A Framework for Linking Urban Traffic and Vehicle Emissions in Smart Cities

Clark Hathaway
University of Tennessee, Knoxville
chatawa@vols.utk.edu

Sebastian Mobo
University of Tennessee, Knoxville
smobo@vols.utk.edu

***Abstract*—We present a methodology to perform an analysis of traffic flows and the resultant vehicle exhaust emissions. We design and implement an analysis and visualization framework that includes metrics and algorithms for adjusting, transforming, relating, and visualizing data, as well as statistical methods for quantifying correlations. Our findings from applying this framework to the Chicago Loop reveal the relationships between traffic congestion, building distributions, and vehicle emissions. These insights can provide communities with decision-making tools for urban design and smart cities.**

1. Introduction

Urban traffic flows are complex phenomena influenced by a diverse range of factors, such as road network topology, physical structures, and human behaviors [1]. These traffic flows, in turn, not only consume large amounts of energy, but impact their environment by emitting heat and gases into their surroundings. However, accurately modelling these relationships can be challenging, as it requires disparate data sets to be unified with an interdisciplinary approach. Fusing these datasets into a coherent analysis of traffic patterns could guide efforts toward reducing traffic emissions and improving traffic energy efficiency.

We develop a methodology to understand the relationship between traffic patterns and emissions. To analyze relevant datasets, we design metrics and algorithms that reconcile their potential disparities and analyze their relationships. During our research, we also develop a suite of data analysis, visualization, and validation tools to prepare and analyze traffic and environment data. Our methodology and tools can be universally applied to study any traffic scenario for which environment and vehicle information (e.g., vehicle positions, road network layout, vehicle features) are available. The Python and Rust implementation of the framework is open source[1], portable across platforms, and able to handle realistic data volumes.

In order to apply our methodology to a realistic scenario, we examine traffic in the Chicago Loop with data provided by Oak Ridge National Laboratory (ORNL)[2]. The data consist of a time series of simulated traffic flows depicting vehicle locations, emissions quantities per road, building footprints, and the layout of the road network. Using our framework, we can address the differences in how these data are spatially organized and temporally sampled, characterize relevant patterns, and correlate emissions with environmental variables.

1. https://github.com/TauferLab/UrbanTrafficFramework_20
2. https://smc-datachallenge.ornl.gov/challenges-2020/challenge-4-2020/

2. Methodology

Our methodology includes four steps: preparation and validation of the data, mapping vehicles to buildings, mapping emissions to spatial cells, and correlation analysis between vehicle and emissions.

2.1. Data Preparation

We deal with heterogeneous data that need some preparation for our analysis. To this end, we validate, transform, and extend the data in the interest of constructing a unified topology of the urban environment.

2.1.1. Traffic Data Validation. Almost every dataset requires some form of preprocessing or cleaning, and the vehicle position data we study here is no exception; we found that some vehicles in the data came with unrealistic locations (e.g., they were located in lakes or rivers). The vehicle position data used in our work includes both geographic coordinates and IDs of travelled roads. By taking advantage of this redundancy, we recompute any vehicle's unrealistic position by inference from the known road network layout as a preprocessing step. In other words, we relocate the impacted vehicles to more realistic locations.

2.1.2. Building Footprints Preparation. We leverage the GeoJSON building footprints dataset maintained by Microsoft from all 50 states[3] to consider about 2,600 buildings within the Chicago Loop. While this format is general, we still need to extend the building footprints so that those buildings can be combined with other geographic data. To this end, we redefine each building (i.e., GeoJSON entity) in terms of a bounding box and centroid. This representation allows us to effectively link both vehicle locations and distributed emissions values to bounding boxes, thus merging vehicle locations and emissions into a common topology.

3. Available at https://github.com/microsoft/USBuildingFootprints

2.2. Vehicle-Building Mapping

To consider relationships between vehicles and their destinations, we adapt an agent-based approach in which individual vehicle agents are mapped to nearby buildings. As efficiency and speed are major concerns given the size of these types of data, our method organizes buildings within the mapping region into a k-dimensional tree, where each building is keyed by the centroid of its footprint, and then uses this tree to search for the building closest to each agent efficiently. This approach is straightforward to implement, is parallelizable, and performs with an acceptable run-time.

2.3. Dispersion of Traffic Emissions

Our analysis uses data generated by the MOVES-Matrix vehicle emissions simulator which outputs an emissions quantity associated with each road segment in a given network. Because emissions in this form are difficult to associate with other features in the data, we associate emissions with points in space by dividing the region of interest into cells and modelling how emissions flow between the cells. In our model, cells that overlap with roads become sources that add emissions to nearby cells. The emissions that are added to a given nearby cell are inversely proportional to the squared distance between the cell and its source cells. Our process models the dispersal of emissions from individual roads into the surrounding environment, and allows us to incorporate emissions into the urban topology. In our analysis, we consider heat emissions from vehicle exhaust, though this step can easily be adapted for other types of emissions.

2.4. Correlation Analysis

As part of our analysis method, we define two metrics per building by aggregating results: for each building, we count the *number of vehicles* mapped to it and sum up *emissions values* from the dispersion model over the cross-sectional area described by the building's footprint. Patterns linking vehicles and emissions can be identified qualitatively with visualizations and quantitatively with a statistical analysis.

Visualizations involve heatmaps of the emissions overlaid on a map of building footprints colored by mapped vehicle counts for specific instants in time. These maps indicate the degree to which traffic and emissions hotspots are co-located.

For our statistical analysis, we fit a linear regression model using normalized and transformed versions of these metrics, predicting emission concentrations from mapped vehicle counts for each building. To make these models more consistent with each other and with the data, we normalize both metrics by dividing by the building footprint area for each data point to account for building size, take the square root of the normalized vehicle counts prior to fitting, and ignore outliers based on a simple emissions concentration threshold.

3. Results

After applying our framework and analysis methods to the Chicago Loop, we qualitatively infer from our visualizations that for a typical day, high-activity areas converge in the west between 07:00 to 12:00, with smaller hot spots appearing in the northeast and southeast.

Quantitatively, we find a weak correlation between mapped vehicle counts and emission concentrations from 08:00 to 11:59 ($r = 0.151$ to 0.220, $p < 0.01$) that seems to peak around 09:00; meanwhile, emission concentrations for hours past 12:00 are simply too low and too sparse for our models to fit reliably. This sparsity is likely due to our data neglecting commutes from work to home. Finally, although there is activity at 12:00 to analyze, emissions concentrations are lower and sparser, with a weaker correlation ($r = 0.119$, $p = 0.032$).

4. Conclusions

We present an effective solution for investigating the relationships between urban traffic flows and vehicle emissions. We propose a general methodology to integrate vehicle and environment data sets into a single coherent analysis of traffic patterns. Specifically, we design methods for computing vehicle-emission maps by mapping both vehicles and emissions to buildings; we create metrics to analyze the correlation between traffic and emissions; and we provide portable open-source implementations of the associated analysis and visualizations that are capable of processing realistic data.

We apply our framework to a specific scenario covering the Chicago Loop area. We conclude that the volume of traffic flow has a weak positive correlation with emissions around buildings. This suggests there are additional factors affecting this relationship that could be reflected in the methodology in future work, such as taking into account all the buildings in the area surrounding a vehicle and accounting for building height or mapped vehicle types, which could affect emission concentrations.

Acknowledgment

Supported by IBM Shared University (SUR) Award and NSF awards IIS 1841758 and CCF 1841758.

The authors wish to thank Silvina Caíno-Lores, Travis Johnston, and Michela Taufer for their mentorship during this project.

References

[1] A. Berres, P. Im, K. Kurte, M. Allen-Dumas, G. Thakur, and J. Sanyal, "A Mobility-Driven Approach to Modeling Building Energy," in *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019*, pp. 3887–3895, Institute of Electrical and Electronics Engineers Inc., dec 2019.