Better Journey-to-Work Measurement

A Cities21 Climate Protection Proposal for CARB/LUSCAT
By Steve Raney. February 11, 2009, Palo Alto, CA. Updated March 13, 1PM

Summary:

The Urban Land Institute's report, "Growing Cooler: The Evidence on Urban Development and Climate Change," shows that carbon from driving will increase even with more Priuses. Hence dense, vibrant, smart growth transit-oriented in-fill development is essential to achieve 2020 carbon targets. Pioneering state legislation including AB32 and SB375 further reinforces the need for smart land use.

The CA Climate Action Team and the Bay Area Metropolitan Transportation Commission (MTC) are very sincere about reducing VMT/carbon via smart growth to protect the climate. Both must have very high quality data to measure the progress of climate protection implementation. Otherwise, both will just be proposing a series of projects without being able to ascertain if they work or not. If the projects don't work, both need to know rapidly, so that they can change course to meet 2020 carbon targets. Nationally, land use related measurement is primitive. State and federal governments really don't know what is happening. MTC has always been the nation's MPO (metropolitan planning organization) measurement leader, but even MTC data is lacking.

To protect the climate, we need innovation in measurement of journey to work information: home origination address and work destination address. We need 95% or better coverage of all Bay Area (and California) workers and all extreme commuters (those who live outside of the 9-county Bay Area but commute in), and we want that 95% coverage updated EVERY year and made available within three months of data collection. More than 50% of household VMT occurs from commuting.

This proposal meets data anonymization/protection standards found in European Parliament Privacy Directive 95/46/EC, the strictest privacy law to date.

Details:

Current journey to work data is deficient

At the Feb 24, MTC/ACT (Northern California Association for Commuter Transportation) SB375 Conference, Ted Droettboom, MTC/ABAG/Air District Joint Policy Committee said "We're looking for unambiguous measures of VMT. Many of our flawed regional models are based on 1990 behavior measures. We are forced to rely on traffic counts supplied by cities that lack budget to count traffic. We don't know VMT in this region with any degree of certainty. We have to have better VMT measures."

The Census Transportation Planning Package journey-to-work flow data produced from the 2000 Census Long Form questionnaire provides a "1 in 6" sample but that data will no longer be available. Replacing the Long Form is the American Community Survey (ACS) which is a "rolling annual sample" that has much smaller sample size and wider confidence intervals. Considering that coupled with the confidentiality data release policies of the Census Bureau, usable flow data will not be produced other than that which can be modeled and synthesized (see: http://trbcensus.com/notes/NCHRP_CTPP_Data_Proposal_Oct2008.pdf). The preliminary research to attempt to develop these data has not yet begun.

The LEHD (Local Employment Housing Dynamics) data on the other hand holds the promise of providing 80 percent or better coverage for journey to work data. However, there are some serious data issues with the current CA implementation the resultant data flawed for the task at hand. Hence there is no identifiable data set in which to turn for precise queries.

As far as work trips. A little bit more than 50% of household VMT is in commuting. See Jonathan Rose and Calthorpe: (http://www.cities21.org/HH_NRG_consumption.htm, http://www.cities21.org/HomeEnergyUseJonathanRoseLLC.xls). The

average annual Bay Area commute is about 6,720 miles (14 mile one-way commute and 240 commuting days). To meet 2050 CO2 goals, we surely need to cut average one-way commute distance dramatically.

Cities21 has highlighted a number of data quality problems with current government datasets. The identification of these problems led directly to a $60,000 third party study funded by the National Academy of Sciences: "NAS NCHRP 08-36/task 81, Enhancing the American Community Survey Data as a Source for Home-to-Work Flows." http://www.trb.org/trbnet/ProjectDisplay.asp?ProjectID=2403 .

The 2006 Journal of the American Planning Association (JAPA) Robert Cervero / Michael Duncan paper argues for emphasizing efforts to reduce jobs/housing distance to reduce VMT. The article is "Which Reduces Vehicle Travel More: Jobs-Housing Balance or Retail-Housing Mixing?" in the Autumn 2006 JAPA. It's not that Cervero is arguing against smart growth to minimize VMT on the 84% of non-work trips, he's just saying that we have to place a higher priority on the 16% of journey-to-work trips to reduce VMT. Hence, journey to work data is very important.

Recommendation: Collect "full population" journey to work data via state income tax forms

Proposed is a NEW dataset with full population O/D, without personally identifiable characteristics. Other datasets contain personally sensitive information such as income and race, and, because of legitimate data restrictions, cannot provide answers to the new, precise queries envisioned.

We need to modify CA Income Tax forms (just slightly) to collect work address data (including two work addresses for married couples filing jointly), to provide 95% or better coverage of CA journey to work, updated each year. Once state law has been changed, then the data can be collected by the State Labor Market Information (LMI) office, and made available to MTC. Budget must be provided at the State level and for MTC.

The resultant journey-to-work database will be "anonymized" to the point where no "personally identifiable" data (including no home addresses) is stored. LMI should establish procedures to anonymize the data and safely delete the personally identifiable source data.

Acceptable anonymization techniques:

- Geocode addresses to a 1/5 mile by 1/5 mile grid. The diagonal of a 1/5 mile grid is 1/4 mile, so is sufficiently fine-grained for bus stop planning.
- Geocode addresses and then randomly add +/- 100 meters displacement in X and in Y.
- Zero out the last two digits of street addresses, anonymizing street addresses to block level

Alternate recommendation: Use DMV rather than Tax Forms

- Collect O/D annually or biannually via the Department of Motor Vehicles. There is a move afoot to collect odometer readings via the DMV. O/D information should also be collected.

Heretofore impossible queries made possible with better data.

- It's 2010. The state Housing and Community Development Department's Regional Housing Needs Assessment policy to balance jobs/housing in affluent, job-rich suburbs is in place. How effective is this policy?
- MTC's 2035 Draft Regional Transportation Plan, Chapter 4 Investments, Section "Take Bold Steps Toward Focused Growth," explains a program to reduce VMT via a strong TOD in-fill policy. This MTC program is called the FOCUS Priority Development Area (PDA) policy. (The SB375 version of this program is called the Sustainable Communities Strategy.) How many workers moved into these PDAs in 2012?
- It's 2017, RHNA, RTP, and SB375 SCS are all working in harmony. Measure effectiveness of this policy cocktail.
- San Ramon and Dublin were the fastest growing residential communities in the Bay Area in 2007. 6,000 new housing units were added. What is the distribution of work destinations for

these new residents? Is the average journey to work distance shorter or longer than we expect? Are our new policies working like we expected?

- We've added a super new master planned community in Tracy. Their marketing brochure promised that this would be an exceptionally green place, with solar on every rooftop. What's the journey to work distribution look like for those 2,000 new 3,000 square foot single family homes?
- We implemented policies to reduce commute distance in 2010. How effective were these policies? We then made the policy stronger. How effective were the new policies in 2012?
- Provide a map of the commute distribution of Bay Area extreme commuters, covering 95% or more of those commuters.
- Is there a need for subscription commute bus service from Manteca to Bishops' Ranch?
- Using NJIT's algorithmic bus route optimization software, where should we place bus stops to attract the most riders?
- We have a new Alameda County dynamic ridesharing service. Where should we target our marketing efforts?
- By May 2013, answer the question: In 2012, where were the new housing units built for Bay Area workers?
- A natural disaster (Loma Prieta or Northridge earthquake), infrastructure failure (2007 Minnesota I-35W bridge collapse) or terrorist act creates a temporary reduction in regional transportation system capacity, reducing the ability of workers to travel to their jobs, reducing economic output. Given better knowledge of journey to work flows, design rapid capacity responses to alleviate the capacity reduction. Given better knowledge of commute flows, better design an emergency evacuation plan.

It is expected that land use policies will tighten over time. Climate objectives become more aggressive from 2020 to 2035 to 2050, requiring new land use policies. The need for precise queries will grow with the need to measure efficacy of new land use policies.

Additional Benefit:

- Provides very useful and accurate input data for MTC and VTA travel demand forecasting models. Makes modeling better.

Optional Longitudinal Data Enhancement

- Personal time series data would lead to a whole new series of analytical capabilities. Internally, in the secure data vault, encrypted social security numbers (SSN) may be kept. The encryption technique should be chosen to be especially secure: AES 256 or better. When time series is released outside of the vault for analysis, a unique, random ID replaces the encrypted SSN, allowing time series analysis with no danger of external SSN disclosure. (Thanks to John Ciccarelli for this enhancement.) US Census calls the anonymization of SSN a “PIK” for Protected Identification Key.
- New longitudinal analysis enabled would include improved calculation of “employment-at-one-location duration” and residential duration. The data we have on this flawed and national-level statisticians are the first to admit this. For a sampling of employment and residential duration data, see: http://www.cities21.org/workerHsngDetails.htm#duration .

“Big Sister” and Preventing Sinister Data Uses

Past data breaches of sensitive datasets with personally identifiable data underscore the need to make this journey to work dataset NOT have any personally identifiable data in it. Creating a large, precise journey to work dataset might create images in the minds of some of “Big Brother.” So, regrettably, we must define a new term to position this effort against Big Brother. Whereas in Orwell’s book, “1984,” Big Brother tracked personal information to reduce personal liberty, “Big Sister,” uses better data to more accurately measure and adjust climate protection and emergency evacuation efforts. Climate protection efforts enhance human experience. Better-planned evacuation procedures can save lives.

There are two sensitive uses of full population O/D data. Improved data makes these uses slightly more effective, but not tremendously more effective.

1) The first sensitive use of anonymized block-level or 1/5 mile by 1/5 mile grid commute flow data that has arisen is the ability of one competitive employer, for example Yahoo, to conduct a direct mail employment recruiting campaign to lure Google workers in a given geographic area. Yahoo has no visibility into the job functions that Google worker-residents have, nor does Yahoo know street address of workers. Hence, Yahoo would have to blanket the entire block with direct mail to reach Google employees. For Yahoo, this is only a slight improvement in direct mail recruiting efficacy, because desirable tech worker residential locations are well-known already from TAZ-level datasets.
2) The second sensitive use is also Human Resources related. There are thousands of Genentech (soon to be Roche) biotech workers in South San Francisco. A real-estate developer looking to build a new industry-specific biotech center asked to examine the "to South San Francisco" journey to work pattern to understand where best to locate the center to "steal" Genentech employees. For such analysis, the improvement from CTPP3 TAZ-level sampled data to full population block-level data is slight.

Retailers use O/D data to help figure out where to locate retail business to attract the optimal shopping demographic. This use of O/D data is not considered sensitive.

Journey to work Background Information

The study of journey to work is a bit of a specialized field. One example of some of the things that we do with journey to work data can be found in the Bay Area Business Park Catalog (part of a larger EPA study): http://www.cities21.org/BABPC/ . A three-paragraph description follows:

> We have identified 17 Bay Area suburban major employment centers, 13 in Silicon Valley. The 17 centers are mostly traditional suburban office parks with many tech workers. Exceptions to traditional office parks include: a) Emeryville is an edge city with more than 1MM square feet of retail and extensive residential, b) Stanford University encompasses the University, the regional Stanford Shopping Center, Stanford Hospital, and downtown Palo Alto, c) SJC is the San Jose airport major activity center, d) Walnut Creek is a suburban downtown with dense employment.
>
> Each center has at least 15,000 jobs. The 17 centers support a total of 594,000 jobs. SOV commute mode share varies from 85% to 65%. The Stanford University job center stands out with 16.8% of commuters biking or walking to work. The other 16 job centers clump between 4.9% and 0.6% bike/ped commute mode share. Stanford's programs to put housing by jobs are shown as a singular success in the high-mileage world of suburban job centers.
>
> Commute distance appears longer than was previously thought. A mean "crow flies" one-way commute distance (for Stanford Research Park) of 14 miles translates into roughly 18.2 driving miles. Other commute surveys report Silicon Valley commute distance of 14 miles. The CTPP3 data used in this EPA study uses a larger sample than other studies and has less "self selection bias." This result may point out that the high income workers in job centers live farther away than typical suburban workers, or it simply may point out that other phone surveys underreport commute distance, because higher income workers are more likely to hang up on tele-market researchers.

This proposal was derived from meetings with: Nanda Srinivasan (Consultant, CTPP and National Household Travel Survey), Ed Christopher, (FHWA, Chair TRB Census Transportation Committee), Elaine Murakami, (FHWA, Mgr, CTPP and National Household Travel Survey), Chuck Purvis, (MTC), Eileen Rohlfing, (State Employment Devt Dept, Labor Market Information Division).

A letter of support from UC Berkeley's Director of the Insititute of Urban and Regional Development, Robert Cervero, is appended at the bottom.

Should there be CA state regulatory/legislative interest in pursuing this proposal, it will be worth exploring the following:

- Proposal analysis by members of the national Transportation Research Board Subcommittee on Census Data for Transportation Planning, ABJ30(1)
- Commitment by Google's Google Earth division to create a data layer of this origination/destination information, to make it very easy to conduct GIS analysis with this data (and combine with other analysis layers such as HOV and transit networks.)
- Analysis by MTC and ABAG
- Commitment by NJIT to apply their algorithmic bus route planning tool to this data.

Legislative Comments / Details

1. It's worthwhile to consider whether "LEHD and QCEW" data (Longitudinal Employment and Household Dynamics / Quarterly Census of Employment and Wages) can provide fine-grained journey to work data to answer the precise new queries suggested above. The approach of using LEHD/QCEW would surely be politically simpler than modifying state tax returns.

LEHD and QCEW are explained in Elaine Murakami's paper:
http://www.edthefed.com/papers/lehd_process_draft.pdf .

To develop fine grained commute flow data from LEHD/QCEW, the following should occur:

- California should adopt MN's QCEW MWR (Multiple Workplace Reporting) reporting, where the Unemployment Insurance file links employee SSN to a specific worksite. SSN data linked to worksite address is essential. When, for other states, this data is not provided, then LEHD uses algorithmic assignment of flows (synthetic flows). Such algorithmic assignment causes precise land use queries to be invalid. Further, SSN data allows for longitudinal studies.
- Even with MN level QCEW MWR, CA data will be quite flawed, because of the large percentage of CA tech worker contractors who will have their contracting service address reported, rather than their work site address reported. Further, there is a higher-than-average number of self-employed workers in CA. These workers are not covered by Unemployment Insurance, so are "invisible" to LEHD. MN funds State Employment Development Division work in contacting and cajoling employers to accurately report their MWR data. It is important to fund this type of work. Accurate QCEW MWR does not just magically appear without state intervention.
- A special initiative to collect federal employee data (for federal employees working in CA) should be undertaken as these workers are not covered by unemployment insurance.
- State worker data is often poorly reported, so funding should be provided to get state worker data reported properly.
- LEHD adds noise to the residential location of O/D pairs, generally moving a location +/- two blocks. Fred Anderson has created a table that shows that 90% of these residential moves end up in the same census tract. For the full population O/D dataset, we've proposed 3 techniques for similar anonymization, with slightly less geographical displacement

2. College students who work part-time

College students away at school often use their parents' address for income tax reporting purposes. For a student with a daily journey to college 9 months out of the year and some part-time job commuting, what data do we really want to collect? We really want to collect the recurring daily journey, in this case the journey to college. This is especially important for "commuter colleges" where students drive to school with average one-way distances of 10 miles. This is less important for colleges where students reside on or close to campus. The personal O/D data reporting could explain what is desired and instruct such students to report their journey to college.

3. Will people report their O/D information accurately on state tax returns?

Within the tax form instructions, there should be a short but compelling explanation of the benefits of reporting the data with a guarantee that the data will not be misused. The labeling of the field on the actual tax form should probably use a few words of assurance: “work street address (for transportation and evacuation planning).”

References

“Privacy-Protecting Commute Shed Study,” Transportation Research Board Conference, January 2003, Steve Raney. http://www.cities21.org/CommuteShed_TRB_111502.doc

ABSTRACT: A new methodology has been developed for collecting fine-grained employee commute origination data from employers in major employment centers. Many U.S. multinational firms have adopted the European Parliament Privacy Directive 95/46/EC, the strictest privacy law to date. The methodology discussed complies with this Directive. “Fine-grained” is defined as providing sufficient resolution (approximately 1,000 feet) to assist transportation planning of individual bus stops. Directive 95/46/EC provides for “anonymization” of data to where the data subject is no longer identifiable. Individual address data is aggregated to a 1/5 by 1/5 mile grid at the employer site using commercial geographic information systems software. Once the grid data is taken from employer premises, data is combined with those of other employers, providing further anonymization.

8,200 out of approximately 20,000 worker addresses have been collected from the Stanford Research Park employment center in Palo Alto, California using this methodology. While the sprawling spatial distribution of these addresses challenges many Transportation Demand Reduction strategies, planned transit system improvements should result in a significant patronage increase.

This methodology could be automated and applied nationally by the U.S. Census Bureau as part of their Transportation Planning Package, creating higher quality transportation data for improved investment decision making, ridesharing, and transit routing.

BERKELEY IURD INSTITUTE OF URBAN AND REGIONAL DEVELOPMENT

March 9, 2009

State Senate Transportation and Housing Committee
Sacramento, California

Subject: Support of Steve Raney's request for "better journey to work data"

Dear Committee Members:

I am writing to support Steve Raney's request to build a rich longitudinal database that would enable California policy analysts to monitor and track trends in journey to work. His request for quick-turnaround annual surveying of most Bay Area workers and extreme commuters would provide the kind of scientific, no-nonsense data needed to evaluate performance of proactive strategies that promote jobs-housing balance, VMT-reducing TOD, and other sustainable-growth strategies as found in HCD's "Regional Housing Needs Assessment", MTC's "2035 Regional Transportation Plan", and SB375's "sustainable community strategies" initiative. Without a rich time series of journey-to-work data, connecting VMT and travel-behavior outcomes to planning and policy initiatives will become difficult, which invariably poses the risk that political factors will win out over objective, scientific data in informing policy practices.

Such data will also be useful to a number of current and future research projects that are being carried out through research centers on the Berkeley campus, including the Institute of Urban and Regional Development, the University of California Transportation Center (UCTC), and the Center for Sustainable Transportation, among others. For example, I am currently conducting research that aims to associate the transformation of low-density office parks to more mature, mixed-use centers with reduced car dependence and work-trip VMT per employee. A database like that outlined by Mr. Raney would be invaluable to this and related studies.

In close, I strongly and enthusiastically support Mr. Raney's proposal. The investment in filling this critical data gap will pay off handsomely in many years to come – in the form of policies and practices that genuinely reduce the environmental footprint of the urban transport sector throughout California.

Sincerely,

Robert Cervero, Director, Institute of Urban and Regional Development
Professor, Department of City and Regional Planning
Director, University of California Transportation Center