



**CALIFORNIA
ENERGY COMMISSION**



**CALIFORNIA
natural
resources
AGENCY**

Energy Research and Development Division

FINAL PROJECT REPORT

High-Fidelity Solar Power Monitoring and Forecasting for Utility- Scale Solar Farms

**Gavin Newsom, Governor
March 2020 | CEC-500-2020-010**

PREPARED BY:

Primary Authors:

Carlos F. M. Coimbra
Hugo T. C. Pedro
David P. Larson
Mengying Li
Jeremy Orosco

University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093

Contract Number: EPC-14-008

PREPARED FOR:

California Energy Commission

Silvia Palma-Rojas, Ph.D.

Project Manager

Jonah Steinbuck, Ph.D.

Office Manager

ENERGY GENERATION RESEARCH OFFICE

Laurie ten Hope

Deputy Director

ENERGY RESEARCH AND DEVELOPMENT DIVISION

Drew Bohan

Executive Director

DISCLAIMER

This report was prepared as the result of work sponsored by the California Energy Commission. It does not necessarily represent the views of the Energy Commission, its employees or the State of California. The Energy Commission, the State of California, its employees, contractors and subcontractors make no warranty, express or implied, and assume no legal liability for the information in this report; nor does any party represent that the uses of this information will not infringe upon privately owned rights. This report has not been approved or disapproved by the California Energy Commission nor has the California Energy Commission passed upon the accuracy or adequacy of the information in this report.

PREFACE

The California Energy Commission's Energy Research and Development Division supports energy research and development programs to spur innovation in energy efficiency, renewable energy and advanced clean generation, energy-related environmental protection, energy transmission and distribution and transportation.

In 2012, the Electric Program Investment Charge (EPIC) was established by the California Public Utilities Commission to fund public investments in research to create and advance new energy solutions, foster regional innovation and bring ideas from the lab to the marketplace. The California Energy Commission and the state's three largest investor-owned utilities—Pacific Gas and Electric Company, San Diego Gas & Electric Company and Southern California Edison Company—were selected to administer the EPIC funds and advance novel technologies, tools, and strategies that provide benefits to their electric ratepayers.

The Energy Commission is committed to ensuring public participation in its research and development programs that promote greater reliability, lower costs, and increase safety for the California electric ratepayer and include:

- Providing societal benefits.
- Reducing greenhouse gas emission in the electricity sector at the lowest possible cost.
- Supporting California's loading order to meet energy needs first with energy efficiency and demand response, next with renewable energy (distributed generation and utility scale), and finally with clean, conventional electricity supply.
- Supporting low-emission vehicles and transportation.
- Providing economic development.
- Using ratepayer funds efficiently.

High-Fidelity Solar Power Monitoring and Forecasting for Utility-scale Solar Farms is the final report for the High-Fidelity Solar Power Forecasting Systems for the 392 MW Ivanpah Solar Plant (CSP) and the 250 MW California Valley Solar Ranch (PV) project (CEC-EPC-14-008) conducted by University of California San Diego. The information from this project contributes to the Energy Research and Development Division's EPIC Program.

For more information about the Energy Research and Development Division, please visit the [Energy Commission's research website](http://www.energy.ca.gov/research/) (www.energy.ca.gov/research/) or contact the Energy Commission at 916-327-1551.

ABSTRACT

This project introduced novel forecasting methods for the solar resource and solar power generation by addressing critical gaps in the current forecasting methods. The project focused on technologies that rely on solar irradiance such as concentrated solar power, concentrated photovoltaics and tracking photovoltaics. Power generation for these technologies relies on direct normal irradiance, which is the component of irradiance that requires the most specific development due to high sensitivity to cloud cover and aerosol content in the atmosphere. The critical need for accurate direct normal irradiance forecasting tools is evident by the relative scarcity of native forecasting algorithms for this irradiance component in the scientific literature and its absence from most numerical weather prediction models.

Researchers developed tools to monitor cloud cover that cause solar irradiance fluctuations across the solar plant field. The project team designed and used a new generation of low-cost solar instruments for monitoring and forecasting the solar resource for utility-scale solar farms. The team designed the instruments to provide reliable low-cost telemetry for real-time forecasts and covering detailed patches of the solar field nonintrusively.

The project team created several forecasting models depending on the forecast horizon (such as intrahour, day-ahead) and operational target (for example, reducing monthly imbalances) for two utility-scale solar farms in California. The project demonstrated that the forecasting accuracy for direct normal irradiance and plane-of-array irradiance across all time scales is improved using forecasting models that blend local and remote telemetry. The project results demonstrated the importance of having a rich set of input data to improve forecasting. This effect was especially important for the intrahour time scale, for which the use of multiple solar sensors, sky images, and high-resolution satellite images was found to be essential to reducing forecasting errors.

This report outlines the development and performance of the models. Furthermore, the report also addresses the applicability of the tools to other solar projects of different sizes and weather patterns.

Keywords: DNI, POA irradiance forecasting, solar generation forecasting, CSP, PV tracking, sensor network

Coimbra, Carlos F. M., Hugo T. C. Pedro, David P. Larson, Mengying Li and Jeremy Orosco. 2020. *High-Fidelity Solar Power Monitoring and Forecasting for Utility-Scale Solar Farms*. California Energy Commission. Publication Number: CEC-500-2020-010.

TABLE OF CONTENTS

	Page
PREFACE	i
ABSTRACT	ii
TABLE OF CONTENTS	iii
LIST OF FIGURES	v
LIST OF TABLES	vii
EXECUTIVE SUMMARY	1
Introduction.....	1
Project Purpose.....	1
Project Approach.....	2
Project Results.....	3
Recommendations.....	4
General Conclusion.....	4
Technology/Knowledge Transfer/Market Adoption	5
Benefits to California	5
CHAPTER 1: Introduction	7
CHAPTER 2: Project Approach	9
Model Development.....	9
CHAPTER 3: Testbeds and Data	11
Testbeds.....	11
Grid-Connected Solar Plants Greater than 100 Megawatts.....	11
Medium-Sized Solar Plants (About 1 Megawatt).....	11
Data Sources	12
Telemetry from Solar Farms.....	12
Sensor Network.....	13
Sky Images.....	13
Satellite Images	13
Numerical Weather Prediction	14
CHAPTER 4: Wireless Sensor Network.....	16

Design Requirements	16
Design Components	16
Design Iterations.....	18
Central Receiver Node	19
Deployment at CVSR	19
Phase 1	20
Phase 2	20
Phase 3	21
WSN Performance	21
Weatherproofing	21
Energy Usage.....	21
Wireless Communication	21
Data Storage.....	21
Data Quality.....	22
WSN Cost Reduction.....	23
CHAPTER 5: Resource and Generation Forecasting	24
Intrahour Forecasting	24
DNI Forecast for Ivanpah.....	24
POA Irradiance Forecast for CVSR	27
PO Forecast for CVSR	28
Intraday Forecast.....	32
Day-Ahead Forecast	35
Real-Time Implementation.....	38
Applicability to Other Solar Projects.....	40
A Database Infrastructure to Implement Real-Time Solar and Wind Power Generation Intrahour Forecasts	41
Assessment of Machine Learning Techniques for Deterministic and Probabilistic Intrahour Solar Forecasts.....	41
Direct Power Output Forecasts from Remote Sensing Image Processing	42
CHAPTER 6: Short-Term Forecast for Market Preparedness	43
Results	46
Model Training and Development	47

CHAPTER 7: Wind Surface Forecasting	52
CHAPTER 8: Technology/Knowledge/Market Transfer Activities.....	59
Solar Forecasting	59
Low-Cost Sensors.....	59
Technical Reports and Publications.....	59
General Outreach	61
Conference Presentations	61
American Geophysical Union (AGU) Fall Meeting.....	61
International Heat Transfer Conference	62
American Meteorological Society (AMS) Conferences and Symposiums	62
2018 Electric Program Investment Charge (EPIC) Symposium	63
CHAPTER 9: Benefits to Ratepayers	64
CHAPTER 10: Conclusions/Recommendations	66
GLOSSARY AND ACRONYMS	68
REFERENCES	70
APPENDIX A: FORECASTING TOOLS	A-1
APPENDIX B: PAPER ABSTRACTS.....	B-1

LIST OF FIGURES

	Page
Figure 1: Grid-Connected Solar Plants	11
Figure 2: Sky Images at Ivanpah.	13
Figure 3: Satellite Images for CVSR and Ivanpah.	14
Figure 4: NAM Nodes Near Ivanpah and CVSR.....	15
Figure 5: WSN Components and Deployment.....	17
Figure 6: Using the WSN at CVSR	20
Figure 7: Correlation Between Measured PO and WSN Irradiance Data.....	22
Figure 8: Correlation Between PO Irradiance and PO (Circuit 8).....	23

Figure 9: Schematic for the Intrahour DNI Forecast Based on Sky Classification.....	25
Figure 10: CMIP Data Versus POA Ground Measurements.....	29
Figure 11: Robustness Analysis With Ground Telemetry Features.....	31
Figure 12: Robustness Analysis with Ground Telemetry and Satellite Features	31
Figure 13: Comparative Mean Performance for Ground and Ground + Satellite Features	32
Figure 14: A Diagram of the Overall Forecast Method Using Satellite Images.....	33
Figure 15: Example of the POA Irradiance Forecast for Several Days.....	38
Figure 16: Project Website	39
Figure 17: Snapshot of the Real Time Day-ahead DNI Forecast for Ivanpah’s Tower 1 40	
Figure 18: California ISO hour-ahead market timeline.	44
Figure 19: Forecast Imbalances for April to November 2016	44
Figure 20: Hybrid Model Selection.....	48
Figure 21: Model Selection	49
Figure 22: Forecast Imbalance Comparison	50
Figure 23: Power Reduction due to High Wind	52
Figure 24: Histograms for Wind Data	53
Figure 25: NAM Wind Speed and Wind Gust Forecast.....	54
Figure 26: Example of kNN Correction for Wind Forecast.....	55
Figure 27: KNN Wind Forecast Errors as a Function of the Hour of the Day	57
Figure 28: Histogram Comparison for the Measured and Forecasted Wind Gusts.....	57
Figure 29: Measured and Forecasted Wind Gust for Days with High Wind.....	58
Figure 30: Dynamic TV Display of the Project Data and Results	61
Figure A1: Single Axis Tracking Geometry.....	A-2
Figure A2: Power Output Clear-Sky Model for CVSR	A-4
Figure A3: KNN Model Schematic	A-7
Figure A4: Timescale Resolution Analysis.....	A-9
Figure A5: Memory Feature Set Performance Analysis	A-11

LIST OF TABLES

	Page
Table 1: Components of Each WSN Device	19
Table 2: Clear-Sky Criteria Threshold Values for GHI and DNI.....	26
Table 3: Performance of Intrahour DNI Forecast for Ivanpah.....	27
Table 4: Forecast Performance of Intrahour POA Irradiance Forecast for CVSR.....	28
Table 5: Forecast Performance for Intraday Forecast for DNI and POA Irradiance	34
Table 6: Forecast Performance for the Day-Ahead DNI Forecast for Ivanpah	36
Table 7: Forecast Performance for the Day-Ahead POA Irradiance Forecast for CVSR..	37
Table 8: Error Metrics for the Forecasted Mean Wind Speed and Wind Gust	56

EXECUTIVE SUMMARY

Introduction

Solar energy is an abundant and renewable resource that can be harnessed to meet California's demand for electrical energy and reduce the electric grid carbon footprint. Mandates like the Renewables Portfolio Standard as well as decreasing costs for solar panels have accelerated the use of systems that convert solar energy into useful electrical energy.

As more variable renewable generation such as solar and wind comes on-line, the probability of unwanted and disruptive fluctuations in voltage and frequency increases for the electric power grid. Integrating solar resources successfully into the grid requires coordination among solar farm managers, utility companies, and independent system operators. In California, such efforts resulted in operational strategies like the requirements for solar plants to provide day-ahead generation forecasts for scheduling. It has also resulted in new market structures, such as the Western Energy Imbalance Market. This is a recent real-time bulk power trading market that allows the broader dispatch of power to consumers in western states in case of renewable overproduction. The success of these measures relies, in large part, on using forecasting models for this variable generation. Solar and wind forecasts are the lowest-cost, enabling technologies for addressing grid reliability and stability when there is a large amount of variable renewable electricity generation. Forecasting is also critical to reducing the need to overbuild energy storage or keep high levels of spinning resources online. These forecasts allow effective real-time management of supply. Various prediction models for solar power generation developed in recent years have improved forecasting accuracy, but despite these efforts, the state-of-the-art solar forecasting models still lack the accuracy to optimally integrate the levels of renewable resources required to meet California's energy and climate goals.

Project Purpose

Weather—particularly cloud cover, ground wind speeds, and atmospheric aerosols—can suddenly and significantly affect power generated by centralized solar farms. Therefore, in the past years, solar forecasting, has emerged as one of the main tools to deal with weather-related solar generation variability. Accurate forecasts reduce requirements for reserves, ancillary resources, and energy storage systems. Such "insurance" expenses associated with resource variability increase the cost of adding solar energy resources to the electric power grid. Increasing the accuracy of solar forecasting models for variable generation helps California to achieve statutory goals for renewable energy generation. In addition, high-fidelity forecasts decrease the likelihood of curtailment from large central plants substantially, thus increasing the effective amount of renewables. High-fidelity forecasts are useful for power plant siting, planning, and operations; utility scheduling; and real-time system operations and dispatching.

While forecasting models for fixed photovoltaics (PV) technologies are relatively well developed, the same is not true for directional components of the solar irradiance (the power per unit area received from the Sun in the form of radiation) that are critical for concentrating solar technologies, such as concentrated solar power or concentrated photovoltaics. In these technologies, sunlight is concentrated to increase the radiant energy available for conversion into electricity. Power generation for concentrated and tracking technologies relies on the direct normal irradiance and plane-of-array components of the total irradiance. These components are substantially more difficult to forecast because of the associated sensitivity to cloud and aerosol transmittance effects. The small number of irradiance forecasting algorithms published in the scientific literature readily illustrates the unavailability of forecasting models for these quantities.

This project addressed critical technological gaps in forecasting that rely on direct normal irradiance and plane-of-array irradiance, and it substantially increased the existing knowledge and resources available to continued improvement of solar forecasting. The major technological needs addressed in this project are high-fidelity direct normal irradiance and plane-of-array irradiance forecasting models for time horizons ranging from 5 minutes to 48 hours, and accurate resource-to-power models for utility-scale central power plants.

In addition to improving modeling and software, this project designed, manufactured, and installed low-cost, nonintrusive, scalable, and autonomous solar sensors for analyzing solar variability within utility-scale solar fields. The researchers used this low-cost hardware to accurately monitor solar irradiance and provide data for the spatially resolved forecasting algorithms. The nonintrusive aspect of this solution is essential, as physical modifications to the infrastructure (such as digging trenches for power and Ethernet connections) would discourage the deployment of these sensors in grid-connected solar farms.

The improvements developed by this project promote the integration of new concentrated solar power (CSP) power plants in areas of high direct normal irradiance resource such as the U.S. Southwest, including the southeast and Central Valley portions of California. Furthermore, the tools developed in this project apply to any solar harvesting system and are suitable for addressing issues related to the uncertainty of solar energy at several spatial and temporal time scales.

Project Approach

This project used accurate forecasting tools for two main variables: the solar irradiance, also called the solar resource, and the solar plant power output. It also considered several ancillary or additional variables (such as ground wind speed, temperature, humidity, and atmospheric aerosols) as inputs. The forecast horizons considered range from 5 minutes to 48 hours. In general, the models used evolutionary machine learning as a multilayered framework capable of incorporating and processing diverse forms of data inputs to generate adaptive predictions enhanced for each solar microclimate.

Much of the research developed in this project directed efforts to the 392 megawatt (MW) Ivanpah Solar Energy Generation Systems (Ivanpah), one of the largest CSP plants in the world. The portion of the project dedicated to Ivanpah focused on developing and validating tools capable of monitoring and forecasting the local solar resource accurately so that resource-to-power models are suitable for operation, scheduling, and market participation. These are models that estimate the plant's production as a function of the measured or forecasted solar irradiance. The accuracy of the resource-to-power models is especially critical and challenging during times when the solar resource is highly variable due to the presence of cloud and aerosol effects. The proposed forecasting tool considers the position of the sun, clouds, and solar plant such that the forecast of solar resources is accurate even when weather conditions induce high variability for the solar resource. This complex system sought to reduce uncertainties associated with operating, regulating, and scheduling for solar plants. The researchers used lessons learned from the Ivanpah phase of the project to produce a forecast system for the single-axis 250 MW California Valley Solar Ranch. This forecast system was an important project component given that single-axis PV tracking plants are common in the United States. The model is appropriate as the foundation for building similar models of other single-axis tracking PV plants and was adapted for fixed-PV plants.

The success of the project is measured by increasing levels of accuracy of the existing forecasting system for all horizons, using standard error metrics (mean absolute error, mean bias error, root mean square error, and forecasting skill) and ramp rate, ramp intensity, and ramp-following metrics. Ramp metrics are important, because they quantify the forecast's ability to predict weather-dependent steep changes (the ramps) in the solar resource. In this evaluation, ramp rate and intensity quantify the magnitude of the irradiance change whereas ramp-following metrics quantify the forecast's ability to anticipate ramps at the correct time.

A second research direction in this project is the design, fabrication, and testing of a low-cost wireless sensor network. This design is a solution for the lack of direct solar resource measurements often encountered in larger grid-connected solar farms.

Project Results

Overall, the project demonstrated that the forecasting accuracy for direct normal irradiance and plane-of-array irradiance across all time scales (intrahour, intraday, day-ahead) is improved using forecasting models that blend local and remote telemetry. The project results demonstrated the importance of having a rich set of input data to improve forecasting. This effect was especially important for the intrahour time scale, for which the use of multiple solar sensors, sky images, and high-resolution satellite images was found to be essential to reducing forecasting errors.

The project team used the tools developed in this project in two testbeds (Ivanpah and California Valley Solar Ranch) to predict direct normal irradiance, point of array

irradiance, and power generation. In both cases, the forecasting accuracy improved relative to baseline models. Furthermore, the researchers successfully applied those models to other locations, such as a 1MW PV solar farm in southern California. This showed that these models are of general applicability and suitable for other grid-connected solar farms.

In addition, the team demonstrated a successful design of a low-cost, nonintrusive, scalable, and autonomous wireless sensor network. The team designed and built the wireless sensor network at the University of California, San Diego and installed it at California Valley Solar Ranch. A yearlong test of the wireless sensor network revealed that the proposed design meets the success criteria for data accuracy, data availability and data coverage. It also revealed an important limitation, namely, the need of good line-of-sight between nodes. Nevertheless, the proposed wireless sensor network can be readily adopted by other solar projects with minimal changes.

Recommendations

Based on the work performed during this project, the project team identified a few research areas for renewable energy forecasting with the potential of promoting higher penetrations of carbon-neutral, weather-dependent renewable resources into the California power grid. These recommendations include:

- The inclusion of detailed analysis of resource and generation predictability on the solar assessment for plant siting.
- A study of scenarios with multiple penetrations of solar and wind in different microclimates and how new solar and wind capacity in these regions can contribute to offset atmospheric carbon emissions.
- The integration of new satellite imagery provided by the new generation of geostationary operational environmental satellite in solar resource and generation forecasts.
- The development of smart and hybrid remote-sensing/numerical weather-prediction models that cover the entire state of California in real time for the short-term prediction (0-6 hours) of solar and wind resources.

General Conclusion

The main project conclusions are the following:

- The wireless sensor network designed for this project was found to be a low-cost solution for obtaining spatially resolved irradiance over solar farms.
- Accurate direct normal irradiance forecasts are more difficult to obtain than plane-of-array irradiance forecasts. Rapidly changing weather patterns and cloud formation/dissipation affect direct normal irradiance more than plane-of-array irradiance.

- Higher forecasting skill for intrahour power output predictions require careful and innovative feature engineering - a process that extracts new data from the original raw values. The new features can then be used as predictors to in machine learning tools, such as XGBoost.
- The tools developed in this project are suitable to be easily and effectively implemented in other solar projects.
- Successful forecasting of potentially dangerous wind conditions for heliostats and PV panels can be achieved based on numerical weather prediction outputs that are calibrated against historical ground telemetry using machine-learning tools.

Technology/Knowledge Transfer/Market Adoption

The project team developed algorithms that address gaps in solar harvesting technologies that rely on direct normal irradiance and plane-of-array irradiance. These technologies include CSP, concentrated PV, and tracking PV solar plants. The project team continues to promote the forecasting algorithms through journal papers, conference presentations, and direct communication with researchers and industry.

High-resolution monitoring of solar irradiance over a solar farm is frequently prevented by the sparsity of local telemetry available. Often just a few irradiance sensors monitor a solar field with areas in the tens of square kilometers. The low-cost, autonomous, and wireless sensor developed in this project can be easily used over a large solar field. The team collaborated with the industry at California Valley Solar Ranch to test the sensors in a real-world scenario.

The models developed in this project relate closely to the overall goal of a U.S. Department of Energy (U.S. DOE) Solar Forecasting II funding program in which the research team is also participating. The U.S. DOE project aims to obtain significant forecasting error reductions for direct normal irradiance and global horizontal irradiance forecasts for intraday and day-ahead temporal horizons. Lessons learned in the Energy Commission's project will contribute to substantially reducing the risk in accomplishing the goals for the U.S. DOE Solar Forecasting II program. Moreover, solar resource and generation forecasting is moving from point forecasts to probabilistic forecasts. The former provide a single value for the prediction, whereas the latter also provide a probabilistic score for the likelihood of the forecast (similar to the cone of uncertainty used in hurricane forecasting). These models often rely on large ensembles of point forecasts. The models developed in this work naturally find a place in this framework, and the team will use the acquired knowledge in developing new probabilistic forecasts.

Benefits to California

The research team successfully demonstrated tools to increase the forecasting reliability for CPS and PV tracking solar generation. These tools have the potential to simplify and accelerate the integration of carbon-free technologies into the existing California power grid and market structure. Accurate solar generation forecasting has the potential to

help ensure a reliable power supply and higher quality power and enable the growth and integration of grid-connected solar energy resources into the electric grid. The project team identified several areas that will benefit from this project:

- Enhanced capacity of utility-scale CSP plants due to improved prediction of resource and power output.
- Improved integration of solar assets with utility and California Independent System Operator operations.
- Increased ability of solar plants to absorb short-term ramps and maintain solar production, leading to lower overall operation costs and lower consumer cost per solar kilowatt-hour (kWh).
- Decreased number of unscheduled power outages because of inaccurate forecasting of solar variability.
- Improved use of ancillary generation resources (for example, services dispatched to ensure that there is sufficient energy generation to meet load) including lower use of peak fossil fuel-powered plants.
- Improved public health from increased levels of renewable generation displacing the use of fossil fuel-based generation over the lifetime of higher-capacity solar plants.
- Reduced renewable energy curtailment and resulting lower consumer cost per solar kWh.

CHAPTER 1:

Introduction

Weather, particularly cloud cover, ground wind speeds and atmospheric aerosol, can suddenly and significantly affect power generated by centralized solar farms. The electricity market system requires higher commitment of ancillary or additional resources, storage systems and spinning reserves to offset the variability of the solar resource. Such “insurance” costs associated with resource variability compromise the increase in penetration of solar energy resources in the electric grid. Thus, there is a direct relationship between increasing the accuracy of solar forecasting models for variable generation and the ability of California to achieve statutory goals for renewable energy penetration. In addition, high-fidelity forecasts decrease the chance of forced outages from large central plants substantially, thus increasing the effective penetration of renewables. This project implemented the first large-scale, high-fidelity solar forecasting models for large capacity central plants in the world, which will directly improve the ability of utilities and grid regulators to incorporate the output of these plants with reduced ancillary costs and increased confidence.

This project introduced a new generation of forecasting methods for the solar resource (global horizontal irradiance (GHI); direct normal irradiance (DNI); and plane-of-array irradiance (POA)) and solar power generation (MW), which addresses critical gaps in state-of-the-art in solar forecasting. The project focused on technologies as concentrated solar power (CSP), concentrated photovoltaics (CPV) and tracking PV. Power generation for these technologies relies either solely or strongly on the DNI, which is the most difficult component of the solar resource to predict. DNI is also the component of irradiance that requires the most specific development due to its high sensitivity to cloud cover and aerosol content in the atmosphere. This critical need is evident by the relative scarcity of native DNI forecasting algorithms discussed in the scientific literature, and the absence of DNI information from the majority of the numerical weather prediction (NWP) models. In contrast, forecast algorithms for GHI for fixed PV applications abound in the literature, and are not only well developed but also well-funded at state and federal levels.

A large portion of the applied research activities proposed in this project centered on the 392 MW Ivanpah Solar Thermal plant, currently the largest CSP plant in the world, and responsible for 30 percent of all the CSP capacity in the United States. The activities focused primarily on the development and validation of tools capable of monitoring and forecasting the local solar resource and the power generation of each of Ivanpah’s three towers accurately, from five minutes out to 48 hours in the future. In addition to developing accurate solar resource and ground wind forecasting, the comprehensive forecasting platform forecasts total power generation. The primary goal of this complex system was to reduce uncertainties associated with operation,

regulation and scheduling for the Ivanpah plant. A direct outcome of integrating DNI forecasting in the plant control system was that the DNI forecasts would be used as a feed-forward input to the Solar Field Controls System (SFINCS) that enables the heliostat field to position itself ahead of transients caused by weather variability. This integration enhances the ability of plant managers to keep the steam generation online during periods of weather transients that cause the plant to trip, thus maximizing generation capacity and improving grid stability.

Because PV solar farms are more common than CSP plants in the United States, this project also developed forecasting and RTP models for a large-scale, single-axis PV tracking plant. UCSD and NRG implemented and used an operational forecast system for the 250 MW California Valley Solar Ranch (CVSR).

Achieving the objectives of this project required close monitoring of cloud cover, aerosols that result in solar irradiance fluctuations across the solar plant field. Therefore, the project team installed a new generation of low-cost solar instruments recently developed at UCSD by the Coimbra Energy Group for the monitoring and forecasting of the solar resource for utility-scale solar farms. These instruments designed had the explicit goals of providing low-cost reliable telemetry for real-time forecasts, and covering detailed patches of the solar field non-intrusively.

In summary, this project aimed to:

1. Improve DNI and POA irradiance forecasting accuracy
2. Improve the power output (PO) forecasts for the Ivanpah CSP and CVSR plants
3. Install and test a new generation of forecasting systems at these two plants to enable better control strategies for partially cloudy conditions (allowing operators to “ride” the transients without tripping the plants, particularly at Ivanpah Solar Thermal)
4. Develop and install PO forecast models for Ivanpah and CVSR
5. Generalize the lessons learned to other/future CSP plants and to other solar technologies (CPV, PV), and
6. Assess the benefits of these new tools to the California ratepayers.

CHAPTER 2:

Project Approach

The success of this project required using accurate forecasting tools for two main variables: the solar resource (DNI, POA irradiance) and the solar plant power generation, plus several ancillary variables (ground wind speed, temperature, humidity, atmospheric aerosol, etc.). The forecast horizons range from five minutes out to 48 hours. The team built the new generation models on the comprehensive forecasting algorithms developed by the Coimbra research group. The models use evolutionary stochastic learning as a multi-layered framework capable of ingesting and processing diverse forms of data inputs in order to generate adaptive predictions that are optimized for each solar microclimate. The forecasting skill of these hybrid stochastic methods (ANN = Artificial Neural Networks; KNN = k-Nearest Neighbors, SVM = Support Vector Machines, etc.) depends on several free parameters: the ANN architecture (number of layers, number of neurons per layer, neuron connectivity), the number of neighbors and the distance function for KNNs, the selection of input variables and input preprocessing, among others. In the data rich scenario for the two testbeds, where irradiation, meteorological, and cloud data are available, it is not always evident which variables to include as inputs for the forecasting engine. To address this difficulty, the project team developed an approach where stand-alone forecasting engines evolved to be optimal regardless of the size of the parameter space. This is done with genetic algorithms as master controllers of the evolutionary method.

This general framework has been tried and tested for many different and widely diverse solar microclimates in the past eight years and has been the basis for over many scientific publications, including more than 20 journal papers in the last three years. This adaptive approach was applied to all forecast horizons to produce a fully integrated, concatenated forecasting system.

Model Development

The workflow to prepare a forecasting model was the following:

1. Feature development, selection, and assessment (i.e., feature engineering),
2. As a starting point, the team chose a statistical or machine learning (ML) model. This project considers models such as Ordinary Least Squares (OLS), Support Vector Regression, Artificial Neural Networks, etc.
3. Using historical data, the free parameters or hyperparameters are determined in the training stage. The data used in the training is known as the training dataset. The goal of this task is to tune the models' structure for least forecasting error.
4. Once the models are trained, they are tested using an independent dataset. That is, data that the models were not exposed to in the training stage. This dataset is

known as the testing dataset and it is used to assess the models' performance as they are exposed to new data. Error metrics for the testing data are very important because they anticipate the models' performance in real-time. The innovation of this project was mainly included in items 1 and 2 in the list.

CHAPTER 3:

Testbeds and Data

Testbeds

The tools and algorithms developed in this project were tested for several solar farms of different size and harvesting technology (e.g. CSP, PV, PV tracking). Two reasons motivate this approach. Firstly, CSP plants rely on DNI and the forecasting models for this variable are not as mature as the ones for GHI. Secondly, this allows showing that the work here presented can be replicable in many other solar projects.

Grid-Connected Solar Plants Greater than 100 Megawatts

For this scale, the project team collaborated with NRG that operates two large-scale solar farms in California: the 250 MW California Valley Solar Ranch and the 392 MW Ivanpah Solar Electric Generating Systems (Figure 1). Ivanpah Solar Thermal plant is one of the largest CSP plants in the world, and responsible for 30 percent of all the CSP capacity in the United States in 2015. Because photovoltaics (PV) solar farms are more common than CSP plants, this project also developed PO forecasting models for a large-scale, single-axis PV tracking plant. The solar farm used in this case was the 250 MW California Valley Solar Ranch.

Figure 1: Grid-Connected Solar Plants



Left: The 392 MW CSP Ivanpah Solar Thermal power plant. *Right:* The 250 MW tracking PV California Valley Solar Ranch (CVSR) power plant.

Photo Credit: NRG

Medium-Sized Solar Plants (About 1 Megawatt)

This project demonstrated the applicability of the models developed for Ivanpah and CVSR for other locations. The project team obtained several years' worth of irradiance and PO data from two PV installations in southern California. These are: Canyon Crest Academy (CCA) in San Diego and La Costa Canyon (LCC) in Carlsbad. Both plants have

non-tracking PV panels at a fixed 5° incline, with AC nameplate capacities of 1 MW each.

Data Sources

Once the testbeds were determined, the project team procured the data necessary for the forecasting algorithms. The approach for data collection aimed at creating a data-rich scenario that can be leveraged with machine learning algorithms to improve forecasting. The data used in this project consisted of time-series data for several variables (e. g. irradiance, power), images from local sky cameras and satellites, publicly available data, etc.

Telemetry from Solar Farms

The most important data set consisted of telemetry from the solar farms. These data are essential to train the forecasting algorithms and to validate the resulting models. In general, telemetry data are in the form of time series, that is, a value associated with a time stamp. The two utility-scale solar farms have several weather sensors and irradiance sensors that log data in to a PI server¹ from OSIsoft. The PI server collects, stores, and organizes data from the solar farms and can be queried to retrieve any data logged. For both solar farms, UCSD collaborated with NRG's IT personnel to establish automatic data retrieval. At Ivanpah, this project took advantage of a preexisting connection that transfers data from the PI server at Ivanpah to a MySQL database at UCSD via an intermediate virtual machine and an attendant task that runs every 5 minutes. Although this solution worked for Ivanpah, the team decided to follow a more direct approach for CVSR. To this end, UCSD acquired its own PI server to establish a PI-to-PI connection. Once the two servers were connected, data logged at CVSR was immediately available at UCSD.

Using these data handshakes, the project team collected a large set of time series for both locations:

- Power output from the three towers at Ivanpah and from the whole CVSR. At CVSR it was possible to obtain PO data from each one of the nine separate circuits that form the plant. These data were collected at minute and sub-minute resolutions.
- Irradiance data from multiple sensors. At Ivanpah, DNI (GHI) was available from nine (three) sensors, three (one) for each one of the three towers. For CVSR, POA irradiance data was collected from three weather stations. GHI data was also available from one of the weather stations.
- Weather data was also collected from the two locations. These include wind speed and direction, precipitation, humidity, pressure, etc.

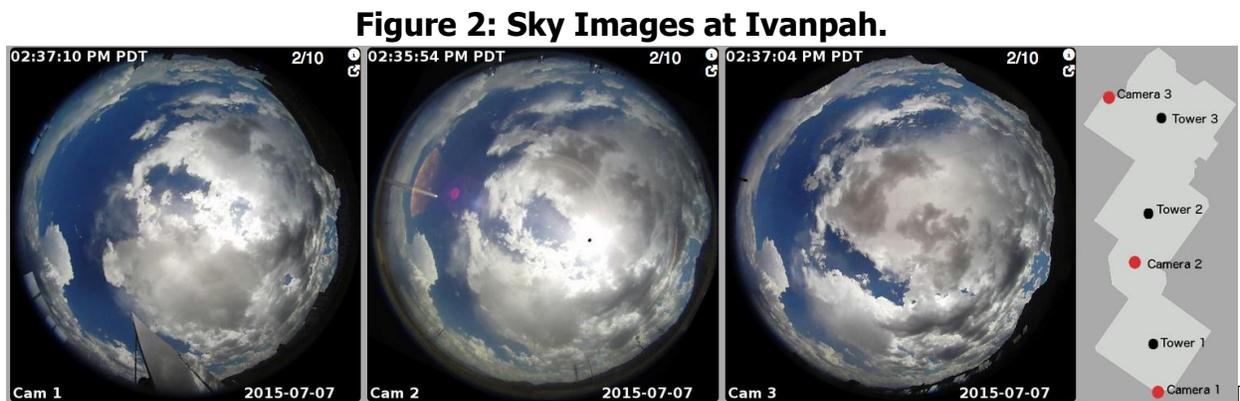
¹ <https://techsupport.osisoft.com/Products/PI-Server>

Sensor Network

The project team conducted an exploration of a dense network of irradiance sensors to improve PO monitoring and forecasting. To achieve this goal, the project team designed, built and tested a wireless sensor network (WSN). Using this tool, the project team obtained POA irradiance for ~20 nodes over CVSR. These data were recorded at 15 second resolution.

Sky Images

The short-term fluctuations in the solar irradiance, which include large and steep ramps, are dictated almost exclusively by cloud cover. A single optically thick cloud can bring the power output of a concentrated solar farm from full capacity to zero and back in a matter of minutes. Some of the intrahour models for DNI and POA irradiance use ground telemetry and sky images. The images are used to provide features, such as image color and texture, which are then used as predictors for a ML model. For this purpose, the project team installed four cameras at the two solar farms. Three cameras were installed at Ivanpah (one for each tower). Figure 2 shows one image per camera and a map of the cameras' location. The fourth camera was installed at CVSR on top of the Operations and Management (O&M) building. These cameras capture 3-megapixel (MP) sky images at 1-minute resolution. These are network cameras that were configured to transfer the images via File Transfer Protocol (FTP) to the UCSD servers in real-time.



Sky images from the three heliostats fields in Ivanpah. These images are processed to extract cloud cover information to generate the shortest-term intrahour DNI forecasts.

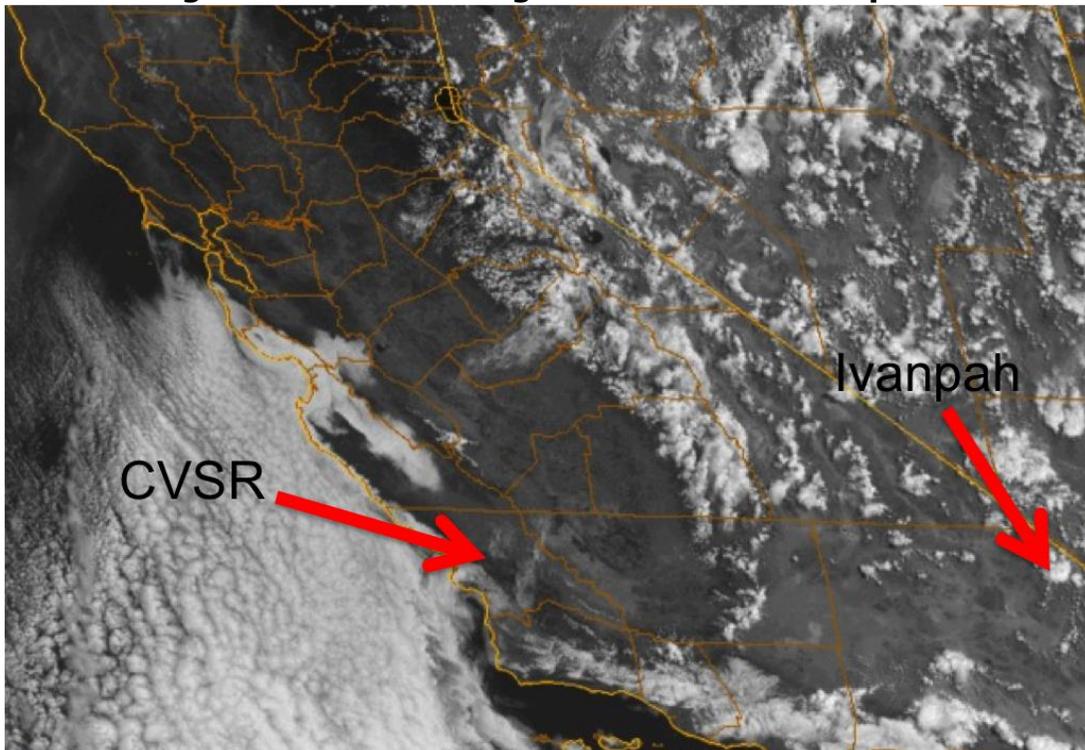
Source: Elaborated by the authors of University of California, San Diego (2019).

Satellite Images

For horizons longer than 30 minutes, sky cameras provide little information due to the limited field-of-view. In that case, the forecasting models capitalize on combining the strengths of Machine Learning (ML) for time series predictions with the relevance of information available in satellite images. To this end, the project team acquired images from the Geostationary Operational Environmental Satellite system (GOES), which are

operated by the United States' National Environmental Satellite, Data, and Information Service (NESDIS). More specifically, visible wavelength images from GOES-15, which is currently designated as GOES-WEST. The visible image channel is centered at $0.63 \mu\text{m}$, with a spatial resolution of 1 km and a temporal resolution of one image per 30 minutes. Figure 3 shows an example of one such image with the location of the two solar farms annotated.

Figure 3: Satellite Images for CVSR and Ivanpah.



An example of a satellite image obtained from the GOES-West website. Historical satellite images were obtained for the same period as the DNI ground data. Images such as this one will be processed to extract cloud cover information to assist in the intraday forecast for Ivanpah and CVSR.

Source: Elaborated by the authors of University of California, San Diego (2019). Satellite image from GOES-West.

Numerical Weather Prediction

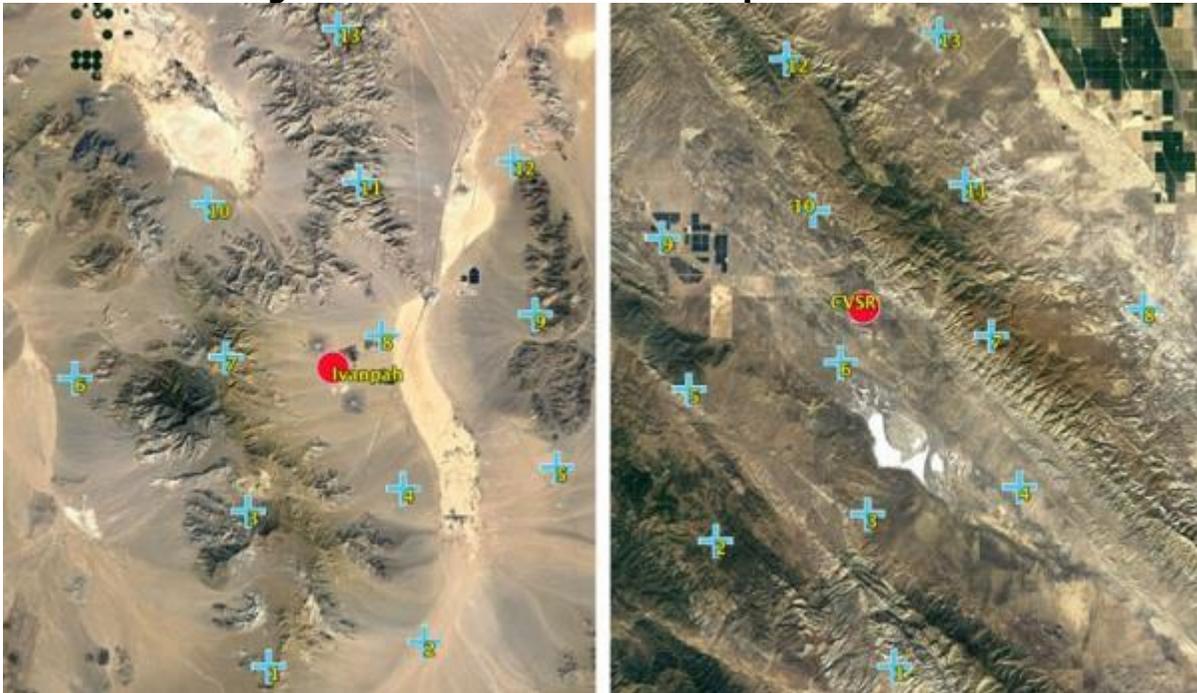
Forecasting models for several hours ahead to day(s) ahead require information about future weather conditions. For these horizons, ground telemetry and satellite images are not enough to anticipate irradiance and solar generation behavior. Thus, in order to have some idea of next day(s) weather, the models use publicly available information from numerical weather prediction (NWP) models. Specifically, data from the North American Mesoscale Forecast System (NAM). The NAM is an NWP model provided by the National Oceanic and Atmospheric Administration (NOAA) on a $12 \text{ km} \times 12 \text{ km}$ spatial grid that covers the continental United States. Forecasts are generated four

times daily at 00Z, 06Z, 12Z and 18Z, with hourly temporal resolution for 1 to 36-hour horizons and 3 h resolution for 39 to 84 h horizons. Several variables are extracted from the NAM files and used in the forecasts for this project:

- Downward shortwave radiative flux (DSWRF) [W/m²] at the surface, a synonym for GHI and denoted as $G_{NAM}(t)$;
- Total cloud cover (TCDC) [%] denoted as $C_{NAM}(t)$ where the entire atmosphere is treated as a single layer.
- Surface Wind speed and wind gusts forecasts.

These variables are obtained for the grid points in the vicinity of the locations of interest. Figure 4 identifies the nodes near Ivanpah and CVSR.

Figure 4: NAM Nodes Near Ivanpah and CVSR



The NAM forecasted data from these nodes is used in the day-ahead forecasts for these locations.

Source: Elaborated by the authors of University of California, San Diego (2019) using Google Earth.

CHAPTER 4:

Wireless Sensor Network

The devices developed under this project provide a solution to the need for high-density ground telemetry at low cost, with each unit costing less than \$500 in components. Each device is solar powered and communicates wirelessly, thereby removing the need for modifying or expanding the power and networking infrastructure of the site. Additionally, the wireless network topology and device design enable rapid reconfiguration of the sensor deployment, e.g., the addition of more device nodes. The sections below detail the design requirements that guided the development of the sensors and the design iterations from proof-of-concept to a final product and the field-testing operations at CVSR.

Design Requirements

The design of the wireless sensor network focused on five main requirements: low-cost, wireless, solar powered, automated and scalable.

Low-cost: Current options for irradiance measurement systems are prohibitively expensive for many interested parties. Prices for even the most basic systems start at several thousands of dollars, which restricts the number of systems installed at a site. By reducing the cost of the network, one can ensure that 1) the devices are affordable to be installed at more sites and 2) the number of devices installed increase at each site.

Wireless: Although all power plants have some form of standard internet connectivity, networking is rarely available across the entire site. Using wireless communication means it is possible to install devices at more locations without the need for additional network infrastructure.

Solar powered: As with networking, not all areas of a site will have power outlets. Powering the devices using solar opens up more options for where to install the devices and makes the devices robust to issues at the ground, e.g., maintenance work that disrupts power access.

Automated: Although power plants such as CVSR have staff on site, for forecasting applications, latency between data measurement and collection needs to be minimized. By automating the operation of the devices, latency is minimized.

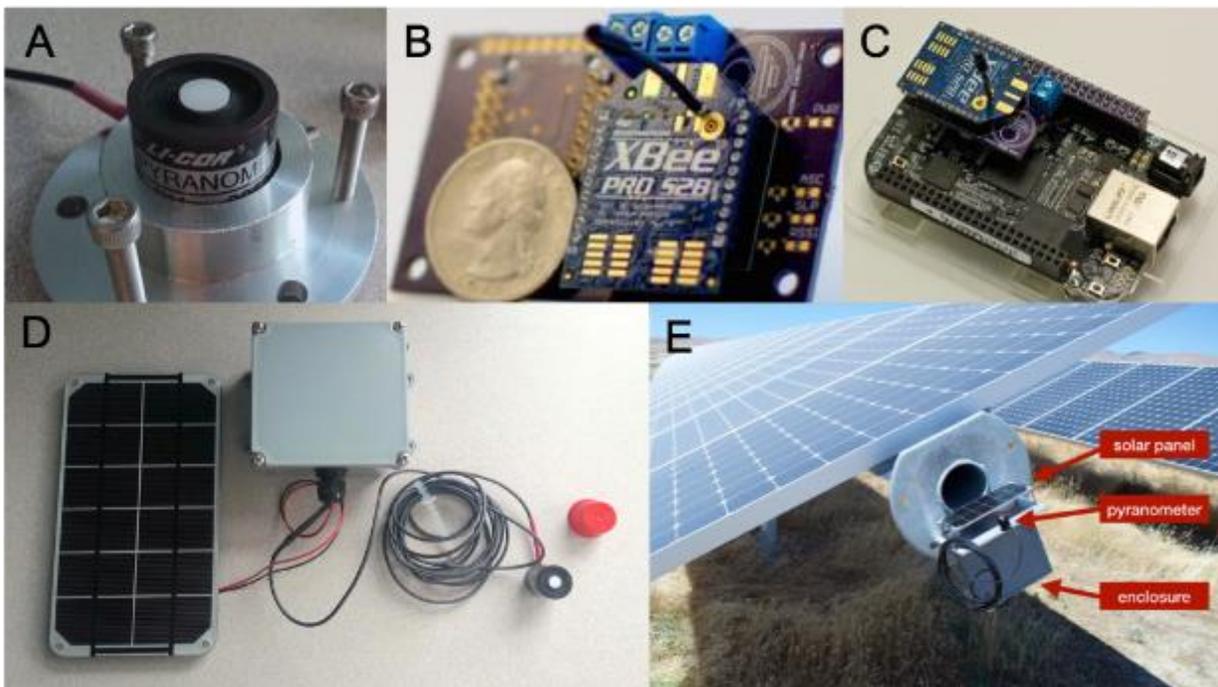
Scalable: A major goal of the project in developing a network of sensors is the ability to grow the density of sensors as needed. Therefore, the design of the network should minimize the difficulty in adding sensor nodes and enable a high density of sensors.

Design Components

The design requirements identified guided the choices in components for the system.

- Irradiance sensor: To balance cost versus performance, the WSN uses an off-the-shelf pyranometer. Specifically, the Licor LI-200, a photodiode-based pyranometer (see Figure 5A). Retailing for less than \$300, the LI-200 provides an economical option for irradiance measurement. Additionally, compared to other options, e.g., the Eppley PSP, the LI-200 is compact in size.
- Wireless communication: The distance between the sensor nodes and the central receiver is likely to exceed 100 meters. To meet this distance requirement, the sensor nodes use XBee-PRO ZigBee radio frequency (RF) modules from Digi. The XBee modules provide a maximum range of 3.2 km (2 miles) line-of-sight and data rates up to 250 Kbps. The modules have a standard serial data interface and can be programmed via USB with an adaptor board (Fig. 5B).

Figure 5: WSN Components and Deployment



(A) Licor LI-200 pyranometer with a custom aluminum mounting base. (B) final WSN PCB design, featuring a Global Positioning System (GPS) module for automatic positioning data and a 16-bit analog-to-digital (ADC) chip for high-resolution irradiance measurements. A quarter is included in the photo as a size reference. (C) The central receiver node, consisting of a BeagleBone Black mounted with a custom PCB (purple board) that provides an interface for the XBee wireless module (blue board; top left). (D) A WSN node ready to be installed at the field. (E) The sensor nodes attach to the PV tracking systems using pre-existing mounting points on the face plates at the end of each row. The weatherproof enclosure protects the electronics, while still providing easy access for maintenance and repair. The solar panel on top of the enclosure provides power to the node, but the panel is positioned so that it does not interfere with the pyranometer used for the POA irradiance measurements.

Photo Credit: University of California, San Diego (2019).

- **Computation:** Each sensor node requires a computing device to measure the irradiance and transmit the data via the XBee module. As the nodes only need to push data, a microcontroller is sufficient. Thus, the sensor nodes are equipped with the Adafruit Trinket line of microcontrollers (see Fig. 5B). The Trinket devices provide the same functionality and support as Arduino devices, but in a lower cost, size and energy usage format. Additionally, the Trinket devices are Arduino compatible and have the full support of Adafruit, a well-known electronics manufacturing company based in the United States.
- **Power:** The sensor nodes are powered using a combination of a photovoltaic (PV) panel and a backup battery. Due to low power requirements of the other components, the sensor nodes can be powered off of a single 3.4 W, 6V PV panel (approximately 4.72 in × 8.7 in, Fig. 5D). A 3.7 V, 2500 mAh Lithium Ion Polymer (LiPo) battery provides backup energy usage.

Design Iterations

After an initial proof-of-concept using commodity off-the-shelf (COTS) components, the design was converted to a printed circuit board (PCB) form-factor. The conversion to PCBs was necessary to further reduce costs, energy usage and size, while also enabling future mass production of the devices. The final version of the sensor is equipped with Adafruit Trinket Pro. This microcontroller provides additional input/output (I/O) options as compared to the Adafruit Trinket (the microcontroller used in the proof-of-concept devices). Additionally, to enable rapid debugging and repair, the PCB was designed such that the microcontroller and XBee module are mounted using header pins.

The final version, denoted as v2, has two main differences as compared to the original (v1) PCB: the addition of a Global Positioning System (GPS) module and a high-resolution (16-bit) analog-to-digital (ADC) chip. The GPS provides automatic positioning of the devices, which enables re-arranging of the physical locations of the devices without needing to manually log their information. By adding the 16-bit ADC, the devices can measure irradiance with greater precision, approximately 0.1 Wm^{-2} instead of the 10 Wm^{-2} from the microcontroller's built-in ADC.

Table 1 provides a breakdown of the costs of the WSN PCB. Note that the costs are approximate values, as prices decrease as with bulk purchases and manufacturing. Also, if a larger PV panel and/or battery becomes necessary, the cost of components included in the WSN devices will still be under \$500 per unit.

Table 1: Components of Each WSN Device

Component	Purpose	Approx. price
Pyranometer	irradiance sensor	\$200
Microcontroller	computation/control	\$10
XBee	wireless communication	\$30
PV panel	power	\$40
Battery	power	\$15
GPS	positioning data	\$40
Custom PCB	WSN base	\$5
Other components	WSN base	\$20
	Subtotal	\$360

“WSN base” refers the core portion of the WSN device, which ties together the entire system. Note that the costs are approximate values. Prices decrease for bulk orders and manufacturing.

Source: Elaborated by the authors of University of California, San Diego (2019).

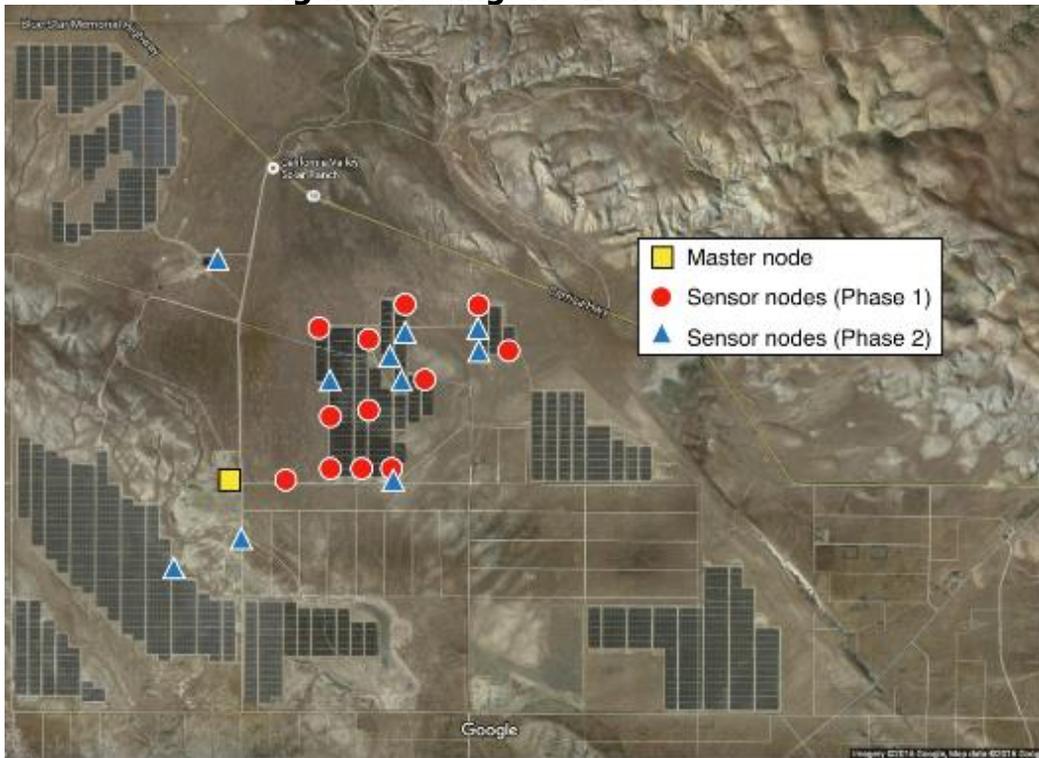
Central Receiver Node

In the system design, the network of WSN devices push their data to a central receiver node. The central receiver node is based on a BeagleBone Black minicomputer running a Linux operating system. A custom PCB provides an interface between the BeagleBone and an XBee wireless module, as well as an optional connection for a pyranometer. As data is received through the XBee wireless network, the data is logged to the filesystem of the BeagleBone. The data is then forwarded to a remote server or database at UCSD. The current configuration uses the BeagleBone's built-in ethernet port for external network connectivity.

Deployment at CVSR

The WSN field test deployment at CVSR occurred in three phases. Figure 5E shows one of the sensors installed in CVSR. To best assess the performance of the WSN design, the sensors were installed with a focus on the 40 MW Circuit 8 near the center of CVSR. Figure 6 shows the physical locations of the sensor nodes installed during Phases 1 to 3. The team installed the master node, which receives data from the sensor nodes, on the roof of the Operations and Management (O&M) building near the center of the site.

Figure 6: Using the WSN at CVSR



The master node was installed on the roof of the Operations and Management (O&M) building to ensure optimal wireless performance, while also enabling easy access to CVSR’s internal wired network. The sensor nodes markers denote locations where a sensor node was initially installed during either Phase 1 (red circle) or Phase 2 (blue triangle). To avoid confusion, locations where a node was replaced during Phase 2 and 3 are not shown.

Source: Elaborated by the authors of University of California, San Diego (2019) using Google Earth.

Phase 1

The first use of the WSN to CVSR occurred in July 2016. A total of 12 nodes were installed to CVSR, with a master receiver node installed on the roof of the O&M building. The sensor nodes were configured for 1 second sampling in order to benchmark the communication and operational performance of the WSN at CVSR.

Phase 2

In December 2016, the UCSD team visited CVSR and installed an additional 10 nodes to the WSN. Based on the analysis of data from the nodes installed in Phase 1, the sampling period was decreased from every 1 second to every 15 seconds. Additionally, the nodes were retrofitted to use external antennas with the XBee wireless transmitters, which helped improve the wireless performance and reliability. During this trip, the UCSD team also performed maintenance the previously installed sensor nodes and installed a sky camera to provide images of the sky for monitoring and forecasting applications.

Phase 3

The UCSD team traveled to CVSR in May 2017 to install software changes and perform maintenance on the pre-existing nodes. The software changes were related to improving the wireless performance of the XBee modules and to decreasing the energy consumption of the nodes. Specifically, the changes enabled the nodes to enter a low-power mode at nighttime, thus saving battery power for daytime cloudy periods. Sensor nodes from Phase 1 and 2 which did not exhibit energy consumption or wireless issues were not modified during this trip.

WSN Performance

Weatherproofing

Despite an uncommonly wet winter season in 2017, no sensor nodes have sustained weather-related damage. Moisture was discovered inside a single node during the May 2017 visit to CVSR, but the device was otherwise operating normally. Several nodes were dirtied by bird droppings and dust, but the contaminants did not penetrate inside the weatherproof enclosure and had a negligible effect on operations.

Energy Usage

On sunny or partly cloudy days, the PV panel on each node provides sufficient power for the on-board electronics. However, extended periods of sustained overcast weather (> 5 days), as observed during Winter 2017, led to some nodes experiencing occasional periods of power loss. In order to prevent loss of data during the early hours of the day, the nodes installed during phase 3 were configured to enter a low-power mode during the night.

Wireless Communication

The performance of the XBee wireless modules varied between nodes based on their position within the PV arrays. Part of the variation in performance was due to obstructions in the line-of-sight of the XBee modules, e.g., hills and PV panels. In an attempt to improve wireless performance, the nodes installed during phase 2 and 3 included external antennas for the XBee modules. Additionally, the density of sensor nodes was increased to enable additional multi-hop paths for the XBee wireless modules to reach the master node on the O&M building.

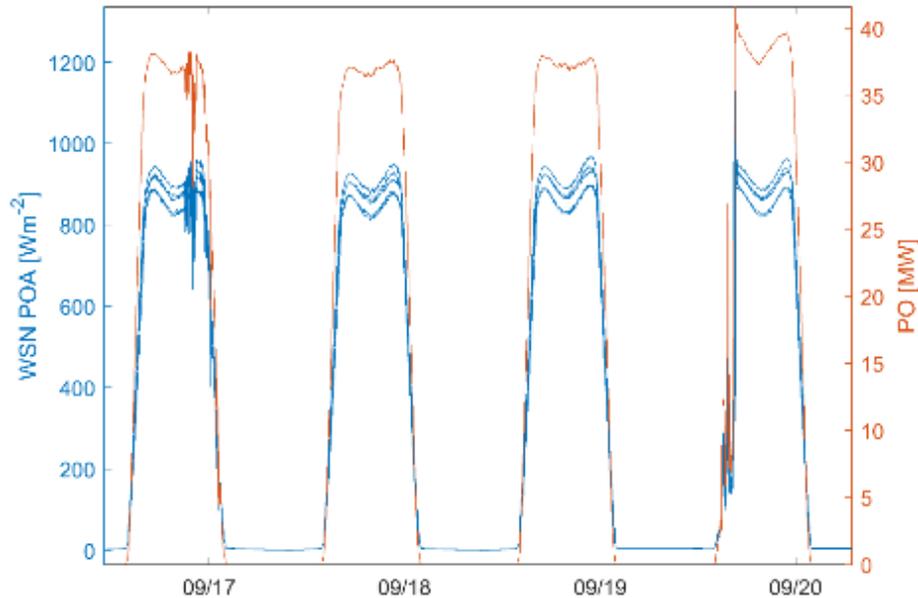
Data Storage

The WSN was designed such that the sensor nodes push their data at regular intervals to the master node, which collects, organizes and temporarily stores the data. Due to the design decisions outlined above, the individual sensors nodes do not have local backup storage, but the master node contains enough storage space to retain over 2 months of sensor node measurements. At the time of this report, the WSN master node has been operating for over two years and has not had a data storage related failure.

Data Quality

Figure 7 shows four days of measured power output (PO) from the 40 MW Circuit 8 and POA irradiance measurements from a subset of five of the WSN nodes. The POA from the WSN nodes exhibits the same dynamics as the PO time-series, while also providing a finer grain viewpoint on the intra-site variability. It should be noted that no post-processing was applied to the PO or POA time-series in the plot.

Figure 7: Correlation Between Measured PO and WSN Irradiance Data

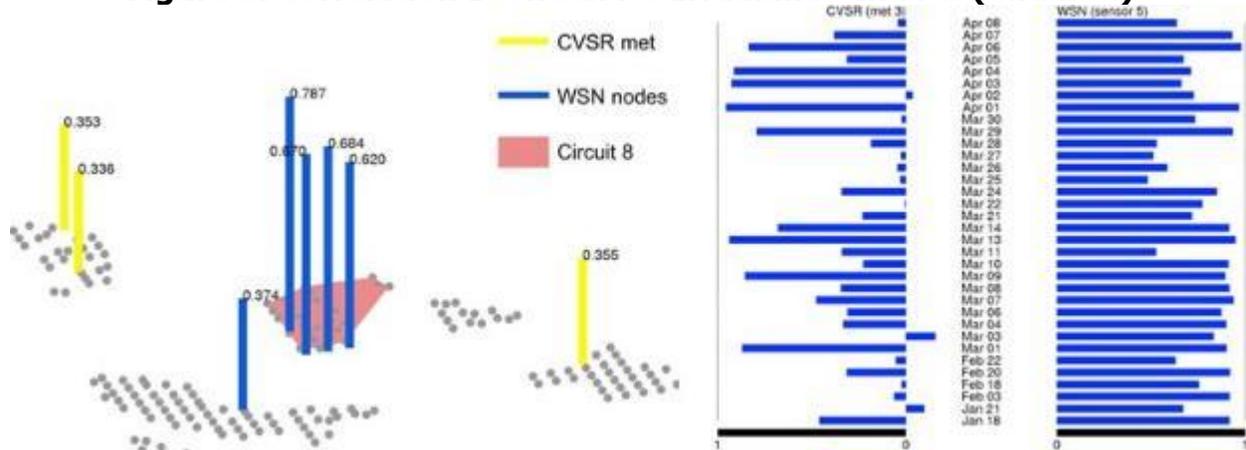


Sample time-series of the power output (PO) of the 40 MW Circuit 8 subset of CVSR compared to the plane-of-array (POA) irradiance measured by five WSN nodes. The POA irradiance measured by the WSN nodes exhibits the same dynamics as the PO time-series.

Source: Elaborated by the authors of University of California, San Diego (2019).

A study performed with data from 2017 revealed that the WSN POA irradiance data shows much higher correlation with Circuit 8 PO than any of the three POA measurements from the preexisting weather stations (installed and maintained by CVRS's personnel) as seen in Figure 8. The left panel in the figure shows that the correlation values between Circuit 8 PO and the POA irradiance from several WSN nodes (blue) are almost two times higher than the correlation values between power generation and POA irradiance from the weather stations. These results are a consequence of the distance from the sensors and Circuit 8 and demonstrate that a proper irradiance monitoring in large solar farms requires more than a few sensors. The right panel plots daily correlation values that again show much higher correlation between WSN POA irradiance and PO.

Figure 8: Correlation Between PO Irradiance and PO (Circuit 8)



Left: Correlation values between Circuit 8 power generation and POA irradiance measurements obtained from the WSN (blue) and the weather stations (yellow). The values for the WSN nodes nearer Circuit 8 show much higher correlation. **Right:** Correlation values for met. Station 3 and sensor 5 of the WSN.

Source: Elaborated by the authors of University of California, San Diego (2019).

WSN Cost Reduction

As stated, the goal for this task was to develop a low-cost solution for the acquisition of spatially resolved solar irradiance over the solar plant. Based on the values listed in Table 1, the deployment of 25 sensors over CVSRS amounted to approximately $\$500 \times 25 = \$12,500$ in equipment and materials. Commercial alternatives for a single sensor with datalogger, start at approximately \$1,000 and can run up to \$3,900 (Lave *et al* (2015)). Thus, the solution found in this project represents a cost reduction of at least 50 percent relative to commercial alternatives. This value is very conservative given that alternatives require both power and ethernet. The low-cost WSN does not, as reported above. Therefore, the deployment of commercial alternatives would be limited to points, in the power plant, where both are available, or, it would require significant investment to extend the power and ethernet infrastructure to the selected locations.

CHAPTER 5:

Resource and Generation Forecasting

Intrahour Forecasting

The intrahour forecast includes the shortest forecast horizons in this work, but it is the most challenging type of forecast in solar energy applications. This stems from the fact that at these very short horizons it is necessary to predict the average behavior of the solar irradiance, and also to predict large and sudden changes: the ramps. Given that, clouds moving in and out of the solar field are the key factor in determining these ramps, the intrahour models rely on sky images. These images are essential to track the cloud motion, estimate the shadowing of the solar field, and predict fluctuations in the solar generation.

DNI Forecast for Ivanpah

There are three power generation towers at Ivanpah, each of which has three DNI sensors and one fisheye sky camera. Two global horizontal irradiance (GHI) sensors are installed on Tower 1 and Tower 3. The high-resolution data was averaged into five-minute windows across the three towers. This operation results into a five minute-resolution time series of solar irradiance for the whole solar plant. Night values (solar elevation angle smaller than 0°) and abnormal measurements such as negative values that result from erroneous sensor readings were filtered out. Irradiance data for short periods with no measurements are interpolated from existing data.

The forecasting model forecasts the clear-sky index of DNI instead of DNI irradiance. The clear-sky index $k_t(t)$ is defined as the ratio of DNI and clear-sky DNI, $k_t(t) = B(t)/B_c(t)$, where $B(t)$ is the actual irradiance and $B_c(t)$ is the clear-sky irradiance. The clear-sky irradiance $B_c(t)$ is calculated from selected clear-sky models, which are presented in Appendix A.

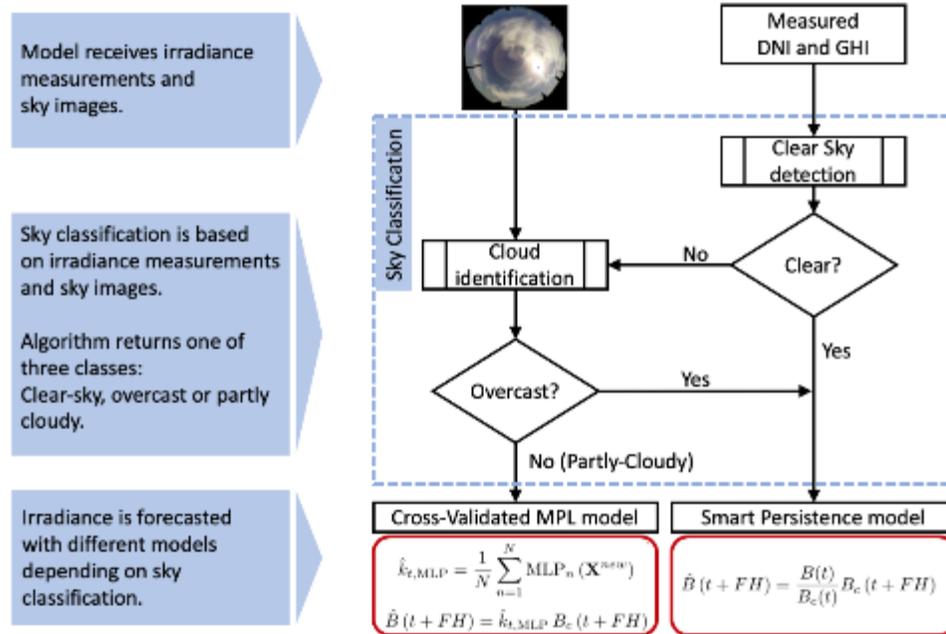
Different forecasting schemes are applied to different sky conditions (as shown in Fig. 9). For cloud-free (clear) and overcast periods, the smart persistence model is applied. For partly-cloudy periods, a pre-trained MLP is applied (Appendix A). The sky classification methods are presented below in more detail.

Sky Classification

The key component of this forecasting model is the accurate sky classification, in terms of cloud cover, that allows switching between the different sub-models. In this model the sky is classified as clear, partly-cloudy (patches of clear sky and isolated clouds) and overcast (sky mostly covered by clouds). The model first determines the sky class based on the latest irradiance data available, since clear-sky instances are easily identifiable by comparing the measured data against the clear-sky model. If a data section is classified as not clear the model uses the corresponding sky image to assess the cloud

coverage. Using that information these periods are classified as one of the two non-clear classes: overcast or partly-cloudy.

Figure 9: Schematic for the Intrahour DNI Forecast Based on Sky Classification



Different models are applied depending on cloud cover. The model produces the irradiance forecasts for different intrahour forecast horizons (FH).

Source: Elaborated by the authors of University of California, San Diego (2019).

Clear-Sky Detection

The clear-sky periods are identified based on a statistical model originally developed by Reno and Hansen (2012) for GHI observations. This method uses five criteria to compare a period of N GHI measurements to a corresponding clear-sky GHI for the same period. The time period is deemed *clear* if threshold values for all the five criteria are met. By expanding the work of Reno and Hansen (2012) to include thresholds for both GHI and DNI, the project team worked on a more reliable clear-sky detection model. This expansion results in a total of 10 criteria (5 for GHI plus 5 for DNI) that must be met for a period to be classified as clear sky. These criteria are applied to a 10-min sliding window as suggested by Reno and Hansen. The threshold values for GHI and DNI are listed in Table 2. A detail description of each criterion can be found in Reno and Hansen (2012).

Table 2: Clear-Sky Criteria Threshold Values for GHI and DNI

Criterion	GHI	DNI
Difference of mean value	100 W/m ²	200 W/m ²
Difference of max value	100 W/m ²	200 W/m ²
Difference of length of the line	50	100
Difference of the variance of step-changes	0.01	0.015
Maximum deviation from the clear-sky slope	10 W/m ²	15 W/m ²

Source: Elaborated by the authors of University of California, San Diego (2019).

Overcast Period Detection

Overcast periods are detected based on the sky images. From each image the algorithm extracts the red and blue color channels and computes the normalized red to blue ratio or NRBR. NRBR is calculated as $(R - B)/(R + B)$ where R and B denote the Red and Blue components of the RGB image. Tests have shown that overcast sky images have a unimodal histogram of NRBR with a relatively small standard deviation (Li *et al*/2011). Using this fact, a data point is denoted as overcast if the corresponding sky image has a standard deviation of NRBR smaller than 0.085.

Partly-cloudy Period Detection and Cloud Detection

Partly-cloudy periods are simply those periods that fail the tests described above for clear and overcast periods. However, for these periods it is very important to identify clouds from the sky images since they will determine the fluctuations in the irradiance. For this purpose, this project uses Minimum Cross Entropy method (MCE). The MCE is an adaptive image processing method based on a simple thresholding image segmentation algorithm (see Chu *et al* (2014)). The value of the MCE threshold is based on the histogram of the $(R - B)$ distribution of the sky image to achieve minimum cross entropy. Once the threshold is determined, pixels with higher $(R - B)$ than that value are classified as cloud and the remaining are classified as clear sky.

Using this cloud cover classification algorithm, the model describe in Fig. 9 is then applied to the data from Ivanpah. The forecasting metrics for the testing set are listed in Table 3.

Table 3: Performance of Intrahour DNI Forecast for Ivanpah (W/m²)

Forecast Horizon (mins)	SPM – MBE	SPM – MAE	SPM – RMSE	HM – MBE	HM – MAE	HM – RMSE	HM – Forecast Skill (%)
AP: 5	0.5	27.3	67.5	0.0	27.7	61.9	8.3
AP: 10	0.9	40.3	94.2	-1.0	39.8	83.1	11.8
AP: 15	1.4	47.6	109.3	-1.2	46.9	97.3	11.0
AP: 20	1.9	53.3	119.6	-2.3	53.5	108.0	9.8
CP: 5	-0.3	34.8	78.2	-1.0	35.6	70.3	10.1
CP: 10	-1.0	49.0	105.7	-4.2	48.3	88.6	16.2
CP: 15	-1.6	55.1	118.1	-5.9	54.0	98.8	16.4
CP: 20	-1.6	59.7	125.8	-8.8	60.0	106.4	15.4

AP = all periods; CP = cloudy periods; SPM = Smart Persistence Model; HM = Hybrid Model.

Source: Elaborated by the authors of University of California, San Diego (2019).

POA Irradiance Forecast for CVSR

Intrahour forecast models developed for POA irradiance for CVSR are based on meteorological data collected from the weather stations installed at the solar farm. These instruments collect POA irradiance, temperature and wind speed every minute. These variables are used to train a Support Vector Regression model (Appendix A) to predict POA irradiance 5, 10, 15, and 30 minutes ahead of time. The error metrics for this model and the persistence model applied to the testing set are listed in Table 4. The table shows good forecasting skills for all the intrahour horizons studied.

Table 4: Forecast Performance of Intrahour POA Irradiance Forecast for CVSR

Forecast: Model Var.	Forecast: FH	Error Metrics: MBE (Wm^{-2})	Error Metrics: MAE (Wm^{-2})	Error Metrics: RMSE (Wm^{-2})	Skill (%)
Pers. POA	5 min	2.60	35.70	84.27	-
SVR. POA	5 min	-11.12	56.73	68.73	19
Pers. POA	10 min	5.78	53.23	112.84	-
SVR POA	10 min	-9.81	64.82	77.12	32
Pers. POA	15 min	9.52	65.13	127.81	-
SVR POA	15 min	-14.86	65.53	79.17	38
Pers. POA	30 min	23.65	94.73	162.63	-
SVR POA	30 min	-7.67	72.28	83.98	49

Source: Elaborated by the authors of University of California, San Diego (2019).

PO Forecast for CVSR

Intrahour power output forecasts for CVSR were focused on the 40WM circuit 8 with the goal to participate in the intrahour energy market (power from other circuits is traded differently).

The baseline feature set used for the development and analysis of the intrahour forecasts is comprised of current and historical (lagged) values: POA irradiance measurements from telemetry station 2, and PO for circuit 8.

The fidelity realized by the intrahour forecasts depends critically on two enhancements to the baseline feature set that can be utilized both independently to achieve differing degrees of forecast performance. These are

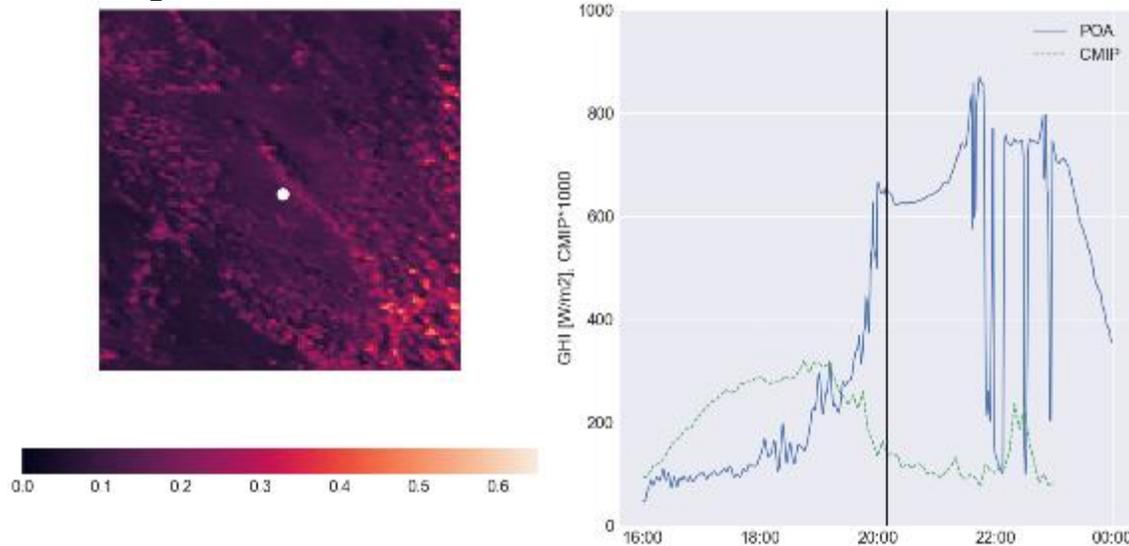
1. features derived from memory operators, and
2. satellite-derived features.

The former is obtained as numerically evaluated fractional derivatives of a corresponding time series (see Appendix A) while the latter are products of the Geostationary Operational Environmental Satellite (GOES) system. Each feature set is compiled as 5-minute backward (i.e., causal) averages of the raw data. The feature vectors include historical (i.e., lagged) values as well as a most-recent-minute column. The latter—a 1-minute causal average of the most recent raw data—is included because it was found to provide a significant benefit beyond the 5-minute averaged data, especially at shorter (<15 minute) horizons.

The satellite-derived feature sets are products of the GOES-West and GOES-East satellites. These features include:

- GHI derived from GOES-West images using the NOAA/NESDIS Operational Cloud Processing System (CLAVR-x) (Heidinger et al. (2014)) that are publicly available in real-time².
- Cloud and Moisture Imagery Product (CMIP)³ from the new satellite GOES-EAST. Figure 10 shows a CMIP map and it illustrates its relationship with POA at CVSR.

Figure 10: CMIP Data Versus POA Ground Measurements



Relationship between GOES-R CMIP data and POA data at CVSR. Left: the CMIP map over CVSR (white dot). Right: POA and CMIP time series for Jan 20, 2018. The black vertical line indicates the time at which the map on the left was captured. Large CMIP values indicate clouds over CVSR and low values indicate clear-sky. This relationship is clearly evident in the time series variations between 18:00 and 20:00 in Coordinated Universal Time (UTC).

Source: Elaborated by the authors of University of California, San Diego (2019).

A brief feasibility analysis immediately revealed that using satellite features for the node closest to CVSR does not provide any substantial benefit beyond the ground telemetry features. However, features derived from the first and second statistical moments of the nearest-geographical-neighbor (NGN) node distribution do yield a substantial benefit. This is an expected result, since the effect of the NGN data is to increase the topological “reach” of the model, which is realized in terms of an increase in performance at longer horizons.

² https://cimss.ssec.wisc.edu/clavrx/google_earth_main.html

³ https://www.goes-r.gov/products/ATBDs/baseline/Imagery_v2.0_no_color.pdf

The final model selected for forecasts in this task is that obtained when utilizing a machine learning technique known as gradient boosting, which is explained in more detail in Appendix A. The team selected the gradient boosting method for this work since the more traditional MPL structure struggled to extract benefits from some of the input data. It was observed that by placing multiple MPLs into a naïve boosting configuration, and then cascading the differing memory features in at each round, an independent performance increase was obtained for each fractional order. The gradient boosting technique is an obvious candidate for capitalizing on this result, since it has recently received great notoriety and wide adoption within the machine learning community due to its robust performance with both regression and classification tasks.

Robustness Analysis

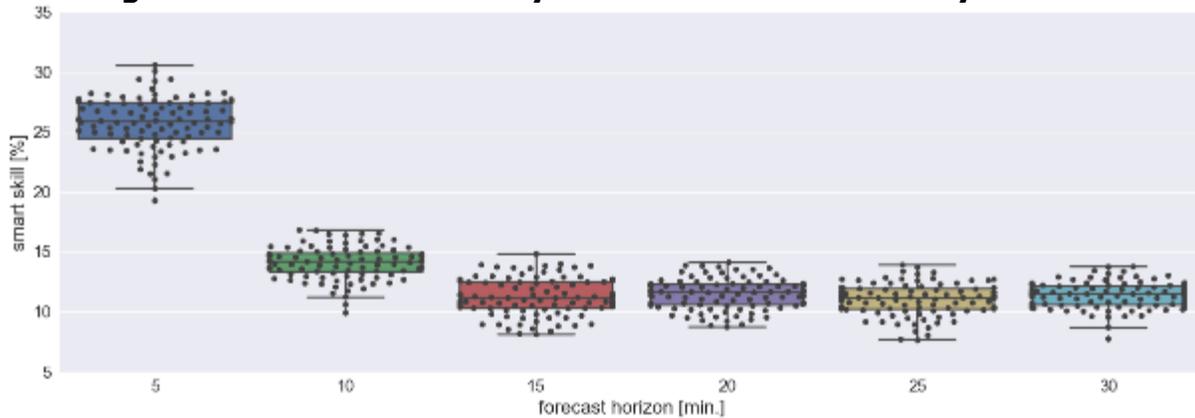
Since the forecasting fidelity is dependent on the model, which is itself dependent on input data having a stochastic component, the performance at each time horizon is best characterized as a distribution of skills. It is in this spirit the analysis of model robustness was undertaken.

To facilitate the analysis, distributions were developed in the following way:

1. Validation data set is randomly shuffled into two disjoint sets:
 - a. training (70 percent)
 - b. testing (30 percent)
2. Model is trained utilizing training data
3. Forecast skill is computed and stored
4. Steps 1-3 are repeated 100 times.

The resulting set of 100 forecast skills then represents a performance distribution associated with the validation data set. It should be noted that model hyperparameters were obtained on data that is entirely independent from the validation data. This analysis was performed at each time horizon. Figure 11 provides analysis results for models utilizing ground telemetry features and the corresponding memory operator features. Figure 12 provides analysis results for models utilizing ground telemetry features, satellite features, as well as memory operator features corresponding to both.

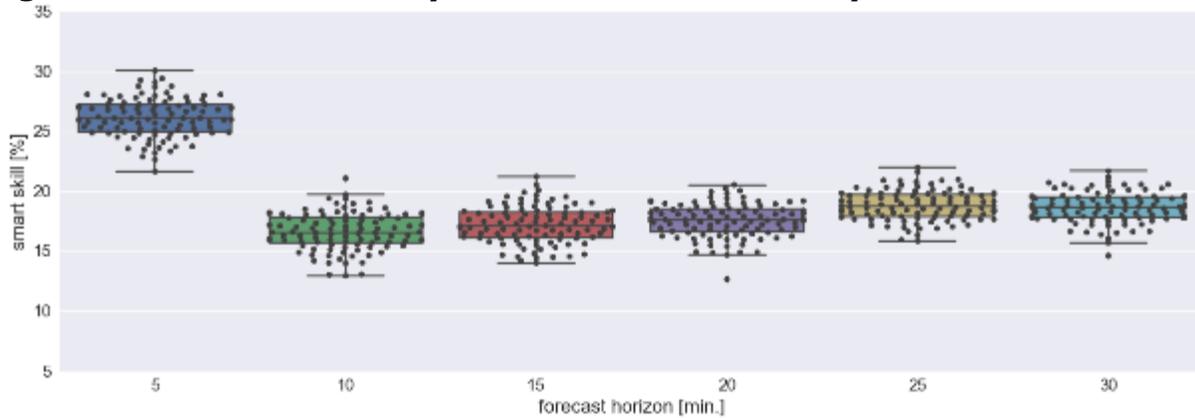
Figure 11: Robustness Analysis With Ground Telemetry Features



The box and swarm plots demonstrate performance distributions for forecasts with the XGB model structure utilizing ground telemetry and corresponding memory operator features. The forecast skill “levels off” at longer horizons, indicating that a fixed performance enhancement (i.e., beyond a smart persistence model) is realized at horizons greater than 10 minutes.

Source: Elaborated by the authors of University of California, San Diego (2019).

Figure 12: Robustness Analysis with Ground Telemetry and Satellite Features

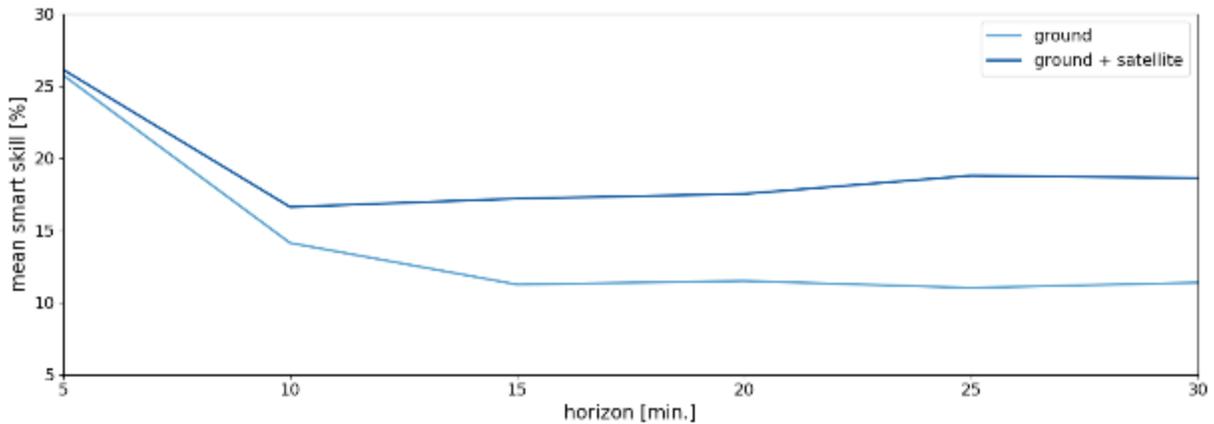


The box and swarm plots demonstrate performance distributions for forecasts with the XGB model structure utilizing ground telemetry features, satellite-derived features, and memory operator features corresponding to both. For horizons greater than 5 minutes, the results indicate a substantial additional benefit beyond those produced with ground telemetry.

Source: Elaborated by the authors of University of California, San Diego (2019).

A general trend that can be observed in Figures 11 and 12 is the improvement in forecast fidelity obtained at longer horizons when the satellite-derived features are included. This benefit is diminishing for the 5-10-minute horizons. In Figure 13, plotting the mean skills from the previous two figures has delineated this trend.

Figure 13: Comparative Mean Performance for Ground and Ground + Satellite Features



The plots of mean forecast skill (obtained from the distributions in the previous two figures) demonstrate the performance enhancement achieved with the use of satellite-derived feature sets. In each case, the corresponding memory operator features have also been used.

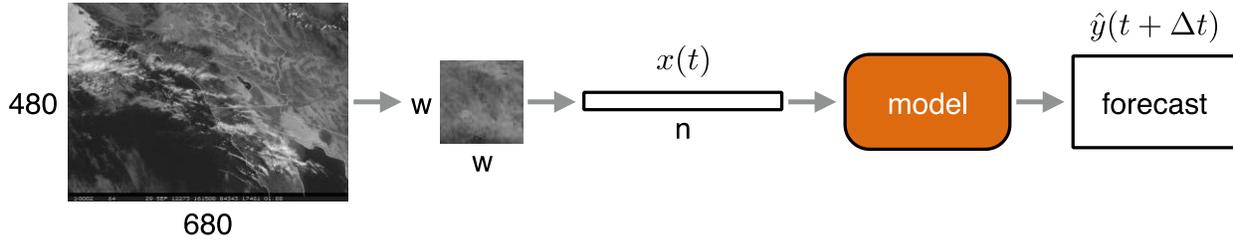
Source: Elaborated by the authors of University of California, San Diego (2019).

The trend observed in the figure makes sense, since the satellite-derived features provide NGN statistics for the temporal evolution of the weather system, which is expected to be useful primarily at longer horizons. If the information provided locally by ground telemetry is consistent with that provided by the satellite, then the performance augmentation should vanish at shorter horizons, which is precisely the empirical result. Furthermore, the mean skills at the 5-minute horizon were obtained from performance distributions resulting from independently randomized training-testing splits, so that the asymptotic agreement at shorter horizons also reinforces the consistency and robustness of the method used to optimize the hyperparameters.

Intraday Forecast

Similar to intrahour forecast the intraday solar forecast also targets the forecasting of large ramps in the solar resource. However, for these horizons the cloud information in sky images is too short-lived (sky images are useful for horizons below 30 minutes). An alternative to sky images that allows for a much larger field of view are the images acquired by the GOES-West satellites as explained in Chapter 3. Using these images the forecasting models learn a mapping $f(\circ)$ between a set of real-valued satellite-derived features at the current time t and ground conditions at some future time $t + \Delta t$ (Figure 14).

Figure 14: A Diagram of the Overall Forecast Method Using Satellite Images



First, a 480×680 pixel satellite image at time t is cropped down to a $w \times w$ region of interest, centered around the target site. The cropped image is then transformed into a n -length feature vector $x(t)$ and fed into a forecast model, e. g., SVR, to produce a prediction of the target output $\hat{y}(t + \Delta t)$.

Source: Elaborated by the authors of University of California, San Diego (2019).

To accomplish these goals, two mapping functions are considered: a linear model based on Ordinary Least-Squares (OLS) and a non-linear model based on Support Vector Regression (SVR) (Appendix A).

The satellite image inputs are derived from the GOES geosynchronous satellite system. A $w \times w$ square region, centered on the target site, is extracted from each image and then flattened into a vector \tilde{x} with $n = w^2$ elements. Each image is then normalized to remove brightness and contrast variations that depend on the time of the day at which the image was captured.

$$x = \frac{\tilde{x} - \text{avg}(\tilde{x})}{\sqrt{\text{var}(\tilde{x}) + 10}}$$

where \tilde{x} is the unprocessed 8-bit gray scale image vector, avg is the arithmetic mean value, and var is the variance. After normalization, the image vectors are stacked to create a matrix X that is used as input to the forecasting models.

The forecast models (OLS and SVR) are trained to predict backwards-averaged, hourly values for horizons of one to six hours. For each model, the optimal hyperparameters are chosen using cross-validation on the training set. The forecast performance of the models for both DNI at Ivanpah and POA at CVSR are summarized below (Table 5).

Table 5: Forecast Performance for Intraday Forecast for DNI and POA Irradiance

Forecast: Mod. Var.	Forecast: FH (hour)	Forecast: Loc.	Error Metrics: MAE (Wm^{-2})	Error Metrics: MBE (Wm^{-2})	Error Metrics: RMSE (Wm^{-2})	Skill (%)
OLS DNI	1	Ivp	173.5	26.1	222.1	35.2
SVR DNI	1	Ivp	144.5	-5.4	201.6	41.2
OLS DNI	2	Ivp	197.9	37.2	251.4	31.4
SVR DNI	2	Ivp	167.2	-2.2	228.0	37.8
OLS DNI	3	Ivp	216.9	49.7	272.4	31.9
SVR DNI	3	Ivp	188.4	4.2	248.7	37.8
OLS DNI	4	Ivp	233.0	59.4	284.5	34.2
SVR DNI	4	Ivp	205.1	14.0	263.5	39.1
OLS DNI	5	Ivp	240.9	66.0	289.9	37.6
SVR DNI	5	Ivp	216.9	25.1	272.9	41.3
OLS DNI	6	Ivp	243.8	71.8	290.6	41.7
SVR DNI	6	Ivp	222.9	36.5	275.0	44.9
OLS POA	1	CVSR	165.7	-15.0	211.2	5.1
SVR POA	1	CVSR	135.3	-12.7	184.5	17.1
OLS POA	2	CVSR	172.0	-25.2	220.2	25.7
SVR POA	2	CVSR	145.8	-18.0	198.1	33.1
OLS POA	3	CVSR	179.2	-36.1	230.3	36.4
SVR POA	3	CVSR	159.6	-51.9	217.9	39.9
OLS POA	4	CVSR	183.3	-34.5	237.7	42.9
SVR POA	4	CVSR	174.1	-43.2	235.1	43.5
OLS POA	5	CVSR	187.3	-38.3	242.7	47.2
SVR POA	5	CVSR	180.9	-50.3	244.4	46.8
OLS POA	6	CVSR	192.4	-42.4	246.8	51.0
SVR POA	6	CVSR	179.4	-59.0	243.0	51.7

Source: Elaborated by the authors of University of California, San Diego (2019).

For all OLS and SVR models, the input satellite images are cropped to 64×64 square pixel regions, centered on the target sites. The image size of 64×64 was found to be the optimal based on forecast performance metrics. Larger image sizes (such as 128×128) require more storage and computational time, but do not provide significant increases in forecast accuracy. The SVR-based forecasts achieve the overall best performance across the intraday horizons, with forecast skill scores up to 45 percent and 52 percent for Ivanpah and CVSR, respectively.

Day-Ahead Forecast

The last category of models developed in this task are models that predict the solar resource for the next day or days. The goal for these models is not to predict the solar irradiance to a fine granularity, instead the models predict longer trends in the solar resource. For instance, forecast what is the aggregated level of irradiance for the following day, the onset and length of cloudy periods for the next day, etc. Such predictions are very important to determine the expected power generation for the next day(s) as requested by the regulating entities whose job is to schedule resources to balance the electric grid in California.

The day-ahead forecast models use the NAM data and ground telemetry (Chapter 3) to predict hourly irradiance for forecast horizons ranging from 24 to 48 hours. Four different day-ahead forecasts models were tested in this project:

- Smart persistence model as defined in Appendix A.
- A non-linear model based on the NAM cloud cover forecast to predict DNI for Ivanpah defined as:

$$\hat{B}_{NL}(t + \Delta t) = (1 - C_{NAM,nearest}(t + \Delta t))^{0.76} B_c(t + \Delta t).$$

For CVSR the best non-linear model is obtained by using the NAM forecasted GHI:

$$\hat{P}_{NL}(t + \Delta t) = G_{NAM,nearest}(t + \Delta t)^{1.17} P_c(t + \Delta t).$$

In these equations the subscript *nearest* indicates the NAM node closest to the power plant. That is node six and eight for CVSR and Ivanpah, respectively (Fig. 4 in Chapter 3).

- An MLP forecast model similar to the one introduced above for the intrahour DNI forecast that uses as inputs the two NAM variables and the current irradiance measurement.

$$\hat{I}_{MPL}(t + \Delta t) = \text{MPL}(\vec{G}_{NAM}(t + \Delta t), \vec{C}_{NAM}(t + \Delta t), I(t)).$$

- An optimized nearest neighbor (KNN) forecast model that uses as inputs the two NAM variables for the NAM nodes in the vicinity of the solar farm.

$$\hat{I}_{KNN}(t + \Delta t) = \text{KNN}(\vec{G}_{NAM}(t + \Delta t), \vec{C}_{NAM}(t + \Delta t)).$$

In the MPL and KNN models *I* denotes DNI in the case of Ivanpah and POA irradiance in the case of CVSR. The vectors \vec{G}_{NAM} and \vec{C}_{NAM} contain the GHI and cloud cover NAM forecast for the all the nodes extracted.

The exponents in non-linear models and the MLP free parameters are determined using the training dataset for Ivanpah and CVSR. The KNN model has no free parameters and, as such, requires no training. All models are then validated with an independent test dataset. The error metrics for the forecast horizons ranging from 24 h to 48 h are listed in Table 6 for the DNI forecast for Ivanpah.

Table 6: Forecast Performance for Day-Ahead DNI Forecast for Ivanpah

Forecast: Mod.	Forecast: Var.	Forecast: FH (hour)	Error Metrics: MBE (Wm^{-2})	Error Metrics: MAE (Wm^{-2})	Error Metrics: RMSE (Wm^{-2})	Skill (%)
Pers.	DNI	24-27	34.19	227.29	337.72	-
NL	DNI	24-27	-70.48	147.09	225.46	33.2
MLP	DNI	24-27	30.11	142.75	198.17	41.3
Pers.	DNI	28-31	35.00	249.58	357.64	-
NL	DNI	28-31	-74.31	154.18	235.00	34.3
MLP	DNI	28-31	38.22	156.10	207.95	41.9
Pers.	DNI	32-35	45.04	232.11	325.12	-
NL	DNI	32-35	-103.50	165.57	237.96	26.8
MLP	DNI	32-35	42.13	160.31	208.11	36.0
Pers.	DNI	36-39	79.51	159.45	235.30	-
NL	DNI	36-39	-69.18	116.29	163.79	30.4
MLP	DNI	36-39	18.10	82.49	123.42	47.5
Pers.	DNI	40-43	83.75	212.73	313.74	-
NL	DNI	40-43	-54.35	121.00	182.31	41.9
MLP	DNI	40-43	19.98	99.51	155.49	50.4
Pers.	DNI	44-48	58.20	219.75	333.55	-
NL	DNI	44-48	-54.93	130.49	204.97	38.6
MLP	DNI	44-48	21.17	119.62	178.30	46.5

Source: Elaborated by the authors of University of California, San Diego (2019).

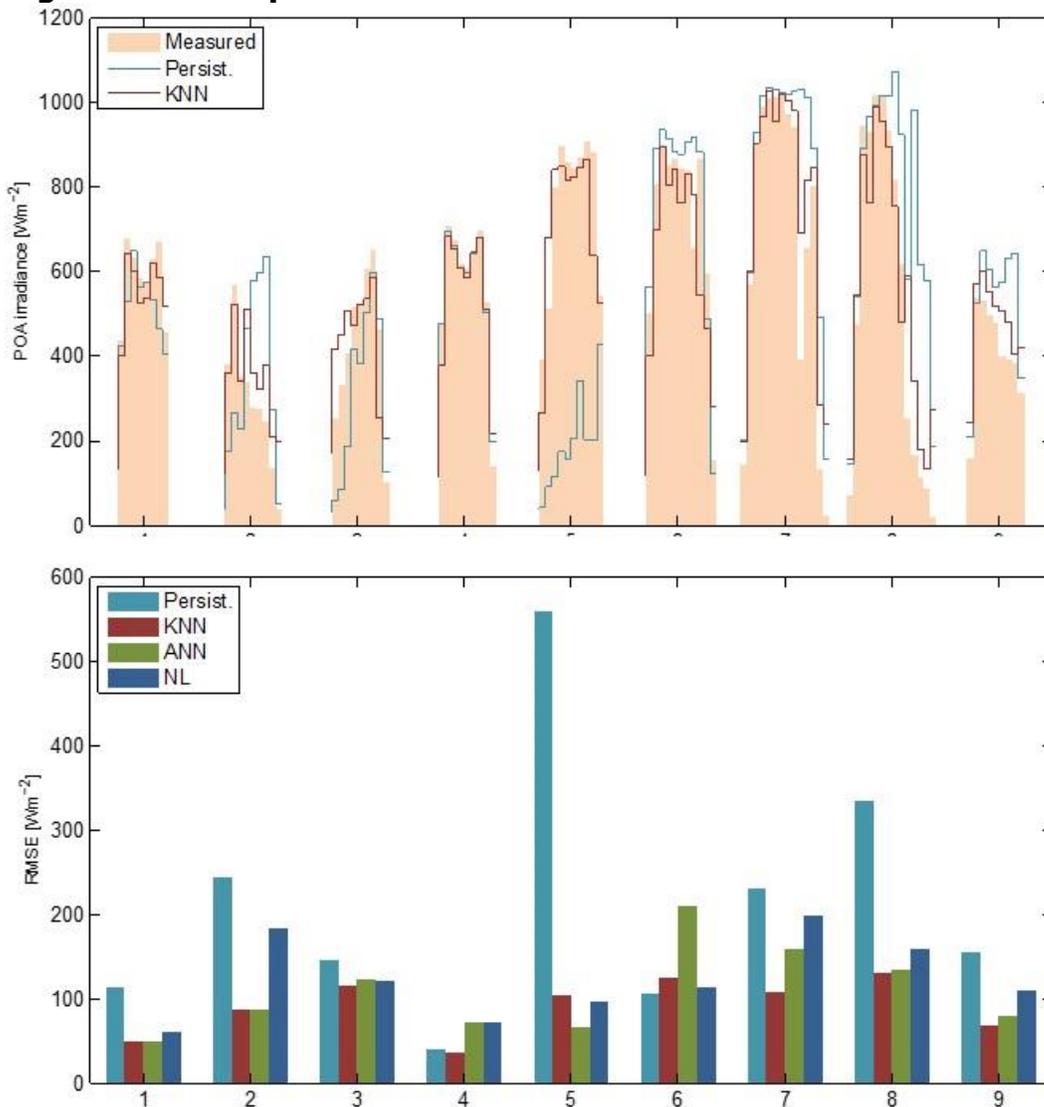
Table 7 shows the same results for the POA irradiance forecast for CVSR. The values in these tables show that substantial forecast skills (above 50 percent) are achieved by the machine learning models (MPL and KNN). Figure 15 (top) shows nine days in the testing set that illustrate the output from the Persistence model and the KNN model when predicting the POA irradiance for CVSR 24 hours ahead. The shaded area shows the measured data. The figure illustrates the benefits from the KNN model. For instance, the KNN model can predict well the overall behavior of the POA irradiance 24 hours ahead of time. Such performance is very useful for cloudy days, such as days eight and night. The persistence model, shows the expect behavior: it performs well for extended periods of clear sky but shows no forecast skill once the weather becomes unstable. Day five in the figure, illustrates well the poor behavior of the persistence model. The KNN model does not suffer from this issue since it uses the forecasted weather from NAM. For a quantitative analysis of the forecast performance, the bottom panel in Figure 15 shows the daily values of RMSE for the two forecast models and for the NL and MPL. This figure shows that the KNN model produces the best forecast for most days. The same information is also conveyed by Table 7.

Table 7: Forecast Performance for the Day-Ahead POA Irradiance Forecast for CVSR

Forecast: Mod.	Forecast: Var.	Forecast: FH (hour)	Error Metrics: MBE (Wm⁻²)	Error Metrics: MAE (Wm⁻²)	Error Metrics: RMSE (Wm⁻²)	Skill (%)
Pers.	POA	24-27	-3.19	147.82	245.56	-
KNN	POA	24-27	-0.11	58.02	98.37	59.94
MLP	POA	24-27	-.038	60.75	102.10	58.42
NL	POA	24-27	10.86	92.35	155.57	36.65
Pers.	POA	28-31	13.56	186.96	285.59	-
KNN	POA	28-31	5.89	59.49	102.09	64.25
MLP	POA	28-31	10.64	60.11	101.45	64.48
NL	POA	28-31	8.16	98.40	160.22	43.90
Pers.	POA	32-35	32.22	202.01	289.68	-
KNN	POA	32-35	4.17	65.94	108.54	62.53
MLP	POA	32-35	8.39	65.31	105.22	63.68
NL	POA	32-35	23.56	113.87	165.40	42.90
Pers.	POA	36-39	16.82	142.84	206.39	-
KNN	POA	36-39	-32.66	53.65	77.94	62.24
MLP	POA	36-39	-44.07	59.25	88.15	57.29
NL	POA	32-35	-61.66	84.18	142.66	30.88
Pers.	POA	40-43	17.90	164.43	247.61	-
KNN	POA	40-43	-19.12	53.91	88.19	64.39
MLP	POA	40-43	-22.47	59.31	97.47	60.63
NL	POA	32-35	-44.51	81.37	145.25	41.34
Pers.	POA	44-48	-3.01	149.85	242.01	-
KNN	POA	44-48	-6.63	56.86	94.79	60.83
MLP	POA	44-48	-8.15	60.77	100.13	58.63
NL	POA	44-48	-22.64	86.73	149.37	38.28

Source: Elaborated by the authors of University of California, San Diego (2019).

Figure 15: Example of the POA Irradiance Forecast for Several Days



The top figure shows the measured data, the Persistence and the KNN forecast for the 24 ahead forecast. The bottom figure shows the RMSE for four forecast models for the same days in the top figure.

Source: Elaborated by the authors of University of California, San Diego (2019).

Real-Time Implementation

To turn the forecasting models useful for the solar farms, its real-time deployment is necessary. The real time implementation relies on the having live data from the two solar farms. During this project, the research team created an online platform that:

1. received real time data from the sensors at the solar farms;
2. computed the solar resource predictions for the different forecast horizons using the models described above;

3. updated the forecasting plots in the web portal and transmit the forecast to the solar farms.

The team create a web portal that allows tracking several of the forecasts in real-time as means to demonstrate this technology. To access the portal, please go to <http://coimbra-server3.dynamic.ucsd.edu/index.html>. Figure 16 shows a snapshot of the portal's homepage and Figure 17 shows a snapshot of the real time day-ahead DNI forecast for Ivanpah's tower 1.

Figure 16: Project Website

High-Fidelity Solar Power Forecasting Systems for the 392 MW Ivanpah Solar Plant (CSP) and the 250 MW California Valley Solar Ranch (PV)

Executive Summary

This project addresses critical technological gaps for increasing substantially the accuracy of solar forecasting for solar tracking power plants such as CSP, CPV and PV tracking.

The proposal focuses on developing high-fidelity direct normal irradiance (DNI) and plane of array (POA) irradiance forecasting methods, and on developing accurate irradiance-to-power models through advanced computational methods, enhanced telemetry and numerical weather predictions.

This project will deliver models that increase the accuracy and reliability of forecasts, assist in the management of variable and intermittent resources, and reduce the costs of solar generation by reducing utility scale costs associated with day-ahead scheduling, as well as reducing plant operation costs associated with intra-day variability.

Selected products for this project

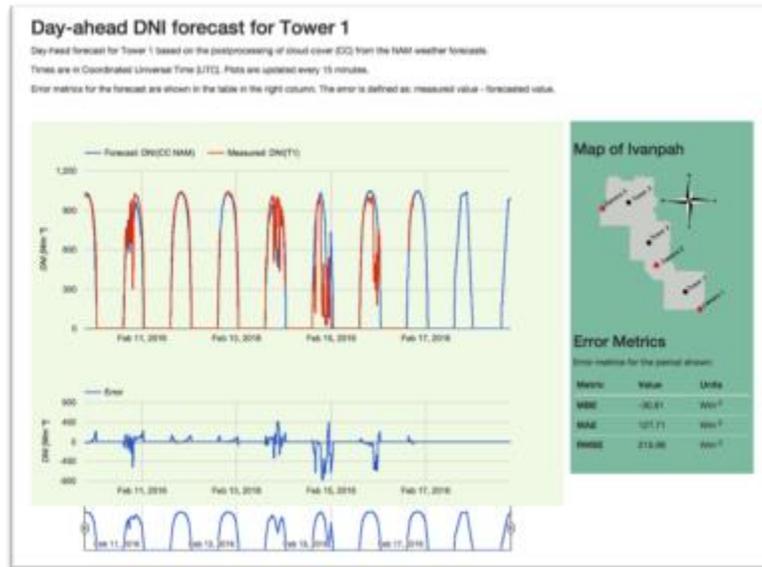
- Real-time measured data for CVSR
 - POA, PD and cloud fraction computed from sky images
 - POA irradiance measurements
 - GHI measurements from network of sensors
 - Power Output measurements
- Real-time forecasts for CVSR
 - Intra-hour PD forecast for CVSR circuit (8 H0 MW)
- Real-time measured data for Ivanpah
 - All Towers: wind speed measurements
 - All Towers: DNI measurements
 - All Towers: sky-images



Snapshot of the web portal that display real time telemetry and forecasts for this project.

Source: Elaborated by the authors of University of California, San Diego (2019).

Figure 17: Snapshot of the Real Time Day-ahead DNI Forecast for Ivanpah’s Tower 1



Source: Elaborated by the authors of University of California, San Diego (2019).

Applicability to Other Solar Projects

One of the goals for this project was to demonstrate the applicability of the tools developed in this project (for Ivanpah and CVSR) to other solar locations. The project team applied these tools and models to three cases:

- The short-term PO production forecast for a 1 MW solar farm.
- The intrahour DNI and GHI forecasts using sky image information.
- The intraday PO production forecast for two 1 MW solar farms.

Each case-study resulted in a peer-reviewed publication supported by the project:

1. H. T. C. Pedro, E. Lim and C. F. M. Coimbra (2018) "A Database Infrastructure to Implement Real-Time Solar and Wind Power Generation Intra-Hour Forecasts," *Renewable Energy* (123), pp. 513–525.
2. H. T. C. Pedro, C. F. M. Coimbra, M. David and P. Lauret (2018) "Assessment of Machine Learning Techniques for Deterministic and Probabilistic Intra-Hour Solar Forecasts," *Renewable Energy* (123), pp. 191–203.
3. D. P. Larson and C. F. M. Coimbra (2018) "Direct Power Output Forecasts from Remote Sensing Image Processing," *ASME Journal of Solar Energy Engineering–Transactions of the ASME* 140(2), Article Number: 021011.

The two cases related to the forecasting of power output (papers 1 and 3) from 1 MW plants were carried out in the medium size testbeds described in Chapter 2. The work for paper 2 was done with data from Folsom, California. This section summarizes the results and conclusions from those case-studies. A detailed description of the models

can be found in the three references listed and in the interim project report *Applicability of the Developed Tools for Other Solar Power Plants*.

A Database Infrastructure to Implement Real-Time Solar and Wind Power Generation Intrahour Forecasts

This study presents a simple forecasting database infrastructure implemented using the open-source database management system MySQL. This proposal aims at advancing the myriad of solar forecast models present in the literature into a production stage by implementing a MySQL infrastructure that collects the raw data, filters unrealistic values, classifies the data, and produces forecasts automatically and without the assistance of any other computational tools.

The performance of this method is demonstrated by creating intrahour (15-, 30-, 45-, and 60-minutes) power output forecasts for a 1MW photovoltaic installation in Southern California. Several machine learning forecast models are implemented (persistence, auto-regressive and nearest neighbors) and tested. Both point forecasts and prediction intervals (PIs) are generated with this methodology. Quantitative and qualitative analyses of solar power forecasts were performed for an extended testing period of four years. Results show a good and robust performance for the proposed forecasts. This test case leverages two of the technologies used in this project: the machine-learning methodology K-Nearest-Neighbors (kNN) and the open-source database management system MySQL.

This work demonstrated a MySQL database infrastructure to acquire, filter, classify, and forecast solar power generation data. It can be concluded that MySQL can be used for much more than just storing data for solar forecast models. Any forecast model that can be expressed in an algebraic form can be implemented in this manner. By adapting the data classification and forecast procedures presented in this work, it is possible to deploy many of the forecast models present in the literature in real-time. This can be very relevant in the production of stand-alone forecasting units that provide local monitoring and forecast renewable generation. For instance, in the case of solar generation, a simple hardware configuration consisting of a portable computing unit (such as BeagleBone) and photodiode sensor (similar to the central node for the WSN) could host the proposed methodology.

Assessment of Machine Learning Techniques for Deterministic and Probabilistic Intrahour Solar Forecasts

This work compares the performance of machine learning methods kNN and gradient boosting (GB) in intrahour forecasting GHI and DNI. The models predict the GHI and DNI and the corresponding prediction intervals. The data used in this work include pyranometer measurements of GHI and DNI and sky images. Point forecasts are evaluated using bulk error metrics such as RMSE and forecasting skill.

Results show that the machine learning models achieve significant forecast improvements over the reference smart persistence model. The reduction in the RMSE translates into forecasting skills ranging between eight percent and 24 percent, and 10 percent and 30 percent for the GHI and DNI testing set, respectively. Regarding the point forecasts, the GB method performs better than the kNN method when sky image features are included in the model. These observations are in line with what was observed in the development of intrahour forecasts for CVSR, as reported.

Direct Power Output Forecasts from Remote Sensing Image Processing

A direct method for intraday forecasts (1–6h ahead) of power output (PO) from photovoltaic (PV) solar plants is proposed and tested. The forecasting method uses publicly available images from geosynchronous satellites to predict PO directly without resorting to intermediate irradiance (resource) forecasting. Forecasts are evaluated using four years (January 2012–December 2015) of hourly PO data from two non-tracking, 1 MW PV plants in California. For both sites, the proposed method achieves forecasting skills ranging from 24 percent to 69 percent relative to reference persistence model results, with root mean square error (RMSE) values ranging from 90 to 136kW across the studied horizons.

The presented method enables direct prediction of PO of operational solar farms from satellite imagery. This has two major benefits: first, by directly forecasting PO, it avoids the need for intermediate irradiance forecasts and resource-to-power modeling, which are typically site specific; second, by removing the need for additional data dependencies, such as meteorological telemetry, it ensures that the method is generalizable to power plants with varying levels of pre-existing instrumentation. Together, these two features make the proposed forecasting method well posed to be applied to a range of operational PV power plants.

CHAPTER 6:

Short-Term Forecast for Market Preparedness

California is the leading solar state in the country of installed grid-connected solar capacity to date, with projections to achieve even higher renewable penetration in the next decades. In the pursuit of this goal the grid operator, California Independent System Operator or California ISO, has created the Participating Intermittent Resource Program (PIRP) to facilitate the integration of grid-connected variable energy resources (VERs) such as wind and solar. Recognizing that the prediction of generation from VERs entails a level of irreducible error, PIRP nets positive and negative deviations over each month and charges intermittent generators based on this value (deviation between generated and scheduled or forecasted energy). Monthly imbalances between the scheduled energy (based on the forecast) and the generated energy are settled based on locational marginal prices (LMP):

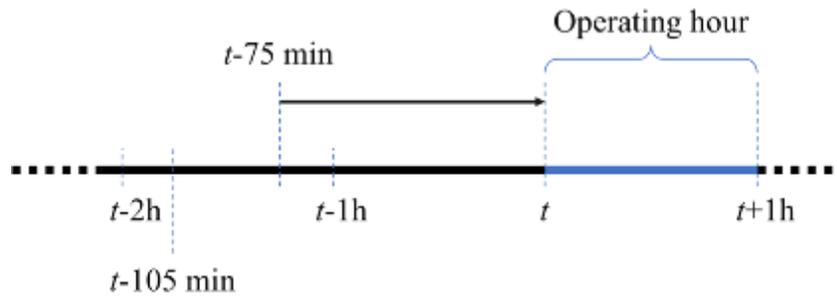
$$C = \sum_h (E_{sch,h} - E_{gen,h}) * LMP_h$$

where $E_{sch,h}$ is the scheduled energy production (usually in MWh) and $E_{gen,h}$ is the generated value for hour h and LMP_h is the respective locational marginal price. The summation in the previous equation is taken over all hours in a given month. This settlement strategy was designed to minimize penalties to the PIRP participants since it is based on the bias between the forecasted and actual values, which should approach 0 when averaged over the month for accurate forecasts.

PIRP participants without the capability to generate their own generation forecast use California ISO's centralized solar forecasting service (Letendre *et al.* (2014)) for large utility-scale plants and are charged a forecasting fee of \$0.10 per MWh. The centralized forecast was introduced in 2011 and is provided by AWS Truepower. It uses satellite-derived data, data from NWP models and recent measurements from the solar farm. Figure 18 illustrates the California ISO real-time market schedules, along with the publishing time for centralized solar forecasts. The centralized hour-ahead forecast is delivered 105 minutes before the operating hour (OH) and the hour-ahead scheduling and bids must be submitted 75 minutes before the respective OH. At the time of this project, CVSR participated in the PIRP program and was charged both the forecasting fee and monthly imbalances. Thus, the purpose of the market preparedness activities was to develop an alternative to California ISO's centralized solar forecasting for CVSR. An alternative accurate forecast would have two immediate advantages:

- Reduce the net monthly imbalances;
- Waive the forecasting fee.

Figure 18: California ISO hour-ahead market timeline.

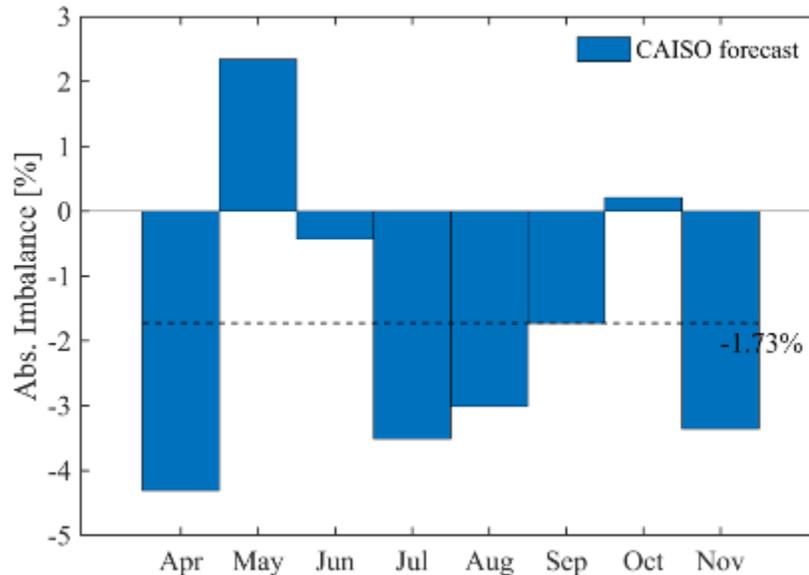


Forecasts for the operating hour (OH) t must be delivered 75 minutes before the start of the hour and cover the whole hour. The centralized hour-ahead forecast is delivered 105 minutes before the operating hour.

Source: Elaborated by the authors of University of California, San Diego (2019).

The first goal is the most important one, since the alternative forecast must be able to reduce imbalances with respect to California ISO’s forecast otherwise there is no incentive to modify current operations. To assess this criterion CVSR provided to UCSD relative net monthly imbalances from April to November 2016 (Figure 19).

Figure 19: Forecast Imbalances for April to November 2016



The figure shows the monthly imbalances between the centralized hour-ahead forecast delivered 105 minutes before OH and the generated power output at CVSR. A negative imbalance indicates that the forecasted PO was less than what was generated. The dashed line indicates the average imbalance for the testing period.

Source: Elaborated by the authors of University of California, San Diego (2019).

The values plotted in the figure are computed as follows, for each month i :

$$rIMB_i = \frac{(E_{sch,i} - E_{gen,i})}{E_{gen,i}} \times 100 [\%]$$

where the generated and scheduled energy for that month are, respectively

$$E_{gen} = \int_{t_1}^{t_2} P(t) dt = \sum_{h=1}^{N_{hours}} P_h$$

$$E_{sch} = \int_{t_1}^{t_2} \hat{P}(t) dt = \sum_{h=1}^{N_{hours}} \hat{P}_h$$

where P_h and \hat{P}_h are the measured and the forecasted power for a given OH. The figure illustrates that the monthly imbalances fluctuate considerably between close to zero percent for October and above four percent in April. On average there is a -1.73 percent imbalance. Thus, the main goal was to demonstrate an alternative forecasting model that can achieve lower monthly net imbalances. Given that the LMPs for this period are not available, the forecasting objective concerns only in reducing the imbalance $(E_{sch,i} - E_{gen,i})$. Not knowing the LMP does not prevent a comparison between the different forecast since LMP are set by California ISO ahead of time and are not depend on the generation forecast of a single market participant.

Data used consists of data collected from CVSR through the PI-to-PI connection set-up previously. To access the benefits to CVSR, the forecasting imbalance is compared against imbalances reported by CVSR for April to November 2015. Data from December 2016 onwards was used for training the models.

Several forecasting models were trained and tested with the goal to minimize monthly imbalances for CVSR. A key tool was the accurate clear-sky model explained in Appendix A for the power output of the solar farm. Using this model (P_c), the metered power output is normalized as $k_p(t) = P(t)/P_c(t)$. Although simple, this normalization technique can improve the accuracy of the predictions since the forecast model is left to predict the deviation with respect to the deterministic clear-sky model. Researchers used the normalized power as the dependent variable several forecast models:

- Persistence forecasts:

$$\hat{P}_{per,w}(t + 80 \text{ min}) = k_{p,w}(t) \times P_{cs}(t + 80 \text{ min})$$

where $k_{p,w}(t)$ is the average of k_p over a window with length w that precedes time t . As shown in (Pedro and Coimbra, 2015) optimizing this averaging window can result in better persistence forecasts. This approach also takes advantage of the fact that the PO data is available at a high granularity of 1-minute.

- NAM-derived forecast:

$$\hat{P}_{NAM}(t + 80 \text{ min}) = k_{G,NAM}(t + 80 \text{ min}) \times P_c(t + 80 \text{ min})$$

This model takes advantage of the GHI forecasts produced by NAM (an NWP model) produced and at 12 pm Coordinated Universal Time (UTC) or 4 am Pacific Standard Time (PST) (CVSR's time zone). The model uses GHI values with a resolution of 1h to produce PO forecast for the next OH by multiplying GHI's clear-sky index ($k_{G,NAM}$) by the respective PO clear-sky value.

- ANN forecast model:

$$\hat{P}_{ANN}(t + 80 \text{ min}) = k_{P,ANN}(features(t), NAM(t + 80 \text{ min}), geometry) \times P_c(t + 80 \text{ min})$$

In this model, an MPL is trained to predict the deviation with respect to the clear-sky model $k_{P,ANN}$ for the next OH. In order to predict this value, the ANN use data features derived from the most current PO, the NAM forecasts and also solar geometry.

- Hybrid model: Preliminary study of the forecasting performance showed that, as anticipated, the forecasting error reduces when ground data is available. However, models that do not use ground data, such as the ones derived from NAM forecasts, can be more accurate in some cases. One of those cases is in the forecasting of generation at the first hours of the day. For these forecasts there is none or very little daytime ground data available, thus reducing the accuracy of the models. An improved forecast can be obtained by merging the two types of forecasts. Here, that is accomplished by using the following model:

$$\hat{P}_{HYB}(t + 80 \text{ min}) = \begin{cases} \hat{P}_{NAM}(t + 80 \text{ min}) & \text{if } \alpha_s < \alpha_{s0}, \quad \varphi_s < 180^\circ \\ \hat{P}_{base}(t + 80 \text{ min}) & \text{otherwise} \end{cases}$$

where α_s is the solar elevation, φ_s is the solar azimuth, and α_{s0} is the threshold that controls when the forecast switches between ground data forecasts and NAM data forecasts. The condition $\varphi_s < 180^\circ$ simply guarantees that the NAM correction is only applied in the morning.

Note that, the ANN model used is an MPL as described in Appendix A. The same appendix also provides more information about the persistence model.

Results

All the candidate models were tested using training data. The best performing model is then selected and applied to the testing period between April and November of 2015.

Model Training and Development

The team developed four different persistence models in this work to take advantage of the 1-minute data available. The models simply differ in how much data is used to define the value that persists into the future. In mathematical terms, that affects how $k_{P,w}(t)$ in the persistence model is computed. Four backward averages of increasing window size were implemented here with $w = \{15, 30, 45, 60\}$ minutes. NAM-derived forecast was created using the 4 am GHI NAM forecast for the whole day. The GHI from NAM was converted to the respective clear-sky index by dividing its value by a clear-sky model for GHI by Ineichen and Perez that is explained in the Appendix A. With that value, PO can be readily computed. This model requires no training and is always available when NAM data is available. The MPL forecast was trained with two different sets of inputs:

- Features from $k_{P,w}(t)$ including backward averages and standard deviation, solar geometry (α_s and φ_s). This model is designated as ANN1 in the figures below.
- The same as above plus NAM-derived forecasts. This model is designated as ANN2.

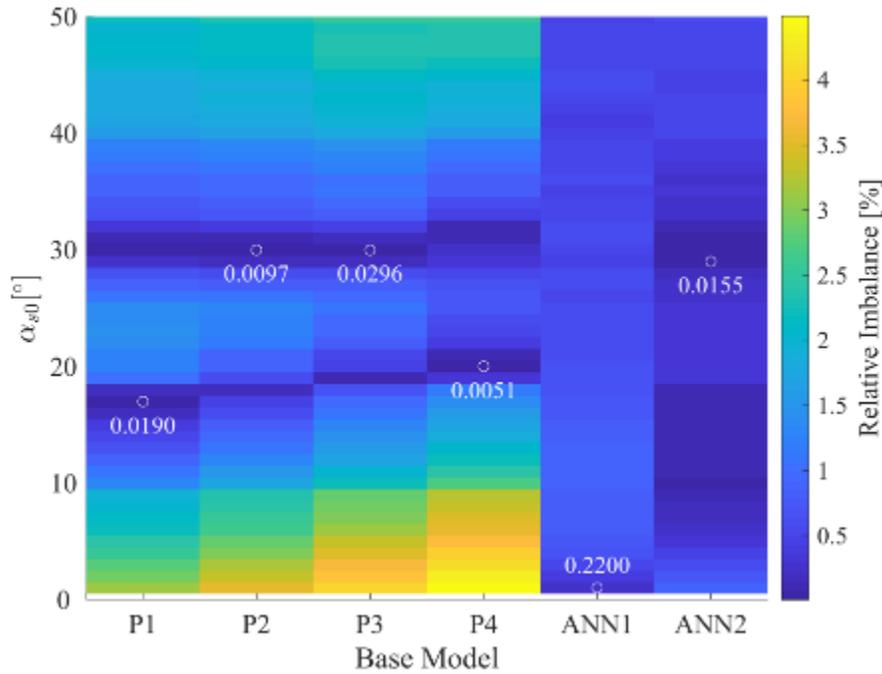
The ANNs used here are MPLs with a single hidden layer with 30 neurons. Due to the random nature of the ANN training (ANN parameters are initialized randomly), each ANN is trained 10 times separately. The final ANN forecast is the average of the 10 runs.

Finally, the hybrid model is also computed using for \hat{P}_{HYB} . Its accuracy depends on the base model and the solar elevation threshold α_{s0} . Figure 20 was created to determine the best base model and α_{s0} based on results for the training data. The figure was created by changing the base model and α_{s0} independently and computing the average relative imbalance for the resulting hybrid model. Figure 20 shows these values in the form of a heatmap where the best models (the ones with the lowest average relative imbalance) show in dark blue. For each base model (x-axis) the figure also shows the α_{s0} that results in the lowest imbalance and the respective value.

The figure shows that the hybrid model imbalance is sensitive to the α_{s0} value, especially when the base model is persistence. Hybrid models based on the ANN forecasts show lower sensitivity. Regardless, the figure indicates that the best hybrid model is based on the fourth persistence model ($w = 60$ minutes) with $\alpha_{s0} = 19^\circ$. Thus, the equation for the hybrid model becomes:

$$\hat{P}_{HYB}(t + 80 \text{ min}) = \begin{cases} \hat{P}_{NAM}(t + 80 \text{ min}) & \text{if } \alpha_s < \alpha_{s0}, \quad \varphi_s < 180^\circ \\ \hat{P}_{per,60}(t + 80 \text{ min}) & \text{otherwise} \end{cases}$$

Figure 20: Hybrid Model Selection

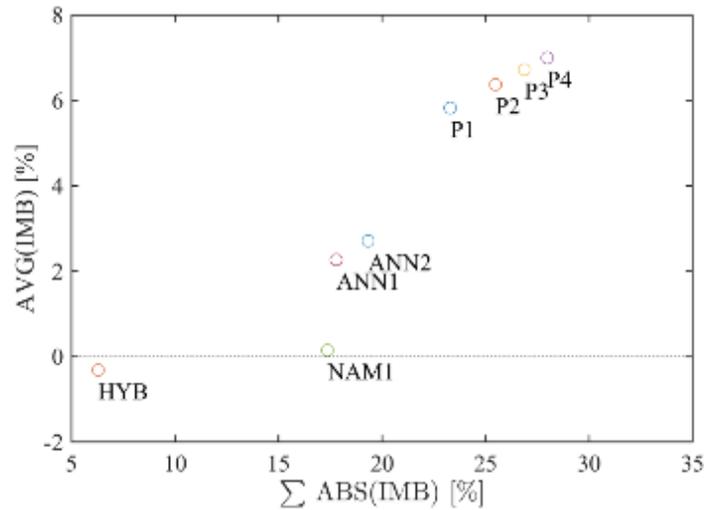


Heatmap used to select the hybrid model based on training data relative imbalance between forecasted and generated PO. The x-axis indicates the base model modified with NAM forecasts and the y-axis indicates the solar elevation threshold. The with circles and respective label indicate the lowest imbalance obtained for the respective base model.

Source: Elaborated by the authors of University of California, San Diego (2019).

Once all these models are determined and trained there are eight candidate models. The selection of the model to be compared against California ISO forecast imbalance was done based on Figure 21. The figure shows the average monthly imbalance on the y-axis and the sum of the absolute monthly imbalances on the x-axis for all candidate models applied to training dataset. The best model is the one that minimizes these quantities; thus, it is the one closer to the origin (0,0). The figure clearly shows that, as anticipated, that model is the hybrid model.

Figure 21: Model Selection

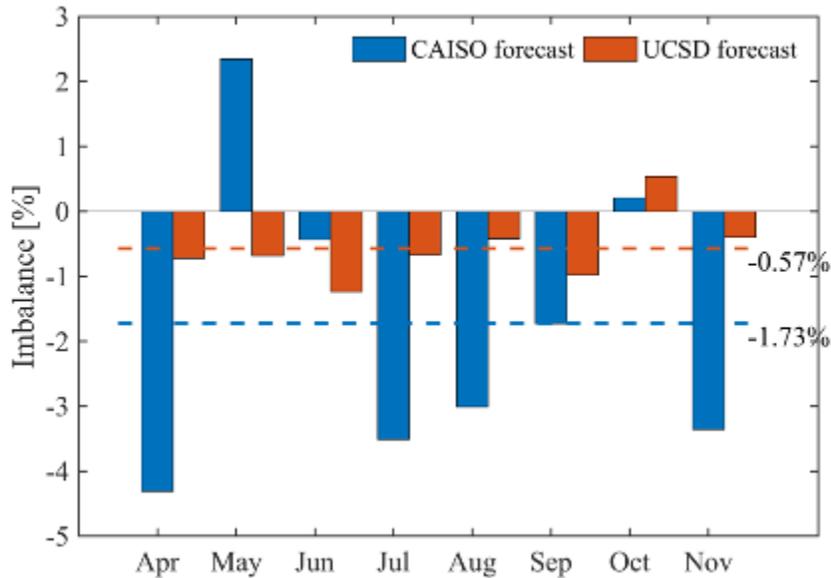


Model selection based on training data. The scatter plot shows the total absolute monthly imbalance (x-axis) and the average monthly imbalance (y-axis) for the candidate models (P=Persistence, HB=hybrid).

Source: Elaborated by the authors of University of California, San Diego (2019).

The hybrid model was then applied to the testing data between April and November 2015. After producing the power output forecast \hat{P}_{HYB} the relative monthly imbalance is computed with the equations provided at the top of this chapter and compared to the California ISO forecast imbalances reported by CVSR. Figure 22 compares the imbalances that results from these two forecasts. The figure shows that the \hat{P}_{HYB} forecasts reduces large imbalances (e.g. April, July and November) and only underperforms in a few cases where the imbalances are small (e.g. June and October). The figure also shows the average imbalance for the whole period incurred by the two forecasts (dashed lines and annotation); these are -0.57 percent and -1.73 percent for UCSD's and California ISO's forecasts, respectively.

Figure 22: Forecast Imbalance Comparison



Monthly imbalances for the centralized hour-ahead forecast (blue) and UCSD’s forecast (orange). The dashed line and respective label indicate the average monthly imbalances for the validation period.

Source: Elaborated by the authors of University of California, San Diego (2019).

These results clearly indicate that it is possible to reduce imbalances for PIRP participating solar farms. This conclusion is based primarily on the results shown in Figure 22 that indicate a 67 percent reduction in the monthly imbalance for the period between April and November of 2015. This result is based a single bulk error metric. A more granular error inspection would be necessary to validate fully the forecast improvement. Unfortunately, such study was not possible given that non-disclosure agreements prevented the research team from having access to the time-resolved California ISO centralized forecast. Nevertheless, given that the economic burden for CVSR is determined based on the monthly imbalance presented in the previous section, the model attained the goal of preparing CVSR to the California ISO real-time market.

As mentioned above, without having access to California ISO’s forecast and methodology is not possible to pinpoint accurately the reason to the forecast improvement. However, some explanations are below:

- The use of the most up-to-date ground telemetry: California ISO’s forecast is published 105 minutes before the OH, which dictates that the latest data is at least 105 minutes old. This value is probably even higher to account for data transmission and forecast creation. UCSD’s forecast takes advantage from using data as recent as 80 minutes before the OH. The 5-minute gap between the 80 minutes and 75 minutes is included in this study to account for time necessary to retrieve data from the PI server, compute the forecast and transmit the data

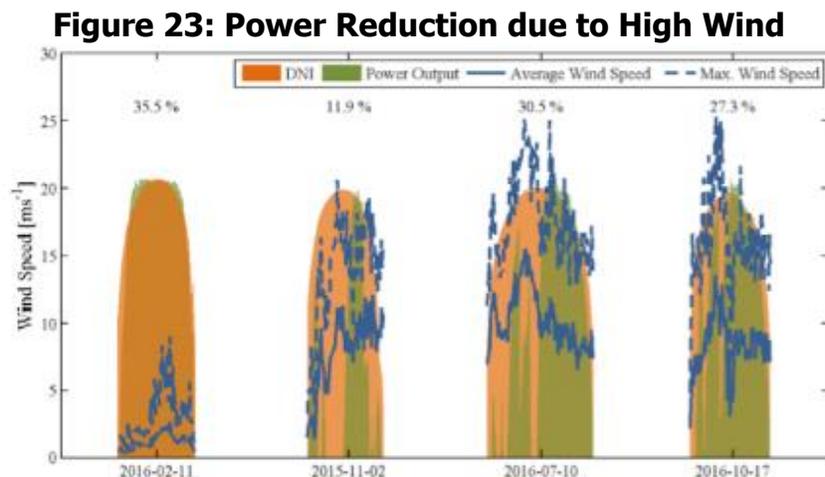
back to the PI server. With some care this gap could be reduce to under 1 minute, which could lead to further (but small) reduction of the imbalances.

- The use of a properly calibrated clear-sky model for the solar farm power plant: To participate in the PIRP solar producers must deliver 30 days of telemetry to the forecasting provider (Blatchford and Zack (2004)). This is a small sample to create a properly calibrated clear sky model (for instance, it does not contain data for every season). The UCSD forecast takes advantage of more than one-year' worth of data to create the clear sky model for the power produced at CVSR.

CHAPTER 7: Wind Surface Forecasting

Surface wind speed is a key variable for the operation of the solar farms studied in this project. High winds cause forces and moments on the irradiance collecting surfaces (PV panels or heliostats) and support and tracking structure. For this reason, surface wind velocity must be monitored and forecasted. In the event that a user-specified maximum wind speed is reached, the collecting surfaces move into a protective stowed position - usually parallel to the ground. These events effectively shut down the plant and their prediction is determinant to the accurate power production forecast aimed at in this project.

Although the collecting surfaces and structure are designed to sustain extreme weather events with wind gusts larger than 50 ms^{-1} (Strachan and Houser (1993)), in practice the systems are moved into a stowed position at conservative wind speeds around 20 ms^{-1} . The effect of wind gusts larger than 20 ms^{-1} on the power produced at Ivanpah can be seen in Figure 23.



Power reduction at Ivanpah's Tower 1 due to heliostat stowing during days with high wind speeds. The first day illustrates a day with calm wind as reference. The percentage values indicate the capacity factor for the respective day. DNI and PO curves are scaled for comparison.

Source: Elaborated by the authors of University of California, San Diego (2019).

The first day shown in the figure illustrates a day with low wind speed in which the plant operates normally. The following three days illustrate clear-sky days (in which one would expect full power production) with reduced power due to the high wind speeds. This simple analysis shows the importance of properly forecasting wind gust at the solar farm. Without this information, the most accurate forecast models that rely on solar irradiation will predict full power production for the last three days in the figure. To improve the power output forecast models, and more importantly, to improve the

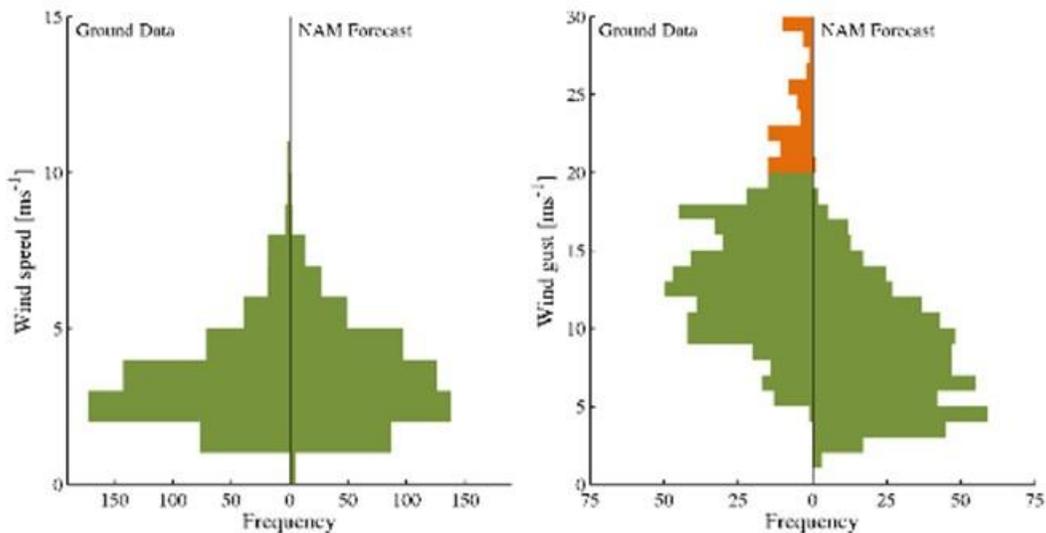
prediction of large wind gust that may damage the heliostats the project team have developed several wind forecast models based on local telemetry and numerical weather prediction (NWP) wind forecasts.

The first step in this task was to study the accuracy of the NWP wind forecasts, namely the North American Mesoscale (NAM) Forecast. Wind speed and wind gusts forecasts were obtained from the NAM repositories available daily at 4 am Pacific Time (PST). These forecasts consist of hourly values for average wind speed and wind gusts out to 24 hours ahead of time.

As discussed in Chapter 3 shows, the project uses data from the 13 nodes around Ivanpah. The eighth node is used as the reference or baseline forecast since it is the closest to Ivanpah. The other nodes provide additional input data to the forecast models that, as demonstrate below, help reducing the forecast error. The forecast performance is measured in reference to the ground data. In this case, data from the wind speed sensors installed at Ivanpah and synchronized with the NAM forecasts.

Histograms for the two data sets are compared in Figure 24. These histograms indicate the relative frequency of the daily averaged wind speed and daily wind gust in 1 m/s bins. The histograms for the average wind speed (Fig. 24 (left)) are similar but the histograms for the wind gust (right) are not. In particular, NAM forecast has just one day (indicated by the orange color) for which the wind gust reaches 20 m/s, whereas the sensors on the ground measured 74 days in the same period.

Figure 24: Histograms for Wind Data



Histograms for the daily averaged wind speed (*left*) and daily wind gust (*right*). The side-by-side plots compare the histogram for ground data against the histogram for the NAM forecasts. The orange color indicates the critical instance for which the wind speed or wind gusts surpass 20 m/s.

Source: Elaborated by the authors of University of California, San Diego (2019).

In summary, the NAM forecast is a promising baseline forecast since it predicts well the average wind speed at the solar farm. It severely under predicts the critical large wind gusts but that can be corrected with some advanced machine learning tools as explained below.

Figure 25 shows the mean bias error (MBE), and root mean square error (RMSE) for the NAM wind average and wind gust as a function of the hour of the day. These figures were obtained with the validation dataset. The bottom plot in the figures shows the measured data for the same period. The heat map and whiskers allow observing how the measured data varies as a function of the hour of the day. In general, the wind speed varies quite substantially at any time of the day resulting in larger errors. The goal of the forecast models developed in this project was to reduce these errors for all hours of the day. These models ingest the NAM data to produce a new wind forecast that takes into account localized effects via the use of local ground data in the training stage.

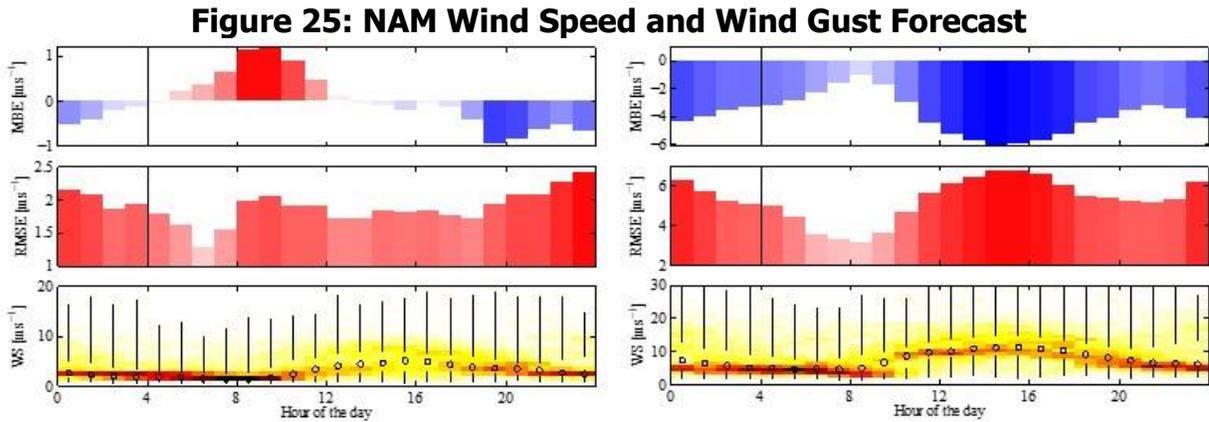


Figure 25: NAM Wind Speed and Wind Gust Forecast
 MBE (top) and RMSE (middle) for the NAM hourly average wind speed (left) and hourly wind gust (right) as a function of the hour of the day at Ivanpah for the validation set. The bottom panels show the measured average hourly wind speed (left) and wind gust (right). The heat map indicates the density of the measured data and the whiskers identify the 0, 25, 50, 75 and 100 percentiles. The vertical black bar indicates when the forecasts is produced (4 am daily PST).

Source: Elaborated by the authors of University of California, San Diego (2019).

Given that the baseline forecast model follows the trend of the measured data but substantially under predicts it, the first model developed consists of a global bias correction based on the hour of the day. In this model the forecasted average wind speed (\hat{W}), and wind gust (\hat{G}) are given by:

$$\hat{W}_b(\text{HOD}) = W_{\text{NAM}}(\text{HOD}) - B_W(\text{HOD})$$

and

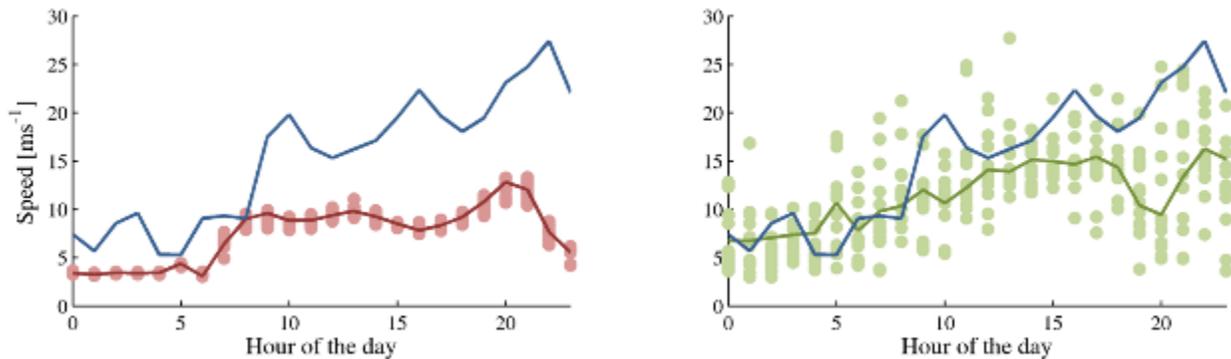
$$\hat{G}_b(\text{HOD}) = G_{\text{NAM}}(\text{HOD}) - B_G(\text{HOD})$$

where HOD is the hour of the day ($HOD = \{0, 1, \dots, 23\}$), W_{NAM} and G_{NAM} are the NAM average wind and wind gust forecast, and B_W and B_G are the bias errors computed with the whole training dataset as a function of HOD.

The second type of model developed for this problem is based on the well-established machine learning algorithm K-nearest neighbors (KNN) explained in Appendix A. In this case, the nearest neighbors are identified by computing the distance between the current NAM forecast and the historical NAM forecast. Given that there is data from several NAM nodes available for this operation, an optimization algorithm is used to find the set of inputs that best describes the similarity between current and past data. Once the list of nearest neighbors is determined the model assesses how the NAM forecast performed in the past for the selected conditions and corrects the new forecast based on that information.

Figure 26 illustrates how the KNN forecast for the wind gust is computed. The left figure shows the NAM gust forecast (red line) which under predicts the measured wind gust by a large amount. The dots indicate the NAM forecast from the nearest-neighbors. These values do not improve the forecast, however each one of these values is associated with a measured wind gust in the training data set. Those values are shown as the green dots on the plot on the right. By averaging those values for each hour of the day the model obtains the green curve which is much closer to the actual wind gust than the NAM forecast.

Figure 26: Example of kNN Correction for Wind Forecast



Left: the current NAM forecast (red line), the NAM forecast from the nearest-neighbors (red dots), and the measured wind gust (blue line). **Right:** the corrected wind gust based on all the nearest-neighbors (green dots), and the final adjusted forecast (green line).

Source: Elaborated by the authors of University of California, San Diego (2019).

Based on this idea the KNN forecasts for the average wind speed and wind gust are computed as:

$$\widehat{W}_k(HOD) = \frac{1}{K} \sum_{i=1}^K W_{NAM}(HOD) - (W_{NAM}(t_k) - W(t_k))$$

and

$$\hat{G}_k(\text{HOD}) = \frac{1}{K} \sum_{i=1}^K G_{\text{NAM}}(\text{HOD}) - (G_{\text{NAM}}(t_k) - G(t_k))$$

where K is the number of nearest-neighbors and $(W_{\text{NAM}}(t_k) - W(t_k))$ and $(G_{\text{NAM}}(t_k) - G(t_k))$ are bias errors in the training dataset associated with the nearest-neighbors identified by the time stamps t_k . This model is similar to the first type of models but instead of using error information from all the training set it uses only the information from the nearest-neighbors. This way the model learns from the past performance under similar conditions and corrects the NAM forecast accordingly.

Results

The results in Table 8 are calculated using validation data since that reflects more accurately the real time performance of the models. The initial performance analysis for the forecast models developed in this work is based on the error metrics given in Appendix A and listed in the table.

Table 8: Error Metrics for the Forecasted Mean Wind Speed and Wind Gust (ms⁻¹)

	Avg Wind Speed: NAM	Avg Wind Speed: \hat{W}_b	Avg Wind Speed: \hat{W}_k	Wind Gust: NAM	Wind Gust: \hat{G}_b	Wind Gust: \hat{G}_k
MBE	-0.033	-0.11	-0.040	-3.84	-1.00	-0.45
MAE	1.38	1.32	1.21	4.31	2.72	2.37
RMSE	1.91	1.83	1.70	5.42	3.73	3.23
Skill (%)	-	4.28	10.97	-	31.27	40.50

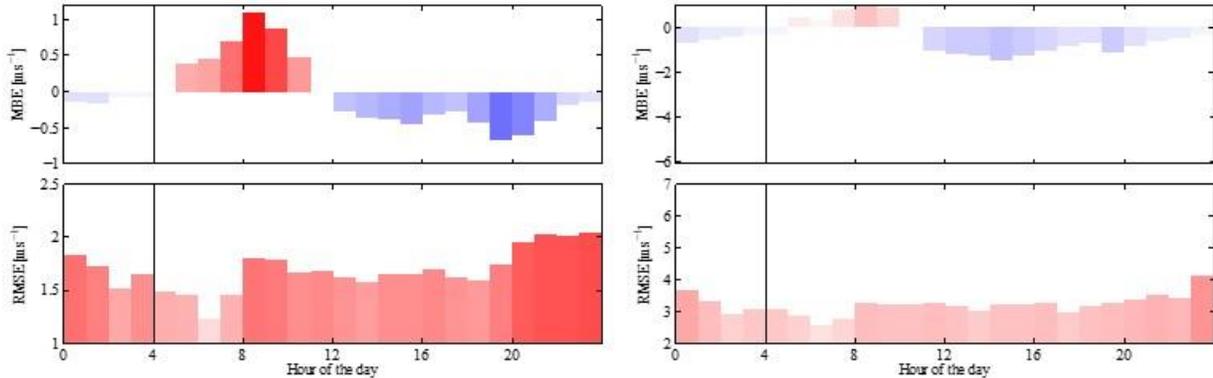
\hat{W}_b and \hat{G}_b refer to the models corrected with bias from the whole training set. \hat{W}_k and \hat{G}_k refer to the KNN forecast models. NAM refers to the error metrics for the baseline NAM model.

Source: Elaborated by the authors of University of California, San Diego (2019).

The results show that the forecast models implemented in this work improved the baseline forecast in almost all metrics. In terms of forecast skill, the wind speed forecasts show modest improvements of around 10 percent for the KNN model on average. This was expected since, as the analysis above showed, the NAM forecast is accurate in predicting this variable. In terms of wind gust, which is the key variable in identifying dangerous events for the power plant, the models developed under this project showed much higher forecasting skills. The bias correction results in an improvement of 31 percent and the KNN forecast reaches 41 percent skill. Figure 27 shows the improvements for the KNN forecasts as a function of the hour of the day.

Figure 27 depicts the daily wind gust histogram, which is another way to illustrate the improvements in the KNN gust forecast. Figure 28 compares the histogram for the KNN forecast against the histogram for the measured data. The histograms are very similar and the improvement relative the results for the NAM baseline gust forecast shown in Fig. 24 (right) is clear.

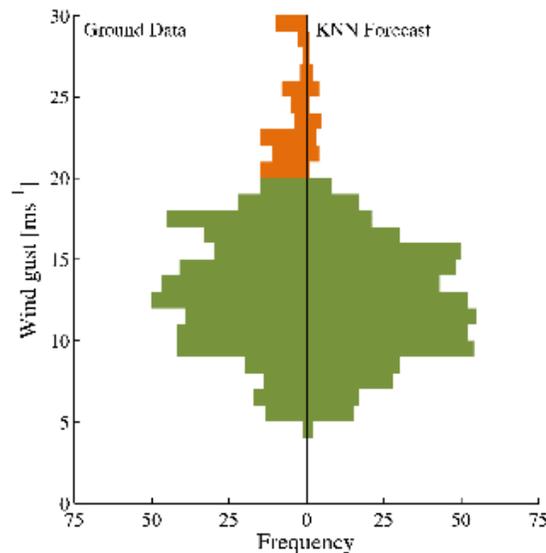
Figure 27: KNN Wind Forecast Errors as a Function of the Hour of the Day



Left: MBE (top) and RMSE (bottom) for the KNN hourly average wind speed forecast as a function of the hour of the day at Ivanpah for the validation set. *Right:* same but for the KNN hourly wind gust forecast.

Source: Elaborated by the authors of University of California, San Diego (2019).

Figure 28: Histogram Comparison for the Measured and Forecasted Wind Gusts

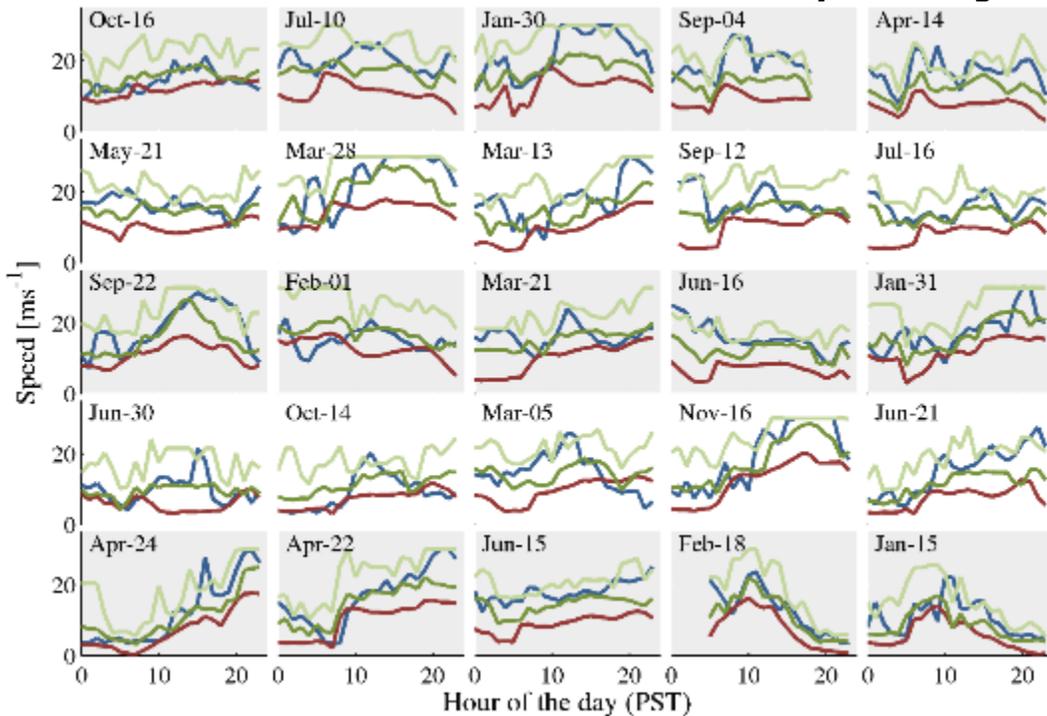


Histogram for the daily wind gust. The side-by-side plot compares the ground data against the KNN gust forecast. The orange color indicates the critical instance for which wind gusts surpass 20 m/s.

Source: Elaborated by the authors of University of California, San Diego (2019).

A final analysis of the wind forecast consisted in analyzing days for which the measured wind gusts surpass the 20 m/s critical value. Figure 29 shows the measured wind gusts, the baseline NAM forecast, the KNN gust forecast, and the maximum KNN gust forecast for 25 days with wind gusts above 20 m/s. The improvement in the forecasting of wind gusts is clear for most days. The large bias shown by the baseline NAM forecast is reduced, and more importantly, the maximum gust predicted by the KNN gust forecast matches the measured gust closely in most days. With these models, it is possible to create an accurate warning system that can flag potentially dangerous wind conditions for the heliostats at Ivanpah.

Figure 29: Measured and Forecasted Wind Gust for Days with High Wind



Several days for in which the wind gusts surpassed the critical value of 20m/s. The colors identify different data: measured data (blue line), NAM forecast (red line), KNN MOS forecast (dark green line) and max. value for the KNN MOS forecast (light green line).

Source: Elaborated by the authors of University of California, San Diego (2019).

CHAPTER 8:

Technology/Knowledge/Market Transfer Activities

The key technologies developed under this contract are in the area of solar forecasting and low-cost sensors for solar irradiance monitoring.

Solar Forecasting

The project team developed algorithms that address gaps in solar harvesting technologies that rely on Direct Normal Irradiance (DNI) and plane of array (POA) irradiance. These include concentrated solar power (CSP), concentrated PV (CPV) and tracking PV solar plants. To demonstrate these tools, the project team has collaborated with two larger solar producers in California:

- Ivanpah Solar Energy Generation Systems (Ivanpah), the largest CSP plant in the world.
- The 250 MW California Valley Solar Ranch (CVSR) which is a single axis PV tracking plant that represents the majority of solar plants in the United States.

The project team is continuing to promote the forecasting algorithms through journal papers, conference presentations, and direct communication with researchers and industry.

Low-Cost Sensors

Currently, high resolution monitoring of solar irradiance over a solar farm is frequently prevented by the sparsity of local telemetry available. Often just a few irradiance sensors monitor a solar field with areas in the tens of square kilometers. In this project, the project team developed a low-cost, autonomous and wireless sensor that are suitable to be install easily over a large solar field. The team collaborated with the industry at CVSR to test the sensors in a real-world scenario.

Technical Reports and Publications

Journal publications are technical reports that are peer-reviewed to ensure high quality and visibility through the journal database. Work associated DNI and POA irradiance forecast intrahour forecast models, advancement in the use of forecasting tools, and application of the developed tools to other solar plants (Non-Concentrating) generated 10 peer reviewed papers during the project. Appendix B contains the paper abstracts of the peer-reviewed papers.

These papers addressed major research questions in the field of solar energy forecasting:

- Real time solar power generation forecasting implementation:
 - H. T. C. Pedro, E. Lim and C. F. M. Coimbra (2018) A Database Infrastructure to Implement Real-Time Solar and Wind Power Generation Intra-Hour Forecasts, *Renewable Energy* (123), pp. 513–525.
- Machine learning tools for solar irradiance forecasting:
 - H. T. C. Pedro, C. F. M. Coimbra, M. David and P. Lauret (2018) Assessment of Machine Learning Techniques for Deterministic and Probabilistic Intra-Hour Solar Forecasts, *Renewable Energy* (123), pp. 191–203.
 - Y. Chu and C. F. M. Coimbra (2017) Short-Term Probabilistic Forecasts for Direct Normal Irradiance, *Renewable Energy* (101), pp. 526-536.
 - Y. Chu, M. Li and C. F. M. Coimbra (2016) Sun-Tracking Imaging System for Intra Hour DNI Forecasts. *Renewable Energy* (96), Part A, pp. 792-799.
 - H. T. C. Pedro and C. F. M. Coimbra (2015), Nearest-Neighbor Methodology for Prediction of Intra-Hour Global Horizontal and Direct Normal Irradiances, *Renewable Energy* (80) pp. 770-782.
- Solar power generation from satellite images:
 - D. P. Larson and C. F. M. Coimbra (2018), *Direct Power Output Forecasts from Remote Sensing Image Processing*, *ASME Journal of Solar Energy Engineering-Transactions of the ASME* 140(2), Article Number: 021011.
- Clear-sky solar irradiance:
 - M. Li, Y. Jiang, and C. F. M. Coimbra (2017), *On the Determination of Atmospheric Longwave Irradiance under All-Sky Conditions*, *Solar Energy* (144), pp. 40–48.
 - R. H. Inman, J. G. Edson and C. F. M. Coimbra (2015), *Impact of Local Turbidity Estimation on Clear Sky Models for Direct Normal Irradiance*, *Solar Energy* (117) pp. 125-138.
- Solar resource characterization and forecasting benchmarking:
 - H. T. C. Pedro and C. F. M. Coimbra (2015) "Short-Term Irradiance Forecastability for Various Solar Micro-Climates," *Solar Energy* (122), pp. 587–602.
- Cloud optical depth evaluation and forecasting:
 - M. Li, Y. Chu, H. T. C. Pedro and C. F. M. Coimbra (2016), *Quantitative Evaluation of the Impact of Cloud Transmittance and Cloud Velocity on the Accuracy of Short-Term DNI Forecasts*, *Renewable Energy* (86) pp. 1362–1371.

General Outreach

UCSD receives frequent visits from industry stakeholders, funding agency directors, educators and researchers (national and international). The project team conducts brief outreach sessions (30 minutes to 1 hour) where the team promote the work done in this project. For this purpose, the project team prepared a series of dynamic TV displays (Figure 30), where several of the deliverables for this project are presented and disseminated.

Figure 30: Dynamic TV Display of the Project Data and Results



Photo Credit: University of California, San Diego (2019).

Conference Presentations

The project team presented several posters and conference papers in several technical conferences: American Geophysical Union Fall Meeting, International Heat Transfer Conference, and American Meteorological Society (AMS) Conferences and Symposiums.

American Geophysical Union (AGU) Fall Meeting

The AGU Fall Meeting is the largest Earth and space science meeting in the world. "The meeting offers a unique mix of more than 20,000 oral and poster presentations, a broad range of keynote lectures, various types of formal and informal networking and career advancement opportunities, and an exhibit hall packed with hundreds of exhibitors showcasing new and relevant research tools and services that meet the professional needs of our attendees year after year."⁴

⁴ <https://fallmeeting.agu.org/2016/welcome/>

In the 2016 AGU Fall Meeting Hugo Pedro presented a poster about the forecasting for large centralized solar farms such as CVSR and Ivanpah. In the same conference the PhD student Yinghao Chu presented a poster about the intrahour forecasting of DNI.

International Heat Transfer Conference

“The International Heat Transfer Conferences (IHTC) are the world’s premier conferences for scientists and engineers in the heat and mass transfer research community. The conferences are convened every four years to exchange the latest heat and mass transfer information. IHTC-16 asked important questions about how we can better serve society and make this world a better place for all people. We face major issues such as ensuring sustainable development, healthy ageing, sufficient food for all, and economic growth that need scientific and technological solutions.”⁵

At the 2018 IHTC conference in Beijing, PhD student Mengying Li presented a new model, related to work for task 5, to evaluate clear sky radiation:

M. Li, Z. Liao and C. F. M. Coimbra (2018), *An Efficient Spectral Model for Evaluation of Clear-Sky Atmospheric Longwave Radiation*, International Heat Transfer Conference 16, Beijing, China.

American Meteorological Society (AMS) Conferences and Symposiums

“Founded in 1919, the American Meteorological Society (AMS) is the nation’s premier scientific and professional organization promoting and disseminating information about the atmospheric, oceanic, and hydrologic sciences. Our more than 13,000 members include researchers, educators, students, enthusiasts, broadcasters and other professionals in weather, water, and climate.”⁶

AMS organizes regular conferences and symposiums to disseminate the latest work in scientific areas related to solar forecasting. PhD students Mengying Li and Zhouyi Liao presented two papers in 2018 in AMS events:

M. Li, Z. Liao and C. F. M. Coimbra (2018) “Spectral Model for Clear-Sky Longwave Surface Irradiance,” American Meteorological Society 31st Conference on Climate Variability and Change, Austin, TX.

Z. Liao, M. Li and C. F. M. Coimbra (2018) “Nonisotropic Aerosol Scattering Effects on Longwave Irradiance,” American Meteorological Society 10th Symposium on Aerosol–Cloud–Climate Interactions, Austin, TX.

⁵ <https://www.ihtc16.org/index.php>

⁶ <https://www.ametsoc.org/index.cfm/ams/about-ams>

2018 Electric Program Investment Charge (EPIC) Symposium

The Epic Symposium is a one-day event that highlights research projects using innovative energy technologies. This event also offers informative poster sessions and networking opportunities with high-powered speakers, industry leaders, and researchers. One member of the research team participated in the panel "Accurate Forecasting to Support the Modern Grid". The panel addressed questions such as: What instrumentation was the most beneficial to forecast accuracy? What is the optimal instrumentation set for different power plant locations and characteristics?

CHAPTER 9:

Benefits to Ratepayers

The research team successfully demonstrated tools to increase the forecasting reliability for CPS and PV tracking solar generation. These tools have the potential to facilitate and accelerate the integration of carbon-free technologies into the existing California power grid and market structure. Accurate solar generation forecasting could help ensure a reliable power supply and higher quality power and enable the growth and integration of grid-connected solar energy resources into the electric grid. The project team has identified several areas that will benefit from this project:

- Enhanced capacity of utility-scale CSP power plants due to improved prediction of resource and power output; better integration of solar assets with utility and ISO operations;
- Increased ability to absorb short-term ramps and maintain solar production lead to lower overall operation costs and lower consumer cost per solar kWh;
- Decreased number of unscheduled power outages due to solar variability;
- Better use of ancillary generation resources (for example, services dispatched to ensure that there is sufficient energy generation to meet load) including lower utilization of peak fossil fuel powered plants;
- Improved health effects from increased levels of renewable generation displacing fossil fuel-based generation over the lifetime of higher capacity solar plants; and
- Reduced renewable energy curtailment.

In 2018, the solar power capacity at utility scale in California was more than 15 percent of the total power capacity of the state. This means that, on average, the power consumption of more than 6 million California residents is entirely provided by solar power plants. This figure includes a compensation for additional spinning reserves to take into account the variability of both solar and wind power plants. This project was sponsored by the Energy Commission at ~\$1 million over three years, which means that the total cost to the average ratepayer (considering only the equivalent fraction that is entirely covered by solar year-round) was less than 6 cents for the entire 3-year project, or less than the cost of 1kWh by any energy source.

Clearly, forecasting research not only can substantially affect the ability of power plant managers, utility companies and the California ISO to reduce solar costs to all ratepayers, but it can also allow for much higher penetration of renewables. In addition, according to the cost estimates, even developing new forecasting methods and techniques at the most fundamental level (the most expensive part), is a great investment to ratepayers. A single ramping event in a large-scale power plant that is

predicted correctly (for example, a DNI ramp of $800\text{Wm}^{-2}/\text{minute}$) can return the investment of a whole year, if the power plant is warned early enough to avoid tripping. The same is true with curtailment management.

CHAPTER 10:

Conclusions/Recommendations

The following conclusions are drawn from the work:

- This project demonstrated the design, assembling, deployment and testing of a low-cost wireless sensor network (WSN). The project team evaluated the performance of the sensor network of sensors over approximately one year of data. The results show strong performance in terms of data quality and data availability. This technology is suitable for deployment at any solar farm to supplement local data acquisition, which often has been limited to just a few sensors.
- Accurate DNI forecasts are more difficult to obtain than POA irradiance forecasts. This results from the fact that DNI is much more sensitive to clouds and aerosols than POA irradiance, resulting in a higher variability. Rapidly changing weather patterns and cloud formation and dissipation lead to changes in solar irradiance and generation that can exceed 60-80 percent over a few minutes. Sudden fluctuations of large magnitude are challenging for the forecasting algorithms. Additionally, the results demonstrated that, for most forecast horizons the models developed in this project show large forecasting skills (above 30 percent reduction in forecast error relative to reference models). For POA irradiance, several forecasts show skills above 50 percent.
- The intrahour PO forecasting at CVSR shows that higher forecasting skills require careful and innovative feature engineering - a process that extracts new data from the original raw values. The new features can then be used as predictors to in machine learning tools, such as XGBoost.
- The researchers applied the tools developed for the two utility-scale solar farms (Ivanpah and CVSR) in smaller (1MW) testbeds. From these results, all algorithms and tools developed are suitable to implement easily and effectively in other solar projects.
- A tailor-made solution for CVSR using the most up-to-date data available and a carefully calibrated clear-sky model results in forecasts that substantially reduce the net imbalances relative to the centralized California ISO forecast.
- The baseline NAM forecast (the NAM forecast for the node closest to Ivanpah) severely under predicts the wind gust magnitude. Results for the forecast models developed by this project overcome this deficiency in most cases. With these forecasts, it is possible to create an accurate warning system that can flag potentially dangerous wind conditions for the heliostats at Ivanpah. These models are general and are suitable for any other locations.

Based in this work, the project team offers the following recommendations for future work:

1. Development of a map of predictability ("forecastability") for different solar microclimates in California. Such a map, when combined with other constraints (distance to distribution lines, distance to populations, etc.), is critical for siting and proper integration of large-scale solar power plants into the grid.
2. A study of scenarios with multiple penetrations of solar and wind in different regions, with an analysis of how new solar and wind capacity in these regions can contribute to offset atmospheric carbon emissions.
3. Integration of the new NOAA/NASA satellite (GOES 17) images into the forecasting engines described in this work. The GOES-17 images started streaming in late 2018 and represent a substantial improvement over the GOES-West images, both in terms of refreshing rates and spatial resolution.
4. A study of the optimal distribution of low-cost sensors and the potential of using the sensors outside of the solar fields for early cloud detection.
5. Finally, the development of smart and hybrid remote sensing/numerical weather prediction models that cover the entire state of California in real time for the short-term prediction (0-6 hours) of solar and wind resources, with local optimization for regions with high concentration of solar and/or wind power plants.

GLOSSARY AND ACRONYMS

Term/Acronym	Definition
EPIC	Electric Program Investment Charge
Smart Grid	Smart Grid is the thoughtful integration of intelligent technologies and innovative services that produce a more efficient, sustainable, economic, and secure electrical supply for California communities.
CVSR	California Valley Solar Ranch
MSE, MAE, MBE	Mean Squared Error, Mean Absolute Error, Mean Biased Error. Common error metrics used to assess the accuracy of the forecasted values. These values quantify the forecast error. Lower values indicate better forecasts.
s	Forecasting skill. Metric that compares the forecasting accuracy of a new model relative to some benchmark result. $s > 0$ indicates a model that performs better than the benchmark ($s = 1$ indicates a perfect forecast); $s = 0$ indicates a model that perform as well as the benchmark; $s < 0$ indicates a model that performs worse than the benchmark.
DNI, B	Direct Normal or Beam Irradiance. The irradiance component that takes into account only light rays that come in a straight line from the direction of the sun at its current position in the sky.
GHI, G	Global Horizontal Irradiance. The amount of terrestrial irradiance falling on a surface horizontal to the surface of the earth.
POA, I	Plan of array Irradiance. Similar to GHI but in this case the surface can be at any angle. For a horizontal surface it is equal to GHI.
CSP	Concentrated Solar Power. A system to generate solar power that uses mirrors or lenses to concentrate a large area of sunlight onto a small area.
CPV	Concentrated Photovoltaics. A photovoltaic technology to generate solar power. It uses lenses or curved mirrors to focus sunlight onto small, highly efficient solar cells.
NWP	Numerical Weather Prediction. Algorithms that use mathematical models of the atmosphere and oceans to predict the weather based on current weather conditions.
SFINCS	Solar Field Controls System
PO	Power Output. The electric power generated by the solar plant.
NAM	North America Mesoscale forecast system. A numerical weather prediction model run by National Centers for Environmental Prediction for short-term weather forecasting.
ML	Machine Learning. A method of data analysis that automates analytical model building.

Term/Acronym	Definition
ANN	Artificial Neural Network. One of the main tools used in machine learning. They are tools for finding patterns which are too complex or numerous for a human programmer to extract and teach the machine to recognize.
KNN	K-Nearest Neighbors. A classical machine learning algorithm that relies in finding similar conditions in the historical data to produce the forecasting.
MLP	Model Multilayer Perceptron. A type of ANN.
SVR, SVM	Support Vector Regression, Support Vector Machines. Machine learning algorithms that analyze data for classification and regression analysis.
OLS	Ordinary Least Squares. A type of linear least squares method for estimating the unknown parameters in a linear regression model.
XGBoost	Open-source software library that provides a gradient boosting framework. An implementation of a popular machine learning algorithm for classification and regression.
WSN	Wireless Sensor Network. A collections of irradiance sensors installed over the solar plant, connected wirelessly to a central node that collects the data.
O&M	Operations and Management
GOES	Geostationary Operational Environmental Satellite system. Satellites that support weather forecasting, severe storm tracking, and meteorology research.
k_t, k_p	Clear-sky index. The ratio between actual solar irradiance and the solar irradiance under clear sky conditions.
B_c, I_c, P_c	Clear-sky value for DNI, irradiance, PO. Models that estimate the clear sky values at a certain time of the day.
RGB, NRBR	Red Green Blue, normalized red to blue ratio. Color data from sky images.
$\Delta t, FH$	Forecasts Horizon. The length of time into the future for which forecasts are to be prepared.
\hat{B}, \hat{P}, \dots	Forecasted data. The data produced by the forecasting models.
W, G	Average wind speed, wind gust.
Heliostat	An instrument consisting of a mirror mounted on an axis moved by clockwork by which a sunbeam is steadily reflected in one direction.

REFERENCES

- Blatchford, J., & Zack, J. W. (2004). California ISO's Participating Intermittent Resource Program (PIRP): Description and Initial Results. Proc. Global WINDPOWER.
- Braun, J.E., and Mitchell, J.C. (1983). Solar geometry for fixed and tracking surfaces. *Solar Energy* 31, 439–444.
- Chang, C.-C., and Lin, C.-J. (2011). LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.* 2, 27:1–27:27.
- Chen, T., and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (New York, NY, USA: ACM), pp. 785–794.
- Chu, Y., Pedro, H.T.C., Nonnenmacher, L., Inman, R.H., Liao, Z., and Coimbra, C.F.M. (2014). A Smart Image-Based Cloud Detection System for Intrahour Solar Irradiance Forecasts. *J. Atmos. Oceanic Technol.* 31, 1995–2007.
- Heidinger, A.K., Foster, M.J., Walther, A. and Zhao, X., (2014). The pathfinder atmospheres–extended AVHRR climate dataset. *Bulletin of the American Meteorological Society*, 95(6), pp.909-922.
- Hoff, T.E., Perez, R., Kleissl, J., Renne, D., and Stein, J. (2012). Reporting of irradiance modeling relative prediction errors. *Progress in Photovoltaics: Research and Applications*
- Ineichen, P., and Perez, R. (2002). A new airmass independent formulation for the Linke turbidity coefficient. *Solar Energy* 73, 151–157.
- Lave, M., Reno, M.J., Stein, J., and Smith, R. (2015). Low-cost solar variability sensors for ubiquitous deployment. In *2015 IEEE 42nd Photovoltaic Specialist Conference (PVSC)*
- Letendre, S., Makhyoun, M., & Taylor, M. (2014). Predicting solar power production: irradiance forecasting models, applications and future prospects. *Solar Electric Power Association, Tech. Rep.*
- Marquez, R., and Coimbra, C.F.M. (2012). Proposed Metric for Evaluation of Solar Forecasting Models. *J. Sol. Energy Eng.* 135, 011016–011016.
- Pedro, H.T.C., and Coimbra, C.F.M. (2012). Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy* 86, 2017–2028.

- Pedro, H.T.C., and Coimbra, C.F.M. (2015). Nearest-neighbor methodology for prediction of intra-hour global horizontal and direct normal irradiances. *Renewable Energy* 80, 770–782.
- Perez, R., Ineichen, P., Seals, R., Michalsky, J., and Stewart, R. (1990). Modeling daylight availability and irradiance components from direct and global irradiance. *Solar Energy* 44, 271–289.
- Remund, J., Wald, L., Lefèvre, M., Ranchin, T., and Page, J.H. (2003). Worldwide Linke turbidity information. In *Proceedings of ISES Solar World Congress 2003*
- Reno, M.J., and Hansen, C.W. (2016). Identification of periods of clear sky irradiance in time series of GHI measurements. *Renewable Energy* 90, 520–531.
- Strachan, J.W., and Houser, R.M. (1993). Testing and evaluation of large-area heliostats for solar thermal applications (Albuquerque, NM: Sandia National Laboratories).

APPENDIX A: FORECASTING TOOLS

This appendix details several of the tools used in this project.

Clear-Sky Models

A good clear-sky model for the irradiance has a big impact in the forecast error. These models remove well predicted variations in the irradiance (such as seasonal and daily cycles) leaving the forecast tools to deal with the challenge of predicting variations due to localized weather conditions. In this project, clear-sky normalization is applied to several data: irradiance (GHI, DNI and POA irradiance) and power generation.

General Irradiance Clear-Sky Model

In several cases this project uses a well know irradiance clear-sky model for GHI and DNI. This model is the implementation of the algorithm proposed in the work of Ineichen and Perez (2002), which requires Linke Turbidity as an input. Turbidity measures the haziness of the atmosphere due to suspended particles or aerosols. Maps of monthly average Linke Turbidity developed by Remund et al. (2003). In other cases, clear-sky models are developed specifically for a location or variable. These are explained next.

DNI Clear-Sky Model for Ivanpah

In the case of DNI at Ivanpah, the general model explained above was deemed not accurate enough. For this reason, the project team developed an empirical model to compute the clear-sky irradiances based on data collected at the solar farm. Data was selected based on the irradiance measurements and sky images, as the clearest days in the historical data. A polynomial curve fit is then used to estimate clear-sky DNI as a function of the solar elevation angle α_s : $B_c = \sum_{i=1}^6 a_n \alpha_s^n$ with the following coefficients $a_n = \{-8.859 \times 10^{-4}, 8.026 \times 10^{-2}, -3.726 \times 10^{-3}, 1.039 \times 10^{-4}, -1.671 \times 10^{-6}, 1.42 \times 10^{-8}, -4.916 \times 10^{-11}\}$.

Given that, even during cloudless days, the irradiance (GHI and DNI) can be attenuated by aerosols, the previous equations is by including a turbidity correction coefficient $C_{T,B}$. The larger the turbidity, the greater is the irradiance attenuation and the smaller will be the correction factor $C_{T,B}$. With this correction, the final clear-sky model for DNI becomes

$$B_c = C_{T,B} \sum_{i=1}^6 a_n \alpha_s^n.$$

The factor $C_{T,B}$ is estimated daily from periods of clear sky. If no such periods exist $C_{T,B}=1$.

For Ivanpah, clear-sky GHI is estimated in a similar way. In this case, with a fourth order polynomial

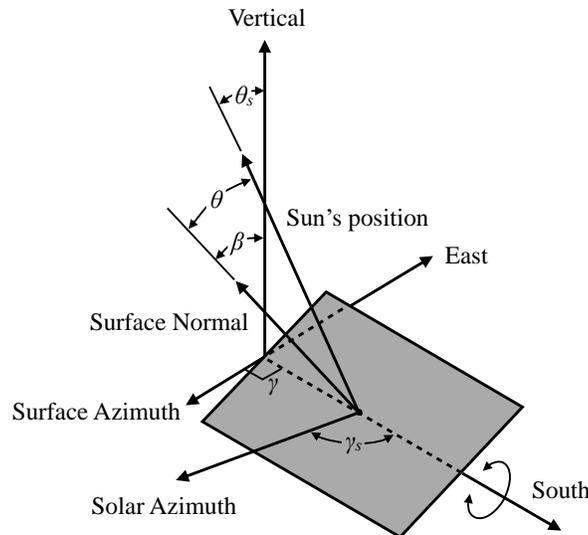
$$G_c = C_{T,G} \sum_{i=1}^4 b_n \alpha_s^n,$$

with $b_n = \{-5.266 \times 10^{-3}, 1.025 \times 10^{-2}, 3.478 \times 10^{-4}, -5.982 \times 10^{-6}, 2.703 \times 10^{-8}\}$. The turbidity correction coefficient for GHI ($C_{T,G}$) is estimated the same way as the one for DNI.

POA Irradiance and PO Clear-Sky Model for CVSR

The development of a power output clear sky model for CVSR involves the modeling of clear sky irradiance components that can then be projected onto the tilted surface of an optimally oriented solar panel. These values are computed with the work of Ineichen and Perez (2002), mentioned above. After the clear sky model has been calculated, optimal orientation of the panel must be determined. It is well known that the DNI on a panel is maximized when the angle of incidence is minimized, within the constraints of the tracking apparatus. In the case of CVSR, see Fig. A1, the constraints are such that the single axis of rotation is parallel to a vector pointing due south, which greatly simplifies the expressions for the incidence angle as a function of surface tilt and azimuth angles for optimally tracking single-axis panels provided by Braun and Mitchell (1983).

Figure A1: Single Axis Tracking Geometry



Geometry necessary for determining plane of array (POA) irradiance for the single axis tracking panels at CVSR.

The minimization of the angle of incidence is performed in two steps: first, determine the optimal surface slope; and second, determine the angle of incidence using the solar zenith and azimuth angles.

For the geometry illustrated in Fig. A1, the incidence angle is minimized when

$$\frac{d(\cos\theta)}{d\beta} = \sin\theta_s \cos(\gamma_s - \gamma)\cos\beta - \cos\theta_s\sin\beta = 0$$

where θ is the angle of incidence, β is the surface slope, θ_s is the solar zenith angle, γ_s is the solar azimuth angle, and γ is the surface azimuth angle. As a result of the tracking constraints at CVSR, the surface azimuth $\gamma = \pm 90^\circ$. Substituting this value and solving for the surface slope gives

$$\beta = \tan^{-1}[\tan\theta_s\sin\gamma_s]$$

which relates the surface slope to the solar geometry. Many sources contain the following trigonometric relationship for the angle of incidence θ and solar and surface angles

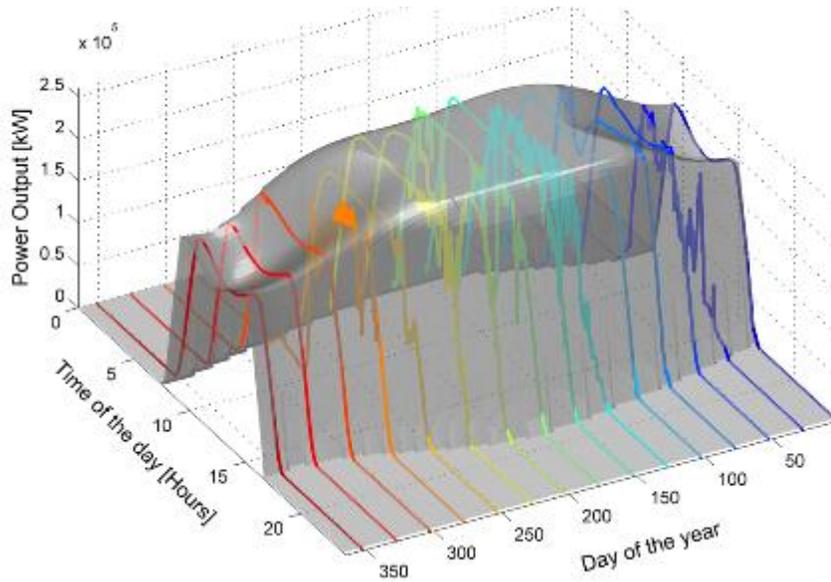
$$\theta = \cos^{-1}[\cos\beta\cos\theta_s + \sin\beta\sin\theta_s\sin\gamma_s]$$

Once the clear sky irradiance components have been modeled and the angle of incidence is known, the plane of array irradiance is calculated by projecting the irradiance components onto the optimally oriented surface. This is accomplished using the work of Perez et al. (1990), which presents several models to estimate DHI impinging on tilted surfaces of arbitrary orientation. Using the results from the Perez model for plane of array DHI in conjunction with the clear sky DNI calculated from the Ineichen model and angle of incidence, the plane of array irradiance can be calculated as

$$I_{\text{POA}} = I_{\text{D}}\cos\theta + I_{\text{d,POA}}$$

where I_{POA} is the plane of array irradiance, I_{D} is the clear-sky DNI, and $I_{\text{d,POA}}$ is the diffuse clear sky plane of array irradiance. As a final step, the plane of array irradiance is transformed into power output by linearly scaling by a factor of equal to the nameplate capacity of the solar farm in W; i.e., 2.5×10^5 and curtailed for values that exceed the capacity. Figure A2 shows the clear-sky model for the 250MW solar farm. The figure also shows the measured PO for some clear-sky and cloudy days to demonstrate the clear-sky model accuracy.

Figure A2: Power Output Clear-Sky Model for CVSR



The gray surface indicates the clear-sky power output as a function of the day of the year and the time of the day (solar time). The colored lines show actual measurements from the solar plant. The new model accounts for the tracking of the sun by the solar panels. This model is a fully deterministic model (only requires the date and time as inputs) and is used to improve the power output forecast at CVSR.

Error Metrics

The forecast performance is established by analyzing the forecast error, that is the difference between the measured and forecasted data. In this project, the models are evaluated using error metrics recommended in forecast literature: mean absolute error (MAE), mean bias error (MBE), root mean square error (RMSE), and forecast skill (s) (see Marquez and Coimbra (2012) and Hoff *et al.* (2012)).

MBE is a measure of systematic errors (or bias). An unbiased forecast model, that is, one that on average does not under or over predicts irradiance will have an MBE value close to 0. It is an important metric because understanding the overall forecast bias (over or under-forecasting) would allow power system operators to better allocate resources for compensating forecast errors in the dispatch process.

MAE is the magnitude of the forecasting error. In this metric all errors contribute equally regardless of their magnitude.

RMSE is a measure of random errors and, unlike MAE, it penalizes larger errors. In this sense this is a very important metric since one of the goals of this project is to reduce the instances of large error.

Forecast skill s is a convenient way to benchmark new forecast models. It measures their performance relative to the persistence model. It is a good metric to compare models applied to different variables (DNI, POA irradiance, solar generation, etc.) since

the intrinsic variability of the data is reflected in the RMSE for the persistence model. For example, a model that returns very low RMSE for a dataset comprised of only clear-sky days will have a very low skill since the RMSE for the persistence model will also be very low. The ideal forecast will have a low RMSE and high forecast skill.

Mathematically these error metrics are defined as:

$$\text{MBE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

$$s = 1 - \frac{\text{RMSE}}{\text{RMSE}_p}$$

where y_i is the true value, \hat{y}_i is the forecasted value and RMSE_p is the RMSE of the reference forecast model persistence forecast. In all the results presented in this report nighttime values are discarded. These metrics can be applied to any subset of the data. In the results presented they are applied to the testing dataset, since those are the results representative of the model's performance when applied to live data.

Forecasting Models

Smart Persistence

As mentioned above, forecast skill is evaluated against the smart persistence model. The smart persistence model assumes the clear-sky index k_t (the ratio of measured irradiance or power generation to the clear-sky value) persists over the forecast horizon (Δt). Therefore, the predicted value is given by,

$$\hat{y}_p(t + \Delta t) = k_t(t) \times y_c(t + \Delta t)$$

where the subscript p denotes the persistence model and y_c is the one of the clear-sky models introduced above. The variable y denotes any of the irradiance components or power generation used in this work. It should be noted that this is the general form for the persistence model. In the work presented above a few simple variations are used (for instance, $k_t(t)$ can be defined with the latest data point available or it can use

several of the last values). In those cases, the differences relative to this model are clearly stated.

Linear and Non-Linear Least-Squares

These models are explored in this project as forecasting models. Generically, the forecast is produced from a linear or non-linear combination of the input variables. For the linear model, this means that the forecast is calculated as

$$\hat{y}_p(t + \Delta t) = a_0 + a_1x_1(t) + \dots + a_nx_n(t)$$

where x_1, x_2, \dots, x_n are the input data (current telemetry, sky images, etc.), The linear regression fits a linear model with coefficients a_0, a_1, \dots, a_n to minimize the residual sum of squares between the observed data, and the predicted data by the linear approximation. The non-linear version of this model is very similar, however in the right-hand-side of the previous equation there are non-linear terms such as $a_1x_1(t)^{b_1}$. In this case the minimization problem must determine the free-coefficients a and b .

MPL Models

Multi-layer perceptron (MLP) is a type of artificial neural network that consists of one input layer, several hidden layers and one output layer. Each hidden layer consists of several neurons - the processing elements. The model used in this project, follows from previous work done by members of the project team, in which the MPL structure is optimized using a genetic algorithm (Pedro and Coimbra (2012), Chu *et al* (2014)).

The MLP models are trained against historical data with the goal minimize the RMSE between the MLP outputs training measured data for the target variable (DNI, PO irradiance, PO). To obtain new prediction the model simply uses new measured data as input. A cross-validation method is implemented with the MLP to generate more accurate predictions. The training dataset is divided into N subsets (10 in this case). Then, one subset is used for validation and the rest subsets are used to train the MLP. The trained model is assessed on the validation subset and a RMSE is computed by comparing the predictions and the validation targets. This process is repeated N times, each time a different subset is used as the validation set. The final prediction is the average of the N trained models.

Support Vector Regression (SVR)

SVR is a form of Support Vector Machines (SVM) used for non-linear regression tasks. SVMs are a set of supervised learning methods used for classification, regression and outlier's detection. The advantages of SVMs are:

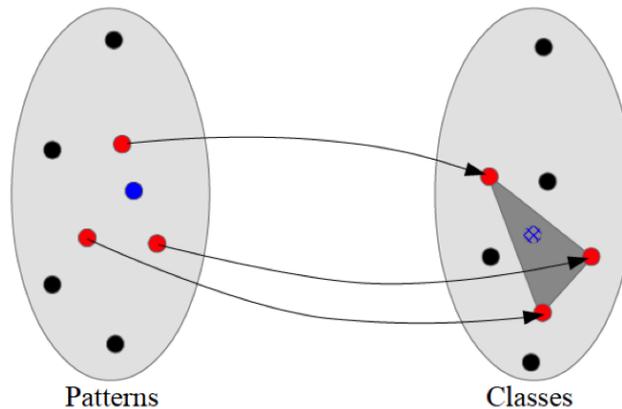
- Effective in high dimensional spaces as in the cases where there are multiple predictors for one forecast (e.g. telemetry, sky images, NWP to predict DNI).
- Still effective in cases where number of dimensions is greater than the number of samples, although that does not occur in this project.

- Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.
- Versatile: different Kernel functions can be specified for the decision function.

This project used this tool as implemented in LIBSVM (Chang and Lin (2011)) in several tasks. SVR models are trained in the same way as MPL models, as indicated in the previous section.

K-Nearest Neighbors (KNN)

Figure A3: KNN Model Schematic



The blue circle on the left represents a new forecast instance. In this case it represents the NAM forecast. The red circles represent the 3 nearest neighbors. The forecast - the hatched blue circle, is computed as the average of the 3 classes, in this case the measured historical irradiance corresponding to the 3 blue circles.

The KNN method is one of the simplest and most robust methods of pattern recognition in machine learning. In contrast to the MPL models that use the training data to estimate the free parameters in the model, the KNN uses the training data as the model and no free parameters need to be determined. The KNN forecast is based on identifying similarities between current data and historical data (Figure A3). This forecast is computed in three steps:

1. Identification of the list of the nearest neighbors;
2. Retrieval of historical irradiance data corresponding to those neighbors;
3. Aggregation of those data to produce a forecast.

In this work the nearest neighbors are identified by computing the distance between the current data (telemetry, NWP forecast, etc.) and historical data. Once the list of nearest neighbors is determined the model selects the corresponding irradiance or PO values and aggregates those values using a simple arithmetic average.

Gradient Boosting

The final type of forecast models used in this project is that obtained when utilizing a machine learning technique known as boosting. Boosting is achieved by training many

weak learners—simple models yielding high bias predictions—in succession: each new learner is trained based on the results of the previous learner. The overall model is then defined as a weighted ensemble of the weak learners. The ensemble constitutes a strong learner since its predictions yield a bias much lower than those of the individual weak learners. A general representation of this idea may be written as the recursion

$$F_{m+1}(x) = F_m(x) + \gamma_m h_m(x),$$

where F_m is the model at iteration (also, “round”) m , γ_m is the weighting attached to weak learner h_m , and x is a vector of features at a given time step. If a prediction is $\hat{y}_m = F_m(x)$, then the corresponding residual is $r_m = y - \hat{y}_m$, where y is the true observation. In theory, the recursion is terminated and the final model F_M is obtained when a time series vector of corresponding residuals r_M is approximately white (i.e., uncorrelated with itself for nonzero lags). In practice, the final model is obtained when the addition of weak learners worsens the model performance on a *testing* data set, which is a stricter requirement that reduces the tendency to overfit the model to noise dynamics.

When the weak learners are trained on the negative gradient of the loss function (i.e., in a direction that minimizes the loss function), the method is known as gradient boosting. To illustrate this idea, consider a standard quadratic loss function $(1/2)(y - F_m(x))^2$, which has the negative gradient $-(y - F_m(x)) = -(y - \hat{y}_m) = -r_m$. Chen and Guestrin (2016) recently developed a highly optimized backend for the efficient implementation of gradient boost models. This framework, leveraged in this work, is known as eXtreme Gradient Boosting (XGB) or XGBoost.

Features Derived Using Memory Operators for PO at CVSR

The features derived using fractional derivatives are best understood from an abstracted viewpoint. Whereas a standard derivative provides information about the instantaneous change behaviors of a given system, a fractional derivative provides information about historical behaviors of a system. Accordingly, it may be thought of as a derivative that possesses memory, or more simply, as a memory operator. If $D^{(\cdot)}$ denotes differentiation to order (\cdot) , then the appropriate fractional derivative definition is the Caputo derivative:

$$D^p f(t) \approx \int_{t_0}^t \frac{(t-s)^{m-p-1}}{\Gamma(m-p)} D^m f(s) ds,$$

where $f(t)$ represents a time series evaluated at the timestamp t , the derivative order $p \geq 0$ is a real number, and m is an integer such that $m - 1 < p < m$. The gamma function Γ is a generalization of the factorial function. The expression is generally approximate since a true memory operator should be evaluated for all times since the system was last quiescent. For irradiance time series, this implies that a true memory operator should be evaluated over all historical values since the last true nighttime

value, t_{dawn} (i.e., just before daybreak). However, since this is not computationally practical, a “finite memory” approximation is implemented:

$$t_0 = \max(t_{dawn}, t - \Delta),$$

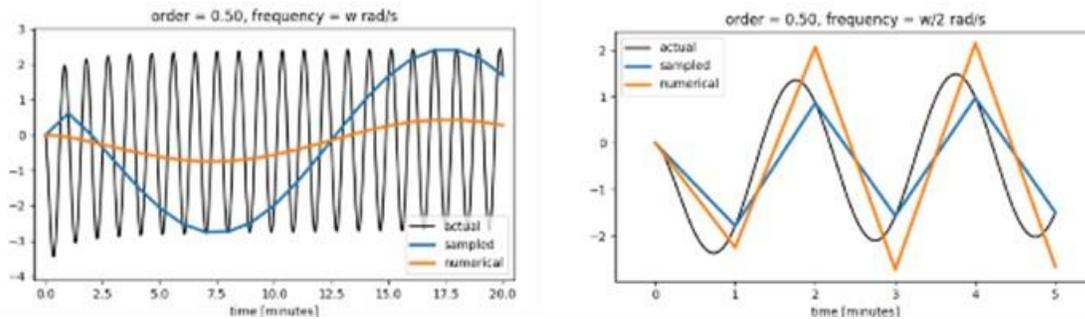
where Δ is the desired evaluation interval cutoff (i.e., the length of the finite memory). A brief analysis revealed that for $\Delta \geq 180$ minutes, the error in the approximation is negligible. Numerical evaluation of the Caputo derivative is achieved by weighted summation

$$D^p f_N \approx \sum_{n=0}^N a_n f_n,$$

where $f_n = f(t_n)$ and the historical weighting factors a_n are determined by an appropriate numerical differintegration algorithm. Fractionally integrated feature sets were computed using a similar method applied to the Riemann-Liouville fractional integral.

The numerical weighting algorithm was developed proprietarily for this work based on the requirement that the accuracy must be congruent with the characteristic dynamical timescales being observed. Since the primary feature sets used in this work are based on a 5-minute backward-averaged discretization the method should provide a commensurate level of accuracy at this timescale. An analysis of the dynamical timescale resolution, sampling rate, and convergence accuracy of the algorithm is given in Figure A4.

Figure A4: Timescale Resolution Analysis



The plots show the alignment of three relevant timescales when evaluating the $p = 0.5$ order derivative of a sinusoid: simulated dynamics being observed (actual), sensor sampling (sampled), and convergence of the numerical algorithm (numerical). The characteristic frequency of the dynamics is halved in each plot from left to right, top to bottom. The upper right plot represents sampling in accordance with the Nyquist-Shannon sampling theorem, for which the periodicity of the dynamics is accurately recovered. Sampling at roughly 1/5 the rate of the dynamics resolves both the periodicity and the amplitude of the signal. In each case, it is demonstrated that the accuracy of the memory operator is commensurate with that of the sampled signal.

The figure demonstrates that the algorithm for numerically evaluating the memory operator provides accuracy commensurate with the sensor sampling under for simulated sinusoidal dynamics having a period w . It is also shown that the frequency dependence of the dynamics is completely recovered when they are sampled at least as fast as $w/2$, which is a consequence of the Nyquist-Shannon sampling theorem. The amplitude is also accurately reproduced for sampling at least as fast as $\approx w/5$. This is important, since the baseline feature sets are computed from 5-minute backward averages of the raw data. Thus, the analysis in Figure A4 indicates that the memory operator features should be evaluated from data having at least 1-minute resolution. The necessary data was obtained by taking 1-minute backward averages of the raw telemetry.

In order to assess the effects of operator memory and order of differentiation or integration on the forecast performance, a case study was undertaken wherein forecasts were generated for all combinations of Δ and p generated from the baseline features with

- $\Delta \in \{10, 30, 60, 120, 180\}$ minutes, and
- $p \in [-2, 2]$ in increments of 0.1.

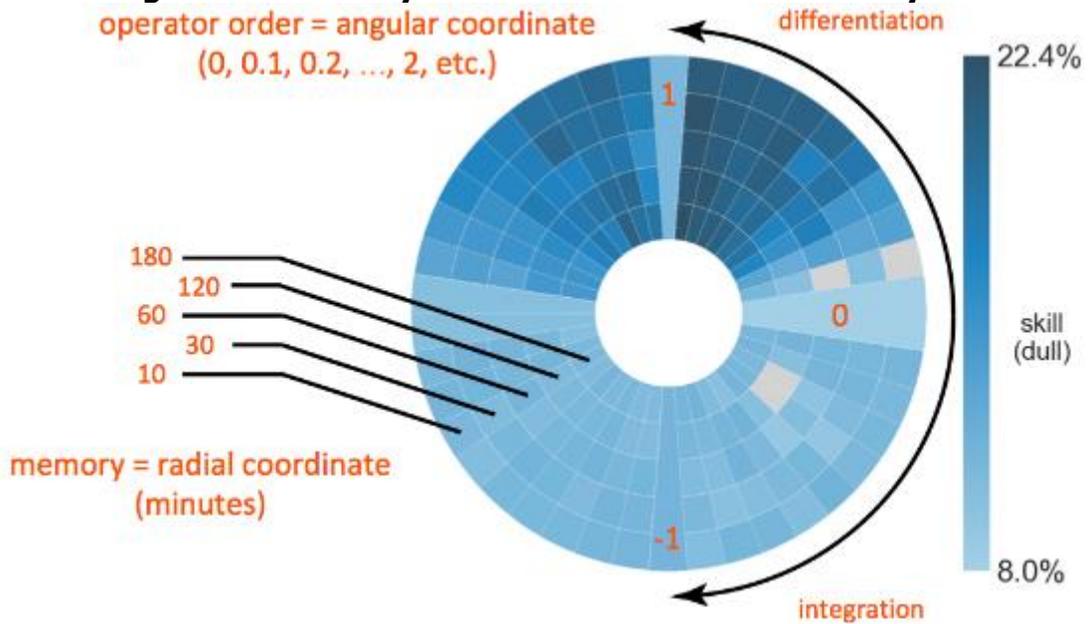
Sample results of this analysis for the 5-minute forecast horizon are demonstrated in Figure A5.

Among the many conclusions that can be drawn from the figure it can be said that:

1. integrated features do not generally yield a significant additional benefit,
2. integer order operators (i.e., simple finite differences) do not generally yield a significant additional benefit,
3. no apparent systematic improvement in performance is obtained by variation of the finite memory parameter Δ , and
4. there are evidently two distinct regions of fidelity at $p = 0.9$ and $p = 1.3$.

An identical analysis performed on the 10-30 minute forecast horizons produced similar results, with diminishing performance enhancements at increasing time horizons. The results of these analyses helped to inform both the feature set and model structure selection used further on.

Figure A5: Memory Feature Set Performance Analysis



Each contiguous cell in the diagram represents the performance of an individual forecast as determined by a naïve (or “dull”) persistence model with a 5-minute forecast horizon. Forecast performance resulting from a model utilizing only the baseline feature set determines the minimum on the performance scale and is represented in the diagram by the zero-order cell. All other cells represent baseline features along with the indicated feature set combination. Gray-colored cells designate feature set combinations for which no increase in performance was obtained

APPENDIX B:

PAPER ABSTRACTS

Abstracts for the journal papers based on this project:

H. T. C. Pedro, E. Lim and C. F. M. Coimbra (2018) "A Database Infrastructure to Implement Real-Time Solar and Wind Power Generation Intra-Hour Forecasts," *Renewable Energy* (123), pp. 513–525.

Abstract: This paper presents a simple forecasting database infrastructure implemented using the open-source database management system MySQL. This proposal aims at advancing the myriad of solar and wind forecast models present in the literature into a production stage. The paper gives all relevant details necessary to implement a MySQL infra-structure that collects the raw data, filters unrealistic values, classifies the data, and produces forecasts automatically and without the assistance of any other computational tools. The performance of this methodology is demonstrated by creating intra-hour power output forecasts for a 1 MW photovoltaic installation in Southern California and a 10 MW wind power plant in Central California. Several machine learning forecast models are implemented (persistence, auto-regressive and nearest neighbors) and tested. Both point forecasts and prediction intervals are generated with this methodology. Quantitative and qualitative analyses of solar and wind power forecasts were performed for an extended testing period (4 years and 6 years, respectively). Results show an acceptable and robust performance for the proposed forecasts.

H. T. C. Pedro, C. F. M. Coimbra, M. David and P. Lauret (2018) "Assessment of Machine Learning Techniques for Deterministic and Probabilistic Intra-Hour Solar Forecasts," *Renewable Energy* (123), pp. 191–203.

Abstract: This work compares the performance of machine learning methods (k-nearest-neighbors (kNN) and gradient boosting (GB)) in intra-hour forecasting of global (GHI) and direct normal (DNI) irradiances. The models predict the GHI and DNI and the corresponding prediction intervals. The data used in this work include pyranometer measurements of GHI and DNI and sky images. Point forecasts are evaluated using bulk error metrics while the performance of the probabilistic forecasts are quantified using metrics such as Prediction Interval Coverage Probability (PICP), Prediction Interval Normalized Averaged Width (PINAW) and the Continuous Ranked Probability Score (CRPS). Graphical verification displays like reliability diagram and rank histogram are used to assess the probabilistic forecasts. Results show that the machine learning models achieve significant forecast improvements over the reference model. The reduction in the RMSE translates into forecasting skills ranging between 8 percent and 24 percent, and 10 percent and 30 percent for the GHI and DNI testing set,

respectively. CRPS skill scores of 42 percent and 62 percent are obtained respectively for GHI and DNI probabilistic forecasts. Regarding the point forecasts, the GB method performs better than the kNN method when sky image features are included in the model. Conversely, for probabilistic forecasts the kNN exhibits rather good performance.

D. P. Larson and C. F. M. Coimbra (2018) "Direct Power Output Forecasts from Remote Sensing Image Processing," ASME Journal of Solar Energy Engineering—Transactions of the ASME 140(2), Article Number: 021011.

Abstract: A direct methodology for intra-day forecasts (1–6 h ahead) of power output (PO) from photovoltaic (PV) solar plants is proposed. The forecasting methodology uses publicly available images from geosynchronous satellites to predict PO directly without resorting to intermediate irradiance (resource) forecasting. Forecasts are evaluated using four years (January 2012–December 2015) of hourly PO data from 2 nontracking, 1 MWp PV plants in California. For both sites, the proposed methodology achieves forecasting skills ranging from 24 percent to 69 percent relative to reference persistence model results, with root-mean-square error (RMSE) values ranging from 90 to 136 kW across the studied horizons. Additionally, we consider the performance of the proposed methodology when applied to imagery from the next generation of geosynchronous satellites, e.g., Himawari-8 and geostationary operational environmental satellite (GOES-R).

M. Li, Y. Jiang, and C. F. M. Coimbra (2017) "On the Determination of Atmospheric Longwave Irradiance Under All-Sky Conditions," Solar Energy (144), pp. 40–48.

Abstract: In this work we review and recalibrate existing models and present a novel comprehensive model for estimation of the downward atmospheric longwave (LW) radiation for clear and cloudy sky conditions. LW radiation is an essential component of thermal balances in the atmosphere, playing also a substantial role in the design and operation of solar power plants. Unlike solar irradiance, LW irradiance is not measured routinely by meteorological or solar irradiance sensor networks. In most cases, it must be calculated indirectly from meteorological variables using simple parametric models. Under clear skies, fifteen parametric models for calculating LW irradiance are compared and recalibrated. All models achieve higher accuracy after grid search recalibration, and we show that many of the previously proposed LW models collapse into only a few different families of models. A recalibrated Brunt-family model is recommended for future use due to its simplicity and high accuracy (rRMSE = 4.37 percent). To account for the difference in nighttime and daytime clear-sky emissivities, nighttime and daytime Brunt-type models are proposed. Under all sky conditions, the information of clouds is represented by cloud cover fraction (CF) or cloud modification factor (CMF, available only during daytime). Three parametric models proposed in the bibliography are compared and calibrated, and a new model is proposed to account for the

alternation of vertical atmosphere profile by clouds. The proposed all-sky model has 3.8–31.8 percent lower RMSEs than the other three recalibrated models. If GHI irradiance measurements are available, using CMF as a parameter yields 7.5 percent lower RMSEs than using CF. For different applications that require LW information during daytime and/or nighttime, coefficients of the proposed models are corrected for diurnal and nocturnal use.

Y. Chu and C. F. M. Coimbra (2017) “Short-Term Probabilistic Forecasts for Direct Normal Irradiance,” *Renewable Energy* (101), pp. 526–536.

Abstract: A k-nearest neighbor (kNN) ensemble model has been developed to generate Probability Density Function (PDF) forecasts for intra-hour Direct Normal Irradiance (DNI). This probabilistic forecasting model, which uses diffuse irradiance measurements and cloud cover information as exogenous feature inputs, adaptively provides arbitrary PDF forecasts for different weather conditions. The proposed models have been quantitatively evaluated using data from different locations characterized by different climates (continental, coastal, and island). The performance of the forecasts is quantified using metrics such as Prediction Interval Coverage Probability (PICP), Prediction Interval Normalized Averaged Width (PINAW), Brier Skill Score (BSS), and the Continuous Ranked Probability Score (CRPS), and other standard error metrics. A persistence ensemble probabilistic forecasting model and a Gaussian probabilistic forecasting model are employed to benchmark the performance of the proposed kNN ensemble model. The results show that the proposed model significantly outperform both reference models in terms of all evaluation metrics for all locations when the forecast horizon is greater than 5-min. In addition, the proposed model shows superior performance in predicting DNI ramps.

Y. Chu, M. Li and C. F. M. Coimbra (2016) “Sun-Tracking Imaging System for Intra-Hour DNI Forecasts” *Renewable Energy* (96), Part A, pp. 792–799.

Abstract: A Sun-tracking imaging system is implemented for minimizing circumsolar image distortion for improved short-term solar irradiance forecasts. This sky-imaging system consists of a fisheye digital camera mounted on an automatic solar tracker that follows the diurnal pattern of the Sun. The Sun is located at the geometric center of the sky images where the fisheye distortion is minimized. Images from this new system provide more information about the circumsolar sky cover, which provides critical information for intra-hour solar forecasts, particularly for direct normal irradiance. An automatic masking algorithm has been developed to separate the sky area from ground obstacles and the image edges for each image that is collected. Then numerical image features are extracted from the segmented sky area and are used as exogenous inputs to MultiLayer Perceptron (MLP) models for direct normal irradiance forecasts. Sixty-seven days of irradiance and image measurements are used to train, optimize, and assess the MLP-based forecast models for solar irradiance. The results show that the

MLP forecasts based on the newly proposed sky-imaging system significantly outperform the reference models in terms of statistical metrics and forecast skill, particularly for shorter horizons, achieving forecast skills 18 percent–50 percent higher than the skills of a reference MLP-based model that is based on a zenith-pointed, stationary sky-imaging system.

M. Li, Y. Chu, H. T. C. Pedro and C. F. M. Coimbra (2016) "Quantitative Evaluation of the Impact of Cloud Transmittance and Cloud Velocity on the Accuracy of Short-Term DNI Forecasts," Renewable Energy (86) pp. 1362–1371

Abstract: Ground based sky imaging and irradiance sensors are used to quantitatively evaluate the impact of cloud transmittance and cloud velocity on the accuracy of short-term direct normal irradiance (DNI) forecasts. Eight representative partly-cloudy days are used as an evaluation dataset. Results show that incorporating real-time sky and cloud transmittances as inputs reduces the root mean square error (RMSE) of forecasts of both the Deterministic model (Det) (16.3 percent~ 17.8 percent reduction) and the multi-layer perceptron network model (MLP) (0.8 percent ~ 6.2 percent reduction). Four computer vision methods: the particle image velocimetry method, the optical flow method, the x-correlation method and the scale-invariant feature transform method have accuracies of 83.9 percent, 83.5 percent, 79.2 percent and 60.9 percent in deriving cloud velocity, with respect to manual detection. Analysis also shows that the cloud velocity has significant impact on the accuracy of DNI forecasts: underestimating the cloud velocity magnitude by 50 percent results in 30.2 percent (Det) and 24.2 percent (MLP) increase of forecast RMSE; a 50 percent overestimate results in 7.0 percent (Det) and 8.4 percent (MLP) increase of RMSE; a $\pm 30^\circ$ deviation of cloud velocity direction increases the forecast RMSE by 6.2 percent (Det) and 6.6 percent (MLP).

H. T. C. Pedro and C. F. M. Coimbra (2015) "Short-Term Irradiance Forecastability for Various Solar Micro-Climates," Solar Energy (122), pp. 587–602.

Abstract: The purpose of this work is to present a simple global solar irradiance forecasting framework based on the optimization of the k-nearest-neighbors (kNN) and artificial neural networks algorithms (ANN) for time horizons ranging from 15 min to 2 h. We apply the proposed forecasting models to irradiance from five locations and assessed the impact of different micro-climates on forecasting performance. We also propose two metrics, the density of large irradiance ramps and the time series determinism, to characterize the irradiance forecastability. Both measures are computed from the irradiance time series and provide a good indication for the forecasting performance before any predictions are produced. Results show that the proposed kNN and ANN models achieve substantial improvements relative to simpler forecasting

models. The results also show that the optimal parameters for the kNN and ANN models are highly dependent on the different micro-climates. Finally, we show that the density of large irradiance ramps and time series determinism can successfully explain the forecasting performance for the different locations and time horizons.

R. H. Inman, J. G. Edson and C. F. M. Coimbra (2015) "Impact of Local Turbidity Estimation on Clear Sky Models for Direct Normal Irradiance," *Solar Energy* (117) pp. 125–138.

Abstract: Clear-sky modeling is of critical importance for the accurate determination of Direct Normal Irradiance (DNI), which is the relevant component of the solar irradiance for concentrated solar energy applications. Accurate clear-sky modeling of DNI is typically best achieved through the separate consideration of water vapor and aerosol concentrations in the atmosphere. Highly resolved temporal measurements of such quantities is typically not available unless a meteorological station is located in close proximity. When this type of data is not available, attenuating effects on the direct beam are modeled by Linke turbidity-equivalent factors, which can be obtained from broadband observations of DNI under cloudless skies. We present a novel algorithm that allows for a time-resolved estimation of the average daily Linke turbidity factor from ground-based DNI observations under cloudless skies. This requires a method of identifying clear-sky periods in the observational time series (in order to avoid cloud contamination) as well as a broadband turbidity-based clear-sky model for implicit turbidity calculations. While the method can be applied to the correction of historical clear-sky models for a given site, the true value lies in the forecasting of DNI under cloudless skies through the assumption of a persistence of average daily turbidity. This technique is applied at seven stations spread across the states of California, Washington, and Hawaii while using several years of data from 2010 to 2014. Performance of the forecast is evaluated by way of the relative Root Mean Square Error (rRMSE) and relative Mean Bias Error (rMBE), both as a function of solar zenith angle, and benchmarked against monthly climatologies of turbidity information. Results suggest that rRMSE and rMBE of the method are typically smaller than 5 percent for both historical and forecasted CSMs, which compare favorably against the 10–20 percent range that is typical for monthly climatologies.

H. T. C. Pedro and C. F. M. Coimbra (2015) "Nearest-Neighbor Methodology for Prediction of Intra-Hour Global Horizontal and Direct Normal Irradiances," *Renewable Energy* (80) pp. 770-782

Abstract: This work proposes a novel forecast methodology for intra-hour solar irradiance based on optimized pattern recognition from local telemetry and sky imaging. The model, based on the k-nearest-neighbors (kNN) algorithm, predicts the global (GHI) and direct (DNI) components of irradiance for horizons ranging from 5 min up to 30 min, and the corresponding uncertainty prediction intervals. An optimization

algorithm determines the best set of patterns and other free parameters in the model, such as the number of nearest neighbors. Results show that the model achieves significant forecast improvements (between 10 percent and 25 percent) over a reference persistence forecast. The results show that large ramps in the irradiance time series are not very well capture by the point forecasts, mostly because those events are underrepresented in the historical dataset. The inclusion of sky images in the pattern recognition results in a small improvement (below 5 percent) relative to the kNN without images, but it helps in the definition of the uncertainty intervals (especially in the case of DNI). The prediction intervals determined with this method show good performance, with high probability coverage (≈ 90 percent for GHI and ≈ 85 percent for DNI) and narrow average normalized width (≈ 8 percent for GHI and ≈ 17 percent for DNI).