Model To Estimate Passenger Origin-Destination Pattern on a Rail Transit Line

SHINYA KIKUCHI AND VIJAY PERINCHERRY

A method that develops a passenger origin-destination (O-D) table for a transit line is presented. The input to the model is the boarding and alighting counts at stations, and the output is the estimated passenger volume for each station pair. The model can make use of the analyst's knowledge of passenger volumes for selected station pairs if it is available in an approximate range. The O-D volumes are estimated to minimize the expected error by locating each estimate as close to the center of the feasible solution space as possible; this is accomplished by a linear programming method. The estimates can be revised iteratively by incorporating the analyst's knowledge of the passenger travel pattern. Examples include the case for which only the nondirectional boarding and alighting counts are available.

The boarding and alighting counts at stations and the origin-destination (O-D) travel pattern are the basic data for analyzing the demand for a transit line. These data provide the basic information of the number of passengers traveling between stations, which can be used to determine stopping schemes, fare structures, and schedules and to serve as the data for general system planning. The boarding and alighting counts at each station can be obtained without major difficulty as part of the routine activities of a transit agency. The O-D volumes (the number of passengers traveling between specific stations), on the other hand, are not easy to obtain. They require more-elaborate surveys of tracking passengers from their boarding stations and alighting stations. Such surveys are generally expensive to conduct, and accuracy depends on the sample size. For transit lines with many passengers—such as rail rapid transit lines—subjectively estimated passenger trip patterns, perhaps in the form of a range, may be available for selected station pairs based on past surveys and the experience of the analyst. In this paper, we propose a method that develops an estimated O-D table of a transit line using approximate information of selected O-D pairs as well as the boarding and alighting counts.

Mathematically, the proposed method estimates the elements of a passenger O-D table by solving an indeterminate system of linear equations and inequalities. The set of linear equations represents the conservation of flow equations based on the boarding and alighting counts, and the set of linear inequalities represents the information on some of the unknown parameters given by ranges of values. The elements of the O-D table are derived to minimize the expected error between "true" value and predicted value. The expected error is minimized when the estimate is at the center of the feasible range of the true value. A measure that indicates how close an estimate is to the center of the feasible range is developed. For each estimate, its closeness to the midpoint is measured, and the sum of the measures is maximized using a linear programming formulation.

Civil Engineering Department, University of Delaware, Newark, Del. 19716.

The proposed model is suited for estimating an O-D pattern for a heavily used transit line, in which the analyst has a general idea about discernible flow pattern of certain station pairs, based on general knowledge, previous surveys, and planning data. The method is characterized by its ability to estimate an O-D table based on the boarding and alighting counts at stations and estimated ranges of O-D volumes for some pairs. The method can also be used to estimate a bidirectional O-D table when the boarding and alighting counts are available only for the total of the bidirectional movement; for example, the data collected at the fare gates of rail transit stations.

PREVIOUS WORK

Similar problems are found among papers dealing with the development of an O-D table in four subjects: the context of the travel demand forecasting process, the passenger travel pattern on a transit line, intersection turning volumes, and freeway travel patterns. The goal common to all applications is the identification of the elements of the O-D matrix given the row and column totals (trip generation and attraction) and the information on the elements of the matrix. In developing the O-D table of a transit line, the approaches may be grouped into two types: one based on the improvement of a "seed" (or a priori) O-D matrix; and the other, which does not use the seed matrix, based on the analogy to the fluid flow.

Ben-Akiva et al. compared iterative proportional fitting (IPF), constrained generalized least-squares method (CGLS), constrained maximum likelihood estimation (CMLE), and fluid analogy method using the actual transit ridership data (*1*). The first three methods require a seed matrix, and the elements of the seed matrix are iteratively revised to satisfy the conservation of flow principle. The IPF method revises the value of the elements iteratively to obtain a balance between the boarding and alighting counts. Furth also applied this method to estimate intersection turning movements and compared the estimates and the observed values (*2*). The CGLS and CMLE methods make certain assumptions about the relationship between the true value and the sampled value and solve optimization models that take into account the conservation of flow. A large number of models and discussions are presented on the estimation of an O-D table in the context of travel demand forecasting, including entropy maximization and minimization of information (*3*), maximum likelihood (*4*), and generalized constrained least-squares model (*5*).

The fluid analogy method requires no seed O-D table; it uses only the boarding/alighting counts. It assumes a certain rule by which boardings and alightings are related at each station. At a station, passengers are equally likely to alight after they have traveled on the vehicle for at least a minimum distance. The ratio between the actual number of alightings and the total passengers eligible to alight is applied to the boarding passengers at each of the previous stops to determine the O-D pattern. Simon and Furth also show an application of the fluid analogy method to estimate an O-D table of a bus line (*6*). Furth further studied the procedure of updating an O-D table by multiproportional method after obtaining the initial matrix by the fluid analogy method (*7*). Although the fluid method is simple, straightforward, and easy to apply, its problem is the rigidity of the assumption. It lacks the mechanism to consider the travel pattern unique to a line. Another problem is that it cannot logically be applied to the case of a bidirectional O-D table; in other words, the input data must be the directional boarding/alighting counts. If the boarding/alighting counts are made at rail transit stations and directional separation of counts is not possible, applicability of the fluid analogy to rail transit O-D table development is questionable.

Additional literature on estimating an O-D table of a linear movement pattern without a seed matrix includes works by Stokes and Morris, who use simplified maximum likelihood estimates on a two-way contingency table (*8*), and Nihan and Davis, who show, among several approaches, a nonrecursive ordinary least-squares model for estimating the trip pattern on a freeway based on in-out counts at ramps (*9*). It requires the operation of the inversion of a large matrix and many total sets of data on total boardings and alightings along the line.

None of the models described above has the ability to incorporate the approximate information that the analyst may be able to provide. The effective use of such information requires a model that can incorporate approximate seed volumes for some O-D pairs in addition to boarding and alighting counts. The approximate volumes may be in the form of a range of values; for example, "10 to 50 percent of passengers boarding at Station A travel to B," or "less than (or more than) 60 percent of the passengers boarding at Station A should travel to B." A process that interacts with the analyst and incrementally improves the solution is also desirable. For example, if some elements of the derived O-D table do not look reasonable, the analyst can generate a second O-D table after revising the initial ranges of estimates.

PROBLEM AND BASIC EQUATIONS

Problem

For a transit line with a fixed number of stations, one-way passenger volume for every station pair is to be estimated for a given period. The following data are known to the analyst:

1. The numbers of boardings and alightings at each station for one direction of vehicle movement for the period in question (later we will show an example in which the boarding and alighting counts are available only for two-way volume).
2. Some knowledge of the O-D pattern of the passengers using the line. The degree of the analysts' knowledge may vary among the station pairs. For certain station pairs they may be confident, whereas for some other pairs they may not have any idea. The knowledge of the travel pattern for some station pairs may be expressed as "Between x and y percent of the passengers boarding at Station A travel to Station B." If no knowledge is available, the range is "between 0 and 100 percent."

Basic Equations

Consider one direction of vehicle movement on a transit line that has n stations, including both terminals, and denote a_{ij} as the number of passengers boarding at Station i who travel to Station j. The number of passengers alighting at Station j must be equal to the sum of the passengers who board at prior stations and travel to Station j, and each passenger who boards at Station i must alight at one of the stations $i + 1$ to n. The following relationships exist between the boarding passengers and the alighting passengers (these may be called the conservation of flow equations).

$$\sum_{i=1}^{j-1} a_{ij} = Q_j \qquad \text{for } j = 2,3, \ldots, n \tag{1}$$

and

$$\sum_{j=i+1}^{n} a_{ij} = P_i \qquad \text{for } i = 1,2,3 \ldots, n-1 \tag{2}$$

where P_i is the number of passengers boarding at Station i during the analysis period, and Q_j is the number of passengers alighting at Station j for the same period.

The problem is to estimate the values of a_{ij}s that satisfy Equations 1 and 2. Since there are $n(n - 1)/2$ unknowns and $2(n - 1) - 1$ equations in Equations 1 and 2 (one of the equations can be derived from the remaining equation), a unique set of solutions can be obtained only when $n = 3$ (this is the case with one intermediate station). When n is greater than 3, the problem becomes an indeterminate system of linear equations; thus, normally, many sets of solutions exist.

If the approximate volumes are available for selected O-D pairs, they are expressed as ranges as follows:

$$s_{1(i,j)} \leq a_{ij} \leq s_{2(i,j)} \qquad \text{for } (i,j) = (1,2) \ldots, (n-1,n) \tag{3}$$

where $s_{1(i,j)}$ and $s_{2(i,j)}$ are lower and upper bounds of the estimated range for a_{ij}, respectively. If it is more realistic to assume that the range is given in percent of P_i, then $s_{1(i,j)}$ and $s_{2(i,j)}$ can be computed on the basis of the estimated percents of P_i. If no external bound is given to a_{ij}, the lower and upper bounds, $s_{1(i,j)}$ and $s_{2(i,j)}$ of a_{ij}, are determined by Equations 1 and 2 as

$$s_{2(i,j)} = \min\,[Q_j, P_i] \tag{4}$$

$$s_{1(i,j)} = \max\left[P_i - \sum_{k=i+1\,(k\neq j)}^{n} s_{2(i,k)},\ Q_j - \sum_{\ell=1(\ell\neq i)}^{j-1} s_{2(\ell,j)},\ 0\right] \tag{5}$$

We now have a problem that has $n(n-1)/2$ unknowns, with $2(n-1)-1$ equations (Equations 1 and 2) and $n(n-1)/2$ inequalities (Equation 3), which bind the solution space of the unknowns.

APPROACH

The problem is to solve for a_{ij} from the set of expressions that are Equations 1 through 3. Our approach is to identify the values for a_{ij} that would result in the least expected error between the true value and the predicted value. Before solving the problem, let us consider the following simple two-variable problem as an example.

Two-Variable Example

Suppose that the values of two parameters, x and y, are to be determined when the following conditions are given:

$$\alpha x + \beta y = w \tag{6}$$

$$a \leq x \leq b \tag{7}$$

$$d \leq y \leq e \tag{8}$$

where α, β, a, b, d, and e are constants greater than or equal to zero.

Graphically, the feasible region for x and y lies on the line segment AB shown in Figure 1. From Equations 6 through 8 combined, the values of x and y are bound by

$$v_{1x} \leq x \leq v_{2x} \tag{9}$$

$$v_{1y} \leq y \leq v_{2y} \tag{10}$$

The set of (x,y) values that corresponds to the midpoint of line AB represents the "safest" estimates for x and y because at this point the expected error from the true value is minimized. This expected error is the expected difference between the estimated and the true value, assuming that the location of the true value is unknown and anywhere between A and B. If the location of the true value is assumed to be uniformly distributed over the line AB, it can be proved that the expected value of the distance between the estimated and the true values is minimum when the estimate is at the center of the line.

Let us now introduce artificial variables c_x and c_y, which are defined as

$$0 \leq c_x \leq z_x \tag{11}$$

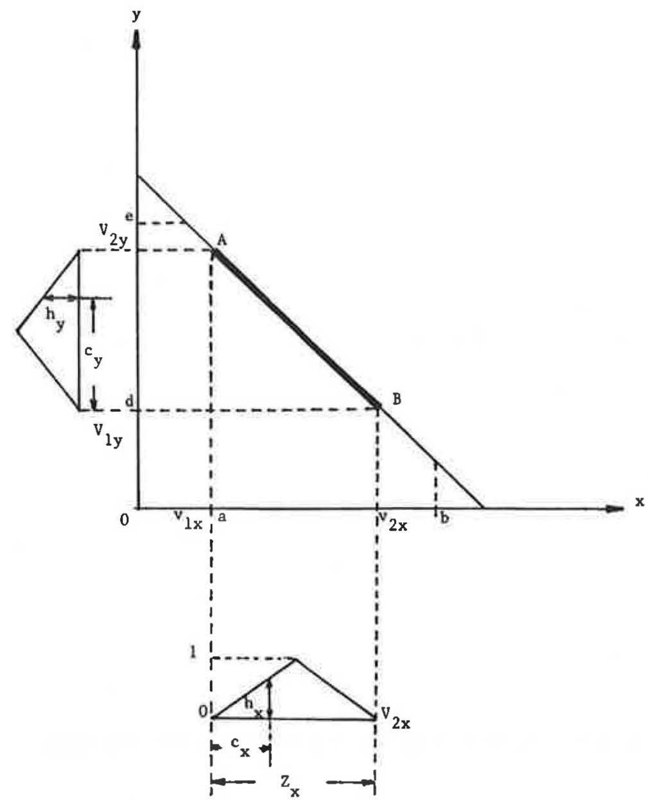

FIGURE 1 Feasible region and explanation of *h*.

$$0 \leq c_y \leq z_y \tag{12}$$

where

$$z_x = v_{2x} - v_{1x} \tag{13}$$

$$z_y = v_{2y} - v_{1y} \tag{14}$$

where z_x and z_y represent the sizes of the feasible regions of x and y, respectively. Our task is to locate the value of x and y as close to the middle of z_x and z_y, respectively, as possible.

Assume variables h_x and h_y, which represent the measure of how close the values of x and y are to the middle of z_x and z_y, respectively, and let h_x and h_y follow triangular functions, as shown in Figure 1. The functions peak at the middle of z_x and z_y and the peak values are 1.

Let us now express x and y as

$$x = v_{1x} + c_x \tag{15}$$

$$y = v_{1y} + c_y \tag{16}$$

The degree that x and y are close to the middle of the z_x and z_y is measured respectively by

$$h_x = \min\left\{\frac{2c_x}{z_x},\ 2 - \frac{2c_x}{z_x}\right\} \tag{17}$$

$$h_y = \min\left\{\frac{2c_y}{z_y},\ 2 - \frac{2c_y}{z_y}\right\} \tag{18}$$

Therefore, the values of x and y that are closest to the middle of z_x and z_y can be found by maximizing $h_x + h_y$ and also maximizing the minimum value of h_x and h_y. This forms the following linear programming (LP) problem:

Objective:

$$\max h_x + h_y \text{ and } \max\{\min(h_x, h_y)\} \tag{19}$$

subject to

$$\alpha(\nu_{1x} + c_x) + \beta(\nu_{1y} + c_y) = w \tag{20}$$

$$\frac{2c_x}{z_x} \geq h_x \tag{21}$$

$$2 - \frac{2c_x}{z_x} \geq h_x \tag{22}$$

$$\frac{2c_y}{z_y} \geq h_y \tag{23}$$

$$2 - \frac{2c_y}{z_y} \geq h_y \tag{24}$$

$$c_x, c_y, h_x, h_y \geq 0$$

In practice, the $\max\{\min(h_x,h_y)\}$ objective in Equation 19 can be accommodated by setting additional constraints of $h_x \geq h_z$ and $h_y \geq h_z$, where h_z is a threshold that defines the minimum value of h_x and h_y. The value of h_z is provided externally on a trial-and-error basis. The LP model here is identical to the formulation of fuzzy LP formulation in which satisfaction of the decision maker, as represented by h_z, is to be maximized under constraints.

Multivariable Formulation

We now expand the formulation to the problem defined by Equations 1 through 3. First, we redefine the boundary of the feasible region of each variable based on Equations 1 through 3 as

$$\nu_{1(i,j)} \leq a_{ij} \leq \nu_{2(i,j)} \tag{25}$$

Since a_{ij} appears once each in Equations 1 and 2, and all coefficients and the value of the right-hand side of the equations are positive, $\nu_{1(i,j)}$ and $\nu_{2(i,j)}$ can be systematically determined after incorporating the range defined in Equation 3.

We now introduce a slack variable, c_{ij} for a_{ij}, which corresponds to c_x (or c_y) in the two-variable example. This variable represents the distance between the lower boundary of the feasible range and the estimated value of the variable. Using the same approach as mentioned in the two-variable case, we define the slack variable, c_{ij}, for each a_{ij} as follows:

$$a_{ij} = \nu_{1(i,j)} + c_{ij}$$

$$0 \leq c_{ij} \leq z_{ij} \tag{26}$$

where z_{ij} $[= \nu_{2(i,j)} - \nu_{1(i,j)}]$ is the size of the range of a_{ij}.

We then set up a function h_{ij} such that

$$h_{ij} = \min\left\{\frac{2c_{ij}}{z_{ij}}, 2 - \frac{2c_{ij}}{z_{ij}}\right\} \tag{27}$$

as was the case in Equations 17 and 18.

The value of h_{ij} is a measure of how close c_{ij} is to the center of $[\nu_{1(i,j)}, \nu_{2(i,j)}]$. The value of h_{ij} lies between 0 and 1, and the closer the value of h_{ij} is to 1, the closer the obtained a_{ij} is to the midpoint of the feasible range.

The formulation of the model that corresponds to Equations 17–22 is

$$\max \sum_i \sum_j \mathrm{h}_{ij} \tag{28}$$

subject to

$$\sum_i [\nu_{1(i,j)} + c_{ij}] = Q_j \quad \text{for all } j \text{ (from Equation 1)} \tag{29}$$

$$\sum_j [\nu_{1(i,j)} + c_{ij}] = P_i \quad \text{for all } i \text{ (from Equation 2)} \tag{30}$$

$$\frac{2c_{ij}}{z_{ij}} \geq h_{ij} \quad \text{for all } i,j \text{ (from Equation 27)} \tag{31}$$

$$2 - \frac{2c_{ij}}{z_{ij}} \geq h_{ij} \quad \text{for all } i,j \text{ (from Equation 27)} \tag{32}$$

$$h_{ij} \geq h_z \quad \text{for all } i,j \text{ (from Equation 19)} \tag{33}$$

$$c_{ij}, h_{ij} \geq 0 \quad \text{for all } i,j \tag{34}$$

The inputs to the above LP formulation are $\nu_{1(i,j)}$, Q_j, P_i, z_{ij}, and h_z, where constraint $h_{ij} > h_z$ in Equation 33 acts as $\max\{\min[h_{ij}$, for all $(i,j)]\}$, as defined in the second equation of Equation 19. h_z is an externally provided value ($0 < h_z < 1$). Equation 33 ensures that the minimum value of h_{ij} is greater than at least h_z. It is solved for c_{ij} and h_{ij}. The O-D volume, a_{ij}, is obtained by $\nu_{1(i,j)} + c_{ij}$, according to Equation 26. h_{ij} indicates the degree of closeness of a_{ij} to the center of the range.

The existence of the solution for this LP model depends on the range of the estimated value for c_{ij}, z_{ij}, as expressed in Equation 26. If the solutions cannot be obtained, a different range must be supplied or the current range should be relaxed, and the process should be repeated. If no range is given, other than the one determined by Equations 4 and 5, one should always get a set of solutions. This is the solution for which no external estimates are given.

To compensate for the possible error of the analyst's estimates, more than one analyst may be employed to provide different sets of estimated ranges, and the procedure discussed is repeated for each set of estimated ranges. The average of the results may be used as the aggregate measure of the passenger O-D pattern.

The procedure can be briefly summarized in the following eight steps:

1. Obtain the boarding (P_i) and alighting (Q_j) counts at each Station i.

2. Estimate the range of passenger volume for trips between i and j: minimum $v_{1(i,j)}$; the range z_{ij}. If the approximate range is not available, $v_{1(i,j)} = s_{1(i,j)}$ from Equation 5 and $z_{ij} = \min(P_i,Q_j) - s_{1(i,j)}$.
3. Determine acceptable value of h_z.
4. Formulate an LP model according to Equations 28 through 34, and solve for c_{ij}.
5. Prepare the O-D table. The O-D volume for the station pair $i-j$ is $v_{1(i,j)} + c_{ij}$.
6. Inspect the O-D table and identify the station pairs whose values do not match the analyst's subjective feeling.
7. Introduce new ranges for these O-D pairs (which may be based on subjective judgment), and adjust the ranges according to Step 2.
8. Repeat Steps 3 through 6 until the O-D volumes do not conflict with the analyst's observation and feeling.

EXAMPLES

The estimated O-D volumes of the proposed method are compared with the actual travel data of two transit lines: one, the Lindenwold line in Philadelphia, Pa., and the other, a new people-mover line in Yokohama, Japan. The results of the proposed model are also compared with the ones derived from the fluid analogy method.

Example 1: Lindenwold Line O-D Volume

In the Lindenwold line example, the estimated O-D matrix of the Lindenwold line is compared with the actual data obtained from the 1979 O-D survey. The Lindenwold line is a rail rapid transit line that traverses between Philadelphia and Lindenwold, New Jersey, and is operated by Delaware River Port Authority. There are 13 stations on the line including the two end stations. The travel pattern of the passengers focuses to and from Philadelphia; it collects passengers to Philadelphia for its westbound travel and distributes them from Philadelphia in its eastbound travel. The actual O-D data provided to us by Delaware River Port Authority (*10*) are adjusted from the sample survey of 3,226 counts, and the adjusted O-D table is a symmetric table with the total number of 40,532 daily trips in both directions, which was based on a sample survey of 3,226 passengers.

A model is constructed according to the formulation shown in Equations 28–34. The O-D table of the line is estimated by more than one run of the LP model. Starting with the case that has no information other than the boarding and alighting counts at each station, each run incorporates additional information on the estimated range for c_{ij} for selected (i,j) pairs.

The following runs were tested:

Run 1. Run 1 is based on the boarding and alighting counts only—all P_is and Q_js (no subjective estimates for a_{ij}) only. Having obtained results from run 1, the O-D volumes for selected station pairs that do not appear reasonable are adjusted based on partial information for those station pairs; ranges of values considered are based on the general travel patterns obtained from the total boardings and the total alightings. The following ranges (in proportion of the total boarding at i) are incrementally incorporated for each of the subsequent runs:

Run 2. $0.2 \leq (a_{i,13}/P_i) \leq 1.0$, for all i; and data for Run 1.
Run 3. $0.1 \leq (a_{i,8}/P_i) \leq 1.0$, for all i; and data for Run 2.
Run 4. $0.15 \leq (a_{i,11}/P_i) \leq 1.0$, for all i; and data for Run 3.

These values of the ranges are determined considering the number of alightings at Stations 13, 8, and 11. For example, Station 13 has the highest proportion (approximately 25 percent: 5,162 ÷ 20,264) of the passengers alighting; thus, a rough range of "greater than 20 percent or $0.2 \leq a_{i,13}/P_i \leq 1.0$" is selected for Run 2. Similarly, the subsequent runs incorporate additional ranges to selected elements on the basis of the boarding and alighting counts at stations.

Table 1 compares the actual volumes with the result of Run 4, the upper value of each cell being the result of Run 4 and the lower value being the actual volume. Table 2 shows the changes in the accuracy of the estimates as additional information, represented by the ranges in each run, is incorporated. It is seen that the number of matrix elements within a given margin of error increases as more information is incorporated. In Run 4, all elements are within the margin of error of 500 (which is less than 2.5 percent of the total passenger volume).

Table 2 also compares the results of the fluid analogy method with those of Runs 1, 2, 3, and 4 of the proposed model. The performance of the two methods is compared using the values of the correlation coefficient and the slope of the least-squares fit for the relationship between the actual and estimated values. With each run, the result of the proposed model improves, but the fluid analogy method yields a slightly better set of estimates than the proposed method (under Run 4) based on the performance indicators. This improvement may be attributed to the fact that the travel pattern of the eastbound Lindenwold line is similar to the fluid flow from a high point to low points, because most passengers board at stations in Philadelphia (Stations 1 through 4) and travel to the remaining stations.

Example 2: Yokohama's Transit Line

A second example is based on the O-D data of a newly built automated people-mover system in Japan. The system is outside Yokohama, and it has 14 stations including the end stations. Both end stations are connected to the stations of a heavily used rail transit line. Unlike Lindenwold line, a one-way movement of the train performs two major functions: it distributes passengers from the starting terminal to the stations on the middle of the line, and it collects passengers from these middle-of-the-line stations and transports them to the other terminal. The O-D data (surveyed December 14, 1989), were obtained from the computerized ticket validation counts.

An analysis similar to that of Example 1 is performed for this line. Because in this case we have the complete actual O-D data, we tested three cases: each direction of movement separately and both directions combined. The following are the inputs used for the runs.

TABLE 1 LINDENWOLD LINE: 1979 O-D DATA EASTBOUND

Destination Stations

	1	2	3	4	5	6	7	8	9	10	11	12	13	Total
1	0	0 0	2 4	60 103	7 12	404 256	219 180	1250 1237	504 607	994 701	1100 1104	548 939	1811 1760	6903
2		0	2 0	51 17	1 5	85 167	64 118	500 497	132 213	149 278	400 409	628 331	700 679	2714
3			0	9 1	1 0	58 34	58 37	150 161	58 90	58 81	100 97	135 117	150 162	780
4				0	9 2	568 680	426 494	1192 1170	882 553	594 713	1167 1166	928 958	1500 1533	7269
5					0	19 0	19 14	55 83	20 29	20 23	40 40	20 5	70 70	264
6						0	63 9	50 49	106 199	106 137	225 215	365 275	500 533	1417
7							0	8 7	48 67	48 27	90 91	147 151	300 298	641
8								0	13 11	13 20	35 36	17 15	100 97	179
9									0	4 7	10 9	24 23	10 10	49
10										0	7 8	19 22	10 9	37
11											0	1 0	10 11	11
12												0	0 0	0
13													0	0
	0	0	4	121	19	1137	852	3204	1769	1989	3175	2836	5162	20,264

Notes: The upper number in each cell is the estimated O-D volume for Run No. 7.
The lower number in each cell is the actual O-D volume.

TABLE 2 SUMMARY OF PERFORMANCE OF MODEL

	Cumulative Number of Elements within Margin of Error (E)[1]						Performance measure Actual vs. Estimated	
	E≤20	E≤50	E≤100	E≤500	E≤1000	E≤2000	Correlation[2] Coefficient	Slope[3]
Lindenwold-Westbound (Total elements 78)								
LP Model Run 1	34	45	58	76	78	–	0.911	0.899
LP Model Run 2	36	49	59	77	78	–	0.911	0.897
LP Model Run 3	41	54	64	77	78	–	0.931	0.924
LP Model Run 4	44	65	70	78	–	–	0.976	0.974
Fluid Analogy Model	45	62	68	78	–	–	0.984	0.988
Yokohama-Westbound (Total elements 91)								
LP Model Run 1	54	66	73	90	90	91	0.773	0.684
LP Model Run 2	69	84	90	91	–	–	0.987	0.975
Fluid Analogy Model	61	74	82	91	–	–	0.928	0.921
Yokohama-Eastbound (Total elements 91)								
LP Model Run 1	55	65	70	90	90	91	0.752	0.698
LP Model Run 2	66	83	85	91	–	–	0.984	1.022
LP Model Run 3	67	83	88	91	–	–	0.977	0.991
Fluid Analogy Model	63	73	83	90	91	–	0.928	0.942
Yokohama-Both Direc. (Total elements 182)								
LP Model Run 1	122	137	143	179	180	182	0.822	0.813
LP Model Run 2	127	145	157	181	182	–	0.923	0.903
LP Model Run 3	131	150	162	182	–	–	0.955	0.938

Notes: 1. E = |actual volume – (minus) estimated O-D volume|

2. Correlation coefficient of the regression line of the relationship between the actual and estimated volumes.

3. Slope represents the gradient of the regression line (y=ax).

- Westbound

Run 1. Boarding and alighting counts at all stations (all P_s and Q_js).

Run 2. $0.05 \leq (a_{1,14}/P_i) \leq 0.15$, and data for Run 1.

- Eastbound

Run 1. Boarding and alighting counts at all stations (all P_is and Q_js).

Run 2. $0.1 \leq a_{14,1}/P_{14} \leq 0.3$, and data for Run 1.

Run 3. $0.5 \leq a_{14,6}/P_{14} \leq 1.0$, and data for Run 2.

- Bidirectional

Run 1. Boarding and alighting counts at all stations (all P_is and Q_js).

Run 2. $0 < a_{1,14}/P_i < 0.2$, $0 < a_{14,1}/P_{14} < 0.2$, and data for Run 1.

Run 3. $0.1 < a_{6,14}/P_6 < 0.4$, $0.1 < a_{8,14}/P_8 < 0.4$, and data for Run 2.

To determine the ranges shown subjectively the distributions of total alighting volumes and boarding volumes are examined. The results of the runs are shown in Table 2. In all cases, as more information on selected O-D pairs is incorporated, the accuracy of the estimate improves significantly; particularly, the change in the performance from run 1 to run 2 is significant. Run 2 of the westbound O-D table is a result incorporating a range to only one element (*1,14*). Table 3 shows the estimated and actual O-D tables for the westbound and the eastbound movements separately. The upper value of each cell is the estimated value, and the lower value is the actual value. The estimated volumes for the westbound are based on the results of Run 2 and for the eastbound, Run 3.

As for Example 1 (the Lindenwold line), we compare the estimates obtained using the proposed method with those obtained using the fluid analogy method in Table 2. In this example, the estimates using the proposed method are found to perform better than those using the fluid analogy method in terms of the number of elements within a given margin of error and the performance measures. As seen in Table 2, the results of the coefficient of correlation and the slope of the least-squares fit for the relationship between the actual and estimated volumes indicate that the proposed model (after additional information) yields better estimates than the fluid analogy method. This may be caused by the unique passenger travel characteristics of this line, as described, which has a less "fluid" passenger flow pattern. In addition, a bidirectional O-D table was estimated by the proposed model using the total boarding/alighting counts at each station. The results

TABLE 3 YOKOHAMA PEOPLE-MOVER O-D TABLE (ESTIMATED AND ACTUAL)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
1		368 368	924 928	844 849	970 919	992 1025	819 846	861 877	497 577	13 25	91 131	125 163	90 188	1096 797	7690 7693
2	400 400		14 11	14 9	14 18	14 17	14 10	14 8	14 4	3 0	14 4	14 5	14 13	48 98	599 597
3	1011 1056	53 9		33 34	33 63	33 38	33 16	33 15	33 24	2 2	16 17	23 23	28 31	170 178	1507 1506
4	720 763	8 7	78 38		24 43	24 7	24 27	24 50	24 33	2 1	17 15	23 15	24 32	138 106	1140 1137
5	975 942	8 11	24 58	18 16		36 15	36 47	36 18	36 28	2 2	16 16	23 25	28 22	271 294	1519 1517
6	1135 1147	8 9	24 22	17 12	23 20		35 18	35 55	35 46	2 1	16 15	23 23	28 22	295 294	1686 1684
7	645 888	8 9	24 32	58 30	123 41	157 18		33 16	41 12	2 1	16 12	23 20	28 20	300 366	1470 1465
8	823 853	8 8	24 44	17 29	24 59	28 61	139 17		46 6	2 1	16 10	23 18	28 26	502 560	1690 1692
9	589 608	8 4	24 13	17 37	24 58	28 43	20 25	75 8		2 1	16 5	23 12	28 12	308 350	1173 1176
10	18 22	2 0	2 1	2 1	2 5	2 3	2 2	1 1	5 2		1 1	1 4	8 2	12 17	58 61
11	107 104	8 5	11 20	11 15	11 25	11 17	11 13	10 10	33 6	1 0		6 5	37 8	44 76	301 304
12	139 130	8 3	14 28	14 14	14 20	14 21	14 23	14 19	22 15	1 3	24 3		35 7	35 63	348 349
13	178 165	8 15	18 27	17 22	18 28	18 27	18 20	18 25	15 13	12 0	17 9	17 5		211 211	565 567
14	965 646	38 90	246 209	171 167	239 223	312 382	204 309	306 364	226 266	16 29	41 71	67 80	175 175		3006 3011
Total	7725 7724	543 538	1432 1431	1233 1235	1519 1522	1682 1674	1369 1373	1465 1466	1037 1032	60 66	304 309	391 398	551 558	3430 3433	

Notes: 1. The upper right matrix represents the westbound movement
2. The lower left matrix represents the eastbound movement
3. The upper number of each cell is the estimated value and the lower number is the actual O-D volume
4. For westbound (Stations 1→14) movement, the estimated values are the results of Run 2
5. For eastbound (Stations 14→1) the estimated values are the results of Run 3
6. Due to rounding of estimated volumes there are slight differences between the sums of actual and estimated volumes on each

of the correlation analysis between the actual and estimated O-D volumes for this case are also shown in Table 2. A comparison with the fluid analogy method is not performed because the fluid analogy cannot logically be applied to the bidirectional case.

CONCLUSIONS

This paper has presented a method that estimates the O-D pattern of passenger travel along a transit line. The input to the model is the boarding and alighting counts at each station and estimated ranges of passenger O-D volumes for selected station pairs. The estimated ranges may be given by an analyst who is familiar with the O-D pattern along the line. The ranges may also be inferred from past O-D surveys, from analyst observation, or from values derived by other O-D estimating methods. Although the proposed method is an approximate method, the examples demonstrate that it can yield reasonably accurate estimates of the O-D pattern and at least the same level of accuracy as the fluid method. Unlike the fluid method, the proposed method can improve the estimates based on incomplete information on the O-D pattern. It is particularly interesting to notice how quickly the estimates improve by incorporating loose ranges on only one or two O-D pairs.

The method solves an indeterminate system of linear systems with the aid of information on the ranges of the values of selected unknown parameters. The advantage of the proposed method is that analysts can incorporate estimated O-D information (in a range) for only those pairs for which they have some confidence. The method is also suited for the transit lines in which the fluid analogy travel pattern is hard to justify, such as the case of bidirectional O-D.

The method can be used not only for estimating the O-D table of a transit line but also for a number of other applications; for example, (a) the distribution of the duration of stay at a parking lot can be estimated for the counts of vehicles entering and exiting the lot over the period; (b) the vehicle travel pattern along a freeway or an arterial can be estimated from the traffic counts at entrances and exits at the ramps or at intersections; and (c) the characteristics of the bypass traffic can be estimated for a small city when the inbound and outbound traffic volumes on each of the roads leading to the city are known and the planner supplies the estimated values of the bypass traffic between two road pairs. In general, the method determines the cause-and-effect relations of a system: the causes are passenger boardings at various stations, and the effects are the alightings at various stations. The travel pattern derived is the relationship between these boardings and alightings.

REFERENCES

1. M. E. Ben-Akiva, P. P. Macke, and P. S. Hsu. Alternative Methods to Estimate Route Level Trip Tables and Expand On-Board Surveys. In *Transportation Research Record 1037*, TRB, National Research Council, Washington, D.C., 1985, pp. 1–11.
2. P. G. Furth. A Model of Turning Movement Propensity. In *Transportation Research Record 1287*, TRB, National Research Council, Washington, D.C., 1990.
3. H. J. Van Zuylen and L. G. Willumsen. The Most Likely Trip Matrix Estimated From Traffic Counts. *Transportation Research*, Vol. 14B, 1980, pp. 281–293.
4. H. Spiess. A Maximum Likelihood Model for Estimating Origin-Destination Matrices. *Transportation Research*, Vol. 21B, No. 5, 1987, pp. 395–412.
5. C Hendrickson and S. McNeil. Estimation of Origin-Destination Matrices with Constrained Regression. In *Transportation Research Record 976*, TRB, National Research Council, Washington, D.C., 1984, pp. 25–32.
6. J. Simon and P. G. Furth. Generating a Bus-Route O-D Matrix from On-Off Data. *Journal of Transportation Engineering*, Vol. 111, No. 6, 1985, pp. 583–593.
7. P. G. Furth. Updating Ride Checks with Multiple Point Checks. In *Transportation Research Record 1209*, TRB, National Research Council, Washington, D.C., 1989, pp. 49–57.
8. R. W. Stokes and D. E. Morris. Application of an Algorithm for Estimating Freeway Trip Tables. In *Transportation Research Record 976*, TRB, National Research Council, Washington, D.C., 1984, pp. 21–23.
9. N. L. Nihan and G. A. Davis. Recursive Estimation of Origin-Estimation Matrices from Input/Output Counts. *Transportation Research*, Vol. 21B, No. 2, 1987, pp. 149–163.
10. *Lindenwold Line Passenger Survey*, Volume 1. UMTA Project IT-09-0073. Delaware Valley Regional Planing Commission, 1979.

Publication of this paper sponsored by Committee on Rail Transit Systems.