THESIS

DISAGGREGATION OF NET-METERED ADVANCED METERING INFRASTRUCTURE

DATA TO ESTIMATE PHOTOVOLTAIC GENERATION

Submitted by

Wendell Jay Stainsby

Department of Systems Engineering

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2019

Master's Committee:

    Advisor: Peter Young

    Daniel Zimmerle
    Patricia Aloise-Young

ABSTRACT

DISAGGREGATION OF NET-METERED ADVANCED METERING INFRASTRUCTURE

DATA TO ESTIMATE PHOTOVOLTAIC GENERATION

Advanced metering infrastructure (AMI) is a system of smart meters and data management systems that enables communication between a utility and a customer's premise, and can provide real time information about a solar array's production. Due to residential solar systems typically being configured behind-the-meter, utilities often have very little information about their energy generation. In these instances, net-metered AMI data does not provide clear insight into PV system performance. This work presents a methodology for modeling individual array and system-wide PV generation using only weather data, premise AMI data, and the approximate date of PV installation. Nearly 850 homes with installed solar in Fort Collins, Colorado, USA were modeled for up to 36 months. By matching comparable periods of time to factor out sources of variability in a building's electrical load, algorithms are used to estimate the building's consumption, allowing the previously *invisible* solar generation to be calculated. These modeled outputs are then compared to previously developed white-box physical models. Using this new AMI method, individual premises can be modeled to agreement with physical models within $\pm 20\%$. When modeling portfolio-wide aggregation, the AMI method operates most effectively in summer months when solar generation is highest. Over 75% of all days within three years modeled are estimated to within $\pm 20\%$ with established methods. Advantages of the AMI model with regard to snow coverage, shading, and difficult to model factors are discussed, and next-day PV prediction using forecasted weather data is also explored. This work provides a foundation for disaggregating solar generation from AMI data, without knowing specific physical parameters of the array or using known generation for computational training.

Most importantly, my nuclear and extended family has always provided a vibrant and positive environment to grow up and live in. My four grandparents instilled into my genes the love of many things that are needed for being a better human being: generosity, academics, reading, open-mindedness, traveling, the University of Florida, camping, and four-legged companions. Our parents raised Aaron and I the right way and have provided a motivating template of how to live a meaningful life. Mom and Dad: I am grateful for your infinite, supportive attitude through my winding journey of this life. My younger and smarter brother Aaron: you have been an inspiration as you've grown up and now as you work through your graduate photonics degree in Germany. I look forward to witnessing your continued accomplishments. Mom, Dad, Aaron, thank you. I love you.

If you are still reading this, you are about to be in the thick of it. Let's go.

DEDICATION

To my Grandparents:

Frances Elaine Wintersteen, Judith Barbara Kasler, Irving Jay Goffman, Wendell Nichols Stainsby

I am inspired to build a better world.

And Toby, of course.

TABLE OF CONTENTS

LIST OF TABLES

viii

LIST OF FIGURES

NOMENCLATURE

NOTABLE ABBREVIATIONS

AC	Alternating Current
AMI	Advanced Metering Infrastructure
AOI	Angle of Incidence
DC	Direct Current
DHI	Diffuse Horizontal Irradiance
DNI	Direct Normal Irradiance
E	Energy
GHI	Global Horizontal Irradiance
GW	Gigawatt
kW	Kilowatt
kWh	Kilowatt-hours
MWh	Megawatt-hours
PVL	Photovoltaic Library Model
TOD	Time-of-Day Pricing
TWh	Terrawatt-hours

VARIABLES

E_a	Extraterrestrial Irradiance (watts/m^2)
E_{qt}	Equation of Time (minutes)
E_{SC}	Solar Constant (watts/m^2)
E_b	Energy, Beam Irradiance (watts per time)
E_d	Energy, Diffuse Irradiance (watts per time)
E_g	Energy, Ground Reflected Irradiance (watts per time)
E_{POA}	Energy, Plane of Array (watts per time)
k_t	Clearness Index
k_d	Diffuse Fraction
P_{AC}, P_{DC}	Alternating/Direct Current Power (watts)
T_{solar}	Solar Time (minutes)
$T_{correction}$	Temperature Correction Factor
T_C	Temperature Coefficient
T_M	Module Temperature (°C)
WS	Wind Speed (m/s)

GREEK AND LATIN

P	Estimated Delivered Energy
Δ	Delivered Energy
Π	Received Energy
A	Delivered Energy After PV
σ	Standard Deviation
ε	Absolute Percentage Error
ω	Weighting of an Error
θ_Z	Solar Zenith Angle
θ_T	Array Tilt Angle
θ_A	Solar Azimuth Angle
θ_{array}	Array Azimuth Angle
θ_{hr}	Hour Angle
Γ	Relative Percentage Error
η	Efficiency
λ	Latitude
γ	Fractional Year

SUBSCRIPTS AND SUPERSCRIPTS

$\overline{x}$	Mean
$\hat{x}$	Median
x_{max}	Maximum

Chapter 1

Introduction

1.1 Motivation

By 2019, the U.S. had 70 GW_{DC} of solar capacity with around 20% of that being residential rooftop solar [1, 2]. Between 2014 through 2018, the compounded annual growth rate of small-scale (less than 1 MW) photovoltaic (PV) generation in the United States was 28.2% [3]. The 2 millionth solar array was installed in early 2019 and within 5 years over 2.5% of all U.S. homes will have a solar installation [4, 5]. Additionally, the National Renewable Energy Laboratory (NREL) has found that the technical solar potential of suitable residential buildings was over 700 GW_{DC}, which could generate 926 TWh/year, a quarter of the United States' total annual electricity consumption [6].

The growth of solar has increased interest in better estimates of real-time generation from solar arrays for use by distribution operators to manage grid resources [7]. Larger commercial and utility-scale solar deployments are usually directly metered, providing actual generation typically at high time resolution and minimal latency. Thus, both system owners and electrical system operators normally know the actual energy produced by large solar arrays. These data can be used to accurately simulate expected generation.

However, residential solar systems in the USA are typically net-metered by the distribution operator, as 38 states, Washington D.C., and four overseas territories have net-metering policies by 2019 [8]. In traditionally metered systems with PV, system operators have access only to net load for the billing period, obscuring any information about the PV generation. When available, digital smart meter data generally reports only net-load or net-generation (generation net of load) for each meter reading. Since generated PV offsets concurrent load in a net-metered building, it is difficult to separate behind-the-meter (BTM) generation from load. Therefore, distribution utilities are essentially blind to PV production at the vast majority of residential solar sites around the

country. This lack of insight into solar generation complicates solar forecasting and the balancing of dispatchable generation with demand [9].

To this end, input-intensive physical models have been developed but require intimate knowledge of each system, which a utility may not have. The difficulty in modeling a portfolio of residential assets with physical-models lead to work investigating the estimation of PV output based on net-metered smart meter data.

1.2 Advanced Metering Infrastructure

Advanced metering infrastructure (AMI) is a system of communication and data collection between a building's electric service infrastructure and the providing utility. By 2017, there were nearly 80 million buildings with AMI in the United States, and approximately 88% of these were installed in residential buildings [10].

A major component of AMI is the smart meter, which enables access to smart grid services as a two-way interface between the utility and building [11]. These smart grid services include demand response, dynamic pricing, and home energy management with generation or storage. Wired or wireless communication systems in the AMI connect the smart meters with a central utility provider platform that manages storage, two-way communication, and price signals [12]. Typically, the smart meters record electricity consumption at 5-, 15-, 30- or 60-minute intervals, but meters recording at finer temporal resolution do exist [11].

Utilities are installing AMI because they reduce operating costs for metering and billing. Manned sorties to read analog meters are no longer necessary, saving on labor and fuel costs [11]. Smart meters also allow for remote control of connection/disconnection in much quicker time than deploying someone to the field [11]. Smart meters also provide outage detection and voltage monitoring. Although push-back from some citizens over the installation of smart meters revolves around the perception of risks to health and data privacy, opt-out rates are generally low [11].

The growth in smart meters has led to huge amounts of high-granularity electric consumption data, leading to research in areas that was previously more difficult or not possible. As [13]

notes, the increase in large AMI datasets has led to new work involving descriptive, predictive, and prescriptive analytics. Load forecasting for better asset management and distribution-grid operation can be enabled by this mass of information. Premise level information can be aggregated to feeder- or substation-level to provide a bottom-up approach of load analysis and forecasting. As smart meter penetration increases, valuable insight will be gained through ever expanding energy data analytics.

As mentioned, the nature of net-metered systems does not easily allow for the measurement of distributed generation (DG) assets such as rooftop solar. Research has been conducted to attempt to disaggregate generation from AMI data; these methods and others will be discussed in Section 1.3.

1.3 Current Predictive Methods

1.3.1 Machine Learning

Several authors have proposed machine learning (ML) methods to estimate PV production [14–20]. These techniques rely on training a model using separately metered PV and/or total load data from residences, or use PV production from nearby metered arrays as a proxy for residential PV production [7, 21]. One ML program, DeepSolar, even identifies and learns solar panel locations and sizes from satellite imagery data [22]. Others use weather-based models or require known generation [23, 24], or are hybrid models that use both physical and ML approaches [25, 26].

The array configuration data required for each system estimation varies between methods, from no data to specific physical and site parameters. Several approaches report root mean squared errors between 20%-50%, and most require access to data from actual or representative systems in which PV generation is metered independently from load.

Shaker et al. [27] suggest using the output from a small number of known PV arrays to extrapolate power production across a greater number of BTM solar systems within the same area. However, this type of modeling requires the utility to know the total capacity of residential sys-

tems within the area, and implicitly assumes that the small number of monitored arrays represent all variation in system configurations for the unmonitored arrays.

1.3.2 Physical Models

Directly estimating generation for a solar array is also complex. Modeling requires numerous variables, including weather, physical array orientation, and system and electrical characteristics. In addition, variable external factors such as seasonal tree shading, snow coverage, ground reflectance, and panel soiling also impact power output. If these data are available, there are well established "white-box" models that utilize these physical inputs with current or forecasted weather to estimate solar generation. These tools are often utilized for estimating the output of large solar arrays and have achieved accuracy to within $\pm 10\%$ of the actual generation of various solar arrays studied [28].

Two of the U.S. Department of Energy's National Laboratories have developed open-source physical PV performance tools. A well-known model has been developed by Sandia National Laboratories (SNL), and SNL has a long history of developing PV performance models starting with PVFORM in the 1980's [29]. This original, simplistic model used typical meteorological year (TMY) solar and weather data and incorporated module operating temperature [30]. In 1994, SNL presented a physical model that has been updated and validated over the course of 25 years using countless measured datasets across the spectrum of commercially available PV modules products [30]. This model and related work has grown into the PV Performance Modeling Collaborative (PVPMC), a collaborative group of industry professionals facilitated by SNL. The PVPMC contains a PV Library, allowing open-access to the document library and software code for the modeling tool [31].

Another popular interactive modeling tool is PVWatts developed by the National Renewable Energy Laboratory. PVWatts is geared to providing aggregated, long-term estimates of energy production, not to estimate exact production over hourly or daily periods. Released in 1998, the original version was largely based on the PVFORM model from Sandia [28]. PVWatts is a tool that

uses several assumptions about system losses to limit complexity for user-friendly operation [28]. Because PVWatts aims to provide long-term estimates, it uses TMY weather data - output is a *representative estimate* of actual system performance [28]. NREL's Version 5 manual states up to $\pm 10\%$ error for annual energy totals and up to $\pm 30\%$ error for monthly totals due to variation from using historical typical weather data may occur. While useful for monthly or annual insight, the tool is not effective for finer resolution modeling; this is consequential as it would not be appropriate for hourly or daily modeling.

Both models use a minimum number of inputted PV system specifications such as geographical coordinates, size, tilt, azimuth, losses, and component efficiencies. PVWatts does have default values for every input parameter (even assuming a residential system to be a 4 kW array facing due South) and only truly requires a location from geographic coordinates [28]. However, for useful model output, very specific and correct system information is required.

1.3.3 Disaggregation from AMI Data

Although physical models have achieved reasonable accuracy, the inputs and assumptions can be onerous and their complexity cumbersome to the end-user. As noted in [18], distribution utilities often do not have access to this specific data for each residential array, either because the data were not collected during the permitting process or because the *as installed* configuration of the system deviates from the reported configuration. In addition, system or site modifications over time may also impact PV production.

The difficulty of modeling a large portfolio of net-metered residential assets using physical models stimulated interest in estimating PV output based on smart meter (AMI) data. Disaggregating PV generation from the net-generation/net-load measurements recorded by the smart meter is the crux of this type of estimation.

Tabone et al. built an open-source model that aggregated load as a function of hourly ambient temperature and time of day and utilizes a nearby irradiance proxy such as known PV generation [7, 21]. In this work, one year of the Pecan Street household-level generation and total load

data set from 52 households with PV were used, which is not a typical AMI dataset. The actual PV generation from each building was known and subtracted from the observed building load to calculate net-load at the meter. Three known array generation data sets were used as a solar resource proxy to train the model on 8 houses over the span of 10 days and this method led to a root mean squared error between 20-50% for average daily PV generation.

SunDance was developed by Chen and Irwin to disaggregate solar generation data from net-metered data without knowing the array's actual generation for use as training data [25, 26]. The method uses ML to infer the tilt, azimuth, and overall system efficiency from the AMI data to build a custom model of maximum *clear sky* solar generation and maps weather data to an expected reduction from the maximum solar potential during a clear sky day. This model still uses the underlying physical characteristic relationships to estimate maximum generation potential and subtract reduction from each time step. Hybridizing a physical modeling approach with inputs estimated from machine learning on net-metered data achieves a similar accuracy as other ML methods that use solar generation training data. The resulting mean absolute percentage error (MAPE) for SunDance is between 20-40% on 100 premises analyzed.

In a related method to SunDance, SolarCast utilizes historical generational data and array location, and does not require specific physical orientation and site information [18]. This "black box" model estimates tilt, orientation, and external parameters such as shade, snow cover, and soiling through automated ML. SunDance then uses adaptive learning to predict the power output from forecasted weather. 116 houses were analyzed, with MAPE's also generally spread between 20-40%.

Yang et al. created a hybrid ML method using weather data and observed PV generation [14]. Historical PV generation is classified via observed weather, then ML is used to train the weather-generation sets based on similar days. These similar days are based on 3-hour irradiance, maximum temperature, and probability of precipitation. A fuzzy inference method is used to forecast next-day generation based on forecasted weather as well as categorical weather descriptions. While

6

estimation errors on periods using observed weather are below 10% from actual generation, this method requires large amounts of actual PV generation and is modeled from a single 5 kW$_{DC}$ array.

Wang et al. [15] also decomposed load and PV output using historical weather data with estimated equivalent capacity, tilt, and azimuth of a representative system. This model then uses an artificial neural network to map between day type and weather. The model uses synthetic actual load and PV generation data from an NREL tool to verify performance. The drawback is that individual or groups of PV systems are not modeled, rather MAPEs (2-22%) for PV penetration scenarios (varying from 0-20% penetration) are compared to other methods.

These antecedent works have proposed AMI disaggregation but these methods are hindered by their data requirements. Tabone et al. [7] uses atypical AMI data from separately metered building and PV from a publically available dataset for model training. SolarCast by Iyengar et al. [18] requires at least one month of historical generation data for each premise, which is not realistic for a portfolio of PV systems.

Although Wang et al. [15] uses a representative PV proxy, the work is of particular interest due to the mapping of categorical and numerical variables such as day type (week day or weekend), hour of day, month, week, and ambient air temperature. Like others, Yang et al. [14] uses actual generation for training, and only uses one PV array for forecasting generation. However, Yang's work does open to the door to modeling PV generation based on similar periods of time, a method which will be leveraged in a new AMI model described in this work.

1.4   Objectives and Overview

As discussed in the review of PV estimation methods, there are shortcomings from these countless modeling approaches. The key challenges with these proposed methods include *a)* acquiring long time series of local generation data for training, *b)* understanding if the solar resource used for training is representative of current and future PV installations, *c)* acquiring the skills to train the model - likely a challenge for most distribution utilities, and *d)* depending upon the method, having reasonably accurate records of the location, size and as-built physical characteristics of individual

PV arrays. Many of the models described were developed and tested using only small data sets, including a single 5-kW array [14], 8 houses over the span of 10 days [7], 15 days of training data from one PV system [16], or only superficially describe training methods and data sets [19]. Another method requires metered PV output from over 300 sites [24], an arduous data set to replicate. Because these works use uncommonly available data sets and typically only train and/or model from a few PV systems over short periods of time, further work to improve net-metered PV modeling is appropriate.

This work proposes an alternative method that will estimate PV production using open-source software and three data sources that are commonly accessible by distribution utilities: *a)* historical and, if desired, forecasted weather data, *b)* net-metered, sub-hourly, AMI data, and *c)* the approximate month of PV system installation. This method does not require training using known PV generation, nor does it attempt to estimate array-specific [25] or a representative system's geometry [15] as attempted by others. Using the model as a primitive forecasting tool for next-day generation will also be explored. This approach is deployed to estimate individual system and aggregated generation using 1-4 years of AMI data from nearly 850 net-metered residential systems. By using the known basic configurations of these systems, the method's performance is tested with a physical model to compare generation estimates.

The work is organized as follows: Chapter 2 briefly provides a detailed description of the data sources used for this analysis. Chapter 3 explicates the methodology of building the AMI model. Chapter 4 presents necessary first experimental results. Chapter 5 describes the process of building a validation tool based on a physical model. Chapter 6 presents computational results, figures, and statistics of the work and discusses the impact and validity of the findings and method employed. Chapter 7 briefly discusses the model projecting generation from next-day weather forecasts. Chapter 8 concludes on the current AMI model, and offers a recommendation for future work.

Chapter 2

Description of Data Sources

For this analysis, we utilize data from Fort Collins, Colorado, USA, a high-elevation community in the central United States with highly variable weather conditions. Electricity service is provided by the City of Fort Collins Municipal Utilities (FCU), a city-owned municipal distribution company with a sole-provider contract with Platte River Power Authority (PRPA). PRPA generates and transmits power to FCU infrastructure for distribution.

2.1 AMI Data

Fort Collins Utilities received $18M from the U.S. Department of Energy through the American Recovery and Reinvestment Act of 2009 to deploy smart meters city-wide and expand SCADA and automation capabilities in the distribution system [32]. Beginning in 2014, FCU installed AMI meters on existing premises, and the AMI system become operational at the beginning of 2015 [32]. FCU had approximately 76,000 premise smart meters in January 2019. The AMI meters report on 15-minute intervals, and all residential PV systems are net-metered. FCU provided all gathered AMI data to Colorado State University (CSU) researchers - a total of approximately 11 billion meter readings. Four data fields were of interest for this research:

1. Premise identification number

2. Timestamp

3. Delivered energy (kWh)

4. Received energy (kWh)

Each premise has a unique ID number, which can be tied to geospatial and housing data. A timestamp is recorded in 15-minute increments and has a timezone offset from Coordinated Universal Time (UTC). Fort Collins is located in the Mountain time zone, this offset oscillates between

7 and 6 hours from UTC based on Daylight Saving Time. The timestamps are coded at the end of each interval, meaning a timestamp of 16:15 represents the power flow read by the meter between 16:01 and 16:15 in the afternoon. Delivered energy refers to energy provided to the building by the grid during the 15-minute interval while received energy refers to energy exported by the building back to the distribution grid during the time interval, both expressed in kilowatt-hours (kWh). Only premises that have generation capabilities such as rooftop solar should ever have received values greater than zero. For premises with PV, received and delivered values can be present in the same timestamp as power flow can rapidly change direction (e.g. generation drops due to moving cloud, or consumption increases from an appliance that is turned on). As discussed in the previous chapter, measured received energy values greater than zero only occur when the building's consumption is less than the generation, exporting excess power back to the distribution grid.

Figure 2.1 below illustrates the recorded AMI data. Solid boxes represent metered readings and dotted boxes represent unknown values.
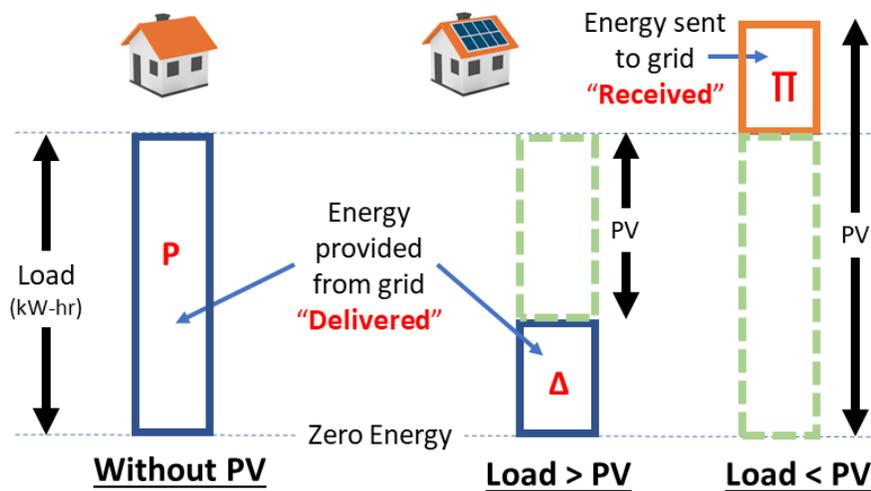
Figure 2.1: AMI Data Visualization. Box P represents the delivered energy to the premise before PV installation. Box Δ represents energy delivered to the premise when some load is offset by unmeasured PV generation (dashed green boxes). Box Π represents excess generated energy received by the distribution grid from the premise when generation exceeds load.

Figure 2.2 displays ten days of energy consumption of a residential home before solar was installed. Once a solar array is installed, energy flows bi-directionally across the meter. Figure

2.3 exhibits this interaction: generation flowing back to the distribution grid (orange), and energy consumed from the grid (blue).
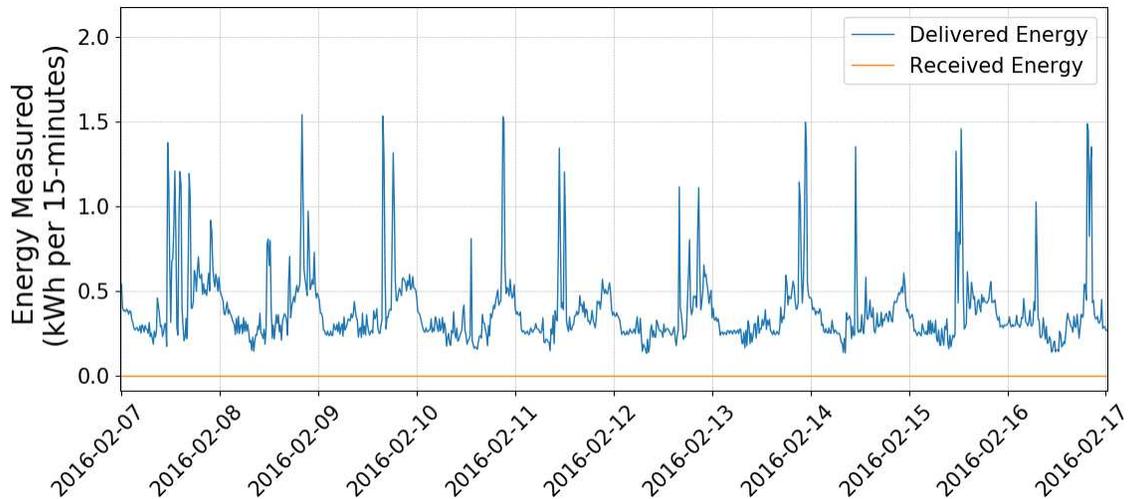
Figure 2.2: Pre-PV AMI Data. Ten days of energy consumption from a residential premise before rooftop solar was installed. As expected, received energy measurements are consistently zero. Late afternoon/evening peaks are visible.
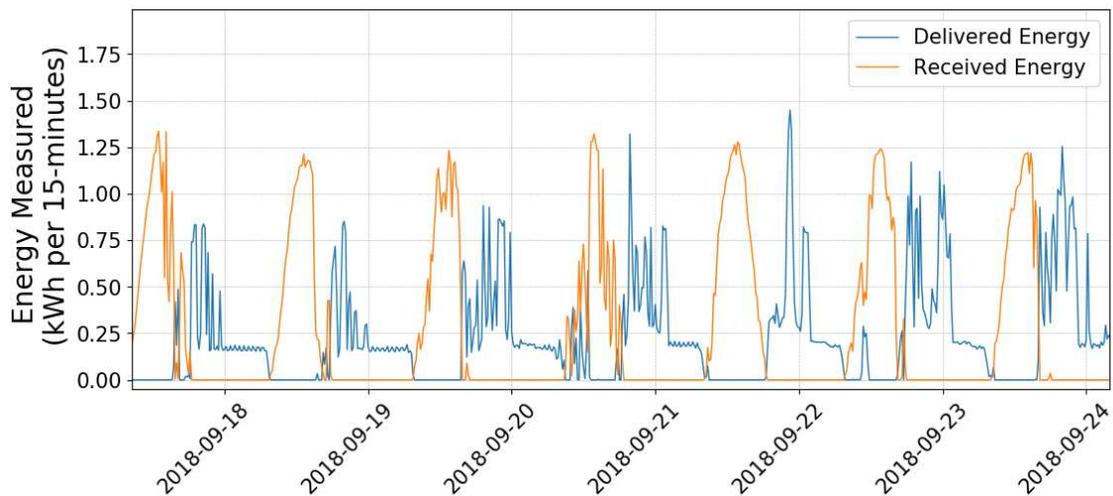
Figure 2.3: One week of AMI Data with PV. Displayed for a residential premise with rooftop solar. During days of good weather, there are clear solar production curves that are immediately followed by the predictable evening spikes in consumption.

Figure 2.4 portrays one day of energy consumption and generation for a residential premise with solar. The conundrum of invisible generation is observable around mid-day where generation and consumption nearly cancel each other out in the AMI data.
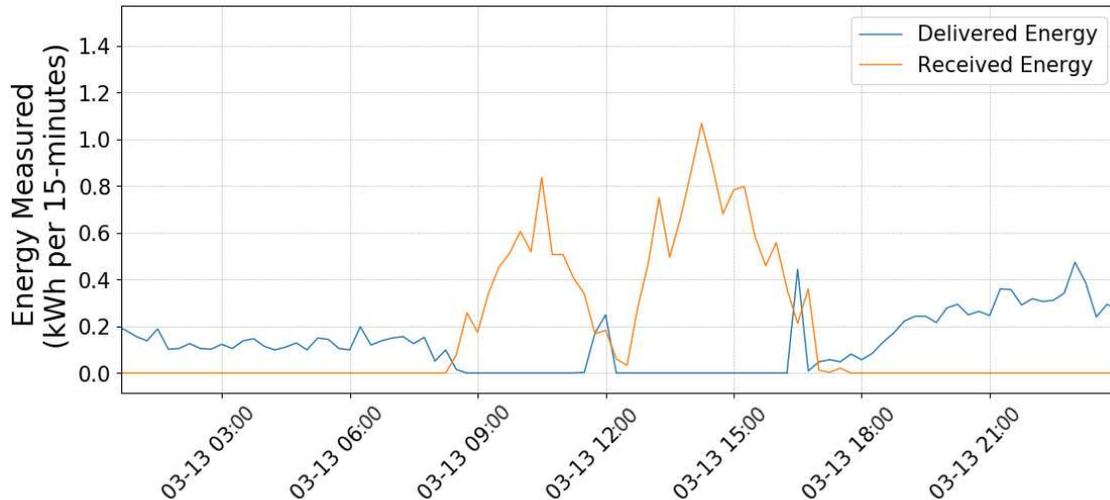
Figure 2.4: Invisible Generation in AMI Data. 24-hours of 15-minute AMI data for a residential premise with rooftop solar. As is evident by the dual-peaks of PV generation being exported back to the grid, something occurred during the middle of the day. One scenario is that cloud cover significantly reduced generation, leading to energy consumption from the grid (around noon in blue). Another scenario is that load increased greatly, consuming most of the generation behind-the-meter, and decreasing exported energy mid-day. Due to the nature of net-metering, it is impossible to know which case or combination occurred solely based on the AMI data.

2.2 Weather Data

High granularity weather data were downloaded from an internet platform compiled by the Department of Atmospheric Sciences at Colorado State University. This Department operates a weather station located on the main campus of CSU in Fort Collins adjacent to the Lory Student Center - a reasonably central location to FCU's distribution system [33]. The weather station provides 10-minute measurements of temperature, global horizontal irradiance (GHI), relative humidity, wind speed and direction, air pressure, and soil temperature [33]. GHI is the total solar irradiance arriving on a horizontal surface on Earth. For the scope of this work, temperature (°C),

wind speed (m/s), and GHI (watts/m^2) were obtained from January 1, 2015 through the end of March 31, 2019.

Fort Collins snowfall data was obtained from the National Oceanic and Atmospheric Administration's Climate Data online portal [34]. The database provided daily snowfall (inches) and daily accumulated snowfall (inches).

2.3   Distributed Generation Assets

Fort Collins Utilities also provided records of each photovoltaic system installed in the City's distribution network. The *DG Aassets* provided the following information as shown below:

Table 2.1: Distributed Generation Assets. Fields of information about PV systems interconnected in the City.

Field	Units
Premise ID Number	
Installed Capacity	kilowatts$_{DC}$
Date of Installation	
Inverter Manufacturer	
Inverter Model/Quantity	
Array Azimuth	degrees
Array Tilt	degrees
Billing Rate Class	

From 1987 though Q1 2019, 1,389 PV systems, totaling a capacity of 13.1 MW$_{DC}$, were installed within the City. Of these, 1,298 (93%) PV systems were installed on residential premises, with an average residential system capacity of 5.5 kW$_{DC}$. Residential premises accounted for 7.3 MW$_{DC}$ of installed capacity, approximately 55% of total installed capacity within the City. Around 285 older systems, mostly installed before 2015, contained missing array tilt and/or azimuth values in the City's records. Using Google Earth and Google Street View, these values were estimated and replaced the vacant values. Figure 2.5 below illustrates the growth in solar in Fort Collins, Colorado.
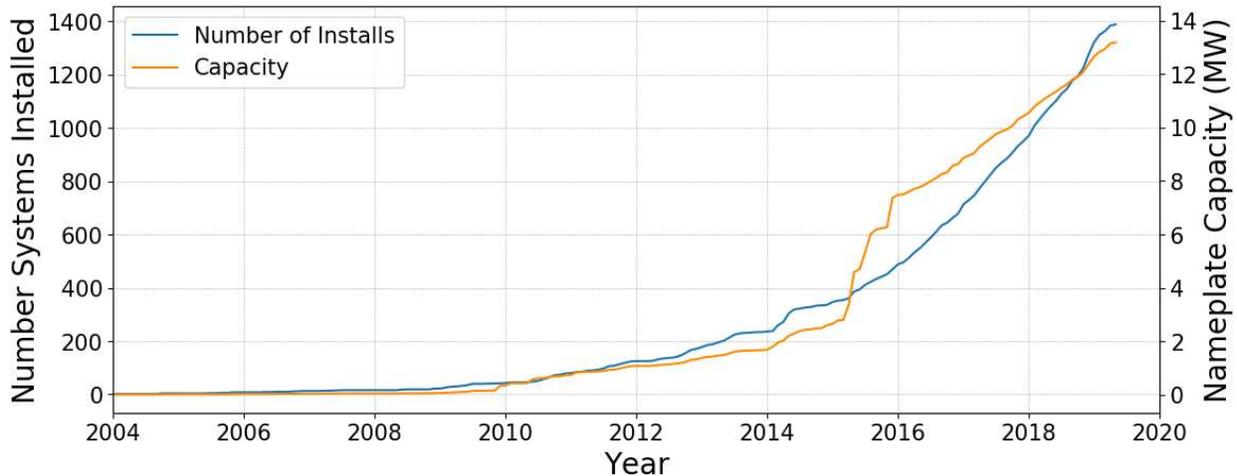
Figure 2.5: Historical PV Assets in Fort Collins. Cumulative PV assets in Fort Collins with cumulative installed capacity (megawatts$_{DC}$).

2.4 Privacy

AMI data is very sensitive because it can provide specific information about a consumer's energy usage. Behavioral patterns such as daily schedule, number of occupants, types of electronics in the premise, or even the use of medical equipment may be discernible from the data [35]. Colorado State University and the City of Fort Collins entered into a data sharing agreement that specified privacy considerations and managed the terms of any potential data sharing outside of the University's research team.

The *DG Assets* data was also treated as sensitive and not shared outside of the research group. It should be noted that the City has a transparency initiative with an *Open Data* platform, and detailed data about buildings with PV is accessible online [36].

Chapter 3

AMI Model Methodology

All work presented in this thesis was performed at the Powerhouse Energy campus of Colorado State University. The Powerhouse hosts a variety of research initiatives, mostly centered on clean and efficient power generation and use. This chapter explicates the AMI model methodology, Chapter 4 provides direct experimental results, while Chapter 5 describes the construction of a validation tool to attempt to answer the research objectives. With this tool, Chapter 6 presents computational results and discusses the findings. Finally, Chapter 7 briefly addresses the AMI model to forecasting next-day load and PV generation.

The modeling work was written in Python utilizing the open-source integrated development environment PyCharm. The Pandas software library for python was heavily utilized for data manipulation and analysis, and Matplotlib was used for plot-making. Other open-source software will be mentioned in proceeding sections.

3.1 Basic Framework

The underlying assumption behind this model is that given similar conditions, electricity usage in a building does not change after a PV system is installed. From this, based on comparable points in time, the difference in delivered energy between pre- and post-PV periods can be calculated as the assumed offset from behind-the-meter PV generation. Factoring out sources of variability in electrical load by the comparison of times that have similar conditions enables this analysis.

Addressing and discussing the use of this assumption is necessary. Prior work has shown that some segments of the PV-adopting population actually slightly increase their overall consumption after system installation [37]. Recent work studying Arizonan solar owners found that overall consumption actually rebounded by up to 18% higher than pre-PV consumption in summer months. It should be noted this specific work was conducted in a region with likely higher cooling and lower heating needs. The study also found that consumers in more liberal areas (like Fort Collins)

experience lower or no rebound effects, which is consistent with existing literature [38]. Qui et al. found that more environmentally-aware adopters, identified by progressive voter registrations as a proxy, had a slight decrease (up to 6%) in overall consumption after system installation [37]. It should also be noted that unknown efficiency or appliance upgrades may also skew results from studies on household consumption that have previously installed solar [37]. Thus, literature shows that consistent consumption patterns before/after PV installation are sometimes (but not always) present. However, for this research, consistent consumption is assumed to be the case for solar adopters in Fort Collins, Colorado.

Weather impacts both building energy demand and solar generation. In buildings there is temperature-dependent energy demand driven by heating or cooling [39], and periods of similar temperature should have similar heating/cooling loads [40]. Additionally, solar irradiance is the energy source for a photovoltaic cell and available solar resource dictates the cell's output. For these reasons, temperature and GHI are considered in the pairing of timestamps for reconstructing building load.

The basic approximation from this model for PV generation is $P - \Delta + \Pi$ for any time duration based on Figure 2.1. The values (Δ) and (Π) are measured and recorded in the AMI data. It is up to the model's algorithm to match similar timestamps from pre- and post-PV periods, which are used to calculate P, the estimate for total load for the building. Reconstructing this baseline building consumption (total load) from net-metered AMI data represents the core contribution this work.

Weather information is inputted into the model to build a dictionary containing the target timestamps and their comparable timestamps. A Python dictionary is a data structure used to store key, value pairs, which allows for very efficient lookup of the values based on the key. In this case, the keys are target timestamps, and the values are the comparable timestamps for the target.

The model then queries AMI data for the premise being analyzed, and uses the approximate date of PV system installation to separate the AMI data into pre- and post-PV groups. Finally, a series of algorithms attempts to estimate the building load from the groups of AMI data, using

16

timestamp keys to isolate the comparable periods of AMI data. From this, PV generation is then estimated, and Figure 3.1 below depicts the block diagram of this methodology.
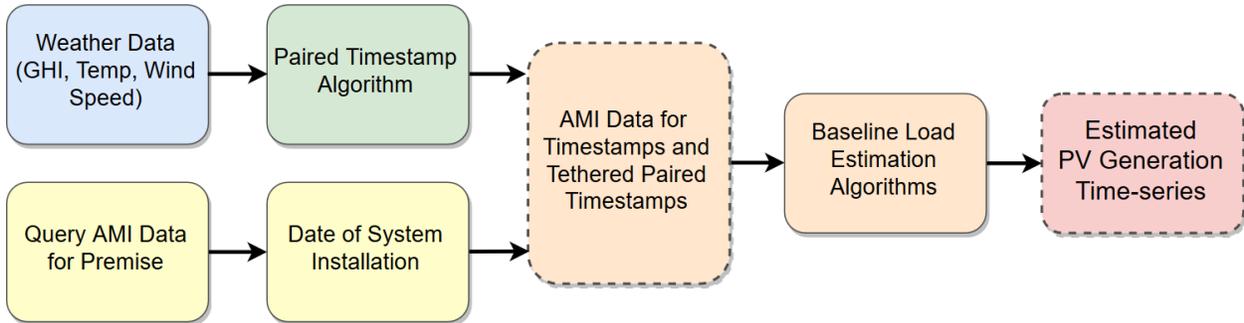
Figure 3.1: Block Diagram of AMI Model.

3.2 Paired Timestamps

First, the 10-minute weather data was interpolated to 15-minute granularity in order to eventually pair with the corresponding AMI timestamp. After interpolation, each timestamp contained the date, hour, minute, measured GHI, wind speed, and temperature for the 15-minute interval.

For computational efficiency, PV modeling occurs during the periods when the sun is above the horizon. Only timestamps in which the sun is at least one degree above the horizon are included in the modeling. These timestamps are identified by a sequence of solar position algorithms that will be discussed in Section 5.2 and Appendix B.1.

Comparable timestamps must be within 15 days of the day of year of the target timestamp. This creates a 30-day range from which the rest of this timestamp algorithm uses. This range contains days that are similar in seasonal weather, with similar lengths of day and solar patterns.

From this 30-day subset, the following requirements are added:

1. The target timestamp occurs either during a weekday or weekend. Comparable timestamps must be the same as the target timestamp, whether a weekday or weekend.

2. Comparable timestamps must be within +/- 4 hours of the target timestamp's hour of day (0-23).

3. The sample standard deviation (σ) of the temperature and irradiance are calculated from the subset.

4. The temperature of each potential comparable timestamp must be within $0.3 \times \sigma$ of the temperature subset.

5. The irradiance of each potential comparable timestamp must be within $0.4 \times \sigma$ of the irradiance subset.

   Steps 4 and 5 are designed to remove timestamp outliers based on weather parameters.

6. Finally, major American holidays are removed as they occur during each year from 2015-2019. Holidays are removed because they represent days in which large parts of the population are not in their normal daily routine. Some businesses are closed during the week, and residential consumption may be dissimilar from a typical day. The removed holidays are New Year's Day, Memorial Day, Independence Day, Labor Day, Thanksgiving (both Thursday and Friday), Christmas Eve, Christmas Day, and the day after Christmas (Boxing Day).

The period of analysis from 01/01/2015 through 03/31/2019 contains over 72,000 daylight timestamps. Upon completion of the comparable timestamp algorithm, the median and mean number of similar timestamps for a given target timestamp was 118 and 129, respectively. The range of the number of comparable timestamps was 0 to 460. Cases with less than 3 comparable timestamps for a target timestamp are addressed in Subsection 3.3.3. Figure 3.2 visualizes the matching of comparable timestamps.

3.3   Model Structure

3.3.1   Query AMI Data

Using pgAdmin4, an open-source and easily usable PostgreSQL database management tool, the AMI data for a premise with PV was queried from the secure database hosted at CSU. The
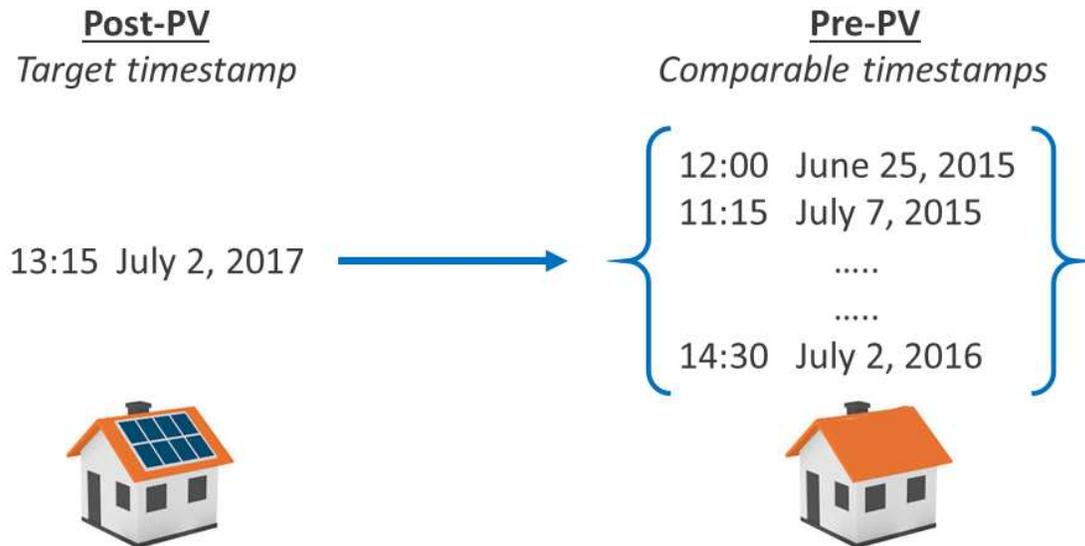
Figure 3.2: Comparable Timestamps Visualization. The pairing of comparable timestamps based on similar weather and temporal factors. This theoretical PV system could have been installed in early 2017. To re-create building load, post-PV installed timestamps are matched with a set of similar timestamps, only which occurred before the building had solar. Using the pre-PV timestamp data enables the total building load to be estimated for the post-PV timestamp.

SQL query is keyed off the premise ID number, and the queried data is ordered by ascending timestamp.

This AMI data is then divided into pre- and post-PV AMI datasets based on the commissioning date of the system from the *DG Assets*. A 20-day buffer in each direction around the date of installation is added to eliminate uncertainty in the exact day the system became operational. This ensures the system was not producing energy during the pre-PV period and was functioning during the post-PV period, or at the very least *could* be functioning. The algorithm does not correct for the possibility that a system is offline for any reason post-installation. Indeed, this is a generally a strength of the algorithm if it were utilized for an extended period: more matching days before/after installation would eventually compensate for a system that was taken offline permanently, whereas a physical model assumes the system is operating at all times.

The 20-day buffer is required to ensure that there was no PV production during the pre-installation period. Figure 3.3 below illustrates the necessity of addressing this potential issue by

showing a system with an incorrect commissioning date. Having a cushion safeguards the model from analyzing incorrect energy consumption patterns.
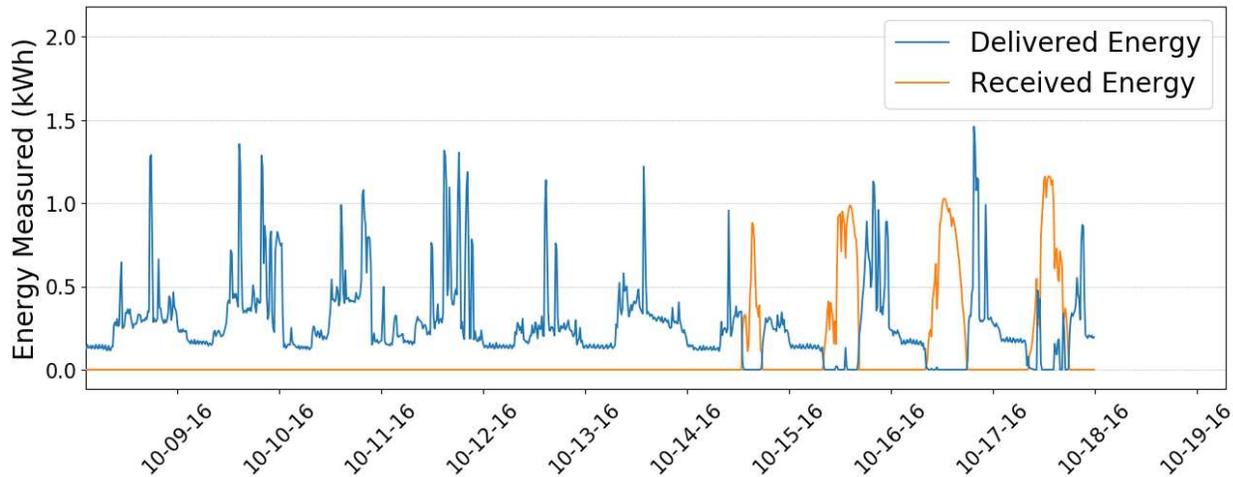
Figure 3.3: Inaccurate Commissioning Date. According to City records, the system was officially commissioned on October 28, 2016. However, plotting the *pre-PV* delivered and received AMI data displays generation starting on October 14, 2016.

House G, with one of the sets of known generation data that will be explored and utilized later, also exhibits this incorrect commission date incident. Records establish the system install date on August 29th, however received energy is recorded beginning on August 8th of the same month. Recorded commissioning errors can exist in both directions, thus the 20 day buffer before and after the recorded commissioning date. It is not known if these commissioning date errors only occur in one direction (systems operating before official commissioning date), therefore a buffer in both directions was used.

3.3.2   Calculating Baseline Energy Load

The sets of pre- and post-PV installation AMI data allow the estimation of pre-PV baseline load, box P in Figure 2.1.

Each target timestamp contains a set of comparable timestamps; using the date of install with a 20 day buffer in either direction, only comparable timestamps with pre-PV AMI data are grouped into a set of timestamps known as *comparable timestamps before PV*, referred to as P . The

target timestamps only ever represent times after the system is operational and the comparable timestamps only ever represent times before PV was installed.

Median and mean energy delivered to the premise during similar points in time are calculated from P, and symbolized as P_{med} and $\overline{P}$. The energy received by the grid (Π) during the *comparable timestamps* is also obtained. This energy should always be zero because there is no generation at the premise before the PV system is installed. This data check can be used to identify an incorrectly recorded date of installation or malfunctioning smart meter.

Additionally, the quantity of corresponding P is obtained for each target timestamp. If a target timestamp is tethered to less than three *comparable timestamps*, the aforementioned P_{med} is calculated as the prior target timestamp's P_{med} value. The prior timestamp is defined as $timestamp_{t-1}$ and represents the 15-minute period preceding the target timestamp's interval.

Another term calculated from the AMI data will be discussed in the following section and is used in irregular conditions. *Comparable timestamps **after** PV*, written as A, are the group of comparable timestamps tethered to the target timestamp that occur after solar is installed. The median delivered energy value from A is calculated as A_{med}. If there are no post-PV comparable timestamps, A_{med} is null.

Here, all the inputs that are needed to calculate the projected energy output from the PV system for each 15-minute period are available. The baseline equation for estimated generation (kWh) is:

$$\text{Gen} \; = \; P_{med} \; - \; \Delta \; + \; \Pi \tag{3.1}$$

Where P_{med} is the Pre-PV$_{Delivered}$ median, Δ is the delivered energy for premise at timestamp t, and Π is the received energy from premise at the timestamp.

3.3.3 Irregular Conditions

Equation 3.1 is unable to logically calculate PV generation for all timestamps. There are numerous irregularities that necessitate amended methods of calculating the PV generation for a

given timestamp. For example, if there are less than three comparable timestamps associated with a timestamp, the delivered and received data from the previous timestamp is used in substitution.

The primary irregular condition is when Δ is greater than the P_{med} term. Intuitively, energy consumed from the grid will *generally* be lower after a premise has installed PV. However, there are instances where occupancy or usage lead to post-installation timestamps with large amounts of grid consumed energy. This scenario leads to a negative first term in Equation 1 that reduces generation from the Π energy value. If the difference between P_{med} and Δ is greater than the value of Π, a negative value will be calculated for PV generation. Logically, negative PV generation does not exist, thus this condition must be addressed.

Analysis was conducted on the frequency of timestamps where actual delivered energy (Δ) was greater than the P_{med} (estimated building load). Calculations from 846 residential premises spanning 1-3.25 years show the average and median percentages of these irregular timestamps are 8.3% and 7.7%, respectively.

Steps are taken in the following series of estimation equations to handle these situations, and are ordered from least to most proffered method of calculation. Additionally, if there are less than 3 comparable timestamps, the previous timestamp's value for P_{med} is used.

The process for calculating the estimated energy output begins at a default condition of zero kWh for the timestamp and progresses through four calculations. A generation estimate of zero is clearly the least desirable method of prediction while method (**d**) is the most desirable equation. Generation is estimated to be:

a) If $\Pi > 0$:

$$\Pi \tag{3.2}$$

b) If $P_{med} > A_{med}$:

$$P_{med} - A_{med} + \Pi \tag{3.3}$$

c) If $\overline{P} > \Delta$:

$$\overline{P} \: - \: \Delta \: + \: \Pi \tag{3.4}$$

d)   If $P_{med} \: > \: \Delta$:

$$P_{med} \: - \: \Delta \: + \: \Pi \tag{3.5}$$

These equations are used to calculate the estimated energy generation for each target timestamp that occurs after a PV system is installed at an individual premise.

3.3.4   Snow Losses

Unlike a physical model, no additional intervention is needed for handling snowfall on PV panels. Because the AMI model utilizes similarities in building energy load before and after PV installation, a lack of generation from snow would innately be carried in the AMI data. Referring to equation 3.1, the model assumes that building load for comparable periods of time will be similar to a given timestamp's energy load. If snow is covering the array, the delivered energy would then be similar to the delivered energy from the comparable periods before the premise had PV. The difference in these terms would be small; adding the near-zero received (produced) energy would result in a low estimate for PV generation. Model prediction performance during periods of snowfall will be explored in Section 4.3.2.

The AMI model does not use precipitation in the timestamp matching process. When rain or snow occurs the irradiance will significantly decrease, thus matching periods of precipitation with other high-cloudiness/precipitation periods.

3.4   Modeling Consumption Without PV

Another application of the timestamp matching AMI model is to predict electricity consumption when PV is not being produced. This method allows the estimation of consumption of any non-PV building, allowing a grid operator to estimate required load based solely on weather and previous AMI data. Because no load is offset by PV, the equation for estimated energy consumption during t is:

$$Consumption_t = Comparable\ Delivered_{median} \tag{3.6}$$

While estimating non-PV building load is not conducted in this work, consumption modeling will be discussed as areas of future work.

3.5   Forecasting PV Generation

PV generation estimates are of most value when the model provides forecasted future energy generation in real-time. An overarching goal of this research group was to explore the AMI model using forecasted weather data to project next-day generation.

3.5.1   Access to Weather Forecasts

SNL has built a set of python-coded functions that allow for the easy retrieval of forecast data for use in PV generation modeling. These functions utilize Unidata's Siphon library to access real-time forecasted weather data hosted by University Corporation for Atmospheric Research (UCAR) in Boulder, Colorado. UCAR hosts the Thematic Real-time Environmental Distributed Data Services (THREDDS) which provides these real-time and archived datasets for research and education [41]. Among these datasets include forecasts from the Global Forecast System (GFS), the North American Model (NAM), High Resolution Rapid Refresh (HRRR), and the Rapid Refresh model (RAP) [42]. The table below provides details on each of these weather models.

Table 3.1: Weather Forecasting Models.

Model	Geographic Resolution	Time Resolution	Forecast Window	Area
GFS	0.25 & 0.5 degree	3-hour	to 7 days	Any location on Earth
HRRR	3 km	1-hour	to 30 hours	Continental U.S. only
NAM	20 km	1-hour	to 4 days	North America
RAP	20 km, 40 km	1-hour	to 36 hours	Most of North America

3.5.2 North American Model

The North American Mesoscale Forecast System is a model run by the National Centers for Environmental Prediction. Four models are produced each day (every six hours) [43]. The parameters available from the dataset query are: air temperature, wind speed, GHI, DNI, DHI, total cloud, low cloud, mid cloud, and high cloud coverage [43]. Because of the model's multi-day range at 1-hour granularity, this model was selected for forecasting use.

3.5.3 Single Premise PV Forecasting

Weather forecast data for the upcoming 48 hours was queried by a forecasting function in the AMI model. For the purposes of this modeling effort, only air temperature (°C), wind speed (m/s), and GHI (watts/m^2) were utilized. The closest coordinate intersection modeled by the NAM was approximately 12 kilometers (km) from the geographic coordinates of downtown Fort Collins.

In the same manner as described in Section 3.3, the AMI model compared forecasted timestamps to past comparable timestamps. For this analysis of the NAM forecast, 48 target timestamps (one for each hour of forecasted weather) are matched to historic weather timestamps spanning back to January 1, 2015. Using the same AMI model method, estimated output from comparable timestamps are calculated.

The forecasted weather data was then inputted into the CSU-PVL model (described later in this work) with the specific system parameters from each PV asset (tilt, azimuth, capacity, date of installation, rate class) to create a PVL estimated time-series. Finally, these two outputs were plotted for each premise. Preliminary results from AMI forecasting are discussed in Chapter 7.

It should noted again that the timestamp matching does not utilize precipitation. The forecast model requires enhancement by adding a precipitation component. This element, combined with ambient temperature below freezing, would allow for a snowfall trigger that could be applied to the following day(s) to simulate snow coverage on panels. This is one element of this work's PV forecasting that needs further refinement.

Chapter 4

Experimental Results Part 1

This chapter discusses the experimental modeling results and findings for four systems as compared to their actual recorded generation.

4.1   Known Generation

Ideally, known generation from all or many residential arrays would be used to verify the model's performance. Unfortunately, large sets of actual generation from smaller PV systems outside of research datasets are not commonly available. Actual generation from four small-scale net-metered PV systems in Fort Collins were obtained and used to initially validate the accuracy and robustness of the AMI model. Table 4.1 details these net-metered systems.

Table 4.1: Known Generation Data Sources.

Site Name	Capacity (kW$_{DC}$)	Premise Type	Granularity	Duration
House G	5.04	Residential	Daily	13 months
House Z	6.00	Residential	Hourly	2 months
Timberline	20.59	Non-residential	Hourly	8 months
Church	53.46	Non-residential	Daily	12 months

House G had a PV system installed in mid-2016 with a tilt and azimuth of 26.5° and 160°, respectively. House Z hosts a system installed in early 2019 with a tilt and azimuth of 27° and 244°, respectively. The Timberline array consists of a ground-mounted system with 180° azimuth and 30° tilt. A local church in Fort Collins was outfitted with solar in fall 2016. The church has two arrays, 11.9 kW with a tilt of 21° and azimuth of 90°, along with a south facing 41.6 kW array at a 16° tilt.

These premises are a diverse and representative mix of net-metered solar assets. Most obviously, they span across a range of installed capacity and building types with different usages and

load profiles. Residential buildings typically have two daily peaks; a small morning spike and more substantial afternoon/evening increase in consumption. They also exhibit different energy needs during the work week and on weekends. The Timberline Recycling Facility is a city-owned building that is open 360 days a year and is served by a general service net-metered electricity connection. The load profile at this premise is likely be characterized by consistent energy usage during business hours every day of the week, with major fluctuations driven by HVAC needs. In contrast, the energy needs of the Church are likely low during the week and vary on weekends depending on the different services and one-off events that the site hosts.

These arrays also provide an assortment of orientation configurations for the AMI model to be compared against. The Church has a due east facing sub-array, while House Z faces west-southwest, with the Timberline and House G generally south facing systems. The tilt of these systems vary between 16° and 30°. Having a variety of PV arrangements helps build confidence in the robustness of the model on arrays of other potential designs.

4.2   Error Terms

There are several ways to prepare the data that can provide different views of the statistical agreement and consistency of the AMI model to recorded PV generation. While the AMI model operates sub-hourly, generation comparisons are conducted at hourly and daily resolutions.

The absolute percentage error, referred to as ε in this work, is a robust and commonly used error term metric for comparing a modeled output to an known value. It can be used across different PV systems with varying capacities. ε is defined as:

$$\varepsilon_t\ (\%)\ =\ 100\ \times\ \frac{AMI_t\ -\ Actual_t}{Actual_t} \tag{4.1}$$

Where AMI_t and $Actual_t$ are the estimated and actual values (kWh) at time interval, t.

With modeled and ground-truth data both in time-series form, the mean ($\bar{\varepsilon}$) and median ($\hat{\varepsilon}$) of the ε errors are useful to describe algorithm performance over the duration modeled. The $\bar{\varepsilon}$ is defined as:

$$\bar{\varepsilon}\ (\%) \ = \ \frac{100}{n} \ \times \ \sum_{t=0}^{n} \frac{AMI_\mathrm{t} \ - \ Actual_\mathrm{t}}{Actual_\mathrm{t}} \tag{4.2}$$

The $\hat{\varepsilon}$ is often reported in this work because it is more resilient to outliers than the $\bar{\varepsilon}$ and is defined as:

$$\hat{\varepsilon}\ (\%) \ = \ median(\varepsilon_1, \varepsilon_2, ... \ \varepsilon_n) \tag{4.3}$$

Where ε is an absolute percentage error and *n* is the number of ε being considered.

A weighted error term (ω) is also calculated because errors of similar magnitude are not of the same importance. A 10% ε at 7:00 am when generation is low is not as significant as a 10% ε in the mid-afternoon when the system is maximizing production. The weighted error term minimizes errors that occur during hours of lower generation and amplifies those that occur during greater energy production. This is achieved by scaling ε by the ratio of the estimated generation during the hour to the maximum estimated hourly generation for the system. The weighted error equation for any hourly error is:

$$\omega_\mathrm{t}\ (\%) \ = \ \varepsilon_\mathrm{t} \ \times \ \frac{AMI_\mathrm{t}}{AMI_\mathrm{max}} \ \times \ 100 \tag{4.4}$$

Where AMI_max is the maximum estimated hourly value (kWh) from the AMI model; and AMI_t is the estimated value (kWh) at the hourly time interval. The weighted error term is only calculated at hourly granularity, it is not used for analysis of daily generation. Throughout this work, a negative error value always corresponds to the AMI model value being lower than the value it is being compared to. Colloquially, a positive error means the AMI model overestimated relative to the other value.

28

The hourly mean ($\overline{\omega}$) and median ($\hat{\omega}$) weighted percentage error terms are found by the following equations:

$$\overline{\omega}\ (\%)\ =\ \frac{100}{n}\ \times\ \sum_{t=0}^{n} \omega_t \tag{4.5}$$

$$\hat{\omega}\ (\%)\ =\ median(\omega_1, \omega_2, ... \omega_n) \tag{4.6}$$

4.3   Results

The time-series of modeled and actual generation from each of the four premises are plotted together to graphically compare the two. Hourly generation at the Timberline and House Z arrays enables hourly analysis, while House G and the Church can only be compared at daily resolution due to data reporting limitations from the data acquisition system.

Residential modeling was characterized by daily $\overline{\varepsilon}$ between -4% to -26%, meaning that the AMI model was underestimating generation on average. Hourly generation was not available for House G, but for House Z the hourly $\hat{\varepsilon}$ was -26.5%.

Interestingly, total generation aggregated over longer periods of time correlated better to observed generation than at daily granularity. Over the course of one year at House G, the total sum of actual generation was 6,570 kWh while the AMI model estimated 6,492 kWh, a difference of -1.2%. Throughout the two months analyzed for House Z, the actual generation was 885 kWh while the AMI model estimated 673 kWh, a difference of -23.8%. However, it is known that the occupant of House Z is an energy engineer affiliated with this research and was aggressively shifting and minimizing loads to minimize consumption during higher pricing periods. Speculatively, the winter months modeled also may not have provided consistent weather and generation for evaluation of this newer system (installed February 2019).

Figure 4.1 depicts the actual and modeled daily generation values for three months from the House G system. Figure 4.2 portrays hourly recorded and modeled values from House Z.
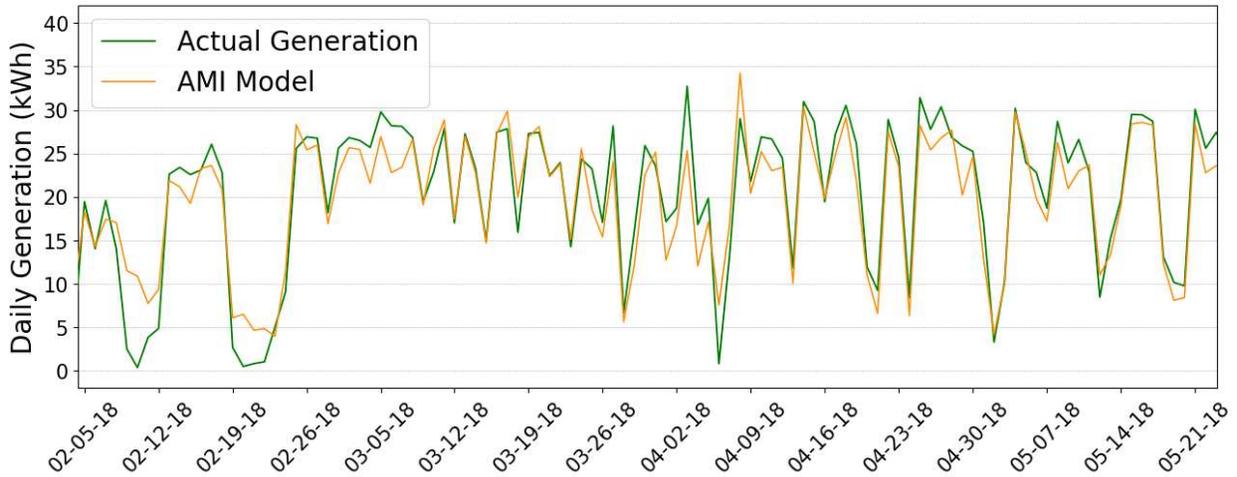
Figure 4.1: House G Daily Generation. 3 months of estimated daily generation aligned with recorded generation. The daily $\hat{\varepsilon}$ was -3.9%, that is, the AMI model typically underestimated daily generation by around 4%. As the generation data was accrued by day, hourly errors cannot be calculated.
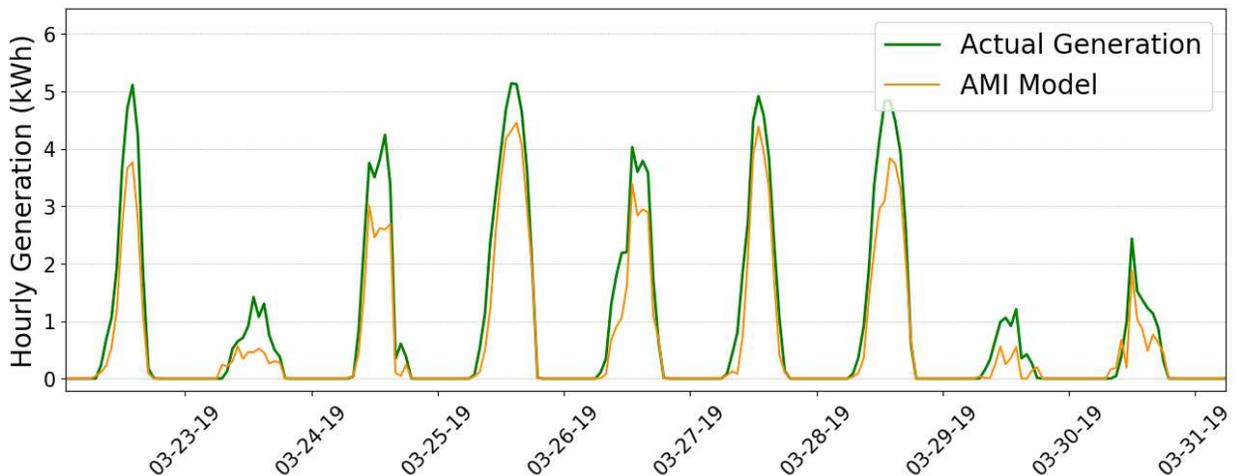
Figure 4.2: One Week of Hourly Generation Comparison for House Z. This hourly plot displays modeled and recorded hourly PV generation. The model underestimates during both sunny and cloudy (days 2,8,9) periods. The AMI model was found to have a $\hat{\varepsilon}$ of -26.5% and -25.8% at hourly and daily granularity, respectively.

Statistically and graphically (Figures 4.3 and 4.4) the model performed best at Timberline. At both hourly and daily resolution, agreement between modeled and actual data appear to be quite strong. This may suggest that consumption patterns are very consistent at this municipally-operated commercial site and did not diverge after PV system commissioning. Data from the

Timberline facility shows over the course of eight months analyzed, the total sum of actual generation 16,607 kWh, while the AMI model estimated a total of 16,735 kWh, a difference of only +0.8%.
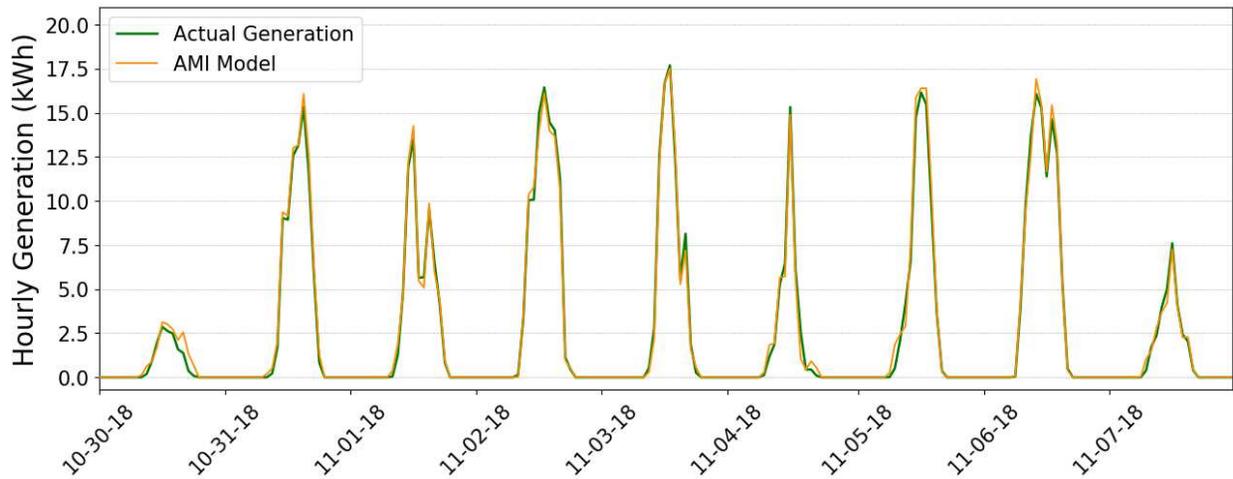
Figure 4.3: Timberline Hourly Generation. Shown is the hourly generation estimate from the AMI model plotted with actual generation. An hourly $\hat{\varepsilon}$ of 1.7% was observed. An $\overline{\omega}$ and $\hat{\omega}$ of +4.0% and +0.4% were calculated.
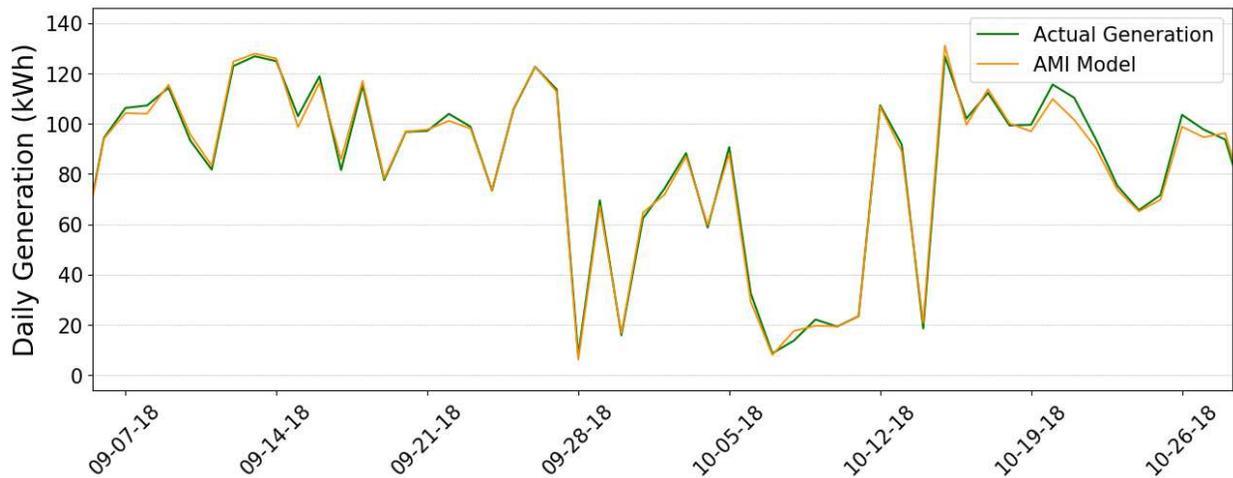
Figure 4.4: Timberline Daily Generation. Shown is the daily generation estimate from the AMI model plotted with actual generation. For the AMI model, a daily $\hat{\varepsilon}$ and $\overline{\varepsilon}$ unweighted error of +0.3% and +3.6% were observed.

The model accuracy for the Church was not as unblemished as the other non-commercial site at Timberline. At the Church the daily $\hat{\varepsilon}$ was -7.4% and inspection of Figure 4.5 illustrates the typical underestimation error. Churches typically only have on-site activity a few days a week, and load may be highly variable even between the weekend services and events. Different community events may consume energy in very different quantities, leading to more uncertainty from estimating generation from past consumption data.
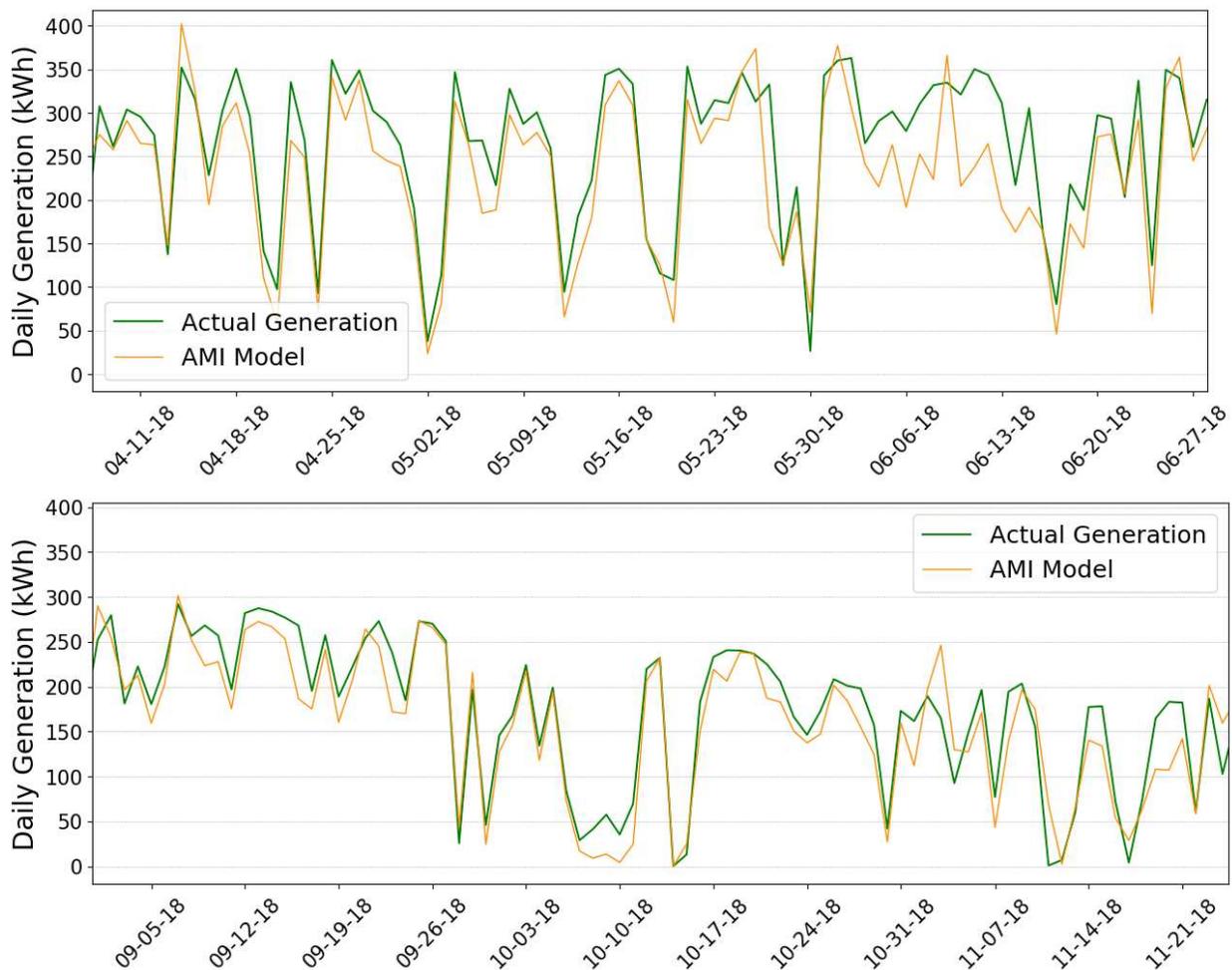
Figure 4.5: Daily Generation Comparison for the Church. These daily plots display modeled and recorded generation through various points of 2018. The model typically underestimates during both sunny and cloudy periods. The AMI model was found to have a daily $\hat{\varepsilon}$ of -7.4%.

The raw AMI data can provide insight into the consumption patterns of these buildings. From Timerbline's AMI data (Figure 4.6 below), it can be seen that there is a consistent afternoon/evening

usage around 2.5 kW, with an overnight base load of around 2 kW. Of course, daytime consumption is mostly hidden due to BTM generation usage; days of low generation such as Jan. 9th, 16th, and 18th do display periods of daytime delivered energy. As shown in Figure 4.7, Church activity is relatively easy to discern as weekend spikes in consumption, with green arrows indicating Sundays. Wednesday night services are also visible during the week in Figure 4.7. The consumption on other days is usually less than array generation and thus not shown in the AMI data. Because Timberline probably has more consistent day-to-day activities, logically the AMI model would perform better at that site than at the Church.
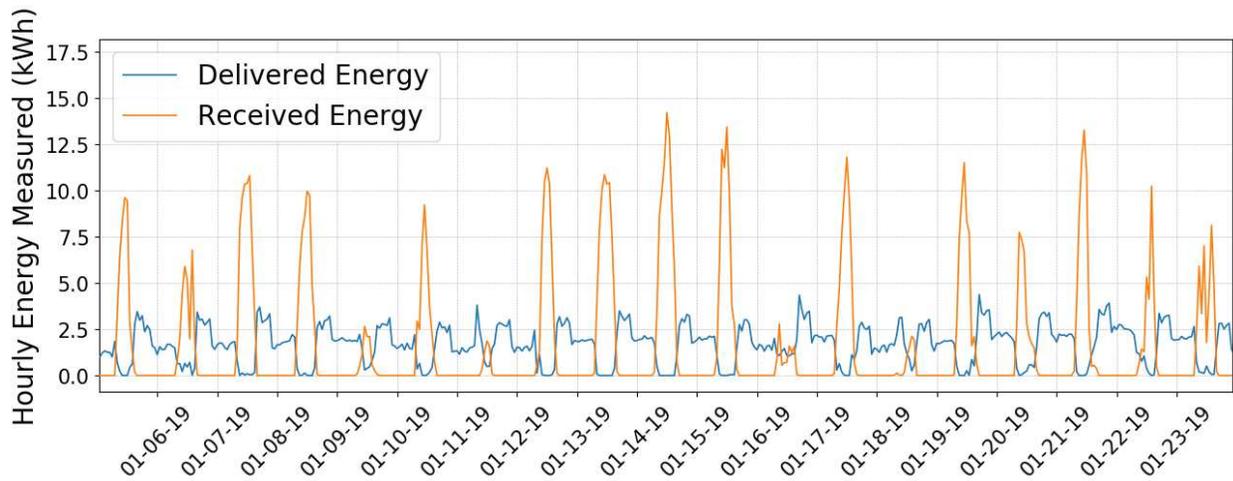
Figure 4.6: Timberline Hourly AMI Data. Several days of invisible generation offsetting delivered energy needs are visible on Jan. 9th, 16th, and 18th, 2019.
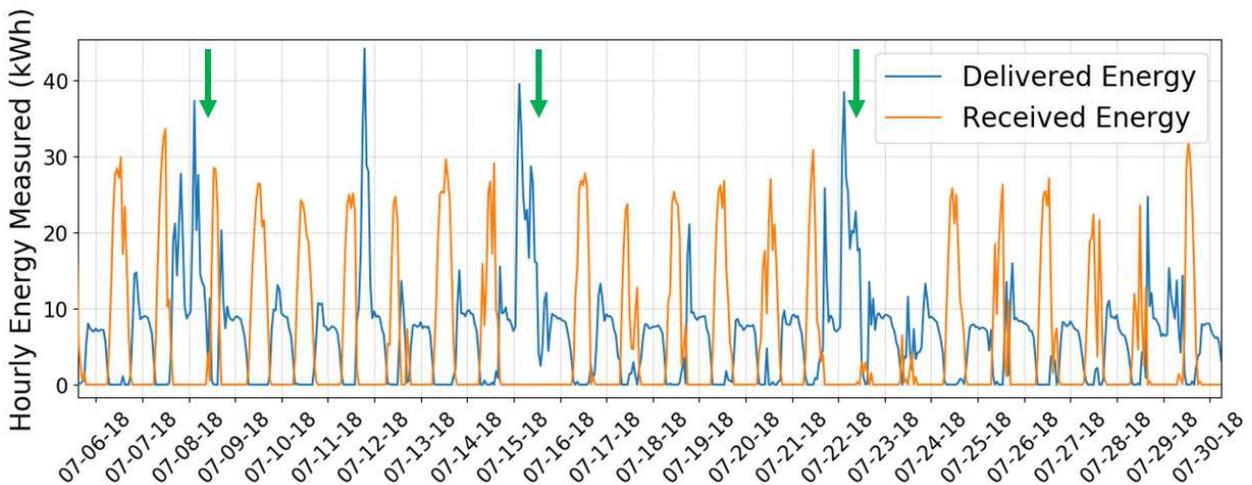
Figure 4.7: Church Hourly AMI Data. Sundays are denoted by green arrows.

4.3.1 Monthly Estimates

For individual premises, modeling monthly generation provides a useful interval of estimation. Most utility customers in the U.S. receive a monthly bill for total consumed energy (kWh) from approximately the previous 30 days, and the Energy Information Administration reports monthly electricity usage statistics [44]. This kWh total is converted to a dollar amount through various billing structures. A residential solar system owner would likely want to compare their actual monthly generation with predicted monthly generation to understand system performance and expectations for the payback period of the array. Thus, aggregated actual and estimated monthly values are compared and discussed in this section.

In Figure 4.8 the aggregated monthly errors from each site are plotted together. As shown, 27 of 35 months plotted are within $\pm$ 10%. The greatest variance from known generation totals occur from the two residential premises. Interestingly, the larger systems are better modeled at a monthly level. This prediction accuracy to actual generation may suggest meaningful monthly or annual predictions are possible with this model.
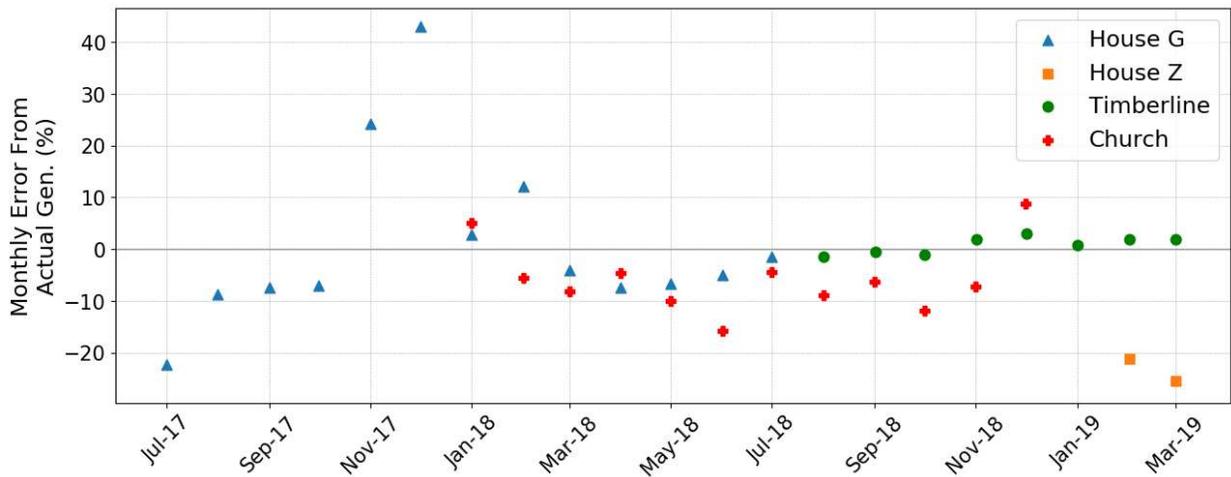
Figure 4.8: Aggregated Monthly Errors. 22 months of aggregated monthly generation errors presented for 4 sites with known generation. Winter months appear to be the most difficult to consistently model. The AMI model typically underestimates monthly generation by 0-10%.

34

The cause of the large spike in error for House G in November and December 2017 is not initially apparent. Closer inspection of the AMI data reveals that the smart meter at this premise was malfunctioning for about 4 weeks spanning the two months. This artificially low delivered energy (Δ) is subtracted from the P_{med} term in Equation 3.1 drastically increases the generation estimate during the affected period. This overestimation is evident by the large error toward the AMI model for those two months. Figure 4.9 below displays this incident with delivered values plotted on top of received energy values.
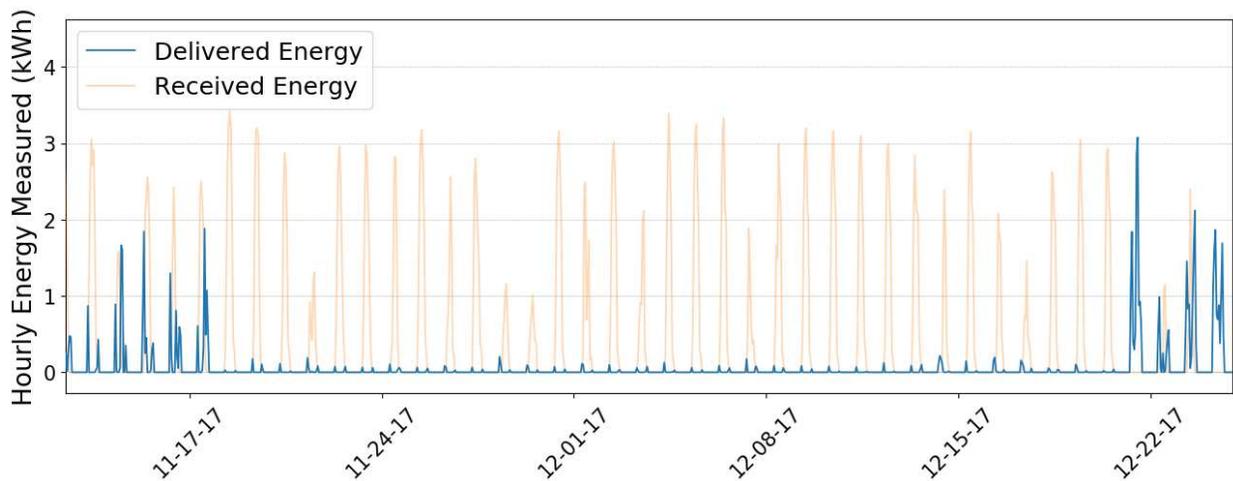
Figure 4.9: Missing AMI Data at House G. Nearly one month of delivered energy values appear to be missing, denoted by the blue line. The orange line represents recorded hourly received energy to the grid.

Data quality issues from this small sample size have significant impact on monthly error results, and AMI data quality and cleaning issues are discussed in Section 6.5. While more known data for comparison is needed, the general trend of the AMI model slightly underestimating generation is visually apparent in Figure 4.8. The causation of this underestimation pattern is not fully understood and may be a result of behavioral changes in energy use after solar is installed. This explanation would imply energy consumption at a premise rising slightly after installation, causing the overall estimated PV generation term decrease slightly. As mentioned in Section 3.1, this pattern has been been observed in behavioral research as the rebound effect after PV is installed [37].

Nonetheless, in summary, these monthly comparisons reinforce the notion that the AMI model achieves reasonable precision within $\pm$ 10% from actual generation.

4.3.2   Snowfall Performance

Estimating PV generation during or after snow events is tedious and experimental. Cloud coverage and precipitation distribution are unpredictable and highly variable across small areas, lending to uncertainty in prediction. Because post-snowfall temperature and sunlight is a key driver of snow melting, the following days' weather are important to determine panel clearing and resumption of unobstructed generation. The inherent nature of the AMI model positions it to be well-suited at predicting generation reduction due to snow coverage. The AMI data describes what is happening within the building's load, not just on the roof under snow. Net-load is inclusive of panel snow coverage and other losses, by-passing the various physical model external variables that must be factored in.

The following plots display modeled and actual generation for two snowfall events at two different arrays. An accurate estimate of generation per the AMI model is displayed in Figure 4.10 below. Between October 6-11, 2018, a mix of rain, light snow, and heavy cloud cover was observed, greatly reducing generation over the six-day period as shown by measured generation plotted in green. Reduced generation during these overcast and pluvious days is shown in the AMI estimate (orange), considered to be in high agreement with known generation.
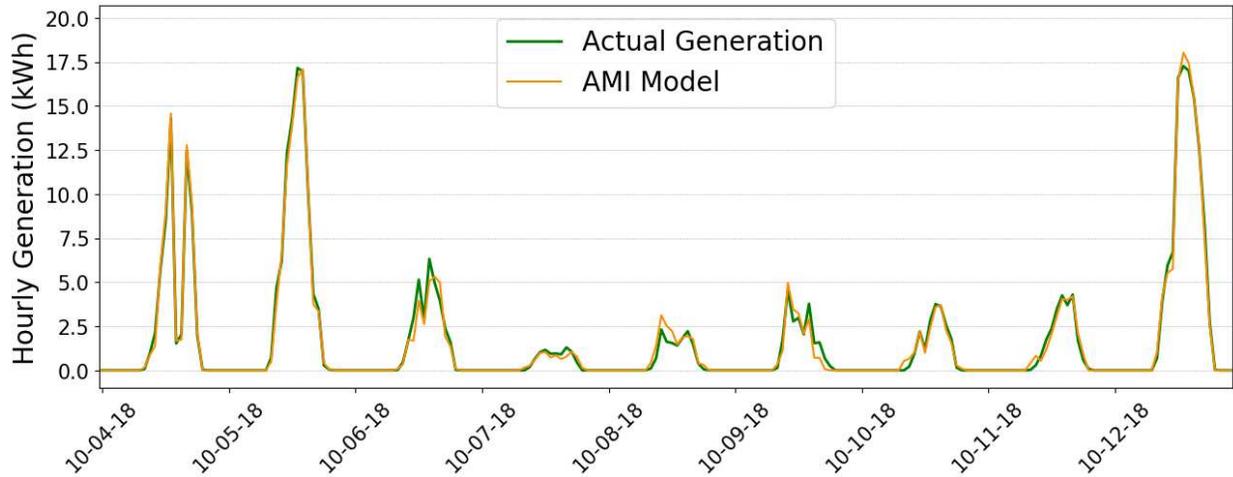
Figure 4.10: Model Performance During Precipitation Period at Timberline Site. Poor weather was observed from October 6 through October 11, 2018 in Fort Collins, which is reflected in the AMI model and recorded generation data.

Figure 4.11 displays a heavy snow event on March 2, 2019 in which Fort Collins received over 7-inches of snow. In addition to no generation on March 2nd, two succeeding days of negligible generation due to snow coverage are shown by the AMI model and recorded generation data.
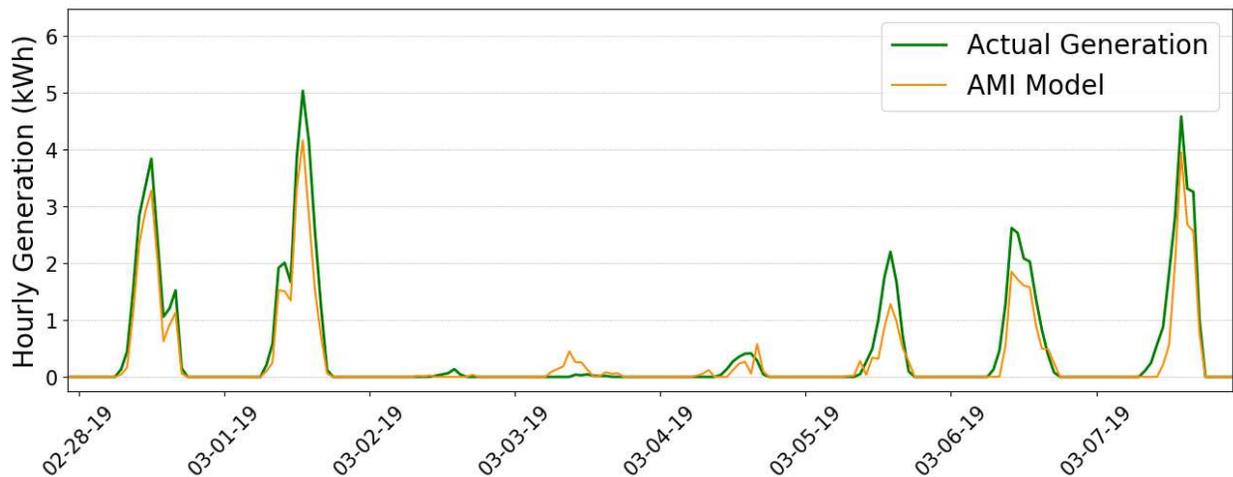
Figure 4.11: House Z Generation during Snow Event. On March 2, 2019, Fort Collins received over 7 inches of snow.

The AMI model is responsive to the effects of snow coverage during and after the snowfall event. While by no means perfect, the AMI model's structure produces generation estimates that are not "fooled" by high irradiance periods while an array lies under inches of unmelted snow. This capability provides a competent alternative to prior snowfall modeling approaches: AMI snowfall performance compared to physical models will be discussed in Section 6.4.

4.4   AMI Validity

Overall, the AMI model accomplishes the objective of achieving reasonable accuracy via comparison to known generation. Although fine and aggregated errors exist, nothing in these results suggest the experimental method is deeply flawed. A foundation for disaggregation of net-metered data was constructed and tested on two residential and two commercial systems representative of a variety of array configurations.

To test the robustness of the model across thousands of systems in a distribution network, greater effort was expended on the remainder of the modeling activities in this work. Without the luxury of actual generation data, a surrogate for output was needed. Consequently, a detailed and input-intensive physical model was built, which acted as a proxy for known production. Generation estimates born from the AMI data of 846 PV systems in Fort Collins were compared against this physical-modeled output. Chapters 5 and 6 describe and discuss the outcome of this portion of the work.

Chapter 5

Development of PVL

FCU does not have access to actual generation data for nearly 1,300 residential PV systems in Fort Collins. Because FCU has recorded basic array characteristics, a physical model to estimate generation provided a proxy for these premises. This physical model was used as a validation tool by comparing the PVL output to the AMI model's output for these arrays.

Prior PV performance models developed by Sandia National Laboratories were chosen as a template for this physical model. The modeling components in this thesis work were based on the Sandia National Labs PV Performance Modeling Collaborative. All the information from the PVPMC is open-source and available online [31]. This PV library provided the methodology framework for the model, named "CSU_PVL", built by the CSU research team.

The CSU-PVL model is a physical model that incorporates solar irradiance, thermal, electrical, mechanical, and optical components [30, 31, 45]. The model also built off prior work by CSU research associate John Bleem [46], who developed a similar solar model based on the PVPMC and related sources [30, 31, 45, 47, 48]. Inputs into CSU-PVL are the same weather data in Chapter 2, along with physical inputs for each array such as tilt, azimuth, and capacity$_{DC}$ from the City's *DG Assets*. Performed on each premise with PV, the model's estimated generation time-series provided surrogate generation output to determine the precision of the AMI model in the absence of known generation data.

The following sections in Chapter 5 superficially discuss the model's development based on Bleem's prior work [46]. Appendix B provides a more detailed step-by-step description, along with the assumed variable values and coefficients used in the CSU-PVL model.

5.1   PVL Model Inputs

The PVL model was designed to accept weather data at any temporal resolution. The weather parameters needed are the timestamp's GHI, wind speed, and ambient air temperature.

There are three functions within the model that are run in series order. First, the solar positioning algorithms calculate the solar azimuth and zenith at each inputted timestep. Next, the Erbs model is used to derive DHI and DNI from the measured GHI. Finally, energy output is calculated incorporating temperature correction, system efficiencies, losses, and array capacity. Coefficient values can be found in Appendix B.

5.2 Plane of Array Irradiance

The fundamental action in calculating PV cell performance is to ascertain the irradiance incident on the surface plane of the array (POA) as a function of time. This POA irradiance is a function of the position of the sun, the orientation of the array, the solar resource, albedo, and coverage losses such as shading, soiling, or snow cover. The equation for energy from POA irradiance is shown below:

$$E_{POA} = E_b + E_g + E_d \tag{5.1}$$

The two major components of E_{POA} energy reaching the PV cell are the beam component (E_b) and the irradiance diffused through the atmosphere (E_d). Beam energy is from irradiance that is directly normal from the sun to a horizontal surface on Earth, which is known as direct normal irradiance (DNI). DNI is used to calculate E_b based on Equation 5.2. Diffused horizontal irradiance (DHI) entering the atmosphere is used to calculate E_d in Equation 5.5. and relies on empirical modeling to determine. Figure 5.1 illustrates the two major components of incoming solar irradiance. The ground reflected component of irradiance (E_g) is not considered in this version of the PVL model.

5.2.1 Solar Position Calculation

Direct normal irradiance (DNI) is the incident irradiance caused by direct beam from the sun - i.e. does not include any atmospheric scattering or ground reflection. The beam component is calculated as:

$$E_b = DNI \times cos(AOI) \tag{5.2}$$
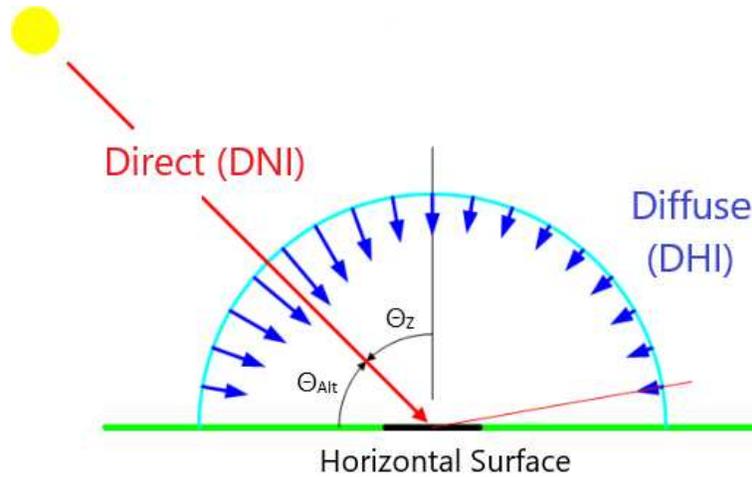
Figure 5.1: Direct and Diffuse Solar Irradiance. The solar zenith angle (θ_Z) is shown, which can be used to calculate the solar altitude angle ($90° - \theta_Z$). Image adapted from [49].

Where AOI is the angle of incidence between the sun's direct normal ray and the PV cell, found as:

$$AOI = cos^{-1} \left[cos(\theta_Z) \times cos(\theta_T) + sin(\theta_Z) \times sin(\theta_T) \times cos(\theta_A - \theta_{array}) \right] \quad (5.3)$$

Where θ_A and θ_Z are the solar azimuth and solar zenith angles. θ_T and θ_{array} are the tilt and azimuth angles of the PV array, in which an azimuth of $180°$ is a system facing due south.

The solar zenith is measured as the angle of the sun in the sky relative to a location on the surface of earth. A zenith of $0°$ is defined as the sun being directly overhead, while a zenith angle of $90°$ occurs when the sun is directly horizontal to the observer on the horizon. The solar azimuth angle is the sun's relative direction from an observer on Earth, with north having an azimuth of $0°$ increasing in the clockwise direction from north. Calculating the solar zenith and solar azimuth for a given point in time is a complex process that is described in greater detail in Appendix B. Figure 5.2 displays the relationship of these solar angles.
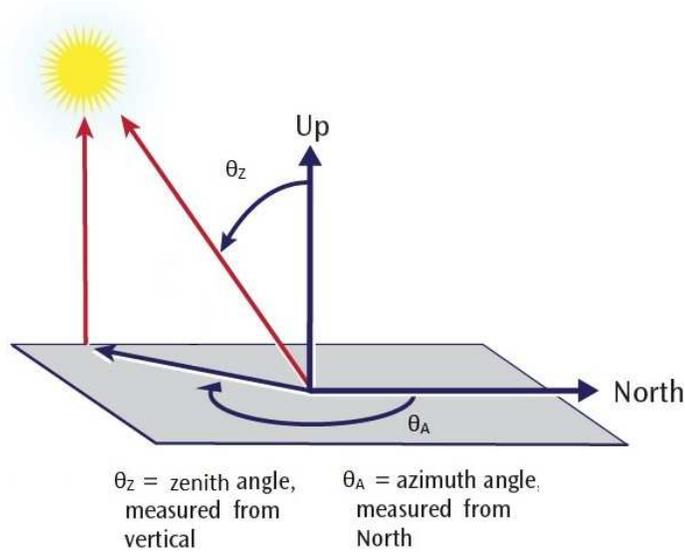
Figure 5.2: Solar Zenith and Solar Azimuth Angles. Visual definition of the solar zenith angle (θ_Z) and solar azimuth angle (θ_A). Image from [50].

5.2.2   Calculating DNI and DHI from GHI

Typical measurements of solar irradiance consist of the global horizontal irradiance and do not individually measure DNI and DHI. The relationship between the two components depends on the clearness of the sky at the given point in time; on clearer days more of the irradiance will arrive as direct beam radiation.

Empirical models have been developed using decomposition models to estimate the diffuse fraction (k_d) of irradiance based on a clearness index. The index of clearness through the atmosphere (k_t) is defined as as the ratio of irradiance reaching the surface to available extraterrestrial irradiance. The diffuse fraction (k_d) is calculated from a piecewise range of k_t values. The PVPMC provides three decomposition models; the Erbs model was selected for this work because it was developed with data collected from weather stations in the U.S.A. at latitudes near that of Fort Collins [51]. Further explanation of Erbs and the derivation of Erbs decomposition constants can be found in Appendix B.2.

Once the empirical k_d is derived, DHI reaching a flat surface is calculated using GHI:

$$DHI = GHI \times k_d \qquad (5.4)$$

To calculate the diffuse radiation on a tilted surface, another empirical model has been developed [52]:

$$E_d = DHI \times \frac{1 + cos(\theta_T)}{2} + GHI \times \frac{(0.012\theta_Z - 0.04) \times (1 - cos(\theta_T))}{2} \qquad (5.5)$$

Where θ_Z is the solar zenith angle and θ_T is the tilt angle of the PV array. Further, DNI is calculated as:

$$DNI = (GHI - DHI) \times cos(\theta_Z) \qquad (5.6)$$

And can be used to calculate the beam component E_b solving with Equation 5.2. Once E_{POA} is calculated as the sum of E_b and E_d, factors that affect the conversion of E_{POA} to alternating current electricity from the cell and inverter are considered.

5.3   Temperature Effect on Modules

PV cell performance is impacted by the module temperature as higher panel temperatures produce lower energy output, especially in commonly-used silicon-based cells [46]. Based on a model developed by SNL, module temperature is primarily governed by heat from solar irradiance, ambient air temperature, and cooling effects from wind [53]. These inputs calculate a temperature correction later factor applied to the power output from the PV cell. The temperature correction factor is defined as:

$$T_{correction} = 1 + T_C \times (T_M - T_{test}) \qquad (5.7)$$

Where T_C is a calculated temperature coefficient, T_M is the calculated module temperature, and T_{test} is the standard test temperature of 25 °C. The derivations for T_C and T_M and needed empirical coefficients are found in Appendix B.3.

5.4   Equipment Efficiencies and Losses

Array generation is also a function of various system characteristics, some of which may not typically be known to the modeler and must be estimated. The panel and inverter efficiencies (η_{panel} and $\eta_{inverter}$) are denoted as fractional efficiencies out of 1. Other system losses such as wiring losses, aging degradation, and inverter clipping also effect the final output from the system.

Another required input into the PVL model is the PV cell area, A_{cell}, which is the total area of PV cells within an array (m^2). If the specific module cell area's are not known, approximations must be made in order to convert Capacity$_{DC}$ into PV cell area. The general equation for calculating cell area is:

$$CellArea\ (m^2)\ =\ Capacity_{DC}\ \times\ 1000\ \times\ (Area_{module}/Wattage_{module}) \tag{5.8}$$

Where Capacity$_{DC}$ is in kilowatts, module area in m^2, and module rated power in watts. A utility will often not know the specifications of the modules installed in the array, requiring approximations to convert Capacity$_{DC}$ into cell area.

The model used an estimate of η_{panel} = 16.2% and a module power rating of 250 watts, based on NREL's 2018 PV System Benchmark report. From an average phsyical panel size of 1.63 m^2, an average model area of 1.55 m^2 was calculated. This is from the assumption that the cell area of a panel encompasses 95% of the total panel area. Physical panel area was found by comparing numerous commercially available modules and calculating the average area of these panels. These assumptions are explained in greater detail in Appendix B.4.

Inverter clipping is another factor that causes a reduction in power generated by the array. Formally known as inverter saturation, inverter clipping occurs when the DC power from the array is greater than the maximum input for the inverter [54]. While the DC rating of an array commonly

exceed the input capacity of the inverter to maximize long-term energy yield, at a certain point the inverter will regulate power input and clip potential generation from the array [54]. Because the City's PV records contained inverter quantity and model, the maximum AC output (AC$_{max}$) from the premise can be determined from the different inverter specification sheets. When modeled AC output (P$_{AC}$) exceeded AC$_{max}$, the output was reduced to AC$_{max}$ if coherent inverter information was available.

Figure 5.3 below displays inverter clipping during summer days at CSU's Student Recreation Center (SRC), which hosts a 544 kW$_{DC}$ system. The horizontal red line expresses the maximum AC output of the array, which is 437 kW$_{AC}$ based on the site's inverter data.
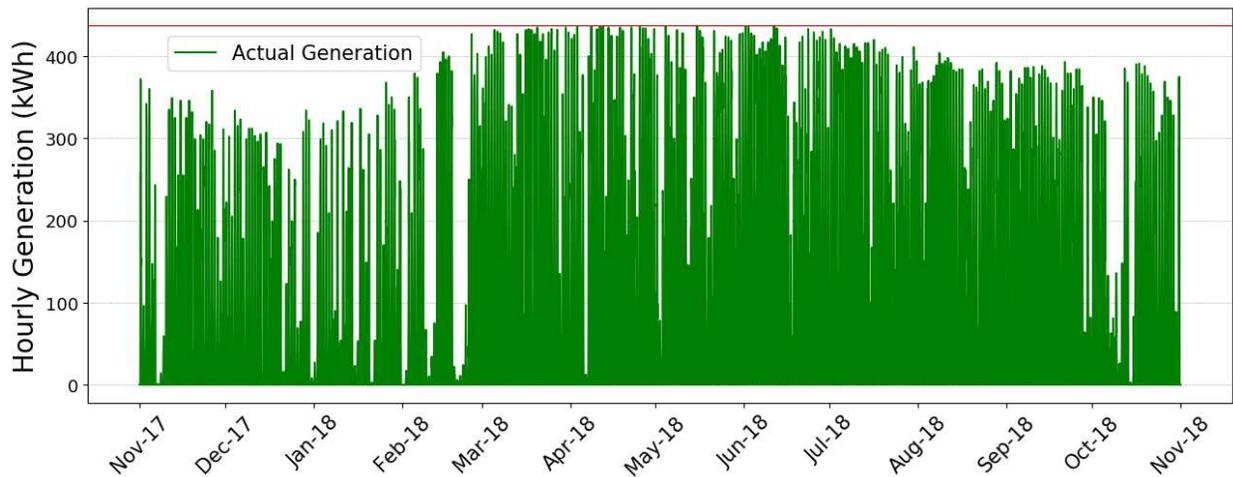
Figure 5.3: Inverter Clipping at the CSU SRC. DC components are often over-sized to maximize generation yield.

From *DG Assets*, inverter information enables an AC$_{max}$ value to be determined for 88% of premises. In Fort Collins, the average inverter load ratio was only 1.02 from these premises. If inverter information was not available or incomplete, inverter clipping and maximum AC power output were not considered for the premise modeled.

Inverter efficiencies were also available from the numerous inverter product specification sheets. Over 120 inverters are present in Fort Collins and their California Energy Commission (CEC) efficiency values were compiled. The CEC rating is a standard designed to allow a less complex and

more accurate method of calculating inverter efficiency [55]. The average CEC efficiency present in Fort Collins was calculated to be 96.1%. A rounded value of 96% was used in the model for η_{inverter}.

To encompass wiring losses, aging degradation of equipment, soiling, outages, and other miscellaneous losses, additional system losses (SL) were set to 8.7% based on prior PV modeling experience from [46]. Table B.1 contains a comprehensive list of assumed values and coefficients used the PVL model.

5.5 Power Production

As outlined by SNL and Bleem, AC power (P_{AC}, in watts) can be calculated using the previously determined variables as discussed in this chapter [46, 56]. AC generation is found as:

$$P_{AC} \; = \; [\, P_{DC} \times (1 \; - \; \frac{SL}{100}) \; \times \; (\, 1 \; - \; \frac{\eta_{\text{inverter}}}{100} \,) \, / \, 1000 \,] \tag{5.9}$$

Where SL is system losses (%), η_{inverter} is inverter efficiency (assumed as 96%), and P_{DC} is DC power from array side of the inverter. P_{DC} is found as:

$$P_{DC} \; = \; [\, E_{POA} \; \times \; \eta_{\text{panel}} \; \times \; Area_{cell} \; \times \, [\, 1 \; + \; T_C \; \times \; (T_{cell} \; - \; T_{test}) \,] \,] \tag{5.10}$$

Where E_{POA} is the plane of array energy found in Equation 5.1. T_{cell}, T_{test}, and T_C are the cell temperature (°C), reference test temperature (25 °C), and the temperature coefficient, respectively. $Area_{cell}$ is the cell area (m^2), and η_{panel} is panel efficiency, which is assumed to be 16.2% in this work.

5.6 Snow Coverage Reduction

Typically, physical models approach snowfall correction by using the tilt, ambient temperature, and plane of irradiance to determine snow sliding down a panel [57]. This phenomena is displayed on the CSU Powerhouse's array (tilt = 15°) in Figure 5.4 below.

Figure 5.4: Powerhouse Array with Snow. Partial coverage of a PV array after a snowfall event in Fort Collins. Photo taken May 9, 2019 at the CSU Powerhouse Energy Campus by Wendell Stainsby.

A slide amount is calculated to determine exposed panel surface based on an empirically derived sliding coefficient and a function of the steepness of array tilt [58]. As the panel and environment warm, snow melts and/or slides further off the panels, and generation increases until the snow has expired.

This snowfall correction method is used by NREL's System Advisor Model (SAM) for PV projects. While this snowfall correction was shown to improve the SAM estimates by 4-8%, it is an imperfect process [57]. To this point, NREL's report on the validation and application of the PV snow coverage model states [57]:

> *The snow model is observed to both over-predict and under-predict energy estimates in an unforeseeable fashion on a monthly, daily, or hourly basis. This is expected behavior, however, since Marion et al. (2013) states that the original model performs*

well on an annual average despite the fact that "large differences between modeled and measured energy losses should be expected for monthly or shorter time periods". For this reason, results from the model implemented in SAM should only be factored into annual considerations and not applied to monthly or shorter time periods.

Unrelated prior work from the Department of Energy determined rough estimates of system output reduction over time based on the tilt of the panel [59]. Brench determined average energy losses for 30° and 40° panels for various snow coverage and weather scenarios [59]. These energy reductions were extrapolated from flat (0°) to 45° systems and applied to the hourly generation estimates. For this work, daily snowfall data from NOAA was utilized. Snowfall reduction was triggered if an hour occurred during a day which recieved at least 1-inch of snow that day, or had at least 3-inches of accumulated snow from a prior event during the calendar day.

The following table displays the different hourly snowfall reduction factors based on extrapolation of panel tilt from [59]. The panel tilts listed represent the upper bin limit of the angle for the reduction factor.

Table 5.1: Energy Loss Based on Panel Tilt.

Tilt (°)	Loss Factor (%)
5	92.5
10	83.0
15	73.5
20	64.0
25	54.5
30	45.0
46	26.0

The snowfall modeling reduction effect is displayed in the figures below, which depict two snow events in March 2019 with actual generation data from House Z. On March 2, 2019 Fort Collins received 7.3-inches of snow that persisted as accumulated snow until March 6, 2019. From Figure 5.5, the generation reduction from panel coverage in spite of good weather conditions is shown. March 4th & 5th were sunny days, and the uncorrected PVL model (dashed red) grossly

overestimates generation compared to actual generation (green). While a corrected PVL estimate (blue) still overestimates generation, it is a substantial improvement from the uncorrected output estimate.
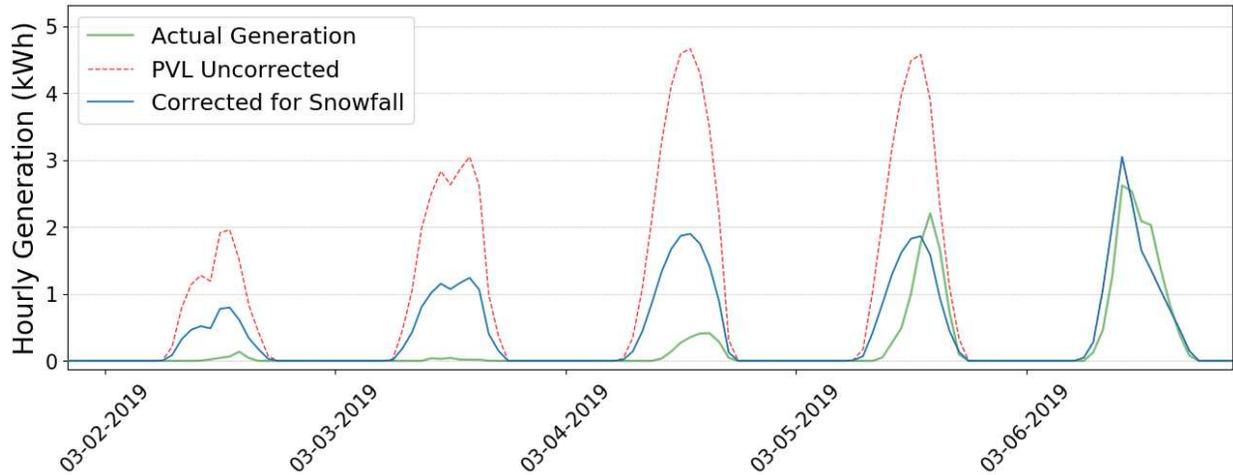
Figure 5.5: Multiday Snowfall Effects on PVL Model. The effect of snow coverage on modeled generation is shown. Uncorrected generation has poor performance during ensuing days with inundated panels. Snowfall correction from March 3-5 was conditional on daily accumulated snow measured to at least 3-inches.

On March 13, 2019 Fort Collins received 2.8-inches of snow that melted away by the next day and only March 13th triggered snowfall reduction. While the weather dictated the uncorrected PVL model to estimate low generation, the corrected estimate aligns even better with the actual generation. Further, good alignment with between actual and uncorrected estimated generation is achieved the following day, confirming that the snow did in fact melt by March 14th.
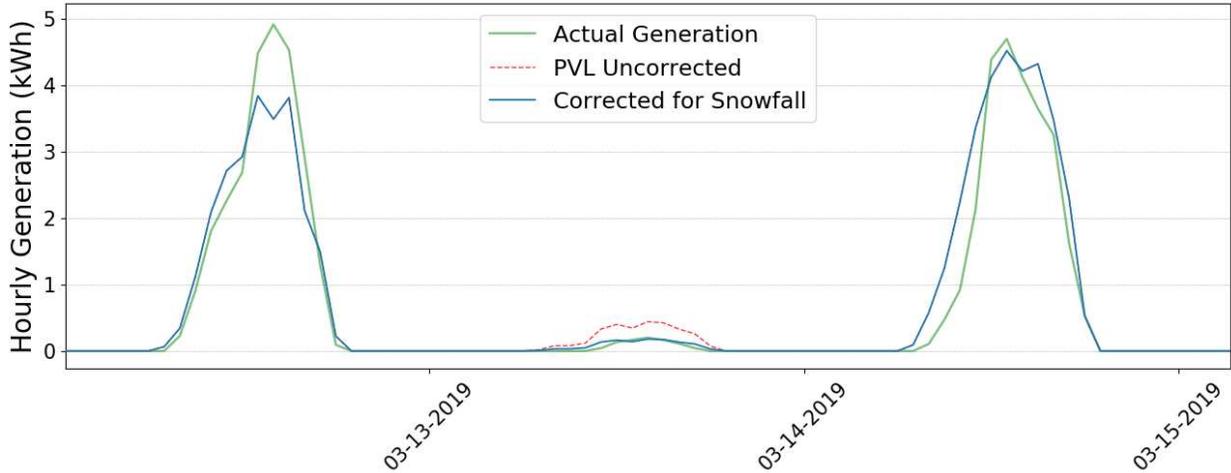
Figure 5.6: Single Day Snowfall Effects on PVL Model. High accuracy is achieved for the corrected estimated relating to recorded generation. As snow melts away by the next-day, recovery of system production is impressive.

As previously stated by NREL researchers, modeling snowfall generation reduction at fine temporal resolution is very difficult to perform. Nonetheless, this work attempts to incorporate snowfall reduction that consistently improves the estimate. Greater improvement than what was already achieved by the PVL snowfall correction would require arduous modeling that is outside of the scope of this work.

5.7   PVL Verification

The PVL model was verified with known generation from ten array sites, and the details of these systems are shown below in Tables 5.2-5.4. While sites such as the CSU Powerhouse and the 222 City building are also net-metered, they were not used to verify the AMI model in Section 4.3 because the systems were installed before the January 1, 2016 cutoff. Larger systems in Table 5.4 are independently metered and were also installed before 2016.

Table 5.2: Residential Known Generation Sources.

Site Name	Capacity (kW$_{DC}$)	AC$_{max}$ Output	Granularity
House G	5.04	4.5	Daily
House Z	6.00	5.4	Hourly

Table 5.3: Commercial Known Generation Sources.

Site Name	Capacity (kW$_{DC}$)	AC$_{max}$ Output	Granularity
Timberline	20.59	19.72	Hourly
CSU Powerhouse	21.6	20	Hourly
Church	53.46	43.2	Daily
Firehouse Alley Garage	93.84	90	Hourly
222 City Building	103	99.1	Daily

Table 5.4: Utility-scale Known Generation Sources.

Site Name	Capacity (kW$_{DC}$)	AC$_{max}$ Output	Granularity
CSU Vet Teaching Hospital	220.2	184	Hourly
CSU Student Rec. Center	544.7	437.6	Hourly
Bella Site	971	784	15-minute

Analysis was conducted using the PVL model along with actual generation, and the outputs were statistically compared. The first statistic that was used was total generation error, the integrated amount energy produced during the durations modeled expressed in kWh or MWh. The second statistic reported was $\hat{\varepsilon}$, the median value of daily error between PVL and actual generation. The two figures on the following page present these errors, both expressed as percentages, for the ten sites. Each marker on the scatter plots is scaled to the system's capacity to illustrate the range of system size in this analysis. The systems are color-coded to group similar building and size types: residential (blue), commercial (red), and utility-sized (green).

Total generation error, a measure of long-term precision, is shown in Figure 5.7. The two residential premises displayed the highest PVL overestimation, but there does not appear to be a correlation of system-size and this measure of accuracy. All systems but the Bella Site expressed positive total generation error, meaning that the PVL model overestimated the other nine systems, ranging from 0.3%-13%.
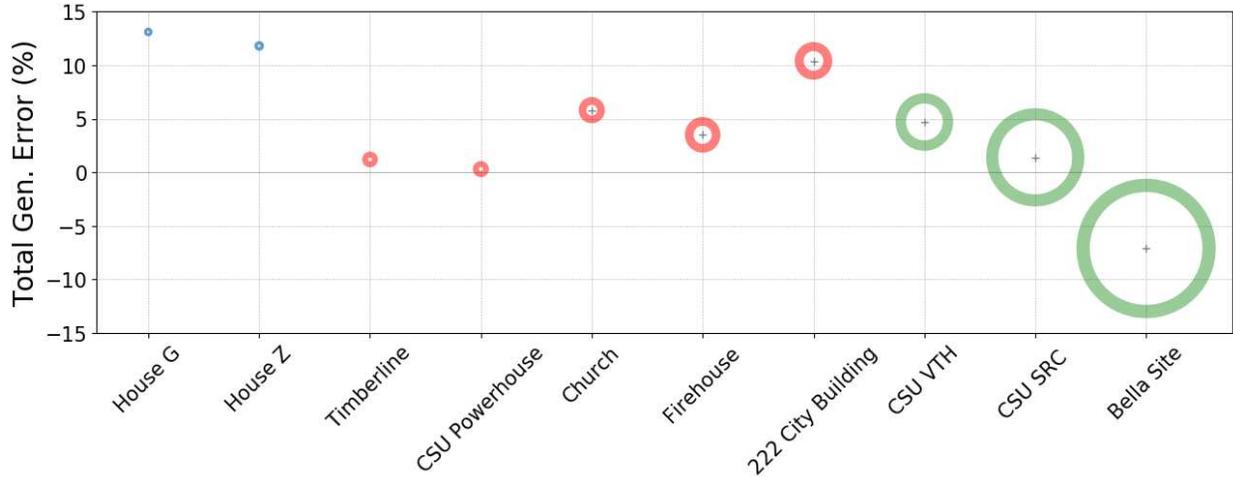
Figure 5.7: Total Generation Error from PVL Model for 10 Sites. The residential premises display the highest PVL overestimation, but there does not appear to be a trend in error based on array size.
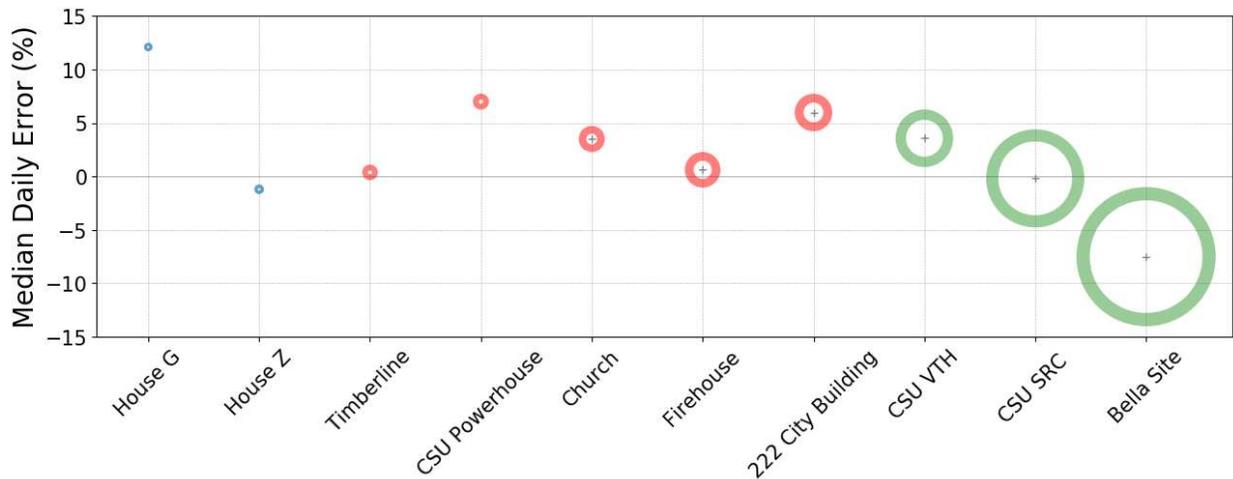
Figure 5.8: Median Daily Error from PVL Model for 10 Sites. The median of daily error is $\pm 10\%$ for all systems except House G, which has known shading present.

All but one system had a $\hat{\varepsilon}$ daily error less than 10% from actual generation as displayed in Figure 5.8. A noticeable difference between the two plots is the negative $\hat{\varepsilon}$ value for House Z while the premise had an positive and relatively large total generation error.

Several of the systems in this Actual-to-PVL analysis are detailed individually in the following pages. House Z, as previously described, has a 6 kW_{DC} array. Total recorded generation over the period analyzed was 888 kWh, with a modeled generation total of 993 kWh, a difference of +11.8%. The $\hat{\varepsilon}$ at daily granularity was -1.2%. and Figure 5.9 below displays these outputs.
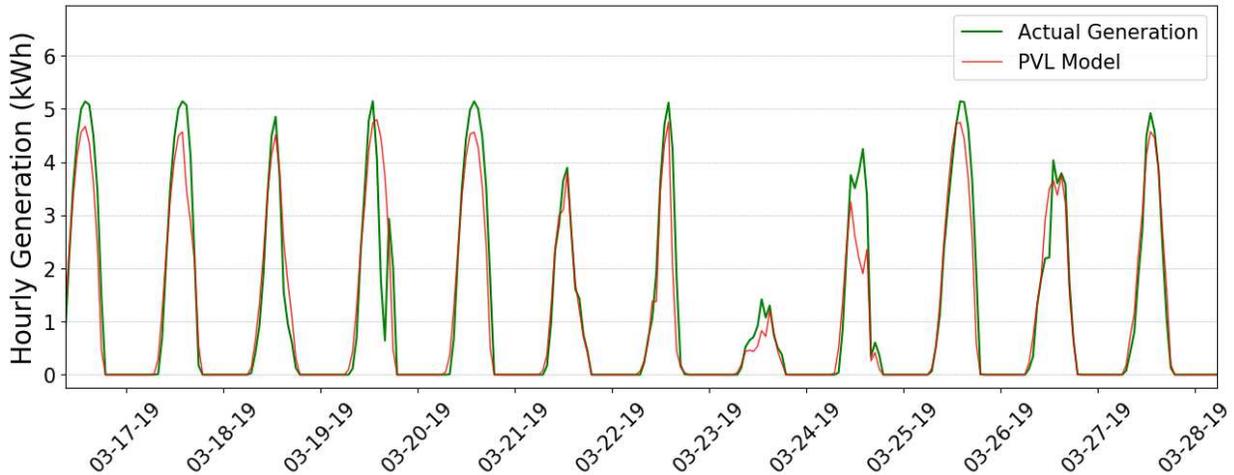
Figure 5.9: House Z Actual and PVL Generation. House Z exhibited a negative $\hat{\varepsilon}$ value, indicating that PVL is likely typically underestimating with a few outlier days of overestimation.

Timberline has a 20.59 kW$_{DC}$ array. Total recorded generation over the period analyzed was 16,590 kWh, with a modeled generation total of 16,789 kWh, a difference of +1.2%. The $\hat{\varepsilon}$ at daily granularity was 0.4%, showing model precision at this site. Figure 5.10 below depicts Timberline's actual and PVL generation outputs. The CSU Powerhouse houses a 21.6 kW$_{DC}$ array. Total
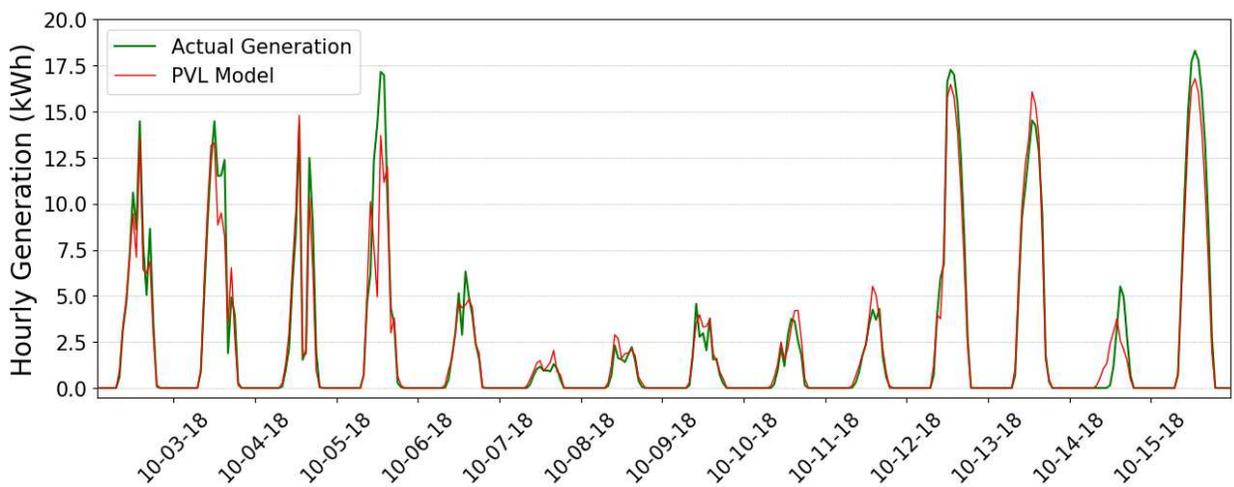
Figure 5.10: Timberline Actual and PVL Generation. Cloudy days (low GHI) are shown, with effective PVL sensitivity in the hourly generation plot.

recorded generation over the period analyzed was 30,767 kWh, with a modeled generation total of

30,859 kWh, a difference of +1.2%. The $\hat{\varepsilon}$ at daily granularity was +7%. Recorded and estimated output is found in Figure 5.11.
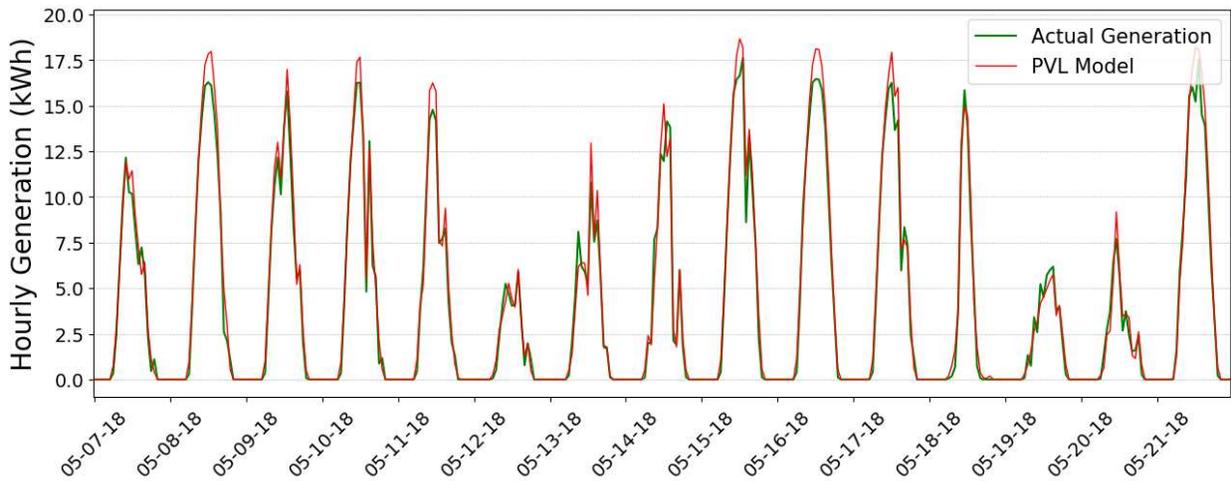
Figure 5.11: Powerhouse Actual and PVL Generation. Hourly generation is shown, with slight PVL overestimation visible during many high-production days.

The Firehouse Alley Garage is a 94 kW$_{DC}$ system flat-mounted system on top of a parking garage in downtown Fort Collins. Total recorded generation over the period analyzed was 91.7 MWh, with a modeled generation total of 94.9 MWh, a difference of +3.5% and plotted below in Figure 5.12.
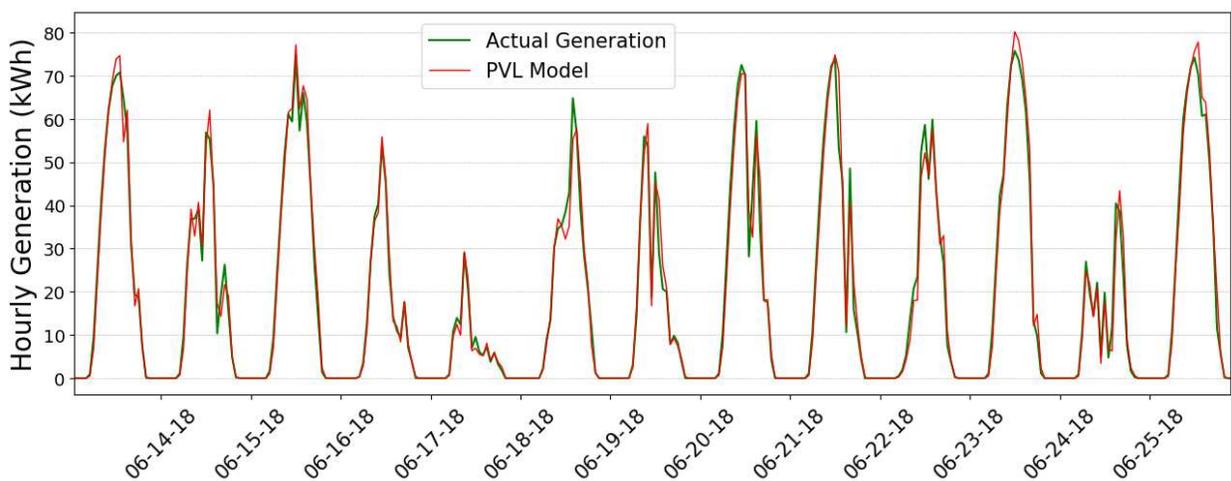
Figure 5.12: Firehouse Garage Actual and PVL Generation. While hourly generation is shown, the $\hat{\varepsilon}$ at daily granularity was 0.62%.

222 Laporte Avenue is a city building that houses several departments of Fort Collins City staff. The building hosts a 103 kW$_{DC}$ at a low tilt angle of 5°. Total recorded generation over the period analyzed was 127.0 MWh, with a modeled generation total of 140,279 MWh, a difference of +10.4%. A daily $\hat{\varepsilon}$ of 5.9% was observed and daily output is shown in Figure 5.13 on the following page.
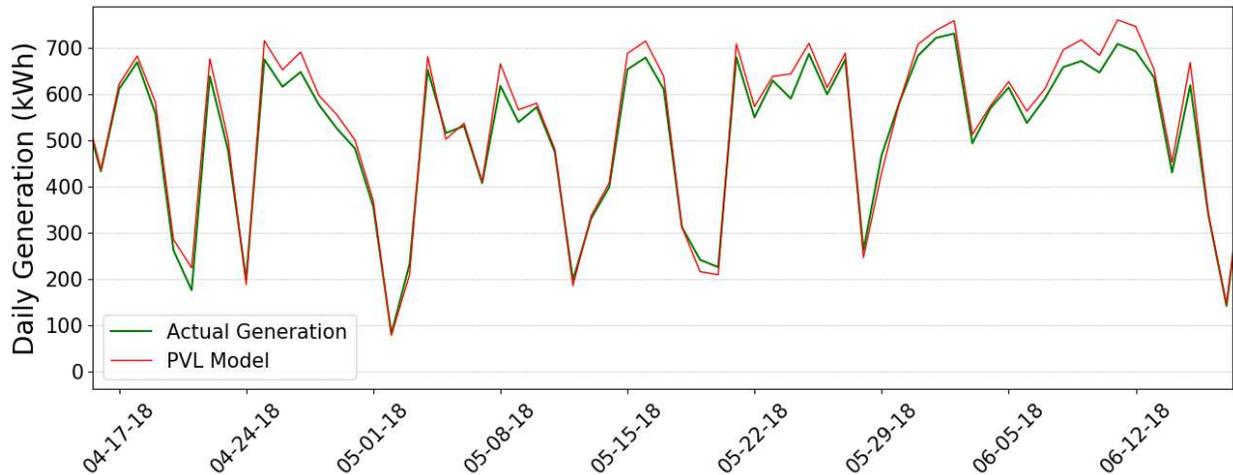
Figure 5.13: 222 Laporte Avenue Actual and PVL Generation. Daily generation comparison is shown.

Colorado State University's Veterinary Teaching Hospital hosts a 220 kW$_{DC}$ system. Total recorded generation over the period analyzed was 312.7 MWh, with a modeled generation total of 327.5 MWh, a difference of +4.7%. A daily $\hat{\varepsilon}$ of 3.6% was observed and daily output is shown in Figure 5.14.
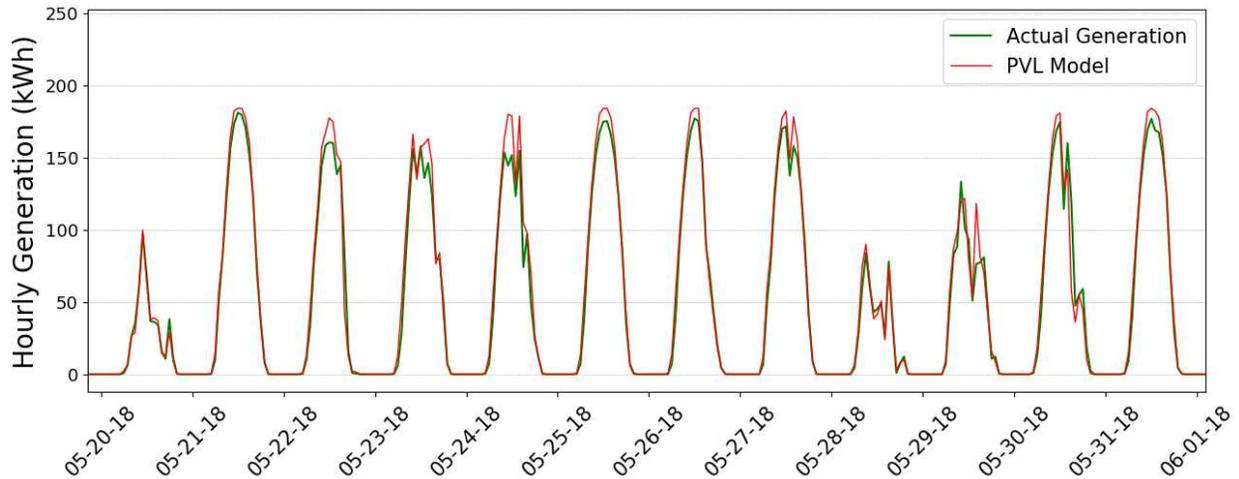
Figure 5.14: CSU Veterinary Teaching Hospital Actual and PVL Generation. Hourly generation comparison is shown.

The CSU Student Recreation Center (SRC) has 544 kW$_{DC}$ array with a maximum AC output of 437 kW$_{AC}$. Total recorded generation over the year was 760.7 MWh, with a modeled generation total of 771.4 MWh, a difference of +1.4%. The $\hat{\varepsilon}$ at daily granularity was -0.2%. The unweighted hourly $\hat{\varepsilon}$ was 1.8%, with hourly $\overline{\omega}$ and $\hat{\omega}$ of 22% and 0.0006%, respectively. Clearly, the average error terms are much more susceptible to outlier generation mis-estimates. Two figures of the SRC array are presented below. Figure 5.15 displays hourly alignment between known and modeled generation while Figure 5.16 displays daily generation agreement.
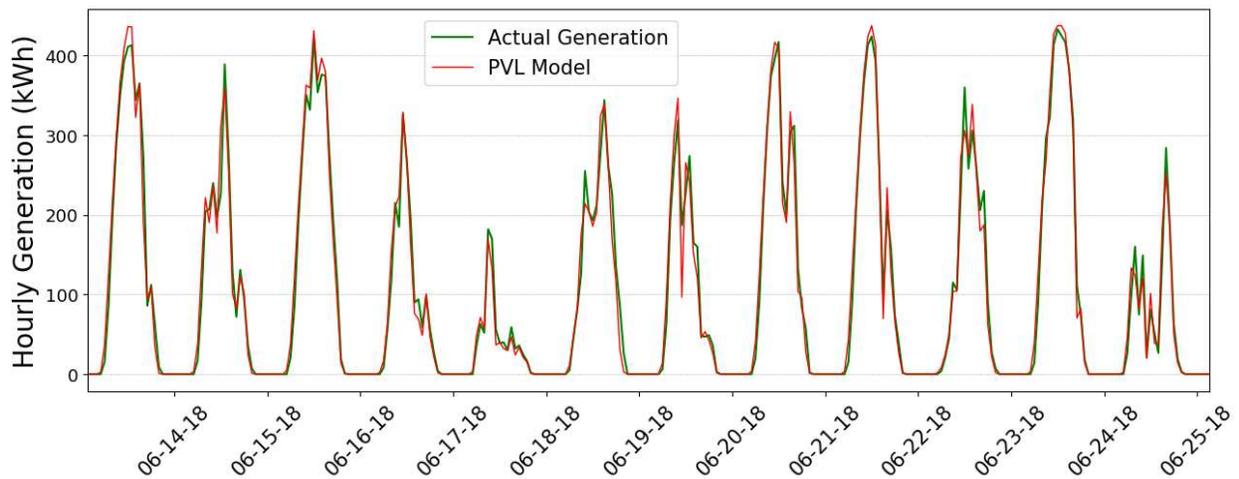
Figure 5.15: CSU SRC Actual and PVL Generation Hourly.

56
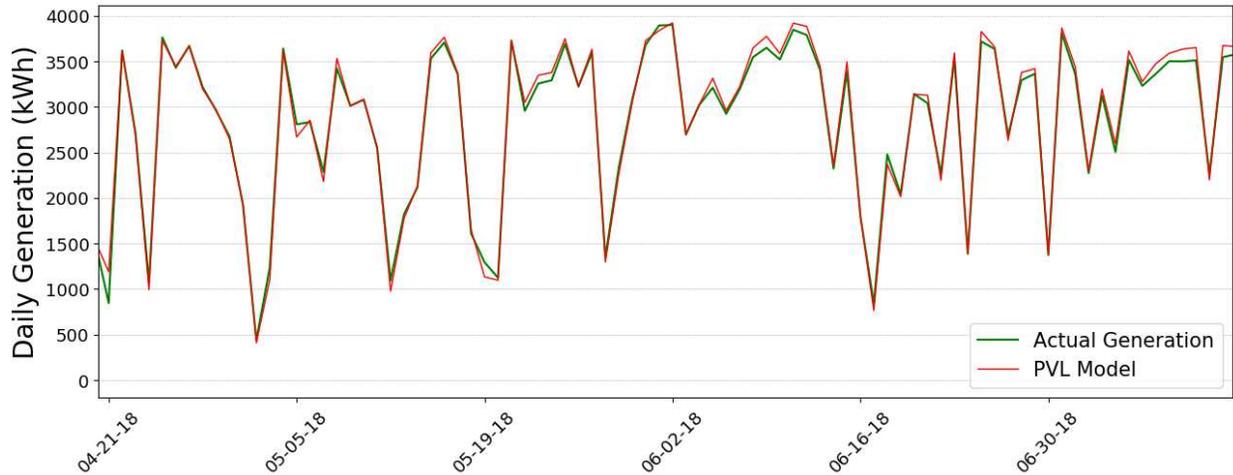
Figure 5.16: CSU SRC Actual and PVL Generation Daily.

As shown, the PVL model both under- and overestimated to actual generation among these systems. Easily identifiable causes are the use of assumed values for module rated power, inverter efficiency, etc. that certainly vary between arrays. While these site-specific inputs could have been used to better model these ten sites, they were not used because these site-specific details will not be available for the majority of modeled systems. The intent of this verification was primarily to build confidence that the stock PVL model with standard assumptions/coefficients was capable of modeling this variety of systems to within reasonable accuracy. Therefore, the ten systems being modeled within ±10% of actual generation was sufficient for this confirmation of viable operation of the PVL model.

5.8 PVL Uncertainties and Limitations

This PVL model does not encompass *every* element that influences physical modeling. Peripheral, minor components of high modeling difficulty are not included for the sake of time and resources, and exemplify the need for alternative methods of estimating PV generation.

For example, light reflected from the ground onto the array, known as albedo, is not modeled and would increase energy production by a few percent. The equation for albedo (E_g) is defined as:

$$E_g \ = \ GHI \ \times \ albedo \ \times \ [\ 1 \ - \ cos(\theta_T)\ ]/\ 2 \tag{5.11}$$

Where θ_T is the array tilt, GHI is the irradiance value, and albedo is a normalized value for ground reflectivity. NREL's PVWatts uses a default albedo input value of 0.2 [28]. For context, a system with a tilt of 45° and an albedo value of 0.2 would see a POA energy increase of nearly 3%. This ground reflected component of POA energy will be added in future work as the CSU-PVL model is refined.

Incorrectly recorded system configurations in the City's records also implant large PVL errors into the results. When premises have multiple sub-arrays, the azimuth and tilt values for the premises appear to sometimes be the average of the two (or three). For example, on a home with a gabled roof and sub-arrays facing both east (90°) and west (270°), an average azimuth value of 180° may be recorded. For houses with several hipped and gabled planes, multiple arrays of various tilts may be represented by one value. Also affecting energy generation was the uncertainty from the unknown tilt/azimuth values that were estimated through satellite analysis. Depending on the direction and magnitude of estimation error, generation may increase or decrease. Section 6.2.1 discusses many of these types of error and provides examples of the occurrences.

Better aging losses also can be factored in to replicate real world wear and tear. Another component of modeling that would certainly trim down generation estimates is shading. Shading is variable for every site and the amount of shading loss also fluctuates through the year due to foliage or structure locations. As will be shown and discussed in the following chapter, the PVL model generally overestimated generation relative to the AMI model. It is speculated these shading losses, inverter operating limitations and other losses that are not fully accounted for in the physical model lead to this model diversion.

Chapter 6

Computational Results Part 2: Validation

In this chapter the agreement precision between the AMI and PVL models are presented. Evidence of adequate precision between the AMI and PVL model, and that the AMI model is valid alternative to the PVL model are presented in this chapter. Individual premise-level estimations will be examined, followed by a discussion of portfolio-wide aggregation of residential systems within the City of Fort Collins.

846 residential premises are included in this two-model computational investigation. Although there were almost 1,300 residences with PV, data requirements trim down this group. First, at least one full year of pre-PV AMI data is required for AMI model operation based on a full year of comparable periods of time. Accordingly, only PV systems installed after Jan. 1, 2016 are included, providing the 2015 year of data. Building off this precondition, arrays installed on newly constructed homes also cannot be currently analyzed due to lack of historical consumption data. Consequently, 438 premises with a combined capacity of 2.1 MW_{DC} are excluded, including 54 systems on newly constructed buildings. An additional 14 residences that have batteries installed at the premise are also excluded. Nonetheless, the 846 systems analyzed had a combined capacity of 5 MW_{DC}, representing 65% of residential systems and 70% of installed residential capacity in Fort Collins. Figure 6.1 displays cumulative growth of residential PV in Fort Collins. The shaded area represents the capacity of systems included in this analysis.
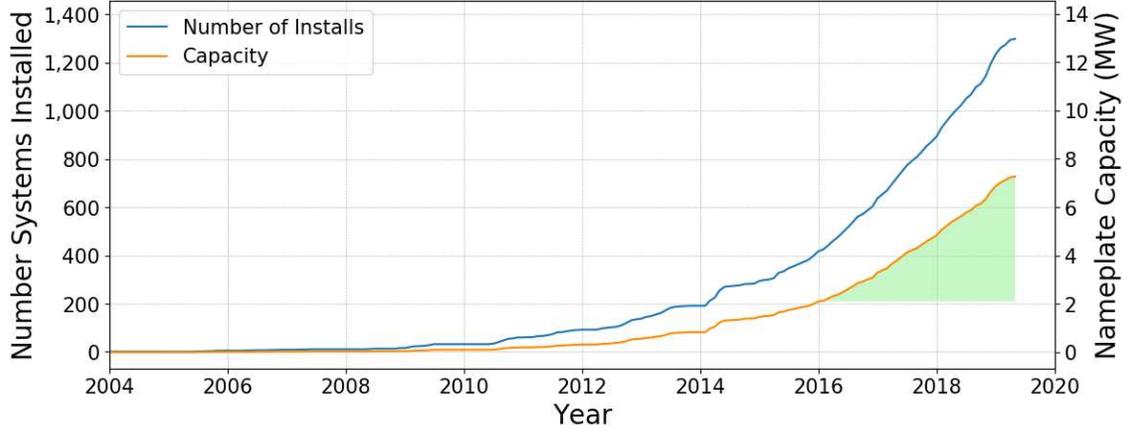
Figure 6.1: Cumulative Residential PV Growth. Historical trend of residential PV in Fort Collins, CO. The shaded area represents the capacity of the 846 premises included in this validation analysis.

6.1 Error Terms

The previous mean and median absolute percentage errors ($\bar{\varepsilon}$, $\hat{\varepsilon}$) rely on employing a ground-truth value in their equations. Now, because both the AMI and PVL models provide estimations, measures of absolute accuracy cannot be calculated. Contrarily, relative error (RE) is a measure of precision: the absolute error between the two estimates relative to the magnitude of the estimates. This relative error provides feedback for how close in agreement the two observed values are. Like the prior error terms, relative error is expressed as a unitless percentage, denoted as Γ and defined as:

$$\Gamma\ (\%)\ =\ 2\ \times\ \frac{AMI_{\mathrm{t}}\ -\ PVL_{\mathrm{t}}}{AMI_{\mathrm{t}}\ +\ PVL_{\mathrm{t}}}\ \times\ 100 \tag{6.1}$$

Where t is a time period in which the premise had a PV system installed. Γ can be calculated at any granularity from 15-minute, hourly, daily, to annually. From the numerator in Equation 6.1, a positive Γ equates to the AMI estimate being greater than the PVL estimate. All premises in this analysis were modeled through March 31, 2019.

The mean relative error ($\overline{\Gamma}$) and median relative error ($\hat{\Gamma}$) are also used in place of $\bar{\varepsilon}$ and $\hat{\varepsilon}$, with $\overline{\Gamma}$ denoted as:

$$\overline{\Gamma} \ (\%) \ = \ \frac{1}{n} \ \times \ \sum_{}^{n} \ \Gamma \tag{6.2}$$

And Γ_{md} presented as:

$$\hat{\Gamma} \ (\%) \ = \ median \ (\ \Gamma_1, \Gamma_2, ... \ \Gamma_n \ ) \tag{6.3}$$

Where Γ are the relative error percentages and *n* is the number of Γ being considered.

A weighted relative error is also utilized for hourly resolution, similar to the previously defined weighted absolute error. The term attempts to minimize errors between the two models that occur during hours of lower generation and amplifies those that occur during greater energy production periods. A weighting factor is applied to each relative error that is proportional to the estimated energy generation versus the maximum estimated generation of the array. In order to calculate the more conservative weighted Γ, the estimated generation for any time step (E_t), is defined as the larger estimated value of the two modeled values. The maximum estimated hourly energy produced by the system over the whole period of analysis (E_{max}), is used as the denominator of the weighting term. As the ceiling of predicted energy generation, all other estimates and associated errors are weighted in proportion of their magnitude versus E_{max}. The definition of weighted relative error for any hourly error is:

$$\omega\Gamma_t \ (\%) \ = \ \Gamma_{\mathrm{t}} \ \times \ (\ \frac{E_{\mathrm{t}}}{E_{\mathrm{max}}} \ ) \ \times \ 100 \tag{6.4}$$

Where t is an hour of analysis between the two models. The mean and median weighted relative error ($\omega\overline{\Gamma}$, $\omega\Gamma_{\mathrm{md}}$) were calculated at hourly resolution for each time-series of the analysis.

6.2 Individual Residential Systems

AMI model precision with the PVL model were first evaluated across unique premises to attempt to quantify meaningful characteristics of the comparisons. For each premise analyzed, the

two estimated generation time-series' were plotted together and the various relative error statistics were assembled.

601 premises each had a daily Γ_{md} between $\pm20\%$, while 526 each had a daily $\overline{\Gamma}$ between $\pm20\%$. 511 premises were part of both $\pm20\%$ groups. Across the 846 premises, the average daily $\overline{\Gamma}$ and Γ_{md} was found to be -14.1% and -8.8%, respectively, toward PVL. Average hourly errors across all premises had a $\omega\overline{\Gamma}$ of -3.7% and an $\omega\Gamma_{md}$ of -1.5%.

An example of this two model plot is shown in Figure 6.2, displaying the daily output from both models for a 9.15 kW$_{DC}$ system. From this array, the average daily Γ of -10.5%, a daily Γ_{md} of -7.5%, and a total generation Γ of -4.5% were computed.
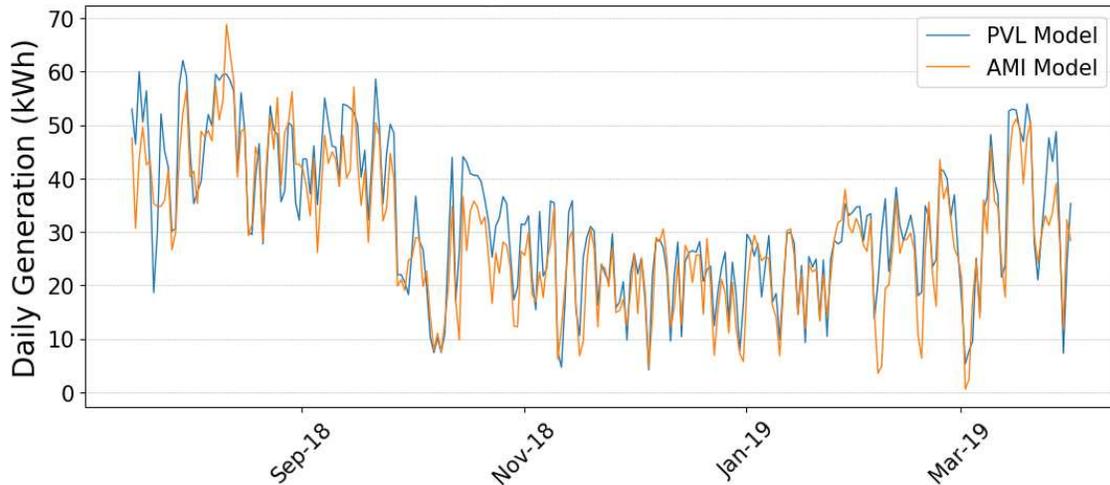
Figure 6.2: Daily Time-Series Outputs Plotted. Daily model comparison for a 9.15 kW$_{DC}$ array installed in July 2018 with AMI (orange) and PVL (blue) estimations.

The following pages contain three additional premise plots to display the variation in duration and precision across premises in this work. Long periods of poor precision are show in Figures 6.4 and 6.5. The causes for divergence are discussed in Section 6.5 and causes for PVL model overestimation were also discussed in Section 5.8.
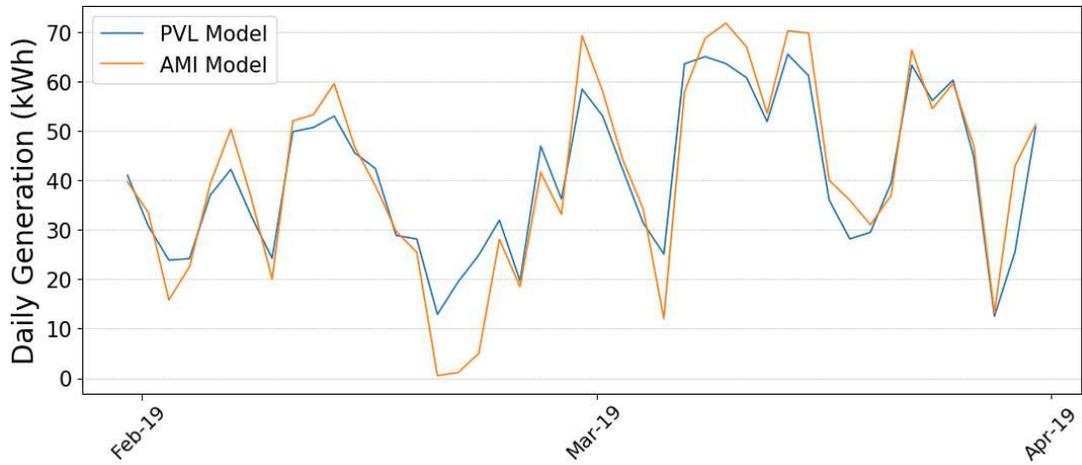
Figure 6.3: 12.2 kW system installed February 2019. AMI (orange) and PVL (blue) model estimations are shown.
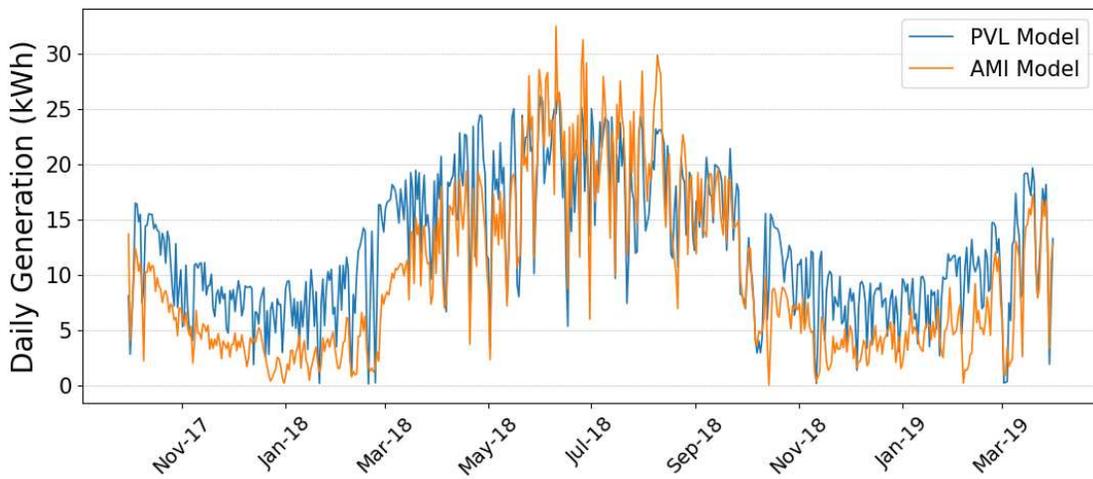
Figure 6.4: 3.6 kW system installed September 2017. PVL overestimation in winter periods is apparent.
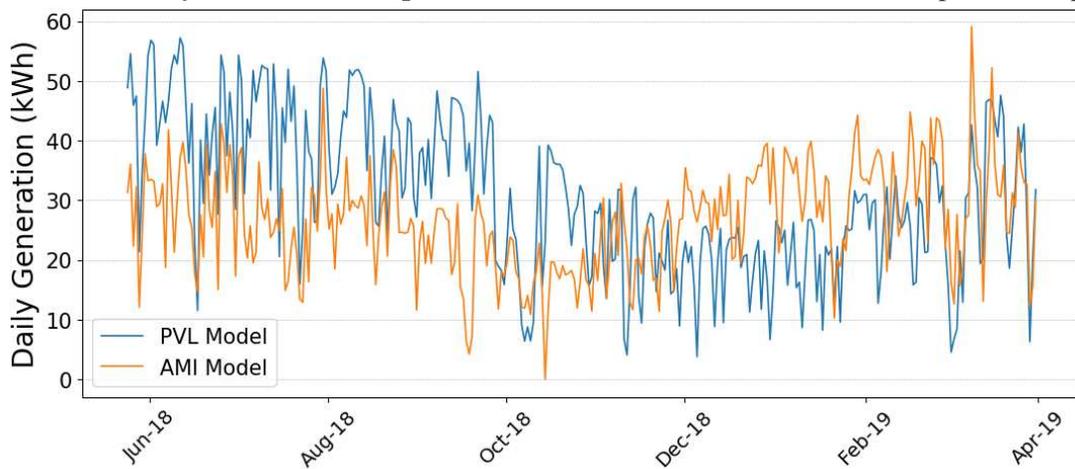
Figure 6.5: 7.9 kW system installed May 2018. Very poor model precision is displayed.

6.2.1 Total Lifetime Error

Integrating the two generation curves provided cumulative values of total generation produced during each system's applicable period of production. These two totals were used to find the lifetime relative error between the two modeled outputs for each system. The individual lifetime errors of each premise provide insight into longer-term precision between the models.

The average lifetime Γ for all premises was -4.5% and a scatter plot of individual lifetime errors is presented in Figure 6.6 below. 584 premises had a lifetime Γ of $\pm20\%$ between models, shown in green. On the right tail of the errors (AMI estimating greater), 39 premises were between 20% and 50% and only four had worse alignment than that. However, the left error tail suggests a skew toward PVL overestimating lifetime error as 25% of *all* premises had a lifetime Γ magnitude of more than 20% toward PVL.
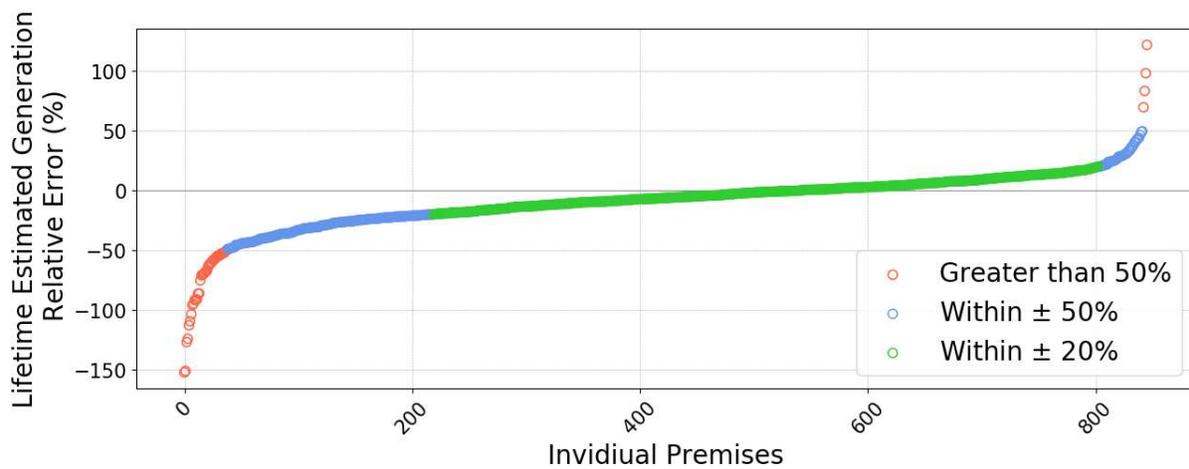
Figure 6.6: Lifetime Generation Sorted Relative Errors. Total relative error provides an analogous metric of precision for the different durations that each array was in operation. The bias toward PVL estimation is visible on the left side of the plot.

The worst errors may be attributed to several factors that are reoccurring themes in this chapter. Speculating, the PVL model can grossly overestimate numerous systems because system misinformation undermines the input parameters. Incorrectly recorded array characteristics such as size, tilt, and azimuth can drastically dictate the PV generation. Further, changes to the system such as array modifications or upgrades may not have been recorded in the City's records. Additionally,

two key external factors are only partially addressed by the PVL and other physical models. As discussed in Chapter 5.6, snow coverage reductions must be incorporated during days of snow fall and, more importantly, periods after snow events when the weather is favorable but the panels remain inundated by accumulated snow. However, these modeling reductions are complex and irregular across events; a general reduction has been built into the PVL model but it has only limited sensitivity to different snow and array conditions. Critically, modeling array shading from buildings or trees is an intricate and tedious task by and of itself, and accounting for variable shading reduction at hundreds of premises is not feasible in this work or for a utility.

Because the AMI model accounts for reduced generation due to shading, as it uses measured data, many premises with shading would have lower AMI estimates. Since the PVL model does not currently consider shading, this may be a major reason for the numbers of premises with greater PVL estimates.

From the worst premises in red on the left-side of Figure 6.6, spot checks were conducted to verify installed configuration versus recorded system configured. Figure 6.7 illustrates one major cause of modeling error, when an array is split between multiple roof pitches and the recorded tilt/azimuth are not accurate for the installed system. The house on the left had east and west facing sub-arrays, and a the recorded value appears to split the difference between these. The house on the right has three arrays: ones facing east and west, and a larger one facing south. As shown, the recorded azimuth for the entire array was due south. Shading is also evident from a south-easterly direction.

Other recording errors are shown on the left house in Figure 6.8, where the recorded azimuth was a completely different direction than the installed system. While the system would have been better suited with the 140° orientation, field conditions sometimes lead to installed arrays on less-optimal roof pitches that are clear of obstructions and easier to install the system.

In the right side of Figure 6.8, an array with considerable shading is shown, which would be another leading cause of PVL providing much greater estimation estimates as evidenced by the left-side of the lifetime error plots.

Figure 6.7: Sub-arrays with Incorrect Recorded Values. Sub-arrays on different roof planes are shown with the recorded azimuth value. Yellow arrows depict actual and recorded array azimutsh and the red dashing denotes roof ridges. Both premises were located in the left red grouping on Figure 6.6.

Figure 6.8: Causes of Greater Estimation by PVL. Field conditions do not always match recorded system configuration, shown by the left premise. Array shading is a serious issue at other premises. Yellow arrows depict actual and recorded array azimuths and the red dashing denotes roof ridges. Both premises were located in the left red grouping on Figure 6.6.

In summary, these exogenous components of generation modeling likely lead to greater estimation in the PVL model's values for many systems. Modeling limitations for the PVL model were also discussed in Section 5.8.

6.2.2 Monthly Aggregation

As outlined previously, modeling individual premises at a monthly timespan is a beneficial exercise that is relatable to consumers with rooftop solar. Each premise's two modeled outputs were separately aggregated by calendar month for performance comparison and monthly relative error was calculated.

For insight into the monthly agreement, the average of the monthly Γ for each system was calculated from the set of available individual monthly Γ present through the lifetime for each system. Further, from each premise's set of monthly Γ, a 90% empirical confidence interval of monthly error was constructed for each premise. That is, an empirical confidence interval was constructed from the distribution of the premise's monthly errors. The following plot contains the average Γ and 90% empirical confidence interval of monthly error for each premise.
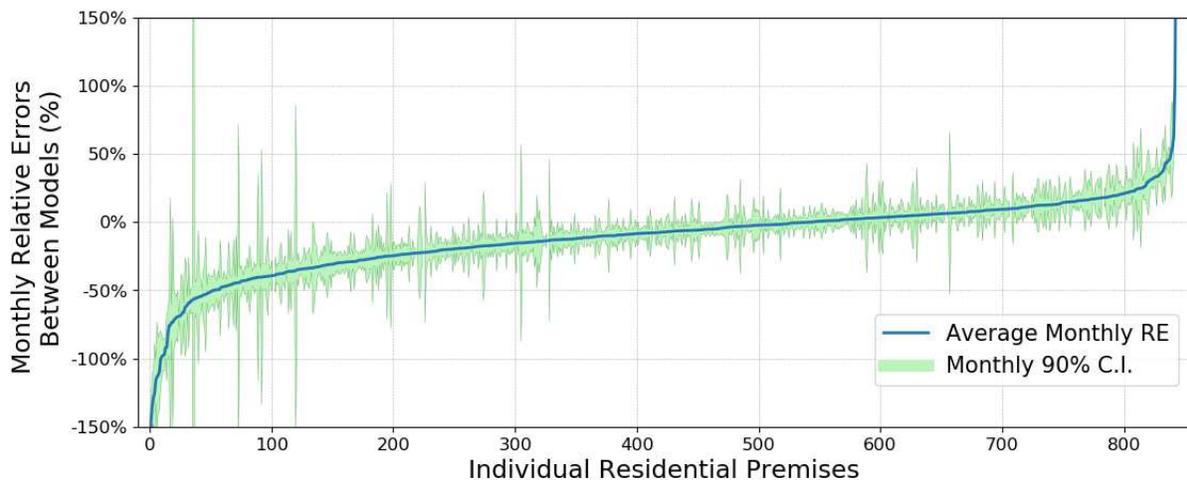
Figure 6.9: Average Monthly Error Between Models. The x-axis is ordered by ascending average monthly error. The green shaded area is the 90% empirical confidence interval of monthly error for each premises.

A negative Γ corresponds to the monthly PVL model estimate being greater than the monthly AMI estimate, and vice versa for positive error values. 544 premises had negative average monthly Γ while 301 had positive average monthly errors, indicating a slight bias toward the PVL model estimating higher generation than AMI. The premises in Figure 6.9 are ordered by ascending aver-

age monthly relative error. As the lifetime of data varies between systems, the number of months in which production estimation data is available also varies. A system installed in April 2016 has 35 monthly estimation sets and 35 monthly REs, while a system installed in January 2019 would only have 3 estimation sets and 3 errors.

From the figure, there is a gently sloping monthly error curve bounded by two elbow regions on the curve extremities where the error greatly increases. Just as with the lifetime error plot from Figure 6.6, in Figure 6.9 the precision deteriorates on the left side of the error curve as PVL persistently estimates higher generation. Visually there are months with very large empirical intervals, these tend to lie on the left-side of the plot. These inconsistencies are assumed to be from the same fundamental reasons as previously discussed: system misinformation, along with unaccounted shading, variable snowfall losses, and other minor losses.

6.3 Portfolio-Wide Residential Aggregation

Aggregation of many residential systems is more insightful from the perspective of a distribution operator such as Fort Collins Utilities. This section discusses the relative errors between the two models aggregated to both hourly and daily resolution across all premises and grouped by calendar month. When results are grouped monthly, months are color-coded by a qualitatively-derived "season" as found in Table 6.1. All premises in this analysis were modeled through March 31, 2019.

Table 6.1: Seasonal Categories.

Season	Corresponding Months
Winter	December, January, February, March
Spring	April, May
Summer	June, July, August, September
Fall	October, November

The following box and whisker shapes plotted each represent all of the days or daylight hours of each calendar month, as specified. The whiskers represent the 5th and 95th percentiles of the

data. That is, if a boxplot of daily Γ are presented, the whiskers represent the inner 90th percentiles of the month's daily Γ. Presented below, the boxes show the 25th, median, and 75th quartile of each month's set of daily Γ. The lower box edge represents the 25th percentile, the middle line represents the median, and the upper box edge represents the 75th percentile of the data.
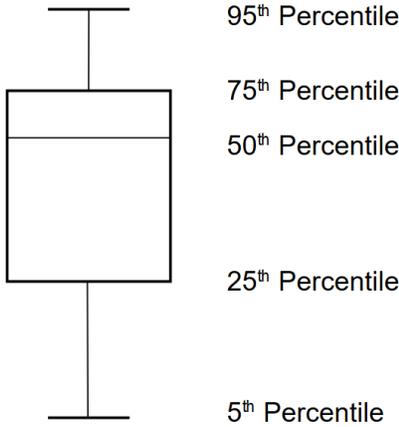
Figure 6.10: Boxplot Diagram. Daily or Hourly Γ's are the input data used to construct each month's box and whisker figure, shown in Figure 6.11. The whiskers represent the inner 90th percentile of each month's errors.

The months of January 2016, February 2016, and March 2016 were discarded from analysis because these months had fewer than 30 premise sampling groups, which is typical of a minimum number of samples in statistical analysis. January 2016 only had 7 premises, February 2016 had 10 premises, and March 2016 had 29 premises. In contrast, months occurring in 2019 incorporated over 800 PV systems for modeling.

6.3.1 Daily Aggregation

The daily generation time series from both models for each premise were aggregated to create a portfolio-wide daily time series of generation. Fundamentally, this plot is the purest visualization of relevant modeled data for a utility: the modeled generation of a portfolio of assets. Generation increases as more systems are interconnected. As mentioned, in Fort Collins there are were also

438 residential PV systems excluded; their generation (2.1 MW$_{DC}$) is not represented on the plots or in this analysis.

Figure 6.11 depicts the increasing generating capacity with seasonal trends as systems are continually installed. Generation per day can be normalized by installed capacity to highlight the periods of the year which produce more energy. This allows a comparison between different summer months across the three years shown in the figure below. The normalized generation potential per month is displayed as the purple line in the top of Figure 6.12.
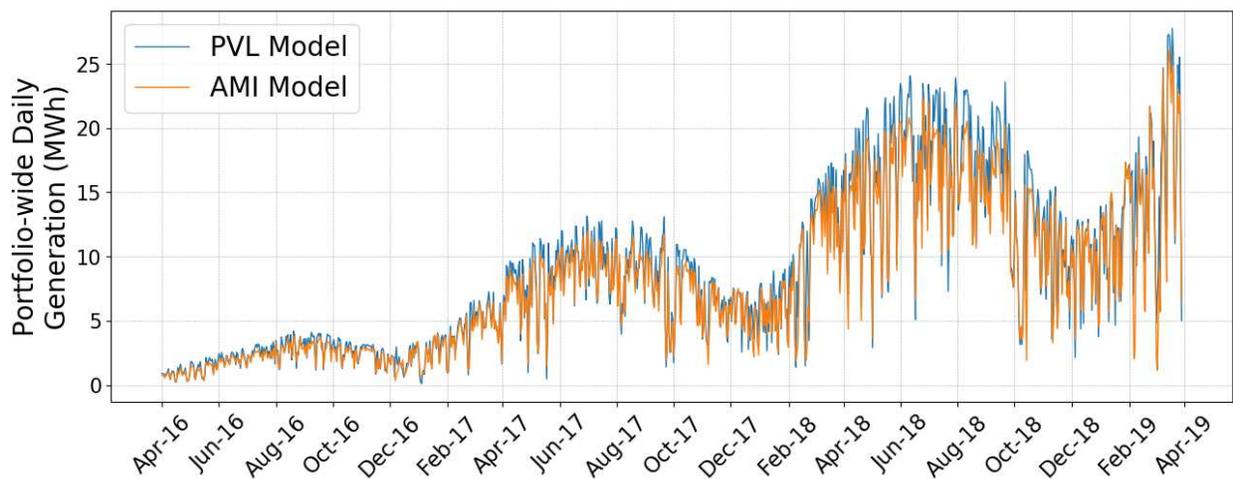
Figure 6.11: Daily Aggregated Portfolio Generation. Through the end of Q1 2019, a cumulative 846 residential arrays are considered with a final cumulative capacity of 5 MW$_{DC}$.

The plot below displays the individual daily Γ between estimating system-wide PV production from only AMI data versus estimating system-wide PV production from a physical model. Daily Γ's are grouped by calendar month. The y-axis shows the percent relative error with positive errors representing the AMI model overestimating compared to the PVL model. The bold numbers across the top represent the number of premises available for analysis during that month. The right y-axis shows the estimate of PV production normalized by installed capacity during each month. The boxes are ordered left to right from least to highest monthly normalized production. The average daily Γ was -7.7% from 1,093 days of modeled generation comparisons, and is displayed as the red horizontal line.
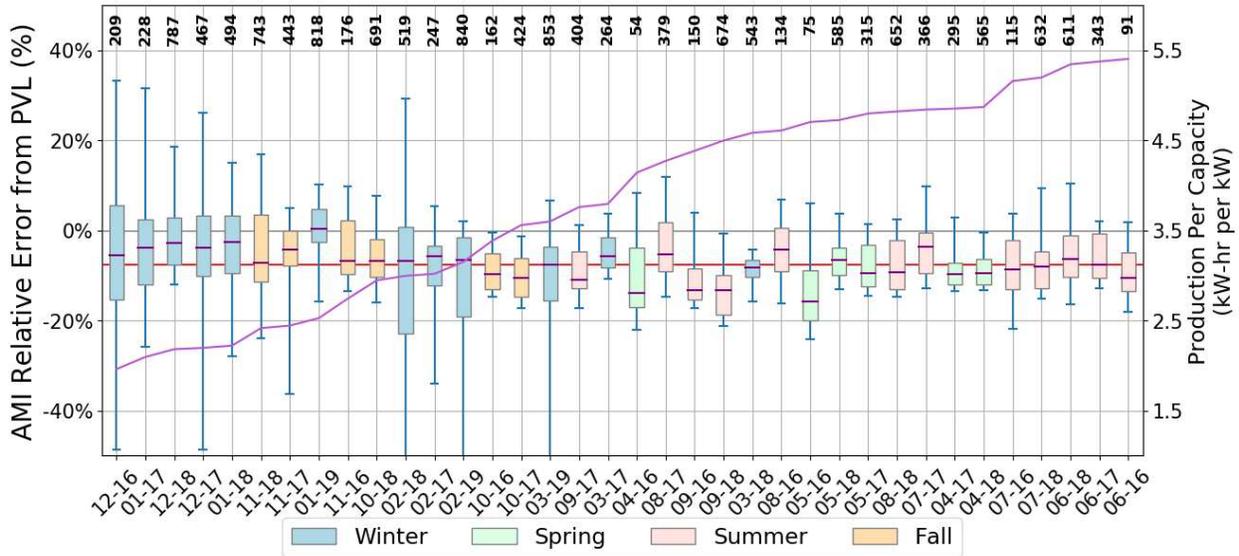
Figure 6.12: Portfolio-wide Daily Aggregation Errors Grouped by Month. Months are ordered estimated normalized production. The blue interval represents the 90% empirical interval of each month's daily errors.

Portfolio-wide precision is confirmed at the daily aggregation level. In every month of analysis, 50% or more of all days in the given month had an inter-model precision within ±20%. 10 months had at least 50% of all days within ±10%. Further, 21 of 36 months contained an inner 90% empirical interval of modeled days that were within ±20% of each model. 9 of the 10 highest producing months were within this group, and this consistency, especially in higher producing months, builds confidence in the AMI model's capabilities.

Even at good precision, it is clear statistically and visually from the boxplots that most days exhibit PVL overestimation bias. 25 months contained 75% or more days that had negative Γ's, meaning bias was toward PVL estimating greater generation relative to the daily output of the AMI model. Among these 25 months were the 12 most-productive solar months. All but one month (Jan. 2019) had a negative Γ_{md}, and the average monthly Γ_{md} was -7.5%.

In another measure of precision, 28 months had at least 50% of days spread to within 20% of each other, calculated as the difference in Γ between the 25th and 75th percentile. Further, 12 months had at least 90% of days spread within only 20% of each other, calculated as the difference

between the 5th and 95th percentile values. Four of these were summer months, four were spring, and two each were fall and winter months.

Another observation is that fewer numbers of premises aggregated do not impact the precision between models. April 2016 (54 systems), May 2016 (75 systems), and June 2016 (91 systems) exhibited 90% empirical ranges of 30%, 30%, and 20%, respectively. That is, 90% of all days in those months had Γ agreement within those ranges, which is not unreasonable precision.

A trend emerges that winter months have less agreement and alignment between the two models. 8 of 12 winter months contained a 90% interval that spanned outside of $\pm20\%$ Γ precision, and the six worst (widest) 90% intervals belonged to winter months. Intuitively, winter months also represent periods that account for some of the lowest amounts of normalized PV generation. The purple line displays normalized production estimated for each calendar month. From the figure, correlation of months of lower production having larger error bounds is observable; higher production months on the right of the figure tend to have tighter error intervals.

This is consequential because loads are higher in summer months and more solar generation offsetting some load has greater significance. To exemplify this, Platte River Power Authority encountered the 2018 peak load hour on July 10, 2018 from 5-6pm at 686 MW [60]. The lowest monthly peak load hour occurred during April 6, 2018 at 412 MW. Winter 2018 monthly peak loads never exceeded 479 MW, 30% less than the annual peak seen in the summer. With these higher loads, predicting load offset due to PV generation during summer months then is inherently more significant for managing generation assets and operational awareness.

6.3.2 Hourly Aggregation

This section discusses the hourly relative errors between the AMI and PVL models. The hourly time series of the two generation estimates for each premise were aggregated to create portfolio-wide hourly time series.

Because of the weighting, hourly errors in lower-producing months are minimized while errors in high-generation periods are not. For the aggregated weighting factor, the normalized generation

72

per installed capacity is used instead of generation because the capacity changes over time, and hourly generation potential changes drastically from 2016 to 2019.

The plot below displays the individual hourly weighted relative errors ($\omega\Gamma$) between estimating portfolio-wide PV production from only AMI data versus a physical model. Hourly $\omega\Gamma$ are grouped by calendar month. The y-axis shows the percent relative error with positive errors representing the AMI model overestimating compared to the PVL model. The bold numbers across the top represent the number of premises available for analysis during that month. The boxes are ordered left to right from least to highest monthly normalized production. The hourly $\omega\overline{\Gamma}$ was -3.0% from 12,831 hours of modeled generation, and is displayed as the red horizontal line.
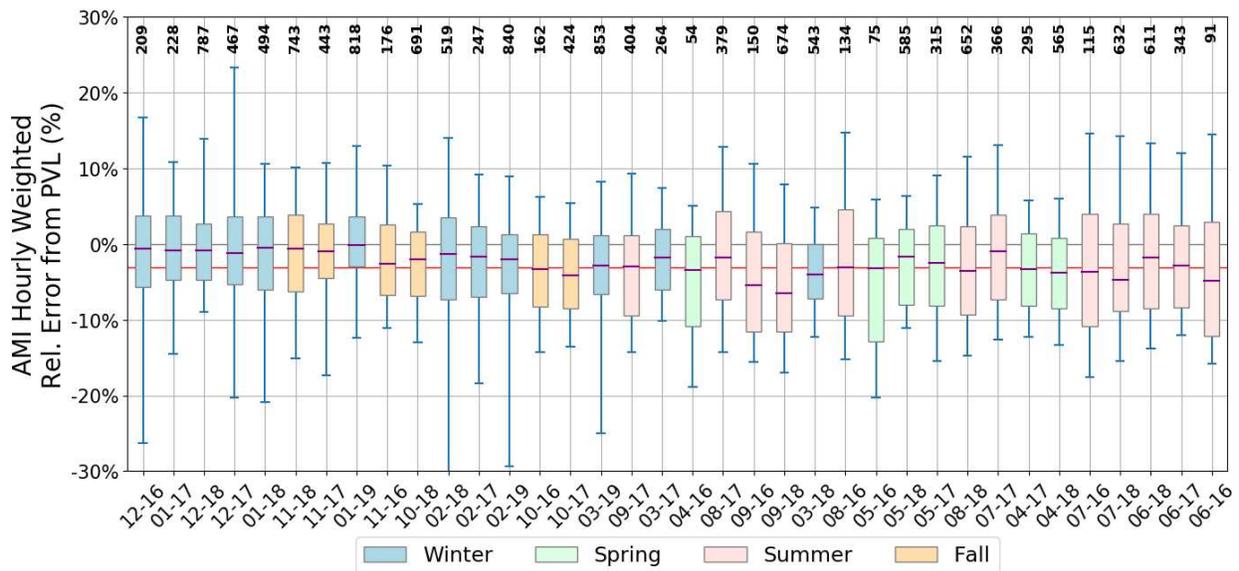
Figure 6.13: Portfolio-wide Weighted Hourly Aggregation Errors Grouped by Month. Months are ordered estimated normalized production. The blue interval represents the 90% empirical interval of each month's hourly errors.

Because weighted errors are used with normalized generation, descriptive statistics regarding the two models' precision are not as straightforward. However, it is still possible to gain insight into the direction of bias and precision among different periods of the year.

The $\omega\Gamma_{\mathrm{md}}$ for every month was negative, reinforcing that was PVL institutionally estimating greater generation than AMI. 29 months had at least 90% of hours within $\pm20\%$ $\omega\Gamma$, adding to

confidence in precision between models at the most important hours of generation. The PVL bias is evident by 16 upper bounds (95th percentile) that do not exceed 10% $\omega\Gamma$ toward the AMI model, while only one lower bound (5th percentile) does not exceed -10%$\omega\Gamma$ toward PVL (December 2018).

The propensity of winter months being the most difficult to model is reinforced by the hourly boxplot. Comparing 90% percentile spread magnitude, the worst six months are winter months, and six of seven months with a 90% percentile spread that exceed $\pm20\%$ $\omega\Gamma$ were winter months. Although much less obvious due to the weighting scaling, some lower generation months on the left of the figure have noticeably larger error bounds than higher production months on the right.

The purpose of comparing hourly errors is for consideration of the model to be useful at sub-daily intervals for a utility. This initial work shows that normalized, weighted errors are lower, thus a perceived achievement of better agreement between models. However, raw hourly errors inherently fluctuate and deviate more, and hourly granularity and may be a less useful modeling interval due to higher uncertainty. What can be concluded is that at both daily and hourly resolution, the AMI model performs with better precision at higher producing periods with the PVL proxy. Performance difficulty with snow/winter months will be discussed in the following section.

6.4 Snowfall Performance

Apparent difficulty of modeling precision in winter months may be less of an issue with the AMI model and more to do with performance limitations in snowfall correction from contemporary physical models.

The first two figures below show actual generation, the PVL and AMI modeled outputs, and solar irradiance during two snowfall events. Even with snowfall reduction in the PVL model, the ability of the AMI model to better predict generation after snowfall is evident. In Figure 6.14 the day of snowfall (Feb. 6, 2019) is modeled relatively correctly by both models. This is expected because the weather during a snow storm has very low GHI and temperature values that both models would pick up on the lack of generation.

However, due to snow coverage, the following day exhibits the divergence in model accuracy. Because it was a sunny on Feb. 7th (GHI plotted in dashed pink), the *corrected* PVL model still estimates up to around 14 kW of generation during mid-day. In fact, building load was not offset by nearly any PV generation - thus the difference between delivered energy and comparable delivered energy was minimal. This enabled the AMI model to predict very low generation with relatively good accuracy to actual generation.
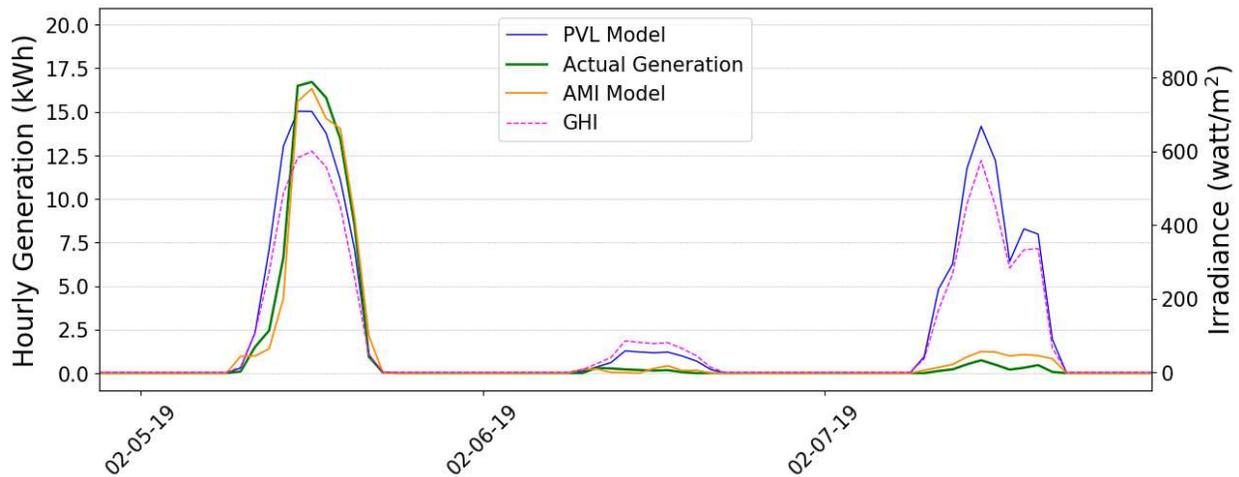
Figure 6.14: Comparison of Model Performance with Snow Coverage at Timberline Array. Hourly recorded (green) and estimated generation (orange, blue) at the Timberline array, with solar irradiance (dashed pink). On Feb. 6, 2019, Fort Collins received 2.8-inches of snow. Snowfall correction factors as discussed in Section 5.6 have been applied to the PVL estimate.

Figure 6.15 displays two additional days of panel coverage and the modeling impacts at the Timberline array. On March 2, 2019 Fort Collins received more than 7-inches of snow. The PVL model estimated large amounts of generation on the following two days because the weather was fair on March 3rd and sunny on March 4th. The hourly AMI model's estimate (orange) aligns much better with recorded generation (green) than the PVL model.
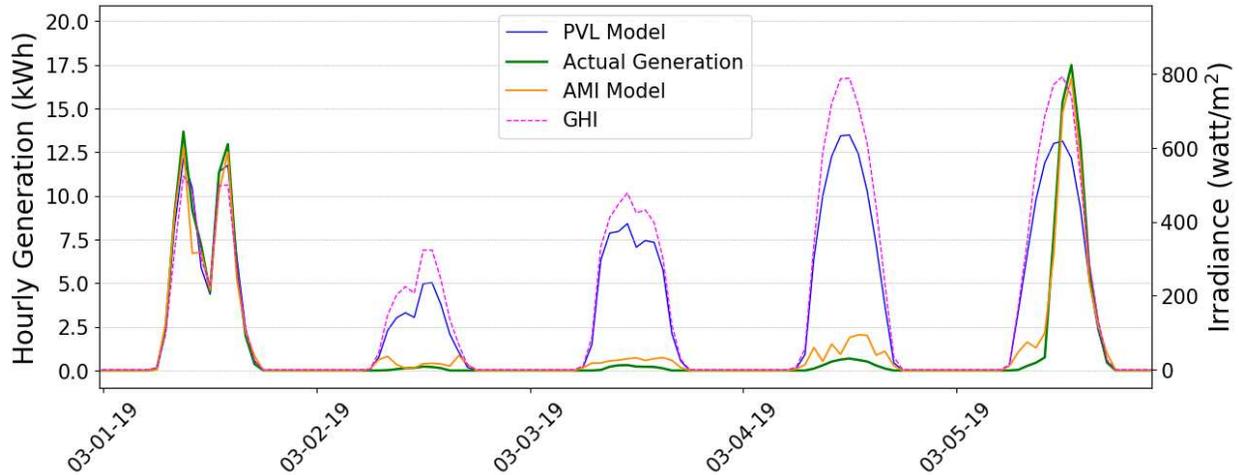
Figure 6.15: Comparison of Model Performance with Snow Coverage at Timberline Array. Hourly recorded (green) and estimated generation (orange, blue) from Timberline, with solar irradiance (dashed pink). Snowfall correction factors as discussed in Section 5.6 have been applied to the PVL model.

Shown below in Table 6.2 are annual statistics of days that triggered snowfall correction calculations, with 2019 data only including the first quarter of the year. From 2016 through Q1 2019 (the period compared between AMI and PVL), only 83 of 1,176 possible days triggered snow correction action by the PVL model. At worst, snowfall correction occurs at up to 10% of the days of a given year (as seen in 2016). Because generation is low in winter months, these 10% of days likely generate less than 10% of annual generation. This means that while snowfall prediction is important, a sensible method of correction can mitigate a reasonable portion of inaccuracy within a physical model.

Table 6.2: Annual Snowfall Statistics and Correction Events. NOAA records for Fort Collins, Colorado and number of unique correction events.

Year	Days of New Snow (> 1 inch)	Avg. Depth (inches)	Accumulated Snow Days (> 3 inches)	Avg. Depth (inches)	Unique Events
2015	16	2.86	40	4.31	43
2016	10	5.47	42	5.81	44
2017	13	2.61	10	4.40	17
2018	12	2.69	9	3.66	14
2019*	6	3.14	6	4.17	8

* denotes partial year of data for 2019.

76

While physical models readily model the reduced insolation occurring *during* snowfall, these models have difficulty modeling the reduction in PV output after snow events when solar insolation is high but the panels are covered by accumulated snow. These modeling reductions are complex and irregular across events. The inability of this basic snow-correction in the PVL model is apparent from the time-series generation plots. To further visualize the correlation between snowfall and large PVL overestimation errors, aggregated hourly $\omega\Gamma$ have been plotted against days of snowfall. The weighting during low-production winter months decreases the negative relative error; they would actually be even more pronounced if not weighted. The upper blue plot displays the aggregated hourly weighted relative error, with negative spikes of PVL estimating greater generation shown. The lower orange plot aligns daily snowfall. The visual trend between snowfall and greater disagreement between models is apparent.
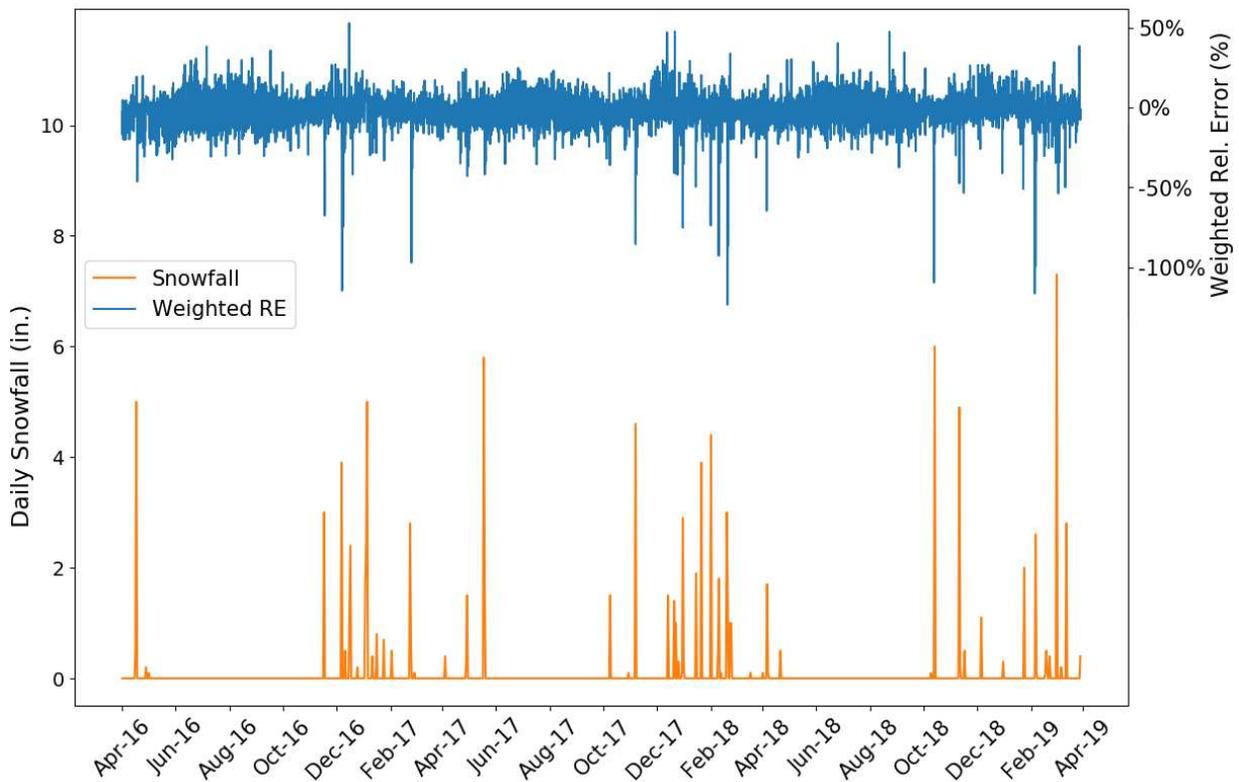
Figure 6.16: Snowfall Correlation with Model Disagreement. Aggregated hourly $\omega\Gamma$ plotted in blue above, with daily snowfall (inches) plotted in orange.

Although the AMI model does not match periods based on precipitation, it is responsive to the fact that delivered energy is higher on a sunny day when snow is covering the array panels than the delivered energy *would* have been had the array producing power. The AMI model intrinsically captures reduced generation from structural factors – shading, soiling, or general aging degradation – and provides some insight into the snow reductions. Because snowfall events are discrete events, they are more quantifiable and discernible phenomena than a complex shading metric.

6.5   Model Limitations and Uncertainties

The total reliance on AMI data leads to the exclusion of premises if there is not 12 months of pre-PV data available. The consequence is that older systems or systems on newly constructed buildings cannot be analyzed as the method currently stands. Although a proxy for pre-PV AMI data perhaps *could* be constructed from similar premises, it would increase involvement and bog down the process. Relief from input-intensive modeling requiring specific known data was one of the original intents of conducting this work and must be kept in view.

A rebuttal to this issue notes that solar is being installed at a rapid rate in Fort Collins, as shown by Figure 2.5 and Figure 6.1. Most future systems will be installed on already-built structures, meaning that the number of premises that cannot be modeled will be an ever-decreasing fraction of the total asset pool. As time progresses, these few hundred legacy systems will comprise a smaller minority of solar systems in the City.

One way to avoid this exclusion going forward is for new premises to be required to provide generation data access from internet-tied inverters at these net-metered arrays. Automated data collection architecture will be critical to minimize the effort and time required to work with these datasets. It will be up to distribution utilities to weigh the privacy and effort considerations for implementing some type of data sharing requirement.

Uncertainty was also derived from external sources to the model. The model operates under the assumption that energy consumption habits do not significantly change once PV is installed. As addressed, some research suggests that this assumption may not always hold true, but in some or

many instances there may not be significant changes. Efficiency upgrades or changes in appliances would also externally impact the consumption patterns of the premise, and the occurrence of these events was not known. Additionally, there were two other interventions that would may cause the premise's energy usage to change that were not accounted for in this current methodology. First and foremost, it was not known when occupancy changed in any of the homes with PV. As homes are rented or sold, new occupants indubitably will change the energy consumption patterns found in the AMI data and lead to greater uncertainty and error in the modeling. Regardless of whether this information is known, actually accounting for it may eventually prove to be too complicated to be justified.

A second external force that influenced home energy consumption was the November 2018 implementation of Time-of-Day (TOD) electricity rates in Fort Collins. Changing the rate structure from fixed prices to a TOD price is inherently designed to change energy consumption patterns.

Shifting electricity use from on-peak to off-peak times would greatly impact hourly model performance, and accounting for these possible behavioral shifts was not factored into this AMI procedure. It is possible that daily aggregation of estimated generation from AMI data is more resilient to the effects of inter-hour pricing than hourly modeling. Understanding the behavioral impact of this price change provides numerous other areas of future work with the City of Fort Collins.

Additionally, this work is not exempt from data quality concerns. There were rare instances were an individual smart meter was malfunctioning or the data acquisition system did not record the data to the premise account properly. Missing fragments of AMI data are obviously detrimental on an individual premise modeling level but are likely washed out when many systems are aggregated together. The addition of a function to identify and interpolate missing data is appropriate. Another beneficial data quality exercise would be exploration of premises without recorded solar installs that have AMI data containing *received* kWh values.

Moreover, while this analysis was conducted on a significant number of premises, the investigation only occurred upon one geographical location. Partnering with other distribution operators

that have installed smart meters could be pursued for future work to characterize robustness across areas.

Although there are limitations and various layers of uncertainty, the qualitative agreement with the physical model and to known generation data speaks to the model's value. Whereas individual system accuracy varies, the AMI model's ability to estimate portfolio-wide daily PV generation to precision within $\pm$ 20% is a strength, especially in higher generation months. While there is surely room for refinement, this AMI model does provide fundamental insight into a method to disaggregate PV generation from net-metered load data.

Chapter 7

AMI-based PV Forecasting

The PV forecasting arm of the AMI model was in the preliminary stage of development at the time of writing. Early findings are discussed and several examples of forecasted generation are presented. At this phase of development, the initial results appear promising and future work in this area is needed.

7.1   Preliminary Model Runs

Known generation from House Z was compared to forecasted generation from the AMI model. As seen in the other estimates from this premise, the AMI model underestimated compared to actual energy produced. Over the two forecasted days, the model estimated a total generation of 45 kWh while the system actually produced 59 kWh, and the average hourly error was -20%. Figure 7.1 below displays the two day forecast.
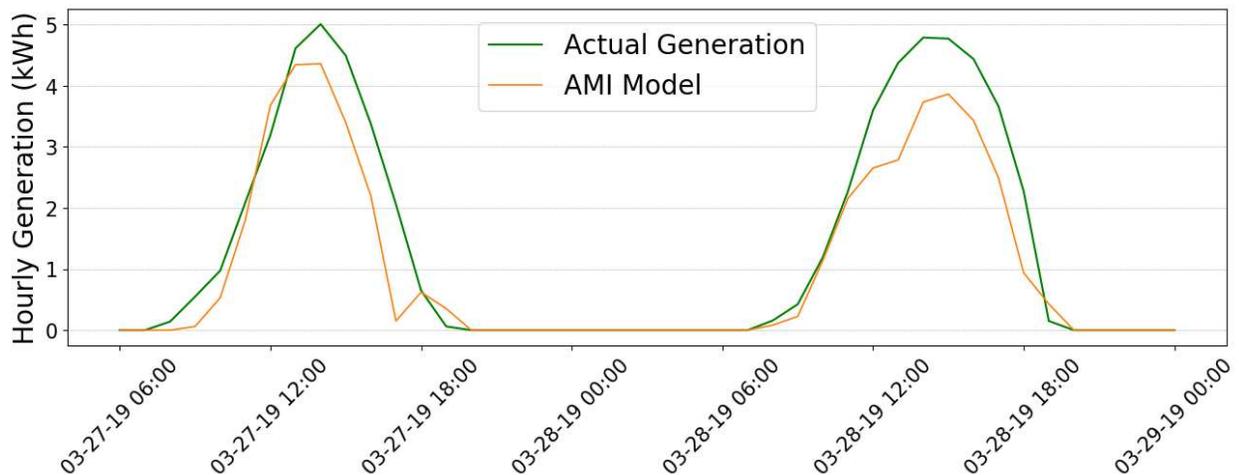
Figure 7.1: Actual and Forecasted Generation from House Z. Two days of recorded and forecasted generation from a 48-hour weather forecast.

As previously discussed, actual generation from most premises is not known, therefore the PVL model is useful for comparison to the AMI model's output. Using the same forecasted weather inputs from the NAM, the two models were run through numerous sets of 48-hour forecasts. Two of these forecasting runs are displayed below in Figures 7.2 and 7.3.
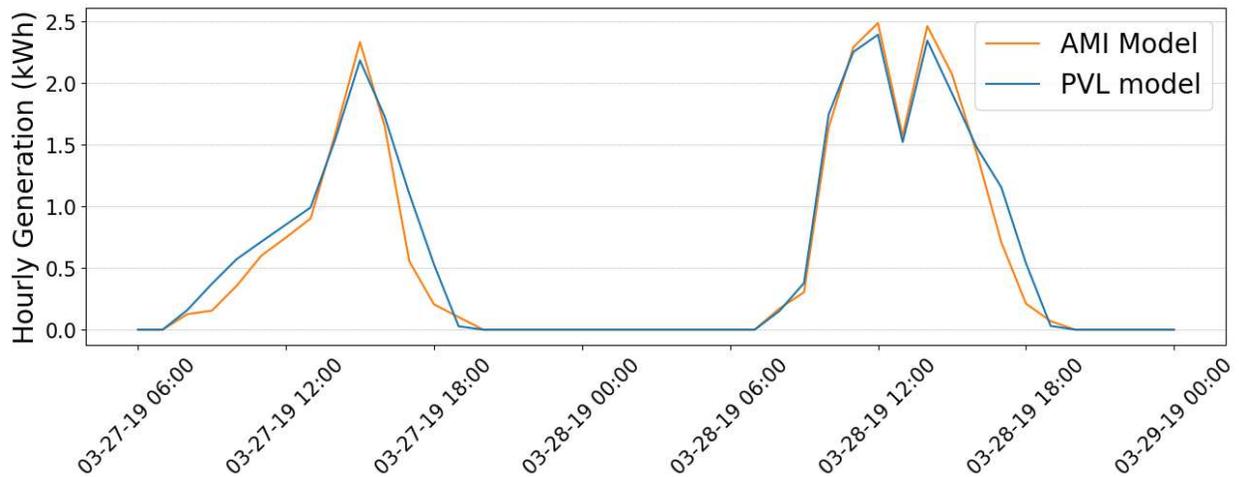
Figure 7.2: 48-hour generation forecast from AMI and PVL models. A 3.25 kW$_{DC}$ array using both models to forecast generation for the next 48-hours.
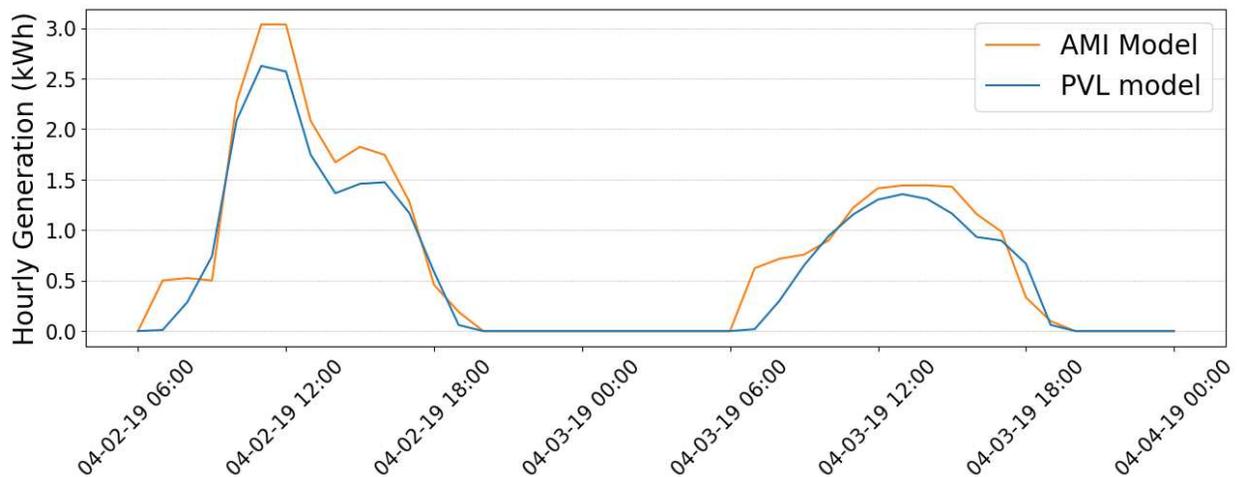
Figure 7.3: 48-hour generation forecast from AMI and PVL models. A 5.1 kW$_{DC}$ system using both models to forecast generation for the next 48-hours.

Actual generation from only House Z is available in real-time for use with forecasted weather. While initially modeling efforts may suggest precision between the PVL and AMI output, the

quality of the inputted forecast data is not known. The accuracy of these models against actual generation has not been studied enough and should be pursued further.

7.2   Portfolio-wide PV Forecasting

Aggregated forecast comparisons between the PVL and AMI models with the fleet of systems in the City has not yet been conducted. This thesis work focuses primarily on the creation of the AMI methodology, with resulting comparisons based on actual generation and the PVL estimates. These analyses utilize observed weather data, eliminating uncertainty from the weather variables.

However, using forecasted weather data introduces greater uncertainty in the modeling. Days with more accurate weather forecasts will obviously provide better PV forecasts. Because daily aggregation appears to be more accurate than hourly aggregation in the general AMI model, day-ahead generation may be more easily attainable than next-day hourly estimates. Depending on the accuracy of traditional temperature-based regression models, the AMI model may provide more accurate forecasts than what some utilities currently operate. Further work is needed in this space to understand the usability of these potential forecasting products.

Chapter 8

Conclusions and Future Work

High-confidence estimates of renewable generation are needed in the delicate and dynamic balancing of load with intermittent and dispatchable generation by grid operators. Estimating PV generation from thousands of small, behind-the-meter systems, is an important and non-trivial task using existing physical models or algorithms that require actual generation data. Thus, as rooftop solar deployment increases, the ability to disaggregate invisible generation and total load from net-metered AMI data will be necessary for better management of the distribution of electricity.

Validation of the matching timestamp algorithm from this study enables greater information capabilities for utility operators and planners. The model utilizes readily available AMI data, and requires very little other data, unlike contemporary tools. This is advantageous for use by utilities that previously have missing, or no recorded residential PV install data – and serve as an example of other areas where utilities lack performance data. Unlike physical models, the AMI model is responsive and will adjust over time to changes in system performance, such as snow coverage, changes in shading, or decommissioned systems; a physical model will not. This type of model is an example of how AMI data analytics provides adaptable estimates of PV performance.

At its core, the comparable timestamp methodology is an exercise in estimating building load and can be useful in many other energy services applications. Creating before/after periods of AMI data allows analysis around an intervention date, such as the implementation of time-of-day pricing, efficiency education, weatherization, or building appliance upgrades. Energy load estimation on thousands on non-PV associated buildings can be performed from datasets like the one used. Perhaps of even greater value than PV forecasting is next-day load forecasting for a City's population of buildings, successfully leveraging end-use AMI data for more informed distribution.

Further, the timestamp matching method provides new avenues of theft detection using AMI data by utilities. Additionally, for residents interested in comparing their expected and actual

PV generation, the foundation of a viable tool is built without using an input-intensive physical model. Next-day PV forecasting techniques were superficially explored in this work and may be a promising new tool in this space. The method and computational results build a foundation for much future research on disaggregation of BTM generation from net-metered loads for use in PV and load forecasting.

In summary, this work shows that AMI data can be leveraged to predict energy generated over varying durations to a similar consistency and confidence as more technically-dense physical models. Using these simple algorithms and approach, one can estimate PV system generation, while knowing nearly nothing about the systems themselves.

8.1   Future Work

This study revealed a promising new method for PV estimation and the groundwork is laid for meaningful subsequent work. Given the numerous shortcomings of this initial experimentation, there is plentiful opportunity for improvement and model refinement. Foremost, more sophisticated sensitivity analysis of the timestamp matching algorithms is needed to optimize constraints around removing variability in energy use. Once this method is shored up, expected improved model performance positions the model for succeeding projects. Partnering with distribution utilities from other cities that have installed smart meters to examine replicability of the AMI model in other geographic locations should also be pursued.

As this methodology can be used to estimate building load, analysis of historical or future load is poised for greater research. In late 2018 Fort Collins transitioned to time-of-day pricing; a planned project with this CSU research team and the City will examine the consumption patterns within this first year of implementation. By estimating building load after TOD rates, actual energy consumption can be compared to find if the rates apply a behavioral change by shifting customers usage to off-peak periods.

Finally, future work should investigate the accuracy of PV generation estimates based on next-day weather forecasts. If the model shows estimation precision comparable to physical models

or current utility regressions, it will provide a valuable prediction instrument to utility planners overseeing a fleet of behind-the-meter solar systems. In closing, an accurate and real-time PV forecasting tool using AMI and weather data is a quintessential objective that motivated this work and should be seen to fruition.

Bibliography

[1] F. Mayes, "More Than Half of Small-scale PV Generation Comes From Residential Rooftops," *U.S. Energy Information Administration*, June 1, 2017.

[2] Wood Mackenzie Power & Renewables, "U.S. Solar Market Insight, Executive Summary: 2018 Year in Review," Solar Energy Industries Association, Report, March 2019.

[3] U.S. Energy Information Administration, "Electric Power Monthly with Data for February 2019," U.S. Energy Information Administration, Report, April 2019.

[4] Solar Energy Industries Association. (May 9, 2019) United States Surpasses 2 Million Solar Installations. [Online]. Available: https://www.seia.org/news/united-states-surpasses-2-million-solar-installations

[5] C. Roselund, "Room for Residential Solar to Reach 41 GW by 2025," *PV Magazine*, Dec. 11, 2018.

[6] P. Gagnon, R. Margolis, J. Melius, C. Phillips, and R. Elmore, "Rooftop Solar Photovoltaic Technical Potential in the United States. A Detailed Assessment," *National Renewable Energy Laboratory*, Jan. 2016.

[7] M. Tabone, S. Kiliccote, and E. C. Kara, "Disaggregating Solar Generation Behind Individual Meters in Real Time," in *Proceedings of the 5th Conference on Systems for Built Environments*, ser. BuildSys '18.   New York, NY, USA: ACM, 2018, pp. 43–52.

[8] M. Cleveland and G. Andersen. State Net Metering Policies. [Online]. Available: http://www.ncsl.org/research/energy/net-metering-policy-overview-and-state-legislative-updates.aspx

[9] B. Tarroja, F. Mueller, J. D. Eichman, and S. Samuelsen, "Metrics for evaluating the impacts of intermittent renewable generation on utility load-balancing," *Energy*, vol. 42, no. 1, pp. 546 – 562, 2012.

[10] U.S. Energy Information Administration. (May 22, 2019) How many smart meters are installed in the United States, and who has them? [Online]. Available: https://www.eia.gov/tools/faqs/faq.php?id=108&t=3

[11] Office of Electricity Delivery and Energy Reliability, "Advanced Metering Infrastructure and Customer Systems," U.S. Department of Energy, Report, September 2016.

[12] R. R. Mohassel, A. S. Fung, F. Mohammadi, and K. Raahemifar, "A survey on advanced metering infrastructure and its application in smart grids," in *2014 IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE)*, May 2014, pp. 1–8.

[13] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3125–3148, May 2019.

[14] H. Yang, C. Huang, Y. Huang, and Y. Pai, "A Weather-Based Hybrid Method for 1-Day Ahead Hourly Forecasting of PV Power Output," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 3, pp. 917–926, July 2014.

[15] Y. Wang, N. Zhang, Q. Chen, D. S. Kirschen, P. Li, and Q. Xia, "Data-driven probabilistic net load forecasting with high penetration of behind-the-meter pv," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3255–3264, May 2018.

[16] S. Chow, E. Lee, and D. Li, "Short-term prediction of photovoltaic energy generation by intelligent approach," *Energy and Buildings*, vol. 55, pp. 660–667, 12 2012.

[17] H. Shaker, H. Zareipour, and D. Wood, "A Data-Driven Approach for Estimating the Power Generation of Invisible Solar Sites," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2466–2476, Sep. 2016.

[18] S. Iyengar, N. Sharma, D. Irwin, P. Shenoy, and K. Ramamritham, "A Cloud-Based Black-Box Solar Predictor for Smart Homes," *ACM Trans. Cyber-Phys. Syst.*, vol. 1, no. 4, pp. 21:1–21:24, Aug. 2017.

[19] R. Mohan, T. Cheng, A. Gupta, V. Garud, and Y. He, "Solar Energy Disaggregation using Whole-House Consumption Signals." NILM Workshop, June 2014.

[20] N. Sharma, P. Sharma, D. E. Irwin, and P. J. Shenoy, "Predicting Solar Generation From Weather Forecasts Using Machine Learning," *2011 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pp. 528–533, 2011.

[21] E. C. Kara, M. Tabone, C. Roberts, S. Kiliccote, and E. M. Stewart, "Estimating Behind-the-meter Solar Generation with Existing Measurement Infrastructure: Poster Abstract," in *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*, ser. BuildSys '16, 2016, pp. 259–260.

[22] J. Yu, Z. Wang, A. Majumdar, and R. Rajagopal, "DeepSolar: A Machine Learning Framework to Efficiently Construct a Solar Deployment Database in the United States," *Joule*, vol. 2, no. 12, pp. 2605 – 2617, 2018.

[23] P. Gotseff, J. Cale, M. Baggu, D. Narang, and K. Carroll, "Accurate Power Prediction of Spatially Distributed PV Systems Using Localized Irradiance Measurements," in *2014 IEEE PES General Meeting - Conference Exposition*, July 2014, pp. 1–5.

[24] A. DeBenedictis and T.E. Hoff and S. Price and C.K. Woo, "Statistically Adjusted Engineering (SAE) Modeling of Metered Roof-top Photovoltaic (PV) Output: California Evidence," *Energy*, vol. 35, no. 10, pp. 4178 – 4183, 2010.

[25] D. Chen and D. Irwin, "SunDance: Black-box Behind-the-Meter Solar Disaggregation," in *Proceedings of the Eighth International Conference on Future Energy Systems*, ser. e-Energy '17, 2017, pp. 45–55.

[26] ——, "Black-box Solar Performance Modeling: Comparing Physical, Machine Learning, and Hybrid Approaches," *SIGMETRICS Performance Evaluation Review*, vol. 45, no. 2, pp. 79–84, Oct. 2017.

[27] H. Shaker, H. Zareipour, and D. Wood, "Estimating Power Generation of Invisible Solar Sites Using Publicly Available Data," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2456–2465, Sep. 2016.

[28] A. P. Dobos, "PVWatts Version 5 Manual," National Renewable Energy Laboratory, Report, 2014.

[29] D. Menicucci, "PVFORM: A New Approach to Photovoltaic System Performance Modeling," in *Presented at the 18th IEEE Photovoltaic Specialists Conference, Las Vegas, Nevada*, 21 Oct. 1985.

[30] D.L. King, W.E. Boyson, J.A. Kratochvill, "Photovoltaic Array Performance Model," Sandia National Laboritories, Report, December 2004.

[31] PV Performance Modeling Collective. (May 23, 2019) Sandia Module Temperature Model. [Online]. Available: https://pvpmc.sandia.gov/modeling-steps/2-dc-module-iv/module-temperature/sandia-module-temperature-model/

[32] U.S. Departmnet of Energy. (May 22, 2019) Recovery Act Selection For Smart Grid Investment Grant Awards. [Online]. Available: https://www.energy.gov/sites/prod/files/edg/media/SGIGSelections_Category.pdf

[33] Colorado State University. (May 22, 2019) Fort Collins Weather Station. [Online]. Available: https://ccc.atmos.colostate.edu/~autowx/

[34] NOAA National Centers for Environmental Information. (June 3, 2019) Climate data online. National Oceanic and Atmospheric Administration. [Online]. Available: https://www.ncdc.noaa.gov/cdo-web/

[35] B. Murrill, E. Liu, and R. Thompson, "Smart Meter Data: Privacy and Cybersecurity," Congressional Research Service, Report, February 3, 2012.

[36] City of Fort Collins Open Data. (June 3, 2019) Fort Collins Solar Installations. Cityof Fort Collins. [Online]. Available: https://opendata.fcgov.com/Environmental-Health/Fort-Collins-Solar-Installations/3ku5-x4k9

[37] Y. Qiu, M. E. Kahn, and B. Xing, "Quantifying the Rebound Effects of Residential Solar Panel Adoption," *Journal of Environmental Economics and Management*, vol. 96, pp. 310 – 341, 2019.

[38] D. L. Costa and M. E. Kahn, "Energy Conservation "Nudges" and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment," *Journal of the European Economic Association*, vol. 11, no. 3, pp. 680–702, 06 2013.

[39] R. Haas, H. Auer, and P. Biermayr, "The Impact of Consumer Behavior on Residential Energy Demand for Space Heating," *Energy and Buildings*, vol. 27, no. 2, pp. 195 – 205, 1998.

[40] S. Yuan, W. Stainsby, M. Li, K. Xu, M. Waite, D. Zimmerle, R. Feiock, A. Ramaswami, and V. Modi, "Future Energy Scenarios with Distributed Technology Options for Residential City Blocks in Three Climate Regions of the United States," *Applied Energy*, vol. 237, pp. 60 – 69, 2019.

[41] UCAR/Unidata, "Integrated Data Viewer (IDV) version 3.1." [Online]. Available: http://doi.org/10.5065/D6RN35XM

[42] W. F. Holmgren, C. W. Hansen, and M. A. Mikofski, "Pvlib Python: a Python Package For Modeling Solar Energy Systems," *Journal of Open Source Software*, vol. 29, no. 884, p. 3, 2018.

[43] NOAA National Centers for Environmental Information. (June 3, 2019) North American Mesoscale Forecast System. National Oceanic and Atmospheric Administration. [Online]. Available: https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/north-american-mesoscale-forecast-system-nam

[44] U.S. Energy Information Administration, "Electric Sales, Revenue, and Average Price," U.S. Energy Information Administration, Report, October 12, 2018.

[45] J. S. Stein, W. F. Holmgren, J. Forbess, and C. W. Hansen, "PVLIB: Open source photovoltaic performance modeling functions for Matlab and Python," in *2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC)*, June 2016, pp. 3425–3430.

[46] John Bleem, "PV Solar Modeling Guidelines Version 3.1," Colorado State University, Report, May 28, 2019.

[47] I. Reda and A. Andreas, "Solar Position Algorithm for Solar Radiation Applications," *Solar Energy*, vol. 76, no. 5, pp. 577 – 589, 2004.

[48] NOAA Global Monitoring Division, *General Solar Position Calculations*, 2019.

[49] M. M. Ismail, W. R. Anis, and R. Ghoneim, "Comparative Study Between Silicon & Gallium Arsenide on Grid PV System," *International Journal of Advanced Research*, vol. 4, pp. 2001–2017, March 2016.

[50] PV Performance Modeling Collective. (June 10, 2019) Sandia Module Sun Position. [Online]. Available: https://pvpmc.sandia.gov/modeling-steps/

[51] ——. (June 10, 2019) Sandia Module Sun Piecewise Decomposition Models. [Online]. Available: https://pvpmc.sandia.gov/modeling-steps/1-weather-design-inputs/ irradiance-and-insolation-2/direct-normal-irradiance/piecewise_decomp-models/

[52] M. S. Lave, W. Hayes, A. A. Pohl, and C. W. Hansen, "Evaluation of Global Horizontal Irradiance to Plane-of-Array Irradiance Models at Locations Across the United States," *IEEE Journal of Photovoltaics*, vol. 5, pp. 597–606, 2015.

[53] PV Performance Modeling Collective. (June 10, 2019) Sandia Module Temperature Model. [Online]. Available: https://pvpmc.sandia.gov/modeling-steps/2-dc-module-iv/ module-temperature/sandia-module-temperature-model/

[54] P. Grana. (September 8, 2017) Rethinking Inverter Clipping. [Online]. Available: https://www.solarpowerworldonline.com/2017/09/folsom-rethinking-inverter-clipping/

[55] J. Newmiller, W. Erdman, J. S. Stein, and S. Gonzalez, "Sandia Inverter Performance Test Protocol efficiency weighting alternatives," in *2014 IEEE 40th Photovoltaic Specialist Conference (PVSC)*, June 2014, pp. 0897–0900.

[56] PV Performance Modeling Collective. (June 10, 2019) DC to AC Conversion. [Online]. Available: https://pvpmc.sandia.gov/modeling-steps/dc-to-ac-conversion/

[57] J. M. Freeman and D. S. Ryberg, "Integration, Validation, and Application of a PV Snow Coverage Model in SAM," National Renewable Energy Laboratory, Report, August 2017.

[58] B. Marion, R. Schaefer, H. Caine, and G. Sanchez, "Measured and Modeled Photovoltaic System Energy Losses from Snow for Colorado and Wisconsin Locations," *Solar Energy*, vol. 97, pp. 112–121, 11 2013.

[59] B. L. Brench, "Snow-Covering Effects on the Power Output of Solar Photovoltaic Arrays," U.S. Department of Energy, Report, December 1979.

[60] C. of Fort Collins. Monthly peak hour history. [Online]. Available: https://www.fcgov.com/peakload/

[61] R. Fu, D. Feldman, and R. Margolis, "U.S. Solar Photovoltaic System Cost Benchmark: Q1 2018," National Renewable Energy Laboratory, Report, November 2018.

[62] Modernize.com, "Average Solar Panel Dimensions and Sizes," June 2019. [Online]. Available: https://modernize.com/home-ideas/29543/average-solar-panel-dimensions-and-sizes

Appendix A

AMI Model Software Package

Example code as a Jupyter Notebook is uploaded and available in a GitHub repository.

These are located at < https://github.com/wstainsby/AMI-Model-2019/ >.

Appendix B

PVL Modeling Functions

This chapter explicates the physical model developed by CSU researchers for use as a validation tool. First, the solar positioning algorithms calculate the solar azimuth and zenith at each inputted timestep. Next, the Erbs model is used to derive DHI and DNI from the measured GHI. Last, energy output is calculated incorporating temperature correction, system efficiencies, losses, and array capacity. The table below contains coefficients and parameters used in the model. The following sections delve further into these functions of the model.

Table B.1: PVL Model Coefficients

Parameter	Symbol	Value	Units
Panel Efficiency	η_{panel}	16.2	%
Module Wattage	Wattage_{module}	250	watts
Inverter Efficiency	$\eta_{inverter}$	96	%
Module Area	Area_{module}	1.55	m^2
Miscellaneous Losses	SL	8.7	%
Temperature Coefficient	T_C	-0.0043	
Solar Constant	E_{SC}	1,367	watts per m^2

B.1 Solar Position Calculations

The solar zenith angle (θ_Z) and the solar azimuth angle (θ_A) rely on calculations for the the equation of time (E_{qt}), solar declination angle (θ_d), hour angle (θ_{hr}), and latitude angle (λ):

$$\theta_Z = cos^{-1}[sin(\lambda) \times sin(\theta_d) + cos(\lambda) \times cos(\theta_d) \times cos(\theta_{hr})] \tag{B.1}$$

and the solar azimuth angle (θ_A) is found as:

$$\theta_A = Arctan2 \{ sin(\theta_{hr}) / [cos(\theta_{hr}) \times sin(\gamma) - tan(\theta_d) \times cos(\gamma)] \} + 180 \tag{B.2}$$

where both angles are in degrees. The equation of time is the calculated difference between solar time and locate time and is a function location and time of year. The equation of time is calculated as:

$$E_{qt} = \{\ 229.18 \times [0.000075\ +\ 0.001868 \times cos(\gamma)\ -\ 0.032077\ \times\ sin(\gamma)$$
$$-\ 0.014615 \times\ cos(2 \times \gamma)]\ -\ 0.040849\ \times\ sin(2 \times \gamma)\ \} \tag{B.3}$$

where E_{qt} is in minutes and γ is the fractional year:

$$\gamma\ =\ (2 \times \pi/365)\ \times\ [n - 1 + (hour - 12)\ /\ 24] \tag{B.4}$$

where *n* is the day of the year and *hour* is the hour in standard time. To calculate hour angle (θ_{hr}), calculating solar time (T_{solar}) is required. From [48], solar time (in minutes) is found as:

$$T_{solar}\ =\ [\ hour\ \times 60\ +\ min\ +\ sec/60\ +\ \textit{time offset}\ ]\ /\ 60 \tag{B.5}$$

where hour is the *hour* of the timestamp, *min* is the minute, and *sec* is the second. Time offset between local and solar time is found as:

$$\textit{time offset}\ =\ E_{qt}\ +\ 4\ \times\ longitude\ -\ 60\ \times\ timezone \tag{B.6}$$

where longitude is in degrees, and timezone is in hours from Coordinated Universal Time (Mountain Standard Time is 7 hours from UTC). With T_{solar} calculated, the hour angle can now be found in radians:

$$\theta_{hr}\ =\ (\pi\ /\ 12)\ \times\ (T_{solar}\ -\ 12) \tag{B.7}$$

At this point, θ_A and θ_Z can be calculated and are used to calculated the angle of incidence in Equation 5.3.

B.2 Erbs Decomposition Model

As global horizontal irradiance (GHI) is the typically measured form of irradiance, relationships between the irradiance reaching Earth's surface and extraterrestrial irradiance are used to estimate the diffuse (DHI) and direct (DNI) irradiance components. This section is directly based on Bleem's work [46].

A clearness index (k_t) is used to decompose DHI and DNI from GHI. The more clear the sky is, the greater the proportion of beam radiation from the incoming irradiance. The clearness index is defined as:

$$k_t = \frac{GHI}{E_{aH}} \tag{B.8}$$

where GHI is total horizontal irradiance reaching the Earth's surface and E_{aH} is the horizontal component of extraterrestrial irradiance. E_{aH} is determined from the extraterrestrial radiation E_a reaching the top of the Earth's atmosphere on a plane normal to the Sun. This radiation varies through the year and is empirically calculated as:

$$
\begin{aligned}
E_a = E_{SC} \times [&1.00011 + 0.034221 \times cos(B) + 0.00128 \times sin(B) \\
&+0.000719 \times cos(2 \times B) + 0.0000077 \times sin(2 \times B)]
\end{aligned}
\tag{B.9}
$$

where:

$$B = 2 \times \pi \times \frac{n}{365} \tag{B.10}$$

and where n is the day of year and E_{SC} is the solar constant equal to 1,367 watts/m^2. The horizontal component of the extraterrestrial irradiance is now:

$$E_{aH} = E_a \times cos(\theta_Z) \tag{B.11}$$

where θ_Z is the zenith angle of the Sun.

Once k_t is calculated, three curve fits are used across the range of k_t values to calculate the diffuse fraction, k_d. The Erbs model constructed these three curve fit equations:

$$k_d \; = \; 1 - 0.09 \; \times \; k_t \qquad\qquad\qquad k_t \leq 0.22 \qquad \text{(B.12)}$$

$$k_d \; = \; 0.9511 - 0.1604 \times k_t + 4.388 \times k_t^2 - 16.638 \times k_t^3 + 12.336 \times k_t^4 \qquad \text{(B.13)}$$
$$0.22 < k_t \leq 0.80$$

$$k_d \; = \; 0.165 \qquad\qquad\qquad\qquad k_t > 0.80 \qquad \text{(B.14)}$$

Once k_d is derived, DHI reaching a flat surface is calculated in Equation 5.4.

B.3   Temperature Correction Factor

A temperature correction factor is found and applied to the DC energy output from the array:

$$T_{correction} \; = \; 1 \; + \; T_C \; \times \; (T_M \; - \; T_{test}) \qquad \text{(B.15)}$$

Where T_C is a temperature coefficient, T_M is the module temperature, and T_M is the standard test temperature of 25 °C. Different module materials dictate use of various derived constants and are used in Equation B.16 to calculate module temperature [53]:

$$T_M \; = \; E_{POA} \; \times \; (e^{a+b \times WS}) \; + \; T_A \qquad \text{(B.16)}$$

Where EPOA is the solar irradiance incident to the module (watts/m^2), T_A is the ambient air temperature (°C), and WS is the wind speed (m/s). Constants a and b are empirical constants based

on the module material and construction, along with the panel mounting configuration [53]. The resulting T_M is used to calculate an estimate for the temperature coefficient, T_C, which used to correct PV generation output. $T_{correction}$ is defined in Equation B.15 and ΔT is shown in Table B.2 below:

$$T_{cell} \; = \; T_M \; + \; (\; \frac{E_{POA}}{1000} \;) \; \times \; \Delta T \tag{B.17}$$

The following table is adapted from the Sandia PVPMC temperature model literature and contains values for various module and mounting types [53].

Table B.2: Temperature Coefficients for Different PV Configurations

Module Type	Mount Configuration	a	b	ΔT (°C)
Glass/cell/glass	Open rack	-3.47	-0.0594	3
Glass/cell/glass	Close roof mount	-2.98	-0.0471	1
Glass/cell/polymer sheet	Open rack	-3.56	-0.0750	3
Glass/cell/polymer sheet	Insulated back	-2.81	-0.04554	0
Polymer/thin-film/steel	Open rack	-3.58	-0.113	3

B.4   Panel Efficiency and Size

Figure B.1 is adapted from NREL's 2018 PV system benchmark report [61]. Shown are annual residential module power and efficiency data from the California Net Energy Metering Database, which provides a large dataset of PV statistics. For the PVL model, an average module power rating of 250 watts and efficiency of 16.2% were chosen as input constants.
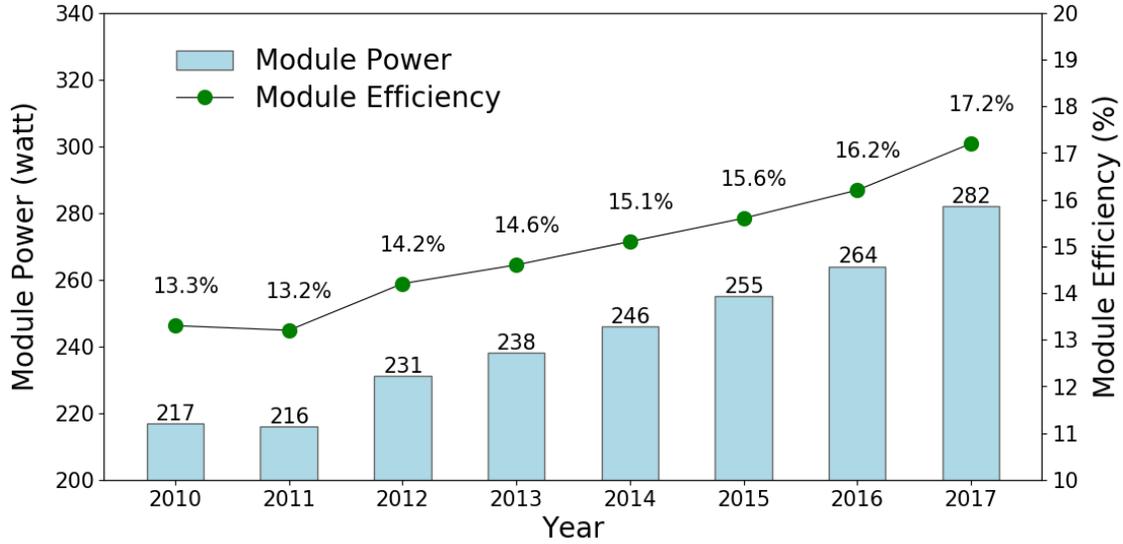
Figure B.1: Residential Module Power and Efficiency Trends. 8 years of PV equipment trends from California's NEM database from [61].

The assumed panel area was determined from a list of ten common solar panel products [62]. The dimensions of this group of ten panels are shown below in Table B.3 with brand, model, and panel area. Based on the panel areas in the right-most column, the median of these was 1.63 m², thus a panel area of 1.63 m² was used in this PV model. An anecdotal proportion of 95% of a panel area being cell area was used to calculate an assumed cell area of 1.55 m² as shown in Table B.1.

Table B.3: Common Residential Solar Panel Dimensions from [60].

Brand	Model	Length (m)	Width (m)	Area (m²)
Trina Solar	TSM-260-PD05	1.651	0.991	1.63
Canadian Solars	CS6K-275	1.651	0.991	1.63
Jinko Solar	JKM250P-60	1.651	0.991	1.63
JA Solar	JAM6 60/250-270	1.651	0.991	1.63
Hanwha	Q.Pro-G4 255-265	1.676	0.991	1.66
First Solar	Series 5	1.854	1.194	2.21
Yingli Green	YLM 60 Cell	1.651	0.991	1.63
Suntech	STP 280S-20Wem	1.651	0.991	1.63
ReneSola	JC270 S-24	1.651	0.991	1.63
SunPower	E20-327	1.557	1.046	1.63