Enhancing Ozone Prediction using Environmental and Social Justice Data

Antonio Gamboa [1], Lucretia Gant [2], Timothy Gant [3], Andrea Glover [4], Julianna Long[5]

[1] Garey High School, Pomona, CA, [2] Chamblee High School, Chamblee, GA, [3] Lithonia High School, Lithonia, GA, [4] Druid Hills High School, Atlanta, GA, [5] Stone Ridge School of the Sacred Heart, Bethesda, MD

Abstract

Preliminary results suggest a potential correlation between the ground-level ozone and socially vulnerable neighborhoods. By including the Center for Disease Control's Environmental Justice Index (EJI) as an additional principal component in the model we are able to produce a more accurate and socially informative distribution of ground-level ozone. Our prediction function is an effective tool to investigate and guide policies aimed to improve social justice.

This research project demonstrates that using bi-variate splines over a triangulation and an adaptive principal component selection is an effective method for daily ozone level predictions in a city. We demonstrate this model using the previous days' ozone concentration together with the EJI, a prediction was made for the ozone levels around Atlanta for a day.

Modified Autoregressive Model

Assume that the ozone concentration over $\mathcal{D}$ is a random surface, X, with known values at N locations. We consider the ozone concentration for each hour of a day for one city to be a functional, $f(X)$, of the ozone concentration surface, X, over $\mathcal{D}$ on the previous day. We use the bi-variate spline space $S_d^r(\triangle)$ with smoothness $r > 0$ and degree $d > r$ over a triangulation, $\triangle$, of $\mathcal{D}$ to approximate X using the given measurement values at the N locations.

Now we would like to solve $\hat{A}c = \hat{b}$ for c which will yield a coefficient vector for a spline S_α such that $f(X) \approx \langle S_\alpha, S_X \rangle$. To do this we:

1. Fit each X_i with a Penalized Least Squares (PLS) spline S_{X_i}.
2. Use S_{X_i} to calculate the empirical covariance $\widetilde{\Gamma_n}$ and cross-covariance and $\widetilde{\Delta_n}$.
3. Compute the Singular Value Decomposition (SVD) of $\widetilde{\Gamma_n}$ and analyze eigenvalues (also known as principal components) to find the number, k_n, that meets a percent variance explained threshold of 0.999.
4. Compute the prediction function $\widetilde{\alpha}_{PCR} = \sum_{j=1}^{k_n} \frac{\Delta_n(\widetilde{v_j})}{\widetilde{\lambda_j}}\widetilde{v_j}$.

Then we compute the EJI over the same domain $\mathcal{D}$ as a surface, Y, with known values at M locations. We use the same bi-variate spline space $S_d^r(\triangle)$ to approximate Y using the given measurement values at the M locations. Next we include this surface as an additional principal component to shape the prediction function as follows:

1. Fit the surface Y with a PLS spline S_Y.
2. Use S_Y to augment the prediction function $\widetilde{\alpha}_{PCR} = \sum_{j=1}^{k_n} \frac{\Delta_n(\widetilde{v_j})}{\widetilde{\lambda_j}}\widetilde{v_j} + \hat{\lambda} S_Y$

As seen in fig. 1, this enables us to include finer scale data in the prediction function. There are less than 10 EPA stations that measure ozone in the chosen region, and thousands of census tract measurements used to generate the EJI.

The Prediction Function

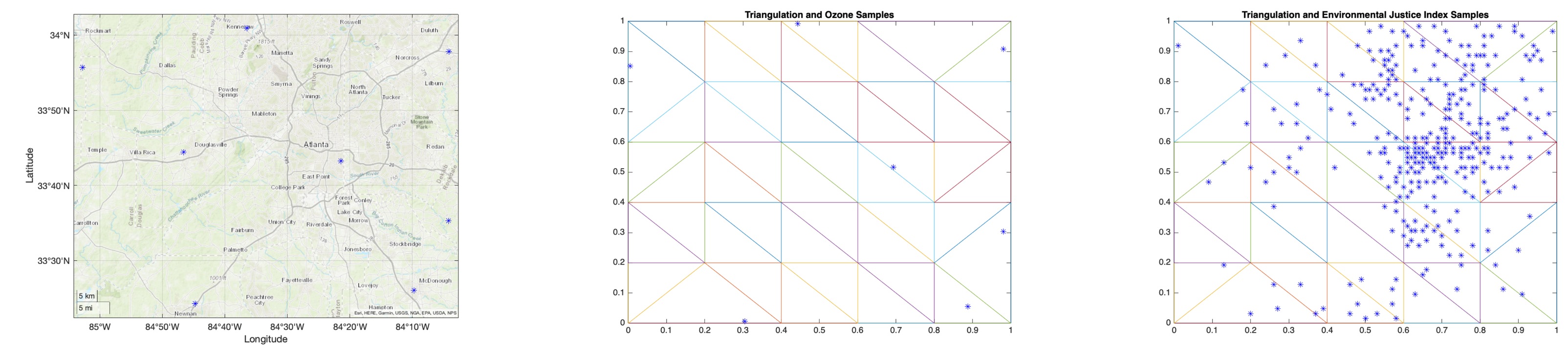

Fig. 1: This is a plot of the regions over the domain, the left figure is of the metro Atlanta area, the center figure is the triangulation of EPA ozone stations over the domain and the right is the triangulation of EJI for census tracts over the domain

Through our modified autoregressive model we are able to better predict the ground-level ozone concentration for a given day. The model is implemented in MATLAB learning over thirty previous days of hourly ozone concentration measurements. The model was able to effectively predict ozone concentrations from a few geographically sparse samples of ozone. See fig. 5.

The predictive model was implemented over a geographical rectangle that includes the city of Atlanta and its surroundings (fig. 1). The 2016 ozone was measured by the EPA's Air Quality System (AQS) and aggregated by the Tropospheric Ozone Assessment Report (TOAR). The EJI was generated by the United States Census Bureau. The geographical boundaries and datasets were selected for a sufficient density of surface-level ozone measurements in the Atlanta area.

Fig. 2 shows the surface fit to the EJI over the Atlanta area domain, while fig. 3 shows the ozone level prediction function made using only previous ozone measurements. These two surfaces are combined to create fig. 4, which shows the prediction function of ozone levels based on both previously measured ozone and the EJI over the Atlanta region.

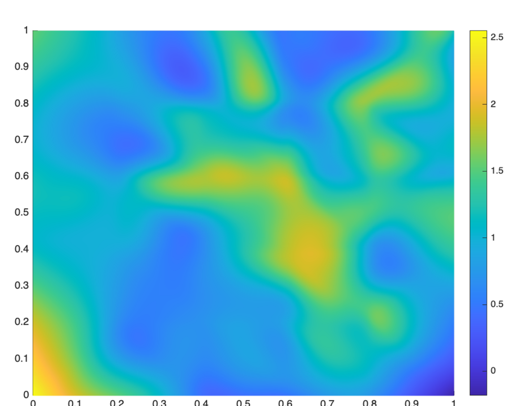

Fig. 2: Environmental Justice Index for Atlanta

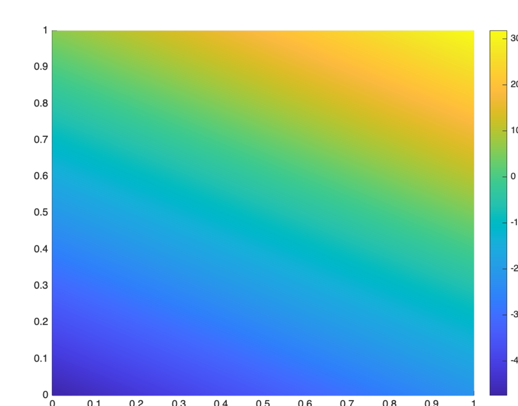

Fig. 3: Ozone-based Prediction function for Atlanta using only O3

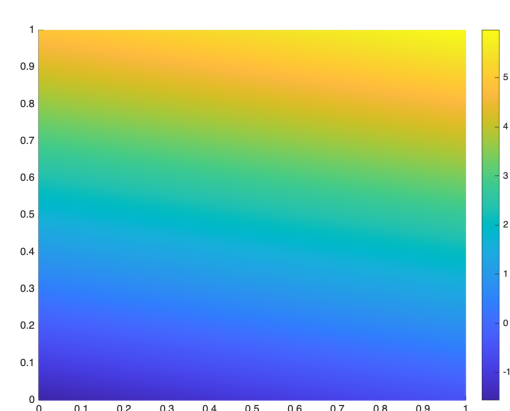

Fig. 4: Ozone Prediction Function for Atlanta based on EJI and O3

References

[1] Centers for Disease Control, Prevention, and Agency for Toxic Substances Disease Registry. *Technical Documentation for Environmental Justice Index*. URL: `https://www.atsdr.cdc.gov/placeandhealth/eji/docs/EJI-2022-Documentation.pdf`. (accessed: 06.21.2023).

[2] B Ettinger S Guillas MJ Lai. "Bivariate splines for ozone concentration forecasting". In: *Environmetrics* 23 (May 2012), pp. 317–328.

Results

To evaluate the efficacy and accuracy of the model, we chose one location (Emory University) and one day (July 10, 2016), to compare the actual measurement and the predictions made with and without EJI. Fig. 5 shows a comparison between the actual measurement and two predictions for each hour of the day. Here we see that the inclusion of the EJI values increased the accuracy of our predictive model.

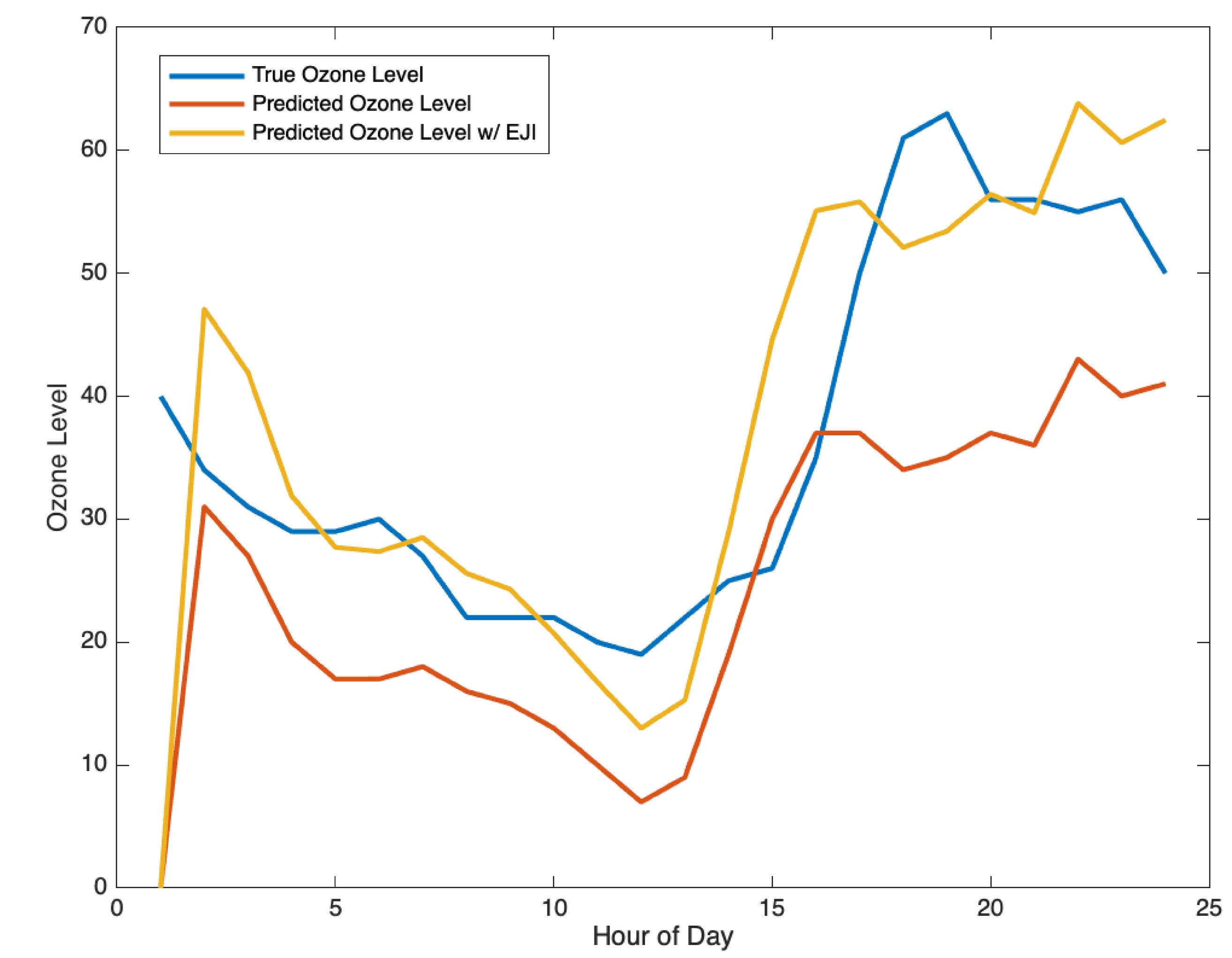

Fig. 5: The daily ozone prediction for a single location based the prediction functions in Fig. 3 and Fig. 4

Conclusion

Bringing social justice to communities that are impacted by unidentified factors depends on the guidance resulting from scientific research that exposes and discovers inequalities. The modified autoregressive model used here to evaluate ground-level ozone allows us to incorporate neighborhood level data where there may not be EPA stations. Providing a method to better predict low-level ozone for those locations.

Further work to improve and verify the model includes analyzing the accuracy of predictions over a longer period of time and in a variety of locations. In our report, we have investigated the Atlanta vicinity; however, these methods can be applied to any location within the United States. In addition, this method utilizes the EJI as a an additional source of variation, but the prediction can be augmented by using any other geospatial data.

Acknowledgements

This work is supported in part by the US NSF award DMS-2051019. We would like to thank Dr. Bree Ettinger for her mentorship.