

Can We Reduce Air Pollution Using Sensor Data?

Prithvi Yarlagadda
Data Science Institute

pyarlagadda@saintpeters.edu — 1 (201) 936 4393



Abstract

Pollution is when natural resources like Air, Water, Land and other parts of environments starts to become unsuitable or unsafe to use. Air pollution is one of the most dangerous pollutions humans are facing right now. Air pollution may result in various harmful diseases and mortality rate will also fall due to this. There are many factors which results in air pollution, humans are also one of the factors for causing air pollution and important factor. Mostly air pollution is caused due to motor vehicles and many recent electronic devices. There are many ways to calculate the air pollution, sensors can be used to understand the percentage of different parameters which cause the air pollution.

Introduction

The dataset[1] is from Citypulse website, where it provides with the air pollution dataset of Brasov in Romania. There are 449 different locations where data is collected. Dataset consists of 449 different csv files, each file for each location. Each csv file have parameters like location (latitude and longitudes), timestamp (it is recorded for every 5 minutes), ozone level, particulate matter level, carbon monoxide level, sulphur dioxide and nitrogen dioxide. Dataset is from the dates August 1 2014 to September 30 2014, 2 months of data which is recorded every 5 minutes. Using this dataset we can find some pattern in the air pollution and help in bringing that down. There are some ideas or implementations which we follow through in this paper for better understanding and conclude with some insight.

From the dataset we can find the locations and their respective gas emissions. From these factors we can find some patterns like correlation between time and pollution. We can also perform time-series calculations[5] and can predict the pollution on that particular location at this particular interval of time. From this analysis we can mark the highly polluted areas and may avoid frequent visits in these related areas. These can be calculated by the dataset we have, there are 449 locations and each location have one separate file. Each file has around 18000 records with location and emission values. So each location can be taken at a time and sum all the values. These can be categorized into different time series like calculate for monthly, daily and hourly[4]. From this we can find a pattern and may predict for the next day or hour. These can be very helpful in understanding the pollution distribution and help people in avoiding those areas. We can use Brasov city data and use the location data and find the values of the pollution and show case the highly polluted areas in the city by heatmap[3]. Using these heatmap we can find the highly polluted areas so that public can avoid those in case of travelling. This can also help in taking measures for reducing the pollution in those areas, may be by planting more trees or making vehicles to take alternate route. Some ideas of show casing the polluted areas are by making specific areas green where it is not polluted and red where it is highly polluted[2].

Main Objectives

1. Locate sensors and show on map from dataset[1].
2. Show normal traffic rate in those locations.
3. show how population is distributed among those locations.
4. Use timestamp from the data and show case how it is distributed over time.
5. Time Series plot for all five factors of pollution causing.
6. Estibulum est purus, ultricies cursus volutpat
7. Nullam at mi nisl. Vestibulum est purus, ultricies cursus volutpat sit amet, vestibulum eu.
8. Praesent tortor libero, vulputate quis elementum a, iaculis.

Materials and Methods

Fusce magna risus, molestie ut porttitor in, consectetur sed mi. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque consectetur blandit pellentesque. Sed odio justo, viverra nec porttitor vel, lacinia a nunc. Suspendisse pulvinar euismod arcu, sit amet accumsan enim fermentum quis. In id mauris ut dui feugiat egestas. Vestibulum ac turpis lacinia nisl commodo sagittis eget sit amet sapien.

Results

Pollution is relatively proportional to the population of the area. For example less densely populated areas have less transportation so this result in less gas emissions and which results in less pollution. So correlation between pollution and population is also considered in study. This makes huge difference in understanding and finding the patterns. If we did not consider the population and calculated the pollution, we cannot rely on that results and it is not a good model to consider for further calculations or predicting the pollution. One limitation for this approach is that we need to have access to the population dataset and we need to have location data too for that population. We have the pollution dataset and locations based on sensors. But we do not have sensors for calculating the population.

Even if we had, it would not be same locations as the sensor data. The population and pollution locations will operate on different areas. So this approach cannot be accurate enough, though we can find the city population and can calculate. But this will be for whole and cannot be relied on it.

Tried different regression models to come up with better prediction of the air pollution. In my research and I am going to consider these methods to understand how it is distributed and predicted and I will implement those on our dataset. I also consider the parameters like correlation between area and pollution, population and pollution. Because we cannot say that these part of city have high pollution based on parameters, there could be more vehicles running in these areas at some particular time interval and which results in these kind of increase in air pollution. So population should also be a main factor in deciding and predicting the air pollution.

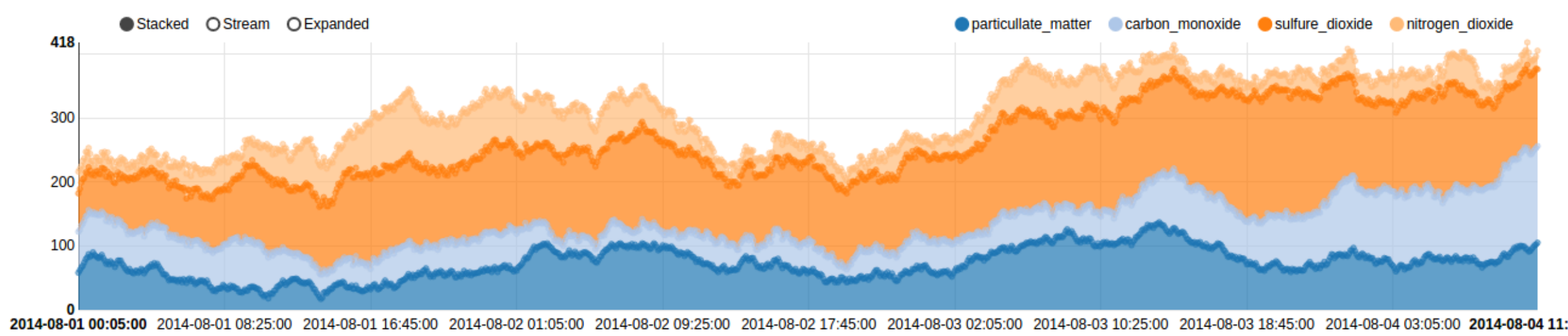


Figure 1: Time series for variables of Air pollution causing factors, ozone level, particulate matter, carbon monoxide, sulphur dioxide. They are distributed similarly across selected time frame

These research also considered different factors varying from simple highways to as complicated as emission of gases with respect to speed. These all factors play major role in air pollution. In our research we consider the location, time and different gases as we have access to this data only and perform different regression models and test which one is best suited and gives more accurate results in predicting the pollution for next hour or next day. This paper will also provide with the areas which are highly polluted. This can be used in various field to make the world better place by implementing some measures reduce the pollution or showing the public what are the dangerous places to live or travel regarding air pollution.

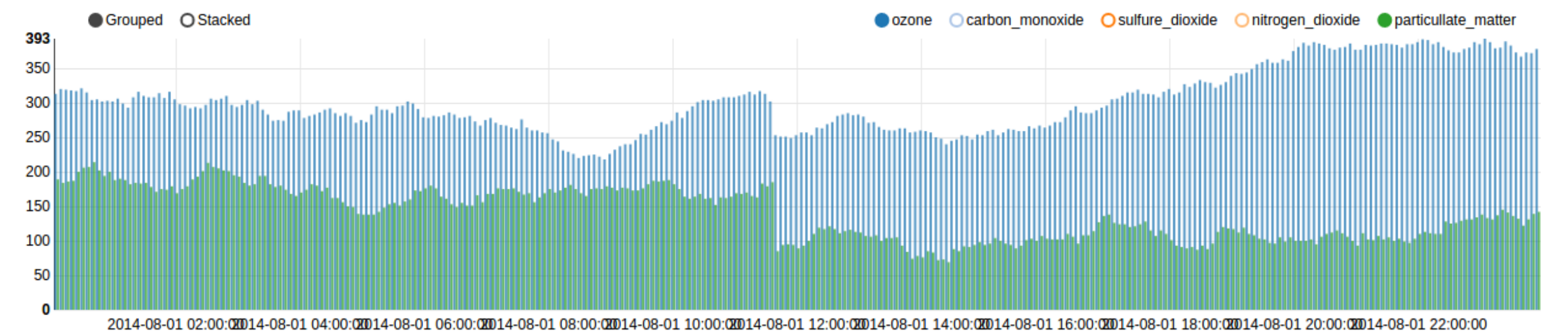


Figure 2: Time series plot showing Ozone Vs Particulate matter. This two variables have most correlation between them. The Ups and Downs on the graph for both of them are similar around afternoon

Conclusions

- Better understanding of how improvement of transport technology and take steps to save environment
- Decrease vehicles pollution by technology like hybrid cars and electric vehicles.
- Install sensors and maintain traffic-related air quality updated.
- Real-time traffic related air pollution can be maintained.

References

[1] city-pulse data. smart city data. <http://iot.ee.surrey.ac.uk:8080/datasets.html/>.

[2] Corneliu Cofaru. Strategies of developing road transport by controlling automotive emissions to reduce local and global environment impacts. In *Proceedings of 6th IASME/WSEAS International Conference on Energy & Environment*, pages 23–25, 2011.

[3] F Costabile and I Allegrini. A new approach to link transport emissions and air quality: An intelligent transport system based on the control of traffic air pollution. *Environmental Modelling & Software*, 23(3):258–267, 2008.

[4] Stelian Tarulescu, Horia Beles, Adrian Soica, Nicolae Chioreanu, and Tudor Mitran. Predict air pollution in brasov city with regression models. In *Applied Mechanics and Materials*, volume 659, pages 617–622. Trans Tech Publ, 2014.

[5] Stelian Tarulescu, Radu Tarulescu, and Adrian Soica. Approximation of the carbon monoxide concentration resulting from the road traffic using experimental measurements. *Annals of DAAAM & Proceedings*, pages 1865–1867, 2009.