Bicycle Route Choice Data Collection using GPS-Enabled Smartphones

Billy Charlton *(corresponding author)*
San Francisco County Transportation Authority
100 Van Ness Avenue, 26th Floor
San Francisco, CA 94102
billy.charlton@sfcta.org
(415) 522-4816

Elizabeth Sall
San Francisco County Transportation Authority
100 Van Ness Avenue, 26th Floor
San Francisco, CA 94102
elizabeth.sall@sfcta.org
(415) 522-4810

Michael Schwartz
San Francisco County Transportation Authority
100 Van Ness Avenue, 26th Floor
San Francisco, CA 94102
michael.schwartz@sfcta.org
(415) 522-4800

Jeff Hood
San Francisco County Transportation Authority
100 Van Ness Avenue, 26th Floor
San Francisco, CA 94102
(415) 522-4800
jeffnhood@gmail.com

Keywords: Bicycle Route Choice, GPS, Data Collection, iPhone

Word Count:

Text:	3,800
Tables: 3 x 250 =	750
Figures: 1 x 250 =	250
Total Word Count:	**4,800**

Submitted for Presentation at the 2011 Transportation Research Board Annual Meeting.
Original Submission Date: August 1, 2010; Revised November 15, 2010.

ABSTRACT

The proliferation of consumer-grade smartphones with global positioning system (GPS) location capabilities opens a new data collection method for researchers. When confronted with a lack of data on bicycle routes preferred by local cyclists, the San Francisco County Transportation Authority (SFCTA) developed a freely downloadable iPhone/Android smartphone "app" called CycleTracks to collect actual bicycle routes traversed by city cyclists. Cooperation with local bicycle advocacy groups, along with social media and email campaigns, encouraged use of the app by regular citizenry. Several rounds of pre-release testing showed that making the app start up quickly, and minimizing battery usage during recording, were critical to getting good data. Once installed on a user's smartphone, a single "tap" would start and stop recording a bicycle trip; after completing a trip, the app automatically uploaded the track to a central database/web server, via the phone's built-in data plan. Approximately 5,000 usable bicycle trips were collected from hundreds of users in the region. Demographic data was optionally provided by some users, and showed a bias toward frequent cyclists, and toward male users (even more than cycling is already male-dominated in the region's most recent household travel survey). A bicycle route choice model developed using the data revealed sensitivity to slope, presence of bike lanes and/or bike route designations, trip purpose, and gender. The bike route choice model is now being integrated into San Francisco's regional travel model.

INTRODUCTION

Nonmotorized modes such as walking and biking play an increasing role in daily travel in major metropolitan areas, accounting for 9% and 4% respectively of work tours by San Francisco residents in 2000 (Bay Area Travel Survey 2000). Because of their relatively low costs, negligible emissions, and high potential public health and welfare benefits, the San Francisco County Transportation Authority (SFCTA) and many other public agencies are actively seeking ways to promote even more nonmotorized tripmaking. Accordingly, attention to nonmotorized modes within the regional travel demand models has increased. It is now standard practice in many urban areas to include walking and biking within mode choice models; however, due to lack of data, little is known about the routes and actual infrastructure that are used. There are two main benefits to adding a bicycle route choice model to a regional travel demand model: (1) decisionmakers can geographically target bicycle-related capital investments and analyze operational improvements; and (2) travel demand models can quantify the improved accessibility benefits from nonmotorized infrastructure, leading to better forecasts.

SF-CHAMP is San Francisco's tour-based travel demand model, used citywide to quantify benefits and understand the implications of various transportation projects, plans, and policies (Outwater & Charlton 2006). Like many advanced travel models, SF-CHAMP is currently able to forecast the quantity of bicycle trips, but cannot assign these trips to specific streets. To remedy this deficiency, the California Department of Transportation (Caltrans) awarded SFCTA a State Planning and Research Grant to develop a bicycle route choice component for SF-CHAMP. The results of that research are in the companion paper (Hood et al. n.d.).

This paper focuses on the data collection effort in support of the route choice model. Predicting a selected route from a set of available routes requires sufficient data on the routes cyclists actually prefer. This paper details SFCTA's strategy of utilizing the satellite-based global positioning system (GPS) capability of consumer-grade smartphones.

REVIEW OF ROUTE CHOICE DATA COLLECTION TECHNIQUES

Before committing to a smartphone-based strategy, several data collection techniques from existing bicycle route choice studies were reviewed: (1) web-based stated preference surveys; (2) route recall; (3) personal GPS devices; and (4) bicycle-mounted GPS devices.

Web-based stated preference surveys have the advantage of being quick and inexpensive. In addition, the stated preference nature allows the surveyor to force the respondent to make certain tradeoffs to lower the needed sample size, and base those tradeoffs on respondents' answers to previous questions. However, estimating a model from such forced tradeoffs is not nearly as reliable as from a good revealed choice data set where many biases can be eliminated. (Sener, Eluru, and Bhat 2008)

Non-GPS methods such as route recall are costly on a per-record basis, and are prone to human error in description and translation. McDonald and Burns (2001) used route-recall for respondent's "most recent commuting route." An additional problem with this methodology is that it is unlikely to capture much stochasticity among individuals that occurs in the route selection process over the course of several days.

Special-purpose GPS devices are small and lightweight enough to carry on-person for long time periods. Price has also come down significantly in the past ten years and individual units can be obtained for less than $100. However, if the user is not continually prompted for mode information the GPS points for bicycle modes can be confused with various other modes. Meghini et al (2009) gleaned 2,657 bike routes from a personal GPS dataset with 11,000 total trips and 2,435 people. Dill and Gliebe used bicycle-mounted GPS devices to record bicycle routes of 164 adults in Portland, Oregon. The bike-mounted GPS eliminated much of the data-cleaning required by the personal GPS devices, but researchers still had to trim portions of routes where the cyclist was on transit. The participants used the bike-mounted GPS device for seven days, at the end of which the project team retrieved the unit and downloaded the GPS data. Doherty (2009) provided smartphones to participants to track second-by-second GPS locations, which were regularly uploaded wirelessly to a remote server and stored in a MySQL database. The smartphones operated continuously for 17 hours at a time with the help of extra battery packs. Wireless data uploads negated the need for researchers to spend much time in the field retrieving data.

The techniques discussed above were contrasted with SFCTA's project needs and constraints. The SFCTA team had the following desired features for a data collection instrument: revealed preference as opposed to stated-preference; GPS-native data format to reduce error and researcher hours; and inexpensive to deploy in terms of time (researcher hours) and money (capital investment). These attributes pointed to smartphone-based GPS data collection that could wirelessly transmit data to an SFCTA server. Instead of providing devices to participants, the team could also leverage the high smartphone penetration in San Francisco and allow participants to download the application to their own phone.

The iPhone was chosen as the initial deployment platform because it is easy to develop applications for, has a high market penetration, and has a userbase already accustomed to downloading "apps." After the initial version was operational, SFCTA also developed an Android version, which allowed coverage of all four major cell service providers. A BlackBerry version was also considered but ultimately dropped, because not all BlackBerry phones have access to GPS, there is no Google Maps programmatic interface available for BlackBerry, and because BlackBerry users are not as accustomed or willing to download third party apps. Given more time and resources, a BlackBerry version could be produced in the future, beneficial given its continued status as a widely-used smartphone platform in the U.S.

CONCERNS REGARDING DATA BIAS

Two main concerns arose immediately upon choosing this approach: (1) would the app be expensive to produce and then difficult to distribute; and (2) would the data be unacceptably biased in some manner. There was no way to know the answer to the first question without trying, so that concern was tabled. Regarding bias, the team considered many possibilities.

Smartphone users might choose different routes than non-smartphone users

It is known that smartphone users are generally more affluent than owners of less expensive phone models. Income is correlated in many ways to travel behavior, so it is reasonable to expect that this income disparity could affect results. The population of San Francisco is quite affluent (in general terms) compared to the national average, so in some ways this bias is almost citywide regardless of phone type. But, as the revealed-choice data is only being used to identify tradeoffs between the types of routes preferred (and thus the attributes of those routes), the real question here is whether smartphone users would prefer different *types* of routes than non-smartphone users. For example, would a smartphone user dislike hills more or less than a non-smartphone user? Would owning an iPhone make a user more likely to prefer bike lanes? Ultimately our design team decided that this seemed like a possibility, but not one which would bias results any more than selection bias in other techniques.

The team decided to request "personal info" including age and gender to help document and analyze the user base.

Users who download this app might be different from smartphone users who do not
Clearly, cyclists interested in improving civic life for other cyclists are most likely to use the app, and they could easily have different route preferences than infrequent cyclists. The team tried to reach out to both groups, as is described below. The app also requested typical cycling frequency to help researchers understand the user base.

Users might use the app for certain trips or certain routes, to "game" results
With any uncontrolled data experiment, this is a risk. The team decided to weight the dataset inversely proportional to the number of submitted trips by any one user, to help combat any concerted effort to skew results.

While this discussion in no way eliminates the possibilities for bias, the team felt that the risks were understandable, and the method worth pursuing. The next section describes the CycleTracks application, which SFCTA developed to collect the bicycle route choice data.

THE "CYCLETRACKS" IPHONE/ANDROID APPLICATION

Design goals for the app were few and simple: (1) It must be free and quick to download and install; (2) It must be as easy to use as possible, with minimum tapping/clicking necessary to get started, so even casual cyclists can use it; (3) Every track recorded should be uploaded immediately to our central database using the phone's built-in data plan, so the user doesn't have to manually intervene, sync, or upload anything; (4) It must not run down the user's battery; (5) It needs a catchy name. After brainstorming countless ideas, the team chose CycleTracks. There are no physical "cycle tracks" in San Francisco (i.e. separate rights-of-way protected from street traffic by parking), so there was no risk of confusion between our app CycleTracks and the term used by planners.

Freely available on the iTunes "App Store" and the Android Market, any user can download and install CycleTracks. To make the app something cyclists would want to download, care was taken to make the app not just useful for research purposes but also useful and possibly fun for cyclists. To that end, in addition to recording and sending GPS data to SFCTA servers, the app allows users to view maps of all the routes they have recorded, and track their distance and speed on each trip.

The CycleTracks user experience is designed to be as unobtrusive as possible. To record a trip, cyclists just tap **START** (see Figure 1). The app then connects to GPS satellites and begins recording the trip. The Android version records the GPS track in the background, and allows the user to continue using their phone for other purposes while recording. As iPhones did not have background GPS capabilities at the time of this research, the iPhone version automatically locks itself to prevent the participant from accidentally tapping something, dims the backlight to save battery life, and displays a timer to let the user know how long it has been since the trip started. (The most recent iPhones have finally caught up to Android in feature parity, and a modified version of the iPhone code could conceivably allow background operation for those newest phones.)

When a participant finishes their bike trip, they tap **FINISH,** choose a trip purpose from the eight options (Commute, Work-Related, School, Social, Shopping, Errands, Exercise, Other), and then tap **SAVE t**o upload the route to the SFCTA server. The user also has the option to discard the trip instead of saving it.

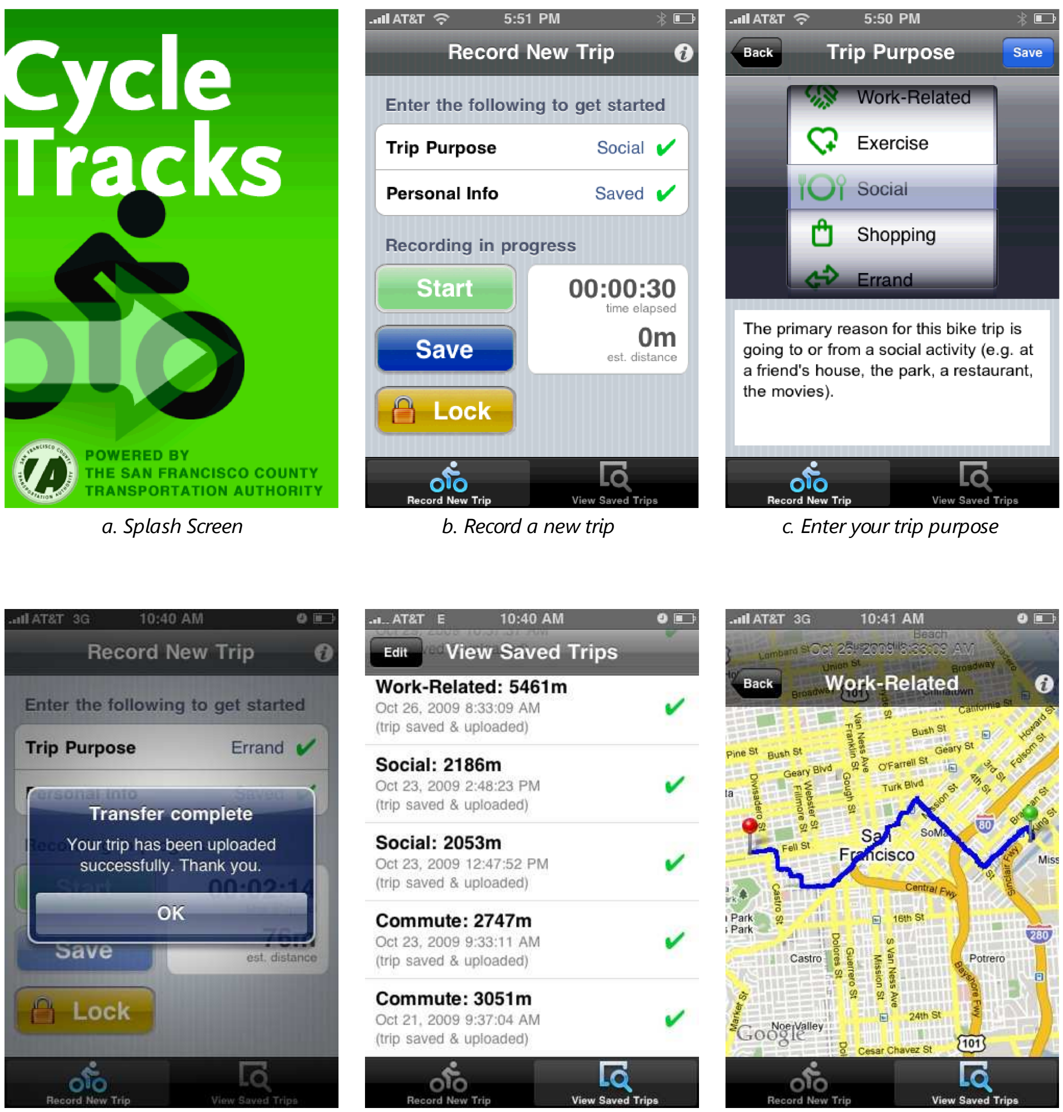
GPS data is saved locally on the device throughout the trip, and only uploaded at completion. Thus, a live data connection is not required during recording of the trip, although

GPS accuracy is contingent on a clear view of the sky. Any trips that did not successfully upload for any reason, can be re-uploaded later and are marked with an exclamation point. The participant will be prompted to upload them, or they will be automatically uploaded the next time a trip is recorded.

As a convenience to the user, all trips can be reviewed in a Google Maps view (Figure 1f) which also shows the trip purpose and length.

A beta testing period revealed battery usage issues correlated with heavy GPS radio usage. Recognizing that participants are voluntary, several modifications avoid draining the battery. To remind participants that it is collecting data, CycleTracks makes a "bicycle bell" noise and vibrates after an initial 15 minutes of GPS data collection and every five minutes thereafter. Additionally, CycleTracks turns itself off if the phone battery life gets below 10%, allowing participants enough battery life to make phone calls.

Figure 1. Screenshots of the CycleTracks Application (iPhone version shown)



a. Splash Screen *b. Record a new trip* *c. Enter your trip purpose*

d. Success! Trip uploaded. *e. List of saved trips for review* *f. Map of your bike trip*

SERVER SIDE DATABASE

CycleTracks uploads trip data to a MySQL database on an SFCTA server, via an Apache webserver and the "Javascript over network" (JSON) data transfer protocol. Three data tables are stored: Person (identified by their iPhone unique device ID), Trips, and GPS Coordinates. Data stored in each of these tables is outlined below.

Note that the only user-identifiable fields in the collected data are the phone hardware "IMEI" number and a voluntarily-provided email address. The IMEI field needs to be collected so that multiple trips by the same person can be identified and linked to the anonymous user ID, and is scrubbed from the final analysis data.

Trip data typically takes just a few seconds to upload from the phone to the database. All route-based analysis can be performed on the raw GPS coordinate data at a later time.

Table 1: Person Table

User ID	Numeric identifier for the person record
Created	Creation date/time for this user record
Device ID	IMEI number of phone hardware (stripped from final data)
Home ZIP	Home ZIP code*
School ZIP	School ZIP code*
Work ZIP	Primary workplace ZIP code*
Gender	Male/Female*
Age	Age in years of this person*
Biking Frequency	Daily / Several per week / Several per month / <1 per month*
Email	Email address for raffle & future contact*

* Fields with an asterisk are optionally provided by user

Table 2: Trip Table

Trip ID	To match to GPS Coordinates Table
User ID	To match to Person Table
Start Time	Time stamp for when user taps "start"
End Time	Time stamp for when user taps "stop"
Number of Coordinates	Number of non-null coordinates in this trip
Trip Purpose	Selected and confirmed by user (see Figure 1c)

Table 3: GPS Coordinates Table

Trip ID	To match to Trip Table
Time	Time stamp of when the GPS coordinate was taken. In addition to travel time and speed calculations, it is used to order points to determine route.
Latitude	Latitude and Longitude are stored with ten decimal points
Longitude	
Altitude	In meters
Estimated Accuracy	Provided by GPS system, in meters

TESTING RESULTS

In beta testing of the app, accuracy is usually sufficient in most Bay Area locations to identify the street being traversed. However, two cases repeatedly caused GPS problems: (1) at the start of a trip, the GPS receiver may not have a fully accurate lock on enough satellites to provide high-quality coordinates. The phone attempts to place the location anyway, resulting in a noisy start to many trips. This problem is usually rectified after about one minute of use. (2) A downtown "urban canyon effect" is noticeable in the high-rise portion of downtown San Francisco. The shadows of skyscrapers definitely block out some satellite signals, resulting in less accurate pinpointing.

Other than those two issues, the data stream seemed to be coming in cleanly and without unexpected problems.

CYCLETRACKS USAGE PROMOTION AND RECRUITMENT

Given its design as a general purpose app for everyday use by regular citizens (i.e. no official recruitment stage for participants), and given the unique nature of app without any real precedent from other agencies, the research team did not know how many people would realistically download and use the app, or whether the collected data would be skewed by very narrow interests using it.

From the very early stages of the project, SFCTA performed outreach to local bicycle advocacy organizations including the Bay Area Bicycle Coalition and the San Francisco Bicycle Coalition, along with sister agencies with an interest in the data.

The campaign to generate interest also included multiple newsletters and blog entries, press releases to local media and transportation-related blogs, a Facebook group, and several staff interviews on local and public radio.

To encourage use, these announcements included notice of multiple iTunes gift cards for users who submit at least one valid trip using the app. iTunes gift cards were a "no-brainer" incentive since fully 100% of users are known to already use the iTunes program.

Finally, once the app was released, the interviews completed, and people started using it, staff just sat back and... waited.

DATA COLLECTION AND ANALYSIS

Data collection commenced on November 12, 2009 and ended on April 18, 2010. In that time, 7,096 trips were uploaded by 1,083 users. A cleaning process was used to filter out problematic or unusable trips and focusing on San Francisco trips only, as we have the richest set of network attributes for San Francisco trips. This left 5,178 trips, which were cleaned and smoothed using the fuzzy logic method of Schussler and Axhausen (2009b). After processing, fully 3,034 bicycle stages from 2,777 original traces uploaded by 366 unique users were successfully matched to the network.

Of the respondents who provided demographic information, the mean age was 34, and 79% were male. The most recent household travel diary information for the Bay Area is from year 2000 and shows 63% of cyclists were male. Thus the data appears more biased toward men. In addition, 60% of respondents reported cycling daily, with 34% cycling several times per week, 7% several times per month. No one reported using the app but cycling less than once per month. Thus, the data is biased toward frequent cyclists. This is important for researchers and planners, as the cycling preferences of frequent cyclists could quite conceivably be different from those of infrequent cyclists. If, as at the SFCTA, planners wish to increase the usage of bicycling, knowing the preferences of infrequent cyclists is paramount.

No other characteristics were collected for which we suspected bias (such as income), because we expected a poor response rate due to privacy concerns. While media and press often asked if there was concerned about only collecting data from smartphone users, the team believed sample bias of smartphone users is a negligible problem for route choice modeling. Route choice is conditioned on already having chosen a destination and mode, and should depend less on demographics and more on characteristics of the alternative. In other words, there is not much reason to suspect that smartphone users prefer bike lines, or dislike hills, any more or less than other groups.

A side note on privacy issues: while the research team went to great lengths to ensure that no user-identifiable information was collected, and made that effort clear to participants, privacy concerns never came up when discussing the app with actual users. In this era of Facebook, Foursquare, and Twitter updates every ten minutes, it seems that location-based data is happily given up for a multitude of reasons by a sizeable portion of our society.

MODEL ESTIMATION

The companion paper by Hood & Sall (n.b.) describes the route choice model estimation procedure and results in extensive detail. Estimation results were produced using a doubly-stochastic choice set generated from network attributes, and with the CycleTracks route included as the "chosen" route.

The following variables were included in the final model specification:

- length of route (coef -1.05, t-stat -11.80)
- turns per mile (coef -0.21, t-stat -12.15)
- proportion of route on wrong-way links (coef -13.30, t-stat -19.87)
- proportion on bike paths (coef 1.89, t-stat 6.17)
- proportion on bike lanes (coef 2.15, t-stat 17.69)
    - Infrequent cyclists (coef 1.85, t-stat 44.94)
- proportion on bike routes (coef 1.85, t-stat 3.14)
- Average up-slope (coef -0.50, t-stat -0.50)

In addition, infrequent cyclists prefer bike lanes even more than frequent cyclists; women dislike up-slope more than men dislike it; and all cyclists dislike up-slope more while commuting than for other trip purposes. When comparing marginal rates of substitution (MRS) for the length of path on street versus various bike facilities, bike lanes have the lowest MRS with 0.49 compared to regular streets, while bike paths have 0.57 and signed bike routes have MRS of 0.92.

This translates to a user benefit of riding on a bike lane of $0.98 per mile of trip, assuming an average value of time of $15 per hour and a speed of 10 miles per hour. Thus the model predicts that people are willing to bike out of their way to reach a bike lane facility; behavior that is borne out in the field.

CONCLUSIONS AND NEXT STEPS

From our perspective this study showed the clear viability of smartphone-based GPS data collection for this specific purpose. Given the issues above regarding participant self-selection, care must be taken when expanding beyond these very targeted data collection tasks. The study team was concerned about the bias of only surveying smartphone users, since there is no reason to suspect that smartphone users dislike hills (for example) more than people without smart-phones. However, for other types of data collection, this bias issue is probably quite real and must be addressed.

The route-choice model developed using this data is already in production at SFCTA and shows promise, but it is clear that more work is needed. The existing destination and mode choice models are not informed by the route-choice "logsums", thus the trip tables coming out of the models don't take slope or bike facilities into account. It doesn't take much imagination to realize that a bike trip table which is blind to street slope is not going to be very accurate. So, a clear next step is to incorporate the logsums of the bike route-choice model into the earlier steps in the model, thus improving the bike trip tables themselves.

As for CycleTracks itself, SFCTA is considering expanding its use to other modes such as pedestrian use, or possibly in before & after studies of new bicycle facilities.

BIBLIOGRAPHY AND WORKS CITED

Sener, I.N., N. Eluru and C.R. Bhat (2008) An analysis of bicycle route choice preferences using a web-based survey to examine bicycle facilities, Working Paper, Department of Civil Engineering, University of Texas, Austin.

Dill, J. and J. Gliebe (2008) Understanding and Measuring Bicycling Behavior: A Focus on Travel Time and Route Choice, Final Report. Oregon Transportation Research and Education Consortium Portland, OR.

Doherty, Sean T. (2009) Multi-sensor Monitoring of Human Activity and Physiology in the Built Environment. Presented at the Annual Meeting of the Transportation Research Board, Washington, DC.

Hood, J., Sall, E., & Charlton, B. (n.d.), A GPS-based Bicycle Route Choice Model for San Francisco, California. Submitted for presentation at the 3rd Conference on Innovations in Travel Modeling 2010, a Transportation Research Board Conference.

McDonald, C.H. and E.K. Burns (2001) Cycling to Work in Phoenix: Route Choice, Travel Behavior, and Commuter Characteristics (Paper 01-2526). Presented at the Annual Meeting of the Transportation Research Board,Washington, DC.

Menghini, G., N. Carrasco, N. Schüssler and K.W. Axhausen (2009) Route choice of cyclists: discrete choice modelling based on GPS-data, Arbeitsberichte Verkehrs- und Raumplanung, 544, IVT, ETH Zurich, Zurich.

Outwater, M. & Charlton, B. (2006), The San Francisco model in practice: Validation, testing, and application, in `Innovations in Travel Demand Modeling: Summary of a Conference', Vol. 2, Transportation Research Board, Washington, DC.