IIE Transactions

ISSN: 0740-817X (Print) 1545-8830 (Online) Journal homepage: http://www.tandfonline.com/loi/uiie20

# An efficient formulation of the flow refueling location model for alternative-fuel stations

Ismail Capar & Michael Kuby

*IIE Transactions* (2012) **44**, 622–636

ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/0740817X.2011.635175

# An efficient formulation of the flow refueling location model for alternative-fuel stations

ISMAIL CAPAR[1,*] and MICHAEL KUBY[2]

[1] *Texas A&M University, Industrial Distribution Program, College Station, TX 77843, USA*
*E-mail: capar@tamu.edu*
[2] *Arizona State University, School of Geographical Sciences and Urban Planning, Tempe, AZ 85287, USA*
*E-mail: mikekuby@asu.edu*

Received June 2010 and accepted August 2011

The Flow-Refueling Location Model (FRLM) locates a given number of refueling stations on a network to maximize the traffic flow among origin–destination pairs that can be refueled given the driving range of alternative-fuel vehicles. Traditionally, the FRLM has been formulated using a two-stage approach: the first stage generates combinations of locations capable of serving the round trip on each route, and then a mixed-integer programming approach is used to locate $p$ facilities to maximize the flow refueled given the feasible combinations created in the first stage. Unfortunately, generating these combinations can be computationally burdensome and heuristics may be necessary to solve large-scale networks. This article presents a radically different mixed-binary-integer programming formulation that does not require pre-generation of feasible station combinations. Using several networks of different sizes, it is shown that the proposed model solves the FRLM to optimality as fast as or faster than currently utilized greedy and genetic heuristic algorithms. The ability to solve real-world problems in reasonable time using commercial math programming software offers flexibility for infrastructure providers to customize the FRLM to their particular fuel type and business model, which is demonstrated in the formulation of several FRLM extensions.

**Keywords:** Alternative-fuel vehicles, facility location, flow capturing, mixed-integer programming

## 1. Introduction

Alternative-Fuel Vehicles (AFVs) that can be powered by electricity, natural gas, hydrogen, biodiesel, or ethanol have been promoted as a possible solution to the problem of reducing the adverse effects of extracting, importing, refining, and burning petroleum. AFVs face many disadvantages when compared with conventional gasoline- and diesel-powered vehicles, including a high initial investment cost, limited driving range, and, in particular, the scarcity of available stations for their refueling or recharging (Melendez, 2006; Romm, 2006). To attract consumers to purchase and use AFVs, station infrastructure must be deployed in convenient locations that are coordinated with each other. One approach for doing so is the Flow-Refueling Location Model (FRLM), originally developed by Kuby and Lim (2005) for optimally locating a system of alternative-fuel stations to refuel vehicles on the routes that consumers normally drive. The FRLM is a path-based model that locates $p$ stations to maximize the traffic flow of origin–destination (O–D) trips that are potentially refuelable given a maximum driving range of vehicles on a single tank of fuel or battery charge. Longer trips may require *combinations* of facilities on the path to serve an O–D trip that cannot be completed by refueling at a single station. In the original Kuby–Lim model, all possible feasible combinations of facilities that can serve each round trip are generated exogenously and input to the model as the options for covering or capturing each demand.

In this research, we propose and test a radically new Mixed Binary Integer Programming (MBIP) model for solving the FRLM that makes it possible for any feasible combination of facilities on a given path to refuel the flow on that path but does so without using combinations in the formulation. Although our formulation creates a much larger and more complex model, it solves the FRLM to optimality as fast as or faster than the published heuristic algorithms (Lim and Kuby, 2010) are able to obtain solutions. Furthermore, it is able to solve the FRLM on much larger networks than the original Kuby–Lim formulation can handle because of the proliferation of the number of facility combinations. Although computation speed, network size, and optimality are important considerations for analysts, the advantages of the new formulation extend beyond those characteristics.

*Corresponding author

The ability to generate and solve a FRLM using commercial mathematical programming software offers a valuable degree of flexibility to infrastructure providers to customize the FRLM to its particular fuel and business model or to relax some of the limiting assumptions of the basic FRLM. For instance, electric vehicles can be charged at home, and possibly at work, which means that many trips would start with a full charge, which changes the assumptions for determining whether a round trip can be completed. Electric charging stations can offer different voltages and therefore different charging times, while battery switch stations offer fast replacement that make longer trips with multiple "re-energizing" stops feasible. Station capacities were not included in the original FRLM but were introduced to the mixed-integer linear programming (MILP) formulation by Upchurch *et al*. (2009). For stations with long refueling times, with on-site fuel generation, or without large on-site storage, capacity may be an important issue to consider in planning the initial network, while for other fuels it may not affect location for a number of years. Some infrastructure providers may need to also service depot-based fleet demands, which could require that the model serve point-based or tour-based demands. As more manufacturers begin to produce vehicles, there will be a diverse fleet of vehicles with different driving ranges, rather than the single range assumption in the FRLM. Also, if early station networks are being partially funded by government subsidies, there may be equity or military considerations. There could be multiple objectives for the deployment or even multiple decision makers involved. Thus, as these examples illustrate, there is a need in the alternative-fuel industry for an infrastructure deployment model that can be flexibly adapted to a variety of circumstances.

In the remaining sections of this article, we review the closely related literature in Section 2 and introduce the formulation in Section 3. In Section 4, we compare results for several networks of different sizes with the original MILP formulation of the FRLM and with the previously published greedy and genetic algorithm heuristics. Then, in Section 5, we formulate several extensions to the FRLM to demonstrate how the new formulation can be flexibly adapted to different fuels and business models. Finally, we provide concluding remarks and future work in Section 6.

## 2. Literature review

Several different approaches have been developed to determine optimal locations of refueling stations. Upchurch and Kuby (2010) classified these approaches into three broad types. The first group of studies follow in the tradition of the well-known $p$-median model (Hakimi, 1964; ReVelle and Swain, 1970), which minimizes the total weighted distance traveled from each node to its closest facility. The $p$-median model locates fuel stations near where people live, and studies have shown that consumers tend to refuel near their homes (Sperling and Kitamura, 1986; Kitamura and Sperling, 1987). Nicholas *et al.* (2004), Nicholas and Ogden (2006), and Greene *et al*. (2008) applied the $p$-median model to locate fuel stations using heuristics. The "fuel travel-back" approach of Lin *et al.* (2008) is structurally similar to the $p$-median model but uses vehicle-miles traveled data rather than population data to weight the demand. Goodchild and Noronha (1987) used the $p$-median objective as one of the objectives in a multi-objective model for rationalizing gasoline station networks.

Goodchild and Noronha (1987) employed a second objective that demonstrates a second general approach to locate fuel stations: location on high-traffic routes to maximize the exposure to passing traffic. Their second objective maximizes the sum of traffic flows passing by all stations. Similarly, in a geographic information system–based suitability analysis for a national network of hydrogen stations, Melendez and Milbrandt (2005) required that stations be on roads with at least 20 000 vehicles per day. Traffic-count methods, however, may count the same trips multiple times, even if the drivers only refuel once, and thus they could recommend locating stations on consecutive links of a busy highway.

The third general approach identified by Upchurch and Kuby (2010) is also based on traffic flows but eliminates double counting of trips by using path flows instead of arc flows. Originating with Hodgson (1990) and Berman *et al.* (1992), a new branch of location analysis has arisen in which the fundamental units of demand are individual trips through a network rather than residential locations. The objective of the Flow-Intercepting Location Model (FILM) is to maximize the number of trips intercepted by locating $p$ facilities on a network. The model mimics the mathematical structure of the maximal covering model (Church and ReVelle, 1974), with the exception that the node-covering criterion is replaced by a flow-intercepting one. Like the max cover model, the FILM allows each demand to be satisfied only once, regardless of how many facilities can satisfy it. Zeng (2009) comprehensively reviewed the rapidly growing literature on these types of path-based models.

FILMs redefine consumer convenience and locate facilities "on the way" to where people are going, rather than near where they live. The applicability of the concept to driver behavior and fueling stations is confirmed by the survey of Kitamura and Sperling (1987), which found that whereas most drivers preferred to refuel near home, only 7% did so via a single-purpose round trip from home to station. Kuby *et al.* (2009) have argued that the preference to refuel near home is a luxury afforded by the ubiquity of gasoline stations. During the early transition phase to alternative fuels when alternative-fuel stations will be sparse, it may be more convenient for drivers to refuel far from home if there is a station right on their route rather than to drive to a station that is near home but out of their way. In addition, unlike the $p$-median model, which allocates each demand node to its closest facility, the path-based FILM allows more than one facility to satisfy the demand of a given origin node, because each origin node has trips to many destinations.

The application of a path-based flow-intercepting approach to locating refueling stations, however, raises a major complication. The basic FILM assumes that one facility anywhere on the path can fully satisfy the demand, which will often not be the case for AFVs. Due to their limited driving range, several stations may be required to complete longer trips without running out of fuel—and those stations will need to be properly spaced. Kuby and Lim (2005) extended FILMs to take these factors—driving range, multiple stations, and proper spacing—into account. The authors first created combinations of stations that can refuel a round trip on a path given the driving range of AFVs. Next, given the combination of stations, Kuby and Lim (2005) formulated the problem such that a traffic flow is captured only if all facilities in a feasible combination are opened:

$$\max \sum_{q \in Q} f_q y_q, \tag{1}$$

subject to:

$$\sum_{h \in H} b_{qh} v_h \geq y_q \quad \forall q \in Q, \tag{2}$$

$$a_{hi} x_i \geq v_h, \quad \forall h \in H, i \,|a_{hi} = 1, \tag{3}$$

$$\sum_{i} x_i = p, \tag{4}$$

$$x_i, v_h, y_q \in \{0, 1\}, \quad \forall i, h, q, \tag{5}$$

where:

$q$ = index of O–D pairs (and the shortest paths between them);
$Q$ = set of all O–D pairs;
$h$ = index of combinations of facilities;
$H$ = set of all potential facility combinations;
$i$ = index of potential facility locations;
$f_q$ = flow volume on the shortest path between O–D pair $q$;
$b_{qh}$ = a coefficient equal to one if facility combination $h$ can refuel O–D pair $q$, zero otherwise;
$a_{hi}$ = a coefficient equal to one if facility $i$ is in combination $h$, zero otherwise;
$p$ = the number of facilities to be located;
$y_q$ = one if $f_q$ is captured, zero otherwise;
$v_h$ = one if all facilities in combination $h$ are open, zero otherwise;
$x_i$ = one if a facility is located at $i$, zero otherwise.

The objective function (1) and the $p$-facility constraint (4) are unchanged from the FILM. Constraints (2) make refuelability contingent on an eligible combination $h$ of stations being opened ($v_h = 1$), and constraints (3) require that *all* of the stations $i$ in combination $h$ must be built ($x_i = 1$) for the combination to be considered open.

A number of extensions and applications of the FRLM have been developed. Incorporating a driving range limitation means that the nodes of the network are no longer a finite-dominating set, and thus Kuby and Lim (2007) created and tested methods for adding candidate sites along arcs. They found that additional candidate sites on arcs are unlikely to lead to increased coverage in urban areas, but as the range of the vehicle shrinks relative to the length of network arcs, additional sites can improve the objective, especially as $p$ increases and the model tries to serve the hard-to-cover paths. Upchurch *et al.* (2009) added station capacities to the original FRLM formulation, which required not only a new capacity constraint but the conversion of the $y_q$ "coverage" variable to a $y_{qh}$ allocation variable that assigns paths $q$ to combinations $h$ in a system-optimal way. Lim and Kuby (2010) developed greedy and genetic heuristic algorithms for solving the FRLM.

Kuby *et al.* (2009) applied the FRLM to metropolitan Orlando and the state of Florida and introduced weighting of trips by distance (longer AFV trips count more than short trips because they replace more conventional fuel) or likelihood of consumer adoption of AFVs (e.g., in neighborhoods with higher education and incomes, more vehicles per household, and longer commutes). Upchurch and Kuby (2010) compared the FRLM and the $p$-median model on a number of different dimensions for the same Florida and Orlando networks. The $p$-median model tended to spread the stations across the network favoring the denser clusters of population but without regard to locating on major thoroughfares. In contrast, the FRLM located stations at high-traffic intersections that intercept large flow volumes while trying not to cannibalize the flows intercepted by other stations. On the inter-city network, the FRLM tended to first develop clusters of stations in large metropolitan conurbations such as Miami–Ft. Lauderdale–Palm Beach in which they refuel the shorter and heavier O–D flow volumes between cities within the conurbation and also work together to refuel longer trips requiring multiple refueling stops. Upchurch and Kuby (2010) also compared the ability of the two models to produce solutions that also perform well on the other's objective. The FRLM's solutions performed better in minimizing average home-to-station distances than the $p$-median's solutions did in intercepting and refueling flows. Finally, they investigated how the solutions of both models "nest" as the station network expands. The FRLM's solutions were extremely stable as $p$ increases, with very few substitutions made as new stations are added. In contrast, there was quite a bit of turnover of sites in $p$-median's solutions from small to large $p$.

In one respect, the most important part of the FRLM lies outside of the formulation itself. Kuby and Lim (2005) developed an algorithm to determine the feasible combinations of facilities that can enable a driver to repeatedly complete a round trip from their origin to their destination and back without exceeding their driving range and running out of fuel. This algorithm determines exogenously the $b_{qh}$ coefficients that define whether facility combination $h$ can refuel O–D pair $q$. However, even with a feature that eliminates redundant superset combinations (e.g., combination {3, 5, 7} can be safely eliminated if combination {3, 7} can also refuel the path without location 5), the number of combinations to evaluate and incorporate into the model for a single very long path can be astronomical. For this reason, Lim and Kuby (2010) were unable to even *generate* the MILP model for a Florida network with 302 nodes, let alone solve it, in order to compare their heuristics to the global optimum.

## 3. Problem definition and formulation

Let $G(N, A)$ be a transportation network, where $N$ is the set of nodes (i.e., origins, destinations, and junctions) and $A$ is the set of arcs. While all nodes in $N$ are eligible candidate sites for stations, the set of O–D nodes can be a subset of $N$. Thus, an unpopulated road junction can be included as a candidate site but need not be included as an O–D node. Next, given a set of O–D pairs ($Q$) with a non-negative traffic flow ($f_q$), the set of nodes visited while traveling on path $q$ ($N_q$), and vehicle range ($R$), the FRLM is defined as the problem of locating $p$ facilities on the network $G(N, A)$ to maximize the total traffic flow refueled. Traffic flow between an O–D pair $q$ is considered as refueled only when vehicles leaving the origin can reach the destination and return back to the origin without running out of fuel. Before presenting the problem formulation, we discuss related assumptions and present additional notation.

Similar to the previous FILM/FRLM-related research, it is assumed that traffic flow and the path between O–D pairs are known in advance. In FILM/FRLM problems, traffic is assigned a single path—usually the shortest path (e.g., Hodgson (1990); Berman *et al.* (1992); Kuby and Lim (2005) and Zeng (2009)). From a problem formulation perspective, the proposed model can easily be extended to multiple paths; hence, this assumption is not restrictive. Although in some instances, the flow information may not be readily available, it can be obtained from a traffic demand matrix or through an O–D estimation method (Cascetta and Nguyen, 1988). Therefore, the assumption that the traffic flow is known in advance is reasonable as well.

We explain the concept of refueling and effect of range through a very simple example using three different ranges on the network given in Fig. 1. Assuming that only nodes are candidate sites, when $R < 50$, there is no chance for vehicles to complete the trip between the O–D pair because vehicles cannot travel the arc between $x_2$ and $x_3$. When $R = 50$, the infrastructure provider can choose one of two alternative solutions by opening $\{x_1, x_2, x_3, x_4\}$ or $\{x_1, x_2, x_3, x_5\}$. Under both solutions, vehicles refuel at $x_1$, $x_2$, and $x_3$. If $x_4$ is built, vehicles refueling at $x_4$ can reach the destination and return to $x_4$. Next, after refueling at $x_4$, vehicles can travel back to the origin by refueling again at $x_3$ and $x_2$. Similarly, we can see that when $R = 60$, the infrastructure provider does not need to build $x_1$ anymore because a vehicle refueled at $x_2$ (while traveling back from the destination) can reach the origin and have enough fuel to travel to $x_2$ when a new trip is next started. When $R = 254$, a single station at any node is enough to refuel the path because even after refueling at $x_1$ ($x_5$) vehicles will have enough fuel to reach $x_5$ ($x_1$) and return to $x_1$ ($x_5$).

Although all possible combinations of range, number, and location of facilities are not iterated in the above examples, they provide three key observations about formulation of such a problem. First, if there is no station built at the origin (e.g., $R = 60$) then there should be at least one station built within the $R/2$ distance from the origin node so that it can be reachable by half a tank. Note that we refer to "tank" of "fuel" generally, but the model also applies to batteries containing an electric charge or more generally to any form of energy storage. Second, if there is a station built at a location, the next built facility should be within the range (e.g., $R = 50, 60$). Finally, if the vehicle range is greater than or equal to two times the path length (e.g., $R = 254 = 2(30 + 50 + 32 + 15)$), a single station at any point can refuel the entire path.

*New parameters, indexes, and sets*

$R$ = the range of alternative fuel vehicles;
$m, r$ = indices of the order of candidate sites on a given path;
$t$ = an index of the state of a candidate site, $t = 0$ (station built), or 1 (station not built);
$D_q$ = the length of the shortest path of an O–D pair $q$;
$M_q$ = the number of candidate sites on path $q$ beyond the origin but not within half the range $R$ of the destination of path $q$; that is, in the distance interval $(0, D_q - R/2)$ on path $q$; if $(D_q - R/2 \leq 0)$, then $M_q = 0$;

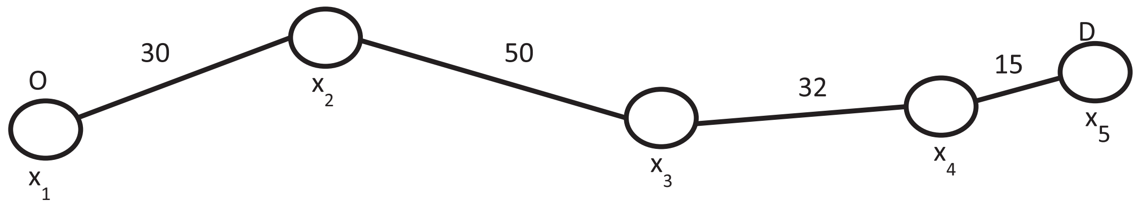


**Fig. 1.** An example network with a single O–D pair.

$d_{m,r}^q$ = the distance between the $m$th and $r$th candidate locations in $N_q$;

$k_m^q$ = the index $i$ corresponding to the $m$th candidate site in $N_q$, $m = 1$ (origin), 2, 3, ..., $n_q$ (destination);

$N_q$ = the set of candidate sites on a path $q$, *now sorted in sequential order* from origin to destination;

$N_{m,t}^q$ = the set of candidate sites accessible from the $m$th candidate site on a path $q$, where

$$N_{m,t}^q = \begin{cases} \{N_q \mid d_{m,r}^q \le R, r > m\}, & \forall q \in Q, \quad m = 1, 2, \ldots, M_q, t = 1, \\ \{N_q \mid d_{m,r}^q < R, r > m\}, & \forall q \in Q, \quad m = 2, \ldots, M_q, t = 0, \\ \{N_q \mid d_{m,r}^q \le R/2, r > m\}, & \forall q \in Q, \quad m = 1, t = 0. \end{cases}$$

*New decision variables*

$b_{m,0}^q$ = 1 if $x_{k_m^q} = 0$ and vehicles on path $q$ would have enough fuel remaining at the $m$th candidate site to be able to get to the next open fuel station on $q$ without running out of fuel, and zero if $x_{k_m^q} = 0$ and it cannot get to an open fuel station.

$b_{m,1}^q$ = 1 if $x_{k_m^q} = 1$ and vehicles on path $q$ would have enough fuel after refueling at the $m$th candidate site to be able to get to the next open fuel station on $q$ without running out of fuel, and zero if $x_{k_m^q} = 1$ and it cannot get to an open fuel station.

The term $k_m^q$ is a pointer to a particular candidate site $i$, and each $x_i$ may correspond to multiple $x_{k_m^q}$ because a candidate site can be on many different paths with varying location order indexes $m$.

$$\max \sum_{q \in Q} f_q y_q, \tag{1}$$

subject to:

$$b_{m,t}^q + (-1)^t x_{k_m^q} \le 1 - t, \quad \forall q \in Q, t \in \{0, 1\}, \quad m = 1, 2, \ldots, M_q, \quad \text{if } M_q \ne 0, \tag{6}$$

$$b_{m,t}^q - \sum_{n \in N_{m,t}^q} x_n \le 0, \quad \forall q \in Q, t \in \{0, 1\}, \quad m = 1, 2, \ldots, M_q, \quad \text{if } M_q \ne 0, \tag{7}$$

$$b_{m,1}^q - \sum_{n \in x_{k_m^q} \cup N_{m,1}^q} x_n \le 0, \quad \forall q \in Q, m = M_q + 1, \tag{8}$$

$$b_{m,0}^q = 0, \quad \forall q \in Q, \quad m = M_q + 1, \tag{9}$$

$$\sum_{m=1}^{M_q+1} \sum_{t \in \{0,1\}} b_{m,t}^q = (M_q + 1) y_q, \quad \forall q \in Q, \tag{10}$$

$$\sum_{i \in N} x_i = p, \tag{4}$$

$$y_q \in \{0, 1\}, \; x_i \in \{0, 1\}, \; b_{m,0}^q \in \{0, 1\}, \; b_{m,1}^q \in \{0, 1\}. \tag{11}$$

The formulation in this article is radically different than the original formulation of the FRLM. Although it implements a similar logic and functionality as the original model, it eliminates the need for facility combinations by incorporating the underlying logic of the exogenous algorithm presented in Kuby and Lim (2005) for determining the feasible combinations for each path into the constraints of the MILP model itself. This is done by introducing $b_{m,0}^q$ and $b_{m,1}^q$ variables for $M_q + 1$ points along every path $q$. These variables indicate whether a station exists ($b_{m,1}^q$) or not ($b_{m,0}^q$) at that node along the path and whether a driver at that node could reach another station further down the path without running out of fuel ($b_{m,t}^q = 1$) or not ($b_{m,t}^q = 0$). Then, using these variables, if every considered node along the path is refuelable by virtue of being able to reach another station without running out of fuel, then the full path is refuelable, and feasible spacing of stations along paths is guaranteed without the use of combination variables.

The objective function (1) is identical to that in the original FRLM. Constraint (6) governs the relationship between the $b$ variable for the $m$th node on path $q$ and the $x$ variable for *that same node* to make sure that the right case of $b_{m,t}^q$ is allowed to equal a value of one. Constraints (7) govern the relationship between the $b$ variable for the $m$th node on path $q$ and the $x$ variables *further along the path*, to make sure that the $b$ variable is zero unless another refueling station can be reached from the $m$th node on path $q$. There are different cases of (7) depending on whether the $m$th node is the origin of the path and whether there is a station at the node. Constraints (8) and (9) deal with the case of nodes within half the range of the destination of path $q$ ($m = M_q + 1$). They make sure that a trip is counted as refuelable if the traffic flowing through the ($M_q$)th candidate site can travel to its destination and make it back to the ($M_q$)th candidate site by virtue of the fact that at least one facility is located in this remainder segment of the path.

In constraint (10), the traffic flow on path $q(f_q)$ can be completely refueled if the sum of the refueled points (with/without an open facility at the considered location) on the path is equal to the number of the points considered ($M_q + 1$). Constraint (4) is the same as in the original FRLM and requires that exactly $p$ fuel stations are built. Constraint (11) requires that all variables are binary variables.

Although we formulated FRLM as a binary integer programming (BIP) model, the $b_{m,0}^q$ and $b_{m,1}^q$ variables can be relaxed to continuous variables between zero and one to reduce the computational time, and the proposed model becomes a MBIP model.

## 4. Numerical experiments

We ran experiments using networks from the literature (i.e., a Florida state highway network (see Fig. 2) from Kuby *et al.* (2009) and Lim and Kuby (2010) and a 55-node

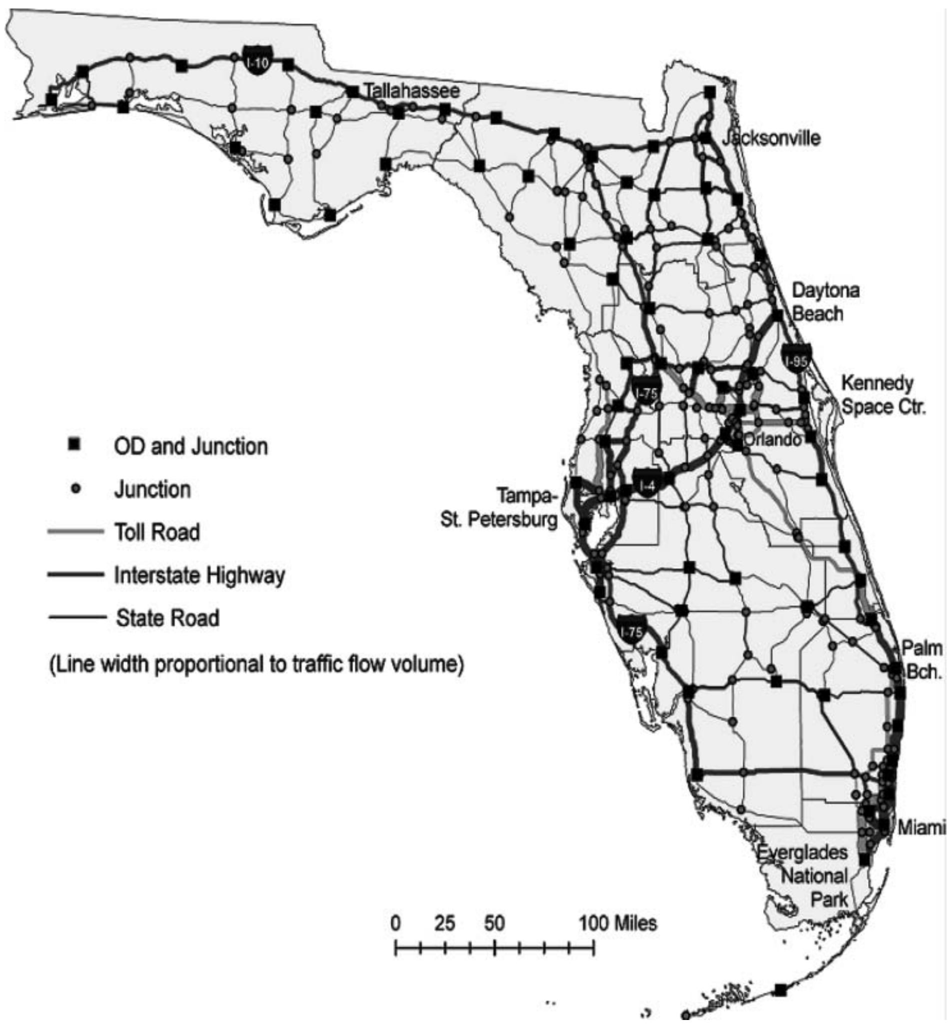


**Fig. 2.** The Florida state highway network.

network utilized by Hodgson *et al.* (1994)) and randomly generated networks. While the Florida state highway network provides an opportunity to observe how solutions nest on a real map and creates a chance to compare the solution time and quality with already published heuristics solutions, randomly generated networks and the 55-node network from Hodgson *et al.* (1994) allow us to compare two modeling approaches and evaluate the performance of the proposed model in more detail. We ran experiments using Xpress-MP Version 2008a on a computer with Windows 7 64-Bit OS, Intel Core *Q*-720 CPU at 1.6 GHz with 6 GB of RAM. Xpress-MP was run for a maximum of 1 h for each problem instance.

The Florida state highway network has 495 arcs and 302 junctions (vertices), each of which serves as a candidate site. Of the 302 candidate sites, 74 are O–D nodes for trips. Note that because the model ensures that the return trip is refuelable, by extension so are the round trips starting at either end (Kuby and Lim, 2005). The Florida network of 74 O–D nodes requires 2701 unique O–D pairs excluding the intra-zonal trips. The road network includes all interstate, toll, and U.S. highways as well as important state highways. The 74 O–D nodes generally represent one node per county, although some highly populated counties are disaggregated into several nodes and some adjacent low-population counties are combined into a single node. The 55-node network has 129 arcs and it is assumed that each node is a candidate site for a service station.

To obtain random problem instances, for a given network size of $s$, we initially generated $s$ nodes in $[0, 1000]^2$ according to a continuous uniform distribution. We assumed Euclidean distance between nodes and, using the algorithm presented in Kruskal (1956), we determined the minimum spanning tree. Then, we added $s$ additional arcs to the minimum spanning tree by connecting the $s$ closest node pairs that were not directly connected with an arc. We also randomly chose $w$ O–D nodes out of $s$ nodes and used the algorithm presented in Dijkstra (1959) to find the shortest path among $w(w-1)/2$ unique O–D pairs. Finally, we used $R = 250$ for experiments with random networks.

In our experiments with random networks, we generated two random instances of networks with $s = 75$, 100, 250, 500, 750, and 1000; $w = 40$ and 80 nodes. To generate the flow between O–D pairs, we first randomly generated a population density for each node and then used the following spatial gravity model provided by Fotheringham and O'Kelly (1989) to estimate the traffic flow for each OD pair:

$$f_q = \frac{e_o^q \, e_d^q}{D_q^2}$$

where $e_o^q$ ($e_d^q$) is the population at the origin (destination) node. Finally, to avoid solving problem instances for every $p$ value, we determined the minimum number of service stations needed to cover between 20% and 100% with 20% increments by solving a set covering version of the problem by replacing the objective function with $\min \sum_{i \in N} x_i$, removing constraint (4), and adding a constraint in the form of $\sum_{q \in Q} f_q y_q \geq O_c$, where $O_c$ is the objective flow coverage. This is an example of the modeling flexibility afforded by a math programming solution method.

In the following subsections, we first compare the initial FRLM formulation proposed by Kuby and Lim (2005) with the new formulation both in terms of input sensitivity and computational results. Next, we present experiments on the Florida state highway network to analyze the performance of the proposed model and compare its solution time/quality with existing heuristics, namely, greedy, greedy with substitution, and genetic algorithms. Finally, we show some additional experiments to evaluate the performance of the proposed formulation on variety of networks.

### 4.1. *Comparison of FRLM formulations*

As mentioned earlier, the formulation of the FRLM presented by Kuby and Lim (2005) requires pre-generation of a complete list of candidate site combinations that can refuel each path $q$ (excluding redundant supersets). When Lim and Kuby (2010) attempted to generate the path-combinations for the Florida state highway network, the computer (Pentium 4, 3.2 GHz, 1 GB RAM) started running low on memory after 13 h of computation, by which time it had generated combinations for only 39 out of 2701 O–D pairs (1.4% of all paths). Hence, generating the complete list of combinations and the paths each one can refuel has proven a great challenge so far. Our modeling approach answers these challenges by providing a fundamentally different formulation that makes it possible to solve large problems to optimality.

Before providing computational results, we discuss the two models in terms of sensitivity to input data. First, let us focus on Fig. 1 again. Given that the distance between O–D pair ($D_q$) is 127, assuming $R = 60$, the minimum number of stations required is equal to three ($lu = \lceil D_q/R \rceil$). Hence, the combination generation scheme utilized by Kuby and Lim (2005) basically needs to populate:

$$\sum_{r=3}^{5} \binom{5}{r} \left( \sum_{r=lu}^{n_q} \binom{nq}{r} \right)$$

combinations initially and then eliminate redundant supersets to generate the combination variables ($b_{qh}$). Therefore, the input generation scheme is not directly limited by the network size but rather by the number of nodes on the shortest path between an O–D pair. A single O–D pair with a high number of nodes in between and a vehicle with a short range can run a computer out of memory or take a very long time to generate the combinations. It is clear that the generation of combinations and elimination of redundant combinations is in itself a combinatorial problem and the number of combinations for a path increases exponentially as the number of nodes between O–D pairs increases.

On the other hand, the new formulation adds only two new variables and two constraints (8) and (9) or four constraints (6) and (7) for each new node included on a path. Consequently, the size of the MBIP model grows linearly as new nodes are introduced to a network, which is the main reason for being able to formulate and solve large networks.

To compare the FRLM formulations, we used the 55-node network with $w = 40$ and $R = 15$. We also generated random networks as previously specified and used $R = 250$ as mentioned earlier. For networks with 100 or more nodes, we were not able to generate the combination variables or solve the resulting problems because the computer (with six times more RAM than available to Lim and Kuby (2010)) ran out of memory. For example, it took over an hour to generate 147 483 combinations for a 100-node network; however, Xpress-MP gave an "out of memory" error while loading the model. Hence, our experimental results are based on 55- and 75-node networks. In Table 1, we present details for the networks considered as well as the

**Table 1.** Details of pre-processing for the model presented in Kuby and Lim (2005)

| | *55 Nodes* | *75 Nodes* | *100 Nodes* |
|---|---|---|---|
| Number of O–D pairs | 780 | 780 | 780 |
| *Max. number of nodes on a path | 10 | 21 | 23 |
| *Avg. number of nodes on a path | 4.9 | 9.6 | 10.8 |
| Number of combinations | 5055 | 24 107 | 147 483 |
| Time to generate combinations (s) | 56 | 338 | 4590 |

*Includes origin and destination nodes

**Table 2.** Comparison of solution time and memory requirements

| | *Kuby and Lim (2005)* | | *MBIP* | |
|---|---|---|---|---|
| | *55 Nodes* | *75 Nodes* | *55 Nodes* | *75 Nodes* |
| Avg. solution time (s) | 7 | 20 | 25 | 37 |
| Maximum memory used (MB) | 52 | 115 | 47 | 81 |

resulting average number of combinations and the time to generate combinations.

Based on Table 1, when the network size is increased from 55 nodes to 75 nodes, the time it takes to generate combinations increased from 56 s to 338 s, on average. Considering that the number of O–D pairs is the same for these networks, the reason for the time increase is due to the increase in the number of nodes between O–D pairs. The exponential increase becomes obvious when the network size increases from 75 to 100 nodes. We observed a 13-fold increase in the time for generating the combinations when the average/max number of nodes increased only slightly.

In terms of solution times, all 40 problem instances are solved to optimality in less than 5 min using both formulations. In general, the Kuby–Lim formulation performed slightly faster and solved the problems in 14 s on average compared to 31 s for the new formulation. Although the new formulation required more time, it required less computer resources. For example, as shown in Table 2, the MBIP formulation required around 30% more memory for the larger 75-node network. Hence, the MBIP formulation provided in this article requires less computer resources and is scalable to large problem instances.

### 4.2. *Experiments with the Florida state highway network*

Due to the time-consuming process of generating combinations, Lim and Kuby (2010) developed heuristic algorithms (i.e., greedy (Greedy), greedy-adding with $n$ substitution (G_Sub-1, -2, and -3) and genetic algorithm (GA)) to solve large-scale problems. We re-ran the heuristics (the same code used by Lim and Kuby (2010) on the same computer as the Xpress-MP model to have a fair comparison in terms of solution time advancement—(faster CPU, higher RAM, etc.). Consequently, the solution times reported for heuristics in this article are significantly less than those reported by Lim and Kuby (2010). We solved the FRLM problem with two different objective functions: maximize trips refueled (flow) and maximize Vehicle-Miles Traveled (VMT). Under VMT, the number of trips in the objective function is multiplied by the path distance. We ran experiments for $p = 1$ to $p = 25$ with an increment of one and $p = 30$ to $p = 60$ with an increment of 10.

Figure 3 shows a typical solution of the FRLM for $p =$ 5, 10, and 15 and a driving range of 100 miles, maximizing VMT. Three of the first five stations are located in the greater Miami region. This corridor is ideally suited for introducing new fueling infrastructure because it has high volumes of short-distance inter-city trips that are arranged linearly in a narrow corridor using many of the same highways. Thus, these three stations can work together to refuel longer trips requiring multiple stops between Palm Beach and South Miami Heights (see covered routes for $p = 5$). The other two stations are placed in central Orlando and central Tampa, making it possible to complete single-stop trips to those centers from origins less than 50 miles away or a multi-stop trip from St. Petersburg all the way to Daytona and back. (Note that the Orlando station is *not* optimal for $p = 5$ when maximizing trips instead of VMT. In that case, four stations are located in the Miami region, which adds more total trips but fewer long trips.) Returning to the VMT solution displayed in Fig. 3, the $p = 10$ solution adds three more stations to the dense Orlando–Tampa–Wildwood I-4/I-75/Florida Turnpike triangle, as well as two more stations in the southeastern corridor. When $p = 15$, three of the new stations are added to the northern half of the I-95 corridor, making long-distance trips (which are more heavily weighted by the VMT objective) possible all the way from South Miami to Jacksonville (700 miles round trip).

This example also confirms the strong "nesting" or "stability" properties of FRLM solutions first noted in Upchurch and Kuby (2010). With one exception, the stations that were optimal with smaller $p$ remain optimal with larger $p$. Only the $p = 5$ station at Pompano Beach loses its optimal status when $p$ is increased to 10, when two new stations are added to its north and south.

We use solution time and percentage of traffic that can be refueled as the basis of comparison in the analysis that follows. For the MBIP model solved with Xpress-MP, the maximum solution time was set at 1 h: five runs ($p = 20, 21,$ 23, 24, and 25) reached the limit before global optimality was reached or confirmed.

The numerical experiment shows that the proposed model solves very efficiently, in terms of both computation time and solution quality, using over-the-counter software (Table 3). The last column ($\Delta\%$) indicates the absolute percentage improvement in flow refueled when the MBIP's results were compared with the best results achieved using heuristic approaches. The average improvement was 0.22% with a maximum improvement up to 0.53%. In seven runs—mostly with small $p$—one or more heuristics also found the optimal solution.

**Table 3.** Results of MBIP model and heuristics approaches utilized by Lim and Kuby (2010) for flow refueling objective function, with $R = 100$ miles

| | *Greedy* | | *G_Sub-1* | | *GA* | | *MBIP* | | |
|---|---|---|---|---|---|---|---|---|---|
| *p* | *Trips (%)* | *Time* | *Trips (%)* | *Time* | *Trips (%)* | *Time* | *Trips (%)* | *Time* | Δ% |
| 1 | 25.64 | 0 | 25.64 | 0 | n/a | n/a | 25.64 | 15 | 0 |
| 2 | 40.50 | 1 | 40.50 | 2 | n/a | n/a | 40.50 | 8 | 0 |
| 3 | 52.60 | 2 | 52.60 | 5 | n/a | n/a | 52.60 | 8 | 0 |
| 4 | 58.96 | 3 | 58.96 | 9 | n/a | n/a | 59.02 | 10 | 0.06 |
| 5 | 64.89 | 4 | 64.89 | 15 | 64.43 | 278 | 64.89 | 13 | 0 |
| 6 | 70.69 | 5 | 70.69 | 22 | n/a | n/a | 70.69 | 9 | 0 |
| 7 | 74.34 | 6 | 74.34 | 32 | n/a | n/a | 74.34 | 11 | 0 |
| 8 | 77.04 | 8 | 77.04 | 43 | n/a | n/a | 77.04 | 16 | 0 |
| 9 | 79.32 | 9 | 79.32 | 57 | n/a | n/a | 79.67 | 20 | 0.35 |
| 10 | 80.87 | 11 | 80.87 | 73 | 80.89 | 1168 | 81.22 | 72 | 0.33 |
| 11 | 82.31 | 12 | 82.31 | 91 | n/a | n/a | 82.67 | 86 | 0.36 |
| 12 | 83.71 | 14 | 83.71 | 112 | n/a | n/a | 84.06 | 150 | 0.35 |
| 13 | 85.06 | 16 | 85.06 | 135 | n/a | n/a | 85.41 | 177 | 0.35 |
| 14 | 86.27 | 18 | 86.27 | 158 | n/a | n/a | 86.64 | 276 | 0.37 |
| 15 | 87.40 | 20 | 87.40 | 183 | 87.7 | 2187 | 87.84 | 242 | 0.14 |
| 16 | 88.48 | 22 | 88.48 | 212 | n/a | n/a | 89.00 | 573 | 0.52 |
| 17 | 89.44 | 24 | 89.46 | 243 | n/a | n/a | 89.99 | 271 | 0.53 |
| 18 | 90.31 | 25 | 90.36 | 280 | n/a | n/a | 90.88 | 1698 | 0.52 |
| 19 | 91.15 | 28 | 91.36 | 321 | n/a | n/a | 91.72 | 1260 | 0.36 |
| 20 | 91.84 | 30 | 92.15 | 364 | 92.35 | 3508 | 92.39 | 3600 | 0.04 |
| 21 | 92.52 | 32 | 92.87 | 412 | n/a | n/a | 93.10 | 3600 | 0.23 |
| 22 | 93.08 | 35 | 93.44 | 461 | n/a | n/a | 93.80 | 1237 | 0.36 |
| 23 | 93.65 | 37 | 94.00 | 515 | n/a | n/a | 94.26 | 3600 | 0.26 |
| 24 | 94.20 | 39 | 94.54 | 574 | n/a | n/a | 94.81 | 3600 | 0.27 |
| 25 | 94.74 | 42 | 95.01 | 636 | 95.27 | 6587 | 95.29 | 3600 | 0.02 |

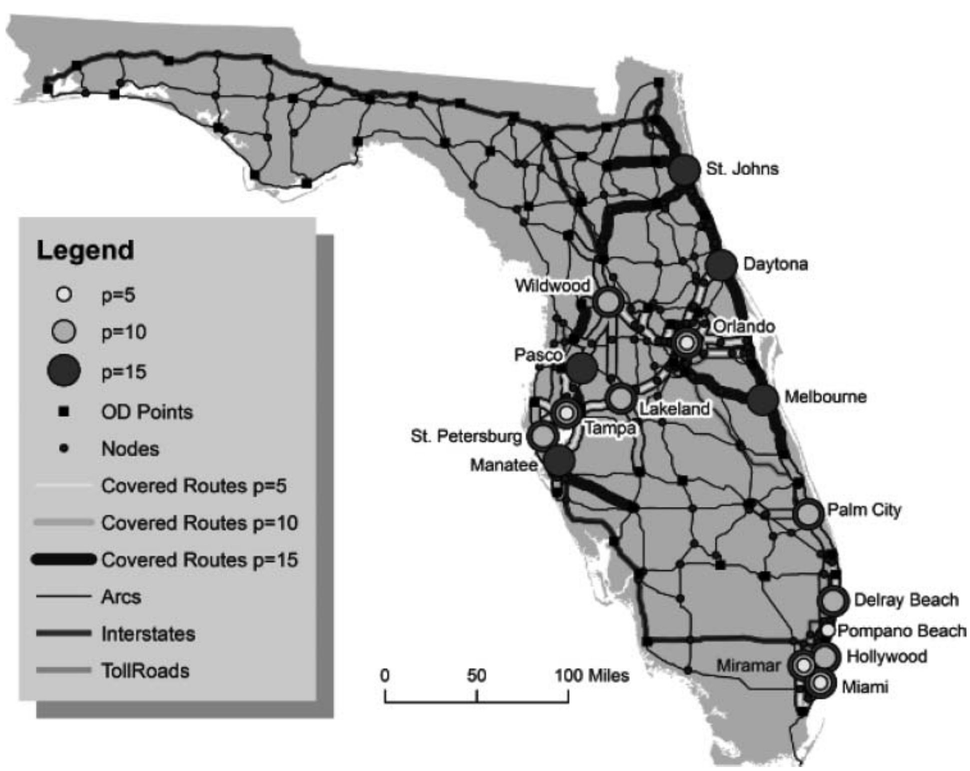


**Fig. 3.** Solution for the Florida state highway network for $p = 5$, 10, and 15 under VMT objective when $R = 100$ miles.

**Table 4.** Results of MBIP model and heuristics approaches utilized by Lim and Kuby (2010) for VMT objective function with $R =$ 100 miles

| | *Greedy* | | *G_Sub-1* | | *GA* | | *MBIP* | | |
|---|---|---|---|---|---|---|---|---|---|
| *p* | *Trips (%)* | *Time* | *Trips (%)* | *Time* | *Trips (%)* | *Time* | *Trips (%)* | *Time* | Δ% |
| 1 | 21.46 | 0 | 21.46 | 0 | n/a | n/a | 21.46 | 18 | 0 |
| 2 | 30.74 | 1 | 30.74 | 2 | n/a | n/a | 30.74 | 10 | 0 |
| 3 | 37.58 | 2 | 37.58 | 4 | n/a | n/a | 37.58 | 17 | 0 |
| 4 | 44.38 | 3 | 44.38 | 8 | n/a | n/a | 44.38 | 20 | 0 |
| 5 | 49.52 | 4 | 49.52 | 13 | 49.52 | 344 | 50.46 | 25 | 0.94 |
| 6 | 53.56 | 5 | 53.56 | 20 | n/a | n/a | 54.51 | 54 | 0.95 |
| 7 | 57.41 | 6 | 57.41 | 28 | n/a | n/a | 58.20 | 68 | 0.79 |
| 8 | 61.10 | 8 | 61.10 | 39 | n/a | n/a | 61.56 | 135 | 0.46 |
| 9 | 63.31 | 9 | 63.31 | 50 | n/a | n/a | 64.69 | 213 | 1.38 |
| 10 | 65.73 | 10 | 65.73 | 63 | 66.90 | 1098 | 66.90 | 397 | 0 |
| 11 | 68.03 | 11 | 68.03 | 78 | n/a | n/a | 69.57 | 360 | 1.54 |
| 12 | 70.05 | 13 | 70.68 | 95 | n/a | n/a | 71.62 | 609 | 0.94 |
| 13 | 71.95 | 14 | 72.58 | 114 | n/a | n/a | 73.52 | 1888 | 0.94 |
| 14 | 73.79 | 16 | 74.42 | 135 | n/a | n/a | 75.50 | 910 | 1.08 |
| 15 | 75.38 | 17 | 76.18[a] | 208 | 77.54 | 2383 | 77.54 | 3267 | 0 |
| 16 | 77.00 | 19 | 77.92[a] | 235 | n/a | n/a | 79.39 | 725 | 1.47 |
| 17 | 78.45 | 21 | 80.14[b] | 320 | n/a | n/a | 81.18 | 799 | 1.04 |
| 18 | 79.84 | 23 | 81.88 | 249 | n/a | n/a | 82.54 | 2313 | 0.66 |
| 19 | 81.20 | 25 | 83.24 | 286 | n/a | n/a | 83.84 | 3600 | 0.6 |
| 20 | 82.51 | 26 | 84.52 | 325 | 84.63 | 3716 | 85.12 | 3600 | 0.49 |
| 21 | 83.79 | 28 | 85.71 | 368 | n/a | n/a | 86.26 | 3600 | 0.55 |
| 22 | 84.98 | 30 | 87.25[b] | 598 | n/a | n/a | 86.98 | 3600 | −0.27 |
| 23 | 86.17 | 33 | 88.49[a] | 638 | n/a | n/a | 88.56 | 3600 | 0.07 |
| 24 | 87.35 | 35 | 89.69[a] | 746 | n/a | n/a | 89.38 | 3600 | −0.31 |
| 25 | 88.40 | 37 | 90.57[b] | 874 | 90.17 | 6006 | 90.32 | 3600 | −0.25 |

[a,b] The best results achieved with G_Sub-2,3.

Although the Greedy approach is fastest in terms of solution time, the solutions obtained are often the worst of the four methods—although sometimes it finds the same solution as the others. FRLM problems are strategic decisions, not operational; hence, solving these problems in a couple of seconds, as the Greedy algorithm does, may not be as important as improving the solution 1–2% in an hour. In addition, the solution time of the model can compete with that of G_Sub-1. For example, 7 out of 20 optimally solved problems were actually solved faster using Xpress-MP. (Note: only G_Sub-1 information is provided in Table 3 because it generated solutions with the same objective function value within less time compared to G_Sub-2 and Sub-3.) Finally, compared with the GA, the new MBIP found equally good or better solutions for reported $p$ values but did so one to two orders of magnitude faster.

We also used a VMT objective function on the Florida state highway network. As Table 4 shows, the performances of the heuristics were not as good when they were applied to the VMT objective. Greedy and G_Sub-$n$ heuristics obtained the optimal solutions for only four problems. In addition, the gap between the optimal solutions and results of Greedy heuristics increased to 1.52% on average, and the gap was higher than 2% for 10 problem instances. G_Sub-$n$ heuristics performed better compared to Greedy with an average gap of 0.63; however, they took 14 times longer compared to Greedy. The G_Sub-2 and Sub-3 together outperformed MBIP in three instances with an average improvement of 0.28% and solution times of 12–14 min compared with 1 h for MBIP. The gap also widened for GA under the VMT objective function. While the performance of the heuristic approaches are data dependent and it may be difficult to know which heuristic should be used for which type of data, the MBIP performed similarly under both objective functions based on the number of optimal solutions (20–18) obtained within a 1-h time frame.

Although solving the MBIP model using Xpress-MP without any additional input is promising, we observed that solution times can be improved by strengthening connections among $b^q_{m,t}$ variables and providing a list of prioritized variables and a cutoff value for branch and bound. We specifically utilized the following proposition to add additional constraints for a tighter formulation and used the solution of the Greedy heuristic to prioritize variables and determine a cutoff value for branch-and-bound.

**Table 5.** Summary of improvements achieved after utilizing the proposition (PR) and the solution of Greedy for variable prioritization and cutoff values (G-PC)

| | *Flow refueled* | | | *VMT* | | |
|---|---|---|---|---|---|---|
| | *MBIP* | *MBIP PR* | *MBIP PR+G-PC* | *MBIP* | *MBIP PR* | *MBIP PR+G-PC* |
| Number of opt. solution found | 20 | 20 | 25 | 18 | 21 | 25 |
| Avg. solution time–opt. solutions | 307.6 | 304 | 114 | 657 | 229 | 204 |
| Number of solutions improved | — | 5 | 5 | — | 6 | 6 |
| Avg. improvement (%) | — | 0.06 | 0.08 | — | 0.23 | 0.30 |

**Proposition 1.** *If* $x_{k_{m_i}^{q_i}} = x_{k_{m_j}^{q_j}}$ *and* $N_{m_i,t}^{q_i} = N_{m_j,t}^{q_j}$ *for* $q_i \neq q_j$, *then* $b_{m_i,t}^{q_i} = b_{m_j,t}^{q_j}$.

**Proof.** If $x_{k_{m_i}^{q_i}} = x_{k_{m_j}^{q_j}}$ and $N_{m_i,t}^{q_i} = N_{m_j,t}^{q_j}$, then either constraints (6) and (7) are used for both variables ($m_k \leq M_{qk}, k = i, j$) or constraints (8) and (9) are used for both variables ($m_k = M_{qk}+1,\ k = i, j$). In these cases, it is straightforward to show that the constraints for $b_{m_i,t}^{q_i}$ and $b_{m_j,t}^{q_j}$ are identical to each other. ■

To utilize the proposition, constraints in the form of $b_{m_i,t}^{q_i} - b_{m_j,t}^{q_j} = 0$ were added to the original formulation. As a result, the number of constraints increased considerably (from 169 577 to 1715 421). Although the number of constraints increased ninefold, as Table 5 shows, the problems were solved faster on average after utilizing the proposition, and it was possible to show that solutions for $p = 19$ and $p = 20$ under the VMT objective in Table 4 were optimal solutions. We observed more benefits of using the proposition for VMT problems since there was more room for improvements. In addition, after using the proposition, the best known solutions also improved for all problem instances except $p = 23$ under VMT ($-0.02\%$ worse result than before). These findings are consistent with the concept of more highly structured problems being more integer-friendly (ReVelle, 1993). Overall, utilizing the proposition helped reduce the solution time while improving the quality of solution.

When results of the Greedy heuristics were used to prioritize variables and provide a cutoff value for branch-and-bound, the solution time decreased significantly. Consequently, all $p = 1$ to 25 problems are solved optimally within 1 h with an average solution time of 114 s under the flow refueling objective and 204 s under the VMT objective function.

Finally, using these observations, we solved $p = 30$ to 60 with an increment of 10. The results show that the new MBIP formulation performs well for large values of $p$ as well. Although we were not able to obtain the optimal results for large values of $p$ using the VMT objective function within 1 h, the optimality gap is very small (0.04% to 0.63%). It is important to note that although the Greedy heuristic is a very fast method with solution times that grow linearly as $p$ increases, G_Sub-$n$ heuristics' solution times increase at a higher rate at each iteration at $p$ gets larger, due to the fact that more substitution alternatives need to be considered. Therefore, G_Sub-$n$ time performance worsens rapidly compared with the optimal MBIP formulation. For example, it took 6759 s to construct the solution for G_Sub-1 and 15 484 s for G_Sub-2 to build the $p = 60$ solution under the maximize flow refueled objective. The situation was even worse for the GA, which took over 40 000 s to get the solution for the same problem. In addition, although improvements were not very large (0.07 to 0.31%), the optimal formulation provided better results than all heuristics combined for all $p$ values between 30 and 60. Hence, the MBIP performs better compared to G_Sub-$n$ and GA in terms of time performance as well.

### 4.3. *Additional numerical experiments*

In this section, we use random networks to evaluate the performance of the proposed formulation. We continued comparing the performance of the new formulation with heuristics approaches; however, due to extensive time required to run GA and G_Sub-2,3, we focused only Greedy and G_Sub-1. The experiments' results provided in Table 6 show that solving problems takes more time either when the number of O–D pairs increases or when the number of nodes increases. In addition, as these two inputs are increased, generally speaking, the number of problems optimally solved goes down and the optimality gap increases. In the table, we present both the gap reported by Xpress-MP at the end of an hour (e.g., Min. gap) and the gap between the optimal solution and best solution found within an hour (e.g., Min. opt. gap). To find the optimal solution for problems, we used additional information about problems. For example, $p$ values were generated by solving a set covering problem (see Section 4); hence, providing this information to the solver as additional input helped find the optimal solutions (e.g., the solution should cover a certain percentage of flow), and we allowed problems to run until the optimal solutions were found, which took up to 8 h.

Similar to the Florida state highway network results, experiments based on random networks also point out that, although Greedy is fast, the optimality gap can be very large compared to the MBIP formulation. For example,

**Table 6.** Computational results from experiments with random networks when the MBIP formulation is used

| | *250 Nodes* | | *500 Nodes* | | *750 Nodes* | | *1000 Nodes* | |
|---|---|---|---|---|---|---|---|---|
| *Number of O–D nodes* | *40* | *80* | *40* | *80* | *40* | *80* | *40* | *80* |
| Number of optimal solutions | 9 | 9 | 8 | 8 | 9 | 9 | 7 | 6 |
| Min. solution time | 4 | 14 | 4 | 27 | 6 | 31 | 6 | 48 |
| Avg. solution time | 390 | 570 | 855 | 849 | 773 | 866 | 1268 | 1584 |
| Min. gap (%) | 2.62 | 2.88 | 2.57 | 4.73 | 7.7 | 5.83 | 3.94 | 0.57 |
| Max. gap (%) | 2.62 | 2.88 | 9.52 | 5.13 | 7.7 | 5.83 | 8.65 | 7.54 |
| Avg. gap (%) | 2.62 | 2.88 | 6.05 | 4.93 | 7.7 | 5.83 | 6.1 | 4.21 |
| Min. opt. gap (%) | 1.53 | 1.42 | 0.49 | 1.01 | 3.18 | 4.24 | 0.88 | 0 |
| Max. opt. gap (%) | 1.53 | 1.42 | 4.68 | 2.86 | 3.18 | 4.24 | 4.74 | 2.93 |
| Avg. opt. gap (%) | 1.53 | 1.42 | 2.96 | 1.93 | 3.18 | 4.24 | 2.57 | 0.97 |

while the overall optimality gap for MBIP is 2.35%, it is 16.48% for Greedy, and 12.7% for G_Sub-1. In addition, as Fig. 4 depicts, the time needed to find a solution using G_Sub-1 heuristics increases rapidly, which makes it impractical to use the G_Sub-$n$ heuristics for the FRLM when large $p$ values are considered. Although it is proven that the Greedy algorithm returns a solution within 37% of the optimal solution for location problems and flow interception problems (Nemhauser and Wolsey, 1978; Berman *et al.*, 1995), as presented in Table 7, the maximum optimality gap was 38.8%. This was due to the fact that the Greedy algorithm does not consider the relationship among open facilities to capture a larger flow requiring multiple facilities along a path. Experiments also show that G_Sub-1 may not provide a better result than Greedy for some problem instances. This is again related to the myopic nature of these heuristics where an improvement through a substitution at one iteration may degrade the performance in the future. Finally, it is worth mentioning that whereas the MBIP formulation provided the optimal solution for 65 out of 80 problem instances in an hour, the Greedy and G_Sub-1 heuristics together were not able to find even a single optimum solution for the test problems.

## 5. Model extensions

Having a mixed-integer programming formulation of the FRLM offers a new degree of flexibility to infrastructure providers to add features to the FRLM that customize the model to their fuel.

One important capability to add is fixed costs ($E_i$) that vary by location $i$. Land costs can vary widely by candidate site, especially in downtown areas and at major freeway interchanges where stations could capture large flow volumes traveling in all directions. In this case, the $p$ facility constraint (4) could be changed to a budget ($B$) constraint (12):

$$\sum_{i \in N} E_i x_i \leq B. \tag{12}$$

Budget constraints have been introduced in hierarchical facility models (Moore and ReVelle, 1982) and reserve-design models (ReVelle *et al.*, 2002) but have been found to reduce the "integer friendliness" of models (ReVelle, 1993). This simple constraint is not so simple to introduce into greedy/substitution algorithms and GAs that rely on crossover and mutation mechanisms that add, substitute, or exchange stations one at a time.

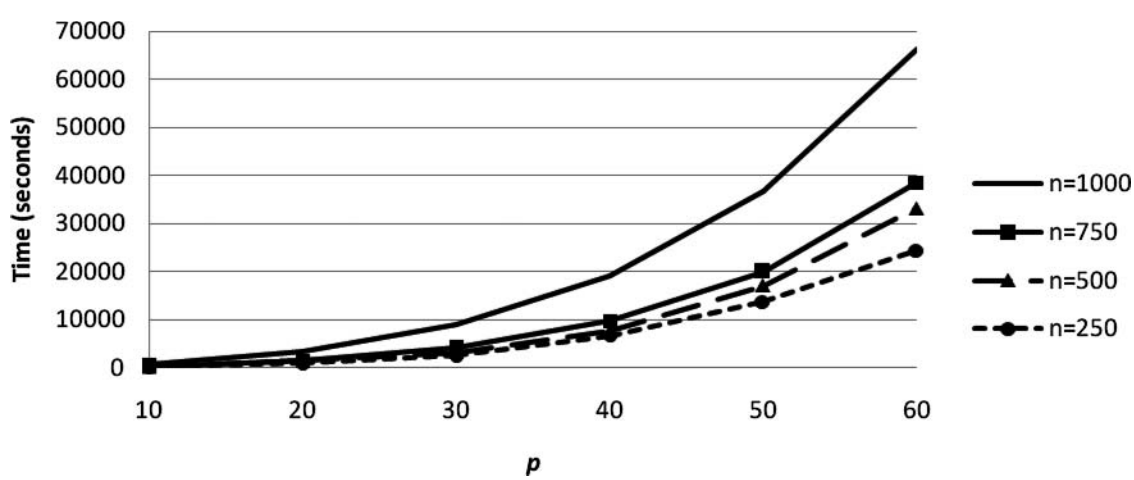


**Fig. 4.** Change in solution time of G_Sub-1 heuristics for various network sizes when $w = 80$.

**Table 7.** Computational results for random networks when Greedy and G_Sub-1 are used

| | *250 Nodes* | | *500 Nodes* | | *750 Nodes* | | *1000 Nodes* | |
|---|---|---|---|---|---|---|---|---|
| *Number of O–D nodes* | *40* | *80* | *40* | *80* | *40* | *80* | *40* | *80* |
| Greedy | | | | | | | | |
| Number of opt. solutions | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Min. solution time | 8 | 61 | 19 | 84 | 14 | 90 | 17 | 150 |
| Max. solution time | 68 | 652 | 158 | 911 | 125 | 989 | 114 | 1587 |
| Avg. solution time | 34 | 311 | 81 | 439 | 62 | 458 | 60 | 755 |
| Min. opt. gap (%) | 9.89 | 2.08 | 7.81 | 9.09 | 4.36 | 10.15 | 15.11 | 9.32 |
| Max. opt. gap (%) | 38.8 | 16.28 | 15.67 | 24.68 | 28.73 | 27.13 | 36.21 | 14.83 |
| Avg. opt. gap (%) | 19.87 | 9.83 | 11.07 | 16.47 | 15.57 | 19.05 | 27.58 | 12.37 |
| G_Sub-1 | | | | | | | | |
| Number of opt. solutions | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Min. solution time | 47 | 452 | 82 | 692 | 85 | 741 | 99 | 1836 |
| Max. solution time | 1658 | 24 382 | 3172 | 33 194 | 3232 | 38 455 | 3581 | 66 251 |
| Avg. solution time | 655 | 9305 | 1320 | 13 007 | 1278 | 14 308 | 1373 | 25 983 |
| Min. opt. gap (%) | 5.77 | 1.63 | 5.06 | 6.17 | 4.36 | 8.86 | 4.93 | 6.57 |
| Max. opt. gap (%) | 38.84 | 15.08 | 7.81 | 18.72 | 25.42 | 27.72 | 29.2 | 13.19 |
| Avg. opt. gap (%) | 18.26 | 6.89 | 6.2 | 12.65 | 13.8 | 19.47 | 14.64 | 9.67 |

For multiple decision makers or decision makers with multiple objectives, the new formulation can be modified for multi-objective analysis. For instance, given the demonstrated interest in locating stations close to residential populations, either the $p$-median (Hakimi, 1964) or max cover (Church and ReVelle, 1974) objective could be used as a second objective with the FRLM. A number of researchers have already used the $p$-median model, which minimizes total person-miles traveled from population centers to stations (Goodchild and Noronha, 1987; Nicholas and Ogden, 2006; Upchurch and Kuby, 2010). A multi-objective model maximizing O–D trips refueled and minimizing person-miles from population nodes to stations would require changing the objective function in Equation (13) and adding Constraints (14) to (16) to the formulation:

$$\max \alpha \sum_{q\in Q} f_q y_q - (1-\alpha) \sum_{i\in N} \sum_{j\in s} O_j d_{ij}\, U_{ij}, \tag{13}$$

subject to:

$$\sum_{i\in N} U_{ij} = 1, \quad \forall j \in S, \tag{14}$$

$$U_{ij} \le x_i, \quad \forall i \in N, j \in S, \tag{15}$$

$$U_{ij} \in \{0, 1\}, \tag{16}$$

where $\alpha$ is the weight on the maximizing-trips objective in the FRLM; $O_j$ is the population of node $j$; $d_{ij}$ is the distance from $i$ to $j$; $S$ is the set of population nodes, and $U_{ij}$ is an allocation variable that has a value of one if population node $j$ is allocated to station $i$ and is zero otherwise.

While the $p$-median model has received more attention in the fuel station location literature, the max cover model might be more appropriate, especially at regional scales of analysis, because it will not allocate a population node to a far-distant station even if there are no other closer stations. The max cover model considers a Euclidean radius or a maximum travel time or distance to specify which stations are capable of "covering" which demand nodes. It then maximizes the population that can be covered with a given number of stations. The max cover model could be combined with the FRLM in a multi-objective model by changing the objective function to Equation (17) and adding Equations (18) and (19):

$$\max w \sum_{q\in Q} f_q y_q + (1-w) \sum_{j\in S} O_j Z_{j,} \tag{17}$$

subject to:

$$Z_j \le \sum_{i\in N} e_{ij} x_i, \quad \forall j \in S, \tag{18}$$

$$Z_j \in \{0, 1\}, \tag{19}$$

where $Z_j$ is the covering variable that has a value of one if demand node $j$ is covered and zero if not, and $e_{ij}$ are the covering matrix coefficients that have a value of one if station $i$ can cover demand node $j$ and zero if not.

## 6. Conclusions and future research

The FRLM has unique characteristics compared with traditional node-based location models; hence, a new research area has arisen in recent years. For formulating the FRLM, previous research has used a two-stage approach: generating combinations of candidate locations and then using these combinations to locate $p$ facilities to maximize the flow refueled or recharged. The two-stage approach using facility combinations has been applied to small networks successfully and pointed directions for further research. However, as recent applications to larger networks showed, generating the combinations and incidence matrices is time-consuming, and unless a much more efficient way of generating combinations can be found, the applicability of existing formulations to large-scale problems is limited. Therefore, heuristics approaches have been developed for large networks to optimize the location of fuel stations on a network given the limited range of AFVs.

In this research, we present a fundamentally new way of formulating the FRLM, which eliminates the need to generate facility combinations as input data and makes it possible to solve large problems that were impossible to solve with the previous formulation. Although the proposed formulation eliminates the need for developing heuristics approaches for mid-sized networks, we introduced a proposition to create a tighter formulation, which reduces the time required to solve FRLM problems, improves the solution quality, and makes it possible to solve problems on substantially larger networks. In addition, the results showed that the gap between the optimal solution and previously presented heuristics in the literature can be several percentage points.

The model presented here changes the way the FRLM is formulated and enables analysts to solve large-scale problems to optimality within a short period of time. Equally important, it provides modelers with the greater flexibility afforded by MIP formulations. Side constraints, budget constraints, and multiple objectives can be easily added to an MIP formulation in a commercial program such as Xpress-MP or Cplex with just a few lines of code. Given the variety of technologies, driving ranges, and business models across the variety of liquid, gaseous, and electric-powered vehicles vying to replace gasoline-fueled vehicles, the ability to easily customize a FRLM to a particular company, technology, or region will be a valuable contribution of this new formulation.

The proposed model can be extended in several ways. First, alternative model(s) can be created to take advantage of the observations behind the proposition, which may help to reduce the problem size as well as improve the solution time. Second, it was assumed here that each facility has unlimited capacity; hence, a model that solves the capacitated version of the FRLM (Upchurch *et al.*, 2009) more efficiently would be helpful. Finally, although our formulation is solved in a few minutes using over-the-counter software, developing heuristics approaches or decomposition techniques may be necessary to solve very large problems (e.g., the entire U.S.A. network).

## References


Berman, O., Bertsimas, D. and Larson, R.C. (1995) Locating discretionary service facilities, II: maximizing market size, minimizing inconvenience. *Operations Research*, **43**(4), 623–632.

Berman, O., Larson, R.C. and Fouska, N. (1992) Optimal location of discretionary service facilities. *Transportation Science*, **26**, 201–211.

Cascetta, E. and Nguyen, S. (1988) A unified framework for estimating or updating origin/destination matrices from traffic counts. *Transportation Research Part B: Policy and Practice*, **22**(11), 437–455.

Church, R. and ReVelle, C. (1974) The maximal covering location problem. *Papers of the Regional Science Association*, **32**, 101–118.

Dijkstra, E. (1959) A note on two problems in connexion with graphs. *Numerische Mathematik*, **1**, 269–271.

Fotheringham, A.S. and O'Kelly, M.E. (1989) *Spatial Interaction Models: Formulations and Applications*, Kluwer, Dordrecht, the Netherlands.

Goodchild, M.F. and Noronha, V.T. (1987) Location-allocation and impulsive shopping: the case of gasoline retailing, in *Spatial Analysis and Location-Allocation Models*, Ghosh, A. and Rushton, G. (eds), Van Nostrand Reinhold, New York, NY, pp. 121–136.

Greene, D.L., Leiby, P.N., James, B., Perez, J., Melendez, M., Milbrandt, A., Unnasch, S., Rutherford, D. and Hooks, M. (2008) Analysis of the transition to hydrogen fuel cell vehicles and the potential hydrogen energy infrastructure requirements. Oak Ridge National Laboratory Report No. ORNL/TM-2008/30), Oak Ridge National Laboratory, Oak Ridge, TN.

Hakimi, S.L. (1964) Optimum locations of switching centres and the absolute centres and medians of a graph. *Operations Research*, **12**, 450–459.

Hodgson, M.J. (1990) A flow capturing location allocation model. *Geographical Analysis*, **22**, 270–279.

Hodgson, M.J., Rosing, K.E. and Storrier, A.L.Y. (1994) Testing a bicriterion location-allocation model with real-world network traffic: the case of Edmonton, multicriteria analysis, in *Proceedings of the 11th International Conference on Multi Criteria Decision Making, 1–6 August 1994, Coimbra, Portugal*, Climaco, J. (ed.), Springer-Verlag, New York, pp. 484–495.

Kitamura, R. and Sperling, D. (1987) Refueling behavior of automobile drivers. *Transportation Research, Part A: General*, **21**, 235–245.

Kruskal, J.B. (1956) On the shortest spanning tree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, **7**(1), 48–50.

Kuby, M. and Lim, S. (2007) Location of alternative-fuel stations using the flow-refueling location model and dispersion of candidate sites on arcs. *Networks and Spatial Economics*, **7**, 129–152.

Kuby, M., Lines, L., Schultz, R., Xie, Z., Kim, J. and Lim, S. (2009) Optimal location strategy for hydrogen refueling stations in Florida. *International Journal of Hydrogen Energy*, **34**, 6045–6064.

Kuby, M.J. and Lim, S. (2005) The flow-refueling location problem for alternative-fuel vehicles. *Socio-Economic Planning Sciences*, **39**, 125–145.

Lim, S. and Kuby, M. (2010) Heuristic algorithms for siting alternative-fuel stations using the flow-refueling location model. *European Journal of Operational Research*, **204**, 51–61.

Lin, Z., Ogden, J., Fan, Y. and Chen, C.-W. (2008) The fuel-travel-back approach to hydrogen station siting. *International Journal of Hydrogen Energy*, **33**, 3096–3101.

Melendez, M. (2006) Transitioning to a hydrogen future: learning from the alternative fuels experience. National Renewable Energy Laboratory Technical Report NREL/TP-540-39423, National Renewable Energy Laboratory, Golden, CO.

Melendez, M. and Milbrandt, A. (2005) Analysis of the hydrogen infrastructure needed to enable commercial introduction of hydrogen-fueled vehicles. National Renewable Energy Laboratory Technical Report CP-540-37903, National Renewable Energy Laboratory, Golden, CO.

Moore, G.C. and ReVelle, C. (1982) The hierarchical service problem. *Management Science*, **28**, 775–780.

Nemhauser, G.L. and Wolsey, L.A. (1978) Best algorithms for approximating the maximum of a submodular set function. *Mathematics of Operations Research*, **3**(3), 177–188.

Nicholas, M.A., Handy, S.L. and Sperling, D. (2004) Using geographic information systems to evaluate siting and networks of hydrogen stations. *Transportation Research Record*, **1880**, 126–134.

Nicholas, M.A. and Ogden, J. (2006) Detailed analysis of urban station siting for California hydrogen highway network. *Transportation Research Record*, **1983**, 121–128.

ReVelle, C. (1993) Facility siting and integer-friendly programming. *European Journal of Operational Research*, **65**, 147–158.

ReVelle, C.S. and Swain, R. (1970) Central facilities location. *Geographical Analysis*, **2**, 30–42.

ReVelle, C.S., Williams, J.C., and Boland, J.J. (2002) Counterpart models in facility location science and reserve selection science. *Environmental Modeling and Assessment*, **7**, 71–80.

Romm, J. (2006) The car and fuel of the future. *Energy Policy*, **34**, 2609–2614.

Sperling, D. and Kitamura, R. (1986) Refueling and new fuels: an exploratory analysis. *Transportation Research, Part A: General*, **20**, 15–23.

Upchurch, C. and Kuby, M. (2010) Comparing the *p*-median and flow-refueling models for locating alternative-fuel stations. *Journal of Transport Geography*, **18**(6), 750–758.

Upchurch, C., Kuby, M.J. and Lim, S. (2009) A capacitated model for location of alternative-fuel stations. *Geographical Analysis*, **41**, 85–106.

Zeng, W. (2009) *GIS Based Facility Location Planning*, VDM Verlag, Saarbrücken, Germany.

## Biographies

Ismail Capar received his B.S. and M.S. degrees in Industrial Engineering from Istanbul Technical University and Sabanci University, respectively. Upon completion of his Ph.D. in Industrial and Systems Engineering at Mississippi State University, he joined the Industrial Distribution Program at Texas A&M University as an Assistant Professor in 2007. His current research focuses on location and pricing decisions for alternative-fuel vehicles including electric vehicles and transportation issues during events of national significance. He teaches courses in the areas of supply chain management, logistics, and quality issues in industrial distribution. He is a member of ASEE, APICS, IIE, and INFORMS.

Michael Kuby received an A.B. from the University of Chicago in 1980 and a Ph.D. from Boston University in 1988, both in Geography. He is a Professor in the School of Geographical Sciences and Urban Planning at Arizona State University, where he has taught since 1988. He is also Director of the Interdisciplinary Certificate Program in Transportation Systems. His research interests are in transportation, energy, and facility location, with a focus on optimization models for planning sustainable infrastructure systems. He is Location Area Editor for *Networks and Spatial Economics* and serves on the editorial boards of the *International Regional Science Review* and the *Journal of Transport Geography*. His work has been funded by the U.S. Department of Energy, the U.S. Maritime Administration, NASA, the NSF, and the World Bank.