Predicting Electricity Generated at Wind Farms in the California ISO

Peter Meleney

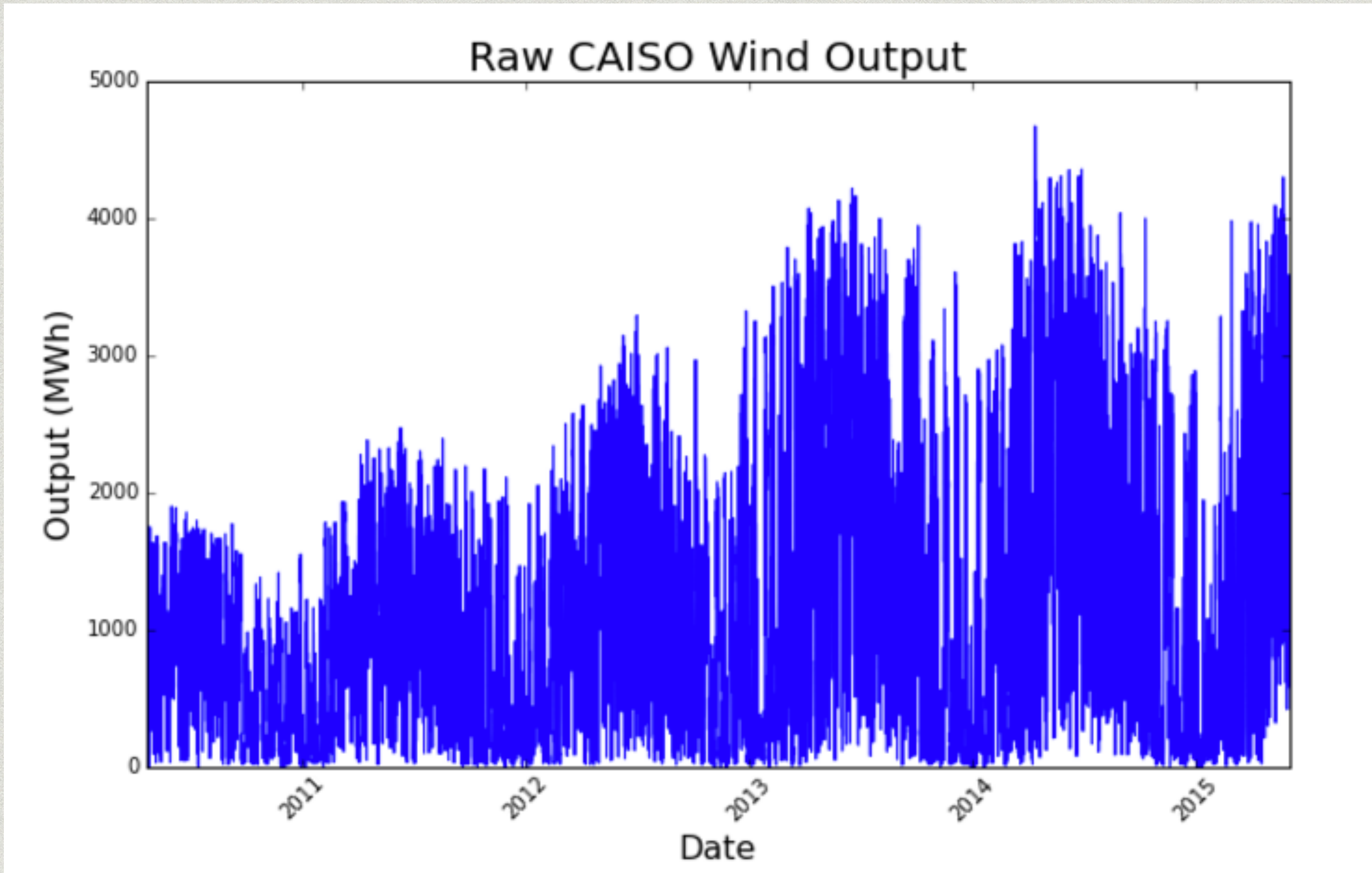
A General Assembly Data Science Project
Instructor: Mason Gallo
Final Presentation

10/21/15

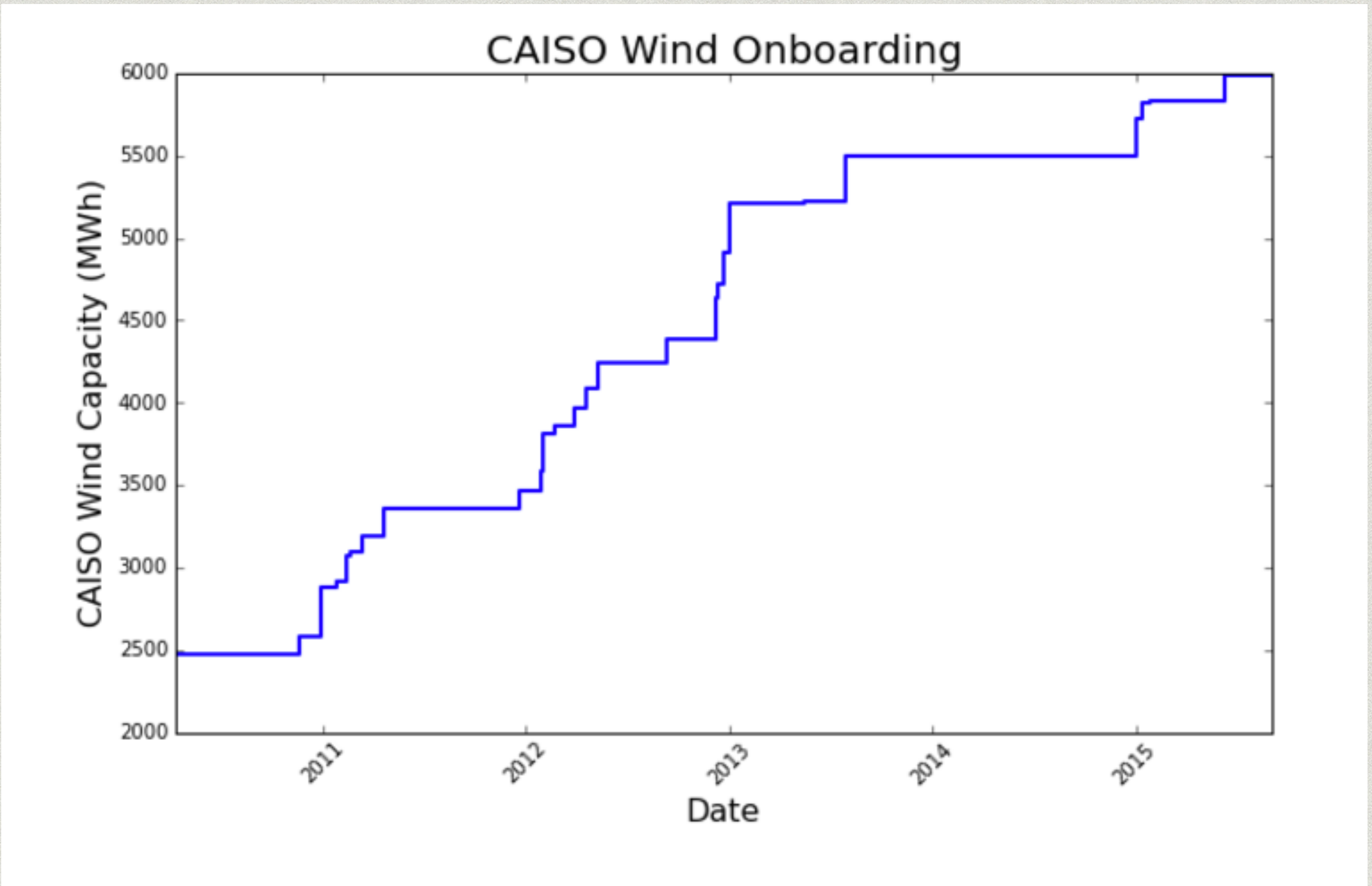
Purpose

- Predict how much **electricity** (kWh) will be generated at wind farms based on the **weather**.
- This is **important** because better the prediction lead to lower electricity **cost**, and **reduces GHG** emissions.

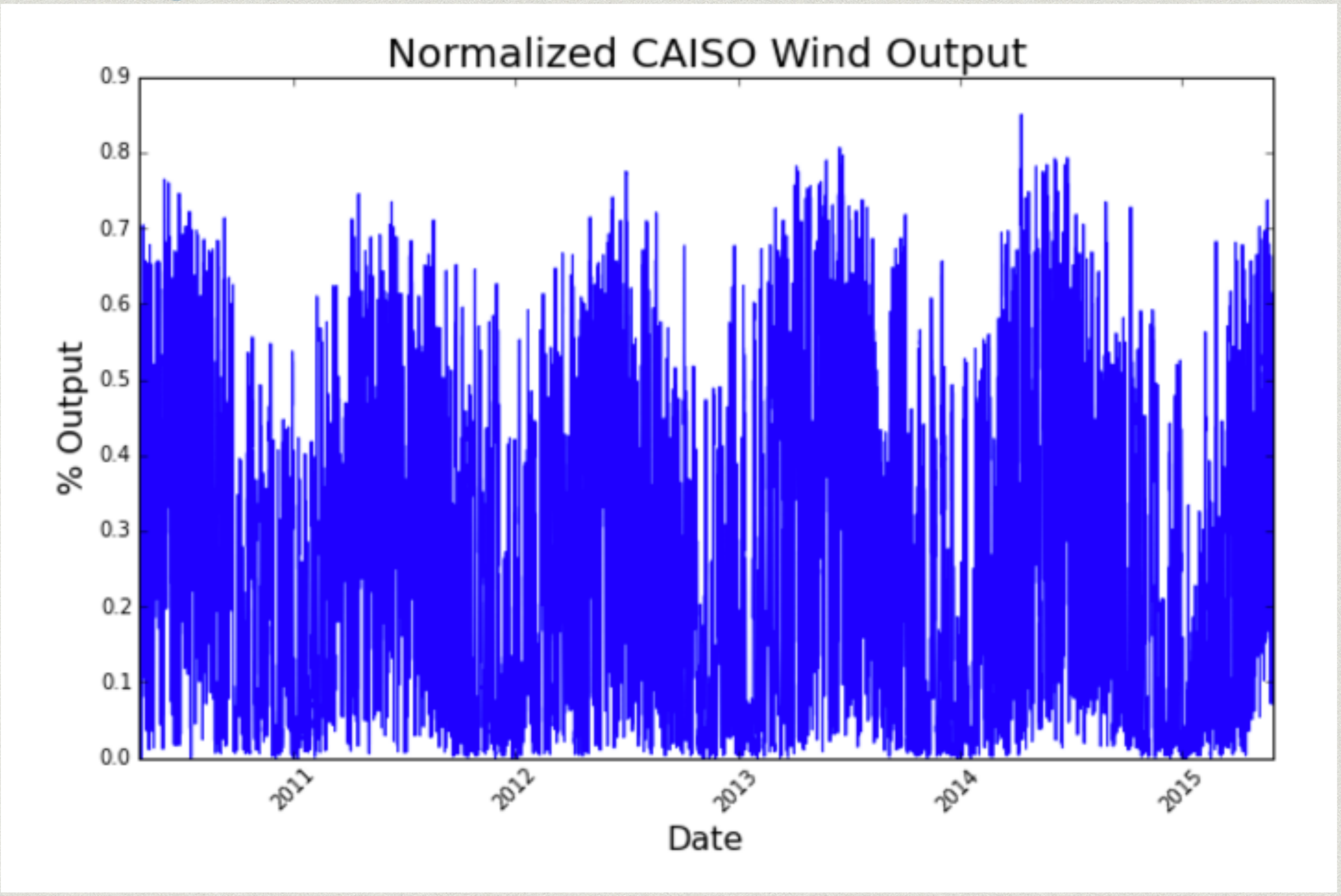
Target Normalization



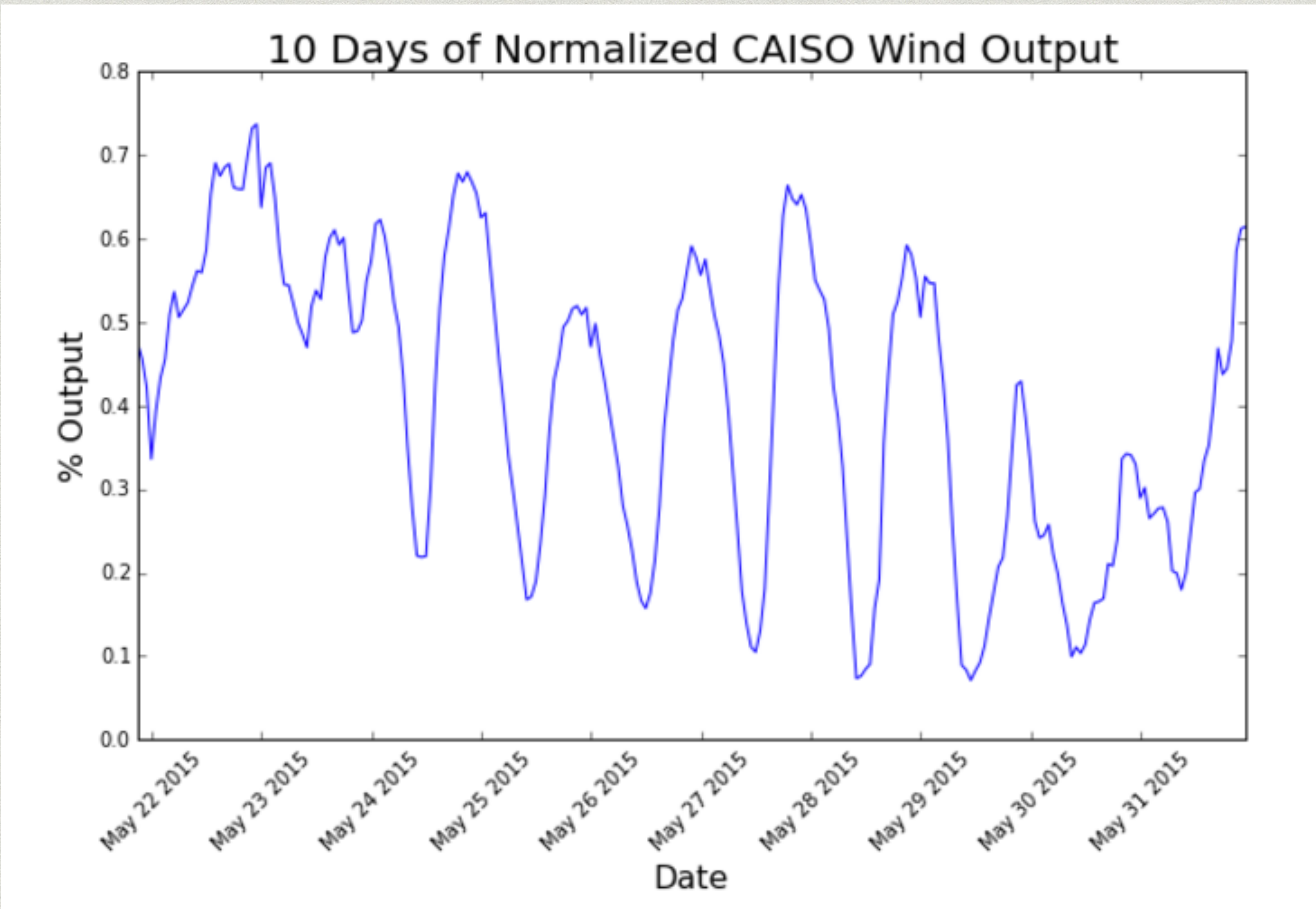
Target Normalization



Target Normalization



Target Normalization



Theory

Available Wind Energy:

$$E = \boxed{\frac{1}{8}\pi d^2}\ \boxed{\rho U^3}$$

$$\boxed{d = rotor\ diameter\ (m)}$$

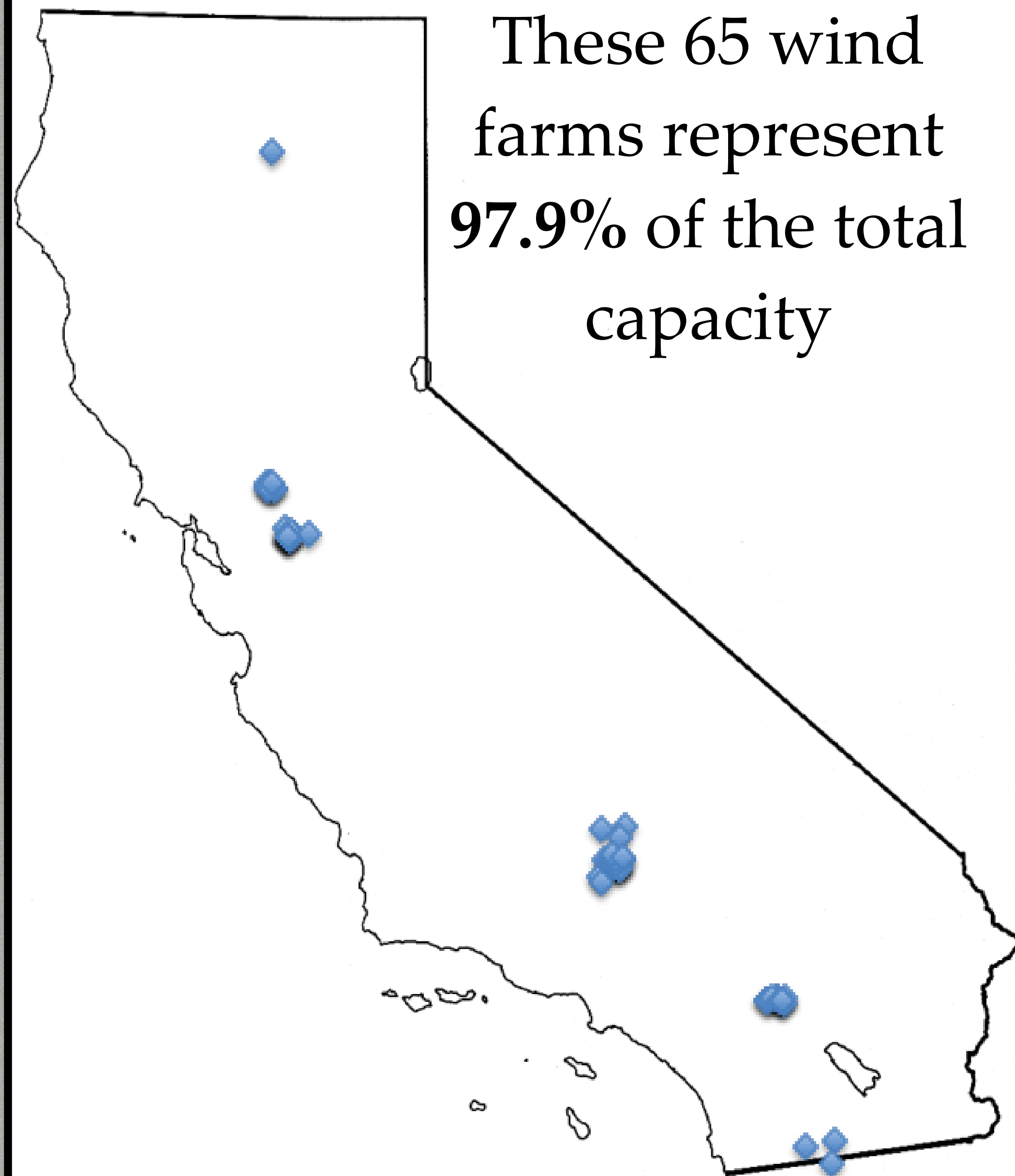
$$\rho = air\ density\ (kg/m^3)$$

$$U = wind\ speed\ (m/s)$$

$$U = f(Speed,\ Gust,\ Direc.)$$

$$\rho = f(T,\ P,\ hum.)$$

NOAA Data



These 65 wind farms represent **97.9%** of the total capacity

NOAA Data

1. Redding Muni. Airport

n = 1

C = **102 MW**

2. Port Chicago Station

n = 9

C = **1,009 MW**

3. Livermore Muni. Airport

n = 13

C = **584 MW**

4. Tehachapi Muni. Airport

n = 26

C = **3,210 MW**

5. Palm Springs Intl. Airport

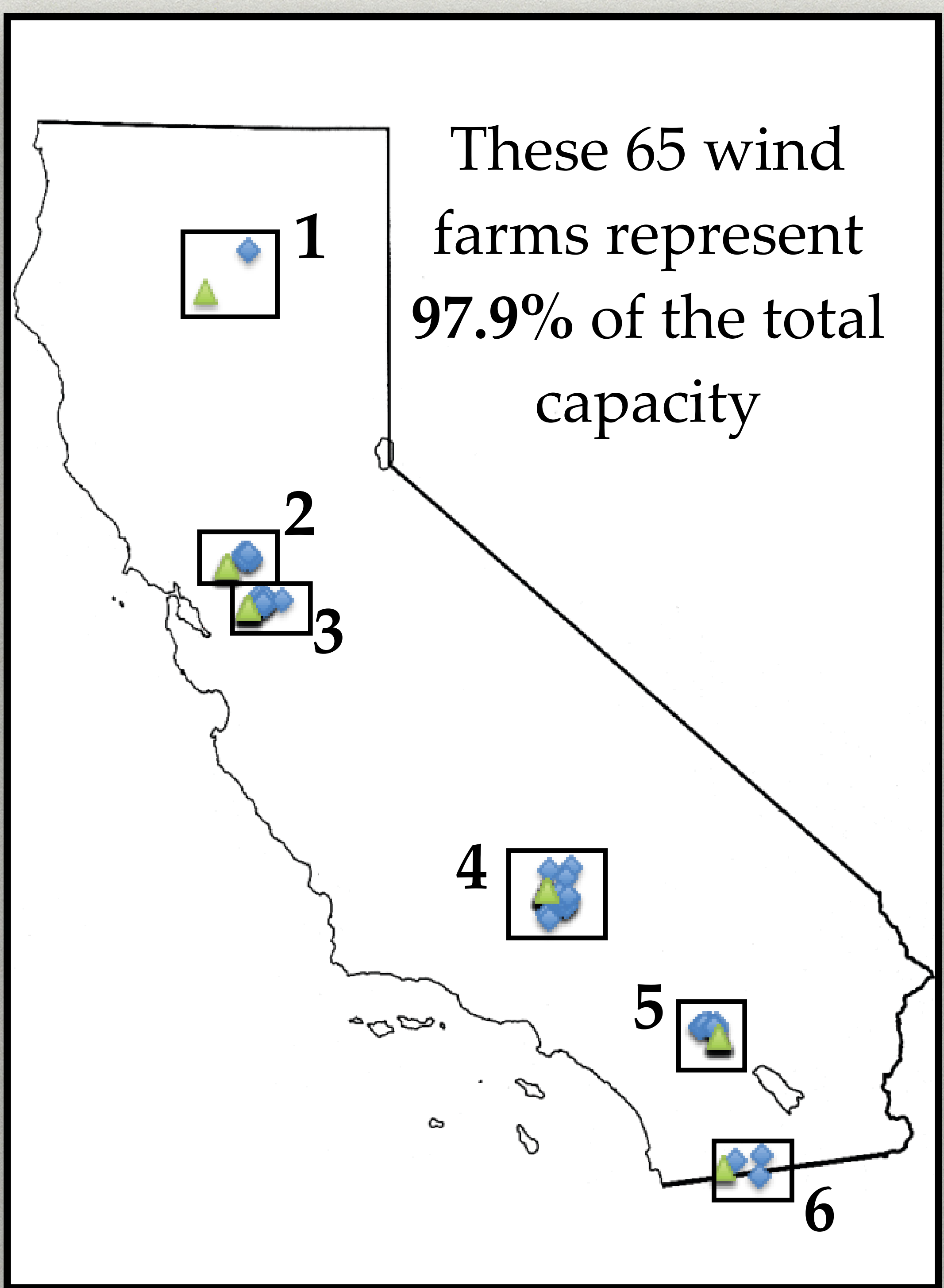
n = 13

C = **609 MW**

6. CAMPO Station

n = 3

C = **470 MW**



Simple Model Features

Feature	Reason for Inclusion
Past Output	Simple Regression

Complex Model Features

Feature	Reason for Inclusion
Past Output	From Simple Regression
(Wind Speed)^3	From Theory
(Gust Speed)^3	From Theory
Air Density	From Theory, Calculated as *f*(humid., T, P)
Max Sun Angle	Tracks Annual changes, i.e. “natural” time

Model Selection

Random Forest

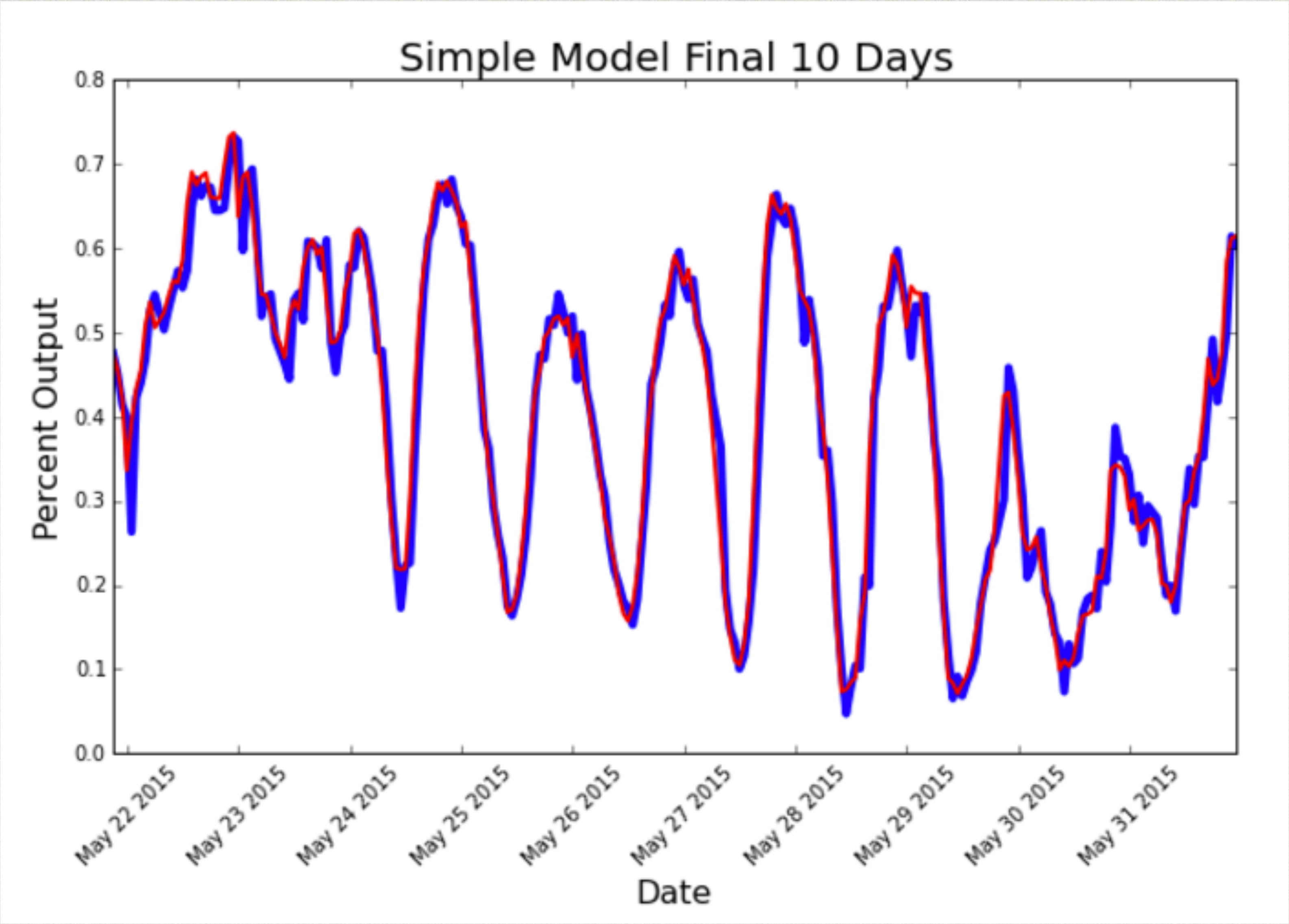
Pros:

- Regression - can predict a number
- Handles outliers well
- Can model complex, non-linear relationships
- Features interact, and may be correlated
- Dataset may have outliers

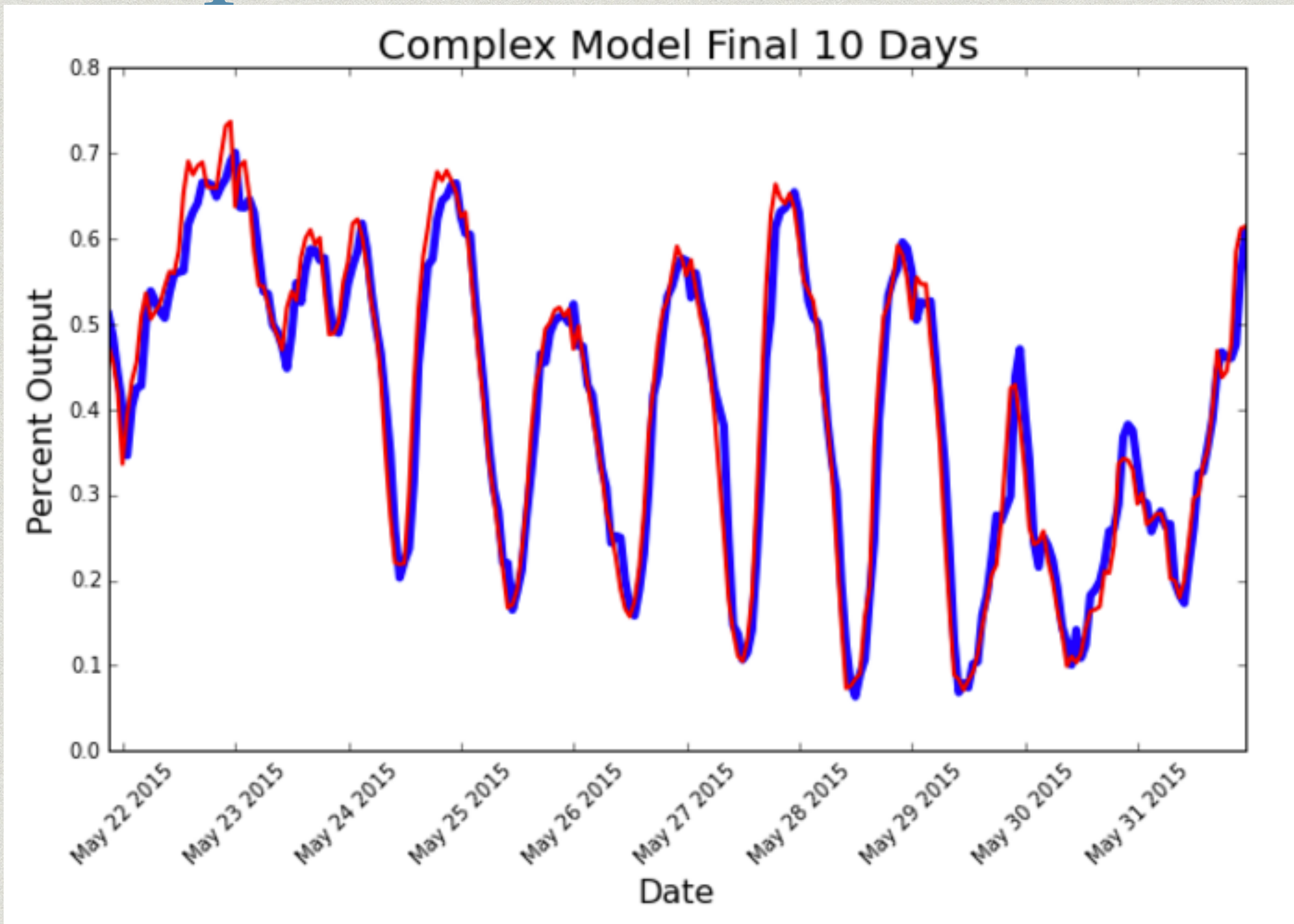
Cons:

- Takes significant computation time

Simple Model Results



Complex Model Results



Conclusions

- The simple model outperformed the complex model:
  - MSE_Simple = **11.9, 3.4%**
  - MSE_Complex = **14.6, ~3.7%**
  - MSE_Dummy = **402.9, ~20.1%**
- I have **more work** I want to do on this project
  - Use Support Vector Regression (**SVR**)
  - Try additional complex models (**feature selection**) that may supersede the simple model.