Using Machine Learning to Benefit Oil and Gas Industry in Challenging Meteorological Conditions

Elizabeth Spicer[1] (easpicer@ou.edu), Petra Klein[1], Sean Crowell[2], Wesley T. Honeycutt[1]

[1]: Univ. of Oklahoma, [2]: LumenUs Scientific Solutions

Introduction

The energy industry must adapt to meet increasing demand while reducing environmental impact to appease customers and regulators. Increasing process efficiency can support efforts to meet growing demands, but the mechanisms to do so are not always easy to identify. Even when utilizing best engineering and business practices, costly inefficiencies like fugitive emissions difficult to control. Unaddressed leaks can contribute to failing to meet regulations, potentially incurring fines, and lead to significant product losses. Tools for leak detection and repair (LDAR) are not always perfect.

Big data and machine learning (ML) improve LDAR capabilities. Satellite and ground data are useful in helping detect fugitive emissions. When communicated properly with industry professionals, LDAR data reduces historically hemorrhaged process materials lost to the environment. Many holes exist when combining these toolsets which ML is capable of filling. Two solutions will be discussed in this work:

1. How ML with ground-based sensing can identify possible fugitive emissions in complex industrial and urban environments,
2. How continuous monitoring of pollutants from space will be analyzed by ML.

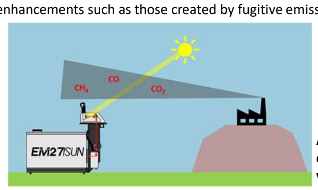
Data Collection in Houston

In the 2022 GeoCarb-TRACER Campaign, as a part of the DOE Atmospheric Radiation Measurements (ARM) facility's TRacking Aerosol Convection interactions ExpeRiment (TRACER), ground-based EM27/SUN spectrometers observed carbon-based pollutants methane (CH_4), carbon dioxide (CO_2), and carbon monoxide (CO) in Houston. The goals of this project:

1. Monitoring carbon-based pollutants and attempt to attribute anomalous concentrations,
2. Contextualize emissions by deploying alongside aerosol, boundary layer, and other instruments,
3. Validate space-based remote sensors.

The GeoCarb EM27/SUN

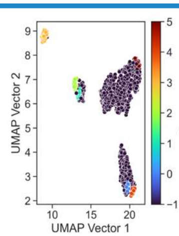
The EM27/SUN operates by following a light source– the sun– throughout the day and recording near-IR radiation. Pollutants like CH_4, CO_2, and CO absorb solar radiation in well-defined frequency ranges. Changes in radiation intensity are due to pollutant concentrations in the column of air observed by an EM27/SUN. Therefore, the instrument sees background concentrations of CH_4, CO_2, and CO as well as local enhancements such as those created by fugitive emissions.



A cartoon of how the EM27/SUN collects data about pollutants by viewing solar radiation.

Machine Learning Methods

1. __Identify when pollutants change rapidly__: For each pollutant measured by the EM27/SUN, data are statistically isolated during the time periods when the pollutant concentration has changed rapidly, likely due to local, anthropogenic sources.
2. __Simplify the data:__ Data are then fed through a dimensional reduction algorithm called UMAP, or Uniform Manifold Approximation and Projection (McInnes et al. 2018).
3. __Cluster to find important details:__ The now three-dimensional data are then clustered using Density-Based Clustering of Applications with Noise (DBSCAN) to identify groups of data that are similar (Ester et al. 1996).



Five clusters identified for high variability CO_2 data

Machine Learning Results

__Isolating Unique Data__

Of the high variability data clusters, the clusters associated with highest and lowest concentrations are isolated. Mathematically, clusters are analyzed if they meet the following criteria:

$$\bar{X}_{cluster} = \bar{X}_{dataset} \pm s_{dataset}$$

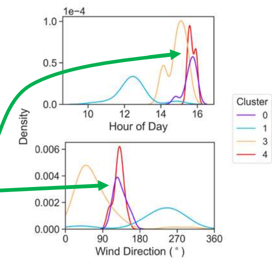
where $\bar{X}_{cluster}$ is the mean concentration of the cluster
$\bar{X}_{dataset}$ is the mean concentration of the high variability dataset
$s_{dataset}$ is the standard deviation of the high variability dataset

For CO_2, two clusters (0 and 4) show interesting results:

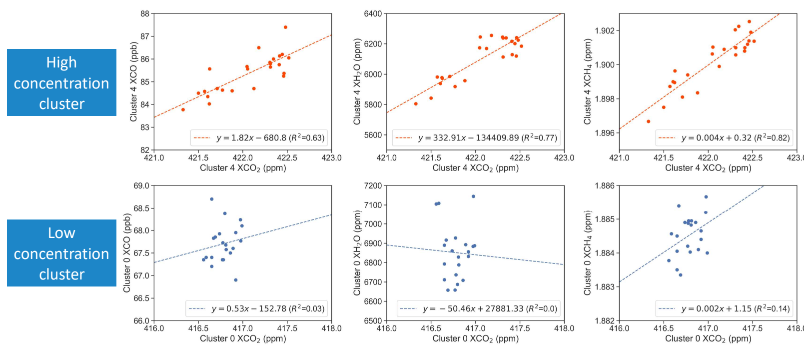
The usefulness of this method is highlighted in clusters 0 and 4. For these clusters,

- The high concentration CO_2 event occurred at the __same time of day.__
- The data were collected at the __same site__.
- The occurred during the __same low-level wind direction__.

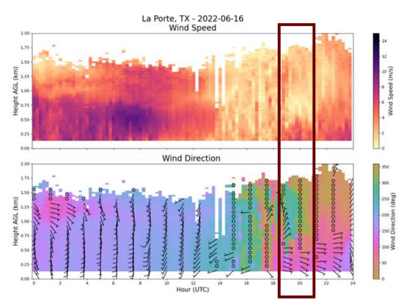
We observed high and low concentrations of CO_2 under similar conditions. Is the high CO_2 due to an anthropogenic emission?



Densities of variables in the high and low concentration CO_2 data clusters



Correlation of CO_2, CO, CH_4, and H_2O for the high and low concentration clusters. CO_2 is positively correlated with CO, CH_4, and H_2O in the high concentration cluster, but has no correlation in the low concentration case.



Sea breeze passage around 19-20 UTC as indicated by the wind speed increase and direction shift. Plot provided by Michelle Spencer from the TRACER-CUBIC grant.

__Investigating Correlations to Determine Potential Pollutant Sources__

Plotting the concentrations of CO, CH_4, and water vapor with respect to CO_2 (above) demonstrates that these concentrations are linearly correlated with a positive regression slope, so larger values of one gas implies larger values of the others as well. For the low concentration cluster, there is no correlation between the variables.

One plausible reason for this correlation is an abnormal process upset involving combustion.

- The area of deployment, La Porte, TX, includes many flared O&G and petrochemical processes
- Flares combust hydrocarbons, producing CO_2 and H_2O.
- However, all flare combustion is rarely complete, producing CO and leaving uncombusted CH_4.

Further meteorological analysis shows that the spike in pollutant concentrations coincides with a sea breeze passage, which could blow pollutants from a release into the instrument's path.

Discussion

This ML method provides an automated approach to isolating interesting data, allowing users to identify trends and outliers in large amounts of data using meteorology principles. This can help fill gaps in existing LDAR technology. While ML principles can be applied to any data, remote sensing instruments like the EM27/SUN provide unique advantages when identifying fugitive emissions.

__Benefits of Using Remote Sensing Data__

- Remote sensors, especially space-based ones, can be deployed for long periods of time, building extensive data records. The GeoCarb instrument, if launched, could provide near-continuous monitoring in the Americas, monumentally increasing the volume of data that can be used for monitoring emissions.
- While instruments like Optical Gas Imaging (OGI) cameras are useful at detecting leaks, they do not provide quantitative information about the release. EM27/SUN provide concentration data that can be used to quantitatively estimate leak size. Therefore, __operators can use these data to prioritize the biggest leaks__.
- Preliminary data suggest that __EM27/SUN data can be used to identify small leaks as well as large ones, not limiting its applications to only major releases__. In a field test at an oil well superpad, EM27/SUN data were used to identify a small leak that was later confirmed with a handheld sensor.

Aerial view of the EM27/SUN deployment that helped identify a tank leak

- __EM27/SUN instruments can collect reliable data in any ambient temperature and wind condition with the right enclosure__. Temperature and strong winds have been shown to impact the accuracy of OGI cameras.

Future Work

More analysis is necessary to solidify the efficacy of this method. Continued deployments of remote sensors are necessary to limit emissions on large scales. A successful launch of the GeoCarb instrument would revolutionize this work.

The probability of detection for EM27/SUN instruments is being investigated in collaboration with a partner company that specializes in LDAR evaluations. The GeoCarb EM27/SUN was deployed during controlled natural gas releases alongside mobile LiCOR 7810, C1-C10 GC, as well as handheld and drone-based FLIR OGI cameras. Publications are in progress resulting from this work.

Left: FLIR images from one controlled release during EM27/SUN deployment
Right: Members of the controlled release project with the GeoCarb EM27/SUN

Acknowledgements and References

The ML work was funded by NASA through the GeoCarb Mission under award 80LARC17C0001. EM27/SUN leak and detection limit work were funded under OCAST IP21.2-016 and OU BIC 1.0 - XGEM. Thank you to the organizers of TRACER and all volunteers for the GC-TRACER campaign.

Please see the QR code on the right for a full list of references.