COMPUTATIONAL INTELLIGENCE-BASED PHOTOVOLTAIC SYSTEM PERFORMANCE MODELING IN SNOW CONDITIONS

By:

Behzad Hashemi

Thesis Presented to

Département d'informatique et d'ingénierie

Université du Québec en Outaouais (UQO)

In partial fulfillment of the requirements for the degree of Doctor of Philosophy Ph.D.

June 2023

Board of Examiners

This thesis has been evaluated by the following board of examiners:

Thesis Supervisor:	Prof. Shamsodin Taheri Département d'informatique et d'ingénierie, Université du Québec en Outaouais.
Thesis Co-Supervisor:	Prof. Ana-Maria Cretu Département d'informatique et d'ingénierie, Université du Québec en Outaouais.
Committee President:	Prof. Mohand Said Allili Département d'informatique et d'ingénierie, Université du Québec en Outaouais.
Internal Examiner:	Prof. Ahmed Lakhssassi Département d'informatique et d'ingénierie, Université du Québec en Outaouais.
External Examiner:	Prof. Hasan Mehrjerdi Department of electrical and computer engineering, Royal Military College of Canada.

This thesis was presented and defended in the presence of the board of examiners and the public on June 8, 2023, at the département d'informatique et d'ingénierie.

ABSTRACT

Photovoltaic (PV) energy is the direct conversion of sunlight into electricity using solar cells. It is a rapidly growing industry with a wide range of applications, including residential, commercial, and utility-scale projects. With the global will to address environmental concerns caused by fossil fuel consumption in the energy sector and make a paradigm shift to renewable energies, PV technology is getting more popular even in northern regions where solar radiation intensity is lower. One of the other challenges of PV installations in northern areas is the longer duration of winter with snowfall events that cause snow accumulation on the PV panels and a significant reduction in their performance. Hence, analyzing the performance of PV systems in snow conditions is of significant importance.

Unlike conventional power generation units, PV systems are known to be intermittent energy resources. It means that PV systems cannot always consistently produce the desired amount of energy at all hours of the day. This originates from the fact that the intensity of solar radiation varies between sunrise and sunset due to the earth's spin. While this variation is simply predictable, there are some other factors attenuating solar insolation reaching the surface of PV panels that cannot be easily predicted. These factors include shading by cloud coverage, nearby vegetation, buildings, etc., and soling by dust, pollen, snow, etc. Moreover, the efficiency of solar cells in photo-current production and the dissipation of power in electric components of the PV system affect the electric power delivered to the grid. For this reason, evaluating the performance of PV systems in terms of different types of power losses can be very insightful for PV systems' operators. On the other hand, having available an accurate model for the power generation of PV systems compatible with the local environmental conditions can be very beneficial for power grid operators to maintain optimal operation of the grid.

Integrating the effects of the aforementioned factors into empirical PV power generation models is challenging, demands full knowledge of underlying physical-electrical principles, and results in complex equations that require detailed measurements for each factor. On the other hand, recent developments in computational intelligence techniques have provided a new approach to data-driven modeling. These techniques can inherently learn from large amounts of historical data of PV systems and predict their performance in various conditions. This can help to develop accurate models of PV

systems' performance relying on widely available meteorological parameters. In this thesis, a systematic approach to detailed power loss calculation for PV systems in snow-prone regions is first proposed. The power losses are then modeled using computational intelligence techniques. Moreover, the challenge of PV power prediction in snow conditions is addressed using computational intelligence techniques. Finally, novel short-term PV power forecasting models compatible with snow conditions are proposed and their applications in two power grid-related energy management problems are investigated.

This research could be regarded as an important contribution to the development of reliable models of PV systems in snow conditions and further studies of the applications of the models in optimal energy management of power grids.

RÉSUMÉ

L'énergie photovoltaïque (PV) est la conversion directe de la lumière du soleil en électricité à l'aide de cellules solaires. Il s'agit d'une industrie en croissance rapide avec un large éventail d'applications, y compris des projets résidentiels, commerciaux et à grande échelle. Avec la volonté mondiale de répondre aux préoccupations environnementales causées par la consommation de fossiles combustibles dans le secteur de l'énergie et de faire un changement de paradigme vers les énergies renouvelables, la technologie PV devient de plus en plus populaire même dans les régions du nord où l'intensité du rayonnement solaire est plus faible. L'un des autres défis des installations photovoltaïques dans les régions nordiques est la longue durée de l'hiver avec des épisodes de chutes de neige qui provoquent une accumulation de neige sur les panneaux photovoltaïques et une réduction significative de leurs performances. Par conséquent, l'analyse des performances des systèmes PV dans des conditions de neige est d'une importance significative.

Contrairement aux unités de production d'électricité conventionnelles, les systèmes PV sont connus pour être des ressources énergétiques intermittentes. Cela signifie que les systèmes photovoltaïques ne peuvent pas toujours produire de manière constante la quantité d'énergie souhaitée à toutes les heures de la journée. Cela provient du fait que l'intensité du rayonnement solaire varie entre le lever et le coucher du soleil en raison de la rotation de la Terre. Bien que cette variation soit simplement prévisible, il existe d'autres facteurs qui atténuent l'insolation solaire atteignant la surface des panneaux photovoltaïques et qui ne peuvent pas être facilement prédits. Ces facteurs comprennent l'ombrage par la couverture nuageuse, la végétation à proximité, les bâtiments, etc., et la couverture par la poussière, le pollen, la neige, etc. De plus, l'efficacité des cellules solaires dans la production de photo-courant et la dissipation de puissance dans les composants électriques du PV système affecte la puissance électrique fournie au réseau. Pour cette raison, l'évaluation des performances des systèmes PV en termes de différents types de pertes de puissance peut être très utile pour les opérateurs de systèmes PV. D'autre part, se disposer d'un modèle précis pour la production d'électricité des systèmes PV compatible avec les conditions environnementales locales peut être très bénéfique pour les opérateurs de réseaux électriques afin de maintenir un fonctionnement optimal du réseau.

L'intégration des effets des facteurs mentionnés ci-dessus dans les modèles empiriques de production d'énergie PV est un défi, exige une connaissance complète des principes physico-électriques sous-jacents et aboutit finalement à des équations complexes qui nécessitent des mesures détaillées pour chaque facteur. D'autre part, les développements récents des techniques d'intelligence computationnelle ont fourni une nouvelle approche de la modélisation basée sur les données. Ces techniques peuvent intrinsèquement apprendre à partir de grandes quantités de données historiques de systèmes PV et prédire leurs performances dans diverses conditions. Cela peut aider à développer des modèles précis des performances des systèmes PV en s'appuyant sur des paramètres météorologiques largement disponibles. Dans ce projet de recherche, une approche systématique du calcul détaillé de la perte de puissance pour les systèmes PV dans les régions sujettes à la neige est d'abord proposée. Les pertes de puissance sont ensuite modélisées à l'aide de techniques d'intelligence computationnelle. De plus, le défi de la prédiction de la puissance PV dans des conditions de neige est relevé à l'aide de techniques d'intelligence computationnelle. Enfin, de nouveaux modèles de prévision de puissance PV à court terme compatibles avec les conditions de neige sont proposés et leurs applications dans deux problèmes de gestion de l'énergie liés au réseau électrique sont étudiées.

Cette recherche peut être considérée comme une contribution importante au développement de modèles fiables de systèmes PV dans des conditions de neige et à d'autres études sur les applications des modèles dans la gestion optimale de l'énergie des réseaux électriques.

ACKNOWLEDGEMENT

This thesis was done under the supervision of Prof. Shamsodin Taheri and the co-supervision of Prof. Ana-Maria Cretu. First of all, I would like to express my deepest gratitude to my supervisors for their unwavering support, patience, open-mindedness, guidance, and encouragement throughout my Ph.D. journey. Their expertise and knowledge have been instrumental in shaping my research and my future career and endeavors.

I am also grateful to my committee members, Prof. Mohand Said Allili, Prof. Ahmed Lakhssassi, and Prof. Hasan Mehrjerdi, for their valuable insights and constructive feedback throughout the process.

Finally, I would like to thank my parents and my brother for their unwavering love and support during my Ph.D. studies. Without their encouragement and support, this journey would not have been possible.

DEDICATION

This thesis is dedicated to

My family and friends

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network.
AUC	Area Under Curve.
CUSUM	Cumulative Sum.
CVaR	Conditional Value-at-Risk.
DAM	Day-Ahead Energy Market.
DAMCPs	Day-Ahead Energy Market Clearing Prices.
DHI	Diffuse Horizontal Irradiance.
DNI	Direct Normal Irradiance.
DSO	Distribution System Operator.
EMD	Empirical Mode Decomposition.
EMS	Energy Management System.
GAF	Gate Activation Function.
GASVM	Genetic Algorithm-Based Support Vector Machine.
GBT	Gradient Boosting Tree.
GHI	Global Horizontal Irradiance.
LR	Learning Rate.
LSE	Load Serving Entity.
LSTM	Long Short-Term Memory.
MAE	Mean Absolute Error.
MBE	Mean Bias Error.
MILP	Mixed-Integer Linear Programming.
MPC	Model Predictive Control.
MPPT	Maximum Power Point Tracking.
MSE	Mean Squared Error.
NHU	Number of Hidden Units.
NREL	National Renewable Energy Laboratory.
NWP	Numerical Weather Prediction.
NYISO	New York Independent System Operator.
PC	Principal Component.
PCA	Principal Component Analysis.
PCSs	Public Charging Stations.
PDF	Probability Density Function.
PEV	Plug-In Electric Vehicle.
PM	Particulate Matter.
POA	Plane of Array.
PR	Performance Ratio.
PV	Photovoltaic.
PVPF	Photovoltaic Power Forecasting.

RERs	Renewable Energy Resources.
RMSE	Root Mean Squared Error.
ROC	Receiver Operating Characteristic.
RTC	Real-Time Cost of DSO.
RTM	Real-Time Energy Market.
RTMCPs	Real-Time Energy Market Clearing Prices.
SAF	State Activation Function.
SoC	State of Charge.
STC	Standard Test Condition.
SV	Sophisticated Verification.
SVM	Support Vector Machine.
SVR	Support Vector Regression.
TCC	Total Charging Cost of PEVs.
TOU	Time of Use.
V2G	Vehicle-to-Grid.
WPP	Wind Power Producer.
XGboost	Extreme Gradient Boosting.

CHAPTER 1

INTRODUCTION

1.1 Overview

Over the past few years, the increasing concerns about global warming and its devastating effects on the environment have resulted in a paradigm shift in the energy sector. The fossil fuel-based electricity sector, which is responsible for almost 25% of greenhouse gas (GHG) emissions globally [1], is facing a major transformation towards renewable energy resources. According to the report of the international energy agency (IEA), the share of renewables in the global electricity generation jumped to 29% in 2021, up from 27% in 2019 [2]. China, the U.S., Brazil, India, Germany, Japan, and Canada have been the seven leading countries in the installed renewable energy capacity in 2021 [3]. Among different types of renewable energies, photovoltaic (PV) energy has been one of the fastest-growing renewable energy resources with an increasing share in the world electricity generation, growing from almost 0% in 2010 to 3.4% in 2021 [2]. The main drives of such a rapid share have been the significant drop in the price of PV modules (from $106 per watt in 1976 to $0.38 per watt in 2019 [4]), the considerable improvement in the efficiency of the PV modules (from 18% in 1990 to 24.4% in 2020 for silicon-based modules [5]), low maintenance costs, strong persistence, and high reliability.

One challenge of the integration of PV systems into electric grids originates from the intermittency of the power generated by these systems. Unlike conventional power generation units, the output power of PV systems is not controllable and may vary based on the weather conditions during the day. Besides the other meteorological factors, such as ambient temperature, the available PV energy highly depends on the amount of solar radiation reaching the PV panels. The power output of a PV system can be significantly affected by any factor that causes a blockage for irradiance reaching the surface of the panels and the amount of the power reduction is proportional to the intensity of this blockage. Cloud coverage in high altitudes and soiling/shading factors close to the panels, such as shading by vegetation or nearby buildings and soling by dust or snow in snow-prone regions, are examples of such blockage. On the other hand, the power output

of a PV system depends on the efficiency of the overall energy conversion system and the amount of power loss in its different components. Considering all the aforementioned complications and the factors affecting PV power generation, having an accurate model for PV systems is of crucial importance in optimal decision-making by PV system owners and power grid operators.

As the first study of this thesis, computational intelligence-based modeling of PV systems' performance will be presented. This will be performed by proposing a methodology for calculating different types of power losses in PV systems based on historical datasets of electrical and meteorological parameters and developing several models for each type of power loss using various computational intelligence techniques. In the second study, computational intelligence-based modeling of PV systems' power generation in snow conditions will be discussed by proposing several snow-cover prediction models and a novel approach for improving PV power prediction accuracy for snow-covered panels. As the third study, PV power forecasting (PVPF) in snow conditions and its applications in the decision-making problems of power grids will be presented. Two PVPF models will be proposed: one requires external weather forecasts as inputs while the other is independent of such exogenous data. The application of these models will be investigated in two different power grid scheduling problems.

This thesis provides a methodology for a comprehensive analysis of PV systems' performance which can help PV system owners to have a better understanding and a clear perspective on the short- and long-term operations of the system. Moreover, the models proposed in this thesis can be used as the basis for developing PV systems' design tool packages or be integrated into the power systems' operation and management platforms.

1.2 Motivation of Research

PV energy is generated by the conversion of sunlight into electricity using semiconducting materials, with silicon being used in the vast majority of today's PV cells. PV potential is an important metric that represents the expected lifetime average electricity production (in kWh) produced per kilowatt of installed PV capacity in a location and significantly depends on the intensity of solar irradiance reaching that location [6]. Although the PV potential in higher latitudes is lower compared to regions close to the equator, a major part of the PV power capacity is installed in countries

located in these regions, such as Japan, Germany, the United Kingdom, France, and Canada [7]. As an example, the installed PV capacity in Canada jumped from 281 MW in 2010 to 3,000 MW in 2020, which is almost 11 times over 10 years [8] [9]. That is while winters are snowy, harsh, and long in Canada, and on average, the major cities in this country experience snowfall or snow cover on the ground in almost 30% of the days in a year [10].

Increasing the penetration level of PV energy in power grids is a promising solution for decarbonizing the electricity sector. However, it brings new challenges for the grid operators in terms of the modeling and evaluation of the power grid operation. The overall performance of a PV system in converting solar energy into electricity depends not only on the efficiency of each individual part of the system but also on the weather and environmental conditions. This requires a comprehensive evaluation of different types of power losses in the system. The intermittent nature of PV power generation is the other challenge in decision-making for grid operators. This is even worse for PV installations in northern areas where PV power generation can be significantly disrupted during winter by snowfall events. Hence, modeling the PV system's performance and power generation regarding the characteristics of the system and weather conditions is a key element in the power grid scheduling and operation.

Presenting an accurate model for PV systems' performance in different weather conditions has been a hot topic over the past few years. This mainly includes the prediction of power generation and power losses in PV systems. PV system researchers are contributing to the achievement of this goal by proposing models that provide predictions as close as possible to the actual behavior of PV systems. While many models have been so far proposed by researchers for different types of PV systems in different climatic conditions, there are fewer studies focusing on installations in snow-prone regions. That is while modeling the performance of PV systems in snow conditions has its own challenges due to the probable coverage of the panels with a snow layer. Filling this research gap and addressing the problem of PV performance modeling in snow conditions is the main motivation of the current thesis.

1.3 Problem Statement

Modeling the performance of a PV system in terms of different types of power losses and also power generation can be categorized into three methods, namely

analytical modeling, statistical modeling, and machine learning. While analytical modeling requires great knowledge about the underlying physics and accurate relationships between the PV system's performance and all the affecting factors, statistical modeling tries to represent these relationships by developing parametric models according to the data records of the essential parameters. On the other hand, machine learning-based models utilize computer algorithms that automatically learn from a huge amount of historical records of all effective parameters and make accurate predictions of the PV system's performance using the capabilities of computational intelligence techniques. Considering the advantages of the last method, computational intelligence-based modeling of PV systems' performance and power generation and its application in power grids' scheduling for PV installations in snow-prone areas is the main focus of this thesis.

1.3.1 Problem 1: PV System Power Losses Prediction in Snow Conditions

The performance of a PV system depends on different types of power losses which can reduce the overall power production of the system. These losses can be categorized into two major groups, i.e. PV array capture losses and system losses. The array capture losses are associated with the array side, for example, the attenuation of the incoming light (soiling, snow cover, shading, reflection, etc.), temperature dependence, electrical mismatching, and parasitic resistances in PV modules. System losses are caused by the conversion system, for instance, wiring, inverters, and transformers [11]. The amounts of these losses depend not only on the efficiency of the PV arrays and the electrical equipment but also on the weather conditions and meteorological parameters in the location of the system installation. Therefore, analyzing the performance of a PV system in terms of various types of power losses is of significant importance from operational and planning viewpoints. In addition, the possibility to know the current amounts of losses and having an available estimation of the future values of these losses can help the PV system owners to have a clear perspective on the long-term operation of the system and plan for maintenance or other solutions.

Calculating accurate values of the power losses in a PV system requires not only access to the characteristics of all equipment of the system but also having the knowledge of underlying physical and electrical concepts and accurate measurements of all meteorological and environmental parameters affecting the PV system performance at

the installation location. Considering data and information availability issues for a majority part of PV installations, having a simplified approach to provide a fair estimation of different types of power losses can be very helpful. Moreover, machine learning techniques can be utilized to automatically capture the interrelationships between the power losses and meteorological parameters and present a reliable model of the losses that can be used conveniently by end users.

1.3.2 Problem 2: PV System Power Generation Prediction in Snow Conditions

While PV systems may have lower efficiency in northern regions due to lower levels of solar radiation, their highly intermittent power generation during winter due to snowfalls can bring challenges for power systems with high penetrations of PV power [12]. The power generated by a PV module highly depends on the solar radiation reaching the surface of the module [13]. On the one hand, a layer of snow can cause an obstruction of solar irradiance. The amount of radiation passing through a snow layer depends on many factors including the depth of the snow layer, the shape of the snowflakes, and the density of the snow layer. On the other hand, the power generated by a partially snow-covered PV module depends not only on the characteristics of the snow layer but also on the interconnections of the PV module, and the wirings in the PV panels.

Considering all the aforementioned factors is of significant importance in PV power prediction for fully or partially snow-covered systems. However, some of them cannot be simply measured or quantified and the complex relationship between them and the PV power is not easily interpretable. This is the reason why PV power modeling is harder in snow conditions. This can be a major issue for PV installations in northern snow-prone areas that experience long winters with huge amounts of snowfall such as Canada. That is while having an accurate model for PV power generation is a key factor in optimal operation and planning studies of renewable energy-based power grids. Considering the capabilities of computational intelligence techniques, historical data of PV systems in snow-prone regions, as a valuable source of information presenting the actual effect of snowfall events on PV power generation in terms of the intensity and duration of power reductions, can be used to develop PV power prediction models compatible with snow conditions.

1.3.3 Problem 3: PV System Power Generation Forecasting in Snow Conditions

One of the sources of uncertainty in the scheduling of power grids with considerable penetration levels of renewable energies originates from the intermittent power output of renewable energy resources (RERs), such as PV systems, that can be effectively tackled by means of computational intelligence techniques. The intermittency of renewable energies considerably influences the reliability and stability of power systems' operation and economic dispatch. Hence, reliable PV power forecasting can highly decrease this uncertainty, enhance stability, and improve economic viability. Therefore, accurate PV power forecasting is a hot and interesting research topic in the field of power grid scheduling. The future time period for output power forecasting or the time duration between the actual and effective time of prediction is the forecast horizon which is categorized as very short-term, short-term, medium-term, and long-term forecasting with prediction periods of some seconds up to 30 minutes, 30 minutes to one day, a day to a month, and a month to a year, respectively [13]. As the length of the forecast horizon increases, the accuracy of the PVPF modeling approach decreases because cloud cover and distribution, which are strongly correlated with solar irradiance, cannot be predicted with considerable precision for extended time periods due to its inherent stochastic nature. Power forecasting for PV installations in snow-prone areas is even more challenging due to the probability of full/partial snow cover formation on the panels, the uncertainty in snow removal events by melting or sliding, and the intensity of the power reduction caused by snow covers. PV system researchers have not fully addressed this issue and there is a good opportunity to fill this gap by developing PVPF models compatible with snow conditions using computational intelligence techniques and evaluating their performance in power sector applications.

One of the main applications of reliable and accurate short-term PV power forecasting is in the day-ahead scheduling of electric distribution systems with PV installations and the scheduling problems linked to that. In the past few years, renewable energies and plug-in electric vehicles have attracted much attention to address the environmental concerns of global warming. The intermittent nature of RERs together with the uncertain and large charging load of plug-in electric vehicles (PEVs) in high penetration levels demand a highly flexible power grid. This flexibility, however, can be effectively achieved by applying a smart charge and discharge management to the PEVs so that the vehicles are considered not only as controllable loads, that can be shifted to

off-peak hours, but also as distributed generation units, that can provide technical support for the power grid through the vehicle-to-grid (V2G) technology [14]. This charge-discharge management, which requires the infrastructure of a smart grid, can be performed either in a passive way by the PEV owners with the help of financial incentives, or in an active way by the power grid operators and PEVs aggregators with the aim of reducing the operation cost [15], increasing the profit [16], improving the power grid operation condition [17], reducing the environmental footprints, or a combination of them [18] [19]. With the number of PEVs increasing on the road, a competitive market is emerging for PEVs aggregators that, as independent entities, provide smart charging solutions for PEV owners. This has made the optimal decision-making of a PEVs aggregator an interesting research topic [20].

Another example of the application of short-term PV power forecasting is the energy management system of microgrids. Microgrids appear to be a prominent solution for the optimized operation of power systems with high penetrations of renewable energies. A microgrid can be considered a small-scale power system with a cluster of loads, distributed renewable-conventional generators, and energy storage units operating together with energy management, control, and protection devices [21]. As a key component in realizing the optimal operation of a microgrid, the energy management system can be defined according to the International Electro-Technical Commission (IEC) standard as a computer system providing a software platform for basic support services and several applications for functionalities required for the efficient operation of electrical generation and transmission facilities to guarantee adequate security of energy supply at the lowest cost [22]. Model predictive control (MPC) is known to be a reliable solution to implement the optimal energy management strategy in microgrids. MPC solves an optimization problem at each sampling time in order to determine the control signals that result in the minimum operation cost while maintaining the demand-supply balance and considering technical and physical limitations. It consists of an optimization block, where the optimal decisions are made, and a forecasting block, where the forecasts of the uncertain variables are generated. Having snow conditions-compatible PVPF models is essential to guarantee the optimal operation of renewable energy-based microgrids in snow-prone regions.

1.4 Objectives of Research

As stated before, accurate modeling of the performance and power generation of PV systems in snow conditions can be an important factor in the optimal decision-making of power grid operators. While developing analytical and statistical models compatible with snow conditions is a big issue, computational intelligence-based models can automatically capture the behavior of the system in snow conditions based on historical data, facilitate the modeling procedure, and increase the modeling accuracy. The goal of this thesis is to develop computational intelligence-based models for PV power losses prediction, PV power generation prediction, and short-term PV power forecasting applicable to PV installations in snow-prone areas. These models can be made available to end-users, such as electric system operators, in the form of computer program tools. To this end, the following specific objectives are presented:

- **Objective 1:** Develop a systematic procedure of power losses calculation for PV systems and model them using computational intelligence techniques;
- **Objective 2:** Propose a novel computational intelligence-based approach to PV power generation modeling in snow conditions;
- **Objective 3:** Propose a novel computational intelligence-based method of short-term PV power forecasting compatible with snow conditions and investigate its applications in the power sector.

1.5 Methodology

In order to achieve the objectives stated in the previous section, this study is accomplished based on data gathering, data preparation, model training, and model validation using different types of computational intelligence techniques. The proposed methodologies are briefly described as follows.

1.5.1 Methodology of Objective 1

The first objective of this thesis is to develop and implement a comprehensive approach for calculating/estimating and modeling various types of power losses in PV systems using computational intelligence techniques. This objective will be achieved through the following two parts. At first, snow loss estimation and modeling are investigated in part I and then, the study is extended over all different types of power losses of PV systems in part II.

In the first part, a three-stage model is developed to calculate the snow loss of a PV farm in Ontario, Canada. PV yield determination in the first stage, system power loss calculation in the second stage, and snow loss extraction in the third stage constitute the aforementioned model. Then, the daily snow loss values together with the daily meteorological measurements for snowfall, temperature, humidity, irradiance, and wind speed on the farm site over the duration of 4 years are used to develop a series of snow loss prediction models using machine learning techniques. Regression trees, gradient boosting trees, random forest, feed-forward artificial neural networks, long short-term memory networks, and support vector regression models are among the implemented algorithms.

In the second part, a systematic approach to PV system power losses calculation is developed. This approach can extract and calculate the amounts of different types of power losses in a PV system based on the monitored field data of the main electrical parameters and some meteorological parameters measured at the PV site. Both array capture losses (including temperature loss, mismatching and soiling losses, low irradiance, spectral, and reflection losses, module quality degradation, and snow loss) and system losses (including inverter loss, cabling loss, inverter power limitation loss, and maximum power point tracking losses) are analyzed in detail in the proposed approach. The overall performance together with the power losses of a 1.44 kW rooftop PV system located in Denver, CO, are investigated by applying the proposed approach to the system's historical data recorded over an 8-year period. A novel approach to modeling each type of PV system power loss and performance based on computational intelligence techniques is also proposed and introduced as a promising solution to tackle the complexity of classical calculations. The proposed models, which are more comprehensive and more generalizable due to the large amount of data used to train them, can predict the future values of the losses only based on the main meteorological parameters. To this end, two widely-used and well-performing computational intelligence techniques including gradient boosting trees and long short-term memory networks are used to build the prediction models for each type of power loss. The performance of the proposed models is then validated on the calculated power losses of the roof-top PV system. In addition, the prediction models are applied to another PV system with different technical characteristics in different climatic conditions to evaluate whether the models developed for a system are applicable to another one or not.

1.5.2 Methodology of Objective 2

The second objective of this thesis is to develop and implement a systematic analysis and computational intelligence-based modeling of PV power generation in snow conditions. This objective will be achieved through the following two parts. First, several snow cover prediction models are developed in part I. The study is then completed in part II by proposing a PV power prediction method by categorizing data based on snow cover conditions.

In the first part, computational intelligence-based snow cover prediction for PV systems is proposed. This 3-step approach detects the most probable full/partial snow covers on the panels and labels (2-class binary labeling) the data points for the snow cover conditions. The historical dataset of hourly power generation and 16 meteorological parameters for a PV system in Canada over almost three years is used as the case study. Then, various computational intelligence-based classification models are developed over the dataset under study. The meteorological parameters are used as features and the snow cover condition (binary classes representing snow-covered/clean panels) is the target variable. The best snow cover prediction model is chosen by comparing the average test accuracy. In order to evaluate the performance of this model in the case of other PV systems, it is tested on the unseen data of two other PV systems in Canada.

In the second part, the effects of 16 meteorological variables on PV power prediction in snow-related conditions are investigated. The analysis is performed on the hourly historical data of 17 PV systems across Canada with an aggregated time period of more than 55 years. First, three datasets, i.e. snow-free condition, snow condition, and snow-cover condition datasets, are extracted from the full datasets of each PV system. The values of snowfall, snow depth on the ground, and snow cover on the panels are used to extract these datasets. The computational intelligence-based 3-step approach, introduced in the previous part, is used to detect the most probable full/partial snow covers on the panels. Then, principal component analysis is performed to reduce the dimensionality of the datasets and recognize the most important attribute variables. Moreover, various computational intelligence techniques are used to model PV power in each snow-related condition. A comprehensive comparison of the performance of the models is then provided to investigate how accurate the models are when there is

snowfall, snow accumulation on the ground, or snow cover on the panels. The performance of the best computational intelligence-based models is also compared with the well-known Marion model, a modified version of the Marion model, classic PV power calculation, and a computational intelligence model combined with the snow cover detection and quantification of the Marion model to prove the effectiveness of the proposed approach.

1.5.3 Methodology of Objective 3

The third objective of this thesis is to develop and implement computational intelligence-based short-term PV power forecasting in snow conditions and investigate its applications in power grids' scheduling. This objective will be achieved through the following two parts, where a PVPF model based on exogenous weather forecasts is developed for day-ahead scheduling of a PEVs aggregator in part I, and a novel PVPF approach independent of external weather forecasts is proposed for energy management of a microgrid in part II.

In the first part, the optimal day-ahead scheduling of a PEVs aggregator offering a novel grid-support service is investigated. The proposed service is designed based on the participation profitability, meaning that the aggregator provides the support as far as it is beneficial for the PEV owners. The objective is to minimize the total daily charging cost of the vehicles, by participating in the day-ahead energy market (DAM) and the real-time energy market (RTM), and fulfill the PEV owners' charging demand. This is achieved by developing a two-stage stochastic programming approach to tackle uncertain driving patterns and real-time market clearing prices (RTMCPs). Day-ahead market clearing prices (DAMCPs) can be estimated with high accuracy due to being close to the market clearing process. It is assumed that the local distribution system operator (DSO) owns a grid-connected PV plant. The DSO's day-ahead predictions of the hourly PV power generation may not be accurate, especially on a snowy day. Snow loss affects the output of PV systems significantly and has a highly uncertain nature. The DSO's inaccurate day-ahead market bids, which originate from errors in day-ahead load and PV power predictions, confront the DSO with the highly volatile RTMCPs. To deal with this issue, a novel local out-of-market balancing service is proposed in this part which enables the aggregator to provide the extra energy demanded by the grid in real-time and reduce the PEVs' charging costs by the profit earned through the service. Integrating the proposed

balancing service into the aggregator's day-ahead scheduling requires day-ahead predictions of the hourly values of the grid load demand and the PV power generation which are obtained by developing two computational intelligence-based predictors. A risk assessment is also performed by integrating the risk-averse equations into the optimization problem.

In the second part, an energy management system (EMS) based on a two-stage model predictive control for microgrids in snow-prone areas is developed. The system covers the responsibilities of the tertiary and secondary control levels. The microgrid is assumed to have a considerable penetration of PV energy which might be significantly affected by snowfall events. In order to address the uncertainties pertaining to the PV power generation and load demand of the microgrid, short-term PV power and load forecasting models compatible with snow conditions based on computational intelligence techniques are proposed and embedded in the developed EMS. These models are independent of external weather forecasts and only rely on the local measurements of the main meteorological and electrical parameters. The objective is to minimize the operation cost of the microgrid while maintaining the load-supply balance.

1.6 Statement on the Originality of the Thesis

Many studies can be found in the literature on modeling PV systems' performance and power generation using computational intelligence techniques. However, there is no comprehensive study on this problem for PV systems in snow-prone areas focusing on snow conditions. Therefore, the current thesis addresses this research gap by evaluating the problem from different perspectives, proposing new methodologies and models, and validating them using different case studies. The contributions of the objectives of this thesis are presented in the following.

The contributions of each part of the first objective of this thesis, which focuses on computational intelligence-based modeling of PV power losses in snow conditions, can be listed as follows:

Part I of Objective 1:

- Developing a 3-stage model to estimate snow loss of PV farms based on the hourly measurements of the meteorological parameters and PV yield;

- Proposing computational intelligence-based prediction models to predict daily snow loss of PV farms solely based on meteorological parameters.

Part II of Objective 1:

- Developing a novel systematic approach for detailed power losses calculation for PV systems based on monitored field data;
- Proposing a novel comprehensive and generalizable computational intelligence-based approach for modeling each type of PV system power loss that is capable to predict the future values of the losses using the main meteorological parameters;
- Validating the performance of the developed prediction models in the case of a PV system with different technical specifications and on-site climatic conditions.

The contributions of each part of the second objective of this thesis, which investigates computational intelligence-based modeling of PV power generation in snow conditions, can be listed as follows:

Part I of Objective 2:

- Proposing a computational intelligence-based 3-step approach to detect the most probable full/partial snow covers on PV panels;
- Developing various computational intelligence-based snow-cover prediction models solely based on meteorological parameters.

Part II of Objective 2:

- Conducting a systematic analysis of the effects of 16 meteorological parameters on the PV power generation in snow-prone areas focusing on snow-related conditions;
- Developing a computational intelligence-based prediction approach for the hourly PV power generation based on snow-related conditions and providing a comprehensive comparison of the models' performance with respect to the existing solutions in the literature.

The contributions of each part of the third objective, which addresses computational intelligence-based short-term PV power forecasting in snow conditions and evaluates its applications in power grids' scheduling, can be listed as follows:

Part I of Objective 3:

- Proposing a snow loss-aware short-term PV yield predictor using long short-term memory networks to model the hourly power generation of snow-covered PV systems based on meteorological parameters;
- Developing a stochastic programming approach with a comprehensive availability model of PEVs for both residential night-charging and public intraday-charging for optimal day-ahead scheduling of an independent PEVs aggregator;
- Proposing and validating a novel local balancing service offered by the PEVs aggregator to the DSO in order to provide the extra real-time energy demand of the grid by the PEVs and benefit the PEV owners financially.

Part II of Objective 3:

- Proposing a novel short-term PV power forecasting approach based on computational intelligence techniques compatible with snow conditions and independent of exogenous weather forecasts;
- Developing an MPC-based EMS for photovoltaic energy-penetrated microgrids in snow-prone areas;
- Validating the performance of the developed EMS in different weather conditions against a developed heuristic control method of the microgrid.

1.7 Thesis Outline

The individual chapters of this thesis are organized as follows.

Chapter 2 is a literature review. This chapter conducts a comprehensive review of the methods developed for PV systems' power losses calculation and modeling, the models built for PV power prediction in snow conditions, the papers propose various types of PVPF models, the studies address PEVs aggregators' day-ahead scheduling problem, and the approaches used for optimal energy management in microgrids.

Chapter 3 provides a brief explanation of the computational intelligence techniques used in this thesis. These techniques include linear regression, naïve Bayes, logistic regression, k-nearest neighbors, decision tree, random forest, gradient boosting tree, support vector machine, artificial neural network, and long short-term memory network. The three objectives of this thesis will be accomplished in the next chapters; so that, the first and second parts of the first objective in chapters 4 and 5, the first and second parts of the second objective in chapters 6 and 7, and the first and second parts of the third objective in chapters 8 and 9 provide all the proposed computational intelligence-based PV performance modeling in details.

Chapter 4 investigates the estimation of daily values of snow loss based on historical data of a PV system. Several computational intelligence-based snow loss prediction models are also developed in this chapter.

Chapter 5 proposes a systematic approach for the calculation of different types of power losses based on historical data of a PV system. Then, each type of power loss is modeled using computational intelligence techniques.

Chapter 6 develops a 3-step method for detecting the presence of a full/partial snow cover on the panels of PV systems. Computational intelligence-based snow cover prediction models are then proposed in this chapter.

Chapter 7 conducts a detailed analysis of over 55 years of historical data on electrical and meteorological parameters for 17 PV systems across Canada. Then, computational intelligence-based PV power prediction based on snow-related conditions categorization is proposed as an effective approach to power generation modeling for PV systems in snow-prone regions.

Chapter 8 studies the day-ahead scheduling problem of a PEVs aggregator offering a type of balancing service to the local distribution grid. This chapter develops a computational intelligence-based PVPF model which is compatible with snow conditions and provides reliable forecasts to guarantee the optimal decisions of the aggregator.

Chapter 9 addresses the optimal energy management of a renewable energy-based microgrid located in a snow-prone region. Model predictive control is implemented to maintain the optimal operation of the microgrid in all conditions. A novel computational

intelligence-based PVPF model independent of exogenous weather forecasts is proposed in this chapter which is integrated into the forecasting block of the MPC.

Chapter 10 summarizes the work presented in this thesis, highlights the research contributions, and outlines the topics for future investigation.

1.8 List of Publications

A list of the papers derived from this thesis, which are published so far or have been submitted, is given as follows:

Journal Papers:

[R1] **B. Hashemi**, S. Taheri, and A.M. Cretu, "Computational intelligence-based energy management system of microgrids in snow prone regions," (Under Review)

[R2] **B. Hashemi**, S. Taheri, A.M. Cretu, and E. Pouresmaeil, "Computational intelligence based PEVs aggregator scheduling with support for photovoltaic power penetrated distribution grid under snow conditions," *Electric Power Systems Research*, vol. 214, Jan. 2023.

[R3] **B. Hashemi**, S. Taheri, and A.M. Cretu, "Systematic Analysis and Computational Intelligence Based Modeling of Photovoltaic Power Generation in Snow Conditions," *IEEE Journal of Photovoltaics*, vol. 12, no. 1, Jan. 2022.

[R4] **B. Hashemi**, S. Taheri, A.M. Cretu, and E. Pouresmaeil, "Systematic Photovoltaic System Power Losses Calculation and Modeling Using Computational Intelligence Techniques," *Applied Energy*, vol. 284, Feb. 2021.

[R5] **B. Hashemi**, A.M. Cretu, and S. Taheri, "Snow Loss Prediction for Photovoltaic Farms Using Computational Intelligence Techniques," *IEEE Journal of Photovoltaic*, vol. 10, no. 4, pp. 1044-1052, July 2020.

Conference Papers:

[R6] **B. Hashemi**, A.M. Cretu, S. Taheri, "Computational Intelligence Based Snow Cover Prediction for Photovoltaic Systems," in *2021 IEEE Electrical Power and Energy Conference (EPEC)*, Toronto, ON, Canada, Oct. 2021.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter starts with a description of the challenges of PV systems performance modeling and the existing opportunities. Then, a literature review pertaining to the objectives of the thesis, including strengths and weaknesses of state-of-the-art methods, is presented. In this regard, a review of the calculation and modeling of various types of power losses in PV systems is presented. Then, the existing works in the literature on modeling PV systems power generation in snow conditions are reviewed. Finally, a comprehensive literature review of PV power forecasting methods and their applications in power systems-related scheduling, i.e. day-ahead PEVs aggregators scheduling and microgrids energy management, is provided at the end of this section.

2.2 Challenges and Opportunities in PV Systems Performance Modeling

The performance of PV systems depends on both the efficiency of each individual part of the system, including PV panels and inverters, and the weather and environmental conditions, including solar irradiance, ambient temperature, and shading/soiling on the surface of the panels caused by clouds, vegetation, buildings, dust, snow, etc. Mathematical modeling of a PV system's power losses and power generation by extracting the underlying relationships between all electrical and meteorological factors and representing them in the form of empirical equations, which is called analytical modeling, is a very complex task. This not only requires a vast knowledge about the underlying physical and electrical concepts, energy conversion rules, and the characteristics of all parts of the system but also demands accurate measurements and observations of all factors affecting the system's performance. Eventually, the extracted equations might be so complex and rely on many external parameters that make it difficult to implement or might be so simplified that cannot accurately model the behavior of the system in altering conditions.

On the other hand, the statistical modeling of the PV system's performance and power generation can address many issues of analytical modeling. Statistical models can

identify the mathematical relationships between the independent and dependent variables in an automatic manner. Using the capabilities of computational intelligence and machine learning techniques, computer programs can be developed that are able to learn from historical data without explicitly being programmed. These techniques use algorithms and statistical models to analyze and draw inferences from patterns in data. Hence, the modeling of the PV system's performance and power generation can be performed with limited knowledge about the system's specifications and the underlying electrical-physical concepts. Moreover, computational intelligence-based models can capture the nonlinear behavior of the system under study and make predictions based on a limited number of input variables. This can be very beneficial for PV systems in snow-prone areas where the effect of snow cover on the panels adds to the complexity of the modeling. In the following, a comprehensive literature review is presented on all three topics addressed in this thesis.

2.3 Review of PV Power Losses Calculation and Modeling in Snow Conditions

PV power losses calculation and the use of computational intelligence/machine learning techniques for modeling some particular types of power losses in PV systems have been investigated in a good number of papers in the literature. Performance metrics such as performance ratio and efficiency have been widely used in the literature to present the effects of the total power losses in PV systems. The authors of [23] evaluate the annual performance loss rates of five different grid-connected PV technologies based on historical data of field measurements in Italy, Cyprus, and Australia. In [24], the power output, efficiency, and relative efficiency losses are calculated and compared for five different types of PV modules based on monitored data collected in a temperate mountain climate. The authors investigate the influence of the high relative air humidity on the PV conversion efficiency during time intervals with a clear sky and low amounts of incident solar irradiance.

Some papers only investigate the array capture losses and system losses and do not go further into details. In [25], the performance of a 1.72 kW PV system in Ireland is studied. The monthly and annual values of the final yield, reference yield, array yield, system losses, array capture losses, cell temperature losses, PV module efficiency, system efficiency, inverter efficiency, performance ratio, and capacity factor are

calculated for the system. The array capture losses and the system losses together with some other performance measures for a 2.07 kW PV system in Norway are calculated in [11]. In [26], the experimental results on the performance of a 302.4 kW grid-connected PV power plant in Djibouti operating under dusty, desert maritime climate conditions are provided. The daily performance metrics including the reference yield, array yield, final yield, performance ratio, system losses, and array capture losses are calculated in this study. The paper also provides a quantitative estimate of losses due to soiling. The authors of [27] evaluate the performance of an 11.2 kW grid-connected PV system in India in terms of PV module efficiency, array yield, final yield, inverter efficiency, performance ratio, capture losses, and system losses. The performance ratio, system losses, and array capture losses of a 1 kW PV system in Poland are investigated over a year in [28], while in [29], the performance of a standalone PV system in a remote island in China is studied in terms of the performance ratio, efficiency of panels and inverter, array capture losses, and system losses using 2 years of recorded data.

PV simulation computer software is used in some papers to obtain the values of different types of losses in PV systems. PVsyst software, a software package developed by the Energy Group at the University of Geneva in Switzerland, is utilized in [30] and [31] to analyze the performance and power losses of PV systems under study. TRNSYS software is another tool used in [32] to analyze the performance of a 7.2 kW PV system. The authors of [33] investigate the performance of a 960 kW PV system in terms of the monthly average of the energy yields, losses, and efficiency using both System Advisor Model and PVsyst software. The performance parameters of a rooftop PV system are measured and analyzed in [34] by using two software tools including PVsyst and HelioScope.

Some papers propose approaches for calculating one particular or multiple types of power losses as part of array capture or system losses based on monitored field data. The authors of [35] mainly focus on soiling (dust) effects on the PV yield by testing a PV system installed near a mine in a dry area. They propose a framework that works by dividing the recorded power data and the output of an ideal model of the produced power into daily frames and assessing the frames considering the effects of daily cloud and rain levels. A model for calculating the soiling losses of PV panels is presented in [36], which uses ambient airborne particulate matter (PM) concentrations, both PM10 and PM2.5, the tilt of the PV array including if the array is tracking, and rain data to estimate soiling

losses over time. In [37], a method for calculating PV yield reduction due to dust deposition on the surface of panels is proposed based on optical analysis. The calculation is carried out based on the changes in daily PM10 and rainfall. In [38], a method termed the stochastic rate and recovery method is proposed for estimating the soiling losses directly from the PV yield without the need for precipitation data. The method automatically detects soiling intervals in a dataset, then stochastically generates a sample of possible soiling profiles based on the observed characteristics of each interval. The authors of [39] present analytical expressions for the annual energy losses on DC and AC cables, the annual energy losses on step-up transformers, and the optimum transformer size for fixed-mounted PV plants of any size. The sophisticated verification (SV) method developed in [40] can estimate six types of loss rates (shading effect, losses due to incident angle, load mismatch, efficiency decrease by temperature, inverter losses, and other losses) using system specifications, such as latitude, longitude, inclination angle, azimuth, system rating, temperature coefficient, and measured operational data (inclined-plane irradiation, array output, system output, and module temperature). The authors of [41] conduct a performance and loss analysis for residential PV systems using the SV method. The performance of different system configurations is quantitatively investigated and compared by providing annual ratios of different losses. In [42], one hundred and two environmental and meteorological parameters are selected and their effects on the performance of 20 soiling stations installed in the USA are investigated to determine their effectiveness in predicting the soiling losses in PV systems. Considering the significance of both particulates and rainfall in predicting soiling ratios and soiling rates, a two-variable regression model is proposed in this paper. A continuous non-linear method is presented in [43] to model array long-term performance for six PV arrays in Italy, China, Brazil, and Switzerland. The proposed performance degradation is obtained using three continuous non-linear functions of time which allows considering the effects of different degradation factors including light-induced degradation.

Another type of power loss for PV systems in snow-prone areas is snow loss which is caused by partially or fully covering the surface of the panels with a snow layer. This causes an obstruction for the solar irradiance reaching the surface of the panels. Snow loss calculation is addressed in several research papers. Some papers investigate the problem experimentally by doing only some measurements on testbeds exposed to snow. In [44], the outputs of a PV module for two different days with the same meteorological

conditions with and without snow on the panels are measured and compared. In [45], the outputs of two sets of modules, one exposed to snow and the other cleaned manually are measured. Finally, they are subtracted to calculate the snow loss. The authors of [46] investigate PV snow loss as a function of tilt angle and snow sliding obstruction geometry. They developed a testbed consisting of several modules with different tilt angles and presented two equations for calculating the PV yields with and without snow cover. Some papers provide simple models to predict the snow loss of PV panels. A linear regression model and a non-linear empirical model are introduced in [47] which relate monthly snow loss to monthly snowfall, tilt angle, number of snow events per month, average air temperature, the plane of array insolation, and average relative humidity. The snow loss is obtained using the historical data for two sets of PV modules, one in normal condition and the other manually cleaned from the snow. A time series model is proposed in [48] to allow for the prediction of snowfall effects. This model is built over the data obtained by subtracting the measured PV outputs from the values of a standard model based on meteorological data. The results of this work show that the proposed model does not work accurately in predicting the daily snow loss of the PV system.

Modeling the PV system power losses using computational intelligence methods is a reliable solution to tackle the complexity of classical calculations. These models can also be used to predict the future values of the losses based on simple input data. The authors of [49] describe a method for creating a dataset of the power loss under shade conditions and investigating the logic behind the selection of the scenarios. An artificial neural network (ANN) model is then proposed for estimating the shading loss. The ratio of the array area covered by shade to the total array area, the ratio of diode-protected cell strings impacted by shade to the total number of cell strings in the system, and the ratio of module strings impacted by shade to the total number of parallel module strings in the system are the features of the model. The normalized power from the 2-diode model is the target. An artificial neural network approach is proposed in [50] for modeling the relationship between environmental variables and daily change in the Cleanness Index, which is a measure of the performance loss due to the soiling of PV modules. Ambient dust mass concentration (PM10), wind speed, wind direction, ambient temperature, and relative humidity are used as inputs in this study. In [51], an artificial soiling experiment is applied to a PV panel using five soil samples collected from five different locations in

India. Then, a regression model is developed to analyze the particle size influence on the soiling loss. An artificial neural network model is also built based on experimental data to predict the power loss at various levels of irradiance. The authors of [52] investigate the effect of soiling on the performance of large-scale PV plants by comparing the power output of the system before and after clean-ups. Then, Bayesian neural network-based models are proposed and compared with the well-known regression method to model this effect.

While computational intelligence algorithms are implemented in the literature to model PV yield [53] or a particular type of PV losses based on historical data, there is no work on employing such methods to predict snow loss of PV panels [54]. Moreover, while the calculation of various performance metrics, array capture losses, system losses, and their different subsets is addressed widely in the literature for PV installations in different climatic conditions, there is no work that investigates all types of power losses (including snow loss) together for a PV system in a snow-prone area by presenting a clear calculation approach based on historical field data. As well, while computational intelligence/machine learning techniques have been used extensively to model PV power generation or some particular types of PV power losses, to the best of our knowledge, there is no work in the literature that implements computational intelligence techniques for modeling all different types of PV power losses together based on meteorological parameters. Hence, developing a systematic approach for PV power losses calculation and modeling based on computational intelligence techniques can fill the aforementioned gaps and be an interesting research topic.

2.4 Review of PV Power Generation Modeling in Snow Conditions

PV power prediction and forecasting have been extensively studied in the literature for various time resolutions [55] [56] and for PV systems with various sizes [57] [58]. To this end, a series of physical and data-driven models have been proposed [13]. Developing empirical equations for PV power prediction [59], computational intelligence-based prediction and forecasting models such as decision tree-based models [60], support vector regression [61], feed-forward and deep recurrent neural networks [62], and many different hybrid models like wavelet packet decomposition-based long short-term memory networks [63], autoregressive integrated moving average combined with deep belief network [64], extreme learning machine with vector functional link and

stochastic configuration network [65], genetic algorithm-based extreme learning machine [66], ant colony optimization-based support vector machine [67], etc. are just a small part of the work done in this field.

Nevertheless, there are few papers addressing PV power prediction considering snowfall events. The authors of [68] propose a scenario-based binary decision-making model of a snow cover existence on PV panels. In the case of a probable snow cover, PV power is considered to be zero. In [48], a linear regression model is proposed with the main purpose of daily, monthly, and annual snow loss calculation. The predictions of the PV power in winter show large errors. An empirical approach aiming at improving PV power predictions during periods of snow cover based on a simple persistence model is proposed in [69]. The authors of [12] develop a snow cover detection and quantification model, called the Marion model, which identifies when snow slides down the PV panel based on an empirical relationship between plane-of-array irradiance and air temperature, and determines how far the snow slides based on the PV array tilt angle. The PV power generation is then calculated based on the number of snow-free strings. A version of the Marion model by replacing the front plane-of-array irradiance with the absorbed irradiance is built in [70] to reduce the uncertainty in identifying the beginning of snow sliding. The authors of [71] improve the accuracy of the Marion model by proposing a snow sliding coefficient that not only depends on the system tilt angle but also on the snow depth. The authors of [72] develop a neural network-based black-box approach for hourly PV power prediction which is claimed to be adapted to dynamic environmental factors such as dust and snow. However, no result is presented for predictions in snow conditions. In [73], a long short-term memory network-based model is proposed for hourly PV power prediction. No specific data exploration or modeling has been provided in snow conditions. An artificial neural network-based model is proposed in [74] for the daily power prediction of PV systems in Alberta, Canada. The authors simply use an empirical snow adjustment factor based on the system's tilt angle and month of operation. The predictions in winter do not show good accuracy. The authors of [54] present an extensive literature review on modeling the PV systems' performance in snow conditions. The lack of a detailed study on the hourly PV power modeling in snow conditions using computational intelligence techniques is concluded based on this paper. Other papers, such as [75] and [76], only refer to the large prediction errors due to snowfalls.

According to the best of our knowledge, there is no work on PV power modeling focusing particularly on snow conditions and snow-covered PV systems to improve prediction accuracy in snow-related conditions. Moreover, there is no study on developing computational intelligence-based models for the prediction of the presence of full/partial snow covers on the panels. Considering the research gaps in this area, there is great research potential for the aforementioned topics.

2.5 Review of PV Power Forecasting in Snow Conditions and Its Applications

Regarding the significant effects of certain meteorological parameters on the power generation of PV systems, PVPF models fall into two categories based on their inputs. The first category of the models depends on weather forecasts and uses them as their inputs. These forecasts may come from the results of numerical weather prediction (NWP) models or other models. These PVPF models can be physical models based on empirical equations relating the weather parameters to the PV power or machine learning-based models developed over historical data. In this case, the performance of the PVPF models is subject to inaccuracies in the forecasted weather parameters. According to the literature, day-ahead forecasts frequently use NWP outputs. The literature search demonstrates that about 57% of intra-day and approximately 79% of day-ahead forecasting approaches employed NWP variables [13]. The reason is that NWP improves accuracy when time horizons increase, feeding future meteorological trends to the models. It is worth mentioning that the NWP models combine meteorological information and atmosphere model equations for arriving at predictions. The second category of the models is independent of external weather forecasts and only uses the current and past values of the local measurements of the electrical and meteorological parameters to forecast the future values of the PV power. These models can be time series-based models, e.g. autoregressive moving average (ARMA) model, or machine learning-based models. In contrast to the first category, the performance of these models only depends on the models themselves and the measurement accuracy of the parameters.

There are many studies in the literature focusing on computational intelligence-based short-term PV power forecasting. Various types of machine learning techniques, e.g. regression tree, gradient boosting trees, random forest, support vector regression, multi-layer perceptron, deep recurrent neural networks, and many hybrid models, various

types of optimization techniques to find the optimal parameters of the models, e.g. grid-search, particle swarm optimization, genetic algorithm, and ant colony optimization, different types of pre-processing methods, e.g. wavelet transform, trend-free time series, empirical mode decomposition, self-organizing map, normalization, and singular spectrum analysis, and different types of post-processing methods, e.g. anti-normalization and wavelet reconstruction, have been implemented in the literature to develop PVPF models [13] [77] [78]. In [79], a genetic algorithm-based support vector machine (GASVM) is proposed to forecast the hourly power generation of a PV system in Australia based on the measurements of solar irradiance and ambient temperature. The GASVM model classifies the historical weather data using an SVM classifier initially and later it is optimized by the genetic algorithm using an ensemble technique. The PVPF is done by an ensemble creating a separate classifier for each PV range and then polls all the classifiers at once with regard to their respective designated PV ranges. In [80], a long short-term memory network with two steps for modifying the PV power forecasts is proposed. The inputs of the model are the local direct normal irradiance and ambient temperature. First, an independent day-ahead PVPF model based on the long short-term memory (LSTM) network is established. Second, a modification method is proposed to update the forecasting results of the model based on time correlation principles regarding different patterns of PV power in the forecasting day. Third, a partial daily pattern prediction framework is proposed to provide accurate daily pattern prediction information of particular days, which is used to guide the modification parameters. The authors of [81] propose a physical hybrid artificial neural network model for day-ahead power forecasting for a PV system in Italy. The model uses the forecasts of ambient temperature and solar irradiance as well as the corresponding power output of the five-parameter PV model (where optimization is used to find the parameters of the model based on the historical data) as the inputs. The results show that the forecasting errors over winter are larger. A 3-stage short-term PVPF approach formed by combining the empirical mode decomposition (EMD) technique, sine cosine algorithm (SCA), and extreme learning machine (ELM) technique is proposed in [82] for a PV system in India that receives the measurements of plane-of-array irradiance and module temperature as the inputs. At the initial phase of the proposed technique, a de-noised series is obtained by adopting a signal filtering strategy based on the EMD technique. Next, three different time interval data series are opted for the training and forecasting stage. In [83], four different LSTM-based

short-term PVPF models are proposed, two of which use convolutional neural networks to obtain higher levels of abstraction, and the other two use multiple LSTM layers. The measurements of solar irradiance, ambient temperature, and wind speed are used as the inputs of the models. The authors of [84] develop a physical hybrid artificial neural network short-term PVPF model by combining the deterministic clear sky solar radiation algorithm and the stochastic ANN model receiving the forecasts of 8 meteorological parameters from an external source as the inputs of the model. The model will partly overcome the inaccuracies of the weather forecasts.

Using computational intelligence techniques to forecast hourly PV power generation has been widely addressed in the literature [53]. In this context, short-term power generation forecasting for snow-covered PV panels is more complicated due to the significant effect of snow on the performance of the panels which depends not only on the snow depth but also on snow type, snow cover shape, etc. As discussed in the previous section, PV power prediction in snow conditions has been addressed in a few papers. However, there is no study on computational intelligence-based PV power forecasting focusing on snow conditions [54]. Other papers, such as [75], do not consider snow cover in their models and only refer to the effect of snow as the source of errors in forecasting. Considering this research gap, short-term PV power forecasting in snow conditions can be an interesting research topic.

As stated in chapter 1, one application of PV power forecasting can be found in the scheduling problem of a PEVs aggregator interacting with a renewable energy-based power grid. In this context, the optimal day-ahead scheduling of a PEVs aggregator participating in day-ahead electricity markets has been widely addressed in the literature. The day-ahead energy market enables the participants to avoid significant price volatility in the real-time energy market by locking in energy prices before the operating day. Thus, deploying an optimal day-ahead schedule is a vital part of optimizing the operation of an electricity market participant. However, there are uncertain variables that can affect day-ahead schedules. As an example, the availability of PEVs the next day to receive charge/discharge control signals is a source of uncertainty that a PEVs aggregator has to deal with. A well-known and reliable solution to consider the effects of uncertainties in day-ahead decision-making problems is multi-stage stochastic programming where the uncertain variables are modeled as a set of possible scenarios and corrective actions are taken at each stage after the realizations of the scenarios are known [85]. The multi-stage

stochastic programming has been implemented in many studies to obtain optimal day-ahead schedules of PEVs aggregators.

In [86], the optimal bidding strategy of a PEVs aggregator participating in the day-ahead and real-time energy markets and the frequency regulation market is modeled as a two-stage stochastic programming approach. It aims at minimizing the aggregator's costs considering the PEV- and market-related scenarios. The authors of [16] investigate the optimal scheduling of a PEVs aggregator bidding in the day-ahead energy and reserve markets. The uncertainties are characterized using two-stage stochastic programming that maximizes the aggregator's profit. A two-stage hierarchical profit-enhancing PEV-aggregation system is proposed in [87] for the day-ahead scheduling and real-time operation of a PEVs aggregator. This approach minimizes the energy consumption cost in electricity markets while enhancing the profits of both PEV owners and energy suppliers. The authors of [88] investigate the day-ahead scheduling of a parking lot aggregator using a three-stage stochastic-based structure by modeling three trading floors of the electricity market including day-ahead, adjustment, and balancing markets. The problem is split into three layers where the electricity market model, parking lot aggregator model, and PEVs model constitute the layers. In [89], a stochastic and dynamically updated two-stage multi-period optimal bidding strategy is developed for a PEVs aggregator. The authors of [90] propose a two-stage programming approach for the day-ahead scheduling of a parking lot aggregator that is developed as a two-level model. It maximizes the aggregator's profit in the first level and minimizes the grid operation cost in the second level. In [91], the day-ahead scheduling of a PEVs aggregator participating in the energy and ancillary service markets is addressed by a two-stage stochastic programming approach. A risk-constrained stochastic approach is proposed in [92] for a PEVs aggregator's optimal participation in day-ahead and reserve markets by involving the risk-related uncertainties through the downside risk constraints. This provides the aggregator with decisions that are made by considering various quantities for risk. Stochastic programming has been widely used for optimal scheduling of PEVs aggregators, however, the published literature either does not consider both the night charging at home and the intraday charging at public charging stations or they do not present a clear methodology for that.

The integration of PEVs into power grids provides a great opportunity to exploit the charge-discharge controllability and flexibility of the vehicles in grid-support services

and facilitate higher penetrations of renewable energies in power systems. In this context, PEVs' participation in market-based services, such as reserve or frequency regulation services, has been widely investigated in the literature [93] [94]; however, a limited number of papers address the capabilities of independent PEVs aggregators in providing out-of-market energy services for local distribution grids or other energy entities by investigating the affordability of such services for both sides [95]. These out-of-market services can also be beneficial in case market participation is not achievable for PEVs aggregators due to the market structure or the aggregated PEVs' capacity size. In [96], the day-ahead scheduling of a PEVs aggregator offering balancing services for a wind power producer (WPP) is investigated. The authors of [91] investigate the optimal charging strategy of a PEVs aggregator under incentive and regulatory policies of the DSO. The aggregator provides voltage regulation and power loss cost reduction services for the local DSO. The joint operation of a fleet of PEVs with a wind power producer is studied in [97], in which the PEVs aggregator counterbalances WPP fluctuations. In [98], the day-ahead scheduling of a renewable power producer is addressed where the real-time deviations from the day-ahead bids are compensated by demand response and PEVs aggregators. A similar method is proposed in [99] for a wind farm where the formation and scheduling of a virtual power plant by integrating the electric vehicles and flexible loads is investigated. The proposed method aims to minimize the deviation of the wind power generation capacity from the final amount of cleared power in the electricity market. A flexible penalty contract between a PEV charging station and a DSO in terms of voltage security is introduced in [100]. The authors of [101] propose a strategy to provide voltage regulation services by PEVs aggregators for a distribution grid with a grid-connected PV plant. In [17], a coordinated management system is proposed for PEVs that enables the aggregator to offer grid-support services to avoid grid overload and local voltage violation issues. The authors of [102] develop a PEVs aggregator scheduling approach as a bilevel optimization framework that minimizes distribution system congestion. A distributed privacy retaining model for PEVs' charge and discharge management in the presence of distributed generators in a distribution network is proposed in [103] where the network losses reduction, the aggregator's cost minimization, and the distributed generators' profit maximization are considered in a single framework. In [104], an optimal scheduling algorithm for an aggregator based on the benefits of the distribution network, aggregator, and PEV owners is proposed. The

algorithm maximizes the aggregator's profit while satisfying the minimum requirements of the demand response capability set by the grid operator and the satisfaction required by the PEV owners. An important design issue in some of the aforementioned services is the lack of a reasonable definition for the interactions between the PEVs aggregators and the service receivers. It is important to note that an independent PEVs aggregator is not responsible for the grid operation condition and providing grid-support services unless it is beneficial for the PEV owners. Considering the potential created by the PEVs' integration into the grid, there is also an opportunity to design new grid-support services to benefit both the aggregator/PEV owners and the grid operator the most and reduce the challenges originating from the intermittency of renewable energy resources, which is investigated in this thesis.

This thesis also investigates the application of PV power forecasting in the energy management of microgrids. In some microgrids, the main challenge is to achieve a safe operation by keeping the supply-demand balance, while in others an optimal operation considering economic criteria is aimed to achieve. In the first case, energy management is usually carried out by heuristic algorithms [105]. The Hysteresis band control method is proposed for microgrids' EMS in several papers such as [106] [107] [108] [109] due to its simple structure and ease of implementation. In the case of optimizing the microgrid operation, however, the problem is more difficult. In this case, model predictive control is a reliable solution.

MPC-based microgrid EMSs have been widely investigated in the literature [22] [110] [111]. In [112], the application of MPC for energy management in an islanded microgrid with PV generation and a hybrid storage system is assessed. The objective is to improve the utilization of renewable generation, the operational efficiency of the microgrid, and the reduction in the rate of degradation of storage systems. Load and PV power forecasting models are not presented in this study. The authors of [113] propose an MPC-based model for the management of a microgrid that allows the automatic grid connection to provide ancillary services to the main grid, such as selling the excess of renewable generation and purchasing energy to charge the battery. The forecasts of PV and wind energies are performed based on the physical models and the forecasted weather data obtained from a mesoscale meteorological model. A multi-stage EMS for participating in the deregulated electricity market considering the cost of market participation is proposed in [114]. The proposed EMS, which is based on a two-stage

MPC, forecasts the microgrid load and PV power generation using long short-term memory networks and minimizes the operation cost. The authors of [115] develop an MPC-based approach for controlling the heating, ventilation, and air conditioning systems in a multiple dwelling unit building, including a local renewable power generator, an energy storage system, and electric vehicles. The proposed approach reduces the electric bill of the building and improves the matching performance between the local generation and consumption. In [116], a two-stage MPC-based EMS is developed for a microgrid with PV, battery storage, fuel cell, and electrolyzer systems. The paper proposes a procedure for changing the parameters of the MPC for the hydrogen-based system in real-time to improve its effectiveness. PV power and load forecasts are considered to be perfect and no model is presented for them. An MPC-based EMS is proposed in [117] to improve the hosting capacity of PV systems and electric vehicles in a stand-alone microgrid with an energy storage system. No model is developed to provide forecasts of PV power, EV availability, and load. The authors of [118] propose a two-level model predictive controller enhanced by two data-driven modules. The model identification module aims to continuously improve the accuracy of the MPC's internal model and the cost estimator module adjusts the cost function to satisfy the required marks of the annual self-consumption rate imposed by the grid code at minimum cost. The load and PV power forecasting models are not introduced in this paper. A probabilistic microgrid dispatch problem is studied in [119] where the predictions of the load and the renewable energies are given in the form of intervals. A hybrid method combining scenario-selected optimization and reserve strategy using the MPC framework is proposed. The forecasting models are not presented in the paper.

After doing an extensive literature review and according to the best of our knowledge, there is no paper on computational intelligence-based short-term PV power forecasting focusing on snow conditions. Moreover, there is no work on optimal day-ahead scheduling of a PEVs aggregator offering out-of-market balancing services to support the local DSO. In addition, there is no study evaluating the performance of an EMS for PV-penetrated microgrids in snow-prone areas and proposing an MPC-based EMS approach compatible with snow conditions. Considering the research gaps in this area, there is a great opportunity to conduct more studies on the aforementioned topics.

CHAPTER 3

COMPUTATIONAL INTELLIGENCE TECHNIQUES

3.1 Introduction

The focus of this thesis is on the implementation of computational intelligence techniques to model PV systems' performance and power generation in snow conditions. Hence, several types of machine learning techniques will be used multiple times to address the introduced problems. Therefore, a brief explanation will be presented in this chapter about the implemented techniques.

3.2 Methodologies

A large area within artificial intelligence is machine learning. A computer program is said to learn some task from experience and by the use of data if its performance at the task improves with experience/data, according to some performance measure. Machine learning involves the study of algorithms that can extract information automatically and build a model based on sample data (known as training data) in order to make predictions or decisions without being explicitly programmed to do so. Machine learning techniques fall into three categories, i.e. supervised, unsupervised, and semi-supervised learning [120]. Supervised learning is defined as learning a function that maps an input to an output based on training data that involves both attribute and target variables. Through methods like classification and regression, supervised learning uses patterns to predict the values of the label on additional unlabeled data. Classification is the problem of predicting a discrete class label output (target) while regression is the problem of predicting a continuous quantity output. Unsupervised learning finds previously unknown patterns in datasets consisting of input data without labeled output. The most common unsupervised learning method is clustering. Semi-supervised learning is a hybridization of supervised and unsupervised techniques.

In this thesis, several supervised machine learning techniques are used for the introduced regression and classification problems that are briefly explained in the following [120] [121] [122].

3.2.1 Linear regression

Linear regression is a simple model that is widely used as a benchmark in machine learning to evaluate the performance of other more complicated models in a prediction problem. It provides minimal performance to be accepted by a predictor since a line or a multi-dimensional plane is simply fitted on the data to represent the relationship between variables. In this case, the target variable can be modeled as a linear combination of one or more features as follows [120]:

$$\hat{y} = C_0 x_0 + C_1 x_1 + C_2 x_2 + \ldots + C_n x_n = \boldsymbol{C}\boldsymbol{x} \tag{3.1}$$

where $\hat{y}$, x_n, and C_n are the modeled target variable, n^{th} input feature, and n^{th} coefficient, respectively. x_0 is 1. $\boldsymbol{C}$ and $\boldsymbol{x}$ are also the corresponding vectors. In order to obtain the best coefficients to minimize the mean squared error (MSE) of the model, the following formulation can be used where y is the actual target variable [120]:

$$\boldsymbol{C} = (\boldsymbol{x}^T \boldsymbol{x})^{-1} \boldsymbol{x}^T \boldsymbol{y} \tag{3.2}$$

3.2.2 Naïve Bayes

A Naive Bayes classifier is a probabilistic model based on Bayes' Theorem with the naive assumption of conditional independence between every pair of attribute variables given the value of the target variable. This model is usually used as a benchmark in classification problems. By using Maximum A Posteriori (MAP) estimation, the prediction of the target class ($\hat{y}$) for the input variables x_l is determined as follows [120]:

$$\hat{\mathrm{y}} = argmax\, P(y) \prod_{l=1}^{p} P(x_l | y) \tag{3.3}$$

where $P(y)$ and $P(x_l|y)$ can be simply calculated as the ratio of the number of data points with the target class y to the total number of data points and the ratio of the number of data points with the target class y and the l^{th} attribute variable taking the value of x_l to the number of data points with the target class y, respectively.

3.2.3 Logistic Regression

Logistic regression is a type of nonlinear regression model to capture how a target class varies with the attribute variables in a classification problem. Unlike linear regression that fits a straight line or hyperplane, logistic regression uses the logistic

function to squeeze the output of a linear equation between 0 and 1. It has an S-shaped curve with the following function [122]:

$$y = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 x_{i.1} + \beta_2 x_{i,2} + \cdots + \beta_p x_{i,p}))} \tag{3.4}$$

where β_0 is the intercept and β_1 to β_p are the coefficients of the attribute variables.

3.2.4 K-Nearest Neighbors

K-nearest neighbors is a simple method as a type of instance-based learning in a classification problem that stores instances (data points) of the training dataset and classifies new instances based on a distance or similarity/dissimilarity measure. The prediction is obtained from a simple majority vote of the K nearest neighbors of each data point. The Euclidean distance (d) is a common measure of dissimilarity between two data points x_i and x_j as follows [120]:

$$d(x_i, x_j) = \sqrt{\sum_{l=1}^{p} (x_{i,l} - x_{j,l})^2} \quad , \qquad i \neq j \tag{3.5}$$

The larger this distance is, the more dissimilar the two data points are.

3.2.5 Decision/Regression Tree

A decision tree, or a regression tree in the case of the continuous values of the target variable, is mainly composed of a root node, non-leaf nodes, leaf nodes, and branches. The training data points in the root node are split into two or more sub-nodes based on a splitting criterion. The optimal splitting criterion is calculated at a certain value of a particular attribute variable based on a cost minimization problem. The data points of j^{th} attribute variable in the data at node m (R_m) are divided into two regions R_m^1 and R_m^2 based on Recursive Binary Splitting as follows:

$$R_m^1(\theta) = \{x | x_j \leq s\}, R_m^2(\theta) = \{x | x_j > s\}, \theta = (j, s) \tag{3.6}$$

where x_j represents the j^{th} attribute variable. A cost function (C) is used to obtain the quality of a candidate splitting criterion (θ) as follows [121]. The Gini index and mean squared error are popular functions to measure the quality of a split in a decision tree and a regression tree, respectively. The best splitting criterion can be found as follows:

$$\theta_{best} = argmin_{\theta}\left[\frac{N_m^1}{N_m}C(R_m^1(\theta)) + \frac{N_m^2}{N_m}C(R_m^2(\theta))\right] \quad (3.7)$$

where θ_{best} represents the best splitting criterion. N_m, $N_m{}^1$, and $N_m{}^2$ are the number of data points in R_m, $R_m{}^1$, and $R_m{}^2$, respectively. The procedure of dividing the data points into two subsets is iterated to reach the desired tree depth. Finally, the predicted values of the leaf (terminal) nodes are set to the dominant class (in a decision tree) or the mean target value (in a regression tree) of the leaf nodes.

3.2.6 Random Forest

A random forest is based on the concept of decision/regression trees [122]. The first step is to take several samples from the training data. The number and the size of the samples are hyperparameters that can be adjusted. Moreover, sampling should be done randomly and this can be performed with or without replacement. After producing all the samples, a tree should be trained over each sample data. In this step, all features can be evaluated to build the trees or even a random subset of the features with more than one member can be selected. The purpose of these two sources of randomness is to decrease the variance of the model. After producing all the trees over all samples, the test dataset will be fed into the model to obtain the predicted target values using each tree. In the end, the final predicted target value is calculated by averaging the outputs of all the trees for each data point in the test data. Combining diverse trees in random forests results in a reduced variance, sometimes at the cost of a slight increase in bias. However, the variance reduction is often significant which yields an overall better model.

3.2.7 Gradient Boosting Tree

Gradient boosting tree (GBT) is an ensemble of decision/regression trees built in a stage-wise manner. In the first stage, a tree is developed based on the training data. Then, the training data is fed into the tree and the predicted values of the target variable are obtained. The error corresponding to each data point *i* of the training data is calculated as follows [122]:

$$ErrP_i^1 = MT_i - PT_i^1 \ , \ \forall i \in N \quad (3.8)$$

where $ErrP^1$, MT, PT^1, and N are the error of prediction in stage 1, the measured target value, the predicted target value in stage 1, and the number of the data points in the training data, respectively. In the next stage, a tree will be built to predict the values of

$ErrP^1$. The training data is fed into the new tree to get the predicted errors. The new prediction is calculated as follows [122]:

$$PT_i^2 = PT_i^1 + (LR \times P_i^{ErrP1}) \ , \ \forall i \in N \tag{3.9}$$

where PT^2, P^{ErrP1}, and LR are the predicted value of the target in stage 2, the predicted error in stage 1, and the learning rate which can have a value between zero and one, respectively. Then, the new error of prediction is calculated by subtracting the measured target value from the last stage predicted target value. This procedure continues in iterations until a stopping criterion is met. The number of stages over which the error values and new predictions are obtained can be considered as the stopping criterion so that after a particular number, the procedure does not continue.

3.2.8 Support Vector Machine/Regression

Support vector machine (SVM) is a well-known classification method based on the best separation of the data points from different classes. The best separation is achieved by the hyperplane that has the largest distance to the nearest training data points of any class in a high dimensional space. The generalization error of the classifier is lower when the margin is larger. So, a decision boundary at a particular margin of tolerance from the hyperplane is determined such that the support vectors, as the data points closest to the hyperplane, are within that boundary. In the case of the prediction problem, a linear support vector regression (SVR) defines a linear relationship (as a line or hyperplane) between the vector of the attribute variables ($\boldsymbol{x}_n$) for each data point n and the corresponding value of the target variable as follows [121]:

$$f(\boldsymbol{x}_n) = \boldsymbol{x}_n^T \times \boldsymbol{\beta} + b \quad , \forall n \in N \tag{3.10}$$

The best values of the vector $\boldsymbol{\beta}$ and the scalar b are selected to minimize the expected risk of prediction which is obtained by minimizing both the empirical prediction error and the VC (Vapnik–Chervonenkis)-dimension of the function (a measure of the complexity of the function) at the same time according to the structural risk minimization principle [120]. To this end, a margin of tolerance (ε) is set and a decision boundary at ε distance from the original hyperplane is determined such that the data points closest to the hyperplane (support vectors) are within that boundary. In the case of the classification problem, the goal is to find $\boldsymbol{\beta}$ and b such that the prediction given by $sign(\boldsymbol{x}^T \times \boldsymbol{\beta} + b)$ is correct for most samples. This is formulated as a quadratic optimization problem. In the

case of a nonlinearly separable problem, slack variables ξ_n and ξ_n^* are used for each observation $n \in N$ to avoid the infeasibility of the constraints. Non-negative slack variables are provided to account for training data points that fall outside of the ε-insensitive zone. The inclusion of slack variables leads to the following optimization formulation [121]:

$$\begin{aligned} Minimize \quad & J(\beta) = \frac{1}{2}\beta^T\beta + C\sum_{n=1}^{N}(\xi_n + \xi_n^*) \\ s.t. \quad & y_n - \hat{y}_n \le \varepsilon + \xi_n \quad \forall n \in N \\ & \hat{y}_n - y_n \le \varepsilon + \xi_n^* \quad \forall n \in N \\ & \xi_n \ge 0 \text{ and } \xi_n^* \ge 0 \quad \forall n \in N \end{aligned} \tag{3.11}$$

where y_n and $\hat{y}_n$ are the actual and the predicted target values, respectively. The constant C is a positive value that controls the penalty imposed on observations that lie outside the margin of ε and helps to prevent overfitting. When the problem is nonlinearly separable, the non-linear model can also be implemented since some problems cannot adequately be described using a linear model. A nonlinear model is obtained by using a nonlinear kernel function to transform the p-dimensional data into a higher l-dimensional feature space where data can be trained using a linear model. This concept is shown in Fig. 3.1 [123]. The polynomial and Gaussian functions are among the common non-linear kernel functions.

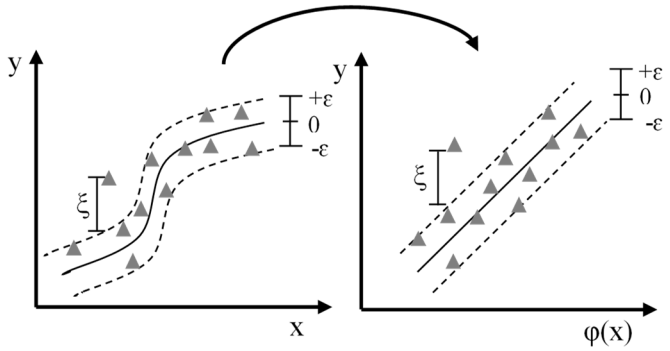


Figure 3.1. Performance of a non-linear kernel-based SVR.

3.2.9 Artificial Neural Network

Artificial neural networks (ANNs) are systems inspired by the organization and functionality of the human brain that are intended to replicate the way that humans learn. ANNs consist of input and output layers, as well as (in most cases) one or more hidden layers with several processing units. The function of a processing unit, with the structure shown in Fig. 3.2, can be formulated as follows [120]:

$$o = f(net_j) = f(\sum_{i=0}^{p} w_{j,i}x_i) \tag{3.12}$$

where o is the output of the unit, x_i is the i^{th} input of the unit, P in the number of inputs, $w_{j,i}$ is the connection weight of input i in unit j, and f is the activation function. In order to design an ANN to map the inputs to the outputs, the following steps should be implemented.

In the first step, the architecture of the model should be determined. This includes choosing the number of inputs, the number of outputs, the number of hidden layers, the number of processing units in each hidden layer, the activation functions of the processing units, and the type of the model including feed-forward ANN or recurrent ANN. In the next step, the connection weights should be determined. The back-propagation learning method is a good choice for this purpose. This method aims at searching for a set of connection weights (including biases) that minimizes the output error. The search is called the gradient descent search and it changes the weights in order to reduce the output error after passing the inputs through the ANN in iterations.

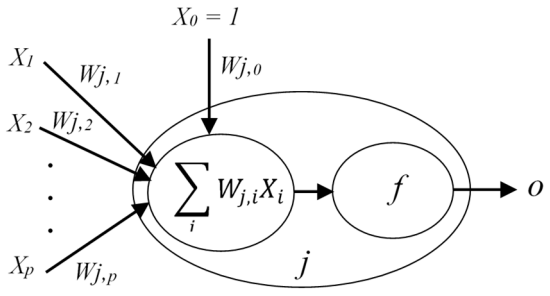


Figure 3.2. A processing unit in an ANN model.

3.2.10 Long Short-Term Memory Network

Long short-term memory (LSTM) network is a type of recurrent artificial neural network that can learn long-term dependencies in the data [124]. It is one of the most promising computational intelligence techniques to use when dealing with a sequence of data in the form of time series. LSTM network, as a recurrent ANN, has a chain-like architecture that contains several cells and gates. Information is retained by the cells and the memory manipulations are done by the gates which include forget gates, input gates, and output gates. The equations for the gates in an LSTM block are as follows [124]:

$$\begin{aligned} i_t^{gate} &= \sigma(\omega_i[h_{t-1}, x_t] + b_i) \\ f_t^{gate} &= \sigma(\omega_f[h_{t-1}, x_t] + b_f) \\ o_t^{gate} &= \sigma(\omega_o[h_{t-1}, x_t] + b_o) \end{aligned} \tag{3.13}$$

where i^{gate}, f^{gate}, and o^{gate} are the outputs of the input gate, forget gate, and output gate, respectively. σ, ω, and b represent sigmoid function, weights of the gates, and biases of the gates. x is the input while h is the output of the block. By using the sigmoid function, 0 means the gates are blocking everything, and 1 means the gates are allowing everything to pass through them. The equations for the cell state (c), candidate cell state ($\hat{c}$), and the final output (h) are as follows [124]:

$$\begin{aligned} \widehat{c_t} &= \tanh(\omega_c[h_{t-1}, x_t] + b_c) \\ c_t &= f_t^{gate} \times c_{t-1} + i_t^{gate} \times \widehat{c_t} \\ h_t &= o_t^{gate} \times \tanh(c_t) \end{aligned} \tag{3.14}$$

After calculating the candidate cell state and then the cell state, the cell state knows at any time step t that what it should forget from the previous state (i.e. $f_t^{gate} \times c_{t-1}$) and what it should consider from the current state (i.e. $i_t^{gate} \times \hat{c}_t$). Finally, the cell state is filtered and passed through the activation function which determines what portion should appear as the output of the LSTM block at time step t. The architecture of the LSTM network is shown in Fig. 3.3 [124].

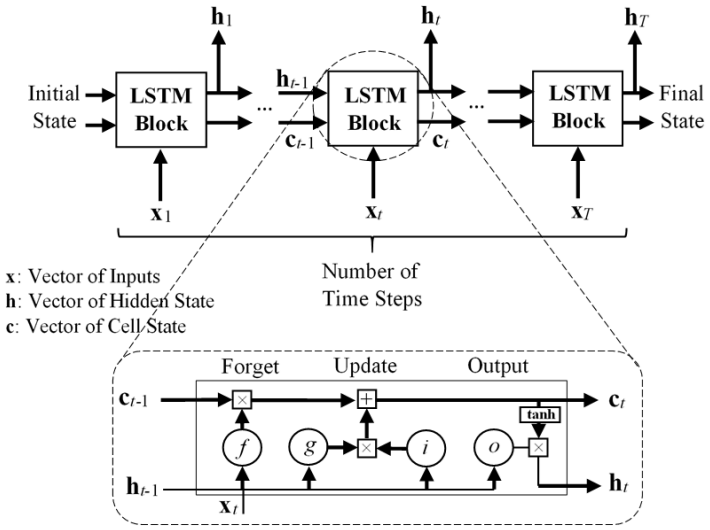


Figure 3.3. The architecture of LSTM networks.

3.2.11 Performance Metrics

The performance analysis of the developed models can be carried out by using the assessment metrics. Some of these metrics, that are used in this thesis, are presented in the following [122]. In all the equations, P^{ACT} and P^{PRD} are measured and predicted values of the target variables, respectively. N is the total number of data points over which the predictions are made.

- Mean Squared Error (MSE):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(P_i^{PRD} - P_i^{ACT})^2 \tag{3.15}$$

- Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i^{PRD} - P_i^{ACT})^2} \tag{3.16}$$

- Mean Absolute Error (MAE):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|P_i^{PRD} - P_i^{ACT}\right| \tag{3.17}$$

- Mean Bias Error (MBE):

$$MBE = \frac{1}{N}\sum_{i=1}^{N}(P_i^{PRD} - P_i^{ACT}) \tag{3.18}$$

- Accuracy:

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{N} \tag{3.19}$$

All the aforementioned metrics, except Accuracy that is used in the case of a classification problem, are used for assessing the performance of regression models. The lower value of MSE, RMSE, MAE, and MBE implies the higher accuracy of a regression model. However, a higher value of Accuracy is considered desirable. Unlike MAE, MSE and RMSE penalize large prediction errors and are most useful when large errors are undesirable. Unlike MSE, both RMSE and MAE have the same unit as the target variable and are more interpretable. The greater the difference between them, the greater the variance in the individual prediction errors. If RMSE is equal to MAE, then all the prediction errors are of the same magnitude. MSE is a differentiable function that makes it easy to perform mathematical operations in comparison to a non-differentiable function like MAE. Therefore, in many models, RMSE is used as a default metric for calculating loss functions. MBE indicates the average bias in the model to recognize if any steps need to be taken to correct the model bias.

CHAPTER 4

COMPUTATIONAL INTELLIGENCE-BASED PREDICTION OF SNOW LOSS

4.1 Introduction

Snow loss, as the power loss caused by a snow cover on the PV panels, is a very complex phenomenon in PV systems that not only depends on the local weather conditions but also more complicated parameters like the shape and depth of the snow layer on every single panel, arrays' tilt angle, albedo radiation, etc. This is the reason why this chapter copes with this complexity by formulating a prediction problem using computational intelligence techniques (that are able to capture inherently this complexity), while also capitalizing on the widely available data. In order to calculate the snow loss values for a PV system based on the historical data of the system, a methodology is developed in this chapter. The historical dataset should contain the measurement records of the technical parameters of the PV system including the output voltage and current (or the output power) of the inverters/PV modules as well as the module temperature together with the on-site measurements of plane-of-array irradiance on the surface of the modules. In the following, the dataset used to conduct this study, the methodology implemented to estimate snow loss values, and the developed snow loss prediction models will be explained.

4.2 System Under Study

This study is carried out on the historical data of a PV farm located in Ontario, Canada. The hourly data records of the electrical parameters of the farm including the output voltage (V) and the output current (A) of all inverters as well as module temperatures (°C) together with the meteorological on-site measurements including irradiance (W/m^2), ambient temperature (°C), and wind speed (m/s) are collected for a 4-year period. Data records of humidity (%) and snowfall (cm) are also obtained from the historical databases provided by the Canadian government website [125]. Then, the outliers are detected and removed. Outliers, as data points that differ largely from the norm of the data, tend to produce large errors during the training process. In the first step of the outlier detection, some incorrect data points, which can be easily detected by

visualizing and considering the nature of each aforementioned input variable, are removed. Then, considering the nature of the variables as time series, cumulative sum (CUSUM) control charts are used to detect outliers for all variables [120]. This method not only detects outliers but also large and small shifts in the variable mean. Unconditional mean is also used for filling the missing data.

4.3 Snow Loss Calculation

In order to calculate the snow loss values, a 3-stage model shown in Fig. 4.1 is implemented. In the first stage, the expected power outputs of PV modules are calculated using a standard equation [126]:

$$P_{exp} = P_{stc} \times (1 - T_{coef}(25 - T)) \times \frac{G}{G_{stc}} \tag{4.1}$$

where P_{exp} is the expected output power of the module in W, P_{stc} is the standard test condition output power of the module in W, T_{coef} is the temperature coefficient in 1/°C, T is the cell temperature in °C, G is the irradiance in W/m^2, and G_{stc} is the standard test condition irradiance of the module in W/m^2 which equals 1,000. Then, the actual generated power of the modules is calculated using the following equation:

$$P_{act} = U \times I \tag{4.2}$$

where P_{act} is the actual generated power in W, U is the output voltage of the module in V, and I is the output current of the module in A. After calculating the expected and the actual output power for each module, the total power loss is obtained in the second stage using the following equation:

$$L_{total} = P_{\text{exp}} - P_{act} \tag{4.3}$$

where L_{total} is the power loss caused by soiling/snow, DC cabling, the maximum power point tracking (MPPT), low irradiances, and mismatching caused by the interconnection of the solar cells or modules which do not have identical properties or which experience different conditions from one another. In the third stage, the following equation is used to obtain the snow loss of each module and finally, the values calculated for all modules are added to present the snow loss of the PV farm.

$$L_{snow} = L_{total} - C \times P_{exp} \tag{4.4}$$

where L_{snow} is the snow loss in W and C is a constant representing the percentage of losses from DC cabling, the MPPT, low irradiances, and mismatching and is considered here as 6.6%.

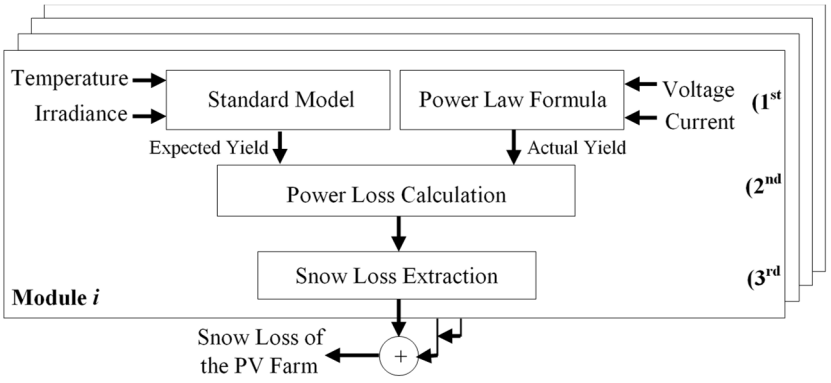


Figure 4.1. 3-Stage model for snow loss calculation.

After obtaining the snow loss values of the farm over the 4-year period, daily average values of all parameters are calculated and the data is normalized between 0 and 1 using the Min-Max formulation as follows:

$$\widehat{x_i} = \frac{x_{min} - x_i}{x_{min} - x_{max}} \tag{4.5}$$

where $\widehat{x_i}$ is the normalized value of x_i and x_{Min} and x_{Max} are the minimum and the maximum values of x, respectively.

The final prepared dataset contains 1,461 data observations of daily average values for 6 variables including snow loss, snowfall, irradiance, ambient temperature, humidity, and wind speed. Line graphs illustrating the components of the data are shown in Fig. 4.2 for all the parameters. The seasonality in the data can be easily recognized. The graph of irradiance (Fig. 4.2(b)) also shows a great variation which is normal regarding the nature of this parameter. As indicated in Fig. 4.2(e) and Fig. 4.2(f), there is daily snow loss in the system if there is snowfall on that day or on previous days due to snow accumulation. The correlation values are presented in Table 4.1 which shows how the pairs of the parameters are correlated. As can be seen, the most correlated parameter to the snow loss is the snowfall with a positive correlation.

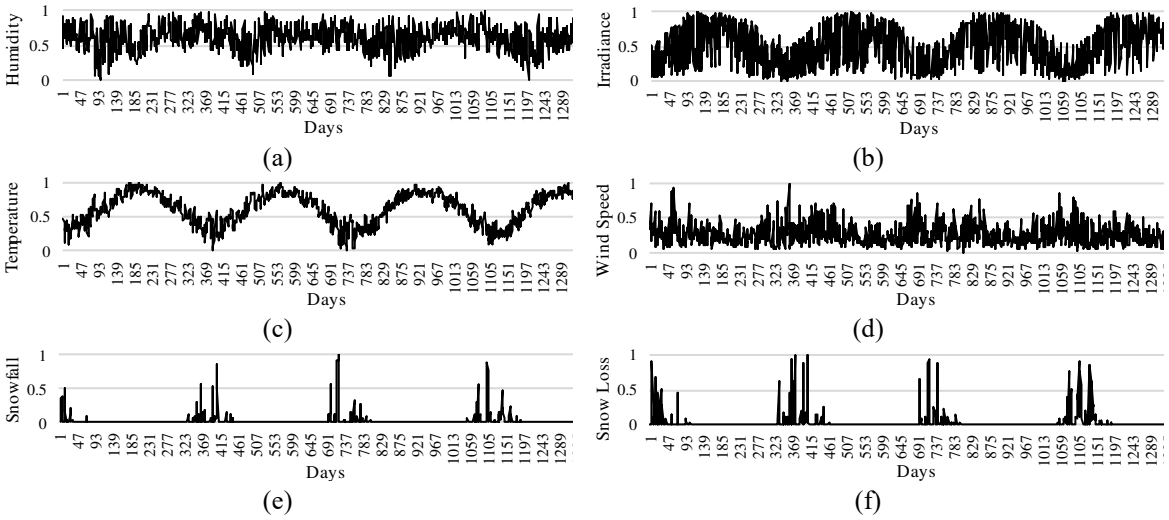


Figure 4.2. Line graphs of the normalized data: (a) humidity, (b) irradiance, (c) ambient temperature, (d) wind speed, (e) snowfall, and (f) snow loss.

Table 4.1. Linear correlation values for different pairs of variables in the final dataset

	Humidity	Irrad.	Temp.	Wind Speed	Snowfall	Snow Loss
Humidity	1					
Irradiance	-0.658	1				
Temperature	0.026	0.364	1			
Wind Speed	0.020	-0.211	-0.161	1		
Snowfall	0.072	-0.136	-0.342	0.037	1	
Snow Loss	0.081	-0.152	-0.388	0.012	0.771	1

4.4 Snow Loss Prediction

The final prepared dataset with the daily values of the variables is used to build computational intelligence-based snow loss prediction models using machine learning techniques. This is a supervised learning approach, with a multiple univariate regression problem with 5 features including humidity, irradiance, ambient temperature, wind speed, and snowfall, and one numeric target variable which is the snow loss. The techniques used to build the prediction models are regression tree, random forest, gradient boosting tree, support vector regression, artificial neural network, and long short-term memory network. These techniques together with a linear regression model, that is used as a benchmark, are applied to the dataset. Finally, the performance of the models is compared to find the best snow loss predictor.

4.5 Numerical Results

The data corresponding to 5 features (humidity, irradiance, ambient temperature, wind speed, and snowfall) and one target variable (snow loss) is used to develop the

models. Two cases are studied as follows: As a popular approach, in case 1, the first 75% of the data (the first 3 years' records) is used to train the models and the rest 25% is used to test them. Since the available dataset is relatively small and training over 75% of the data in the first case may lose some important information in the rest of the data, a second case is also investigated where the whole data can contribute to training the models. In case 2, 5-fold cross-validation is performed over the whole data; as such, the dataset is divided randomly into 5 segments with equal sizes, and from 5 different combinations of the segments, 4 segments are used to train the model and the other one is used to test the generated model. Finally, the performance score of the system is obtained by averaging the scores of the models over all 5 combinations. In order to obtain the optimal hyperparameters, a search is implemented for the models in both cases. The performance analysis of the models is performed in this study by using the MSE and the MBE metrics. Models are developed in MATLAB software.

4.5.1 Case 1: Model Development and Evaluation for 75%-25% Holdout Case

The summary of all the developed models together with the corresponding hyperparameters giving the minimum MSE over the test data and the MBE of each model output is presented in Table 4.2. The models are discussed in the following.

- **Models Based on Linear Regression**

In the case of linear regression with one variable, the minimum MSE with the value of 0.0116 belongs to the model built using snowfall values only, since according to Table 4.1, this feature is the most correlated one to the target variable. C_1 and C_0 are the slope and the intercept of the fitted line, respectively and are presented in the second row of the table. When using more than one feature in the model, the minimum MSE with the value of 0.0110 belongs to the model built using all features, where MSE is a little reduced by 0.0006 compared to the previous model. C_1 to C_5 are coefficients of snowfall, temperature, irradiance, humidity, and wind speed, respectively and are presented in the third row of the table. Linear regression models are widely used as a benchmark in machine learning to evaluate the performance of other complicated models. Hence, they provide minimal performance to be accepted by a classifier/predictor since a line or a multi-dimensional plane is simply fitted on the data to represent the relationship between variables.

Table 4.2**.** Performance metrics and optimal hyperparameters for snow loss prediction models in case 1

Model		**Min. MSE on Test Data**	**MBE on Test Data**	**Parameters**						
Linear Regression	Single Variable	0.0116	-0.006	Using Snowfall only	C_0 = 0.0086	C_1 = 1.1360				
	Multiple Variables	0.0110	-0.007	Using All Features	C_0 = 0.0117	C_1 = 1.0796	C_2 = -0.0743	C_3 = 0.0281	C_4 = 0.0519	C_5 = -0.0059
Regression Tree		0.0067	-0.001	Depth = 3						
Gradient Boosting Tree		0.0052	-0.003	Error Cal. = 17 Iterations	*LR* = 0.60	Depth = 3				
Random Forest		0.0064	-0.005	Number of Samples = 10000	Using all features	Depth = 8				
ANN	One-Layer	0.0094	-0.007	1 Processing Unit	Activation Function = Sigmoid	Number of Epochs = 500	*LR(k) = 1 /(1+(k/35))*		Stochastic Gradient Descent	Configuration = 5-4-3-2
	Multi-Layer	0.0084	-0.006	4 Hidden Layers						
	LSTM	0.0062	0.005	17 Hidden Units	Gate Act. Function = Sigmoid	State Act. Function = tanh	*LR(k) = 0.1 / ((k/34) × 2)*		Number of Epochs = 274	Mini-Batch Size = 3
SVR	Linear	0.0108	0.001	Kernel = Linear	Kernel Scale = 0.39	ε = 0.032	*C* = 0.02	*b* = 0.019		
	Non-Linear	0.0072	-0.006	Kernel = Gaussian	Kernel Scale = 0.33	ε = 0.031	*C* = 3.20	*b* = 0.14		

➢ **Models Based on Regression Tree**

Several regression trees with different maximum depths were developed and their performance was investigated. MSE values for the test and training data corresponding to each tree are illustrated in Fig. 4.3. As can be observed, the minimum MSE is 0.0067 which is obtained using a 3-layer tree. Using deeper trees leads to an increase in MSE due to overfitting. This model works better compared to the linear regression because, considering the nature of tree-based models, it performs a comprehensive analysis of the consequences along each branch and considers all possible outputs of a decision. The built-in feature selection process of tree-based models also has provided the superiority of these models for the problem under study.

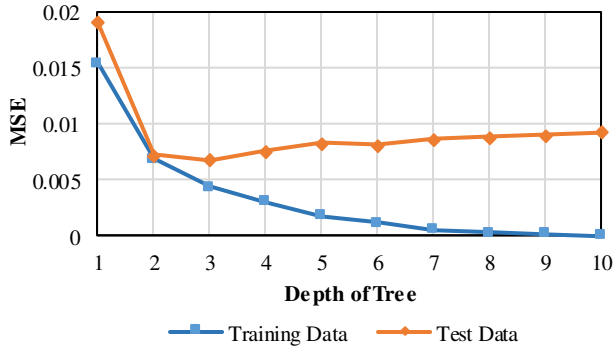


Figure 4.3. MSE for the test and training data using regression trees.

➢ **Models Based on Gradient Boosting Tree**

Different gradient boosting trees were developed using different values of hyperparameters, including the number of iterations for calculating and modeling the errors, learning rate, and depth of the trees. Finally, the minimum MSE using the best model is obtained as 0.0052 which is 0.0015 lower compared to the regression tree model. As can be seen in Table 4.2, this model has the best performance and the minimum MSE among all models. As it has been proven in many applications, gradient boosting trees usually have an outstanding performance compared to many other methods. That is because they build an ensemble of shallow and weak successive trees, where each tree improves the performance of the previous one. Fig. 4.4 shows the MSE values for the test data using the optimal LR = 0.6 for 1- to 5-layer trees and different numbers of iterations. The minimum MSE is obtained using a 3-layer tree with 17 iterations of error calculation. Using deeper trees overfits the data. Moreover, trees with

lower depth cannot model the data very well. Increasing the number of iterations of error calculation than a particular number also increases MSE due to overfitting.

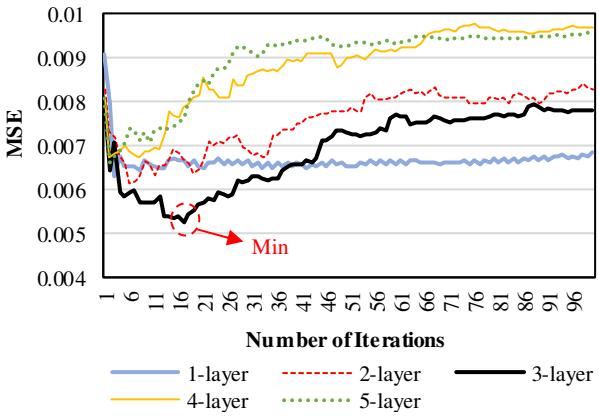


Figure 4.4. MSE for the test data using gradient boosting tree with *LR* = 0.6.

- **Models Based on Random Forest**

Implementing random forest requires a number of random samples from the training data (data samples with the same size of the training data that are randomly chosen with replacement are considered) which is large enough to represent all characteristics of the main dataset. Building models with different numbers of random samples showed that 10,000 is the best choice for this parameter. Moreover, the best performance was obtained when all features were used to build the models. As shown in Fig. 4.5 and presented in Table 4.2, the minimum MSE over the test data using the best random forest model is 0.0064. This model is designed using 10,000 samples with random selection (with replacement) and the maximum depth of 8 for the trees built over the samples when all features are available to produce the trees. As can be seen, this model also works well and overfitting does not happen due to the nature of this technique. However, it has a little bit weaker performance compared to the gradient boosting tree because unlike random forests, gradient boosting trees start the combination process of the results at the beginning, instead of at the end and this helps to improve the shortcomings of the produced weak learners.

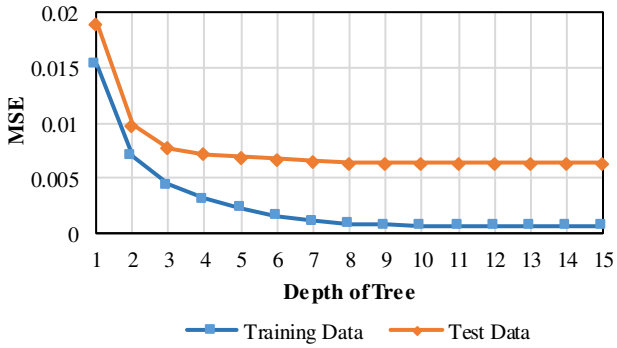


Figure 4.5. MSE for the test and training data using random forest with 10,000 samples and considering all features.

➢ Models Based on Artificial Neural Network

Various ANN models with different structures were developed and their performance was investigated. The first one is a simple feed-forward ANN with one hidden layer and one processing unit. MSE values for the test data using this structure and different types of activation functions, including sigmoid, tanh, LReLU, and swish, are shown in Fig. 4.6. As it is expected, sigmoid achieves a better performance among all the others. The minimum MSE obtained using this model is 0.0094. The other model is a multilayer ANN. As presented in Table 4.2, the minimum MSE using this model is 0.0084 which was obtained using 4 hidden layers in a triangle shape with 5, 4, 3, and 2 processing units in each layer. In this model, the sigmoid activation function achieved the best performance over the dataset, too. Moreover, the stochastic learning approach obtained better results compared to the batch and mini-batch learning for both models. A decreasing learning rate, as presented in Table 4.2, was applied to both models which reduces according to the iteration number (k).

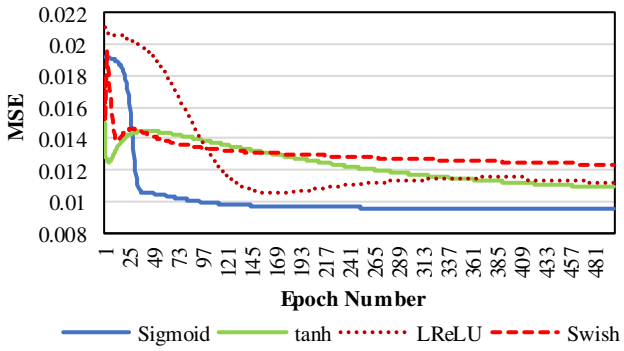


Figure 4.6. MSE for the test data using feed-forward ANN with one hidden processing unit and different activation functions.

Considering the nature of the data, an LSTM network is developed as well. This model, as a recurrent neural network, may achieve better performance when dealing with data in sequence. As presented in Table 4.2, the minimum MSE obtained using this model is 0.0062 which is 0.0032 lower than the one-layer ANN and 0.0022 lower than the multi-layer ANN. This value is still 0.001 higher than that of the gradient boosting tree; however, the LSTM network performs well since we are dealing with time series in this application. It shows that considering the correlation between data records is of significant importance. Feed-forward ANNs also perform poorer compared to tree-based models over the dataset because they usually require training over a large dataset to produce an accurate result.

- **Models Based on Support Vector Regression**

Both linear and non-linear SVR models were developed for the dataset. Considering that the problem is nonlinearly separable, the linear SVR does not have good performance over the data so that the minimum MSE over the test data obtained using this model is 0.0108. The parameters of this model are presented in Table 4.2. By implementing non-linear SVR using both Gaussian and polynomial kernel functions, it was found that the model built using Gaussian function can better reduce the error of prediction compared to the linear SVR. MSE value for the best non-linear SVR is 0.0072 which is 0.0036 lower than that of the linear SVR.

As it was discussed and can be seen in Table 4.2, the best three models developed over the dataset are those built using gradient boosting tree, LSTM, and random forest, respectively. The actual and the predicted snow loss values (as a percentage of the expected power output of the PV modules) for the test data using these three best models are indicated in Fig. 4.7. We can notice that all three models have acceptable performance in predicting the daily snow loss values for one year ahead. Moreover, the MBEs on the test data for the models in Table 4.2 show a small range of positive/negative values (considering the normalized data between 0 and 1) from the optimal MBE = 0. The positive/negative values indicate an accumulated overestimation/underestimation of the outputs on the test data. The table shows that the annual snow loss of the PV system can be predicted with acceptable accuracy.

It is important to note that removing that part of the data which corresponds to the months that never experience snowfall during the 4-year period and retraining the

prediction models using the new dataset does not have a considerable effect on the performance of the models in predicting daily snow loss values and the priority of the models in terms of accuracy; however, the MSE values won't be the same as those presented in Table 4.2. Therefore, we preferred to build the prediction models considering the whole year data which has a better generalization.

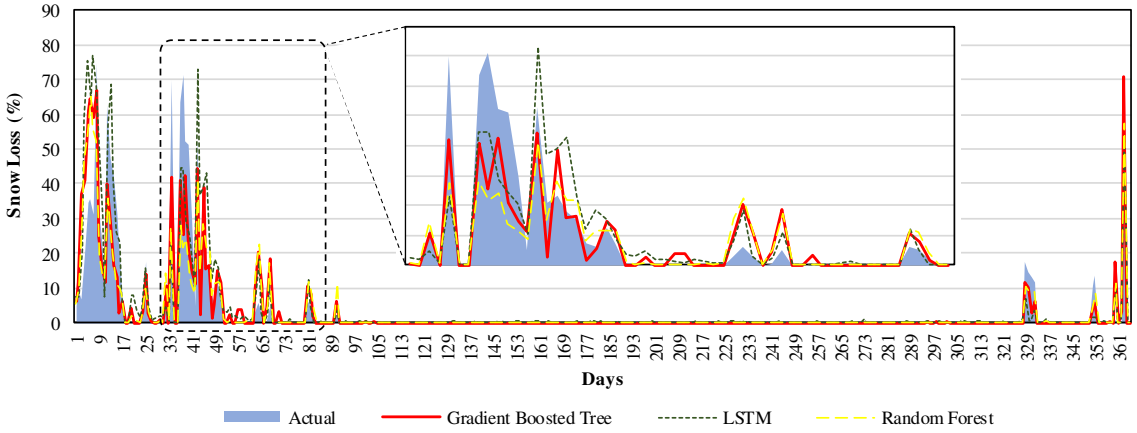


Figure 4.7. Actual and predicted snow loss for the test data using three best models in case 1.

4.5.2 Case 2: Model Development and Evaluation for Cross-Validation Case

In this section, results corresponding to the 5-fold cross-validation implementation over 4 best models including the regression tree, gradient boosting tree, random forest, and non-linear SVR are investigated. Considering the nature of the LSTM network, cross-validation cannot be applied to it. The MSE values of cross-validation together with the parameters of the models are presented in Table 4.3. The parameters are optimized over the test datasets to achieve the minimum prediction error.

Table 4.3. Performance metrics and optimal hyperparameters for snow loss prediction models in case 2

Model	Regression Tree	Gradient Boosting Tree	Random Forest	Non-Linear SVR
Min. MSE on Test Data	0.0053	0.0042	0.0049	0.0056
MBE on Test Data	0.0001	-0.0005	0.0004	-0.0011
Parameters	Depth = 3	Error Cal. = 39 iterations	Number of Samples =10000	Kernel = Gaussian
		LR = 0.4	Using all features	Kernel Scale = 1.66
		Depth = 3	Depth = 6	ε = 0.016
				C = 903
				b = -2.14

As can be seen in Table 4.3, the minimum MSE for the regression tree model is obtained using a 3-layer tree which equals 0.0053 and is 0.0014 lower than that of the same model in case 1 (Table 4.2). Hence, the accuracy of this model is higher than that developed in case 1. Cross-validation is also applied to the model developed using the gradient boosting tree. Using 3-layer trees together with a higher number of iterations of error calculation gives noticeably better performance compared to the related model in case 1. MSE value of this model for the test data is 0.0042, 0.001 lower than that of the same model in case 1 and the minimum MSE over all other models in case 2. Moreover, the model developed using the gradient boosting tree has the best performance among all models in both cases. The random forest is the other model implemented using cross-validation. The MSE value using this model is 0.0049 which shows a better performance compared to the same model in case 1. The final model in this case is made using a non-linear SVR. The MSE value of cross-validation using this model is 0.0056 which is higher compared to the others in this case and has a poorer performance. The MBE values presented in the third row of Table 4.3 also show that models built in this case provide predictions with significantly lower MBE compared to those in case 1, which is a result of better prediction performance.

It can be easily observed from the results presented in Table 4.2 and Table 4.3 that the models developed by applying cross-validation in case 2 have better performance compared to those in case 1. Since 25 percent of the data is not used to build the models in case 1, the information in that part of the data is not utilized in the training procedure; while applying cross-validation gives the opportunity to the whole data to contribute to developing the model. By comparing the performance of different models in both case 1 and case 2, it can be found that the best model for predicting the daily snow loss over the prepared data is the one designed using gradient boosting trees in both cases. Applying cross-validation to this model achieves the minimum MSE among all models developed in this study. It is worth mentioning that in case of implementing 4-fold cross-validation, dividing the whole data into 4 folds instead of 5 folds, each fold containing 25 percent of the data to be similar as much as possible to the 75%-25% holdout in case 1, the new MSE values obtained using the new trained models are very similar to those presented in Table 4.3.

The procedure implemented in this study to build snow loss prediction models can be easily applied to any other PV system with similar or different characteristics.

However, the models' parameters may be different considering various characteristics of the PV system like tilt angle, arrays' surface material, etc. Moreover, in case of including other input variables, the models should be retrained to ensure a better adaptation to the new context; however, the process itself remains the same. It is also important to mention that in a real PV farm, as data becomes available, one will have to retrain the models to capture new trends and thus improve the prediction capability.

4.6 Conclusion

Snow loss calculation and prediction for a PV farm were addressed in this chapter. This energy loss is mainly due to the accumulation of snow on the surface of PV panels which can significantly reduce the energy yield of the system. Calculation of the snow loss for a PV farm located in Ontario, Canada having 4-year historical data corresponding to different technical and meteorological parameters was performed using a 3-stage model. Moreover, several snow loss prediction models were developed using machine learning algorithms and their performance was compared in two cases including 75%-25% holdout and cross-validation. The dataset consists of five meteorological parameters and the snowfall is the most correlated one to the snow loss. Five machine learning algorithms including regression tree, gradient boosting tree, random forest, feed-forward and recurrent artificial neural networks, and support vector regression were used to develop snow loss prediction models. According to the numerical results, the model built using the gradient boosting tree achieved the best performance with the MSE value of 0.0052 in the first case. The LSTM network is the second-best choice with an MSE value of 0.0062. In the cross-validation case, the best performance again belongs to the model developed using the gradient boosting tree with the MSE value of 0.0042. However, there is a small difference (0.0007) between this value and the one obtained using the random forest. Applying cross-validation on the models gives better results compared to the 75%-25% holdout case due to developing the models using the whole dataset.

CHAPTER 5

COMPUTATIONAL INTELLIGENCE-BASED PREDICTION OF PV SYSTEM POWER LOSSES

5.1 Introduction

The performance of a PV system depends on different types of electrical losses which are mainly categorized into array capture and system losses. In fact, these losses are very complex in nature and the values of each one depend on many electrical and meteorological factors among which some may not be available or cannot even be normally measured. Not only the data availability is the main issue in calculating the accurate values of each type of power loss in a PV system, but also a very comprehensive analysis is required to extract the exact values of each one from the total calculated power losses. In order to cope with this complexity and provide a reliable solution for obtaining a fair estimation of the future values of the losses, computational intelligence-based modeling of each type of power loss using the available data is proposed and investigated in this chapter. The developed models not only can capture the complex nature of the losses to simplify the next implementations but also capitalize only on the widely available data as the inputs to the models to predict the future values of the losses.

In order to calculate various types of power losses for a PV system based on its historical data, a methodology is implemented in this chapter. Then, several computational intelligence-based models are developed for each type of power loss. The PV system under study, the proposed power losses calculation methodology, and the developed power losses prediction models are explained in the following.

5.2 System Under Study

The PV system under study is a rooftop installation located in Denver, Colorado. The historical data of the main electrical and meteorological parameters of the system over an 8-year period from July 2012 to Dec. 2019 is extracted from the website of the National Renewable Energy Laboratory (NREL) [127]. This system consists of 18 panels tilted at 40°, produced by Shell Solar company. The panel type is “Eclipse 80” and the

nominal output power of each one is 80 W. Hence, the nominal output power of the overall system is 1,440 W. The manufacturer of the inverter used in this system is SMA Solar Technology. The extracted dataset consists of the measured values of the main parameters including AC current (A), AC voltage (V), and AC power (W) in the output side of the inverter, DC current (A), DC voltage (V), and DC power (W) in the input side of the inverter, ambient temperature (°C), module temperature (°C), and plane-of-array (POA) irradiance (W/m^2). The measurements are taken every minute. In addition, the daily values of the snowfall, snow depth, and average wind speed near the PV site are extracted from the website of the National Ocean and Atmospheric Administration (NOAA) [128].

As the first step of the data preparation, the available data was observed using visualization tools and the outliers, as records that differ largely from the norm of the data for a specific parameter, were detected and removed from the dataset. Then, the incorrect data were recognized according to the nature, the normal values, and the acceptable range of the variations over time and finally omitted. CUSUM control charts were used for this purpose which not only detect incorrect data but also shifts in the mean of the parameter [120]. Moreover, there were many missing values in the dataset; as such, some days or even months were without recorded measurements and hence, eliminated from the final dataset. After applying the data cleaning process, the total number of the data points in the main dataset is 3,608,818 (minutes of recorded measurements of the parameters).

5.3 PV Power Losses Calculation

Calculating the values of different types of power losses for a PV system based on a historical dataset of the aforementioned electrical and meteorological parameters needs a systematic approach to achieve as accurate results as possible. Hence, a systematic PV system power losses calculation approach is presented in this chapter which provides an estimation of the various types of losses in the system based on a dataset of the main parameters. The block diagram of this approach is shown in Fig. 5.1. Each block of the diagram is described in the following subsections.

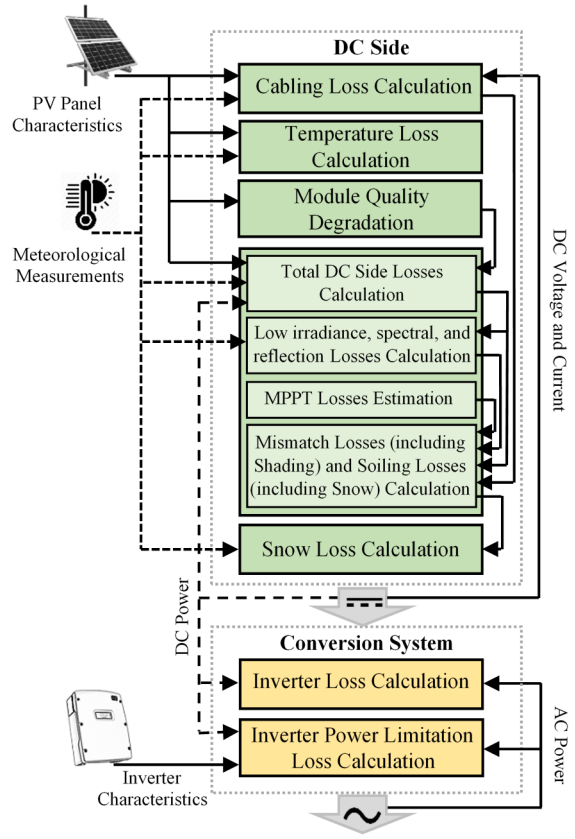


Figure 5.1. The block diagram of the proposed approach of PV system power losses calculation.

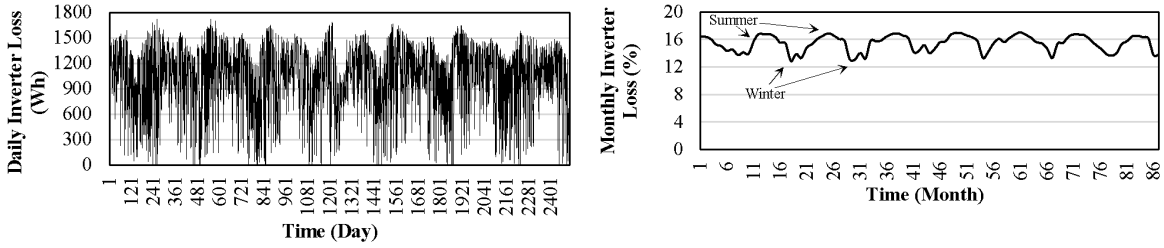
5.3.1 Inverter Loss Calculation

Inverters convert DC power provided by arrays of PV modules to AC power compatible with the utility power grid. The inverter loss contains the switching and the ohmic losses in the switching devices through which PWM techniques are applied to the inverter. The inverter loss can be obtained using the following equation:

$$P^{Inv\ Loss} = P^{Inv\ Input} - P^{Inv\ Output} \tag{5.1}$$

where $P^{Inv\ Loss}$, $P^{Inv\ Input}$, and $P^{Inv\ Output}$ are the power loss, the input power (DC side), and the output power (AC side) of the inverter, respectively. For the system under study, the values of the inverter loss have been calculated for each data point and averaged over each hour. Then, the daily and the monthly values of the inverter loss have been obtained by adding all the hourly values over each day (total of 2,502 days) and month (total of 87 months), respectively. The line graphs of the daily inverter loss and the monthly percentage of the inverter loss over the 8-year period under the study are shown in Fig.

5.2 (a) and (b), respectively. As one can notice, the inverter loss has a greater value in summer in comparison to winter.



(a) (b)

Figure 5.2. Line graphs of (a) the daily inverter loss and (b) the monthly percentage of the inverter loss over the 8-year period for the PV system in Denver.

5.3.2 Inverter Power Limitation Loss Calculation

Today's PV systems typically show an inverter being underrated with respect to the installed module power. This design option is chosen intentionally, as a PV generator rarely operates at its nominal power. However, under some cold and sunny conditions, the inverter might limit the system output power. Inverter power limitation loss can be simply detected by visualizing the input and output power of the inverter using a line graph. In the case of a limitation imposed by the inverter, a cap at the value of the maximum power of the inverter will be placed on the values of the output power over the period of the year when the maximum power production is expected. By investigating the dataset of the PV system under study, it has been found that the value of this type of power loss is zero and the power limitation by the inverter never happened over the 8-year period.

4.3.3. DC and AC Cabling Losses Calculation

In the case that only the measurements of the inverter-side parameters are available, a technique is developed to estimate the DC wiring losses of the PV system. Most of the inverters used in PV systems include a maximum power point tracker within their controllers to track the maximum power point of their input. For a given temperature and insolation level, the maximum output current (I_m) and voltage (V_m) of a PV module can be estimated as follows [129]:

$$I_m = (C_0 G/G_{STC} + C_1 (G/G_{STC})^2) \times (I_{mo} + K_{Im}(T_c - 25)) \tag{5.2}$$

$$V_m = V_{mo} + C_2 a V_t \ln(G/G_{STC}) + \frac{C_3 (a V_t \ln(G/G_{STC}))^2}{N_s} + K_{Vm}(G/G_{STC})(T_c - 25) \tag{5.3}$$

where $V_t=N_skT_c/q$, k=1.3806503×10^{-23} J/˚C is the Boltzmann constant, q=1.60217646×10^{-23} C is the electron charge, and T_c and G are the PV module temperature in ˚C and POA irradiance in W/m^2, respectively. G_{STC} is the irradiance at the standard test condition (STC) equal to 1,000 W/m^2. Other parameters of the model can be extracted from the manufacturer's datasheet of the PV system which are shown in Table 5.1 [130]. The potential difference over the DC wiring (ΔV_{dc}) is calculated by subtracting the measured DC voltage from the calculated maximum power point voltage as follows:

$$\Delta V_{dc} = V_m \times N_s - V_{dc,mes} \tag{5.4}$$

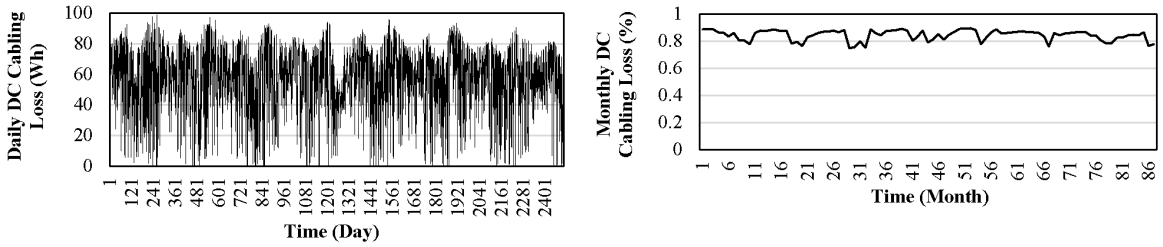
where $V_{dc,mes}$ is the measured DC voltage at the input of the inverter. The DC voltage and the DC current at the input side of the inverter are available from the dataset. The percentage of DC cabling loss ($P^{DC\ Cabling\ Loss}$) can be then estimated as follows:

$$P^{DC\ Cabling\ Loss} = \frac{\Delta V_{dc} \times I_{dc,mes}}{V_m \times I_m \times Q_{deg}} \tag{5.5}$$

where $I_{dc,mes}$ is the measured DC current at the input of the inverter. Q_{deg} is a coefficient to represent the quality degradation of the PV modules. Several data points of the dataset are examined in the aforementioned methodology to obtain an average value for the DC cabling loss which is around 1% of the measured DC power. The line graphs of the daily DC cabling loss and the monthly percentage of the DC cabling loss over the 8-year period are shown in Fig. 5.3 (a) and (b), respectively. Since the PV system under study is a small-scale rooftop system, the DC cabling loss is relatively small compared to that of large-scale PV farms.

Table 5.1. Parameters of the DC cabling loss calculation for the PV system in Denver

Parameter	Value	Parameter	Value
a	1.452	K_{Vm}	-0.42 %/˚C
C_0	1.002	K_{Im}	-0.09 %/˚C
C_1	-0.002	N_s	6
C_2	-0.5279	V_{mo}	33.2 V
C_3	-9.615	I_{mo}	2.41 A



(a) (b)

Figure 5.3. Line graphs of (a) the daily DC cabling loss and (b) the monthly percentage of the DC cabling loss over the 8-year period for the PV system in Denver.

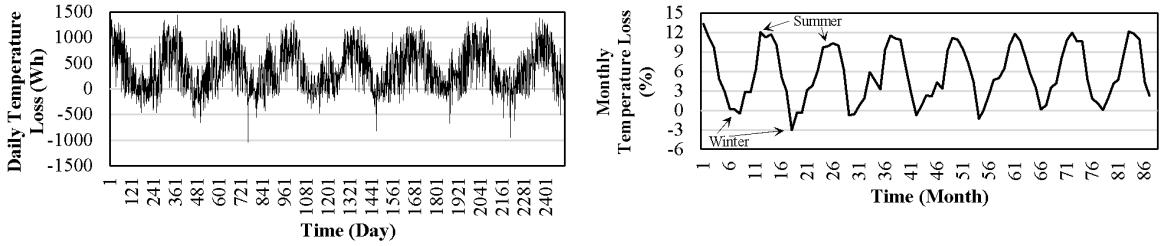
In the case that the measurements of the input current and voltage of the final step-up transformer are available, the AC cabling loss can be simply calculated by multiplying the current by the voltage difference between the output side of the inverter and the input side of the transformer. The AC cabling loss cannot be calculated using the available dataset under study since we do not have the data for the input/output current and voltage of the final step-up transformer.

5.3.4 Temperature Loss Calculation

The temperature of PV panels depends on several factors such as ambient temperature, irradiance level, and wind speed. The operating temperature plays a key role in the PV conversion process. Both the electrical efficiency and the power output of a PV module depend on the operating temperature. The performance of solar cells decreases significantly as the temperature increases. On the contrary, lower temperatures are known to improve the efficiency of the PV modules. The power generation variation due to temperature also depends on the type of the solar panel. Losses caused by the temperature ($P^{Temp\ Loss}$) are calculated as follows [25]:

$$P^{Temp\ Loss} = P_{STC} \times \left(\frac{G}{G_{STC}}\right) \times Q_{deg} \times C_{Temp} \times (25 - T_c) \tag{5.6}$$

where P_{STC} is the maximum power of the PV system at STC and C_{Temp} is the temperature coefficient. Therefore, temperature losses will be positive when the module temperature exceeds 25˚C; otherwise, the losses are negative. The line graphs of the daily temperature loss and the monthly percentage of the temperature loss for the PV system under study over the 8-year period are shown in Fig. 5.4 (a) and (b), respectively. As can be observed, the temperature loss is negative or close to zero in winter while it has a large positive value in summer.



(a) (b)

Figure 5.4. Line graphs of (a) the daily temperature loss and (b) the monthly percentage of the temperature loss over the 8-year period for the PV system in Denver.

5.3.5 Module Quality Degradation Calculation

The quality of PV panels decreases over time. According to the datasheet of the panels under study, a power output warranty of 90% of the nominal output power rating during the first ten years and 80% during twenty-five years is ensured. The manufacturer ensures 97% of the rated power in the first year and a reduction of -0.7%/year through year 25. Thus, a linear performance curve can be represented for the quality degradation of the PV modules. The historical data available on the website of NREL for the PV system under study contains measurements from 2006. Hence, it is assumed that this PV system is installed in 2006 and the quality degradation factors for the years 2012 to 2019 will be 92.8%, 92.1%, 91.4%, 90.7%, 90%, 89.3%, 88.6%, and 87.9%, respectively.

5.3.6 Estimation of Mismatch Losses and Soiling Losses

Mismatch losses are caused by the interconnection of solar cells or modules which do not have identical properties or experience different conditions from one another. In fact, the output of the entire PV module under worst-case conditions is determined by the solar cell with the lowest output. This mismatch can be caused for instance by partial shading of the modules. Moreover, the effect of dust, pollen, snow, and other contaminant accumulation on PV modules, commonly referred to as soiling, is an important environmental factor that causes reduced energy generation of PV systems. As the mismatch and soiling losses cannot be directly calculated from the available datasets, a methodology with the following steps is developed to estimate these losses.

Step 1) The expected DC side power is calculated for each data point using the available dataset and the effects of the module quality degradation and the temperature loss are applied. Then, the total DC side energy losses without the quality degradation and the temperature losses are obtained as follows [131]:

$$
\begin{aligned}
P^{Total\ DC\ Loss} &= P^{Inv\ Input} - P_{DC}^{Exp} \\
&= P^{Inv\ Input} \\
&- \left(P_{STC} \times \left(\frac{G}{G_{STC}} \right) \times Q_{deg} \times \left(1 - C_{Temp} \times (25 - T_c) \right) \right)
\end{aligned}
\tag{5.7}
$$

where $P^{Total\ DC\ Loss}$, $P^{Inv\ Input}$, and $P^{Exp}{}_{DC}$ are the total DC side power losses, the measured power at the DC side of the inverter, and the expected DC power of the system, respectively.

Step 2) A technique is developed to obtain the losses due to low irradiance, spectral, and reflection effects. PV modules respond only to a restricted range of wavelengths, limited at long wavelengths by the material bandgap, and at short wavelengths by material absorption. Reflection or optical losses mean that some parts of the light are reflected by the surface of PV panels and therefore cannot contribute to the photo-current production. Hence, at low irradiance levels, the efficiency of the PV module is significantly lower than at STC. This effect can be simply observed by visualizing the values of the total DC side losses and the irradiance for a sunny day which is indicated in Fig. 5.5 for a typical day. To estimate the losses due to low irradiance, spectral and reflection effects, the following procedure is developed.

- The average value of the total DC side losses on several sunny days (i.e. when irradiance is between 800 and 1000) is calculated;
- The difference between this average value and the total DC side losses in low irradiance levels likely represents the low irradiance, spectral, and reflection losses.

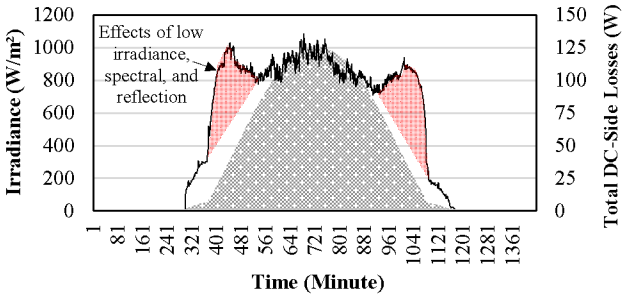


Figure 5.5. Graph of the irradiance (area graph) and the total DC-side losses (line graph) on a given sunny day (10/07/2019) for the PV system in Denver.

For the system under study, these losses occur at irradiance levels lower than 600-800 W/m^2. By doing the calculation for some selected days, it has been found that these losses account for around 3.5% of the daily measured DC power.

Step 3) The DC cabling loss and an estimation of the MPPT losses are also considered. Malfunction of the MPPT controller can lead to considerable losses. These losses depend on the type of the MPPT controller incorporated in the control system of the inverter and cannot be directly calculated from the available dataset. According to different references [132] [133], total MPPT losses of 1.5% could be considered.

Subtracting the losses calculated in step 2 and step 3 from the losses obtained in step 1 represents the mismatch and soiling losses. By applying the aforementioned procedure to the dataset under study, the values of the mismatch and soiling losses have been calculated and the daily values of the losses for the 8-year period are indicated in Fig. 5.6. In fact, this is the total mismatch and soiling losses which include shading, dust, dirt, and snow losses.

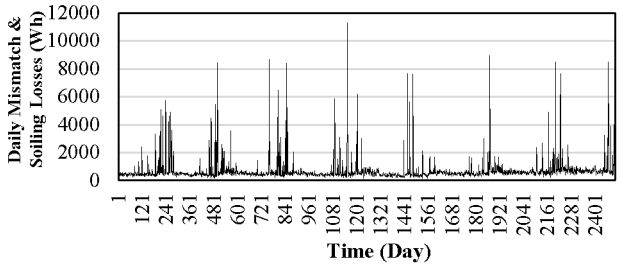
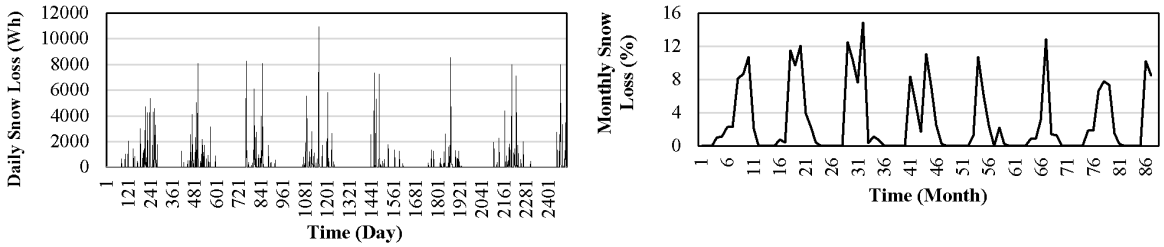


Figure 5.6. Line graph of the daily mismatch and soiling losses over the 8-year period for the PV system in Denver.

5.3.7 Snow Loss Calculation

The accumulated snow/ice on the modules has a negative effect on the PV system performance and decreases the output power. This power reduction, which is referred to as snow loss, cannot be directly calculated from the available dataset. To estimate the snow loss, the average value of the mismatch and soiling losses are obtained during months without snow. The comparison between the mismatch and soiling losses when there are snowfalls and the aforementioned average value for the months without snow can be considered a measure of snow loss. Moreover, the daily values of the snowfall and the snow depth near the location of the PV system are used to ensure that the obtained snow loss values really correspond to the snow cover on the surface of the panels. The

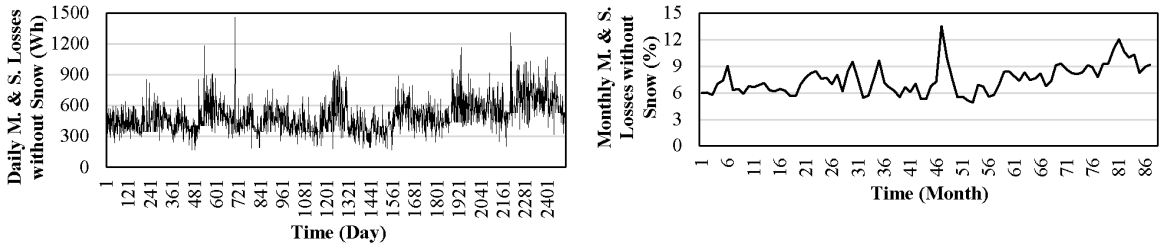
line graphs of the daily snow loss and the monthly percentage of the snow loss for the system under study over the 8-year period are shown in Fig. 5.7 (a) and (b), respectively. As can be seen, the snow loss rarely has a positive value during a year since it depends on snowfall events; however, it has larger peaks compared to other types of losses.



(a) (b)

Figure 5.7. Line graphs of (a) the daily snow loss and (b) the monthly percentage of the snow loss over the 8-year period for the PV system in Denver.

In order to obtain the mismatch and soiling losses without the snow loss, we can simply subtract the snow loss values from the total mismatch and soiling losses. The daily values and the monthly percentage of the mismatch and soiling losses without snow loss for the system under study are shown in Fig. 5.8 (a) and (b), respectively. For the system under study, this type of loss has had an increasing trend in the last years of operation.



(a) (b)

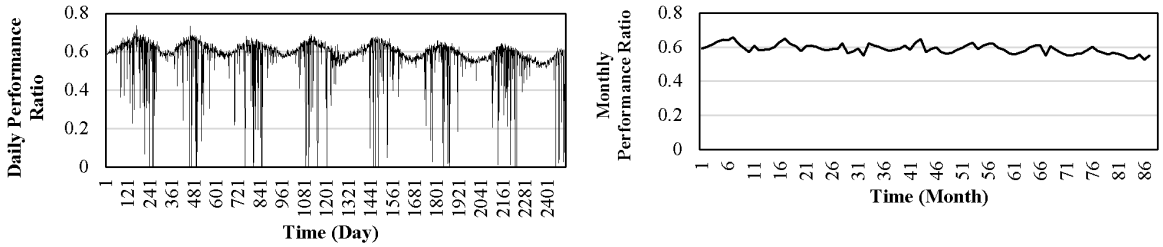
Figure 5.8. Line graphs of (a) the daily mismatch and soiling losses without snow and (b) the monthly percentage of the mismatch and soiling losses without snow over the 8-year period for the PV system in Denver.

5.3.8 Performance Ratio Calculation

The performance ratio (PR) measures how effectively the system converts sunlight collected by the PV panels into AC energy delivered to the grid. In fact, this metric quantifies the overall effect of different losses. The PR is the ratio of the electricity generated to the electricity that would have been generated if the system consistently converted sunlight to electricity at the level expected from the DC nameplate rating, and can be obtained as follows [25]:

$$PR = \frac{\sum_{t}^{T} P_t^{Inv\ Output}}{\sum_{t}^{T} \left(P_{STC} \times \left(\frac{G_t}{G_{STC}} \right) \right)} \tag{5.8}$$

where T is the period over which the PR value is calculated. The values of the daily and the monthly PR have been calculated for the system under study over the 8-year period and are indicated in Fig. 5.9 (a) and (b), respectively. As can be seen, PR has had a decreasing trend over the period under study so that it has dropped from about 0.65 in the first years to about 0.55 in the last years. Moreover, the daily PR has been highly affected by snow loss in winter.



(a) (b)

Figure 5.9. Line graph of (a) the daily and (b) the monthly PR values for the PV system in Denver.

Finally, the average annual percentages of different types of losses (from the expected PV yield) for the PV system under study have been calculated and are illustrated in Table 5.2.

Table 5.2**.** Average annual percentages of different PV system losses for the PV system in Denver

Technical Losses	**Values**
Temperature losses	5.7 %
Module quality degradation	From the datasheet
Low irradiance, spectral, and reflection losses	3%
Electrical mismatch (including shading) and soiling losses (including dirt, dust, …) without snow	7.5%
Snow loss	2.9%
MPPT losses	1.3%
DC cabling losses (including wiring, connections, and parasitic resistances)	0.9%
Inverter loss	15.4%
Performance Ratio	0.6

5.4 PV Power Losses Prediction

The first step of the model-building procedure is data normalization. Normalizing the variables makes their values lie within similar ranges and avoids the larger influence

of features with large values in the training step. Data is normalized between 0 and 1 using equation (4.5). The line graphs of the normalized values of the main meteorological parameters over the 8-year period in Denver are indicated in Fig. 5.10. Since the DC cabling loss, low irradiance, spectral, & reflection losses, and MPPT losses have been calculated as a particular percentage of the DC yield of the system, their normalized values are the same as the DC yield. The line graphs of the normalized values of the power losses are similar to those presented in the previous sections but with variations between 0 and 1.

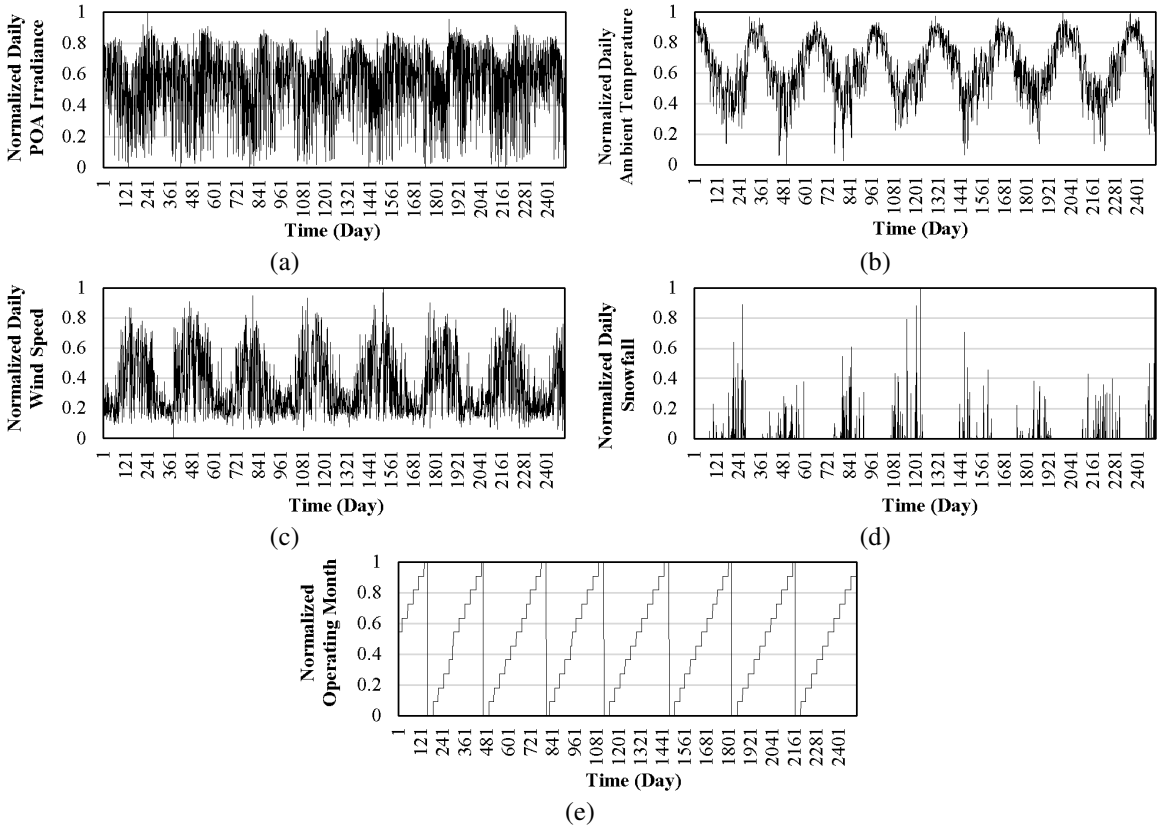


Figure 5.10. The line graphs of the normalized values of the main meteorological parameters for the PV system in Denver.

After normalizing the dataset with a daily temporal resolution, the computational intelligence-based PV power losses prediction models are developed over the data. In order to design the models, a supervised learning approach is implemented. It is a multiple univariate regression problem with 5 features including irradiance, ambient temperature, wind speed, snowfall, and operating date, and numeric target variables which are different types of power losses. According to the literature (such as [134]) and after trying different types of techniques, the two best and most well-known

computational intelligence techniques including gradient boosting tree and long short-term memory network are selected to model each PV power loss. Linear regression models are also developed for each type of loss as a benchmark to compare the performance of the models. Therefore, three models are developed for each target variable including the inverter loss, DC cabling loss, temperature loss, low irradiance, spectral, & reflection losses, MPPT losses, snow loss, and mismatch & soiling losses without snow loss. Computational intelligence-based modeling is also applied to the daily DC yield and daily PR values.

5.5 Numerical Results

In order to build the prediction models, MATLAB software is used and Bayesian optimization is implemented to find the best and optimal hyperparameters of the models. The total number of data points (daily calculated/observed values) over the 8-year period under study is 2,502. By dividing the whole dataset into five sections, the first four sections (80% of the dataset) of data are used to train the models and the last section (20% of the dataset) is used to test the models. Hence, the training dataset contains the first 2,002 data points (14/07/2012 to 01/07/2018) and the test dataset contains the last 500 data points (02/07/2018 to 02/12/2019). The models are trained using the training data and then evaluated on the test data. The performance of the models is evaluated by using the widely utilized metric of root mean squared error.

Two cases are studied in this section. In case 1, the computational intelligence modeling is applied to the dataset of the system under study (daily values) and the prediction models are built for each type of power loss as well as the daily PR values. In case 2, the best models in case 1 for each type of power loss are applied to the historical data of another PV system with different technical characteristics located in different climatic conditions to evaluate whether the models developed for a particular PV system are applicable to another system or not.

5.5.1 Case 1: Model Development and Evaluation for Main PV System

Three models consisting of the linear regression model, GBT model, and LSTM network are developed for each target variable and their optimal hyperparameters which result in the best performance are obtained. The RMSE values on the test data resulting from implementing each model with its optimal hyperparameters for each target variable are presented in Table 5.3. The line graphs of the predicted values of each target variable

using the best model together with their real calculated values in the test dataset are shown in Fig. 5.11.

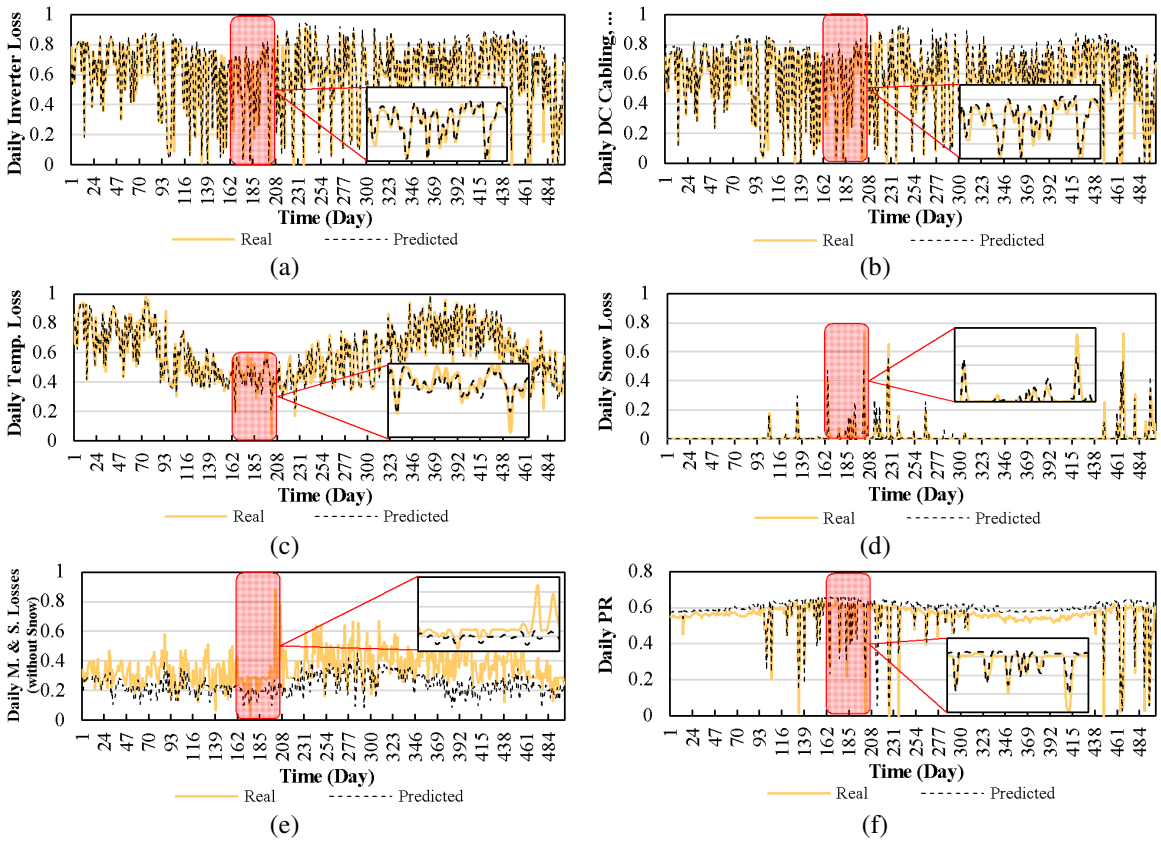


Figure 5.11. The line graphs of the real and the predicted values of the target variables using the best prediction models in case 1 (PV system in Denver).

As can be observed in the first row of Table 5.3, the developed LSTM model for the inverter loss has a better performance with a lower RMSE value compared to the linear regression and GBT models. The RMSE obtained using this model is 0.0387 while the other models have resulted in a higher RMSE on the test data. This model has 14 hidden units with sigmoid gate activation functions and tanh state activation functions. The initial learning rate (LR) is 0.1 which is halved after every 99 epochs. The best result has been obtained after 352 epochs with a mini-batch size of 26. The predicted values of the inverter loss in the test data using the developed LSTM model are shown in Fig. 5.11 (a). As can be seen, the predicted values are very close to the real ones which proves the accurate performance of the prediction model.

Table 5.3**.** Performance measure and optimal hyperparameters for power losses prediction models of the main PV system (case 1)

Target Variable	Model	RMSE on Test Data	Optimal Hyper-parameters [a]					
Inverter Loss	Regression	0.0592	C_1 = 0.9754	C_2 = 0.1622	C_3 = 0.0445	C_4 = -0.4103	C_5 = -0.0109	b = -0.0406
	GBT	0.0510	Depth of Trees = 2		Learning Rate = 0.6		Error Calculation = 188 Iterations	
	LSTM	**0.0387**	NHU [b] = 14	GAF [c] = Sigmoid	SAF [d] = tanh	*LR(k) = 0.1 / ((k/99) × 2)*	Epochs = 352	Mini-Batch Size = 26
DC Cabling, Low Irradiance, and MPPT losses	Regression	0.0648	C_1 = 1.0082	C_2 = 0.0155	C_3 = 0.0578	C_4 = -0.4434	C_5 = 0.0025	b = 0.0002
	GBT	0.0574	Depth of Trees = 2		Learning Rate = 0.4		Error Calculation = 95 Iterations	
	LSTM	**0.0469**	NHU = 3	GAF [b] = Sigmoid	SAF [c] = tanh	*LR(k) = 0.1 / ((k/99) × 2)*	Epochs = 614	Mini-Batch Size = 5
Temperature Loss	Regression	0.0583	C_1 = 0.3625	C_2 = 0.5815	C_3 = -0.0866	C_4 = 0.0758	C_5 = 0.0171	b = 0.0260
	GBT	0.0346	Depth of Trees = 3		Learning Rate = 0.1		Error Calculation = 138 Iterations	
	LSTM	**0.0332**	NHU = 17	GAF [b] = Sigmoid	SAF [c] = tanh	*LR(k) = 0.1 / ((k/98) × 2)*	Epochs = 607	Mini-Batch Size = 34
Snow Loss	Regression	0.0548	C_1 = 0.0661	C_2 = -0.1473	C_3 = -0.0240	C_4 = 0.4418	C_5 = 0.0063	b = 0.0725
	GBT	**0.0283**	Depth of Trees = 4		Learning Rate = 0.4		Error Calculation = 5 Iterations	
	LSTM	0.0316	NHU = 26	GAF [b] = Sigmoid	SAF [c] = tanh	*LR(k) = 0.1 / ((k/42) × 2)*	Epochs = 285	Mini-Batch Size = 19
Mismatch and Soiling Losses without Snow	Regression	0.1539	C_1 = 0.1108	C_2 = -0.0269	C_3 = -0.0787	C_4 = -0.2244	C_5 = -0.0333	b = 0.2377
	GBT	**0.1449**	Depth of Trees = 3		Learning Rate = 0.9		Error Calculation = 9 Iterations	
	LSTM	0.1483	NHU = 21	GAF [b] = Sigmoid	SAF [c] = tanh	*LR(k) = 0.1 / ((k/68) × 2)*	Epochs = 616	Mini-Batch Size = 31
Performance Ratio	Regression	0.0866	C_1 = 0.1312	C_2 = 0.0581	C_3 = 0.1127	C_4 = -0.6803	C_5 = 0.0037	b = 0.4430
	GBT	**0.0557**	Depth of Trees = 4		Learning Rate = 0.1		Error Calculation = 49 Iterations	
	LSTM	0.0656	NHU = 28	GAF [b] = Sigmoid	SAF [c] = tanh	*LR(k) = 0.1 / ((k/38) × 2)*	Epochs = 62	Mini-Batch Size = 41

[a] Hyper-parameter of the linear regression models including C_1, C_2, C_3, C_4, C_5, and b are the coefficients for POA irradiance, ambient temperature, wind speed, snowfall, operating month, and the intercept of the model, respectively.
[b] Number of Hidden Units. [c] Gate Activation Function. [d] State Activation Function.

The results obtained from the models developed for the DC cabling loss, low irradiance, spectral, and reflection losses together with MPPT losses are presented in the second row of Table 5.3. As mentioned earlier, these losses have been calculated as a particular percentage of the daily DC yield and hence, they are modeled using a single target variable. Similar to the inverter loss, the LSTM model developed for the aforementioned losses has a better performance compared to the linear regression and GBT models. The RMSE obtained using the LSTM model is 0.0469 which is a network with 3 hidden units. The activation functions and the learning rate are similar to those in the model developed for the inverter loss. However, the best result has been obtained after 614 epochs with a min-batch size of 5. The line graphs of the predicted and the real values of the target variable on the test data using the LSTM model in Fig. 5.11 (b) also prove the effectiveness of the model in predicting the future values of the target variable.

The results of the models developed for the temperature loss in the third row of Table 5.3 also show that the LSTM model achieves superior performance over the other developed models. However, the resulting RMSE using this model (0.0332) is very close to the RMSE obtained using the GBT model (0.0346). This LSTM model consists of 17 hidden units and sigmoid and tanh as the gate and state activation functions, respectively. The initial learning rate is 0.1 which is halved after every 98 epochs. A total number of epochs of 607 and a mini-batch size of 34 are the other optimal hyper-parameters. The line graphs presented in Fig. 5.11 (c) also show that the LSTM model has finely captured the behavior of the temperature loss and can predict the values of the target variable on the test data very well.

Based on the results obtained from the snow loss prediction models in the fourth row of Table 5.3, the GBT model has resulted in the lowest RMSE value on the test dataset (0.0283). This means that the tree-based model of GBT has had better performance compared to the linear regression model and even the LSTM model. This GBT model consists of trees with a maximum depth of 4. The optimal learning rate is 0.4. The best result has been obtained after 5 iterations of calculating the errors. The predicted values of the snow loss in the test dataset using the GBT model are indicated in Fig. 5.11 (d). As can be seen, the snow loss has been predicted very well using the prediction model based on GBT. It should be noted that the snow loss has a more complex nature compared to the other losses investigated so far since not only it depends on the daily snowfall but also on the snow depth and some other sophisticated factors such as the type

of snowflakes and the shape of the snow cover which cannot be measured easily and are not used in this work, either.

Certainly, the mismatch and soiling losses are the most difficult type of loss investigated in this study for modeling. It is mainly due to the complexity in the nature of these losses and also the fact that some other factors such as the depth and shape of dust on the panels, which are not widely available or not even normally measured, mainly affect the amount of these losses. As can be seen in the fifth row of Table 5.3, all models have poor performance in predicting the values of these losses. However, the RMSE obtained using the GBT model is a bit lower compared to the other models. The line graphs in Fig. 5.11 (e) also show that even this model could not predict the values of these losses with good accuracy. The model, however, could capture the main trend in the values of the target variable.

The results obtained from the PR prediction models are presented in the last row of Table 5.3. The GBT model with the RMSE value of 0.0557 has had a better performance in comparison with the other two models. This model consists of trees with a maximum depth of 4. The best learning rate and the optimal error calculation iterations are also 0.1 and 49, respectively. The line graphs of the predicted and the real values of the daily PR using the GBT model in Fig. 5.11 (f) also prove the acceptable performance of this prediction model.

As can be seen in Table 5.3, the linear regression models have relatively poor performance for all target variables compared to the other two models (GBT and LSTM). This shows that implementing the proposed computational intelligence modeling has been effective in obtaining more accurate predictions of the target variables.

5.5.2 Case 2: Model Evaluation on a Second PV System

In this section, the previously developed power loss prediction models are used for a different PV system to evaluate how well the models can predict the values of the daily power losses for the new system. This analysis attempts to demonstrate whether the loss prediction models built based on the historical data of a particular PV system (which work well for that PV system) are applicable to another PV system with different technical characteristics and climatic conditions or not.

To this end, the only available dataset with an acceptable quality of data and required parameters belongs to a PV system located in Las Vegas, NV. This dataset is extracted from the website of NREL [127]. The nominal power of the PV system is 68.4 kW. The type of the panels is NU-U240F1 manufactured by Sharp. The tilt angle of the panels is 5°. The total number of panels is 285 (each of 240 W). A 50 kW inverter is used in the system which is manufactured by SatCon Technology. Since the previously developed models are going to be tested on the new system, selecting the data records over only a year is a reasonable choice. By analyzing the dataset which contains monitored field data from 2011 to 2020, it has been found that the data recorded in 2012 has the best quality with a smaller number of anomalies and missing data. Therefore, the data for this particular year has been prepared and cleaned. The data is available from 05/01/2012 to 30/12/2012 (i.e. for 360 days).

By implementing the proposed approach of PV power losses calculation presented in Fig. 5.1, different types of losses have been calculated for the new PV system. DC cabling loss is calculated as 1.5% of the DC yield. In contrast to the first PV system, the inverter of the new system limits the power when the DC yield is more than 50.3 kW. Hence, the inverter power limitation loss is not zero. Since this type of power loss was zero for the first PV system, no prediction model was built for that. Moreover, the low irradiance, spectral, and reflection losses are about 1% which is lower compared to the first PV system. MPPT losses are again assumed to be 1.5%. In addition, there was no snowfall in Las Vegas in 2012. So, the snow loss was zero over this year.

The best prediction models developed in case 1 (according to Table 5.3, the LSTM model with 14 hidden units for the inverter loss, the LSTM model with 3 hidden units for the DC cabling, low irradiance, and MPPT losses, the LSTM model with 17 hidden units for the temperature loss, the GBT model with a maximum depth of 4 for the snow loss, the GBT model with a maximum depth of 3 for the mismatch and soiling losses, and the GBT model with a maximum depth of 4 for the performance ratio) were utilized to predict different types of losses for the new PV system. The prediction models used for each target variable together with the RMSE values obtained on the new test system are presented in Table 5.4. The line graphs of the calculated PV system losses together with the predicted values of each type of loss for the second PV system using the best prediction models in case 1 are shown in Fig. 5.12.

Table 5.4. Performance measure of the best prediction models for the second PV system in Las Vegas (case 2)

Target	Model	RMSE on Test Data
Inverter Loss	LSTM	0.2917
DC Cabling, Low Irradiance, and MPPT Losses	LSTM	0.0721
Temperature Loss	LSTM	0.1625
Mismatch & Soiling Losses without Snow	GBT	0.4245
Performance Ratio	GBT	0.0640

As can be seen in Table 5.4, the best performance has been obtained for the target variables of the PR and the DC cabling, low irradiance, and MPPT losses with RMSE values of 0.0640 and 0.0721, respectively. The largest RMSEs resulted from applying the developed GBT and LSTM models to mismatch and soiling losses and the inverter loss with values of 0.4245 and 0.2917, respectively. The temperature loss prediction model has achieved acceptable performance.

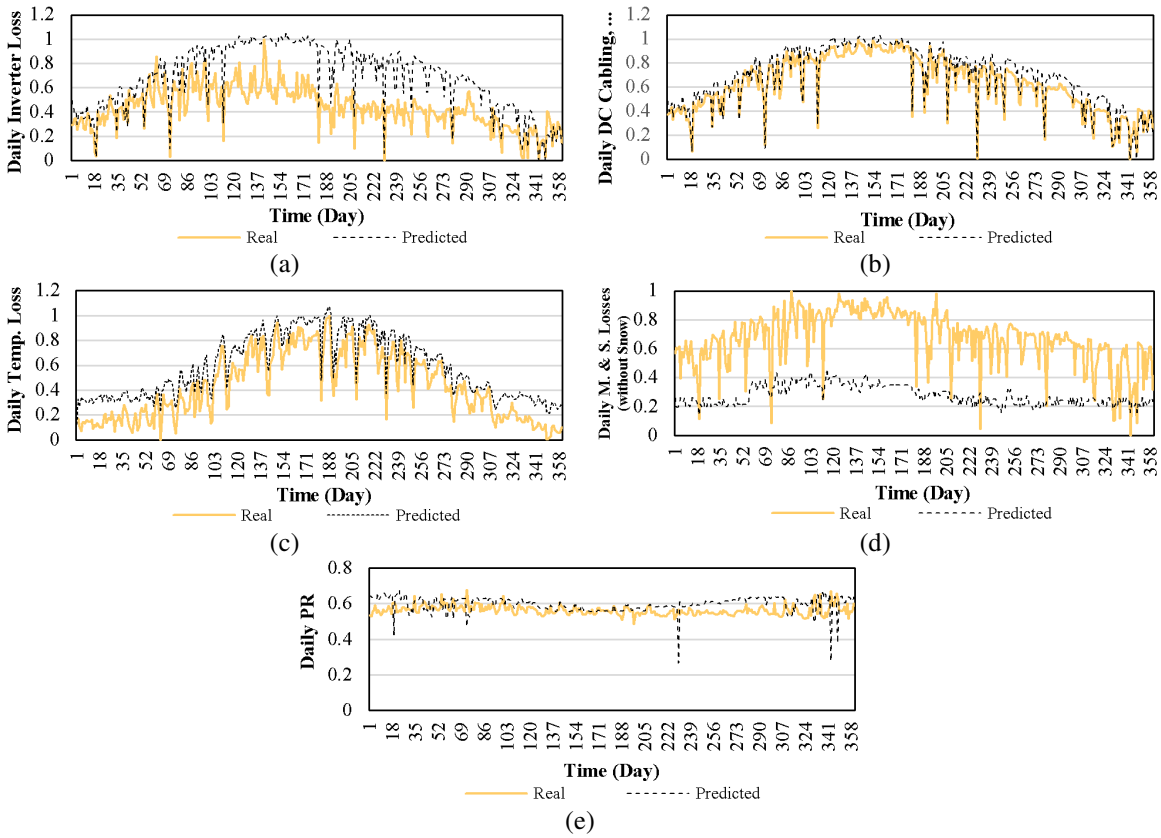


Figure 5.12. The line graphs of the real and the predicted values of the target variables using the best prediction models in case 2 (PV system in Las Vegas).

The line graphs presented in Fig. 5.12 also prove that the inverter loss has not been predicted very well. This means that the inverter loss depends highly on the

characteristics of the inverter itself and different inverters can have different behavior in the same condition. So, the inverter loss prediction model developed for a particular PV system may not be applicable to another one. The DC cabling, low irradiance, and MPPT losses are predicted with good accuracy. Since these losses have been calculated as a particular percentage of the DC yield, it shows that PV arrays have very similar behavior in the DC power generation and the losses related to that. As can be seen in Fig. 5.12 (c), the model has captured the trend in the values of the temperature loss and can have acceptable performance. The large RMSE value obtained in predicting the mismatch and soiling losses can be easily recognized from the line graphs in Fig. 5.12 (d). There is a large difference between the predicted and the real values. The line graphs in Fig. 5.12 (e) also show that the predicted values of the PR are very close to the real values except for several data points.

In order to evaluate the performance of the models developed in case 1 if they were also exposed to the data from the second PV system, we retrain the models by a part of the 1-year dataset of the second PV system. To this end and considering the nature of the data as time series, we use the first half of the data (the first 180 data points) for retraining the models and the last half of the data (the last 180 data points) for testing the models. The models are exposed to the new training data for 200 iterations and tested using the new test data after each iteration. The RMSE values over the new test dataset obtained from all the models after retraining them are shown in Fig. 5.13. The line graphs of the predicted values of the target variables over the new test data before and after retraining the models for the optimal number of iterations are indicated in Fig. 5.14.

As one can notice in Fig. 5.13, the minimum RMSEs over the new test dataset (the last half of the dataset for the second PV system) for the LSTM-based inverter loss prediction model, LSTM-based DC cabling, low irradiance, and MPPT losses prediction model, LSTM-based temperature loss prediction model, GBT-based mismatch and soiling losses prediction model, and GBT-based PR prediction model are 0.0469, 0.0753, 0.0573, 0.0754, and 0.0263 obtained by retraining the models after 21, 17, 96, 12, and 13 iterations, respectively. The aforementioned RMSEs, however, are significantly lower compared to those presented in Table 5.4 before retraining the models, except for the LSTM-based DC cabling, low irradiance, and MPPT losses prediction model, it may not be reasonable to compare them since they have been obtained over test datasets with different numbers of data points. Nevertheless, the effect of retraining the models by a

relatively small training dataset (containing 180 data points) of the second PV system on improving the prediction of the target variables for this PV system can be observed from the line graphs in Fig. 5.14.

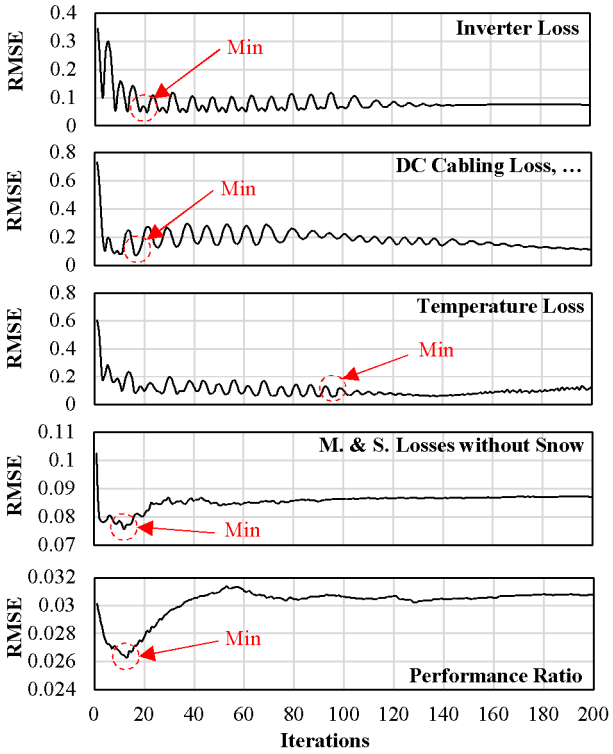


Figure 5.13. The RMSE values over the new test dataset (the last 180 data points for the second PV system) after retraining the models using the new training dataset (the first 180 data points for the second PV system) over 200 iterations.

In general, the results obtained in case 2 prove that the LSTM-based inverter loss prediction model, LSTM-based DC cabling, low irradiance, and MPPT losses prediction model, LSTM-based temperature loss prediction model, GBT-based mismatch and soiling losses prediction model, and GBT-based PR prediction model which were built based on the data of the main PV system can be utilized for another PV system with different characteristics in different climatic conditions. The prediction models can perform much better if they are retrained by a small dataset of the new PV system under the analysis.

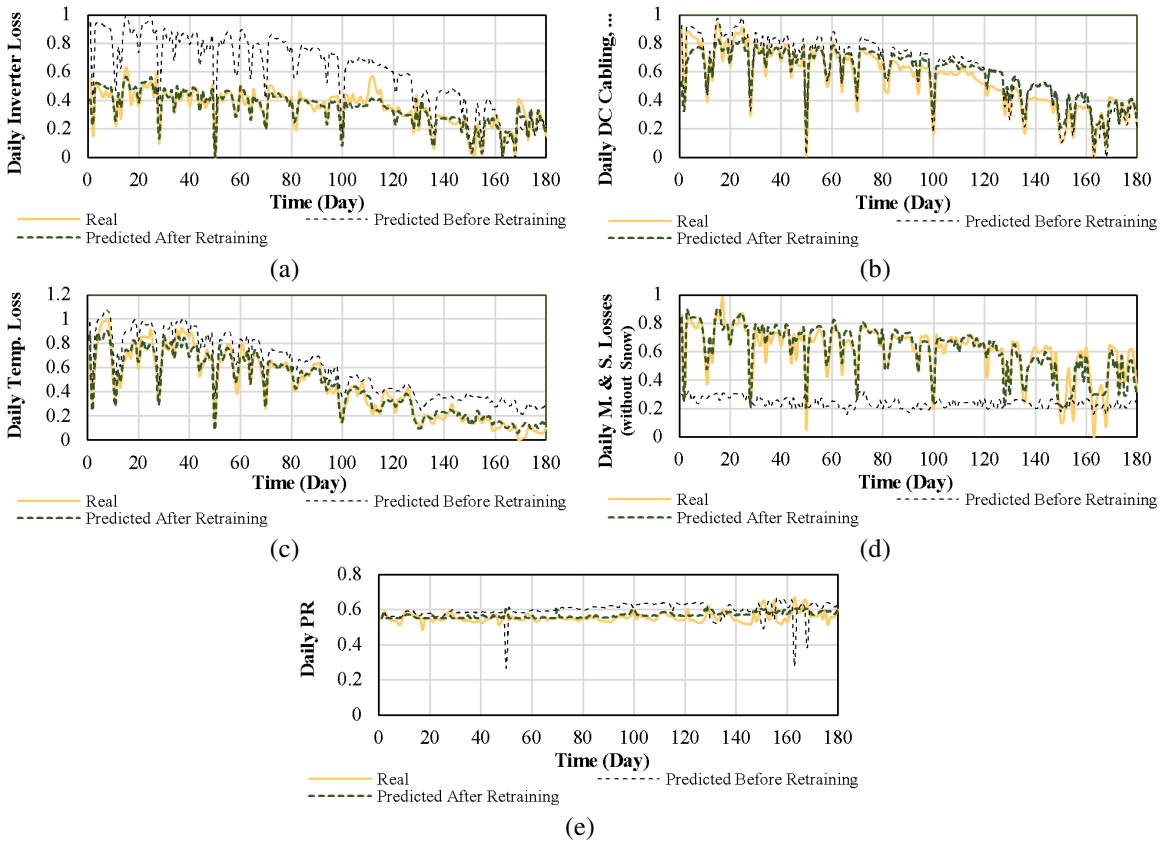


Figure 5.14. The line graphs of the real and the predicted values of the target variables using the best prediction models in case 2 before and after retraining the models with the first half of data for the second PV system in Las Vegas.

5.6 Conclusion

In this chapter, a systematic PV system power loss calculation approach was proposed. By implementing this approach, different types of power losses in PV systems, including both array capture losses (i.e. temperature loss, mismatching and soiling losses, low irradiance, spectral, and reflection losses, module quality degradation, and snow loss) and system losses (i.e. inverter loss, cabling loss, inverter power limitation loss, and MPPT losses) can be extracted from the historical data of the main electrical and meteorological parameters of the system. The proposed approach was applied to 8 years of recorded data for a 1.44 kW rooftop PV system located in Denver, CO, and daily, monthly, and annual values of all aforementioned power losses were calculated for the system. The inverter of this system had a low efficiency over the 8-year period with an average annual loss of 15.4%. The average annual performance ratio of the overall system was also calculated as 0.6 which indicates the relatively poor performance of the system. In addition, a computational intelligence modeling of the PV system power losses

was proposed in the chapter. Two well-known techniques, namely gradient boosting tree and long short-term memory network, were implemented to build prediction models for each type of loss solely based on some main meteorological parameters as the inputs of the models. The developed LSTM models were demonstrated to predict the values of the inverter loss, the DC cabling, low irradiance, and MPPT losses, and the temperature loss in the test data more accurately (with RMSE values of 0.0387, 0.0469, and 0.0332) compared to the GBT models. In contrast, the developed GBT models for the snow loss, mismatch and soiling losses, and PR performed better in comparison with the LSTM models with RMSE values of 0.0283, 0.1449, and 0.0557, respectively. These models were also applied to another dataset for a 68.4 kW PV system located in Las Vegas, NV, to demonstrate their generalization capability. According to the obtained results, the developed models can predict the values of the target variables very well if they are retrained on a small dataset of the new PV system under analysis.

CHAPTER 6

COMPUTATIONAL INTELLIGENCE-BASED PREDICTION OF SNOW COVER

6.1 Introduction

The power generation of PV systems in snow-prone areas is highly volatile during long harsh winters due to shading by snow covers. In this chapter, the snow cover detection and prediction problem for PV systems in snow-prone areas is investigated. Unlike electrical variables such as voltage, current, or power, no tool is usually used to detect the presence of a full or partial snow cover on PV panels. Even in the case of small-scale PV systems, no sensor is utilized for measuring the on-site values of the meteorological variables. So, the records of the snow cover conditions for PV systems are not available and the data recorded by the closest weather stations, which are usually some kilometers away from PV systems, provide the other meteorological variables. Therefore, a procedure is required to detect the hours when a probable snow layer exists on the surface of the panels and label the dataset. This information can be used as an input for PV power prediction models and improve prediction accuracy. In the following, first, the case study of this chapter is explained and then, the proposed methodology is presented.

6.2 System Under Study

In order to conduct the study, the dataset of the DC power generation [135] and 16 meteorological variables for a rooftop PV system in Edmonton, Alberta, Canada has been extracted. The nominal power, tilt angle, and orientation of the PV system are 5.5 kW, 20°, and south, respectively. The temporal resolution of the data is one hour with measurements from 2016-11-09 to 2019-12-31. The meteorological variables are global horizontal irradiance (GHI), direct normal irradiance (DNI), diffuse horizontal irradiance (DHI), ambient temperature, dew point temperature, relative humidity, wind direction, wind speed, visibility, atmospheric pressure, cloud cover, rain, snowfall, snow depth on the ground, solar zenith, and solar azimuth. For the three variables of rain, snowfall, and snow depth, only the total daily values are available. The publicly available online databases of the National Renewable Energy Laboratory [136], Environment and Natural

Resources Canada [10], and Canada Weather Stats [137] are used to extract the historical datasets of the aforementioned meteorological variables.

The detection and removal of the outliers and incorrect values in the dataset are performed as the first step. Then, the K-nearest neighbors technique is used to impute missing data points. In order to integrate the daily values of rain, snowfall, and snow depth to the rest of the data with the temporal resolution of one hour, the hourly values of the three aforementioned variables are assumed to be constant on each day, equal to the daily value. Based on the nature of photo-current generation in PV cells, there is no power generation if the amount of solar radiation reaching the surface of the panel is zero. Hence, all data points with solar irradiance equal to zero are dropped from the dataset since, in this condition, snow cover prediction has no value. Finally, the cleaned dataset is used for the modeling. The total number of data points in the dataset is 13,523. The main statistical characteristics of the variables are presented in Table 6.1. As can be seen, the solar irradiance has no value greater than 1,000 W/m^2 considering the location of the system in a relatively high latitude. Moreover, the maximum snow depth on the ground is 45 cm with an average value of 5.8 cm which shows the huge amount of snowfall in this region whose maximum value is 18 cm. It also can get as cold as -35°C in winter which slows down the melting of snow.

Table 6.1**.** Statistical characteristics of the historical dataset of the main PV system

Variables	**Mean**	**S.D.**	**Min**	**Median**	**Max**
Power (W)	1228.2	1378.4	0	604.5	5069.2
GHI (W/m^2)	265.6	233.8	1	198	916
DNI (W/m^2)	341.4	319.8	1	276	986
DHI (W/m^2)	107.9	90.9	1	79	428
Ambient Temp. (°C)	7.2	11.9	-35.3	9	32.8
Dew Point Temp. (°C)	1.1	10.3	-39	2	20
Relative Humidity (%)	68.2	16.5	23.9	69.7	100
Wind Direction (10s°)	22.2	9.5	0	23	36
Wind Speed (km/h)	14.4	8.2	0	13	73
Visibility (km)	20.3	6.3	0	24.1	24.1
ATM Pressure (kPa)	93	0.7	90.1	93	95.3
Rain (mm)	1.1	3.5	0	0	44.6
Snowfall (cm)	0.3	1.3	0	0	18
Snow Depth (cm)	5.8	10.2	0	0	45
Cloud Cover (0-10)	7.4	3.1	0	9	10
Solar Zenith (°)	66.8	17.5	30.4	69	103.8
Solar Azimuth (°)	161	63.8	32.5	160.4	289.1

6.3 Snow Cover Detection

As mentioned before, no measurement records are usually available for a partial/full snow cover on the PV panels. Hence, a 3-step computational intelligence-based approach with the structure shown in Fig. 6.1 is proposed to find hours with a probable snow cover, label the dataset, and build computational intelligence-based snow cover prediction models.

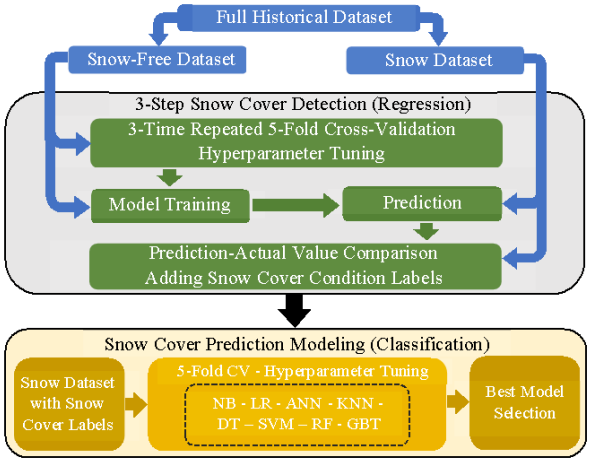


Figure 6.1. The framework of the proposed computational intelligence-based snow cover detection and prediction approach.

In the first step, two datasets, namely snow-free condition and snow condition datasets, are extracted from the full dataset. The snow-free condition dataset is constituted by the data points with snowfall and snow depth equal to zero. The snow condition dataset is formed by the data points with snowfall or snow depth greater than zero. Then, a computational intelligence-based model is developed on the normalized snow-free condition dataset with the PV power target variable and the meteorological attribute variables. This is achieved by implementing a 3-time repeated random 5-fold cross-validation approach together with hyperparameter tuning based on Bayesian optimization. After developing and comparing the performance of various models, the gradient boosting tree-based model had the minimum prediction error with a test RMSE of 0.0745 and a train RMSE of 0.0355. In the second step, the best model, gradient boosting tree in this case, is trained over the entire snow-free condition dataset. Then, the model is used to predict the target values of the snow condition dataset. In the third step, the actual values of the PV power generation are compared with the corresponding predictions. If an actual value is much lower than the corresponding prediction, a snow cover is detected on the panels and the data point is labeled as 1. Otherwise, the data

point is labeled as 0. In order to prevent false detections, snowfall events have been also taken into account.

The line graphs of the PV power and the snow cover condition are shown in Fig. 6.2. 5,050 data points in the full dataset represent snow conditions, among which 3,100 data points are detected with probable snow cover. Considering the snow cover condition as the target variable of the main problem, the correlation coefficients between the target variable and the PV power as well as the meteorological variables are indicated in Fig. 6.3. As can be seen, there is a relatively strong negative correlation between the PV power and the snow cover condition. Although the PV power won't be used in the snow cover prediction models, this correlation shows that the snow cover predictions can be an important and effective input for PV power prediction and forecasting models. The most correlated meteorological variables with the snow cover condition are ambient temperature, dew point temperature, GHI, visibility, and daily snow depth, respectively. While the correlations for temperature and GHI are negative, snowfall and snow depth have positive correlations.

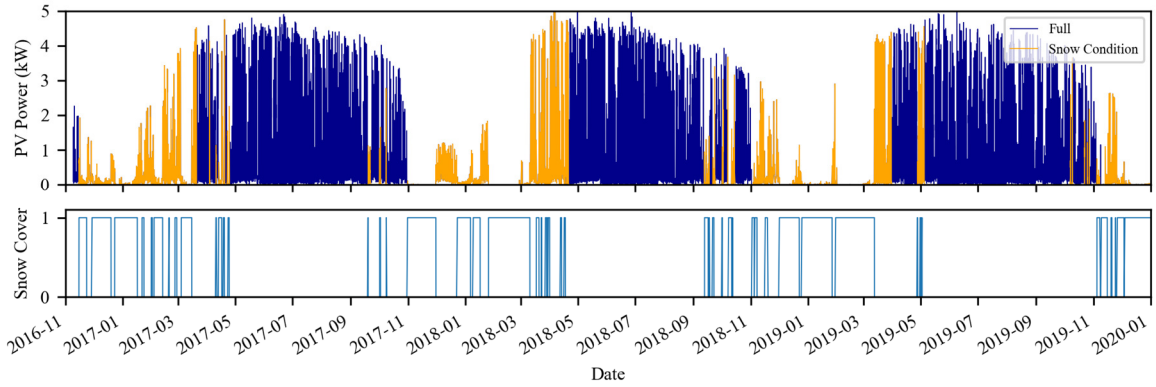


Figure 6.2. Line graphs of the PV power and snow cover condition in the full dataset.

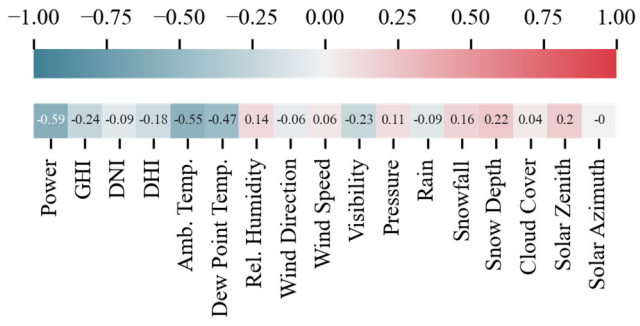


Figure 6.3. Linear correlations between the snow cover labels and other variables in the snow condition dataset.

6.4 Snow Cover Prediction

After obtaining the snow cover labels using the proposed approach in the previous section, they are added to the historical dataset of the system. The updated dataset will be used for developing computational intelligence-based snow cover prediction models. Considering the snow cover binary labels as the target variable, the snow cover prediction is a classification problem. As such, several computational intelligence techniques, i.e. naïve Bayes, logistic regression, k-nearest neighbors, decision tree, random forest, gradient boosting tree, support vector machine, and artificial neural network, are used to develop snow cover prediction models. All 16 meteorological variables are used as the attribute variables. The target variable is the hourly snow cover condition (binary labels). Regarding the fact that if there is no snowfall and no snow accumulation on the ground (both variables are zero), no snow cover is expected on the panels, training and testing of the models are performed on the snow condition dataset. As mentioned earlier, the snow condition dataset can be simply extracted from the full historical dataset of the system by choosing only the data points whose snowfall or snow depth on the ground has values bigger than zero. All 16 meteorological variables are used as the attribute variables. The target variable is the hourly snow cover condition (binary labels). Finally, the performance of the models will be compared to find the best snow cover predictor.

6.5 Numerical Results

Considering the small size of the snow condition dataset and in order to have all the data points of this dataset contributed to training the models, 5-fold cross-validation is implemented. The dataset is divided randomly into 5 equally-sized segments. Then, 5 different combinations of the segments are formed, four of which are used in training the model, and the one remaining is used in testing the model. The performance score of the model is the average of the scores over all 5 combinations. Bayesian optimization-based hyperparameter tuning is implemented to obtain the best hyperparameters of the models. Python programming language and Scikit-learn library are used to build the models.

6.5.1 Case 1: Model Development and Evaluation for Main PV System

The aforementioned models are developed over the dataset and their optimal hyperparameters are presented in ascending order of their accuracy in Table 6.2. As can be seen, the naïve Bayes model, which is usually considered a benchmark, has the minimum test accuracy with a value of 0.7846. The performance of the logistic regression

model is a bit better with a test accuracy of 0.8055. C represents the inverse of the regularization strength and β_0 to β_{16} are the intercept and the coefficients of GHI, DNI, DHI, ambient temperature, dew point temperature, relative humidity, wind direction, wind speed, visibility, pressure, rain, snowfall, snow depth, cloud cover, solar zenith, and solar azimuth, respectively. A better test accuracy, 0.8861, has been obtained by the ANN model. The rectified linear unit (relu) function, which returns max(0, x), is used as the activation function. The learning rate is adaptive which keeps the learning rate constant at 0.001 as long as the training loss keeps decreasing; otherwise, it is divided by 5. The model has three hidden layers with 25, 50, and 25 processing units in each layer. The K-nearest neighbors model is the next model with a test accuracy of 0.8897. The best value of K is 4 and the data points are weighted by the inverse of their distance. The developed decision tree with a maximum depth of 20 has a bit better performance with a test accuracy of 0.8978. The next test accuracy, 0.9216, has been obtained by the non-linear SVM model with Radial Basis Function (RBF) as the kernel. C, gamma, and b are the regularization parameter, Kernel coefficient, and intercept of the model, respectively. The random forest model is the next model with a test accuracy of 0.9448. This model, formed by 100 trees, uses the whole dataset and 7 features to build each tree with a maximum depth of 24. Finally, the best performance among all the developed models has been obtained by the gradient boosting tree model. The test accuracy of this model is 0.9580 which is the maximum value in this study. This model is constituted of 300 trees of a depth of 10 for the boosting stages. The training accuracy of the model is 1.

Table 6.2**.** Accuracy of the snow cover prediction models and their optimal hyperparameters

Model	Test Accuracy	Training Accuracy	Optimal Hyperparameters
Naïve Bayes	0.7846	0.7829	-
Logistic Regression	0.8055	0.8069	C = 0.572, β_0, …, β_{16} = 6.899, -2.41, 1.1, 0.22, -5.98, -5.09, 2.97, -0.35, 1.0, -0.96, -0.88, -1.39, 3.39, 1.16, 0.73, -1.61, -0.22
Artificial Neural Network	0.8861	0.9221	Activation Function = relu, Batch Size = 1, Learning Rate = Adaptive, Max Iteration = 50, Hidden Layer Sizes = (25,50,25)
K-Nearest Neighbors	0.8897	1	K = 4, Weights = Inverse of Distance
Decision Tree	0.8978	0.9993	Max Depth = 20
Support Vector Machine	0.9216	0.9924	C = 6.288, Gamma = 4.904, Kernel = RBF, b = 0.082
Random Forest	0.9448	1	Bootstrap = False, Max Depth = 24, Max Features = 7, Number of Estimators = 100
Gradient Boosting Tree	**0.9580**	1	Learning Rate = 0.299, Max Depth = 10, Number of Estimators = 300

Considering the gradient boosting tree as the best snow cover prediction model, the importance given by this model to each feature is shown as a value between 0 and 1 in Fig. 6.4. It should be noted that the sum of all the feature importance values is 1. As can be observed, ambient temperature has been recognized as the most important feature with an importance value of 0.3424. After that, snow depth has an importance of 0.1781, almost half of that for ambient temperature. Snowfall, pressure, and solar azimuth have importance values of 0.0742, 0.0706, and 0.0605, respectively. The importance values of the other features are less than 0.05.

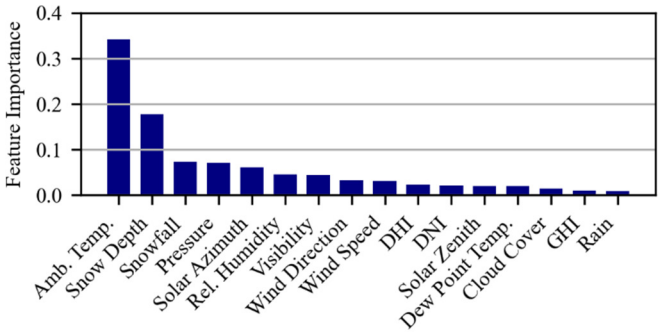


Figure 6.4. The values of the feature importance in the gradient boosting tree model.

6.5.2 Case 2: Model Evaluation on Additional PV Systems

The best snow cover prediction model in this study is tested on the historical datasets of two other roof-top PV systems which are located in Edmonton, Canada, have tilt angles of 25° and 30°, and are facing south. The historical datasets of the new systems contain the hourly values of the same variables as the main system over almost a 4-year period. Similar to the proposed procedure, the snow condition datasets of the new PV systems are extracted and the snow cover labels are obtained and added to the data points. Then, the gradient boosting tree model with the hyperparameters presented in Table 6.2 is trained over the entire snow condition dataset of the main PV system. Finally, the trained model is tested on the unseen snow condition datasets of the new PV systems. The accuracy of the model over the datasets of the new systems is 0.80 and 0.78. This shows that the model trained on the dataset of the main PV system not only has excellent accuracy for the system itself, but also it has an acceptable performance in the case of other PV systems and it is applicable to them, too.

6.6 Conclusion

Data-driven snow cover prediction using computational intelligence techniques for PV systems in northern snow-prone areas was proposed in this chapter. First, the historical dataset of the hourly values of power generation and 16 meteorological variables for a roof-top PV system in Edmonton, Canada, was extracted and cleaned. Then, the presence of a full/partial snow cover on the panels was recognized using a 3-step computational intelligence-based approach and the data points were annotated by binary labels. Finally, 8 computational intelligence-based snow cover prediction models were developed using 5-fold cross-validation and Bayesian optimization-based hyperparameter tuning, where the snow cover condition was the target and 16 meteorological variables were the input features. The results showed that the gradient boosting tree model achieved the best performance with a test accuracy of 0.9580. The random forest model was in the second position with a test accuracy of 0.9448. The test accuracy of the benchmark naïve Bayes model was 0.7846. For the gradient boosting model, ambient temperature, snow depth, and snowfall were the three most important features. By testing the developed gradient boosting tree model with the optimal hyperparameters on the unseen data of two other PV systems, test accuracies of 0.80 and 0.78 were obtained. Overall, the results proved the efficiency of the proposed approach in hourly snow cover prediction for PV systems based only on the main meteorological variables.

CHAPTER 7

COMPUTATIONAL INTELLIGENCE-BASED PREDICTION OF PV POWER GENERATION IN SNOW CONDITIONS

7.1 Introduction

Despite the reasonable performance of the physical PV power prediction models in snow-free conditions, they provide inaccurate results in snow conditions. PV power prediction in snow conditions is a challenging task due to the adverse effect of a partial or full snow cover on the surface of the panels. This effect, which depends on many snow-related factors, cannot be easily formulated as an empirical equation. Implementing sophisticated empirical equations to tackle the modeling in snow conditions needs more information about the snow-related parameters which are not usually available or are difficult to measure in order to make a prediction. In this chapter, a new approach is presented for PV power prediction in snow conditions which benefits from the capabilities of computational intelligence techniques and is based on the categorization of the data points in terms of the snow-related conditions. The computational intelligence-based models have the advantage to perform modeling using a limited number of widely available meteorological data and the minimum knowledge about the underlying physical, electrical, and meteorological principles. The models can be trained to provide fair results in various conditions, such as snow conditions.

7.2 Systems Under Study

In order to conduct the study, the historical datasets of the electrical and meteorological variables for 17 small-scale PV systems in 11 major cities across Canada have been extracted. The nominal power and tilt angle of the PV systems range from 5 kW to 303 kW and 10° to 60°, respectively. All modules of each PV system have identical tilt angles and directions. Most of the PV systems have panels facing south. The detailed specifications of the PV systems are presented in Table 7.1. The data is the hourly measurements of the variables with an aggregated duration of more than 55 years. The electrical variable is the output DC power of each PV system [135]. The datasets also contain the measurements of 16 meteorological variables i.e. global horizontal irradiance,

Table 7.1. Specifications of the PV systems under study

Location	Calgary	Edmonton				Fredericton	Halifax		Ottawa	Regina	Saskatoon		Sherbrooke	Sudbury		Toronto	Winnipeg
PV System Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Array Type [a]	RRM	SRM	SRM	SRM	RRM	RRM	SRM	SRM	SRM	SRM	SRM	RGM	RRM	RGM	RRM	SRM	SRM
Tilt (°)	10	35	30	18	25	30	49	26	20	25	20	60	30	30	45	10	30
Azimuth (°)	180	180	180	180	180	175	133	185	175	180	180	180	180	175	180	162	230
Rating (kW)	153	12.96	6.89	5.58	11.96	100.08	4.69	10.23	4.95	7.42	13.76	24	302.72	11.52	10.08	9.97	8.25
Panel Manufacturer	ReneSola	Canadian Solar	HES	Hanwha Solar	Conergy	Heliene	Canadian Solar	Canadian Solar	Conergy	Solarfun	LG Solar	Hanwha Solar	Canadian Solar	Canadian Solar	Renogy	Silfab	LG Solar
Panel Model	255	CS6P-270	265	HSL 72S-310	PE 260P	72M-360	CS6U-335M	CS3K-310MS	PM 225P	SF220-30-P225	NeON 320 BW	HSC-250-60P	CS1H-320MS	CS6X-320P	RNG-80	SLG 285M	LG 275S1C
Module Number	600	48	26	18	46	278	14	33	22	33	43	96	946	36	126	35	30
Module Orientation [b]	L	L	(6)L-(20)P	P	L	L	P	P	P	P	P	P	L	L	L	P	P
Vertical Modules	1	4	3	2	1	3	2	3	3	3	3	2	1	3	1	3	3

[a] SRM: Standoff Roof Mount, RRM: Rack Roof Mount, RGM: Rack Ground Mount
[b] L: Landscape, P: Portrait

direct normal irradiance, diffuse horizontal irradiance, ambient temperature, dew point temperature, relative humidity, wind direction, wind speed, visibility, atmospheric pressure, cloud cover, rain, snowfall, snow depth on the ground, solar zenith, and solar azimuth. Since the hourly values of rain, snowfall, and snow depth have not been available, the total daily values of these variables are used. The meteorological variables are extracted from three publicly available online databases of the National Renewable Energy Laboratory [136], Environment and Natural Resources Canada [10], and Canada Weather Stats [137].

For the dataset of each PV system, outliers and incorrect values are detected and removed. Then, data imputation using K-nearest neighbors is performed to find the best value for each missing data point considering the rest of the dataset. For the three parameters of rain, snow, and snow depth, the daily values are repeated every 24 hours. Considering the nature of PV cells, the power generation of a PV system is zero if the solar irradiance on the surface of the cell is zero. Hence, all data points with solar irradiance equal to zero are dropped from the datasets. The total number of data points in all the datasets is 242,796.

7.3 Snow-Related Conditions Datasets Extraction

Regarding the main focus of this chapter on snow conditions, three datasets are extracted from the full dataset of each PV system based on the snow-related parameters. The first dataset, the snow-free dataset, contains the data points of the full dataset in the snow-free condition when snowfall and snow depth are both zero. The second dataset, the snow dataset, contains the data points of the full dataset in the snow condition when snowfall or snow depth is greater than zero. The third dataset, the snow-cover dataset, contains the data points of the snow dataset in the snow-cover condition when the surface of the panels is most probably partially or fully shaded by a snow cover.

The extraction of the first two datasets is simple since it only depends on the values of the available snow-related parameters. However, building the third dataset demands a procedure to detect a probable snow cover on the panels. To this end, a procedure similar to the 3-step computational intelligence-based approach developed in chapter 6 is used. The structure of this approach is shown in Fig. 7.1. In the first step, a computational intelligence-based model is trained on the snow-free dataset of each PV system considering the power as the target variable and the meteorological variables as the

attribute variables. To this end, a 3-time repeated random 5-fold cross-validation together with hyperparameter tuning based on Bayesian optimization is implemented. By developing different models for 17 datasets, the best performance has been obtained using gradient boosting tree-based models with an average test RMSE of 0.078 ± 0.011 and an average train RMSE of 0.044 ± 0.014 considering the normalized data between 0 and 1. In the second step, the best models are trained over the entire snow-free dataset and a prediction is made over the snow dataset for each PV system. In the third step, a comparison is made between the measured and the predicted values of the PV power. If a measured value is much lower than the predicted value, there is a probable snow cover on the panels. The snowfall events have been also considered to avoid false detections. To this end, the following algorithm is used in step 3. In order to ensure the fairness of the snow cover labels, all the related data together with the labels have been visualized and explored to recognize and correct the suspicious labels.

Algorithm: to label hours with probable full/partial snow cover

---

Input: predicted (P^{PRD}) and actual (P^{ACT}) PV power in snow dataset, daily snowfall (*SNF*), daily snow depth (*SND*).

Output: full/partial snow cover (*SC*) labels (0/1: without/with snow cover).

Initialization:

for $d = 1$ (day) to D **do**
  for $h = 1$ (hour) to H_d **do**
    if $P^{PRD}_{d,h} - P^{ACT}_{d,h} \geq \varepsilon$ **then**
      if $h = 1$ **and** $SC_{d-1,Hd} = 1$ **then** $SC_{d,h} = 1$.
      else if $h > 1$ **and** $SC_{d,h-1} = 1$ **then** $SC_{d,h} = 1$.
       else if $SND_d > SND_{d-1}$ **or** $SNF_d > 0$ **then** $SC_{d,h} = 1$.
      else $SC_{d,h} = 0$.
    else $SC_{d,h} = 0$.
  end for
end for
return $SC_{d,h}$

ε: The value of this parameter can be determined experimentally by trial and error to provide the most reliable and reasonable snow cover labels and may be different for the PV systems based on the quality and accuracy of the data in their datasets.

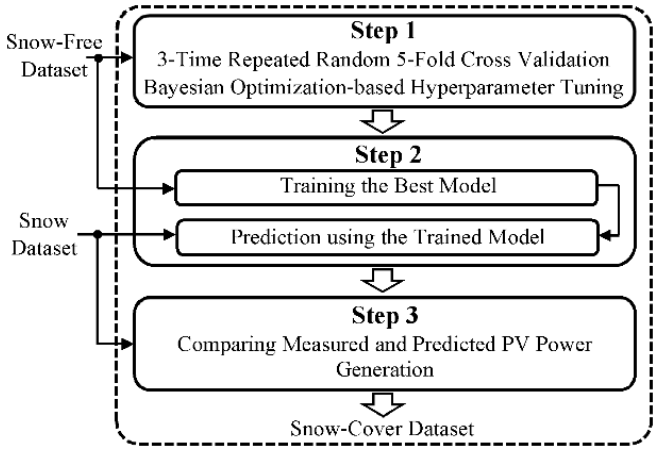


Figure 7.1. Three-step approach for extracting the snow-cover datasets from the full datasets.

7.4 Data Exploration and Dimensionality Reduction

Considering the three aforementioned extracted datasets, four datasets are available for each PV system i.e. full, snow-free, snow, and snow-cover datasets. On average, the snow and snow-cover datasets constitute 32% and 13% of the full datasets, respectively. It also means that almost half of the data points of the snow datasets pertain to the snow-cover condition. The mean values of 16 variables in the four datasets of all 17 PV systems are presented in Table 7.2. As can be seen, there is a considerable difference in the values of irradiance components between different datasets. In the snow datasets (winter time), irradiance is significantly lower, and in the snow-cover datasets, the values are even smaller. While the average hourly temperature over the full datasets is 8.28°C, this variable has an average of -5.26°C in the snow datasets. It is even colder, more humid, and cloudier when there is snowfall or snow cover on the ground. The average daily snowfall and snow depth in the full datasets are 0.49 and 5.01 cm, respectively, which shows the considerable amount of annual snowfall and huge accumulation of snow on the ground in Canada.

The line graphs of the aggregated normalized values of the power generation of all 17 PV systems are demonstrated for each dataset in Fig. 7.2. The considerable effect of a snow cover on the PV power can be easily observed.

Table 7.2. Mean values of the variables of each dataset

Variables	Datasets			
	Full	**Snow-Free**	**Snow**	**Snow-Cover**
GHI (W/m^2)	292.11	337.15	196.57	166.28
DNI (W/m^2)	363.87	382.98	323.33	281.69
DHI (W/m^2)	111.60	124.92	83.33	79.76
Ambient Temp. (°C)	8.28	14.66	-5.26	-7.92
Dew Point Temp. (°C)	2.07	7.26	-8.96	-11.18
Relative Humidity (%)	68.77	65.47	75.76	77.28
Wind Direction (10s°)	20.73	20.62	20.96	20.36
Wind Speed (km/h)	16.52	16.28	17.05	16.74
Visibility (km)	22.78	23.79	20.65	18.29
ATM Pressure (kPa)	96.08	96.14	95.93	95.56
Rain (mm)	1.64	2.14	0.59	0.48
Snowfall (cm)	0.49	0.00	1.28	2.18
Snow Depth (cm)	5.01	0.00	15.65	17.79
Cloud Cover (0-10)	6.92	6.80	7.16	7.54
Solar Zenith (°)	64.87	61.10	72.86	74.45
Solar Azimuth (°)	161.22	157.59	168.91	167.74

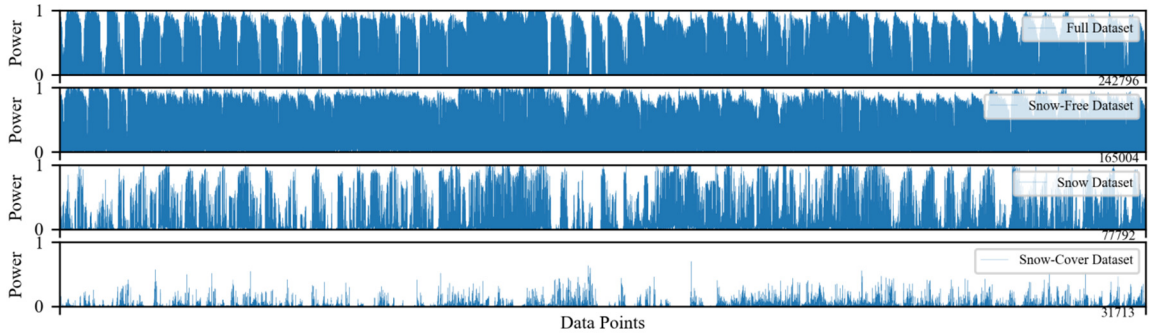


Figure 7.2. Line graphs of the aggregated full, snow-free, snow, and snow-cover datasets for all 17 PV systems.

A comprehensive comparison is made for the correlations between the power generation of each PV system and the meteorological variables. Pearson's coefficient is calculated for the variables in every four datasets of each PV system and is shown in the form of violin plots in Fig. 7.3. As can be seen, the most correlated attribute variable to the PV power in all the conditions is GHI. In the snow-free dataset, this correlation is very strong (close to 1 with a very small variance); however, it is considerably weaker in the snow and snow-cover datasets with average values of 0.75 and 0.5, respectively, and bigger variances. After GHI, the biggest average positive correlations belong to the other irradiance components. Relative humidity and solar zenith have the largest negative correlations. These variables have weaker correlations in the snow and snow-cover datasets. While the correlations for ambient temperature are positive with an average of

0.4 in the full and snow-free datasets, this variable may have a negative correlation in the snow-cover condition where its average value is zero. It is worth mentioning that considering the nature of the semiconducting materials of PV cells, ambient temperature is expected to have a negative correlation with PV power. However, these correlations in Fig. 7.3 are highly affected by the positive correlation of ambient temperature with solar radiation. While the correlations for rain, snowfall, and snow depth for the full datasets are negative, they can have a positive value in the snow and snow-cover conditions.

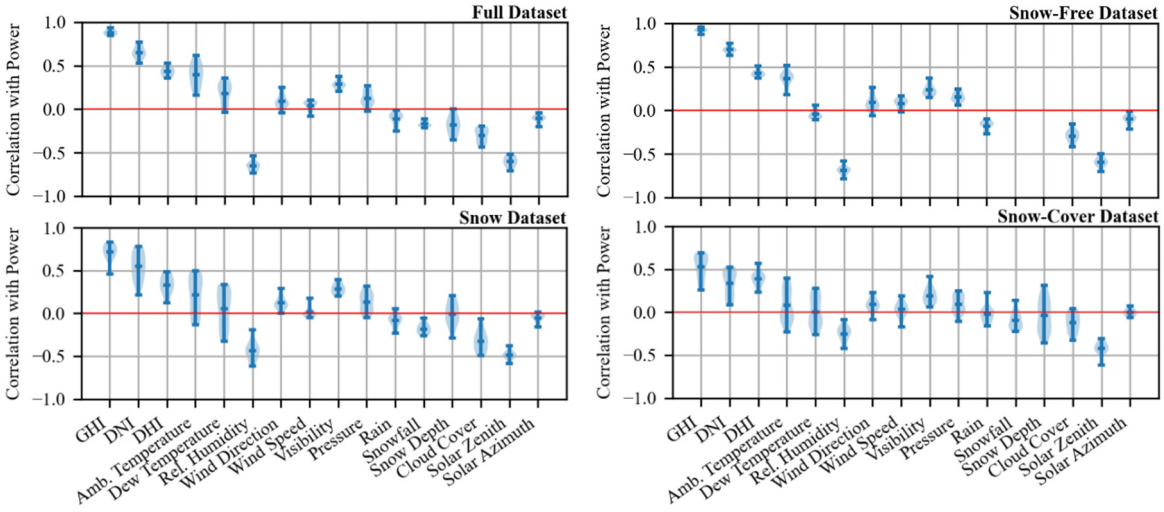


Figure 7.3. Correlation coefficients between the meteorological variables and PV power generation considering all four datasets of 17 PV systems.

Generally, most of the correlation coefficients between the meteorological variables and the PV power are weaker with a larger variance in the snow-cover datasets compared to the snow datasets and the same in the snow datasets compared to the snow-free datasets. As a result, on one hand, snow accumulation on the surface of the panels can make modeling a more difficult problem. On the other hand, it reveals the reason behind the poorer performance of the prediction models in snow conditions if they are trained based on snow-free conditions.

Principal component analysis (PCA) is a well-known linear transformation for reducing the dimensionality of a dataset so that the variables in the low-dimensional dataset are uncorrelated [120]. This helps to obtain a lower-dimensional dataset while preserving the variation of the data as much as possible. By calculating 16 principal components (PCs) for the full, snow, and snow-cover datasets and 14 PCs for the snow-free dataset, PCA is performed over the four aggregated datasets of all 17 PV systems.

The variance ratios for the PCs in each dataset are indicated in Fig. 7.4. The variance ratio indicates the amount of information or variance each PC holds. As can be observed, the first PC holds 19% to 25% of the variance in the data with the highest value for the snow-free dataset and the lowest value for the snow dataset. In order to preserve 60% of the information, the first 4 or 5 PCs should be used.

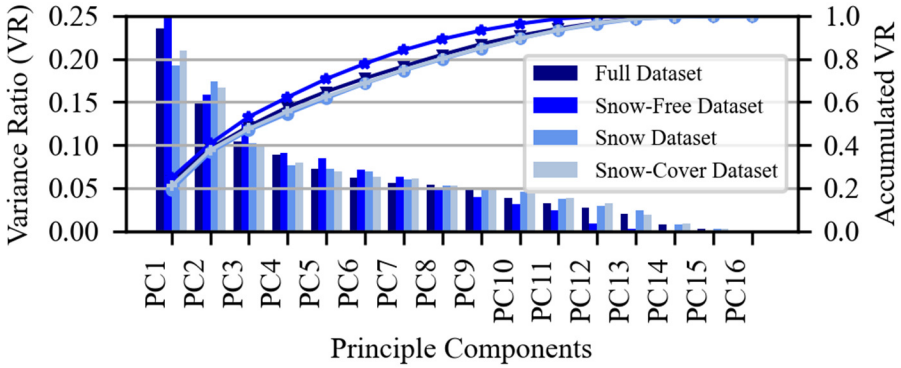


Figure 7.4. Variance ratio for the principal components in the PCA-based datasets.

The biplots for the first two PCs in all four datasets are shown in Fig. 7.5. A biplot indicates the transformed data points and how much weight the variables in the main dataset have on a particular PC and what their directions are. In the full dataset, GHI, ambient temperature, and zenith make the biggest contributions to PC1, and dew point temperature and DNI have the largest impact on PC2. Azimuth, wind speed, and pressure have the smallest contributions to the first PCs in the full dataset. In the snow-free dataset, GHI, humidity, zenith, and DNI have the biggest contributions to PC1, respectively; while, PC2 is mostly affected by dew point temperature and ambient temperature. In the snow dataset, GHI, DNI, and relative humidity have the largest contributions to PC1. Ambient temperature, dew point temperature, and zenith are the most important variables to form PC2 in this dataset. In the snow-cover dataset, ambient temperature and dew point temperature have the biggest impact on PC1; while, GHI, zenith, and DHI have the largest contributions to PC2. By comparing all four biplots, it can be observed that GHI, zenith, ambient, and dew point temperature have the biggest contributions to forming the first PCs in all four datasets. Moreover, the first two PCs in the snow-cover dataset are formed by larger contributions of fewer variables, with the largest belonging to GHI and zenith.

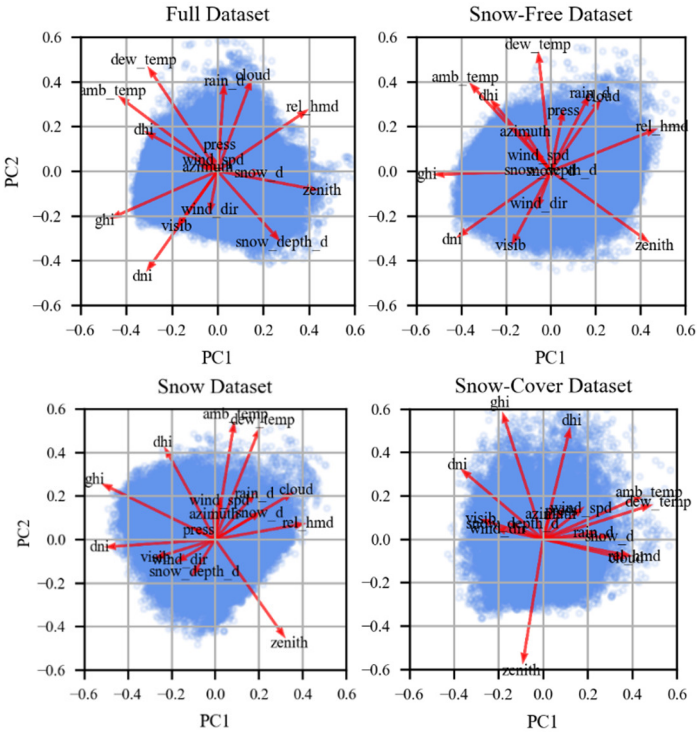


Figure 7.5. Biplots for the first two principal components in the PCA-based datasets.

7.5 PV Power Prediction

After categorizing the data based on snow-related conditions and extracting the snow-free, snow, and snow-cover datasets from the full datasets of all PV systems under study, all datasets are used to develop PV power prediction models using computational intelligence techniques. Separate models will be developed for each full and snow-related condition dataset of each PV system. Considering the PV power as the target variable, the PV power prediction is a regression problem. As such, several computational intelligence techniques, i.e. linear regression, regression tree, random forest, gradient boosting tree, support vector regression, and artificial neural network, are used to develop PV power prediction models. All 16 meteorological variables are used as the attribute variables. The target variable is the hourly PV power. Finally, the performance of the models is compared to find the best PV power predictor. Based on the proposed methodology, categorizing the data into snow-related conditions and using separate PV power prediction models developed specifically for each condition should improve the prediction accuracy.

7.6 Numerical Results

In order to enable the contribution of all the data points of a dataset in training the models, 5-fold cross-validation is implemented. To this end, the dataset is divided randomly into five segments with equal sizes. Then, five different combinations of the segments are formed, four of which are used to train a model, and the one remaining is used to test the trained model. The final performance score is obtained by averaging the scores of the models over five combinations. In order to obtain the optimal hyperparameters of each model, a Bayesian optimization-based hyperparameter tuning is implemented. Python programming language and Scikit-learn library are used to develop the models. In order to provide a comprehensive comparison, five case studies are investigated for the datasets of each PV system as follows:

- Case 1: Separate snow-related conditions models;
- Case 2: Predictions in the snow condition;
- Case 3: Predictions in the snow-cover condition;
- Case 4: Separate PCA-based snow-related conditions models;
- Case 5: Comparison with the existing models.

It is important to note that all the datasets are normalized between zero and one using a scale based on the full datasets.

7.6.1 Case 1: Separate Snow-Related Conditions Models

In this case, all the computational intelligence techniques mentioned in the previous section are separately applied to every four datasets of each PV system. The values of the test RMSE and MAE for the best models are presented as box plots in Fig. 7.6 where the median and mean values are indicated by an orange line and a green diamond, respectively. As can be seen, the best performance has been obtained by the GBT models followed by the RF models in all four datasets. While the errors in the snow-free datasets are slightly lower than the full datasets, the figure shows the poorest performance for the models developed for the snow datasets. These large errors with big variances show that modeling in snow conditions is a challenging task. The errors of the models in the snow-cover datasets are considerably smaller due to the significantly lower magnitude of the PV power when there is a snow cover on the panels, as can be seen in Fig. 7.2. The MAE values also confirm the best performance of the GBT and RF models. In the snow-free

datasets, however, the SVR models have the best accuracy. MAEs are smaller than RMSEs due to large variances in the prediction errors.

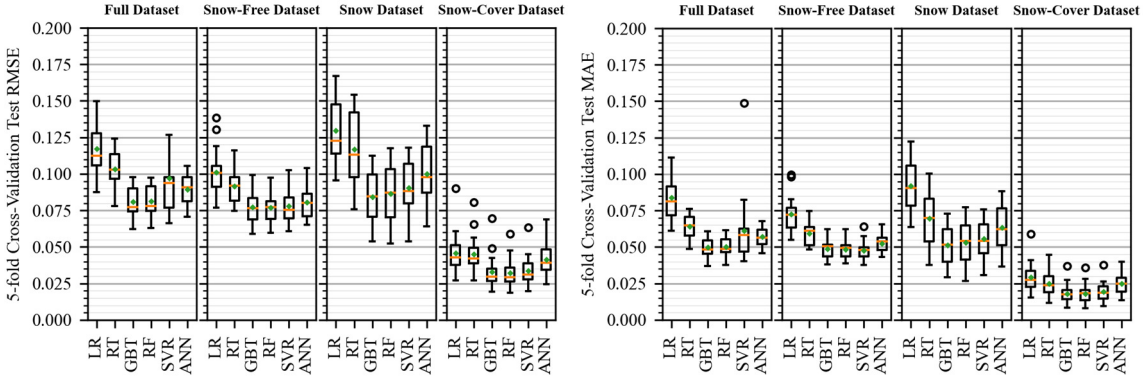


Figure 7.6. 5-fold cross-validation test RMSEs and MAEs for the best models built for four datasets of 17 PV systems.

The average values of the feature importance for the best GBT models considering all the PV systems are indicated in Fig. 7.7. They show how helpful the attribute variables are in predicting the target variable. It can be seen in the figure that the most important variable is GHI, with the maximum average value of 0.88 for the snow-free datasets. However, GHI is less important in the snow and snow-cover datasets where the importance of the other attribute variables such as ambient temperature, relative humidity, and snow depth has increased.

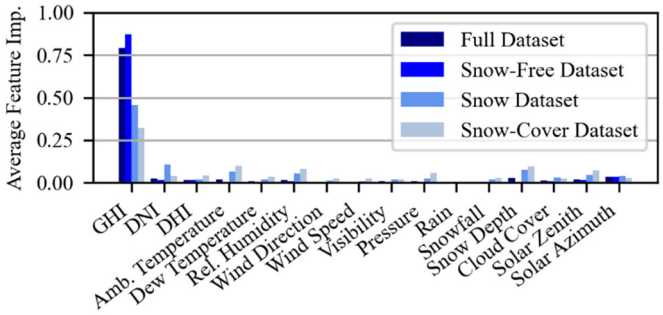


Figure 7.7. Average feature importance in the best GBT models in case 1.

7.6.2 Case 2: Predictions in Snow Condition

In this case, the models with the optimal hyperparameters in case 1 are used to make a prediction on the snow datasets. It shows how accurately the models trained on the full datasets and the snow-free datasets perform in the snow condition. To this end, the snow datasets are randomly divided into five folds. A fold is selected as the test dataset. The full datasets after dropping the data points of the test datasets as well as the

complete snow-free datasets are used for training. All the models with the optimal hyperparameters in case 1 are trained on the training datasets and tested on the test datasets. The same procedure is repeated for each fold. Finally, the average test RMSEs and MAEs over all five folds are calculated. These values together with those for the models built specifically for the snow datasets (same as case 1) are demonstrated in Fig. 7.8.

As can be observed, the best performance in the snow condition has been obtained by the specific models built on the snow datasets followed by those trained on the full datasets. The models trained on the snow-free datasets have worse accuracy in the snow condition since, unlike the full datasets, there is no data pertaining to the snow condition in the snow-free datasets, and the snow and snow-free datasets have different characteristics (as can be seen in Fig. 7.3). Based on both RMSEs and MAEs, the closest performance to the models built on the snow datasets themselves has been obtained by the GBT models trained over the full datasets. The smaller MAEs compared to RMSEs show higher variances in the prediction errors.

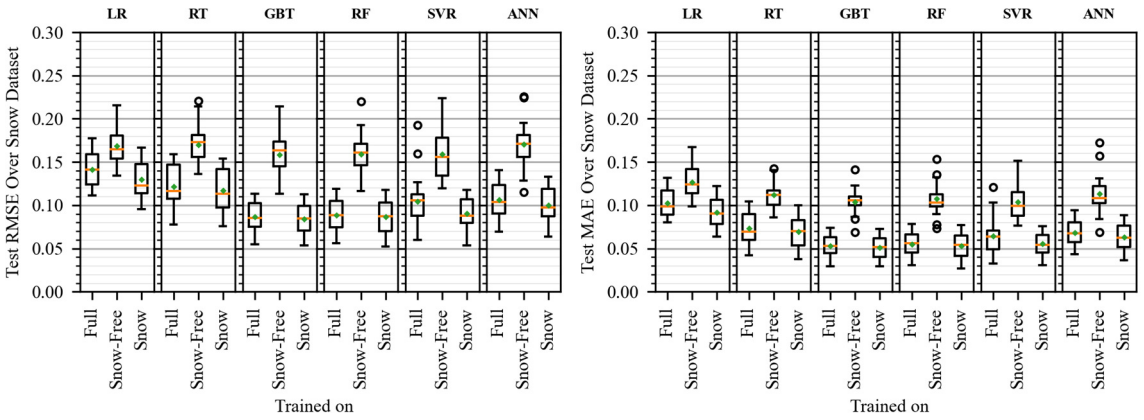


Figure 7.8. Test RMSE and MAE values for the models with the best hyperparameters trained on the Full and snow-free dataset and tested on the snow dataset together with 5-fold cross-validation values for the best models trained and tested on the snow dataset.

7.6.3 Case 3: Predictions in Snow-Cover Condition

In this case, the same procedure in case 2 is applied to the snow-cover datasets. It shows how accurately the models trained on the full, snow-free, and snow datasets perform in the snow-cover condition. Similar to case 2, the test data points in every 5 folds are dropped from the training datasets to make a fair evaluation of the models. The test RMSEs and MAEs of the models together with those for the models built for the snow-cover datasets (same as case 1) are shown in Fig. 7.9.

As can be seen, the worst performance has been obtained by the models trained on the snow-free datasets. It shows that the characteristics of the data in the snow-cover condition differ largely from the snow-free condition. The models trained on the full datasets have better performance since they have seen similar data points in the snow-cover condition in the training stage. However, the models trained on the snow datasets have slightly better accuracy in the snow-cover condition. Nevertheless, the predictions made by the models specifically built on the snow-cover datasets have significantly smaller errors. It shows how beneficial it is to detect the data in the snow-cover condition and develop specific models for them.

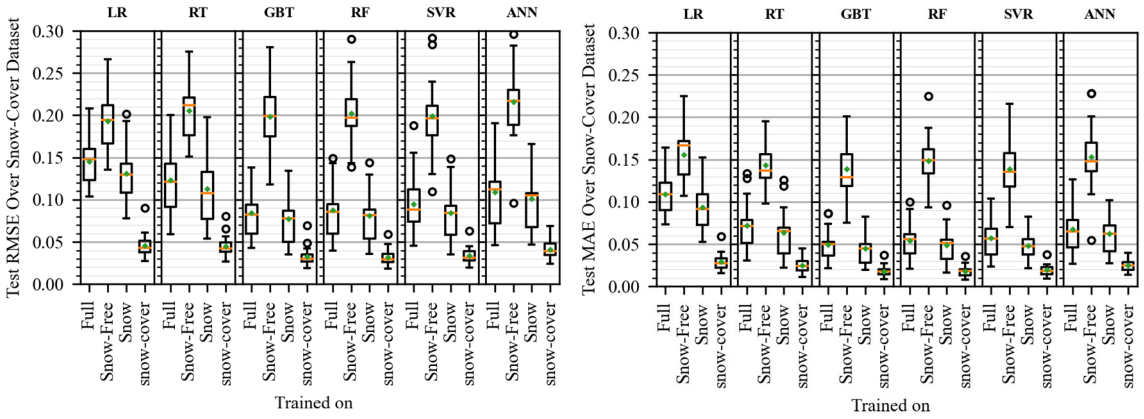


Figure 7.9. Test RMSE and MAE values for the models with the best hyperparameters trained on the Full, snow-free, and snow dataset and tested on the snow-cover dataset together with 5-fold cross-validation values for the best models trained and tested on the snow-cover dataset.

7.6.4 Case 4: Separate PCA-Based Snow-Related Conditions Models

In this case, the accuracy of the best models in case 1 (GBTs) is compared with the accuracy of the models built for the PCA-based datasets using 5-fold cross-validation. Therefore, GBTs with hyperparameter tuning are trained on the datasets formed by the first, two-first, three-first, four-first, and five-first PCs. The test RMSEs, indicated in Fig. 7.10, show that increasing the number of PCs gradually improves the accuracy so that using all the PCs gives the same results as in case 1. Based on the average RMSEs, the accuracy of the linear regression models in case 1 has been reached by using the first four, five, three, and three PCs in the full, snow-free, snow, and snow-cover datasets, respectively. The models based on the first PC result in accuracy almost half of the best ones using all the PCs.

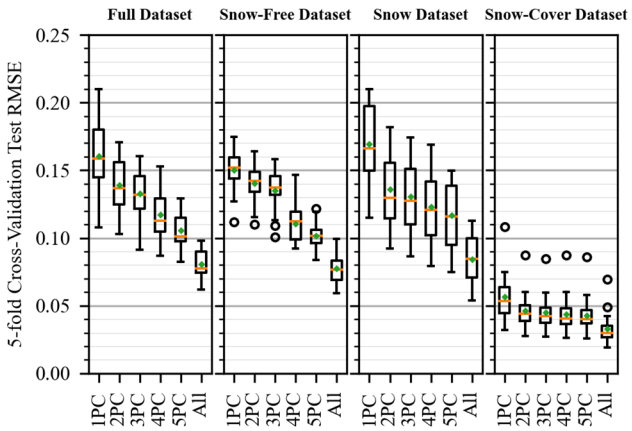


Figure 7.10. 5-fold cross-validation test RMSEs for the best GBT models built for each PCA-based dataset of 17 PV systems using the first PCs.

7.6.5 Case 5: Comparison with Existing Models

In this case, the well-known Marion model [12], the modified version of the Marion model proposed by Øgaard et. al. [71], the classic PV power calculation, a simplified gradient boosting tree model (without considering snow-related parameters), and a combination of the simplified gradient boosting tree and the snow cover detection and quantification based on the Marion model are implemented and applied to the available datasets to perform a comparison between their performance and that of the proposed approach. Since the Marion model requires the logical layout of the PV system in the power calculation step and due to the lack of this information for most of the systems under study, only four PV systems, i.e. system numbers 2, 4, 10, and 16, are selected to apply the models. These systems are formed by 4, 1, 1, and 3 strings with a tilt angle of 35°, 18°, 25°, and 10° which are installed with 4, 2, 3, and 3 vertical modules in the landscape, portrait, portrait, and portrait modes, respectively. The main datasets of the systems contain records for 37, 38, 16, and 55 months, respectively.

At first, the hourly plane-of-array irradiance and cell temperature are calculated for the main dataset (dataset involving all 24 hourly values of the parameters) of each system using the PVLIB Python based on the isotropic sky diffuse model and the Sandia array performance model, respectively [138]. According to the Marion model, the values of the snow cover percentage are determined based on the daily values of snow depth on the ground, the snow sliding condition, and the sliding distance (which depends on the tilt angle). The power generation of the PV systems is then calculated based on the POA irradiance, cell temperature, and the strings with no snow cover. It is important to note

that unlike the work in [12] with accurate local measurements, the meteorological data in the current study have been gathered from different publicly available online sources based on the records of the weather stations several kilometers away from the PV systems' locations or extracted from satellite images with low accuracy. Hence, calculating and modeling the values of PR25 in the Marion model has been inaccurate and almost impossible. Thus, only the quality degradation coefficient of the panels and a constant coefficient representing the other array capture losses, obtained by comparing the measured PV power with the modeled PV power on several days in the snow-free condition, have been considered. In order to be able to compare the performance of the Marion model with the proposed approach (separate computational intelligence-based modeling for each snow-related condition), the predictions of the Marion model corresponding to data points with solar irradiance equal to zero are discarded and the rest, called the full dataset, is split into snow-free, snow, and snow-cover datasets based on the indices obtained for the snow-related datasets of each PV system in section 7.3.

Moreover, a modified version of the Marion model proposed by Øgaard et. al. [71] (called the Øgaard model here) is implemented. Unlike the Marion model, the Øgaard model uses a snow sliding distance which not only depends on the tilt angle of the system but also on the snow depth. The same procedure as the Marion model is applied to the datasets of the four PV systems using the Øgaard model.

In order to evaluate the application of the snow cover detection and quantification of the Marion model in the computational intelligence-based PV power prediction, GBT models with optimal hyperparameters obtained in case 1 for the full datasets of the four PV systems are trained after dropping the snowfall and snow depth parameters to exclude the effect of snow on the power prediction. To this end, the full dataset of each PV system is linearly split into 5 folds. Then, a model is trained on 4 folds and makes a prediction on the other fold. Finally, the predictions on all 5 folds are put together. We call this model simplified GBT since it is trained on the datasets after dropping 2 snow-related features. Then, the number of snow-free strings based on the snow cover detection and quantification of the Marion model is used to modify the power predictions.

The classic PV power calculation without considering snow effects based on the PVWatts DC model [126] is also applied to the datasets to give a better insight into the performance of the models. The prediction errors for each model are calculated based on

the real values of the PV power and shown in Fig. 7.11 together with the RMSEs of the GBT models developed specifically for each snow-related condition in case 1. As can be seen, the accuracy of the models is very similar in the snow-free condition for all four systems; however, the GBT models perform slightly better since the classic, Marion, and Øgaard models are very sensitive to the accuracy of the input data. In the snow condition, the Marion model has resulted in a better prediction compared to the classic model (almost half of the error for system 10); while the Øgaard model has performed poorer except for system 10. The simplified GBT model has resulted in reduced prediction errors; while adding the Marion snow cover detection and quantification concept into this model has increased the prediction errors except for system 10. For the snow datasets of all four systems, the proposed approach had the best accuracy. The prediction errors in the snow-cover condition, as the main focus of this study, provide the most valuable insight into the performance of the models. As expected, the classic model has the lowest accuracy. The Marion model can reduce the prediction error to even one-third of that when a snow cover is probable on the panels. The Øgaard model has further reduced the RMSEs except for system 16. The simplified GBT has a slightly poorer performance compared to the Marion model; while combining this model with the Marion snow cover detection procedure has resulted in accuracy close to that of the Øgaard model. It can be clearly seen in the figure that the GBT models developed specifically for the snow-cover condition using the proposed approach present the best performance with significantly lower prediction errors compared to the other models in the case of all PV systems.

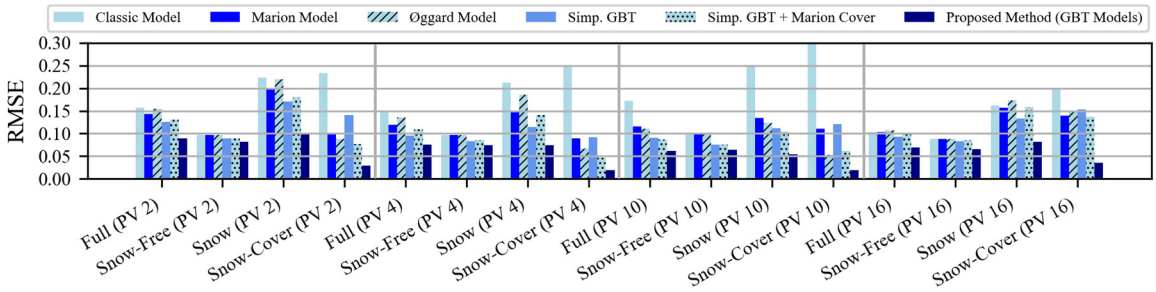


Figure 7.11. The prediction errors of the proposed approach, Marion model, Øgaard model, etc. for all snow-related datasets of four chosen PV systems in case 5.

Generally, the advantages of the proposed approach of PV power prediction in snow conditions compared to the other models analyzed in this study are as follows:

- Providing better PV power prediction accuracy, especially in the case of a probable full/partial snow cover;

- Requiring no knowledge of the PV system technical specifications and the underlying electrical/physical formulations and principles, while these are implicitly captured by the proposed models;
- Achieving superior performance compared to the other models when very accurate data measured at the system's location is not available.

7.7 Conclusion

PV power prediction in snow conditions has been addressed in this chapter. Comprehensive data analysis and computational intelligence-based modeling have been applied to the data of the hourly power generation and 16 meteorological variables of 17 PV systems across Canada. The snow-free, snow, and snow-cover conditions datasets have been extracted from the full dataset of each PV system using a proposed 3-step snow-cover detection approach to perform the study. As the results showed, GHI has been the most correlated variable with the PV power in all the conditions. The GBT and RF models have achieved the best performance for all four datasets. However, the prediction accuracy has been lower in the snow condition. It has been shown that the prediction errors in the snow and especially snow-cover conditions are smaller by using specific models built for the snow and snow-cover datasets compared to those built for the full and snow-free datasets. The results also show significantly lower prediction errors especially in the snow-cover condition using the proposed approach compared to the Marion model and some other models. This proves the effectiveness of the probable snow cover detection and distinct prediction models built for the snow-related conditions in PV power modeling for systems in snow-prone areas.

CHAPTER 8

COMPUTATIONAL INTELLIGENCE-BASED FORECASTING OF PV POWER IN SNOW CONDITIONS AND ITS APPLICATION IN A PEVS AGGREGATOR SCHEDULING

8.1 Introduction

A PEVs aggregator, as an intermediate entity between the PEV owners and the power grid, offers high-tech smart solutions to reduce the charging costs for the PEV owners. This can happen by taking control of the charging or even discharging processes of the vehicles and applying a smart strategy utilizing the smart grid communication infrastructure. This requires the aggregator to be equipped with an optimal management strategy in the form of computer software in which forecasting renewable power generation can play an important role. In this chapter, the methodology proposed to achieve the first part of the third objective, defined as computational intelligence-based PV power forecasting for a PEVs aggregator scheduling problem with support for photovoltaic power penetrated distribution grid under snow conditions, is described.

8.2 Framework of the Problem

In this study, the optimal day-ahead scheduling of an independent PEVs aggregator providing smart charging solutions for the PEVs in a town in a snow-prone region is investigated. The distribution grid operated by the local DSO is equipped with a large PV plant. The aggregator is assumed to be able to participate in the day-ahead and real-time energy markets through the local DSO in order to provide affordable services for its customers. The direct participation of such a small-scale PEVs aggregator in the market is assumed to be impractical due to the market structure and requirements in terms of the minimum size of the participants. Instead, the aggregator is committed by the local DSO to avoid charging surges and grid overloads due to the simultaneous charging of a large number of PEVs. Such an agreement is beneficial for the aggregator financially and for the DSO technically. This requires the aggregator to adapt its charging policy to the

technical limitations of the grid. The main framework of the problem indicating the interactions between all the participants is shown in Fig. 8.1.

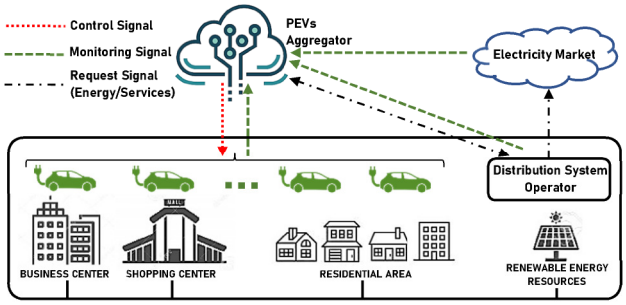


Figure 8.1. Main framework of the problem.

The structure of the energy market is assumed to be similar to that of the New York Independent System Operator (NYISO). The DSO, as a load-serving entity (LSE), can purchase energy by submitting hourly bids to the DAM. Considering the market rules in terms of the minimum size of the participants, the aggregator submits its energy requirements through the local DSO. So, the hourly aggregated values of the grid's net energy demand and the PEVs' charging/discharging power are submitted as the day-ahead bids. The differences in the load consumption of the LSEs as compared to the day-ahead bids are settled at the RTM. Therefore, the DSO has to purchase the extra real-time energy demand (for both the grid and PEVs if there is a difference between the day-ahead bids and the real-time demand) at the RTMCPs.

Before the DAM closure, the aggregator needs to submit its day-ahead hourly energy requests to the DSO. In order to obtain an optimal day-ahead schedule, the aggregator implements a computational intelligence-based approach to minimize its expected cost for the next operating day. Considering the uncertain nature of the driving patterns and energy market prices, the aggregator's decision-making is developed by a two-stage stochastic programming approach [85]. This is a reliable method to address uncertainties by means of probable scenarios of the input variables. The objective is to minimize the charging cost of the vehicles while it is constrained to the satisfaction of the PEV owners in terms of the final battery state of charge (SoC) at the departure time.

From the DSO's viewpoint, the inevitable errors in the day-ahead predictions of the hourly PV power generation (especially for a snowy day) and the hourly load demand of the grid cause inaccuracy in the DSO's day-ahead scheduling. In the case of underestimating the net load of the grid (the aggregated customers' load minus the PV

power generation) and a consequent difference between the actual demand and the day-ahead bids, and due to the lack of flexibility, the DSO has to purchase the extra amount of the demanded energy from the RTM. This exposes the DSO to highly volatile RTMCPs. In this condition, the PEVs aggregator is able to provide the extra energy demanded by the grid locally through an out-of-market balancing service which is proposed in this chapter. This can happen by manipulating the charging and discharging processes of the vehicles in real-time thanks to the flexibility in the PEVs' charging demand. The aggregator provides this service as far as it is beneficial for the PEV owners in terms of the charging cost. The day-ahead bids of the DSO together with the actual values of the grid load demand and the PV plant power generation determine the extra energy demanded by the DSO in real time. The proposed balancing service requires the aggregator to consider the uncertain nature of both the grid load demand and the PV power generation as the realization of the probable scenarios in its day-ahead scheduling problem. Before submitting the bids to the DAM, the DSO notifies the aggregator of its bids. Considering the DSO's day-ahead bids, the probable scenarios of the uncertain variables, and the balancing tariff (that the DSO pays to the aggregator for the energy provided by the vehicles in real-time to support the grid), the aggregator submits its optimal decisions for purchasing/selling energy to the DSO. In real-time, the aggregator notifies the DSO of the supported percentage of the grid's energy shortage that can be provided by the vehicles through the balancing service. This local out-of-market service is expected to be a suitable tool for small-scale PEVs aggregators that cannot participate directly in the markets, to reduce the charging costs of their customers, and to ensure a good source of flexibility for the DSO in providing its real-time extra energy demand and avoiding highly fluctuating RTMCPs.

8.3 Mathematical Formulation of the Problem

This section presents the mathematical formulation of the aggregator's day-ahead scheduling problem as a two-stage stochastic programming approach. The developed approach consists of two stages, namely here-and-now (the first stage) and wait-and-see (the second stage). Each stage represents a point in time where decisions are made or where uncertainties partially or fully vanish [85]. In the first stage, optimal decisions of purchasing/selling energy from/to the DAM are made before the realization of the scenarios. In the second stage, optimal decisions on (i) the charging/discharging rates of the connected PEVs, (ii) purchasing energy from the RTM, and (iii) providing the

balancing service for the DSO, are made after knowing the actual realization of the scenarios. Each decision in the first stage is represented by a single variable, while there are separate variables for each decision in the second stage considering the realization of the scenarios. The inputs of the developed approach are the values of DAMCPs, characteristics of the PEVs, in terms of the battery capacity, nominal charging rate, etc., topology of the grid, and a set of scenarios for each uncertain variable, including RTMCPs, driving patterns, load demand of the grid, and PV plant power generation. Fig. 8.2 shows the schematic of the proposed two-stage stochastic programming approach in detail. The problem is formulated as a mixed-integer linear programming (MILP) approach. The subscripts d, t, ω, l, and i/j represent PEV, time step, scenario, grid line, and grid node numbers belonging to the sets D, T, Ω, L, and N, respectively. The objective function and the constraints forming the proposed stochastic programming approach are as follows:

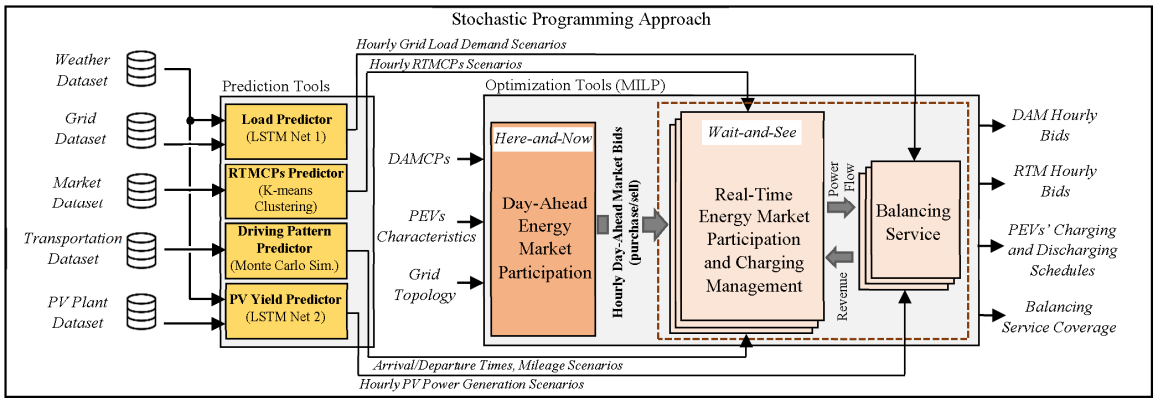


Figure 8.2. Structure of the developed stochastic programming approach for the aggregator's day-ahead scheduling.

8.3.1 Objective Function

The objective of the problem is to minimize the expected total daily charging cost of all PEVs. The revenue obtained from providing the proposed balancing service is added to the objective function. The deterministic equivalent of the stochastic programming approach is as follows:

$$Minimize\ Z = \text{Cost}^{DAM} + \sum_{\omega=1}^{\Omega} \pi_{\omega} \times \left[\left(\text{Cost}_{\omega}^{RTM} + \text{Cost}_{\omega}^{Batt.\ Deg.} \right) - \sum_{t=1}^{T} \text{Revenue}_{\omega,t}^{Balancing} \right] \quad (8.1)$$

where Z is the objective function. Cost^{DAM} and Cost^{RTM} are the costs ($) of purchasing energy from DAM and RTM, respectively. $\text{Cost}^{Batt.\ Deg.}$ is the battery degradation cost for all PEVs due to discharging ($). $\text{Revenue}^{Balancing}$ is the revenue obtained by providing the proposed balancing service ($). π_{ω} is the probability of each scenario ω. The terms of the objective function (Z) are as follows:

- Day-Ahead Market Cost

$$\text{Cost}^{DAM} = \sum_{t=1}^{T} (\text{Price}_t^{DAM}) \times \text{Bid}_t^{DAM} \quad (8.2)$$

where Price^{DAM} is DAMCP ($/kWh). Bid^{DAM} is the hourly bid of the aggregator in the DAM. This cost is positive/negative if energy is purchased/sold. The aggregator's bids are added to the DSO's bids and then submitted to the market.

- Real-Time Market Cost

$$\text{Cost}_{\omega}^{RTM} = \sum_{t=1}^{T} (\text{Price}_{\omega,t}^{RTM}) \times \text{Bid}_{\omega,t}^{RTM} \quad , \quad \forall \omega \quad (8.3)$$

where Price^{RTM} is RTMCP ($/kWh). Bid^{RTM} is the aggregator's hourly bid in the RTM. In the case that the aggregator cannot meet its hourly commitment in real-time based on the day-ahead sold energy, this is considered as an increase in the LSE's demand in real-time. So, the RTMCPs should be paid by the aggregator to the RTM through the DSO for this deviation from the commitment. This cost is considered to be always positive because the DSO participates in the market as an LSE and normally submits the major part of its bids in the DAM.

- Battery Degradation Cost

$$\text{Cost}_{\omega}^{Batt.\ Deg.} = \sum_{d=1}^{D} \sum_{t=1}^{T} \left(\frac{c^{b} \times Cap_{d}^{Batt.} + c^{L}}{L^{c} \times Cap_{d}^{Batt.} \times DoD} \right) \times \frac{P_{\omega,d,t}^{dch}}{\eta^{ch}}, \quad \forall \omega \tag{8.4}$$

Discharging causes degradation of the batteries and its cost can be calculated based on the discharging power using equation (8.4) [92]. c^{b}, c^{L}, L^{c}, DoD, $Cap^{Batt.}$, η^{ch}, and P^{dch} represent the battery cost ($/kWh), labor cost for replacing the old battery with a new one ($), the life cycle of the battery, depth of discharge, battery capacity, charger efficiency, and the discharging power, respectively.

- Balancing Service

In order to integrate the proposed balancing service into the scheduling problem of the aggregator, a revenue term is added to the objective function that reduces the total charging cost of the PEVs based on the amount of energy they provide for this service. This allows the PEVs to provide part of the extra energy demanded by the DSO in real-time which is beneficial for them based on the structure shown in Fig. 8.3. The effect of this revenue can be applied later as a reduction to the charging bills of those PEVs participated in the service in the billing calculation process. The following equations form the mechanism of the proposed service.

$$\text{Revenue}_{\omega,t}^{Balancing} \leq \sum_{lll=1}^{LLL} Tariff_{\omega,lll,t}^{Balancing} \times Ppos_{\omega,lll,t}^{Balancing}, \quad \forall \omega, t \tag{8.5}$$

$$\sum_{d=1}^{D} (P_{\omega,d,t}^{dch} - P_{\omega,d,t}^{ch}) + (\text{Bid}_{t}^{DAM} + \text{Bid}_{\omega,t}^{RTM}) = Pdch_{\omega,t}^{PEVs}, \quad \forall \omega, t \tag{8.6}$$

$$Pdch_{\omega,t}^{PEVs} = \sum_{lll=1}^{LLL} Ppos_{\omega,lll,t}^{Balancing} - \sum_{lll=1}^{LLL} Pneg_{\omega,lll,t}^{Balancing}, \quad \forall \omega, t \tag{8.7}$$

$$\text{Bid}_{t}^{DAM} + \text{Bid}_{\omega,t}^{RTM} \leq \sum_{d=1}^{D} (P_{\omega,d,t}^{ch}), \quad \forall \omega, t \tag{8.8}$$

where $\text{Revenue}^{Balancing}$ is the revenue obtained by offering the balancing service to the DSO ($). $Ppos^{Balancing}$ is the hourly supplied power by the PEVs into the grid exclusively for the balancing service. $Tariff^{Balancing}$, as the tariff of the service, can be determined using a pricing mechanism that maximizes the obtained profit by the service. *lll* denotes the segment number for $Pdch^{PEVs}$ in Fig. 8.3 (here, LLL = 2). The revenue of the

balancing service is simply calculated in (8.5) by multiplying the service tariff by the hourly supplied power by the PEVs into the grid exclusively for this service. In order to be able to calculate this supplied power, the PEVs' market participation is also needed to be considered and distinguished from the PEVs' participation in the service. Moreover, the revenue has to be calculated for the supplied power up to the requested power by the DSO so that $Ppos^{Balancing} \leq$ the requested power by the DSO. These are formulated in equations (8.6)-(8.8). The equations needed to divide the values of $Pdch^{PEVs}$ into two segments using auxiliary variables are added to the problem.

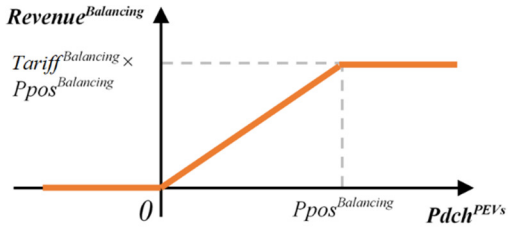


Figure 8.3. Structure of the proposed balancing service.

8.3.2 Constraints

The presented objective function is constrained to the following equations.

- Charging and Discharging Rates

$$P^{ch}_{\omega,d,t} \leq P^{charger}_{d} \text{ and } P^{dch}_{\omega,d,t} \leq P^{charger}_{d} \times \eta^{ch}, \qquad \forall \omega, d, t \tag{8.9}$$

P^{ch} and P^{dch} are both positive variables and are limited by the nominal charging rate of the battery charger.

- State of Charge Variations

$$SoC_{\omega,d,t} = SoC_{\omega,d,t-1} + \frac{(P^{ch}_{\omega,d,t} \times \eta^{ch} - P^{dch}_{\omega,d,t}/\eta^{ch})}{Cap^{Batt.}_{d}} \times 10^{2}, \qquad \forall \omega, d, t \tag{8.10}$$

The battery SoC at each hour is determined by the SoC at the previous hour and the energy exchange with the grid, which depends on the charging/discharging power of the battery at that hour.

- State of Charge Bounds

$$20 \leq SoC_{\omega,d,t} \leq 100, \qquad \forall \omega, d, t \tag{8.11}$$

In general, it is recommended to avoid discharging a battery if its SoC is lower than 20% due to battery degradation.

- PEV Owners' Satisfaction in Terms of Final SoC

$$SoC_{\omega,d}^{Arr} + \sum_{t=1}^{T} \frac{(P_{\omega,d,t}^{ch} \times \eta^{ch} - P_{\omega,d,t}^{dch}/\eta^{ch})}{Cap_{d}^{Batt.}} \times 10^2 = SoC_{\omega,d}^{Final\ Guar.}, \qquad \forall \omega, d \qquad (8.12)$$

This constraint is applied to the vehicles that are connected long enough to the grid to reach the desired SoC. These vehicles will be fully charged and get ready for the next trip. The other vehicles will be charged with a constant charging rate.

- Aggregator's Bids

$$\text{Bid}_{t}^{DAM} + \text{Bid}_{\omega,t}^{RTM} \geq \sum_{d=1}^{D} (P_{\omega,d,t}^{ch} - P_{\omega,d,t}^{dch}), \qquad \forall \omega, t \qquad (8.13)$$

This constraint guarantees that the aggregator's hourly bids meet the total hourly energy demand of the PEVs; so that, the total energy purchased from the DAM and RTM at each hour has to cover the charging load of the vehicles.

- Power Flow Equations

$$\begin{cases} p_{l,t,\omega}^{s} = g_l V_{i,t,\omega} - g_l V_{j,t,\omega} + \left(\frac{1}{2}\right) p_{l,t,\omega}^{loss} - b_l \delta_{i,t,\omega} + b_l \delta_{j,t,\omega}, \\ q_{l,t,\omega}^{s} = -b_l V_{i,t,\omega} + b_l V_{j,t,\omega} + \left(\frac{1}{2}\right) q_{l,t,\omega}^{loss} - g_l \delta_{i,t,\omega} + g_l \delta_{j,t,\omega}, \end{cases} \quad \forall l, t, \omega \qquad (8.14)$$

$$\begin{cases} P_{i,t,\omega}^{g} - P_{i,t,\omega}^{d} - \sum_{l \in L} A_{il} . p_{l,t,\omega}^{s} - \sum_{l \in L} A'_{il} . p_{l,t,\omega}^{loss} = 0, \\ Q_{i,t,\omega}^{g} - Q_{i,t,\omega}^{d} - \sum_{l \in L} A_{il} . q_{l,t,\omega}^{s} - \sum_{l \in L} A'_{il} . q_{l,t,\omega}^{loss} + q_{i}^{shunt} = 0, \end{cases} \quad \forall l, t, \omega \qquad (8.15)$$

$$\begin{cases} p_{l,t,\omega}^{loss} = \left((p_{l,t,\omega}^{s})^2 + (q_{l,t,\omega}^{s})^2\right) \times r_l, \\ q_{l,t,\omega}^{loss} = \left((p_{l,t,\omega}^{s})^2 + (q_{l,t,\omega}^{s})^2\right) \times x_l, \end{cases} \quad \forall l, t, \omega \qquad (8.16)$$

The linearized AC power flow model is extracted from [139]. The active-reactive power flows through the grid's lines and the active-reactive power equilibrium in the grid's nodes are formulated in (8.14) and (8.15), respectively. Equation (8.16), which calculates the active-reactive power losses in the grid's lines based on the flowing power

and the resistance-reactance of the lines, will be linearized using the piecewise linear approximation method based on the following set of equations [139].

$$p^{S}_{l,t,\omega} = \sum_{ll=1}^{LL} \beta^{+}_{ll,l,t,\omega} - \sum_{ll=1}^{LL} \beta^{-}_{ll,l,t,\omega} \ , \qquad \forall \text{ l,t, } \omega \tag{8.17}$$

$$(p^{S}_{l,t,\omega})^2 = \sum_{ll=1}^{LL} \alpha_{ll} \times \beta^{+}_{ll,l,t,\omega} + \sum_{ll=1}^{LL} \alpha_{ll} \times \beta^{-}_{ll,l,t,\omega} \ , \qquad \forall \text{ l,t, } \omega \tag{8.18}$$

$$(p^{S}_{l,t,\omega}/M) \le \mathrm{h}_{l,t,\omega} \le 1 + (p^{S}_{l,t,\omega}/M) \ , \qquad \forall \text{ l,t, } \omega \tag{8.19}$$

$$\begin{cases} 0 \le \beta^{+}_{ll,l,t,\omega} \le h_{l,t,\omega} \times \bar{\beta}, \\ 0 \le \beta^{-}_{ll,l,t,\omega} \le (1 - h_{l,t,\omega}) \times \bar{\beta}, \end{cases} \qquad \forall ll, l, \text{t, } \omega \tag{8.20}$$

$$\begin{cases} (\beta^{+}_{ll,l,t,\omega} - \bar{\beta})/M \le h^{+}_{ll,l,t,\omega} \le 1 + (\beta^{+}_{ll,l,t,\omega} - \bar{\beta})/M, \\ (\beta^{-}_{ll,l,t,\omega} - \bar{\beta})/M \le h^{-}_{ll,l,t,\omega} \le 1 + (\beta^{-}_{ll,l,t,\omega} - \bar{\beta})/M, \end{cases} \qquad \forall ll, l, \text{t, } \omega \tag{8.21}$$

$$\begin{cases} \beta^{+}_{ll,l,t,\omega} \le h^{+}_{ll-1,l,t,\omega} \times \bar{\beta}, \\ \beta^{-}_{ll,l,t,\omega} \le h^{-}_{ll-1,l,t,\omega} \times \bar{\beta}, \end{cases} \qquad \forall ll, l, \text{t, } \omega \tag{8.22}$$

Equations (8.19)-(8.20) and (8.21)-(8.22) guarantee the upper and lower bounds of the linear sections and the connection of the sections, respectively. Similar equations are also used for the reactive power flows.

- Grid Normal Operation Commitment

$$p^{S}_{\text{"1"},t,\omega} \le P^{Max} \ , \qquad \forall \text{ t, } \omega \tag{8.23}$$

As mentioned earlier, the aggregator is committed to maintaining the normal operation of the grid by avoiding the main transformer overload due to the PEVs' charging load. This helps keep the load peak of the grid low and prevents excessive voltage drops in the grid.

8.3.3 Risk Assessment

Despite the aforementioned risk-neutral decision-making formulation, a risk assessment using the popular conditional value-at-risk (CVaR) as the risk measure is performed to evaluate the proposed approach considering the risk of variability associated with the aggregator's cost [85]. The objective function of the risk-averse approach is defined in equation (8.24). All the constraints of the risk-neutral formulation together with equations (8.25) and (8.26) constitute the new optimization problem.

$$Minimize\ Z^{risk-averse} = (1-\varphi) \times Z + \varphi \times \left((\frac{1}{1-\gamma}) \sum_{\omega=1}^{\Omega} \pi_\omega \times s_\omega - \xi \right) \quad (8.24)$$

$$\text{Cost}^{DAM} + \text{Cost}_\omega^{RTM} + \text{Cost}_\omega^{Batt.Deg.} - \sum_{t=1}^{T} \text{Revenue}_{\omega,t}^{Balancing} + \xi \leq s_\omega, \ \ \forall \omega \in \Omega \quad (8.25)$$

$$s_\omega \geq 0, \quad \forall \omega \in \Omega \quad (8.26)$$

where φ and γ are the risk aversion parameter and confidence level, respectively. s and ξ are the risk-associated continuous variables.

8.4 Modeling of Uncertain Input Variables

Concerning the aggregator's scheduling as a stochastic programming approach, two scenario-generation methods for driving patterns and RTMCPs are presented. Considering the proposed balancing service, two predictors for the load demand of the grid and PV power generation are proposed.

8.4.1 Driving Patterns of PEV Owners

The uncertain driving patterns can be modeled by three attributes namely the arrival and departure times (when a PEV arrives/departs at/from a charging station), and the daily mileage (total distance a PEV travels on a day). All PEVs can get connected to the grid through residential chargers and some are randomly selected to get connected in public charging stations (PCSs) during the day. Two types of PCSs, i.e. business center PCSs and shopping center PCSs, are considered. The arrival and departure times of the PEVs are generated randomly based on the widely used normal probability density functions (pdfs). Daily mileages are generated using a Lognormal pdf. The parameters of the pdfs for each type of charging station are presented in Table 8.1.

Table 8.1. Parameters of the pdfs related to driving patterns

Station	Residential		Business PCSs		Shopping PCSs		Daily Mileage
Parameter	Arr. Time	Dep. Time	Arr. Time	Dep. Time	Arr. Time	Dep. Time	
Mean	20	7.5	8.5	14	16	18	3.715
Stand. Dev.	1.5	0.75	0.5	0.5	1	1	0.6

The stochastic driving patterns are generated using a Monte Carlo simulation based on the proposed flowchart shown in Fig. 8.4. At first, arrival and departure times at/from home are generated for all vehicles. Then based on the number, type, and capacity of the

PCSs, some PEVs are randomly selected to get charged at PCSs. The arrival and departure times of the selected PEVs at/from the PCSs are finally generated considering the previously generated arrival and departure times.

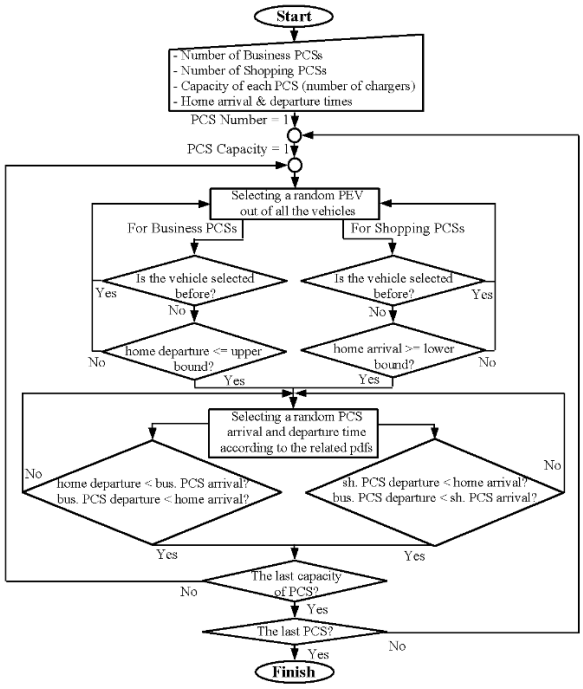


Figure 8.4. Flowchart of stochastic driving patterns of PEV owners.

The state of charge (*SoC*) of a PEV's battery at its arrival time is calculated based on the generated driving patterns:

$$SoC^{Arr} = \begin{cases} SoC^{LDep} - \dfrac{Mil \times E}{Cap^{Batt.}} \times 100, & Mil < 0.8AER \\ 20\%, & Mil > 0.8AER \end{cases} \tag{8.27}$$

8.4.2 Real-Time Market Clearing Prices

Considering the participation of the aggregator in the RTM, the probable scenarios of RTMCPs should be integrated into the problem. The RTMCPs are normally highly volatile since they depend on many factors including the uncertain behavior of the market participants. Thus, K-means clustering is implemented to generate the scenarios based on K centroids of the clusters over historical data. The probability of each scenario is calculated based on the number of members of each cluster. This method enables the aggregator to consider all various probable RTMCP patterns in a historical dataset.

The historical data of the NYISO is used for this purpose [140]. A one-time market clearing process at the beginning of the 24-hour scheduling horizon is considered to reduce complexity. 23 Feb. 2015, which is a winter day, is arbitrarily chosen as the operating day. The hourly RTMCPs for 15 days before that date constitute the dataset and 5-means clustering is applied to that.

8.4.3 Load Demand of the Grid

Integrating the proposed balancing service to the scheduling problem of the aggregator demands day-ahead predictions of the hourly load demand of the grid which are obtained here using a computational intelligence technique. In order to provide real historical data on the hourly load demand of a town, a dataset corresponding to the city of Dayton, Ohio is extracted [141] which was the closest city to New York with a publicly available dataset. 23 Feb. 2015 (Monday) is chosen as the operating day. Considering the periodic patterns in the load demand of the grid, its high correlation with ambient temperature, days of the week, and months of the year, and its lower uncertainty compared to other uncertain input variables in this study, the load demand prediction model is built based on a historical data containing the most similar data points to the operating day. This enables us to build a model faster with the most relevant data in a smaller dataset. In this way, the model can be rebuilt for each operating day by shifting the dataset by one day. To this end, the data for 30 days before the operating day, and 30 days before and 30 days after the same date in the previous year (2014) constitute the dataset that covers measurements over 3 months. The hourly values of ambient temperature are extracted for the city of Dayton [128]. The attribute and target variables are:

- Att. 1: Weekday/weekend numbers (1 for Monday to 7 for Sunday);
- Att. 2: Hours (1 to 24);
- Att. 3: Temperature;
- Target: Hourly load demand of the grid.

The line graphs of the load and temperature in the dataset are shown in Fig. 8.5. Considering the nature of the data as time series, a long short-term memory network, as a recurrent deep learning model, is used to build the load demand predictor [124] which can capture long-term dependencies in the data. The first 2016 data points (84 days) are used to train the model and the next 168 data points (the next 7 days) are used as the

validation set. Bayesian optimization is implemented to select optimal hyperparameters based on the validation set. Then, the model is used to make a prediction on the unseen data of the operating day. The mean squared errors together with the optimal hyperparameters of the model are presented in Table 8.2. The line graphs of the real and predicted normalized values of the target variable on the validation set and the test set are shown in Figs. 8.6 (a) and 8.6 (b), respectively. The MSE value on the test set is 0.00204. The figures demonstrate the ability of the model to predict the load demand of the grid.

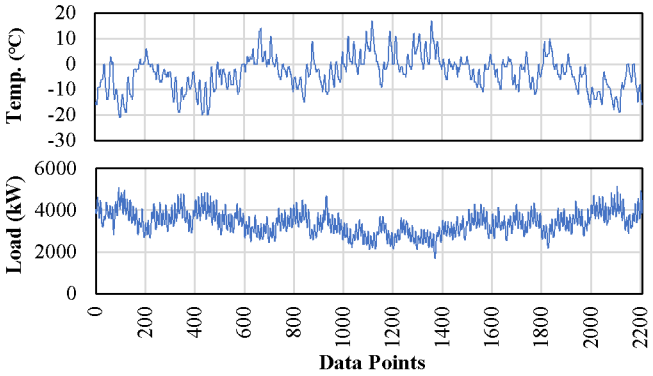


Figure 8.5. Line graphs of the load demand and temperature in the historical dataset of the load demand predictor.

Table 8.2. Parameters of the grid load demand predictor

Model	Validation MSE	Test MSE	Num. of Hidden Units	Initial Learning Rate	Learning Rate Drop Period	Learning Rate Drop Factor	Maximum Epochs	Mini-Batch Size
Load Predictor (LSTM Net 1)	0.00214	0.00204	34	0.1	97	0.5	589	9

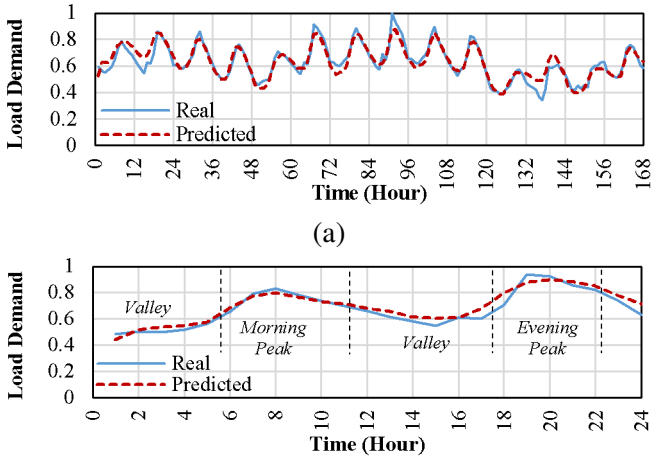


Figure 8.6. The real and predicted values of the hourly load demand using the LSTM model on (a) the validation set and (b) the test set.

8.4.4 PV Power Generation

Another source of uncertainty considering the proposed balancing service originates from the PV plant. PV yield prediction can be even more challenging on a snowy day, not only due to the errors in the day-ahead predictions of the meteorological parameters but also because of the complexity of obtaining the exact values of the power generation for snow-covered PV panels. Snow loss can be very large and variable depending on several factors such as the snow depth, the type of snow, the shape of the snow cover, etc. Therefore, a computational intelligence-based snow loss-aware PV power generation predictor is built that allows the aggregator to consider all probable scenarios for the next snowy day.

Historical data of the hourly PV yield and the meteorological parameters are extracted for a PV system in New York [127]. A snowy winter day (17 Feb. 2015) is chosen as the operating day. Considering the periodic patterns in the power generation of a PV system over years and due to the significant dependence of the PV power on weather conditions and meteorological parameters, the PV power prediction model in this study is built based on historical data containing data points in the most similar weather conditions to the operating day. To this end, the data of 30 days before the operating date, and 30 days before and 30 days after the similar date in the past year (17 Feb. 2014) and the past 2 years (17 Feb. 2013) are selected to constitute the dataset, which covers measurements over 5 months. This approach to building the dataset has been validated after training models based on datasets with different sizes and comparing their performance. Since the data is gathered over three consecutive years, the annual trend, such as the effect of increasing system power losses, is also maintained. The attribute and target variables are:

- Att. 1: Hours of the day (1 to 24);
- Att. 2: Hourly ambient temperature;
- Att. 3: Hourly global horizontal irradiance;
- Att. 4: Daily snowfall;
- Att. 5: Daily snow depth;
- Target: Hourly PV power generation.

The line graphs of the electrical and meteorological parameters in the dataset are shown in Fig. 8.7. The last 24 data points (the last day) are used as the test set. 720 data

points (30 days) before the operating day are used as the validation set. The first 2,928 data points (122 days) are used as the training set. The LSTM network together with the Bayesian optimization is implemented to build the optimal model. Then, the model is used to make a prediction on the unseen data of the operating day in the test set. The mean squared errors together with the optimal hyperparameters of the model are presented in Table 8.3. The line graphs of the real and predicted normalized values of the target variable on the validation set and the test set are shown in Figs. 8.8 (a) and 8.8 (b), respectively. The MSE value on the test set is 0.00055.

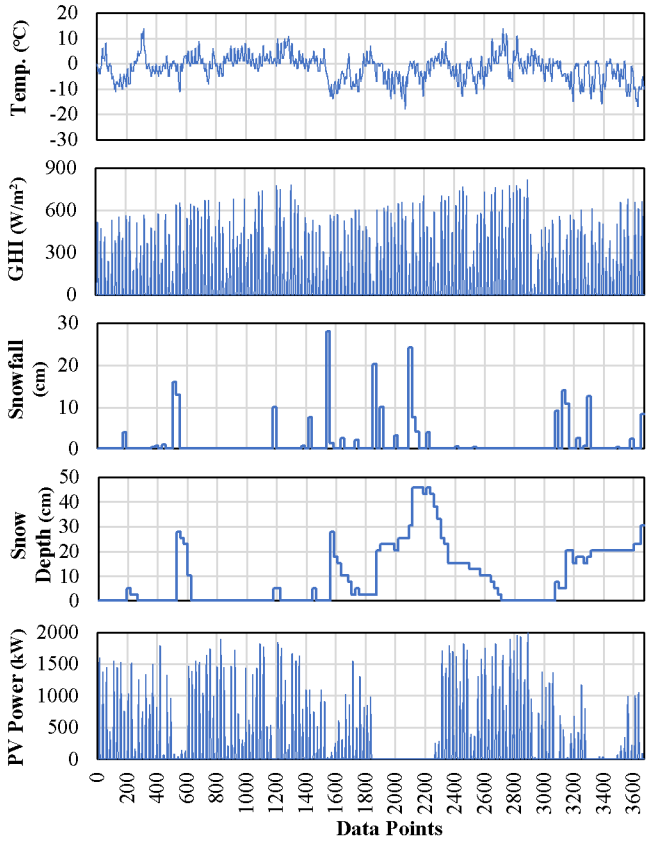


Figure 8.7. Line graphs of the electrical and meteorological parameters in the historical dataset of the PV yield predictor.

Table 8.3**.** Parameters of the PV yield predictor

Model	Validation MSE	Test MSE	Num. of Hidden Units	Initial Learning Rate	Learning Rate Drop Period	Learning Rate Drop Factor	Maximum Epochs	Mini-Batch Size
PV Predictor (LSTM Net 2)	0.00409	0.00055	87	0.1	100	0.5	75	48

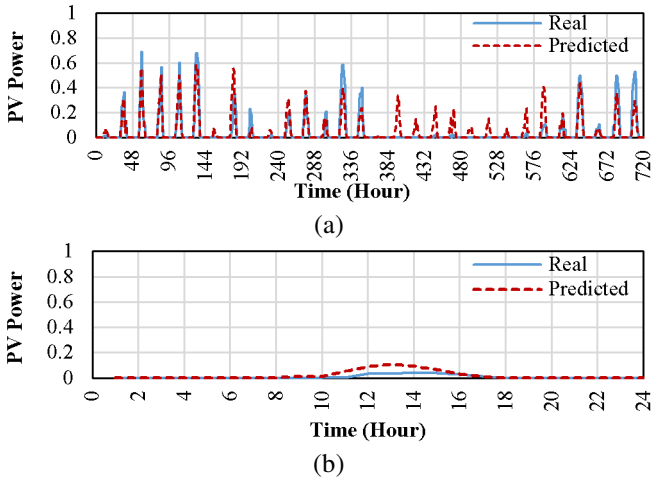


Figure 8.8. The real and predicted values of the hourly PV power generation using the LSTM model on (a) the validation set and (b) the test set.

8.5 Numerical Results

The topology of the 20.8 kV distribution grid, where PEVs are distributed, is adapted from the IEEE 34-node test system [142]. The main feeder is connected to the upstream grid by a 5.5 MW transformer. Three business PCSs (connected to nodes 5, 10, and 19) and two shopping PCSs (connected to nodes 14 and 18) are considered in the grid. Personal residential chargers are connected to the other nodes (except nodes 1 and 24). A 2.3 MW PV plant owned by the DSO is connected to node 24.

In order to generate the scenarios of the hourly PV power generation and the hourly load demand of the grid, 1000 scenarios of each uncertain attribute variable including temperature, GHI, snowfall, and snow depth are generated by adding a random value from a normal distribution to the real recorded values on the operating day. It is assumed that the predicted values of temperature are more accurate for the first hours of the day (± 1°C) compared to the last hours (± 4°C). Moreover, the variations of the hourly GHI at noon (± 200 W/m^2) are assumed to be bigger than the variations after sunrise or before sunset (± 50 W/m^2). It is also assumed that the values of daily snowfall may change between 0 and 15 cm. The snow depth can be simply calculated based on the snowfall and the snow depth on the previous day. The real values together with 1000 generated scenarios of these parameters are shown in Fig. 8.9. These scenarios are fed into the models and 1000 scenarios of the hourly PV power generation and the hourly load demand are generated. Then, a scenario reduction procedure based on Kantorovich distance is implemented to reduce the number of scenarios to lower the computational burden. Finally, the 5 most probable scenarios are selected and shown in Fig. 8.10.

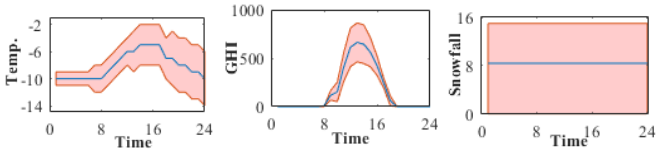


Figure 8.9. Real values (blue line) and 1,000 generated scenarios of (a) hourly temperature, (b) hourly GHI, and (c) daily snowfall.

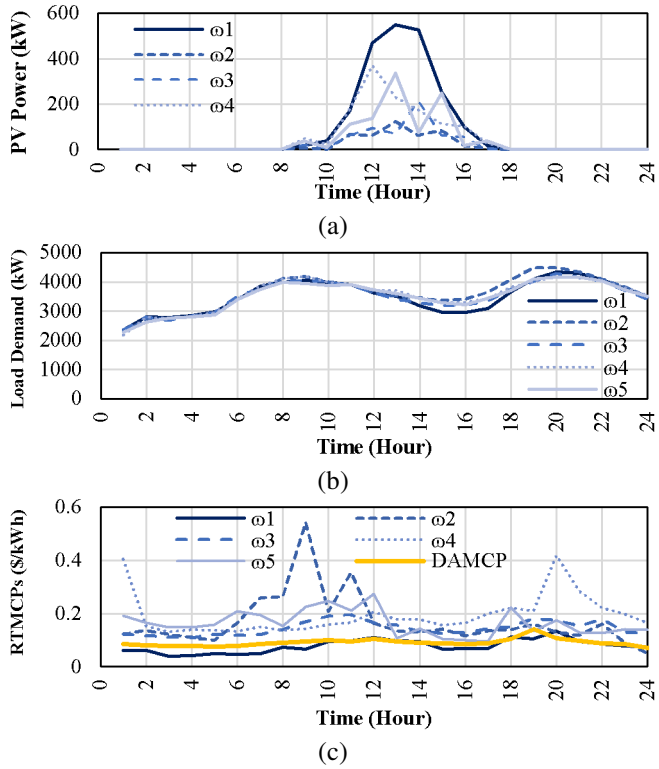


Figure 8.10. Scenarios of (a) PV power, (b) load demand, and (c) RTMCPs.

Eight different types of PEVs are considered [143]. Table 8.4 presents the properties of the PEVs. The penetration level of PEVs is defined as the ratio of the number of consumers with a PEV to the total number of consumers in the grid. Considering the residential load, the coincidence factor, and the average load peak of a consumer, the number of PEVs at the penetration level of 40% is 1000. The Roulette Wheel Mechanism is used to locate the PEV owners' houses randomly among the residential nodes. Moreover, 70 vehicles are randomly selected to be charged in each PCSs.

Finally, 25 final scenarios are formed by combining each RTMCPs scenario with five scenarios of the other three uncertain input variables. The proposed MILP problem is

solved using the CPLEX solver in GAMS software. Five cases are studied to prove the effectiveness of the proposed approach.

Table 8.4. Properties of the PEVs

PEV Type Number	1	2	3	4	5	6	7	8
Battery Capacity (kWh)	100	80	60	40	20	15	10	5
All Electric Range (km)	490	450	320	220	110	75	45	23
Market Share (%)	13	58	6	5	4	2	9	3
Charger Rate (kW)	11.5	11.5	7.2	6.6	6.6	6.6	3.3	3.3

8.5.1 Cases 1 and 2: Uncoordinated and Off-Peak Charging

The PEV owners plug in their vehicles as soon as they arrive home. In the uncoordinated charging mode, charging starts immediately with the maximum charging rate until the battery is fully charged (charging up to 90% of SoC is assumed, regarding the linear section of the charging curve, if the PEV is plugged in long enough). Time-of-Use (TOU) tariffs of the DSO are used to calculate the charging costs. In the off-peak charging mode, the charging process is delayed until midnight since the lower step of the TOU tariffs starts at 12 a.m. The charging bills also include a fixed monthly fee for a separate meter. In the winter of 2015, electricity consumers in New York paid a monthly fee of 19.87 $ and 1.36 cents (1:00 to 7:00)-7.13 cents (8:00 to 24:00) for every kWh of energy consumed [140].

The expected hourly charging load of all PEVs together with the expected hourly load demand of the grid, the expected hourly voltage magnitude of the last node (node 34), and the arrival and departure SoCs of the vehicles are shown in Fig. 8.11. In case 1, the PEVs' charging load peaks at 8 p.m. when most vehicles are coming back home and it has a gradual increase. This peak in case 2 happens at 1 a.m. and is more than three times bigger than that in case 1. As a result, the voltage drop at node 34 in case 2 is more severe compared to case 1; however, its duration is longer in case 1. Hence, a voltage drop across the last nodes of the feeder and the overload of the main feeder transformer at peak hours are undesirable consequences of the uncoordinated and off-peak charging of a large number of PEVs. The extra load incurred by the business and shopping PCSs to the grid during the day is not significant since the vehicles are mainly getting fully charged at home. Moreover, the majority of the PEVs have been charged to the predefined 90% of SoC. The expected daily charging costs of all PEVs in cases 1 and 2 are 1,301 $ and 867 $, respectively. Thus, off-peak charging is more beneficial for PEV owners since a major

part of the charging process occurs at the lowest level of the TOU tariff; however, the grid experiences worse technical challenges in this case.

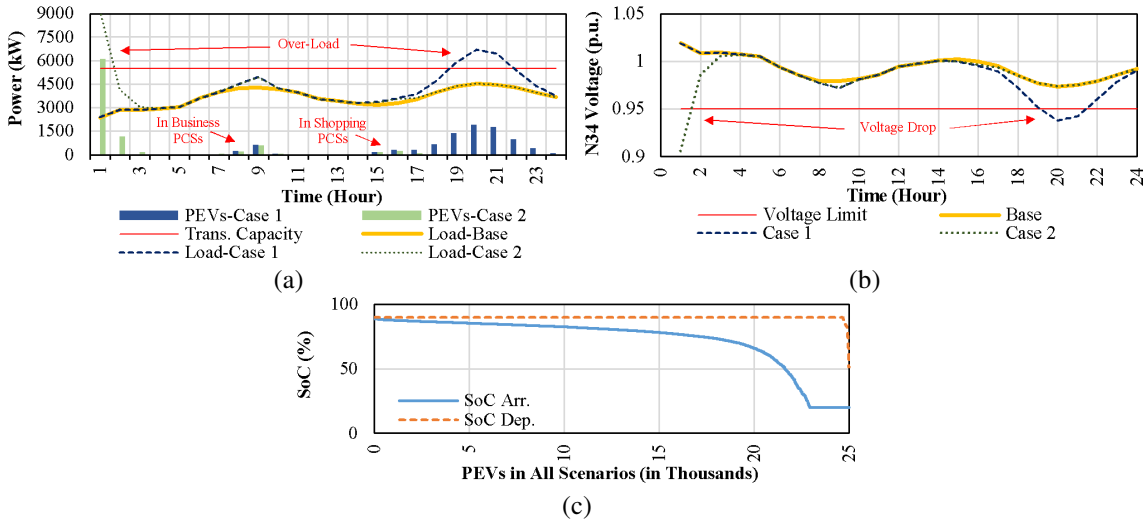


Figure 8.11. Case 1 and case 2 results: (a) Expected charging and main feeder load, (b) expected voltage magnitude of node 34, and (c) arrival and departure battery SoCs.

8.5.2 Case 3: Coordinated Charging (the basic model)

In this section, the developed stochastic programming approach (without the balancing service) is investigated. The results have been obtained considering the expected price of brand-new batteries in 2023 [144] and are shown in Fig. 8.12. The major part of the charging load has been shifted to the off-peak hours when the price of energy in the market is lower. This shows the significant effect of the DAMCPs on the established schedule of the aggregator for the next operating day. Unlike cases 1 and 2, the resulting load shift avoids both main feeder overloads and voltage drops lower than 0.95 p.u. across the grid (Figs. 8.12(a) and 8.12(b)) by distributing the charging load of the vehicles over six hours, obtained by the aggregator's commitment to preventing charging surges. Moreover, PEVs have been discharged at 7 p.m. when DAMCP is higher. This reduces the overall charging cost of the vehicles by contributing to load serving at peak hours. Considering the highly uncertain RTMCPs, it is clear in Fig. 8.12 (c) that the aggregator tends to provide the major part of the PEVs' energy demand from DAM to guarantee a minimum expected charging cost for the next operating day. As can be seen in Fig. 8.12 (d), up to 530 kWh of the daily charging demand has been purchased from the RTM at hours with lower RTMCPs in each scenario while in scenarios 1, 6, 11, 16, and 21, almost all the charging demand of the PEVs has been purchased from DAM. The difference in the RTM purchase values for every five consecutive scenarios in Fig.

8.12 (d) originates from the difference in the driving patterns and consequently the charging requirements.

As indicated in Fig. 8.12 (e), the expected daily charging cost of all PEVs is 610 $ (53% lower than case 1 and 30% lower than case 2), ranging from 663 $ in scenario 19 to 584 $ in scenario 21. The aggregator's cost in DAM (purchases minus sells) and its expected cost in RTM are 497 $ and 28 $, respectively. The expected degradation cost of all batteries is 85 $. The aggregated discharging power in Fig. 8.12 (a) shows that the vehicle-to-grid is affordable only when the price of energy is high enough to compensate for the battery degradation cost and charging/discharging power losses. So, batteries' technology improvement and price reduction can play a key role in making V2G services more affordable and popular. In general, the coordinated charging approach can reduce the charging cost of PEVs compared to uncoordinated and off-peak charging and at the same time, guarantee the normal operation of the grid.

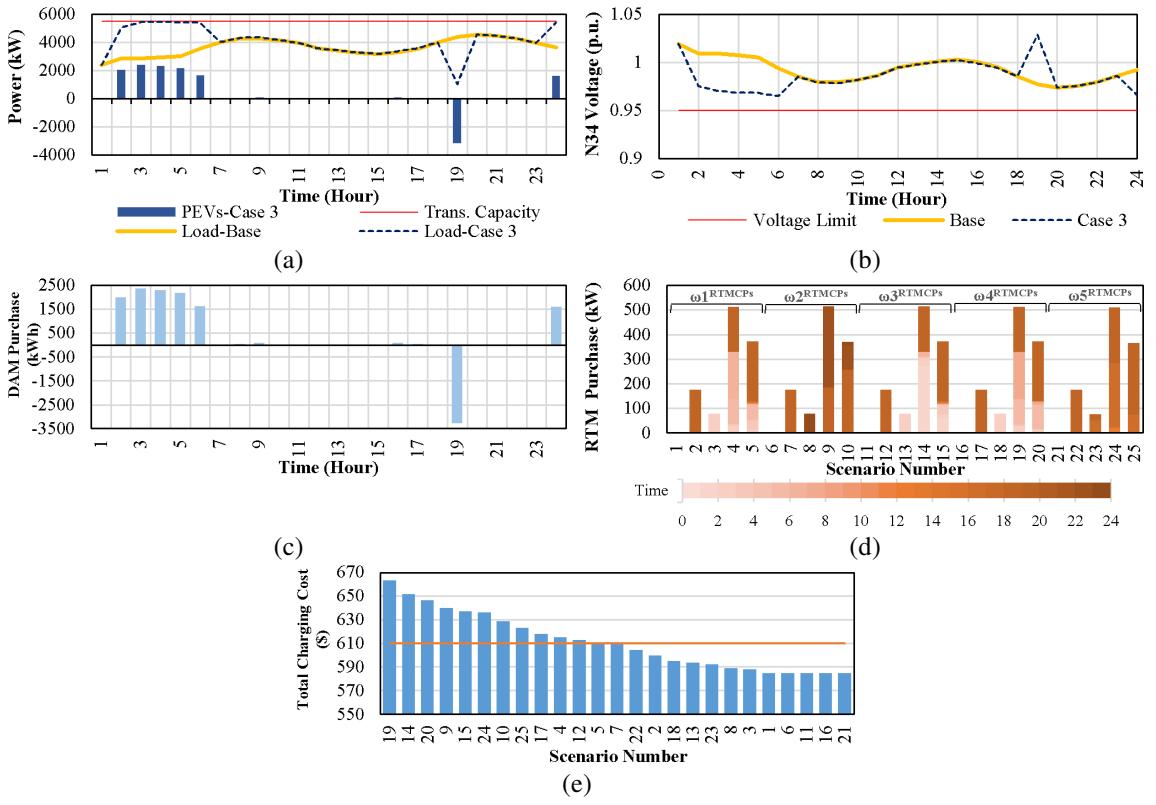


Figure 8.12. Case 3 results: (a) Expected charging power and main feeder load, (b) expected voltage magnitude of node 34, (c) aggregator's DAM energy purchases, (d) aggregator's RTM energy purchases, and (e) total charging cost in each scenario.

8.5.3 Case 4: Coordinated Charging with Balancing Service

The aggregator's day-ahead scheduling with the proposed balancing service is investigated in case 4. A pricing mechanism is used to calculate $Tariff^{Balancing}$ in equation (8.5) as a multiple of RTMCP unless RTMCP is higher than a particular threshold. This prevents the reflection of the surges and large values of RTMCPs in $Tariff^{Balancing}$. This threshold can be considered as a multiple of DAMCP and used as a cap for $Tariff^{Balancing}$. Moreover, due to the highly downward trend in the prices of brand-new batteries, two cases of ½ battery prices (0.5*BP) and without battery degradation cost (0*BP) are also investigated.

The DSO's hourly balancing service requests for the first 5 scenarios are shown in Fig. 8.13 (a) which reflects the effect of the combination of 5 PV yield scenarios and 5 grid load demand scenarios. The peak and the largest variability of the requests happened at noon when the PV plant may not be able to supply the grid as expected in the snow conditions. The expected value of the total charging cost (TCC) reduction for the PEVs together with the DSO's real-time cost (RTC), as the sum of the costs of the balancing service and the purchase of the remaining amount of the demand from the RTM, reduction thanks to the balancing service compared to case 3 are shown for three cases of battery prices in Fig. 8.13 (b). A grid search has been performed to find the best $Tariff^{Balancing}$ to maximize the expected value of the aforementioned reductions. The maximum expected DSO's RTC reduction can be achieved by $Tariff^{Balancing}$ = 1.1×RTMCPs and 1.9×DAMCPs as a cap and the maximum expected TCC reduction can be achieved by $Tariff^{Balancing}$ = 0.8×RTMCPs and 1.3×DAMCPs as a cap. As can be seen in Fig. 8.13 (b), decreasing the battery prices can increase the benefit obtained by implementing the balancing service where a 14% reduction in the DSO's RTC and a 12% reduction in the PEVs' TCC have been obtained. This is achieved by a higher contribution of the PEVs to support the balancing requests, from an average of 70% in the case of 1*BP to an average of 85% in the case of 0*BP, which is indicated in Fig. 8.13 (c). Smaller DSO's balancing requests have resulted in almost complete support of the balancing service by the PEVs in scenarios 1, 6, 11, 16, and 21. The percentage of TCC and DSO's RTC reductions in the case of 0*BP for each scenario together with their expected values (horizontal dash lines) are shown in Fig. 8.13 (d). The maximum and the minimum reductions have been obtained in scenarios 21-25 and 11-15 using $Tariff^{Balancing}$ based on the fifth and the third scenarios of RTMCPs, respectively. As can be seen in the

results of this case, implementing the proposed local balancing service can be beneficial for both the DSO and the PEVs aggregator. In fact, this type of grid-support service can enable small-scale aggregators to contribute to facilitating the higher penetrations of intermittent renewable energies in local power grids. Similar to case 3, the normal operation of the grid has been obtained based on the aggregator's schedule by avoiding the charging surges over the operating day.

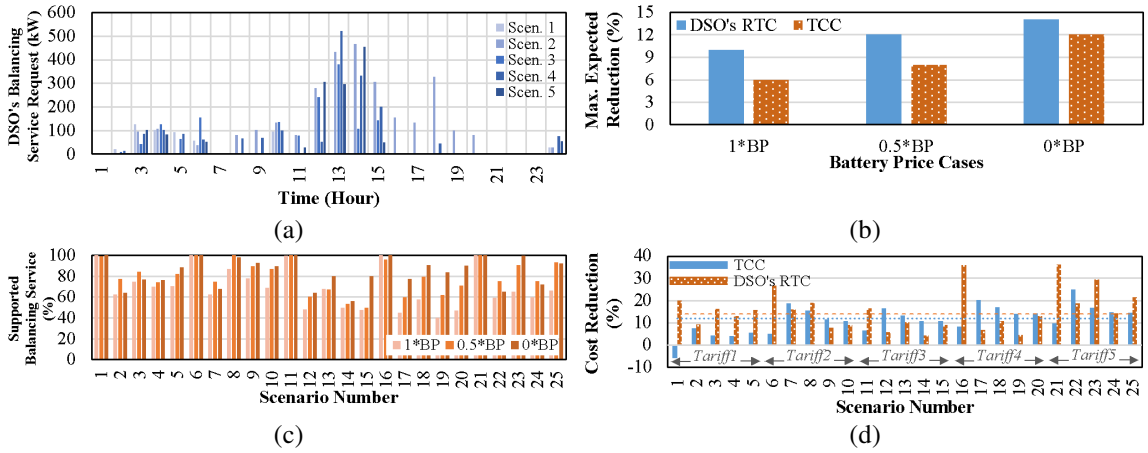


Figure 8.13. Case 4 results: (a) DSO's balancing service requests, (b) maximum expected TCC and DSO's RTC reductions, (c) supported balancing service, and (d) TCC and DSO's RTC reductions in the case of ignoring the battery degradation cost.

8.5.4 Case 5: Risk-Averse Coordinated Charging with Balancing Service

In this case, the results of the risk-averse formulation based on equations (8.24)-(8.26) are investigated. To this end, the battery degradation cost is fully considered (battery price case of 1*BP), $Tariff^{Balancing}$ is set to have the most expected TCC reduction, and the confidence level (γ), which categorizes high-cost and low-cost scenarios, is set as 0.9. The cumulative distribution functions for three cases of the weighting parameter φ, which materializes the trade-off between the expected cost and the risk aversion, are illustrated in Fig. 8.14 (a).

As can be seen, the expected TCC and CVaR, which denotes the expected cost of the worst scenarios, in the risk-neutral formulation (with φ=0) are 578 $ and 625$, respectively. As expected, considering the risk measure and increasing its weight increases the expected cost and reduces the CVaR; so that, with φ=1, they are very close to each other (expected cost=616 $ and CVaR=617 $). So, the risk-averse formulation has reduced the cost of encountering the worse scenarios at the expense of increasing the expected TCC. As can be seen in Fig. 8.14 (b) for risk-averse results with φ=1 and γ=0.9,

the PEVs have less supported the balancing service requests in most of the scenarios after considering the risk in the optimization problem. A sensitivity analysis is also performed for different values of the confidence level (γ) from 0.5 to 0.99 with φ=1. The results are shown in Fig. 8.14 (c). The confidence level plays an important role in capturing the adverse effect of the worst scenarios. By increasing the confidence level, the expected total charging cost will rely on a fewer number of high-cost scenarios ((1- γ)×100%). So, the expected cost and the CVaR both have increased by raising the confidence level.

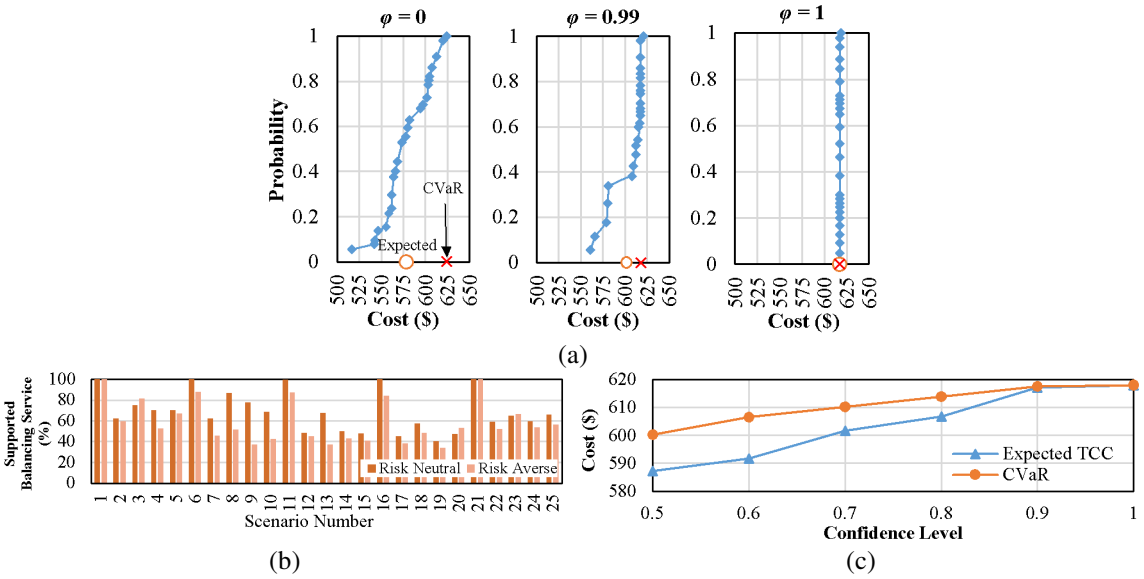


Figure 8.14. Case 5 results: (a) Cumulative distribution function for risk-neutral and risk-averse approaches, (b) supported balancing service with risk-neutral and risk-averse formulations, and (c) the effect of confidence level on CVaR and expected TCC.

In general, the results prove the effectiveness of the proposed computational intelligence-based scheduling approach in reducing the charging cost of the PEVs and maintaining the normal operation of the grid compared to the un- or semi-coordinated approaches. Moreover, the integrated out-of-market balancing service is proved to be a useful tool for reducing the challenges that a grid operator deals with in high penetrations of intermittent renewable energies and to be beneficial for the aggregator/PEV owners. To clarify the differences between the case studies, a brief benchmarking is performed in Table 8.5, in which the levels of achievements are ranked from low (*) to high (***). As can be observed, the developed stochastic programming approach with the proposed balancing service (cases 4 and 5) is the most desirable from different aspects.

Table 8.5. Benchmark for comparing the case studies

	Case 1	Case 2	Case 3	Case 4	Case 5
PEV Owners' Satisfaction (E. Range)	***	***	***	***	***
PEVs' Charging Costs Reduction	-	*	**	***	***
Guarantee Grid Reliability (Overload)	-	-	***	***	***
DSO's Operation Cost Reduction	-	*	**	***	***
Risk Aversion Scheduling	-	-	-	-	***

8.6 Conclusion

This chapter addresses the issue of day-ahead scheduling of a PEVs aggregator that participates in both day-ahead and real-time energy markets on behalf of the PEV owners through the local DSO. The aggregator can also provide the DSO's extra energy demand in real-time, which originates from the differences between the DSO's day-ahead bids and the actual load demand of the renewable power penetrated distribution system, through a proposed out-of-market balancing service, and reduce the charging costs of the vehicles. A two-stage stochastic approach with a comprehensive availability model of PEVs for both residential night-charging and public intraday-charging has been developed that minimizes the total charging cost of the PEVs. Computational intelligence methods have been used to model the uncertain input variables on a snowy day and generate the scenarios. The results show that the proposed day-ahead scheduling approach can reduce the expected total charging cost by 53% and 30% compared to the uncoordinated and off-peak charging modes, respectively. It also guarantees the normal operation of the grid. Moreover, the proposed balancing service can be beneficial for both the DSO and the PEV owners. Depending on the price of the service and brand-new batteries, it can reduce up to 25% of the total charging cost and 36% of the DSO's real-time cost while the expected reductions are 12% and 14% compared to the same condition without the service, respectively. Moreover, the risk-averse formulation can reduce the cost of encountering the worst scenarios while it increases the expected charging cost significantly.

CHAPTER 9

COMPUTATIONAL INTELLIGENCE-BASED FORECASTING OF PV POWER IN SNOW CONDITIONS AND ITS APPLICATION IN A MICROGRID ENERGY MANAGEMENT SYSTEM

9.1 Introduction

The energy management system of a microgrid is a system mainly responsible for scheduling and power sharing between distributed energy resources and, therefore, comprises secondary and tertiary control levels. A model predictive control-based EMS can adequately optimize the operation of the microgrid by incorporating a feedback mechanism, which allows the system to face uncertainty and disturbances, handle physical constraints, such as storage capacity, incorporate generation and demand forecasts, and consider future behavior of the system, which is of crucial importance for systems that depend on demand and renewable energy generation. The methodology developed to achieve the second part of the third objective is described in this chapter.

9.2 Framework of the Problem

This study implements a snow conditions-compatible MPC-based EMS for a microgrid in a snow-prone region. The commercial microgrid is assumed to be grid-connected, have a relatively large level of PV energy penetration, and be equipped with a battery storage system, as shown in Fig. 9.1. The EMS receives the measurements of the electrical and meteorological parameters in real-time and generates the optimal control signal for the battery charge controller by minimizing a cost function based on the system model. The nature of the MPC is based on the selection of the best input sequences that results in the minimum cost over a future time horizon constrained to some criteria. The receding horizon principle shapes the basis of the MPC so that the first input of the best-calculated sequence is applied to the system and the procedure is repeated when the state information is renewed at the next time step [145].

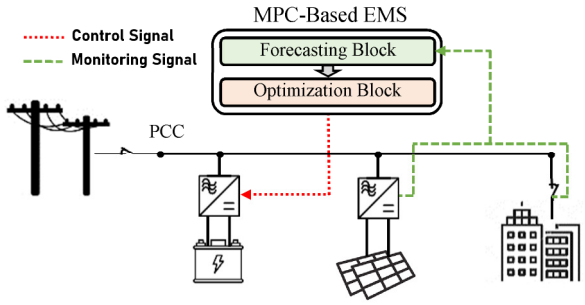


Figure 9.1 The main framework of the MPC-based EMS of the microgrid

The MPC is formed by three steps [146]. In the first step, the dynamic model of the system is used to predict the outputs of the system over the prediction horizon considering the predicted values for the uncertain input variables. In the second step, an optimization problem with the aim of keeping the output of the system as close as possible to the reference trajectory is solved to generate the optimal sequence of the control signals over the control horizon. In the third step, the first signal of the control sequence is sent to the system while the next control signals are discarded. After each time step, these three steps are repeated and the control signals are brought up to date. The new control signal may be different from the previously calculated one since it is updated based on the new information becoming available.

In this study, a two-stage MPC is implemented which corresponds to the tertiary and secondary controllers of the microgrid. The tertiary controller is in charge of the economical optimization of the microgrid. This controller receives the forecasts of the load demand and PV power generation from the snow condition-compatible forecasting block and outputs the optimal dispatch schedules for the energy exchange with the main grid and the battery charge/discharge rates for all the instants belonging to the schedule horizon. These schedules are then sent to the secondary controller. The secondary controller has an intermediate position between the tertiary controller, which generates the day-ahead schedules, and the primary controller, which operates autonomously with a fast response on the electronic power converters. The secondary controller adjusts the set points sent to the primary controllers according to the optimal schedules and the real-time situation. This control layer uses the measurements of the PV power generation and the load demand of the microgrid to obtain the net power of the microgrid. So, there is no need to have an energy forecast model since the energy values are assumed to be constant during the prediction horizon.

9.3 Forecasting Block of the EMS

Considering the two sources of uncertainty, including the PV power generation and load demand of the microgrid, the forecasting block of the EMS consists of two models. These computational intelligence-based forecasting models are designed to provide high accuracy, especially in snow conditions. All the models are developed in Python using Scikit-Learn and TensorFlow libraries.

9.3.1 Load Demand Forecasting

A short-term load demand forecasting approach is developed which provides 24 hours ahead hourly forecasts of the microgrid's electric load. Considering the periodicity and predictability of the electricity consumers' behavior, short-term electric load forecasting is a relatively straightforward task. In this section, a load forecasting model based on an autoregressive long short-term memory network [124], with the schematic shown in Fig. 9.2, is developed. The model generates the forecasts in one-hour steps and uses each forecast as input for generating the next forecast. In this way, the interdependency of the consecutive hourly forecasts as a time series is considered and the forecasting error is decreased.

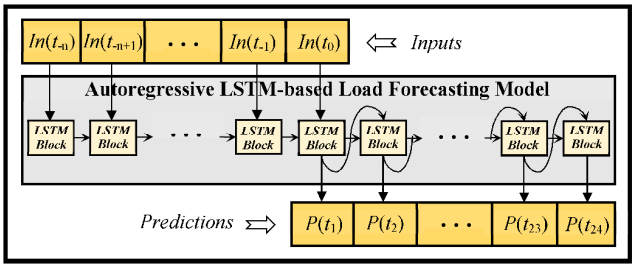


Figure 9.2. The schematic of the developed load demand forecasting model.

In order to develop the load forecasting approach, the historical load data of the commercial section of the city of Dayton, OH, together with the local data of ambient temperature, which has a considerable correlation with the load demand, for a period of almost 7.5 years, from 14/07/2012 to 02/12/2019, is extracted from the website of the Applied Energy Services Ohio [141] and the National Renewable Energy Laboratory [136]. The data is cleaned by removing the outliers and the missing data points are imputed using the k-nearest neighbor method [120]. Then, the equivalent date-time, weekday numbers, and holidays are added to the dataset. The daily and yearly periodicity is extracted and substituted for the date-time values using sine and cosine transforms. The

holidays are substituted by one binary label and the weekday numbers are substituted by seven binary labels generated based on one-hot encoding. The dataset is split into 70% train-15% validation-15% test sets so that the test set remains unseen when building the model. Then, the data is normalized based on the training set. In order to use the past data as the input of the model and generate the forecasts, the sliding window method is utilized to make windows with an input width of 7×24, to cover a whole week pattern, including all the aforementioned parameters, and a target width of 24, including the load demand. The optimal hyperparameters of the model are obtained by hyperparameter tuning based on Bayesian optimization and evaluating the resulted mean squared error on the validation set. The input width of the sliding windows is also considered a hyperparameter. After finding the best model, it is trained over the full train-validation set and tested on the test set. The performance scores of the main model on the test set together with that of several benchmark models are presented in Table 9.1. The models listed in the table are described as follows: model (1): uses load values of seven days (7×24 hours) ago as the forecasts for the next 24 hours, model (2): 24 linear regression models with the current values of the attribute variables as inputs, model (3): a feed-forward artificial neural network (ANN) model with the current values of the attribute variables as inputs, model (4): a feed-forward ANN model with rolling window-based inputs of the past target values, model (5): a 1-d convolutional neural network (CNN) model with rolling window-based inputs, model (6): an LSTM layer followed by a feedforward output layer with rolling window-based inputs, model (7): 24 LSTM models with rolling window-based inputs, each for forecasting a specific hour in the next 24 hours, and model (8): the proposed autoregressive LSTM-based model which is the best one identified among the aforementioned models.

As can be seen in the table, the autoregressive LSTM model has the best performance with the smallest forecasting error (MSE of 0.0025). The optimal values of the input width of the sliding windows, the number of LSTM units, and the learning rate of the developed autoregressive LSTM model are 96, 240, and 0.001, respectively.

Table 9.1 Performance scores for the proposed and benchmark load forecasting models

	Models	Scores Over Test Set			Optimal Hyperparameters
		MSE	MAE	RMSE	
1	Last-7 Days Repeat Model	0.0070	0.0557	0.0834	-
2	Single-Shot Linear Model	0.0103	0.0793	0.1016	-
3	Single-Shot ANN Model	0.0037	0.0446	0.0609	Hidden Layers=1, Hidden Units=120, Learning Rate=0.001, Activation=ReLU
4	Multi-Shot ANN Model	0.0056	0.0562	0.0750	Hidden Layers=1, Hidden Units=336, Learning Rate=0.001, Activation=ReLU
5	Convolutional Neural Network	0.0039	0.0460	0.0621	Window Input Width=12, Filters=228, Learning Rate=0.001, Activation=Sigmoid
6	Long Short-Term Memory Network	0.0032	0.0415	0.0566	Window Input Width=3, LSTM Units=240, Learning Rate=0.01
7	24 LSTM Network-based Model	0.0033	0.0412	0.0570	Different Values for Each Model
8	**Autoregressive LSTM Network**	**0.0025**	**0.0355**	**0.0501**	Window Input Width=96, LSTM Units=240, Learning Rate=0.001

9.3.2 PV Power Forecasting

PV power forecasting is considered a more challenging task compared to load forecasting since it gets highly affected by cloud movements and uncertain weather conditions such as snowfall events in the hours ahead. A novel short-term PVPF approach is proposed in this section which considers the effect of snow layer formation on the panels. The schematic of the proposed approach is shown in Fig. 9.3. This approach receives the on-site measurements of the meteorological and electrical parameters, and generates the PV power forecasts over the next 24 hours with a temporal resolution of 1 hour. Snow-cover detection, snow-cover forecasting, base PVPF, and snow-cover conditions PVPF constitute the proposed approach.

In order to develop the PVPF approach, the historical electrical data of a PV system in Denver, CO, together with the local weather data for a period of almost 7.5 years, from 14/07/2012 to 02/12/2019, is extracted from the website of the National Renewable Energy Laboratory [127] and the National Ocean and Atmospheric Administration [128]. The data is cleaned by removing the outliers and the missing data points are imputed using the k-nearest neighbor method. The features include the AC power of the PV system, ambient temperature, PV module temperature, plane-of-array irradiance, snowfall, snow depth on the ground, wind speed, wind direction, humidity, air pressure, and date time. All the features, except snowfall and snow depth with daily values, have a temporal resolution of one hour. Wind speed and wind direction are converted into a wind vector, which is easier for the models to interpret, by calculating its x and y

components based on the magnitude and direction of the wind. The daily and yearly periodicity is extracted and substituted for the date-time values using sine and cosine transforms. The daily values of snowfall and snow depth are repeated over the hours of a day.

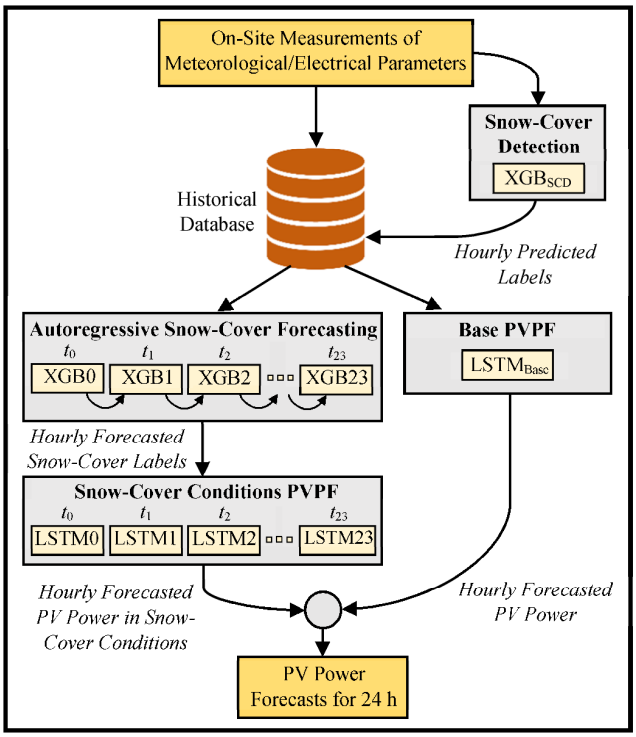


Figure 9.3. The schematic of the proposed PV power forecasting approach.

9.3.2.1 Snow-Cover Detection Approach

The first step of developing the proposed PVPF model is to detect the hours when there is a partial/full snow cover on the panels because the presence of a snow layer is not detected and recorded most of the time for PV systems. This step is of great importance since in regions with harsh winters, the PV power generation can be highly disrupted by snow. First, the expected PV power is calculated using the PVWatts DC model [126] by considering the degradation and temperature losses as follows:

$$P_t^{PV\ Exp} = P^{STC} \times \alpha_t^{Quality\ Deg} \times \frac{G_t^{POA}}{G^{STC}} \times (1 - \alpha^{Temp} \times (Temp^{STC} - Temp_t^{Module})) \quad (9.1)$$

where $P^{PV\ EXP}$ is the expected PV power, P^{STC} is the PV power in standard test conditions with solar irradiance (G^{STC}) and module temperature ($Temp^{STC}$) of 1000 W/m^2 and 25°C,

respectively. G^{POA} is the solar irradiance reaching the surface of the modules with a temperature of $Temp^{Module}$. $\alpha^{QualityDeg}$ and α^{Temp} are the quality degradation and temperature coefficients, respectively. Then, the rest of the PV power loss, which includes snow loss, is calculated by subtracting the expected power from the actual PV power. By considering these values together with the values of snowfall, snow depth, temperature, and POA irradiance, the binary snow cover labels are manually generated and added to the dataset. The labels are one when a solar irradiance blockage by a probable partial/full snow cover increases the power loss of the PV system and zero otherwise. After labeling the dataset, a snow-cover prediction model based on extreme gradient boosting (XGBoost) [147] is developed that receives the current values of the AC power and the main meteorological parameters and detects the presence of a snow layer on the panels. The XGBoost model is identified as the best snow-cover predictor after developing and comparing several models based on other computational intelligence techniques. In order to develop the model, the dataset is split into 85% train-15% test sets so that the test set remains unseen even when developing the other models. Then, the data is normalized. The snow-cover labels when POA irradiance is lower than 20 W/m^2 are considered as zero since, considering that PV power is close to zero at this range of irradiance, it improves the performance of the snow-cover detection model. The positivity rates of the labels in the train and test sets are 2.2% and 3.3%, respectively.

Hyperparameter tuning based on 3-fold cross-validation and Bayesian optimization is implemented to find the best XGBoost-based snow-cover detection model. Then, the model is tested on the test set. The optimal learning rate, maximum depth, and the number of estimators of the model are 0.305, 9, and 100, respectively. The confusion matrices for the train and test sets considering 0.5 as the threshold are shown in Fig. 9.4. As can be seen, the model detects all the labels in the train set correctly while in the case of the test set, there are some falsely detected labels. The accuracy, precision, and recall scores of the model on the test set are 0.9965, 0.9457, and 0.9457, respectively. Using the developed snow-cover detection model, the snow-cover labels are generated in real-time and added to the dataset.

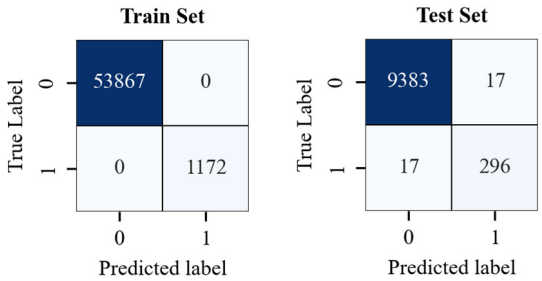


Figure 9.4. Confusion matrices for the snow-cover prediction model.

9.3.2.2 Snow-Cover Forecasting Approach

PV power forecasting can be very challenging when the power generation is probable to get disrupted by the presence of a snow cover on the panels in the next few hours. Hence, forecasting the snow-cover conditions should help to improve the PVPF performance for PV systems in snow-prone regions. This can be done by feeding the snow-cover forecasts over the forecasting horizon into the PVPF approach and effectively making corrections to the next 24 hours' power forecasts. As can be seen in Fig. 9.3, an autoregressive snow-cover forecasting approach based on XGBoost is proposed where 24 XGBoost models are developed to provide forecasts of the snow-cover conditions for the next 24 hours. Each model presents the forecast for a specific hour in the future. The models not only receive the current values of the PV power, the meteorological parameters, and the snow-cover label as their inputs but also the snow-cover forecast of the model built for the previous hour of the forecasting horizon to improve the accuracy. In order to develop the models, the dataset is split into 85%-15% train-test sets and standardized. Hyperparameter tuning based on 3-fold cross-validation with Bayesian optimization is implemented to find the optimal hyperparameters. The optimal hyperparameters together with the performance score of each model are presented in Table 9.2. As can be observed, the best performance scores are obtained by the first model which forecasts the snow-cover label for the first hour of the forecasting horizon. The performance scores of the other models decrease gradually; so that, the drop and fluctuation of the performance scores increase among the models built for the last hours.

The overall accuracy, precision, and recall scores on the train set are 0.997, 0.948, and 0.937, respectively. The overall accuracy, precision, and recall scores on the test set are 0.992, 0.916, and 0.840, respectively. The confusion matrices for the train and test sets considering 0.5 as the threshold and flatting the vector of 24 forecasts are shown in Fig. 9.5. As can be seen, the proposed approach has good performance in forecasting the

snow-free conditions (when the label is zero), as it is the most dominant condition over the dataset. Considering this fact, it would be normal that the number of falsely negative (zero) forecasted labels is higher than the number of falsely positive (one) forecasted labels. Nevertheless, the proposed approach is able to forecast more than 93% and 84% of the snow-cover labels correctly in the train and test sets, respectively.

Table 9.2 Hyperparameters and performance scores of the snow-cover forecasting models

Model Num.	Optimal Hyperparameters			Performance Scores			
	Max. Depth	Learning Rate	Num. Estimators	Accuracy	Precision	Recall	AUC ROC
1	6	0.185	83	0.998	0.983	0.961	0.999
2	5	0.194	100	0.996	0.963	0.920	0.999
3	9	0.130	100	0.997	0.976	0.936	0.998
4	9	0.253	85	0.996	0.963	0.916	0.996
5	10	0.176	43	0.995	0.972	0.897	0.995
6	7	0.113	43	0.992	0.931	0.827	0.994
7	10	0.337	67	0.995	0.968	0.881	0.994
8	7	0.036	100	0.991	0.891	0.843	0.988
9	9	0.118	70	0.994	0.971	0.862	0.988
10	7	0.075	88	0.992	0.958	0.805	0.988
11	9	0.204	45	0.994	0.978	0.853	0.988
12	8	0.036	100	0.990	0.914	0.789	0.976
13	10	0.111	100	0.994	0.960	0.849	0.989
14	6	0.139	59	0.990	0.889	0.798	0.988
15	3	0.001	1	0.984	0.786	0.693	0.912
16	10	0.363	100	0.994	0.961	0.872	0.990
17	3	0.011	1	0.988	0.849	0.795	0.902
18	3	0.005	95	0.983	0.751	0.725	0.911
19	1	1.000	100	0.994	0.952	0.888	0.991
20	5	0.162	16	0.988	0.844	0.795	0.958
21	10	0.150	89	0.994	0.948	0.888	0.989
22	7	0.195	16	0.989	0.868	0.798	0.982
23	6	0.033	72	0.984	0.786	0.706	0.983
24	8	0.735	74	0.993	0.928	0.872	0.993

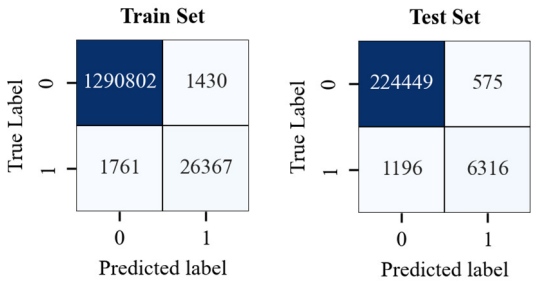


Figure 9.5. Confusion matrices for the snow-cover forecasting approach.

9.3.2.3 PV Power Forecasting Approach

First, a base PVPF model using LSTM networks is proposed to provide the base hourly PV power forecasts for the next 24 hours in all conditions. LSTM networks learn long-term dependencies in the data and are known as a promising computational intelligence technique to use when dealing with a sequence of data in the form of time series [124]. This model is trained over the full train set. Second, a snow-cover conditions PVPF approach is proposed to provide PV power forecasts specifically for the hours with a forecasted snow cover on the panels. In fact, the output of the snow-cover forecasting approach determines for which hours the base power forecast has to be replaced by the snow-cover conditions power forecast. The proposed snow-cover conditions PVPF approach consists of 24 models based on LSTM networks. Each model provides power forecasts for a specific hour in the future and is trained only over those data points with a positive snow-cover label.

In order to build the base and snow-cover conditions PVPF models, the data is split into 55%-30%-15% train-validation-test sets and normalized. Then, the sliding window method is utilized to generate the windows with an input width of 4×24, including all the meteorological parameters and PV power, and a target width of 24, including PV power. The optimal hyperparameters of all the LSTM models are obtained by hyperparameter tuning based on Bayesian optimization and evaluating the resulting MSEs on the validation sets. The input width of the sliding windows is also considered a hyperparameter. After finding the best model, it is trained over the full train-validation set and tested on the test set. The optimal hyperparameters together with the performance scores of the base and snow-cover conditions PVPF models on the test set are presented in Table 9.3. As can be seen, the smallest forecasting errors among the snow-cover conditions PVPF models, with MSE, MAE, and RMSE of 0.0145, 0.0658, and 0.1204, belong to the first model that forecasts the PV power in the next hour. The errors highly increase for the models that provide forecasts for the other hours ahead. This comes from the fact that making forecasts for the near future is easier than the far future. Moreover, the forecasting errors of the base PVPF model are considerably lower than those of the snow-cover conditions PVPF models, except for the first model. The reason is that the base model is trained and tested over the full data, the majority of which is acquired in snow-free conditions, while the other models are trained and tested only on the data in

snow-cover conditions. This shows that PVPF is more challenging when the panels are partially/fully covered by a snow layer.

Table 9.3 Hyperparameters and performance scores of the PVPF models

Model Num.		Optimal Hyperparameters			Performance Scores		
		Windows' Input Width	LSTM Units	Learning Rate	MSE	MAE	RMSE
Base		96	24	0.001	0.0145	0.0658	0.1204
Snow-Cover Conditions PVPF	1	12	101	0.01	0.0067	0.0538	0.0820
	2	12	31	0.01	0.0280	0.1136	0.1672
	3	12	31	0.01	0.0403	0.1314	0.2007
	4	12	31	0.01	0.0397	0.1344	0.1992
	5	24	211	0.01	0.0509	0.1602	0.2257
	6	48	111	0.001	0.0572	0.1658	0.2391
	7	96	61	0.01	0.0518	0.1601	0.2276
	8	96	171	0.001	0.0612	0.1660	0.2475
	9	24	181	0.001	0.0524	0.1593	0.2289
	10	96	1	0.01	0.0521	0.1628	0.2283
	11	96	1	0.01	0.0459	0.1476	0.2141
	12	48	111	0.001	0.0559	0.1537	0.2364
	13	24	1	0.01	0.0404	0.1414	0.2010
	14	96	231	0.001	0.0757	0.1814	0.2752
	15	12	31	0.01	0.0525	0.1625	0.2291
	16	96	1	0.001	0.0499	0.1592	0.2233
	17	3	1	0.01	0.0475	0.1461	0.2180
	18	48	111	0.001	0.0474	0.1490	0.2177
	19	48	111	0.001	0.0534	0.1604	0.2312
	20	72	1	0.01	0.0442	0.1364	0.2103
	21	24	1	0.01	0.0449	0.1391	0.2118
	22	24	1	0.001	0.0469	0.1553	0.2165
	23	3	1	0.001	0.0486	0.1568	0.2204
	24	3	111	0.001	0.0494	0.1721	0.2223

By combining the forecasts of the base and snow-cover conditions PVPF models, the performance scores on the test set are calculated and presented in Table 9.4. The first, second, and third rows of the table show the scores calculated on the full test set, the part of the test set with data points in snow-free conditions, and the part of the test set with data points in snow-cover conditions, respectively. As can be seen, the MSE of the combined PVPF approach is reduced by almost 42% (from 0.0815 to 0.0474) in snow-cover conditions compared to the base model.

Combining all the models, including the snow-cover detection model, snow-cover forecasting models, base PVPF model, and snow-cover conditions forecasting models, results in the proposed snow condition-compatible PVPF approach which improves power forecasting for PV systems in snow-prone regions. The performance scores of the

proposed approach are compared with several other PVPF approaches in Table 9.5. The benchmark models have the same definition as those of Table 9.1. As can be seen, the best overall performance is obtained using the proposed approach. It has the smallest MSE in snow-cover conditions and a very close forecasting error to the LSTM model in snow-free conditions.

Table 9.4 Overall performance scores of the base and combined PVPF approaches

Subsets of Test Set	Base PVPF			Base PVPF + Snow-Cover Conditions PVPF		
	MSE	MAE	RMSE	MSE	MAE	RMSE
Full	0.0145	0.0658	0.1204	0.0134	0.0633	0.1158
Snow-Free	0.0122	0.0604	0.1107	0.0122	0.0604	0.1107
Snow-Cover	0.0815	0.2265	0.2856	0.0474	0.1485	0.2178

Table 9.5 Performance scores for the proposed and benchmark PVPF approaches

	Models	MSE over Test Subsets			Optimal Hyperparameters
		Full	Snow-Free	Snow-Cover	
1	Last-Day Repeat Model	0.0282	0.0250	0.1238	-
2	Single-Shot Linear Model	0.0239	0.0225	0.0680	-
3	Single-Shot ANN Model	0.0161	0.0141	0.0749	Hidden Layers=1, Hidden Units=120, Learning Rate=0.001, Activation=ReLU
4	Multi-Shot ANN Model	0.0153	0.0135	0.0701	Hidden Layers=5, Hidden Units=all 24, Learning Rate=0.001, Activation=ReLU
5	Convolutional Neural Network	0.0149	0.0131	0.0667	Window Input Width=48, Filters=120, Learning Rate=0.001, Activation=Sigmoid
6	Long Short-Term Memory Network	0.0145	0.0122	0.0815	Window Input Width=96, LSTM Units=24, Learning Rate=0.001
7	24 LSTM Network-based Model	0.0149	0.0130	0.0694	Different Values for Each Model
8	Autoregressive LSTM Network	0.0147	0.0132	0.0601	Window Input Width=3, LSTM Units=228, Learning Rate=0.001
9	**Proposed Snow Conditions Compatible Model**	**0.0138**	**0.0123**	**0.0556**	-

9.4 Optimization Block of the EMS

The schematic of the developed two-stage MPC-based EMS is shown in Fig. 9.6. As shown in the figure, the EMS interacts with the microgrid sending reference set points to the different components of the microgrid and acquiring measures from the system through a local net area (LAN) communication. The tertiary controller is devoted to the schedule of the microgrid, where the long-term issues are solved. In this step, the control decisions are taken based on the generation and demand forecasts. The selected schedule horizon (*SH*) corresponds to the 24 hours of the day, discretized with a sample period of Ts = 1 h. The real operational scenario differs from the forecast, which requires a second

step where optimal power sharing among all the components of the microgrid must be carried out. This corresponds to the secondary control level of the microgrid. At this control level, the control decisions are taken with a faster sampling time of Ts = 5 min and a shorter horizon of CH = 20 min. The main objective of this control stage is tracking the schedule established in the tertiary controller. The developed EMS can be formulated as follows [145].

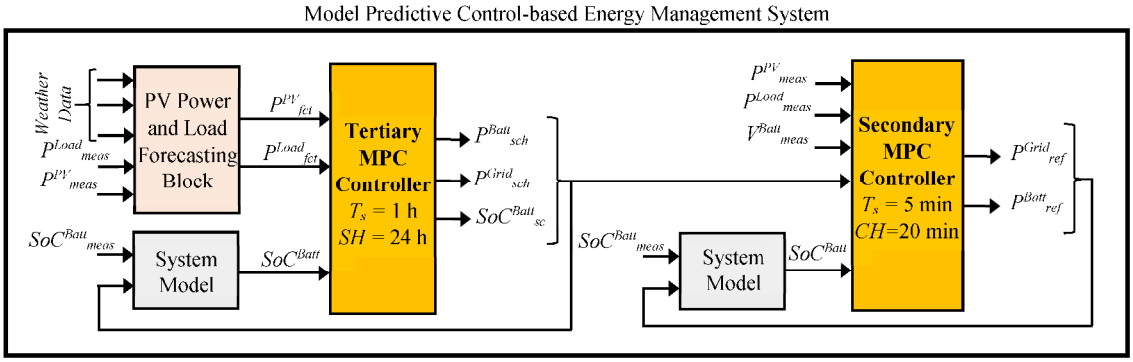


Figure 9.6. The schematic of the developed MPC-based EMS of the grid-connected microgrid.

9.4.1 System Model

The system model is an important and inseparable part of the MPC for both the tertiary and secondary control levels. The system model corresponds to the dynamic linear model of the state variables, given by the level of the stored energy in the batteries and other energy storage systems. The evolution of the state of charge of the battery and the equations describing the relationships between the SoC and the charging/discharging power of the battery form the system model which is formulated for the time instant t as follows:

$$SoC_{t+1}^{Batt} = SoC_t^{Batt} + \frac{Pch_t^{Batt} \times \eta ch^{Batt} \times T}{Cap^{Batt}} - \frac{Pdch_t^{Batt} \times T}{\eta dch^{Batt} \times Cap^{Batt}}; \quad (9.2)$$

$$Pch_t^{Batt} \geq 0 \ \& \ Pdch_t^{Batt} \geq 0; \quad (9.3)$$

$$P_t^{Batt} = Pdch_t^{Batt} - Pch_t^{Batt}; \quad (9.4)$$

$$\delta dch_t^{Batt} + \delta ch_t^{Batt} = 1; \quad (9.5)$$

where SoC^{Batt}, Cap^{Batt}, P^{Batt}, $Pch^{Batt}/Pdch^{Batt}$, and $\eta ch^{Batt}/\eta dch^{Batt}$ are the state of charge, capacity, net exchanged power, charging/discharging power, and charging/discharging efficiency of the battery, respectively. T is the time step length. Equation (9.2) specifies

that the battery SoC at each time depends on the SoC at the previous instance and the energy that the battery absorbed from/injected into the electrical system. δch^{Batt} and δdch^{Batt} are binary variables defining the charging and discharging states of the battery, respectively. The charging and discharging states are 0 and 1, respectively, if $P^{Batt} \geq 0$. This statement can be converted into mixed-integer inequalities as follows:

$$\varepsilon \leq -P_t^{Batt} - (Pmin^{Batt} - \varepsilon) \times \delta dch_t^{Batt} \leq \infty; \tag{9.6}$$

$$-\infty \leq -P_t^{Batt} + Pmax^{Batt} \times \delta dch_t^{Batt} \leq Pmax^{Batt}; \tag{9.7}$$

where $Pmin^{Batt}$ and $Pmax^{Batt}$ are the minimum and maximum power exchange rates of the battery corresponding to the charging and discharging processes, respectively. On the other hand, $Pdch^{Batt}$ equals P^{Batt} when the binary discharging state (δdch^{Batt}) is 1. This can be formulated by the following mixed-integer inequalities:

$$-\infty \leq Pdch_t^{Batt} - Pmax^{Batt} \times \delta dch_t^{Batt} \leq 0; \tag{9.8}$$

$$0 \leq Pdch_t^{Batt} - Pmin^{Batt} \times \delta dch_t^{Batt} \leq \infty; \tag{9.9}$$

$$-\infty \leq Pdch_t^{Batt} - P_t^{Batt} + Pmin^{Batt} \times (1 - \delta dch_t^{Batt}) \leq 0; \tag{9.10}$$

$$0 \leq Pdch_t^{Batt} - P_t^{Batt} + Pmax^{Batt} \times (1 - \delta dch_t^{Batt}) \leq \infty; \tag{9.11}$$

Considering (9.3) to (9.5), having the above linearized equations for $Pdch^{Batt}$ can satisfy the constraints and there is no need to repeat them for Pch^{Batt}. The SoC and net exchange power of the battery are bounded as follows:

$$SoCmin^{Batt} \leq SoC_t^{Batt} \leq SoCmax^{Batt}; \tag{9.12}$$

$$Pmin^{Batt} \leq P_t^{Batt} \leq Pmax^{Batt}; \tag{9.13}$$

Usually, the minimum SoC is intentionally chosen to be larger than 0.2 to avoid the detrimental effects of over-discharging the battery. The maximum SoC is technically limited by 1; however, it is considered as 0.9 here to ignore the highly nonlinear part of the battery charging curve.

9.4.2 Tertiary Controller – Economical Optimization

The formulation of the mixed-integer linear programming implemented in the tertiary controller over the schedule horizon is as follows.

9.4.2.1 Cost Function

The cost to be minimized by the tertiary controller can be formulated as the sum of the cost of energy exchange with the main grid and the cost related to the use of the battery.

$$Min.TCCF_t = \sum_{k=1}^{SH/T^S} \left(Cost_{t+k|t}^{Grid} + Cost_{t+k|t}^{Batt}\right) \tag{9.14}$$

$$Cost_{t+k|t}^{Grid} = \left(-Tariff_{t+k}^{Sale} \times Psale_{t+k|t}^{Grid} + Tariff_{t+k}^{Pur} \times Ppur_{t+k|t}^{Grid}\right) \times T^s; \tag{9.15}$$

$$Cost_{t+k|t}^{Batt} = \frac{CC^{Batt}}{2 \times LC^{Batt}} \left(Pch_{t+k|t}^{Batt} + Pdch_{t+k|t}^{Batt}\right) \times T_s; \tag{9.16}$$

where $TCCF$, $Cost^{Grid}$, and $Cost^{Batt}$ are the tertiary control cost function, the cost of energy exchange with the grid, and the degradation cost of the battery. $Tariff^{Sale}$/$Tariff^{Pur}$ and $Psale^{Grid}$/$Ppur^{Grid}$ are the electricity sale/purchase price and the power sold/purchased to/from the grid, respectively. CC^{Batt} and LC^{Batt} are the capital cost of the battery and the labor cost for replacing the battery, respectively. SH and T^s are the scheduling horizon and the scheduling time step, respectively; so that, $k \in [1{:}SH/T^S]$.

9.4.2.2 Constraints

The cost function of the programming approach implemented in the tertiary controller is constrained to the following equations:

- Power Exchange with the Grid

$$Psale_{t+k}^{Grid} \geq 0 \ \ \& \ \ Ppur_{t+k}^{Grid} \geq 0; \tag{9.17}$$

$$P_{t+k}^{Grid} = Ppur_{t+k}^{Grid} - Psale_{t+k}^{Grid}; \tag{9.18}$$

$$\delta pur_{t+k}^{Grid} + \delta sale_{t+k}^{Grid} = 1; \tag{9.19}$$

where P^{Grid} is the net power exchange with the grid. δpur^{Grid} and $\delta sale^{Grid}$ are binary variables that cannot be one at the same time. δpur^{Grid} is 1 if $P^{Grid} \geq 0$. This can be converted into the following mixed-integer inequalities:

$$\varepsilon \leq -P_{t+k}^{Grid} - (Pmin^{Grid} - \varepsilon) \times \delta pur_{t+k}^{Grid} \leq \infty; \tag{9.20}$$

$$-\infty \leq -P_{t+k}^{Grid} + Pmax^{Grid} \times \delta pur_{t+k}^{Grid} \leq Pmax^{Grid}; \tag{9.21}$$

where $Pmin^{Grid}$ and $Pmax^{Grid}$ are the minimum and maximum permitted power exchange rates with the grid. On the other hand, $Ppur^{Grid}$ equals $P^{Grid} \times \delta pur^{Grid}$. This can be formulated by mixed-integer inequalities as follows:

$$-\infty \leq Ppur_{t+k}^{Grid} - Pmax^{Grid} \times \delta pur_{t+k}^{Grid} \leq 0; \tag{9.22}$$

$$0 \leq Ppur_{t+k}^{Grid} - Pmin^{Grid} \times \delta pur_{t+k}^{Grid} \leq \infty; \tag{9.23}$$

$$-\infty \leq Ppur_{t+k}^{Grid} - P_{t+k}^{Grid} + Pmin^{Grid} \times (1 - \delta pur_{t+k}^{Grid}) \leq 0; \tag{9.24}$$

$$0 \leq Ppur_{t+k}^{Grid} - P_{t+k}^{Grid} + Pmax^{Grid} \times (1 - \delta pur_{t+k}^{Grid}) \leq \infty; \tag{9.25}$$

Considering (9.17) to (9.19), having the above linearized equations for $Ppur^{Grid}$ can satisfy the constraints and there is no need to repeat them for $Psale^{Grid}$. The power exchange with the grid is bounded as follows:

$$Pmin^{Grid} \leq P_{t+k|t}^{Grid} \leq Pmax^{Grid}; \tag{9.26}$$

The upper and lower bounds of the energy exchange with the grid can be assigned by the rated capacity of the main feeder connecting the microgrid to the upstream grid or the main feeder transformer.

- Energy Equilibrium

$$Pfct_{t+k|t}^{PV} - Pfct_{t+k|t}^{Load} + P_{t+k|t}^{Grid} + P_{t+k|t}^{Batt} = 0; \tag{9.27}$$

where $Pfct^{PV}$ and $Pfct^{Load}$ are the forecasted values of the PV power generation and load demand of the microgrid at $t+k$. The above equation guarantees the demand-supply balance at all times. In addition, the system model equations are added to the above programming approach.

9.4.3 Secondary Controller – Power Sharing

The formulation of the secondary controller can be presented as quadratic programming over the control horizon as follows.

9.4.3.1 Cost Function

The cost function of the secondary controller is characterized by two terms associated with the grid interaction and the battery operation. The grid-related term incurs a high penalty for the deviation of the power exchange with the grid in real time with respect to the power scheduled by the tertiary controller. The battery-related term

considers the deviations in the energy and power references and prevents high current values in the charging and discharging processes.

$$Min.SCCF_t = \sum_{k=1}^{CH/T^c} \left(Penalty_{t+k|t}^{Grid} + Penalty_{t+k|t}^{Batt}\right) \tag{9.28}$$

$$Penalty_{t+k|t}^{Grid} = W^{Grid} \times (P_{t+k|t}^{Grid} - Psch_{t+k|t}^{Grid})^2; \tag{9.29}$$

$$\begin{aligned} Penalty_{t+k|t}^{Batt} = {} & We^{Batt} \times \left(SoC_{t+k|t}^{Batt} - SoCsch_{t+k|t}^{Batt}\right)^2 \\ & + Wp^{Batt} \times \left(P_{t+k|t}^{Batt} - Psch_{t+k|t}^{Batt}\right)^2 \\ & + Wdeg^{Batt} \times \left(\frac{Pch_{t+k|t}^{Batt} + Pdch_{t+k|t}^{Batt}}{Vdc_t^{Batt}}\right)^2 \\ & + Wripple^{Batt} \times \left(\frac{\Delta Pch_{t+k|t}^{Batt} + \Delta Pdch_{t+k|t}^{Batt}}{Vdc_t^{Batt}}\right)^2; \end{aligned} \tag{9.30}$$

where *SCCF*, $Penalty^{Grid}$, and $Penalty^{Batt}$ are the secondary control cost function, grid-related penalty, and battery-related penalty, respectively. W^{Grid}, We^{Batt}, Wp^{Batt}, $Wdeg^{Batt}$, and $Wripple^{Batt}$ are the weights to penalize the power exchange deviations with the grid, the energy deviations of the battery, the power deviations of the battery, the battery degradation, and the battery power variations (ripple), respectively. $Psch^{Grid}$, $SoCsch^{Batt}$, and $Psch^{Batt}$ are the tertiary control schedules for the power exchange with the grid, SoC of the battery, and the battery power, respectively. Vdc^{Batt} is the voltage at the DC bus of the battery. *CH* and T^c are the control horizon and control time step, respectively; so that, $k \in [1{:}CH/T^C]$.

9.4.3.2 Constraints

The cost function of the programming approach implemented in the secondary controller is constrained to the following equations:

- Power Exchange with the Grid

$$Pmin^{Grid} \leq P_{t+k|t}^{Grid} \leq Pmax^{Grid}; \tag{9.31}$$

Similar to (9.26), the energy exchange with the grid should be limited between the specified upper and lower bounds.

- Energy Equilibrium

$$Pmeas_t^{PV} - Pmeas_t^{Load} + P_{t+k|t}^{Grid} + P_{t+k|t}^{Batt} = 0; \quad (9.32)$$

where $Pmeas^{PV}$ and $Pmeas^{Load}$ are the measured PV power generation and load demand of the microgrid, respectively. It is assumed that the measurements of the PV power and load demand can be used as the forecasts for the next few minutes over the control horizon. Similar to the programming approach of the tertiary controller, the system model equations are added to the above optimization approach.

9.5 Numerical Results

A renewable commercial microgrid, with the schematic shown in Fig. 9.1, connected to the upstream power grid via a 5kV feeder is simulated on the Simulink/MATLAB software. A PV plant with a rated power of 1200 kW and a battery storage system with a rated capacity of 2500 kWh and a rated power of 400 kW are connected to the AC bus of the microgrid. Considering the connection of the microgrid to the upstream electric distribution grid, time-of-use tariffs are assumed to be applied to the energy exchange with the upstream grid. By taking the TOU structures of several utilities in the U.S. into consideration, the price of energy is categorized here into three levels including off-peak (from 12 a.m. to 5 a.m.), mid-peak (from 6 a.m. to 12 p.m. and from 9 p.m. to 11 p.m.), and peak (from 1 p.m. to 8 p.m.) levels with an energy price of 7, 12, and 27 cents/kWh, respectively. The programming approaches of the tertiary and secondary controllers of the MPC-based EMS are implemented in Python programming language using the Pyomo library and are solved by the CPLEX solver. These together with the Python-based forecasting models are integrated into the Simulink model of the microgrid; so that, they receive the measurements of the electrical and meteorological parameters in real time, perform the calculations, and send the control signal to the battery charge/discharge controller.

In order to evaluate the performance of the proposed EMS in different weather conditions, three weather cases of sunny, cloudy, and snowy days are investigated. The PV panels are subjected to a snow cover on the snowy day. The real and forecasted values of the PV power generation using the proposed PVPF method for three arbitrarily chosen days with the aforementioned conditions are shown in Fig. 9.7. The real and forecasted values of the load demand using the developed load forecasting approach are also indicated in Fig. 9.8. These plots show the values from one day before to one day after the operating day to include all the forecasts for the day under study. As can be seen,

the forecasted PV power for the sunny day in Fig. 9.7 (a) is very close to the actual PV power generation. The forecasts are subject to larger errors for the cloudy day, as shown in Fig. 9.7 (b). Considering the focus on snow conditions, the PV power forecasts obtained by the base LSTM model and the proposed PVPF approach are compared in Fig. 9.7 (c). As can be seen, the base LSTM model does not consider the possibility of snow cover formation on the panels in the future hours which results in large day-ahead forecast errors. On the other hand, the proposed approach can appropriately forecast this possibility and modify the PV power forecasts based on the snow-cover condition. The day-ahead load demand forecasts in Fig. 9.8 show a good agreement with the actual values.

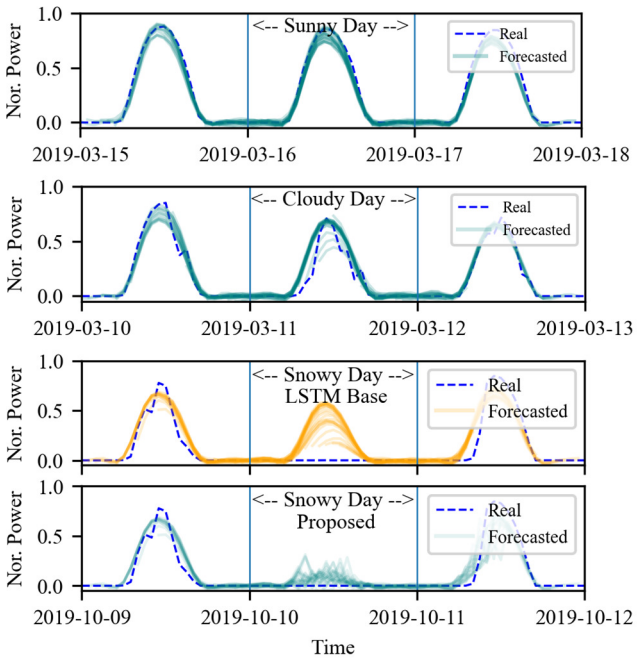


Figure 9.7. Real and day-ahead hourly forecasts of the PV power generation on the given sunny, cloudy, and snowy days.

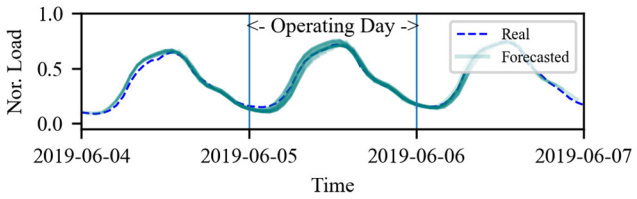


Figure 9.8. Real and day-ahead hourly forecasts of the microgrid load demand on a given operating day.

In the following, three cases are studied to allow us to evaluate the effectiveness of the proposed MPC-based EMS, especially in snow conditions. Concerning the predetermined initial and final SoCs of the battery in the simulations and in order to

consider all the effective day-ahead forecasts for the main operating day, the simulations are performed on three consecutive days so that the focus is on the day in the middle.

9.5.1 Case 1: Heuristic Control Method of The Microgrid

In the first case, a heuristic control method is developed for the EMS of the microgrid which is used as a benchmark. This method generates the control signal of the battery according to the battery SoC, PV power generation, load demand of the microgrid, and time of day. The algorithm of this control method is as follows:

Algorithm: to control charge/discharge of the battery

Input: measurements of load demand (P^{Load}), PV power (P^{PV}), battery SoC (SoC^{Batt}), and time of day ($Time$).

Output: battery charge/discharge power (P^{Batt}) ($P^{Batt} \leq 0$: charge and $P^{Batt} \geq 0$: discharge).

Initialization:

$P^{Batt} = 0$. (idle mode)
while $Time < 6$ **and** $SoC^{Batt} < SoCmax^{Batt}$ **then**
    $P^{Batt} = Pmin^{Batt}$. (off-peak charging mode)
end while
while $6 \leq Time < 17$ **and** ...
($(SoC^{Batt} < SoCmax^{Batt}$ **and** $P^{Load} < P^{PV})$ **or** ...
($SoC^{Batt} > SoCmin^{Batt}$ **and** $P^{Load} > P^{PV})$) **then**
    if $P^{Load} < P^{PV}$ **then**
        if $P^{Load} - P^{PV} < Pmin^{Batt}$ **then** $P^{Batt} = Pmin^{Batt}$.
        else $P^{Batt} = P^{Load} - P^{PV}$.
    if $P^{Load} > P^{PV}$ **then**
        if $P^{Load} - P^{PV} > Pmax^{Batt}$ **then** $P^{Batt} = Pmax^{Batt}$.
        else $P^{Batt} = P^{Load} - P^{PV}$.
end while
while $17 \leq Time < 24$ **and** $SoC^{Batt} > SoCmin^{Batt}$ **then**
    $P^{Batt} = Pmax^{Batt}$. (peak discharging mode)
end while

The developed heuristic method is added to the simulated microgrid and its performance is evaluated for the aforementioned sunny, cloudy, and snowy days with the same microgrid load demand curve. The measurements of the electrical parameters, including the voltage at the point of common coupling (PCC), PV power generation, load demand of the microgrid, power exchange with the grid, charge/discharge power of the battery, and battery SoC for the three weather conditions are shown in Fig. 9.9. The initial and final battery SoCs over the three days of the study are considered as 0.2 (or 20%). The following analysis focuses on the main operating day, which is the day in the middle.

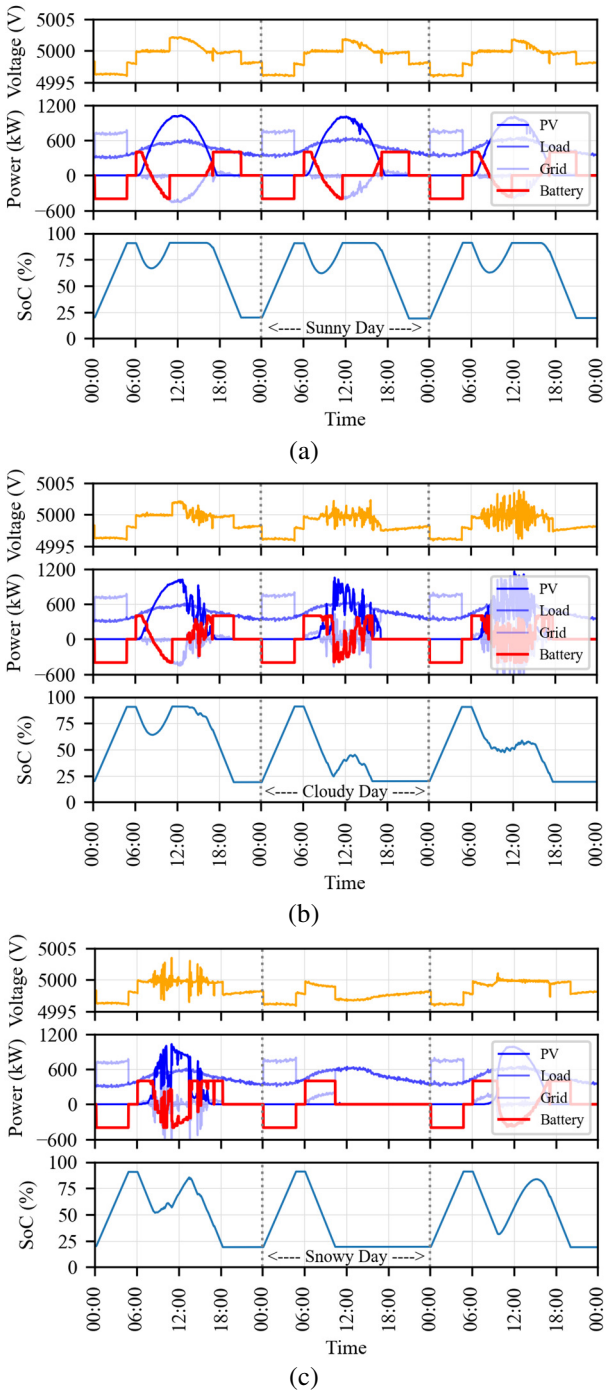


Figure 9.9. Microgrid operation using the heuristic control method on the (a) sunny day, (b) cloudy day, and (c) snowy day.

As can be seen in all the plots, the battery is charged with the nominal charging rate by absorbing energy from the upstream grid from midnight to 5 a.m., when the price of energy is lower, and reaches $SoCmax^{Batt}$. From 6 a.m. to 5 p.m., when the price of energy is higher, the battery is discharged into the microgrid to compensate for the energy

shortage. When the PV power exceeds the microgrid demand during this period, the battery is charged again with the existing energy surplus. According to Fig. 9.9 (a), on the sunny day, the battery is discharged from 6 a.m. to around 8 a.m., charged again until it gets fully charged before 12 p.m., and stays idle until around 4 p.m. With the decrease of the PV power at 4 p.m., the battery injects the energy back to supply the microgrid and prevent the energy purchase from the grid at the peak level of the TOU tariff. The discharge continues even after 5 p.m. until the battery SoC reaches its permitted lower limit. As indicated in Fig. 9.9 (b), the battery starts being discharged at 6 a.m. on the cloudy day but unlike the sunny day, the PV power is so low until 10:30 a.m. that the battery continues discharging until its SoC reaches 25%. The battery stores the extra energy generated by the PV system until 1 p.m. but it injects back this energy into the microgrid due to the highly volatile PV power production caused by clouds. This causes the battery SoC to reach 20% at 4 p.m. So, the battery is not able to support the microgrid in the evening when the price of energy is high. On the snowy day, according to Fig. 9.9 (c), the PV power is almost zero due to solar irradiance blockage by the snow cover on the panels. Hence, the battery starts being discharged at 6 a.m. until its SoC reaches 20% at around 10:30 a.m. Since then, the battery stays idle at $SoCmin^{Batt}$ because there is no PV energy surplus and the microgrid load is fully supplied from the grid.

Considering the power exchange between the microgrid and the upstream grid, the TOU tariffs, and the battery degradation cost, the operation cost of the microgrid is calculated. The cumulative operation cost of the microgrid for the three weather conditions is illustrated in Fig. 9.10, for the case that no opportunity for selling energy back to the grid exists ($Tariff^{Sale} = 0$), and in Fig. 9.11, for the case that the microgrid is paid for injecting energy into the grid. In the latter case, it is assumed that $Tariff^{Sale} = Tariff^{Pur} + 2$. As can be seen in Fig. 9.10, the microgrid has the lowest operation cost on the sunny day and the highest on the snowy day; while, the operation cost on the cloudy day is somewhere in the middle. This originates from the fact that the availability of free renewable energy can significantly reduce the operation cost by decreasing the reliance on the upstream grid. On the sunny day, the operation cost increases gradually over the day. On the cloudy day, the cost increase is intensified after 5 p.m. when the PV power gets zero and the battery is not able to support the microgrid. On the snowy day, the slope of the curve gets larger after 1 p.m. by starting the peak level of the TOU tariff. The operation cost of the microgrid on the sunny, cloudy, and snowy days are 555$, 1134$,

and 1833$, respectively. The battery degradation cost on the sunny, cloudy, and snowy days are 83$, 80$, and 60$, respectively.

By considering the opportunity of selling energy back to the grid, the operation cost of the microgrid can be reduced. As indicated in Fig. 9.11, the operation cost on the sunny day, which equals 309$, is much lower than that of Fig. 9.10. The availability of the PV energy surplus and injecting it into the grid between 12 p.m. and 4 p.m. helps to reduce the operation cost. On the cloudy day, the microgrid does not have the opportunity to sell a considerable amount of energy to the grid except for a very short time. Hence, the operation cost with the amount of 1085$ is only a little lower than that in Fig. 9.10. The operation cost on the snowy day is similar to that in Fig. 9.10 since the microgrid is not able to sell energy to the grid. Considering the similar strategies of the controller with and without the energy sale opportunity, the battery degradation costs are the same in both cases.

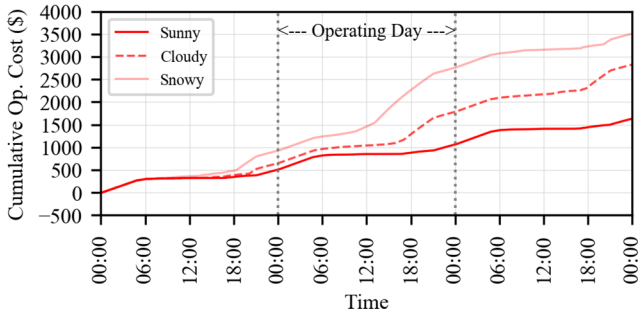


Figure 9.10. Cumulative microgrid operation cost using the heuristic control method without the energy sale opportunity.

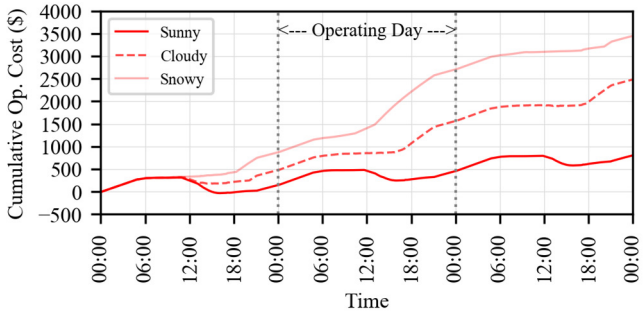


Figure 9.11. Cumulative microgrid operation cost using the heuristic control method with the energy sale opportunity.

9.5.2 Case 2: Proposed EMS without Energy Sale Opportunity

In this section, the microgrid operation by implementing the proposed MPC-based EMS is investigated when the microgrid is not paid for injecting energy back into the

grid. The measurements of the electrical parameters for the sunny, cloudy, and snowy days are indicated in Fig. 9.12. Similar to the heuristic control method, the initial and final battery SoCs over the three days of the study are considered as 20%. On the sunny day, the battery is charged from 20% to around 35% when the price of energy is low before 6 a.m. With the rise of the energy price, the battery is discharged until its SoC gets to 20%. As the PV power increases, the battery is charged by the energy surplus until the amount of the PV power decreases to the load demand of the microgrid. Then, the battery returns the energy back to the microgrid until it gets fully discharged in order to cover the load demand when the energy price is high. In this case, the battery is potentially able to get charged to 90% of its capacity with the surplus of the PV power; however, the SoC reaches up to 85% in the simulation. This might be due to the small forecasting errors of the PV power by the EMS. As can be seen in Fig. 9.12 (b), the battery starts being charged sooner on the cloudy day at 1 a.m. since the PV power forecasts show a lower availability of the PV energy compared to the sunny day. Moreover, the battery SoC gets to 90% at 12 p.m. by receiving energy partly from the grid and the PV system. The battery injects the energy into the microgrid afterward to support the load demand whenever the PV production drops. The discharge continues to avoid the high energy price of the upstream grid until the SoC reaches 20% at 8 p.m. Unlike the sunny and cloudy days, the battery starts the charging process sooner on the snowy day at 12 a.m. and the charging continues without interruption until the SoC reaches almost 90% at around 7:30 a.m. It means that the battery absorbs energy from the grid and stores it over all the hours of the lower level of the tariffs. The battery has small activities until 4 p.m. and then, it starts being discharged to support the microgrid during the high level of the tariffs until its SoC reaches 20% at 7:30 p.m. This strategy originates from the fact that the controller can appropriately forecast the low availability/unavailability of the PV energy on the snowy day and implements the optimal approach for minimizing the operation cost of the microgrid by purchasing energy from the grid when the energy price is low, storing and preserving the energy in the battery, and injecting the energy back into the microgrid when the price of energy is high.

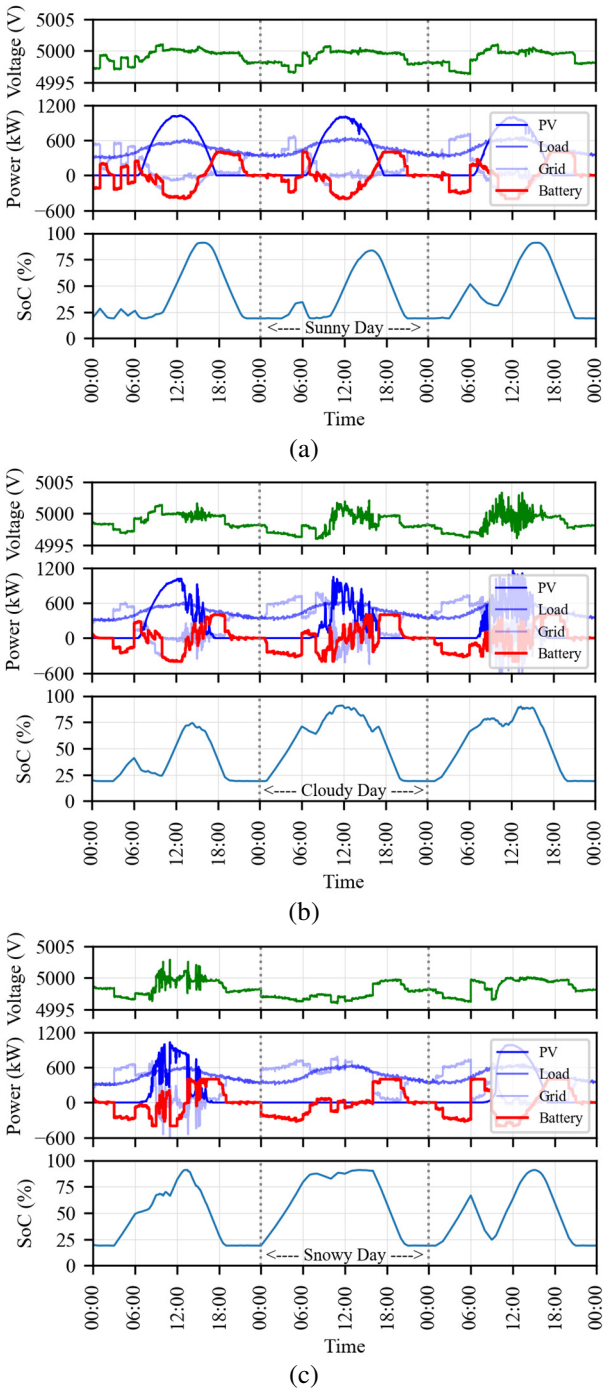


Figure 9.12. Microgrid operation using the proposed EMS on the (a) sunny day, (b) cloudy day, and (c) snowy day without the energy sale opportunity.

The cumulative operation cost of the microgrid by implementing the proposed EMS for the case without the energy sale opportunity is indicated in Fig. 9.13. Similar to the case of the heuristic controller, the operation cost of the microgrid on the sunny day is lower, on the snowy day is higher, and on the cloudy day is in the middle. Unlike the case of the heuristic controller, the cost increases at almost the same pace over the day, the

curves are smoother, and the cost rise intensification between 5 p.m. and 8 p.m. on the cloudy day and between 1 p.m. and 8 p.m. on the snowy day does not occur. The reason is that the controller can oversee the PV energy availability in the hours ahead and get prepared for the upcoming situation by reducing the energy purchased from the upstream grid during high energy prices. The operation cost of the microgrid using the proposed approach on the sunny, cloudy, and snowy days are 528$, 1017$, and 1598$, which are 27$, 117$, and 235$ lower compared to those of the heuristic approach, respectively. The large cost reduction on the snowy day proves the importance of implementing a snow condition-compatible EMS. The battery degradation cost on the sunny, cloudy, and snowy days are 68$, 74$, and 65$, respectively, which are 15$ and 6$ lower on the sunny and cloudy days and 5$ higher on the snowy day compared to the heuristic controller.

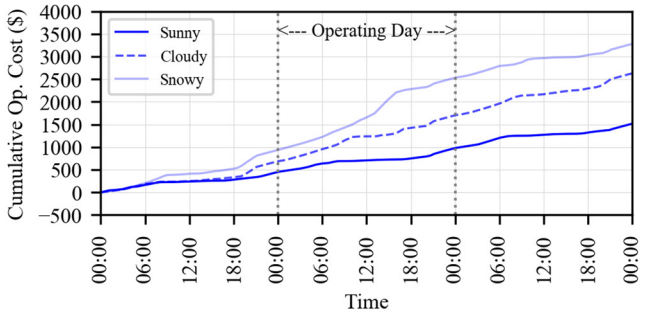


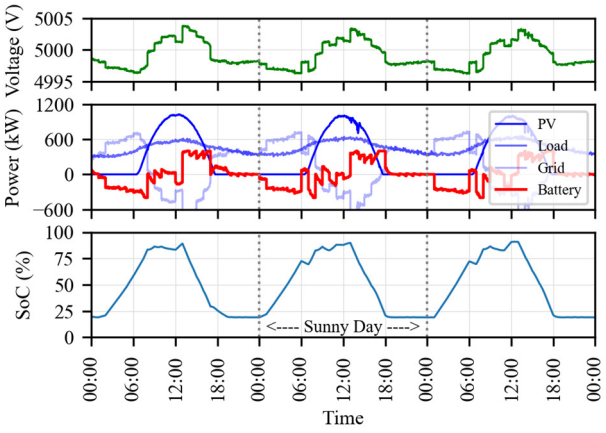
Figure 9.13. Cumulative microgrid operation cost using the proposed EMS without the energy sale opportunity.

9.5.3 Case 3: Proposed EMS with Energy Sale Opportunity

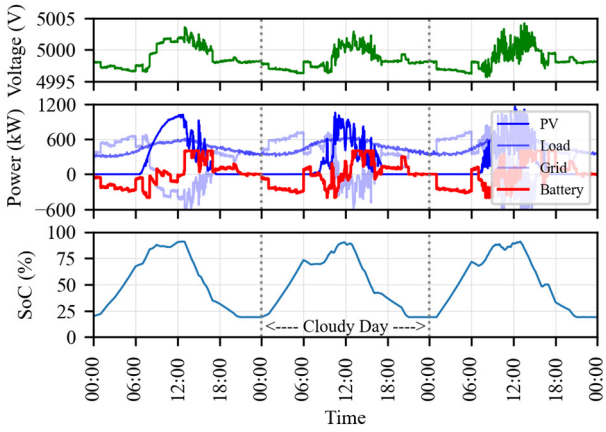
By implementing the proposed MPC-based EMS and considering the microgrid's opportunity to sell energy back to the upstream grid, the operation of the microgrid is investigated on the sunny, cloudy, and snowy days and the results are shown in Fig. 9.14. As can be seen in Fig. 9.14 (a), the battery is charged by absorbing energy from the grid up to 75% of its capacity on the sunny day before 6 a.m. when the price of energy is low. Then, it is slightly charged and discharged until the SoC reaches 90% mainly from the surplus of the PV energy at 1 p.m. Unlike the controller strategy in Fig. 9.12 (a), where the battery is mainly charged from the PV energy at noon, the controller charges the battery mainly from the grid in the morning at a lower price and sells the PV energy surplus to the grid at a higher price. Since the microgrid can obtain revenue from selling energy to the grid, the controller starts discharging the battery at 1 p.m., when the higher level of the energy tariffs begins, despite the fact that there is a PV energy surplus in the

microgrid. The discharge continues until the battery SoC reaches 20% at 6 p.m. The charging strategies taken by the controller on the cloudy and snowy days are very similar to those on the sunny day. The battery is charged up to 75% of its capacity before 6 a.m., fully charged from the PV energy surplus before 12 p.m., and discharged to 20% of its capacity until the peak-level tariff ends at 9 p.m. on the cloudy day. On the snowy day, however, the battery is charged up to almost 70% of its capacity before 6 a.m. and then, gradually charged up to 90% of its capacity from the grid before 1 p.m. That is while the potential optimal solution could be charging the battery to a higher SoC level during the off-peak level of the tariffs so that the battery relies on a smaller amount of energy purchased from the grid at the mid-peak level of the tariffs to get fully charged. Failure to implement this optimal strategy might be due to the small PVPF errors of the EMS and overestimating the PV power generation at the early hours of the day for the period between 6 a.m. and 12 p.m. The battery stays idle until 4 p.m. and then, is discharged to 20% of its capacity before the peak level of the tariffs ends.

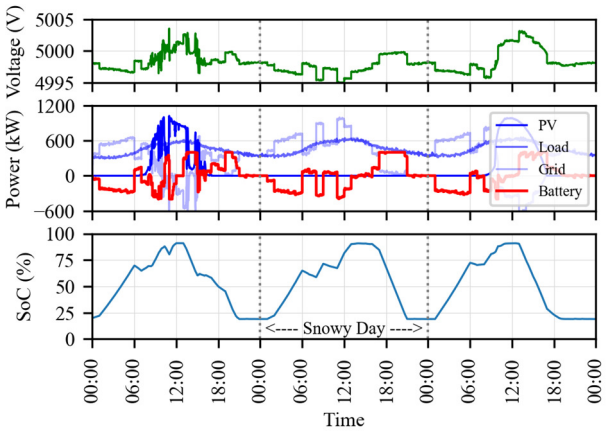
The cumulative operation cost of the microgrid by implementing the proposed MPC-based EMS when considering the microgrid's opportunity to sell energy back to the upstream grid is indicated in Fig. 9.15. By comparing the curves for the sunny and cloudy days on this figure with those of the heuristic control method on Fig. 9.11, it can be seen that the microgrid receives a larger amount of revenue by selling energy to the grid in the afternoon so that the daily operation cost gets negative (obtaining profit) right before 6 p.m. The cost rise has a larger pace before 12 p.m. on the sunny day compared to that of Fig. 9.13 in the case without the sale opportunity due to the larger energy purchased from the grid during this period. The curve on the snowy day is very similar to the one of Fig. 9.13. The operation cost using the proposed approach and considering the sale opportunity on the sunny, cloudy, and snowy days are 285$, 919$, and 1623$, which are 24$, 166$, and 210$ lower compared to those of the heuristic approach, respectively. The battery degradation cost on the sunny, cloudy, and snowy days are 65$, 66$, and 69$, respectively, which are 18$ and 14$ lower on the sunny and cloudy days and 9$ higher on the snowy day compared to the heuristic controller.



(a)



(b)



(c)

Figure 9.14. Microgrid operation using the proposed EMS on the (a) sunny day, (b) cloudy day, and (c) snowy day with the energy sale opportunity.

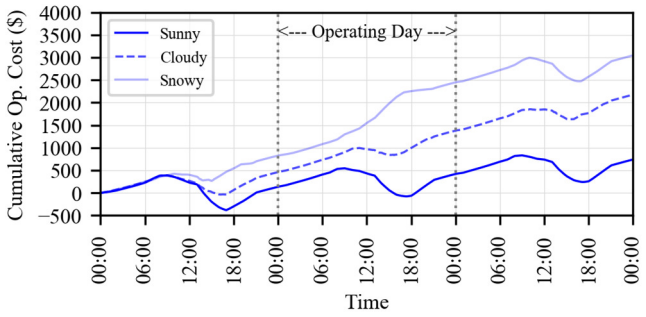


Figure 9.15. Cumulative microgrid operation cost using the proposed EMS with the energy sale opportunity.

In general, the microgrid MPC-based EMS has the ability to oversee the future operation condition of the microgrid based on the forecasts of the uncertain input variables and the system model, and adopt the optimal charging strategy for the battery in order to minimize the operation cost of the microgrid. As indicated in the case studies, implementing the proposed EMS can reduce the operation cost of the microgrid compared to the developed heuristic control method. The percentage of the cost reduction for the sunny, cloudy, and snowy days considering the operation cost over all three days of the simulation and over the main operating day are shown in Fig. 9.16. As can be seen, the proposed EMS can reduce the 3-day microgrid operation cost by up to 7% (on the sunny day) and 12.5% (on the cloudy day) for the two cases of without and with the sale opportunity, respectively. By focusing on the main operating day, the microgrid operation cost is reduced by up to 13% (on the snowy day) and 15% (on the cloudy day) for the cases without and with the sale opportunity, respectively. The amounts of cost reduction in Fig. 9.16 prove not only the effectiveness of the proposed EMS on sunny and cloudy weather conditions but also its compatibility with the snow conditions.

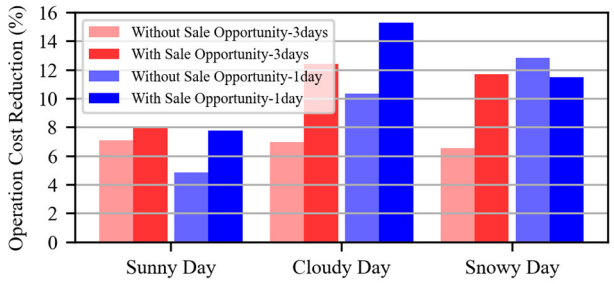


Figure 9.16. Microgrid operation cost reduction percentages using the proposed MPC-based EMS compared to the heuristic control method.

9.6 Conclusion

This chapter addresses a computational intelligence-based energy management system for microgrids in snow-prone regions. This system, which is based on a two-stage model predictive control, benefits from a novel short-term PV power forecasting approach compatible with snow conditions and a load demand forecasting model. The proposed PVPF approach consists of an XGBoost-based snow-cover detector, an autoregressive XGBoost-based snow-cover forecaster, and two LSTM-based PV power forecasters which can reduce the forecasting error by almost 42% in snow-cover conditions compared to a base LSTM model. The load forecasting model is also developed based on an autoregressive LSTM network which provides day-ahead forecasts with errors as small as 5%. The EMS is formed by a tertiary controller, which is in charge of the economical optimization of the microgrid, and a secondary controller, which adjusts the set points sent to the primary controllers. The proposed EMS is applied to a simulated grid-connected microgrid with a large PV system and a battery storage system and its performance is compared with a heuristic controller. As the results show, the proposed approach appropriately oversees the microgrid operation condition, adopts the optimal charging strategy, and reduces the microgrid operation cost by at least 5% (on a sunny day) and up to 15% (on a cloudy day) compared to the heuristic controller. In snow conditions, the proposed EMS makes the largest operation cost reduction which equals 235$ (13%).

CHAPTER 10

CONCLUSION AND FUTURE WORK

10.1 Concluding Remarks

Over the past few years, the devastating effects of global warming on the environment and the severe climate change, that threatens human life, have become more apparent to people all over the world. Consequently, governments and communities are getting more aware and ready to widely switch to green electricity to reduce their carbon footprint. This has resulted in a paradigm shift in the electricity sector towards renewable energy resources among which photovoltaic energy has gained significant popularity due to the low maintenance costs, strong persistence, and high reliability of PV systems. It has been even expected that PV systems become significantly popular in the near future driven by the downward trend in the price of PV technology. Currently, PV systems not only are being installed in countries with high solar potential but also are getting popular in northern areas where winters are very long and harsh. The performance of PV systems highly depends on the meteorological parameters and can be negatively affected in harsh weather conditions. Power generation of PV systems in northern snow-prone areas can be reduced by snow accumulation on the panels or even completely disrupted. Hence, modeling the performance, power losses, and power generation of PV systems in snow conditions is a more challenging task. In this thesis and for the first time in the literature, according to the best of our knowledge, we addressed this issue by developing data-driven PV power losses and power generation models for systems in snow-prone areas using the capabilities of computational intelligence techniques. The challenging problems addressed in this thesis are categorized into the following topics:

- Systematic analysis and computational intelligence-based modeling of PV system power losses in snow conditions;
- Systematic analysis and computational intelligence-based modeling of PV system power generation in snow conditions;
- Computational intelligence-based short-term PV power forecasting in snow conditions and its applications in power grids' scheduling.

In order to achieve the objectives of the introduced topics, different methodologies were proposed and their effectiveness was validated for the case of several PV systems in snow-prone regions. The major findings of this thesis can be summarized as follows:

1) Estimating the amounts of different power losses in PV systems can give a better insight into the performance of the systems and help system operators in making short- and long-term decisions about necessary maintenance and modifications to the system. Among different types of power losses, PV systems in northern regions can be exposed to large amounts of snow loss in winter. This energy loss is mainly due to the accumulation of snow on the surface of PV panels which can significantly reduce the PV energy yield and hence, the profitability of the system. Moreover, the other types of power losses in PV panels, conversion systems, and connections, which depend on meteorological, electrical, physical, and environmental factors, reduce the efficiency of the PV system. Hence, monitoring and modeling different types of power losses in PV systems can be of significant importance. While calculating the values of different power losses requires a vast knowledge of the underlying principles and can be a challenging task, computational intelligence techniques seem to be able to capture this complexity automatically, provide reliable models of the power losses, and ease the power losses estimation for end users. This has been proved through two studies in this thesis.

   - In the first study, a methodology for independently estimating the daily values of snow loss for PV systems based on a 3-stage model and using the historical records of the main electrical and meteorological parameters was developed and different types of computational intelligence-based snow loss prediction models were proposed. The effectiveness of the proposed approach was validated in the case of a PV farm located in Ontario, Canada. Among different developed models, gradient boosting trees followed by random forests showed the minimum error of snow loss prediction in a cross-validation case.

   - In the second study, a methodology for systematic calculation of PV power losses was developed by which different types of power losses in PV systems, including both array capture losses (i.e. temperature loss, mismatching and soiling losses, low irradiance, spectral, and reflection losses, module quality degradation, and snow loss) and system losses (i.e. inverter loss, cabling loss,

inverter power limitation loss, and MPPT losses) can be extracted from the historical data of the main electrical and meteorological parameters of the system. Then, computational intelligence-based modeling of PV system power losses was proposed to build prediction models for each type of power loss solely based on some main meteorological parameters as the inputs of the models. The proposed approach was validated in the case of a 1.44 kW rooftop PV system located in Denver, CO. The developed models were also applied to the case of a 68.4 kW PV system located in Las Vegas, NV, to demonstrate their generalization capability. For the losses associated with the inverter, DC cabling, low irradiance, MPPT, and temperature, LSTM networks showed superior performance. For performance ratio, snow loss, and mismatch and soiling losses, gradient boosting trees presented the minimum prediction errors. This study proves the capability of computational intelligence techniques in capturing the complex nature of different power losses in PV systems.

2) Power generation of PV systems highly depends on some meteorological parameters. Modeling the PV power based on these parameters help power grid operators in getting a clearer view of the operation condition of the grid and evaluating it in different weather conditions. Since the power generation of PV systems in northern regions can get negatively affected by snowfalls, considering this effect in the modeling procedure is very important and of course, challenging. However, computational intelligence techniques are expected to address this issue effectively. This was investigated by conducting two studies.

- A key step in PV power prediction in snow conditions is the detection of snow covers on PV panels. PV power generation is disrupted by snowfall events only if snow accumulation occurs on the panels. This power reduction continues until the snow cover is cleared up by melting or sliding. In the first study, data-driven snow cover prediction modeling using computational intelligence techniques for PV systems in northern snow-prone areas was proposed. In order to recognize the presence of a full/partial snow cover on the panels based on the historical records of the electrical and meteorological parameters of PV systems, a 3-step computational intelligence-based approach was proposed to annotate datasets by binary labels. Then, various snow cover prediction models

were developed. The performance of the proposed approach was validated in the case of a roof-top PV system in Edmonton, Canada. The best developed snow cover predictor was also tested in the case of two other PV systems to demonstrate its generalization capability. The gradient boosting tree model demonstrated the best performance by correctly predicting 96% of the snow-covering hours. Snow-cover predictions provide the basis for the next study on PV power prediction in snow conditions.

- In the second study, computational intelligence-based PV power prediction in snow conditions was addressed. The proposed methodology was based on the categorization of the data as snow-related conditions, i.e. snow-free, snow, and snow-cover conditions, and developing separate prediction models for each condition. Comprehensive data analysis and computational intelligence-based modeling were applied to the case of 17 PV systems across Canada. It was shown that the prediction errors in the snow and especially snow-cover conditions are smaller by using specific models built for the snow and snow-cover datasets compared to those built for the full and snow-free datasets. The results also showed significantly lower prediction errors, especially in the snow-cover condition, using the proposed approach compared to the well-known Marion model and some other models. This proves the effectiveness of the probable snow cover detection and distinct PV power prediction models built for the snow-related conditions.

3) Short-term forecasting of the power generation of PV systems for hours/days ahead is a very challenging and important task in the scheduling problem of power grids. The accuracy of this task can affect the grid operators' decisions in the electricity markets, unit commitment, economic dispatch, and finally the operation cost of the system. Computational intelligence-based short-term PV power forecasting models can be developed in two ways: one dependent on weather forecasts from external sources and the other independent of any external inputs. Both of these approaches were addressed in this thesis for PV systems in snow-prone areas and their applications in typical power grid scheduling problems were investigated.

- In one study, the day-ahead scheduling of a PEVs aggregator that participates in both day-ahead and real-time energy markets on behalf of the PEV owners

through the local DSO was addressed. The aggregator can also provide the DSO's extra energy demand in real-time through a proposed out-of-market balancing service, and reduce the charging costs of the vehicles. A two-stage stochastic approach with a comprehensive availability model of PEVs for both residential night-charging and public intraday-charging was developed that minimizes the total charging cost of the PEVs. Computational intelligence-based PV power and load demand predictors dependent on meteorological forecasts were developed to generate scenarios on a snowy day. This study showed that building prediction models based on historical data containing data points in the most similar weather conditions to the operating day can be a smart solution to reduce large prediction errors caused due to less frequent but high-impact events, such as snow covers over a year. The models then can be updated regularly as new data getting available. Moreover, the results showed that integrating the proposed balancing service into the charge scheduling approach of the aggregator is effective in reducing the total charging cost of the PEVs and reducing the real-time cost of the distribution system operator.

- In another study, a model predictive control-based energy management system for a PV-penetrated microgrid in a snow-prone area was proposed. By using a two-stage MPC, the EMS is responsible for the secondary and tertiary control levels of the microgrid. The EMS consists of a forecasting block, where novel computational intelligence-based PV power and load forecasting models compatible with snow conditions and independent of exogenous weather forecasts make short-term forecasts of the PV power generation and load demand of the microgrid, and an optimization block, where the optimal control signals are generated for controllable DERs and storage systems in the microgrid. This study showed that using LSTM models is an appropriate choice for PV power and load forecasting. However, it can result in large PV power forecast errors when a snow cover is expected for future hours. As proposed in this study, the PVPF performance can be improved by integrating the baseline PV power forecasts of the LSTM model and the snow-cover forecasts generated by XGBoost trees into a single framework and updating the PV power forecasts for the hours with a probable snow-cover using LSTM-based PVPF models trained specifically for snow conditions. Moreover, the developed MPC-based

EMS is proven to reduce the microgrid operation cost not only in snow conditions but also on sunny and cloudy days compared to a heuristic control method.

10.2 Research Perspectives

Based on the research work in this thesis, several recommendations for future studies are presented here:

- Extracting large amounts of historical data of PV systems with various specifications and developing comprehensive models of PV power losses and power generation prediction which generalize over a wide range of PV systems;
- Accurate snow-cover and snow-removal events prediction and forecasting based on images taken from the surface of PV panels using a fixed camera;
- Ultra short-term PV power forecasting in snow conditions using computational intelligence techniques;
- Investigating the performance of convolutional and residual neural networks for PVPF in snow conditions based on Scaleogram graphs of wavelet transforms;
- Collaborative PV power forecasting in snow conditions for a group of PV installations in a wide geographical area;
- Implementing multi-stage (more than two) MPC-based EMS of microgrids by developing PV and load forecasting models for smaller prediction horizons to improve the performance of the EMS;
- Developing computational intelligence-based models to forecast the long-term performance and end-of-life of PV systems.

REFERENCES

[1] United States Environmental Protection Agency, “Sources of Greenhouse Gas Emissions,” 2019. [Online]. Available: https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions#colorbox-hidden.

[2] International Energy Agency, “Global Energy Review 2022,” 2022. [Online]. Available: https://www.iea.org/reports/global-energy-review-2022/renewables.

[3] International Renewable Energy Agency, “Renewable Capacity Statistics 2022,” 2022. [Online]. Available: https://www.irena.org/publications/2022/March/Renewable-Capacity-Statistics-2022.

[4] “Why did renewables become so cheap so fast?” Aug. 2021. [Online]. Available: https://ourworldindata.org/cheap-renewables-growth.

[5] National Renewable Energy Laboratory, “Photovoltaic Device Performance Calibration Services,”. [Online]. Available: https://www.nrel.gov/pv/pvdpc/.

[6] Natural Resources Canada, “Photovoltaic potential and solar resource maps of Canada,” 2020. [Online], Available: https://www.nrcan.gc.ca/our-natural-resources/energy-sources-distribution/renewable-energy/solar-photovoltaic-energy/tools-solar-photovoltaic-energy/photovoltaic-potential-and-solar-resource-maps-canada/18366.

[7] International Energy Agency, “IEA PV Snapshot 2019” May 2020. [Online]. Available: https://iea-pvps.org/.

[8] Natural Resources Canada, “Renewable energy facts,” [Online]. Available: https://www.nrcan.gc.ca/science-and-data/data-and-analysis/energy-data-and-analysis/energy-facts/renewable-energy-facts/20069.

[9] Canadian Renewable Energy Association, “Forecast: The future is bright for renewable energy in Canada,” Jan. 2021. [Online]. Available: https://renewablesassociation.ca/forecast-the-future-is-bright-for-renewable-energy-in-canada/.

[10] Environment and Natural Resources, Government of Canada, [Online]. Available: https://climate.weather.gc.ca/.

[11] MS. Adaramola and EET. Vågnes, “Preliminary assessment of a small-scale rooftop PVgrid tied in Norwegian climatic conditions,” *Energy Convers. Manag.*, vol. 90, pp. 458-465, Jan. 2015.

[12] B. Marion, R. Schaefer, H. Caine, and G. Sanchez, “Measured and modeled photovoltaic system energy losses from snow for Colorado and Wisconsin locations,” *Sol. Energy*, vol. 97, pp. 112-121, Nov. 2013.

[13] R. Ahmed, V. Sreeram, Y. Mishra, and M.D. Arif, “A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization,” *Renew. Sustain. Energy Rev.*, vol. 124, May 2020.

[14] T. U. Solanke, V. K. Ramachandaramurthy, J. Y. Yong, J. Pasupuleti, P. Kasinathan, and A. Rajagopalan, "A review of strategic charging–discharging control of grid-connected electric vehicles," *J. Energy Storage*, vol. 28, Apr. 2020.

[15] Q. Huang, X. Wang, J. Fan, S. Qi, W. Zhang, and C. Zhu, "V2G optimal scheduling of multiple EV aggregator based on TOU electricity price," in *2019 IEEE Int. Conf. Environ. Electr. Eng. - 2019 IEEE Ind. Commer. Power Syst. Europe*, June 2019, pp. 1-6.

[16] M. Alipour, B. Mohammadi-Ivatloo, M. Moradi-Dalvand, and K. Zare, "Stochastic scheduling of aggregators of plug-in electric vehicles for participation in energy and ancillary service markets," *Energy*, vol. 118, pp. 1168-1179, Jan. 2017.

[17] M. S. H. Nizami, M. J. Hossain, and K. Mahmud, "A coordinated electric vehicle management system for grid-support services in residential networks," *IEEE Syst. J.*, vol. 15, no. 2, pp. 2066-2077, June 2021.

[18] H. Shahinzadeh, J. Moradi, M. Longo, W. Yaïci, and S. Azani, "Integration of parking lot capacity in retail energy and reserve market mechanism," in *10th Int. Conf. Smart Grid (icSmartGrid)*, June 2022, pp. 275-279.

[19] L. Bagherzadeh, H. Shahinzadeh, and G. B. Gharehpetian, "Scheduling of distributed energy resources in active distribution networks considering combination of techno-economic and environmental objectives," in *2019 Int. Power Syst. Conf.*, Dec. 2019, pp. 687-695.

[20] S. Kerscher and P. Arboleya, "The key role of aggregators in the energy transition under the latest European regulatory framework," *Int. J. Electr. Power Energy Syst.*, vol. 134, Jan. 2022.

[21] D.E. Olivares, et al., "Trends in microgrid control," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1905–1919, May 2014.

[22] P. Sharma, H. D. Mathur, P. Mishra, and R. C. Bansal, "A critical and comparative review of energy management strategies for microgrids," *Appl. Energy*, vol. 327, Dec. 2022.

[23] P. Ingenhoven, G. Belluardo, G. Makrides, GE. Georghiou, P. Rodden, L. Frearson, et al., "Analysis of photovoltaic performance loss rates of six module types in five geographical locations," *IEEE J. Photovolt.*, vol. 9, no. 4, pp. 1091–1096, July 2019.

[24] B. Burduhos, I. Visa, A. Duta, and M. Neagoe, "Analysis of the conversion efficiency of five types of photovoltaic modules during high relative humidity time periods," *IEEE J. Photovolt.*, vol. 8, no. 6, pp. 1716–1724, Nov. 2018.

[25] LM. Ayompe, A. Duffy, SJ. McCormack, and M. Conlon, "Measured performance of a 1.72kW rooftop grid connected photovoltaic system in Ireland," *Energy Convers. Manag.*, vol. 52, pp. 816–825, Feb. 2011.

[26] DH. Daher, L. Gaillard, M. Amara, and C. Menezo, "Impact of tropical desert maritime climate on the performance of a PV grid-connected power plant," *Renew. Energy*, vol. 125, pp. 729–737, Sept. 2018.

[27] R. Sharma and S. Goel, "Performance analysis of a 11.2 kWp roof top grid-connected PV

system in Eastern India," *Energy Rep.*, vol. 3, pp. 76–84, Nov. 2017.

[28] SM. Pietruszko and M. Gradzki, "Performance of a grid connected small PV system in Poland," *Appl. Energy*, vol. 74, pp. 177–184, Jan. 2003.

[29] T. Ma, H. Yang, and L. Lu, "Long term performance analysis of a standalone photovoltaic system under real conditions," *Appl. Energy*, vol. 201, pp. 320–331, Sept. 2017.

[30] D. Okello, EE. van Dyk, and FJ. Vorster, "Analysis of measured and simulated performance data of a 3.2kWp grid-connected PV system in Port Elizabeth, South Africa," *Energy Convers. Manag.*, vol. 100, pp. 10–15, Aug. 2015.

[31] NM. Kumar, RP. Gupta, M. Mathew, A. Jayakumar, and NK. Singh, "Performance, energy loss, and degradation prediction of roof-integrated crystalline solar PV system installed in Northern India," *Case Stud. Therm. Eng.*, vol. 13, March 2019.

[32] B. Quesada, C. Sanchez, J. Canada, R. Royo, and J. Paya, "Experimental results and simulation with TRNSYS of a 7.2kWp grid-connected photovoltaic system," *Appl. Energy*, vol. 88, pp. 1772–1783, May 2011.

[33] M. Malvoni, A. Leggieri, G. Maggiotto, PM. Congedo, and MG. De Giorgi, "Long term performance, losses and efficiency analysis of a 960kWP photovoltaic system in the Mediterranean climate," *Energy Convers. Manag.*, vol. 145, pp. 169–181, Aug. 2017.

[34] S. Goel and R. Sharma, "Analysis of measured and simulated performance of a grid connected PV system in eastern India," *Environ. Dev. Sustain.*, vol. 23, pp. 451–476, Jan. 2020.

[35] JW. Zapata, MA. Perez, S. Kouro, A. Lensu, and A. Suuronen, "Design of a cleaning program for a PV plant based on analysis of energy losses," *IEEE J. Photovolt.*, vol. 5, no. 6, pp. 1748–1756, Nov. 2015.

[36] M. Coello and L. Boyle, "Simple model for predicting time series soiling of photovoltaic panels," *IEEE J. Photovolt.*, vol. 9, no. 5, pp. 1382–1387, Sept. 2019.

[37] F. Liu, Z. Zhang, Y. Zhao, Z. Zhu, W. Pan, L. Wang, et al., "A method of calculating the daily output power reduction of PV modules due to dust deposition on its surface," *IEEE J. Photovolt.*, vol. 9, no. 3, pp. 881–887, May 2019.

[38] MG. Deceglie, L. Micheli, and M. Muller, "Quantifying soiling loss directly from PV yield," *IEEE J. Photovolt.*, vol. 8, no. 2, pp. 547–551, March 2018.

[39] KD. Malamaki and CS. Demoulias, "Analytical calculation of the electrical energy losses on fixed-mounted PV plants," *IEEE Trans. Sustain. Energy*, vol. 5, no. 4, pp. 1080–1089, Oct. 2014.

[40] T. Oozeki, T. Izawa, K. Otani, and K. Kurokawa, "An evaluation method of PV systems," *Sol. Energy Mater. Sol. Cells*, vol. 75, pp. 687–695, Feb. 2003.

[41] Y. Ueda, K. Kurokawa, K. Kitamura, M. Yokota, K. Akanuma, and H. Sugihara, "Performance analysis of various system configurations on grid-connected residential PV systems," *Sol. Energy Mater. Sol. Cells*, vol. 93, pp. 945–949, June 2009.

[42] L. Micheli and M. Muller, "An investigation of the key parameters for predicting PV soiling losses," *Prog. Photovolt.: Res. Appl.*, vol. 25, pp. 291–307, Jan. 2017.

[43] L. Abenante, F. De Lia, R. Schioppo, and S. Castello, "Non-linear continuous analytical model for performance degradation of photovoltaic module arrays as a function of exposure time," *Appl. Energy*, vol. 275, Oct. 2020.

[44] E. Adiyasuren, U. Damba, and B. Tsedensodnom, "Comparison of power generation from solar panel with various climate condition and selection of best tilt angles in Ulaanbaatar," in *Proc. Int. Forum Strategic Technol.*, Ulaanbaatar, Mongolia, 2013, pp. 519-521.

[45] L. Powers, J. Newmiller, and T. Townsend, "Measuring and modeling the effect of snow on photovoltaic system performance," in *Proc. 35th IEEE Photovolt. Spec. Conf.*, Honolulu, HI, USA, 2010, pp. 973–978.

[46] N. Heidari, J. Gwamuri, T. Townsend, and J. M. Pearce, "Impact of snow and ground interference on photovoltaic electric system performance," *IEEE J. Photovolt.*, vol. 5, no. 6, pp. 1680–1685, Nov. 2015.

[47] T. Townsend and L. Powers, "Photovoltaics and snow: An update from two winters of measurements in the SIERRA," in *Proc. 37th IEEE Photovolt. Spec. Conf.*, Seattle, WA, USA, 2011, pp. 3231–3236.

[48] R. W. Andrews and J. M. Pearce, "Prediction of energy effects on photovoltaic systems due to snowfall events," in *Proc. 38th IEEE Photovolt. Spec. Conf.*, Austin, TX, USA, 2012, pp. 3386–3391.

[49] B. Meyers, M. Mikofski, and M. Anderson, "A fast parameterized model for predicting PV system performance under partial shade conditions," in *IEEE 43rd Photovolt. Specialists Conf. (PVSC)*, Portland, OR, USA, 2016, pp. 3173–3178.

[50] W. Javed, B. Guo, and B. Figgis, "Modeling of photovoltaic soiling loss as a function of environmental variables," *Sol. Energy*, vol. 157, pp. 397–407, Nov. 2017.

[51] S. Pulipaka, F. Mani, and R. Kumar, "Modeling of soiled PV module with neural networks and regression using particle size composition," *Sol. Energy*, vol. 123, pp. 116–126, Jan. 2016.

[52] A. Massi Pavan, A. Mellit, D. De Pieri, and SA. Kalogirou, "A comparison between BNN and regression polynomial methods for the evaluation of the effect of soiling in large scale photovoltaic plants," *Appl. Energy*, vol. 108, pp. 392–401, 2013.

[53] R. Blaga, et al., "A current perspective on the accuracy of incoming solar energy forecasting," *Prog. Energy Combustion Sci.*, vol. 70, pp. 119–144, Jan. 2019.

[54] R. E. Pawluk, Y. Chen, and Y. She, "Photovoltaic electricity generation loss due to snow—A literature review on influence factors, estimation, and mitigation," *Renewable Sustain. Energy Rev.*, vol. 107, pp. 171–182, Jun. 2019.

[55] M. Monfared, M. Fazeli, R. Lewis, and J. Searle, "Fuzzy predictor with additive learning for very short-term PV power generation," *IEEE Access*, vol. 7, pp. 91183-91192, July 2019.

[56] N. Son and M. Jung, "Analysis of meteorological factor multivariate models for medium- and long-term photovoltaic solar power forecasting using long short-term memory," *Appl. Sci.*, vol. 11, no. 1, Jan. 2021.

[57] S. Rajabalizadeh and S. M. M. Tafreshi, "A practicable copula-based approach for power forecasting of small-scale photovoltaic systems," *IEEE Syst. J.*, vol. 14, no. 4, pp. 4911-4918, Dec. 2020.

[58] X. Meng, F. Gao, T. Xu, K. Zhou, W. Li, and Q. Wu, "Inverter-data-driven second-level power forecasting for photovoltaic power plant," *IEEE Trans. Ind. Electron.*, vol. 68, no. 8, pp. 7034-7044, Aug. 2021.

[59] G. G. Kim et al., "Prediction model for PV performance with correlation analysis of environmental variables," *IEEE J. Photovolt.*, vol. 9, no. 3, pp. 832-841, May 2019.

[60] S. Choi and J. Hur, "An ensemble learner-based bagging model using past output data for photovoltaic forecasting," *Energies*, vol. 13, 2020.

[61] K. Mahmud, S. Azam, A. Karim, S. Zobaed, B. Shanmugam, and D. Mathur, "Machine learning based PV power generation forecasting in Alice Springs," *IEEE Access*, vol. 9, pp. 46117-46128, March 2021.

[62] A. Mellit, A. Massi Pavan, and V. Lughi, "Deep learning neural networks for short-term photovoltaic power forecasting," *Renew. Energy*, vol. 172, pp. 276-288, July 2021.

[63] P. Li, K. Zhou, X. Lu, and S. Yang, "A hybrid deep learning model for short-term PV power forecasting," *Appl. Energy*, vol. 259, Feb. 2020.

[64] J. Zhang, Z. Tan, and Y. Wei, "An adaptive hybrid model for day-ahead photovoltaic output power prediction," *J. Clean. Prod.*, vol. 244, 2020.

[65] W. Liu and Y. Xu, "Randomised learning-based hybrid ensemble model for probabilistic forecasting of PV power generation," *IET Gener. Transm. Distrib.*, vol. 14, no. 24, pp. 5909-5917, Dec. 2020.

[66] Y. Zhou, N. Zhou, L. Gong, and M. Jiang, "Prediction of photovoltaic power output based on similar day analysis, genetic algorithm and extreme learning machine," *Energy*, vol. 204, Aug. 2020.

[67] M. Pan, C. Li, R. Gao, Y. Huang, H. You, T. Gu, and F. Qin, "Photovoltaic power forecasting based on a support vector machine with improved ant colony optimization," *J. Clean. Prod.*, vol. 277, Dec. 2020.

[68] L. B. Bosman and S. B. Darling, "Performance modeling and valuation of snow-covered PV systems: Examination of a simplified approach to decrease forecasting error," *Environ. Sci. Pollut. Res.*, vol. 25, no. 16, pp. 15484–15491, Jun. 2018.

[69] E. Lorenz, D. Heinemann, and C. Kurz, "Local and regional photovoltaic power prediction for large scale grid integration: Assessment of a new algorithm for snow detection," *Prog. Photovolt., Res. Appl.*, vol. 20, no. 6, pp. 760–769, Sep. 2012.

[70] R. E. Pawluk, M. Rezvanpour, Y. Chen, and Y. She, "A sensitivity analysis on effective

parameters for sliding/melting prediction of snow cover on solar photovoltaic panels," *Cold Reg. Sci. Technol.*, vol. 185, May 2021.

[71] M. B. Øgaard, B. L. Aarseth, Å. F. Skomedal, H. N. Riise, S. Sartori, and J. H. Selj, "Identifying snow in photovoltaic monitoring data for improved snow loss modeling and snow detection," *Sol. Energy*, vol. 223, pp. 238-247, July 2021.

[72] S. Iyengar, N. Sharma, D. Irwin, P. Shenoy, and K. Ramamritham, "A cloud-based black-box solar predictor for smart homes," *ACM Trans. Cyber-Phys. Syst.*, vol. 1, no. 4, pp. 1-24, Oct. 2017.

[73] E. Zhu and D. Pi, "Photovoltaic generation prediction of CCIPCA combined with LSTM," *Complexity*, vol. 2020, Sep. 2020.

[74] H. Awad, M. Gül, K. M. E. Salim, and H. Yu "Predicting the energy production by solar photovoltaic systems in cold-climate regions," *Int. J. Sustain. Energy*, vol. 37, no. 10, pp. 978-998, Nov. 2018.

[75] D. Koster, F. Minette, C. Braun, and O. O'Nagy, "Short-term and regionalized photovoltaic power forecasting, enhanced by reference systems, on the example of Luxembourg," *Renew. Energy*, vol. 132, pp. 455-470, Mar. 2019.

[76] S. Pelland, G. Galanis, and G. Kallos, "Solar and photovoltaic forecasting through post-processing of the Global Environmental Multiscale numerical weather prediction model," *Prog. Photovolt: Res. Appl.*, vol. 21, no. 3, pp. 284-296, May 2013.

[77] A. Mellit, A. Massi Pavan, E. Ogliari, S. Leva, and V. Lughi, "Advanced methods for photovoltaic output power forecasting: A review," *Appl. Sci.*, vol. 10, no. 2, Jan. 2020.

[78] H. Wang, Y. Liu, B. Zhou, C. Li, G. Cao, N. Voropai, and E. Barakhtenko, "Taxonomy research of artificial intelligence for deterministic solar power forecasting," *Energy Convers. Manag.*, vol. 214, June 2020.

[79] W. VanDeventer, E. Jamei, G. S. Thirunavukkarasu, M. Seyedmahmoudian, T. K. Soon, B. Horan, S. Mekhilef, and A. Stojcevski, "Short-term PV power forecasting using hybrid GASVM technique," *Renew. Energy*, vol. 140, Sept. 2019.

[80] F. Wang, Z. Xuan, Z. Zhen, K. Li, T. Wang, and M. Shi, "A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework," *Energy Convers. Manag.*, vol. 212, May 2020.

[81] A. Niccolai, A. Dolara, and E. Ogliari, "Hybrid PV power forecasting methods: A comparison of different approaches," *Energies*, vol. 14, no. 2, Jan. 2021.

[82] M. Kumar Behera and N. Nayak, "A comparative study on short-term PV power forecasting using decomposition based optimized extreme learning machine algorithm," *Eng. Sci. Technol. an Int. J.*, vol. 23, no. 1, pp. 156-167, Feb. 2020.

[83] F. Succetti, A. Rosato, R. Araneo, and M. Panella, "Deep neural networks for multivariate prediction of photovoltaic power time series," *IEEE Access*, vol. 8, pp. 211490-211505, Nov. 2020.

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

[85] A. J. Conejo, M. Carrion, and J. M. Morales, *Decision Making Under Uncertainty in Electricity Markets*, vol. 1, New Year, Springer, 2010.

[86] R. Habibifar, A. A. Lekvan, and M. Ehsan, "A risk-constrained decision support tool for EV aggregators participating in energy and frequency regulation markets," *Electr. Power Syst. Res.*, vol. 185, Aug. 2020.

[87] Z. Liu, Q. Wu, K. Ma, M. Shahidehpour, Y. Xue, and S. Huang, "Two-stage optimal scheduling of electric vehicle charging based on transactive control," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2948-2958, 2019.

[88] M. K. Daryabari, R. Keypour, and H. Golmohamadi, "Stochastic energy management of responsive plug-in electric vehicles characterizing parking lot aggregators," *Appl. Energy*, vol. 279, Dec. 2020.

[89] Y. Vardanyan and H. Madsen, "Optimal coordinated bidding of a profit maximizing, risk-averse EV aggregator in three-settlement markets under uncertainty," *Energies*, vol. 12, no. 9, May 2019.

[90] M. Shafie-Khah, P. Siano, D. Z. Fitiwi, N. Mahmoudi, and J. P. S. Catalão, "An innovative two-level model for electric vehicle parking lots in distribution systems with renewable energy," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 1506-1520, Mar. 2018.

[91] B. Hashemi, M. Shahabi, and P. Teimourzadeh-Baboli, "Stochastic-based optimal charging strategy for plug-in electric vehicles aggregator under incentive and regulatory policies of DSO," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3234-3245, Apr. 2019.

[92] M.W. Tian, S.R. Yan, X.X. Tian, M. Kazemi, S. Nojavan, and K. Jermsittiparsert, "Risk-involved stochastic scheduling of plug-in electric vehicles aggregator in day-ahead and reserve markets using downside risk constraints method," *Sustain. Cities Soc.*, vol. 55, April 2020.

[93] M. Amjad, A. Ahmad, M. H. Rehmani, and T. Umer, "A review of EVs charging: From the perspective of energy optimization, optimization approaches, and charging techniques," *Transp. Res. D Transp. Environ.*, vol. 62, pp. 386-417, Jul. 2018.

[94] A. Arias-Londoño, O. D. Montoya, and L. F. Grisales-Noreña, "A chronological literature review of electric vehicle interactions with power distribution systems," *Energies*, vol. 13, no. 11, Jan. 2020.

[95] N. Bañol Arias, S. Hashemi, P. B. Andersen, C. Træholt, and R. Romero, "Distribution system services provided by electric vehicles: Recent status, challenges, and future prospects," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 12, pp. 4277-4296, Dec. 2019.

[96] M. González Vayá and G. Andersson, "Self scheduling of plug-in electric vehicle aggregator to provide balancing services for wind power," *IEEE Trans. Sustain. Energy*, vol. 7, no. 2, pp. 886-899, Apr. 2016.

[97] M. H. Abbasi, M. Taki, A. Rajabi, L. Li, and J. Zhang, "Coordinated operation of electric vehicle charging and wind power generation as a virtual power plant: A multi-stage risk constrained approach," *Appl. Energy*, vol. 239, pp. 1294-1307, Apr. 2019.

[98] P. Afzali, M. Rashidinejad, A. Abdollahi, and A. Bakhshai, "Riskconstrained bidding strategy for demand response, green energy resources, and plug-in electric vehicle in a flexible smart grid," *IEEE Syst. J.*, pp. 1-8, Mar. 2020.

[99] H. Shahinzadeh, J. Moradi, Z. Pourmirza, E. Kabalci, M. Benbouzid, and S. M. Muyeen, "Optimal operation of distributed flexible generation sources incorporating VPP framework in market environment considering uncertainties, " in *2022 IEEE Kansas Power Energy Conf.*, April 2022, pp. 1-5.

[100] J. Kim, S. W. Kim, Y. G. Jin, J.-K. Park, and Y. T. Yoon, "Optimal coordinated management of a plug-in electric vehicle charging station under a flexible penalty contract for voltage security," *Energies*, vol. 9, no. 7, pp. 538–552, Jul. 2016.

[101] Y. Liu and H. Liang, "A discounted stochastic multiplayer game approach for vehicle-to-grid voltage regulation," *IEEE Trans. Veh. Technol.*, vol. 68, no. 10, pp. 9647-9659, Oct. 2019.

[102] B. S. K. Patnam and N. M. Pindoriya, "DLMP calculation and congestion minimization with EV aggregator loading in a distribution network using bilevel program," *IEEE Syst. J.*, vol. 15, no. 2, pp. 1835-1846, June 2021.

[103] A. Tazikeh Lemeski, R. Ebrahimi, and A. Zakariazadeh, "Optimal decentralized coordinated operation of electric vehicle aggregators enabling vehicle to grid option using distributed algorithm," *J. Energy Storage*, vol. 54, Oct. 2022.

[104] H. Ren, A. Zhang, F. Wang, X. Yan, Y. Li, N. Duić, M. Shafie-khah, and J. P.S. Catalão, "Optimal scheduling of an EV aggregator for demand response considering triple level benefits of three-parties," *Int. J. Electr. Power Energy Syst.*, vol. 125, Feb. 2021.

[105] W. Gu, Z. Wu, and X. Yuan, "Microgrid economic optimal operation of the combined heat and power system with renewable energy," in *IEEE PES General Meeting*, Minneapolis, MN, USA, 2010, pp. 1-6.

[106] D. Giaouris, et al., "Performance investigation of a hybrid renewable power generation and storage system using systemic power management models," *Energy*, vol. 61, pp. 621–635, Nov. 2013.

[107] D. Ipsakis, S. Voutetakis, P. Seferlis, F. Stergiopoulos, S. Papadopoulou, and C. Elmasides, "The effect of the hysteresis band on power management strategies in a stand-alone power system," *Energy*, vol. 33, no. 10, pp. 1537–1550, Oct. 2008.

[108] K. Rouzbehi, A. Miranian, JI. Candela, A. Luna, and P. Rodriguez, "Intelligent voltage control in a DC micro-grid containing PV generation and energy storage," in *IEEE PES T&D conf. expos.*, Chicago, IL, USA, April 2014, pp. 1–5.

[109] C. Ziogou, D. Ipsakis, C. Elmasides, F. Stergiopoulos, S. Papadopoulou, P. Seferlis, and S. Voutetakis, "Automation infrastructure and operation control strategy in a stand-alone power system based on renewable energy sources," *J. Power Sour.*, vol. 196, no. 22, pp.

9488-9499, Nov. 2011.

[110] F. Garcia-Torres, A. Zafra-Cabeza, C. Silva, S. Grieu, T. Darure, and A. Estanqueiro, "Model predictive control for microgrid functionalities: Review and future challenges," *Energies*, vol. 14, no. 5, Feb. 2021.

[111] J. Hu, Y. Shan, J. M. Guerrero, A. Ioinovici, K. W. Chan, and J. Rodriguez, "Model predictive control of microgrids – An overview," *Renew. Sust. Energ. Rev.*, vol. 136, Feb. 2021.

[112] U. R. Nair and R. Costa-Castelló, "A model predictive control-based energy management scheme for hybrid storage system in islanded microgrids," *IEEE Access*, vol. 8, pp. 97809-97822, May 2020.

[113] D. P. e Silva, J. L. Félix Salles, J. F. Fardin, and M. M. Rocha Pereira, "Management of an island and grid-connected microgrid using hybrid economic model predictive control with weather data," *Appl. Energy*, vol. 278, Nov. 2020.

[114] Y. Wang, W. Dong, and Q. Yang, "Multi-stage optimal energy management of multi-energy microgrid in deregulated electricity markets," *Appl. Energy*, vol. 310, March 2022.

[115] E. Rezaei and H. Dagdougui, "Optimal real-time energy management in apartment building integrating microgrid with multizone HVAC control," *IEEE Trans. Industr. Inform.*, vol. 16, no. 11, pp. 6848-6856, Nov. 2020.

[116] F. K/bidi, C. Damour, D. Grondin, M. Hilairet, and M. Benne, "Multistage power and energy management strategy for hybrid microgrid with photovoltaic production and hydrogen storage," *Appl. Energy*, vol. 323, Oct. 2022.

[117] K-S. Ryu, D-J. Kim, H. Ko, C-J. Boo, J. Kim, Y-G. Jin, and H-C. Kim, "MPC based energy management system for hosting capacity of PVs and customer load with EV in stand-alone microgrids," *Energies*, vol. 14, no. 13, July 2021.

[118] D. Y. Yamashita, I. Vechiu, and J.P. Gaubert, "Two-level hierarchical model predictive control with an optimised cost function for energy management in building microgrids," *Appl. Energy*, vol. 285, March 2021.

[119] J. Cheng, D. Duan, X. Cheng, L. Yang, and S. Cui, "Probabilistic microgrid energy management with interval predictions," *Energies*, vol. 13, no. 12, June 2020.

[120] N. Ye, *Data Mining: Theories, Algorithms, and Examples*. Boca Raton, FL, USA: CRC Press, 2013.

[121] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Netherlands: Elsevier Science, 2003.

[122] B. Boehmke and B. M. Greenwell, *Hands-on Machine Learning With R*. Boca Raton, FL, USA: CRC Press, 2019.

[123] U.K. Das, K.S. Tey, M. Seyedmahmoudian, M.Y. Idna Idris, S. Mekhilef, B. Horan, and A. Stojcevski, "SVR-based model to forecast PV power generation under different weather conditions," *Energies*, vol. 10, no. 7, pp. 876-893, Jul. 2017.

[124] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[125] Environment and Climate Change Canada, "Canadian climate normals 1981–2010," Oct. 2019. [Online]. Available: https://climate.weather.gc.ca/climate_normals/index_e.html.

[126] A. P. Dobos, "PVWatts version 5 manual," *Nat. Renewable Energy Lab.*, Golden, CO, USA, Tech. Rep. NREL/TP-6A20-62641, Sep. 2014.

[127] "Photovoltaic data acquisition," [online] Available: http://maps.nrel.gov/pvdaq/.

[128] U.S. Local Climatological Data, "National Ocean and Atmospheric Administration (NOAA)," [online] Available: http://www.ncdc.noaa.gov/.

[129] S. Hosseini, S. Taheri, M. Farzaneh, H. Taheri, and M. Narimani, "Determination of photovoltaic characteristics in real field conditions," *IEEE J. Photovolt.*, vol. 8, no. 2, pp. 572–580, Mar. 2018.

[130] "Shell PowerMax Eclipse 80-c," [online] Available: https://www.shell.com.

[131] M. Khenar, S. Hosseini, S. Taheri, A. Cretu, E. Pouresmaeil, and H. Taheri, "Particle swarm optimisation-based model and analysis of photovoltaic module characteristics in snowy conditions," *IET Renew. Power Gener.*, vol. 13, no. 11, pp. 1950–1957, Aug. 2019.

[132] A. H. M. E. Reinders, V. A. P. van Dijk, E. Wiemken, and W. C. Turkenburg, "Technical and economic analysis of grid-connected PV systems by means of simulation," *Prog. Photovolt.: Res. Appl.*, vol. 7, no. 1, pp. 71–82, Jan. 1999.

[133] C. W. A. Baltus, J. A. Eikelboom, and R. J. C. Van Zolingen, "Analytical monitoring of losses in PV systems," in *14th European Photovolt. Sol. Energy Conf.*, Barcelona, Spain, 1997, pp. 1-5.

[134] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, "A review of deep learning for renewable energy forecasting," *Energy Convers. Manag.*, vol. 198, Oct. 2019.

[135] SolarEdge Inc., [Online]. Available: https://www.solaredge.com/.

[136] National Solar Radiation Database, National Renewable Energy Laboratory, [Online]. Available: https://nsrdb.nrel.gov/.

[137] Canada Weather Stats, Environment and Climate Change Canada, [Online]. Available: https://www.weatherstats.ca/.

[138] W. F. Holmgren, C. W. Hansen, and M. A. Mikofski, "pvlib python: A python package for modeling solar energy systems," *J. Open Source Softw.*, vol. 3, no. 29, pp. 884-886, Sep. 2018.

[139] A. Safdarian, M. Fotuhi-Firuzabad, and M. Lehtonen, "Integration of price-based demand response in DisCos' short-term decision model," *IEEE Trans. Smart Grid*, vol. 5, no. 5, pp. 2235–2245, Sep. 2014.

[140] New York Independent System Operator Inc., Rensselaer, NY, USA. [Online]. Available:

http://www.nyiso.com.

[141] Dayton Power and Light Company, Dayton, OH, USA. [Online]. Available: https://www.dpandlpowerauctions.com/LoadData.aspx.

[142] W. H. Kersting, "Radial distribution test feeders," in *Proc. IEEE Power Eng. Soc. Winter Meeting*, Columbus, OH, USA, 2001, pp. 908–912.

[143] Transportation Research Center at Argonne National Laboratory, [online] Available: https://www.anl.gov/es/light-duty-electric-drive-vehiclesmonthly-sales-updates.

[144] "Battery Pack Prices Cited Below $100/kWh for the First Time in 2020, While Market Average Sits at $137/kWh," [Online], Available: https://about.bnef.com/blog/battery-pack-prices-cited-below-100-kwhfor-the-first-time-in-2020-while-market-average-sits-at-137-kwh/.

[145] C. Bordons, F. Garcia-Torres, and M.A. Ridao, *Model predictive control of microgrids*. vol. 358, Springer, 2020.

[146] EF. Camacho and CB. Alba, *Model predictive control*. Springer science & business media, Jan. 2013.

[147] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proc. 22nd acm sigkdd int. conf. knowledge discovery data mining*, San Francisco, CA, USA, Aug. 2016, pp. 785-794.