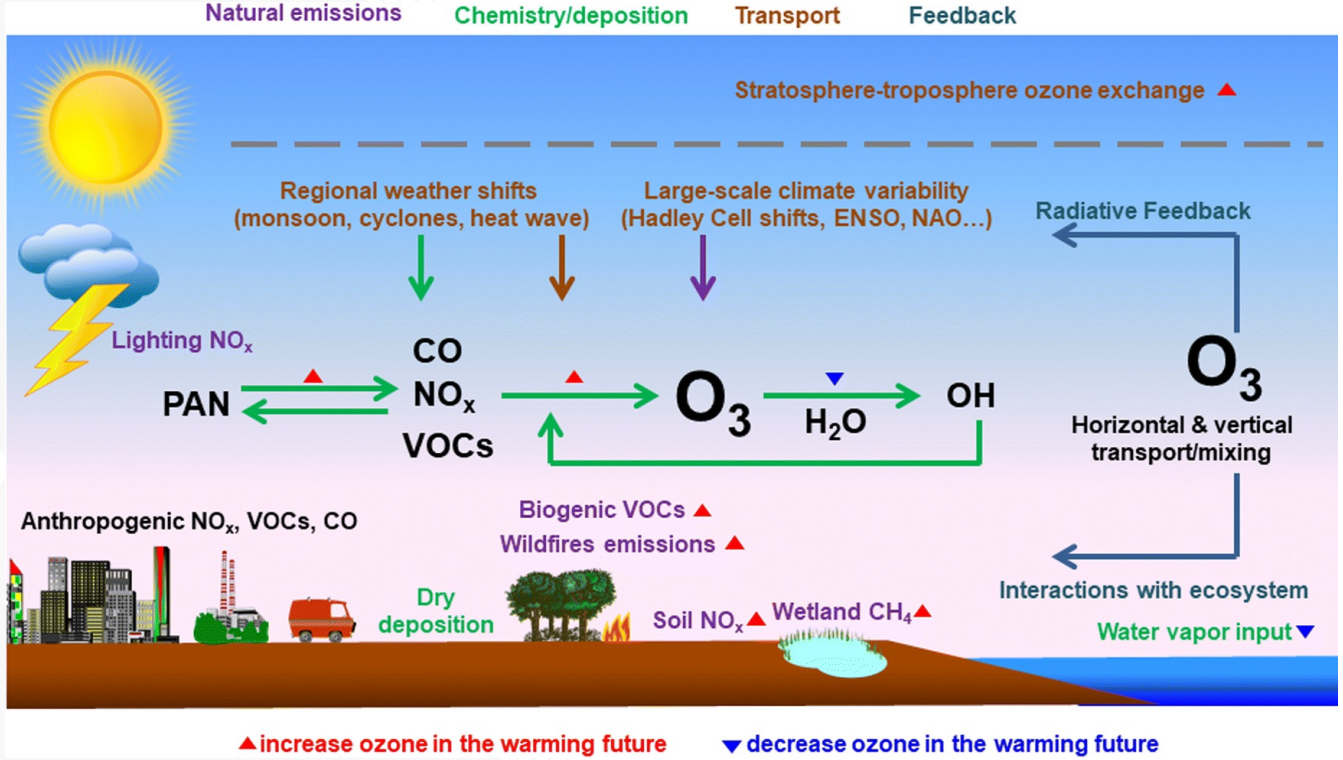
Ozone Concentration Prediction

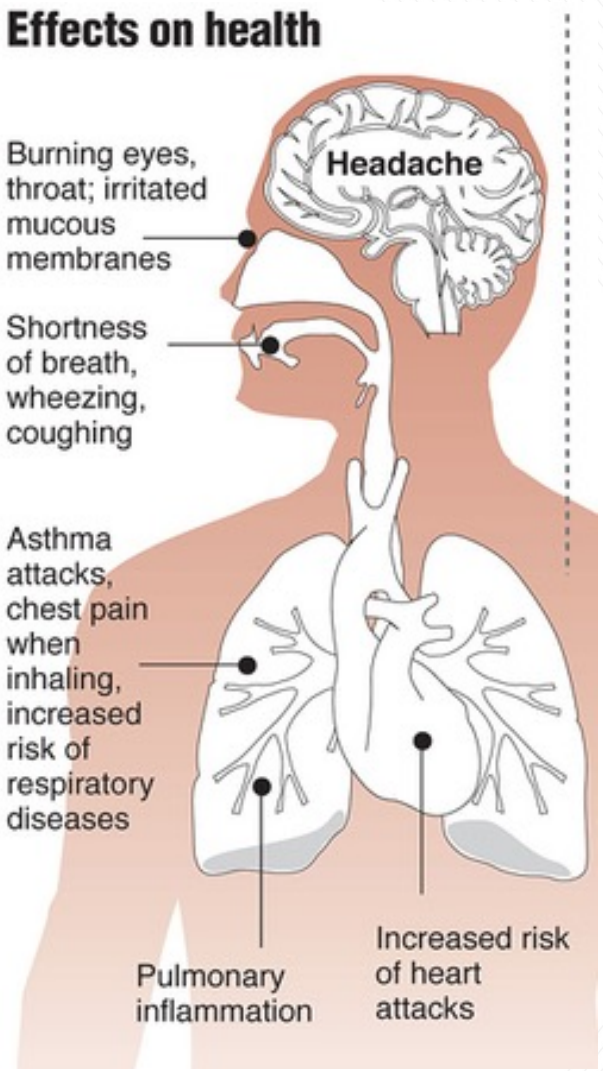
Lijing Zhai, Minyu Ma, Zephren Collison, Zongrun Li

Outline

- Introduction and Background
- Problem definition
- Methods
- Potential results and Discussion
- References

Introduction and Background





Since ozone concentrations depend on both weather conditions and emissions especially VOC (volatile organic carbon) and NO_x (NO_2 + NO), empirical model such as machine learning models trained by observation data can be conducted to predict ozone levels.

Ozone is considered as an air pollutant which causes respiratory and other health effects on humans.

Georgia Tech
CREATING THE NEXT

Problem definition

- Goal

We will focus on the data on South Coast Air Basin (SoCAB) in California during 1980 to 2020, building a robust ML model to predict ozone concentration, which is useful for ozone controlling and public health.

- Data Source
  - The ozone concentration data and weather data:

    https://aqs.epa.gov/aqsweb/airdata/download_files.html
  - geographic information:

    https://landfire.gov/lcp.php

Methods

In this project, we will provide two routines to train the empirical models.

One is that we use principal component analysis to reduce dimensions of data and then use polynomial regression to predict ozone by using selected features.

The other is that we use self-organizing map to reduce dimensions and then use artificial neural network to predict ozone.

Potentially, training models by using all features directly can be used to compare with these two routines. Random forest is another potential algorithm since it has better performance for discrete relationships than supervised algorithms we used above. To avoid overfitting and estimate the uncertainty of our model, data withholding will be conducted.

Potential Results and Discussion

Potential Results

We will build 2-3 different models based on several machine learning algorithms and evaluate them with their accuracy and robustness. From previous work on ozone research, we believe that the models will perform good.

Discussion

There are several potential applications for the models.

One is that spatial distribution of ozone will be predicted by applying the model for all places in SoCAB. The ozone levels for every place in the whole region can be used for estimating ozone exposure which is the product of ozone concentration and population. The ozone exposure is helpful for medical research related to ozone.

Another is that we can simulate ozone levels under different emission conditions. The simulation results can be used for environmental engineer to decide emission control policies.

References

[1] Balmes, J. R. (1993). The role of ozone exposure in the epidemiology of asthma. *Environmental health perspectives*, *101*(suppl 4), 219-224.

[2] Qian, Y., Henneman, L. R., Mulholland, J. A., & Russell, A. G. (2019). Empirical development of ozone isopleths: Applications to Los Angeles. *Environmental Science & Technology Letters*, *6*(5), 294-299.

[3] Andras, P. (2013). Function approximation using combined unsupervised and supervised learning. *IEEE transactions on neural networks and learning systems*, *25*(3), 495-505.

[4] Friedman, J. H. (1994). An overview of predictive learning and function approximation. *From statistics to neural networks*, 1-61.

[5] Jerrett, M., Burnett, R. T., Pope III, C. A., Ito, K., Thurston, G., Krewski, D., ... & Thun, M. (2009). Long-term ozone exposure and mortality. *New England Journal of Medicine*, *360*(11), 1085-1095.