An Alternative Method to Estimate Balancing Factors for the Disaggregation of OD Matrices

UCI-ITS-WP-13-5

Neda Masoud
Fatemeh Ranaiefar
Michael McNally, Ph.D
Daniel Rodriguez-Roman
Stephen G. Ritchie, Ph.D

Institute of Transportation Studies
University of California, Irvine; Irvine, CA 92697-3600, U.S.A.

nmasoud@uci.edu, franaief@uci.edu, mmcnally@uci.edu,
d.rodriguez@uci.edu, sritchie@uci.edu

August, 2013

Institute of Transportation Studies
University of California, Irvine
Irvine, CA 92697-3600, U.S.A.
http://www.its.uci.edu
An Alternative Method to Estimate Balancing Factors for the Disaggregation of OD Matrices

Neda Masoud
Ph.D. Student
Department of Civil and Environmental Engineering and Institute of Transportation Studies
University of California, Irvine
Irvine, CA, USA, 92697
nmasoud@uci.edu

Fatemeh Ranaiefar
Ph.D. Candidate
Institute of Transportation Studies
University of California, Irvine
Irvine, CA, USA, 92697
franaief@uci.edu

Michael McNally
Professor
Department of Civil and Environmental Engineering and Institute of Transportation Studies
University of California, Irvine
Irvine, CA, USA, 92697
mmcnally@uci.edu

Daniel Rodriguez-Roman
Ph.D. Student
Department of Civil and Environmental Engineering and Institute of Transportation Studies
University of California
Irvine, CA, USA, 92697
d.rodriguez@uci.edu

Stephen Ritchie
Professor
Department of Civil and Environmental Engineering and Institute of Transportation Studies
University of California, Irvine
Irvine, CA, USA, 92697
sritchie@uci.edu
Revision Submitted: July 31, 2013
Word Count: 5,510
Tables and Figures: 1 Tables + 6 Figures = 1750 words
Total Word Count: 7,260

To be considered for Presentation and Publication in the 93rd TRB Annual Meeting only

Submitted to the Joint Call for Papers sponsored by AT015, AT025 and ADB30(1):
Quantitative freight transportation and logistics modeling.
ABSTRACT

The solution algorithms for the family of flow distribution problems, which include (1) the trip distribution problem of travel forecasting, (2) the OD estimation from link counts problem, and (3) the trip matrix disaggregation problem, are usually based on the Maximum Entropy (ME) principle. ME-based optimization problems are hard to solve directly by optimization techniques due to the complexity of the objective function. Thus, in practice, iterative procedures are used to find approximate solutions. These procedures, however, cannot be easily applied if additional constraints are needed to be included in the problem.

In this paper a new approach for balancing trip matrices with application in trip matrix disaggregation is introduced. The concept of generating the most similar distribution (MSD) instead of the Most Probable Distribution of Maximum Entropy principle is the basis of this approach. The goal of MSD is to minimize the deviation from the initial trip distribution, while satisfying additional constraints. This concept can be formulated in different ways. Two MSD-based objective functions have been introduced in this paper to replace the ME-based objective function. One is the Sum of Squared Deviations MSD (SSD-MSD), and the other is Minimax-MSD. While SSD-MSD puts more emphasis on maintaining the base year trip shares as a whole, Minimax-MSD puts more emphasis on maintaining the share of each individual element in the trip table.

The main advantage of replacing the entropy-based objective functions with any of these functions is that the resulting problems can include additional constraints and still be readily solved by standard optimization engines. In addition, these objective functions could produce more meaningful results than entropy-based functions in regional transportation planning studies, as shown in the case study and some of the examples in the paper. Several examples and a case study of the California Statewide Freight Forecasting Model (CSFFM) are presented to demonstrate the merits of using MSD-based formulations.

Keywords: trip distribution, maximum entropy, trip matrix disaggregation, Most Similar Distribution
INTRODUCTION

The trip distribution problem is a familiar component of passenger and freight transportation planning models. It may appear in the form of trip matrix disaggregation for sub-region analysis, trip table generation based on traffic counts, or trip matrix balancing to match forecasts of zonal productions and attractions. The most widely used methods in each of these cases are directly or implicitly based on Maximum Entropy (ME) theory. This principle implies that in the absence of prior information, the probability distribution which best represents the current state of knowledge is the one with the largest entropy. In other words, if there are no preferences among various outcomes, then all the possible outcomes are assumed to have uniform probability (1).

ME-based formulations are too complex to be directly solved by optimization engines. Multiple iterative methods founded on ME theory have been developed to solve different forms of the trip distribution problem and have been widely used due to their ease of computation and fast convergence rate. However, the restrictive framework of the iterative procedures doesn’t allow an exhaustive use of the information in hand.

Given the significant increase in computational speeds and wide-spread availability of powerful open source optimization packages, the long-held “benefits” of ME-based methods no longer outweigh their possible errors in final trip distribution results. It is time to evaluate the impacts of using ME-based methods on final trip distribution results and potentially use alternative methods that make use of all the available accessory data sources. This area of research has received little attention to the best of the authors’ knowledge.

The goal of this paper is to investigate alternative methods by which to approach the trip distribution problem, with a focus on its application in freight transportation. Two alternative models are proposed to estimate or adjust trip tables in situations where ME-based methods may not be applicable or provide valid results. The proposed models can be easily handled by available optimization packages, include modal accessibility constraints, and maintain consistency with aggregate observed trip tables while also minimizing deviations from original trip shares when estimating balancing factors for future trip tables.

We believe that the theoretical arguments made in this paper and the practical significance of using the alternative models combined implies that these models could be suitable to be used in practice where appropriate.

In the next section, methods used to solve distribution models with a focus on ME methods are reviewed. This is followed by an examination of the impacts of ME assumptions on trip table estimation including example cases and presentation of proposed models. Finally, the California Statewide Freight Forecasting Model (CSFFM) is presented as an important application of the proposed methods.

LITERATURE REVIEW

In the transportation literature, the term distribution models generally refers to the methods used in the second step of the Four Step Model for travel forecasting, where the total number of trips beginning and ending in each zone is known and the goal is to find the number of trips between each pair of zones. In a matrix representation, the objective is to populate the cells of a matrix given row and column marginal values.

In this paper, this family of models is extended to include not only those estimated from marginal values of the matrix itself, but also those methods that use any other available information at any level for estimation, that is both estimating at coarser levels based on detailed
data, as in traffic OD estimation, or estimating at finer levels based on aggregate data as in
matrix disaggregation. Based on this definition, the “matrix disaggregation problem” and the
“OD estimation problem” can be categorized as distribution problems.

ME-based models have been the method of choice when working with limited
information to find the most likely solution amongst many alternates and have been widely used
in distribution models. These three defined problems deemed the “distribution problem” will be
presented in detail with an emphasis on the solution methods that use ME.

The first problem is the estimation of a trip table where forecast productions and
attractions and a base year trip table are available. Each zone may have two different growth
factors for in- and out-bound trips. To obtain the future trip table a number of iterative methods
have been proposed. The most well-known of these methods were developed by (and named
after) Furness (2) and Fratar (3). It was later shown that the Furness model can be derived from a
doubly constrained optimization problem using ME as the objective function.

Gravity models are one of the first theory-based methods to be used for trip distribution
modeling. Assuming that the total productions and consumptions by all the zones in the network
are known, gravity models distribute these marginal values in trip tables estimating the most
likely number of trips between each origin and destination, given known impedance values for
each OD pair. Wilson (4) derived the classic gravity model from the Lagrangian form of an
optimization problem using the ME objective function and production and attraction constraints.
A number of authors have attempted to formulate more general versions of the classical gravity
model. For example, Fang and Tsao (5) generalized the classical gravity model to a “self-
deterrent” gravity model by adding a “congestion term” to the exponent in the formulation of
gravity model. Additionally, De Grange et al. (6) included a term in the exponent to account for
accessibility to destinations.

Another theoretical model for trip distribution is the intervening opportunities model.
This model was initially proposed by Stouffer (7), but the theoretical formulation known today
was developed by Schneider (8). The basic idea behind the intervening opportunities model is
that destination choice is not explicitly related to distance, but rather to the accessibility to
opportunities that can satisfy the trip’s purpose. The main difference between the intervening
opportunities model and the gravity model is in the use of the distance variable. While in the
latter model distance is a continuous variable, it is treated as an ordinal variable in the former
model. The intervening opportunities model is not widely used by practitioners, due to the less
well known and more complicated theoretical basis, and lack of suitable software. Wilson (4)
also showed that the intervening opportunities model can be derived from an ME-based
formulation.

While aggregate distribution models are estimated at the zonal level, disaggregate
approaches are estimated at the individual household level. As opposed to aggregate approaches
where the output of the distribution model is the number of trips between zones, disaggregate
methods estimate the probability that an individual chooses a specific destination to satisfy the
need for an activity. Discrete choices models are usually used for the purpose of disaggregate
estimation of trip distribution (9). Wilson (10) also showed that multinomial logit models can be
derived from an ME-based optimization.

The next problem is matrix disaggregation. In transportation modeling, fine disaggregate
level data can be expensive or impractical to obtain, so many transportation models are
calibrated at coarser, aggregate levels and then disaggregated using local data. The limited
availability of disaggregate data is a more significant issue in freight transportation due to the
proprietary nature private sector data that is of interest in freight modeling. The Freight Analysis Framework (FAF) (11) is the only national publicly available commodity flow database. However, FAF zones are very aggregate (there are only 123 domestic regions in the last version of FAF).

Several attempts have been made to disaggregate FAF flows to county or sub-county levels. For instance, Alam and Majed (12) proposed a methodology to disaggregate the FAF Truck OD matrices to the county level using proportional allocation techniques based on land use or employment data at the county level and total truck VMT within that county to facilitate freight assignment on interstate and state highways. Opie et al. (13) developed several weighting schemes to disaggregate FAF flows proportionally. The recommended scheme was based on NAICS 3-digit employment (14). They also used a tri-proportional method to balance the final disaggregate matrix. Cambridge Systematics (15) and Viswanathan et al. (16) proposed similar methodologies to divide FAF2 regional commodity flow data for all commodities into county level flows by commodity. They developed robust linear regression equations to generate total production and attraction for each county. These regressions were used to guide the development of factors for each commodity for the disaggregation of freight flow productions and attractions.

ALTERNATIVE MODEL FOR THE TRIP DISTRIBUTION PROBLEM

Two alternative models are proposed to replace the ME-based models that serve as the basis of the three types of distribution problems introduced above. These alternative methods will be of interest especially when other data regarding capacity, constraints on transportation facilities, or aggregate trip matrices are available. However, since the case study in this paper is about trip distribution and disaggregation, these example models are presented accordingly.

Maximum Entropy (ME) Model for Matrix Balancing

The iterative process proposed by Furness (2), also known as the bi-proportional method, has been very popular and, in fact, is included in the majority of transportation software packages as the standard procedure by which to estimate balancing factors for trip distribution matrices. Similar to doubly constrained growth factor models, the bi-proportional method requires target production \((P_i \forall i \in Z)\) and target attraction \((A_j \forall j \in Z)\) estimates, and an initial trip table \((t_{ij})\), where \(Z\) is the set of zones in the networks.

The iterative solution process produces an approximate solution based on the Lagrangian of an optimization problem known as the “maximum entropy transportation model” or the “maximum entropy special interaction model”, presented in Eqs.1.1 to 1.4:

\[
\begin{align*}
\text{Min } & \quad G_{ME} = \Sigma_{ij} T_{ij} \left( \log \frac{t_{ij}}{t_{ij}} - 1 \right) & \text{Eq. 1.1} \\
\text{Subject to } & \quad \Sigma_j T_{ij} = P_i \forall i \in Z & \text{Eq. 1.2} \\
& \quad \Sigma_i T_{ij} = A_j \forall j \in Z & \text{Eq. 1.3} \\
& \quad T_{ij} \geq 0 \forall i,j \in Z & \text{Eq. 1.4}
\end{align*}
\]

Although this iterative fitting method has been used widely in transportation, in some cases there is merit in replacing it with other methods.
Most Similar Distribution (MSD) Model

In multi-regional transportation analysis, it is common to assume that the entire system is stable and any relationships, transactions, or flows among regions are in equilibrium. This concept can be applied to trip distribution by assuming that future trip shares are the same or close to present trip shares (the shares \([s_{ij}]\) for a trip table \([t_{ij}]\) can be obtained by normalizing the trip table to 1, i.e. by dividing the elements of the table by the sum of the table elements, \(t_{ij}\)). Under this formulation, instead of maximizing entropy as is the case for ME-based methods, the goal is to minimize the difference between the \([T'_{ij}]\) and \([t_{ij}]\) trip shares, that is, to generate the MSD given the production and attraction constraints. The objective function in MSD-based trip distribution problem can be formulated in different ways depending on the goals of the researcher and attributes of initial trip distribution. In the following section two formulations are presented and compared with several examples.

Sum of Squared Deviations MSD (SSD-MSD)

The first proposed formulation is to minimize sum of the squared deviations between the estimated future year and base year trip shares. The model formulated for trip distribution in freight and passenger models is presented in Eqs. 2.1 to 2.5.

\[
\text{Min } G_{SSD-MSD} = \sum_{ij} (\delta_{ij})^2 \quad \text{Eq. 2.1}
\]

Subject to:
\[
\sum_i T_{ij} = P_i \quad \forall i \in Z \quad \text{Eq. 2.2}
\]
\[
\sum_j T_{ij} = A_j \quad \forall j \in Z \quad \text{Eq. 2.3}
\]
\[
\delta_{ij} = \frac{T_{ij}'}{\sum_i T_{ij}'} - \frac{t_{ij}}{\sum_j t_{ij}} \quad \forall i, j \in Z \quad \text{Eq. 2.4}
\]
\[
T_{ij} \geq 0 \quad \forall i, j \in Z \quad \text{Eq. 2.5}
\]

The \(T'_{ij}\) terms in the above formulation are the decision variables and represent the trip table in the forecast year. Eqs. 2.2 and 2.3 ensure that the forecast year production and attraction constraints are satisfied. Clearly in order for the problem to be feasible, total production should be equal to total attraction (i.e. \(\sum_i P_i = \sum_j A_j\)). The \(\delta_{ij}\) variables measure the difference between the base and forecasted trip shares. The objective function of the SSD-MSD problem attempts to minimize sum of the squared deviations between the estimated future year and base year trip shares.

If constraints 2.3 and 2.4 allow, then the objective function (2.1) is minimized when the \(\delta_{ij}\) terms are all equal in value. So the same change in share of two cells in the matrix, say \(\delta_{ij}=0.01\), could have different implications depending on the base year shares. For example, if the base year share is 0.1, \(\delta_{ij}=0.01\) means that the forecast year share is 0.11. Similarly, for a base year share of 0.01, a \(\delta_{ij}=0.01\) means that the future year share is 0.02. In the first case, the share has changed from 0.1 in the base year to 0.11 in the forecast year, a change of only 10 percent. However, in the second case, the share has changed from 0.01 in the base year to 0.02 in the forecast year, a change of 100 percent. This is an advantage for regional analysis where zones with small volumes of flow between them are expected to have more fluctuation in volumes of flow in the future, compared to zones that already have strong and stable flow between them.

One advantage of having a simple mathematical form of an objective function is that the constraints of the problem are not restricted to production and consumption constraints. Any
capacity constraint that can be formulated as a linear function of decision variables or any
relationship between flows of different OD pairs can be easily added to the problem, without
having a considerable impact on the time complexity of reaching the optimal solution. Although
such constraints can also be added to more complicated non-linear objective functions like ME,
solving such non-linear problems via standard optimization packages is not as easy. Usually to
solve such non-linear problems, algorithms are developed based on the Lagrangian form of the
problem and specific optimization techniques such as column generation are used to solve the
resulting sub-problems.

In the next section, a simple numerical example is presented to compare SSD-MSD and
ME closely. The other possible and desired formulation for objective function will be discussed
later in the paper.

**Example 1: Comparison of SSD-MSD and ME** Assume that there exists one origin and two
destination zones in a network, and the initial value of each cell in the trip table is \((t_{11}, t_{12}) =
(0.5, 0.5)\), in units of hundred trips (half the production of zone 1 goes to zone 2 and the other
half is intra-zonal flow). Also, for simplicity, the only constraints applied are the non-negativity
constraints. The surfaces of the solution space of the SSD-MSD and ME models are displayed in
Figure 1.

![Figure 1](image)

**FIGURE1** Comparing the ME and SDS-MSD solution space

As shown in Figure 1(b), the second derivative of the SSD-MSD function is a constant,
and therefore \(G_{SSD-MSD}\) is a symmetric function and its curvature doesn’t depend on the value of
the decision variables, while the curvature of \(G_{ME}\) changes with the decision variables. Clearly in
this case both models are minimized at \((T_{11}, T_{12}) = (0.5, 0.5)\), returning the initial trip shares.

Expanding the example further, the effect of adding constraints to both optimization
problems will be analyzed next. The first constraint added to the problems is \(T_{11} + T_{12} \leq 0.5\).

Figures 2(a) and 2(b) show the contours of the two objective functions, \(G_{ME}\) and \(G_{SSD-MSD}\)
respectively, with the added constraint. Both problems are minimized at \((T_{11}, T_{12}) = (0.25, 0.25)\)
(which can be normalized to trip shares of (0.5,0.5), maintaining the original trip shares.

Now consider the initial scenario, but with different base year trip table elements
\((t_{11}, t_{12}) = (0.8, 0.2)\). In this scenario, the constraint \(T_{11} + 0.29 T_{12} \geq 1\) is added to both
problems.
In this case, the two problems do not return the same optimal solution. The dotted blue lines in Figures 2(c) and 2(d) show the initial trip rates, and the dotted red lines show the optimal solutions after adding the constraint. The optimal solution by the SSD-MSD model occurs at (1.92,0.44), which can be normalized to trip shares of (0.81,0.19), very close to the initial trip share (0.8,0.2). The ME model, however, changes the initial trip share from (0.8,0.2) to (0.57,0.43). In fact, the entropy function leans toward a uniform distribution of changes in number of trips.

---

**FIGURE 2 Example 1: contours of ME and SSD-MSD objective functions under different constraints**

**Minimax-MSD**

In SSD-MSD, the sum of the squared deviations from the initial trip shares is minimized. This implies that some cells may experience large deviations while other cells experience little or no deviation. A less discriminating application of the same principle is performed using the Minimax objective function. The objective function of this problem minimizes the maximum
deviation of the resulting trip shares from the base year trip shares in the form: \( \text{Min} \ Max |\delta_{ij}| \).

Since an absolute value cannot be handled directly by linear optimization software, mathematical transformations are needed. The transformed form of the problem is presented in Eq. 3.1 to 3.7.

\[
\begin{align*}
\text{Min} & \quad G_{G_{\text{Min} \text{Max, MSD}}} = G_{MM} \\
\text{Subject to} & \\
\Sigma_j T_{ij} &= P_i \quad \forall i \in Z \quad \text{Eq. 3.2} \\
\Sigma_i T_{ij} &= A_j \quad \forall j \in Z \quad \text{Eq. 3.3} \\
\delta_{ij} &= \frac{T_{ij}}{\Sigma_j T_{ij}} - \frac{t_{ij}}{\Sigma_i t_{ij}} \quad \forall i, j \in Z \quad \text{Eq. 3.4} \\
\delta_{ij} &= \delta_{ij}^+ - \delta_{ij}^- \quad \forall i, j \in Z \quad \text{Eq. 3.5} \\
G_{MM} &\geq \delta_{ij}^+ + \delta_{ij}^- \quad \forall i, j \in Z \quad \text{Eq. 3.6} \\
T_{ij}, \delta_{ij}^+, \delta_{ij}^- \geq 0 &\quad \forall i, j \in Z \quad \text{Eq. 3.7}
\end{align*}
\]

Eqs. 3.2 and 3.3 are the production and consumption constraints. Eq. 3.4 defines \( \delta_{ij} \) as the difference between the base year and forecast year trip shares. Eq. 3.5 sets \( \delta_{ij} \), to be the difference between two positive variables, \( \delta_{ij}^+ \) and \( \delta_{ij}^- \). This transformation is required to turn the objective function \( \text{Min} \ Max |\delta_{ij}| \) into the form \( x (\delta_{ij}^+ + \delta_{ij}^-) \). Eq. 3.6 is added to further simplify the objective function. The term \( \text{Max} (\delta_{ij}^+ + \delta_{ij}^-) \) in the objective function is replaced by \( G_{MM} \), and Eq. 3.6 is added to ensure that \( G_{MM} \) represents the max value of \( (\delta_{ij}^+ + \delta_{ij}^-) \)s over all the \( i \) and \( j \) indices. Lastly, Eq. 3.7 enforces the non-negativity constraints.

**Example 2: Application of MSD-based models in OD disaggregation** Assume that the aggregate trip table for a study area comprising two zones (Z1 and Z2) is available and that a disaggregate sub-regional analysis of the network is required. Assume Z1 is to be divided into three zones (z1, z2, and z3) and Z2 is to be divided into two zones (z4 and z5) as displayed in Figure 3(a). A trip table of the disaggregate network with five zones is available from previous sub-regional studies, and there is reason to believe that the trip shares have not changed in the intervening time period. Therefore, this trip table represents the base year trip table for an MSD-based optimization problem. The aggregate trip table for Z1 and Z2 and the base year disaggregate trip table for z1, z2, ... z5 are presented in Figures 3(b) and 3(c).

The goal is to readjust the disaggregate trip table such that it complies with the corresponding aggregate trip table, while disturbing the base year disaggregated trip shares as little as possible. Towards this end, the SSD-MSD problem is formulated in Eqs. 4.1 to 4.7.
3(a) Update the zoning system in the network

<table>
<thead>
<tr>
<th>Zone</th>
<th>1</th>
<th>2</th>
<th>Prod.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>Attr.</td>
<td>18</td>
<td>13</td>
<td>31</td>
</tr>
</tbody>
</table>

3(b) Observed aggregated trip table

<table>
<thead>
<tr>
<th>Zone</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Prod.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>19</td>
</tr>
<tr>
<td>Attr.</td>
<td>13</td>
<td>11</td>
<td>18</td>
<td>16</td>
<td>14</td>
<td>72</td>
</tr>
</tbody>
</table>

FIGURE 3. Matrix disaggregation example. Color coded aggregate and disaggregate matrices

\[
\text{Min } G_{SSD-MSD} = \sum_{i=1}^{5} \sum_{j=1}^{5} \delta_{ij}^2 \tag{4.1}
\]

Subject to:
\[
\sum_{i=1}^{3} \sum_{j=1}^{3} T_{ij} = 10 \tag{4.2}
\]
\[
\sum_{i=1}^{3} \sum_{j=4}^{5} T_{ij} = 7 \tag{4.3}
\]
\[
\sum_{i=4}^{5} \sum_{j=1}^{3} T_{ij} = 8 \tag{4.4}
\]
\[
\sum_{i=4}^{5} \sum_{j=4}^{5} T_{ij} = 6 \tag{4.5}
\]
\[
\frac{T_{ij}}{\sigma_Z} + \delta_{ij} = \frac{T_{ij}}{\sigma_Z}, \quad i, j = 1, 2, 3, 4, 5 \tag{4.6}
\]
\[
T_{ij} \geq 0, \quad i, j = 1, 2, 3, 4, 5 \tag{4.7}
\]

In Eq. 4.7, \( \sigma_Z = 31 \), the sum of the observed trips for Figure 3(b), and \( \sigma_Z = 72 \), the sum of the base trips in Figure 3(c).

The solution is presented in Figure 4(a). Figures 4(b) and 4(c) display the trip shares of the base and current years, respectively. \( \delta_{ij} \)'s, the difference between these two shares, are presented in Figure 4(d). The Minimax-MSD formulation for this problem includes constraint Eqs. 4.2 through 4.7 along with the additional constraints in Eqs. 5.2, 5.3, and 5.4 shown below.

\[
\text{Min } G_{\text{Minimax-MSD}} = Z \tag{5.1}
\]

Subject to:
\[
Z \geq \delta_{ij}^+ + \delta_{ij}^-; \quad i, j = 1, 2, 3, 4, 5 \tag{5.2}
\]
\[
\delta_{ij} = \delta_{ij}^+ - \delta_{ij}^-; \quad i, j = 1, 2, 3, 4, 5 \tag{5.3}
\]
\[
\delta_{ij}^+, \delta_{ij}^- \geq 0; \quad i, j = 1, 2, 3 \tag{5.4}
\]

Eqs. 4.2-4.7
The optimal values of $\delta_{ij}$ from solving problem 5 are presented in Figure 4(e). Figure 4 shows that the maximum change in the optimal trip shares generated by Minimax-MSD equals 0.0067, while the maximum change in trip share generated by the SSD-MSD is 0.074. This shows the primary merit of the Minimax-MSD objective function: all cells carry equal weight.

CASE STUDY: CALIFORNIA STATEWIDE FREIGHT FLOW MATRIX

The MSD model was used in the California Statewide Freight Forecasting Model (CSFFM) as the only viable method that can address all the modeling concerns of its developers. CSFFM is a commodity based interregional freight transportation model which focuses on rail and truck movements. It is based on the FAF3 public data source (11). The model covers 215 Freight Analysis Zones (FAZ) including 96 zones in California and 118 zones outside California in the United States. In less populated areas, FAZs are represented by counties and in the more developed central and southern parts of California, FAZs are sub-counties (Figure 5). There are also 15 commodity groups in the model. Details of data preparation and the commodity generation process are provided in Ranaiefar et al. (17).
CSFFM has a hybrid generation/distribution model. The goal of this model is to disaggregate FAF flows from five FAF regions in California to 97 FAZs for the base year and then to forecast flows for years 2020 and 2040 from base year estimations and scenario assumptions. Initially, in the total generation model the total production and attraction of each zone for all commodities including imports and exports were estimated simultaneously using Structural Equations Modeling (SEM) (17, 18). Next, a domestic commodity flow matrix was estimated by a SEMCOD Model (Structural Equations for Multi-Commodity OD Distribution). SEMCOD is a spatial interaction econometric model, also known in the transportation literature as a direct demand model, which combines trip generation and distribution steps in conventional four-step demand models. It estimates an initial flow matrix for the base year for all commodity groups simultaneously using SEM techniques. SEMCOD uses demographic, economic, land-use, and composite impedances between zones to estimate an initial 215 by 215 OD matrix for each commodity group. The difference between marginals of domestic flow matrices and total productions and attractions estimated by the total generation model is assumed to represent total exports and imports, respectively.

SEMCOD and total generation models are estimated independently. An additional procedure is needed to synchronize these models to (1) ensure consistency with the aggregate FAF domestic flow matrices and total imports and exports for FAF geographic regions; (2) ensure that marginals of domestic commodity flow matrices are less than total production and attraction estimated by the total generation model, and (3) ensure modal accessibility and capacity constraints are not violated in disaggregated flow matrices. A customized SSD-MSD model is used to achieve these goals.

FAF is a multi-dimensional commodity flow matrix (one matrix for each commodity group). These matrices are very sparse. On average 67% of cells are zero or the flow volume is...
less than 1 k-ton/year for each commodity group. Cells with very small flows are subject to
greater measurement error and sampling bias. Also intra-zonal flows account for more than 50%
of the total flows for each commodity group. Given these properties of FAF commodity flow
matrices and available information on modal facilities in each FAZ, using a Furness process to
estimate balancing factors is not suitable. The SSD-MSD model presented in Eq. 6 considers all
the constraints while minimizing the deviation from initial domestic commodity distributions
estimated by the SEMCOD model.

Other data sources used in this study included a detailed California rail carload waybill
sample (19), seaport data from the US Army Corps of Engineers waterborne database (20) and
land port data from the North American Transborder Freight Data (21).

\[
\begin{align*}
\text{Min} & \quad \sum_{i,j} B^2(\delta_{ij}^*)^2 + \sum_i (\delta_i^P)^2 + \sum_j (\delta_j^A)^2 \\
\text{Subject to:} & \\
\sum_j T_{ij} &= P_i - Exp_i \forall i \in z & \text{Eq. 6.2} \\
\sum_i T_{ij} &= A_j - Imp_j \forall i \in z & \text{Eq. 6.3} \\
\sum_{i \in I} \sum_{j \in J} T_{ij} &= F_{ij} \forall i, J \in Z & \text{Eq. 6.4} \\
\sum_i P_i &= Prod_i^F \forall i \in Z & \text{Eq. 6.5} \\
\sum_j P_j &= Attr_j^F \forall j \in Z & \text{Eq. 6.6} \\
\frac{\sum_{i,j} T_{ij}}{\sum_{i,j} F_{ij}} + \delta_{ij}^x &= \frac{T_{ij}}{\sum_{i,j} F_{ij}} \forall i, j \in z & \text{Eq. 6.7} \\
\frac{P_i}{\sum_i Prod_i^F} + \delta_i^P &= \frac{Prod_i}{\sum_i Prod_i} \forall i \in z & \text{Eq. 6.8} \\
\frac{A_j}{\sum_j Attr_j^F} + \delta_j^A &= \frac{Attr_j}{\sum_j Attr_j} \forall j \in z & \text{Eq. 6.9} \\
\sum_{i \in I} \sum_{j \in J} \gamma_{ij}^R T_{ij} &\geq F_{ij}^R \forall I, J \in Z & \text{Eq. 6.10} \\
\sum_{i \in I} \sum_{j \in J} \gamma_{ij}^W T_{ij} &\geq F_{ij}^W \forall I, J \in Z & \text{Eq. 6.11} \\
T_{ij} &\geq 0 \quad \forall i, j \in z & \text{Eq. 6.12} \\
P_i &\geq 0 \quad \forall i \in z & \text{Eq. 6.13} \\
A_j &\geq 0 \quad \forall j \in z & \text{Eq. 6.14}
\end{align*}
\]
The parameters and variables for the SSD-MSD model are explained in Table 1. Note that lower case \( i \) and \( j \) refer to FAZs whereas upper case \( I \) and \( J \) refer to FAF regions. In addition, “domestic” flows refer to flows inside the United States (excluding import and export) whereas “total” flows include foreign import and export. Variables \( \delta \) are auxiliary decision variables and are used to make the formulation more readable.

**TABLE 1** list of parameters and decision variables in problem 6

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Z )</td>
<td>Set of FAF zones</td>
</tr>
<tr>
<td>( z )</td>
<td>Set of FAZ’s</td>
</tr>
<tr>
<td>( B )</td>
<td>Number of FAZ’s (215)</td>
</tr>
<tr>
<td>( Exp_i )</td>
<td>Export for FAZ ( i, i \in z )</td>
</tr>
<tr>
<td>( Imp_i )</td>
<td>Import for FAZ ( i, i \in z )</td>
</tr>
<tr>
<td>( F_{ij} )</td>
<td>Domestic FAF flow between FAF zones ( I ) and ( J; I, J \in Z )</td>
</tr>
<tr>
<td>( Prod^F_i )</td>
<td>Total production for FAF zone ( I \in Z )</td>
</tr>
<tr>
<td>( Attr^F_i )</td>
<td>Total attraction for FAF zone ( I \in Z )</td>
</tr>
<tr>
<td>( Prod_i )</td>
<td>Total production for FAZ ( i; i \in Z ) obtained from the total generation model</td>
</tr>
<tr>
<td>( Attr_i )</td>
<td>Total attraction for FAZ ( i; i \in Z ) obtained from the total generation model</td>
</tr>
<tr>
<td>( t_{ij} )</td>
<td>Initial trip table between FAZs ( i ) and ( j; i, j \in Z ) obtained from SEMCOD</td>
</tr>
<tr>
<td>( y^R_{ij} )</td>
<td>Rail availability binary; ( i, j \in Z )</td>
</tr>
<tr>
<td>( F^R_{ij} )</td>
<td>Rail flow from FAF zone ( I ) to FAF zone ( J; I, J \in Z )</td>
</tr>
<tr>
<td>( y^W_{ij} )</td>
<td>Water mode availability binary; ( i, j \in Z )</td>
</tr>
<tr>
<td>( F^W_{ij} )</td>
<td>Water flow from FAF zone ( I ) to FAF zone ( J; I, J \in Z )</td>
</tr>
<tr>
<td>( s_{ij} )</td>
<td>Priori probability distribution; ( i, j \in Z )</td>
</tr>
</tbody>
</table>

For the optimization problem (Eq. 6) to be feasible, it is necessary for Eq. 6.15 to hold:

\[
\sum_{i \in Z} Prod^F_i - \sum_{i \in Z} Exp_i = \sum_{j \in Z} Attr^F_j - \sum_{j \in Z} Imp_j = \sum_{i,j} F_{ij}
\]  

Eq. 6.15
The objective function (Eq. 6.1) minimizes the sum of deviations from initial trip shares. There are three main initial shares involved in CSFFM. The first shares come from the domestic trips estimated by the SEMCOD model ($\delta_{ij}^Z$ and $T_{ij}$ are the variables related to this distribution). The second and third shares come from the total productions and attractions estimated by the total generation model. The total productions are represented by $P_i$ and $\delta_i^P$ terms and total attractions are represented by $A_j$ and $\delta_j^A$ terms. The first term in the objective function has a coefficient of $B^2$ for scaling purposes.

Eqs. 6.2 and 6.3 enforce equivalence of the domestic marginal values. Eq. 6.4 ensures that observed flows between FAF zones are equivalent at the disaggregate level. Eqs. 6.5 and 6.6 ensure that the observed FAF total production and attractions are equivalent. Eqs. 6.7, 6.8 and 6.9 generate auxiliary decision variables ($\delta$) that show the deviation of the optimal domestic, total production, and total attraction trip shares from their corresponding base trip shares.

In the new zoning system, not all the disaggregated zones (FAZs) contain rail yards, seaports, or both. Binary parameters $y^R_{ij}$ and $y^W_{ij}$ determine the availability of rail and water modes between zones $i$ and $j$ ($i, j \in z$), respectively. Eqs. 6.10 and 6.11 convey that the proportion of the flow between two FAF zones $I$ and $J$ that moves by the rail or water modes should go through FAZs $i \in I$ and $j \in J$ that have rail or water accessibility between them. Eqs. 6.12-6.14 are the non-negativity constrains.

Figures 6(a) and 6(d) compare the results of the SSD-MSD and SEMCOD models. The horizontal axes in these figures show the element IDs of the matrices. The element ID for the element $C_{mn}$ of a matrix can be obtained from Eq. 7.

$$ID_{mn} = \sum_{i=1}^{m-1} S_{j=1}^{i} + \sum_{i=1}^{n} 1$$

Eq. 7

The vertical axes of these two figures show the cumulative shares generated by each model. The shares [$S_{ij}$] for an OD matrix are obtained by dividing the elements of the matrix by the sum of the matrix elements so that values are normalized between zero and one. For an element $c_{mn}$ in the normalized matrix, the cumulative share can be obtained by:

$$CS_{mn} = \sum_{i=1}^{m-1} S_{j=1}^{i} s_{mn} + \sum_{i=1}^{n} s_{mn}$$

Eq. 8

**Evaluation of Results**

An element-by-element comparison of the cumulative shares of the cells in the base trip table given by the SEMCOD model and the final trip table estimated by SSD-MSD model provides a visual tool to assess the performance of the SSD-MSD model. Ideally, the two trip tables would be the same, and the two graphs corresponding to the cumulative shares of the base and final trip tables would coincide. However, this is not the case in practice due to the presence of constraints. Figures 6(a) and 6(b) show the cumulative shares given by SEMCOD model and those generated by the SSD-MSD. Since SSD-MSD should consider the modal accessibility constraints and consistency with FAF observed flows, the two graphs do not exactly coincide. Therefore there is a gap between them and the spikes in the graphs, which show the share of the corresponding elements, are not exactly similar. However, the results are more promising for commodity group five (6(a)) than commodity group one (6(d)).
FIGURE 6 Effectiveness of SSD-MSD in maintaining trip shares
Figures 6(b), 6(c), 6(e), and 6(f) compare the zonal production and attractions given by the total generation model and those generated by the SSD-MSD. The horizontal axis in these figures show the zone IDs (215 FAZs), and the vertical axis show the cumulative shares generated by the SSD-MSD and total generation models. Here, because there is only one constraint on zonal productions and attractions (matching FAF observations), the cumulative shares generated by the SSD-MSD and total generation models show a smaller gap, and follow a more similar pattern in spikes.

Figure 6 implies that SSD-MSD is in general successful in maintaining the initial domestic and total trip shares given by the SEMCOD and total generation models. As also shown in Example 1, MSD-Based Models outperform ME-based models in maintaining the original distribution. The main merit of SSD-MSD model is the possibility to consider all the modal accessibility constraints in generating the final trip table. In addition, the final disaggregated trip table is consistent with aggregate FAF flow tables.

CSFFM could not use iterative procedures such as bi-proportional fitting and still address the modal accessibility constraints and be consistent with FAF observed flows. Also the size of the problem prohibited solving an ME-based formulation directly with optimization engines. Therefore, MSD was the only viable solution for the purposes of this study. The contribution of the SSD-MSD in this study was not only in generating more desirable results, but in generating results that could at least satisfy the most basic constraints.

CONCLUSION

A new approach for balancing trip matrices with application in trip distribution and matrix disaggregation has been introduced. The concept of generating the most similar distribution (MSD) instead of Maximum Entropy (ME) principle is the basis of these models. The goal of MSD is to minimize the deviation from the initial trip distribution, considering existing constraints. Two MSD-based objective functions, SSD-MSD and Minimax-MSD were introduced.

ME-based formulations are too complex to be directly solved by optimization engines, and the restrictive framework of the iterative procedures used to solve them doesn’t allow an exhaustive use of the information in hand. The main advantage of replacing the entropy-based models with any of these new models where appropriate is that the resulting problems can accommodate additional constraints and still be easily solved by optimization engines. In addition, these models could produce more meaningful results than entropy-based models in regional transportation planning studies as demonstrated in Example 1 of this paper.

Given the improvements in computational speeds, the authors believe that the theoretical arguments made in this paper and the practical significance of using the alternative models combined implies that these models could be suitable to be used in practice where appropriate.

A case study of the California Statewide Freight Forecasting Model (CSFFM) was presented to show the merits of using MSD-based formulations. In this study balancing factors for 3 initial trip shares pertaining to the domestic flow, total production, and total consumption at a fine zonal level were estimated, while respecting the observed flows between the more aggregate FAF zones, and satisfying the modal accessibility constraints. These goals could not be addressed in conventional bi-proportional matrix fitting techniques such as the Furness method. In addition, the size of the problem prohibited solving an ME-based formulation directly with optimization software. Hence, the MSD-based models were determined to be the only viable options in this study. Some of the attractive features of using the SSD-MSD model for
CSFFM were demonstrated by generating and comparing graphs of the cumulative base and optimal shares.

ACKNOWLEDGEMENTS
The research reported in this paper was supported by the California Department of Transportation. The authors gratefully acknowledge the assistance provided by Doug MacIvor, Kalin Pacheco and Diane Jacobs of the California Department of Transportation and Iris You, Andre Tok, and Pedro Camargo from UCI. The contents of this paper reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California. This paper does not constitute a standard, specification, or regulation.

REFERENCES


