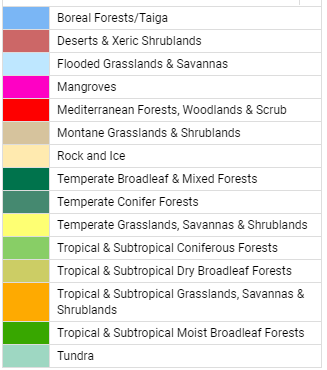
Features

* Application Based
  + Provide forecasting of dangerous Ozone levels
  + Alerts for current high ozone levels and high certainty future ones
  + Bake in functionality with latitude and longitude location rather than exact
* Codebase
  + Test multiple Machine Learning models
  + State that we made a new model;s
  + Train
    - Use seasonal models rather than comprehensive annual models
    - One model for each major biome type
      * 
      * From <https://ecoregions.appspot.com/> using GIS data
    - One model for region type (Urban, Suburban, Rural)
      * No because these different city types closer correlate to
    - Total of 144 Models (Way too much) for just one model tested
  + Test
    - Determine city type, biome,

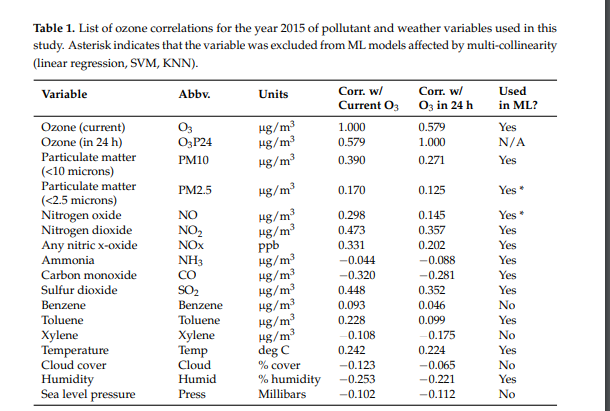
Presentation:

* Use correlation datasat to determine top 5
* Use only 5 pollutants to reduce the quality of sensor required to take data and as such maximum usage

General Notes

* “Indeed, combining PDE-based weather forecasting systems with ML-based methods for the prediction of air pollutants may combine the best of both approaches [10].”
* This study compares eight machine learning algorithms, which are representative of the categories of ML methods and are those in popular use: linear regression, KNN, SVM, Decision Trees, Random Forest, AdaBoost, XGBoost, and LSTM.
* Feature selection was used to reduce the total available training data set to 12 pollutant and 4 weather variables. Parameter tuning was then used to optimize the performance of each method individually
* In order to prepare the data fields for the regression models, invalid values with not-anumber (nan) entries were replaced with a linear interpolation between the surrounding valid values. These interpolated values were tracked, and never exceeded 3% of the full data.
* Weather Station and Pollutant station need to be close to one another (i.e. <10 km)

Parameters to look at:

* Air Quality: the increase of their concentrations negatively affects the quality of the air. We calculated the correlations between the features, of the air quality and we found a high correlation value between PM2.5, PM10, and CO as shown in Fig. 8.
* Meteorological feature Weather parameters (atmospheric temperature, atmospheric pressure, wind speed, wind direction, and relative humidity) affect air quality. For example, high wind speed will reduce the concentration of PM2.5, high humidity generally worsens air pollution, and high air pressure generally results in good air quality [50, 51]. Therefore, meteorological parameters are of prime importance for the task of forecasting air quality (Fig. 9)
  + Ensemble Methods (AdaBoost, Random Forest, XGBoost)
* List
  + CO
  + NOX (NO, NO2)
  + VOCs
  + Secondary pollutants (expand upon)
  + Sunlight
  + Air temperature inversions and general temp
  + Humidity
  + Previous Day Ozone
  + Particulate Matter (2.5, 5, 10)
* 

What the fuck does this mean

* Spatial analysis We performed the spatial correlation between Aotizhongxin station (target) and other adjacent stations. We used Pearson correlation to select the correlated PM2.5 monitoring stations around the target. The results are shown in Fig. 10. All correlation values are above 0.80 indicate that there is a strong spatial correlation between the selected stations. The data set has been split into two, a training set and a test set. 80% (28,052 h) of the dataset was taken as a training set. The remaining 20% (7012 h) becomes the test set used to test the model and analyze its accuracy.

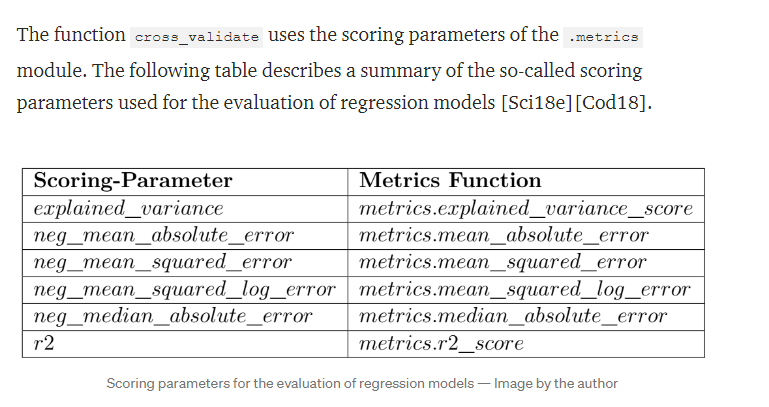
Methods

* Find Stations:
  + https://aqicn.org/data-platform/covid19/airquality-covid19-cities.json
* Get Current Data Air Quality
  + <https://aqicn.org/json-api/doc/>
* Get Historical Data Air Quality

Models

* Linear Regression
* Polynomial Regression
* Robust Regression — RANSAC
* Decision Tree
* Random Forest
* Gaussian process regression
* Support Vector Regression
* K Nearest Neighbors
* SVM
* AdaBoost
* XGBoost

Testing

* Use cross validation to prevent overfitting from skewing results
* 
  + Error minimize vs score greater

Other Methods

* Backfilling data collection to minimize computational intesntiy
* Aqi not used due to non-standardized measurements across globe