Impact of Distributed Generation of Solar Photovoltaic (PV) Generation on the Massachusetts Transmission System

by
Arvind Simhadri

B.S.E., University of Calgary, Calgary, Canada, 2006

Submitted to the MIT Sloan School of Management and the Engineering Systems Division in partial fulfillment of the requirements for the degrees of
Master of Science in Systems Engineering
and
Master of Business Administration
in conjunction with the Leaders for Global Operations Program at the MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2015

© Arvind Simhadri, MMXV. All rights reserved.
The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

Signature redacted

the MIT Sloan School of Management and the Engineering Systems Division

Signature redacted

May 8, 2015

Certified by .........................

James L. Kirtley Jr., Thesis Supervisor
Professor of Electrical Engineering

Certified by .........................

Georgia Perakis, Thesis Supervisor
William F. Pounds Professor of Management Science, MIT Sloan School of Management

Approved by .........................

Munther A. Dahleh
William A. Coolidge Professor of Electrical Engineering and Computer Science
Chair, ESD Education Committee

Approved by .........................

Maura Herson
Director, MBA Program, MIT Sloan School of Management
Impact of Distributed Generation of Solar Photovoltaic (PV) Generation on the Massachusetts Transmission System

by

Arvind Simhadri

Submitted to the MIT Sloan School of Management and the Engineering Systems Division on May 8, 2015, in partial fulfillment of the requirements for the degrees of Master of Science in Systems Engineering and Master of Business Administration

Abstract

After reaching 250 megawatt direct current (MW dc) of solar photovoltaic (PV) generation installed in Massachusetts (MA) in 2013, four years ahead of schedule, Governor Deval Patrick in May of 2013 announced an increase in the MA solar PV goal to 1,600 MW by 2020 ([13]). However, integration of such significant quantities of solar PV into the electric power system is potentially going to require changes to the transmission system planning and operations to ensure continued reliability of operation ([14]). The objective of this project is to predict the distribution of solar PV in MA and to develop a simulation framework to analyze the impact of solar generation on the electric power system.

To accomplish this objective, we first developed a prediction model for solar PV aggregate and spatial long term distribution. We collected solar PV installation data and electricity consumption data for 2004 to 2014 for each ZIP code in MA. Additional information such as population, land availability, average solar radiance, number of households, and other demographic data per ZIP code was also added to improve the accuracy of the model. For example, ZIP codes with higher solar radiance are more likely to have solar PV installations. By utilizing machine learning methods, we developed a model that incorporates demographic factors and applies a logistic growth model to forecast the capacity of solar PV generation per ZIP code.

Next we developed an electrically equivalent model to represent the predicted addition of solar PV on the transmission system. Using this model, we analyzed the impact of solar PV installations on steady-state voltage of the interconnected electric transmission system. We used Siemens PTI’s PSS/E software for transmission network modeling and analysis. Additionally, we conducted a sensitivity analysis on scenarios such as peak and light electricity consumption period, different locations of solar PV, and voltage control methods to identify potential reliability concerns. Furthermore, we tested the system reliability in the event of outages of key transmission lines, using N-1 contingency analysis.

The analysis identified that the voltage deviation on transmission system because of adding 1,600 MW dc of distributed solar PV is within +/- 5% range. Based on the analysis
performed in this thesis, we conclude that the current MA transmission system can operate reliably after the addition of the expected 1,600 MW dc of solar PV.

As National Grid acquires information on solar installations, new data will improve the ability and accuracy of the prediction model to predict solar PV capacity and location more accurately. The simulation framework developed in this thesis can be utilized to rerun the analysis to test the robustness of the electric transmission system at a future date.

Thesis Supervisor: James L. Kirtley Jr.
Title: Professor of Electrical Engineering

Thesis Supervisor: Georgia Perakis
Title: William F. Pounds Professor of Management Science, MIT Sloan School of Management
Acknowledgments

First and foremost I would like to thank my advisors Prof. Georgia Perakis and Prof. James Kirtley for their invaluable insight, guidance, and support on this project.

In addition to my advisors, the work on this paper was done with extensive collaboration with Ludovica Rizzo, Masters of Science student in the Operations Research Department, and the National Grid team including Janny Dong, Per-Anders Lof, and Jack Martin. I want to thank all of them for their continued inputs to the project and guidance throughout the project.

I would also like to thank my fellow LGO and Sloan classmates for making my graduate experience at MIT and Sloan one of the most memorable experiences.

Finally, I would like to thank my friends, family, and wife, Pragati Chopra, for their encouragement, love, and support during my two years at MIT.
Contents

1 Introduction to Impact of Solar PV Generation on the Electric Transmission System 13
   1.1 Project Motivation and Scope .................................................. 14
   1.2 Literature Review ........................................................................ 15
       1.2.1 Solar PV Distributed Generation Forecasting .......................... 15
       1.2.2 Impact of DG on the Electric Transmission System ................. 16
   1.3 Thesis Organization and Contribution ........................................... 17

2 Equivalencing the Distributed Generation and Distribution System 19
   2.1 System Description and Assumptions ........................................... 22
       2.1.1 System Assumptions .............................................................. 24
   2.2 Equivalent Model for Distributed Generation ............................... 26
   2.3 Derivation of Equations to Equivalence the Distributed Generation Network 28
       2.3.1 Equivalencing of the Secondary Distribution Network ............... 29
       2.3.2 Equivalencing the Primary Feeder Branch Network .................... 31
   2.4 Conclusion ................................................................................. 34

3 Predictive Model for Solar PV Location 36
   3.1 Historical Data and Trends ............................................................ 36
   3.2 Rooftops ..................................................................................... 39
       3.2.1 Data .................................................................................... 40
       3.2.2 Model .................................................................................. 42
       3.2.3 Estimation ............................................................................ 46
List of Figures

1-1 Solar PV Generation Growth in US ([21]) .................................................. 14

2-1 Simplified representation of Power System Network ([16]) .................... 20
2-2 Simplified representation of Power System with Distributed Generation ... 20
2-3 Simplified representation of Distribution Network ................................. 21
2-4 Simplified representation of Solar PV Rooftop Panel Interconnection to Grid 23
2-5 Simplified representation of Solar PV Farm Interconnection to Grid .......... 24
2-6 Equivalencing of the Distributed Generation Network ............................. 28
2-7 (a) Secondary Distribution Branch - Electrical Representation (b) Equivalent Model .......................................................... 29
2-8 (a) Distribution Feeder Branch - Electrical Representation (b) Equivalent Model .......................................................... 31

3-1 Number of installations per quarter in MA ........................................... 37
3-2 Evolution of Penetration Level for 20 ZIP codes .................................. 37
3-3 Capacity Installed per ZIP code in 2013 .......................................... 38
3-4 Distribution of capacity for rooftops ..................................................... 40
3-5 Number of installations per ZIP code .................................................. 40
3-6 Number of rooftop installations per ZIP code in 2013 ........................... 41
3-7 Load zones for rooftop prediction ......................................................... 42
3-8 Shape of logistic growth ................................................................. 43
3-9 Number of installations and prediction for 01824. The circles represent the data and the line represents the fitted model ......................... 49
3-10 Number of installations and prediction for 02790. The circles represent the data and the line represents the fitted model.

3-11 Prediction for 2020 for 150 ZIP codes. The colors represent the prediction and the size of the circles represents the installations until 2013.

3-12 Cumulative capacity installed in every load zone.

3-13 Load zones for farm prediction.

3-14 Current installations and fitted model for the 6 load zones. The circles represent the actual cumulative capacity and the red curve represent the fitted model.

3-15 Using Arc GIS Software to Calculate Land Usage Per Town.

3-16 Total solar PV prediction in MA in 2020.

4-1 Example of Distributed Generation Modeling in PSS/E.

4-2 Changes in Transmission Voltage - Summer Peak Case.

4-3 Changes in Transmission Voltage - Light Load Case.

4-4 Changes in Transmission Voltage - Shoulder Peak Case.
List of Tables

2.1 Feeder Length in each Load Zone ............................................................ 25
3.1 Columns of Open PV data set ................................................................. 37
3.2 Performance of the predictive model ......................................................... 48
3.3 Capacity installed before 2014 and predictions for 2020 for the six load zones 54
3.4 Weight Assigned per Land Use Type ......................................................... 55
3.5 Predicted capacity of solar PV farm per town in Load Zone - Boston ........ 56
3.6 Sample mapping of transmission buses in MA .......................................... 57
6.1 Demographic features considered ............................................................ 69
6.2 Parameters estimated from formulation 3.5 ............................................. 70
6.3 Fitted models for farms, C in MW, t in weeks (starting from 2008-1-1) .... 71
6.4 PSS/E Load Flow Analysis Parameters .................................................. 71
6.5 Transformer Impedance Calculation for Rooftop PV and Solar PV farms .. 72
Chapter 1

Introduction to Impact of Solar PV Generation on the Electric Transmission System

Electric power system operators and utilities worldwide have raised concerns about the impact of high-penetration of renewables, like wind and solar generation, on the electric grid reliability ([3], [5], [14] and [1]). In 2011, Germany experienced electric power system stability issues due to the high penetration of Solar photovoltaic (PV) installations. It was estimated that 315,000 solar PV inverters in Germany must be retrofitted at a cost of approximately $300M in order to address the reliability concerns ([6] and [7]).

Stability of the US electric power system is of particular concern here because of the significant growth seen in the solar PV industry in the past decade (see Figure 1-1). Between 2004 and 2014, the US PV generation has increased from 1-2 megawatt (MW) to approximately 16,000 MW hour. Additionally, the annual energy outlook report published by the US Energy Information Administration (EIA) states that solar is the fastest growing source of renewable energy in US with an annual growth rate of 6.8%. Furthermore, the report predicts that 77 GW of renewable generation capacity will be added up until 2040 with 44% (33900 MW) of that from solar generation (including solar thermal and solar PV). Of the 33.9 GW of solar generation, 31000 MW is predicted to be from solar PV installations ([22]).
1.1 Project Motivation and Scope

In Massachusetts (MA), after reaching the previous goal of having 250 MW dc of solar PV generation installed in 2013, four years ahead of schedule, Governor Deval Patrick in May of 2013 announced an increase in MA solar PV goal to 1,600 MW dc by 2020 ([13]). Note that in our analysis we applied a DC-to-AC derate ratio of 83% to the MA goal of 1,600 MW dc. This ratio was applied because PV system designers typically oversize their solar panel array with respect to their inverter by a factor of 1.2 ([12] and [9]).

To promote future solar PV growth, MA has established strong incentives for renewable energy production that have led to significant cost reductions in solar electricity, making clean energy more accessible to MA businesses and residents. The amount of solar energy installed in MA has increased over 80 times from the 3 MW installed in 2007. In 2012, MA became the 6th largest solar PV producer in the United States ([4]). If the new goal is reached, 1,600 MW dc of installed solar will be sufficient to generate enough electricity to power approximately 240,000 homes annually and reduce the greenhouse gas emissions
produced by about 166,000 cars ([30]).

The objective of this project is to research technical issues, as seen from a transmission system perspective, related to such a large increase of solar PV installations. Introducing such significant quantities of generation to the power system could have an impact on the reliability of the MA electric transmission system. In this thesis we develop a simulation framework to analyze the issues related to voltage control and reactive power from the integration of 1,600 MW dc of solar PV generation on the MA electric transmission system.

1.2 Literature Review

1.2.1 Solar PV Distributed Generation Forecasting

Most of the solar PV generation interconnected to the electric power system grid in the past decade has been through Distributed Generation (DG) technology. DG refers to power produced at the point of consumption. DG resources, or distributed energy resources (DER), are small-scale energy resources that typically range in size from 3 kilowatts (kW) to over 10 megawatts (MW) or larger. A typical household’s peak demand is about 3.5 kW, so the smaller resources are used by residential customers, while the larger systems are typically used by commercial and industrial customers. In addition to PV, DERs can include small wind turbines, combined heat and power (CHP), fuel cells, microturbines, and other sources. However, more than 90 percent of installed distributed generation in the United States today is solar ([29]). This type of DG is directly interconnected to the electrical distribution network ([2]). Since DG of solar PV could be located in diverse locations such as residential rooftops, landfills, or golf courses, it makes it difficult for the utility companies to forecast the location and capacity of solar PV penetration. Therefore, it is a challenge for the utility companies to plan ahead for potential system updates that may be required to maintain continued system reliability.

In Massachusetts, the electric power system grid operation is overseen by ISO-NE (Independent System Operator - New England), which is a Regional Transmission Organization (RTO). In December 2014, ISO-NE created a Distributed Generation Forecast Working
Group (DGFWG) to better understand the impact of high penetration of renewables in the New England (NE) region ([9]). This group provides input to NE’s long term DG forecast and examines the challenges and solutions associated with large scale DG integration in NE.

On Feb 27th 2015, the DGFWG group presented a draft of 2015 solar PV forecast for NE ([9]). The forecast from the DGFWG group used a qualitative approach to predict solar PV capacity per state based on factors such as state policy goals, funding, and recent trends in solar PV development in each state. The study assumed that the state policy goal of 1,600 MW dc will be reached by 2020 and the increase in capacity per year is estimated based on the following anticipated policy factors:

1. Planned reduction of federal tax credits in 2016 will promote increased development through 2016

2. Program stabilizes from 2017-2020 until goal of 1,600 MW dc is achieved

Furthermore, the forecast from DGFWG states that a reasonable representation of the geographical locations of existing and future PV installation is required for appropriate modeling. It also lays out a proposed plan for forecasting geographic distribution of solar PV using the available solar PV installation application queue data (this queue is a list of applications submitted to utilities for interconnection of solar PV generation).

Based on our literature review, we believe that the forecast from the DGFWG group is the first attempt to forecast the capacity and location of distributed generation of solar PV in Massachusetts.

1.2.2 Impact of DG on the Electric Transmission System

The need for electric transmission system analysis is being driven by a redistribution of generation on the electric power system grid. To replace conventional generation resources, such as coal and gas, with more efficient and/or cleaner renewable resources like solar PV could require changes to the transmission system to ensure continued reliability.

A joint study conducted by the North American Electric Reliability Corporation (NERC) and California Independent System Operator (CAISO) ([10]) analyzes the reliability of electric power system with high penetration of renewables. The study shows that the voltage
on the electric transmission network is often degraded because of the lack of voltage control ability on the DG technology. To address these issues, there is a need to improve the visibility and forecast for DG resources and to establish a requirement to regulate the aggregated voltage of point of interconnection of DG. ([10]). Furthermore, the study states that modeling smaller generator units with often unknown proprietary control models can manifest into large modeling errors and an incorrect understanding of system behavior. Standard, valid, generic, and non-confidential power-flow (steady-state) models for DG are needed and must be developed to enable accurate system representation.

Another study conducted by the National Renewable Energy Laboratory (NREL) developed a simulation database to allow analysis of system behavior and interactions with high levels of PV penetration under realistic conditions ([11]). Based on the analysis and observation, the study recommended that there should be future research in developing models that are accurate enough to estimate aggregated behavior of solar PV systems for electric system planning.

The literature review above highlights the importance of creating a standard and generic model to simulate the aggregated behavior of solar PV on the electric transmission system. However, to the best of our knowledge, there is currently no study that analyzes the impact of solar PV penetration on the interconnected MA electric transmission system.

### 1.3 Thesis Organization and Contribution

In this thesis we developed a simulation framework to predict the impact of adding 1,600 MW dc of distributed solar PV generation on the existing transmission grid. Currently, all utilities, including National Grid, perform transmission system impact studies for individual solar PV interconnection requests greater than 5 MW. However, as of now there has not been any study that analyzes the transmission system with all the predicted solar PV interconnected to the MA electric power system at the same time. Note that our analysis considers all service territories in MA including the ones outside of National Grid's operating territory.

In Chapter 2, we develop an electrical model to calculate the equivalent impedances
between the source of DG and the distribution bus at each substation. Due to the unavailability of detailed distribution network model in the existing transmission system model, we have proposed several assumptions that allow us to simplify and generalize the distribution network based on few significant parameters such as the distribution feeder length and the impedance of transformers. While models currently exist that attempt to simulate DG interconnections, our solution provides a standard model that can be applied across the MA transmission system.

In Chapter 3, we predict the location and amount of solar PV DG in each ZIP code/town in MA based on the assumption that the 2020 installed capacity in MA would be 1,600 MW dc - the target set by the MA Governor. The total generation predicted is then assigned to a corresponding substation based on proximity. If there is more than one substation in the area, then the total generation amount is evenly split among all the substations. As highlighted in the literature review, even though the DGFWG group has developed a way to predict the solar PV capacity in MA, our model provides a prediction at a new level of granularity that is currently not available. Also, because our methods are completely data driven they can be continuously improved as additional data become available.

In Chapter 4, we combine Chapter 2 and Chapter 3 to model DG resources on the existing transmission system model. Then we test different electric system loading scenarios on this model to predict the behavior and impact of solar PV DG. This model allows National Grid to analyze the impact of 1,600 MW dc of solar generation on the transmission network. Based on this analysis, National Grid can predict areas of the electric transmission network that may have reliability concerns.
Chapter 2

Equivalencing the Distributed Generation and Distribution System

In the mid 20th century, centralized power generating plants have become an integral part of the electric grid. These centralized plants were large generating facilities located close to generation resources, such as waterfall or coal supplies, and generally far from the populated load centers. The power generated was transmitted through the transmission grid at high voltages to a distribution substation located closer to the load center. At the distribution substation the voltage was stepped down to a lower voltage and distributed to the end users through the distribution network. Figure 2-1 shows a typical power system connection from generation to end user.
However, in the recent years there has been a large increase in DG. DG typically use renewable energy sources such as solar power or wind power. This type of generation can be located closer to the load center and connected to the power system grid through the distribution network rather than the transmission network. Figure 2-2 shows a power system network with distributed generation.

Figure 2-1: Simplified representation of Power System Network ([16])

Figure 2-2: Simplified representation of Power System with Distributed Generation
In this section, we develop a simplified model to represent the impact of DG on the electric transmission grid. Figure 2-3 below shows a simplified representation of a distribution network connected to a distribution bus at the substation. The Substation Distribution Bus is the point of interconnection of the distribution network to the transmission grid. The model we have developed represents the impedance in the distribution network. As shown in Figure 2-3 each distribution network consists of several distribution feeders and each feeder has secondary distribution networks that connect to the end customers. There are two types of distributed generation considered in our analysis: rooftop solar PV generation (here classified as installations less than 20 kW AC), and solar PV farms which includes the commercial building installations and the solar PV farms (here classified as installation above 20 kW AC).

![Figure 2-3: Simplified representation of Distribution Network](image)

The model we developed in this Chapter adapts the methodology outlined in ([8]). The rest of this chapter describes the steps we used to derive the equivalent impedance of the distribution system network. First, we describe the DG system under consideration and the assumptions we made to derive the model. Secondly, we show the derived equations that represent the DG resources connected to the distribution network. Finally, the last section
provides a proof of derivation for the equations highlighted in step two.

2.1 System Description and Assumptions

Power system substation transmission buses in New England are typically operated at voltage levels of 69 kV to 345 kV. A transmission substation transformer steps the voltage down to a range of 13.2 - 34.5 kV, which is the interconnection point to the distribution feeder. This interconnection point is shown in Figure 2-3 as the "Substation Distribution Bus". There are typically 3-5 distribution feeders connected to each distribution bus at a substation. The distribution feeder voltage is further stepped down by a pole-top distribution transformer to 240/480 V and connected to the end use customer through the secondary distribution branch.

For our analysis we have considered two types of distributed solar PV generation: Rooftop solar PV and solar PV Farms.

Rooftop Solar PV

Figure 2-4 shows a typical solar rooftop panel connection to the grid. Based on the solar PV installation data we have, we consider solar PV generation with capacity between 5 to 20 kW as rooftop solar PV. Several rooftop solar PV generation unit from the customers are then connected to the pole-top transformer. Pole-top transformers are then connected back to the distribution feeder as shown in 2-4. The length of branches in the secondary distribution network is much shorter than the distribution feeder branch, and the impedance at the pole-top transformers are much greater than the secondary distribution branch. Thus, in our analysis we assume that the secondary branch impedance are negligible. The only impedance considered in the secondary distribution branch is the pole-top transformer impedance.
Solar PV Farms

For our analysis, these installations include the medium size commercial solar PV (e.g., on the rooftop of shopping centers) and large farms built for intensive electricity generation. Figure 2-5 shows a typical solar PV farm connection to the grid. A solar PV farm in MA can range anywhere from 25 kW for commercial installations to over 15 MW for larger solar PV farms. We have made the assumption that all solar PV farms connect directly to the distribution feeder through a Low Voltage/Medium Voltage transformer. Also, for the same reasons as mentioned in the rooftop solar PV case above, we only consider the impedance of the step up transformer for the secondary distribution branch.
2.1.1 System Assumptions

Our derivation of the equivalent impedance for the distribution network is based on the apparent power losses (real power losses and reactive power losses). To simplify and standardize the general equation for the circuit, we have made the following assumptions about the distribution network:

1. Branch Impedance: All distribution feeder branches in MA are modelled as having resistance and reactance values of 0.197 ohm/mile and 0.643 ohm/mile respectively. These typical values have been provided by Distribution Planning at National Grid.

2. Branches per distribution feeder: Typically each distribution bus has three distribution feeder interconnections. Even though the number of distribution feeders per distribution bus can vary significantly, based on input from Distribution Planning at National Grid.
Grid we have chosen a typical value of three feeders per distribution bus.

3. Distribution Feeder Type: Each distribution feeder is modelled as having only one type of distributed generation. For example, each distribution feeder either has all solar rooftop installations or all solar PV farm installations. Therefore, if a substation is predicted to have interconnection to both solar PV rooftop and solar PV farms, then we use one feeder for the farm interconnection and two other feeders for the rooftop PV interconnection. This approximation allows us to still represent the predicted rooftop and farm PV generation, but at the same time simplifies the circuit design.

4. Distribution feeder length: The typical length of distribution feeder is determined based on the location of the feeder. We have divided the location into three broad categories. See Table 2.1 below. Note that the load zones are defined in Figure 3-7.

<table>
<thead>
<tr>
<th>Load Zones</th>
<th>Feeder Length (miles)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boston</td>
<td>5</td>
</tr>
<tr>
<td>SEMA, LowerSEMA, NorthShore, CentralMass</td>
<td>10</td>
</tr>
<tr>
<td>WesternMass</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 2.1: Feeder Length in each Load Zone

5. Transformer Impedance: We calculated an average transformer impedance for rooftop solar PV and for solar PV farms. The corresponding average value was used for all the PV rooftop and PV farm transformer impedance in our analysis.

*Rooftop Solar PV Transformer Impedance:*
As shown in Figure 2-4, rooftop solar PV is connected to the electric power system through the pole-top transformer. Based on input from the Distribution Planning group at National Grid, we assume that the typical pole-top transformer has a 25 kVA nameplate rating. Additionally, we assumed a typical pole-top transformer impedance of 2.5% and X/R ratio of 1.31. The typical impedance and X/R ratio values were taken from GE transformer ratings given in ([28]). Based on these assumptions, the pole-top transformer impedance is equal to:
\[ \text{Average Resistance Rooftop} = 21 \ p.u. \text{ on 100MV Asystem} \quad (2.1) \]

\[ \text{Average Reactance Rooftop} = 28 \ p.u. \text{ on 100MV Asystem} \]

The calculation for the resistance and reactance values are shown in Table 6.5 in the Appendix.

**Solar PV Farm Transformer Impedance:**

As shown in Figure 2-5, solar PV farms are connected to the electric power system through a network of parallel transformers. Each transformer is connected to an array of solar panels with 1 MW of generation. Based on input from the Distribution Planning group at National Grid, we assume that the typical solar PV farm transformer has a 500 kVA transformer nameplate rating. Additionally, we assumed a typical solar PV farm transformer impedance of 5% and X/R ratio of 99. Based on these assumptions, the solar PV farm transformer impedance is equal to:

\[ \text{Average Resistance Farm} = \frac{0.1}{\text{Total MW of Generation}} \ p.u. \text{ on 100MV Asystem} \]

\[ \text{Average Reactance Farm} = \frac{9.9}{\text{Total MW of Generation}} \ p.u. \text{ on 100MV Asystem} \quad (2.2) \]

The calculation for the resistance and reactance values are shown in Table 6.5 in the Appendix.

Based on these assumptions, we derived an equivalent circuit for the distribution network represented in the following sections.

**2.2 Equivalent Model for Distributed Generation**

The equivalent model for the DG network is represented in terms of the following parameters:

\[ Z_b = \text{The equivalent impedance per secondary distribution branch} \]
\( Z_{dt} \) = The impedance of secondary distribution transformer  
\( n \) = Number of PV installations per branch  
\( Z_{fs} \) = Impedance of the branch between the secondary distribution transformers  
\( m \) = Number of secondary distribution transformers  
\( k \) = Number of feeders per substation distribution bus

The secondary distribution branch impedance, \( Z_S \), is represented by the equation (2.3)

\[
Z_S = \left( \frac{Z_b}{n} \right) + Z_{dt} \quad (2.3)
\]

The primary feeder branch impedance, \( Z_F \), is represented by the equation (2.3)

\[
Z_F = \frac{Z_S}{m} + \frac{(m + 1) \times (2m + 1) \times Z_{fs}}{6m} \quad (2.4)
\]

The total impedance, \( Z_T \), per substation distribution bus is represented by the equation (2.5)

\[
Z_T = \frac{Z_F}{k} \quad (2.5)
\]

Figure 2-6 highlights the parameters and the equations used to represent the distribution network and shows the equation derived above for each section of the network. The proof of derivation for the equations (2.3), (2.3), and (2.5) is described in the next section.
Using this standard equation, we were able to expand the existing electric power system model in the PSS/E software to represent the DG interconnection on the network. The equations derived here allow us to represent the various types of DG networks as a standard equation in terms of parameters such as the distribution feeder length and the impedance of transformers. For example, the distribution networks located in Western MA are much longer than the distribution networks in Boston. This difference in feature in the network is captured by the impedance $Z_{f,s}$ values, which is a function of the length of the network.

2.3 Derivation of Equations to Equivalence the Distributed Generation Network

In this section we first show the derivation of the equation for a secondary distribution network. Using the secondary distribution network model, we then derive the equation for a primary feeder branch network connected to a distribution bus at a substation.
2.3.1 Equivalencing of the Secondary Distribution Network

In this step we equivalence the secondary distribution network. The secondary distribution network can be represented as parallel branches of solar PV installations connected to the same node as shown in Figure 2-7. Each branch has a unique impedance and is connected to a group of solar PV installations. In Figure 2-7 we have considered a node with a three branch configuration.

![Diagram of secondary distribution network]

Figure 2-7: (a) Secondary Distribution Branch - Electrical Representation (b) Equivalent Model

Note that the current shown in each branch is a phasor quantity, as follows:

\[ I_m = I_m \angle \theta_m \]

In all the equations below we represent phasor quantities as boldfaced variables as shown above.

The total output current from the secondary distribution network is calculated as

\[ I_s = I_1 + I_2 + I_3 \]

(2.6)

The apparent power losses in each branch is calculated as:

\[ S_{Z_1} = I_1 * I_1^* * Z_1 \]
\[ S_{Z_2} = I_2 * I_2^* * Z_2 \]
\[ S_{Z_3} = I_3 * I_3^* * Z_3 \]
\[ S_{Z_{dt}} = I_s^2 * Z_{dt} \]

(2.7)
The Total loss is given by:

$$S_{Z_S} = I_S^2 \cdot Z_S$$  \hspace{1cm} (2.8)

Using the above equations we can derive the following equation:

$$I_S^2 \cdot Z_S = I_1^2 \cdot Z_1 + I_2^2 \cdot Z_2 + I_3^2 \cdot Z_3 + I_S^2 \cdot Z_{dt}$$  \hspace{1cm} (2.9)

To further simplify equation 2.9 we make the following assumptions:

1. Total predicted generation per secondary distribution branch is equally distributed among all the installations. Therefore, the MW of solar PV generated in each installation is the same.

$$G_1 = G_2 = G_3 = \ldots = G_n = G$$  \hspace{1cm} (2.10)

2. The current injected by each generator is assumed to be identical in magnitude and phase

$$I_1 = I_2 = I_3 = \ldots = I_n = I$$

based on equation (2.6) : $I_S = n \cdot I$  \hspace{1cm} (2.11)

where n is equal to the number of branches with solar PV generation

3. The impedance from each generator to the point of interconnection are equal.

$$Z_1 = Z_2 = Z_3 = Z_b$$  \hspace{1cm} (2.12)

Based on the above assumptions, we can generalize equation 2.9 as:

$$I_S^2 \cdot Z_S = I^2 \cdot (Z_1 + Z_2 + Z_3 + \ldots + Z_n) + I_S^2 \cdot Z_{dt}$$

using equation 2.11

$$I_S^2 \cdot Z_S = (I_S/n)^2 \cdot \sum_{m=1}^{n} Z_m + I_S^2 \cdot Z_{dt}$$  \hspace{1cm} (2.13)
From the previous equation we can derive the impedance of the secondary circuit as:

\[ Z_s = \frac{Z_b}{n} + Z_{dt} \]  

(2.14)

where \( Z_b \) represents the equivalent impedance per secondary distribution branch and \( Z_{dt} \) represents the impedance of secondary distribution transformer.

### 2.3.2 Equivalencing the Primary Feeder Branch Network

In this step we use the equivalent impedance for the secondary distribution network derived in step 2 to derive an equivalent impedance for a single distribution feeder branch. A distribution feeder branch can be represented as a combination of series and parallel branches of the secondary distribution network as shown in Figure 2-8.

![Diagram of distribution feeder branch](image)

**Figure 2-8:** (a) Distribution Feeder Branch - Electrical Representation (b) Equivalent Model

In Figure 2-8, \( Z_F \) represents the impedance in the feeder branch between two secondary distribution nodes. \( I_S \) and \( Z_S \) represent the equivalent current and impedance for the secondary distribution network as calculated from step 2.
The total output current in the distribution feeder is calculated as

\[ I_F = I_{S1} + I_{S2} + I_{S3} \]  \hspace{1cm} (2.15)

The voltage drop across each impedance can be calculated as

\[ \Delta V_{ZS1} = I_{S1} * Z_{S1} \]
\[ \Delta V_{Zfs1} = I_{S1} * Z_{fs1} \]
\[ \Delta V_{ZS2} = I_{S2} * Z_{S2} \]
\[ \Delta V_{Zfs2} = (I_{S1} + I_{S2}) * Z_{fs2} \]
\[ \Delta V_{ZS3} = I_{S3} * Z_{S3} \]
\[ \Delta V_{Zfs3} = (I_{S1} + I_{S2} + I_{S3}) * Z_{fs3} \]  \hspace{1cm} (2.16)

To further simplify equation 2.16 we make the following assumptions:

1. The aggregated predicted generation per secondary distribution network are assumed to be equal to each other.

2. Each distribution feeder can only have one type of distributed generation: either all solar rooftop installations or all solar PV farm installations.

3. Each secondary distribution network is identical in terms of total generation, number of branches (n) and total impedance.

\[ n_1 = n_2 = n_3 = \ldots = n_m = n \]
\[ I_{S1} = I_{S2} = I_{S3} = \ldots = I_{Sm} = n * I_S \] based on equation (2.6)
\[ Z_{S1} = Z_{S2} = Z_{S3} = \ldots = Z_{Sm} = Z_S \]
where m = number of secondary distribution branches on the distribution feeder  \hspace{1cm} (2.17)

4. Every secondary distribution network on the feeder branch is located at equal distances.

\[ Z_{fs1} = Z_{fs2} = Z_{fs3} = \ldots = Z_{fsm} = Z_{fs} \]  \hspace{1cm} (2.18)
Based on the previous assumptions and equation (2.16) we can compute the loss in each branch as

\[ S_{Zs1} = \Delta V_{Zs1} * I_{S1}^2 = I_{S1} * I_{S1}^2 * Z_{S1} = n_1^2 * I^2 * Z_{S1} \]

\[ S_{Zs2} = \Delta V_{Zs2} * I_{S2}^2 = I_{S2} * I_{S2}^2 * Z_{S2} = n_2^2 * I^2 * Z_{S2} \]

\[ S_{Zs3} = \Delta V_{Zs3} * I_{S3}^2 = I_{S3} * I_{S3}^2 * Z_{S3} = n_3^2 * I^2 * Z_{S3} \]

\[ (2.19) \]

\[ S_{Zf1} = \Delta V_{Zf1} * I_{S1}^2 = I_{S1} * I_{S1}^2 * Z_{f1} = n_1^2 * I^2 * Z_{f1} \]

\[ S_{Zf2} = \Delta V_{Zf2} * (I_{S1} + I_{S2})^2 = (n_1 + n_2)^2 * I^2 * Z_{f2} \]

\[ S_{Zf3} = \Delta V_{Zf3} * (I_{S1} + I_{S2} + I_{S3})^2 = (n_1 + n_2 + n_3)^2 * I^2 * Z_{f3} \]

Also, based on the above assumptions and equation (2.15), the total distribution feeder current is equal to

\[ I_F = (n_1 + n_2 + n_3) * I \]  (2.20)

The equivalent power loss in the circuit can be calculated as

\[ \text{Total Loss in distribution feeder} = S_{ZF} = I_F^2 * Z_F \]  (2.21)

By equating the total loss in the distribution feeder to the sum of the losses in individual branches, we can find the equivalent impedance value of the distribution feeder network.

\[ (I_F^2 * Z_F) = I^2 [(n_1^2 * Z_{S1} + n_2^2 * Z_{S2} + n_3^2 * Z_{S3}) + (n_1^2 * Z_{f1} + (n_1 + n_2)^2 * Z_{f2} + (n_1 + n_2)^2 * Z_{f3}) + (n_1 + n_2 + n_3)^2 * Z_{f3})] \]

Using equation 2.20
\[ Z_F = \frac{1}{(n_1 + n_2 + n_3)^2} \left[ (n_1^2 \cdot Z_{S1} + n_2^2 \cdot Z_{S2} + n_3^2 \cdot Z_{S3}) + (n_1^2 \cdot Z_{f1} + (n_1 + n_2)^2 \cdot Z_{f2} + (n_1 + n_2 + n_3)^2 \cdot Z_{f3}) \right] \]

The above equation can be generalized as:

\[ Z_F = \frac{1}{(\sum_{i=1}^{m} n_i)^2} \left( \sum_{i=1}^{m} n_i^2 \cdot Z_{S_i} + \sum_{i=1}^{m} \left( \sum_{j=1}^{i} n_j \right) \cdot Z_{f_{si}} \right) \quad (2.22) \]

This equation can be further simplified based on assumptions made in (2.17) & (2.18)

\[ Z_F = \frac{1}{n^2 \cdot m^2} \left[ (n^2 \cdot m \cdot Z_S) + (n^2(1 + 2^2 + 3^2 + \ldots + m^2) \cdot Z_{f_s}) \right] \]
\[ Z_F = \frac{n^2}{n^2 \cdot m^2} \left[ (m \cdot Z_S) + ((m \cdot (m + 1) \cdot (2m + 1)/6) \cdot Z_{f_s}) \right] \quad (2.23) \]
\[ Z_F = \frac{1}{m} \left[ (Z_S) + ((m + 1) \cdot (2m + 1)/6) \cdot Z_{f_s}) \right] \]
\[ Z_F = \frac{Z_S}{m} + \frac{(m + 1) \cdot (2m + 1) \cdot Z_{f_s}}{6m} \]

where \( Z_F \) represents the equivalent impedance per distribution feeder branch, \( Z_S \) represents the secondary distribution network impedance and \( Z_{f_s} \) represents the impedance of segment of distribution feeder branches between the secondary distribution networks.

### 2.4 Conclusion

In this Chapter we derived an equivalent model to represent the impact of DG on the electric transmission grid. This simplified equation to represent DG on the electric power system was created because the current transmission system model used by the utilities in NE currently do not have the capability to model the complete distribution network. The equivalent model derived in this section allows us to expand the existing electric transmission system model and represent the interconnection of DG resources to the electric transmission grid.

National Grid is currently investigating a single power system model that contains both
the Transmission and the Distribution network. However the software used for this modeling, Grid Lab D, and the detailed modeling of the network on the software are still in preliminary stages. The time line for this implementation is still a few years away. In the mean time National Grid can use the simplified model of the distribution network presented in this Chapter for analyzing the impact of DG resources.
Chapter 3

Predictive Model for Solar PV Location

In MA, the total installed solar PV capacity was at 250 MW DC in 2013. The goal set by the MA Governor is to reach 1,600 MW DC by 2020. We built our model based on the assumption that 1,600 MW DC will be reached by 2020. We believe that this is a good assumption because historically MA was able to achieve its earlier goal of 250 MW earlier than expected.

In this section, we develop a model that predicts the geographical location of 1,600 MW of solar PV installations in MA by 2020. The prediction model is based on historical data, together with geographic and demographic information about the different areas in the state.

The prediction model was developed in the following order:

1. Collect historical data and trends on the prediction parameters
2. Develop a prediction model for rooftop solar PV installations
3. Develop a prediction model for solar PV farm installations

3.1 Historical Data and Trends

Historical data was collected from “The Open PV Project” ([17]), an open source project sponsored by the National Renewable Energy Laboratory that records all the Solar Panels
installations across the country. For every installation, this database records the date, the location (GPS coordinates and ZIP code), the capacity and the cost of the installation.

<table>
<thead>
<tr>
<th>Date</th>
<th>GPS coordinates</th>
<th>ZIP code</th>
<th>Capacity (kW DC)</th>
<th>Cost($)</th>
</tr>
</thead>
</table>

Table 3.1: Columns of Open PV data set

Data from early in the year 2000 until the end of 2013 from the database was used to develop our model. In MA, this corresponds to almost 10,000 installations for a total capacity of 250 MW DC.

![Number of Installations per Quarter in MA](image1.png)

**Figure 3-1:** Number of installations per quarter in MA

![Evolution of Penetration Level for 20 ZIP codes](image2.png)

**Figure 3-2:** Evolution of Penetration Level for 20 ZIP codes

From Figure 3-1, we can notice that even though some solar PV installations are recorded in the first half of the year 2000, solar panels start becoming widely installed only after 2007. Since 2007, the installation rate has been increasing exponentially and there were over 1,500 installations per quarter in 2012. Only 6% of the solar PV in MA had been installed before 2007. Figure 3-2 shows the penetration level as a function of time for 20 different ZIP codes in MA.

The penetration level at time $t$ is defined as:

$$\frac{\# \text{Installations before time } t}{\# \text{ Installations in 2013}}$$

(3.1)
For each of the 20 ZIP codes highlighted, we observe a similar exponential growth in terms of the number of installations. Note that Figure 3-2 shows the increase in penetration level rather than the cumulative number of installations. This was done to normalize the data and plot all the ZIP codes on the same scale. The solar power generation capacity in 2013 varies significantly across ZIP codes (see Figure 3-3).

The capacity of each installation ranges between 0.5 kW and 5 MW DC. These two extreme values correspond to a small residential rooftop solar PV and a large solar farm. Therefore, the data set was split into two categories based on the size of installation:

- **Rooftop PV:** All the installations that are less than 20 kW fall in this category. These installations represent 90% of the installations up to 2013 but only 20% of the total capacity installed in 2013. The average capacity installed is 5 kW. This type of installation corresponds to individual home owners who decide to install solar PV on their roofs. The number of these installations vary significantly across ZIP codes.

- **PV Farms:** All the installations above 20 kW fall in this category. These installations represent 80% of the installed capacity. This includes the medium size commercial solar PV (e.g., on the rooftop of shopping centers) and large farms built for intensive electricity generation. These installations are not frequent but represent a large fraction of the capacity installed.
Using the collected data, we have built two separate models for rooftop solar PV and solar PV farms to capture this heterogeneity. For the rooftop installations, we have built a model to predict the number of installations at the ZIP code level. Whereas for the solar PV farms, due to the limited number of installations in MA, we divided MA in six load zones and predicted the aggregate capacity installed in every zone.

In the next two sections, we present the prediction models and results.

### 3.2 Rooftops

“Rooftop PV” solar panels are installed in residential areas by home owners on their rooftop or in the proximity of their home. Thanks to the well conceived incentives and tax rebates implemented in MA in the recent years, the applications for this type of solar panels have been exponentially increasing. Based on our analysis of data, there are a certain factors that are favorable for installing rooftop panels: enough solar radiance (no tree coverage for example), a roof oriented south, a owner occupied house. Furthermore, the installation of solar panels requires a significant upfront cost. Looking at historical data, we can observe that the number of installations and the adoption rate change significantly across ZIP codes. For example, in the Greater Boston area, the installation rate started increasing early in 2008 and slowed down by the end of 2012. Whereas, in Central MA the installations started appearing much later around 2010 but the rate has been increasing exponentially since then. We believe that these differences can be explained based on demographic and geographic characteristics of the ZIP code. Intuitively, in areas like Boston suburbs, with a medium density of houses and with a high proportion of educated and more affluent homeowners, the installations begin early but then a capacity saturation was reached quickly. In less densely populated areas, the first installations may occur later, but there is more available land for future growth in residential construction potentially adding more rooftop installations.
3.2.1 Data

First, we analyzed the historical installation data from the "Open PV" database. We notice that 90% of "Open PV" installations are below 20 kW. Also, their mean capacity is 5.8 kW and most of the installations are between 3 and 10 kW (see Figure 3-4). Because of the relatively small variability in installation size, the number of installations per ZIP code was predicted rather than the capacity. The output of the prediction model was converted into capacity by multiplying it by the mean capacity of an installation. Figure 3-5 and 3-6 illustrate the large variability in the number of installations before 2013 per ZIP code. The majority of the ZIP codes have less than 15 installations, but some others have more than 100 (mainly in the Boston area). We considered only ZIP codes that have more than 20 installations up to 2014. This allowed us to build an effective predictive model that shows a trend on the number of installations over time. This criteria resulted in 150 ZIP codes out of the 530 ZIP codes, and captured 70% of the rooftop solar installations in MA. Research has shown that there is a distinctive pattern of geographical clustering in the diffusion of solar technology in California ([18]). Furthermore, research also shows that there may be peer effect, whereby previous choices of owner's neighbors influence the decision to adopt solar PV panel ([18]). Therefore, we believe that moving forward a similar trend will be seen
in MA and these 150 ZIP codes will see maximum increase in rooftop solar PV installation. Therefore, in our implementation of the model, we restricted our analysis to these 150 ZIP codes.

Additionally, we collected geographic and demographic information at a ZIP code level from open source database and selected features that could explain the differences in historical trends per ZIP code. Data was extracted from the 2010 census collected from the Government Census Bureau website. For our analysis, we assume that these ZIP codes’ characteristics do not change significantly during the time frame of our study (approximately 15 years).

The features we extracted can be split into four major categories:

- **Solar Radiance:** Annual and Monthly Average, Minimum, Maximum and Standard Deviation
- **Households:** Number of Residents (2010 census), Population Density, Number of Houses, and Number of Owner Occupied Houses
- **Demographic:** Household Income, Age Distribution, Political Orientation, Education Levels
- **Load Zones:** We divide MA in 6 main load zones illustrated in Figure 3-7.

Figure 3-6: Number of rooftop installations per ZIP code in 2013
An exhaustive list of the features we considered is reported in the Appendix in Table 6.1.

3.2.2 Model

Exponential Growth  As illustrated in Figure 3-2, the number installations in every ZIP code has increased exponentially since 2007. However, the growth is different across ZIP codes in terms of how fast it increases, and when it starts increasing exponentially. It is important to note that the historical data represented Figure 3-2 is in the early stage of solar PV installations in MA. The total capacity currently installed in MA is 250 MW but the goal for 2020 is 1,600 MW DC. We believe that this goal will be reached. This is a reasonable assumption because MA reached its earlier goal 250 MW earlier than expected because of the policies implemented by the government.

An exponential model for solar PV installation has been used to model the installation growth in California ([24]) where 65% of the total solar PV in the United States was installed as of 2010. In MA, even though an exponential model is a good approximation for the current growth rate, but this may not be appropriate for predictions on a 10-20 years scale. The growth rate is extremely high on historical data but it is likely to decrease in the future when the novelty of the incentives will have less impact. For this reason, we investigated a logistic model, which we believe to be more realistic to predict the solar PV installations in 2020.
**Logistic Model**  We used a logistic growth model to fit the number of rooftop solar PV installed in every ZIP code. The logistic model is widely used in the biology literature to model population growth. It allows us to capture an exponential growth rate in the first period and a successively declining growth rate. In the biology application, cells placed in a rich environment reproduce very fast initially but, when they reach a certain number, the resources become rare and the population growth decreases. A good introduction to the logistic growth model, its estimation methods and applications in biology can be found in ([25]). The same approach can be used to model the penetration rate of solar panels.

![Shape of logistic growth](image)

**Figure 3-8: Shape of logistic growth**

The logistic function is represented in Figure 3-8 and is characterized by the equation

\[ N(t) = \frac{M}{1 + e^{-(\beta t + \beta_0)}} \]

where \( t \) is the time, \( N(t) \) is the number of installations before time \( t \) and \( M, \beta, \beta_0 \) are the three characteristic parameters. \( M \) is the maximum capacity: \( M = \lim_{t \to \infty} N(t) \), \( \beta \) measures the speed of the growth rate and \( \beta_0 \) is an intercept that captures the time when the exponential growth starts. In the solar panels application, a large value of \( \beta \) represents a fast growth, and a small value of \( \beta_0 \) represents a growth that starts early.

Logistic growth model was used for every ZIP code and it was assumed that the three parameters depend on the ZIP code demographic characteristics. Before defining the model further, we introduce some useful notations. Let \( t \) be the index of time and \( z \) the index of a ZIP code. Then we define

- \( N_{t,z} \): Number of installations in ZIP code \( z \) at time \( t \)
- \( X_z \in \mathbb{R}^p \): vector of demographic characteristics (see Table 6.1) for ZIP code \( z \)
For every $z$ the number of installations follow a distinct logistic growth model with parameters $(M_z, \beta_z, \beta_{0,z})$ that are linear combinations of the features vector $X_z$.

Therefore, for every $(t,z)$

$$N_{t,z} = \frac{M_z}{1 + e^{-(\beta_z t + \beta_{0,z})}}$$

(3.2)

with:

$$M_z = m \cdot X_z$$

$$\beta_z = \beta \cdot X_z$$

$$\beta_{0,z} = \beta_0 \cdot X_z$$

where $x \cdot y$ denotes the inner product of $x$ and $y$ and $(m, \beta, \beta_0) \in \mathbb{R}^p$ are three parameter vectors that need to be estimated.

This model relies on the assumption that the number of installations follow a logistic growth for every ZIP code but with different parameters. These parameters depend linearly on the ZIP code’s demographic characteristics. Note that different demographic characteristics can have different impacts on $m$, $\beta$ and $\beta_0$. For example, a ZIP code with many houses will have a large $M_z$. Population, income, and political orientations can lead to an early adoption rate $(\beta_{0,z})$. We assume that $M_z$, $\beta_z$ and $\beta_{0,z}$ are linear combinations of the demographic characteristics of the ZIP code, thus the effects of the demographic characteristics are captured by only three vectors of parameters. Choosing a unified model (captured with this linear relationship) rather than estimating a distinct model for every ZIP code has several advantages. The small number of parameters allows us to build a robust model that can handle outliers and can make predictions for ZIP codes with fewer observations so far. This also helps in extracting meaningful insights on the effect of the different demographic characteristics; such information can be very useful for policy making.

**Constraints** Values of $(m, \beta, \beta_0)$ were estimated to best fit the historical data and satisfy some additional constraints driven from a prior knowledge on the penetration level in 2020.

Firstly, we know that the goal for 2020 is to have 1,600 MW DC of installed solar PV capacity in MA. As mentioned earlier, we assume that the incentives will be adjusted over
time to reach 1,600 MW DC but not to exceed it. In addition, we assume that the ratio of capacity for rooftop/farms will be consistent to the ratio that exists today (20% of the capacity is from rooftop solar PV and 80% of the capacity from solar PV farms). This means that in 2020 approximately 320 MW of rooftop capacity will be installed. (This number can be translated into number of installations by dividing 320 MW by the mean capacity of a rooftop solar PV).

Secondly, there is an upper and lower bound on the maximum penetration level that can be reached in every ZIP code. A simple lower bound is that the maximum number of installations ($M_z$) has to be greater than the penetration level in 2013.

Experts can add more sophisticated upper bounds on $M_z$. For our model, the upper bound on penetration level cannot exceed a certain fraction of the number of households or a certain percentage of the land in the ZIP code.

To simplify the problem, we approximate the penetration level in 2020 over the entire state by the penetration level at $t = +\infty$. This assumption is justified by the fact that our training set represents only the early stage of the growth process and 2020 is far in the future. We will check the validity of this assumption on data later.

With this assumption, we can write the constraints as linear constraints on $M_z$:

"The maximum total number of installations in MA is approximately $N_{max}$" becomes:

$$ (1 - \epsilon)N_{max} \leq \sum_z M_z \leq (1 + \epsilon)N_{max} $$

(3.3)

$\epsilon$ is used to capture the uncertainty in the final number of installations, in our application we use it as an uncertainty level of 10% ($\epsilon = 0.1$). Similarly, "The maximum total number of installations in each $z$ is less than $\alpha$% of the number of houses" becomes:

$$ M_z \leq \alpha N_{houses,z} \quad \forall z $$

(3.4)

In the next section, we formulate an estimation setting that is able to incorporate constraints that can be written as linear inequalities on the parameters $m, \beta$ and $\beta_0$. 

45
3.2.3 Estimation

We show, in this section, how to estimate the parameters $(m, \beta, \beta_0)$ by non-linear least squares and incorporate the constraints presented in the previous paragraph.

Consider a set of pairs (ZIP codes, time) denoted $(z, t)$ and assume that for each pair the number of installations in ZIP code $z$ until time $t$ ($N_{z,t}$) is observed. Then we can estimate the common parameters of the logistic growth function by non-linear least squares, by solving the following optimization problem:

$$
\min_{m, \beta, \beta_0 \in \mathbb{R}^p} \sum_{z,t} \left( N_{z,t} - \frac{mX_z}{1 + e^{-\left(\beta X_z + \beta_0 + X_z\right)}} \right)^2
$$

subject to

$$(1 - \epsilon)N_{max} \leq \sum_z mX_z \leq (1 + \epsilon)N_{max} \quad (3.5)
$$

$$
mX_z \leq \alpha_h N_{houses,z} \quad \forall z
$$

$$
mX_z \geq \alpha_t N_{2013,z} \quad \forall z
$$

$N_{houses,z}$ represents the number of houses in $z$ and $N_{2013,z}$ represents the number of installations in $z$ before 2013.

The objective function of (3.5) represents the sum of the squared prediction errors. The constraints presented in the previous section are added as linear constraints on the parameter $m$. This optimization formulation can handle any other type of linear constraints on the parameters. Least squares is a standard approach for parametric data fitting. The goal is to find the value of the parameters that minimize the sum of the squared residuals (prediction errors). When the prediction error is a linear function of the parameters, this problem has a closed-form solution. When the residuals are nonlinear in the parameters, as it is the case here, iterative approaches are used where at each iteration the system is approximated by a linear system ([26]).

For this implementation, we set $t = 0$ in January 2008 and we measure the time in terms of weeks. For every ZIP code $z$, we consider the number of installations ($N_{t,z}$) every 10 weeks. This scale of time was used in order to have a large number of observations and to
avoid a very short time step where $N$ does not increase from one observation to the next. We also set $\alpha_h = 0.5$ which corresponds to 3 times the maximum penetration level in 2013, and $\alpha_t = 1.5$. These values can be adjusted with more insights on the growth behavior.

Recall that the vector $X_z$, that denotes the demographic characteristics of $z$, appears three times in the logistic model $(M_z, \beta_z, \beta_0)$. Nevertheless, we do not assume that the same features have to appear in the three coefficients. We used a greedy backward selection approach for feature selection (see [26] for more details). We first solve problem (3.5) with all the features collected in Table 6.1, and then successively eliminated the non-significant features using standard t-tests.

### 3.2.4 Results

Our objective is to build a robust model that is able to simulate the growth of the number of installations in every ZIP code. In order to accomplish this, we need to analyze ZIP codes with a sufficiently large number of installations before 2013. As illustrated in Figure 3-2, ZIP codes are extremely heterogeneous and most of them have only a few installations in our data set. We solve problem (3.5) on the subset of 150 ZIP codes with the greatest number of installations across MA. The same model can be solved for all MA in the future when additional data is available. The significant features and their coefficients are reported in Table 6.2 in the Appendix. Note that different features influence the parameters $(m, \beta, \beta_0)$. Recall that $m$ represents the maximum number of installations, $\beta$ represents the “speed of the growth” and $\beta_0$ represents the time when “the exponential growth starts”.

Qualitative insights can be derived from the results shown in Table 6.2. A high solar radiance, a low population density, a large number of households and a strong democratic orientation have a positive impact on the limit of the number of installations. Whereas, a high median income and a low cost of installation increase the speed of growth. Notice also that the load zones have a significant impact on the three parameters, revealing differences across MA that were not already captured by the other features.

We used three metrics listed below to evaluate the performance of the predictive model. Let $\hat{N}_{t,z}$ denote the predicted value for $(z, t)$. 

47
1. The Mean Absolute Percentage Error:

\[ MAPE = \sum_{z,t} \frac{|N_{t,z} - \hat{N}_{t,z}|}{N_{t,z}} \]  

2. Weighted Mean Absolute Percentage Error:

\[ WMAPE = \frac{\sum_{z,t} |N_{t,z} - \hat{N}_{t,z}|}{\sum_{z,t} N_{t,z}} \]  

3. The correlation between predicted and actual value.

To evaluate the model’s in and out of sample performance, we trained the model on a random subset of 100 ZIP codes (and compute the “In sample” metrics) and then computed the “Out of sample” metrics on the remaining 50 ZIP codes. Finally, recall that our model is trained on “Open PV” data that records the installations until the beginning of 2013. We collect data about the installations in 2013 and the first half of 2014 from a separate source [19] and evaluated the performance of our model on forecasting the installations until 2014. The results are reported in Table 3.2.

<table>
<thead>
<tr>
<th></th>
<th>In sample</th>
<th>Out of sample</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAPE</td>
<td>0.35</td>
<td>0.37</td>
<td>0.48</td>
</tr>
<tr>
<td>WMAPE</td>
<td>0.35</td>
<td>0.31</td>
<td>0.43</td>
</tr>
<tr>
<td>Correlation</td>
<td>88%</td>
<td>78%</td>
<td>75%</td>
</tr>
</tbody>
</table>

Table 3.2: Performance of the predictive model

We observe a strong correlation between the prediction and the actual number of installations, both in sample and out of sample. Furthermore, the Absolute Percentage Errors have satisfying levels considering that the number of installations is low (the average is 20). It is important to notice that there is not a significant drop in the prediction performance between in sample and out of sample. This reveals that our model is robust and does not
overfit the data. Nevertheless, there is room for improvement in the prediction accuracy. The same model, when trained with a larger amount of installation data will be able to significantly outperform the current accuracy.

Figure 3-9: Number of installations and prediction for 01824. The circles represent the data and the line represents the fitted model.

Figure 3-10: Number of installations and prediction for 02790. The circles represent the data and the line represents the fitted model.

Figure 3-11: Prediction for 2020 for 150 ZIP codes. The colors represent the prediction and the size of the circles represents the installations until 2013.
Figures 3-9 and 3-10 represent the data and the fitted model for two ZIP codes. For both figures, the observations since 2013 seem to follow an exponential growth. The model is able to capture the future decrease in the growth rate, especially in Figure 3-10 where we can distinguish the shape of the logistic function. Figure 3-11 is a map of the 150 ZIP codes in the training set. The color represents the predictions for 2020 and the size of the circles represents the installations until 2013. We can observe that most of the dark red circles are in the Boston area, which is not surprising because Boston is an area with the greatest number of residential apartments/houses with roofs. We can also observe a certain correlation between size and color. On average a large number of installations before 2013 implies a large number of installations in 2020. This is not surprising because, as mentioned earlier, 2013 is at the beginning of the growth period and most of the regions are still very far from their capacity limit. At the same time, we also observe few areas with smaller dark red circles. These are areas predicted by our model to have a significant increase in installations by 2020 but do not have a large number of installations currently.

3.3 Farms

We considered as farms all the installations that have a capacity of more than 20 kW. They range from commercial solar PV (e.g., installed on the rooftop of shopping malls) to large photo voltaic power stations (a few installations reach 1 MW in our data set). These installations produce the majority of the photovoltaic power in MA, but because of their size, there are not many such installations. Solar farms are generally installed in rural areas because of large land requirement. Former landfills, golf courses or agriculture fields are also good candidates to be converted to solar farms.

Based on the criteria described above, in our data set approximately 900 installations are considered as farms. However, there is a large variability in their capacity. Thus, in this context, it is key to be able to predict the total capacity installed rather than the number of installations. The installation of a solar farm comes from a different decision process than for a rooftop solar PV. Rooftop solar power generation comes from a multitude of small panels
in residential areas by individual consumers. In most cases, home owners only wait a couple of months between applying for authorization and having a working solar panel installed on their rooftop. As seen in previous section, we can build a good forecast model by looking at historical data and demographic information on the corresponding areas. But, solar farms have more restricting geographical constraints and have long installation times. For these reasons, we use an aggregated approach combined with geographical data to predict the solar farms installations in 2020.

### 3.3.1 Model

We do not have enough data points to build an effective predictive model at a local scale (ZIP code). Therefore, a two step approach was used to predict the geographical location of solar farms installed until 2020.

1. We divided MA in six geographical areas and predicted the capacity installed in every zone by extrapolating the current trend

2. For each geographical area, we use geographical data to distribute the total capacity across towns.
Prediction per load zone

Figure 3-12: Cumulative capacity installed in every load zone

Figure 3-13: Load zones for farm prediction

Figure 3-12 shows the cumulative capacity installed in each of the load zones from 2010 to 2014. We can clearly distinguish two growth modes. On one hand, Boston, NorthShore and WesternMassW have a “slow growth” that can be approximated by a linear or polynomial function. This makes sense because Boston is a very densely populated area where there is no room for large solar farms. The installations over 20 kW are mainly small commercial solar PV. Also, even though WesternMassW and NorthShore are low population density areas, these areas have fewer substations. Thus WesternMassW and NorthShore are not very attractive for solar farms installations. On the other hand, SEMA, CentralMass and WesternMassE follow a very fast paced growth that can be approximated by an exponential or a logistic function. These areas have low or medium population density and are well connected to the grid.

We fit a distinct model for each load zone, and use it to extrapolate the number of installations in 2020. For the “slow growing” areas we fit a linear or a polynomial function, versus for “fast growing” areas we fit an exponential or a logistic function. The data and resulting fitted curves are reported in Figure 3-14. The corresponding coefficients are reported in Appendix Table 6.3.
In WesternMassE the actual growth can only be captured by an exponential function. Nevertheless, this model would lead to a significantly excessive capacity (much higher than the expected total of 1,600 MW DC in MA) installed by 2020. Therefore, we used a constraint that 1,600 MW DC has to be installed in MA by 2020 to correctly estimate the capacity installed in this region: first we forecast the capacities for the other zones and for the rooftop solar PV separately and then attribute the rest of the capacity to WesternMassE. Table 3.3 summarizes the capacity installed in 2014 and the predictions for 2020.
<table>
<thead>
<tr>
<th>Load Zone</th>
<th>Capacity Installed in farms (MW)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2014 (data)</td>
</tr>
<tr>
<td>Boston</td>
<td>19.8</td>
</tr>
<tr>
<td>NorthShore</td>
<td>30</td>
</tr>
<tr>
<td>WesternMassW</td>
<td>36.9</td>
</tr>
<tr>
<td>CentralMass</td>
<td>153.8</td>
</tr>
<tr>
<td>SEMA</td>
<td>153</td>
</tr>
<tr>
<td>WesternMassE</td>
<td>145</td>
</tr>
</tbody>
</table>

Table 3.3: Capacity installed before 2014 and predictions for 2020 for the six load zones

**Distribution of Predicted PV within Load Zones**

In this step we distribute the predicted capacity per load zone using geographical data of each load zone. We assume that the availability of areas that have certain land type characteristics such as open land, golf courses, capped landfills, high density of commercial buildings, and medium density residential housing will determine the installation of solar farms.

We extracted the land usage data per town using the Geographic Information System (GIS) data provided by the state of MA ([23]). Figure 3-15 shows a sample land data visualization in Arc GIS software. Arc GIS's "Union (Analysis)" functionality was used to overlay the town boundary data for MA and the land type data for each town. This function breaks up the land types that spread across town boundaries into two separate land pieces and assigns the town ID to each land area. Using the new area map created by the "Union" functionality, we were able to calculate the land usage type within a town. The land usage types that were considered are: Open Land, Commercial Area, Golf Course, Medium Density Housing, and Capped Landfill.
Figure 3-15: Using Arc GIS Software to Calculate Land Usage Per Town

We then assigned a weight (expected % of capacity in each town) to each land type based on our best opinion. It is important to note that these weights can be easily updated to rerun the analysis if required. The assigned weights for the results shown in this thesis are shown in Table 3.4

<table>
<thead>
<tr>
<th>Land Type Data</th>
<th>Weight (Expected % of capacity in each type)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open Land</td>
<td>50%</td>
</tr>
<tr>
<td>Golf Course</td>
<td>10%</td>
</tr>
<tr>
<td>Commercial Area</td>
<td>5%</td>
</tr>
<tr>
<td>Medium Density Housing</td>
<td>5%</td>
</tr>
<tr>
<td>Unused Landfill</td>
<td>30%</td>
</tr>
</tbody>
</table>

Table 3.4: Weight Assigned per Land Use Type

Based on the land use data and the weight assigned per land type we predicted the installed capacity of solar PV farms as shown in equations 3.8 and 3.9. First we calculate the fraction of each type of land available in each town in the load zone. For example, the fraction of open land available in Arlington is equal to the open land area in Arlington divided by the total open land available in load zone - Boston. The values shown in the Table 3.5 for the fraction of capacity per town, was calculated as shown in equations 3.8 and 3.9 below.
\[
\text{FracCapacity}(\text{Arlington}) = \text{FracOpenLand}_{\text{Arl}} \times W_{\text{OpenLand}} \\
+ \text{FracGolfCourse}_{\text{Arl}} \times W_{\text{GolfCourse}} \\
+ \text{FracCommercial}_{\text{Arl}} \times W_{\text{Commercial}} \\
+ \text{FracMResidential}_{\text{Arl}} \times W_{\text{MResidential}} \\
+ \text{FracLandfill}_{\text{Arl}} \times W_{\text{Landfill}} \\
\]

\[
\text{Capacity}(\text{Arlington}) = \text{FracCapacity}(\text{Arlington}) \times \text{TotalCapacityPredictedinBoston} \\
\]

Table 3.5: Predicted capacity of solar PV farm per town in Load Zone - Boston

Table 3.5 above shows the calculation for load zone Boston. Total solar PV farm capacity was predicted for towns in other load zones using the same principle.

### 3.4 Prediction Result: Solar PV Rooftop + Solar PV Farms

Based on the rooftop and solar PV farm prediction, we obtain the total capacity of solar PV predicted per town. However, we need to map the predicted solar PV to the towns that have a substation. Therefore, we mapped all the substations in MA to the geographically
nearest town. Table 3.6 below shows a sample set of substation buses in the system that were mapped to the corresponding town/ZIP code based on its location. Next, we predicted the solar PV generation for each town that contained a substation. For towns with no substation, the predicted solar PV for that town was assigned to the geographically nearest town with a substation.

<table>
<thead>
<tr>
<th>Bus Number</th>
<th>Bus Name</th>
<th>Substation Name</th>
<th>Town</th>
<th>Zip Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>111301</td>
<td>ARSENE 13.200</td>
<td>Arsen 634</td>
<td>FAIRHAVEN</td>
<td>02719</td>
</tr>
<tr>
<td>113361</td>
<td>ASHBURNHM 113.800</td>
<td>ASHBURNHAM 610</td>
<td>ASHBURNHAM</td>
<td>01430</td>
</tr>
<tr>
<td>113363</td>
<td>ASHBURNHM 213.800</td>
<td>ASHBURNHAM 610</td>
<td>ASHBURNHAM</td>
<td>01430</td>
</tr>
<tr>
<td>116589</td>
<td>ASHFIELD 23.000</td>
<td>Ashfield</td>
<td>ASHFIELD</td>
<td>01330</td>
</tr>
<tr>
<td>115509</td>
<td>AVON 13.800</td>
<td>AVON UNIT 68</td>
<td>AVON</td>
<td>02332</td>
</tr>
<tr>
<td>113375</td>
<td>AYER R1 13.800</td>
<td>AYER</td>
<td>AYER</td>
<td>01432</td>
</tr>
<tr>
<td>113382</td>
<td>AYER R2 13.800</td>
<td>AYER</td>
<td>AYER</td>
<td>01432</td>
</tr>
<tr>
<td>111042</td>
<td>BAKER ST 24.000</td>
<td>Baker St. #10</td>
<td>BOSTON</td>
<td>02108</td>
</tr>
<tr>
<td>113082</td>
<td>BARRE MA T1 13.800</td>
<td>BARR 604</td>
<td>BARRE</td>
<td>01005</td>
</tr>
<tr>
<td>113081</td>
<td>BARRE MA T2 13.800</td>
<td>BARR 604</td>
<td>BARRE</td>
<td>01005</td>
</tr>
<tr>
<td>114711</td>
<td>BARTHOLOMW 123.000</td>
<td>BARTHOLOMW STREET</td>
<td>PEABODY</td>
<td>01960</td>
</tr>
<tr>
<td>114712</td>
<td>BARTHOLOMW 223.000</td>
<td>BARTHOLOMW STREET</td>
<td>PEABODY</td>
<td>01960</td>
</tr>
<tr>
<td>115731</td>
<td>BATES STREET 13.800</td>
<td>BATES 115</td>
<td>FALL RIVER</td>
<td>02720</td>
</tr>
</tbody>
</table>

Table 3.6: Sample mapping of transmission buses in MA

Figure 3-16 shows the total solar PV including rooftop solar PV and farms in the towns that have a substation. The size of the circle indicates the amount of solar PV expected in the town. The number below the circle represents capacity in MW, and the number below the town name represents the number of substations located in the town. To predict the total generation per substation, we assume that the total generation per substation is equally divided between all the substations in a town.
Figure 3-16: Total solar PV prediction in MA in 2020
3.5 Conclusion

In this Chapter we predict the location and amount of solar PV DG in each ZIP code/town in MA based on the assumption that the 2020 installed capacity in MA would be 1,600 MW DC - the target set by the MA Governor.

We have developed separate prediction models for the solar PV rooftop panels and the solar PV farms. The total predicted generation from the model was then assigned to the corresponding substation in the proximity. The prediction models are completely data driven and can be continuously improved as additional data become available.

Based on our literature review, we believe that currently there are no prediction models to predict the capacity and the geographical location of solar PV in MA to the level of granularity presented in this Chapter. Therefore, this model will assist the Transmission Planning group at National Grid to better understand and plan for the impact of DG of solar PV on the MA electric power system network.
Chapter 4

Electric Transmission System Model with Distributed Generation of Solar PV - Design, Analysis, and Results

In this section, we add the predicted distributed generation of solar PV to the existing transmission system model in PSS/E. The model developed in this section can be used with any power system software, however we use PSS/E because it is currently used by National Grid and ISO-NE (Independent System Operator-New England) to model the New England (NE) electric transmission system. This model is then used to analyze the reliability of the transmission system when 1,600 MW DC of distributed generation is added to the system.

4.1 Design

The predicted solar PV per substation was added to its corresponding distribution bus as shown in Figure 4-1. The circuit on the left shows the PSS/E model without any distributed generation. The circuit on the right represents the updated circuit where we add two generators, one for rooftop solar PV and one for solar PV farm. The impedance in the bus connecting distributed generator to the distribution bus was calculated using the distributed generation equivalence equations derived in Chapter 2 of this thesis.
There are over 300 distribution buses in the MA system. Therefore, to speed up the process and to enable repeatability we created a Python script that automatically adds a distributed generation circuit with the required generation and impedance. The input to the Python script is given through a spreadsheet that lists the bus number and the corresponding generation and impedance for each node.

### 4.2 Analysis and Results

All our analysis was performed on steady state conditions. First, we obtained the current NE transmission system model from ISO-NE for three system loading conditions: Light Load, Shoulder Peak, and Summer Peak. Using this power system model from ISO-NE we ran a power flow study to conduct a Voltage analysis and N-1 contingency analysis for each of the loading cases.

A power-flow study, is a numerical analysis of the flow of electric power in an interconnected system. A power-flow study usually uses simplified notation such as a one-line diagram and per-unit system, and focuses on various aspects of AC power parameters, such as voltages, voltage angles, active power and reactive power. Results from the power flow study calculates the voltages and currents that different parts of the system are exposed to.

A N-1 contingency analysis tests the resulting voltages and power flows when one of the elements in the electric power system grid is taken out of service.
Voltage Analysis

The following steps were taken to analyze the changes in system voltages when distributed generation is added to the system:

1. First we created a base case by distributing 1,600 MW DC equally among all substation distribution buses in each of the system loading condition models received from ISO. In the second run of this analysis we distributed the 1,600 MW DC based on the prediction model developed in Chapter 3 of this thesis.

We used a Python script to build this system. Note that 20% of the total solar PV capacity was assigned to rooftop solar PV and 80% was assigned to solar PV farms. This is based on the assumption that the current proportion holds true in 2020.

2. Switched off 1,600 MW dc of generation in the New England power system so that the power flow in the grid interfaces with the neighboring systems such as New York and New England does not change.

3. Ran the steady state system load flow analysis on the interconnected system. The load flow parameter settings in PSS/E are shown in Table 6.4 in the Appendix.

4. Analyzed the changes in the substation bus voltages to the corresponding bus voltages before solar was added.

5. Repeated the analysis for different voltage setpoints at the solar PV generator buses. Schedule voltage is the voltage set point used by each PV generator in the simulation.

The change in voltage for the three system loading conditions Light Load, Shoulder Peak, and Summer Peak are shown in Figures 4-2, 4-3, and 4-4 correspondingly. In each figure the upper chart shows the case with 1,600 MW dc equally distributed among all substation distribution buses. Whereas, the lower chart shows the case with 1,600 MW DC distributed based on the prediction model developed in Chapter 3 of the thesis. The x-axis lists the buses with maximum change in voltage after addition of solar PV to the system, and the y-axis shows the magnitude of per unit voltage at the bus. The red color bar shows the base level voltage when no solar is added to the system, whereas the orange, green, and pink bars
show the per unit voltage at different voltage schedules after solar PV has been added to the electric power system.

Figure 4-2: Changes in Transmission Voltage - Summer Peak Case
Figure 4-3: Changes in Transmission Voltage - Light Load Case

Figure 4-4: Changes in Transmission Voltage - Shoulder Peak Case
We notice that the changes in voltage in all the cases due to the addition of the solar PV generation are within a +/- 5% limit, which is the typical accepted range of electric power system operations. Note that the plots above only represent the buses with maximum transmission voltage changes. All the other buses in the system have an equal or a lower change in voltage.

**N-1 Contingency Analysis**

The N-1 contingency test on select major transmission lines to explore potential reliability concerns. The transmission lines selected for contingency analysis are at various voltage levels and spread out in MA. We repeated the analysis for the three loading scenarios provided by ISO.

Contingency analysis did not show any reliability concerns such as line overloads or voltage issues. Therefore, based on our analysis the transmission system, with 1,600 MW DC distributed solar PV added in the distribution system, is reliable even if one of the above transmission lines are taken out of service one at a time.

### 4.3 Conclusion

The work presented in this chapter combines the equivalent distribution network impedance derived in Chapter 2 with the prediction model developed in Chapter 3 to build the electric power system network in MA with the predicted solar PV generation resources. The model presented adds the predicted DG resources to the existing transmission model in PSS/E software.

Based on our load flow study on the MA transmission system, we conclude that with the addition of 1,600 MW DC the voltage levels in the MA electric transmission system have limited impact and would be within the operating range of +/- 5%. Furthermore, our contingency analysis concludes that the MA electric transmission system can operate reliably in the event of an outage of major transmission lines.
Chapter 5

Conclusions and Future Work

5.1 Generic Simulation Framework

The DG of solar PV in MA has increased over 80 times since 2007 ([4]) and is anticipated to grow exponentially in the next decade driven by green energy initiatives. Interconnection of such large quantities of solar PV to the electric power system grid poses potential reliability risks to the transmission network. To analyze this risk we have built a forecasting framework of well assessed demand projections with machine learning algorithms on the one hand, and a network analysis model to simulate the interconnection of DG of solar PV to the electric power grid on the other hand. Our literature review from Chapter 1 suggests a lack of such comprehensive modeling at this time. This thesis is a concentrated effort to fill the void by creating a model in collaboration with National Grid using real time data such as electricity consumption and Massachusetts area demographic data. The simulation framework is generic in nature and can be adapted for network impact analysis in other geographical region.

This thesis presents a methodology to model the DG interconnection on the transmission network, to predict the capacity and location of solar PV generation in MA, and to build a simulation framework that can be used to analyze the impact of DG of solar PV on transmission network in MA.

We first derived an equivalent electric circuit model for the distribution network to add the solar PV distributed generation resources to the existing transmission system model built in PSS/E that is used by National Grid. The model developed in this thesis provides
a simplified and standard way to represent the impedance and the aggregated distributed generation capacity of solar PV at each transmission substation. We then developed a forecasting model to predict the capacity and geographical distribution of solar PV generation in MA by the year 2020.

The forecasting model is the first of its kind to allow for the prediction of solar PV demand at the ZIP code level. We have built two separate models for the PV farms and PV rooftop installations, because the trends and factors that impact the two categories are different. Analysis of data in MA shows that the parameters, high solar radiance, low population density, larger number of residential houses, and strong democratic political party orientation have a positive impact on total capacity of solar PV installations we can expect in a region. Additionally, locations with high median income and low cost of installation have a strong correlation to the speed of growth of solar PV installations in a region.

5.2 Application of the Generic Simulation Framework to Predict Impact of DG of solar PV in MA

We used the prediction model to aggregate the total DG expected per ZIP code in MA. To test the MA electric transmission grid, we added 1,600 MW DC of solar PV from the forecasting model to the existing electric transmission system model in PSS/E. The generation resources were modeled in PSS/E using the equivalent circuit model derived for the distribution network. Using this network model, we ran an electric power flow study to analyze the steady state voltages at both transmission and distribution buses. The new voltage levels were compared with the voltage levels at the buses when no solar was added. The voltage level comparison derived from our simulation model affirmed that the MA transmission system can operate reliably with the addition of 1,600 MW DC of solar PV.

We also ran selected contingency analysis to test the system reliability in the event of outage of crucial transmission line segments. For the selected contingency scenarios that were tested, the transmission network performance was acceptable and within the established power system ratings.
In conclusion, our analysis of the MA transmission system shows that the addition of distributed solar PV to the extent of 1,600 MW DC has limited impact on the transmission voltages (voltage changes are between 1-4%, which are within the acceptable range). The forecasting model that we have developed is an iterative learning model, which enables future real time data from various geographic areas to be input into the model in order to derive a more precise solar PV location capacity forecast. As demonstrated, the model also enables us to perform contingency analysis to identify potential overload or voltage issues that might need further attention and monitoring.

5.3 Future Work

The forecasting model for solar PV developed in this thesis can be iteratively improved by adding more data to the model. As more solar PV installation data becomes available, the prediction model can be re-trained to improve the accuracy of prediction.

Additionally, our electrical model to represent the distribution network is built on a set of simplifying assumptions. Based on the current transmission and distribution system model used by National Grid and ISO-NE, the simplifying assumptions render the application of our equivalent electrical model to be more practical. However, potential future work on distribution network modeling could be undertaken to simulate a more complex model without recourse to the simplifying assumptions. National Grid is currently working on developing one such model that would be able to provide a detailed representation of the interconnected transmission and distribution network. However the software used for this modeling, Grid Lab D, and the detailed modeling of the network on the software are still in preliminary stages.

Another possibility of future work is to package the model by providing a comprehensive user dashboard such that the application could be extended to other regions such as New York (NY) or Rhode Island (RI).
Chapter 6
Appendix

<table>
<thead>
<tr>
<th>Solar Radiance (kWh/m²/day)</th>
<th>Jan.Min</th>
<th>July.Max</th>
<th>Annual.Average</th>
<th>Annual.Min</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Households</strong></td>
<td>X2010.Population</td>
<td>Land.Sq.Miles</td>
<td>Pop.density</td>
<td>Number.households</td>
</tr>
<tr>
<td></td>
<td>Owner.occupied.households</td>
<td>MeanCostperCap</td>
<td>average cost ($) per kW of capacity installed in the ZIP Code</td>
<td></td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td>some.college</td>
<td>Residents with some college education</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bachelor_degree</td>
<td>Residents with a bachelor degree</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>over50</td>
<td>Number of Residents above 50 years old</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>percent_over50</td>
<td>over65</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Median.age</td>
<td>Mean.age</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Income</strong></td>
<td>MedianIncome</td>
<td>MeanIncome</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Political Orientation</strong></td>
<td>percent.democrat</td>
<td>percent.republican</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>percent.green</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Load Zones</strong></td>
<td>Boston, WesternMass, CentralMass,SEMA, LowerSEMA, Northshore</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.1: Demographic features considered
<table>
<thead>
<tr>
<th>m</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan.Min</td>
<td>3.87</td>
</tr>
<tr>
<td>July.Max</td>
<td>149</td>
</tr>
<tr>
<td>X2010.Population</td>
<td>-0.0144</td>
</tr>
<tr>
<td>MedianIncome</td>
<td>-0.00403</td>
</tr>
<tr>
<td>MeanIncome</td>
<td>0.0028</td>
</tr>
<tr>
<td>Percent.Democrat</td>
<td>0.194</td>
</tr>
<tr>
<td>some_college</td>
<td>-0.0439</td>
</tr>
<tr>
<td>Bachelor_degree</td>
<td>0.0477</td>
</tr>
<tr>
<td>over50</td>
<td>0.112</td>
</tr>
<tr>
<td>percent_over50</td>
<td>1570</td>
</tr>
<tr>
<td>Median.age</td>
<td>-27.9</td>
</tr>
<tr>
<td>over65</td>
<td>-0.122</td>
</tr>
<tr>
<td>Number.households</td>
<td>0.00343</td>
</tr>
<tr>
<td>Boston</td>
<td>1570</td>
</tr>
<tr>
<td>WesternMass</td>
<td>-6.12</td>
</tr>
<tr>
<td>CentralMass</td>
<td>583</td>
</tr>
<tr>
<td>SEMA</td>
<td>152</td>
</tr>
<tr>
<td>LowerSEMA</td>
<td>-92.5</td>
</tr>
<tr>
<td>intercept</td>
<td>114</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\beta_0$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual.Min</td>
<td>0.0876</td>
</tr>
<tr>
<td>Jan.Min</td>
<td>-0.571</td>
</tr>
<tr>
<td>some_college</td>
<td>$1.85 \times 10^{-5}$</td>
</tr>
<tr>
<td>Bachelor_degree</td>
<td>$-2.34 \times 10^{-5}$</td>
</tr>
<tr>
<td>Number_households</td>
<td>$1.4 \times 10^{-5}$</td>
</tr>
<tr>
<td>Owner_occupied_household</td>
<td>$2.45 \times 10^{-5}$</td>
</tr>
<tr>
<td>MeanCostperCap</td>
<td>0.000945</td>
</tr>
<tr>
<td>Boston</td>
<td>-1.76</td>
</tr>
<tr>
<td>WesternMass</td>
<td>0.847</td>
</tr>
<tr>
<td>CentralMass</td>
<td>-0.254</td>
</tr>
<tr>
<td>SEMA</td>
<td>-0.167</td>
</tr>
<tr>
<td>LowerSEMA</td>
<td>1.35</td>
</tr>
<tr>
<td>intercept</td>
<td>-10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\beta$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MedianIncome</td>
<td>$1.8 \times 10^{-8}$</td>
</tr>
<tr>
<td>MeanCostperCap</td>
<td>$-3.9 \times 10^{-7}$</td>
</tr>
<tr>
<td>CentralMass</td>
<td>-0.00283</td>
</tr>
<tr>
<td>intercept</td>
<td>0.0343</td>
</tr>
</tbody>
</table>

Table 6.2: Parameters estimated from formulation 3.5
Table 6.3: Fitted models for farms, C in MW, t in weeks (starting from 2008-1-1)

Quadratic model

\[ C = at^2 + bt + c \]

<table>
<thead>
<tr>
<th>Region</th>
<th>Coefficients</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>Boston</td>
<td>-1.4 (10^{-4})</td>
<td>0.1683</td>
<td>-20.23</td>
</tr>
<tr>
<td></td>
<td>(2.4 (10^{-5}))</td>
<td>(0.122)</td>
<td>(1.44)</td>
</tr>
<tr>
<td>NorthShore</td>
<td>0.15</td>
<td>-27.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7.07 (10^{-3}))</td>
<td>(1.83)</td>
<td></td>
</tr>
<tr>
<td>WesternMassW</td>
<td>0.178</td>
<td>-283</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.391 (10^{-3}))</td>
<td>(1.099)</td>
<td></td>
</tr>
</tbody>
</table>

Logistic model

\[ C = \frac{M}{1 + e^{-(\beta_0 + \beta\ t)}} \]

<table>
<thead>
<tr>
<th>Region</th>
<th>Coefficients</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>(\beta_0)</td>
<td>(\beta)</td>
</tr>
<tr>
<td>SEMA</td>
<td>245</td>
<td>-8.41</td>
<td>0.0257</td>
</tr>
<tr>
<td></td>
<td>(10.2)</td>
<td>(0.11)</td>
<td>(5.78 (10^{-4}))</td>
</tr>
<tr>
<td>CentralMass</td>
<td>267</td>
<td>-7.79</td>
<td>2.29 (10^{-2})</td>
</tr>
<tr>
<td></td>
<td>(17.5)</td>
<td>(9.5 (10^{-2}))</td>
<td>(5.8 (10^{-4}))</td>
</tr>
</tbody>
</table>

Table 6.4: PSS/E Load Flow Analysis Parameters

<table>
<thead>
<tr>
<th>Case</th>
<th>Area Interchange</th>
<th>Transformer LTCs</th>
<th>Phase Angle Regulators</th>
<th>Switched Shunts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>Disabled</td>
<td>Stepping</td>
<td>Regulating</td>
<td>Regulating</td>
</tr>
<tr>
<td>Contingency</td>
<td>Disabled</td>
<td>Locked</td>
<td>Locked</td>
<td>Locked</td>
</tr>
</tbody>
</table>

71
<table>
<thead>
<tr>
<th>Transformer Impedance Calculation</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>1. Rooftop PV Impedance Calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three phase Impedance, Resistance, and Reactance for the 25kVA transformer is:</td>
</tr>
<tr>
<td>Impedance, Z% = 2.60%</td>
</tr>
<tr>
<td>Resistance, R% = 1.60%</td>
</tr>
<tr>
<td>Reactance, X% = 2.10%</td>
</tr>
<tr>
<td>Note: Impedance values are taken from GE transformer ratings [28]</td>
</tr>
</tbody>
</table>

Converting to single phase Impedance, Resistance, Reactance values at 100 MVA base system

| Impedance, Z = (0.026*100*1000/25)*(1/3) = 35 ohm p.u on a 100 MVA base system |
| Resistance, R = (0.016*100*1000/25)*(1/3) = 21 ohm p.u on a 100 MVA base system |
| Reactance, X = (0.021*100*1000/25)*(1/3) = 28 ohm p.u on a 100 MVA base system |

<table>
<thead>
<tr>
<th>2. Solar PV Farm Impedance Calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impedance per 500 kVA transformer is 5%</td>
</tr>
<tr>
<td>Assuming R is 0.01 % of Z, and X is 0.99% of Z</td>
</tr>
</tbody>
</table>

Converting impedance to 1000 MVA base system

| Impedance, Z = (0.05*100*1000/500) = 10.0 ohm p.u on a 100 MVA base system |
| Resistance, R = (0.05*100*1000/500) * 0.01 = 0.1 ohm p.u on a 100 MVA base system |
| Reactance, Z = (0.05*100*1000/500) * 0.99 = 9.9 ohm p.u on a 100 MVA base system |

Each transformer is connected to 1MW of solar array.
Several transformers with 1MW clusters of solar PV are connected in parallel

Resulting Impedance, Resistance, and Reactance are as follows:

| Impedance, Z = 10 / Total_MW_of_Generation ohm p.u on a 100 MVA base system |
| Resistance, R = 0.1 / Total_MW_of_Generation ohm p.u on a 100 MVA base system |
| Reactance, X = 9.9 / Total_MW_of_Generation ohm p.u on a 100 MVA base system |

Table 6.5: Transformer Impedance Calculation for Rooftop PV and Solar PV farms
Bibliography


