Data Analytics for Solar Energy Management

Lipyeow Lim[1], Duane Stevens[2], Sen Chiao, Christopher Foo[1], Anthony Chang[2], Todd Taomae[1], Carlos Andrade[1], Neha Gupta[1], Gabriella Santillan[2], Michael Gonsalves[2], Lei Zhang[2]

[1]Information & Computer Sciences Department, University of Hawaii at Manoa

[2]Atmospheric Sciences Department, University of Hawaii at Manoa

[3]Department of Meteorology and Climate Science, San Jose State University

Abstract

The recent surge of solar energy generation systems on both the residential and commercial level in the state of Hawaii has catapulted Hawaii to one of the states with the highest penetration of grid-linked solar photovoltaic systems. It is well known that the variability in solar energy generation can introduce a variety of problems for Hawaii's grid operators. Unlike traditional oil-based generator based systems, solar energy generation is largely beyond the control of the operators as they are dependent on a myriad of factors like the weather. This adds a level of unpredictability to the net energy generation on the grid and further exacerbates the delicate balancing act that takes place between energy consumers and generators in order to ensure the stability of the grid.

To alleviate some of this uncertainty, energy operators can (and do) use various dynamic models such as the Weather Research and Forecast (WRF) model for predicting the evolution of local weather. These dynamic models utilize observations of many different types, including surface sensors that we focus on, as initial conditions to project forward in time a prediction of the evolving weather using the laws of physics. The laws of physics are typically represented as nonlinear partial differential equations in space and time, but require further approximations such as discretization of differential equations, data assimilation, and subgrid scale parameterizations of physical processes. The errors inherent in these approximations lead us to ask whether statistical methods alone (such as data mining) without explicit dynamics might be useful in characterizing relationships among sensor data and solar irradiance, both in space and time.

Here we have leveraged this sensor data using simple machine learning and data mining techniques to aid with the prediction and analysis of the generative capabilities of solar energy in Hawaii. In this paper, we discuss the preprocessing, prediction and analysis techniques we applied and report on the results of the application of those techniques on sensor data from several weather stations on the island of Oahu.

For preprocessing, we have applied various techniques for dealing with the seasonality (diurnal and annual) of the time series sensor data and found that those techniques when used with the appropriate prediction and analysis techniques do help to improve accuracy.

For prediction, we have applied linear regression and a non-linear cubist (decision tree) techniques and found that reasonable short term irradiance forecasts are possible with only a handful of dependent variables.

For analysis, we have applied a variety of analysis techniques including clustering, Markov chains, and information-theoretic dependency measures. Clustering analysis on daily irradiance vectors has yielded clusters that match the intuitive understanding of grid operators. Further analysis of those clusters over time using Markov chains yielded interesting insights into the persistence of weather/irradiance patterns at the temporal granularity of days. Further information-theoretic analysis also yielded interesting probabilistic relationships between irradiance and several weather-related variables.

We believe our prediction and analysis techniques can be easily adapted to other regions and can yield analysis results that are insightful for grid operators.