Investigating Sources of $PM_{2.5}$ in Texas

Archana Dayalu (adayalu@aer.com), Chase Calkins, Jennifer Hegarty, John Henderson, Matthew Alvarado

Verisk Atmospheric and Environmental Research, Lexington MA, USA

I. Introduction

Understanding sources of fine particulate matter ($PM_{2.5}$) is important for Federal Clean Air Act § 179b demonstrations, exceptional event demonstrations, and other high-priority policy objectives.

In this work, we investigated the sources of fine particulate matter ($PM_{2.5}$) pollution in urban areas of Texas. Using air quality site data spanning 2015-2021, obtained from the Texas Commission on Environmental Quality (TCEQ), we:

- ❖ Developed methods to rapidly flag and evaluate sudden increases in hourly $PM_{2.5}$ measurements
- ❖ Used positive matrix factorization (PMF) for a selection of sites to determine relative source contributions to total $PM_{2.5}$
- ❖ Investigated the impact of biomass burning smoke and mineral dust

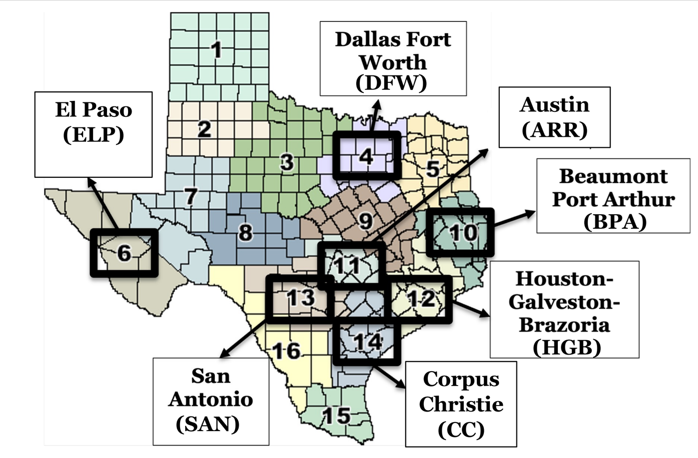

Figure 1. Map of main Texas regions included in this study.

II. $PM_{2.5}$ Anomaly Flagging and Investigation

Isolation Forest (IF) Technique: Used for $PM_{2.5}$ Anomaly Detection

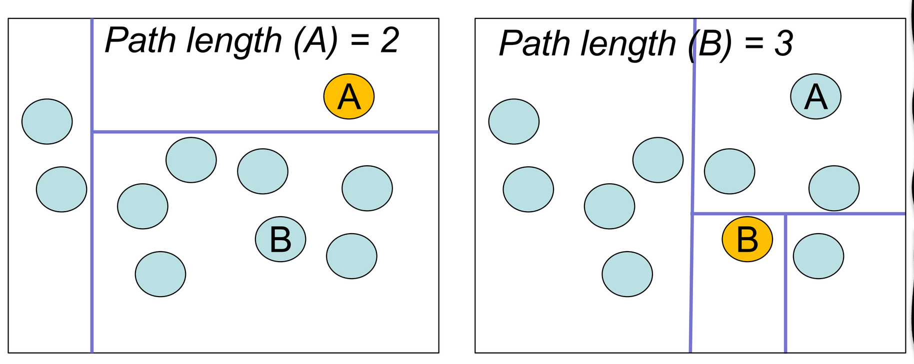

Figure 2. Illustration of IF technique. The number of line segments (blue) to isolate a point is the path length.

- Unsupervised outlier detection algorithm developed by Liu et al. (2008); here applied to rapidly flag sudden increases in hourly $PM_{2.5}$ measurements
- Useful in large datasets that require effective processing with lower compute costs
- The number of partitions required to separate the data point from the larger set is referred to as the path length; the fewer iterations it takes to separate an individual data point, the shorter its path length and the greater likelihood that it is an outlier (anomaly) (Fig. 2)
- We used IF methodology from the 'IsolationForest' module of Python's sklearn toolkit (Pedregosa et al., 2011)

$PM_{2.5}$ Anomalies

- We flag anomalies in $PM_{2.5}$ on days that exceed the 24-hour NAAQS threshold of 35 μg/m³.
- Anomalies that occurred at several monitors were flagged as regional (-2) or multiregional (-3) events, while anomalies at isolated monitors were flagged as local (-1) events (Fig. 3a).

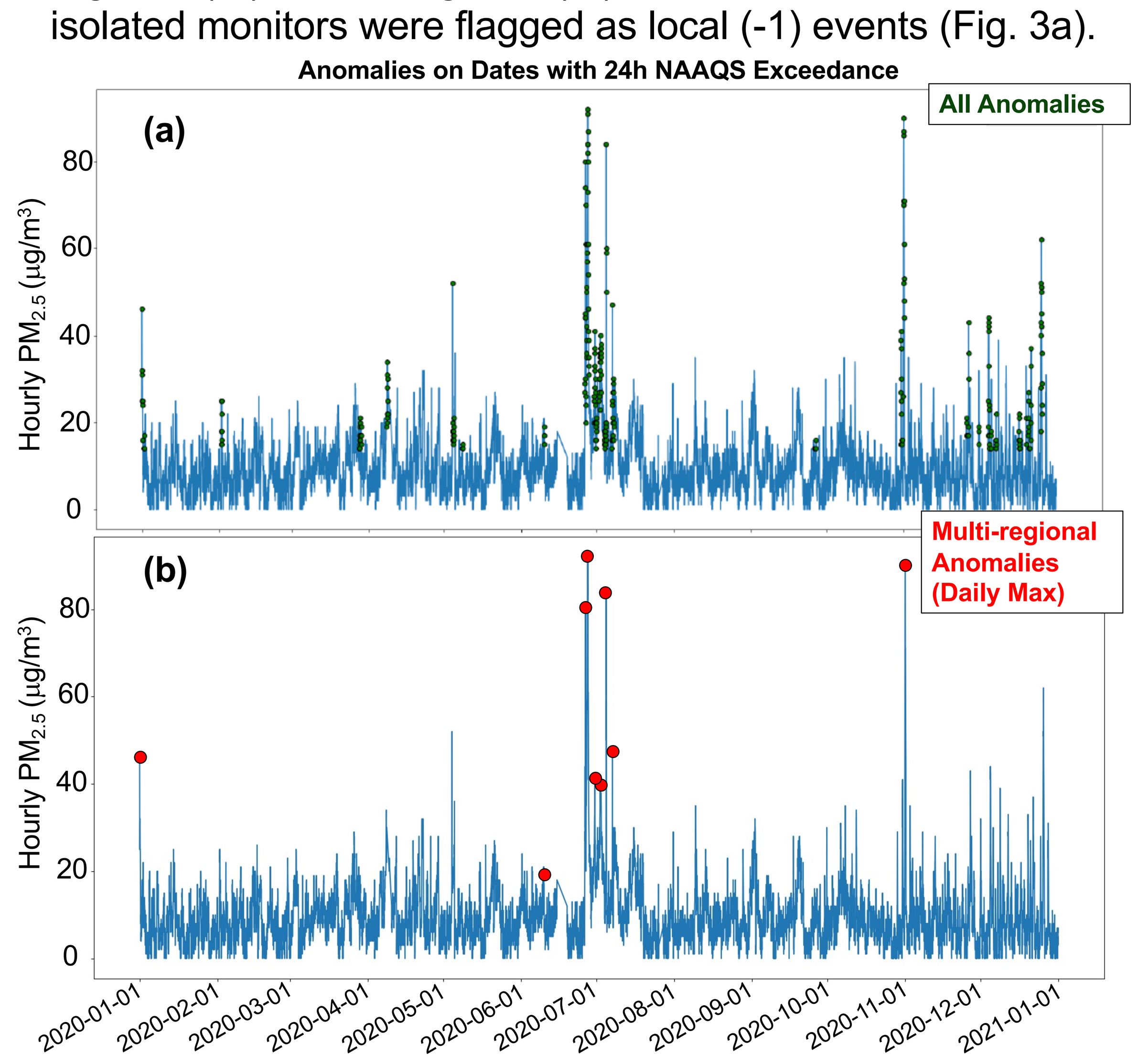

Figure 3. Hourly $PM_{2.5}$ anomaly flagging in Texas Austin Round Rock (ARR) region, (a) all anomalies; (b) multi-regional anomalies only.

- In general, multi-regional and regional anomalies were associated with significantly higher hourly $PM_{2.5}$ concentrations than local anomalies by about 25% and 10% respectively.

Table 1. Summary of hourly $PM_{2.5}$ results by anomaly type aggregated across all regions

Anomaly Type	Mean (μg/m³)	Lower Range (μg/m³)	Upper Range (μg/m³)	SD (μg/m³)
NA/Normal	6.94	0	15.1	3.45
Local (-1)	18.4	13.3	655	7.52
Regional (-2)	20.5	13.3	719	11.6
Multi-Regional (-3)	23.1	13.3	998	13.2

- We further focused our analysis on the maximum concentration recorded for each anomaly type in each region from 2012-2021 (Fig. 3b; Table 2).
- We used HYSPLIT back trajectories to qualitatively examine if the air parcel was likely associated with long-distance transport of dust and biomass smoke, or local sources of $PM_{2.5}$.

Table 2. $PM_{2.5}$ anomaly maxima (μg/m³) for subset of Fig. 1 regions illustrated later in this presentation. HYSPLIT 24-h backward trajectories and 24-h backward dispersions were conducted for each of the points below and merged with high resolution road maps and NEI 2020 point source emissions locations. Smoke and dust maps were also obtained for these data points. See Panel III.

Date	Hour (LST)	Lat	Lon	$PM_{2.5}$ (μg/m³)	Anom Type	AQS Code	Region
20151017	20	29.92	-93.91	355	-1	482450021	BPA
20120414	14	31.75	-106.40	719	-2	481410055	ELP
20190101	0	30.26	-97.72	626	-3	484530326	ARR
20190101	1	27.80	-97.43	254	-3	483550032	CC
20190101	1	29.77	-95.22	558	-3	482011034	HGB

III. $PM_{2.5}$ Source Attribution

BPA
Local Anomaly 2015-10-17
355 μg/m³

ELP
Regional Anomaly 2012-04-14
719 μg/m³

ARR, CC, HGB, (SAN)
Multiregional Anomaly 2019-01-01
254-626 μg/m³

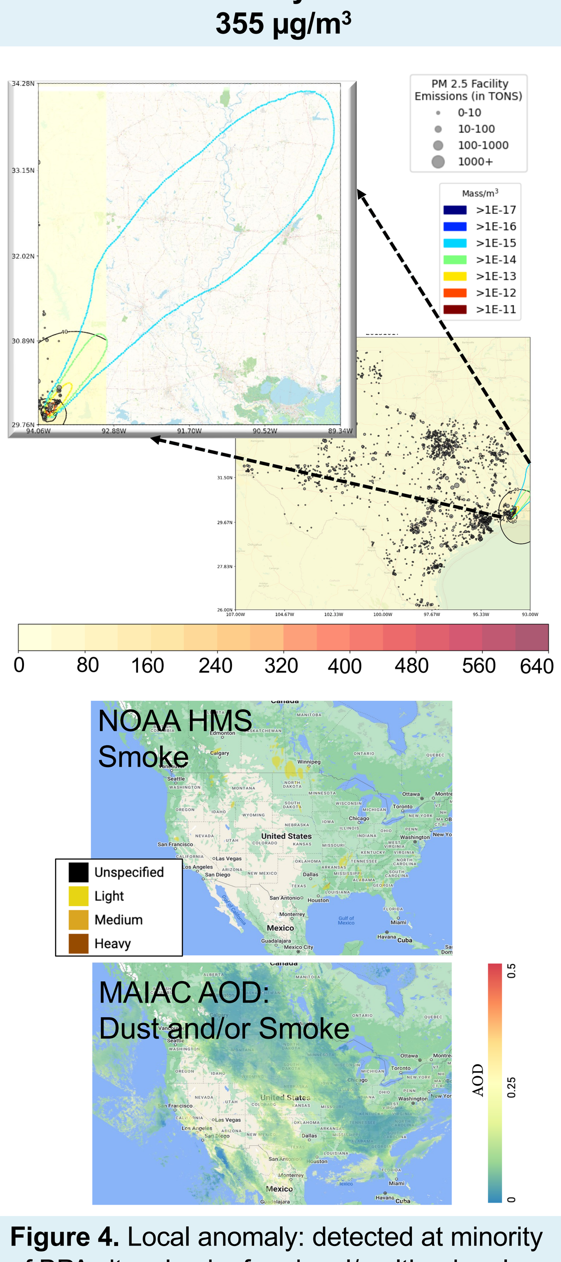

Figure 4. Local anomaly: detected at minority of BPA sites. Lack of regional/multiregional smoke and/or dust sources further suggests local sources dominate (e.g., point sources, roadways).

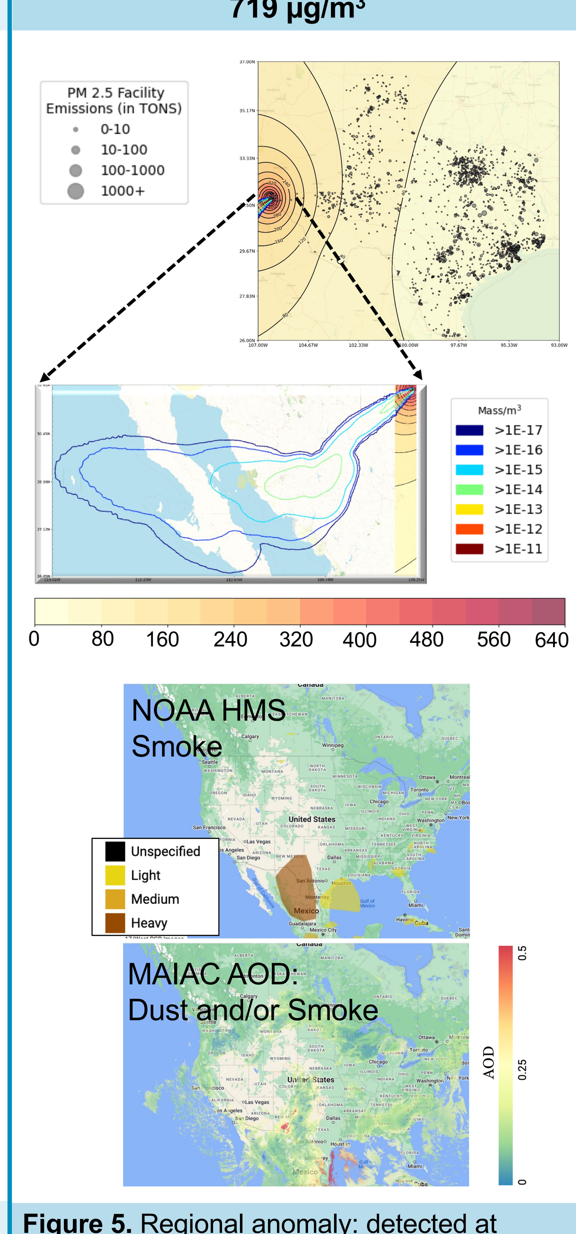

Figure 5. Regional anomaly: detected at majority of ELP sites, but not beyond. Air mass trajectories indicate heavy smoke transported from Mexico; overall suggests international smoke dominates source.

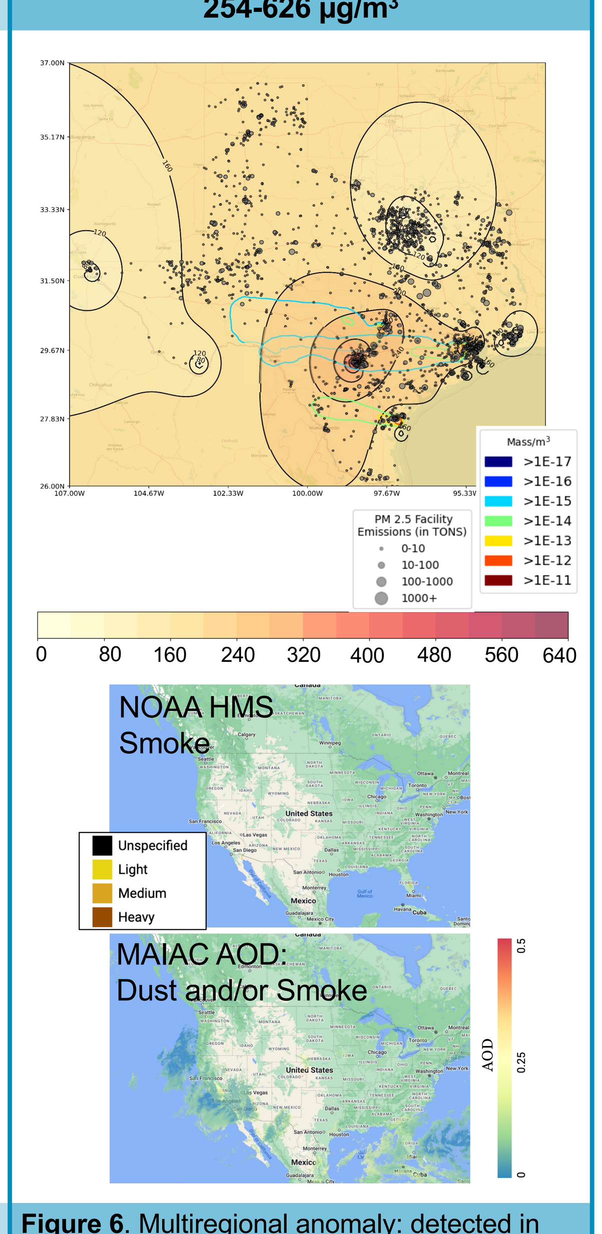

Figure 6. Multiregional anomaly: detected in majority of Texas regions. Air mass trajectories, $PM_{2.5}$ spatial gradient, and absence of smoke/dust all suggest a San Antonio source.

Contributions to Total $PM_{2.5}$: PMF Case Study at two DFW Sites (2018-2021)

- Speciated $PM_{2.5}$ monitoring data provided by the TCEQ from Dallas Hinton (AQS ID 481130069) and Deer Park (AQS ID 481011039) monitoring sites.
- U.S. EPA PMF Model (Version 5.0) used to identify source types and factors contributing to measured $PM_{2.5}$ species; U.S. EPA's SPECIATE repository used to identify source types associated with PMF factors.
- Independent check on both PMF and anomaly categorization technique (Fig. 7).
  - ✓ Regional and multiregional anomalies largely correlated with soil/dust (e.g., summer Saharan Dust episodes) and agricultural burning sources (e.g., Mexico)
  - ✓ Local anomalies broadly associated with higher relative contributions from diesel exhaust

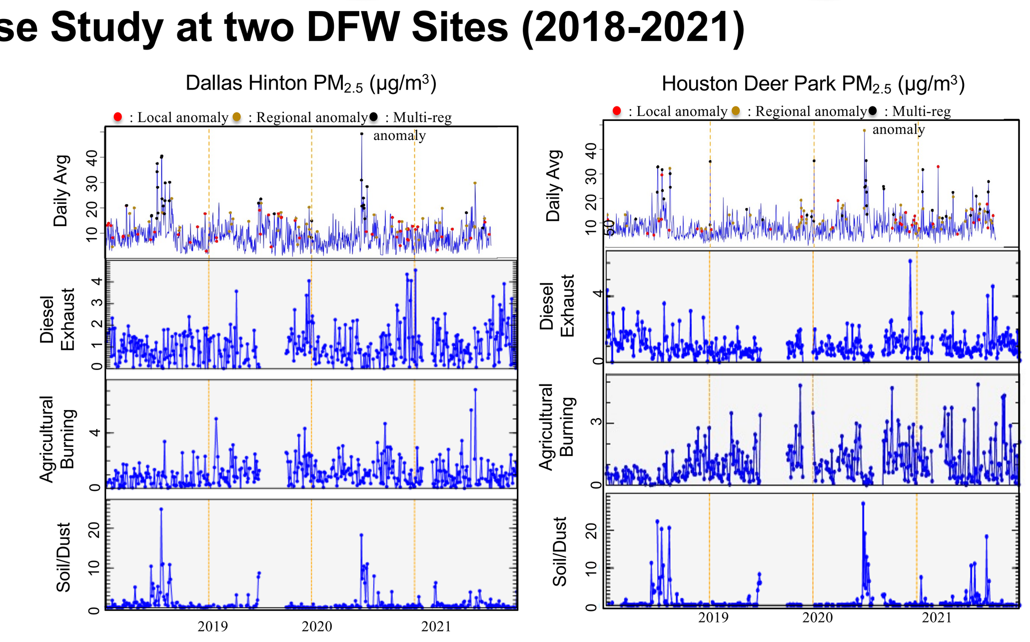

Figure 7. Dallas Hinton (Left) and Houston Deer Park (Right) Site Daily Average $PM_{2.5}$. Top panels: $PM_{2.5}$ flagged with anomaly dates and categories (data through 2021 Sept.). Remaining panels: Normalized $PM_{2.5}$ broken into diesel exhaust, agricultural burning, and soil/dust factors via PMF by Hegarty et al. (2023).

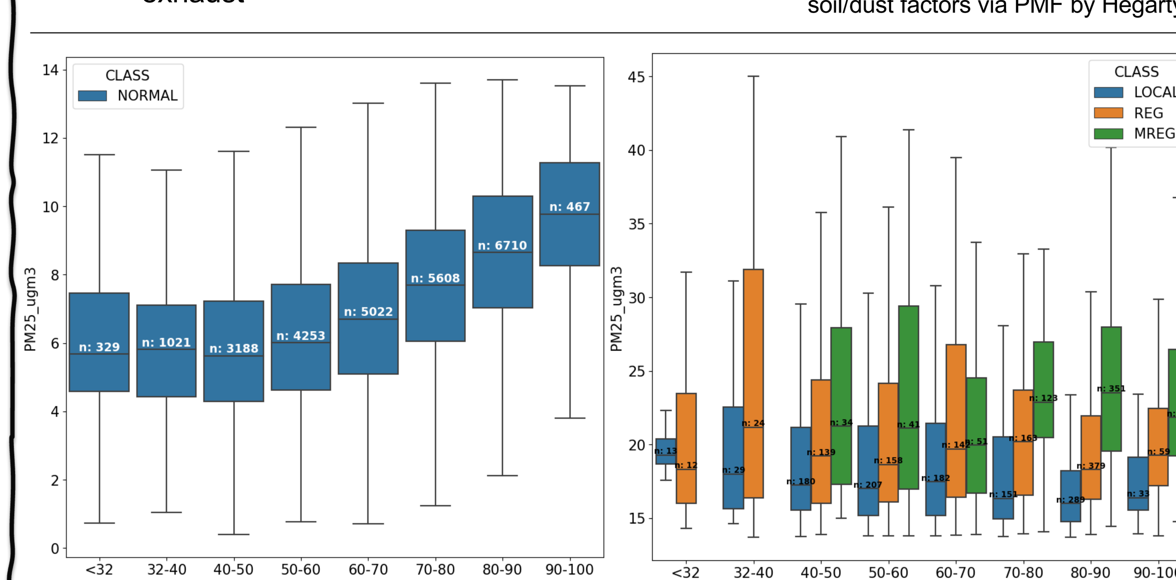

Figure 8. DFW region $PM_{2.5}$ as a function of air temperatures (deg F). Plot of normal $PM_{2.5}$ (left) and anomalous $PM_{2.5}$ (right) binned by air temperature. Outliers are removed for clarity.

- As with all other regions, lower air temperatures are typically associated with higher $PM_{2.5}$ anomalies (especially local and regional anomalies). This pattern suggests contributions from winter heating sources.
- Anomalous $PM_{2.5}$ relation to temperature is distinct from normal $PM_{2.5}$ response, which tends to increase with temperature. This indicates different processes dominating baseline $PM_{2.5}$ formation/sources.

IV. Conclusions and Future Work

- The Isolation Forest technique is effective for $PM_{2.5}$ anomaly detection and exploration.
- Bulk analysis suggests regional and multi-regional anomalies are associated with specific wind directions and transport patterns; anomalies generally increase as temperature drops.
- Detailed analysis suggests anomaly categories – i.e., local, regional, multi-regional – are robust and accurately reflect the nature of the $PM_{2.5}$ source emission.
- Using examples from two sites (Dallas Hinton and Houston Deer Park), we combined $PM_{2.5}$ anomaly data with PMF-type analysis on sources of $PM_{2.5}$. We found that local anomalies tend to be correlated with diesel exhaust, while regional- and multi-regional anomalies tended to be correlated with agricultural burning and/or dust events.

References & Acknowledgements

Dayalu *et al.* (2023), *Investigating Sources of PM2.5 in Texas with Multiple Techniques*, Final Report to Texas Commission on Environmental Quality (TCEQ) for Work Order No. 582-23-42350-016 under TCEQ Contract No. 582-19-90498, June 30.

Hegarty *et al.* (2023), *Positive Matrix Factorization Analysis of the Dallas Hinton and Deer Park Speciated Monitoring Data*, Final Report to Texas Commission on Environmental Quality (TCEQ) for Work Order No. 582-23-43374-018 under TCEQ Contract No. 582-19-90498, June 30.

Liu, F.T., Ting, K.M., and Zhou, Z.-H, "Isolation Forest," *2008 Eighth IEEE International Conference on Data Mining*, Pisa, Italy, 2008, pp. 413-422, doi: 10.1109/ICDM.2008.17.

NOAA HMS Smoke Explorer. https://globalfires.earthengine.app/view/hms-smoke

Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011, Scikit-learn: Machine Learning in Python, https://scikit-learn.org/stable/modules/outlier_detection.html#isolation-forest

This presentation is based on work supported by the State of Texas through a contract from the Texas Commission on Environmental Quality (TCEQ). The conclusions are the authors' and do not reflect TCEQ policy.