



CTPP Status Report



May 2013

U.S. Department of Transportation
Federal Highway Administration (FHWA)
Bureau of Transportation Statistics (BTS)
Federal Transit Administration (FTA)



AASHTO Standing Committee on Planning
TRB Census Subcommittee

Census Transportation Planning Products (CTPP) AASHTO Update

Penelope Weinberger, AASHTO,

Pweinberger@ashto.org

2013 marks an exciting year for the CTPP community. With the CTPP 2006 – 2010 data delivered to us in May, and the change of the CTPP program from an ad hoc program to an ongoing technical services program, we are looking forward to changes and challenges! The CTPP Oversight Board met on February 26, 2013 via web conferencing and discussions included:

- 2006-2010 CTPP: Some CTPP data for the first three states (Alabama, Rhode Island, and District of Columbia) was delivered from the Census Bureau to the data access software vendor. The complete data delivery from the Census Bureau is expected by May 2013 and the data release to public with the data access software is scheduled for August 2013.
- CTPP New Directions Task Force: The resolution to establish the CTPP on-going technical service program was approved by AASHTO Executive Board meeting in November 2012. The New Directions Task Force convened to plan the future on-going program, starting with a revised charter, a

American FactFinder2 (AFF) Self Tutorial

Because of recent changes to AFF, revised tutorials are now available at:

http://www.fhwa.dot.gov/planning/census_issues/ctpp/training/af2sf1.cfm

American FactFinder2 Equal Employment Opportunity (EEO) Self Tutorial is available at:

http://www.fhwa.dot.gov/planning/census_issues/ctpp/training/af2eeo.cfm

new mission and a first look at future budgets and funding requests. The task force is heartened to hear the program inspires confidence, and we know the new smaller on-going program will be well supported by states and MPOs. FHWA has informally approved a waiver of local match when using SPR funds for future CTPP efforts.

- E-Learning Modules: CTPP has published five e-learning modules and they can be accessed at: <http://ctpp.transportation.org/Pages/elearningmodules.aspx>. The Disclosure Proofing of ACS Data in CTPP module will be released in June 2013.

Reports of Census Data Subcommittee at TRB

Mara Kaminowitz, Baltimore Metropolitan Council, Mkaminowitz@baltometro.org

The Census Data sub-committee of the Urban Transportation Data and Information Systems Committee met at the 2013 Transportation Research Board annual meeting. The theme of the meeting was the new Census Application Programming Interface (API). The first presentation was an introduction to the Census API, and covered what the new API is, how it works, and who might want to use it. The second presentation was by Catherine Lawson, State University of New York (SUNY) Albany, and chair of the TRB Urban Data Committee, which uses the Census API to access data for transit analysis. This application ensures that everyone is working with the same data, and can be easily updated when new Census products are released.

Following the presentations there was a discussion of the resource needs and issues brought about by this new technology. The Census API is primarily for computer programmers and web developers. There will be a growing need to develop these skills within the data community and to make programmers aware that this tool exists. Additionally the Census API, along with other open-source data platforms, presents challenges to determining data validity and appropriate uses of different data sets. It is important to ensure that new products created from the Census API appropriately identify the data set and year and disclose if the developer has altered the data in any way.

The meeting concluded with an update from the Census Bureau and AASHTO on the upcoming CTPP data release. With so much new data and technology, the Census Data sub-committee is looking forward to an exciting year.

Improving Access to Microdata files

Elaine Murakami, FHWA Office of Planning, Elaine.murakami@dot.gov

This issue of the CTPP Status Report includes two articles about different research centers that allow for microdata access to survey or administrative records. Sometimes what you need is not available in pre-defined tables such as Summary File 1 from the decennial Census, or available in American FactFinder for a 1-year ACS summary, or a 3-year or 5-year CTPP summary. The Census Bureau's Research Data Center, described in the article by Mark Fossett, has 15 different locations across the United States from which to access the data. The Transportation Secure Data Center described by Jeff Gonder is a virtual data center that does not require travel to the National Renewable Energy Laboratory in Colorado.

Protecting individual confidentiality is a critical component to these research centers, and in both cases, a proposal must be submitted and accepted before data access is granted. Related articles and webinars are listed as following:

- The article of "PUMS and PUMAs" by Elaine Murakami and the link is http://www.fhwa.dot.gov/planning/census_issues/ctpp/status_report/sr0409.cfm
- The webinar of "How to use IPUMS at the University of Minnesota" http://ctpp.transportation.org/Pages/webinar_directory.aspx
- The Metropolitan Travel Survey Archive located at the University of Minnesota <http://nexus.umn.edu/Papers/SurveyArchive.pdf>

Census Bureau Research Data Centers for Transportation Planners

Mark Fossett, Texas Census Research Data Center, m-fossett@tamu.edu

Research Data Centers (RDC) provide secure access to restricted-use demographic and economic data for statistical purposes to qualified researchers with approved research projects. RDCs are established in collaboration with leading universities and research organizations under the auspices of Joint Statistical Project Agreements. Under these agreements, RDCs are locally sited U.S. Census Bureau facilities, staffed by Census Bureau employees, which meet all physical and computer security requirements for access to restricted-use data. RDCs are found at several locations around the country including: the Atlanta and Chicago Federal Reserve Banks, the National Bureau of Economic Research (NBER) in Boston, the Universities of California at Berkeley and Los Angeles, Stanford University, the United States Census Bureau in Suitland Maryland, the University of Michigan, Cornell University, Baruch College, Duke University, RTI International, the University of Minnesota, the University of Washington, and Texas A&M University. To find out if your institution is partnered with an RDC, go to: www.census.gov/ces/rdcresearch/rdcpartners.html.

Many datasets that are available through an RDC may be of interest to transportation planners. The following data sets may hold particular relevance.

- The American Community Survey (ACS): with an approved RDC project, microdata, including detailed geographies (e.g., census tract, block group, and block) for residence and place of work, are available for analysis.
- The Commodity Flow Survey (CFS): with an approved RDC project, a researcher would be able to link data from the CFS to those from other Economic datasets such as those featured here: <http://www.census.gov/ces/dataproducts/economicdata.html>.

- The Longitudinal Employer- Household Dynamics (LEHD): most of the public LEHD data files are by state but the dataset available in the RDCs provide access to the microdata. Some LEHD data contain Federal Tax Information (FTI). Use of LEHD data containing FTI requires approval by both the Internal Revenue Service (IRS) and the Census Bureau.

The first and most important step in exploring the possibilities of working toward a RDC project is to contact the RDC Administrator at your RDC location: <http://www.census.gov/ces/main/contact.html>. Because projects involve access to restricted data, the proposal development and review processes are extensive, and in some cases, may require review from a sponsoring agency as well as by the Census Bureau. Interested researchers should consult with the local RDC Administrator as early in the proposal development process as possible so he or she can make you aware of all relevant issues and provide an overview of the applicable proposal requirements. Also note that, once a project is approved, researchers undergo a background investigation to gain clearance to enter the secure lab. Proposals are developed in coordination with the local RDC Administrator who transmits the proposal to Census Bureau for review when it is ready. The full process of proposal development, Census review, security review (initiated only after receiving project approval) usually takes 6-12 months with exact length of review varying depending on the type of data requested. The Census Bureau Center for Economic Studies (CES) provides information about the data available, the proposal format, and other relevant issues. See the following links for general information about the Center for Economic Studies <http://www.census.gov/ces/> and for more specific information about the RDC network <http://www.census.gov/ces/rdcresearch/index.html>. All analysis for approved projects is conducted in the secure computing lab at the local RDC. The data are hosted by the CES in Bowie, Maryland with researchers accessing the data via thin-client terminals in the local secure labs. The thin-client terminals use Red Hat Linux. The software for conducting analysis is loaded from CES servers; available programs include

SAS, Stata, R, MatLab, Gauss, HLM, and SUDAAN for statistical analysis and QGIS, GeoDa, and GRASS for GIS analysis. As projects progress to successful completion, research results are subject to Census review to assure that confidential information is never disclosed. Researchers need to be mindful of such issues to plan for successful projects. The payoff for well-chosen projects is that the rich

data allow researchers to develop better answers to the important questions they are investigating, often advancing the debate on the particular subject well beyond the previous state. The author would like to thank Bethany DeSalvo from Texas Census Research Data Center for her contribution to this article.

Transportation Secure Data Center – Real World Data for Planning, Modeling and Analysis

Jeff Gonder, National Renewable Energy Laboratory, Jeff.Gonder@nrel.gov

To resolve the inherent conflict between preserving survey respondent privacy and making vital transportation data more broadly available, the U.S. Department of Energy's National Renewable Energy Laboratory (NREL) and the U.S. Department of Transportation (DOT) have launched the free, web-based Transportation Secure Data Center (TSDC) at www.nrel.gov/tsdc. Unlike other sensitive data archives to which users must physically travel, TSDC users may access microdata through a secure online connection from the comfort of their own desks after completing a simple application process.

Data Available through the TSDC

The repository includes data from value pricing/tolling and travel surveys collected from the municipal to federal level using global positioning system (GPS) devices. The millions of data points available through the TSDC include second-by-second GPS readings from many studies. NREL screens data for missing values and adds metadata to assure quality and supply context.

Table 1: Sample Data Sets

Data Source	Date(s) Collected	Vehicles	Vehicle Days	Study Length
Atlanta Regional Commission Travel Survey*	2011	1,653	8,589	7 days
Chicago Regional Household Travel Inventory *	2007	408	1,773	7 days
Puget Sound Regional Council Traffic Choices Study	2004-2006	484	145,273	18 months
Southern California Association of Governments Regional Travel Survey	2001-2002	624	1,208	2 days
Texas Department of Transportation Travel Surveys from Austin, Houston, San Antonio and many other cities	2002-2011	3,404	5,258	1-2 days

*Also includes wearable GPS component to capture other travel modes.

Valuable Geospatial Analysis Resource

Individual data collection and analysis projects can cost millions of dollars, so reusing the results enables more effective utilization of limited public funds. The TSDC provides web-based access to valuable transportation data for many applications, including:

- Transit planning
- Travel demand modeling
- Congestion mitigation research
- Emissions and air pollution modeling
- Vehicle energy and power analysis
- Climate change impact studies

- Homeland Security evacuation planning
- Alternative fuel station planning
- Validating transportation data from other sources
- Toll and pricing research

Two Levels of Clearance

While the detailed geographic and time/speed resolution makes GPS data extremely valuable, associated privacy concerns often discourage collecting agencies from sharing it with other researchers. The TSDC’s two-level access makes this data available while maintaining participant anonymity.

Cleansed data is readily available for download from the website. This publicly downloadable data includes high-level summary statistics, vehicle and participant demographic information, and second-by-second speed profiles (with latitude/longitude detail removed).

Detailed spatial data is made available online through a secure web portal. After completing a simple application and obtaining approval, researchers may access the GPS data files. Users are prohibited from copying or transferring raw data, but they are able to

conduct statistical and geographic analyses from the microdata records and to generate aggregated results for removal from the secure environment. The following table lists example features available in the environment, and users may import additional software tools and reference data.

Reference information – such as the underlying road network, demographic, and economic grid data – helps support geographic information system (GIS) analyses. Controlled access, secure storage, and the support of NREL’s legal and cyber security offices provide additional safeguards for this data.

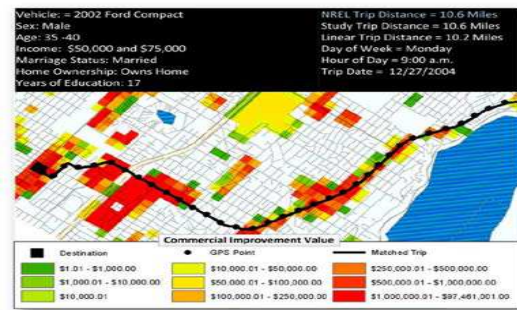


Figure 1 Example Analysis of TSDC Spatial Data

Table 2: Examples Features Available in the TSDC Environment

Included Features	Provided Tools/Reference Files
Database Querying	pgAdmin/PostgreSQL
Statistical Analysis	R Python(x,y)
GIS Visualization	ArcGIS QGIS uDig
Demographic, Economic, and Land-Use Data Layers at Various Summary Levels	UrbanSIM Census 2010 American Community Survey (ACS)

More Information

NREL continues to build TSDC data sets. Visit www.nrel.gov/tsdc to subscribe for periodic e-mail updates when new data sets and features become available in the TSDC.

To discuss options for joining NREL as a partner in the TSDC, to apply for spatial data clearance, or for more information on the project, contact NREL’s Jeff Gonder at (303) 275-4462 or Jeff.Gonder@nrel.gov or DOT’s Elaine Murakami at Elaine.Murakami@dot.gov.

Industry by Occupation Tables from CTPP, ACS PUMS and EEO

Travel demand modelers are interested in total employment by occupations and industries as they are critical inputs for travel demand models, particularly the work location choice model and destination choice model. However, many sources of employment provide total employment and do not distinguish employment by occupation. For example, Quarterly Census of Employment and Wages (QCEW) provided by Bureau of Labor Statistics, which is regarded as authoritative public employment source, only provides industry information, but not distinguish occupation categories.

Quarterly Census of Employment and Wages (QCEW) from Bureau of Labor Statistics (BLS) and two databases produced by the Census Bureau Longitudinal Employment Household Dynamics Program (LEHD) – Quarterly Workforce Indicators (QWI) and LEHD Origin Destination Employment Statistics (LODES) and private sources like InfoUSA are classified by North American Industry Classification System (NAICS). Check the following link for a list of sources for employment data: http://www.fhwa.dot.gov/planning/census_issue_s/ctpp/status_report/sr1209.cfm

We list 3 ways to get a tabulation of industry by occupation: (1) 2006-2008 CTPP, and soon the 2006-2010 CTPP; (2) ACS PUMS; and (3) the EEO file.

1. 2006-2008 CTPP includes table B22222 Occupation (25) by Industry (15) for workplace location, and is currently available using the 2006-2008 ACS records at the County and Place geographic level. It will be available at much smaller geographic scale in the 2006-2010 CTPP, expected out in a few months. However, the industrial codes have been aggregated in the CTPP table.
2. If you need full NAICS detail, you may prefer to use the ACS PUMS, at PUMA level residential geography and apply the same distributions to a workplace geography (note: each PUMA must have 100,000 residential population).

3. Another option is using Equal Employment Opportunity (EEO) data; the most recent EEO data are based on ACS 2006-2010. EEO contains a list of crosstabs by industry and occupation for workplace location (Table Set 10 is industry by occupation tabs and the table list for EEO can be found here: <http://www.census.gov/people/eeotabulation/data/eeotables20062010.html>). EEO is only available for counties and places and is subject to minimum population thresholds¹.

In 2013, ARC will use the assumed consumption of labor by occupation for each industry that is used in PECAS (the Production, Exchange and Consumption Allocation System) as size term coefficients in the work location choice model. That is, ARC will match workers by occupation with the kinds of jobs that they are likely to work in by industry. There are several key issues involved:

1. The occupation categories used in the Activity-Based Model synthetic population and in the land use model (PECAS) need to be consistent.
2. A base-year TAZ file with employment by industry must be consistent with PECAS employment by industry.
3. PECAS provides a region-wide control total of workers by PECAS occupation category for each year the Activity-Based Model will be run, to ensure that the synthetic population has consistent numbers of industry by occupation compared to the figures.

This article is contributed by Guy Rousseau from Atlanta Regional Commission (ARC) and Ju-Yin Chen from Virginia Department of Transportation and compiled by Liang Long from Cambridge Systematic. If you have any questions, please contact Dr. Liang Long by e-mail at liang.long@dot.gov

¹ Depending on content, some tables have minimum residence population thresholds for some geographic summary levels, either 50,000 or 100,000. Population thresholds are always based on the residence population, even for tabulations at the worksite geography.

Incremental Integration of Land Use and Activity-based Travel Modeling: Using CTPP2000 for Model Validation and Calibration

*Liming Wang, Portland State University,
lmwang@pdx.edu*

*Paul Waddell, University of California,
Berkeley, waddell@berkeley.edu*

*Maren Outwater, Resource Systems Group,
moutwater@rsginc.com*

The purpose of this research is twofold. Firstly, we propose an incremental approach that balances the risks and benefits of moving operational models in new directions. This paper addresses replacement of the home-based work destination choice model within the 4-step travel model system with a pair of choice models at the individual worker level, implemented as long-term choices in the linked land use model system. Secondly, our models also provide a way to derive matches between workers and their workplace with commonly available data. These matches complement synthetic populations and provide a key input for activity-based travel models. Our models predict whether a worker will choose to work at home on a long-term basis, and if not then choose an out-of-home job. These models link an individual worker to a specific job at a workplace, and therefore directly predict commuting patterns. We present the model specification, estimation results, and more specifically results of validating the models against observed commuting data from the Census Transportation Planning Products (CTPP). The model reproduces observed commuting flows quite well, and computational performance is fast, in spite of operating at the individual worker and job level.

Data

The data used in this study is from the Puget Sound region, the Seattle metropolitan area of

Washington State. The 2006 Puget Sound Regional Council (PSRC) Activity Survey was the main data source for model estimation. This data was augmented by parcel data and accessibility measures from the PSRC travel model, as summarized in Table 3. The activity survey data was geo-coded to parcel, as were business establishments from the QCEW unemployment insurance records from the State of Washington. The travel diary and household information provided details on the current and previous residence location, and the current and previous workplace location. By combining these data sources, we were able to identify for each worker whether they worked at home or not, and if not, their specific work location to the level of a parcel. Since we had created a micro-level table of the jobs at each location based on the number of employees listed for each business establishment, we assigned the workers to a specific job at these workplaces.

The Models

The HBW travel demand is derived from two closely related choices: whether these workers will work at-home, and if not, where they will work, conditional on household residential locations, and on individuals in the household being identified as workers.

Although the two outcomes we want to predict are related, we posit the model as a choice of workplace, conditioned on a prior choice of whether to work at home. The logic is both pragmatic, to keep the model structure simple and robust, and also based on a behavioral expectation that the decision of a worker to be an at-home worker is often motivated less by the quality or accessibility of outside employment opportunities than by personal characteristics like education (and opportunities to generate a consulting or home-based business), or desire to be able to care for smaller children at home while working.

Table 3: Data Sources

Data	Source	Description
Estimation		
PSRC Activity Survey	PSRC	2006 Household activity and travel survey for the central Puget Sound region
Business establishment data	QCEW unemployment insurance records of Washington State	Individual business establishment geocoded to parcel location
Travel model data	PSRC Travel Model	Zone to zone travel times by mode for a.m. peak, from the 2006 travel model; network
Puget Sound parcel data	Data assembled and processed from appraisal data of King, Kitsap, Pierce and Snohomish Counties	Year 2005 data
Calibration and Validation		
Puget Sound parcel data	Same as above	Year 2000 reflected by removing buildings built after 2000 from the 2005 database
Synthetic Population	Census PUMS, SF3, Parcels	Adapted Household Synthesizer to generate households by block group and assign to individual buildings
Commuting Flows	Census Transportation Planning Products (CTPP)	CTPP 2000 commuting flows, aggregated to 19 x 19 traffic analysis districts

Calibration and Validation

Due to the fact that the share of at-home workers in the Activity Survey (8.45%) is different from that in the CTPP validation target (4.32%), the constant term in the Work-at-home Choice Model is calibrated so that the predicted share of at-home workers will be consistent with the target share (4.32%). This difference may be due to the inconsistency in the definition of work-at-home. We calibrated the constant term to meet the CTPP target of 4.32%.

To validate the models, we wish to examine their combined capacity to predict home to work commuting flows. To do this, we have used the CTPP journey to work data for 2000, a year 2000 travel model and network, and a synthetic population for the entire Puget Sound region. This is not a small sample application, but rather an application to the full set of synthetic workers in the region, approximately 1.6 million.

Figure 2 summarizes the results of this validation exercise. The trip-length frequency distributions, shown in subfigure a and b, are very similar. The deviations of the predicted

from observed values for origin-destination trip flows are shown in subfigures c and d. The predictions fall very close to the diagonal, and the distribution of the errors is very tightly spiked at zero, with a very narrow distribution. In short, the model predictions are very close to the observed commuting flows, and do not require substantial k-factoring or other forms of constraints to achieve a high level of predictive accuracy.

The results also offer a significant improvement over the existing trip distribution model for predicting HBW trips, which is a gravity model as is common in four-step models. Subfigures e and f of Figure 2 reflect the comparable validation results of the gravity model. It shows a visually less robust result than our new models, and summarized in subfigures c and d. A summary measure, the Root Mean Squared Error (RMSE), also confirms what is evident in the graphs: assuming the CTPP commute flows as ground true, the RMSE for the gravity model is 2558.65, whereas for the new models it is 1578.82, representing a 38.3% improvement.

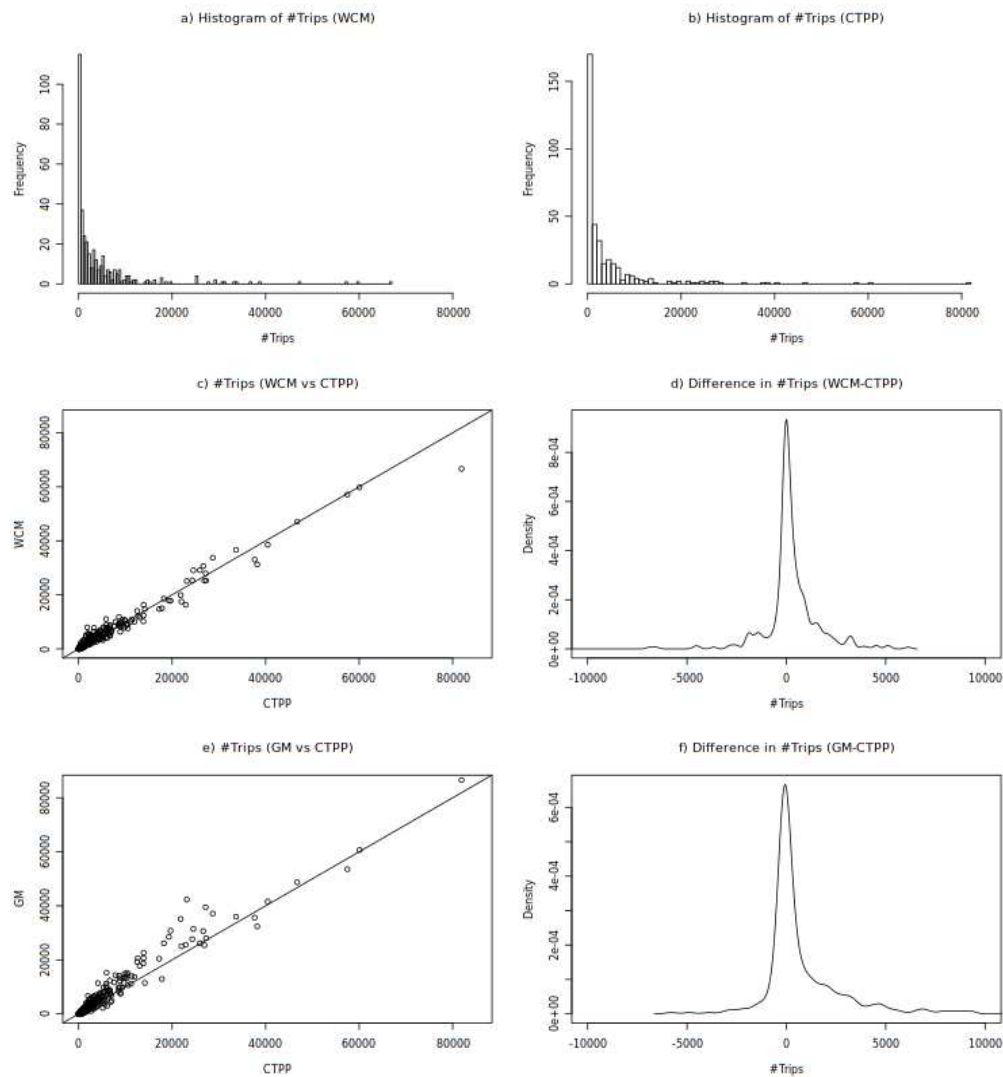


Figure 2: Validation results for the new models (WCM) (a), compared to CTPP (c, d), and PSRC Gravity Model (GM) results compared to CTPP (e, f).

Conclusions

This research presents the results of an incremental step in the direction of integrating land use modeling and activity-based travel modeling into operational use, and offers a strategy that is relatively straightforward to adapt to existing model systems. It offers some novel features, including an individual work-at-home model and a workplace choice model that matches individual workers to individual jobs. The models are relatively simple in their construction and specification, yet both the statistical estimation results and the aggregate

validation results appear quite robust. The models are able to recreate aggregate district level commuting flows from CTPP quite well, and significantly improve the predictive capacity of the PSRC travel demand model.

This is an excerpted version of Wang, Liming, Paul Waddell, and Maren Outwater, 2011. *Incremental Integration of Land Use and Activity-based Travel Modeling: Workplace Choices and Travel Demand*, Transportation Research Record: Journal of the Transportation Research Board, Vol 2255, pp 1-10.

CTPP Contact List

CTPP Hotline – 202/366-5000

E-mail: ctpp@dot.gov

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

PH: 202/624-3556

E-mail: pweinberger@aaashto.org

Jennifer Toth, AZDOT

Chair, CTPP Oversight Board

PH: 602/712-8143

E-mail: JToth@azdot.gov

Susan Gorski, MI DOT

Vice Chair, CTPP Oversight Board

PH: 517/335-2958

E-mail: gorskis@michigan.gov

Census Bureau: Social, Economic and Housing Statistics Division

Alison Fields

PH: 301/763-2456

E-mail: alison.k.fields@census.gov

Brian McKenzie

PH: 301/763-6532

E-mail: brian.mckenzie@census.gov

FTA

Ken Cervenka

PH: 202/493-0512

E-mail: ken.cervenka@dot.gov

BTS

Li Leung

PH: 202/366-0634

E-mail: li.leung@dot.gov

FHWA

Elaine Murakami

PH: 206/220-4460

E-mail: elaine.murakami@dot.gov

Ed Christopher

PH: 708/283-3534

E-mail: edc@berwyned.com

Liang Long

PH: 202/366-6971

E-mail: liang.long@dot.gov

TRB Committees

Catherine Lawson

Urban Data Committee Chair

PH: 518/442-4775

E-mail: lawsonc@albany.edu

Clara Reschovsky

Census Subcommittee Co-Chair

PH: 202/962-3332

E-mail: creschovsky@mwcog.org

Mara Kaminowitz

Census Subcommittee Co-Chair

PH: 410/732-0500

E-mail: mkaminowitz@baltometro.org

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 e-mails to be batched in a daily digest. The website also includes an archive of past e-mails posted to the listserv.