Transportation Research Part D 74 (2019) 48–64

Contents lists available at ScienceDirect

Transportation Research Part D

journal homepage: www.elsevier.com/locate/trd

A multi-scale framework for fuel station location: From highways to street intersections

Qunshan Zhao[a,*], Scott B. Kelley[b], Fan Xiao[c], Michael J. Kuby[a]

[a] Spatial Analysis Research Center, School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ 85287-5302, USA
[b] Department of Geography, University of Nevada, Reno, Reno, NV 89557, USA
[c] National Marine Environmental Forecasting Center, 100081 Beijing, China

ARTICLE INFO

Keywords:
Alternative fuel
Infrastructure
Hydrogen
Path
Flow
Connecticut

ABSTRACT

Electric drive vehicles (plug-in electric vehicle or hydrogen fuel cell vehicles) have been promoted by governments to foster a more sustainable transportation future. Wider adoption of these vehicles, however, depends on the availability of a convenient and reliable refueling/recharging infrastructure. This paper introduces a path-based, multi-scale, scenario-planning modeling framework for locating a system of alternative-fuel stations. The approach builds on (1) the Flow Refueling Location Model (FRLM), which assumes that drivers stop along their origin-destination routes to refuel, and checks explicitly whether round trips can be completed without running out of fuel, and (2) the Freeway Traffic Capture Method (FTCM), which assesses the degree to which drivers can conveniently reach sites on the local street network near freeway intersections. This paper extends the FTCM to handle cases involving clusters of nearby freeway intersections, which is a limitation of its previous specification. Then, the cluster-based FTCM (CFTCM) is integrated with the FRLM and the DFRLM (FRLM with Deviations) to better conduct detailed geographic optimization of this multi-scale location planning problem. The main contribution of this research is the introduction of a framework that combines multi-scale planning methods to more effectively inform the early development stage of hydrogen refueling infrastructure planning. The proposed multi-scale modeling framework is applied to the Hartford, Connecticut region, which is one of the next areas targeted for fuel-cell vehicle (FCV) market and infrastructure expansion in the United States. This method is generalizable to other regions or other types of fast-fueling alternative fuel vehicles.

1. Introduction

The transportation sector became the largest source of CO_2 emissions in the United States in February 2016, surpassing electric power generation (US Department of Energy, 2016). Of the five energy-use sectors (including residential, commercial, and industrial), transportation is also the only one for which CO_2 emissions continue to rise. Not coincidentally, transportation is the sector that relies most heavily on a single form of energy, with petroleum accounting for over 90% of total consumption (US EIA, 2018). Unlike the other four sectors where energy-using equipment is stationary, transportation vehicles are mobile and depend on a widespread refueling infrastructure. For most alternative-fuel vehicles (AFVs) in most regions, the refueling infrastructure is relatively sparse or even non-existent. Early stage AFV refueling infrastructure development is essential and necessary to encourage

* Corresponding author at: Urban Big Data Centre, School of Social and Political Sciences, University of Glasgow, Glasgow G12 8RZ, UK.
E-mail addresses: qszhao@asu.edu, Qunshan.Zhao@glasgow.ac.uk (Q. Zhao), scottkelley@unr.edu (S.B. Kelley), xiaofan10243@163.com (F. Xiao), mikekuby@asu.edu (M.J. Kuby).

https://doi.org/10.1016/j.trd.2019.07.018
Received 30 April 2019; Received in revised form 18 July 2019; Accepted 20 July 2019
Available online 25 July 2019
1361-9209/

individual car buyers and freight companies to consider switching from vehicles that operate with liquid petroleum fuels to AFVs. In order to best encourage adoption broadly across regions, initial stations should be located where they can provide the greatest benefit.

In the United States, California led the way as the first commercial market for fuel-cell vehicles (FCVs) and hydrogen refueling infrastructure. For the initial rollout in 2015, California funded six stations and drivers bought or leased 200 FCVs, rising to 25 stations and 925 FCVs in 2016, 31 stations and 2473 FCVs in 2017, and 39 station and 6558 vehicles as of March 1, 2019 (California Air Resources Board, 2018). The New York-Boston corridor is one of the regions targeted for the next rollout of FCVs and stations, and Connecticut has an active hydrogen coalition ("Connecticut Hydrogen-Fuel Cell Coalition," n.d.). As in California, an initial network of hydrogen fueling stations is needed before introducing the first FCVs for the consumer market in Connecticut.

Planning an effective network of alternative-fuel stations is a multi-scale location problem. Alternative-fuel vehicle drivers have diverse travel needs, and the ability to refuel on a variety of trips across a region is an essential component of a convenient refueling network. Once an effective refueling network is built, stations can be used for neighborhood refueling, urban trips, and long-distance travel in the region. In addition, drivers use a hierarchy of roads from local streets to arterials to limited-access highways (freeways) to complete trips across the region, and generally want to reach refueling stations with a minimum of wasted travel time. Thus, it is important to build the refueling station at locations that are easy to reach on the local street network. This paper proposes a multi-scale method for planning a starter network of stations to serve local, metropolitan, and inter-city trips, considering travel and station access on all classes of roads. The method is based on the flow-refueling location model (FRLM), in which drivers make trips from origins to destinations and stop along their shortest paths to refuel, considering the driving range of vehicles in determining if the stations can cover the set of trips in the network (Kuby and Lim, 2005). The method proposed here for multi-scale planning integrates the FRLM with an extension that allows deviations from shortest paths (Kim and Kuby, 2013, 2012). The typical FRLM and DFRLM modeling approach works well for recommending generalized station locations on a simplified regional road network, but does not check whether suitably accessible nearby sites exist where stations could be built. To address this, we incorporate and extend the Freeway Traffic Capture Method (FTCM), which measures how conveniently drivers can reach street intersections surrounding freeway intersections for all possible travel routes passing through them, to identify accessible sites on the local street networks (Kelley, 2017). Using these integrated methods, we recommend initial sets of five and ten stations for the greater Hartford, Connecticut region. The method is generalizable to other regions and other fast-refueling AFVs.

2. Literature review

Many different methods have been proposed for planning a system of hydrogen or other alt-fuel stations. One common way to categorize location models is by the spatial units representing the demand for fuel. Classic location models such as the *p*-median (Revelle and Swain, 1970) and max-cover models (Church and ReVelle, 1974) use zonal demand aggregated to centroids. Points representing areal units such as cities, towns, neighborhoods, or census tracts are weighted according to their populations, number of vehicles, or likelihood to purchase FCVs. In median-type station location models, these weighted demands are then assigned to their closest station (He et al., 2016; Nicholas et al., 2004), while in covering models, stations must be "close enough" in terms of a critical distance or travel time threshold for being able to serve a demand point (Frade et al., 2011; Stephens-Romero et al., 2010).

Another group of models locates stations to serve passing traffic, not proximity to homes or companies. Higher amounts of traffic volume, measured in terms of average annual daily traffic (AADT) or vehicle-miles traveled (VMT), signals more demand for fueling (Boostani et al., 2010; Brey et al., 2016; Goodchild and Noronha, 1987; Lin et al., 2008). While it makes sense to locate stations on busy roads or at the intersection of busy roads, several kinds of demand-counting errors can arise because drivers link together a sequence of arcs into a route. Thus, the same vehicle may be in the traffic count on adjoining arcs but would be unlikely to refuel on both arcs. Likewise, an arc-based model will count all of an arc's demand as covered by a nearby station even if some of the traffic on the arc is making a long-distance trip that cannot be completed without other stations being located.

Path-based models were developed to handle the multi-arc nature of road travel. Also known as flow-based or trip-based models, they use origin-destination routes as the fundamental units of demand. An origin-destination flow matrix defines the demand weights by the volume of passenger vehicle trips or freight tonnage. In the original flow-capturing location model (FCLM) by Hodgson (1990) and the flow-intercepting (FILM) model by Berman et al. (1992), weighted demands are counted only once. Kuby and Lim (2005) extended the FCLM/FILM to the case of refueling stations by introducing an explicit vehicle driving range and considering that multiple stations may be needed to complete a long-distance route. In the basic flow-refueling location model (FRLM), only the shortest or least-travel-time path is considered, and an O-D flow is not considered covered unless there are stations along the path that enable a driver to complete the round trip without running out of fuel. Kim and Kuby (2012, 2013) then introduced the deviation flow-refueling location model (DFRLM) to allow drivers to deviate off the shortest path to reach a refueling station. Numerous studies have developed faster solution methods for the FRLM (Capar et al., 2013; MirHassani and Ebrazi, 2012; Yıldız et al., 2016). Other researchers applied and extended the FRLM and DFRLM by weighting flow volumes by trip length which yields VMT or tkm measures (Lines et al., 2007) and by considering station capacities (Upchurch et al., 2009), multiple driving ranges (de Vries and Duijzer, 2017), congestion (Fan et al., 2017), trucking (Fan et al., 2017), and regional equity (Hong and Kuby, 2016; Kuby et al., 2016).

Choosing the appropriate model for optimizing a set of stations depends on numerous factors: the type of vehicles, the geographic scale of the region, the refueling or recharging speed, the driving range, whether home refueling or recharging is commonplace, and the behavior of drivers. There is behavioral and survey evidence supporting the assumptions underlying node-based approaches (Sperling and Kitamura, 1986), though often these are based on surveys of gasoline drivers who have hundreds or thousands of stations to choose from in large cities (Kuby, 2019). Kuby et al. (2013) surveyed drivers in Southern California who were refueling at

CNG stations, comparing their driving and refueling trips to drivers surveyed at nearby gasoline stations. They found that the CNG drivers refueled at higher tank levels, at the same stations more frequently, farther from home, more frequently in the middle of trips, and more on work-based trips, compared with gasoline drivers. Kelley and Kuby (2013) found that, when no station exists that is both closest to home and most on the way, CNG drivers favored the station with the least deviation off their shortest path by a 10:1 margin over the station closest to home. This suggests that early AFV adopters adapt to the greatly reduced station availability by changing their pre-AFV refueling habits.

There are limits to the recommendations provided by these modeling approaches, though. Most station location models use a relatively coarse representation of a region's road network to locate a limited number refueling stations across a city or region (Jochem et al., 2016; Sathaye and Kelley, 2013; Wang and Lin, 2009). Roads included in these networks are typically highways, freeways, and major arterials that are represented as arcs, while their intersections are represented as nodes. These nodes represent the set of candidate sites at which refueling stations can be built. In the case of selecting a node that represents an intersection of at least one limited-access highway, however, an additional step is necessary in order to identify suitable refueling station sites. Stations cannot be built in the middle of freeway intersections but instead must be located on nearby surface streets, which require navigating freeway entrances and exits, access ramps, and congested roadways to reach. Travel in these environments can be complex and time-consuming and can vary greatly from one freeway intersection to another. Freeway intersections, however, are critical to station planning. Trip-based approaches, in particular, frequently select these nodes because of the high path volumes flowing through them (Kuby et al., 2009). To systematically identify effective station sites near major freeway intersections, Kelley (2017) developed the Freeway Traffic Capture Method (FTCM) to help address the inherent scale dependency between regional optimization models and local road networks, where stations must be built. This study found only 7% of street intersections near each of the 72 freeway intersections in Los Angeles could be conveniently reached by drivers on all possible travel paths that passed through the freeway intersection, signaling that an ad-hoc station site selection process near freeway intersections is unlikely to identify a convenient location.

While helpful in providing key stakeholders with a tool to identify best station sites near freeway intersections, the FTCM has a number of limitations. First, its intent is to define a best station location on the local street network for each optimization network node that represents a freeway intersection. Therefore, it assesses each freeway intersection independently. In a given regional freeway network, however, there may be multiple freeway intersections near one another. In such cases, one station on the local street network may be a suitable location for travel paths passing through multiple freeway intersections. Particularly in limited refueling infrastructures where only a few stations can be built, it may be best to consider general areas where a number of major highways, freeways, and arterials intersect as a cluster. Sites chosen on the local street network could then be suitable for all paths traveling through the cluster. Second, the FTCM has not yet been applied to the outputs from a regional scale optimization model to determine whether convenient sites exist on the street network near a selected optimization node. This is a priority research consideration for applying station planning methods to a relatively coarse representation of a road network.

3. Methods

This paper proposes a multi-scale method for planning a network of fuel stations. The method integrates the FRLM and DFRLM models with a modified Cluster-based Freeway Traffic Capture Method (CFTCM) introduced here. Table 1 summarizes the abbreviations, terminology, and notation used in the multi-scale method.

Fig. 1 shows the methodology framework that integrates the FRLM, DFRLM, and CFTCM methods across multiple scales for planning a robust refueling infrastructure. Phase 1 is a filtering process that runs the FRLM and DFRLM without any restriction of candidate sites on a network of major urban and regional roads. Twenty-four scenarios are systematically analyzed based on the most critical modeling assumptions: vehicle driving ranges (100 and 150 miles), number of hydrogen stations to build (5, 10 or 15 stations in the early phase of development), choice of objective function (Max trips or Max VMT), and model (FRLM or DFRLM). Following Phase 1, any optimization node that was never optimal in any of the 24 scenarios is eliminated from subsequent phases of the analysis, while any node that was optimal under any scenario is then evaluated by the CFTCM.

The CFTCM process begins by grouping any remaining optimization nodes located on freeways into clusters if they are close enough to each other. Then, street intersection nodes around each optimization node in the cluster are evaluated to determine if all possible travel on all routes through the cluster can reach them conveniently. Each optimization node is then given the score of its best-performing street network node, and the best optimization node in the cluster is selected to represent the cluster in Phase 2. Using this reduced set of optimization nodes as candidate sites, we again run the 24 FRLM and DFRLM scenarios. Based on the Phase 2 scenario results, we then adapt Owen and Daskin's (1998) *scenario-planning approach*, in which "the objective is to determine robust facility locations which will perform well (according to the defined criteria) under a number of possible parameter realizations" (p. 435). Here, we determine the best-performing set based primarily on how frequently locations are optimal, while also considering how frequently they occur in diverse scenarios. In other words, consistent appearance of a station in only one similarly defined set of scenarios is not considered to be as robust as a station that occurs with similar frequency across a range of scenarios.

Finally, in Phase 3, we evaluate the chosen system of stations according to a more limited set of FRLM and DFRLM scenarios to check how well they perform together under different assumptions. Depending on those results, the final set of stations could be revised if necessary.

Table 1
Abbreviations and Definitions.

Abbreviations		FRLM Notation	
AADT	average annual daily traffic	y_q	1 if path q is refuelable 0 otherwise
AFV	alternative fuel vehicle	X_k[1]	1 if facility k is open 0 otherwise
CFTCM	Cluster Freeway Traffic Capture Method	f_q	flow volume on path q
CNG	Compressed natural gas	d_q	distance of path q
DFRLM	Deviation Flow Refueling Location Model	p	number of facilities to open
FCV	fuel cell vehicle	A_q	set of arcs on round trip q
FCLM/FILM	Flow Capturing/Intercepting Location Model	Z_{ij}^q	set of nodes that can cover a vehicle crossing directional arc ij on path q given the assumed driving range of vehicles
FRLM	Flow Refueling Location Model	N	set of network nodes (optimization nodes)
FTCM	Freeway Traffic Capture Method		
O-D	origin–destination	E	set of existing or planned stations
VMT	vehicle miles traveled	**CFTCM Notation**	
		k	local street intersection node
DFRLM Notation		A_{kh}	score for street intersection node k in cluster h ($0 \le A_{kh} \le 1$)
A, B, C	parameter settings in the deviation decay function	t_{ijh}	shortest travel time (in minutes) from artificial origin i to artificial destination j through cluster h
DT	deviation time (in hours)	t_{ikj}	shortest travel time (in minutes) from artificial origin i to local street network intersection node k (Fig. 2) to artificial destination j through cluster h
DT_{max}	Maximum deviation time	X_{ikjh}	1 if t_{ikjh} - $t_{ijh} \le \Delta_{max}$, 0 otherwise
t_q	travel time of fastest path for O-D pair q	I_h	set of inbound origins for cluster h
Terminology		J_h	set of outbound destinations for cluster h
Optimization Node	a node in the simplified regional network used in the optimization models	N_{ih}	number of possible freeway directions through cluster h[2] starting from artificial origin i
Street Network Node	a node that represents an intersection on the detailed local street network surrounding freeway intersections	Δ_{max}	maximum deviation from shortest travel time (in minutes) through cluster h
Freeway intersection or confluence	a location where two or more limited-access freeways intersect or converge	c_{max}	maximum network travel distance between optimization nodes in the same cluster (miles)
Cluster	a subset of optimization nodes that (a) connect to at least one limited-access freeway arc, and (b) are within c_{max} network miles of each other. *Note*: an optimization node may be a member of more than one cluster, and a cluster may contain only a single optimization node.	b_{max}	maximum buffer distance (in miles) from optimization nodes to street network nodes

[1] The definitions of *i, j,* and *k* are different in the FRLM and CFTCM notation, but in each model *k* refers to the potential station locations being evaluated at that geographical scale.

[2] The N_{ih} notation is needed because all *ij* combinations of directions of travel may not be possible through a freeway intersection cluster.

3.1. The flow-refueling and deviation flow-refueling location models

This section introduces the two location optimization models (FRLM and DFRLM). The first model, the standard FRLM, requires pre-generation of a single shortest (fastest) path for each origin-destination (O-D) path q. In this paper, we used the arc cover-path cover (AC-PC) formulation of the FRLM (Capar et al., 2013), an exact mixed-integer programming (MIP) solution method that is fast and globally optimal (Arslan and Karaşan, 2016).

$$Max \sum_{q \in Q} f_q(d_q) y_q \tag{1}$$

Subject to:

$$\sum_{k \in Z_{i,j}^q} X_k \ge y_q \ \forall\ q \in Q,\ a_{i,j} \in A_q \tag{2}$$

$$\sum_k X_k = p \tag{3}$$

$$X_k = 1 \ \forall\ k \in E \tag{4}$$

$$y_q, X_k \in \{0, 1\} \ \forall\ q \in Q,\ k \in N \tag{5}$$

The objective function (1) maximizes the origin-destination flow volume or VMT that can be covered, i.e., the trips that are refuelable given the station locations, the shortest paths, and the driving range. The objective can be generated with or without

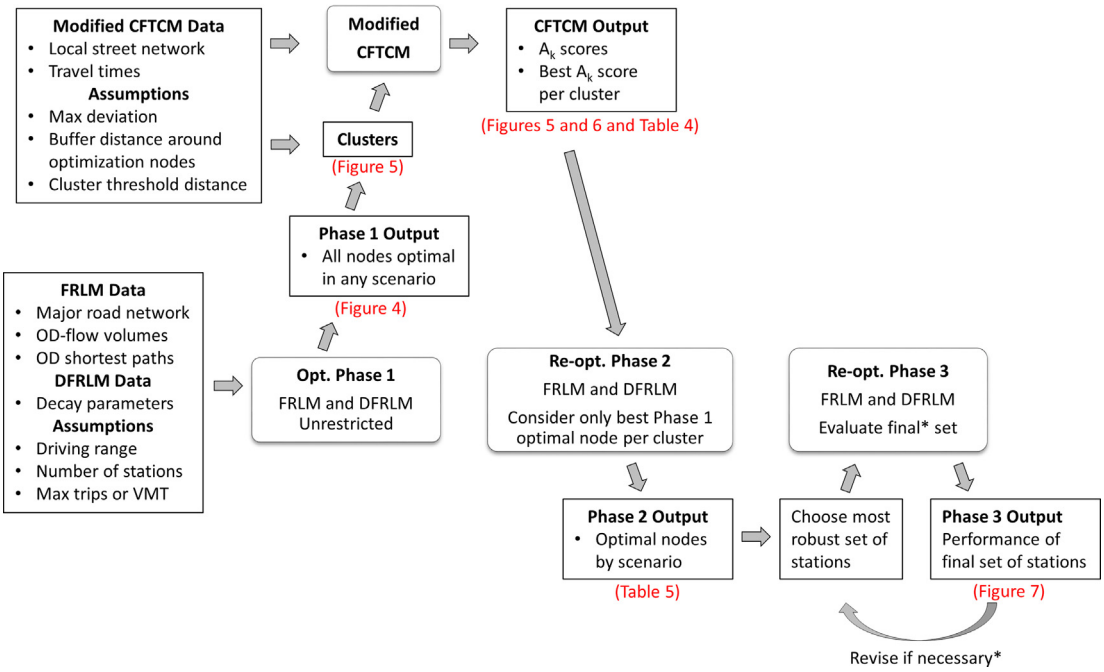


Fig. 1. Multi-scale refueling infrastructure planning framework.

multiplying by the path length, d_q. The standard form without d_q multiplies flow volume of path q by a 0–1 variable indicating whether the shortest path is covered or not. Summed over all paths q, this maximizes total covered trips and counts all trips equally, thus assuming that drivers care about accomplishing the purpose of a trip regardless of how far away that activity is. By including d_q in the objective function, the trip volumes are weighted by their distances, thus maximizing total covered VMT and emphasizing longer trips and conventional fuel replaced. Both objectives are used in this paper.

Constraints (2) ensure that the y_q variables equal 1 only if the round trip on path q can be completed without running out of fuel. If stations are located on the path that can fully cover each directional arc ij, then and only then is the full round-trip path considered covered. The exogenously determined set Z_{ij}^q includes all stations "downstream" of directional arc ij on round-trip path q that are within driving range of the far end (j) of the arc. The downstream part of the round-trip path is allowed to wrap around the origin and/or destination, e.g., in the round-trip path 1-2-3-4-3-2-1, the directional arc 1–2 could be refueled by a station at node 3 if the distance 3-2-1-2 does not exceed the driving range (see Capar et al., 2013 for details). All refueling stops on a round-trip path are assumed to top up the tank to full. Constraint (3) requires that exactly p stations are opened. Constraints (4) force open any stations k in the set E of existing and planned stations. Constraints (5) define y_q and X_k as 0–1 variables, whereby trips are either covered or not and stations are either open or not.

In contrast with the FRLM, the DFRLM allows drivers to detour from their fastest path to access one or more open stations while traveling from their origin to their destination. Kim and Kuby (2012) originally formulated the DFRLM as an exact MIP model that requires enumeration of a set of possible deviation paths. Because of the very large number of possible deviation paths even for a single O-D pair, the MIP was not solvable on large real-world networks, leading Kim and Kuby (2013) to develop a greedy-substitution heuristic algorithm. The model allows any feasible deviation path to be taken, subject to a restriction on the maximum difference in absolute or relative distance or travel time from the shortest path. Furthermore, the model includes a penalty function that reduces the covered flow volume as a function of the absolute or relative size of the deviation, which can take linear, exponential, inverse distance, and sigmoidal functional forms. Subject to the maximum deviation and penalty function, any kind of deviation is allowed, including deviations that leave the shortest path and return to it by the same route, leave the shortest path and return to it at a later point, or take a completely different path through the open station(s). At each stage of the greedy-substitution algorithm a new station is added after testing all possible station additions and substitutions. For any given set of stations being evaluated, an artificial network is constructed for each O-D pair using the set of stations plus the origin and destination nodes. Between any pair of nodes in this artificial network, an artificial arc is added if the shortest path between them satisfies the driving range conditions. On each of these artificial networks, the shortest path is by definition feasible given the open stations and the driving range specified. Like the FRLM, the DFRLM's objective can maximize covered trips or VMT, and the model can force open the set of existing or planned stations. Additional algorithm details can be found in Kim and Kuby (2013).

An important point for both FRLM and DFRLM models used here is that the objective function value is reported in terms of the percent of demand that is *potentially* refuelable given the assumed driving range. The stations in these models are uncapacitated, and are designed to provide basic geographic coverage for as many potential early FCV adopter trips as possible in the initial rollout stage, assuming that initial penetration rates are extremely low and station capacity is not yet an issue.

3.2. The cluster-based freeway traffic capture method

The original specification of the FTCM assessed the accessibility of street network nodes surrounding major intersections of two or more freeways. Each freeway intersection, however, was analyzed separately. In this study, we extend the FTCM to evaluate a cluster of nearby freeway intersections simultaneously. These clusters are broadened to also include nearby freeway-street intersections that may have been chosen in Phase 1. The CFTCM generates a score for each street network node that measures its ability to serve as a convenient station site for all travel paths that pass through a cluster of optimization nodes where multiple highways, freeways, and major street intersections intersect in close proximity. The data inputs to the CFTCM include: (1) a set of optimization nodes that can be filtered to a smaller set after evaluation, (2) a simplified optimization network of major streets and highways, and (3) a detailed street and highway network. Both of the network's datasets contain travel speeds and lengths along all arcs. Then, the user sets a number of flexible parameters. These include: (1) the maximum deviation (Δ_{max}) in minutes required to reach a street network node from a shortest travel time path through the cluster h, (2) the maximum network travel distance (c_{max}) between optimization nodes in the same cluster, and (3) a maximum Euclidean buffer distance (b_{max}) from optimization nodes to street network nodes.

The CFTCM first produces a set of clusters and the optimization nodes that belong to them, then generates accessibility scores A_{kh} for each street network node in each cluster. Then, summary statistics for all street network nodes within each cluster are generated. These scores are used to narrow the optimization nodes to a smaller set of candidate nodes that are sufficiently convenient for drivers through each cluster of optimization nodes.

3.2.1. Optimization node clusters

Given a set of network optimization nodes, such as those identified by Phase 1 of this analysis, all nodes adjoining at least one limited-access highway or freeway arc are evaluated by the CFTCM. For each node, the approximate center of its individual highway or street intersection is stored as point data in a GIS environment. Next, the CFTCM groups nearby optimization nodes into a cluster if they are within c_{max} network travel miles of one another. There is no limit to the number of nodes in a cluster, and individual optimization nodes can belong to multiple clusters. Clusters consisting of only one optimization node are also allowed, in which case the CFTCM simplifies to the original FTCM metric in Kelley (2017) for individual freeway intersections.

Three categories of optimization nodes do not need to be assessed by the CFTCM. First, for nodes adjacent to only arterial roads or surface streets, stations can be built directly at these intersections without drivers having to navigate limited access highways or freeways. Second, the locations of existing or planned station locations are already fixed. Third, long-distance inter-city trip destinations are used to represent the entirety of metropolitan areas outside the primary study area.

3.2.2. CFTCM specification

The CFTCM identifies the best locations on the local street network within close proximity of optimization node clusters. Using a detailed street network dataset in GIS, all street intersections, k, are selected that lie within the maximum Euclidean buffer distance, b_{max}, of any optimization node in cluster h. Next, the analyst must identify a set of artificial origins, I_h, and destinations, J_h, for trips that pass through the cluster. These O-D points are placed along all inbound and outbound major roadways that are included in the simplified optimization network that pass through any of the optimization nodes in the cluster. These inbound-outbound points are placed beyond the buffer distance, b_{max} that defines the extent of cluster h (Fig. 2). Shortest travel times, t_{ij}, are computed and stored for all ij pairs ($i \neq j$). Then, the CFTCM computes all ikj shortest travel times (t_{ikj}), inserting each node k within b_{max} distance of any optimization node in cluster h as an intermediary stop between i and j. The travel time difference between t_{ikj} and t_{ij} is compared against a maximum deviation threshold value, Δ_{max}. If $t_{ikjh} - t_{ijh} \leq \Delta_{max}$, then the tracking variable $X_{ikjh} = 1$ for that travel path, and 0 otherwise. This is repeated for all ij pairs and street network nodes k in all clusters h. The CFTCM then computes a continuous score $0 \leq A_{kh} \leq 1$ for all nodes k within each cluster h representing the fraction of travel paths that pass through node k with a deviation less than Δ_{max} for all possible travel paths through cluster h:

$$A_{kh} = \frac{\sum_{i}^{I_h} \sum_{j \in N_{ih}}^{J_h} X_{ikjh}}{\sum_{I_h}^{I_h} \sum_{j \in N_{ih}}^{J_h}} \tag{6}$$

The CFTCM, like the original specification of the FTCM, is an enumeration procedure. One key difference between them is the greater number of possible ij travel paths to account for in each cluster h. In the original specification and application of the FTCM, the maximum number of artificial origins and destinations for any freeway intersection in greater Los Angeles was five. However, by grouping nearby freeway-freeway and freeway-street intersections into a single cluster and by also considering major arterial routes that carry traffic through the cluster, the sets I_h and J_h include as many as eight inbound and outbound points.

3.2.3. Selection of best optimization node within a cluster

After the CFTCM evaluates all street network nodes near an optimization node selected by any of the initial regional modeling scenarios, these scores are used to identify the best optimization node for each cluster h. To do so, we first determine if there are street network nodes where $A_{kh} = 1.0$ in each cluster. If so, the optimization node that is the nearest neighbor to the most street network nodes where $A_{kh} = 1.0$ is selected as the best optimization node in the cluster. If there are no street network nodes in the cluster where $A_{kh} = 1.0$, we then consider optimization nodes' proximity to the best A_{kh} scores in the cluster in similar fashion, so long as the score is still relatively high. For this study, $A_{kh} > 0.8$ is considered the threshold for "good" street network node scores, as 80% represents a clear majority of through-routes. Similar to the procedure above, we select the optimization node that is the nearest

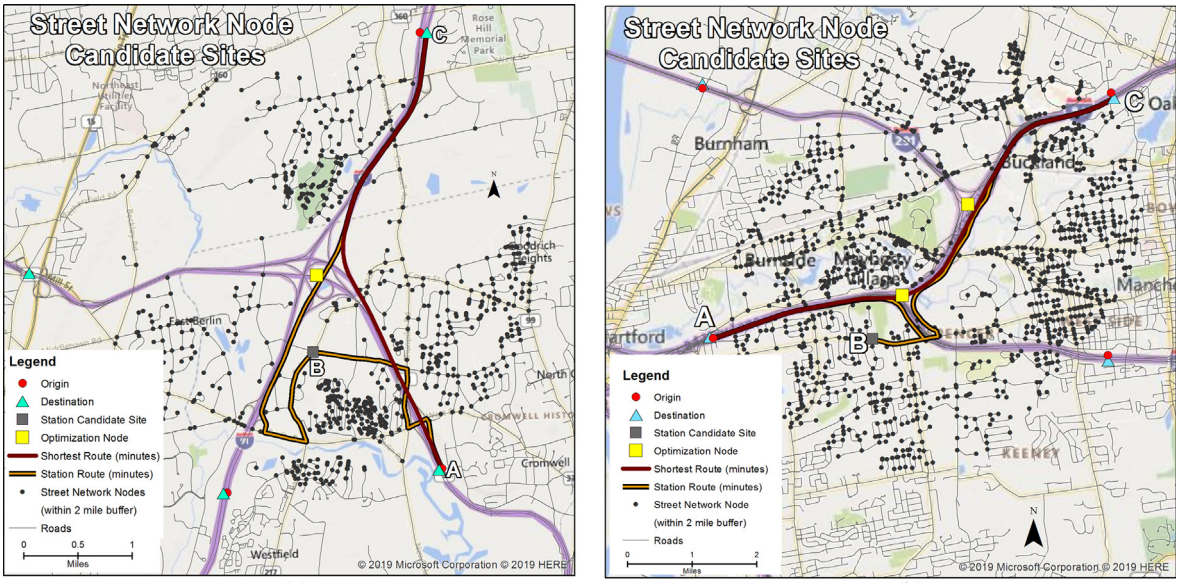


Fig. 2. (a) Example of CFTCM evaluation for a street network node k (node B) in a cluster h consisting of a single optimization node. The red line shows the direct route from A to C without a stopping at the street network node, and the brown line shows the fastest route from A to C via B. Similar calculations are repeated for all directional pairs and for all street nodes within the buffer. (b) Example of evaluating least travel-time paths for inbound and outbound origins and destinations for a cluster consisting of multiple optimization nodes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

neighbor to the greatest number of street network nodes with the highest A_{kh} score observed in the cluster, so long as the value is greater than 0.8. If there is no single node where $A_{kh} > 0.8$, then we recommend that none of the optimization nodes in the cluster be considered for station locations, as this indicates that drivers will have a difficult time leaving the freeway network, reaching a nearby station on the local street network, and continuing on their journey.

4. Data

4.1. Network

To implement the multi-scale refueling infrastructure planning framework, we obtained 2015 data for the transportation network, origin-destination trip matrix, and traffic analysis zones (TAZ) from Capitol Region Council of Governments (CRCOG) with help from the Connecticut Center for Advanced Technology's Hydrogen Fuel Cell Advancement Program. We chose the Greater Hartford area as our case study area because it is slated to be one of the next regions for hydrogen refueling infrastructure development after California. The transportation network includes interstates, freeways, arterials, and rural roads. To simplify the network and make it compatible with the FRLM/DFRLM models, we extracted the major roads in similar fashion to Kuby et al. (2009) and changed the connectivity of exits on the Massachusetts Turnpike to reflect the limited accessibility. The final simplified road network used in the analysis has 1344 arcs and 806 junctions/candidate sites. To represent long distance trips to New York, Boston, Albany, and Providence, we further extend the existing road network to these cities using highway travel distances from Google Maps.

4.2. Preprocessing

4.2.1. TAZ aggregation and shortest path routes

Since the number of TAZs in the original dataset (1829) exceeded the number of junctions in the simplified road network (806), we aggregated TAZs to capture all the trips from the region using a multi-step procedure (Table 2). The aggregation algorithm is

Table 2
TAZ aggregation methods.

	dH/dS < 1.25	dH/dS < 2 and dH/dS > 1.25	dH/dS > 2
Choose street junction	Never	Only if dH > 4 miles (6437 m)	Always
Choose highway junction	Always	Always	Only if dH < 1 miles (1609 m)

Note: dH: distance to highway junction; dS: distance to street junction.

based on common-sense rules for where drivers in each TAZ would most frequently join the major road network: at the nearest street junction or the nearest highway junction. The steps are as follows:

1. Calculate distance from TAZ centroids to their nearest street junction *and* nearest highway junction.
2. For each TAZ, calculate the ratio of these two distances (distance to highway junction/distance to street junction).
   a. If the ratio is less than 1.25, assign the TAZ to the nearby highway junction.
   b. If the ratio is between 1.25 and 2, assign the TAZ to the highway junction, unless the nearest highway junction is > 4 miles away from the TAZ centroid, in which case assign the TAZ to the nearest street junction.
   c. If the distance ratio is greater than 2, assign the TAZ to the street junction unless the highway junction is within 1 mile of the TAZ centroid, in which case assign the TAZ to the highway junction.
3. Aggregate TAZs that are assigned to the same junction.

The overall strategy favors highway junctions because we believe highway junctions capture the majority of the traffic flows in the urban area, and these locations should be given priority for the early phase of infrastructure planning. After TAZ aggregation, the number of TAZs (and the corresponding O-D nodes) dropped from 1829 to 514. Next, we aggregated the original 1829 × 1829 O-D trip matrices to a 514 × 514 trip matrix. Note that nearby Springfield, Massachusetts was not included in the analysis: Springfield is not a member of the Capitol Region Council of Governments or part of the Hartford Census Combined Statistical Area.

Using ArcGIS Network Analyst, we calculated the shortest paths among all pairs of the 514 OD nodes based on posted speed limits. Following Kuby et al. (2009), we added a 15% speed penalty for non-freeway roads. The final model included 264,196 O-D pairs and routes.

4.2.2. Assumptions

In the FRLM and DFRLM, the vehicle driving range typically reflects an agreed-upon maximum spacing of stations rather than the technical driving range of a fully fueled vehicle. In this paper, we run models with driving ranges of 100 and 150 miles, even though most FCVs on the market in early 2019 have ranges of 300–400 miles. The smaller assumed driving range provides a safety margin in case of side trips, incomplete fills, air conditioning usage, hill climbing, and station reliability issues. Although the criteria for the US FAST Act Alternative Fuels Corridor Designation is spacing every 100 miles (USFHWA 2018), spacing of 150 miles would still enable a vehicle with a more than 300 mile range to reach the next station on a long-distance trip if its first station were temporarily closed. For the DFRLM, the key assumptions are the shape, slope, and maximum deviation of the deviation penalty function. We used the sigmoidal function form in (7), which is similar but not identical to the sigmoid function in Kim and Kuby (2012).

$$g(DT)_{sigmoid} = \frac{1}{1 + AC^{\left(B \cdot (100(\frac{DT}{tq}) - DT_{max}\right)}} \tag{7}$$

- The A parameter mainly shifts the function to the left or right, and as such controls the presence or absence of the upper and lower plateaus of the sigmoid function.
- B controls the "tipping point" where the rate of decline hits a maximum steepness and the fractional demand is 0.5.
- C controls the gradualness or steepness of the distance decay.

The precise parameters of the sigmoidal functional form were calibrated to produce the decay function shown in Fig. 3: $A = 0.015$, $B = 2.06$, and $C = 1.046$, and $DT_{max} = 0.33$ h. Of course, different drivers will have varying willingness to deviate from their shortest paths, but we believe this to be a reasonable approximation.

Although some existing work directly incorporates driver's travel behaviors and path choices into the modeling of refueling

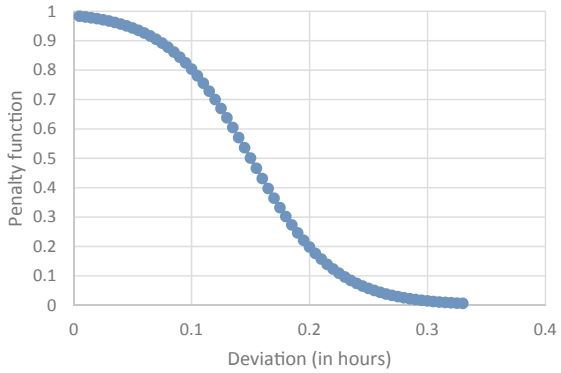


Fig. 3. Deviation decay function assumed. Based on the deviation findings in Kuby et al. (2013), the plateau of relative indifference to small deviations was set so that 80% of drivers would be willing to make up to a 6 min deviation, after which the willingness should drop rather steeply, up to a maximum deviation of 20 min.

station locations (Guo et al., 2016), we do not incorporate this factor directly into the DFRLM model. Instead, we use the deviation decay function to show a generalized willingness of drivers to deviate from their pre-defined shortest path to the nearby refuelling stations. The drivers' path choices and travel behaviors are influenced by the nearby refuelling station locations if they are willing to deviate their shortest path and refuel.

The third set of assumptions relate to the CFTCM. The Δ_{max} parameter, the maximum driver deviation in minutes from any inbound path to a street network node and back to the outbound path in a cluster, was tested at six minutes to be consistent with the end of the indifference plateau in Fig. 3, and also tested with a more conservative assumption of five minutes. The c_{max} threshold, the maximum distance between optimization nodes in the same cluster, is set to three network travel miles, also consistent with the 6-min rule (at freeway travel speeds). Finally, b_{max}, the buffer distance for street network nodes around optimization nodes, is set to two miles. Ultimately, the Δ_{max} parameter will largely influence the A_{kh} scores; the b_{max} and c_{max} parameters serve to limit the computation time by precluding calculation of travel times to the exponentially increasing number of street intersections that are too far from the center of a cluster to meet the Δ_{max} requirement for all inbound-outbound travel directions.

5. Results

5.1. Regional-scale station location using FRLM and DFRLM

The FRLM and DFRLM were solved on an Apple iMac running under Parallels with Windows 7 with 1600 MHz, DDR3 with 3.4 GHz Intel Core i7 and 12 GB RAM allocation. The FRLM was solved using FICO Xpress 7.8 (64-bit). The DFRLM was solved in ArcGIS Desktop 10.0 using the greedy substitution algorithm in Kim and Kuby (2013) with three substitutions, coded in Microsoft Visual Studio 2010 using C#. Phase 1 FRLM models solved in 2–541 s, while DFRLM runs solved three orders of magnitude slower (5K–197K s). For Phase 2, despite fewer candidate sites, FRLM runs solved more slowly (0–1266 s), while DFRLM computations times fell by 1–2 orders of magnitude to 171–3614 s. In Phase 3, with all integer location variables fixed at 1 or 0, FRLM runs took 0–72 s and DFRLM runs took 16–70 s.

The CFTCM was constructed in Python 2.7, using the Network Analyst extension in ArcGIS 10.6.1. The program ran on a virtual machine in the Windows Server 2012 environment with 256 GB of RAM, using a 2.6 GHz Intel Core i7 processor. Solution times ranged from 45 min to 4 h per cluster, depending on the number of travel directions and number of street network nodes.

In Phase 1, we ran the FRLM using all 806 optimization nodes as candidate sites, while for the DFRLM we use the same set of nodes excluding 6 highway-highway intersections. For each model, we generated 12 scenarios for all combinations of two driving ranges (100 and 150 miles), two objective functions (Max Trips and Max VMT), and three station budgets (5, 10, and 15 stations. At the time this analysis was performed, there was one existing station at the Sun Hydro headquarters near Wallingford and one planned station north of downtown Hartford at the Pride Travel Center, located near a bus depot and an aggregation of automobile dealers. These locations were considered "fixed" in all scenarios. Table 3 shows the Phase 1 coverage results for 5, 10, and 15 stations. DFRLM can certainly cover more trips or VMT because deviation is allowed. Fig. 4 shows the frequency with which stations were selected in the 24 Phase 1 scenarios.

5.2. Cluster-based FTCM analysis

Following the Phase 1 output, we applied the CFTCM to narrow down the set of optimization nodes to consider for Phase 2. In Phase 1, 46 different optimization nodes were selected in at least one of the 24 modeling scenarios. Of these 46 nodes, the CFTCM evaluated 36 different nodes within a total of 19 clusters (Table 4, Fig. 5). Of these 19 clusters, 13 clusters contained 2–4 optimization

Table 3
Phase 1, 2, and 3 optimization coverage results comparison.

Number of Stations	Phase	FRLM				DFRLM			
		Driving range 100		Driving range 150		Driving range 100		Driving range 150	
		Max Trips	Max VMT	Max Trips	Max VMT	Max Trips	Max VMT	Max Trips	Max VMT
P = 5	Phase 1*	16.98	56.23	17.06	65.46	43.36	59.68	44.57	73.50
	Phase 2**	16.76	56.23	16.80	65.46	43.34	59.62	44.72	73.62
	Phase 3***	15.48	45.07	16.15	49.81	43.34	59.38	43.47	60.21
P = 10	Phase 1	30.08	72.15	30.16	74.23	64.36	84.28	63.48	85.69
	Phase 2	28.55	70.71	28.63	73.10	64.16	84.34	63.20	85.74
	Phase 3a (Oakland)	27.07	54.14	27.52	57.49	64.16	70.80	64.20	71.07
	Phase 3b (Sturbridge)	24.54	52.99	26.44	71.46	59.55	69.30	61.08	85.57
P = 15	Phase 1	39.76	78.27	39.83	79.68	75.08	90.22	75.11	90.96
	Phase 2	37.03	77.01	37.25	78.36	75.07	90.28	73.41	90.98

* Phase 1 candidate sites: Using all nodes for FRLM; using all nodes except multi-freeway intersections for DFRLM.
** Phase 2 candidate sites: Best optimization node from each CTFCM cluster only.
*** Phase 3 candidate sites: Five or ten final recommended nodes only.

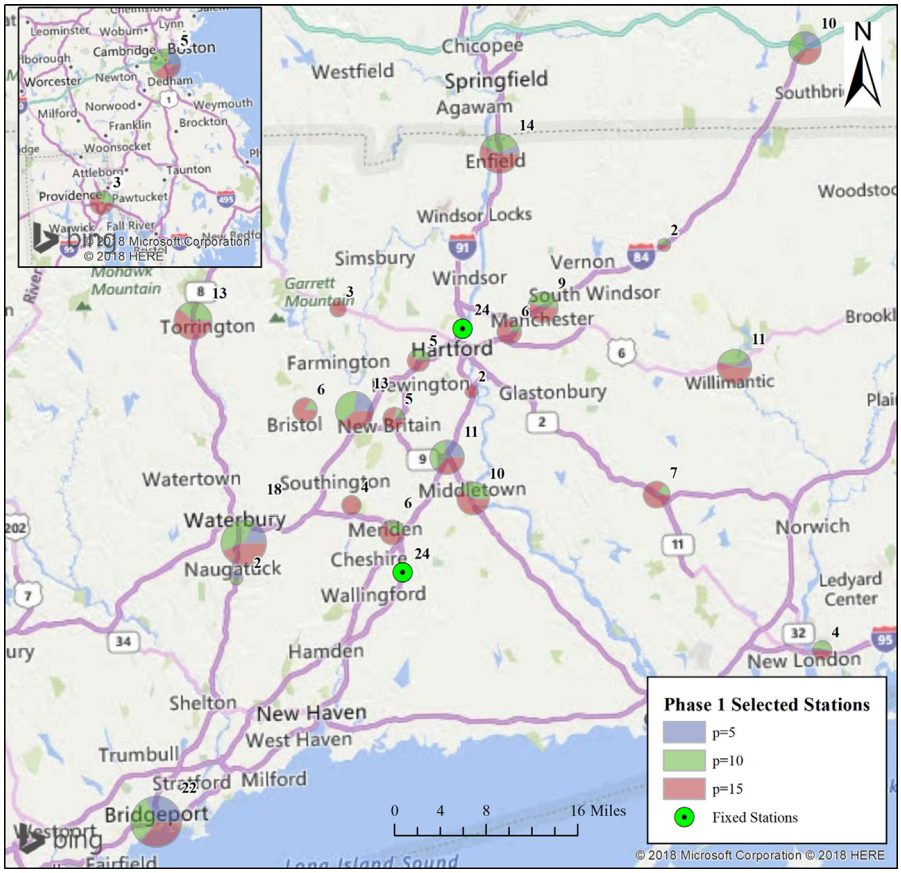


Fig. 4. Station locations selected in Phase 1 by FRLM and DFRLM, by p = 5, p = 10, and p = 15 scenarios. Circle size and numerical label indicates the number of times the station was optimal across the 24 scenarios.

Table 4
CFTCM summary statistics.

Cluster name	Number of optimization nodes	Major roads	Travel paths	Street network nodes	Δ_{max} = 6 min		Δ_{max} = 5 mins	
					A_{kh} Max.	n A_{kh} = 1.0	A_{kh} Max.	n A_{kh} = 1.0
Torrington	1	US-202, CT-8, CT-4	30	1025	1.00	192	1.00	91
Waterbury	4	I-84, CT-8, CT-69	12	2454	1.00	467	1.00	218
Southington	1	I-84, CT-10	10	845	1.00	160	1.00	78
Plainville	3	I-84, CT-72, CT-10	12	1865	1.00	31	1.00	10
Pratt's Corner	1	I-84, CT-322	12	464	1.00	68	1.00	45
Meriden	2	I-91, I-681, US-5, CT-15	38	1370	1.00	119	1.00	32
New Britain	3	I-84, CT-9, CT-72, CT-372	28	1780	0.89	0	0.80	0
West Hartford	1	I-84	12	1379	1.00	86	1.00	11
East Berlin	2	I-91, CT-9, CT-3, CT-372	20	1112	1.00	4	1.00	1
Wethersfield	2	I-91, US-5, CT-3	24	1282	1.00	10	0.92	0
Middletown	2	CT-9, CT-17, CT-66, CT-3	30	1285	1.00	5	0.93	0
Enfield	2	I-91, CT-190, CT-220	18	1046	1.00	289	1.00	190
Hilliardville	3	I-84, I-291, I-384	36	1877	0.78	0	0.78	0
Oakland	3	I-84, CT-83	26	1524	0.88	0	0.88	0
Colchester	2	CT-2, CT-11	30	361	1.00	22	1.00	3
Willington	2	I-84, CT-32, CT-74	28	207	1.00	21	1.00	4
Willimantic	2	US-6, CT-195, CT-32	30	846	0.93	0	0.93	0
Sturbridge	1	I-90, I-84	6	289	1.00	76	1.00	57
Platt Mills	1	CT-8, CT-63	6	1171	1.00	104	1.00	40

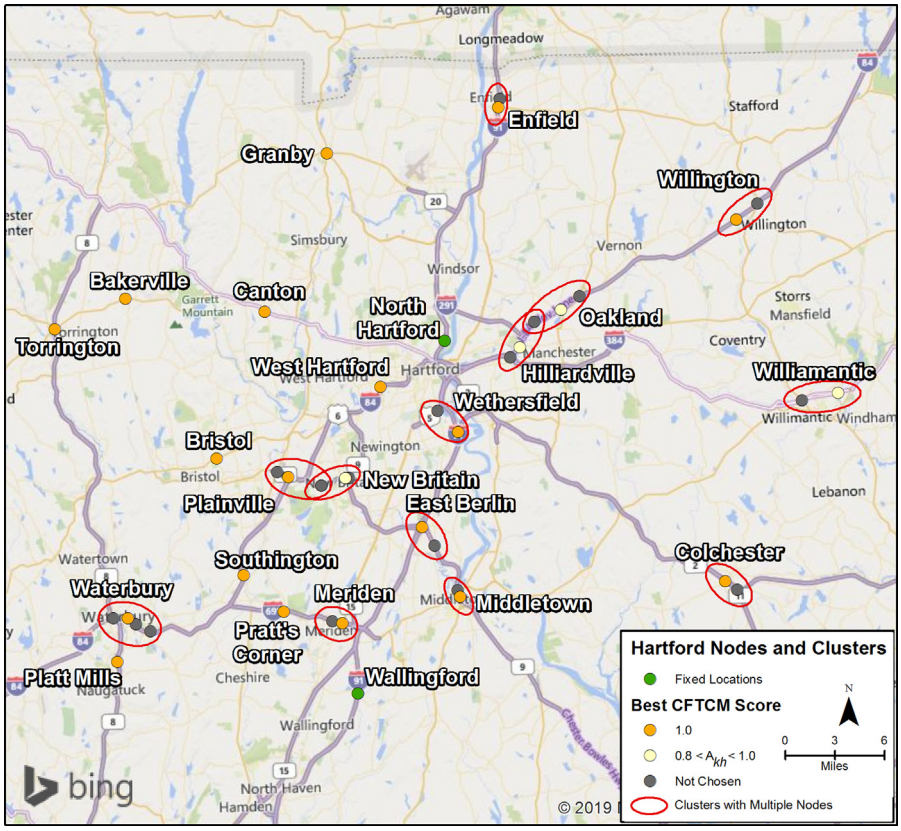


Fig. 5. Map of optimization nodes chosen for Phase 2 of the analysis. The best optimization node in each cluster with multiple nodes (indicated by the red ovals) is identified in orange if $A_{kh} = 1.0$, or yellow if $0.8 < A_{kh} < 1.0$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

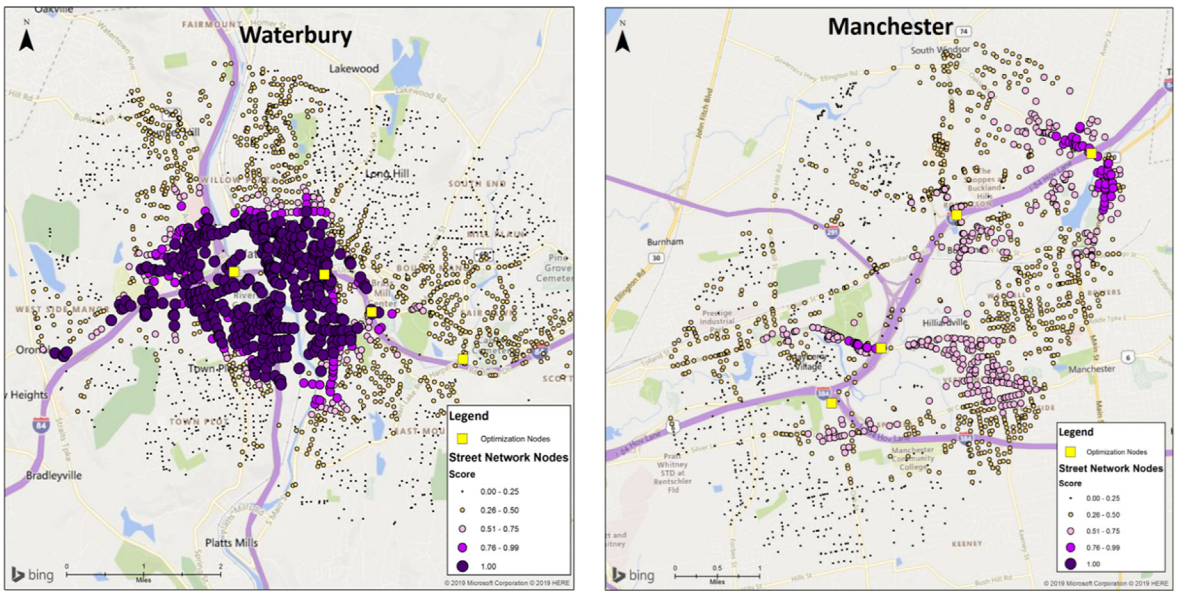
nodes, while six clusters had only one, and there were two optimization nodes shared between clusters. The remaining 10 optimization nodes were not evaluated by the CFTCM. In addition to the two fixed stations, four stations were occasionally selected in Boston, Providence, Bridgeport, and New Haven, which were included as highly generalized locations of possible refueling stations for inter-city trips from Hartford. An additional four are located in Bristol, Canton, Bakerville, and Granby at the intersection of surface streets that are not adjacent to any limited access highway or freeway nodes, which precludes them from CFTCM assessment.

Clusters generally had a number of promising street network nodes in close proximity to optimization nodes. Only four of the 13 clusters with multiple optimization nodes lack a street network node capable of being reached with a six-minute deviation or less for all possible inbound-outbound routes passing through them, and the lowest "best" score for a street network node in a cluster was 0.78. For sensitivity analysis, we lowered the deviation to five minutes. In this case, the original group of nine clusters with at least one street intersection with $A_{kh} = 1.0$ assuming $\Delta_{max} = 6$ was reduced to seven with $\Delta_{max} = 5$. Meanwhile, of the original group of four clusters with maximum A_{kh} scores below 1.0, the maximum score did not fall any further when lowering Δ_{max} to 5 min.

Using the decision rules outlined in Section 3.2.3, we then selected the best optimization node from each of the 13 clusters with multiple optimization nodes. These are the nodes that are passed to Phase 2 of the analysis. This process also removed 18 optimization nodes identified in Phase 1 from further consideration. When visualizing the distribution of A_{kh} scores within clusters, we observe that in most cases, the best street network node scores were noticeably nearer to one of the optimization nodes, making the selection of a best node in the cluster a straightforward process. The relative position of the chosen optimization node within the cluster differed, though, largely as a function of the value of the maximum A_{kh} scores. For clusters with multiple $A_{kh} = 1.0$ street network nodes, these nodes tended to be in the middle of the cluster (Fig. 6a). In contrast, for clusters that lacked at least one $A_{kh} = 1.0$ score, these "best" sites were often located farther from the center of the cluster (Fig. 6b). For these clusters without a location capable of capturing all travel paths through them, the best street network nodes were often relatively close to one or more freeway entrances and exits. These sites were therefore able to capture a consistent subset of passing travel paths that included travel along the adjacent freeway, but were unable to capture at least one shortest travel path elsewhere in the cluster.

5.3. Phase 2 analysis and choice of top ten stations

In Phase 2, we execute the same FRLM and DFRLM scenarios as in Phase 1, but with the 18 optimization nodes that were not the



a) well-connected intersection. b) poorly connected intersection.

Fig. 6. Distribution of A_{kh} scores for street intersection nodes around optimization nodes in clusters with multiple street network nodes where $A_{kh} = 1.0$ (Fig. 6a), and street network nodes where the maximum A_{kh} score < 1.0 (Fig. 6b).

best sites in their clusters removed from consideration. After this was complete, 26 different optimization nodes were identified as optimal in at least one of the 24 scenarios. We then narrowed this set of 26 nodes to two final lists: 1) the best five initial recommended station locations for the Hartford region (including the existing and planned stations at Wallingford and North Hartford), and 2) the best ten, which includes the best five (Table 5, Fig. 7). These numbers are chosen because they represent reasonable scenarios of station counts that would be initially recommended for a region interested in developing an infrastructure to support HFCV adoption with a limited budget.

Fig. 7 shows the spatial distribution of these nodes across the Hartford region, and Table 5 shows how frequently stations were selected in either a DFRLM or FRLM scenario. To determine which nodes to include in the recommended sets of 5 and 10 stations, we first considered the overall frequency of a node's occurrence across the 24 scenarios. In addition, we placed a high degree of importance on how often a node was chosen in the more selective $p = 5$ or $p = 10$ Phase 2 solutions. Next, we prioritized relatively consistent performance in both DFRLM and FRLM models, VMT and trip maximization, and both vehicle ranges. We considered these nodes to be more robust locations than those that were never selected under certain assumptions. For these reasons, we selected Bridgeport, Waterbury, and Plainville to accompany Wallingford and North Hartford in the set of five best locations. These three nodes are selected more frequently in $p = 5$ or $p = 10$ solutions than any of the other 21 nodes. Each appears with relatively equal frequency between DFRLM and FRLM scenarios, between VMT and trip scenarios, and between scenarios where the range is 100 or 150 miles.

For the set of ten best locations, additional factors were considered. Torrington and Enfield are each ranked seventh in occurrence in $p = 5$ or $p = 10$ outputs, and sixth in overall occurrence. They also appear consistently across scenario parameters. Willimantic occurs slightly less often than these two but still in the top 10 of total frequency and in $p = 5$ or $p = 10$ frequency. This location, east of Hartford, is near both the Eastern Connecticut State University and the University of Connecticut. After these three, the last two nodes included in stations 6–10 are cases where each decision came down to two distinct pairs of stations that were favored by specific scenario types.

In the first case, we chose a station in Middletown instead of East Berlin for three primary reasons. East Berlin occurred more frequently in $p = 5$ or $p = 10$ solutions than Middletown, but there are important limitations to its inclusion in the set of best ten locations. First, East Berlin appeared only in FRLM solutions, while Middletown occurs in both. Second, the CFTCM indicates that there are no street network nodes near the intersection of Interstate 91 and Connecticut Highway 9 at which to build a site near East Berlin that can capture passing traffic for all or most travel directions passing through the area, while there are more promising sites near Middletown. Finally, we note that the predetermined inclusion of stations at Wallingford and North Hartford makes East Berlin less important to the initial arrangement of stations in the region. This is particularly true for the existing station in Wallingford, which can capture all flows passing along Interstate 91 if drivers are allowed to deviate to the parallel Wilbur Cross Parkway.

The second case involved two stations along Interstate 84 between downtown Hartford and Massachusetts: Sturbridge, Massachusetts and Oakland, CT (near the suburban Manchester, CT shopping complex). Sturbridge is selected slightly more frequently in $p = 5$ or $p = 10$ solutions than Oakland, but there is a substantial difference between scenarios that maximize trip capture or VMT capture. The station at Sturbridge is an important site for refueling trips to Boston, and would be an essential component of a

Table 5
Phase 2 results for recommended stations.

Location	Notes	Total out of 24 Scenarios	DFRLM	FRLM	p = 5	p = 10	p = 5 and p = 10	p = 15	Max Trips	Max VMT	Range 100	Range 150	Selected Most Robust Nodes	Rank
Bridgeport I-95/CT-8		22	10	12	7	7	14	8	10	12	12	10	Top 5	1
Waterbury I-84/Baldwin St.		18	8	10	3	7	10	8	10	8	10	8	Top 5	1
Plainville I-84/CT-72	Freeway Intersection	13	7	6	4	5	9	4	7	6	7	6	Top 5	1
Wallingford I-91/CT-68	Fixed	24	12	12	8	8	16	8	12	12	12	12	Top 5	1
N. Hartford I-91/Jennings Rd.	Fixed	24	12	12	8	8	16	8	12	12	12	12	Top 5	1
Torrington CT-8/CT-202		14	6	8	0	6	6	8	8	6	7	7	Top 10	2
Middletown CT-9/CT-66.		10	8	2	0	4	4	6	6	4	5	5	Top 10	2
Enfield I-91/CT-90		14	9	5	1	5	6	8	7	7	7	7	Top 10	2
Willimantic US-6/CT-195		11	9	2	1	4	5	6	6	5	6	5	Top 10	2
Oakland I-84/CT-30		9	5	4	0	4	4	5	8	1	4	5	Top 10 Alt1	3
Sturbridge, MA I-90/I-84	Freeway Intersection	10	3	7	3	3	6	4	1	9	3	7	Top 10 Alt2	3

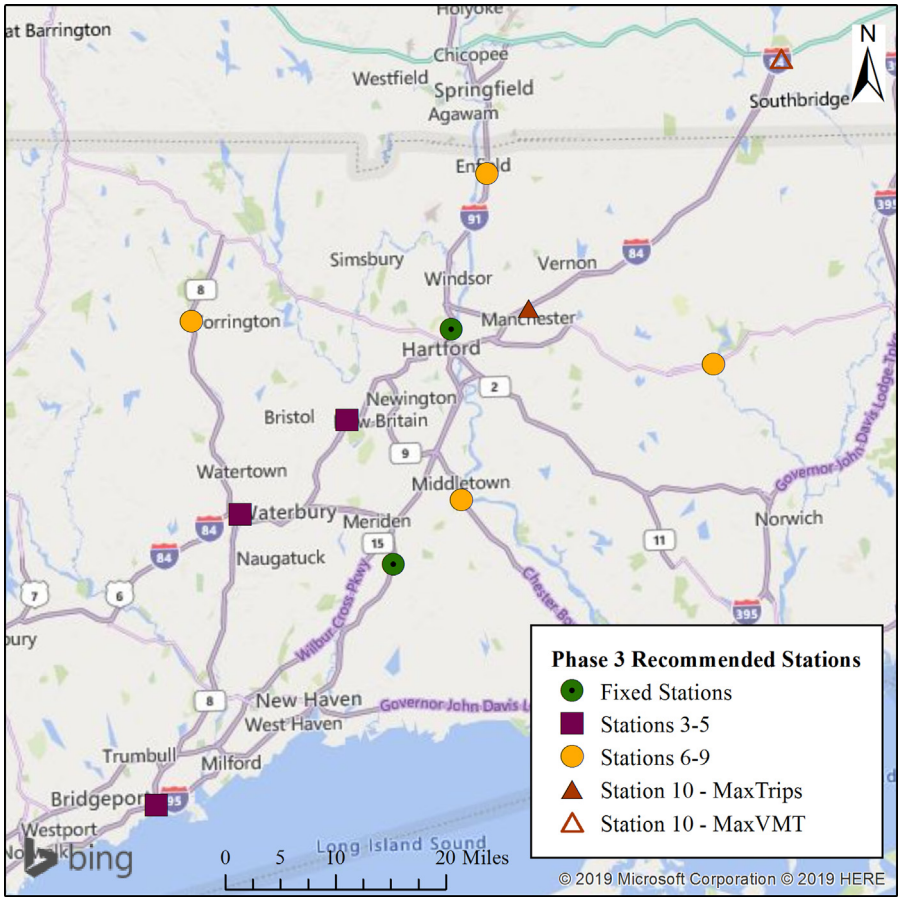


Fig. 7. Final set of recommended stations evaluated in Phase 3, with two options shown for the 10th station.

refueling infrastructure that includes Massachusetts and Connecticut simultaneously. However, the station site in Oakland is more important to local trips within the Hartford region, and is located immediately adjacent to the Buckland Hills Mall. When reviewing the CFTCM scores for this cluster, the Sturbridge site has better sites on the street network that can capture all passing travel directions, though it is important to note that Manchester is part of a large cluster that must satisfy many more travel paths through its nearby area than the isolated rural node at Sturbridge. Notably, the street intersections near the Oakland station node have higher FTCM scores than any other optimization node in the cluster, even though the node is not in the center of the cluster. If the intent, then, is to capture more trips, the Oakland node would be most appropriate, while Sturbridge would be best for maximizing VMT capture.

5.4. Phase 3 testing of recommended stations

The final phase of the analysis evaluates how well the recommended sets of five and ten stations perform as a fixed group under all range, objective, and FRLM/DFRLM scenarios (Table 3). That is, the five or ten recommended x_k variables are set to 1. In this analysis, the best recommended set can do is match the performance of the optimized solution. This occurs once, where the top five was also the optimal solution of the DFRLM scenario where $p = 5$ and vehicle range = 100 and the number of covered trips are maximized. This is notable because this is arguably the single most important scenario for planning the first five stations, since it (a) uses a conservative range, (b) emphasizes more basic local trips, and (c) takes into account that with only five stations many early adopters will need to (and be willing to) make short deviations to refuel. The other $p = 5$ trip maximization solutions are also extremely close to the Phase 2 objective values. The recommended five underperform the best $p = 5$ solutions compared with three of the four Phase 2 solutions that maximized VMT, which is not surprising. Without a station in Sturbridge or Boston to facilitate the round trip to Boston and back, a significant amount of inter-city trips are not refuelable. Given that most early FCV adopters in California own a second conventional car in the household that can be used for longer trips, a connecting station to another state is not a high priority for the first five stations (California Air Resources Board, 2018; Lopez Jaramillo et al., 2019).

A similar pattern emerges in the Phase 3 analysis of the recommended group of 10 stations. Especially with Oakland as the 10th station, the recommended 10 perform an average of only 1.9% less than the optima in Phase 2 for maximizing refuelable trips, but 19.5% worse for maximizing refuelable VMT. When choosing Sturbridge, MA as the 10th station, the final 10 stations perform much

better on the two VMT-maximizing scenarios with a driving range of 150 miles—only 1.2% below the Phase 2 customizable solutions. When the assumed driving range is lowered to 100 miles, the Sturbridge solutions average 21.5% less than the best Phase 2 VMT objectives. For trip maximization, the Sturbridge solutions perform 8.1% worse than the best Phase 2 objectives, and 6.3% worse than the final 10 stations using Oakland.

6. Discussion

The final choice between Oakland, CT and Sturbridge, MA for the 10th station comes down to whether VMT or trips are prioritized, and whether government subsidies dictate against an out-of-state station. If one assumes that most early adopters would have a second conventional car in the household even at the 10-station stage, it would argue for prioritizing the station in the Oakland-Manchester cluster.

Several limitations should be kept in mind when evaluating these results. The existing Sun Hydro station in Wallingford is currently not open to the public, and this location is particularly problematic in the FRLM runs because more O-D shortest paths in our dataset take the lower-capacity Wilbur Cross Parkway, which parallels the higher-capacity I-91 in the vicinity of Wallingford.

Other limitations concern inter-city trips. We initially obtained an O-D trip matrix that included flows between Hartford and Springfield, MA, which is just across the Massachusetts border from the northern boundary of the Hartford Metropolitan Statistical Area. However, the size of TAZs and flow volumes were incompatible. If Springfield were included, one or more optimal stations may have shifted to that region. For modeling trips to neighboring metropolitan areas, we included simplified routes departing from a single exit point from the Hartford region to each of Boston MA, Albany NY, Providence RI, and New York NY. In this fashion, we attempted to factor in both intra-city and inter-city trips into infrastructure planning. Since stations can serve both types of demand this is important to consider, but it remains a challenge to realistically integrate trips at these different scales. Given the importance of state subsidies in California's roll-out of hydrogen stations, the coordination between neighboring states in New England creates the potential for long-distance gaps and/or cross-border duplication.

Previous applications of the FRLM and DFRLM to infrastructure planning have tended to focus on only one scale of trips at a time: intra- or inter-city, but not both. An exception is Kuby et al. (2009), which solved a variety of scenarios on two separate networks: an inter-city Florida network and an intra-city network for the Orlando metropolitan area. The authors then recommended sets of stations at each scale, making certain that the inter-city stations from the statewide model were included in the local set of stations at the exact same locations. However, all stations in the Orlando recommendations were not included in the statewide network. This paper has attempted to improve on their method with this new hybrid approach.

In our opinion, the inclusion of inter-city trips in the flow matrix makes it even more important to include the street-level CFTCM analysis. Given that the stations are serving inter-city highway trips, it becomes even more important that any station near a confluence of limited-access highways be reachable conveniently by as many inbound-outbound travel paths as possible, and hopefully by all such travel paths ($A_{kh} = 1.0$). The CFTCM, then, can identify locations that serve as many trips passing through the cluster from across the region as possible that are also convenient for those living or traveling nearby.

The six-minute deviation threshold used in this study for the CFTCM is subject to uncertainty. That metric was observed in a sample of CNG drivers in southern California who also refueled in a sparse AFV refueling infrastructure (Kelley and Kuby 2013), though it is possible that early HFCV adopters in Hartford would react differently to a different network of stations. We also did not vary deviation thresholds based on the number of stations assumed to be built when running the FTCM, and it is possible that this threshold would be lower if more stations were present in a region. The metric also only evaluates accessibility in terms of travel time deviation reduction, though other factors are important to drivers' refueling station choices, including proximity to nearby amenities, perceptions of safety, and congestion.

It is important to note that the CFTCM provides a standard performance metric for street network intersections, and not parcels upon which stations would actually be built. In order to effectively conduct this final piece of the station site selection process, other factors not considered in the specification of the CFTCM, such as zoning and land use data, property values, and municipal ordinances that would impact station construction at the site, would be necessary. These could be directly incorporated into future specifications of the CFTCM or conducted separately in GIS analysis after the workflow described in this study is completed. Such an effort would require gathering a variety of this data from different municipalities across the study area. Additionally, we evaluate deviations for travel on the regional network and on the local street network near clusters separately. It is unclear to what extent drivers make this distinction, or if sensitivity to leaving the freeway before returning to it on a travel path changes the willingness to deviate.

7. Conclusions

The main contribution of this work is to demonstrate the importance of integrating modeling approaches for optimal location of refueling infrastructure across street-level, intra-city, and inter-city scales. Given the high flow volumes that can potentially be refueled at stations located near the confluence of multiple freeways, and the demonstrated willingness of AFV drivers to refuel mid-route, the ability to deviate from inbound freeways to stations and back to outbound freeways for all possible through-routes becomes an important consideration for limited infrastructure planning. We find that with this multi-scale, integrated path-based planning approach, a small number of alt-fuel stations are capable of providing basic (uncapacitated) coverage to large number of trips for the Central Connecticut region. In our final model results with 150 miles vehicle driving range by the DFRLM, 5 stations cover 43% of trips and 60% of VMT, 10 stations cover 64% of trips and 71% of VMT if planned to facilitate local travel (max trips), and 10 stations cover 61% of trips and 86% of VMT if planned to facilitate regional travel (max VMT). This paper has illustrated one method for

integrating across these different models and scales in a timely real-world application. The multi-stage modeling process also ensures the robustness of station selections across different modeling scenarios. This newly proposed planning method is generalizable to other regions or other types of fast-fueling alternative fuel vehicles.

Future work on multi-scale modeling could follow a number of promising directions. An obvious next step is to develop a framework for suitability analysis of sites with similarly high A_{kh} scores, taking into account zoning, safety, parcel size, land value, cost, and existing land use. Indeed, constructing stations in areas proximal to amenities that attract both local and regional travel that are near freeway intersections—such as shopping malls—with zoning compatible with station development may best support early adopter refueling needs from across the region and encourage them to purchase HFCVs, and a variety of local stakeholder knowledge of the Hartford region would be essential to informing this next step. In this paper, DFRLM runs were computationally intensive. Incorporating some recent computational developments in DFRLM solution methods could allow more detailed geographic networks at both the intra- and inter-city scales in the same model runs. The CFTCM is also a computationally intensive enumeration procedure, and developing effective heuristics to more efficiently identify promising sites near clusters would be a promising research direction. Computerized iteration between the two optimization scales is also potentially promising. Undoubtedly, there are other ways that the FTCM or CFTCM could be integrated with the FRLM and DFRLM for multi-scale infrastructure planning. Instead of using the CFTCM as a filtering step, it might also be possible to incorporate A_{kh} scores directly in the FRLM or DFRLM.

Acknowledgements

This material is based upon work supported by the Seed Research Grant from Institute for Social Science Research at Arizona State University. The authors would like to thank Dr. Jong-Geun Kim and Mr. Shuyao Hong for the software set up efforts and all the anonymous reviewers for their insightful comments and suggestions on the earlier version of this manuscript. We also thank Joel Reinbold of the Connecticut Fuel Cell Coalition, and Erik Snowden and Ming Zhao of the Capitol Region Council of Governments, for their help obtaining network and trip data. We thank FICO for providing Xpress through their Academic Partnership Program.

References

Arslan, O., Karaşan, O.E., 2016. A Benders decomposition approach for the charging station location problem with plug-in hybrid electric vehicles. Transp. Res. Part B: Methodol. 93, 670–695. https://doi.org/10.1016/j.trb.2016.09.001.

Berman, O., Larson, R.C., Fouska, N., 1992. Optimal location of discretionary service facilities. Transp. Sci. 26, 201–211.

Boostani, A., Ghodsi, R., Miab, A.K., 2010. Optimal location of compressed natural gas (CNG) refueling station using the arc demand coverage model. In: Presented at the 2010 Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation (AMS). IEEE, pp. 193–198.

Brey, J.J., Brey, R., Carazo, A.F., Ruiz-Montero, M., Tejada, M., 2016. Incorporating refuelling behaviour and drivers' preferences in the design of alternative fuels infrastructure in a city. Transp. Res. Part C: Emerg. Technol. 65, 144–155.

California Air Resources Board, 2018. Hydrogen Fueling Infrastructure Assessments.

Capar, I., Kuby, M., Leon, V.J., Tsai, Y.-J., 2013. An arc cover–path-cover formulation and strategic analysis of alternative-fuel station locations. Eur. J. Oper. Res. 227, 142–151. https://doi.org/10.1016/j.ejor.2012.11.033.

Church, R., ReVelle, C., 1974. The maximal covering location problem. Papers Reg. Sci. Assoc. 32, 101–118.

Connecticut Hydrogen-Fuel Cell Coalition [WWW Document], n.d. URL http://chfcc.org/ (accessed 2.7.19).

de Vries, H., Duijzer, E., 2017. Incorporating driving range variability in network design for refueling facilities. Omega 69, 102–114.

Fan, Y., Lee, A., Parker, N., Scheitrum, D., Dominguez-Faus, R., Myers Jaffe, A., Medlock III, K., 2017. Geospatial, temporal and economic analysis of alternative fuel infrastructure: the case of freight and U.S. natural gas markets. Energy J. 38. https://doi.org/10.5547/01956574.38.6.yfan.

Frade, I., Ribeiro, A., Gonçalves, G., Antunes, A.P., 2011. Optimal location of charging stations for electric vehicles in a neighborhood in Lisbon, Portugal. Transp. Res. Rec. 2252, 91–98. https://doi.org/10.3141/2252-12.

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

Guo, Z., Deride, J., Fan, Y., 2016. Infrastructure planning for fast charging stations in a competitive market. Transp. Res. Part C: Emerg. Technol. 68, 215–227. https://doi.org/10.1016/j.trc.2016.04.010.

He, S.Y., Kuo, Y.-H., Wu, D., 2016. Incorporating institutional and spatial factors in the selection of the optimal locations of public electric vehicle charging facilities: a case study of Beijing, China. Transp. Res. Part C: Emerg. Technol. 67, 131–148. https://doi.org/10.1016/j.trc.2016.02.003.

Hodgson, M.J., 1990. A flow capturing location allocation model. Geogr. Anal. 22, 270–279. https://doi.org/10.1111/j.1538-4632.1990.tb00210.x.

Hong, S., Kuby, M., 2016. A threshold covering flow-based location model to build a critical mass of alternative-fuel stations. J. Transp. Geogr. 56, 128–137.

Jochem, P., Brendel, C., Reuter-Oppermann, M., Fichtner, W., Nickel, S., 2016. Optimizing the allocation of fast charging infrastructure along the German autobahn. J. Bus. Econ. 86, 513–535. https://doi.org/10.1007/s11573-015-0781-5.

Kelley, S., 2017. AFV refueling stations and the complexity of freeway intersections: the scale dependency of network representation. Int. J. Geogr. Info Sci. 31, 346–363. https://doi.org/10.1080/13658816.2016.1202416.

Kelley, S., Kuby, M., 2013. On the way or around the corner? Observed refueling choices of alternative-fuel drivers in Southern California. J. Transp. Geogr. 33, 258–267. https://doi.org/10.1016/j.jtrangeo.2013.08.008.

Kim, J.-G., Kuby, M., 2013. A network transformation heuristic approach for the deviation flow refueling location model. Comput. Oper. Res. 40, 1122–1131.

Kim, J.-G., Kuby, M., 2012. The deviation-flow refueling location model for optimizing a network of refueling stations. Int. J. Hydrogen Energy 37, 5406–5420. https://doi.org/10.1016/j.ijhydene.2011.08.108.

Kuby, M., 2019. The opposite of ubiquitous: How early adopters of fast-filling alt-fuel vehicles adapt to the sparsity of stations. J. Transp. Geogr. 75, 46–57. https://doi.org/10.1016/j.jtrangeo.2019.01.003.

Kuby, M., Capar, I., Kim, J.-G., 2016. Efficient and equitable transnational infrastructure planning for natural gas trucking in the European Union. Eur. J. Oper. Res. 257, 979–991. https://doi.org/10.1016/j.ejor.2016.08.017.

Kuby, M., Lines, L., Schultz, R., Xie, Z., Kim, J.-G., Lim, S., 2009. Optimal location strategy for hydrogen refueling stations in Florida. Int. J. Hydrogen Energy 34, 6045–6064. https://doi.org/10.1016/j.ijhydene.2009.05.050.

Kuby, M.J., Kelley, S.B., Schoenemann, J., 2013. Spatial refueling patterns of alternative-fuel and gasoline vehicle drivers in Los Angeles. Transp. Res. Part D: Transp. Environ. 25, 84–92. https://doi.org/10.1016/j.trd.2013.08.004.

Kuby, M.J., Lim, S., 2005. The flow-refueling location problem for alternative-fuel vehicles. Soc. Econ. Plann. Sci. 39, 125–145. https://doi.org/10.1016/j.seps.2004.03.001.

Lin, Z., Ogden, J., Fan, Y., Chien-Wei, Chen, 2008. The fuel-travel-back approach to hydrogen station siting. Int. J. Hydrogen Energy 33, 3096–3101. https://doi.org/

10.1016/j.ijhydene.2008.01.040.
Lines, L., Kuby, M., Schultz, R., Xie, Z., Lim, S., Kim, J.-G., Clancy, J., 2007. Location strategies for the initial hydrogen refueling infrastructure in Florida. Presented at the Proceedings of the National Hydrogen Association Annual Hydrogen Conference.
Lopez Jaramillo, O., Stotts, R., Kelley, S., Kuby, M., 2019. Content analysis of interviews with hydrogen fuel cell vehicle drivers in Los Angeles. Transp. Res. Rec. https://doi.org/10.1177/0361198119845355. 0361198119845355.
MirHassani, S., Ebrazi, R., 2012. A flexible reformulation of the refueling station location problem. Transp. Sci. 47, 617–628. https://doi.org/10.1287/trsc.1120.0430.
Nicholas, M.A., Handy, S.L., Sperling, D., 2004. Using geographic information systems to evaluate siting and networks of hydrogen stations. Transp. Res. Rec. 1880, 126–134. https://doi.org/10.3141/1880-15.
Owen, S.H., Daskin, M.S., 1998. Strategic facility location: a review. Eur. J. Operat. Res. 111, 423–447. https://doi.org/10.1016/S0377-2217(98)00186-6.
Revelle, C.S., Swain, R., 1970. Central facilities location. Geogr. Anal. 2, 30–42. https://doi.org/10.1111/j.1538-4632.1970.tb00142.x.
Sathaye, N., Kelley, S., 2013. An approach for the optimal planning of electric vehicle infrastructure for highway corridors. Transp. Res. Part E: Logistics Transp. Rev. 59, 15–33. https://doi.org/10.1016/j.tre.2013.08.003.
Sperling, D., Kitamura, R., 1986. Refueling and new fuels: an exploratory analysis. Transp. Res. Part A 20A, 15–23. https://doi.org/10.1016/0191-2607(86)90011-7.
Stephens-Romero, S.D., Brown, T.M., Kang, J.E., Recker, W.W., Samuelsen, G.S., 2010. Systematic planning to optimize investments in hydrogen infrastructure deployment. Int. J. Hydrogen Energy 35, 4652–4667.
Upchurch, C., Kuby, M.J., Lim, S., 2009. A capacitated model for location of alternative-fuel stations. Geogr. Anal 41, 85–106.
US Department of Energy, O. of E.E. and R.E., 2016. Fact #948: October 24, 2016 Carbon Dioxide Emissions from Transportation Exceeded those from the Electric Power Sector for the First Time in 38 Years - Dataset.
U.S. Energy Information Administration, 2018. Monthly Energy Review (No. April 2018). US Energy Information Administration, Washington, D.C.
Wang, Y.-W., Lin, C.-C., 2009. Locating road-vehicle refueling stations. Transp. Res. Part E: Logistics Transp. Rev. 45, 821–829. https://doi.org/10.1016/j.tre.2009.03.002.
Yıldız, B., Arslan, O., Karaşan, O.E., 2016. A branch and price approach for routing and refueling station location model. Eur. J. Oper. Res. 248, 815–826. https://doi.org/10.1016/j.ejor.2015.05.021.