Sparse Subspace Clustering Methods for Tolling Zone Definition within Distance-based Toll Optimization

Antonis F. Lentzakis
NCS Hub, NCS Pte Ltd.
5 Ang Mo Kio Street 62, Singapore 569141
Email: aflentz@mit.edu

Ravi Seshadri
DTU Management, Technical University of Denmark
Bygningstorvet, 116, 116A
Kongens Lyngby, Denmark
Email: ravse@dtu.dk

Moshe Ben-Akiva
Intelligent Transportation Systems Lab, MIT
77 Massachusetts Avenue, Room 1-181
Cambridge, M.A., 02139, U.S.A.
Email: mba@mit.edu

Word Count: 5953 words + 3 table(s) x 250 = 6703 words + 6 figures

Prepared for presentation at 101st Transportation Research Board Annual Meeting, Washington D.C., and being included on Annual Meeting Online.
Submission Date: August 1, 2021

ABSTRACT

Congestion pricing is a standard approach to mitigate traffic congestion in a number of urban networks around the world. The advancement of satellite technology has spurred interest in distance-based congestion pricing schemes, which obviate the need for fixed infrastructure such as gantries that are used in area- and cordon-based pricing. Moreover, distance-based pricing has the potential to more effectively manage traffic congestion. In the context of distance-based congestion pricing, we propose the use of sparse subspace clustering methods employing Elastic Net optimization (SSCEL) and Orthogonal Matching Pursuit (SSCOMP), for the definition of tolling zones. These tolling zone definitions are then used within a simulation-based framework for real-time predictive distance-based toll optimization to examine network congestion and performance of the tolling schemes. Within this framework, for a given definition of tolling zones, tolling function parameters are optimized in real-time using a simulation-based Dynamic Traffic Assignment (DTA) model. Guidance information generation and predictive optimization are integrated and behavioral models behavioral responses to the information and tolls along dimensions of departure time, route, mode, and trip cancellation. For the evaluation of network performance we make use of Travel Speed Index (TSI) data from the real-world Boston Central Business District urban network and demonstrate that the definitions of tolling zones derived from the sparse subspace clustering are effective means of operationalizing real-time distance-based toll optimization schemes and can positively impact overall network performance, showing improvements in average travel time and social welfare relative to the baseline.

Keywords: Distance-based Toll Optimization, Density-based Clustering, Sparse Subspace Clustering

INTRODUCTION

Traffic congestion is a serious issue world-wide, which results in large costs to travelers, the environment and economy. Congestion was estimated to result in a total of 5.5 billion hours of time delay and 2.9 billion gallons of fuel expenditure in urban areas in the United States between 2000 and 2010 (*1*) and the costs of congestion were projected to increase from $121 billion in 2011 to $199 billion in 2020. Mitigating congestion is always a high-priority and also impacts transportation network reliability, driver's comfort and traffic safety. Congestion pricing is a standard approach for congestion mitigation that influences traveler behavior along several dimensions: trip making and frequency, mode, destination, time of day, route, and so on. Traditional approaches to congestion pricing include Facility-based and area-based schemes (*2*) that rely on physical infrastructure such as gantries or gates for vehicle detection. Unfortunately, the reliance on fixed physical infrastructure makes it difficult to modify or relocate the charging areas or zones. Moreover, these schemes can result in efficiency in terms of congestion mitigation since they do not differentiate toll charges based on the associated externalities or congestion caused (due to differing distances traveled or time spent in congestion). The aforementioned disadvantages of area- and facility-based pricing and the advancement of Global Navigation Satellite Systems (GNSS) have focused attention on usage-based tolling wherein toll charges depend on the distance-traveled or the time spent in congestion (see (*3*) and (*4*) for a detailed discussion on the comparative performance of distance- and time-based schemes). Singapore is in the process of transitioning to such a GNSS-based electronic road-pricing scheme (ERP2) (*5*).

Past research on area and cordon-based real-time toll optimization has typically applied reactive approaches (where the optimization of tolls is not based on forecasts of future traffic conditions, but rather on prevailing traffic conditions) for small corridor networks and there are few studies that adopt a predictive approach in the context of large networks (*6, 7*). A more detailed discussion of cordon and area-based real-time toll optimization may be found in (*7*). As noted previously, in contrast with cordon- and area-based schemes, distance-based tolling schemes involve partitioning the network into zones, and levying a toll within each zone that is a function of distance traveled (linear toll functions are considered in (*8–10*), and piece-wise linear functions are used in (*11–13*)). Distance-based toll optimization problems have largely been formulated as simulation-based optimization problems (*10, 14, 15*), non-linear programs (*9*) and mathematical programs with equilibrium constraints or MPEC (*11, 12*), which are solved by global optimization approaches (*11*), meta heuristics (*12, 14*), reinforcement learning (*8*) and feedback controllers (*10, 15*). With the exception of (*14*), these approaches are based on prevailing network conditions (i.e., they are reactive as opposed to proactive), and do not consider elastic demand or the integration of guidance information generation.

Several studies have also examined the partitioning of networks utilizing flow, speed and density data (*16–18*) for the design of traffic management schemes utilizing the Network Fundamental Diagram (NFD) concept. Although area- and cordon-based pricing has been studied in great detail (*19–22*), distance-based pricing in particular has only recently received attention on idealized networks (*23*), using nested regions (*10*) and at the link-level (*24*).

With the exception of (*14*), there has been limited research on systematic approaches for the definition of tolling zones within distance-based toll optimization strategies. Due to the increasing significance of distance-based road pricing in traffic network management and operations, this paper addresses the problem of how to define tolling zones and proposes the application of sparse subspace clustering methods to define parsimonious sets of tolling zones. The performance of

these methods is evaluated within a framework for real-time toll optimization which generates predictive optimized distance-based toll strategies combined with guidance information. This paper contributes to the existing literature in the following respects:

1. We apply sparse subspace clustering methods for the definition of tolling zones that are feature variant, and utilize location coordinates and normalized link speeds.
2. The proposed subspace clustering methods are evaluated using a framework for real-time distance-based predictive toll optimization on the Boston CBD network and yield insights into their performance (and suitability for deployment) against other density-based clustering methods.

FRAMEWORK, MODEL AND PROBLEM DESCRIPTION

In this section, we summarize the real-time distance-based predictive toll optimization framework (more details may be found in (*14*)), the optimization problem formulation, the proposed clustering methods for tolling zone definition and the algorithmic solution for the optimization problem.

Framework

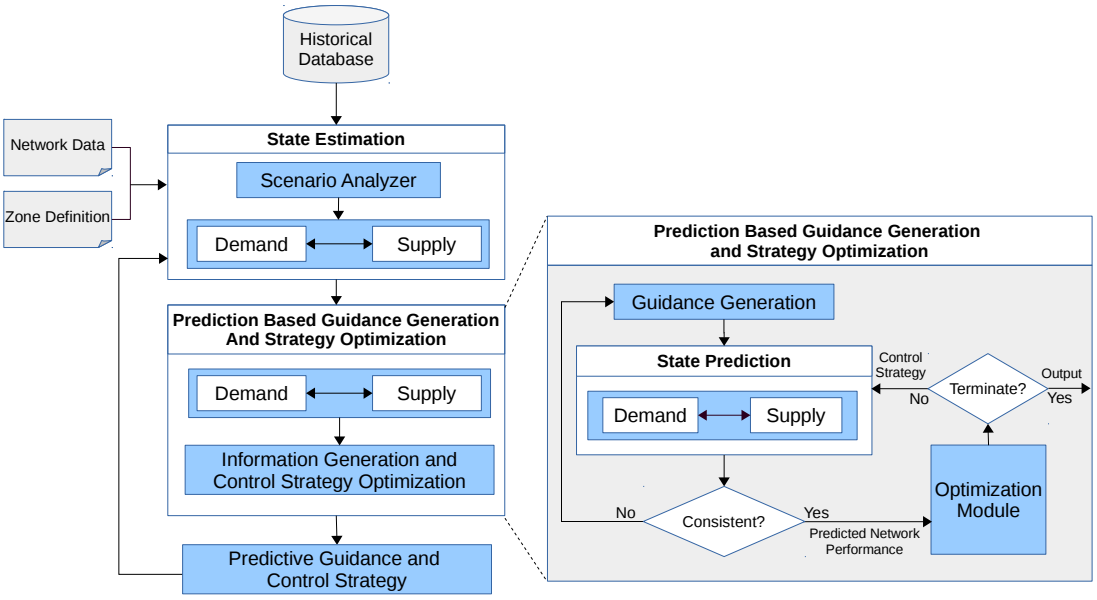


FIGURE 1: Real-time distance-based predictive toll optimization framework

The framework, shown in Figure 1, uses DynaMIT2.0 - a simulation-based Dynamic Traffic Assignment (DTA) system developed at the MIT Intelligent Transportation Systems Lab, (*25, 26*). DynaMIT2.0 employs a rolling horizon approach involving two key modules, state estimation and state prediction. The state estimation process involves using a combination of historical data, real-time traffic surveillance data, and prevailing network control strategies (such as distance-based toll optimization) to estimate the current state of the network. It used detailed models of demand (pre-trip models of departure time, route and mode choice), supply (mesoscopic traffic simulator that combines speed-density relationships and a deterministic queuing model) and their interactions. Following this, the state prediction module generates forecasts of traffic conditions for a

pre-specified prediction horizon (origin-destination demands and supply parameters are forecasted for the future using an autoregressive process). The strategy optimization and guidance generation modules in conjunction use the state predictions to first, optimize control strategies for the prediction horizon and second, generate guidance information (traveler information) for the prediction horizon. The evaluation of candidate control strategies makes use of network predictions and guidance information that are consistent, i.e., the guidance information is as close as possible to actual predicted network travel times (see Figure 1 and (*25*) for more on this aspect of consistency).

Problem Description

We represent the transportation network of interest as a directed graph $\mathscr{G} = (\mathscr{N}, \mathscr{A})$, where $\mathscr{N}$ denotes the set of n network nodes and $\mathscr{A}$ denotes the set of m links. The network is partitioned into $l = 1 \dots L$ tolling zones, where every zone l is defined by a subset of network links $\mathscr{A}_l \subseteq \mathscr{A}$. For each zone l, we define a tolling function $\phi_l(\theta_l^t, D_l)$ that maps distance traveled within the zone l, D_l to the toll amount; θ_l^t is a vector of parameters that defines the tolling function in time interval t. Further, it is assumed that the toll payable in a zone is bounded, i.e $\tau_{LB} \leq \phi_l(\theta_l^t, D_l) \leq \tau_{UB}, \forall l = 1, 2..., L \; \forall t = 1, 2, ..., T$.

Denote the length of the state estimation interval in DynaMIT by Δ (usually 5 minutes) and assume that the prediction horizon is composed of H such intervals so that the size of the prediction horizon is $H\Delta$. We assume that the prediction horizon and the optimization horizon are identical. Further, the tolling function parameters do not vary within a given time interval of size Δ and these tolling intervals coincide with DynaMIT estimation intervals. For an arbitrary estimation interval $[t_0 - \Delta, t_0]$, let $\theta^h = (\theta_1^h, \theta_2^h ... \theta_L^h)$ represent the vector of tolling function parameters for the time period $[t_0 + (h-1)\Delta, t_0 + h\Delta]$ where $h = 1, ..., H$. Accordingly, for the current optimization horizon, the decision variables are $\theta = (\theta^1, \theta^2, ..., \theta^H)$.

Consider the set of vehicles $v = 1, ..., V$ that are on the network during the prediction horizon $[t_0, t_0 + H\Delta]$. For each vehicle v, we denote the experienced trip travel by time tt^v and the predictive guidance information by $\mathbf{tt^g} = (\mathbf{tt_i^g}; \forall i \in A)$, where $\mathbf{tt_i^g}$ represents a vector of time dependent travel times for link i . Note that the vehicle travel times $\mathbf{tt} = (tt^v; v = 1, ..., V)$ are obtained from the state prediction module of DynaMIT2.0, which we characterize through a single constraint that represents the coupled demand and supply simulators as:

$$G(\mathbf{x^p}, \boldsymbol{\gamma^p}, \mathbf{tt^g}, \theta) = \mathbf{tt} \tag{1}$$

Where $\mathbf{x^p}$, $\boldsymbol{\gamma^p}$ represent the forecasted demand and supply parameters for the prediction horizon, and θ is the vector of tolling function parameters. As noted previously, the state prediction module ensures consistency between $\mathbf{tt^g}$ and $\mathbf{tt}$.

Pre-trip Model

The pre-trip response of users to the travel time guidance and distance-based tolls is modeled using a path-size nested logit model with heterogeneous value of time (illustrated in Figure 2) that captures decisions of mode choice, trip cancellation, departure time and path (notation is provided in Table 1). We provide a brief description of the model here for completeness (more details may be found in (*14*)). In response to pre-trip information and tolls, a traveler may alter his/her habitual travel pattern, which may include changing mode, canceling trip, changing departure time or path, or changing departure time and path. This results in elastic total demand w.r.t. traffic congestion.

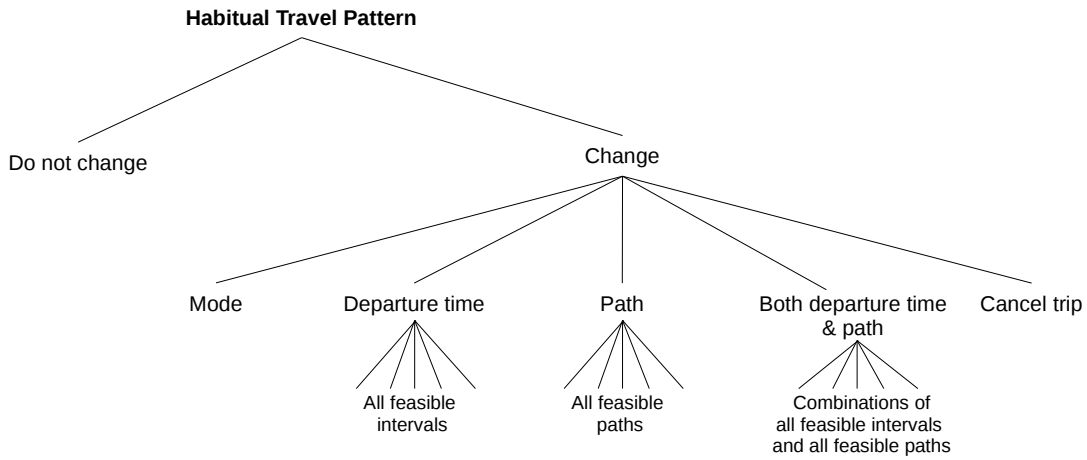


FIGURE 2: Pre-trip Behavior Model

TABLE 1: Pre-trip Model - Abbreviations

Abbreviation	Variable
β_{CM}	Alternative Specific Constant (ASC) for change of mode to transit
β_{CT}	ASC for canceling trip
β_{CDT_d}	ASC for departure in time interval d
c_m^v	monetary cost for traveling with non-private (transit) mode
c_{dp}^v	toll charge for departure via path p in interval d
c_p^v	toll for switching to path p
t_m^v	travel time associated with non-private (transit) mode
t_{dp}^g	travel time (guidance) for departure via path p in interval d
$at_{d'p'}^{hab}$	arrival time (habitual)
at_{dp}^g	arrival time (predicted) for departure via path p in time interval d
β_c^v	monetary cost coefficient
β_t^v	travel time coefficient
β_E	schedule delay early coefficient
β_L	schedule delay late coefficient
PS_p	path size variable
C_*	utility relating to number of left turns/signalized intersections and path length
ε_*	error component

The options of mode modeled are private car (drive alone) and public transit. The utility of change to a transit for vehicle v is given by:

$$U^v(CM) = \beta_{CM} + \beta_c^v c_m^v + \beta_t^v t_m^v + \varepsilon_m \quad (2)$$

The utility of departing at time interval d and choosing path p for vehicle v is given by:

$$\begin{aligned} U_{dp}^{v} &= \beta_{CDT_{dp}} + \beta_{c}^{v} c_{dp}^{v} + \beta_{t}^{v} t_{dp}^{g} \\ &\quad + \beta_{E} \max(at_{d'p'}^{hab} - at_{dp}^{g}, 0) + \beta_{L} \max(at_{dp}^{g} - at_{d'p'}^{hab}, 0) \\ &\quad + log(PS_{p}) + C_{dp} + \varepsilon_{dp} \\ where: & \\ c_{dp}^{v} &= \sum_{l=1}^{L} \phi_{l}(\eta_{p,l}^{v}) \end{aligned} \tag{3}$$

The utility of canceling trip altogether is given by:

$$U^{v}(CT) = \beta_{CT} + \varepsilon_{CT} \tag{4}$$

Thus, the probability of vehicle v choosing alternative c within the choice set C is given by:

$$P^{v}(c|C) = \frac{e^{\mu V_{c}^{v}}}{\sum_{a \in C} e^{\mu V_{a}^{v}}} \tag{5}$$

where V_c^v is the systematic utility given by $V_c^v = U_c^v - \varepsilon_c$ and μ is a scale parameter .

The objective function for the toll optimization problem, formulated from the standpoint of the traffic regulator, is total social welfare (SW), which is the sum of the consumer surplus and the producer surplus. In this context, the consumer surplus (CS) is defined as the sum of the experienced utilities across all travelers and the producer surplus is the net revenue, denoted by TP, which is simply the toll revenue minus variable costs (fixed costs are ignored), $TP = TR - VC$. We assume that the variable costs are a proportion of the toll revenue (the proportionality factor is denoted by $\alpha < 1$). Thus, the social welfare is given by:

$$\begin{aligned} SW &= CS + TP \\ &= \sum_{v=1}^{V} \frac{U^{v}}{|\beta_{c}^{v}|} + \left[(1-\alpha) \times \sum_{v=1}^{V} c^{v} \right] \end{aligned} \tag{6}$$

The absolute value of β_c^v is used to translate CS into dollar equivalents. The distance-based toll optimization problem is formulated as a simulation-based optimization problem in Equation 7, where the objective is social welfare, the decision variables are the vector of tolling function parameters for the current optimization horizon, and the constraints are toll bounds and the DTA model system.

$$\begin{aligned} \mathbf{DTOP}: & \max_{\theta} \left[\sum_{v=1}^{V} \frac{U^{v}}{|\beta_{c}^{v}|} + (1-\alpha) \times \sum_{v=1}^{V} c^{v} \right] \\ & \mathbf{s.t.} \\ & G(\mathbf{x^p}, \boldsymbol{\gamma^p}, \mathbf{tt^g}, \theta) = \mathbf{tt} \\ & \tau_{LB} \leq \phi_{l}(\theta_{l}^{h}, D_{l}^{v}) \leq \tau_{UB}, \forall v = 1,2,...,V; l = 1,2...,L; h = 1,2,...,H \end{aligned} \tag{7}$$

Definition of Tolling Zones using SSCEL, SSCOMP, OPTICS and HDBSCAN*,

One of the inputs with significant impact to our distance-based tolling system performance is the tolling zone definition. This input specifies which links belong to each tolling zone and the number of zones. Each tolling function $\phi_l(\theta_l^h, D_l^v)$ corresponds to one tolling zone. Past literature for partitioning urban traffic networks used datasets based on speed, flow, density (*16–18, 27*) and, more recently, marginal cost toll data (*14*). In our case, the travel speed index (TSI) is used, a widely used quantitative indicator that employs link speed normalization (*28*), given the fact that identical link speed levels might reflect different traffic conditions. Elhamifar and Vidal (*29*), inspired by compressed sensing (*30*), introduced Sparse Subspace Clustering (SSC), which makes use of the self-expressiveness property to construct the affinity matrix, upon which spectral clustering is applied, in order to derive the underlying subspaces. While subspace clustering methods have been used extensively for, among others, temporal video segmentation, switched system identification, (*31, 32*), only recently, has this technique come to the attention of the transportation research community. Zhang et al. (*33*) employed SSC to classify spatiotemporal taxi patterns with regards to their passenger searching behavior. For our experiments we selected to compare two Sparse Subspace Clustering variants, SSCEL and SSCOMP, employing Elastic Net optimization (*34*) and Orthogonal Matching Pursuit (*35*) respectively, against two well-known hierarchical density-based clustering methods, OPTICS (*36*), (Ordering Points To Identify the Clustering Structure), and HDBSCAN* (*37*), (Hierarchical Density-Based Spatial Clustering of Applications with Noise) .

Sparse Subspace Clustering Methods

For the Sparse Subspace Clustering application, we selected two variants,SSCEL and SSCOMP. We exploited the property of self-representation to learn the affinity matrix, to be subsequently used in our implementation of spectral clustering. Data self-expressiveness (*38*) describes the fact that a data point found in a union of subspaces can be represented as the linear combination of other data points, expressed through the following optimization problem:

$$\begin{aligned} &\min_{C} \|\mathbf{C}\|_1 \\ &\textbf{s.t.} \\ &\mathbf{X} = \mathbf{XC} \\ &diag(\mathbf{C}) = 0 \end{aligned} \tag{8}$$

Where $\mathbf{X} \in \mathbf{R}^{D \times N}$ is the data point matrix and $\mathbf{C} \in \mathbf{R}^{N \times N}$ is the self-expression coefficient matrix. In practice, however, solving N such problems over N variables may be computationally expensive for large N. If we choose to express the optimization problem as follows:

$$\begin{aligned} &\min_{c_j} \|\mathbf{x}_j - \mathbf{X}\mathbf{c}_j\|_2^2 \\ &\textbf{s.t.} \\ &\|\mathbf{c}_j\|_0 \leq \mathbf{k} \\ &diag(\mathbf{C}) = 0 \end{aligned} \tag{9}$$

We can efficiently solve it using the Orthogonal Matching Pursuit algorithm, as described in (*35*). Orthogonal Matching Pursuit selects a single column of $\mathbf{X}$ each time, $\mathbf{x}_j$, such that the absolute value of the dot product with the residual $\mathbf{c}_j$ is maximized and the coefficients are computed until k columns are selected. Subsequently, we learn the affinity matrix $\mathbf{W}$ through data self-representation

as: $\mathbf{W} = |\mathbf{C}| + \left|\mathbf{C}^T\right|$. Alternatively, we may employ Elastic Net regularization for scalable subspace clustering. Following (*34*), we used an active set algorithm that efficiently solves the elastic net regularization subproblem, which follows below, by capitalizing on the geometric structure of the elastic net solution:

$$\min_{c_j} \lambda \|\mathbf{c}_j\|_1 + \frac{1-\lambda}{2}\|\mathbf{c}_j\|_2^2 + \frac{\gamma}{2}\|\mathbf{x}_j - \mathbf{X}\mathbf{c}_j\|_2^2 \tag{10}$$

Where $\lambda \in (0,1]$ and $\gamma > 0$. In the majority of solution approaches for the Subspace Clustering problem, after learning the affinity matrix, spectral clustering is applied to the resulting matrix to derive the final clustering.

Hierarchical Density-based Clustering Methods

Hierarchical density-based clustering methods are gaining traction among the research community, exhibiting robustness during parameter selection and being able to cope with clusters characterized by large inter-cluster density variability, unlike their non-hierarchical predecessor, DBSCAN (*39*). OPTICS utilizes hyperparameters ε and κ, representing the maximum ball radius with each data point at its center and the minimum density threshold, respectively. Assuming a metric space (X,d) comprising of a set of data points $X = \{x_1, x_2, ..., x_n\}$, a data point x is considered to be a core point with respect to ε and κ if its ε-neighborhood $N_\varepsilon(x)$ contains a minimum of κ data points. Two core points x_i, x_j are ε-reachable with respect to ε and κ if they are both contained within each others ε-neighborhood. Two core points x_i, x_j are density-connected with respect to ε and κ if they are directly or transitively ε-reachable. A cluster is the largest possible group of data points, where each two points are considered connected in terms of density. In OPTICS data points are assigned a core distance $d_{\text{core}}^{\varepsilon,\kappa}(x)$ to the κ-th nearest neighbor, for varying degrees of density. The reachability-distance $d_{\text{reach}}^{\varepsilon,\kappa}(x_i, x_j)$ is the maximum between the core distance of x_i and the distance between data points x_i,x_j. A single global ε' value is used to extract a flat clustering.

HDBSCAN* is similar to OPTICS with parameter $\varepsilon = \infty$ and a different technique, based on cluster stability, is utilized for flat clustering. In the case of HDBSCAN*, we have $d_{\text{core}}^{\kappa}(x_i)$ representing the κ-th nearest neighbor distance. For a fixed κ and a range of possible ε values, the mutual reachability distance $d_{\text{mreach}}^{\kappa}(x_i, x_j)$ is used to generate a complete hierarchy of clusterings. Thus, for any fixed ε value, the clustering produced by DBSCAN at a given level in the hierarchy is the clustering obtained for the corresponding ε value. The selected hierarchical density-based clustering methods result in clusterings where some data points are considered noise. A feasible definition of tolling zones, however, must involve the assignment of all data points to clusters. In order to address this issue, we perform a secondary assignment where all noise data points are assigned to cluster exemplars (using minimum Euclidean distance).

Clustering Performance Metrics

As clustering performance metrics, the Silhouette Coefficient (SC) (*40*) and the Davies-Bouldin index (DB)(*41*) were selected. SC is the average for the entire dataset of the silhouette, which measures cohesion and separation for each cluster and ranges from $[-1,1]$, where -1 represents incorrect clustering, 0 represents overlapping clusters and 1 represents highly dense clustering. DB is a function of the ratio of intra-cluster scatter to inter-cluster separation. DB values closer to 0 indicate a better clustering result.

Clustering Method Results

It would be preferable that the selected clustering methods produce clustering results that are of high quality, according to our previously presented internal evaluation indices, but also do not preclude any sort of practical application, due to high computational cost, incurred on the distance-based toll optimization framework. This translates into a static tolling zone definition (non-varying during the simulation), with a reasonably low number of tolling zones. For our dataset, besides spatial coordinates, we decided to use the travel speed index (TSI) for each link as an additional feature, calculated as follows:

$$\mathrm{TSI}_i = 1 - \frac{v_i}{v_i^f} \tag{11}$$

Where v_i, v_i^f the link speed and free flow speed for link i respectively. Simulated speed data at the segment and link level, obtained from a calibrated DynaMIT2.0 of the Boston CBD (*42*), were used to derive the tolling zone definitions. The Boston CBD network, shown in Figure 3, has 846 nodes, 1746 links, 3085 segments, 5057 lanes and 13080 Origin-Destination pairs.

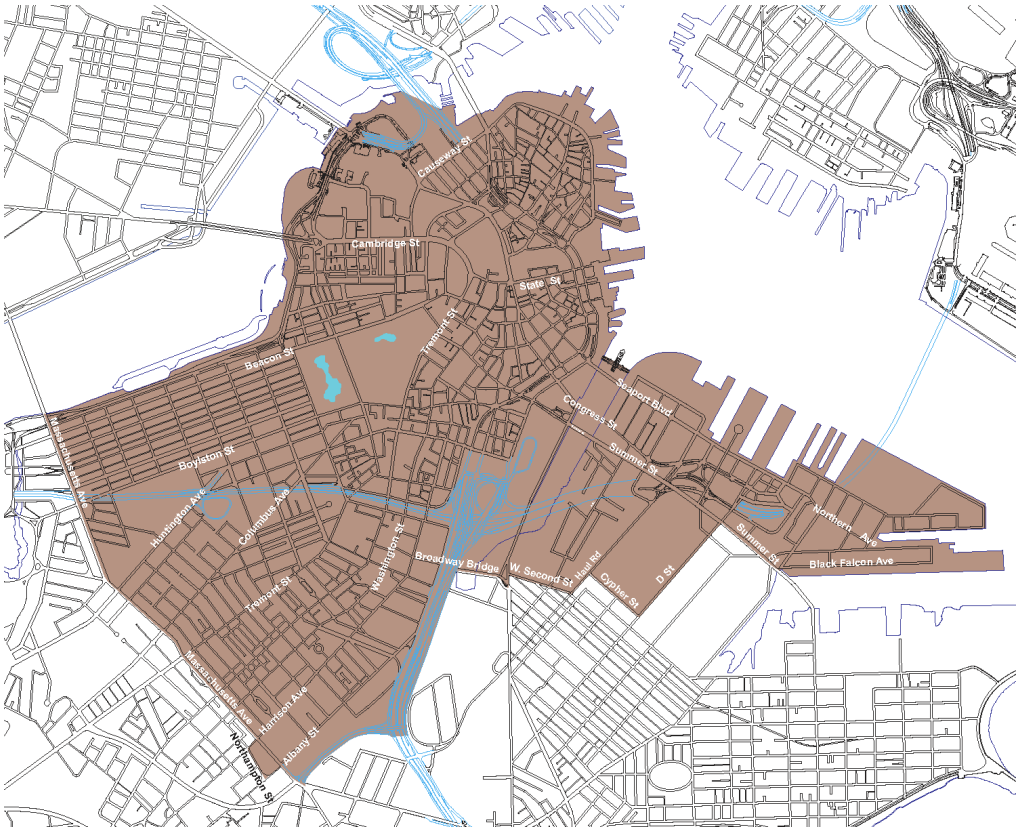


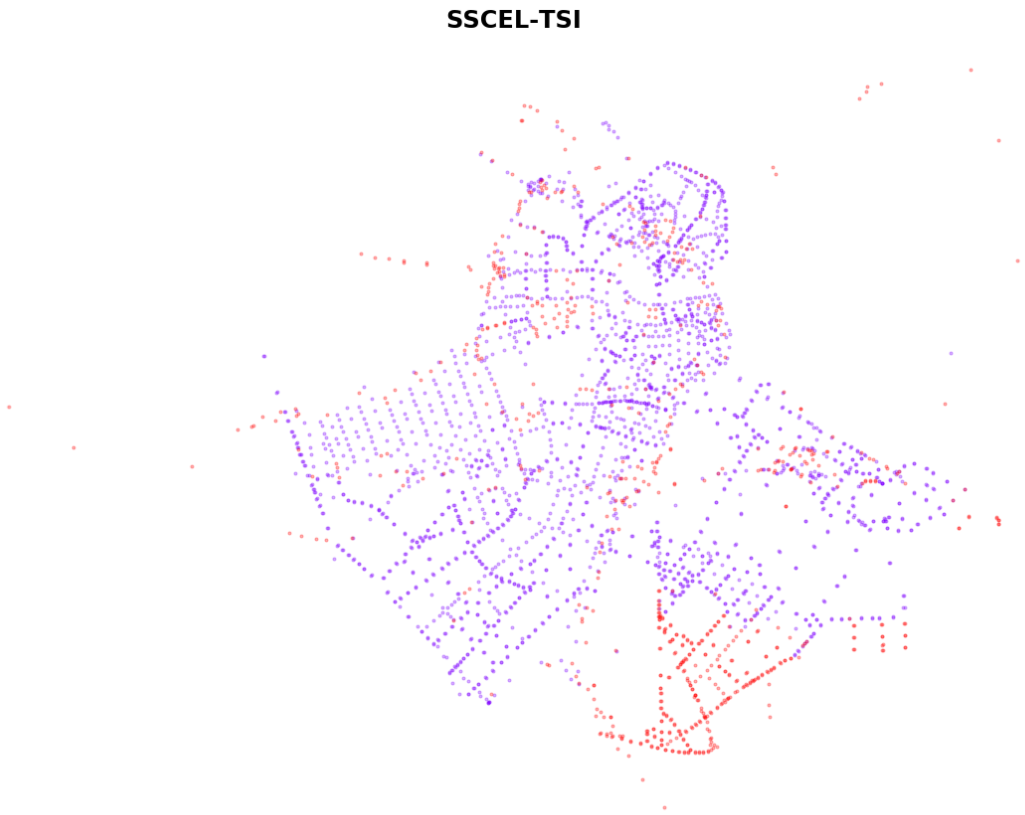
FIGURE 3: Boston CBD Network

In the case of the hierarchical density-based clustering methods implementation (OPTICS, HDBSCAN*), we used the average of TSI across 5-minute intervals to account for the dynamics of TSI variation. In the case of the sparse subspace clustering methods, we were able to use the entirety of our dataset (i.e., all time intervals) to learn the affinity matrix via the SSCEL and SS-COMP algorithms, since self-representation is amenable for use of datasets with spatiotemporal attributes (*43, 44*). In the case of static partitioning schemes derived offline, this is a preferable

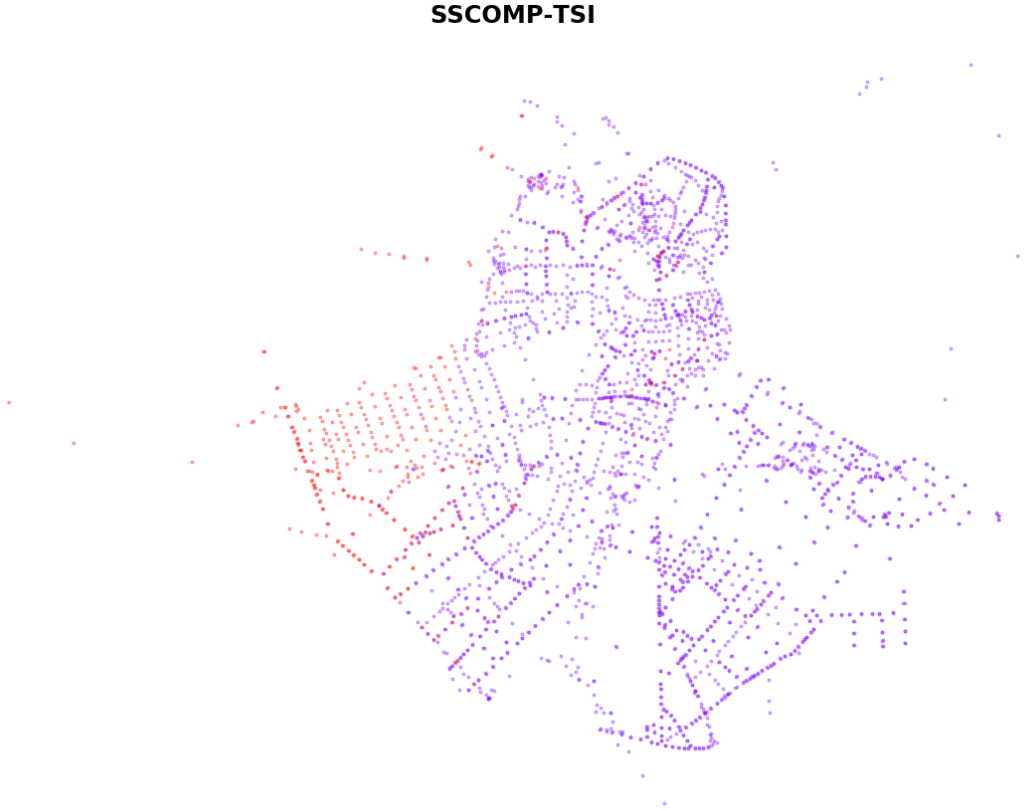
alternative to using the average of TSI across specific intervals, however, spatial compactness may be affected, leading to lower quality clustering results, as represented by our internal evaluation indices. While high quality clustering results, according to our previously presented internal evaluation indices, are desirable, for any sort of practical application, due to the high computational cost incurred on our framework, lower quality clustering results may also be taken into consideration. For each clustering method, the resulting tolling zone definition and corresponding performance index results,are presented in Figures 4,5.

Solution Algorithm

Due to the highly non-convex nature of the objective function in 7, we apply a real-coded Genetic Algorithm (GA) to solve the optimization problem in 7. More details on the GA algorithm may be found in (*14*). Computational performance is enhanced by utilized parallelization wherein the evaluations of different candidate individuals within an iteration of the GA are performed in parallel.

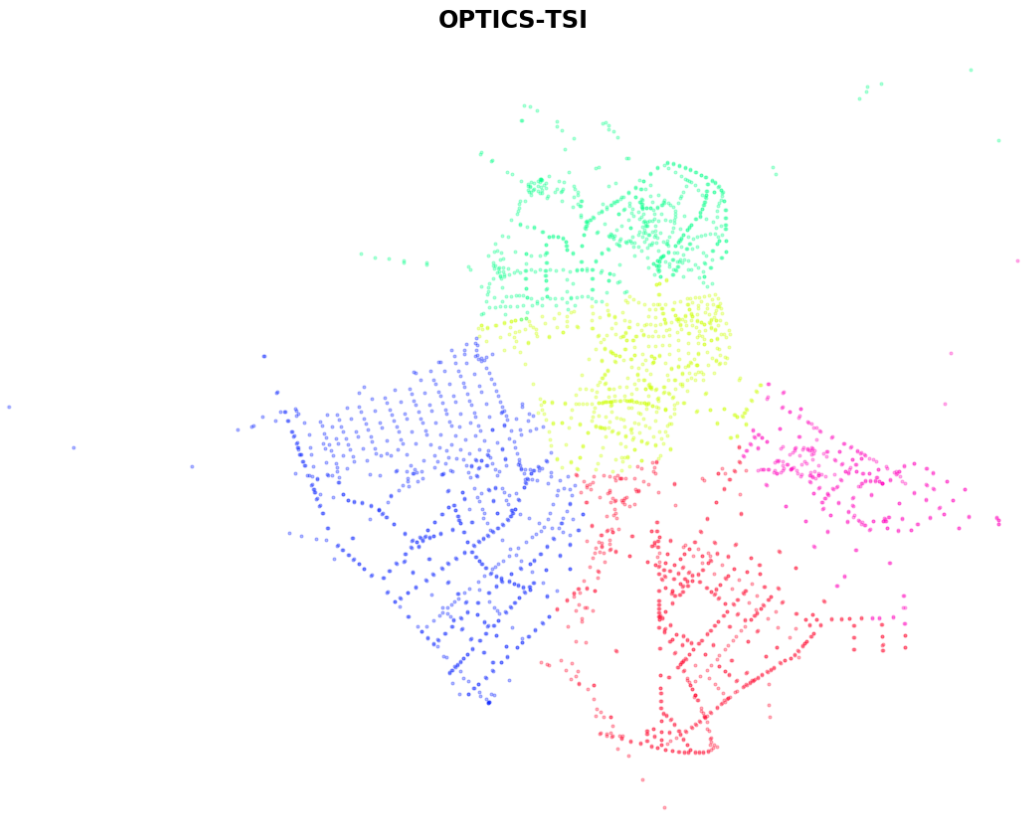


(a) 2 zones derived using SCCEL for feature TSI, with SC=0.114, DB=3.925

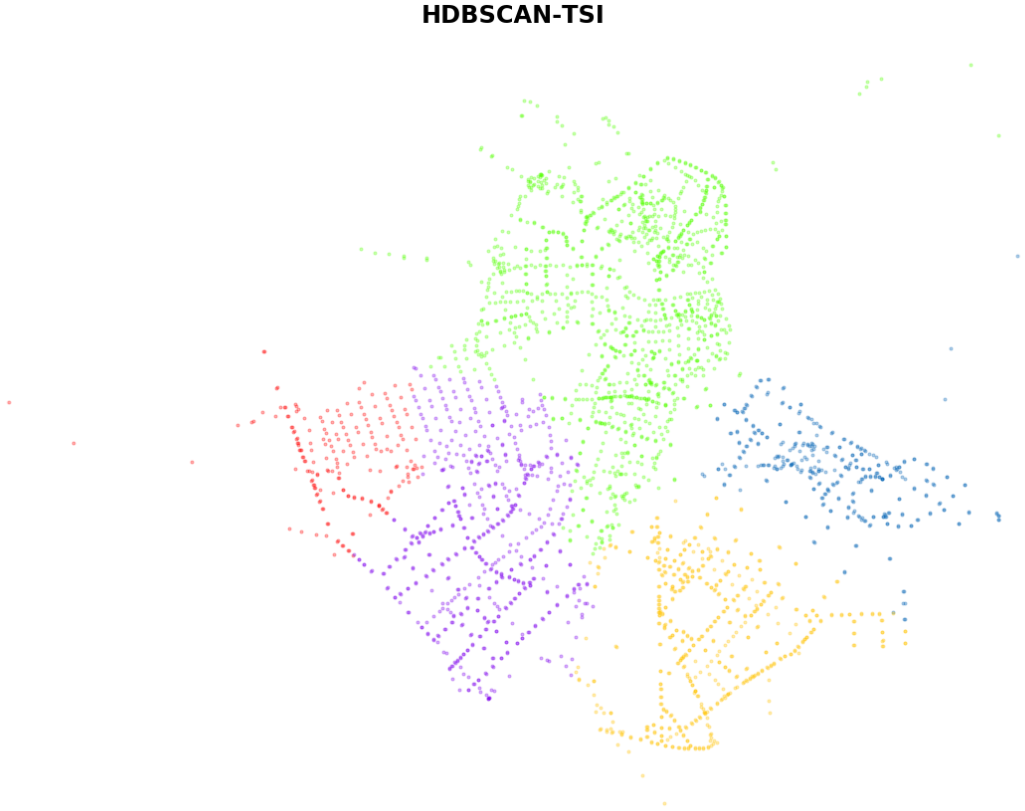


(b) 2 zones derived using SSCOMP for feature TSI, with SC=0.198, DB=1.276

FIGURE 4: Clustering results and tolling zones (Sparse Subspace Clustering)



(a) 5 zones derived using OPTICS for feature TSI, with SC=0.387, DB=0.858



(b) 5 zones derived using HDBSCAN* for feature TSI, with SC=0.400, DB=0.776

FIGURE 5: Clustering results and Tolling zones (Hierarchical Density-based Clustering)

EXPERIMENTS AND RESULTS

Experiment Settings

The Boston CBD network, presented in Figure 3, will be the focus of our experiments, calibrated within DynaMIT2.0 both for demand and supply (*45*). We decided upon a bounded linear tolling function (i.e $\phi_l(\theta_l^t, D_l) = \theta_{l1}^t + \theta_{l2}^t D_l$; and $0 \leq \phi_l(\theta_l^t, D_l) \leq 1.5$). The simulation period is from 06:00-09:00. Estimation interval is 5 minutes and prediction horizon is 30 minutes. Vehicles with habitual departure time within 06:00-09:00 (which time may change corresponding to traffic conditions information), are used for evaluating network and individual driver performance. Warm-up and cool-down periods of 15 min were used with distance-based tolling disabled, to ensure that all vehicles are able to finish their trips. The five scenarios for our case study are summarized in Table 2. All scenarios include predictive distance-based tolling employing our toll optimization framework from Section 3, apart from the base scenario (B0) representing our baseline. It should be noted that the tolling function parameters are updated in 5-minute intervals.

The base scenario **B0** was calibrated to replicate prevailing traffic conditions in the Boston CBD (refer (*45*) for more details). Scenarios **B1, B2, B3, B4** employ predictive distance-based tolling and differ only in the definition of the tolling zones. In scenario **B1** (termed **SSCEL-TSI**), tolling zones are defined based on TSI data using SSCEL. In scenario **B2** (termed **SSCOMP-TSI**), tolling zones are defined based on TSI data using SSCOMP. In scenario **B3** (termed **OPTICS-TSI**), tolling zones are defined based on TSI data using OPTICS, and finally, in scenario **B4** (termed **HDBSCAN*-TSI**) they are based on TSI data using HDBSCAN*. Scenarios **B0**-**B4** are evaluated on three performance measures, total social welfare (SW), consumer surplus (CS) and average travel time (TT) to capture overall societal benefits, together with the impact on individual travelers.

TABLE 2: Simulation scenarios

Scenario	Tolling Scheme	Description
B0	No Toll	*No tolling scheme in place*
B1	Predictive distance-based	*Tolling zones derived from: SSCEL-TSI*
B2	Predictive distance-based	*Tolling zones derived from: SSCOMP-TSI*
B3	Predictive distance-based	*Tolling zones derived from: OPTICS-TSI*
B4	Predictive distance-based	*Tolling zones derived from: HDBSCAN*-TSI*

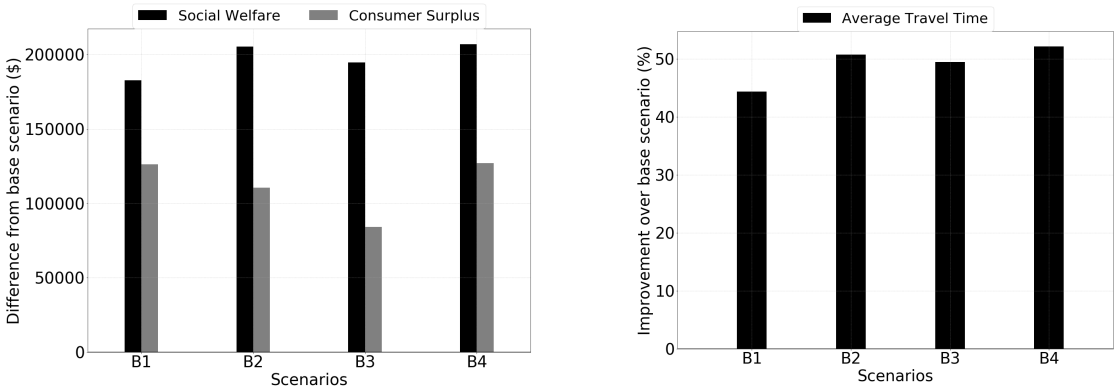
Results

The performance measures for all simulation scenarios are summarized in Table 3, athe differences in SW and CS (in $ amounts) of scenarios **B1, B2, B3, B4** relative to the base scenario B0 are presented in 6a, and the relative performance in terms of average travel time (% improvement) over the base scenario B0 is illustrated in Figure 6b. From Table 3, **B1, B2, B3, B4** exhibit an increase between $182623.5 - $206866.5 and $84083.7 - $127062.4, for SW and CS respectively, relative to **B0**. The average SW gain per traveller, relative to the no toll case is found to be around $1.69 for those acquired via sparse subspace clustering and around $2.15 for tolling zone definitions acquired via hierarchical density-based clustering. Larger variation of differences in CS confirms that the performance of distance-based tolling schemes are significantly affected by the definition of the tolling zones. Relative difference for scenario **B1**, in terms of SW, seem to be the lowest, however, the relative difference in terms of CS is comparable to that of scenario **B4**, the overall best performing scenario. This is due to the fact that scenario **B1** presents with very low revenue, a direct result

of the low quality clustering exhibiting lack of spatial compactness. What is most impressive is the TT performance improvement illustrated in Figure 6b, relative to the base case **B0**. It is evident that the Boston CBD area would benefit from an application of a predictive distance-based tolling scheme, with average travel time TT improvements of up to 52% (relative to **B0**). While it should be stated that predictive distance-based tolling schemes with OPTICS-, HDBSCAN*-, SSCEL- and SSCOMP- derived tolling zone definitions deliver substantial network performance benefits when compared to the no toll scenario, scenario **B4** with HDBSCAN*-derived tolling zone definition had the highest impact on distance-based tolling optimization performance. It should be noted that, scenario **B2** with only 2 tolling zones derived from SSCOMP resulted in very comparable levels of performance to the top performing scenario **B4**, and in cases where computational effort poses a significant hurdle for practical implementation, it would be preferable to use SSCOMP.

TABLE 3: Performance measures

	Scenarios			
Metrics	$B1$	$B2$	$B3$	$B4$
SW ($)	182623.5	205345.1	194744.3	206866.5
CS ($)	126194.4	110504.2	84083.7	127062.4
TT (s)	172.2	152.3	156.3	147.9



(a) Difference in SW, CS for scenarios **B1, B2, B3, B4** relative to **B0**

(b) Percentage Improvement in TT for scenarios **B1, B2, B3, B4** relative to **B0**

FIGURE 6: Performance results for scenarios **B1, B2, B3, B4** relative to **B0**

CONCLUDING REMARKS AND FUTURE RESEARCH

In this paper we investigated the use of sparse subspace clustering methods to define tolling zones for distance-based tolling schemes, and their impact on traffic network performance using a predictive real-time distance-based toll optimization framework. Experiments were conducted on the real-world urban network of the Boston Central Business District. We determined that the best network performance comes from the use of a distance-based tolling zone definition derived from HDBSCAN*, when using Travel Speed Index data. Performance using only 2 tolling zones acquired via the SSCOMP sparse subspace clustering variant was found to be comparable to that of

a 5-zone definition derived from HDBSCAN*, so, in cases where minimizing computational effort is one of the primary objectives, it should be considered as a viable alternative. Despite the fact that all clustering approaches produced tolling zone definitions which, as part of our framework, contributed to significant performance gains, when compared to the No Toll case, we observed large differences in performance between0 tolling zone definitions acquired via the sparse subspace clustering variants. Specifically, for this particular dataset, the SSCEL variant of sparse subspace clustering produced low quality clustering, due to the low degree of spatial compactness. This warrants further investigation, however, overall, tolling zone definitions derived from both types of clustering methods, yielded significant benefits on network performance. In conclusion, we have determined that the best choice to systematically derive a tolling zone definition, for this particular network dataset, is the HDBSCAN* clustering method.

In future work, we aim to evaluate alternate clustering methods for systematic tolling zone definition as part of the distance-based tolling optimization framework. We are also in the process of investigating alternative solution approaches, including Bayesian and Surrogate Optimization, and comparing toll optimization framework performance to that of our currently used solution approach.

ACKNOWLEDGEMENTS

This research was supported by the National Research Foundation of Singapore through the Singapore-MIT Alliance for Research and Technology's FM IRG research programme.

AUTHORS CONTRIBUTION

The authors confirm contribution to the paper as follows: study conception and design: A.F. Lentzakis, R. Seshadri, M. Ben-Akiva; analysis and interpretation of results: A.F. Lentzakis, R. Seshadri; draft manuscript preparation: A.F. Lentzakis, R. Seshadri. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

[1] Litman, T., *Congestion Costing Critique Critical Evaluation of the Urban Mobility 2014 Report*. Victoria Transport Policy Institute, 2014.

[2] de Palma, A. and R. Lindsey, Traffic congestion pricing methodologies and technologies. *Transportation Research Part C: Emerging Technologies*, Vol. 19, 2011, pp. 1377–1399.

[3] Smith, M. J., A. D. May, M. B. Wisten, D. S. Milne, D. Van Vliet, and M. O. Ghali, A comparison of the network effects of four road-user charging systems. *Traffic Engineering and Control*, Vol. 35, 1994, pp. 311–315.

[4] Bonsall, P. W. and I. A. Palmer, Do time-based road-user charges induce risk-taking? - results from a driving simulator. *Traffic Engineering and Control*, Vol. 38, 1997, pp. 200–203.

[5] LTA, *Tender Awarded to Develop Next Generation Electronic Road Pricing System*, 2016.

[6] Gupta, S., R. Seshadri, B. Atasoy, F. C. Pereira, S. Wang, V.-A. Vu, G. Tan, W. Dong, Y. Lu, C. Antoniou, and M. Ben-Akiva, Real time optimization of network control strategies in dynamit2. 0. In *Transportation Research Board 95th Annual Meeting*, 2016, 16-5560.

[7] Gupta, S., R. Seshadri, B. Atasoy, A. A. Prakash, F. Pereira, G. Tan, and M. Ben-Akiva, Real-Time Predictive Control Strategy Optimization. *Transportation Research Record*, 2020.

[8] Zhu, F. and S. V. Ukkusuri, A reinforcement learning approach for distance based dynamic

tolling in the stochastic network environment. *Journal of Advanced Transportation*, Vol. 49, 2015, pp. 247–266.

[9] Yang, L., R. Saigal, and H. Zhou, Distance-based dynamic pricing strategy for managed toll lanes. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2283, 2012, pp. 90–99.

[10] Gu, Z., S. Shafiei, Z. Liu, and M. Saberi, Optimal distance-and time-dependent area-based pricing with the Network Fundamental Diagram. *Transportation Research Part C: Emerging Technologies*, Vol. 95, 2018, pp. 1–28.

[11] Liu, Z., S. Wang, and Q. Meng, Optimal joint distance and time toll for cordon-based congestion pricing. *Transportation Research Part B*, Vol. 69, 2014, pp. 81–97.

[12] Meng, Q., Z. Liu, and S. Wang, Optimal distance tolls under congestion pricing and continuously distributed value of time. *Transportation Research Part E: Logistics and Transportation Review*, Vol. 48, 2012, pp. 937–957.

[13] Sun, X., Z. Liu, R. Thompson, Y. Bie, J. Weng, and S. Chen, A multi-objective model for cordon-based congestion pricing schemes with nonlinear distance tolls . *Journal of Central South University*, Vol. 23, 2016, pp. 1273–1282.

[14] Lentzakis, A. F., R. Seshadri, A. Akkinepally, V.-A. Vu, and M. Ben-Akiva, Hierarchical density-based clustering methods for tolling zone definition and their impact on distance-based toll optimization. *Transportation Research Part C: Emerging Technologies*, Vol. 118, 2020, p. 102685.

[15] Gu, Z. and M. Saberi, A Simulation-Based Optimization Framework for Urban Congestion Pricing Considering Travelers' Departure Time Rescheduling. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, IEEE, 2019, pp. 2557–2562.

[16] Ji, Y. and N. Geroliminis, On the spatial partitioning of urban transportation networks. *Transportation Research Part B: Methodological*, Vol. 46, No. 10, 2012, pp. 1639–1656.

[17] Lentzakis, A. F., R. Su, and C. Wen, Time-dependent partitioning of urban traffic network into homogeneous regions. In *Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on*, IEEE, 2014, pp. 535–540.

[18] Saeedmanesh, M. and N. Geroliminis, Dynamic clustering and propagation of congestion in heterogeneously congested urban traffic networks. *Transportation Research Procedia*, Vol. 23, 2017, pp. 962–979.

[19] Geroliminis, N. and D. M. Levinson, Cordon pricing consistent with the physics of overcrowding. In *Transportation and Traffic Theory 2009: Golden Jubilee*, Springer, 2009, pp. 219–240.

[20] Zheng, N., R. A. Waraich, K. W. Axhausen, and N. Geroliminis, A dynamic cordon pricing scheme combining the Macroscopic Fundamental Diagram and an agent-based traffic model. *Transportation Research Part A: Policy and Practice*, Vol. 46, No. 8, 2012, pp. 1291–1303.

[21] Zheng, N., G. Rérat, and N. Geroliminis, Time-dependent area-based pricing for multimodal systems with heterogeneous users in an agent-based environment. *Transportation Research Part C: Emerging Technologies*, Vol. 62, 2016, pp. 133–148.

[22] Simoni, M., A. Pel, R. Waraich, and S. Hoogendoorn, Marginal cost congestion pricing based on the network fundamental diagram. *Transportation Research Part C: Emerging Technologies*, Vol. 56, 2015, pp. 221–238.

[23] Daganzo, C. F. and L. J. Lehe, Distance-dependent congestion pricing for downtown zones. *Transportation Research Part B*, Vol. 75, 2015, pp. 91–99.

[24] Simoni, M. D., K. M. Kockelman, K. M. Gurumurthy, and J. Bischoff, Congestion pricing in a world of self-driving vehicles: An analysis of different strategies in alternative future scenarios. *Transportation Research Part C: Emerging Technologies*, Vol. 98, 2019, pp. 167–185.

[25] Ben-Akiva, M., H. N. Koutsopoulos, C. Antoniou, and R. Balakrishna, *Fundamentals of Traffic Simulation*, New York, NY, chap. 10-Traffic Simulation with DynaMIT. International Series in Operations Research and Management Science, 2010.

[26] Lu, Y., R. Seshadri, F. Pereira, A. OSullivan, C. Antoniou, and M. Ben-Akiva, Dynamit2.0: Architecture design and preliminary results on real-time data fusion for traffic prediction and crisis management. In *Proceedings of IEEE 18th International Conference on Intelligent Transportation Systems*, Spain, 2015, pp. 2250–2255.

[27] Gu, Z. and M. Saberi, A bi-partitioning approach to congestion pattern recognition in a congested monocentric city. *Transportation Research Part C: Emerging Technologies*, Vol. 109, 2019, pp. 305–320.

[28] Li, Y. and J. Xiao, Traffic peak period detection using traffic index cloud maps. *Physica A: Statistical Mechanics and its Applications*, 2020, p. 124277.

[29] Elhamifar, E. and R. Vidal, Sparse subspace clustering. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2009, pp. 2790–2797.

[30] Lee, H., A. Battle, R. Raina, and A. Y. Ng, Efficient sparse coding algorithms. In *Advances in neural information processing systems*, 2007, pp. 801–808.

[31] Rao, S., R. Tron, R. Vidal, and Y. Ma, Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 10, 2009, pp. 1832–1845.

[32] Bako, L., Identification of switched linear systems via sparse optimization. *Automatica*, Vol. 47, No. 4, 2011, pp. 668–677.

[33] Zhang, K., Y. Chen, and Y. M. Nie, Hunting image: Taxi search strategy recognition using Sparse Subspace Clustering. *Transportation Research Part C: Emerging Technologies*, Vol. 109, 2019, pp. 250–266.

[34] You, C., C.-G. Li, D. P. Robinson, and R. Vidal, Oracle based active set algorithm for scalable elastic net subspace clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3928–3937.

[35] You, C., D. Robinson, and R. Vidal, Scalable sparse subspace clustering by orthogonal matching pursuit. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3918–3927.

[36] Ankerst, M., M. M. Breunig, H.-P. Kriegel, and J. Sander, OPTICS: ordering points to identify the clustering structure. In *ACM Sigmod record*, ACM, 1999, Vol. 28, pp. 49–60.

[37] Campello, R. J., D. Moulavi, and J. Sander, Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining*, Springer, 2013, pp. 160–172.

[38] Elhamifar, E. and R. Vidal, Sparse subspace clustering: Algorithm, theory, and applications. *IEEE transactions on pattern analysis and machine intelligence*, Vol. 35, No. 11, 2013, pp. 2765–2781.

[39] Schubert, E., J. Sander, M. Ester, H. P. Kriegel, and X. Xu, DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems (TODS)*, Vol. 42, No. 3, 2017, p. 19.

[40] Rousseeuw, P. J., Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, Vol. 20, 1987, pp. 53–65.

[41] Davies, D. L. and D. W. Bouldin, A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 1, No. 2, 1979, pp. 224–227.

[42] Lu, L., Y. Xu, C. Antoniou, and M. Ben-Akiva, An enhanced SPSA algorithm for the calibration of Dynamic Traffic Assignment models. *Transportation Research Part C: Emerging Technologies*, Vol. 51, 2015, pp. 149–166.

[43] Pham, D.-S., S. Budhaditya, D. Phung, and S. Venkatesh, Improved subspace clustering via exploitation of spatial constraints. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2012, pp. 550–557.

[44] Hashemi, A. and H. Vikalo, Evolutionary self-expressive models for subspace clustering. *IEEE Journal of Selected Topics in Signal Processing*, Vol. 12, No. 6, 2018, pp. 1534–1546.

[45] Azevedo, C. L., R. Seshadri, S. Gao, B. Atasoy, A. P. Akkinepally, E. Christofa, F. Zhao, J. Trancik, and M. Ben-Akiva, Tripod: sustainable travel incentives with prediction, optimization, and personalization. In *the 97th Annual Meeting of Transportation Research Board*, 2018.