Sandia National Laboratories

LDRD LABORATORY DIRECTED RESEARCH & DEVELOPMENT

Using Machine Learning for Predictive Modeling of Weather Impacts on Utility-scale Photovoltaic Systems

Hector Mendoza[1*] and Nicole D. Jackson[1]

[1]Sandia National Laboratories, *hmendo@sandia.gov

Motivation

- Solar energy is projected to be a critical source of energy for the United States in the future
- Climate change is expected to increase the frequency and severity of extreme weather events
- As more energy generation is expected from photovoltaic (PV) sites, it is important for operators and planners to **forecast impacts of [extreme] weather events** on PV energy production
- **Accurate predictive models** are important to enable efficient grid operation

Current Approaches & Limitations

- SNL developed predictive PV energy model shows high error compared to measured data
- The persistence method as well as statical, machine learning, and hybrid approaches have been used for forecasting PV
- Support vector machines used to predict solar intensity

> Limited locations:
Few sites or a single region in a country are used

> Limited times series of data:
Common to have 0.5, 1, or only 2 years of site-level data

> Simplified weather events:
combination of sunny, cloudy, rainy, heavy rain; clear sky versus overcast

Study Objectives

Study Focus:
Develop a targeted predictive energy model for a single type of extreme weather event for utility-scale PV sites across multiple states in the US

Methods

Study Sites

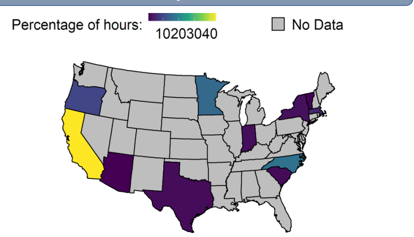

Figure 1. Distribution of production hours from PVROM.

Table 1. Summary of PVROM sites used in this study.

Descriptor	Value
Number of sites	118
Total site capacity in GWDC (GWAC)	1.4 (1.1)
Record ranges	02/2018 – 03/2019
Number of hours	476,567
Percent of hours above 77F (25C)	23.8
Percent of hours above 90F (32.2C)	8.3
Percent of hours above 100F (37.8C)	1.4

Work Flow

PV **Production** Data from PVROM

Site-level **Weather** Data from NSRDB

Data fusion of PV production data with weather data

Model development via machine learning:
- Baseline ML model based on Gunda & Cai (2019)
- New ML-based models using updated data and RF model

Model selection:
≥ **25% reduction** in mean absolute error compared to baseline ML model

Sensitivity analysis

Figure 2. Overview of datasets, data processing, and modeling approach used in this study.

Preliminary Study Findings

Baseline Model Re-Analysis

- Variables: irradiance, DC:AC ratio, ambient temperature, array type, hours, month, site age, PV climate zone

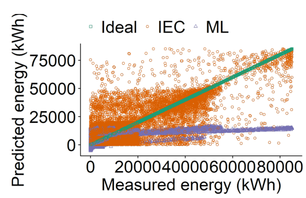

Figure 3. Comparison between prior models.

Table 2. Baseline model selection criteria.

	Mean Absolute Error (kWh)	
Descriptor	Original	Target
Overall	2595	1946
T > 90F	2707	2030

Models in Development

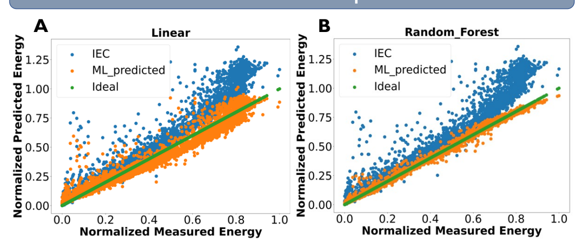

Figure 4. ML models in development using a) linear regression, and b) random forests for a subset of PVROM data.

Table 3. Comparison of preliminary model performance.

Model	MAE (kWh)	MAE (normalized)
IEC	1318	0.1114
Linear	520	0.0439
RF	123	0.0104

Ongoing and Future Work

- Understand larger variance in ML models for lower values of energy generated by PV systems
- Pending automated hyperparameter optimization for existing models to reduce predictive model variance
- ML models improve upon baseline standards (e.g., IEC), but there is still room for improvement
- Use deep learning to capture non-linear dependencies of PV energy production on weather, site parameters

U.S. DEPARTMENT OF ENERGY
NNSA National Nuclear Security Administration

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.
SAND No. SAND2022-11237 C

Sandia National Laboratories