

Quantifying Air Pollution without High Quality Measurements

(Inspired from conversations with researchers at UC Irvine)

Issue Statement

Research grade air quality sensors are not densely distributed. This complicates quantifying how much different communities are exposed to different air pollutants.

However there are substantial amounts of additional data that one could use to train a model to predict pollution in the locations that don't have research grade sensors, such as temperature, proximity to pollution sources, elevation and so on. Moreover purpleair.com maintains a database of consumer-grade particulate matter sensor measurements. Though in principle these sensors do not produce measurements of the same quality as the EPA's sensors, they produce information that could be valuable features in a statistical learning model.

Proposal:

Build a model that predicts monthly, daily and / or hourly air pollution measurements on the basis of any features you can bring to bear on the problem.

Possible data sources:

- <https://www.epa.gov/outdoor-air-quality-data>
- <https://www2.purpleair.com> (Note we can give you instructions on how to directly download data from PurpleAir's servers.)