**RELEVANCE**

 Most of the air pollution that is responsible for public health emergencies in Delhi every winter is caused by crop burning in neighboring states, scientists at Harvard University say in a recent study.

The study conducted by Harvard John A. Paulson School of Engineering and Applied Sciences (SEAS), Harvard University puts to rest several questions on the extent to which stubble burning in north-western states including Punjab and Haryana contribute to the already severe pollution in Delhi.

Although Delhi witnesses a spike in particulate matter PM2.5 limits round the year, the ambient PM2.5 concentrations show spikes in October-November, which is also the peak season for paddy harvesting, when abundant crop residue is burnt by farmers to prepare for next crop.

Satellite data from Nasa was used to identify hotspots of active fires to model how much of the pollution is coming from crop fires. The data was fed in an algorithm which accounts for geography, wind patterns, and physics to predict how far and in what direction smoke particles travel.

Since the harvesting season coincides with the post-monsoon conditions which favours stagnation of wind in northern India, these conditions allow smoke to slowly permeate throughout the Indo-Gangetic region including Delhi. This smoke mixes with existing pollution from cars and factories creating a thick, deadly haze.

Despite a national ban and regulations by National Green Tribunal Act of 2010, the practice continues as farmers and government remain at loggerheads over a cost-effective alternative.

“A relationship between pollution and mortality is well known," said Mickley, who co-authored the paper. Studies have shown that Delhi suffers from diseases related to air pollution at a rate which is 12 times higher than the national.

“Adverse effects of fire emissions need to continue to be seriously considered as population of Delhi continues to grow, leaving more people at risk," said the researchers, adding that the information would provide policymakers with a quantitative sense of the consequences of current agricultural burning practices in order to inform decision making.

**WORK PLAN**

*WHAT WE HAVE ALREADY DONE*

* Collection OF DATASETS.
* Establishing relationships between datasets.
* Mapping of datasets using python libraries.

*WHAT ARE WE PLANNING TO DO*

**Familiar with tools and equipment used**

* Supervised Learning

*Supervised Learning*

In Supervised learning, you train the machine using data which is well **"labeled**." It means some data is already tagged with the correct answer. It can be compared to learning which takes place in the presence of a supervisor or a teacher.

A supervised learning algorithm learns from labeled training data, helps you to predict outcomes for unforeseen data.

Successfully building, scaling, and deploying [accurate](https://www.datarobot.com/wiki/accuracy) supervised machine learning models takes time and technical expertise from a team of highly skilled data scientists. Moreover, [Data](https://www.datarobot.com/wiki/data-science/) scientist must rebuild [models](https://www.datarobot.com/wiki/model/) to make sure the insights given remains true until its data changes.

In this project we have given our machine some datasets and with the help of which we establish relationships between various attributes of the datasets.

**Data Set used**

* <http://berkeleyearth.lbl.gov/air-quality/local/India/NCT/Delhi>

Berkeley Earth is independently operated, funded primarily by [unrestricted educational grants](http://berkeleyearth.org/funders/)

The Berkeley Earth Surface Temperature Study has created a preliminary merged data set by combining 1.6 billion temperature reports from 16 preexisting data archives. Whenever possible, we have used raw data rather than previously homogenized or edited data. After eliminating duplicate records, the current archive contains over 39,000 unique stations. This is roughly five times the 7,280 stations found in the Global Historical Climatology Network Monthly data set (GHCN-M) that has served as the focus of many climate studies. The GHCN-M is limited by strict requirements for record length, completeness, and the need for nearly complete reference intervals used to define baselines. We have developed new algorithms that reduce the need to impose these requirements (see [methodology](http://berkeleyearth.org/methodology/)), and as such we have intentionally created a more expansive data set.

These data are based on the regional interpolation of real-time observations by ground-level monitoring stations. As the intent is to capture regional variations in air quality, be aware that individual air quality monitors may report somewhat higher or lower values for PM2.5 concentrations than the values represented by the local averages reported here. Since real-time data is used, these time series should be regarded as preliminary and subject to change. Automated quality control processes are used to identify erroneous data, but additional corrections may be necessary. In addition, the number of stations available and their spatial distribution is likely to change over time. Be aware that changes in the station network may have introduced biases in these results.

* https://earthdata.nasa.gov/faq/firms-faq#ed-user-guides

Fire detection is performed using a contextual algorithm that exploits the strong emission of mid-infrared radiation from fires. The MODIS algorithm examines each pixel of the MODIS swath, and ultimately assigns to each one of the following classes: missing data, cloud, water, non-fire, fire, or unknown.

**Research Paper**

* **Agricultural Burning and Air Quality over Northern India: A Synergistic Analysis using NASA’s A-train Satellite Data and Ground Measurements**

In the recent years, New Delhi, the capital city of India, has ranked among the most polluted cities in the world regarding its air quality related to the submicron Particulate Matter (PM2.5). Using NASA’s A-train satellite data (MODIS, OMI, and CALIOP), ground-level PM2.5 measured in New Delhi (2013–2016), and back-trajectory calculations, we show that the PM2.5 over New Delhi is strongly affected by the agricultural fires in the northwestern Indian states of Punjab and Haryana during the post-monsoon season (October and November). The mass concentration of PM2.5 escalates from ~50 µg m–3 measured prior to the onset of residue burning in early October to as high as 300 µg m–3 (24-hour averaged, 7- day running mean) during the peak burning period in early November. A linear regression analysis reveals that the variations in PM2.5 over New Delhi can be attributed to the concurrent changes in the satellite retrievals of fire counts and aerosols over the crop burning area. The back-trajectory analysis shows that most clusters (> 80%) of the northwesterly flow near the ground intercepted the crop burning region before arriving at the receptor location in New Delhi; this further corroborates the transport patterns inferred from the satellite data. A 15-year long satellite record (2002–2016) reveals an increasing trend in agricultural fires (~617 per year) and aerosol loading (0.031 and 0.04 per year in aerosol optical depth and UV aerosol index) in November. Increasing levels of crop residue burning and resulting particulate matter pollution at an alarming rate over northern India is a pressing concern demanding corrective measures to substantially reduce or completely diminish the crop burning through an effective residue management system.

*Feature extraction and training model.*

In this research paper, the dataset used is NASA’s A-train Satellite Retrievals. We use multiple pieces of information on fires and smoke retrieved from NASA’s A-train satellite sensors. The MODIS sensors aboard the Terra and Aqua platforms detect fire spots at 1 × 1 km2 spatial resolution globally on a daily basis. The fire detection is performed using a contextual algorithm (Giglio et al., 2003, 2016) that exploits the strong emission of mid-infrared radiation from fires.

**CONCLUSION**

By this paper we want to showcase the adverse effects burning of crops have on air pollution level not in just that place, but also on the neighboring cities as well as states. New Delhi has long struggled with the worst air quality of any major city in the world, according to World Health Organization data.

**References**

* <https://earthdata.nasa.gov/faq/firms-faq#ed-user-guides>
* <http://berkeleyearth.lbl.gov/air-quality/local/India/NCT/Delhi>