Machine Learning Techniques for Predicting Ozone Concentration using Surface Meteorological and Vegetative Conditions

Austin Dixon[1], Ann Podleski[1], Nichole Gosselin[2]

[1]Harris-Stowe State University, [2]Saint Louis University

Background

Despite regulations placed on pollutants, tropospheric ozone levels are projected to increase in polluted regions with a warming climate [1]. Although ozone projections vary regionally and seasonally, many studies show ozone concentration increases in peak pollution season along with an increasing peak season length [1]. Concentrations are typically higher in rural areas downwind of major pollution centers, often the locations of agricultural production [2] [3].

Ozone is a secondary pollutant formed through a series of reactions involving ultraviolet radiation from the sun and multiple primary pollutants such as methane and nitrogen oxides and volatile organic compounds. These ingredients are often the result of automobile and factory pollution, so tend to occur more in crowded, urban areas. Additionally, the reaction occurs more rapidly under clear sky, allowing for the necessary ultraviolet light and under warmer temperatures which increase the reaction rate. Once ozone is formed, it can be carried downwind, dependent on wind speed and direction. Additionally, ozone levels are related to relative humidity in the troposphere because of plant's stomatal response to vapor pressure deficit (VPD). In areas where there are many plants, ozone will enter the plants through their stomata, particularly on days with lower VPD. Hence, tropospheric ozone levels are dependent on multiple factors including many meteorological conditions, amount of plants nearby, as well as the availability of primary pollutants.

Ground ozone monitoring stations distributed across the globe provide critical ground truth for not just validation of satellite-based ozone products but also to understand the local effects of ozone on our environment. However, there is often a data gap in the ground or satellite-based observations of background ozone due to instrument failure or satellite data availability. Hence, methods that reliably fill this data gap are important for better understanding ambient ozone concentrations and its impacts on human and plant health.

We hypothesized that we can fill in gaps in tropospheric ozone data using ozone data from nearby sites along with available meteorological data.

GO3 Project

The GO3 Project allows for citizen science tropospheric ozone investigations. Students and teachers in schools, universities, and public education centers across the globe have installed ozone monitors alongside weather stations. Data is uploaded every 15-minutes to a website (http://go3project.com/network2/ozone-data.html) allowing for sharing and analysis of ozone data from multiple locations [6]. Details on the equipment and project can found online at https://twobtech.com/education.html.

This study uses data from three GO3 station locations: The Saint Louis Science Center, Southwestern Illinois College, and Grant's Farm. These three sites are accompanied by ozone gardens, educational exhibits to visualize the effects of ozone on plants. A map of the locations along with pictures from each site are shown in the accompanying figures. Because a large portion of data was missing in 2015, the other two sites were used to fill the data gap.

Three St. Louis Area sites. Location of Ozone Gardens with Ozone and Weather data collection (GO3)

Planting the SWIC Ozone Garden, June 2013

St. Louis Science Center (Planetarium) Ozone Garden (2013)

Ozone Garden Grant's Farm, July 2013

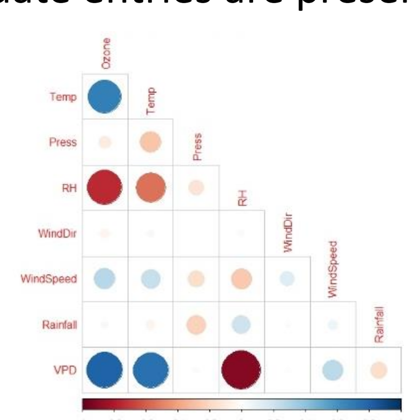
Methods

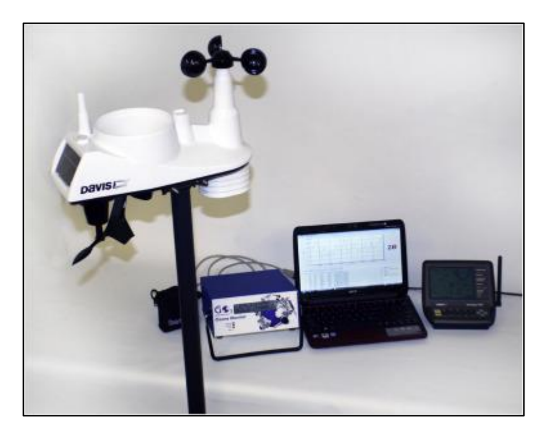
Variables: The weather station measures temperature, pressure, relative humidity, wind speed, wind direction, and precipitation. We also calculated Vapor Pressure Deficit (VPD), the difference between saturation vapor pressure and actual vapor pressure. VPD can also be important predictor of midday Ozone levels, especially during summer [7].

We used a supervised machine learning algorithm to perform a Linear Regression model to predict Ozone at the Planetarium based on various climate measurements and Ozone levels at Grant's Farm or Southwestern Illinois College (SWIC). The machine learning algorithm was applied to the dates where all three sites had data. Python package Sklearn [8] was used to implement the algorithm. We specified a training set of 70% of the data. We did 10,000 iterations and kept the model with the largest Coefficient of Determination (R^2). This was done separately for Grants Farm data and for SWIC.

Grant's Farm data was used if it was available for the missing dates. If that date was also missing from Grant's Farm, data from SWIC was used to predict Ozone at the Planetarium site. There is a gap in Ozone and weather data at Planetarium for 2 months in 2015 (July and August), and so we tested our imputing plan on 2014 data where the missing date entries are present.

Correlations Between climate variables





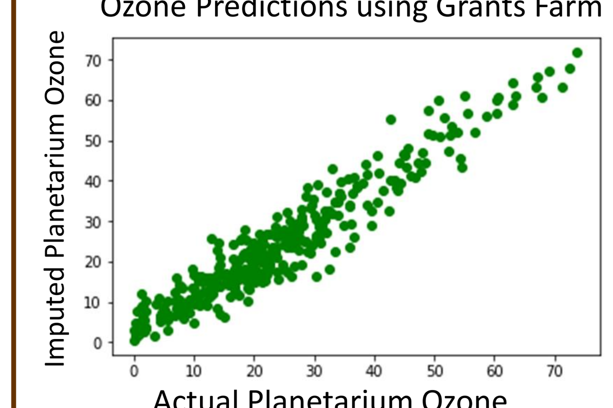
The GO3 Project Package with Davis Vantage Vue weather station, 2B Technologies Model 106-L ozone monitor

Results

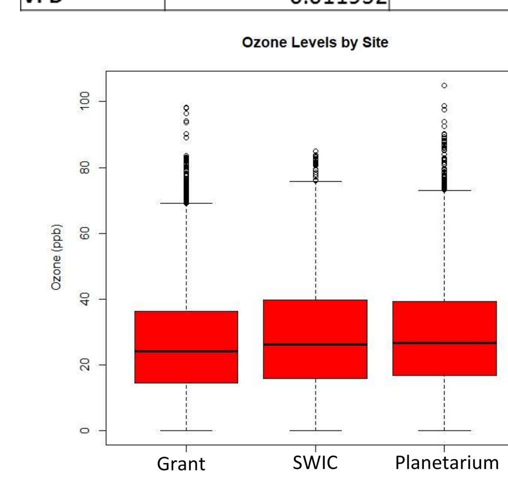
The mean difference between Actual and Imputed Ozone levels was 2.0 ppb. Although the range of difference is between -25.6 and 38.7, fifty percent of the differences lie between -1.6 ppb to 5.2 ppb. (Actual Ozone levels range between 0 and 100 in this dataset.)

We also looked at the actual and imputed by date.

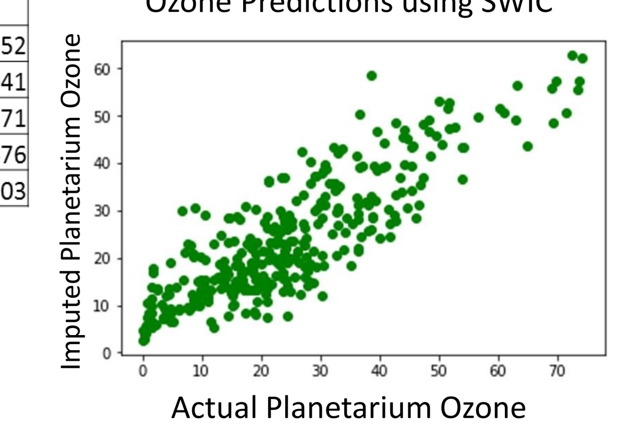
Ozone Predictions using Grants Farm

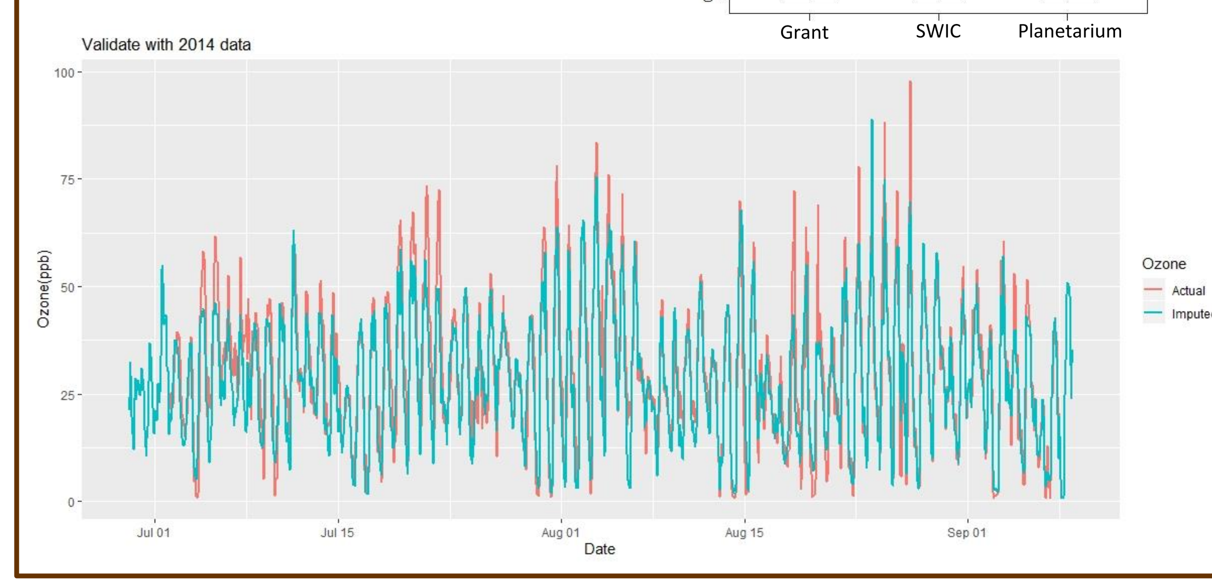


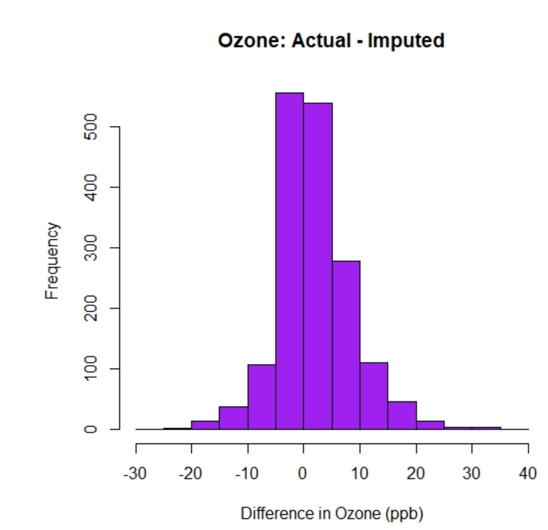
	Grants Farm (R^2 =0.92)	SWIC (R^2 = 0.73)
Variable	Estimate	Estimate
Constant	-5.001896	10.69852
Ozone	0.863258	0.578041
Temperature	0.10026	0.177171
RH	0.00867	-0.165476
VPD	0.011952	-0.13503



Ozone Predictions using SWIC







References

1. Kirtman, B., et al., Near-term climate change: projections and predictability. 2013.
2. Fowler, D., et al., Ground-level ozone in the 21st century: future trends, impacts and policy implications. Royal Society Science Policy Report, 2008. 15(08).
3. Xu, J., et al., Measurements of ozone and its precursors in Beijing during summertime: impact of urban plumes on ozone pollution in downwind rural areas. Atmospheric Chemistry and Physics, 2011. 11(23): p. 12241-12252.
4. Finlayson-Pitts, B. J., and J. N. Pitts Jr. "Atmospheric chemistry of tropospheric ozone formation: scientific and regulatory implications." *Air & Waste* 43.8 (1993): 1091-1100.
5. Kavassalis, Sarah C., and Jennifer G. Murphy. "Understanding ozone-meteorology correlations: A role for dry deposition." *Geophysical Research Letters* 44.6 (2017): 2922-2931.
6. Ellenburg, Jessa A., et al. "Global Ozone (GO3) project and AQTreks: Use of evolving technologies by students and citizen scientists to monitor air pollutants." *Atmospheric Environment: X* (2019): 100048.
7. Kavassalis, S.C.,and J.G. Murphy, "Understanding ozone-meteorology correlations; A role for dry deposition, Geophys. Res Lett., 44, (2017) 2922 – 2931
8. Fabian Pedregosa,et.al, "Scikit-learn: Machine Learning in Python", *Journal of Machine Learning Research*, 12, 2825-2830 (2011)

Discussion and Future Research

We hope to extend this research through use of deep neural networks that use meteorological conditions, proximity to urban locations, available ozone data, and satellite derived leaf area index to improve ozone predictions in areas where data is unavailable.

More complete records of ozone pollution will help improve calculations of trends in both "polluted" and background concentrations of ambient ozone, something we would also like to evaluate.

Acknowledgements

Minority Science and Engineering Improvement Program (MSEIP) P120A160064 and Dr. Hence, Bernadette at U.S. Department of Education. This work was also supported by Harris-Stowe State University and Saint Louis University. Special thanks to Dr. Vasit Sagan, Associate Professor of Earth and Atmospheric Sciences, Saint Louis University.