The SmartBay project: connected mobility in the SanFrancisco Bay Area

Andrew A. Campbell
Mogeng Yin
Sid Feygin
Alexei Pozdnuhkov

Abstract The novel mobility-as-a-service paradigm, enabled by ICT and mobile computing, is changing the transportation landscape quicker than traditional data sources, such as travel surveys, are able to reflect. This is particularly true in the San Francisco Bay Area as the influx of citizens and businesses to the city, the volatility of job markets, evolving demographics and internal migration further increase the variability of mobility patterns. It is more important than ever to measure, realistically model and forecast travel demand usig new data streams that most accurately reflect current travel and activity behaviors. The Smart Bay project extends the state-of-the-art in activity-based simulations by incorporating the anonymized data stream from the cellular network infrastructure. The efficacy of cellulary data in activity-based simulations is evaluated by comparing three scenarios: a control scenario using a traditional HHTS-based activity model, an activity model based only on cellular data, and a hybrid model using both the HHTS and cellular models.

1 Introduction

Informed transportation policy, planning, and operations decision making requires accurate forecasting of travel demand. Travel demand models are traditionally estimated using either stated or revealed preference travel survey data. The revealed preference survey is preferred, conducted in the form of an individual or household travel diary. These data collection methods provide a rich set of features for

Andrew A Campbell
University of California, Berkeley, e-mail: andrew.campbell@berkeley.edu
Mogeng Yin
University of California, Berkeley, e-mail: mogengyin@berkeley.edu
Sid Feygin
University of California, Berkeley, e-mail: sid.feygin@berkeley.edu
Alexei Pozdnukhov
University of California, Berkeley, e-mail: alexeip@berkeley.edu

survey participants, but are limited in four critical ways. 1) One-off travel surveys are expensive, both in terms of monetary and time costs. 2) Sample sizes are very small compared to the populations they are supposed to represent. 3) Travel surveys are not longitudinal. In most instances, a travel diary is collected for only one or two weeks, making it difficult to capture seasonal effects. 4) Travel surveys are conducted infrequently. For example, the most recent household travel survey of the nine counties of the San Francisco metropolitan region was conducted in 2000 (BATS, the Bay Area Travel Survey) [16]. This data set still serves as the modeling baseline for many county and city agencies (though the data are augmented with even smaller project-specific studies) [15]. Given the accelerating pace of global urbanization, metropolitan economies and land-use evolve too rapidly to use traditional travel survey methods for building demand models. Transportation practitioners need more temporally relevant data sources.

The rise of ubiquitous mobile computing has created an opportunity to use large-scale passively collected location data that addresses these four shortcomings. Signals such as geotagged social media posts and cellular network data, provide a continuous, longitudinal stream of travel behavior information at a very large sampling scale. In particular, mobile phone data are a very attractive source due to both the scale of adoption and frequency of usage. In the United States, it is estimated that adult cell phone adoption rate exceeds 90% [19]. Cellular network providers continuously log antenna connection events for billing purposes. These logs provide a near real-time data stream of approximate location. Since the data are linked with a unique account identifier, this stream becomes a longitudinal sample of mobility data for the lifetime of the users account. Large volumes of cellular data records (CDRs) are generated daily and can be queried with relatively little time latency (as compared to travel surveys). This allows for the training and calibration of travel demand models using very large-sample, longitudinal data that is temporally accurate.

We sought to quantitatively evaluate the efficacy of using cellular data to train and calibrate travel demand models in the context of the San Francisco Bay Area. The metropolitan region consists of nine counties and 7.2 million inhabitants. An experimental approach was taken. As a baseline for model performance comparison, we used the Metropolitan Transportation Commissions (MTC) Travel Model to generate activity chains for a synthetic population. We used MATSim to simulate mobility and generate volume and link performance estimates for model evaluation. The baseline MTC Travel Model results were compared with two models using cellular data. One model used only cellular data to generate activity chains. The other was a hybrid model that combined the MTC Travel Model, American Community Survey census data, and cellular data.

2 Background

Cellular data has been used extensively to develop models of travel demand, and to a lesser extent, network performance. On the demand side, several studies employ cellular data for analysis of human mobility and activities [20] [9] [11] [12]. These projects focus on measures of travel distance (e.g. radius of gyration), or prediction of labeled location sequences. On the supply side, there is limited research utilizing cellular data for transportation network analysis. A recent submission to MIT's NetMob conference Data for Development challenge explores using cellular data for estimating highway traffic volumes in Senegal [13]. Traffic volumes on rural roads are predicted using linear and support vector regression. This work is limited by only being able to model links that are sufficiently covered by the cellular network and remote enough to avoid confounding with pedestrian traffic.

By implementing a full integrated travel model (ITM), the SmartBay project utilizes both state of art demand and supply modeling. A machine learning Activity-Based Travel Demand Model (ABTDM) is trained on cellular data to generate sequences of activities, locations, and timings for a large synthetic population. Traffic conditions are simmulated and analysed at all locations and levels of the road network through the Agent-Based Microsimulation (ABM) traffic assignment model. The two halves of the ITM opperate in an iterative feedback cycle, Fig. 1. The system reflects the most up-to-date travel behaviors of the local population and is sensitive to supply-side changes enabling what-if testing for policy and operations decision makers. For a full discussion of the advantages of ABTDMs and their integration with network assignment models, we refer the reader to the 2015 report by Castiglione et al. [7].

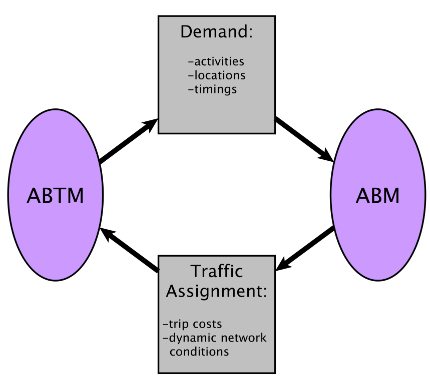

Fig. 1 The ITM Cycle

3 Data

Implementing an integrated travel model requires a very large outlay of work-hours for data collection, cleaning, integration and tuning. In public agencies, this is often a continuous work in progress as data must be ever updated to reflect the current state of the population and transportation supply.

The SmartBay project encompasses the entire San Francisco Bay Area: a 7,000 sq-mi metropolitan region spanning the nine counties under the jurisdiction of the Association of Bay Area Governments. The 2014 population of the Bay Area was estimated to be 7.5M residents [5]. Location data for a very large sample of the population was collected by a major cellular network provider as part of their routine billing practices. This was used to train the IOHMM demand model. Crowdsourced and public agency data was collected to build and callibrate the network model.

3.1 Demand Data

An ABTDM for demand generation is one half of the ITM framework. We employed two demand models: the MTC Travel Model, and the IOHMM developed by our team. Both models are used to generate input datasets for MATSim: files describing the daily activity sequence for every agent in the population. The data sources for training and callibrating the models differ greatly. The MTC model used datasets collected and published by government agencies ranging from federal to local. Some of the datasets are over a decade old at the time of this writing. The IOHMM approach used CDRs from the actual case study time perdiod, summer 2015.

3.1.1 MTC Travel Model

We use the MTC Travel Model as a baseline demand model. It is an ABTDM developed by Parsons Brinckerhoff, Inc. under contract for the MTC. It is a member of the Coordinated Travel - Regional Activity Modeling Program (CT-RAMP) family of models [10]. The model development, calibration and validation process is described in a 2012 report [15]. Agent populations were synthesized using historical and forecasted census and socio-economic distributions. The 2000 US Census Public Use Microdata Sample (PUMS) was used for generating empirical distributions of eight person types and four household types employed by the model. Aggregated TAZ-level socio-economic distributions from the year 2000 were provided by the Association of Bay Area Governments (ABAG). The baseline model used population distributions from the year 2000. Future scenario populations were drawn using iterative proportional fitting (IPF) with forecasted distributions of TAZ-level person and household categories and socio-economic variables.

The activity segmentation was based on the 2000 Bay Area Travel Survey (BATS). The 16 original activity categories from BATS were aggregated into 10 types for the Travel Model. All major agent decision making, excepting traffic assignment were modeled using a sequence of multinomial logit choices ranging in scope from work and school location to intra-tour mode choices. MTC Travel Model was calibrated and validated against the Caltrans State Highway traffic count databases.

For our baseline modeling benchmark, we sample agent unimodal activity chains directly from the MTC Travel Model.

3.1.2 IOHMM Demand Model Data

Contrary to the more traditional datasets employed by the MTC Travel Model, the IOHMM developed by our team only uses CDRs collected by a major cellular network provider in the San Francisco Bay Area. These data have two advantages: they are collected during the actual case study time period, and the sampling frame is very large percentage of the population, on the scale of 25%. Here we briefly give a high-level overview of the data and processing. For a complete discussion of the data processing and IOHMM model, we refer the reader to Yin et al. (2016), [14].

For accounting and operations purposes, the cellular network providers collect call data records (CDRs). They are timetamped logs that record data connection and handoff events. Everytime a phone makes or receives a call, or sends or requests data, a CDR is logged. Logs are also created when a phone is in motion and the connection is handed off from one antenna to the other. The CDR data entries contain, among many other data points, unique identifiers for the antenna and the user account. The locations of the network antenna are known to the provider. Thus, the CDRs for a single user constitues a rough trace of his or her location history.

The raw CDRs are converted into sequences of stay points using a four step process. First, they are are grouped using a density-based clustering algorithm such as K-Means, Fig. 2. Second, an “oscillation graph” is constructed. A user at a fixed location may be handed off between multiple local antennas during the process of network load balancing. The oscillation graph is used to identify antenna clusters that frequently work in concert and to infer the actual location of the user. Third, outlier oscillations are filtered out. And fourth, stay locations are filtered out when the duration is below a predefined threshold.

The obvious challenge with using the CDR data is that it is not labeled. Unlike the traditional travel surveys used by the MTC model, the CDR data sets do not identify activity types or even their exact locations. The four steps described above

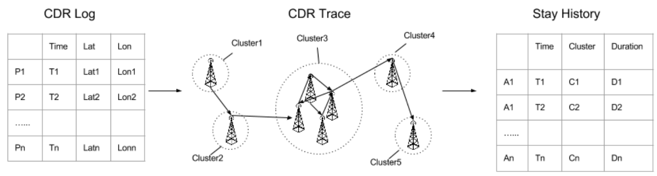

Fig. 2 Raw CRDs to Stay Histories [14]

are used to infer locations. Home and work activities are identified using a simple heuristic based on where a user most frequently is observed during typical home and work hours. All other activity labels are learned by the IOHMM. The activities are the latent states of the model and are characterized by the distributions of time of day, duration, and distances from home and work. Again, we refer the curious reader to Yin et al. (2016) for a detailed discussion of the IOHMM model development and results [14]. As we shall see later in this paper, the MTC and IOHMM demand models produce similar aggregate measures, but that the IOHMM data produces different localized behaviors such as county-to-county commute flows. We believe these reflect changes in the population and land use not reflected in outdated data sources.

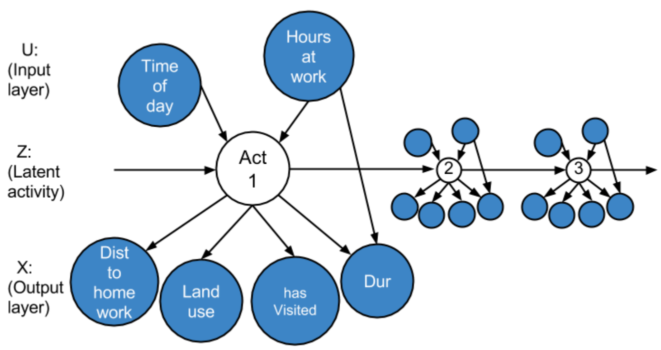

Fig. 3 IOHMM Specification [14]

3.2 *Transportation Network Data*

The MATSim ABM traffic assignment model requires a detailed representation of the actual road network. The MATSim road network was created using OpenStreetMap (OSM) road network data, downloaded in July, 2015 [17]. The user-generated OSM data offers very complete coverage in major metropolitan regions as well as rich feature sets including: link distance, number of lanes, speed limit, and hierarchical road classification. A manual inspection of sampled freeway links in California found the OSM features to be accurate.

The data was clipped and filtered using Osmosis, an open source Java application for editing OSM data. The OpenStreetMap Standards and Conventions define tags for classifying roads hierarchically. There are 14 tags which encompass nearly all road links in the dataset. These range from motorway and trunk down to residential and smaller hierarchical classes [18]. We found that for the Bay Area, the residential links constitute 74% of all links in the network. By filtering out the residential links, we were able to greatly improve the computational running time of MATSim without compromising regional-scale demand patterns. It is possible to maintain residential links for a localized area for future studies which require accurate neighborhood-level traffic reproduction. However, limiting factors, such as the realism of MATSims queueing, traffic signal, and physics engines call into question the efficacy of including the lowest hierarchy links in the network

The MATSim network is represented as a set of straight-line links and the nodes that connect them. To maintain realistic travel time skims, attributes of the original OSM network are preserved as attributes of the links: length, free-flow travel speed, and capacity (as imputed from the number of lanes). The final network used in the Smart Bay studies includes 564,368 links and 352,011 nodes.

3.3 Network Performance Data

While academic journal publications focus on the theoretical contributions, the work-hours required to collect and clean data and callibrate simmulations cannot be understated. Calibration of link performance functions and validation of the traffic volumes was conducted using freeway traffic counts from the Caltrans Performance Monitoring Systems (PeMS) [6]. We downloaded all 5-minute count data for Caltrans District 4 for the complete months of June, July and August 2015. Data for a total of 3,322 vehicle detector stations (VDS) was collected. It consists of hundreds of text files requiring almost 10GB of storage. They data were downloaded and processed using PeMS Tools, and open source Python package developed by one of the authors [3].

Although there as thousands of VDSs in District 4, there are many issues in the data quality to consider. During the study period, less than half of the stations were fully functional. Although the PeMS system employs heuristics to impute missing data, we wanted to calibrate and validate against only against true data. To this end, we employed six levels of filtering:

1. **Date Range** - Only stations that were active during the study period were used.

2. **Link Matching** - Only sensors that could be matched to within 100 feet of a link in the MATSim network were kept.
3. **Missing Data** - Although a station may be active during the study period, many have significant periods of no reported counts. We removed all stations with more than one hour of missing data.
4. **Boundary Buffer** - To avoid edge effects, we removed any stations that were within 10-km of the study are boundary.
5. **Outlier Detection** - There are frequent "false outliers" in the PeMS data. These days are unusual due to errors in the PeMS reporting system instead of actual changes in traffic patterns. For example, a station may be reporting 100% of lanes are observed, but it will report the same volume of traffic in every time bin for that day. Another common error is gross under reporting of flow even when no accidents or construction were present. We used a one-class support vector machine (detailed below) to identify and remove stations with false outliers.
6. **Fraction Observed** - The 'Observed' attribute in the data describes the fraction of lanes for which the detectors were working during that reporting period. We only kept stations where at least 50% lanes were observed during every day of the study period.

After applying the six filters, only 774 usable stations remained. In Table 1 we see the results of sequentially applying the filters. It is important to note that each removed station is only attributed to one filter. Once a filter was triggered, the station was removed and the following filters were not applied. Although the total number of removed stations would stay the same, changing the order of filter applications would change the results in Table 1.

Table 1 VDS Filtering Results

Date Range	206
Link Matching	119
Missing Data	216
Boundary Buffer	185
Outlier Detection	200
Fraction Observed	1,622

3.3.1 One-Class SVM Outlier Detection

A one-class SVM was used to identify "false outliers" among each station that satisfied the first four filters. For a single station, we created scaled and centered feature matrix, X. A single day of hourly volumes constitutes one point in a 24-dimension space. The feature matrix is thus in $\mathbb{R}^{n \times 24}$ space, where n is the number of days in the study period. We train the SVM on X to define a separation plane. The most typical days fall within the space defined by the plane and have a positive distance from the plane. The outliers fall outside and have negative distance. The more negative the distance from the separation plane is, the more unusual a day is. From Figure 4, we see that 99% of days are closer than -5.0 from the plane.

We manually inspected the data for 10 of the stations found to have days beyond the -5.0 distance. It was decided that if more than 5% of days for that station were beyond that distance, then the station would be removed.

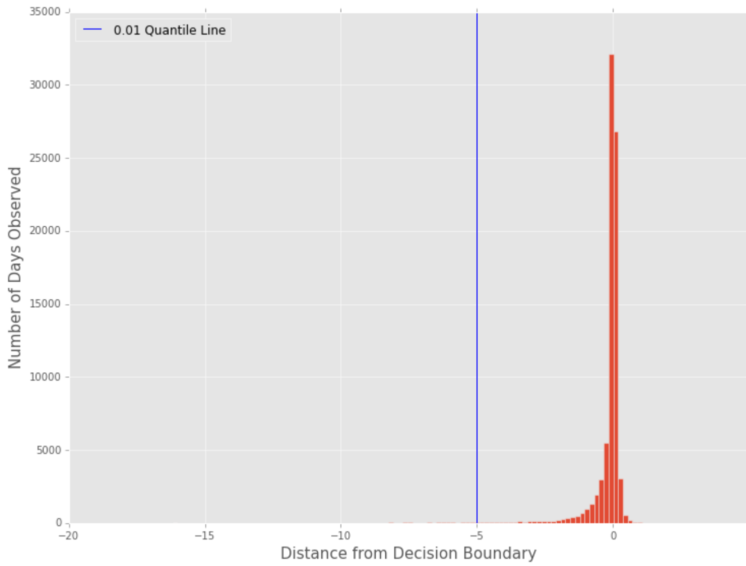

Fig. 4 One-Class SVM Outlier Detection - Distribution of Distances from Separation Plane

4 ITM Implementation

The SmartBay ITM is implemented as two separate interfacing models, Fig. 1. An ABDTM generates travel demand and traffic assignment is executed by the MATSim ABM. We evaluate two different ABTDM demand models: the MTC Travel Model, Sec. 3.1.1, and the IOHMM developed by our team, Sec. 3.1.2. Both demand models are used to generate activity sequences for one million agents. The demand models should represent typical weekday travel behavior for the San Francisco Bay Area during the summer of 2015.

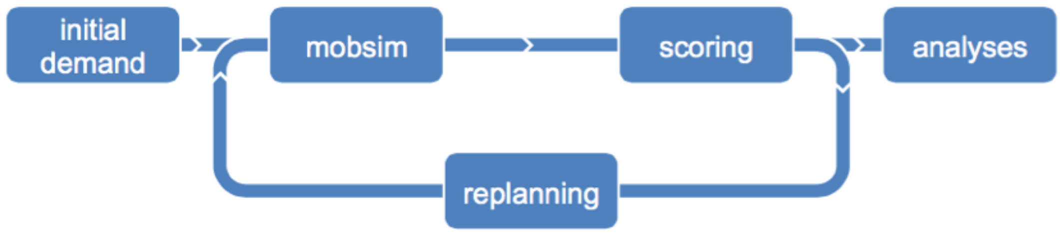

Fig. 5 The MATSim Cycle [4]

4.1 MATSim Platform

The MATSim (Multi-Agent Transport Simulation) platform is an agent-based activity model that performs microscopic modeling of traffic (using queue-based link performance functions) and agent decision making [4]. The MATSim run cycle, Figure 5, is an iterative process whereby agents make adaptations to routing, activity timing, and other optional choices until convergence is reached. As input, each agent is assigned an activity chain (initial demand), consisting of activity types, timing and locations. During the mobility simulation (mobsim), the agents travel the network, interact, and experience congestion that lowers their overall utility scores for the day. During the replanning phase, a subset of agents may adapt their routes and activity timings. For our simulations, we restricted replanning adaptation to random selection of 10% of the population during each iteration. Many other forms of adaption are possible with MATSim, but for this project we have restricted adaptation to timing and routing. Agents incur a negative penalty for deviating from their original activity timings, so dramatic shifts in activity start and end times are not possible. Rerouting agents are allowed to update their routes to the new shortest path, based on the loaded network conditions in the most recent mobility simulation.

We used the MTC and IOHMM models to generate initial demand for a typical weekday. The scenarios simulate a single 24-hour day for 1M agents, scaled up to represent the total driving population. For this initial implementation, we cast all agent demand into private passenger-car equivalents. Thus each simulated car only carries a single agent, but it may represent more than one passenger trip.

4.2 Microsimulation Calibration

For calibration, we used the full set of 774 PeMS sensor stations that passed the filtering procedure in Sec. 3.3. Simulation performance was evaluated in terms of aggregate volume error and aggregate travel times.

The aggregate volume error was calculated using the root mean squared error (RMSE) calculated over all sensors over all hours of a 24-hour day:

$$RMSE = \sqrt{\frac{\sum_{s\in\mathscr{S}}\sum_{h=1}^{24}(\alpha\hat{y}_{s,h} - y_{s,h})^2}{\sum_{s\in\mathscr{S}}\sum_{h=1}^{24} 1}} \tag{1}$$

where $\mathscr{S}$ is the set of all vehicle detector stations used in calibration, and $\hat{y}_{s,h}$ and $y_{s,h}$ are the simulated and measured counts for station s at hour h. The scalar coefficient α is a MATSim tuning parameter called the Counts Scale Factor. Since the number of agents may not be the same as the real population, the Counts Scale Factor is used scale simulated counts to best match the real world counts. The Counts Scale Factor has no impact on the actual microsimulation and is thus adjusted post-simulation to minimize the RMSE. Taking the derivative of equation 1, the optimal factor, α^*, is found:

$$\alpha^* = \frac{\sum_{s\in\mathscr{S}}\sum_{h=1}^{24}\hat{y}_{s,h}y_{s,h}}{\sum_{s\in\mathscr{S}}\sum_{h=1}^{24}\hat{y}_{s,h}\hat{y}_{s,h}} \tag{2}$$

MATSim provides two main levers for calibrating link performance during the microsimulation: the Flow Capacity Factor (FCF) and the queue Storage Capacity Factor (SCF). The Flow Capacity Factor dictates how many vehicle may pass through a link during a simmulated time step. The Storage Capacity Factor controls how many vehicles can be stored in queue before no more vehicles can enter the link. The role of the two factors is succinctly described by Nurhan et al. (2003): "link outflows are constrained both by the flow capacity of the link itself and by space limitations on the receiving link" [8].

There is a no deterministic method for flow and storage factor values, but a good initial guess is to use the agent-to-population ratio. With our demand models, we have 750k commuters, out of a total 3.5M for 2015 [1]. Rounding down, we chose 0.21 as the initial guess for both the flow and storage factors. We gridded the factors from 0.15–0.25 and ran a total of 36 callibration experiments using the MTC AB-DTM to generate demand, Fig. 6 and Fig. 7.

The counts RMSE, as defined in Eq. (1), is minimized for $(FCF, SCF) = (0.17, 0.25)$. However, at this value the home-to-work average commute time is 69-minutes. This is too high. The MTC estimates that in 2015 the average home-to-work commute time was 30-minutes [2]. For our final configuration, we chose $(FCF, SCF) = (0.23, 0.25)$. At this point, the RMSE is only 3.1% higher and the home-to-work commute time is reduced to 39-minutes. The MTC commute time estimate is based on aggregate self-reported measures from the deccenial Census and American Community Survey. A better approach to travel time validation would be

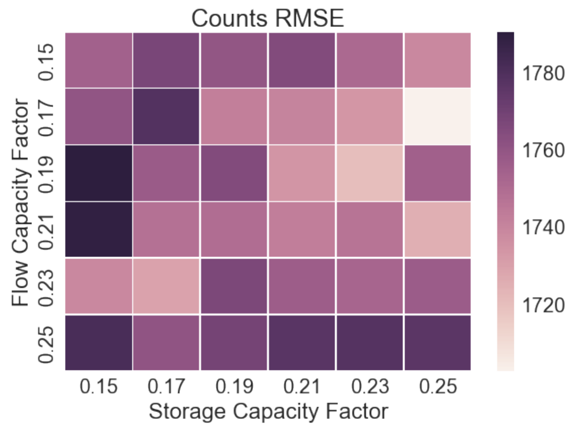

Fig. 6 MATSim Traffic Counts Callibration

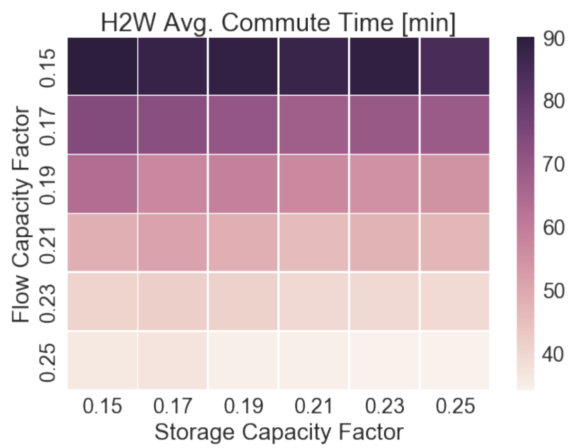

Fig. 7 MATSim Travel Time Callibration

to gather link-level actual travel times and minimize the RMSE in the same fashion we did for counts.

5 Results and Conclusions

NEW INTRO PARGRAPH: This is the is the first ever ITM using an ABDTM trained only on cellular data. It performs comporably to the MTC model without using any external data. Here are all the great measure we have...

5.1 Validation

The SmartBay project models traffic at a level of a large urban region. For this reason, we validated against macroscopic measures of volume, travel time, delay, distance and spatial distribution. Since the MTC Travel Model is used used as our control treatment, we validate against the measures the MTC uses for evaluating the performance of the Bay Area road system.

5.1.1 Regional Gateway Screenlines

We validated simulation traffic volume on major freeways at regional gateway screenlines located at county boundaries. This reproduces Tables 69 and 70 in the MTC Travel Model calibration and validation report [15]. These tables describe AM peak and PM peak predicted and observed flows for a typical weekday. The AM peak is defined as 6:00 - 09:59, and the PM peak is 16:00 - 18:59. In Tables 2, and 3, we have used our PeMS typical weekday profiles for the observed values and the MATSim simulated volumes for the predicted counts. We have included 'NA' placeholders for locations that were included in the MTC report, but for which no PeMS data was available. In many instances, there were no filtered sensors available at the exact location specified in the report. If a sensor did exist within 10km of the location without any major on or off ramps in between, it was used as a stand in for the original screenline location.

References

1. Commute Patterns | Vital Signs.
2. Commute Time | Vital Signs.
3. neverforgit/PeMS_tools.
4. Andreas Horni, Kai Nagel, and Kay Axhausen, editors. *The Multi-Agent Transport Simulation: MATsim*. April 2016.
5. California Department of Finance. E-1 Population Estimates for Cities, Counties, and the State - January 1, 2015 and 2016, May 2016.
6. Caltrans. Caltrans PeMS, 2015.
7. Joe Castiglione, Mark Bradley, and John Gliebe. Activity-Based Travel Demand Models: A Primer. *SHRP 2 Report*, (S2-C46-RR-1), 2015.
8. Nurhan Cetin, Adrian Burri, and Kai Nagel. A large-scale agent-based traffic microsimulation based on queue model. In *IN PROCEEDINGS OF SWISS TRANSPORT RESEARCH CONFERENCE (STRC), MONTE VERITA, CH*, pages 3–4272, 2003.
9. Eunjoon Cho, Seth A Myers, and Jure Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1082–1090. ACM, 2011.
10. William Davidson, Peter Vovsha, Joel Freedman, and Richard Donnelly. CT-RAMP family of activity-based models. In *Proceedings of the 33rd Australasian Transport Research Forum (ATRF)*, 2010.

Screenline Facility	Observed	Predicted	Predicted Less Obsv'd	Percent Difference
Bay Area Bridges				
US-101, Golden Gate Bridge (S)	NA	NA	NA	NA
I-80, SF/Oakland Bay Bridge	19,290	23,659	4,369	22.6%
Cal-92, San Mateo/Hayward Bridge (W)	23,272	23,923	651	2.8%
Cal-84, Dumbarton Bridge (N)	12,673	21,104	8,431	66.5%
I-580, Richmond/San Rafael Bridge (E)	13,788	13,183	-605	-4.4%
I-80, Carquinez Bridge (E)	17,864	18,428	564	3.2%
I-680, Benicia/Martinez Bridge	17,864	18,428	564	3.2%
Cal-160, Antioch Bridge	NA	NA	NA	NA
Bay Area Bridges Sub-Total	***104,751***	***118,725***	***13,974***	**13.3%**
San Francisco / San Mateo Line				
US-101, Bayshore Freeway (N)	21,981	38,556	16,575	75.4%
Cal-35, Skyline Blvd. (N)	NA	NA	NA	NA
Cal-1, Junipero Serra Blvd. (N)	NA	NA	NA	NA
I-280, Foran Freeway	NA	NA	NA	NA
SFM/SM County Line Sub-Total	***21,981***	***38,556***	***16,575***	***75.4%***
San Mateo / Santa Clara County Line				
Cal-82, El Camino Real (N)	NA	NA	NA	NA
US-101, Bayshore Freeway (N)	17,168	16,749	-419	-2.4%
I-280, Serra Freeway (N)	NA	NA	NA	NA
SM / SC County Line Sub-Total	***17,168***	***16,749***	***-419***	***-2.4%***
Santa Clara / Alameda County Line				
I-680, at Scott Creek Road (N)	27,700	22,849	-4,851	-17.5%
I-880, Nimitz Freeway (N)	18,759	34,790	16,031	85.5%
SC / Ala Line Sub-Total	***46,459***	***57,639***	***11,180***	***24.1%***
Alameda / Contra Costa County Line				
I-580, Knox Freeway	17,467	20,094	2,627	15.0%
I-80, Eastshore Freeway	38,447	45,406	6,959	18.1%
Cal-24, Caldecott Tunnel (E)	34,367	29,654	-4,713	-13.7%
I-680, in Dublin/San Ramon	33,483	28,017	-5,466	-16.3%
Ala / CC County Line Sub-Total	***123,764***	***123,171***	***-593***	***-0.5%***
Solano / Napa County Line				
Route 29, nodatapa-Vallejo Highway (N)	NA	NA	NA	NA
Solano / Sonoma County Line				
Route 37, Sears Point Road	7,526	11,858	4,332	57.6%
Napa / Sonoma County Line				
Route 121, Carneros Highway (N)	NA	NA	NA	NA
Route 128, Calistoga-Healdsburd Rd. (E)	NA	NA	NA	NA
Sonoma / Marin County Line				
US-101, Redwood Highway (N)	27,152	27,659	507	1.9%
Screenline Totals	**348,801**	**394,357**	**45,556**	**13.1%**

Table 2 Screen Line Validation for AM Peak

Screenline Facility	Observed	Predicted	Predicted Less Obsv'd	Percent Difference
Bay Area Bridges				
US-101, Golden Gate Bridge (S)	NA	NA	NA	NA
I-80, SF/Oakland Bay Bridge	33,366	33,622	256	0.8%
Cal-92, San Mateo/Hayward Bridge (W)	32,521	31,053	-1,468	-4.5%
Cal-84, Dumbarton Bridge (N)	16,084	23,611	7,527	46.8%
I-580, Richmond/San Rafael Bridge (E)	18,035	12,729	-5,306	-29.4%
I-80, Carquinez Bridge (E)	24,755	28,965	4,210	17.0%
I-680, Benicia/Martinez Bridge	24,755	28,965	4,210	17.0%
Cal-160, Antioch Bridge	NA	NA	NA	NA
Bay Area Bridges Sub-Total	***149,516***	***158,945***	***9,429***	***6.3%***
San Francisco / San Mateo Line				
US-101, Bayshore Freeway (N)	26,505	41,233	14,728	55.6%
Cal-35, Skyline Blvd. (N)	NA	NA	NA	NA
Cal-1, Junipero Serra Blvd. (N)	NA	NA	NA	NA
I-280, Foran Freeway	39,310	31,953	-7,357	-18.7%
SFM/SM County Line Sub-Total	***65,815***	***73,186***	***7,371***	***11.2%***
San Mateo / Santa Clara County Line				
Cal-82, El Camino Real (N)	NA	NA	NA	NA
US-101, Bayshore Freeway (N)	22,410	20,592	-1,818	-8.1%
I-280, Serra Freeway (N)	NA	NA	NA	NA
SM / SC County Line Sub-Total	***22,410***	***20,592***	***-1,818***	***-8.1%***
Santa Clara / Alameda County Line				
I-680, at Scott Creek Road (N)	31,600	34,601	3,001	9.5%
I-880, Nimitz Freeway (N)	20,208	43,304	23,096	114.3%
SC / Ala Line Sub-Total	***51,808***	***77,905***	***26,097***	***50.4%***
Alameda / Contra Costa County Line				
I-580, Knox Freeway	24,495	25,900	1,405	5.7%
I-80, Eastshore Freeway	51,146	60,269	9,123	17.8%
Cal-24, Caldecott Tunnel (E)	44,810	42,137	-2,673	-6.0%
I-680, in Dublin/San Ramon	45,101	38,975	-6,126	-13.6%
Ala / CC County Line Sub-Total	***165,552***	***167,281***	***1,729***	***1.0%***
Solano / Napa County Line				
Route 29, nodatapa-Vallejo Highway (N)	NA	NA	NA	NA
Solano / Sonoma County Line				
Route 37, Sears Point Road	11,234	8,852	-2,382	-21.2%
Napa / Sonoma County Line				
Route 121, Carneros Highway (N)	NA	NA	NA	NA
Route 128, Calistoga-Healdsburd Rd. (E)	NA	NA	NA	NA
Sonoma / Marin County Line				
US-101, Redwood Highway (N)	46,394	27,643	-18,751	-40.4%
Screenline Totals	**512,729**	**534,404**	**21,675**	**4.2%**

Table 3 Screen Line Validation for PM Peak

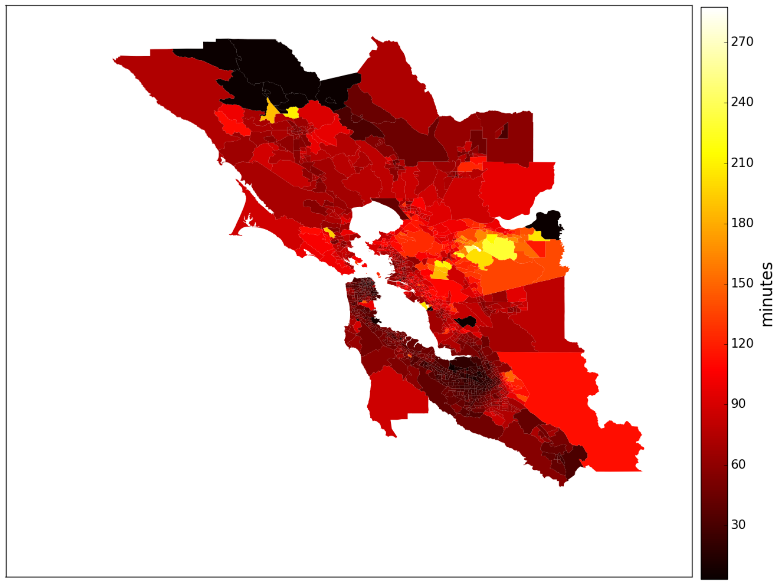

Fig. 8 Total Commute Times by TAZ

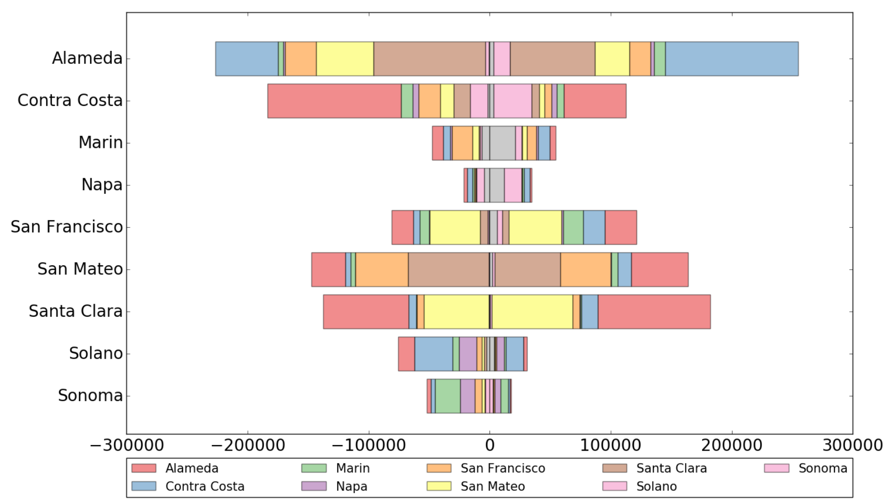

Fig. 9 County-to-County Commute Flows. Negative numbers are outbound flows. Positive numbers are inbound flows.

Table 4 Home-to-Work Total Commute Times [minutes]

	Total Time	Delay Time	Time in Congestion
Alameda	78.0	61.1	59.6
Contra Costa	125.1	100.8	98.9
Marin	81.5	63.0	61.3
Napa	61.9	41.4	39.8
San Francisco	38.8	26.7	25.4
San Mateo	49.1	34.0	32.5
Santa Clara	46.0	32.5	31.3
Solano	80.2	58.6	56.8
Sonoma	84.0	61.9	60.6
TOTALS	69.9	52.6	51.2

Table 5 Home-to-Work Total Commute Distances [miles]

	Total Dist	Dist in Congestion
Alameda	15.7	4.9
Contra Costa	21.3	7.0
Marin	16.8	5.6
Napa	17.6	4.1
San Francisco	11.0	2.9
San Mateo	13.9	3.9
Santa Clara	12.3	3.4
Solano	20.4	4.9
Sonoma	17.6	5.6
TOTALS	15.5	4.6

11. Vanessa Frias-Martinez, Jesus Virseda-Jerez, and Enrique Frias-Martinez. On the relation between socio-economic status and physical mobility. *Information Technology for Development*, 18(2):91–106, April 2012.
12. Marta C. Gonzlez, Csar A. Hidalgo, and Albert-Lszl Barabsi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, June 2008.
13. Tony Liang and Vanessa Frias-Martinez. Cars and Calls: Using CDR Data to Approximate Official Traffic Counts. In *NetMob D4D Challenge submissions*, 2015.
14. Mogeng Yin, Madeline Sheehan, and Danqing Zhang. A Generative Model of Urban Activities from Cellular Data [DRAFT]. May 2016.
15. MTC. Travel Model Development: Calibration and Validation, May 2012.
16. MTC - Data Portal. Bay Area Travel Survey (BATS), 2000.
17. OpenStreetMap. Mapzen metro extracts - San Francisco Bay, July 2015.
18. OpenStreetMap Wiki. Highways, March 2016.
19. Pew Research Ceneter. Device Ownership Over Time, April 2015.
20. Dashun Wang, Dino Pedreschi, Chaoming Song, Fosca Giannotti, and Albert-Laszlo Barabasi. Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1100–1108. ACM, 2011.