San Francisco Travel Demand Forecasting Model Development

Executive Summary

Final Report

prepared for
San Francisco County Transportation Authority

prepared by
Cambridge Systematics, Inc.

Updated by:
San Francisco County Transportation Authority

October 1, 2002

Table of Contents

1.0 Introduction

Overview

The San Francisco County Travel Demand Forecasting Model (San Francisco Model) was developed for the San Francisco County Transportation Authority (SFCTA) to provide detailed forecasts of travel demand for various planning applications. These applications included developing countywide plans, providing input to microsimulation modeling for corridor and project-level evaluations, transit planning, and neighborhood planning. The objective was to accurately represent the complexity of the destination, temporal and modal options and provide detailed information on travelers making discrete choices. These objectives led to the development of an activity-based model that uses a synthesized population as the basis for decision-making rather than zonal-level aggregate data sources. The activity-based model has nine primary components.

Most of the model components were estimated using household survey data collected by the Metropolitan Transportation Commission (MTC) for San Francisco residents only. Each model component was calibrated using various observed data sources, then the full model was validated using traffic count and transit ridership data for each of five time periods. The model is applied as a focused model, which combines trip-making from the entire Bay Area (derived from the MTC's BAYCAST trip tables) with the travel demand from San Francisco residents produced by the activity-based model.

Contents of this Report and Related Reports

This executive summary discusses all nine model components and provides an overview of the data required to run the model. It is designed to provide an overview of the process and a brief summary of the results. There were numerous technical reports developed during the process; these should be referred to for more detail. The primary reports are listed below:

- Data Development
- Population Synthesis
- Vehicle Availability Model
- Tour and Trip Generation and Time-of-day Models
- Destination Choice Models

- Tour and Trip Mode Choice Models
- Visitor Model
- Model Validation
- MTC Consistency

Background

The model system was designed to use the "full day pattern" activity modeling approach, first introduced by Bowman and Ben-Akiva at MIT.[1] The first practical application of that approach was done for Portland METRO,[2] with part of the funding provided by the Federal Travel Model Improvement Program (TMIP). The Portland model system has since undergone a number of further developments.[3] The San Francisco models are very similar in design to the most recent Portland models. Key features of the model system are:

- The use of tours as a key unit of travel.
- Joint modeling of various tours made within a person's day.
- Breaking down each tour into a chain of linked trips.
- Microsimulation of the travel for each individual in the population.

To allow the San Francisco models to be developed and implemented within a relatively short time frame, some simplifications were made relative to the Portland models.

Purpose

The main motivation in developing activity-based models is the need to better understand traveler's responses to transportation policies. There are a number of ways that activity-based models provide these capabilities:

- Activity-based models can account for tradeoffs for auto ownership based on the employment location of the primary worker in the household. This is a significant factor

[1] John Bowman and Moshe Ben-Akiva. "The day activity schedule approach to travel demand analysis". Paper presented at the 78th Annual Meeting of the Transportation Research Board, Washington, 1999.

[2] Mark Bradley, John Bowman, T. Keith Lawton, Moshe Ben-Akiva, Tom Rossi and Yoram Shiftan, "A System of Activity-Based Models for Portland, Oregon". Report prepared for the Federal Highway Administration Travel Model Improvement Program. Washington, D.C. 1998.

[3] Mark Bradley, John Bowman, and T. Keith Lawton, "A Comparison of Sample Enumeration and Stochastic Microsimulation for Application of Tour-Based and Activity-Based Travel Demand Models." Presented at the European Transport Conference. Cambridge, England, 1999.

for auto ownership in a transit-rich environment such as San Francisco. According to the 1990 Census, San Francisco has the second highest percentage of transit usage in the U.S. and the third highest percentage of other modes for travel to and from work[4].

- Activity-based models can account for tradeoffs between making additional stops on the primary tour or making additional tours by defining the primary tour, tour type, and number of stops simultaneously. For example, a policy that encourages alternative modes for travel to work can be evaluated in relationship to the tradeoffs for making intermediate stops (to run errands) to and from work using alternative modes. This is achieved by including modal accessibility measures in the full-day pattern tour models.

- Activity-based models can account for tradeoffs between trip chaining and time of day by evaluating time of day decisions at the tour level rather than the trip level. These time of day decisions are made simultaneously for the outbound and return portions of tours and are based on the tour type and number of stops. Pricing policies (such as parking or toll policies) can be tested more accurately by including these tradeoffs between the need to travel for purposes that are time-dependent (such as day care or work) and the desire to avoid peak period pricing.

- Activity-based models can also account more reliably for the complexities involved in multi-mode trip-making. Travel modes are affected by decisions to travel in a round trip rather than an individual trip segment and certain modal options have multiple options for modes within a round trip. The San Francisco Model was developed to address tradeoffs for modes for the full tour, as well as the tradeoffs for modal options of trips within a tour.

The San Francisco Model was developed as a microsimulation model rather than a sample enumeration model because of the significant advantage that much more detail could be retained in the output. Microsimulation models forecast travel by modeling a set of actual or synthetic individuals or households that represent the entire population. A "synthetic" sample is created which is composed of a hypothetical set of people or households with characteristics that as a whole match the overall population, and the model considers these decision-makers choices individually. In contrast, a "sample enumeration" model enumeration relies on the modeling of behavior for a representative sample of the population, rather than the full population. These results are then expanded to represent the entire population.

The simulated trips generated by the microsimulation model can be aggregated to perform equity analysis, create trip matrices or provide input to traffic simulations. This process also provides flexibility in developing work-based tours and intermediate stop locations with much less processing time. The primary disadvantage to this approach is that there is less stability in the results due to the random draws in the simulation process, but using larger synthetic samples can counteract this.

---

[4] Yoram Shiftan and Thomas F. Rossi, "Moving Toward Activity-Based Models – Experience in Four U.S. Areas," submitted to the Transportation Research Board, July 2000.

2.0 Model System Overview

The main feature of the "full day pattern" approach is that it simultaneously predicts the main components of all of a person's travel across the day. This includes the frequency of five types of tours:

- Home-based work primary tours
- Home-based education primary tours
- Home-based other primary tours
- Home-based secondary tours
- Work-based sub-tours

A home-based tour includes the entire chain of trips made between leaving home and arriving back at home. The "primary" home-based tour is defined as the main home-based tour made during the day. If a worker makes a work tour or a student makes an education tour, then that is always the primary tour. If there are no work or education tours, the primary tour is the tour with the highest priority activity at the destination (shopping/personal business, followed by social/recreation, followed by serve passenger). If there are two or more tours with the same activity priority, then the one with the longest duration of stay at the destination is the primary tour. All other home-based tours are designated as "secondary" tours. A special type of tour is a work-based "sub-tour", defined as the entire chain of trips made between leaving the primary workplace and returning back to that workplace in the same day. By using tours as a key unit of travel, we capture the interdependence of different activities in a trip chain. This provides a better understanding of non-home-based trips, especially in the case of the work-based sub-tours that represent a significant proportion of non-home-based travel.

Figure 2.1 is an illustration of a possible "full day pattern" of a San Francisco resident worker. The Figure shows that the worker makes three tours during the day: a primary home-based work tour, a work-based subtour, and a secondary home-based tour. These tours, which always begin and end at the same destination, are composed of a series of trips. The tours are comprised of a total of seven trips, and the trips are labeled with numbers to indicate the order in which the trips occur. The worker begins their day at home. Their first trip takes them from home to an intermediate stop en route to work (perhaps to address a child care or personal business requirement, although the specific purpose of this intermediate stop is not modeled). The second trip of the day takes the traveler from this intermediate stop to their workplace destination. While at work they begin second tour, a work-based subtour. Note that this sub-tour begins prior to completion of the first home-based work tour. The third and fourth trips of the day take the traveler to and from work to this work-based subtour destination. This destination may be to address either a work or personal business matter, but again the specific purpose is not modeled. The fifth trip of the day returns the traveler from their workplace location to their home. Once home, they begin the third tour of the day, a secondary home-based tour. Trips six and seven take the traveler from home to their destination and back home again."

The study area for the model is the 9-county San Francisco Bay Area, which is represented by the Metropolitan Transportation Commission (MTC) regional travel demand forecasting model (BAYCAST). The study area is divided in two parts, so the San Francisco activity-based model can be used to predict travel by San Francisco County residents, while the BAYCAST model can be used to predict travel by residents from the other eight counties.

Figure 2.2 presents a schematic diagram of the San Francisco Model system. This diagram includes the model components and data inputs for these components. A synthesized population of San Francisco residents is input to each model component to estimate choices for work location, vehicle availability, and tours and trips by time-of-day, destination and mode of travel. The synthesized tours and trips are aggregated to represent flows between traffic analysis zones before traffic assignment. A separate model of visitor travel is estimated to incorporate trips made by tourists and business travelers visiting San Francisco County. The model system also incorporates other trips made by non-San Francisco residents by merging regional trip tables into the process for assignment.

Figure 2.1 Illustration of "Full Day Tour Pattern"

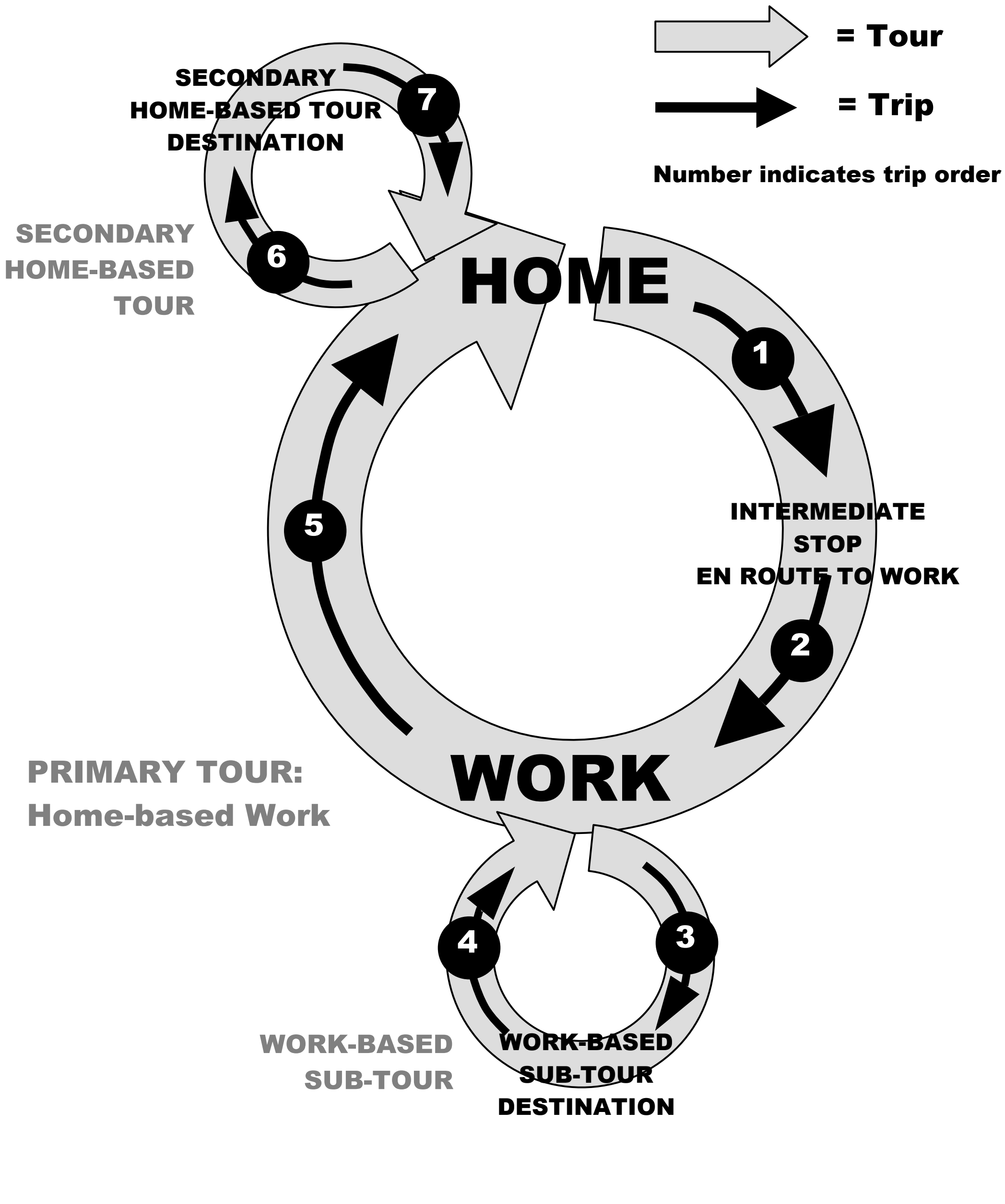
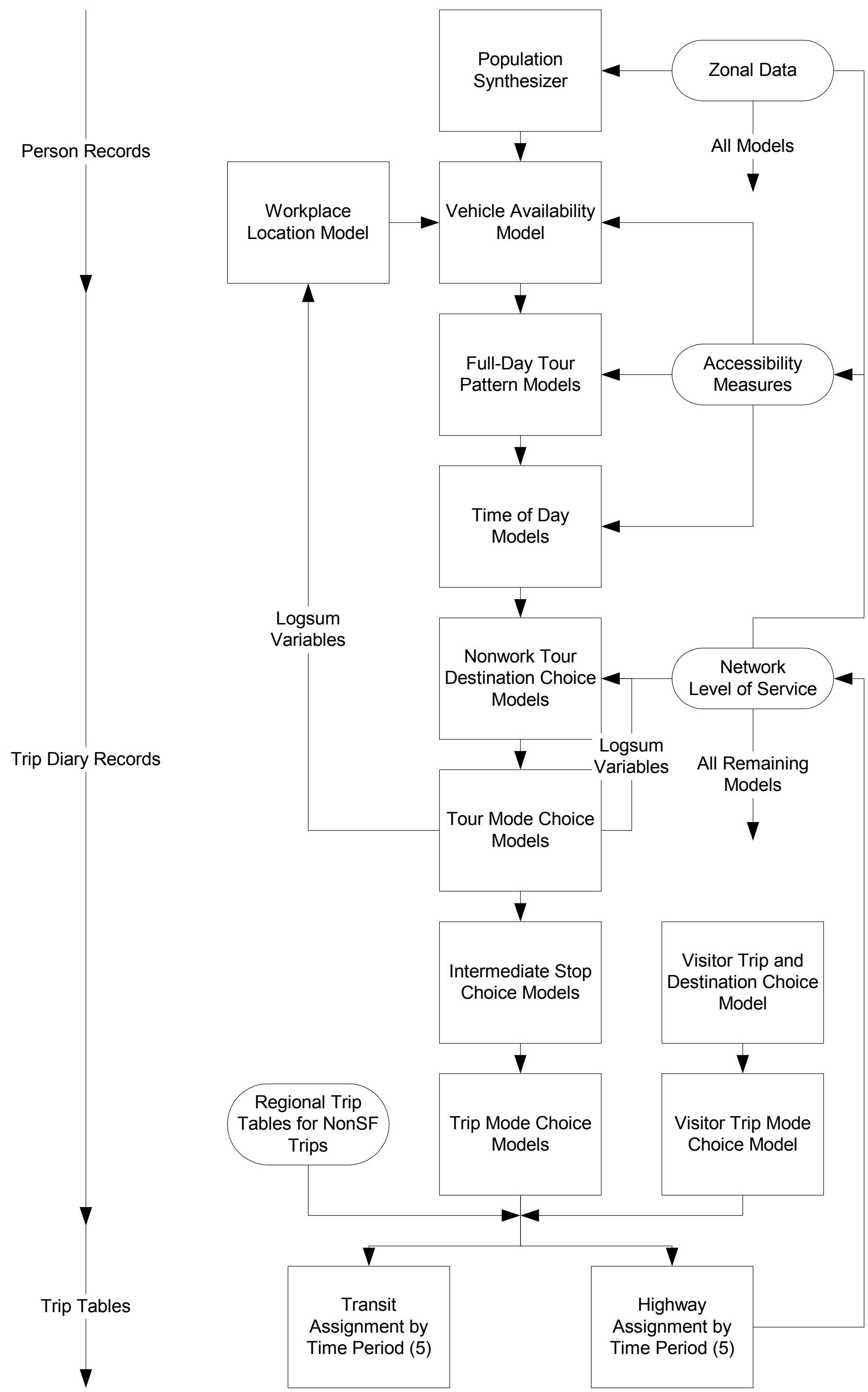


Figure 2.2 Model System



3.0 Data Development

There were three primary areas of data development: data collected as part of the stated preference survey, the development of the synthetic population data, and data used as input to the San Francisco model. There are individual reports for each of these areas. An overview of these data is provided below.

Stated Preference Survey

The stated preference survey was conducted for 609 households in San Francisco in June, 1999 to collect data on transit and auto travel characteristics. The primary focus of the survey was to collect preference data on transit reliability, crowding and personal security and auto parking availability and cost. The survey was conducted by Corey, Canapary and Galanis and the design of the survey was completed by Mark Bradley Research and Consulting, with other members of the Cambridge Systematics team.

The purpose of the survey was to provide data that can be incorporated into the mode choice model estimation process, in the areas of transit reliability, crowding and personal security and auto availability and cost. The analysis of these data was conducted as part of the mode choice model process.

Synthetic Sample Generation

A prototypical sample of persons and households was generated for San Francisco County using three primary data sources: the U.S. Census Public Use Microdata Sample (PUMS), the population and employment data developed for San Francisco County, and other socioeconomic data developed for the MTC. There is a hierarchy of zonal systems for these three datasets:

- Six Public Use Microdata Areas (PUMAs), containing
- 127 MTC Traffic Analysis Zones (MTAZs), containing
- 766 San Francisco Traffic Analysis Zones (SFTAZs).

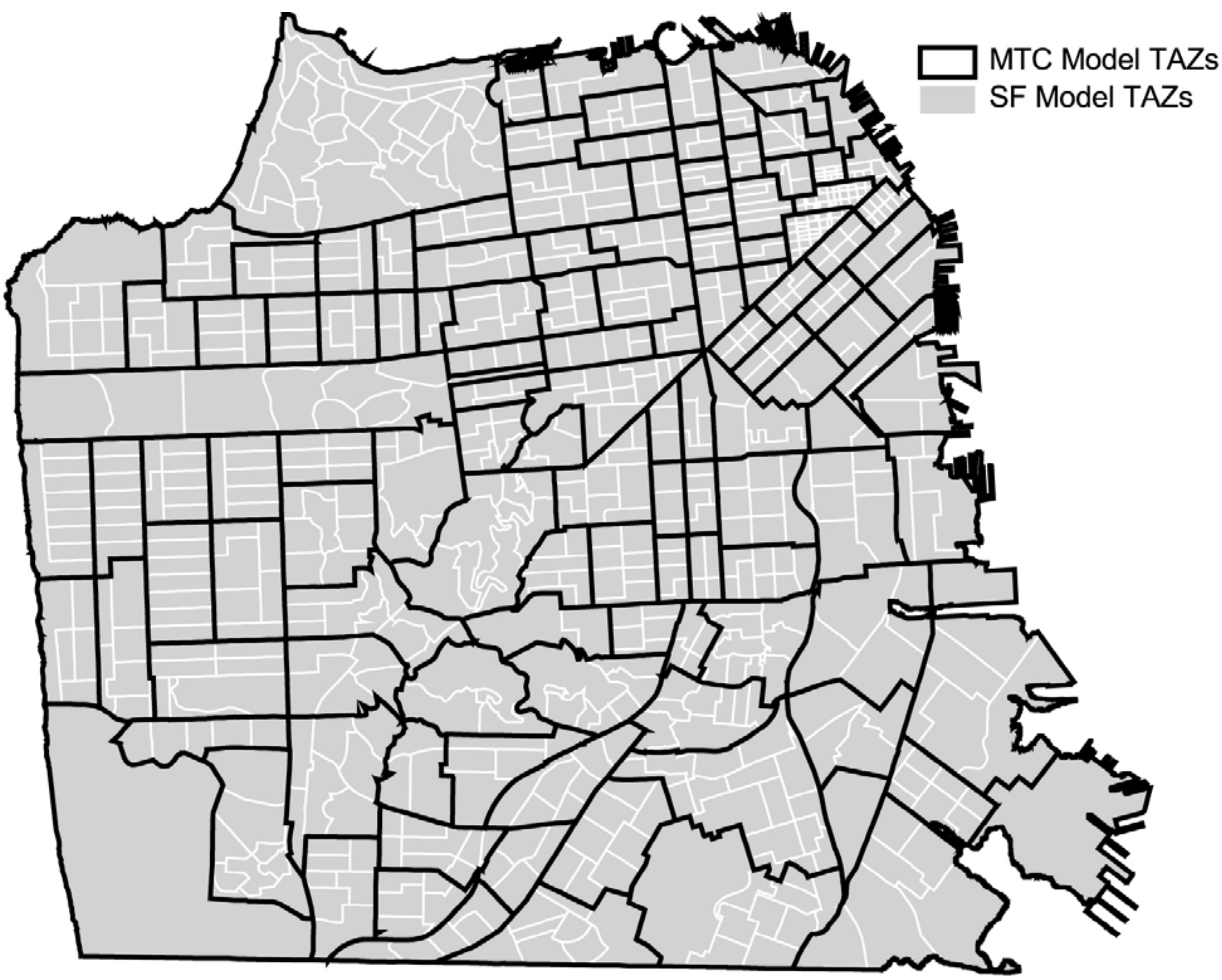
Figure 3.1 shows the boundaries of the SFTAZs and MTAZs. The PUMAs are not shown because they are relatively large areas used to preserve the anonymity of long form respondents.

The prototypical sample contains marginal distributions across three dimensions:

- Household size and number of workers (nine categories);
- Household income (four categories); and
- Age of head of household (three categories).

There are a total of 108 possible combinations of the above dimensions (9x4x3). The nine categories for household size/number of workers were chosen because they efficiently distinguish between important household life-cycle groups. The specific breakdowns for income and age were chosen because they correspond to categories that are available in the MTC future year land use files, so updating the populations to future years can be kept consistent with MTC breakdowns within zones. Also, all of these categorizations are compatible with the Census tables available in the Census Transportation Planning Package (CTPP) Urban Element.

Figure 3.1 Map of San Francisco Model & MTC regional model TAZ boundaries



Other Model Data

Aggregate Zonal Data

Some of the data used by the model components are aggregate zonal data developed as either necessary inputs or because these are desired for testing planning policies. Table 3.1 provides a list of these aggregate variables and the model components that use these variables. The socioeconomic data were developed from parcel-level data aggregated to traffic analysis zones and adjusted to match control totals, as follows:

- The San Francisco Planning Department provided a current parcel database and a current business and employment database. The parcel database provides current estimates of residential units at the block and lot level and the business and employment database contains current estimates of employment by type at the block and lot level. These are aggregated to the traffic analysis zones.
- The San Francisco Planning Department, the Presidio Trust, the San Francisco Redevelopment Agency and the Port of San Francisco maintain lists of new development projects under construction, approved, and under review, as well as information on development potential for major area plans. These are used to allocate forecast data by traffic analysis zone.
- The Association of Bay Area Governments' *Projections '98* was used as a control total for countywide forecasts of population and employment. The San Francisco Planning Department has subsequently updated these forecasts to reflect the *Projections 2000* data.

The employment data in San Francisco uses a different categorization compared to the MTC data. The original MTC databases classified employment by six categories – retail, service, other, agricultural, manufacturing and trade. The new San Francisco socioeconomic databases classified employment by a different set of six categories:

- Cultural, institutional and educational services (CIE),
- Medical and health services (MED),
- Management, information, and professional services (MIPS),
- Production, distribution and repair (PDR),
- Retail and entertainment (RETAIL), and
- Visitor (VISITOR).

These employment categories were defined by the San Francisco Planning Department in the 1998 Citywide Land Use Study. Most models retained the distinctive employment categories, but some used a common set of categories across all areas, where basic information on the SIC codes falling under each category was used to regroup the MTC categories into four San Francisco categories – PDR, MIPS, Retail and Service.

Pedestrian environment factors (PEF) were developed to evaluate urban design projects and estimate changes in pedestrian and bicycle modal options. PEFs will allow local planners to:

- Quantify base year variables related to the pedestrian environment by geographic area (traffic analysis zone, area type) that can be used for transportation, transit, and land use planning and modeling;
- Develop a policy variable to measure the potential impacts of improved pedestrian systems and expected growth (in vehicles, population, employment) that will likely impact future travel demand; and
- Incorporate pedestrian factors into the travel demand modeling process to assess integrated land use and transportation policies/alternatives.

Eight members of the San Francisco Pedestrian Advisory Group, made up of staff from local agencies and private enterprises, collected relevant data and allocated the results to the PEF traffic analysis zone (TAZ) system established for the effort.

Parking cost and availability have a significant impact on vehicle availability and mode choice for travel to destinations in San Francisco. To support this analysis, the supply, cost, and availability of parking were developed from a variety of sources, including parking surveys, a small sample stated preference survey, parcel data and aerial photographs of on-street parking.

Area type was used as an aggregate zonal variable in a number of the model components, as well as a network variable. Area type was derived from the MTC regional TAZs and is classified into six categories – Core Central Business District (CBD), CBD, Urban Business District (UBD), Urban, Suburban, and Rural. Within San Francisco County, all the TAZs fall within the first four categories, implying that all suburban and rural classifications within the region are outside the City.

Table 3.1 Aggregate Zonal Data Used in the San Francisco Model

Type	Description	Model Components That Variables are Used in
Socioeconomic Data	Total number of Households	Destination Choice Models
	CIE Employment	Destination Choice Models
	MED employment	Destination Choice Models
	MIPS employment	Destination Choice Models
	PDR employment	Destination Choice Models
	RETAIL employment	Destination Choice Models
	VISITOR employment	Destination Choice Models
	Total Employment	Destination Choice Models
Pedestrian Environment Factors	Network Continuity/Integrity	Mode Choice Models
	Ease of Street Crossing	Mode Choice Models
	Perception of Safety and Personal Security	Mode Choice Models
	Urban Vitality	Vehicle Availability Model
	Topological Barriers	Mode Choice Models
Parking Data	Parking Availability Index	Vehicle Availability Model
	Average Parking Cost for work trips (based on 8 hours)	Mode Choice Models
	Average Parking Cost for other trips (based on 1 hour)	Mode Choice Models
Hotel & School Data	Number of Hotel Rooms	Visitor Model
	High School Enrollment	Destination Choice Models
	College Full Time Enrollment	Destination Choice Models
	College Part Time Enrollment	Destination Choice Models
	Number of School Buildings	Destination Choice Models
	School Area (square footage)	Destination Choice Models

Level of Service and Accessibility Variables

Accessibility variables are developed for different combinations of employment within a travel time bandwidth, using the following categories:

- Within 15, 30 and 45 minutes time intervals and half-mile distance interval;

- Number of jobs in service (CIE and MH), retail, MIPS and PDR employment categories; and
- Auto and transit travel times during AM, midday, PM, and evening time periods.

These accessibility variables are included in the vehicle availability, full-day pattern tour, and time-of-day model components and are based on assignments of regional trip tables for each of the time periods. There are five time periods that are modeled, but the early AM and evening time periods are assumed to operate in free flow conditions for this purpose. The destination and mode choice models use the mode choice logsum variable for level of service.

The highway and transit networks that were developed to support this model are regional networks that have full level of network detail within the County of San Francisco. All streets and all transit routes within San Francisco are included. There are 51,000 links and 1,739 zones in the full system.

4.0 Travel Model Development

■ Overview

Table 4.1 lists the sequence of models and gives an idea of the key endogenous variables in each one. Endogenous variables are variables that are defined and produced within the model system, as opposed to exogenous variables which are provided as external inputs into the model system. Examples of an endogenous variable are travel times to and from every zone to every other zone. We do not have observed data for each zone-to-zone movement (there are over 3 million). Examples of a related exogenous variable are the transportation networks used in the model. These detailed descriptions of transit and roadway alignments and capacity are based on observed data about roadway lengths, transit stop locations and many other relevant transportation system characteristics. It is these "exogenous" networks that are used to develop the "endogenous" travel times between all zones.

The first model in the sequence is the workplace location model, which predicts the work locations for any workers in the household, along with the residence location and variables such as household size and income. This model feeds into a model to predict the number of vehicles available to the household. The vehicle availability model is described in the next section. Both workplaces and vehicle availability feed into the "full day tour pattern" model, described later in this section. The pattern model predicts the main components of the entire day's travel for the individual, and provides key inputs to the subsequent models: primary tour time-of-day, tour and trip destination choice, and tour and trip mode choice.

A feature of this model system that differs from previous ones is that workplace location is modeled at the beginning of the decision process. This structure reflects the fact that work location is a longer-term decision relative to choosing time of day or mode or non-work travel destinations. By putting workplace location at the top of the structure, the location of workplaces relative to home can be used as input in other models such as vehicle ownership and frequency of travel.

Table 4.1 Sequence of Models for San Francisco County Residents (I = input, O = output, (W) = work purpose only)

	Primary home-based tours					Secondary & work-based tours					Trips (tour segments)			
Model	purp	chain	time	dest	mode	num	chain	time	dest	mode	num	time	dest	mode
Workplace location choice				O (W)										
Household vehicle availability choice				I (W)										
Full day tour pattern choice	O	O		I (W)		O					O			
Primary tour time period choice	I	I	O			I								
Secondary and work-based tour chain type	I	I				I	O				O			
Secondary and work-based tour time periods	I		I			I		O						
Trip time period	I		I					I			I	O		
Non-work tour primary destination choice	I	I	I	I (W) O		I	I	I	O					
Tour main mode choice	I	I	I	I	O	I	I	I	I	O	I			
Intermediate stop location choice	I	I	I	I	I	I	I	I	I	I	I	I	O	
Trip mode choice	I	I	I	I	I	I	I	I	I	I	I	I	I	O

Tour purposes (purp)	Tour chain types (chain)	# tours or stops (num)	Trip time periods (time)	Destinations (dest)	Tour modes(mode)
Work Education Other	No intermediate stops Stop(s) before primary Stop(s) after primary Stop(s) both ways	None 1 2 3 4 or more	Early (3-6 am) AM peak (6-9 am) Midday (9 am-3:30 pm) PM peak (3:30-6:30 pm) Evening (6:30 pm-3 am)	1739 traffic analysis zones (For each tour/trip, a stratified sample of 40 zones is used for estimation)	Car drive alone Car passenger Walk Bike Walk to transit Drive to transit

■ Vehicle Availability Model

The vehicle availability model is a multinomial logit model that predicts the vehicles available in each household for each San Francisco resident. Given the location of the household, the characteristics of the household members, and the primary work place location of each of its workers, the model estimates the probabilities of having none, one, two, or three or more vehicles available.

A large number of households (42.9%) in San Francisco in 1990 had only one vehicle and the average number of vehicles for all households was 1.16. The number of vehicles is defined as automobiles plus trucks; also available in the survey data are the numbers of motorcycles, mopeds and bicycles owned by the household, but these were not included in the number of vehicles available for household travel. The model was limited to four alternatives (0, 1, 2, or 3+ vehicles available) because of the relatively small number of households with four or more vehicles available (1.8%). The average number of vehicles in the fourth alternative (households with three or more vehicles available) was 3.36.

Information was assembled from a number of sources to create the estimation data set. For example, the household survey came from MTC, population and employment datasets were developed by the consultant team working with Planning Dept data, Pedestrian Environment Factors were developed by SFCTA staff with assistance from staff of other city departments and consultant team, and parking costs based on small survey undertaken by consultant team. The structure of this data set is a file with one record for each San Francisco household in the travel survey, with data on income, location, and the age and employment status of the various household members. (Driver's license status was not used in estimation, because it is not available in the PUMS Census data used to apply the models.) The household file was supplemented by adding zonal data, level of service data, and accessibility data. The zonal data included population, households, and employment by type, area in square miles, area type, pedestrian environment factor, and parking costs. The level of service data included both auto and transit travel times and costs between the residence zone and each household member's workplace. The accessibility data included measures of how many jobs of various types could be reached by transit or car in various travel time bands.

■ The Full Day Pattern Models

As Table 4.1 indicates, the full day pattern model predicts:

- The purpose class of the primary home-based tour (work, education, other, or none)
- The trip chain type of the primary home-based tour (1 or more stops before, after, neither, or both)
- The number of home-based secondary tours (0, 1, or 2+)

Section 2.0 and Figure 2.1 of this report describe and illustrate the concept of the full-day tour pattern model

For purposes of estimation, four different person types were defined:

- **Children**: Anyone under age 16
- **Working adults**: Anyone age 16+ who has employment status "employed full time" or "employed part time" OR whose primary tour was for work, with a stay of at least 2 hours at the primary destination
- **Student adults**: Anyone age 16+ who has employment status "full time student" or "part time student" OR whose primary tour was for education, with a stay of at least 2 hours at the primary destination
- **Other adults**: All other people age 16+

Using only San Francisco residents from the 1990 BATS survey with valid full-day travel diaries, the total sample includes 3519 person-days, of which about 63% are working adults, 11% are student adults, 21% are other adults, and only 8% are children age 5 to 15. This percentage of children seems low, but may be a particular feature of San Francisco county residents.

The tours and purpose types are defined in such a way that primary work tours can only be made by working adults and primary education tours can only be made by students or children. Other primary tours and secondary tours can be made by all person types. During the 3519 person-days, there are 4176 tours in total, an average of about 1.2 tours/day. Working adults make more than 1 tour per person-day on average, students and children make slightly less than 1 tour per person-day, and other adults travel the least. Overall, slightly over half of all tours are made either to work or are base from work ("sub-tours"). Also, 63% of all tours are made by a person who makes a work tour during the day. This figure emphasizes the importance of the work tour in determining travel patterns – not only for the trip chain to and from work, but also in affecting secondary tours made during the same day.

Tour Time-of-day Models

The next type of model is the time-of-day model for primary home-based tours. This model is conditional on the type of pattern predicted by the full day pattern model. It predicts the period when the traveler leaves home to begin the primary tour, simultaneously with the period when the traveler leaves the primary destination to return home. The periods used for the models are defined as:

- Early (3:00 AM to 5:59 AM)
- AM peak (6:00 AM to 8:59 AM)
- Midday (9:00 AM to 3:29 PM)

- PM peak (3:30 PM to 6:29 PM)
- Late (6:30 PM to 2:59 AM)

Excluding overnight tours, of which there are almost none in the data, there are 15 possible combinations of these 5 periods.

The estimation data set contains the chosen departure time period combinations for each of the three types of primary tours. There are virtually no Education or Other tours that begin in the Early period before 6 AM. There are some Work tours that begin that early, with almost all of those ending either in the Midday or PM peak period. Also, tours beginning after 3:30 PM in the PM peak or Late periods are very rare for Work and Education, but more common for Other primary tours.

The majority of Work tours are AM peak-PM peak, with most of the remaining tours in AM peak-Midday, AM peak-Late, Midday-PM peak and Midday-Late. Almost all Education tours are either AM peak-Midday or AM peak-PM peak. Other primary tours are the most heterogeneous, with Midday-Midday and Midday-PM peak being the most common combinations.

The input variables available for use in the tour time-of-day models are the same as for the day pattern models. In addition, the time-of-day models are conditional on the predictions from the day pattern models, so a few further input variables are possible:

- Whether or not intermediate stops are made during the tour
- Whether or not work-based sub-tours are made
- Whether or not other secondary home-based tours are made

Including these endogenous variables ensures that certain types of tours and tour patterns are more likely to occur at certain times of day. For instance, tour legs with intermediate stops are more likely to be made during periods when most stores are open, and primary tours that are followed by secondary tours may tend to be made earlier in the day.

In general, there is strong relationship between the day pattern type (primary tour purpose, stops before and after the main destination, secondary tours) and the time periods during which the tour is made. This result reinforces the importance of modeling pattern choice in order to predict realistic shifts in time of day distributions.

■ Destination Choice Models

Destination choice models perform the same general function that trip distribution models, such as the gravity model, do in the traditional four-step modeling process. Unlike traditional models, trip productions are determined first using day-pattern models and are the starting point for the destination choice models. Trip attractions are not determined explicitly and instead are represented using size or opportunity indicators such as employment by category. Thus, destination choice models determine not only the trip interchanges but also the total attractions for each zone.

Two types of models are estimated – tour-level models that determine the primary destination and trip-level models that capture the choice of intermediate stops on a tour, with the latter type of model dependent on the former. Each tour leaving home (home-based) or work (work-based) is modeled to have a number of stops ranging from one to nine – the primary destination and a maximum of four stops on each half-tour. Following the hierarchy of trip purposes, and depending upon the time or distance traveled, one of these stops is classified as the primary destination. All other stops on the tour are considered to be intermediate stops made on the way to or from the primary destination. In the destination choice models, each Transportation Analysis Zone (TAZ) is a potential alternative, whose attributes and accessibility determine its utility, and therefore its probability of selection as the chosen alternative. A tour-level and trip-level model was estimated for each purpose, for a total of eight separate models. For the purpose of developing the tour-level models, no differentiation is made among primary and secondary tours.

The work location choice model is at the "top" of the model system (see Figure 2.2). Therefore, this model is conditional only on the variables in the PUMS-based sample, including residence location, household characteristics, and person characteristics. The primary destination choice models for the other purposes come further down the model system (again, see Figure 2.2), and are conditional on the predicted vehicle availability, tour type (number of intermediate stops), and times of day (the time periods of the forward and backward half-tours) for the tour. The trip-level intermediate stop location models are applied after all other tour-level models are applied.

The study area has 1,739 TAZs, 766 of which are within San Francisco County. For trips originating in the City, each of the 1,739 TAZs is a potential location for destination choice. Since it is unwieldy to estimate models with such a large number of choices, the number of alternatives was limited to 40. In similar efforts for Boise, Idaho[5] and New Hampshire[6], this was found to be reasonable and practical.

In order to provide for a good representation of the observed behavior, the 'stratified importance sampling' technique was utilized for sampling the 40 potential alternatives from among the available 1,739 TAZs. This technique has been successfully applied previously for models of Portland[7], Boise, and New Hampshire. The choice set is first divided into a number of strata and each stratum is assigned a different level of importance, which determines the number of alternatives to be sampled from that stratum. This renders the sample to be reflective of the preferential choices of the decision-makers. In order to randomize the sample and make it representative of all alternatives available, a correction factor is applied to each

---

[5] Shiftan, Y. "A Practical Approach to Model Trip Chaining," *Transportation Research Record,* No. 1645, pp.17-23, 1999.

[6] Cambridge Systematics, Inc, *New Hampshire Statewide Travel Model System – Model Documentation Report*, prepared for the New Hampshire DOT, July 1998

[7] Cambridge Systematics, Inc., *Origin Location Choice Models for Portland International Airport Study – Documentation,* prepared for Portland METRO, October 1997.

sampled alternative. This factor is defined as the negative logsum of the selection probability of the alternative in the stratum to which it belongs. The logsum of the selection probability represents the composite probability of the choices within the stratum.

For this study, the strata were defined by the San Francisco County border (whether the trips are internal or external to the county limits), the origin and destination TAZ area type, and the travel time to the destination. The import assigned to each stratum is based on the observed distribution of trips. The specific sampling approach used for preparing the estimation data set depends on the tour purpose:

- For Work tours and Other purpose tours, the sampling approach is based on the area type distribution within San Francisco County and on travel time outside San Francisco County.
- For Education tours, the travel times from survey responses (excluding external destinations, were grouped into quartiles, and zones were sampled equally from each quartile range, ensuring the availability of a school in each zone
- For Work-Based tours, which are typically short both in activity duration and in trip length, the sampling procedures are based on their observed travel time distributions.

For all purposes, destinations internal and external to San Francisco were sampled in proportions found in the survey data. The travel time quartiles for Education- and Work-Based Tours were obtained by attaching peak and off-peak highway and transit travel times to the survey records. For the intermediate stop location choice models, the travel time information was used to create quartile ranges from each of which 10 zones were sampled. For these models, the travel time refers to the additional time incurred in traversing the extra stop on the way to or from the origin.

■ Mode Choice Models

The mode choice models developed for the San Francisco Model determine the mode for tours, and also for all trips made as part of tours, and are the basis for the accessibility measure (logsum) used in the tour destination choice models. The mode choice models differ from traditional "trip-based" mode choice models in that there are two distinct sets of mode choice models. The tour mode choice model determines the primary mode for the tour, while the trip mode choice models determine the mode for each individual trip made on that tour, based on the mode chosen for the tour. There is one of each model (tour and trip) for each tour purpose (Work, School, Other, and Work-Based); however, only the Work tour and trip mode choice models are described below.

An analysis of trips by mode revealed a significant percentage of transit trips and non-motorized (walk and bike) trips made by San Francisco residents. It also showed that there are a number of transit trips made by more than one transit mode; i.e., local bus access to BART. San Francisco can be considered a 'transit-rich' environment, where most residents have walk-access to transit, and the cost of parking is relatively high, accompanied by a limited supply. Based on this data, the following modes were defined and coded for trip mode choice model estimation:

- Drive-Alone
- Shared-Ride 2 (persons per vehicle)
- Shared-Ride 3+ (persons per vehicle)
- Walk
- Bike
- Walk-Local-Walk
- Walk-MUNI Metro-Walk
- Walk-Premium-Walk
- Walk-Premium-Auto
- Auto-Premium-Walk
- Walk-BART-Walk
- Walk-BART-Auto
- Auto-BART-Walk

These modes are used to model individual trips that occur on a tour. A more general tour mode was also defined; to control for the possible combinations of modes used for trips made on a tour. An analysis of the combinations of modes that occur on tours revealed that just one mode is used for most tours, and that much of the mode-switching that occurs on tours occurs in modal combinations including walk or auto-passenger with transit. This analysis was used to guide the definition of the tour modes and the structure of the tour and trip models. Tour modes were defined which allow the traveler to switch between modes where such behavior is most common. The tour modes were coded according to the following rules:

- **Auto Driver –** Tours that consist of trips primarily made by the driver of an automobile. If any trip on a tour was auto driver, the tour mode was coded as Auto Driver.
- **Auto Passenger –** Tours that consist of trips made entirely by passengers of automobiles. Walk trips on Auto Passenger tours were maintained.
- **Walk –** Tours that consist entirely of trips whose mode is walk.
- **Bike –** Tours that consist entirely of trips whose mode is either walk or bike.
- **Walk-Transit –** Tours that consist of trips made by transit passengers or combinations of transit and auto passengers. Walk trips on Walk-Transit tours were also maintained.
- **Drive-Transit –** Tours that consist of trips made by transit passengers where the access mode or egress mode is auto, or combinations of drive-transit, walk-transit, and Auto Passenger trips.

The tour mode definitions listed above allow the traveler to use walk as a mode for trips on any tour, and allow the traveler to switch between transit modes and auto-passenger modes for trips on transit tours. Table 4.2 shows the trip modes allowed for each type of tour mode.

Table 4.2 Trip Modes Allowed by Tour Mode

Trip Mode	Tour Mode					
	Driver	Walk	Bike	Passenger	Walk-Transit	Drive-Transit
Drive Alone	X					
Share-2	X			X	X	X
Share-3+	X			X	X	X
Walk	X	X	X	X	X	X
Bike			X	X	X	X
Walk-Local					X	X
Walk-MUNI					X	X
Walk-Premium					X	X
Walk-BART					X	X
Drive-Premium						X
Drive-BART						X

Figure 4.2 Illustration of Tour and Trip Mode Choices

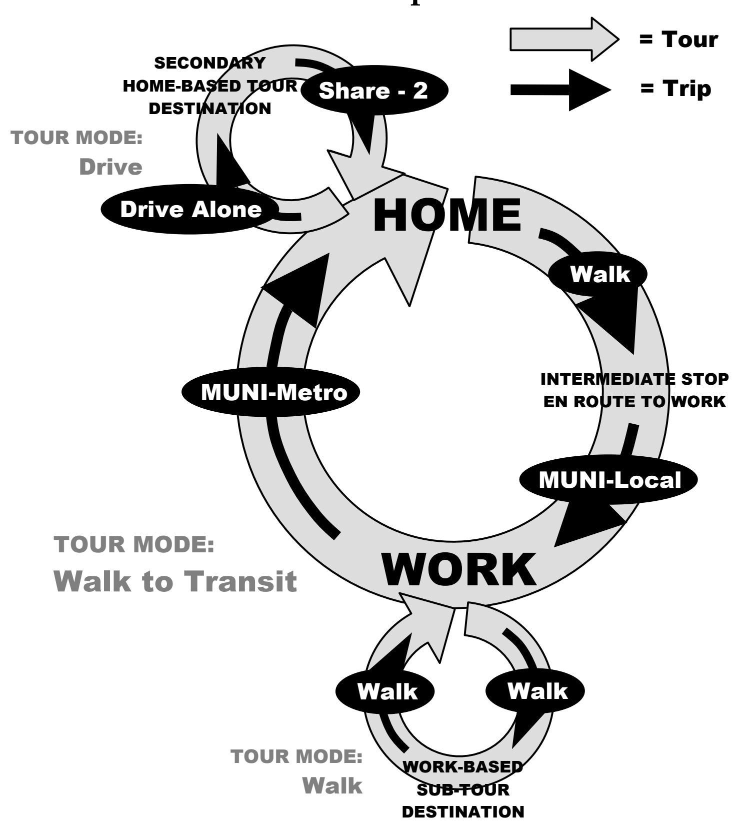


Figure 4.1 is a variation of the earlier figure showing a possible full day pattern for a San Francisco worker. This figure illustrates the function of the tour mode choice model and the trip mode choice model. The tour model choice mode has predicted the primary mode for each tour. In this example, walk to transit was selected as the primary mode for the home-based work tour, while walk was selected as the primary mode for the work-based subtour, and drive was selected as the primary mode for the secondary home-based tour. However, a traveler may use a combination of modes for each of the trips on the tour. For example, the worker may walk to their intermediate stop location en route to work, and from there pick up a local MUNI bus. However, they may choose to take the MUNI Metro on the way home. The traveler is not obliged to return by bus to their origin, even if they used bus to get to their destination. For the work-based subtour trips, the worker walks to and from their destination, so in this case there is no variation between the tour and trip mode choices. For the secondary home-based tour, the workers drives alone to the destination and shares a ride home - they have picked up a passenger at this intermediate stop. The mode choice models are structured in such a way to avoid illogical choices. For example, the worker would not drive to their secondary home-based tour destination, and then choose to walk home

The activity-based model estimation differs from traditional trip-based model estimation in at least two distinct ways. First, the models go beyond the traditional assumption that all work travel occurs in the peak period and non-work travel occurs in the off-peak period. The San Francisco Models use the actual time period (early AM, AM peak, midday, PM peak and evening) that travel occurred, in an attempt to more accurately reflect the travel conditions and modes available during that time period.

Furthermore, the tour models include the round-trip travel (both half-tours) characteristics for the tour. The LOS characteristics however do not include trips made to intermediate stops, since the location of these stops is not known when the tour mode choice model is applied.

One potentially innovative aspect of the mode choice models that was tested but ultimately not included was the inclusion of reliability and crowding as explicit variables in the transit utility functions. These variables were included in a stated-preference telephone survey of 407 transit users in San Francisco. Logit analysis was used to estimate tradeoffs between in-vehicle time, frequency of service, reliability (defined as the percent of days that the vehicle arrives five or more minutes late), and crowding ("low" = plenty of seats available, "medium" = few seats available, but plenty of room to stand, "high" = no seats available, standing room is crowded). It was estimated that improving the percent of vehicles arriving on time by 10 percent (e.g., once every two weeks) is equivalent to reducing the typical wait time (half the headway) by four minutes for commuters, or three minutes for non-commuters. It was also estimated that improving the level of crowding from "high" to "low" is equivalent to reducing the typical wait time by five minutes for commuters and nine minutes for non-commuters. Thus, relative to commuters, non-commuters are less sensitive to delay but more sensitive to crowding, on average. For application, the reliability and crowding for each link in the transit network was coded using observed system data collected by SFCTA. The tradeoffs estimated between these variables and wait time were applied in performing transit assignment and found to be not coincident with the observed boardings. As a result, these variables were not used in model application.

■ Visitor Models

Visitor models were estimated to predict the visitor trips by mode for San Francisco tourist destinations. These models were estimated using available visitor survey data collected in 1995 and 1998. The visitor models estimate the number of visitors to 29 destinations for each of three modes.

The model is developed as a series of multinomial logit (MNL) models that estimate the utility derived by a visitor in visiting a particular attraction/destination and in choosing a particular mode. Overall, 29 key visitor destination choices were modeled as a function of the Logsum variable and other destination specific information. Modal choices were determined based on utility functions specific to each mode. The maximum likelihood method of estimation was adopted to yield consistent and asymptotically efficient parameter estimates of the model.

The visitor models were developed by estimating destination choice and trip generation from the visitor surveys that were available in the San Francisco region. These models compared favorably to similar models developed in Honolulu. Mode choice models were borrowed from the Honolulu model development effort, since these tourist markets are somewhat similar and because the Honolulu model is one of the only visitor models estimated from visitor survey data. The visitor survey data in San Francisco did not have the available data needed to estimate mode choice models. Time-of-day factors were estimated from available traffic count data at select tourist destinations in San Francisco. These were applied to generate trip tables for each of the five time periods: early AM, AM peak, midday, PM peak and evening.

5.0 The Microsimulation Framework

The activity-based models in this project were incorporated into a series of C++ programs, described in the following paragraphs. Note that at present there are no standard software packages to apply these types of models, although it would certainly be feasible to standardize the types of programs we have developed and to integrate them into model application software packages (TransCAD, TP+, etc.). For the San Francisco Model, the visitor model, the peak spreading model and the highway and transit assignments are done with a pre-existing software package (TP+). The programs for each model component are described in the following sections.

■ Synthetic Sample Generator (Population Synthesis)

This program generates a full synthetic population for a base year or forecast year. Households are categorized simultaneously by household size/workers (9 classes: 1 person /0 workers, 1/1, 2/0, 2/1, 2/2, 3+/0, 3+/1, 3+/2, 3+/3+), age of head of household (3 classes: under 35, 35-61, 62 or over) and income class (4 groups). This gives 9 x 3 x 4 = 108 different household types. For each of the 766 TAZs in San Francisco County, we have the marginal distributions (observed or predicted) across the 3 categorizations. Iterative proportional fitting is used to translate those marginals down to level of the 108 combined categories. We also have additional information from the 1990 PUMS 5% micro-data sample, where we can observe the distribution across each of the 108 household types within each of the 6 PUMAS in the county. This distribution can be used to provide starting fractions for the IPF, as well as an additional target for fitting.

Once the appropriate number of households of each of the 108 types in each TAZ is determined, then individual households of those types are randomly sampled (with replacement) from the PUMS dataset for the appropriate PUMA region. This gives us a record for each household and person in the TAZ, while matching all of the marginal distributions as closely as possible. For San Francisco County, the resulting sample contains roughly 300,000 households.

■ Workplace Location Model

This program is run for each worker in the synthetic sample. Using the residence zone and a stratified sample of 40 possible work zones, program first applies the work tour mode choice model to calculate a mode choice accessibility logsum across all modes to each alternative work location. (Since we have not yet predicted the work tour type at this point, we assume an AM peak-PM peak work tour with no intermediate stops in either direction to calculate the logsum.) These logsums are fed into the work location choice model (similar to a destination choice model), to predict the probability of choosing each work zone. Using those probabilities, a single workplace zone is drawn randomly. This location, along with the car travel time and distance to reach it, is appended to the person record.

Vehicle Availability Model

All person records for a household are read in, and the predicted workplaces are compared to determine the maximum travel time to work. The vehicle availability model is applied to predict the probability of owning 0, 1, 2 or 3+ autos, and a single choice is drawn stochastically (randomly). The predicted number of vehicles is appended to the person and household records.

Full Day Tour and Trip Pattern

For each person record in the sample, the full day pattern model for the appropriate person type is applied to calculate the probability for each pattern alternative, and a random Monte Carlo procedure is used to predict a single pattern. The Monte Carlo process selects the pattern from the probability distributions of patterns for the given decision-maker.

If the "no travel" alternative is chosen, nothing further is predicted for that person. Otherwise, the primary tour time-of-day model for the appropriate tour purpose (work, education or other) is applied to calculate the probability for each time period combination, and a single combination is drawn stochastically. At that point a number of simple classification models are applied to "fill in extra details" in the tour and trip pattern:

- The exact number of secondary tours (2, 3, or 4) if 2+ is predicted.
- The exact number of work-based sub tours (1, 2, 3 or 4) if 1+ are predicted
- The trip chain types for any secondary or work-based tours, conditional on the predicted purpose and chain type for the primary tour.
- The time periods for any secondary or work-based tours, conditional on the predicted time periods for the primary tour.
- The exact number of intermediate stops (1, 2, 3 or 4) for any half-tour where 1+ stops are predicted.
- The departure time period from any intermediate stop, conditional on the predicted time periods for the tour and the predicted number of stops on the half-tour.

At this point, the predictions are written as a series of tours, each with:

- The tour purpose (work, school, other)
- The tour type (primary, secondary or work-based)
- The number of trip segments within each half-tour
- The time period in which each trip segment begins
- All household and person information for the person making the tour

Tour Primary Destination and Main Mode Choice

For each tour record, the tour main mode choice model is applied to a stratified sample of 40 alternative destination zones, and mode choice logsums and probabilities are calculated. The logsums are used in the primary destination choice models to predict the destination of any non-work tours. (The destination for work tours is already known.) Using the probabilities, a single primary destination and tour mode are drawn at random and added to the tour record.

Intermediate Stop Location and Trip Mode Choice Models

For each tour record, the intermediate stop location model is applied for any stops along the tour other than the primary destination. The stop location(s) are predicted conditional on the tour purpose, the tour origin and primary destination and the trip departure time. The model is applied to a stratified sample of destinations and a single one is chosen. Then, the trip mode choice model is applied for each stop along the tour, conditional on the predicted tour main mode and the origin, destination and time period of the trip. The tour is then split up into two or more trip records, and a record is written for each trip, with origin zone, destination zone, time period, tour purpose, and household and person identifiers. This output is essentially a simulated trip diary, which can be aggregated into trip matrices for highway and transit assignment, and can also be used in combination with the relevant person and household data to perform segmentation and equity analyses.

6.0 Model Validation

Details of the model validation results are in the corresponding model validation report. Highlights of these results are presented here for travel behavior and trip assignment.

■ Travel Behavior Validation

Travel behavior was validated by comparing travel data in a household travel survey to related travel data in the travel demand forecasting model. For the validation of the 1998 SFCTA regional travel demand forecasting model, we compared the trip data in the 1990 Census, the 1990 MTC household survey data with the same data in the model.

The model components were calibrated individually using various observed data sources, including the decennial census, household surveys, observed traffic counts and transit ridership, vehicle registrations, and many other sources. The specific sources used to calibrate each individual model are described below. This effort involved calibrating each model separately, then reviewing highway and transit assignment results for each of the five time periods to make additional adjustments in the model components. The adjustments were all made to constants within the models, there were no adjustments to model coefficients. Highlights of results of the calibration are summarized below for each model component.

Vehicle Availability

The vehicle availability model was calibrated primarily on two key variables, number of workers per household and super-district, using the 1990 Census as the primary source of observed data. A second validation test was used to evaluate the total number of vehicles estimated by the vehicle availability model compared to Department of Motor Vehicle (DMV) estimates of auto registrations. These data were different by 5 percent. Unfortunately, the 1990 MTC survey, which was used to estimate the model, contained different results for vehicle availability than the 1990 Census. Since, the 1990 Census has a much larger sample size; these data were used to calibrate the vehicle availability model. The results, therefore, have indirect effects on the market segmentation of autos and workers that were carried out in the mode split model.

Full-Day Pattern Tour Models

The full-day pattern tour models were calibrated by converting tours to trips and comparing these to the 1996 MTC household survey of San Francisco and Bay Area residents, expanded to match the 1998 population. The MTC survey trips were summarized as only those weekday trips in the survey that had an origin and destination within San Francisco County. The comparison of trips was developed from the full-day pattern tour model by reallocating the following "trips" from each "tour" for comparison purposes. The 1996 MTC Survey was used because the number of trips within San Francisco County was very low in the 1990 MTC Survey because of under-reporting of trips that occurred in this survey. The under-reporting of trips is not consistent across time periods or across trip purposes, which may have influenced model estimation that was based on the 1990 MTC survey. The differences between trips by time period was confirmed with initial assignments by time periods using the un-calibrated San Francisco model that revealed the off-peak time periods were significantly under-estimated compared to traffic counts. The vast majority of under-reporting of trips in the 1990 MTC survey were in other tours. A comparison of the calibrated San Francisco model trips to the 1996 MTC survey by tour type and time of day shows that the all trips by tour type and by time of day are within +/- 10 percent compared to the 1996 MTC survey.

Trip rates per household were compared by trip purpose and time of day. Trip rates overall are similar, but the trips per household by trip purpose are quite different. The San Francisco model differentiates between trips to work or school with an intermediate stop from those without an intermediate stop and thus has fewer trips identified as work or school trips and many more trips identified as non-home-based. The comparison of trip rates across time period is reasonable, except that early AM and evening time periods are somewhat under-estimated compared to the MTC survey. This is most likely a result of the model estimation process, which was based on the 1990 MTC survey that showed significantly fewer trips in these time periods.

Destination (Primary and Intermediate Stop) Choice Models

The destination choice models were calibrated against the 1990 MTC survey data for primary destinations by purpose and trip length frequency distributions. The results reflect very reasonable allocation of destinations among four areas of the City and those destinations located outside the City. Another evaluation of work locations is the estimate of employment that results from the work location model compared to actual employment by neighborhood. Because some of these data were not actually observed, these results were considered reasonable when compared to estimated values by neighborhood. The biggest differences were the two neighborhoods in the Core business district, which were underestimating employment, but calibration results also show that the destinations in the core are within three percent for each tour type and are actually overestimated in these results.

The destination choice model was also calibrated by comparing trip length and duration frequency distributions. The observed trip lengths are derived from the 1990 MTC survey and reported as the average time and distance to/from the primary destination. These results show reasonable average trip lengths for all tour types. Trip duration frequency distributions were evaluated to determine reasonable by tour purpose. Observed and estimated values of trip duration by travel time increment reflect reasonable comparisons.

The validation of the intermediate stop choice model was challenging because similar models of destination choice have not included separate validation of the intermediate stop choice component for comparison. The validation test was to review the total tour length by tour purpose compared to the observed values. Distance was selected as the primary validation test for this model to isolate the location of the destination from the congestion effects during a particular time period. The results of this validation test are that both work and other tours are over-estimated slightly by the model, while work-based tours are under-estimated. Additional calibration adjustments to try and reconcile these differences were not pursued because further adjustments would have negatively impacted the results of the highway assignments by time period.

Mode Choice (Tour and Trip) Models

The tour and trip mode choice models were calibrated by tour purpose. Alternative-specific constants for each mode were adjusted to match observed modal shares from the 1990 MTC Household Survey. The structure of the activity-based models require that tour models are calibrated first to match tours by mode and market segment, then trip models are calibrated to match trips by trip mode and tour mode. The trips resulting from applying the calibrated alternative-specific constants were then assigned to highway and transit networks and compared to observed traffic counts and transit boardings by mode. The calibration results for tour and trip modes show a very close match between estimated and adjusted observed tours and trips by mode and purpose.

Initially, estimated transit boardings were discovered to be much higher than observed boardings, particularly for local bus and MUNI Metro transit modes. There are four possible reasons for the transit over-estimation; there may be too many trips generated by the pattern models (too many trips going in to mode choice); the transfer rate may be too high; the calibration targets observed in the 1990 MTC survey may be incorrect; or, the observed transit boardings may be too low.

A comparison of estimated versus observed traffic volumes on the highway network confirmed that the number of trips generated by the pattern models was reasonable when compared to independent estimates of travel. An analysis of the estimated transfer rates also confirmed that the number of estimated transfers for San Francisco residents is reasonable. Therefore, it was concluded that either the transit calibration target values generated from the household survey were too high or the observed transit boardings are low. Because the transit boardings are calculated annually by MUNI, they were held constant and both the observed and estimated transit shares were adjusted to better match boardings.

Trip Assignment Validation

There are two primary modes for assignment validation: highway and transit. These were validated separately using observed volumes of vehicles and passengers on the highway and transit systems, respectively. Assignment validation at the county level was completed using aggregated volumes by corridor (identified by screenlines), type of service (facility type, mode or operator), size (volume group), and time period. Speeds and travel times are also used in highway and transit validations to ensure that these are accurately represented in the models.

Highway Assignment

The highway assignment results were compared for five individual time periods and the average daily results. The individual time period results were consistent with the daily results for each validation test, so the daily results were reported to represent all tests. Highlights of these results are presented below:

- The target error for screenlines was +/-10 percent. By this criterion, 8 of 10 screenlines met the criteria.
- A comparison of observed and estimated traffic volumes by facility type shows model volumes that are close to acceptable guidelines.
- A comparison of observed and estimated traffic volumes by volume group shows that all volume groups are well within the validation targets, except for the lowest volume group where one expects the higher level errors to occur. We believe that the high level of error on these lower volume facilities is greater in the San Francisco Model than it would be in other models because of the inclusion of all streets in the network, which may contribute to loading problems at the local level.
- A comparison of the average daily observed and estimated volumes by area type shows that the four area types are well within the target tolerance by area type.
- The validation targets by time period were original set by facility type, area type, screenline and neighborhood group, but the results did not vary significantly by time period. Only the evening time period is outside the 10 percent target range and this under-estimation was recognized during model calibration and significantly improved from initial estimates.
- Volumes by gateway were a critical part of the model calibration effort since the gateway volumes had significant impact on the volumes within San Francisco County. The gateway volumes were originally derived directly from MTC trip tables, stratified by time period, mode and purpose, but these did not produce volumes by time period at each gateway that were accurate enough for local planning purposes. We believe this is because the MTC assignments are daily assignments and the San Francisco model produces daily assignments by summing the five time period assignments. As a result, the MTC trip tables were adjusted to match traffic counts by time period at each gateway crossing.

- The model volumes and traffic counts were also summarized by neighborhood. These neighborhoods are used in various neighborhood planning. The target error for neighborhoods was set at +/- 20 percent and only one neighborhood falls outside this range: Laurel Heights.

Transit Assignment

San Francisco resident trip tables were constructed from trip mode choice model outputs and assigned to transit networks by time period and detailed mode. Non-San Francisco resident trip tables were constructed from MTC trip tables and assigned to identical transit networks by time period and MTC transit mode. Across all modes, the San Francisco models are within 5% of observed transit boardings. However, there are some distinct differences by time of day.

The time periods utilized by MUNI to report observed 1998 MUNI transit boardings are inconsistent with those used for the San Francisco models. The MUNI boardings by time period were converted to SFCTA time periods by applying conversion factors to total daily boardings by route. The time period specific boardings are therefore computed numbers, not observed, and therefore not as reliable as actual observed data. Additionally, though certain transit routes truly only run for a few hours in the early AM period, they may be included in the early AM skims which extend from 3:30 AM to 6:00 AM. Therefore the number of trips that are exposed to these routes may be inconsistent in many cases with the true number of trips that have the option of utilizing the routes. This inconsistency is also observed in the evening period.

Finally, it is shown that estimated bus boardings are significantly greater than observed boardings in the AM Peak period. As previously discussed in the section on Mode Choice calibration, the tour and trip transit shares were reduced in an attempt to better match bus boardings. However, matching the number of AM bus boardings within 5% would require a 30% reduction in Work transit tours compared to the observed 1990 MTC Household Survey data. An independent estimate of Census Journey-to-Work data indicates that the observed transit share of Work tours (35%) is reasonable. Therefore the observed Work Walk-Transit share was held constant, causing an over-estimation of AM period local bus trips.

A comparison of estimated versus observed transit boardings by route group, ranked by daily boardings shows a reasonable match between estimated and observed boardings, especially given the overall 8% over-estimate of boardings. The transfer rates were compared to estimated values (since actual values were not available) and are reasonable; bus mode transfer rates are lower than MUNI Metro and BART, and are typical of transfer rates for other urban areas.

7.0 Summary

Model Calibration and Application

The development of an activity-based microsimulation model for San Francisco was completed in approximately one year, for approximately the same resources required from a more traditional four-step modeling approach, with significant improvements in accuracy and capability of the models to test policies. The validation process took approximately 6 months longer to complete. The project did show that activity-based microsimulation models are practical and useful and can be developed from available data. There are significant advantages when comparing these models to four-step models, and when considering the range of policies that can be tested.

The models and application framework described herein were developed and implemented in a relatively short period of time. During the calibration and application of the model system, we learned a great deal about the behavior of the component models. Because the models can portray a wider variety of behavioral responses than most models (changes in trip chaining and departure times, trip suppression and induction, etc.), it takes some additional care to interpret the results, but this should result in more realistic forecasts of traveler responses to policy decisions.

To meet the tight time frame of the study, we made some simplifications relative to the application of this same approach at Portland METRO. This leaves us with some clear ways that the model system can be enhanced in the future:

- The "other" tour purpose can be broken down to estimate separate models for finer categories, such as maintenance and discretionary travel, and perhaps further sub-purposes such as serve passenger.
- Explicit logsum feedbacks can be used more extensively in the model system. For example, instead of using relatively simple accessibility measures in the tour pattern, tour time-of-day and vehicle availability models, the logsums from tour mode choice and destination choice can be carried up to these higher level models, as is done in the Portland models.
- Logsums from the intermediate stop location choice models can be carried up to use in the tour mode choice and destination choice models. For example, as intermediate stops become more accessible from a given destination using a given mode, the probability of choosing that destination and mode may increase.

Note that each of these enhancements, particularly the last two, add a good deal of complexity to the model estimation and application process, as well as requiring a great deal more computer time to run. By making such enhancements incrementally, it would be possible to see how much the forecast results are changed, and thus whether the additional complexity seems warranted.

Conclusions

The activity-based microsimulation approach has many advantages over traditional four-step travel forecasting models. Whereas traditional travel forecasting models are often disaggregate in estimation and aggregate in application, the microsimulation approach is disaggregate through the whole process, allowing the incorporation of a multitude of socioeconomic variables, as demonstrated successfully in this project. This project has also shown that it is possible to extend the microsimulation and activity-based approach beyond the academic arena and into practice.

The activity-based models described herein are able to reflect modal characteristics in destination choice, through the use of logsums in tour destination choice. The incorporation of tour characteristics, such as number of stops, is found to have a very significant impact on the choice of mode. This is usually unaccounted for in traditional models, potentially leading to inaccurate model indications in response to tested policies.

While the microsimulation approach greatly simplifies choice model application, it also introduces variation due to the nature of the stochastic process involved in choosing alternatives. There is a need for more research into how the distribution of random numbers drawn as part of this process leads to variation in model predictions both by destination and mode, and how these predictions influence both micro-level variables such as link-loadings and transit-boardings, and macro-level measures of system performance such as vehicle-miles of travel and vehicle emissions.

Acknowledgments

We wish to acknowledge the many contributors to the design, development and implementation of the San Francisco Travel Demand Forecasting Model as well as to the SFCTA staff who conceived the idea of developing an activity-based model for San Francisco. First, we would like to acknowledge José Luis Moscovich (with SFCTA), who developed the initial concept and succeeded in funding the initial efforts. We would also like to recognize Joseph Castiglione (SFCTA project manager) for his hard work, support and probing questions.

The consultant team includes those involved in the development of the data (Kevin Tierney, Kazi Ullah, Erin Vaca, Rob Hranac, Steve Decker, Frank Goetzke and Louisa Yue from Cambridge Systematics, Sally Nielson from Hausrath Economics Group, Mike Aronson from CCS and Ross Maxwell from Parsons Brinckerhoff); those involved in the development of the

model (Nageswar Jonnalagadda, Earl Ruiter, Arun Kuppam and Kiersten Haskell from Cambridge Systematics, William Davidson, Joel Freedman and Jim Hicks from Parsons Brinckerhoff, Mark Bradley from Mark Bradley Consulting & Research, T. Keith Lawton from Lawton Consulting and John Douglas Hunt from University of Calgary) and finally those involved in the implementation of the model (Vamsee Modugula from Cambridge Systematics and Larry Seiders from Urban Analysis Group). The consultant Project Manager was Maren L. Outwater from Cambridge Systematics.