Air Quality Predictions

Bailey Bjornstad

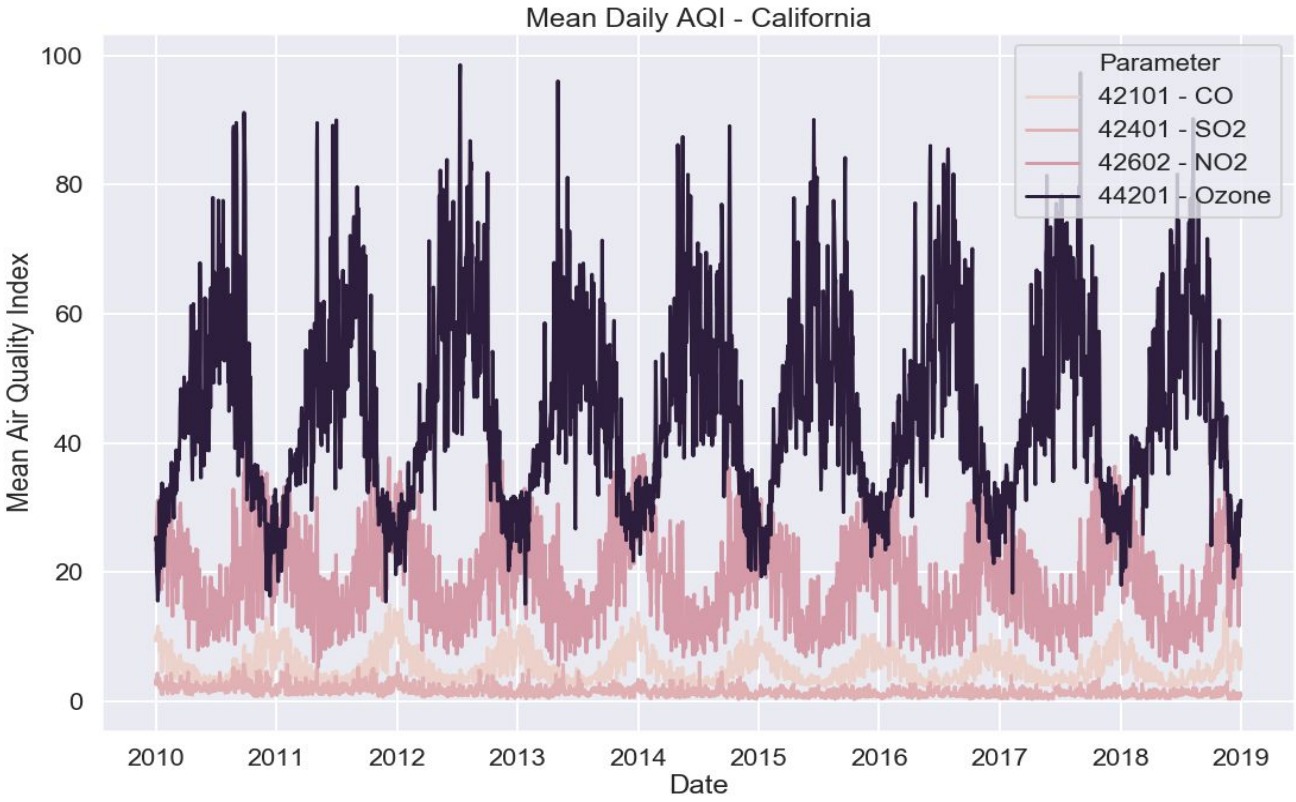
Project Introduction

- Continuation of previous work with EPA's Air Quality System database.
- Goal: Predict future levels of air pollutants
  - Air Quality Index: A relative measurement of the level of an air pollutant, computed from measurements, and defined by the EPA
  - Parameter: A particular pollutant

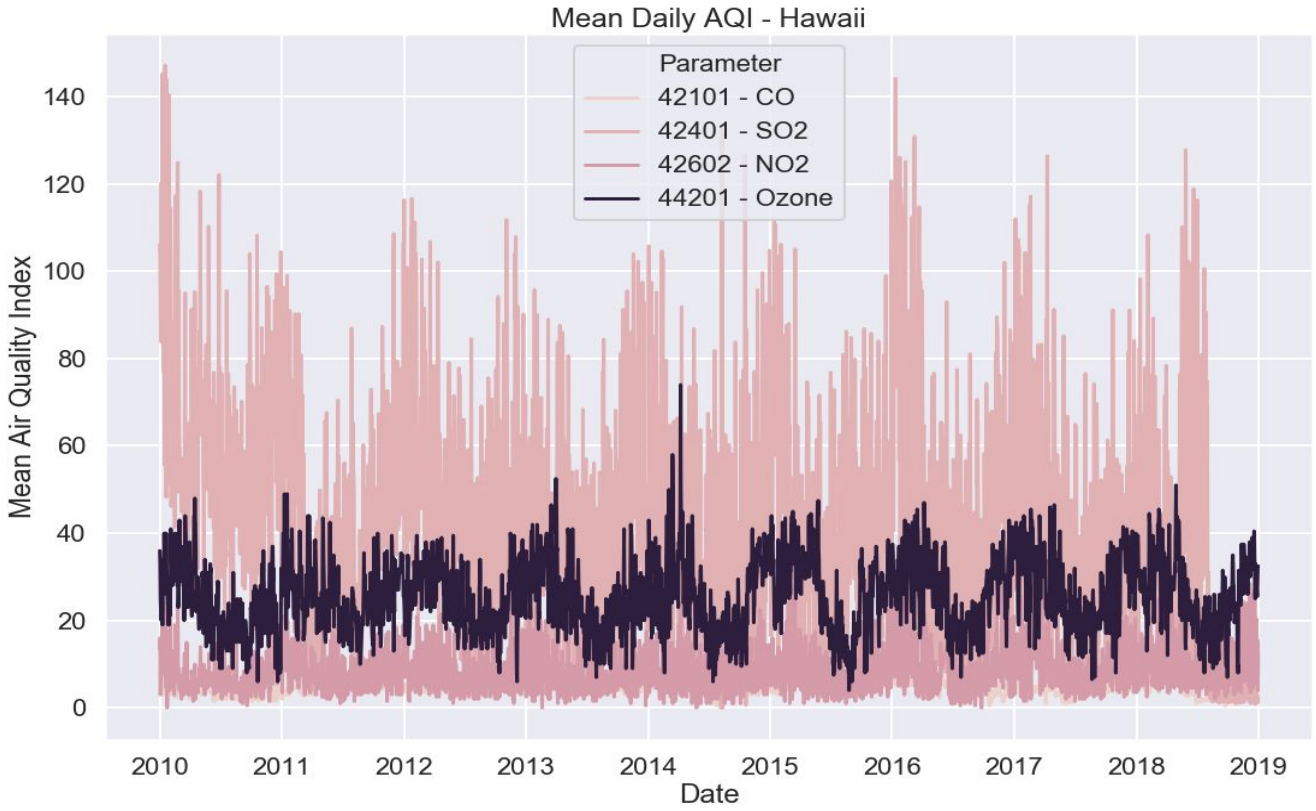
Database Information

- EPA's Air Quality System Database
- Daily summary data for four "Criteria Gasses" with well defined AQI:
  - Carbon Monoxide (CO)
  - Sulfur Dioxide (SO2)
  - Nitrogen Dioxide (NO2)
  - Ozone (O3)
- Collected from 2010-2018
- Some 7.5+ M filtered data points
- Daily aggregation
  - Within State
  - Within Parameter

Seasonality - Visualizations



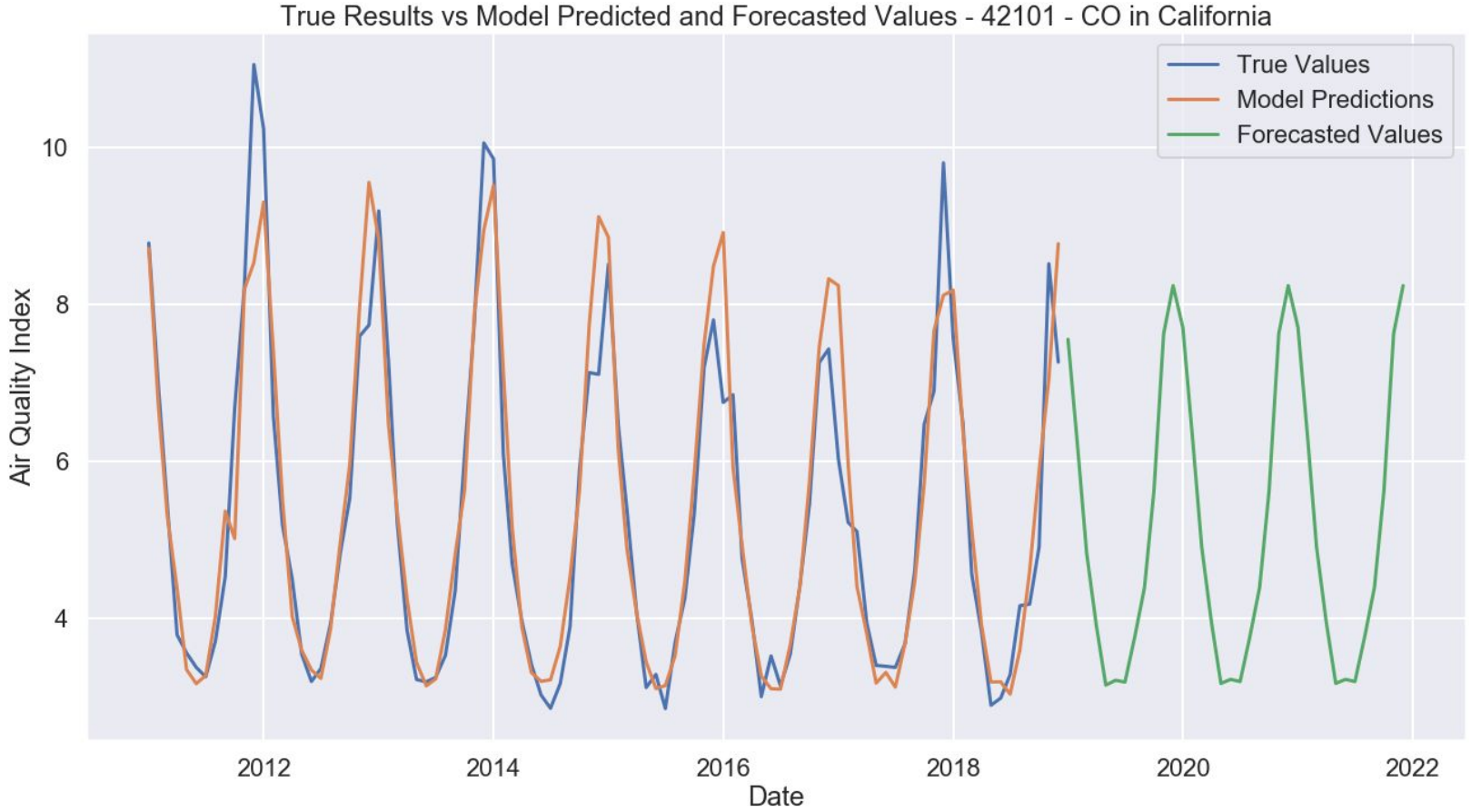
Seasonality - Visualizations



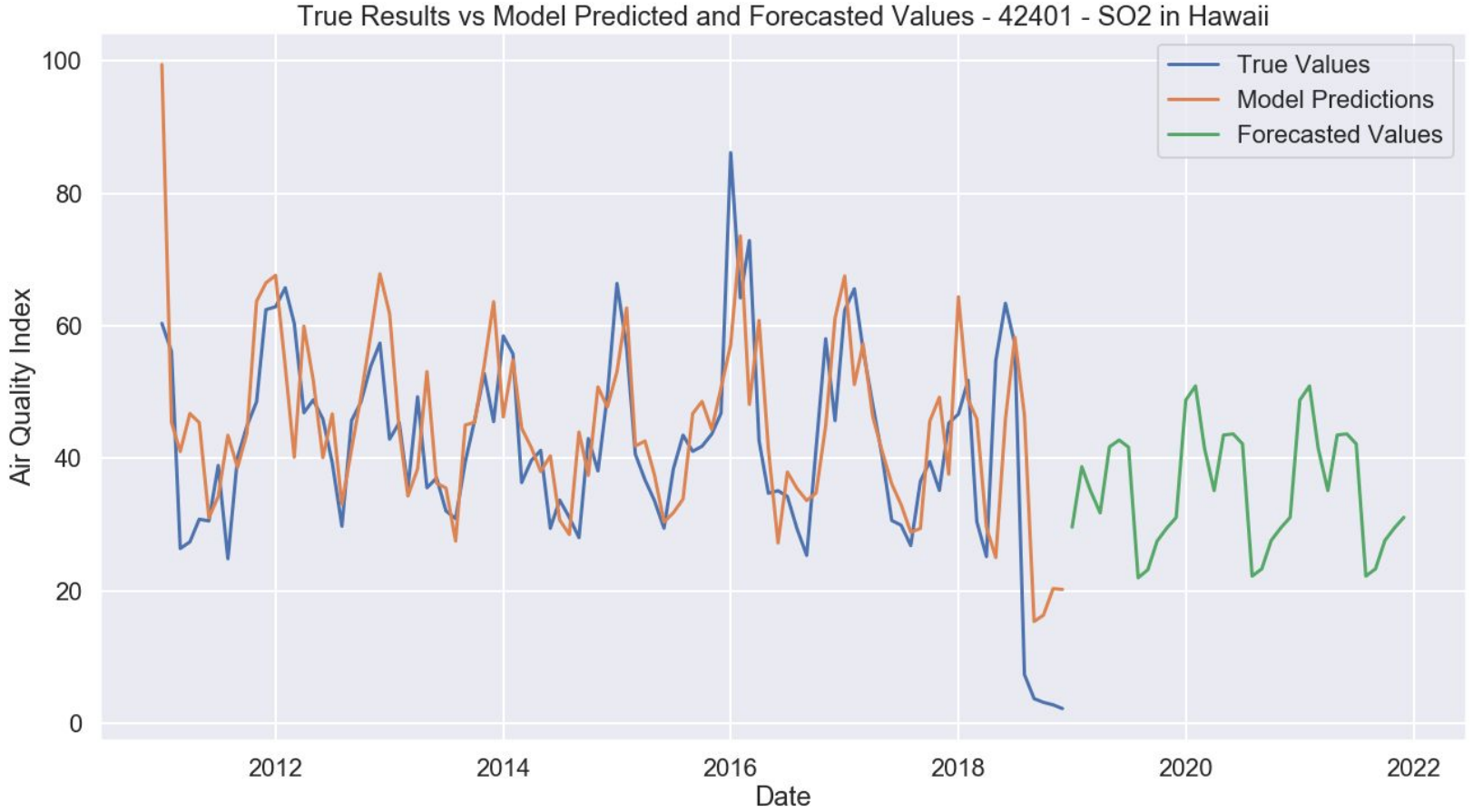
Modeling

- SARIMAX Modeling with Statsmodels
  - Statistical tests to ensure data meets assumptions for time series modeling
  - Monthly Resampling for Smoothing
  - Model fit for each parameter in each state
  - Metrics stored (AIC, BIC, MSE)
  - Models saved for future predictions

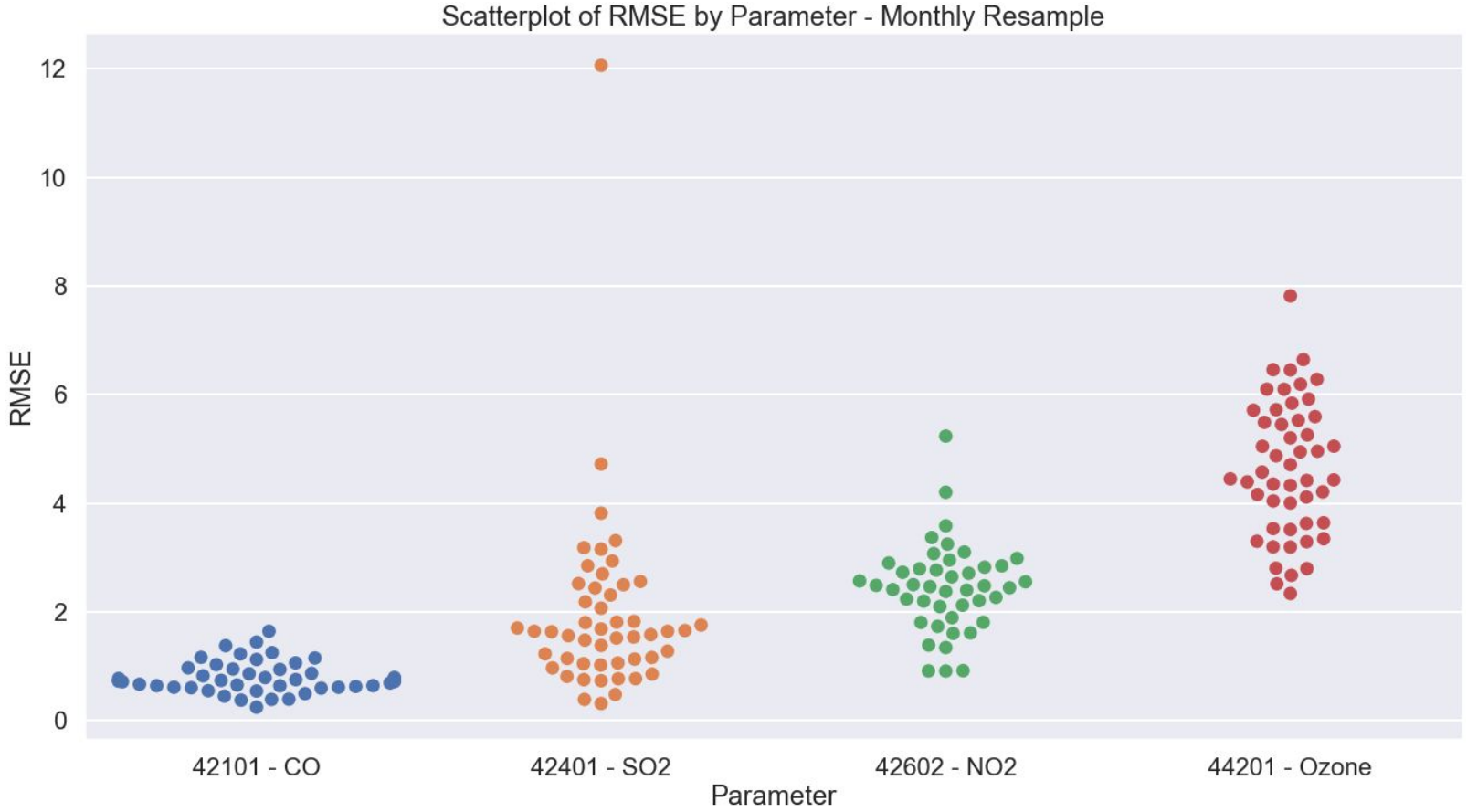
Modeling Results - Examples



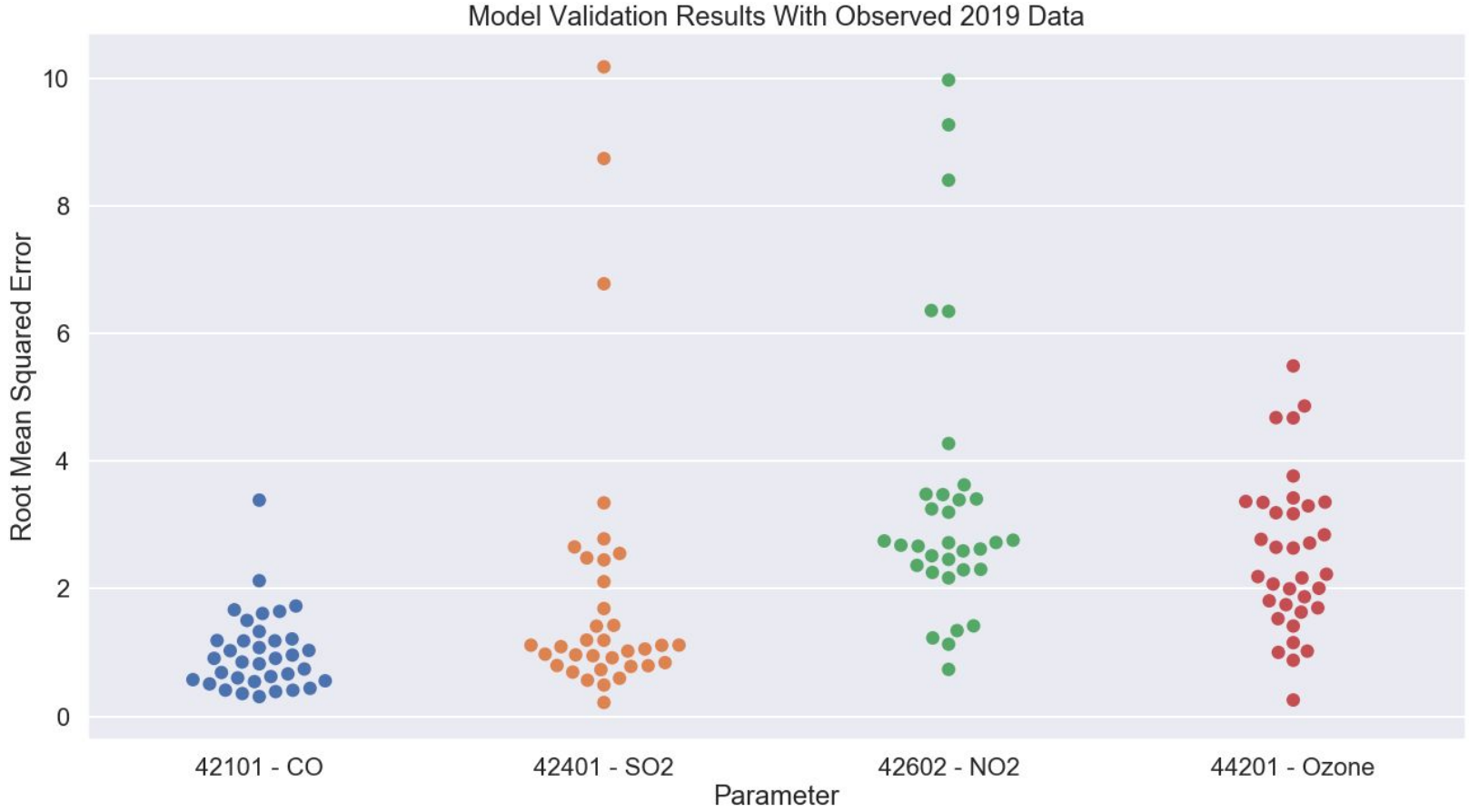
Modeling Results - Examples



Modeling Results - Error Distributions



2019 Validation Results - Error Distributions



Some Findings

- Error distribution is slightly more spread in the case of sulfur dioxide and nitrogen dioxide
- Error distribution is relatively tight in the case of carbon monoxide
- Perhaps this indicates a higher level of localization for carbon monoxide, as models are harder to predict in the case of sulfur dioxide and nitrogen dioxide.
    - Indicates a higher degree of “stochasticness” in the observed measurements.

Future Work

- Integration of more historical data
- Additional parameter fitting