1-1-2019

Solar Forecasting and Integration in Power Systems

Mohana Alanazi
University of Denver

SOLAR FORECASTING AND INTEGRATION IN POWER SYSTEMS

_____

A Dissertation

Presented to

the Faculty of the Daniel Felix Ritchie School of Engineering and Computer Science

University of Denver

_____

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

_____

by

Mohana Alanazi

August 2019

Advisor: Dr. Amin Khodaei

Author: Mohana Alanazi
Title: SOLAR FORECASTING AND INTEGRATION IN POWER SYSTEMS
Advisor: Dr. Amin Khodaei
Degree Date: August 2019

Abstract

Renewable energy resources are becoming critical players in the electricity generation sector, primarily due to viability in combating global warming, effectiveness in reducing pollution caused by fossil fuel based generation, and diversifying energy mix to ensure energy security and sustainability. Solar energy is one of the most common types of renewable energy that has grown rapidly over the past decade and is anticipated to grow even faster in the future. Power supply from renewable resources is forecasted to surpass other types of generation in a foreseeable future. Numerous factors, including but not limited to the dropping cost of solar technology, environmental concerns, and the state and governmental incentives, have made the path for a rapid growth of solar generation. However, increased generation from renewable resources exposes the power system to more vulnerabilities, conceivably due to their variable generation, thus highlighting the importance of accurate forecasting methods. An accurate solar forecasting method, which takes into account generation variability and is able to identify associated uncertainty, can support a reliable and cost-effective deployment. More and more large-scale solar PV farms are expected to be integrated in the existing grids in the foreseeable future in compliance with the energy sector renewable portfolio standards (RPS) in different states and countries. The integration of large-scale solar PV into power systems, however, will necessitate a system upgrade by adding new generation units and transmission lines.

This dissertation proposes a forecasting model that aims to enhance the forecasting result and reduce errors. The proposed model utilizes a new approach to overcome some of significant challenges in solar generation forecasting. The model includes different data processing stages in order to ensure the quality of the data before it is fed to the forecasting tool. The model undergoes further enhancement such as forecasting methods combination and multilevel measurements application. Numerical simulations exhibit the merits of the proposed method through testing under different weather conditions and case studies. Moreover, a co-optimization generation and transmission planning model is proposed to maximize large-scale solar PV hosting capacity. The solution of this model further determines the optimal solar PV size and location, along with potential required PV energy curtailment. Numerical simulations study the proposed co-optimization planning problem with and without considering the solar PV integration and exhibit the effectiveness of the proposed model.

Acknowledgements

I would like to extend my sincere gratitude and appreciation to my advisor Dr. Amin Khodaei for his valuable guidance and endless support throughout my Ph.D. studies. Without his guidance and persistence support, this dissertation would not have been possible. Dr. Khodaei has been a source of inspiration and his motivation and immense knowledge has not only helped me in my research skills but also in my daily life skills. I cannot say thank you enough for your tremendous support and valuable advices.

I would also like to thank my PhD committee members, Dr. Andy Goetz, Dr. Wenzhong Gao, and Dr. Mohammed Matin for serving as my committee members and for their time, brilliant comments and suggestions. Also, I would like to thank my colleagues and lab mates for their support throughout my PhD studies for helping me to strive towards my goal.

A special thanks to my family; my parent, brothers and my sisters. Words cannot express how grateful I am. Your support and prayers for me was what sustained me thus far. To my wife (Hala) and my kids (Yasser, Yazeed, Saif, and Eleen), I am very grateful for your continued patience, sacrifice, and understanding while I spent countless hours working in my research. I dedicate this work to you all.

Also, I cannot forget to express my great appreciation to the Al-Jouf University which provided the opportunity to pursue my studies overseas through the funding and support. Finally, I would like to thank the Saudi Arabian Cultural Mission (SACM) for their support during my study period in the United States.

Table of Contents

List of Figures

List of Tables

Nomenclature

Chapter Two:

GHI_{hist}	Hourly historical global horizontal irradiance
GHI_{CSK}	Hourly historical clear sky
$GHI_{forecast}$	Forecasted global horizontal irradiance
GHI_{dev}	The difference between GHI_{hist} and GHI_{CSK}
GHI_{tNorm}	Normalized Fitted dataset
GHI_{norm}	Normalized GHI dataset
GHI_{denorm}	Denormalized GHI dataset
T	Index for year
H	Index for hour
N	Number of samples
w, v	Weights between neurons

Chapter Three:

GHI_{actual}	Target global horizontal irradiance
$\overline{GHI}_{actual}$	Average target GHI
$GHI_{forecast}$	Forecasted global horizontal irradiance
N	Number of GHI values
q	Back shift operator
t	Index for hour
$u(k)$	ARMAX input
$v(k)$	Disturbance/error
$y(k)$	ARMAX output
Y_t	Detrended solar time series

Chapter Four:

P_{actual}	Actual solar generation
$P_{forecast}$	Forecasted solar generation
$\overline{P}_{actual}$	Average actual solar generation
N	Number of samples
E_C	Calculated error at customer level
E_F	Calculated error at feeder level
E_S	Calculated error at substation level

Chapter Five:

Indices:

t	Index for year
h	Index for hour

q	Index for day
s	Superscript for solar PV
b,m,n	Index for bus
j	Index for generation units, and transmission lines
E	Index for existing units and lines
C	Index for candidate units and lines
$\sim$	Index for forecasted parameters
$\wedge$	Index for known variables

Sets:

EG	Set of existing generation units
EL	Set of existing transmission lines
CL	Set of candidate transmission lines
CG	Set of candidate generation units
G_b	Set of generation units connected to bus b
L_b	Set of transmission lines connected to bus b

Parameters:

B	Bus-line incidence matrix
X	Transmission line reactance
IC	Capital cost
d	Discount rate
L	Large positive number
D	Load demand
CT	Construction and commissioning time
RU	Generation unit ramp up rate
RD	Generation unit ramp down rate
c	Marginal generation cost
α	Normalized forecasted solar output
ε	Small positive and predefined threshold
δ	Target annual solar PV generation curtailment

Variables:

P	Unit generation output
P^s	Solar PV capacity
PL	Line flow
SL_1, SL_2	Nonnegative slack variables
x	Investment state
θ	Voltage angle
λ, π	Dual variable

Chapter Six:

RR	Ramp Rate
$P(t)$	Power at time t
Δ_t	Time difference

Chapter One: Introduction

Renewable generation has become a viable source that can provide sustainable and inexpensive supply of electricity, due to significant technological advances and many local and national incentives. However, generation from renewable resources has confronted variety of challenges, mainly because of the inherently variable generation. Such variability is caused by various climatic parameters such as temperature, air pressure, cloudiness, etc. [1]. An accurate forecast of generation of these resources will provide the power system operator with the ability to track and regulate any generation variability form variable resources through dispatching controllable generation resources in a coordinated fashion [1].
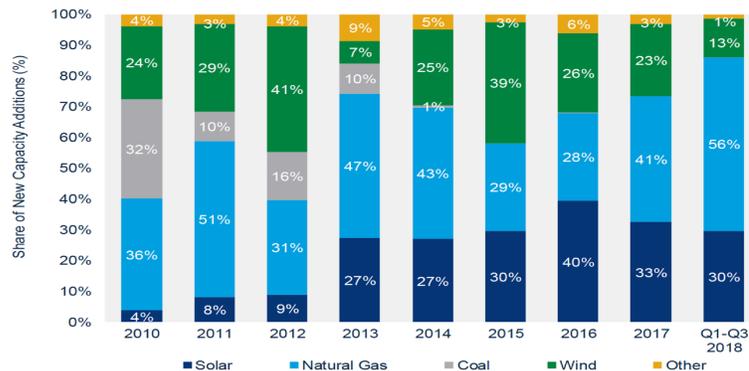
Figure 1.1 The new added U.S. electric generation from 2010 to 2018 [2].

The growth in the amount of generation from renewable energy resources has been unprecedented. This growth is driven by the environmental concerns associated with CO_2 emissions and the global warming, as well as the state and governmental support of

1

renewable resources, combined with the falling cost of the renewable energy technology. As shown in Fig. 1, 30% of the newly added generation capacity in the U.S. in the first three quarters of 2018 came from solar with a total of 6.5 GW, 51% of it being utility-scale PV. The cumulative installed solar PV in the U.S. is currently at 60 GW and this number is expected to double over the next four years. However, the generation variability of these resources, especially at high levels of penetration, may cause an inimical effect on power system operation.

One of the main challenges is therefore to perform as accurate as possible forecasts to overcome potential variability and uncertainty of renewable generation [1]. Variability refers to constant fluctuations in renewable generation while the uncertainty refers to inability to achieve a fully accurate forecast including the time, the duration, and the magnitude of variability. Fig. 1.2 depicts the difference between the variability and the uncertainty. As shown in this figure, variability is associated with actual data while uncertainty is related to forecasted data [2].
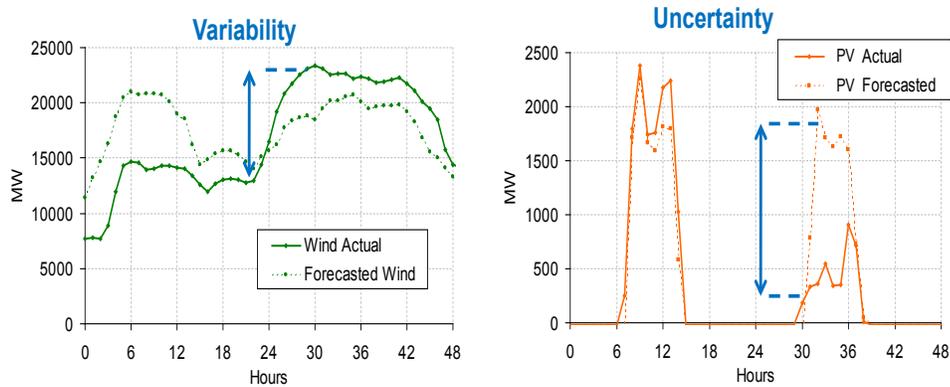
Figure 1.2 Difference between variability and uncertainty [3].

Renewable energy forecasting takes a pivotal role in this situation on the reliable and cost-effective power system operation. A forecast with acceptable accuracy will

2

reduce the uncertainty and provide system operators with the ability to plan ahead and control any fluctuations form renewable generation by increasing or decreasing generation from controllable generation units in a more efficient way.
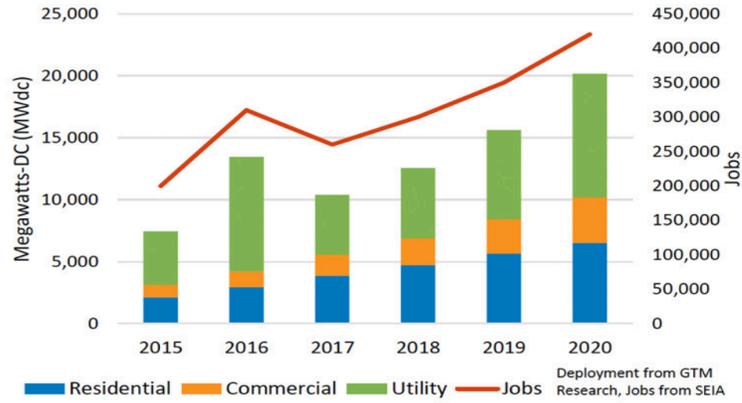
Figure 1.3 U.S. solar energy deployment [4].

In December 2015, legislation was singed to extend the solar Investment Tax Credit (ITC) through 2020 in the Unites States. The ITC extension will result in more than 72 GW of PV to be deployed from 2016 through 2020. Fig. 1.3 depicts the projected solar PV deployment in this timeframe [4]. With such increase in solar generation installations, a proper forecasting is needed to help power system operators safely integrate solar generation and accordingly optimize electricity production and system management.

1.1 Solar Forecasting Challenges and Techniques

Forecasting is not only essential for variable generation, but also useful in load forecasting. In addition, some energy economics quantities such as the electricity price should be forecasted to help with grid's operation, maintenance, and planning [5]. An accurate renewable generation forecast will provide benefits by (i) Minimizing penalties and charges due to imbalance of generated power, (ii) Providing a good knowledge of

3

future energy market trading, and (iii) Helping to carry out reliable operation and maintenance planning [6].

1.1.1 Solar Forecasting Challenges

The current forecasting methods have confronted variety of challenges that are the source of high forecasting error. When comparing the load forecasting errors to solar forecasting errors, it is clear that the solar forecasting is less accurate due to several reasons: First, the time series of solar irradiance is less predictable compared to load forecasting. This is because of the non-stationarity nature of the solar data. The non-continuity of the solar data pattern due to weather changes has imposed significant limitations to forecasting models [7]. So, during the clear sky conditions (sunny) as in Fig.1.4 (a), it is obvious that the patterns are noticeable and the forecasting error is less in these conditions. However, if the weather conditions vary, the pattern of the time series is hardly predictable as in Fig. 1.4 (b).

The second challenge that imposes limitations to forecasting model is the change of daytime hours from one day to another during the forecasting horizon (i.e., the sunset and the sunrise). As shown in Table 1.1, the daytime hours change every day. This will impact the pattern of the time series and hence increase the error relatively. The third challenge is the lack of long-term historical solar irradiance data. The long-term forecasting usually requires a long range of dataset to be trained so the model can extract patterns of the time series.

4

(a)

(b)
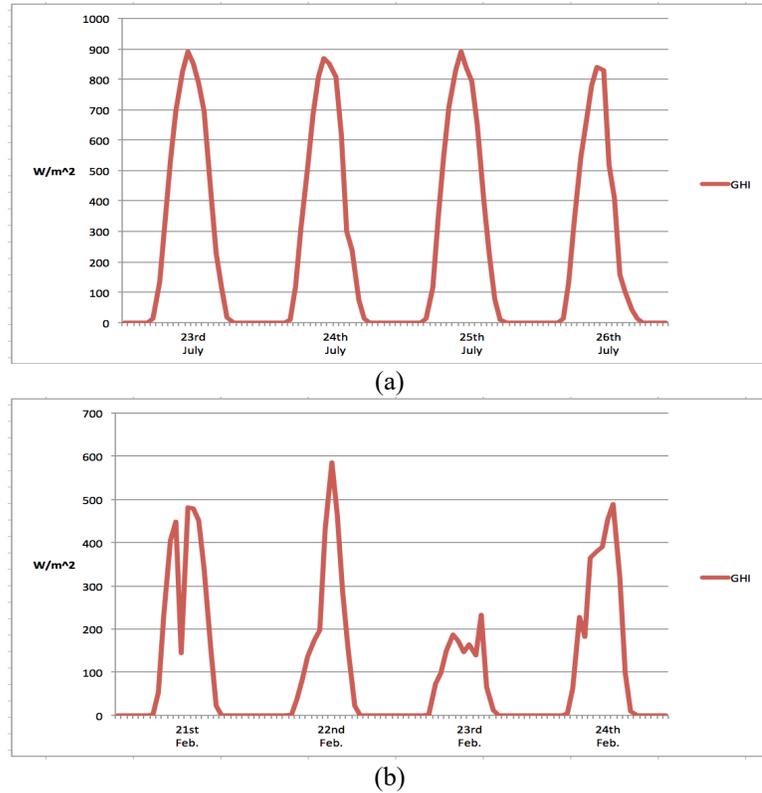
Figure 1.4 Solar irradiance for four consecutive days (a) Sunny, (b) Partly cloudy.

Therefore, the solar irradiance is highly dependent on cloud cover and the daytime hours. In addition, the solar irradiance shows a weak stationarity character in terms of a repetitive pattern. Such variations in pattern will cause difficulty to predict any future changes in solar irradiance and limits the application of historical data to highly correlated ones. Also, the forecasting model requires a large historical dataset to be trained and such amount of data is most likely difficult to be found as solar measurements are limited [7].

Table 1.1 Sample of Selected Sunrise and Sunset time and Daytime hours

Day	Sunrise	Sunset	Daytime hours
22-Jan	7:15	17:06	9 h, 51 min
18-Jul	5:46	20:24	14 h, 38 min
3-Nov	6:30	16:54	10 h, 24 min

5

1.1.2 Solar Forecasting Techniques

Forecasting methods can be categorized into three different methods: Physical, Statistical, and Hybrid [8]. Physical models tend to be good for long term forecasting. Two common physical models are the NWP and the satellite sky imagery. The NWP is based on the physics of the atmosphere which uses current observations of the weather and processes this data to predict the future states using super computers. The satellite and cloud imagery based model is a physical forecasting model that analyzes clouds [9], [10]. Under low sun elevations, low irradiance conditions, and high spatial variability, the errors of satellite and cloud images can increase significantly. In [11] a 17% RMSE for half hour cloud index forecast and 30% RMSE for 2 hours forecast is regressive (AR), moving average (MA), or both (ARMA). The persistence model is the simplest way for forecasting which basically predicts the future value, assuming it is the same as the previous value. Time series models are based on the historical data and defined as a sequence of observations measured over time, such as the hourly, daily, or weekly. It is a stochastic process as observations could be random. Hybrid models merge two forecasting techniques to improve the forecast accuracy. The basic idea of hybrid models is to overcome any deficiency of using an individual model, such as regression models, to take the advantage of each individual model and combine them to reduce forecast errors. For instance, the NWP model can be combined with the ANN by feeding the outputs from the NWP as input to the ANN. In [12], a hybrid model is developed using the satellite imaging as inputs to ANN.

1.2 Solar Forecasting State of the Art

There are extensive studies on solar generation forecasting based on variety of methods with the common goal of minimizing the forecasting error. Using an efficient forecasting tool, power system operators will be able to schedule generation, obtain operating reserves, and administrate changes in loads and power outputs economically. Solar forecast is used in power industry to shape generation portfolios. For instant in Germany, the traded solar capacity in energy market is 38 GW. Such amount of capacity would robustly have an effect on market prices [13]. Forecasting is commonly performed using physical or statistical models, or a combination of the two. Physical models rely on the physics of the atmosphere by observing the current weather conditions and predict the future climate parameters using supercomputers. Physical models are appropriate for long-time horizons and can be divided into two categories of numerical weather prediction (NWP) and the satellite/cloud imagery. The NWP utilizes the current observations in atmosphere, which are processed to produce hundreds of meteorological elements such as temperature and solar irradiance through a process called assimilation. There are different NWP models such as global forecast system (GFS) and the ERA5 by the European center for medium range weather forecast [14]. In [15], different NWP models are analyzed to predict 14 hours of GHI, where the resulted root mean square error (RMSE) ranges from 20% to 40%. The satellite/cloud imagery helps understand the cloud motion by knowing the cloudiness with high spatial resolution. By understanding the cloud motion, the cloud position can be predicted, and thus the solar irradiance can be forecasted [10].

The statistical models require a large set of historical data in order to form a relationship between input and other important factors to forecast the output. These models rely on different mathematical algorithms to recognize the time series trends and patterns. The common statistical models are persistence model, which predicts the next value based on the previous value, autoregressive (AR), moving average (MA), autoregressive and moving average (ARMA), and artificial neural networks (ANNs). Time series models aim to predict the future sequence of observations using historical data over various time horizons such as hourly, daily, or weekly. As the observations could be random, the time series is referred as a stochastic process [16]. Time series models focus only at the patterns of the data. In order to forecast a time series, these patterns should be identifiable and predictable. One of the most widely-used time series models is the ARMA model, which was created by Peter Whittle in 1951 and thoroughly developed and explained by Box-Jenkins in 1971. The ARMA model can be represented mathematically as following:

$$X_t = \sum_{i=1}^{p} \Phi_i * X_{t-i} + \sum_{j=1}^{q} \beta_j * \varepsilon_{t-j} + \varepsilon_t \qquad (1.1)$$

Here Φ_i is coefficient for AR part, β_j is the coefficient for MA part, ε_t is the white noise, and p and q are the orders of the AR and the MA, respectively. In [17], for one-hour ahead forecast, the ARMA model improves the accuracy, the mean square error (MSE), over the persistence model by almost 44.38%. It should be noted that the time series has to be stationary before it is fed to the ARMA model [18]. The artificial neural network (ANN) is another viable statistical forecasting method. The ANN is inspired by the idea of the human biological neural system. These models have the ability to process

8

a complex nonlinear time series and find the relationship between the input and the target output using different training and learning algorithms. There are different types of ANNs such as the recurrent neural network (RNN), feed forward neural network (FFNN), and radial basis function neural network (RBFNN). A detailed review of different ANNs for solar forecasting applications is provided in [19]. Hybrid models have become more attractive to forecasters as they offer the combined advantages and reduce the limitations of other methods. Hybrid models are basically combining two forecasting methods in order to get a better forecasting accuracy. In [20], a hybrid model of a variety of forecasting models is proposed to predict the next 48 hours solar generation in North Portugal. The hybrid model has shown an improvement of 57.4% over the persistence model and 34.06% over the statistical model. Some of the viable solar forecasting models in the literature can be found in [21]–[33].

There are a few major sources of error in solar forecasting: (i) the time series of solar irradiance is unpredictable, caused mainly by weather changes and partial/full cloud cover. As a result, the solar time series is considered non-stationary. Solar has a clear pattern of hourly solar irradiance in clear sky days while it has fluctuations and changes in the patterns due to climate changes. Statistical methods, such as learning-based models used in forecasting, require the time series to be stationary; and (ii) the solar irradiance changes every day based on the duration of the day and sunrise/sunset times. When the duration is different, the historical data cannot be easily used to forecast solar irradiance. For example, the solar data from one day before, one week before, or one month before cannot be suitable to make a viable forecast, while the data from one year before (on the same exact date) is useful as it has similar sunrise/sunset times. This drawback limits the

number of available data points to a set of selected points, which may not be adequate to perform an accurate forecast [34][35].

1.3 Solar PV Integration to Power Systems

By 2023, over 14 GW of solar PV capacity is expected to be installed annually [2]. Such large-scale solar PV integration poses multiple challenges to system control and operation due to the specific characteristics of the solar generation, including variability and uncertainty. Solar PV is considered an intermittent resource due to its output variations. Its generation is also uncertain as there is a lack of ability to perfectly predict the variations [36]-[37]. An extensive study on key challenges imposed by the integration of large-scale solar PV to the transmission grid is discussed in [38], where being 'zero inertia' is further discussed as one of the main concerns.

The study in [39] discusses the adverse effects of low system inertia on the system's angle and frequency stability due to large-scale solar PV integration. Other challenges discussed in [38] associated with large-scale solar PV integration are voltage regulation and stability. A study is conducted in [40] to exhibit the negative impact of large-scale solar PV on voltage stability when integrated in the Western Electricity Coordinating Council (WECC) system. In [41], authors analyze system stability in case of large-scale and distributed solar PV generation, where it is concluded that the distributed solar PV generation is preferred over large-scale integration from a stability viewpoint.

Various studies discuss the grid performance and the operational economics when the solar PV is integrated at different levels of penetration. In [42], authors study the operation and the benefits of 25% solar energy penetration in the Western

Interconnection, which is a large portion of the WECC operated by a group of utilities in Arizona, Colorado, Nevada, New Mexico, and Wyoming. The paper concludes that it is operationally feasible by the Western Interconnection to adopt 25% solar energy penetration if specific operational practices and infrastructure changes are applied to the grid. A study conducted by the National Renewable Energy Laboratory (NREL), titled "the Eastern Renewable Generation Integration Study (ERGIS)" investigates the impact of 30% renewable energy (solar and wind) penetration on the Eastern Interconnection (EI) [43]. The study exhibits the technical potential for EI to accommodate up to 400 GW solar and wind generation. The study reports multiple findings with respect to solar PV and wind integration to EI, and accordingly concludes that at high penetration of solar PV and wind, the operation of thermal and hydro generation changes, different operation patterns are formed at sunrise and sunset, and transmission flows experience a rapid change.

The market of wind and solar PV experiences unprecedented growth in the last decade due to renewable energy polices. In the near future, some states aim to achieve a target renewables portfolio standard (RPS), where utilities have to ensure that a percentage from the electricity, they sell comes from renewable energy resources. For instant, Hawaii requires to accomplish 100% RPS by 2045 [44]. A study is presented in [45] to evaluate the integration of 50% renewables to meet the renewables portfolio standard (RPS) in California in 2030.

In the past decade, an extensive research has been conducted on how to optimally size and integrate solar PV in the distribution grid, which is referred to as hosting capacity. In [44], an optimization-based method is presented to determine the hosting

capacity for distributed generation (DG) resources including solar PV, considering various performance measures. The use of active and reactive power control strategies to increase the hosting capacity through testing different solar PV inverters is demonstrated in [45]. The study in [46] presents a dynamic solar hosting capacity calculation in microgrids in case of transition from the grid-connected mode to the islanded mode. More hosting capacity studies can be found in [47]–[50].

The problem of generation and transmission expansion planning has been investigated in many previous studies. Authors in [51] provide a framework to analyze transmission expansion planning problem from various perspectives including mathematical models and fundamental concepts, available software tools, and educational opportunities. The study in [52] proposes two models to evaluate output power associated with large-scale wind turbines and solar PV. A probabilistic generation expansion planning model is studied in [53] while taking solar PV variability and generator outage possibility into account. This study, however, does not consider unit commitment and transmission line limits. The effect of solar PV penetration on system reliability is investigated in [39], where it concludes that strict performance requirements are needed in case of high solar PV penetration to keep the system reliable. The study in [54] analyzes the impact of large-scale wind and solar PV on net load, where it shows that negative net load leads to renewable generation curtailment. Authors in [55] highlight the effect of DERs on design, operation, and regulation of transmission systems.

Although extensive research exists on generation and transmission expansion planning problem, only a few studies are available in the literature in which the concept

of large-scale solar PV planning is discussed. The study in [56] presents an optimization-based model for large-scale solar PV planning from a private investor perspective so as to pave the way for the investor in making decisions for PV sitting, sizing, and the time of investment. Leveraging a Differential Evolution algorithm, the study in [57] proposes a least-cost generation expansion planning model with solar power plants. Reference [58] studies large-scale solar PV in order to address economic, energy security and environmental challenges confronting power systems. A probabilistic generation portfolio modelling tool is further presented with the objective of minimizing cost and $CO2$ emission.

1.4 Dissertation Outline

This dissertation is organized as follows:

Chapter 2 proposes a new approach to overcome one of the most significant challenges in solar generation forecasting, i.e., the limited availability of the stationary data sets. This challenge is addressed by converting the non-stationary historical solar irradiance data into a stationary set, which will be further validated using an ADF test. This conversion will be followed by a neural network-based forecasting and proper post-processing steps. Numerical simulations exhibit the performance of the proposed method

Chapter 3 proposes a two-stage day-ahead solar forecasting method that breaks down the forecasting into linear and nonlinear parts, determines subsequent forecasts, and accordingly, improves accuracy of the obtained results. To further reduce the error resulted from nonstationarity of the historical solar radiation data, a data processing approach, including pre-process and post-process levels, is integrated with the proposed method.

Chapter 4 presents a solar photovoltaic (PV) generation forecasting model based on multi-level solar measurements and utilizing a nonlinear autoregressive with exogenous input (NARX) model to improve the training and achieve better forecasts. The proposed model consists of four stages of data preparation, establishment of fitting model, model training, and forecasting. The model is tested under different weather conditions. Numerical simulations exhibit the acceptable performance of the model when compared to forecasting results obtained from two-level and single-level studies

Chapter five proposes and develops a co-optimization generation and transmission expansion planning model with the objective of maximizing large-scale solar PV hosting capacity. In the proposed model, dispatchable generation units and transmission lines are expanded in a way that the system will be able to host a maximum capacity for large-scale solar PV penetration. A decomposition approach is applied to coordinate planning and operation problems, and further ensure the computational efficiency of the proposed model.

Chapter six investigates the effect of the solar PV large ramp rate in the operation of the IEEE 33-bus test system. A one-hour time horizon with a resolution of 1 minute of solar PV output has been used to investigate the possible voltage fluctuations caused by the solar generation ramp rate. The simulation results show that mitigation of the ramp rate can reduce the standard deviation in daily voltage values compared to the case when the ramp rates are not mitigated.

Chapter Two: Day-ahead Solar Forecasting Using Time Series Stationarization

2.1 The Proposed Forecasting Model

The global horizontal irradiance (GHI) is the main solar irradiance component considered in this dissertation. The GHI is the total amount of irradiance received by the surface horizontal to the ground. It consists of both Direct Normal Irradiance (DNI) and Diffuse Horizontal Irradiance (DHI). The GHI used in this dissertation is available to public [59] for different locations in the U.S provided by National Renewable Energy Laboratory (NREL). Two different GHI data sets were used which are the historical GHI and the clear sky GHI. The clear sky GHI represents the peak GHI that could be received at the surface during a clear sky condition.

The proposed model consists of three stages to forecast the GHI and as shown in Fig. 2.1. The model includes three stages of data pre-processing, forecasting, and data post-processing. After the output of the model is obtained, errors are calculated, using multiple error criteria, to determine the solution accuracy. These three stages are presented in detail in the following:

2.1.1 Stage 1: Data Pre-Process

Before data are fed into the forecasting tool, the data set is modified under the data pre-processing stage. The pre-processing includes: removing the offset, removing nighttime GHI values, detrending, and normalization. Removing the offset is achieved by

subtracting the historical GHI from the clear sky GHI using (2.1). The GHI scattered by cloudiness or other factors is represented by the deviated GHI that is the result in (2.1) as shown in Fig. 2.2.

$$GHI_{dev.}(t,h) = GHI_{CSK}(t,h) - GHI_{his.}(t,h) \qquad h \notin nighttimehours \qquad (2.1)$$

$$GHI_{norm.}(t,h) = GHI_{dev.}(t,h) / GHI_{CSK}(t,h) \qquad h \notin nighttimehours \qquad (2.2)$$

The resultant data in (1) are a function of time and location that reflects all meteorological data that affects solar irradiance such as cloudiness and aerosol. The next process after removing offset is to remove the nighttime GHI values. The solar irradiance varies during the daytime and it is zero during the nighttime. By knowing the sunrise and sunset for each day, this step can be accomplished regardless of the length of the daytime. The sunrise and sunset times are fixed for each specific day for the same location over the years. The complete list of sunrise and sunset times can be found in [60]. This resultant data from previous two steps are introduced to model fitting in order to detrend the solar time series. The solar time series is a non-stationary time series that has to be stationarized before it is fed to the forecasting tool.
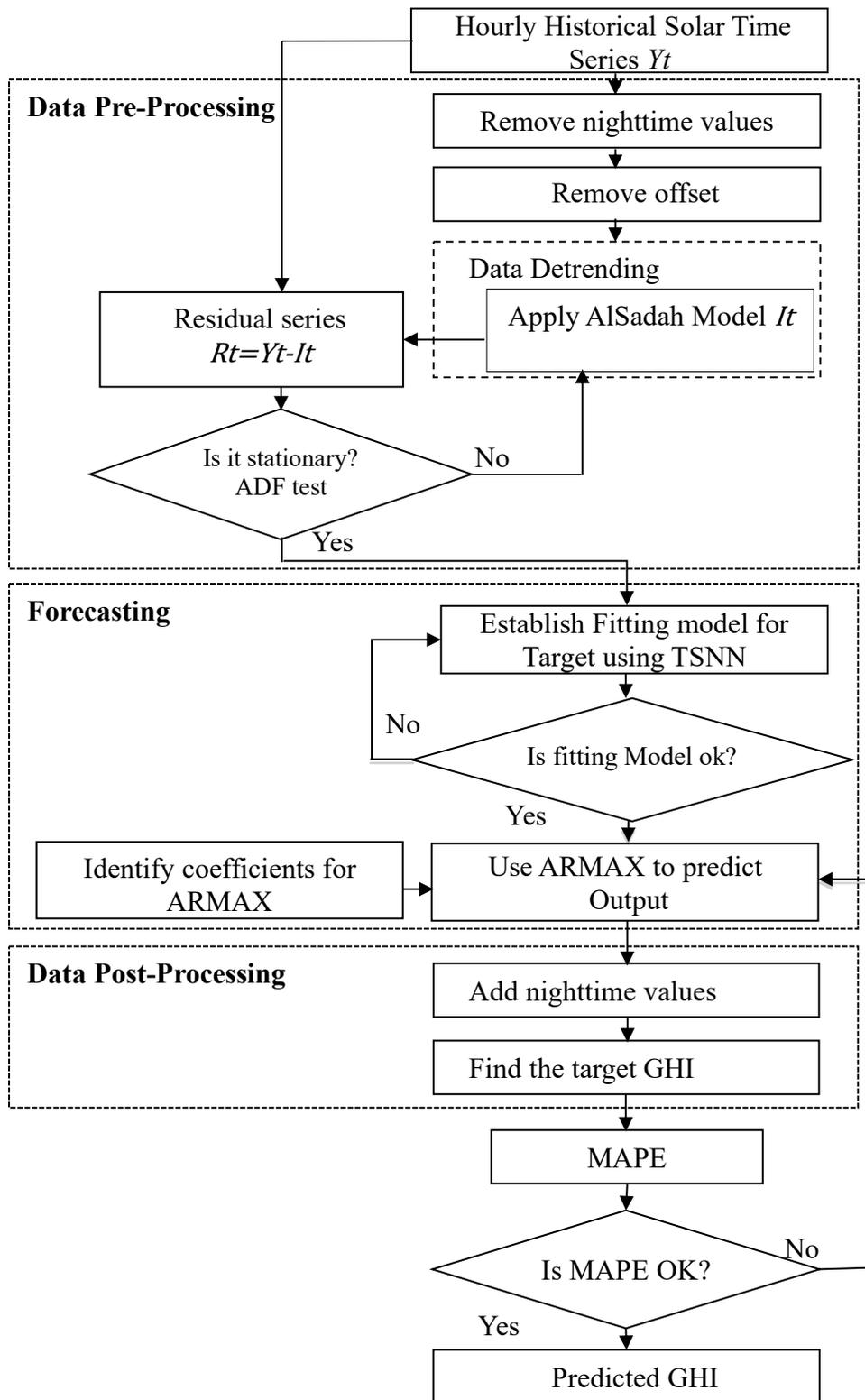
Figure 2.1 The flowchart of proposed model for GHI forecasting

17
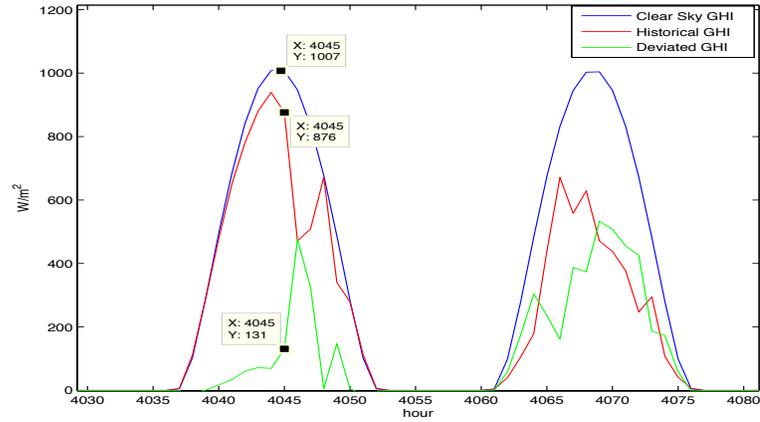
Figure 2.2 GHI difference between clear sky GHI and historical GHI

Various detrending models are discussed in [61] and [62] and in both, authors have concluded that the Al-Sadah's model, which is also referred as high order polynomial model, works more efficiently than other detrending models such as Jain, Baig, and Kaplains.

The high order polynomial model can be represented as follows [63]:

$$I_t = a_0 + a_1 h + a_2 h^2 + ... + a_n h^n \qquad h \notin nighttimehours \qquad (2.3)$$

Where constants $a_0, a_1 ... a_n$ can be found by fitting the actual data of solar time series using least squares regression analysis. After the model is analyzed, a stationary test is applied to residuals to check if the data is stationary or not.

The test used in this dissertation is the Augmented Dickey–Fuller (ADF) test [34]. The ADF test examines whether there is a unit root in the time series. If there is a unit root, which means that the test result is above the critical value, the null hypothesis should be accepted and the time series is not stationary, otherwise the null hypothesis is rejected and the time series is stationary. The ADF model can be presented as follows:

18

$$\Delta x_t = \mu + \beta_t + \rho x_{t-1} + \delta_1 \Delta x_{t-1} + \delta_2 \Delta x_{t-2} + \ldots + \delta_p \Delta x_{t-p} + e_t \tag{2.4}$$

Here μ is a constant called drift, β a coefficient that represents the trend, p is the number of lags or the order of the autoregression process, and e_t represents random variables with zero mean. The detrending is tested first on the hourly average of the month since the specific day's series is less predictable due to variations caused by different factors. Then, if the fitting model is stationary, the model is applied to the historical hourly GHI to predict day-ahead GHI.

The last step in data pre-processing is normalization. Some data sets have extreme values that could lead to distorting the forecasted result. Also, by using normalization, all data sets will be under the same reference scale, thus the variability as a result of changing solar irradiance peak will be eliminated. The normalization in actual data is performed by dividing the resultant data from previous steps over the associated clear sky solar irradiance in the same day as in (2.2). The outputs from (2.2) are normalized GHI values ranging from 0 to 1. Also, normalization is performed in the data output from fitting model using (2.5). Data normalization ensures the quality of the input data before it is fed to the forecasting model.

$$I_{tNorm}(t,h) = I_t(t,h) / GHI_{CSK}(t,h) \qquad h \notin nighttimehours \tag{2.5}$$
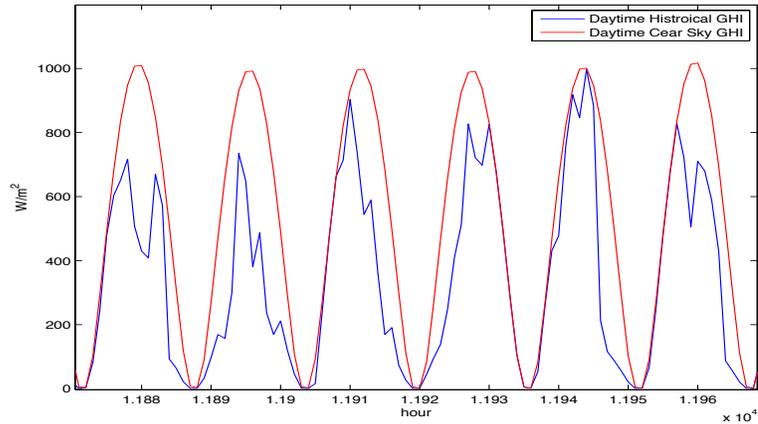
Figure 2.3 Daytime hourly GHI in cloudiness and clear days

2.1.2 Stage 2: Forecasting

In this stage, the pre-processed data are introduced to the forecasting tool, here the neural network (NN). The historical detrended GHI data are fed to the model as an input and the actual GHI is fed as a target. The NN model is established by entering the number of hidden layers. The historical data is trained using different training methods and the error is analyzed. The training process continues until the error between the forecasted and the actual GHI is minimized given the inputs, weights, number of hidden nodes and layers as illustrated in (2.6). The actual N outputs and the forecasted outputs are $GHI_{norm.k}$ and $\hat{GHI}_{norm.k}$, respectively. The weights between neurons arew and v. If the desired accuracy is not achieved, i.e., an acceptable error, the step is repeated by using a different NN structure. The feed-forward neural network (FFNN) is the simplest model in NN and is used in this dissertation due to its promising results in forecasting. The objective of this method is to forecast N outputs given m inputs as represented in (2.7). However, any other method can be utilized without loss of generality.

20

$$\min E(w,v) = \frac{1}{2}\sum_{k=1}^{N}(GHI_{norm.k} - \hat{GHI}_{norm.k})^2 \tag{2.6}$$

$$(y(t_{m+1}), y(t_{m+2}),..., y(t_{m+n})) \approx f(x(t_1), x(t_2),..., x(t_m)) \tag{2.7}$$

2.1.3 Stage 3: Data Post-Processing

The resultant forecasted data from stage two represent the daytime GHI values in normalized form. So, stage three is the reverse of the first stage. Stage three includes three processes: denormalization, adding nighttime hours, and calculating the forecasted GHI.

The processed data is denormalized through multiplying the hourly peak clear sky GHI by the hourly output GHI from stage two as in (2.8). Then, the nighttime values are added in the second process using sunrise and sunset times in addition to the daytime duration. The final step is to subtract the previous resultant GHI data from the clear sky GHI to obtain the actual forecast values using (2.9).

$$GHI_{denorm.}(t,h) = GHI_{NN}(t,h) * GHI_{CSK}(t,h) \qquad \forall h \tag{2.8}$$

$$GHI_{forecast}(t,h) = GHI_{CSK}(t,h) - GHI_{denorm.}(t,h) \qquad \forall h \tag{2.9}$$

2.2 Numerical Studies

A day-ahead forecast under various weather conditions is performed to show the significance of the stationary data sets in improving the forecast accuracy. MAPE is calculated to evaluate the performance under each case. The following cases are studied:

Case1: Forecast using stationary data.

Case2: Forecast using non-stationary data.

Case 1: Forecast using stationary data.

The hourly solar GHI used in the simulation is obtained from NREL [59]. First, the hourly actual and clear sky irradiance monthly average (for March 2010) is calculated to fit the high order polynomial as in (2.3). Then the data undergoes the first two processes in stage one which are removing the nighttime hours and removing the offset. The resultant data is fitted using Al-Sadah's model as shown in Fig. 2.4.
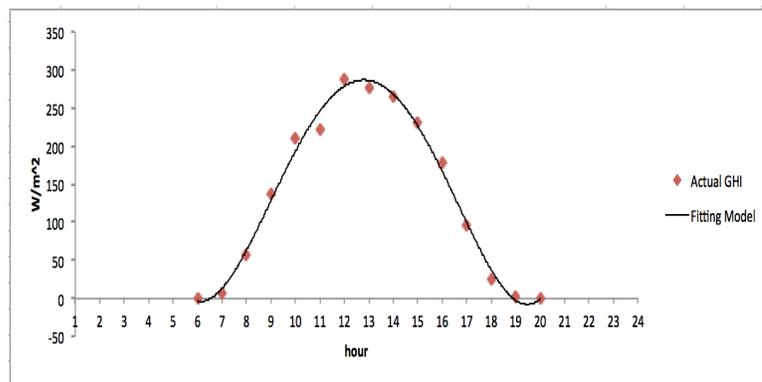
Figure 2.4 Actual hourly average for month of March and the fitting model series

The residual, which is the difference between the actual series and fitting series, is computed. Fig. 2.5 depicts the residual series after detrending by the Al-Sadah's model.
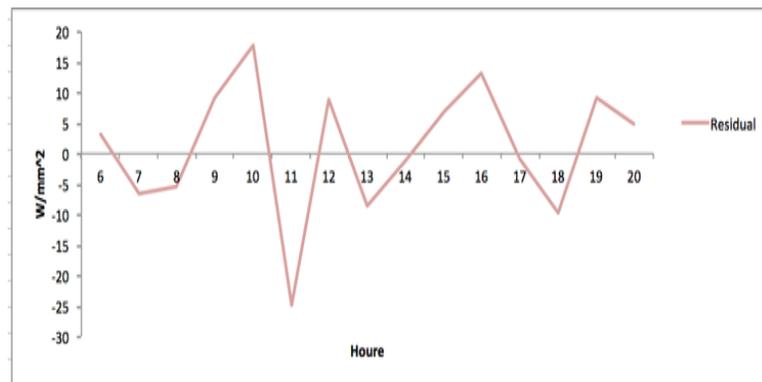
Figure 2.5 The residual after detrending

In the ADF test to check if the series is stationary or not, "daftest" command in Matalb was used to test the stationarity using the significance level of 0.05. The test result is summarized in Table 2.1.

Table 2.1 The ADF Test for The Detrending Model-Case1

Statistical Power	Significance level	Test result	Critical value	RMSE	NRMSE
0.001	0.05	-5.12	-1.957	4.30	0.032

The test result is below the critical value and that means there is no unit root in the data tested, so the null hypothesis should be rejected and the time series is stationary. The statistical power shows the probability that the detrended time series has unit root. As shown in Table 2.1, the probability in the detrended model is very low which confirms the conclusion.

The forecasting accuracy is checked using a variety of error measures, including the following three indicators: root mean square error (RMSE), normalized root mean square error (NRMSE), and mean absolute percent error (MAPE):

$$MAPE = \frac{1}{n}\sum_{t=1}^{N}\left|\frac{GHI(t)_{actual} - GHI(t)_{forecast}}{GHI(t)_{actual}}\right| * 100 \tag{2.10}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{N}(GHI(t)_{actual} - GHI(t)_{forecast})^2} \tag{2.11}$$

$$RMSE = \frac{\sqrt{\frac{1}{n}\sum_{t=1}^{N}(GHI(t)_{actual} - GHI(t)_{forecast})}}{\frac{1}{n}\sum_{t=1}^{N}(GHI(t)_{actual}} \tag{2.12}$$

The model is further tested under cloudy conditions to check if the model can detect variations due to cloud movement. In this case, the data from May 4th was detrended and used as an input to the forecasting tool in order to predict solar GHI of May 5th, which was a cloudy day. Fig. 2.7 depicts the forecasted solar.

Multiple days are forecasted using the proposed method under different weather conditions. First, the data of March 7th, 2010 is detrended and then used as historical data to train the forecasting model and to predict the next day, i.e., March 8th 2010, which was a partially cloudy day. The forecast result is depicted in Fig.2.6.
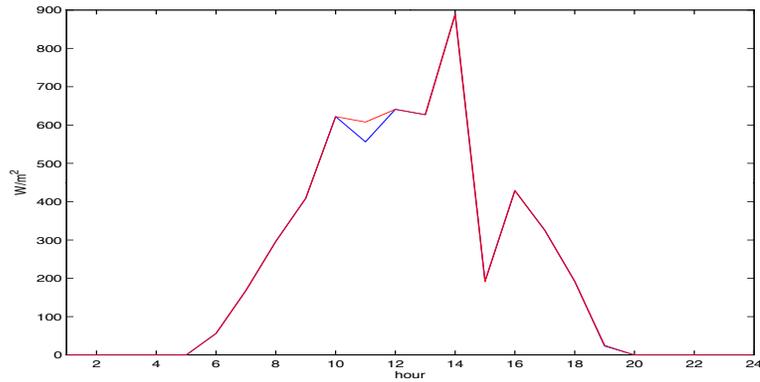
Figure 2.6 The actual and predicted for March 8th, 2010
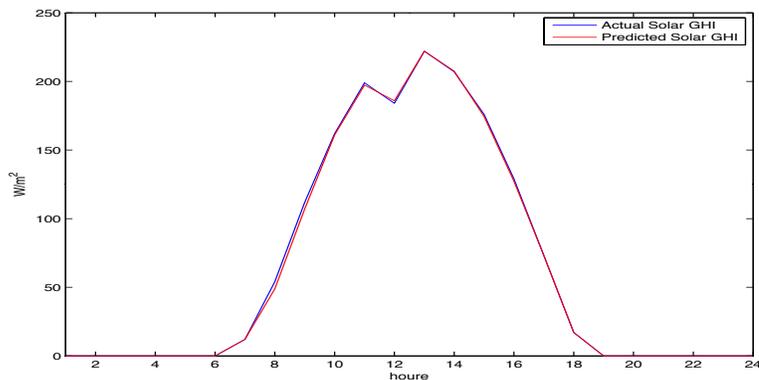
Figure 2.7 The actual and predicted for May 5th, 2010

24

Finally, the model was tested under clear sky condition (sunny). August 13th was selected to be the target and the data from previous day, August 12th, was detrended and used as input to train the model and forecast output as shown in Fig. 2.8. Table 2.2 summarizes the performance of the forecasts using different accuracy indicators.
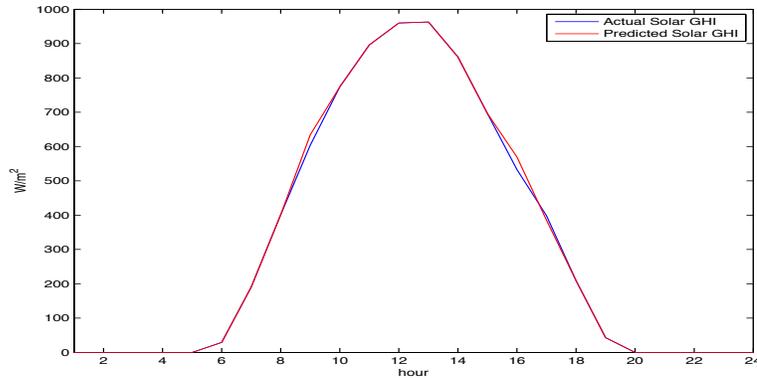
Figure 2.8 The actual and predicted for August 13th 2010

Table 2.2 Forecast Performance under Different Weather Conditions-Case1

Weather Condition	Day	MAPE (%)	RMSE (W/m^2)	NRMSE
P. Cloudy	March 8th	0.811	1.609	0.025
Cloudy	May 5th	0.799	10.582	0.047
Sunny	August 13	0.639	10.066	0.032

Case 2: Forecast using non-stationary data.

The hourly GHI data for March 2010 is tested, using ADF, for checking stationarity and the output results are summarized in Table 2.3. As presented, the test result is above the critical value, which indicates that there is a unit root and null hypothesis should be accepted. As a result, the available solar irradiance is a non-stationary time series. This data is directly fed to the neural network forecasting tool to

25

forecast GHI values for different test dates under various weather conditions. The resultant MAPE is summarized in Table 2.4. As the obtained results indicate, the forecast is more accurate under clear sky conditions, i.e., a sunny day, compared to other days with cloud cover, conceivably due to the GHI variability during cloudy days. Moreover, the error in cloudy and partly cloudy days is almost the same, but it cannot be generalized for other days with similar weather conditions.

Table 2.3 The ADF Test for Hourly Average GHI For March 2010 – Case 2

Statistical Power	Significance level	Test result	Critical value
0.47	0.05	-0.465	-1.9567

Table 2.4 Forecast Performance – Case 2

Weather Condition	Day	MAPE (%)
Partly Cloudy	April 8	2.624
Cloudy	May 5	2.670
Sunny	August 13	1.117

2.3 Discussion

As shown in Table 2.2, the MAPE for the day-ahead forecast is below 1%. The new model has shown a promising result for 24 hours ahead forecast under different climate conditions. Moreover, even with existence of variations due to clouds in some days, the proposed model has shown the ability to detect such fluctuations. The NRMSE is also below 0.05, which indicates that the model has presented a very good result and can potentially outperform many of the current forecasting methods. The results show that the application of stationary data set could reduce the MAPE ranges by as much as

42% in sunny days and 70% in cloudy days. The high accuracy of the forecast would allow power system operators to predict any sudden fluctuations in the solar output and perform the proper control actions. The model has different features when compared to the existing methods:

- The model converts the solar irradiance time series to a stationary time series and apply ADF test to check the stationarity.

- The model has shown a significance decrease in MAPE when compared to exiting methods.

- The new model performs well under different weather conditions, including cloudy conditions.

- The new model reduces the size of the time series to almost half by removing nighttime values, which accelerates the time needed for simulations.

Chapter Three: Two-Stage Hybrid Day-Ahead Solar Forecasting

3.1 The Architecture of The Forecasting Model

Fig. 3.1 depicts the architecture of the proposed decomposed forecasting model. The forecasting model uses two cascaded stages based on Nonlinear Autoregressive Neural Network (NARNN) and the Autoregressive Moving Average with Exogenous Input (ARMAX). The main advantage of this decomposed model is to process both linear and nonlinear parts of the solar time series. NARNN deals with the nonlinear part of the forecasting and is used to predict a fitting model based on the historical stationary solar data. On the other hand, ARMAX considers the linear part of the forecasting and is used to forecast solar irradiance using the predicted fitting model as an input. These two stages along with data processing is explained in the following:
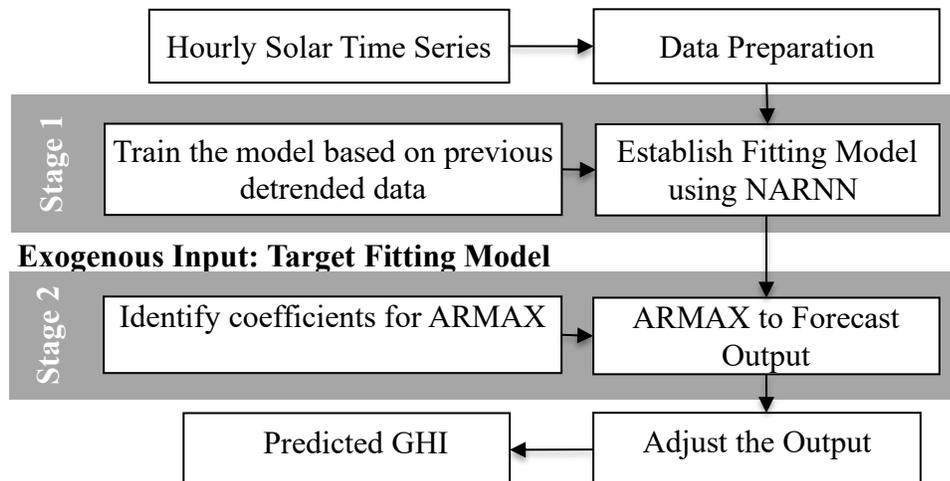
Figure 3.1 The architecture of the proposed forecasting model.

3.1.1 Data Preparation and Preprocessing

The solar data considered in this model are the horizontal global irradiance (GHI) that represent the solar irradiance received at horizontal surface on the ground. The historical GHI data are analyzed and undergone several preprocessing steps to ensure the quality of the data.

The GHI data preparation includes removing GHI nighttime values, removing offset, and detrending. The historical GHI is subtracted from the maximum received GHI under clear sky conditions to remove the offset. The resultant data from this process represent solar irradiance scattered by clouds or other factors. The resultant data are more dependent on the location and time that reflects other meteorological data. The next step is to consider only daytime hours, as solar output at nighttime hours is zero. Reducing the size of the data is achieved by eliminating the nighttime values provided that the time of sunrise and sunset, which will further reduce the simulation time.

Statistical models require data set to be stationary before it is applied to forecasting model. The output data from the two previous steps are detrended and then tested using the Augmented Dickey–Fuller (ADF). The ADF test validates the stationarity of the time series. The ADF test checks if there is a unit root. If test result comes below a defined critical value, then the time series is stationary and the null hypothesis should be rejected. Otherwise, the time series is considered nonstationary and the null hypothesis should be accepted. Different detrending models are addressed and compared in [61].

The detrending model used in this dissertation is Al-Sadah's model as represented mathematically in (3.1). Constants $a_0, a_1,..., an$ are determined using the least square regression analysis to fit the actual data set:

$$Y_t = a_0 + a_1 h + a_2 h^2 + ... + a_n h^n$$

(3.1)

Where h is the local time. After the data are tested for the stationarity and verified, it will be normalized to obtain a number between 0 and 1. Normalization is an important step to ensure all data sets are under same reference scale, and to eliminate any variability due to the changes in the peak of the clear sky irradiance. More detail on GHI data preprocessing can be found in [64].

3.1.2 Forecasting – Stage 1: Nonlinear Autoregressive Neural Network

The NARNN model is a time series model that requires a large set of historical data. In order to train the model and predict the fitting model, a large set of the previous hourly stationary data from the target day are used. One key issue is that the larger number of days that are used for the training, the more accurate the prediction will be. The NARNN is presented in (3.2) where d is the number of previous hourly samples, determined through trial and error.

Once the fitting model is achieved, it is introduced as an exogenous input to the second stage of the forecasting, i.e. ARMAX. The predicted fitting model form this stage plays an active role in forecasting of the next stage, where a more accurately predicted fitting model in the first stage better ensures a more precise forecasting result in the second stage. The architecture of the NARNN is depicted in Fig. 3.2.

$$y(t) = f(y(t-1), y(t-2),..., y(t-d))$$

(3.2)

30
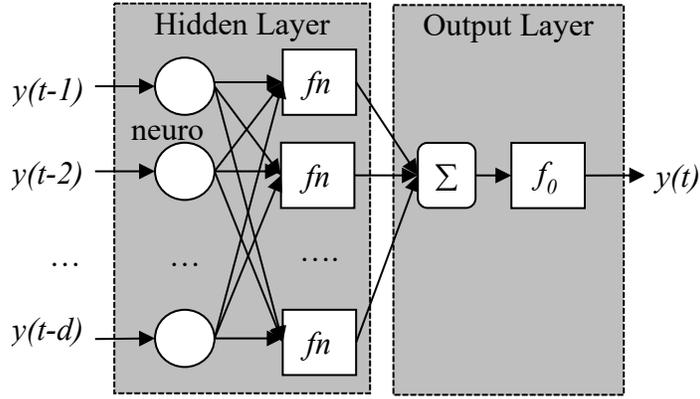
Figure 3.2 The architecture of the NARNN

In order to evaluate the performance of the fitting model the coefficient of determination R^2, which ranges between 0 and 1, is calculated as in (3). If $R^2 = 1$, the NARNN is able to predict the fitting model without any error, while $R^2 = 0$ means that the NARNN is not able to predict the fitting model and further training is needed.

$$R^2 = 1 - \left[\frac{\sum_{t=1}^{N} (GHI(t)_{actual} - GHI(t)_{forecast})^2}{\sum_{t=1}^{N} (GHI(t)_{actual} - \overline{GHI(t)}_{actual})^2} \right] \qquad (3.3)$$

3.1.3 Forecasting – Stage 2: Autoregressive Moving Average with Exogenous Input

The autoregressive moving average model with exogenous input includes two parts and can be mathematically represented as in (3.4).

$$A(q)y(k) = B(q)u(k) + C(q)v(k) \qquad (3.4)$$

Where, $A(q) = 1 + a_1 q^{-1} + \ldots + a_n q^{-n}$, n is the order and $a_1, \ldots, a_n$ are coefficients for the AR part. $B(q) = b_1 + b_2 q^{-1} + \ldots + b_m q^{-m+1}$, m is the order and $b_1, \ldots, b_m$ are coefficients for the input. $C(q) = 1 + c_1 q^{-1} + \ldots + c_r q^{-r}$, r is the order and $c_1, \ldots, c_r$ are coefficients for the MA part.

31

In order to find the coefficients for both AR and MA parts, the previous day is used to train the ARMAX model and estimate coefficients. The order of the ARMAX can be identified using the partial and autocorrelation plots. More detail on how to estimate the order of ARMAX model can be found in [65]. In addition, the Akaike Information Criterion (AIC) can be used to determine the order of ARMAX model. The AIC is modeled under different ARMAX orders and the best order is the one with the lowest AIC [66]. However, in this paper another method is used to find the orders of the AR and the MA. It is assumed that the orders of both the AR and the MA are the same, and then the error value is calculated by increasing the orders. The point with the least error for test data is considered as the best order for the ARMAX model. Fig. 3.3 shows the procedure of finding the order of the ARMAX model.

Once the ARMAX is developed, the fitting model predicted from Stage 1 is introduced as an exogenous input to this stage and the target output is forecasted. To achieve the final GHI, the forecasted output resulted from the ARMAX is adjusted as well.
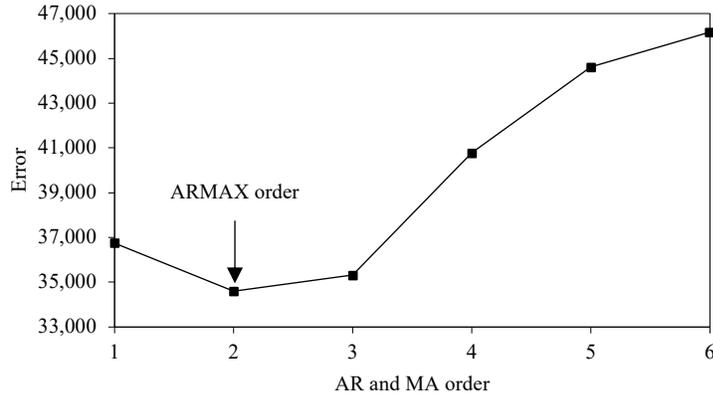
Figure 3.3 Determining AR and MA orders based on error value.

3.1.4 Data Finalization

The output data from the second stage are in normalized form and represent GHI values excluding nighttime hours. So, in order to adjust the predicted GHI values, the GHI data are normalized and nighttime values are added. Finally, the forecasted GHI is calculated.

The performance of the model is accordingly evaluated by calculating the Normalized Root Mean Square Error (NRMSE) as in (3.5).

$$NRMSE = \frac{\sqrt{\frac{1}{N} * \sum_{t=1}^{N}(GHI(t)_{actual} - GHI(t)_{forecast})^2}}{\frac{1}{N} * \sum_{t=1}^{N}GHI(t)_{actual}} \tag{3.5}$$

3.2 Numerical Simulations

The hourly GHI data for the Denver International Airport are used for forecasting [59]. The proposed model is applied to three test days under different weather conditions, and the R^2 and the NRMSE are computed to evaluate output result in stage 1 and 2, respectively. The coefficient R^2 is used to evaluate the performance of the fitting model

33

resulted from the NARNN, and accordingly the NRMSE is applied in the ARMAX model to evaluate the efficiency of the forested GHI. In order to present the effectiveness of the proposed two-stage forecasting model and show the role of the data detrending, the following cases are studied:

Case 1: Forecasting using the proposed model with stationary data.

Case 2: Forecasting using the proposed model with nonstationary data

Case 3: Forecasting using only one stage instead of the proposed two-stage method

Case 1: Using Two-Stage Model with Stationary Data.

In this case, stationary data with the proposed two-stage forecasting model are used for solar forecasting. In this respect, first, the historical GHI data undergo mentioned processes to ensure stationarity before the data are fed to NARNN. Second, the NARNN is trained based on the stationary data to establish the target fitting model. Fig. 3.4 depicts the fitting model predicted by the NARNN model for a cloudy day. The calculated R^2 is 0.90, which reveals that the predicted fitting model is close to the target and the fitting model is well predicted in order to be fed to Stage 2. Third step is to introduce the fitting model to the Stage 2 forecasting, i.e., the ARMAX model. The ARMAX model is developed using order 2 for both AR and MA. It should be noted that the previous day is used to train the model and identify ARMAX coefficient. The ARMAX model forecasts the output as shown in Fig. 3.5. The NRMSE is calculated as 0.085.

In order to evaluate the efficiency of the proposed model, Case 1 is further applied to two additional days. The predicted fitting model by Stage 1 and the forecasted

GHI resulted from Stages 2 for the partly cloudy day are depicted in Figs. 3.6 and 3.7, respectively. Similarly, the fitting model and the forecasted GHI for the sunny day are respectively shown in Figs. 3.8 and 3.9. Table 3.1 summarizes the obtained R^2 and the NRMSE for each of the studied days. As tabulated in Table 3.1, the NARNN detects the fitting model and the ARMAX forecasts the GHI quite accurately. Moreover, the proposed two-stage model with stationary data can accurately forecast not only the sunny days, but also the cloudy and partly cloudy ones.

Table 3.1 R^2 and NRMSE for the Fitting Model and forecasted GHI in case 1.

Weather Condition	R^2	NRMSE
Cloudy	0.90	0.085
Partly Cloudy	0.91	0.100
Sunny	0.86	0.048

Case 2: Using Two-Stage Model with Data.

Data stationarity has a significant Nonstationary role in forecasting solar irradiance. This case aims at investigating the effect of using stationary data in the proposed method. In this regard, the proposed two-stage method is utilized but instead of feeding stationary data to the model, the nonstationary data are used. The simulation processes including Stages 1 and 2 (fitting model predicted by NARNN and the ARMAX model) are completely executed for Case 2 without the pre-processing. Table 3.2 compares the NRMSE values in this case with the same three days as in Case 1. The NRMSE values for the two-stage method using nonstationary solar data are higher compared to NRMSE values computed before. That means even though the cascaded

two-stage method is a useful approach to deal with nonlinear and linear parts of the forecasting, the data stationarity plays a critical role in the accuracy of the results.

Case 3: Using One-Stage Forecasting Method.

To show the merits of the two-stage method over a single-stage method, the forecasting is performed using only one stage, here the NARNN. In this case, the stationary data are applied to forecast the target days. That is, similar to Case 1, the historical GHI data undergo preprocess to ensure stationarity before the data is used in the NARNN. It is then trained based on the historical data set considering a similar number of previous hourly samples, as in Cases 1 and 2. Table 3.2 shows the NRMSE values for the single-stage method comparing with the proposed two-stage method in Case 1. The two-stage method in Case 1 has a better performance in solar forecasting as the NRMSE values are much less than this case, exhibiting the advantages of the two-stage method over a single-stage method. These results advocate on the merits of decomposing model to reap the benefit of both linear and nonlinear parts in the proposed model.

Table 3.2 The NRMSE for Different Case Studies

Weather Condition	NRMSE (Two-Stage model and stationary data)	NRMSE (Two- Stage model and nonstationary data)	NRMSE (One stage model NARNN)
Cloudy	0.085	0.3007	0.512
Partly Cloudy	0.100	0.6799	0.9899
Sunny	0.048	0.212	0.301
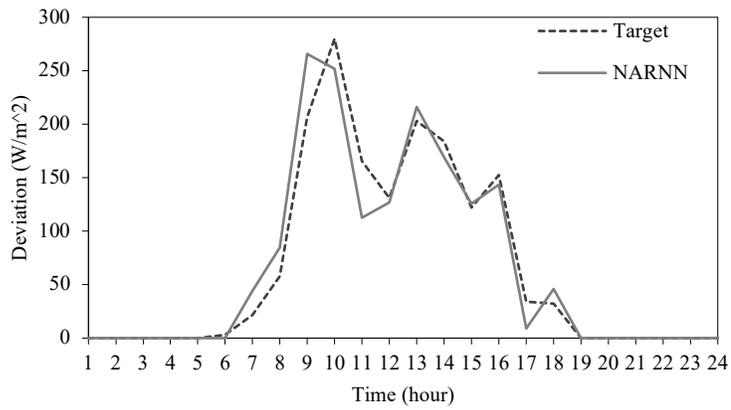
Figure 3.4 The fitting model for the cloudy day using NARNN.
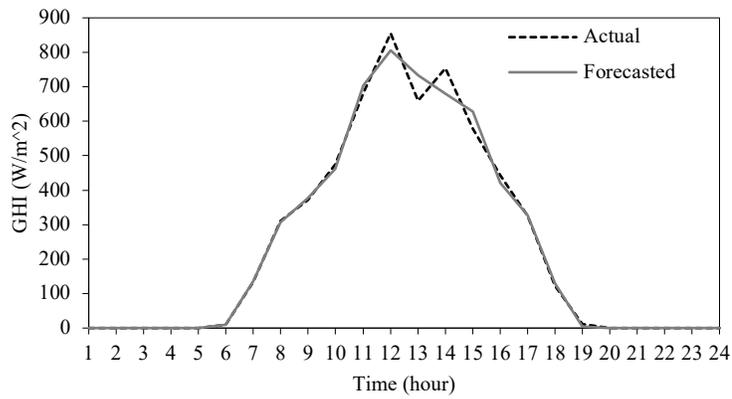
Figure 3.5 Forecasted GHI for the cloudy day using ARMAX.
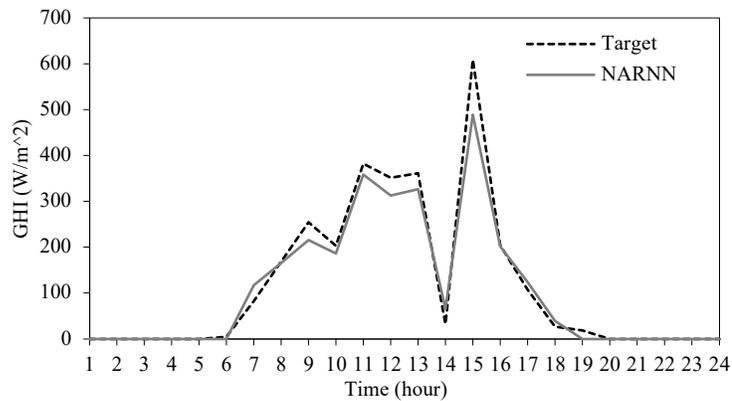
Figure 3.6 The fitting model for the partly cloudy day using NARNN.
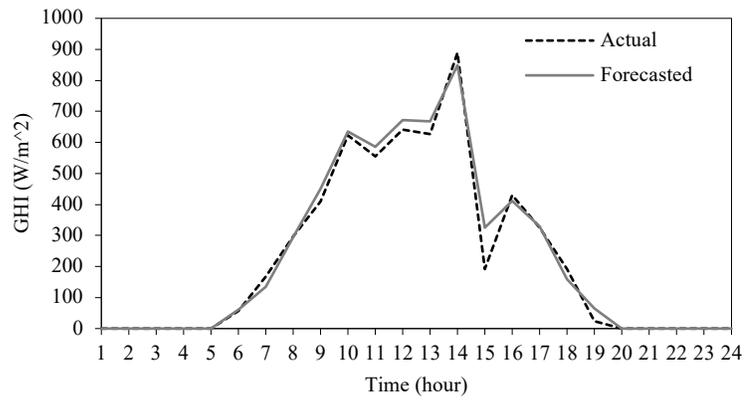
Figure 3.7 Forecasted GHI for the partly cloudy using ARMAX.
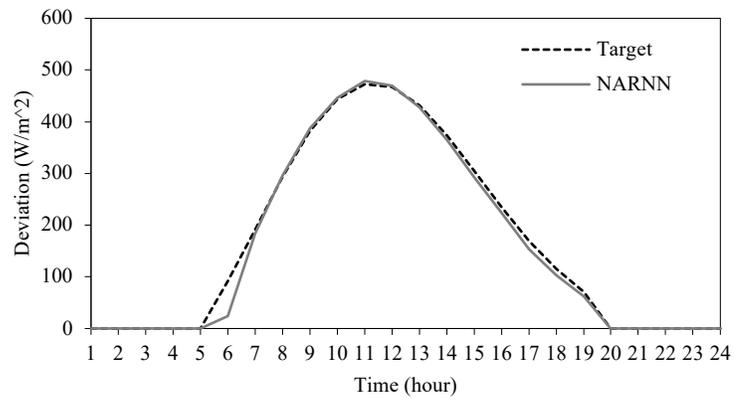
Figure 3.8 The fitting model for the sunny day using NARNN.
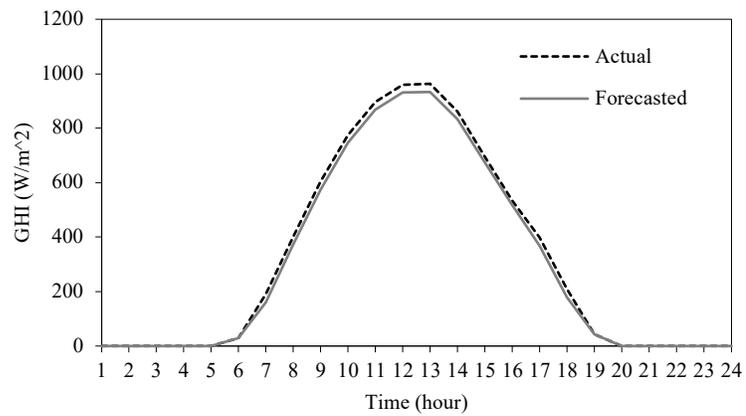
Figure 3.9 Forecasted GHI for the sunny day using ARMAX.

3.3 Discussion

The maximum NRMSE for one day ahead forecast using the proposed two-stage model and stationary data under different weather conditions is 0.1, which is a promising result. In the proposed model, the minimum NRMSE is achieved in a sunny day, which is quite expected as the trends of the solar time series under clear sky conditions are more predictable. Nevertheless, the two-stage model accuracy has improved by 71% to 85% when using stationary data compared to nonstationary data. Finally, the two-stage model outperforms the single-stage model and reduces the NRMSE by almost 83% to 90%.

Chapter Four: Day-Ahead Solar Forecasting Based on Multi-level Solar Measurements

4.1 Forecasting Model Outline and Architecture

Fig. 4.1 depicts the three levels of solar PV measurements: customer, feeder, and substation. The proposed model aims to outperform the forecast applied at each solar measurement level. The forecast in each level is performed using a nonlinear autoregressive neural network. The mean absolute percent error MAPE is accordingly calculated as in (4.3) for each level, and denoted as E_C, E_F, and E_S for customer, feeder, and substation, respectively. This model aims to reduce the forecasting error to be less than the minimum of E_C, E_F, and E_S. Fig. 4.2 depicts the three datasets, which are processed under different stages and explained in the following.
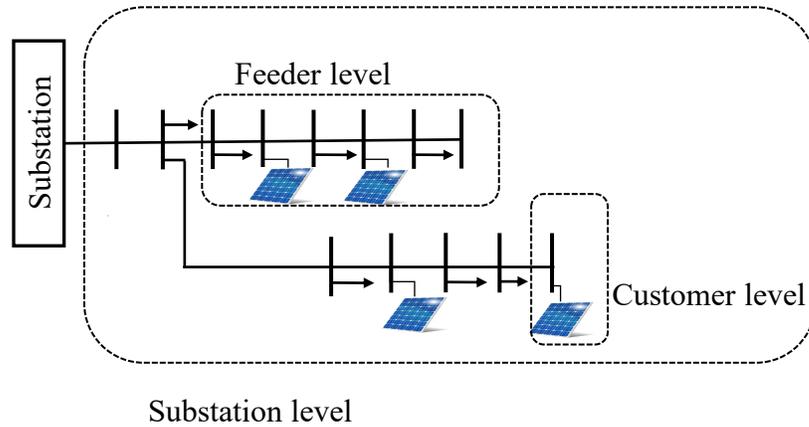
Figure 4.1 Multi-level Solar PVs installed at different locations.
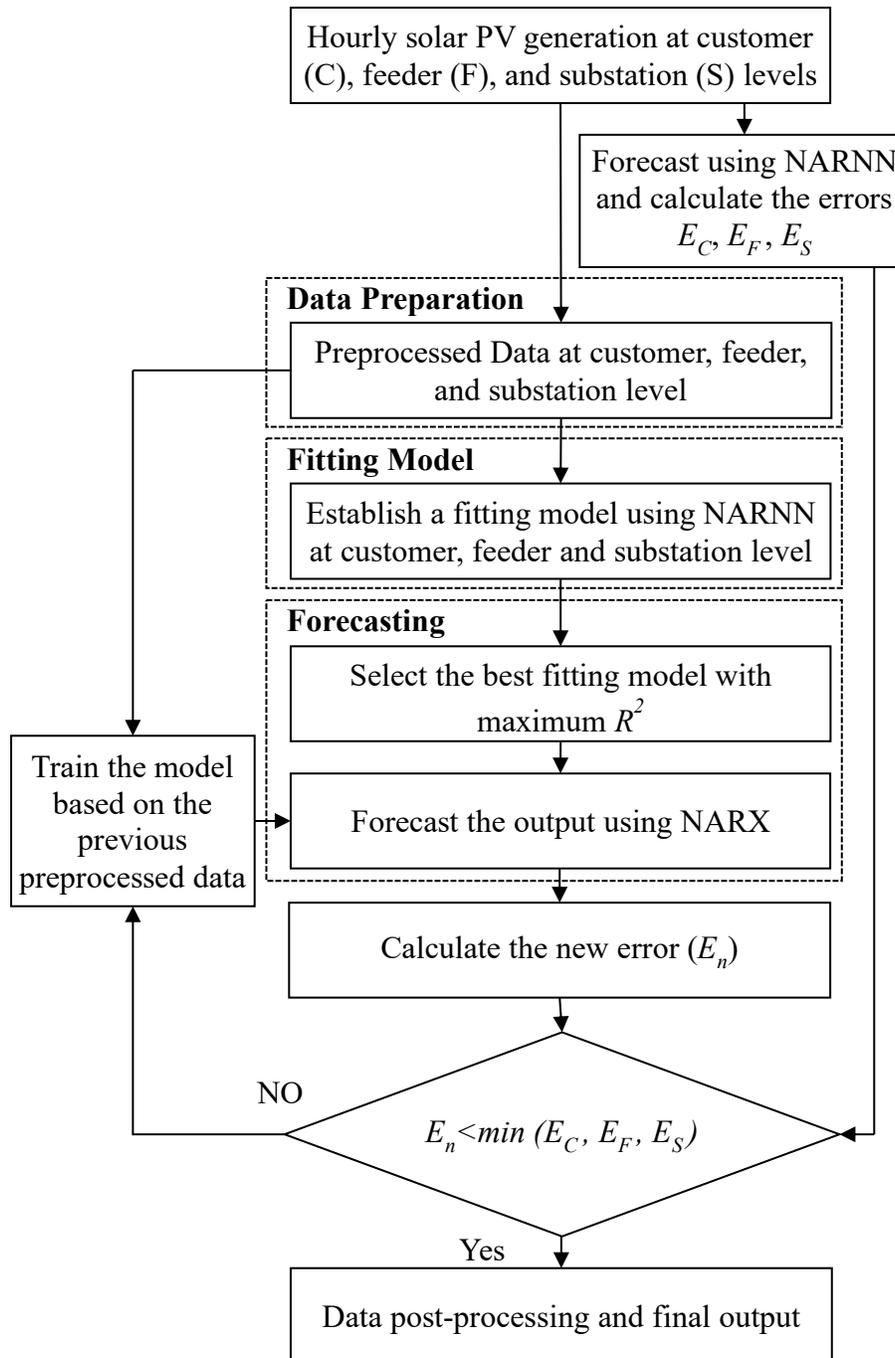
Figure 4.2 The flowchart of the multi-level solar PV generation forecasting.

4.1.1 Data Preprocessing and Adjustment

The data used in the simulation represent the total solar PV generation. The data preparation includes removing offset, normalization, removing nighttime values, and

stationarization. More detail about data preprocessing can be found in [64]. The data preparation is to ensure the quality of dataset before it is inputted to the forecasting model. This step includes the simulation of maximum power generated from solar PV at clear sky conditions. This is achieved by simulating the maximum solar PV generation at clear sky conditions using the system advisory model (SAM) provided by National Renewable Energy Laboratory (NREL) [67]. The maximum solar irradiance along with different metrological inputs in clear condition are fed to SAM in order to simulate the maximum solar PV generation. Fig. 4.3 presents the flowchart for data preparation.
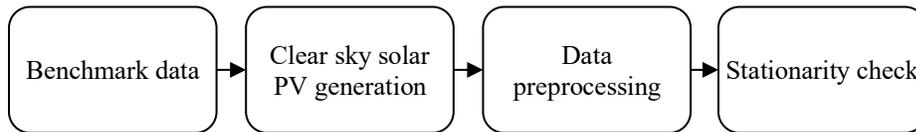
Figure 4.3 The flowchart for data preparation.

4.1.2 Fitting Model

By using NARNN, the fitting model is created for each level. In this respect, the NARNN model utilizes a large set of historical data in order to train the model and then forecast the output. It is applied to the three datasets, including customer, feeder, and substation in order to establish the three fitting models. The best fitting model among the three is selected using the coefficient of determination R^2. The coefficient of determination examines the proportion variance of the predicted fitting model. The coefficient of determination can be expressed mathematically as in (1), where $\overline{P(t)}_{actual}$ is the average of the actual data over the number of samples. The R^2 ranges from 0 to 1, where 0 represents that the fitting model is not predictable, and 1 means that the NARNN

is able to predict the fitting without any error. So, the best selected fitting model among the three is the one with maximum R^2.

$$R^2 = 1 - \left[\frac{\sum_{t=1}^{N}(P(t)_{actual} - P(t)_{forecast})^2}{\sum_{t=1}^{N}(P(t)_{actual} - \overline{P}(t)_{actual})^2} \right] \tag{4.1}$$

4.1.3 NARX Forecasting

NARX is a time series model that predicts the output using historical values $y(t)$ as well as inputs $x(t)$. The NARX model is presented in (4.2), where d is the number of considered historical values. The fitting model is fed as an input to NARX along with the previously preprocessed data. The NARX is trained and the output is forecasted. Fig. 4.4 depicts the architecture of the NARX. The goal is to forecast a day-ahead solar PV generation with a new error E_n, which is less than the minimum of the three errors as shown in the flowchart in terms of a condition.

$$y(t) = f(x(t-1),..,x(t-d),y(t-1),..,y(t-d)) \tag{4.2}$$

4.1.4 Data Post-processing and Accuracy Check

The output from the forecasting model is post-processed by denormalizing, adding nighttime values, and calculating the final solar output as explained in detail in [64]. MAPE and the root mean square error (RMSE) are calculated as in (4.3) and (4.4), respectively.

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{P(t)_{actual} - P(t)_{forecast}}{P(t)_{actual}}\right| \tag{4.3}$$

$$RMSE = \sqrt{\frac{1}{N} * \sum_{t=1}^{N}(P(t)_{actual} - P(t)_{forecast})^2}$$ (4.4)
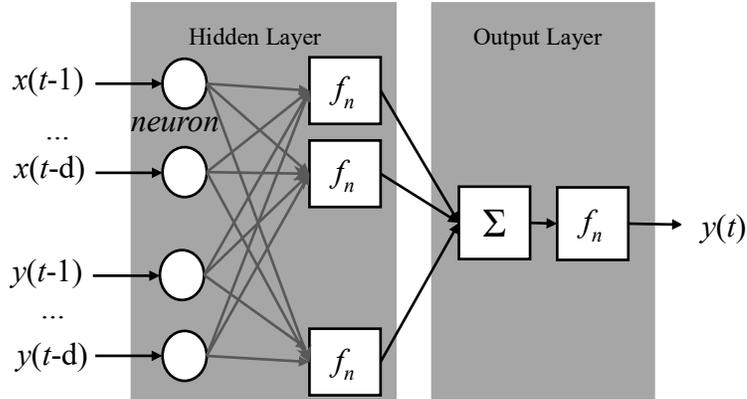
Figure 4.4 The architecture of the NARX

4.2 Numerical Simulations

The hourly solar PV generation of three levels including customer (C), feeder (F), and substation (S), for a specific area in Denver, Colorado are utilized to perform forecasting. The data used in this model are available in [68]. The customer level data are considered as the aggregated customers' solar PVs generation for a selected area. The feeder level data are the aggregated solar PVs generation for each feeder, in which four feeders are considered in this study. Finally, the substation level data are the solar PV generation measured at the substation level. In order to demonstrate the merits of the proposed model, the following four cases with various weather conditions are investigated:

Case 1: Forecast using NARNN for each level without data processing.

Case 2: Forecast using NARX with three-level measurements and data processing.

Case 3: Forecast using NARX with two-level measurements and data processing.

44

Case 4: Forecast using NARX with single-level measurement and data

processing.

Case 1: In this case, by leveraging NARNN, day-ahead solar PV generation is forecasted for all three levels, while ignoring data processing. The calculated MAPE and RMSE for the customer, feeder, and substation levels with different weather conditions are listed in Table 4.1. As highlighted in Table 4.1, the customer level forecast has achieved the minimum MAPE as well as RMSE for the selected weather conditions. This case is considered as a base case, in which the calculated values are utilized in order to demonstrate the effectiveness of using the three-level measurements for forecasting. The objective in the following case is to apply the proposed model in order to get a new error that is less than the minimum achieved under this case.

Table 4.1 Case 1: MAPE and RMSE for the Considered Datasets

Dataset level	Sunny		Cloudy		Partly Cloudy	
	RMSE (kW)	MAPE (%)	RMSE (kW)	MAPE (%)	RMSE (kW)	MAPE (%)
Customer	44.58	4.47	20.54	6.04	36.11	4.09
Feeder	48.81	8.29	23.12	7.34	38.59	6.73
Substation	70.77	10.41	43.65	10.13	53.03	10.37

Case 2: In this case, three-level measurements are preprocessed in order to ensure the quality of the training data fed to the forecasting model. This case includes three forecasting stages: establishing the fitting model from each measurement level using the NARNN, training the NARX model using the previously preprocessed datasets, and forecasting the solar PV generation using the three-level measurements and the fitting

model as input. The fitting model with the minimum MAPE and the maximum R^2 is selected as input to NARX. Table 4.2 exhibits how well the fitting model is established in terms of R^2 and MAPE for the three-level measurements under different weather conditions. As highlighted in Table 4.2, the fitting model established by using the customer level measurement outperforms the ones established by using the feeder and substation measurements. In order to show the merit of using three-level measurements for the same location, the three measurements along with the best fitting model are fed as inputs to NARX to forecast the solar PV generation. The forecast is simulated for the same selected days in Case 1. Table 4.3 exhibits the MAPE and RMSE for the selected days. The forecast errors in this case are less than the minimum achieved in Case 1. Fig. 4.5, 4.6 and 4.7 depict the forecasted and actual solar PV generation for the considered sunny, cloudy, and partly cloudy days, respectively.

Table 4.2 The Fitting Model MAPE and R^2 for the Considered Levels

Dataset level	Sunny		Cloudy		Partly Cloudy	
	MAPE (%)	R^2	MAPE (%)	R^2	MAPE (%)	R^2
Customer	2.39	0.9987	2.29	0.9956	3.49	0.99
Feeder	3.78	0.996	2.80	0.9943	4.02	0.988
Substation	5.95	0.9873	4.98	0.9821	5.37	0.986

Case 3: In this case only two measurements at customer and feeder levels are used for forecasting. The preprocessed data along with the best selected fitting model are fed to NARX. The forecasting performance of this case is shown in Table 4.3. In sunny day, this case has reduced the MAPE compared to Case 1 by 47%. In cloudy and partly

cloudy weather conditions, Case 3 has reduced the MAPE compared to Case 1 by 61% and 19%, respectively.

Case 4: To exhibit the effectiveness of the three-level measurements, Case 2 is repeated, but only one measurement (customer level) is included as an input to NARX. Similar to the previous case, the best fitting model based on MAPE and R^2 is fed to NARX along with preprocessed customer level measurement. Table 4.3 shows the forecast error using NARX with single-level measurement comparing to NARX with three-level measurements, two-level measurements, and the minimum forecast error among the single-level measurement using NARNN without data processing. A single-level measurement considerably improves the results over NARNN method; however, achieve not as good solution as in two previous cases with three- and two-level measurements.

Table 4.3 The MAPE for Different Case Studies

Weather Condition	Minimum MAPE (Using NARNN without data processing)	MAPE (Using NARX and three-level processed data)	MAPE (Using NARX and two-level processed data)	MAPE (Using NARX and single-level processed data)
Sunny	4.47	1.67	2.38	3.14
Cloudy	6.04	2.10	2.36	2.44
Partly Cloudy	4.09	2.69	3.30	3.39
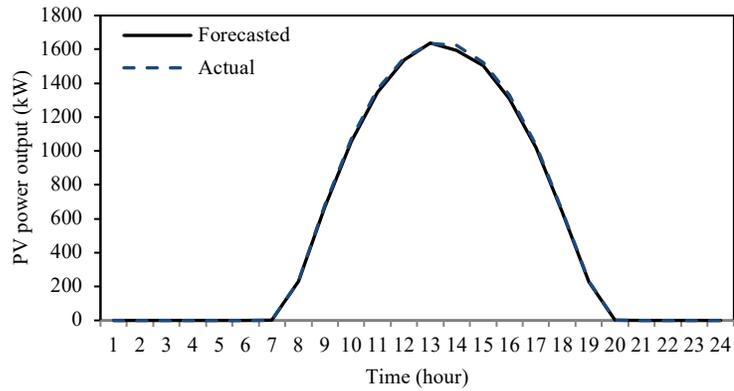
Figure 4.5 Actual and forecasted solar PV generation in a sunny day
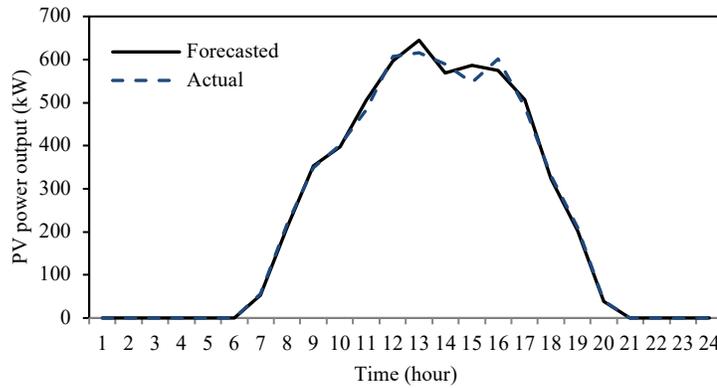
Figure 4.6 Actual and forecasted solar PV generation in a cloudy day
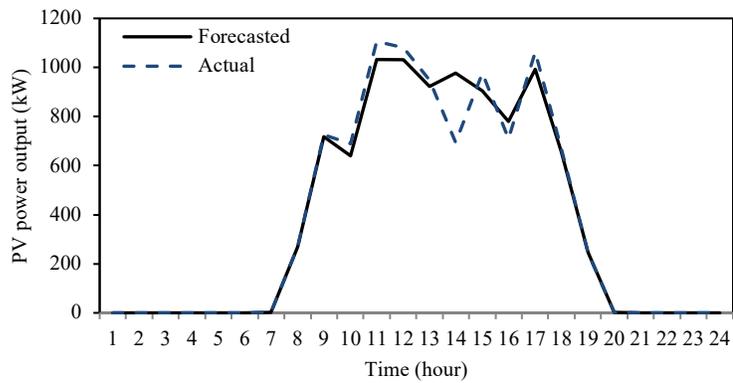
Figure 4.7 Actual and forecasted solar PV generation in a partly cloudy day

4.3 Discussion

As shown in Table 4.3, the model minimizes the forecast error to outperform the minimum error reported at customer level. The proposed model has reduced the error

compared to the minimum error in Case 1 by 63%, 65%, and 34% for sunny, cloudy, and partly cloudy weather conditions, respectively. Moreover, the merit of using three-level measurements is shown by comparing the forecast error using the proposed model with applying two-level measurements to the model as in Case 3. The MAPE is reduced by 30%, 11%, and 18% for sunny, cloudy, and partly cloudy weather conditions, respectively. The three-level measurement also outperforms Case 4 in which only single-level measurement are included. The three-level measurements model has reduced the MAPE by 47%, 14%, and 21% for sunny, cloudy, and partly cloudy weather conditions, respectively. The previous cases have shown that forecasting performance is greatly impacted by the historical data used to train the model. Multiple historical data for a specific location along with an appropriate data processing will improve the training step and minimize the forecasting error.

Chapter Five: Co-Optimization Generation and Transmission Planning for

Maximizing Large-Scale Solar Integration

5.1 Co-optimization Generation and Transmission Model Outline

Fig. 5.1 depicts the proposed co-optimization generation and transmission planning model for maximizing large-scale solar PV hosting capacity. The objective of this problem is to minimize the investment cost, for new generation units and transmission lines, plus the system operation costs. The planning problem aims at providing new generation units and transmission lines required for increasing PV hosting capacity. In other words, the system is upgraded to maximize the amount of solar PV that can be integrated to the grid. The objective is subject to prevailing operation and planning constraints associated with generation units, transmission lines, and solar PV, which will be discussed in detail in the next section.

The planning problem is analyzed on an annual basis. A year is broken down into several days at which the maximum solar variability is expected to occur. The number of days is regarded as a tradeoff between the computational complexity and the accuracy of the proposed planning model.

The Benders decomposition is applied to help with the computational complexity of the proposed problem. The Benders decomposition is widely applied in long-term expansion planning problems as discussed in [69]–[72]. In this dissertation, the planning

problem is decomposed into a master problem and two subproblems. The master problem determines optimal investment plan for new generation units and transmission lines, and the subproblems provide feasibility check and optimal operation.

The optimal plan determined in the master problem is sent to the subproblems. The first subproblem will minimize the system network violations based on the calculated plan. If the feasibility check fails in subproblem 1, a feasibility cut is formed and sent back to the master problem to revise the solution of the next iteration of the master problem. The optimal operation is calculated in subproblem 2. In this subproblem, the operation cost is minimized considering the prevailing system operation constraints. The optimality is checked by calculating the upper bound of the original planning problem and comparing it with the lower bound which comes from the master problem. If the solution is not optimal, an optimality cut is generated and added to the master problem for the next iteration. This process will continue iteratively until a secure and optimal planning solution is achieved.

Capacity factor is one of the key factors associated with solar PV sizing. Capacity factor for solar PV generation unit is a percentage defined as the amount of energy produced by solar PV during some period of time over its total energy it produces if it runs at its full output for the same period.[73]. Due to climate conditions and zero solar irradiance at nighttime, solar PV is commonly operated at low capacity factor ranging from 10% to 25% [74]. Another decisive factor in integrating large-scale solar PV is its generation variability. Due to sudden changes in solar PV generation, the system operator may not be able to accommodate the variabilities which leads to curtailing some of the

PV generation. In order to represent the capacity factor of solar PV as well as its variability in the planning model, the normalized solar generation is used. This normalized solar generation is obtained from long-term forecasts, where to find the actual PV generation, this normalized generation is multiplied by annual PV installed capacity.
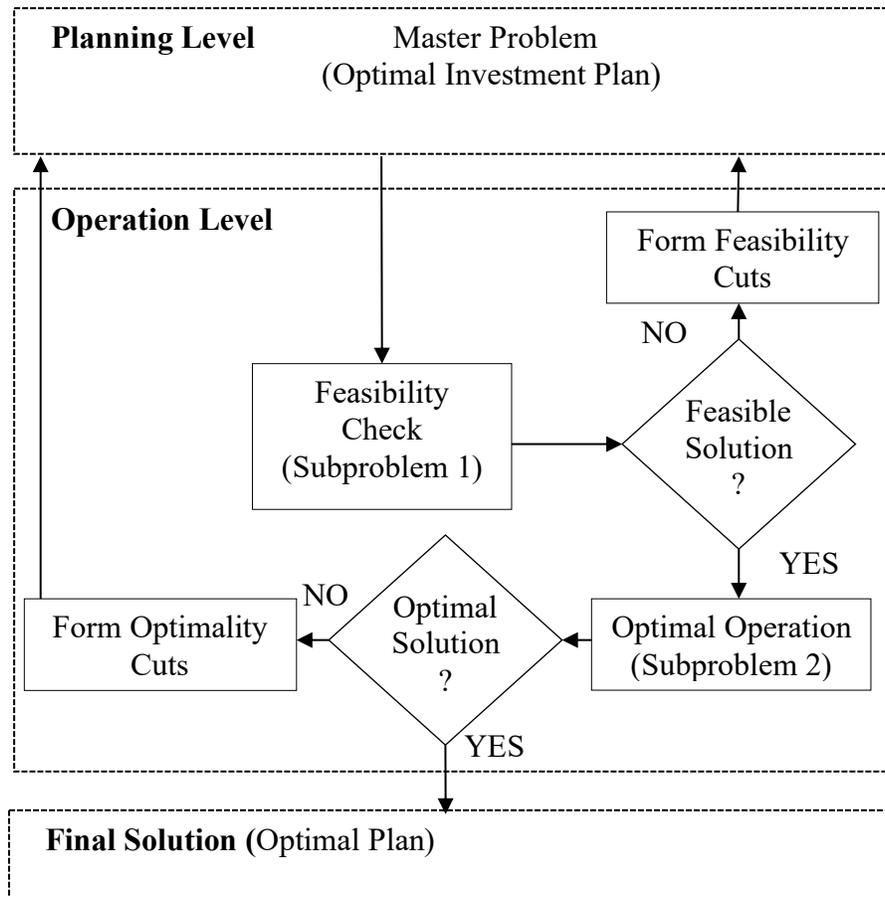
Figure 5.1 The architecture of the proposed planning model.

5.2 Problem Formulation

The objective of the proposed model over the planning horizon is to minimize the planning cost of new generation units and transmission lines required for increasing large-scale solar PV hosting capacity. The planning cost of the new solar PVs is ignored

in this study and only the planning cost to upgrade the system in order to accommodate more solar PVs is considered. The objective function consists of the investment cost of new generation units, new transmission lines, as well as the system operation cost over the planning horizon.

The ultimate solution for this problem economically determines the size and the installation time for adding new generation units, transmission lines, and the maximum possible large-scale solar PV capacity. The solution steps of the proposed model as follows:

A) Step 1 (optimal investment plan)

The first step is to calculate the optimal investment plan and projected operation cost as shown in (5.1). By ignoring the investment cost of new large-scale solar PVs, the system aims at maximizing the installed solar PV. The size and the location of the expanded generation units and transmission lines are optimally determined by the model in each year. $\kappa t = 1/(1+d)^{t-1}$ is the present-worth value coefficient, introduced to evaluate the objective function in terms of discounted cost. The projected operation cost will be achieved from the optimality cuts calculated in the optimal operation subproblem. This term will be considered 0 in the first iteration.

The objective function in (5.1) is subject to investment constraints (5.2)-(5.4). The construction and commissioning time associated with installing new generation units and transmission lines are considered as in (5.2). To ensure there is no recurrence in calculating the capital cost, once either of the candidate generation units or transmission lines are installed, the corresponding investment state will be fixed as 1 for the remainder

of the planning horizon as in (5.3). The main purpose of the proposed model is to increase the large-scale solar PV hosting capacity on an annual basis. The total installed solar PV capacity at each year should be greater than or equal to the installed solar PV size in the previous year as in (5.4). Nevertheless, in each year, the size and the location of the installed PV is calculated in an optimal fashion.

$$\min \mathrm{MP}$$

$$MP \geq \sum_{t} \sum_{j \in \mathrm{CG,CL}} \kappa_t IC_j P_{jt}^{\max,C} (x_{jt} - x_{j(t-1)}) + \sum_{t} \kappa_t \Gamma_t \qquad (5.1)$$

$$x_{jt} = 0 \qquad\qquad \forall t < \mathrm{CT}_j, \forall j \in \mathrm{CG,CL} \qquad (5.2)$$

$$x_{j(t-1)} \leq x_{jt} \qquad\qquad \forall j \in \mathrm{CG,CL}, \forall t \qquad (5.3)$$

$$P^s_{b(t-1)} \leq P^s_{bt} \qquad\qquad \forall b, \forall t \qquad (5.4)$$

At high solar PV penetration, the system is expected to experience over-generation which accordingly violates the system load-supply balance. To tackle this obstacle, generation curtailment may be required in the system. Although the curtailment reduces the capacity factor of the renewable resource, it can alleviate the over-generation and balance the system. Curtailment can be as a result of over-generation (i.e., the renewable generation exceeds the demand), congestion in transmission lines, or voltage and interconnection issues. The Electric Reliability Council of Texas ERCOT curtailed around 17% of wind generation in 2009 [75]. In 2014, ERCOT completed a project to add new transmission lines of a capacity of 19 GW in order to accommodate the expansion in the renewable energy resources [76]. The CAISO predicts at 40% and 50% RPS, the generation curtailment is about 6.5% and 9% of renewable output, respectively

[75]. The curtailment of variable energy resources can be reduced by including a battery energy storage systems (BESS). In [77], 99% reduction in wind energy curtailment is achieved by including a BESS. In this dissertation, a target value for solar PV curtailment is introduced in the model.

B) Step 2 (feasibility check):

Once the optimal planning decisions for installing new units and transmission lines are made in the master problem, the new system topology is sent to subproblem 1 to examine the feasibility of the proposed plan. This task is accomplished by minimizing the potential power mismatches via introducing two nonnegative slack variables to the load balance equation at each bus. The objective is to minimize the system violations based on the master problem solution by minimizing the sum of these slack variables as in (5.5).

Equation (5.6) shows load balance equation at bus b, where *SL1* and *SL2* represent virtual generation and virtual load, respectively, where virtual load translates to generation curtailment meaning the solar PV generation must be curtailed. Since there is no energy storage system incorporated in the problem, a generation curtailment from solar PV is expected especially at higher solar PV penetrations. The investment states associated with generation unit and transmission line as well as the optimal size of solar PV are obtained from the optimal planning master problem. These calculated variables are substituted for local variables in order to obtain related dual variables (5.7) and (5.8). This problem is subject to existing and candidate generation unit and transmission line constraints. These constraints represent the capacity of existing and candidate generation

units (5.9) and (5.10), the existing and candidate units ramp up and ramp down rate limits (5.11) and (5.12). The DC power flow calculation for the existing and candidate lines is presented in (5.13) and (5.14), respectively. The existing and candidate transmission line flows are respectively presented in (5.15) and (5.16). The phase angle of reference bus is defined in (5.17).

$$\text{Min } r_t = \sum_q \sum_h \sum_b (SL_{bhqt,1} + SL_{bhqt,2}) \tag{5.5}$$

$$\sum_{j \in G_b} P_{jhqt} + \sum_{j \in L_b} PL_{jhqt} + \tilde{\alpha}_{bhqt} \hat{P}^s_{bt} - D_{bhqt} + SL_{bhqt,1} - SL_{bhqt,2} = 0 \quad \forall b, \forall h, \forall q \tag{5.6}$$

$$x_{jt} = \hat{x}_{jt} \leftrightarrow \lambda_{jt} \qquad \forall j \in \text{CG,CL} \tag{5.7}$$

$$P^s_{bt} = \hat{P}^s_{bt} \leftrightarrow \pi_{bt} \qquad \forall b \tag{5.8}$$

$$0 \leq P_{jhqt} \leq P^{\max,E}_j \qquad \forall j \in \text{EG}, \forall h, \forall q \tag{5.9}$$

$$0 \leq P_{jhqt} \leq P^{\max,C}_j \hat{x}_{jt} \qquad \forall j \in \text{CG}, \forall h, \forall q \tag{5.10}$$

$$P_{jhqt} - P_{j(h-1)qt} \leq RU_j \qquad \forall j \in \text{EG,CG}, \forall h, \forall q \tag{5.11}$$

$$P_{j(h-1)qt} - P_{jhqt} \leq RD_j \qquad \forall j \in \text{EG,CG}, \forall h, \forall q \tag{5.12}$$

$$PL_{jhqt} = \frac{\theta_{mhqt} - \theta_{nhqt}}{x_{mn}} \qquad \forall j \in \text{EL}, \forall h, \forall q \tag{5.13}$$

$$\left| PL_{jhqt} - \frac{\theta_{mhqt} - \theta_{nhqt}}{x_{mn}} \right| \leq L(1 - \hat{x}_{jt}) \quad \forall j \in \text{CL}, \forall h, \forall q \tag{514}$$

$$\left| PL_{jhqt} \right| \leq PL^{\max,E}_j \qquad \forall j \in \text{EL}, \forall h, \forall q \tag{5.15}$$

$$\left| PL_{jhqt} \right| \le PL_j^{\max.C} \hat{x}_{jt} \qquad \forall j \in CL, \forall h, \forall q \tag{5.16}$$

$$\theta_{bhqt} = 0 \qquad b = \mathrm{ref}, \forall h, \forall q \tag{5.17}$$

The curtailment is included in the model by introducing the target value (δ_t) in the Benders cut (5.18). If the proposed objective in (5.5) is less than or equal to the target value for solar PV generation curtailment, the problem will move forward to the optimal operation subproblem. Otherwise, the Benders cut (5.18) will be formed and added to the master problem for the next iteration. The target value for solar PV generation curtailment is expected to increase gradually as the solar PV penetration grows through the planning horizon.

Here λ and π are dual values of constraints (5.7) and (5.8), respectively. The Benders cut (5.19) demonstrates that the violation could be mitigated by revising the investment plan. In other words, this cut recalculates the capacity signals for the investment of new generating units, new solar PVs, and transmission lines in case the existing ones cannot satisfy the system feasibility. Nevertheless, the iterative procedure continues until a secure plan that satisfies the system feasibility is obtained.

$$r_t + \sum_j \lambda_{jt}(x_{jt} - \hat{x}_{jt}) + \sum_b \pi_{bt}(P^s_{bt} - \hat{P}^s_{bt}) \le \delta_t \qquad \forall j \in CG, CL, \forall t \tag{5.18}$$

C) Step 3 (optimality check):

After the feasibility of the calculated investment plan is ensured, the optimality of the solution will be checked in the optimal operation subproblem. The objective of the optimal operation subproblem is to minimize the operating cost for every year as shown in (5.19).

$$\min Q_t = \sum_q \sum_h \sum_j \kappa_t c_{jhqt} P_{jhqt} \qquad\qquad \forall j \in \text{EG,CG} \qquad\qquad (5.19)$$

Subject to (9)-(17) and (20).

$$\sum_{j \in G_b} P_{jhqt} + \sum_{j \in L_b} PL_{jhqt} + \tilde{\alpha}_{bhqt} \hat{P}^s_{bt} - \hat{SL}_{bhqt,2} = D_{bhqt} \quad \forall b, \forall h, \forall q \qquad\qquad (5.20)$$

The load balance equation is presented in (20), where the solar PV generation curtailment calculated in Subproblem 1, and then introduced in the load balance equation.

If the proposed plan is not optimal, a Benders cut will be formed and added to the master problem for the next iteration. Leveraging the proposed Benders cut (5.21), the lower bound of objective function in the master problem is restricted.

$$\Gamma_t \geq \hat{Q}_t + \sum_j \lambda_{jt}(x_{jt} - \hat{x}_{jt}) + \sum_b \pi_{bt}(P^s_{bt} - \hat{P}^s_{bt}) \qquad \forall j \in \text{CG,CL}, \forall t \qquad (5.21)$$

An optimal solution of the co-optimization generation and transmission planning problem is calculated through the iterative process amongst the master problem and subproblems. The master problem solution is regarded as the lower bound for the optimal solution. Accordingly, the upper bound for the original problem is calculated by utilizing the result from the optimal operation subproblem as in (5.22). This solution provides the upper bound of the objective function of the co-optimization generation and transmission planning. This upper bound is utilized to check the optimality of the solution, so that the stopping criterion is specified on the basis of this solution. The optimal solution for the proposed co-optimization generation and transmission planning problem is obtained once the lower and upper bounds are converged, according to a convergence criterion as in (5.23).

$$Y = \sum_t \sum_j \kappa_t IC_{jt} P_{jt}^{\max,C}(x_{jt} - \hat{x}_{j(t-1)}) + \sum_t \sum_b \kappa_t IC^s(P_{bt}^s - \hat{P}_{b(t-1)}^s) + \sum_t \kappa_t \Gamma_t$$

$$\forall j \in CG,CL \tag{5.22}$$

$$\frac{Y - MP}{Y + MP} \leq \varepsilon \tag{5.23}$$

5.3 Numerical Simulations

Four cases based on a modified six-bus test system as well as the IEEE 118-bus system are analyzed to demonstrate the effectiveness and the performance of the proposed co-optimization generation and transmission planning model. The proposed model is formulated as mixed integer linear programming (MILP) and solved in a high performance computing server with Intel Xeon E7 2.3 GHz processor and 96 GB RAM using CPLEX 12.6.

A ten-year planning horizon is considered. The six-bus system data is available in [37]. The candidate generation units and transmission lines data are provided in Tables 5.1, and 5.2, respectively, where a set of four candidate generation units and four candidate transmission lines are regarded as planning options. The investment cost for solar PV is ignored to maximize its deployment. The annual load growth is considered 5% per year. The forecasted yearly peak load for the six-bus system is listed in Table 5.3, where the load is distributed amongst buses 3, 4, and 5 at the rate of 20%, 40%, and 40%, respectively. The data associated with the modified IEEE 118-bus system are provide in [37]. The system has 118 buses, 54 units, and 186 branches. In order to reduce the computational burden, days which have worst solar PV ramping rate are considered. The following cases are studied for the six-bus system:

Case 1: Solar PV integration ignoring the co-optimization generation and transmission planning.

Case 2: Solar PV integration supported by generation expansion planning.

Case 3: Solar PV integration supported by transmission expansion planning.

Case 4: Co-optimization generation and transmission planning.

An additional case is considered to study a relatively bigger test system:

Case 5: Co-optimization generation and transmission planning to support solar PV integration for the IEEE 118-bus system.
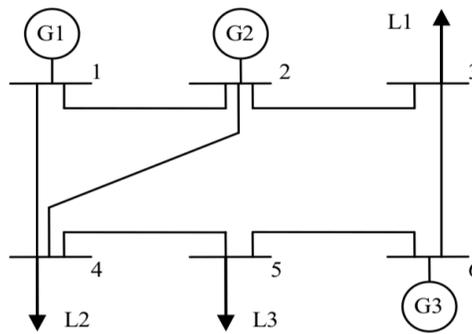
Figure 5.2 IEEE Six-Bus System

Table 5.1 Existing and Candidate Generation Unit Data of Six-Bus System

Unit No.	Bus No.	Generating Capacity (MW)	Investment Cost ($/kW)	Operation Cost ($/MWh)	Commissioning Time (Year)
1	1	100	Existing	15	-
2	2	100	Existing	18	-
3	2	50	Existing	23	-
4	1	100	200	15	3
5	2	80	270	21	2
6	2	60	250	24	2
7	3	20	250	24	1

Table 5.2 Existing and Candidate Transmission Line Data of Six-bus System

Line No.	From Bus	To Bus	X (p.u)	Capacity (MW)	Investment Cost ($/kW)	Commissioning Time (Year)
1	1	2	0.17	70	Existing	-
2	2	3	0.037	70	Existing	-
3	1	4	0.258	80	Existing	-
4	2	4	0.197	80	Existing	-
5	4	5	0.037	50	Existing	-
6	5	6	0.14	80	Existing	-
7	3	6	0.018	80	Existing	-
8	1	2	0.17	70	23	2
9	2	3	0.258	70	23	2
10	2	4	0.258	80	24	2
11	3	6	0.018	70	24	2

Table 5.3 Forecasted Yearly Peak Load of Six-bus System

Year	1	2	3	4	5
Peak (MW)	209	219	230	241	254
Year	6	7	8	9	10
Peak (MW)	266	280	294	308	324

Case 1: The large-scale PV hosting capacity is calculated without considering any system upgrade. A total solar capacity of 123.95 MW is installed at bus 5 in year 1. Load is mainly supplied by unit 1 as the least expensive unit. Unit 2 is the next economic unit after unit 1, however it is partially dispatched due to congestion in line 2-3. The system accommodates a considerable amount of solar PV in the first six year to reach a total capacity of 162.55 MW (with a breakdown of 141.78 MW and 20.77 MW at buses 5 and 4, respectively). In year 7, however, the system has neither adequate generation to supply

the load nor adequate network capacity to accommodate additional solar PV, therefore it experiences a load curtailment of 12.65 MWh. In the first seven years, the system experiences solar PV generation curtailment mainly due to overgeneration. However, once there is no further installation of solar PV and considering the load growth, the curtailment is reduced to zero. By the end of the planning horizon the total installed solar PV capacity reaches 163 MW which supplies 21% of the total annual load.
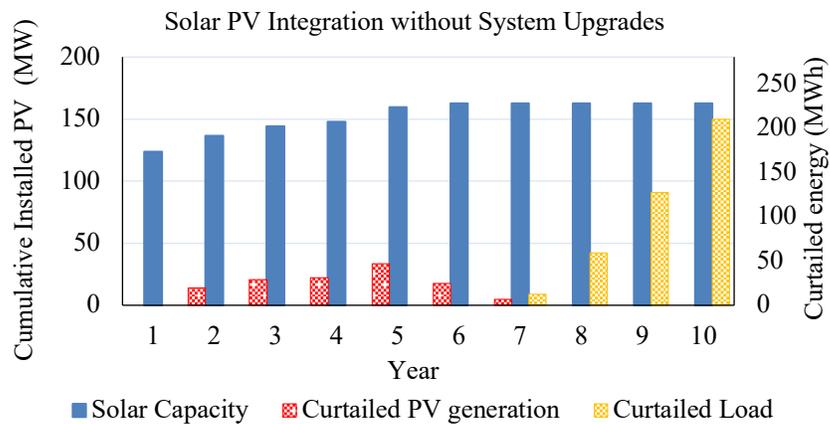
Figure 5.3 Installed solar capacity and curtailed load and solar energy in Case 1

Case 2: Solar PV integration is supported by generation expansion planning in this case, i.e., the proposed planning model is used while ignoring transmission line installation. The load is mainly supplied by unit 1 and when solar PV generation is available, generation from both units 2 and 3 goes to zero in most times. The system keeps accommodating more solar PV while maintaining the feasibility of transmission flows to reach a total of 157.53 MW at bus 3 in year 5. About 25% of the load supply in year 5 comes from solar PV, where the system curtails a total of 150 MWh of solar PV energy due to excess generation. In year 6, a total of 4.92 MW and 12.80 MW is installed at buses 3 and 5, respectively. In year 7, a new generation capacity needs to be installed

to satisfy the load growth especially at hours where solar PV generation is not available. In year 8, line 5-6 is congested which causes unit 2 to dispatch at its maximum capacity in order to supply the load at bus 4. Due to repeated congestion at both lines 2-3 and 5-6, the system has to curtail a total of 150 MWh of solar PV energy in year 8. In year 9, when solar PV generation is not available or low, units 1 and 2 are dispatched at their maximum capacity. Moreover, since lines 1-2, 2-3 and 5-6 are congested, a new generation needs to be installed to supply the load at bus 4. The available options are to install at bus 1 or 2. The installation of a new unit at bus 2 would cause more congestion between buses 2 and 3. As a result, the model selects unit 4 to be installed. Once candidate unit 4 is installed, unit 2 reduces its generation as it is cheaper to supply the load from the new installed unit.   By the end of the planning horizon, the total installed solar PV capacity reaches 192 MW which supplies 26% of the annual load, i.e., a 5% increase compared to Case 1 at the expense of high investment cost of $ 17.3 M.
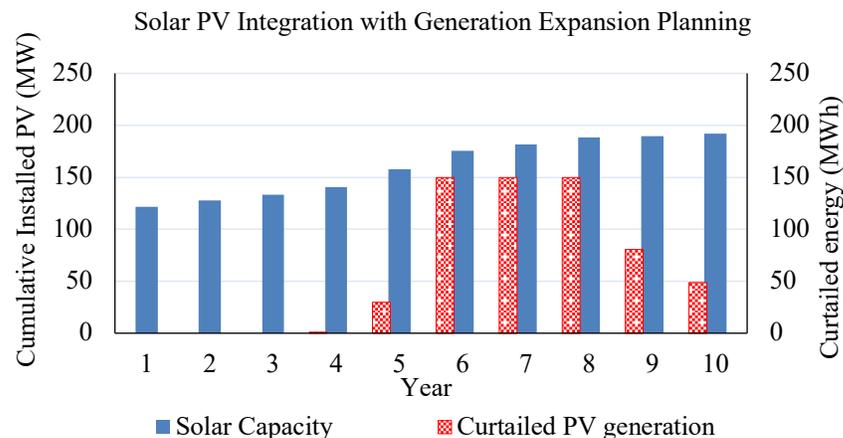
Figure 5.4 Installed solar capacity and curtailed solar energy in Case 2

Case 3: Solar PV integration is supported by transmission expansion planning in this case, i.e., the proposed planning model is used while ignoring generation unit

installation. Similar to previous cases, line 2-3 often experiences a congestion, causing a reduction in generation of unit 2. In year 1, the total installed PV capacity is 122.36 MW. Candidate lines 2-3 and 1-4 are installed in the second year to increase the network capacity and hence allow more large-scale solar PV installations. In year 4, a total of 86.16 MW solar PV capacity is installed at bus 5. Also, candidate line 1-2 is installed in year 4 to increase the network capacity. In year 7, the system needs to install new generation units to meet the demand growth; however, since no generation expansion is considered in this case, the system curtails 7.45 MWh of the load. Compared to Case 1, the load curtailment is reduced due to the increase in solar PV hosting capacity driven by the increase in the network capacity. By the end of the planning horizon, the total installed solar PV capacity is 266 MW, which supplies 34%, a much higher percentage than in previous cases. The planning cost for adding the new lines is $ 4.58 M.
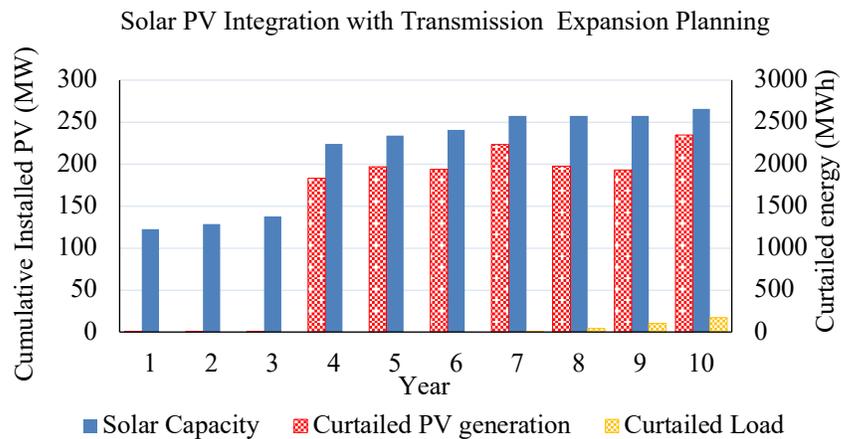
Figure 5.5 Installed solar capacity and curtailed load and solar energy in Case 3

Case 4: In this case, the proposed co-optimization generation and transmission planning model is employed to maximize solar PV capacity. Fig. 5.7 shows the annual installed solar PV over the planning horizon.

In the first year, the total installed solar PV capacity is 122.55 MW where 120.09 MW and 2.46 MW are installed at buses 5 and 6, respectively. To remove congestion in line 2-3, candidate line 2-3 is installed in year 2. Line 4-5 is congested at the first year caused by the installation of solar PV at bus 5. During nighttime hours or the unavailability of solar generation due to weather conditions, unit 1 is dispatched at its maximum capacity and the remaining is supplied by units 2 and 3. However, during the afternoon periods where solar PV generation is at its maximum, the load is supplied by solar PV as it has no operation cost. Accordingly, generation from units 1 and 2 is reduced and the generation from units 3 is dropped to zero. In year 1, 23% of the annual demand is supplied by solar PV. The total solar energy curtailed in year 1 is 30.42 MWh, which is mainly due to solar PV over-generation. In year 2, the total solar PV capacity is increased to 150.15 MW where 8.19 MW and 19.41 MW are installed at buses 5 and 6, respectively. In year 2, 27% of the annual demand is supplied by solar PV. A total of 5.3 GWh solar PV energy is curtailed in year 2 due to excess solar PV generation, representing 1.78% of its total annual energy generation.

A total of 15.04 MW of solar PV capacity is added in year 3. In year 4, an additional of 144.89 MW is installed at bus 3, supplying 39% of the annual demand. In year 5, a total of 100.69 MW solar PV is installed at bus 5 which increases the annual demand supplied by solar PV to 40%. The installed capacity of the solar PV meets the load growth which delayed any additional unit to be installed to meet the load growth. In year 5, line 3-6 experiences congestion which results in additional curtailment of solar PV generation. As a result, the candidate line 4 is installed. As the hosting capacity of the

solar PV increases, the system experiences more solar PV energy curtailment due to excess generation. In year 5, the total solar PV energy curtailed over the year is 10.78 GWh. In year 6, the system accommodates more solar PV where an additional 14.86 MW is installed at bus 3. In year 7, the system needs to install new generation capacity in order to meet the load growth especially at nighttime hours when solar irradiance is not available. Considering the remaining years in the planning horizon, candidate units 6 and 7 are the available options. Unit 6 is installed as it has higher ramp up/down limits, which can manage any substantial ramps in the net load caused by the solar PV variability. At the end of the planning horizon, the solar PV reaches a total capacity of 451 MW and supplies 40% of the annual demand. The total planning cost for reinforcing the system with new lines and units is $ 14.3 M. In year 10, the total curtailed solar PV energy is 10.14 GWh, which represents 34% of the total solar PV generation.

Table 5.4 Candidate Unit and Line Installation Year for Six Bus-System

	Candidate Unit				Candidate Line			
	4	5	6	7	1-2	2-3	2-4	3-6
Case 2	9	-	-	7	-	-	-	-
Case 3	-	-	-	-	4	2	2	-
Case 4	-	-	7	-	-	2	-	5
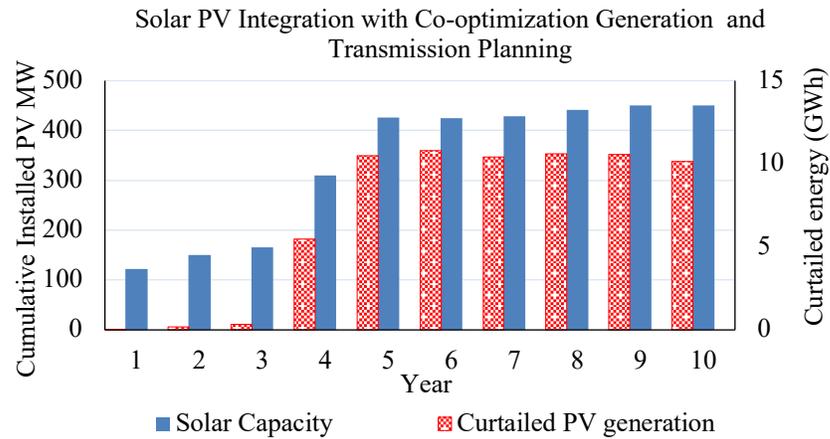
Figure 5.6 Installed solar capacity and curtailed solar energy in Case 4

Case 5: To demonstrate the effectiveness of the proposed model on a relatively larger system, the proposed co-optimization generation and transmission planning problem is solved for the IEEE 118-bus system. The list of existing and candidate generating units and transmission lines for the IEEE 118-bus system is available in [37]. The existing generation capacity is 7,500 MW. The peak load is 4090 MW and the load growth is considered to be 5%.

In the first year, a total of 1.35 GW of solar PV is installed. Candidate lines 80-99 and 94-100 are installed in year 1 in order to accommodate the anticipated increase in solar PV penetration. At the second year, a total of 0.8 GW is added to the system at buses 37, 56, 100 and 113 and candidate lines 8-30 and 17-113 are further installed. Candidate units 2, 3 and 11 are installed in year 2 at the buses 10, 12, and 113 in order to mitigate any ramps caused by solar PV variability. Unit 9 is installed in year 5 in order to transfer power to the loads due to congestion in line 8-5. A 2.5 GW of solar PV is added in year 6. To elevate this congestion and provide solar PV generation with sufficient access to transmission lines capacity, candidate lines 77-82 and 82-83 are installed in

67

year 6. Also, the total curtailed solar PV energy is 2,835 GWh, which represents 23% of the generated solar PV.

The total cumulative solar PV installed by the end of the planning horizon is 7.9 GW which supplies 38% of the annual load. The total planning cost in order to maximize the solar PV hosting capacity and accommodate such large-scale of solar PV is $100 M.

Table 5.5 Summary for Six-Bus system Cases

	Case 1	Case 2	Case 3	Case 4
Solar PV penetration	21 %	26 %	34%	40 %
Total Planning Cost ($ M)	0	17.3	4.58	14.3
Installed PV (MW)	163	192	266	451

Chapter Six: Investigating the Voltage Fluctuation Caused by Solar PV Generation

Variability in Distribution Grids

The installation of solar PV at both distribution and transmission levels has been streamlined due to considerable enhancements in power electronics. However, at high penetration levels the market value of this resource drops as discussed in [79]. The paper reveals that the value of solar can decline up to 50-80% at 15% penetration. It further discusses various integration options in order to mitigate the drop in the variable generation value. Three factors that negatively impact solar, especially at high penetration levels, are listed in [80] as the low capacity credit, the reduction in utilizing dispatchable resources, and the over generation produced by variable generation.

The negative impacts of solar generation in transmission and distribution result from two key characteristics: the variability and uncertainty of solar generation. The first characteristic is the intermittent supply in solar generation due to solar irradiance variations and other metrological factors such as the movement of the clouds. The second one is the difficulty to predict in advance the time, the duration, and the magnitude of this variability [64]. These two characteristics must be managed by the power system operator to maintain the system adequacy and reliability, while minimizing the curtailment from solar generation. In [81], a study presents the value of solar PV with the use of energy storage   to ensure the availability of the solar power throughout the day and to reduce the variability to allow solar generation to be integrated into the grid with less

curtailment. Solar PV is considered a non-dispatchable energy source due to its output variability. Therefore, large-scale solar PV integration may cause instability and power quality issues in power systems due to fluctuations in generation output. Power system operators should be ready to clear any imbalances in the power system caused by under/over generation from solar PV by dispatching other dispatchable resources, utilizing demand response, and/or curtailing the PV generation. These solutions commonly result in a higher operational cost. Solar forecasting is considered a solution for predicting solar generation variability; however, it greatly depends on the forecasting accuracy, especially at high resolutions. The energy storage system is also proposed as a solution to capture solar PV fluctuations. In [82], the author studies the solar PV fluctuation and determines the sizing requirements for an energy storage unit to mitigate its fluctuation. A 1.2 MW PV in Hawaii is studied in [83] to investigate the effect of the unmitigated ramp rate. The study has revealed that the ramp rate could reach up to 63%/minute of the rated capacity. The integration of the electric vehicle chargers can improve the solar PV integration into the grid by reducing the solar PV output ramp rate as explained in [84]. Hundreds of distributed batteries are used in a control algorithm that was designed in [85] to smooth the PV generation. The trade-off between the smoothing and the size of the energy storage is also discussed. The Puerto Rico Electric Power Authority (PREPA) has included a ramp rate limit of 10%/min of the rated capacity for both wind and solar generation. Likewise, the Germany transmission and distribution operator has imposed the same ramp rate limit [86].

This chapter investigates the impact of solar ramp rate on the IEEE 33-bus test system by integrating a 1.04 MW solar PV system. This chapter exhibits the changes in the voltage when the ramp rate limitation has been applied to solar output. The ramp rate limitation is 10%/minute of the installed capacity. The system is studied for one selected hour, representing the highest ramps.

6.1 Data analyses

The solar irradiance and generation with different ramp rates of one cloudy day is shown in Figure 6.1(a). It is clear from the figure that ramps are small in morning and evening hours due to clear weather conditions. However, ramp rate in the middle of the day is high due to solar irradiance variability. Passing clouds can cause ramp rates to reach up to 500 kW/min as shown in Figure 6.1(b). Ramp rate can be calculated under different time scales using (6.1). The ramp rate should be curtailed at a maximum of 10% of the installed capacity in this work.

$$RR = \left| \frac{p(t) - p(t + \Delta t)}{\Delta t} \right| \qquad (6.1)$$
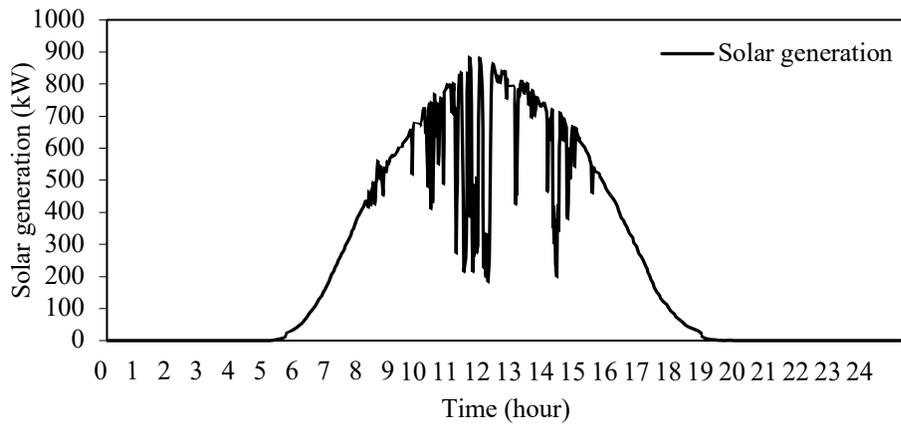
(a)
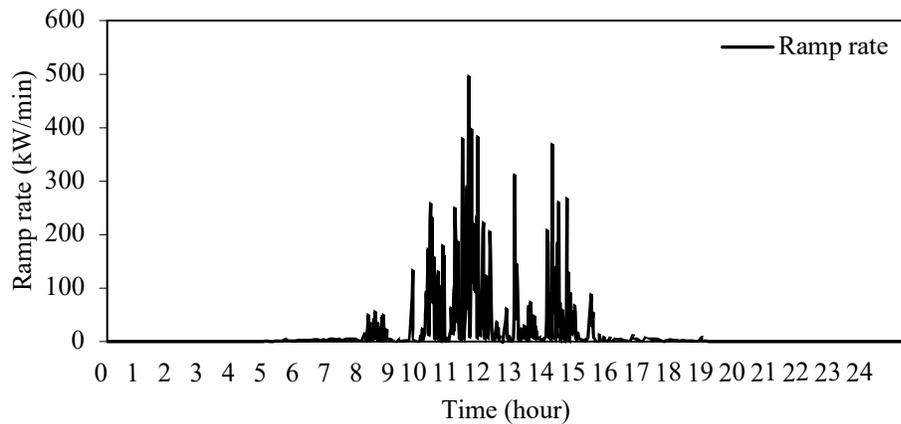
(b)

Figure 6.1 One-minute resolution solar (a) generation. (b) Ramp rate.

6.2 Case Study

Figure 6.2 depicts the total number of the ramps that exceeded the ramp rate limit. A total of 49 ramps is experienced when the ramp rate limit is set to 10%/min of the installed capacity where 37% of these ramps have occurred between 11:30 AM and 12:30 PM. If the ramp rate limit is set to 20%/min of the total installed capacity, the total number of ramps is reduced to 20. After 40%/min ramp rate limit, the number of

violations is zero. Figure 6.3 exhibits actual solar generation and ramp rate along with the

mitigated solar generation and ramp rate for the time between 11:30 and 12:30.
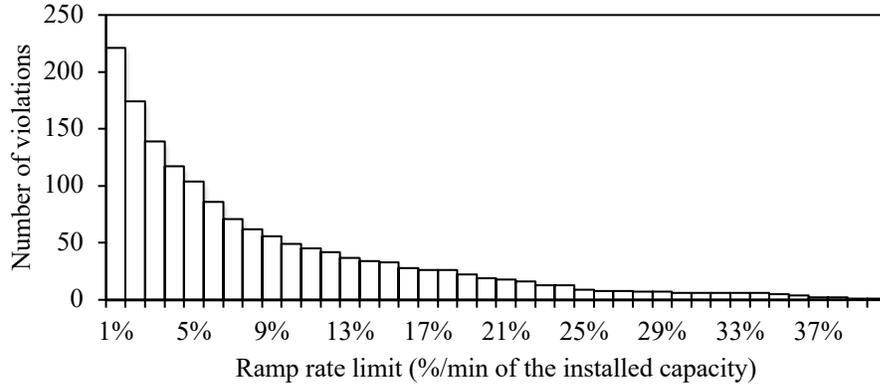
Figure 6.2 The number of ramp rate volitions as ramp rate limit increases.

In order to investigate the impact of solar power ramp rate mitigation, the IEEE

33-bus test system shown in Figure 6.4 is used to run the power flow for a selected hour

(11:30am to 12:30pm) and calculate the voltage changes at each bus. A 1.04 MW solar

PV is connected to bus 18. The load is considered to be constant at this hour, with a value

of 5.6 MW. It is assumed that solar has no voltage control. The following cases are

studied:

Case 1: Voltage deviation at each bus without solar generation ramp rate

   mitigation.

Case 2: Voltage deviation at each bus with solar generation ramp rate mitigation.
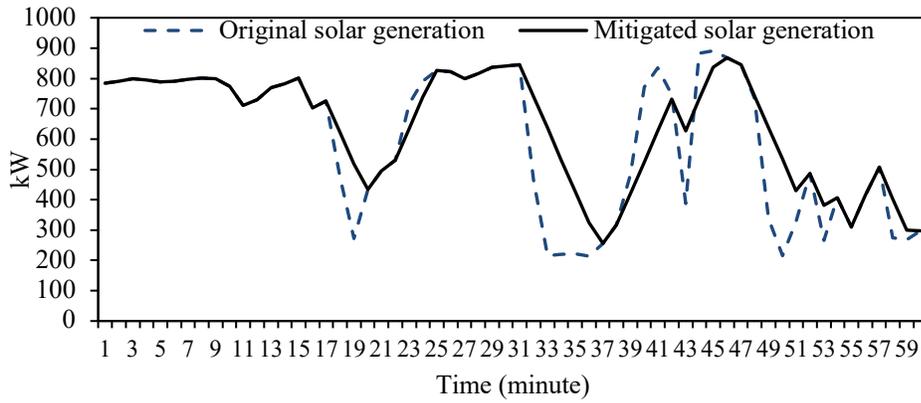
(a)
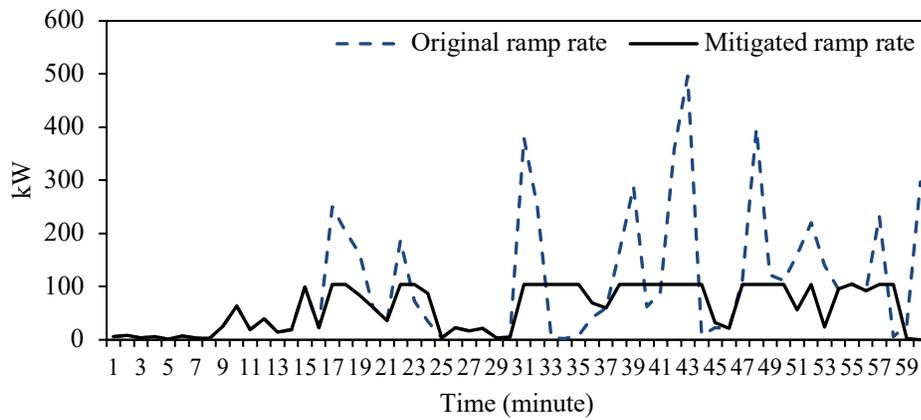
(b)

Figure 6.3 Original versus mitigated (a) solar generation. (b) solar ramp rate
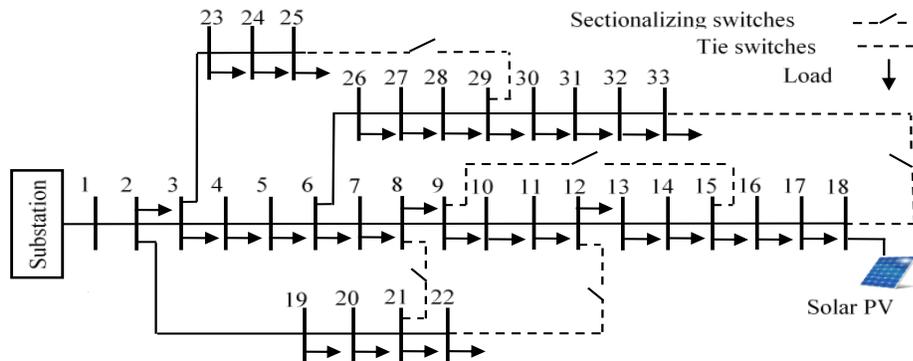
Figure 6.4 IEEE 33-bus test system with PV connected to bus 18

The solar generation is adjusted as in Figure 6(a) to mitigate the solar generation ramp rate. It is assumed that ramp rate, with negative and positive slopes, is corrected by

charging and discharging the energy storage to reduce the ramp rate to a maximum of 104 kW/min. The energy storage sizing requirement is ignored in this thesis assuming that the existing energy storage is optimally sized and placed to limit and capture the solar ramp rate of larger than 104 kW/min. In [87] the solar gneration ramp rate is mitigated by controling the voltage using a capcitor. Other techinques can be used to mitigate the solar generation ramp rate, which are not discussed in this thesis.

Case 1: Voltage deviation at each bus without solar generation ramp rate mitigation

Figure 6.1(b) depicts the ramp rate for a selected day. The highest ramp rate occurs between 11:30am and 12:30pm where the ramp rate has reached 496 kW/min. To run the simulation, the solar generation data for the hour in Figure 6.2(a) is selected as solar generation connected to bus 18. The solar generation is considered in this study with unity power factor so no reactive power is generated by the solar PV. In this hour, 18 violations to ramp rate limit, which is 104 kW/min, have been recorded. The power flow is simulated under this case without mitigating the ramp rate. Table 6.1 depicts the standard deviation for each bus over the hour and the highest deviation is on bus 18 where the PV is connected. In the simulation, bus one is considered an infinite bus with the bus voltage fixed to 1 pu. All the buses are still within the voltage limits used in the power flow. Figure 6.5 illustrates the changes in the voltage at bus 18 for each minute. It is clear that the trend for changes in voltage at bus 18 follows the ramp rate trend.

Table 6.1 Standard Deviation for the voltage at each bus – Case 1

Bus#	Standard deviation	Bus#	Standard deviation	Bus#	Standard deviation	Bus#	Standard deviation
2	0.000159	10	0.008644	18	0.018148	26	0.003724
3	0.001008	11	0.008964	19	0.000159	27	0.003738
4	0.001635	12	0.009573	20	0.000160	28	0.003800
5	0.002286	13	0.011965	21	0.000160	29	0.003846
6	0.003713	14	0.012846	22	0.000160	30	0.003865
7	0.004048	15	0.013790	23	0.001013	31	0.003888
8	0.005222	16	0.014973	24	0.001023	32	0.003893
9	0.006924	17	0.017006	25	0.001027	33	0.003895

Case 2: Voltage deviation at each bus with solar generation ramp rate mitigation

Under this case, the solar generation ramp rate is mitigated as shown in Figure 6.3(b) and limited to 104 kW/min. The ramp rate mitigation reduces the standard deviation in the voltage at each bus compared to case 1. Bus 18 still has the highest standard deviation, but it is reduced by almost 23% compared to Case 1. The buses closest to PV have experienced a bigger reduction in the standard deviation compared to the ones that are further from PV. Results further indicate that standard deviation increases gradually as buses get closer to bus 18.
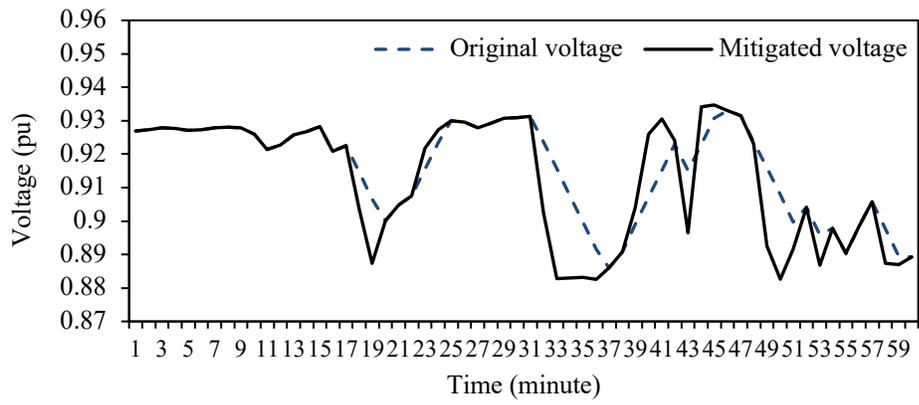
Figure 6.5 Voltage at bus 18 with/without ramp rate mitigation.

Chapter Seven: Conclusion and Future Work

7.1 Conclusion

The shift from conventional generation to renewable energy resources in an effort to reduce emissions has led to a rapid proliferation of renewable resources especially solar photovoltaic (PV) in power systems. Power supply from renewable resources is on a global rise where it is forecasted that renewable generation will surpass other types of generation in a foreseeable future. However, such increase in the renewable resources, mainly solar and wind, exposes the power grid to more vulnerabilities due to the variability and uncertainty associated with these resources. In this dissertation, chapters two, three, and four addressed different approaches to enhance the forecasting accuracy and hence reduce the error. Chapters five and six discussed the integration of solar PV to power system.

A novel model for forecasting short-term solar irradiance was proposed in chapter two. The proposed model converts the non-stationary solar time series data to stationary time series data using high order polynomial model for detrending and further validating the stationarity by calculating residuals. The employed polynomial fitting model had an order of 4 to 6. The proposed model achieved a MAPE under 1%, which was a promising result. The high accuracy of the model would allow system operators to detect any sudden changes in solar output and prepare for proper actions. The FFNN was used to

perform the forecasting on the stationary data. This work was extended to include hybrid forecasting methods that could potentially outperform single methods. The hybrid model can detect the linear and non-linear components in solar time series and overcome any deficiency for individual models. A solar forecasting model based on decomposed linear and nonlinear statistical methods was proposed in chapter three. The proposed model benefited from NARNN in Stage 1 forecasting and ARMAX in Stage 2 forecasting combined with a carefully developed data processing approach. The model was simulated to forecast three days under different weather conditions of sunny, partly cloudy, and cloudy. The maximum resultant NRMSE was obtained as 0.1, for a partly cloudy day, which shows the acceptable performance of the proposed model. To exhibit the effectiveness of the two-stage model, three cases were further studied, comparing the two-stage model with a single-stage model, which clearly demonstrated improvements in NRMSE. The importance of the data stationarity in improving forecasting accuracy was moreover investigated. Furthermore, to include different level of solar measurements a day-ahead solar PV generation forecast model based on multi-level measurements was proposed in chapter four. The proposed model demonstrated an improvement in forecasting accuracy by reducing the MAPE from 14% to 47% for various weather conditions, compared to the case when only single-level measurements were included. It was further seen that the data preprocessing was an important step to ensure the quality of the data before it was used in the training process. The numerical studies revealed that training the forecasting model without data preprocessing might adversely impact the forecasting accuracy. The proposed preprocessing model could potentially reduce the

MAPE by 34% to 65%. It was further shown that the three-level measurements help achieve a better forecasting accuracy compared to two-level measurements.

Improving the forecasting accuracy of solar PV generation would boost the integration of solar PV. Such increase in solar PV integration mandates an upgrade in the power grid to safely accommodate high solar PV penetration. A co-optimization generation and transmission planning model to maximize large-scale solar PV integration was proposed in chapter five. The Benders decomposition method was used to tackle the computational complexity of the model. The proposed model was analyzed through numerical simulations on a small-scale six-bus system as well as relatively large 118-bus test system. The obtained results exhibited that maximizing the large-scale solar PV hosting capacity necessitates a system upgrade. New transmission lines and generation units were to be installed to ensure system flexibility. With proper investments in system upgrade, the studied test system could accommodate up to 40% solar PV by the end of the planning horizon. It was further concluded that reinforcing the system with only transmission lines upgrade would decrease the solar PV penetration to 34%. Moreover, the solar PV penetration would decrease by 14% when the system is only reinforced with generation units upgrade, advocating that a co-optimization planning is much more effective than individual upgrades of generation and transmission. The results further advocated that solar PV energy curtailment is an inherent part of the large-scale solar PV proliferation. This energy curtailment is mainly caused by the lack of adequate system capacity, either in generation or transmission, to fully support solar PV generation, as well as potential overgeneration at times of low load. To study the impact of solar

integration in the operation level, the effect of the solar generation ramp rate was investigated in the IEEE 33-bus test system. The power flow was simulated under two different cases. First, the solar ramp rate was not mitigated while in the second case the solar generation ramp rate was mitigated to limit the ramp rate<10%/min of the installed capacity. The solar generation between 11:30am and 12:30pm had the highest ramp rate. A total of 18 violations to ramp rate limit have occurred at this hour. The ramp rate mitigation reduced the standard deviation for bus 18 where the PV is connected by almost 23%. In the case of higher solar penetration, the solar ramp rate may cause a negative impact on the system voltage. Mitigating the solar generation ramp rate reduces any deviation caused in the system voltage, especially at the bus where the PV is connected.

7.2 Future Work

This dissertation focuses on improving the forecasting for solar PV generation and investigating the integration of solar PV to power system. Different forecasting approaches presented here along with demonstrations of their merits and efficacy. The proposed forecasting models can be further enhanced by including multiple meteorological parameters such as cloud cover, solar irradiance, and temperature along with three-level measurements as inputs to forecasting tool.

Also, the co-optimization generation and transmission planning model can be extended to include BESS in order to reduce the solar PV energy curtailment and also supply the load at night time. The model can be applied to high resolution solar PV data such as 5-min interval data in order to capture solar PV variability. Cost saving due to the reduction of the carbon dioxide CO_2 can be further incorporated in the proposed model.

References

[1] S. Watetakarn and S. Premrudeepreechacharn, "Forecasting of solar irradiance for solar power plants by artificial neural network," in *Smart Grid Technologies-Asia (ISGT ASIA), 2015 IEEE Innovative*, 2015, pp. 1–5.

[2] A. Perea, C. Honeyman, C. Smith, S. Rumery, and A. Holm, "⌘U.S. Solar Market Inisight Executive Summary Q2 2018," GTM Research and the Solar Energy Industries Association, Jun. 2018.

[3] E. Ela, V. Diakov, E. Ibanez, and M. Heaney, "Impacts of variability and uncertainty in solar photovoltaic generation at multiple timescales," *Contract*, vol. 303, pp. 275–3000, 2013.

[4] "IMPACTS OF SOLAR INVESTMENT TAX CREDIT EXTENSION." SEIA, 18-Dec-2015.

[5] V. P. Singh, K. Vaibhav, and D. K. Chaturvedi, "Solar power forecasting modeling using soft computing approach," in *Engineering (NUiCONE), 2012 Nirma University International Conference on*, 2012, pp. 1–5.

[6] "The importance of wind forecasting - Renewable Energy Focus." [Online]. Available: http://www.renewableenergyfocus.com/view/1379/the-importance-of-wind-forecasting/. [Accessed: 04-Sep-2014].

[7] Y. Zhang, M. Beaudin, R. Taheri, H. Zareipour, and D. Wood, "Day-Ahead Power Output Forecasting for Small-Scale Solar Photovoltaic Electricity Generators," *IEEE Trans. Smart Grid*, pp. 1–1, 2015.

[8] H. M. Diagne, M. David, P. Lauret, and J. Boland, "Solar irradiation forecasting: state-of-the-art and proposition for future developments for small-scale insular grids," American Solar Energy Society, 2012.

[9] S. Cros, O. Liandrat, N. Sébastien, and N. Schmutz, "Extracting cloud motion vectors from satellite images for solar power forecasting," in *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*, 2014, pp. 4123–4126.

[10] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," *Renew. Sustain. Energy Rev.*, vol. 27, pp. 65–76, Nov. 2013.

[11] A. Hammer, D. Heinemann, E. Lorenz, and B. Lückehe, "Short-term forecasting of solar radiation: a statistical approach using satellite data," *Sol. Energy*, vol. 67, no. 1, pp. 139–150, 1999.

[12] R. Marquez, H. T. C. Pedro, and C. F. M. Coimbra, "Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs," *Sol. Energy*, vol. 92, pp. 176–188, Jun. 2013.

[13] A. Tuohy *et al.*, "Solar Forecasting: Methods, Challenges, and Performance," *IEEE Power Energy Mag.*, vol. 13, no. 6, pp. 50–59, Nov. 2015.

[14] J. Kleissl, *Solar Energy Forecasting and Resource Assessment*. Academic Press, 2013.

[15] J. Remund, R. Perez, and E. Lorenz, "Comparison of solar radiation forecasts for the USA," in *Proc. of the 23rd European PV Conference*, 2008, pp. 1–9.

[16] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renew. Energy*, vol. 37, no. 1, pp. 1–8, Jan. 2012.

[17] R. Huang, T. Huang, R. Gadh, and N. Li, "Solar Generation Prediction using the ARMA Model in a Laboratory-level Micro-grid," in *Smart Grid Communications (SmartGridComm), 2012 IEEE Third International Conference on*, 2012, pp. 528–533.

[18] G. P. Nason, "Stationary and non-stationary times series," *Stat. Volcanol. Spec. Publ. IAVCEI*, vol. 1, pp. 000–000, 2006.

[19] A. Mellit and S. A. Kalogirou, "Artificial intelligence techniques for photovoltaic applications: A review," *Prog. Energy Combust. Sci.*, vol. 34, no. 5, pp. 574–632, Oct. 2008.

[20] J. M. Filipe, R. J. Bessa, J. Sumaili, R. Tome, and J. N. Sousa, "A hybrid short-term solar power forecasting tool," in *Intelligent System Application to Power Systems (ISAP), 2015 18th International Conference on*, 2015, pp. 1–6.

[21] M. Tucci, E. Crisostomi, G. Giunta, and M. Raugi, "A Multi-Objective Method for Short-Term Load Forecasting in European Countries," *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3537–3547, Sep. 2016.

[22] A. Mellit, M. Benghanem, and S. A. Kalogirou, "An adaptive wavelet-network model for forecasting daily total solar-radiation," *Appl. Energy*, vol. 83, no. 7, pp. 705–722, Jul. 2006.

[23] H. T. C. Pedro and C. F. M. Coimbra, "Assessment of forecasting techniques for solar power production with no exogenous inputs," *Sol. Energy*, vol. 86, no. 7, pp. 2017–2028, Jul. 2012.

[24] C. Paoli, C. Voyant, M. Muselli, and M.-L. Nivet, "Forecasting of preprocessed daily solar radiation time series using neural networks," *Sol. Energy*, vol. 84, no. 12, pp. 2146–2160, Dec. 2010.

[25] D. Yang, P. Jirutitijaroen, and W. M. Walsh, "Hourly solar irradiance time series forecasting using cloud cover index," *Sol. Energy*, vol. 86, no. 12, pp. 3531–3543, Dec. 2012.

[26] R. Marquez and C. F. M. Coimbra, "Intra-hour DNI forecasting based on cloud tracking image analysis," *Sol. Energy*, vol. 91, pp. 327–336, May 2013.

[27] C. W. Chow *et al.*, "Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed," *Sol. Energy*, vol. 85, no. 11, pp. 2881–2893, Nov. 2011.

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

[29] C. Chen, S. Duan, T. Cai, and B. Liu, "Online 24-h solar power forecasting based on weather type classification using artificial neural network," *Sol. Energy*, vol. 85, no. 11, pp. 2856–2870, Nov. 2011.

[30] P. Bacher, H. Madsen, and H. A. Nielsen, "Online short-term solar power forecasting," *Sol. Energy*, vol. 83, no. 10, pp. 1772–1783, Oct. 2009.

[31] S. Chai, Z. Xu, and W. K. Wong, "Optimal Granule-Based PIs Construction for Solar Irradiance Forecast," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 3332–3333, Jul. 2016.

[32] R. H. Inman, H. T. C. Pedro, and C. F. M. Coimbra, "Solar forecasting methods for renewable energy integration," *Prog. Energy Combust. Sci.*, vol. 39, no. 6, pp. 535–576, Dec. 2013.

[33] H. S. Jang, K. Y. Bae, H.-S. Park, and D. K. Sung, "Solar Power Prediction Based on Satellite Images and Support Vector Machine," *IEEE Trans. Sustain. Energy*, vol. 7, no. 3, pp. 1255–1263, Jul. 2016.

[34] D. Dickey and W. Fuller, "Likelihood ratio statistics for autore- gressive time series with a unit root," 1981, vol. Econometrica V. 49 No. 4, pp. 1057–1072.

[35] M. Alanazi, A. Alanazi, and A. Khodaei, "Long-Term Solar Generation Forecasting," presented at the 2016 IEEE/PES Transmission and Distribution Conference and Exposition (T&D), Dallas,TX, 2016, pp. 1–5.

[36] M. Alanazi, M. Mahoor, and A. Khodaei, "Two-stage hybrid day-ahead solar forecasting," *ArXiv Prepr. ArXiv170608699*, 2017.

[37] M. Alanazi, M. Mahoor, and A. Khodaei, "Day-Ahead Solar Forecasting Based on Multi-Level Solar Measurements," in *2018 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*, Denver, CO, USA, 2018, pp. 1–9.

[38] R. Shah, N. Mithulananthan, R. C. Bansal, and V. K. Ramachandaramurthy, "A review of key power system stability challenges for large-scale PV integration," *Renew. Sustain. Energy Rev.*, vol. 41, pp. 1423–1436, Jan. 2015.

[39] S. Achilles, S. Schramm, and J. Bebic, "Transmission System Performance Analysis for High-Penetration Photovoltaics," NREL/SR-581-42300, 924641, Feb. 2008.

[40] S. Eftekharnejad, V. Vittal, Heydt, B. Keel, and J. Loehr, "Impact of increased penetration of photovoltaic generation on power systems," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 893–901, May 2013.

[41] B. Tamimi, C. Canizares, and K. Bhattacharya, "System Stability Impact of Large-Scale and Distributed Solar Photovoltaic Generation: The Case of Ontario, Canada," *IEEE Trans. Sustain. Energy*, vol. 4, no. 3, pp. 680–688, Jul. 2013.

[42] D. Lew, N. Miller, K. Clark, G. Jordan, and Z. Gao, "Impact of high solar penetration in the western interconnection," *Contract*, vol. 303, pp. 275–3000, 2010.

[43] A. Bloom *et al.*, "Eastern renewable generation integration study," *Natl. Renew. Energy Lab. Gold. CO Tech Rep NRELTP-6A20-64472*, 2016.

[44] G. Barbose, "U.S. Renewable Portfolio Standards '2018 Annual Status Report,'" Lawrence Berkeley National Laboratory, Nov. 2018.

[45] "Investigating a Higher Renewables Portfolio Standard in California," Energy and Environmental Economics, Inc, Executive Summary, Jan. 2014.

[46] Z. S. Hosseini, A. Khodaei, E. A. Paaso, M. S. Hossan, and D. Lelic, "Dynamic Solar Hosting Capacity Calculations in Microgrids," presented at the CIGRE US National Committee 2018 Grid of the Futur, p. 12.

[47] D. Q. Hung, N. Mithulananthan, and R. C. Bansal, "Analytical strategies for renewable distributed generation integration considering energy loss minimization," *Appl. Energy*, vol. 105, pp. 75–85, May 2013.

[48] P. S. Georgilakis and N. D. Hatziargyriou, "Optimal Distributed Generation Placement in Power Distribution Networks: Models, Methods, and Future Research," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3420–3428, Aug. 2013.

[49] L. Mokgonyana, J. Zhang, H. Li, and Y. Hu, "Optimal location and capacity planning for distributed generation with independent power production and self-generation," *Appl. Energy*, vol. 188, pp. 140–150, Feb. 2017.

[50] F. Capitanescu, L. F. Ochoa, H. Margossian, and N. D. Hatziargyriou, "Assessing the Potential of Network Reconfiguration to Improve Distributed Generation Hosting Capacity in Active Distribution Systems," *IEEE Trans. Power Syst.*, vol. 30, no. 1, pp. 346–356, Jan. 2015.

[51] J. Quintero, H. Zhang, Y. Chakhchoukh, V. Vittal, and G. T. Heydt, "Next Generation Transmission Expansion Planning Framework: Models, Tools, and Educational Opportunities," *IEEE Trans. Power Syst.*, vol. 29, no. 4, pp. 1911–1918, Jul. 2014.

[52] M. Marinelli, P. Maule, A. N. Hahmann, O. Gehrke, P. B. Norgard, and N. A. Cutululis, "Wind and Photovoltaic Large-Scale Regional Models for Hourly

Production Evaluation," *IEEE Trans. Sustain. Energy*, vol. 6, no. 3, pp. 916–923, Jul. 2015.

[53] F. D. Munoz and A. D. Mills, "Endogenous Assessment of the Capacity Value of Solar PV in Generation Investment Planning Studies," *IEEE Trans. Sustain. Energy*, vol. 6, no. 4, pp. 1574–1585, Oct. 2015.

[54] H. Shaker, H. Zareipour, and D. Wood, "Impacts of large-scale wind and solar power integration on California′s net electrical load," *Renew. Sustain. Energy Rev.*, vol. 58, pp. 761–774, May 2016.

[55] I. J. Perez-Arriaga, "The Transmission of the Future: The Impact of Distributed Energy Resources on the Network," *IEEE Power Energy Mag.*, vol. 14, no. 4, pp. 41–53, Jul. 2016.

[56] W. Muneer, K. Bhattacharya, and C. A. Canizares, "Large-Scale Solar PV Investment Models, Tools, and Analysis: The Ontario Case," *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 2547–2555, Nov. 2011.

[57] K. Rajesh, A. Bhuvanesh, S. Kannan, and C. Thangaraj, "Least cost generation expansion planning with solar power plant using Differential Evolution algorithm," *Renew. Energy*, vol. 85, pp. 677–686, Jan. 2016.

[58] P. Vithayasrichareon and I. F. MacGill, "Valuing large-scale solar photovoltaics in future electricity generation portfolios and its implications for energy and climate policies," *IET Renew. Power Gener.*, vol. 10, no. 1, pp. 79–87, Jan. 2016.

[59] "Solar Data 1991-2010 Site #725650." [Online]. Available: http://rredc.nrel.gov/solar/old_data/nsrdb/1991-2010/hourly/siteonthefly.cgi?id=725650. [Accessed: 16-Sep-2014].

[60] "Denver International Airport, CO, United States Of America Sunrise and Sunset times - weather.co.uk." [Online]. Available: http://uk.weather.com/climate/sunRiseSunSet-Denver-International-Airport-CO-USCO0425?month=1. [Accessed: 16-Sep-2014].

[61] W. Ji and K. C. Chee, "Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN," *Sol. Energy*, vol. 85, no. 5, pp. 808–817, May 2011.

[62] K. Benmouiza and A. Cheknane, "Small-scale solar radiation forecasting using ARMA and nonlinear autoregressive neural network models," *Theor. Appl. Climatol.*, vol. 124, no. 3–4, pp. 945–958, May 2016.

[63] F. H. Al-Sadah, F. M. Ragab, and M. K. Arshad, "Hourly solar radiation over Bahrain," *Energy*, vol. 15, no. 5, pp. 395–402, 1990.

[64] M. Alanazi and A. Khodaei, "Day-ahead Solar Forecasting Using Time Series Stationarization and Feed-Forward Neural Network," presented at the North American Power Symposium (NAPS), 2016, Denver, 2016, pp. 1–6.

[65] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time series analysis: forecasting and control*. Hoboken, N.J.: John Wiley, 2008.

[66] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, 1974.

[67] "System Advisor Model (SAM) |." [Online]. Available: https://sam.nrel.gov/. [Accessed: 10-Aug-2017].

[68] S. Pfenninger and I. Staffell, "Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data," *Energy*, vol. 114, pp. 1251–1265, Nov. 2016.

[69] A. Khodaei, S. Bahramirad, and M. Shahidehpour, "Microgrid Planning Under Uncertainty," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2417–2425, Sep. 2015.

[70] A. Khodaei and M. Shahidehpour, "Microgrid-Based Co-Optimization of Generation and Transmission Planning in Power Systems," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1582–1590, May 2013.

[71] A. Khodaei, "Provisional Microgrid Planning," *IEEE Trans. Smart Grid*, vol. 8, no. 3, pp. 1096–1104, May 2017.

[72] A. Khodaei, M. Shahidehpour, and S. Kamalinia, "Transmission Switching in Expansion Planning," *IEEE Trans. Power Syst.*, vol. 25, no. 3, pp. 1722–1733, Aug. 2010.

[73] S. H. Madaeni, R. Sioshansi, and P. Denholm, "Comparison of Capacity Value Methods for Photovoltaics in the Western United States," NREL/TP-6A20-54704, 1046871, Jul. 2012.

[74] "EIA" U.S. Energy Information Administration, "Electric Power Monthly with data for November 2018," p. 274, 2019.

[75] L. Bird, J. Cochran, and X. Wang, "Wind and Solar Energy Curtailment: Experience and Practices in the United States," NREL/TP--6A20-60983, 1126842, Mar. 2014.

[76] W. Lasher, "The Competitive Renewable Energy Zones Process," p. 12, 2014.

[77] A. Alanazi and A. Khodaei, "Optimal battery energy storage sizing for reducing wind generation curtailment," presented at the Power & Energy Society General Meeting 2017 IEE, 2017, pp. 1–5.

[78] "IIT data." [Online]. Available: http://motor.ece.iit.edu/data/.

[79] L. Hirth, "The market value of variable renewables," *Energy Econ.*, vol. 38, pp. 218–236, Jul. 2013.

[80] F. Ueckerdt, R. Brecha, and G. Luderer, "Analyzing major challenges of wind and solar variability in power systems," *Renew. Energy*, vol. 81, pp. 1–10, Sep. 2015.

[81] A. Mills and R. Wiser, "Strategies for Mitigating the Reduction in Economic Value of Variable Generation with Increasing Penetration Levels," 2014.

[82] J. Schnabel and S. Valkealahti, "Energy Storage Requirements for PV Power Ramp Rate Control in Northern Europe," *Int. J. Photoenergy*, vol. 2016, pp. 1–11, 2016.

[83] J. Johnson, B. Schenkman, A. Ellis, J. Quiroz, and C. Lenox, "Initial operating experience of the la ola 1.2-MW photovoltaic system," *Sandia Natl. Lab. Tech. Rep. SAND2011-8848*, 2011.

[84] J. Traube *et al.*, "Mitigation of Solar Irradiance Intermittency in Photovoltaic Power Systems With Integrated Electric-Vehicle Charging Functionality," *IEEE Trans. Power Electron.*, vol. 28, no. 6, pp. 3058–3067, Jun. 2013.

[85] M. J. Reno, M. Lave, J. E. Quiroz, and R. J. Broderick, "PV ramp rate smoothing using energy storage to mitigate increased voltage regulator tapping," in *Photovoltaic Specialists Conference (PVSC), 2016 IEEE 43rd*, 2016, pp. 2015–2020.

[86] V. Gevorgian and S. Booth, *Review of PREPA technical requirements for interconnecting wind and solar generation*. National Renewable Energy Laboratory, 2013.

[87] N. Kakimoto, H. Satoh, S. Takayama, and K. Nakamura, "Ramp-Rate Control of Photovoltaic Generator With Electric Double-Layer Capacitor," *IEEE Trans. Energy Convers.*, vol. 24, no. 2, pp. 465–473, Jun. 2009.

List of Publications

- **M. Alanazi**, A. Alanazi, and A. Khodaei, "Long-Term Solar Generation Forecasting," IEEE PES Transmission and Distribution Conference, Dallas, TX, 2016.

- A. Alanazi, **M. Alanazi**, and A. Khodaei, "Managing the Microgrid Net Load Variability, IEEE PES Transmission and Distribution Conference, Dallas, TX, 2016.

- **M. Alanazi**, and A. Khodaei, "Day-Ahead Solar Forecasting Using Time Series Stationarization and Feed-Forward Neural Network," IEEE North American Power Symposium, Denver, CO, September 2016.

- **M. Alanazi**, A. Khodaei, S. Bahramirad, and A. Paaso, "Investigating the Impact of Time Series Stationarization on Day-Ahead Solar Forecasting," CIGRE Grid of the Future Symposium, Philadelphia, PA, October 2016.

- **M. Alanazi**, M. Mahoor and A. Khodaei, "Multilevel Day-Ahead Solar Forecasting [abstract]," IEEE PES General Meeting, Chicago, IL, Jul. 2017.

- **M. Alanazi**, M. Mahoor and A. Khodaei, "Two-Stage Hybrid Day-Ahead Solar Forecasting," IEEE North American Power Symposium, Morgantown, WV, Sep. 2017.

- **M. Alanazi**, A. Khodaei, "Investigating the Voltage Fluctuation Caused by Solar PV Generation Variability in Distribution Grids," CIGRE 2017 Grid of the Future Symposium, Cleveland, OH, October 2017.

- **M. Alanazi**, M. Mahoor, and A. Khodaei, "Day-ahead Solar Forecasting Based on Multi-level Solar Measurements," IEEE PES Transmission and Distribution Conference, Denver, CO, 2018.

- **M. Alanazi**, M. Mahoor, and A. Khodaei, "Co-Optimization Generation and Transmission Planning for Maximizing Large-Scale Solar PV Integration," Int. J. Electr. Power Energy Syst., 2019.