Technical Memorandum for MAPC Research Brief:

"Racial disparities in the proximity to vehicle air pollution sources in the MAPC region."

Prepared by Conor Gately, Senior Land Use and Transportation Analyst, MAPC

For further inquiries please contact: cgately@mapc.org

Data Availability

A shapefile of the 250-meter gridded population data with associated PPI scores is available for download at the MAPC DataCommon: https://datacommon.mapc.org/browser/datasets/413

Development of Pollution Proximity Index (PPI)

In this study, we analyzed the spatial patterns of residents living in close proximity to roads with the highest levels of vehicle air pollution emissions across the MAPC region. We focus on households living within 150 meters of high-emission roads and explore how the demographics of these populations changes as the pollution intensity of the roadways increases. To do this, we developed the Pollution Proximity Index (PPI), a weighted score that accounts for five factors: the volume, speed, and fleet mix of traffic on a given roadway; distance of residents to that roadway; and proximity to other high-emitting roadways nearby. High PPI values are found in areas with multiple high-traffic roadways — in particular, freeways — and urban arterial roads that carry a large volume of diesel-powered freight vehicles.

The PPI was constructed using estimates of nitrogen oxide (NOx) emissions from on-road vehicle sources at roadway scales generated by *Gately et al. (2017)*. A detailed description of the methods used to calculate these emissions is provided in later sections of this memo. NOx emissions are strongly correlated with emissions of ultrafine particles (UFP), with measurement studies showing that the concentrations of NOx and UFP follow similar exponential decay curves as the distance from the roadway edge increases (*Zhu et al., 2002*).

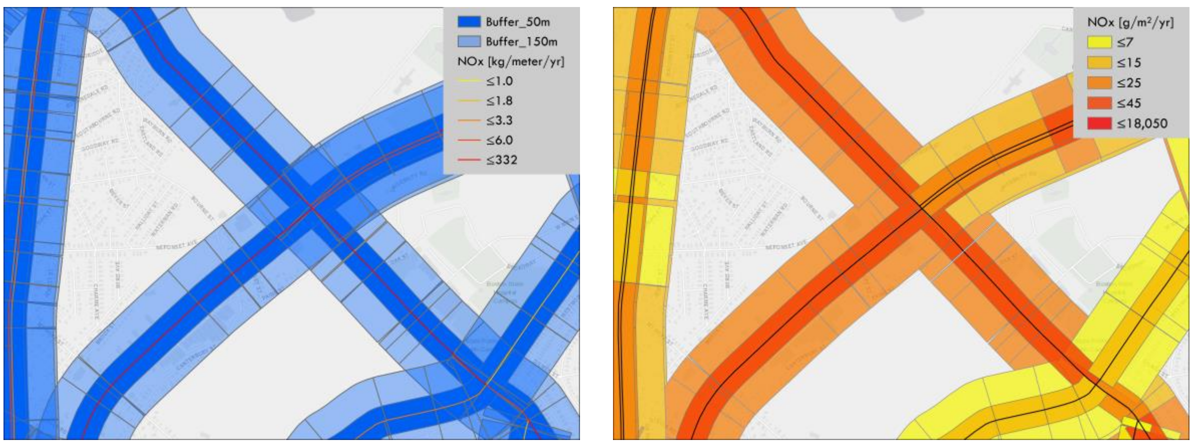
As the original emissions inventory was calculated for the year 2012, we chose to scale these emission forward to the year 2017. We compared the annual average daily traffic (AADT) for each road segment reported in the 2018 MassDOT Road Inventory to the AADT in the 2012 Road Inventory used in the original study. The most recent vintage of AADT data in the 2018 Road Inventory was the year 2017. Emissions of NOx from 2012 were then scaled up or down for each road segment using the relative proportion of 2017 AADT / 2012 AADT. This method for scaling emissions implicitly assumes that the fleet mix and traffic conditions of 2012 was similar to that of 2017. While we acknowledge that this may introduce bias into the estimates, a complete re-run of the modelling methodology used in *Gately et al. (2017)* was beyond the scope of this study. New data from INRIX on vehicle speeds across the state of Massachusetts has recently become available, and MAPC intends to pursue a more comprehensive analysis of statewide vehicle pollutant emissions in the coming year.

For clarity and computational efficiency, we limited our analysis to roads with >500 vehicles per day.

A literature review identified the key distances from the roadway edge where concentrations of UFP are highest are with the first 50 meters, and within 150 meters (*Zhu et al., 2002; Zhou and Levy, 2007; Fuller et al., 2012*). Beyond 150 meters, concentrations of UFP rapidly approach background levels. To identify populations living within these distances from roads in the region, we created spatial buffer features from

0 - 50 meters and from 50 - 150 meters from the MassDOT Road Inventory road segments using ESRI's ArcGIS Pro software.

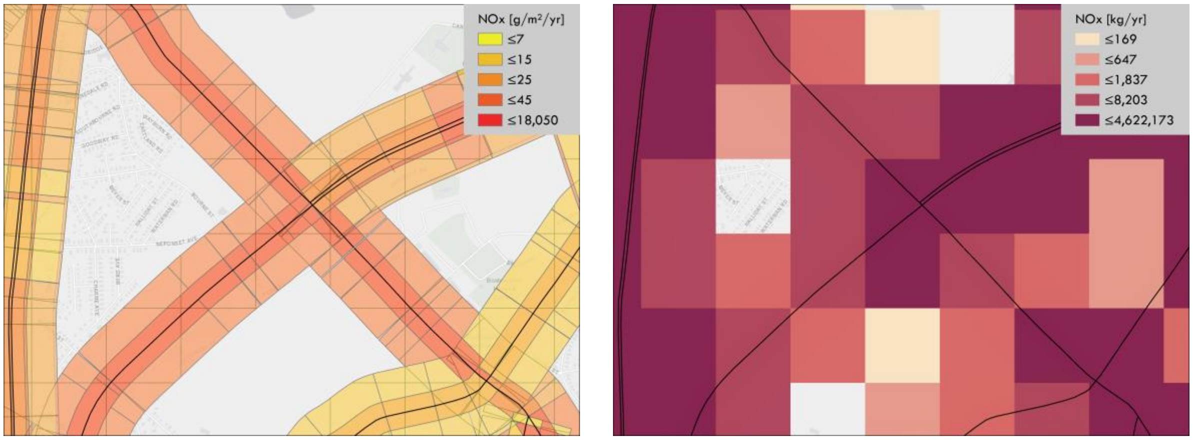
We distribute half of the annual NOx emissions from each road segment uniformly throughout the 0 – 50 meter buffer, and the other half of emissions uniformly throughout the 50-150 meter buffer. As the latter buffer comprises twice the area as the former, this amounts to a weighting of emissions for the inner buffer of double that of emissions in the outer buffer. This proportion reflects measurement data on the distribution of UFP concentrations as roadway distance increases. Zhu et al. (2002) showed that at a distance of 50 meters from a roadway edge, the concentration of UFP observed in the atmosphere declines to 50% of the maximum concentration observed at 0 meters from the roadway edge. At a distance of 150 meters, UFP concentrations are observed to approach background concentrations. Thus we assign the inner 50 meter buffer as having twice the potential pollution intensity as the outer 50 -150 meter buffer area. Figure 1 shows an example of the multiple buffer area geometries.



A B

Figure 1. Example of overlapping buffer areas surrounding major roads. A) shows the buffer polygons at 0-50 meters and 50-150 meters from the road centerlines. Emissions of NOx in kilograms per meter per year are shown for each road segment. Roadways with multiple parallel segments (i.e. divided highways) are given a set of buffers for each line segment. B) shows the buffer polygons with emissions assigned as grams of NOx per square meter per year, as described above.

Overlapping buffer polygons were then intersected, and emissions per square meter were summed for all overlapping geometries. The result is a set of polygons with values for pollution intensity in kg/m_2 of NOx, reflecting the summed emissions of all roads the were within the 50 or 150 meter buffers. These polygons are then intersected with the MAPC 250-meter gridded demographic population layer to summarize emissions intensity in each grid cell. Total emission intensity for each grid cell is calculated as the product of the areal extent of intersection of each buffer polygon and the NOx per square meter of emissions in that buffer polygon fragment (Figure 2).



A **B**

Figure 2. A) Example of 250-meter grid intersected with overlapping buffer areas. B) Sum of cumulative emissions intensity in grid cells.

We do not consider these raw values of NOx in kilograms per year to be the actual concentrations of NOx in the atmosphere of each grid cell, but rather to reflect the relative intensity of emissions on the roads whose buffers intersect with that grid cell. As such, we calculate the Pollution Proximity Index based on the quintiles of the distribution of these raw values. Grid cells with raw emissions intensities falling within each quintile are assigned the relevant PPI score of 1 through 5. Grid cells with no overlap of any of the road buffers are assigned a PPI of 0. The values of raw emissions intensity corresponding to the quintile breaks are shown in Table 1 below. We emphasize again that these values are not to be construed as representing actual concentrations of NOx or UFP in the atmosphere, but rather serve as a proxy for the potential intensity of exposure to these pollutants in the grid cells nearest to these roadways.

PPI	Raw Values NOx [kg/yr]
0	0
1	0 – 204.9
2	204.9 – 810.4
3	810.4 – 2,570.3
4	2,570.3 – 14,058.1
5	14,058.1 – 10,773,506.2

Table 1. PPI scores are assigned based on the quintile ranges of raw emissions intensity values for grid cells that overlap any of the emissions buffers. Grid cells that do not overlap any of the buffers were given a PPI score of 0.

For each grid cell at each PPI level, population was summed by race and ethnicity using data from the 2010 Census Block that were allocated to each grid cell using the method described below.

Allocation methodology for MA Statewide Household Allocation to 250 km grid

MAPC disaggregated the Census 2010 Block level household and population data by race and ethnicity to a 250-meter grid using point-level information on household location obtained from InfoGroup. This dataset provides latitude and longitude coordinates for the location of households. By intersecting this point-level data with the 2010 Census Block geographies and the 250-meter gridded shapefile, households and population from the Census Blocks by race and ethnicity were distributed into the 250-meter gridded cells as follows:

For InfoGroup household points that intersect census blocks with >0 households.

1. Info-Group housing units and household population associated with 2010 census block geography and 250 km grid id using a spatial join.
2. Info group point data aggregated for each grid-block pair to calculate share of households and household population for each grid with respect to each block.
3. Share of households and household population by race and age for each grid-block pair used to allocate households and household population from each block to the intersecting grid(s).
4. Households and household population by race and age allocated to each grid-block pair grouped by grid id and summed to get grid allocated households and household population.

For remaining Census blocks with >0 households but no InfoGroup points

1. Total of 7,285 census blocks with at least 1 household, or 1 population or 1 housing unit have no info-group points/housing units.
2. 6,418 blocks have intersecting residential land use as designated by the MassGIS land use layer. Allocation of households for these blocks to grids done based on residential land use share for each grid-block pair.
3. Residential land use shapefile intersected with census blocks and then with 250m grid. Each section of residential land use associated with grid and block id.
4. Area of residential land use calculated and summed to get residential land area by block.
5. Land use grouped by grid-block id pair similar to InfoGroup housing units (as described above) to get share of residential land use.
6. Share of residential land use used to allocate households and population in households to the grid-block pair.
7. Households and household population allocated to each grid-block pair grouped by grid id and summed to get grid allocated households and household population.

Allocation for these blocks was first checked for error and in cases where the census block had households and other attributes allocated but no development from aerial imagery - the neighboring blocks were allocated the blocks' attributes. Blocks having more than 100 households were analyzed manually and checked for error. The edited set of blocks were intersected with the grids and allocated on an area basis.

Additional Details of Emissions Inventory Methodology

The following schematic (Figure 3) shows the main data sources and analysis flow used to calculate emissions for each roadway segment in the MassDOT Roadway Inventory for the MAPC region.

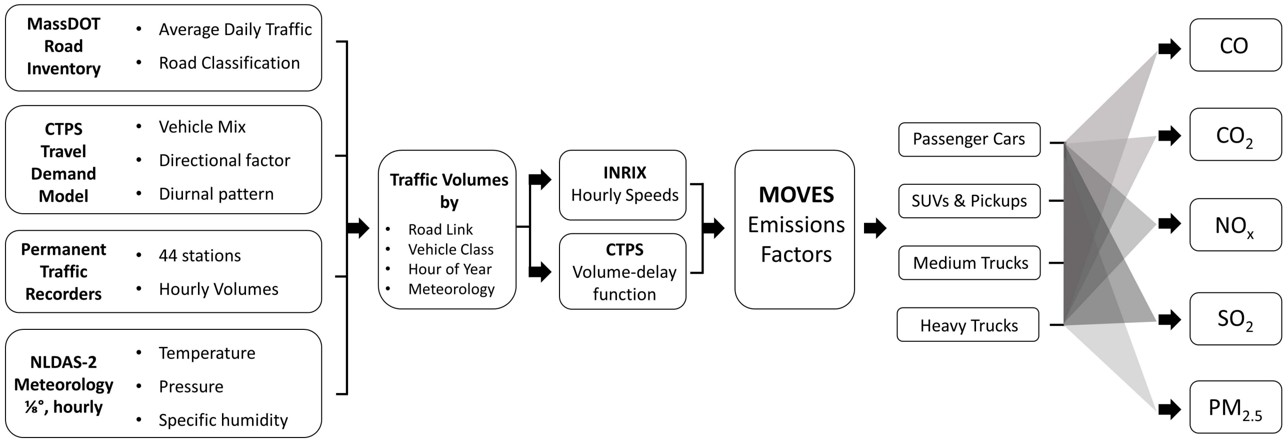


Figure 3. Data fusion methodology for emissions calculations. The four major input datasets are spatially merged to generate hourly traffic volumes for 5 vehicle types for each of the 280,424 road segments in the domain, coupled to the hourly ambient meteorology for each road. Traffic flow speeds are assigned to segments using INRIX data (where available), or else imputed using *Equation 1*. Emission factors for CO, CO_2, NO_2, NO_x, and $PM_{2.5}$ are calculated using MOVES2014a for all combinations of vehicle types, speed intervals, and temperature and humidity regimes, and then applied to the hourly traffic volumes of each vehicle class accordingly.

Hourly emissions of NOx were calculated for each road segment in the study region. These emissions were estimated by integrating data on average daily traffic counts (ADT) and hourly traffic volumes from MassDOT short-term and permanent count stations, data on the composition of the vehicle fleet at different time periods and on different roadway classes from the CTPS travel demand model, vehicle speed data from the transportation data provider INRIX, ambient meteorology from the NLDAS-2 reanalysis data product, and emissions factors from the EPA MOVES 2014a vehicle emissions model.

Emissions were calculated for the year 2012, for each hour of which (indexed by *h*) we estimated the flux of six pollutants (CO, SO_2, NO_2, NO_x, $PM_{2.5}$ and CO_2, indexed by *p*) emitted by vehicles on each of 280,424 road segments (indexed by *l*). Pollutant species are emitted by *v* types of vehicles, traveling at speeds that we discretize into *s* 5-mph intervals. Each road segment's hourly emission flux (**q***) is the product of the vehicle kilometers traveled (**k***) on it and an emission factor (**f***). Emissions factors are calculated for every combination of v and s, and are specified as a function of spatially and temporally varying temperature (**T**) and relative humidity (**H**):

(1)

The suite of emission factors that encompass the range of vehicle types, travel speeds, and meteorological conditions observed in our study domain was derived using the MOVES vehicle emission simulator. Our meteorological variables were obtained from the North American Land Data Assimilation System (NLDAS-2), which reports 0.125° gridded hourly temperature and specific humidity. MOVES was run for every county in our study area using vehicle fleet age distributions and fuel formulation distributions provided by the Boston MPO. Output from MOVES included emissions factors for each pollutant stratified by vehicle type, road type, fuel type, vehicle speed, ambient air temperature, and relative humidity.

The original emissions inventory was calculated for the year 2012. To scale these emission forward to the year 2017, MAPC compared the annual average daily traffic (AADT) for each road segment reported in the 2018 MassDOT Road Inventory to the AADT in the 2012 Road Inventory used in the original study. The most recent vintage of AADT data in the 2018 Road Inventory was the year 2017. Emissions estimates from 2012 were then scaled up or down for each road segment using the relative proportion of 2017 AADT / 2012 AADT.

The full details of how the emissions inventory was developed are available in [Gately et al., 2017](). A summary of key data sources is described below.

INRIX Traffic Speeds

The INRIX database of average hourly vehicle speed by segment is derived from thousands of mobile phone and vehicle GPS devices, which are then aggregated by INRIX to calculate average travel speed on over 60,000 road segments in our study domain at 5-minute intervals for the year 2012. The INRIX road network separates vehicle travel on each road segment by the direction of travel, to account for the variation in daily traffic patterns on roads that experience distinct directional patterns of traffic activity depending on the time of day. The georeferenced Massachusetts Road Inventory (MRI) road network forms the spatial basis for our emissions model but lacks a crosswalk to unique road segment IDs in the INRIX database. We therefore merged the two datasets in a GIS using a proximity-based spatial join. Manual validation of road segments in the merged dataset was undertaken by Boston MPO CTPS staff for major roadways. This procedure was completed for all road segments by the authors of Gately et al., 2017.

Traffic Volumes

The MRI shapefile contains estimates of ADT volumes for each road segment, as well as the number of lanes in the segment and the roadway functional class as defined by FHWA. ADT estimates were missing for some road segments, mostly local roads. However, estimates of total vehicle-miles travelled on local roads in the Boston Urbanized Area are available from the FHWA Highway Statistics Series Table HM-71 (FHWA 1980-2012). We used these aggregate totals to assign an average ADT to each Boston Urbanized Area local road with missing observations in the MRI. The procedure is as follows. We first subtracted total VMT from the local roads that do have a reported ADT from the total VMT in Table HM-71, and then divide the remaining VMT by the total length of all the local roads that are missing VMT in the Urbanized Area, and then divide again by 366 days, to get an average daily traffic volume for all local roads.

CTPS Travel Demand Model

We joined output from the CTPS highway assignment travel demand model (TDM) year 2012 base run to the MRI attribute table. The highway assignment model is implemented using the TransCAD traffic modeling software and is continuously updated and maintained by CTPS as part of its mission to model and forecast local and regional transportation system demand. The model output we use is the final component of a larger model set which follows the well-known four-step transportation modeling procedure:

1. Trip generation: estimation of the number of daily trips in Eastern Massachusetts, based on travel survey data, vehicle fleet data, and demographic information. Freight trips are estimated separately from non-freight trips. Trip generation is estimated for four time periods: AM peak (6am – 10am), Midday (10am – 3pm), PM peak (3pm – 7pm), and Nighttime (7pm – 6am).
2. Trip distribution: using the same data on land-use and travel patterns, the location of trip origins and destinations is estimated and aggregated to traffic analysis zones (TAZs). Freight trips are again assigned separately from non-freight trips. Data output from the distribution step are

matrices of the freight and non-freight trips by origin-destination (O-D) pairs for all TAZs in the domain.

3. Mode choice: assignment of trips in the O-D matrices created in step 2 to different travel modes. Non-freight trips are divided amongst passenger cars, SUVs and pickups, public transit (buses and light and heavy rail), bicycling, and walking. Freight trips are divided amongst large and medium class trucks (i.e. combination tractor trailer and single-unit trucks, respectively).
4. Route assignment: allocation of trips to individual segments of the relevant transportation network (roads, heavy/light rail, cycling infrastructure, etc.) using optimization techniques to minimize total travel delay while satisfying constraints of origin-destination travel demands and link capacities. Additional details of the CTPS model are available at: http://www.ctps.org/data/html/studies/other/Travel_Modeling_101.htm

We utilize the TDM model output consisting of estimated vehicle travel for four representative periods of a weekday (): AM peak (6am-10am), mid-day (10am-3pm), PM peak (3pm-7pm), and night (7pm-6am), stratified by vehicle type and road segment direction. The model also provides information on key characteristics of each segment: its capacity () and the coefficients (and) of its link-performance function, which follows a modified Bureau of Public Roads volume-delay formulation with free flow speed :

(Equation 1)

For roads without records in the INRIX database, the above equation was used to impute vehicle speeds (. CTPS has developed customized coefficients (and) for *Eqn. S1* that vary according to road functional class and location within the urban area. These coefficients have been calibrated against traffic counts and 'floating car' data on vehicle speeds at multiple locations across the region. Generally, the values for range from 0.8 to 1.25, and the values for range from 4 to 5.5.

Simulated traffic volumes are not used directly, as the output of the model does not cover all MRI road segments, and many local roads are excluded to keep the assignment problem computationally tractable. For consistency, MRI AADT for each link is taken as the control total, and CTPS model output used to disaggregate this volume by time of day, travel direction, and vehicle type. Vehicle types are stratified into five aggregate classes: passenger cars (gasoline powered), passenger trucks (SUVs and pickups, both gasoline powered), medium-size trucks (gasoline-powered), medium-size trucks (diesel-powered), and heavy trucks (diesel-powered). We aggregate buses into the "heavy truck" vehicle class. Modeled volumes by vehicle class in each of the four time periods were divided by the modeled total daily link volume (by direction), generating a vector of shares by vehicle class, direction, and time of day that were then used to split MRI AADT. MRI road segments missing from the CTPS model were assigned the characteristics of the nearest modeled road segment belonging to the same functional class.

Hourly Time Structure

To temporally disaggregate MRI AADT we use a large dataset of hourly traffic counts from 62 permanent traffic recorders (PTRs) across the study region (Massachusetts Department of Transportation 2014). Counts are obtained from inductive loop sensors embedded in the roadway surface that continuously monitor traffic throughout the year. For each PTR station we divide the vehicle counts at each hour by total annual count for the year 2012 to calculate an hourly share. We assign each MRI roadway link the hourly traffic profile of the closest PTR. Road segments' hourly traffic volumes are then estimated by multiplying AADT by 366 days and the hourly share.

INRIX Vehicle Speed Assignment

All MRI links matching INRIX road segments were assigned the INRIX mean hourly speed for each hour of the year. Computational tractability necessitated aggregation of the raw 5-min INRIX speeds to an hourly time step. For a link belonging to a particular roadway functional class, free-flow speed was imputed as the mean of the 85th percentile of speeds calculated for all roads in that functional class in the INRIX database. The resulting values of were assigned to MRI segments missing observations in the INRIX database.

MOVES Emission Factors

Emissions factors were calculated using the EPA's Motor Vehicle Emissions Simulator (MOVES2014a). MOVES contains default values for many of the parameters used to determine emission rates. EPA strongly recommends that users include as much local data as possible when running the model, so as to minimize biases due to mismatch between the parameter defaults and actual local parameter values. MOVES' spatial resolution is limited to the county scale in "inventory" mode, but finer resolution estimates of emissions can be calculated in "emission factor" mode, which produces an output table of grams of pollutant emissions per vehicle kilometer travelled (VKT) for a range of vehicles and fuel types. We generate emissions factors for CO_2 as well as four other air pollutants (CO, NO_2, NO_x, and $PM_{2.5}$).

Emissions factors are highly sensitive to the specification of atmospheric conditions, vehicle make, model, age and fuel, and to the speed of travel, with MOVES requiring inputs characterizing all of these variables for the study domain. CTPS, in the course of modeling and forecasting the air quality impacts of regional transportation, has developed a set of dedicated input parameters specific to Eastern Massachusetts. We use data on vehicle fleet composition (vehicle type and age) derived by CTPS from vehicle registration data obtained from the State Registry of Motor Vehicles. CTPS also provided us with data on the fuel formulation for motor fuels sold in Middlesex County, MA. We augment these custom inputs with a table that covers the full range of meteorological conditions in the year 2012, obtained from NLDAS-2 reanalysis at 0.125° resolution of gridded hourly surface temperatures at 2 meter elevation and specific humidity.

References

Fuller CH, Brugge D, Williams P, Mittleman M, Durant JL, Spengler JD (2012) Estimation of ultrafine particle concentrations at near-highway residences using data from local and central monitors, Atmos. Environ. 57 (1994) 257–265.

Gately CK, Hutyra LR, Peterson S, Sue Wing I (2017) Urban emissions hotspots: quantifying vehicle congestion and air pollution using mobile phone GPS data. Environ. Pollut. 229, 496e504. https://doi.org/10.1016/j.envpol.2017.05.091.

Zhou Y and Levy, JI (2007) Factors influencing the spatial extent of mobile source air pollution impacts: a meta-analysis. BMC Public Health 7, 89. https://doi.org/10.1186/1471-2458-7-89

Zhu, Y, Hinds WC, Kim S, Sioutas C (2002) Concentration and Size Distribution of Ultrafine Particles Near a Major Highway, Journal of the Air & Waste Management Association, 52:9, 1032-1042. https://doi.org/10.1080/10473289.2002.10470842