March 2022

Data-Driven Control, Modeling, and Forecasting for Residential Solar Power

Akansha Singh Bansal
University of Massachusetts Amherst

DATA-DRIVEN CONTROL, MODELING, AND FORECASTING FOR RESIDENTIAL SOLAR POWER

A Dissertation Presented

by

AKANSHA SINGH BANSAL

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 2022

Electrical and Computer Engineering

DATA-DRIVEN CONTROL, MODELING, AND FORECASTING FOR RESIDENTIAL SOLAR POWER

A Dissertation Presented

by

AKANSHA SINGH BANSAL

Approved as to style and content by:

David Irwin, Chair

Prashant Shenoy, Member

Michael Zink, Member

Jay Taneja, Member

Christopher V. Hollot, Department Chair
Electrical and Computer Engineering

DEDICATION

To Maa, Papa, and Trapit.

ACKNOWLEDGMENTS

I would like to thank many people who have helped me start and finish my Ph.D. journey. I am extremely grateful to my advisor Dr. David Irwin. He has been very supportive, encouraging, and generous with his feedback and insights throughout my Ph.D. He taught me how to pick research problems, think critically about them from first principles, and how to communicate them to a wider audience. I really value his willingness to be open to new ideas and provide me the freedom to pick and pursue the research topics of my interest. This helped me grow as an independent researcher. I will always cherish his ongoing mentorship throughout my career.

Thanks to the members of my dissertation committee for their time, energy, and feedback in helping me improve this thesis: Dr. Prashant Shenoy, Dr. Michael Zink, and Dr. Jay Taneja. My Ph.D. journey would not have started if not for Dr. Srinivasan Prabhu (IIT-K/IIIT-B) who was the first to introduce me to the idea of applying to graduate school. I will always be grateful to him. My master's thesis advisors, Dr. Jyotsna Bapat (IIIT-B) and Dr. Debabrata Das (IIIT-B) with whom I had my first research exposure and who have spent hours reading my paper drafts. They have been a major influence on me and I am extremely thankful to them.

I am grateful for the time I spent at NTU Singapore which was also a pre-Ph.D. gap year I took right after finishing my Master's. That experience helped me identify the research area that I wanted to explore further in my Ph.D. I would also like to acknowledge my lab members and the staff members at UMass Amherst for their help and support throughout my Ph.D. Also, I would like to thank the wonderful colleagues and mentors I had the privilege of working with as an intern at the Oak Ridge National Lab, during the summer of 2021.

I want to thank my family for their unconditional love and support without whom I would be lost. My mother, who has helped me to keep going even when things were not progressing in my favor. My father, who has always taught me that I still have lots left to learn, both personally and professionally. My brother, Puneet, and his wife Shikha, for always having my back and their two lovely sons Amey and Arth.

Finally, I am deeply appreciative of my best friend and husband, Trapit, who is my pillar of support and I am profoundly grateful for his companionship as a friend, colleague, and partner.

ABSTRACT

DATA-DRIVEN CONTROL, MODELING, AND FORECASTING FOR RESIDENTIAL SOLAR POWER

FEBRUARY 2022

AKANSHA SINGH BANSAL

B.Tech, MAHARISHI DAYANAND UNIVERSITY, ROHTAK

M.S., INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY, BANGALORE

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor David Irwin

Distributed solar generation is rising rapidly due to a continuing decline in the cost of solar modules. Most residential solar deployments today are *grid-tied*, enabling them to draw power from the grid when their local demand exceeds solar generation and feed power into the grid when their local solar generation exceeds demand. The electric grid was not designed to support such decentralized and intermittent energy generation by millions of individual users. This dramatic increase in solar power is placing increasing stress on the grid, which must continue to balance its supply and demand despite the potential for large solar fluctuations. To address the problem, this thesis proposes new data-driven techniques for better controlling, modeling, and forecasting residential solar power.

The grid currently exercises no direct control over its connected solar capacity, but instead indirectly controls it by regulating new solar connections. This approach is highly

inefficient and wastes much of the grid's potential to transmit solar. Instead, we propose Software-defined Solar-powered (SDS) systems that dynamically regulate solar flow rates into the grid and design an SDS prototype, called SunShade. Specifically, we introduce a new class of Weighted Power Point Tracking (WPPT) algorithms, inspired by Maximum Power Point Tracking (MPPT), capable of dynamically enforcing both hard and relative caps on solar power, which enables the grid to decouple rate control from admission control. In contrast, to avoid grid regulations entirely, homes can also partially or entirely defect from the grid to fully utilize their solar power without restrictions. We present a switching architecture that enables homes to dynamically switch between a local generator, battery, and solar to co-optimize their cost, carbon footprint, switching frequency, and reliability. We introduce switching policies that reveal a tradeoff between solar utilization and reliability, such that higher solar utilization requires more switching, which can lead to lower reliability.

Enabling better control of intermittent solar also requires improving solar performance models, which infer solar output based on current environmental conditions. Recent solar models primarily leverage data from ground-based weather stations, which may be far from solar sites and thus inaccurate. In addition, these weather stations report cloud cover—the most important metric for solar modeling—in coarse units of oktas. Instead, we propose developing solar models based on data from a new generation of Geostationary Operational Environmental Satellites (GOES-16 and GOES-17) that began launching in late 2017. We develop physical and machine learning (ML) models for solar performance modeling using both derived data products released by the National Oceanic and Atmospheric Administration (NOAA), as well as the satellites' raw multispectral data. We find that ML-based models using the raw multispectral data are significantly more accurate than both physical models using derived data products, such as Downward Shortwave Radiation (DSR), and prior okta-based solar models. The raw multispectral data is also beneficial since it is available at much higher spatial and temporal resolutions—$\sim$1km^2 and every 5 minutes—than

oktas—$\sim$25km^2 and every hour. The accuracy of our ML-based models on multispectral data is also better regardless of whether they are locally trained using data only from a particular solar site or globally trained using data from many solar sites. Since global models can be trained once but used anywhere, they can also enable accurate modeling for sites with limited data, e.g., newly installed solar sites.

Solar forecasting models, which predict future solar output based on environmental conditions also help in better solar control. Accurate near-term solar forecasts on the order of minutes to an hour are particularly important because homes and the grid must be able to adapt to large sudden changes in solar output. Current solar forecasting techniques, which primarily use Numerical Weather Predictions (NWP) algorithms, mostly leverage physics-based modeling. These physics-based models are most appropriate for forecast horizons on the order of hours to days and not near-term forecasts on the order of minutes to an hour. While there is some recent work on analyzing images from ground-based sky cameras for accurate near-term solar forecasting, it requires installing additional infrastructure. We instead propose a general model for solar nowcasting from abundant and readily available multispectral satellite data using self-supervised learning. Specifically, we develop deep auto-regressive models using convolutional neural networks (CNN) and long short-term memory networks (LSTM) that are globally trained across multiple locations to predict raw future observations of the spatio-temporal data collected by the recently launched GOES-R series of satellites. Our model estimates a location's future solar irradiance based on satellite observations, which we feed to a regression model trained on smaller site-specific solar data to provide near-term solar photovoltaic (PV) forecasts that account for site-specific characteristics.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS . . . . . **v**

ABSTRACT . . . . . **vii**

LIST OF TABLES . . . . . **xiv**

LIST OF FIGURES . . . . . **xv**

CHAPTER

1. INTRODUCTION **1**

1.1 Motivation 1
1.2 Thesis Contributions 3

1.2.1 Solar Control 3

1.2.1.1 Controlling Grid-tied Solar Capacity. 4
1.2.1.2 Controlling Off-grid Solar Capacity. 4

1.2.2 Solar Modeling 5

1.2.2.1 Satellite-based Modeling using DSR Data 5
1.2.2.2 Satellite-based Modeling using Multispectral Data 6

1.2.3 Solar Nowcasting 6

1.3 Dissertation Overview 7

2. BACKGROUND **8**

2.1 Rising Solar and its Environmental Impacts 8
2.2 Solar Control and Storage 9

2.2.1 Challenges 10

2.3 Solar Modeling 10
2.4 Solar Predictability 11

3. CONTROLLING GRID TIED RESIDENTIAL SOLAR 12

3.1 Introduction 12
3.2 Background 15

3.2.1 Background 17

3.3 Controlling Solar Flows 20
3.4 SDS Control Mechanisms 21

3.4.1 Absolute Power Capping 22
3.4.2 Weighted Power Point Tracking 23
3.4.3 Search-based Approach 24
3.4.4 Model-based Approach 24

3.5 Implementation 26
3.6 Evaluation 28

3.6.1 Simulation Results 29
3.6.2 Prototype Results 32

3.7 Related Work 36
3.8 Conclusions 38

4. CONTROLLING OFF GRID RESIDENTIAL SOLAR 39

4.1 Introduction 39
4.2 Background 43

4.2.1 Grid Power 43
4.2.2 Whole-Home Natural Gas Generator 44
4.2.3 Solar Power and Batteries 45

4.3 Grid Defection Architecture 46
4.4 Evaluation 49

4.4.1 Cost and Carbon Emissions Analysis 49
4.4.2 Solar Waste Analysis 51
4.4.3 Impact of Battery Size 52
4.4.4 Impact of Aggregating Homes 53

4.5 Related Work 55
4.6 Conclusions 56

5. MODELING SOLAR PERFORMANCE MODELS USING DERIVED PRODUCTS FROM SATELLITE 57

5.1 Introduction 57
5.2 Background 60

5.2.1 Satellite-based DSR 61
5.2.2 Ground-level Cloud Cover Measurements 62
5.2.3 Modeling SURFRAD Data 63
5.2.4 Data-driven Solar Performance Modeling 63

5.3 Satellite-Based Solar Performance Modeling 65

5.3.1 Satellite-Based Models 65
5.3.2 Oktas-Based Models 66
5.3.3 Hybrid Model 67

5.4 Implementation 69
5.5 Evaluation 72
5.6 Related Work 75
5.7 Conclusions 76

6. MODELING SOLAR PERFORMANCE MODELS USING MULTISPECTRAL SATELLITE DATA 77

6.1 Introduction 77
6.2 Background Estimates 81

6.2.1 Prior Approaches 81
6.2.2 Spectral Data Characteristics 85
6.2.3 Local ML Models 86

6.3 Global Model 88
6.4 ML Model Variants 89
6.5 Implementation 90
6.6 Physical Modeling Approaches 91
6.7 Evaluation 92

6.7.1 Performance of Local Models 94
6.7.2 Performance of Global Models 99

6.8 Related Work 101
6.9 Conclusion 102

7. SOLAR NOWCASTING USING SELF-SUPERVISED LEARNING ON MULTISPECTRAL SATELLITE DATA 103

7.1 Introduction . . . 103
7.2 Background . . . 108
7.2.1 Prior Work . . . 109
7.2.2 Basic Methodology . . . 111
7.3 Solar Nowcasting Model Design . . . 114
7.3.1 Spatial Modeling using Convolution Neural Network (CNN) . . . 114
7.3.2 Temporal Modeling using Long Short-Term Memory (LSTM) . . . 115
7.3.3 Training CNN-LSTM Model . . . 116
7.3.4 Forecasting Solar Output . . . 117
7.4 Implementation . . . 118
7.4.1 Satellite Data and Solar Sites . . . 118
7.4.2 Model Hyper-Parameters and Metrics . . . 120
7.5 Evaluation . . . 121
7.5.1 Evaluating ML Models for Spatial Modeling . . . 123
7.5.2 Evaluating CNN-LSTM Models for Spatio-Temporal Modeling . . . 125
7.5.3 End-to-End Solar Nowcasting . . . 127
7.6 Related Work . . . 131
7.7 Conclusion . . . 132
8. FUTURE DIRECTIONS AND CONCLUSION . . . 134
8.1 Conclusion . . . 134
8.2 Future Directions . . . 136
BIBLIOGRAPHY . . . 138

LIST OF TABLES

Table **Page**

3.1 NRMSE comparison for sunny and cloudy weather conditions 33

4.1 Summary of current CapEx, OpEx, lifetimes, and carbon emissions (in kg-CO_2/kWh) of different generation options for grid defectors. 43

6.1 Wavelength for 16 channels of GOES-16 and -17 [26]. 83

7.1 Total number of sequences as well as the number of days of the year that comprise training, validation and testing for the CNN-LSTM model. 121

7.2 Variation in number of data points with respect to different tolerances for different channels. 122

LIST OF FIGURES

Figure **Page**

3.1 I-V curve for a typical solar module, and the effect of changes in lighting and temperature. 17

3.2 Example I-V curve for a solar array formed from wiring multiple solar modules in series (a), in parallel (b), and a combination of the two (c). 18

3.3 Depiction of our SunShade simulator. 25

3.4 Expansion and contraction of the I-V curve for our SunShade prototype, as light intensity changes throughout the day. 27

3.5 Depiction of our SunShade prototype. 29

3.6 SunShade capping the absolute power of a solar panel to 100W. 30

3.7 Example of model-based weighted power capping at 80% of the MPP. 30

3.8 Example of model-based weighted power capping at 80% of the MPP. 31

3.9 Example of search-based WPPT at 80% MPP on a sunny day. 32

3.10 Example of search-based WPPT at 80% MPP on a cloudy day. 32

3.11 NRMSE for periods with different levels of variation. 33

3.12 WPPT of SunShade in a real deployment compared with ideal WPPT based on true MPPT 34

3.13 The RMSE between the ideal WPPT and search-based WPPT as a function of the periodic search interval. 35

3.14 The RMSE between the ideal WPPT and search-based WPPT as a function of the weight. 36

3.15 The RMSE between ideal and search-based WPPT as a function of the voltage step-size. 37

4.1 Grid Defection Architecture 48

4.2 Operational and amortized capital cost for different scenarios in $/year. 49

4.3 Carbon emissions from the same scenarios as in Figure 4.2 for cost. 50

4.4 Price Comparison for Different States 51

4.5 Percent Demand Met by Different Sources 52

4.6 Amortized cost and carbon emissions as battery capacity increases. 53

4.7 Cost Per Unit Power for Grid vs. Total Grid Defection with optimal configuration for aggregation of homes in different states of USA 54

4.8 Amortized Cost vs % Energy Availability for a 15 home cluster 55

5.1 Depiction of bounding solar generation using Equation 5.2. 62

5.2 Relationship between DSR and clear sky irradiance 66

5.3 Scatter plot of normalized solar generation versus normalized satellite DSR (left) and okta-based measurements (right) across many solar sites. 67

5.4 Availability of DSR data product across our 47 solar sites. 71

5.5 DSR availability under different cloud conditions. 72

5.6 MAPE result with SURFRAD Data. 73

5.7 MAPE for satellite-based and okta-based models under overcast cloud conditions. 74

5.8 MAPE for physical and ML models in 2019. 75

5.9 MAPE for satellite-based, okta-based, and hybrid model 2018-19. 76

6.1 Relationship between solar generation at a site, and the values of channels 1, 2, and 3 at the same location. 85

6.2 Diagram of data inputs for different solar performance modeling approaches we implement and evaluate. 93

6.3 Performance comparison across the 29 sites. Data consists of only summer months and the middle of the day. Channel data shows consistently better performance. 94

6.4 Performance comparison of local model under different time-periods. Average over 29 sites is shown. Left shows comparison using MAPE and right using CEP. 95

6.5 Performance comparison of Okta, DSR and Channels. Average across 29 solar sites over summer months and middle of the day is shown. Left shows comparison using MAPE and right using CEP. 97

6.6 Performance comparison of first 3 channels versus 16 channels from GOES-16. Average MAPE across sites is shown. 98

6.7 Combination of GOES-16 and GOES-17. Average MAPE (left) and CEP (right) across sites using 2019 data over summer months and middle of the day. Performance of the models at 5 minutes resolution is also shown. 99

6.8 Performance comparison of different machine learning models. 100

6.9 Performance comparison of global model. Average MAPE on year 2018-2019 data from GOES-16 (left) and year 2019 data from both GOES-16 and GOES-17 (right). Note that the global model performance is on new sites which are not part of training of the regression model. 101

7.1 GOES-16 (left) and GOES-17 (right) coverage area. We use GOES-16 satellite data as it observes the continental U.S. covering all the 25 solar sites considered in this work. 107

7.2 GOES-16 (left) and GOES-17 (right) coverage area. We use GOES-16 satellite data as it observes the continental U.S. covering all the 25 solar sites considered in this work. 108

7.3 Cumulative Distribution Function of changes in solar output at 5, 15 and 60 minute frequency over one year for a representative solar site. Note that a majority of times solar hardly changes between instants (value at 0) and thus modeling the occasional abrupt changes is crucial. 112

7.4 Overview of our complete pipeline. Multispectral satellite data from 25 sites across U.S. is collected. Sequences of observations at a site, comprising of first 3 channels from the satellite (visualized in middle) are input to a CNN-LSTM model that is trained to predict the observation at the center (red pin) for the next time step ($t + 1$). An auto-regressive SVR takes the predicted channels and previous solar output as well as previous temperature to predict the solar output at next step. 113

7.5 MAE for different ML models used to predict the next instant channel values from current instant. CNN model has the lowest error. 123

7.6 MAE for models trained/evaluated at different time intervals. At larger intervals (30-60 minutes) forecasting becomes increasingly harder. 124

7.7 Effect of spatial area on forecast MAE for different models. Using larger area improves forecast for all models and CNN trained using (10×10) area is better. 125

7.8 CNN-LSTM performance at forecasting next instant channel values. Compared to not using previous temporal history (CNN one step), CNN-LSTM leads to significant error reduction on predicting large changes (tolerance> 0) and retains overall better performance compared to persistence model (tolerance$= 0$). 126

7.9 Comparison of CNN-LSTM models using varying amounts of previous temporal information. Note that using 3-4 steps is better than using lesser steps. There is a marked reduction in error when using more than 1 step in the model. 127

7.10 End to end solar forecasting on 10x10 km area, averaged over 25 solar sites over 15 mins. Performance for summer months (May-September) is shown on left and for the full year on the right. Using the predictions from the CNN-LSTM model leads to solar output forecasting with error close to that of using the current satellite observations. Compared to precision model forecasts, this approach is consistently better, specially at predicting when there will be large changes in solar ($\geq 5\%$). 128

7.11 CEP for end to end solar forecasting on 10x10 km area, averaged over 25 solar sites over 15 mins. Performance for summer months (May-September) is shown on left and for the full year on the right. The performance of CEP is similar to that of MAPE where the predictions from the CNN-LSTM model leads to solar output forecasting with error close to that of using the current satellite observations. 128

7.12 Forecast skill score distribution over summer months and full year, along with the distribution of the number of instances at different tolerances. 130

7.13 Forecasting skill score of all the 25 solar sites at 5% tolerance for the full year. 131

CHAPTER 1

INTRODUCTION

Solar energy generation has grown at nearly an exponential rate over that past 30 years, and is now cheaper than the retail price of electricity in many locations [113]. The goal for the U.S. Department of Energy's SunShot initiative is for solar to satisfy 14% of U.S. electricity demand by 2030 and 27% by 2050 [23], or a factor of 10× and 20×, respectively, greater than the 1.4% it satisfied in 2018 [24]. These rapidly falling prices and environmental concerns are in turn driving significant increases in distributed solar generation – both residential and commercial. Nearly all these solar deployments are "grid-tied," enabling them to draw power from the grid when their local demand exceeds solar generation and feed power into the grid when their local solar generation exceeds demand. However, these grid-tied systems impose a burden on the grid to absorb a building's energy surpluses and make up for its energy deficits. This thesis discusses techniques for controlling, modeling and forecasting residential solar which are important to support the increasing penetration of solar in the grid.

1.1 Motivation

Solar has experienced an average annual growth rate of 49% in the last decade with more than 89 gigawatts (GW) of solar capacity installed nationwide, enough to power 16 million homes [36]. This is mainly attributed to federal policies like the tax credits, declining PV costs, and increasing demand for clean electricity. The electric grid was not designed to support such decentralized and intermittent energy generation by millions of individual users. Instead, the grid imposes a rigid top-down hierarchy where large, highly-

regulated utilities generate (and purchase) energy to meet the demand of their customers and maintain grid stability. To do so, utilities continuously balance electricity's supply and demand in real time by regulating generator power output. Since the energy demand (or "load") profile of individual users is stochastic, such real-time balancing is only possible because the sum of load profiles across many users tends to be smooth and highly predictable. As a result, utilities can plan when to activate (or "dispatch") generators in advance to satisfy large increases in demand. Distributed solar generation at large scales fundamentally alters this paradigm by increasing the stochasticity of user load profiles, even when aggregating them. Solar power output can change almost instantaneously, e.g., due to passing clouds making it stochastic and unreliable source of power. While dispatchable generators are mechanical devices that take some time to activate and adjust their power output, thus preventing them from maintaining high power quality when compensating for rapid solar variations. In general, the grid faces significant operational challenges when renewable penetration approaches 10% [122], necessitating additional energy storage or sophisticated demand-side load management.

This increasing amount of solar generation is profoundly changing the grid's operation and the model of utilities. In particular, while utilities operate and maintain the distribution network, they earn most of their revenue from generating electricity, which they currently can do much more efficiently than individual users. However, solar generation differs from fossil-fuel based generation in that it does not benefit as much from economies of scale. As a result, individual homeowners can install solar on their rooftops for closer to the same cost per watt that utilities can install large grid-scale solar farms. Even now, when amortized over a 25-year lifetime, solar power is cheaper than retail electricity rates in much of the U.S., assuming the solar power can be "net metered" and the utility credits users the retail rate for surplus power fed into the grid. Net metering enables consumers to connect solar power to the grid such that it acts as a negative load, causing their meter to run backwards when generating a net power surplus. When combined with government

incentives, the payback period for net metered solar is now well under 10 years in many states. Of course, the more individual users generate their own power, the less revenue utilities earn from generating electricity. Yet, utilities cannot simply decommission their generators (and take a capital loss), since they must still supply the grid's power at night when the sun is not shining. In addition, utilities may need to alter their mix of generators to handle increasing net metered solar installations, and their increased stochasticity, by employing more responsive but less efficient peaking generators. These changes may in-turn increase the cost and decrease the efficiency of grid generators.

In order to increase the share of solar in the energy mix and achieve the clean energy goals, there is a need for better solar control, modeling and forecasting techniques.

1.2 Thesis Contributions

As described above, distributed solar creates fundamental challenges which require better controlling, modeling and forecasting techniques. Data-driven techniques can be studied at different granularities, right from the rooftop solar installations to weather monitoring stations and the satellite data. We summarize our thesis contributions below.

1.2.1 Solar Control

With the rise in distributed generation and demand for energy, the grid is becoming increasingly strained. As thousands of people are installing solar and generating their own energy, consumers are becoming less and less dependent on the utilities for their energy supplies. In order to control this, the state and government introduce "net metering" laws to allow customers with rooftop solar to sell their energy back to the grid. Unfortunately a number of states are lacking the policies to encourage solar; but many are also actively preventing it through policies and restrictions [110]. The current approach to allocating the grid's available solar capacity is a static peak-based first-come-first-serve policy which is highly inefficient, and wastes much of the grid's potential to transmit solar. As these

policies become more and more strict, customers are beginning to defect from the grid either partially or completely in order to employ solar without restrictions.

1.2.1.1 Controlling Grid-tied Solar Capacity.

In Chapter 3, we propose Software-defined Solar-powered (SDS) systems to dynamically control the flow rates of solar into the grid and also present an SDS prototype, called SunShade [118]. SDS focuses not only on managing grid stability but also on fairly sharing the grid's available capacity to accept solar while maintaining grid neutrality. It does so by introducing a new class of Weighted Power Point Tracking (WPPT) algorithms that enforce a relative cap and eliminate the need for policy imposed static limits. The goal of SDS systems is to enable all users to freely connect to the grid and dynamically share its capacity to transmit solar power. This principle represents a form of *grid neutrality*, akin to net neutrality, where the grid treats all solar energy contributions equally without discriminating between users.

1.2.1.2 Controlling Off-grid Solar Capacity.

In the absence of such a system and with increasing policy regulations, we envision solar users considering partially or entirely defecting from the grid so as to employ solar without restrictions and have more control over their energy. The idea of energy independence by combining solar with storage will allow consumers with rooftop solar to meet their own energy demands without selling it back to the grid. In this regard, we present an architecture for grid defection that enables a home to dynamically switch between a local/-generator and solar/battery depending on its power consumption and generation in Chapter 4 [43]. We introduce switching policies that define a tradeoff between power switching and wasted solar: more switching leads to less reliability, but maximizes the use of solar. We further analyze the tradeoffs in terms of number of switches, solar waste, reliability, carbon emissions and total cost.

1.2.2 Solar Modeling

Enabling better control of intermittent solar also requires improving solar performance models that infer solar sites' output based on its physical and environmental characteristics. Solar is intermittent and its output is affected by a number of factors like sky conditions, location, time of the day, site's elevation etc. Even passing clouds have the potential to introduce significant fluctuations in solar energy production, by introducing sudden increases or decreases, often called ramps. Recent solar models primarily leverage data from ground-based weather stations, which are not always available for every solar site. Even if available, these could be located at distance away from the solar sites making their readings inaccurate. In addition, these weather stations report cloud cover—the most important metric for solar modeling—in coarse units of oktas. One okta is equal to one eight of the sky covered in cloud. These further loosely translate to fractions between 0-1 and are measured on an hourly basis.

1.2.2.1 Satellite-based Modeling using DSR Data

The low frequency, resolution and spatial coverage of these ground-based okta measurements make this metric coarse, resulting in inaccuracy which is by far the largest source of error in solar modeling. In Chapter 5 [42] we instead develop and evaluate solar performance models that use satellite-based estimates of downward shortwave (solar) radiation (DSR) at the Earth's surface, which NOAA began publicly releasing after the launch of the GOES-R geostationary satellites in 2017. Unlike public weather data, DSR estimates are available for every 0.5km^2 area. We show that the accuracy of solar performance modeling using satellite data and public weather station data depends on the cloud conditions, with DSR-based modeling being more accurate under clear skies and station-based modeling being more accurate under overcast skies.

1.2.2.2 Satellite-based Modeling using Multispectral Data

DSR data is released only hourly and are often not available, especially under over-cast sky conditions. While the raw multispectral data is available at much higher spatial and temporal resolutions—$\sim$1km^2 and every 5 minutes—than oktas—$\sim$25km^2 and every hour. Further, we observed a strong correlation between multispectral satellite data solar output and show in Chapter 6 [41] that ML-based models using these measurements are significantly more accurate than both physical models using derived data products (DSR), and prior okta-based solar models. We further develop both local and global ML models using data from both GOES-16 and GOES-17. Local models are trained on data from a specific solar site, while global models are trained on normalized data from many solar sites. This also enables accurate modeling for newly installed solar sites with limited historical generation data.

1.2.3 Solar Nowcasting

Finally, such high resolution satellite data opens up the possibility for better forecasting of solar generation. With the rise in residential solar PV installations, accurate solar forecasts are becoming increasingly important to allow seamless integration of solar in the energy supply. Solar forecasting models predict future solar output based on environmental conditions and are basically a two step process - predicting future weather conditions and second translating those predictions into solar output by using solar models. Accurate forecast can help the utility operators to anticipate the solar variations in advance and allow them to schedule the available energy reserves while reducing the need for storage. Near-term solar forecasts on the order of minutes to an hour are particularly important because homes and the grid must be able to adapt to large sudden solar fluctuations introduced by solar variations. The key challenge is to be able to predict the ramp up and down or the small changes in solar generation in the order of minutes to an hour. Current solar forecasting techniques which are primarily based on NWP algorithms, mostly leverage physics-

based modeling. These physics-based models are most appropriate for forecast horizons on the order of hours to days and not near-term forecasts on the order of minutes to an hour. While there is some recent work on analyzing images from ground-based sky cameras for accurate near-term solar forecasting, it requires installing additional infrastructure.

To address this, in Chapter 7 we discuss our self- supervised solar nowcasting models to predict near-term solar output using live multispectral data from GOES-R satellites. Our approach, unlike NWP-based models and ground-based sky cameras, can enable near-term solar forecasting at any location without additional infrastructure. Atmospheric changes, like cloud movement from one location to another over time, directly affects how the solar generation at these sites will change. Such changes are captured in this satellite data by processing an area around the solar site instead of using a single point value from the satellite. We introduce CNN-LSTM models to forecast a residence's solar output in the future based on the satellite's view of the surrounding area. This approach relies and builds on leveraging the spatio-temporal properties of the satellite data. By analyzing time series multi-channel data from these satellites at different resolutions, it is possible to indirectly capture the atmospheric changes thus resulting in accurate near-term solar forecasts.

1.3 Dissertation Overview

We organize the rest of the dissertation as follows. Chapter 2 provides the necessary background on solar control, modeling and forecasting that will serve as the basis for the rest of the thesis. Chapter 3 covers Sunshade, our SDS prototype, and illustrates enforcing static and dynamic solar rate control. Chapter 4 describes feasibility, cost, and carbon emissions of grid defection. Chapter 5 details solar modeling with the use of derived products from satellite and compares its performance against ground-based models. Chapter 6 introduces both local and global solar models using multispectral satellite data. Chapter 7 presents our work on near-term solar forecasting using multispectral satellite data. Finally, chapter 8 concludes the completed work and outlines possible future work.

CHAPTER 2

BACKGROUND

In this chapter, we provide background required for various aspects of this dissertation. We shed some light on the different components related to controlling, modeling and forecasting of residential solar.

2.1 Rising Solar and its Environmental Impacts

Today, the threat to climate change is higher than ever with the dramatic increase in carbon and other greenhouse gas (GHG) emissions. United States is on a path to achieve zero carbon emissions by 2050 while the power grid is set to become carbon free by 2035. This would require power systems to run on renewable energy like solar and wind which has zero carbon waste. Presently, electric grid is the largest contributor to greenhouse gases owning 28% of all emissions [32]. Solar energy is a renewable, zero carbon source of energy generation widely available in every geographic region, holding enormous potential to reduce the GHG emissions. With these growing environmental concerns and falling prices of solar PV modules, there is an increasing interest of utilities and consumers towards solar installations.

As per a report by the Solar Energies Industries Association (SEIA) and Wood Mackenzie, U.S. solar installations are expected to increase by 43% this year and residential PV solar capacity is expected to more than triple by 2024 [35]. With this increase in the installed capacity of the distributed solar, power is becoming less controllable. As per the laws of physics, demand and supply must be balanced at all times and the energy fed into the grid must be equal to the demand at that moment. Feeding power either higher or

lower than required can either result in power surge or power outage, damaging millions of connected electrical devices. This is why there is a need for accurate solar controlling, modeling and forecasting techniques for geographically distributed rooftop solar. Residential solar can only host a certain number of panels accounting for few kilowatts of energy. But these residential solar PV when aggregated over large scales together have the potential to generate more electricity than a power plant. The challenge here is that these rooftop solar are usually spread over large geographical area and governed by varying weather and physical conditions, which in turn affects each panel's solar output at any given time. It hence becomes imperative to accurately control, model and predict these multitude of small solar setups scattered over large area.

2.2 Solar Control and Storage

In 2013, the California ISO introduced "Duck Curve" that essentially depicts the difference in electricity demand and the amount of available solar energy throughout the day. This curve highlighted the potential of solar over-generation when the sun reaches the zenith and how the solar drops towards the evening as the sun sets and when the electricity demand from consumers rises [33]. This breakthrough evidently showed for the first time the problem with intermittent solar generation and how this problem scales up as the share of solar keeps on increasing in the grid.

With the increasing penetration of solar, policies are being scaled up to match renewable energy which now meets 22% of the global electricity demand [34]. Development of solar PV is only possible because of these carefully crafted state and federal policies which help towards their accelerated development. In order for the grid to maintain stability, there needs to be a close match between the utility's share of power and the power feeding in from the consumer side which varies because of the ever changing weather amongst many other factors. While the power demand and the solar generation profiles vary almost in real time, these policies of solar admission to the grid are mostly static and vary by states. These

policies in turn decides the market and tax pricing which in turn motivates or demotivates people from installing solar, selling their energy to the grid or getting defected from the grid.

2.2.1 Challenges

The challenge with solar is to provision for the time of under-generation i.e. time with snow, rain, clouds, and after sunset and also to cater to the problem of over-generation during the periods of high-sunshine. One of the solution in this regard is to install batteries i.e. using batteries to store extra charge during periods of high solar production and use them to meet the power demand during period of low solar generation. Batteries could also eliminate the need for solar curtailment without the fear of over-generation.

Battery storage at grid level seems impractical and expensive solution as a long-term alternative, though it still is a viable option over short-term. Peak energy demand including heating, cooling, etc. can be as much as 20 times the energy consumed on an average day [31]. Transitioning to these periods of high demand with these batteries will not be as easy as with conventional generators where we can increase the fuel intake to get the desired power output. Another problem with batteries is its sizing. Batteries are need to store the surplus solar charge while such peaks in solar generation are only a seasonal phenomenon leaving the batteries idle for the rest of the year. For a future powered by renewable energy, the grid needs to become more flexible and accommodating by better predicting, control and modeling of solar.

2.3 Solar Modeling

Traditionally, the grid operators regulated the generation schedules of the generators by using the load profiles of the consumers. Now, with the proliferation of distributed PV, consumers have become prosumers—by both producing and consuming energy. The lack of information about distributed residential PV output and irradiance data makes it challeng-

ing for the utilities and grid managers to predict how much demand will these prosumers reduce or instead feed back into the grid. This ultimately makes it difficult for them to decide how much electricity the grid should carry, which is a significant challenge. Hence, modeling these individual rooftop solar with respect to their behavior towards changing weather conditions, geographical location and characteristics like panel size, tilt, etc. is extremely important. In the absence of accurate solar output models for each of these individual solar sites, distributors either apply conservative estimates or apply generalized metrics while modeling the solar capacity. Clearly those approaches are not close to accurate and in turn will affect reliable operation of the grid. Accurate solar modeling which is not limited by historical data, or site-specific details can help in providing better solar estimates.

2.4 Solar Predictability

Solar prediction or forecasting allows the grid operators to know in advance the solar generation profiles under variable conditions. This does gives the utilities opportunity for scheduling their mix of generating sources in advance. Solar forecasting is basically a two step process - one, predicting the weather at the next time instants and two, translating those weather predictions into solar output via solar models to predict solar generation in the future. Accurate solar modeling and weahter predictions are both needed for accurate solar forecast. Both over or under estimation of the solar output can lead to power surge or power outage which the grid presently avoids by either imposing penalty charges on the grid market or by implementing reliability curtailment where operators can scale down the solar production to match the forecasted value. Both of these are costly ways of handling the inaccuracies in the prediction [30].

CHAPTER 3

CONTROLLING GRID TIED RESIDENTIAL SOLAR

In this chapter, we address the problem of solar control introduced by the large scale penetration of intermittent solar generation. We propose SDS systems that dynamically regulates the amount of solar power that flows into the grid by introducing fundamental mechanisms for programmatically controlling the size of solar flows. We implement an SDS prototype, called SunShade, and evaluate tradeoffs in the accuracy and fidelity of these mechanisms to enforce limits on solar flows.

3.1 Introduction

The electric grid is in the midst of a profound transformation, as users are increasingly generating their own energy locally from renewable sources rather than purchasing it from electric utilities. This transformation is being spurred by exponential decreases in the cost of solar modules, which have fallen 60% since 2011 [2]. Rapidly falling prices have in turn driven significant increases in distributed solar generation. Nearly all solar deployments are "grid-tied," enabling them to draw power from the grid when their local demand exceeds solar generation and feed power into the grid when their local solar generation exceeds demand. However, these grid-tied systems impose a burden on the grid to absorb a building's energy surpluses and make up for its energy deficits.

The electric grid was not designed to support such decentralized and intermittent energy generation by millions of individual users. Instead, the grid imposes a rigid top-down hierarchy where large, highly-regulated utilities generate (and purchase) energy to meet the demand of their customers and maintain grid stability. To do so, utilities continuously

balance electricity's supply and demand in real time by regulating generator power output. Since the energy demand (or "load") profile of individual users is stochastic, such real-time balancing is only possible because the sum of load profiles across many users tends to be smooth and highly predictable. As a result, utilities can plan when to activate (or "dispatch") generators in advance to satisfy large increases in demand.

Distributed solar generation at large scales fundamentally alters this paradigm by increasing the stochasticity of user load profiles, even when aggregating them. While solar power output can change instantaneously, e.g., due to passing clouds, dispatchable generators are mechanical devices that take some time to activate and adjust their power output, which prevents them from maintaining high power quality when compensating for rapid solar variations. In general, the grid faces significant operational challenges when renewable penetration approaches 10% [122], necessitating additional energy storage or sophisticated demand-side load management.

Due to the challenges above, states generally place hard limits ("caps") on the collective solar capacity that may connect to the grid. However, due to the rapid growth in solar power, users are now starting to hit these caps. For example, Massachusetts reached its cap in the summer of 2015, which immediately halted construction on 134MW of new solar deployments. The legislature did not pass a stop-gap bill to raise the cap until April 2016 [115]. In Hawaii, where 12% of residents have rooftop solar, utilities barred additional residents from installing grid-tied solar for two years until the government recently intervened [49]. Similar tensions now exist in Germany [133], Australia [103], and Italy [58].

The underlying reason for the caps above is that the grid exercises no control over when and how much solar power flows into it, enabling unlimited solar power to flow in from any connected solar-powered system, even when it might compromise grid reliability and power quality. To address the problem, we propose Software-defined Solar-powered (SDS) systems capable of dynamically regulating the power they let "flow" into the grid, similar to how the Internet and other networks strictly regulate data transmission. By adaptively

controlling the size of *solar power flows*, SDS systems provide the grid the necessary tools to control solar power in real-time to balance supply and demand, thereby reducing the need for artificial regulatory caps on solar connections. The current approach to allocating the grid's available solar capacity is essentially a static peak-based first-come-first-serve policy that only approves new solar connections if the grid can handle *all* solar connections operating at their peak capacity. This approach is highly inefficient, and wastes much of the grid's potential to transmit solar power.

The goal of SDS systems is to enable all users to freely connect to the grid and dynamically share its capacity to transmit solar power. This principle represents a form of *grid neutrality*, akin to net neutrality, where the grid treats all solar energy contributions equally without discriminating between users. Just as in the Internet, the rate at which users inject energy should be dynamically regulated to maximize the grid's available solar capacity (i.e., maximize goodput), maintain the supply/demand balance (i.e., prevent congestion collapse), and fairly share the capacity among connected users. Of course, controlling solar power differs in key ways from regulating data transmission. Since data transmission is packet-based, adjusting sending rates is simple and only requires regulating the time between packet transmissions. In contrast, solar power is continuous and thus requires different mechanisms for control. In addition, the maximum "sending rate" of a solar power flow varies continuously in real time based on physical properties, e.g., the weather and the sun, and is not a function of an arbitrary application's demands.

To provide a foundation for SDS systems, in this chapter we propose fundamental software mechanisms to control the "sending rate" of solar flows. Our first SDS mechanism enables software to directly control solar flows by placing an arbitrary *absolute cap* on solar power generation, akin to a network bandwidth cap. However, as we discuss, controlling solar flows using absolute caps may unfairly penalize large, or optimally configured, solar installations, e.g., by restricting the percentage of their maximum power that can flow into the grid. Thus, we introduce a new class of *Weighted Power Point Tracking* (WPPT)

algorithms, akin to weighted proportional-share allocation in networking, that enforce a relative limit on solar output as a fraction of its (dynamically changing) maximum power point. We implement these mechanisms in a prototype SDS system we have developed, called SunShade. In designing SunShade, we make the following contributions.

Software defined Solar-powered Systems. We introduce Software-defined Solar-powered (SDS) systems to enable innovation in the design of higher-level solar transmission protocols to support arbitrarily high solar penetrations in the grid. We contrast SDS systems with existing work on smart inverters and active solar curtailment, which enable specific operational modes and not software programmability.

Solar Flow Control Mechanisms. We design two fundamental mechanisms for enabling software to control solar flow rates, inspired by similar mechanisms in networking and operating systems. Absolute capping enforces hard caps on solar output, while WPPT enforces a relative cap based on a system's changing maximum power point. We define two WPPT variants—model-based and search-based—and examine their tradeoffs in complexity, accuracy, and performance.

Implementation and Evaluation. We implement a prototype SDS system, called SunShade, and evaluate the fidelity of the mechanisms above. In particular, we quantify the effects of variable conditions, such as clouds, passersby, and other shading, on the fidelity of a search-based WPPT algorithm, which must periodically deviate from its cap to discover changes in the MPP that affect the cap's accuracy.

3.2 Background

Solar-powered buildings and homes typically connect to the grid, enabling them to feed power into the grid and draw power from the grid as solar generation fluctuates. Such grid-tied systems are much more efficient (and less expensive) than "off grid" installations because they reduce (or eliminate) the need for local energy storage, which is expensive to install and maintain. Grid-tied systems are also inherently more efficient because they

enable buildings to exchange power to balance demand, i.e., by permitting one building to consume surplus solar power from a neighbor. However, injecting arbitrarily large amounts of intermittent solar energy into the grid is problematic, as utilities must offset any fluctuations in solar generation to maintain grid balance and power quality, e.g., by generating less power during solar surpluses and more power during solar deficits. Unfortunately, installing and maintaining enough energy storage capacity at grid scales to smooth fluctuations remains prohibitively expensive at high solar penetration levels, e.g., $>10\%$ generation from solar.

As a result, governments strictly regulate grid solar connections, typically by setting a hard limit on the maximum solar capacity that may connect to the grid. These limits—set statically through legislation—are generally low. In Massachusetts, the cap that was hit in summer 2015 was set at 4% of peak grid load [115]. Admission control policies for solar installations based on static peak-based capping are highly wasteful because they prevent fully utilizing the grid's capacity to transmit solar power. To understand why, consider that solar-powered systems are rarely generating at peak power due to non-optimal operating conditions, e.g., from clouds, shade from obstructions, early morning or evening hours, non-optimal tilts/orientations, seasonal variations, etc. In fact, solar installations are capable of generating their peak power at only a single instant over the entire year: at solar noon on the summer solstice, assuming clear skies. Thus, generating the same amount of solar power on cloudy days or in the winter requires energy contributions from more solar-powered systems than on sunny summer days. Unfortunately, static peak-based caps prevent new solar connections unless the grid can safely service all systems operating at their peak.

Any approach to dynamically regulating solar output must ensure *fairness* between solar systems. We define an allocation of solar flow rates as fair if over some time period τ, the solar flows are able to contribute the same percentage of their maximum possible generation potential to the grid. This definition of fairness normalizes for the size of the

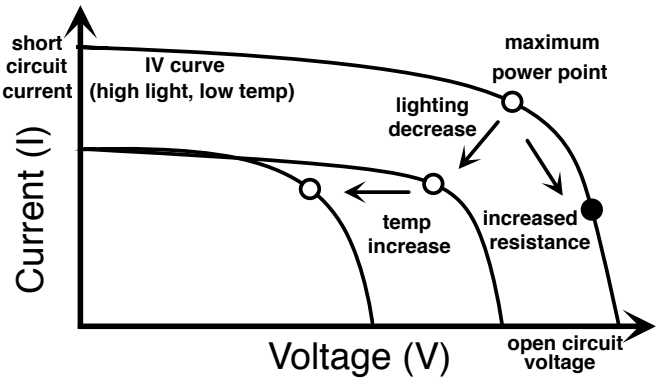
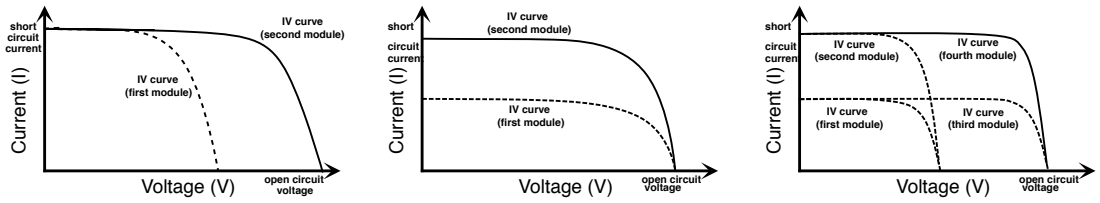


Figure 3.1: I-V curve for a typical solar module, and the effect of changes in lighting and temperature.

deployment, and also incentivizes an optimal configuration. For example, physical characteristics, such as non-optimal tilts and orientations and shade from surrounding buildings, lower a deployment's maximum generation potential and thus its fair allocation. Unfair allocations of solar flows are undesirable because they reduce the compensation users receive for the solar energy they contribute to the grid, and increase users' local energy storage requirements (to store the energy they cannot contribute). This chapter narrowly focusses on introducing new mechanisms to enforce setting proportional flow rates in SDS systems. Using these mechanisms to globally and dynamically regulate solar flows across multiple SDS systems to ensure fairness is outside of our scope.

3.2.1 Background

SDS systems enable software to *control* the size of solar power flows into the grid. Modern solar-powered systems already actively control solar power output within their *inverter*, which converts the DC electricity generated by the solar modules into AC electricity that is synchronized with the grid's AC electricity, e.g., to the same frequency and phase. Inverters typically implement an embedded algorithm for Maximum Power Point Tracking (MPPT) that constantly adjusts the deployment's operating voltage to maximize its power generation, as the current produced by solar modules varies non-linearly with voltage. However,



(a) Array in Series (b) Array in Parallel (c) Array Series-Parallel

Figure 3.2: Example I-V curve for a solar array formed from wiring multiple solar modules in series (a), in parallel (b), and a combination of the two (c).

existing solar inverters generally do not expose such control mechanisms to higher-level software. As we discuss, new smart inverters support other operating modes that implement embedded control algorithms beyond MPPT, e.g., VAR control, voltage/frequency ride-through, etc., but do not permit programmatic control by higher layers. In contrast, we focus on exposing programmatic interfaces to leverage similar inverter control mechanisms for regulating solar power output.

The primary factor that affects a solar deployment's maximum possible production is its solar insolation, i.e. the amount of solar radiation that is incident on the solar modules' area. The amount of solar insolation is affected by numerous variables, including the weather, angle of the sun in the sky (which varies across the day and year), shade from neighboring buildings and trees, modules' tilt and orientation, etc. Given these factors, a typical solar module is capable of operating at a range of different current and voltage levels, which govern its actual power output. The operating region of a solar system is governed by its *I-V curve*, as depicted in Figure 3.1. The figure shows a solar module's output current across a range of voltages (as dictated by the applied resistance), where the solar power output is simply the product of the voltage and current. Due to the nature of the I-V curve, the solar output power *changes* at different operating voltages. Specifically, since the I-V curve is initially flat, as the operating voltage increases, the output current remains virtually unchanged, leading to an *increase* in power output. However, after reaching the knee of the curve, any further increase in operating voltage yields a corresponding reduction in current,

and hence the power begins to drop. Thus, the solar output rises with increasing voltage up to a point and then precipitously drops. As a result, each I-V curve has an optimal operating voltage V_{opt} that maximizes its output.

Note that the precise shape of the I-V curve is dynamic and changes continuously. For example, the maximum power point decreases as the solar insolation decreases, causing the curve to contract along both the x-axis and y-axis as depicted. In addition, the solar cell temperature also affects the shape of the curve, expanding and contracting it along the x-axis. While Figure 3.1 depicts an idealized curve for a single solar module, solar systems are typically composed of multiple modules wired (or "strung") together and connected to a single inverter. In this case, the I-V curve of the aggregate solar circuit is a combination of the I-V curves of each module. Figure 3.2 shows how the combined I-V curve is a composition of each module's I-V curve when wiring modules in series (a), in parallel (b), and a combination of the two (c). In particular, two modules wired in series operate at the same current, but have additive voltage, while two modules wired in parallel operate at the same voltage, but have additive current. The characteristics of each module may then change independently, affecting both the output of the other modules and system's aggregate I-V curve. For example, two connected modules may be installed with different tilts at different orientations, causing a shadow to cover one but not the other. If wired in series, the module producing the lowest current restricts the current generated by the other modules, reducing the entire array's output.

MPPT algorithms dynamically adjust the system's voltage to maximize power generation by operating at the "knee" of the I-V curve as the curve changes. Inverters implement MPPT algorithms using a DC-to-DC buck-boost converter that is able to adjust output voltage to be greater than or less than input voltage. Buck-boost converters typically use pulse-width modulation (PWM) to vary their duty cycle, which also varies the input/output voltage. There is a large body of prior work on developing Maximum Power Point Tracking (MPPT) algorithms—examples include the perturb and observe, current sweep, incre-

mental conductance, and constant voltage ratio algorithms amongst many others. MPPT algorithm design is well-studied and presents many tradeoffs in optimizing power accuracy, convergence speed, implementation complexity, initialization procedures, etc. [46].

The most common MPPT algorithm is the Perturb and Observe (P&O) algorithm. This algorithm perturbs the voltage by a small amount, and then measures the instantaneous current and voltage to calculate the new power (P_t) and compares it to the power P_{t-1} at the previous voltage. If the change in power is positive, it continues to perturb the voltage in the same direction; if the change is negative then it reverses the direction of its search. Simple P&O algorithms use a fixed voltage step size on each iteration, while more sophisticated variations adapt the step size, e.g., proportional to the slope of the P-V curve $\frac{\Delta P}{\Delta V}$, to converge more quickly.

3.3 Controlling Solar Flows

SDS-enabled inverters use the same basic functions as MPPT to regulate solar power flows by operating at points other than the maximum power point (MPP). For example, an inverter could decrease the output below the MPP, or after decreasing output, could then increase output back to the MPP. By controlling the operating voltage, an inverter can precisely control solar output up to its MPP. Note that inverters may be used in conjunction with solar charge controllers, enabling them to either dissipate excess power when operating below the MPP (if there is no battery) or store the excess power in a battery. In the latter case, the solar charge controller circuitry implements the MPPT algorithm.

Interestingly, while solar power is characterized as inherently intermittent, only its MPP is intermittent. Below the variable MPP, solar inverters are capable of making rapid and precise changes in solar output, and generally have more flexibility to rapidly and precisely control power output than mechanical generators, which have physical limitations on the speed at which they can increase or decrease their revolutions. As we have discussed earlier, active control of solar power to support increased solar penetration is an emerging

area. However, prior work focuses largely on coarse solar curtailment during times of peak generation (akin to coarse demand response), and not on exposing mechanisms for fair, fine-grained regulation of solar output to higher-level software.

3.4 SDS Control Mechanisms

We design two useful mechanisms to control solar flows. Our first mechanism enforces an *absolute cap* on solar output, wherein the solar output is capped based on a specified limit. Formally, this cap is specified by a tuple (P_{cap}, t), which imposes an upper limit P_{cap} on the power for a duration t. Our second mechanism is a class of Weighted Power Point Tracking algorithms that enforce relative caps, such that power output is capped as a percentage of the system's time-varying maximum output. Formally, this cap is specified by a tuple (Δ, t) which indicates that power should be limited to a fraction Δ of the maximum power over the duration t, where $0 < \Delta \leq 1$. Note that, since the maximum power output is constantly changing, the absolute power generated from the relative cap also changes. While absolute capping enforces a strict power limit, WPPT enforces a "fair" limit across deployments with different characteristics.

Our SunShade prototype exposes a narrow interface that enables software to set and alter either the absolute cap P_{max} or the weight Δ. Note that this interface *does not* expose direct programmatic control of the voltage, but rather internally determines the appropriate operating voltage to enforce the specified power output. This is akin to software-defined networks that expose forwarding mechanisms to an external controller in the control plane, but do not expose direct control of the data plane's packet processing. Exposing direct control of voltage lowers the barrier to introducing deviant behavior into the grid, and could enable sophisticated grid attacks. In the past, grid interconnection standards have prevented inverters from actively adjusting their power output outside of using MPPT (see IEEE 1547-2003 [73]). However, these standards are changing to permit the basic control functions we

propose (IEEE 1547a-2014 [74]), as solar capacity increases and smart inverters become more commonplace.

3.4.1 Absolute Power Capping

As discussed in the previously, the Perturb and Observe (P&O) algorithm is the most widely used algorithm for tracking the MPP. To support absolute power capping, we adapt the classic P&O algorithm to ensure the solar output operates at or below a specified power limit. Algorithm 1 shows the pseudo-code for setting an absolute power cap on solar output. Similar to the P&O algorithm, the algorithm uses the instantaneous voltage and current to calculate the current power P_t. Power P_t is then compared to the previous power P_{t-1} to determine if there has been a change in power.

To operate at a given power cap P_{cap}, the algorithm simply compares P_{cap} to the current power P_t. If P_t is less or more than P_{cap}, the voltage is perturbed to increase or decrease, respectively, the power. To increase and decrease the magnitude of power, the instantaneous power P_t is compared to previous power P_{t-1}. If the change in power is positive, the algorithm continues to perturb in the same direction, else the direction of the perturbation is reversed. Upon reaching the limit P_{cap}, the instantaneous power output then oscillates around the limit. In this case, a bigger step size results in larger oscillations around the limit, but faster convergence, while a smaller step size has smaller oscillations, but has a slower convergence. As a result, a larger step size is more appropriate in scenarios where conditions are changing rapidly, while a smaller step size is more appropriate under stable conditions. While this algorithm uses a fixed voltage step size, we can also use an adaptive voltage step size proportional to the difference between P_{cap} and P_t to converge faster after large variations. The only external input signal required is the absolute power cap P_{cap}. If the current maximum power point P_{mpp} is higher than the allowed limit, the system adjusts to operate at P_{cap}, else it operates at P_{mpp}.

Algorithm 1 Absolute capping via modified P&O algorithm.

```
if Pcap != P
    if Pcap > P
        %increase power
        if P > Pold
            if V > Vold
                D = Dold - deltaD;
            else
                D = Dold + deltaD;
            end
        else
            if V > Vold
                D = Dold + deltaD;
            else
                D = Dold - deltaD;
            end
        end
    else
        %decrease power
        if P > Pold
            if V > Vold
                D = Dold + deltaD;
            else
                D = Dold - deltaD;
            end
        else
            if V > Vold
                D = Dold - deltaD;
            else
                D = Dold + deltaD;
            end
        end
    end
else D=Dold;
end
```

3.4.2 Weighted Power Point Tracking

Weighted Power Point Tracking (WPPT) caps the power output such that it maintains output at a fixed fraction of the maximum power point P_{mpp}. Software sets the weight Δ between 0 and 1. To strictly enforce a weighted cap at any given time, the system must know the maximum power point P_{mpp} to compute the appropriate weighted power point. Thus, in this case, the absolute power cap is $P_{cap} = \Delta * P_{mpp}$ and changes dynamically over time. The challenge with WPPT is that the maximum power point is not well-known, and there is a cost to finding it. We describe two approaches to determine the P_{mpp} below.

After determining P_{mpp}, we compute the weighted cap ΔP_{mpp} and use our absolute capping P&O algorithm above to maintain the weighted cap.

3.4.3 Search-based Approach

The search-based approach operates by periodically computing the actual maximum power point P_{mpp} using the standard P&O algorithm (or any other MPPT algorithm). The fidelity of the search-based approach in tracking the true weighted power point is a function of the variability in the MPP, the convergence speed of the MPPT algorithm, and the search interval. For example, if MPP variability is high and we search infrequently, then the fidelity will be low, as our weighted cap will not accurately reflect the actual cap (which is changing as a function of the MPP). In contrast, if variability is low, and we search frequently, then we will deviate from the weighted cap in searching, and thus generate more power than the cap dictates, and introduce spikes in the system's power output.

Our search-based approach invokes the P&O algorithm at a configurable interval with a specified search duration until the algorithm oscillates for multiple iterations around the maximum power point. The algorithm then halts and returns the P_{mpp}. The algorithm then uses the absolute capping algorithm to find the voltage V_{cap} that results in an absolute power cap of ΔP_{mpp}. This voltage will change based on weather conditions and time. The search-based approach may observe the changes in power at V_{cap} to determine when the MPP has deviated from previous MPP and trigger a new search. Our implementation supports setting a fixed search interval or an adaptive search interval that triggers a new search once V_{cap} has deviated by a configurable threshold.

3.4.4 Model-based Approach

The search-based approach has accuracy limitations because it must deviate from the true weighted cap to find the current MPP. These deviations may reduce accuracy to unacceptable levels under highly variable conditions that requires frequent searching. An alternative approach is to compute the MPP based on a model of the solar system's maxi-

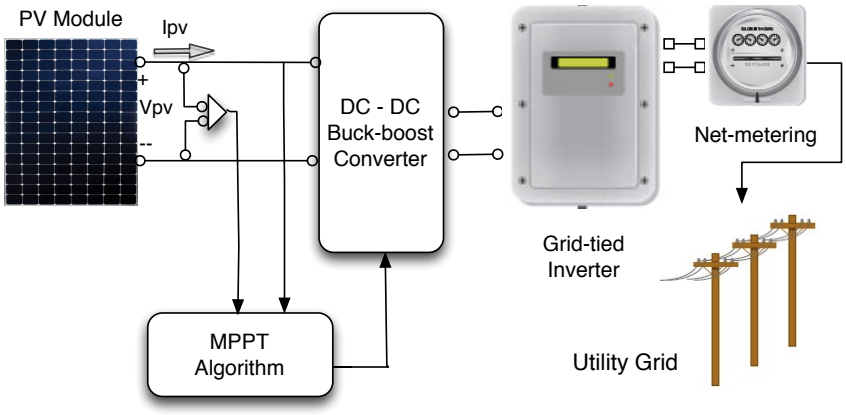


Figure 3.3: Depiction of our SunShade simulator.

mum output at any given time. This model may be constructed either empirically based on data collected by the inverter and weather sensors, or analytically given specifications of the solar panels, including their type, tilt, orientation, wiring, etc.

Building an empirical model has the advantage of not requiring a priori knowledge of the deployment, since the system behavior is learned from empirical observations. However, empirical models take time to build, as they require collecting data on the maximum power point under many different environmental conditions. To build such an empirical model, the inverter would use conventional MPPT to operate at its maximum power point for a long period of time to collect current and voltage values under many different ambient conditions with different temperatures and solar radiation levels. While the temperatures and solar radiation levels could be estimated from a local weather station, e.g., via Weather Underground, it is more accurate to link the inverter with external temperature and solar radiation sensors to record actual conditions. After recording the current and voltage at the maximum power point for many different values of temperature, solar radiation, and time, it can use standard techniques to build a model that predicts the current, voltage, and maximum power point for any values of temperature, solar radiation, and time [76].

The amount of time and data an inverter must collect to build an accurate model varies with each location. For example, in San Diego, CA, where the climate is nearly constant

year-round at 24°C and sunny, an accurate model will take little time to build, while in a highly variable climate, such as in the Northeast, it might take an entire year. As a result, we focus on an approach that analytically models a solar deployment based on its specific characteristics. While this approach requires configuring the inverter with details specific to each deployment, it requires no extended period of operation to collect training data. In the model-based approach, we are given the deployment specifications that dictate a model of the I-V curve for the deployment based on solar radiation and temperature. Again, we assume that the inverter uses external sensors to measure solar radiation, e.g., using a pyranometer, and temperature at the location. The model then infers the MPP based on the radiation and temperature levels. Note that no model is perfect; thus, the fidelity of this approach is ultimately a function of the model's accuracy. Another drawback of the model-based approach is that it requires irradiance and temperature sensors, which add to the cost and complexity of a solar deployment.

3.5 Implementation

We implement our SDS rate control mechanisms in simulation and in a small-scale SunShade prototype. Our simulation leverages Matlab's Simulink library (SimPowerSystems) for simulating a solar deployment's output based on its electrical characteristics, irradiance, and temperature.

Matlab includes a flexible solar cell model that we configure to match the panel from our small-scale prototype, and an implementation of the P&O MPPT algorithm that tracks maximum solar output as a function of solar radiation and temperature. The model shows a close, albeit imperfect, fit to the published data provided by the manufacturer. In order to compute the power cap P_{mpp}, our algorithm measures irradiance and temperature from the sensors, which it provides as input to the model. Our simulator implements absolute power capping and the two WPPT variants from the previous section by modifying the existing P&O algorithm in Matlab. Our simulator takes as input, data traces gathered from

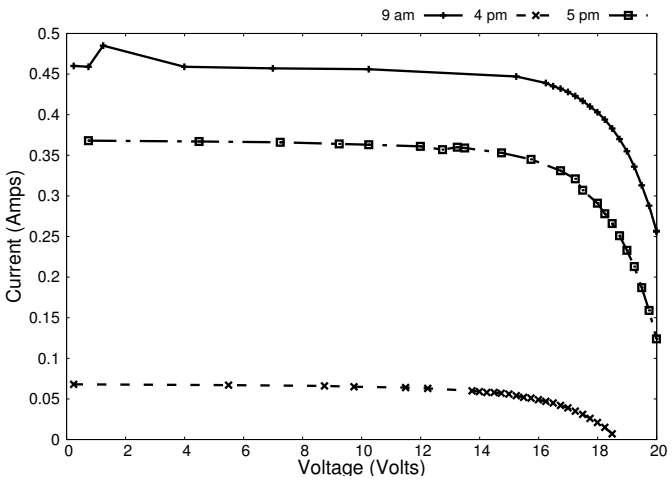


Figure 3.4: Expansion and contraction of the I-V curve for our SunShade prototype, as light intensity changes throughout the day.

a pyranometer that measures solar radiation, and a temperature sensor. We are also able to generate synthetic traces of solar radiation and temperature for the simulator to test the fidelity of the mechanisms above under arbitrary conditions. Note that Matlab simulations are considered highly accurate and frequently used as the only means of evaluating new MPPT algorithms in the power systems community. Figure 3.3 depicts the circuit diagram of our SunShade simulator.

In addition to our simulations, we also construct a SunShade prototype to evaluate its mechanisms under realistic conditions. For our prototype, we connect voltage and current sensors between a load and a small solar panel (rated at a ideal peak capacity of 25W), which independently measure voltage and current. Rather than employ an embedded buck-boost converter, we use a programmable load—the BK Precision 8500 Programmable Load—to control the panel's operating voltage. Using the programmable load enables rapid experimentation by allowing us control operating voltage remotely from a server via `python`, rather than embedding such control into the buck-boost converter's firmware. Note that the programmable load is functionally equivalent to the buck-boost converter, and uses the same PWM mechanism to vary the panel's operating voltage. The primary difference is that the minimum reaction time—the time between two changes in voltage—

on the programmable load is $\sim$100ms due to the latency imposed by the serial connection. We could reduce this latency to near that provided by a typical buck-boost converter, e.g., tens of milliseconds, by using a modern I/O interface, such as USB. Figure 3.4 shows current and voltage from our SunShade prototype that reflect the expansion and contraction of the I-V curve, similar to Figure 3.1 from , as the intensity of light changes over the day.

Finally, evaluating WPPT's fidelity requires comparing its results to the actual weighted power point dictated by the real MPP. To support such comparisons, we construct an additional parallel prototype to run MPPT that uses the same solar panel as SunShade. We then place the two systems directly adjacent to each other so they are subject to nearly identical solar conditions. Figure 3.5 shows a picture of our SunShade prototype with its key components labeled. The algorithm logic runs on a Raspberry Pi, which connects to both an external current sensor—to read changes in current—and to a programmable load to programmatically alter the solar panel's operating voltage. To support model-based WPPT with sensors, our prototype also includes a pyranometer (for sensing solar radiation) and a temperature sensor.

3.6 Evaluation

We first evaluate the performance and fidelity of SunShade's mechanisms in simulation, as our simulator is able to support a much wider range of experimentation, i.e., covering a range of conditions, compared to our prototype. We then examine the performance and fidelity of our SunShade prototype. To evaluate SunShade's fidelity, we use the Normalized Root Mean Squared Error (NRMSE), which is a common metric for quantifying the difference between two time-series. We compute the NRMSE between SunShade's capped values and the ideal values. In this case, an NMRSE closer to one is better, as it indicates the two time-series are similar. The equation for NMRSE is below, where $||$ denotes the 2-norm of the time-series vector.

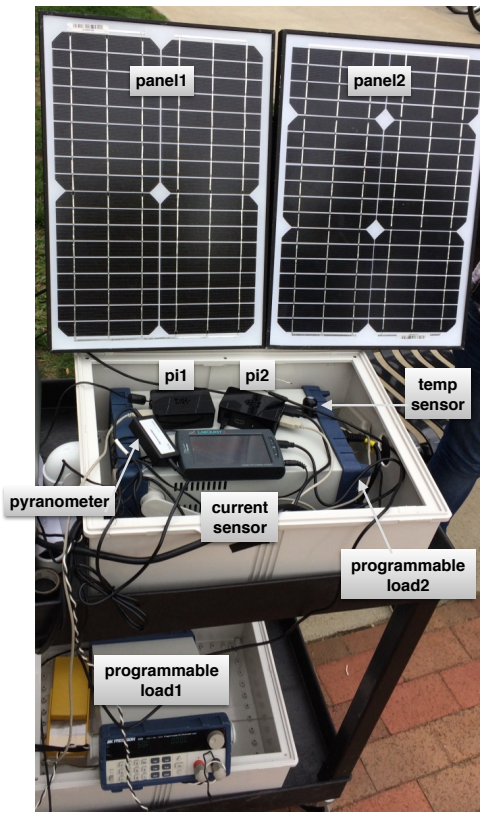


Figure 3.5: Depiction of our SunShade prototype.

$$\text{NRMSE} = 1 - \frac{||actual - estimated||}{||actual - mean(estimated)||} \tag{3.1}$$

3.6.1 Simulation Results

Figure 3.6 demonstrates SunShade's performance in simulation using absolute power capping on a cloudy day. In this case, the clouds are not strong enough to cause the panel output to drop below the cap for most of the middle part of the day. As a result, the system has a near-steady 100W power output with minor oscillations around 100W due to the P&O-based capping algorithm that constantly perturbs voltage searching for the cap. As expected, absolute capping is relatively straightforward as it requires no external knowledge about the MPP. In contrast, WPPT is more challenging, as it requires a prediction (or search) of the MPP.

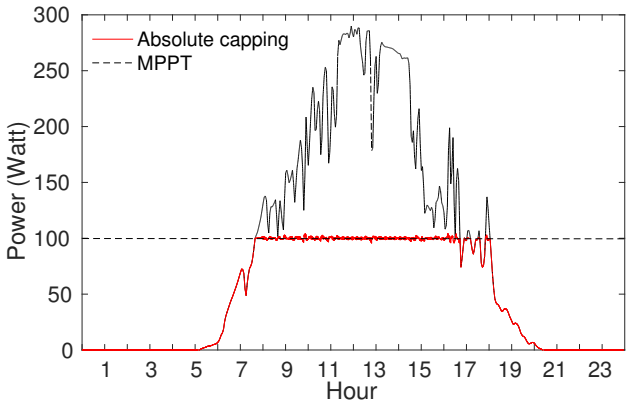


Figure 3.6: SunShade capping the absolute power of a solar panel to 100W.

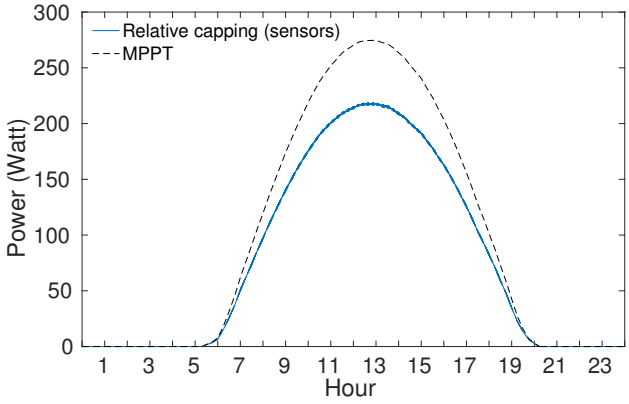


Figure 3.7: Example of model-based weighted power capping at 80% of the MPP.

Figure 3.7 shows the performance of model-based WPPT (relative capping using sensors) at 80% of the MPP in simulation. Since our model, as with most models, is most accurate when the sun is shining, the 80% cap is a near perfect reflection of 80% of the MPP in real-time. Figure 3.8 shows the same model-based weighted power capping at 80% on a cloudy day with significant variations in power. The fidelity of WPPT is slightly less (in terms of NRMSE) under cloudy conditions, as the model may be less accurate and there is more time spent searching for the weighted cap. Likewise, Figures 3.9 and 3.10 show the performance of the search-based algorithm (relative capping using no sensors) for the same simulated days as above. In this case, we conduct a search at a fixed interval every

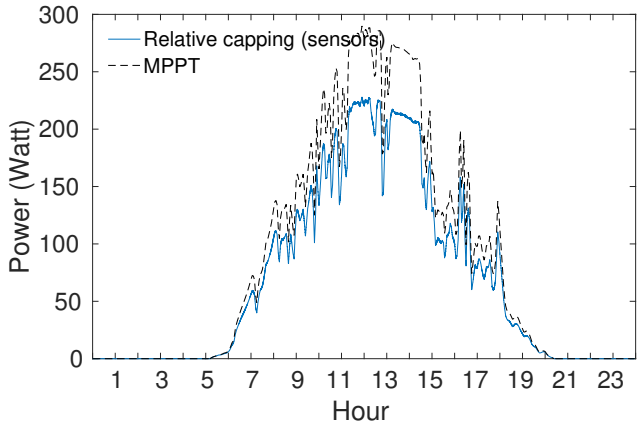


Figure 3.8: Example of model-based weighted power capping at 80% of the MPP.

15 minutes. The results show that the search-based algorithm frequently deviates from the relative cap to find the MPP in order to reset the cap.

Table 3.1 shows the NMRSE for each of the mechanisms. The results show that absolute power capping has the highest fidelity, i.e., is closest to the ideal power time-series, since it is the simplest mechanism. In addition, weighted power capping using an accurate model is close to the performance of absolute capping, since it is able to accurately adjust the power cap in real-time without deviating from it. The degradation in fidelity of the approach in our simulations stems primarily from searching for the cap, similar to how MPPT algorithms must search for the MPP. Finally, the last row shows that the NRMSE for search-based WPPT is the lowest for both the sunny and cloudy day. The search-based approach on the sunny day is only slightly lower than the other approaches, despite conducting a search every 15 minutes. However, on the cloudy day the performance is less than the other algorithms, as the 15 minute search interval is too long relative to the frequency of environmental variations.

While the experiments above demonstrate the behavior of our power capping mechanisms for representative sunny and cloudy days, Figure 3.11 demonstrates how the fidelity of capping changes for the different mechanisms as the frequency of variations increases. In this graph, we subject SunShade to synthetic power fluctuations that increase in frequency

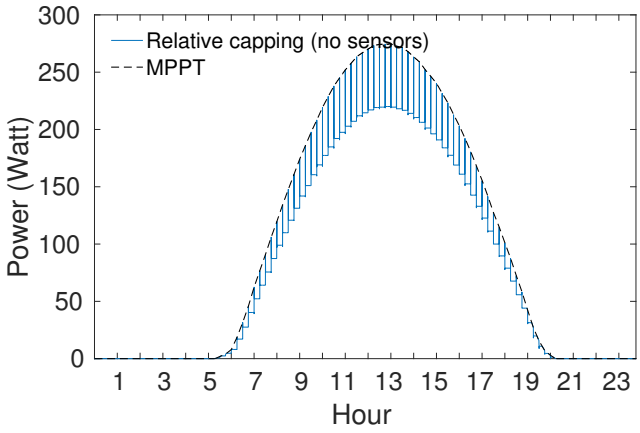


Figure 3.9: Example of search-based WPPT at 80% MPP on a sunny day.

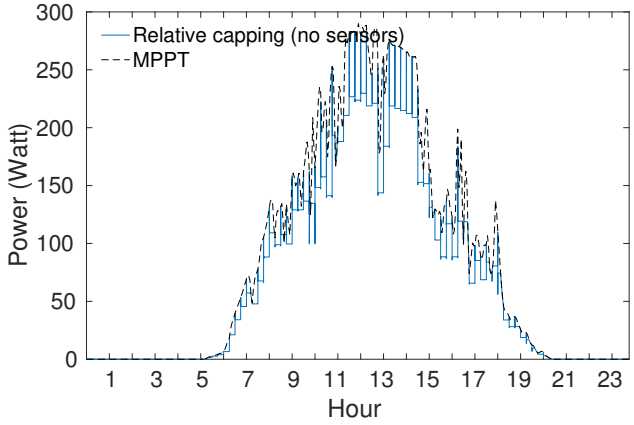


Figure 3.10: Example of search-based WPPT at 80% MPP on a cloudy day.

along the x-axis. Thus, the higher the frequency on the x-axis the more variance in the power output of the solar panel. The graph shows that for both absolute power capping and model-based WPPT (relative capping using sensors) are much less sensitive to variations than the search-based WPPT (relative capping using no sensors).

3.6.2 Prototype Results

Due to the complexity of modeling a solar module, we use our SunShade prototype to primarily evaluate the search-based WPPT algorithm. Figure 3.12 shows power measurements from our SunShade prototype running over a three hour period on a relatively sunny

Algorithm	Sunny	Cloudy
Absolute	.9888	.9843
WPPT (model-based)	.9874	.9826
WPPT (search-based)	.9677	.8628

Table 3.1: NRMSE comparison for sunny and cloudy weather conditions

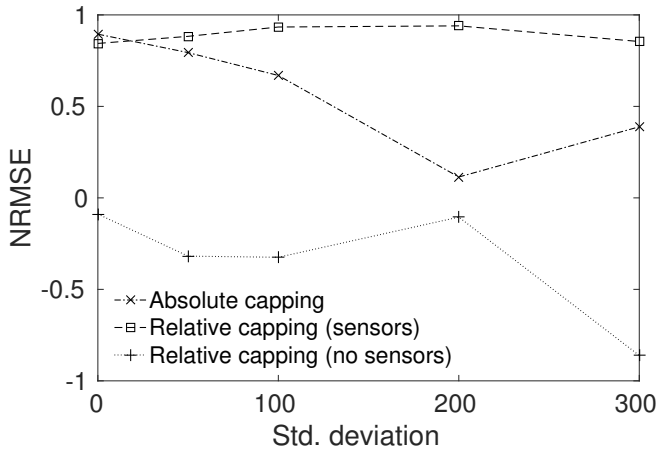


Figure 3.11: NRMSE for periods with different levels of variation.

day. The graph includes results from our two systems running in parallel: one running the search-based WPPT (in black) and one running a typical P&O MPPT algorithm (in red). From the MPPT measurements, we compute the ideal WPPT value. In this case, we set the WPPT weight Δ to be 0.5 or 50% of the maximum value. As the graph shows, WPPT periodically searches (every 15 minutes here) for the MPPT to set a new WPPT weight, causing its power output to increase until it converges to the maximum power point, and then to decrease.

At some time periods, clouds cause the WPP to diverge from the ideal between a search interval. For example, near 11am there are passing clouds that cause the MPPT to decrease. During this time period, the WPP also diverges more from the ideal between each search interval. In contrast, after 11:30am, there is little change in the MPP, and thus the WPP tracks the ideal WPP nearly perfectly. This experiment also stressed our WPP implementation due to people passing by the prototype and briefly shading it. This shading is evident in

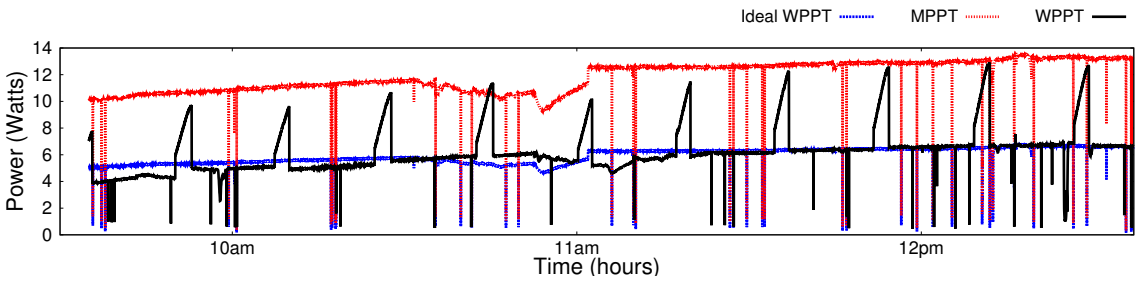


Figure 3.12: WPPT of SunShade in a real deployment compared with ideal WPPT based on true MPPT

the graph from the periodic dips in both the MPPT and the WPPT. Since these dips caused power output to drop to near zero, they affected both the MPPT and WPPT algorithm even within a search interval. In these cases, WPPT was unable to maintain its power cap due to little available power, causing it to match the MPPT algorithm even without searching for the new MPPT. Note that in a few cases, the passersby shaded the two adjacent panels unevenly, causing only one of them to drop its output.

In addition to the illustrative experiment above, we also examine the impact of changing the search interval, voltage step size, and weight on the fidelity of WPPT compared to the ideal WPPT. Figure 3.13 shows how the Root Mean Squared Error between WPPT and the ideal WPPT differs as a function of the search interval, which ranges from 30 seconds to three minutes. For each datapoint, we run the WPPT algorithm for 15 minutes with a default voltage step size of 0.5, a default interval of one minute, and a default weight of 50%. Each graph then adjusts one dimension and observes the effect on the RMSE. The proper search interval is a function of the current conditions and the variability of solar output. Under high solar variations, a smaller interval is better, as the ideal weighted cap is changing frequently, while under near constant solar output, a longer interval is better since the MPP is not changing. Here, as the search interval increases, the error also increases, since the weather conditions during this experiment were partly cloudy.

Figure 3.14 then plots the RMSE as a function of the weight setting. Since lower weights deviate more from the MPP, they take longer to search for the MPP and deviate

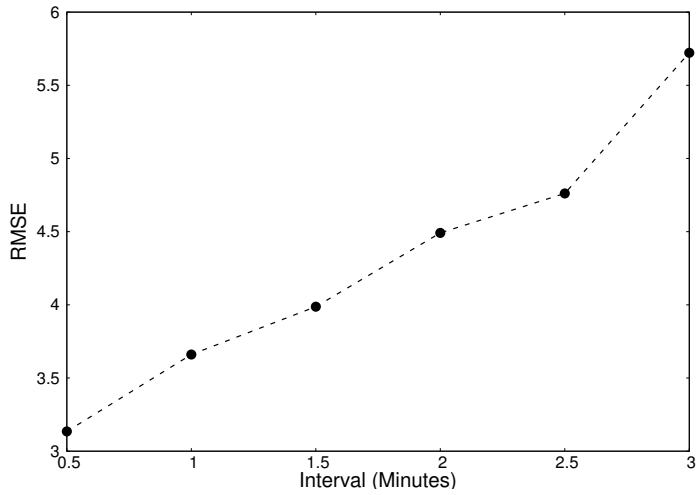


Figure 3.13: The RMSE between the ideal WPPT and search-based WPPT as a function of the periodic search interval.

more from the ideal WPP. The graph demonstrates this trend as the lower the weight setting, the higher the RMSE for our WPPT tracking algorithm. One way to address this issue for lower weights is to save the previous MPP value and immediately start searching from the previous value, rather than from current voltage setting. Of course, this approach is not ideal during highly variable conditions, where the MPP might change significantly. Finally, Figure 3.15 plots the voltage step size as a function of WPP. Similar to the interval above, the tradeoff in the step size is dependent on the conditions. Under highly variable conditions that require more frequent searching, a larger step size is more desirable as it makes each search faster. Even though the larger step size decreases accuracy, since conditions are highly variable this is outweighed by the faster search time. In contrast, under stable conditions that require fewer searches, a smaller step size is better, since it finds a more accurate weight, which is important because searches occur infrequently. Here, since conditions were variable, we see that, as we increase the voltage step size, the RMSE increases.

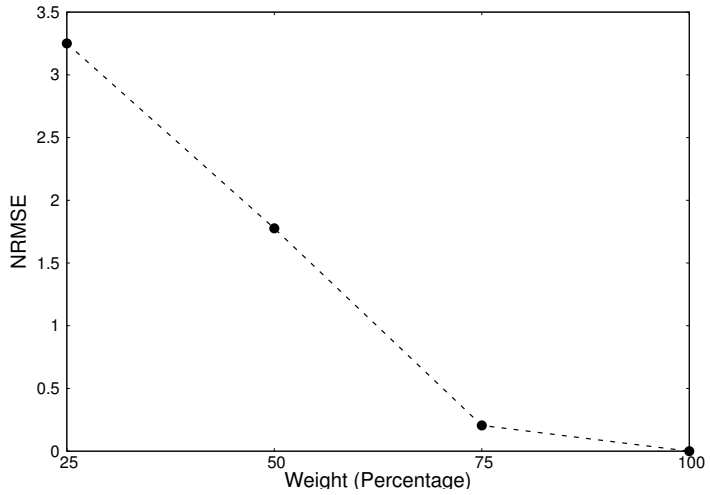


Figure 3.14: The RMSE between the ideal WPPT and search-based WPPT as a function of the weight.

3.7 Related Work

While prior work has also advocated applying Internet design principles to the overall electric grid [69, 81], SDS systems focus narrowly on applying these design principles to solar systems for multiple reasons: in particular, solar power is the predominant source of distributed generation, is growing rapidly, and is programmatically controllable. Since solar cells are silicon-based semiconductors, their output can be programmatically controlled between zero and their maximum output based on the intensity of light.

SDS and SunShade is related to prior work on active solar power curtailment. However, this chapter primarily focuses on sensing and responding to specific situations where an inverter may need to reduce or eliminate solar power output. For example, all grid-tied inverters are able to sense a power outage and reduce output to zero to prevent energizing downed power lines. In addition, there is significant prior work on reducing solar power output during over-voltage situations [86,88,125–127]. This research differs from our work in that it focuses on specific algorithms and policies embedded in the inverter that respond to specific situations. Instead, our goal is to expose programmatic interfaces to fundamental mechanisms for rate limiting solar power. As a result, SDS decouples mechanism from the

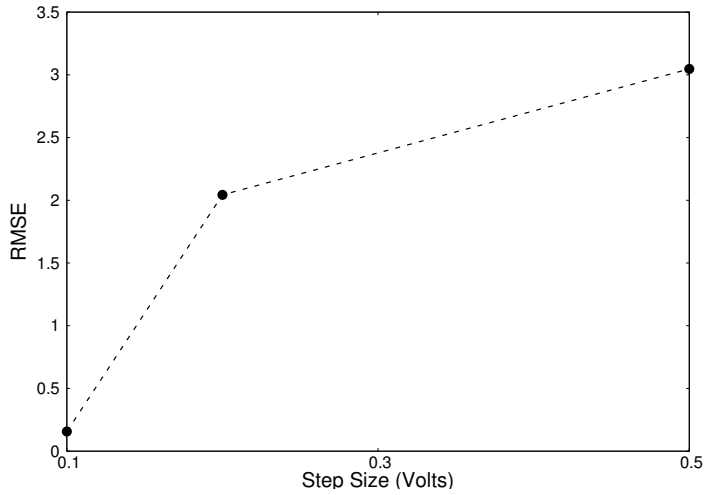


Figure 3.15: The RMSE between ideal and search-based WPPT as a function of the voltage step-size.

policy: while its mechanisms could be used to cap power in response to increased voltage, they could also be used in other contexts.

Recent advancements in smart solar inverters have also recognized the potential benefits of controlling solar power to support grid operation [128]. In contrast, similar to TCP, SDS focuses, not only on managing solar generation to support grid stability, but also on fairly sharing the grid's available capacity to accept solar power with the goal of maintaining grid neutrality. In addition, prior work typically focuses on low-level power systems and power electronics issues. However, the introduction of smart inverters with sophisticated controls is raising the grid's level of abstraction. Similar to the Internet, we expect the grid to evolve into a layered architecture, where the physical layer (layer 2) addresses challenges in power systems and electronics and the higher layers address challenges in capacity management, fair-sharing, quality-of-service, etc. This chapter demonstrates mechanisms necessary to address these higher-level problems.

3.8 Conclusions

This chapter proposes Software-defined Solar-powered (SDS) systems to control the flow rates of solar power into the grid. The goal of SDS systems is to eliminate the need for policies that artificially cap the number and size of solar deployments that can connect to the grid, and instead dynamically rate- limit them if they exceed capacity in real time. To provide a foundation for SDS systems, we present two fundamental software mechanisms to control solar flow sending rates, including an absolute capping mechanism and a class of WPPT algorithms that enforce a relative cap. We implement a prototype SDS system, called SunShade, and evaluate the fidelity of these mechanisms for controlling solar flow rates, and their tradeoffs in terms of accuracy and responsiveness. In the next chapter we discuss our contributions towards controlling off-grid solar.

CHAPTER 4

CONTROLLING OFF GRID RESIDENTIAL SOLAR

As we discussed in the previous chapter, the limitations imposed by the government to limit the new grid connections are likely to become more restrictive over time in many areas as solar disrupts the utility business model. Thus, to employ solar without restrictions, users may increasingly need to defect from the grid. Unfortunately, batteries alone are unlikely to become cost-efficient at enabling grid defection for the foreseeable future. To address the problem, we explore using a mixture of solar, batteries, and a whole-home natural gas generator to shift users partially or entirely off the electric grid. In this chapter, we assess the feasibility and compare the cost and carbon emissions of such an approach with using grid power, as well as existing "net metered" solar installations.

4.1 Introduction

Distributed solar generation on rooftops has been rising rapidly due to a continuing decline in the cost of solar modules. Solar is already the fastest growing segment of U.S. energy generation, with capacity increasing by 40.5% in 2017 alone and accounting for 2% of U.S. generation, and much more in some states, including California (15%), Hawaii (12%), Nevada (11%), and Vermont (12%) [135]. Worldwide, solar capacity increased 50% in 2018 alone [130]. The rapid rise is due to the falling costs of solar modules, which are decreasing much faster than forecasted, and do not appear to be abating any time soon [114]. Thus, many expect solar power to become the dominate source of electricity by the end of this century [114]. Policies in some states are accelerating this trend.

For example, California recently instituted a policy that *requires* new buildings to include rooftop solar [101].

The increasing amount of solar generation will profoundly change the grid's operation and the business model of utilities. In particular, while utilities operate and maintain the distribution network, they earn most of their revenue from generating electricity, which they currently can do much more efficiently than individual users. However, solar generation differs from fossil-fuel based generation in that it does not benefit as much from economies of scale. As a result, individual homeowners can install solar on their rooftops for closer to the same cost per watt that utilities can install large grid-scale solar farms. Even now, when amortized over a 25-year lifetime, solar power is cheaper than retail electricity rates in much of the U.S., assuming the solar power can be "net metered" and the utility credits users the retail rate for surplus power fed into the grid. Net metering enables consumers to connect solar power to the grid such that it acts as a negative load, causing their meter to run backwards when generating a net power surplus. When combined with government incentives, the payback period for net metered solar is now well under 10 years in many states.

Of course, the more individual users generate their own power, the less revenue utilities earn from generating electricity. Yet, utilities cannot simply decommission their generators (and take a capital loss), since they must still supply the grid's power at night when the sun is not shining. In addition, utilities may need to alter their mix of generators to handle increasing net metered solar installations, and their increased stochasticity, by employing more responsive but less efficient peaking generators. These changes may in-turn increase the cost and decrease the efficiency of grid generators. Thus, state governments and utility commissions typically place tight restrictions on users' ability to connect solar to the grid, as well as the compensation they receive for the energy it generates.

These restrictions vary widely by state. For example, in Massachusetts, the state places a hard cap on connected solar capacity under a state-sponsored incentive program. Upon

reaching the cap, the legislature must pass a new law to raise the cap, and update the incentives, e.g., generally by steadily decreasing them over time. The process of drafting the legislation is highly political and involves negotiations among multiple stakeholders, including legislators, utilities, solar installers, environmental groups, etc. The last round of negotiations after hitting the cap in 2016 took more than 9 months during which time new installations were prevented from connecting to the grid [115]. Other states have a similar process. For example, Hawaii prevented residents from connecting solar for 2 years due to similar negotiations [49, 95].

While Massachusetts and Hawaii offer incentives to install and connect solar, some states actually penalize consumers for connecting solar. For example, Alabama requires residential customers with solar to either pay a fee of $5/kW per month of installed solar capacity or pay 6× the standard electricity rate during peak summer hours [102]. In addition, unlike Massachusetts, Hawaii, and many other states, which credit consumers the retail rate for their surplus solar power, Alabama credits them 3-4× *less than* the retail rate. As a result of these solar disincentives, Alabama currently has only 48 residential solar customers statewide [102]. Many other states have similar policies that discourage solar adoption.

Clearly, the policies above directly influence solar energy's rate of growth. As solar penetration rises, utilities are likely to negotiate more strongly to reduce solar incentives, or even create disincentives, to preserve their revenue and business model, especially in states where utilities hold more political sway. Even in states, such as Massachusetts, that offer generous solar incentives, these incentives are decreasing over time as solar adoption increases, impacting utility revenue and operational costs. As these incentives decrease, we envision solar users considering partially or entirely defecting from the electric grid. Grid defection would enable users to install as much solar as they wish without limitations. Of course, the problem is that users can no longer rely on the grid to balance electricity's supply and demand, requiring them to store excess solar power and make up for deficits in

solar power using batteries. However, even if a solar-powered home consumes net zero energy over a year, much of the energy is generated during the summer months, which leaves large deficits on many winter days. Installing enough battery capacity to shift summer solar generation to make up for winter deficits would be prohibitively expensive, likely requiring a battery with greater than 1MWh capacity for the average U.S. home.

Instead, we envision the likely path for grid defection to be using a mix of solar power, batteries, and a whole-home natural gas generator. The cost of generating power from natural gas generators has rapidly decreased due to the steep drop in natural gas prices over the last decade. In contrast, utilities are locked into multi-decade investments in large coal plants that are less efficient and incur high fuel costs. Grid defection differs from prior work on off-grid buildings, which does not consider using backup generators, since it largely focuses on remote regions not connected to gas infrastructure. In exploring the feasibility, cost, and carbon emissions of defecting from the grid, we make the following contributions.

Power Generation Tradeoffs We analyze tradeoffs of using different forms of generation and storage, including grid power, natural gas generation, solar power, and batteries, in terms of their average costs, carbon emissions, lifetime, reliability, and operational constraints .

Grid Defection Architecture We present an architecture for grid defection. The architecture enables a home to dynamically switch between a local/generator and solar/battery depending on its power consumption and generation. We define switching policies that capture a tradeoff between power switching and wasted solar: more switching leads to less reliability, but maximizes the use of solar energy. We then compare the cost and carbon emissions of grid defection for representative home with using grid power based on current cost estimates based on policies from multiple states .

Implementation and Evaluation We implement a trace-driven simulator to evaluate the cost and carbon emissions of grid defection for a wide range of homes that differ in their

Generator	*CapEx ($)*	*OpEx ($/kWh)*	*Life (years)*	*CO_2*
Grid Power	0	0.10-0.32	∞	0.45
Natural Gas	7,656	0.187-0.374	$\gg$10	0.45-0.90
Solar	21,980	0	25	0
Battery	4,550	0	10	0

Table 4.1: Summary of current CapEx, OpEx, lifetimes, and carbon emissions (in kg-CO_2/kWh) of different generation options for grid defectors.

power consumption, solar generation, battery power and energy capacity, and generator capacity.

4.2 Background

To assess the feasibility of grid defection, we first need to understand the cost and carbon emissions of different forms of energy generation and storage, including grid power, local natural gas generation, solar power, and small-scale batteries.

4.2.1 Grid Power

The cost and carbon emissions of grid power are highly dependent on location, the local utility and its mix of generation sources, and the electricity rate structures, e.g., flat rate, time-of-use, peak demand charges, etc. The Energy Information Administration (EIA) estimates 62.7% of electricity generation comes from fossil fuels (coal, oil, and natural gas), 20.0% from nuclear, 17.1% from renewables, and 0.3% from other sources [13]. The total estimated generation from these sources was 4,015 billion kWh in 2017 with total associated CO_2 emissions being 1,821 million metric tons [11]. For this chapter, we translate these averages to our own carbon emission estimate of 0.45 kg-CO_2 per kWh for grid power. Similarly, average grid electricity cost for residential users varies widely by region from a high of $0.32/kWh in Hawaii to a low of $0.10/kWh in Washington, with an average of $0.13/kWh [9].

4.2.2 Whole-Home Natural Gas Generator

Estimating the cost of a whole-home natural gas generator is more challenging than grid power. Standby generators capable of powering a home are increasingly common to provide power during grid outages. These generators are widely available at local home improvement stores at low cost. For example, a 11kW generator with an automatic transfer switch (ATS) currently costs ∼$3100 [6]. These generators connect to a home's electrical panel via the ATS, which is able to sense a power outage, start the generator, and then automatically switch the home's power source from the grid to the generator. There is typically a 30 second delay between sensing an outage and switching to generator power, since the generator requires some time to start up after sensing an outage. The ATS also senses when grid power returns and automatically switches home power back to the grid, and then shuts down the generator. There is typically no loss of power when switching from the generator back to the grid. The generator connects directly to a home's natural gas pipeline, so there is no need for fueling the generator.

Unfortunately, standby generators are designed to only provide power for roughly 200 hours per year, or 3000 hours over their lifetime. In contrast, prime power generators are designed to provide reliable power continuously with an estimated lifetime for a natural gas microturbine being 50-80k hours, or 6-9 years continuous operation. Of course, since grid defectors will not operate generators continuously, they should last significantly longer. Currently, there are no home-scale ($<$20kW) prime power natural gas generators on the market to provide a cost estimate, which is currently an impediment to grid defection. As a result, we use the EIA estimate of $696 per kW of installed capacity for a prime power natural gas generator [10]. Thus, we estimate a 11kW prime power generator would cost $7,656 to install. We view this estimate as conservative, since EIA is a more expensive synchronous generator, which grid defectors would not require.

In addition to its capital cost, the generator also requires natural gas. The average price of natural gas in 2017 was $10.98 per thousand cubic feet, although this price varies

throughout the year [12]. Generator efficiency varies based on load, and ranges from 10%-40% efficient, with higher loads being more efficient. Unfortunately, most homes operate at low load levels $<$1kW most of the time relative to their peak load, resulting in lower efficiencies near 10-20% [4]. This yields an average cost of \$0.187-0.374/kWh of delivered electricity. Since this price is slightly above grid power prices, defection to a natural gas generator is not economically feasible. The carbon emissions of natural gas when burned as a fuel are 5.3 kg CO_2 per therm (or 29.3 kWh at 100% efficiency). At 10-20% efficiency, this translates to $0.45-0.90$ kg-CO_2/kWh. Thus, grid power is slightly cleaner and cheaper than a standalone natural gas generator due to operating its generators at higher load levels that are more cost- and carbon-efficient.

4.2.3 Solar Power and Batteries

Solar power costs also vary widely by region based on the amount of sunlight. The average Levelized Cost of Energy (LCOE) for solar in the U.S. for residential systems is estimated at \$0.129-0.167/kWh without any government subsidies [98], although the precise cost is a function of location and size. LCOE represents the net present value of the unit cost of electricity over a solar installation's lifetime, including the hardware cost of the modules and inverters, as well as the labor cost to install the system. Solar's LCOE is steadily declining with the Department of Energy's goal as part of the SunShot initiative to reach \$0.03/kWh by 2030 [23]. The LCOE amortizes solar's capital cost based on the energy it will generate over its lifetime, which is typically estimated at 25 years (based on manufacturer warranties). The capital cost of solar is currently estimated at \$3.14 per watt installed, which translates to \$31,400 for a system with 10kW rated generation capacity [3]. We reduce this by 30% based on the federal tax credit, resulting in a capital cost of \$21,980. Of course, the operational cost of solar is effectively zero as it requires no fuel. Solar generation has zero carbon emissions.

Unfortunately, LCOE assumes that all energy is used regardless of when it is generated. Of course, solar generation varies over each day and throughout the year. To fully utilize solar without using the grid requires battery-based energy storage. The Tesla Powerwall 2.0 costs $6500 with installation and a 10-year warranty, is designed for daily charging and draining in conjunction with solar, and has a capacity of 14kWh with a round-trip efficiency of 89% [8]. As above, the Powerwall benefits from the 30% federal tax credit, resulting in a capital cost of $4,550. However, the Powerwall has a power constraint of 5kW continuous power and 7kW peak, which is not large enough to concurrently run high-power appliances, such as an air conditioner, clothes dryer, and electric oven. In addition, solar plus 14kWh battery capacity is not nearly enough capacity to defect from the grid, even for a net zero home. Net zero homes at higher latitudes, as in the United States, generate much more energy during the summer than in the winter. Thus, homes must either install enough batteries to shift summer generation to the winter, e.g., 1MWh or 71 Powerwalls, or over-provision solar to generate enough power over the winter. Either case requires over-provisioning, causing excessive capital costs for solar or batteries.

Summary Table 4.1 summarizes our cost and carbon emissions estimates for grid power, local natural gas, solar, and battery. The natural gas estimate is based on the 11kW prime power natural gas generator described above, the solar estimate is based on a 10kW solar installation, and the battery estimate is based on the Tesla PowerWall 2.0 with 14kWh capacity.

4.3 Grid Defection Architecture

Figure 4.1 depicts our grid defection architecture, which includes two power sources: a battery charged by a rooftop solar array, and a whole-home natural gas generator. The power sources connect to a smart ATS that is able to programmatically switch the home's power between the sources. While switching the power source from a battery to the generator requires some delay to start-up the natural gas generator, we assume this delay is brief,

e.g., 1 minute, such that the battery can provide power over this time to prevent power losses at switch-over points. We assume the battery charges from solar, and not the natural gas generator, as charging the battery from the natural gas generator is inefficient. Functionally, our smart ATS is similar to those for backup standby generators already in use. The primary difference is that it requires a policy to determine when to switch between the two power sources.

The switching policy presents a tradeoff between minimizing the wasted solar power, and minimizing the switches between sources. We consider a few simple policies.

Minimizing Switches. Minimizing switches is important, since it decreases wear and tear on the generator, as frequent generator start-ups and shutdowns can reduce its lifetime and reliability. A switching policy that minimizes switches always waits until the battery is at full capacity before switching to the battery, and then waits to switch again until the battery is fully depleted. The problem with this policy is that it has the potential to waste solar energy. In particular, when the battery is at full capacity, and solar generation exceeds our power consumption, then there is no additional capacity to store surplus solar, and we must shed it. Solar charge controllers shed solar by increasing the applied voltage, which reduces the current and the resulting solar power generated to 0 [118]. As a result, waiting until the battery is full to switch increases the likelihood of wasting solar energy, especially in the summer when generation may significantly exceed consumption.

Minimizing Solar Waste. In contrast, we minimize solar waste by *always* switching to the battery whenever it stores any excess solar power. This policy consumes solar energy whenever it is generated, significantly increasing the number of switches as we frequently switch to and from the battery and its state-of-charge remains near empty. In particular, cloudy days may incur numerous switches over the day.

Balanced Policy. We also examine a balanced switching policy that takes advantage of the regularity of solar energy, and switches based, in part, on the net rate of generation/-consumption rather than the battery's state-of-charge. In particular, if there is any excess

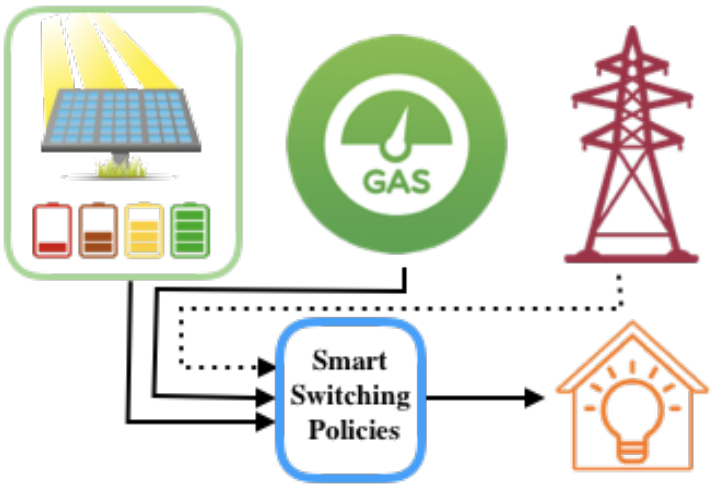


Figure 4.1: Grid Defection Architecture

power in the battery once the sun sets, we drain the battery completely before switching to the natural gas generator. In the morning, we switch back to using the battery after the sun rises (assuming we depleted the battery overnight) once the rate of solar generation exceeds our rate of consumption. Our intuition is that the rate of solar generation will increase over the day, even when cloudy, such that once the generation exceeds consumption at the start of the day, it is likely to continue to exceed it until the afternoon.

Our intuition above also exploits typical electricity usage patterns, which experience peaks in the morning and evening (due to to the use of high-power kitchen appliances), and a lull in usage during the middle of the day. Once switching to the battery, we only switch back to the generator once the battery's capacity is depleted. We switch back to the battery again once generation exceeds consumption, which generally does not occur until the next day. Of course, our system must also switch when exceeding the power limit of the power sources. For our battery, unless otherwise specified, we use a 7kW limit based on the Powerwall's specifications, such that we switch to the generator if power exceeds 7kW. Likewise, our generator has a 11kW limit, such that consuming greater than 11kW triggers an outage that deactivates appliances.

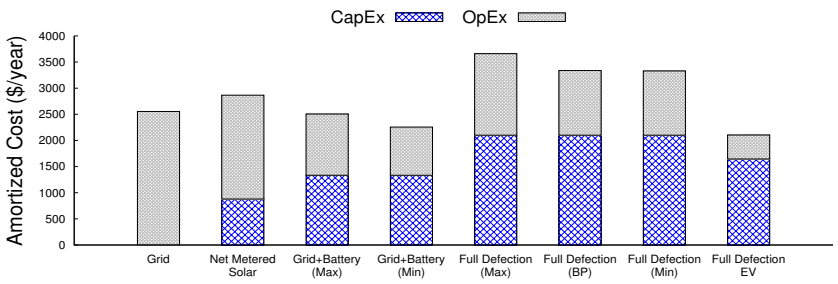


Figure 4.2: Operational and amortized capital cost for different scenarios in $/year.

4.4 Evaluation

4.4.1 Cost and Carbon Emissions Analysis

Figure 4.2 shows the results for the *operational and amortized capital cost* (over 25 years) on the y-axis, and the different power scenarios and switching policies on the x-axis. We view these results as conservative, as they assume current costs are static with no technological improvements. As shown, full grid defection using a residential battery, such as a Tesla Powerwall, is more expensive than using grid power, while using grid power with net metering offers the cheapest option. However, as discussed earlier, grid power with net metering is likely to be disincentivized over time. Some states already enforce such disincentives that make net metered solar much more expensive than the current results. Under such scenarios, users may be forced to defect from the grid to use local solar energy. The graph also shows the tradeoff between the different switching policies with the policy that minimizes solar waste (and maximizes switching) resulting in lower costs than the policy that minimizes switching. The balanced policy from has a similar cost as the policy that minimizes solar waste by maximizing switching.

The graph also shows an alternative where users leverage the battery in an EV rather than purchase a separate battery. In this case, we assume a battery capacity of 75kWh equivalent to the capacity of a Tesla Model 3 with extended range. If we exclude the cost of the EV from the system's capital costs, then this scenario already offers lower costs (by 17.6%) than using grid power. Thus, as EVs become more prevalent, the incentive for

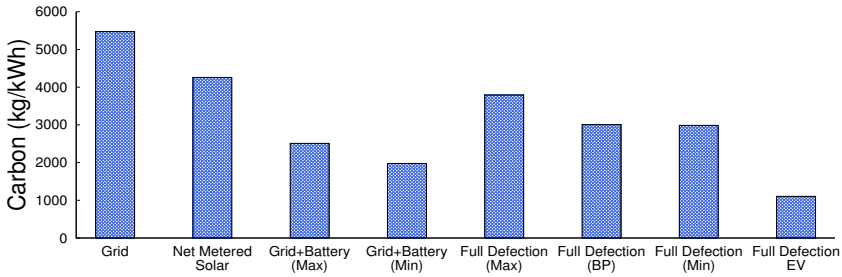


Figure 4.3: Carbon emissions from the same scenarios as in Figure 4.2 for cost.

grid defection increases, especially if solar net metering policies become less attractive. Finally, given that full grid defection is still more costly than using grid power, we analyze an approach that keeps solar disconnected from the grid by switching between the grid and a solar-powered home battery. This approach, labeled grid+battery above, essentially removes the natural gas generator and uses the grid as a backup source of power. We apply the same switching policies from the previous section, and find that this partial defection approach also results in lower costs than using grid power in both cases.

Figure 4.3 then shows the equivalent graph for average *carbon emissions* over the same period under the same scenarios. As before, the graph shows each scenario on the x-axis and carbon emissions on the y-axis. As shown, even though the natural gas generator incurs slightly more carbon emissions than grid power, our grid defection scenarios use it much less than grid power resulting in a significant decrease (by 45.4%) in carbon emissions. Note that carbon emissions when using net metered solar are only $\sim$20% lower than with grid-only because net metering still requires importing a significant fraction of grid power, especially in the winter. As before, using the grid as a backup power source by partially defecting with a solar-powered battery results in even lower carbon emissions, since the grid's carbon emissions are less than those of the natural gas generator. Finally, using the EV as a backup battery results in the lowest carbon emissions, since the size of the EV battery is much greater than the 14kWh Tesla Powerwall. Note that carbon emissions show a different trend than cost, with the grid having significantly more carbon emissions. Thus,

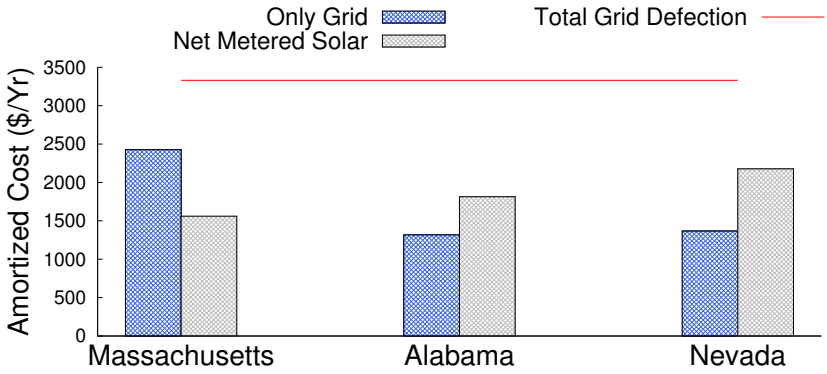


Figure 4.4: Price Comparison for Different States

if governments were to price carbon emissions, the financial incentive for grid defection would likely increase.

Figure 4.4 differentiates the incentive for grid defection in different states based on their solar (dis)incentive policy for that state. Since Alabama and Nevada's policies do not incentivize solar, the cost of net metering solar in Alabama and Nevada is more than in Massachusetts. Interestingly, a higher cost for net metering solar implies a stronger incentive to defect from the grid. That is, the cost of net metered solar in these states is closer to the red line that indicates the cost of grid defection.

4.4.2 Solar Waste Analysis

We compared the solar energy waste using the default 14kWh battery capacity for the different policies from, including minimizing switching (Min), minimizing solar waste (Max) and the balanced policy, assuming full grid defection. We found the following amount of solar waste for these three policies: Min (6483 kWh), Balanced (5327 kWh) and Max (5300 kWh). We see that the minimum switching results in the highest solar waste, since it always waits for the battery to be full before discharging. Thus, upon switching to the full battery, if the rate of generation exceeds the rate of consumption, the excess cannot be stored in the battery and will thus be wasted. In contrast, the maximum switching and balanced policies have close to the same solar waste, since these policies focus on maxi-

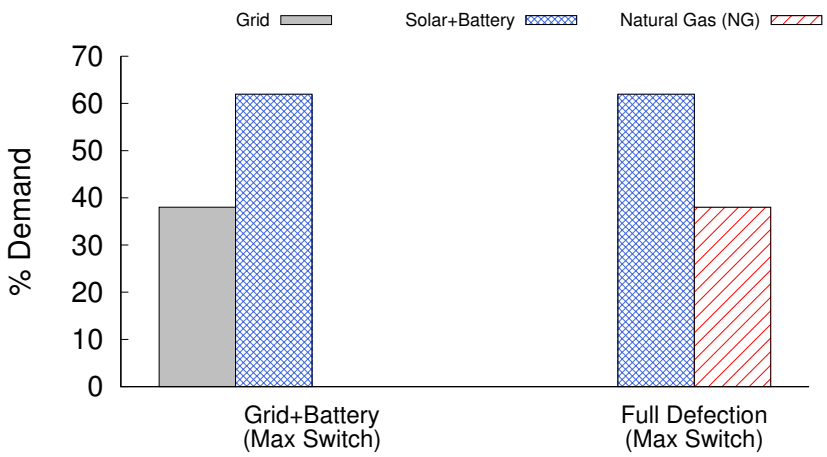


Figure 4.5: Percent Demand Met by Different Sources

mizing solar usage by always immediately switching to consume solar when it is available. In this case, all the policies waste a significant amount of solar, which demonstrates that a larger battery may be cost-effective.

Figure 4.5 then shows the percentage of demand met by the grid, battery, and natural gas generator under both partial grid defection (when using the grid as a backup source of power) and full grid defection (when using the natural gas generator as a backup source of power). In both cases, we focus on the maximum switching policy. The graph shows nearly 40% of the demand cannot be met by solar energy even though more than that amount of solar energy is being wasted – due to the asymmetry in solar generation between summer and winter.

4.4.3 Impact of Battery Size

Based on the results above, we also examined the impact of battery capacity on solar waste by increasing the capacity from 7kWh to 77kWh, assuming the same amount of solar generation and demand. Figure 4.6 compares the amortized cost and carbon emissions for these different battery capacities. As we increase the battery size, the carbon emissions decrease due to less use of the natural gas generator to satisfy demand. However, a battery

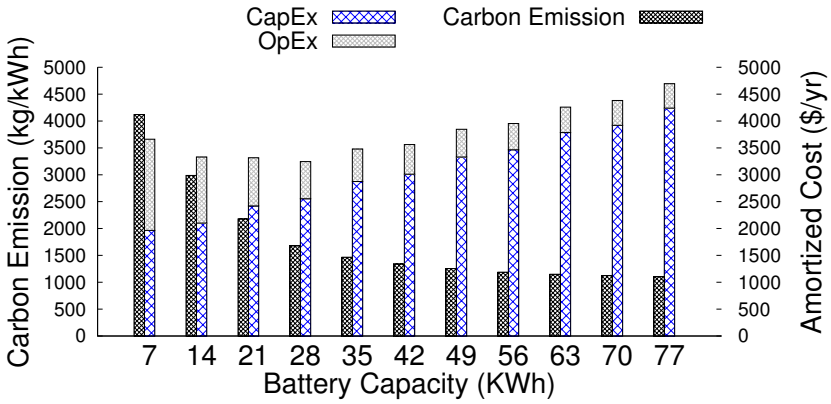


Figure 4.6: Amortized cost and carbon emissions as battery capacity increases.

capacity of 28kWh minimizes the amortized cost, where the increase in capital costs from the battery is more than offset by using it to reduce solar waste, which reduces the operating costs. The optimal amortized cost is closer to the grid power costs. In this case, full grid defection incentivizes the use 2 Powerwall batteries.

4.4.4 Impact of Aggregating Homes

We also considered the benefits of aggregating homes into small clusters, which are able to share solar energy and natural gas generation. These small clusters are akin to microgrids. To simulate this, we used data from multiple homes in different regions. For a set of homes, we create a cluster by aggregating the demand across the homes and aggregating their solar installations. We found that when considering multiple homes it is also important to consider the size of the solar installation and the capacity of natural gas installation, in addition to the battery size. We thus find the optimal feasible configuration of battery size, solar installation, and natural gas generator capacity to minimize the cost. The solar installation capacity is found by scaling (from $(0, 1]$) the combined available installation for the set of homes considered.

Figure 4.7 shows the effect of clustering 5 different homes in the states of Texas, Massachusetts and California, that have similar usage and solar generations. We observe that in comparison to a single home, a cluster of homes show a more competitive reduction

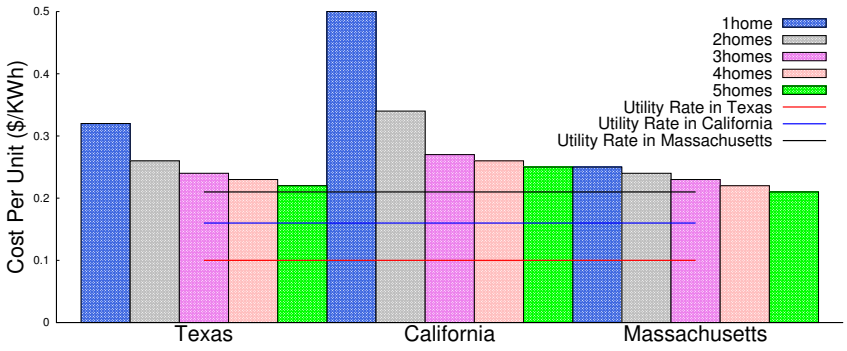


Figure 4.7: Cost Per Unit Power for Grid vs. Total Grid Defection with optimal configuration for aggregation of homes in different states of USA

in cost for full grid defection in comparison to using grid power. This is because we are not scaling the parameters linearly in proportion to the number of homes. For instance, the optimal parameters for a cluster of 5 homes in MA was found to be a 42kWh battery, solar scaling 0.6× combined solar installation of 5 homes, and 33 kWh capacity of natural gas. Thus, these 5 homes are able to multiplex the available energy more effectively than any single home and also benefit from the smoothing of demand and generation due to the aggregation.

Hence we can conclude that aggregating even a small number (5) of homes makes grid defection more feasible as compared to a single home. In particular the cost for grid defection for a cluster of 5 homes in MA is more favorable than the utility cost in MA.

Above, we ensured continuous power with no outages. However, if we permit some power outages, we can significantly reduce costs. To see how much, for a cluster of 15 homes, we calculate the amortized cost ($/kWh) for different percent of energy availability in Figure 4.8. We also show the flat rate utility price for three states, which have different grid costs. This shows that full grid defection could be cheaper than the grid cost if we allow for lower energy availability. We can see from the figure that at 75.75% energy availability, the $/kWh is lower than the utility rate in MA and CA, while at 95.60% energy availability the $/kWh is still lower than the utility rate in MA. The graph also shows that

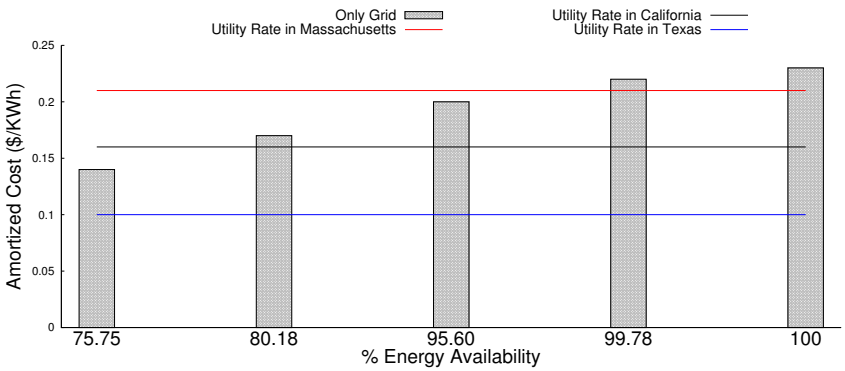


Figure 4.8: Amortized Cost vs % Energy Availability for a 15 home cluster

states with higher grid prices, such as MA, are closer to incentivizing grid defection than states, such as TX, with lower grid prices.

4.5 Related Work

Battery prices are dropping rapidly, which has made it already economical for many commercial customers to reduce their peak consumption levels. Grid defection is beginning to make economic sense both in industrial and residential sectors due to these drops in battery costs and studies have predicted that grid defection may become a viable option in only a few years [62]. Recent work [70, 121], similar to ours, compares a grid-tied residential solar system with an off-grid solar-plus-battery system at locations in the United States, and estimates the costs and carbon emissions. However, this chapter does not consider the use of a natural gas generator. Prior work also examines optimizing energy storage capacity and load scheduling to improve reliability in islanded operation in residential sectors [67], which we leverage in our analysis. Of course, our work does not consider the impact of grid defection on grid power costs. Prior work studies the implications of widespread disconnection from the grid using only solar [82] and examines policies to grid operators develop other sources of revenue rather than increasing energy prices.

4.6 Conclusions

The declining cost of solar generation is leading to an increase in grid solar capacity. However, this is also leading to utilities restricting access to connect solar to the grid, and reducing the compensation for it. As a result, in the future, the most viable way to use solar energy may be to defect from the grid. Thus, in this chapter, we proposed an approach for total grid deflection for residential homes using a combination of solar with/without battery, natural gas and electric vehicles. We presented different policies for smart switching between these power sources with tradeoffs in terms of number of switches, solar waste, reliability, carbon emissions and total cost. We analyzed these tradeoffs using a trace driven simulator for a single home as well as a cluster of homes. For these scenarios we considered homes from Massachusetts, California and Texas and compared the feasibility of grid versus total grid defection. Our analysis indicates that, based on the net metering policies in different states, complete grid-defection is financially attractive for a cluster of homes—even currently—in some states, and in all cases yields less carbon emissions. In the next chapters, we discuss solar modeling using various ground-based and satellite-based approaches that can enable better solar control.

CHAPTER 5

MODELING SOLAR PERFORMANCE MODELS USING DERIVED PRODUCTS FROM SATELLITE

In order to have better control over solar, we also need accurate solar performance models that infer solar power output in real time based on the current environmental conditions. These are an important prerequisite for many advanced energy analytics. In this chapter, we develop and evaluate solar performance models that use satellite-based estimates of downward shortwave (solar) radiation (DSR) at the Earth's surface, which NOAA began publicly releasing after the launch of the GOES-R geostationary satellites in 2017.

5.1 Introduction

Solar energy generation has grown at nearly an exponential rate over that past 30 years, and is now cheaper than the retail price of electricity in many locations [113]. The goal for the U.S. Department of Energy's SunShot initiative is for solar to satisfy 14% of U.S. electricity demand by 2030 and 27% by 2050 [23], or a factor of 10× and 20×, respectively, greater than the 1.4% it satisfied in 2018 [24]. Reaching these targets will require improving solar performance models, which infer solar power output in real time based on current environmental conditions. These models are a prerequisite for a wide range of energy analytics, including solar forecasting, energy disaggregation, and grid simulations, that are necessary for grid operations and planning to accommodate higher solar penetrations.

To address the problem, recent work develops sophisticated data-driven modeling techniques that automatically derive a solar performance model for small-scale sites from public weather data, and thus are more scalable than prior manual approaches [44, 52, 53, 85].

Once built, the model estimates a site's solar output at any time given the current weather conditions. Such data-driven models are highly accessible and useful for modeling any solar site in the U.S., as they rely only on well-known physical models of solar generation and public weather station data that is released in real-time for every location in the U.S. by the National Weather Service (NWS).

Unfortunately, using public weather station data has two primary drawbacks: not all solar sites are near a public weather station, and public weather station data generally quantifies cloud cover—the most significant metric that affects solar—using highly coarse and imprecise measurements [14, 15]. This measurement is in *oktas* and is often taken using a circular sky mirror placed on the ground that divides the sky into eight equal slices, such that the number of slices that contain a cloud translates to the number of oktas. The NWS then quantifies cloud cover using textual descriptions that map to a specific range of oktas. For example, "scattered clouds" maps to 3-5 oktas [14]. The imprecision of cloud cover measurements is by far the largest source of inaccuracy in large-scale data-driven solar performance modeling. Of course, while more accurate cloud cover measurements are possible using a pyranometer, which directly measures solar irradiance at the Earth's surface, only a few public weather stations include pyranometers.

Recently, National Oceanic and Atmospheric Agency (NOAA) in the U.S. has begun, as of 2018, releasing data products derived from a new generation of remote sensing geostationary satellites—the GOES-R series [18]. One of the secondary data products is the Downward Shortwave Radiation (DSR) that is incident at the Earth's surface, which estimates both the direct and diffuse solar radiation. Thus, DSR estimates the solar radiation available at the surface to generate solar power [16]. DSR is derived from the raw satellite data using a state-of-the-art algorithm that analyses the reflectance measurements of GOES-R's Advanced Baseline Imager (ABI) [5]. These DSR estimates account for cloud albedo, or the solar radiation reflected by clouds, and atmospheric conditions, and are available for any 0.5-2km^2 area within the satellite's view.

In contrast, the distance between a solar site and the nearest public weather station varies widely, and can be up to dozens of kilometers. In addition, unlike coarse okta measurements, DSR is a fine-grained measurement. Thus, using satellite data for solar performance modeling has the potential to address the drawbacks of public weather station data. However, satellite data also has drawbacks. While public weather station measurements are taken at the surface and represent ground truth, satellite measurements are taken from geostationary orbit, which is $35,800$km above the Earth's surface. Satellites also can only measure the solar radiation reflected by the top of clouds, but cannot accurately assess cloud depth, height, or temperature, all of which affect the radiation that reaches the Earth's surface. As a result, unlike public weather station data, satellite data does not represent ground truth. Thus, while oktas provide coarse but direct measurements of surface radiation, satellites provide fine-grained but indirect measurements.

In this chapter, we develop and evaluate a solar performance model that uses DSR, and compare it to a similar modeling framework that uses oktas. In doing so, we show how to integrate satellite data into an existing data-driven solar performance model from prior work [44], and examine multiple model variants that i) incorporate satellite data in lieu of public weather station data, and ii) use a combination of both. We will make our satellite-based modeling framework publicly available along with a solar and DSR dataset from nearly 50 sites that we have curated as part of our evaluation. To the best of our knowledge, this is the first use and evaluation of DSR for solar performance modeling, in part, because NOAA only began making this data product available in 2018.

This chapter identifies strengths and weaknesses in using DSR satellite data for solar performance modeling. In particular, and contrary to our intuition, we find that using satellite-based DSR measurements *does not* improve the accuracy of solar performance models compared to using public weather station data. While DSR estimates provide slightly better accuracy during mostly clear skies, the estimates are much worse under overcast conditions. In most cases, DSR measurements are not even available during over-

cast periods due to these known limitations in accuracy under these conditions [5]. Thus, despite DSR's promise in other areas, especially long-term climate modeling, using public weather station data for solar performance modeling yields similar accuracy and is much more accessible. We do show that a hybrid approach that strategically uses satellite DSR data during mostly clear skies can modestly increase accuracy. In performing our data analysis, we make the following contributions.

Satellite Data Background. We present background on the GOES-R series of satellites and the DSR data product, including its availability, accessibility, and ground truth accuracy. We also curate a new dataset that consists of hourly readings of solar generation, cloud cover in oktas, and DSR estimates for each of the 47 solar sites we analyze in our evaluation.

Exploiting DSR for Solar Performance Modeling. We show how to modify an existing data-driven solar performance model that uses cloud cover measurements from public weather stations to instead use satellite-based DSR measurements. We then illustrate salient differences between okta- and satellite-based measurements for a representative solar site. We define multiple model variants that combine satellite and okta data in different ways to understand their strengths and weaknesses.

Implementation and Evaluation. We implement the solar performance models above and evaluate them across the 47 solar sites in our dataset. Our evaluation shows that a physical model that uses okta-based measurements yields similar accuracy as using satellite DSR data, and that a hybrid approach can offer a modest improvement in accuracy.

5.2 Background

We provide background on measuring the impact of clouds using DSR and oktas, as well as on data-driven solar modeling.

5.2.1 Satellite-based DSR

There has been significant prior work on inferring solar irradiance incident at the Earth's surface using satellites. Much of this work, including the Heliosat family of algorithms [47, 48,108], infers solar irradiance from visible satellite imagery, assuming a pixel's intensity is related to cloud cover. In contrast, GOES-R satellites include an Advanced Baseline Imager (ABI) that takes images of the Earth across 16 different spectral bands, which include two visible channels, four near-infrared channels, and ten infrared channels [20]. These 16 bands compare to only 5 bands from the previous generation of weather satellites and offer $4\times$ greater spatial and $5\times$ greater temporal resolution [19]. Specifically, the spatial resolution of contiguous U.S. (CONUS) is 0.5-2km^2 and the temporal resolution is every five minutes.

NOAA publicly releases the raw spectral data in near real time, as well as a large number of higher-level data products derived from this raw data. The raw data only began being released in 2018 (for GOES-16) and 2019 (for GOES-17) with higher-level data products being released later. This chapter focuses specifically on a Level 2b+ data product that estimates the *downward shortwave radiation* (DSR) at the Earth's surface [16], which includes the ground-level direct and diffuse solar radiation in the visible, infrared, and near-infrared spectrums. Solar cells convert some fraction of DSR to electrical power based on their physical characteristics, e.g., power conversion efficiency, temperature coefficient, tilt, orientation, etc. DSR derives from a sophisticated physical model built on lower-level data products, e.g., for cloud optical depth, particle size, height, etc., that estimates cloud albedo and the atmosphere's composition, and represents the state-of-the-art in estimating radiation at the Earth's surface. That said, the DSR documentation quantifies its accuracy, which can vary widely depending on many factors, including the cloud characteristics, solar zenith angle, and latitude [16].

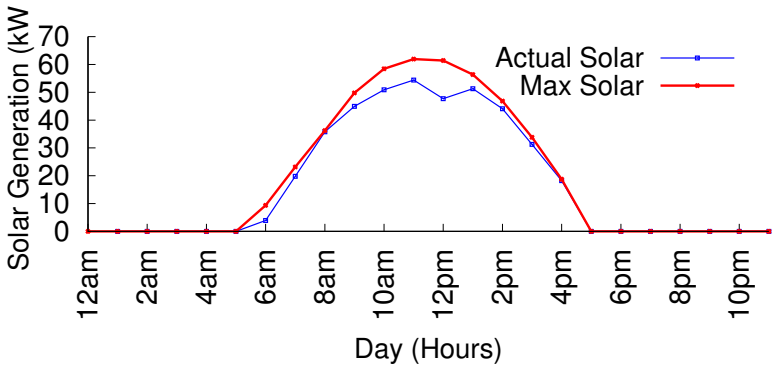


Figure 5.1: Depiction of bounding solar generation using Equation 5.2.

5.2.2 Ground-level Cloud Cover Measurements

Prior work on data-driven solar performance modeling combines clear sky solar irradiance models with ground-level cloud cover measurements in oktas, which are publicly available, to infer surface irradiance. Public weather stations typically report cloud cover as one of five weather strings, including clear skies (CLR), few clouds (FEW), scattered clouds (SCT), broken clouds (BKN), and overcast skies (OVC). These strings map directly to specific ranges of okta values [14]. Specifically, CLR maps to 0-1 oktas, FEW maps to 1-3 oktas, SCT maps to 3-5 oktas, BKN maps to 5-7 oktas, and OVC maps to 7-8 oktas. Prior work captures the relationship between cloud cover measured in oktas, and the clear sky index (CSI), which is the ratio between the actual irradiance at the surface divided by the irradiance at the surface under clear skies. For example, in prior work, Kasten and Czeplak derived the empirical model below, which is widely used in textbooks [80].

$$CSI = 1 - 0.75 \times n^{3.4} \tag{5.1}$$

Here, n represents the fraction of cloud cover, e.g., by taking the midpoint of the okta range and dividing by 8. Chen et al. recently refined this empirical model using a much larger dataset [52]. Note that clear sky solar irradiance is a deterministic function of location, i.e., latitude and longitude, and time, and can thus be accurately estimated without any external inputs [92]. There are many software libraries, such as pysolar [1] and pvlib [38],

that compute the clear sky solar irradiance given a location and timestamp. Thus, we can infer ground-level solar irradiance simply by multiplying the clear sky solar irradiance by the CSI from Equation 5.1 above, which is based on the cloud cover reported by public weather stations.

5.2.3 Modeling SURFRAD Data

While we talked about DSR and Okta and their role in modeling solar irradiance, there is another very important ground irradiance measurement vis Surface Radiation Budget Network or SURFRAD. SURFRAD was etablished in the year 1993 through NOAA with a mission is to support climate research with long-term measurements of surface radiation budget over the United States. Currently there are eight SURFRAD stations across United States operating over climatologically diverse regions in the United States. Data from these SURFRAD stations are distributed in near real time by anonymous FTP and the World Wide Web. The frequency of data availability from these SURFRAD stations is 1 minute average and it contains a large number of variables including but not limited to downwelling/upwelling global solar, direct-normal solar and net solar. These SURFRAD stations although record very accurate measurements, cannot be used to create general solar models because of being localized and site-specific.

5.2.4 Data-driven Solar Performance Modeling

This chapter builds on a simple data-driven solar performance modeling approach from prior work to quantify the accuracy of using satellite DSR estimates to infer solar generation [44, 52, 53]. We briefly summarize this approach, which we show how to modify in to incorporate DSR estimates. As input, the approach only requires a site's location and some historical generation data. The approach leverages the fact that a solar site's generation is always bounded by its maximum generation $P_{max}(t)$ described by the physical model below.

$$P_{max}(t) = I_{clearsky}(t) \times k \times (1 + c \times |T_{baseline} - T_{air}(t)|) \times$$
$$[cos(90 - \Theta) \times sin(\beta) \times cos(\Phi - \alpha) + sin(90 - \Theta) \times cos(\beta)] \quad (5.2)$$

Here, $I_{clearsky}(t)$ is the clear sky solar irradiance at time t, and k is the solar site's efficiency parameter, which is a product of its size and solar conversion efficiency. Since solar conversion efficiency is a function cell temperature, the model multiplies k by an additional term. Here, c is the solar modules' temperature coefficient, while $T_{baseline}$ represents the baseline temperature when the conversion efficiency is k. Solar efficiency varies linearly with temperature, so the model multiplies the absolute value of the difference between the current temperature $T_{air}(t)$ and the baseline by the temperature coefficient. Typical values of c are $\sim 0.5\%$, such that efficiency increases this amount for every 1C drop in temperature. Finally, the lower term captures the impact of solar geometry: Θ and α represent the Sun's zenith and azimuth angles, respectively, while β and Φ represent the solar modules' tilt and orientation angles, respectively. These solar angles are a function of location and time, and can be computed using a library.

Prior work describes an efficient method for searching for values of k, c, $T_{baseline}$, β, and Φ in the equation above that yield the closest upper bound on the historical generation data [44]. The insight is that under clear skies, solar generation should conform to the model above (for some constant values of k, c, $T_{baseline}$, β, and Φ), while under cloudy skies, solar generation should be strictly less than the model above. Figure 5.1 depicts an example of bounding generation data using Equation 5.2.

After bounding the equation above to the data, we can compute $P_{max}(t)$ at any time t. The model then leverages the relationship below, which follows directly from Equation 5.2.

$$\frac{P_{actual}(t)}{P_{max}(t)} \sim \frac{I_{actual}(t)}{I_{clearsky}(t)} = CSI \quad (5.3)$$

To understand the relationship, observe that the only change in Equation 5.2 when computing actual solar output under cloudy skies is that we must replace $I_{clearsky}(t)$ with the actual solar irradiance under cloudy skies (assuming no change in temperature). All other parameters are independent of the cloudiness. As a result, when dividing $P_{actual}(t)$ by $P_{max}(t)$, everything on the right side of Equation 5.2 cancels out, which simply leaves $I_{actual}(t)/I_{clearsky}(t)$ or the clear sky index CSI.

Thus, given a model of $P_{max}(t)$ and the CSI, we can infer $P_{actual}(t)$ by simply multiplying the CSI by $P_{max}(t)$.

5.3 Satellite-Based Solar Performance Modeling

We show how to use the data-driven solar modeling framework from the previous section to leverage both oktas and DSR, as well as a hybrid approach that uses both. Importantly, we use the same approach described earlier for each of these solar performance models, but where the clear sky index (CSI) is computed from different sources. In particular, satellite-based modeling computes the CSI using DSR data, while the okta-based approach computes it from the improved Kasten-Czeplank model [52]. As a result, any differences in modeling accuracy are only due to changing this input. In addition, since these models derive directly from the physical relationships inection, they do not take into account any site-specific characteristics. Thus, for comparison, we also develop solar performance models using machine learning (ML) that can learn site-specific characteristics.

5.3.1 Satellite-Based Models

Our satellite-based model is simple: to derive CSI, we take DSR directly from NOAA and divide it by a solar site's clear sky irradiance based on its location and time using a clear sky model. Note that this is a purely physical model that does not perform any regression to learn the relationship between DSR and solar output. Figure 5.2 shows the clear sky irradiance and DSR at a particular site. The graph shows DSR in watts per meter

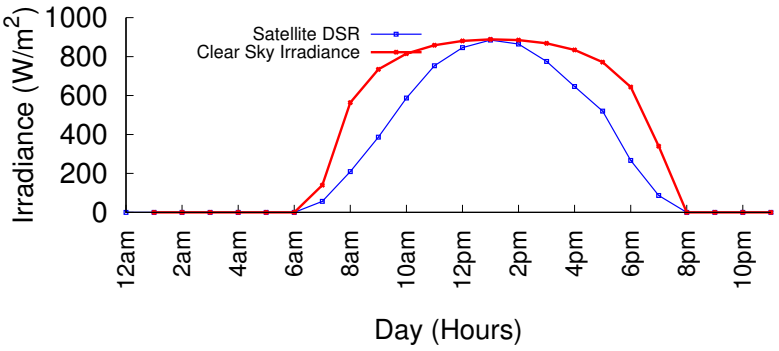


Figure 5.2: Relationship between DSR and clear sky irradiance

squared (W/m^2) and the corresponding clear sky irradiance over a representative clear day. The graph demonstrates that the clear sky irradiance is a strict upper bound on the DSR, such that the values are close when the sky is clear as expected. Interestingly, the values are nearly equal at solar noon, while the clear sky irradiance is slightly greater than DSR before and after solar noon.

Figure 5.3(left) shows the relationship between normalized DSR, or $DSR/I_{clearsky}$, and normalized solar generation, or P_{actual}/P_{max}, across many locations. The graph shows the normalized DSR on the x-axis and the normalized solar generation on the y-axis. As the graph shows, the the relationship is roughly linear, albeit noisy. This noise is largely due to inaccuracy in the DSR measurement, but may also result from unaccounted variables in our model, such as the presence of shading and topography at a solar site. We evaluate this relationship more fully for DSR in evaluation. A benefit of this approach, as discussed earlier, is that DSR is available every 0.5-2km^2, and thus provides more precise measurements than weather stations.

5.3.2 Oktas-Based Models

In our oktas-based model, we compute the CSI using the ground-level cloud cover measurements provided by public weather stations. In this case, we use the mapping of each weather string—CLR, FEW, SCT, BKN, and OVC—to an oktas range. Since the

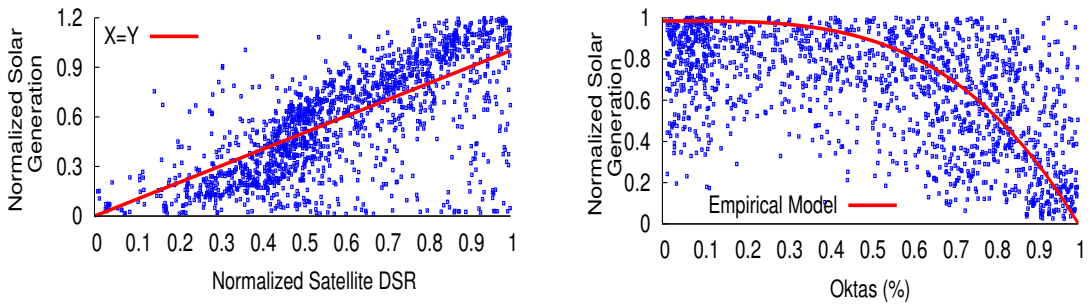


Figure 5.3: Scatter plot of normalized solar generation versus normalized satellite DSR (left) and okta-based measurements (right) across many solar sites.

NWS only specifies a coarse range for these values [14], we simply use the average of each range and map it to a number when computing the CSI. As explained above, we use this okta value for CSI in our data-driven solar model to infer the solar output. Again, this is a pure physical model that does not require learning a model from generation and weather data that is specific to a site. Figure 5.3(right) shows the relationship between oktas and the actual CSI for a particular location. Here, the x-axis is the ground-level cloud cover measurements (okta) and the y-axis is the actual CSI derived from the solar data (as discussed in background). The graph shows that the okta-based measurements, while also noisy, do roughly follow the expected trend of the empirical models defined in prior work [52, 80]. In this case, the increased noise is largely due to the coarseness and imprecision of okta-based cloud cover measurements, which are derived from weather stations that are an unknown distance from each site.

5.3.3 Hybrid Model

As we show in evaluation, the satellite-based DSR model tends to be much less accurate than the oktas-based model when the cloud conditions are broken (BKN) or overcast (OVC), and slightly more accurate otherwise. The DSR documentation explicitly states this limitation of DSR, and, as a result, often does not even provide DSR readings when skies are cloudy [5]. To address this problem, we also design a hybrid solar performance model

that leverages both oktas and DSR. This model uses the ground-level public weather station data as a filter by using the observed cloud cover to decide whether to use the satellite-based model or the oktas-based model. When the ground-level observation is CLR, FEW, or SCT, we use the satellite-based model (as described above), while we use the oktas-based model when the ground-level observation is BKN or OVC. This approach combines the best attributes of both models.

The satellite-based, oktas-based, and hybrid models above all use physical models that are general and applicable to all solar sites. We also develop machine learning (ML) models that are specific to the characteristics of each site. ML models naturally capture unmodeled variables that are unique to each site, such as shading, which our physical models above cannot capture. However, one drawback of ML models is that they require sufficient data for training. Since some characteristics, such as shading, change throughout the year due to the seasons, this may require multiple years of data for solar models.

We train our ML models based on historical energy generation from each solar site. As with the hybrid model, these models combine DSR and oktas as input variables, as well as the clear sky irradiance and time-of-day/year. The dependent output variable is the site's actual solar generation under these conditions. We train our ML models using data from all of 2018, and use the data in 2019 for testing. Since DSR from the GOES satellites only recently became available, this is the maximum amount of training and test data that is available. In addition, since we train our ML models on each site individually, they implicitly incorporate site-specific physical characteristics that affect solar generation, which the physical models above do not, including the site-specific impact of non-ideal solar geometry (i.e., different panel tilts and orientations) and shading. The ML models are purely a black box and do not incorporate any of the physical models above in their training.

We evaluate two different common ML models: decision tree and support vector machines (SVMs). Decision trees are a flow chart-like structure where each internal node rep-

resents a test on a feature for classification and each leaf node represents a class label, while the branches represent features responsible for the class labels. In our decision tree, we used 10-fold cross-validation to select the tree depth from a maximum depth of 20 to avoid over-fitting. We also compare with SVMs, which attempt to fit as many datapoints with the kernel function while limiting margin violations. Under SVM with regression, we define a margin of tolerance (ϵ), a regularization co-efficient C, and use the radial basis function (RBF) as the kernel. The tolerance ϵ and co-efficient C are estimated using 10-fold cross-validation in the following range: $\epsilon \in \{0.01, 0.05, 0.1, 0.2\}$ and $C \in \{1, 10, 100, 1000\}$. For both ML models, we also add the hour of the day as an additional feature.

5.4 Implementation

We implemented the solar performance models as a python module, which we have publicly released We use python's *scikit-learn* ML library to build the ML models in the previous section. We also used the numpy and pandas python packages to decode the NetCDF-formatted satellite data described below. Our module only requires a site's latitude and longitude as input, which it uses to compute clear sky irradiance using pysolar [1], a python library for simulating the solar irradiance at any point on Earth at any time. Similarly, our module programmatically fetches current and historical hourly temperature and cloud cover data from Weather Underground, a commonly used online weather website that maintains historical weather archives. Given a location, Weather Underground automatically determines the nearest weather station to that location. We also have access to two years of solar generation data from 47 homes. While we do not have physical access to all 47 homes, we can visibly observe many of their physical characteristics, e.g., size, shading, tilt, orientation, etc., in satellite imagery.

We use a web service provided by NOAA to access the satellite DSR data. Currently, users must download NetCDF-formatted files from an FTP server or via Amazon S3 buckets, as NOAA does not offer access to it via a web service with a programmatic interface.

NetCDF is a common machine-independent data format for array-oriented scientific data. Users submit requests for data products, such as the ABI L2+ DSR product for the GOES-16 satellite, via NOAA's Archive Information Request System (AIRS) for up to 30 days. Once approved, NOAA sends the user a link via email to download the requested files (typically within an hour). Each file includes data for the entire contiguous U.S. for a single hour. As a result, our python module must decode the NetCDF data, and extract the DSR value for the sites of interest based on their latitude and longitude. Since extracting the DSR value for a site from the NetCDF file is non-trivial, we describe the process below.

Extracting Satellite Data. To extract a site's DSR, we must project the data file onto a geographic map. There is a summary option in each NetCDF file that gives all the variables available in the file. Specifically, the variable *goes_imager_projection* is essential for converting (x, y) coordinates for latitude and longitude in degrees to radians. Our python module uses this variable to extract the satellite sweep, longitude, and satellite height. The projected x and y coordinates equal the product of the scanning angle (in radians) and the satellite height.

Following this projection, we can extract the latitude-longitude pairs in the form of a matrix from the NetCDF file. We calculate the nearest pair of coordinates from this matrix with our specified location using the Vincenty formula [131], which calculates the distance between two points on the surface of a spheroid. For the nearest computed location, we then extract the corresponding DSR value for the latitude-longitude pair.

As with the weather data, the satellite DSR is released hourly. Thus, we focus on solar performance modeling at an hourly resolution. Our python module combines the hourly temperature, cloud cover, satellite DSR, and solar generation for each location into a tabular format, e.g., a CSV file, with a corresponding timestamp for each reading. These data sources are stored in many different formats, particularly with different timestamps and time zones. As a result, our python module normalizes all timestamps and time zones to UTC time. Since our models currently do not account for snow, we focus on periods

with no snow: May to October in 2018 and 2019. Incorporating snow is future work. Our primary metric is the Mean Absolute Percentage Error (MAPE) between our models and the ground truth, where a lower MAPE indicates less error.

$$MAPE = \frac{1}{n}\sum_{t=0}^{n}|\frac{S_t - P_t}{S_t}|$$

Here S_t and P_t are the ground truth and model-inferred solar generation, respectively, at hour t, and n is total number of hourly data points. We use MAPE because it is an intuitive metric that is comparable across solar sites of different sizes. However, note that MAPE is highly sensitive to periods of low absolute solar generation. For example, if solar generation for a 10kW site is only 10W early in the morning, and our model infers 40W, we record a 400% error, even though the 30W error is only 0.3% of the site's capacity. Thus, when evaluating any single solar site, an absolute error metric, such as the Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE) may be more appropriate. However, since our primary focus is comparing across sites with different sizes and characteristics, we continue to use MAPE, and mitigate its drawbacks by focusing on the 10am-3pm time period to eliminate periods that always have low absolute generation. Our primary focus is on the *relative* difference between the MAPEs of models in design and not the absolute value.

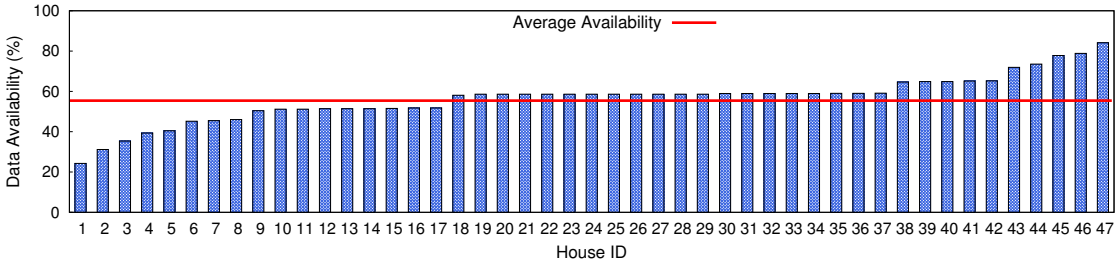


Figure 5.4: Availability of DSR data product across our 47 solar sites.

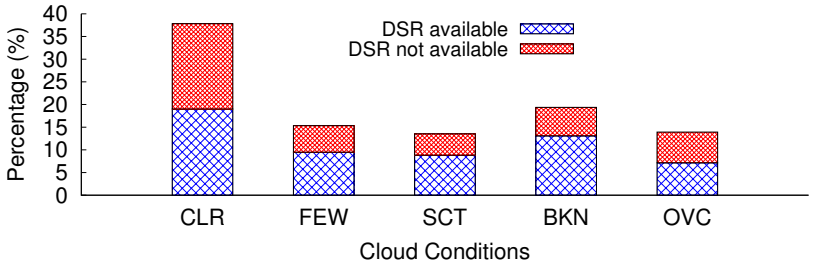


Figure 5.5: DSR availability under different cloud conditions.

5.5 Evaluation

We evaluate our solar performance model using DSR on 47 solar sites. Unfortunately, however, DSR is unavailable during periods when its physical model is too uncertain [5]. On average, across our 47 sites, DSR is only available 55.4% of the time, although this differs across sites. Figure 5.4 shows the data availability across all 47 sites with a horizontal line at the 55.4% average. Figure 5.5 shows DSR's unavailability under different cloud conditions, and shows that this unavailability is higher during clear and overcast skies. This unavailability is currently a drawback to using DSR, especially during overcast skies as modeling solar performance is most important during these periods. Given this lack of availability, we restrict our analysis below to only those periods where DSR is available.

SURFRAD Stations. SURFRAD is a more accurate measurement of ground DSR, but is scarcely available at only few locations (8 location in the U.S.), we picked solar sites close to one of the SURFRAD stations which is in Colorado, Boulder. We analyzed MAPE values for solar generation inference from using SURFRAD values for the sites closer to the Colorado SURFRAD station. As shown in Figure 5.6, on the x-axis we have solar sites in the order of increasing distance from the SURFRAD station. We observe that as the distance of the solar site from the SURFRAD station increases, the SURFRAD result become less and less accurate. While the solar site location closest to the SURFRAD site has highly accurate and better results, even when compared to Oktas. We can see that site 1, which is closest to the SURFRAD site, has the lowest MAPE for using SURFRAD directly

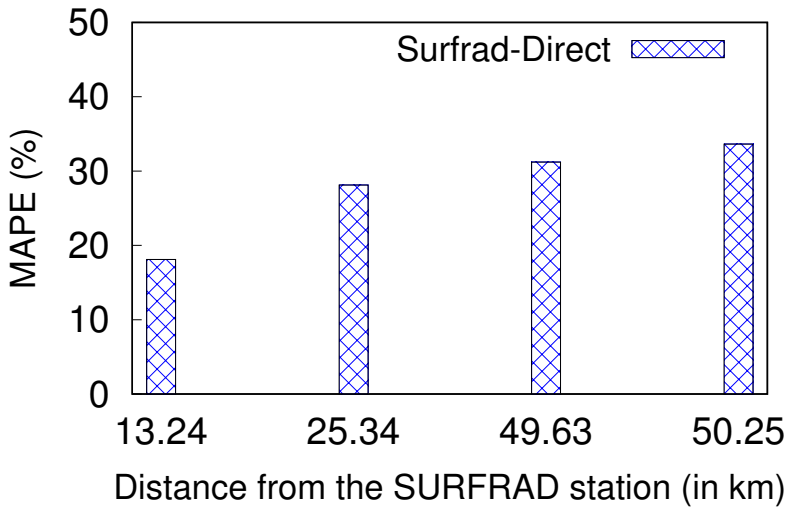


Figure 5.6: MAPE result with SURFRAD Data.

while as we move on to site 3 and site 4 which are further away from the SURFRAD site, the MAPE for SURFRAD increases. Having information from the SURFRAD site is as good as having an irradiance sensor next to the solar site, if the site is close to the SURFRAD station. But since SURFRAD data is only limited to 8 sites across whole United States, and it's accuracy decreases with distance from the site, it is not useful across all geographical locations.

Physical Models. We first analyze the MAPE for our satellite-based, okta-based, and hybrid models from sec:design. Since these are physical models and do not require training, we can use the entire two-year dataset to evaluate their accuracy across all 47 sites. Figure 5.9 shows the overall results, as well as the MAPE under different cloud conditions. We find that, overall, the hybrid approach slightly outperforms the okta-based approach, and, surprisingly, the DSR approach performs the worst. As shown, the inaccuracy of the satellite-based DSR approach is due to its low accuracy during overcast conditions.

To emphasize the point, Figure 5.7 shows the MAPE under overcast cloud conditions for all 47 sites, and demonstrates that this performance for DSR is consistent across almost all of the sites with some sites reporting MAPEs in excess of 100% using DSR. However,

as shown in Figure 5.5, since there are few overcast time periods where DSR is available, this inaccuracy does not contribute a significant amount to the overall results. Under all other cloud conditions, we observe a similar accuracy across the three techniques. Since our hybrid approach uses DSR when skies are not overcast and the okta-based approach otherwise, it slightly outperforms the pure okta-based approach. While our focus is on the relative difference between the models, the absolute MAPEs we find are similar to the okta-based models evaluated in prior work [52].

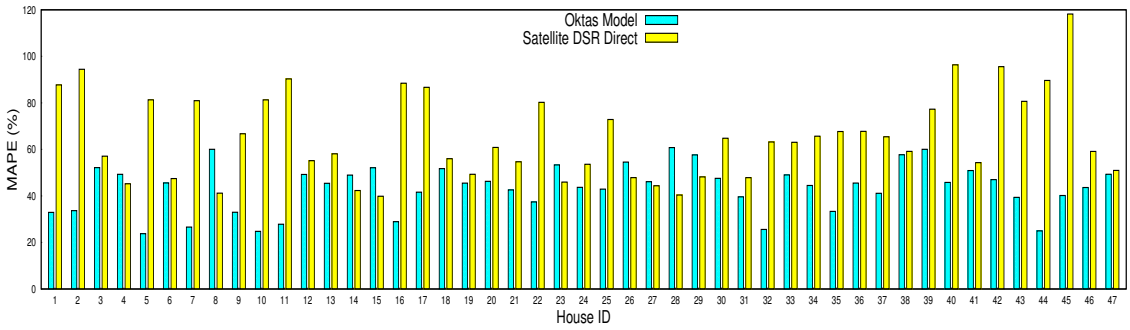


Figure 5.7: MAPE for satellite-based and okta-based models under overcast cloud conditions.

Machine Learning Models. For our ML models, we use 2018's data for training and 2019 for testing our decision tree (DT) and support vector machine (SVM) regression models. Figure 5.8 shows the overall MAPE for both our physical and ML models in 2019 under all cloud conditions, only overcast conditions, and all cloud conditions except overcast. We separate out overcast conditions since they are the most challenging conditions to model. We see that the ML models do not significantly improve upon our hybrid physical model, which does not require training. Overall, the hybrid model performs the best in all three cases, and is slightly better than the oktas-based model. The DT and SVM models actually perform worse than the satellite-based DSR model in overcast conditions. This poor performance may be due to the lack of training data in our dataset, as prior work uses multiple years of training data. Since DSR has only been available for two years, there is limited data available for training our models.

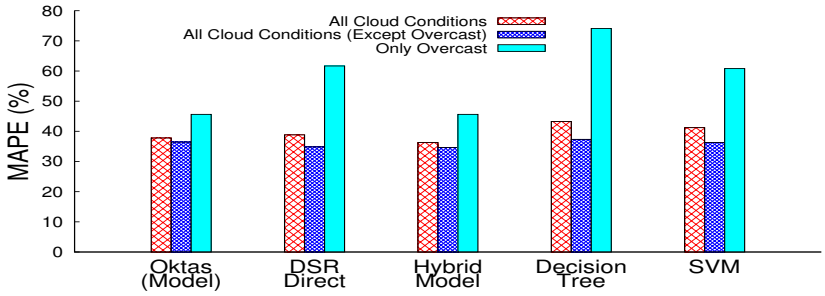


Figure 5.8: MAPE for physical and ML models in 2019.

Key Point. The key takeaway point of our evaluation is that the current DSR data product released by NOAA, which represents the state-of-the-art in satellite-based estimates of surface irradiance, *does not* substantially improve solar performance modeling when compared with using okta-based measurements from weather stations. While DSR is slightly more accurate under non-overcast cloud conditions, it is significantly less accurate under overcast skies. In addition, DSR is also frequently unavailable, which is a significant drawback.

5.6 Related Work

Solar performance modeling that infers a site's solar generation from its location, time, physical characteristics, and weather is a foundation for performing a wide range of solar analytics. There has been significant prior work on solar modeling and forecasting. Recent work on data-driven modeling develops techniques to automatically derive solar performance models for small-scale sites using public data, such as aerial imagery and weather data and thus are more scalable than prior manual approaches [44,52,53,85]. Using satellite data to infer ground-level irradiance has also been well-studied. For example, the Heliosat algorithm [75] is nearly 30 years old and uses visible satellite imagery to infer the global horizontal irradiance based on cloud cover. Our work differs from this and other work on this topic by specifically evaluating NOAA's DSR data product derived from the new generation of GOES satellites. These satellites were not launched until late 2017 and this

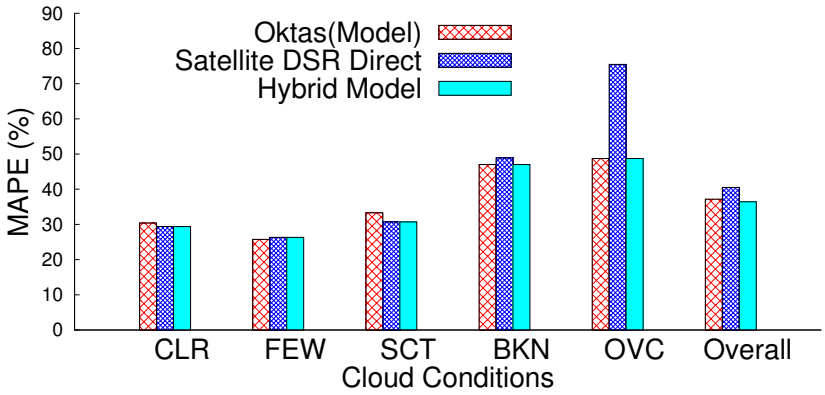


Figure 5.9: MAPE for satellite-based, okta-based, and hybrid model 2018-19.

data product did not become available until 2018. While recent work has compared DSR to ground-level irradiance measurements [66], we know of no work that has evaluated it for solar performance modeling.

5.7 Conclusions

We evaluate the use of DSR estimates from the new generation of GOES satellites for use in solar performance modeling. We show how to leverage DSR for solar performance modeling and compare it with okta-based and ML-based models. We show that the accuracy of satellite-based models depends on the cloud conditions. Surprisingly, our results show that pure satellite-based modeling yields similar accuracy as pure okta-based modeling with a hybrid approach that uses both showing only a modest improvement in accuracy. We also show that ML models are less accurate than physical models, although this may be due to limited training data. In the next chapter,we explore the use of raw multispectral satellite data for solar modeling, rather than the secondary-level DSR data product, especially given DSR's high unavailability. By comparison, the raw multispectral satellite data is always available at a higher resolution (roughly every 5 minutes).

CHAPTER 6

MODELING SOLAR PERFORMANCE MODELS USING MULTISPECTRAL SATELLITE DATA

As we mentioned in the previous chapter, developing accurate solar performance models is increasingly important as the grid's solar capacity rises. The most significant error in existing models is inaccurate estimates of clouds' effect on solar output, as cloud formations and their impact on solar radiation are highly complex. We observed a strong correlation between multispectral satellite data from GOES and solar output, and show in this chapter that training ML-based solar performance models directly on new multispectral satellite data can yield higher accuracy than existing physical models. We further develop both local and global models that are more accurate than prior approaches.

6.1 Introduction

Grid-connected solar capacity continues to grow exponentially at roughly a 20-30% increase per year [25]. This is in line with Swanson's law, which observes that the price of solar PV modules tends to drop 20% for every doubling of production volume [120]. This increase in solar capacity is expected to continue for the foreseeable future with solar power expected to satisfy 25% of global electricity demand by 2050 [23]. Of course, solar's potential is much higher as enough sunlight strikes the Earth's surface in only 1.5 hours to satisfy the world's annual energy consumption [51]. This dramatic increase in solar power is expected to place increasing stress on the electric grid, which must continue to balance supply and demand despite large potential fluctuations in solar power generation that are geographically distributed.

The underlying reason is the mismatch in activation time between solar modules and conventional thermal generators. While solar modules are always active and ramp power up and down nearly instantaneously as clouds pass by, conventional generators may take anywhere from tens of seconds to days to activate depending on their size. As a result, under large solar penetrations, utilities keep many conventional generators active as spinning reserve to quickly offset any dips in solar power. This is both expensive and highly energy-inefficient, and akin to indefinitely maintaining an idling car that is only driven periodically for short distances. While batteries can mitigate some of this inefficiency, they are unlikely to eliminate it at the grid level in the near future due to both high cost and the limited supply of lithium on Earth [72].

A complementary approach to improving grid operations is to improve the accuracy of current and projected solar power output. Solar performance models infer one or more sites' solar output based on their physical and environmental characteristics, and are a basis for a range of solar analytics, including short- and long-term forecasting [50], resource estimation [17], fault detection [7,22], and disaggregation [54,94]. In general, solar power is a well-known function of a module's physical characteristics, e.g., type, wiring topology, inverter, tilt, orientation, location, elevation, etc., and its environment, primarily the time of day, day of year, temperature, and cloud cover. There are many "white box" modeling frameworks, such as PVlib [38] and the U.S. Department of Energy's System Advisor Model (SAM) [63], that enable users to configure their physical and environmental characteristics to estimate solar output. There has also been recent work on "black box" data-driven modeling, such as Solar-TK [44,52], which automatically derive physical characteristics from data, and uses them to estimate solar output based on current or forecasted environmental characteristics.

Unfortunately, the accuracy of these frameworks in estimating solar power is only as good as their input. In general, white-box approaches, such as PVlib and SAM, assume the components of ground-level solar radiation are well-known, e.g, global horizontal irradi-

ance (GHI), direct normal irradiance (DNI), and diffuse horizontal irradiance (DHI). Solar performance modeling is highly accurate given accurate estimates of ground-level solar radiation, derived from either a pyranometer [38] or a nearby solar site. Unfortunately, high fidelity solar radiation data is not widely available at most sites. Thus, prior black-box approaches have estimated ground-level solar radiation using cloud cover estimates commonly reported by weather stations [52, 54]. Unfortunately, the frequency, resolution, and spatial coverage of these cloud cover estimates are coarse and imprecise, which results in significant inaccuracy. Importantly, this inaccuracy in estimating the effect of cloud cover on ground-level solar radiation is *by far the largest source of error* in solar performance models that estimate solar output.

An alternative approach for inferring cloud effects is to use data from satellites. For example, the Heliosat family of algorithms were first introduced in the late 1980s and have been updated since then [47]. These algorithms analyze satellite images in the visible light spectrum, and estimate a "cloud index" by comparing a pixel's actual value with the value it would have under a clear sky. These algorithms generally use physical models that are calibrated from empirical observations of a location, and have grown increasingly more complex as satellite sensors have grown more sophisticated. In particular, the latest generation of U.S. satellites (GOES-16 and GOES-17) include a sensor—the Advanced Baseline Imager (ABI)—that takes images of the Earth with 16 spectral bands, including two visible channels, four near-infrared channels, and ten infrared channels. The ABI is capable of imaging the entire continental U.S. (CONUS) at resolutions ranging from $0.5 - 2$km every 5 minutes.

The U.S. National Oceanic and Atmospheric Administration (NOAA) began releasing both raw data and derived data products from the GOES-16 and GOES-17 in early 2019. As a result, there is now enough raw channel data available to learn solar performance models using machine learning (ML). In this chapter, we show how to develop both local and global solar performance models using ML on multispectral satellite data. Local solar

performance models are trained on data from a specific solar site where the input features include multispectral data and the output is solar power generation, while global models are trained on normalized data from many solar sites. As we show, local models are more accurate, but require local data from each new site for training, while global models are less accurate, but do not require any local data for training.

We compare our ML models above with existing calibrated physical models using both ground-level weather readings and NOAA's estimates of downward shortwave radiation (DSR). The latter estimates also derive from multispectral data, but using a physical model, and represent the state-of-the-art for physical modeling of surface radiation from satellite data. This chapter differs from prior work on estimating solar radiation in that we focus on end-to-end solar performance models that estimate the solar power generation of a particular site (at a specific location and time) using widely available environmental data. Most prior work, including PVlib and SAM, instead decouples estimating surface solar radiation from estimating solar power output based on its physical characteristics, e.g., efficiency, tilt/orientation, shading, temperature coefficient, etc., given surface solar radiation. We focus on end-to-end modeling because it is simpler, and there is less need for decoupling when using ML, as ML training is capable of jointly learning the solar radiation and the effect of a site's physical characteristics.

Our hypothesis is that training ML-based solar performance models on new multispectral satellite data can yield higher accuracy than existing physical models that use either multispectral satellite data or ground-level cloud cover readings. In evaluating our hypothesis, we make the following contributions.

Analyzing Multispectral Satellite Data. We analyze existing multispectral satellite data, and its derived data products, from GOES-16 and GOES-17 that are being made publicly available. We compile a dataset composed of solar generation every 5m-1hr from 29 solar sites at known locations, along with the value of the 16 spectral bands every 5m-1hr, DSR

estimate, temperature, and ground-level cloud cover reading, e.g., clear, scattered, broken, overcast, etc.

ML-based Solar Performance Models. We develop approaches for training both local and global ML models using multispectral satellite data, and compare them with prior approaches that use calibrated physical models. The local models are simple, and trained on a dataset that includes multispectral data as input features and solar generation as the dependent output variable. Instead, the global model requires normalizing each site's solar output to enable training a consistent model across multiple sites.

Implementation and Evaluation. We implement both our ML-based models and existing models and evaluate them on up to 2 years of multispectral data (the maximum that has been released) from the 29 sites. We show that ML-based solar performance models based on multispectral data are much more accurate than weather or DSR-based models, improving the average MAPE across 29 solar sites by over 50% for local models and 25% for global models.

6.2 Background Estimates

Our problem is to develop a solar performance model that infers solar power output for a specific location, time-of-day, and day-of-year given historical solar power output, and the location's environmental data at the same time. Below, we discuss prior approaches that use physical models, but with different types of environmental data as input. We then detail the characteristics of the data sets we use for our learning approach, including the raw channel data gathered by the GOES-16 and GOES-17 satellites.

6.2.1 Prior Approaches

White-box Modeling. Solar performance modeling is a mature area with detailed physical models available that can accurately estimate the power output of a solar system. These models describe how the system's environment (e.g., due to temperature, cloud cover, etc.),

physical characteristics (e.g., wiring topology, conversion efficiency, conversion losses, etc.), and location (e.g., time-of-day, day-of-year, elevation, shading, etc.) affect solar power. White-box modeling frameworks, such as PVlib [38] and SAM [63], require users to configure virtual solar systems that include these details—down to the type of hardware and surface irradiance—and then uses the available physical models to provide a solar estimate. Since solar systems are often highly complex, and surface-level irradiance measurements are often not available, recent work has also explored data-driven approaches to learning the parameters of the physical models solely from historical solar power data [52, 54].

Data-driven Modeling. Data-driven solar modeling approaches estimate surface irradiance by combining well-known clear sky models [27, 96] with simple cloud cover models [52, 80]. Clear sky models accurately estimate surface irradiance based on the Sun's position in the sky, which is deterministic for a given location at a given time-of-day and day-of-year. Simple cloud cover models then translate basic weather station readings of cloud cover, which are made available by the National Weather Service (NWS) for every location in the U.S. These cloud cover readings are coarse observations in units of *oktas*, where 1 okta represents one-eighth of the sky being covered by clouds. The measurements are typically made by placing a circular sky mirror divided into eight slices on the ground, such that any slice that reflects a cloud is 1 okta. Okta-based measurements are typically reported as common string values, such as "clear," "scattered," "broken," and "overcast." Simple data-driven solar modeling uses the reported oktas to estimate a cloud index, which captures the percentage reduction in the clear sky irradiance due to cloud cover. Clearly, measuring cloud cover using oktas is highly imprecise, and thus represents the largest source of error in simple data-driven solar modeling.

Clouds formations are highly complex, and have different impacts on solar radiation depending on their height in the sky and composition. In addition, other atmospheric properties, such as water vapor and aerosol particles, can affect the absorption and scattering of

ABI Band	*Central Wavelength* (μm)	*Spatial Resolution* (km)	*Type*
1	0.47	1	Visible
2	0.64	0.5	Visible
3	0.86	1	Near-Infrared
4	1.37	2	Near-Infrared
5	1.6	1	Near-Infrared
6	2.2	2	Near-Infrared
7	3.9	2	Infrared
8	6.2	2	Infrared
9	6.9	2	Infrared
10	7.3	2	Infrared
11	8.4	2	Infrared
12	9.6	2	Infrared
13	10.3	2	Infrared
14	11.2	2	Infrared
15	12.3	2	Infrared
16	13.3	2	Infrared

Table 6.1: Wavelength for 16 channels of GOES-16 and -17 [26].

solar radiation. These complex effects simply cannot be captured by okta measurements that only range from 1-8.

Satellite-based Modeling. An alternative approach is to estimate a similar cloud index, and surface radiation, using satellite images taken from space. Even early satellite-based imagers were capable of more precision than okta-based measurements. The Heliosat method

was first introduced in the late 1980s [47, 48, 108] to estimate a similar cloud index from visual images of the Earth's surface, and has been improved upon multiple times as satellites have improved. The basic idea is that the more light clouds reflect back in satellite images, the less light reaches the surface. Thus, in visual images, darker pixels represent higher surface irradiance, and lighter pixels represent lower irradiance. Of course, the physical models are highlly complex, as different locations have different ground reflectivities, which can change over time, e.g., due to foliage, snow, roadways, etc. Thus, Heliosat, and methods derived from it, use complex physical models to translate satellite measurements into a surface radiation estimate. The latest methods go well beyond using simple clear sky models, and account for atmospheric changes even under clear skies, such as the Linke turbidity factor. However, many of the latest methods are proprietary, as large-scale solar radiation data is becoming an increasingly valuable commodity for a wide range of applications beyond solar energy modeling [29, 30].

Satellite-based methods also necessarily change as new satellites are launched with new and more advanced sensors. The GOES-16 and GOES-17 are the latest generation of weather satellites launched by the U.S. GOES-16 and GOES-17 became operational in 2017 and 2018, respectively, and cover different regions. GOES-16 covers the east region of the U.S. and GOES-17 covers the west coast and much of the Pacific ocean. The GOES satellites record 16 spectral bands (or channels) for the continental U.S. every 5 minutes at a high spatial resolution (1-2km depending on the channel). By contrast, Heliosat was originally developed for visual images in a single spectral band, while the previous generation of satellites recorded only 5 spectral bands at 15 minutes resolution. Importantly, NOAA makes the satellite data publicly available for download [28]. In addition, NOAA is developing numerous higher-level data products based on the raw spectral data. In particular, the Downward Shortwave Radiation (DSR) product represents the state-of-the-art in estimating surface radiation from satellite data using physical models. The theoretical basis and algorithm for the DSR physical model is described in a 125-page white paper released

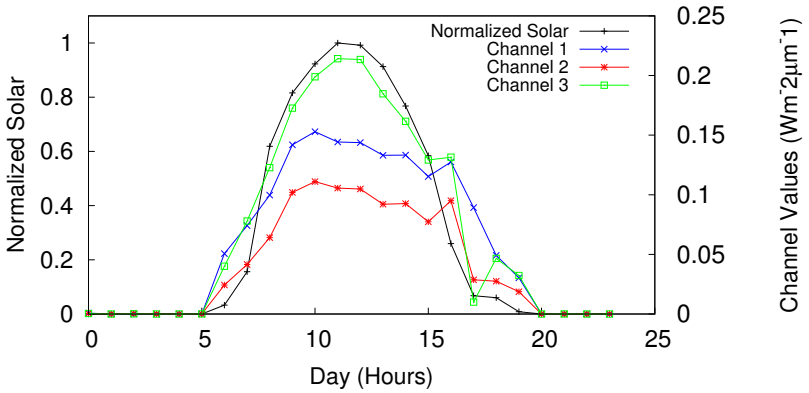


Figure 6.1: Relationship between solar generation at a site, and the values of channels 1, 2, and 3 at the same location.

by NOAA [5]. These DSR estimates are made publicly available hourly, and can be used as input directly into white-box or data-driven models for solar output (or by using it to compute a cloud index). Unfortunately, DSR's physical model cannot often not compute its value under overcast sky conditions, and thus often has missing data points [5].

6.2.2 Spectral Data Characteristics

Table 6.1 shows the centerpoint of spectral bands recorded by GOES-16 and GOES-17, and their wavelengths. Solar cells generate power from wavelengths in the range 0.38μm to 0.75μm, which are described by channels 1 and 2 and part of channel 3. However, while not directly relevant, the other channels may also embed important information about the characteristics of clouds and the atmosphere that could indirectly reveal information about solar generation. A full description of the channels, and what they are capable of sensing, is outside the scope of this chapter. Since we focus on using ML rather than physical modeling for developing our models, the precise meaning of the different channels is not as significant. We simply treat channel data as a "black box" for learning.

Figure 6.1 gives some intuitive sense of the relationship between channel values and a site's solar output on a sunny day with no clouds. The left y-axis shows a site's solar output over a day normalized by the maximum solar output over the day. The right y-

axis then shows the channel values for the same location (within 1km^2 area), which are in Wm$^{-2}\mu$m^{-1}. As shown the channel values follow a similar trend. The relationship with higher channels does not exactly follow the same trend, since they measure longer wavelengths of irradiance that capture properties not reflected in solar output.

/sectionML-Based Solar Models We present both local and global ML solar performance models using GOES satellite imaging data. Local models are trained for each individual solar site using its own data, and only apply to one site. In contrast, global models are trained on data from many solar sites, and are applicable to any new solar site even if data for the site is not available in the training set.

6.2.3 Local ML Models

Our local ML solar performance model is simple: the input features are time-series data of the 16 channel values for a particular solar site location and the location's ambient temperature, while the dependent output variable is the average power generated by the solar site over the same time intervals. The channel values indirectly quantify the surface irradiance, while the temperature is necessary because solar cell conversion efficiency varies with the cell temperature. In general, for every 1°C increase in cell temperature, the efficiency of converting solar irradiance to electrical energy decreases by $\sim 0.5\%$. While we assume a solar site's location is known, and used to determine the associated channel values and temperature, prior work shows how to extract an accurate location directly from solar power data at one-hour or less time resolution [55]. We discuss in detail the process of extracting the 16 channel values for a location from the GOES satellite's NetCDF-formatted data files in Section. We simply retrieve the temperature from Weather Underground, a popular online weather website. Finally, we retrieve solar power data remotely from web-based solar monitoring systems. We have archived multiple years worth of 5 to 15 minute resolution average solar generation data from 29 solar sites.

For each solar site, we curate a training dataset with the timestamp, 16 channel values, temperature, and average solar power generation. We then train an ML model using a support vector machine (SVM). While we could use any ML regression model for training, SVMs have been used for solar modeling in the past, and shown to have higher accuracy than other regression models. SVM is well-suited for a variety of reasons. In particular, it is a non-linear model and our input variables do not have a linear relationship, and we have multiple input features with different magnitudes, units, and ranges. For our SVM, we select a specific error range by setting the margins and a radial function. We define a tolerance margin (ϵ), a regularization co-efficient C, and use the radial basis function (RBF) as the kernel. The tolerance ϵ and co-efficient C are estimated using 5-fold cross-validation on the training data in the following range: $\epsilon \in \{0.005, 0.01, 0.05, 0.1, 0.2\}$ and $C \in \{1, 10, 10^2, 10^3, 10^4, 10^5, 10^6\}$. For our basic model, we perform 5-fold SVM regression for each site individually to evaluate its performance. We use 5-fold evaluation to get a more robust estimate of the performance. We call this a local model because the model is trained separately for each solar site.

A key benefit of our approach above compared to both prior ML-based solar models and prior physical models is its simplicity. We apply a standard ML regression model to a simplistic dataset composed of only three input feature types—timestamp, channel values, and temperature—without using any domain-specific knowledge. As a result, the approach is purely a "black box" that requires only gathering and curating the datasets for training. In contrast, prior ML approaches to solar modeling are much more complex, and not pure black box approaches, because they lack the data necessary to directly infer surface level irradiance [76, 116]. As a result, these approaches must use time and cloud conditions to indirectly estimate surface irradiance. These methods also often mix ML with numerous physical models that describe the effect of temperature, solar geometry, location, and time to improve accuracy. While doing so improves accuracy and reduces the training data necessary to learn the model, it also increases model complexity.

As we have shown earlier, our simple local ML solar performance models, which include no physical models, are significantly more accurate than these prior approaches. The largest source of error in prior approaches stems from the inaccurate measurements of the effect of clouds, which satellite data improves upon. Our ML approach is also able to learn the effects of the physical models above, which mitigates the advantage of using these models.

6.3 Global Model

The local ML models above must be trained for each individual solar site, which requires acquiring sufficient training data to learn the model. In general, roughly one year of training data that captures all of the Sun's positions in the sky across the year is necessary to learn an accurate model. As a result, local ML models have some significant practical limitations. To address this problem, we also develop a global ML model that uses satellite data. Global models, once trained, can be applied to any solar site without retraining the model. These models are less accurate than the local ML models above, but still more accurate than prior approaches that do not use GOES satellite data for estimating the effect of clouds on surface irradiance. The primary reason for the degraded accuracy is that global ML models can conflate the effects of many characteristics that are unique to each solar site when training, including each site's unique shading behavior, geometry (i.e., tilt and orientation), temperature coefficient, wiring topology, inverter type, conversion efficiency, etc. However, as mentioned above, since the effect of these differences is small compared to the effect of inaccurate cloud cover estimates, global ML models are able to maintain higher accuracy than prior approaches.

To develop our global models, we train our models not from data from a single site, but using data from many sites. That is, we combine the training sets for individual sites above together into a large dataset. The only change we make is to the dependent output variable, which in the local models is the solar power output in watts. Since solar sites have different

sizes, and thus different solar outputs, we must normalize this power by the maximum power a site is capable of producing *at the given time*. To do so, we adopt an approach from prior work [54] that bounds a site's solar power curve using the solar irradiance curve from a clear sky model. Prior work shows that this method requires few datapoints, and yields an accurate model of a solar site's maximum solar output under clear skies at any time. Dividing solar power by the maximum generation yields a normalized output across sites in the range $[0, 1]$. We then train the global ML model using the same approach as above.

6.4 ML Model Variants

In addition to defining the basic local and global ML models above, we also experimented with many model variants, which we present the results of in. We describe these variants below.

Varying Resolutions. The spectral satellite data is made available every 5 minutes, enabling us to train models at any resolution greater than 5-minutes. By contrast, DSR and ground-level cloud cover observations are typically reported only every hour. Thus, for comparison, we train our ML models at multiple different time resolutions, including 5 minutes, 15 minutes, and 1 hour. Similarly, for comparison, we increase the resolution of DSR and ground-level cloud cover readings by simply assuming that every 5 or 15 minute interval within an hour has the same value.

Varying Channels. We compare the accuracy of using different numbers of channels. While in our basic model, we use all 16 channels, we also examine the accuracy of using only the first 3 channels that corresponds to the visible region and directly senses the wavelengths converted to solar power.

Multi-Satellite Models. While GOES-16 targets the eastern portion of the U.S. and GOES-17 targets the west coast and Pacific ocean, they both capture data from the entire continental U.S. from different angles. Thus, we augment our basic models above, which primarily

use GOES-16 data since most of our sites are in the eastern part of the U.S., to use both GOES-16 and GOES-17 data. This provides data from two different vantage points in space for the same location. To do so, we simply augment our model above to also include the 3 channels of data from GOES-17.

6.5 Implementation

We implemented our satellite-based ML models using multispectral data in python, along with two competing approaches that apply physical models to ground-level cloud cover readings and DSR. We summarize these competing approaches more below. We used python's *scikit-learn* ML library to train the SVM and other regression models. We collect hourly temperature and ground-level cloud cover readings from Weather Underground, a popular online weather site. For the physical modeling approaches, we use the pysolar [1] library to derive a site's clear sky irradiance based on its location and time. We collect solar power data from 29 sites remotely via their energy meter API. We initially gathered data from 75 sites and filtered sites where we could not verify the solar array in satellite imagery, e.g., from Google, did not have minute-level solar data available, or did not have 2 years worth of solar data. This left us with the 29 sites across U.S. which we analyze.

The GOES-16 and GOES-17 multispectral data is made available by NOAA as netCDF files downloaded from Amazon S3 buckets. We use a script to recursively download the data for specific dates each year along with the description of the ABI product, bucket, and the satellite name. The size of each 5 minute netCDF file is in the range of $\sim$75MB, which requires nearly 16 terabytes to store two years of data from one satellite. Each 5 minute file includes data for all locations. To minimize storage requirements, we filter each file as we download it to extract only the channel data for the locations we are interested in, and discard the rest. The DSR data is also made available by NOAA in the form of netCDF files, but using a different mechanism, which currently requires manually submitting a request and then receiving an FTP link for download. Table provides a summary of these

data sources, their units, and their maximum resolution. In our experiments, we compare accuracy of the models at resolutions coarser than the maximum.

The netCDF files for multispectral and DSR data require some processing to filter out the data for the location of interest. Specifically, our python module reads the *imager_projection* variable to convert (x, y) degree coordinates for latitude and longitude to radians. We then search the file for the latitude-longitude pair that is closest to our location of interest. Since these are geostationary satellites, their rotation matches that of the Earth, enabling us to look at the same part of the file each time. Thus, we read a file and first create a list of closest latitude longitude pair using the Vincenty formula [131], which calculates the distance between two points on the surface of a spheroid. This is done to reduce the computational resources so that the process of finding the nearest location is not repeated for each 5 minute file.

6.6 Physical Modeling Approaches

For comparison, we implemented two physical modeling approaches discussed in . These approaches are distinguished by the input data they use to estimate the effect of cloud cover on surface irradiance. We summarize these approaches below.

Okta-based Approach. This approach is described in prior work [52] and uses ground-level cloud cover readings in oktas to capture the atmospheric and cloud effects on solar power output. In particular, the approach uses a simple formula originally developed by Kasten-Gzeplank [80] to translate an okta-based cloud cover reading into a cloud index, which quantifies the percentage reduction in surface irradiance due to clouds. The approach then essentially multiplies this cloud index by a solar site's estimated maximum output at a given time. This maximum output is modeled in the same way as in our global model by finding the tightest upper bound on the data using a parameterized solar curve from the clear sky model. The tightest upper bound is used instead of a best fit, since the maximum solar generation is dictated by the clear sky model's solar curve. The parameters

applied to the solar curve include an efficiency constant, which captures the solar site's conversion efficiency at 25°C, a temperature coefficient constant, which captures the effect of temperature on efficiency, and constants that capture the solar geometry (e.g., tilt and orientation angles). The approach searches for the parameters that dictate the tight upper bound, which provides an accurate model of a site's maximum solar output. More details are available in prior work [52, 54].

DSR Approach. This approach uses the same physical model as above, but instead of using oktas to compute a cloud index uses the DSR value computed from the GOES-16 satellite. As mentioned in , this DSR value is computed from the channel data using a sophisticated physical model [5], which yields an estimate of the surface radiation. We divide this DSR estimate for a location by the clear sky irradiance to yield a similar cloud index as above, and apply it in the same way. Note that DSR is often not made available, as certain conditions prevent it from being computed accurately, especially under overcast skies. In addition, the DSR technical report [5] evaluates its accuracy for estimating surface radiation and highlights that its accuracy degrades as the cloud cover increases, which are, unfortunately, exactly the times when solar performance modeling is most important.

Both approaches above are deterministic physical models that require calibration, e.g., by fitting known model function parameters to data, and do not require black-box ML training of unknown models. Calibrating parameters for well-known physical models is an advantage compared to using ML to learn these models. Figure 6.2 captures the different inputs, and data processing steps for our ML model and these two physical models.

6.7 Evaluation

We evaluate our ML-based multispectral model and compare it to the physical models described above, on 29 sites across two years, which is currently the maximum data available from the satellites. We use two primary metrics in our evaluation: the Mean Absolute Percentage Error (MAPE) and the Capacity Error Percentage (CEP). The MAPE is com-

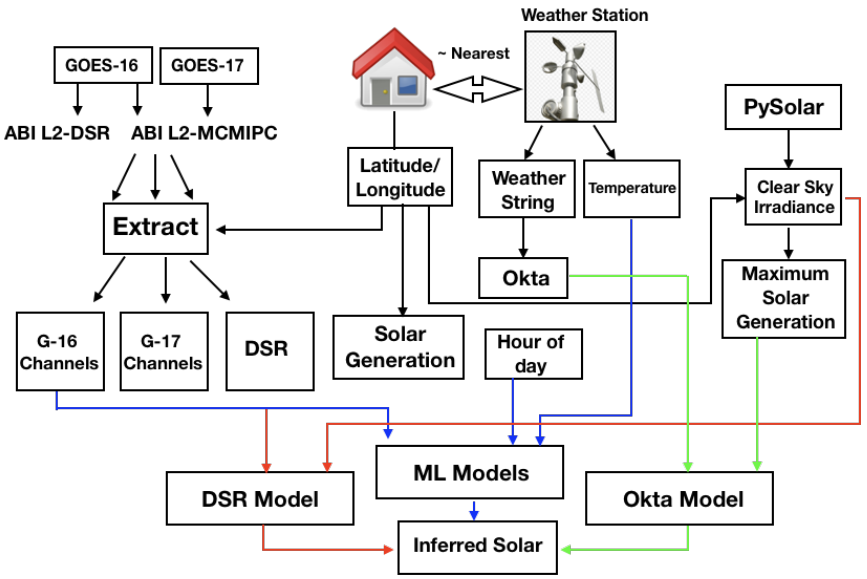


Figure 6.2: Diagram of data inputs for different solar performance modeling approaches we implement and evaluate.

puted as below, where S_t and P_t are the ground truth and model-inferred solar generation, respectively, at time t, and n is total number of temporal data points. The MAPE quantifies the average percentage error across time.

$$MAPE = \frac{1}{n}\sum_{t=0}^{n}|\frac{S_t - P_t}{S_t}|$$

We use MAPE because it is an intuitive metric that is comparable across solar sites of different sizes and configurations. However, MAPE is highly sensitive to periods of low absolute solar generation. For example, if solar generation for a 10kW site is only 10W early in the morning, and our model infers 40W, we record a 300% error, even though the 30W error is only 0.3% of the site's capacity. As a result, MAPE can be significantly affected by these large percentage errors that are actually small and insignificant absolute errors. Thus, we pair the MAPE metric with an absolute error metric, called CEP, that places less weight on these small absolute errors. We define CEP below as the absolute difference in watts between the actual (S_t) and inferred solar generation (P_t) divided by a site's maximum observed capacity (S_{max}). We use CEP instead of other absolute metrics,

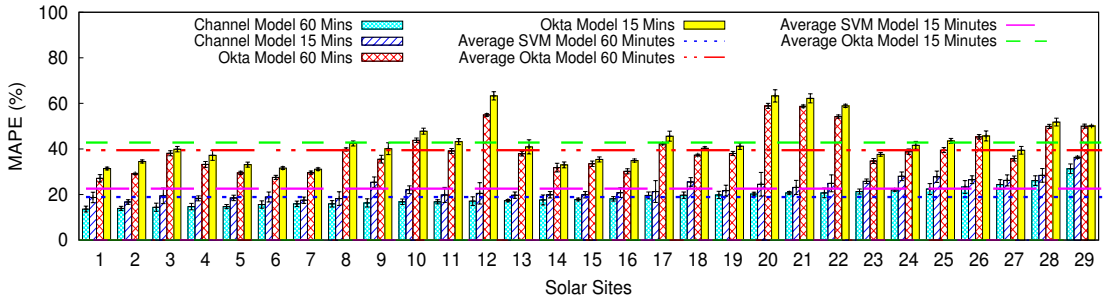


Figure 6.3: Performance comparison across the 29 sites. Data consists of only summer months and the middle of the day. Channel data shows consistently better performance.

such as the Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE), because it is expressed as a percentage and thus is still comparable across solar sites of different sizes.

$$CEP = \frac{1}{n}\sum_{t=0}^{n} |\frac{S_t - P_t}{S_{max}}|$$

6.7.1 Performance of Local Models

We use 5-fold validation in all the experiments. This splits the entire data for each site into 5 sets each of which is used in turn as a test set and the remaining data as the training set. For each site, we compute the average MAPE and CEP for all 5 of the test sets and report the average with standard deviation. We use data from 2018-2019 for all evaluations except where specified.

Analysis on Individual Sites. We first study the performance of the proposed ML model using multispectral satellite data on all 29 solar sites. For this analysis we consider only summer months, May-September, and the middle of the day time period, 10am-3pm. This targets evaluation for sunny periods which are prone to less fluctuations and is the common time period used in evaluation in prior work [44, 52]. The results are shown in Figure 6.3, which shows the MAPE for all the individual sites in inferring solar generation over 15 and 60 minutes time resolution. Note that Okta is only available at a 60 minute resolution and the 15 minute resolutions are obtained by treating the Okta value as constant over the 60 minute period. On the x-axis, we have the different sites in the order of increasing MAPE

under 60 minutes resolution, and the y-axis is the MAPE. For each site on the x-axis, we have four comparisons showing the performance of the Channel and Okta model under 15 and 60 minutes time resolution. The average for the local model across all these sites are shown as flat lines for all the 4 cases.

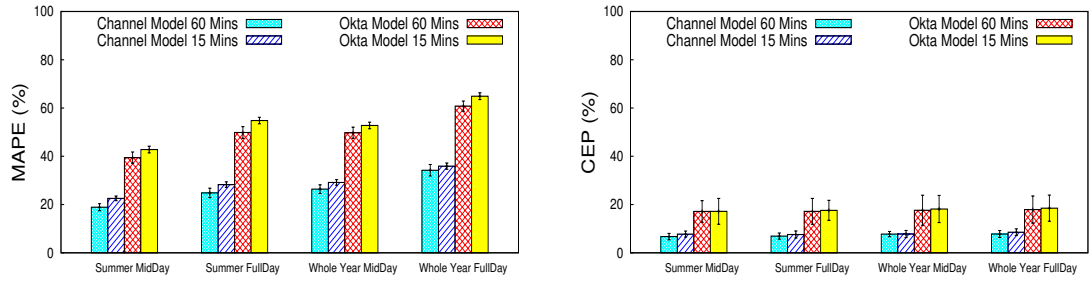


Figure 6.4: Performance comparison of local model under different time-periods. Average over 29 sites is shown. Left shows comparison using MAPE and right using CEP.

We can see that the performance of these models is consistent across all these sites with the percentage error being the lowest for our Channel model. At 60 minutes resolution, which is the minimum resolution for Okta, we observe significantly better solar inference using the Channel model. On average across the sites at 60 minutes resolution, the Channel model gives a MAPE of 18.9% compared to 39.4% from Okta model. At 15 minutes resolution, the Okta performance worsens as expected. The Channel model still performs well with a slightly higher MAPE compared to the performance at 60 minutes resolution. Averaged across the sites at 15 minutes resolution, the Channel model yields a MAPE of 22.6% compared to 42.8% from Okta.

CEP and Performance under Different Conditions. We now analyze the performance for the models under different time periods throughout the year. The averaged result across the sites is shown in Figure 6.4. Summer refers to May-September and mid-day refers to 10am to 3pm. Again, we consider both 15 and 60 minute time resolutions. The trend is the same with 60 minutes Channel Model showing the lowest error and 15 minutes Okta Model showing the highest error under all scenarios. We can also see that the errors are lowest in

the case of summer months and middle of the day time period because this eliminates the period of low solar generation, to which MAPE is sensitive. This is followed by summer months and full day time-period. This period includes data from sunrise to sunset for these summer months, thus eliminating the possibility of snow but still keeping the rainy and cloudy time periods. Furthermore, we can see that when we include the data for the whole year the MAPE further increases. Note that these are the averages across all the sites and contain the mix of sites with snow periods and sites with no snow throughout the year. In all these cases we observe that the Channel Model performs better by almost 50% in comparison to the Okta Model.

Since MAPE is very sensitive to periods of low solar generation, which will be frequent when the entire year is considered, we also analyze the performance with respect to the more balanced Capacity Error Percentage (CEP) metric in Figure 6.4 on the right. Since CEP is normalized with respect to the maximum solar generation for each site, we see that the percentage numbers are considerably lower and performance is comparable for summer mid-day and whole year. For instance, Channel model yields a CEP of 6.7% in the best case sunny scenario and 7.8% across the entire year. Moreover, Channel still leads to substantial error reduction compared to Okta.

DSR Comparison. To evaluate the performance of DSR with the Okta and Channel model, we show the comparison across 2 years worth of data in Figure 6.5. We have again used summer months and middle of the day to show the average results across sites under each model. The number of data points used in this evaluation differ from the other graphs because of the low availability of DSR data. The time periods where DSR value was not available were dropped so as to have the comparison of Okta and Channel model on the same set of points as the DSR. DSR data is only 40% available as compared to the data from the weather station. Interestingly, we find that the DSR performs comparable to the Okta model. DSR analysis has been studied in the past [42].

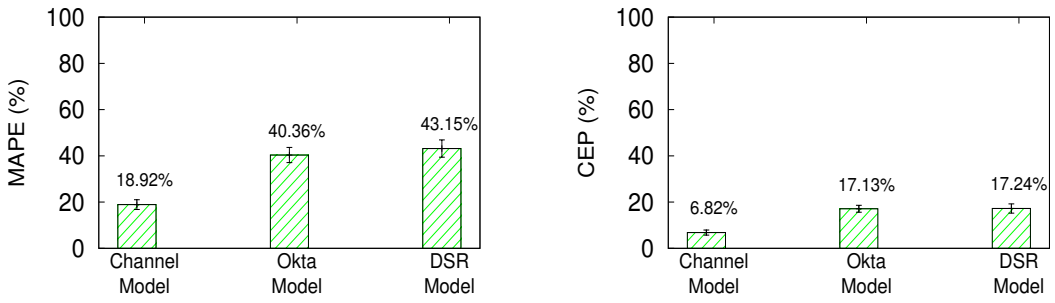


Figure 6.5: Performance comparison of Okta, DSR and Channels. Average across 29 solar sites over summer months and middle of the day is shown. Left shows comparison using MAPE and right using CEP.

Varying Channels. We also evaluate the performance of using the first three channels from GOES-16 versus using all 16 channels. The first three channels correspond to blue, red and veggie (green) bands and together form the visible region. Also we have seen from Figure 6.1, how the first three bands relate to the actual solar output at any given time. The higher channels correspond to higher wavelengths and indirectly contribute to the solar output by embedding information about cloud cover, water vapor, etc. For example, channel 6 and channel 11 have information about cloud particle size and cloud-top phase. Thus, it can be important to look at all the channels while modeling the solar output. In Figure 6.6, we have shown this comparison across different time periods. We can see here that the performance of 3 channels is slightly better than 16 channels when only summer months are analyzed. However, when considering the full year data, 16 channels perform slightly better. This is because summer months only capture the peak of solar generation data while under whole year we have different weather conditions ranging from snow to rain and cloudy periods. At those time periods, using all the16 channels gives better performance.

Multi-Satellite Models. We also analyze the performance when data from both GOES-16 and GOES-17 satellites are combined. As discussed before, the two satellites provide different view points of the same location from space and can provide complimentary infor-

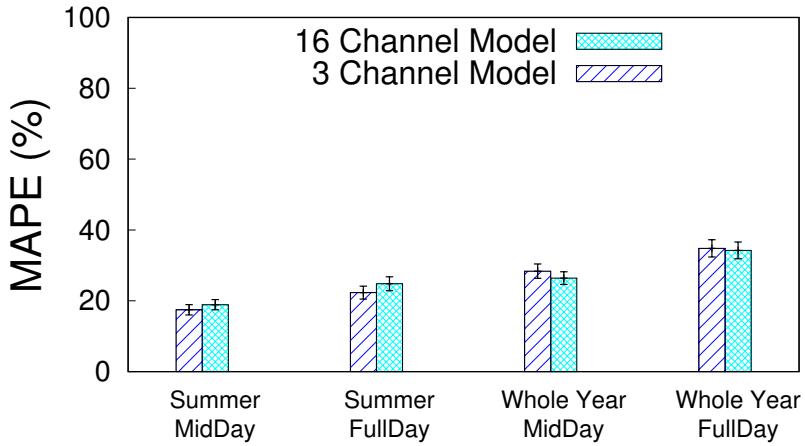


Figure 6.6: Performance comparison of first 3 channels versus 16 channels from GOES-16. Average MAPE across sites is shown.

mation. Figure 6.7 shows the results, comparing both MAPE (on left) and CEP (on right). We have used year 2019 data for this evaluation as GOES-17 data is only available since 2019 [21]. We also evaluate here the performance for the models at 5 minute resolution on this data as the energy meter at the solar sites only store minute-level data for the most recent year. We see that combining the GOES-16 and GOES-17 data further improves the performance for the Channel model indicating that the two satellites provide complimentary information. For instance at 60 minutes resolution, using the combination of satellites improves MAPE from 19.3% to 13.7%. Moreover, even at 5 minute resolution, we observe good performance from the Channel models compared to the Okta model.

Comparing Different Regression Models. While we used SVM regression model in all of the analysis, we also compared different ML based models like decision tree regression, a simple linear regression, and SVM regression. Additional parameters, like the decision tree depth, were also estimated using 5-fold validation on the training data. Figure 6.8 shows the performance of these models. In this case, we perform the evaluation for 15 minute and 60 minute resolution. We can see from the graph that on both 15 and 60 minutes, SVM is performing best and has the highest accuracy, i.e., lowest MAPE. It is also evident that even

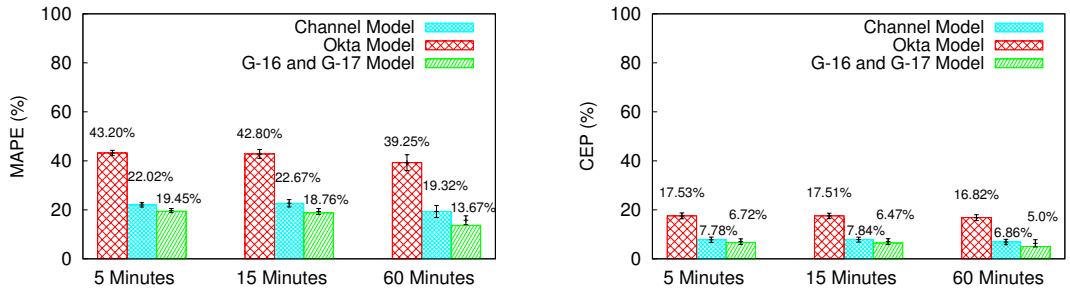


Figure 6.7: Combination of GOES-16 and GOES-17. Average MAPE (left) and CEP (right) across sites using 2019 data over summer months and middle of the day. Performance of the models at 5 minutes resolution is also shown.

a simple regression model performs generally well indicating a strong direct relationship between the channel data and the solar generation.

6.7.2 Performance of Global Models

We now move to the evaluation of the global model. While the local Channel model performs significantly better, it has the downside that it requires at least a year's worth of site-specific data for training. The global model, discussed earlier, overcomes this limitation and builds a general model that is applicable to any new site as long as we have one day's worth of data to calculate the site's physical parameters for maximum solar generation profile [52, 54]. We again use 5-fold validation for all evaluations. This now splits the entire data *based on the sites*, so that each fold consists of 1/5th of the sites. Each fold is then used in turn as a test set and the remaining sites as the training set. For each fold, we compute the average MAPE and CEP across the sites in that fold and report the average with standard deviation.

In Figure 6.9 (left) we have shown our results for the global models. The data for these models are again at 60 minutes resolution for the summer months and middle of the day. The left graph covers only the GOES-16 satellite and hence hourly data of two years. We can see on the y-label the error percentages in the form of MAPE while the x-axis contains the local and global model results. The local model is the individual site-specific model

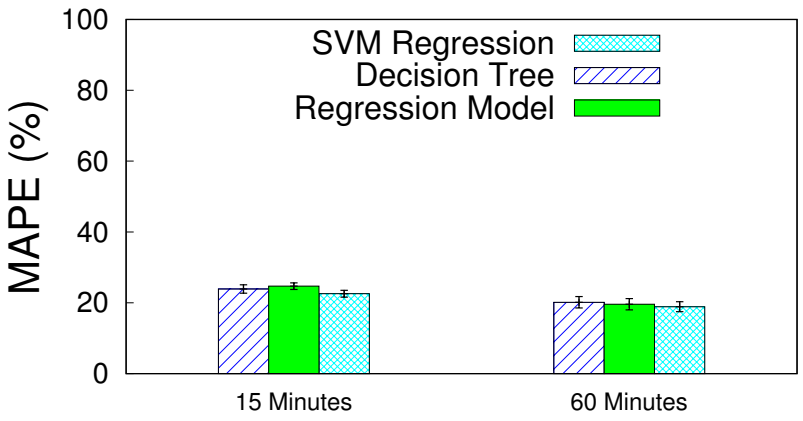


Figure 6.8: Performance comparison of different machine learning models.

evaluated previously and reproduced here for comparison. Note that the local model error is the average across all sites, where some data for each site was used in training the local model while the global model error is on new sites whose data was not used for training the global model. From the graph on the left, we can see that the Okta model has same performance under both local and global setting since it does not learn anything from the data. Comparing the global models, the Channel model still outperforms Okta by a large margin. The error of the global Channel model is higher than the local model. This is expected as the global model does not use any data from the test sites during training and hence will miss site-specific physical parameters, such as shading and location, which are modeled implicitly by the local model.

In Figure 6.9 (right), we have compared the performance of using data from only GOES-16 with the performance of using the combination of GOES-16 and GOES-17 satellites. Note that this only uses year 2019 data. We can see again that the combination model performs better but improvement is not as high as in the case of local model.

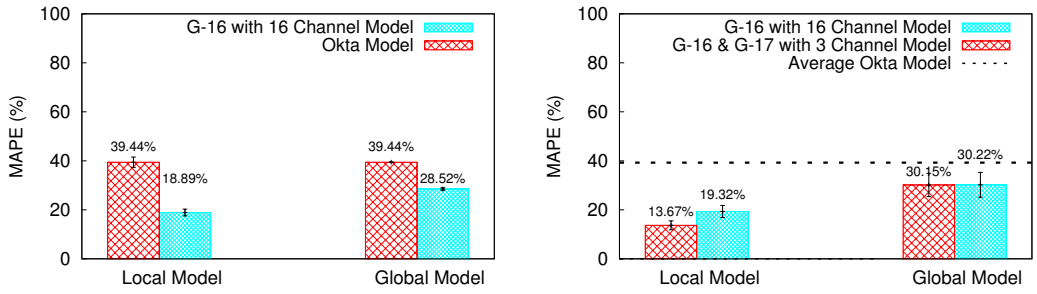


Figure 6.9: Performance comparison of global model. Average MAPE on year 2018-2019 data from GOES-16 (left) and year 2019 data from both GOES-16 and GOES-17 (right). Note that the global model performance is on new sites which are not part of training of the regression model.

6.8 Related Work

This chapter differs from the prior work in multiple ways. First, we use data from the latest generation of U.S. satellites (GOES-16 and GOES-17) launched within the last two years, which includes 16 spectral bands instead of a single band in the visible spectrum. These 16 spectral bands include 3 bands that directly translate to solar generation. This data has much higher temporal and spatial resolution than prior satellites. Second, we do not use any physical models as part of our approach, and instead learn black-box machine learning models. As a result, our approach is highly accessible to those outside atmospheric sciences, which has generally been the domain of solar forecasting. We also use publicly available data from NOAA, so replicating our approach is possible for other researchers. We compare our ML models with a physical model of surface radiation provided by NOAA as a higher level data product called Downward Shortwave Radiation (DSR) [5]. Finally, this chapter also differs from prior work in that we focus on end-to-end modeling of solar power, rather than decoupling models of solar irradiance and solar power generation (given the irradiance). This approach is also more accessible, as real-time solar radiation estimates are not widely available, even though they are required as input to popular modeling frameworks, such as PVlib and SAM.

6.9 Conclusion

In this chapter, we evaluate the performance of using multispectral channel data for solar modeling from the new range of GOES-r series of satellite(s) that cover the east and west CONUS domain of North America. We compare the performance of these channels with conventional ground-level okta-based measurements and a secondary satellite product, DSR, at different time granularities and at different times of the year. Our results show that the multispectral channel data performs better as compared to okta-based cloud measurements and DSR-based approaches by over 50% for local models and 25% for global models. Prior approaches were compared at one hour time granularity and only during sunny conditions whereas we compare our models at finer granularities of 5 and 15 minutes under different conditions throughout the year, with improved results. We also show the merits of combining data from GOES-16 and GOES-17 satellites.

Overall, our results show a strong correlation between satellite data and solar output, and lays a foundation for future work on using multispectral data for solar performance modeling. In the future, this opens up avenues to explore satellite data for better forecasting of solar generation at minute-level resolutions, creating self-supervised solar nowcasting models, to further improve global models using multispectral data, and constructing hybrid models which incorporate both satellite and ground-level measurements for improved performance. We study some of these in the next chapter.

CHAPTER 7

SOLAR NOWCASTING USING SELF-SUPERVISED LEARNING ON MULTISPECTRAL SATELLITE DATA

As we have discussed in the previous chapters, solar energy is now the cheapest form of electricity in history. Unfortunately, significantly increasing the grid's fraction of solar energy remains challenging due to its variability, which makes balancing electricity's supply and demand more difficult. While thermal generators' ramp rate—the maximum rate that they can change their output—is finite, solar's ramp rate is essentially infinite. Thus, accurate near-term solar forecasting, or *nowcasting*, is important to provide advance warning to adjust thermal generator output in response to solar variations to ensure a balanced supply and demand. In this chapter, we discuss self-supervised solar nowcasting models using multispectral satellite data and evaluate our approach for different coverage areas and forecast horizons across 25 solar sites.

7.1 Introduction

Solar is now the cheapest form of electricity in history. As a result, the U.S. Energy Information Administration (EIA) projects that the share of renewable energy from solar and wind in the grid will double to almost 42% by 2050 with solar poised to account for nearly 80% of this increase [60]. This dramatic increase in solar generation is also necessary to mitigate the negative environmental and economic consequences of climate change due to carbon emissions from thermal generators. Unfortunately, significantly increasing the grid's fraction of solar energy, e.g., beyond 50%, remains challenging due to its variability, which makes balancing electricity's supply and demand more difficult. While thermal

generators' ramp rate—the maximum rate they can change their output—is finite, solar energy's ramp rate is essentially infinite. As a result, to maintain the grid's frequency and voltage within a narrow range, utilities will require accurate near-term forecasts of solar generation that provide advance warning to compensate for any changes, either by adjusting thermal generator output or using demand response.

To address the problem, this chapter develops a general model for near-term solar forecasting, or *nowcasting*, from multi-spectral satellite data using deep learning. Specifically, we make use of real-time data obtained from the GOES-R series of satellites, in particular GOES-16. This satellite observes the continental U.S.A in sixteen spectral bands of light, generating rich spatio-temporal data at an unprecedented temporal and spatial resolution: every five minutes for every 0.5-2km^2 area. This data presents an untapped opportunity for solar nowcasting at the continental scale using data-intensive deep learning techniques, which have as yet been limited to sky-imagery data [100, 117, 136, 137] collected using "site-specific" specialized hardware, which inhibits their applicability. In particular, we utilize the intuition that the first three spectral bands of light, corresponding to the visible region, capture information about the solar irradiance and cloud cover at the observed area. Thus, we propose to learn a global deep autoregressive model, i.e., a model that predicts the next observation in a sequence, directly from the raw satellite observations to capture the statistical patterns that are indicative of future solar irradiance.

Our intuition is similar to that of prior work on solar nowcasting using cloud motion vectors. Clouds are the primary reason solar sites' output drops from its clear sky potential, which is largely deterministic based on the ambient temperature, time-of-day, day-of-year, and location. While cloud movements are a function of complex non-linear atmospheric dynamics over long time periods, their movements are more predictable over short periods [57, 89]. Thus, solar forecasting over long periods, e.g., a few hours to days, requires Numerical Weather Prediction (NWP) models [107] that use physical atmospheric models to account for non-linear dynamics. In contrast, solar nowcasting over short time periods

can be easier to model due to the larger influence of more recent changes. In particular, prior work on solar nowcasting has focused on programmatically identifying clouds in satellite or sky images to determine their size, direction, and velocity [57, 89]. Solar nowcasting can use such cloud motion vectors to forecast solar output based on the direction and velocity of clouds. Our approach has a similar intuition, but instead of directly identifying cloud motion vectors for which there is no training data, we train a deep learning model that takes as input historical spatio-temporal satellite observations of a region to infer how spectral satellite data is changing over time and space. Changes in this spectral data implicitly capture cloud movements, as clouds reflect more light, which satellites capture in the spectral data.

Based on our intuition above, we develop self-supervised deep learning models using convolutional neural networks (CNNs) and long short-term memory networks (LSTM) to forecast the next satellite spectral values at the solar sites of interest. These models require historical spectral satellite data over a region surrounding the site as input. We analyze and quantify model accuracy based on both the amount of temporal data, i.e., how far in the past, and the size of the region, i.e., how large of a region, used as input for forecasting 15 minutes in the future. As we show, the more distant the forecast horizon, the larger the historical data and spatial region required, and the lower the accuracy. However, there are rapidly diminishing returns in accuracy improvement, and significant increases in training time and resources, once the historical data and spatial region reach a certain size. We then apply a simpler regression model to infer a specific site's solar output from its spectral forecast data obtained from the self-supervised CNN-LSTM model. We compare our solar nowcasting models with both the accuracy of this model, which infers solar output based on current conditions, as well as a persistence model that assumes that the future power remains unchanged over the forecasting horizon, which also serves as the baseline for comparisons.

Importantly, we condition our analysis above based on the magnitude and frequency of changes in solar output at a given location. Put simply, if a location, such as San Diego, California, U.S.A., is rarely cloudy, then a simple persistence approach that predicts near-term solar output never changes will be highly accurate. Even in highly variable climates, solar output often does not change much over short time periods of 5-30 minutes, which makes simple persistence models appear highly effective. However, accurately forecasting "big" changes in solar output is most important for grid operations, as these are the changes that require an active response. As a result, we specifically focus on the accuracy of forecasting "big" near-term changes in solar power. As we show, the larger the change in solar output, the larger the improvement in accuracy between our deep learning approach compared with others.

Our hypothesis is that solar nowcasting using deep learning models trained on multi-spectral satellite data is both more general and more accurate than prior solar nowcasting approaches, especially at forecasting large changes in solar output. In evaluating our hypothesis, we make the following contributions.

Satellite Data Compilation. We compile a large-scale dataset for 25 solar sites that includes their average solar power generation, ambient temperature, and satellite data across 16 spectral channels for their surrounding region (up to 10km away) every 5 minutes for a year-long period. We use this dataset to train and test our deep learning models, and plan to publicly release it.

Self-supervised Models on Satellite Data. We develop self-supervised deep learning models trained on spectral satellite observations that use convolutional neural networks (CNNs) and long short-term memory networks (LSTM) to forecast the future observations. These models utilize the spatio-temporal observations across all 25 sites (analyzed in this work) for large-scale training. We analyze the importance of both spatial and temporal components, and also compare with other simpler machine learning methods for such self-supervised modeling.

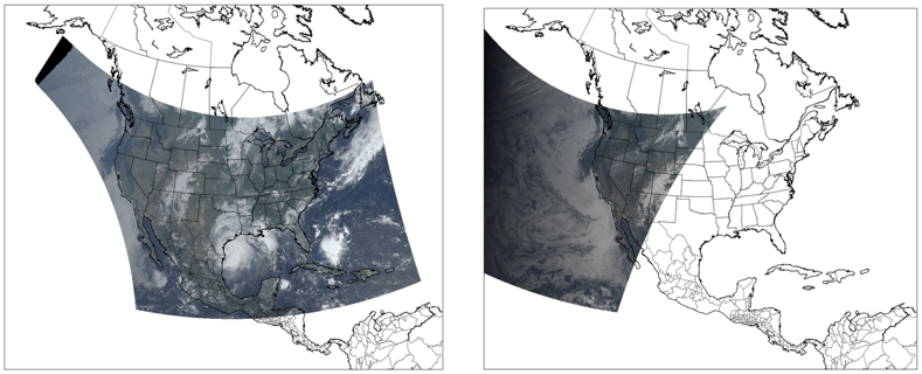


Figure 7.1: GOES-16 (left) and GOES-17 (right) coverage area. We use GOES-16 satellite data as it observes the continental U.S. covering all the 25 solar sites considered in this work.

Deep Solar Nowcasting Models. We demonstrate the utility of self-supervised models for solar nowcasting, which depends on factors like solar irradiance and cloud cover. These models use the forecasted spectral data from the self-supervised model as input to a separate site-specific regression model that predicts a specific site's solar output at 15 minutes in the future from current spectral satellite data. This regression incorporates the effect of physical site characteristics, such as module area, tilt, orientation, and tree cover, on solar output.

Implementation and Evaluation. We implement our models above in python using Tensorflow [37], and train them on a GPU cluster. Given the size of the datasets and complexity of our models, training each model requires $\sim$86 GPU-hours. We evaluate our approach for different coverage areas and forecast horizons across the 25 solar sites, and show that our approach yields errors close to that of a model that uses ground truth observations. We also show that our deep learning models are much more accurate at identifying "big" changes in near-term solar output.

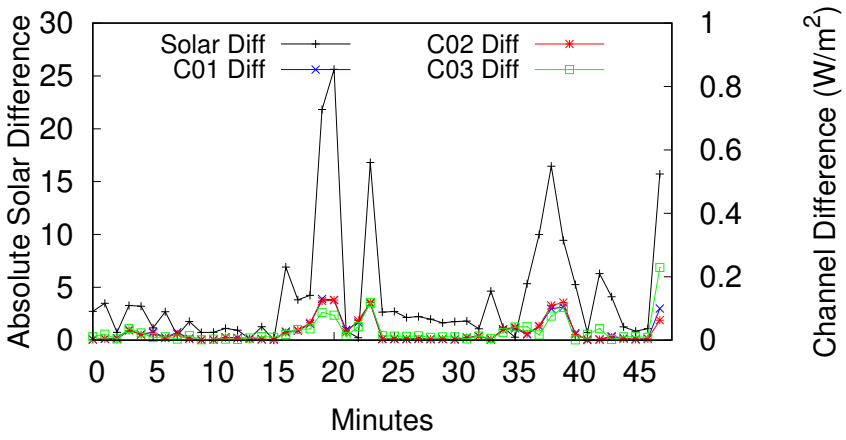


Figure 7.2: GOES-16 (left) and GOES-17 (right) coverage area. We use GOES-16 satellite data as it observes the continental U.S. covering all the 25 solar sites considered in this work.

7.2 Background

The GOES-R series of geostationary satellites started launching in late 2017, and now provides remote sensing data from 16 spectral channels that comprise different ranges of wavelengths of light, as well as numerous secondary derived products, such as estimates of downward shortwave radiation (DSR). Note that solar PV only generates power from the first 3 of these spectral channels, which are mostly in the visible range of light. As a result, our work only trains models on these first three channels. GOES-16 covers the entire U.S., while GOES-17 provides additional coverage for the Western U.S. and the Pacific ocean, as shown in Figure 7.1. Thus, we will use data from GOES-16 as it covers all the solar sites used in this work. The data has both high temporal and spatial resolution, including new spectral readings every 5 minutes for every 0.5-2km^2 area in the U.S., and is made publicly available in near real time. The data is a rich source of information about the environment and climate that is useful for a wide range of applications. Solar forecasting is a particularly compelling application, since solar output correlates directly with the amount of light (of certain wavelengths) that reaches the ground.

7.2.1 Prior Work

Traditionally, solar forecasts depend on some measure of cloud cover to assess the effect of clouds on solar output. Cloud cover is commonly measured by weather stations in units of "oktas," where 1 okta means that one-eighth of the sky is partially covered by clouds. Oktas are measured at ground-level using sky mirrors. Unfortunately, oktas are a coarse and imprecise measure of cloud cover that is typically released by weather stations every hour. In addition, not every solar site is located near a ground-level weather station that reports oktas. Thus, even though cloud cover measurements in oktas are widely available, this data remains an unreliable and inaccurate basis for solar forecasting.

A better basis for solar forecasting is direct ground-level readings of solar irradiance. The U.S. operates the Surface Radiation Budget Network (SURFRAD) [39] within the U.S., which measures and records ground-level solar irradiance at different monitoring sites. These monitoring sites operate in collaboration with the National Oceanic and Atmospheric Administration (NOAA) that also release values of Downward Shortwave Radiation (DSR). Unfortunately, while SURFRAD measurements are precise, they are not widely available, as there are only eight SURFRAD sites maintained in the entire U.S. As a result, we cannot use SURFRAD data as a basis for solar forecasting models. Finally, NOAA releases derived data products from raw GOES spectral data, including DSR [16], which is an estimate of ground-level solar irradiance. Unfortunately, satellite-based estimates of DSR are only released every hour and are often not released during cloudy conditions (when they are most important) [5]. As a result, satellite-based estimates of DSR are also not a reliable basis for solar forecasting. Instead, our work leverages a ML regression model that infers solar output directly from the spectral data, specifically using the channels in the visible range, as the basis for solar forecasting.

Recent work [38,41,42,45,52] has focused significantly on solar performance modeling—inferring current solar output from current environmental conditions—but not forecasting. Solar forecasting is a much more challenging modeling problem, since it must infer, not

only how spectral data correlates with a site's solar output, but also how the spectral data will change over time based on the movements of clouds. Accurately forecasting hours, or days, in the future is challenging because of non-linear atmospheric dynamics that affect cloud movements, and which are not directly captured by GOES data. Such long-term forecasting on the order of multiple hours or days instead require Numerical Weather Prediction (NWP) algorithms [107] that leverage non-linear physical models of the atmosphere, and require more inputs beyond spectral satellite data.

On the other hand, near-term solar forecasting from satellite data is more tractable, since over short time periods, cloud movements are more heavily impacted by recent phenomenon. As a result, models that incorporate historical spectral data across a region have the potential to track changes in the data over time as they move across a region. Prior work on cloud motion vectors [89, 90] has taken this approach in identifying clouds and tracking cloud movements to assist solar nowcasting. However, they largely use physical models, and do not leverage either the latest multispectral data from GOES-R or recent advancements in forecasting using deep learning. The higher resolution data offered by GOES-R admits more accurate, localized, and near-term forecasts compared to prior work based on coarser and less precise data. Similarly, recent advancements in deep learning offer a more automated "black box" approach that does not require manually calibrating physical models for specific data sources or solar sites.

Finally, another line of research [97, 100, 117, 136, 137] utilizes sky-images collected from specialized cameras installed directly at the solar site that continuously capture images of the sky at frequent time intervals. Since this requires specialized hardware to be installed at each solar site of interest, this severely limits the applicability of such models. Indeed, prior work on such methods has only considered a very small number of solar sites in their study, for instance, a maximum of only 2 solar sites in [97, 100, 117, 136, 137]. In contrast, our work utilizes satellite data that is readily available for any location and thus our approach is applicable to any site in the continental U.S.A. We will demonstrate the

utility of our solar nowcasting models on 25 solar sites, an order of magnitude larger than that considered in any of these prior methods.

7.2.2 Basic Methodology

As we discuss in the design section, our approach breaks the solar nowcasting problem into two steps: (1) learning a self-supervised model on the raw satellite data across all sites of interest, and (2) modeling the future solar output at a particular solar site by utilizing predictions from the self-supervised model. The self-supervised model combines a convolutional neural network (CNN) with a long short-term memory (LSTM) for time-series forecasting of spatial multispectral satellite data. CNNs are typically used for analyzing spatial imagery. Multi-spectral satellite data across a region of some size at any moment in time is akin to an image, where the spectral data is equivalent to a pixel value. In contrast, LSTMs have feedback connections that make them well-suited for forecasting temporal data, but cannot be directly applied to spatial data. Thus, combined CNN-LSTMs are useful when the input data has both a spatial and temporal structure, as in solar nowcasting.

Specifically, given a sequence of spectral data covering some area over some number of previous time steps, the self-supervised model is trained to predict the value of the spectral data at the center location corresponding to the solar site's location in the next time step. In our case, we focus on time-steps of 15 minutes in particular. Note that this a self-supervised model as it only makes use of the raw spectral data to predict subsequent samples in that datastream that is already captured by the GOES-16 satellite and does not require any site-specific PV generation data from installed solar sites. This is akin to self-supervised models in ML literature, for instance, language models [104] that predict the next word in a sentence or general auto-regressive models that predict future samples in a sequence [99]. These models can be further specialized to specific supervised tasks of interest.

In order to enable site-specific solar nowcasting, the predictions of the future spectral data at the site's location from the self-supervised model can then be fed to a regression

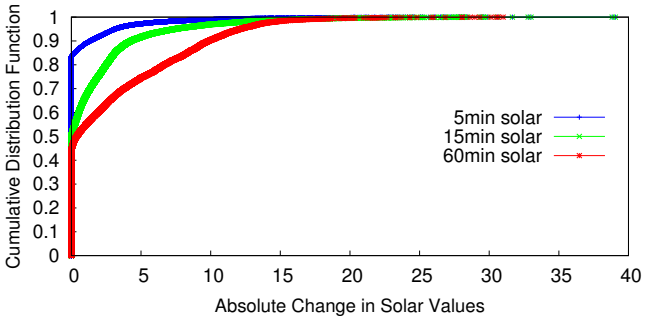


Figure 7.3: Cumulative Distribution Function of changes in solar output at 5, 15 and 60 minute frequency over one year for a representative solar site. Note that a majority of times solar hardly changes between instants (value at 0) and thus modeling the occasional abrupt changes is crucial.

model that infers a site's solar output from given spectral data. These models are capable of incorporating site specific information that affects solar output from current conditions, such as a site's size, orientation, tilt, and shading from obstructions, as we discuss in the design section.

We evaluate our solar nowcasting models above in the evaluation section across 25 solar sites over a year. Our evaluation particularly focuses on the accuracy of our models to predict large changes in solar output over short time periods. These are the changes that are potentially most disruptive to solar-powered systems, including the electric grid. In addition, evaluating solar nowcasting over all time periods obscures the problem, since solar output often does not change much within a 5-15 minute period. As a result, a simple persistence model that predicts solar output never changes over 5-15 minutes are highly accurate when averaged over many time periods, even though they are highly inaccurate, by definition, when any change in solar output occurs. Figure 7.2 illustrates this point by showing the change in solar output every 5 minutes, as well as the first three spectral channel values, over a day for a particular solar site. As shown, most of the time, neither the solar output nor the channel value changes significantly. However, there are a few times within the day that experience significant changes. These significant changes are the ones

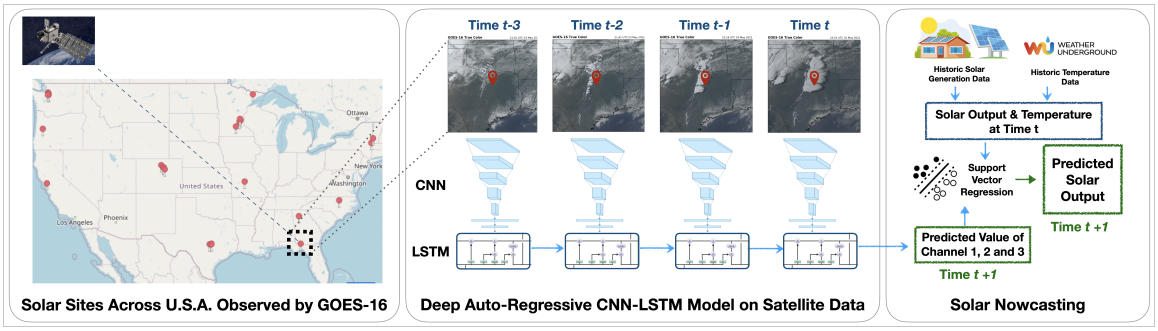


Figure 7.4: Overview of our complete pipeline. Multispectral satellite data from 25 sites across U.S. is collected. Sequences of observations at a site, comprising of first 3 channels from the satellite (visualized in middle) are input to a CNN-LSTM model that is trained to predict the observation at the center (red pin) for the next time step $(t + 1)$. An auto-regressive SVR takes the predicted channels and previous solar output as well as previous temperature to predict the solar output at next step.

that are most important to accurately predict, as they have the most potential to disrupt solar-powered systems and the grid. This graph also demonstrates the correlation between the spectral channel values sensed by the satellite and a site's solar output: they tend to rise and fall in tandem, although the magnitude of the increase and decrease varies over time.

Figure 7.3 then shows a Cumulative Distribution Function (CDF) of the change in solar at 5, 15, and 60 minute periods. This graph shows that, as expected, there are fewer large changes in solar output at small intervals, and the size of the changes are generally larger over longer periods. Thus, accurately predicting the few large changes can be a challenging problem. As a result, we condition our evaluation in the evaluation section on the accuracy of predicting changes in near-term solar output above a specified magnitude. Moreover, note that at 5-minute interval, close to 80% of the data (value at 0) have no change in solar in subsequent time steps. Thus, we choose a 15-minute interval for our study, so that there will be more instances with non-trivial changes in solar.

7.3 Solar Nowcasting Model Design

In this section, we describe our methodology in developing a solar nowcasting model using multispectral satellite data. Figure 7.4 shows an overview of our modeling approach. We first describe our neural network model for self-supervised learning on raw satellite observations, and then describe a simple auto-regressive model for forecasting satellite channel values that can leverage forecasts from the self-supervised model. We then feed these forecasted channel values into a regression model that infers a site's solar output.

Given a set of solar sites of interest, we consider an area of $w \times w$ around each site, and extract the 3 visible channels of satellite data from the GOES-16 satellite. Note that each location l within the area is described by a 3-tuple of channel values. We denote each such 3D image of dimensions $(w \times w \times 3)$ at a location l observed at instant t as $I_t^{(l)}$. We then extract a temporal sequence of these images over time from the satellite, with the target site l always at the center of the image as this is a geo-stationary satellite. This data effectively has four dimensions described by the 2D area (length and width), three channel values, and time. We vary both the area and the amount (and resolution) of historical data we use for training in our models, as the optimal values depend on a location's climate and the target forecast horizon. For example, we could use the previous 4 satellite images for a region of size w with an interval of 15 minutes between images, or the last hour of changes in the spectral data. Larger areas and longer historical time periods increase the training data size, which increases the computational overhead of training. As we discuss, however, there are diminishing returns with respect to improvements in accuracy as these values increase.

7.3.1 Spatial Modeling using Convolution Neural Network (CNN)

Data from the first three channels over an area from the satellite forms a 3D image $I_t^{(l)}$ that we first process using a Convolution Neural Network (CNN) to extract spatial features from the image. A CNN model [84] is the standard neural network architecture used for modeling visual imagery and extracting visual features. CNN models comprise of trainable

convolution filters and pooling operations that together extract spatially invariant features from images. We use multiple layers of convolution filters followed by max pooling layers. The exact CNN architecture is described in the implementation section. The output of processing the image with the CNN model is a k-dimensional feature vector:

$$v_t^{(l)} = CNN(I_t^{(l)}; \theta)$$

where θ represents trainable parameters of the CNN model and $v_t^{(l)}$ is the extracted d-dimensional spatial feature vector for location l at time t. Note that the CNN is trained to extract spatial features that help model the temporal dynamics of the satellite data as explained subsequently.

7.3.2 Temporal Modeling using Long Short-Term Memory (LSTM)

Short-term forecasting of solar power needs to take into account the recent history of changes to solar irradiation at the surface and how it will evolve in a short span of time. We use long short-term memory networks (LSTM) [71] to capture the evolution of the per-instant spatial features extracted from the CNN over time, which is crucial for predicting future satellite channel values. The LSTM is a prominent neural architecture used for modeling sequences of data and is also often employed in time-series forecasting. LSTM makes use of a cell state, that is an internal memory summarizing the previous history at a given time, and a hidden state that's the output of the current time step. Multiple gating mechanisms update the cell state by combining it with the current input and the previous hidden state. In our case, the LSTM update at step t is summarized as:

$$s_t^{(l)}, h_t^{(l)} = LSTM(s_{t-1}^{(l)}, h_{t-1}^{(l)}, v_t^{(l)}; \phi)$$

where ϕ are the LSTM trainable parameters, s_t is the LSTM cell state, h_t is the hidden state, and $v_t^{(l)}$ is the CNN spatial feature vector at time t for location l. Thus recursively

reapplying the same function at every time-step, the LSTM models the evolution of the input features over time.

After processing a sequence of T images through the CNN and LSTM, for instance the satellite imagery over the previous one hour, the final hidden state h_T of the LSTM summarizes the entire sequence. This state is passed through a dense layer with sigmoid output units to predict the value of the visible channels at the site's location for the next step:

$$\hat{C}_{T+1}^{(l)} = \sigma\left(W h_T^{(l)}\right)$$

where W is a $(3 \times k)$ matrix with k being the hidden state dimension and $\hat{C}_{T+1}^{(l)}$ are the predicted values of the 3 channels of satellite at the next time instant.

7.3.3 Training CNN-LSTM Model

The CNN-LSTM model is trained end-to-end in an auto-regressive manner. That is, given the sequence of past images we use the model to compute the predicted values for the next instant and use mean-squared error with respect to the true future satellite values as the loss function:

$$\min_{\theta,\phi} \sum_{l,T} \left\| \hat{C}_T^{(l)} - C_T^{(l)} \right\|^2$$

where C_T^l is the ground-truth satellite observations for all 3 channels at time T and location l, $\hat{C}_T^l$ is the prediction from the CNN-LSTM model as described above, and $\|\cdot\|$ represents the euclidean norm. Note that the prediction is a function of both the CNN spatial extraction model and LSTM temporal extraction model such that backpropagation optimizes the parameters of these models to extract features that can predict future observations well. Thus, using widely available satellite data, we train CNN spatial extraction and LSTM temporal models to capture the dynamics of the multi-spectral satellite data. Note that, in our evaluation, we only train one global model by combining satellite data across multiple sites. This enables modeling of shared statistical properties across sites rather than

overfitting to the peculiar characteristics of an isolated site. Moreover, this also provides a large amount of data for learning a useful CNN-LSTM model which typically do not work well with small datasets.

The drawback of global modeling is that it does not account for unique aspects of any specific location's climate. A local model trained only on data from a specific location is capable of identifying such unique attributes to improve accuracy. For example, in some locations, winds may typically move west to east, while in others, they may typically go in the other direction. However, training our models requires a significant amount of data, and, since the GOES-R satellites only began releasing data a few years ago, there is not yet a large volume of data available for training and testing on any single location. In addition, a global model is beneficial because it does not require re-training, and can be applied to any location.

7.3.4 Forecasting Solar Output

As we discussed earlier that the GOES visible bands are highly correlated with solar irradiance at the surface, this allows accurate inference of solar power output through machine learning models trained on historic generation data with GOES visible bands as inputs. We seek to utilize the same relationship (see Figure 7.2) for solar forecasting by using deep neural networks to implicitly model the short-term changes in the values of the visible bands of GOES.

We use the CNN-LSTM model from the previous section to forecast future values of the visible channels which in turn will help in predicting solar generation, owing to this relationship between channel value and solar generation.

Given a trained CNN-LSTM model that can generate future satellite observations at a given site, we leverage these predictions in a model for solar power forecasting at any solar installation site of interest. We leverage this relationship between visible bands and solar irradiance by considering the following auto-regressive model for forecasting near-term

solar output:

$$P_{t+1}^{(l)} = f(P_t^{(l)}, C_{t+1}^{(l)}, T_t^{(l)}) \tag{7.1}$$

where P_t is the solar power generated, $C_t^{(l)}$ are the satellite channel values and $T_t^{(l)}$ is the temperature at time t. $f(\cdot)$ is a regression model, such as support vector regression, that models the relationship between the input and output variables using historical data. Temperature is an important component of solar generation as solar panel efficiency is sensitive to the surrounding temperature [52]. Note that we use $C_{t+1}^{(l)}$ instead of $C_t^{(l)}$ in (7.1). A major component of change in P_{t+1} from P_t is captured in the change in C_{t+1} from C_t. This complex relationship is modeled using our CNN-LSTM model, described above, which allows us to predict an estimate $\hat{C}_{t+1}^{(l)}$, which is an estimate of true channel values at $t+1$ for the auto-regressive model in (7.1).

The regression model for solar nowcasting in (7.1) is trained using current satellite observations. That is, $f(\cdot)$ is trained using the ground-truth satellite observations at $t+1$ time-step ($C_{t+1}^{(l)}$), historic solar output at the location l, and the historic temperature data at location l. This step does not require the use of the CNN-LSTM model. This is required to train an accurate auto-regressive model that, given true satellite observations, can learn to correctly infer future solar output. Once the regression model is trained, instead of using true future satellite observations, which are unavailable, we replace it with estimates from the CNN-LSTM model ($\hat{C}_{t+1}^{(l)}$) to provide forecasts.

7.4 Implementation

7.4.1 Satellite Data and Solar Sites

The GOES-16 multispectral data is made publicly and freely available by NOAA as netCDF files hosted on Amazon S3 buckets. We use a script to recursively download the data for each date each year along with the description of the ABI product, bucket, domain, and the satellite name. The size of each 5 minute netCDF file is in the range of $\sim$75MB, which requires nearly 16 terabytes to store two years of data from a single

GOES-R satellite. Each 5 minute file includes data across 16 spectral bands covering the entire American subcontinent. To minimize storage requirements, we filter each file as we download it to extract only the relevant channel data for specific area around the locations of interest, and discard the rest.

The netCDF files for multispectral data require some processing to filter out the data for the location of interest. Specifically, our python modules read the *goes_imager_projection* variable to convert (x, y) degree coordinates for latitude and longitude to radians. We then search the file for the latitude-longitude pair that is closest to our location of interest. Since these are geostationary satellites, their rotation matches that of the Earth, enabling us to look at the same part of the file each time. Thus, we read a single netCDF file and first create a list of the closest latitude-longitude pairs using the Vincenty formula [131], which calculates the distance between two points on the surface of a spheroid. This is done to reduce computational resources so that the process of finding the nearest location is not repeated for each 5 minute file.

For all of our analysis, we use data for 25 solar sites across two years. Satellite data for the continental U.S. is extracted for the 2019 year. We restrict modeling to a 10×10 window around the 25 solar sites which constitutes the training data for the CNN-LSTM model for compute efficiency. The 10×10 window covers an area of approximately 10km^2. We average observations in a 15-minute window, which reduces the sequence length for modeling and noise in the data by reducing the number of missing observations and sensor errors. Solar sites used in this work can be seen in Figure 7.4 and are uniformly spread across the continental U.S., including the two coasts and central regions. Since solar modeling only makes sense during the day time, we restrict the satellite data to be from 9am in the morning to 5pm in the evening based on the local time of the solar sites. This yields more than 300,000 5-step sequences of 10×10 images (with 3 channels) at intervals of 15 minutes.

We use 5-fold validation in all the experiments, splitting by day so that test sets have entire days held out for evaluation. This is done for both types of evaluation: evaluating

channel prediction models and evaluating end-to-end solar nowcasting. Table 7.1 shows the training, validation, test split of these satellite observations used in this work.

Solar generation data from the energy meters for the same 25 sites and temperature data from the weather station are obtained from years 2018-19. We restrict generation data to be from 9am to 3pm every day, which is the peak duration of solar generation.

7.4.2 Model Hyper-Parameters and Metrics

CNN model comprises of 2 blocks of convolutions, where each block contains 2 convolution layers with 32 filters of size 3×3 and ReLU activation followed by a max-pooling layer of size 2×2. This is followed by two dense layers with hidden dimension $k = 256$ and ReLU non-linearity between layers. We use a one layer LSTM that takes these 256 dimensional inputs and has a hidden state dimension of 64. Hyper-parameters for this and other ML models considered (decision tree and random forest) were determined using the validation set.

We use three metrics to evaluate the performance of our channel forecasting models and end-to-end solar forecasting models. Mean Absolute Percentage Error (MAPE), which quantifies the average percentage across time.

$$MAPE = \frac{1}{n}\sum_{t=0}^{n} |\frac{A_t - P_t}{A_t}| \qquad MAE = \frac{1}{n}\sum_{t=0}^{n} |A_t - P_t|$$

Here, A_t and P_t represents the actual and predicted values. MAPE, which is often used to quantify the performance in prior work [134], is an intuitive metric and is comparable across solar sites of different installation sizes and configurations. However, it is sensitive to periods of low absolute solar generation and can be significantly affected by small absolute errors. We also used mean absolute error (MAE) to quantify the error in channel modeling given that the first three channels are reflectance values in the range of 0 to 1. In the graphs, we use AE times 100 to plot unless otherwise stated. We also use Capacity Error Percentage (CEP) to quantify the errors of the end-to-end solar forecasting models that we proposed

Data Sets	*Number of points*	*Number of days in a year*
Training	236375	262
Validation	27040	30
Testing	65258	73

Table 7.1: Total number of sequences as well as the number of days of the year that comprise training, validation and testing for the CNN-LSTM model.

in the previous chapter. We define CEP as the as the absolute difference in watts between the actual and forecasted solar generation divided by a site's maximum observed capacity (A_{max}).

7.5 Evaluation

In this section, we evaluate our proposed CNN-LSTM for deep auto-regressive modeling on satellite data and its utility for end-to-end solar nowcasting. First, we evaluate the efficacy of the CNN-LSTM model for predicting future values of satellite channels for a given location. We consider our evaluation along spatial and temporal axes, as well as consider alternative ML models. Then, we utilize the trained CNN-LSTM model obtained from that analysis for solar nowcasting at the site locations. We use the following terminology throughout the evaluation:

Persistence Model: Since all the models considered here predict values for the next instant, typically 15 minutes in the future, a natural baseline is one that assumes there will be no change in the predicted quantity also termed as the persistence model. As we discussed earlier, solar output often changes in small, abrupt bursts and thus a large fraction of the time there is negligible change in near term solar output (see Figure 7.3). Thus, improving over this persistence model prediction is difficult and is of interest in this work.

Tolerance: As across the year, changes in solar (and hence satellite observations) over short durations like 15 minutes are often negligible, we perform all of the analysis in buckets of varying threshold over subsequent changes. That is, we take a tolerance δ and consider only points x_i where subsequent changes were at least δ: $\{x_i| \ |x_{i-1} - x_i| \geq \delta\}$. We evaluate all models over a range of different values of δ to provide a sense of how they perform over both small and big, sudden changes in solar. Table 7.2 lists the fraction of points in the validation data for each tolerance value considered.

Forecast Skill Score: We also use "forecast skill score" (SS) to compare the performance between various methods, which is also used in prior work [134, 136], given by:

$$SS = \left(1 - \frac{\mathcal{E}_{\text{prediction}}}{\mathcal{E}_{\text{baseline}}}\right) * 100\%$$

Here, $\mathcal{E}$ is the error metric used to evaluate performance for every model. If the *prediction model* performs equally well as the *baseline model*, the skill score will be 0. A higher skill score thus means that the prediction model outperforms the baseline model. We will use the skill score to compare the performance between different models. For the skill score, the baseline is always the persistence model as presented above.

Tolerance	***% of points C01***	***% of points C02***	***% of points C03***
0	100	100	100
0.01	65.24	68.33	80.56
0.02	47.32	49.95	64.54
0.05	22.0	24.34	35.39
0.10	8.38	9.63	13.29

Table 7.2: Variation in number of data points with respect to different tolerances for different channels.

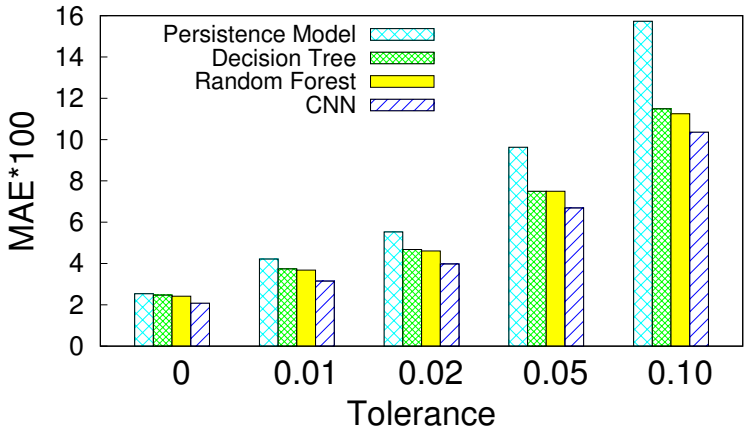


Figure 7.5: MAE for different ML models used to predict the next instant channel values from current instant. CNN model has the lowest error.

7.5.1 Evaluating ML Models for Spatial Modeling

In this section, we consider different choices of ML models for spatial modeling and evaluate their utility compared to using CNN. We also consider choices like the spatial area and its effect on predicting future channel values. For this purpose, we consider three standard ML models: (1) Decision Tree, (2) Random Forest, and (3) Convolution Neural Network (CNN). Note that we only analyze next step predictions given current data (that is a lag-1 time-series models) and consider longer history temporal modeling in the next section. We only show results for channel 1 to avoid repetition as results for other channels are qualitatively similar. To train our decision tree and random forest models, we flatten the $w \times w$ spatial satellite observation into a vector of size $w * w$ that is input to the model.

Figure 7.5 shows the mean absolute error (MAE) for all three models at different tolerances. Here all models are trained over a 10×10 area around each solar site and predict the satellite observations in the next 15 minutes. The point of this graph is to show how the accuracy of a deep learning approach improves relative to that of simpler non-spatial models as the size of the subsequent change increases. At 0, which represents all of the data points, the CNN model is only marginally better than the other models. However, this

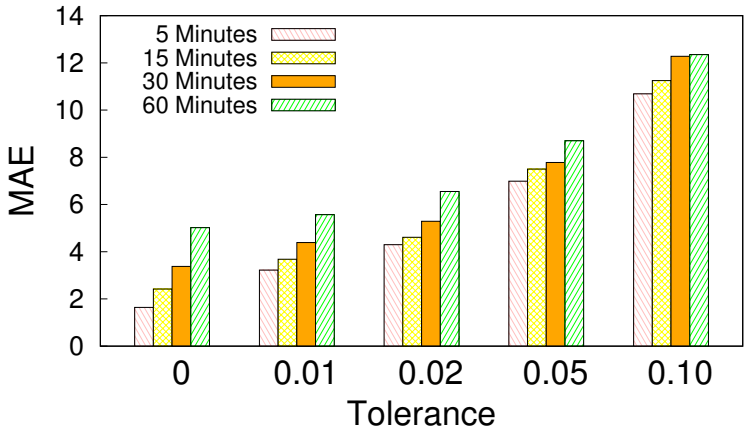


Figure 7.6: MAE for models trained/evaluated at different time intervals. At larger intervals (30-60 minutes) forecasting becomes increasingly harder.

occurs primarily because most of the time there are only small changes in solar over short time periods, as evident from the low error of the persistence model. As we increase the size of the changes we examine, we see that the persistence model predictions, which assume the past is the same as the future, become increasingly worse, while the CNN model remains the best and improves over the other models by a large margin.

Figure 7.6 shows the effect on forecast error for models trained and evaluated at intervals of 5, 15, 30 and 60 minutes. We use the random forest model for this evaluation as training a CNN for every setting is expensive. As expected, predicting further into the future is less accurate, since more changes occur. This discrepancy in accuracy is most evident at 0 tolerance when we include all data points. This occurs because there are few changes in solar output over 5 minutes on average, while on average there are much more significant changes over 60 minutes, including changes due to movement of the sun in the sky. As we increase the tolerance to assess the accuracy of predicting larger changes, as expected, the error increases. However, interestingly, the discrepancy in error actually decreases. That is, the error in predicting a large change 30-60 minutes in the future is more similar to predicting a large change 5-15 minutes in the future.

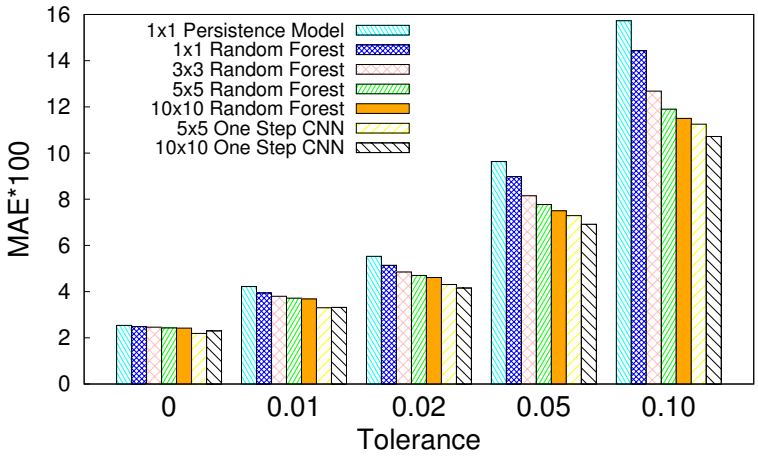


Figure 7.7: Effect of spatial area on forecast MAE for different models. Using larger area improves forecast for all models and CNN trained using (10×10) area is better.

Figure 7.7 then compares the effect of using different sizes of spatial area, focusing on the two best models from Figure 7.5. We train models with areas of 1×1 (i.e. just the site), 3×3, 5×5 and 10×10. The graph shows that increasing the spatial area around the site used by the model results in a large improvement in accuracy. We see that using 10×10 area gives best results, and it is much better than not considering any surrounding area from the site (1×1). Moreover, using the CNN model results in much better spatial processing and improved results over variants of random forests.

7.5.2 Evaluating CNN-LSTM Models for Spatio-Temporal Modeling

We now add the LSTM on top of the 10×10 area CNN model and utilize the previous timesteps as input to the CNN-LSTM model. In this case, we use 4 steps, which means we train the model on a dataset that includes the 4 previous 10×10 spatial regions, corresponding to the past 1 hour of spatial observations. The overall results are shown in Figure 7.8, comparing the CNN-LSTM model with a single step CNN model using 10×10 spatial area. The use of the LSTM drastically improves the results in terms of accurately predicting large changes. Similar to the previous graph, when evaluating across all of the data

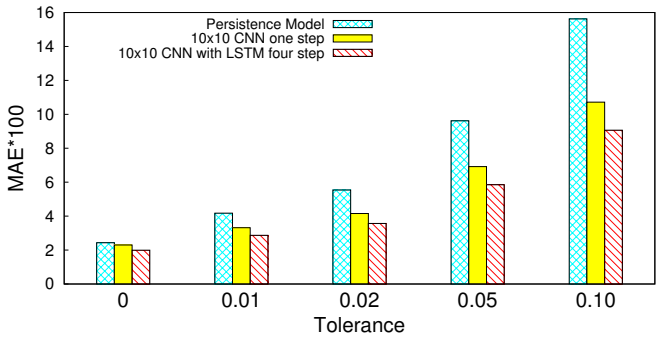


Figure 7.8: CNN-LSTM performance at forecasting next instant channel values. Compared to not using previous temporal history (CNN one step), CNN-LSTM leads to significant error reduction on predicting large changes (tolerance>0) and retains overall better performance compared to persistence model (tolerance$=0$).

(tolerance 0), the improvements over single step CNN are not significant because most of the time there are only small changes in solar output. However, the advantage of the CNN-LSTM becomes apparent when we look at predicting any significant change larger than 0 (tolerance 0.01) and as we increase the threshold of the changes we look at, we observe that our model that covers a 10×10 area and combines a CNN-LSTM leads to larger reduction in errors.

We next evaluate the performance of CNN-LSTM variants in forecasting next time instant channel values. We explore the following temporal variants: CNN using 1-step static image, CNN-LSTM using 1-step static image, CNN-LSTM using 2/3/4 steps of images in the past. Figure 7.9 shows our results compared with the precision model predictions. Incorporating multiple steps of information in CNN-LSTM is better than using the current static image for forecasting, showing the utility of a deep auto-regressive approach. We find that using 3 or 4 steps, i.e., 45 minutes or 60 minutes in the past, perform comparably. Notably, using even just 2 steps leads to a marked reduction in error. This signifies that the model is likely able to infer temporal changes in the satellite data, such as cloud movement, for better prediction.

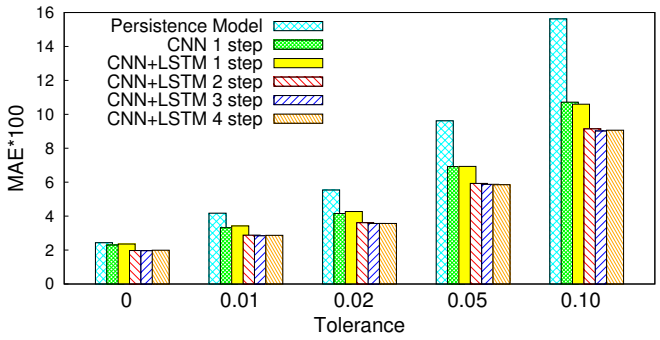


Figure 7.9: Comparison of CNN-LSTM models using varying amounts of previous temporal information. Note that using 3-4 steps is better than using lesser steps. There is a marked reduction in error when using more than 1 step in the model.

7.5.3 End-to-End Solar Nowcasting

The end goal of this work is to use these channel predictions and translate them into an end-to-end solar nowcasting model. In this section, we evaluate the utility of our models for this purpose. Since we want a clear comparision of the benefit of using the self-supervised CNN-LSTM model for solar nowcasting, we use the CNN-LSTM model as a fixed model in the nowcasting part. That is, after the self-supervised learning on raw satellite observations, this model is fixed and not trained further on any site-specific data from the solar installation sites. This enables us to precisely understand the contribution of the predictions from the self-supervised model in solar nowcasting–if the predictions are useful, it will improve nowcasting results over using a persistence baseline's prediction. Moreover, this enables faster computation and cheaper memory overhead as the expensive CNN-LSTM model is not trained on each of the many solar sites.

We use the SVR auto-regressive model, discussed in the solar forecasting section, to forecast 15-minute ahead solar generation. We consider 4 different models to evaluate forecast at time t:

(1) Persistence model: this is again a simple past predicts the future baseline;

(3) CNN-LSTM-SVR: an SVR model using the predictions of CNN-LSTM, that is the fore-

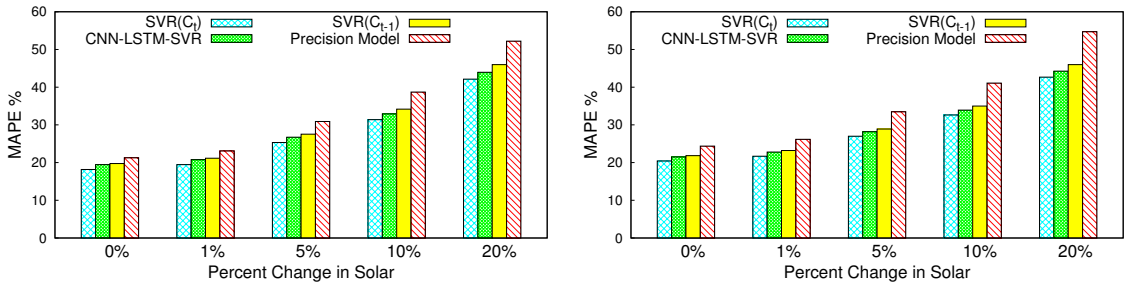


Figure 7.10: End to end solar forecasting on 10x10 km area, averaged over 25 solar sites over 15 mins. Performance for summer months (May-September) is shown on left and for the full year on the right. Using the predictions from the CNN-LSTM model leads to solar output forecasting with error close to that of using the current satellite observations. Compared to precision model forecasts, this approach is consistently better, specially at predicting when there will be large changes in solar ($\geq 5\%$).

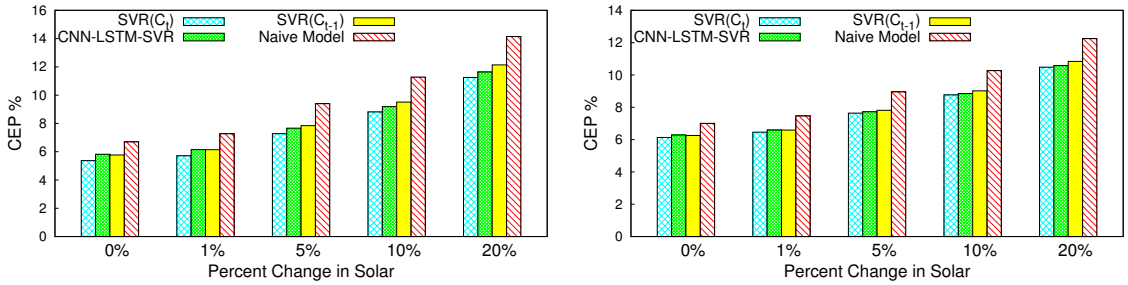


Figure 7.11: CEP for end to end solar forecasting on 10x10 km area, averaged over 25 solar sites over 15 mins. Performance for summer months (May-September) is shown on left and for the full year on the right. The performance of CEP is similar to that of MAPE where the predictions from the CNN-LSTM model leads to solar output forecasting with error close to that of using the current satellite observations.

casted channel values from past values (C_{t-1},C_{t-2},C_{t-3}) from the self-supervised CNN-LSTM model;

(3) SVR(C_{t-1}): using the persistence model on satellite observations, C_{t-1}, as the forecast input to SVR instead of CNN-LSTM forecast, this should be an upper-bound on the error only if the CNN-LSTM model produces useful forecasts;

(4) SVR(C_t): this is a lower-bound on the error that uses the ground-truth satellite observation at the *future* instant and is not a feasible forecast as C_t is unavailable ahead of time.

Note that SVR(C_t) uses the current satellite observations to make predictions. This is a lower-bound on the error of the model given a particular site with some historical data when using SVR auto-regressive models. Estimates of satellite channel values, $\hat{C}_t$ from a model using previous instants observations, will be useful if they lead to an accuracy that is closer-to this performance of using the actual observations. Similarly, SVR(C_{t-1}) corresponds to using persistence model predictions as the future satellite channel estimates, assuming no change in values, and will be an upper bound on the error. Model error should be below this value for the model to be useful for solar nowcasting.

Our results are shown in Figure 7.10, and consider both scenarios of only summer months and the whole year. We include both scenarios, as typically forecasting is easier over summer months, due to largely sunny days, and more difficult over the rest of the year, due to phenomenon like rain, clouds, and snow. The performance of forecasting solar using CNN-LSTM is close to using the ground truth channel values from the future in the model, an upper-bound, hence showing that the approach is useful and accurate for solar forecasting. We have further split the performance of these models into percent changes between successive solar generation values, as shown on the x-axis, where 0 means any change and includes all the values, whereas as 5% means a change of at least 5% in subsequent values and so on. We can also compare the results in the left and right plots of Figure 7.10 in that they both show similar trends but only differ in the MAPE, which is higher for a full year and a little lower for only summer months. Interestingly, we find that these models are not drastically worse when evaluating over non-summer months, which indicates that the models capture rich spatio-temporal phenomenon from the satellite data for accurate modeling. Alternatively, we have also shown the results from the CEP metric on the same data in Figure 7.11. The performance of CEP is similar to that of MAPE, thus showing the usefulness of the forecasting model.

Skill score is a popular metric used to understand the performance of solar forecasting models. For solar nowcasting, the persistence model is the default baseline model that is

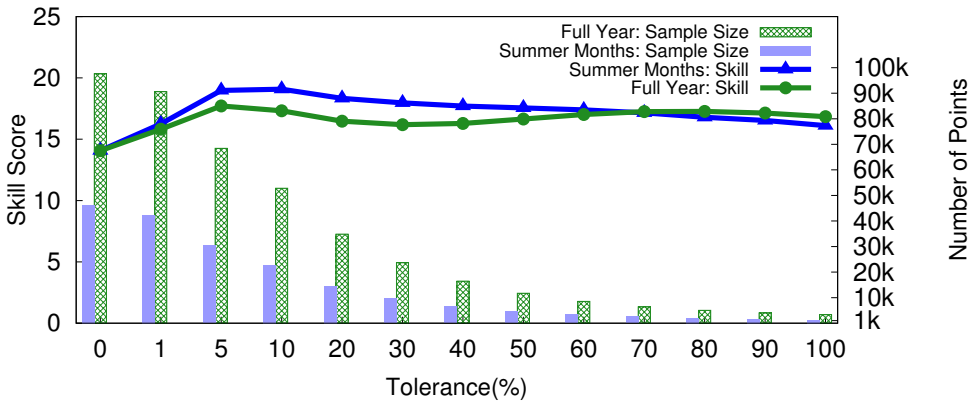


Figure 7.12: Forecast skill score distribution over summer months and full year, along with the distribution of the number of instances at different tolerances.

used in prior work [97,119,134,136]. Figure 7.12 shows the average of the skill score across all the 25 sites at different tolerances. In addition, we show the distribution of the number of data points available for evaluation at the various tolerances through a histogram. We can see that the solar nowcasting model improves over the persistence baseline yielding average skill score in the range of 14-19%. We see that over full year, forecasting skill improves as we increase the tolerance of subsequent changes and then drops a bit at predicting very large changes of more than 5%. Interestingly, skill score is consistently high at predicting very large changes during summer months.

Figure 7.13 then shows the distribution of forecast skill across the 25 solar sites. We can see that the skill varies widely across solar sites, from 14%-27% across the 25 sites. These variations are expected as different sites have different characteristics contributing to errors, including differences in installation capacities, shading from nearby buildings or trees, and widely different climates (snow vs sunny). In particular, note that prior work on solar nowcasting using sky-camera imagery has found that state-of-the-art nowcasting models using deep learning have a skill in the range of 10%-20% [97, 119], as evaluated on only 1 or 2 solar sites (typical in sky-camera nowcasting research). The results in Figure 7.13

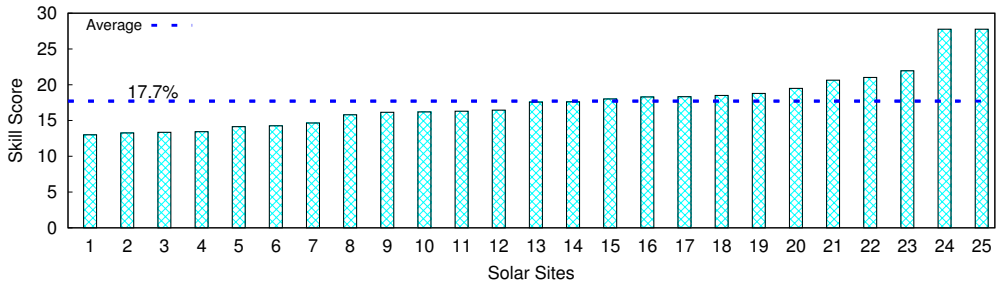


Figure 7.13: Forecasting skill score of all the 25 solar sites at 5% tolerance for the full year.

show that our results are competitive or better, while being highly scalable and cost efficient as they don't require installation of specialized hardware at each solar site of interest.

7.6 Related Work

Forecasting solar PV output is akin to forecasting solar irradiance, since the former strongly correlates with the latter [106]. Numerical Weather Predictions (NWP) algorithms [56, 64, 91, 123], that mostly leverage physics-based modeling, are often used for solar irradiance forecasting. These physics-based models are most appropriate for forecast horizons on the scale of hours to days, and not near-term forecasts on the scale of minutes to an hour [68, 134]. Over long-term horizons, the complex and non-linear evolution of climate patterns can be difficult to model, requiring knowledge of climate processes and the history of many atmospheric events over time that can cause subtle changes.

On the other hand, at shorter time scales of 5 to 60 minutes, machine learning approaches have the potential to implicitly model local changes directly from observational data [109, 134]. While there is recent work on analyzing images from ground-based sky cameras [100, 117, 136, 137] for near-term solar forecasting, it requires installing additional infrastructure at the site. Another alternative is based on estimating cloud motion vectors [57, 89, 90] from satellite images, however ML approaches that more directly model solar irradiance tend to perform better [42, 83]. Our approach differs from recent approaches

in solar nowcasting by forecasting solar irradiance values from multispectral satellite data using a combined CNN-LSTM, which can forecast changes in spatial features over time. We then combine these solar irradiance forecasts with a model that predicts a site's solar output from solar irradiance.

Our approach is a self-supervised approach, where we directly use abundant satellite data for modeling. Such methods have gained increasing popularity in computer vision recently [78]. While such methods have been widely successful, their applications in remote sensing have been limited and their applications to solar modeling have not been explored before. [77] similarly uses self-supervised learning over Landsat images, although their approach is designed for classifying geographical regions and not applicable here. [132] reconstructs visible bands from other bands, in a colorization task, to learn useful representations for land cover classification. Related to our work, [40] uses temporal information for constructing positive-negative pairs for classification of remote sensing data. Unlike this work, they ignore complex spatio-temporal dynamics and auto-regressive modeling, which are more crucial for forecasting. Recently, parallel work [105] utilized generative modeling on radar data for precipitation nowcasting using generative adversarial networks. Compared to this, we focus on a different application of solar nowcasting and are able to demonstrate utility of a simpler model for this application where directly predicting the next instant values is sufficient as they directly correlate with solar irradiance [41] – hence do not necessitate requiring a complex discriminator for learning as in GAN models [65]. Finally, our work is also related to auto-regressive language models, that learn to predict the next word given previous words, which have been very successful for natural language processing [104].

7.7 Conclusion

Our work shows how to apply deep learning to multispectral satellite data to forecast short-term changes in solar output. We propose deep auto-regressive models that com-

bine CNN and LSTMs, trained in a self-supervised manner on abundant satellite data from GOES-R satellites. Such self-supervised training captures rich spatio-temporal dynamics that help improve solar nowcasting and is readily applicable to any solar site of interest, that is captured by the GOES-R satellites, without needing any specialized hardware like prior work. We evaluate our approach for different coverage areas and forecast horizons across a large number of 25 solar sites. Our results are promising and demonstrate that 15 minute forecasts using our approach have an error near that of a solar model using current weather and have forecast skill that is comparable with highly localized methods depending on installation of specialized sky cameras.

While this is a promising first step, we believe there is much progress to be made in this area, as self-supervised learning is a promising approach for rapid progress in this domain due to abundant availability of rich satellite data. The self-supervised learning approach itself can be improved by exploring longer contexts – both in input and output, spatially and temporally – through more sophisticated recent neural networks like Transformers [129] which can model longer range dependencies. Finally, while solar nowcasting is the primary focus here, the self-supervised models presented here should be more generally useful for other applications such as detecting anomalies like wildfires, forecasting cloud cover, precipitation nowcasting, and more.

CHAPTER 8

FUTURE DIRECTIONS AND CONCLUSION

8.1 Conclusion

Distributed solar generation is rising rapidly due to a continuing decline in the cost of solar modules. Nearly all residential solar deployments are grid-tied, enabling them to draw power from the grid when their local demand exceeds solar generation and feed power into the grid when their local solar generation exceeds demand. Utilities need to continuously regulate both the distributed solar and the traditional generators which vary in their activation times. With the digitization of the electrical grid, immense data is being generated at finer time granularities from the grid as well as new range of satellites and ground-based weather monitoring stations. It is hence possible to create data-drive ML models. In this regard, this thesis proposes data-driven techniques for controlling, modeling and forecasting solar for residential homes.

To enable grid-connected solar control, we propose Software-defined Solar-powered (SDS) [118] systems that dynamically regulates the solar flow rates into the grid and also present an SDS prototype, called SunShade. SDS manages grid stability and grid neutrality by introducing a new class of Weighted Power Point Tracking (WPPT) algorithms that enforce a relative cap while eliminating the need for policy imposed hard caps. To allow control for off-grid solar, we consider the benefits of partially or entirely defecting from the grid so as to employ solar without restrictions. We present an architecture [43] for grid defection and implement a trace-driven simulator that enables a home to dynamically switch between a local/generator and solar/battery depending on its power consumption and generation. We further analyze the tradeoffs in terms of number of switches, solar

waste, reliability, carbon emissions and total cost by introducing switching policies: more switching leads to less reliability, but maximizes the use of solar.

In order to exercise better control over the intermittent solar, there is a need for accurate solar performance models which infer solar power output in real time based on the cur- rent physical and environmental conditions. We evaluate the use of DSR estimates from the new generation of GOES satellites for use in solar performance modeling [42]. We show how to leverage DSR for solar performance modeling and compare it with okta-based, SURFRAD (Surface Radiation Budget) Network and ML-based models. We show that the accuracy of satellite-based models depends on the cloud conditions. Surprisingly, our results show that pure satellite-based modeling yields similar accuracy as pure okta-based modeling with a hybrid approach that uses both showing only a modest improvement in accuracy. We also show that ML models are less accurate than physical models, although this may be due to limited training data.

We further explore the use of raw satellite data for solar performance modeling, rather than the secondary-level DSR data product, especially given DSR's high unavailability. By comparison, the raw hyperspectral satellite data is always available at a higher resolution (roughly every 5 minutes). We evaluate the performance of using multispectral channel data for solar modeling from the new range of GOES-R range of satellite(s) that cover the continental domain of North America. We compare the performance of these channels with conventional ground-level okta-based measurements and a secondary satellite product, DSR, at different time granularities and at different times of the year [41]. We also show the merits of combining data from GOES-16 and GOES-17 satellites. Overall, our results show a strong correlation between satellite data and solar output, and lays a foundation for future work on using multispectral data for solar performance modeling.

Further, we explore the use of multispectral satellite data for accurate near-term solar forecasting, or *nowcasting*. We develop a general model for solar nowcasting from abundant and readily available multispectral satellite data using self-supervised learning.

Specifically, we develop deep auto-regressive models using convolutional neural networks (CNN) and long short-term memory networks (LSTM) that are globally trained across multiple locations to predict raw future observations of the spatio-temporal data collected by the recently launched GOES-R series of satellites. Our model estimates a location's future solar irradiance based on satellite observations, which we feed to a regression model trained on smaller site-specific solar data to provide near-term solar photovoltaic (PV) forecasts that account for site-specific characteristics. Our results show that our approach yields errors close to that of a model using ground-truth observations.

8.2 Future Directions

The new-generation constellation of geostationary and polar-orbiting satellites (e.g., GOES/ABI [111] and JPSS/VIIRS [112]) provides unprecedented possibilities for advanced earth observations, scientific discovery and societal benefit. While the visible-spectrum bands of conventional satellites are only useful during the day time, the Day-Night Band on the VIIRS system enables novel sensing of nighttime light, including the ability to observe human activities directly. Paired with a rich array of high-resolution thermal infrared bands, these observations contain new and improved information about the land, ocean, and atmosphere. These observations chart new courses for land-cover mapping, disaster resilience and resource management, modeling and predicting near-term solar power [41, 42], identifying initiation and spread of wildfires, or even monitoring human behaviors such as the impact of the COVID-19 pandemic-based changes to nighttime artificial lighting [61, 87].

Satellite data remains an underused resource in AI methodologies due to several outstanding challenges. One key issue is that whereas satellite data are available in abundance, there is a shortage of processed, labeled data (e.g. correct prediction or classification values) for use in AI training. Here, self-supervised machine learning methods hold a significant potential to extract information from these rich spatio-temporal satellite data in

situations where end-to-end supervised data are limited. Secondly, different satellites have unique characteristics, strengths and weaknesses, which require deep domain knowledge of both the data and their suitability for different applications. For instance, in contrast to standard video sequences, satellite data vary in their spatial, spectral, temporal, and radiometric resolution, and may contain objects with fuzzy boundaries [79]. However, within this diversity there exists rich information content about the underlying processes - waiting to be discovered and leveraged.

In the future, this work can be extended by avoiding the "black-box" solutions in favor of **Explainable AI (XAI)** methodologies [59, 93, 124] which reveal the strategies used by AI methods. Establishing the link between the machine and human intuition, XAI represents the essential bridge to end-user acceptance, particularly in high-stake and time-critical applications.

BIBLIOGRAPHY

[1] PySolar. http://pysolar.org/, 2007.

[2] Solar Energy Industries Association, Solar Energy Facts: 2013 Year In Review. http://www.seia.org, March 5th 2014.

[3] EnergySage, How Much Does a 6,000 Watt (6kW) Solar System Cost in 2018? https://news.energysage.com/6kw-solar-system-compare-prices-installers/, May 30th 2018.

[4] Generac Home Generator Fuel Consumption. https://assurancepower.com/aps-blog/entry/generac-home-generator-fuel-consumption, May 30th 2018.

[5] GOES-R Advanced Baseline Imager (ABI) Algorithm Theoretical Basis Document for Downward Shortwave Radiation (Surface), and Reflected Shortwave Radiation (TOA). Tech. rep., NOAA NESDIS Center for Satellite Applications and Research, November 20th 2018.

[6] Home Depot, 20,000-Watt Air Cooled Standby Generator with Whole House 200 Amp Automatic Transfer Switch. https://www.homedepot.com/p/Generac-20-000-Watt-Air-Cooled-Standby-Generator-with-Whole-House-200-Amp-Automatic-Transfer-Switch-7039/, May 30th 2018.

[7] SolarClique: Detecting anomalies in residential solar arrays. In *COMPASS* (June 2018).

[8] Tesla Powerwall 2.0. https://www.tesla.com/powerwall, May 30th 2018.

[9] U.S. Energy Information Administration, Average Price of Electricity to Ultimate Customers by End-use Sector. https://www.eia.gov/electricity/monthly/epm_table_grapher.php?t=epmt_5_6_a, May 30th 2018.

[10] U.S. Energy Information Administration, Construction Cost Data for Electric Generators Installed in 2015. https://www.eia.gov/electricity/generatorcosts/, May 30th 2018.

[11] U.S. Energy Information Administration, How Much of U.S. Carbon Dioxide Emissions are Associated with Electricity Generation. https://www.eia.gov/tools/faqs/faq.php?id=77&t=3, May 30th 2018.

[12] U.S. Energy Information Administration, Natural Gas Prices. https://www.eia.gov/dnav/ng/ng_pri_sum_dcu_nus_a.htm, May 30th 2018.

[13] U.S. Energy Information Administration, What is U.S. Electricity Generation by Energy Source. https://www.eia.gov/tools/faqs/faq.php?id=427&t=3, May 30th 2018.

[14] Weather.gov Terms. http://www.weather.gov/bgm/forecast_terms, 2018.

[15] NASA, Make a Sky Mirror to Observe Clouds and Contrails. https://mynasadata.larc.nasa.gov/science_projects/make-a-sky-mirror-to-observe-clouds-and-contrails/, June 2019.

[16] Data Products: Downward Shortwave Radiation (Surface). https://www.goes-r.gov/products/baseline-DSR.html, Accessed February 3rd 2020.

[17] Extend: A framework for increasing energy access by interconnecting solar home systems. In *COMPASS* (June 2020).

[18] Geostationary Operational Environmental Satellites-R Series. https://www.goes-r.gov/, Accessed February 3rd 2020.

[19] Instruments: ABI Improvements. https://www.goes-r.gov/spacesegment/abi-improvements.html, Accessed February 3rd 2020.

[20] Instruments: Advanced Baseline Imager (ABI). https://www.goes-r.gov/spacesegment/abi.html, Accessed February 3rd 2020.

[21] SolarAnywhere. https://www.solaranywhere.com/2020/leveraging-goes-17-for-more-accurate-solar-data/, 2020.

[22] SunDown: Model-driven per-panel solar anomaly detection for residential arrays. In *COMPASS* (June 2020).

[23] The SunShot Initiative. https://www.energy.gov/eere/solar/sunshot-initiative/, Accessed February 3rd 2020.

[24] U.S. Energy Information Administration, What is U.S. electricity generation by energy source? https://www.eia.gov/tools/faqs/faq.php?id=427&t=3, Accessed February 3rd 2020.

[25] World now has 583.5 GW of operational PV, April 6th 2020.

[26] ABI Technical Summary Chart. https://www.goes-r.gov/spacesegment/ABI-tech-summary.html, Accessed 2020.

[27] Bird Simple Spectral Model. http://rredc.nrel.gov/solar/models/spectral/, Accessed 2020.

[28] Registry of Open Data on AWS. https://registry.opendata.aws/noaa-goes/, Accessed 2020.

[29] SoDa - Solar Radiation Data. http://www.soda-pro.com/gl/, Accessed 2020.

[30] SolarAnywhere. https://www.solaranywhere.com/, Accessed 2020.

[31] Batteries can't solve the world's biggest energy-storage problem. one startup has a solution. https://qz.com/1133123/batteries-cant-solve-the-worlds-biggest-energy-storage-problem-one-startup-has-a-solution/, Accessed 2021.

[32] Climate change. https://www.seia.org/initiatives/climate-change, Accessed 2021.

[33] Confronting the duck curve: How to address over-generation of solar energy. https://www.energy.gov/eere/articles/confronting-duck-curve-how-address-over-generation-solar-energy, Accessed 2021.

[34] Policies to support solar power deployment. https://cleanenergysolutions.org/policy-briefs/solar, Accessed 2021.

[35] Renewables 2019. https://www.iea.org/reports/renewables-2019/distributed-solar-pv, Accessed 2021.

[36] Solar industry research data. https://www.seia.org/solar-industry-research-data, Accessed 2021.

[37] Abadi, Martín, Barham, Paul, Chen, Jianmin, Chen, Zhifeng, Davis, Andy, Dean, Jeffrey, Devin, Matthieu, Ghemawat, Sanjay, Irving, Geoffrey, Isard, Michael, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)* (2016), pp. 265–283.

[38] Andrews, R.W., Stein, J.S., Hansen, C., and Riley, D. Introduction to the Open Source pvlib for Python Photovoltaic System Modelling Package. In *IEEE Photovoltaic Specialist Conference* (2014).

[39] Augustine, John A., DeLuisi, John J., and Long, Charles N. Surfrad - a national surface radiation budget network for atmospheric research. *Bulletin of the American Meteorological Society 81*, 10 (2000), 2341 – 2358.

[40] Ayush, Kumar, Uzkent, Burak, Meng, Chenlin, Tanmay, Kumar, Burke, Marshall, Lobell, David, and Ermon, Stefano. Geography-aware self-supervised learning. *arXiv preprint arXiv:2011.09980* (2020).

[41] Bansal, A. S., and Irwin, D. See the light: Modeling solar performance using multispectral satellite data. In *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys)* (2020), pp. 1–10.

[42] Bansal, Akansha Singh, and Irwin, D. Exploiting Satellite Data for Solar Performance Modeling. In *SmartGridComm* (November 2020).

[43] Bansal, Akansha Singh, and Irwin, David. On the feasibility, cost, and carbon emissions of grid defection. In *2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)* (2019), pp. 1–7.

[44] Bashir, N., Chen, D., Irwin, D., and Shenoy, P. Solar-TK: A Data-driven Toolkit for Solar PV Performance Modeling and Forecasting. In *MASS* (November 2019).

[45] Bashir, Noman, Chen, Dong, Irwin, David, and Shenoy, Prashant. Solar-tk: A data-driven toolkit for solar pv performance modeling and forecasting. In *2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)* (2019), IEEE, pp. 456–466.

[46] Bendib, B., Belmili, H., and Krim, F. A Survey of the Most Used MPPT Methods: Conventional and Advanced Algorithms Applied for Photovoltaic Systems. *Renewable and Sustainable Energy Reviews 45* (May 2015).

[47] Beyer, H.G., Costanzo, C., and Heinemann, D. Modifications of the Heliosat Procedure for Irradiance Estimates from Satellite Images. *Solar Energy 56*, 3 (1996).

[48] Cano, D., Monget, J., Albuisson, M., Guillard, H., Regas, N., and Wald, L. A Method for the Determination of the Global Solar Radiation from Meteorological Satellite Data. *Solar Energy 37* (1986).

[49] Cardwell, D. New York Times, Solar Power Battle Puts Hawaii at Forefront of Worldwide Changes, April 18th 2015.

[50] Chakraborty, P., Marwah, M., Arlitt, M., and Ramakrishnan, N. Fine-grained Photovoltaic Output Prediction using a Bayesian Ensemble. In *AAAI* (July 2012).

[51] Chandler, David L. Phys.org, Vast amounts of solar energy radiate to the earth, but tapping it cost-effectively remains a challenge. `https://phys.org/news/2011-10-vast-amounts-solar-energy-earth.html`, October 26th 2011.

[52] Chen, D., Breda, J., and Irwin, D. Staring at the Sun: A Physical Black-box Solar Performance Model. In *BuildSys* (November 2018).

[53] Chen, D., and Irwin, D. Black-box Solar Performance Modeling: Comparing Physical, Machine Learning, and Hybrid Approaches. In *Greenmetrics* (June 2017).

[54] Chen, D., and Irwin, D. SunDance: Black-box Behind-the-Meter Solar Disaggregation. In *e-Energy* (May 2017).

[55] Chen, D., Iyengar, S., Irwin, D., and Shenoy, P. SunSpot: Exposing the Location of Anonymous Solar-powered Homes. In *BuildSys* (November 2016).

[56] Chen, Kunjin, He, Ziyu, Chen, Kunlong, Hu, Jun, and He, Jinliang. Solar energy forecasting with numerical weather predictions on a grid and convolutional networks. In *2017 IEEE Conference on Energy Internet and Energy System Integration (EI2)* (2017), pp. 1–5.

[57] Cros, Sylvain, Liandrat, Olivier, Sébastien, Nicolas, and Schmutz, Nicolas. Extracting cloud motion vectors from satellite images for solar power forecasting. pp. 4123–4126.

[58] Delfanti, M., Merlo, M., Pozzi, M., Olivieri, V., and Gallanti, M. Power Flows in the Italian Distribution Electric System with Dispersed Generation. In *International Conference and Exhibition on Electricity Distribution* (June 2009), pp. 1–5.

[59] Ebert-Uphoff, Imme, and Hilburn, Kyle. Evaluation, tuning, and interpretation of neural networks for working with images in meteorological applications. *Bulletin of the American Meteorological Society 101*, 12 (2020), E2149–E2170.

[60] EIA. EIA projects renewables share of U.S. electricity generation mix will double by 2050. `https://www.eia.gov/todayinenergy/detail.php?id=46676`, 2021.

[61] Elvidge, Christopher D, Ghosh, Tilottama, Hsu, Feng-Chi, Zhizhin, Mikhail, and Bazilian, Morgan. The dimming of lights in china during the COVID-19 pandemic. *Remote Sensing 12*, 17 (2020), 2851.

[62] Frankel, D., and Wagner, A. Battery Storage: The Next Disruptive Technology in the Power Sector, June 2017.

[63] Freeman, J.M., DiOrio, N.A., Blair, N.J., Neises, T.W., Wagner, M.J., Gilman, P., and Janzou, S. System Advisor Model (sam) General Description (Version 2017.9. 5). Tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States), 2018.

[64] Gamarro, Harold, Gonzalez, Jorge E., and Ortiz, Luis E. On the Assessment of a Numerical Weather Prediction Model for Solar Photovoltaic Power Forecasts in Cities. *Journal of Energy Resources Technology 141*, 6 (03 2019). 061203.

[65] Goodfellow, Ian, Pouget-Abadie, Jean, Mirza, Mehdi, Xu, Bing, Warde-Farley, David, Ozair, Sherjil, Courville, Aaron, and Bengio, Yoshua. Generative adversarial nets. *Advances in neural information processing systems 27* (2014).

[66] Goodman, S., Schmit, Timothy J., Daniels, Jaime M., and Redmond, Robert J. *The GOES-R series: a new generation of geostationary environmental satellites*. Elsevier, 2020.

[67] Habib, Abdulelah H, Disfani, Vahid R, Kleissl, Jan, and de Callafon, Raymond A. Optimal energy storage sizing and residential load scheduling to improve reliability in islanded operation of distribution grids. In *American Control Conference (ACC), 2017* (2017), IEEE, pp. 3974–3979.

[68] Hao, Yan, and Tian, Chengshi. A novel two-stage forecasting model based on error factor and ensemble method for multi-step wind power forecasting. *Applied energy 238* (2019), 368–383.

[69] He, M., Reutzel, E., Jiang, X., Katz, R., Sanders, S., Culler, D., and Lutz, K. An Architecture for Local Energy Generation, Distribution, and Sharing. In *Energy2030* (November 2008).

[70] Hittinger, Eric, and Siddiqui, Jawad. The challenging economics of us residential grid defection. *Utilities Policy 45* (2017), 27–35.

[71] Hochreiter, Sepp, and Schmidhuber, Jürgen. Long short-term memory. *Neural computation 9*, 8 (1997), 1735–1780.

[72] Hunt, Tam. GreenTech Media, is there enough lithium to maintain the growth of the lithium-ion battery market?, June 20th 2015.

[73] IEEE Standard for Interconnecting Distributed Resources with Electric Power Systems. `https://standards.ieee.org/findstds/standard/1547-2003.html`, 2014.

[74] IEEE Standard for Interconnecting Distributed Resources with Electric Power Systems, Amendment 1. `https://standards.ieee.org/findstds/standard/1547a-2014.html`, 2014.

[75] Inman, Rich H, Pedro, Hugo TC, and Coimbra, Carlos FM. Solar forecasting methods for renewable energy integration. *Progress in energy and combustion science 39*, 6 (2013), 535–576.

[76] Iyengar, S., Sharma, N., Irwin, D., Shenoy, P., and Ramamkritham, K. SolarCast - A Cloud-based Black Box Solar Predictor for Smart Homes. In *BuildSys* (November 2014).

[77] Jean, Neal, Wang, Sherrie, Samar, Anshul, Azzari, George, Lobell, David, and Ermon, Stefano. Tile2vec: Unsupervised representation learning for spatially distributed data, 2018.

[78] Jing, Longlong, and Tian, Yingli. Self-supervised visual feature learning with deep neural networks: A survey. *IEEE transactions on pattern analysis and machine intelligence* (2020).

[79] Karpatne, Anuj, Ebert-Uphoff, Imme, Ravela, Sai, Babaie, Hassan Ali, and Kumar, Vipin. Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering 31*, 8 (2018), 1544–1554.

[80] Kasten, F., and Gzeplak, G. Solar and Terrestrial Radiation Dependent on the Amount and Type of Cloud. *Solar Energy 24*, 2 (1980).

[81] Keshav, S., and Rosenberg, C. How Internet Concepts and Technologies Can Help Green and Smarten the Electric Grid. *CCR 41*, 1 (January 2011).

[82] Khalilpour, Rajab, and Vassallo, Anthony. Leaving the grid: An ambition or a real choice? *Energy Policy 82* (2015), 207–221.

[83] Lago, Jesus, De Brabandere, Karel, De Ridder, Fjo, and De Schutter, Bart. Short-term forecasting of solar irradiance without local telemetry: A generalized model using satellite data. *Solar Energy 173* (2018), 566–577.

[84] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. Backpropagation applied to handwritten zip code recognition. *Neural Computation 1*, 4 (1989), 541–551.

[85] Lee, Stephen, Iyengar, Srinivasan, Feng, Menghong, Shenoy, Prashant, and Maji, Subhransu. DeepRoof: A Data-driven Approach for Solar Potential Estimation Using Rooftop Imagery. In *KDD* (August 2019).

[86] Lew, D., Bird, L., Milligan, M., Speer, B., Wang, X., Carlini, E., Estanqueiro, A., Flynn, D., Gomez-Lazaro, E., Menemenlis, N., Orths, A., Pineda, I., Smith, J., Soder, L., Sorensen, P., Altiparmakis, A., and Yoh, Y. Wind and Solar Curtailment. Tech. rep., National Renewable Energy Laboratory, October 2013.

[87] Liu, Qian, Liu, Wei, Sha, Dexuan, Kumar, Shubham, Chang, Emily, Arora, Vishakh, Lan, Hai, Li, Yun, Wang, Zifu, Zhang, Yadong, et al. An environmental data collection for COVID-19 pandemic research. *Data 5*, 3 (2020), 68.

[88] Lopes, R. Tonkoski L., and El-Fouly, T. Droop-based Active Power Curtailment for Overvoltage Prevention in Grid Connected PV Inverters. In *IEEE ISIE* (July 2010).

[89] Lorenz, Elke, Hammer, Annette, and Heinemann, Detlev. Short term forecasting of solar radiation based on satellite data. *EUROSUN2004 (ISES Europe Solar Congress)* (01 2004).

[90] Lorenz, Elke, and Heinemann, Detlev. Prediction of solar irradiance and photovoltaic power.

[91] Mathiesen, Patrick, and Kleissl, Jan. Evaluation of numerical weather prediction for intra-day solar forecasting in the continental united states. *Solar Energy 85*, 5 (2011), 967–977.

[92] Matthew J. Reno, Clifford W. Hansen, Joshua S. Stein. Global Horizontal Irradiance Clear Sky Models: Implementation and Analysis. Tech. rep., Sandia National Laboratories, March 2012.

[93] McGovern, Amy, Lagerquist, Ryan, Gagne, David John, Jergensen, G Eli, Elmore, Kimberly L, Homeyer, Cameron R, and Smith, Travis. Making the black box more transparent: Understanding the physical implications of machine learning. *Bulletin of the American Meteorological Society 100*, 11 (2019), 2175–2199.

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

[95] Mulkern, A. Scientific American, A Solar Boom So Successful, It's Been Halted, December 20th 2013.

[96] Myers, D. Cloudy sky version of bird's broadband hourly clear sky model. In *Annual conference of the American Solar Energy Society (SOLAR)* (July 2006).

[97] Nie, Yuhao, Sun, Yuchi, Chen, Yuanlei, Orsini, Rachel, and Brandt, Adam. Pv power output prediction from sky images using convolutional neural network: The comparison of sky-condition-specific sub-models and an end-to-end model. *Journal of Renewable and Sustainable Energy 12*, 4 (2020), 046101.

[98] NREL. NREL, News Release: NREL Report Shows Utility-Scale Solar PV System Cost Fell Nearly 30% Last Year. https://www.nrel.gov/news/press/2017/nrel-report-utility-scale-solar-pv-system-cost-fell-last-year.html, September 12th 2017.

[99] Oord, Aaron van den, Li, Yazhe, and Vinyals, Oriol. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748* (2018).

[100] Paletta, Quentin, and Lasenby, Joan. Convolutional neural networks applied to sky images for short-term solar irradiance forecasting, 2020.

[101] Penn, I. New York Times, California Will Require Solar Power for New Homes, May 9th 2018.

[102] Pillion, D. AL.com, Alabama Power Solar Fee Challenged as 'Unjust, Unlawful' in Complaint to PSC, April 26th 2018.

[103] Punyachai, R., Ongsakul, W., and Schmidt, U. Impact of High Solar Penetration on Voltage Profiles in Distribution Systems. In *International Conference and Exhibition on Green Energy for Sustainabile Development* (March 2014), pp. 1–7.

[104] Radford, Alec, Wu, Jeffrey, Child, Rewon, Luan, David, Amodei, Dario, Sutskever, Ilya, et al. Language models are unsupervised multitask learners. *OpenAI blog 1*, 8 (2019), 9.

[105] Ravuri, Suman, Lenc, Karel, Willson, Matthew, Kangin, Dmitry, Lam, Remi, Mirowski, Piotr, Fitzsimons, Megan, Athanassiadou, Maria, Kashem, Sheleem, Madge, Sam, et al. Skillful precipitation nowcasting using deep generative models of radar. *arXiv preprint arXiv:2104.00954* (2021).

[106] Raza, Muhammad Qamar, Nadarajah, Mithulananthan, and Ekanayake, Chandima. On recent advances in pv output power forecast. *Solar Energy 136* (2016), 125–144.

[107] Richardson, Lewis Fry. *Weather prediction by numerical process*. Cambridge university press, 1922.

[108] Rigollier, C., Lefevre, M., and Wald, L. The Method Heliosat-2 for Deriving Shortwave Solar Radiation Data from Satellite Images. *Solar Energy 77*, 2 (2004).

[109] Rolnick, David, Donti, Priya L, Kaack, Lynn H, Kochanski, Kelly, Lacoste, Alexandre, Sankaran, Kris, Ross, Andrew Slavin, Milojevic-Dupont, Nikola, Jaques, Natasha, Waldman-Brown, Anna, et al. Tackling climate change with machine learning. *arXiv preprint arXiv:1906.05433* (2019).

[110] Ryan, Greer. Throwing shade. `https://www.biologicaldiversity.org/news/press_releases/2018/rooftop-solar-04-25-2018.php`, April 2016.

[111] Schmit, Timothy J, Griffith, Paul, Gunshor, Mathew M, Daniels, Jaime M, Goodman, Steven J, and Lebair, William J. A closer look at the ABI on the GOES-R series. *Bulletin of the American Meteorological Society 98*, 4 (2017), 681–698.

[112] Schueler, Carl F, Lee, Thomas F, and Miller, Steven D. Viirs constant spatial-resolution advantages. *International Journal of Remote Sensing 34*, 16 (2013), 5761–5777.

[113] Shah, V., and Booream-Phelps, J. Deutsche Bank Markets Research, Crossing the Chasm. `https://www.db.com/cr/en/docs/solar_report_full_length.pdf`, February 27th 2015.

[114] Shahan, Z. Solar Panel Prices Continue Falling Quicker Than Expected, February 11th 2018.

[115] Shallenberger, K. UtilityDive, Massachusetts Gov. Baker Signs Bill Lifting Solar Cap, Lowering Net Metering Rates, April 11th 2016.

[116] Sharma, N., Sharma, P., Irwin, D., and Shenoy, P. Predicting Solar Generation from Weather Forecasts Using Machine Learning. In *SmartGridComm* (October 2011).

[117] Siddiqui, Talha A., Bharadwaj, Samarth, and Kalyanaraman, Shivkumar. A deep learning approach to solar-irradiance forecasting in sky-videos, 2019.

[118] Singh, Akansha, Lee, Stephen, Irwin, David, and Shenoy, Prashant. Sunshade: Enabling software-defined solar-powered systems. In *2017 ACM/IEEE 8th International Conference on Cyber-Physical Systems (ICCPS)* (2017), IEEE, pp. 61–70.

[119] Sun, Yuchi, Venugopal, Vignesh, and Brandt, Adam. Short-term solar power forecast with deep learning: Exploring optimal input and output configuration. *Solar Energy 188* (08 2019), 730–741.

[120] Swanson, R. A Vision for Crystalline Silicon Photovoltaics. *Progess in Photovoltaics: Research and Appliations 14*, 5 (August 2006).

[121] Thirugnanam, K., Kerk, S. K., Yuen, C., Liu, N., and Zhang, M. Energy management for renewable microgrid in reducing diesel generators usage with multiple types of battery. *IEEE Transactions on Industrial Electronics 65*, 8 (Aug 2018), 6772–6786.

[122] Thomson, P. PRI's The World, When the Grid says 'No' to Wind and Solar Power, this Company's Technology helps it say 'Yes' Again, April 7th 2015.

[123] Tiwari, Soumya, Sabzehgar, Reza, and Rasouli, Mohammad. Short term solar irradiance forecast using numerical weather prediction (nwp) with gradient boost regression. In *2018 9th IEEE International Symposium on Power Electronics for Distributed Generation Systems (PEDG)* (2018), pp. 1–8.

[124] Toms, Benjamin A, Barnes, Elizabeth A, and Ebert-Uphoff, Imme. Physically interpretable neural networks for the geosciences: Applications to earth system variability. *Journal of Advances in Modeling Earth Systems 12*, 9 (2020), e2019MS002002.

[125] Tonkoski, R., and Lopes, L. Impact of Active Power Curtailment on Overvoltage Prevention and Energy Production of PV Inverters Connected to Low Voltage Residential Feeders. *Renewable Energy 36* (June 2011).

[126] Tonkoski, R., Lopes, L., and El-Fouly, T. Coordinated Active Power Curtailment of Grid Connected PV Inverters for Overvoltage Prevention. *IEEE Transactions on Sustainable Energy 2*, 2 (April 2011).

[127] Tonkoski, R., Lopes, L., and Turcotte, D. Active Power Curtailment of PV Inverters in Diesel Hybrid Mini-grids. In *IEEE Electrical Power and Energy Conference* (2009).

[128] Trabish, H. UtilityDIVE, smart Inverters: The Secret to Integrating Distributed Energy onto the Grid?, June 2014.

[129] Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, Uszkoreit, Jakob, Jones, Llion, Gomez, Aidan N, Kaiser, Łukasz, and Polosukhin, Illia. Attention is all you need. In *Advances in neural information processing systems* (2017), pp. 5998–6008.

[130] Vaughan, A. Solar Power Growth Leaps by 50% Worldwide Thanks to US and China, March 7th 2017.

[131] Vincenty, Thaddeus. Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey review 23*, 176 (1975), 88–93.

[132] Vincenzi, Stefano, Porrello, Angelo, Buzzega, Pietro, Cipriano, Marco, Fronte, Pietro, Cuccu, Roberto, Ippoliti, Carla, Conte, Annamaria, and Calderara, Simone. The color out of space: learning self-supervised representations for earth observation imagery. In *2020 25th International Conference on Pattern Recognition (ICPR)* (2021), IEEE, pp. 3034–3041.

[133] von Appen, J., Braun, M., Stetz, T., Diwold, K., and Geibel, D. Time in the Sun: The Challenge of High PV Penetration in the German Electric Grid. *IEEE Power and Energy Magazine 11*, 2 (March 2013), 55–64.

[134] Wang, Huaizhi, Lei, Zhenxing, Zhang, Xian, Zhou, Bin, and Peng, Jianchun. A review of deep learning for renewable energy forecasting. *Energy Conversion and Management 198* (2019), 111799.

[135] Weaver, J. Solar Rises to Nearly 2% of U.S. Generation in 2017. PV Magazine, February 28th, 2018.

[136] Zhang, Jinsong, Verschae, Rodrigo, Nobuhara, Shohei, and Lalonde, Jean-François. Deep photovoltaic nowcasting. *Solar Energy 176* (2018), 267–276.

[137] Zhao, Xin, Wei, Haikun, Wang, Hai, Zhu, Tingting, and Zhang, Kanjian. 3d-cnn-based feature extraction of ground-based cloud images for direct normal irradiance prediction. *Solar Energy 181* (2019), 510–518.