

1 **Bicycle Route Choice Data Collection using GPS-Enabled Smartphones**

2

3 Billy Charlton (*corresponding author*)
4 San Francisco County Transportation Authority
5 100 Van Ness Avenue, 26th Floor
6 San Francisco, CA 94102
7 billy.charlton@sfcta.org
8 (415) 522-4816

9

10 Elizabeth Sall
11 San Francisco County Transportation Authority
12 100 Van Ness Avenue, 26th Floor
13 San Francisco, CA 94102
14 elizabeth.sall@sfcta.org
15 (415) 522-4810

16

17 Michael Schwartz
18 San Francisco County Transportation Authority
19 100 Van Ness Avenue, 26th Floor
20 San Francisco, CA 94102
21 michael.schwartz@sfcta.org
22 (415) 522-4800

23

24 Jeff Hood
25 San Francisco County Transportation Authority
26 100 Van Ness Avenue, 26th Floor
27 San Francisco, CA 94102
28 (415) 522-4800
29 jeffnhood@gmail.com

30

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

32

33 **Word Count:**

34 Text: 3,800

35 Tables: 3 x 250 = 750

36 Figures: 1 x 250 = 250

37 **Total Word Count: 4,800**

38

39 Submitted for Presentation at the 2011 Transportation Research Board Annual Meeting.

40 Original Submission Date: August 1, 2010; Revised November 15, 2010.

41

1 ABSTRACT

2 The proliferation of consumer-grade smartphones with global positioning system (GPS) location
3 capabilities opens a new data collection method for researchers. When confronted with a lack of
4 data on bicycle routes preferred by local cyclists, the San Francisco County Transportation
5 Authority (SFCTA) developed a freely downloadable iPhone/Android smartphone “app” called
6 CycleTracks to collect actual bicycle routes traversed by city cyclists. Cooperation with local
7 bicycle advocacy groups, along with social media and email campaigns, encouraged use of the
8 app by regular citizenry. Several rounds of pre-release testing showed that making the app start
9 up quickly, and minimizing battery usage during recording, were critical to getting good data.
10 Once installed on a user's smartphone, a single “tap” would start and stop recording a bicycle trip;
11 after completing a trip, the app automatically uploaded the track to a central database/web server,
12 via the phone's built-in data plan. Approximately 5,000 usable bicycle trips were collected from
13 hundreds of users in the region. Demographic data was optionally provided by some users, and
14 showed a bias toward frequent cyclists, and toward male users (even more than cycling is already
15 male-dominated in the region's most recent household travel survey). A bicycle route choice
16 model developed using the data revealed sensitivity to slope, presence of bike lanes and/or bike
17 route designations, trip purpose, and gender. The bike route choice model is now being integrated
18 into San Francisco's regional travel model.
19

1 INTRODUCTION

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

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

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

27 REVIEW OF ROUTE CHOICE DATA COLLECTION TECHNIQUES

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

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

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

43 Special-purpose GPS devices are small and lightweight enough to carry on-person for
44 long time periods. Price has also come down significantly in the past ten years and individual
45 units can be obtained for less than \$100. However, if the user is not continually prompted for
46 mode information the GPS points for bicycle modes can be confused with various other modes.
47 Meghini et al (2009) gleaned 2,657 bike routes from a personal GPS dataset with 11,000 total
48 trips and 2,435 people. Dill and Gliebe used bicycle-mounted GPS devices to record bicycle

1 routes of 164 adults in Portland, Oregon. The bike-mounted GPS eliminated much of the data-
2 cleaning required by the personal GPS devices, but researchers still had to trim portions of routes
3 where the cyclist was on transit. The participants used the bike-mounted GPS device for seven
4 days, at the end of which the project team retrieved the unit and downloaded the GPS data.
5 Doherty (2009) provided smartphones to participants to track second-by-second GPS locations,
6 which were regularly uploaded wirelessly to a remote server and stored in a MySQL database.
7 The smartphones operated continuously for 17 hours at a time with the help of extra battery
8 packs. Wireless data uploads negated the need for researchers to spend much time in the field
9 retrieving data.

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

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

27 **CONCERNS REGARDING DATA BIAS**

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

33 *Smartphone users might choose different routes than non-smartphone users*

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

46
47 The team decided to request "personal info" including age and gender to help document and
48 analyze the user base.

1

2 *Users who download this app might be different from smartphone users who do not*

3 Clearly, cyclists interested in improving civic life for other cyclists are most likely to use the app,
4 and they could easily have different route preferences than infrequent cyclists. The team tried to
5 reach out to both groups, as is described below. The app also requested typical cycling frequency
6 to help researchers understand the user base.

7

8 *Users might use the app for certain trips or certain routes, to “game” results*

9 With any uncontrolled data experiment, this is a risk. The team decided to weight the dataset
10 inversely proportional to the number of submitted trips by any one user, to help combat any
11 concerted effort to skew results.

12

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

17 **THE “CYCLETRACKS” IPHONE/ANDROID APPLICATION**

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

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

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

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

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

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

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

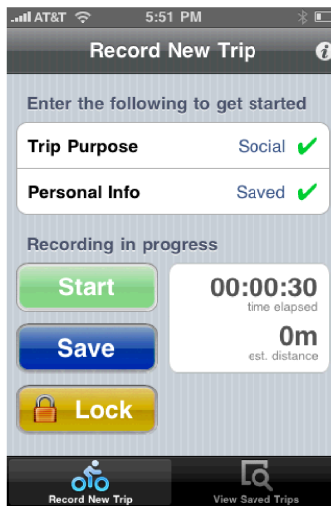
7 A beta testing period revealed battery usage issues correlated with heavy GPS radio
 8 usage. Recognizing that participants are voluntary, several modifications avoid draining the
 9 battery. To remind participants that it is collecting data, CycleTracks makes a “bicycle bell” noise
 10 and vibrates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

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

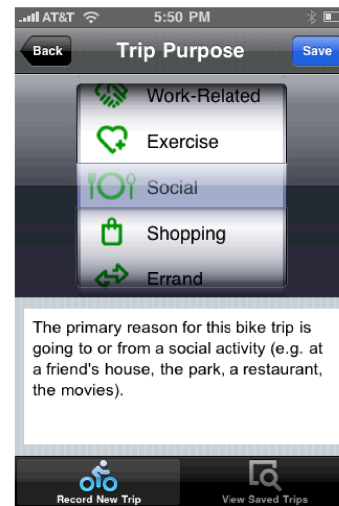
14
15
16
17



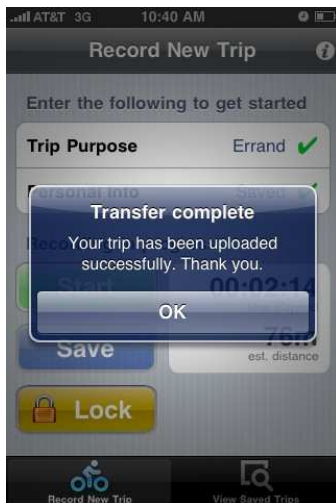
a. Splash Screen



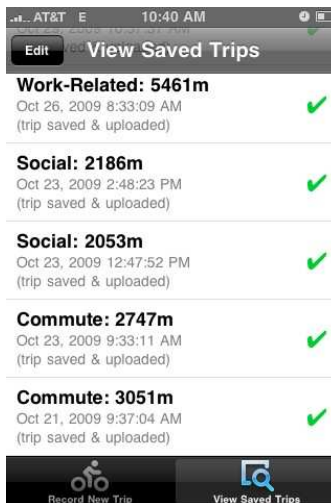
b. Record a new trip



c. Enter your trip purpose



d. Success! Trip uploaded.



e. List of saved trips for review



f. Map of your bike trip

1 SERVER SIDE DATABASE

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

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

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

12

13 **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*

14 * Fields with an asterisk are optionally provided by user

15

16 **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)

17

18 **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

19

1 TESTING RESULTS

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

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

12

13 CYCLETRACKS USAGE PROMOTION AND RECRUITMENT

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

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

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

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

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

30

31 DATA COLLECTION AND ANALYSIS

32 Data collection commenced on November 12, 2009 and ended on April 18, 2010. In that time,
33 7,096 trips were uploaded by 1,083 users. A cleaning process was used to filter out problematic
34 or unusable trips and focusing on San Francisco trips only, as we have the richest set of network
35 attributes for San Francisco trips. This left 5,178 trips, which were cleaned and smoothed using
36 the fuzzy logic method of Schussler and Axhausen (2009b). After processing, fully 3,034 bicycle
37 stages from 2,777 original traces uploaded by 366 unique users were successfully matched to the
38 network.

39 Of the respondents who provided demographic information, the mean age was 34, and
40 79% were male. The most recent household travel diary information for the Bay Area is from
41 year 2000 and shows 63% of cyclists were male. Thus the data appears more biased toward men.
42 In addition, 60% of respondents reported cycling daily, with 34% cycling several times per week,
43 7% several times per month. No one reported using the app but cycling less than once per month.
44 Thus, the data is biased toward frequent cyclists. This is important for researchers and planners,
45 as the cycling preferences of frequent cyclists could quite conceivably be different from those of

1 infrequent cyclists. If, as at the SFCTA, planners wish to increase the usage of bicycling,
2 knowing the preferences of infrequent cyclists is paramount.

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

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

16 17 **MODEL ESTIMATION**

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

22
23 The following variables were included in the final model specification:

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

33
34 In addition, infrequent cyclists prefer bike lanes even more than frequent cyclists; women dislike
35 up-slope more than men dislike it; and all cyclists dislike up-slope more while commuting than
36 for other trip purposes. When comparing marginal rates of substitution (MRS) for the length of
37 path on street versus various bike facilities, bike lanes have the lowest MRS with 0.49 compared
38 to regular streets, while bike paths have 0.57 and signed bike routes have MRS of 0.92.

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

43 44 **CONCLUSIONS AND NEXT STEPS**

45 From our perspective this study showed the clear viability of smartphone-based GPS data
46 collection for this specific purpose. Given the issues above regarding participant self-selection,
47 care must be taken when expanding beyond these very targeted data collection tasks. The study

1 team was concerned about the bias of only surveying smartphone users, since there is no reason
2 to suspect that smartphone users dislike hills (for example) more than people without smart-
3 phones. However, for other types of data collection, this bias issue is probably quite real and
4 must be addressed.

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

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

14
15 CycleTracks is Copyright 2009,2010 San Francisco County Transportation Authority, and is
16 GPL open source software, available at <http://www.sfcta.org/cycletracks>.

17 **BIBLIOGRAPHY AND WORKS CITED**

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

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

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

28
29 Hood, J., Sall, E., & Charlton, B. (n.d.), A GPS-based Bicycle Route Choice Model for San
30 Francisco, California. Submitted for presentation at the 3rd Conference on Innovations in Travel
31 Modeling 2010, a Transportation Research Board Conference.

32
33 McDonald, C.H. and E.K. Burns (2001) Cycling to Work in Phoenix: Route Choice, Travel
34 Behavior, and Commuter Characteristics (Paper 01-2526). Presented at the Annual Meeting of
35 the Transportation Research Board, Washington, DC.

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

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

42