

Low-Cost Investigation Into Levels and Sources of Air Pollution in an NYC Suburban Neighborhood Using Non-Negative Matrix Factorization

Polina M. Goldberg^{1,2,3}, Abhishek Anand³, Daniel M. Westervelt³

¹ *Department of Computer Science, Columbia University, New York, NY 10027, USA.*

² *Department of Biological Sciences, Columbia University, New York, NY 10027, USA*

³ *Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY 10964, USA.*

Air pollution poses a large environmental health threat, accounting for a large proportion of hospital admissions and premature deaths worldwide. While urban air quality is frequently monitored and reported, high-quality suburban air pollution monitoring remains an area of need. In this study, we analyze high temporal resolution data from a suburban site in Palisades, New York, approximately 30 km northwest of New York City (NYC). We deployed a GRIMM Environmental Dust Monitor 180 (GRIMM EDM-180) for reference PM_{2.5} and a QuantAQ Modulair sensor, the latter of which provides measurements of co-pollutants NO, NO₂, O₃, and CO, in addition to PM₁, PM_{2.5}, and PM₁₀. A multiple linear regression model was developed to correct low-cost PM_{2.5} measurements based on temperature and relative humidity, improving agreement with reference-grade data from $r = 0.79$ to $r = 0.81$ and lowering mean absolute error (MAE) from 1.72 $\mu\text{g}/\text{m}^3$ to 1.63 $\mu\text{g}/\text{m}^3$. To identify dominant air pollution sources, we use a non-negative matrix factorization (NMF) method on time-resolved particle size distribution data and gaseous pollutant concentrations. A four-factor NMF solution revealed distinct pollution signatures tied to urban vehicle emissions, along with local combustion activities and distant wildfires. Our results highlight the importance of targeted monitoring and clear air mitigation strategies beyond urban centers.