Hindawi
Journal of Advanced Transportation
Volume 2019, Article ID 5021026, 11 pages
https://doi.org/10.1155/2019/5021026

Research Article

Tour-Based Truck Demand Modeling with Entropy Maximization Using GPS Data

Soyoung Iris You [1] **and Stephen G. Ritchie** [2]

[1] *Innovative Transport Policy Division, Korea Railroad Research Institute, Uiwang 16105, Republic of Korea*
[2] *Department of Civil and Environmental Engineering and Institute of Transportation Studies, University of California, Irvine, CA 92697, USA*

Correspondence should be addressed to Soyoung (Iris) You; syyou@krri.re.kr

Received 3 June 2019; Accepted 9 October 2019; Published 16 December 2019

Guest Editor: Rongjie Yu

An approach is presented to determine the most likely tour distributions and model behavior for investigating drayage truck movements in a coastal region. This was done by implementing a revised form of entropy maximization based on truck tours to model and better understand drayage truck tour behavior at the San Pedro Bay Ports (SPBPs) complex in Southern California. The drayage trucks at the SPBPs have features that are distinct from other commercial trucks. The tour-based entropy maximization model proposed in this paper provides an opportunity to incorporate periodically updated GPS data collected in Southern California into a large-scale tour-based model. With the dataset, four models were estimated by cargo movement: (1) year-based, (2) low period, (3) medium period, and (4) high period models. The findings were consistent with the tour patterns varying by season and by cargo movement. Furthermore, the medium period, which represented relatively steady cargo movement, indicated a better MAPE (mean absolute percent error) than did other models. This proposed approach provides a significant advantage in that the most recent touring information obtained from advanced technologies could be directly applied to the tour-based model and subsequently used to assess various strategies.

1. Introduction

Freight truck movements are complex and distinct owing to the fact that their profits and logistics decisions are greatly affected by the increase in the number of trips. The extensive trip-chaining behavior of freight transportation cannot be represented without considering the dependencies among the trips. This is because these dependencies are heavily linked to the nature of the freight transportation utility and logistics decisions. This is the reason why there is no way to reflect a change in the origin of the following trip in the four-step model, which employs a trip-based approach. For example, let us assume that several trip destinations change as a result of changes to transportation policies. Consequently, analysis results using the trip-based approach represented by the four-step model would inaccurately show changes to a subset of trips that make up the trip chain. In the same vein, Ferdous et al. [1] argued that the tour-based model performed slightly better than the trip-based model in regional-level comparisons.

To properly capture the trip-chaining behavior of commercial vehicle movements, several tour-based model approaches have been proposed. Tour-based models have been developed using an optimization concept [2–5], set of rational discrete choice models [6, 7], phenomenological model [8], and activity-based model concept [4, 5, 9, 10]. Besides modeling, truck trajectory-tracking research has been attempted with various purposes to assess the impact of truck operations [11, 12]. Most tour-based models are built on the decision-making process behind vehicle operations or maximizing vehicle operational utilities. Insufficient data are available for models with large study boundaries because these models require abundant data in the form of detailed truck diaries that indicate the purpose and location of each truck stop. Furthermore, an enormous amount of computation time is typically required. For these reasons, elaborate analysis of tour-based models is difficult to duplicate for large-scale studies, such as freight models for metropolitan planning organizations (MPOs) and state agencies.

From another point of view, tour-based approaches at the network level [13–16] have been introduced to forecast urban freight movements based on trip-chaining characteristics and are more tractable because these require a smaller amount of data and provide faster computational time. Although many types of tour-based models were developed and practical application was attempted, none of the models were successfully implemented. To predict the effects of policy and planning on freight transportation and traffic flow, tour-based approaches need to be evaluated and updated with periodically observed data. Given the limited information of model tours, the theoretical framework of the proposed modeling in this study is based on an entropy-maximizing formulation. Our contribution lies in formulating a model that generates the most likely distribution of tours and represents behavior for investigating drayage truck movements in a coastal region. In other words, this study describes a tour-based entropy maximization formulation to estimate the flow of commercial vehicles on each tour using trip production/attraction by each node, sequentially visited nodes for each tour, and tour impedance from GPS data collected in the San Pedro Bay Ports (SPBPs) study area in Southern California. We carefully interpret the parameters corresponding to tour impedance and some of the major freight-related facilities. The insights and potential uses of a tour-based entropy maximization model of clean port trucks ("Clean trucks" (meeting 2007 model year emission standards) utilized public funds to replace older polluting drayage trucks at the SPBPs.) at the SPBPs generated from sensitivity analysis are also discussed. Finally, the capability of forecasting demand and converting the current four-step model for heavy-duty vehicles in Southern California to a tour-based model is evaluated and discussed.

This paper is organized as follows. First, we review a vehicle tour-based model, entropy maximization, and primal dual method for a convex optimization (PDCO) algorithm, which is followed by a description of SPBP clean truck GPS tour data. Then, the formulation of the tour-based entropy maximization model and calibration of the proposed model are presented. Finally, a potential application in freight-demand forecasting is discussed to conclude our paper.

2. Literature Review

In contrast with passenger transportation models, freight transportation models must address different types of complexities in behavioral and economic features. First, through freight transportation demand analysis, multiple factors such as volume, truck type, weight, length of trip chains, and number of trips have to be thoroughly considered. Second, opportunity costs for different commodities must also be considered. The difficulty is that such characteristics are not fully observable by transportation planners and modelers. Therefore, truck models play an essential role in linking commodity-based models and the need in MPOs to explain truck movements in detail. For example, in Southern California, a special generator model within the four-step approach was developed by the Southern California Association of Governments (SCAG) for heavy-duty vehicles and used by transportation planners for assistance in quick and rational decision making [17]. It is capable of capturing the number of empty and loaded truck trips directly in and out of ports; however, the remaining trips within a given tour are not sufficiently estimated.

Therefore, much effort has been invested into developing vehicle tour-based models. Hunt and Stefan [8] applied a random utility discrete choice model to a truck tour model using truck diary data with the "growing" tour construction approach. To be more specific, the tour-based microsimulation framework consists of six aspects: tour generation, purpose of vehicle and tour, tour start, purpose of next stop, next stop location, and stop duration. Each aspect of this framework is determined by logit choice models. Figliozzi [4] and Figliozzi et al. [5] used a VRP (vehicle routing problem) to analyze the impacts from congestion and technological changes. An interesting finding was that the percentage of empty trips does not influence the overall efficiency of the generated tours, which goes against the conventional wisdom that empty trips are symptomatic of sub-optimal and inefficient resource allocation [16]. Donnelly [3] proposed a commercial vehicle tour model by solving the traveling salesman problem of the empty backhaul. Wang and Holguín-Veras [18] also developed a hybrid micro-simulation modeling framework to generate goods-related vehicle tours that satisfy a known commodity flow O/D matrix in an urban freight network.

In contrast to the above-mentioned disaggregate approaches [3–9, 14–16, 18, 19], a few studies at the aggregate level have also been conducted. Maruyama and Harata [13] utilized a network equilibrium analysis that accounts for trip-chaining behavior. Although the model uses simple network examples, the issues and potential applications were not extended in forecasting freight demand. Wang and Holguín-Veras [20] adopted an entropy maximization method to develop an aggregate tour-based model for urban commercial vehicle movements in an inland city. By using truck travel diary survey data, the study demonstrated that observed tour sets are limited to a couple of days, and each truck has largely one tour each day. Using various factors related to businesses operations, this data from the sampled day were then expanded to the population. On the other hand, You and Ritchie [21] showed that coastal drayage trucks tended to have more than one tour per day, and many tours contained repetitive patterns that generate plenty of similar tours but not the exact same tour (see Figure 1). According to Ruan et al. [19], urban commercial vehicles and long-haul commercial vehicles differ in several aspects: (1) shipment of goods versus delivery of service, (2) distance traveled, (3) multiple stops, and (4) consolidated visits. However, drayage truck movements are distinct from both urban commercial vehicles and long-haul vehicles. In this study, a year of drayage truck GPS data provided a number of tour sets, which include enough pools and varieties of movement types to estimate future tour sets. The only exception to this was whether new attractions and production locations were newly introduced.

Entropy maximization is one of the theoretical bases for trip-based transportation demand and has been incorporated into urban and regional modeling by Wilson [22]. Entropy maximization is also known as minimum information theory because it mostly provides reliable estimates in the given

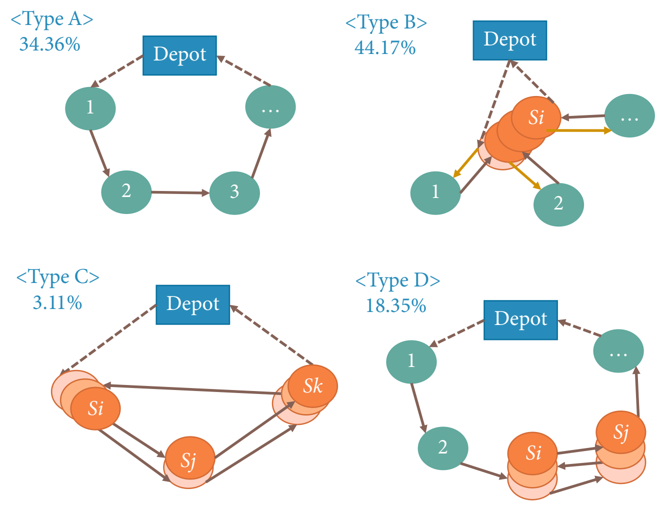

Figure 1: Drayage truck tour types, in which *Si*, *Sj*, and *Sk* are location indicators for the ports of Long Beach and Los Angeles, and near-dock and off-dock intermodal facilities ($i \neq j \neq k$), Source: You and Ritchie (2018).

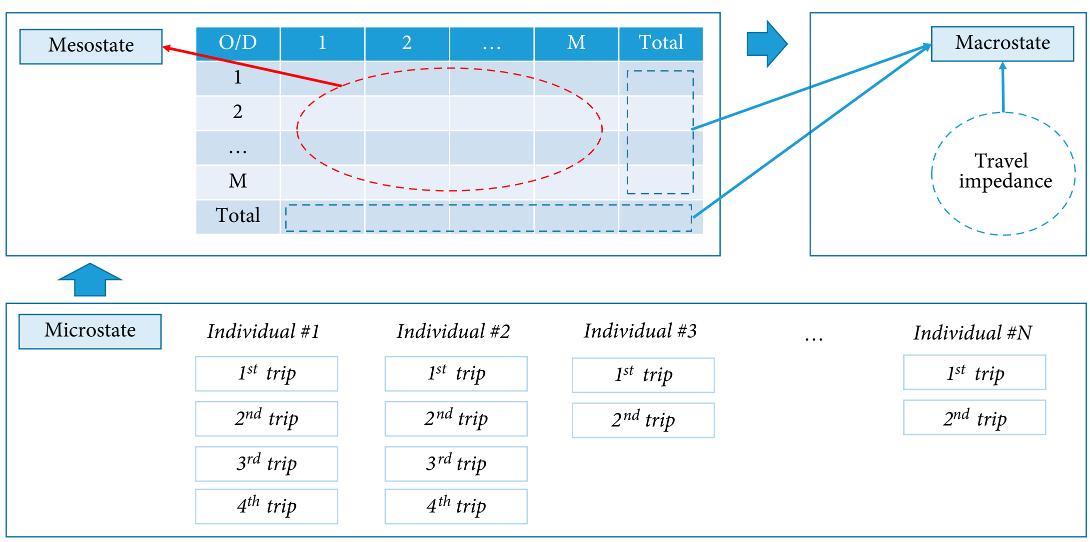

Figure 2: Original system of interest in entropy maximization.

system with limited information. Although the concept of entropy has been used for more than a century, the definition of entropy has often changed. Ludwig Boltzmann defined the measure of thermodynamic disorder in the 19th century, Claude Shannon altered the definition to include information uncertainty in the 20th century, and Jonathan M. Borwein revamped the definition to address barrier functions with superliner growth in the 21st century [23]. As shown in Figure 2, Wilson [22] identified three trip-related states for the "system of interest" as key factors: micro-, meso- (middle-level macro), and macro- (upper-level macro) states. A microstate is a set of individual trip information. A mesostate refers to a trip distribution matrix containing the number of trips between O/D pairs. A macrostate corresponds to an exogenously given total system energy, which is identified by the total number of trip attractions/productions and travel impedance. In entropy maximization, while each microstate is assumed to be equally probable, the most probable states would be elements corresponding to the greatest demand in the mesostate. Therefore, the entropy maximization problem is solved by finding "a state of the system", which is an assignment of individual trips to a trip distribution matrix in accordance with any macrostate constraint. Carrillo, Murillo and Liedtke [24] modeled the formation of colloidal structures and showed possible application to the case of an intermodal terminal in Germany.

After Wilson merged entropy maximization with the four-step planning model, successful application of these components has served as the basis for the entropy maximization method to become one of the most important transportation modeling theories. In the 1980's, count-based trip distribution estimation problems were widely investigated and formulated using the entropy maximization method. In terms of freight modeling, Wang and Holguín-Veras [20] utilized such an entropy maximization method by replacing the trip-based concept with a tour-based concept. In their tour-based entropy maximization method, when enough information is not available, the individual flow of commercial vehicles corresponded to any tour in the network and was expected to be equally probable by considering the constraints of the known aggregate information, namely, trip production by each node, trip attraction by each node, and travel impedance. Furthermore, this involved such factors as travel time and dwell time in each tour.

To solve the problem, the standard linear programming (LP) solution method could not be applied because the entropy maximization problem is a nonlinear problem (NP) with linear constraints. Furthermore, the linear constraints require the enumeration of all possible path flows between each O–D pair, which is computationally prohibitive for a network of realistic size. To solve the count-based trip distribution estimation problem using entropy maximization, Xie et al. [25] used the Frank-Wolfe algorithm for the entropy maximization, but only for a small-scale network. In recent studies, Lee and Fu [26] applied the entropy maximization model to population synthesis in an activity-based microsimulation model and used a quasi-Newton algorithm to solve a large number of dimensions. Boshnakov and Lambert-Lacroix [27] proposed a periodic Levinson-Durbin algorithm, the implementation of which was available with the R package. To solve the entropy maximization model, Li et al. [28] introduced a hybrid intelligent algorithm that assumed (1) the travel costs per unit that flows between different zones are fuzzy variables, and (2) trip productions and attractions are random variables. Many other investigations into the properties of entropy formulations and corresponding solution algorithms for convex problems have been made. Although Bregman's balancing method, a multiplicative algebraic reconstruction technique, and Newton's method are well known, interior methods are the most useful solution methods for large-scale entropy models and require very few primal-dual iterations, even with inexact search directions [29–34].

A primal-dual method for optimization programs with convex objectives (PDCO) is one interior method; however, there are several other methods such as (1) active set, (2) first and second order, (3) penalty, and (4) interior (Barrier) methods [35]. These methods have been widely used to solve inequality-constrained convex optimization problems and are generally based on applying the Newton method to a sequence of equality-constrained problems or to a sequence of modified versions of the Karush–Kuhn–Tucker (KKT) conditions [36–38]. By replacing nonnegativity constraints with equivalent barrier sub-problems, the lower and upper bounds for the decision variables can be set, which enables the specification of the decision variable to be in realistic feasible ranges. The PDCO algorithm has been studied by the System Optimization Laboratory (SOL) at Stanford University. SOL provides the MATLAB files to execute PDCO, which has been updated six times since it was released in 2002 and is adopted in this study [34].

2.1. Tour-Based Entropy Maximization Formulations. Before formulating a tour-based entropy maximization problem, the concept of a "node" needs to be clearly defined in this study. A "node" refers to a traffic analysis cell (TAC) or a stop in a trip/tour distribution. This concept should not be confused with the use of the term node in network problems in trip assignment. Second, it is important to note that the objective of the study is to forecast tour flows using the relationship between trips and tours when trip flows are known. To maintain a policy-sensitive advantage, many agencies have developed freight demand models based upon a commodity-based concept and, in terms of vehicle flows, often end up with a trip-based model. This study excludes the extra step of forecasting tour flows that rely on continuously observed tours and generate tour-based models without expensive or time-consuming travel diary surveys. Instead, it uses currently developed trip-based models and GPS data, both of which are relatively easy to collect as input data. It is also required to define the concepts of tour-related terms and understand the nature of clean drayage truck movements. As shown in Figure 3, a tour is defined as the sequence of stops (or visits) visited by a truck. In general, each drayage truck starts and ends at its depot. A tour consists of multiple trips, which is an individually directed vehicle movement connecting two consecutive stops.

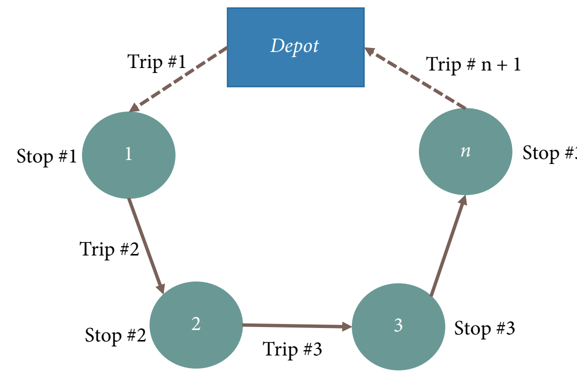

Figure 3: Definition of tour components.

As mentioned earlier, defining intermediate stops in a tour using trip attraction/production constraints in conventional entropy maximization would cause the tour to lose the sequential characteristics of trip chaining. This is because stop sequences are maintained by the travel impedance variables of the microstates. However, owing to the aggregation of individual commercial vehicle tour records to trip attractions/productions, the multiple trip travel impedances are summed into one travel impedance per tour. To capture the realistic and complex structure of each tour, the conventional tour-based entropy maximization model must be expanded to be able to capture sequentially visited nodes. Because a trip as a component of each tour can be explained in a straightforward manner with directional information and the concept of origins and destinations, the conventional entropy maximization model is sufficient to be used for the trip-based distribution model or tour-based model with relatively simplified tour sets throughout sample expansion.

Considering that tours of drayage trucks are unique, the total impedance consists of tour travel time and waiting/transaction time. In particular, waiting/transaction time, which is not considered in trip-based analysis, is as essential as travel time, because a drayage truck is heavily involved in the

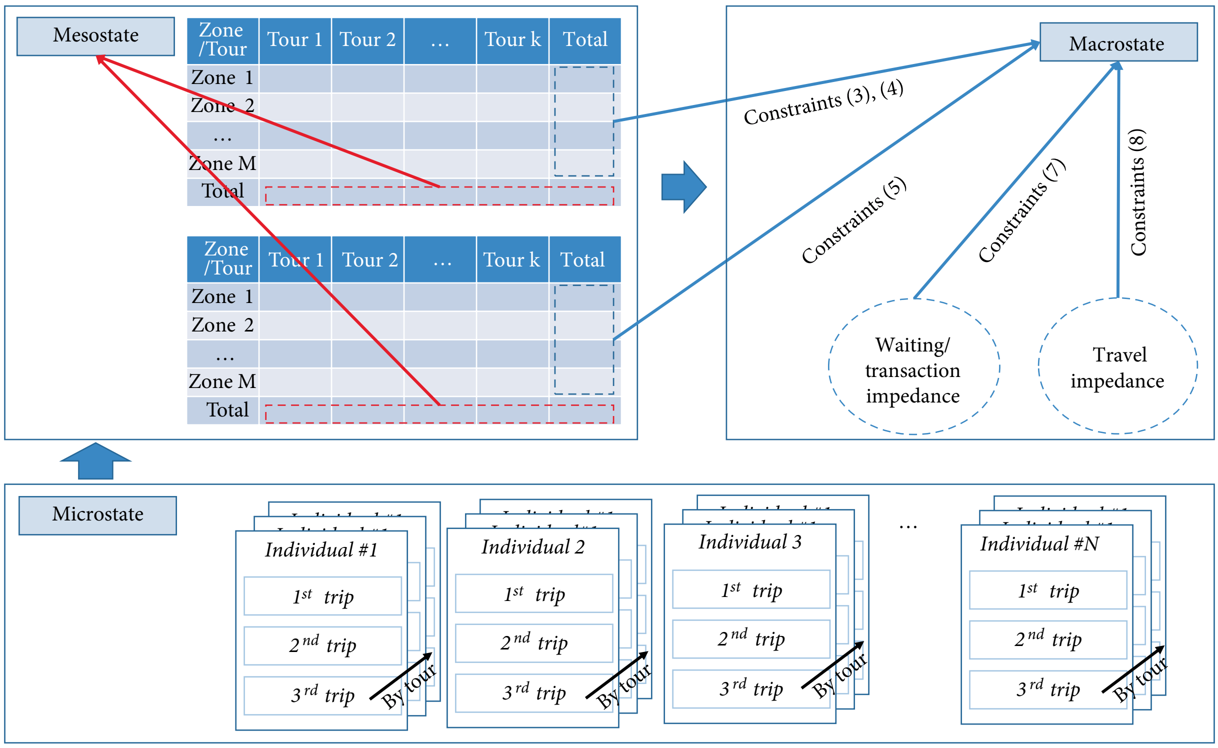

Figure 4: Coastal truck tour-based entropy maximization formulation.

transloading process at each stop. The waiting/transaction time impedance needs to be considered as a separate constraint from tour travel time. The travel time and waiting/transaction time impedances are based on the average of the GPS observed times. The advantage of using the observed time is to capture the underlying conditions, such as continuously congested routes.

The variables for each state used in the proposed model are summarized as follows. The microstate and macrostate variables are fed into the tour-based entropy model to find the best feasible solution at the mesostate, i.e., number of port drayage truck tour flows for a given tour *j*. To account for sequentially visited nodes in each tour, we include both the connectivity of nodes and a data structure, with tours and corresponding trips defined by sequential stops (see Figure 4).

With these assumptions, we revised the mathematical formulation and KKT conditions of the entropy maximization formulations to solve the tour distribution problem of Southern California clean trucks. It should be noted that the drayage truck movements are not mutually exclusive of other commercial vehicles but include the basic tour patterns. Therefore, a revised entropy maximization formulation based on tours is expected to perform well with all types of commercial vehicles.

2.2. Tour-Based Revision of the Entropy Maximization Formulation. The coastal truck tour-based entropy maximization formulation contains one objective function and five constraints, as shown in Equation (1) below.

$$\text{Max}\, W = C_X^{x_1} \cdot C_X^{x_2} \cdots C_X^{x_j} = \frac{X!}{\prod_{j=1}^{J} x_j!}, \quad x_j \geq 0, \quad \forall j \in \{1,2,3,\ldots,J\}. \tag{1}$$

Following Wilson (1970), the objective function can be simplified. Taking the natural logarithm and using the Stirling approximation ($\log x! = x \ln x - x$), the objective function becomes:

$$\text{Min}\, z = \sum_{j=1}^{J} \left(x_j \ln x_j - x_j\right), \quad x_j \geq 0, \quad \forall j \in \{1,2,3,\ldots,J\}. \tag{2}$$

Subject to

$$\sum_{j=1}^{J} a_{ij} x_j = O_i, \quad \forall i \in \{1,2,3,\ldots,N\}, \tag{3}$$

$$\sum_{j=1}^{J} b_{ij} x_j = D_i, \quad \forall i \in \{1,2,3,\ldots,N\}, \tag{4}$$

$$\sum_{j=1}^{J} l_{kj} x_j = L_k, \quad \forall i \in \{1,2,3,\ldots,K\}, \tag{5}$$

$$\sum_{j=1}^{J} c_{Tj} x_j = C_T, \tag{6}$$

$$\sum_{j=1}^{J} c_{Hj} x_j = C_H, \tag{7}$$

where x_j: flow on tour *j*, i.e., number of trucks operating on that tour in the analysis period. a_{ij} (or b_{ij}): Number of trips starting from node *i* (or ending at node *i*) within a given tour *j*. It is one if node *i* is a start node (or an end node) only once in the tour and zero if node *i* is not in the tour. It is more than one if node

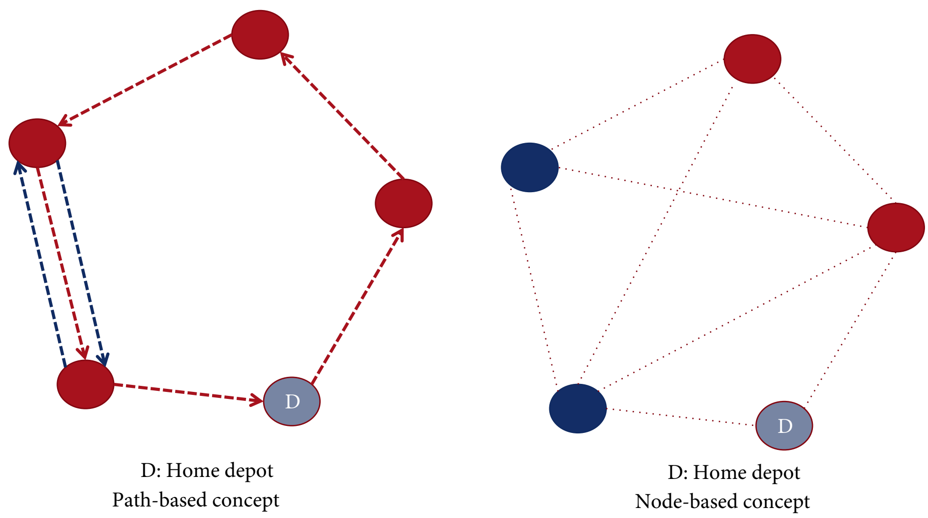

Figure 5: Path-based and node-based concepts.

i is visited more than once in tour j. a_{ij} (or b_{ij}) = 0. O_i (or D_i): total number of departures (or arrivals) at each node i. l_{kj}: the number of times that an O–D pair k is included in a given tour j (since any tour is made up of trips between O–D pairs as links of a trip chain, and repeated trips are possible between an O–D pair). L_k: Total number of truck trips between an O–D pair k. c_{Tj} (or c_{Hj}): the impedance on tour j, which is the travel time (or the handling time). C_T (or C_H): total impedance, which is the overall travel time (or the handling time) in the system.

Equations (1) and (2) of the coastal truck tour-based entropy formulation indicate the objective function to find the most feasible ways to distribute tours by maximizing "the system of interest". Constraint sets (3) and (4) are equality constraints between the total counts of the tour flows passing a node and the total number of trip attractions/productions at the corresponding nodes. When each tour starts from the depot and returns to the same location (closed tour), we can rely on either of the constraint sets (3) or (4) because they are the same for this case; otherwise, such as in the case of an open tour in Ref [18], we should simultaneously consider constraint sets (3) and (4). Constraint set (5) is added to maintain the trip sequences in a tour (or directional information of intermediate stops) and play a role in converting the trips into tours. This indicates that the numbers in the trip distribution table equal the summation of the corresponding trips in each tour. The role of this constraint is to differentiate between, for example, a tour with sequentially visited nodes $A \to B \to C \to D$ and another tour $A \to C \to B \to D$ (path-based concept, see Figure 5). Such tour details were not considered in the original tour-based entropy maximization model (node-based concept, see Figure 5). Although a tour is defined as a node sequence, this concept could be easily annulled in a structure of repeated tours. Therefore, the tour-based model with a node-based concept eventually loses directional information. Constraint sets (6) and (7) are the tour travel and waiting/transaction impedances.

2.3. Karush–Kuhn–Tucker (KKT) Conditions of the Formulations. To understand the characteristics of the entropy maximization formulation and find the optimal solutions, the first-order conditions (or Karush–Kuhn–Tucker (KKT) conditions) and the second-order conditions (or Hessian matrix) are obtained for the two formulations, separately. The formulation of the first-order condition is the Lagrange function shown below in Equation (8).

$$\begin{aligned} L(x, \lambda, \beta) = {} & \sum_{j=1}^{J} \left(x_j \ln x_j - x_j \right) + \sum_{i=1}^{N} \lambda_i^1 \left(O_i - \sum_{j=1}^{J} a_{ij} x_j \right) \\ & + \sum_{i=1}^{N} \lambda_i^2 \left(D_i - \sum_{j=1}^{J} b_{ij} x_j \right) + \sum_{k=1}^{K} \lambda_k^3 \left(L_k - \sum_{j=1}^{J} l_{kj} x_j \right) \\ & + \beta_1 \left(C_T - \sum_{j=1}^{J} c_{Tj} x_j \right) + \beta_2 \left(C_H - \sum_{j=1}^{J} c_{Hj} x_j \right), \end{aligned} \tag{8}$$

where λ_i: Lagrange multiplier associated with the i-th node production constraint; β_i: Lagrange multiplier associated with the total impedance constraint.

The partial derivative of the Lagrange functions with respect to the number of tours x_j and the necessary conditions for x^*, λ^*, β^* to be optimal solutions of the model can be written as follows (Equations (9)–(13)):

$$\begin{aligned} \frac{\partial L(x^*, \lambda^*, \beta^*)}{\partial x_j} = {} & \ln x_j - \sum_{i=1}^{N} \lambda_i^1 a_{ij} - \sum_{i=1}^{N} \lambda_i^2 b_{ij} \\ & - \sum_{k=1}^{K} \lambda_k^3 l_{kj} - \beta_1 c_{Tj} - \beta_2 c_{Hj}, \\ & \forall j \in \{1, 2, \ldots, J\}, \ \forall k \in \{1, 2, \ldots, K\}. \end{aligned} \tag{9}$$

$$x_j^* \frac{\partial L(x^*, \lambda^*, \beta^*)}{\partial x_j} = 0, \qquad \forall j \in \{1, 2, \ldots, J\}. \tag{10}$$

$$\frac{\partial L(x^*, \lambda^*, \beta^*)}{\partial x_j} \geq 0, \qquad \forall j \in \{1, 2, \ldots, J\}. \tag{11}$$

$$\frac{\partial L(x^*, \lambda^*, \beta^*)}{\partial \lambda_i} = 0, \qquad \forall i \in \{1, 2, \ldots, N\}. \tag{12}$$

$$\frac{\partial L(x^*, \lambda^*, \beta^*)}{\partial \beta} = 0. \tag{13}$$

From Equations (9)–(13), the optimal solution can be rewritten as Equation (14). This addresses the number of tours, which is the products from an exponential function of Lagrange multipliers associated with trip attraction/production at each node, trip distribution, travel time, and waiting/transaction time impedance along each tour.

$$x_j^* = \exp\left(\sum_{i=1}^{N} \lambda_i^1 a_{ij} + \sum_{i=1}^{N} \lambda_i^2 b_{ij} + \sum_{k=1}^{K} \lambda_k^3 l_{kj} + \beta_1 c_{Tj} + \beta_2 c_{Hj}\right). \tag{14}$$

The Hessian of the objective function indicates whether the formulation is convex. We conclude that the objective function is convex because the following second-order derivatives are positive definite, whereas the constraints are linear. Therefore, the proposed formulation has a unique optimal set of solutions.

$$\frac{\partial Z^2(x)}{\partial x_{j_1} \partial x_{j_2}} = \begin{matrix} \dfrac{1}{x_j} & \text{for } j_1 = j_2 \\ 0 & \text{otherwise} \end{matrix} \quad \forall j \in \{1, 2, ..., J\}. \tag{15}$$

3. Case Study: Coastal Drayage Trucks in Southern California

3.1. Description of the SPBP Clean Truck GPS Tour Data. As shown in Figure 6, the SPBP is located in Southern California. Our tour data were generated from the SPBP clean truck GPS data using an analysis tool described in Ref. [18]. The framework of GPS data processing consists of 8 steps: (1) selecting all potential O–D stops, (2) identifying truck depots by selecting the greatest major cluster of the last stops of the day, (3) geocoding O–D stops and truck depots, (4) identifying closed/open tours, (5) deleting false-positive stops, (6) condensing pairs of intra-zonal trips caused by transaction and queuing, (7) imposing new tours with a three-hour stop duration before the peak tours, and (8) deleting abnormal pairs of tours/trips. Basically, each coordinate of the GPS trajectory data has been mapped with the road network by minimizing the distances between each coordinate and the set of the arc.

According to our data, most tours of clean trucks at the SPBPs are completed within one day, and one day of travel behavior is not necessarily representative of any other day. The tour data for nearly all of 2010 were chosen for this study for several reasons: (1) to capture many different types of tours, (2) to calibrate and validate the forecasting model, and (3) to compare cargo movement by season. Based on tour distances, clean truck tours can be grouped into three modes: (1) short-haul drayage, (2) local drayage, and (3) regional drayage. The short-haul drayage operation involves very short container movements from two to six miles in length. These short movements consist of cargo movements between the port terminals and the Intermodal Container Transfer Facility (ICTF), which is a near-dock rail terminal, or nearby container yards. Local drayage consists of moves from ports to highly concentrated warehouse and terminals areas, and a major rail yard within 20 miles of the ports. According to the drayage truck origin and destination surveys, approximately 50%–60% port drayage truck activity is captured in this range. Finally, regional drayage includes container moves at distances greater than 20 miles from the ports.

As mentioned earlier in this paper, the SPBP clean trucks exhibit distinct tour behaviors. Drayage truck travel behavior from our GPS tour data in Southern California is relatively complicated because tours contain similarity in terms of stop locations, but the sequence and the number of visits are not exactly the same. As in Ref. [18], approximately 65% of observed tours contain repetitive patterns in a tour because drayage trucks often visit the same stops more than once in a tour. For example, they present repetitive patterns between out-of-depot and returning-to-depot activities in each tour, such as traveling multiple times between the ports and near-dock rail yards.

The other common tour type is to drop by freight facilities before/after visiting the ports at least once. Without defining the sequential information of the intermediate stops in each tour, the estimated tours would not enumerate the existing tours properly.

3.2. Tour-Based Model Estimation and Calibration. In this section, we applied the revised tour-based model to a case study and describe the results. To calibrate the model, 545 clean drayage truck GPS data collected by the SPBP authorities were selected representing 7% of in-service clean trucks, which travel all over California with their base at the SPBPs. For the year 2010, the subject trucks visited a total of 1896 traffic analysis cells (TACs) in the area. These TACs are smaller traffic analysis areas than TAZs.

The GPS tour data were analyzed using the analysis tool developed in Ref. [18]. To avoid the inconsistency of weekday and weekend tour patterns, we focused on tour data collected from weekdays. Approximately 33,000 different tours were identified, representing a total of 83,694 tours observed for the study year. The size of the tour-based entropy maximization problem comes to approximately 57,000 × 33,000.

From the selected year's data, we set up four analysis scenarios: (1) one-year period, (2) extremely busy cargo moving period, (3) busy period, and (4) least busy period. According to the annual emissions inventory report (2010) and You and Ritchie (2012), the period and the corresponding months are defined as follows: (1) low period: January, February, March, and April, (2) medium period: May, November, and December, and (3) high period: June, July, August, September, and October.

As discussed earlier, the PDCO algorithm was used for the tour-based model calibration for three reasons. First, this algorithm is known to be very efficient in solving large-scale entropy maximization problems. Second, it allows the setting of lower and upper bounds for the decision variables, which allows for the specification of feasible ranges. Third, fully-tested MATLAB codes for PDCO algorithm are available, and it is straightforward to apply our entropy maximization problem into the offered codes by generating A and b matrices, as shown in Figure 7.

After calibration using the PDCO algorithm, we evaluate the accuracy of the model. In transportation modeling, the commonly used performance metrics are the mean square error (MSE), mean absolute deviation (MAD), mean percent

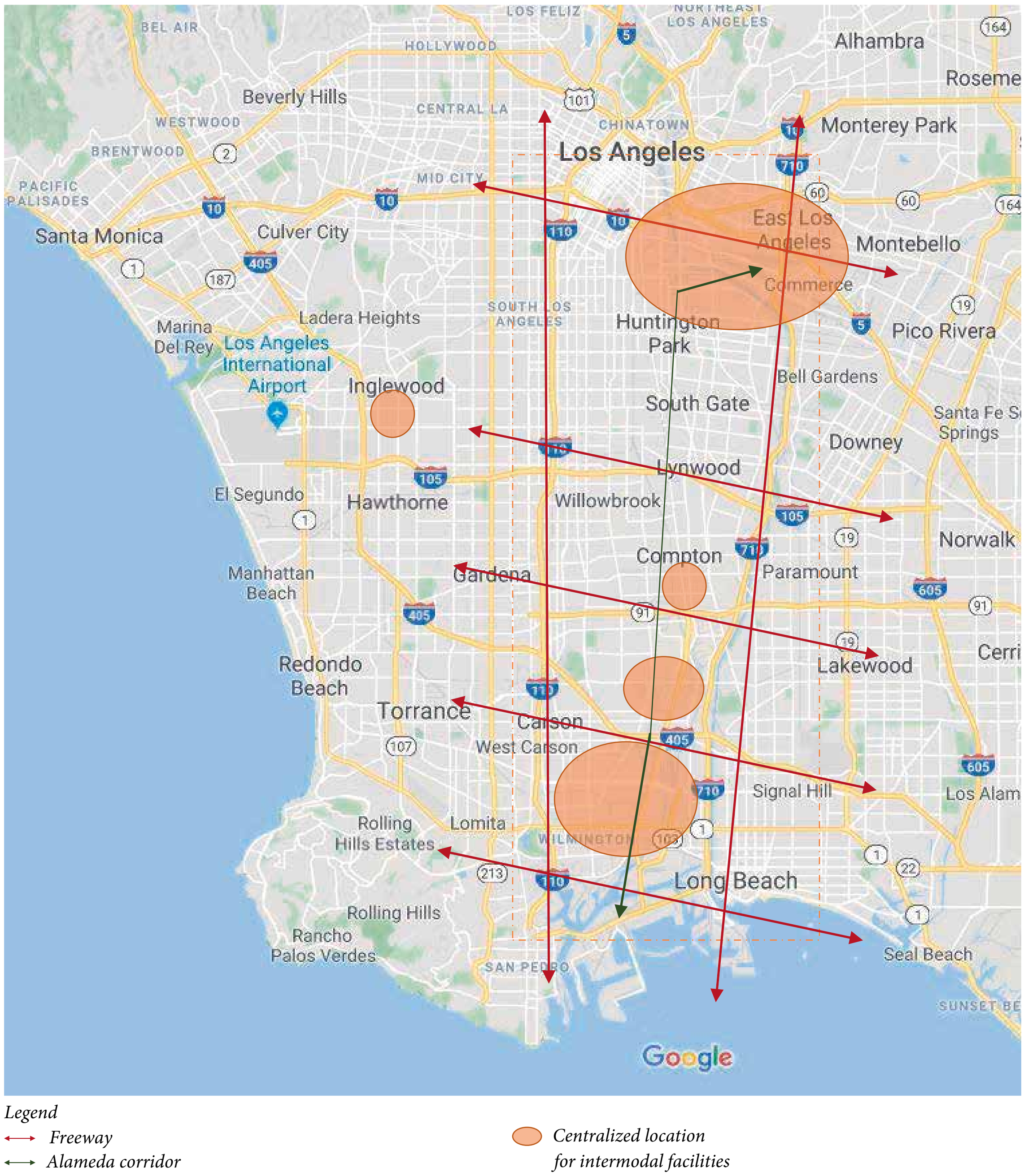

FIGURE 6: Case study site: ports of Long Beach and Los Angeles in Southern California.

error (MPE), and mean absolute percent error (MAPE). Much of the literature has demonstrated the need to use absolute percent error as a basis for comparison to eliminate the effect of the variability observed in most transportation data sets. Therefore, we use the MAPE, and the results indicate a good match between the observed and estimated tour flows. The MAPE can be calculated with Equation (16).

$$MAPE(\%) = \frac{\sum_{i=1}^{n} |PE_i|}{n} = \frac{\sum_{i=1}^{n} |(X_i - F_i)/X_i \times 100(\%)|}{n}, \quad (16)$$

where PE_i: Percentage error. X_i: Observed tour flows on the i-th tour. F_i: Estimated tour flows on the i-th tour. n: Numbers of the different tours that are identified.

The resulting estimated tours are compared to the observed tours for tour travel time, tour transaction time, and total tour time. Frequency distributions are shown in Figure 8. Estimated tours show MAPE values between 2.46% and 3.39%.

As shown in Table 1, the MAPE of the estimated tour flows resulting from the revised tour-based entropy maximization model (with directional constraints) for a year's worth of data is 13.53%. Considering that their tour-based model performed well with the other commercial truck tours in the Denver metropolitan area, drayage truck tours are more sensitive to sequential visits.

Besides the year-based analysis, three other models based on cargo movements are calibrated: (1) low period:less busy months, (2) medium period: busy months, and (3) high period:

Table 1: Estimated results of coastal truck tour-based entropy maximization model.

Estimated result	Year	Low period[2]	Medium period[3]	High period[4]
Measures of accuracy: MAPE[1] with directional constraints (our proposed model)	13.53%	6.19%	4.06%	8.12%
Number of unique tours (total number of tours)	33,148 (83,694)	10,123 (24,224)	9,368 (19,660)	17,431 (39,810)
Tour-travel time related lagrange multipliers (β_1)	−0.5273	−0.4851	−0.3606	−0.5244
Tour-transaction time lagrange multipliers (β_2)	−0.1751	−0.1232	−0.1489	−0.1676

[1]Mean absolute percentage error (MAPE). [2] January, February, March, and April. [3] May, November, and December. [4] June, July, August, September, and October.

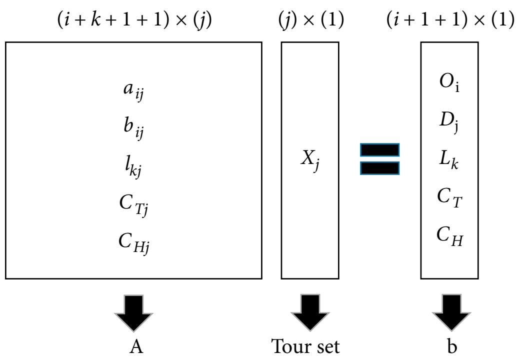

Figure 7: Illustration of A and b matrix in the PDCO algorithm.

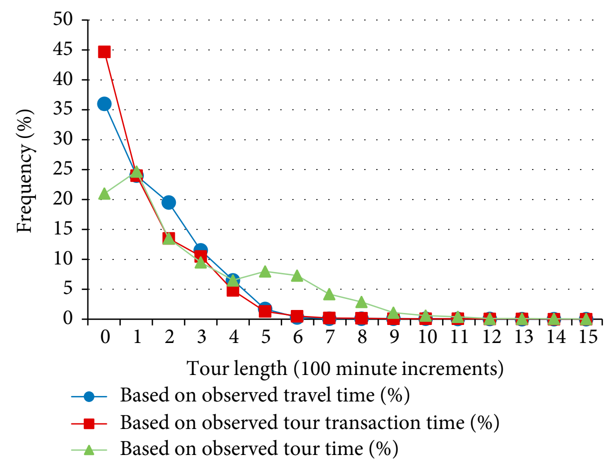

Figure 8: Observed tour length distributions for a year.

busier months. The MAPEs range from 4.06% to 8.12% for the different scenarios, which are better than those for the year-based model. We found even better performance for the medium period than the other periods. This is because the cargo movements for the medium period are steady, while those of the other periods tend to be either increasing or decreasing.

The impedance-related Lagrange multipliers associated with the tour time are more negative, as expected. This indicates that the longer the tour travel time becomes; the less likely tour flows will be made on that tour. Similarly, the tour transaction related Lagrange multipliers are negative as well. As shown in the table above, although the ratio of tour-transaction-time-related Lagrange multipliers do not dramatically change when compared to those of tour-travel time, the tour would be more likely to visit fast-processing freight facilities. In addition, the trip-generation/attraction-related Lagrange multipliers are zone-specific and could be either positive or negative. However, there are distinct patterns observed in the ports of Long Beach and Los Angeles, ICTF, BNSF, and Commerce, which represent special truck trip generation zones. The zone-specific Lagrange multipliers in the major intermodal facilities are positive.

As we know, the proposed model was estimated by several variables. When it comes to comparing periods, the ratio of Lagrange multipliers from different variables is a more proper indicator of how sensitive the impact of variables on each period is.

4. Conclusions

The objective of this paper was to introduce a revised form of entropy maximization based on truck tours to model and better understand drayage truck tour behavior at the SPBPs in Southern California. The drayage trucks at the SPBPs have features distinct from the other commercial trucks. Such distinct features cannot be captured well by the conventional trip-based four-step planning models and conventional tour-based entropy maximization models. Even with a disaggregate tour-based model, it is difficult to evaluate drayage truck behavior, particularly for a large-scale study.

To overcome such obstacles, the main innovations of this study are as follows: First, the coastal truck tour-based entropy maximization model proposed in this paper provides an opportunity to incorporate periodically updated GPS data collected in Southern California into a large-scale tour-based model. This method provides a significant advantage in that the most recent tour information from advanced technologies could be applied directly to the tour-based model to be subsequently used for assessing various strategies. Second, the coastal truck tour-based entropy maximization model is capable of addressing common drayage truck behavior and repetitive patterns. We also found that the parameter corresponding to the wait/transaction time was negative, whereas the inland truck tour model concluded that, for the Denver metropolitan area, it was positive, signifying that longer handling time induced more travel. However, their interpretation could not be supported by the Southern California drayage trucking data because, for example, the ports of Long Beach and Los Angeles have continuously tried to reduce wait/transaction time to accommodate more cargo movement. With the dataset, four models were estimated: (1) year-based, (2) low period, (3) medium period, and (4) high period models by cargo movement. The findings were consistent with the findings in [20] in that the tour patterns varied by season and by cargo movement. Furthermore, the medium period, which represented relatively steady cargo movement, exhibited a better MAPE than those of the other models. Finally, the coastal truck tour-based entropy maximization model with KKT could find the optimal solution in an efficient way, saving exhaustive calculation and effort.

Along with these findings and insights regarding clean trucks at the SPBPs, two possible applications for freight-demand forecasting were discussed, and a numerical test was conducted. The first application converted an existing trip-based forecasting model to the tour-based model forecasting for a future year by assuming that tour sets and travel impedance were the same as the current year. Moreover, the second application utilized tour-based forecasting for assessing policies. Although the numerical studies show potential for use in freight demand forecasting and strategy evaluation, the proposed model may require further studies before being applied to improve the current strategies for policy evaluation.

Data Availability

All data included in this study are available upon request by contact with the author Soyoung Iris You (syyou@krri.re.kr).

Conflicts of Interest

The authors declare that there is no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by a grant from R&D Program of the Korea Railroad Research Institute (KRRI), Republic of Korea.

References

[1] N. Ferdous, C. R. Bhat, L. Vana et al., “Comparison of the four-step versus tour-based models in the context of predicting travel behavior before and after transportation system changes,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2303, no. 1, pp. 46–60, 2012.

[2] W. Wisetjindawat, K. Sano, and S. Matsumoto, “Commodity distribution model incorporating spatial interactions for urban freight movement,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1966, no. 1, pp. 41–50, 2006.

[3] R. Donnelly, “A hybrid microsimulation model of freight flows,” *Innovations in City Logistics*, Nova Science Publishers Inc, Hauppauge, NY, USA, 2008.

[4] M. A. Figliozzi, H. S. Mahmassani, and P. Jaillet, “Pricing in dynamic vehicle routing problems,” in *Transportation Science*, E. Taniguchi and R. G. Thompson, Eds., pp. 249–259, vol. 41, no. 3, pp. 302–318, 2007.

[5] M. A. Figliozzi, “Modeling the impact of technological changes on urban commercial trips by commercial activity routing type,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1964, no. 1, pp. 118–126, 2006.

[6] K. J. Stefan, J. D. P. McMillan, and J. D. Hunt, “Urban commercial vehicle movement model for calgary, alberta, Canada,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1921, pp. 1–10, 2005.

[7] J. Gliebe, O. Cohen, and J. D. Hunt, “A dynamic choice model of urban commercial vehicle and person activity pattern,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2003, no. 1, pp. 17–26, 2007.

[8] J. D. Hunt and K. J. Stefan, “Tour-based microsimulation of urban commercial movements,” *Transportation Research Part B*, vol. 41, no. 9, pp. 981–1013, 2007.

[9] S. I. You, J. Y. Chow, and S. G. Ritchie, “Inverse vehicle routing activity-based urban freight forecast modeling and city logistics,” *Transportmetrica A: Transport Science*, vol. 12, no. 7, pp. 650–673, 2016.

[10] D. Hensher and M. A. Figliozzi, “Behavioral insights into the modeling of freight transportation and distribution systems,” *Transportation Part B*, vol. 41, no. 9, pp. 921–923, 2007.

[11] F. Chen, S. Mingtao, M. Xiaoxiang, and Z. Xingyi, “Assess the impacts of different autonomous trucks’ lateral control modes on asphalt pavement performance,” *Transportation Research Part C: Emerging Technologies*, vol. 103, pp. 17–29, 2019.

[12] A. Xingyi, D. Ziwei, F. Chen, P. Xiaodong, and X. Ming, “Using the visual intervention influence of pavement marking for rutting mitigation II: visual intervention timing based on the finite element simulation,” *International Journal of Pavement Engineering*, vol. 20, no. 5, pp. 573–584, 2019.

[13] T. Maruyama and N. Harata, “Incorporating trip-chaining behavior into network equilibrium analysis,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1921, pp. 11–18, 2005.

[14] W. Wisetjindawat, K. Sano, S. Matsumoto, and P. Raothanachonkun, “Micro-simulation model for modeling freight agent interactions in urban freight movement,” in *CD Proceedings, 86th Annual Meeting of the Transportation Research Board*, Washington, DC, USA, 2007.

[15] G. Liedtke, “Principles of micro-behavior commodity transport modeling,” *Transportation Research Part E*, vol. 45, no. 5, pp. 795–809, 2009.

[16] J. Holguin-Veras and E. Thorson, “Modeling commercial vehicle empty trips with a first order trip chain model,” *Transportation Research Part B*, vol. 37, no. 2, pp. 129–148, 2003.

[17] Southern California Association of Governments (SCAG), *SCAG Regional Travel Demand Model and 2008 Model Validation, 2012*, Southern California Association of Governments (SCAG), Los Angeles, CA, USA, 2012.

[18] Q. Wang and J. Holguín-Veras, “Investigation of attributes determining trip chaining behavior in hybrid microsimulation urban freight models,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2066, no. 1, pp. 1–8, 2008.

[19] M. Ruan, J. J. Lin, and K. Kawamura, “Modeling urban commercial vehicle daily tour chaining,” *Transportation Research Part E*, vol. 48, no. 6, pp. 1169–1184, 2012.

[20] Q. Wang and J. Holguín-Veras, “Tour-based entropy maximization formulations of urban freight demand,” in *CD Proceedings, 88th Annual Meeting of the Transportation Research Board*, Washington, DC, USA, 2009.

[21] S. I. You and S. G. Ritchie, “A GPS data processing framework for analysis of drayage truck tours,” *KSCE Journal of Civil Engineering*, vol. 22, no. 4, pp. 1454–1465, 2018.

[22] A. G. Wilson, *Entropy in Urban and Regional Modeling*, Pion, London, 1970.

[23] J. M. Borwein, "Maximum entropy and feasibility methods for convex and nonconvex inverse problems," *Optimization: A Journal of Mathematical Programming and Operations Research*, vol. 61, no. 1, pp. 1–33, 2012.

[24] D. G. Carrillo Murillo and G. A. Liedtke, "Model for the formation of colloidal structures in freight transportation: the case of hinterland terminals," *Transportation Research Part E*, vol. 49, no. 1, pp. 55–70, 2012.

[25] C. Xie, K. M. Kockelman, and S. T. Waller, "A maximum entropy method for sub-network origin-estimation trip matrix estimation," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2196, no. 1, pp. 111–119, 2010.

[26] D. H. Lee and Y. Fu, "Cross-entropy optimization model for population synthesis in activity-based microsimulation models," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2255, no. 1, pp. 20–27, 2011.

[27] G. N. Boshnakov and S. Lambert-Lacroix, "A periodic levinson-durbin algorithm for entropy maximization," *Computational Statistics and Data Analysis*, vol. 56, no. 1, pp. 15–24, 2012.

[28] X. Li, Z. Qin, L. Yang, and K. Li, "Entropy maximization model for the trip distribution problem with fuzzy and random parameters," *Journal of Computational and Applied Mathematics*, vol. 235, no. 8, pp. 1906–1913, 2011.

[29] S. Erlander, "Entropy in linear programs," *Mathematical Programming*", vol. 21, no. 1, pp. 137–151, 1981.

[30] C. Paige and M. A. Saunders, "Solution of sparse indefinite system of linear equations," *SIAM Journal on Numerical Analysis*, vol. 12, no. 4, pp. 617–629, 1975.

[31] M. Doyle, *A Barrier Algorithm for Large Nonlinear Optimization Problems*, Stanford University, Stanford, CA, USA, 2003.

[32] T. Elfving, "On some methods for entropy maximization and matrix scaling," *Linear Algebra and Its Applications*, vol. 34, pp. 321–339, 1980.

[33] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[34] SOL (Systems Optimization Laboratory), "Free software download," http:// www.stanford.edu/group/SOL/software/pdco.htmlJuly 20, 2011.

[35] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming (Third Edition)*, Springer, Berlin, 2008.

[36] C. Kriches, H. G. Bock, J. P. Schlöder, and S. Sager, "A factorization with update procedures for a KKT matrix arising in direcet optimal control," *Mathematical Programming Computation*, vol. 3, no. 4, pp. 319–348, 2011.

[37] R. Fletcher and T. Johnson, "On the stability of null-space methods for KKT systems," *siam Journal on Matrix Analysis and Application*, vol. 18, no. 4, pp. 938–958, 1997.

[38] A. Griewank, A. Walther, and M. Korzec, "Maintaining factorized KKT systems subject to rank-one updated of hessians and jacobians," *Optimization Methods and Software*, vol. 22, no. 2, pp. 279–295, 2007.