Presidential Initiative on Urban and Place-Based Research
Final Project Report

Project Title

Developing Descriptive and Predictive Causal Models to Study the Impacts of Highway Construction on Ambient Air Quality in the Front Range

Project Team

Principal Investigator (PI) - University of Colorado Denver

Farnoush Banaei-Kashani
Associate Professor
Department of Computer Science and Engineering
University of Colorado Denver
Lab: http://cse.ucdenver.edu/~bdlab/
Homepage: http://cse.ucdenver.edu/~farnoush/
Phone: (303)315-0116
Email: farnoush.banaei-kashani@ucdenver.edu
ORCID: 0000-0003-4102-9873

Co-Investigator (Co-I) - Denver Department of Public Health and Environment (DDPHE), City and County of Denver (CCD)

Michael Ogletree
Air Quality Program Manager
Email: michael.ogletree@denvergov.org

Co-Investigator (Co-I) - Colorado Department of Transportation (CDOT)

Stephen Cohn
Research Branch Manager
Email: steve.cohn@state.co.us

Summary of Goals and Methods

Air pollution is a complex environmental issue that poses a significant risk to public health. Based on studies performed by World Health Organization (WHO), it is believed that almost 99% of the global population breathes air that exceeds the guidelines [1], with developing countries being the most vulnerable and exposed to hazardous pollutants. Long-term exposure to air pollutants such as fine Particulate Matter (PM2.5), Sulfur Dioxide (SO2), and Nitrogen Oxides (NOx) has been linked to severe health issues related to respiratory and cardiovascular diseases [2], adverse birth outcomes [3], and mortality in general [4]. Moreover, the statistics on the death count due to air pollution is staggering. Studies have shown that air pollution accounts for 6.7 million premature deaths in 2019 alone, which is 11.6% of the total global deaths [5], making air pollution one of the leading causes of mortality. The impacts of air pollution are not restricted to people of certain geographical regions, ages, or backgrounds, rather it affects every human being with people possessing pre-existing health conditions with a weaker immune system, elderly individuals, and children being at a higher risk.

Given the widespread impacts and consequences of air pollution, accurate prediction of air pollution levels is of great importance to the management team. With such models, it not only can identify pollution hotspots and accurately predict pollution levels at designated areas but also aid the public in making smaller day-to-day decisions based on short-term predictions to making more important decisions based on long-term predictions. Accurate prediction models can further help inform people of the severity of air pollution and its impacts on health and well-being.

Traditionally, air pollution prediction models were often based on physics that captures the deposition of particles at a target point considering various meteorological and environmental data. Research efforts have been made in the past to accurately interpolate the spread of the pollutants using several physics-guided approaches leveraging various atmospheric models including the Advection-Diffusion equation, Gaussian Plume equation [9], Comprehensive Air Quality Model with Extensions (CALPUFF) [10], American Meteorological Society/Environmental Protection Agency Regulatory Model (AERMOD) [11], and Community Multiscale Air Quality (CMAQ) [12]. Despite these models being widely used and their advantages, physics-based models have limitations, including assumptions on atmospheric conditions and complex interpretability on environmental factors such as topology, and multifaceted pollution sources. On the other hand, there has been growing interest in developing advanced air pollution prediction models via machine learning and deep learning data-driven approaches due to their extraordinary ability to learn complex patterns [6] [7] [8]. Data-driven approaches generally leverage historical data on several factors including emission data, and meteorological data to learn convoluted and intricate patterns.

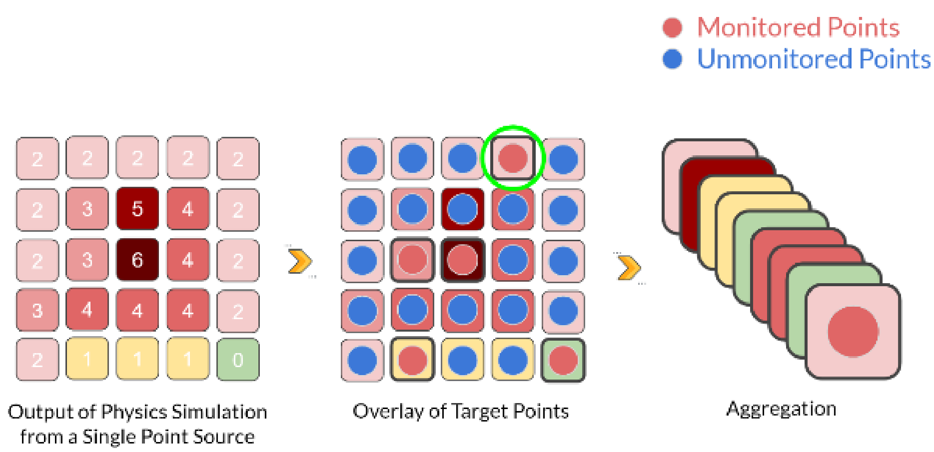
Air pollution is intricate in nature and requires a multifaceted approach. Accordingly, in this study we propose a hybrid model for predicting air pollution at unmonitored locations that combine data-driven methods with physics-based solutions and incorporates quality features extracted from heterogeneous sensor data to ensure the models are aware of data quality, ultimately improving the model's accuracy and prediction capabilities.

Major Findings from Research

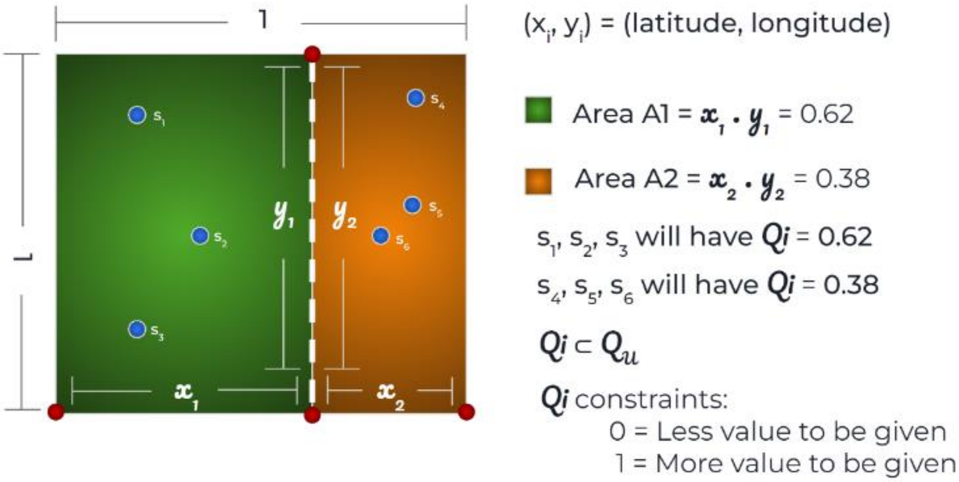
In the first part of this paper, we ran a physics simulation for air pollution dispersion prediction using the Gaussian Plume equation. Then the physics simulation results were processed numerically such that the data-driven model could ingest it. Thus, the processed physics feature was incorporated into data-driven models for the purpose of making a hybrid physics-guided model. Secondly, we extract three quality features with the first one extracted from the performance of the sensor itself, and the remaining two extracted from the spatial context of the available monitoring stations. These features were added to the data-driven model for the purpose of air pollution prediction at any unmonitored location within a spatial grid. The models included in this paper ranged from a simple traditional linear regression model to more complex and novel graphical neural network architecture models. We used available datasets from several sources of varying resolutions and data quality. The model is generic enough to be able to work with various types of air pollutants, however, we focused on and evaluated one of the treacherous air pollutants, i.e., PM2.5.

We observed an improvement in performance across many non-linear models when we started introducing non-linear models such as tree-based and graph-based models. We were able to get the optimal results with graph-based models, and specifically GAT when incorporated with physics, sensor ranking, and spatial proximity features. While the incorporation of physics had some questionable outcomes in general throughout the models, GAT still managed to extract some value from the physics feature.

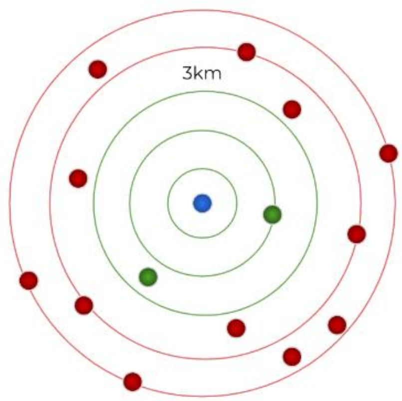
Graphics to Illustrate the Topic



A schematic illustration of a physics-based solution being integrated into a data-driven model as an additional feature from a single point source to a designated (x, y) of a spatial grid



KD-Tree Spatial Data Partitioning for Data Sparsity Quantification



Proximity based Data Sparsity Quantification

List of grants that you or colleagues applied to as a result of this project and if any were funded

We have submitted a proposal to Colorado Department of Transportation, which is currently under consideration:

- Title: "Using Compressive Sensing to Generate Effective Plans for Multi-source, Multi-Modal, and Multi-criteria Sensor Deployments: A Case Study in Air Quality Monitoring to Measure Impact of Highway Construction"
- PIs: Farnoush Banaei-Kashani (PI), Bill Obermann (CCD/DOTI)

Publications and Reports

- The following MS Thesis Report was just submitted to ProQuest to be published:

  Sohil Vaidya, "Physics-Guided Quality-Aware Air Pollution Prediction Model"

- The following conference paper was just submitted to the 10th IEEE International Conference on Data Science and Advance d Analytics (DSAA 2023) for review:

  Sohil Vaidya and Farnoush Banaei-Kashani, "Physics-Guided Quality-Aware Air Pollution Prediction Model", under review.

How many students were funded

One MS student was funded under this project. This student was an undergraduate at the time the project started. Encouraged by the research, the student went on to pursue graduate studies and he recently defended his thesis successfully.

References

1. "WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide", World Health Organization, pp. 290, 2021

2. Guarnieri, M., & Balmes, J. R., "Outdoor air pollution and asthma.", The Lancet, 383(9928), 1581-1592, 2014.

3. Shah, P. S., & Balkhair, T., "Air pollution and birth outcomes: a systematic review." Environment International, 37(2), 498-516, 2011

4. Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., & Pozzer, A., "The contribution of outdoor air pollution sources to premature mortality on a global scale.", Nature, 525(7569), 367-371, 2015.

5. GBD 2019 Risk Factors Collaborators, "Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019" The Lancet, 396(10258), 1223-1249, 2020.

6. Bekkar, A., Hssina, B., Douzi, S. et al, "Air-pollution prediction in smart city, deep learning approach", Journal of Big Data 8, 161, 2021, https://doi.org/10.1186/s40537-021-00548-1

7. Q. Zhang, J. C. K. Lam, V. O. K. Li, and Y. Han, ''Deep-AIR: A hybrid CNN-LSTM framework for fine-grained air pollution forecast,'' 2020, arXiv:2001.11957.

8. Zhao, Z., Wu, J., Cai, F. et al. "A hybrid deep learning framework for air quality prediction with spatial autocorrelation during the COVID-19 pandemic." Sci Rep 13, 1015, https://doi.org/10.1038/s41598-023-28287-8, 2023

9. Turner, D.B., "Workbook of Atmospheric Dispersion Estimates: An Introduction to Dispersion Modeling. 2nd ed.", CRC Press, 1994.

10. USEPA, "User's Guide for the CALPUFF Dispersion Model. Version 7", U.S. Environmental Protection Agency, 2017.

11. USEPA, "User's Guide for the AMS/EPA Regulatory Model (AERMOD). Version 18081", U.S. Environmental Protection Agency, 2017.

12. USEPA, "User's Guide for the Community Multiscale Air Quality Modeling System (CMAQ). Version 5.3.2", U.S. Environmental Protection Agency, 2020.