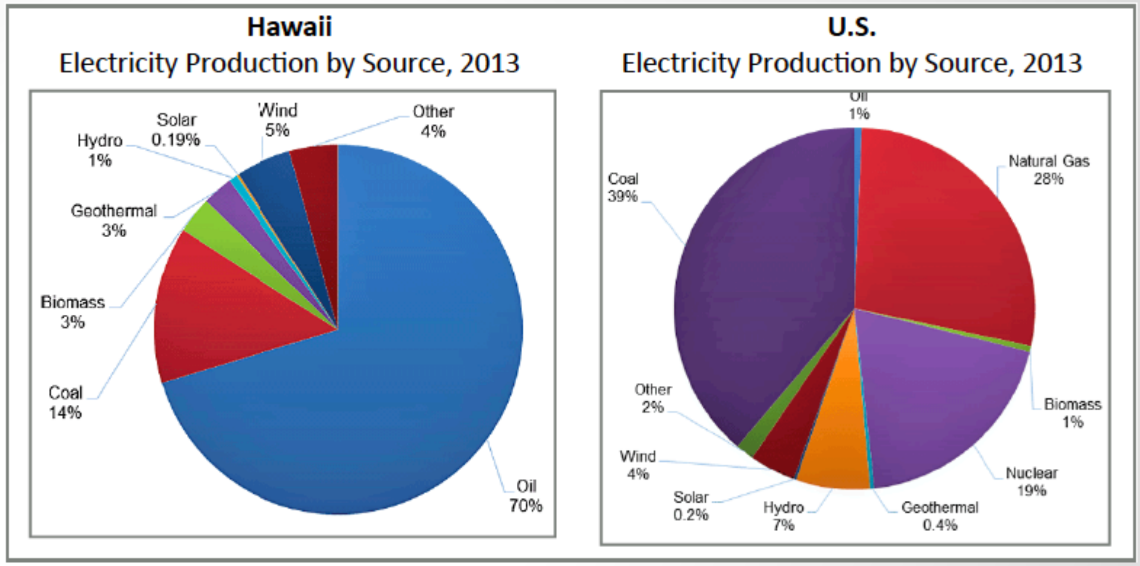
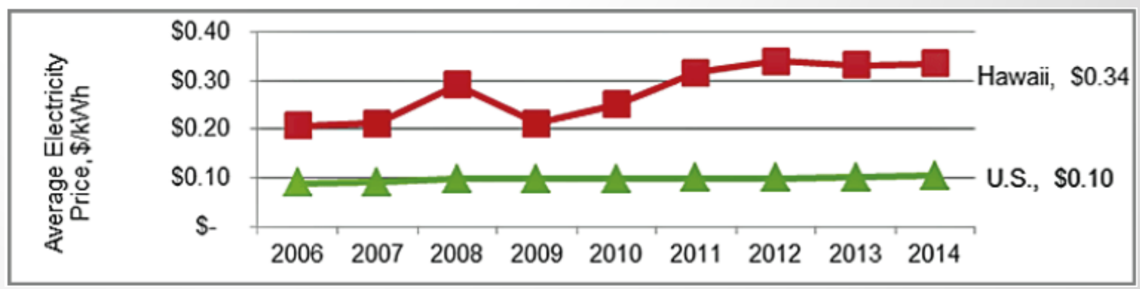
Probabilistic Models for One-Day Ahead Solar Irradiance Forecasting in Renewable Energy Applications

Carlos V. A. Silva, *ICS Dept, University of Hawai`i at Mānoa*
Lipyeow Lim, *ICS Dept, University of Hawai`i at Mānoa*
Duane Stevens, *Atmo. Sc. Dept, University of Hawai`i at Mānoa*
Dora Nakafuji, *Hawaiian Electric Company*

Energy in the State of Hawai`i

- State GOAL: 70% renewables by 2030.
- In 2013, Hawaii relied on oil for 70% of its energy.
- Hawaii's electricity cost is 3 times the US average



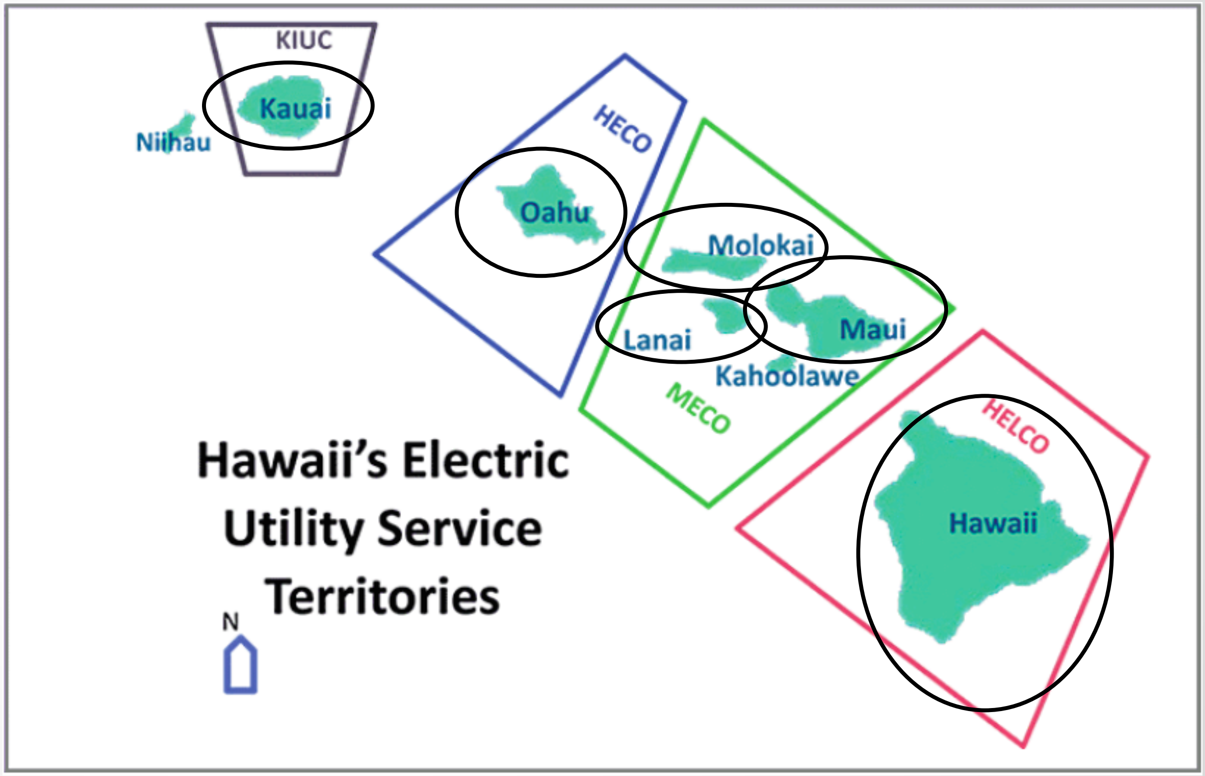


Disconnected Grids

Six independent grids: Kauai, Oahu, Molokai, Lanai, Maui, Hawaii.

UNLIKE MAINLAND

- Cannot sell excess production
- Cannot buy from neighbors to make up generation shortfall



The Problem with Renewables (Solar, Wind)

Operator of the power grids need to ensure that demand is met (while minimizing cost of power supply thereby maximizing profit).

- Demand (Load)
  - Consumers like us!
- Supply (Generation)
  - Conventional power plants
  - Solar/Wind Farms
  - Rooftop Solar

Some uncertainty, but well understood to some degree. In Hawaii, humid and hot weather can create load!

Deterministic, but takes many hours to bring up additional generation units if the load spikes.

Higher uncertainty due to weather, but geographically centralized.

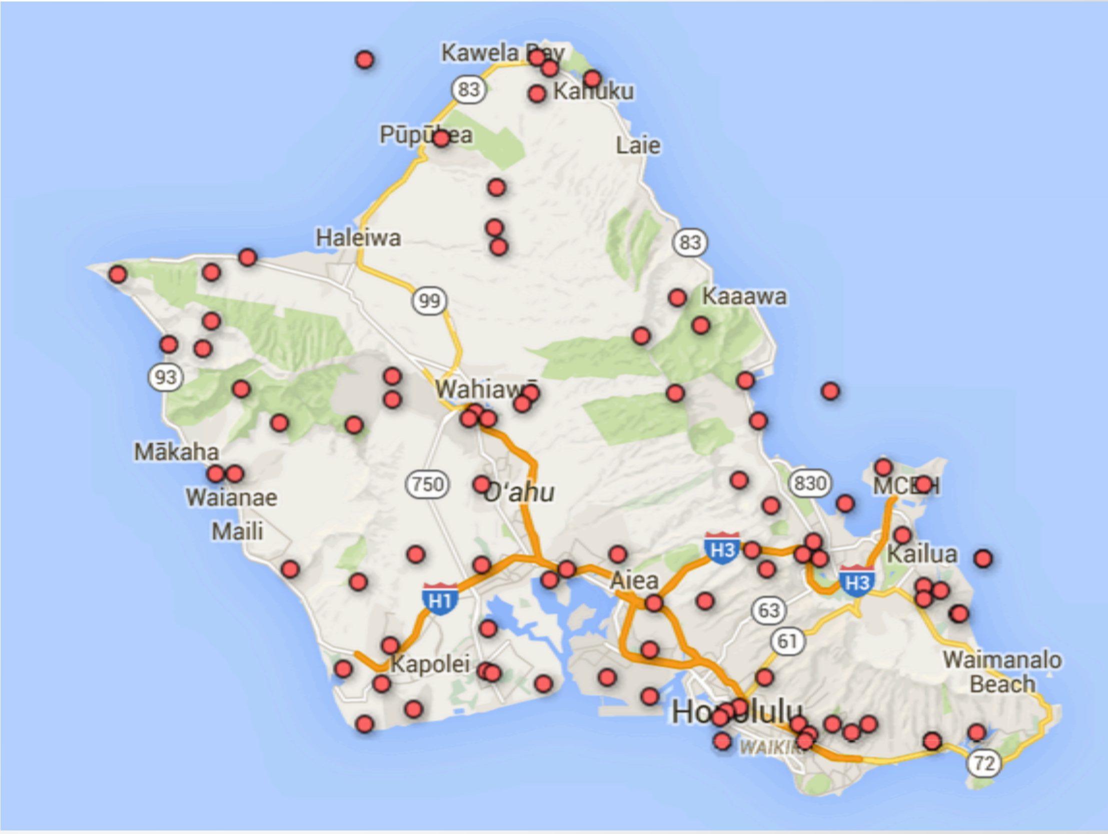
Higher uncertainty due to weather, but geographically distributed.

Weather (solar irradiance & wind) forecasting can lower the uncertainty!

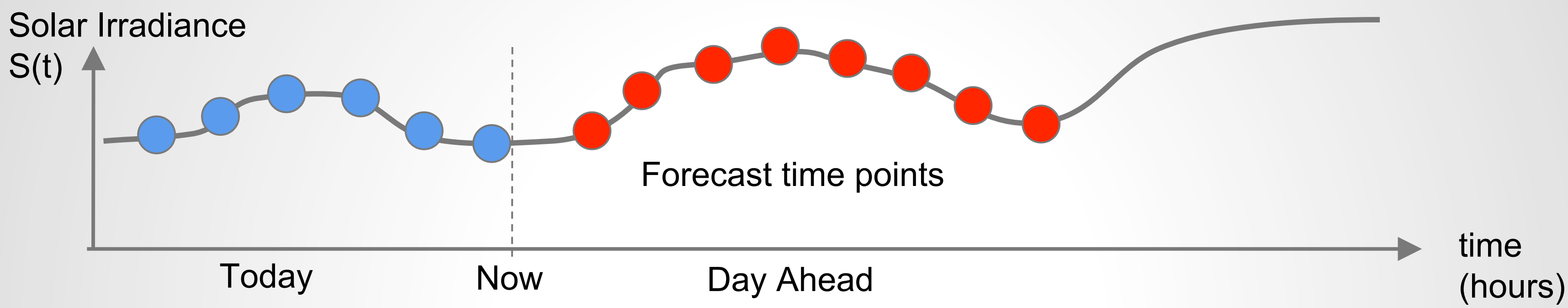
Weather Data Sources on the Island of O'ahu

SCBH1 Variable Description

TMPF Temperature,
RELH Relative Humidity,
SKNT Wind Speed,
GUST Wind Gust,
DRCT Wind Direction,
QFLG Quality check flag,
SOLR Solar Radiation,
TLKE Water Temperature,
PREC Precipitation accumulated,
SINT Snow interval,
FT Fuel Temperature,
FM10_hr_Fuel Moisture,
PEAK Peak_Wind Speed,
HI2424 Hr High Temperature,
LO2424 Hr Low Temperature,
PDIR Peak_Wind Direction,
VOLT Battery voltage



Problem Statement



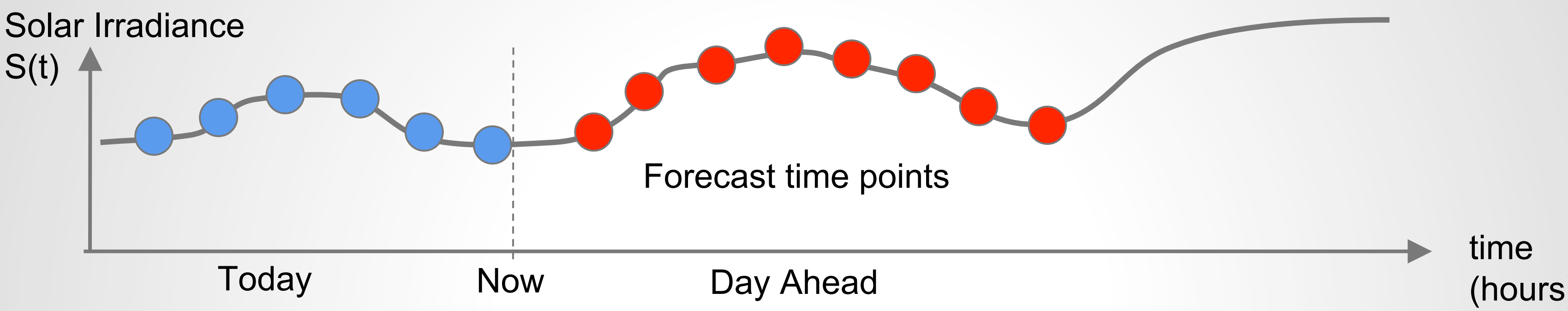
Weather station data (mainly solar irradiance) normalized to hourly samples. Given all sensor data today (sunset), predict the solar irradiance for the next day (8am-5pm).

- Probabilistic Models (including Naïve Bayes)
- Linear Regression

Evaluation Criteria

Mean Absolute Error (hourly)
MAE = Σ | Predicted – Actual |

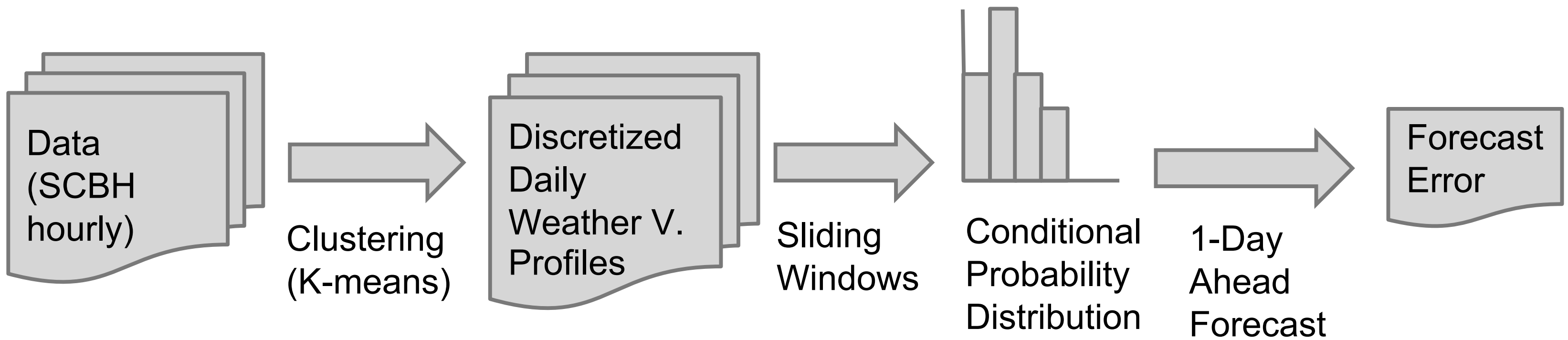
Linear Regression



Construct one LR model for each forecast time point (8am-5pm) the next day:

$$S_{20140214.0900} = c_1 \cdot S_{20140213.1700} + c_2 \cdot S_{20140213.1600} + c_3 \cdot S_{20140213.1500} + \ldots + c_{10} \cdot S_{20140213.0800} + c_{11}. \quad (1)$$

Probabilistic Models: Preprocessing



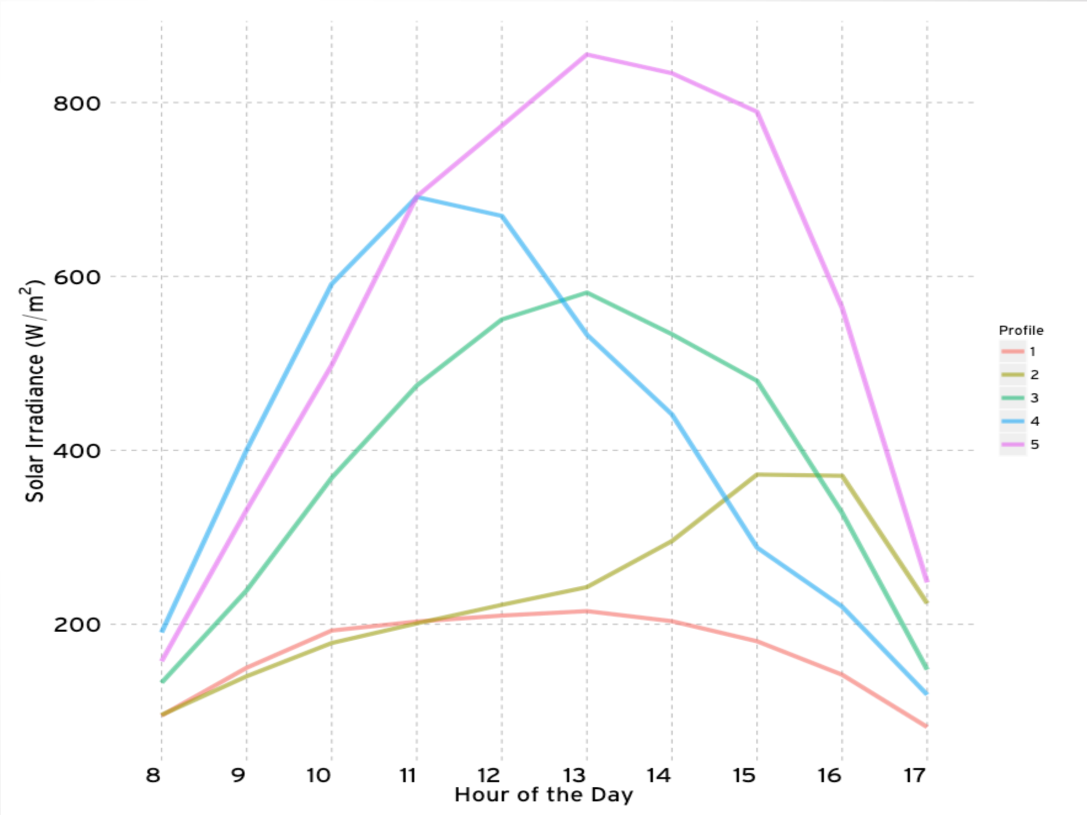
- Use clustering algorithms (K-means) to discretize the solar irradiance for each day into a discrete profile. K=5.
- Hourly data is transformed into a sequence of discrete profile IDs.
- Construct joint probability distributions for sequence assuming stationarity,

$$P(S_t, S_{t-1}, \ldots, S_{t-w+1})$$

Discretized Solar Irradiance Profiles

- Scoffield Station (SCBH1) using data from 2012-2013

- K-means (best of 100 runs)



Probabilistic Models: Prediction

- After getting distributions from historical data
- Naïve Bayes:

$$\hat{s} = \arg\max_{s} P(S_t{=}s) \prod_{i=1}^{w-1} P(S_{t-i}|S_t{=}s)$$

- Fixed-Order Markov models (w is fixed)

$$\begin{aligned}\hat{s} \quad &= \quad \arg\max_{s} P(S_t{=}s|S_{t-1}{=}s_1, S_{t-2}{=}s_2, \ldots, \\ &\qquad S_{t-w+1}{=}s_{w-1}).\end{aligned}$$

Probabilistic Models: Variable Order

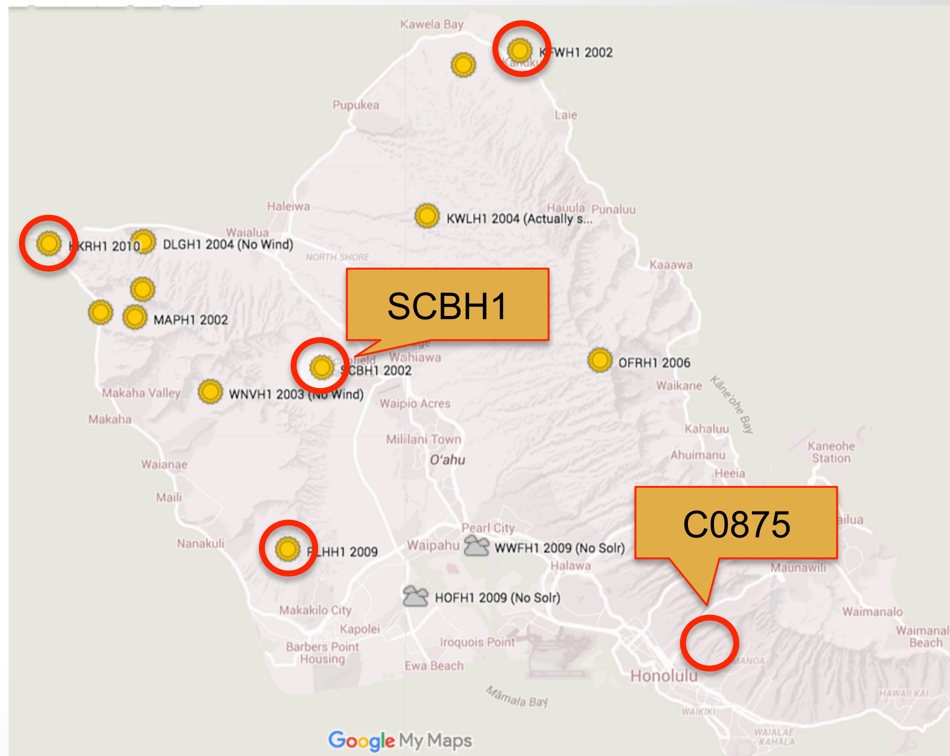
- Fixed-Order: $$\hat{s} \quad = \quad \arg\max_{s} P(S_t{=}s | S_{t-1}{=}s_1, S_{t-2}{=}s_2, \ldots, S_{t-w+1}{=}s_{w-1}).$$

- Variable-Order Markov models (w is chosen dynamically)
  - using entropy $$\hat{w} = \arg\min_{w} H(w)$$
  - Entropy+Support $$\hat{w} = \arg\min_{w} \frac{H(w)}{N(s_1, s_2, \ldots, s_{w-1})}$$

Experiments

Data:
- Training: 2012, 2013
- Testing: 2014
- 5 Stations

Error Measure
- Mean Absolute Error

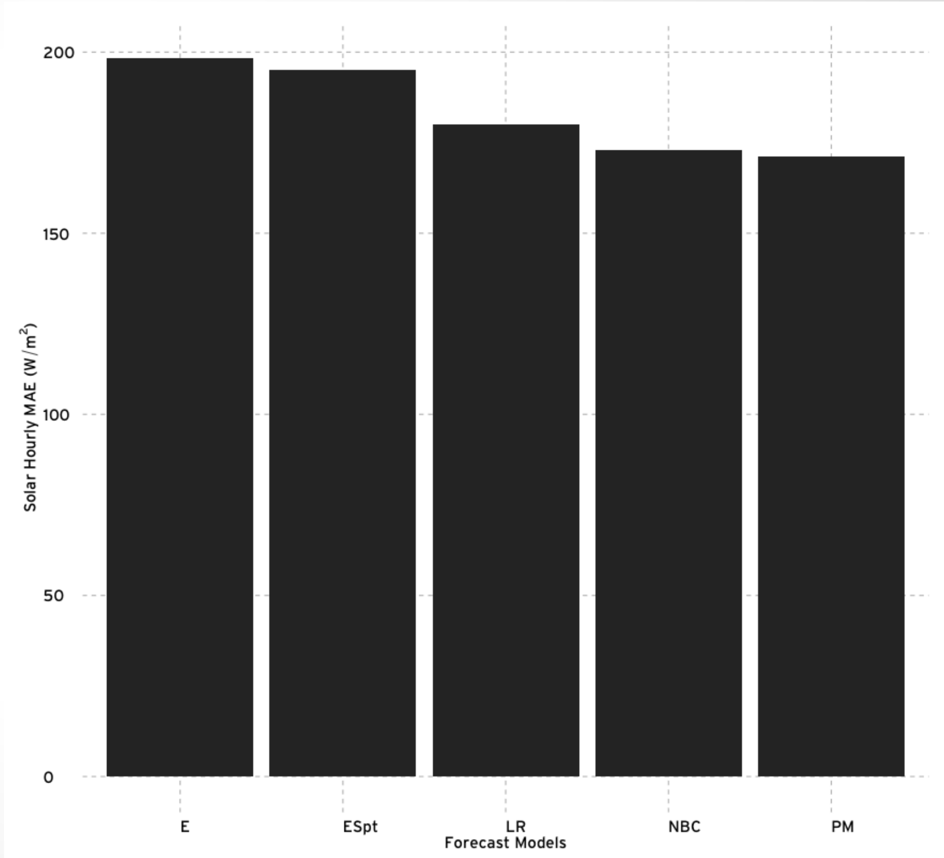


Overall Performance

- SCBH1 station

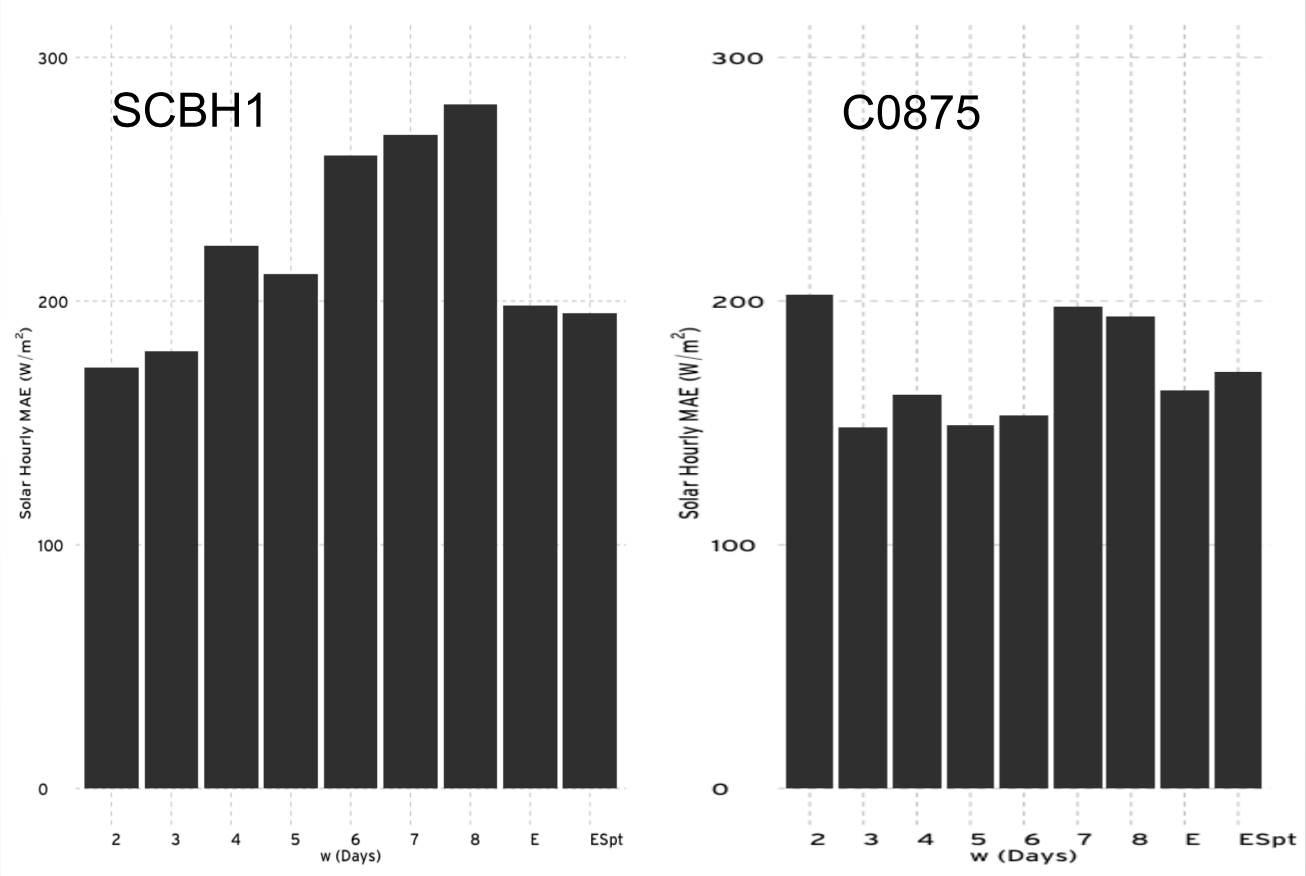
- Probabilistic with fixed w=2 has lowest error

- Despite high average errors, entropy & entropy +support are better predictors of cloudy days



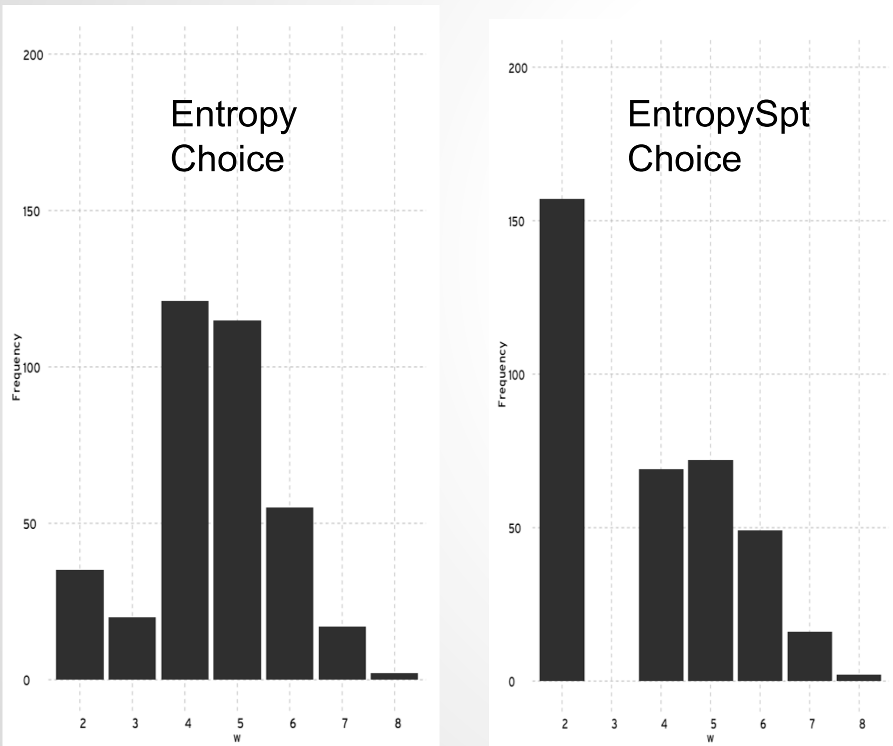
MAE for Probabilistic Methods

Best value for w different for C0875 (and other stations), but still low.

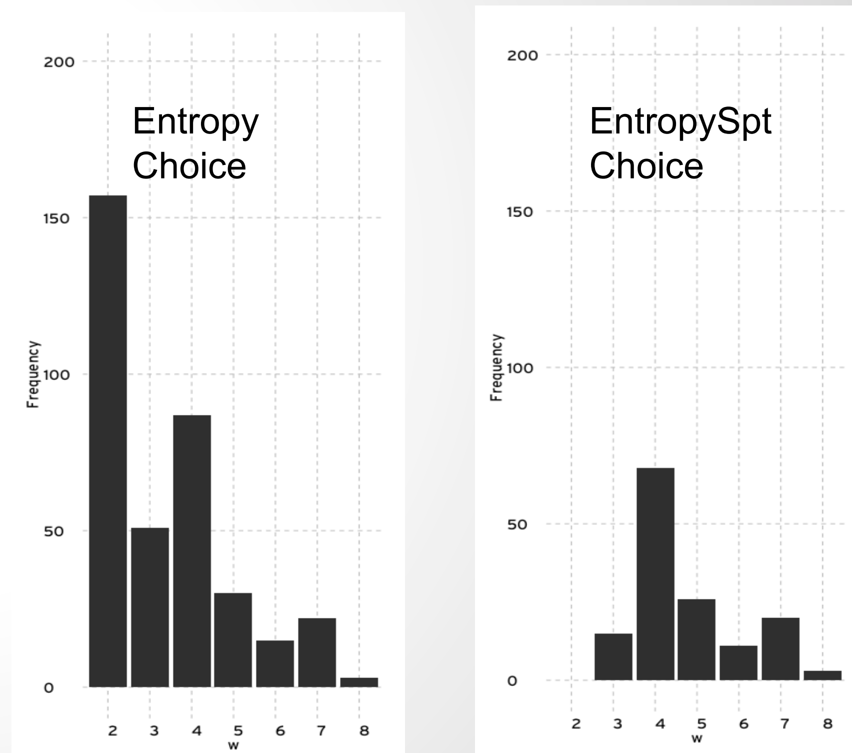


Choice of w

SCBH1



C0875



How much training data ?

Target Year: 2014

Model:

P(St | St-1) ; SCBH1

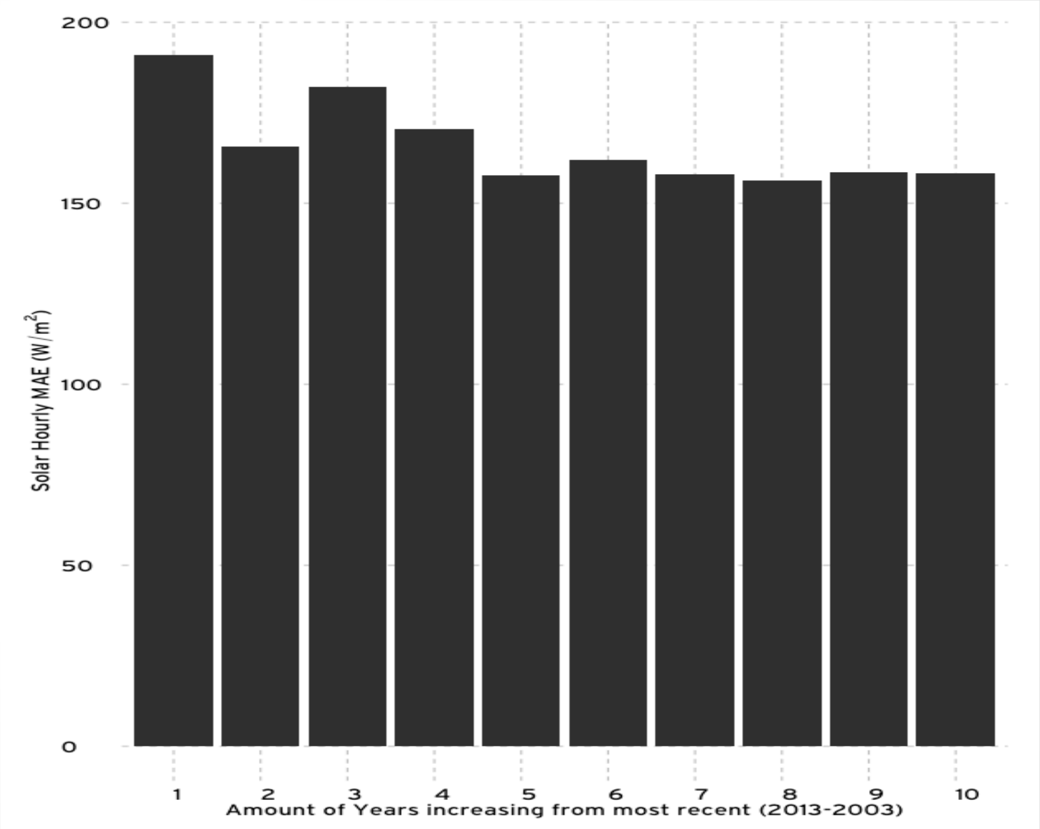
Training Years

(2013)

(2013, 2012)

(2013, 2012 , 2011)

….



Conclusions & Future Work

- Probabilistic models are on average better than linear regression for 1-Day Forecasting

- Small window size works best (Markovian)

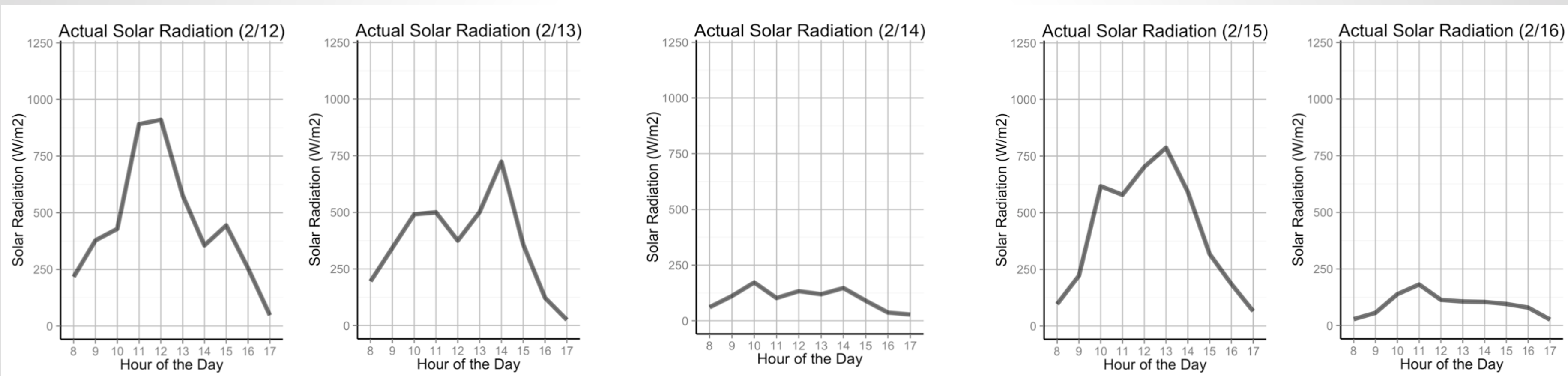
- One to two years of training data sufficient

- Future work : incorporate larger weather features from GFS data

Questions ?

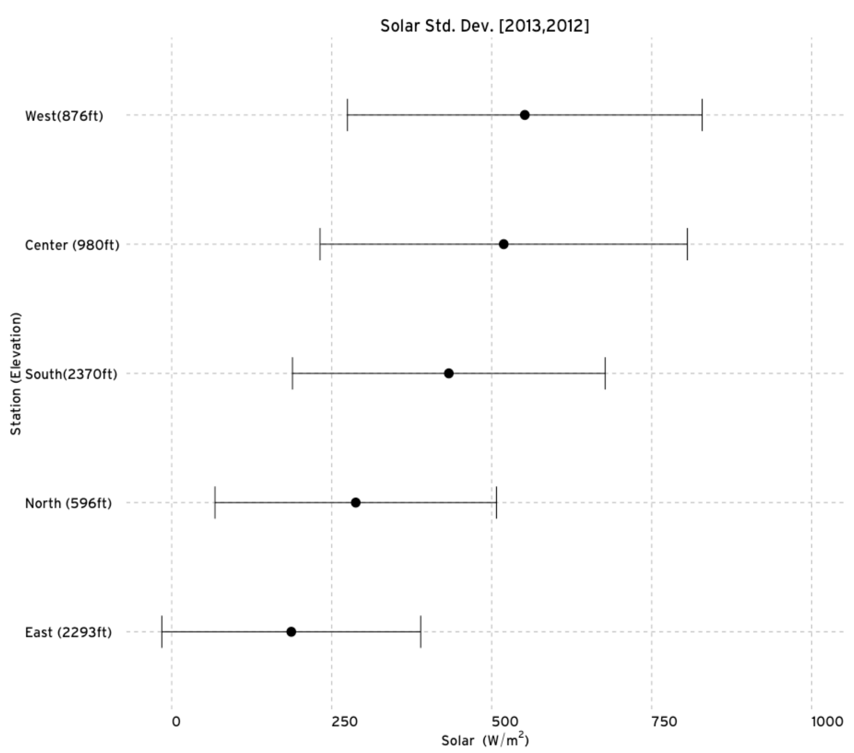
Backup Slides

5-day Sequence of Solar Irradiance

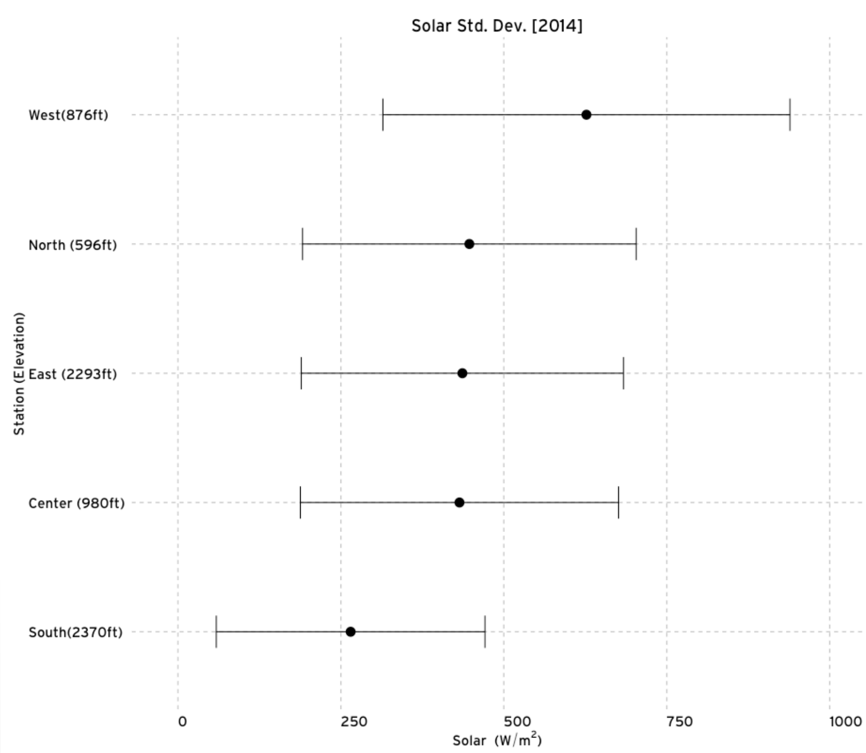


Mean & Std. Dev. for Solar Irradiance

Actual Solar Std. Dev. and Mean on **Training** Data



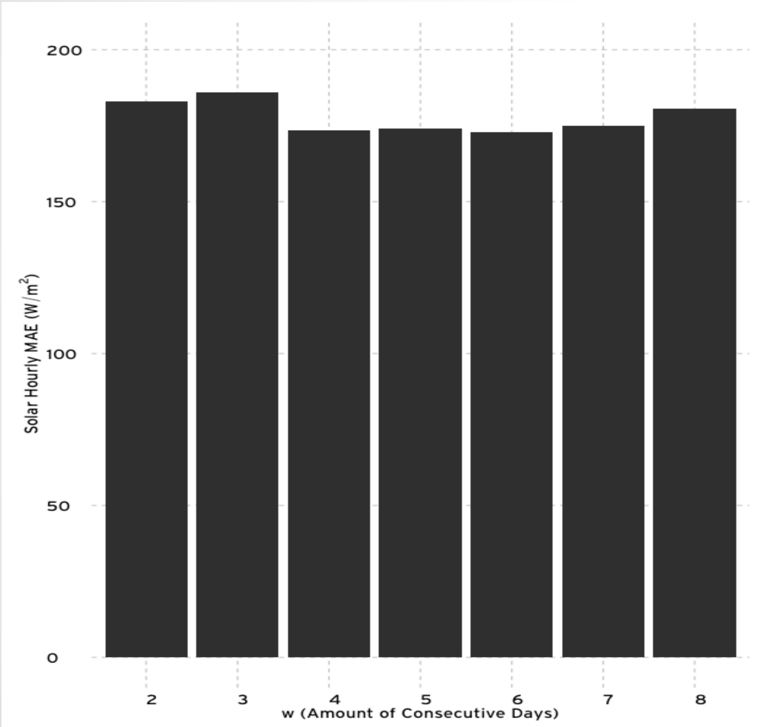
Actual Solar Std. Dev. and Mean on **Test** Data



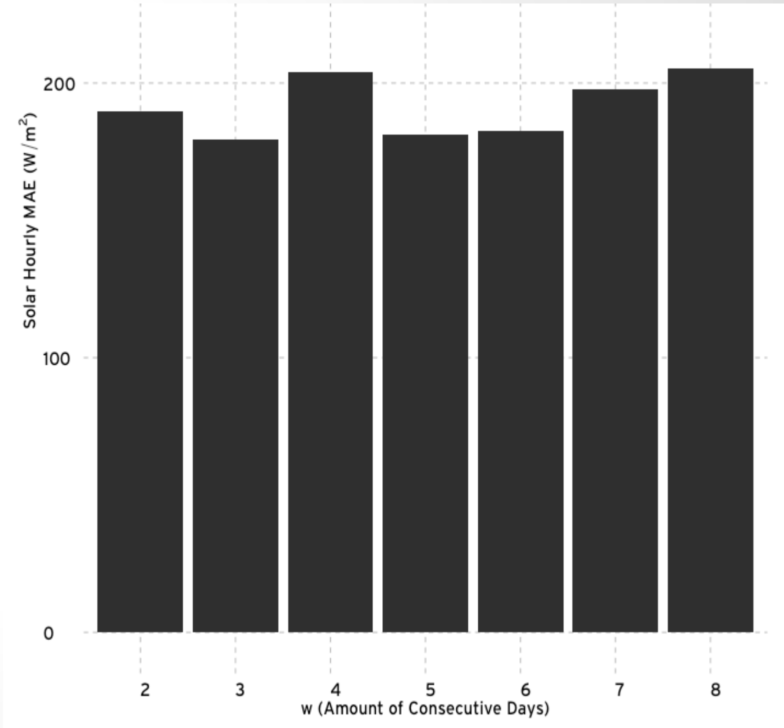
Naive Bayes Classifier

$$\hat{s} = \arg \max_{s} P(S_t{=}s) \prod_{i=1}^{w-1} P(S_{t-i}|S_t{=}s)$$

SCBH1

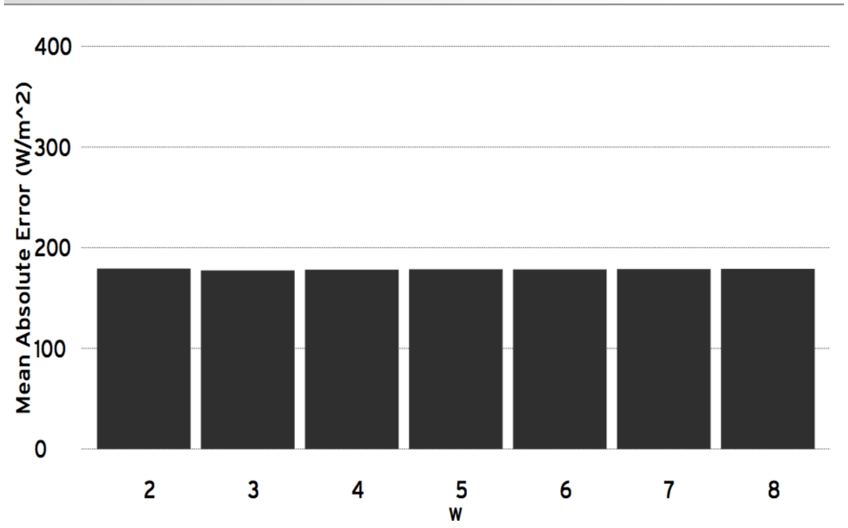


C0875



Linear Regression

SCBH1



C0875