MODELING PASSENGER VOLUMES AND PASSENGER-MILES ON THE MBTA RAIL LINES

Final Report

presented to the

Massachusetts Bay Transportation Authority

by

Peter G. Furth
Northeastern University

June 30, 1997

MODELING PASSENGER VOLUMES AND PASSENGER-MILES ON THE MBTA RAIL LINES

Peter G. Furth, Associate Professor
Northeastern University

ABSTRACT

This report describes of model of passenger flows on the rapid transit system of the Massachusetts Bay Transportation Authority (MBTA), developed at Northeastern University with the support of the MBTA. The model tries to replicate passenger flows that are consistent with three sources of data: (1) detailed entry, exit, and transfer counts, by station and direction, last made in 1989 (1985 for the Green Line); (2) a large-scale passenger survey, last undertaken in 1994; and (3) station entries counts, not by direction, conducted every year. The general approach is to estimate passenger origin-destination (OD) matrices for each hour of the day. Estimates of volume profiles (including peak volumes) and passenger miles are then derived from the estimated flows.

One purpose of the model is to provide an improved method for updating volume profiles from the 1989 / 85 counts to the present, based on current station entries. Volume profiles, particularly in peak hours, and key inputs to determining the needed service frequency and the average level of crowding experienced by passengers. Another purpose is to provide a means of estimating annual passenger-miles as required by FTA Section 15.

The model estimation is done in two stages. In Stage 1, the various historical data sources (the 1989/85 entry, exit, and transfer counts and the 1994 passenger survey) are combined using a maximum likelihood procedure into a "historical seed matrix" (actually, set of fourteen hourly OD matrices). The maximum likelihood procedure smooths out inevitable inconsistencies in the data, and yields estimated OD matrices that, in some sense, best fit the historical data. In Stage 2, the historical seed is updated to the annual station entries counts using a multiproportional fit model. The Stage 2 estimates provide, in some sense, a best fit to the historical seed for the given station entries, with a further constraint that production / attraction balance at each station be preserved.

The model was implemented with two programs, SEEDMAKER for Stage 1 and PROFILEMAKER for Stage 2, ready to be run on a MacIntosh personal computer. PROFILEMAKER can be run each year to provide estimates matching that year's station entries counts (or more often, if station entries are collected more often). SEEDMAKER has been run with the available historical data, and need not be run until the 1989/85 counts or the 1994 survey is replaced.

The passenger flow estimates produced by the model are superior to estimates made by simple factoring because they are internally consistent in terms of flow balance. Their accuracy has been estimated to be reasonable for management reporting and decision making. It can produce Section 15 passenger-miles estimates that satisfy FTA accuracy standards if weekend, as well as weekday, station entries counts are supplied as an input. The model structure is readily adaptable to accepting improved data, such as might be obtained from a new fare collection system or some targeted data collection efforts, which would further improve the accuracy of its estimates.

TABLE OF CONTENTS

Acknowledgements

Our appreciation is extended to Mr. Alan Castaline, Chief Planning Officer of the MBTA, for his support throughout this project in providing us with data and technical feedback. We also appreciate his patience.

Thanks also to Tom Humphrey and the CTPS staff for providing us with data.

Thanks also to two graduate students, Omca Korugan and Huimin Zhao, who spent many hours preparing data files, modifying programs, doing test runs, producing tables, and many other important tasks.

1. INTRODUCTION

This report describes of model of passenger flows on the rapid transit system of the Massachusetts Bay Transportation Authority (MBTA), developed at Northeastern University with the support of the MBTA. The model tries to replicate passenger flows that are consistent with available passenger counts and survey data. Estimates of volume profiles and passenger miles are then derived from the estimated flows.

1.1 Purpose

For this project, passenger flows on the rapid transit system are modeled at the level of station to station origin-destination (OD) flows by one-hour periods. This level of detail can be useful for various planning purposes, such as estimating trip diversions to a new crosstown line. However, the main purpose of this project was to develop volume profiles and passenger-mile estimates. Volume profiles by one-hour time periods, including peak volumes, are important for determining the appropriate amount of service to provide, and to monitor crowding levels. Passenger-mile estimates are necessary for so-called Section 15 reports to the Federal Transit Administration (FTA), and they are a measure of utilization of the system that can be used in year to year comparisons and in economic analyses.

Without a model such as this, methods available to the MBTA for estimating volume profiles are highly inadequate. Full scale counts of boardings and alighting by direction at each station will establish volume profiles, but such counts are too costly to make more often than once every 5 to 10 years. The most recent such count was undertaken by the Central Transportation Planning Staff (CTPS) in 1989 on three of the lines (Red, Orange, Blue) and in 1985 on the Green Line. In the absence of a model of passenger flows, the best that can be done is to scale each line's 1989 or 1985 volume profile to account for changes in entries on each line. Simply scaling like this will not account for shifts in the location of the peak point or in other changes in the peaking pattern. Another estimation method that has been used is to have a checker on a platform estimate volumes on passing trains – a daunting task when trains are crowded, yielding estimates that can only be taken with a grain of salt.

Estimating passenger-miles with confidence has likewise been difficult for the MBTA. One way to determine passenger-miles is from volume profiles (simply multiplying line volume on each segment by segment length). Obviously, estimates made from dubious volume profiles become themselves dubious. Another method is from questioning passengers about their origin and destination stations. Traditional large-scale surveys leave much to be desired in estimating passenger miles due to their low, and highly variable, response rates. Their costly nature also makes it impossible to conduct such a survey annually. Recently, CTPS has inaugurated a different method of surveying for passenger-miles, by orally asking entering passengers to simply declare their destination station. Hopefully this method will prove satisfactory. Nevertheless, this method still involves considerable labor cost, and the MBTA was rightly interested in development of an estimation technique that does not require any special data collection.

1.2 Data Sources

The MBTA has three sources of passenger flow data. These sources vary in their level of detail and timeliness.

1.2.1 Station Entry Counts

The most timely source is station entry counts, conducted annually by the MBTA. The counts are done by having inspectors record turnstile registers every hour. Turnstiles with improperly working registers are closed for the day of the counts. Cash received in the collector's box (for discount fares, e.g. children and elderly) is counted over the entire day and allocated proportionally among the hours. A few stations also have gates at which inspectors allow passengers to enter without passing through a turnstile (either because they have a pass or have exact change to drop into a farebox); gate passages are counted over the whole day and allocated over the hours. While entries counts are done every year, they have the least level of detail. They do not indicate direction of travel. Also, no destination information is obtained.

1.2.2 Passenger Survey

The most detailed source is a full-scale passenger survey. Because of their cost, they are conducted only about once every twenty years. However, the latest survey was quite recent, done in 1994. Questionnaires were distributed to one weekday's worth of entering passengers at every station from 6:30 a.m. to 3:30 p.m., asking, among other things, origin station, destination station, and station at which passengers leave the first train (from which transfer path information can be obtained). Time of day was also coded. There were about 37,000 responses, representing about 20% of the surveyed passengers. As is usual in surveys of this sort, the response rate varies significantly by station, and other factors such as crowding can be expected to create significant variance in the response rate, making the results unfit for use without suitable expansion.

1.2.3 CTPS Passenger Counts

The third source of data is detailed passenger counts last conducted by CTPS in 1989 and (on the Green Line) in 1985. They include passenger ons and offs at each station by direction and time period (15 minute intervals were used). They also include counts of transfer flows at the downtown transfer stations and at Arlington Station. These detailed counts require many man-hours of work, and because of their cost are conducted only once every five to ten years. At high volume stations with multiple entries, often several people are needed to counts ons and offs by direction. At the downtown transfer stations, transfer flows are estimated by comparing and reconciling counts made by dozens of checkers at strategic locations (e.g., a count of people leaving trains on a particular platform and heading down a particular corridor, which could lead to another platform or to an outside exit).

The detailed 1989/85 counts provide sufficient information to define volume profiles on each line in each direction. For the purpose of creating volume profiles, it was helpful to use time periods defined relative to a reference station on each line. The reference station was always one of the downtown transfer stations. For example, the 7-8 a.m. time period on the Red Line represents flows on trains entering the Downtown Crossing station between 7 and 8 a.m. CTPS converted their 15-min station counts into hourly counts for relative time periods in order to create volume profiles for each relative period.

1.3 Overview Of The Model

1.3.1 Two Stages: Generating Historical Seed, And Fitting To Current Entries

The methodology used has two stages. In the first stage, implemented in the program SEEDMAKER, a historical seed matrix is constructed from the 1994 survey and the 1989/85 detailed counts – data that is not collected annually. The historical seed only changes when new detailed counts or a new passenger survey are done. The historical seed matrix is one primary product of this project.

In the second stage, implemented in the program PROFILEMAKER, the historical seed is fit to the annual entries counts to yield an estimate of current passenger flows, from which volume profiles and passenger-mile estimates are generated. The facility to develop these profiles and estimates is the second primary product of this project.

1.3.2 Time Periods And Period Crossovers

The model uses one hour periods, and models the day from 7 a.m. till 9 p.m. (The span of the day covered is limited by the span used for the 1989/85 detailed counts.) The length of period chosen, one hour, is a compromise between a very short period (e.g., 15 min), and a long period. The shorter period would allow for a more exact representation of fluctuations in demand, but are subject to large fluctuations in observed flows because a delay of a few minutes on a train could switch it from one period to the next. Short periods also suffer from a high likelihood of zero observed flow between station pairs on a given day. Long periods are less subject to random fluctuations, but cannot accurately represent systematic variations in the peaking pattern over the day. The one hour length chosen is long enough to be robust with respect to a single train switching periods, and short enough to represent the major peaking pattern. Furthermore, one hour periods are used by the MBTA in various service evaluation and scheduling measures.

The one hour periods used in the model are relative periods, relative to the reference station of each line, as was done by CTPS in their line profiles. For OD flows, the flow in a given period represents the number of travelers beginning their trips on trains that pass their reference station in that period. Therefore, passengers that do not change trains do not cross from one period to another. However, passengers making transfers can cross from one period to another, as illustrated by the following examples. Accounting passengers crossing from one period to another added considerably to the complexity of this project, but doing so was necessary to generate consistent line profiles.

> <u>Example 1.</u> The Orange Line uses Downtown Crossing as its reference station, while the Blue Line uses State as its reference station. It is at State that the Orange and Blue Lines meet. Suppose a passenger boards the Orange Line at Forest Hills on a train that passes through Downtown Crossing at 8:56 a.m. That passenger contributes to the volume on the Orange Line in the 8-9 period. However, by the time the Orange Line train gets to State, it is 8:58. By the time the passenger walks to the Blue Line and waits for the next Blue Line train, it will be about 9:02, so this passenger will

contribute to the Blue Line volume in a different period. In this example, passengers cross from one time period to the later time period. The difference in time at the lines' respective reference stations is 6 min (2 min travel, 1 min walking between platforms, and 3 min waiting). Therefore about 10 percent (6 min / 60 min) of the hourly volume between Forest Hills and the Blue Line will cross into a later period.

Example 2. Consider passengers traveling from the Ashmont branch to the Braintree branch of the Red Line, whose reference station is Downtown Crossing. Downtown Crossing lies 8 minutes north of JFK/UMass station where the two branches meet. Suppose a passenger boards an inbound train that will pass through Downtown Crossing at 8:05 a.m. He gets off at JFK/UMass station at 7:57, and, allowing 1 min to change platforms and 4 min to wait, gets on a Braintree train around 8:02. However, that train would have passed through Downtown Crossing at 7:54. Therefore, while this passenger contributed to the 8-9 a.m. volume on the first (the inbound) train, he contributes to the 7-8 a.m. volume on the second train (the outbound train). The net movement backwards in time is 11 minutes (8:05 to 7:54), so about 18 percent (11/60) of the Ashmont - Braintree travelers will cross into the previous period.

2. NETWORK DESCRIPTION

To adequately model station to station flows and line volumes, a network was formulated that includes lines, station platforms, transfer stations, and segments.

2.1 Lines And Stations

There are four lines (Red, Orange, Blue, Green), numbered 0 to 3. Each line has a unique set of stations. Therefore, a transfer station such as Park is represented by two stations: Park on the Red Line, and Park on the Green Line. On the Green Line, one station called Surface BCD and one called Surface E represent travelers entering and leaving the portals at Kenmore and Symphony, respectively. Altogether, there are 68 stations, numbered 0 to 67.

Each line is assigned a reference station (one of the downtown transfer stations). The period in which a passenger travels is determined by the period in which the train they ride leaves the reference station. The reference stations, which were used by CTPS in the report of their 1989 and 1985 detailed counts, are as follow:

Line	Reference Station
Red	Downtown Crossing
Orange	Downtown Crossing
Blue	State
Green	Park

File STATIONSF lists the stations, indicating for each its line, whether it is a reference station, and its distance and travel time from the end of the line.

2.2 Directions Of Travel Between Station Pairs

Each non-transfer station has two platforms, one for each direction. (Even if physically there is only a single middle platform, we think of it as being two platforms.) The directions of travel are called southbound (leaving Alewife, Lechmere, Oak Grove, and Wonderland) and northbound. For each station pair, there are four platform pairs; however, only one of them is valid. A station-direction matrix, saved as file STADIRF, indicates for each station to station pair which platform or direction pair is valid, as follows:

1: leave the origin station's southbound platform, arrive at the destination station's southbound platform
2: leave southbound, arrive northbound
3: leave northbound, arrive southbound
4: leave northbound, arrive northbound

From symmetry it will be clear that the opposite (reverse direction) of 1 is 4; the opposite of 2 is 2; the opposite of 3 is 3; and the opposite of 4 is 1. For each station pair (i,j), the direction pair indicator for (j,i) must be the opposite of the indicator for (i,j). This symmetry provided a check on the indicator values in the data file.

2.3 Transfer And Branch Transfer Stations

The MBTA system has six downtown transfer stations. Each of them is indexed (0 to 5), and the station file (STATIONF) indicates to which transfer station, if any, each station belongs. For example, Park on the Red Line and Park on the Green Line both belong to the transfer station Park.

In our network there are also three branch transfer stations: JFK/UMass, where passengers transfer between the two branches of the Red Line; Arlington, where passengers transfer between the E branch and the BCD trunk; and Kenmore, where passengers from Surface BCD can transfer to [a different branch that belongs to] Surface BCD.

2.4 Segments And Period Crossovers

As mentioned previously, transferring passengers can cross from one time period to another. For any movement, the period crossovers are three numbers that sum to one describing the fraction of travelers beginning that movement in a given period and ending in the previous period, the same period, or the following period. Stations whose crossover patterns are all identical are grouped into segments. There are 16 segments, numbered 0 to 15, illustrated in Figure 1. The STATIONF file includes, for each segment, the index of the segment to which it belongs. Because segments are not always contiguous (due to branching), it also includes, for stations at the northern end of a segment, the index of the next northern station.

More formally, the period crossovers are defines as follows:

$XOVER_{stx}$ = fraction of travelers going from segment s to segment t whose travel ends x periods before the period in which it began, for x = -1, 0, 1

2.5 Transfer Flows, Transfer Pairs, And Path Shares

Just as at a four way intersection there are 8 possible turning movements (left or right from each of four approaches), so at a transfer station there can be many possible transfer flows. In this context, a transfer flow does not describe passengers' origin or destination; it refers to a flow of passengers, by foot, from one platform to another. While it is possible to calculate passenger miles without knowing transfer flows, a complete volume profile should include transfer flows. And since the 1989/85 detailed counts include observations of most transfer flows, our model had to have a way of calculating them in order to try to match the observed flows to the estimated flows.

The file TFERSF lists and describes the 52 transfer flows, numbered 0 to 51, and describes them by detailing the line, station, and direction passengers are coming from, and the line, station, and direction passengers are going to. In addition to the standard 8 movements at a transfer station, Park and Downtown Crossing include transfer movements between each other via their connecting concourse. Also, some physically possible movements were omitted because nobody makes them (because there is a superior path available); for example, nobody transfers at North Station between the southern sections of the Green and Orange Line, because the same transfer can be made sooner and more easily at Haymarket.

To relate transfer flows to OD flows, for each transfer flow there is a record for each segment pairs (abbreviated as tpairs) contributing to that transfer flow in the file TFERSF. For example, the following segment pairs contribute to the transfer flow at Park from the Red Line southbound to the Green Line northbound: (0,9), (0,10), (0,7), and (0,8). All five pairs begin in segment 0 (Red Line north of Park). The first two pairs are destined for the two Green Line segments north of Park. The last two pairs are destined for the Blue Line, since passengers going from Red Line stations to Blue Line stations contribute to the subject transfer flow.

For some segment pairs passengers use more than one path. Results of the 1994 passenger survey were used to determine the shares of passengers using the alternative paths. For example, for passengers going from the northern side of the Red Line to the northern side of the Blue Line, 84% went via the Green Line, transferring at Park and Government Center, while the remainder went via the Orange Line, transferring at Downtown Crossing and State. The tpair records in file TFERS include the share of that segment pair contributing to the subject transfer flow.

Transfer flows involve some crossover between periods. For example, not everyone leaving a Red Line platform at Park in the 7-8 period reaches the Green Line platform at Park in the same period. For transfer segment pairs other than Red Line - Blue Line pairs, the crossover in the transfer flow is the same as the segment to segment crossover. Because Red Line - Blue Line pairs involve a double transfer, their crossover in the first transfer (e.g. Red Line to Green Line at Park) is not the same as their segment to segment crossover, but instead is the same as the crossover between the origin Red Line segment and the intermediate Green Line segment to which they are transferring.

In addition to the downtown transfer flows, the network has five branch transfer flows, numbered 0 to 4: two at JFK/UMass (Ashmont branch to Braintree branch, and vice versa), two at Arlington, and one at Kenmore (inbound BCD to outbound BCD).

Each has one or more branch transfer pairs (abbreviated bpairs) contributing to it, defined analogously to the tpairs used with regular transfer flows. The original network data for branching transfers is not in a data file, but is coded into the program.

In addition to checking the original data and program files for network data, users of the computer program can check the program's echo files (ECHO.OUT and ECHOM) that echoes not only the original network data, but also internally created network structures such as lists of the various transfer flows leaving at arriving at each transfer station platform.

3. PREPARING OD MATRICES FROM THE 1994 SURVEY

The 1994 passenger survey was coded into records, one per respondent, under the supervision of CTPS. There were approximately 37,000 records. The data was analyzed using SPSS to create frequency tables indicating the number of passengers going between each station pair in three periods of the day. Because the survey ran from 6:30 a.m. to 3:30 p.m., the three periods were a.m. peak (6:30 to 9:30), midday early (9:30 to 12:30) and midday late (12:30 to 3:30), based on the time the passenger entered the system. Overall, the data was found to be very clean, with few invalid records due to misspelling, etc. Fewer than one hundred records were rejected for being invalid.

Some of the survey data pertaining to the Green Line had to be modified to be consistent with the way travel was modeled. Passengers transferring within a surface Green Line branch were excluded (since surface travel is not part of the model), but transfers between branches were retained, including transfers between B, C, and D branches occurring at Kenmore. Transfers between the E branch and the other branches also required some scrutiny. For these movements, the shortest transfer path timewise is at Copley, but this transfer requires exiting the station to the street, crossing the street, and reentering the station. This is a popular path for pass users, who don't have to pay another fare when reentering. A little over 30 percent of the respondents going between the E branch and the BCD trunk and branches reported transferring at Copley. (Nearly all of the remainder transferred at Arlington, where one can transfer within the station.) But in our model such a transfer is not properly a transfer, but a break between two separate trips, because those passengers' exits and entries at Copley are included in the station entry and exit counts. Therefore, respondents reporting a transfer at Copley had their trips recoded as two trips, one ending at Copley and the other beginning at Copley. The file SURVEYF contains the three 68 x 68 survey summary OD matrices.

Because response rates in questionnaire type surveys vary widely between OD pairs, it is naive and incorrect to simply apply a single scalar expansion factor. If that were done, the expansion factor would have been about 7 (implying a response rate of 1/7, or 14%). Because response rates can vary widely between origin stations (due to differences in crowding, trip length, literacy, and cooperativeness), it is common with surveys of this type to apply origin-specific expansion factors, to make the total entries at each station agree with some control counts of station entries. A still more thorough method of expansion is to apply destination-specific expansion factors as well, so that the expanded survey margins agree with control counts for both station entries and exits. Destination-specific expansion was especially called for in this project because the survey was not performed after 3:30 p.m., with the expectation that most travel after 3:30 consists of return trips, for which the destination becomes the origin. Therefore it was important that response biases with respect to both origin and destination be corrected. Both origin and destination specific expansion was done using the biproportional method (also known as iterative proportional fit, or IPF) using the 1989/85 CTPS counts between 7 a.m. and 4 p.m. as control totals for station entries and exits by direction. The outcome of this process is expansion factors such that the expanded survey data matched the 1989/85 CTPS entry and exit counts by direction. The origin and destination station specific expansion factors are presented in Table 1. The pattern exhibited in this table is that expansion factors for the downtown stations tend to be three to five times greater than others (suggesting a very low response rate at these stations). Also, as expected, the origin specific expansion factors are lowest in the northwest corridor, where response rates are traditionally the greatest, and are significantly higher in areas of lower English literacy and fewer white collar workers. The destination specific expansion factors do not vary nearly as much.

A final set of period scale factors was also developed to convert the three-hour periods into which the survey was summarized into one-hour periods for the model. One hour periods between 7 a.m. and 10 a.m. were based on the a.m. summary; those between 10 a.m. and 1 p.m. on the midday early summary; those between 1 p.m. and 4 p.m. on the midday late summary; and those after 4 p.m. on the transpose of the a.m. summary. (The unfortunate difference in 30 min between the boundaries of the one hour periods and the three hour periods is due to differing periods of data collection between the data sources, and is somewhat offset by the fact that the survey data is based on absolute time while the entry and exit control counts were based on relative periods. Before 3:30 p.m. the majority of entries on each line take place before the downtown reference station; therefore, most of the survey data times would have had to be shifted 10 to 20 min later to make them correspond more exactly to the control count times.) The period scale factors were chosen such that, for each one hour period, the corresponding three-hour OD survey summary, after applying origin-specific, destination-specific, and period-specific expansion factors, agrees with the total system entries for that one hour period in the 1989/85 CTPS counts. The period scale factors are presented in Table 2. As a rule, the period scale factors should be around 0.33, since their primary purpose is to convert a three hour summary into a one hour summary. They vary around this value, reflecting peaking within the three hour periods. The factors for the evening periods are particularly low, because the survey summary period used for these periods is the transpose of the a.m. peak, and the period scale factors reflect the ratio between total boardings in the evening period and total boardings in the 3 hour a.m. peak.

4. DETAILED COUNTS FROM 1989 AND 1985

4.1 General Description Of The Counts

In 1985, CTPS performed detailed platform level counts on the Green Line. In 1989, similar counts were performed on the other three lines of the system. They include ons and offs by station and direction. At the downtown transfer stations, simultaneous observations were made at strategic points throughout the station and carefully correlated, yielding estimates of street entries and exits to / from each platform, as well as platform to platform transfer flows. Estimating these flows at the downtown transfer stations was a massive undertaking, with 44 simultaneous observers at Park Street station in 1985 (updated by observations from 6 observers in 1989), 23 at Downtown Crossing, 10 at State, 23 at Government Center, 11 at Haymarket, and 7 North Station.

Table 1. On / Off expansion factors to match the 1989/85 CTPS entry and exit counts

LINE	STATION	ROW (on)	COL (off)
Red	ALEWIFE	3.951	0.743
Red	DAVIS	3.754	0.75
Red	PORTER	3.748	0.888
Red	HARVARD	6.013	0.835
Red	CENTRAL	7.465	1.162
Red	KENDALL	5.195	0.601
Red	CHARLES/MGH	9.63	0.748
Red	PARK	22.497	0.82
Red	DOWNTOWN_X	24.223	0.985
Red	SOUTH_STA	8.38	0.628
Red	BROADWAY	7.458	1.192
Red	ANDREW	10.337	2.095
Red	JFK/UMASS	10.394	0.989
Red	SAVIN_HILL	7.212	1.913
Red	FIELDS_COR	15.57	2.258
Red	SHAWMUT	8.104	2.397
Red	ASHMONT	7.56	1.819
Red	N.QUINCY	6.142	1.025
Red	WOLLASTON	5.782	1.56
Red	QUINCY_CTR	5.418	1.337
Red	QUINCY_A.	4.488	1.207
Red	BRAINTREE	3.196	1.148
Orange	OAK_GROVE	3.278	1.131
Orange	MALDEN_CTR	5.082	0.919
Orange	WELLINGTON	5.971	1.082
Orange	SULLIVAN_SQR	8.253	1.802
Orange	COMMUNITY_CO	3.833	1.568
Orange	NORTH_STA	7.978	0.86
Orange	HAYMARKET	14.373	1.063
Orange	STATE	21.663	0.954
Orange	DOWNTOWN_X	37.531	0.872
Orange	CHINATOWN	22.106	1.199
Orange	NE_MEDICAL_C	18.627	1.02
Orange	BACK_BAY	7.133	0.709
Orange	MASS_AVE	4.759	1.13
Orange	RUGGLES	18.097	1.487
Orange	ROXBURY_X	17.245	1.831
Orange	JACKSON_SQR	15.715	2.719
Orange	STONY_BROOK	5.812	1.429
Orange	GREEN_ST	3.376	0.917
Orange	FOREST_HILLS	7.177	1.025
Blue	WONDERLAND	4.129	1.372
Blue	REVERE_BEACH	9.122	1.876
Blue	BEACHMONT	7.179	2.371
Blue	SUFFOLK_D	9.219	3.386
Blue	ORIENT_H	5.374	1.589
Blue	WOOD_ISLAND	8.045	2.587
Blue	AIRPORT	14.788	1.115
Blue	MAVERICK	13.066	2.268
Blue	AQUARIUM	17.567	0.965
Blue	STATE	19.086	0.695
Blue	GOVERNMENT_C	23.367	0.909
Blue	BOWDOIN	10.585	0.825
Green	LECHMERE	6.46	0.755
Green	SCIENCE_PARK	14.864	1.301
Green	NORTH_STA	5.964	0.69
Green	HAYMARKET	6.673	0.561
Green	GOVERNMENT_C	15.853	0.73
Green	PARK	20.037	1.058
Green	BOYLSTON	11.382	0.8
Green	ARLINGTON	14.877	1.251
Green	COPLEY	9.653	1.056
Green	HYNES/ICA	8.953	0.942
Green	KENMORE	11.773	1.102
Green	SURFACE_BCD	7.289	1.395
Green	PRUDENTIAL	11.444	1.155
Green	SYMPHONY	14.917	1.368
Green	SURFACE_E	16.934	1.678

Table 2. Period scale factors

Time	7:00	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00
factor	0.26	0.38	0.18	0.33	0.34	0.38	0.41	0.46	0.57	0.31	0.35	0.16	0.11	0.06

For estimating line flows on each line, CTPS shifted the counts in time to refer to flows affecting trains passing through their line's reference station (listed in Section 2.1). It is these referenced flows that were used as model inputs. At transfer stations, platform to platform flows were given with reference to both the period of the line from which passengers were transferring and the period of the line to which passengers were transferring; the model uses the former (although they differ little because the time shift between lines is very small for downtown transfers since all four lines' reference stations are very near one another in the downtown).

At Arlington station, transfer flows were measured along with street entries / exits. All transfers there are westbound to eastbound (called northbound to southbound in our model). The transfer counts do not distinguish between flows from E trains to B, C, and D trains and the reverse. There were no transfer counts made in 1985 r 1989 at JFK/UMass or at Kenmore. For Kenmore, an artificial count was generated representing westbound to eastbound (called northbound to southbound in our model) transfers between the B, C, and D branches. Based on results from the 1994 survey, the transfer counts were set equal to 3 percent of the Surface BCD entries in any given period. (An attempt to estimate passenger flows without a Kenmore transfer count resulted in an unrealistically large estimate for Kenmore transfer flows, probably because the model found that the easiest way to reconcile the 1985 counts and the 1989 counts was to keep a large number of surface Green Line users out of the underground system. The artificial transfer count was introduced to give the model a reasonable target for Kenmore transfers. The final estimated flows do not match the target, of course, just like other estimated flows do not exactly match their target, but at least they do not deviate too much from them. No such problem occurred at JFK/UMass, where transfer flows are small.)

Because of the large number of observers needed to make all the detailed counts, they could not, of course, be counted on one day. Rather, the results are a composite of counts made on several days. For this reason, as well as due to measurement error, the counts are not consistent. This means, for example, that total ons do not equal total offs. Rather, each estimate of a passenger flow is an observation of a process in which there is some random variation from day to day.

4.2 Production / Attraction Imbalance

One aspect of the 1989/85 CTPS counts is the large discrepancies they contain in platform level productions and attractions. If every traveler made a simple round trip in the period of observation (7 a.m. to 9 p.m., based on time at the reference stations), there would be a perfect balance between southbound ons and northbound offs at a station, and likewise between northbound ons and southbound offs. Of course, it would be wrong to expect a perfect balance, for three reasons. First, some trips are made outside the study period. This can be expected to make productions somewhat greater than attractions at outer stations, and the reverse at central stations, because of round trips that began during the study period but whose return trips was after 9 p.m. Second, everybody does not make a simple round trip. This can lead to both systematic and random variations. An example of systematic variation is at Ruggles and Jackson Square stations. Passengers in 1989 transferring from bus routes 22 and 29 to the Orange Line generally preferred to transfer at Jackson Square, because that makes the ride to downtown quicker; but on the return trip, they preferred to transfer at Ruggles in order to get a seat. Thus, the data show, as expected, an excess of productions at Jackson Square and an excess of attractions at Ruggles. Random variations, which could lead to a production imbalance in either direction, can be due to people may use transit in one direction, and another mode in the other. Third, the counts were made on different days, and random day to day variations will cause some production / attraction imbalance.

So, although perfect production / attraction balance is not expected, the degree of imbalance still appears large. The imbalances at each platform are shown in Table 3. As one can see, there is a tendency for the outer stations to have an excess of inbound productions, and for central stations to have an excess of attractions. In part, this must be due to trips after 9 p.m., most of which are outbound, that were not counted.

In using the 1989/85 counts as a base for estimating more recent flows, a decision had to be made: whether to force productions and attractions to balance, or to preserve the imbalance observed in the 1989/85 counts. We chose the latter, because of the undeniable systematic imbalances that should be expected due to omitting late evening trips and due to travel patterns favoring some stations for inbound travel and others for outbound travel. Consequently, the model's estimates exhibit some significant production / attraction imbalances. If these imbalances are considered too large, the CTPS counts should be revised in a manner that reduces this imbalance and the model recalibrated to those adjusted inputs.

5. ANNUAL ENTRIES COUNTS

Each year, the MBTA conducts one day of entries counts at its stations. Inspectors test the turnstiles, and close those that are not properly registering entries; then they read the turnstile registers every hour. At most stations, counts are made from 6 a.m. till midnight, although at some stations the first valid count begins at 7 a.m. and the last valid count ends at 9 or 10 p.m. Passengers paying a non-standard fare (e.g., elderly and students) that enter via the collector's box (abbreviated Sbox) are not counted; rather, an average fare per Sbox entry factor, taken from the most recent fare mix study, is applied to convert daily Sbox revenue into passengers. At some stations, inspectors speed entries by allowing passengers with passes and exact change to enter through a "gate." Gate entries are counted, but only over the whole day, not by hour. Sbox and gate entries are distributed by the model over the hours of the day in proportions to hourly turnstile entries.

Because of the nature of the entries counts, there is no data for a few of the stations in the model: the surface E and the surface BCD stations, for which entries occur over many surface stations with on board fare collection, and Science Park, where most fare collection is done on board.

Because of manpower requirements, entries counts are performed over several days, and will therefore exhibit some inconsistency. Like the 1989 counts, they should be seen as observations of a process with random variation. It is probable that human and mechanical error are also present in the counts. Also, it is probable that the proportionate use of the gate and Sbox is not constant all day. When the MBTA acquires new electronic fareboxes, which will accurately record entries every period of every day and eliminate gate and Sbox entries, the quality of entries data will vastly increase.

Table 3

Production / Attraction Imbalance in the 1989 / 85 CTPS Counts

7 a.m. to 9 p.m., based on time at downtown station

		SB	*NB*	*Imbalance*		*NB*	*SB*	*Imbalance*	
Line	Station	*Prod*	*Attr*	*Prod-Attr*	*Percent*	*Prod*	*Attr*	*Prod-Attr*	*Percent*
RED	ALEWIFE	7,713	6,654	1059	14%	-	-	0	0%
RED	DAVIS	6,360	5,180	1180	19%	183	170	13	7%
RED	PORTER	5,386	4,440	946	18%	239	350	-111	-46%
RED	HARVARD	15,078	16,005	-927	-6%	2,871	3,240	-369	-13%
RED	CENTRAL	7,214	7,380	-166	-2%	2,119	2,369	-250	-12%
RED	KENDALL	5,707	6,163	-456	-8%	3,353	3,740	-387	-12%
RED	CHARLES/MGH	4,395	3,940	455	10%	2,826	3,046	-220	-8%
RED	PARK	5,776	4,985	791	14%	4,562	5,422	-860	-19%
RED	DOWNTOWN_X	6,793	8,286	-1493	-22%	6,010	5,999	11	0%
RED	SOUTH_STA	5,231	5,211	20	0%	9,428	8,787	641	7%
RED	BROADWAY	1,316	1,136	180	14%	2,125	2,240	-115	-5%
RED	ANDREW	671	647	24	4%	3,094	2,829	265	9%
RED	JFK/UMASS	1,821	1,770	51	3%	3,707	4,193	-486	-13%
RED	SAVIN_HILL	119	135	-16	-13%	987	1,119	-132	-13%
RED	FIELDS_COR	422	623	-201	-48%	3,767	3,862	-95	-3%
RED	SHAWMUT	24	32	-8	-33%	1,042	1,197	-155	-15%
RED	ASHMONT	-	-	0	0%	7,511	7,882	-371	-5%
RED	N.QUINCY	955	694	261	27%	3,812	3,997	-185	-5%
RED	WOLLASTON	252	257	-5	-2%	3,169	2,897	272	9%
RED	QUINCY_CTR	113	162	-49	-43%	4,723	5,345	-622	-13%
RED	QUINCY_A.	35	28	7	20%	3,513	2,994	519	15%
RED	BRAINTREE	-	-	0	0%	3,136	3,420	-284	-9%
ORANGE	OAK_GROVE	3,660	3,525	135	4%	-	-	0	0%
ORANGE	MALDEN_CTR	6,555	5,998	557	8%	103	105	-2	-2%
ORANGE	WELLINGTON	4,849	5,186	-337	-7%	137	106	31	23%
ORANGE	SULLIVAN_SQ	7,379	7,429	-50	-1%	339	408	-69	-20%
ORANGE	COMMUNITY_C	2,048	2,549	-501	-24%	606	545	61	10%
ORANGE	NORTH_STA	5,902	6,681	-779	-13%	1,502	1,517	-15	-1%
ORANGE	HAYMARKET	2,776	2,700	76	3%	1,735	1,852	-117	-7%
ORANGE	STATE	6,638	8,661	-2023	-30%	5,511	6,717	-1206	-22%
ORANGE	DOWNTOWN_X	4,832	7,033	-2201	-46%	7,462	4,228	3234	43%
ORANGE	CHINATOWN	1,317	1,544	-227	-17%	2,454	2,305	149	6%
ORANGE	NE_MEDICAL	2,005	2,103	-98	-5%	3,103	2,964	139	4%
ORANGE	BACK_BAY	2,850	3,224	-374	-13%	9,969	8,073	1896	19%
ORANGE	MASS_AVE	868	1,057	-189	-22%	2,831	2,709	122	4%
ORANGE	RUGGLES	728	923	-195	-27%	5,691	7,415	-1724	-30%
ORANGE	ROXBURY_X	491	547	-56	-11%	2,411	1,711	700	29%
ORANGE	JACKSON_SQR	266	290	-24	-9%	3,993	2,704	1289	32%
ORANGE	STONY_BROOK	154	159	-5	-3%	1,706	1,357	349	20%
ORANGE	GREEN_ST	149	174	-25	-17%	1,599	1,682	-83	-5%
ORANGE	FOREST_HILL	-	-	0	0%	8,758	8,321	437	5%
BLUE	WONDERLAND	4,646	4,744	-98	-2%	-	-	0	0%
BLUE	REVERE_BCH	1,863	1,493	370	20%	31	31	0	0%
BLUE	BEACHMONT	1,885	1,944	-59	-3%	121	119	2	2%
BLUE	SUFFOLK_D	1,438	1,480	-42	-3%	76	117	-41	-54%
BLUE	ORIENT_H	3,101	3,357	-256	-8%	118	161	-43	-36%
BLUE	WOOD_ISL	1,352	1,080	272	20%	77	83	-6	-8%
BLUE	AIRPORT	3,570	3,350	220	6%	299	318	-19	-6%
BLUE	MAVERICK	6,596	6,481	115	2%	418	491	-73	-17%
BLUE	AQUARIUM	1,516	1,795	-279	-18%	1,524	1,840	-316	-21%
BLUE	STATE	349	292	57	16%	5,078	5,621	-543	-11%
BLUE	GOVT_C	-	21	-21	0%	1,736	2,715	-979	-56%
BLUE	BOWDOIN	-	-	0	0%	2,363	2,753	-390	-17%
GREEN	LECHMERE	3,750	3,575	175	5%	-	-	0	0%
GREEN	SCIENCE_PK	635	776	-141	-22%	99	113	-14	-14%
GREEN	NORTH_STA	3,312	2,822	490	15%	202	301	-99	-49%
GREEN	HAYMARKET	2,217	1,815	402	18%	534	287	247	46%
GREEN	GOVT_C	8,001	9,030	-1029	-13%	2,097	2,288	-191	-9%
GREEN	PARK	12,913	14,685	-1772	-14%	1,468	1,558	-90	-6%
GREEN	BOYLSTON	2,465	3,725	-1260	-51%	1,770	1,522	248	14%
GREEN	ARLINGTON	5,387	5,894	-507	-9%	7,598	7,855	-257	-3%
GREEN	COPLEY	6,026	5,895	131	2%	9,601	9,875	-274	-3%
GREEN	HYNES/ICA	2,288	3,023	-735	-32%	4,266	4,118	148	3%
GREEN	KENMORE	1,619	2,394	-775	-48%	5,926	4,920	1006	17%
GREEN	SURFACE_BCD			#N/A	#N/A	38,003	40,055	-2052	-5%
GREEN	PRUDENTIAL	677	762	-85	-13%	1,459	1,478	-19	-1%
GREEN	SYMPHONY	316	570	-254	-80%	1,777	1,226	551	31%
GREEN	SURFACE_E			#N/A	#N/A	15,793	15,893	-100	-1%
RED	*TOTAL*	75,381	73,728	1653	2%	72,177	75,098	-2921	-4%
ORANGE	*TOTAL*	53,467	59,783	-6316	-12%	59,910	54,719	5191	9%
BLUE	*TOTAL*	26,316	26,037	279	1%	11,841	14,249	-2408	-20%
GREEN	*TOTAL*	49,606	54,966	-5360	-11%	90,593	91,489	-896	-1%
ALL	**TOTAL**	204,770	214,514	-9744	-5%	234,521	235,555	-1034	0%

5.1 Reconciling Absolute And Referenced Times

The entries counts are made in absolute time. There are two options for reconciling them to the model's flows, which are in referenced time (i.e., referenced to a downtown stations). The first is to convert the model flows to absolute time; the second, to convert the entries counts to referenced time. In either option, the conversion is made by shifting flows in time according to the station's distance (in minutes) from the reference station by taking a weighted average of the two original periods contributing to the shifted period.

Each option presents some difficulties. The first, converting flows to absolute time, is theoretically possible, and was the approach originally used. However, numerical results showed that finding flows whose weighted averages matched the absolute entries was mathematically impossible unless negative flows, a practical impossibility, were allowed. Option 1 is very sensitive to absolute entries counts that widely vary from period to period, as occasionally occurs, and cannot at all accommodate entries counts of zero (unless there are zero entries in all periods), which occurred in a few periods at a few stations.

For this reason, the second option is used. Converting entries counts to absolute flows is straightforward if one knows the direction of travel, since the shift between absolute and reference time depends on the direction of travel. Of course, the entries data do not give direction of travel. Therefore this option required that a directional split be estimated, based on the seed matrices derived from the 1994 survey and the 1989/85 detailed counts. For example, if the historical data indicated that 90% of the entries at a station in the 8-9 a.m. period are inbound, the entries are split using this fraction between inbound and outbound, and then the directional flows are shifted (later for inbound, earlier for outbound) according to that station's distance in time from the reference station.

For the downtown transfer stations, the entries data not only do not specify direction; they do not specify the line. For this reason, converting these counts to referenced time would require an additional layer of assumptions based on historical data. However, the downtown stations either are reference stations, or are very close to a reference station, and so the time shift needed would be very small. Therefore the approach taken for the downtown transfer stations was not to shift them in time, and not to split them between lines. In updating the model flows to the entries data at a transfer station, each period's sum of model entries from both lines serving the station are matched to the entries counts.

6. MATHEMATICAL METHOD FOR GENERATING HISTORICAL SEED

The historical seed consists of fourteen station to station OD matrices for one hour periods beginning at 7 a.m. and ending at 9 p.m. It represents the "best" (in some sense) estimate of the historical flow pattern exhibited in four sets of data: ons (from the 1989/85 detailed counts), offs (also from the 1989/85 detailed counts), transfers (also from the 1989/85 detailed counts), and OD flows (from the 1994 passenger survey). The problem of estimating an OD matrix using various data sources has been studied extensively in the literature, and that knowledge, as well as the author's experience in estimating OD matrices in other situations, was used to develop the procedure used in this project.

6.1 Exact Matching Vs. Maximum Likelihood Estimation

The methods of OD estimation can roughly be divided in two groups: those that provide an exact fit to the data, and consequently require that all the data be consistent, and those that attempt to provide the "best" fit to all the available data, without requiring consistency in the data. The first group is simpler, and is appealing in that it provides an exact fit. The problem of requiring consistent data can often be solved by forcing consistency, e.g. by scaling the ons and offs so that total ons equals total offs. This was the originally conceived approach. However, with transfer data as well as ons and offs, and with transferring flows crossing between periods, the problem of adjusting the data to be consistent without biasing the data became almost as complicated as the original problem. Therefore this approach had to be abandoned in favor of a method that can accept inconsistent data and finds a "best" fit.

Of the "best fit" approaches, research (e.g. Ben Akiva et. al. [Transportation Research Record 1037, pp. 1-11, 1985]) has shown that they are very similar in their results, and so the choice can be based more on modeling or mathematical considerations. For our purposes, the method chosen was Maximum Likelihood Estimation (MLE) with OD flows assumed to have a Poisson distribution (a common assumption in transportation modeling), with each data element treated as an observation of the underlying process of random arrivals with background rates that vary by OD pair and by period. The logic of the method is that it finds the set of background rates (in the form of one OD matrix for each period) that were more likely than any other to have given rise to the observations – the on and off counts, the transfer counts, and the OD survey results.

6.2 Weighting The Various Data Sources

The MLE approach is easily adaptable to allowing different weights for the data to reflect differing degrees of measurement accuracy. Essentially, one can treat a day's worth of counts as though it is really is average of several days' data in order to increase that data's importance, or as an average based on counts done during only half the period in question and then expanded in order to decrease the importance that data's importance. The weights used were as follow, based on a subjective assessment of data accuracy:

- Survey data, periods before 4 p.m.: weight = 1 (base level). These observations, expanded as described earlier, are compared with the model's estimated OD flows.
- Survey data, periods after 4 p.m.: weight = 0.5, reflecting increased uncertainty. No data was collected for these periods, but rather the transpose of the a.m. peak period was used.
- Ons: weight = 4, since on counts are usually the most reliable, and because counts do not exhibit non-response bias inevitably present in survey data. These observations are compared with the model's estimated station entries (by direction), obtained by making partial row totals of the estimated OD matrices.
- Offs: weight = 3. These observations are compared with the model's estimated station exits (by direction), obtained by a process similar to making column totals of the estimated OD matrices, complicated somewhat by the phenomenon of period crossovers. Because of randomness in actual period crossovers, the weight is smaller than the weight for ons.

- Transfers and branch transfers: weight = 2, because of the complexity in the process used to determine transfer flows in the 1989/85 CTPS study. These observations are compared with the model's estimated platform to platform transfer rates, obtained by summing estimated OD flows that contribute to the transfer flows, accounting for period crossovers and path shares in cases of path choice.

6.3 Dealing With Boundary Periods

Because transferring passengers cross from one period to another, passengers traveling in the 7 a.m. to 9 p.m. span are not a closed system. Some passengers begin before 7 a.m. (relative to their origin line's reference station), but after transferring are in the 7 - 8 a.m. period. Others begin before 9 p.m., but end after 9 p.m. Branch transfers move the opposite way: they may begin in the 7 - 8 a.m. period, but end before the 7 a.m. period. There are two ways to deal with the boundary periods: (1) give them different crossover fractions, essentially forcing the system to be closed, or (2) including in the model shoulder periods, 6 - 7 a.m. and 9 - 10 p.m., that are for the most part ignored, except insofar as they contribute to flows in the boundary periods (7 - 8 a.m. and 8 - 9 p.m.). The latter method, which is more realistic and flexible, was followed.

The flows in the shoulder periods were not matched to any data, since for the most part there was no data in the shoulder periods. Rather, they were assumed to be a fraction of the neighboring period flows. The a.m. shoulder period flows affect the 7-8 a.m. period only in that passengers who are transferring at a downtown station can be pushed (due the transfer delay) to a later period. The most a transferring passenger can be pushed forward in time is about 10 min, so the ratio we want is the ratio of the flow rates (on an hourly basis) in the (referenced) period 6:50 - 7:00 a.m. to 7-8 a.m. Based on the 1989 CTPS counts, which included counts for some stations in the (referenced) 6:45-7:00 period, that ratio was estimated to be 0.454, using data for inbound boardings only. Meanwhile, the only p.m. shoulder period flows affecting the 8-9 p.m. period are those involving branch transfers at Arlington and JFK/UMass, because, only those flows can result in a crossover to an earlier period. Using 1989 CTPS counts for inbound boardings on the Green Line south of Arlington and on the Red Line south of JFK/UMass, no difference in flow rate between the 9:00 - 9:15 p.m. period and the 8-9 p.m. period was observed, so the p.m. shoulder ratio was 1.0.

6.4 Formulation Of Likelihood Function

The likelihood function is probability that the observed flows actually occurred, given the OD matrices containing the background rates. Its description requires a fair amount of mathematical notation, and is not necessary for a general understanding of the model. The full formulation will be given in a technical paper to be presented shortly.

6.5 Optimization Algorithm

Following standard practice, we maximize the natural logarithm of the likelihood function (known as the log likelihood), because there is a one-to-one correspondence between the likelihood function and its log. The algorithm used to maximize the log likelihood is a gradient search, whose mathematical details will not be spelled out in this report. However, a general description of the optimization procedure will be given in intuitive terms.

What is being estimated is an OD matrix of background flows; actually, one OD matrix for each of 14 periods. From these OD matrices, summary flows can be found by appropriate summation, such as ons at a station by direction in a certain period, or transfer flows from one platform to another. The algorithm compares these estimated summary flows to the observations of those flows, i.e. the observed ons, offs, and transfers. In addition, the model OD flows themselves are compared to the OD flows observed in the 1994 survey, adjusted by the appropriate origin and destination expansion factors and the period scale factor. Each comparison between a model estimate and an observation provides a term that contributes to a huge sum called the log likelihood. The closer the model estimates are to the corresponding observations, the greater the sum. The greatest possible value the log likelihood could have would be if every model estimate exactly equaled its corresponding observation. However, this ideal maximum is possible only if the observations are all consistent with each other, which they are not. The inconsistency requires that some of the estimates be greater than their corresponding observations, and others less. The optimization routine iteratively reduces the discrepancy between estimated and observed flows until only the minimum necessary discrepancy is left, with overestimates just balancing underestimates.

The optimization algorithm was implemented with a C program, and run to obtain the best fitting historical seed. An analysis of that resulting set of matrices is given in the next section.

7. ANALYSIS OF HISTORICAL SEED

The historical seed, consisting of 14 station to station OD matrices for hourly periods from 7 a.m. to 9 p.m. (based on each line's reference station), are found in the file SEEDHIST.1. This section provides a brief analysis of this seed, based upon which current estimates can be made from current entries counts.

7.1 Improvement In Log Likelihood

One measure of success in obtaining a seed that fits the historical data well is the improvement in the log likelihood function. Unfortunately, it is not possible to assign an intuitive meaning to its value. Table 4 presents the value of the log likelihood function for four cases:

- an initial estimate made using entry and exit counts by station and period (but not by direction), assuming independence between origins and destinations.
- another initial estimate made using the survey data expanded by origin, destination, and period specific factors so as to match all day entries by station, all day exits by station, and systemwide entries by period;
- the optimal estimate, i.e. the estimate with the greatest attainable value of log likelihood, making use of entry and exits counts by station, direction, and period, transfer counts by period, and the survey;

- the ideal value, i.e. the value that would have been obtained had all of the data been internally consistent so that it would have been possible for estimated flows to match exactly all of the observed flows.

Comparing these values, one can see the improvement betweenthe initial estimates and maximum likelihood estimation. A statistical (chi-squared) test based on the log likelihood values readily proves that the optimal estimates are "significantly different" in a statistical sense from both of the initial estimates. The realtively small gap between the optimal estimate and the ideal suggests that the problems of internal inconsistency are small.

Table 4
Improvement in Log Likelihood Value

Case	**Assumption**	**Log Likelihood**	**Difference from Optimal**
Independence	Based on entries & exits by station and period, assuming independence between entry and exit station	21.102E+6	-740000
Expanded Survey	Triproportional expansion of the survey to match all day entries by station, all day exits by station, and period total entries	21.535E+6	-307000
Optimal	Maximum likelihood fit to entries and exit counts by station, direction, and period; transfer counts by period; and survey	21.842E+6	0
Ideal	Result if estimated flows exactly matched observed flows, disregarding inconsistency	21.860E+6	18000

7.2 Comparison To Historical Data

The historical seed is supposed to provide the best possible fit to the historical data. While the likelihood function is one measure of how good the fit is, it is worthwhile to provide an informal, visual comparison. For this purpose, the 8 - 9 a.m. period was selected.

First, a comparison of the model estimate's entries and exits by direction to those of the 1989/85 CTPS counts is presented in Table 5. Part (a) shows a comparison for Period 1, which is 7-8 a.m. Part (b) shows a comparison of daily totals. One can see that the estimates match the data rather closely. The maximum likelihood model tries to minimize relative, rather than absolute, discrepancy, and so one can see that differences are very small when the observed value is small, and are proportionately larger when the observed

Table 5-a Comparison of Ons Offs by Directon, Pd 1

			Ons, Direction 0			Ons, Directon 1			Ons, Total			Offs, Directon 0			Offs, Direction 1			Offs, Total		
STA	LINE	STA	Est.	Obs.	Diff (%)	Est.	Obs.	Diff (%	Est.	Obs.	Diff (%)	Est.	Obs.	Diff (%	Est.	Obs.	Diff (%)	Est.	Obs.	Diff (%)
0	Re	ALWF	1168	1204	3	0	0	0	1168	1204	3	0	0	0	236	233	1	236	233	1
1	Re	DAVS	823	846	3	32	34	5	856	880	3	9	9	1	114	113	1	123	122	0
2	Re	PORT	558	567	2	16	16	2	574	583	2	15	15	1	85	83	2	99	98	1
3	Re	HARV	1057	1078	2	76	78	3	1132	1156	2	209	198	5	665	657	1	874	855	2
4	Re	CENT	728	747	3	103	105	2	831	852	2	168	162	3	426	429	1	594	591	0
5	Re	KEND	146	148	2	34	34	0	180	182	1	363	339	7	1062	1081	2	1425	1420	0
6	Re	CHAS	138	141	2	121	116	4	259	257	1	307	295	4	603	613	2	910	908	0
7	Re	PARK	39	39	1	72	71	1	111	110	1	391	355	10	804	808	0	1195	1163	3
8	Re	DTXG	43	43	1	50	49	1	92	92	0	834	801	4	1271	1283	1	2105	2084	1
9	Re	SSTA	185	192	3	691	683	1	877	875	0	1298	1255	3	1057	1056	0	2355	2311	2
10	Re	BWAY	120	126	4	199	194	3	319	320	0	266	256	4	385	395	3	651	651	0
11	Re	ANDR	9	9	4	335	330	1	344	339	2	20	17	16	88	89	1	108	106	2
12	Re	JFKU	135	141	5	331	326	1	465	467	0	403	384	5	308	317	3	711	701	1
13	Re	SAVN	19	20	4	197	196	0	216	216	0	18	17	6	1	1	0	19	18	6
14	Re	FLDS	42	44	4	822	814	1	864	858	1	117	111	5	64	65	1	181	176	3
15	Re	SHAW	5	5	6	272	272	0	277	277	0	16	15	5	5	5	0	21	20	4
16	Re	ASHM	0	0	0	1900	1897	0	1900	1897	0	248	236	5	0	0	0	248	236	5
17	Re	NQCY	110	117	6	697	689	1	806	806	0	357	342	4	114	117	3	470	459	2
18	Re	WOLL	38	41	6	1048	1052	0	1087	1093	1	41	39	6	8	8	1	49	47	5
19	Re	QCTR	19	20	4	971	966	1	990	986	0	220	208	6	15	15	3	235	223	5
20	Re	QADM	14	15	7	595	583	2	609	598	2	129	123	5	3	3	3	132	126	5
21	Re	BRNT	0	0	0	1266	1297	2	1266	1297	2	121	115	5	0	0	0	121	115	5
22	Or	OAKG	1199	1221	2	0	0	0	1199	1221	2	0	0	0	54	55	1	54	55	1
23	Or	MALD	1346	1343	0	8	8	3	1354	1351	0	16	17	7	319	330	3	334	347	4
24	Or	WELL	1517	1530	1	11	11	0	1528	1541	1	11	11	2	160	162	2	170	173	2
25	Or	SULL	1265	1257	1	33	32	2	1298	1289	1	61	62	2	499	510	2	560	572	2
26	Or	CCOL	142	142	0	41	39	4	183	181	1	100	101	1	266	272	2	366	373	2
27	Or	NSTA	813	801	1	79	78	1	891	879	1	427	435	2	678	699	3	1105	1134	3
28	Or	HAYM	179	175	2	55	54	1	233	229	2	367	373	2	529	545	3	896	918	2
29	Or	STAT	38	38	1	28	27	3	66	65	2	1549	1543	0	1231	1252	2	2781	2795	1
30	Or	DTXG	37	36	3	74	72	2	111	108	2	724	716	1	1012	1039	3	1736	1755	1
31	Or	CHTN	25	25	1	54	53	2	80	78	2	390	391	0	205	207	1	594	598	1
32	Or	NEMC	145	144	0	159	155	3	304	299	2	622	630	1	441	456	3	1063	1086	2
33	Or	BBAY	223	225	1	1331	1309	2	1554	1534	1	1371	1402	2	590	614	4	1960	2016	3
34	Or	MASS	57	57	1	321	311	3	378	368	3	222	224	1	105	108	3	327	332	2
35	Or	RUGG	45	45	0	674	659	2	719	704	2	590	592	0	191	197	3	781	789	1
36	Or	ROXB	25	25	0	270	264	2	296	289	2	125	125	0	140	145	3	265	270	2
37	Or	JACK	66	66	1	933	915	2	998	981	2	101	100	1	17	17	2	118	117	0
38	Or	STON	45	46	3	234	227	3	279	273	2	22	22	1	5	5	4	27	27	0
39	Or	GREE	23	24	6	222	215	3	245	239	2	85	86	1	14	15	5	99	101	2
40	Or	FORE	0	0	0	1767	1737	2	1767	1737	2	274	275	1	0	0	0	274	275	1
41	Bl	WON	1249	1320	5	0	0	0	1249	1320	5	0	0	0	51	54	6	51	54	6
42	Bl	REVE	313	325	4	5	5	2	318	330	4	3	3	7	14	14	1	17	17	1
43	Bl	BEAC	440	459	4	5	5	4	445	464	4	8	8	4	24	25	6	32	33	3
44	Bl	SUFF	131	136	4	5	5	2	136	141	3	1	1	10	4	4	5	5	5	4
45	Bl	ORNT	657	686	4	5	5	4	662	691	4	16	15	5	72	77	6	88	92	4
46	Bl	WOO	240	250	4	10	10	3	251	260	4	2	2	10	37	39	6	39	41	5
47	Bl	AIRP	164	170	3	17	16	4	181	186	3	44	43	3	289	309	7	333	352	5
48	Bl	MAVE	1257	1309	4	29	28	5	1286	1337	4	64	61	5	213	226	6	277	287	4
49	Bl	AQUA	50	51	3	35	33	5	84	84	0	289	265	9	187	200	7	476	465	2
50	Bl	STAT	6	6	3	28	26	6	33	32	4	910	841	8	50	54	7	960	895	7
51	Bl	GOVT	5	5	8	12	11	5	16	16	1	372	348	7	2	2	15	374	350	7
52	Bl	BOWD	0	0	0	29	27	8	29	27	8	717	675	6	0	0	0	717	675	6
53	Gr	LECH	488	432	13	0	0	0	488	432	13	0	0	0	423	366	16	423	366	16
54	Gr	SCPK	30	26	17	4	5	12	35	31	13	13	15	15	46	38	20	58	53	10
55	Gr	NSTA	436	381	14	19	21	10	455	402	13	16	19	14	176	149	18	192	168	14
56	Gr	HAYM	257	225	14	60	68	12	316	293	8	23	27	17	159	136	17	181	163	11
57	Gr	GOVT	156	136	14	8	9	8	164	145	13	223	265	16	808	662	22	1032	927	11
58	Gr	PARK	221	194	14	51	56	9	272	250	9	139	161	14	973	804	21	1112	965	15
59	Gr	BOYL	80	71	13	38	42	9	118	113	5	165	184	10	298	250	19	463	434	7
60	Gr	ARLT	330	290	14	67	72	7	396	362	9	1394	1619	14	577	491	18	1971	2110	7
61	Gr	COPL	339	293	16	200	218	8	539	511	5	1256	1467	14	537	449	20	1793	1916	6
62	Gr	HYNE	126	110	14	328	366	10	454	476	5	283	331	14	271	226	20	555	557	0
63	Gr	KENM	119	107	11	310	341	9	429	448	4	444	514	14	141	118	20	586	632	7
64	Gr	SBCD	0	0	0	3276	3616	9	3276	3616	9	1967	2260	13	0	0	0	1967	2260	13
65	Gr	PRUD	78	69	12	105	116	10	182	185	2	193	226	15	105	90	16	298	316	6
66	Gr	SYMP	34	30	13	133	148	10	167	178	6	50	58	14	26	22	19	76	80	4
67	Gr	SURE	0	0	0	1310	1456	10	1310	1456	10	1144	1321	13	0	0	0	1144	1321	13
Avg Abs Error				11			13			20			24			16			26	
Avg % Diff				4			4			3			7			6			4	

Table 5-b Comparison of Ons Offs by Direction over all periods

			Ons, Direction 0			Ons, Direction 1			Ons, Total			Offs, Direction 0			Offs, Direction 1			Offs, Total		
STA	LINE	STA	Est.	Obs.	% Diff	Est.	Obs.	% Diff	Est.	Obs.	% Diff	Est.	Obs.	% Diff	Est.	Obs.	% Diff	Est.	Obs.	% Diff
0	Re	ALWF	7576	7713	2	0	0	0	7576	7713	2	0	0	0	6510	6654	2	6510	6654	2
1	Re	DAVS	6242	6360	2	186	192	3	6428	6552	2	164	170	3	5086	5180	2	5251	5350	2
2	Re	PORT	5297	5386	2	238	242	2	5534	5628	2	337	350	4	4323	4440	3	4660	4790	3
3	Re	HARV	14876	15078	1	2837	2871	1	17713	17949	1	3192	3240	1	15362	16005	4	18554	19245	4
4	Re	CENT	7108	7214	1	2151	2119	1	9258	9333	1	2363	2369	0	7121	7380	4	9484	9749	3
5	Re	KEND	5631	5707	1	3426	3353	2	9058	9060	0	3746	3740	0	5927	6163	4	9673	9903	2
6	Re	CHAS	4288	4395	2	2918	2826	3	7207	7221	0	3101	3046	2	3786	3940	4	6887	6986	1
7	Re	PARK	5579	5776	3	4679	4562	3	10258	10338	1	5531	5422	2	4737	4985	5	10268	10407	1
8	Re	DTXG	6670	6793	2	6216	6010	3	12886	12803	1	6167	5999	3	7864	8286	5	14031	14285	2
9	Re	SSTA	5150	5231	2	9734	9428	3	14883	14659	2	9073	8787	3	4984	5211	4	14057	13998	0
10	Re	BWAY	1288	1316	2	2217	2125	4	3504	3441	2	2319	2240	4	1089	1136	4	3409	3376	1
11	Re	ANDR	657	671	2	3213	3094	4	3871	3765	3	2914	2829	3	618	647	5	3532	3476	2
12	Re	JFKU	1769	1821	3	3866	3707	4	5634	5528	2	4324	4193	3	1665	1770	6	5989	5963	0
13	Re	SAVN	127	129	1	1022	987	4	1149	1116	3	1147	1119	3	128	135	5	1275	1254	2
14	Re	FLDS	413	422	2	3917	3767	4	4330	4189	3	3998	3862	4	589	623	6	4586	4485	2
15	Re	SHAW	68	70	3	1078	1042	3	1146	1112	3	1232	1197	3	33	35	5	1266	1232	3
16	Re	ASHM	0	0	0	7756	7511	3	7756	7511	3	8137	7882	3	0	0	0	8137	7882	3
17	Re	NQCY	914	955	4	3963	3812	4	4877	4767	2	4150	3997	4	648	694	7	4799	4691	2
18	Re	WOLL	244	256	5	3264	3169	3	3508	3425	2	3019	2897	4	239	257	7	3257	3154	3
19	Re	QCTR	116	121	5	4884	4723	3	4999	4844	3	5535	5345	4	152	163	7	5688	5508	3
20	Re	QADM	75	80	6	3621	3513	3	3696	3593	3	3118	2994	4	29	31	7	3147	3025	4
21	Re	BRNT	0	0	0	3206	3136	2	3206	3136	2	3566	3420	4	0	0	0	3566	3420	4
22	Or	OAKG	3675	3660	0	0	0	0	3675	3660	0	0	0	0	3603	3525	2	3603	3525	2
23	Or	MALD	6605	6555	1	106	112	6	6711	6667	1	96	105	8	6085	5998	1	6181	6103	1
24	Or	WELL	4890	4849	1	138	143	4	5027	4992	1	103	106	3	5241	5186	1	5343	5292	1
25	Or	SULL	7473	7379	1	334	339	2	7807	7718	1	397	408	3	7497	7429	1	7893	7837	1
26	Or	CCOL	2083	2048	2	613	606	1	2696	2654	2	544	545	0	2531	2549	1	3075	3094	1
27	Or	NSTA	5977	5902	1	1507	1507	0	7484	7409	1	1482	1517	2	6671	6681	0	8153	8198	1
28	Or	HAYM	2825	2776	2	1715	1735	1	4540	4511	1	1798	1852	3	2701	2700	0	4499	4552	1
29	Or	STAT	6748	6638	2	5577	5511	1	12324	12149	1	6621	6717	1	8447	8661	2	15068	15378	2
30	Or	DTXG	4923	4832	2	7402	7462	1	12325	12294	0	4172	4228	1	6888	7033	2	11060	11261	2
31	Or	CHTN	1342	1317	2	2450	2454	0	3792	3771	1	2259	2305	2	1524	1544	1	3783	3849	2
32	Or	NEMC	2034	2005	1	3109	3103	0	5143	5108	1	2921	2964	1	2066	2103	2	4986	5067	2
33	Or	BBAY	2871	2850	1	10027	9969	1	12898	12819	1	8005	8073	1	3138	3224	3	11142	11297	1
34	Or	MASS	876	868	1	2858	2831	1	3734	3699	1	2671	2709	1	1039	1057	2	3710	3766	1
35	Or	RUGG	738	728	1	5752	5691	1	6490	6419	1	7260	7415	2	912	923	1	8172	8338	2
36	Or	ROXB	497	491	1	2436	2411	1	2932	2902	1	1691	1711	1	536	547	2	2227	2258	1
37	Or	JACK	268	266	1	4052	3993	1	4321	4259	1	2678	2704	1	289	290	0	2967	2994	1
38	Or	STON	159	162	2	1729	1706	1	1889	1868	1	1343	1357	1	153	159	4	1496	1516	1
39	Or	GREE	150	156	4	1618	1599	1	1768	1755	1	1654	1682	2	165	174	5	1818	1856	2
40	Or	FORE	0	0	0	8862	8758	1	8862	8758	1	8221	8321	1	0	0	0	8221	8321	1
41	Bl	WOND	4570	4646	2	0	0	0	4570	4646	2	0	0	0	4385	4744	8	4385	4744	8
42	Bl	REVE	1856	1863	0	74	73	2	1930	1936	0	34	34	0	1388	1493	7	1422	1527	7
43	Bl	BEAC	1869	1885	1	137	131	4	2005	2016	1	123	119	3	1802	1944	7	1925	2063	7
44	Bl	SUFF	1374	1444	5	117	113	4	1491	1557	4	120	120	0	1345	1480	9	1465	1600	8
45	Bl	ORNT	3056	3101	1	136	130	5	3192	3231	1	166	161	3	3102	3357	8	3268	3518	7
46	Bl	WOOD	1333	1352	1	99	95	4	1433	1447	1	85	83	2	1011	1080	6	1095	1163	6
47	Bl	AIRP	3463	3570	3	318	303	5	3782	3873	2	324	318	2	3128	3350	7	3451	3668	6
48	Bl	MAVE	6531	6596	1	448	419	7	6979	7015	1	512	491	4	6003	6481	7	6514	6972	7
49	Bl	AQUA	1436	1516	5	1642	1524	8	3079	3040	1	1892	1840	3	1705	1795	5	3597	3635	1
50	Bl	STAT	330	349	5	5450	5078	7	5780	5427	7	5656	5621	1	271	292	7	5928	5913	0
51	Bl	GOVT	59	70	15	1862	1737	7	1921	1807	6	3128	2715	15	24	28	13	3152	2743	15
52	Bl	BOWD	0	0	0	2633	2423	9	2633	2423	9	3046	2765	10	0	0	0	3046	2765	10
53	Gr	LECH	4072	3750	9	0	0	0	4072	3750	9	0	0	0	3835	3575	7	3835	3575	7
54	Gr	SCPK	693	635	9	108	115	6	801	750	7	101	114	12	828	776	7	929	890	4
55	Gr	NSTA	3552	3312	7	190	204	7	3742	3516	6	274	301	9	3007	2822	7	3281	3123	5
56	Gr	HAYM	2384	2217	8	494	536	8	2877	2753	5	258	287	10	1923	1815	6	2181	2102	4
57	Gr	GOVT	8463	8001	6	1994	2097	5	10457	10098	4	2030	2288	11	9403	9030	4	11433	11318	1
58	Gr	PARK	13532	12913	5	1439	1468	2	14971	14381	4	1447	1558	7	15215	14685	4	16662	16243	3
59	Gr	BOYL	2624	2465	6	1732	1770	2	4357	4235	3	1408	1522	8	3837	3725	3	5244	5247	0
60	Gr	ARLT	5779	5387	7	7408	7598	3	13186	12985	2	7188	7855	8	6143	5894	4	13331	13749	3
61	Gr	COPL	6441	6026	7	9308	9601	3	15749	15627	1	9010	9875	9	6231	5895	6	15240	15770	3
62	Gr	HYNE	2460	2288	7	4121	4266	3	6580	6554	0	3821	4118	7	3146	3023	4	6967	7141	2
63	Gr	KENM	1753	1619	8	5771	5926	3	7525	7545	0	4558	4920	7	2472	2394	3	7030	7314	4
64	Gr	SBCD	0	0	0	36871	38003	3	36871	38003	3	37289	40055	7	0	0	0	37289	40055	7
65	Gr	PRUD	724	677	7	1419	1459	3	2142	2136	0	1377	1478	7	795	762	4	2172	2240	3
66	Gr	SYMP	360	325	11	1721	1777	3	2080	2102	1	1127	1227	8	588	570	3	1716	1797	5
67	Gr	SURE	0	0	0	15360	15793	3	15360	15793	3	14594	15893	8	0	0	0	14594	15893	8
Avg Abs Error				85			94			120			159			118			207	
Avg % Diff				3			3			2			5			4			3	

values are larger. One can also see that the agreement between estimated and observed is better for entries than it is for exits, due to the greater weight given to entries.

A comparison of estimated vs. observed transfer flows is given in Table 6. The table includes a comparison for a sample hour (period 1, 7-8 a.m.) and for the daily total. The agreement between estimated and observed is worse than the on and off agreement, again due to the smaller weight accorded the transfer data. Some of the transfer flow estimates differ widely from their observed values, because improving them would have required worsening the agreement with on, off, or OD data.

Finally, a comparison against the original OD data can be made for the 8-9 a.m. period using the matrices in Table 7. Part (a) is the estimated OD flows, while part (b) is the OD data as expanded; the closeness of the fit can be gauged by comparing part (a) with part (b). It will be obvious that there are many overestimates and many underestimates, each resulting from attempting to reduce discrepancy in a summary measure such as ons, offs, or transfers. The degree of closeness that the model seeks to obtain for a given cell depends on the product of that cell's origin and destination expansion factors, given originally in Table 1. The greater the expansion factor product, the less importance accorded to an OD cell, because a large expansion factor implies a low response rate, and therefore less confidence in the observed value. Station pairs with large expansion factors, such as Jackson Sq. - Back Bay, are not matched as closely as those with small expansion factors, such as Harvard - Back Bay.

In summary, the seed fits the historical data rather well, but there are still some rather large discrepancies due to inconsistencies in the historical data. For this reason, reducing one discrepancy will make another worse, and the model, by optimizing, has found the point at which the total relative discrepancy is minimized.

8. MATHEMATICAL METHOD FOR UPDATING TO ENTRIES

The first stage in modeling passenger flows was the estimation of a historical seed matrix, which is really a set of period specific seed matrices incorporating historical counts. The second stage is updating that historical seed to current entries counts. This section describes mathematically the method used for that update procedure.

8.1 Splitting Entries By Direction And Shifting To Referenced Periods

First, as mentioned earlier, the entries counts at all stations except the downtown transfer stations are split by direction based on the directional split for the corresponding station and period in the historical seed. (The entries counts are recorded in absolute time, while the historical seed is in referenced time, so the period correspondence can be off by up to 20 min, which is close enough because the directional split does not vary rapidly between neighboring periods.) The resulting directional flows are then converted to referenced time by shifting flows according to the travel time to or from the line's reference station, assuming a uniform arrival distribution within the original (absolute) periods. For example, in the 8-9 a.m. period, according to the historical seed, 94.9% of the entries at Harvard are inbound. Therefore the 8-9 a.m. entries at Harvard are split accordingly. Inbound, Harvard is 11.5 min upstream of Downtown Crossing, the Red Line's reference station, and so we assume that 11.5/60th's of the 8-9 a.m. inbound entries will arrive at

Table 6. Comparison of estimated vs.observed transfer flows (Pd 1 and over all periods)

							Period 1			over all periods		
TFER	L_FR	D_FR	S_FR	L_TO	D_TO	S_TO	Est.	Obs.	Diff (%)	Est.	Obs.	Diff (%)
0	Re	0	PARK	Gr	0	PARK	831	575	44	10524	9066	16
1	Re	0	PARK	Gr	1	PARK	280	336	17	3015	3177	5
2	Re	1	PARK	Gr	0	PARK	1923	1523	26	11375	10720	6
3	Re	1	PARK	Gr	1	PARK	396	501	21	2173	2538	14
4	Gr	0	PARK	Re	0	PARK	114	147	23	2669	3093	14
5	Gr	0	PARK	Re	1	PARK	142	170	17	3040	3256	7
6	Gr	1	PARK	Re	0	PARK	600	511	17	10729	10572	1
7	Gr	1	PARK	Re	1	PARK	445	332	34	8454	7422	14
8	Gr	1	PARK	Or	0	DTXG	52	41	28	1060	983	8
9	Gr	1	PARK	Or	1	DTXG	73	76	4	1039	1012	3
10	Re	0	DTXG	Or	0	DTXG	166	154	7	2243	2098	7
11	Re	0	DTXG	Or	1	DTXG	139	128	9	2035	1988	2
12	Re	1	DTXG	Or	0	DTXG	572	584	2	2511	2610	4
13	Re	1	DTXG	Or	1	DTXG	1029	1022	1	5135	5447	6
14	Or	0	DTXG	Re	0	DTXG	643	711	10	3881	4169	7
15	Or	0	DTXG	Re	1	DTXG	331	342	3	2469	2437	1
16	Or	1	DTXG	Re	0	DTXG	247	275	10	3456	3721	7
17	Or	1	DTXG	Re	1	DTXG	546	573	5	3862	3776	2
18	Or	0	DTXG	Gr	0	PARK	133	124	7	797	861	7
19	Or	1	DTXG	Gr	0	PARK	43	48	10	434	396	10
20	Or	0	NSTA	Gr	0	NSTA	68	13	420	458	227	102
21	Or	0	NSTA	Gr	1	NSTA	7	7	1	74	57	31
22	Or	1	NSTA	Gr	1	NSTA	9	1	790	56	21	166
23	Gr	0	NSTA	Or	0	NSTA	3	1	220	41	16	154
24	Gr	0	NSTA	Or	1	NSTA	7	2	245	78	42	85
25	Gr	1	NSTA	Or	1	NSTA	10	7	39	173	247	30
26	Or	0	HAYM	Gr	0	HAYM	173	153	13	1174	1083	8
27	Or	0	HAYM	Gr	1	HAYM	7	9	23	74	116	36
28	Or	1	HAYM	Gr	0	HAYM	29	9	221	464	371	25
29	Or	1	HAYM	Gr	1	HAYM	49	73	32	310	402	23
30	Gr	0	HAYM	Or	0	HAYM	18	25	29	227	280	19
31	Gr	0	HAYM	Or	1	HAYM	7	15	54	78	134	42
32	Gr	1	HAYM	Or	0	HAYM	20	16	24	289	273	6
33	Gr	1	HAYM	Or	1	HAYM	71	33	115	1264	1145	10
34	Or	0	STAT	Bl	0	STAT	29	33	11	163	193	15
35	Or	0	STAT	Bl	1	STAT	139	129	8	1455	1355	7
36	Or	1	STAT	Bl	0	STAT	255	293	13	2892	3843	25
37	Or	1	STAT	Bl	1	STAT	182	164	11	969	734	32
38	Bl	0	STAT	Or	0	STAT	612	567	8	3516	3250	8
39	Bl	0	STAT	Or	1	STAT	230	208	11	1700	1606	6
40	Bl	1	STAT	Or	0	STAT	9	10	12	420	460	9
41	Bl	1	STAT	Or	1	STAT	4	4	10	130	150	14
42	Bl	0	GOVT	Gr	0	GOVT	1509	1049	44	8464	6908	23
43	Bl	0	GOVT	Gr	1	GOVT	101	118	14	628	676	7
44	Bl	1	GOVT	Gr	0	GOVT	4	3	17	245	241	2
45	Bl	1	GOVT	Gr	1	GOVT	1	2	30	89	117	24
46	Gr	0	GOVT	Bl	0	GOVT	7	9	26	34	48	30
47	Gr	0	GOVT	Bl	1	GOVT	16	18	11	472	479	2
48	Gr	1	GOVT	Bl	0	GOVT	90	79	14	421	497	15
49	Gr	1	GOVT	Bl	1	GOVT	480	340	41	9245	7676	20

	pd 1	total
Average absolute error	45	239
Average estimated transfers	257	2330
Average observed transfers	231	2240
Average % difference	19	11

Table 7a. Estimated OD Flows (Pd 2, 8:00 -- 9:00)

	ALWF	DAVS	PORT	HARV	CENT	KEND	CHAS	PARK	DTXG	SSTA	BWA	ANDR	JFKU	SAVN	FLDS	SHA	ASH	NQC	WOLL	QCTR	QAD	BRNT	OAK	MALD	WELL	SULL	CCOL	NSTA	HAY	STAT	DTXG	CHTN	NEM	BBAY	MASS
ALWF		5	2	153	60	180	80	230	338	257	10	2	19				3	20		7		1	1	2	2		2	6		30		5	7	13	2
DAVS	44		19	302	102	165	71	196	178	130	8	3	34					15		9	2	1						2	3	18		2	12	14	
PORT	22	7		197	60	186	48	139	189	135	18	10	15				2	14			5	2			3		3			10			11	12	
HARV	63	14	21		35	186	140	306	382	247	7		34		6		3	14	3	11	2	2		3	7		8	15	2	45			3	15	
CENT	30	30	13	88		94	96	220	174	231	12	10	28		11		10	7	3	5	4	3				6		3	2	29			9	10	
KEND	6	7	9	35	6		5	30	58	38	4		9					3			4	1					2			2		2	3	2	
CHAS	15	14	12	150	39	102		7	40	68	14		5				4	5			2	2	1		2					10			3		
PARK	15	3	5	39	12	31	11			21	8		9				2			2	4	3													
DTXG	2	3		19		13	9				2	8	21		8	5	4	10	9	6	3	3													
SSTA	9	9	9	182	91	354	130	202	50		17	7	53	5	12		8	44		12	4	2		7		10			3	61			4	7	11
BWAY	4			17	6	21	10	46	17	36			13		8		13	13		34	3	5				3		3	3	14			2	7	2
ANDR		7		11	6	17	27	66	139	143	14		9		10		12	13		13	4					13	5	3	3	39			7	6	4
JFKU	1	3	5	14	4	12	9	40	55	92	10			6			4	15		55	10				5		3	11	4	14		6		18	3
SAVN				2	7	4	5	28	63	36	3		11		8		3	2		1				2			3	2		18		3	5	10	
FLDS	3	2		13	3	13	7	104	81	113	15	4	37	12		0	14			2	1			2	2			6	2	57		4		9	
SHAW			1	10	4	6	5	22	47	61		4	12				5								2			4		14				4	
ASHM	7	4	4	51	23	44	45	175	276	352	45	19	130		41	1		7	1	1	1	1			4	7	14	14	19	47		3	13	21	2
NQCY		3	1	19	4	22	40	154	241	280	12	15	48				1		7	21		2		3	2	3		10	8	63		9	7	20	
WOLL	3	1		19	4	23	31	89	163	219	8	7	24					59			18	3		4	1	4	3	22	2	33		2	7	23	2
QCTR	3	3	1	27	13	40	26	135	169	193	26	15	42					134	4		5	1		5	3	11	5	24	18	35		9	15	33	4
QADM	1	2	2	34	4	39	30	171	287	356	2		18		1			40	3	7		5				6	4	15	9	105		11	9	28	
BRNT	7			17	14	29	17	82	136	202	6	3	7					13		3	2			6		5	3	20	6	28		7	3	17	1
OAKG		3	1	11	9	11	4	3		22	3	2	7					9						19		25	21	65	58	394	174	52	58	105	5
MALD	1	4	2	34	33	27	4	3		52	11	9	36			7	9	12		7		1	7		11	46	52	134	150	656	256	91	99	210	27
WELL	1		2	6	9	18	2			33	7		8					10		2		1	3	5			12	81	105	489	198	43	55	161	7
SULL		2		12		22				42			6				7	6		3	2		1	20	2		61	154	235	818	250	66	60	230	13
CCOL	1		5	11	3	1		40			3	4	7	2	5		14	2	3	8	2	3	11	16	7	23		4	3	8	3	4		9	8
NSTA	2			13	8	33				105	6		12					12			2		1	9	11	28	17		4	567	212	59	63	198	31
HAYM	2	3		7	11	7					5		10					6		3	2		2		11	10	11			24	42	12	25	72	18
STAT										51											4			6	5	11		5					6	21	
DTXG																									4	9	9	5	6	83				45	
CHTN				6																						3		3		39	8			6	
NEMC					5	12	9			13			3					3				2		4	2	7		5	9	37	9			17	4
BBAY	5	2		44	10	20	3			15			18				10	11			3		1	40	22	47	13	158	97	947	436	104	172		42
MASS	4	1	2	18	3	15	6	37		21	4		6	3		9		4		3			2	5	5	7	3	39	21	197	167	36	57	74	
RUGG				6		18	8			14					7			13		4		2	1	4	4	6	4	55	13	397	204	22	26	197	50
ROXB	2			4			10			15	2		7		5											2	3	11	2	97	73		28	54	
JACK	7			14	8	28	11			24			8					8						2	2	3	4	36	19	287	188	31	31	159	22
STON		2		25	27	6	3	7		14			5							6			1			5	3	16	19	110	109	26	41	70	17
GREE		1	1	14	10	6	5	16		8		3	3											1		4		13	18	87	85	25	21	45	20
FORE	5	5	9	38	33	20	18	5		18					5		2	10	4	6	2		3	9	7	25	11	79	55	669	614	117	176	271	75
WOND	1			20	4	11				25			3					2		4						4		8			16	30	32	27	6
REVE				5		6	2			3																		10			24	44		13	7
BEAC			2	2						15					10		2	2				2		3	2	6		17			8	8	18	17	
SUFF				2		2				3																					5	17	15	12	
ORNT				9	13	15	2			13		9					2	2			2			6		11	10	22			19	6	32	24	5
WOOD				6		5				5														5				4			4			6	5
AIRP				3		9				7								7						4			56	14			13			12	
MAVE	3	2		25		13				20	5	9	7							3					3		18	17	5		48	32	31	22	23
AQUA						3																		4										24	
STAT																																			
GOVT																																		2	
BOWD										6														2							8		8		
LECH							1			8								2					1				12			16			5		
SCPK																								2											
NSTA	1			14	2	14				8								6			1														
HAYM	1		1	19	4	14							2																						
GOVT	2					6						4			4					6											3	7		0	
PARK																																			
BOYL					7	5			2						2											2									
ARLT	4	4		6	12	24	3			26	2		3							3		3													
COPL	2		4	15	11	20	2		6	25			2					5	2			1													
HYNE	2	2	2	10	15	20	11		11	50	2	4						3		3	2												3		
KENM	5			2	3	11			3	28		6	5					2					1			7									
SBCD	14	8	6	99	44	161	113		6	256	15	7	68	2	4	5	8	39	7	10	10	2	2	13	12	20	30					4	5		
PRUD	2			12		4				12																									
SYMP				11		11				13			5																						
SURE	5	9	3	46	43	106	13		22	241	3		39				3	4							7	9	79					12			
Offs Tot	304	164	143	1968	821	2282	1070	2553	3135	4417	308	163	847	30	147	27	143	606	46	259	104	54	38	213	148	388	486	1116	905	6593	3175	908	1184	2371	416

Table 7a. Estimated OD Flows (Pd 2, 8:00 -- 9:00, continued)

	RUG	ROXB	JACK	STON	GREE	FORE	WON	REVE	BEAC	SUFF	ORNT	WOO	AIRP	MAVE	AQUA	STAT	GOVT	BOW	LECH	SCPK	NSTA	HAY	GOVT	PARK	BOYL	ARLT	COPL	HYNE	KEN	SBCD	PRUD	SYMP	SURE	Ons Tol
ALWF	6	4	2		1	4							4	3	10			3	1		4	4	46		19	110	52	12	4	35	9	3	24	1794
DAVS	10	3	2	2	1	1		2			5		5		3						2	3	26		7	112	55	8	14	57	7		71	1726
PORT	3			1	1	2							2	11	57						2	4	21		4	59	38	8	5	46	5	2	20	1377
HARV	8	9		4	1	1							3		7				2	1	8	11	68		10	82	51		12	40	2	4	23	1918
CENT	4	6			1	2									7					1	6	5	64		6	62	42		2	15	3		21	1373
KEND			2		1								2	6								1	6			9	13		3	7			10	285
CHAS															11								4				7	3	6	12			13	549
PARK																																		164
DTXG																							9				25		5	4				168
SSTA	3									2			7					17	5		1		56			81	54	8	11	62			45	1656
BWAY	1	2			1	1		1							2				3		1		14		1	6	4	4	3	11			5	339
ANDR	3				1						9				9			9	1		1	1	15			7	22	5	3	15	5		12	690
JFKU	2				1								3						1				12		1	15	11		2	8	6		5	464
SAVN													3					7	1				5			4		3		5				241
FLDS						2							4									1	5		5	25	26	4	5	10	4			593
SHAW		3																	1		1		8				18			11			9	251
ASHM	2					2							3	10	7			13	1		5	3	68		13	51	42	8	16	66			16	1697
NQCY	10					2												3	1		1	1	34		11	71	28	17	13	32	6		25	1250
WOLL	2												2		3						1	1	23		9	43	24	9	5	26	5		32	958
QCTR	5						2						5	3	3			15	1	1	1	1	33		11	73	42	18	14	48			30	1304
QADM	2					2		2					3	6	7			22	3		3	2	33		21	99	69	15	16	19			31	1541
BRNT	2				1	1			1				1	3	2			1	1	1	1	2	15		7	43	24	9	5	21	7		10	787
OAKG	24	5												2	4			1	4	2			2		1	9	3	4	5	25			13	1168
MALD	43	10	6		3	4		3			5		8	9	18			10	8	6			2			22	15	6	9	54			21	2248
WELL	25	3	9			3					4				6			3	8				5			18	10	3	12	32	3		9	1406
SULL	35	4				6							8	25	7			5	3							45	20	10	9	42			9	2238
CCOL	4	3	5	3		8					4	17	4	4			35		2				18							6			27	345
NSTA	45	3				15							13	4	14			4																1489
HAYM		3	5		2	6												7																305
STAT						6													6															120
DTXG	5	8	3			5																												183
CHTN						1																											2	67
NEMC	21		7		1	2					3			2				7	3															186
BBAY	67	13	10		1	4		2			8	5	28	3	27			71	10	5										1				2474
MASS	10	14	11		2	18					3		2	3	2			10	5							2		1	2	9			2	846
RUGG			7	3	1	22									12			83	11	7										2				1202
ROXB				2	1	14								8	4		38	9												7				401
JACK	16	10		2	1	9			2				3	9	5			11	3								2			4				966
STON	25	10	12			6		2						6				4									2			4				582
GREE	27	30				9					2		1	5	2			5	2											9				475
FORE	81	46	20	1	12			1					9	14	14		12	58	4	4									2	14				2585
WOND	15					3		1		1	4		10	7	157	700	138	106	8		9			25	15	103	60	18	7	29	13		30	1650
REVE					2		5		26				15		47	168	49	43	4		6			10	4	18	14	3	2	9			8	544
BEAC	7	9					3	2		0	2			4	53	302	68	41	3		4			16	4	27	23	3	2	13			3	703
SUFF							2		3			6			22	112	23	11							1	7	7			6				256
ORNT	10		6		1	2	3	2					11	17	78	425	107	80	5	10	6			23	11	39	41	14	7	30			19	1148
WOOD	6										5				48	128	47	36	3					13	4	9	9		2	10	3		2	363
AIRP		16							9						18	67	13	32				10		39	3	15	11		3	13			3	378
MAVE	13					3	1			1	1		3		54	482	113	104	28		14	6		103	8	51	75	11	12	21			18	1406
AQUA													30	13		8		6		3					3	5	5			2			3	110
STAT								4		1	8		5	9	18			18						2							5			69
GOVT											3		-					5																10
BOWD										1			6	4		4	1				1			1		11								52
LECH	4	11	6		1			1			2		2	4	4	2		4		19	35	29	198	56	10	37	36		6	18	13	9	34	585
SCPK																			5				12	8		2	2						9	39
NSTA											2		2		2			1	19			17	116	70	29	156	65	18	24	47	2		78	696
HAYM																			39		11		12	27	7	69	37	18	12	47	15	5	46	388
GOVT																			27	10					5	50	45	17		61			18	263
PARK														6	21				7		3	2				60	108	32	11	79	9		48	385
BOYL													2	2					1		2	1	10	7			34		10	23			15	126
ARLT								2					3					3	2		3	1	25	54			29	34	29	106	13	13	96	499
COPL											2		2		6			3	6	3	5	6	73	94	13	6		12	46	197		4	46	619
HYNE															11	4			2		5	1	73	180	32	33	14		7	99				597
KENM		3			1										10	7		2	3		3	11	43	127	31	43	66	19		105		22		569
SBCD	5	11	6		1	5	4	2		1	3		24	6	49	8		9	11	7	43	18	725	1832	432	549	685	401	238	126		13	97	6277
PRUD							3								15				2		1	1	28	96	5	11		1		38			50	278
SYMP																			1	2			29	103	32	36	8	2					20	273
SURE													14			23			7	4	7	12	227	395	157	134	79		0	109	80	48		1937
Offs Tol	551	237	118	18	38	170	21	25	42	7	75	27	248	207	856	2439	643	879	268	84	193	154	2163	3281	931	2518	2149	754	598	1944	212	123	1124	60627

Table 7b. Expanded OD Flows (Pd 2, 8:00 -- 9:00)

	ALWF	DAVS	PORT	HARV	CENT	KEND	CHAS	PARK	DTXG	SSTA	WAY	ANDR	JFKU	SAVN	FLDS	HAW	SHM	QCY	WOLL	QCTR	ADM	BRNT	AKG	MALD	WELL	SULL	CCOL	NSTA	AYM	STAT	DTXG	CHTN	EMC	BBAY	MASS
ALWF		7	1	100	65	161	88	212	331	228	14	3	22				5	23		10		2	2	3	3		2	6		21		5	8	13	2
DAVS	24		11	150	84	127	66	147	138	99	10	3	32					15		10	2	2						1	3	10		2	12	12	
PORT	16	4		99	50	144	45	105	147	103	22	9	14				3	13			5	3			3		2			5			10	10	
HARV	42	7	12		42	166	169	294	407	220	14		45		10		8	19	4	18	3	5		4	12		11	16	2	31			5	16	
CENT	32	26	13	47		73	110	184	165	178	24	18	36		19		26	9	4	8	7	7				10		2	3	14			12	10	
KEND	7	7	11	30	5		6	31	66	36	7		12					4			7	2					3			2		2	4	3	
CHAS	27	22	23	184	47	99		6	40	51	31		7				13	8			4	4	4		4					3			4		
PARK	44	6	15	64	20	31	13			32	51		34				16			11	21	20													
DTXG	7	7		38		17	14				11	39	55		42	22	33	28	29	25	11	21													
SSTA	14	12	14	186	85	303	114	110	25		34	13	69	12	22		23	59		21	8	4		9		11			3	3			3	5	7
BWAY	8			33	13	32	17	67	28	41			11		6		26	12		38	3	10				10		5	6	16			3	10	3
ANDR		12		16	9	19	32	55	120	81	19		8		9		29	12		16	5					28	6	3	4	11			8	6	4
JFKU	3	6	14	26	9	17	15	52	78	84	19			8			7	8		42	10				13		6	17	8	11		9		22	4
SAVN				2	10	5	6	27	62	28	3		5		6		5	6		4				3			4	2		10		3	6	10	
FLDS	9	4		30	7	21	13	145	128	97	35	12	18							16	7			5	6			10	6	40		7		13	
SHAW			3	15	7	7	7	23	52	48		6	6												3			5		9				4	
ASHM	13	6	8	77	37	53	60	186	311	292	68	30	77		58	7		27	4	4	3	7			9	16	23	17	31	33		3	18	22	3
NQCY		4	2	23	5	22	44	132	209	199	14	15	21				4			9		3		4	3	4		10	10	33		8	7	18	
WOLL	5	2		26	5	26	38	88	169	182	10	8	15					23			5	3		6	2	8	3	25	2	23		3	9	23	2
QCTR	5	5	2	34	17	43	29	128	164	155	32	17	24					49	6		17	2		8	4	19	6	25	24	24		10	17	32	5
QADM	1	3	3	33	4	35	27	123	198	234	2		7		4			7	3							6	3	12	9	47		8	7	22	
BRNT	10			22	20	34	22	96	165	209	9	5	7					11		3	4			9		9	4	24	9	29		9	4	19	1
OAKG		4	1	10	9	10	4	2		20	4	3	7					10						2		16	20	68	50	371	181	55	63	101	6
MALD	1	4	2	27	22	21	3			35	14	8	32			5	14	12		8		2			10	3	30	120	88	475	229	79	93	168	24
WELL	2		2	6	8	16	2			26	11		9					12		3		3	5	4			11	88	82	446	216	49	65	151	8
SULL		2		10		17				26			6				17	6		4	4		4	29	3		39	154	130	613	252	68	67	189	14
CCOL	1		5	9	2	1					3	3	6	3	3		19	1	2	8	2	3	21	20	9	32		4	2	6	3	3		7	7
NSTA	2			13	7	29				76	11		15					16			4		3	8	13	22	5		3	518	243	73	80	187	38
HAYM	4	4		9	13	7					13		16					11		7	7		6		18	10				26	62	20	45	77	31
STAT																					10			15	18	44		7					8	12	
DTXG																									31	77	45	12	30					61	
CHTN				7																						15		7		24	7			24	
NEMC					8	13	11			9			7					7				8		13	8	38		12	30	41	12			20	8
BBAY	6	2		34	6	15	2			9			16				20	11			3		3	60	41	88	17	168	138	515	298	107	133		37
MASS	4	1	2	14	2	11	4			14	4		5	3		4		4		2			4	7	8	10	3	37	25	117	117	33	42	50	
RUGG				6		12	5			4					16			21		9		8	8	13	15	25	11	85	37	184	156	41	28	98	39
ROXB	5			5			10			8	8		13		15											12	10	23	7	81	80		40	42	
JACK	13			15	7	22	9			11			12					12						5	6	11	9	62	51	182	167	57	37	102	20
STON		2		18	15	4	2			8			4							6			2			8	3	15	23	59	71	24	29	44	10
GREE		1	1	12	7	5	4			6		3	3											1		5		12	20	62	67	23	17	35	15
FORE	6	6	12	34	25	16	14	2		12					6		5	11	4	7	3		9	15	15	54	17	94	87	453	495	141	161	195	59
WOND	1			20	4	10				19			3					2		4						3		5			11	17	22	19	4
REVE				6		6	3			2																		6			15	21		7	4
BEAC			2	2						10					12		5	3				3		3	3	5		9			5	3	11	10	
SUFF				3		2				2																					3	8	11	7	
ORNT				9	12	14	2			9		9					4	2			2			6		7	3	12			11	2	19	14	2
WOOD				8		6				4														6				3			3			4	3
AIRP				5		10				4								12						5			9	5			5			4	
MAVE	7	4		37		15				12	12	21	10							7					5		8	9	5		26	12	20	11	11
AQUA						4																		6										9	
STAT																																			
GOVT																																			
BOWD										3														4							4		4		
LECH							2			8								3					3				4			5			3		
SCPK																								5											
NSTA	2			19	3	16				7								9			3														
HAYM	2		2	25	6	15							3																						
GOVT	4					7						13			14					16															
PARK																																			
BOYL					15	8			4						10											8									
ARLT	8	8		9	20	27	4			21	7		6							8		13													
COPL	5		10	28	21	29	3		11	30			4					11	6			4													
HYNE	3	3	3	11	16	18	10		10	36	4	7						3		5	4												3		
KENM	10			4	5	13			4	25		19	9					5					5			16									
SBCD	21	10	10	106	45	150	108		5	193	26	12	82	5	6	7	20	51	9	15	17	3	6	15	18	25	22					3	6		
PRUD	3			11		3				5																									
SYMP				9		7				4			6																						
SURE	10	14	6	48	37	77	10		13	77	8		51				12	7							14	12	10					8			
Offs Tot	389	205	189	1828	856	2073	1145	2215	3052	3331	554	276	851	31	258	45	313	572	71	344	181	141	86	293	299	666	350	1207	931	4554	2737	918	1142	1918	371

Table 7b. Expanded OD Flows (Pd 2, 8:00 -- 9:00, continued)

	UGG	ROXB	JACK	STON	GREE	FORE	OND	REVE	BEAC	SUFF	ORNT	OOD	AIRP	MAVE	AQUA	STAT	GOVT	OWD	LECH	SCPK	NSTA	AYM	GOVT	PARK	BOYL	ARLT	COPL	HYNE	ENM	SBCD	PRUD	SYMP	SURE	Ons tol
ALWF	9	5	4		1	6							3	3	7			2	2		5	5	40		29	122	54	13	5	57	10	4	38	331
DAVS	13	3	4	4	1	1		3			2		3		1						3	4	19		9	94	45	7	16	74	7		84	335
PORT	4			2	1	3							2	6							3	6	16		5	50	32	7	6	60	5	2	24	189
HARV	17	17		10	2	2							3		4				3	3	14	19	60		18	106	60		20	83	3	6	46	342
CENT	8	10			3	3									3					4	12	10	48		11	75	45		3	32	3		43	212
KEND			5		2								2	9								2	6			12	17		4	14			20	67
CHAS															4								3				8	3	12	31			31	85
PARK																																		
DTXG																													10	13				23
SSTA	5									11			4					3	10		2		21			20	13	3	11	71			37	155
BWAY	4	5			3	3		5							3				9		2		17		2	11	6	5	6	28			14	72
ANDR	6				4						6				4			3	3		3	2	9			5	12	4	4	27	5		20	77
JFKU	6				4								4						3				12		3	20	13		4	22	9		13	84
SAVN													3					5	2				4			3		3		8				14
FLDS						6							7									3	4		14	37	31	6	13	33	7			141
SHAW		6																	2		2		7				16			21			16	53
ASHM	4					3							3	20	6			9	2		10	6	59		23	54	39	8	25	128			29	307
NQCY	17					2												2	2		2	1	24		17	53	20	13	15	46	5		31	200
WOLL	3												2		2						2	1	19		14	41	21	8	7	43	5		48	188
QCTR	9						3						5	5	2			10	2	3	1	1	27		16	64	35	16	18	72			41	263
QADM	3					2		3					2	4	3				4		4	2	21		25	53	38	10	15	19			26	185
BRNT	4				1	1			3				1	6	2			1	1	2	1	3	16		10	53	27	10	7	34	8		16	165
OAKG	35	7												3	4			1	4	2			2		1	8	3	4	5	31			15	66
MALD	60	11	11		5	6		7			3		6	9	9			6	6	3			1			12	8	4	9	59			19	111
WELL	47	4	25			5					4				4			2	7				3			14	7	2	15	47	3		11	100
SULL	70	6				10							7	36	3			3	2							24	10	6	10	61			11	122
CCOL	4	3	8	4		9					2	8	3	3					1											6			22	28
NSTA	99	6				28							14	7	9			3																
HAYM		10	30		10	17												5																
STAT						17													6															
DTXG	42	52	39			29																												
CHTN						9																											14	14
NEMC	95		58		6	7					11			16				6	5															
BBAY	109	15	22		2	6		5			9	7	33	6	21			36	10	4										4				4
MASS	11	10	15		3	20					3		2	4	2			6	4							2		2	4	18			3	29
RUGG			19	10	6	35									13			23	16	9										10				10
ROXB				9	6	27								45	6		6	5												46				46
JACK	36	11		9	5	24			14				7	41	6			5	5								6			25				31
STON	23	4	6			11		4						10				2									2			9				12
GREE	25	19				3					2		1	6	1			3	2											16				16
FORE	105	35	22	4	8			5					12	31	13		5	36	4	4									6	42				48
WOND	14					3		3			5		7	7	166	518	124	105	6		5			13	20	92	51	16	9	39	13		37	277
REVE					6								12		70	113	54	54	3		2			4	8	22	15	3	4	19			17	89
BEAC	8	5					4	5			4			6	66	195	65	45	2		2			6	7	27	20	3	3	23			5	87
SUFF												9			34	78	25	14							3	9	7			15				34
ORNT	9		6		2	2	3	4					7	19	83	275	91	77	3	3	3			9	16	33	32	12	9	43			24	169
WOOD	9														71	96	53	45	2					6	7	11	10		3	21	4		5	62
AIRP		10													33	31	15	46				3			9	21	12		6	39			9	97
MAVE	22					5	7			34	8		17		96	290	140	152	19		3	3		21	20	75	89	14	27	62			50	337
AQUA													30	30		5		11							11	8	7			9			11	47
STAT								27		25	12		8	33	7			6																
GOVT											14																							
BOWD										14			18	27		6										10								10
LECH	4	4	7		2			5			4		3	11	5	2		6		10	22	17	111	36	20	46	41		11	38	17	13	70	256
SCPK																							16	18		7	6						57	70
NSTA											4		3		2			2	21				61	41	51	176	67	19	37	89	3		141	582
HAYM																			23				6	13	12	76	37	19	20	92	18	7	85	366
GOVT																			18						14	53	38	17		185			51	358
PARK														17					34		11	9				38	48	22	25	244	9		128	514
BOYL													5	10					3		9	2	19	14			14		14	42			22	92
ARLT								11					6					5	9		12	3	41	78			18	27	56	276	13	15	218	623
COPL											6		4		11			6	22	14	18	16	131	171	21	14		24	133	722		10	172	1096
HYNE															10	2			5		12	2	87	176	30	43	22		15	266				375
KENM		8			4										13	6		4	10		9	33	69	185	39	84	146	42		56		12		380
SBCD	8	15	15		3	9	8	5		9	4		31	13	43	6		9	29	25	97	37	821	1734	399	658	970	585	339	240		4	60	3254
PRUD							6								8				7		3	2	29	64	3	11				18			73	106
SYMP																			4	15			29	54	18	35	12						29	94
SURE													22			4			34	25	27	40	268	279	108	169	143			54	74	70		618
Offs Tol	948	281	294	52	92	314	30	92	17	92	103	24	300	441	849	1627	577	764	371	123	300	233	2123	2923	1013	2651	2374	934	964	3880	219	145	1936	14117

Downtown Crossing after 9 a.m. Therefore 11.5/60th's of the 8-9 a.m. inbound entries are assigned to the referenced 9-10 a.m. period, and the remainder to the referenced 8-9 a.m. period. Outbound, Harvard is 10.8 min downstream of Downtown Crossing, and so 10.8/60th's of the 8-9 a.m. inbound entries are assigned to the referenced 7-8 a.m. period, and the remainder to the referenced 8-9 a.m. period.

At the downtown transfer stations, which are all within 4 min of their line's reference stations, the model makes no directional split and no shift between periods.

An exception to the assumption of uniform arrivals during a period was made for the periods bordering the boundary periods of the study, i.e. boardings in the 6-7 a.m. period and the 9-10 p.m. "shoulder" periods. Only the arrivals during the part of the shoulder periods close to the boundary periods enter into the estimates, and the arrival rate during that part of the shoulder periods is certainly greater than the average rate in the hour-long shoulder periods. The 1989 CTPS counts included data beginning at 6 a.m. for a few Red Line stations. At stations 10 to 20 min distant from the reference station, about 33% of the 6 - 7 a.m. entries occurred in the 6:45-7:00 a.m. period, implying a rate 1.33 times the average hourly rate. At Braintree (about 28 min distant from the reference station) about 66% of the 6-7 a.m. entries occurred in the final half hour, again implying a rate 1.33 times greater than the average hourly rate.. Therefore, the simple assumption was made that entries in the relevant portion of the 6-7 a.m. period occurred at 1.33 times the average entry rate for the hour. In the evening shoulder period, there was no 9-10 p.m. data in the 1989 CTPS counts. A comparison of 9 - 9:15 counts from 1989 with 9-10 counts from 1995 showed no significant difference in rate, so for the evening shoulder period the hourly rate was assumed to hold constant over the hour.

8.2 First Target: Match Station / Direction / Period Entries Data

Although the entries counts are made different days, they are not mathematically inconsistent because they consist of counts of entries only, not exits. Therefore it is possible to fit the seed exactly to the shifted entries flows. This became then the first target of the updating procedure: match the shifted entries flows for every station, direction (except in the case of downtown transfer stations, for which direction is ignored), and period. This is accomplished by creating for each station / direction / period a balancing factor. For the downtown transfer stations, there is a balancing factor for each station / period. For the few stations without entries data (as mentioned earlier, in 1995 they included Surface E, Surface BCD, and Science Park), there were no such balancing factors.

8.3 Second Target: Match Overall Change In Entries

Because there were a few stations without entries data, it was necessary that there be a single scale factor reflecting an overall increase or decrease in ridership between the seed and the counts. (This would be especially critical if the seed were expressed as a contingency table rather than as passenger counts.) For example, the 1995 entries at stations with entries data represented a drop in ridership of about 8 percent compared to the historical seed. The scale factor is therefore 0.92, which is applied to the entire historical seed. This scaling has no effect on the stations with entries data, since they have their own balancing factors which will compensate for any scaling, but it does affect the stations without entries data in a manner that makes their estimates responsive, at least in a gross manner, to current ridership counts. This then became the second target of the updating procedure: to see that stations without entries data were scaled consistently with the overall trend.

8.4 Third Target: Maintain Production / Attraction Balance

Finally, because the current counts do not include exit counts, there is no guarantee that the degree of production / attraction (P/A) balance present in the historical seed will be maintained without an explicit attempt to do so. For example, if inbound Harvard entries increased in the current data, the seed would be factored to show that increase; but it would not necessarily show any increase over the day in outbound Harvard exits. Therefore, a set of factors was created to preserve the relative P/A balance between entries at a station in one direction and exits in the opposite direction.

As mentioned earlier, there is a considerable degree of P/A imbalance in the historical seed because of imbalance in the historical data which is due, at least in part, to factors that can be expected to continue in the future: people making their return trips after 9 p.m., people preferring one station for transferring from bus in the morning and a different station for transferring to bus in the evening, and so on. Therefore the third target of the updating procedure was not to force exact P/A balance, but rather to maintain the degree of relative P/A imbalance found in the historical seed. For example, if in the historical seed inbound productions at a station are 1.1 times as great as outbound attractions, that ratio will be preserved in the updated estimate. At the downtown transfer stations, the P/A balance factors are not line or direction specific, but simply maintain the P/A balance between total station entries and exits.

At stations with entries data, the daily productions targets are given, and attractions targets are determined simply by dividing by the P/A target ratio. At stations without an explicit productions target, the attractions target is the estimated productions divided by the desired P/A ratio. For each station, then, there is an attraction factor for each direction (except at downtown transfer stations, which have one attraction factor per station). The production and attraction factors are found using the iterative proportional fit method, which is embedded into a larger algorithm that performs the overall scaling mentioned in the previous subsection and that adjust the attractions targets as the productions estimates at stations without entries data change.

8.5 Algorithm

The algorithm used to find the factors and update the historical seed is a generalization of Iterative Proportional Fit. The sequence is as follows:

1. *Initialize* . Set the overall scale factor to 1. Set all estimated OD flows equal to the historical seed. Calculate and store the overall grand total of the estimated OD flows.

2. *Fit to station entries.* For each station / period with entries data, compare the estimated (i.e., from the OD flows) entries by station / direction / period (by station / period for the downtown transfer stations) with the entries data, and scale the contributing OD flows so that their sum matches the entries data.

3. *Establish attractions targets by station and direction.* Aggregate the OD flows to find daily entries by station / direction (by station only for the downtown transfer stations). Divide by the desired P/A factor for that station / direction, yielding the tentative attraction target for that station and the opposite direction. Calculate an adjustment factor equal to the sum of the total daily productions divided by the sum of the tentative attraction targets. Multiply all the tentative attraction targets by this adjustment factor. (Note: for stations with entries data in every period, the tentative attraction targets will not change. However, they will change for stations missing entries data.)

4. *Test for convergence (not performed on the first pass).* If the ratio of grand totals calculated in Step 6 is sufficiently close to 1, and if the attraction targets in Step 3 did not change, the estimate has converged; STOP. Otherwise, continue.

5. *Fit to attractions.* For each station, find the daily attractions total by direction. Compare it to the attractions target, and iteratively scale the contributing OD flows so that the attractions target is matched.

6. *Update the overall scale factor and repeat.* Find the grand total of the estimated OD flows. Multiply all the estimates by the ratio of this grand total to the previous grand total. Then store the new grand total in place of the previous grand total and go to Step 2.

In applying this procedure with 1995 entries counts, approximately 17 iterations were needed. The entire procedure, including loading the network and counts data, required about 4 minutes on a Macintosh PowerPC Performa 6300.

9. ANALYSIS OF UPDATED OD MATRICES

The updated OD matrices have been preserved in the file `fittedOD.xls`. Because the update to station entries produces a perfect fit, and because P/A balance is prescribed, the usual OD matrix summaries applied to the updated OD matrices reveal little. There was an overall scaling downward, with a factor of 0.922. The entries flows all match the entries data, and the productions and attractions for every station / direction pair (every station for the downtown transfer stations) exhibit the same relative imbalance present in the historical seed. A closer look at the updated matrices may be made by looking at the Microsoft Excel file fittedOD.xls (matrices for all periods, with ons and offs by station / period / direction and transfer flows by period).

It may be interesting to note that, for stations without entries data, the updating procedure does not merely multiply their entries by the overall scale factor. They are also affected the production / attraction factors of other stations (as well as their own). For example, Surface E is a popular destination for passengers boarding at Park. If the entries at Park change compared to the historical seed, there will be a change in exits at Surface E. Then, to maintain P/A balance at Surface E, there will be a corresponding change in entries at Surface E. In this way, changes are made to entries at those stations echoing both the overall system change (via the overall scale factor) and changes at particular stations with which they have a lot of interaction. In the updated matrices, the ratios of updated entries to historical seed entries at Surface E, Surface BCD, and Science Park were 0.906, 0.867, and 0.909, respectively.

10. VOLUME PROFILES AND PASSENGER-MILES ESTIMATES

Once a final set of OD matrices is obtained, the results are summarized in volume profiles. Passenger miles estimates are made in two different ways, one from the volume profiles, the other from the OD matrices. Both methods offer different advantages. Fortunately, the estimates do not materially differ.

10.1 Volume Profiles

The program PROFILEMAKER constructs hourly volume profiles for each line from matrices of hourly OD flows. The primary intended use of PROFILEMAKER is to make volume profiles from the OD flows after they have been fitted to match current entries data. However, the program can be used in two other ways as well: (1) it will construct volume profiles from the historical seed, and (2) if the fitted matrices are stored, they can later be used as inputs to PROFILEMAKER directly as a seed for the following year's entries counts.

PROFILEMAKER constructs volume profiles for each line / direction / period, where the periods are the hour-long referenced periods used throughout this study (e.g., the 8-9 a.m. period is the period in which trains on the subject line / direction pass that line's reference station). Each volume profile consists of a list of the stations on the line, showing for each station the departing volume, the entries and exits (from / to the street), and the transfer flows. Each platform to platform transfer flow is identified separately. The stations are listed from south to north for both directions. Branches are correctly dealt with.

The volume profiles are written to files, one for each line / direction (containing volume profiles for 14 periods between 7 a.m. and 9 p.m.), in a format readily importable to a spreadsheet such as LOTUS 1-2-3, with which tables, summaries, and graphical representations in any format can be made. An example of the output is given in Table 8.

10.2 Passenger-Miles Estimates From Volume Profiles

Attached to the volume profiles are estimates of passenger-miles made by multiplying segment volumes by segment lengths. These estimates of passenger-miles should be interpreted as passenger-miles occurring on each line / direction on trains passing through downtown in the each period. Total passenger miles over the 14 hourly periods studied was estimated by line as follows:

Red	NB	392,545
Red	SB	416,525
Orange	NB	166,974
Orange	SB	162,486
Blue	NB	46,760
Blue	SB	50,834
Green	NB	122,483
Green	SB	122,682

Table 8
Sample Volume Profile: Orange Line Northbound, 7-8 a.m.

STATION	OFF	ON	TRANSFERS FROM THE ORANGE LINE						TRANSFERS TO THE ORANGE LINE						DEPART VOLUME	SEG MILES	PASS-MI
			to:	vol	to:	vol	to:	vol	from:	vol	from:	vol	from:	vol			
FORE	0	1506													1506	0.69	1039
GREE	14	190													1682	0.51	858
STON	5	429													2105	0.51	1074
JACK	14	773													2865	0.55	1576
ROXB	95	221													2991	0.52	1555
RUGG	177	489													3302	0.41	1354
MASS	127	552													3727	0.69	2572
BBAY	761	1349													4315	0.55	2373
NEMC	474	80													3922	0.27	1059
CHTN	189	29													3762	0.25	940
DTXG	775	66	Red SB	216	Red NB	557	Green SB	39	Green NB	67	Red SB	244	Red NB	1010	3562	0.25	891
STAT	1256	86	Blue SB	138	Blue NB	163			Blue SB	230	Blue NB	0			2322	0.22	511
HAYM	501	97	Green SB	26	Green NB	58			Green SB	7	Green NB	65			1906	0.27	514
NSTA	720	142	Green NB	10					Green SB	7	Green NB	9			1333	0.81	1080
CCOL	293	69													1109	0.83	921
SULL	491	28													646	1.19	769
WELL	218	11													439	1.75	768
MALD	373	8													73	0.83	61
OAKG	73	0													0	0.00	0
TOTAL	6558	6125	all transfers from Orange Line					1206	all transfers to Orange Line					1639		11.10	19914

These passenger miles estimates are readily convertible into average (passenger) trip lengths for *unlinked trips*. This is the usual practice with bus passenger data, since for bus the key measure of passenger use is unlinked trips or boardings, because in the bus system the farebox serves as the control point. However, on the rail system, the turnstile is the control point, so that *entries* (from the street) is the key measure of system use. Therefore, what is needed for rail is measure of average trip length per entry, which the line / direction volume profiles do not directly provide.

10.3 Passenger-Miles Estimates From OD Flows

PROFILEMAKER estimates passenger miles a second way: directly from the estimated OD flows, multiplying each OD flow by the distance from origin to destination. For trips with alternative paths, a weighted average trip length is used, based on the path shares assumed by the model.

These passenger-mile estimates are aggregated by origin station and direction to yield estimates of passenger-miles attributable to passengers entering a station during a one hour period. Dividing by entries during the same period yields an estimate of average trip length for each entry station in each period. The periods spoken of are referenced periods, as explained earlier; for example, the 8-9 a.m. period at describes passengers boarding Red Line trains at Harvard that will pass or (for outbound passengers) have passed Downtown Crossing between 8 and 9 a.m.

Aggregations are also made to the line, day, and system. The greatest aggregation is the 14-hour day on the system (all four lines), for which the average trip length per entry was found to be 3.591 mi. This figure can be expanded by annual entries to yield an estimate of annual passenger-miles.

10.4 Analysis Of Passenger Miles Estimates By Entry Station And Period

Average trip lengths by entry station and a representative set of periods is displayed in Table 9. As one can see, average trip lengths are greatest for entries at the more distant stations such as Braintree, Alewife, Wonderland, and decrease as one moves inward. At the downtown stations, average trip length is far shorter in the a.m. than in the p.m., since the a.m. entries primarily consists of short trips being made by downtown residents, while in the p.m. entries are dominated by workers returning home.

Table 9
Average Trip Length by Station

Line	Station	7 - 8 a.m. Detail									Other Periods		
		ENTRIES			PASS-MI			AVG TRIP LEN (mi)			AVG TRIP LEN (mi)		
		SB	NB	Tot	SB	NB	Tot	SB	NB	Tot	12-1 p.m.	4-5 p.m.	all day
	PARK			270			580			2.1	2.8	3.4	3.0
	DTXG			181			467			2.6	3.0	4.0	3.7
	NSTA			2326			4082			1.8	2.3	3.4	2.5
	HAYM			982			2161			2.2	2.6	3.0	2.7
	STAT			266			788			3.0	2.3	3.4	3.1
	GOVT			162			356			2.2	2.2	2.7	2.6
Re	ALWF	1797	0	1797	10874	0	10874	6.1		6.1	5.1	5.1	5.6
Re	DAVS	1359	32	1392	6572	31	6602	4.8	1.0	4.7	3.8	4.8	4.5
Re	PORT	1557	30	1587	6595	42	6637	4.2	1.4	4.2	3.5	3.5	3.7
Re	HARV	1977	115	2091	7614	233	7847	3.9	2.0	3.8	3.6	4.2	3.8
Re	CENT	688	131	819	2133	292	2425	3.1	2.2	3.0	2.7	3.6	3.0
Re	KEND	220	49	268	553	130	684	2.5	2.7	2.5	2.4	3.3	3.1
Re	CHAS	179	156	334	394	355	748	2.2	2.3	2.2	2.9	3.4	3.3
Re	SSTA	194	754	948	847	1653	2500	4.4	2.2	2.6	3.1	4.7	4.1
Re	BWAY	58	106	164	329	256	585	5.7	2.4	3.6	3.3	3.3	3.5
Re	ANDR	14	458	472	61	1282	1344	4.3	2.8	2.8	3.4	3.3	3.1
Re	JFKU	75	237	312	313	866	1179	4.2	3.6	3.8	4.5	4.3	4.2
Re	SAVN	22	287	309	31	1125	1156	1.4	3.9	3.7	3.7	2.3	3.3
Re	FLDS	19	499	518	22	2393	2415	1.2	4.8	4.7	5.0	4.4	4.4
Re	SHAW	3	299	302	2	1702	1705	0.6	5.7	5.6	5.2	5.2	5.4
Re	ASHM	0	1344	1344	0	7865	7865		5.9	5.9	5.8	5.1	5.6
Re	NQCY	76	882	958	195	5877	6072	2.6	6.7	6.3	6.0	5.8	6.1
Re	WOLL	23	888	912	73	6563	6635	3.1	7.4	7.3	6.7	6.3	6.9
Re	QCTR	15	1378	1393	23	11760	11783	1.5	8.5	8.5	8.4	8.0	8.3
Re	QADM	9	952	962	17	9506	9523	1.9	10.0	9.9	10.2	7.4	9.6
Re	BRNT	0	939	939	0	11286	11286		12.0	12.0	10.8	10.8	11.6
Or	OAKG	1485	0	1485	9575	0	9575	6.4		6.4	6.2	6.3	6.3
Or	MALD	1714	7	1721	10284	6	10289	6.0	0.8	6.0	6.0	6.3	6.0
Or	WELL	1633	10	1643	6847	21	6868	4.2	2.1	4.2	3.9	5.0	4.2
Or	SULL	1283	26	1309	3587	74	3662	2.8	2.9	2.8	2.8	3.8	3.0
Or	CCOL	268	65	333	1304	190	1494	4.9	2.9	4.5	4.3	4.0	4.1
Or	CHTN	24	29	53	27	39	66	1.1	1.3	1.2	2.6	3.6	3.2
Or	NEMC	71	82	154	98	182	280	1.4	2.2	1.8	2.2	3.3	2.9
Or	BBAY	202	1401	1603	236	2727	2963	1.2	1.9	1.8	2.3	3.4	2.6
Or	MASS	72	535	607	129	1265	1394	1.8	2.4	2.3	2.9	2.7	2.5
Or	RUGG	24	500	524	57	1268	1325	2.4	2.5	2.5	2.9	3.6	3.0
Or	ROXB	15	201	216	30	596	626	2.0	3.0	2.9	3.0	3.9	3.1
Or	JACK	43	706	749	61	2317	2379	1.4	3.3	3.2	3.3	3.7	3.3
Or	STON	45	393	437	43	1562	1605	1.0	4.0	3.7	3.7	4.0	3.8
Or	GREE	9	192	201	7	812	818	0.7	4.2	4.1	4.0	4.5	4.0
Or	FORE	0	1422	1422	0	7059	7059		5.0	5.0	4.6	4.8	4.7
Bl	WOND	729	0	729	4494	0	4494	6.2		6.2	5.9	4.6	6.0
Bl	REVE	224	3	226	1232	1	1233	5.5	0.4	5.5	5.7	5.7	5.6
Bl	BEAC	261	3	264	1368	3	1370	5.2	1.0	5.2	4.7	2.6	4.6
Bl	SUFF	80	2	82	342	3	345	4.3	1.7	4.2	4.3	4.3	4.1
Bl	ORNT	697	4	701	2925	7	2932	4.2	2.0	4.2	4.3	4.3	4.2
Bl	WOOD	252	8	261	807	17	824	3.2	2.1	3.2	3.9	3.4	3.2
Bl	AIRP	274	17	291	929	64	993	3.4	3.8	3.4	3.7	3.2	3.1
Bl	MAVE	934	19	954	2189	27	2216	2.3	1.4	2.3	2.8	3.4	2.6
Bl	AQUA	43	30	73	103	38	141	2.4	1.3	1.9	2.9	3.2	2.9
Bl	BOWD	0	0	0	0	0	0				2.6	2.8	3.1
Gr	LECH	424	0	424	998	0	998	2.4		2.4	3.6	2.7	2.9
Gr	SCPK	26	5	31	62	3	65	2.4	0.5	2.1	1.5	2.8	2.1
Gr	BOYL	19	11	30	30	19	49	1.6	1.7	1.6	1.5	3.5	2.3
Gr	ARLT	176	42	219	229	107	337	1.3	2.5	1.5	2.1	3.3	2.9
Gr	COPL	252	336	587	286	750	1036	1.1	2.2	1.8	1.8	2.1	2.2
Gr	HYNE	121	378	499	95	851	945	0.8	2.2	1.9	2.0	2.6	2.2
Gr	KENM	86	290	376	33	639	672	0.4	2.2	1.8	2.3	2.8	2.5
Gr	SBCD	0	3294	3294	0	7887	7887		2.4	2.4	2.4	3.5	2.7
Gr	PRUD	35	65	100	20	158	178	0.6	2.4	1.8	2.1	2.9	2.2
Gr	SYMP	20	110	130	6	262	268	0.3	2.4	2.1	1.6	2.0	2.0
Gr	SURE	0	1268	1268	0	3529	3529		2.8	2.8	3.1	3.7	3.3
	TOTAL			45001			189214			4.2	3.1	3.6	3.6

10.5 Expanding To Annual Passenger Miles: Late Evenings And Weekends

The fact that the average trip length estimate made by the model covers only the hours 7 a.m. to 9 p.m. should be of no significant concern, because those hours include the preponderance of daily travel, and because trips made before 7 a.m. and after 9 p.m. are usually half of a round trips whose other half occurred in the study period, and therefore their length is fairly represented in the estimate.

However, the model makes no use of weekend data, and using a weekday estimate for weekends is not advised. The weekend travel pattern are different, being more dominated by people who do not own cars, who tend to live closer to the city center and therefore make shorter trips. By the same token, weekend use of outer stations relying heavily on park and ride is far less on weekends than on weekdays, and theses are the stations with the greatest average trip length per entry.

To fairly estimate weekend passenger miles, the PROFILEMAKER model should be run with entries data from the weekend. While the historical seed is based on weekday data, it is still an adequate seed for weekend since the fitting to weekends will correct for changing overall use of stations; the role of the historical seed is to provide information on trip distribution, i.e. how origins and destinations tend to be linked, and this kind of distribution has been shown to be highly invariant between periods and day types. This is in keeping with standard transportation planning theory, in which trip distribution depends primarily on travel time, which does not vary appreciably between weekday and weekend.

If weekend entries data by station cannot be obtained, the next best approach would be use average trip length for the midday period, since midday riders tend to be more like weekend riders than do weekday peak period riders.

11. ACCURACY OF ESTIMATES

This model can be used for estimating many quantities. The key quantities of interest to MBTA management would appear to be the following: (1) peak hour, peak point volumes; (2) transfer flows – both peak hour for capacity analysis, and all day for network analysis; and (3) passenger-miles for Section 15 reporting. The Section 15 estimates, in particular, are subject to an FTA mandated accuracy requirement of ±10% precision (relative tolerance) at the 95 percent confidence level.

11.1 Factors Affecting Accuracy

The first issue affecting accuracy of the model estimates is the quality of the entries data. Because they drive the final estimates, the accuracy of any passenger volume or passenger-mile estimate can only be as good as the accuracy of the entries data. The quality of the entries data suffers on several counts. First, mechanical difficulties with the turnstiles can cause measurement error. Second, the fact that only one day's worth of entries is counted makes the counts subject to sampling error. Third, the program of entries counts does not include weekends, providing no basis for making weekend estimates, which are needed for annual passenger-mile counts. These difficulties can be overcome, of course, with more frequent counting, careful attention to mechanical issues (functioning of turnstile registers), and improved supervision (such as independent checks and providing feedback when counts are questionable). Hopefully, new electronic fareboxes with automatic reporting will eliminate all of these issue in the near future.

The second issue is the quality of the historical data. The large production / attraction imbalances and the poor fit between the 1985 Green Line counts and the 1989 Rapid Transit line counts lowers one's confidence in estimates based on these counts. The quality of this data will hopefully improve when it is updated by the next a large-scale passenger counting effort (comparable to the 1985 and 1989 effort) undertaken by the MBTA. That effort should include checks for production / attraction imbalance.

After data quality issues, the next issue is modeling accuracy – does the model accurately convert entries to passenger flows? This question is difficult to answer objectively, as there is no source of complete "true" data against which to compare model estimates. In the following discussion, modeling accuracy will be assessed subjectively.

11.2 Accuracy Of Peak Volume Estimates

Given entries by station and direction and a general constraint on exits, there are not that many degrees of freedom with respect to peak volume, making the estimates rather robust with respect to errors in the historical data. The key issue for these measures, then, is the directional split and, at the downtown transfer stations, the line split, which the model bases on historical data.

Inbound peak volumes are most closely related to entry counts on the subject line, while outbound peak volumes are most closely related to exit counts. Because entry counts are given, while exit counts are estimated from historical data with a constraint for daily production / attraction imbalance, inbound peak volume estimates will be more accurate than outbound. The only possibility for a large error in inbound peak volumes is if the directional split has changed much since 1989, which is unlikely. Therefore, inbound peak volume estimates are probably accurate to within ±3%, given accurate entries counts. Meanwhile, outbound volumes in a given hour, being driven by exits which are estimated within the model, are probably only accurate to within ±8%. Moreover, the model structure is such that random peaks will tend to be smoothed, as the model is designed to estimate a systematic average. Therefore its estimate of the peak outbound flow for a given day's peak hour will probably be underestimated by a few percent (it is difficult to say how much). On the other hand, there will be no such bias in the estimate of a systematic mean (e.g., mean 5-6 p.m. flow on a given segment).

These levels of accuracy compare favorably with the alternatives now available to the MBTA: visual estimates, which have large measurement error, and simply factoring the flows estimated from the 1989 counts.

11.3 Accuracy Of Passenger-Miles Estimates

For Section 15 reporting, the assumed approach is that an estimate of annual system entries, obtained independently, will be expanded by a model-based average trip length. The accuracy of the result depends on the accuracy of the entries estimate and on the accuracy of the estimated average trip length.

In spite of the various uncertainties affecting an estimate of average trip length, there are few degrees of freedom for an all-day estimate. The two main sources of uncertainty are between day variation in the average trip length, assumed to have a precision of about 3%, and modeling error, assumed to have a precision of about 5%. Therefore an estimate of the average trip length based on a single day's entries counts will have a precision of ±6%, and an estimate based on many day's counts will have a precision of ±5%.

In order to use the model for Section 15 estimating, a one day sample of Saturday and Sunday entries counts by station will be needed, from which average trip lengths for these days can be estimated, which would then be used to expand estimated annual Saturday and Sunday entries. The weekend sampling could be done over a two year cycle, doing Saturday one year and Sunday the next and using the most recently available counts for the Section 15 estimate.

Assuming the annual estimate of systemwide boardings has a precision of ±7 percent (a reasonable assumption), the precision of the annual passenger-miles estimate, at the 95 percent confidence level, will be ±8.9% if the estimate is based on a single day's sample of station entries, and ±8.5% if the estimate is based on four weekday's sample of station entries. In both cases, a single set of Saturday and Sunday station entries no more than two years old is assumed.

11.4 Accuracy Of Transfer Flow Estimates

Estimates of transfer flows are affected by all the factors previously discussed. In addition, they are particularly sensitive to the 1994 survey data and the 1985 / 89 transfer counts. Their smaller absolute magnitudes also makes them more sensitive. In addition, changes in operations (headways, Green Line turnarounds, etc.) and in passenger attitudes (e.g., willingness to use the Orange Line as a substitute for the Green Line) can lead to systematic changes relative to historical patterns. Considering these factors, the margin of error for estimates of daily transfer flows is probably about ±10% for large flows, and ±20% for small flows. For peak hour transfer flows, the margin of error is probably about ±15% for large flows, and ±25% for small flows. These levels of accuracy, while not great in themselves, are still far better than could be obtained by any other means except an extensive survey or labor-intensive transfer counts.

11.5 Improving Model Accuracy With Targeted Data Collection Efforts

To conclude this section, it is worth noting some data collection efforts that would have a high payoff in terms of model accuracy versus cost (however, using these additional data sources would require some model revision to take advantage of them):

- Directional and line splits at major stations. If station entries could be split using current rather than historical data, model accuracy would be improved. Most important in this respect are downtown stations (for which the line split is needed as well as directional split) in the p.m. peak. Next most important are major entry stations in the a.m. peak where reverse direction travel may have grown; examples are Red Line stations between Davis and Charles, Ruggles, Community College, and Sullivan. This data could come from counts or simple interview surveys at the turnstiles.

- Branch transfer counts at Kenmore. The model assumes a value based on the 1994 survey, when a value based on actual counts would be much more reliable.
- Exit counts at key stations in the p.m. peak. Using exit counts at stations with large numbers of exits would improve the accuracy of peak outbound volumes. The model's estimates could still be used for minor stations.

12. CONCLUSION

A model has been developed by which the MBTA can estimate passenger flows and passenger-miles using detailed information obtained from historical counts and surveys and updated with the most recent station entries counts. It makes optimal use of the different data sources and their many dimensions, producing estimates that are far more sensitive to relationships in the historical and current data than would be produced by simple factoring. Its estimates are based on a fully specified flow model, and are therefore internally consistent in terms of flow balance, a feature that is not found (except by "fudging") in estimates produced by simple factoring. The model has been implemented in a ready-to-use C program. The accuracy of its estimates has been determined to be reasonable with respect to operational and planning decisions, and, with the addition of weekend station entries counts, will yield Section 15 passenger-miles estimates that satisfy FTA accuracy standards. The model structure is readily adaptable to accepting improved data, such as might be obtained from a new fare collection system or some targeted data collection efforts, which would further improve the accuracy of its estimates.

APPENDIX A: INSTRUCTIONS FOR USING PROGRAMS SEEDMAKER AND PROFILEMAKER

(These instructions are found in the file `README`.)

Application programs `PROFILEMAKER` and `SEEDMAKER` were delivered along with all of the input files listed below, and samples of the important output files listed below. To run one of the programs, simply double click on its icon.

INPUT FILES THAT SHOULD BE IN THE SAME FOLDER AS PROFILEMAKER:

1. `STATIONSF` (list of stations with network info)
2. `STADIRF` (station / station correspondence matrix)
3. `CROSSOVF` (segment / segment matrix)
4. `TFERSF` (list of transfer movements)
5. `ONOFFF` (ons & offs by period from 1989. Note: `PROFILEMAKER` uses only the first two lines of this file as a way of defining the period boundaries; `SEEDMAKER` uses the entire file.)
6. A file with the historic seed, such as `SEEDHIST.1`
7. (Optional, if you want to update the seed to fit the latest station entries data) A file with station entries, such as `ENTRIES95`. (If you don't use entries data with `PROFILEMAKER`, it can still make volume profiles for the seed matrix (item 6 above).)

PROFILEMAKER'S OUTPUT FILES INCLUDE:

- `echom` – echoes inputs. Also contains updated OD matrices with margins if entries data was supplied as an input file. In the disk transmitted, the output matrices with margins were copied from file echom into an Excel workbook to make viewing easier.
- `Pro_xxx` – a set of 9 volume profile files: 1 per line / direction, plus a summary file. The user names the extension on these files.)
- A user-named passenger-mile output file.
- (optional) A user-named file with the updated OD seed matrices (may later be considered as a historic seed, and used as input to either `PROFILEMAKER` or `SEEDMAKER`)

INPUT FILES THAT SHOULD BE IN THE SAME FOLDER AS SEEDMAKER:

1-5. Same as input files needed for `PROFILEMAKER`

6. `SURVEYF` – survey data file

7. (optional) A file with an historic seed, such as `SEEDHIST.1`, which will serve as a starting point for estimating a new historic seed.

SEEDMAKER'S OUTPUT FILES INCLUDE:

- `echo.out` (an echo of the input)
- `LL_prog` (a log of the progress of the maximum likelihood search)
- A user-named file with the updated OD seed matrices (may later be considered as a historic seed, and used as input to either PROFILEMAKER or SEEDMAKER)

APPENDIX B: SUMMARY OF SOURCE FILES

The source files include:

1. Header files

`EntriesDefines`
`RailDefinesDim`
`RailDefinesDist`
`RailDefinesT`
`RailStructuresT`

2. Main programs

For SEEDMAKER, `seed_main.c`
For PROFILEMAKER, `main_match_abs.c`

3. Many files containing functions (subroutines), most of which are used by both programs. These files are:

`check_sta_dir.c`
`clean_scale.c`
`CREATE OD ARRAYS` (a folder containing 33 small files)
`CREATE PAR ARRAYS` (a folder containing 33 small files)
`DeclSurv0.c`
`DeclSurv1.c`
`DeclSurv2.c`
`DeclTfer_fn.c`
`DeclTpair_fn.c`
`decl_offs_fn.c`
`decl_ons_fn.c`
`dir_fne.c`
`dist_sta_to_sta.c`
`echo_countsKENM.c`
`fill&smooth_entries_obs.c`
`FillSurvArr.c`
`fill_expansionA.c`
`fill_expansionB.c`
`fill_od_functions.c`
`fill_od_functions_pmi.c`
`fitter.c`
`fit_attractions.c`
`fit_rows_to_entries.c`
`hhmm.c`
`LL_functionsKENM.c`
`margins_off.c`
`margins_on.c`
`margins_tfer.c`
`margin_functions.c`
`margin_functions_p.c`
`output_OD_fn.c`
`output_par_fn.c`
`partials&gradientKENM.c`
`pmi_entries.c`

profile.c
ScanBfers.c
ScanOnsOffsTfersKENM.c
ScanPdsTfersN.c
scan_network.c
SetUpODandPar.c
skiptab.c
sta&tsta&xover_fns.c
tfer_match.c

APPENDIX C: SOURCE FILE LISTINGS

Listings of the source files follow. (They have been omitted from all except one copy of this report, as they take about 100 pages.)