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

Sebastian Mobo, Clark Hathaway
Mentored by Silvina Caino-Lores, Travis Johnston, Michela Taufer

THE UNIVERSITY OF TENNESSEE KNOXVILLE
BIG ORANGE. BIG IDEAS.®

Motivation

- Most people are already familiar with the presence of traffic in big cities and the harms it introduces
- Congestion is a well known problem but the irreparable damage done by emissions poses a threat to public health and the environment
- Despite these obvious concerns, the questions researchers ask in tackling them do not come with simple answers

Motivation

The questions we ask are:

- How do traffic patterns affect vehicle emissions?
- Specifically, does traffic congestion induce additional emissions?
- If we can mitigate emissions by reducing congestion, how do we define and measure congestion itself?

Motivation

Other questions include:

- If traffic patterns are driven by human behavior, how do we separate this from other variables?
- And, finally, how do we consider multiple datasets when they are diverse in both
  - their format and
  - method in which they were sourced?

Motivation

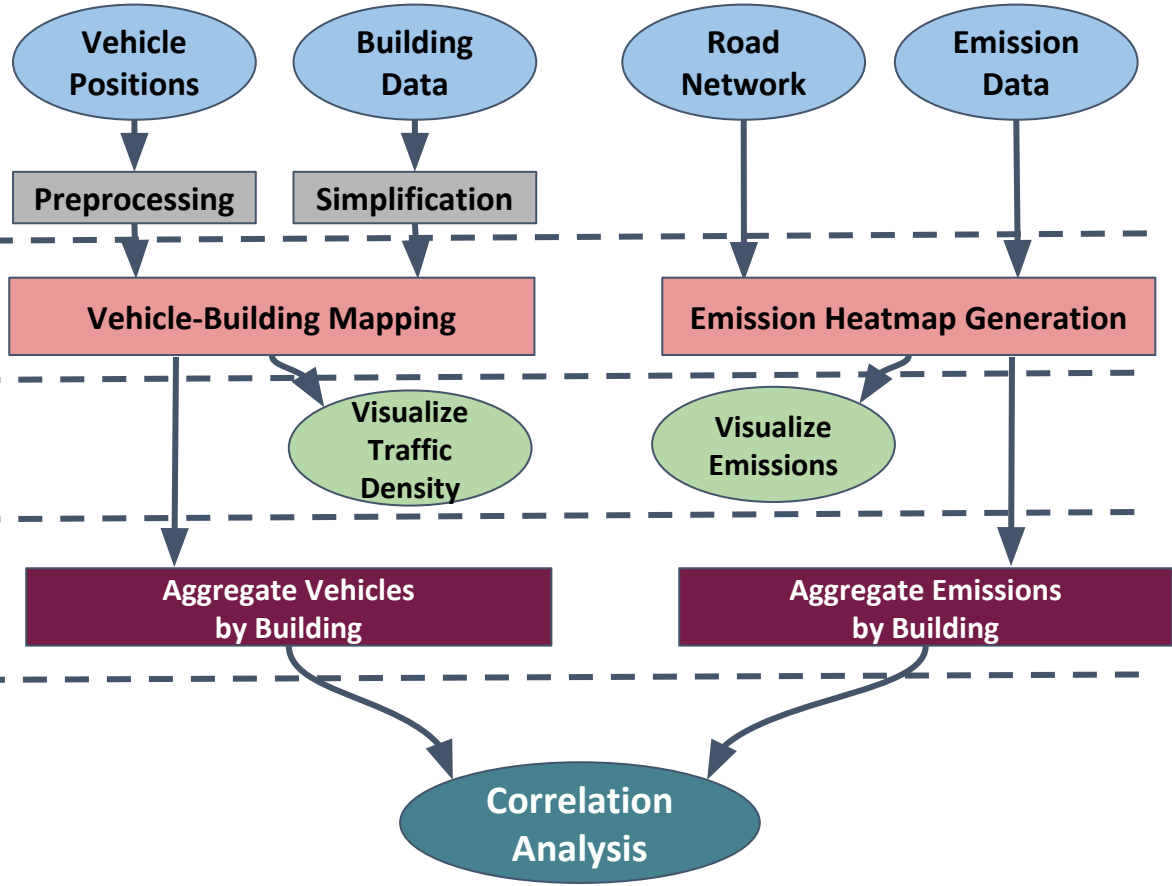
- We hypothesize that **traffic congestion yields a significant increase in vehicle emissions**
- To test this, we will:
  - Develop metrics to measure traffic congestion
  - Perform a correlation analysis to show that emissions are increased when congestion increases

Motivation

- In our exploration of this problem, we develop a methodology to
  - understand the relationship between traffic and emissions
  - provide a generalized framework for future work
- We apply this methodology to data describing the Chicago Loop so we can
  - validate our framework and
  - test our hypothesis

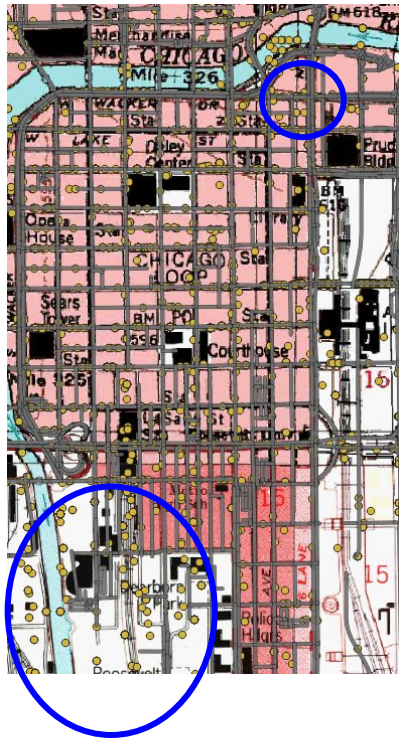
Our Framework

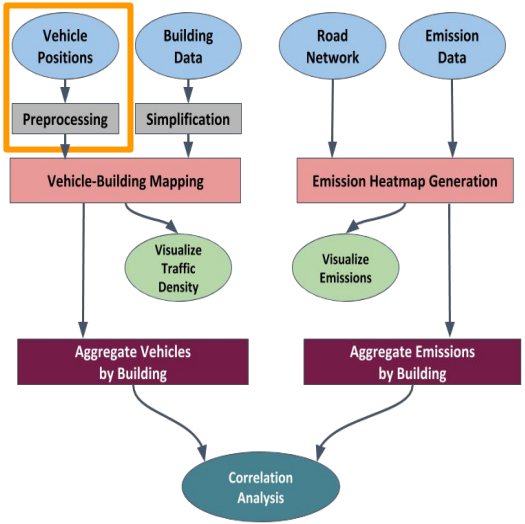
- **Extrapolate** and **transform** the data
- **Fuse** data into a **unified** urban **layout**
- **Visualize** the models
- **Aggregate** metrics over urban layout
- Reveal **meaningful correlations**



Processing Traffic Data

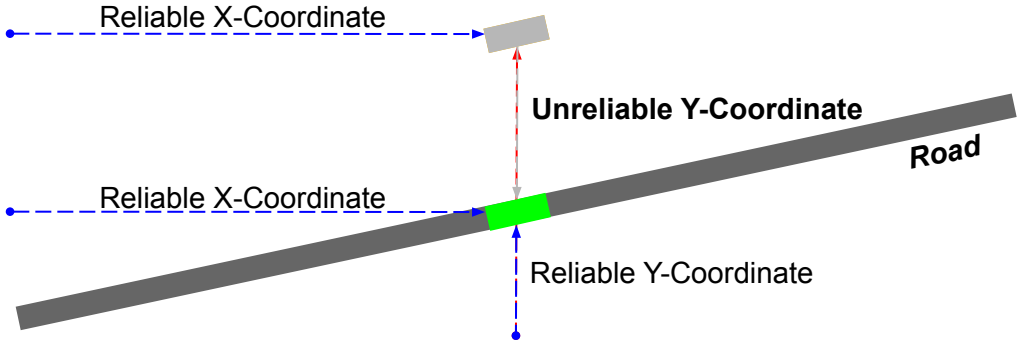
- Our original data had unrealistic locations for vehicles
  - Roughly 50% of vehicles have impossible offsets along roads
  - Another 25% are placed in water, on railroads, etc.
- Pre-processing must validate and make corrections

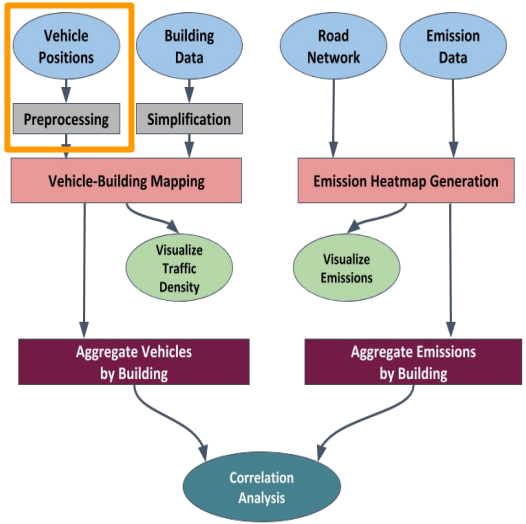




Processing Traffic Data

Solution: take advantage of redundancy in vehicle position records to recompute invalid coordinates.

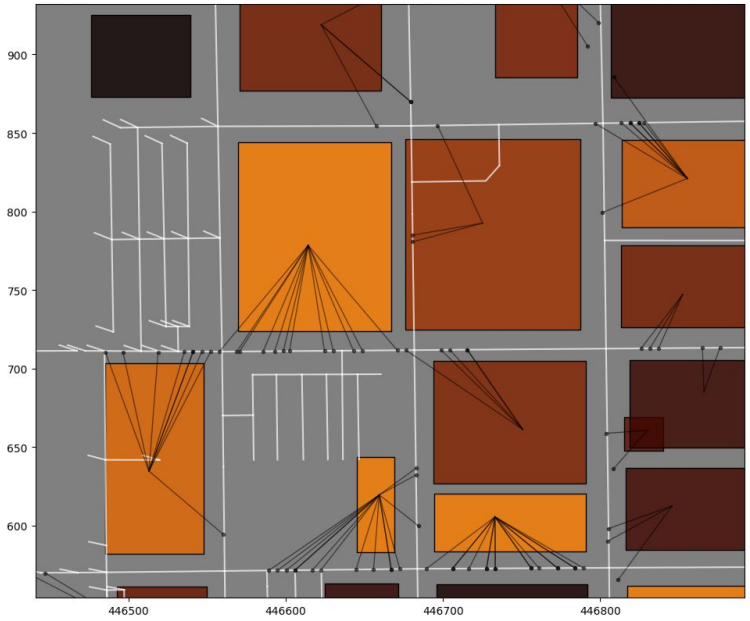


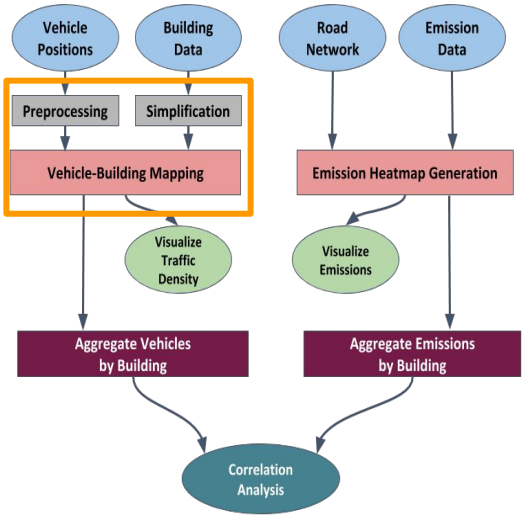


Vehicle-Building Mapping

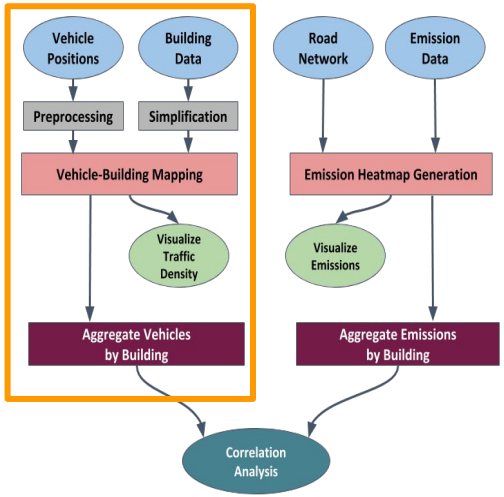
Simplify building footprints by computing centroid and cross-sectional area

Then map vehicles to nearby buildings using k-d trees





Vehicle-Building Mapping

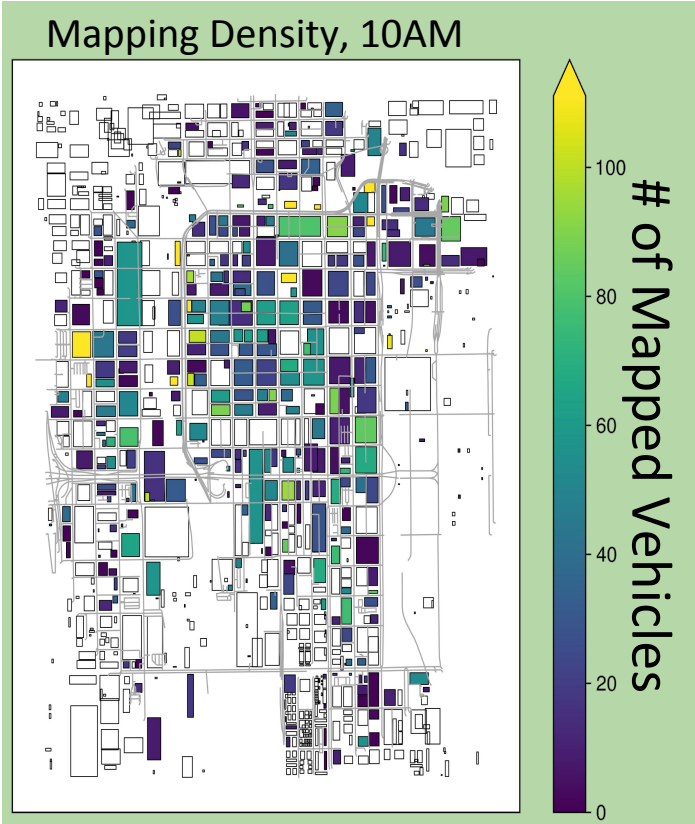


- Vehicle Positions: UTM coordinates and road IDs
- Building Footprint Data: GeoJSON polygons

- Correct inaccurate vehicle positions
- Simplify buildings to reduce computational load

- Map vehicles to buildings using *k*-d trees

- Aggregate mappings into per-building vehicle counts

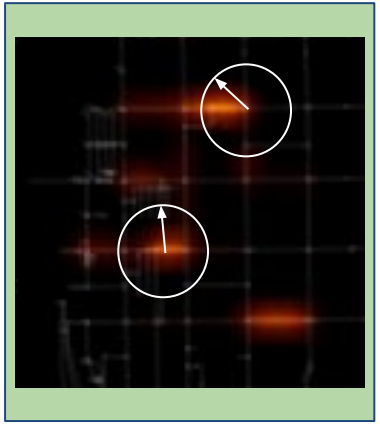


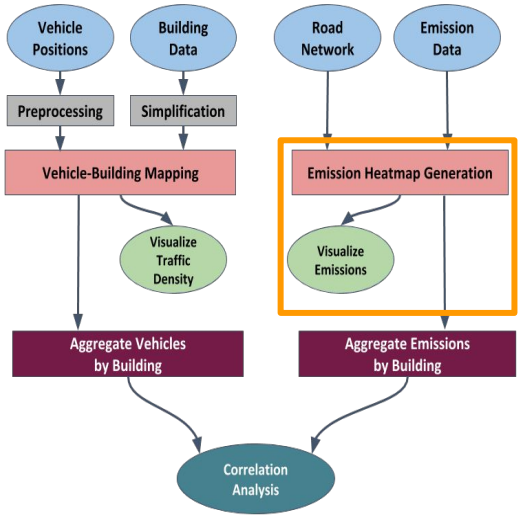
Dispersion of Traffic Emissions

Dispersal of emissions is modeled using a **cell-based heatmap**

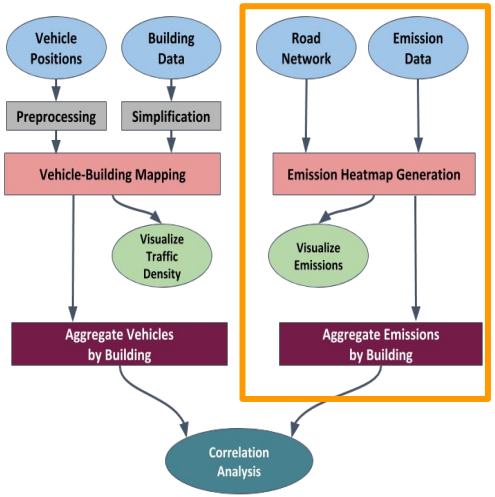
Cells defined over the area of a road become **source cells**

The algorithm computes the value of **1 / r²** for each cell where **r** is the distance from a given **source cell** nearby





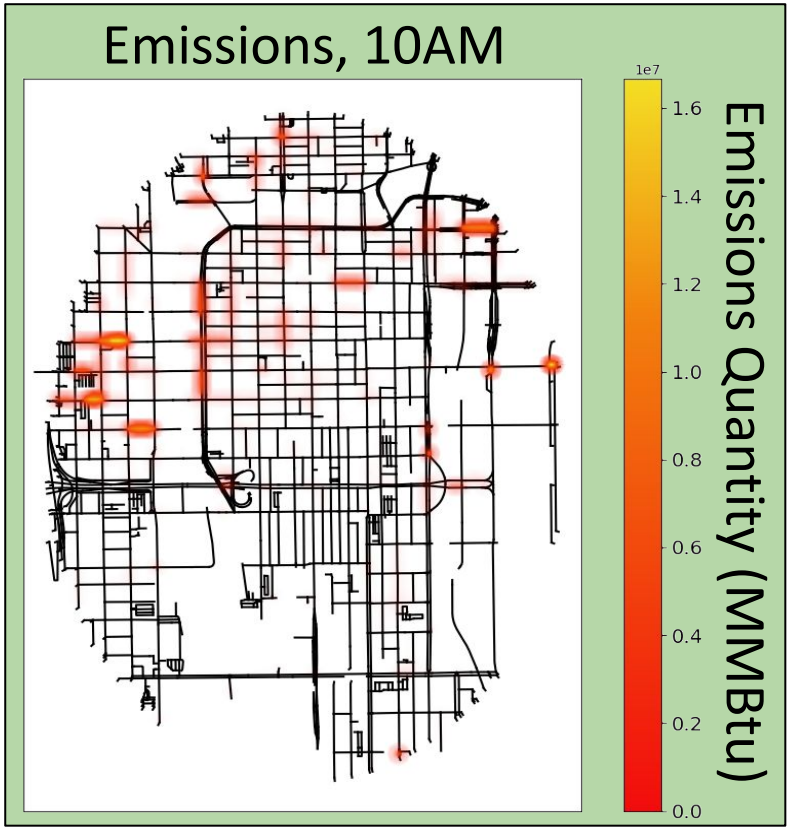
Dispersion of Traffic Emissions



- Road Network: GeoJSON line-strings
- Emissions: per-road heat emission totals

- Model dispersal of emissions from roads using a cell-based heatmap

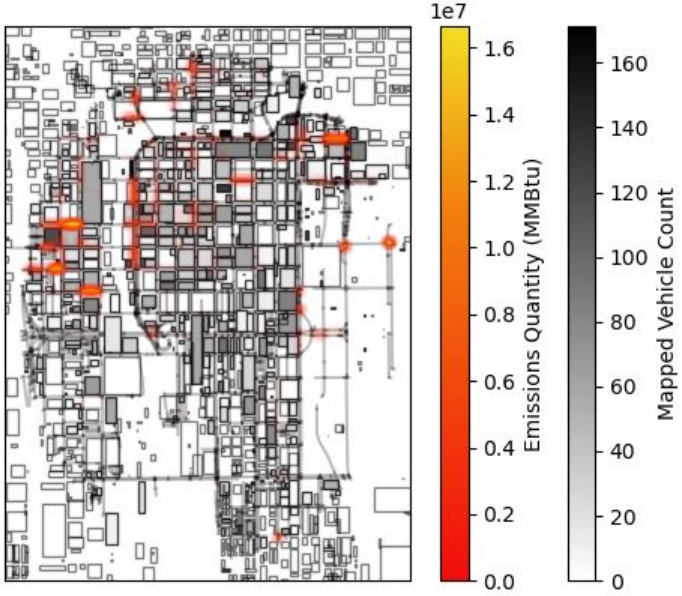
- Aggregate emissions values into per-building emission sums

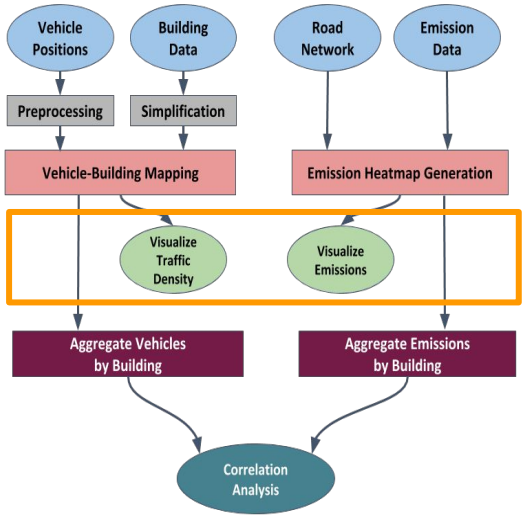


Correlating Traffic and Emissions

Overlay visualizations of the vehicle counts per building and dispersed emissions

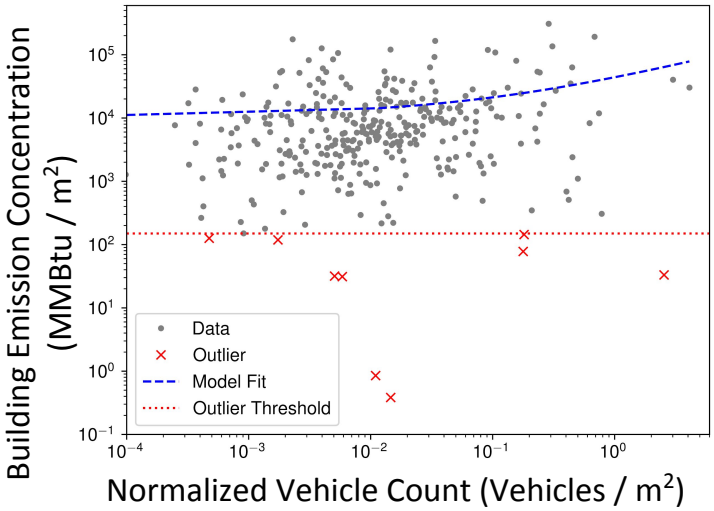
Allows qualitative analysis of emissions and traffic hotspots, and how they coincide

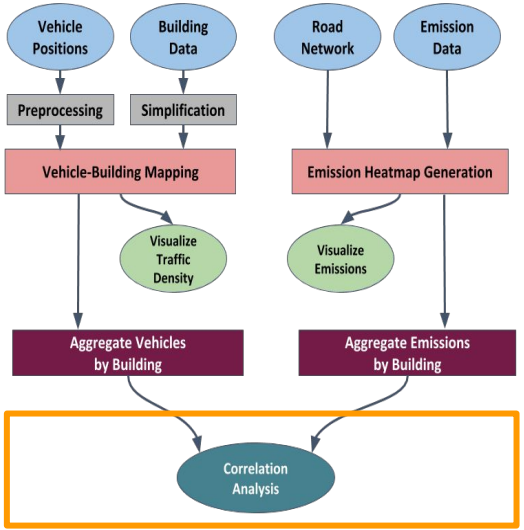




Correlating Traffic and Emissions

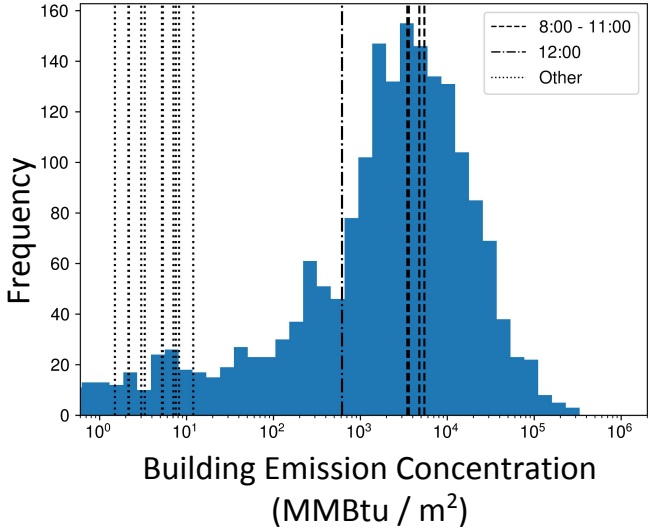
A weak positive correlation exists between mapped vehicle counts and emission concentrations from 08:00 to 11:00 (r = 0.151 to 0.220, $p < 0.01$)

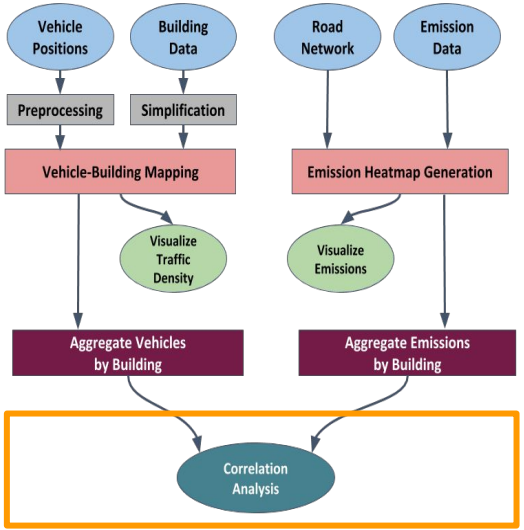




Correlating Traffic and Emissions

Morning “rush hour” effect is also visible when plotting emission concentration values





Artifacts

- Heat map generation tool, which can generate similar images for any GEOJSON road network file and properly formatted CSV file.
- Vehicle mapping tools and library, which take the same inputs as above, plus building data, and compute vehicle-building mappings.

Lessons Learned

- In our analysis of the Chicago Loop we
  - apply methods for **characterizing**, **cleaning**, and **fusing** data about traffic and emissions,
  - find a weak **correlation** between traffic and emissions during morning commute hours, and
  - observe both **spatial** and **temporal** patterns in emissions throughout the area of interest

Future Work

- Model other variables that affect emissions:
  - Building height
  - Vehicle types
  - Weather
- Develop other mapping methods:
  - For example, map vehicles to multiple buildings, based on a distance threshold

Our Datasets

- Very broad, with varying granularity
- Our datasets cover:
  - Built environment
  - Traffic and vehicles
  - Emissions

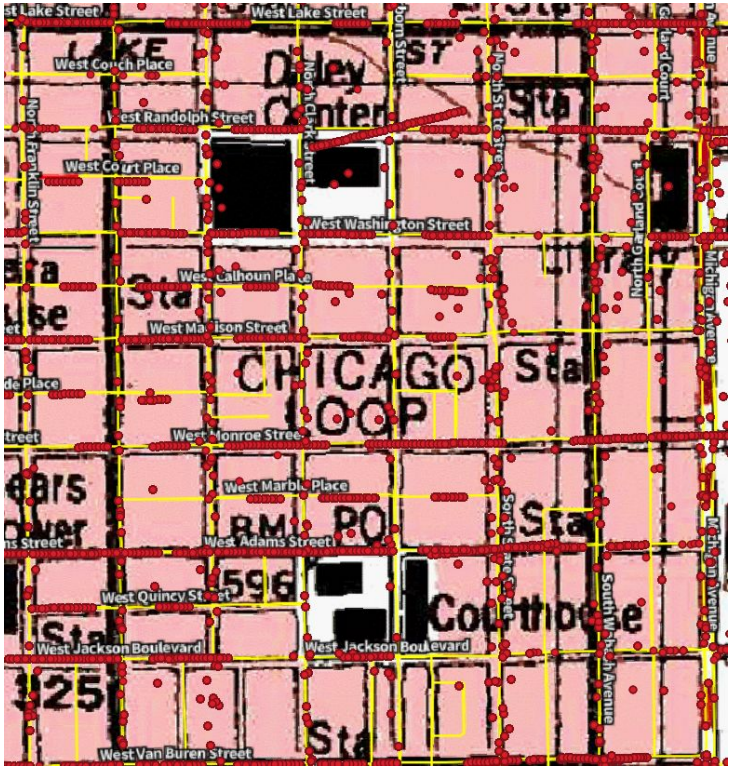
Our Datasets

- Very broad, with varying granularity
- Our datasets cover:
  - Built environment
    - Building footprints
    - Local road network
  - Traffic and vehicles
  - Emissions

GeoJSON road network overlay taken from www.geojson.io, map image provided by MapBox

Our Datasets

- Very broad, with varying granularity
- Our datasets cover:
  - Built environment
  - Traffic and vehicles
    - Traffic simulations
    - Vehicle type survey
    - Commute schedule survey
  - Emissions



Our Datasets

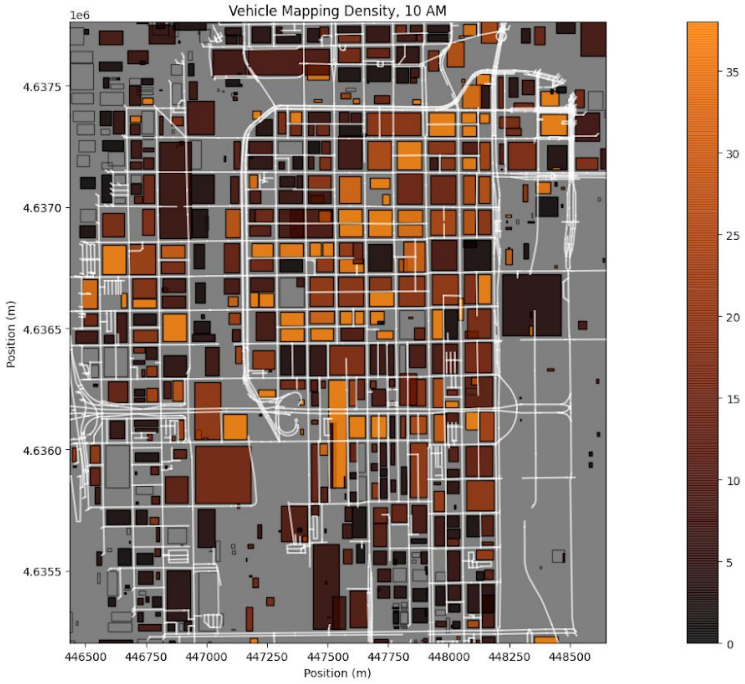
- Very broad, with varying granularity
- Our datasets cover:
  - Built environment
  - Traffic and vehicles
  - Emissions
    - Vehicle heat exhaust
    - Per-link and per-hour

Mapping Vehicles to Buildings

- Being able to associate vehicles with nearby buildings can be useful for:
  - Visualizing the impact of vehicle emissions on nearby buildings
  - Visualizing how urban layout affects traffic congestion
  - Inferring building occupancy from nearby traffic
- Methods for computing mappings need to scale:
  - Our simulated dataset includes around 107,000 different agents that need to be mapped to about 2,600 buildings

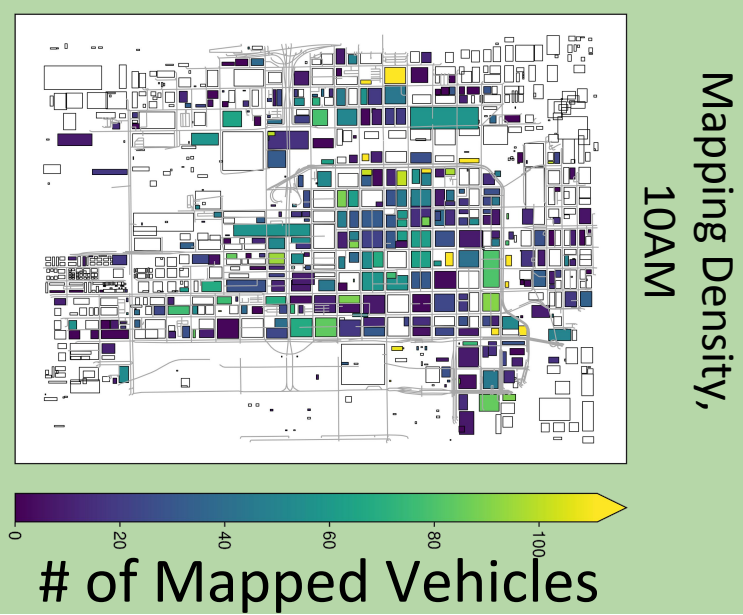
Mapping Vehicles to Buildings

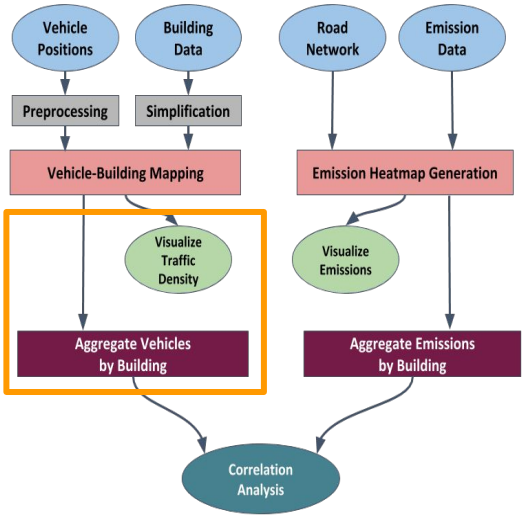
- This is nearest-neighbor search.
- We have investigated two mapping methods so far:
  - A quadtree-based approach developed by the challenge authors, and
  - An alternative approach using k-d trees.
- k-d trees work well enough for the number of vehicles and buildings we have.



Vehicle-Building Mapping

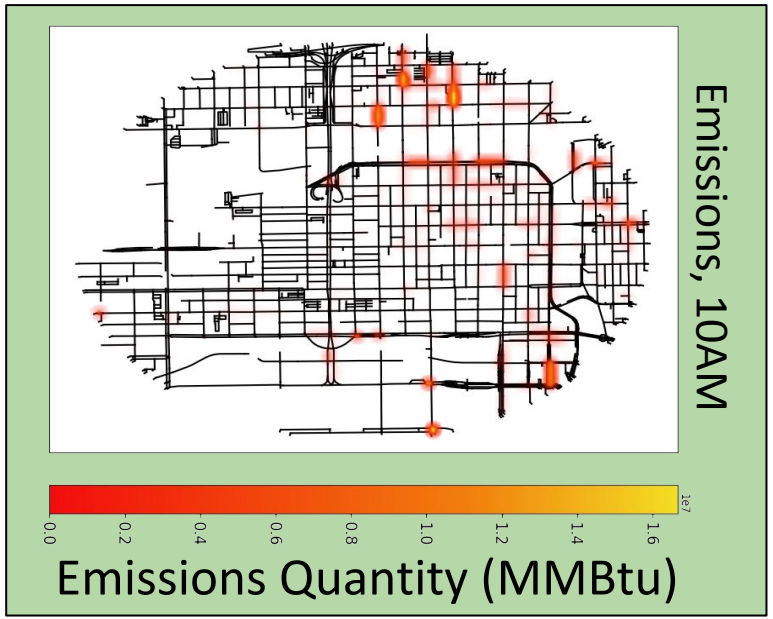
Aggregate these mappings by building to compute traffic density per building

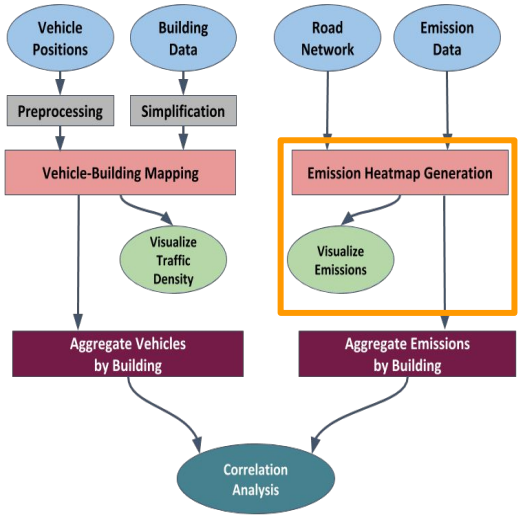




Dispersion of Traffic Emissions

Model output can be visualized with a map

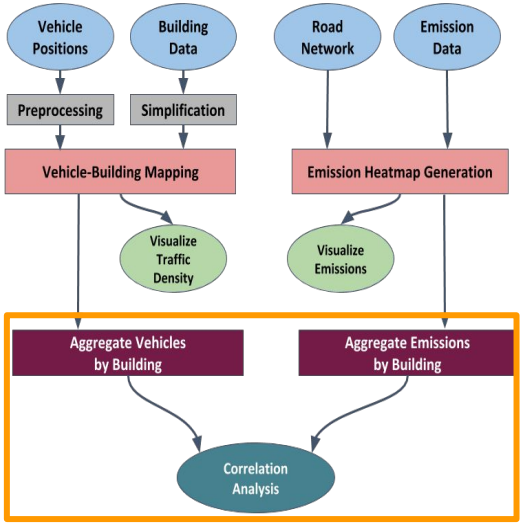




Correlating Traffic and Emissions

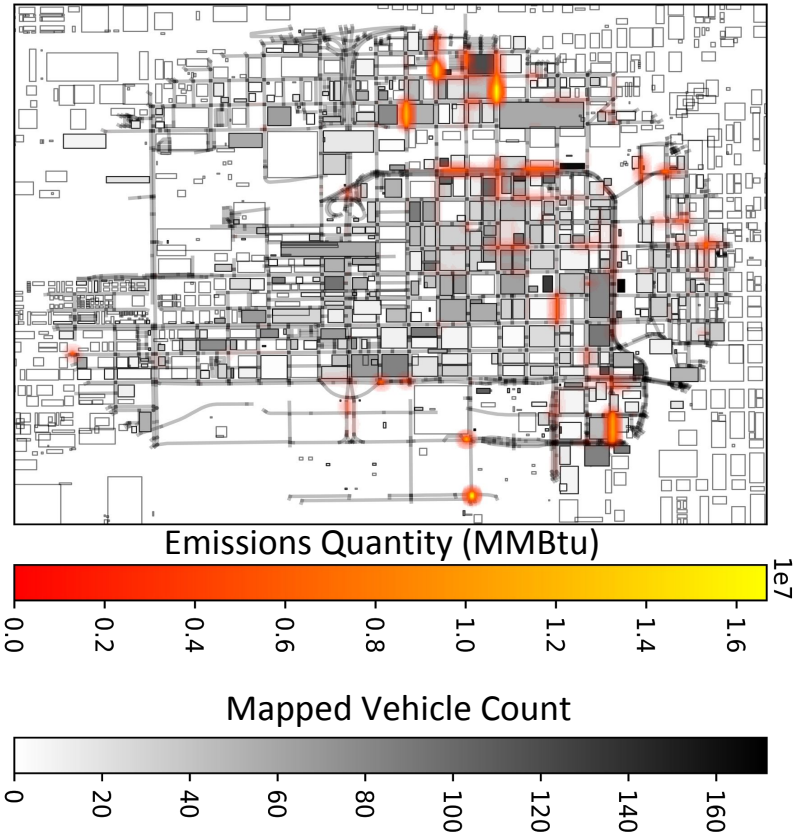
Overall, we compute two metrics for the correlation analysis of traffic and emissions:

- Traffic: Normalized vehicle count
  - Number of vehicles mapped to each building
- Emissions: Emission concentration
  - Sum of emission cell values over building footprints
- Both metrics are divided by building area



Correlating Traffic and Emissions

- It is clear that the traffic simulator did not factor in random traffic patterns
- The correlation we assumed in our hypothesis between traffic density and emissions is apparent
- We can further qualify these observations with our statistical analysis



Outline

1. Establish the motivation behind this research and provide our hypothesis
2. Describe our framework
3. Explore the datasets we use to validate our framework
4. Examine the results of applying our framework and the impact on our hypothesis

Our Initial Approach

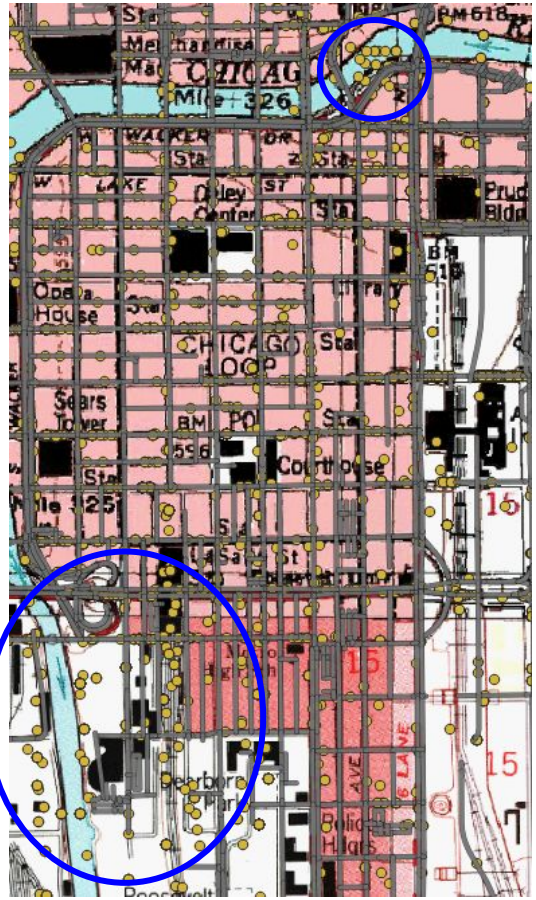
1. We explored the scope of our data and used visualization techniques to characterize it.
2. We determined we wanted to see what traffic patterns existed before addressing the other questions.
3. We hypothesized that if we could measure traffic congestion, that metric would show a correlation with exhaust emissions in the same area.

Problems with Data

- Our traffic simulation data had vehicles in impossible locations.
- This was due to a misconfiguration of the simulator and we do not have a corrected version.
- This simulator, however, only provides part of the data.
- We decided to focus our efforts on the more reliable datasets provided; link volumes, commute schedules, emissions, etc.

Problems with Data

- Problematic vehicle positions include:
  - Positions that aren't aligned to roads
  - Positions that intersect buildings, railways, or bodies of water
  - Positions that don't tally with the current road segment the vehicle is travelling on
- Roughly 50% of vehicles have impossible offsets along the links they are travelling on.
- Another 25% have recorded coordinates that are inconsistent with the layout of the road network.

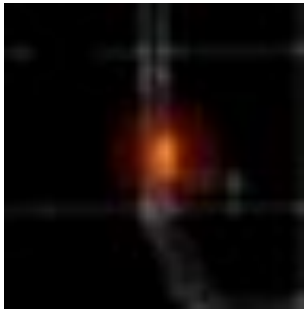


Visualization Tools

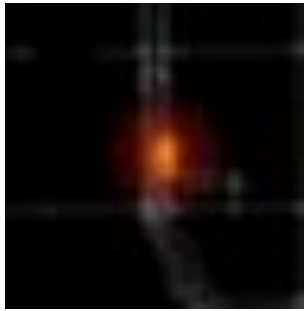
- We developed several scripts for displaying the data we had over the topology of the city.
- This revealed some of the problems we faced early on and allowed us to validate our remaining data.
- This motivated us to develop additional visualizations to demonstrate interesting patterns we found.

Heat Map of Vehicle Emissions

- One of the patterns revealed is that there is no variation from one day to another.
- In fact, there is no numerical difference between these two images, or any two days with at the same hour.



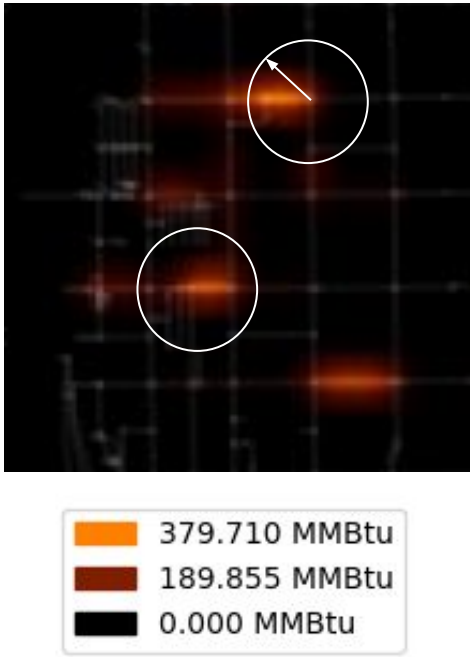
07/04/17 2:00 pm



07/07/17 2:00 pm

Heat Map of Vehicle Emissions

- We assume that emissions dissipate from the source with respect to distance
- The algorithm assumes a relationship of 1 / r^2 where r is the distance from the source



Comparison of Emissions and Mappings

- There are several areas where increased activity with occupants accessing buildings shows increased emissions levels.

Comparison of Emissions and Mappings