Comparing Dynamic User Equilibrium and Noniterative Stochastic Route Choice in a Simulation-Based Dynamic Traffic Assignment Model: Practical Considerations for Large-Scale Networks

Mudabber Ashfaq,1 Ziyuan Gu,2 S. Travis Waller1, and Meead Saberi1

1Research Centre for Integrated Transport Innovation (rCITI), School of Civil and Environmental Engineering, University of New South Wales (UNSW), Sydney, Australia
2Jiangsu Key Laboratory of Urban ITS, Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, School of Transportation, Southeast University, Nanjing 210096, China

Correspondence should be addressed to Meead Saberi; meead.saberi@unsw.edu.au

Received 11 November 2020; Revised 12 April 2021; Accepted 25 April 2021; Published 5 May 2021

Academic Editor: Jing Dong

Simulation-based dynamic traffic assignment (DTA) models play a vital role in transportation planning and operations. While the widely studied equilibrium-seeking DTA including dynamic user equilibrium (DUE) often provides robust and consistent outcomes, their expensive computational cost for large-scale network applications has been a burden in practice. The noniterative stochastic route choice (SRC) model, as a nonequilibrium seeking DTA model, provides an alternative for specific transportation operations applications that may not require equilibrium results after all (e.g., evacuation and major network disruptions) and thus tend to be computationally less expensive, yet may suffer from inconsistent outcomes. While DUE is a widely accepted approach for many strategic planning applications, SRC has been increasingly used in practice for traffic operations purposes. This paper aims to provide a comparative and quantitative analysis of the two modeling approaches. Specifically, a comparison has been made at two levels: link-level flows and network-level congestion patterns. Results suggest that adaptive driving improves the quality of the SRC solution, but its difference from DUE still remains significant at the link level. Results have practical implications for the application of large-scale simulation-based DTA models for planning and operations purposes.

1. Introduction

Simulation-based dynamic traffic assignment (DTA) models are powerful tools for planning and operations of transportation networks. The two main components of a DTA model are equilibrium-seeking route choice and network loading [1]. The literature also clearly distinguishes between two other commonly used terms known as route search and route choice. Route search refers to the filtering of the most appropriate paths from thousands or millions of possibilities in a large-scale network [2]. Route choice is, on the other hand, the selection of certain routes from an origin to a destination in the presence of alternative routes. Route selection can depend on multiple factors; however, the most widely assumed behavior is that users seek to minimize their respective travel time. Route choice is considered as the core of the traffic assignment problem commonly following Wardrop’s principle based on the assumptions that all drivers are homogeneous and completely rational, and they have perfect information on all available paths and their costs [3–7]. While equilibrium-seeking DTA models often provide robust and consistent outcomes, their expensive computational cost in large-scale networks has made their applications in practice limited and an ongoing challenge.

A noniterative stochastic route choice (SRC) model, as a nonequilibrium seeking DTA model, provides an alternative...
approach for specific large-scale transportation operations applications that may not require equilibrium after all such simulation network-level evacuation or major disruptions. Thus, it tends to be computationally less expensive. While DUE is a widely accepted approach for many strategic planning applications, SRC has been increasingly used in practice for traffic operations optimization because of its lower computational cost without proper acknowledgement of its different nature and use case compared to equilibrium-seeking approaches. SRC models allow drivers to choose a route from a set of feasible paths given the utility function of the path and select a path without strictly considering the flows and travel times on other paths. Consequently, the path overlapping probability increases due to lack of iterative process and can increase travel time on certain paths. This problem is often resolved by giving the option to simulated drivers to alter their selected paths in order to decrease their travel times enroute. This behavior, known as adaptive driving, can increase the average network flow and reduces the formation of gridlock in the simulation [8].

Despite the existence of a vast literature on both simulation-based DUE and SRC, very few studies have provided a quantitative comparative analysis of the two modeling approaches and how selection of each could result in significantly different outcomes in practice. A recent study by [9] highlights that care is needed when using deterministic as opposed to stochastic day-to-day models for transport planning and control purposes. They argued that the two models may provide qualitatively similar outcomes only if the mean of the stochastic model behaves similarly to the deterministic model. In this paper, we aim to provide a comparative quantitative analysis of the solution characteristics of the DUE and SRC models applied to a large-scale network of Melbourne, Australia [10]. We also demonstrate the differences between the two modeling outcomes in a simple network traffic signal optimization problem. The paper also explores the impact of adaptive driving on the quality of the SRC model at both the link and network levels. It is very well known in the academic literature that the two models are fundamentally different and serve different purposes in transportation planning and operations. However, given the increasing use of SRC in practice for different applications, in this paper we aim to demonstrate with numerical evidence how use of each of the two modeling approaches affects the estimated traffic volumes and network wide congestion patterns and provides significantly different outcomes if used carelessly for traffic operations optimization.

2. Background

2.1. Simulation-Based Dynamic Traffic Assignment Models. Traffic assignment models including simulation-based dynamic traffic assignment models are extensively used for a variety of applications such as transportation planning and traffic operations and management [4, 5, 11–13]. In the presence of a reliable origin-destination (OD) demand matrix, a reasonable estimation of network and link traffic states can often be achieved in these models. The two main components of the dynamic traffic assignment models, namely, route choice and network loading, aim to capture vehicles’ movements dynamics in a network. Network loading mainly replicates vehicles’ movements in the network on different paths specified by the route choice model.

The route choice problem is the most critical aspect of the traffic assignment problem as it involves multiple behavioral factors especially in a large-scale network. Trying to capture the reasons behind the selection of paths has always been a challenging task due to the heterogeneity of drivers’ preferences. Even under the assumption of similar behavior of drivers, the model faces many challenges such as route search from a huge set of possible routes in a large-scale network and real-time access and provision of traffic conditions. Vehicles’ path-changing enroute due to nonrecurring events such as incidents, vehicle breakdowns, and special events requires special modeling considerations as most existing models in practice apply pre-defined routes before vehicles’ departure [14]. Different equilibrium-seeking models, such as DUE, dynamic stochastic user equilibrium (DSUE), boundedly rational user equilibrium (BRUE) [15–17], multiclass fuzzy user equilibrium (MFUE) [18], prospect-based user equilibrium (PBUE) [19], and behavioral user equilibrium (BUE) [20], adopt unique behavioral assumptions of drivers [1]. On the other hand, drivers’ route choice using a probabilistic approach considering utilities of different alternative routes results in the so-called stochastic route choice (SRC) models that are not necessarily seeking an equilibrium solution [21].

2.1.1. Dynamic User Equilibrium (DUE). DUE aims to achieve the state of user equilibrium in the network for each time interval in the presence of time-dependent demand. The process is typically carried out in an iterative manner for each origin-destination-time (ODT) pair in the network. At each iteration, the set of shortest paths is updated, and path flows are calculated using, for instance, the method of successive averages (MSA) as one of the most widely used techniques [1, 22]. To measure the quality of the DUE solution, the relative gap metric is commonly used for each ODT that quantifies between the total cost of paths used by vehicles and the total cost of the shortest path used by all the vehicles [13]. A relative gap can be defined and expressed as follows to measure the proximity to an equilibrium solution [23]:

\[
R_{\text{gap}}(n) = \frac{\sum_{t=1}^{T} \sum_{s,t} (r,s) \sum_{p} \sum_{\delta \in (r,s)} \sum_{\theta_{rs}^{p}} f_{rs}^{p}(t) \left[ \tau_{rs}^{p} - \theta_{rs}^{p}(t) \right] \sum_{t=1}^{T} \sum_{s,t} (r,s) \sum_{\delta \in (r,s)} \sum_{\theta_{rs}^{p}} \delta(t) \theta_{rs}^{p}(t) \right]}{\sum_{t=1}^{T} \sum_{s,t} (r,s) \sum_{\delta \in (r,s)} \sum_{\theta_{rs}^{p}} \delta(t) \theta_{rs}^{p}(t) \right]},
\]

where \( f_{rs}^{p}(t) \) represents flow on path \( p \) from OD pair \( r,s \) at time \( t \) and iteration \( n \) whereas \( \tau_{rs}^{p} \) and \( \theta_{rs}^{p}(t) \) quantifies the difference in cost experienced by vehicle following path with cost \( \tau_{rs}^{p} \) in comparison to the shortest path cost \( \theta_{rs}^{p}(t) \). \( \delta(t) \) represents the demand at ODT pair of \( r,s \) and \( \theta_{rs}^{p}(t) \) is the set of all available paths for the ODT, and \( \delta \) is the set of all paths for the OD in network independent of time \( t \). When the relative gap reduces below an acceptable threshold, the network in question is considered to reach the equilibrium state.
MSA is a widely used iterative method that tends to slowly converge to the equilibrium state especially in a large-scale network. Therefore, the whole simulation process can be quite computationally expensive even for mesoscopic simulations. This is a major concern for large-scale DUE applications. Some other methods in the literature are known to converge at an improved rate with a smaller number of iterations like WMSA (weighted-MSA), gradient-based methods, simple travel time responsive method, and alternating direction (AD) methods [24–26]. However, the level of convergence in user equilibrium widely depends on the complexity, size of the network, and the criterion for convergence [25] making it difficult to specify which method is better in general. The MSA algorithm used in our study is the modified version proposed by [27] which is considered to be computationally efficient [26]. The WMSA is a variant of MSA, with the only difference being in the step size calculation, proposed by [24]. Where the MSA algorithm iterates by \( 1/n \) of the demand, the implemented WMSA tends to move by \( 2/(n+1) \) where \( n \) is the number of iterations. This modification in the step size sequence allows the solution in some cases to converge quicker. Analytical DUE methods such as [28] tend to perform faster on small networks but are known to be impractical for large-scale networks [29]. Another concern is the degree of validity of the assumed equilibrium-seeking behavior of drivers in a real-world network, as empirical evidence has suggested the existence of the noncooperative or selfish equilibrium at the macroscopic level only [30].

2.1.2. Noniterative Stochastic Route Choice (SRC). The assumption underlying DUE that every driver has perfect knowledge of the path costs to help choose the best alternative does not necessarily hold in practice as there is a possibility of drivers making imperfect decisions due to inadequate knowledge [3]. Also, unknown factors in relation to preference of some routes that cannot be applied to the general behavior makes the intrinsic nature of traffic extremely variable. This perception error can be incorporated into the SRC model by decomposing the path utility into a deterministic and a random term. The probit model assuming a normally distributed random term and the logit model assuming a Gumbel distribution are two widely used methods to solve the SRC problem [7]. The probit model is comparatively computationally expensive due to the absence of a closed-form formula [31]. On the other hand, the logit model has a closed-form structure for easy calculation while having two main drawbacks, one being the issue with highly overlapped path set due to its inability of finding path correlation and the other being the scaling problem of not identifying the heterogeneity in drivers’ perception errors. To overcome the path overlapping problem, modified versions of the multinomial logit model have been introduced, such as the C-logit (which will be used in this paper), path size logit (PSL), and implicit availability/perception (IAP), by introducing a new utility function that tends to choose the spatially separated routes. Scaling problem can be addressed by introducing a scaling factor \( \theta \) [32]. C-logit presents probability \( P_k \) for selecting the paths using

\[
P_k = \frac{e^{\theta (V_k - \text{CF}_k)}}{\sum_{k \in K} e^{\theta (V_k - \text{CF}_k)}},
\]

where \( V_k \) and \( V_l \) are the perceived utility for selected path \( k \) and alternative paths \( l \), respectively and \( \text{CF}_k \) denotes the “commonality factor” of the path \( k \) that describes the degree of overlapping for current path to all the alternative paths chosen. This \( \text{CF}_k \) is calculated using the cost of links \( L_{lk} \) common to paths \( k \) and \( l \), \( L_k \), \( L_l \) the link costs of path \( l \) and \( k \), respectively (expressed in hours), as follows

\[
\text{CF}_k = \beta \cdot \ln \left( \sum_{l \in K} \left[ \frac{L_{lk}}{L_k^{1/2} L_l^{1/2}} \right]^\gamma \right),
\]

Here, \( \beta \) and \( \gamma \) are the two opposite parameters where \( \beta \) displays the importance of utility \( V_k \) and \( \gamma \) with lesser influence plays an opposite role to \( \beta \) and is usually in the range of \([0, 2]\) [23]. SRC model characteristics are applicable to both iterative (e.g., DSUE, DSUE-SRDTC [33–35]) and noniterative techniques. Noniterative SRC is used for all the analysis here as the focus of the paper is on the comparative analysis of iterative user-equilibrium versus noniterative route choice models. Going forward, noniterative SRC would be referred to as SRC.

2.1.3. Adaptive Driving. Compared with the DUE, SRC is much less computationally expensive because it does not require multiple iterations to reach the equilibrium state. However, without seeking equilibrium and with the introduction of perception errors, the quality of the SRC solution cannot be guaranteed and the network might evolve into unexpected gridlock. It has been shown that the network flow can be improved by considering adaptive driving for a proportion of drivers [8, 36]. Here, adaptive driving refers to the ability to alter the assigned path at each time step by choosing the new shortest path according to the prevailing traffic conditions. Our goal in this paper is to study how adaptive driving affects the estimated network flows and thus the quality of the SRC solution in comparison to the DUE solution as it may have significant practical implications if used for traffic operations optimization.

3. Model Description

In this paper, we perform a mesoscopic simulation-based DTA on a subnetwork of Melbourne metropolitan area (Figure 1) covering the city center using an existing calibrated and validated model in AIMSUN [10]. The network consists of 4375 links and 1977 nodes. Links have multiple attributes including the number of lanes, capacity, and the free-flow speed. The geometric configuration of the network is adapted from the Victoria Integrated Transport Model (VITM) including 492 traffic zones. The static OD matrix included 2,173,306 vehicles for a four-hour morning peak period (6 AM–10 AM). A time-dependent OD matrix is then estimated based on the initial static OD matrix and observed counts from Sydney Coordinated Adaptive Traffic System (SCATS) loop detector data from over 1500 actuated signals.
flow-density FD is as follows: level (see for [23] further details). The resulting triangular simplification results in a link fundamental diagram (FD) car following model as the supply model [26, 40–42]. (Y=he validation details, please see [10]. AIMSUN uses a simplified models were calibrated and validated. For calibration and model, please see [10, 37–39]. (Y=he supply and demand methods to calculate DUE, namely, MSA (method of equation represents the undersaturated regime of the traffic where \( V \) and \( k \) denote the speed limit and the density level of a given link. The first term in the right-hand side of the equation represents the undersaturated regime of the traffic while the second term represents the oversaturated regime where \( L \) and \( R \) are the effective length (sum of the vehicle length and the clearance distance) and the reaction time in a link, respectively. The link flow is zero either for \( k = 0 \) or for \( k_{jam} = 1/L \) and is maximum at the intersection of two regimes. Reaction time, here, is the production of the global reaction time and the reaction time factor of the link. The local reaction time factor brings more flexibility in the calibration process.

Real observed traffic count data obtained from loop detectors and SCATS detectors at the signalized intersections is used for comparative analysis and validation of simulated solutions. We use a relative gap of less than 5% to ensure the DUE solution is properly converged based on previous studies in the literature [27, 43–45]. The literature suggests that the relative gap has a high dependency on the size of the network. For a small size network, a very low relative gap close to zero is possible. However, for medium to large-scale networks, approaching a zero relative gap is very difficult. Hence, most practitioners working with large-scale networks tend to choose appropriate relative gap of around 5–10%. AIMSUN provides three different methods to calculate DUE, namely, MSA (method of successive averages), WMSA (weighted-MSA), and gradient-based. Each method has a different approach to converge to the appropriate relative gap (5%); hence their computation performance varies accordingly (see Figure 2). The MSA and WMSA processing times for each iteration are very similar as the underlying basic mechanism is the same with only the difference being in step size calculation, whereas the gradient-based method is more costly per iteration as compared to the other two methods. The second and important factor in the processing time is the number of iterations that is required to converge to the desired solution, in which WMSA performs the best with only 36 iterations as opposed to its competitors, MSA and gradient based with 120 and 121 iterations. Despite improved convergence of WMSA, DUE still took 3214.45 seconds which suggests that the noniterative SRC still outperforms DUE. All the computation metrics are recorded while using the same platform with specifications as follows: processor: Intel (R) Core (TM) i9-9900 CPU @ 3.10 GHz 3.10 GHz, RAM: 64.0 GB.

For SRC, we use the C-logit model and consider different proportions of drivers with adaptive driving. C-logit is used with the default values of scale factor as 1 and commonality factor parameters, \( \beta \) and \( \gamma \), as 0.15 and 1, respectively. These parameters are fixed for all the OD pairs in the matrix throughout the simulation. The model takes into consideration 5 paths initially and, from those, chooses up to three most suitable paths by calculating the shortest path(s) from each OD pair using initial path travel times. At each simulation time interval, paths are recalculated using experienced average link travel times. Adaptive drivers also recalculate their shortest path at each time interval to check for the possible alternative routes for dynamic rerouting.

The distribution of the number of paths used by vehicles in the DUE (see Figure 3) suggests considering three alternative paths in the path assignment of the SRC should be a reasonable assumption. To further strengthen our analysis, we have conducted analysis of running noniterative SRC with three, five, and seven paths with same the number of adaptive drivers and compared the results against real data (see Figure 4). The results using mean absolute error (MAE) suggest that the noniterative SRC achieves more accurate outcomes when three paths are used. When root mean square error (RMSE) and root mean square normalized (RMSN) errors are used, no significant difference is observed in the model outcomes when 3, 5, or 7 paths are used.

4. Results and Discussion

The SRC simulations are performed with different adaptive driving proportions ranging between 0% and 100% with sampling interval of 10%. To maintain consistency, each simulation is run using five different random seed numbers. Comparison has been made at two levels: link-level flows and network-level congestion patterns. At the link level, flows are compared for each link in the network, whereas at the network level congestion patterns across the entire network are analyzed and compared.
4.1. Link-Level Comparison. The average flow of every link in the network for time window (6–10 AM) is calculated and compared under the DUE and SRC scenarios using the mean absolute error (MAE), root mean square error (RMSE), and root mean square normalized (RMSN) which can be calculated as

\[ MAE = \frac{1}{n} \sum_{j(t)=1}^{n} |S_{j(t)} - O_{j(t)}|, \]

\[ RMSE = \sqrt{\frac{\sum_{j(t)=1}^{n} (S_{j(t)} - O_{j(t)})^2}{n}}, \]

\[ RMSN = \frac{\sum_{j(t)=1}^{n} (S_{j(t)} - O_{j(t)})^2}{\sum_{j(t)=1}^{n} O_{j(t)}} = \frac{RMSE}{1/n \sum_{j(t)=1}^{n} O_{j(t)}}, \]

where \( n \) is the total number of links times the time intervals for observation and \( S_{j(t)} \) and \( O_{j(t)} \) represent flow or volume at link \( j \) at time interval \( t \) for simulated (SRC/DUE) and observed (DUE/Real data) values, respectively. Results suggest a decrease in MAE, RMSE, and RMSN (Figure 5) as the proportion of adaptive drivers in the network increases. The stochasticity of link flows as a result of different random

Figure 2: Computational time performance and the relative gap of DUE methods (MSA, WMSA, and gradient-based) with the number iterations represented on the x-axis of both plots, whereas the left plot displays maximum relative gap at each iteration and the right plot represents the progression of processing time in seconds with respect to the number of iterations.

Figure 3: Distribution of the number of DUE paths used by vehicles. X-axis represents unique number of alternative paths used by the DUE in the simulation for different ODT pairs. Y-axis represents the number of ODT pairs. The majority of the simulated vehicles follow 3 or fewer paths.
seed numbers also decreases in the presence of high adaptiveness, suggesting that the network is more robust with more adaptive drivers. However, high variance during peak traffic hours (9–10 AM) suggests time dependency of these solutions (DUE and SRC) (see Figure 6). Increased proportion of adaptive drivers reduces the gap between the two but not significantly.

The SRC with 40% to 80% adaptive drivers produces the closest result to the DUE, in terms of individual link flows, as indicated by MAE, RMSE, and RMSN. Interestingly, a number of links in the network experience flows larger than 1000 veh/hr in the DUE, which, however, exhibit a rather small flows in the SRC with very few adaptive drivers. This is because when the adaptive driving proportion is low, most drivers simply follow their initially assigned paths even if they encounter severe congestion enroute. With a higher proportion of adaptive drivers, flows are improved as also observed and reported in [8]. A comparison between the observed link volumes and the simulated volumes of both DUE and SRC is performed to further validate the results.

![Figure 4](image-url)

**Figure 4:** A comparison of the accuracy of the SRC solution when 3, 5, and 7 paths are used in the assignment with a fixed 30% adaptive driving percentage and five replications against real-world link volumes: (a) MAE, (b) RMSE, and (c) RMSN.
Figure 5: (a) MAE, (b) RMSE, and (c) RMSN for individual link flows from the SRC model with different adaptive driving compared to the DUE. X-axis shows percentage of adaptive drivers in the network for SRC and y-axis is the error in vehicles/hr.
Traffic congestion can be measured using a variety of methods. In this section, we use link speeds to measure the congestion level by comparing the observed speed with the maximum speed (i.e., the speed limit) [46, 47]. For each link in the network, we obtain the simulated speed $v_i(t)$ for each time interval $t$ and the maximum speed $v_i^{\text{max}}$. To quantify congestion as a function of speed, we define

$$r_i(t) = \frac{v_i(t)}{v_i^{\text{max}}},$$

where $r_i(t)$ is the ratio of the link speed $v_i(t)$ over the maximum speed limit $v_i^{\text{max}}$. We then classify links into congested links with $c_i = 1$ and uncongested links with $c_i = 0$ using a threshold $\rho$ [47]:

$$c_i(t) = \begin{cases} 
1, & r_i(t) < \rho, \\
0, & r_i(t) \geq \rho,
\end{cases}$$

where $\rho$ represents a user-defined congestion level. Here, we consider $\rho = 0.2$ and $\rho = 0.5$ (i.e., 20% and 50% of the maximum speed) as two congestion levels. With $\rho = 0.2$, we are capturing links that are severely congested. On the other hand, with $\rho = 0.5$ (Figure 8), any link with a speed less than half of the maximum speed is considered as congested. Consequently, $C(t)$ is defined as the fraction of congested links in the network:

$$C(t) = \frac{\sum_{i=1}^{N} c_i(t)}{N},$$

where $N$ is the total number of links in the network and $C(t)$ curve represents the overall congestion pattern in the network at any time $t$ which evolves throughout the simulation.

The main difference between the DUE and SRC with low percentage of adaptive drivers appears at the onset of congestion as well as in the recovery period. With 60% and 80% of adaptive drivers, the network manages to recover more rapidly similar to DUE as compared to the other adaptiveness levels. However, even with 80% of drivers being adaptive, the SRC model still produces slightly different congestion patterns in the network as compared to the DUE. Traffic congestion obtained from DUE solution tends to peak earlier as compared to SRC and recovers more rapidly as well.

Results also suggest that propagation and recovery rates of congested links, measured as the gradient of the curves shown in Figure 9, in the DUE model are smaller as compared to the SRC model with low percentage of adaptive drivers. Adaptive driving does not necessarily eliminate the propagation of congestion in the network, but it helps the network to recover from the gridlock state more rapidly than without adaptive drivers. Figure 10 shows and compares the spatial distribution of the congested links in the network under both the DUE and SRC scenarios. As expected, a higher dissimilarity is observed in the presence of low adaptiveness in SRC, and vice versa.

4.2. Network-Level Comparison. Traffic congestion can be measured using a variety of methods. In this section, we use link speeds to measure the congestion level by comparing the observed speed with the maximum speed (i.e., the speed limit) [46, 47]. For each link in the network, we obtain the simulated speed $v_i(t)$ for each time interval $t$ and the maximum speed $v_i^{\text{max}}$. To quantify congestion as a function of speed, we define

$$r_i(t) = \frac{v_i(t)}{v_i^{\text{max}}},$$

where $r_i(t)$ is the ratio of the link speed $v_i(t)$ over the maximum speed limit $v_i^{\text{max}}$. We then classify links into congested links with $c_i = 1$ and uncongested links with $c_i = 0$ using a threshold $\rho$ [47]:

$$c_i(t) = \begin{cases} 
1, & r_i(t) < \rho, \\
0, & r_i(t) \geq \rho,
\end{cases}$$

where $\rho$ represents a user-defined congestion level. Here, we consider $\rho = 0.2$ and $\rho = 0.5$ (i.e., 20% and 50% of the maximum speed) as two congestion levels. With $\rho = 0.2$, we are capturing links that are severely congested. On the other hand, with $\rho = 0.5$ (Figure 8), any link with a speed less than half of the maximum speed is considered as congested. Consequently, $C(t)$ is defined as the fraction of congested links in the network:

$$C(t) = \frac{\sum_{i=1}^{N} c_i(t)}{N},$$

where $N$ is the total number of links in the network and $C(t)$ curve represents the overall congestion pattern in the network at any time $t$ which evolves throughout the simulation.

The main difference between the DUE and SRC with low percentage of adaptive drivers appears at the onset of congestion as well as in the recovery period. With 60% and 80% of adaptive drivers, the network manages to recover more rapidly similar to DUE as compared to the other adaptiveness levels. However, even with 80% of drivers being adaptive, the SRC model still produces slightly different congestion patterns in the network as compared to the DUE. Traffic congestion obtained from DUE solution tends to peak earlier as compared to SRC and recovers more rapidly as well.

Results also suggest that propagation and recovery rates of congested links, measured as the gradient of the curves shown in Figure 9, in the DUE model are smaller as compared to the SRC model with low percentage of adaptive drivers. Adaptive driving does not necessarily eliminate the propagation of congestion in the network, but it helps the network to recover from the gridlock state more rapidly than without adaptive drivers. Figure 10 shows and compares the spatial distribution of the congested links in the network under both the DUE and SRC scenarios. As expected, a higher dissimilarity is observed in the presence of low adaptiveness in SRC, and vice versa.

5. Application in Network Traffic Signal Optimization

Network traffic signal optimization is a well-studied problem in the literature [48–51]. Among the more commonly used methods to solve the problem are heuristic and metaheuristic approaches including genetic algorithms, neural networks, feedback control, and simulation-based optimization [50, 52–54]. We are specifically interested in formulating the problem as a bi-level program similar to what has been proposed in [49] where the upper-level problem is signal timing optimization and the lower-level problem is the DTA (equilibrium seeking DUE or nonequilibrium seeking SRC). In fact, the literature on network traffic signal optimization can be grouped into two, in which the first group of studies applies an equilibrium route choice model to solve the signal control problem [55–58] while the second group applies a nonequilibrium route choice model [59–65]. Table 1 provides a summary of the relevant studies in the literature classified based on the network size and utilized traffic assignment model. The aim here is not to discuss or advocate whether, for network traffic signal optimization, the use of an equilibrium route choice is a more reasonable and realistic approach or a nonequilibrium approach.
Rather, we only aim to provide a comparative analysis to shed light on how use of each of the models affects the solution.

Here, we formulate a simple network traffic signal optimization problem in which we consider the signal cycle length as the only decision variable which is assumed to be homogeneous across the selected traffic signals in the study network. To reduce the computational burden of the optimization in this case study, we select 19 traffic signals in the CBD area only to be included in the optimization problem (see Figure 11) while the remainder of the traffic signals in the network are actuated with a pre-determined and fixed cycle length. Note that the simulations are run on the complete network of Melbourne as shown earlier in Figure 1 while the optimization and traffic measurements are only conducted for the CBD area. The goal of this exercise is to analyze the impact of DUE or SRC selection on the developed congestion patterns and the estimated optimal solution.

$$\min_\omega \frac{\sum_{t=1}^r c(t)}{N},$$

subject to

$$c(t) = DTA(\omega),$$

\begin{figure}[h]
\centering
(a) \hspace{1cm} (b) \hspace{1cm} (c) \hspace{1cm} (d)
\caption{Scatterplot of individual link flows: comparing the SRC (with random seed 28,678) with different levels of driving adaptiveness (a) 0%, (b) 20%, (c) 60%, and (d) 80% in comparison to the DUE link flows. X-axis displays link flows estimated by DUE in veh/hr and y-axis represents the link flows estimated by the SRC model in veh/hr.}
\end{figure}
Figure 8: (a) MAE, (b) RMSE, and (c) RMSN for individual link traffic counts from the DUE and SRC with different percentage of adaptive drivers and five replications with different random seeds for each adaptiveness level against observed link volume data. X-axis shows the percentage of adaptive drivers in the network for the SRC and y-axis represents the error as volume (vehicles). DUE error remains constant and is shown as the straight dashed line as no adaptive driving is considered in DUE.
Figure 9: Congestion propagation and dissipation in the network from 06:00 AM–10:00 AM represented by the fraction of congested links $C(t)$ with (a) congestion threshold $\rho = 0.2$, (b) congestion threshold $\rho = 0.5$.

Figure 10: Spatial visualization of congestion patterns in the network at 08:00 AM when $\rho = 0.5$: comparison of the SRC and DUE model outcomes with varying levels of adaptive drivers (a) 0%, (b) 20%, (c) 60%, and (d) 80%. Congested (uncongested) refers to the case in which a link in the network is estimated to be congested (uncongested) in both the DUE and SRC models, while DUE congested and SRC congested refer to the cases in which the link is estimated to be congested in DUE or SRC alone, respectively.
where $\omega$ represents signal cycle length, $\tau$ represents the study time period, and $N$ is number of time intervals between 1 and $\tau$. The upper-level problem (equation (6)) minimizes the average number of congested links in the network over the study time period, while the lower-level problem (equation (7)) represents the simulation-based DTA (SRC or DUE model) with variable cycle length $\omega$ for the selected traffic signals.

To solve the signal optimization problem, involving a simulation-based DTA model, a variety of simulation-based optimization (SBO) methods could be applied. Without loss of generality, SBO refers to the optimization of an objective function subject to certain constraints, both of which are evaluated through stochastic computer simulations. In transportation network design and analysis, SBO is frequently applied in the presence of “black-box” simulation [83–86]. For an overview of different SBO methods, refer to [87]. In this paper, we adopted a naïve response surface method where a parametric model mimicking the simulation input-output mapping was constructed based on results of selected sample points. Specifically, we randomly selected 7 sample points representing different cycle lengths between 60 and 150 sec, for each of which simulations were performed and the objective function was evaluated. Note that SRC was run for five replications with 30% adaptive drivers. Admittedly such a naïve approach lacks a mathematical rigor without accounting for any infill sample points as one would see in the literature [88], but it suffices to highlight the difference between the two modeling approaches as the key emphasis in this paper.

Table 1: A summary of literature on simulation-based traffic signal optimization studies.

<table>
<thead>
<tr>
<th>Traffic assignment</th>
<th>Network size</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deterministic user equilibrium (DUE)</td>
<td>Local (&lt;5 signals or &lt;5 available routes)</td>
<td>Very few studies are found carrying out signal optimization using DUE in relatively small networks with few routing options including [66] that used DUE to calculate the reserve capacity of the network. Meanwhile, [67, 68] implemented DUE on a local network with isolated signal or few signals as well as with a real network of multiple signalized intersections. DUE is the popular choice of traffic assignment for signal optimization in relatively large networks [67–70]. These studies yield efficient techniques for signal optimization based on DUE traffic assignment with some providing a comparison between DUE with SO, concluding significant advantage of DUE for adaptive signals.</td>
</tr>
<tr>
<td></td>
<td>Global (&gt;5 signals or &gt;5 available routes)</td>
<td>SUE is more prominent in localized signal optimization [57, 68, 71, 72]. The study reported in [58] has provided a comparison of the performance of deterministic and stochastic user equilibrium techniques over different networks and observed improved results for SUE in local networks.</td>
</tr>
<tr>
<td>Stochastic user equilibrium (SUE)</td>
<td>Local (&lt;5 signals or &lt;5 available routes)</td>
<td>Very few studies are found in the literature using SUE in relatively large networks including [72] that has recommended the use of DUE on global traffic signal optimization.</td>
</tr>
<tr>
<td></td>
<td>Global (&gt;5 signals or &gt;5 available routes)</td>
<td>References [67, 73, 74] used SO in both large and small networks. Reference [67] argued that DUE is advantageous over SO in designing adaptive signals in a network.</td>
</tr>
<tr>
<td>System optimal (SO)</td>
<td>Local/global</td>
<td>Use of SRC mostly prevails in the studies on a single freeway or highway with few alternate routes available [65, 75–79].</td>
</tr>
<tr>
<td>Stochastic route choice (SRC)</td>
<td>Local (&lt;5 signals or &lt;5 available routes)</td>
<td>A few studies are found in the literature that used relatively large networks and SRC [80–82]. In [82], the selection of route choice model was arbitrary as the main focus of the study on the optimization part of the problem rather than the simulation.</td>
</tr>
</tbody>
</table>

![Melbourne CBD subnetwork. The red boundary shows the extent of the area considered in traffic signal optimization problem with red points representing the selected 19 traffic signals.](image)

Figure 12 summarizes the analysis results for both DUE and SRC models. The formulated objective function value, namely, the average fraction of congested links in the network, varies for different cycle length values. However, the variations for both SRC and DUE models show a similar pattern. When SRC is used, the optimal cycle length is estimated to be 126 seconds while for the DUE the optimal cycle length is 93 seconds. The observed and quantified difference demonstrates how selection of the route choice model affects the optimization solution. Therefore,
considerable care is needed when selecting the route choice model and comparing the results of DUE and SRC in traffic planning and operations applications.

6. Conclusions

Finding dynamic equilibrium is computationally expensive in large-scale networks, but it often provides robust and consistent model outcomes given the equilibrium-seeking behavior. The SRC model introduces a random perception error in the path assignment without seeking the equilibrium state of the network, thereby resulting in a more probabilistic DTA model outcomes. Given the increasing use of SRC model in practice and its lower computational costs, in this paper we aimed to provide a quantitative comparison between the two modeling approaches when applied to a large-scale network and used in traffic operations optimization. Due to the absence of the equilibrium-seeking behavior, the SRC with low adaptive driver’s population tends to produce more congestion in the network and, hence, lower link flows.

Figure 12: Network traffic signal optimization results. (a) Mean fraction of congested links for different signal cycle length with congestion threshold $\rho = 0.5$. Dashed curves represent the fitted surface function. (b) Evolution of $C(t)$ over time for DUE and SRC when $\omega = 108$ sec. (c) Boxplots representing variations in the $C(t)$ over 5 SRC replications when $\omega = 108$ sec.
However, by introducing adaptive driving, the quality of the SRC solution in terms of overall congestion rate is shown to be improved as measured by the increased link flows as well as the faster recovery from congestion. With increasing the proportion of adaptive drivers in the network, the SRC model performs more closely to both DUE and real-world observations in terms of link flows and congestion patterns. However, if the whole population of drivers behave adaptively, the estimated link flows drift away from DUE. These results have practical implications for the application of large-scale DTA models for planning and operations purposes. A simplified case study on network traffic signal optimization was performed and revealed the practical implications of use of each of the route choice models. While DUE is widely used in academic literature in many applications including network design problems, the increasing use of SRC in practice and ongoing discussions among practitioners on the necessity of equilibrium (or lack thereof) in large-scale modeling studies have motivated us to provide evidence on the significant practical differences that each modeling approach provides. We hope the provided discussions and quantitative comparisons in this paper help practitioners with selection of the most appropriate route choice model for their specific purposes.

Data Availability

The simulation-based traffic model is publicly available on GitHub through https://github.com/meeadsaberi/dynamel. The simulation model can be used to replicate the data used in this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References


