Comparison of Interpolation Methods on Various Air Pollutants

Jimmy Zhang
May 6, 2024

Site & Mentors

Science Systems and Applications Inc.

10210 Greenbelt Rd, Lanham, MD 20706

SSAI specialized in data collection, algorithm development, model enhancement and other technical publication support deal with the environment.

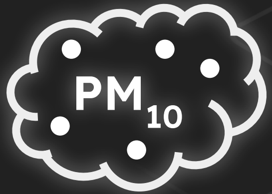
Mentors

- **Jackie Kendall** - Chief Knowledge Officer
- **Makhan Virdi** - Senior Research Scientist
- **Perry Oddo** - Research Scientist

Problem & Rationale

Climate change has caused an increase in smoke days from wildfires and pollen outbreaks, leading to poor air quality. Air quality monitors were developed to help people make informed decisions to reduce their exposure to poor air quality, but the monitors are expensive to maintain. To make air quality data more widely available, interpolation methods were used to predict air pollutant values for areas without an air quality monitor.

Air Pollutants Being Covered



PM2.5

Or particulate matter less than 2.5 µm in diameter accounts for smoke, fine dust particles and other fragments.

Ozone

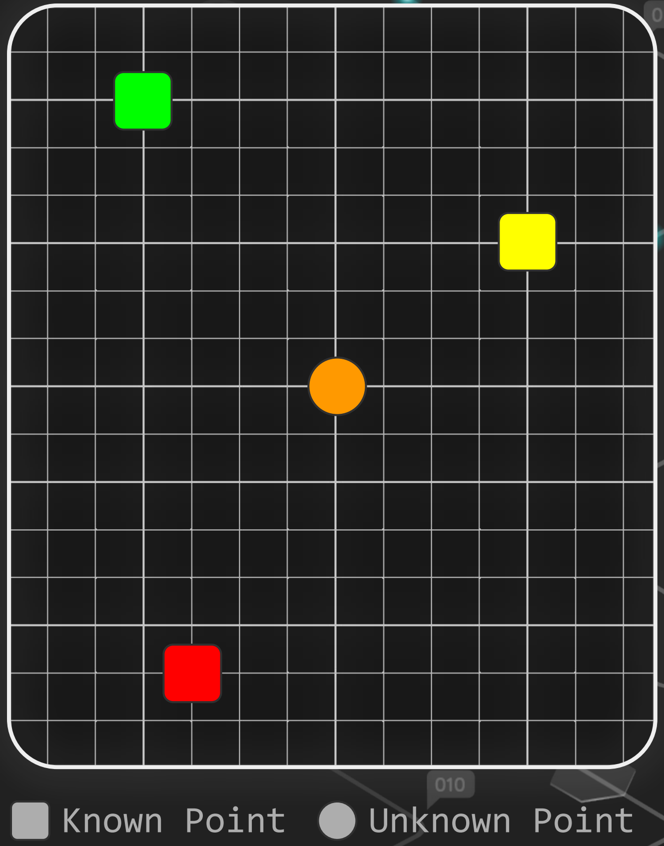
Is known for forming a protective barrier that prevents UV rays but surface level ozone can corrosive to cells due to their very reactive nature.

PM10

Or particulate matter between 10 µm and 2.5 µm in diameter accounts for pollen, bacteria fragments and coarse dust particles.

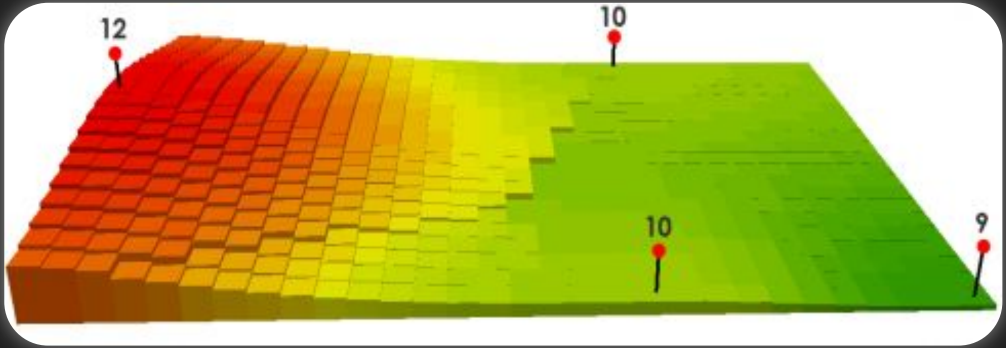
What is interpolation?

Interpolation is a type of estimation where the known points are used to calculate the value of the unknown points based on spatial factors like distance. This study aimed to find the optimal interpolation method for various air pollutants within the bounding box of Maryland. The models being tested are Inverse Distance Weighted (IDW), Linear Kriging, Exponential Kriging, Gaussian Kriging, and Spherical Kriging.



Inverse Distance Weighted

IDW was a deterministic interpolation method that used calculates the weight of the known points as an inverse of distance. The recommended parameter or power of the inverse is 2 so IDW would not create a heatmap of all the values around the same number or have the values concentrated around the known points and everything else is around 0.



3D Heatmap by *GIS Geography*

$$X = \frac{\sum_{n=0}^{n} \left(\frac{Z}{d^{p}} \right)}{\sum_{n=0}^{n} \left(\frac{1}{d^{p}} \right)}$$

KEY

X: Value of Unknown Point

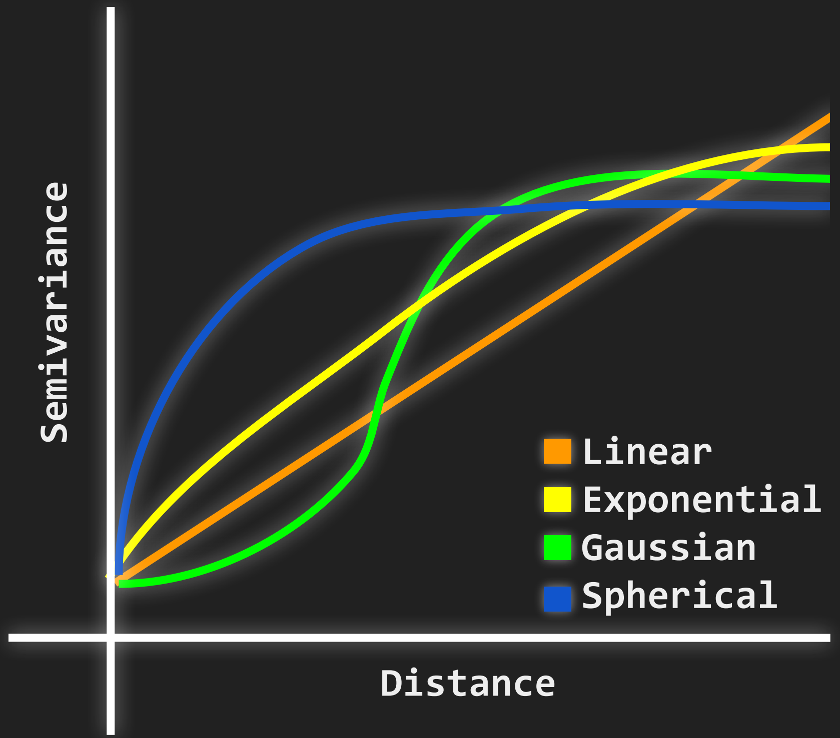
Z: Value of Known Point

d: Distance

p: Parameter

Kriging Methods

Kriging is a probabilistic interpolation that uses a variogram (variability as a function of distance) to calculate the spatial relationship between the known and unknown points. They differ into how the variogram is calculated, thus producing different results. The images on the right shows the variogram for ozone on UTC 2023-09-01 00:00.



Methods

Gathered $PM_{2.5}$, Ozone and PM_{10} data from EPA AirNow

Removed any invalid data (missing or negative values)

Separated the data into based on the air pollutant type

Separated the data into the known data and sample data

Interpolating Point Data

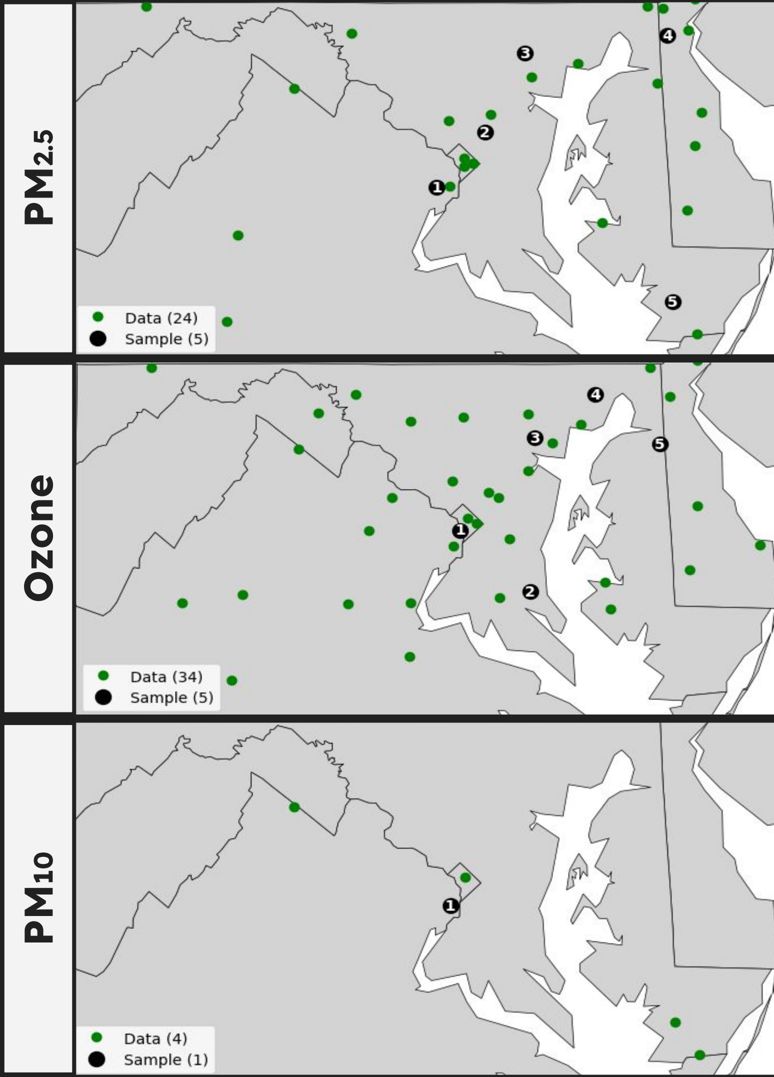
Loop over every UTC hour and uses the known locations and their values to interpolate for each of the sample locations

Find the Optimal Interpolation Method for $PM_{2.5}$, Ozone and PM_{10}

- Use a time graph to see if the interpolation methods could shows a local smoke event and international smoke event
- Use RMSE and R^2 to find how well the interpolation methods could interpolate the actual value of the sample location

Monitor Locations

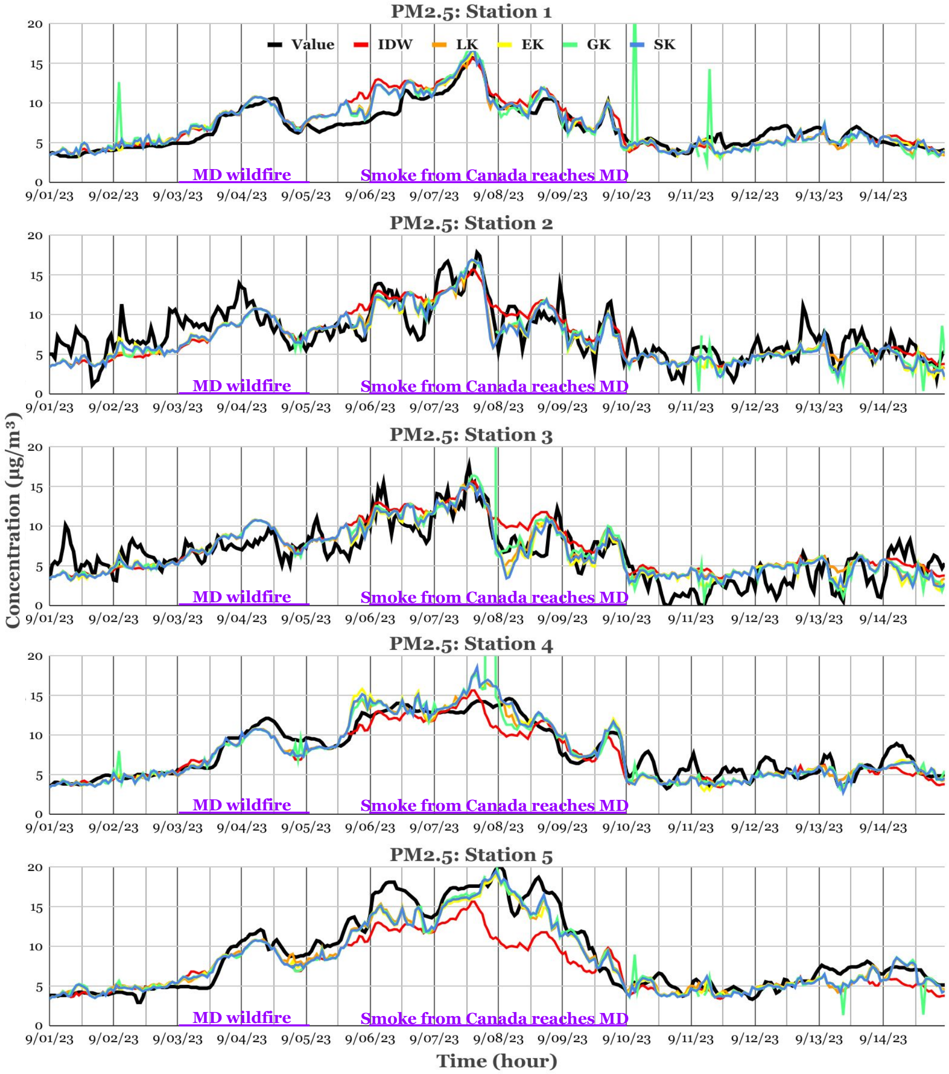
The dataset had 29 $PM_{2.5}$ monitors, 5 PM_{10} monitors, and 39 Ozone monitors. Since $PM_{2.5}$ and Ozone had many monitor locations, 5 monitors were used as the sample for $PM_{2.5}$ and Ozone. There were only 5 PM_{10} monitors which is not enough for a strong spatial correlation (<0.6), so only the central PM_{10} was used as the sample to ensure a reasonable spatial correlation.



Results - $PM_{2.5}$

All the interpolations followed the actual $PM_{2.5}$ trend closely during the two wildfire events except for Gaussian which randomly jumped to a high number and sometimes went into the negatives (removed since it interpolated values beyond limits).

For Station 5, IDW's trend was much lower than the other interpolation method which could be due to a lack of nearby monitors.



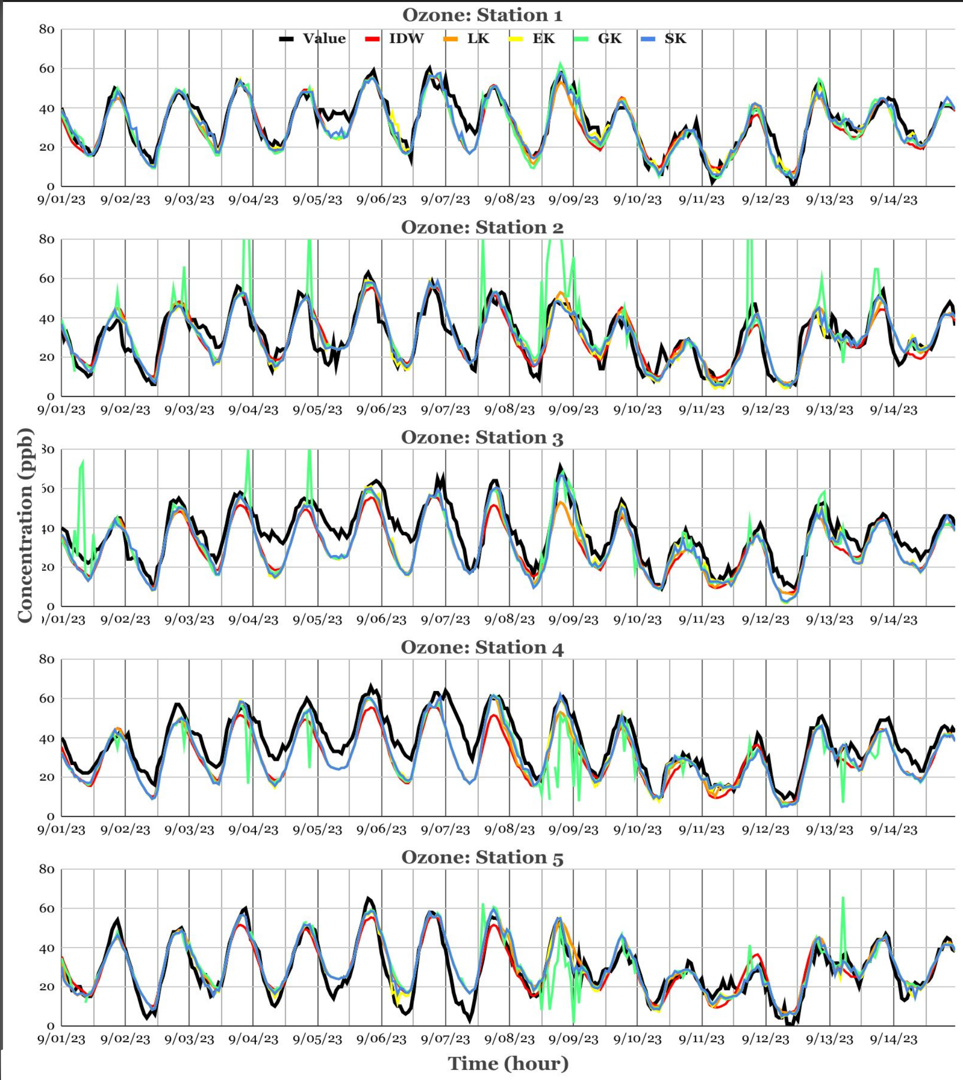
Results - $PM_{2.5}$

PM2.5 Station	IDW		Linear		Exponential		Gaussian		Spherical	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
1	1.282	0.770	1.096	0.832	1.061	0.843	1.703	0.591	1.054	0.845
2	2.012	0.601	1.927	0.634	1.957	0.622	1.975	0.616	1.934	0.631
3	2.158	0.581	2.036	0.627	2.037	0.627	25.890	N/A	2.008	0.637
4	1.388	0.829	1.260	0.859	1.308	0.848	28.068	N/A	1.274	0.856
5	2.749	0.674	1.477	0.906	1.537	0.898	1.597	0.887	1.504	0.902
Average	**1.918**	**0.691**	**1.559**	**0.772**	**1.580**	**0.768**	**11.847**	N/A	**1.555**	**0.774**

For all of the Kriging interpolation methods except for Gaussian, the RMSE and R^2 were very similar for each of the stations. The Kriging methods did better than IDW with Spherical Kriging being the best interpolation method with a RMSE of 1.5548 and an R^2 of 0.7744.

Results - Ozone

Ozone follows cosine wave pattern or that recurs every 24 hours which is known as a diurnal cycle. Gaussian Kriging (green line) also seems to randomly dip or spike for Ozone.



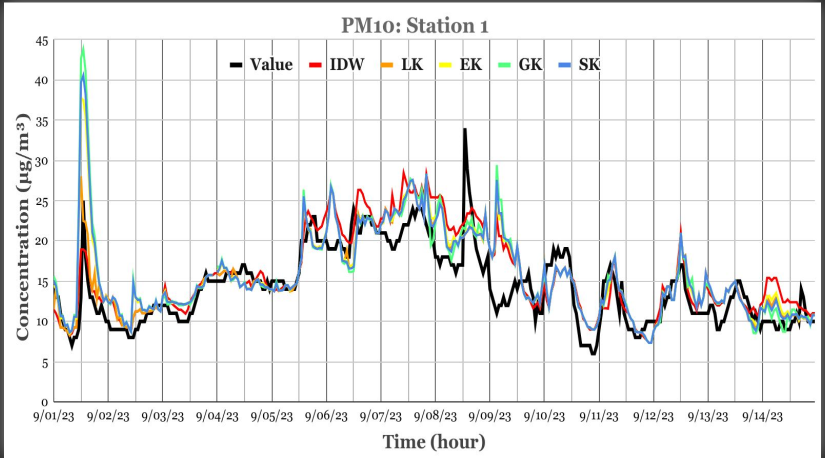
Results - Ozone

Ozone Station	IDW		Linear		Exponential		Gaussian		Spherical	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
1	4.931	0.851	4.647	0.867	4.212	0.891	4.636	0.868	4.045	0.900
2	5.818	0.811	5.487	0.832	5.302	0.843	13.343	N/A	5.281	0.844
3	8.235	0.622	7.883	0.653	7.368	0.697	8.793	0.569	7.385	0.696
4	8.440	0.610	7.702	0.675	7.480	0.693	9.522	0.505	7.415	0.699
5	5.318	0.854	5.059	0.868	4.617	0.890	6.606	0.772	4.776	0.883
Average	**6.548**	**0.749**	**6.156**	**0.779**	**5.796**	**0.803**	**8.580**	**0.678**	**5.781**	**0.804**

The recorded values for all the interpolation methods showed a strong average correlation with values between 0.88–0.75 range and gradually decreased in that order. The optimal interpolation for Ozone was Spherical with an RMSE of 5.7805 and R^2 of 0.8041.

Results - PM_{10}

There was a limited number of stations (only 5) which was not enough for a strong spatial correlation. More linear interpolation methods produced better results, with IDW and Linear having the highest correlation.



PM10 Station	IDW		Linear		Exponential		Gaussian		Spherical	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
1	3.077	0.563	3.160	0.541	3.726	0.362	4.217	0.183	3.926	0.292

Conclusion

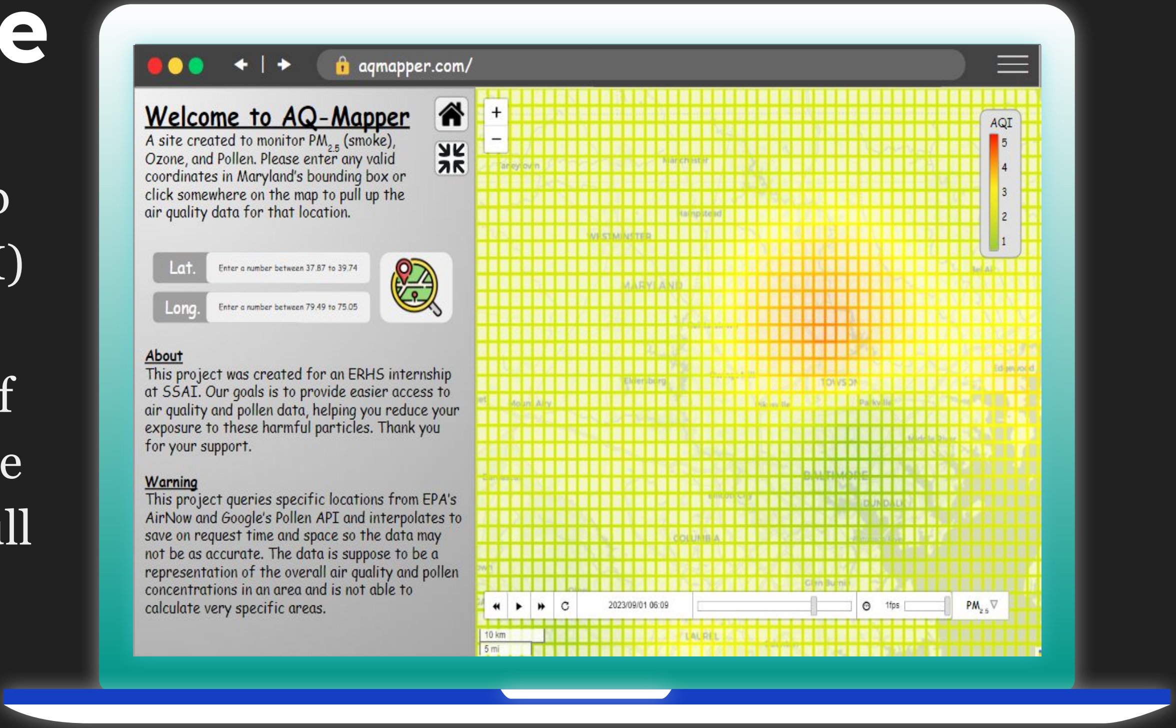
Based on the findings, Spherical Kriging was the best method for air pollutants with more locations/data ($PM_{2.5}$ and Ozone) with an average R^2 of 0.774 and 0.804, respectively. IDW was the best method for air pollutants with a few locations (PM_{10}) with an R^2 of 0.563, closely followed by Linear Kriging with an R^2 of 0.541. Gaussian Kriging was the worst for every air pollutant since the interpolated values spiked randomly and did not follow a predictable trend, making it harder to correct. Some possible sources of error could be incorrect entries in the AirNow data or an incorrectly calculated variogram. Other sources of error could be region-specific factors that could affect the concentrations of the air pollutant like wind speed and terrain which the interpolation methods could not account for.

Future Research

- Creating a website intended to raise public awareness about the significance of air quality and pollen levels.
- Creating the variogram for the Kriging calculation instead of relying on the automatic calculation from PyKrige to make the interpolation more adjustable.
- Testing other interpolation methods or incorporating machine learning to improve the accuracy of heatmap.
- Increasing the number of data sources/monitors.

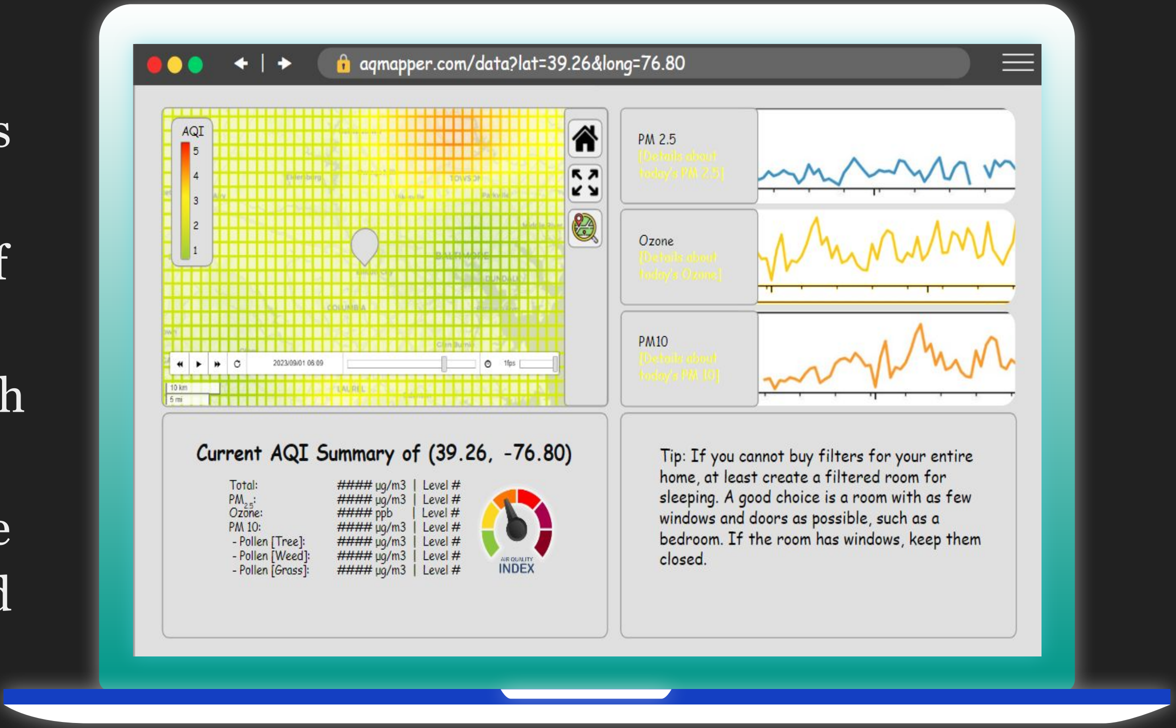
Landing Page

The landing page provided users with a general heatmap of the Air Quality Index (AQI) which was centered at the center of the bounding box of Maryland. The squares on the heatmap can be clicked to pull up the AQI and air pollutant concentration of that area during that time.



Data Page

The data page provided users with detailed information on the current concentrations of various air pollutants values for the selected location. Each column of the AQI table can be clicked to pull up what the level of air quality means and how to reduce exposure to that specific air pollutant.



Acknowledgements

I would like to give special thanks to my mentors, Ms. Jackie Kendall, Dr. Makhan Virdi, and Mr. Perry Oddo from Science Systems and Applications Inc., for providing their valuable guidance. I would also like to thank the Systems Engineer, Mr. James Davis, for the virtual environment to host the website. I would also like to thank my peers, Charlie Renze and Xavier Francois, for their insight throughout the project. I also like to thank Dr. Yau-Jong Twu for her guidance throughout my RP project. Finally, I would like to thank my family for their unwavering support throughout my senior year.

Credits

- Presentation template by SlidesCarnival
- Icons created by Freepik - Flaticon
  - Warning [Title slide] from the “Extreme Weather” pack
  - Ozone [Slide 4] from the “Gas” pack
- Inverse Distance Weighted Example Model from GIS Geography