24-786: Bayesian Machine Learning

Photovoltaic Production Forecasting

Mechanical Engineering
Carnegie Mellon University

Carnegie Mellon University
Scott Institute for Energy Innovation

Kian Gruszczynski, Wolf Thomsen, Clement Wong, Ali Alp Gurer

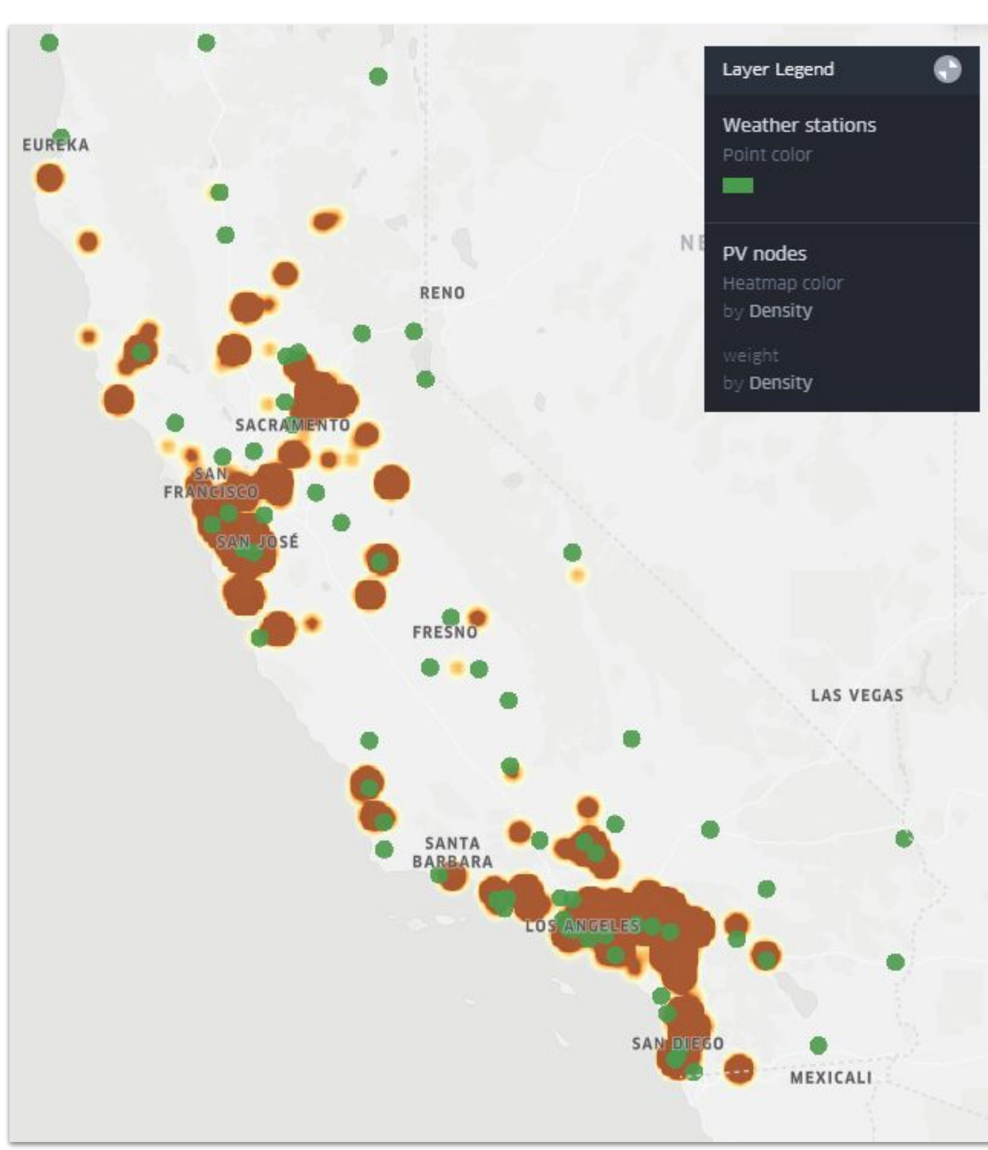
INTRODUCTION

Motivation

One of the biggest challenges for a fully renewable energy supply is the lack of reliability. While traditional energy sources can be controlled, influencing the weather is still out of human control.
With this project, we aim to instead predict the hourly energy production for the next 24 hours. This will allow energy suppliers to optimize the energy grid in advance and plan optimal transport routes.

Origin and Characteristics of the Data

Through a collaboration with the company *Kevala,* we have access to photovoltaic production data from over 200 different postal codes, measured quarter-hourly, and weather data from 70 different weather stations. The density of photovoltaic systems and the individual weather stations are shown in the map below.



METHODS

Data Processing

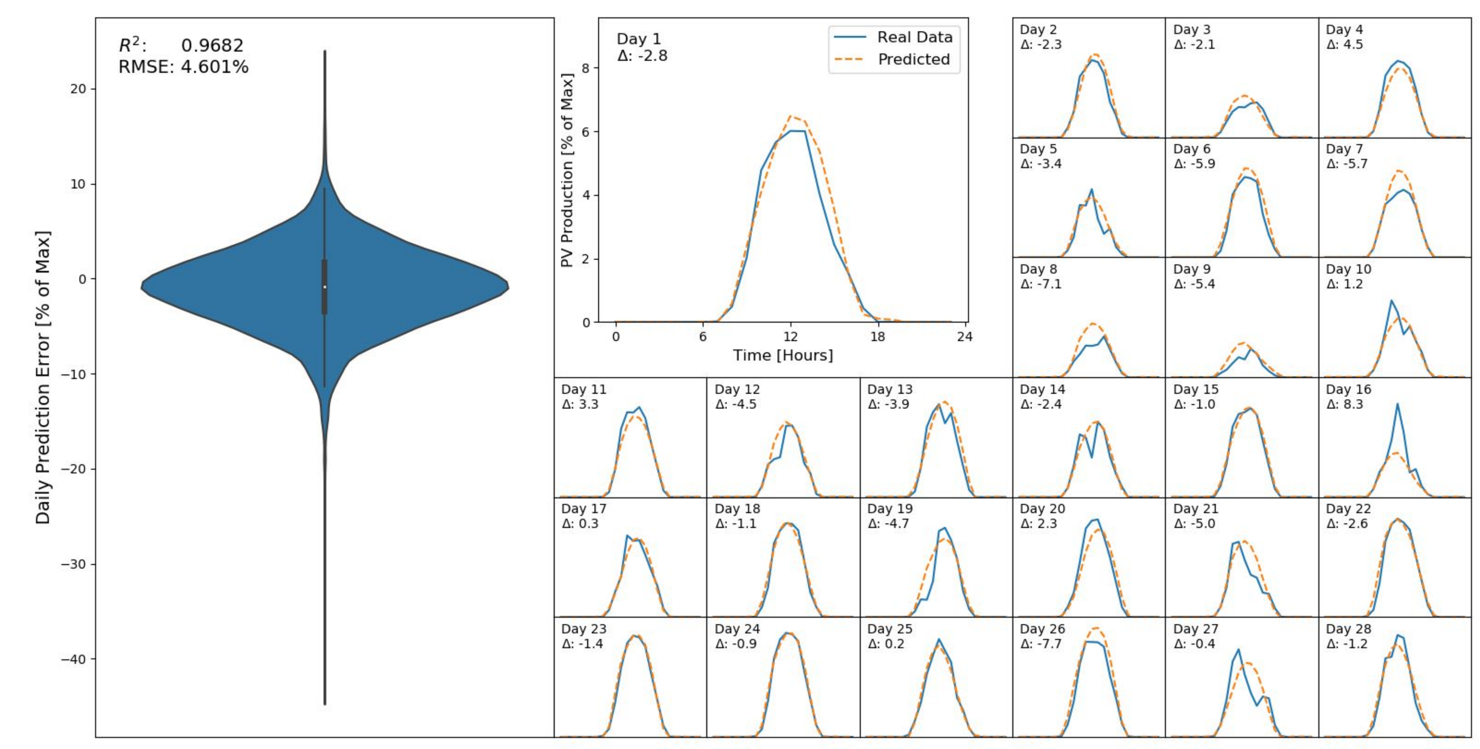
Before developing the models, we first transformed the data into an hourly dataframe of photovoltaic generation and weather data. To do this, we developed a wrapper for the Kevala API that took in yearly chunks of generation data and weather data from the nearest weather station. The data was stitched together and reduced to 1 hour intervals to match the weather data. In addition, we filled instances of missing weather data by interpolating surrounding values. Finally, we normalized the photovoltaic generation data by the maximum daily production potential of that node. The maximum value was provided through the Kevala API.

Model Development

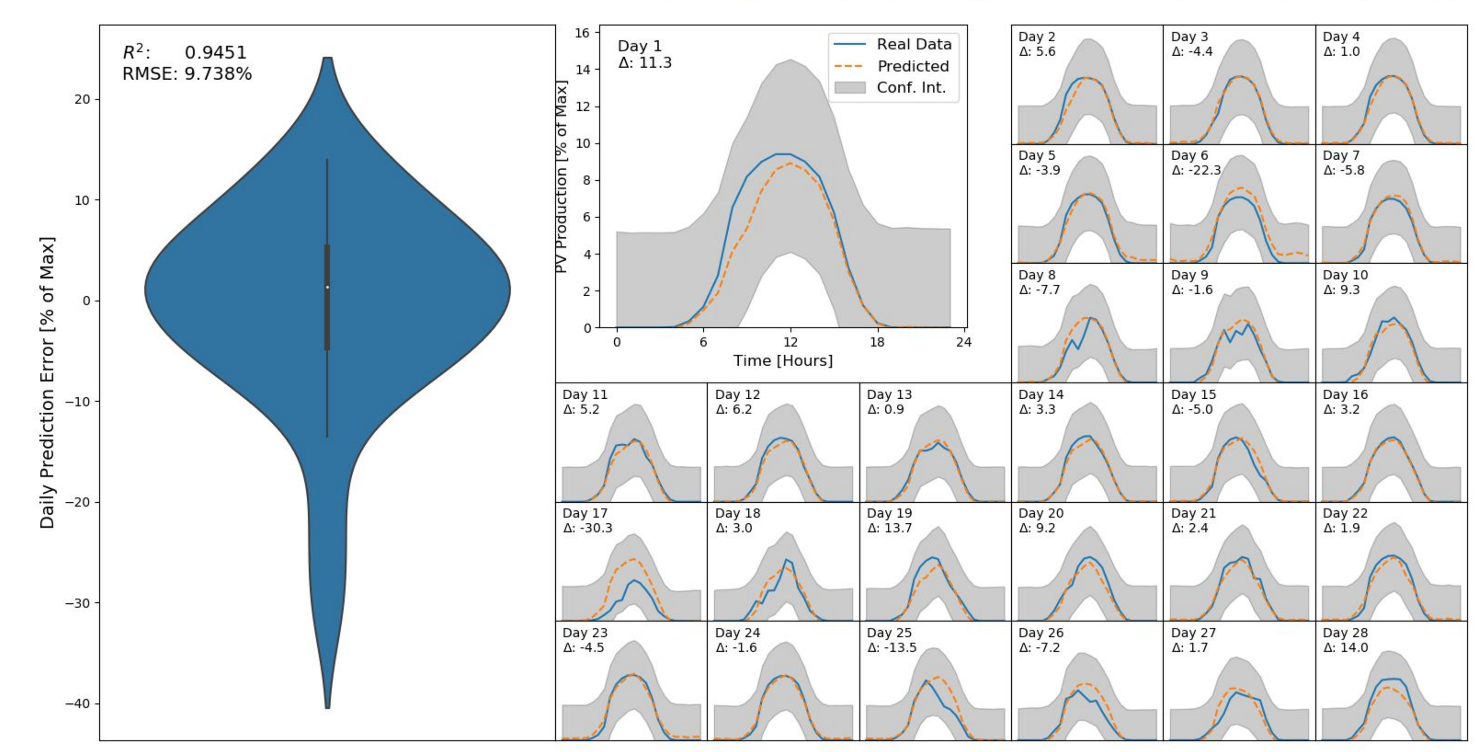
We focused on three model types: SARIMAX, a staple time series model from economics, Dynamic Linear Models that use a Kalman Filter, and feed-forward neural networks. The first two models have the advantage that they require comparatively few parameters and give confidence intervals alongside the prediction. Simultaneously, they cannot model the very complex relationships that neural networks can learn from data.

RESULTS

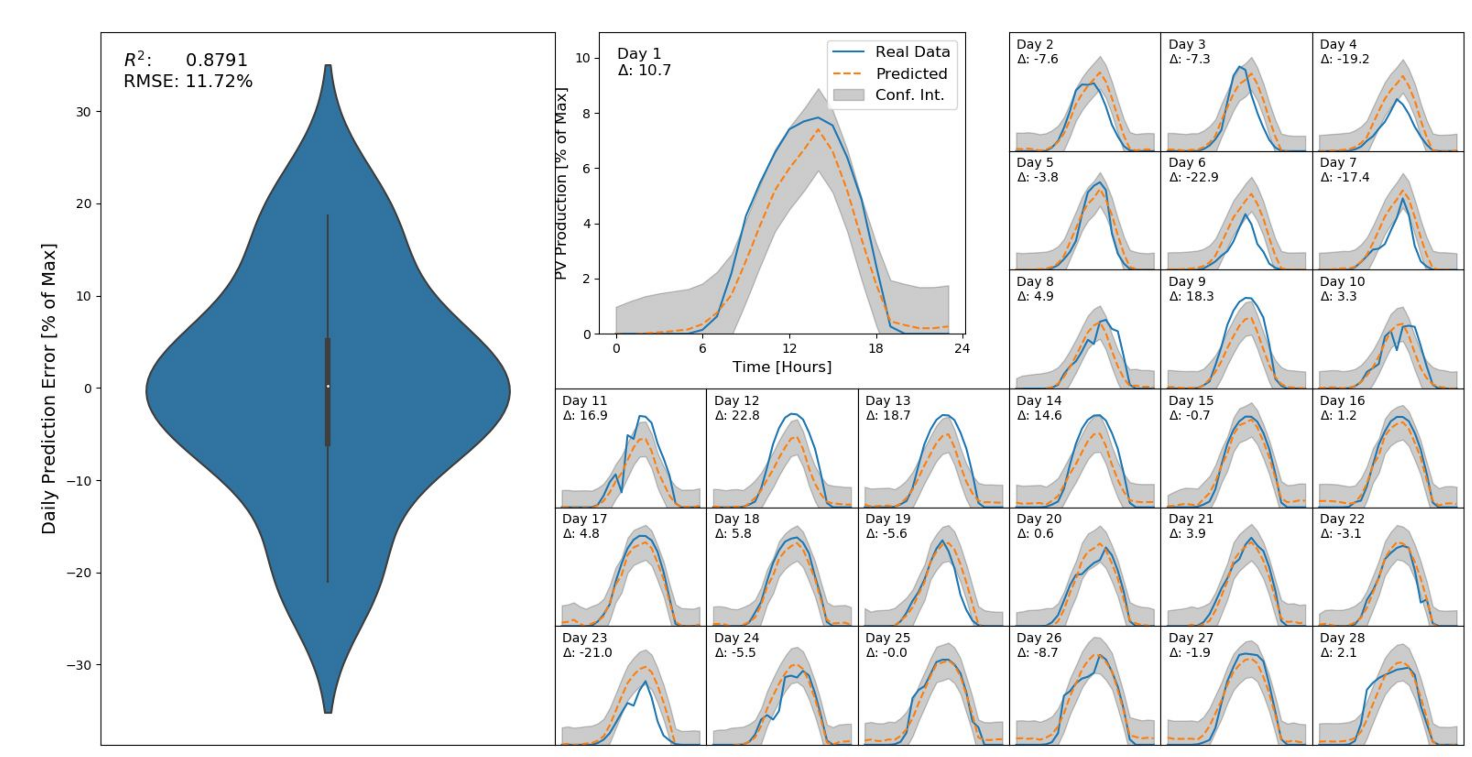
Feed-Forward Neural Network



Dynamic Linear Model



SARIMAX



CONCLUSIONS/ DISCUSSIONS

All three model types perform very well when predicting the total production for 24 hours. Simultaneously, none of the models are able to predict hourly fluctuations. This can be attributed to the quality of the weather data. In particular for the DLM and the SARIMAX models, the coefficients returned for the weather data tend to be small, and only influence the general trend for the day instead of dominating the seasonal part.
For the neural network, this analysis is not possible as easily, but the fact that the model behaves very similar in its predictions indicates that it discovered similar underlying relationships.
Generally, the neural network clearly outperforms the other models, but does this at the cost of not providing confidence intervals. Ideally, by leveraging Bayesian neural networks, we will be able to get the high accuracy of the neural network predictions alongside confidence intervals.
Achieving this will allow robust estimations of the best and worst case scenarios for the energy grid and corresponding planning.