January 2019

CTPP Status Report

U.S. Department of Transportation
Federal Highway Administration
AASHTO Standing Committee on Planning

TRB Census Subcommittee
Bureau of Transportation Statistics
Federal Transit Administration

Census Transportation Planning Product (CTPP) Highlights

Penelope Weinberger, AASHTO, *pweinberger@aashto.org*

New CTPP Mapping and Analysis Tool

The new CTPP mapping and analysis tool is now alive! The new interactive mapping tool is designed to help data users visualize CTPP data and perform thematic analysis easily and quickly. Check it out on our website: *http://data5.ctpp.transportation.org/ctpp/Browse/browsetables.aspx*.

Watch the CTPP tutorials to learn more about how to use this tool and its capabilities: *https://www.youtube.com/playlist?list=PLVnwe7nwy_Vge95BH6TaNpbBG6P39bLCZ*.

Small Geography Data Reporting Webinar

The 2020 Decennial Census marks a new era for CTPP data. All CTPP data released after the 2020 Census will be reported by block group versus Transportation Analysis Zones (TAZ). On October 24, 2018, a webinar hosted by the CTPP Oversight Board and Census Bureau Geography Division covered the 2020 Census Participant Statistical Areas Program (PSAP) block group and tract delineation criteria.

The webinar provided an update on the process initiated to adapt to this geography change, discussions on how the new PSAP criteria will better enable the alignment of block group and TAZ boundaries to sustain the value of CTPP data for modeling, transportation planning and analysis, how to get involved in PSAP, and who is participating in your region.

Webinar recordings and presentations are now available on CTPP website: *https://ctpp.transportation.org/policy-change-on-small-geography/psap-update-on-small-geography-delineation-criteria-oct-24-2018/*.

Welcome New CTPP Oversight Board Members

Mark Grainer, New York State Department of Transportation

Paul Agnello, Fredericksburg Area Metropolitan Planning Organization

Ron Chicka, Metropolitan Interstate Council

The CTPP Oversight Board Roster is available from CTPP website: *https://ctpp.transportation.org/*.

The 2012 – 2016 CTPP Data is Coming Soon!

The new CTPP tabulation production in the Census Bureau is on schedule: the 2012 – 2016 CTPP data will be available in December 2018, and released to public on CTPP website in spring 2019. Stay tuned for the new CTPP data!

AASHTO CTPP New Webpage on Relevant Conference Papers

AASHTO CTPP Program recently launched a new webpage listing CTPP-relevant conference papers at: https://ctpp.transportation.org/census-conference-commissioned-papers/.

If you are aware of any other relevant papers, please contact us through pweinberger@aashto.org.

The Use of CTPP and Census Data in Evaluating Cell Phone Derived Travel Patterns

Kimon Proussaloglou, Cambridge Systematics, Inc., kproussaloglou@camsys.com

Background and Objective

The transportation community has a great interest in how different types of locational data can be analyzed to infer travel patterns in a region to support planning applications that are traditionally based on survey data and regional models. A key component of the ongoing discussion in our field is the need to compare and contrast new forms of data with traditional household surveys, Census-based products such as the CTPP, and regional model outputs. The National Cooperative Highway Research Program (NCHRP) Report 868: Cell Phone Location Data for Travel Behavior Analysis, presents guidelines for transportation planners and travel modelers on how to:

1. Evaluate the extent to which cell phone location data accurately depict travel.
2. Identify whether and how these data can be used to improve our understanding of travel characteristics and our ability to model travel patterns and behavior more effectively.
3. Support practitioners' evaluation of the strengths and weaknesses of anonymized "call detail record" (CDR) locations from cell phone data.

This guidebook is intended for transportation practitioners and agency staff interested in new methods of capturing travel data from cell phones. This emerging field is subject to complexities related to acquiring and analyzing locational data while maintaining privacy in a complex legal and practical framework.

The emergence of these data constitutes a significant opportunity for change in the travel modeling community, with access to detail and volume not previously available. A better understanding of the strengths and weaknesses of locational data is an important step in this direction. Research is needed to explore and evaluate methods used for processing cell phone location data to generate travel behavior information and provide guidelines for the use of these data by transportation planning practitioners.

The case study used travel in the Boston region to compare and contrast traditional travel survey data, regional models, Census data, and cell phone derived CDR data describing regional travel. The questions that were addressed included the following:

- What are the best options for using cell phone CDR data to derive different estimates of travel?
- What are the strengths and weaknesses of cell phone CDR data and how can they best support travel behavior analysis and policy decisionmaking?
- How can cell phone CDR data be used to enhance access to information on travel behavior characteristics necessary for effective model applications?
- What tools and techniques are available for collecting and analyzing cell phone CDR data?
- How can travel modelers overcome practical and legal problems associated with CDR data acquisition, and how can this process respond to privacy requirements?

The Concept of "Ground Truth"

A key issue, which does not have a clear or straightforward answer, is which of the data and modeling sources constitutes "ground truth." In the case study, we compare CDR-derived results with household surveys, Census estimates of commute travel, and well-understood model outputs. However, we need to recognize the different nature of each data source, and the fact that each data source reflects a sample of observations that has its own strengths, weaknesses, assumptions, and errors embedded in it.

A related question includes the **assumptions made and inferences drawn** when analyzing each of these data sources. Weaknesses in analysis approaches reflect different assumptions used to infer activities, travel purposes, travel destinations, modes of travel, and times of day of travel from the underlying data:

- **Weighting of household surveys**. A small sample is collected and weighted based on the regional distribution of socioeconomic characteristics. The implicit assumptions are that the determinants of travel are properly reflected in the market segments defined in the sampling plan, and that enough observations are collected in each cell to properly assess travel within each segment.
- **Weaknesses in model development**. In both trip- and activity-based models, errors are likely to propagate throughout the model components. These errors may reflect limited data for certain market segments, errors or omissions of important variables in model specification, and linkages among models that are not properly reflected in the analysis.
- **Census data.** Journey-to-Work travel flows probably offer the strongest "ground truth" source of data for the **daily commute** market. However, both CTPP and American Community Survey (ACS) data also come from a sample of a region's households. Weaknesses of these data include absenteeism, the reporting of the primary work location only, and lower sampling rates in smaller geographies.
- **CDR data assumptions**. This new source of data benefits from a much larger sample size, the ability to observe the same cell phone device over a long period of time, and the ability to make inferences about a user's activities using repeated observations. At the same time, the value of CDR data also is constrained by the following:
  - Passive or active use of the phone is needed to record travel.
  - There is uncertainty in "stay locations" that are inferred by the analyst.
  - There is limited differentiation among nonwork travel purposes.
  - Trips by members of the same household are not linked.
  - Lack of socioeconomic information prevents analysis by market segment.

In assessing the strengths and weaknesses of traditional and emerging data sources, it is critical to keep in perspective the nuances of each data source and corresponding analysis methodologies, and recognize the uncertainty due to the lack of absolute and definitive "ground truth" estimates.

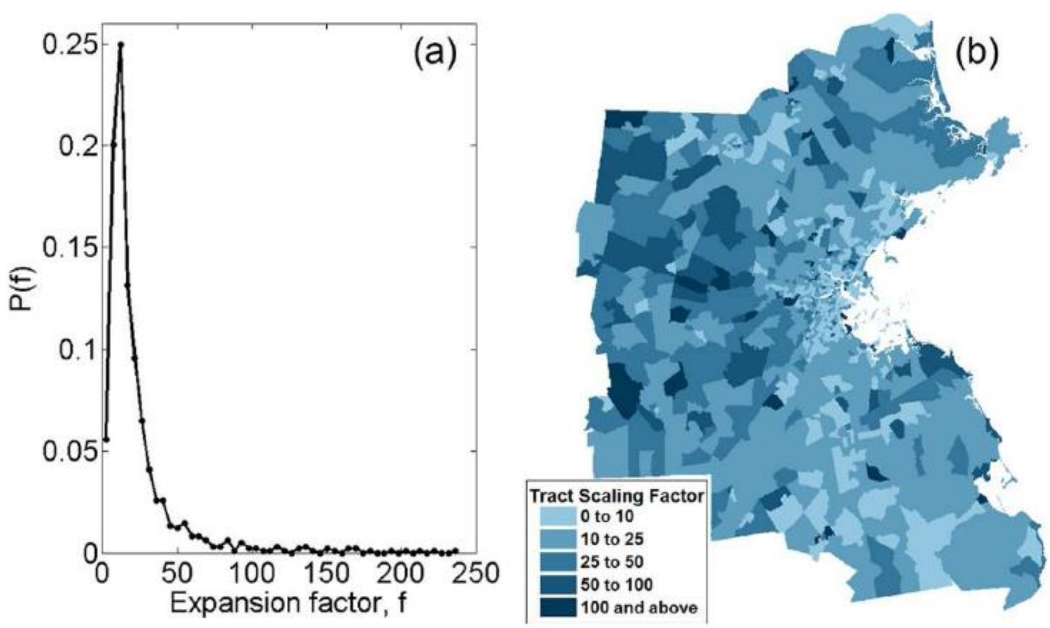
Sample Weighting

The analysis of the CDR data and the weighting of the CDR sample was conducted by the MIT research team led by Professor Marta Gonzalez and Dr. Shan Jiang. In weighting the cell phone CDR data used in this study, the research team filtered out observations with very few visits over the two-month observation period to designated home stay locations. This filter serves the additional purpose of ensuring, with a reasonable degree of certainty, that the designated stay is the user's home—a key assumption in expanding users to the population.

To expand the filtered sample of cell phone users to the total population of the study region, the inferred home stays were aggregated to Census tracts in the Boston region. An expansion factor was calculated for each tract as the ratio of the 2010 Census population and the number of residents identified in the CDR data. There were a few Census tracts with fewer than 10 CDR residents, where the expansion factor was set to zero to ensure that we did not overweight users who may not be representative of a Census tract.

Figure 1 shows the distribution of the expansion factor values. Figure 1 also illustrates the spatial distribution of the expansion factors. Our analysis suggests that the tracts in the suburban western portion of the study area tend to be more heavily weighted than the core central area, which is better represented in the sample (Source: Alexander, L., S. Jiang, M. Murga, and M. C. Gonzalez (2015). Origin-destination trips by purpose and time of day were inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies*, 58, Part B: 240-250).

The availability and analysis of cell phone CDR data for a period greater than two months would most likely require lower expansion factors and result in a better spatial distribution of users. The analogy with traditional surveys is an increase in the sample size and a greater focus on geographic and socioeconomic market segments that improves the representativeness of the sample. There are some other interesting analogies that are worth noting when we contrast the cell phone CDR expansion factors to the sampling weights in a traditional household survey.



Source: Alexander et al., 2015.

Figure 1. Expansion factors for Census tracts

First, the motivation behind the sampling weights in traditional surveys is to reflect differences in making contact with travelers and their willingness to participate in a survey. Both of these steps in traditional surveys require a correction through the development of sample weights that expand the sample to be more representative of the regional population.

- Survey sampling weights reflect and correct for the under- and over-representation of certain geographic and socioeconomic market segments in the survey. Implicit in sample weighting is the need to **adjust the representation** by members of these market segments to avoid a model that under- or over-predicts travel in the region.
- In the case of the cell phone CDR data, the expansion factors are similar to the sampling weights and reflect the **market penetration** and **use of cell phones** during a typical day. Younger, more educated, and more technology savvy cell phone users are more likely to provide traces of their daily routines through their increased use of calls, text messages, and internet data access when visiting websites or receiving passive signals from apps.

Second, the **cell phone expansion weights are smaller** in magnitude compared to the sampling weights of a traditional household survey. This is an expected result since a sampling rate of one percent in a typical traditional survey would correspond to an average expansion factor of about 100 with higher values for hard-to-reach geographic and socioeconomic market segments.

Third, one expects to find **differences due to geography and socioeconomics** both for traditional surveys and for cell phone use.

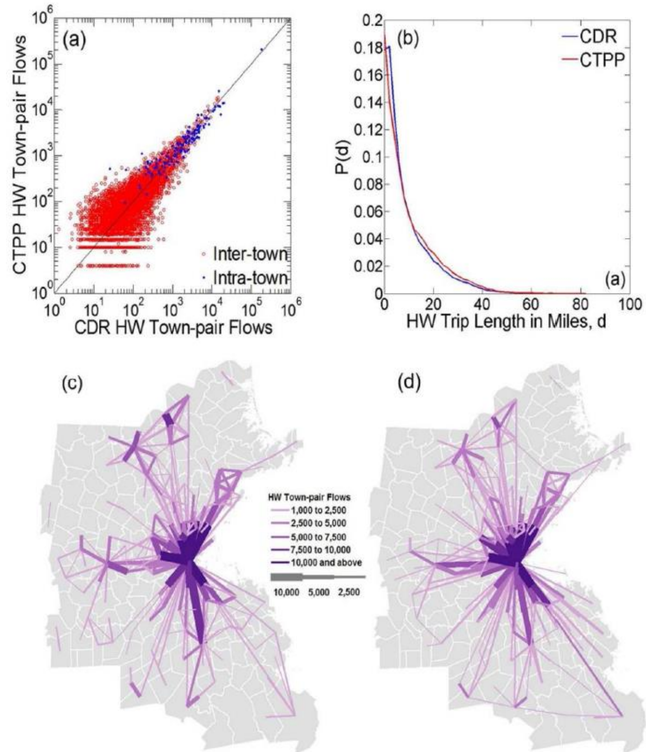
- In **traditional surveys**, large households are harder to reach, as are households with younger members and those that rely less on phone land lines, have lower incomes, or live in urban areas. These hard-to-reach households are expected to have lower response rates compared to the rest of the sample.
- In contrast, the **market penetration** is higher and the **usage of cell phones** is more extensive among younger cohorts of the population and some of these traditionally hard-to-reach segments. As a result, data and analyses that are based on the cell phone sample are likely to better reflect the travel behavior and habits of some of the hard-to-reach segments in traditional household surveys.

Validation

The accurate extraction of users' stays and the proper expansion to the regional population is critical to trip generation and estimates of total travel in a region. Due to the regularity of human behavior (González et al., 2008; Song et al., 2010a and 2010b; and Jiang et al., 2013), users' home stay and, where applicable, their work stay locations were inferred from the CDR data.

A comparison of **home locations** at the town level was made using the 2010 Census data and the raw and expanded CDR data. Since the Census tract-level population was used to expand the data, the number of residents in each town is almost identical to the estimates from the expanded CDR data as expected (Figure 2).

A second comparison focused on **work locations** aggregated at the town level. The distribution of raw workplace data is fairly consistent with the 2006 – 2010 CTPP. The data slope is close to one and the sample expansion method adjusts well for the differences in magnitude across towns. This strong correlation is noteworthy, considering that each user's home and work locations were expanded based on their home location only.



Source: Alexander et al., 2015.

Figure 2. Work trips—travel flows, trip length, and origin-destination trip patterns

A third comparison focused on comparisons between the CDR and the CTPP datasets using town-pair **travel flows (**Figure 2**)**. The correlation between the home-to-work flows for all intra-town and inter-town pairs was 0.99 and 0.95, respectively. The results were stronger for town pairs with many trips with weaker results for pairs with fewer than around 500 daily trips, reflecting the scarcity of data for the smaller markets.

Finally, the trip-length distributions of home-to-work flows derived from the CDR data and the 2006 – 2010 CTPP data also suggest a good match of home to work flows at a town level.

References

Alexander, L., S. Jiang, M. Murga, and M. C. González (2015). Origin-destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies*, 58, Part B: 240-250. ISSN 0968-090X. doi: http://dx.doi.org/10.1016/j.trc.2015.02.018.

González, M. C., C. A. Hidalgo, and A. L. Barabási (2008). Understanding individual human mobility patterns. *Nature*, 453(7196):779–782. URL http://www.ncbi.nlm.nih.gov/pubmed/18528393.

Song, C., T. Koren, P. Wang, and A. L. Barabási (2010a). Modelling the scaling properties of human mobility. *Nature Physics*, 6(10):818–823, 2010a.

Song, C., Z. Qu, N. Blumm, and A. L. Barabási (2010b). Limits of predictability in human mobility. *Science*, 327(5968):1018–1021.

Jiang, S., G. A. Fiore, Y. Yang, J. Ferreira, E. Frazzoli, and M. C. González. A review of urban computing for mobile phone traces: current methods, challenges and opportunities. In *Proceedings of the Second ACM SIGKDD International Workshop on Urban Computing*, page 2. ACM, 2013.

Using CTPP Data for the 2030 District in Philadelphia

Sarah Reinheimer, Duke University, sarah.reinheimer@duke.edu

This article is a synthesis of a 2018 Duke Master's Project (MP) by the author, where CTPP data were used to help a public-private partnership called the 2030 District in Philadelphia measure transportation emissions within Philadelphia. Although this MP focuses on one American city, it is meant to be applicable to American cities across the country.

Cities are only two percent of the world's landmass. Despite this, they "account for more than 70 percent of global CO2 emissions" (C40 Cities, n.d.). Cities also are anticipated to grow, and 66 percent of the world's population are projected to live in urban areas by 2050 (UN DESA, 2014).

Unfortunately, traffic is a key component of city life that many of us are accustomed to, along with the resulting growing transportation emissions. While cities and their subdistricts have started to establish standardized methodologies to measure other sources of CO2 emissions (such as buildings), they have struggled on how to measure transportation in a way that is standardized across regions.

Why Does this Matter?

Measuring transportation GHG emissions is important because, while in the past, policymakers have primarily focused on electric power generation and industry to limit the growth of GHG emissions, transportation emissions today account for 27 percent of U.S. GHG emissions (EPA, 2015). Transportation also is now the fastest growing source of GHG emissions, and there are one-third more vehicles on the road than there were in 1990 (Sorrel, 2016).

The Project and its Importance

This project has three parts, and this article will focus on the first part:

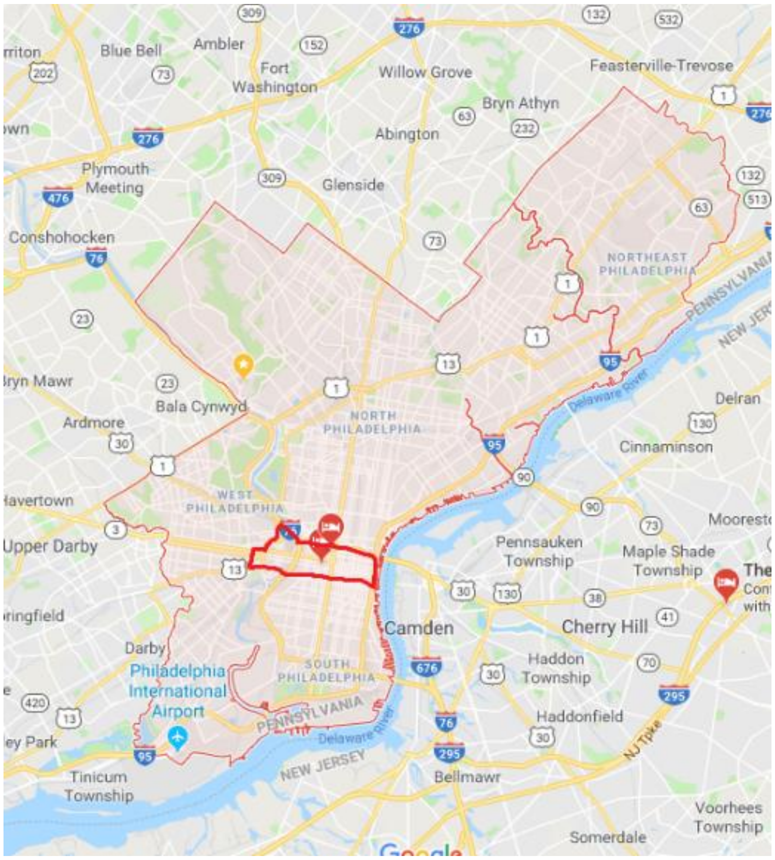
1. Developing a transportation greenhouse gas (GHG) emissions baseline for the Philadelphia 2030 District.
2. Creating a proposed survey for the District to better track transportation GHG emissions moving forward.
3. Creating case studies comparing the 2030 Districts already measuring their transportation emissions.

It is important to establish a transportation GHG emissions baseline for several reasons. Not only would it enable the Philadelphia 2030 District to accurately measure progress at reducing its transportation emissions by 50 percent by 2030, but it also would serve as an example for other 2030 Districts, and add to the literature for standardizing a method all Districts can use, regardless of their size or location in the U.S. This will enable Districts to better measure progress against one another, contribute to the best-practice literature, and determine which policies have an impact.

Methodology for Philadelphia 2030

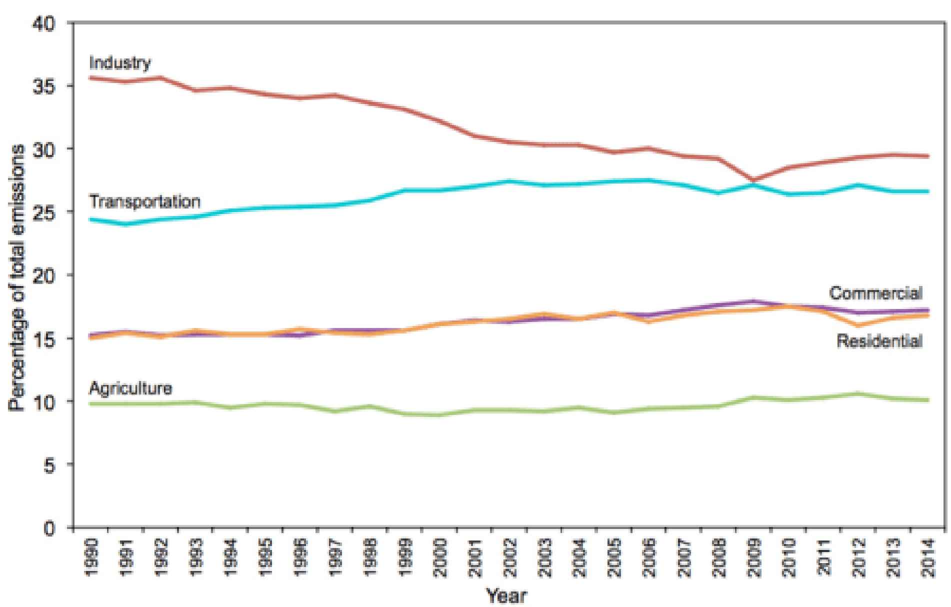
There were five different methodologies considered throughout the process of determining the best way to develop a baseline:

1. The Delaware Valley Regional Planning Commission (DVRPC) 2012 – 2013 Household Travel Survey (HTS).
2. Develop and Distribute a Survey.
3. Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics (LODES) data with the American Community Survey (ACS).
4. On the Map (OTM) with ACS.
5. CTPP with the DVRPC Distance Matrix Data.



Source: Bartolotta, 2018.
Note: The red dots indicate buildings committed to the 2030 District.

Figure 3. Map of Philadelphia with District 2030 drawn in red surrounding commuter region



Source: Sorrell, 2016.

Figure 4. Percentages of total U.S. GHG emissions by sector from 1990 to 2014

Based on discussions with Robert Graff the Office of Energy and Climate Change Initiatives from the DVRPC, methodology five, CTPP with the DVRPC Distance Matrix Data, was used due to data availability, age of data, replicability/fit, the project timeline, and risk of error. Specific reasons methodology five was adopted, include:

- First, the current CTPP dataset is the 2006 – 2010 samples, which is more than 5 years old. The next release for 2012 – 2016 CTPP is not scheduled until early 2019 (AASHTO-CTPP, 2017). While an initial concern was that the CTPP data are a little outdated and, therefore, would not provide enough of an accurate baseline, Mr. Graff suggested that transportation and land use patterns have not changed in Philadelphia, and thus the CTPP data would still be largely accurate.
- Second, inputting and updating the results in the Excel calculator tool developed for this project would not be hard. The biggest advantage of the CTPP is that it provides modes of commuting flows between any two TAZs. Using census TAZs versus census tracts/blocks also presented an advantage, because the majority of metropolitan planning organizations (MPO) and planning agencies use TAZs as a basis for their modeling process. Therefore, it would be easier for other peer MPOs to implement the methodology. DVRPC also had developed and could provide a TAZ-to-TAZ matrix with commuting distances.

As a result, using the A302103 Table from the 2006 – 2010 CTPP dataset and the TAZ-to-TAZ commuting distance matrix, a GHG emissions baseline calculator was developed; the next section discusses it in more detail.

Philadelphia 2030 Baseline Result

Through the use of methodology five in an Excel Workbook Baseline Calculator, it was found that the Philadelphia 2030 District had a transportation GHG emissions baseline of 9.4 kg CO2 per commuter per day. Additional information on the methodology, data assumptions, and data sources can be found in the full report.

Below, Figure 5 shows the mode split in the 2030 District.

Limitations

The CTPP data enabled the creation of a transportations baseline for the 2030 District, which was the biggest challenge they had faced to date. However, there are some weaknesses to this data, for example, since the CTPP is based off the ACS, multimodality is lost, along with only one work location being measured. The spreadsheet matrix measuring distance is measured in highway distance, which means it reflects auto travel, rather than transit. In addition, the spreadsheet matrix is from 2015, while most of the other data is from 2010. However, as there have been no major infrastructure projects in the ensuring time period, this should not fundamentally alter the calculations.

As a result, the recommendation was made that Districts and areas within cities use this data methodology to start to measure transportation emissions; and once they are more established, to pursue a survey methodology.

Comparison

As part of the research, Philadelphia's GHG emissions were compared relative to the other Districts; the results are shown in Figure 6. A challenge to the below data is that each District used different methodologies to measure their transportation emissions.

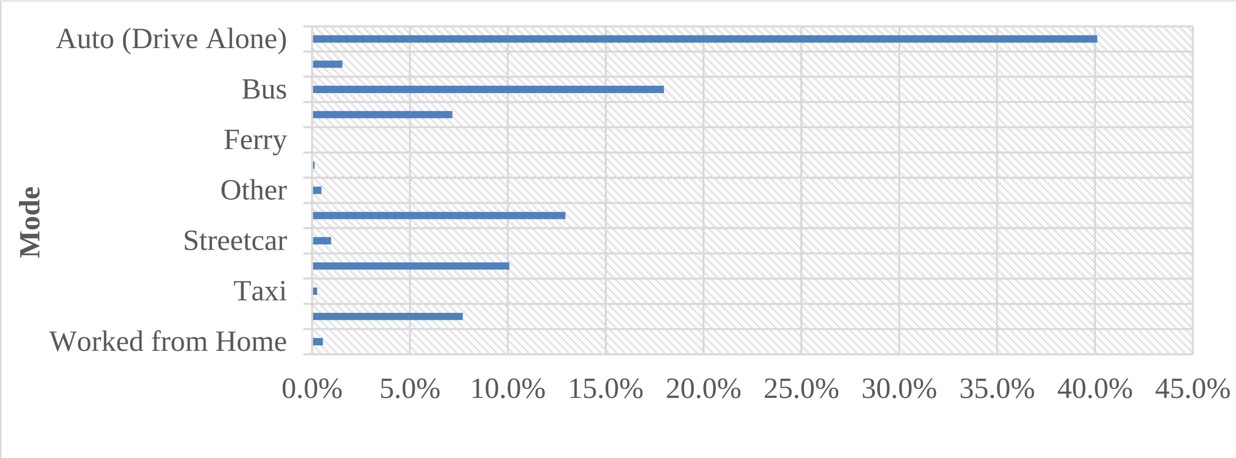


Figure 5. Mode split in district, 2030

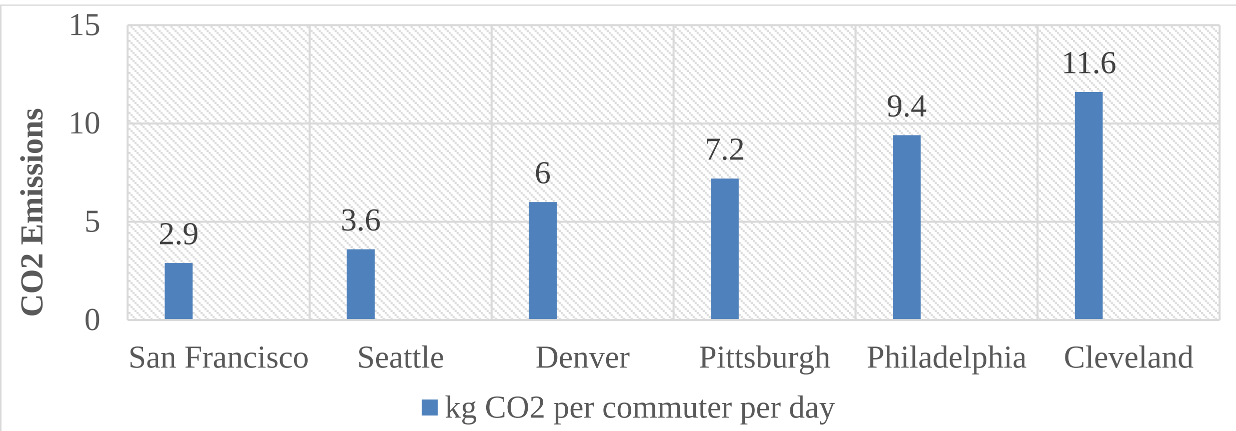


Figure 6. City CO2 baseline transportation rankings

Design and Replicability

The methodology used to establish the 2030 Philadelphia baseline is a methodology that can be replicated moving forward. First, the Excel Workbook Calculator is easy and inexpensive to use. Most importantly, its inputs are easy to change and are customizable to the location. As MPOs are already required to collect emission factors because of the CAA, this data is already available.

Finally, while the baseline calculator is an important first step, it is only a first step. If coupled with a survey methodology, Districts would be able to see the current travel habits of their 2030 District commuters, as well as the nuances of multimodal travel, which is not available through the CTPP data. In addition, survey questions would allow the District to understand some of the rationale behind transit choices, as well as options commuters have.

Interestingly, it was found that many organizations are unaware of the best data to use because they are not always closely connected to the local government organizations who know the best data products for their needs. As a result, nonprofits or other organizations looking to measure their emissions are recommended to focus on collaborating with local MPOs, developing strong relationships and partnerships, standardizing communication and survey methodologies, and—in the long term—developing a centralized approach to obtaining information on the various 2030 Districts, as well as deciding if the 2030 District focus is on adding new buildings or decreasing emissions.

CTPP Contact List

Email: CTPPSupport@camsys.com
CTPP 2006-2010 Data: http://ctpp.transportation.org/Pages/5-Year-Data.aspx
CTPP website: http://www.fhwa.dot.gov/planning/census_issues/ctpp/
FHWA website for Census issues: http://www.fhwa.dot.gov/planning/census_issues
AASHTO website for CTPP: http://ctpp.transportation.org
1990 and 2000 CTPP data downloadable via Transtats: http://transtats.bts.gov/
TRB Subcommittee on census data: http://www.trbcensus.com

AASHTO
Penelope Weinberger
Phone: (202) 624-3556
Email: pweinberger@aashto.org

Jessie Jones, ARDOT
Chair, CTPP Oversight Board
Phone: (501) 569-2201
Email: Jessie.Jones@ahtd.ar.gov

Guy Rousseau, Atlanta Regional Commission
Vice Chair, CTPP Oversight Board
Phone: (404) 463-3274
Email: GRousseau@atlantaregional.com

U.S. Census Bureau: Social, Economic and Housing Statistics Division
Brian McKenzie
Phone: (301) 763-6532
Email: brian.mckenzie@census.gov

Federal Transit Administration (FTA)
Ken Cervenka
Phone: (202) 493-0512
Email: ken.cervenka@dot.gov

Bureau of Transportation Statistics (BTS)
Clara Reschovsky
TRB Census Subcommittee Co-Chair
Phone: (202) 366-2857
Email: clara.reschovsky@dot.gov

Federal Highway Administration (FHWA)
Joseph Hausman
Phone: (202) 366-9629
Email: Joseph.Hausman@dot.gov

TRB Committees
Stacey Bricka
Consultant
Chair, TRB Urban Data Committee
Email: Sgbricka@gmail.com

Mara Kaminowitz
TRB Census Subcommittee co-Chair
Phone: (410) 732-0500
Email: mkaminowitz@baltometro.org

CTPP Technical Support
Jingjing Zang
Phone: (301) 347-9100
Email: CTPPSupport@camsys.com

CTPP Listserv

The CTPP Listserv serves as a web-forum for posting questions, and sharing information on Census and ACS. Currently, more than 700 users are subscribed to the listserv. To subscribe, please register by completing a form posted at: http://www.chrispy.net/mailman/listinfo/ctpp-news.

On the form, you can indicate if you want emails to be batched in a daily digest. The website also includes an archive of past emails posted to the listserv.