Predicting Ground Level Ozone Pollution with Weather Data

Paul Black
Metis-Chicago
April 2017

The Hypothesis.

Tropospheric ozone pollution, measured in parts per million, can be predicted by weather related metrics.

What is tropospheric ozone pollution

The Data.

Weather Data from Wunderground.com
- L.A. (2005 - 2016)

Air Quality Index (AQI) Data from EPA.org
- L.A. (2005 - 2016)

Features include:

- Humidity
- Temperature
- Wind
- Condition

Target: Tropospheric Ozone Pollution

Feature Design

Minimum Temperature

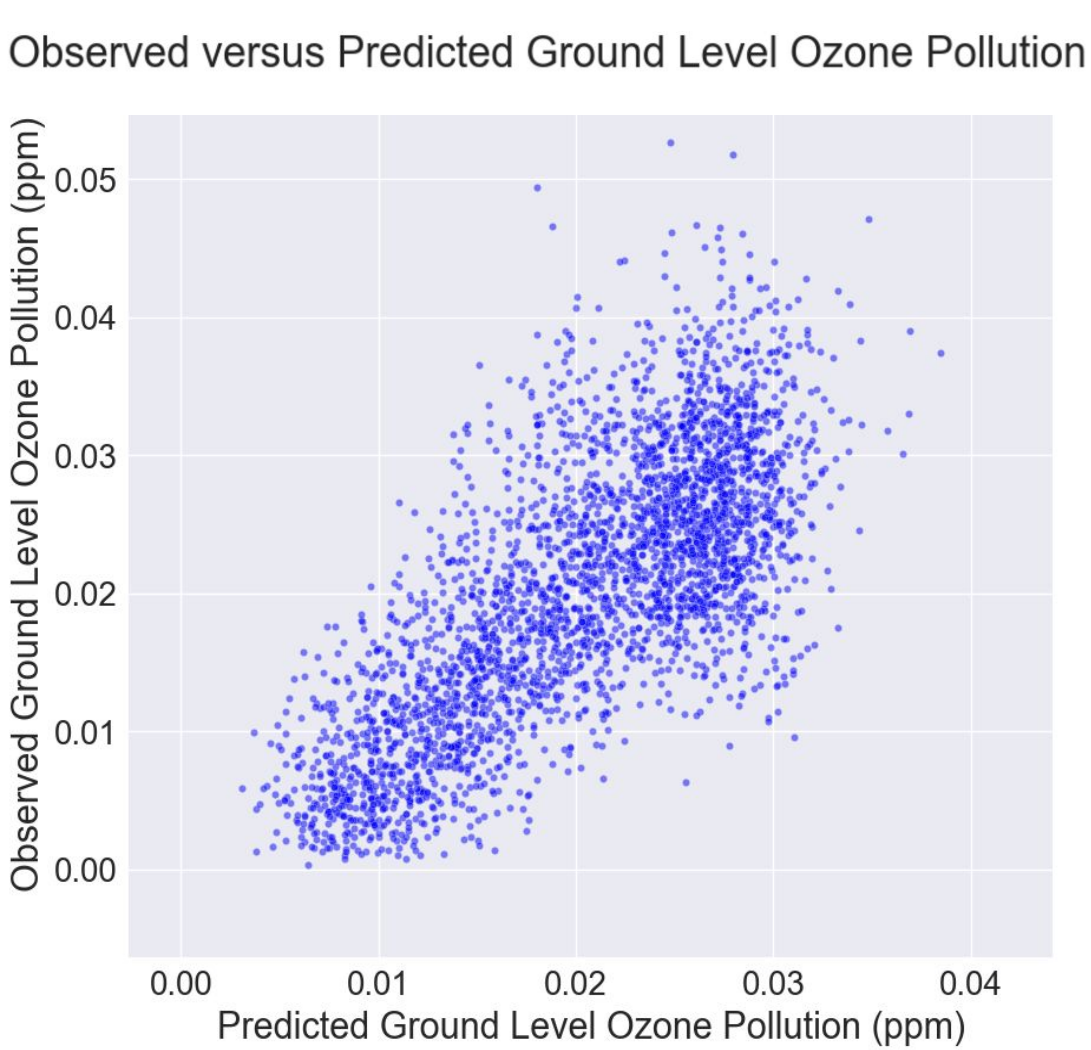
-- Higher min. temp. associated with sunnier days

Max Wind Gust Speed

-- Strongest effect of wind speed data

Overall Condition of Day

-- Quantifying a qualitative measure



Prediction Model -

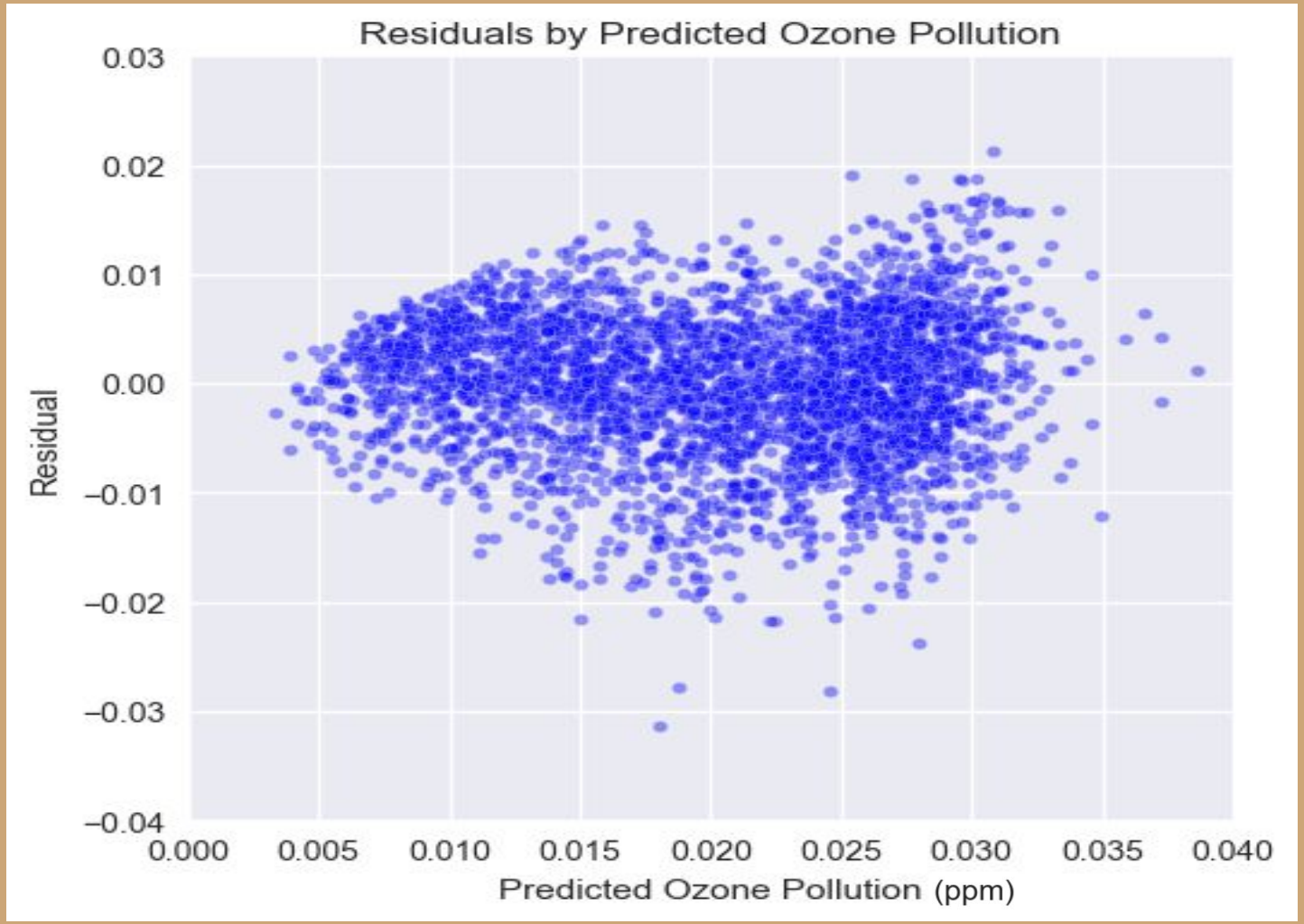
Model Type:
Linear Regression with L2 Regularization

Regularization Parameter:
1e-5

Mean R-Square value from cross validation:
0.5418

Sample Size:
3118 Days

Residuals


Residuals by Predicted Ozone Pollution
Residual
0.03
0.02
0.01
0.00
−0.01
−0.02
−0.03
−0.04
0.000
0.005
0.010
0.015
0.020
0.025
0.030
0.035
0.040
Predicted Ozone Pollution (ppm)

The Findings.

The linear regression model accounts for 54% of variation in tropospheric ozone pollution.

Features:

- Average Humidity (+)
- Dew Point (-)
- Max Wind Gust Speed (+)
- Minimum Temperature (+)
- Precipitation (-)
- Sea Level Pressure (-)
- Total Condition of Day (-)
- Length of Day (+)

The Future.

- Incorporate traffic data
- Globalize Model
  - Geographical and Industrial Features

Fin.