Modeling Ozone with Triangulations of Bivariate Splines

Through a lens of environmental justice

A. Gamboa, L. Gant, T. Gant, A. Glover, J. Long

Emory University Math Department
Research Experience for Teachers

30 June 2023

EMORY

Research Goals:

1. Use sparse sampling stations to show variation in ozone at a neighborhood resolution.
2. Compare the ozone levels in a neighborhood to its social vulnerability.

EMORY

Scientific Background: Tropospheric Ozone

- Ozone high in the atmosphere (the stratosphere) is good
- Surface-level ozone is a health hazard from air pollution.
- "Bad" ozone is created by chemical reactions between oxides of nitrogen (NOx) and volatile organic compounds (VOC).
- Sources include industrial facilities, electric utilities, motor vehicle exhaust, gasoline vapors, and chemical solvents.[1]
- High ozone levels adversely affect the respiratory system and can aggravate lung diseases.[2]

[1] http://www.epa.gov/air/ozonepollution/basic.html

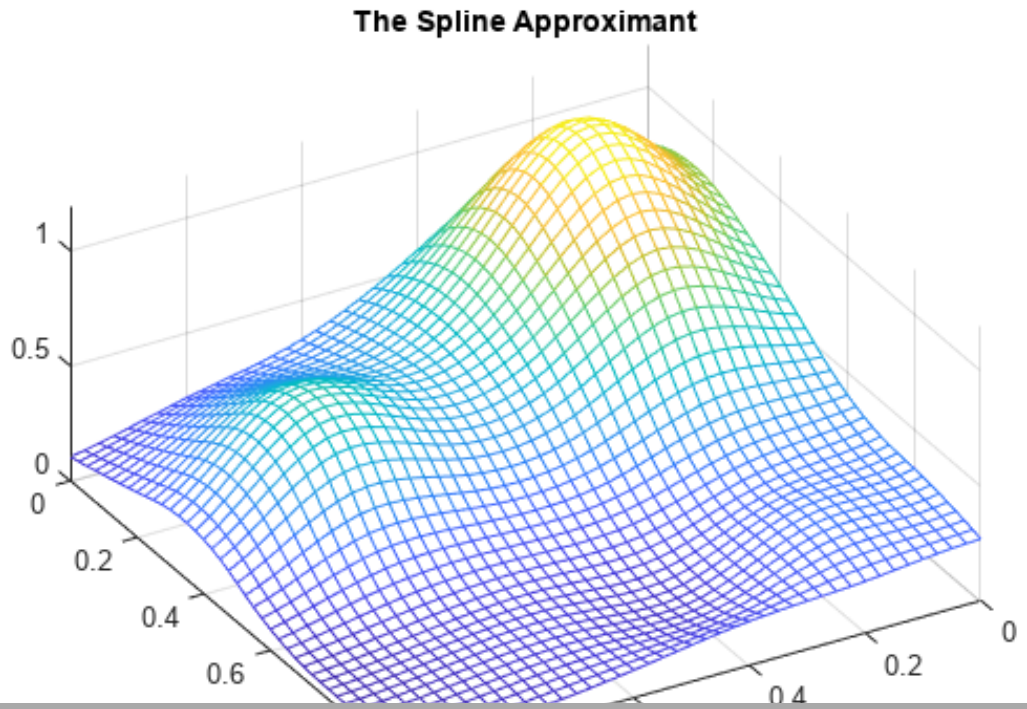
[2] https://www.epa.gov/ground-level-ozone-pollution/health-effects-ozone-pollution

Mathematical Theory: Bivariate Splines

Spline

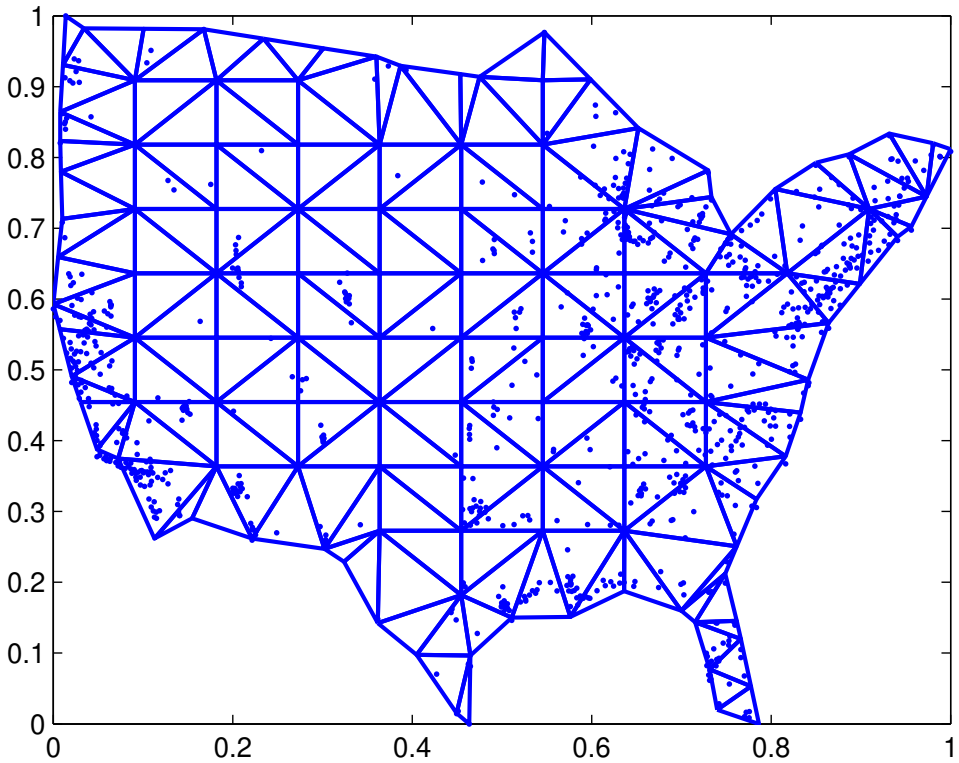
A piecewise polynomial that is continuous and smooth

A bivariate spline is a surface rather than a curve because it is drawn over a 2-dimensional domain



Mathematical Theory: Triangulation

Triangulation separates a given area into triangles with shared vertices (corners). Here is our triangulation of the U.S.



EMORY

Why Triangulation of Bivariate Splines?

Ozone measuring stations are relatively sparse. We are using triangulations of bivariate splines to fit this data so that we can get a resolution that shows variation in ozone levels between neighborhoods or regions in a city.

EMORY

Computational Method

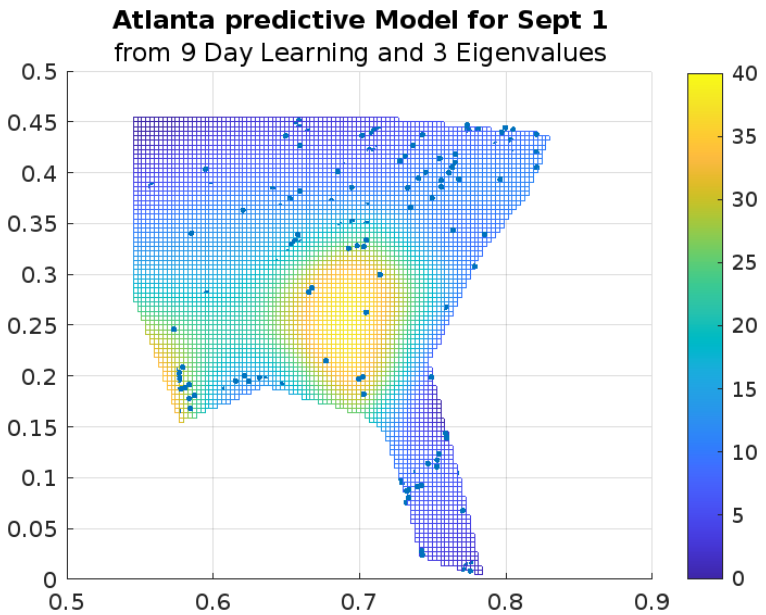
1. Fit historical measured ozone data with a spline surface.
2. Use Principal Component Analysis (PCA) to create a regression model to predict a given date.
3. Calculate the mean percent error to evalauate the success of the model.

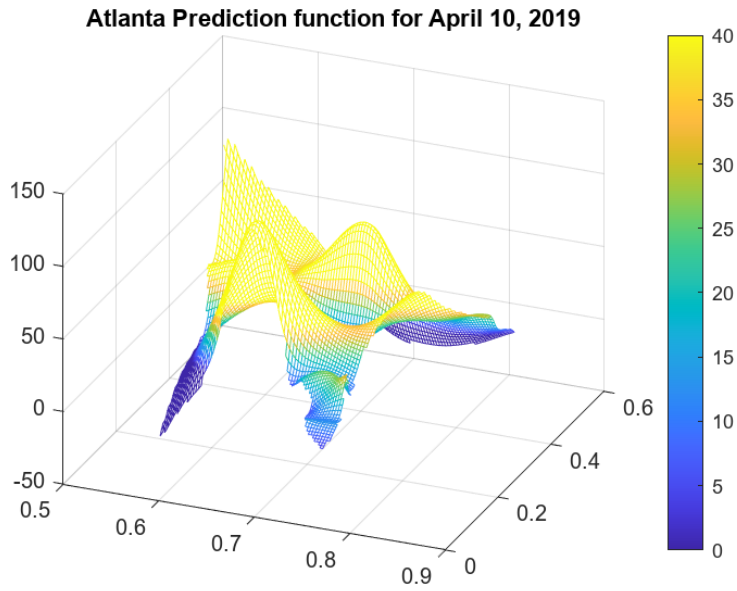
Ozone Datasets

- EPA sensor locations varies throughout the years.
- Satellite Datasets are for a daily fractional columns.
- EPA data is collected sparsely on small areas.
- Instrumentation and measurements lead to inconsistencies.
- Merged datasets over several years are uncommon.
- Merging multiple datasets is a complex task.
- The file extensions varies widely among datasets.

- Therefore, we have concluded that our best dataset is the hourly EPA from 2016.This dataset has more EPA ozone sensors for Atlanta and it also directly correlates to the data used for Social Justice, CDC and by the Census Bureau.

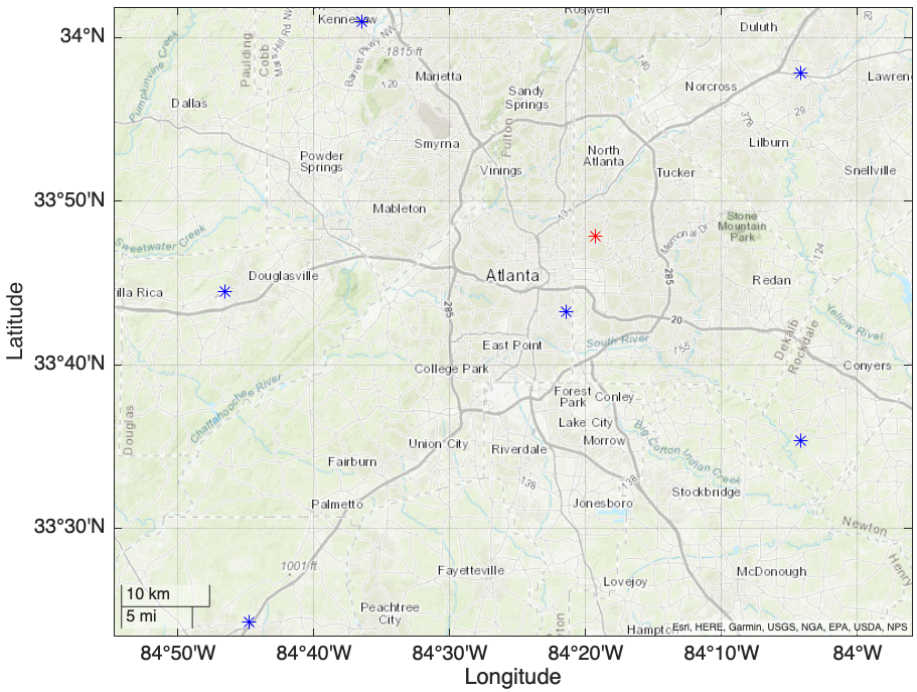
Ozone of South East United States

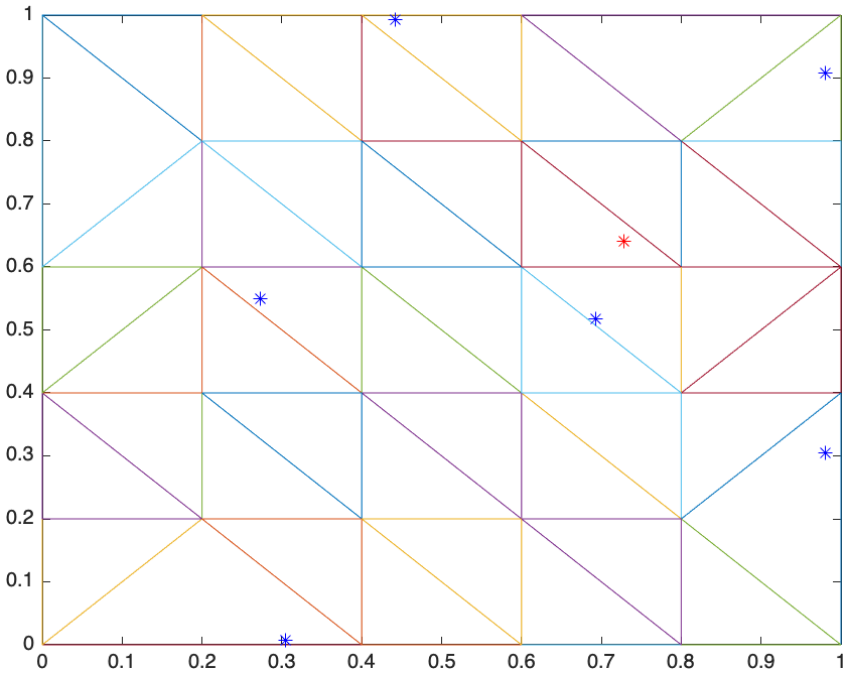




EMORY

Triangulation of Greater Atlanta





EMORY

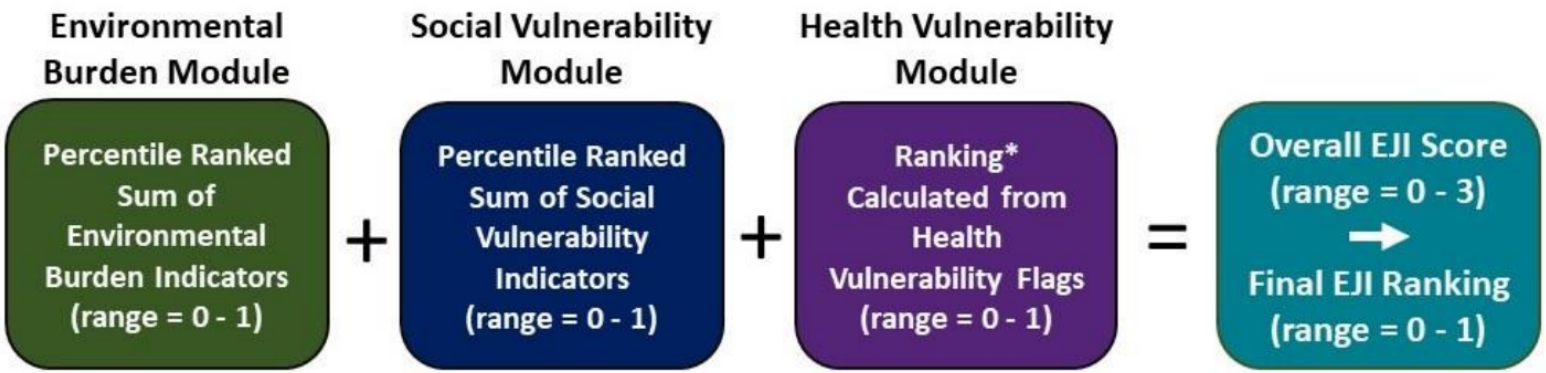
Ozone and Justice

- The CDC, EPA and other institutions have mapped data to census tracts in order to rank communities using indexes for Environmental Justice (EJI) and Social Vulnerability (SVI).
- For the Environmental Justice Index there are 36 factors grouped into 3 themes of Environmental Burdens, Social Vunerability and Health Vunerability.
- Data for Ozone is one of the component factors for environmental burden.
- The number of days that ozone is above the maximum concentration is mapped to our Metro Atlanta area.

EMORY

EJI Component Variables

CDC Environmental Justice Index Components



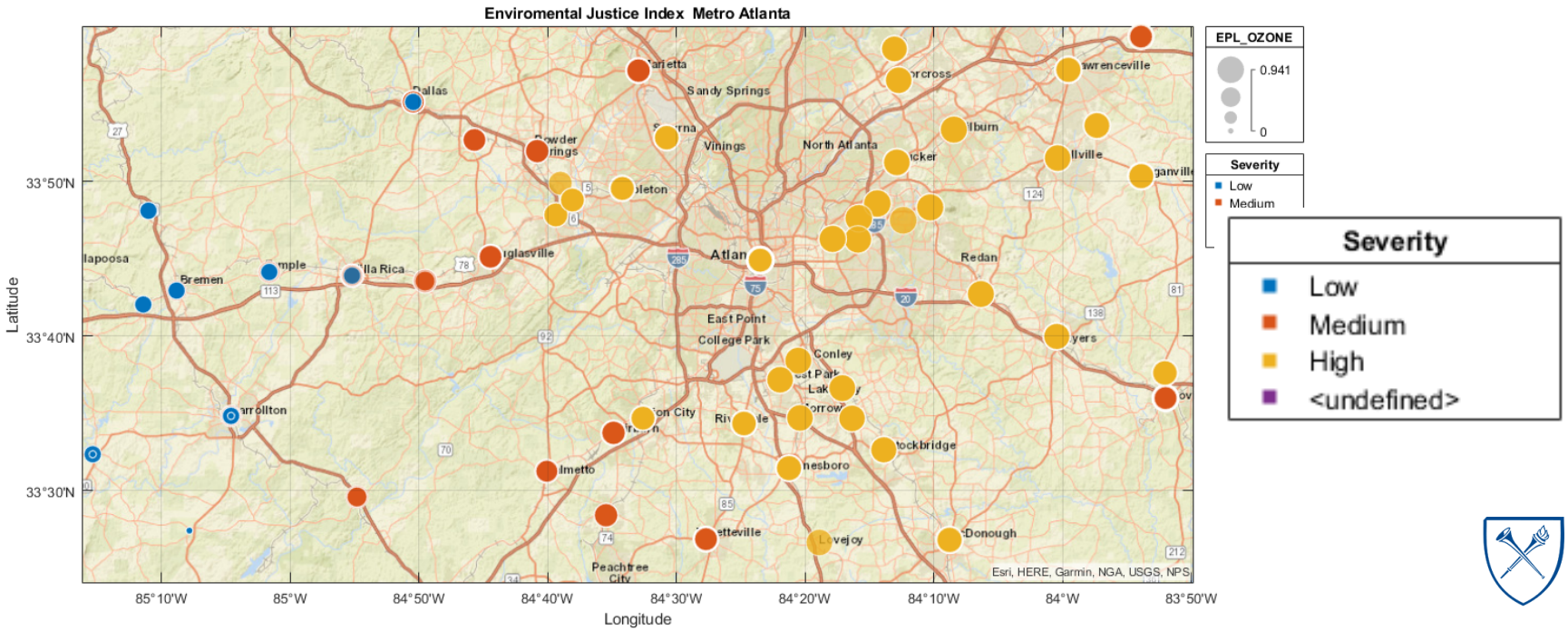
CDC EJI-2022 Documentation

The EJI uses data from the U.S. Census Bureau, the U.S. Environmental Protection Agency, the U.S. Mine Safety and Health Administration, and the U.S. Centers for Disease Control and Prevention to rank the cumulative impacts of environmental injustice on health for every census tract.

EMORY

EJI and Ozone in Greater Atlanta

Severity of numbers of days of Ozone levels above maximum concentration

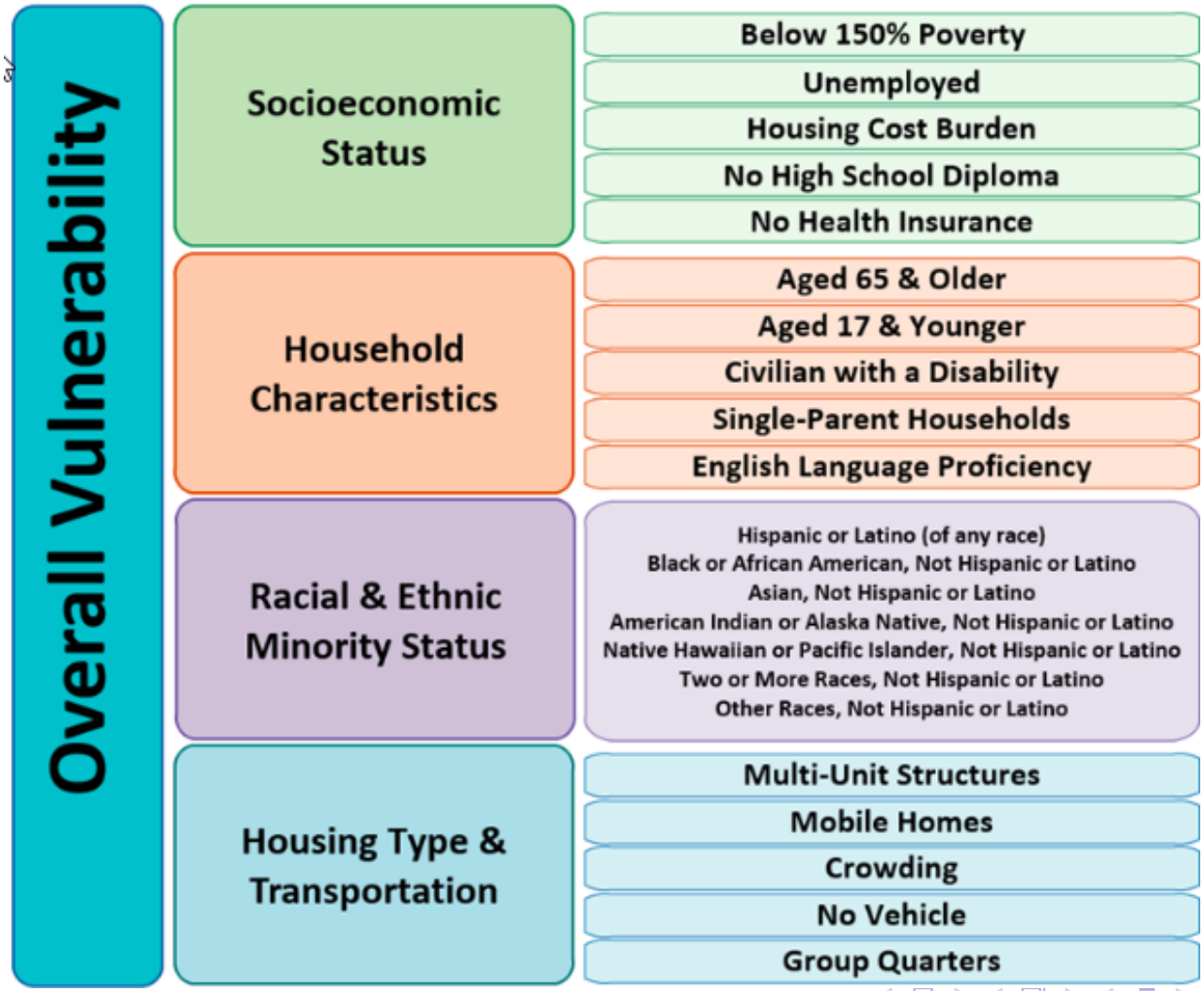


Ozone and Social Justice

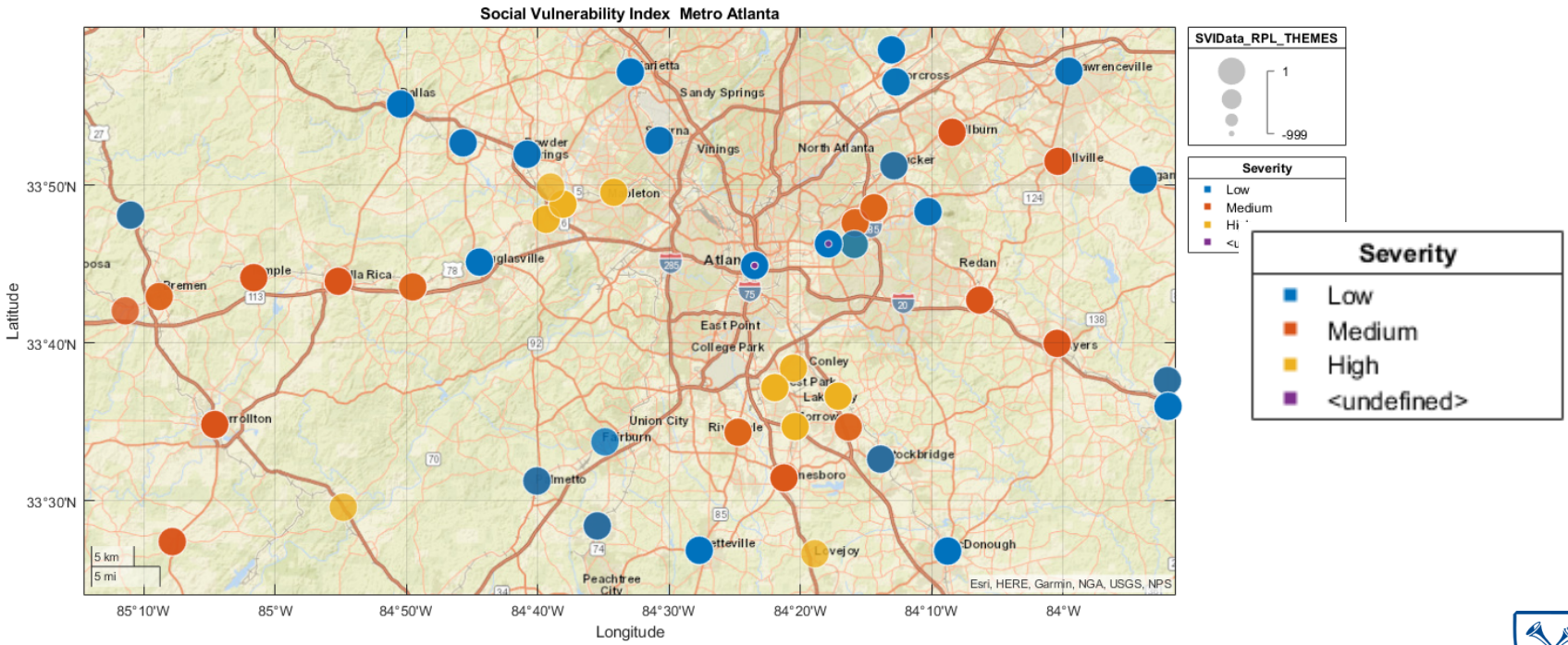
- Social Vulnerability Index (SVI) can be an indicator to be used for social justice.
- SVI is used to map communities that will most likely need support before, during, and after a hazardous event or exposure.
- SVI ranks census tracts, (county subdivisions), based on 16 social factors which are grouped into four related themes.
- Each tract receives a ranking for each of the 16 Census variable and for each of the four themes as well as an overall ranking which is the SVI.

EMORY

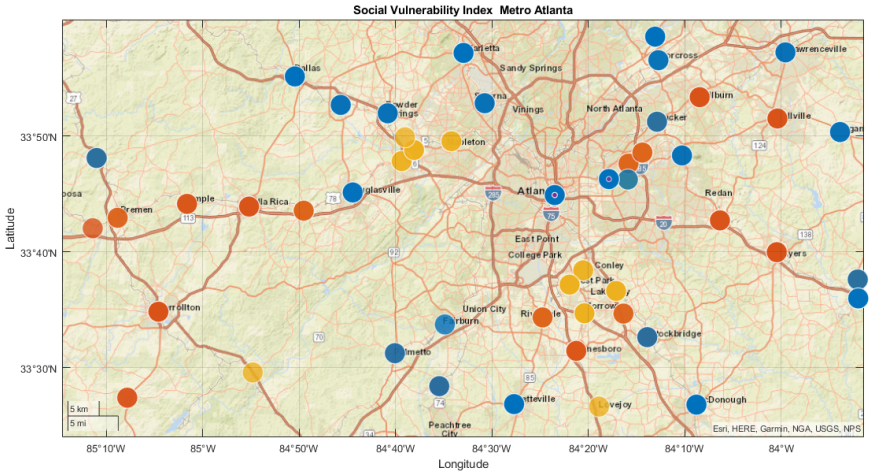
SVI Component Variables

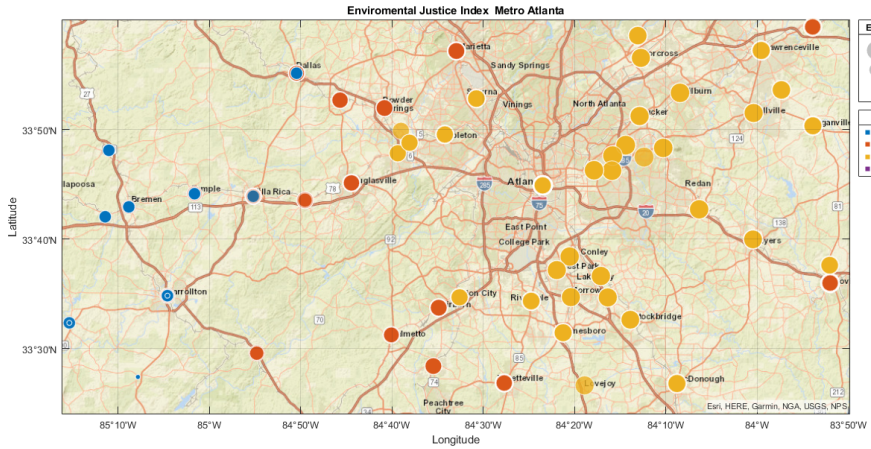


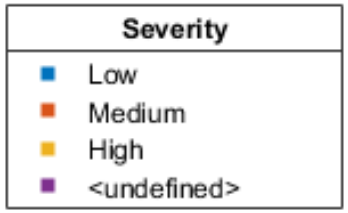
SVI of Greater Atlanta



Eyeball Norm of Ozone and Social Vulnerability







Red and yellow areas indicate high ozone exposure at max concentration in highly social vulnerability.

EMORY

Adaptive Eigenvalue Selection

Eigenvalue method α_n is

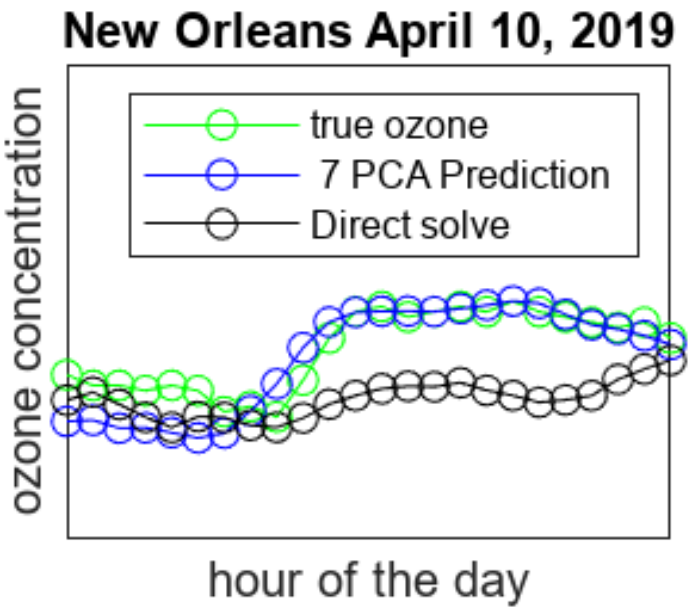
$$\widetilde{\alpha}_{PCR} = \sum_{j=1}^{k_n} \frac{\Delta_n(\widetilde{v}_j)}{\widetilde{\lambda}_j} \widetilde{v}_j + \rho SVI \tag{1}$$

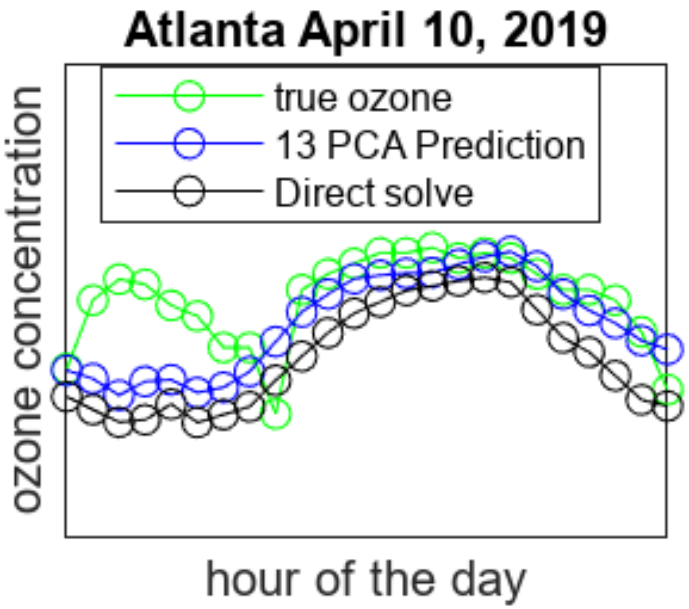
EMORY

Adaptive Selection of Eigenvalues

Our current predictive function can evaluate PCA and direct solve values and auto-select the best eigenvalue for the function.

- These two figures show the Direct solve and the PCA function evaluated with an auto-selective Eigen value and compared to known ozone values.

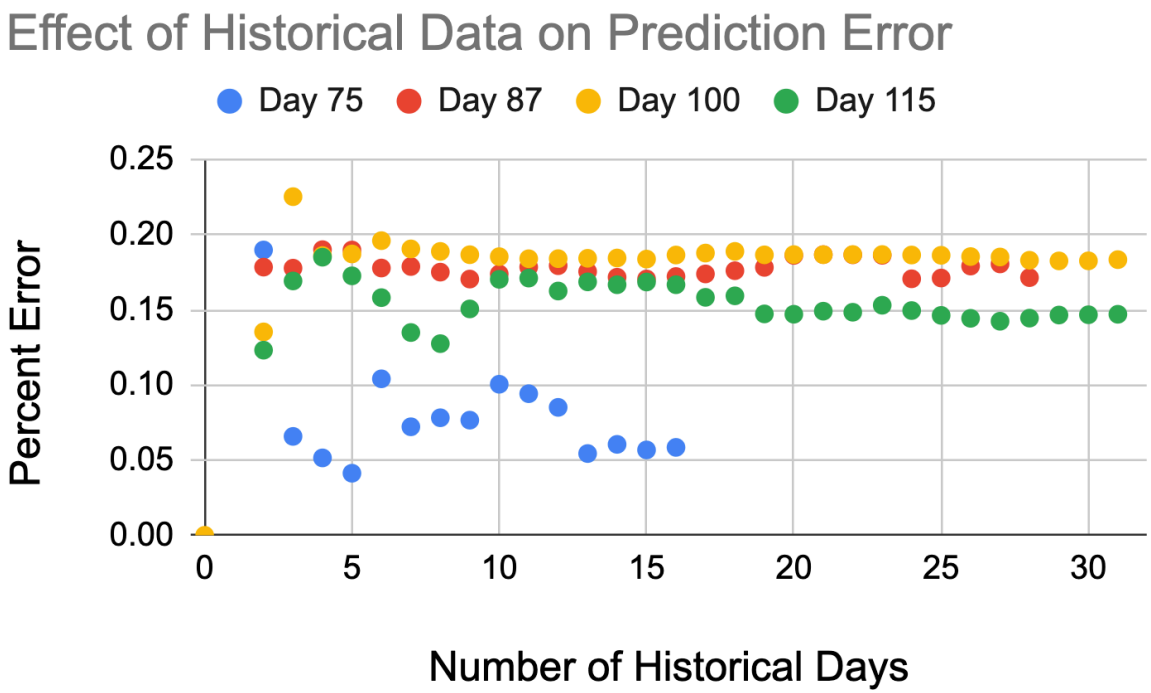




EMORY

How much historical data?

Analysis of the number of historical days using 2 eigenvalues.



Analysis needs to be repeated by either

- Directly solving the matrix rather than using PCA
- Using the adaptive eigenvalue method when it is complete.

Next Steps

- To match the zip code based on census tract to the latitude and longitude corresponding to the hourly ozone known values for Dekalb county.
- To improve our current Matlab functional regression model to pick up neighborhood scale variations
- Assess maximum ozone concentration from code correlated to Zip Code
- Determine the difference between the EPA max concentration and the percent error of our predictive model for ozone concentrations within specific EJI and SVI areas

EMORY

Acknowledgements

- National Science Foundation (NSF)
- Tropospheric Ozone Assessment Report (TOAR)
- Emory Math and Computer Science Department
- Dr. Bree Ettinger

EMORY

References

1. Spatial regression models over two-dimensional manifolds March 2016. B. Ettinger, S. Perotto, L. M. Sangalli Biometrika 103(1):71-88 DOI:10.1093/biomet/asv069

2. Bivariate splines for ozone concentration forecasting B Ettinger, S Guillas, MJ Lai - Environmetrics, 2012

3. Centers for Disease Control and Prevention and Agency for Toxic Substances Disease Registry. 2022 Environmental Justice Index.Accessed June 21, 2023. https://www.atsdr.cdc.gov/placeandhealth/eji/index.html

4. Centers for Disease Control and Prevention. Agency for Toxic Substances and Disease Registry. Geospatial Research, Analysis and Services Program. CDC/ATSDR Social Vulnerability Index 2020 Database Georgia. (https://www.stsdr.cdc.gov/placeandhealth/svi/data documentation download) Accessed on June 21, 2023.

EMORY

...

Thank you for your attention!

EMORY