Forecasting PV Power Time Series Data

Bennet Meyers, Júlio Hoffimann
CS229, Final Project, December 2017

SLAC NATIONAL ACCELERATOR LABORATORY

Team ID: 724

Background

- Solar photovoltaic (PV) systems show variability in output across all time scales, from seasonal changes in available sunlight to a cloud moving in front of the sun in a matter of seconds. PV forecasting techniques have been developed to address this variability, and various techniques exist for forecasting at different time-horizons, from 10s of minutes to days (Sophie Pelland, 2013).
- Short term forecasting is primarily concerned with predicting and understanding rapid changes in power output from a PV system, events called "ramps". Ramp events are large positive or negative deviations from a long-term trend over a short time period (Raffi Sevlian, 2013).
- Correctly identifying an upcoming ramp events is important for maintaining load and generation balance and can be utilized to develop more efficient scheduling algorithms for controlling grid-connected energy storage systems.
- Our data was provided by the Grid Integration Systems and Mobility Lab at SLAC through the Visualization and Analytics for Distributed Energy Resources (VADER) program
- The provided data was an unstructured set of data files from many PV system in the Southern California/Orange County Area, including system AC power at 5-minute intervals and system location (lat/long).

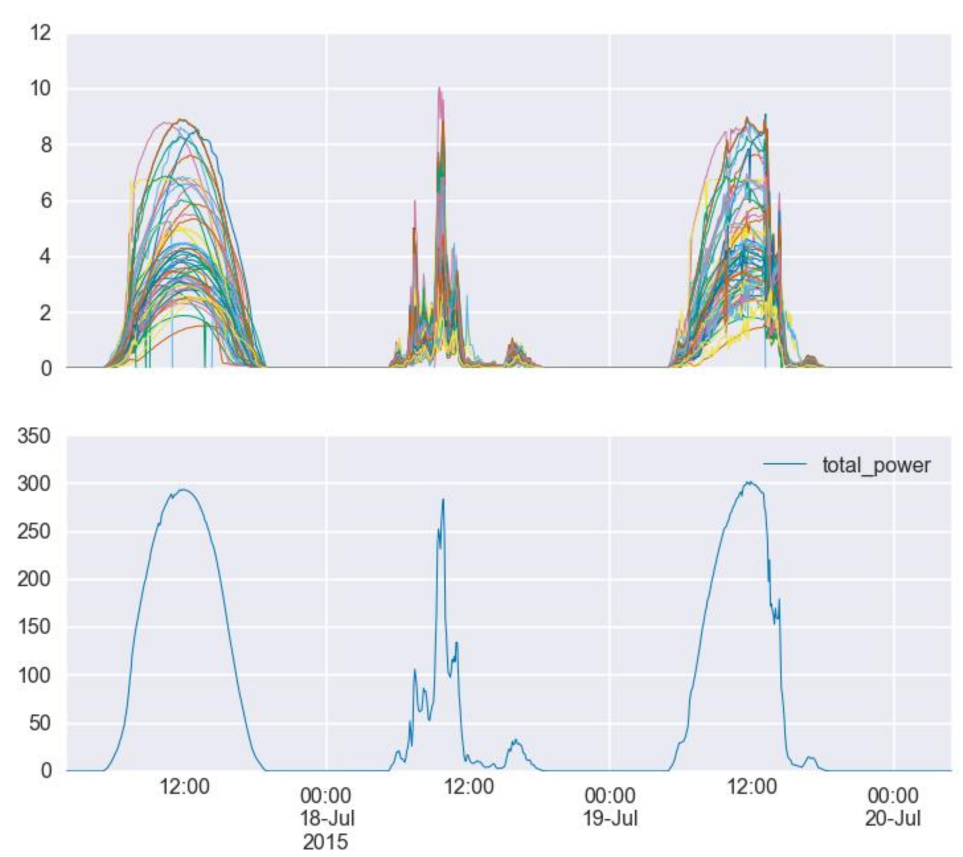

Fig. 1: A selection of three days from the data set, showing the individual system power (top) and the aggregate regional power (bottom).

Research Objectives

- Forecast aggregate PV power in a region on a three-hour time horizon at 5-minutes steps based on observed individual system data from the previous three hours.
- Understand how different machine learning models perform on this problem.

Methodology

Data Collection and Cleaning

- The data were provided as uncleaned and unverified raw data files for all available systems monitored by SunPower in a certain geographic region in Southern California.
- Downselected from 218 sites to 62 sites by enforcing at least two years of data and >100,000 non-null data points.
- The 62 sites translates to 73 separate systems (multiple systems per site)
- 8 systems were subsequently dropped from data set due to data errors
- 65 systems selected for final data set

Methodology (contd.)

Problem Structure

- *One-to-one forecasts:* Use the observed total power for the previous three hours to predict the expected total power for the next three hours
- *Many-to-one forecasts:* Use the observed individual power across all systems for the previous three hours to predict total power for the next three hours

In both cases, the target is *vector-valued*. For a three-hour forecast horizon at 5-minute steps, we are predicting a vector in $\mathbf{R}^{36}$.

Models

- *Persistence (one-to-one):* A naïve forecaster that propagates the last observed value, unchanged, through the forecasting horizon. This is used as a baseline to compare other models to.
- *ARIMA (one-to-one)*: We explored a parameter space of autoregressive parameters between 1 and 30 (the number of lagged terms in the model), moving average parameters between 0 and 2 (the number of lagged error terms in the squared model), and integrative parameters between 0 and 2 (the number of differencing steps, deals with non- stationary data). The data are stationary, but integrative parameters were explored for completeness and to observe the effect on forecasts.
- *Functional Regression (many-to-one):* We implemented a functional regression model that takes a window of data from all systems as features, and returns the summed power as an output. Neighborhood size was picked empirically by choosing the lowest dev error over neighborhood sizes between 1 and 100 (Figs. 2 and 3).
- *Neural Networks (many-to-one):* Using TensorFlow and Keras, we explored many different fully-connected neural networks and began exploration of convolutional neural networks. The best performing fully connected NN has 2 hidden ReLU layers, containing 2000 neurons in the first hidden layer and 1000 neurons in the second, and the cost function includes L2 regularization (lambda=1e-4).

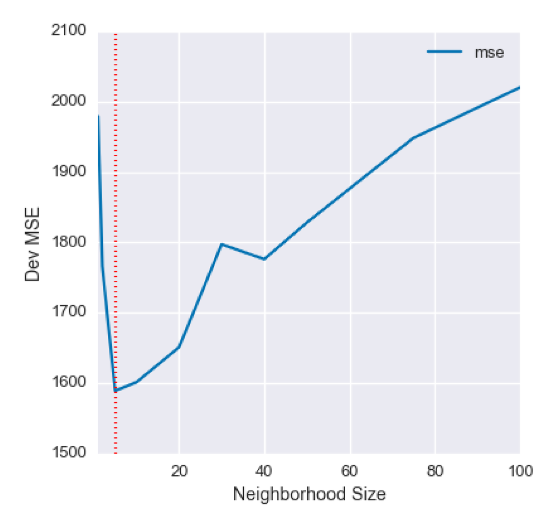

Fig. 2: Neighborhood size hyperparameter search for functional regression with the original data

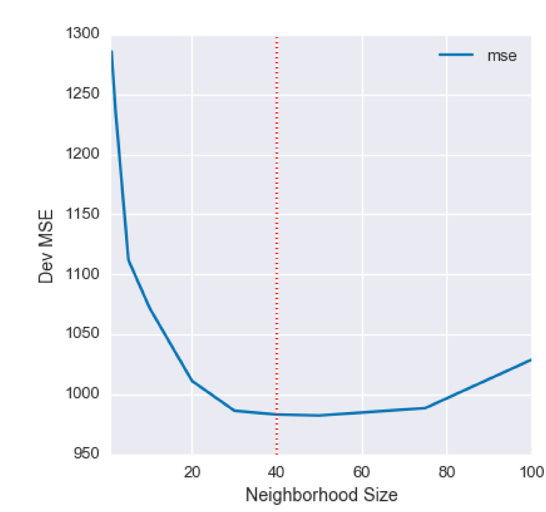

Fig. 3: Neighborhood size hyperparameter search for functional regression with the data detrended with the statistical clear sky fit

Domain Knowledge

- Periodic Detrending:
  - The data experience periodicity on a daily cycle and and a yearly cycle
  - Classic approach: sun position, transposition, and array performance models
  - Data-driven approach:. estimate the clear sky signal directly from observed data.
  - Form data matrix (split by day, each vector in $\mathbf{R}^{288}$), take SVD, take periodic smoothing fits of the daily scale factors for left singular vectors.
  - Construct clearsky with new fitted daily scale factors and left singular vectors
  - Lower dimensional reconstruction possible
- DoY and ToD
  - Day of year and time of day of the start of the forecast window
  - Distinguish between the middle of the night and the early morning

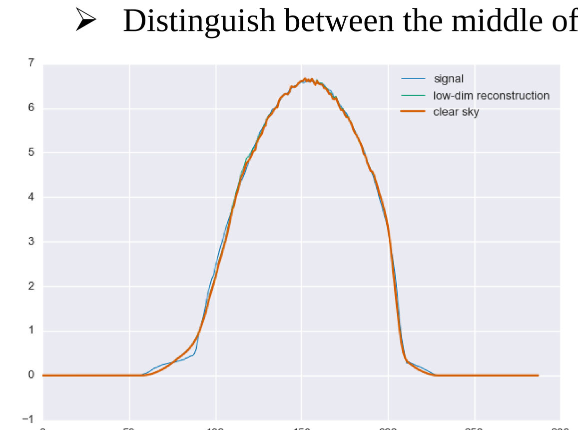

Fig. 4: Statistical clear sky signal compared to sunny day

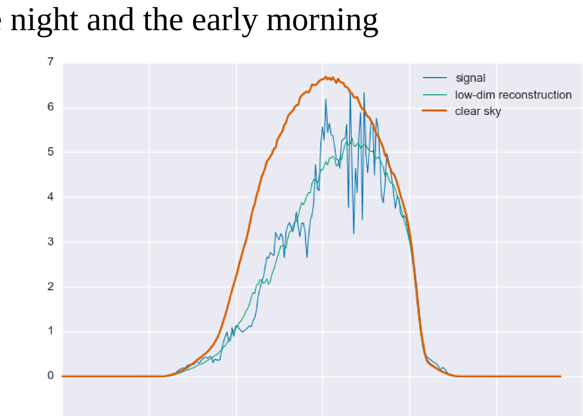

Fig. 5: Statistical clear sky signal compared to cloudy day

Results

Table 1: Mean-squared error on the small, 8-day test set.

Model	Original Data	Detrended Data	Detrended +DoY, ToD
Persistance	5635.0	1592.3	-
ARIMA	3021.9	1227.5	-
Functional Regression	1658.2	1434.1	-
Best Fully Connected NN	1045.5	672.4	510.7

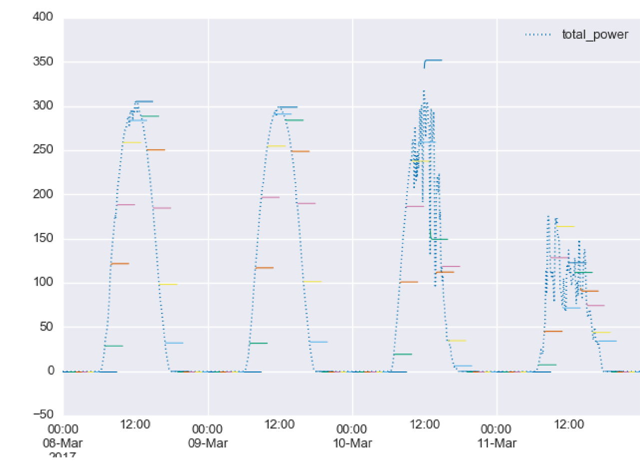

Fig. 7: Persistence forecast

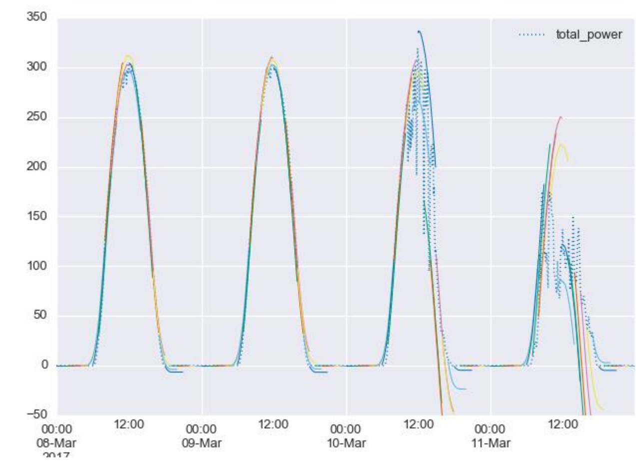

Fig. 8: Persistence forecast on detrended data

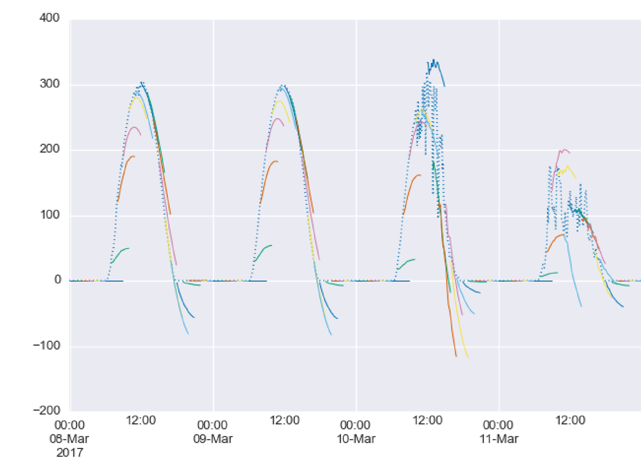

Fig. 9: ARIMA forecast (24, 0, 0)

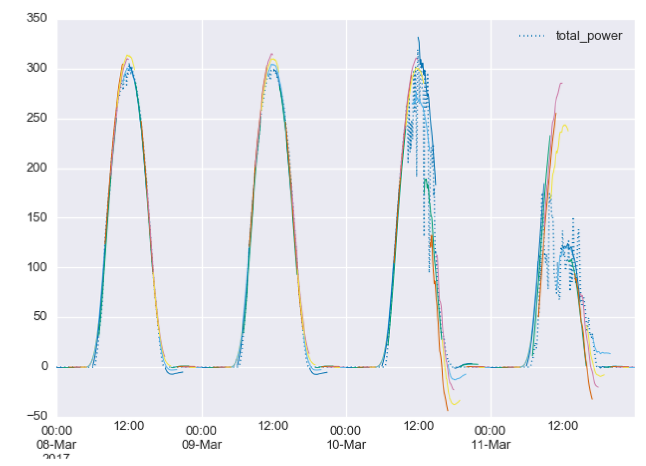

Fig. 10: ARIMA forecast on detrended data (30, 0, 0)

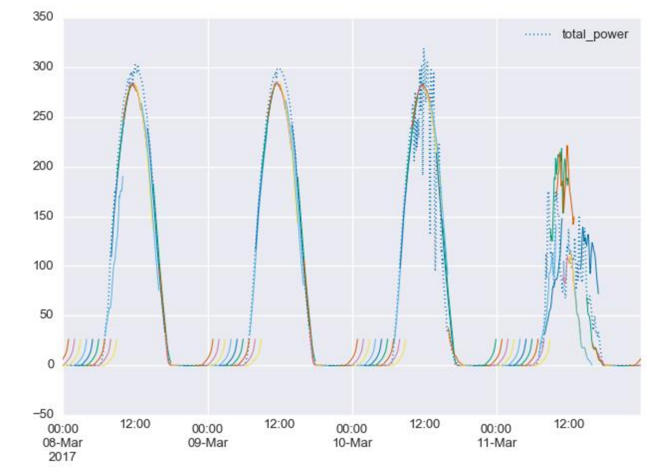

Fig. 11: Functional regression forecast

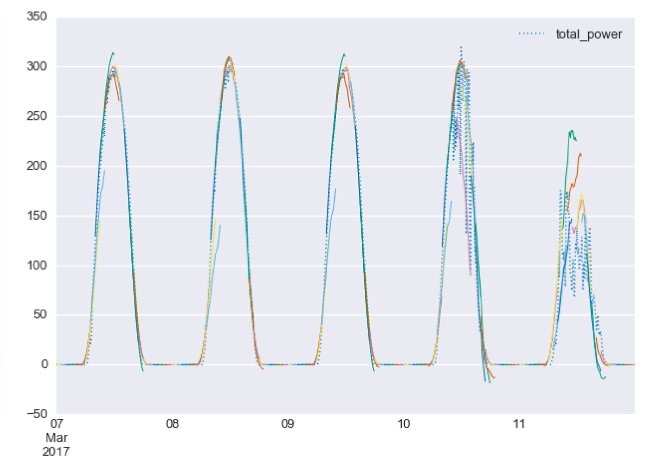

Fig. 12: Functional regression forecast on detrended data

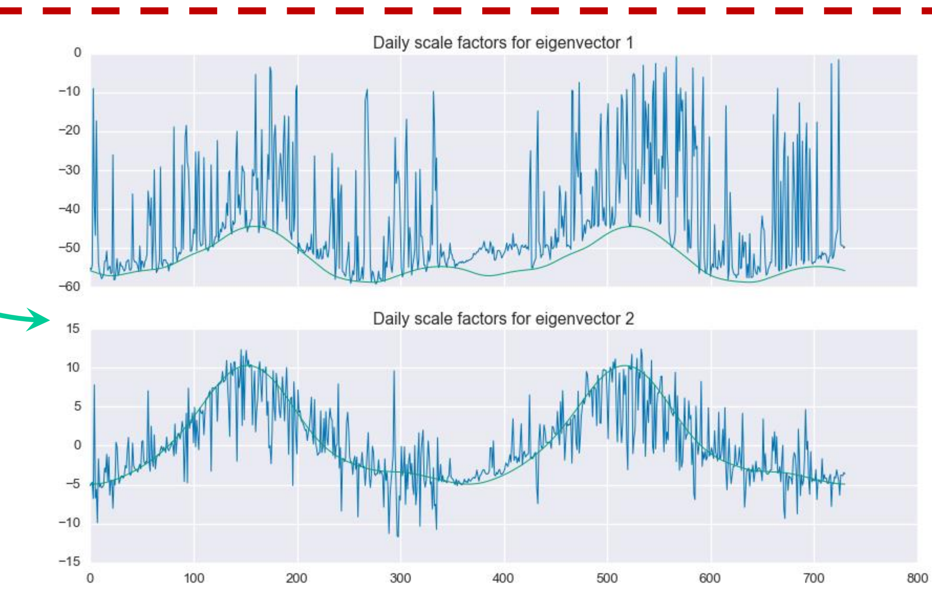

Fig. 6: Daily fitter scale factors for first two left singular vectors for clear sky fit

Results (contd.)

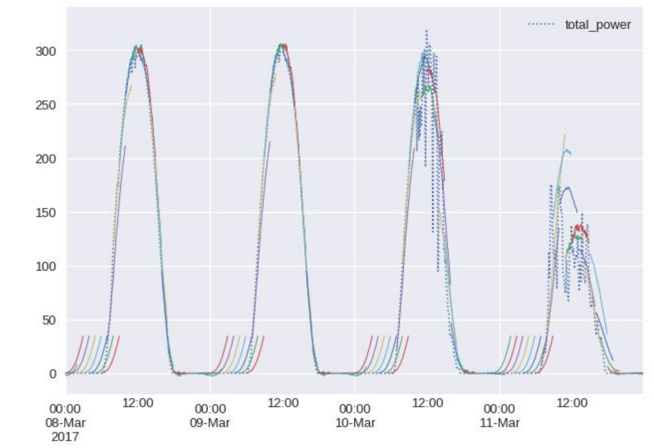

Fig. 13: Dense NN forecast (2000, 1000)

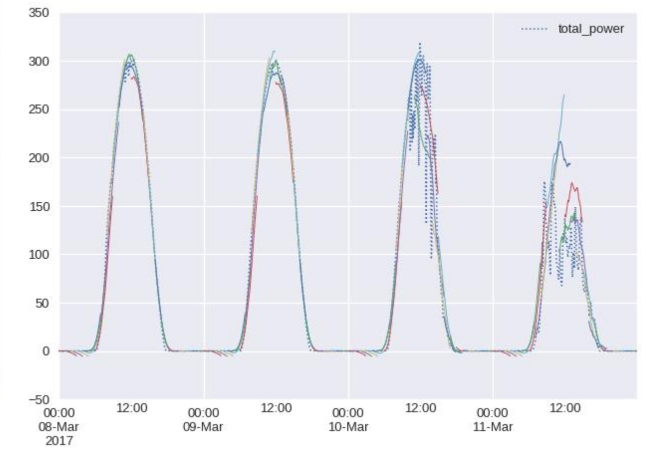

Fig. 14: Dense NN forecast on detrended data

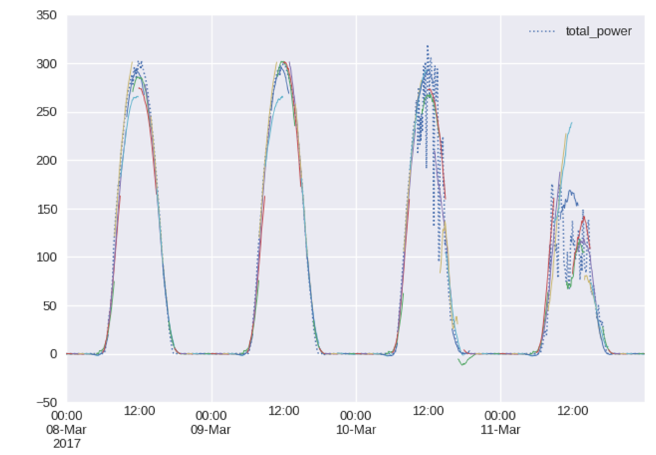

Fig. 15: Dense NN forecast on detrended data with DoY and ToD

Conclusions

- It is very difficult to forecast 5-minute time series data 3 hours into the future with high accuracy (no plots show perfect forecasts for cloudy periods)
- All forecasters struggle with day/night transition when periodic nature of data is unaccounted for
- Clear sky detrending improves MSE for all models
- All explored models outperform persistence forecasting
- Relatively simple NNs easily outperform ARIMA and functional regression models
- Neural networks have much more flexibility in terms of including new, heterogeneous data into the feature set (e.g. DoY, ToD)
- Including DoY and ToD data in the feature set provided additional improvement over clear sky detrending

Next Steps/Ongoing Work

- Successful training and testing of convolutional and recurrent neural networks
- Deeper look at ARIMA models and existing successful forecasters
- Test performance of 1:1 formulation with NNs to show improvement over many-to-one formulation explored in this project
- Inclusion of uncertainty output from predictors (investigating total variation as a good metric)
- More focus on ramp detection as a metric of success rather than overall MSE
- Benchmark against "easier" problems: smaller forecast vectors, larger time-steps
- Better utilization of spatial distribution information

Software

All software and notebooks are available at https://github.com/bmeyers/SolarForecasting

References

Sophie Pelland, Jan Kleissl Takashi Oozeki Karel De Brabandere, Jan Remund. Photovoltaic and Solar Forecasting: State of the Art. 1–40 In International Energy Agency: Photovoltaic Power Systems Programme. 2013.

Raffi Sevlian, Ram Rajagopal. Detection and statistics of wind power ramps. IEEE Transactions on Power Systems, 2013.