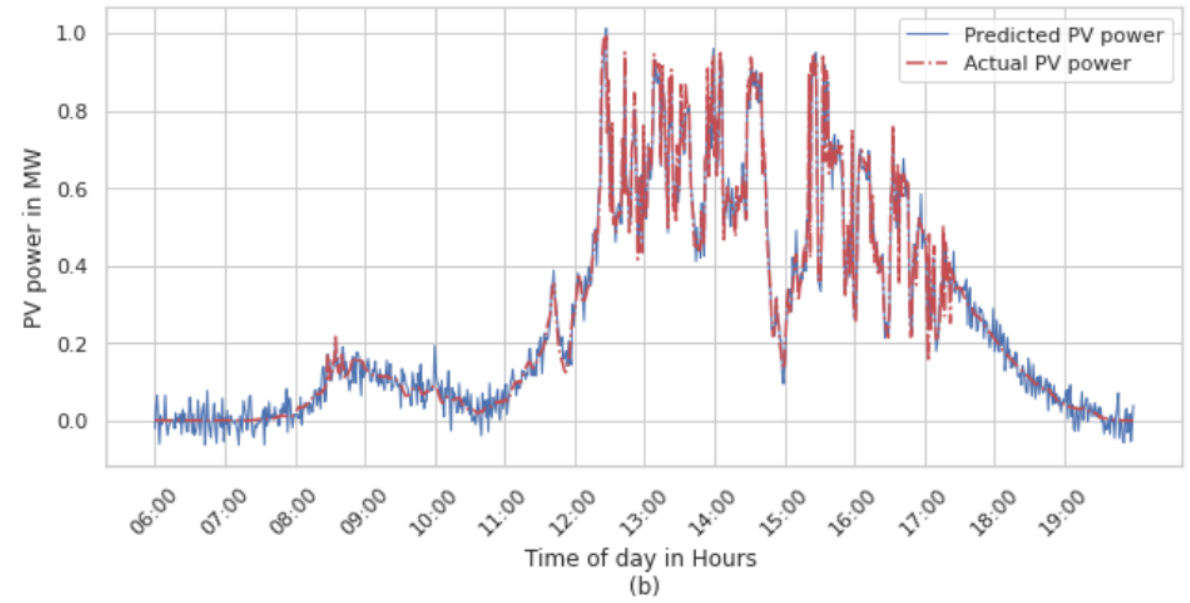


Figure 3. 14 DT estimation for 60 (a), 120 (b) and 180 (c) seconds ahead plotted for testing dataset for a sunny day





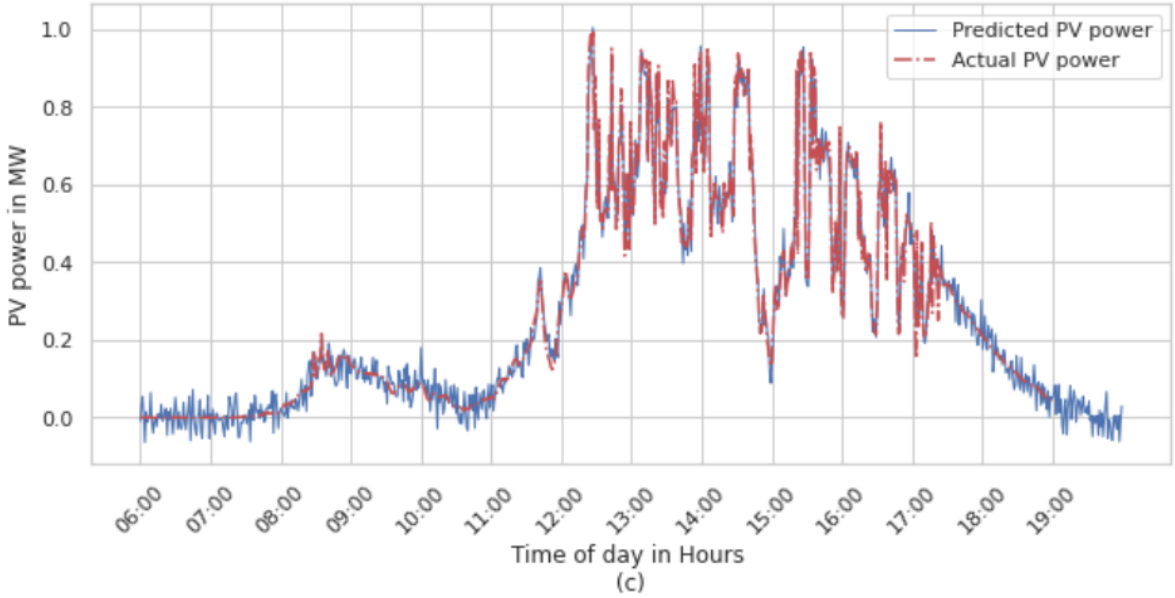


Figure 3. 15 DT estimation for 60 (a), 120 (b) and 180 (c) seconds ahead plotted for testing dataset for a moderately cloudy day

The model's performance is averaged over 25 trials. The model's performance with varying reservoir sizes, compared to the actual PV output for the testing dataset can be seen in Table 3.5, where mean squared error (MSE) (3.2), mean absolute error (MAE) (3.3), and mean absolute percentage error (MAPE) (3.4) of the predicted vs actual values during testing

the model are calculated along with average prediction interval time. Bolded values show the best performance of the models.

Table 3.5 ESN model performance for Testing data

Day Category	*Reservoir neurons (n)*	*Metric*	*60 seconds*	*120 seconds*	*180 seconds*
Sunny	200	MAE	**0.0005**	0.0006	0.0011
	200	MAPE	**0.023**	0.0353	0.0423
	200	MSE	**0.001**	0.0013	0.0014
	200	RMSE	**0.0316**	0.0361	0.0374
Moderately Cloudy	550	MAE	**0.0063**	0.0089	0.012
	550	MAPE	**0.2456**	0.3571	0.3345
	550	MSE	**0.0079**	0.0123	0.0149
	550	RMSE	**0.0889**	0.1109	0.1221
Cloudy	450	MAE	**0.016**	0.0168	0.0194
	450	MAPE	**0.2124**	0.2232	0.2361
	450	MSE	**0.0183**	0.019	0.0204
	450	RMSE	**0.1353**	0.1378	0.1428

Summary

3 data-driven PV DTs of a 1 MW plant were developed using MLPs and ESNs for each day type and tested for real-time estimation and prediction accuracy over multiple cloud-coverage patterns for the months of February to April. The DTs operate with reasonable accuracy while being computationally efficient, and are able to predict or estimate data alongside the physical system. PV DT streams of data are additionally archived and categorized for future optimization and analysis of the system. This methodology can be easily adapted to support new PV installations, rooftop or utility-scale, with varying levels of granularity, complexity, different tilts or new surrounding environmental factors as per availability. Different AI algorithms or frameworks can be utilized if improved accuracy is possible.

CHAPTER FOUR
VIRTUAL WEATHER STATIONS

Introduction

Modeling regional variation in weather over a localized area is an important aspect of DER resource planning and distribution system studies. Optimizing and analyzing the spatial variances in DERs, and in this study, distributed PV (DPV plants), integrated into a distribution network necessitates the availability of high quality, high resolution, spatiotemporal weather data. This data is to further be harnessed for real time or time-series distribution system studies, where minutes or even seconds resolution is needed. Volt-var optimization, power flow studies and DPV planning operations such as voltage profile analysis, hosting capacity studies among others often require quick and easily available estimations of realistic DPV power production over a region, in different configurations, time-scales or placement regions.

Thereby, to provide real-time, multi-time scale weather station measurement inputs to a PV DT and enable quick, computationally efficient estimations of PV power prediction over an area, a means to develop virtual weather stations (VWS) at any location, supported by locally available measurement sites through data reanalysis and intelligent mutations is designed.

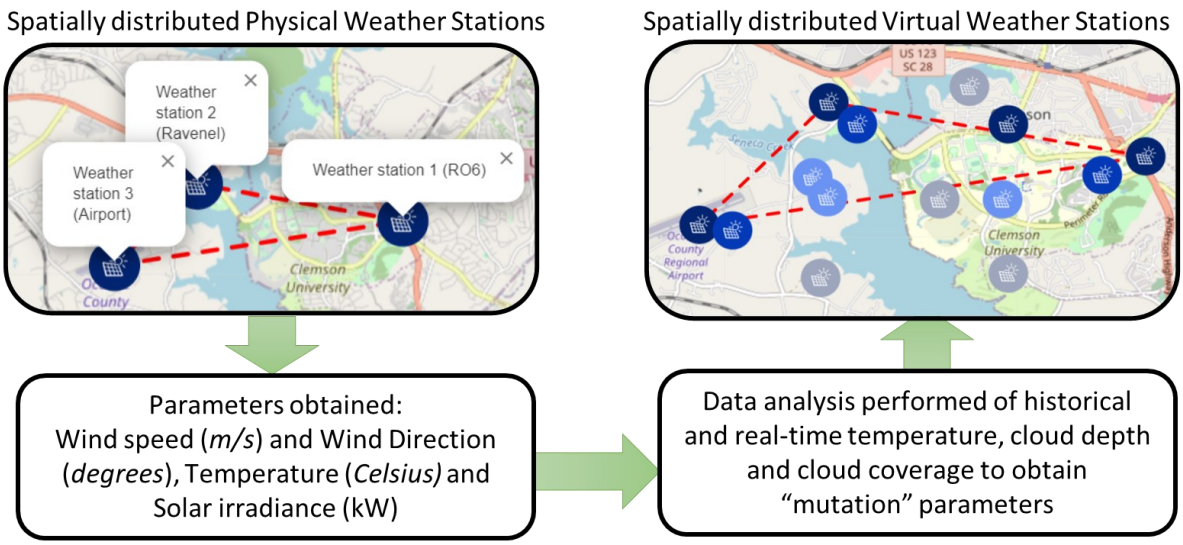


Figure 4. 1 Schematic block diagram for creation of virtual weather stations.

Spatially Distributed Physical Weather Stations (PWS)

While mathematical estimations and interpolations of solar irradiance profiles over geographical locations can be calculated using various solar irradiance models, an intelligent mutation-based approach using reanalysis data is proposed. Reanalysis typically uses numerical models and observational data to generate a numerical weather prediction model that creates a representation of the system. This methodology is less dependent on satellite-based weather forecasting models, can be optimized and restructured for new locations and can be adapted for developing new DERs or load profiles at local regions with similar levels of uncertainty but highly complex process dynamics.

Parameters such as latitude, elevation, surface coverage and meteorological conditions such as cloud coverage, cloud transmissivity, aerosols, relative humidity, and column water vapor typically affect solar irradiance across a large geographical region. They are therefore environmental or climate factors that affect solar irradiance falling on PV panels.

For a smaller, localized area in the range of 20-25 km radius however, fewer variables cause significant variation in solar irradiance as affects potential new PV plants in the area.

The behavior of three important factors affecting solar irradiance over the selected region: cloud cover depth, cloud cover variance and temperature with respect to spatial distribution is identified as main variational factors in the selected region for PV power generation [66]. These parameters are analyzed and leveraged to generate real-time synthetic weather station data streams at new locations. Three existing weather stations that measure wind speed and direction, solar irradiance and temperature are then used as physical systems to develop VWSs at new, user selected geographical locations with intelligently mutated solar irradiance profiles, generating new weather station data in real-time through the development of weather station DTs.

The original dataset utilized to form base inferences is the weather the three stations recorded over a period of 72 days in the spring-summer months. The weather stations use three Campbell Scientific CR300 dataloggers, cables and network links to measure weather data, along with measurement equipment such as pyranometers, anemometers, thermometers and so on. A TCP/IP connection is established with the datalogger servers to query data at required time intervals. The hardware of each weather station can be seen in Figure 4.2.

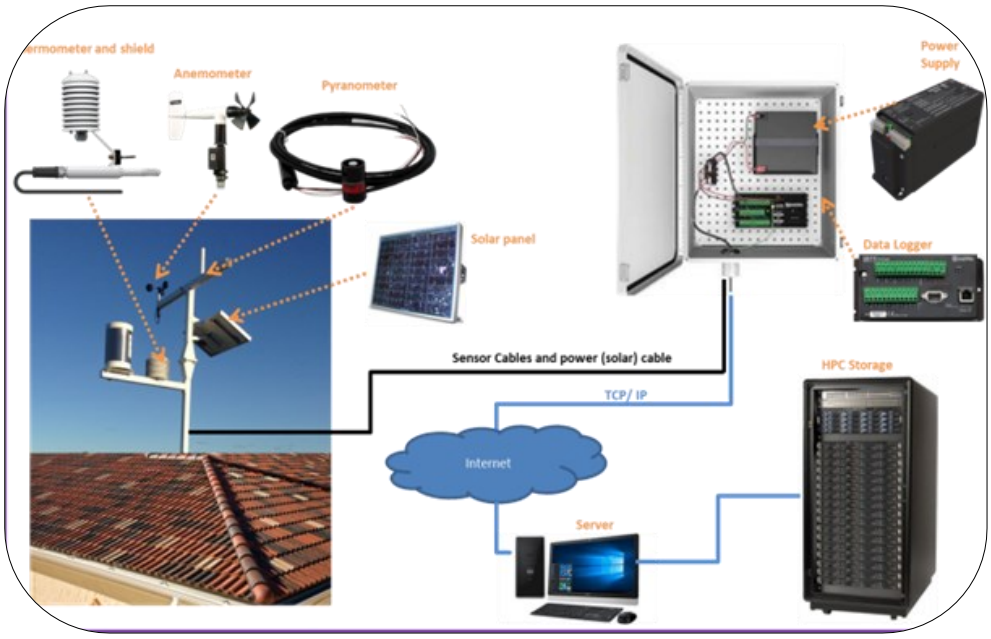


Figure 4.2 Physical Weather stations

Mutation of real-time solar irradiance data instances collected at the weather stations at a geographical location is done through utilizing weighted combination of geographical and meteorological variables such as cloud coverage, cloud depth and temperature. These parameters are utilized to modify the parent datasets, to generate child populations of real-time irradiance data distributed over an estimated geographical location. The child datasets are visualized on a map at the user given locations, creating virtual real-time weather stations as necessary [66]. A single parent irradiance data set can essentially be used to generate multiple realistic offspring data sets and be further used to estimate corresponding power generation over a geographical distribution of virtual PV plant sites. The physical locations of the weather stations and their distance from each other are shown in Figure 4.3, called “Ravenel”, “R06” and “Airport”. The 1 MW PV plant is at site R06, as explained in chapter 3. A circle surrounding the three sites shows the area of potential VWS location placements.

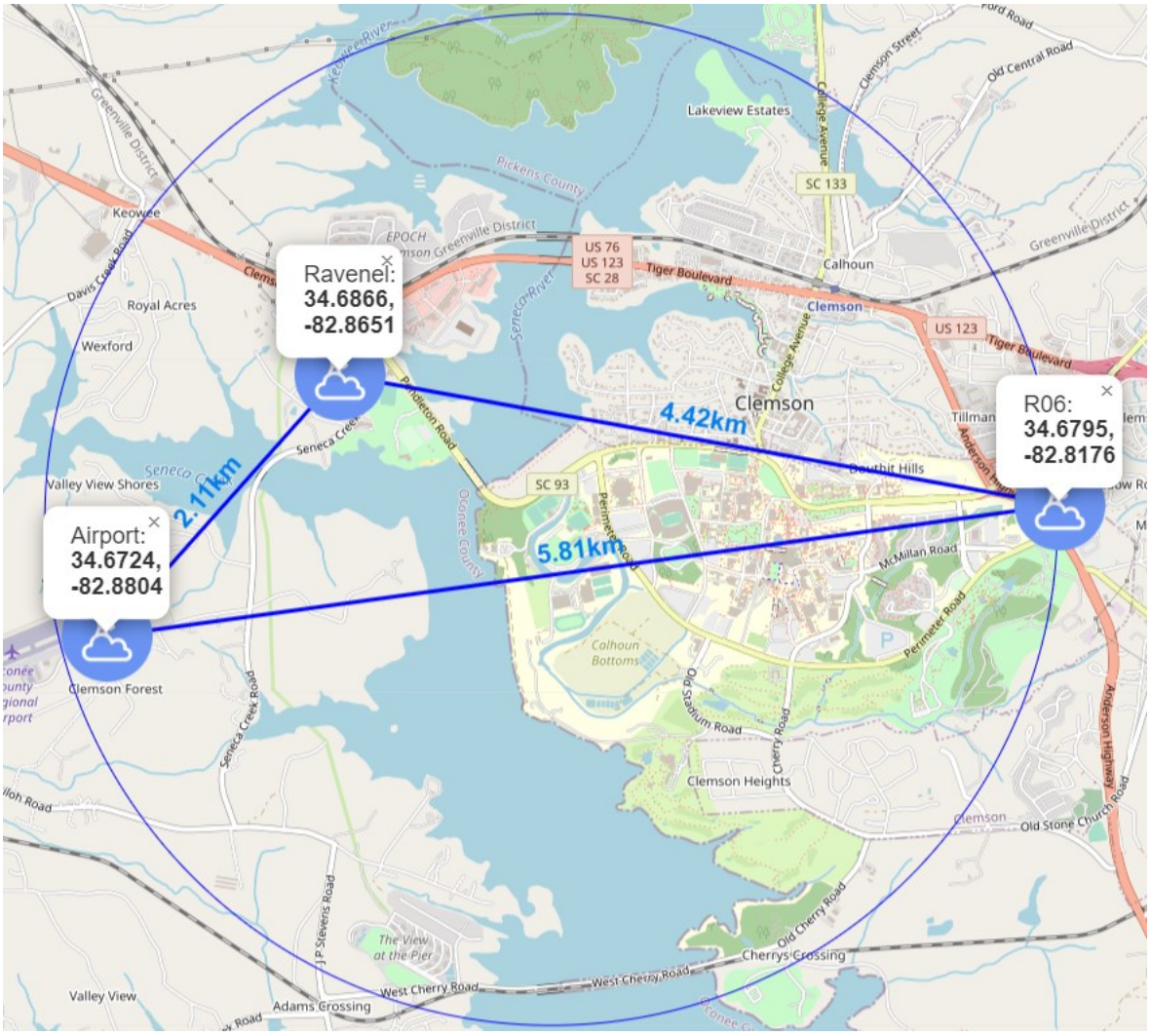


Figure 4. 3 PWSs and distances from each other (Label as ‘Airport’)

Data Analysis

The real-time behavior of the weather stations with respect to each other has been observed and modeled for a particular day using three-month averages of historical data to capture long term trends and daily data to analyze short term trends between the stations.

The resultant graphs are as seen in Figures 4.4 to 4.12. R06, on average, measures slightly more solar irradiance, while Ravenel and the Airport stations experience nearly similar amounts. However, daily measurements show the inherent uncertainty in cloud coverage across the three sites for second or minute intervals along with the variance in overall cloud coverage patterns across the three sites.

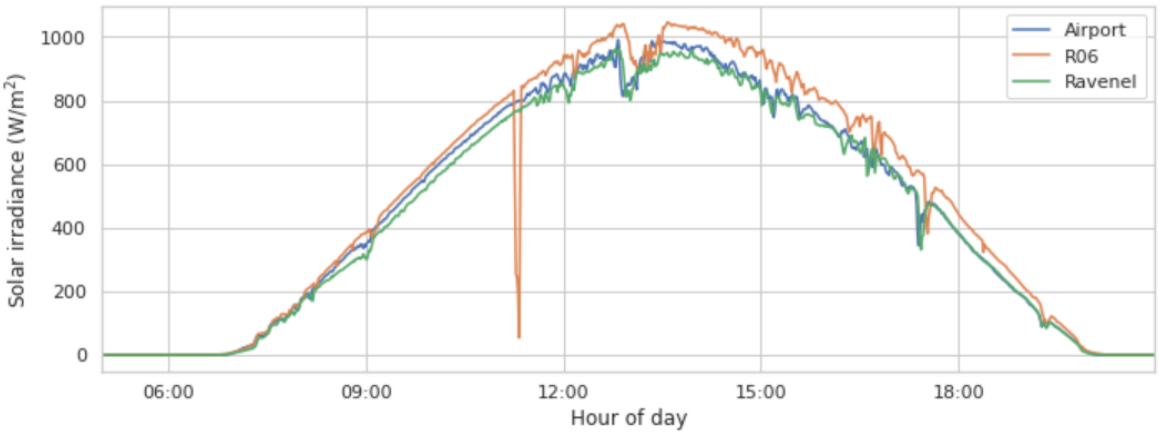


Figure 4. 4 Daily solar irradiance trends for a sunny day

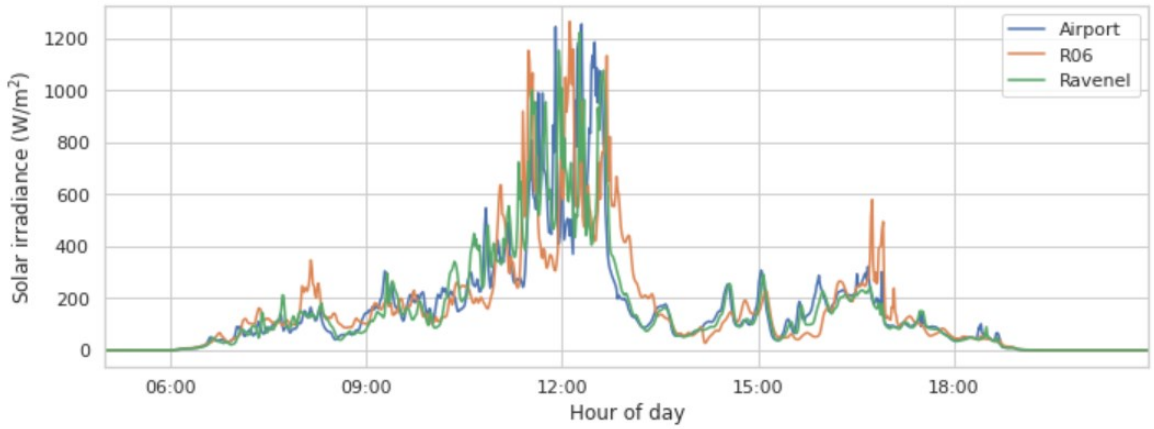


Figure 4. 5 Daily solar irradiance trends for a cloudy day

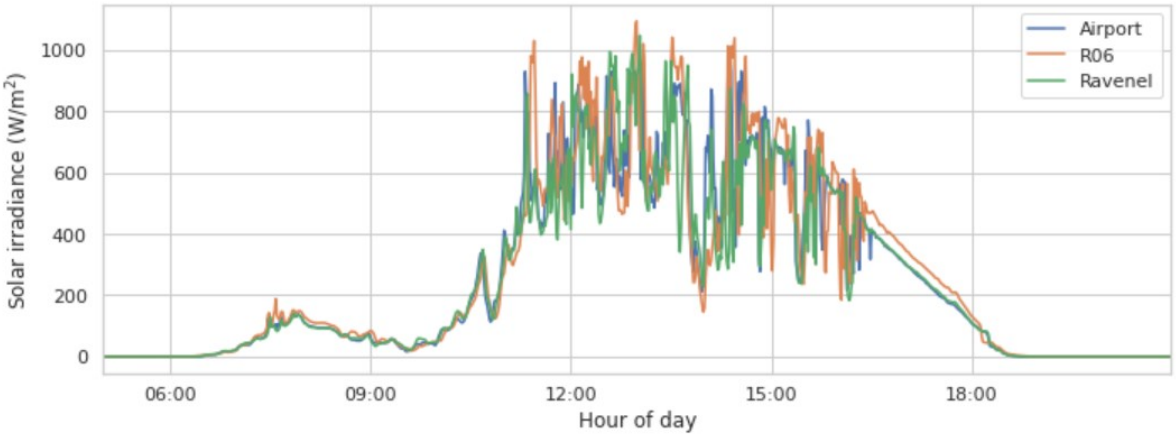


Figure 4. 6 Daily solar irradiance trends for a moderately cloudy day

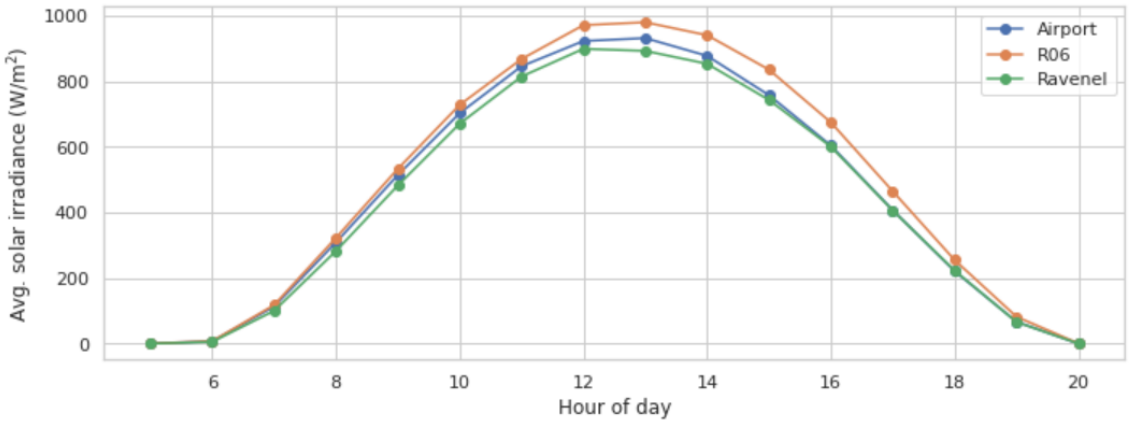


Figure 4. 7 Historical average solar irradiance for sunny days

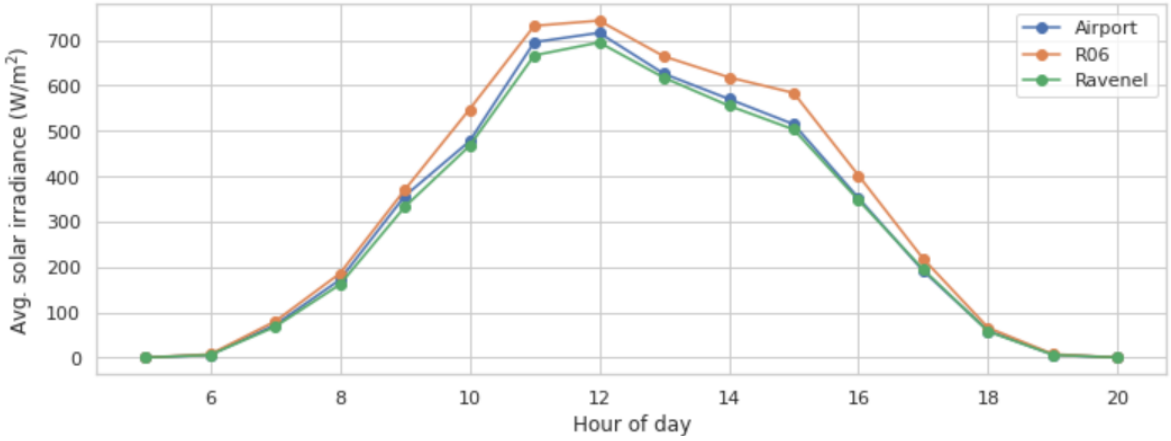


Figure 4. 8 Historical average solar irradiance for moderately cloudy days

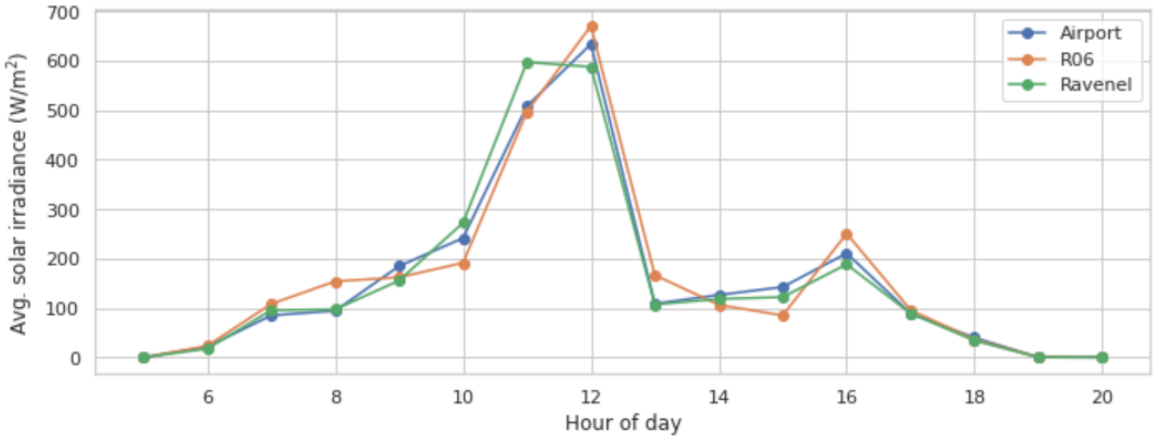


Figure 4. 9 Historical average solar irradiance for cloudy days

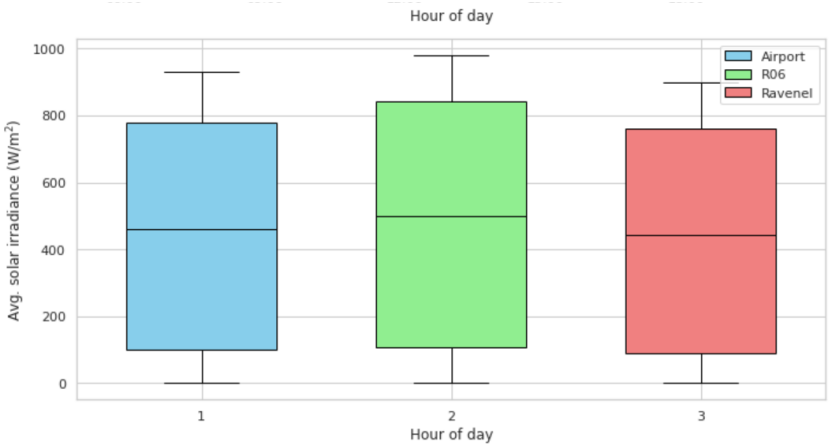


Figure 4. 10 Comparative analysis of variance between stations for sunny days

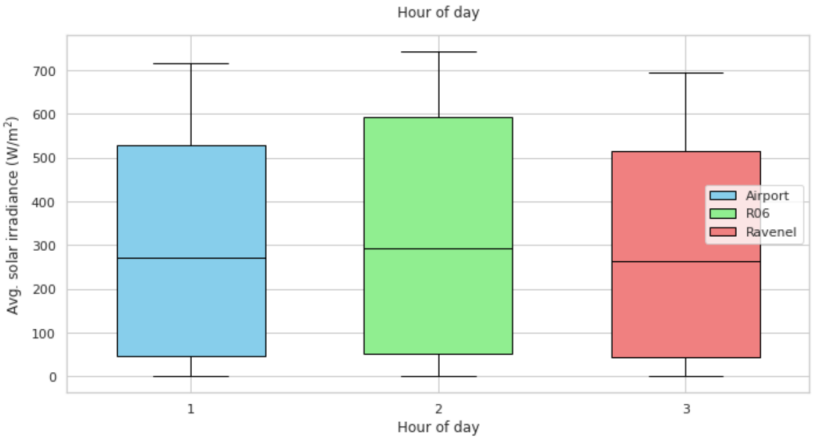


Figure 4. 11 Comparative analysis of variance between stations for moderately cloudy days

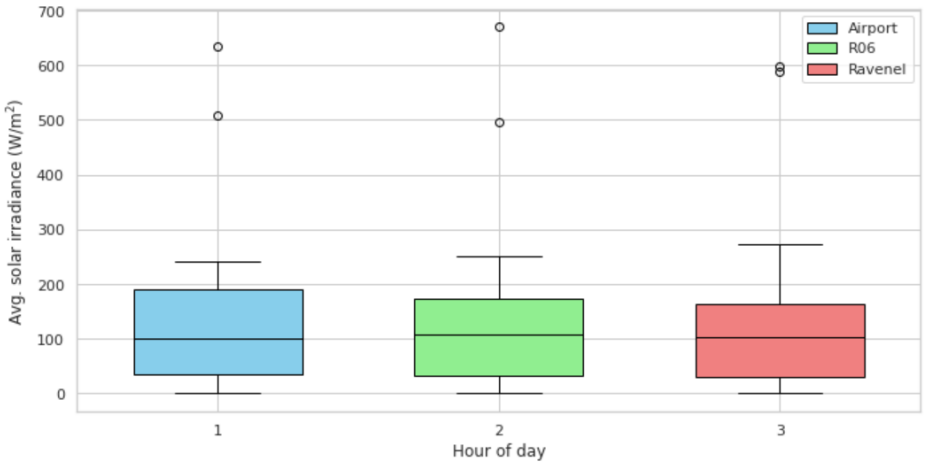


Figure 4. 12 Comparative analysis of variance between stations for cloudy days

The variation of windspeed is additionally unpredictable. However, R06 seems to show, on average, higher wind speeds compared to the other three sites. The temperature variances at the site also lie between 1-3 degrees Celsius variations.

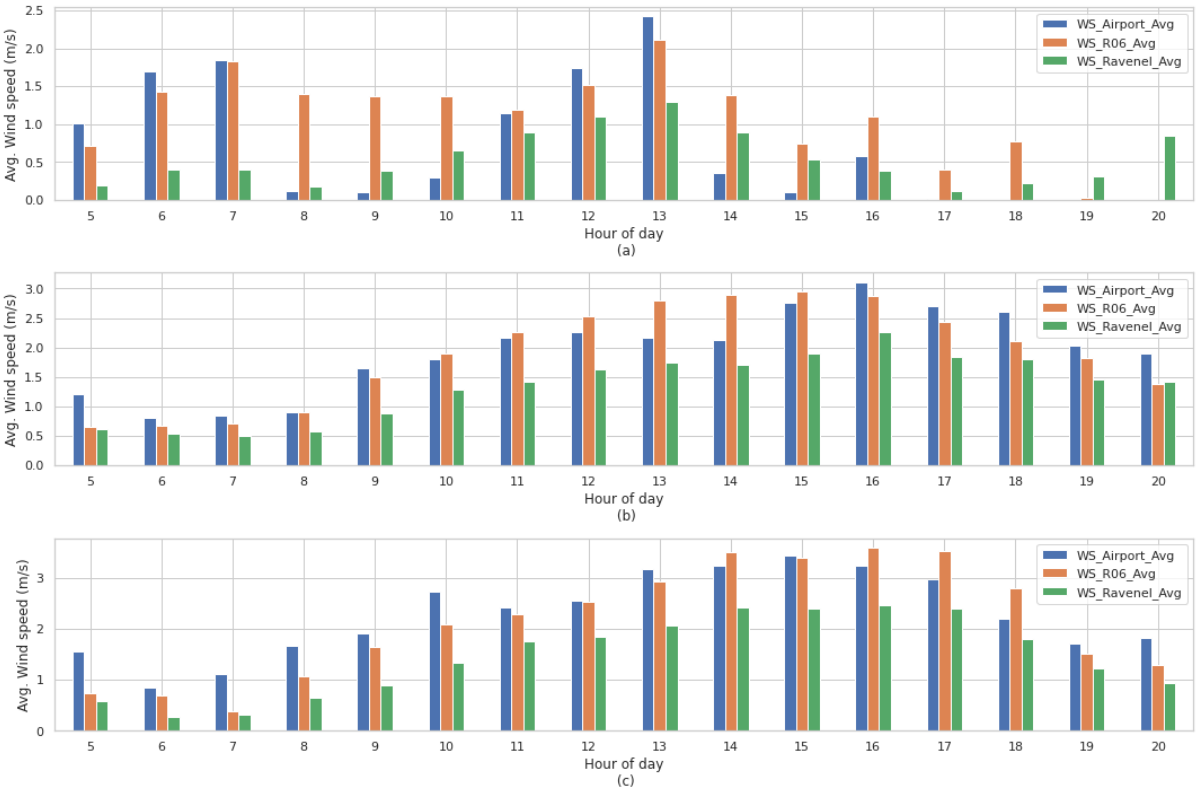


Figure 4. 13 Wind speed variation between the stations Ravenel (a), Airport(b) and RO6 (c)

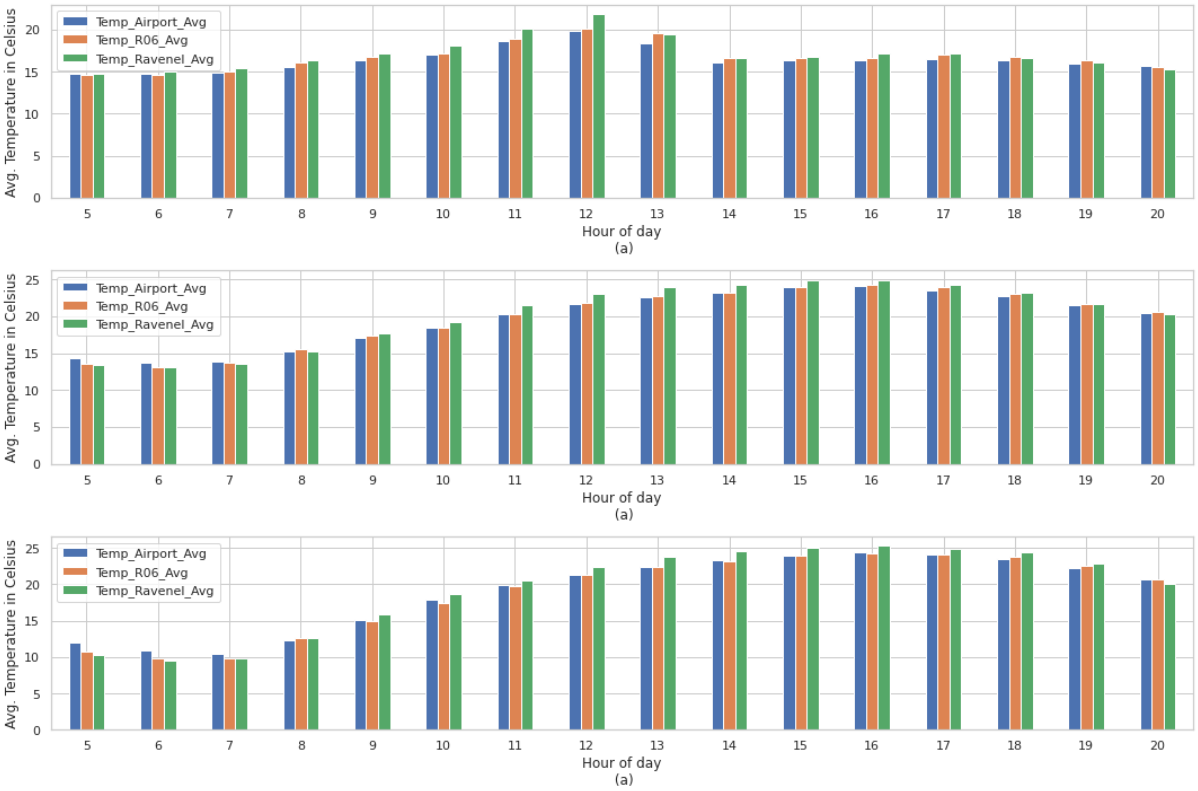


Figure 4. 14 Temperature variation between the stations Ravenel (a), Airport(b) and RO6 (c)

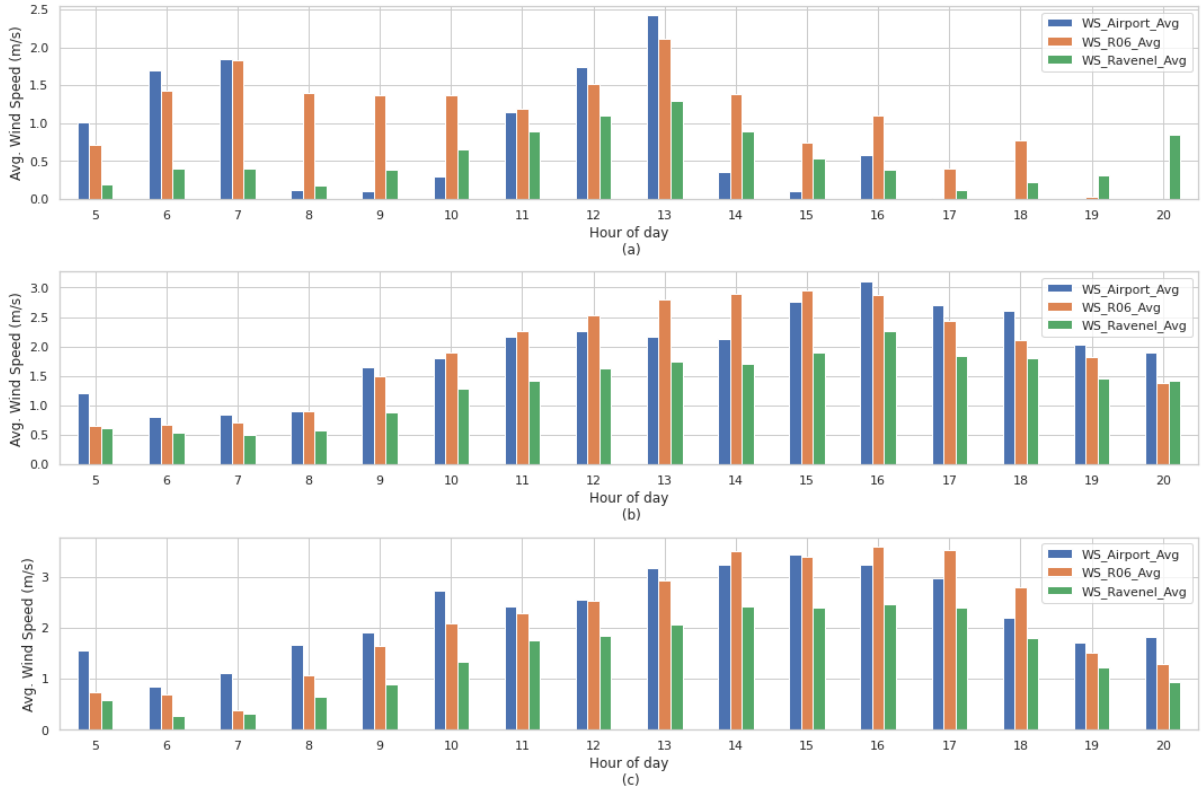


Figure 4. 15 Average Historical Wind speed variation between the stations Ravenel (a), Airport(b) and RO6 (c)

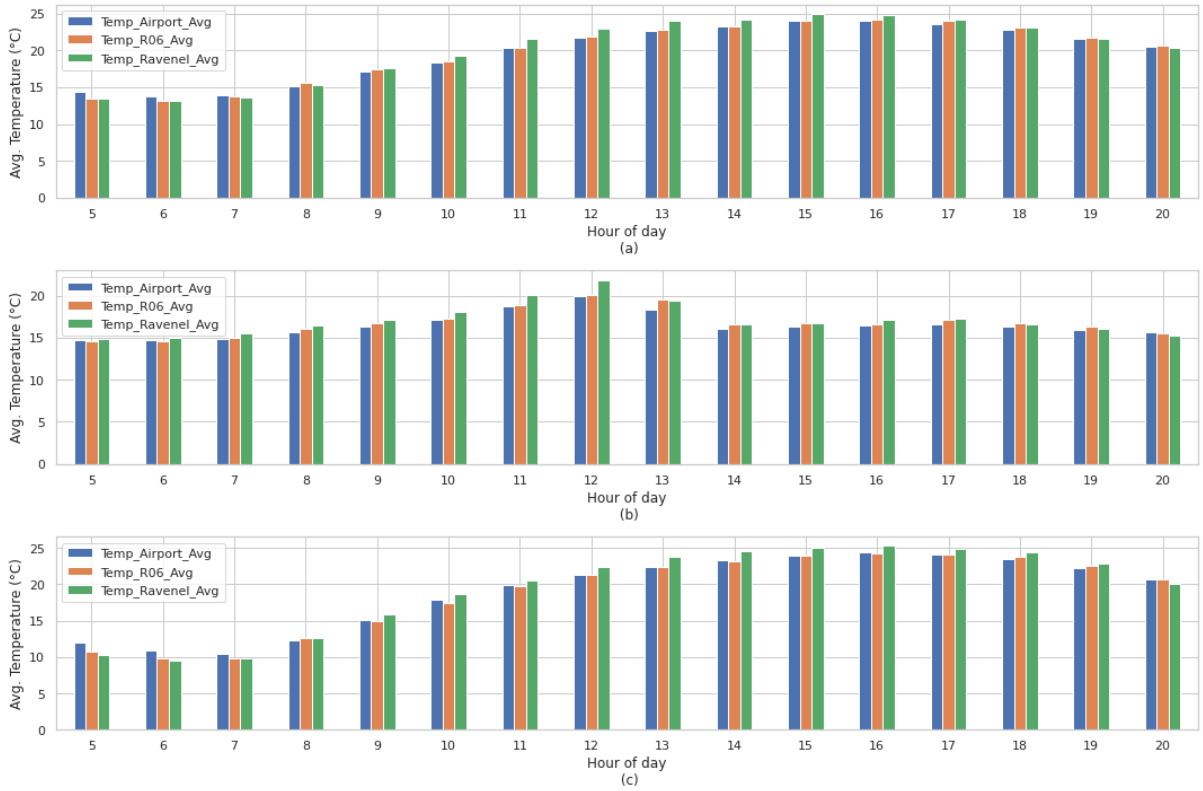


Figure 4. 16 Average Historical Temperature variance between the stations Ravenel (a), Airport(b) and RO6 (c)

Mutation procedure

An arbitrary number of user-selected regions from this area, here 10, are utilized as the locations of the VWSs. Latitudes and longitudes of their locations are needed. The overall procedure involved for mutation is explained in Figure 4. 17

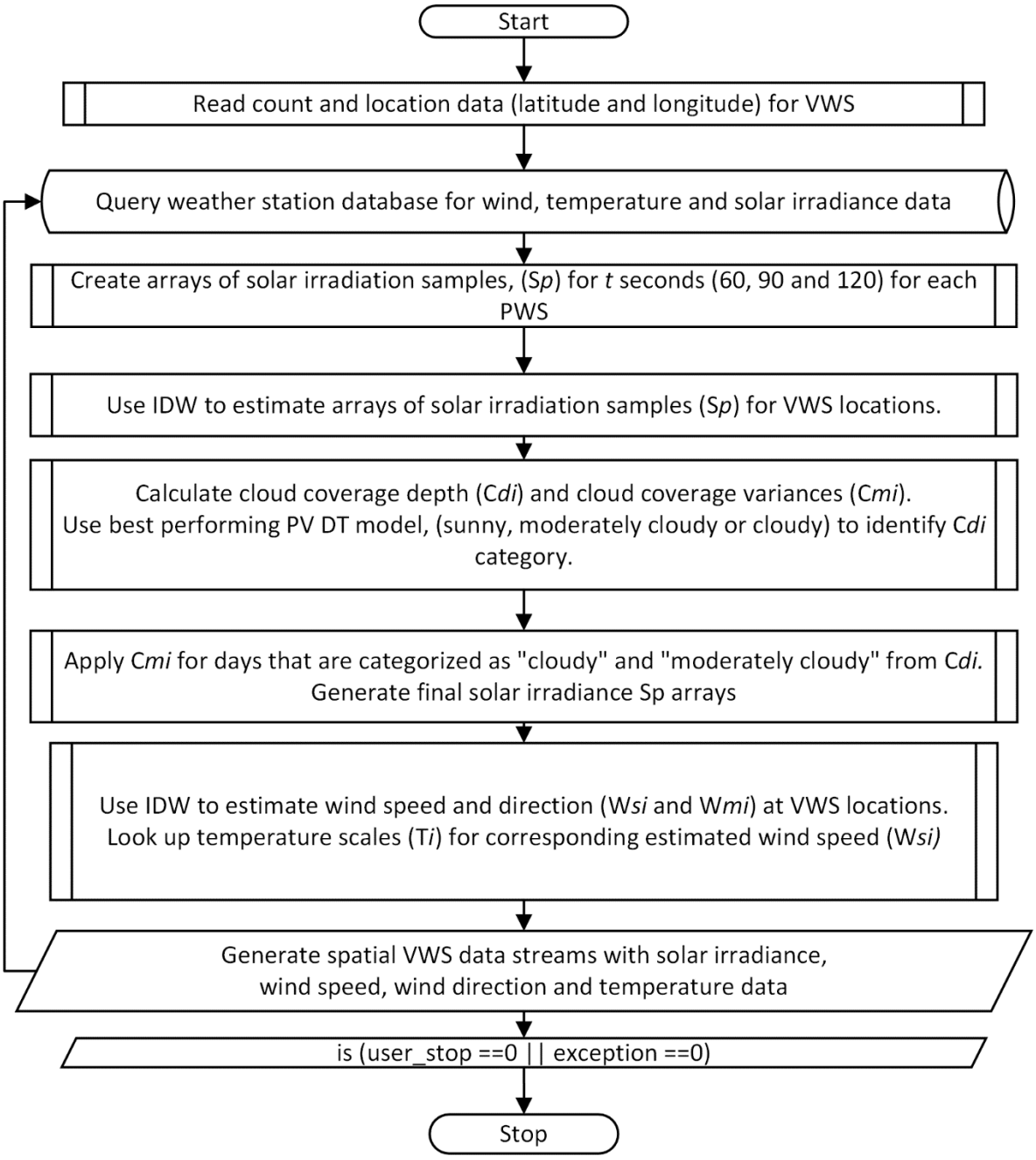


Figure 4. 17 Flowchart for VWSs' data generation

The parameters utilized for and important to VWS data generation are as shown in Table 4.1.

Table 4.1 Parameters utilized for VWS data generation

Sl No.	*Parameters*	*Notation*
1	Solar irradiance arrays of 3 stations consisting of *t* seconds data, here 60,90 and 120 seconds	S_p (4.1)

2	Cloud coverage depth, categorized as “sunny”, “cloudy” and “moderately cloudy”	C_d
3	Cloud coverage variance, with temporospatial variances dependent on location from original site and cloud coverage depth	C_m (4.4)
4	IDW obtained Wind speed in meters/second and wind direction in degrees	W_{msi} (4.6), W_{mdi} (4.7)
5	Temporospatial temperature scale from wind-speed vs temperature look-up table	T_i
6	IDW estimations of solar irradiance data at VWS locations at *t* seconds, here 60,90 and 120 seconds	S_{pred}
7	Mutated solar irradiance data arrays of VWS at *t* seconds, here 60,90 and 120 seconds	S_m (4.5)

Solar irradiance S_p consists of arrays of solar irradiance measurements, S_i in kW/m^2, from each station over a period of time *t*, for example, 60,120 or 180 seconds as shown in (4.1). Inverse Distance Weighting (IDW) is used to estimate base solar irradiance values at VWS based on the approach used in [67]. The formula for IDW can be expressed as in equation where: *V(u)* is the estimated value at the unknown location, V_i is the value at the known location *i* and w_i is the weight assigned to the known location *i*, calculated based on the distance between the known and unknown locations as a squared or cubic distance. Once IDW is applied, 10 S_{pred} arrays with *t* lengths will be identified for each VWS. Here, 3 locations are known and p is set to 2.

$$S_p = [S_1, S_2, \dots S_i] \quad (4.1)$$

$$V(u) = \frac{\sum (Vi * Wi)}{\sum Vi} \quad (4.2)$$

$$Vp = 1 / (x1 - x2)^p \quad (4.3)$$

S_{pred} is the base solar irradiance estimation at each VWS to which the impact of cloud coverage and depth is to be added. To account for cloud impact, historical data as well as parameter inferences obtained from the National Solar Radiation Database (NSRDB) [66] is referenced in the mutation procedure. Two mutation parameters are identified for cloud impact on solar irradiance, the depth of cloud coverage (C_d) and the movement of cloud coverage (C_m).

C_d is a scale ranging from between 0 to 1, where the solar irradiance at time t is divided by the maximum historical solar irradiance for a sunny day at that time step, updated to match real-time conditions at the PWS as the day progresses. For a sunny day identified from physical readings at the weather station, C_d will be closer to 0, indicating lower cloud depth and therefore little to no impact caused by movement of cloud coverage C_m on all virtual weather station output streams. The C_d over a seconds' interval is shown in Table 4.2 and can be further categorized based on best performing PV DT model as explained in chapter 2 into "cloudy", "moderately cloudy" and "sunny" days.

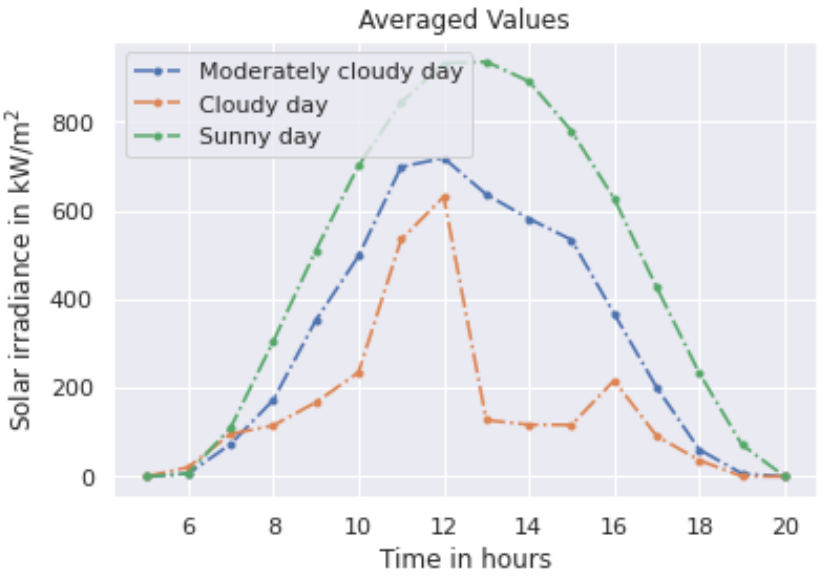


Figure 4. 18 The average of solar irradiance categories over 72 days

Table 4.2 Hourly cloud coverage percentage variations for each day category

Hour of the day (24 hour clock)	***Sunny day (%)***	***Moderately Cloudy day (%)***	***Cloudy day (%)***
7	0.00	33.74	12.86
8	1.00	43.35	62.24
9	0.20	30.87	67.18
10	0.00	29.07	66.50
11	0.00	17.21	36.63
12	0.00	22.89	32.30
13	0.00	32.01	86.38
14	0.30	34.74	86.88
15	0.05	31.48	85.05
16	0.00	41.45	65.45

17	0.30	52.90	78.60
18	0.00	73.97	84.25

Based on the random number generator model referenced in [66], a scaled random distribution based on estimated distance from the plant is used to mimic C_m (4.4) over geographical locations within <20 kms of each other. C_m is obtained from combinations of controlled random sampling of t second IDW solar irradiance estimations, S_{pred} and the addition of up to 5% Gaussian noise as seen in (4.5)

$$C_m = rand(S_{pred}) + \frac{0.05}{\sigma\sqrt{2\pi}} e^{-(S_{pred}-\mu)^2/2\sigma^2} \quad (4.4)$$

Solar irradiance $S_m$at a given n interval is then mutated with the effect of C_m as shown in (4.4).

$$S_m = \begin{cases} [S_{pred}] & S_{pred} > 0 \; and \; Cd \leq 10\% \\ [C_m] & S_{pred} > 0 \; and \; Cd > 10\% \\ 0 & S_{pred} = 0 \end{cases} \quad (4.5)$$

Wind speed and direction is a separate parameter that is seen to be very weakly correlated to cloud cover and thereby does not directly impact solar irradiance patterns affected by cloud coverage [68]. A weak correlation is seen, where for both rural and urban areas, highly windy days are slightly more likely to be clear [69]. However, wind speed affects the temperature at the weather station. IDW is used to calculate W_{msi} and W_{mdi}, or overall wind speed and wind direction at various geographical locations for a given solar

irradiance. The weight is equal to the count of actual weather stations with available real-time measurements, here, 3. The distance calculated is the cubic distance and not the least squared distance.

$$W_{msi} = \sqrt{(\frac{1}{ws}\sum_i^n W_{si}\cos(W_{di}))^2 + (\frac{1}{ws}\sum_i^n W_{si}\sin(W_{di}))^2} \quad (4.6)$$

$$W_{mdi} = \arctan(\frac{\sum_i^n W_{si}\cos(W_{di})}{\sum_i^n W_{si}\sin(W_{di})}) \quad (4.7)$$

A cooling effect factor, W_e, is obtained to vary temperature data obtained at each VWS based on estimated wind speed effects. W_{diff} is the wind speed difference at the VWS station from the average wind speeds at the three PWS at instant *t,* divided by the average wind speeds at the PWS. W_{diff} is used to look up values in the Table 4.3, where each range in W_{diff} corresponds to a cooling effect factor W_e. If W_{diff} is positive at time *t*, the W_e is added to temperature T_i, and if negative, it is subtracted. In this manner real time temperature data is estimated at the VWS locations. A hysteresis-like loop is observed for wind speed variation over temperatures throughout the day, as seen in Figure 4.19. Additionally, a maximum and minimum range of temperature variation of + or - 3 degrees for this site, as observed from historical data analysis of temperature trends, seen in Figure 4.7. From this, based on average values seen in Table 4.3, a look up table is derived, shown in Table 4.4, for average wind speed variation ranges corresponding to drops in VWS temperature

$$W_{diff} = \frac{avg(W_{msi(PWS)}) - W_{msi(VWS)}}{avg(W_{msi(PWS)})} \quad (4.8)$$

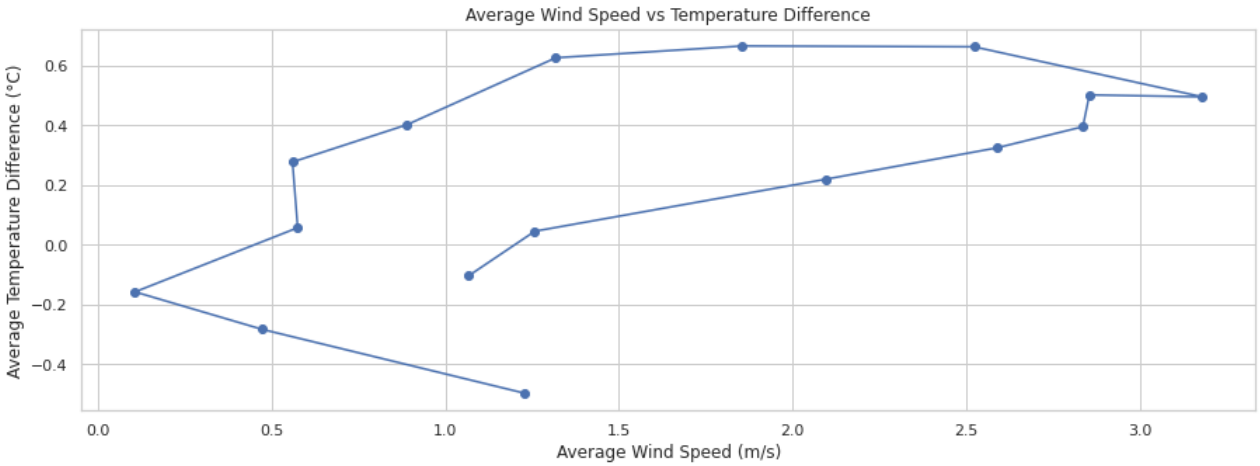


Figure 4. 19 VWS Temperature measurements over a day

Table 4.3 Average temperature difference vs wind speed difference for three stations over ranges

Sl No.	*Wind Speed Range*	*Average Temperature difference between stations (degrees celsius)*	*Average wind Speed difference between stations (m/s)*
1	0.1 - 0.2	0.10729	-0.4988
2	0.2 - 0.3	0.47286	-0.2846
3	0.4 - 0.5	0.56056	-0.1592
4	0.6 - 0.7	0.57554	-0.1053
5	0.8 - 0.9	0.88953	0.04419
6	1.1 - 1.2	1.06719	0.05597
7	1.2 - 1.3	1.22857	0.21896
8	1.2 - 1.3	1.25787	0.27621
9	1.3 - 1.4	1.31845	0.32393
10	1.8 - 1.9	1.85432	0.3945
11	2.0 - 2.1	2.09749	0.40099
12	2.5 - 2.6	2.52157	0.49461
13	2.5 - 2.6	2.5886	0.50103
14	2.8 - 2.9	2.8348	0.62554
15	2.8 - 2.9	2.85343	0.66242
16	3.1 - 3.2	3.17831	0.66504

Table 4.4 Look up table

Sl No.	*Wdiff scale (+/-)*	*Temperature Variation (degrees Celsius) T_i (+/-)*

0	(0.1, 0.4]	-0.3323
1	(0.4, 0.7]	-0.20232
2	(0.7, 1.0]	-0.1258
3	(1.0, 1.3]	0.062736
4	(1.3, 1.8]	0.332292
6	(1.8, 2.1]	0.396208
7	(2.1, 2.4]	0.440125
8	(2.4, 2.6]	0.486667
9	(2.6, 2.9]	0.544938
10	(2.9, 3.2]	0.594583

Results and Discussion

The ten selected locations for VWS are as seen in Table 4.5. The locations plotted onto a map are seen in Figure 4.20.

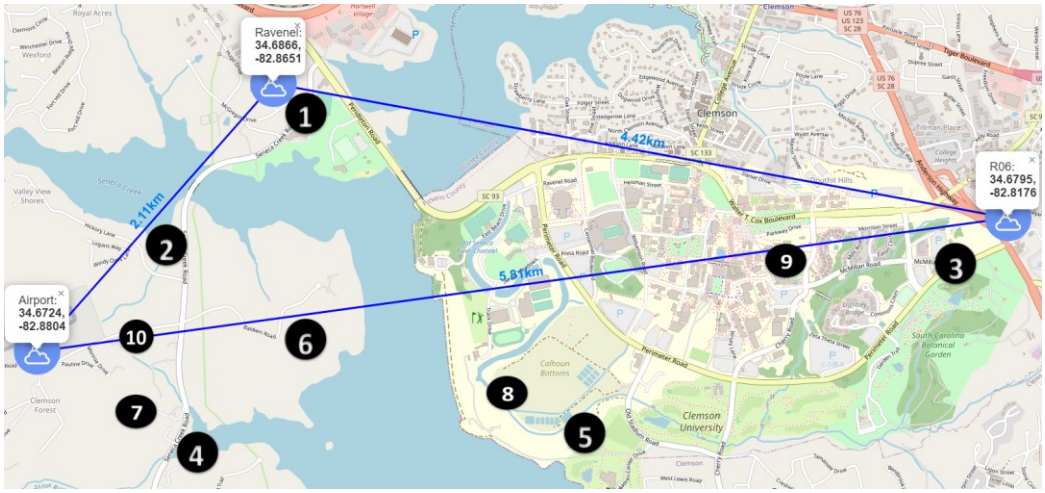


Figure 4.20 Location of PWSs and VWS generated in the region

Table 4.5 Selected locations for VWSs

Station no.	***Latitude (°N)***	***Longitude (°W)***
1	34.685	-82.863
2	34.672	-82.886
3	34.678	-82.818
4	34.667	-82.87
5	34.668	-82.845
6	34.673	-82.863
6	34.669	-82.874
7	34.67	-82.85
8	34.677	-82.832
9	34.6724	-82.8804
10	34.673	-82.874

Validation of proposed approach

The mutation approach explained in the previous section is used to generate weather station DTs at the PWS site locations, and compare estimated vs actual solar irradiance and temperature results. This is to further validate the inferences used in the previous section.

To test the results of the weather station DTs, the locations of the sites are set to the original site locations from Table 4.5. The results of estimated temperature variations can be seen in Figure 4.21. Here, the effects of IDW for solar irradiance estimation are not applied as the distance between the PWS and VWS is 0. The effect of estimated C_m is therefore seen

in the Figures 4.21-4.22. The comparison results of estimated vs actual solar irradiance at a site are seen in Table 4.6.

Some limitations of the validation procedure apply as below:

- IDW cannot be utilized when the distance between the PWS and locations for VWS is 0, i.e., when the same location is set for a PWS and VWS. Therefore, wind speed (W_{msi}) IDW weighting and solar irradiance base inference (S_{pred}) IDW weighting has not been applied.
- Mutation factor (C_m) for sunny day is set to an insignificantly low value. Solar irradiance comparisons when VWS is set to a PWS location are therefore not meaningful for a sunny day
- Temperature (T_i) lookups are performed based on a W_{diff} of the next nearest PWS, which means a larger variance is seen than typical.

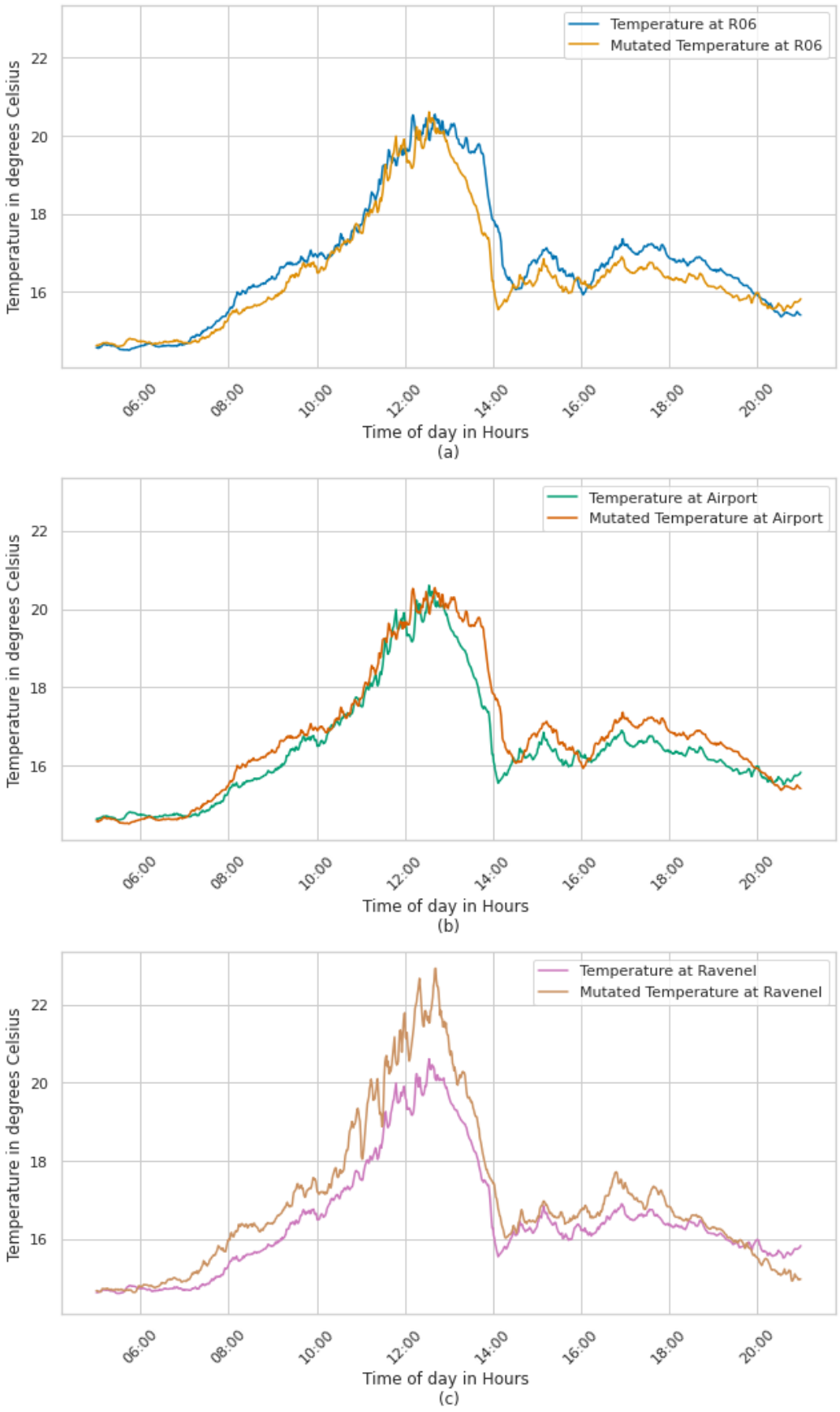


Figure 4.21 PWS vs estimated VWS Temperature measurements over a day for Ravenel (a), Airport (b) and R06 (c)

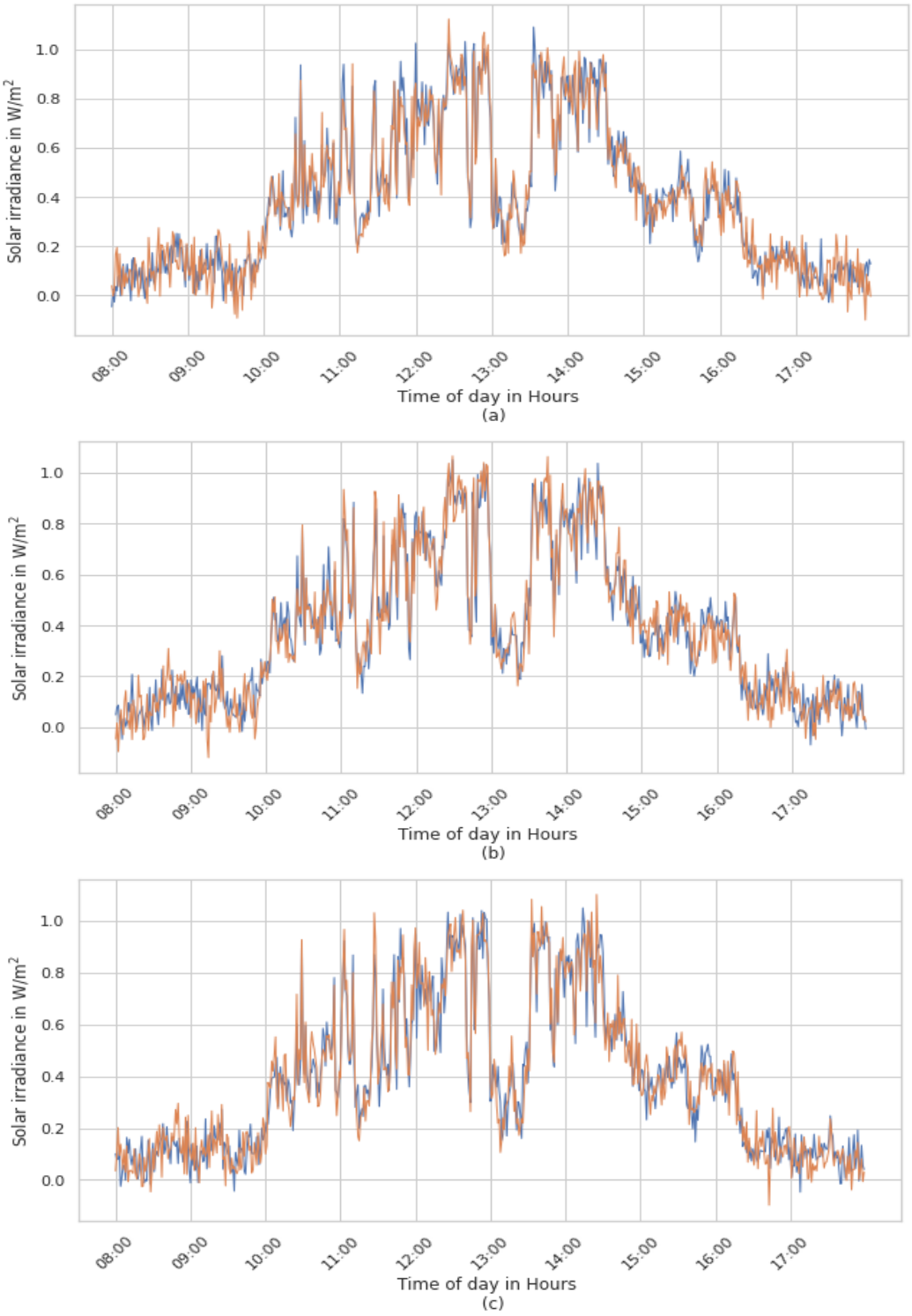


Figure 4. 22 PWS vs estimated VWS solar irradiance measurements over a cloudy day for R06 (a), Ravenel (b) and Airport (c)

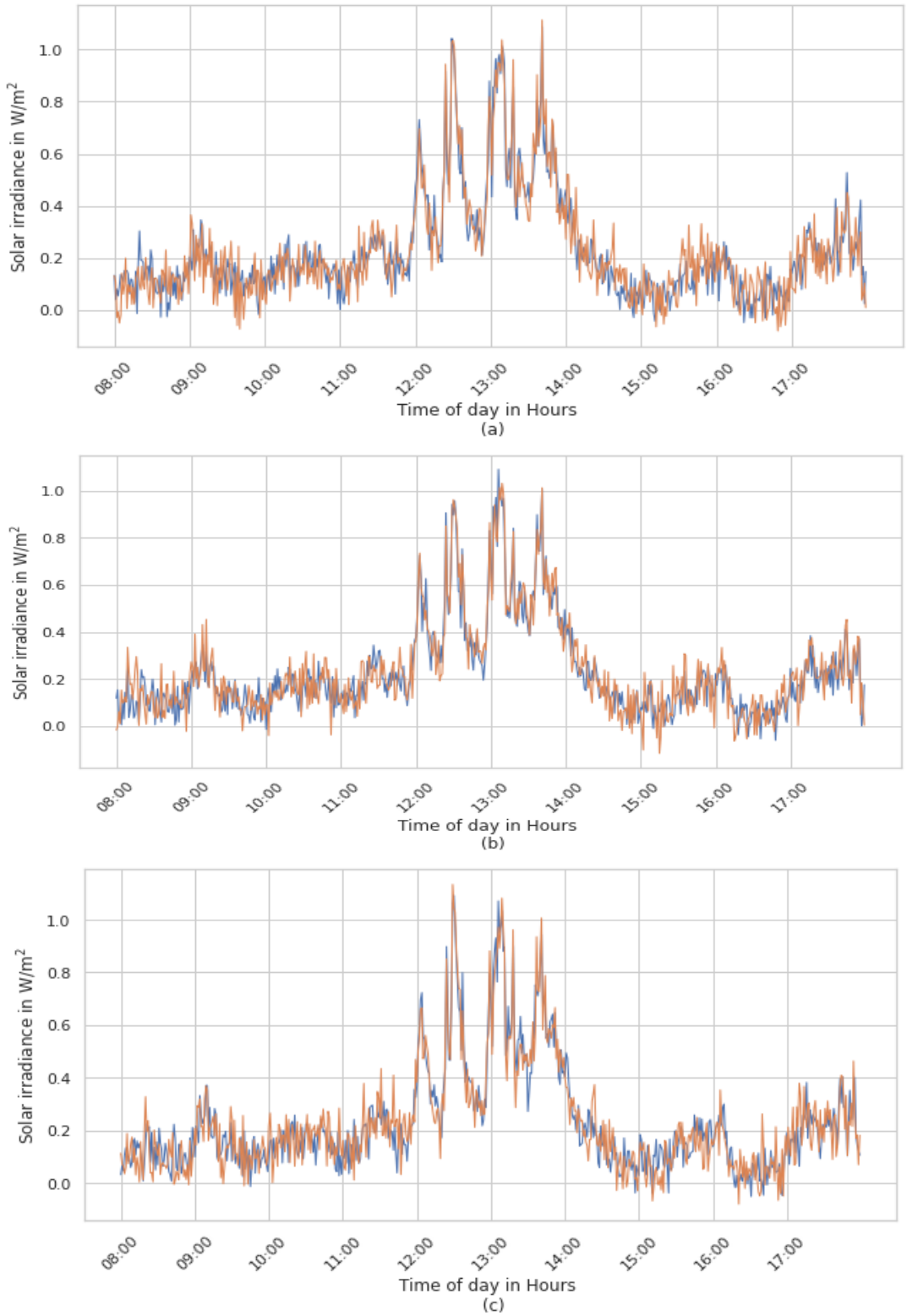


Figure 4. 23 PWS vs estimated VWS solar irradiance measurements for a moderately cloudy Day at each PWS, R06 (a), Ravenel (b) and Airport (c)

Table 4.6 Validation results of VWS weather estimation for solar irradiance

Site	*Day Category*	*MSE*	*MAPE*	*MAE*	*RMSE*
Ravenel	Moderately cloudy	16.8340	11.0596	17.0600	4.1029
R06	Moderately cloudy	15.0150	13.0594	15.0590	3.8749
Airport	Moderately cloudy	13.0160	24.6000	14.0060	3.6078
Ravenel	Cloudy	23.0204	18.1596	21.0160	4.7980
R06	Cloudy	22.0192	14.1594	25.0159	4.6925
Airport	Cloudy	21.0196	19.1600	22.0160	4.5847

VWS are then developed using the methodology outlined in previous sections. The selected stations and their locations are shown as in Table 4.5.

Estimation of Wind speed and Direction

The measured wind speed for each PWS is as shown in Figure 4.24. Wind speed is mutated through the usage of IDW weighting as previously explained. Cubic distances are used to calculate IDW for locations given in Table 4.5.

d

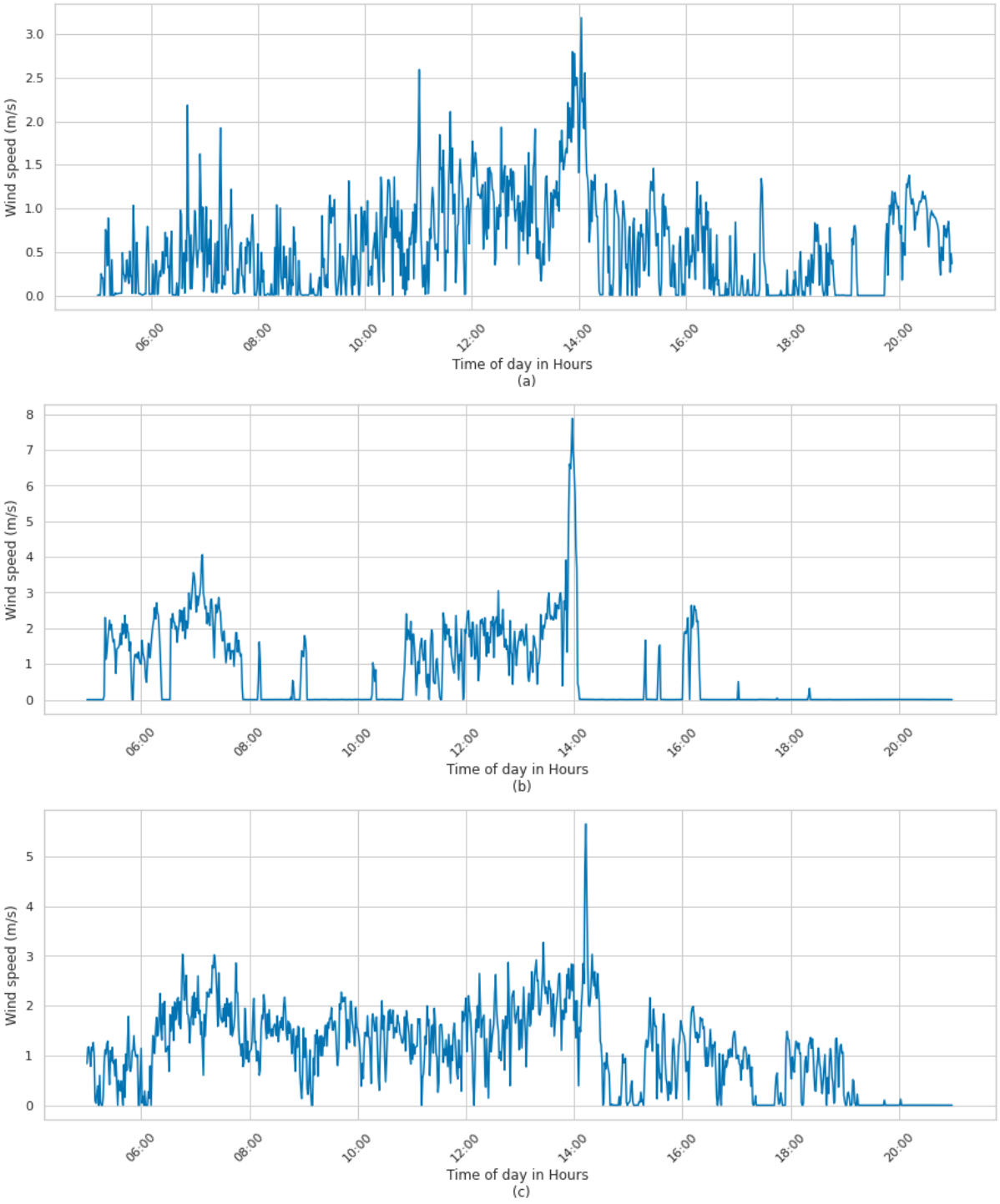


Figure 4. 24 PWS wind speed measurements over a day for Ravenel (a), Airport (b) and R06 (c)

These values are converted to vector values for IDW calculation. A temperature scale variation ranging from ±3°C based on the lookup table for wind speed at time t in Table 4.3 is used to change temperature values at the site. The wind speed values obtained at each time

instant t used to look up temperature cooling effect. Wind speed and direction is obtained in m/s and angles in degrees respectively, shown in Figures 4.25- 4.34.

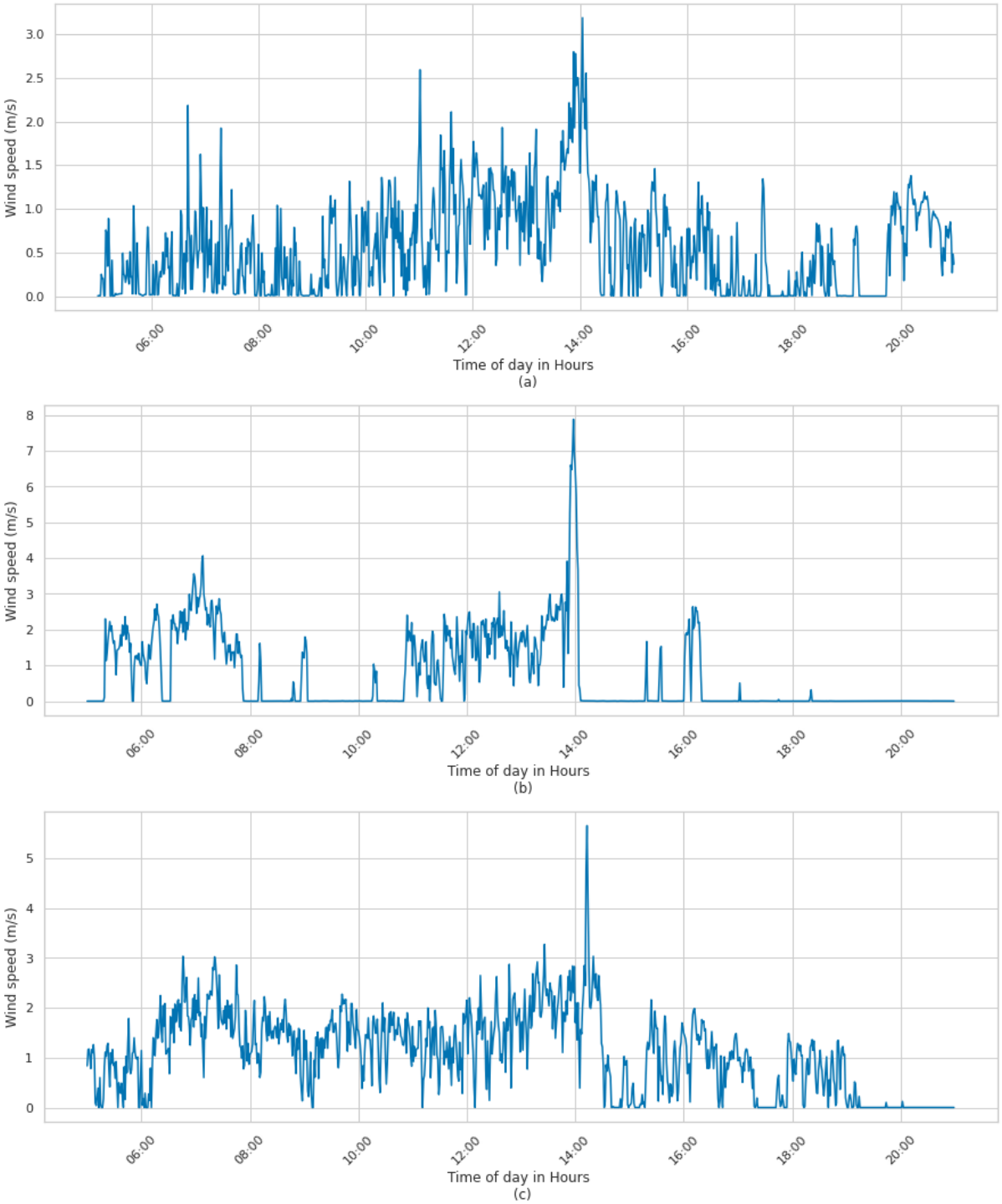


Figure 4.25 VWS wind speed measurements over a day for locations (a)-(c)

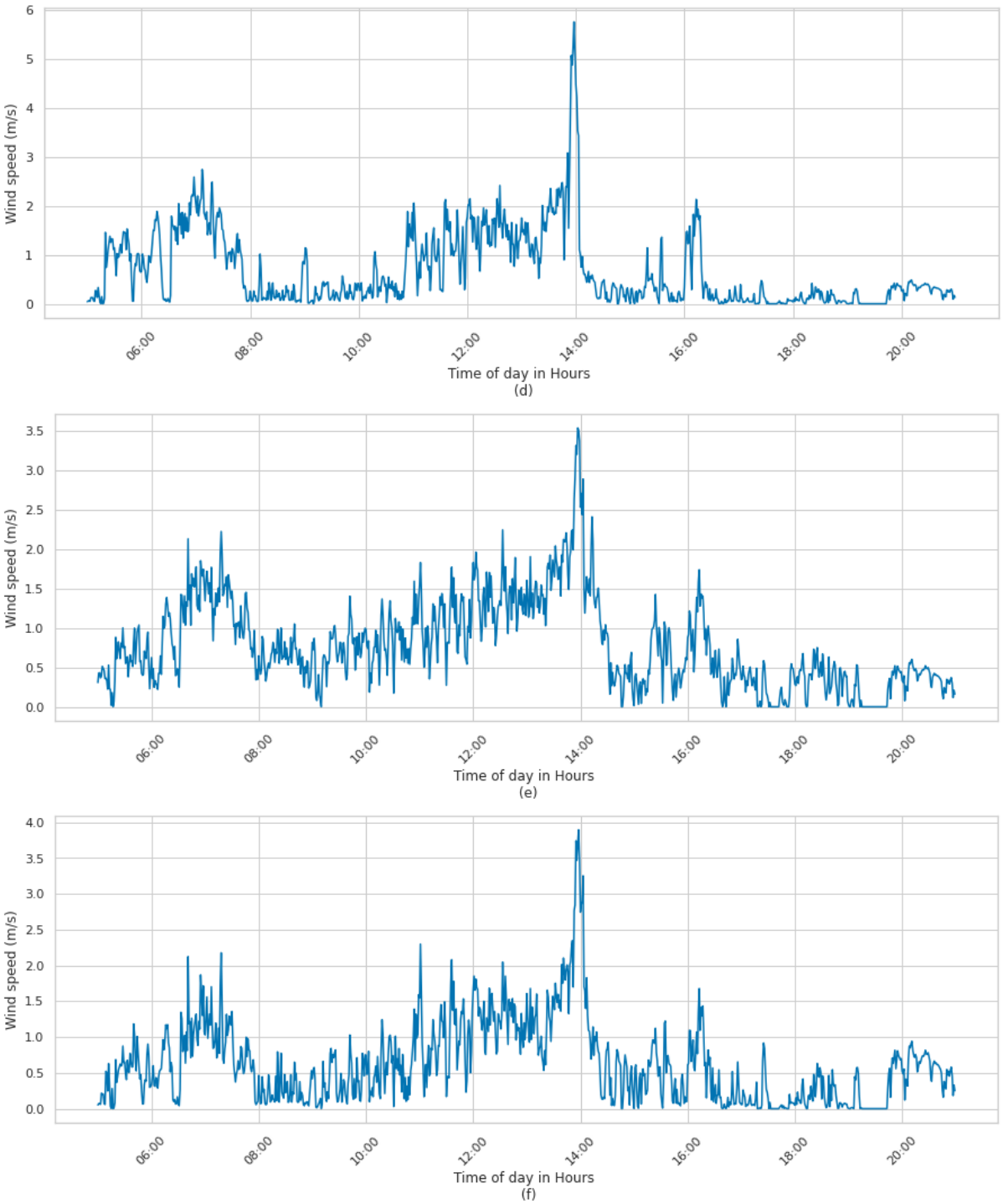


Figure 4.26 VWS wind speed measurements over a day for locations (d)-(f)

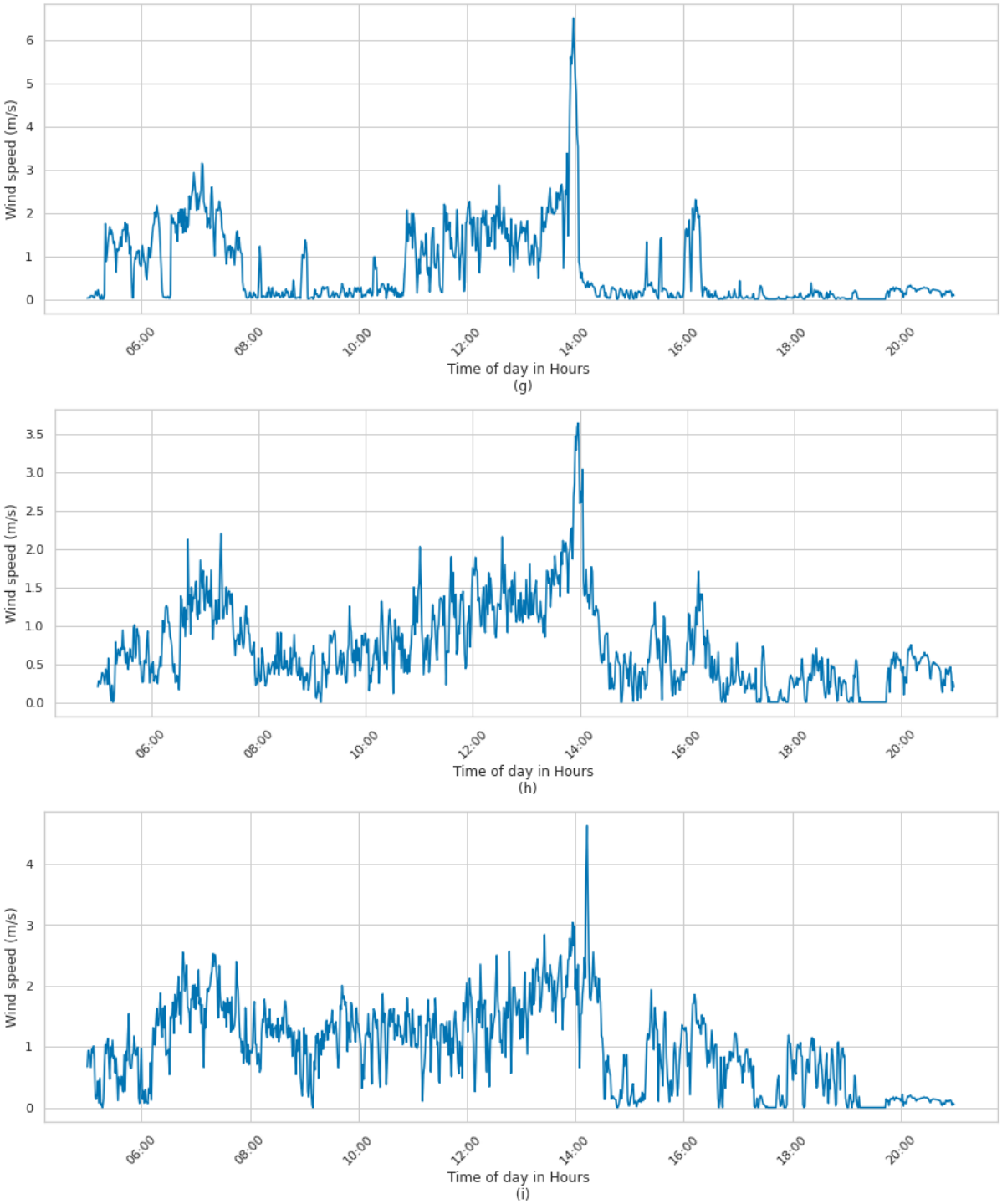


Figure 4.27 VWS wind speed measurements over a day for locations (g)-(gi)

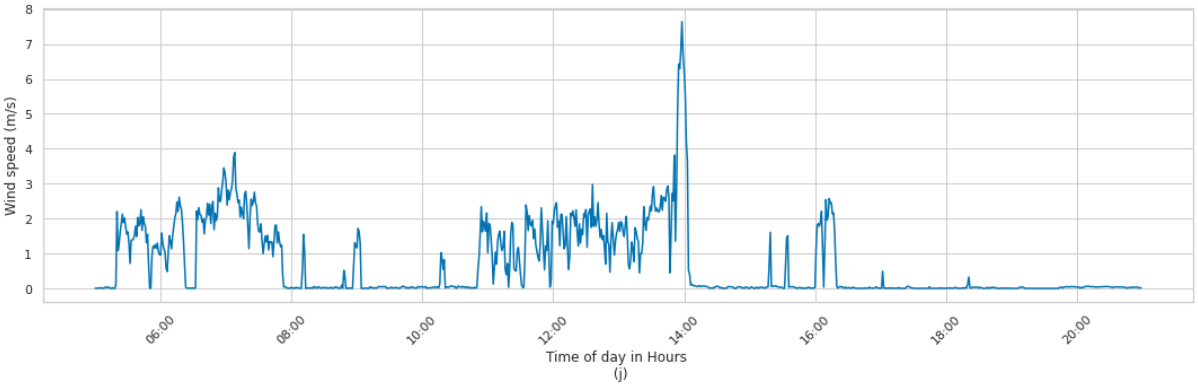
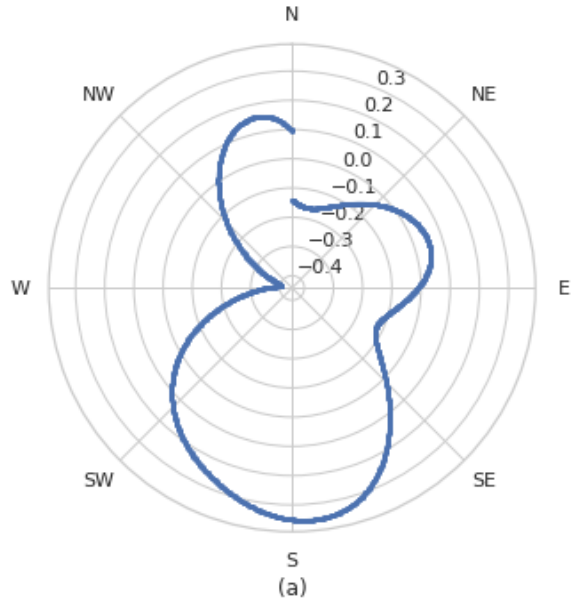
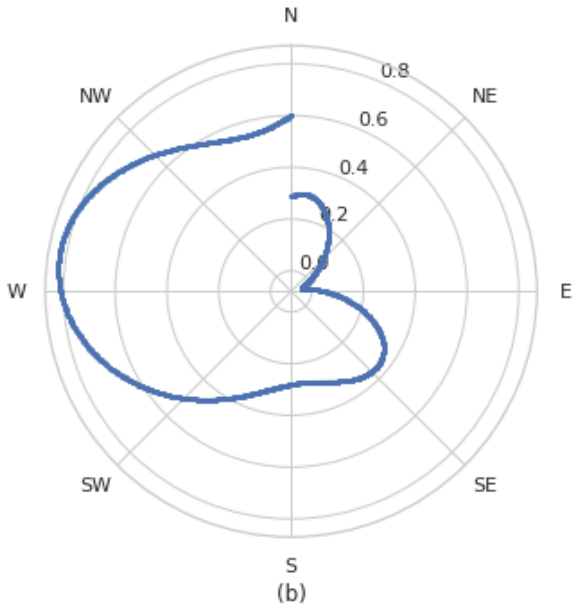
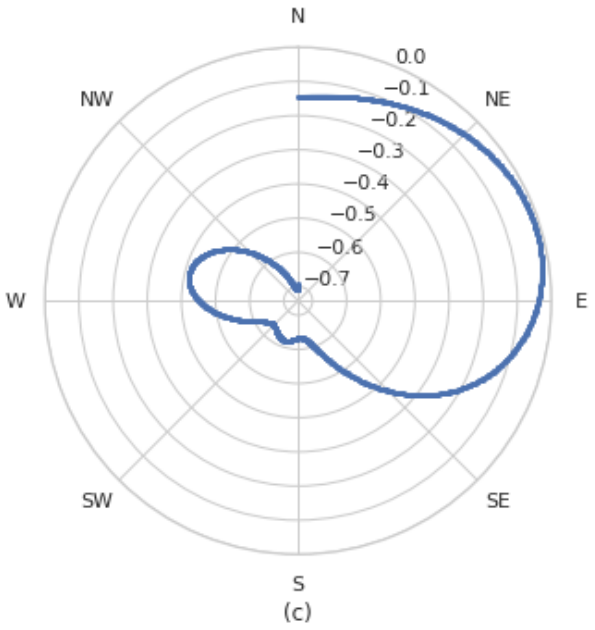


Figure 4.28 VWS wind speed measurements over a day for locations (j)

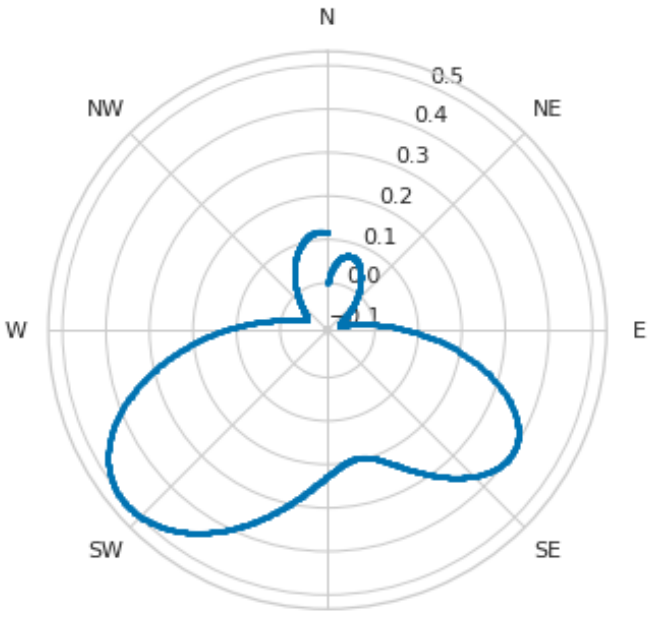




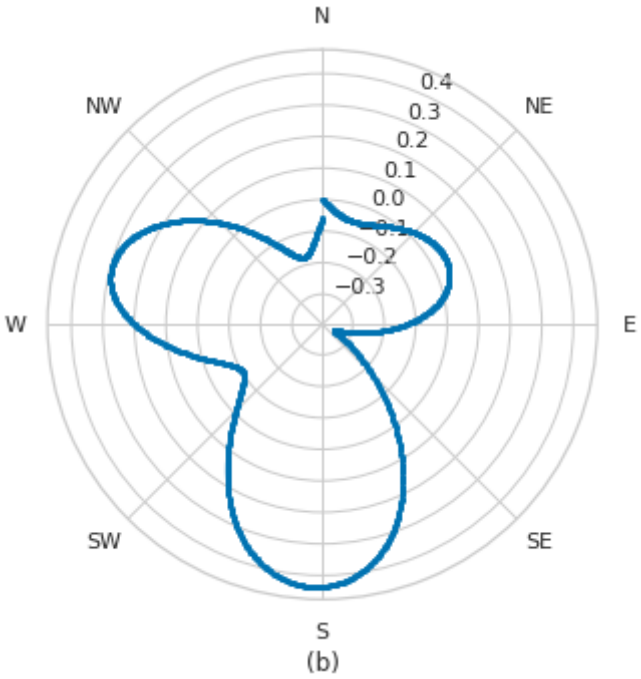


(c)

Figure 4.29 PWS wind speed measurements over a day for Ravenel (a), Airport (b) and R06 (c)

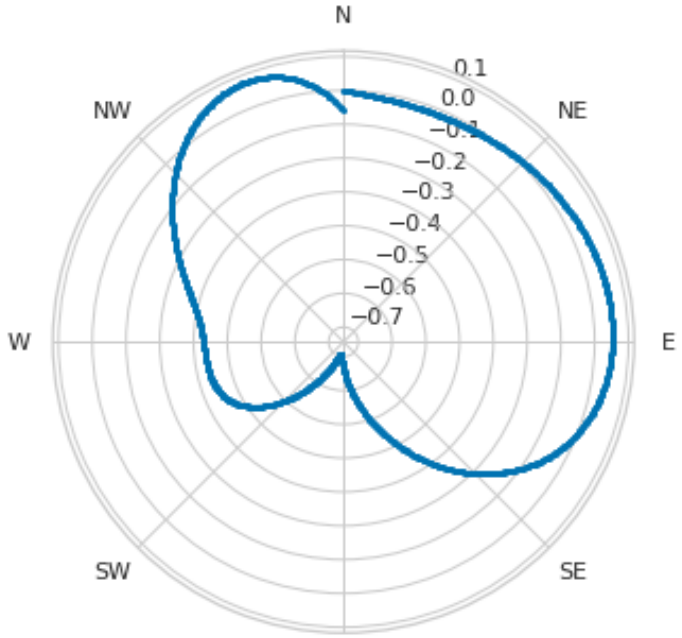


(a)

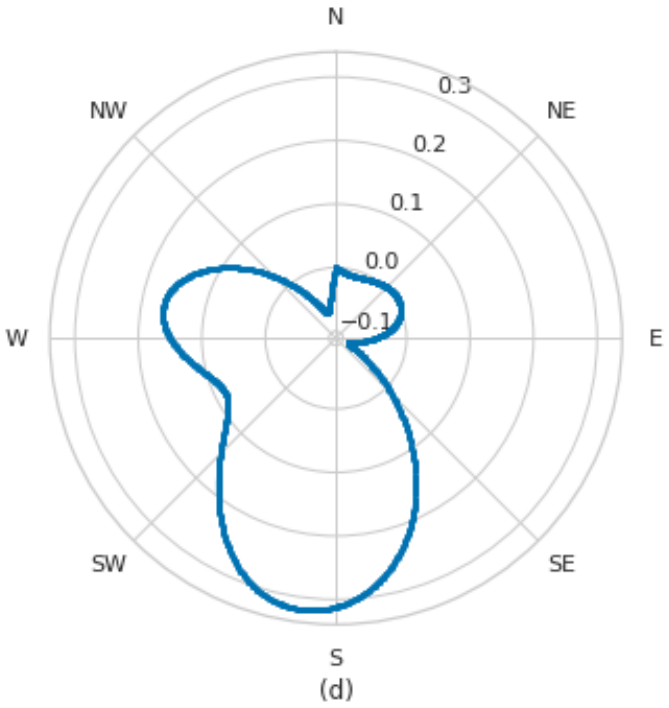


(b)

Figure 4.30 VWS wind speed directions over a day for locations (a)-(b)

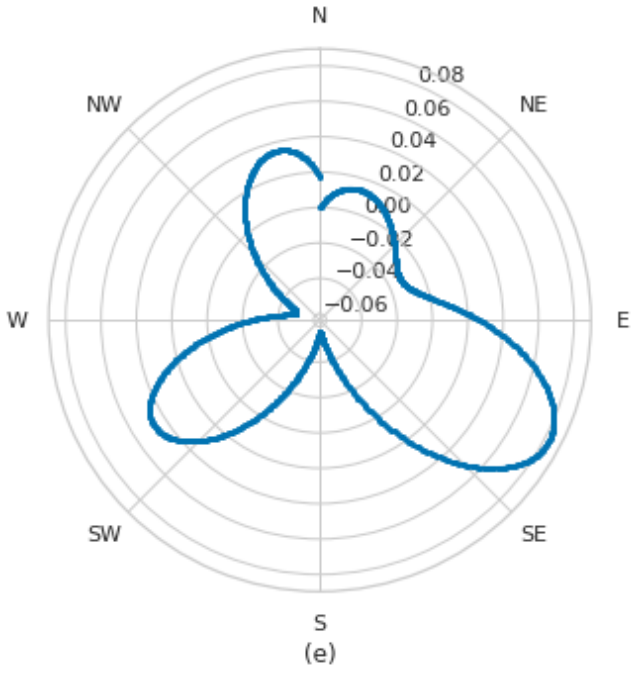


(c)

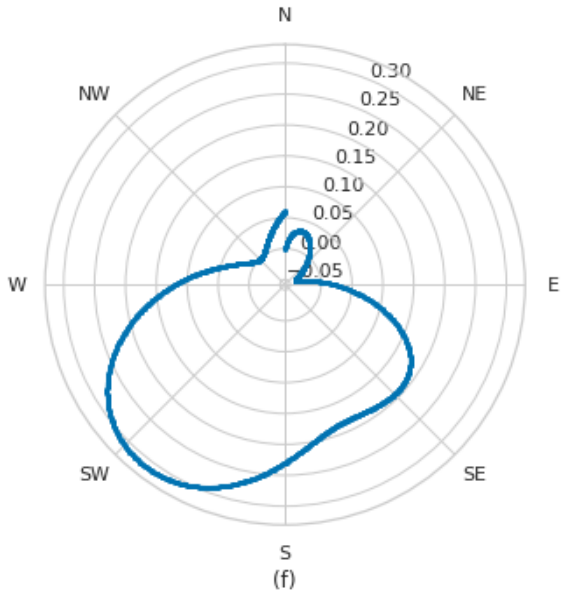


(d)

Figure 4.31 VWS wind speed directions over a day for locations (c)-(d)

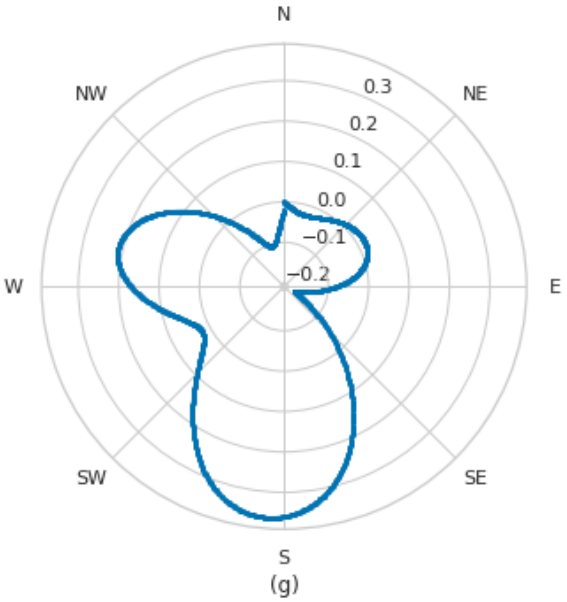


(e)



(f)

Figure 4.32 VWS wind speed directions over a day for locations (e)-(f)



(g)

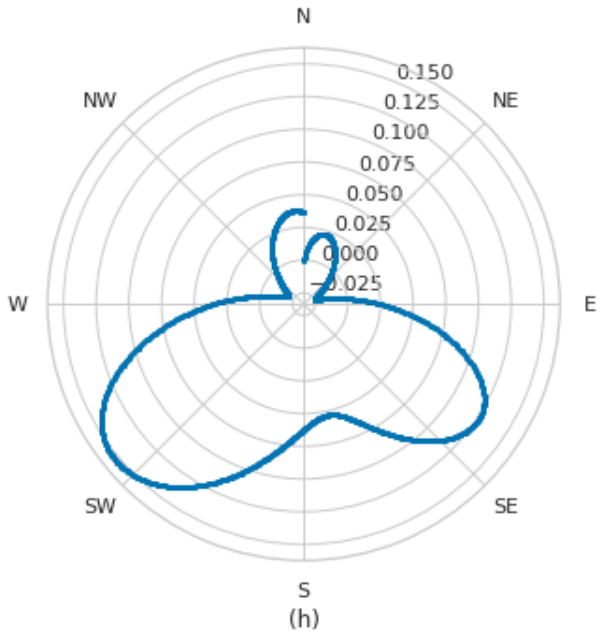
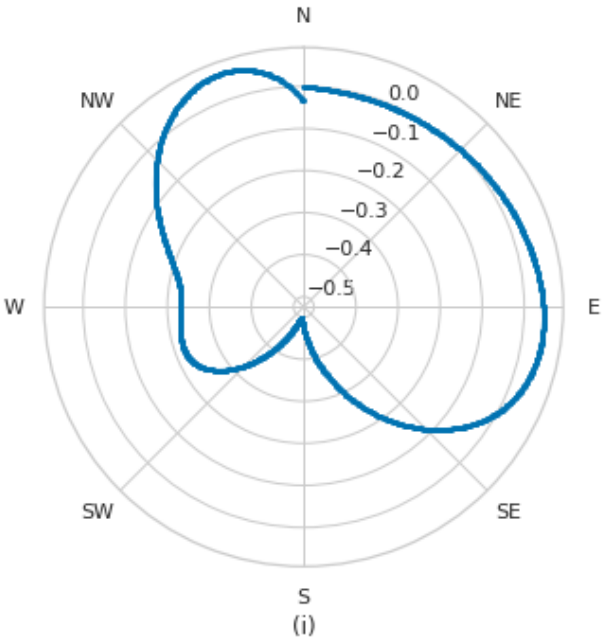


Figure 4.33 VWS wind speed directions over a day for locations (g)-(h)



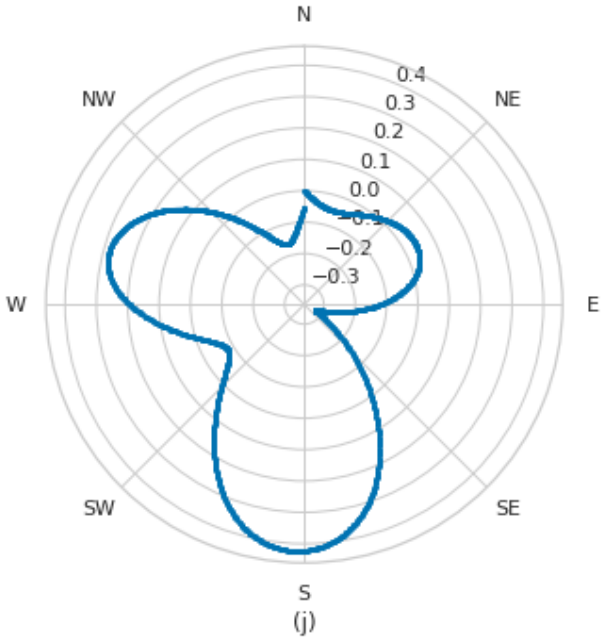


Figure 4.34 VWS wind speed directions over a day for locations (i)-(j)

Figure 4. 35 Physical (blue) vs Virtual (red) weather stations wind directions and speed for a single instant

Estimation of Temperature

Temperature is estimated through the application of a +/- difference of temperature with the measurement obtained from the nearest PWS site. The temperature difference to be applied is looked up from the corresponding wind speed difference from the given nearest PWS, through the lookup table, Table 4.4

The results of applying W_e to temperature, resulting in T_i measurements at the given locations are seen in figure 4.15, 4.16 and 4.17 for a given day

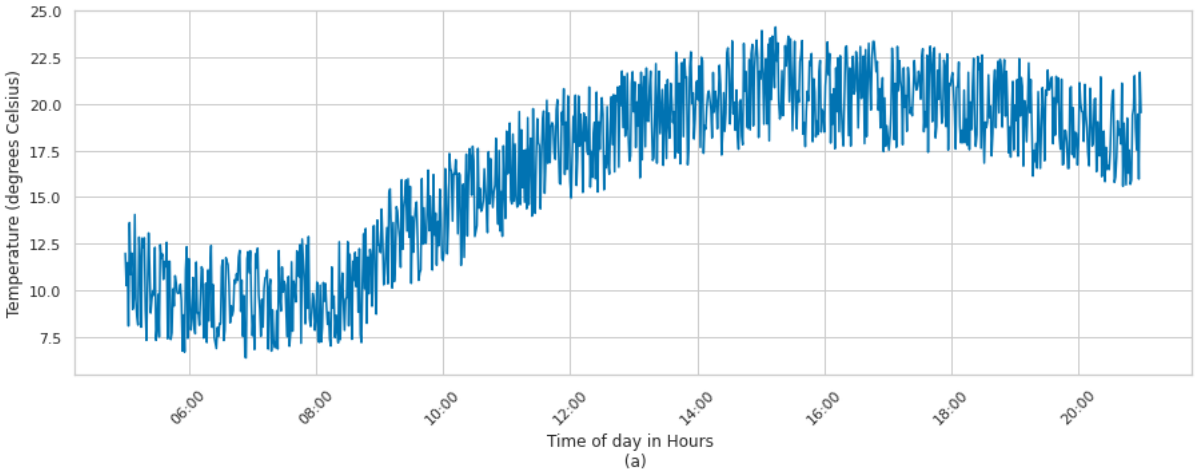
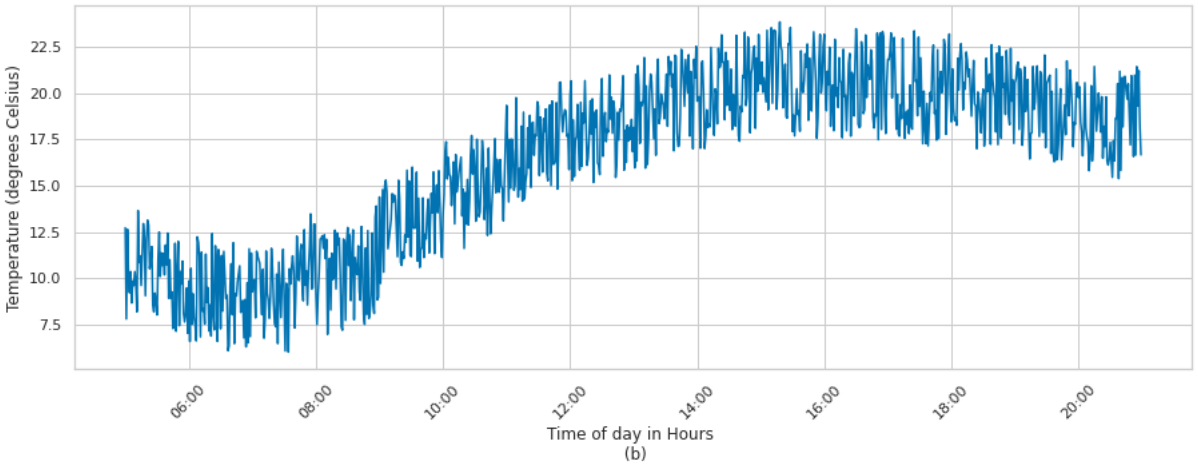
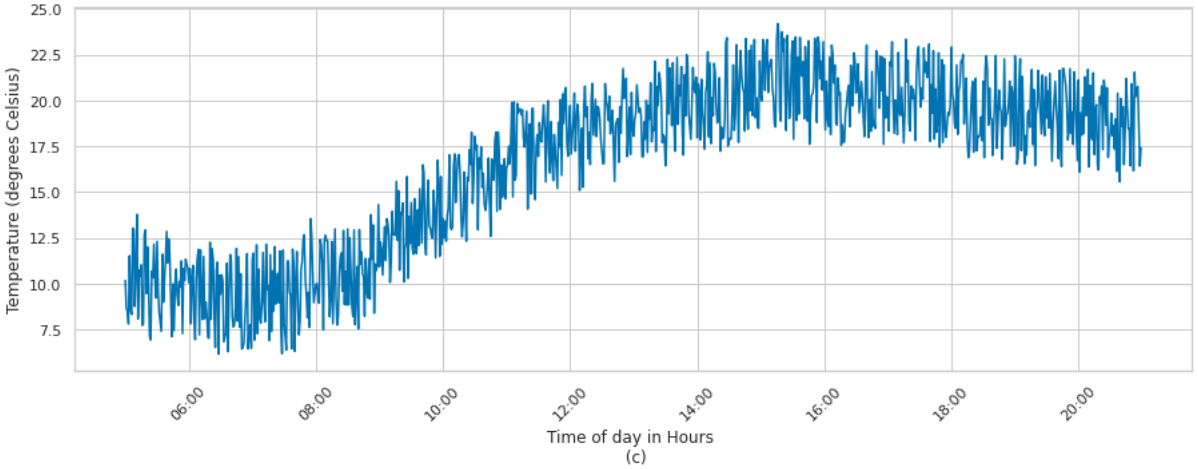


Figure 4. 36 VWS Temperature measurements over a day for locations (a)





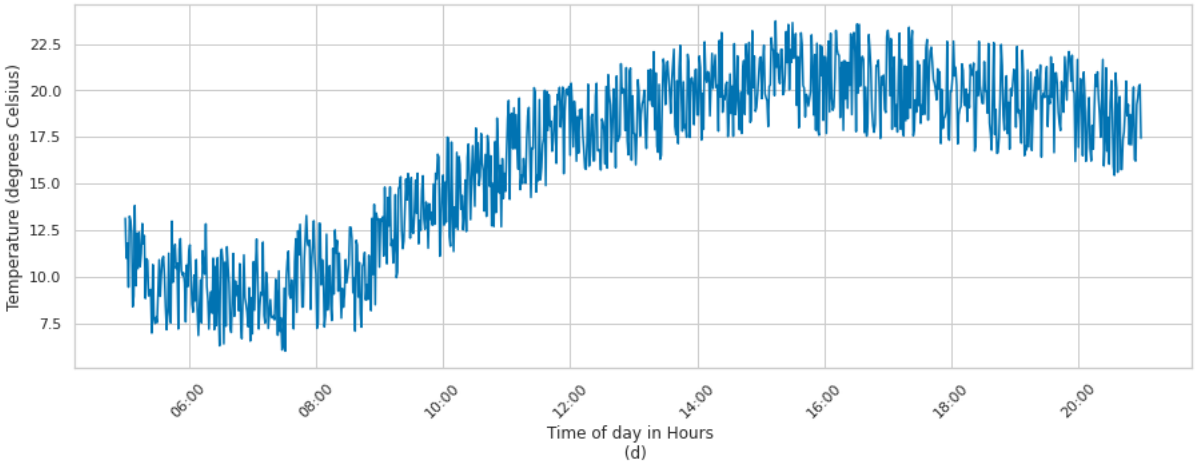
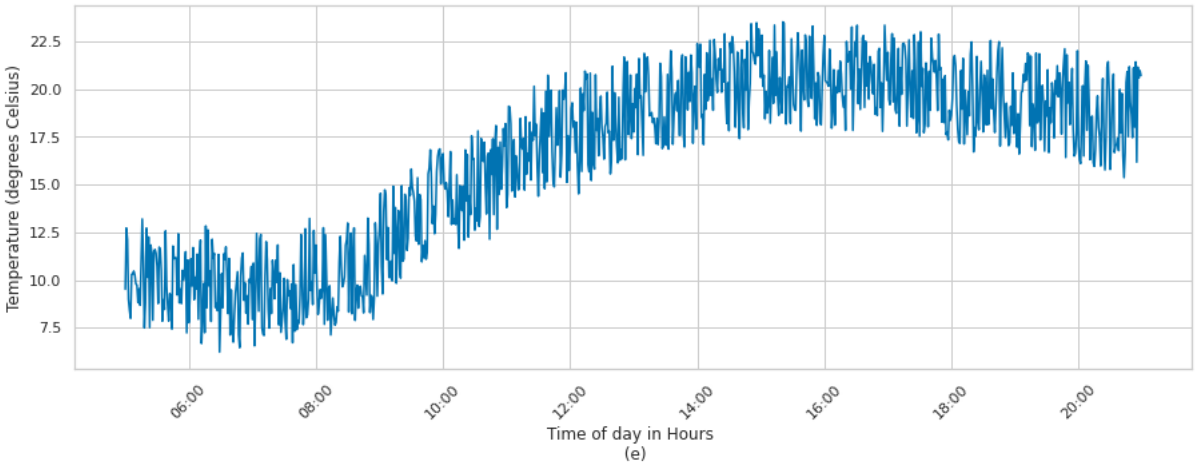
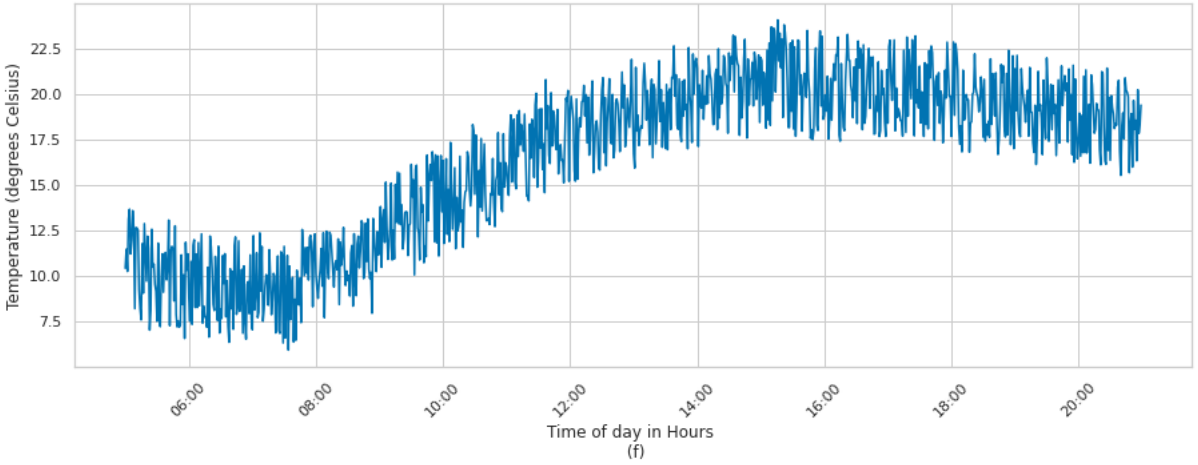


Figure 4. 37 VWS Temperature measurements over a day for locations (b) - (d)





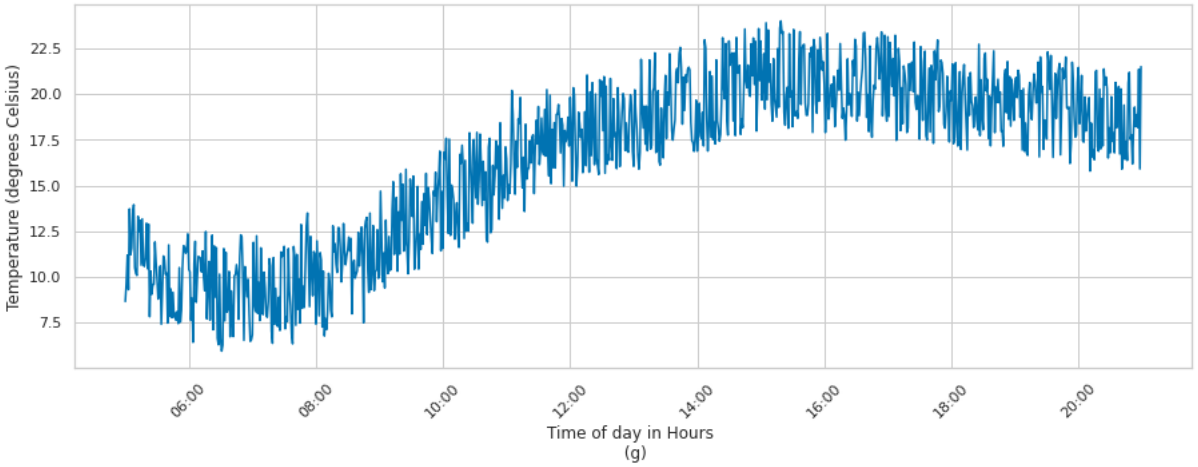
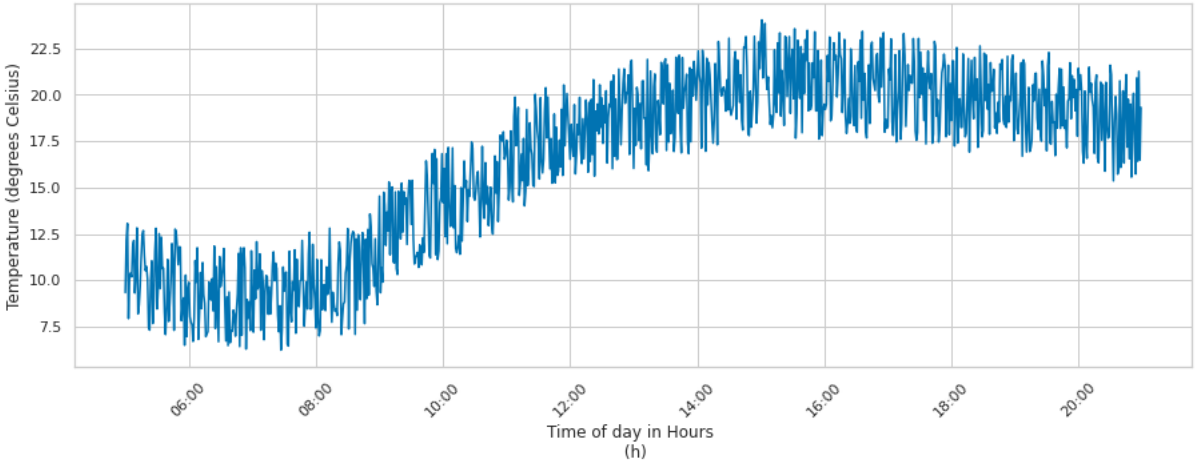
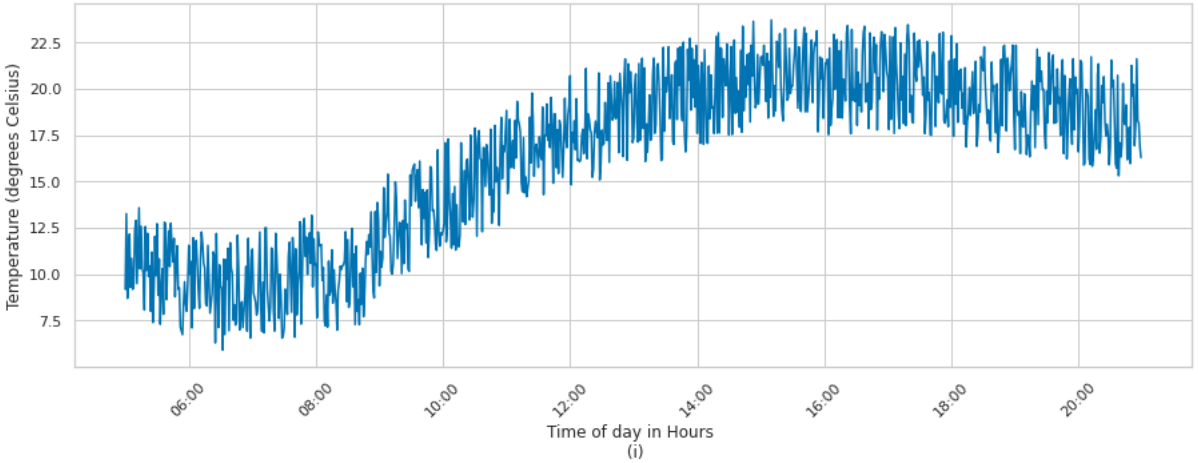


Figure 4.38 VWS Temperature measurements over a day for locations (e) - (g)





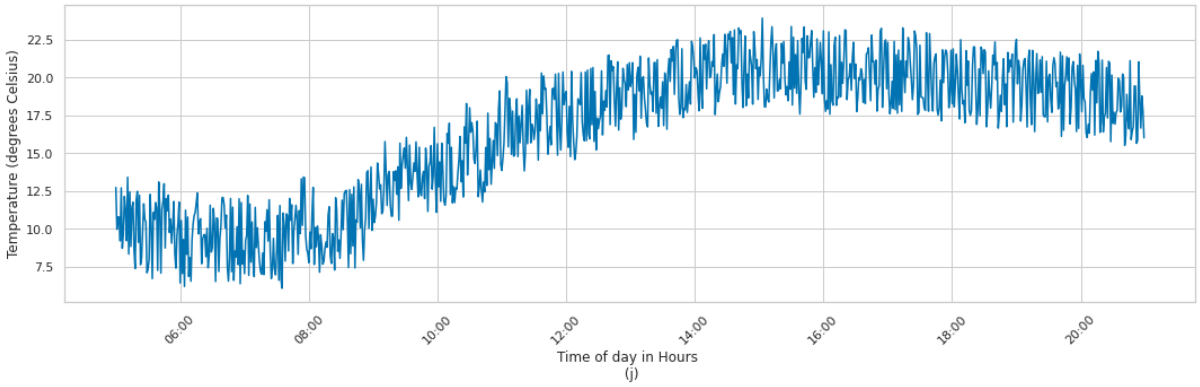


Figure 4.39 VWS Temperature measurements over a day for locations (i) - (j)

Estimation of solar irradiance

10 output mutated solar irradiance populations are therefore obtained through mutating data streamed from the PWS using the equations described previously, above at 60, 90 and 120 second intervals.

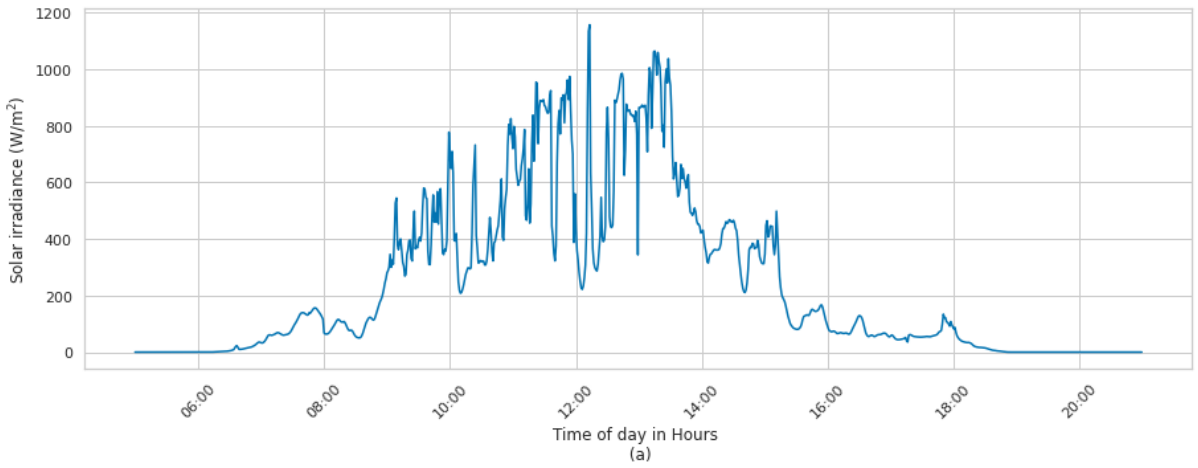
The outputs from IDW estimations, S_{pred} and cloud impact C_d for a corresponding, mutated solar irradiance output S_m for a single time instant t over the 10 locations are as shown in Table 4.6.

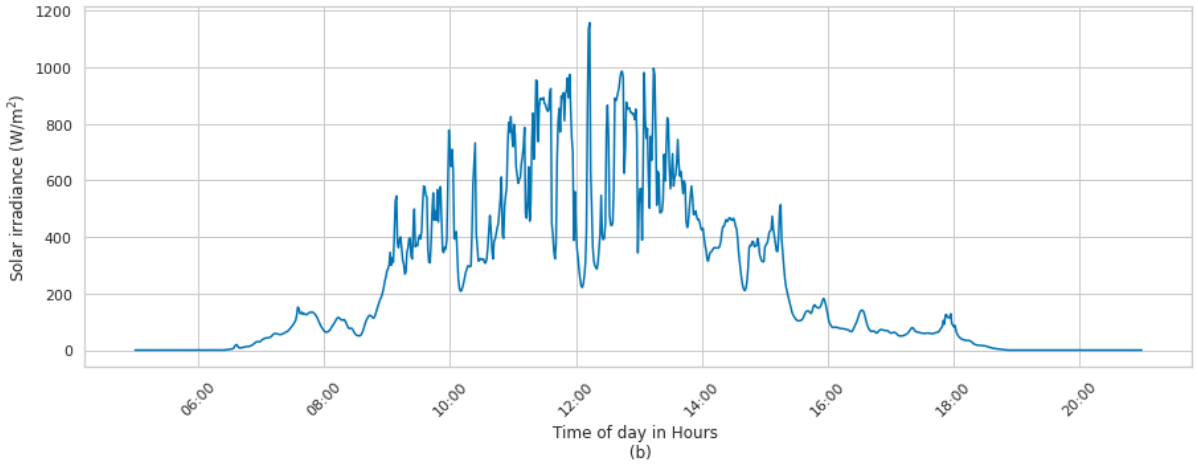
Table 4.6 Mutation parameters for various cloud coverage categories for 60 seconds data

Mutation parameter	*Sunny day*	*Moderately Cloudy day*	*Cloudy Day*
Spred0	59.72	95.9	652.4
Sm0	59.718	95.955	652.544
Spred1	58.14	85.5	584
Sm1	58.145	85.346	583.853
Spred2	55.98	95.9	581.2
Sm2	56.043	95.861	581.236
Spred3	58.14	85.5	584

Sm3	58.203	85.461	584.036
Spred4	58.14	85.5	652.4
Sm4	58.109	85.433	652.323
Spred5	58.14	95.9	584
Sm5	58.141	85.68	652.293
Spred6	59.72	85.5	584
Sm6	58.141	96.08	583.893
Spred7	58.14	95.9	652.4
Sm7	58.116	95.766	584.195
Spred8	55.98	95.9	584
Sm8	56.064	96.031	584.115
Spred9	59.72	85.5	584
Sm9	59.709	85.678	584.158

The solar irradiance finally obtained through the impact of cloud coverage is shown in Figures 4.18, 4.19 and 4.20





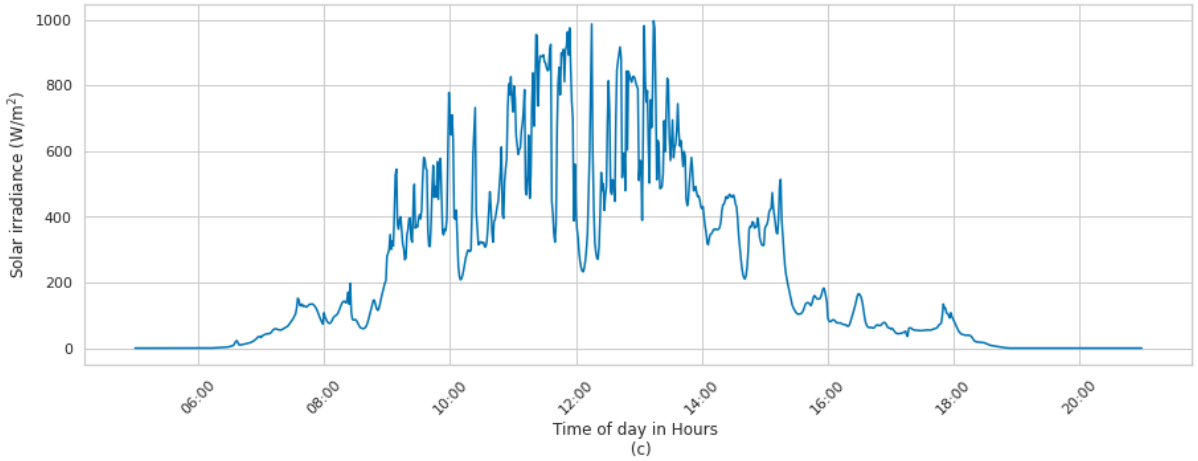
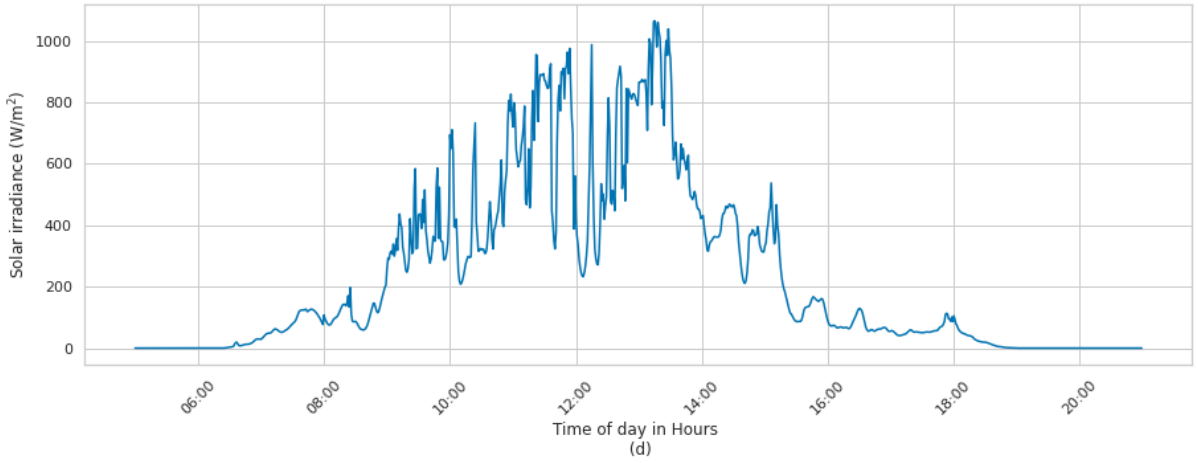


Figure 4.40 VWS Temperature measurements over a moderately cloudy day for locations (a) – (c)



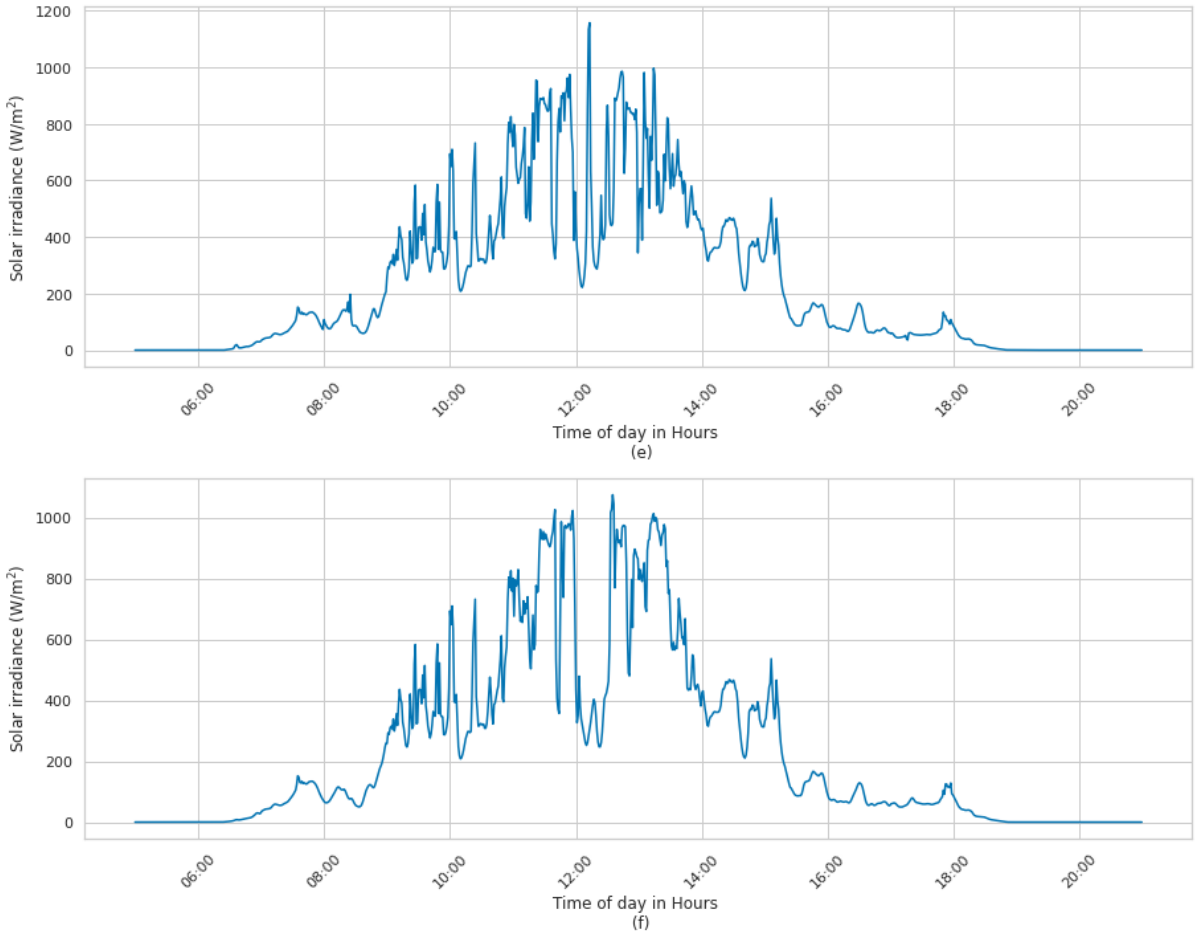
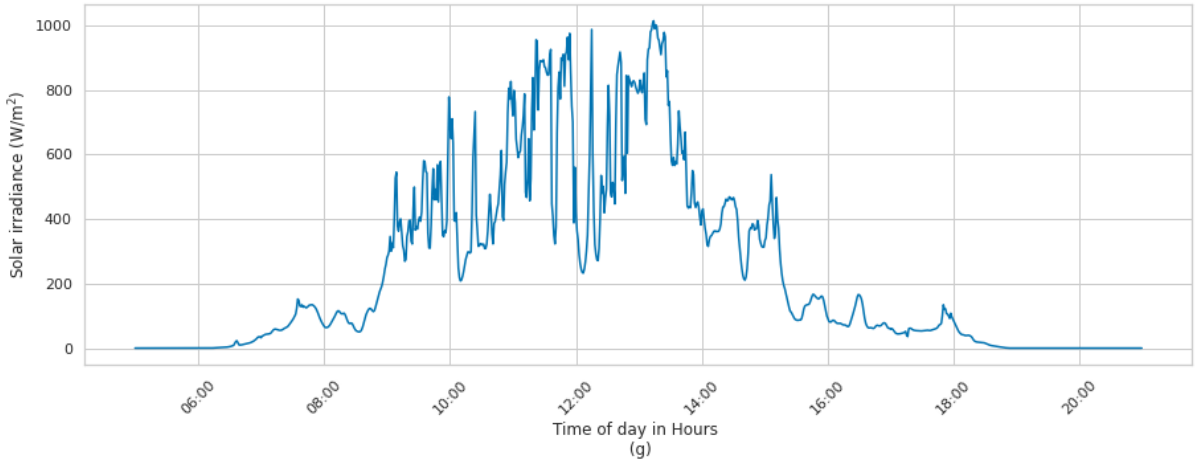


Figure 4.41 VWS Temperature measurements over a moderately cloudy day for locations (d) – (f)



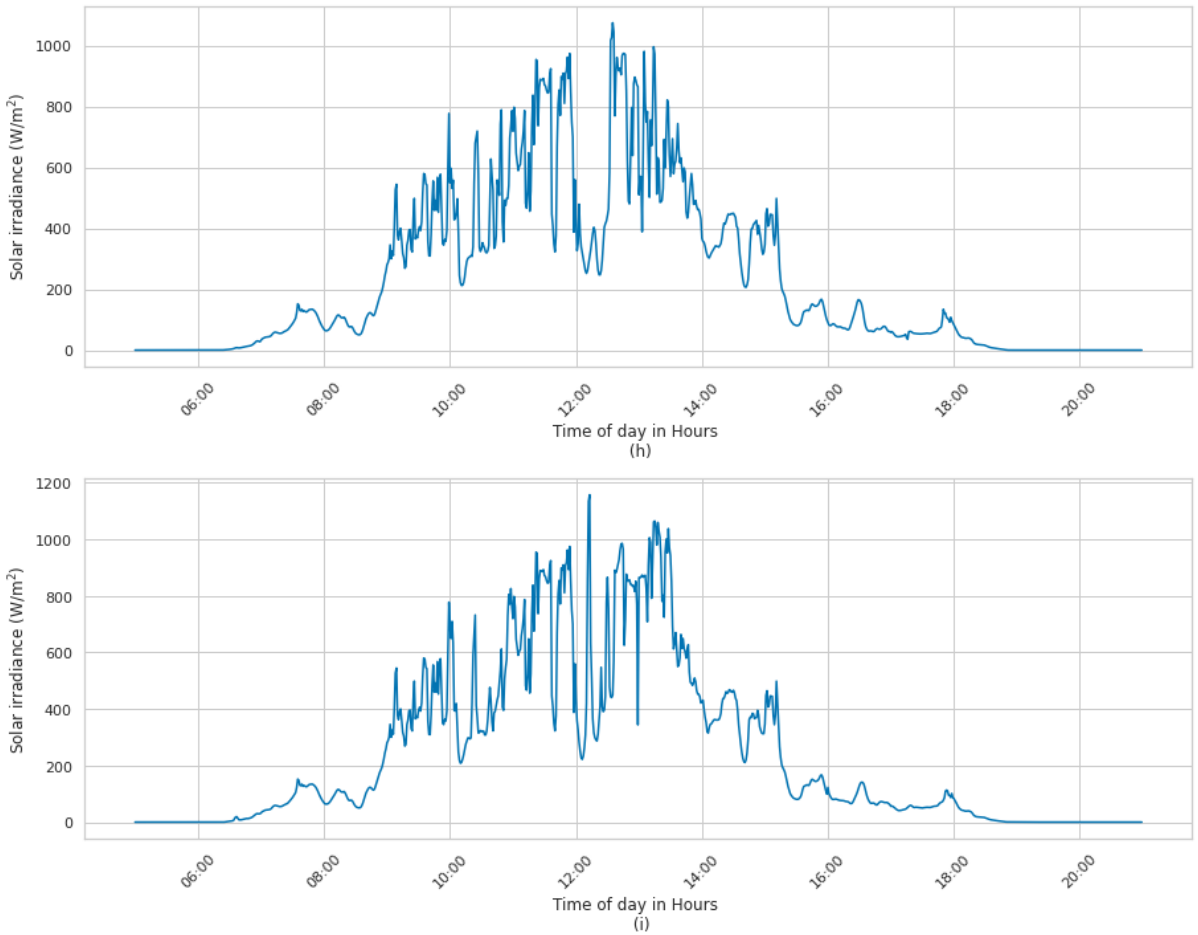


Figure 4.42 VWS Temperature measurements over a moderately cloudy day for locations (g) – (i)

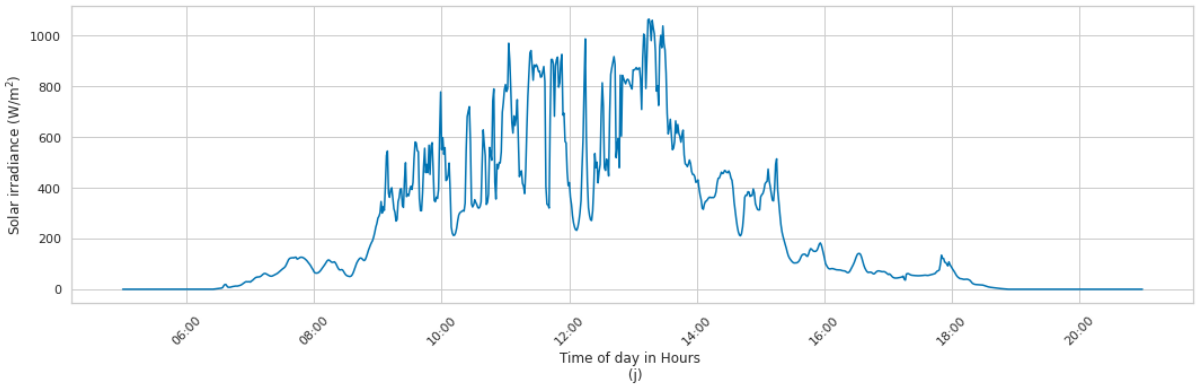


Figure 4.43 VWS Temperature measurements over a moderately cloudy day for locations (j)

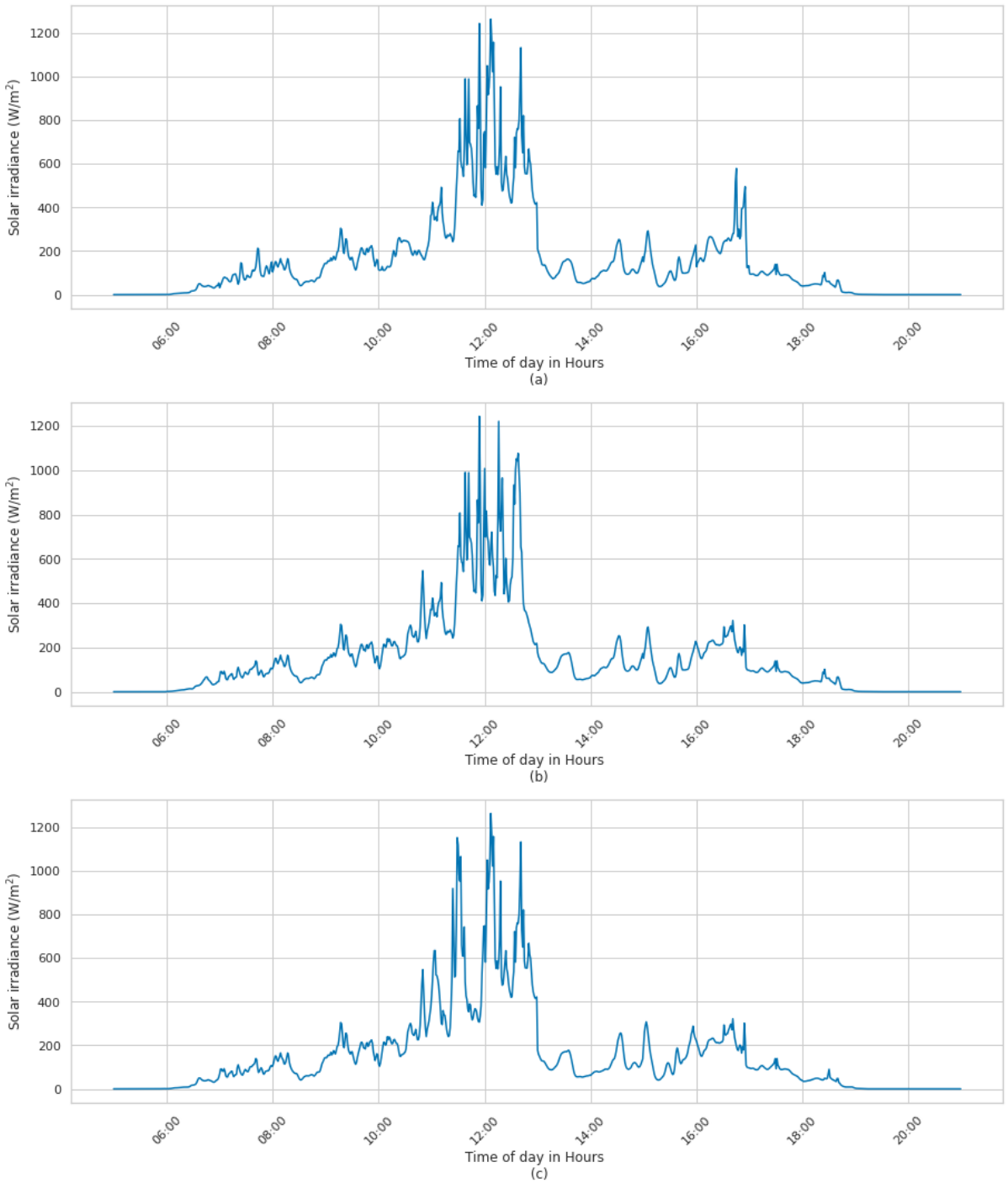


Figure 4.44 VWS Temperature measurements over a cloudy day for locations (a) – (c)

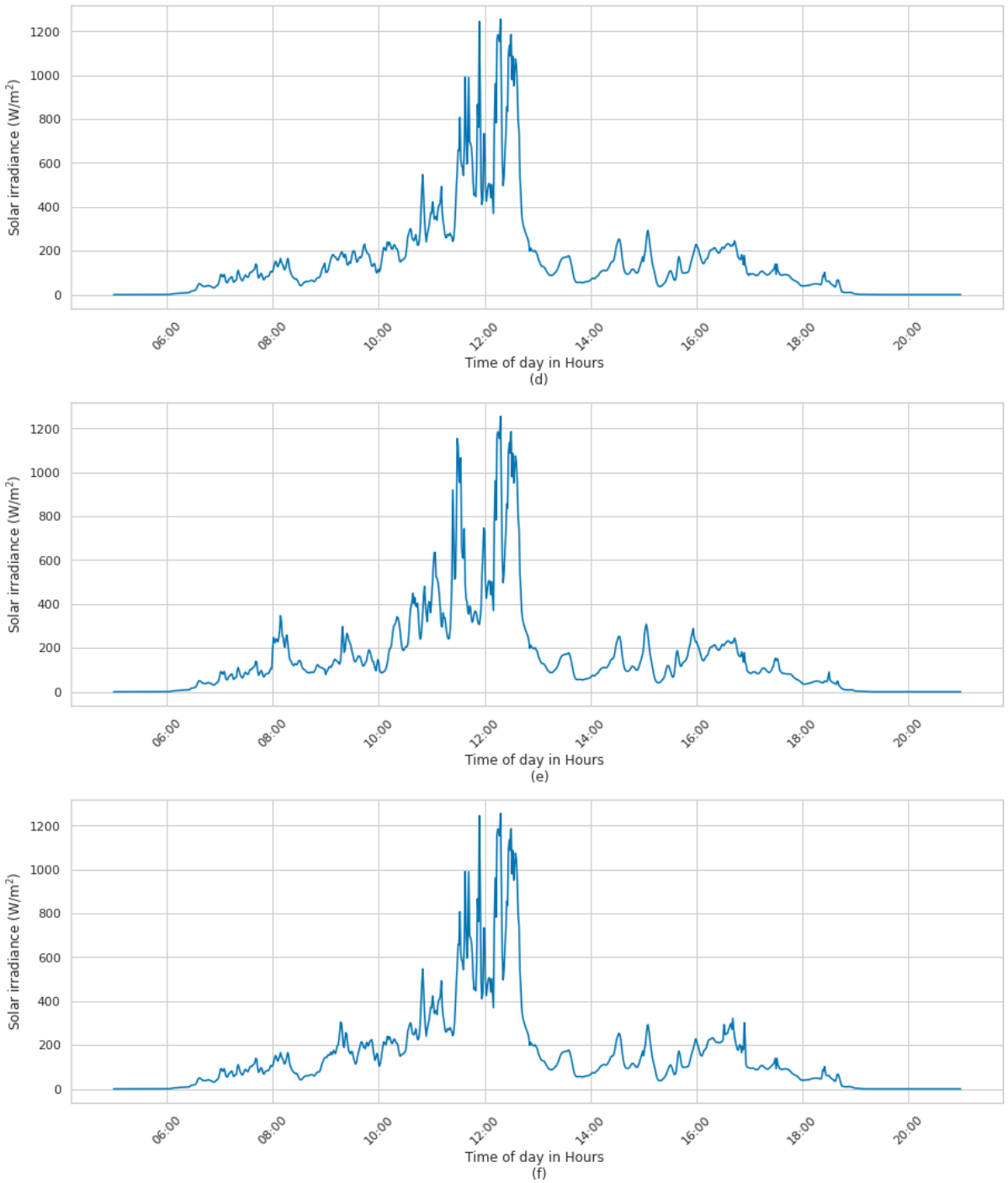


Figure 4.45 VWS Temperature measurements over a cloudy day for locations (d) - (f)

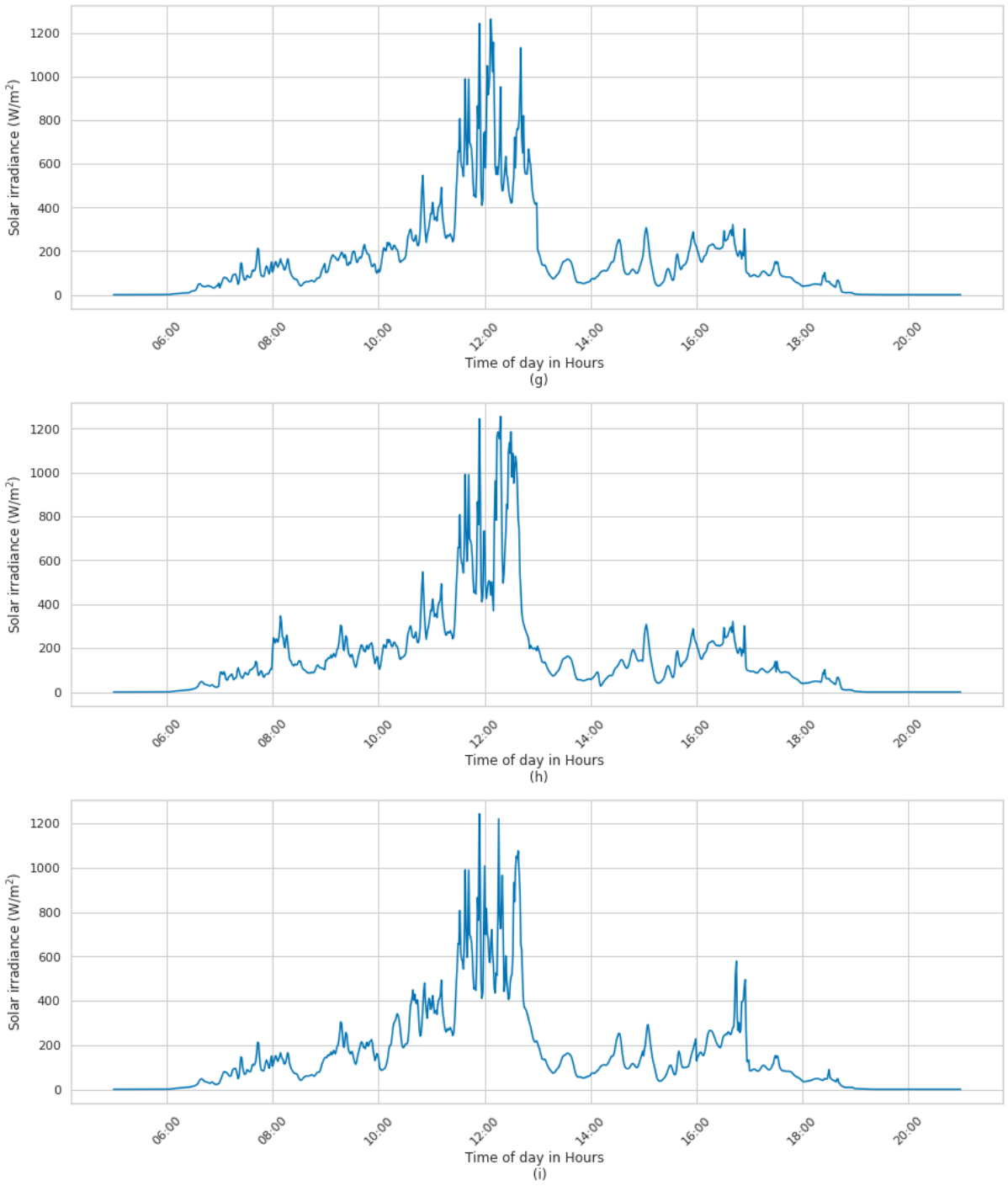


Figure 4.46 VWS Temperature measurements over a cloudy day for locations (g) - (i)

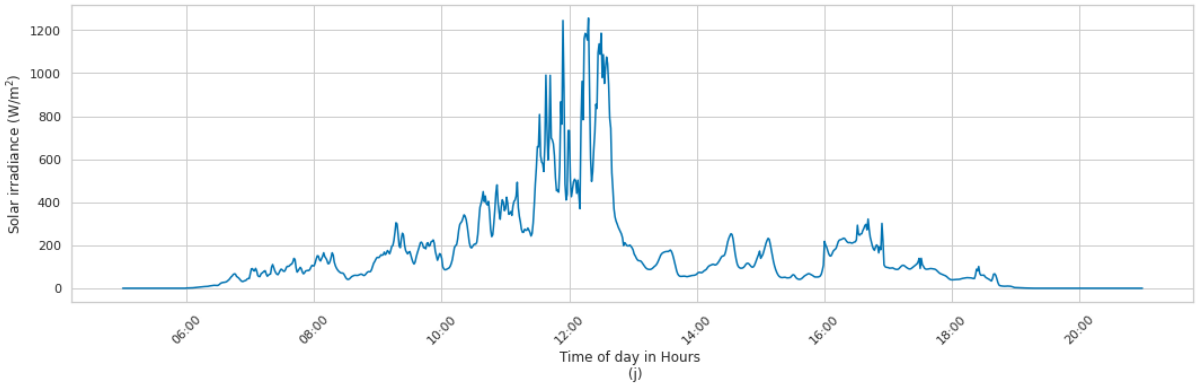


Figure 4.47 VWS Temperature measurements over a cloudy day for locations (j)

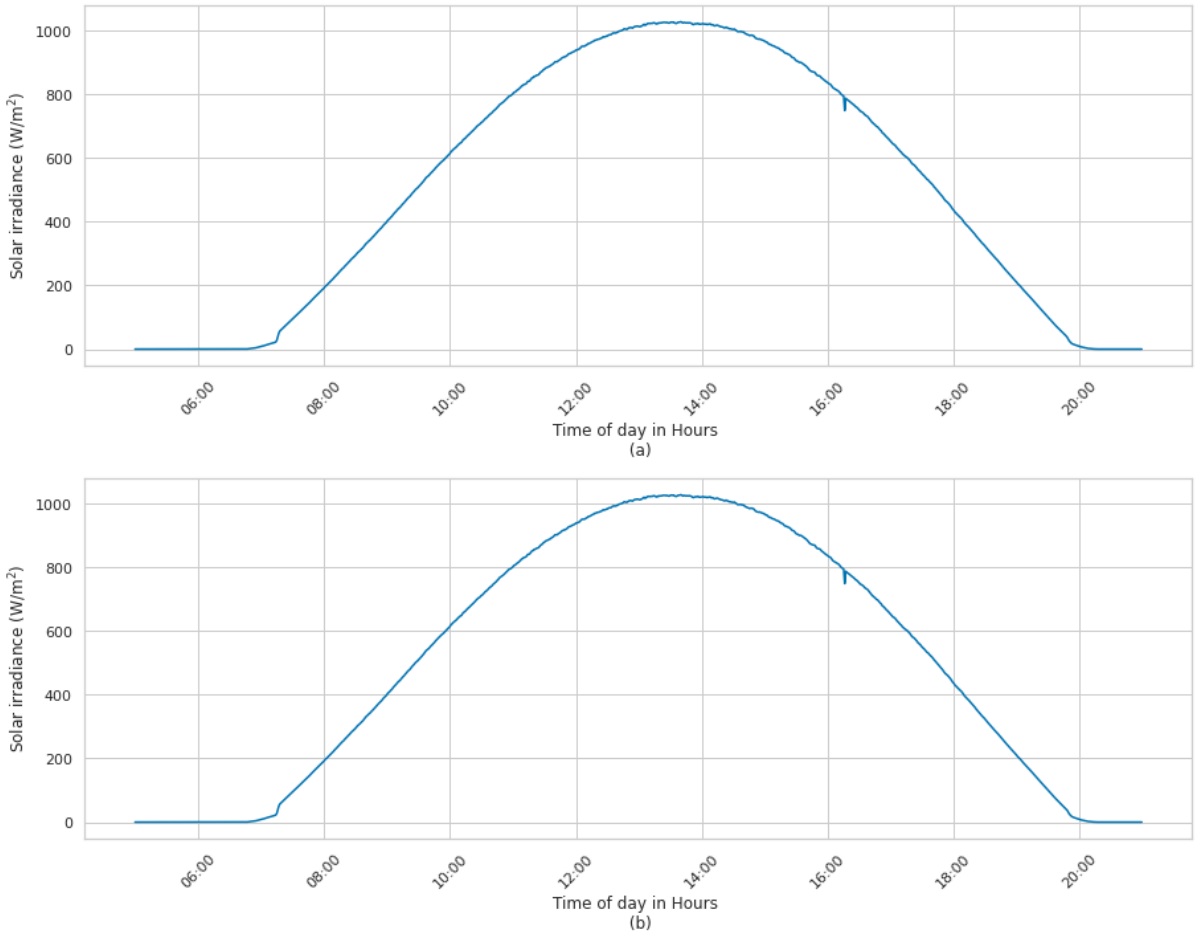


Figure 4.48 VWS Temperature measurements over a sunny day for locations (a)-(b)

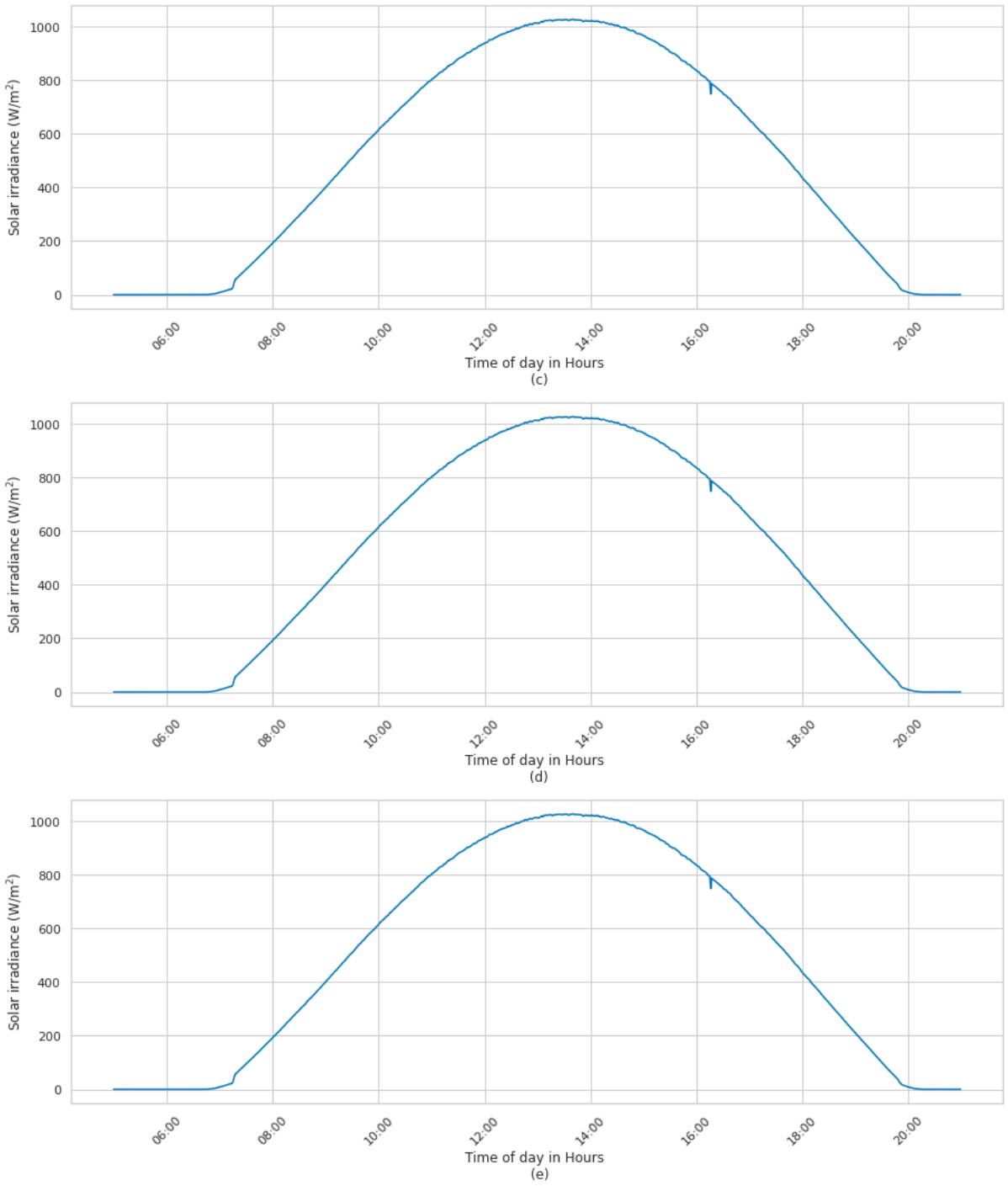


Figure 4.49 VWS Temperature measurements over a sunny day for locations (c)-(e)

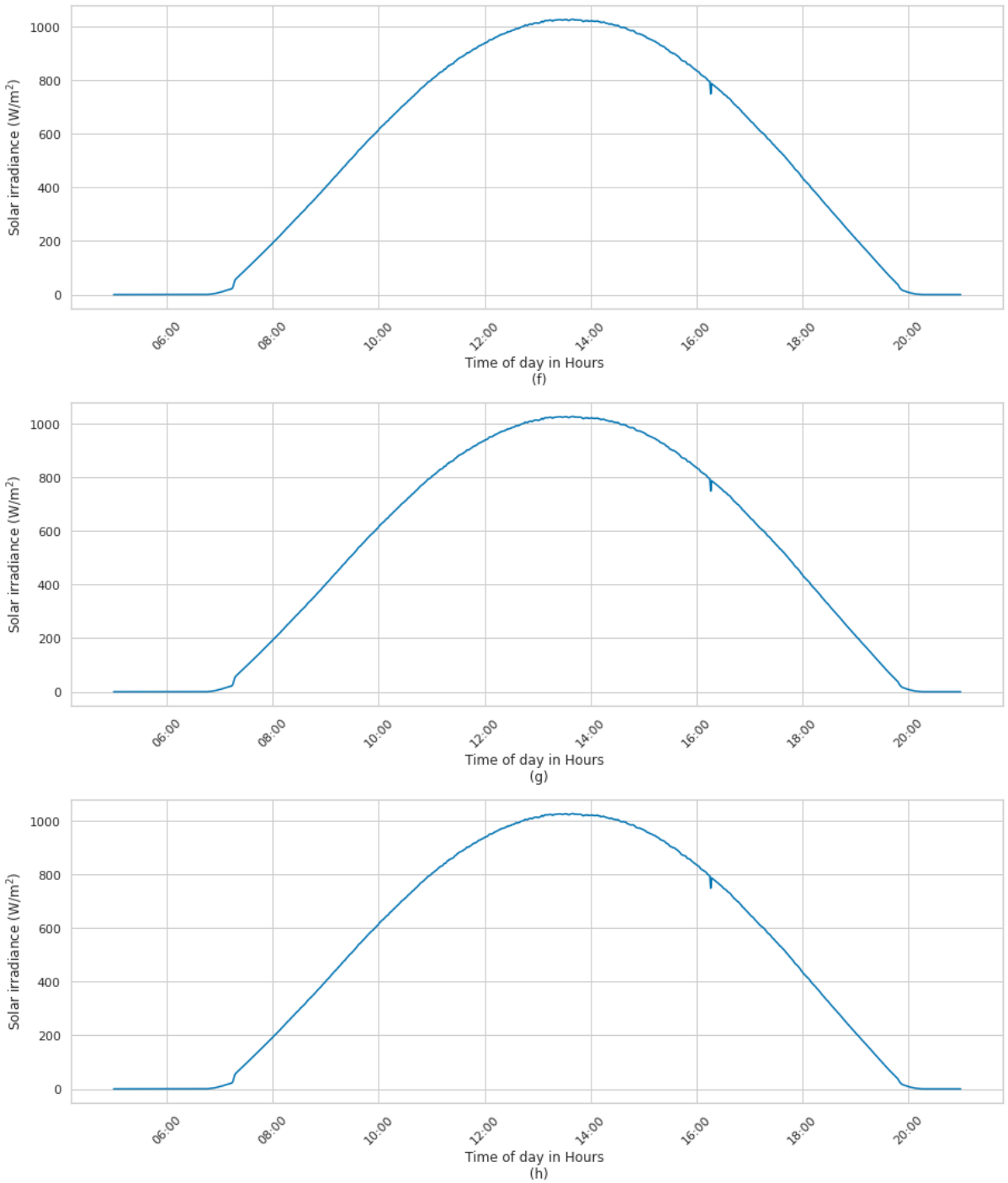


Figure 4.50 VWS Temperature measurements over a sunny day for locations (f)-(h)

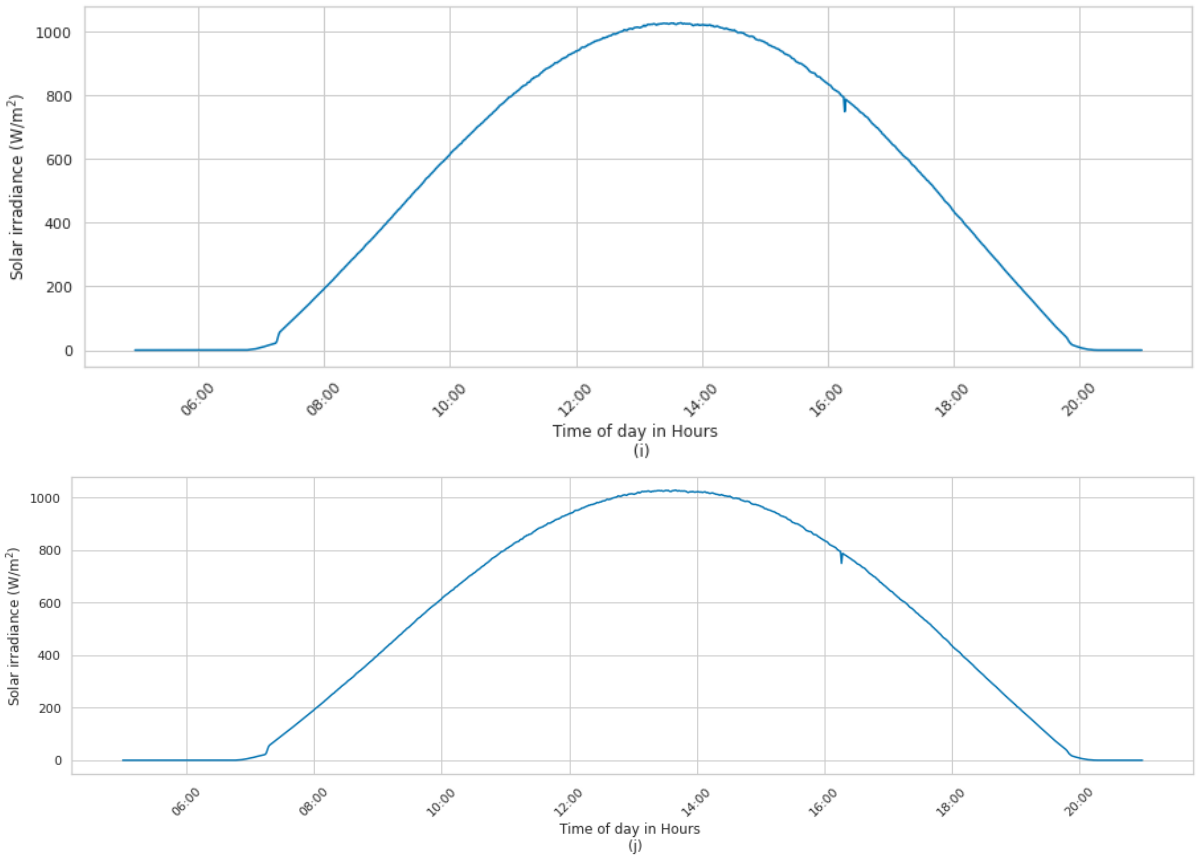


Figure 4.51 VWS Temperature measurements over a sunny day for locations (i)-(j)

Virtual real-time weather stations are developed from these physical data stations by mutating real-time solar irradiance streams from these three locations for real-time simulations and studies. These VWSs are projected to exist at localized geographical locations around the original site as seen in a map in Fig. 4.21, with accompanying real-time output streams.

Summary

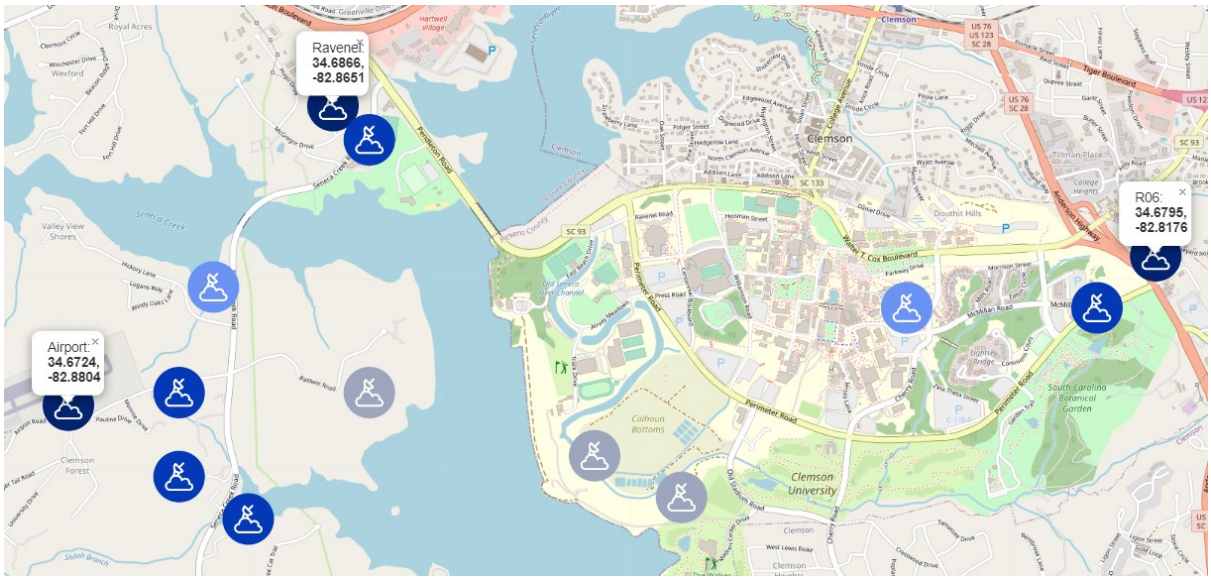


Figure 4.52 VWS and PWS developed over the area, graded from dark to light for closest similarity to the original PWS to furthest

Ten virtual weather station DTs of three existing physical weather stations were developed using intelligent mutations; incorporating inverse distance weighting and data analysis of historical and current data at the site. These VWS generate wind speed, wind direction, temperature and solar irradiance data streams in real-time with respect to real-time data measured at the PWS. This method can incorporate other weather station parameters in the mutation algorithm such as elevation, temperature or humidity by estimating its impact and variance over the location and fine-tuning the estimated output of these VWS compared to the PWS. This approach is additionally simplistic and requires little hardware investment and maintenance in the form of new measurement devices, for instance to infer solar irradiance at the location.

CHAPTER FIVE
SITUATIONAL AWARENESS AND INTELLIGENCE FOR DISTRIBUTED PV PLANTS

Introduction

Amidst the constantly evolving modern electric power distribution system (MEPDS), the high penetration of distributed energy resources (DERs), particularly PV and DPV has been a driving factor of significant operational, management, and planning challenges in the system. This calls for the integration of new technologies and frameworks to effectively address the complexities presented by DPVs through advanced monitoring and control through the leveraging of data-driven DTs, as explored in chapter 2. As further shown in Chapter 3, individual digital twin (DT) models of system entities provide valuable information about PV plant conditions. By utilizing photovoltaic (PV) plant DTs used for estimation and prediction of PV power at a physical site and harnessing data from geographically distributed weather stations (PWS), virtual weather stations (VWS) developed in chapter 4, spatial distribution of PV sites or Virtual RT photovoltaic plants (V-RT-PVPs) can be generated.

V-RT-PVPs can be used to create real-time estimations and short-term forecasts of PV plant power production over the region. Through the exploration of various scenarios involving the availability of VWS, PWS, PV plants, and PV DTs, diverse types of Distributed PV (DPV) sources can be generated, generating situational awareness and situational intelligence of PV sources within the MEPDS that can be harnessed to create CDTs and further provide cognition in the system.

DTs for DPV Situational Awareness and Situational Intelligence

The creation of CDTs for DPV situational awareness and situational intelligence is broken down into three cases:

- Where both PWS and PMU data is available, called the *physical PV source*
- Where only PWS may be available, or *hybrid PV sources*
- Where neither PWS or PMU data is available, called *virtual PV sources*, and DPV power generation needs to be estimated or forecasted in the region.

Here, the Clemson region, (<20 km radius) is chosen as a localized area over which spatial estimations or predictions of PV power is needed. Target Locations for VWS and PWS are seen on the map diagram, in Figure 5.1

Figure 5.1 Target locations for DT based PV power estimation on a map

The exact latitudes and longitudes for these chosen locations for PV power estimation are seen in Table 5.1.

Table 5.1 Target locations for DT based PV power estimation and prediction

Station no.	*Location*	*Latitude (°N)*	*Longitude (°W)*
1	(a)	34.685	-82.863
2	(b)	34.672	-82.886
3	(c)	34.678	-82.818
4	(d)	34.667	-82.87
5	(e)	34.668	-82.845
6	(f)	34.673	-82.863
7	(g)	34.669	-82.874
8	(h)	34.67	-82.85
9	(i)	34.677	-82.832
10	(j)	34.6724	-82.8804

Assumptions made for this study:

While the general approach to creating hybrid and virtual PV sources for a selected geographical region may remain, some generalizations and assumptions have been made for this study

- PV plant panel material, tilt placement, configuration and power capacity of the new PV sources are identical over the location, i.e., a 1 MW PV plant with identical configuration parameters and inverter configuration to the physical R06 site.
- The VWS parameters are tuned to the geographical site selected. In this study, the Clemson-Anderson-Pendleton region with little variation in elevation and good amount of solar irradiance received at the site is chosen.

- Data collection has taken place over spring, meaning the historical data used to train AI models and make mathematical inferences are over spring. Seasonal variations can be incorporated into the model when more amounts of data are collected over the year.

Situational Awareness of PV Power Generation using Estimation DTs

Chapter 2 presents some of the operational and management challenges in an MEPDS. A significant challenge is the growing lack of observability and difficulty in forecasting and planning in active distribution networks, for instance [73]. This means that situational awareness of real-time behavior of various DERs in the system is often difficult to monitor, especially due to lack of measurement data. PV Plant DTs that obtain real-time data from physical and virtual weather stations are fed into PV DT estimation models to facilitate both real-time monitoring and DSO operator situational awareness of PV sources in a system. Data collection over a day can further allow for PV integration into distribution system studies or time series analysis to be performed in the system.

Multiple scenarios may be applicable while trying to obtain realistic PV plant data over a geographical area, for both distribution system studies and for DSO operators as seen in Figure 5.2.

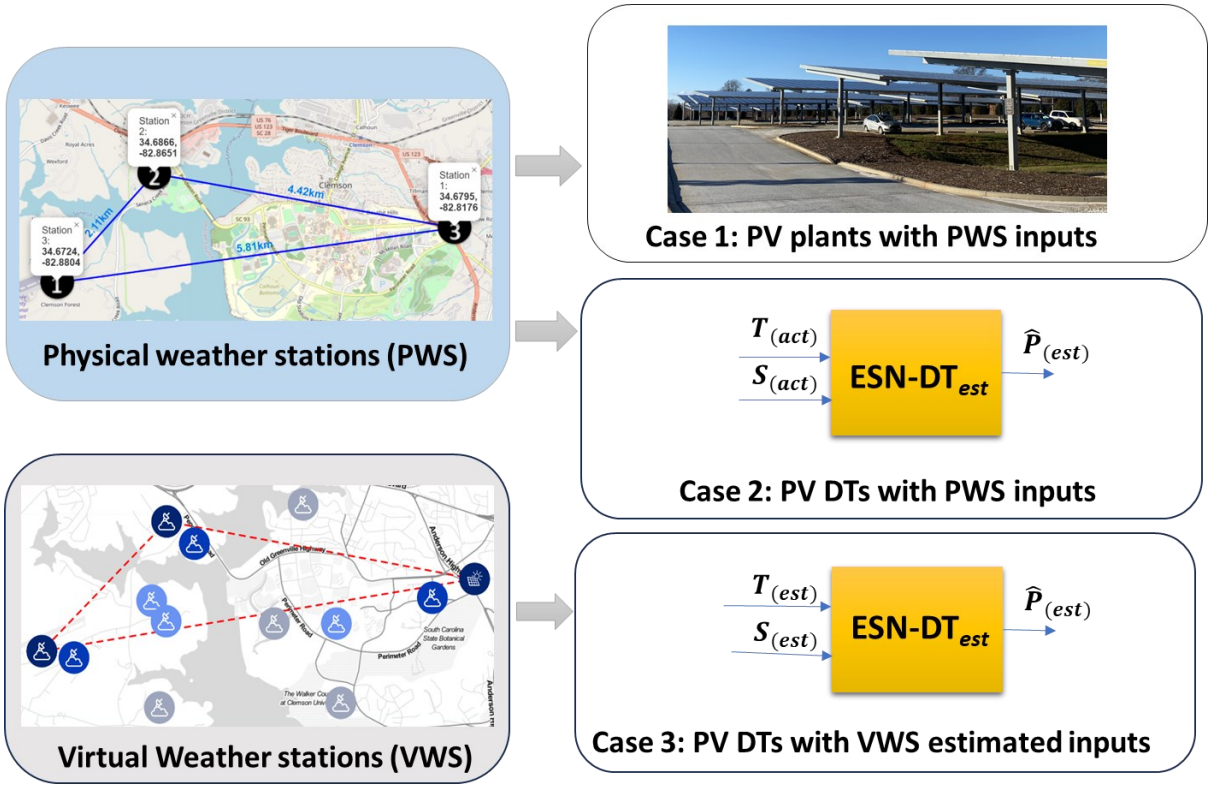


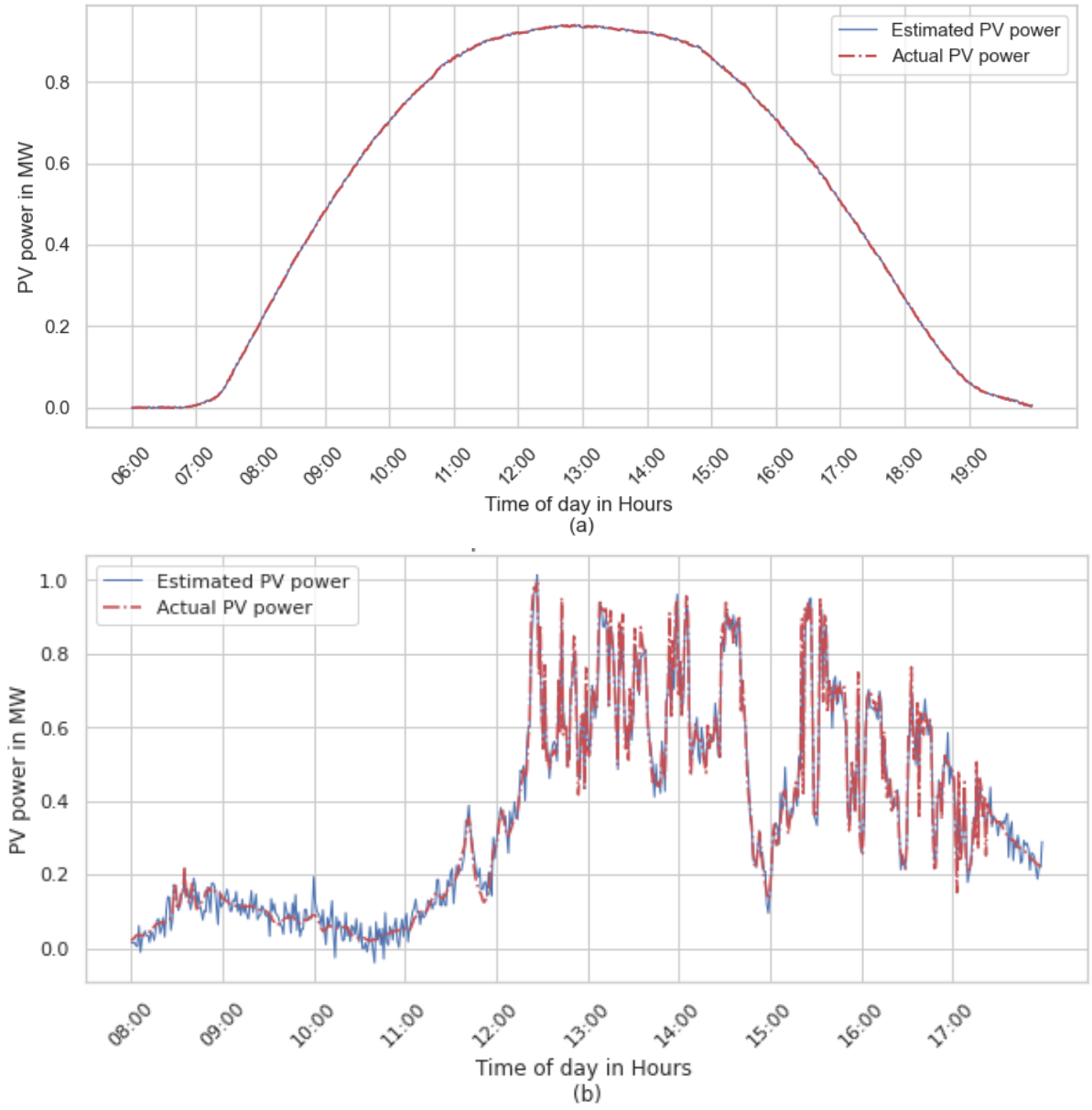
Figure 5. 2 Scenarios for DPV data generation using PV plants, hybrid PV plants and virtual PV plants

Case 1: Physical weather station data streams feeding into existing DPVs or PV locations with measurement data

This case study is used to validate the proposed DT based approach. Here, measurements are present for both PWS streams as well as microPMUs or PMU data from DPVs. The developed PV DTs to estimate DER behavior at these locations can be used to validate and optimize the PV DT performance over long term operation, such as test seasonality, granular process modeling, as well as test new AI, hybrid-AI or physics-based models to better simulate DER behavior in real-time. Additionally, once fully tested, these PV DTs can be used similarly to traditional DTs and utilized for prognostics, condition monitoring and fault detection and faster-than-real-time operation analysis.

The PV DT for estimation of PV power at R06, developed in Chapter 2, is run alongside the physical system and is fed with temperature and solar irradiance from the PWS

at R06 in real-time. The results of PV DT power estimations with real PV plant data, here, 1 MW PV plant at R06 for various day categories is seen in Figure 5.3 and Table 5.2.



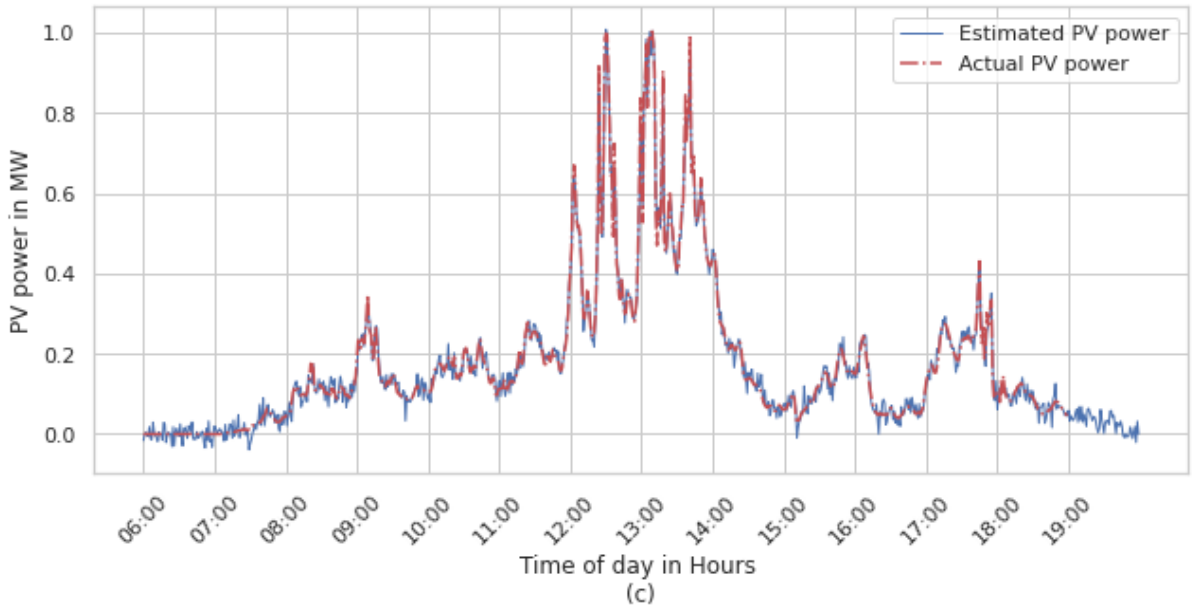


Figure 5. 3 Case 1, Estimated PV power at R06 using PWS for a sunny (a), cloudy (b) and moderately cloudy (c) day

Table 5.2 DT performance for PV power estimation

Day Category	*Dataset type*	*Neurons*	*MAE*	*MSE*	*RMSE*	*MAPE*
Sunny	Training	10	0.0002	0.0003	0.0173	0.0049
Moderately Cloudy	Training	20	0.0087	0.009	0.0949	0.0089
Cloudy	Training	20	0.0092	0.0098	0.0990	0.0135
Sunny	Testing	10	0.0003	0.0002	0.0141	0.0064
Moderately Cloudy	Testing	20	0.0075	0.0087	0.0933	0.0112
Cloudy	Testing	20	0.0090	0.0001	0.0097	0.0142

Case 2: Physical weather station data streams feeding into unknown PV locations or DPVs without measurement data

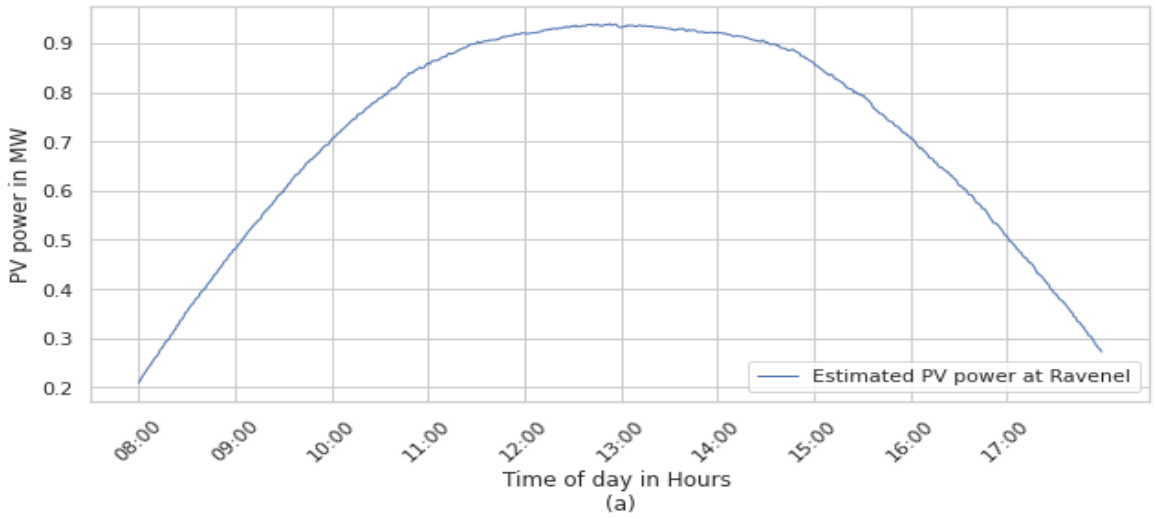
There is often a need to estimate PV power and its impact over the distribution system for locations where micro-PMU, separate meter readings or other measurement data is unavailable or where a PV plant hasn't been setup yet. In this scenario, PWS may be present at the site, or other weather models may be available for estimation of solar irradiance and temperature at the site. This occurs in the case of planning new PV installations or for generation of time-series data for distribution system studies. In such scenarios, realistic spatial estimations of weather data over a given area is necessary to model PV behavior.

To support this scenario, the two available PWS, "Ravenel" and "Airport", locations, seen in Table 5.3, are used to generate spatial estimations of temperature and solar irradiance over the PWS locations, and are used as inputs to estimate DPV performance in real-time. DPV streams obtained through this method can be easily integrated into a distribution testbed or used for advanced monitoring and analysis of PV system behavior and patterns over the region.

Table 5.3 PWS locations

Station no.	*Latitude (°N)*	*Longitude (°W)*
Ravenel	34.69	-82.87
Airport	34.68	-82.82

Estimated PV power over the region for PWS locations at Ravenel and the Airport can be seen in Figures 5.4-5.6



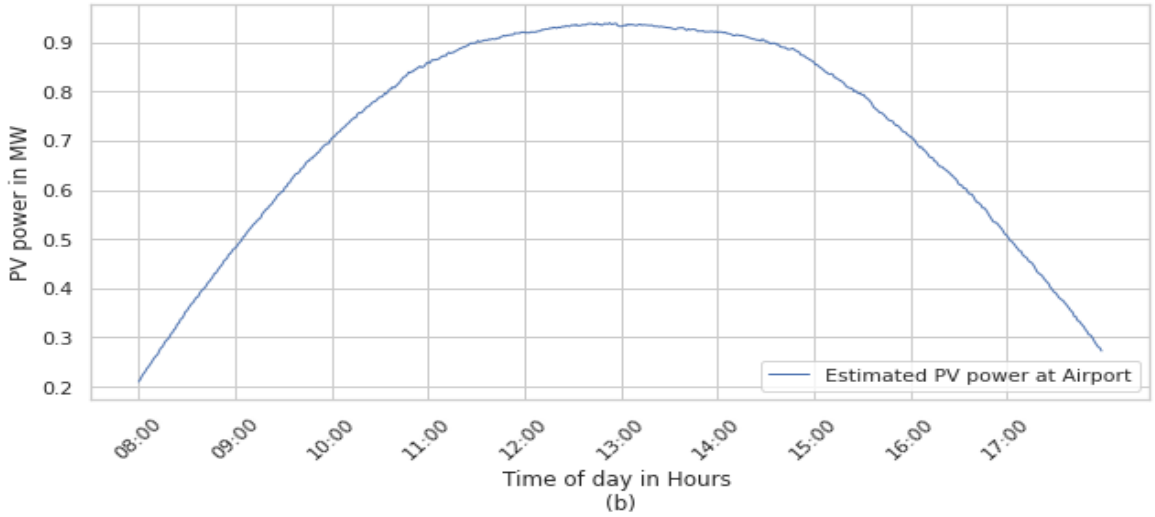


Figure 5.4 Case 2, Estimated PV power using PWS at R06 for Ravenel (a), Airport (b) for a sunny day

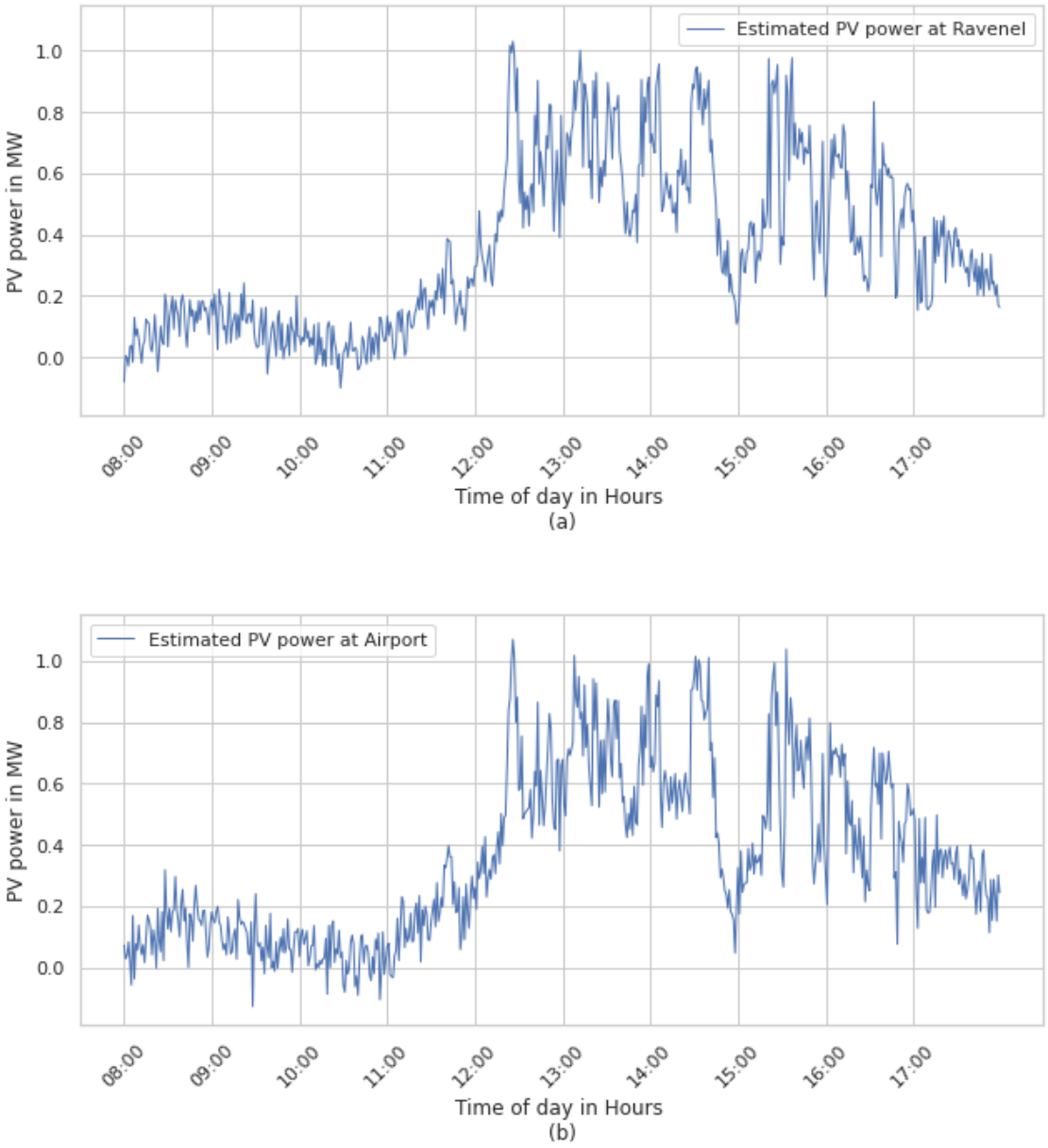


Figure 5.5 Case 2, Estimated PV power using PWS at R06 for Ravenel (a), Airport (b) for a moderately cloudy day

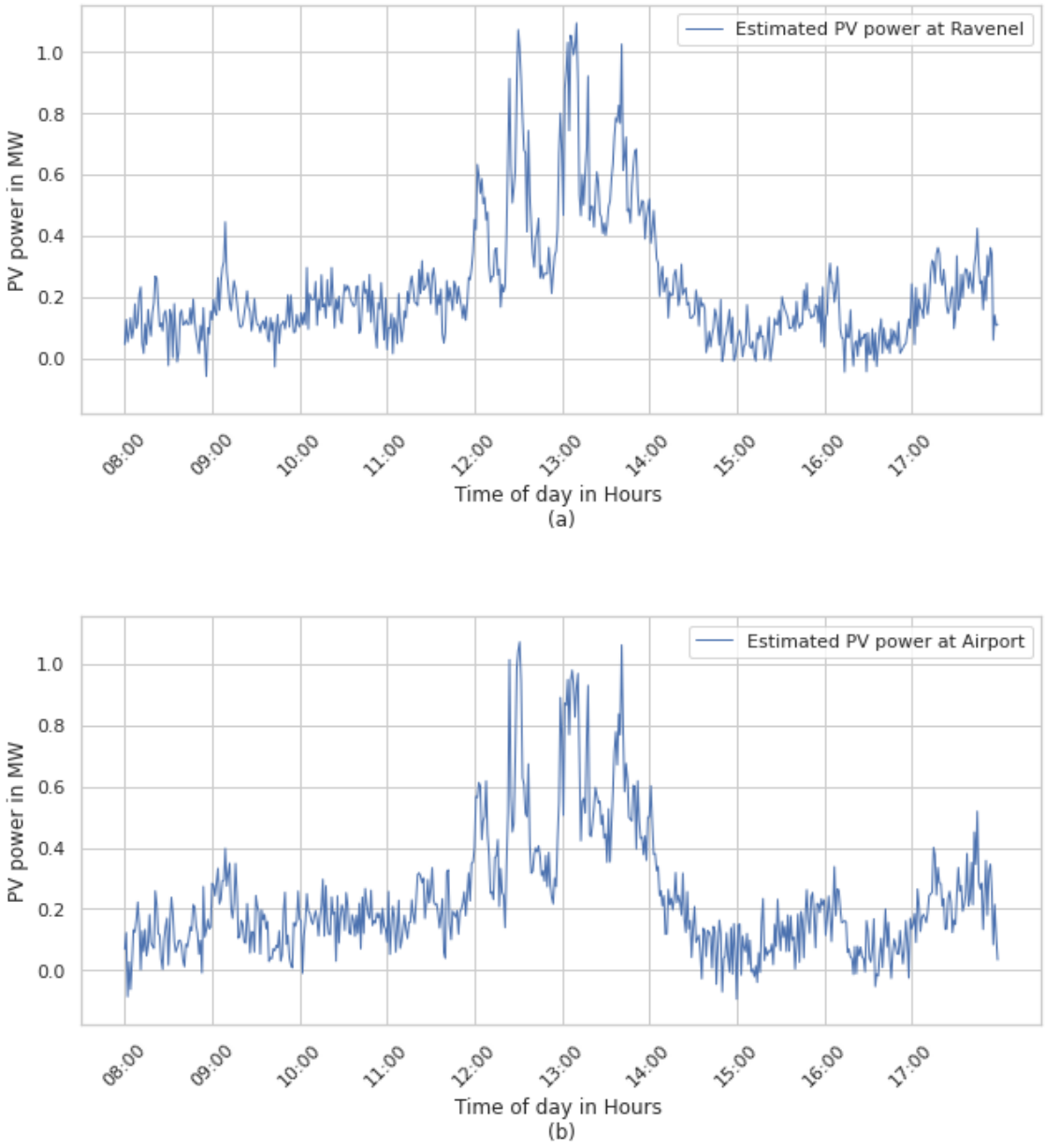


Figure 5.6 Case 2, Estimated PV power using PWS at R06 for Ravenel (a), Airport (b) for a cloudy day

Case 3: Virtual weather station data streams feeding into unknown DPV locations or DPVs without measurement data

Distribution system studies such as in planning studies for hosting capacity, network congestion and the like requires realistic time-series data and estimation of DER behavior in various placement locations. In this scenario both measurement data and PV plant

installations or micro-PMU, separate meter readings or other measurement data is not present in the selected location.

To support this scenario, VWS developed in Chapter 4 can be used to generate spatial estimations of temperature and solar irradiance over the location, and are used as inputs to PV DTs to estimate potential DPV installation behavior in a particular area, and estimate overall PV production in the area in real-time. DPV streams obtained through this method can be further integrated into a distribution testbed for quasi-static time-series analysis to establish benchmarks, perform hosting capacity studies, feeder voltage profile analysis, create generation profiles and so on, in various PV placements and configurations in the area.

Spatial PV power estimations are generated using VWS and PV DTs. Data streams generated over various day categories are seen in Figures 5.7-5.22

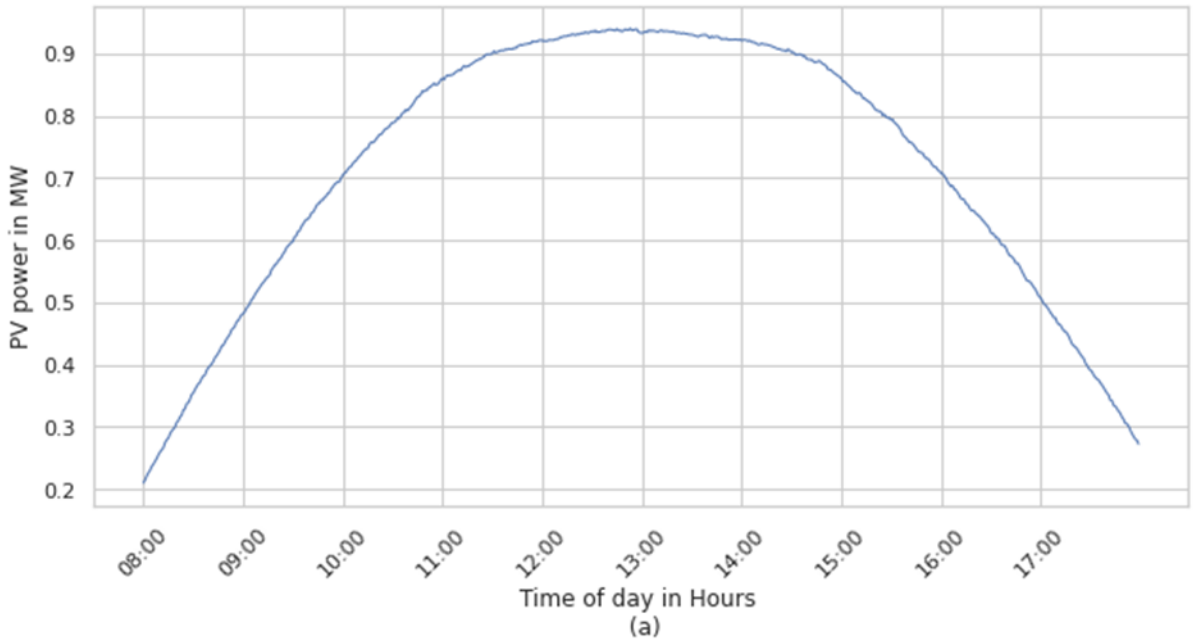
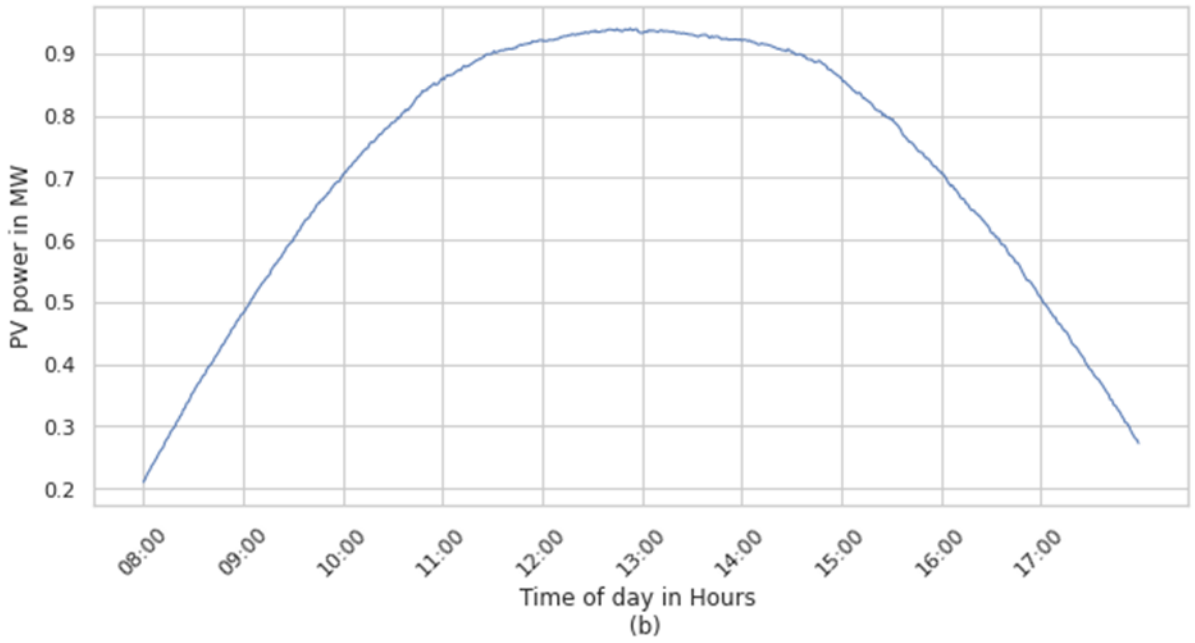


Figure 5.7 Case 3, Estimated PV power using VWS for locations (a) for a sunny day



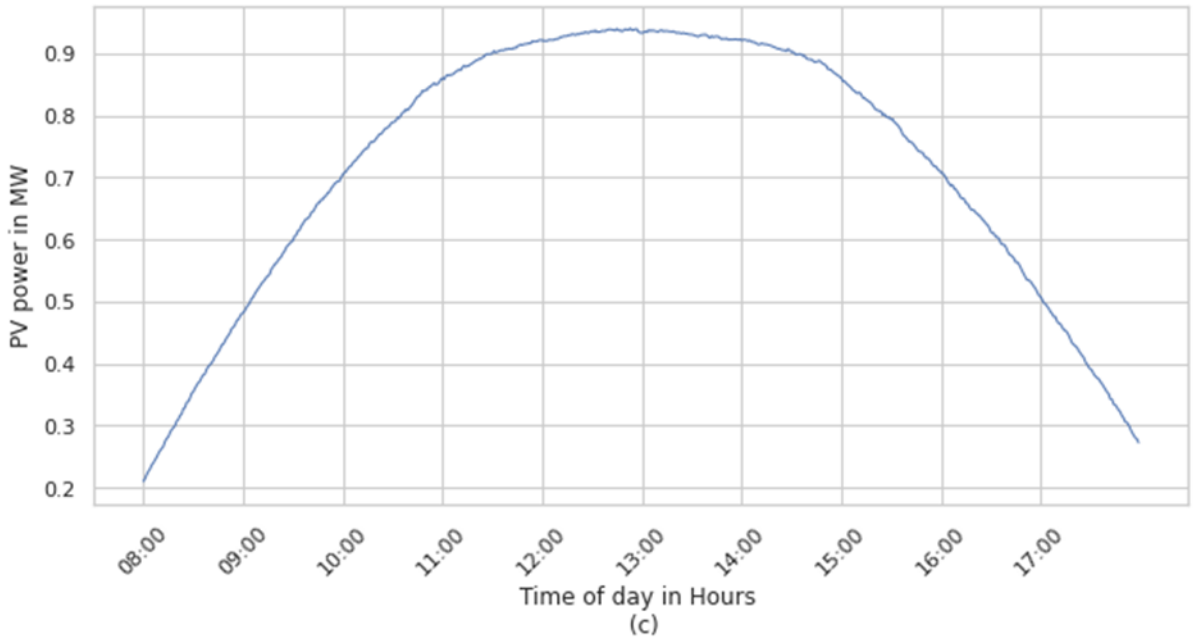
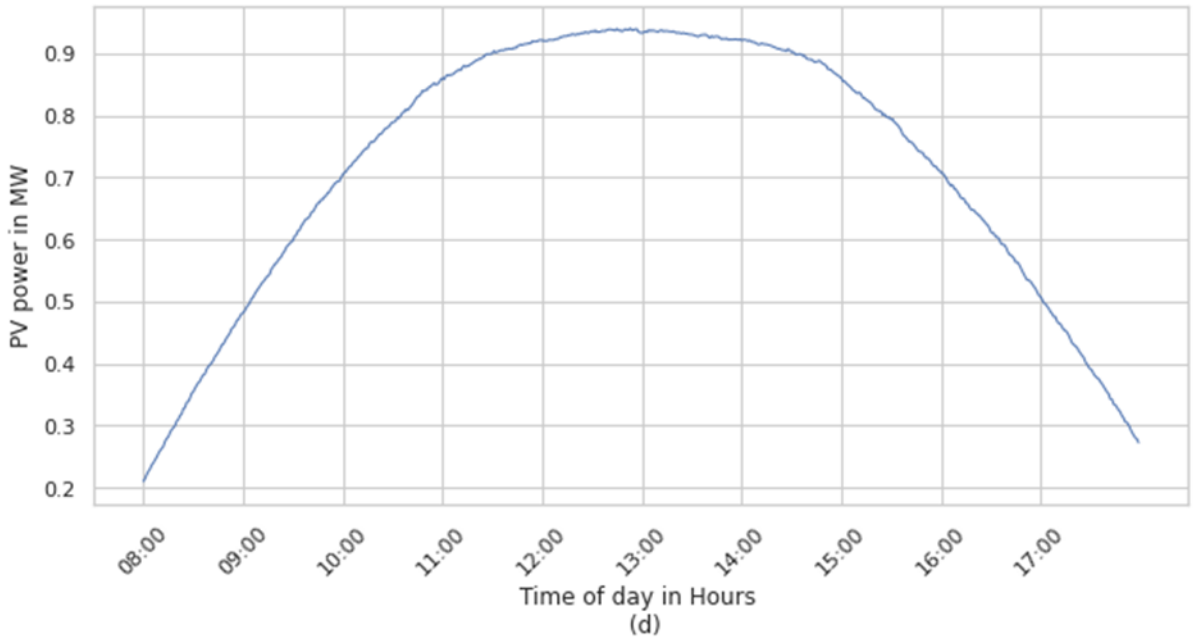


Figure 5.8 Case 3, Estimated PV power using VWS for locations (b)-(c) for a sunny day



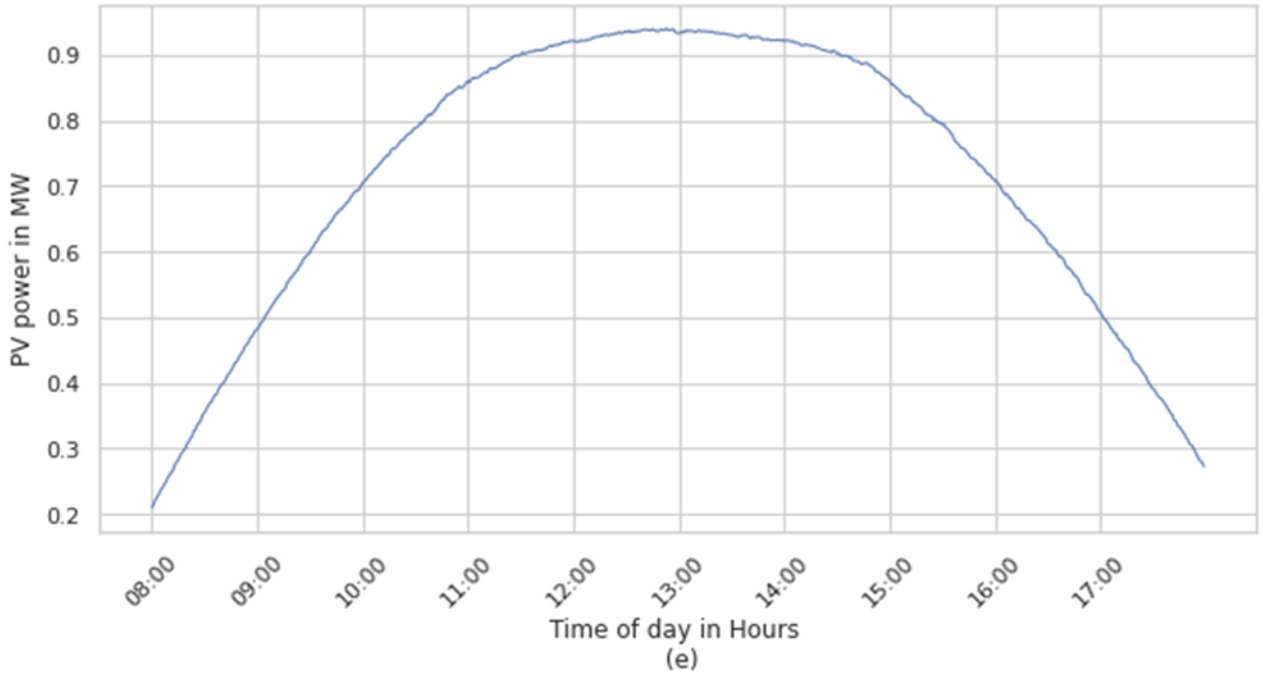
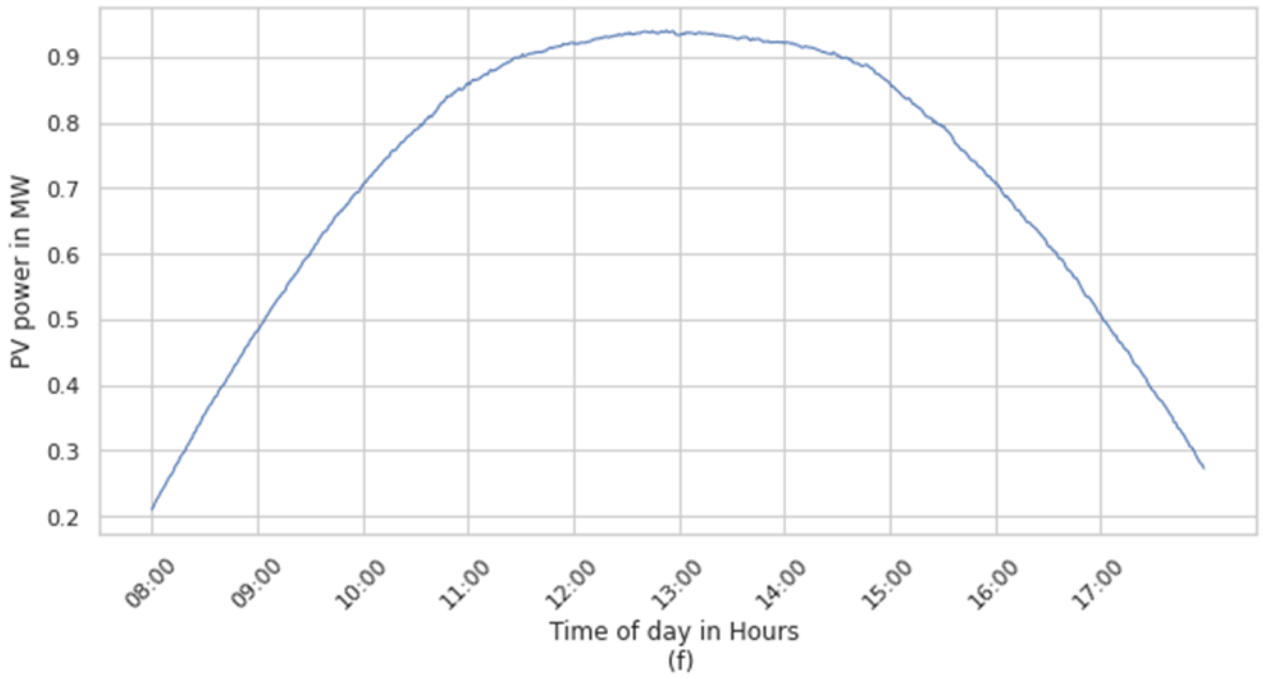


Figure 5.9 Case 3, Estimated PV power using VWS for locations (d)-(e) for a sunny day



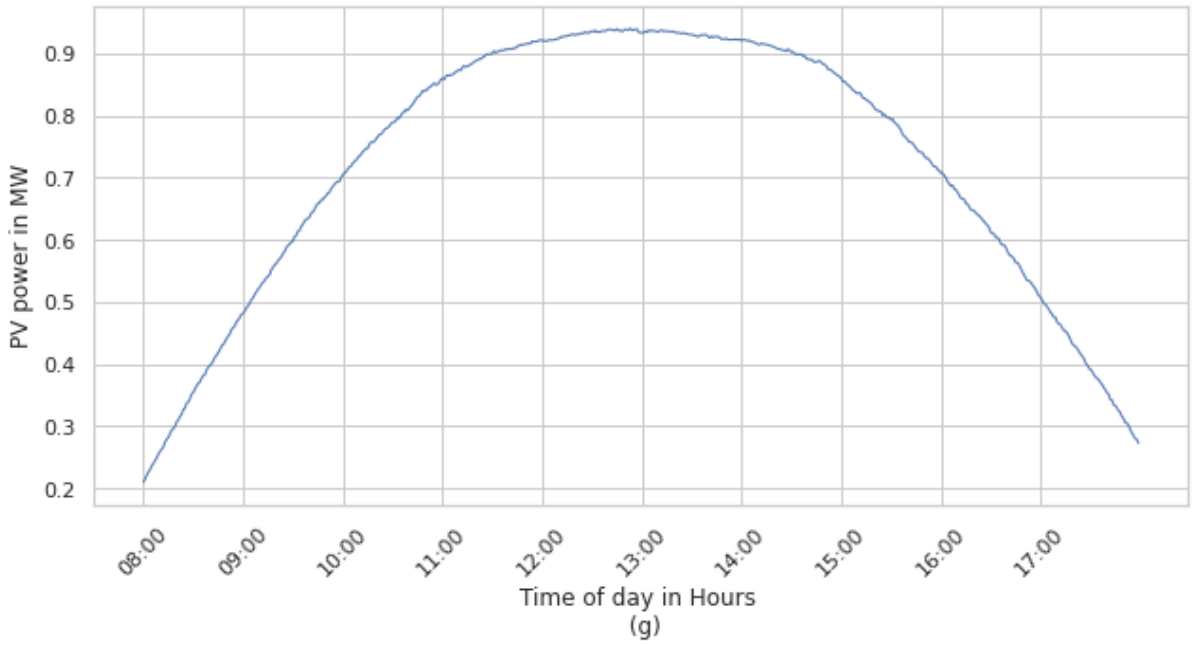
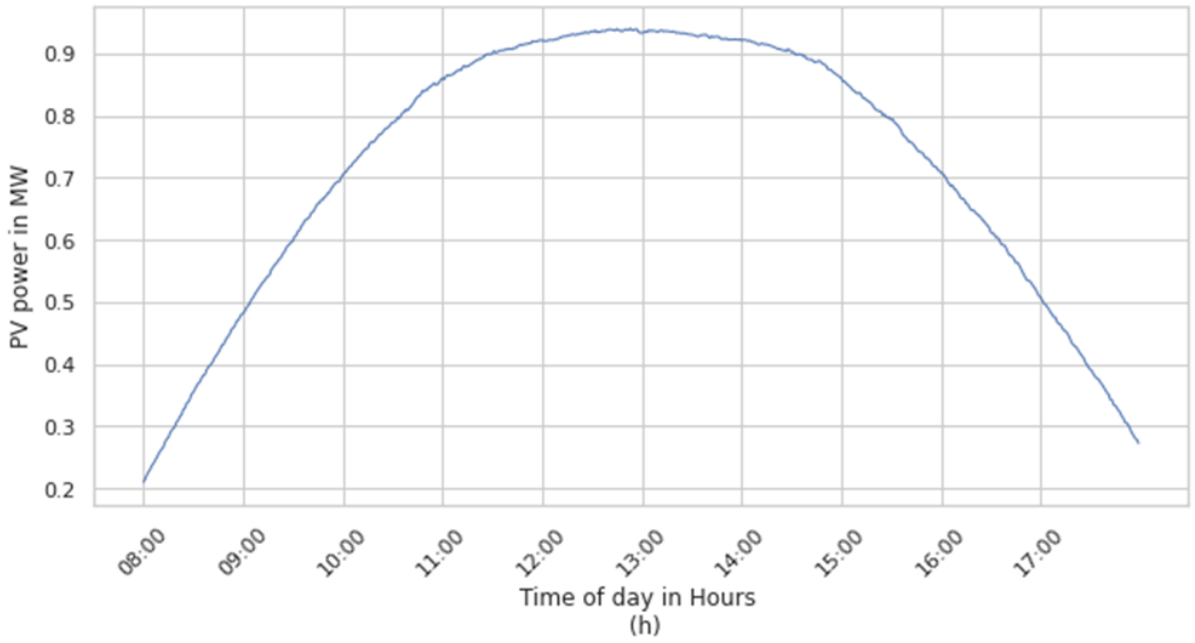


Figure 5.10 Case 3, Estimated PV power using VWS for locations (f)-(g) for a sunny day



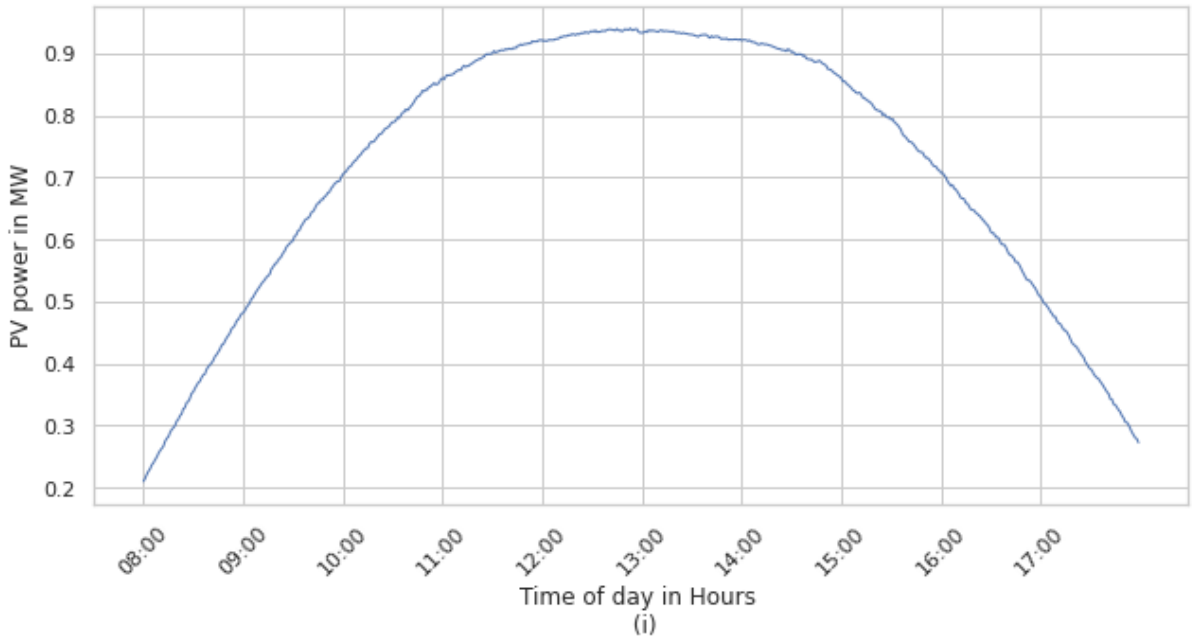


Figure 5.11 Case 3, Estimated PV power using VWS for locations (h)-(i) for a sunny day

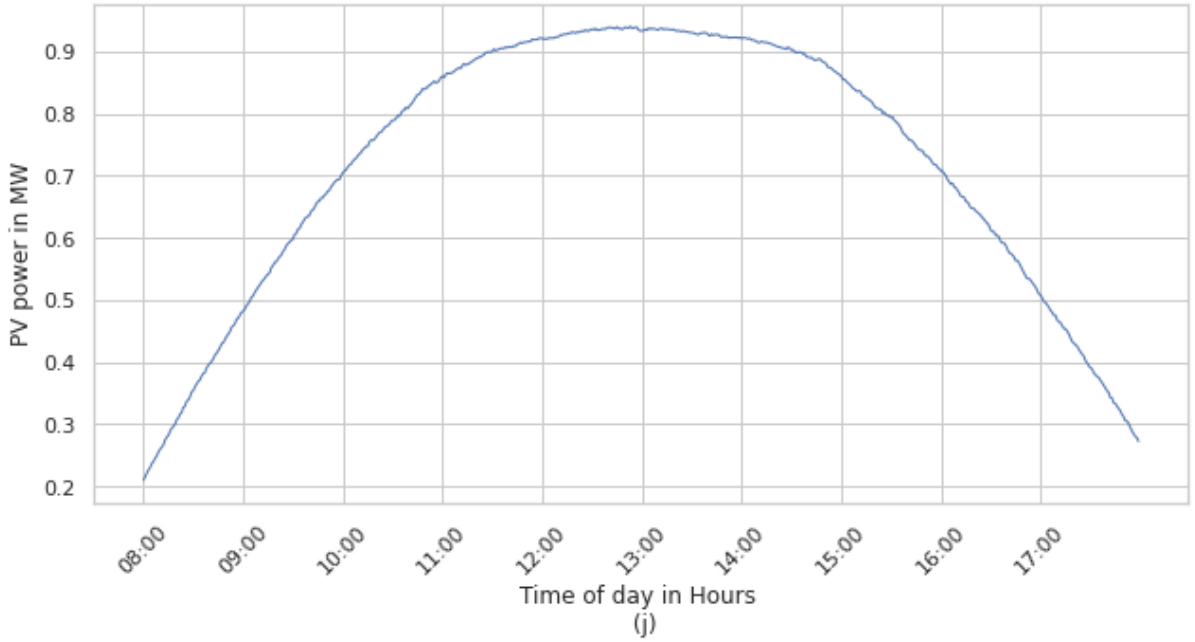
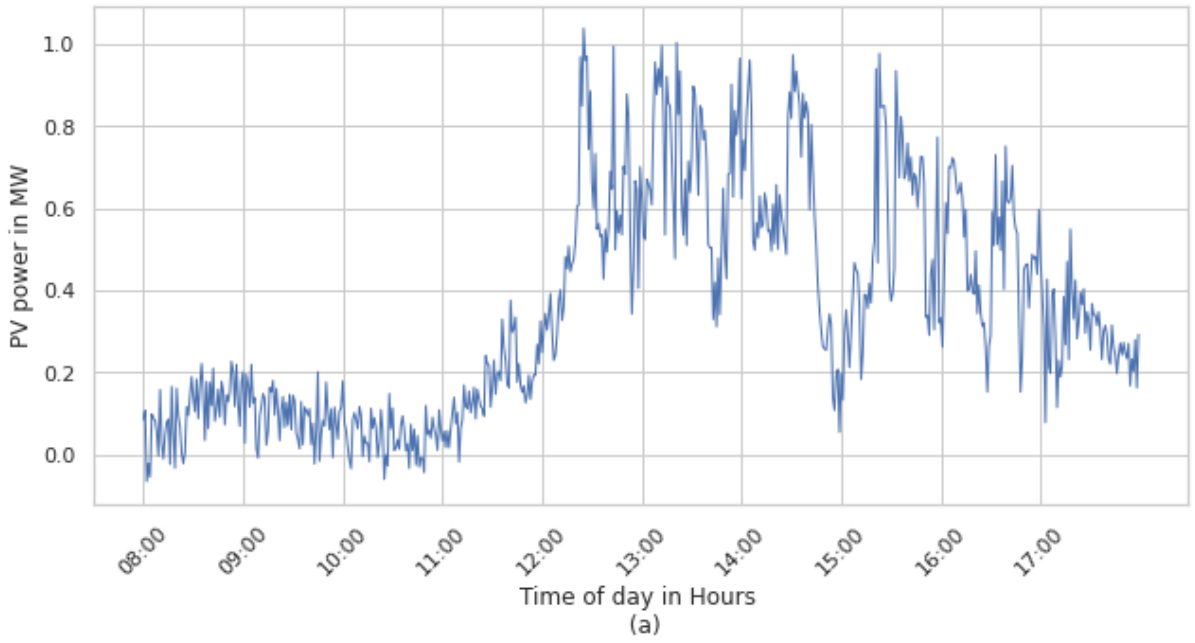


Figure 5.12 Case 3, Estimated PV power using VWS for locations (a)-(j) for a sunny day



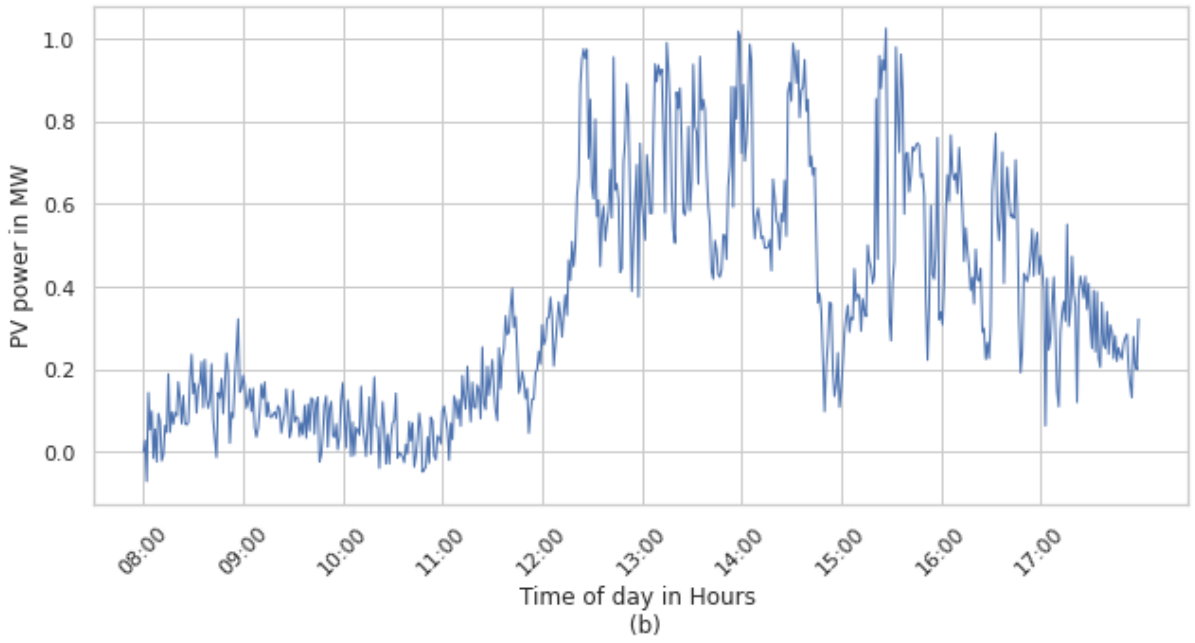
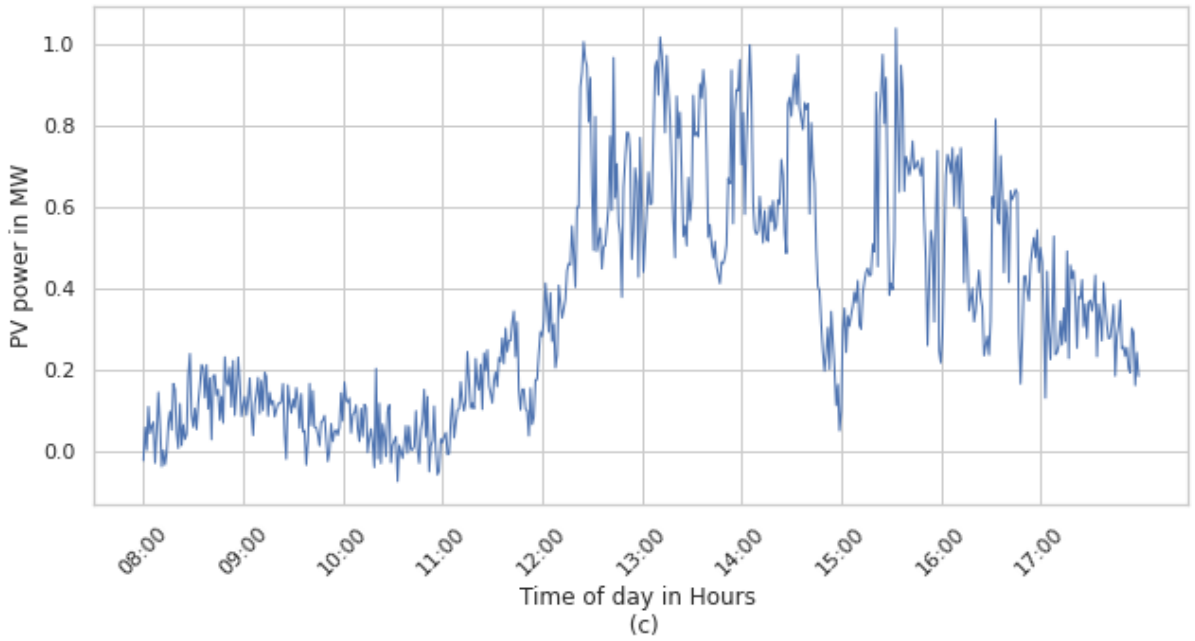


Figure 5.13 Case 3, Estimated PV power using VWS for locations (a)-(b) for a moderately cloudy day



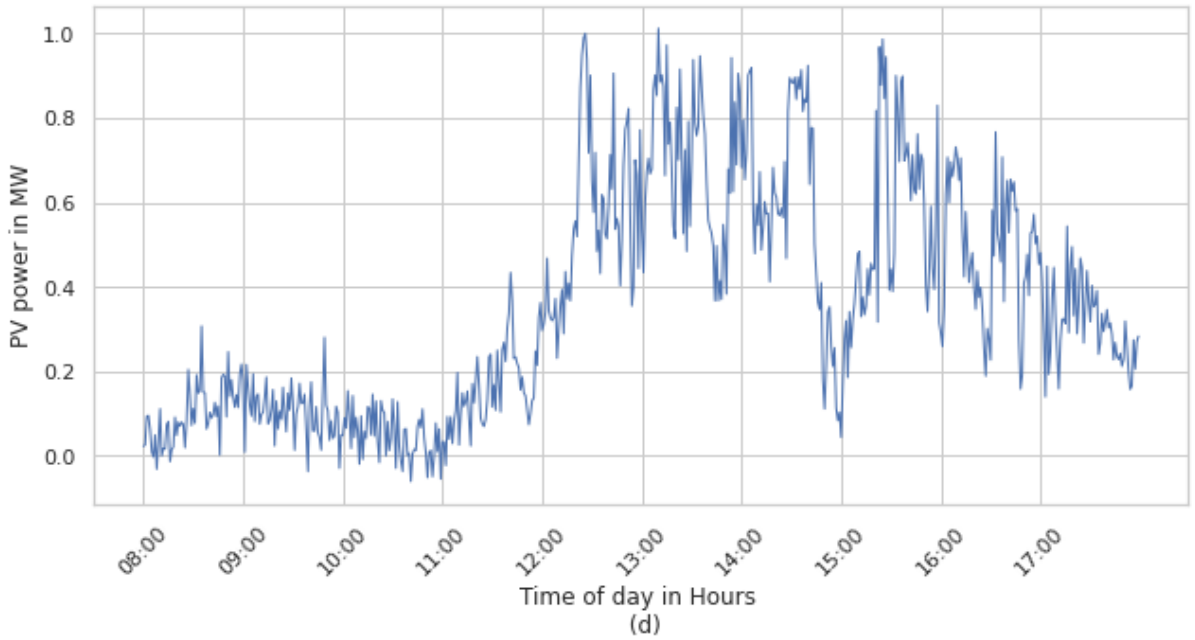
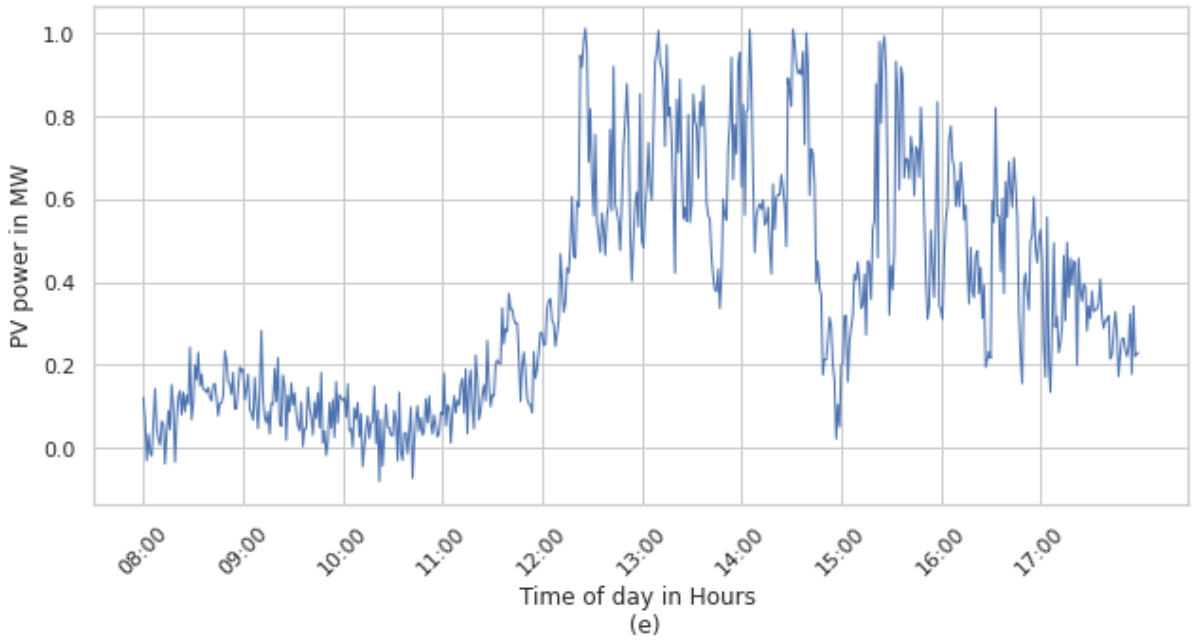


Figure 5.14 Case 3, Estimated PV power using VWS for locations (c)-(d) for a moderately cloudy day



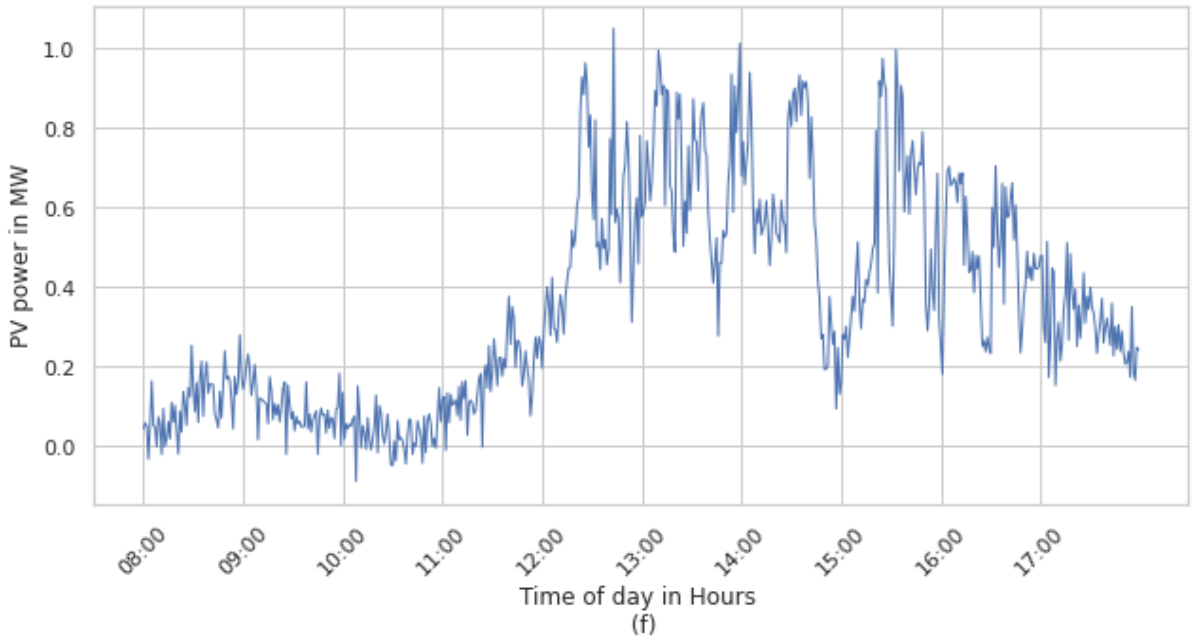
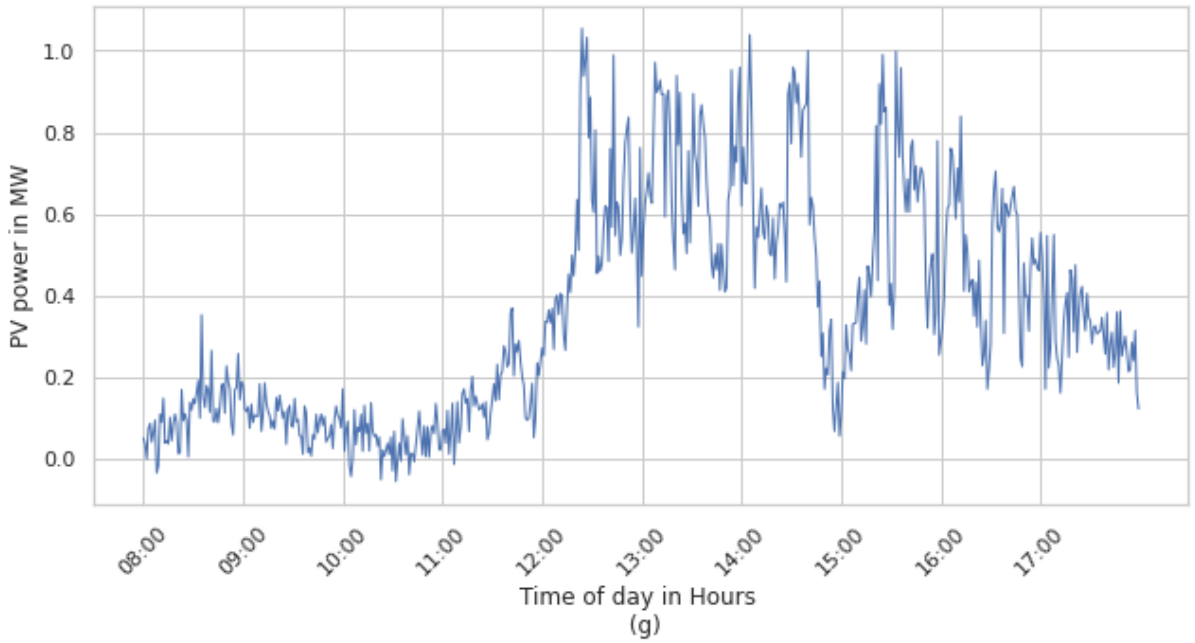


Figure 5.15 Case 3, Estimated PV power using VWS for locations (e)-(f) for a moderately cloudy day



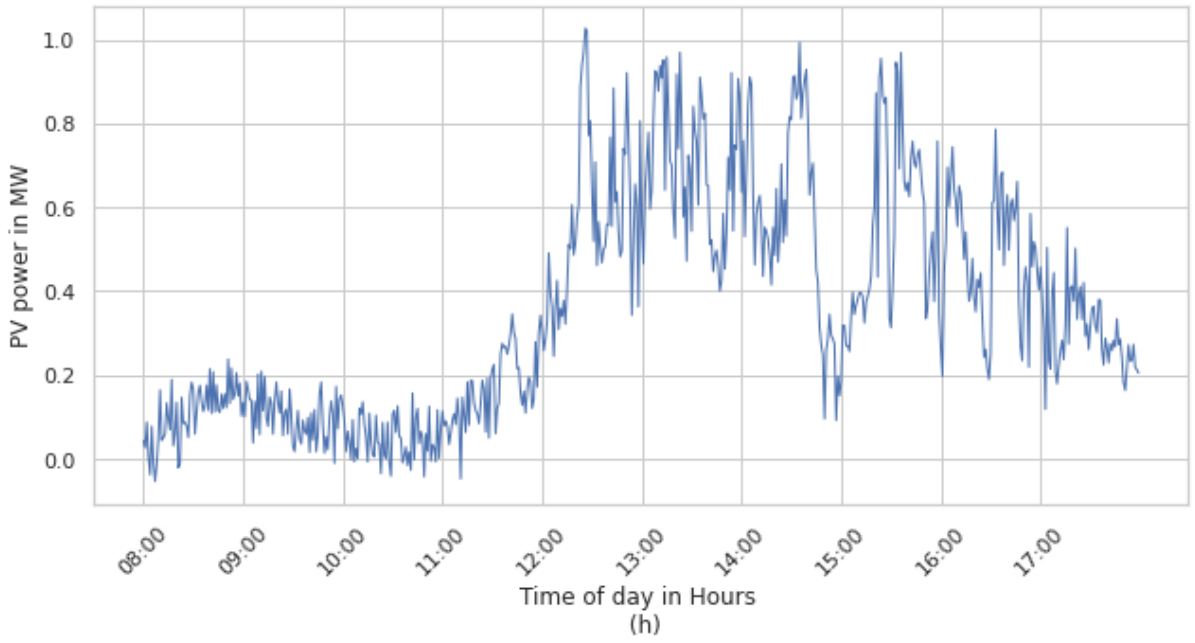
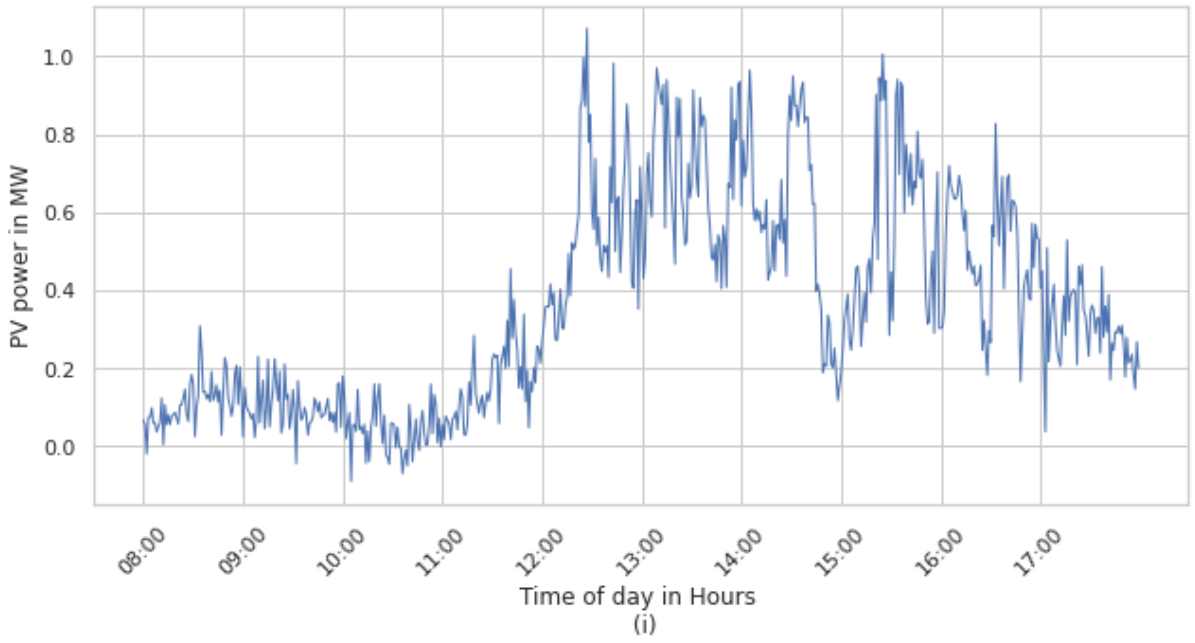


Figure 5.16 Case 3, Estimated PV power using VWS for locations (g)-(h) for a moderately cloudy day



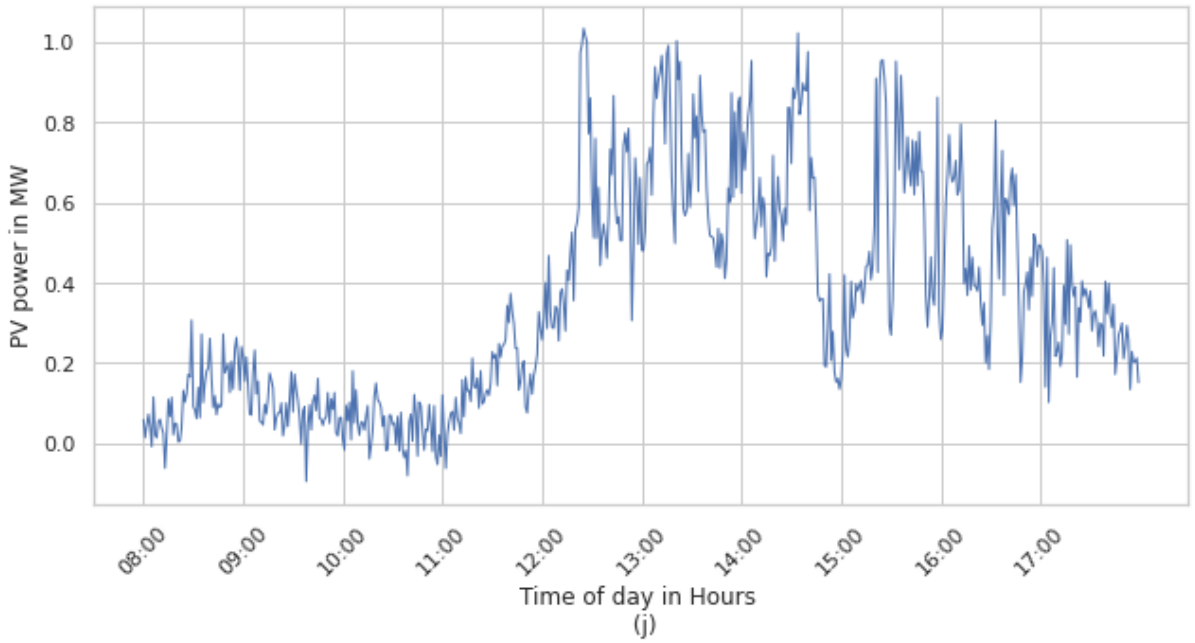


Figure 5.17 Case 3, Estimated PV power using VWS for locations (i)-(j) for a moderately cloudy day

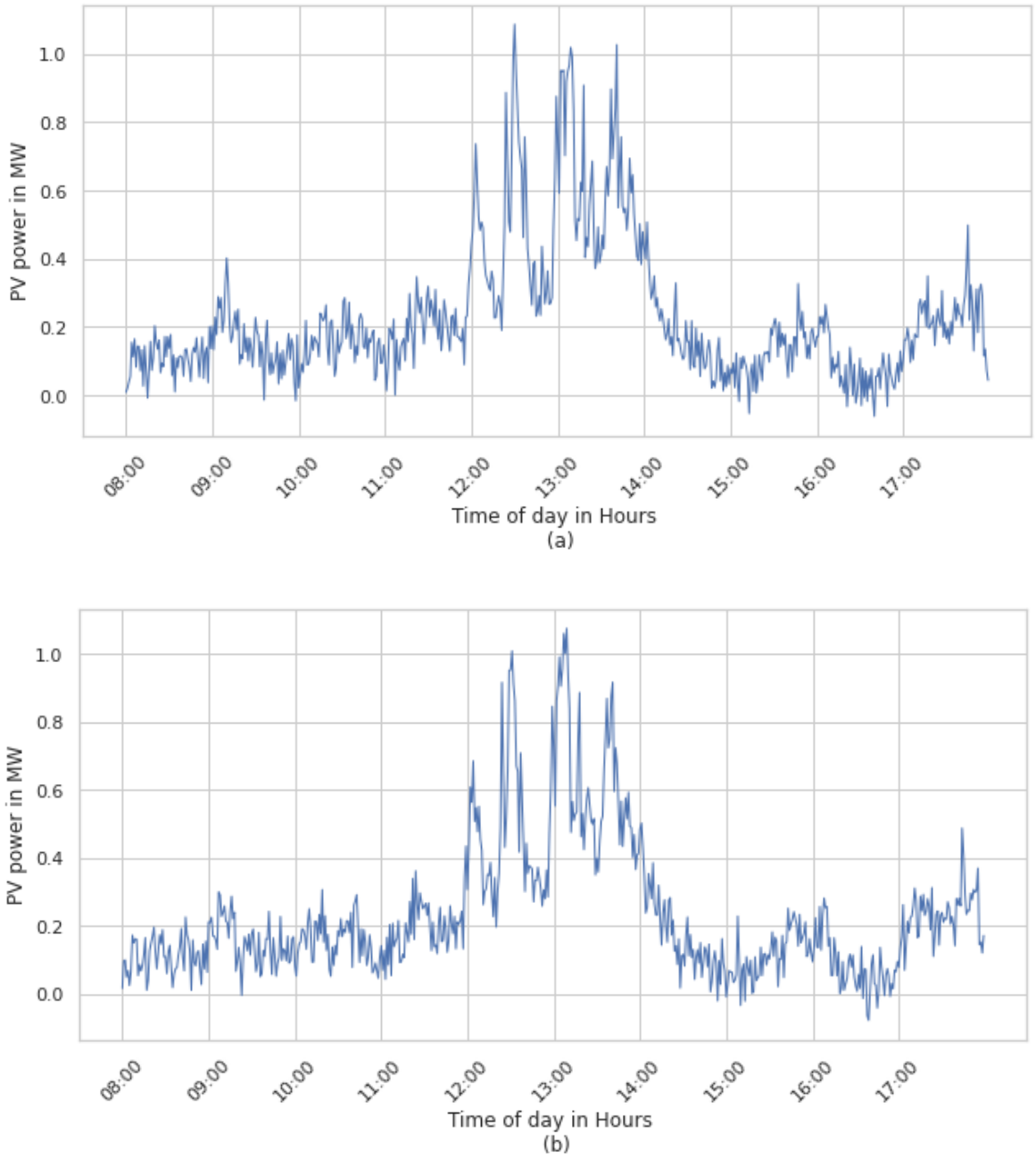
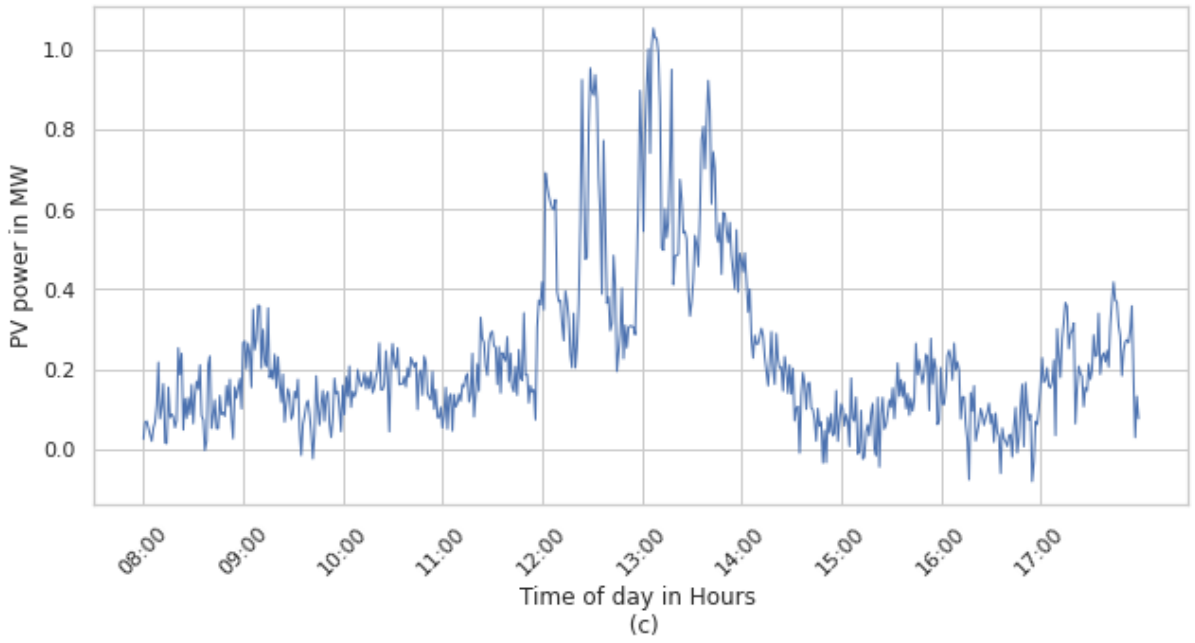


Figure 5.18 Case 3, Estimated PV power using VWS for locations (a)-(b) for a cloudy day



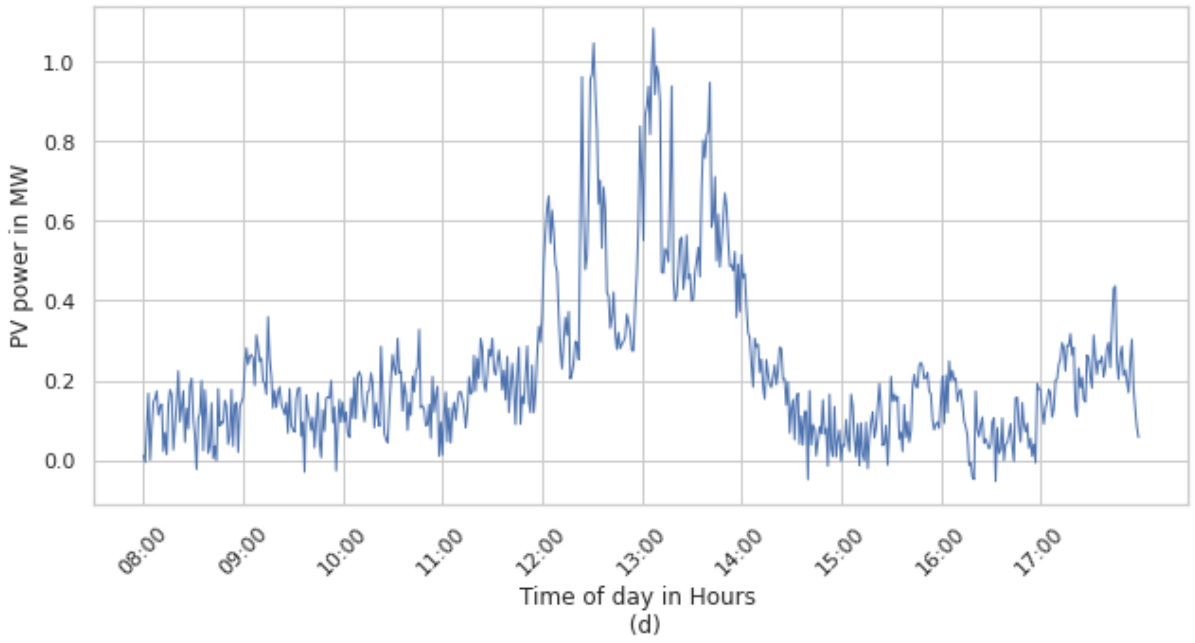
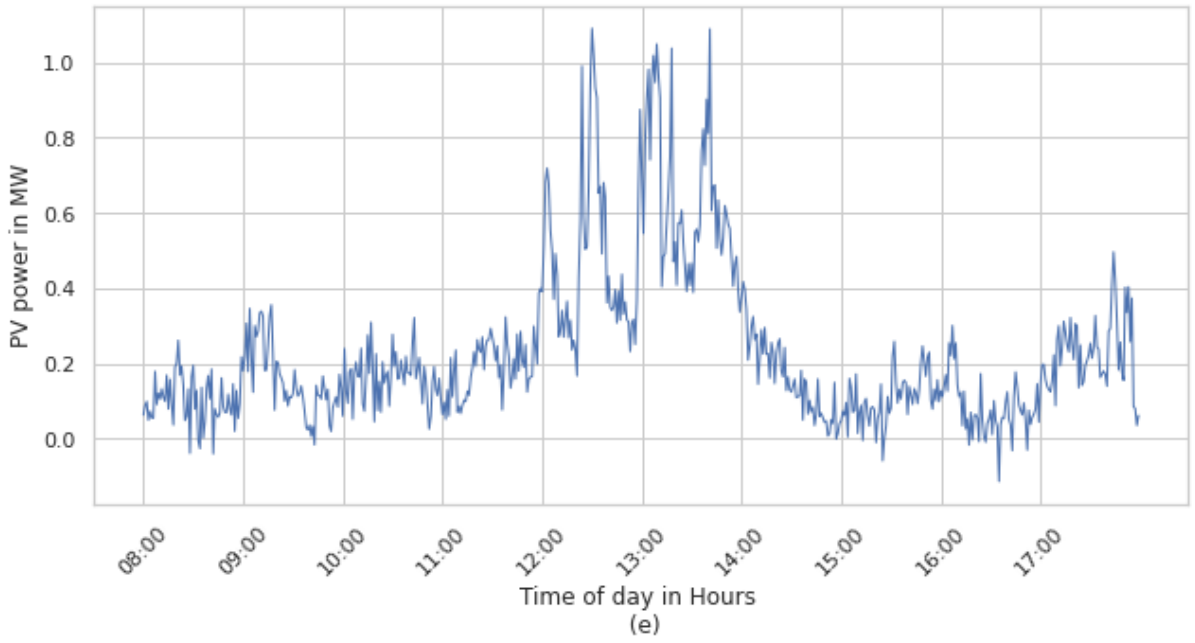


Figure 5.19 Case 3, Estimated PV power using VWS for locations (c)-(d) for a cloudy day



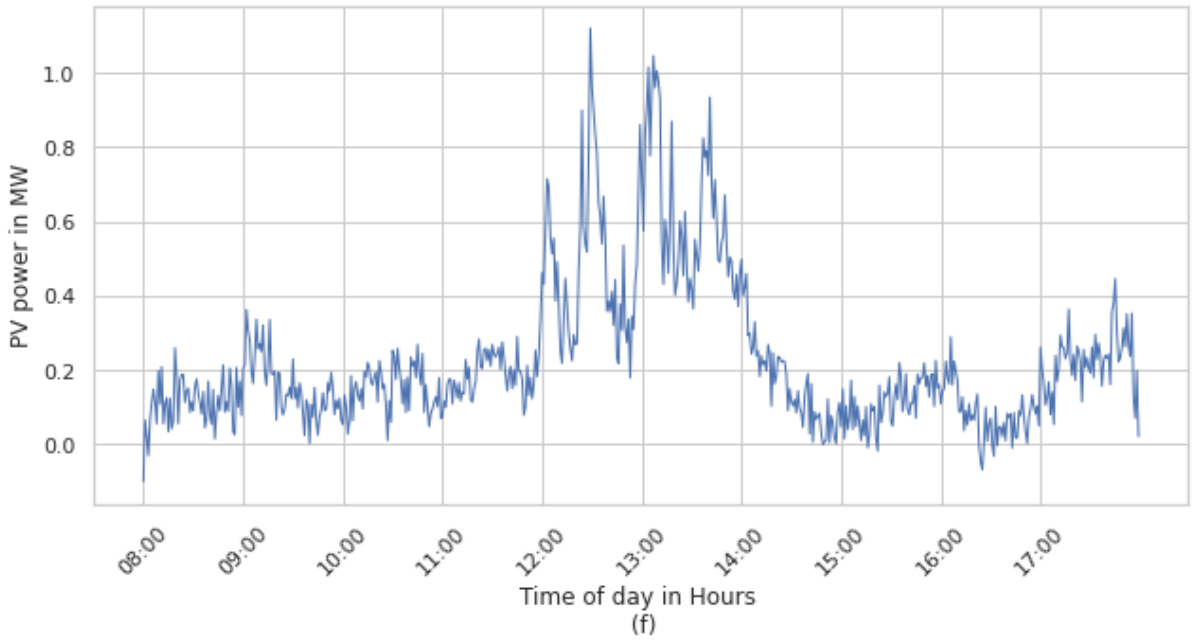
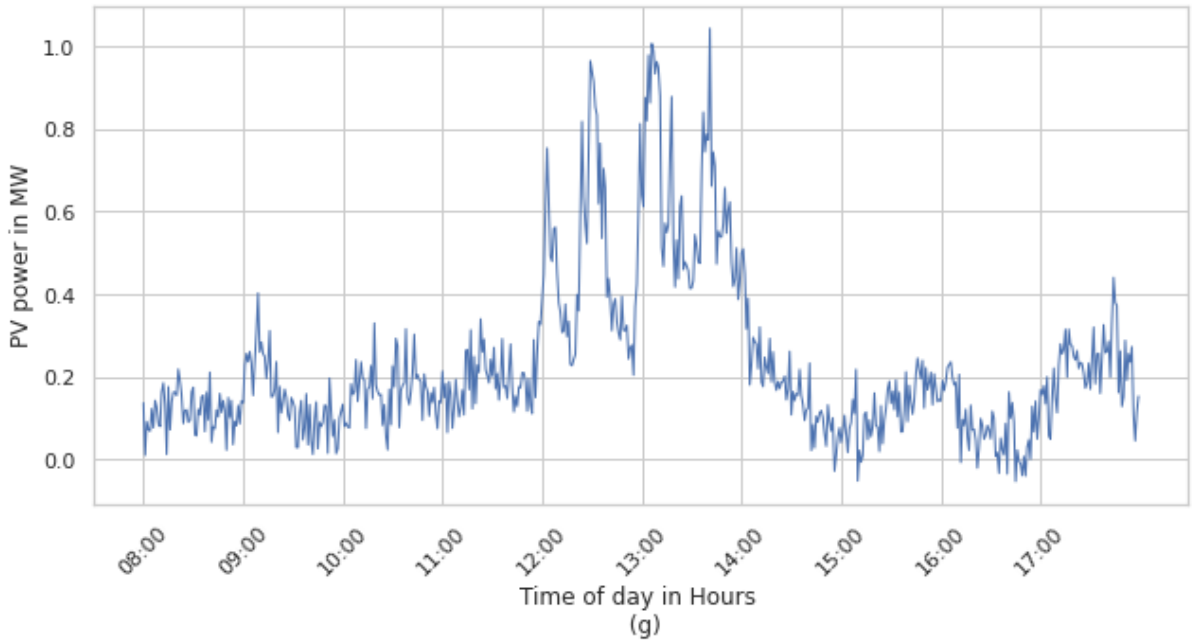


Figure 5.20 Case 3, Estimated PV power using VWS for locations (e)-(f) for a cloudy day



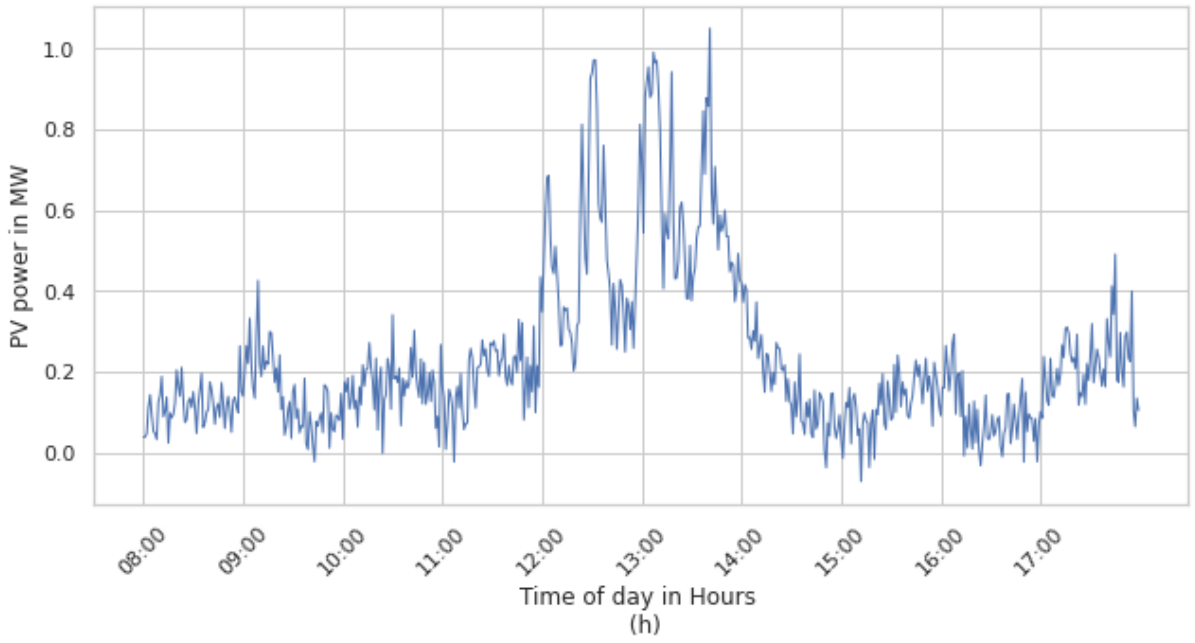
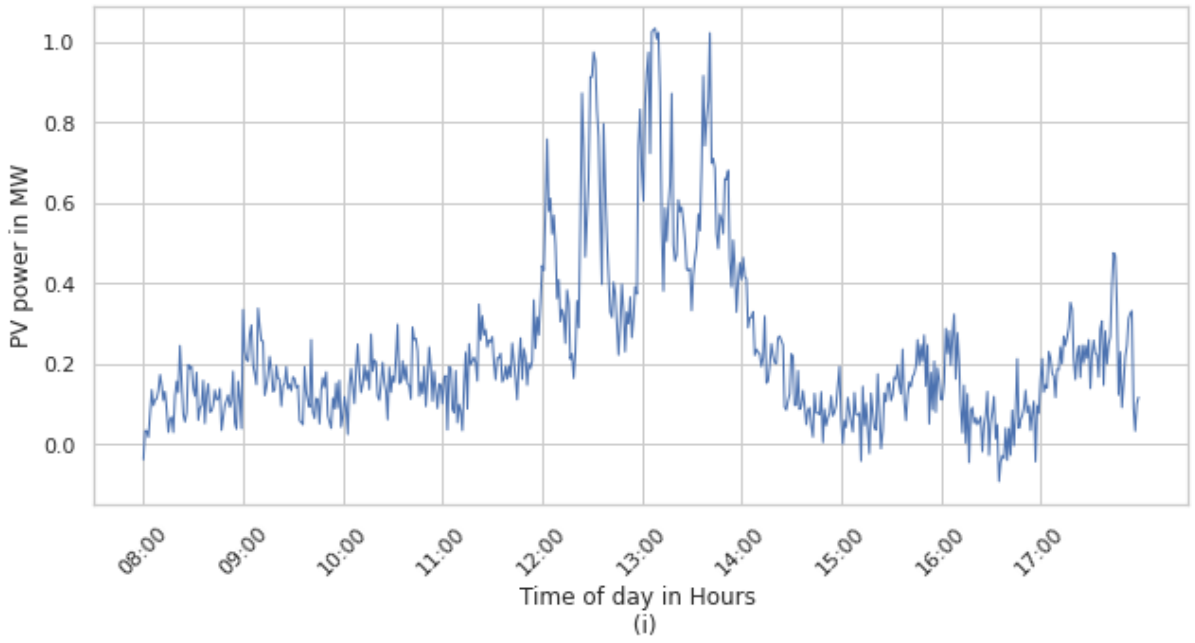


Figure 5.21 Case 3, Estimated PV power using VWS for locations (g)-(h) for a cloudy day



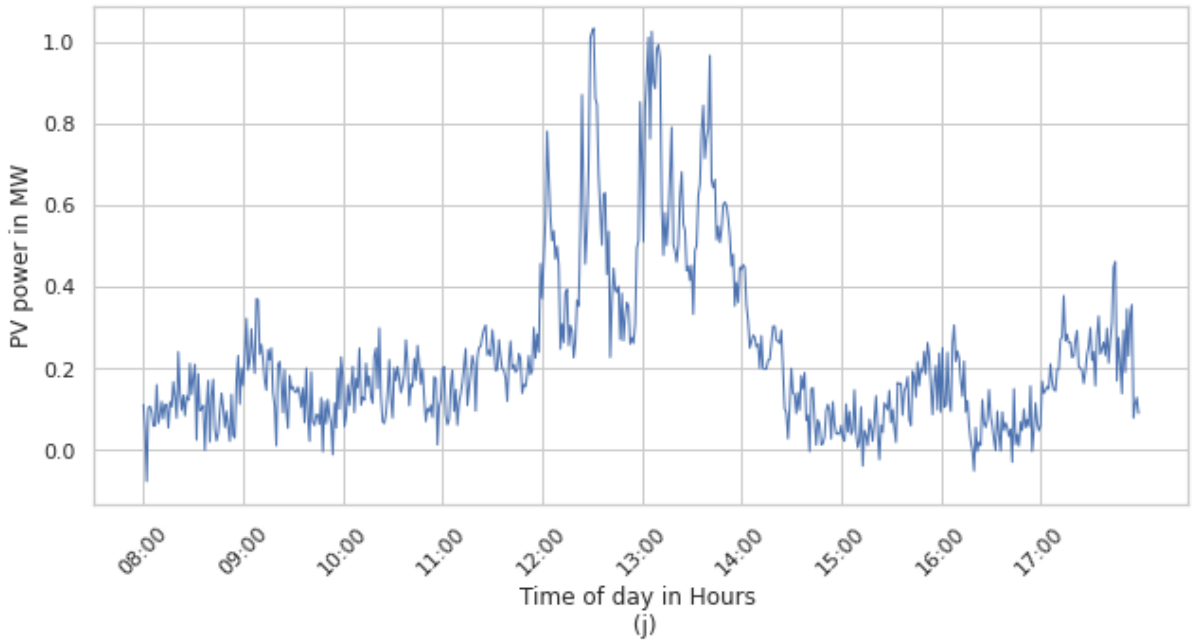


Figure 5.22 Case 3, Estimated PV power using VWS for locations (h)-(i) for a cloudy day

Situational Intelligence of PV Power Generation using Prediction DTs

Operator situational awareness can be established through behavioral modeling of DPVs in various data availability scenarios. However, situational intelligence, or predictive analysis is often needed in the system, especially for short term PV production generation

forecast, feeder voltage analysis, scenario analysis or DSO decision-making wherein multiple DT models can be run in parallel for predictive analysis and volt-var algorithm optimization or control, among other use cases. Here, short term forecasts of PV power production support DERMS applications and other distribution system analyses involved in resiliency, reliability and planning such as hosting capacity studies, voltage profile analyses, distribution system behavior in peak load and light generation, light load and peak generation, and other long-term or time-series analysis performed on DER/ DPV sources integrated into the grid.

The overall workflow for PV ESN predictions at the region using DTs can be seen in Figure 5.10, for a geographical location seen in Figure 5.11.

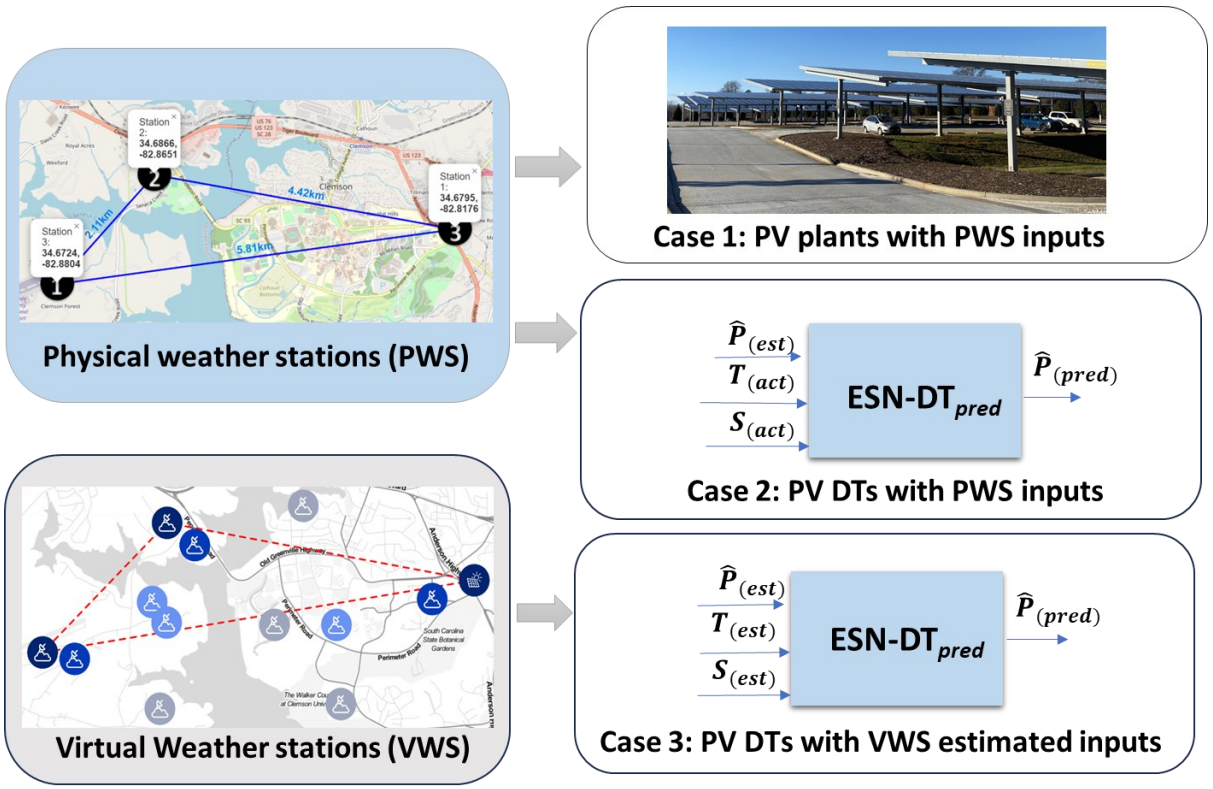
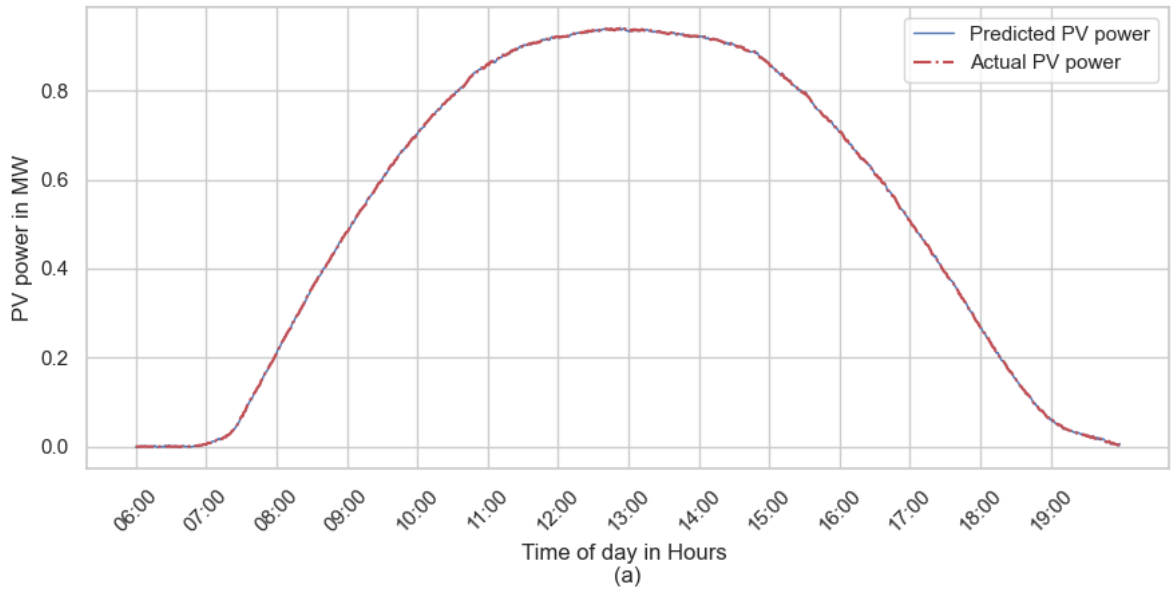


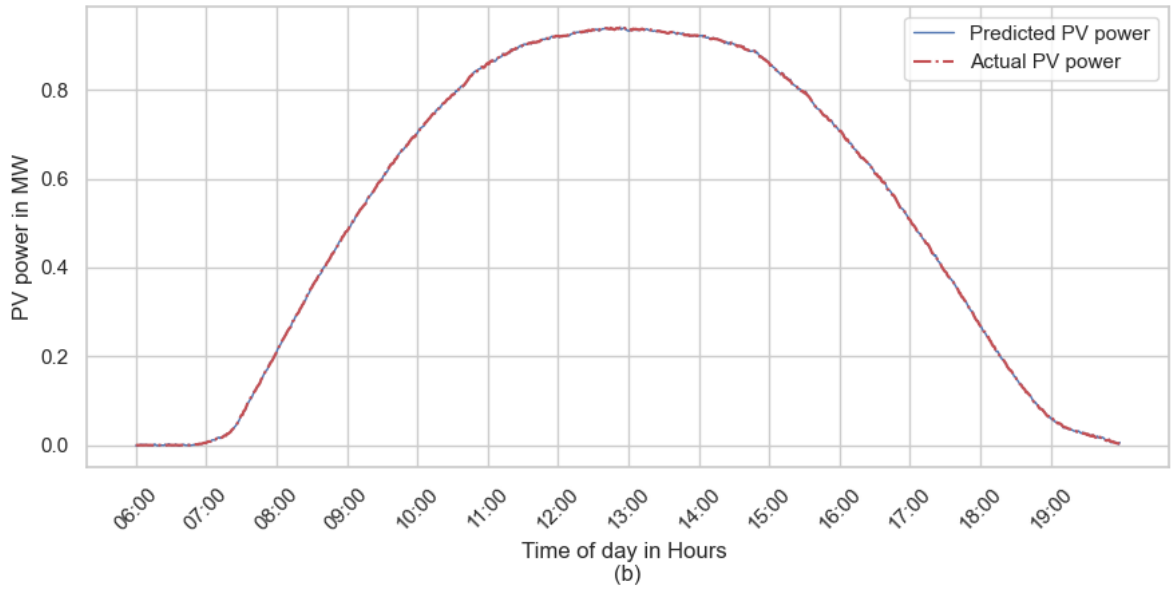
Figure 5.23 Scenarios for DPV power forecasting using PV plant DTs, hybrid PV plants and virtual PV plants

Case 1: Physical weather station data streams feeding into existing DPVs or PV locations with measurement data

In a similar workflow to Case 1 in situational intelligence, short term forecasts can be made over for the PV Plant at R06 utilizing the PV DTs for predictions using input measurements from the PWS present at the site. This is useful for studies such as cloud cover impact, volt-var regulation and optimization support, and parallelized scenario-testing as well as useful information for energy markets, when longer time-scale predictions are made.

The results for various day types and different time-scales are seen in Figures 5.24-5.46 and in Table 5.5.





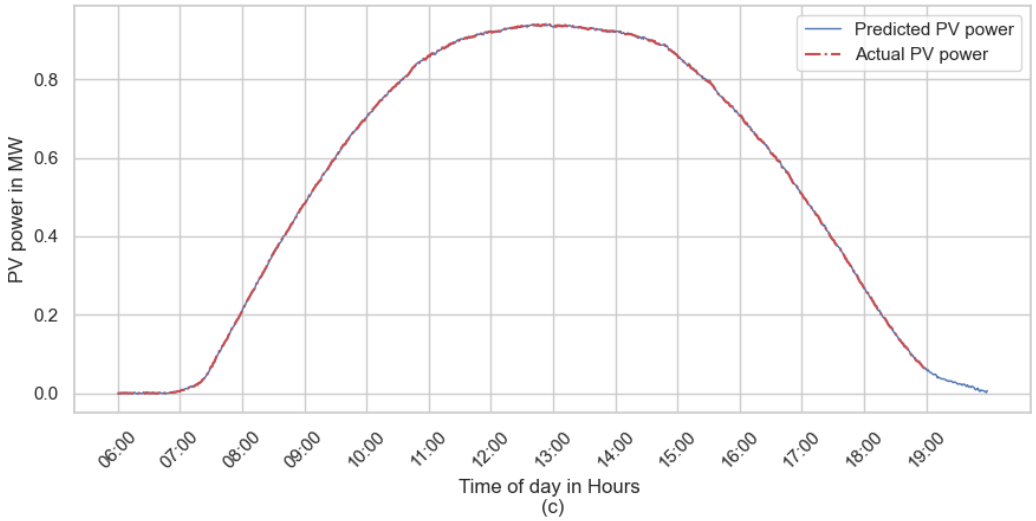


Figure 5.24 Case 1: 60 (a), 120 (b) and 180 (c) seconds ahead predicted PV power at R06 using PWS for a sunny day

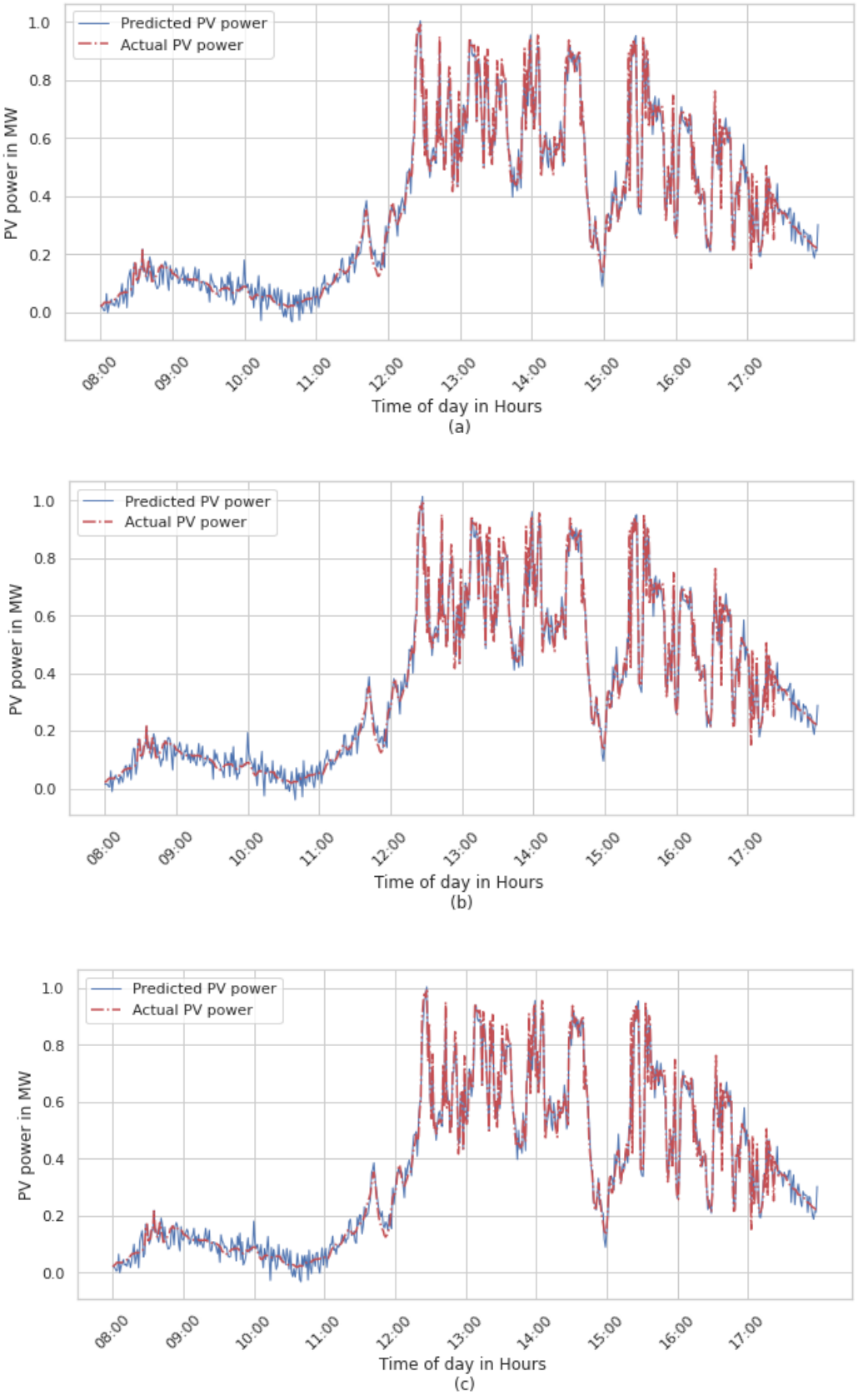
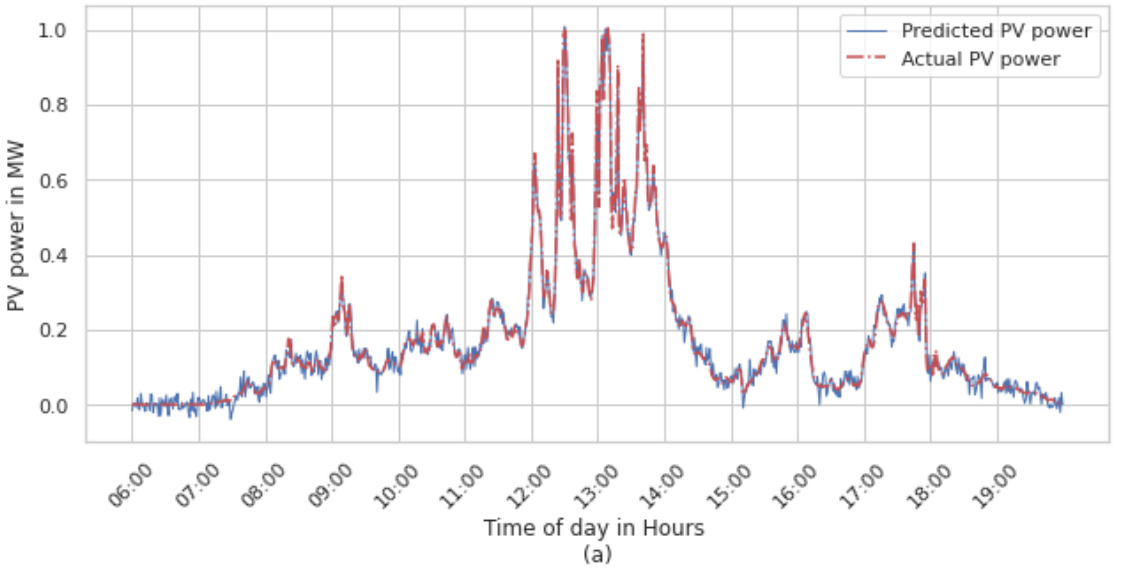
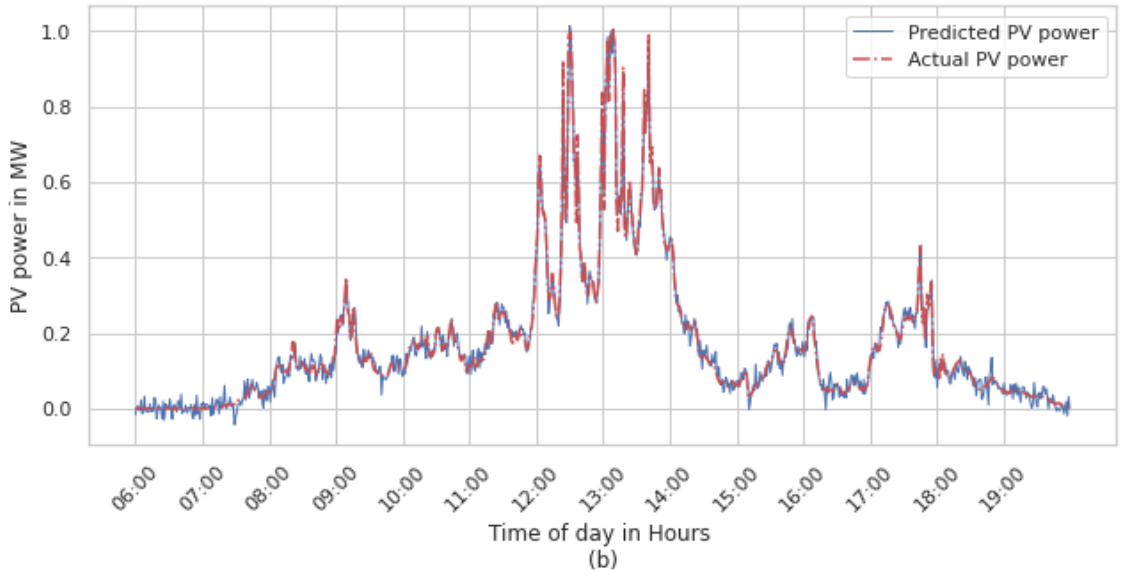


Figure 5.25 Case 1: 60 (a), 120 (b) and 180 (c) seconds ahead predicted PV power at R06 using PWS for a moderately cloudy day





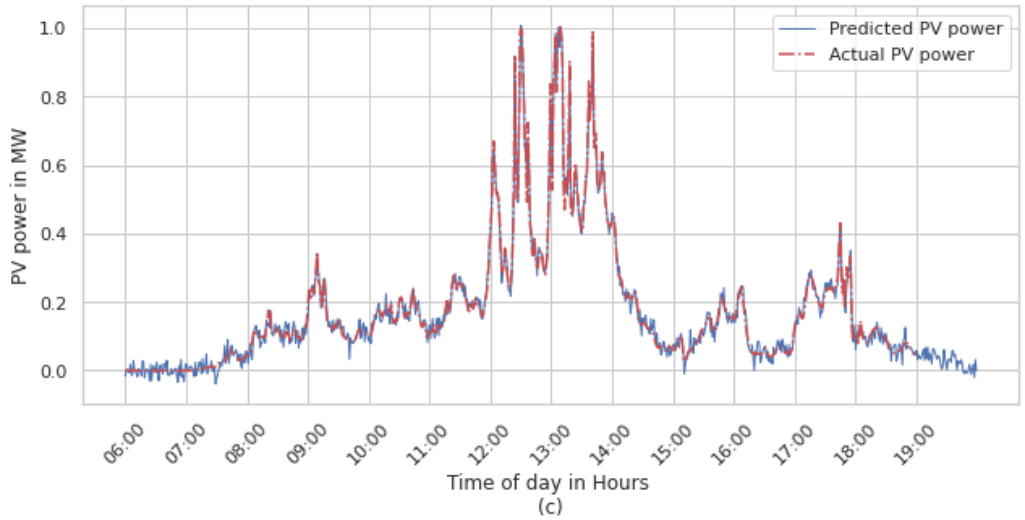


Figure 5.26 Case 1: 60 (a), 120 (b) and 180 (c) seconds ahead predicted PV power at R06 using PWS for a cloudy day

Table 5.4 DT performance for PV power prediction

Day Category	*Reservoir neurons*	*Metric*	*60 seconds*	*120 seconds*	*180 seconds*
Sunny	200	MAE	0.0005	0.0006	0.0011
	200	MAPE	0.023	0.0353	0.0423
	200	MSE	0.001	0.0013	0.0014
	200	RMSE	0.0316	0.0361	0.0374
Moderately Cloudy	550	MAE	0.0063	0.0089	0.012
	550	MAPE	0.2456	0.3571	0.3345
	550	MSE	0.0079	0.0123	0.0149
	550	RMSE	0.0889	0.1109	0.1221
Cloudy	450	MAE	0.016	0.0168	0.0194
	450	MAPE	0.2124	0.2232	0.2361
	450	MSE	0.0183	0.019	0.0204
	450	RMSE	0.1353	0.1378	0.1428

Case 2: Physical weather station data streams feeding into unknown DPVs or PV locations without measurement data

As in Case 2 for situational intelligence, PV power prediction can be done on "hybrid" sites, where a PWS is present, but PV plants have not been installed in the area where PMU, separate meter readings or other measurement data is unavailable. This allows for detailed analysis of PV dynamics for distribution system studies. Figures 5.27- 5.32 show PV predictions at the Ravenel and Airport PWS for various day types and over different prediction intervals.

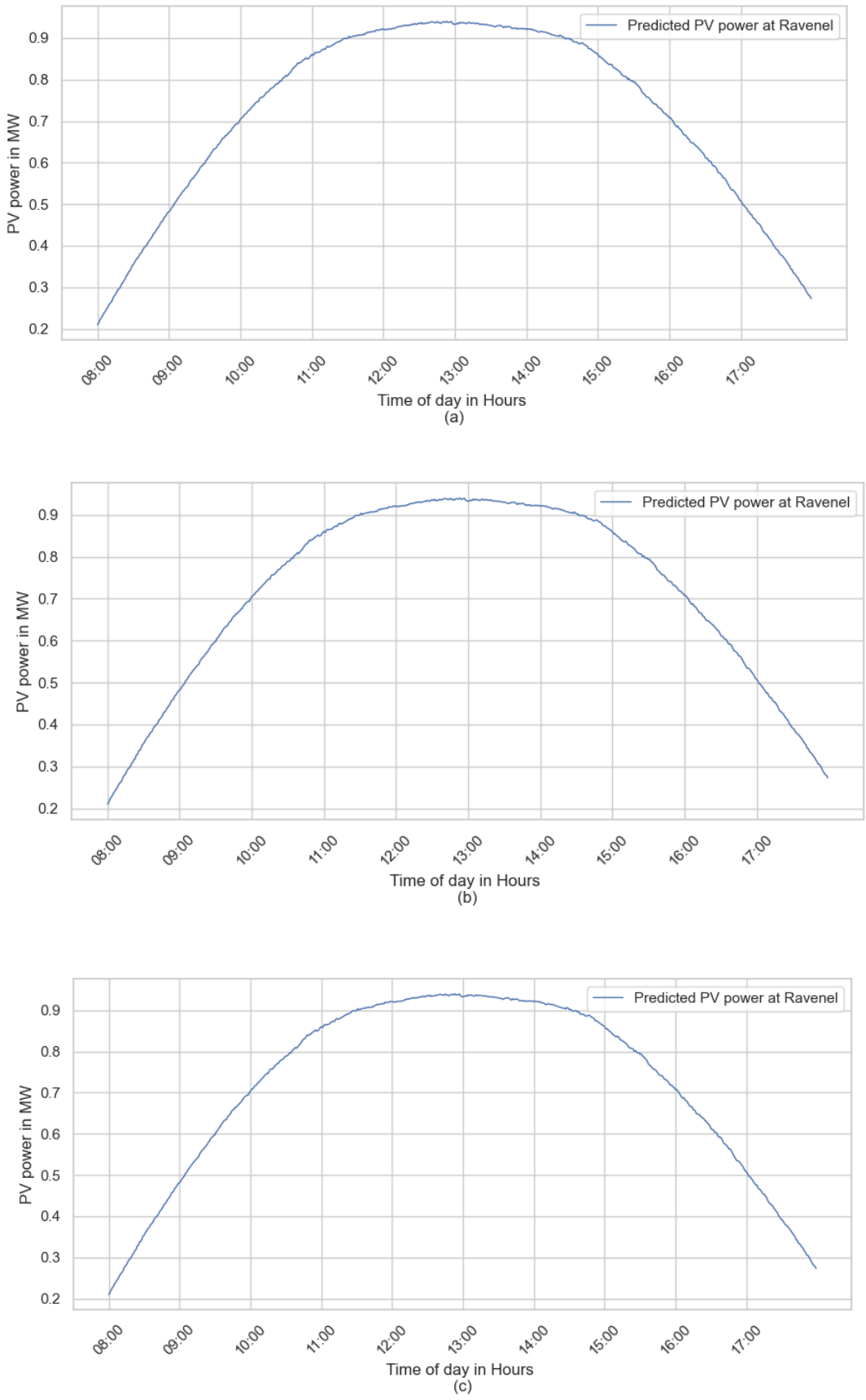
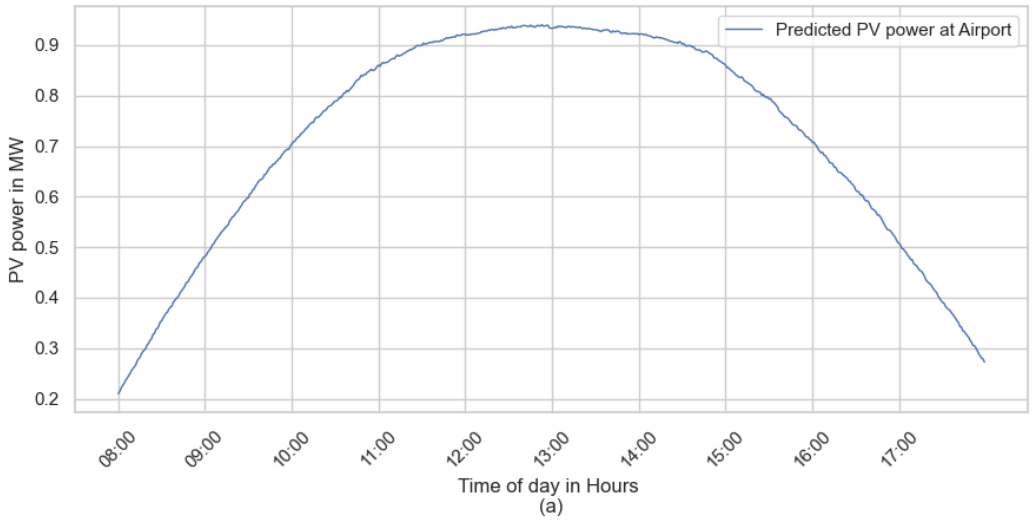
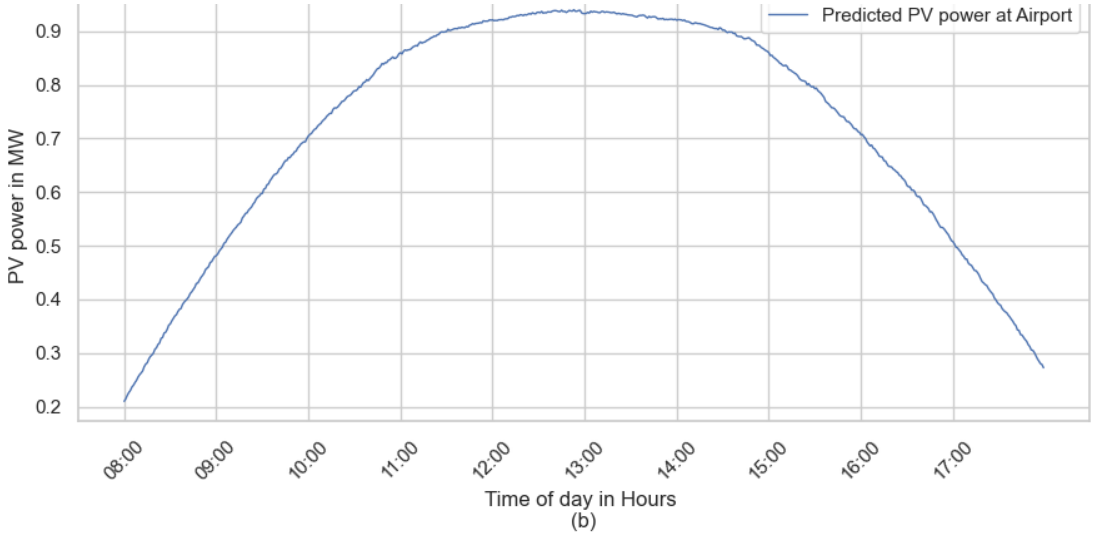


Figure 5.27 Case 2, 60, 120, 180 seconds ahead predicted PV power at Ravenel using PWS

for a sunny day



(a)



(b)

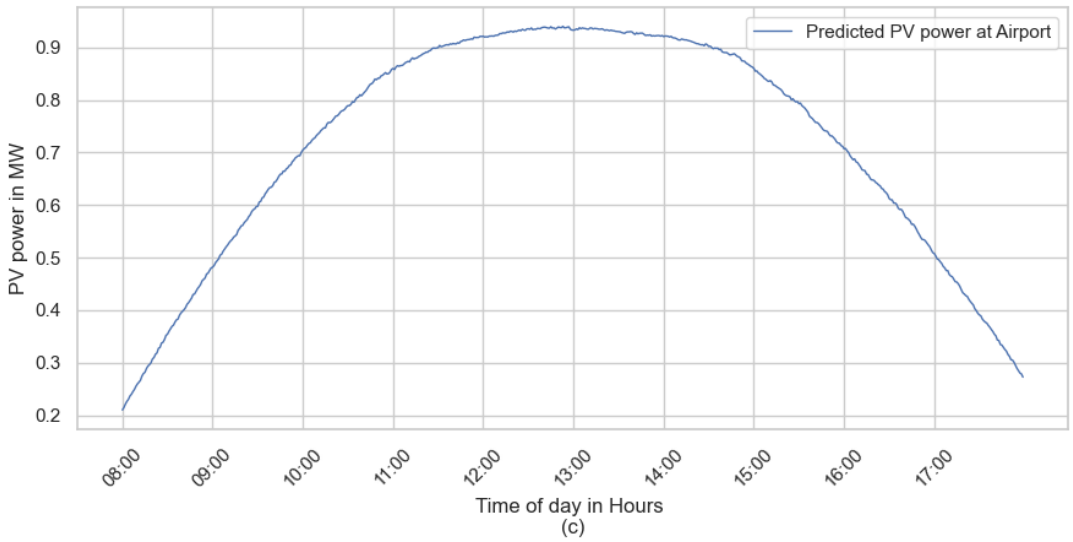


Figure 5.28 Case 2, 60, 120, 180 seconds ahead predicted PV power at Airport using PWS for a sunny day

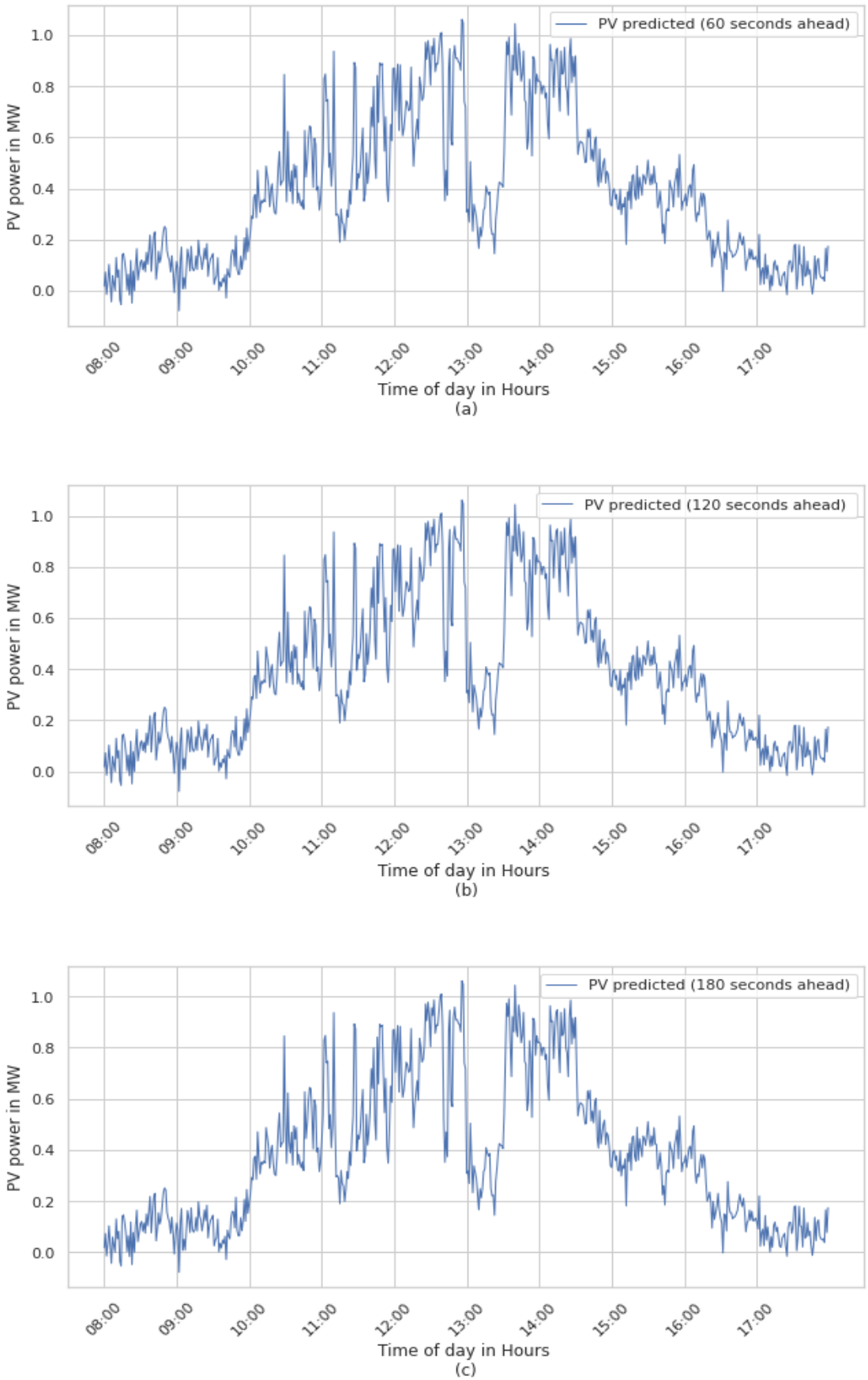
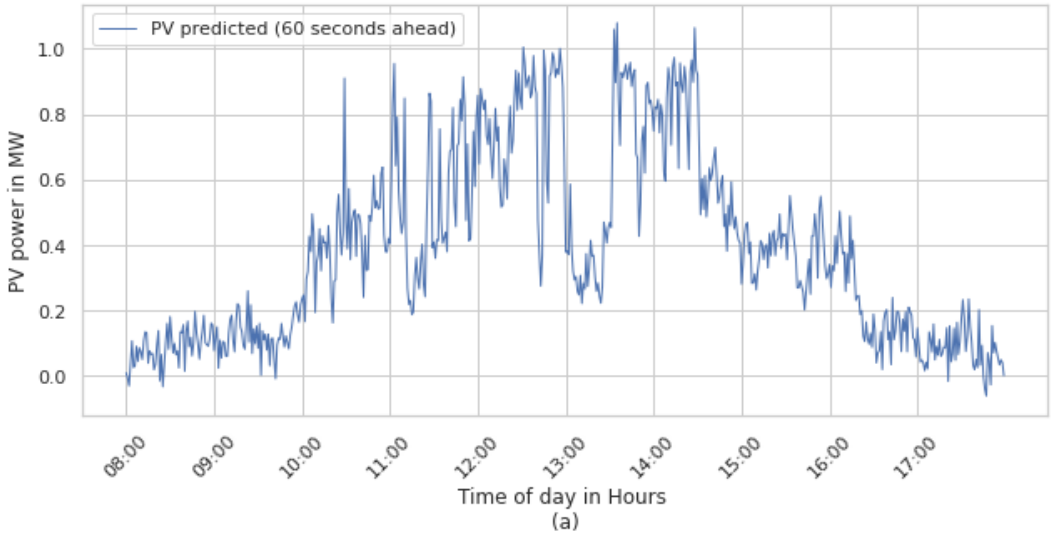
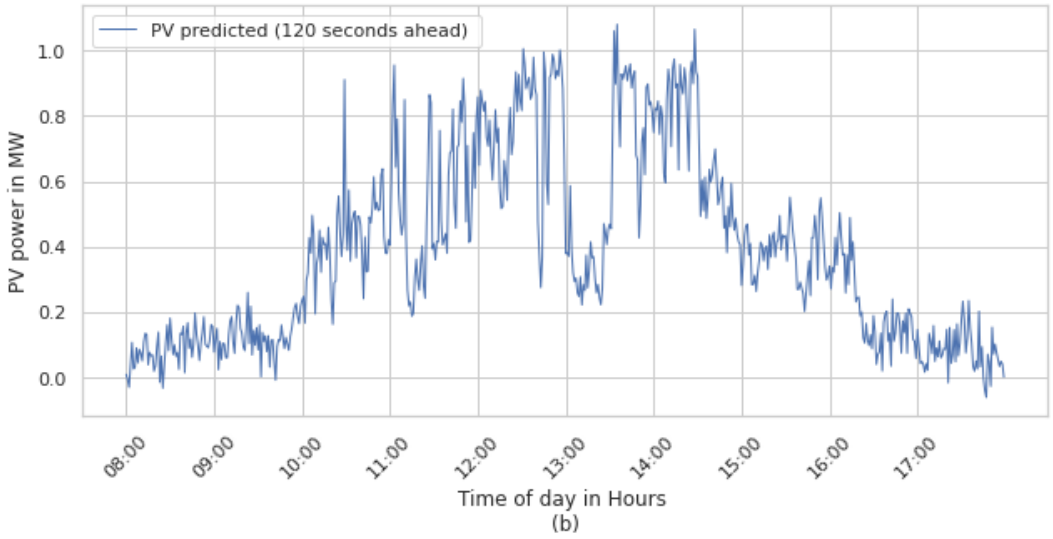


Figure 5.29 Case 2, 60 (a), 120 (b), 180 (c) seconds ahead predicted PV power at Airport using PWS for a moderately cloudy day





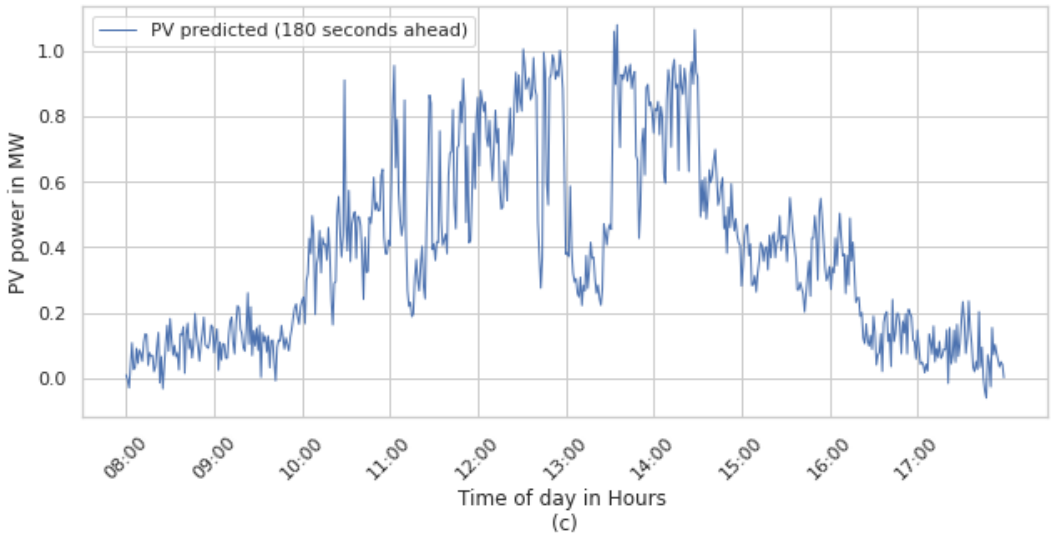
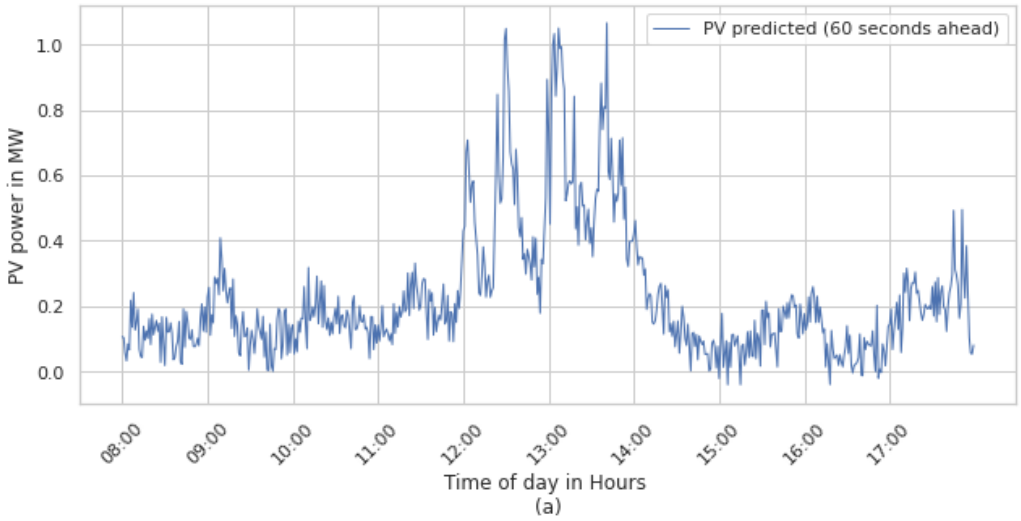
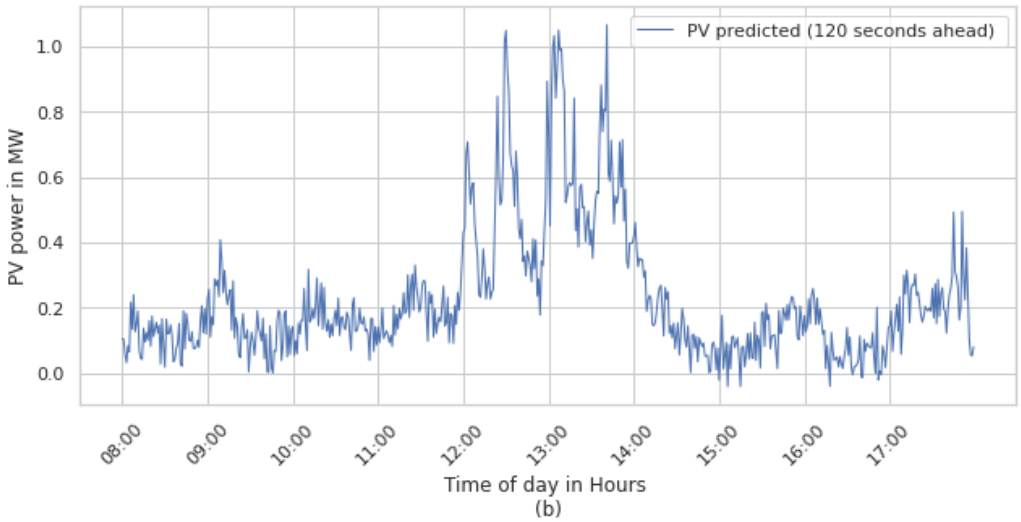


Figure 5.30 Case 2, 60 (a), 120 (b), 180 (c) seconds ahead predicted PV power at Ravenel using PWS for a moderately cloudy day





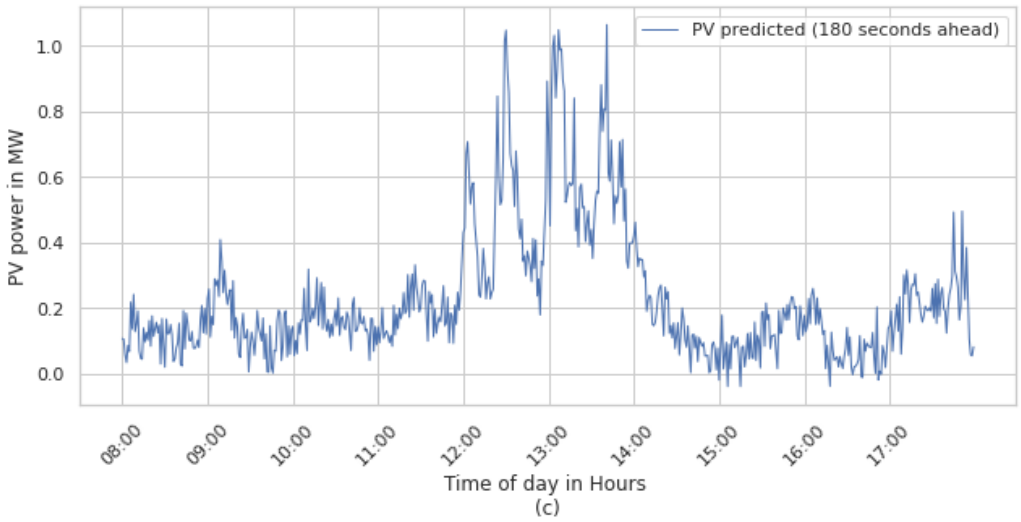


Figure 5.31 Case 2, 60 (a), 120 (b), 180 (c) seconds ahead predicted PV power at Airport using PWS for a cloudy day

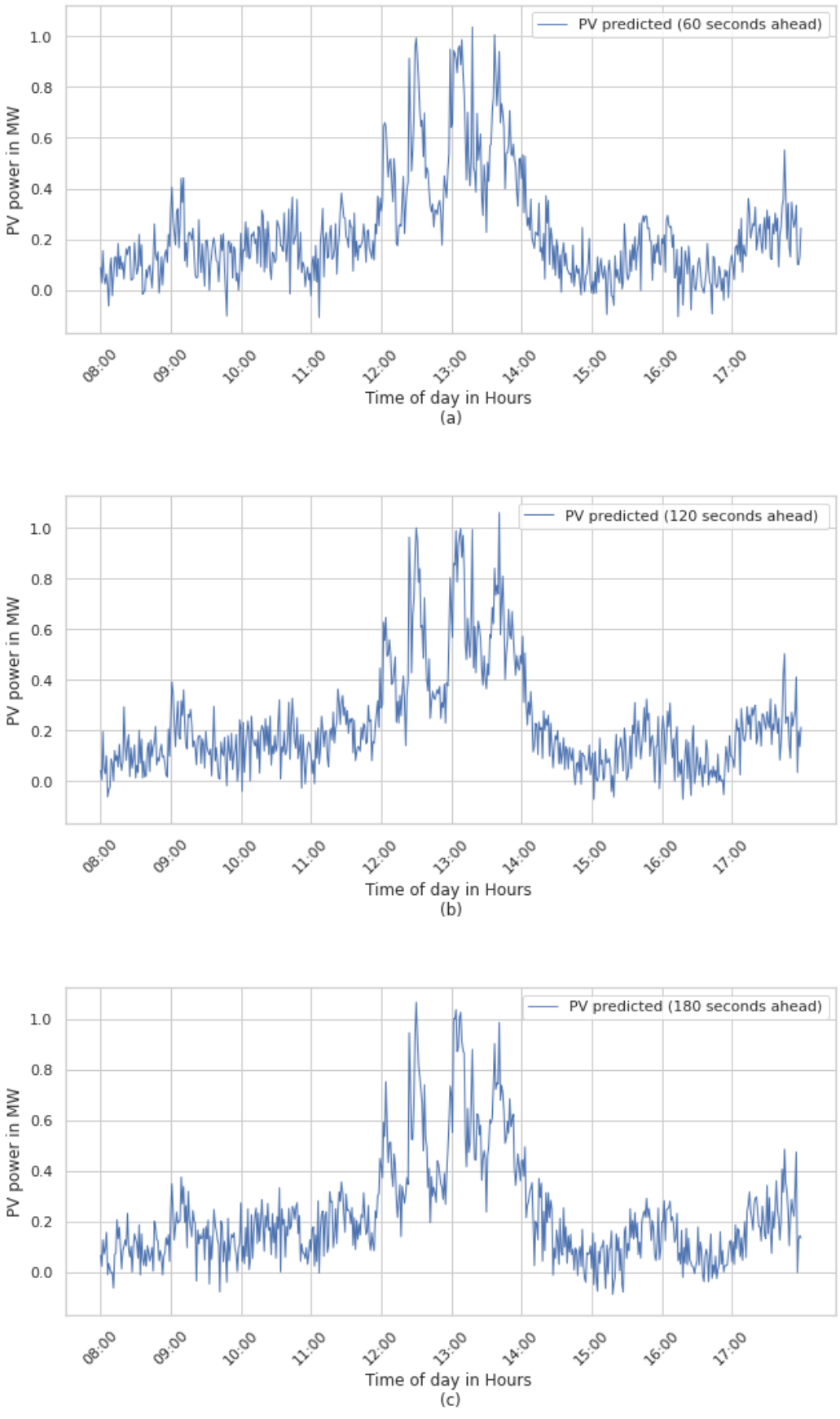


Figure 5.32 Case 2, 60 (a), 120 (b), 180 (c) seconds ahead predicted PV power at Ravenel using PWS for a moderately cloudy day

Case 3: Virtual weather station data streams feeding into unknown DPVs locations or PV locations without measurement data

Finally, as in case 3, PV power can be predicted at VWS sites, where no PV plant has yet been installed or where micro-PMU, separate meter readings or other measurement data is unavailable. These virtual PV sources can be easily adapted to be integrated into a distribution testbed for short term PV forecasts, data collection and general situational intelligence of PV power generation in the area. The results of the study for virtual PV sites for various day types and different time-scales can be seen in Figures 5.33-5.64.

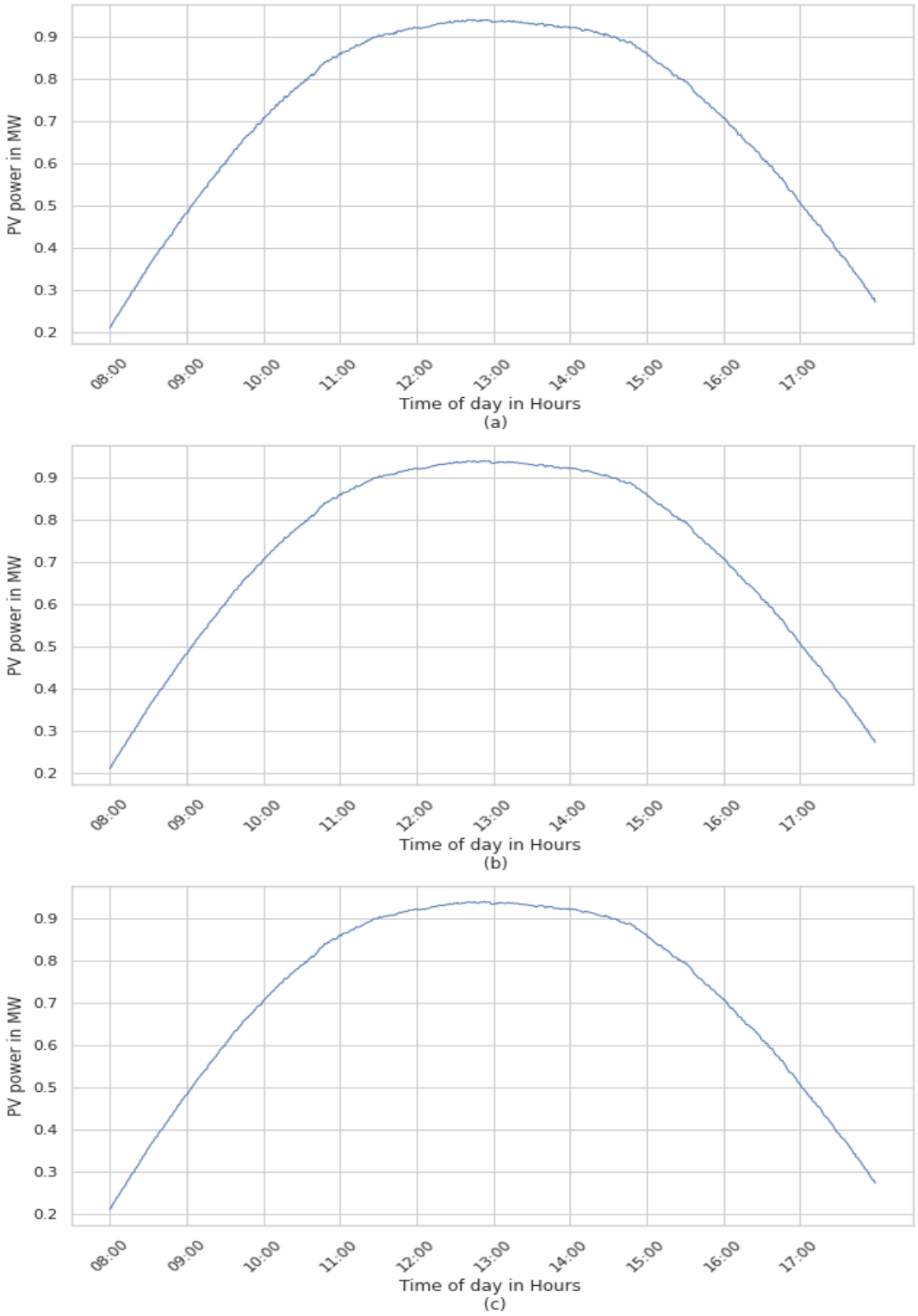
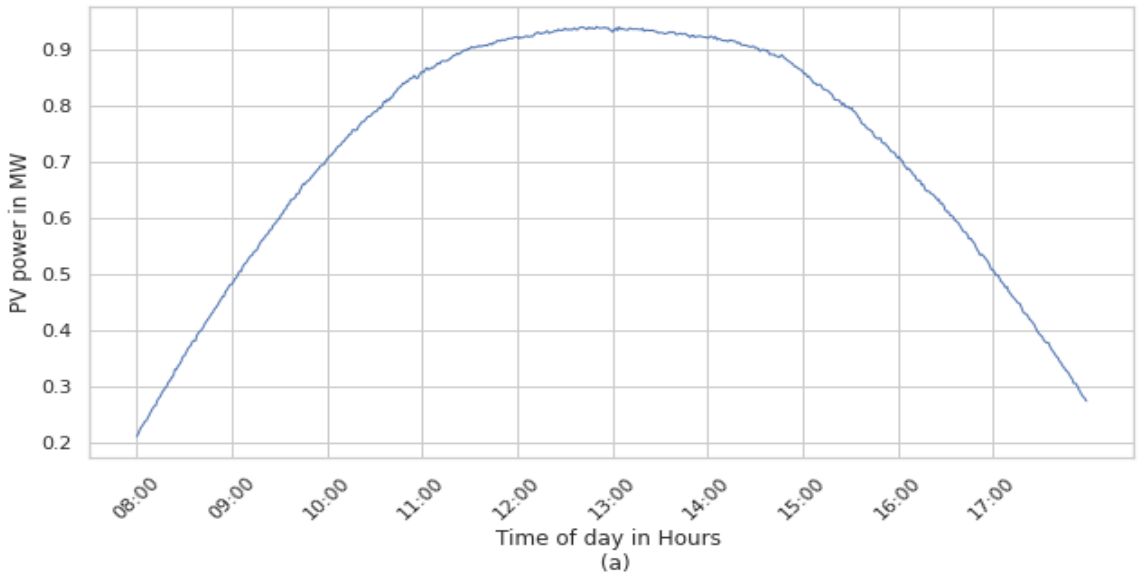
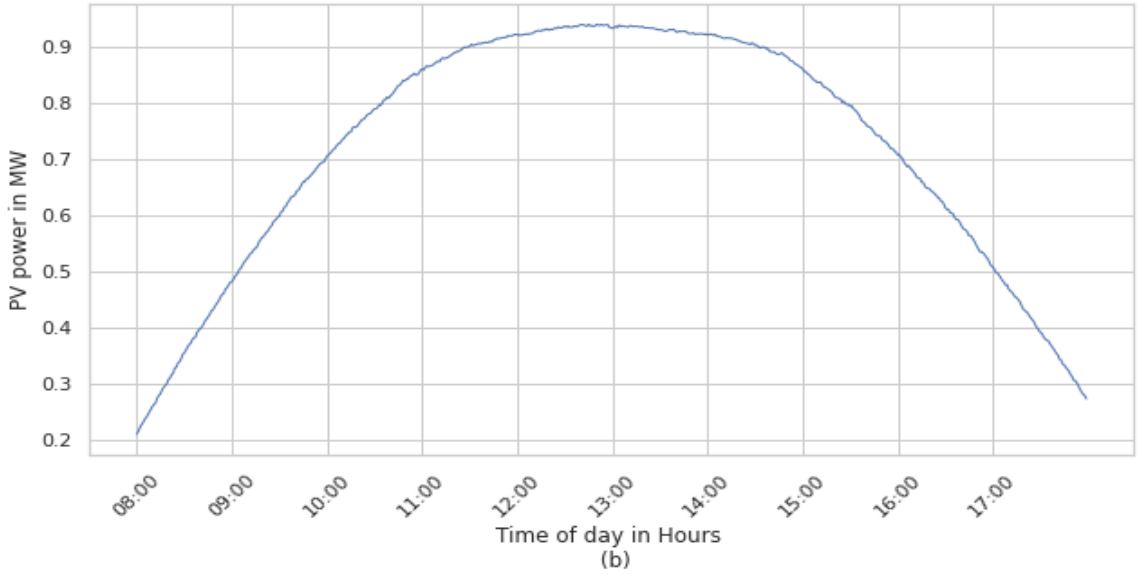


Figure 5.33 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 1 using VWS for a sunny day





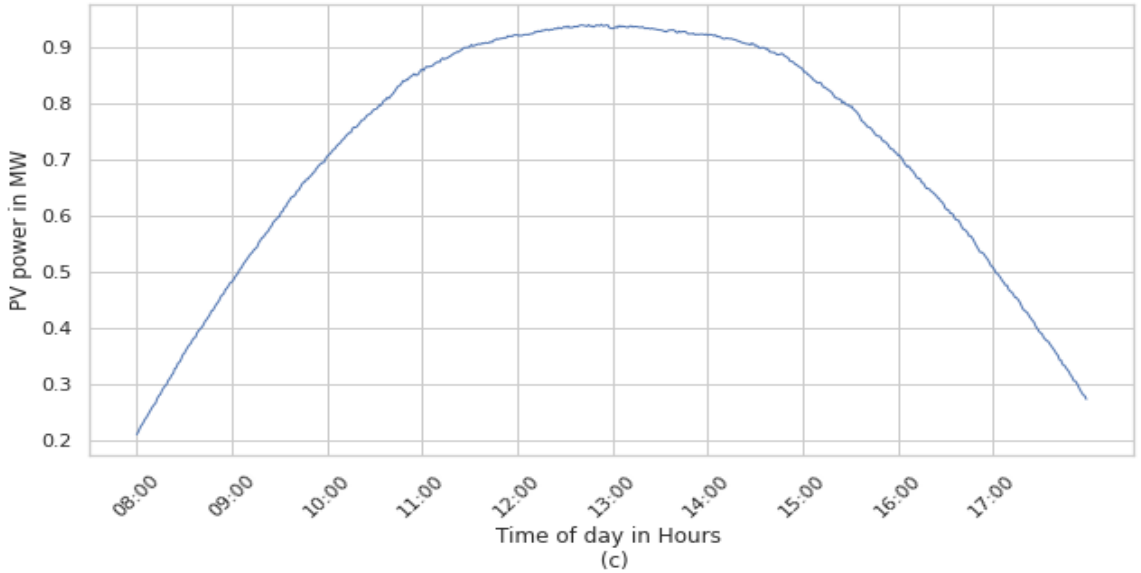


Figure 5.34 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 2 using VWS for a sunny day

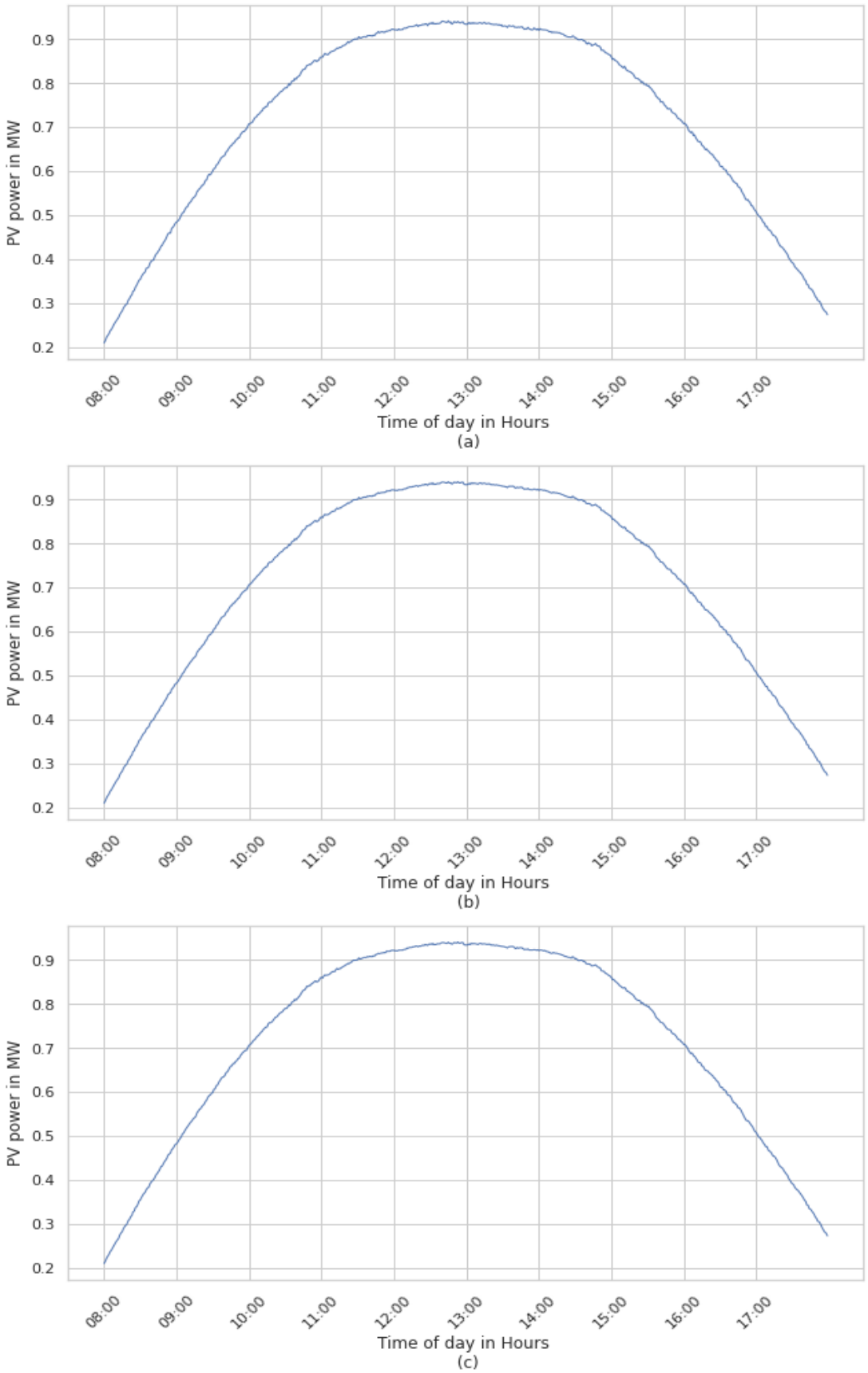


Figure 5.35 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 3 using VWS for a sunny day

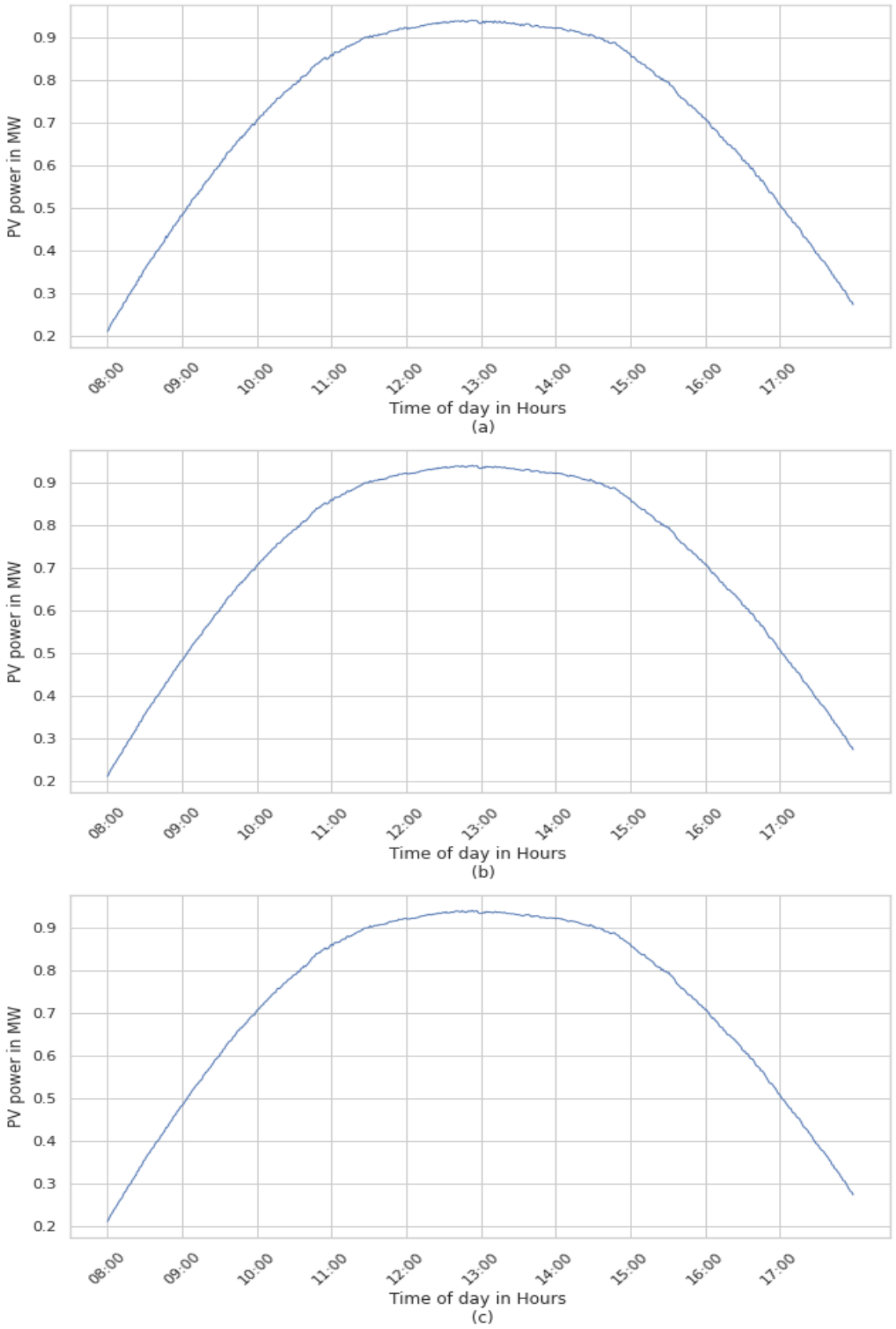
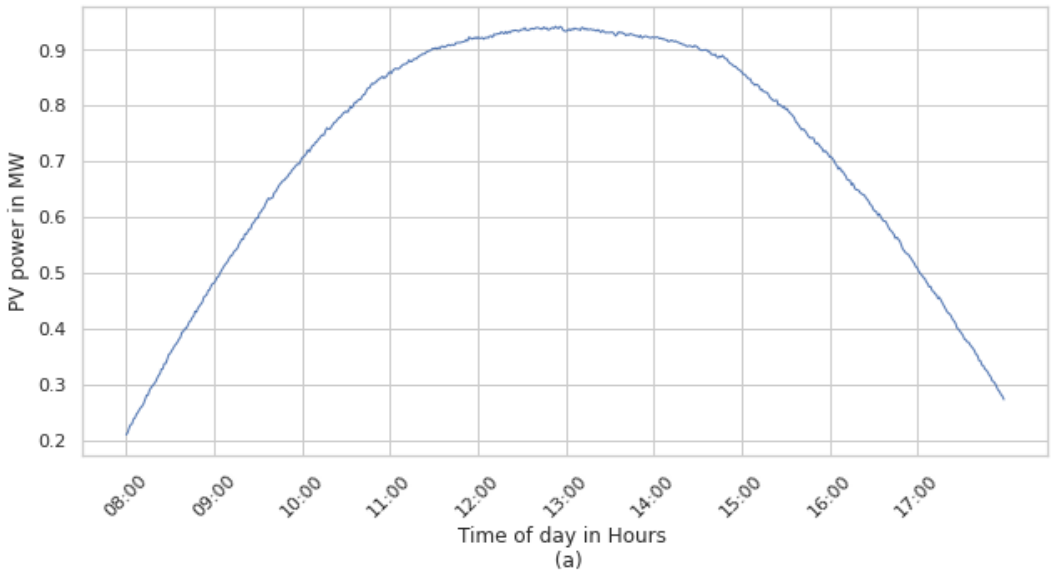
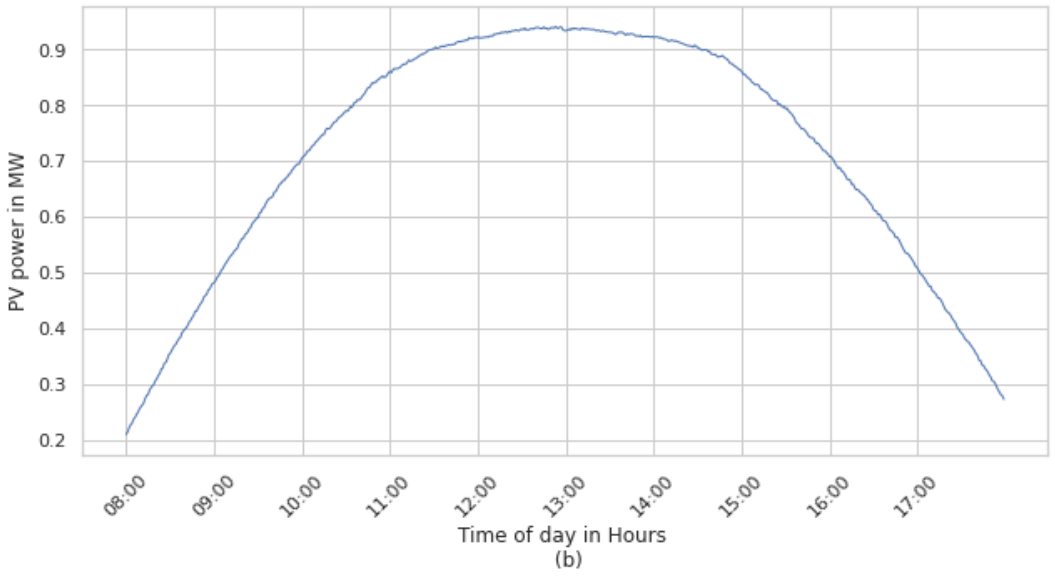


Figure 5.36 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 4 using VWS for a sunny day





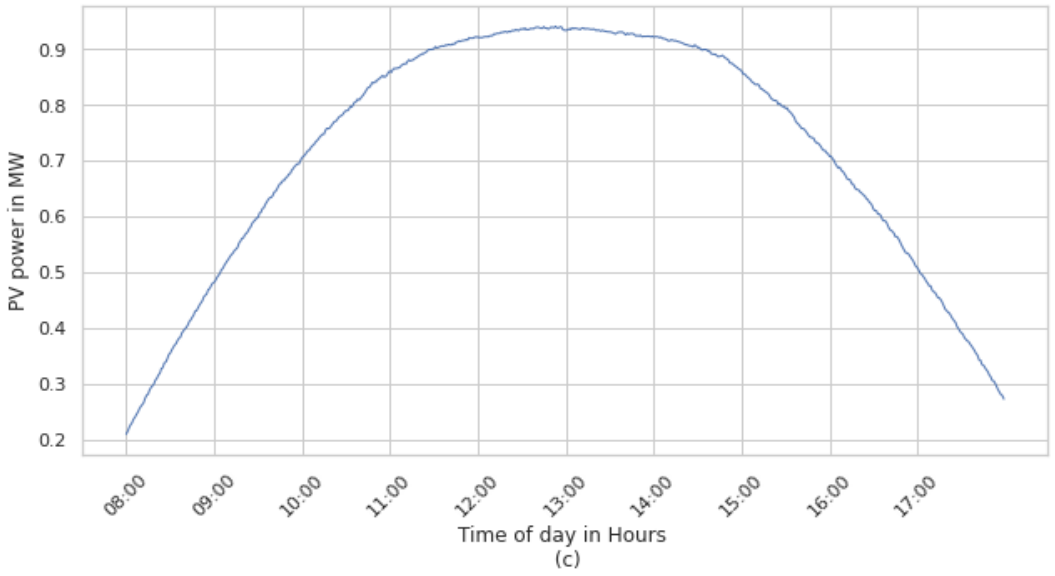


Figure 5.37 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 5 using VWS for a sunny day

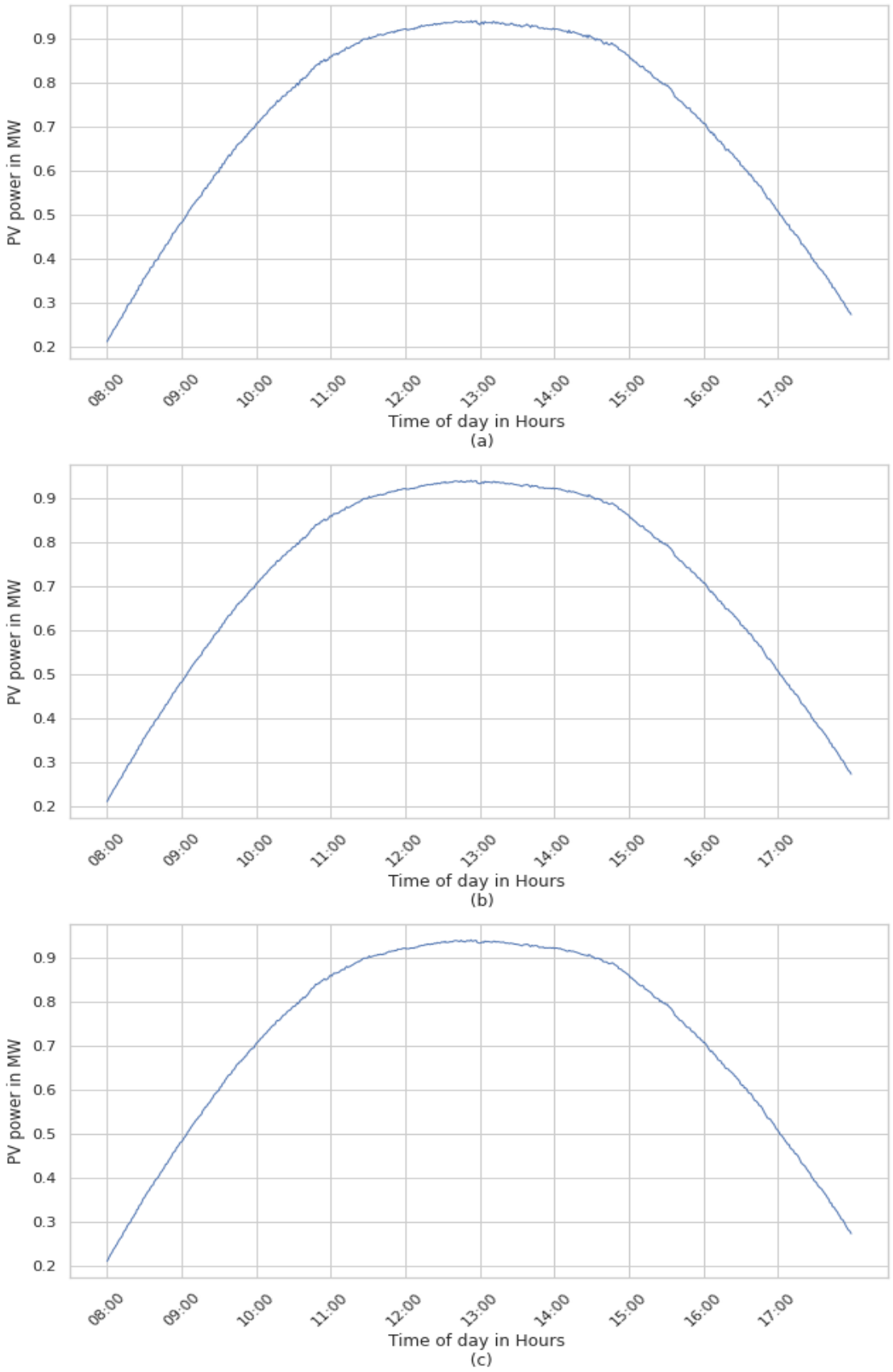


Figure 5.38 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 6 using VWS for a sunny day

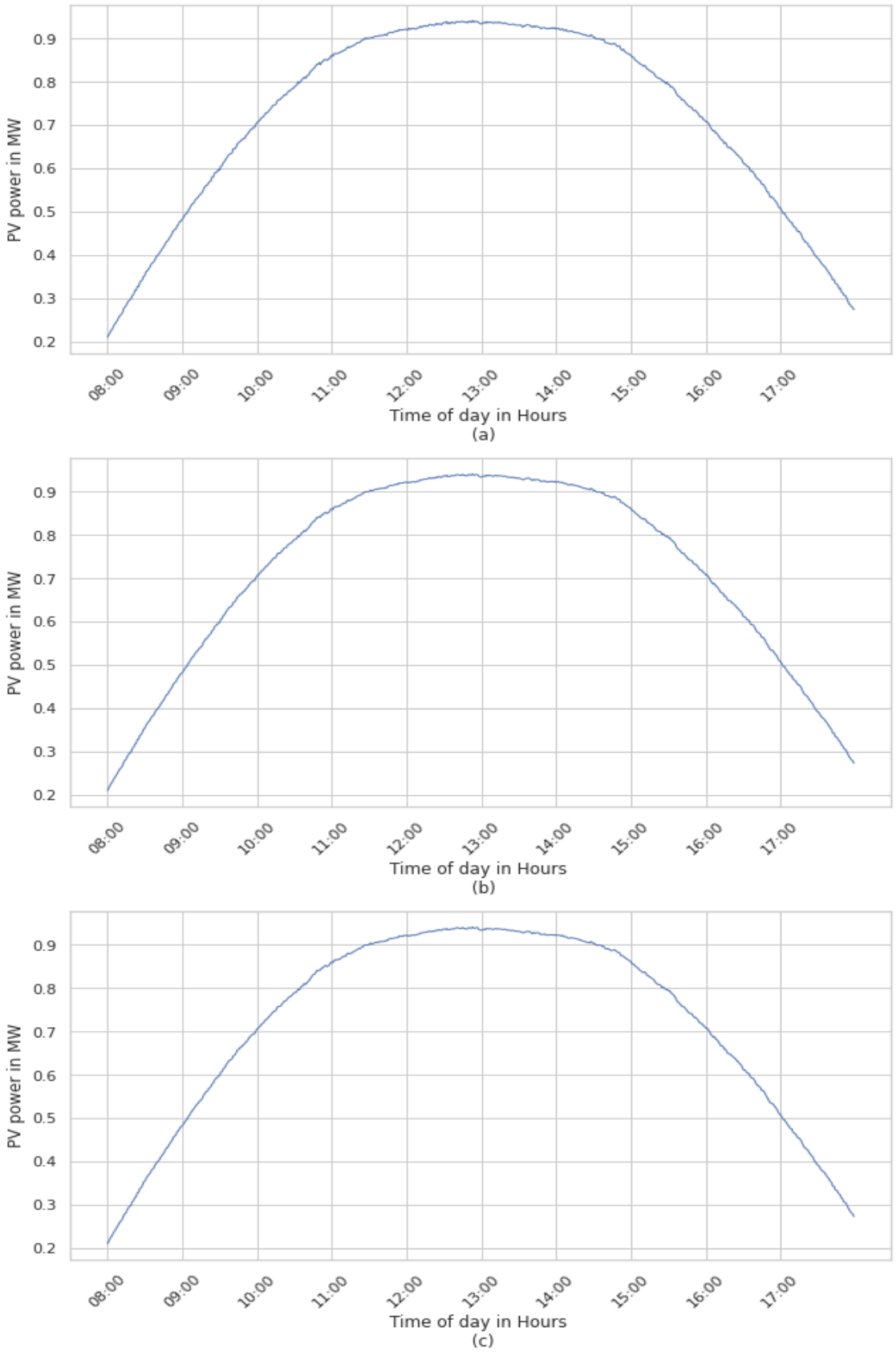


Figure 5.39 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 7 using VWS for a sunny day

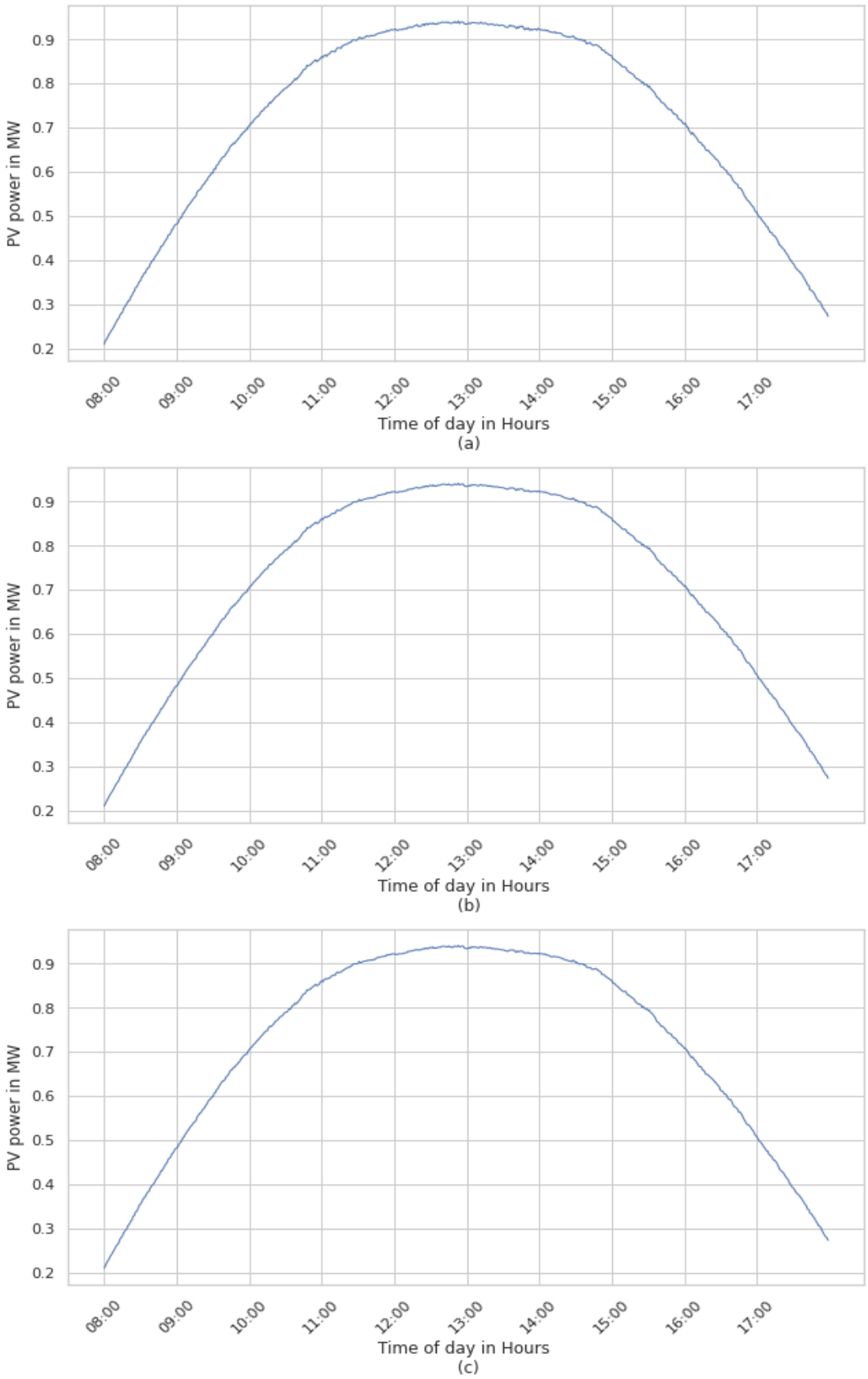


Figure 5.40 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 8 using VWS for a sunny day

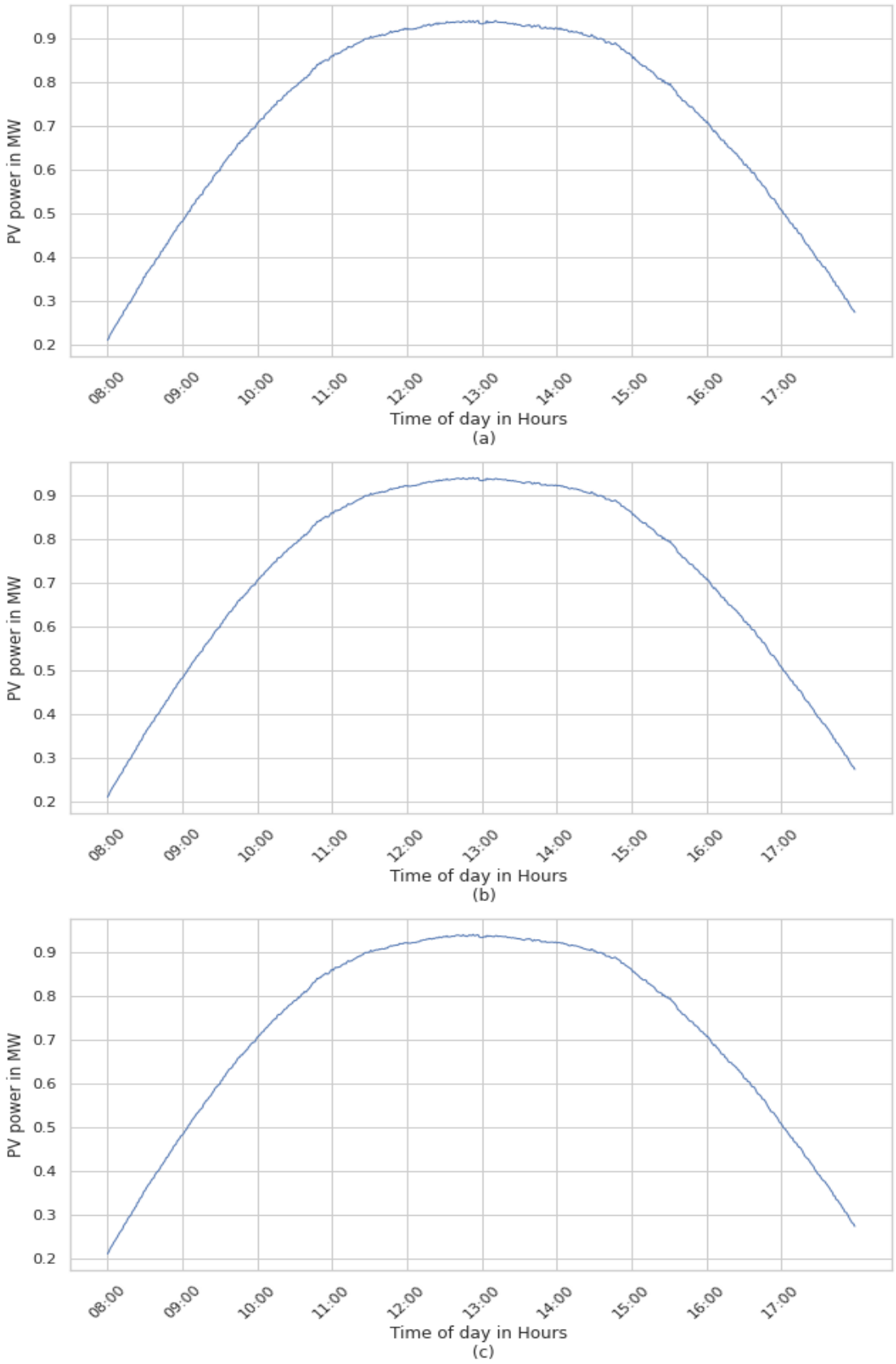
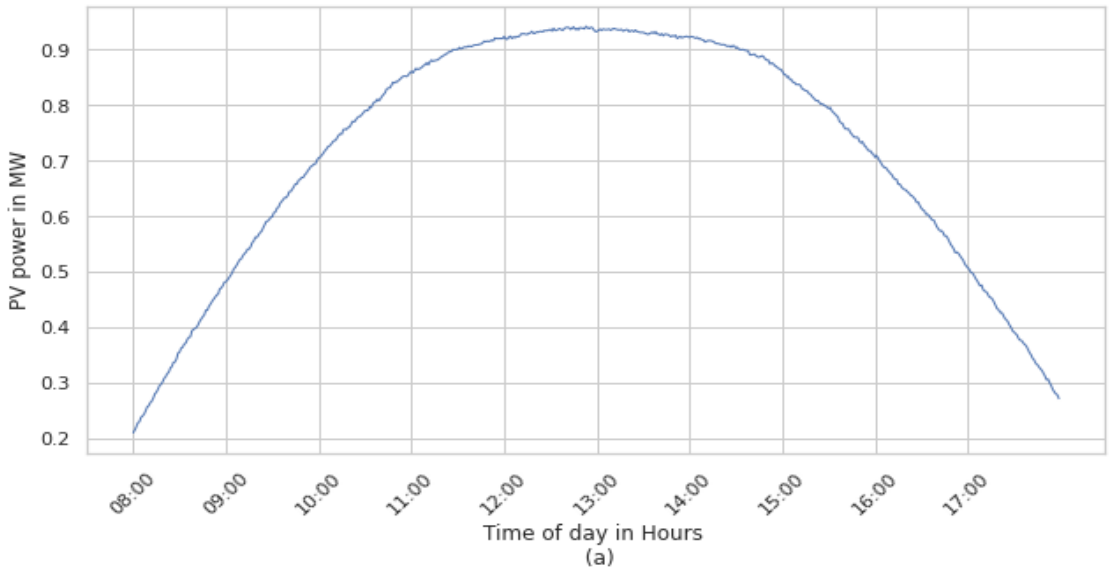
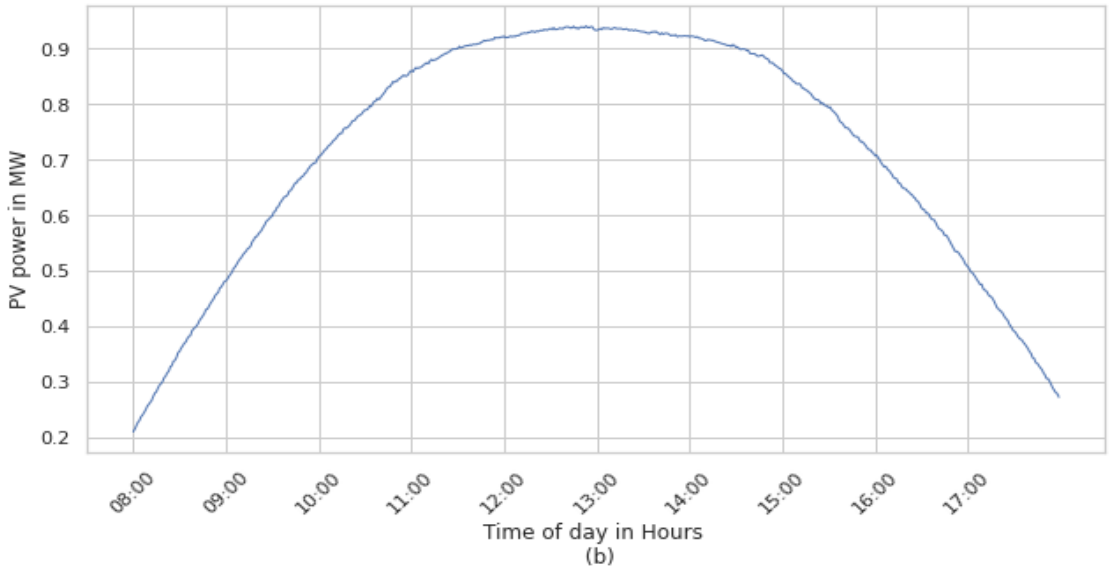


Figure 5.41 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 9 using VWS for a sunny day





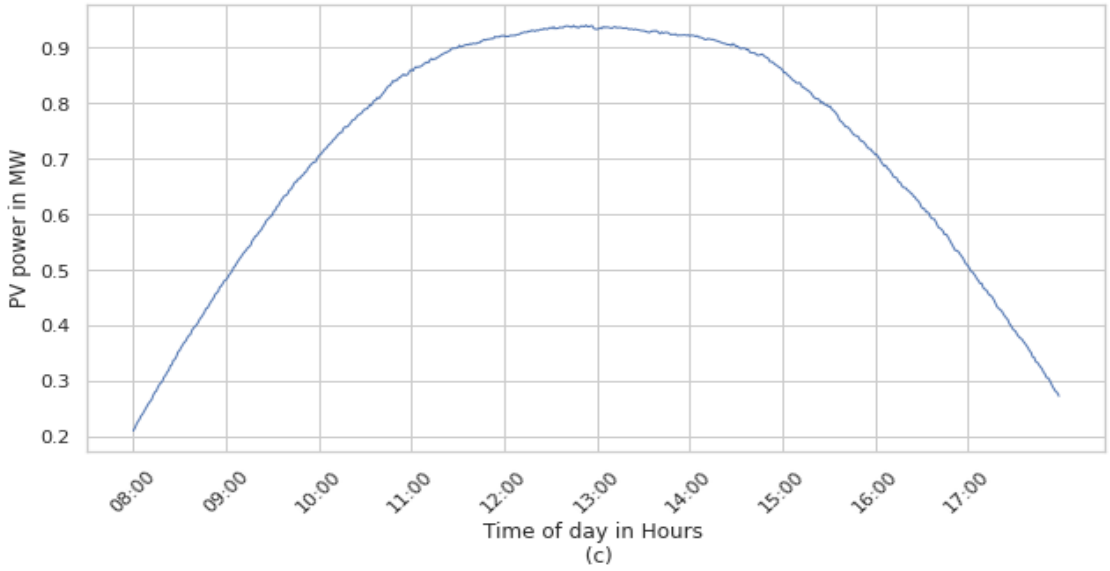


Figure 5.42 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 10 using VWS for a sunny day

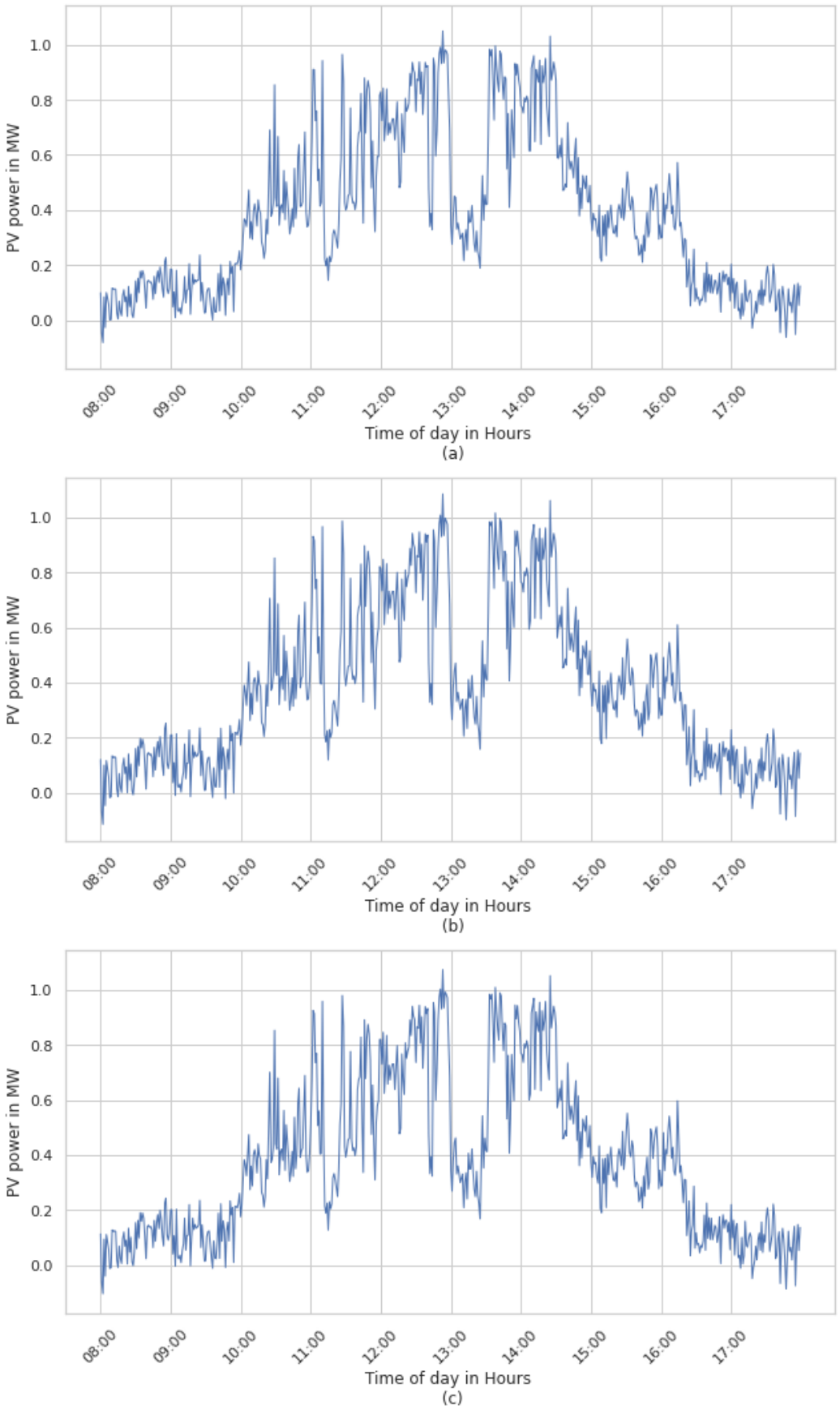


Figure 5.43 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 1 using VWS for a moderately cloudy day

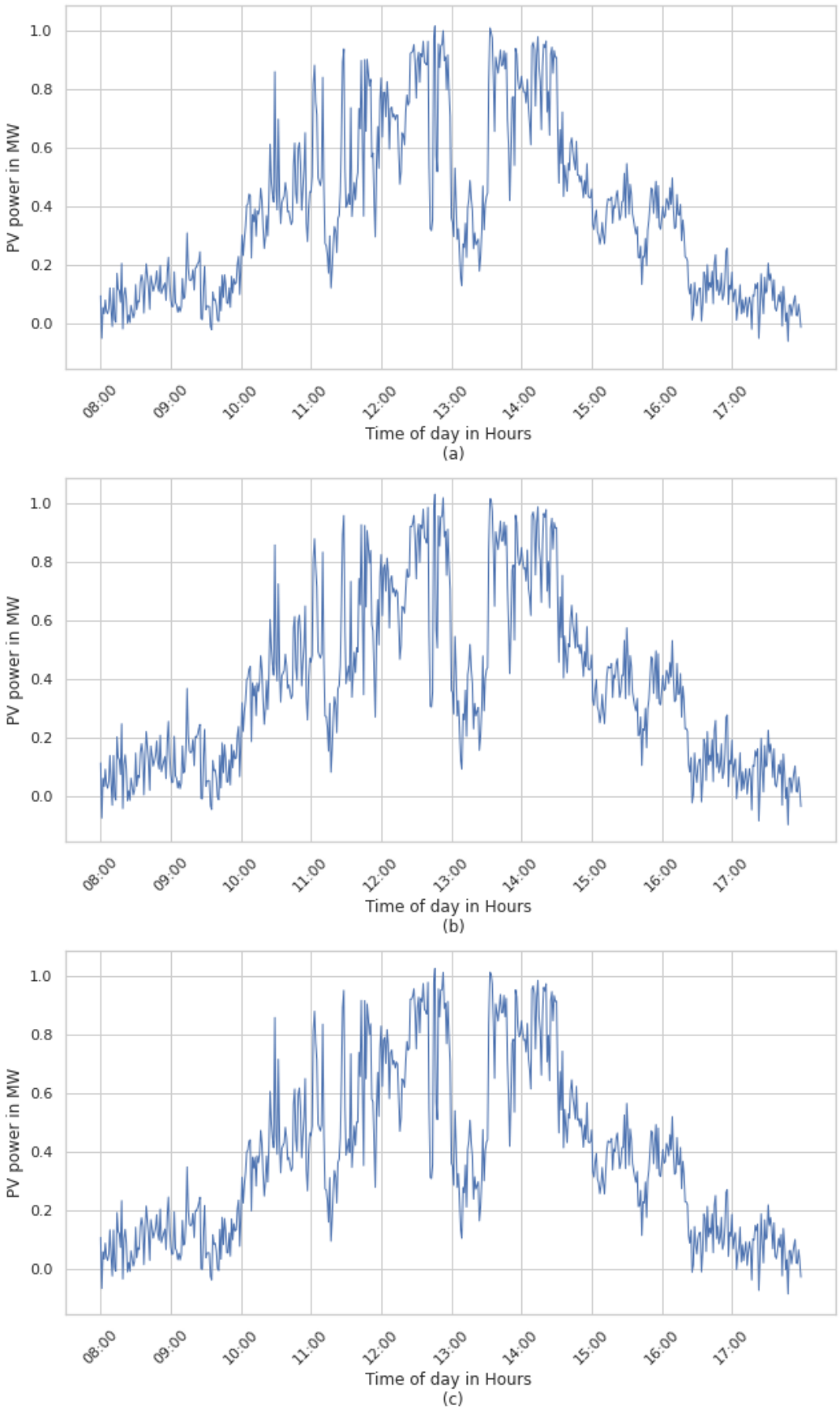


Figure 5.44 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 2 using VWS for a moderately cloudy day

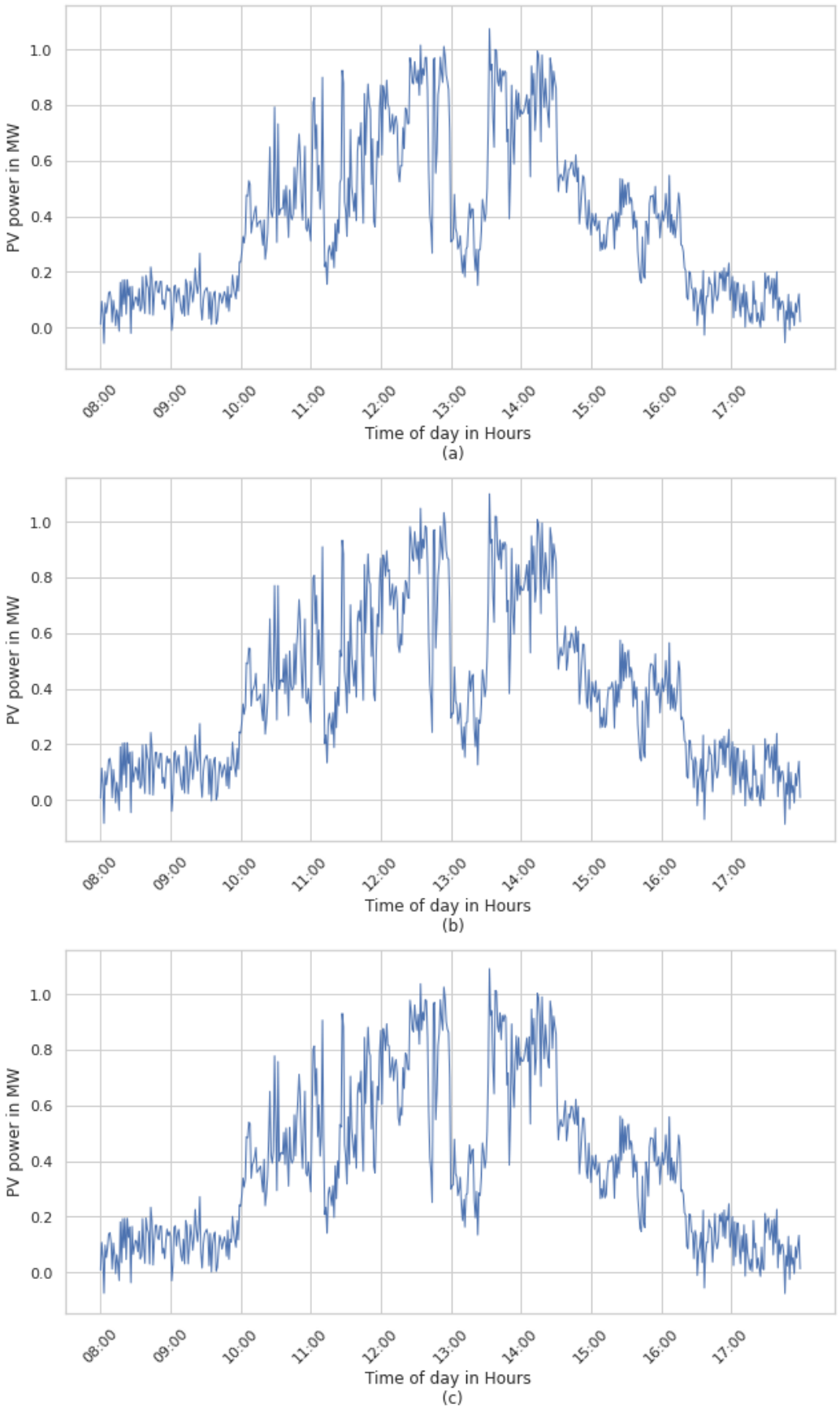


Figure 5.45 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 3 using VWS for a moderately cloudy day

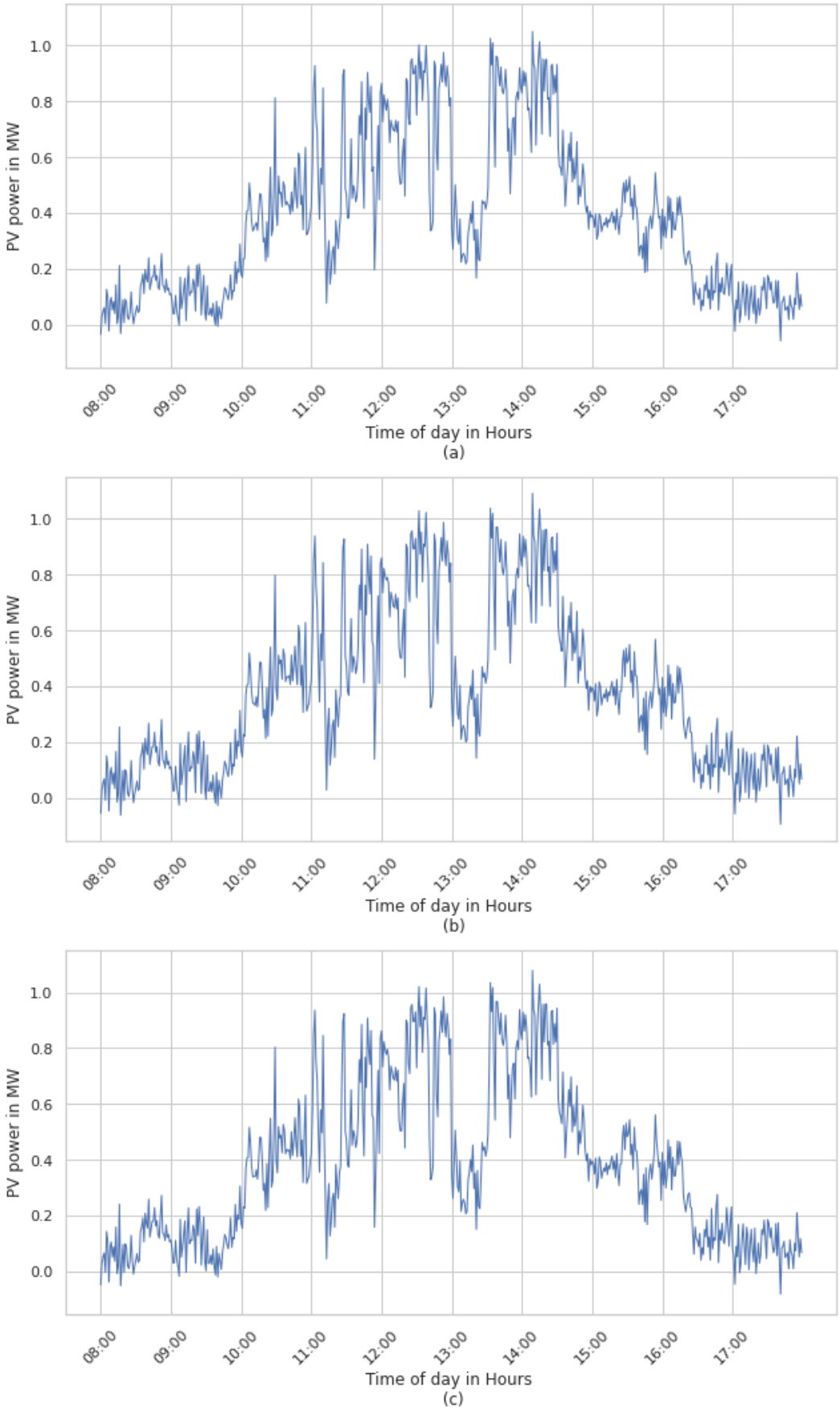


Figure 5.46 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 4 using VWS for a moderately cloudy day

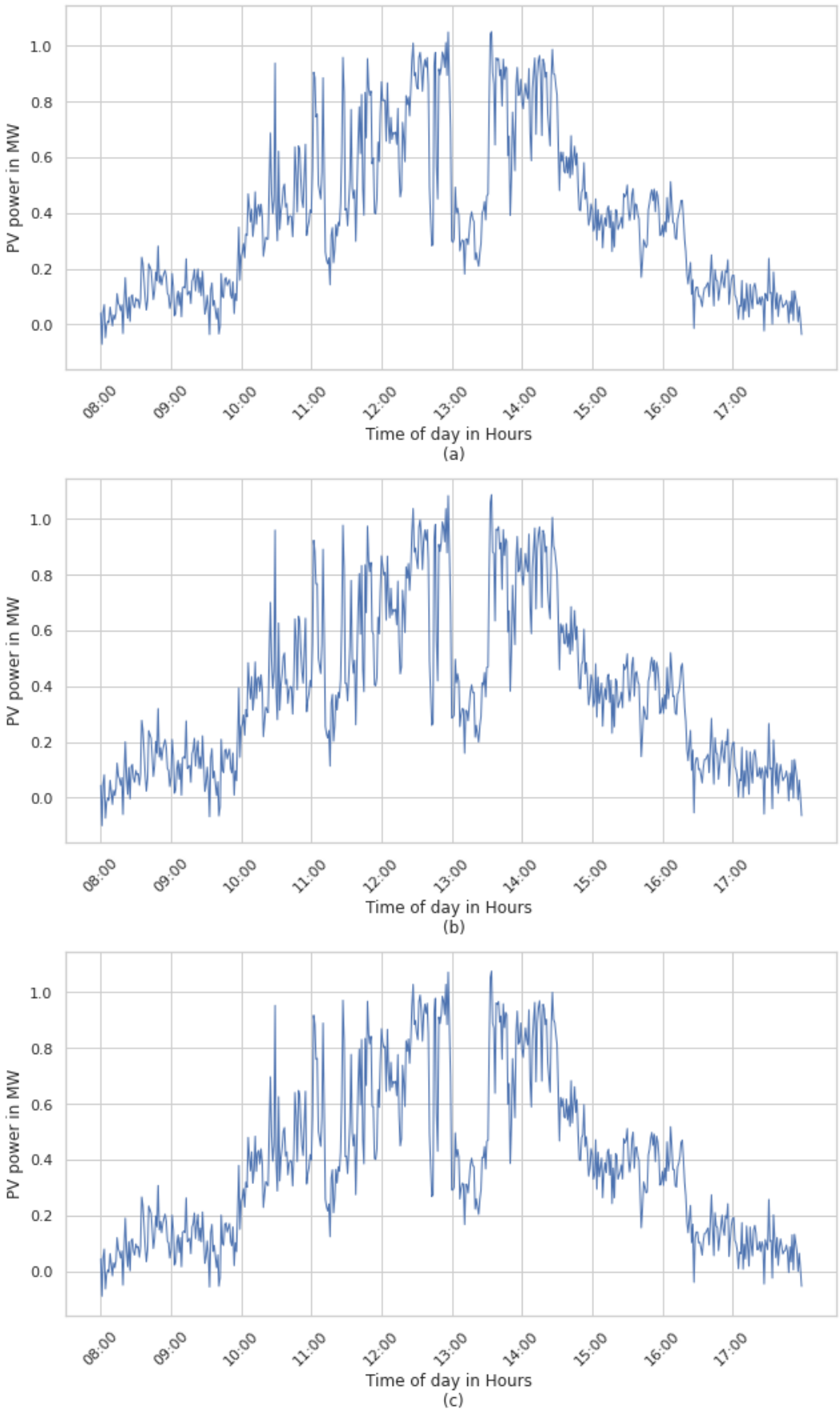


Figure 5.47 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 5 using VWS for a moderately cloudy day

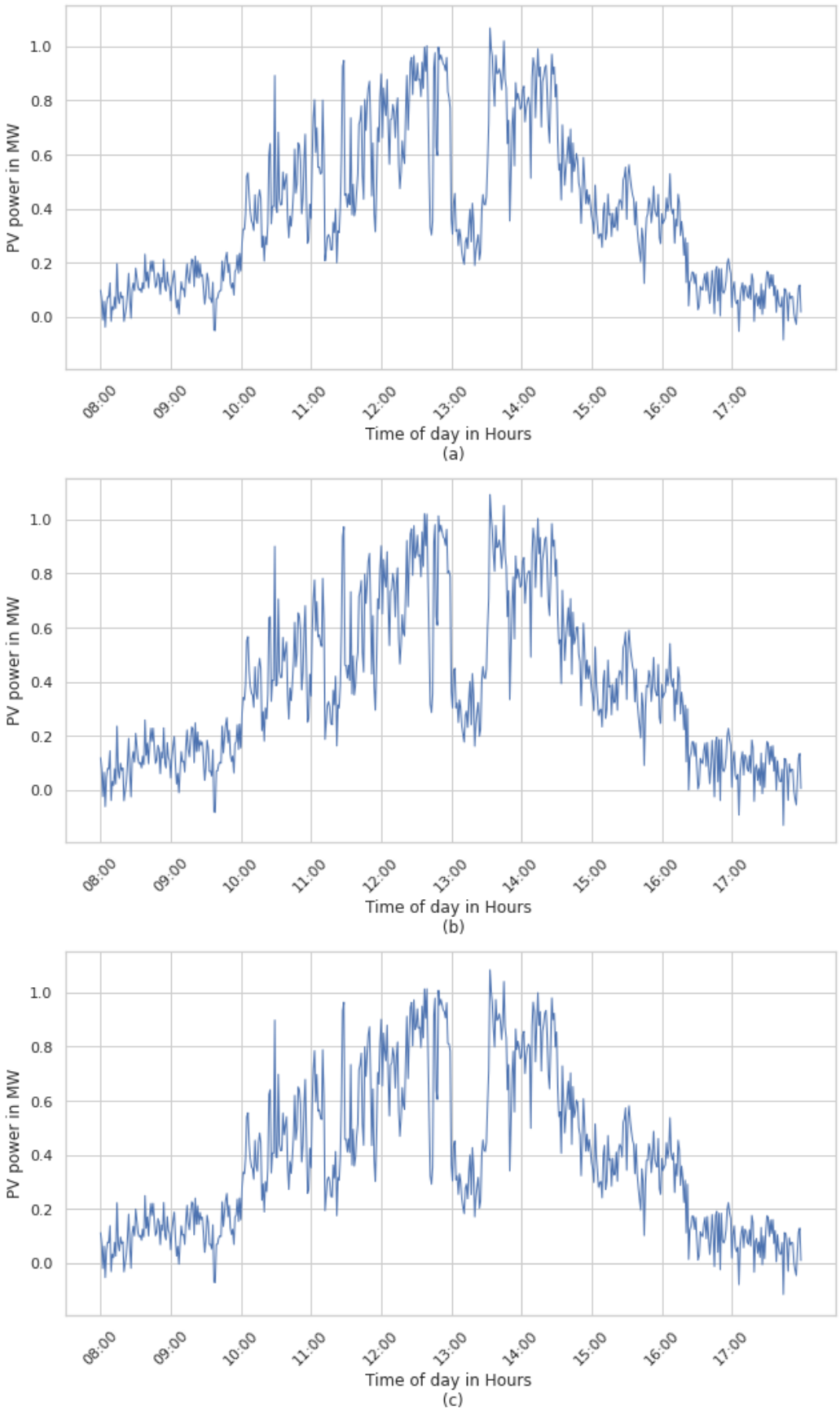


Figure 5.48 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 6 using VWS for a moderately cloudy day

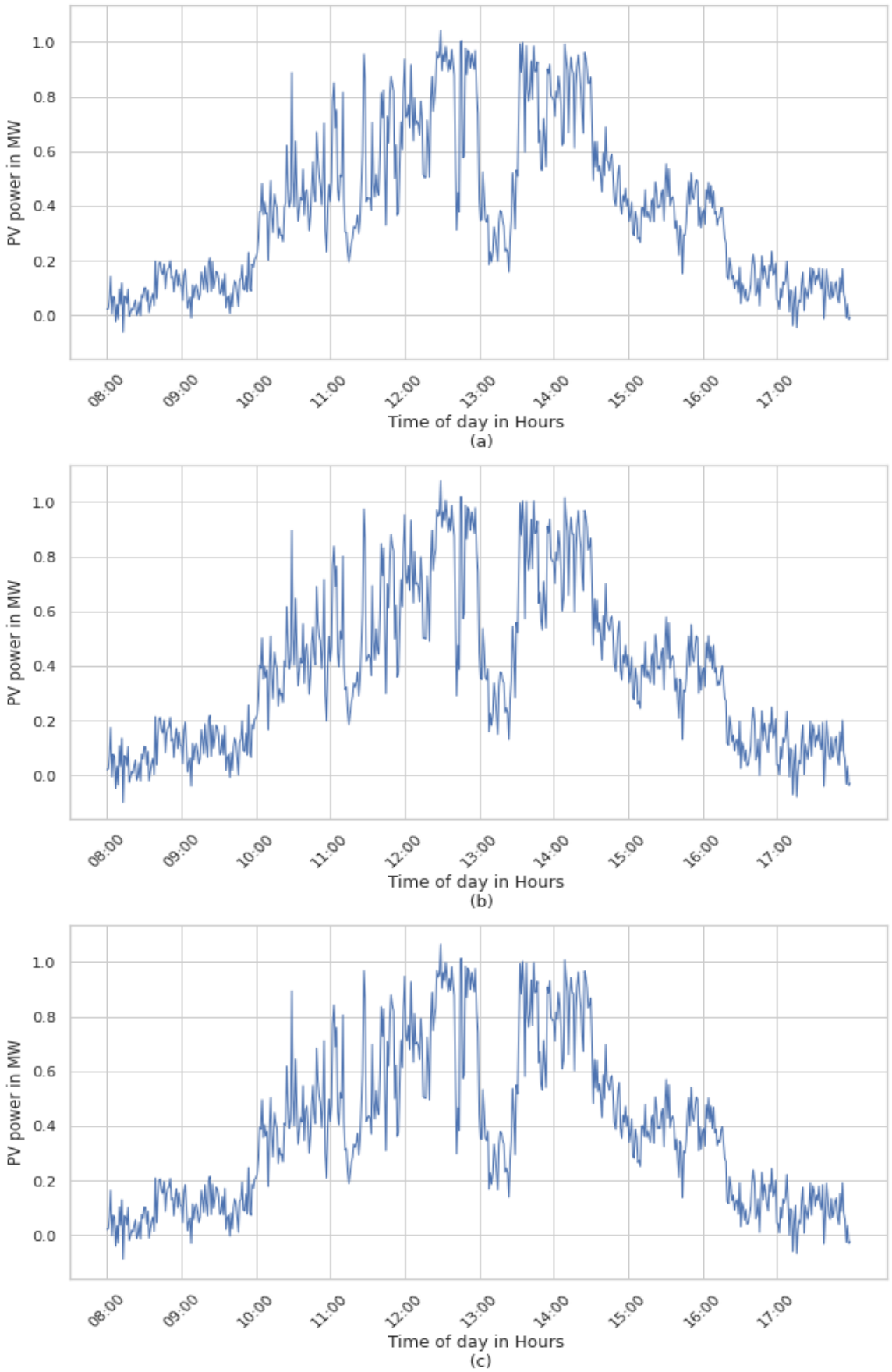


Figure 5.49 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 7 using VWS for a moderately cloudy day

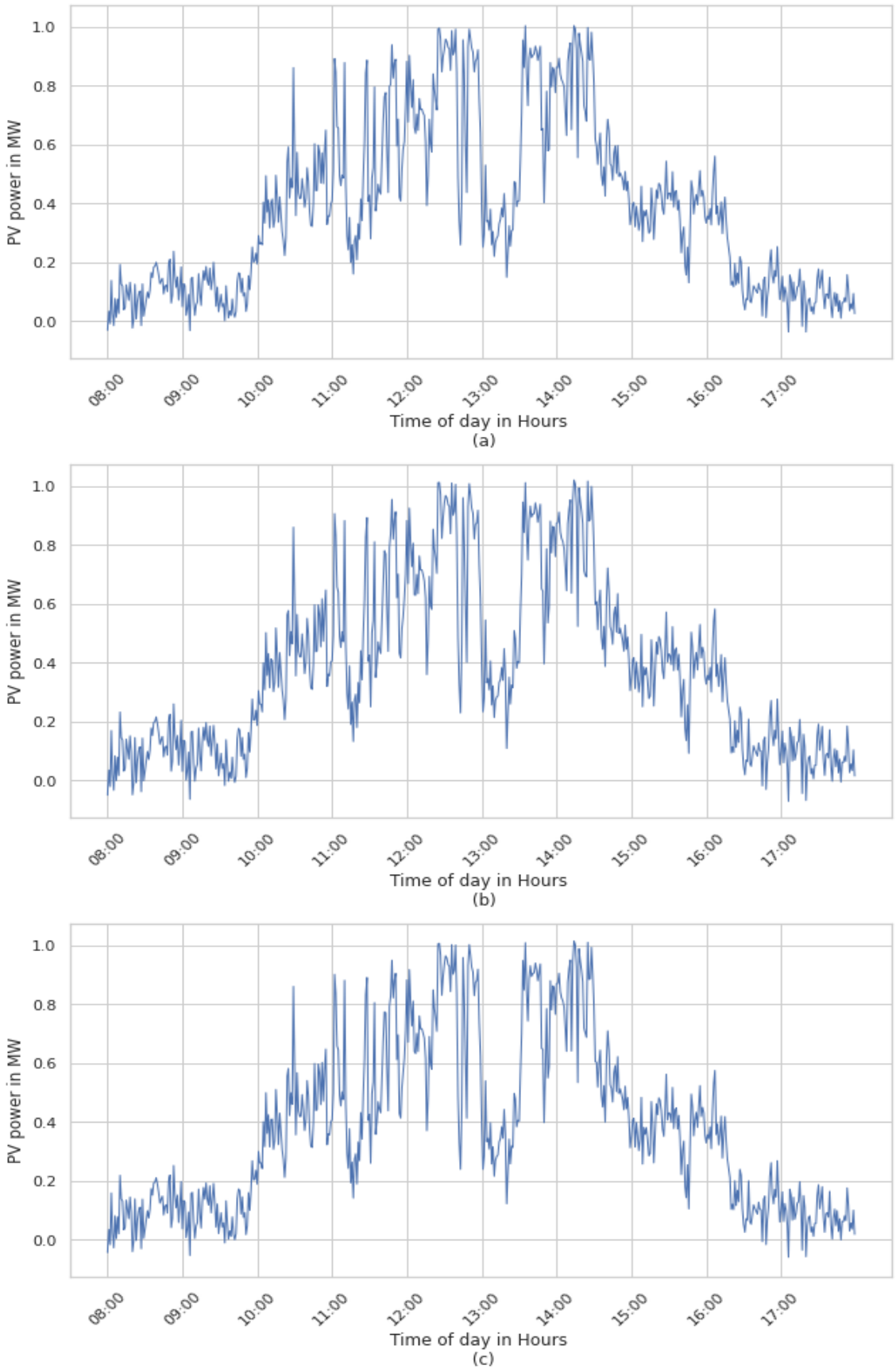


Figure 5.50 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 8 using VWS for a moderately cloudy day

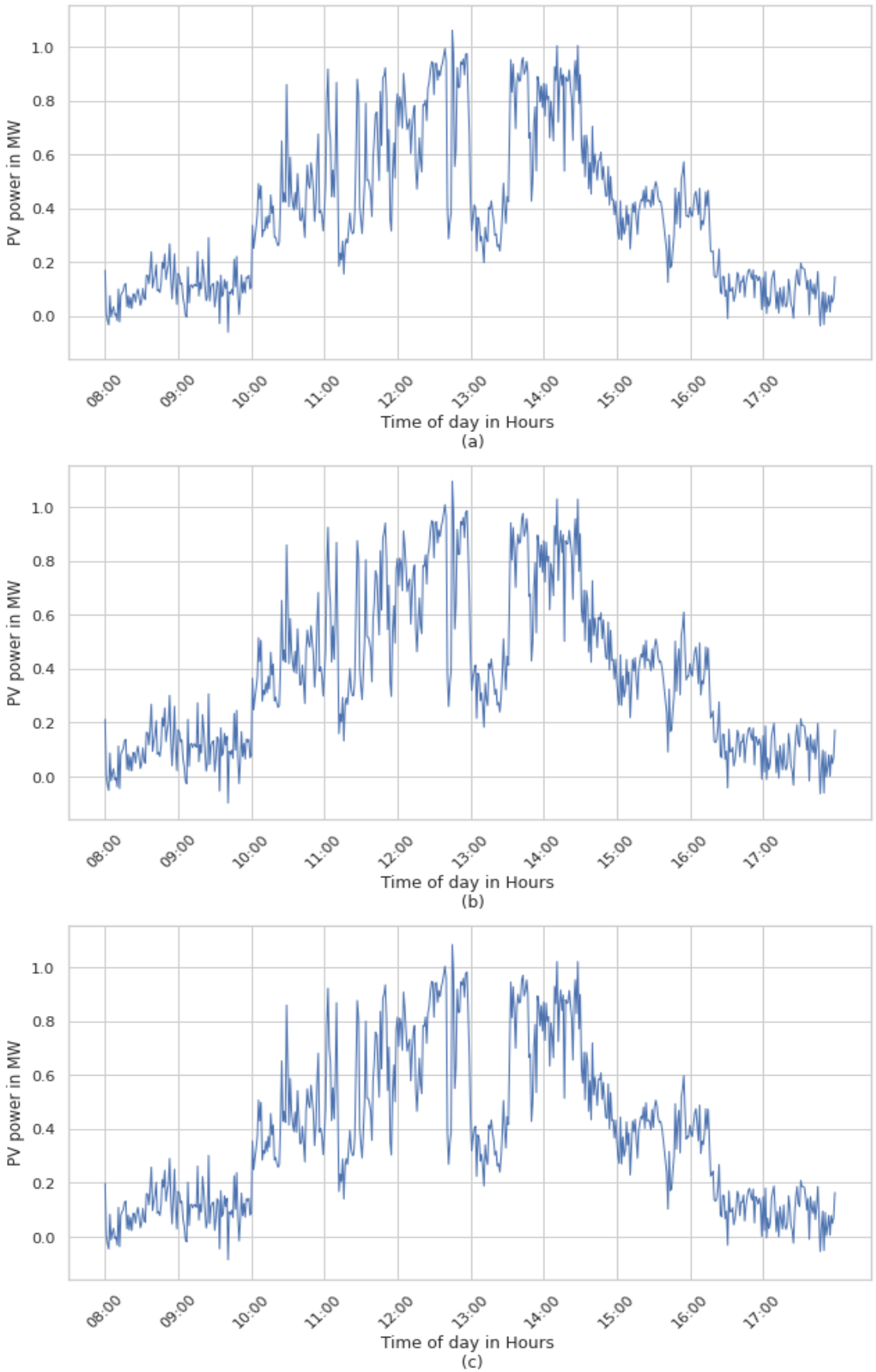


Figure 5.51 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 9 using VWS for a moderately cloudy day

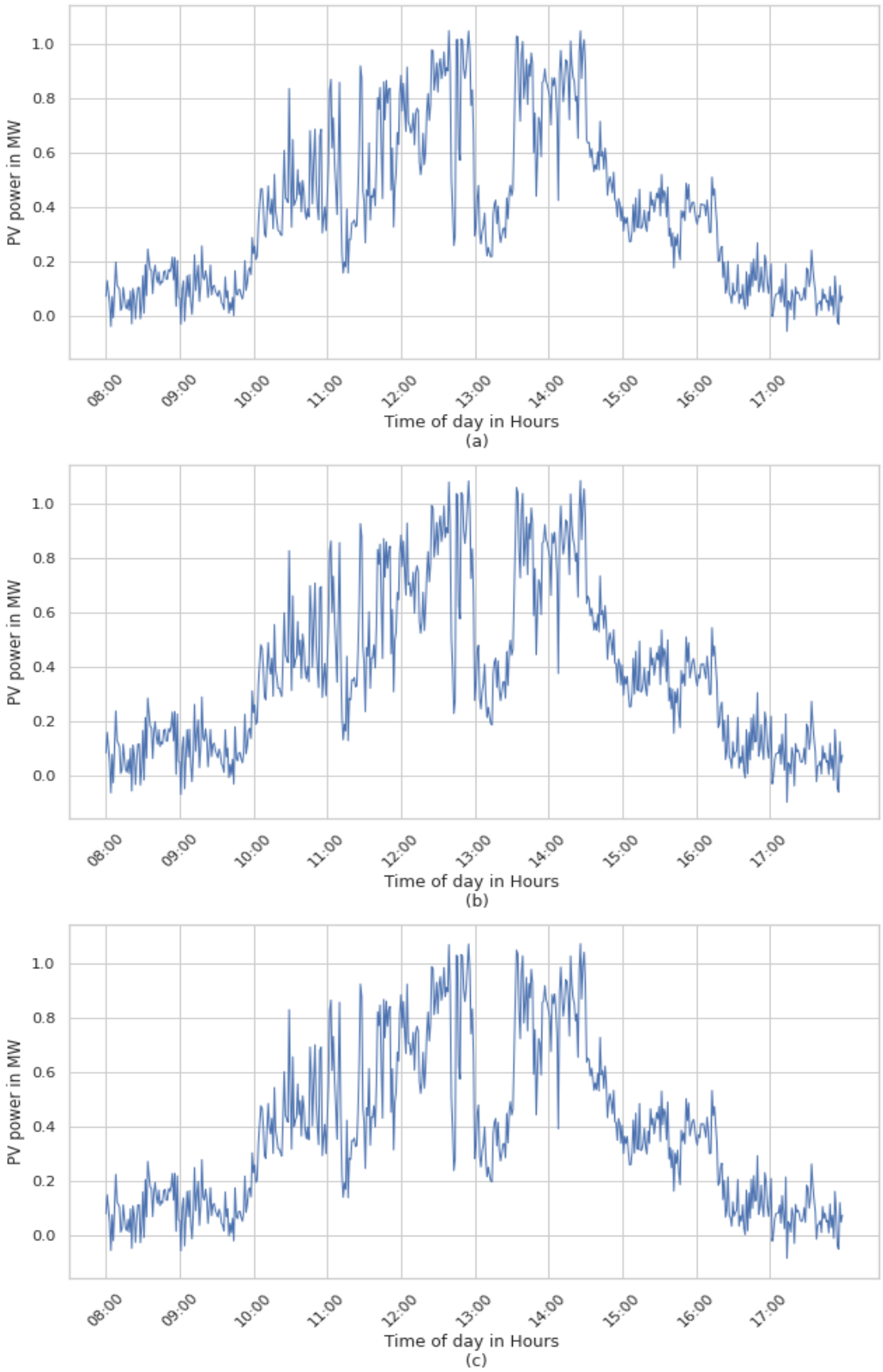


Figure 5.52 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 10 using VWS for a moderately cloudy day

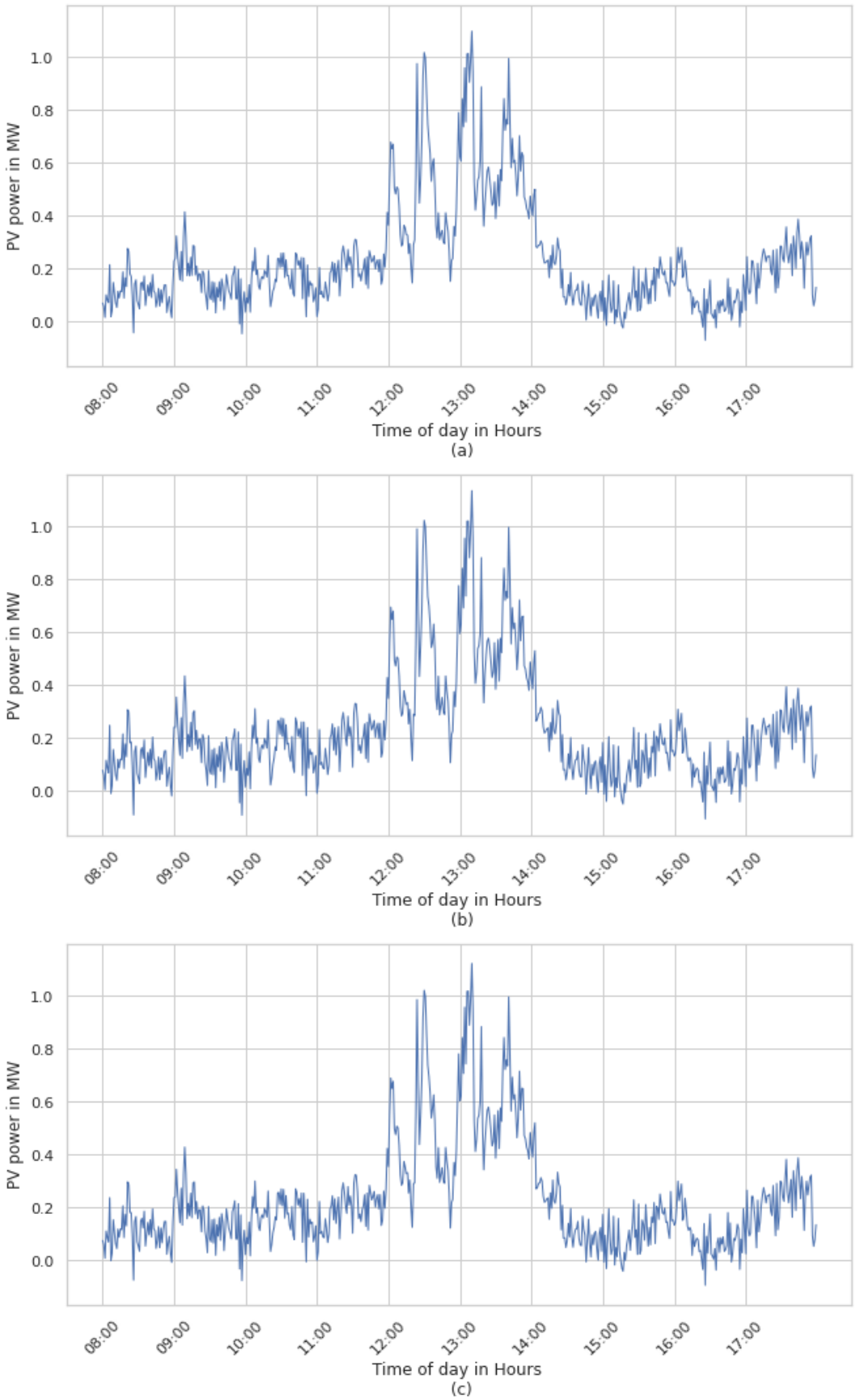


Figure 5.53 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 1 using VWS for a cloudy day

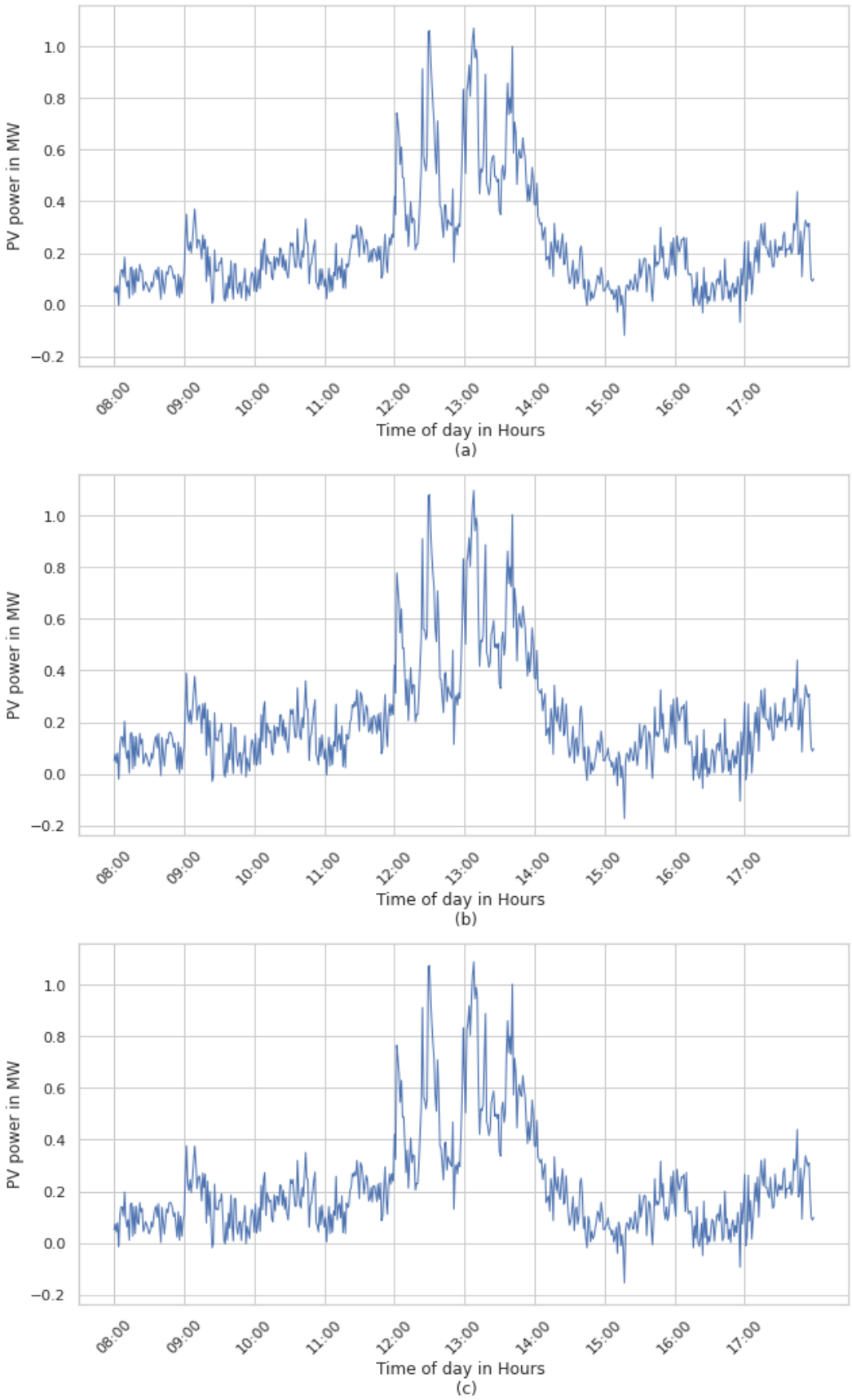


Figure 5.54 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 2 using VWS for a cloudy day

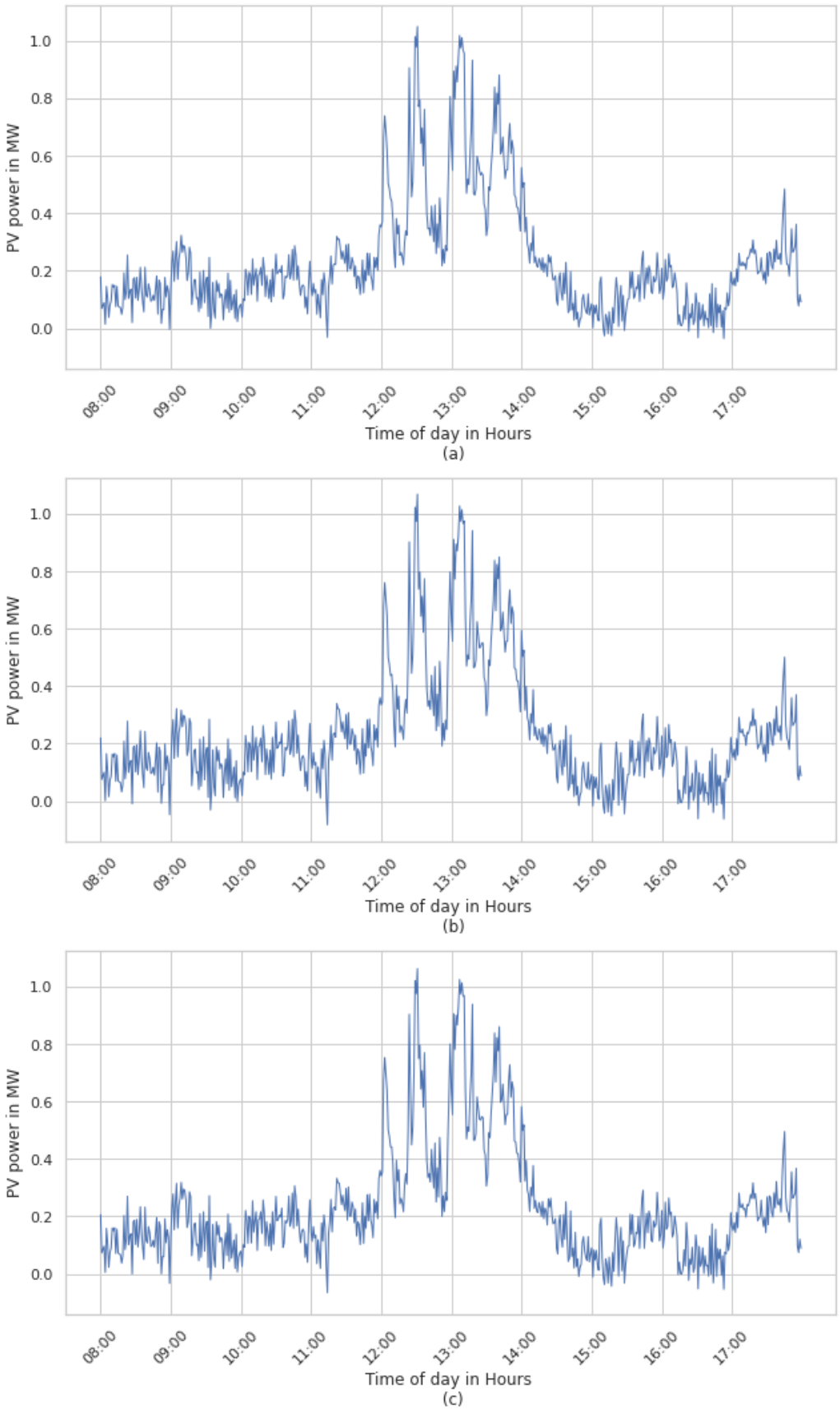


Figure 5.55 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 3 using VWS for a cloudy day

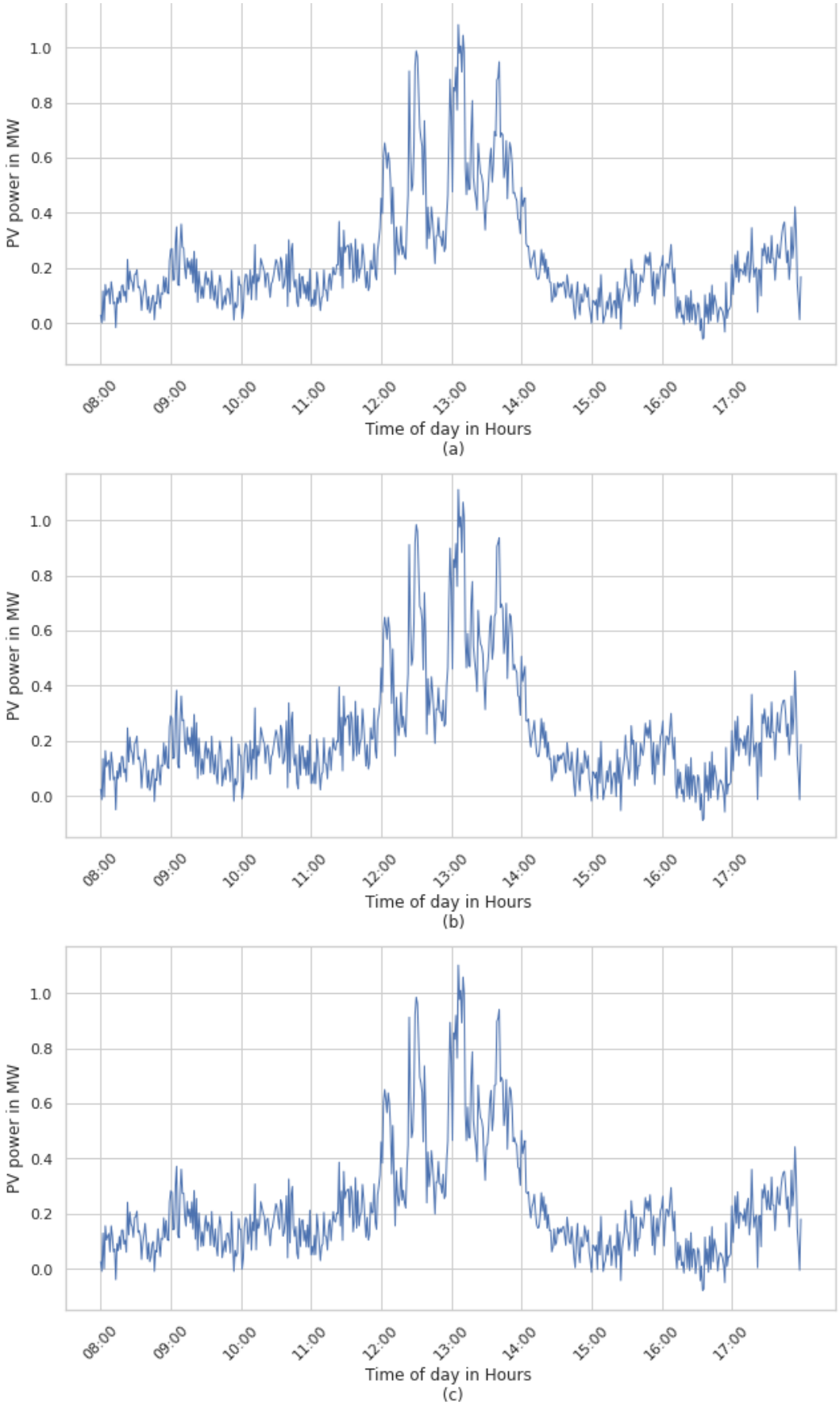


Figure 5.56 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 4 using VWS for a cloudy day

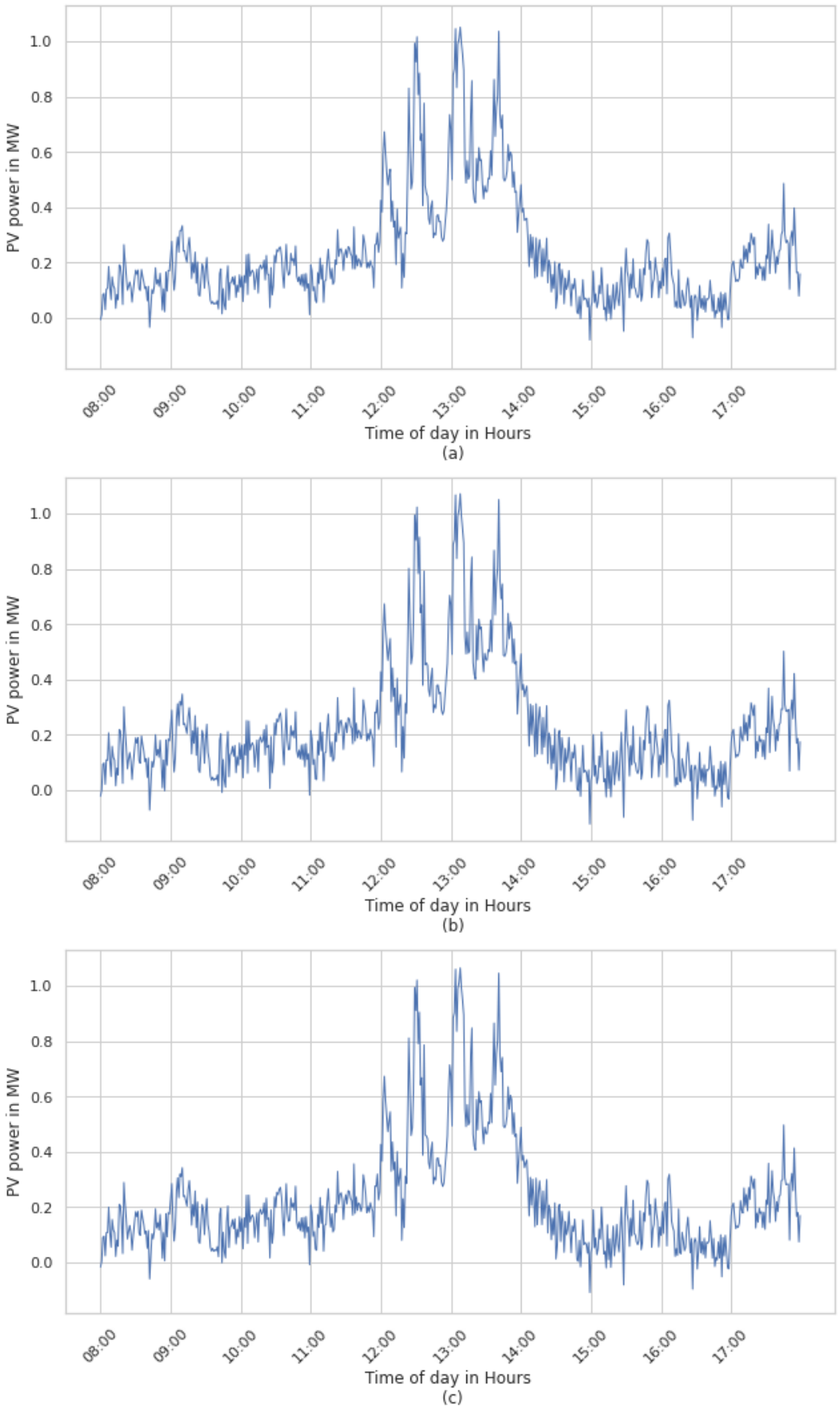


Figure 5.57 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 5 using VWS for a cloudy day

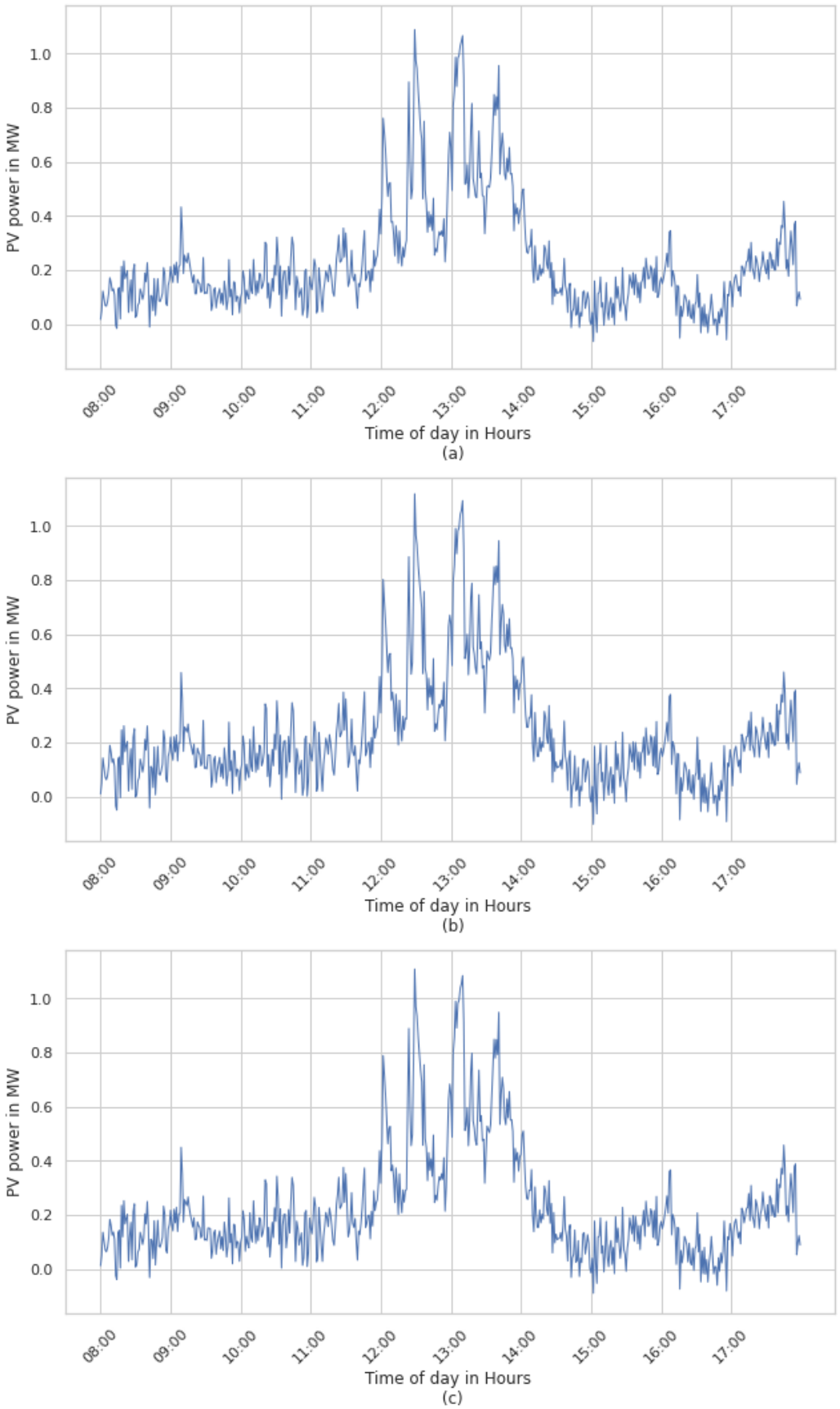
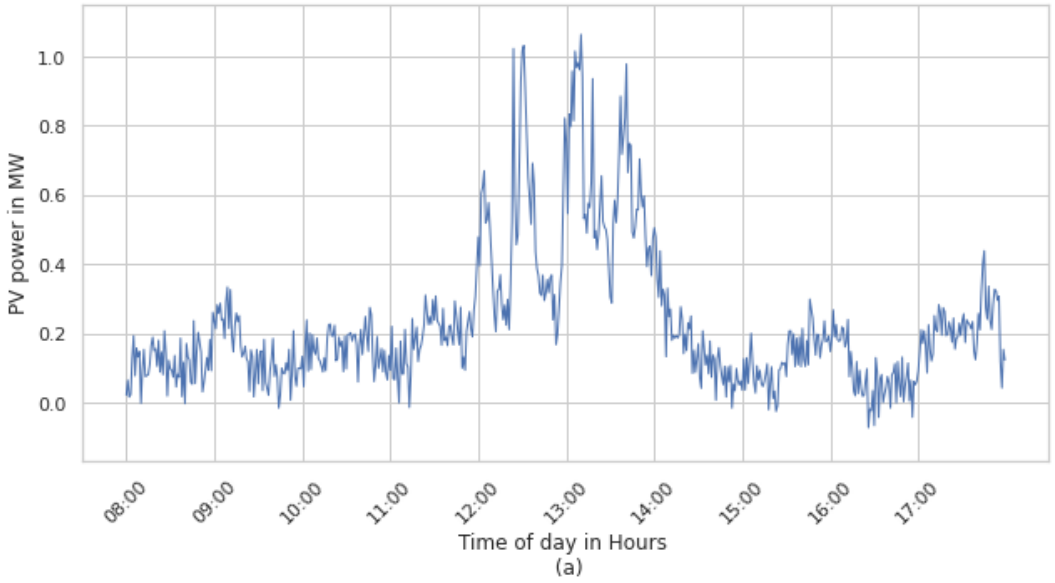
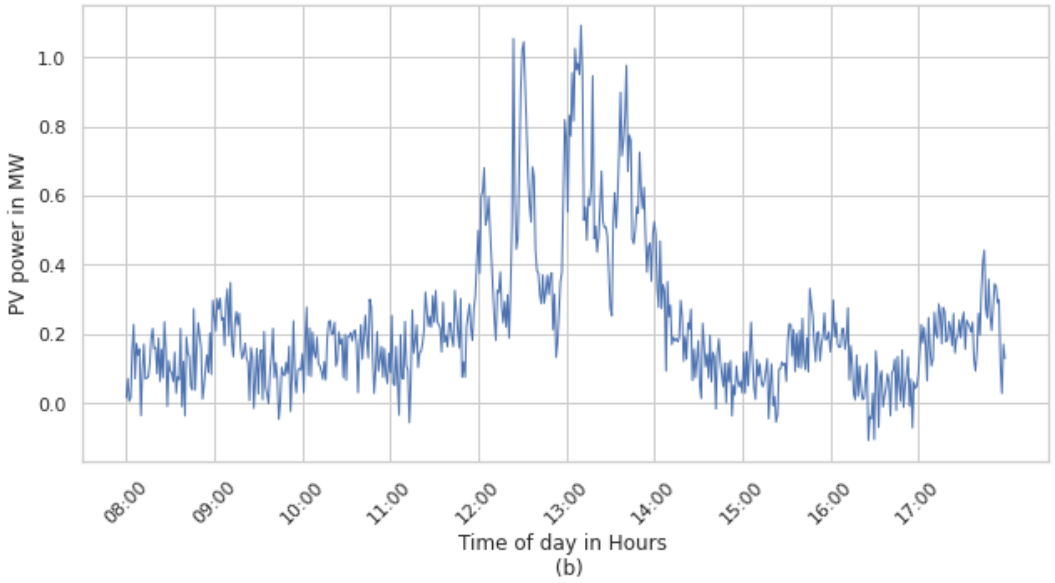


Figure 5.58 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 6 using VWS for a cloudy day





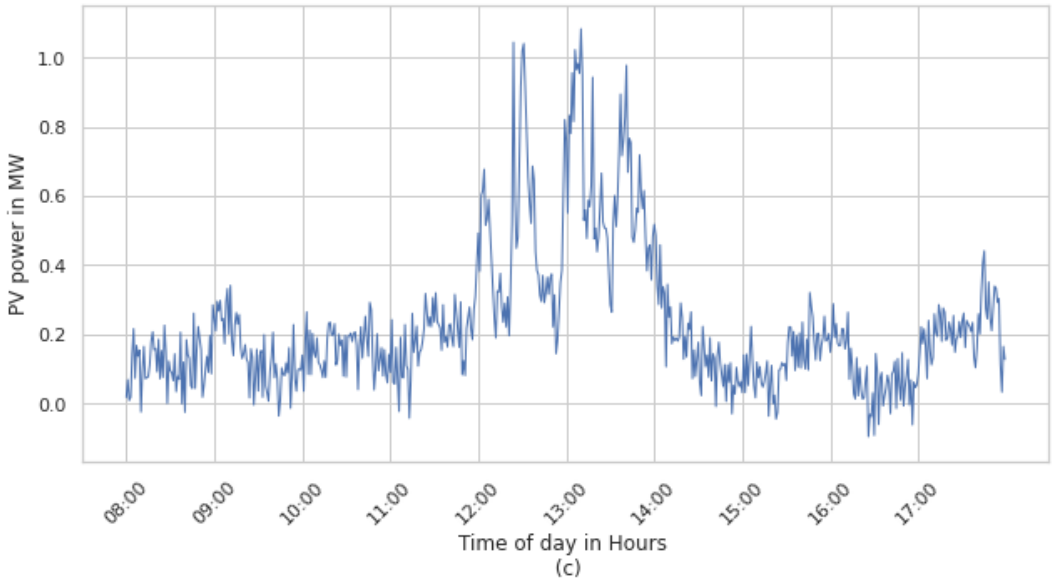


Figure 5.59 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 7 using VWS for a cloudy day

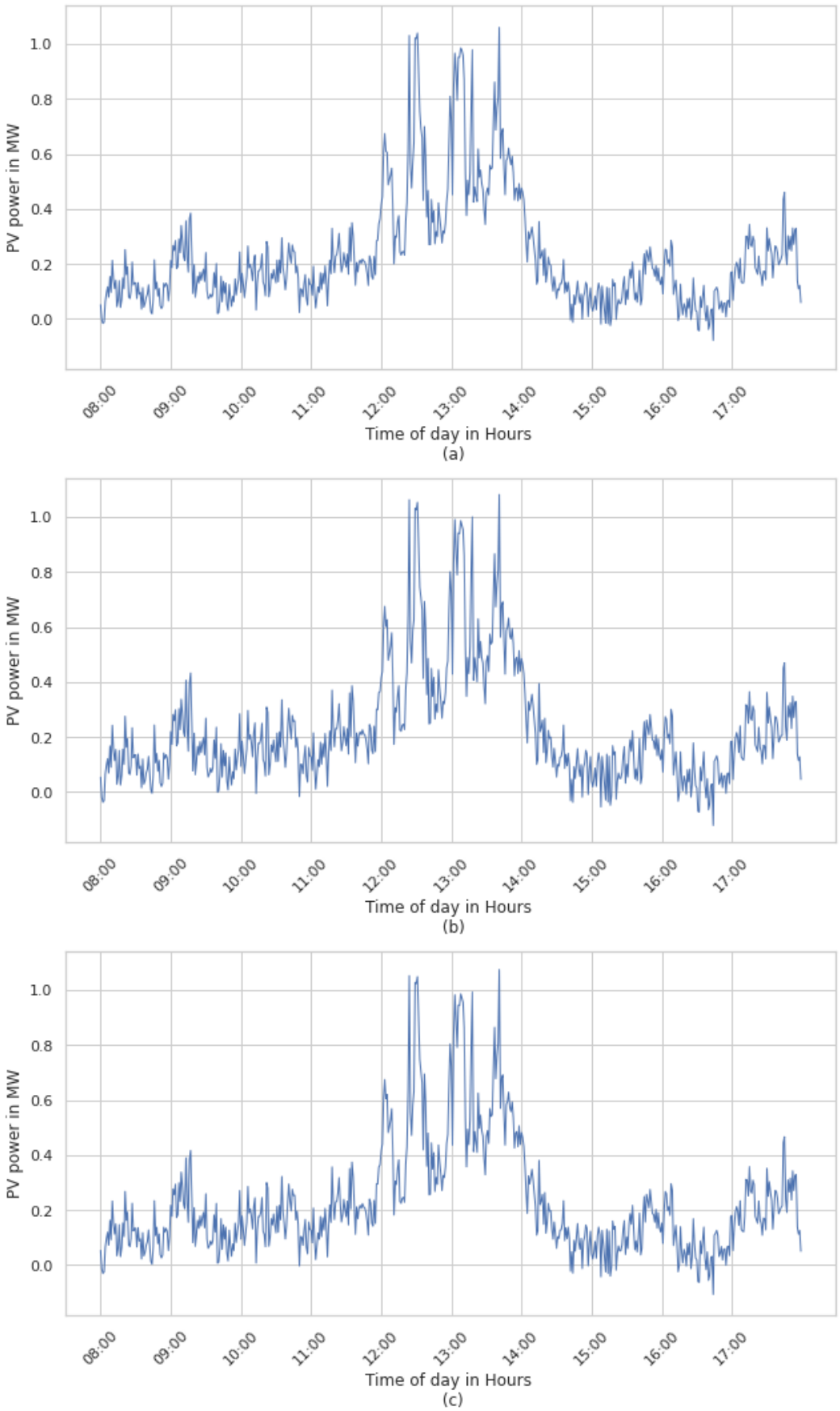


Figure 5.60 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 8 using VWS for a cloudy day

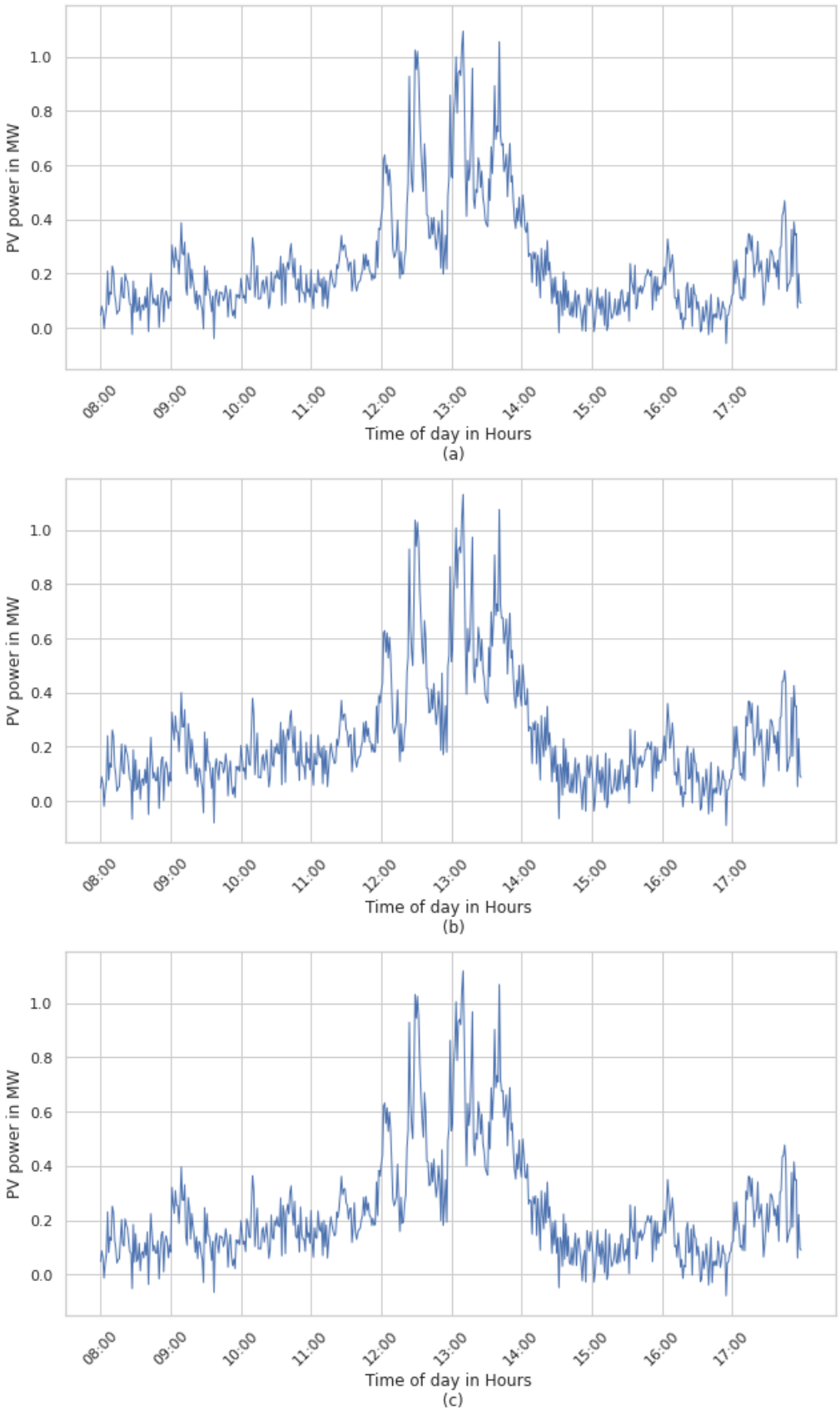


Figure 5.61 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 9 using VWS for a cloudy day

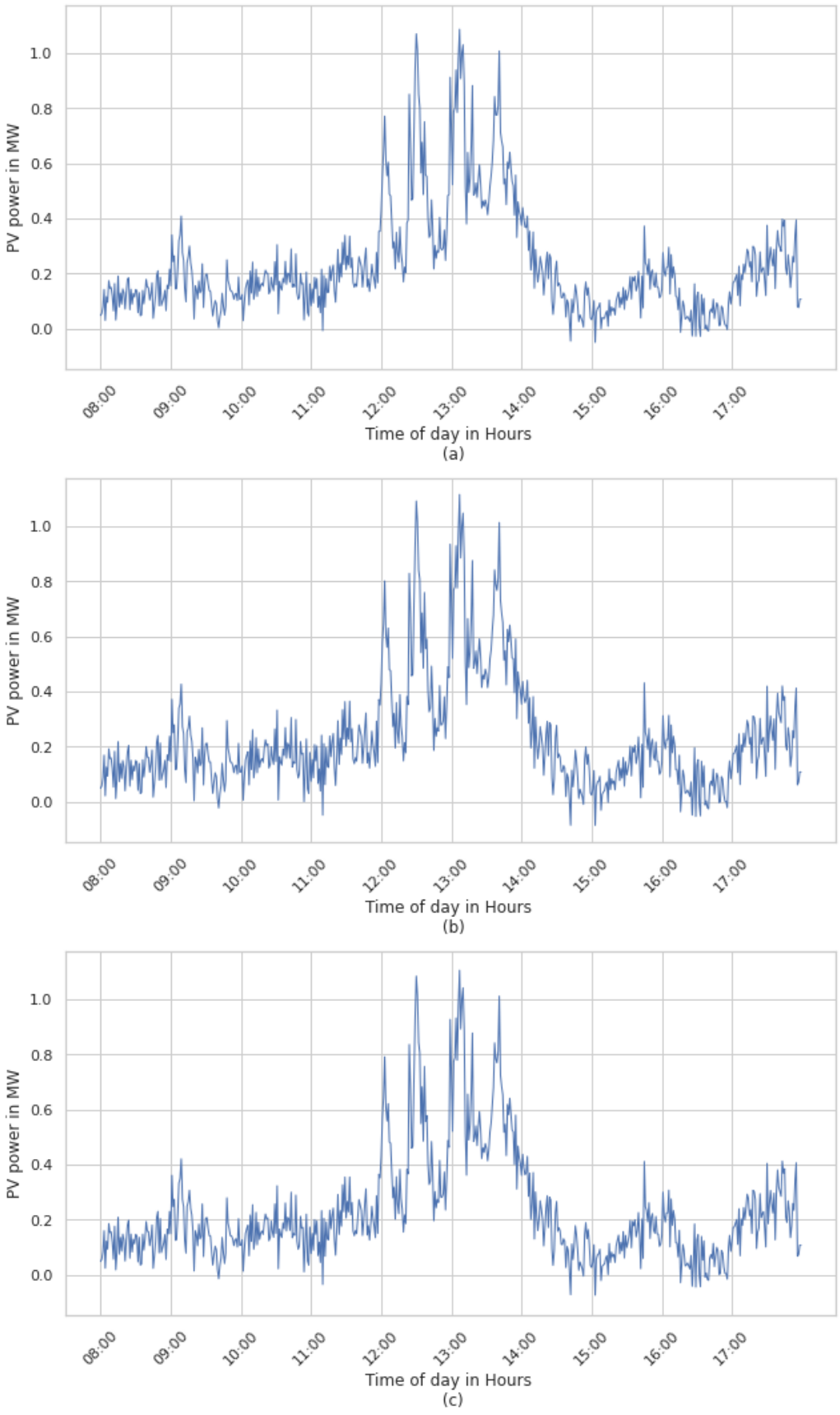


Figure 5.62 Case 2, 60 (a), 120 (b), 180(c) seconds ahead predicted PV power at location 10 using VWS for a cloudy day

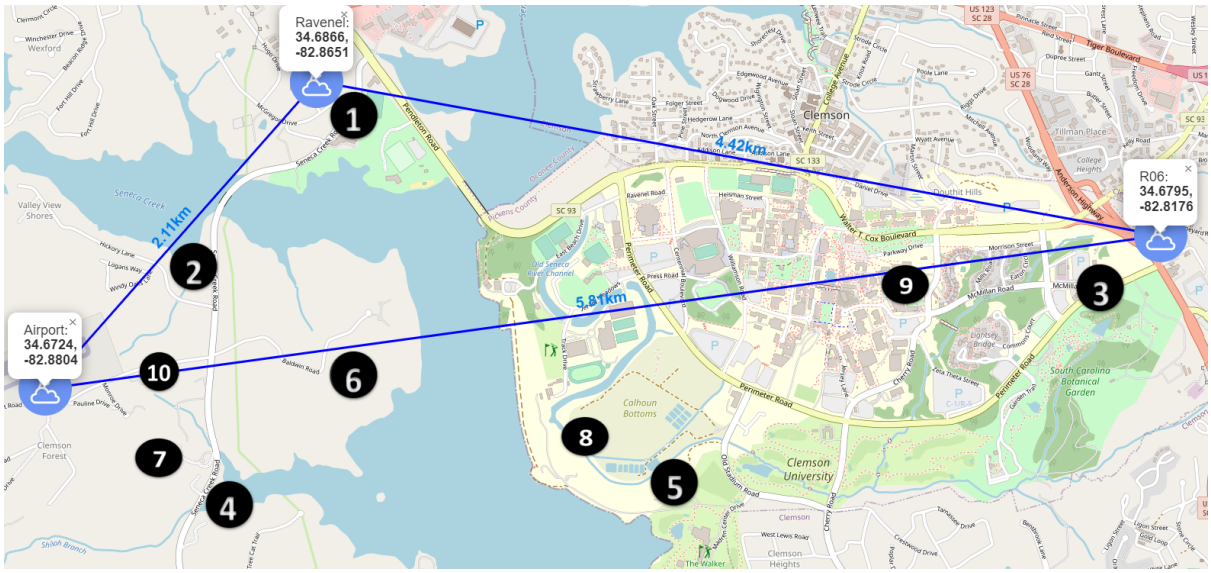


Figure 5.63 Physical, Virtual and Hybrid PV plant forecasting sites

Summary

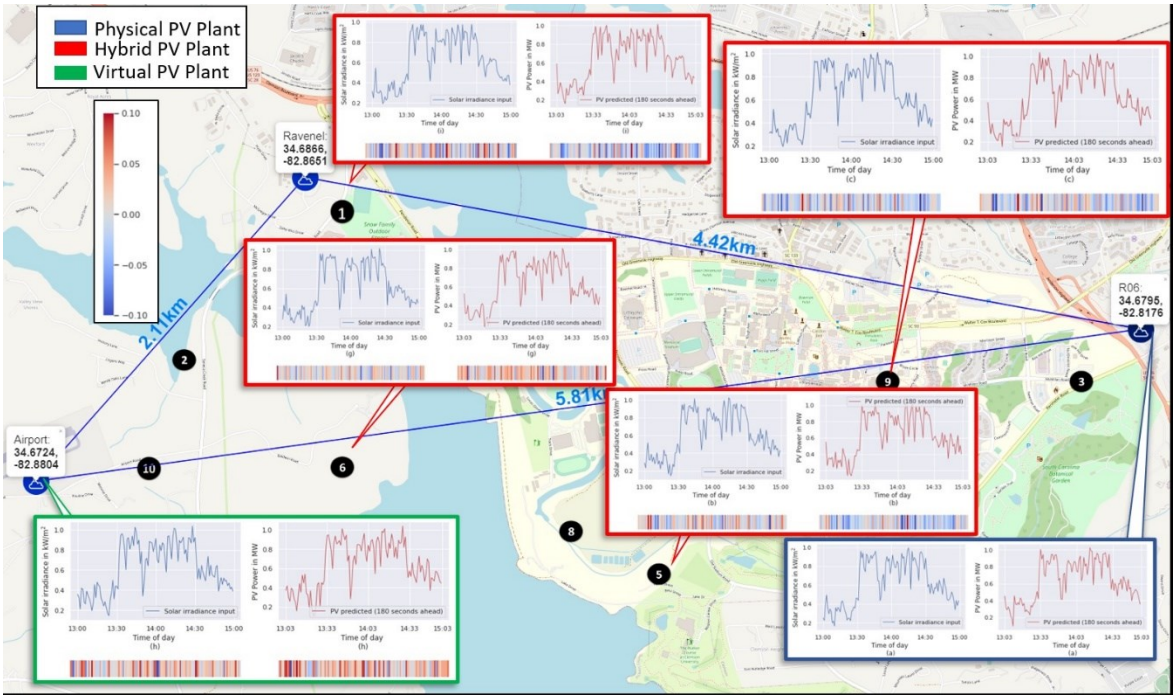


Figure 5.64 Weather Stations and corresponding DPV Sources

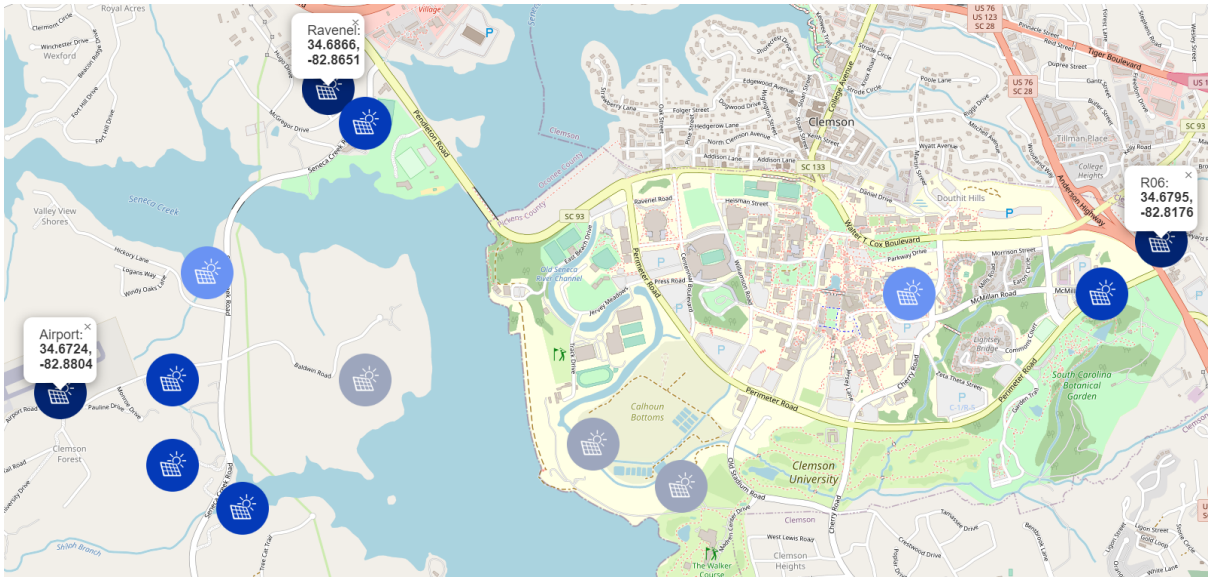


Figure 5.65 PV plants and PV DTs correspondingly graded from light to dark for accuracy of PV estimation and forecasting DTs

10 CDTs of 3 physical weather stations and 10 virtual weather stations were developed using mathematical models and intelligent mutations as seen in Figure 5.64. A similar method can be used to improve visibilities of other DERs sources and loads.

Three cases were examined, where PV sources and measurements are present, where no PVs are present, and where neither PV plants or weather stations are present at the site. Estimation of overall PV production over the region in real-time, considering spatiotemporal and environmental characteristics for user selected locations, provides enhanced observability and situational awareness of PV sources DSO operator or regulator situational awareness. Prediction of overall PV power generation over a region in real-time can additionally be done. Insights generated can be utilized for cloud cover impact mitigation and volt-var optimization, scenario testing and other predictive studies, providing situational intelligence of PV power generation in the area.

CHAPTER SIX
CONCLUSION

Introduction

Data-driven and hybrid DT technology for MEPDS system complexity mitigation, particularly for PV plant integration at the distribution level is necessary. Advanced applications enabling situational awareness and situational intelligence at the MEPDS level are possible through leveraging DT and AI paradigms, in a quick, cost-effective, low complexity manner through available data measurements of weather and PV power outputs at the PV plant sites.

Chapter Summaries

Chapter 1 presented the introduction, problem statement, objectives and chapter-wise contributions of the study and presented a brief, general overview of the thesis

Chapter 2 covered the need for data-driven DTs to truly encapsulate the complexity of system DTs and Asset DTs of systems dependent on stochastic processes such as PV power production for instance. Additionally, the potential of enhancing DTs with AI and utilizing DTs for advanced MEPDS applications to produce cognition in a system was presented

Chapter 3 presented the design and development of a data-driven DT for estimation of PV power at a 1 MW PV plant using multi-layer perceptrons (MLP) and a DT for prediction of PV power at the same site 60,90 and 180 seconds ahead using an echo state network (ESN). The DTs were trained for 72 days worth of historical data collected at the site over the spring and validated against real-time behavior of the plant. These DTs can be further used for condition and alert monitoring, predictive analysis, scenario analysis and for other long-term data collection purposes for future studies.

Chapter 4 presented the design of 10 virtual weather stations producing real-time wind speed and direction, temperature and solar irradiance measurements across a selected region. Historical data, or descriptive analysis was first performed to understand weather station variations of the existing (three) physical weather stations, and further leveraging those inferences to perform calculations such as inverse distance weighting, generate look up tables, and intelligent mutations of actual recorded measurements at the site in real-time.

Chapter 5 showed the utilization of the 10 virtual weather stations and prediction and estimation DTs to generate situational awareness and situational intelligence of PV sources in the distribution system. Various cases, such as prediction and estimation DTs of PV power where a PV plant already exists, at regions where only PWSs exists and at regions where neither PWS or PV plant exists are supported. This enables a DSO, regulator or distribution system planner to understand or predict PV power over a region in real-time, based on their requirements and available equipment. This further enables advanced studies of distribution systems with high PV penetration once integrated into a testbed or RTDS system. The operator can, at a glance, be able to monitor potential behavior of PV sources at different placement locations connected to feeders and understand voltage profiles, distribution system hosting capacity, and system behavior in different loading conditions, along with performing fine-tuned volt-var analysis.

Conclusions of this Thesis

The work can be split into three main focus areas. The first part of the thesis shows the design and development of data-driven DTs using echo state networks and multi-layer perceptrons for PV power estimation and prediction. The second part of this work deals with mathematical and data-driven weather modeling, particularly solar irradiance and temperature

based on wind speed variations to develop virtual weather station and virtual weather station measurements. The third part of the work utilizes combinations of available sources, DT models, and physical, virtual and hybrid stations to show an efficient, scalable means to estimate and forecast PV power production over an area. This supports advanced applications, studies and analyses by enabling situational awareness as well as situational intelligence in the MEPDS.

Future work

Different AI paradigms can be used for DT design along with hybrid-AI methodology, based on system requirements. The fidelity and complexity of the system can be expanded to support new PV plant configurations and sizes.

The modeling of virtual weather stations can be expanded to include new parameters such as elevation, humidity, pollution and dust. AI methodologies can be used to approximate complex functions to generate hybrid weather station DTs and virtual weather station measurements while maintaining a similar approach.

The Physical, Hybrid and Virtual PV plants developed in chapter 5 can be easily integrated into an RTDS model such as RT-Lab or RSCAD in a similar manner to real sources. A TCP/IP connection can be made over Python, sending data into the system. A 34-bus distribution system and network representation can be seen in Figure 6.1. The three cases are as shown.

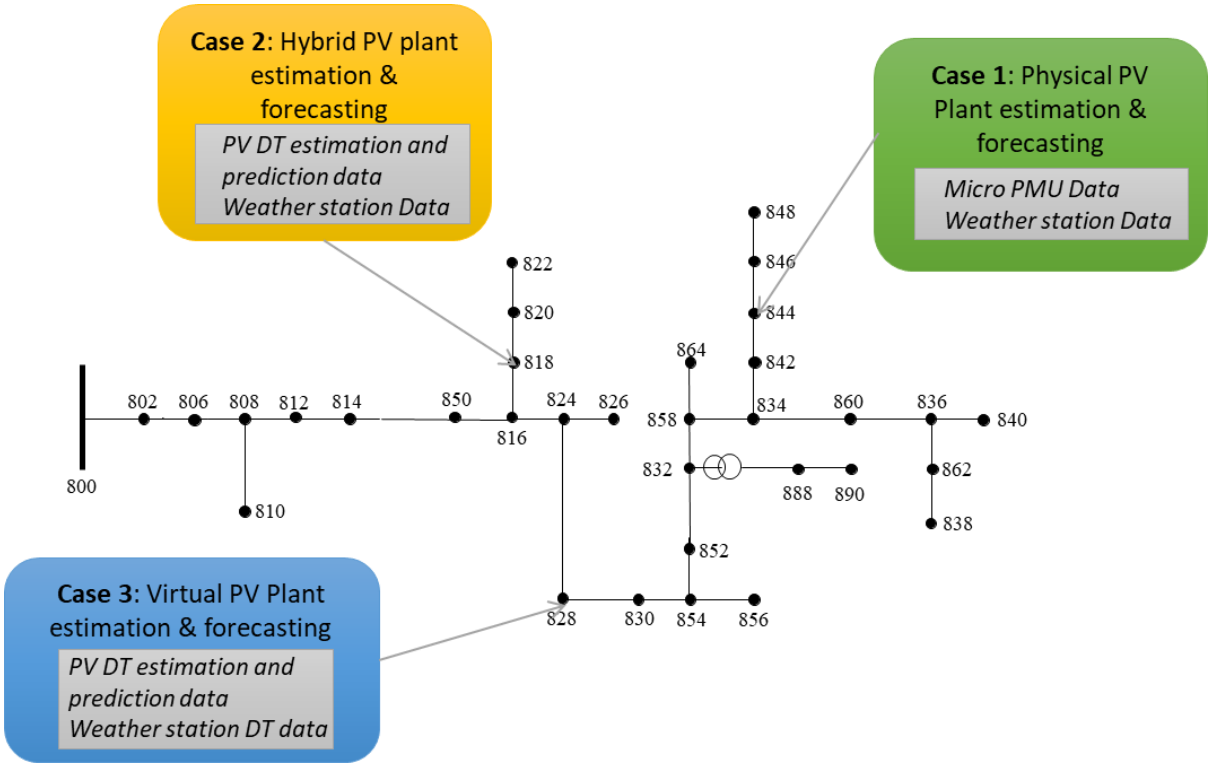


Figure 6.1 Modified IEEE 34 Bus system with DER input

Distribution testbed studies can thereby be done in real-time. The data can also be collected over a long period and integrated with distribution planning tools like OpenDSS or WindMil for time-series simulation supporting, long-term studies.

BIBLIOGRAPHY

[1] "IEEE Standard for Interconnection and Interoperability of Distributed Energy Resources with Associated Electric Power Systems Interfaces," in IEEE Std 1547-2018 (Revision of IEEE Std 1547-2003), vol., no., pp.1-138, 6 April 2018, doi: 10.1109/IEEESTD.2018.8332112.

[2] "Approved Draft Guide to Using IEEE Standard 1547 for Interconnection of Energy Storage Distributed Energy Resources with Electric Power Systems," in IEEE P1547.9/D5.6, May 2022 (Approved Draft) , vol., no., pp.1-83, 22 June 2022.

[3] Cano, Craig. "Ferc order no. 2222: A new day for distributed energy resources." Federal Energy Regulatory Commission (2020).

[4] Ge, Leijiao, et al. "Smart distribution network situation awareness for high-quality operation and maintenance: a brief review." Energies 15.3 (2022): 828.

[5] Chaudhary, Priyanka, and M. Rizwan. "Voltage regulation mitigation techniques in distribution system with high PV penetration: A review." Renewable and Sustainable Energy Reviews 82 (2018): 3279-3287.

[6] Olowu, Temitayo O., et al. "Future challenges and mitigation methods for high photovoltaic penetration: A survey." Energies 11.7 (2018): 1782.

[7] Emmanuel, Michael, and Ramesh Rayudu. "The impact of single-phase grid-connected distributed photovoltaic systems on the distribution network using PQ and PV models." International Journal of Electrical Power & Energy Systems 91 (2017): 20-33.

[8] Horowitz, Kelsey A., et al. "An overview of distributed energy resource (DER) interconnection: Current practices and emerging solutions." (2019).

[9] Zhaoyun, Zhang, and Lv Linjun. "Application status and prospects of digital twin technology in distribution grid." Energy Reports 8 (2022): 14170-14182.

[10] Strezoski, Luka. "Distributed energy resource management systems—DERMS: State of the art and how to move forward." Wiley Interdisciplinary Reviews: Energy and Environment 12.1 (2023): e460.

[11] Dharmawardena, Hasala. "Data-Driven Distributed Modeling, Operation, and Control of Electric Power Distribution Systems." (2022)

[12] Obi, Manasseh, and Robert Bass. "Trends and challenges of grid-connected photovoltaic systems–A review." Renewable and Sustainable Energy Reviews 58 (2016): 1082-1094.

[13] Schauder, Colin. "Impact of FERC 661-A and IEEE 1547 on photovoltaic inverter design." 2011 IEEE Power and Energy Society General Meeting. IEEE, 2011.

[14] Chapagain, Prerak, et al. "Stability impact of IEEE 1547 operational mode changes under high DER penetration in the presence of cyber adversary." 2021 IEEE Green Technologies Conference (GreenTech). IEEE, 2021.

[15] Klyve, Øyvind Sommer, et al. "The value of forecasts for PV power plants operating in the past, present and future Scandinavian energy markets." Solar Energy 255 (2023): 208-221.

[16] Cirjaleanu, Valeriu. "Investigation of Cloud-Effects on Voltage Stability of Distribution Grids with Large Amount of Solar Photovoltaics." (2017).

[17] Evangelopoulos, V. A., Georgilakis, P. S., & Hatziargyriou, N. D. (2016). Optimal operation of smart distribution networks: A review of models, methods and future research. Electric Power Systems Research, 140, 95-106.

[18] E. Glaessgen and D. Stargel, "The digital twin paradigm for future NASA and US Air Force vehicles." In E. Glaessgen and D. Stargel. "The digital twin paradigm for future NASA and US Air Force vehicles." In *53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference 20th AIAA/ASME/AHS adaptive structures conference 14th AIAA*, p. 1818, 2012.

[19] C. Semeraro, M. Lezoche, H. Panetto, and M. Dassisti, "Digital twin paradigm: A systematic literature review." In *Computers in Industry*, *130*, p.103469, 2021

[20] Singh, Maulshree, et al. "Digital twin: Origin to future." Applied System Innovation 4.2 (2021): 36.

[21] He, Rui, et al. "Data-driven digital twin technology for optimized control in process systems." ISA transactions 95 (2019): 221-234.

[22] Zheng, Xiaochen, Jinzhi Lu, and Dimitris Kiritsis. "The emergence of cognitive digital twin: vision, challenges and opportunities." International Journal of Production Research 60.24 (2022): 7610-7632.

[23] Sifat, Md Mhamud Hussen, et al. "Towards electric digital twin grid: Technology and framework review." Energy and AI (2022): 100213.

[24] Sleiti, Ahmad K., Jayanta S. Kapat, and Ladislav Vesely. "Digital twin in energy industry: Proposed robust digital twin for power plant and other complex capital-intensive large engineering systems." Energy Reports 8 (2022): 3704-3726.

[25] Fernandes, Sabryna V., et al. "Digital twin concept developing on an electrical distribution system—An application case." Energies 15.8 (2022): 2836.

[26] M. Grieves and J. Vickers. "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems." In *Transdisciplinary perspectives on complex systems*, Springer, Cham, pp. 85-113, 2017.

[27] Gauce, Diana, et al. "Application of digital twin in medium-voltage overhead distribution network inspection." Remote Sensing 15.2 (2023): 489.

[28] Moutis, Panayiotis, and Omid Alizadeh-Mousavi. "Digital twin of distribution power transformer for real-time monitoring of medium voltage from low voltage measurements." IEEE Transactions on Power Delivery 36.4 (2020): 1952-1963.

[29] Yalçin, Tolga, et al. "Exploiting Digitalization of Solar PV Plants Using Machine Learning: Digital Twin Concept for Operation." Energies 16.13 (2023): 5044.

[30] Tang, Xueyong, et al. "Creating multi-timescale digital twin models for regional multiple energy systems on CloudPSS." 2020 IEEE Sustainable Power and Energy Conference (iSPEC). IEEE, 2020.

[31] Guzman Razo, Dorian Esteban, et al. "A genetic algorithm approach as a self-learning and optimization tool for PV power simulation and digital twinning." Energies 13.24 (2020): 6712.

[32] Wang, Kangshi, et al. "Digital twin based maximum power point estimation for photovoltaic systems." 2022 19th International SoC Design Conference (ISOCC). IEEE, 2022.

[33] Jain, Palak, et al. "A digital twin approach for fault diagnosis in distributed photovoltaic systems." IEEE Transactions on Power Electronics 35.1 (2019): 940-956.

[34] Zhou, Mike, Jianfeng Yan, and Donghao Feng. "Digital twin framework and its application to power grid online analysis." CSEE Journal of Power and Energy Systems 5.3 (2019): 391-398.

[35] Nguyen, Van Hoa, et al. "Digital twin integrated power-hardware-in-the-loop for the assessment of distributed renewable energy resources." Electrical Engineering (2021): 1-12.

[36] Gupta, Sarthak, Spyros Chatzivasileiadis, and Vassilis Kekatos. "Deep Learning for Optimal Volt/VAR Control using Distributed Energy Resources." arXiv preprint arXiv:2211.09557 (2022).

[37] Gui, Yonghao, et al. "Automatic voltage regulation application for PV inverters in low-voltage distribution grids–A digital twin approach." International Journal of Electrical Power & Energy Systems 149 (2023): 109022.

[38] Kharlamova, Nina, Chresten Træholt, and Seyedmostafa Hashemi. "A digital twin of battery energy storage systems providing frequency regulation." 2022 IEEE International Systems Conference (SysCon). IEEE, 2022.

[39] H. Dharmawardena and G. K. Venayagamoorthy, "Distributed Volt-Var Curve Optimization Using a Cellular Computational Network Representation of an Electric Power Distribution System." In *Energies,* 15(12), p.4438, 2022.

[40] Yao, Yiyun, et al. "Coordinated inverter control to increase dynamic PV hosting capacity: A real-time optimal power flow approach." IEEE Systems Journal 16.2 (2021): 1933-1944.

[41] R. Darbali-Zamora, J. Johnson, A. Summers, C. B. Jones, C. Hansen, and C. Showalter. "State estimation-based distributed energy resource optimization for distribution voltage regulation in telemetry-sparse environments using a real-time digital twin." In *Energies,* 14(3), p.774, 2021.

[42] S. V. Fernandes., D. V. João, B. B. Cardoso, M. A. Martins, and E. G. Carvalho. "Digital Twin Concept Developing on an Electrical Distribution System—An Application Case." In *Energies,* 15(8), p.2836, 2022

[43] P.T. Baboli, D. Babazadeh, and D.R.K. Bowatte. "Measurement-based modeling of smart grid dynamics: A digital twin approach." In *2020 10th Smart Grid Conference (SGC)*, pp. 1-6. IEEE, 2020.

[44] W. Danilczyk, Y.L. Sun, and H. He. "Smart grid anomaly detection using a deep learning digital twin." In *2020 52nd North American Power Symposium (NAPS)*, pp. 1-6. IEEE, 2021.

[45] T. Wagner, G. Mehlmann, and M. Richter, "Application of the digital twin concept for a distribution network." In *PESS 2020; IEEE Power and Energy Student Summit*, pp. 1-5. VDE, 2020.

[46] Jain, Palak, et al. "A digital twin approach for fault diagnosis in distributed photovoltaic systems." IEEE Transactions on Power Electronics 35.1 (2019): 940-956.

[47] Sehrawat, Neha, Sahil Vashisht, and Amritpal Singh. "Solar irradiance forecasting models using machine learning techniques and digital twin: A case study with comparison." International Journal of Intelligent Networks (2023).

[48] Z. Zhu, K.W. Chan, S. Bu, B. Zhou, and S. Xia, “Real-Time interaction of active distribution network and virtual microgrids: Market paradigm and data-driven stakeholder behavior analysis.” In *Applied Energy*, *297*, p.117107, 2021.

[49] Saad, Ahmed, Samy Faddel, and Osama Mohammed. "IoT-based digital twin for energy cyber-physical systems: design and implementation." Energies 13.18 (2020): 4762.

[50] Yao, Chuantao, et al. "A Data-driven method for adaptive resource requirement allocation via probabilistic solar load and market forecasting utilizing digital twin." Solar Energy 250 (2023): 368-376.

[51] Song, Xinya, et al. "Application of digital twin assistant-system in state estimation for inverter dominated grid." 2020 55th International Universities Power Engineering Conference (UPEC). IEEE, 2020.

[52] R. Li, W. Wang, and M. Xia, “Cooperative planning of active distribution system with renewable energy sources and energy storage systems.” In *IEEE access*, *6*, pp.5916-5926, 2017.

[53] You, Minglei, et al. "Digital twins based day-ahead integrated energy system scheduling under load and renewable energy uncertainties." Applied Energy 305 (2022): 117899.

[54] K.Z. Oo, K.M. Lin, and T.N. Aung, “Particle Swarm Optimization based optimal reactive power dispatch for power distribution network with distributed generation.” In *International Journal of Energy and Power Engineering*, *6*(4), pp.53-60, 2017.

[55] C. Semeraro, M. Lezoche, H. Panetto, and M. Dassisti, "Digital twin paradigm: A systematic literature review." In *Computers in Industry*, *130*, p.103469, 2021

[56] Howell, Shaun, et al. "Towards the next generation of smart grids: Semantic and holonic multi-agent management of distributed energy resources." Renewable and Sustainable Energy Reviews 77 (2017): 193-214.

[57] Pathiravasam, Chirath, et al. "Spatio-temporal distributed solar irradiance and temperature forecasting." 2020 International Joint Conference on Neural Networks (IJCNN). IEEE, 2020.

[58] SEIA, SEIA. "Solar industry research data." SEIA. https://www. seia. org/solar-industry-research-data (2021).

[59] Yang, Dazhi. "A guideline to solar forecasting research practice: Reproducible, operational, probabilistic or physically-based, ensemble, and skill (ROPES)." Journal of Renewable and Sustainable Energy 11.2 (2019).

[60] Guzman Razo, Dorian Esteban, et al. "A genetic algorithm approach as a self-learning and optimization tool for PV power simulation and digital twinning." Energies 13.24 (2020): 6712.

[61] Yang, Heng, and Weisong Wang. "Prediction of photovoltaic power generation based on LSTM and transfer learning digital twin." Journal of Physics: Conference Series. Vol. 2467. No. 1. IOP Publishing, 2023.

[62] Haupt, Sue Ellen, et al. "Combining artificial intelligence with physics-based methods for probabilistic renewable energy forecasting." Energies 13.8 (2020): 1979.

[63] Madhiarasan, M., and S. N. Deepa. "A new hybridized optimization algorithm to optimize echo state network for application in solar irradiance and wind speed forecasting." World Applied Sciences Journal 35.4 (2017): 596-614.

[64] Venayagamoorthy, Ganesh K., and Bashyal Shishir. "Effects of spectral radius and settling time in the performance of echo state networks." Neural Networks 22.7 (2009): 861-863.

[65] Jayawardene, Iroshani, and Ganesh K. Venayagamoorthy. "Comparison of adaptive neuro-fuzzy inference systems and echo state networks for PV power prediction." Procedia Computer Science 53 (2015): 92-102.

[66] JL Kafka, MA Miller. “A climatology of solar irradiance and its controls across the United States: Implications for solar panel orientation.” Renewable Energy 135 (2019): 897-907.

[67] C. Pathiravasam, P. Arunagirinathan, I. Jayawardene, G. K. Venayagamoorthy and Y. Wang, ”Spatio-Temporal Distributed Solar Irradiance and Temperature Forecasting,” 2020 International Joint Conference on Neural Networks (IJCNN), 2020, pp. 1-8, doi:10.1109/IJCNN48605.2020.9206936.

[68] Morris, C. J. G., Ian Simmonds, and Neil Plummer. "Quantification of the influences of wind and cloud on the nocturnal urban heat island of a large city." *Journal of Applied Meteorology and Climatology* 40.2 (2001): 169-182.

[69] Bett, Philip E., and Hazel E. Thornton. "The climatological relationships between wind and solar energy supply in Britain." *Renewable Energy* 87 (2016): 96-110.

[70] A. Ebrahimi.“"Challenges of developing a digital twin model of renewable energy generators”" 2019 IEEE 2^{8t}h international symposium on industrial electronics (ISIE). IEEE, 2019.

[71] L Massel, N Shchukin, A Cybikov, “"Digital twin development of a solar power plant”" E3S Web of Conferences. Vol. 289. EDP Sciences, 2021.

[72] Yang, Heng, and Weisong Wang. "Prediction of photovoltaic power generation based on LSTM and transfer learning digital twin." Journal of Physics: Conference Series. Vol. 2467. No. 1. IOP Publishing, 2023.

[73] Fang, Zhi, et al. "State estimation for situational awareness of active distribution system with photovoltaic power plants." IEEE Transactions on Smart Grid 12.1 (2020): 239-250.

[74] L Massel, N Shchukin, A Cybikov, "Digital twin development of a solar power plant." E3S Web of Conferences. Vol. 289. EDP Sciences, 2021.

BIOGRAPHY

Deborah George (Student Member, IEEE) received her bachelor's degree in Electrical and Electronics Engineering from BMS Institute of Technology, India in 2018. She enrolled in the Master's degree in Electrical Engineering, focus Intelligent Systems, program offered by Clemson University, South Carolina in 2021. She is a Research Assistant for the Real Time Power and Intelligent Systems Lab (RTPIS), and for VIPR-GS, a research partnership between the Ground Vehicle Systems Center and Clemson University. Her research interests include Digital Twin technology and AI for grid modernization and distribution automation as well as models-based systems engineering for EVs.