**Using Supervised Machine Learning Algorithms**

**for**

**Accurate and Efficient Ground-level Ozone Predictions**

Grade 11

Delaware Valley Science Fair

Dover, Delaware

April 4, 2018

**Abstract**

Ground-level ozone is a hazardous pollutant that silently kills humans, animals, and plants. This type of ozone is formed when oxides of nitrogen (NOx) and volatile organic compounds (VOCs) chemically react in the presence of sunlight and high temperatures. Aside from the reactants of ozone, several other factors such as motor vehicles emissions, fossil fuels, and particulate matter are correlates of ozone.

The US has a network of sensors to monitor and predict ozone levels. These forecasts, however, are chiefly based on current ozone levels. If these forecasts based their models on the sources of ozone such as traffic (motor vehicles), their estimations would be more accurate, as ozone does not maintain a strong correlation with time. Furthermore, the current predictions are far too complex to be implemented in local areas, since they are intended for large regions.

This project attempts to use past data to develop a correlation between ozone levels and the causes of ozone. The expectation is that supervised machine learning algorithms can be developed to accurately and efficiently model ozone levels against the causes of ozone in three different locations and scales.

Following the outlined procedure, a supervised machine learning algorithm was successfully developed. Using this algorithm, ozone levels can now be predicted with higher accuracy than before, and can be predicted efficiently in local areas with no ozone monitors. Lastly, this algorithm could also be used to find accurate correlations with other pollutants, in order to create predictions for them as well.

**Table of Contents**

Abstract . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . i

Key Terms . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1

Purpose . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2

Research Question . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2

Hypothesis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2

Materials . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

Variables . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

Procedures . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

Tables and Figures . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5

Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

Practical Applications . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

Future Research . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

Bibliography . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10

**Key Terms**

Ground-level Ozone: Tropospheric ozone that is not directly emitted into the air, rather it is created by chemical reactions between NOx and VOC.

Supervised Machine Learning: A type of machine learning that pairs an input with a known output, and trains the relationship between the variables so that future outputs may be predicted with just the inputs.

Regression Analysis: A set of statistical processes for estimating the relationships among variables.

Artificial Neural Network: An interconnected group of nodes, similar to the vast network of neurons in a brain.

Python: An interpreted high-level programming language for general-purpose programming.

TensorFlow: An open source Python library for numerical computation using data flow graphs. Nodes in the graph represent mathematical operations, while the graph edges represent the multidimensional arrays (tensors) communicated between them.

Volatile Organic Compounds (VOC): Organic compounds (chemicals that contain carbon and are found in all living things) that easily become vapors or gases.

Oxides of Nitrogen (NOx): A group of gases that are composed of nitrogen and oxygen, mainly nitric oxide (NO) and nitrogen dioxide (NO2).

**Purpose**

This project attempts to develop a predictive model of ozone levels and the causes of ozone. The expectation is that supervised machine learning algorithms can be developed to accurately and efficiently model ozone levels against the causes of ozone in various locations and scales.

**Research Question**

In various locations and scales, can self-correcting supervised machine learning algorithms be developed to accurately and efficiently model ozone levels against the causes of ozone?

**Hypothesis**

Self-correcting supervised machine learning algorithms can be developed to accurately and efficiently model ozone levels against the causes of ozone in various locations and scales. In other words, regression analysis will be used in supervised machine learning algorithms to create a basis for ozone prediction in different areas.

**Materials**

* Computer with PyCharm Professional 2017.3
* Python 3.6.3 interpreter
* TensorFlow 1.5.0 library for Python
* Pandas (Python Data Analysis) library for Python
* Swing library for Java
* Microsoft Excel
* Datasets from a variety of locations

**Variables**

Independent Variables: Correlates and Causes of Ground-level Ozone (NO, VOC, RH, Temp, PM2.5, etc.)

Dependent Variable: Ground-level Ozone

**Procedures**

1. Research and collect data on correlates of ozone.
2. Use correlates of ozone to generate several datasets that can be used for ozone regression analysis, in order to create a basis for predicting ozone levels.
3. Develop a self-correcting Python program which uses supervised machine learning algorithms for the regression analysis of ozone with the TensorFlow library (See Developmental Procedure).
4. Accept or reject the hypothesis.
5. If hypothesis is accepted, test the algorithm against the previously generated ozone datasets, and evaluate its accuracy.
6. Develop a Java program using the Swing library, in order to simplify use for everyone (See Developmental Procedure, item 5).
7. If hypothesis is rejected, propose other possible alternatives.

**Tables and Figures**

**Table 1**: Sample Ozone data modeled against its correlates from New Castle County on 1/1/2017

|  |
| --- |
| **Figure 1**: Sample Ozone data modeled against its correlates from New Castle County on 1/1/2017  **Figure 2**: Actual Ozone data modeled against its predicted values from New Castle County on 1/1/2017 |

**Conclusions**

The hypothesis of this study was as follows:

*Self-correcting supervised machine learning algorithms can be developed to accurately and efficiently model ozone levels against the causes of ozone in various locations and scales.*

This project attempts to develop a predictive model of ozone levels and the causes of ozone. The hypothesis that self-correcting supervised machine learning algorithms can be developed to accurately and efficiently model ozone levels against the causes of ozone in various locations and scales was accepted. The calculated loss was minimal for the large datasets, and thus the algorithms were deemed to be accurate. Hence, these algorithms enable one to predict ozone levels with higher accuracy than before, and with higher efficiency in local areas with few to no ozone monitors.

**Practical Applications**

The findings of this study can revolutionize the accuracy and efficiency of predicting hazardous ground-level ozone data, by developing a basis for ozone prediction in a variety of locations and scales. Therefore, the practical applications achieve the primary goals of the purpose of this project. Furthermore, the supervised machine learning algorithm in an artificial neural network can also be applied to develop models of other hazardous pollutants.

**Future Research**

In the future, unsupervised machine learning in artificial neural networks could be investigated to develop more accurate and efficient self-reliant algorithms for ozone analysis, or even for other data. This could reduce human error, as well as improve data selection, filtering, and correction.

**Bibliography**

# *Air Quality Monitoring Network | Delaware Open Data Portal.* (n.d.). Retrieved from [https://data.delaware.gov/Energy-and- Environment/Air-Quality- Monitoring-Network/b6hy-ss9](https://data.delaware.gov/Energy-and-%20Environment/Air-Quality-%20Monitoring-Network/b6hy-ss9)q

# *The Ozone Problem | Ground-level Ozone | New England | US EPA.* (2017, April 10). Retrieved from <https://www3.epa.gov/region1/airquality/oz_prob.html>

# *Air Emissions Sources.* (2017, June 02). Retrieved from [https://www.epa.gov/air-emissions-inventories/air- emissions-sources](https://www.epa.gov/air-emissions-inventories/air-%20emissions-sources)

# Lopez, J. (2017, August 10). *Are Lower PM2.5 Levels Causing an Increase in Hazardous Ozone Pollution?* Retrieved from [http://www.thatsmags.com/beijing/post/20176/lower-levels- of-pm2- 5-might- be-causing-a-rise- in-harmful- ozone-in- Beijing](http://www.thatsmags.com/beijing/post/20176/lower-levels-%20of-pm2-%205-might-%20be-causing-a-rise-%20in-harmful-%20ozone-in-%20Beijing)

# Pfister, G. G., Wiedinmyer, C., &amp; Emmons, L. K. (2008, October 9). *Wildfires Cause Ozone Pollution to Violate Health Standards, New Study Shows.* Retrieved from [https://www2.ucar.edu/atmosnews/news/916/wildfires-cause- ozone-pollution- violate-health-standards- new-study- shows](https://www2.ucar.edu/atmosnews/news/916/wildfires-cause-%20ozone-pollution-%20violate-health-standards-%20new-study-%20shows)