Data Analytics for Solar Energy Management

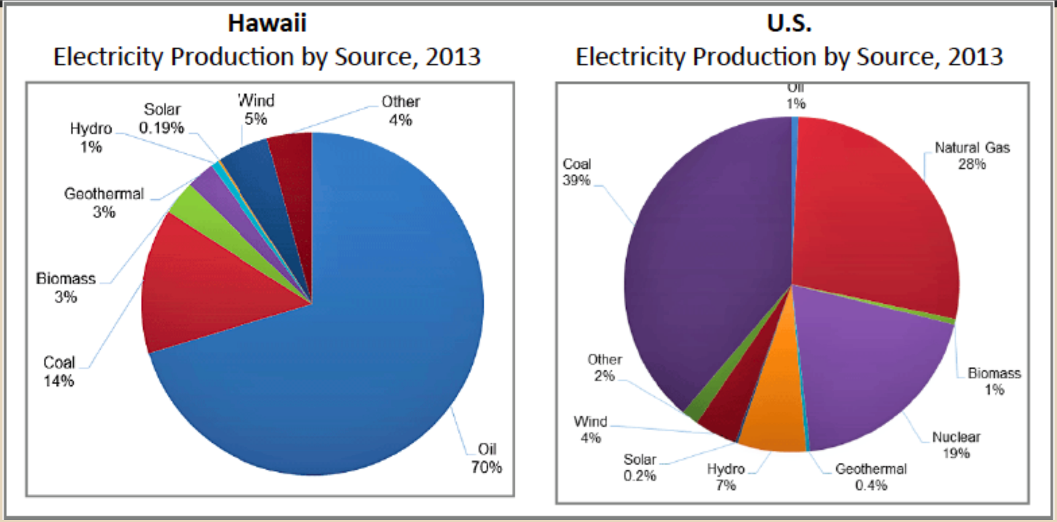
Lipyeow Lim[1], Duane Stevens[2], Sen Chiao[3], Christopher Foo[1], Anthony Chang[2], Todd Taomae[1], Carlos Andrade[1], Neha Gupta[1], Gabriella Santillan[2], Michael Gonzalves[2], Lei Zhang[2]

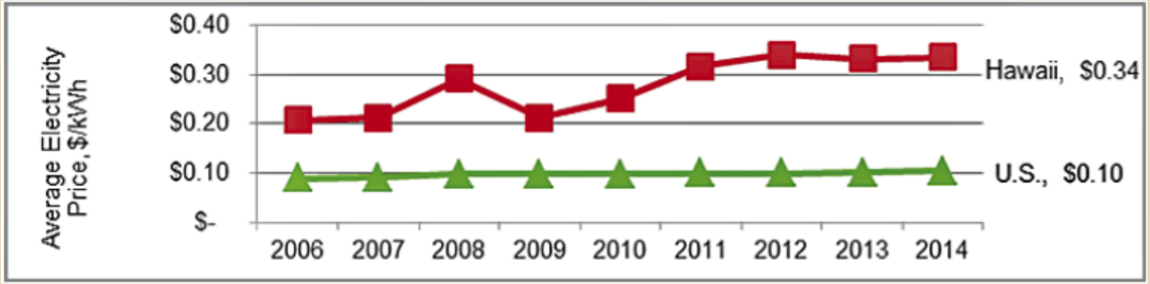
[1] Info. & Comp. Sciences, U. of Hawai`i at Mānoa
[2] Atmospheric Sciences, U. of Hawai`i at Mānoa
[3] Met. & Climate Science, San Jose State University

Energy in the State of Hawai`i

- In 2013, Hawaii relied on oil for 70% of its energy.

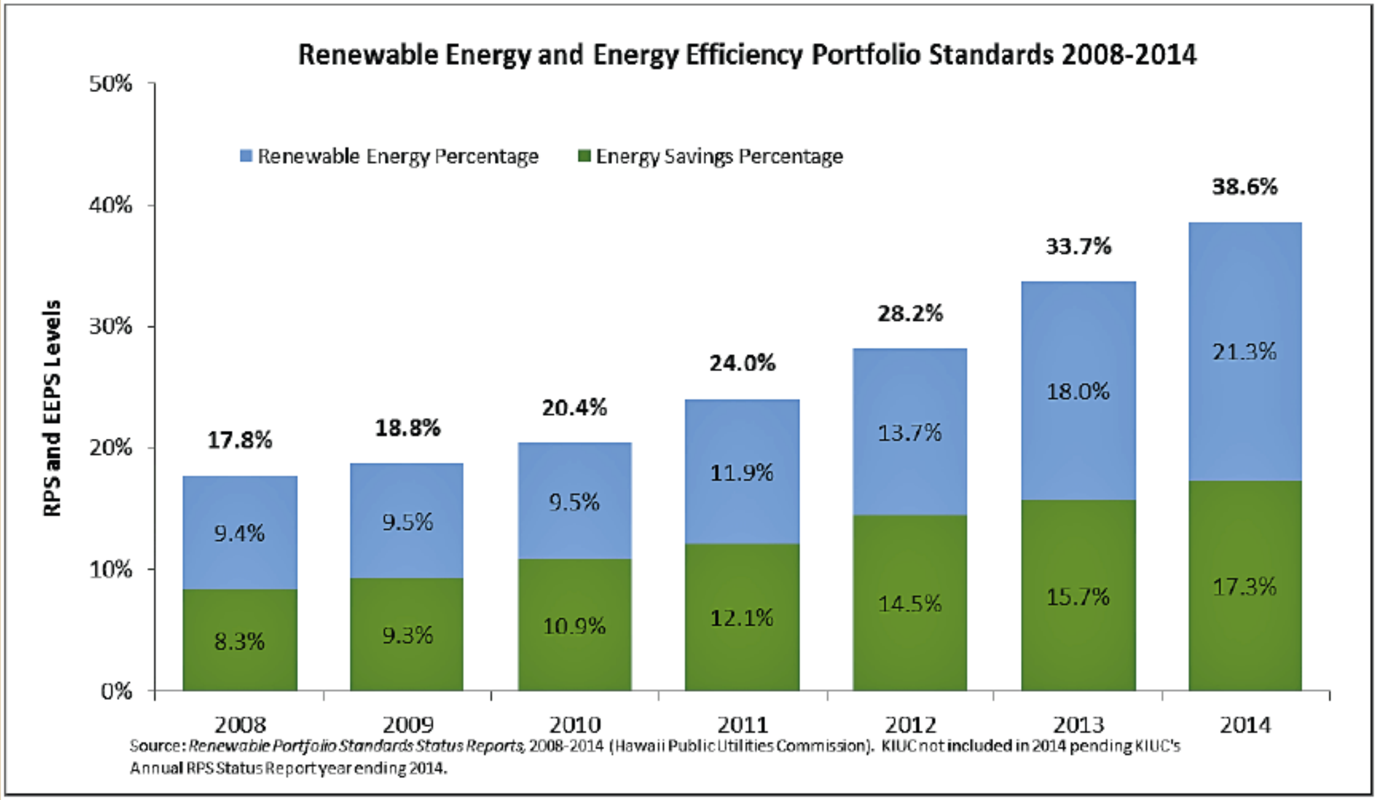
- Hawaii's electricity cost is 3 times the US average





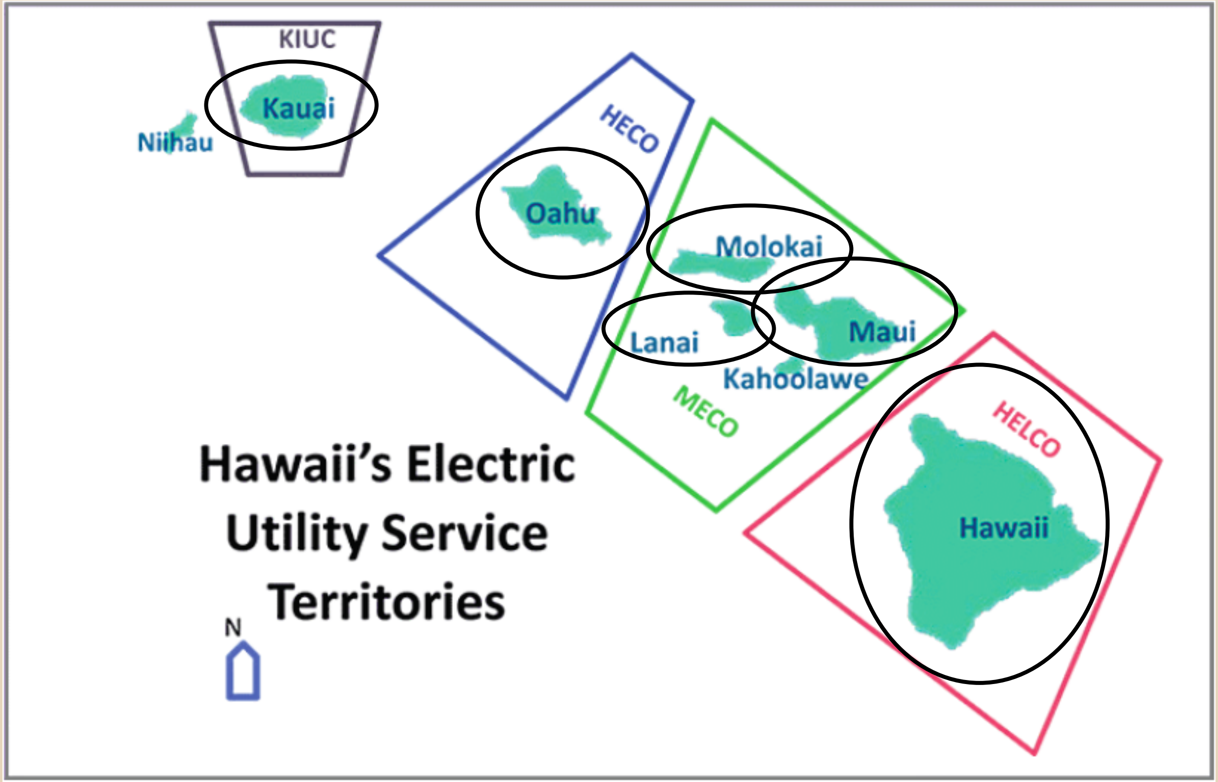
Renewables in the State of Hawai`i

Meet & exceed 70% clean energy by 2030



Disconnected Grids

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



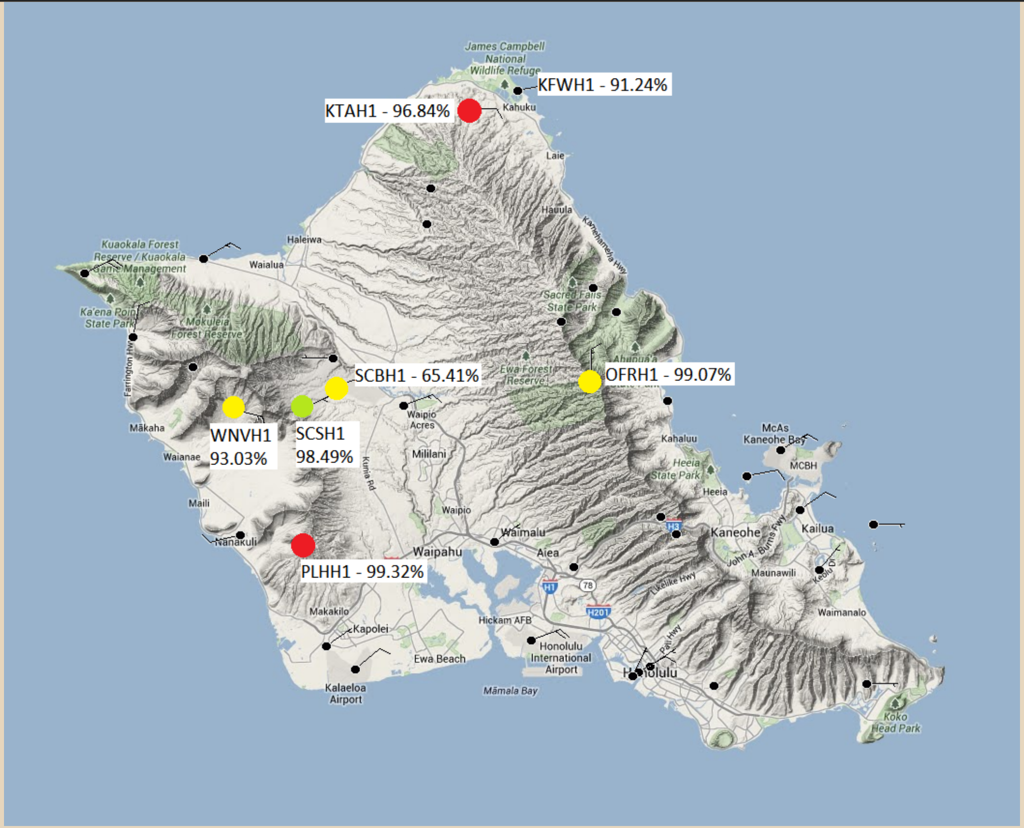
Research Objective

Investigate the use of data-centric methods for predicting solar irradiance at a specific location

- complement not replace NWP (eg. WRF)
- 1-3 hour ahead predictions
- 1 day ahead predictions

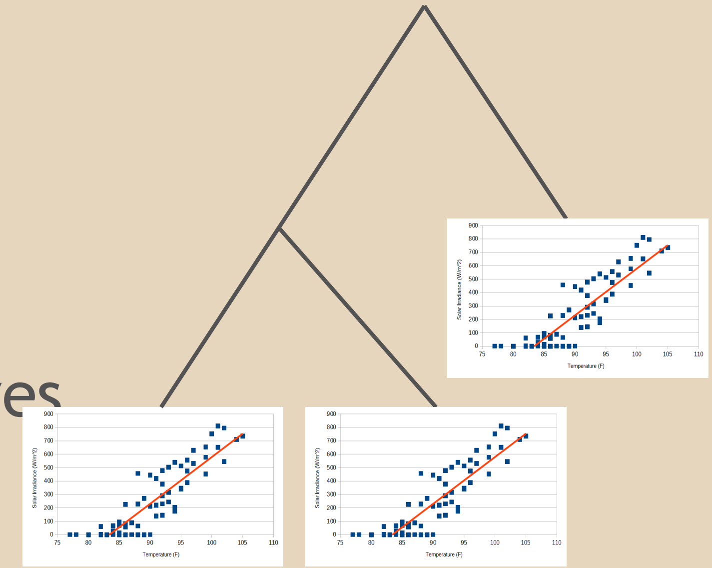
Data Sources

- MesoWest
  - 30 Weather Stations
  - ~10 sensors each
  - 5-60 min sampling interval
- 4 Years of Hourly Data
  - January 1, 2010 to December 31, 2013
- SCSH1, PLHH1 & KTAH1 stations



1-Hour Ahead Predictions

- **Linear Regression**
  - Select top-5 features from diff sensors at diff time at diff neighboring location
- **Cubist Trees**
  - Decision trees with linear regression models at the leaves
- **Normalize data to hourly readings**



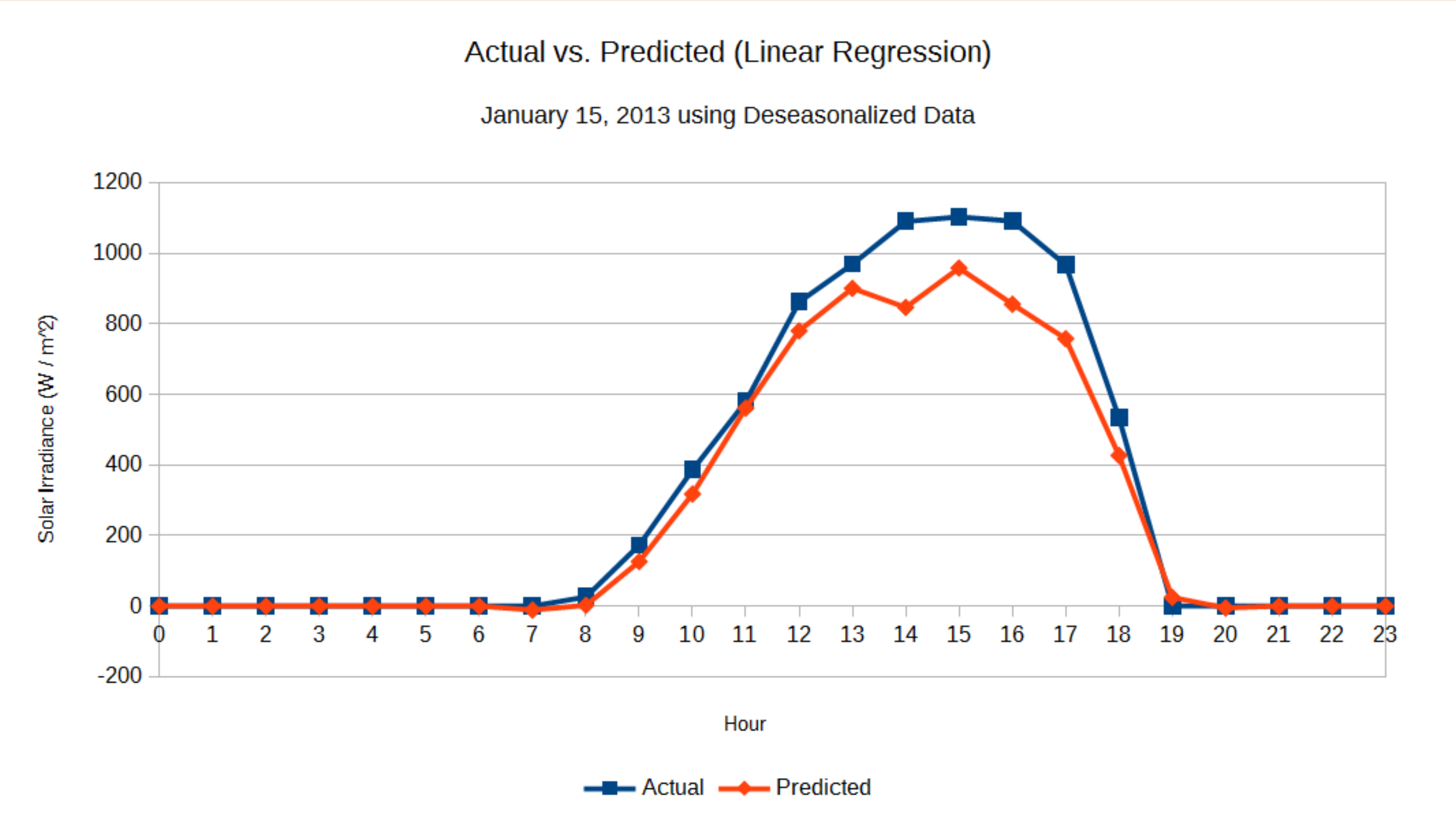
Dealing with Seasonality

Two types of cycles in the (irradiance) data: daily & yearly

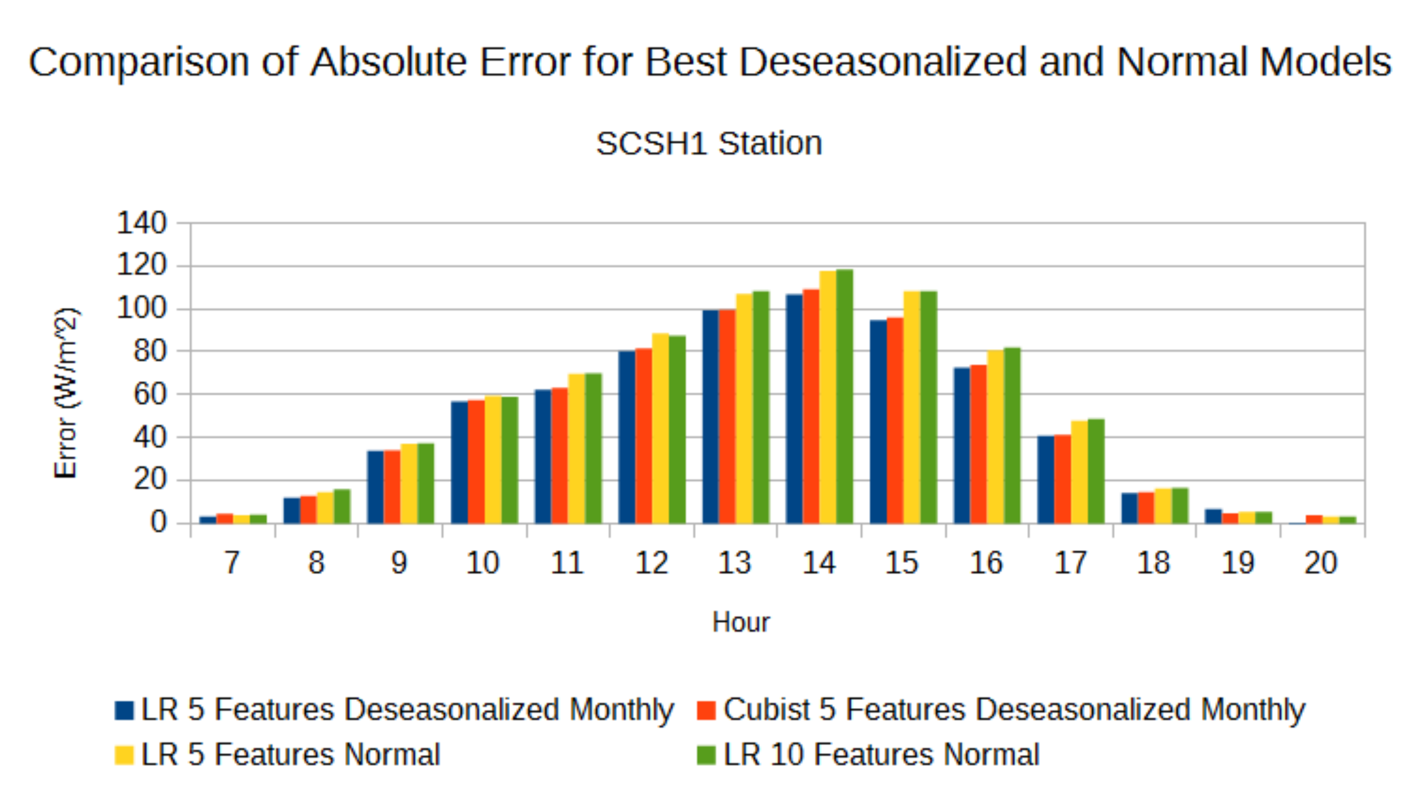
- **Separate models for each “season”**
  - eg. a separate model for each month & hour: Jan 10am
- **Deseasonalize the data**
  - Mean signal: for each day & hour average the values over the 4 hours
  - Subtract the mean signal from the data

On a good day...

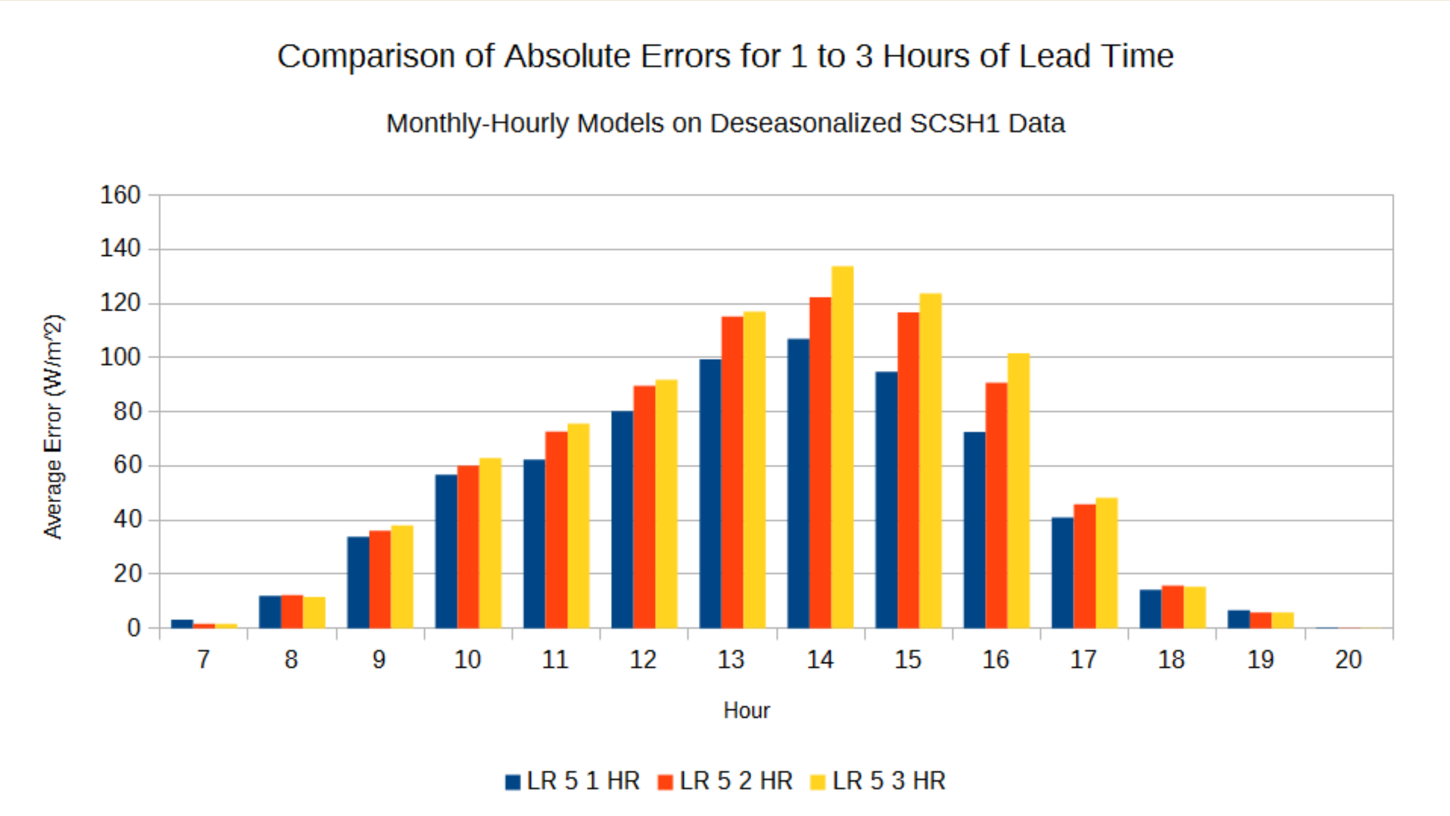
Month-hour with top 5 features



Prediction Errors

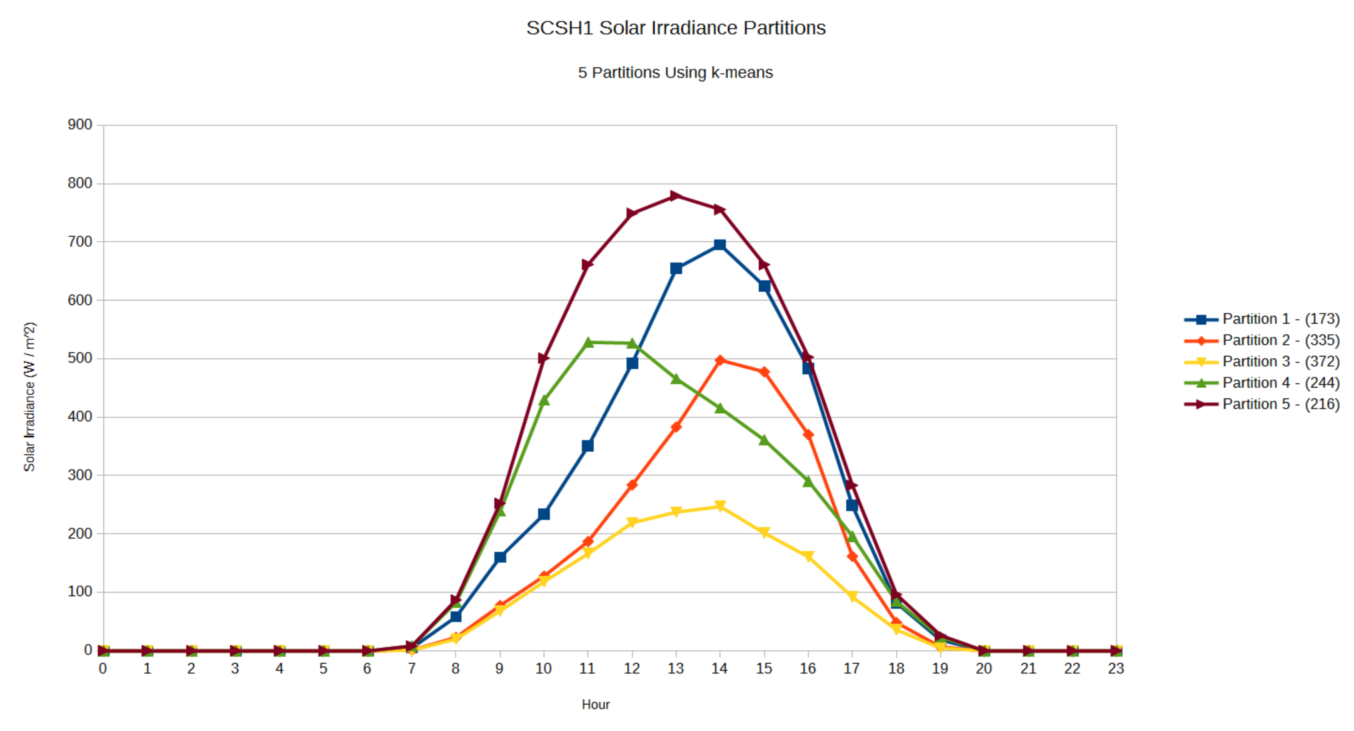


1-3 Hour Ahead Predictions

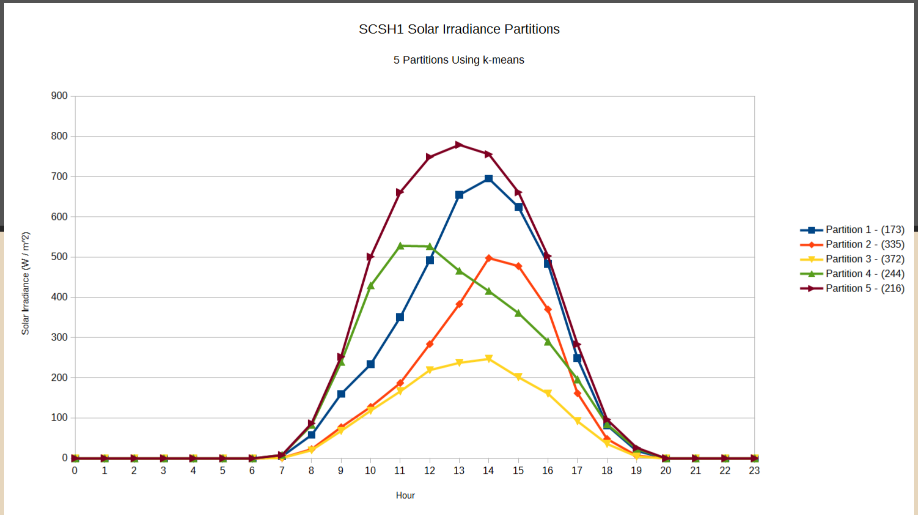


1-Day Ahead Predictions

- Consider granularity of 1 day
- Apply a clustering algorithm
  - k-means
  - PAM
- Examine centroids / medoids



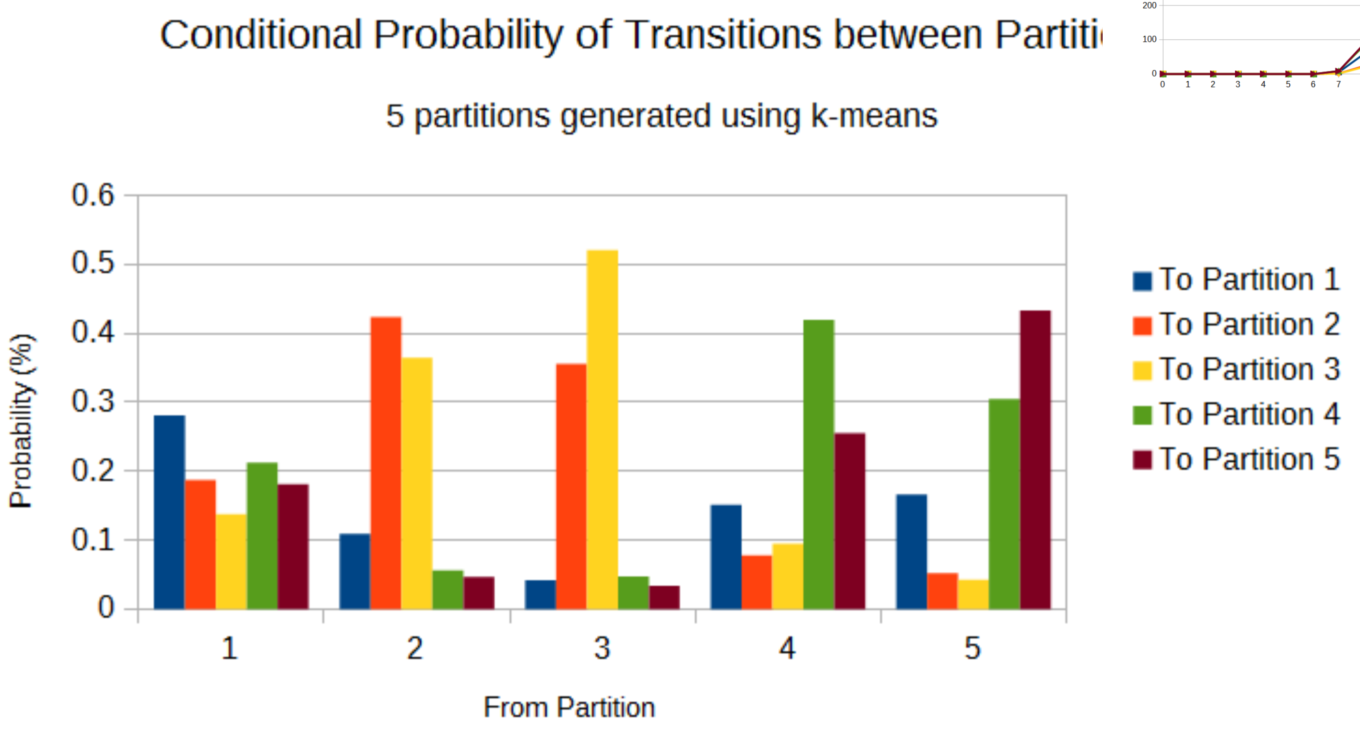
Partition Chains

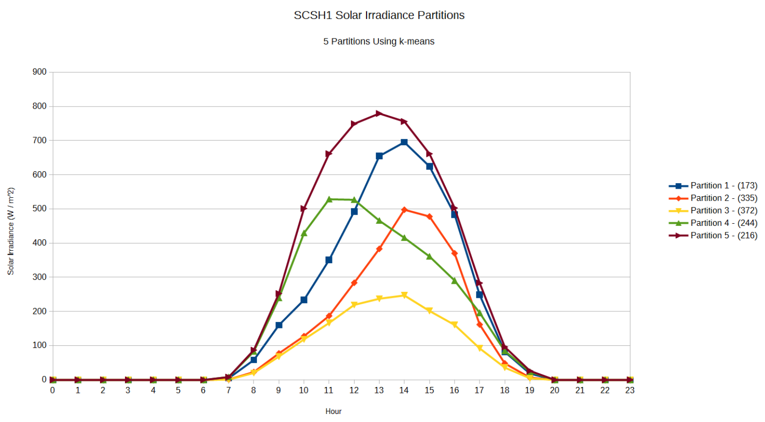


- Procedure
  - Order partition numbers by date
  - Find consecutive days with the same partition number
  - Find the length of these “chains”
- Result:Normally about 2 ~ 3 consecutive days in the same partition

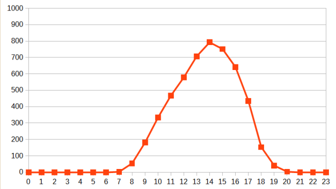
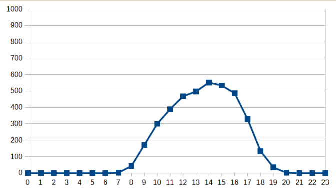
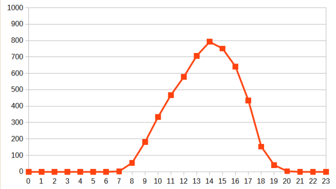
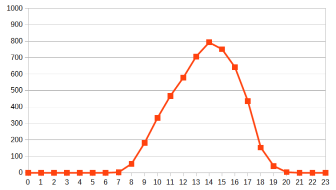
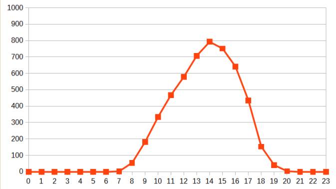
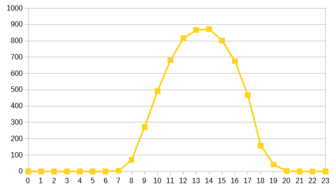
	Partition 1	Partition 2	Partition 3	Partition 4	Partition 5
Average Chain Length	2.286	2.863	3.583	2.732	2.717
Maximum Chain Length	5	11	13	6	11

Partition Transitions

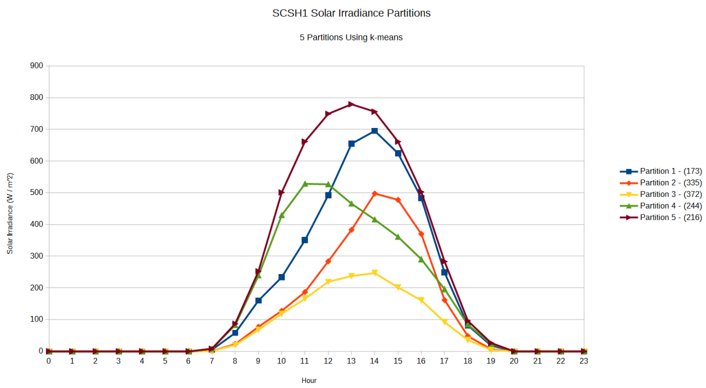




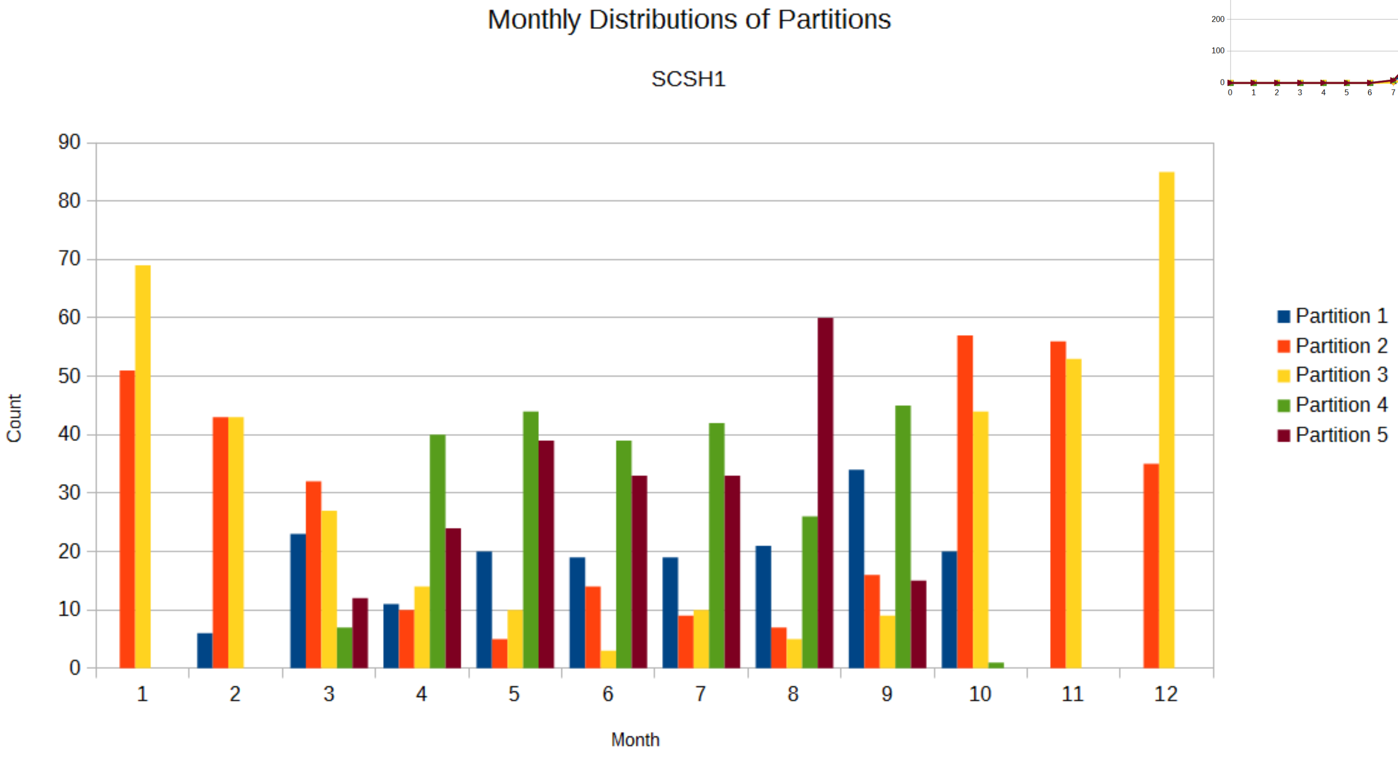
Conditional Probability

1 Day Before	Next / Forecast Day	Probability
		15.38%
		23.53%
		26.70%

vs. Months



SCSH1 Solar Irradiance Partitions
5 Partitions Using k-means
Solar Irradiance (W / m^2)
Hour
Partition 1 - (173)
Partition 2 - (335)
Partition 3 - (372)
Partition 4 - (244)
Partition 5 - (216)



Monthly Distributions of Partitions
SCSH1
Count
Month
Partition 1
Partition 2
Partition 3
Partition 4
Partition 5

Naive Bayes Classifier

- Probabilistic classifier using Bayes' theorem
  - Assumes independence between features
  - $$\text{classify}(f_1, \ldots, f_n) = \underset{c}{\operatorname{argmax}}\ p(C = c) \prod_{i=1}^{n} p(F_i = f_i | C = c).$$
- Feature Selection
  - Relative Humidity, Temperature, Wind, Solar Clusters for target site
  - Greedy
    - Select best number of clusters for each feature
    - Find best combination of features

Setup

- 1 Day and 4 Day lead time
- 3 years training (2010 - 2012)
- 1 year testing (2013)
- PLHH1 & KTAH1
- Hourly
  - Top 5, 10, 20, 30, 50 features
  - 6 hour data window
- Daily
  - Conditional Probability & Naive Bayes
    - Predicting 6 solar irradiance partitions
    - 2 day data window

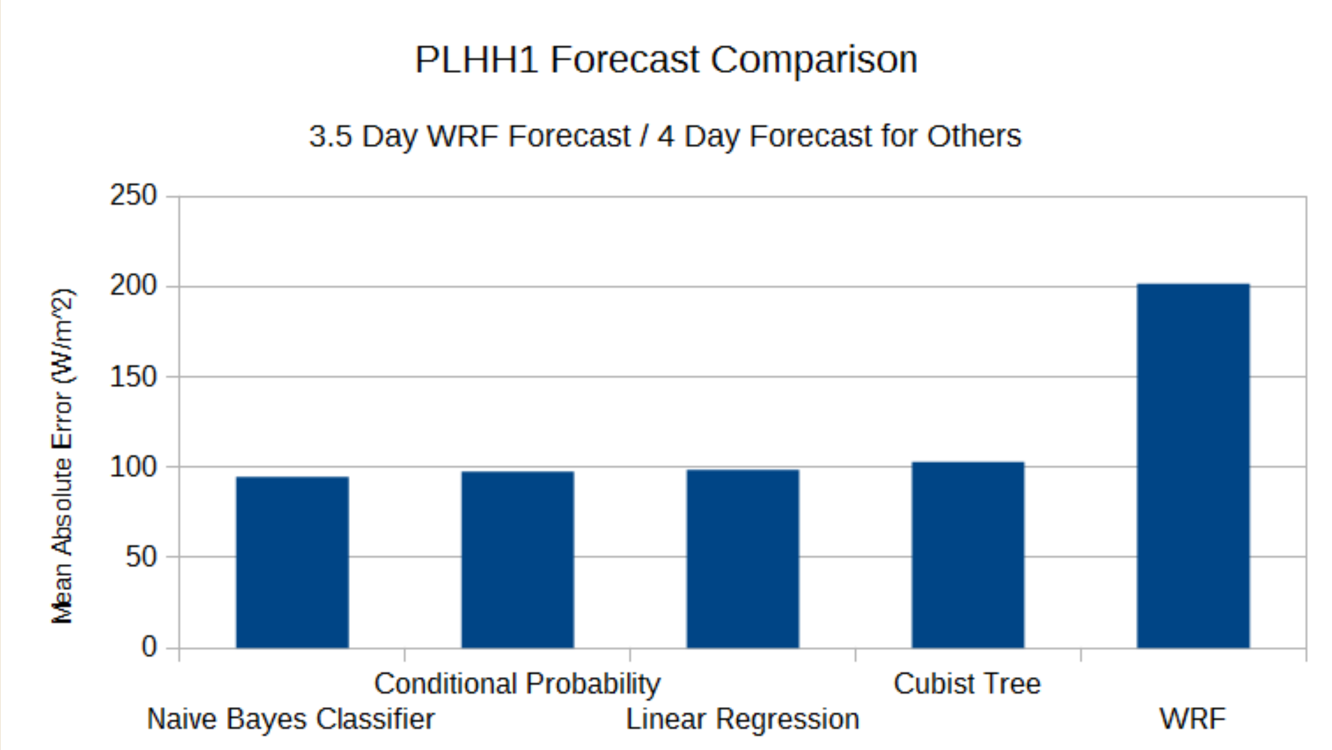
WRF Comparison

- WRF Irradiance Forecasts
  - Run by Prof. Yi-Leng Chen of the Meteorology Department in SOEST
  - Freely available online
  - 3.5 Day Hourly Forecasts
  - 1.5 km resolution
- Find closest grid to stations
- Difference between forecasted and observed

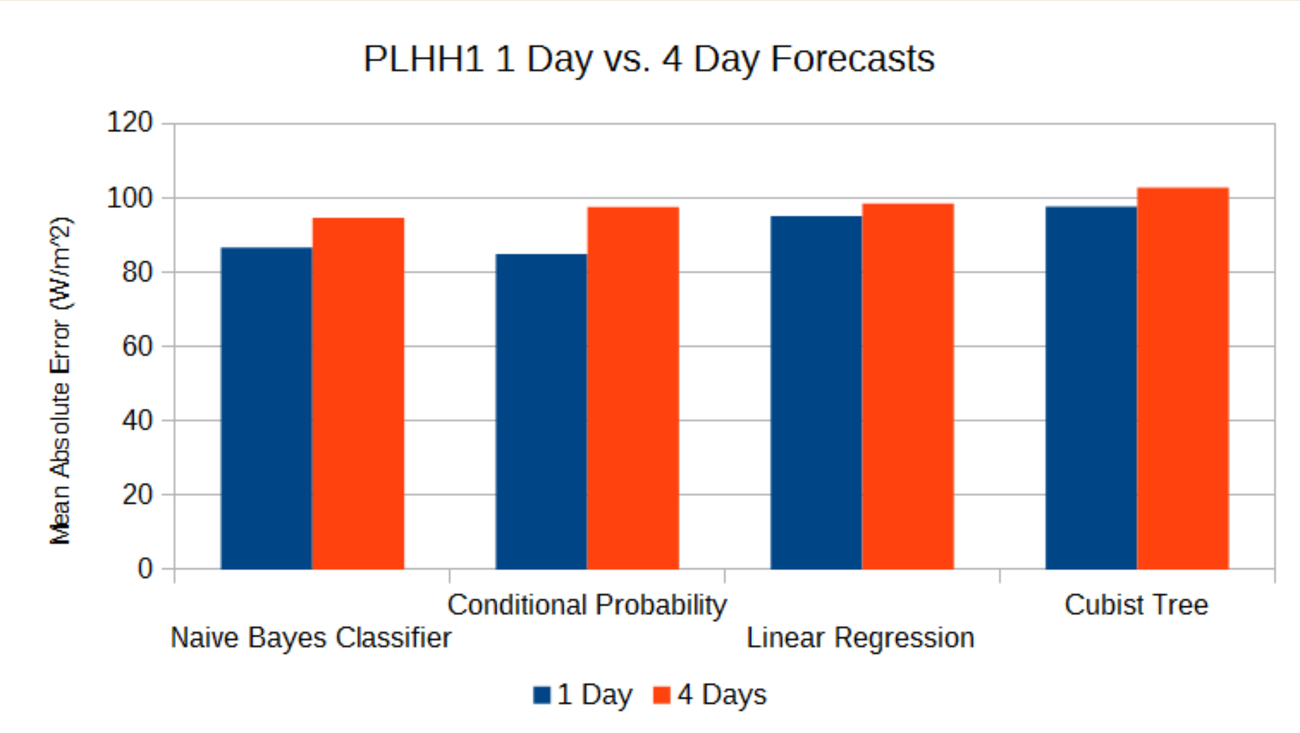
Metric

- Mean Absolute Error = $\frac{1}{N}\sum_{i=1}^{N}|predicted_i - actual_i|$
- WRF & Hourly Forecasts
  - Predicted = Forecasted solar irradiance at the hour
  - Actual = Observed solar irradiance at the hour
- Daily Forecasts
  - Predicted / Actual solar irradiance values obtained from the cluster
- Only daytime hours (7 am - 8 pm) are considered

Data Driven vs. WRF - PLHH1



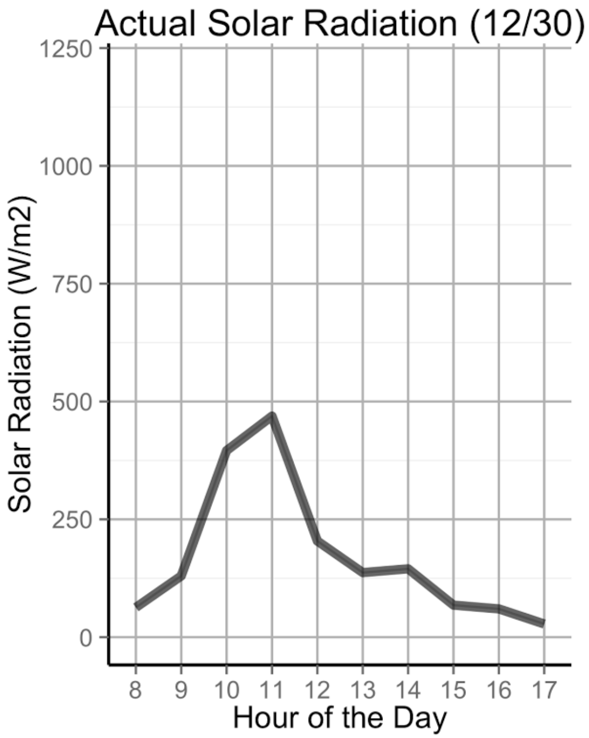
1 Day vs. 4 Days - PLHH1

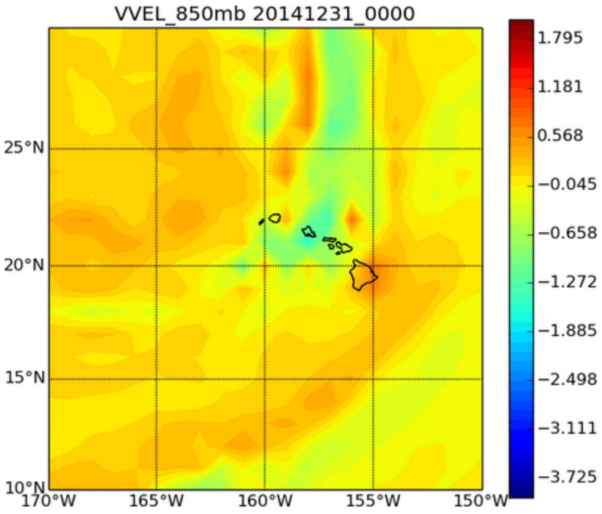


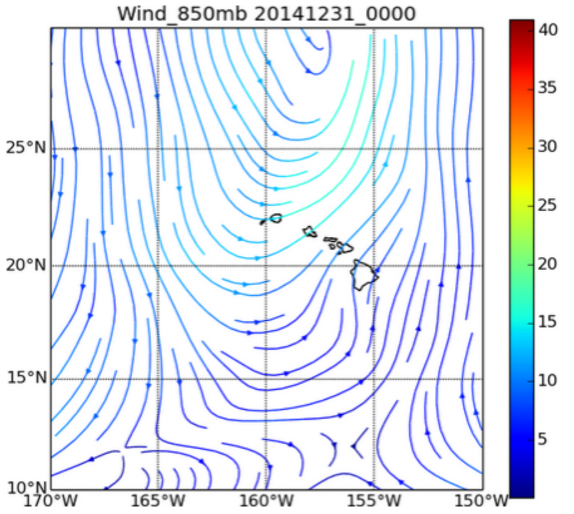
“Rare” Events

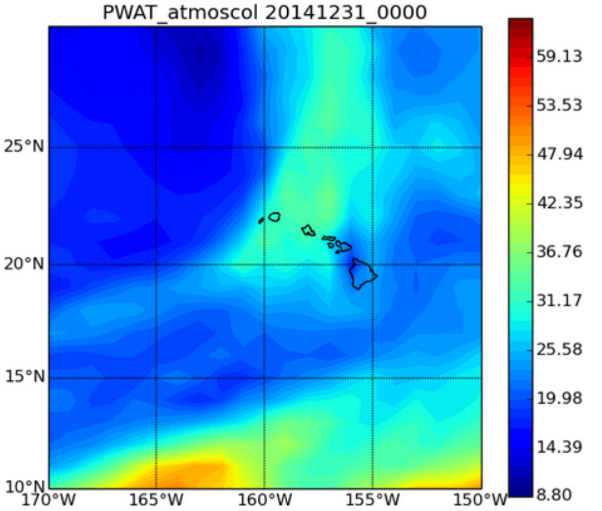
- Similar to outlier analysis
- Several possible definitions depending on how we model what is NOT rare:
    - Infrequent events (phenomenological)
    - Events not predicted well by a given model (statistical or dynamical or both)
    - Events with high disagreement in an ensemble of models

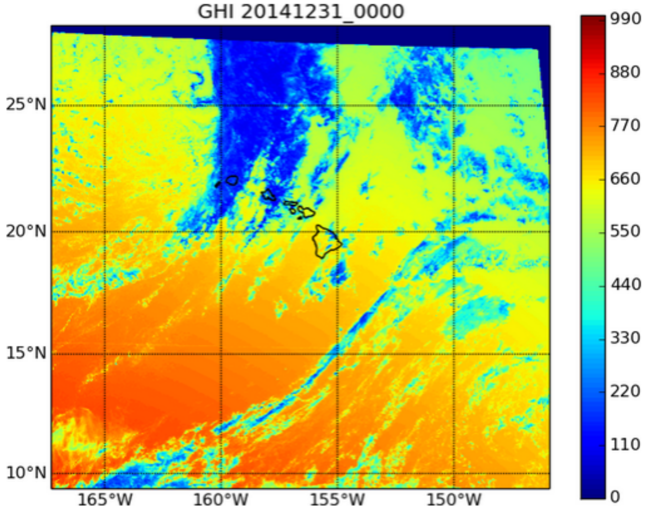
GFS Rare Day: Dec 30, 2014 (- 0 days)







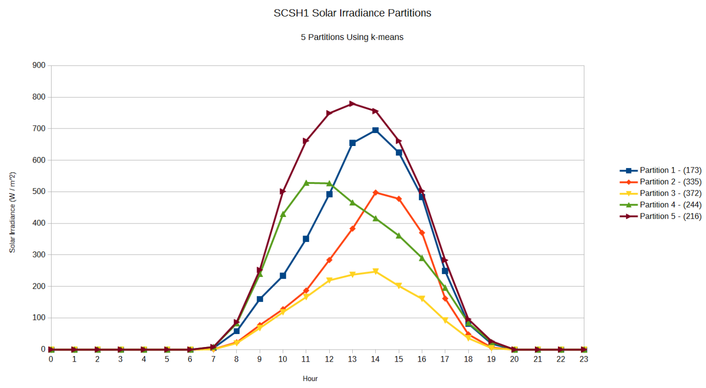


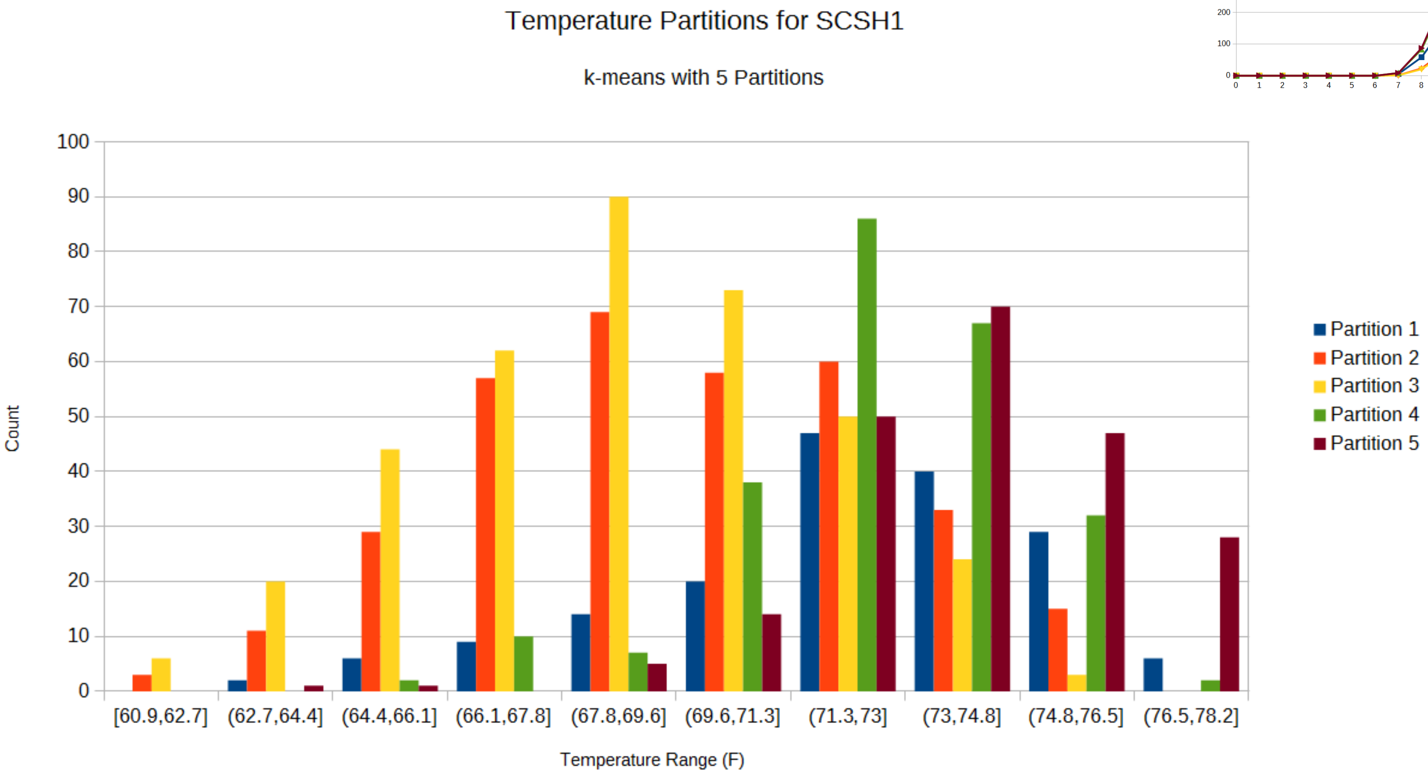


Conclusions

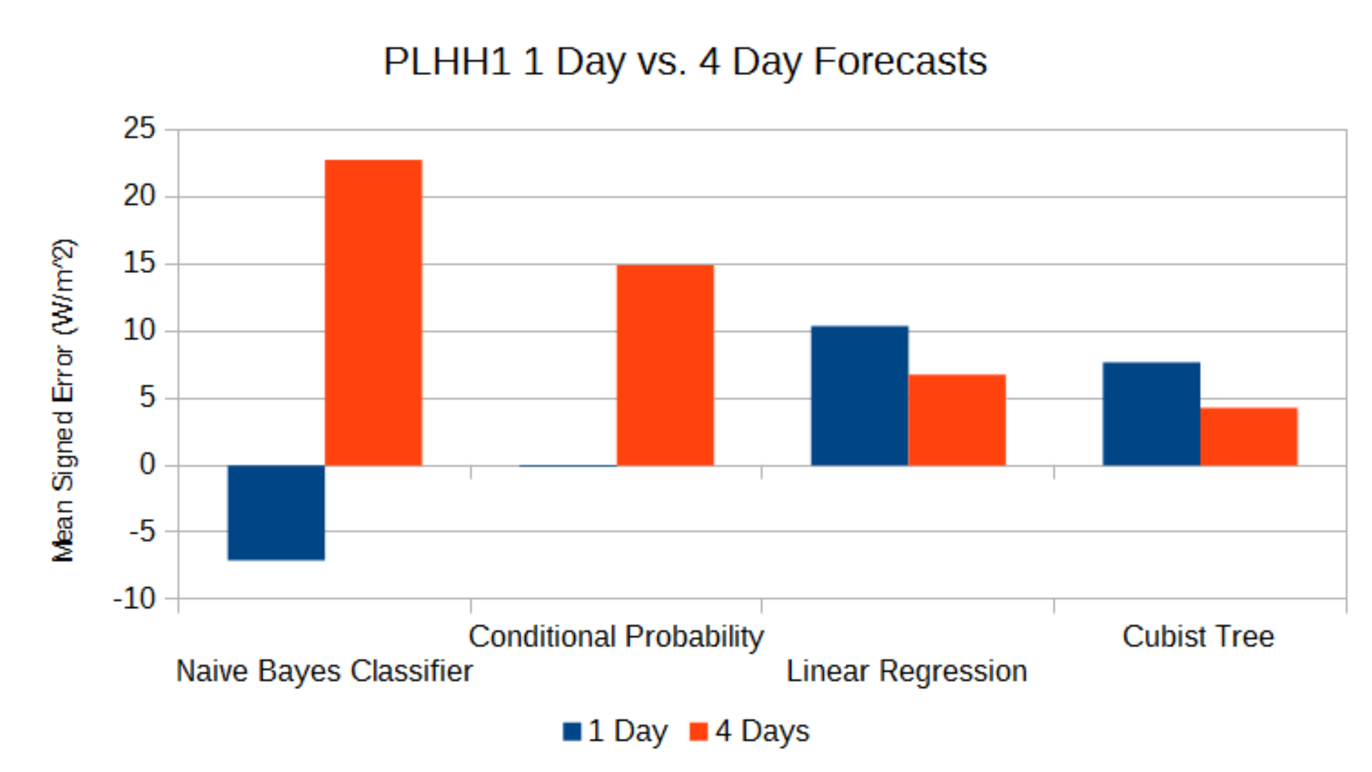
- 1-3H ahead forecasts
  - Linear Regression & Cubist Trees: ~15% error
- 1-3D ahead forecasts
  - Clustering into daily irradiance profiles
  - Interesting analysis using discrete techniques: chains, conditional entropy etc.
  - Discrete prediction techniques: ~15% error
- Outlier analysis
  - Incorporate “signal” from larger scale

vs. Temperature

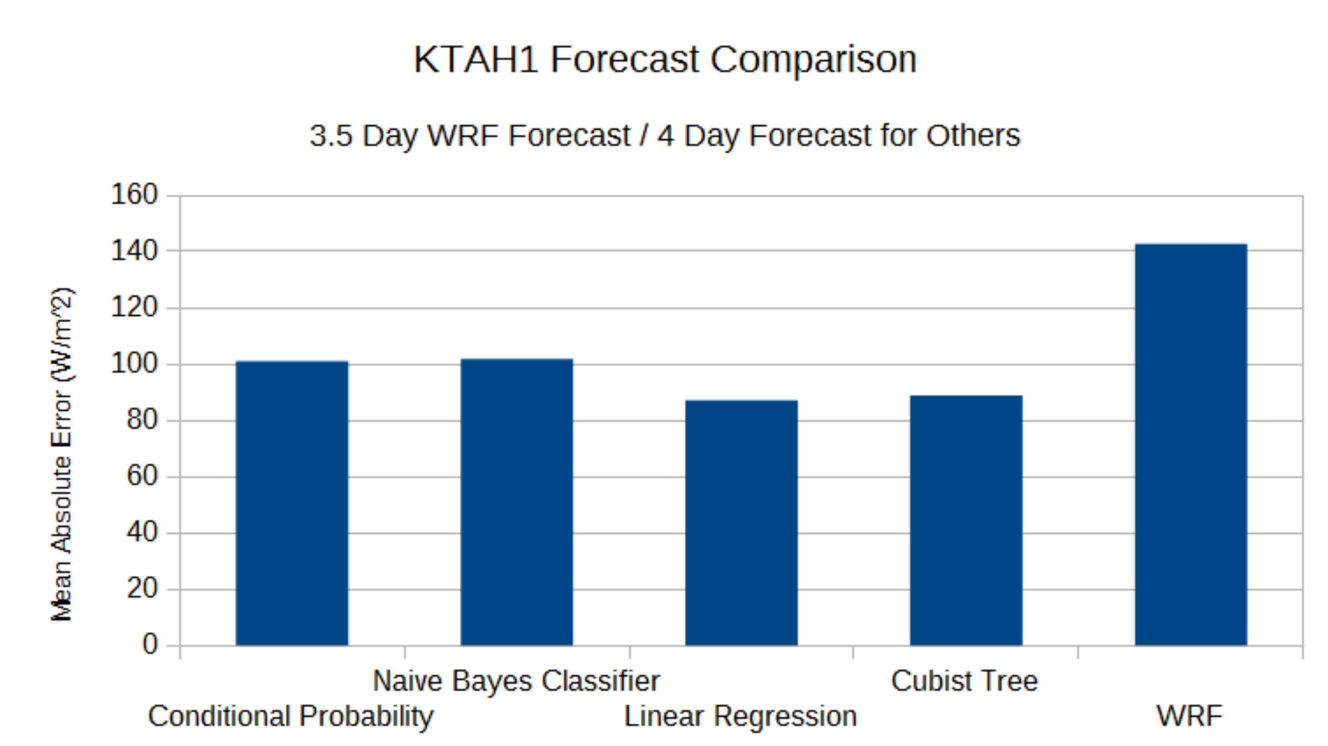




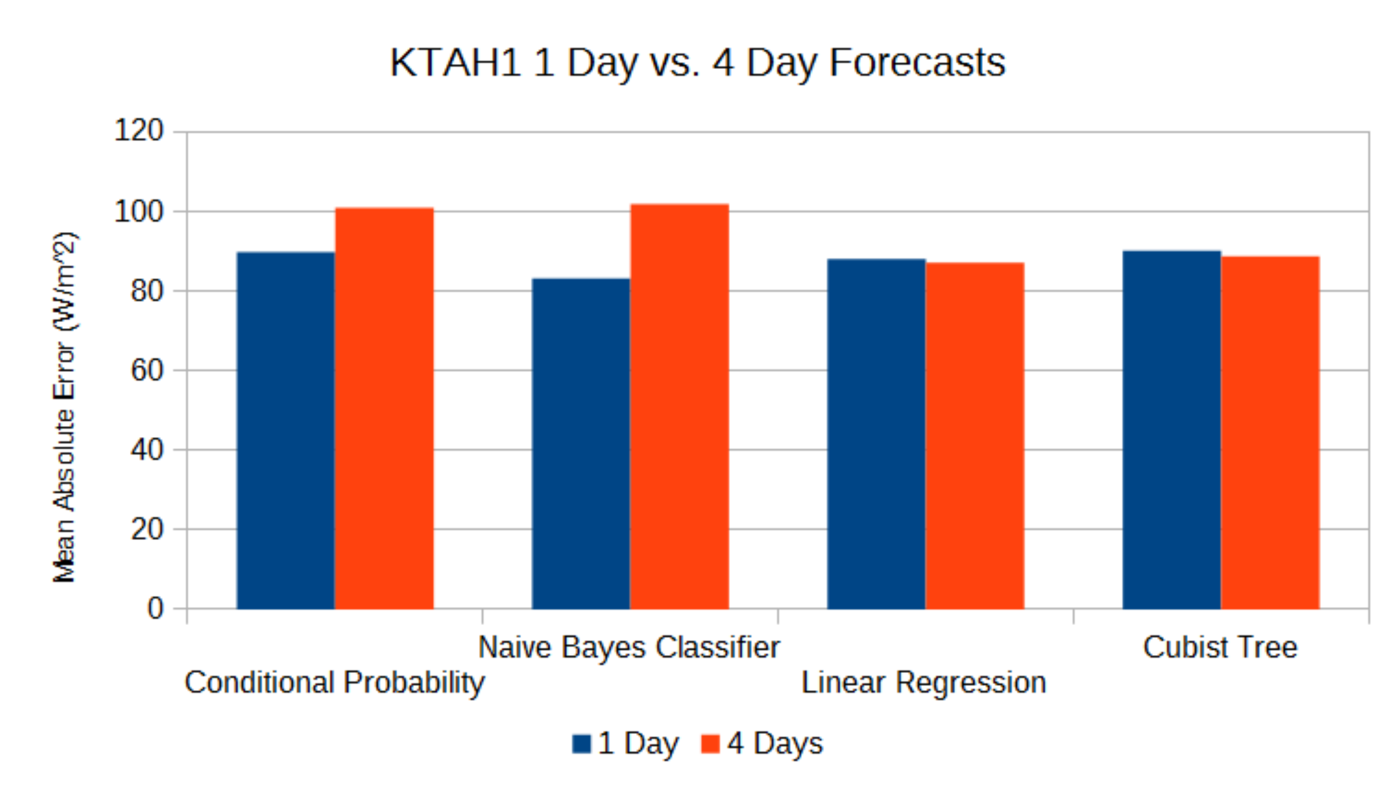
1 Day vs. 4 Days - PLHH1



Data Driven vs. WRF - KTAH1



1 Day vs. 4 Days - KTAH1



1 Day vs. 4 Days - KTAH1