2019

Intra-Day Solar Irradiance Forecasting for Remote Microgrids Using Hidden Markov Model

Abhilasha Bajracharya
South Dakota State University

INTRA-DAY SOLAR IRRADIANCE FORECASTING FOR REMOTE MICROGRIDS USING HIDDEN MARKOV MODEL

BY

ABHILASHA BAJRACHARYA

A thesis submitted in partial fulfillment of the requirements for the

Master of Science

Major in Electrical Engineering

South Dakota State University

2019

INTRA-DAY SOLAR IRRADIANCE FORECASTING FOR REMOTE MICROGRIDS USING HIDDEN MARKOV MODEL

This thesis is approved as a creditable and independent investigation by a candidate for the Master of Science in Electrical Engineering degree and is acceptable for meeting the thesis requirements for this degree. Acceptance of this thesis does not imply that the conclusions reached by the candidates are necessarily the conclusions of the major department.

Reinaldo Tonoski Jr., Ph.D.
Thesis Advisor Date

George Hamer, Ph.D.
Head, Electrical Engineering and Computer Science Date

Dean, Graduate School Date

ACKNOWLEDGEMENTS

I would like to express my immense gratitude to my research advisor, Dr. Reinaldo Tonkoski for allowing me to be a part of his research group and for giving me an opportunity to work in this project. His patient guidance, useful critiques, and unceasing encouragement in the most struggling times are much appreciated. I would also like to offer my special thanks to my co-advisor, Dr. Semhar Michael for her expert advice and guidance that helped me grasp the understanding of the immense knowledge of statistics required to complete this work. I would also like to thank Dr. Timothy M. Hansen for providing me with his invaluable insights that proved to be extremely helpful.

I am eternally grateful to Ujjwol Tamrakar for guiding me throughout the completion of this work. His assistance in helping me learn the research methods and to organize my work is greatly appreciated. I am also grateful towards André Luna for sharing his knowledge during our research meetings, which helped me discover valuable ideas for this thesis. I am particularly grateful towards Mahesh Shrestha for his endless support and motivation. I would also like to thank him for taking his time to look at my work and provide suggestions that helped me view my work from a different dimension.

I also wish to thank Ayush Shakya for providing me with the foundation of my thesis work and Riaz A. Khan for helping me understand the fundamentals of statistics required to move forward with this work. My grateful thanks are also extended to my friends Manisha, Prajina, Jharna, and Jyotshna for their never-ending support and encouragement.

Finally, I am very thankful for my family without whom none of this would have been possible. I would like to thank them for being the prime source of my inspiration.

CONTENTS

ABBREVIATIONS

AC	Alternating Current
AM	Air Mass
BIC	Bayesian Information Criteria
CSI	Clear-Sky Irradiance
DC	Direct Current
DERs	Distributed Energy Resources
DOE	Department of Energy
EM	Expectation Maximization
EMS	Energy Management System
ESS	Energy Storage System
GHI	Global Horizontal Irradiance
HMM	Hidden Markov Model
IEA	International Energy Agency
MAPE	Mean Absolute Percentage Error
MSM	Markov Switching Model
PV	Photovoltaic
RMSE	Root Mean Square Error
SSE	Sum of Squared Errors

LIST OF FIGURES

ABSTRACT

INTRA-DAY SOLAR IRRADIANCE FORECASTING FOR REMOTE MICROGRIDS USING HIDDEN MARKOV MODEL

ABHILASHA BAJRACHARYA

2019

Accurate solar irradiance forecasting is the key to accurate estimation of solar power output at any given time. The accuracy of this information is especially crucial in diesel-PV based remote microgrids with batteries to determine the set points of the batteries and generators for their optimal dispatch. This, in turn, is related directly to the overall operating cost because both an overestimation and an underestimation of the irradiance means additional operating costs for either suddenly ramping up the backup resources or causing under-utilization of the available PV power output. Accurately predicting the solar irradiance is not an easy task because of the sporadic nature of the irradiance that is received at the solar panel surfaces. Handling the dynamic nature of the irradiance pattern requires a strong and flexible model that can precisely capture the irradiance trend in any given location at a given time. Usually, such a robust model requires a lot of input variables like weather data including humidity, temperature, pressure, wind speed, wind direction, etc. and/or large inventory of satellite images of clouds over a long period of time. The expensive sensors and database tools for collecting and storing such huge information may not be installed in remote locations. Therefore, this thesis prioritizes on developing a simple method requiring a minimum input to accurately forecast the solar irradiance for remote microgrids.

Essentially, this thesis is an extension of the work by Shakya et al.[1], which is the implementation of day-ahead solar irradiance forecasting using a Markov switching model. In this work, a hidden Markov model (also known as the Markov switching model) is developed using the past irradiance data, clear-sky irradiance, and Fourier basis functions to generate three energy states: low, medium, and high. Each of the three states corresponds to a different cloud cover conditions and thereby a different level of irradiance that is collected at the solar panels. In this thesis, four different methods are described to select the closest state to the actual irradiance in every hour. In other words, four different intra-day forecasting methods are proposed to forecast the 1-hour ahead irradiance. Updating the forecast every hour greatly reduces the forecast error compared to the beforehand mentioned day-ahead forecast. The proposed methods are named as Past4Hr method, Slope1 method, Slope2 method, and PastHr method. A case study for Brookings, South Dakota, is considered to both train and validate the described methods. Additionally, the performance of each of the methods is compared on the basis of the time of the forecast, the time of the year, and at different years to assess their consistencies. The simulations results show that the performance of the Past4Hr method and the PastHr method outperforms the remaining two.

CHAPTER 1 INTRODUCTION

1.1 Background

Around the world, an estimated of 1.1 billion people do not have access to basic electricity service, according to a report by the International Energy Agency (IEA) [2]. A greater number of this population reside in the developing countries, in remote areas with high poverty rates. The inaccessible topography and poor economy make the grid extension to such locations impractical and almost impossible. In such cases, separate isolated grids known as the remote microgrids have been used to power such communities.

The U.S. Department of Energy (DOE) defines microgrid as "a group of interconnected loads and distributed energy resources within clearly defined electrical boundaries that acts as a single controllable entity with respect to the grid. A microgrid can connect and disconnect from the grid to enable it to operate in both grid-connected or island mode [3]." Remote microgrids, on the other hand, are special types of microgrids that are isolated and are not connected to the grid at all. They are small-scale power supply systems that are aimed to power small communities, which may be a number of households, rural communities, and commercial or industrial sites. Presently, many remote communities in the world are powered by such remote microgrids that are primarily dependent on diesel for generating electricity. For instance, out of 292 remote communities (mostly indigenous) in Canada, about 86% of them are predominantly based on diesel generators [4]. These communities collectively consume 90 million to 120 million liters of diesel per year, which is often shipped or even flown in from long distances, only to last for a few months. Relying so heavily on shipments of fossil fuels is

extremely expensive, susceptible to fuel shortages, adds to greenhouse gas emissions, and increases risks of environmental and public health hazards from air pollution and storage tank leaks. To alleviate this heavy reliance on fuel, several studies on integrating renewable energy sources have been made, particularly utilizing solar energy because of its clean nature and abundance.

Most commonly, diesel-based microgrids with photovoltaic (PV) integration have energy storage systems (ESS) such as a battery bank, incorporated into the system to overcome issues with PV integration regarding efficiency and reliability, as in [5]–[7]. However, the high initial cost, limited life cycle, and difficulty in frequent transportation to remote locations pose additional challenges if not used optimally. Therefore, an energy management system (EMS) is required to optimally dispatch the distributed energy resources (DERs) like batteries and generators and provide an effective means to meet the load requirements. Usually, the EMS of such renewable-integrated microgrids consists of two layers – a day-ahead scheduling layer and a real-time dispatch layer, which schedules and controls the operation of the diesel generators and the battery. One example of such two-layered EMS as shown in Figure 1.1 is presented in [8]. Typically, the scheduling module needs to calculate the power set points of the dispatchable generators and the battery, 24 hours ahead, to achieve the minimum cost of operation. From the set points obtained, the real-time dispatch module dispatches the resources accordingly. In this thesis, we are focused primarily on the scheduling module. This scheduling should be done based on the availability of the dispatchable resources, PV power forecast, and the load demand forecast. Thus, forecasting of PV power output for the next day is a crucial task in the EMS of a PV-integrated remote microgrid. The PV power output is a function

of the total amount of shortwave radiation received at a surface horizontal to the ground, called the global horizontal irradiance (GHI). Therefore, the fundamental step in PV power prediction, as stated in [9], is to forecast the GHI. From this point onwards, forecasting of the GHI will simply be referred to as "solar forecasting".

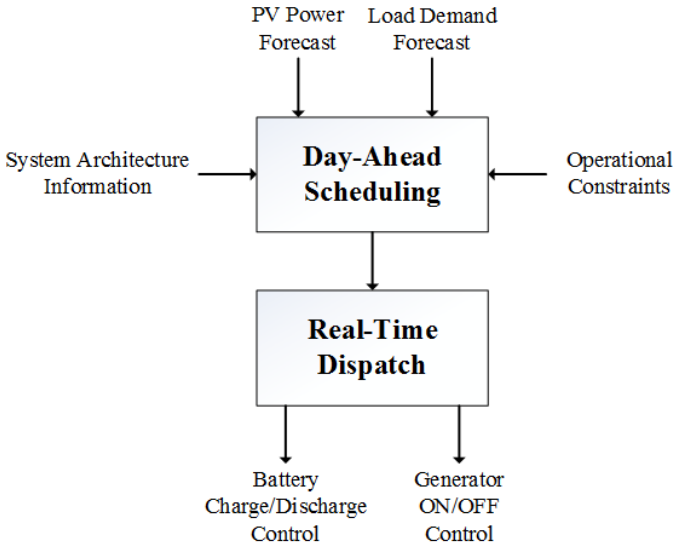

Figure 1.1. Energy management system of PV-integrated microgrid.

Accurate solar forecasting is particularly required for the EMS of remote microgrids to estimate the solar power output to allocate the resources better and to avoid unnecessary costs. If there is an overestimation in the amount of solar power output, then the backup generators need to be ramped up on short notice. This means additional operating costs. Likewise, when there is an underestimation, full potential of solar power output cannot be utilized. Accurate forecasting of PV power output has shown to improve the PV power utilization and thus result in savings in operational costs. One such instance is shown in study [10] where a deep learning method was used for solar power and load forecasting for optimal load dispatch of a community microgrid. In this case, a deep recurrent neural network with long short-term memory units (DRNN-LSTM) model developed to forecast

the PV power output has shown to reduce the daily costs by 8.97 %. Therefore, an accurate prediction of the solar irradiance and hence the solar power output is the key to reducing the operational costs of microgrids by proper scheduling of the dispatchable resources.

The state-of-the-art in solar forecasting requires heavy information of meteorological data and satellite images, which may not be readily available in remote locations. A novel solar forecasting technique based on Markov switching model (MSM) has been proposed in [11]. This model only uses that past irradiance data from local measurements for day-ahead forecasting. The past measurements are required only for fitting the model once. Then after, any other past data (except the current day irradiance measurements) are not required for forecasting. Thus, this method is suitable for remote microgrids without prior instrumentation for weather data measurements. This thesis is an extension of the aforementioned Markov-based model, which explores several methods of intra-day forecasting to improve the forecasting accuracy. The first method simply adds a moving window to the work in [11] to update the hourly forecast. The second two methods use the derivative of the irradiance values in the past two hours to detect the hourly rise or fall in solar irradiance using two separate algorithms. Furthermore, the fourth method simply uses the actual irradiance value from the current hour to estimate a value closest to the actual for the next hour.

1.2 Previous Work

This section accounts the literature related to solar forecasting both in the context of remote microgrids and the commonly used practices as a whole. Focusing on forecasting

for remote microgrids and the challenges involved, this chapter will also explore the state-of-the-art of daily and hourly solar forecast.

Shakya et al. in 2017 presented a method of solar irradiance forecasting in remote microgrids using MSM, alternatively known as the hidden Markov model (HMM). This method used locally available past dataset of solar irradiance to predict the day-ahead irradiance for scheduling the energy resources in remote microgrids. For fitting the MSM, Fourier basis expansions, which capture the daily and yearly periodic nature of irradiance, were used along with the clear-sky irradiance (CSI) and the past irradiance measurements. The model generated 3 states representing sunny, mildly cloudy, and extremely cloudy days for each day. Then, the day-ahead forecast was made based on the first four daylight hours – the state closest to the actual (measured) irradiance in the first four hours determined the forecast for the day while the forecast for those four hours was the state from the day before. The mean absolute percentage error (MAPE) for the case study in Brookings, South Dakota for the years 2001 to 2005 ranged between 30.7%-32.9%. Similarly, the root mean square error (RMSE) computed was higher for summer months. This method was minimalistic in the sense that it utilizes only the locally available past irradiance data and does not require prior records of meteorological data, which may not be available in remote locations. In this regard, the proposed Markov model -based method is suitable for economic dispatch of energy resources in remote microgrids [1].

The solar forecasting technique presented in [1], however, has a few shortcomings. One of the limitations of this study was that the first four daylight hours may not necessarily be the best representative of the entire day. Similarly, a huge error would occur if the actual irradiance in the first four daylight hours is drastically different from that of

the previous day. Such significant errors would affect the scheduling of the microgrid resources, which may not lead to the most optimal dispatch following substantial operating costs. Essentially, updating the forecast every hour could accommodate the changing conditions within a day that was not able to be have been captured using only the first few hours of the day. Therefore, we are investigating the quality of forecast for other times of the day, which led to the extensive research of literature on short-term forecasting with multiple updates within the day. The notable ones amongst such studies are summarized in the paragraphs that follow.

Feng et al. in 2018 developed an unsupervised clustering-based (UC-based) 1-hour ahead solar forecasting. This method is comprised of 3 parts – GHI time series unsupervised clustering, pattern recognition, and UC-based forecasting. First, the daily GHI is clustered by an optimized cross-validated clustering (OCCUR) method that determines the optimal number of clusters and the best clustering results. Then, the category of a particular day is determined using the first four hours' data in the forecasting stage using the support vector machine pattern recognition (SVMPR). Finally, 1-hour ahead GHI is forecasted by the most suitable models in different clusters that have been built by a two-layer machine learning based multi-model (M3) forecasting framework. The proposed method was validated using 1-year of data with 13 solar features from three information sources some which include GHI, clear sky GHI, CSI, sky imaging features, temperature, relative humidity, wind speed, and wind direction. The simulation results show that the UC-based models outperform the non-UC models with the same architecture by about 20%. Similarly, the M3-based models also outperform the single-algorithm machine learning (SAML) models by about 20% [12].

Wang et al. in 2018 used an image phase shift invariance (IPSI) based cloud motion displacement vector (CMDV) calculation method for ultra-short-term solar power forecasting with one minute time-scale. This method comprised of three stages to forecast the solar power in a minute time scale. First, a Fourier phase correlation theory (FPCT) method was used to calculate multiple different CMDVs from the corresponding consecutive image pairs acquired through different synchronous rotation angles compared to the original sky images. Second, a centroid iteration strategy was used to generate the final CMDV from all the calculated CMDVs based on its density and distance distribution. Finally, the impact of different rotation angle resolution on the final CMDV was analyzed for parameter estimation. This method eliminated random errors of the FPCT method and improved both the accuracy and reliability compared to the original FPCT method [13].

Akarslan et al. in 2018 proposed a novel one-hour ahead short term solar irradiance forecasting methods based on Angstrom-Prescott (A-P) type models. In this work, five models were developed that utilized the past irradiance data, the extraterrestrial irradiance, and the clearness index to generate the hourly forecast. The accuracy of the A-P models highly rely on the region-dependent coefficients, which are determined empirically. The models were validated against a linear and a quadratic A-P type equation-based models using a dataset from three different regions. The proposed methods proved to outperform the conventional A-P type equation-based model [14].

Sheng et al. in 2018 proposed a short-term solar power forecasting based on weighted Gaussian process regression (WGPR). This method was successful in handling the outliers in the measured meteorological data by introducing a weighted local outlier factor (W-LOF) approach. Data samples containing 8 attributes, which includes

photosynthetic active radiation (PAR), ambient temperature, relative humidity, wind speed, wind direction, solar radiation, and precipitation, were used for a 5-min forecast period to cope with the rapid weather changes in the tropical rainforest climate of Singapore. A validation data of 5 consecutive days and the corresponding 5-min ahead measured PV output was used for verification. The proposed method was compared against commonly used forecasting techniques – artificial neural network (ANN), least square support vector machine (LSSVM), and Gaussian process machine learning (GPR), using the same dataset. The simulation results showed that W-LOF-based WGPR method performed better than the other listed in terms of accuracy. The experimental results also showed better accuracy than the commonly used methods under different weather conditions [15].

Miller et al. in 2018 developed a short-term solar irradiance forecasting (1 hour ahead) via satellite/model coupling. This satellite-based, model fusion technique coupled operational geostationary satellite retrievals of cloud microphysical and macrophysical properties to a numerical model to advect clouds and find their shadows on the ground. This cloud information is used to initialize a radiative transfer model that computes the direct and diffuse-sky solar insolation at both shadow locations and intervening clear-sky regions. The model performance was validated against Surface Radiation (SURFRAD; [16]) network observations and was found to outperform the persistence-based forecasting of GHI under all conditions by approximately 10 W/m^2 [17].

Wang et al. in 2018 developed a least absolute shrinkage and selection operator (LASSO) and long short term memory (LSTM) integrated temporal model for

short-term forecasting of short-term solar intensity based on meteorological data. It is a combination of a simple time series model, data clustering, a statistical model, and machine learning. This model takes into account every aspect inside the short-term solar forecasting problem, which includes string time correlation, weather complexity, linear, and non-linear relationships. Different forecasting models were generated for every cluster using the k-means++ in which the TSM captured the strong time correlation, the LSTM learned the non-linear relationship, and the LASSO captured the linear relationship. Meteorological data including temperature, wind chill, humidity, dew-point, wind speed, wind direction, and rainfall were collected every 5 minutes in two different locations, which was used for training, evaluation, and testing. The forecast was validated against the benchmark methods before integration under two different scales – 5 minutes and 30 minutes. The test results show that the proposed method had good precision in forecasting the short-term solar irradiance. However, the prediction accuracy decreased with increasing time scale [18].

Bouzgou et al. in 2019 introduced a fast short-term global irradiance forecasting with wrapper mutual information methodology (WMIM) from an existing historical time series. This method relies on two measures. First is a mutual information measure (MIM) that selects an optimal input set that includes the most significant values of the existing time series. Next is an ANN of the extreme learning machine (ELM) to forecast the future GHI. This new variable selection framework for forecasting the GHI time series is directed specifically towards arid sites with a lower frequency of cloud coverage. Experiments were run at different forecasting horizons (intra-day and intra-hour) and compared against three dimensionality reduction scenarios – full space (latest 50 variables), partial space

(latest 5 variables), and the principal component analysis (PCA). Results show that the proposed method is superior to the other variable selection methods for up to 3 hours ahead for the test in one location and up to six hours ahead in the second one [19].

There are a lot of recent research works in the development of advanced and accurate methods for short-term forecasting of solar irradiance. However, the primary focus on the energy management system of remote microgrids has not been a key objective in many of them. A summary of the works accounted is presented in Table 1.1. As can be seen, most of the methods described in the literature [12]–[15], [17]–[19] make use of weather variables and satellite images. This information may not necessarily be available in a remote community where road accessibility on top of electrical service is still a major challenge. Thus, the main priority of this thesis is to develop an accurate irradiance forecasting method that requires the least prior-installed instrumentation. Achieving this goal is especially challenging because the intermittent nature of the solar irradiance demands a robust model that is capable of handling the uncertainties and requires a large dataset with multiple variables. On the other hand, Hidden Markov models, otherwise known as Markov switching models or dependent mixture models, have proven to be robust for modeling dynamic systems [20]–[25]. Study [1] is an application of such a method in solar irradiance forecasting. Nevertheless, there is still room for improvement in this method in terms of reducing the forecasting error.

Table 1.1. State-of-the-art of various types of forecasting methods.

Forecasting Method		Data Required	Forecasting Horizon	nRMSE (%)	MAPE (%)	Limitations
Markov-based [1]		Global horizontal irradiance (GHI)	1 day	-	31.80 %	Relatively high forecast error
Machine learning [12], [19]		GHI, direct normal irradiance (DNI), direct horizontal irradiance (DHI), clearness index, temperature, pressure, relative humidity, wind speed, wind direction, sky images	1 h to 6 h	8 to 12 %	7 to 11 %	Data required may not be readily available in remote locations
Hybrid	**Satellite-based + NWP [13], [17]**	Sky images	1 min to 1 h	11 to 34 %	-	
	Statistical + Machine learning [18]	Temperature, wind chill, humidity, dew-point, wind speed, wind direction, rainfall	5 min to 30 min		5 %	
Empirical [14]		GHI, clearness index, extra-terrestrial irradiance	1 h	9 to 33 %	-	
Statistical [15]		GHI, photosynthetic active radiation (PAR), ambient temperature, relative humidity, wind speed, wind direction, precipitation	5 min	5 to 11 %	-	

1.3 Motivation

Accurate forecasting of solar irradiance plays a crucial role in determining the optimum dispatch of DERs in a remote microgrid and hence reduce the operating costs. Therefore, the accuracy of various methods of forecasting with hourly update that uses the energy states from the HMM as input is investigated.

1.4 Objective

The main objectives of this thesis are stated below:

(a) develop and evaluate various methods of hourly solar irradiance forecasting based on the states from the HMM,

(b) validate and analyze the forecast with the past irradiance data, and

(c) evaluate the performance of the presented methods on the basis of the time of

forecasting, the time of the year, and their forecasting consistency.

1.5 Thesis Outline

This thesis is organized as follows: Chapter 2 begins with the introduction of the typical structure and management of a remote microgrid. This chapter also presents the theoretical background on the computation of inputs for the HMM, the basic operating principle of the HMM, and the details the metrics used for model validation. Chapter 3 presents the methodology used for generating the intra-day forecast. This chapter also introduces the dataset, details the model fitting procedure, and describes various methods of hourly forecasting. Chapter 4 presents the results and analysis of the forecasting performance of the described methods at different forecasting times and at different times of the year. The validation of the individual proposed methods are also presented in this chapter. Lastly, Chapter 5 presents the conclusion, limitations, and future developments related to the research thesis.

CHAPTER 2 THEORY

Chapter 2 describes the typical structure of remote microgrids along with the EMS required for the optimization of the available resources. As solar irradiance forecasting is a crucial component in the EMS of any microgrid, this chapter also details the theoretical framework for generating the solar irradiance using the method used in this thesis, which is the HMM. This includes the description of the operating principle of the forecasting model along with the input parameters for the model, and the methods used for performance evaluation of the model.

2.1 Structure of remote microgrids

A typical remote microgrid consists of diesel generators, battery banks, and renewable energy sources (like a PV system) to power load for an isolated community. As shown in Figure 2.1, all these components are controlled by a central controller, which is the EMS of the remote microgrid. Conventionally, diesel generators are the primary source of energy for remote microgrids. However, diesel generators have a high operating cost due to the high cost of fuel, fuel transportation and storage cost, and high maintenance cost [5]. In such a scenario, a PV system could help reduce the generator-fuel consumption and hence save the cost in fuels. However, this results in the operation of the generators at low load thus lowering the generator efficiency [26]. Apart from this, the addition of PV in diesel-based microgrids is further challenging as PV is a non-dispatchable energy source and does not correlate with the load demand. Therefore, typically a smaller generator is used in conjunction with an energy storage system like battery banks to minimize the variability in PV generation [27], [28]. Thus, a balance

between lowering the operating cost and maintaining the overall system efficiency is the utmost requirement in a remote microgrid.

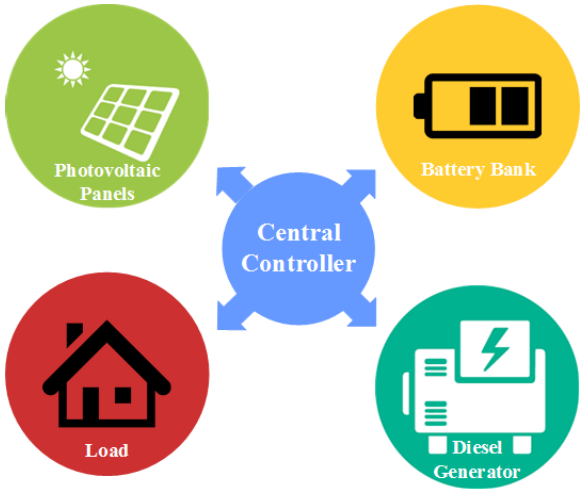

Figure 2.1. A typical remote microgrid with diesel generator, battery bank, and photovoltaic panel serving an isolated load.

This thesis is focused on exploring different methods of intra-day solar irradiance forecasting for remote microgrids, thus the detailed operating principle of only the PV system is presented here.

2.1.1 Photovoltaic system

A PV system basically consists of an array of solar panels connected in series and parallel to match the consumer-end voltage. The solar arrays convert the photon energy from the solar radiation to direct current (DC) electricity. Additionally, solar inverters are also integrated with the solar panels to convert the DC to the alternating current (AC) to enable usage of the generated electricity for a wider range of applications.

The PV power output is a function of the total amount of shortwave radiation received at a surface horizontal to the ground, which is known as the global horizontal irradiance (GHI). This GHI has a diurnal pattern because of the rotation of the Earth

around its own axis. Similarly, the amount of radiation received by the panels at a particular time also varies over the year due to the revolution of the Earth around an elliptical orbit, which is also responsible for the change in seasons. Thus, the amount of electrical power output ($P_{PV,t}$) from a PV array at any time t is given by Equation 2.1.

$$P_{PV,t} = \eta G_t A \tag{2.1}$$

Here, η is the efficiency of the solar panels, G_t is the amount of GHI falling on the surface of the solar panels at time t (in W/m^2), and A is the total effective area of the solar panels (in m^2). Thus, for a particular PV system, the solar irradiance collected at the panel surface determines the electric power output generated by the system.

2.2 Energy management system of remote microgrids

An EMS provides the necessary control for the coordination and efficient operation of the DERs in the remote microgrids like the batteries, generators, and the PV system. The EMS ensures the optimum utilization of these DERs so as to effectively meet the load demands with the least operating cost. A typical EMS for a microgrid consists of a scheduling layer and a dispatch layer [29], which is also adopted by the remote microgrids.

2.2.1 Day-ahead scheduling

Scheduling refers to the process of allocating the dispatchable resources to operate over a certain period of time. For this, the optimal power set points of all the available DERs need to be determined ahead of time. Generally, a day-ahead approach is adopted

by remote microgrids in their scheduling layer. This layer selects the most optimum resources one day ahead of the operation using the forecasted values of the PV power output and the load demand for the next day. Using these forecast values, the power set points for the dispatchable resources (diesel generators and batteries) are determined.

2.2.2 Real-time dispatch

The real-time dispatch layer utilizes the power set points from the scheduling layer to optimally dispatch the diesel generators and the batteries in real-time. Apart from this, the dispatch layer also handles the deviation from the pre-determined set points to ensure power system reliability. This is done by taking necessary corrective actions such as PV power or load curtailment, running spinning reserves, rescheduling, etc.

2.3 Solar irradiance forecasting

Solar irradiance forecast is one of the information required by the scheduling layer of the EMS of a remote microgrid for the optimal dispatch of the available resources. This work is based on study [1], the overview of which is illustrated in Figure 2.2. This method was developed using MSM, alternatively, known as HMM that utilizes only the available past irradiance data during the fitting process. This thesis work is an extension of the MSM-based forecasting method.

The local measurements of the solar irradiance were collected from an online database known as “solaranywhere.com”. First, CSI, which is the irradiance on a cloudless sky, is computed. Then, two Fourier basis expansions are fitted to capture the daily and yearly variability of the solar irradiance. This information is taken as inputs to the MSM. Multiple MSMs with a different number of predefined states are fitted using these inputs.

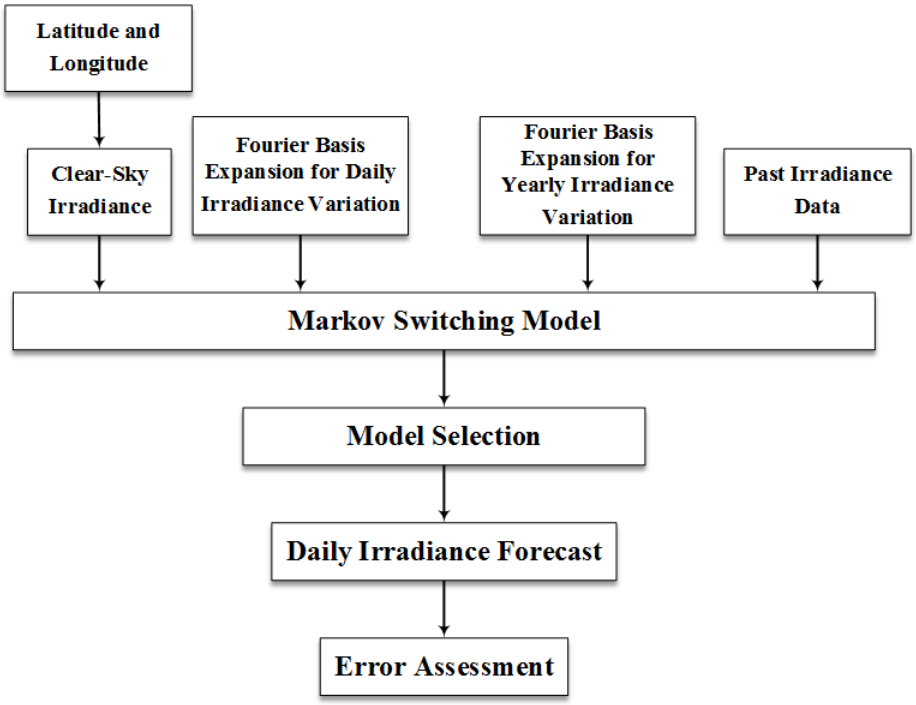

Figure 2.2. Block diagram for solar irradiance forecasting using MSM.

Then, the best model with the optimum number of states is selected using the Bayesian information criteria (BIC). The BIC gives a numeric index to each of the models, which accounts for the goodness of fit and the complexity of the model. The best model with the least BIC index is found for the model with three states. These three states are defined as high, medium, and low corresponding to sunny, mildly cloudy, and extremely cloudy days. The daily irradiance forecast is then computed using the least square error between the three states and the actual irradiance during the first four daylight hours. The details of these steps that were used in this thesis are described in the following sub-headings:

2.3.1 Computation of clear-sky irradiance (CSI)

The instantaneous energy received by the Earth from the Sun can be expressed in terms of solar irradiance (W/m^2). In a cloudless day, the amount of irradiance received at a given location is reduced in intensity as it passes through the atmosphere. Thus, the

attenuation in a clear-sky day is only due to scattering and absorption, and the irradiance in such days are referred to as clear-sky irradiance [30]. The amount of this clear-sky irradiance can be used as a basis in solar irradiance forecasting models.

The reduction of intensity with decreasing altitude is typically assumed to be directly proportional to the increase in air mass [31], which is the ratio of the path length of the Sun's rays that travel through the atmosphere to the minimum path length occurring at noon [30]. For a plane-parallel atmosphere, the air mass (AM) is a function of the secant of the zenith angle (θ_z). An empirical relation given in [31] can be used to determine the clear sky irradiance as follows:

$$CSI = 1367(0.7)^{AM^{0.678}} W/m^2 \tag{2.2}$$

Here, 1367 W/m^2 is the adopted solar constant, which is the average of the irradiance at the highest level of atmosphere. It can, therefore, be inferred that the zenith angle is of utmost importance in computing the clear-sky irradiance and needs to be calculated first.

2.3.1.1 Calculation of zenith angle (θ_z)

For a cloudless day, the amount of irradiance received depends primarily on the position of the Sun relative to the location under evaluation, which is defined by the zenith angle. The zenith angle is the angle of incidence of the solar radiation on a horizontal surface [32] as shown in Figure 2.3. This angle between the Sun's incident beam radiations to the vertical is a function of the time of the year and the time of the day and can be defined as:

$$\cos\theta_z = \sin\delta\sin\phi + \cos\delta\cos\phi\cos\mu \tag{2.3}$$

Here, ϕ is the latitude that defines a geographic coordinate by specifying an angular location north or south of the equator. The declination, δ and the hour angle, μ are described in the sub-headings that follow.

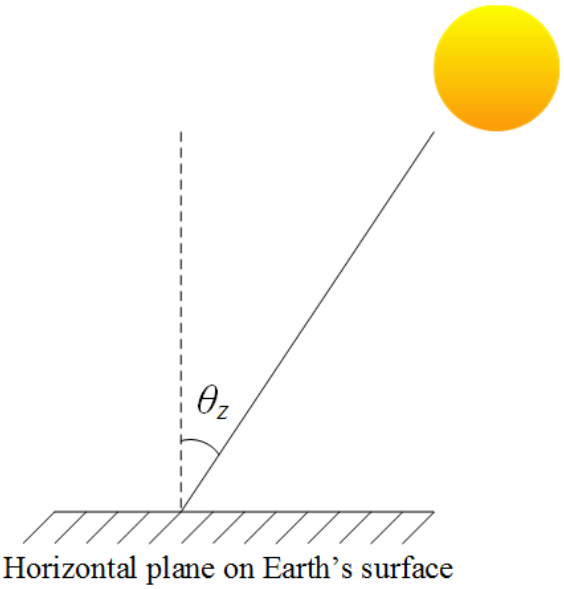

Figure 2.3. Zenith angle made by the Sun on a horizontal surface of the Earth.

Declination (δ): Declination is the angular position of the Sun at solar noon with respect to the equatorial plane as can be seen in Figure 2.4. As the Earth revolves around the Sun, this angle varies throughout the year and therefore gives a measure of the time of the year. Thus, the declination for any day (n) can be approximated using Cooper's equation (1969) as given in [32].

$$\delta = 23.45^0 \sin\left(\frac{360}{365}(284+n)\right) \tag{2.4}$$

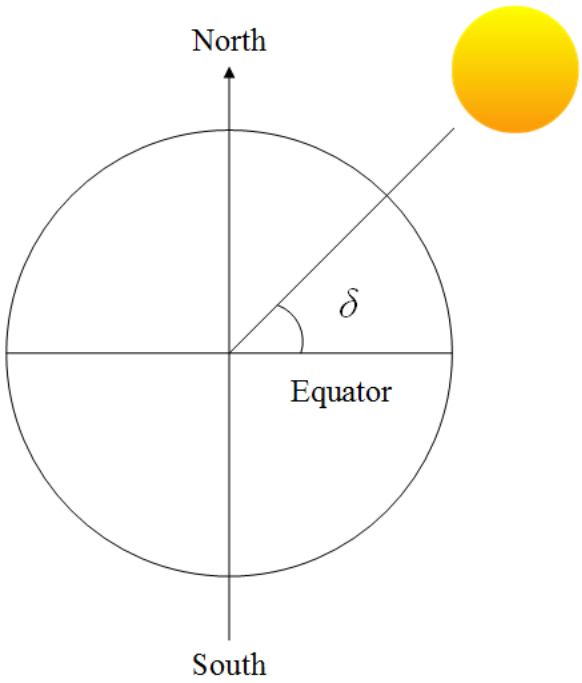

Figure 2.4. Solar declination.

Hour angle (μ): Just as the declination tells the time of the year, the hour angle quantifies the time of the day. The hour angle is the angular displacement of the Sun east or west of the local meridian due to the rotation of the Earth on its axis at the rate of 15^0 per hour. In other words, it is the difference between the solar noon and the desired time of the day in terms of a 360^0 rotation in 24 hours. That makes its value to be negative for the morning and positive for the afternoon. As shown in [33], if T is the time of the day with reference to the solar midnight on a 24-hour clock, the hour angle can be mathematically represented as:

$$\mu = \frac{12 - T}{24} \times 360^0 \tag{2.5}$$

2.3.2 Computation of Fourier basis expansions

The periodicity of solar irradiance on a daily and yearly span can be characterized by Fourier basis expansions. A Fourier series is an expansion of a periodic function in terms of an infinite sum of sines and cosines as follows [34].

$$\hat{x}(t) = \beta_0 + \beta_1 \sin \omega t + \beta_2 \cos \omega t + \beta_3 \sin 2\omega t + \beta_4 \cos 2\omega t + \ldots \tag{2.6}$$

A Fourier series as in (2.6) can also be defined by a weighted sum of the basis functions, $\phi_0(t) = 1$, $\phi_{2r-1}(t) = \sin r\omega t$, and $\phi_{2r}(t) = \cos r\omega t$ such that the series can be alternately represented as follows.

$$\hat{x}(t) = \sum_{m=0}^{n} \beta_m \phi_m(t) \tag{2.7}$$

Here, ω is the angular frequency that defines the time period as $\omega = 2\pi/T$ and r, in the basis functions, represents the number of pairs of the sine and the cosine terms or simply the number of basis functions that make up the series.

The amount of solar irradiance received at a specific location is variable throughout the day because of the rotation of the Earth about its polar axis. Likewise, there is also variation throughout the year due to the revolution of the Earth around the Sun in an elliptical orbit. This variability of the solar irradiance, throughout the day and the entire year, makes it reasonable for the irradiance to be represented as a periodic function of time. In [1], these daily and yearly variations are represented by 8 and 6 Fourier basis functions, respectively. Here, the values of the coefficients are determined by fitting a model.

2.3.3 Fitting the hidden Markov model (HMM)

An HMM (or MSM) is used to capture the dynamic nature of the solar irradiance. There are two components in an HMM. First is the observation sequence, which is the actual irradiance measurements in this case. Second is a set of hidden states that correspond to the cloud cover conditions. Three hidden states have been defined as high, medium, and low, which in fact, are also irradiance values at a particular time of the day that represent sunny, mildly cloudy, and extremely cloudy conditions. Here, the irradiance forecast is made by first estimating the irradiance values corresponding to each state throughout the day and then choosing the state that is closest to the observation.

Let $\boldsymbol{Y} = [y_1, y_2, \ldots, y_T]$ be a sequence of observations of past irradiance data. Also, at a given time point, let the observations be distributed according to a mixture model with K hidden states given by $\boldsymbol{S} = (S_1, \ldots, S_T)$, where, $S_t \in \{1, \ldots, K\}$ for $t = 1, \ldots, T$. Here, the covariates or the independent variables of the model are the CSI, and 8 daily and 6 yearly Fourier basis expansions. Then, let $\boldsymbol{X} = [\boldsymbol{X_1}, \ldots, \boldsymbol{X_T}]$ denote the matrix of covariates over a time series of length T. Now, the set of model parameters is given by $\boldsymbol{\theta} = [\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\Pi}, \boldsymbol{\beta}]$. Here, $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{\Pi}$, and $\boldsymbol{\beta}$ represent the probability transition matrix, the emission matrix, the initial state matrix, and the model coefficients respectively. In this thesis, only the model coefficients are being used for estimating the states and hence in the final forecast. Now, the joint likelihood of the sequence of observations ($\boldsymbol{y}$) and hidden states ($\boldsymbol{S}$), given the covariates ($\boldsymbol{X}$) and the model parameters ($\boldsymbol{\theta}$) is given by:

$$L(\boldsymbol{y}, \boldsymbol{S} | \boldsymbol{X}, \boldsymbol{\theta}) = P(S_1 = i | \boldsymbol{X_1}) f_{S_1}(y_1 | \boldsymbol{X_1}) \prod_{t=2}^{T} p_{ij}(\boldsymbol{X_t}) f_{S_t}(y_t | \boldsymbol{X_t}) \tag{2.8}$$

Here, $f_{S_1}(y_1|\boldsymbol{X_1})$ describes a vector of conditional densities $\boldsymbol{B} = P(y_t|S_t = j, \boldsymbol{X_t})$ for $j = 1, \ldots, K$. Similarly, $p_{ij}(\boldsymbol{X_t}) = P(S_{t+1} = j|S_t = i; \boldsymbol{X_t})$ represents the transition probability from state i to j given the covariate $\boldsymbol{X_t}$. The model parameters are then estimated using the expectation-maximization (EM) algorithm. The EM algorithm consists of two steps – E-step and M-step, which perform the expectation and the maximization, respectively as described in the following steps.

1. Input past irradiance data as observation sequence $\boldsymbol{y}$.
2. Define K number of states.
3. Initialize model parameters: $\boldsymbol{\theta} = [\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\Pi}, \boldsymbol{\beta}]$
4. E-step: Estimate expected value of each state using model parameters $\boldsymbol{\theta}$ and observation sequence $\boldsymbol{y}$.
5. M-step: Re-estimate model parameters to maximize joint likelihood described in Equation 2.8.
6. Repeat E-step and M-step until model parameters $\boldsymbol{\theta}$ are adjusted to a pre-specified threshold.

In this way, the required model parameter β is retrieved from the EM algorithm. Here, β constitutes the coefficient for the CSI, and 8 and 6 coefficients for the daily and yearly Fourier basis expansions, respectively.

2.3.4 State estimation using linear regression model

A linear regression model as shown in Equation 2.9 is used to estimate the irradiance values corresponding to each state. The model coefficients β for each of the k

states obtained by training the HMM along with the model covariates are used to estimate the irradiance values $\hat{y}_t$ for each state k at any time t.

$$\hat{y}_t(S_t = k) = \beta_{k,1} CSI(t) + \sum_{i=1}^{p} \beta_{k,1i} \phi_{1i}(t) + \sum_{j=1}^{q} \beta_{k,2j} \phi_{2j}(t) + \varepsilon_k(t) \tag{2.9}$$

Here, $CSI(t)$ is the clear-sky irradiance at time t with its corresponding coefficient $\beta_{k,1}$. ϕ_{1i} and ϕ_{2j} are the yearly and daily Fourier basis functions, respectively with a corresponding total of p and q components. Also, $\beta_{k,1i}$ denotes the coefficient corresponding to the i^{th} component of the Fourier basis function for the yearly variation of irradiance at state k. Similarly, $\beta_{k,2}$ denotes the coefficient corresponding to the j^{th} component of the Fourier basis function for daily variation of solar irradiance at state k. Finally, $\varepsilon_k(t)$ is a random noise component that is assumed to be normally distributed with a mean of zero and a constant variance σ_k^2. Thus, $\hat{y}_t(S_t = k)$ gives the estimate of the irradiance value corresponding to each state k at any instant of time t.

2.3.5 Irradiance forecasting method

Once the states are estimated for each day, the daily forecast is generated by choosing the best state looking into the first four daylight hours of the target day. This is done by calculating the sum of squared errors (SSE) between each of the states and the actual irradiance in the first four daylight hours. Mathematically, the SSE is calculated as follows:

$$SSE(S_t = k) = \sum_{t=1}^{4} (y_t - \hat{y}_{k,t})^2 \tag{2.10}$$

Here, y_t is the actual irradiance and $\hat{y}_{k,t}$ is the irradiance of each of the three states (given by k) in the four-hour window.

2.3.6 Error metrics for evaluation of forecasting performance

The two standards error metrics for evaluating the performance of forecasting are the RMSE and the MAPE. The RMSE gives the standard deviation of the prediction errors. It quantifies the deviation of the forecast from the actual observation. Thus, larger deviations have stronger weights than smaller ones. This error is expressed in the same unit as the unit of the observation. The MAPE, however, gives the forecast error as the percentage of the actual observation. In other words, it is a proportional error in terms of the actual observation. These errors are calculated using expressions shown in Equations (2.11) and (2.12).

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2} \tag{2.11}$$

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\frac{(y_t - \hat{y}_t)^2}{y_t} \times 100\% \tag{2.12}$$

Here, $(y_t - \hat{y}_t)$ is the error in point forecast with y_t and $\hat{y}_t$ representing the actual observation and the forecast, respectively at a time t. On the other hand, T represents the total number of observations in the period within which the error is to be calculated.

CHAPTER 3 METHODOLOGY

Chapter 3 describes the processes that were followed to reach the objectives of this thesis. The HMM-based technique described in [1] will be used to investigate four different methods of intra-day solar irradiance forecasting. The methods will be tested against the actual irradiance measurements from Brookings, South Dakota. The comparison of forecasting accuracy with the different intra-day forecasting methods described will be done and presented in Chapter 4.

3.1 Overview of solar irradiance forecasting process

The basic scheme for forecasting the solar irradiance is based on [1] as shown in Fig. 3.1. The first step is to fit an HMM with four inputs – CSI, daily and yearly Fourier basis expansions, and the past solar irradiance data. Then, the coefficients of the model, known as the Markov coefficients, are used in a linear regression model to estimate the daily patterns of each of the chosen states. Then the next step is to choose the best state for every hour to carry out the intra-day forecasting, which in our case refers to updating the forecast every hour. Finally, the forecast error is calculated by comparing against the actual irradiance using the error metrics described in Section 2.3.6.

3.2 Dataset

The dataset used for fitting and evaluating the forecasting model is acquired from SolarAnywhere®, which is a web-based tool that offers hourly GHI, direct normal irradiance (DNI), and diffuse horizontal irradiance (DHI) for various locations in the United States. The irradiance data that the tool provides has an hourly temporal resolution

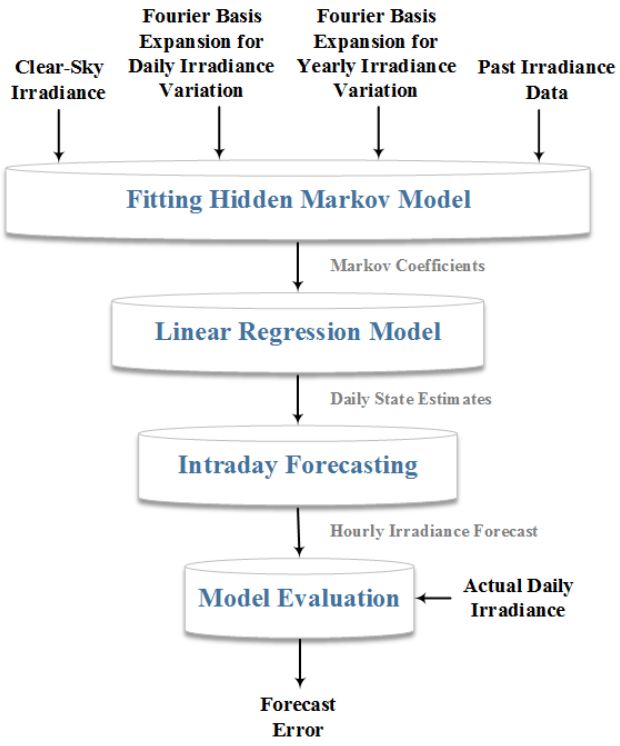

Figure 3.1. Overview of the methodology for forecasting solar irradiance .

and an area resolution of 100 km^2. In this thesis, a case study for Brookings, South Dakota, is performed taking only the GHI for analyzing the model. While the entire dataset considered consists of the years 2000 to 2011, only the years 2008 to 2010 were used for fitting the HMM because of no missing data in that period. The remaining years, however, were used for model evaluation.

3.3 HMM fitting

An HMM was fitted in *R* version 3.4.4, which is a free software environment for statistical computing and graphics. The inputs for fitting the model are the CSI, the daily and yearly Fourier basis expansions, and the past irradiance data for Brookings, South Dakota taken for the years 2008 to 2010. The CSI was obtained from [1], which uses Equation 2.2 and the Fourier components were determined using a package in *R* for

functional data analysis called the *FDA*. The model fitted consists of three states namely high, medium, and low that represents irradiance level corresponding to sunny, mildly cloudy, and extremely cloudy days. As a result of model fitting, the Markov coefficients were obtained, as shown in Table 3.1. They give the corresponding coefficients for CSI, 8 daily, and 6 yearly Fourier components for each of the three states.

Table 3.1. Markov coefficients obtained from hidden Markov model

State (k)	CSI	sin1	cos1	sin2	cos2	sin3	cos3	sin4	cos4	sin1.1	cos1.1	sin2.1	cos2.1	sin3.1	cos3.1
k=1	0.2	0.3	-6.4	3.8	1.8	2.2	-2.2	-0.2	-1.7	-11.3	7.0	13.3	12.8	4.8	7.5
k=2	0.5	4.5	-44.9	-1.9	-21.7	-3.1	-6.4	-3.1	0.5	-14.1	69.4	60.8	121.5	27.6	35.2
k=3	0.6	37.3	-132.1	3.6	-19.6	-3.7	-0.5	-1.6	-0.1	-19.7	101.5	102.4	274.2	37.6	57.4

3.4 Linear regression model for state estimation

After fitting the HMM, the Markov coefficients are used to fit a linear regression model to estimate the daily pattern of the three states throughout the year. The coefficients of CSI and daily and yearly Fourier components from Table 3.1 are used to compute the daily state estimates y_t for each state k as shown in Equation (2.9). The estimated three states – high, medium, and low, shown in Figure 3.2, portray three different types of cloud coverage condition that could occur within a day. So, the next task is to select the appropriate state in every hour that best defines the cloud coverage condition and hence the irradiance level for that hour. For this purpose, four different methods for intra-day forecasting are proposed, which are described under the following heading.

3.5 Intra-day forecasting methods

To select the proper state in every hour in a day, four methods for intra-day forecasting have been developed. All the methods only consider the daylight hours of each day for prediction. The daylight hours in a day are determined from the CSI values for the

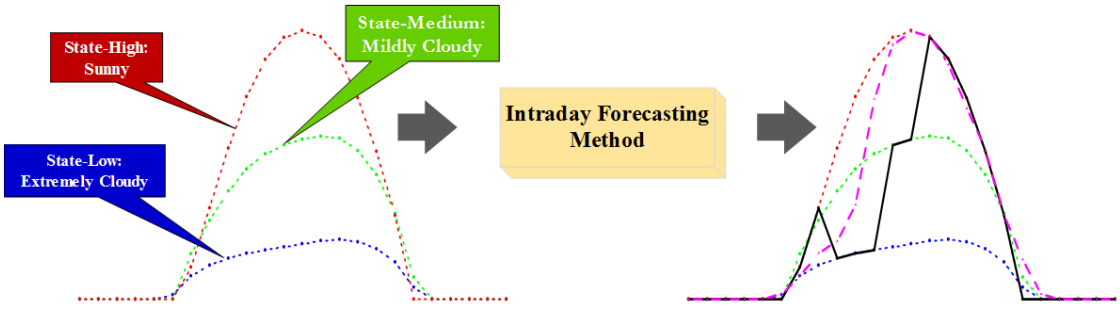

Figure 3.2. Selection of hourly state using intra-day forecasting methods.

day. A specific initialization is made for the forecast of the early daylight hour(s) for each method. Then, the forecast for the next hour is carried out based on the specific algorithm pertaining to the method. Therefore, as the day progresses, the forecast for the following hours is updated until when it stops an hour before the last daylight hour.

3.5.1 Past four-hour method (Past4Hr)

This method considers the irradiance data from the past four hours for making the forecast. Figure 3.3 illustrates the flowchart for intra-day forecasting using the Past4Hr method. The main idea behind this method is to have a four-hour window and choose the state that is closest to the actual irradiance (in that window) as the forecast for the next hour. For this, each of the three states are compared with the actual irradiance to find the closest state by using least square method. In other words, the SSE is computed for each state by using Equation (3.1) and the state with the least SSE would give the forecast for the following hours. However, the state in the first four-hour window is found using the

least square method for the same window from the previous day.

$$SSE(S_t = k) = \sum_{t=1}^{4} (y_t - \hat{y}_{k,t})^2 \tag{3.1}$$

Here, y_t is the actual irradiance and $\hat{y}_{k,t}$ is the irradiance of each of the three states (given by k) in the four-hour window. The intra-day forecast is generated by moving this 4-hour window until the last daylight hour.

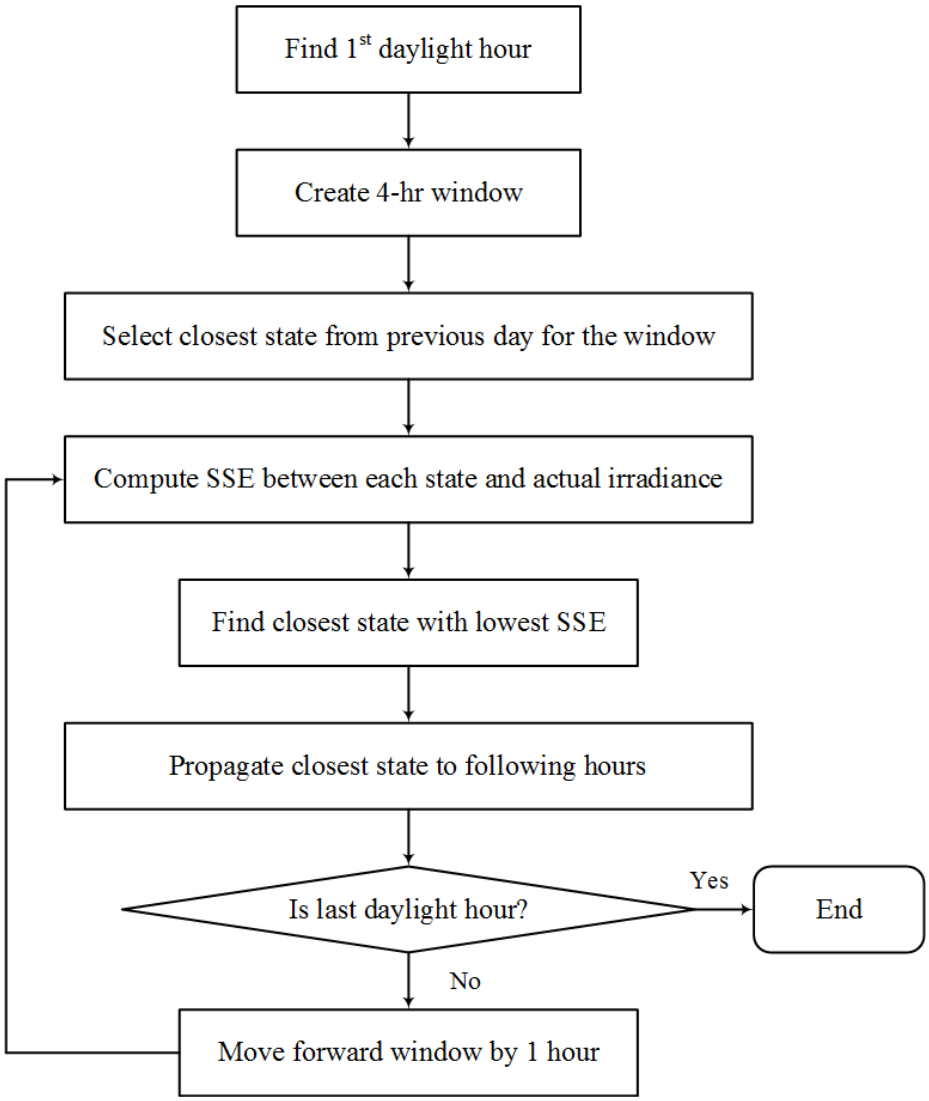

Figure 3.3. Flow chart of intra-day forecasting using Past4Hr method.

3.5.2 Slope1 method

This method uses the slope to detect the hourly rise and fall of irradiance. Here, the slope refers to the first derivative or the rate of the change in irradiance in the past two hours from the time of forecast. Figure 3.4 illustrates the flowchart for intra-day forecasting using the Slope1 method. Here, the hourly forecast is made by comparing the slopes derived from the past two hours from the point of the forecast. For this, the irradiance values of the actual measurement and all three states from the past two hours are taken to compute the slope as shown in Equation 3.2. Then, the state with the slope closet to that of the actual irradiance is chosen as the forecast for the upcoming hours. On the other hand, the first two daylight hours are initialized to be in the low state for every day for simplicity. This assumption would not be detrimental to the forecast as there would not be a significant difference in the irradiance values between the three states in those early hours.

$$m_t = \frac{\Delta y_t}{\Delta t} \tag{3.2}$$

Here, Δy_t is the difference in irradiance values between two consecutive hours for either of the three states or the actual irradiance. In this case, y_t will always be unity as two consecutive hours are being considered.

3.5.3 Slope2 method

This method also utilizes the slope to identify if the irradiance in the next hour is going up or down. Figure 3.5 illustrates the flowchart for intra-day forecasting using the Slope2 method. First, the state closest to the actual, at the point of the forecast, is determined. Then, the slope of the state is calculated from the past two hours. Next, a

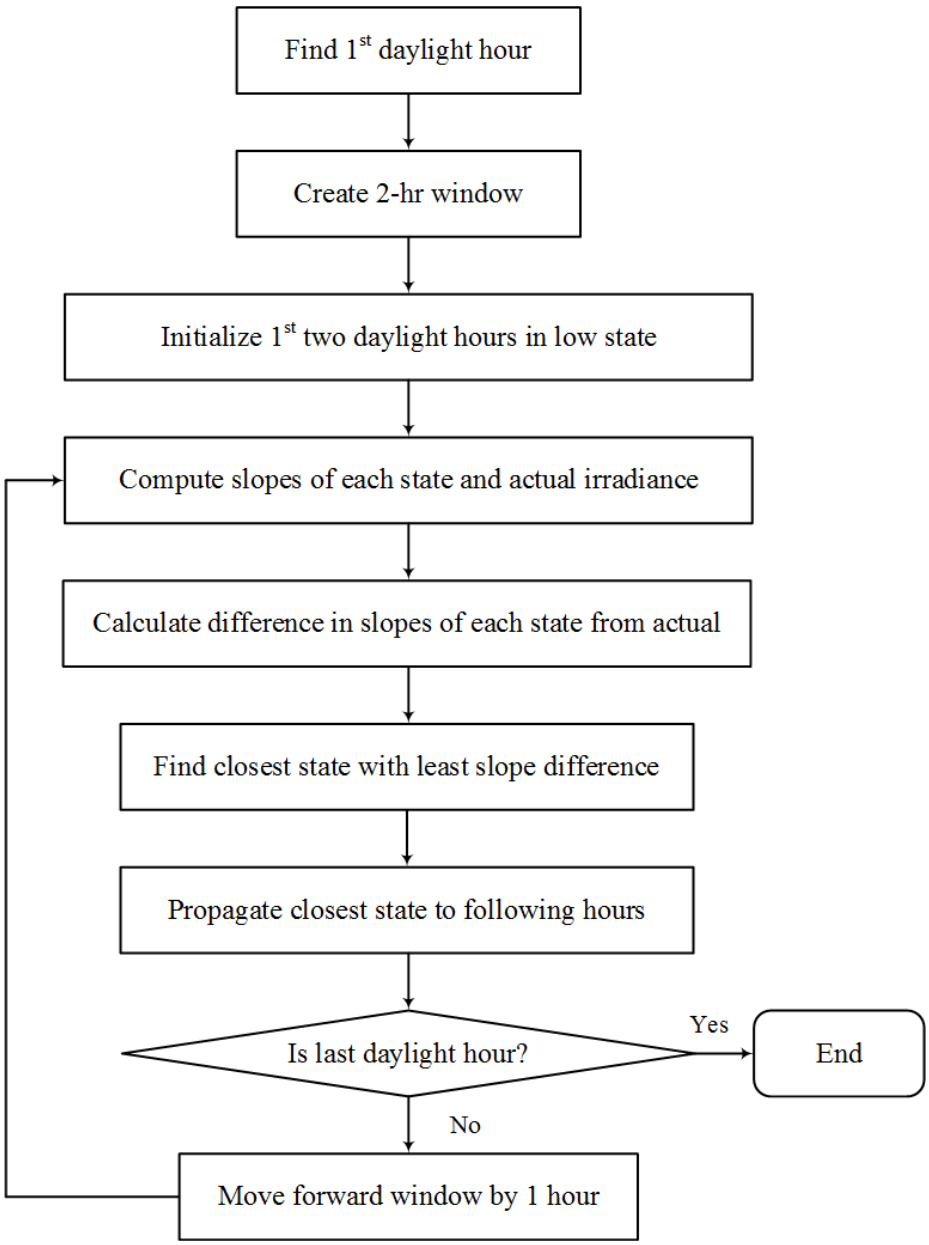

Figure 3.4. Flow chart of intra-day forecasting using Slope1 method.

decision band is created above and below the slope of the closest state. If the slope of the actual irradiance is greater than the upper band, the state in the next hour goes a level higher. On the contrary, if the slope of the actual is lower than the lower band, the next hour is a level lower. However, if the slope of the actual lies in between the two decision bands, the same state is persisted in the following hours. The closest state is found by merely taking the absolute difference of each of the states from the actual irradiance. Meanwhile, the decision band is formed by taking the slopes of the closest state from the past four days. For this, a standard deviation (sd) of the slopes derived from the irradiance in the same interval from the past four days are computed. Then, the slope of the closest state in the current day $\pm 2sd$ gives the upper and lower bounds. Like in the previous method, the first two hours of the day are assumed to be in the low state for the entire year.

3.5.4 Past hour method (PastHr)

This method takes into account the irradiance value from the past hour to make the forecast. Figure 3.6 illustrates the flowchart for intra-day forecasting using the PastHr method. Here, irradiance values of each state from the past hour is compared with that of the actual irradiance in the same hour. The state with the least difference will give the forecast for the subsequent hours. Furthermore, the first daylight hour is initialized to be in the low state like in the previous two methods.

3.6 Evaluation of forecasting performance

The overall forecasting performance using the aforementioned methods are evaluated using the standard error metrics described in Section 2.3.6. In this thesis, the

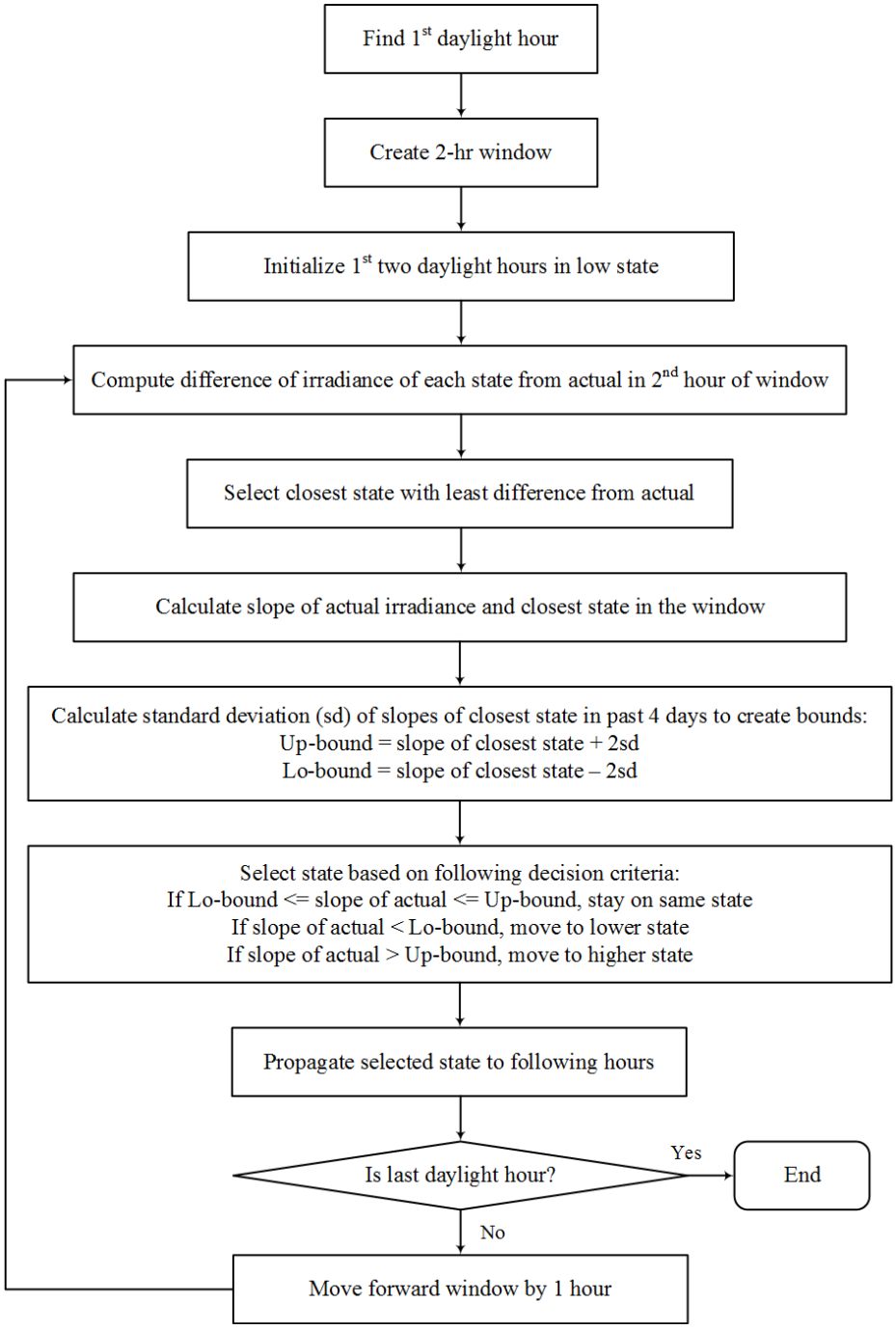

Figure 3.5. Flow chart of intra-day forecasting using Slope2 method.

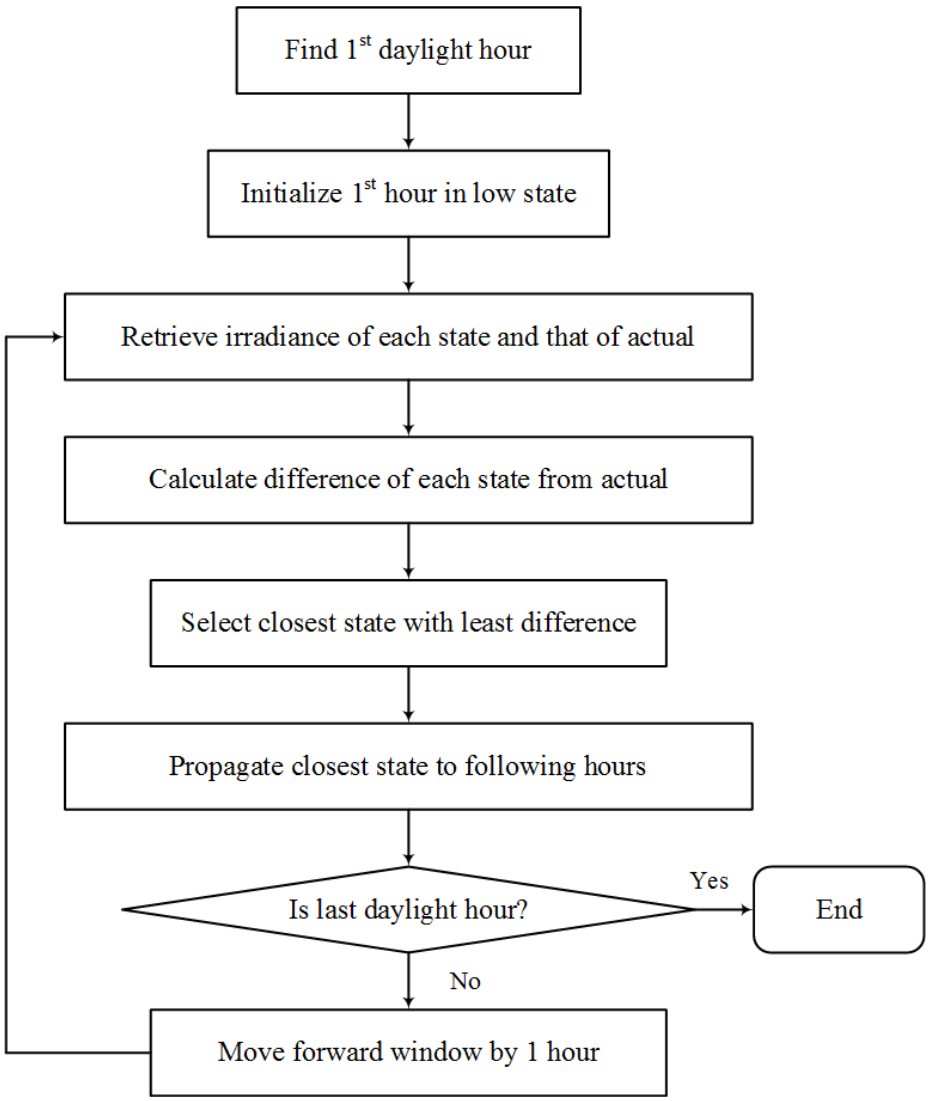

Figure 3.6. Flow chart of intra-day forecasting using PastHr method.

forecasting performance is measured in two ways by calculating – (a) daily errors and (b) yearly errors. The daily error gives an idea of how far off the forecasted irradiance and thus the solar power output are from the actual solar power that is available after the point of forecast. Moreover, the yearly error gives a bigger picture of the forecasting strength of different methods to enable comparison on a broader scale.

For calculating both error types, the number of observations taken is only the points or hours after the forecast until the last daylight hour. For instance, a forecast is to be

made for a day with daylight hours from 5 am to 9 pm. If the forecast is made at 8 am, then the points of observations beyond this time, which is from 9 am to 9 pm, will only be considered in calculating the daily error. Thus, the total time period under consideration is 13 hours. Now for the same day, this time period shrinks to 12 hours when the forecast is made at 9 am. Thus, there will be individual error values associated with each hour of the forecast, even for the same day. Now for the yearly error, the sum of these daily time periods at each forecast time gives the total number of observations for the forecast made at that particular hour.

CHAPTER 4 RESULTS AND ANALYSIS

This chapter includes the results of intra-day solar forecasting using different methods described in Section 3.5. For the validation of the forecasting methods, the forecasts were compared against the historical data from the year 2011 because of the availability of a complete dataset during the year. The methods investigated were validated by daily plots and by analyzing their annual error rates. Furthermore, the performance of each method was evaluated on the basis of (a) the forecasting time of the day, (b) the time of the year, and (c) the consistency over multiple years.

4.1 Validation of the proposed methods

An example day of August 13, 2011 is chosen to validate the proposed forecasting methods. For this, the three states - high, medium, and low, are represented by the red, green, and blue curves. The curve in magenta depicts the actual irradiance while the black curve indicates the forecasted irradiance. The daily forecast starts from the first daylight hour until the last one and is updated every hour. However, in this section, forecast for only three specific times - 8 am, 11 am, and 2 pm, are being demonstrated.

4.1.1 Results from Past4Hr method

Figures 4.1 to 4.3 show an example of how intra-day forecasting is carried out using the past four-hour method (Past4Hr). The example shown is for August 13, 2011, when the daylight hours are from 5 am to 9 pm. Here, Figure 4.1 shows the first forecasting window that covers 5 am to 8 am as shown by the transparent orange rectangle. In this window, the low state has the least square error due to which the low state is propagated in

the subsequent hours. It can be noticed that the forecasted state in the first four-hour window is high. This is because the high state has the least square error in the 5 am to 8 am window in the previous day. In summary, the forecast at 8 am shows that the period from 9 am to the last daylight hour will be in the low state judging from the past four-hour window. However, the state changes to medium and then to high after forecasting at 11 am and 2 pm respectively, as shown by Figures 4.3 and 4.3. After 2 pm, the state remained high for all remaining hours. The RMSE and MAPE for the forecast made at 8 am were calculated to be 424.8 W/m^2 and 69.3%. Nevertheless, the errors reduced to 211.5 W/m^2 and 37.2% when forecasted at 11 am, which further reduced to 53.5 W/m^2 and 34.9% when forecasted at 2 pm. The yearly errors at different forecasting times using this method for the year 2011 are tabulated in Table 4.1.

Table 4.1. Yearly errors at different forecasting times using Past4Hr method.

	8:00 AM	11:00 AM	2:00 PM
RMSE (W/m^2)	199.7	136	102.4
MAPE (%)	63.1	53.7	67.8

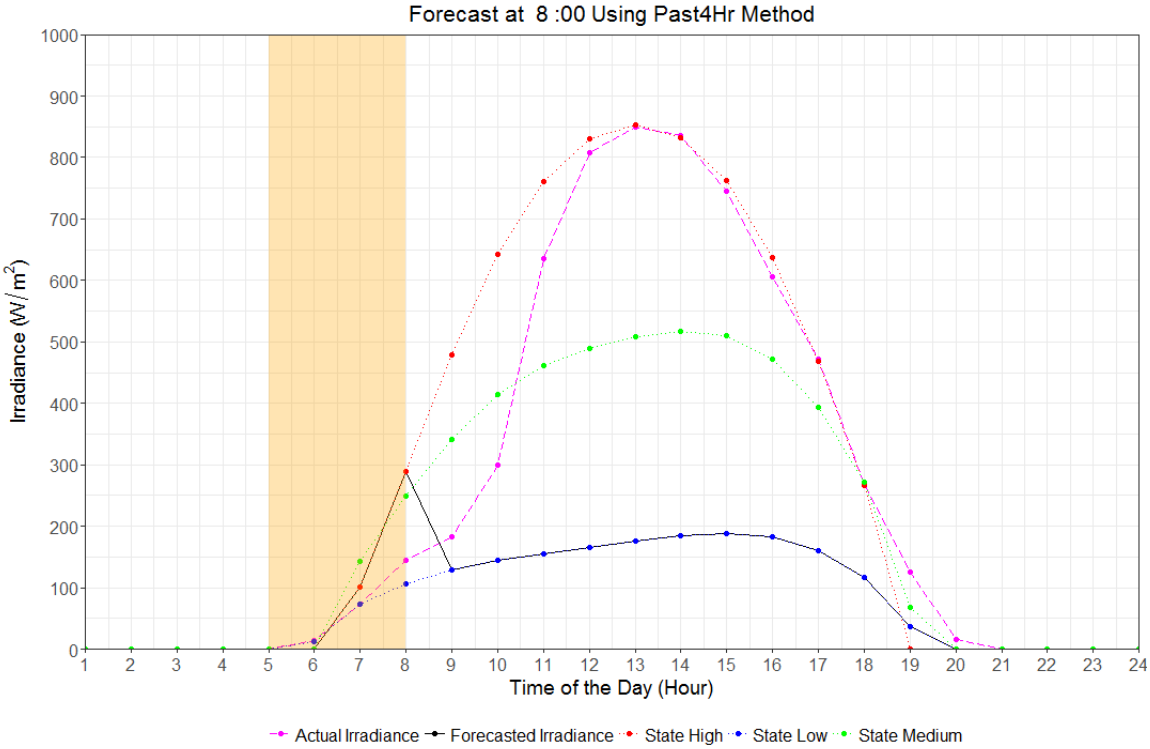

Figure 4.1. Forecasting at 8 am using Past4Hr method for August 13, 2011.

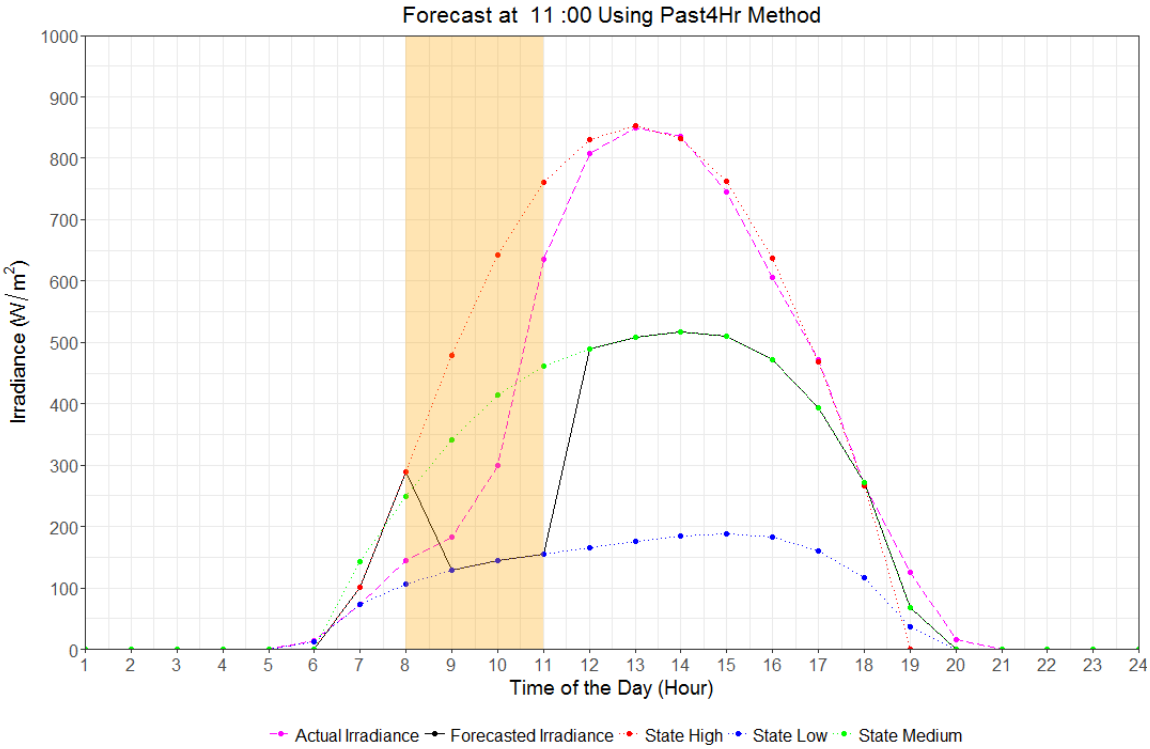

Figure 4.2. Forecasting at 11 am using Past4Hr method for August 13, 2011.

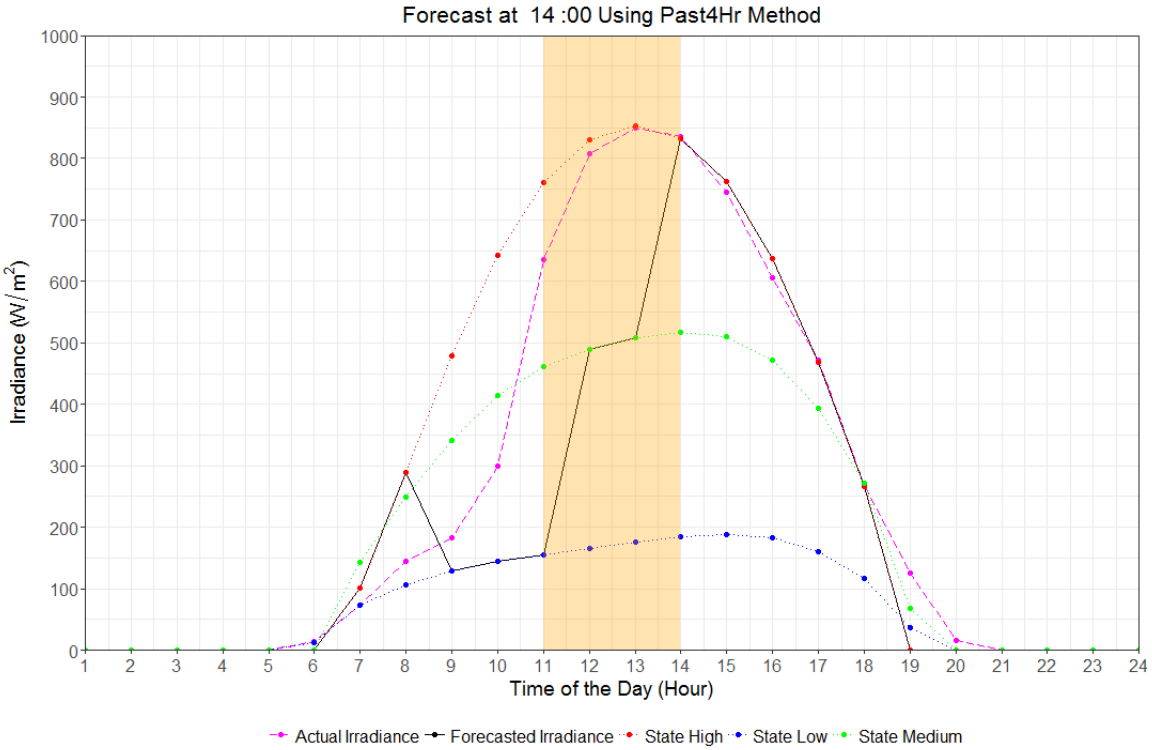

Figure 4.3. Forecasting at 2 pm using Past4Hr method for August 13, 2011.

4.1.2 Results from Slope1 method

Figures 4.4 to 4.6 display intra-day forecasting using the Slope method. Figure 4.4 shows a forecast made at 8am using the irradiance values from 7 am to 8 am window as shown by the transparent orange rectangle. Here, the irradiance values in the window are used to compute the slopes of the actual irradiance and the three states. During this period, the slope of the medium state was closer to the actual irradiance as shown in Table 4.2. Due to this, the medium state is propagated in the following hours. The forecast is updated every hour in the similar way. For forecast made at 11 am, as shown in Figure 4.5, the state closest to the actual irradiance is the high state. Therefore, for the hours beyond 11 am, the forecast is at high state even though it stayed briefly at medium at 9 am and then back to low at 10 am. Now, in the 1 pm to 2 pm window, as can be seen in Figure 4.6, only

the actual irradiance and the high state, have downward slopes. The values of these slopes were found to be -13 Wm^2h and -20.88 Wm^2h as shown in Table 4.2. Therefore, the high state is again the closest state to the actual irradiance. The RMSE and MAPE for the forecast made at 8 am were calculated to be 198.1 W/m^2 and 40.6%. Nevertheless, the errors reduced to 44.3 W/m^2 and 23.6% when forecasted at 11 am. However, the errors slightly rose to 53.5 W/m^2 and 34.9% when forecasted at 2 pm. The yearly errors at different forecasting times using this method for the year 2011 are tabulated in Table 4.3.

Table 4.2. Forecasting of state for the next hour using Slope 1 method.

Forecast Time	Actual Irradiance Slope (W/m^2h)	Low State Slope (W/m^2h)	Medium State Slope (W/m^2h)	High State Slope (W/m^2h)	Forecasted State
8:00 AM	72	33.18	106.41	187.56	Medium
11:00 AM	335	11.02	46.94	118.32	High
2:00 PM	-13	9.17	9.82	-20.88	High

Table 4.3. Yearly errors at different forecasting times using Slope1 method.

	8:00 AM	11:00 AM	2:00 PM
RMSE (W/m^2)	183.6	147	147.3
MAPE (%)	63.7	56.3	87.7

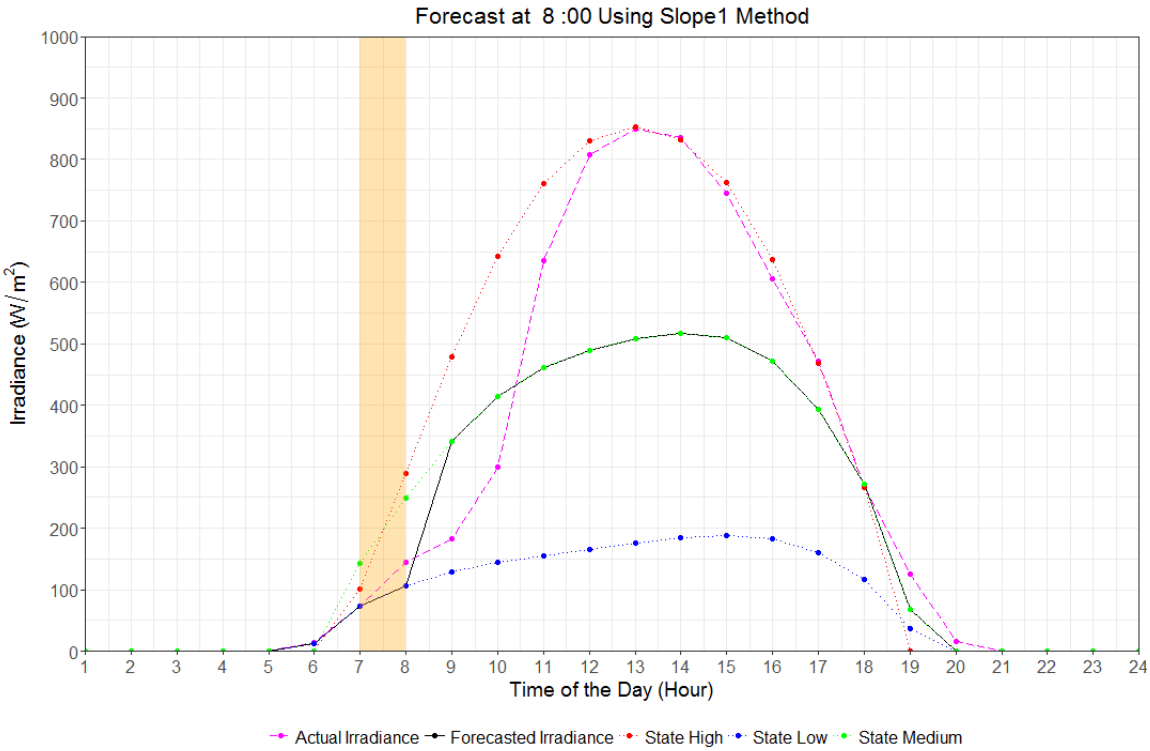

Figure 4.4. Forecasting at 8 am using Slope1 method for August 13, 2011.

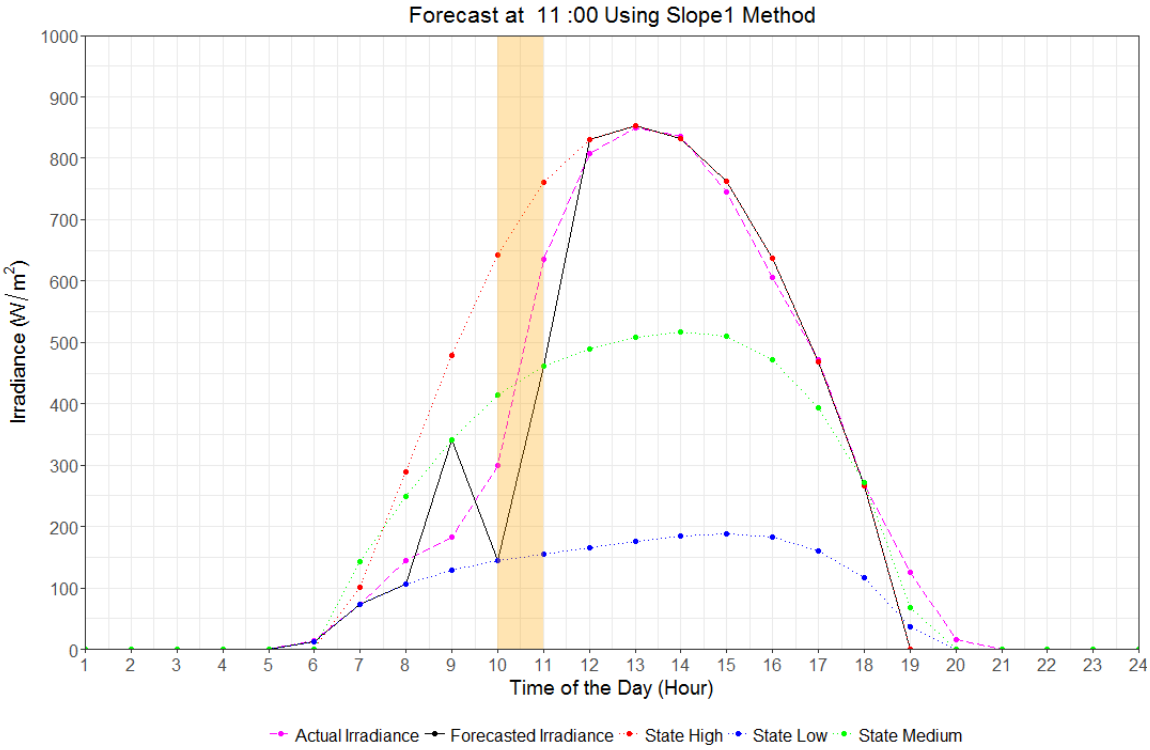

Figure 4.5. Forecasting at 11 am using Slope1 method for August 13, 2011.

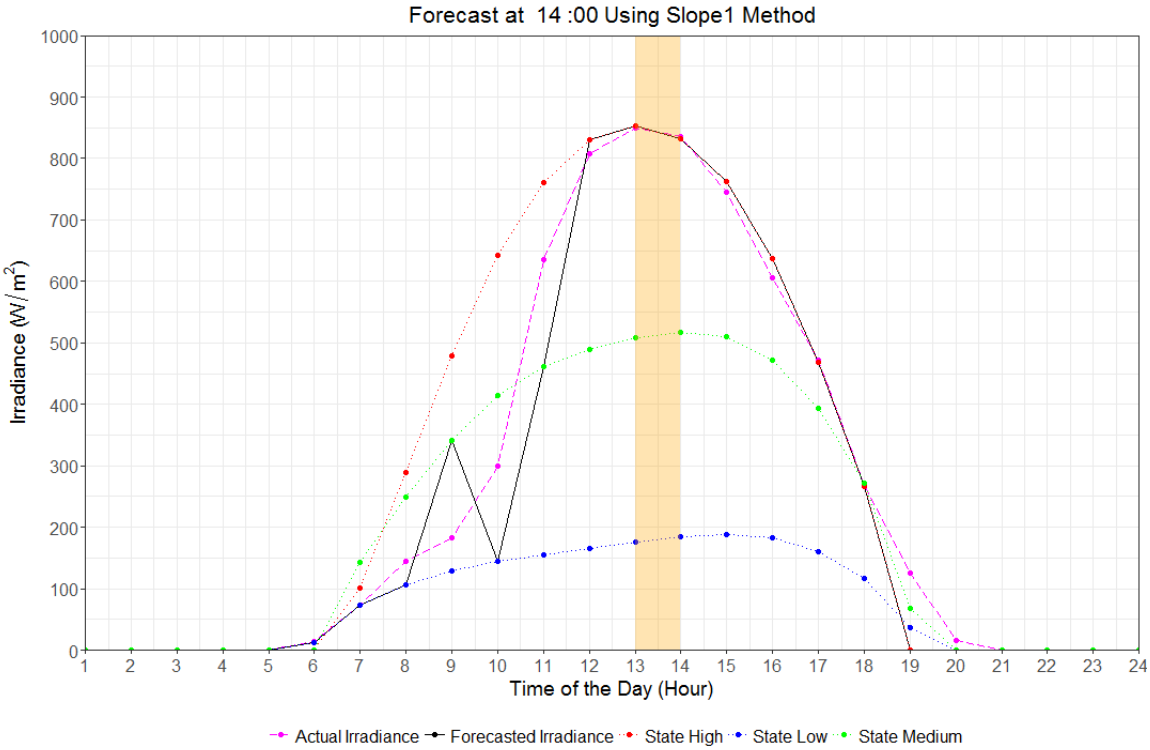

Figure 4.6. Forecasting at 2 pm using Slope1 method for August 13, 2011.

4.1.3 Results from Slope2 method

Figures 4.7 to 4.9 demonstrate an example of intra-day forecasting using the Slope2 method. The forecast for 8 am is shown in Figure 4.7. At 8 am, the low state is closest to the actual irradiance, which has a slope of 33.2 W/m^2h. Therefore, irradiance data of low state from the past four days for 7 am and 8 am are fetched to calculate the slopes in this window for the past four days. Then, the standard deviation of the slopes for this state is computed to form the decision bands. In this period, the slope of actual irradiance is 72 W/m^2h while the $2sd$ that helps form the band is 0.9 W/m^2h. Clearly, the slope of actual irradiance is greater than the upper bound as shown in Table 4.4. Due to this, the next hours will go a state higher to the medium state. Similarly, the slope of bounds and the forecasted state at other hours are shown in Table 4.4. The RMSE and MAPE for the

forecast made at all three times for this day - 8 am, 11 am, and 2 pm, were calculated to be the same as for the forecast using Slope1 method. This is because the states chosen for the three forecast times were same for the example day chosen, which was medium, high, and high. The yearly errors at different forecasting times using this method for the year 2011 are tabulated in Table 4.5.

Table 4.4. Forecasting of state for the next hour using Slope 2 method.

Forecast Time	Closest State	Closest State Slope (W/m^2h)	Lower Bound Slope (W/m^2h)	Upper Bound Slope (W/m^2h)	Actual Slope (W/m^2h)	Forecasted State
8:00 AM	Low	33.2	32.3	34.1	72	Medium
10:00 AM	Medium	71.6	71.1	72.1	117	High
12:00 PM	High	68.5	68.4	68.6	173	High
3:00 PM	High	-70.23	-70.6	-69.9	-91	Medium

Table 4.5. Yearly errors at different forecasting times using Slope2 method.

	8:00 AM	11:00 AM	2:00 PM
RMSE (W/m^2)	175.5	160.4	118.3
MAPE (%)	67.7	67.9	68.3

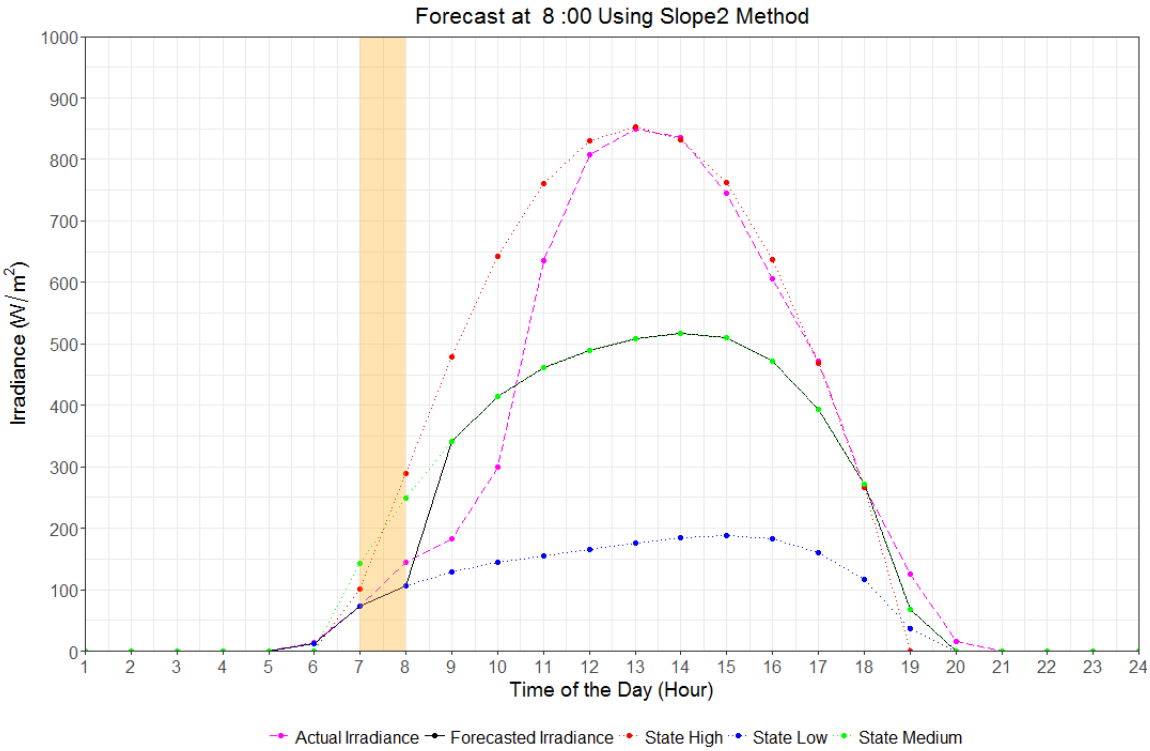

Figure 4.7. Forecasting at 8 am using Slope2 method for August 13, 2011.

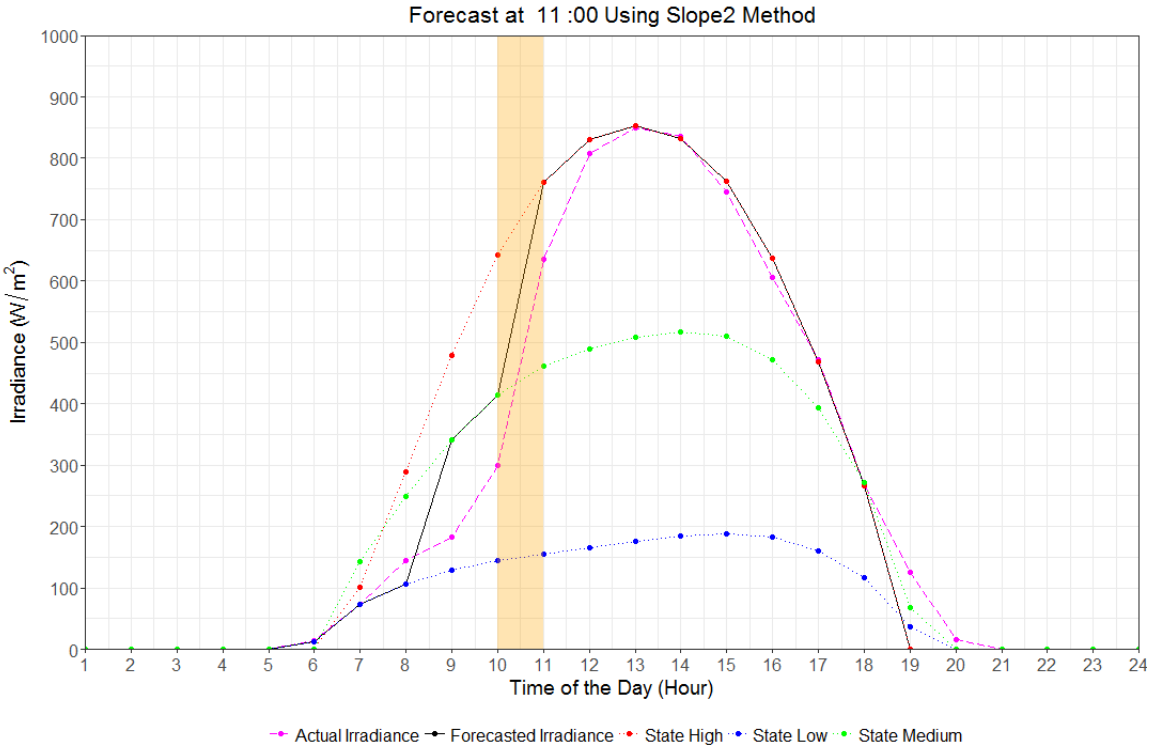

Figure 4.8. Forecasting at 11 am using Slope2 method for August 13, 2011.

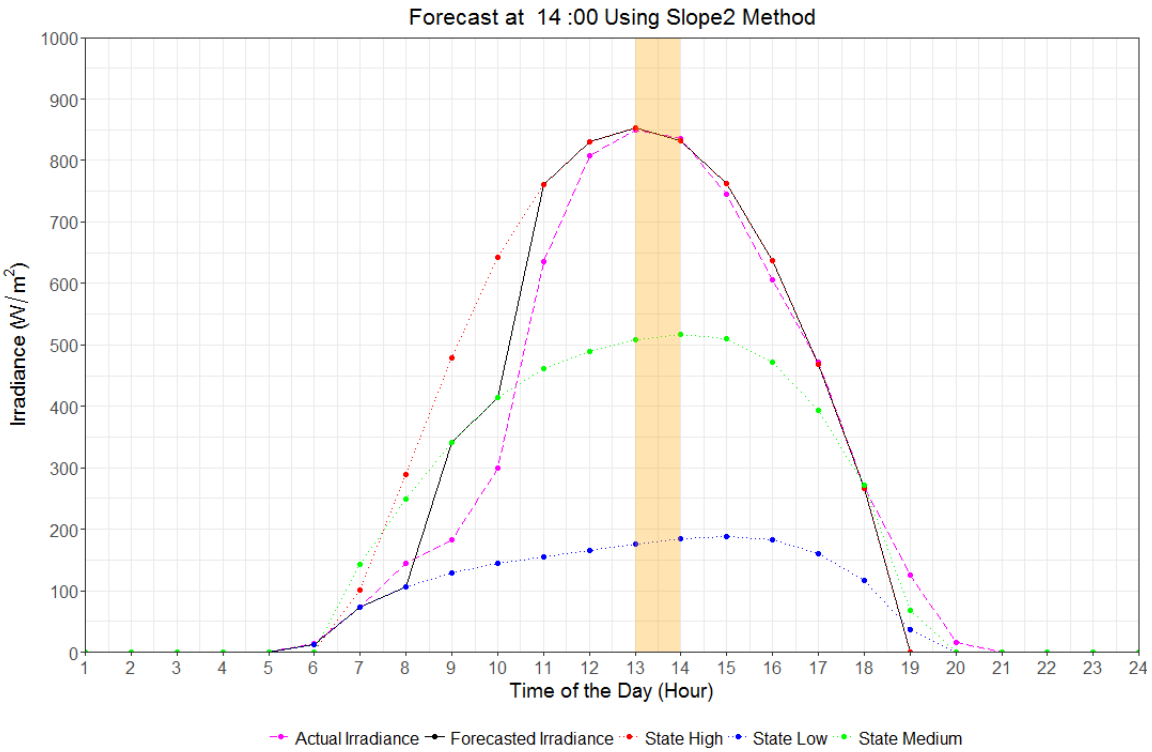

Figure 4.9. Forecasting at 2 pm using Slope2 method for August 13, 2011.

4.1.4 Results from PastHr method

Figures 4.10 to 4.12 give the illustration of intra-day forecasting using the Past Hour method. The first plot shows the forecast made at 8 am. At 8 am, the low state is closest to the actual due to which the following hours remain in the low state. However, at 10 am, it changes to medium and finally to high when forecasted 11 am, as shown by Figure 4.11 and maintains in the same state. The RMSE and MAPE for the forecast made at 8 am were calculated to be 424.8 W/m^2 and 69.3%. Nevertheless, the errors reduced to 44.3 W/m^2 and 23.7% when forecasted at 11 am. However, the errors slightly rose to 53.5 W/m^2 and 34.9% when forecasted at 2 pm. The yearly errors at different forecasting times using this method for the year 2011 are tabulated in Table 4.6.

Table 4.6. Yearly errors at different forecasting times using PastHr method.

	8:00 AM	11:00 AM	2:00 PM
RMSE (W/m^2)	192.2	129.7	92.5
MAPE (%)	62.8	52.6	63.4

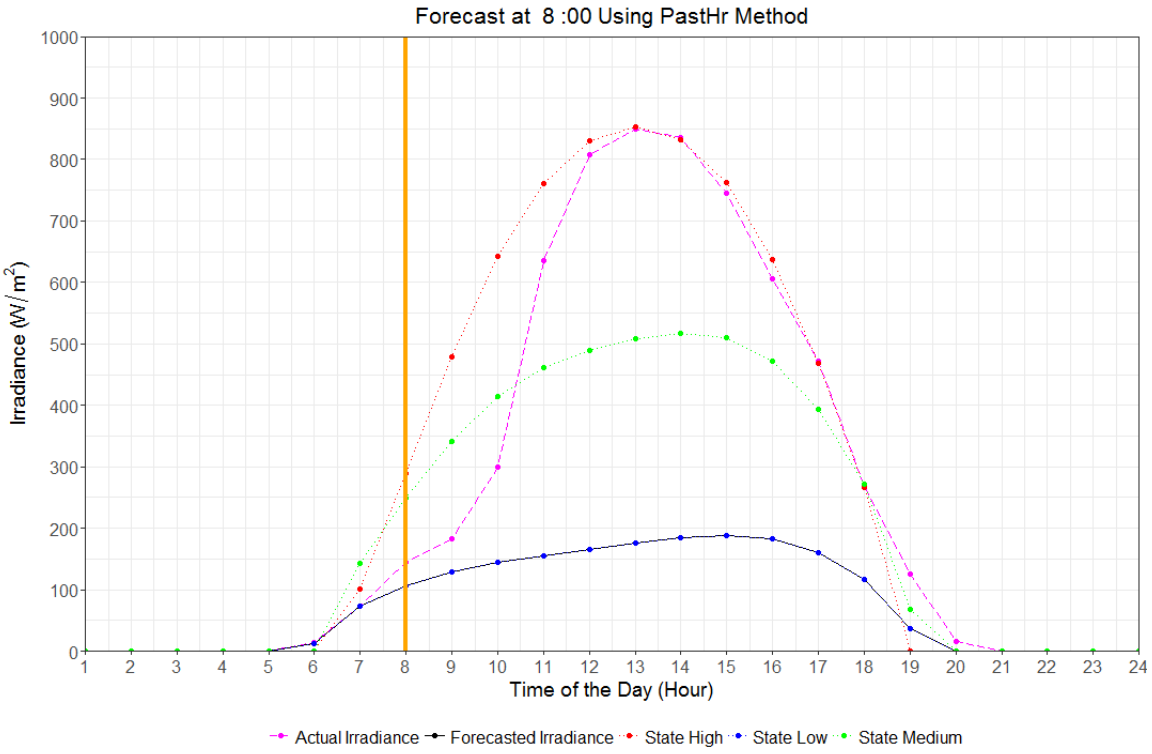

Figure 4.10. Forecasting at 8 am using PastHr method for August 13, 2011.

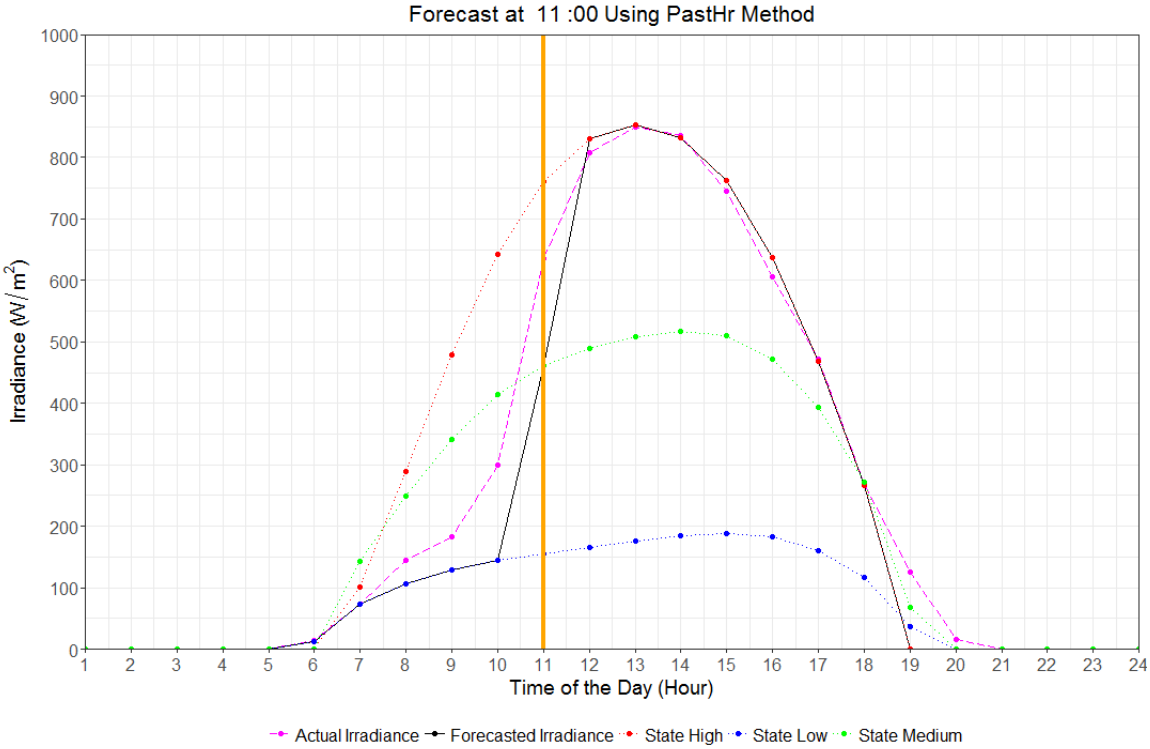

Figure 4.11. Forecasting at 11 am using PastHr method for August 13, 2011.

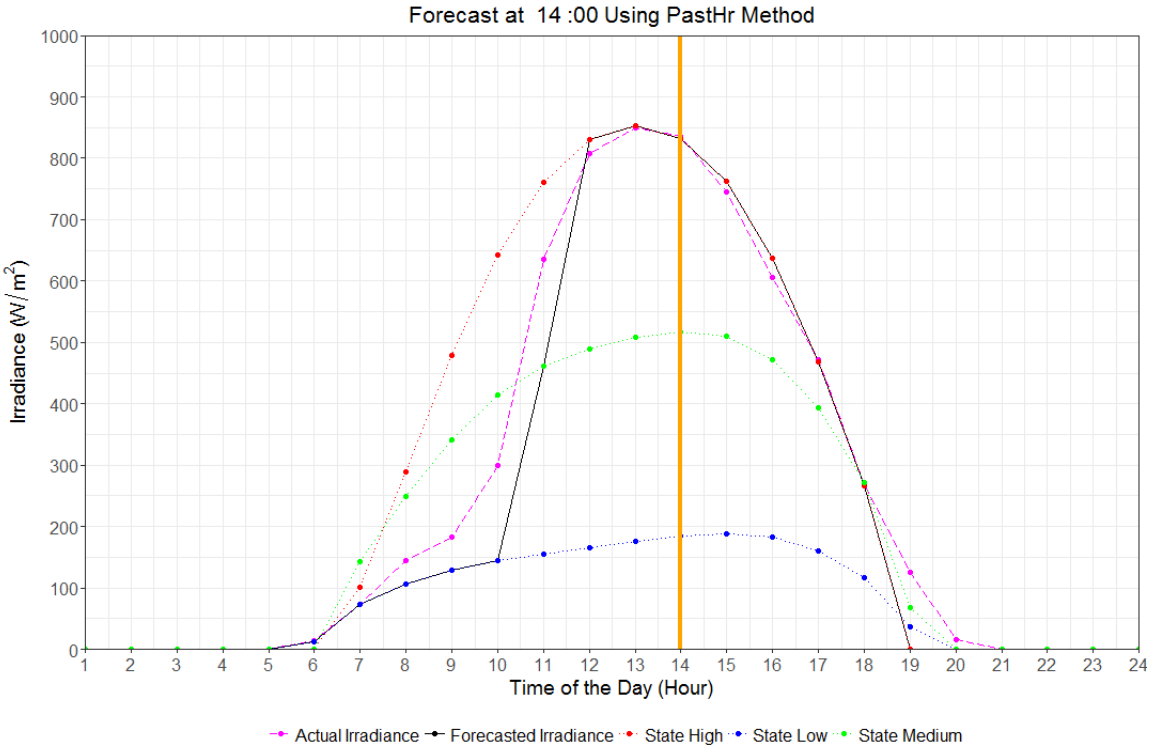

Figure 4.12. Forecasting at 2 pm using PastHr method for August 13, 2011.

4.2 Performance evaluation of forecasting methods

The performance of each forecasting method is evaluated on the basis of (a) the forecasting time of the day, (b) the time of the year, and (c) the consistency over multiple years.

4.2.1 On the basis of forecasting time of the day

The performance of the proposed four methods were evaluated at different forecasting times – 8 am, 11 am, and 2 pm. This was done for the year 2011 because of the availability of complete data points. At first, daylight hours were determined for each day in the year. Then, the forecasted irradiance values beyond the point of forecast to the last daylight hour were extracted for each day to compute the daily RMSE and MAPE for each forecasting time. This can be more clear from the example day in Section 4.1 when the daylight hours are from 5 am to 9 pm. For the forecasting time at 8 am, the forecasted values from 9 am to 9 pm are used to compute the daily errors. Such daily errors were calculated for the forecast made at 11 am and 2 pm in a similar manner for all four methods for each day of the year. The distribution of the daily RMSE for the proposed methods for the forecast made at the three times is illustrated by the violin plots in Figures 4.13 to 4.15. Each violin plot also incorporates a boxplot where the ends of the box represent the first and the third quartile and the middle line gives the median of the daily errors over the year while the black dot represents the mean.

The summary of the plots is tabulated in Table 4.7. It can be noticed that when forecasted early at 8 am, both the average and the median of the distribution of RMSE is the least for the Slope2 method. At both 11 am and 2 pm, however, both the mean and the

median of the daily RMSE distributed over the year is the least for the PastHr method. It can further be deduced that with the progression of the day, the forecasting error is reduced because the forecast is updated to a closer state with each hourly update. Only the Slope1 method deviates from this trend as the average of the forecasting error is greater at 2 pm than at 11 am. Furthermore, it can be inferred that the error rates are more consistent over the year for the forecast in the afternoon. In other words, the variability of error is reduced with the hourly update of the forecast which is supported by the shorter length of the boxplot.

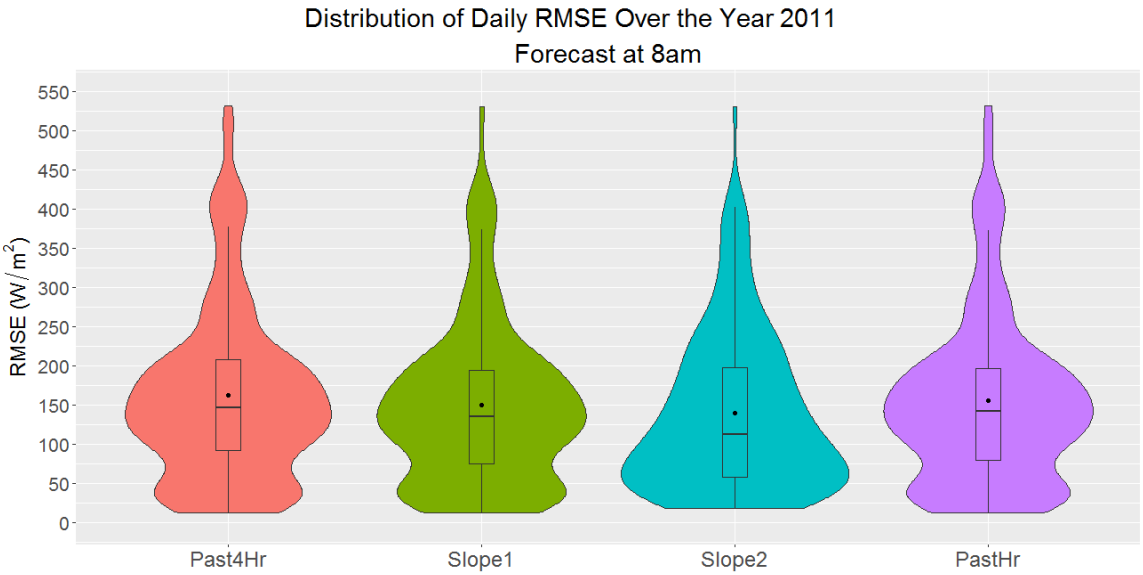

Figure 4.13. Distribution of daily RMSE over the year 2011 for the forecast at 8 am.

Table 4.7. Mean and median of daily RMSE at different forecast times for different methods for the year 2011.

Forecast Time	**8 am**		**11 am**		**2 pm**	
Method	**Mean RMSE** (W/m^2)	**Median RMSE** (W/m^2)	**Mean RMSE** (W/m^2)	**Median RMSE** (W/m^2)	**Mean RMSE** (W/m^2)	**Median RMSE** (W/m^2)
Past4Hr	163.5	147.6	104.3	77.5	79.2	63.7
Slope1	150.9	135.3	106.1	**68.3**	112.5	86.1
Slope2	**140.6**	**112.9**	127.9	107.1	96.1	82.5
PastHr	156.6	142.9	**99.6**	**68.3**	**74**	**61.5**

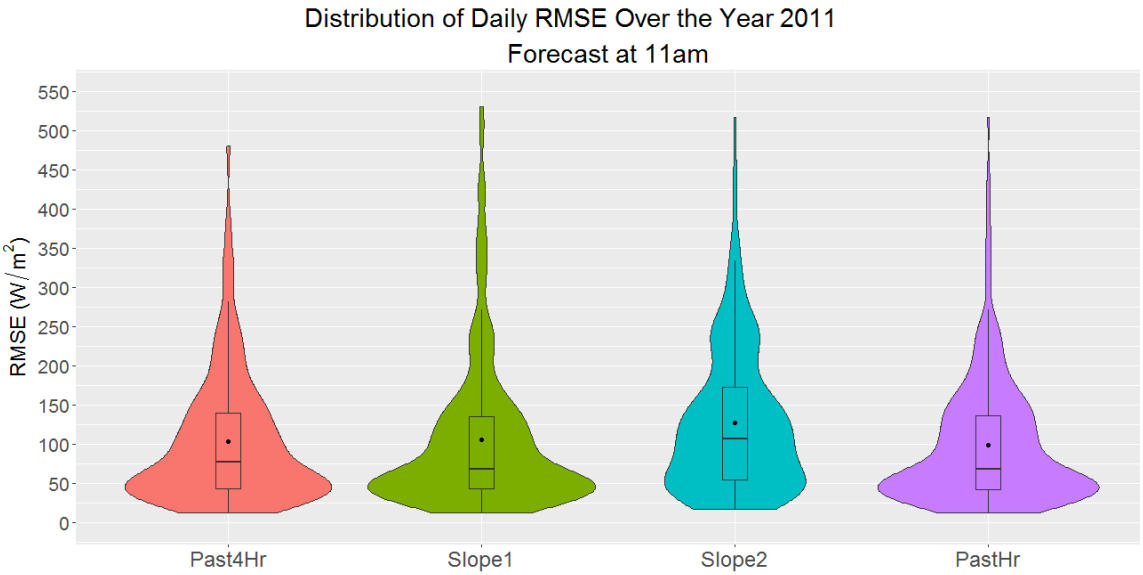

Figure 4.14. Distribution of daily RMSE over the year 2011 for the forecast at 11 am.

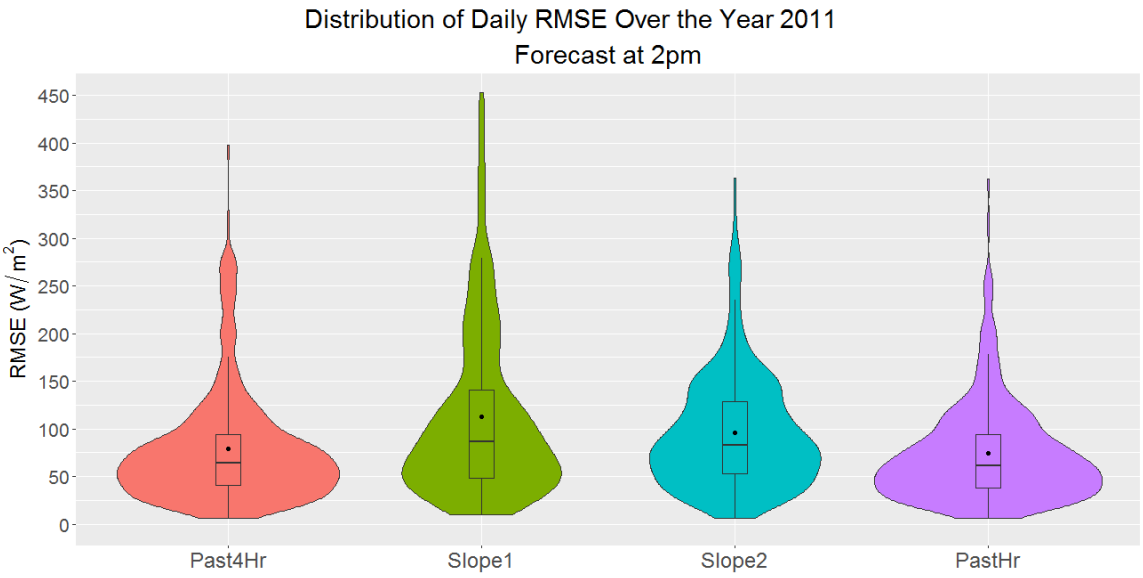

Figure 4.15. Distribution of daily RMSE over the year 2011 for the forecast at 2 pm.

4.2.2 On the basis of the time of the year

This section deals with the performance of each method on a monthly basis at different forecasting times. First, the daily error rates are calculated in the same way as discussed in Section 4.1. Then, the distribution of the daily errors by month is evaluated using box plots for each method. Figure 4.16 shows such a distribution of daily RMSE over the months for the forecast made at 8 am. For earlier forecast like this, the daily errors are higher and spread out for each month for all the methods. This is because the first few hours of the day are not always the best representation of the following hours. As can be seen from Figure 4.17, daily RMSEs are significantly reduced in most of the months for the forecast at 11 am. This is supported by the lower median values indicated by the middle bar within the box plot. Also, the forecasting is more consistent over the month, which is indicated by the shorter box lengths. However, the forecasting is more variable and with a higher error during the summer months (May to July) for all the forecasting methods. With further update at 2 pm, as shown in Figure 4.18, the errors are further reduced with lower median values and less variable for most months.

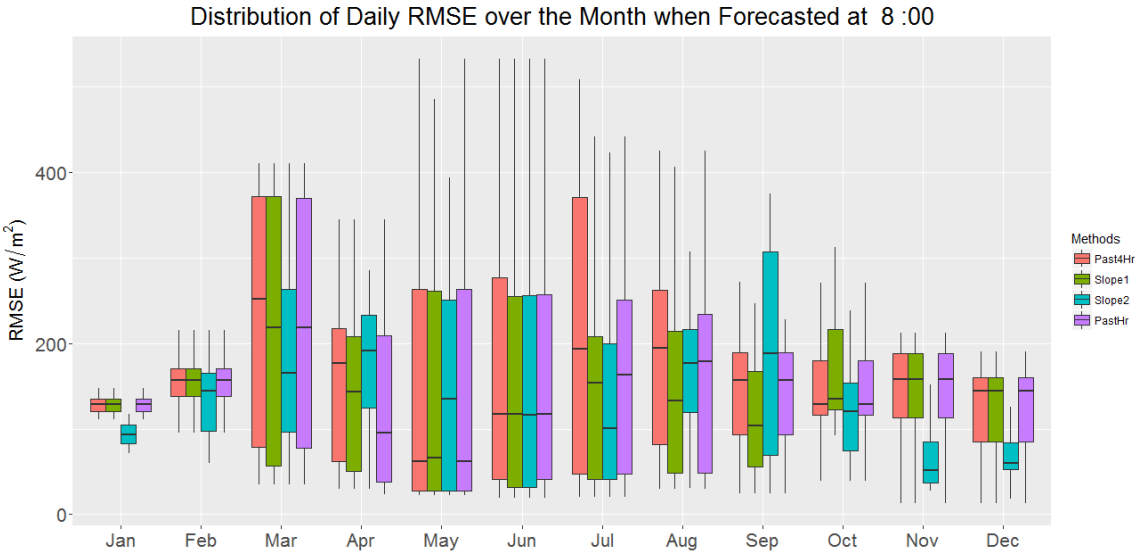

Figure 4.16. Distribution of daily RMSE over the month when forecasted at 8 am.

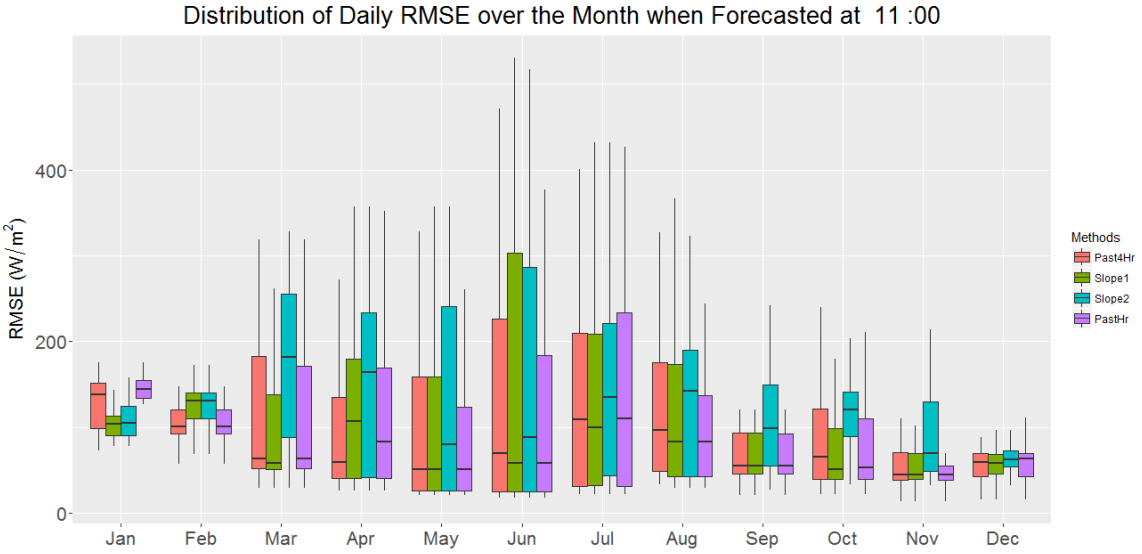

Figure 4.17. Distribution of daily RMSE over the month when forecasted at 11 am.

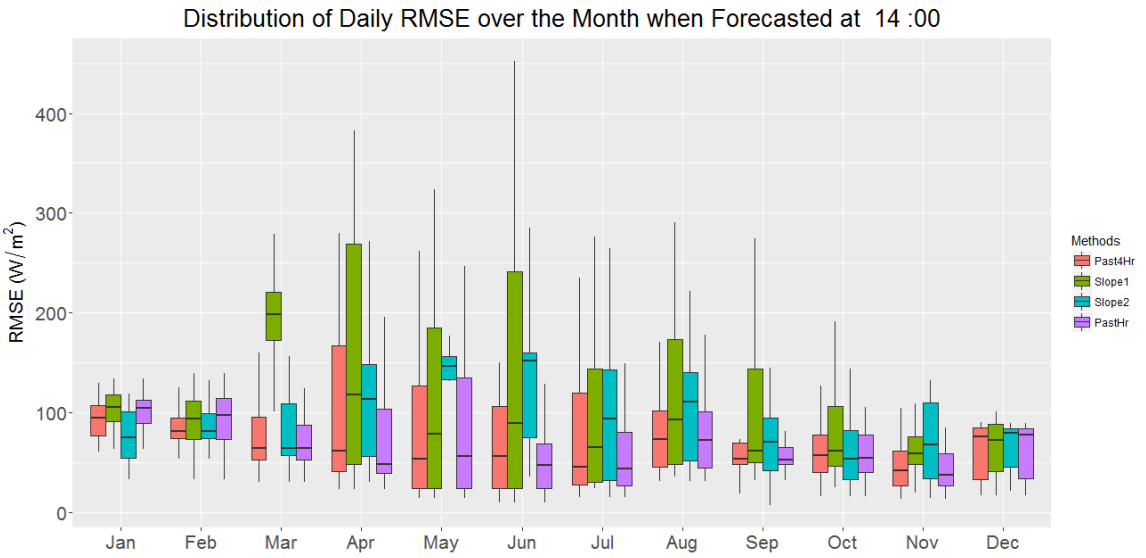

Figure 4.18. Distribution of daily RMSE over the month when forecasted at 2 pm.

It can also be inferred that the interquartile range of errors is larger for the Slope 1 method, mostly during the months of April to August. This is because of two main reasons. First, the Slope 1 method is not very efficient in predicting the correct state when there is a sudden rise or fall in irradiance as in a cloudy day in summer as shown in Figure 4.19. For instance, in Figure 4.20, the actual irradiance shown by the magenta curve is suddenly falling from 1 pm to 2 pm. While the level of irradiance within that interval is closer to the low state, only the high state has a falling slope similar to that of the actual irradiance. Because this method chooses the state with the slope closest to that of the actual irradiance, the high state is wrongly selected causing a large RMSE of 427.6 W/m^2 and a MAPE of 420.2 %. Secondly, during summer seasons, the difference in the level of irradiance between each state is larger. Due to this, the selection of a wrong state brings about a larger error than it would in other seasons. This tendency to pick the incorrect state is probable also in the Slope 2 method. This is because the method relies on the irradiance data from the past 4 days, which may not always be in favor of taking the

correct decision. For this method, the RMSE and MAPE for the forecast at 2 pm are 47.9 W/m^2 and 62.9 %, respectively. The Past4Hr and PastHr methods, however, deal with this issue more appropriately as shown in Figure 4.21. Both these methods update the hourly states based on the most recent data from the same day. Due to this, the selection of state is more accurate with an RMSE of 47.9 W/m^2 and a MAPE of 62.9 % for both the methods.

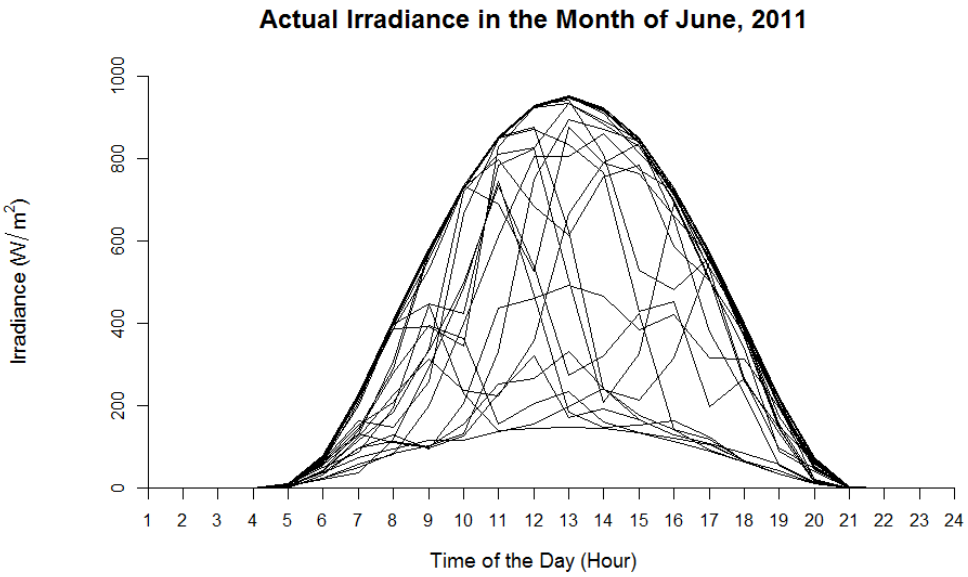

Figure 4.19. Irregular daily irradiance pattern in the month of June.

4.2.3 On the basis of consistency over multiple years

The performance of each of the models are further validated and compared for different years. For this purpose, the Pearson's coefficient of correlation is used. The coefficient of correlation is ranged between -1 to +1 where a higher positive value means a stronger correlation between the actual irradiance and the forecasted irradiance.

In this thesis, the irradiance forecasts are updated every hour. In other words, it can be said that solar irradiance is forecasted one hour ahead of time. Thus, the overall

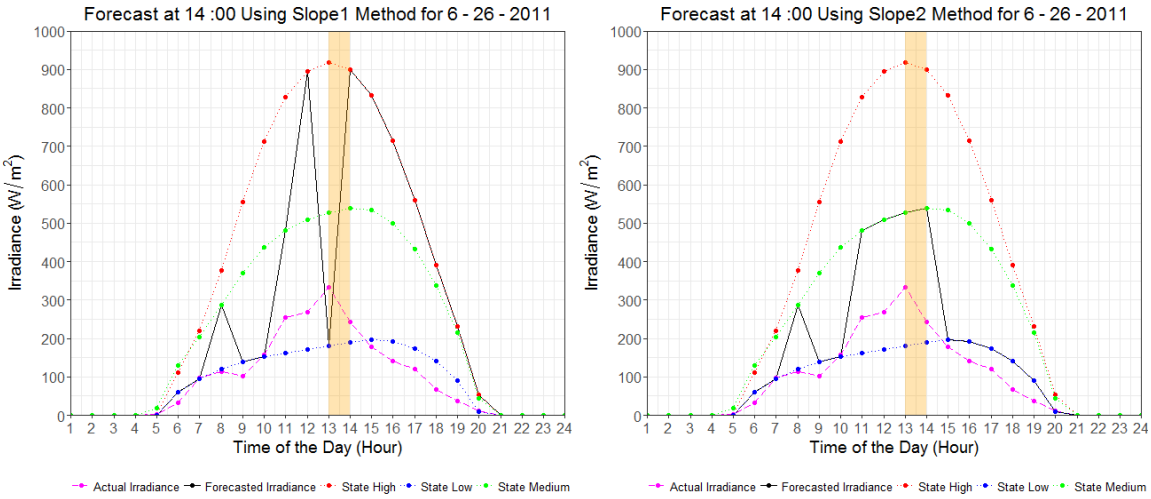

Figure 4.20. Larger error on a cloudy day in summer by Slope 1 and Slope 2 methods.

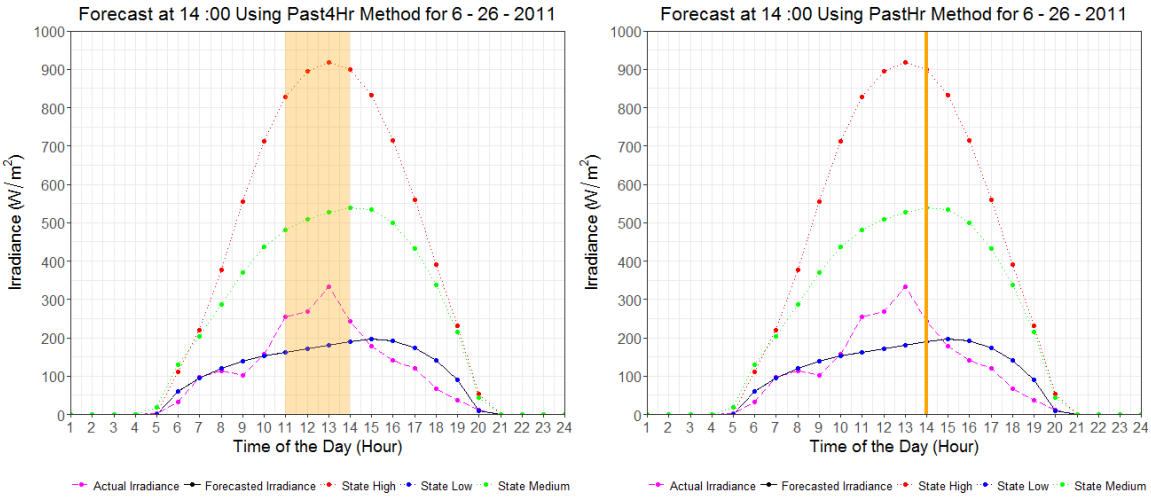

Figure 4.21. Smaller error on a cloudy day in summer by Past4Hr and PastHr methods.

forecast for the day is the time series of the irradiance forecasts developed until the last daylight hour. Due to this, the energy states selected every hour until the last daylight hour is considered as the overall forecast for the day, which is compared with the actual irradiance for that day. In this way, the correlation between the hourly forecasts over a year is compared with the hourly actual irradiance for the same year. For instance, Figure 4.22 compares the hourly actual irradiance and hourly forecast over the year 2011. The scatter plots of the forecasted irradiance versus the actual irradiance for the Past4Hr and the PastHr methods are more concentrated as opposed to the more dispersed plot for the slope methods, thus demonstrating a stronger correlation of the forecast with the actual measurements. Pearson's coefficient of correlation was computed for several years for all the methods to assess the consistency in forecasting over multiple years.

Figure 4.23 shows a bar chart illustrating the coefficient of correlation between the actual irradiance and the forecast for all four methods for the years 2000 to 2007 and 2011 (years except the ones used for training). As can be seen, all four methods have the correlation coefficient values that remain consistent over the years they were tested for. Therefore, the performance of all methods are consistent over multiple years.

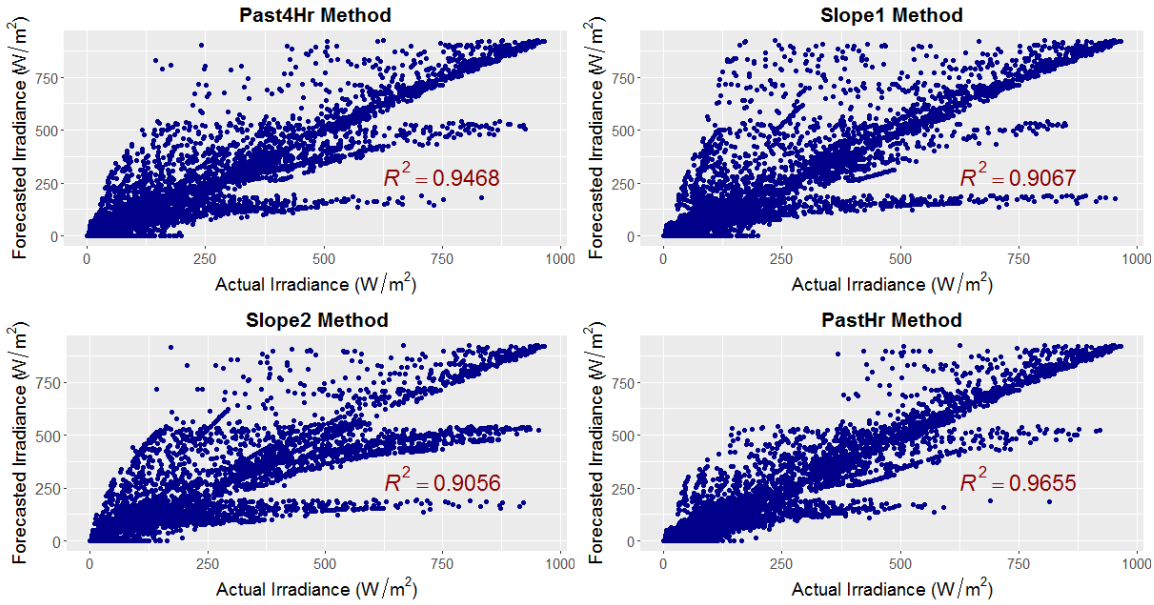

Figure 4.22. Scatter plots of forecasted irradiance vs. actual irradiance for the year 2011 for the proposed methods.

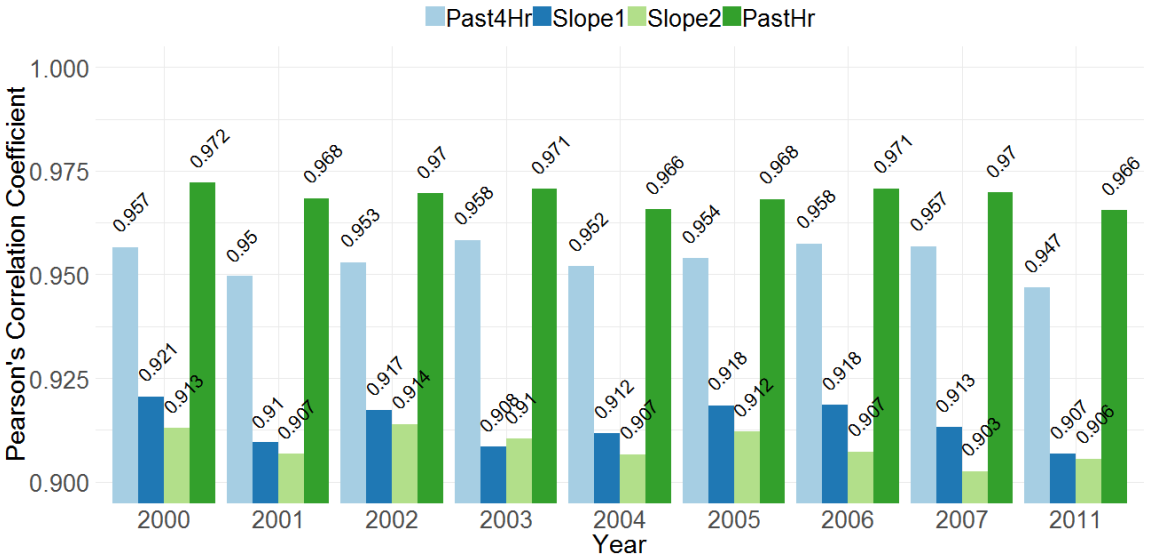

Figure 4.23. Comparison of Pearson's coefficients of correlation for multiple years.

CHAPTER 5 CONCLUSIONS

5.1 Conclusions

This work is an extension of the solar irradiance forecasting method using the HMM [1]. This thesis explores the accuracy of four different methods of intra-day forecasting that use the energy states from the HMM used as the input. All the methods are based on the HMM to generate three energy states corresponding the cloud coverage conditions and hence different irradiance levels. The input to the HMM are only the past irradiance data, clear sky index, and the Fourier basis expansions – all of which do not require complex instrument and large datasets. The Past4Hr method and the PastHr method, as the name suggests, directly utilizes the past irradiance data from the same day to select a state one hour ahead of time. While the Past4Hr method uses the least squared error of the irradiance data from the past 4 hours, the PastHr only makes use of the immediate past irradiance to find the closest state. On the other hand, the Slope1 and Slope2 methods use the irradiance data from the past two hours to compute the first derivative of the irradiance values. Using two different algorithms, the rise and fall of irradiance are detected to select the state closest to the actual irradiance as the forecast for the next hour.

The past irradiance data from Brookings, South Dakota was used to validate each of the forecasting methods. The year 2011 was chosen for validation because of the availability of a complete dataset. The RMSE and MAPE were used as error metrics for validation. It was found for all the methods that the error rates reduced with the hourly update of the forecast. Three forecasting times – 8 am, 11 am, and 2 pm, were considered

to evaluate the annual performance of the proposed methods on the basis of the forecasting time. The Slope2 method worked best for the forecast at 8 am while the PastHr method stood out from the rest of the methods for the remaining forecasting times. Monthly distribution of daily errors was graphed into violin plots for each method to evaluate their performance according to the time of the year. Again, the PastHr method performed consistently better than the rest for most of the months. Nevertheless, a relatively higher error was observed during summer months, particularly for the Slope2 method. This is because the algorithm that the method used could not accommodate for the sudden rise or fall of the irradiance in the proper selection of the state. This matters significantly during summer because of larger irradiance difference between the consecutive states. In other words, wrong selection of state by even one level results in a huge error. Similarly, the consistency in the performance of the proposed methods was also evaluated using the correlation between the hourly forecasted values and the actual irradiance values. Pearson's correlation coefficient was computed for the years 2000 to 2007 and 2011, which excludes the years used for training (2008 to 2010). All four methods performed consistently as supported by the approximately equal correlation coefficient values for all the years considered. Taking into account the simplicity of the method, lower forecasting error, and the requirement of minimal input dataset for modeling, a hidden Markov model with PastHr method for solar irradiance forecasting can be used as an effective measure to forecast solar irradiance for remote microgrids.

5.2 Future Work

In this work, it was observed that hourly updating the irradiance forecast would result in better accuracy. At the same time, it was also discerned that the technique used for intra-day forecasting in this thesis could not accommodate the sudden rise or fall of irradiance. Instead of making switches between the estimated states or energy levels, a more advanced approach of intra-day forecasting using machine learning techniques could be implemented. This would also help solve the issue of higher error in the summer months due to selection of the state that is further from the actual value. Furthermore, the forecasted solar irradiance data could be used in a microgrid testbed to run the optimization for the EMS of the remote microgrid.

REFERENCES

[1] A. Shakya, S. Michael, C. Saunders, D. Armstrong, P. Pandey, S. Chalise, and R. Tonkoski, "Solar irradiance forecasting in remote microgrids using Markov switching model," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 3, pp. 895–905, 2017.

[2] "Energy access outlook 2017: From poverty to prosperity," International Energy Agency (IEA), 2018.

[3] D. T. Ton and M. A. Smith, "The US department of energy's microgrid initiative," *The Electricity Journal*, vol. 25, no. 8, pp. 84–94, 2012.

[4] J. Wilt, *Canada's commitment of $220 million to transition remote communities off diesel a mere 'drop in the bucket'*, Accessed: March 8, 2019. [Online]. Available: https://thenarwhal.ca/canada-s-commitment-220-million-transition-remote-communities-diesel-mere-drop-bucket/.

[5] S. Pelland, D. Turcotte, G. Colgate, and A. Swingler, "Nemiah valley photovoltaic-diesel mini-grid: System performance and fuel saving based on one year of monitored data," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 1, pp. 167–175, 2012.

[6] S. Chalise, F. B. Dos Reis, J. Sternhagen, and R. Tonkoski, "Power management strategies for microgrids with high penetration of renewables," in *Fifth International Conference on Power and Energy Systems*, IEEE, 2013, pp. 1–6.

[7] P. Tandukar, L. Bajracharya, T. M. Hansen, R. Fourney, U. Tamrakar, and R. Tonkoski, "Real-time operation of a data center as virtual power plant considering battery lifetime," in *SPEEDAM 2018 – Proceedings: International Symposium on Power Electronics, Electrical Drives, Automation and Motion*, IEEE, 2018, pp. 81–86.

[8] S. Chalise, J. Sternhagen, T. H. Hansen, and R. Tonkoski, "Energy management of remote microgrids considering battery lifetime," *The Electricity Journal*, vol. 29, no. 6, pp. 1–10, 2016.

[9] E. Lorenz, T. Scheidsteger, J. Hurka, D. Heinemann, and C. Kurz, "Regional PV power prediction for improved grid integration," *Progress in Photovoltaics: Research and Applications*, vol. 19, no. 7, pp. 757–771, 2010.

[10] L. Wen, K. Zhou, S. Yang, and X. Lu, "Optimal load dispatch of community microgrid with deep learning based solar power and load forecasting," *Energy*, vol. 171, pp. 1053–1065, 2019.

[11] A. Shakya, S. Michael, C. Saunders, D. Armstrong, P. Pandey, S. Chalise, and R. Tonkoski, "Using Markov switching model for solar irradiance forecasting in remote microgrids," in *2016 IEEE Energy Conversion Congress and Exposition (ECCE)*, IEEE, 2016, pp. 1–7.

[12] C. Feng, M. Cui, B. M. Hodge, S. Lu, H. Hamann, and J. Zhang, "Unsupervised clustering-based short-term solar forecasting," *IEEE Transactions on Sustainable Energy*, vol. PP, 2018.

[13] F. Wang, Z. Zhen, C. Liu, Z. Mi, B. M. Hodge, M. Shafie-khah, and J. P. Catalão, "Image phase shift invariance based cloud motion displacement vector calculation method for ultra-short-term solar PV power forecasting," *Energy Conversion and Management*, vol. 157, pp. 123–135, 2018.

[14] E. Akarslan, F. O. Hocaoglu, and R. Edizkan, "Novel short term solar irradiance forecasting models," *Renewable Energy*, vol. 123, pp. 58–66, 2018.

[15] H. Sheng, J. Xiao, Y. Cheng, Q. Ni, and S. Wang, "Short-term solar power forecasting based on weighted gaussian process regression," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 1, pp. 300–308, 2018.

[16] J. A. Augustine, J. J. DeLuisi, and C. N. Long, "SURFRAD—a national surface radiation budget network for atmospheric research," *Bulletin of the American Meteorological Society*, vol. 81, no. 10, pp. 2341–2358, 2000.

[17] S. D. Miller, M. A. Rogers, J. M. Haynes, M. Sengupta, and A. K. Heidinger, "Short-term solar irradiance forecasting via satellite/model coupling," *Solar Energy*, vol. 168, pp. 102–117, 2018.

[18] Y. Wang, Y. Shen, S. Mao, X. Chen, and H. Zou, "LASSO and LSTM integrated temporal model for short-term solar intensity forecasting," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2933–2944, 2018.

[19] H. Bouzgou and C. A. Gueymard, "Fast short-term global solar irradiance forecasting with wrapper mutual information," *Renewable Energy*, vol. 133, pp. 1055–1065, 2019.

[20] M. Hassan and B. Nath, "Stock market forecasting using hidden Markov model: A new approach," in *5th International Conference on Intelligent Systems Design and Applications (ISDA'05)*, IEEE, 2005, pp. 192–196.

[21] A. M. González, A. M. n. S. Roque, and J. García-González, "Modeling and forecasting electricity prices with input/output hidden Markov models," *IEEE Transactions on Power Systems*, vol. 20, no. 1, pp. 13–24, 2005.

[22] P. Jiang, X. Liu, J. Zhang, and X. Yuan, "A framework based on hidden Markov model with adaptive weighting for microcystin forecasting and early-warning," *Decision Support Systems*, vol. 84, pp. 89–103, 2016.

[23] Y. Liu, L. Ye, H. Qin, X. Hong, J. Ye, and X. Yin, "Monthly streamflow forecasting based on hidden Markov model and gaussian mixture regression," *Journal of Hydrology*, vol. 561, pp. 146–159, 2018.

[24] T. Wang, J. Zhuang, J. Buckby, K. Obara, and H. Tsuruoka, "Identifying the recurrence patterns of nonvolcanic tremors using a 2-d hidden Markov model with extra zeros," *Journal of Geophysical Research: Solid Earth*, vol. 123, no. 8, pp. 6802–6825, 2018.

[25] S. Yun, C.-S. Son, S.-H. Lee, and W.-S. Kang, "Forecasting of heart rate variability using wrist-worn heart rate monitor based on hidden Markov model," in *2018 International Conference on Electronics, Information, and Communication (ICEIC)*, IEEE, 2018, pp. 1–2.

[26] R. Tonkoski, "Impact of high penetration of photovoltaics on low voltage systems and remedial actions," Ph.D. dissertation, Concordia University, Montréal, Québec, Canada, 2011.

[27] M. D. Bachi, "Economic dispatch and demand side management in diesel hybrid mini-grids," Master's thesis, Concordia University, Montréal, Québec, Canada, 2012.

[28] B. Kroposki, K. Burman, J. Keller, A. Kandt, J. Glassmire, and P. Lilienthal, "Integrating high levels of renewables into the Lanai electric grid," National Renewable Energy Lab (NREL), Golden, CO, 2012.

[29] F. Katiraei, R. Iravani, N. Hatziargyriou, and A. Dimeas, "Microgrids management," *IEEE power and energy magazine*, vol. 6, no. 3, pp. 54–65, 2008.

[30] C. Nemes, "A clear sky irradiation assessment using the European Solar Radiation Atlas model and Shuttle Radar Topography Mission database: A case study for Romanian territory," *Journal of Renewable and Sustainable Energy*, vol. 5, no. 4,

[31] A. B. Meinel and M. P. Meinel, *Applied Solar Energy: An Introduction*, 3rd ed. Addison-Wesley, 1977.

[32] J. A. Duffie and W. A. Beckman, *Solar Engineering of Thermal Processes*, 4th ed. John Wiley and Sons, 2013.

[33] R. A. Messenger and J. Ventre, *Photovoltaic Systems Engineering*, 2nd ed. CRC Press, 2004.

[34] E. W. Weisstein, *Fourier series*, Accessed: Mar. 7, 2019. [Online]. Available: http://mathworld.wolfram.com/FourierSeries.html.