



Caltrans Division of Research,  
Innovation and System Information

# Research

# Notes

Planning  
Policy  
Programming

SEPTEMBER 2015

Project Title:  
UTC - Demand Forecasting and  
Activity-Based Mobility Modeling  
from Cell Phone Data (UCCONNECT)

Task Number: 2793

Start Date: March 24, 2015

Completion Date: March 31, 2016

Task Manager:  
Frank Law  
Project Manager  
frank.law@dot.ca.gov

## Demand Forecasting and Activity-Based Mobility Modeling from Cell Phone Data

Exploring and developing methods for processing cell phone location data to generate travel behavior information for demand modeling, activity-based traffic micro-simulation.

### WHAT IS THE NEED?

As the field of transportation rapidly evolves, influenced by private sector and powered by information technologies (IT) and mobile technologies, it is more important than ever to be able to measure and realistically model travel demand in near real-time. Quantifiable models based on bottom-up individual-level data are required to guide top-down policy regulation and governance in the face of these fast changes.

Fortunately, as mobile devices and localization technologies have become ubiquitous over the past decade, mobility data has increasingly grown in volume giving rise to new opportunities for high accuracy activity-based demand modelling.

In the context of this project, we focus on cell phone location data as signal-derived geographic location data collected by telecom operators. The project explores and develops the methods for processing cell phone location data to generate travel behavior information with the focus on demand modeling, activity-based traffic micro-simulation, and discrete choice models.

### WHAT ARE WE DOING?

The major tasks for this project are:

- 1) **Data handling and bias correction** - Specifically, this task will develop a methodology for demographic and sampling biases correction towards reliable large-scale transportation demand applications based on cell phone data.
- 2) **Method development** - Develop a method of activity inference and activity chain generation for individual users based cell phone records. We will implement the methods in software and describe their use in a technical report.



DRISI provides solutions and  
knowledge that improves  
California's transportation system

- 3) **Agent-based traffic micro-simulation** - The calibrated micro-simulation will be validated on the loop detector data available from PEMS. Agent-based traffic micro-simulation in the SF Bay Area on a typical work day will be implemented in software based on MATSim environment. A technical report describing model calibration and validation will be prepared, as well as a publication for TRB Annual Meeting and ACCESS Magazine.
- 4) **Aggregation schemes:** Produce aggregation schemes allowing the use of non-parametric activity sequences inferred in Task 2 for specification and parameter estimation in Discrete Choice Models. Our initial work will concentrate on the mode choice for primary trips (home to work/school commute) that is most easily derived from cell phone data by disregarding all but two main (and most certain) hidden states with longest duration.

### WHAT IS OUR GOAL?

This project will contribute to developing methodologies and providing new source of data for demand modeling.

### WHAT IS THE BENEFIT?

The following benefits can be listed:

- Novel methods to produce up-to-date, cost-effective activity based demand models.
- Methodology of sampling bias correction in travel behavior derived from cell phone data.
- Machine learning tools for mode, destination and activity inference from massive scale non-pervasive individual mobility sensing.
- Demonstrated potential of data-rich activity-based micro-simulation models for demand forecasting and 'what if' scenario evaluation.
- Demonstrated potential of the new up-to-date source of data for forecasting demand variability in face of evolving demographics.

### WHAT IS THE PROGRESS TO DATE?

Held kickoff meeting and work in this task began March 24, 2015.

We processed two weeks of cellular data and detected home and work locations for a sample of over 1M commuters within 9 counties of SF Bay Area (TAZ1454). We developed a correction procedure based on spatial regression and iterative proportional fitting to rescale data from cellular network level onto TAZ units. Evaluated MATSim micro-simulations, made initial comparisons with observed (PEMS) volumes. Identified the need for data samples from longer time periods in order to infer secondary activities.

Results dissemination: prepared a promotional video for AT&T research showcase.