Energy-Efficient Urban Traffic Management: A Microscopic Simulation-Based Approach

Carolina Osorio, Kanchana Nanduri
Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 {osorioc@mit.edu, kanchana@mit.edu}

Microscopic urban traffic simulators embed the most detailed traveler behavior and network supply models. These simulators represent individual vehicles and can therefore account for vehicle-specific technologies. They can be coupled with instantaneous fuel consumption models to yield detailed network-wide fuel consumption estimates. Nonetheless, there is currently a lack of computationally efficient optimization techniques that enable the use of these complex integrated models to design sustainable transportation strategies.

This paper proposes a methodology that combines a stochastic microscopic traffic simulation model with an instantaneous vehicular fuel consumption model. The combined models are embedded within a simulation-based optimization algorithm and used to address a signal control problem that accounts for both travel times and fuel consumption. The proposed technique couples detailed, stochastic, and computationally inefficient models, yet is an efficient optimization technique. Efficiency is achieved by combining simulated observations with analytical approximations of both travel time and fuel consumption.

This methodology is applied to a network in the Swiss city of Lausanne. Within a tight computational budget, the proposed method identifies signal plans with improved travel time and fuel consumption metrics. It outperforms traditional methodologies, which use only simulated information or only analytical information. The case study illustrates the added value of combining simulated and analytical information when performance metrics with high variance, such as fuel consumption, are used. This method enables the use of disaggregate instantaneous vehicle-specific information to inform and improve traffic operations at the network-scale.

Keywords: simulation-based optimization; traffic signal control; microscopic simulation; energy-efficiency

History: Received: January 2013; revision received: November 2013; accepted: March 2014. Published online in Articles in Advance February 27, 2015.

1. Introduction

The International Energy Agency (IEA) has estimated that over 50% of oil use worldwide is from transport and that three-quarters of the energy used in the transport sector is consumed on the roads (IEA 2012). The IEA also projects that without strong new measures, road transport fuel use will double between 2010 and 2050. The fuel consumed every day on our roads not only contributes to the depletion of a valuable natural resource but the linear relationship between fuel consumption and CO₂ means that urban traffic plays a role in global warming. Thus, there is a need to understand how the use of existing urban transportation infrastructure can be enhanced to reduce energy consumption. Signal control remains a viable solution in this regard: The re-timing effort involved is low cost and the environmental benefits can be realized within a short time span.

Over the past decade state-of-the-art traffic, fuel consumption, and emissions simulators have been developed independently, coupled and extensively used to evaluate the impacts of various transportation projects on traffic, fuel consumption, and the environment. Nonetheless, there is currently a lack of computationally efficient optimization techniques that enable the use of these complex integrated models to design sustainable transportation strategies.

This paper proposes a methodology that combines detailed traffic and fuel consumption models to design traffic signal control strategies that improve traditional traffic metrics (e.g., average travel times) while reducing total fuel consumption. The main challenge is to use the most detailed yet inefficient models, while simultaneously deriving a computationally efficient methodology.

This paper focuses on the development of computationally efficient simulation-based optimization (SO) techniques, which can yield solutions within a tight simulation budget (defined as a maximum number of simulation runs or run time). Such techniques respond to the needs of practitioners by allowing them to address real problems in a practical manner. In the remainder of this section, we present a
brief description of the main types of traffic and fuel consumption models.

Traffic simulators can be primarily classified as macroscopic, mesoscopic, or microscopic according to their modeling scale. For reviews see Barceló (2010); Boxill and Yu (2000); Algers et al. (1997). Macroscopic simulators use models that describe the progression of traffic along links as a function of average link speed, flow, and density. Macroscopic models are therefore flow-based and provide an aggregate representation of traffic. Vehicles along a link are collectively represented. This leads to models that have few parameters to calibrate, are computationally efficient, and well suited to study large-scale networks. These advantages come at the cost of a nondetailed (aggregate) description of traffic.

Microscopic models represent individual travelers and/or vehicles. Driver-specific characteristics (e.g., socio-economic attributes) and vehicle-specific characteristics (e.g., vehicle type, vehicle technology) can be accounted for. Microscopic models provide a highly detailed representation of network flows and use disaggregate behavioral models (e.g., departure-time choice, mode choice, lane changing, car-following) to describe the reaction of individual drivers towards network components, traffic conditions, and adjacent drivers. Such detail leads to data- and computationally-intensive models. Mesoscopic traffic simulators lie on the spectrum between microscopic and macroscopic simulators.

Similarly, fuel consumption models can primarily be categorized as macroscopic or microscopic. Macroscopic models estimate fuel consumption based on average speed/acceleration. However, fuel consumption depends on the spatial-temporal variations of both speed and acceleration. For example, past work has shown how common average speed/acceleration profiles may arise from different instantaneous profiles, and therefore lead to different fuel consumption levels (Rakha et al. 2000). Microscopic models rely on instantaneous (e.g., second-by-second) speeds and accelerations of individual vehicles.

Integrated microscopic traffic and microscopic fuel consumption models are particularly suitable when studying the impact of network changes on fuel consumption; they account for detailed and complex vehicle-to-vehicle and vehicle-to-supply interactions. The detail of such models comes with an increased computational evaluation cost, as well as a greater challenge to embed them within an optimization framework.

The next section reviews related work. Section 3 presents the proposed SO methodology. The method is then applied to address various traffic signal control problems that are formulated in §3.1. Results from a Lausanne city case study are analyzed (§4), followed by concluding remarks (§5).

2. Literature Review

The interactions between traffic operations and fuel consumption have been extensively investigated over the past three decades (since the seminal work of Robertson 1983). In this section, we review recent work that has coupled traffic simulators with fuel consumption simulators to address urban transportation problems. For early references, we refer the reader to the review presented in Liao and Machemehl (1998).

Ikeda, Kawashim, and Oda (1999) modify the usual objective function (called the performance index) of the macroscopic TRANSYT (Robertson 1969) signal control technique to explicitly account for fuel consumption. A case study considering a 10-link linear network in Yokohama, Japan, is carried out. Li et al. (2004) develop a macroscopic fuel consumption and emissions model that is combined with a macroscopic traffic model. The combined model is used to evaluate the impact of a set of predetermined cycle length values of a signal plan for one intersection in Nanjing, China.

Zegeye et al. (2010) couple the macroscopic traffic model METANET (Messmer and Papageorgiou 1990) with a microscopic fuel consumption and emissions model VT-Micro (Rakha, Ahn, and Trani 2004). As a macroscopic model, METANET provides average link speeds and accelerations. These are plugged into VT-Micro (at every simulation step) as if they were instantaneous vehicle speeds. The combined models are embedded within a dynamic control framework. They consider a dynamic speed limit problem along with an objective function that combines three metrics: total travel time, total fuel consumption, and total CO2 emissions. Their case study considers a hypothetical 12 km 2-lane freeway.

Cappiello (2002) couples a mesoscopic traffic model (Bottom 2000) with a microscopic fuel consumption and emissions model. They consider a hypothetical 14-link network, and evaluate the travel time and fuel consumption performance of a set of predetermined variable message sign strategies. Williams and Yu (2001) use the macroscopic traffic model DYNAMIC (Yu 1994) along with a macroscopic fuel consumption model. They consider two hypothetical networks with, respectively, one and two signalized intersection(s) and evaluate the fuel consumption performance of several predetermined cycle lengths.

To provide a more accurate representation of the interaction between vehicular fuel consumption and supply changes, microscopic traffic and microscopic fuel consumption models have been coupled. Stathopoulos and Noland (2003) use the microscopic models VISSIM (PTV 2008) and CMEM (Scora and Barth 2006) albeit not in an optimization context. The coupled models are used to evaluate the travel time, fuel consumption, and emission impacts of a set of predetermined scenarios for two hypothetical transportation projects (capacity...
expansion of an arterial bottleneck and synchronization of traffic signals) on a hypothetical linear network with three signalized intersections.

Rakha, Ahn, and Trani (2004) develop a microscopic fuel/emissions model known as VT-Micro. This model is used in Rakha et al. (2000) along with the microscopic simulator INTEGRATION (Van Aerde 1999) to evaluate the performance of predetermined signal plans considering a linear network with four links and simple demand profiles.

To our knowledge, the only work that has integrated microscopic traffic and microscopic fuel consumption models to perform optimization is that of Stevanovic et al. (2009). They integrate VISSIM (PTV 2008) with CMEM (Scora and Barth 2006), and embed the coupled models within the signal optimization tool VISGAOST (Stevanovic et al. 2008). Their case study considers a network of two arterials with 14 signalized intersections in Park City, Utah. They investigate various formulations of the signal control problem (e.g., objective functions consider throughput, stops, delay, fuel consumption, or CO₂ emissions). Their problems have over 100 signal control variables. For each problem they run a total of 60,000 simulation runs (12,000 signal plans evaluated across five simulation replications each). This is a flexible approach, yet it is not designed to address problems under tight computational budgets.

To summarize, traffic models coupled with fuel consumption models have been applied at macroscopic, mesoscopic, and microscopic scales in traffic management. Microscopic simulators incorporate disaggregate behavioral models, making them ideal for scenario-based analysis and accurate estimation of network performance measures such as fuel consumption and travel time. However, the use of microscopic simulators coupled with detailed fuel consumption models has been mainly limited to evaluating the effect of a set of predetermined alternatives. This can primarily be attributed to the challenges faced when integrating microscopic simulators in an optimization framework. The outputs from the simulator are stochastic and nonlinear with possibilities for numerous local minima. Also, a large number of simulation replications are needed to derive accurate estimates of the objective function, thus driving up computational costs.

This paper proposes a simulation-based optimization technique that uses integrated microscopic traffic and fuel consumption models to address signal control problems. The SO technique can identify signal plans with improved performance within a few simulation runs, i.e., it is efficient. For instance, for the Lausanne city case study considered in §4, which has nine signalized intersections and over 50 signal control variables, the technique identifies signal plans with improved performance within a total of 150 simulation runs. The signal plans derived are shown to reduce both average travel time and total fuel consumption during the evening peak period.

Additionally, this work illustrates how highly variable outputs from traffic simulators, such as fuel consumption, can be efficiently used for optimization. This method enables the use of disaggregate instantaneous vehicle-specific information to inform and improve traffic operations at the network scale.

3. Methodology

3.1. Fuel-Efficient Signal Control Problems

In this section, we formulate the two different traffic signal control problems that are addressed in this paper. For a review of traffic signal control terminology and formulations, we refer the reader to Osorio (2010, Appendix A). In this paper, we consider a fixed-time, also called time of day or pretimed, control strategy. These strategies use historical traffic patterns to derive a fixed signal plan for a given time period. The signal control problem is solved offline. The signal plans of multiple intersections are determined jointly. The decision variables are the green splits (i.e., green times) of phases of the different intersections. All other traditional control variables (e.g., cycle times, offsets, stage structure) are assumed fixed.

To formulate this problem we introduce the following notation:

- \( c_i \): cycle time of intersection \( i \);
- \( d_i \): fixed cycle time of intersection \( i \);
- \( x(j) \): green split of phase \( j \);
- \( x_L \): vector of minimal green splits;
- \( J \): set of intersection indices; and
- \( P_j(i) \): set of phase indices of intersection \( i \).

The problem is formulated as follows:

\[
\min_x f(x; p) \equiv E[F(x; p)]
\]

subject to

\[
\sum_{j \in P_j(i)} x(j) = \frac{c_i - d_i}{c_i}, \quad \forall i \in J,
\]

\[
x \geq x_L,
\]

where the continuous and deterministic decision vector \( x \) consists of the green splits for each phase. Constraints (2) ensure that for a given intersection the available cycle time is distributed among all phases. Green splits have lower bounds (Equation (3)), which are set to four seconds in this work (following the Swiss transportation norms (VSS 1992)).

The complexity of this problem comes from the simulation-based objective function \( f \), which is the expected value of a stochastic network performance measure, \( F \) (e.g., trip travel time, total fuel consumption). The probability distribution function of \( F \) depends on \( x \) and on deterministic exogenous simulation parameters \( p \) (e.g., network topology, total network demand).
A given simulation run yields a realization of the random variable $F$, which involves sampling for each vehicle (or traveler) from the numerous probability distributions that account for uncertainty in, for instance, traveler behavior (e.g., route choice for individual drivers) or traffic generation (e.g., headways of vehicles entering the network). Hence, for any network performance measure, $E[F(x; p)]$ is an intricate function of $x$. Additionally, there is no closed-form expression available for $E[F(x; p)]$; we can only derive estimates for it. Deriving an accurate estimate for a given point $x$ requires running numerous replications of the simulator, each of which are computationally costly to run.

Our past work has considered traditional objective functions, such as expected trip travel time (Osorio and Bierlaire 2013; Osorio and Chong 2015; Chen, Osorio, and Santos 2012). In this paper, we consider more challenging objectives that account for both expected trip travel time as well as vehicle-specific information, such as expected total fuel consumption. The latter depends on instantaneous traffic conditions, and on the underlying vehicle type and technology.

This paper considers two different objective functions: (i) a combination of expected travel time (denoted $f_T$) and expected total fuel consumption (denoted $f_{FC}$); (ii) expected fuel consumption ($f_{FC}$). The first objective function is denoted $f_{T, FC}$ and is defined as a convex combination of $f_T$ and $f_{FC}$

$$f_{T, FC} = wf_T + (1 - w)f_{FC},$$

where $w$ is a weight parameter ($0 \leq w \leq 1$). In the case study of this paper (§4), we consider various weight parameter values.

### 3.2. Simulation-Based Optimization Framework

To address the problem (1)–(3), we use the simulation-based optimization framework of Osorio and Bierlaire (2013), which we refer to as the initial framework. The initial framework considers generally constrained continuous optimization problems. The constraints have analytical differentiable expressions, but there is no analytical expression of the objective function. The latter is defined implicitly by the simulator.

The initial framework is a metamodel SO technique. Each iteration of the SO algorithm considers a given point $x$ and proceeds through two main steps. First, it collects a sample of simulated observations of $F(x; p)$ and estimates $f(x; p)$ by the sample average. This estimate along with estimates at other points in previous iterations, are used to fit an analytical approximation of the objective function. The latter is called the metamodel or surrogate model. Second, the metamodel is used to solve a signal control problem, and to derive a trial point (e.g., new signal plan). The performance of the trial point is then evaluated with the simulator (first step), and the process iterates until the computational budget is depleted. The initial framework uses a derivative-free trust region algorithm based on the algorithm of Conn, Scheinberg, and Vicente (2009). That is, at each iteration of the SO algorithm the trial point is derived by solving a trust region subproblem. The trust region subproblem of the optimization problem (1)–(3) is given in Appendix B.

At every iteration of the SO algorithm, the initial framework builds a metamodel, which is an analytical approximation of the objective function. The metamodel takes the following form:

$$m(x, y; \alpha, \beta, q) = \alpha Q(x, y; q) + \phi(x; \beta),$$

where the following notation is used:

- $x$: decision vector;
- $Q$: approximation of the objective function derived by the analytical traffic (queueing) model;
- $\phi$: polynomial quadratic in $x$ with diagonal second derivative matrix;
- $y$: endogenous queueing model variables; and
- $\alpha, \beta$: metamodel parameters.

The metamodel is a combination of a physical component $Q$ and a generic (also called general-purpose) component $\phi$. The latter is a quadratic polynomial in $x$. Its functional form is problem-independent, yet it can asymptotically provide an excellent local fit to any objective function, and hence ensures asymptotic convergence properties. The physical component $Q$ is an analytical approximation of the objective function provided by a physical model, e.g., a macroscopic urban traffic model or a macroscopic fuel consumption model. It provides a problem-specific global approximation of the objective function. It provides closed-form continuous expressions for the objective function and for its first-order derivatives. Therefore, it surmounts the main limitations of the simulator, and thus improves the computational efficiency of the SO algorithm. The metamodel can be interpreted as an analytical approximation of the objective function provided by the physical component and corrected by both a scale coefficient and an additive quadratic error term.

The parameters of the metamodel, $\alpha$ and $\beta$, are fitted by using the simulated observations and solving a least squares problem (see Osorio and Bierlaire 2013 for details). Hence, the metamodel combines information from a stochastic simulation-based model and an analytical model. For the SO framework to be efficient the analytical model should yield a differentiable and highly efficient approximation of the objective function (Equation (1)). To efficiently address the signal control problem (1)–(3), we need an analytical traffic and an analytical fuel consumption model that together yield a good and highly efficient approximation of the objective functions.
Past work has considered traditional objective functions, mainly expected trip travel time, which can be relatively well approximated by macroscopic models. Nonetheless, the use of less traditional, as well as vehicle-technology dependent, objective functions remains a challenge. This work considers objective functions that account for vehicle-specific fuel consumption. The latter is highly dependent on both vehicle types, vehicle technologies, and instantaneous vehicle acceleration and speeds, which have complex spatial-temporal variations in congested networks. This paper considers tight computational budgets (i.e., few simulation runs are allowed), where deriving a suitable approximation of the relationship between city-wide fuel consumption and signal plans is an even greater challenge.

3.3. Models

3.3.1. Traffic Models. We use the same traffic models as in the initial framework.

Microscopic simulation model. We use a microscopic traffic simulation model of the Swiss city of Lausanne (Dumont and Bert 2006). It is calibrated for evening peak period traffic. This model accounts for the behavior of individual drivers within the network. Trips are generated based on an origin-destination matrix, along with a headway model. Driver behavior is modeled using car following, lane changing, gap acceptance, and route choice models. It is implemented with Aimsun software (TSS 2011).

Macroscopic analytical model. We use an analytical urban traffic model based on the finite capacity queueing theory. Its formulation (given in Appendix A) is derived in Osorio (2010, Chapter 4), which is based on the more general queueing network model of Osorio and Bierlaire (2009). This traffic model combines ideas from finite capacity queueing theory, national transportation norms, and other urban traffic models. It models each lane of an urban network as a set of finite capacity queues. The model uses the finite capacity queueing theory notion of blocking to describe how congestion arises and spatially propagates through the network. It analytically approximates how upstream and downstream queues interact. This is a very simple traffic model, which approximates the delay vehicles incur due to queueing, yet does not account for vehicles acceleration or deceleration. It is formulated as a system of nonlinear equations (System (A1) in Appendix A). For a road network represented by \( n \) queues it is implemented as a system of \( 5n \) nonlinear differentiable and highly tractable equations. The model complexity is linear in the number of queues, which makes it a scalable model suitable for large-scale networks.

3.3.2. Fuel Consumption Models. Microscopic simulation model. We use the microscopic fuel consumption model embedded in Aimsun (v6.1). This detailed model accounts for the time spent by each vehicle in the network during each simulation time step in each of four operating modes: idling, deceleration, acceleration, and cruising. The fuel consumed during the idling, deceleration, and acceleration modes is derived from Ferreira (1982, pp. 4–16), while the fuel consumed during the cruising mode is derived from Akçelik (1983, pp. 51–53).

During a given simulation time step the fuel consumed by a given vehicle \( j \) is given by

\[
FC_j = C^j t^j_1 + C^j t^j_2 + \left( C^j + C^j a_j v_j \right) t^j_3 \\
+ \left[ C^j \left( 1 + \frac{v_j}{2 (V_m^j)^3} \right) + C^j v_j \right] t^j_4, \tag{6}
\]

where the following notation for a given vehicle \( j \) is used:

- \( FC_j \): fuel consumed during a given simulation time step;
- \( C^j \): idling fuel consumption rate;
- \( C^j_2 \): decelerating fuel consumption rate;
- \( C^j_1, C^j_2, C^j_3, C^j_4 \): other vehicle-specific parameters provided by the vehicle manufacturer;
- \( V_m^j \): speed at which vehicle fuel consumption is minimum (vehicle-specific constant);
- \( t^j_M \): time spent in a given mode \( M \in \{ I \) (idling), \( D \) (deceleration), \( A \) (acceleration), \( C \) (cruising)\); and
- \( v_j, a_j \): instantaneous speed and acceleration.

On the right-hand side of Equation (6), capital letters are used to denote the exogenous vehicle-specific parameters. In this paper, we use the parameters corresponding to a 1994 Ford Fiesta (UK DOT 1994). All vehicles in the simulation are of this model. This assumption can be easily relaxed. Microscopic simulators represent individual vehicles, thus accounting for the specific technologies and performance of various fleet compositions is straightforward.

The simulation time step \( \Delta t \) is such that \( \Delta t = t^j_1 + t^j_2 + t^j_3 + t^j_4 \). In the considered simulation, \( \Delta t = 0.75 \) seconds. That is, every 0.75 seconds, the simulator identifies for each vehicle its current operating mode (idling, deceleration, acceleration, or cruising), its instantaneous speed, and its instantaneous acceleration, and derives from Equation (6) the corresponding vehicle-specific instantaneous fuel consumption.

Macroscopic analytical model. The purpose of the analytical model is to provide an analytical, tractable (e.g., differentiable) macroscopic (i.e., aggregate) approximation of the microscopic fuel consumption model. We consider the entire simulation period (e.g., evening peak period). For this time period, the expected fuel consumption per vehicle on link \( \ell \), \( E[FC_{\ell}] \), is approximated by using a simplified version of the Akçelik (1983) model. This is given by

\[
E[FC_{\ell}] = \left(C^\ell (1 + \frac{E[V_{\ell}^3]}{2 (V_m^\ell)^3}) + C^\ell_4 E[V_{\ell}] \right) E[T_{\ell}], \tag{7}
\]
where $C^3$, $C^4$, $V^m$ are vehicle parameters, $E[V_\ell]$ is the expected vehicle speed on link $\ell$, and $E[T_\ell]$ is the expected vehicle travel time on link $\ell$.

Equation (7) is derived by making the following two simplifications to the microscopic fuel consumption model. First, we account for a single operating mode, which is the cruising mode. That is, we assume that throughout the entire trip, the vehicle is in cruising mode. Second, we define the fuel consumption in cruising mode as a function of average vehicle speed (instead of instantaneous speed). If different vehicle types are to be used, similar macroscopic fuel consumption approximations can be derived for each vehicle type.

The approximations for $E[T_\ell]$ and $E[V_\ell]$ are derived as follows. The average link travel time $E[T_\ell]$ is derived by applying Little’s law (Little 1961) to the underlying queue:

$$E[T_\ell] = \frac{E[N_\ell]}{\lambda_\ell(1 - P(N_\ell = k_\ell))},$$

where $N_\ell$ is the number of vehicles in queue $\ell$, $\lambda_\ell$ is the arrival rate to the queue, $k_\ell$ is the space capacity of the queue, and $(1 - P(N_\ell = k_\ell))$ is the probability that the queue is not full (i.e., it can accept flow from upstream). For finite capacity queues the flow entering link $\ell$ is given by $\lambda_\ell(1 - P(N_\ell = k_\ell))$. For a description of how to apply Little’s law to finite capacity queues, refer to Tijms (2003, pp. 52–53).

Both $\lambda_\ell$ and $P(N_\ell = k_\ell)$ are endogenous variables of the macroscopic traffic model. The expected number of vehicles, $E[N_\ell]$, is given by:

$$E[N_\ell] = \rho_\ell \left( \frac{1}{1 - \rho_\ell} - \frac{(k_\ell + 1)\rho_\ell^{k_\ell}}{1 - \rho_\ell^{k_\ell+1}} \right),$$

where $\rho_\ell$ is known in queuing theory as the traffic intensity, an endogenous variable of the macroscopic traffic model. The derivation of Equation (9) is detailed in Osorio (2010, pp. 69–70).

The average speed is approximated using the fundamental relationship that relates the average flow $q$ to average (space-mean) speed $v$ and density $k$: $q = kv$, which in this context is given by:

$$\lambda_\ell(1 - P(N_\ell = k_\ell)) = \frac{E[N_\ell]}{L_\ell} E[V_\ell],$$

where $L_\ell$ denotes the length of link $\ell$. In this equation the flow of a link is given as in Equation (8) by $\lambda_\ell(1 - P(N_\ell = k_\ell))$, and the average link density is approximated by $E[N_\ell]/L_\ell$.

We then use $E[FC_\ell]$ of Equation (7) to approximate the expected total fuel consumption in the network by:

$$E[FC] = \left( \sum_{\ell \in \mathcal{L}} E[FC_\ell] \right) \gamma,$$

where $\mathcal{L}$ is the set of all links in the network and $\gamma$ is the expected number of trips during the considered simulation period (given, for instance, by the origin-destination matrix). This approximation may overestimate fuel consumption because multiplying the expected fuel consumption per vehicle by the total demand is similar to assuming that all vehicles have traveled along all links. An alternate approximation such as $\sum_{\ell \in \mathcal{L}} (E[FC_\ell] \gamma_\ell)$ (where $\gamma_\ell$ is the expected demand for link $\ell$) may be more accurate. To summarize, the analytical approximation of the expected total fuel consumption in the network is obtained by solving Equations (7)–(11).

4. Case Study

4.1. Experimental Setup

As described in §3.3.1, in this study we use a model of the city of Lausanne that represents evening peak period traffic (Dumont and Bert 2006). We consider the first hour of the evening peak period (5–6 p.m.). The network under consideration is in the city center and is delimited by a circle in Figure 1. The detailed network is displayed in the left plot of the figure. This network contains 47 roads and 15 intersections of which 9 are
signalized. The signalized intersections have a cycle time of 90 or 100 seconds and a total of 51 variable phases. This is a complex constrained simulation-based optimization problem.

The queueing model of the network consists of 102 queues. The trust region subproblem (formulated in Appendix B) consists of 621 variables with their corresponding lower bound constraints, 408 nonlinear equality constraints, 171 linear equality constraints, and one nonlinear inequality constraint. This is a high-dimensional simulation-based optimization problem.

We compare the performance of the following three optimization methods. The models to which each of these methods resort are also summarized in Table 1.

- The proposed approach, denoted $Am$.
- A traditional SO metamodel method, where the metamodel consists only of a quadratic polynomial with diagonal second derivative matrix (i.e., the metamodel consists of $\phi$ given in Equation (5)). This approach therefore uses simulation information but does not use information from the analytical traffic model (i.e., it does not have a physical component). This approach is denoted $A\phi$.
- A method that uses only the analytical traffic model, and does not use any simulated information (i.e., the objective function is given by $Q$ in Equation (5)). This method is denoted $AQ$.

We compare the performance of these three methods for the two different objective functions $f_{T,FC}$ and $f_{FC}$.

We consider different types of initial points, i.e., an existing fixed-time signal plan for Lausanne city (for details see Dumont and Bert (2006)), and randomly drawn feasible signal plans. The latter are uniformly drawn from the feasible region defined by Equations (2) and (3). We draw uniformly from this space using the code of Stafford (2006).

For methods $Am$ and $A\phi$, we define the computational budget as a maximum of 150 simulation runs that can be carried out. That is, the algorithm begins with no simulated information; once it has called the simulator 150 times it stops. The point considered as the current iterate (best point found so far) is taken as the proposed signal plan. This is a very tight computational budget, given the dimension and complexity of the problems considered.

The derivation of a proposed signal plan involves calling the simulator. Given the stochastic nature of the simulation outputs, for a given initial point, the methods $Am$ and $A\phi$ are run five times (allowing each time for a maximum of 150 runs). We then compare the performance of all five proposed signal plans.

We evaluate the performance of a proposed signal plan as follows. We embed the proposed signal plan within the Lausanne simulation model. We then run 50 simulation replications, which yield 50 observations of the performance measures of interest (travel time, fuel consumption). For a given performance measure, we plot the cumulative distribution function (cdf) of these 50 observations.

### 4.2. Results

We first compare the performance of the three methods for a combined objective function $f_{T,FC}$. The weight parameter $w$ is set to $4/7$, as defined in Li et al. (2004). Figure 2 considers a given initial point, and displays

---

**Table 1** Traffic and Fuel Consumption Models Used by Each of the Compared Methods

<table>
<thead>
<tr>
<th>Traffic model</th>
<th>Fuel consumption model</th>
</tr>
</thead>
<tbody>
<tr>
<td>(simulation-based)</td>
<td>(analytical)</td>
</tr>
<tr>
<td>(microscopic)</td>
<td>(macroscopic)</td>
</tr>
<tr>
<td>$Am$</td>
<td>✓</td>
</tr>
<tr>
<td>$A\phi$</td>
<td>✓</td>
</tr>
<tr>
<td>$AQ$</td>
<td>✓</td>
</tr>
</tbody>
</table>

---

**Figure 2** (Color online) Performance of the Signal Plans Derived with the $f_{T,FC}$ Objective Function and a (First) Random Initial Signal Plan

- (a) Empirical cdf $F(x)$: Total fuel consumption in network (L)
- (b) Empirical cdf $F(x)$: Average travel time in network (min)
the performance of various signal plans in terms of total fuel consumption (left plot) and average travel time (right plot). Each plot displays 12 cdf curves. The five solid black (resp. dashed red) cdfs correspond to the cdfs of the signal plans proposed by Am (resp. Aφ). The green cdf corresponds to the signal plan proposed by AQ, and the blue is that of the initial signal plan. Each cdf curve consists of 50 observations that correspond to the 50 simulation replications. The optimization methods were initialized with a random (uniformly drawn) initial signal plan (blue cdf).

Figure 2(a) indicates that three of the signal plans derived by Am outperform, in terms of total fuel consumption, all other signal plans, and in particular all those derived by Aφ. The performance of the other two signal plans derived by Am is similar to that of the plans derived by Aφ. The signal plans derived by Am also have reduced variability compared to those of Aφ. For both Am and Aφ, all proposed plans show improved performance compared to the initial plan; four of five show improved performance compared to the signal plan proposed by AQ. Figure 2(b) considers the same signal plans and displays their performance in terms of average travel time. Similar conclusions hold.

Figure 3 carries out the same experiment as Figure 2, but considers a different random (uniformly drawn) initial signal plan. Figure 3(a) displays the cdfs of total fuel consumption. The four best signal plans are derived by Am; these plans also have the smallest variance in total fuel consumption. The fifth plan derived by Am shows the worst performance. Similar conclusions hold when evaluating the signal plans in terms of average travel time (Figure 3(b)).

Figure 4 considers an existing signal plan for the city of Lausanne as the initial plan. Four of the five plans derived by Am are among the best signal plans. The fifth shows similar performance to the existing Lausanne plan. The performance of one of the signal plans is shown in Table 1.
plans derived by \( A\phi \) is worse compared to the initial plan. The signal plan derived by \( AQ \) shows similar performance to that of the initial signal plan. The same conclusions hold when evaluating the signal plans in terms of average travel time (Figure 4(b)).

In Figures 2–4, 13 of the 15 plans derived by \( Am \) outperform the plans derived by \( AQ \), in terms of both fuel consumption and travel time. This shows the added value of using both analytical and simulated information as opposed to only analytical information.

The information provided in Figures 2(a), 3(a), and 4(a) is summarized in Figure 5(a). The latter displays two cdfs, one for all signal plans derived by \( Am \) (solid black) and one for all those derived by \( A\phi \) (dashed red). Each cdf consists of all fuel consumption observations displayed in the three previously mentioned figures (i.e., each cdf consists of \( 50 \times 5 \times 3 = 750 \) total observations). Similarly, Figure 5(b) summarizes the information provided in Figures 2(b), 3(b), and 4(b). In both cases, the signal plans proposed by \( Am \) outperform those proposed by \( A\phi \). These figures show that there is added value in complementing the simulated information with analytical information.

Figure 6 considers the same initial point as Figure 2 and the objective function \( f_{T, FC} \). Each plot displays the network of interest. Each link is colored according to the average (over 50 replications) fuel consumption per vehicle (in liters). The top plot considers the performance under the initial signal plan, whereas the bottom plot displays the performance of a signal plan derived by \( Am \). This figure illustrates the fuel consumption reductions that can be achieved at the link level by the proposed signal plan.

Figures 7–10 consider a signal control problem with an objective function that accounts only for total fuel consumption \( (f_{FC}) \). This is a challenging problem, as fuel consumption is a highly variable metric. Additionally, we are attempting to identify a signal plan with 51 variable phases (dimension of the decision vector) by calling the simulator at most 150 times.

To evaluate the performance of the signal plans, we proceed as earlier. That is, we evaluate their performance both in terms of fuel consumption and travel time. Figure 7 considers the same initial plan as Figure 2. Figure 7(a) indicates that of the top four signal plans with best performance, three are proposed by \( Am \). Four of the plans proposed by \( Am \) outperform four of
the plans proposed by $A\phi$. These four plans also lead to reduced variability. Both $Am$ and $A\phi$ derived one signal plan with poorer performance than the initial plan. When evaluating these plans in terms of their travel times (Figure 7(b)), four of the plans proposed by $Am$ are among those with the best performance, whereas the fifth performs similar to the initial plan.

Figure 8 considers the same initial plan as in Figure 3. Figure 8(a) compares the total fuel consumption of the signal plans. Here the top four signal plans with best performance are proposed by $Am$. These plans also have reduced variability compared to those proposed by $A\phi$ and to the initial signal plan. Similar conclusions hold when evaluating these signal plans in terms of average travel time (Figure 8(b)).

Figure 9 considers the existing Lausanne signal as the initial plan (i.e., same initial plan as Figure 4). For both fuel consumption (Figure 9(a)) and travel time (Figure 9(b)) four of the top five signal plans are proposed by $Am$. As before, the plans proposed by $Am$ lead to reduced variability compared to those proposed by $A\phi$.

Figure 10 summarizes the information provided in Figures 7–9. Each plot of Figure 10 considers the same two performance measures as the plots in Figures 7–9. For each performance measure, Figure 10 displays two cdfs, one for all signal plans derived by $Am$ (solid black) and one for all those derived by $A\phi$ (dashed red). Each cdf therefore consists of all observations (of fuel consumption or travel time) displayed in the three previously mentioned figures (i.e., each cdf consists of $50 \times 5 \times 3 = 750$ total observations). This figure indicates that the signal plans proposed by $Am$ outperform those proposed by $A\phi$ in terms of both total fuel consumption and average travel time.

These figures show that there is added value in combining simulated and analytical information. Additionally, because fuel consumption observations strongly depend on individual vehicle attributes and complex...
local traffic dynamics, they have high variability. Thus, an algorithm that uses only simulated information is typically at a disadvantage compared to one that combines suitable analytical and simulated information.

We perform a sensitivity analysis for the weight parameter $w$ of the travel time component of the objective function (Equation (4)). We consider the weight values: $w \in \{1, 6/7, 4/7, 2/7, 0\}$. For each weight value, we proceed as earlier: we solve the corresponding optimization problem once for $AQ$, and five times for each of the simulation-based methods ($Am$ and $A\phi$) with a computational budget of 150. We initialize the algorithms with the first random initial plan used in Figures 2 and 7.

Figure 11 displays the corresponding cdfs. Each row of plots corresponds to a given weight value. The left (resp. right) column plots display total fuel consumption (resp. average travel time). The results for weight $w = 4/7$ are displayed in Figure 2. The results for weight $w = 0$ are displayed in Figure 7, and re-displayed here to facilitate the comparison. The cdfs of Figure 11 indicate that 18 of the 20 signal plans proposed by $Am$ outperform those proposed by $A\phi$ both in terms of fuel consumption and travel time. Most also outperform $AQ$. In other words, the method that combines analytical and simulation-based information outperforms the methods that resort to only analytical or only simulation-based information.

The simulation-based method $A\phi$ yields signal plans with similar performance for $w = 1$ (first row: Figures 11(a) and 11(b)). For $w = 1$, the objective function consists of only travel time and does not account for fuel consumption. For any other weight value ($w < 1$), fuel consumption is accounted for, and the five signal plans proposed by $A\phi$ differ significantly in performance. This shows the complexity of solving a problem with a high-variance simulation-based objective function. For $w < 1$, the five signal plans proposed by $Am$ show lower variability in performance than those proposed by $A\phi$. This illustrates the added value of
Figure 11  (Color online) Performance of the Signal Plans Derived with Different Weight Parameter, $\omega$, Values and a (First) Random Initial Signal Plan
using analytical information to address high-variance simulation-based problems.

5. Conclusion

The numerical results show that when accounting for high-variance performance measures, such as fuel consumption, there is added value in combining information from the simulator with approximations derived from analytical macroscopic traffic models. The proposed methodology outperforms both traditional methodologies, which use only simulated information (Aφ) or only analytical information (AQ). The proposed methodology systematically achieves reductions in both travel time and fuel consumption, and does so within a very tight computational budget.

With energy efficiency being a growing concern for the transportation industry, this paper demonstrates how detailed traffic and vehicle-performance simulation tools can be coupled and used to design traffic management strategies that improve network-wide performance metrics.

Ongoing work addresses traffic management problems that account for environmental performance measures that are more complex to approximate analytically, and also have high variability. We are addressing SO signal control problems that account for greenhouse gas emissions and various pollutant emissions (Osorio and Nanduri 2015). Additionally, to better address problems with high-variance performance measures, we are developing simulation-budget allocation strategies that determine the number of simulation replications (i.e., sample size) to allocate across various points (Chingcuanco and Osorio 2013). The technique, known as a subset selection procedure, performs well for small sample size problems, such as those considered in this paper. Preliminary results with traditional signal control problems indicate that subset selection procedures can significantly enhance the computational efficiency of SO algorithms.

The importance of incorporating fuel consumption and emissions in our framework is exemplified by, for instance, United States federal regulations such as the Clean Air Act and the Surface Transportation Efficiency Act, which place increasing responsibility on city and regional agencies to account for and achieve their environmental targets. The success of traffic management strategies now depends on their demonstrated ability to prevent further degradation of air quality in the surrounding areas. However, as is currently the practice, the use of integrated traffic-fuel-emissions models is primarily restricted to observing the effect of predetermined alternatives on emissions and/or fuel consumption. The optimization framework presented in this paper enables practitioners to go beyond evaluation purposes and systematically identify alternatives with improved urban-scale performance.

Appendix A. Analytical Queueing Network Model

The physical component of the metamodel is an analytical and differentiable urban traffic model. Each lane of an urban road network is modeled as a set of finite capacity queues. In the following notation the index $i$ refers to a given queue. We refer the reader to Osorio (2010, Chapter 4) and to Osorio and Bierlaire (2009) for details.

$\gamma_i$: external arrival rate;
$\lambda_i$: total arrival rate;
$\mu_i$: service rate;
$\bar{\mu}_i$: unblocking rate;
$\mu_i^{eff}$: effective service rate (accounts for both service and eventual blocking);
$\rho_i$: traffic intensity;
$P^f_i$: probability of being blocked at queue $i$;
$k_i$: upper bound of the queue length;
$N_i$: total number of vehicles in queue $i$;
$P(N_i = k_i)$: probability of queue $i$ being full, also known as the blocking or spillback probability;
$p_i^f$: transition probability from queue $i$ to queue $j$; and
$D_i$: set of downstream queues of queue $i$.

The queueing network model is formulated as follows:

$$
\lambda_i = \gamma_i + \sum_j p_{ij} \lambda_j (1 - P(N_j = k_j)) / (1 - P(N_i = k_i)) \quad (A1a)
$$

$$
\frac{1}{\mu_i} = \sum_{j \in D_i} \lambda_j (1 - P(N_i = k_i)) (1 - \lambda_j) / \mu_j \quad (A1b)
$$

$$
\frac{1}{\mu_i^{eff}} = 1 / \mu_i + p_i^f \frac{1}{\bar{\mu}_i} \quad (A1c)
$$

$$
P(N_i = k_i) = \frac{1 - p_i}{1 - \rho_i} p_i^f \quad (A1d)
$$

$$
p_i^f = \sum_j p_{ij} P(N_j = k_j) \quad (A1e)
$$

$$
\rho_i = \lambda_i / \mu_i^{eff} \quad (A1f)
$$

The exogenous parameters are $\gamma_i$, $\mu_i$, $p_{ij}$, and $k_i$. All other parameters are endogenous. When used to solve a signal control problem (as in this paper), the flow capacity of the signalized lanes become endogenous, which makes the corresponding service rates, $\mu_i$, endogenous.

Appendix B. Trust Region Subproblem

At a given iteration $k$, the SO algorithm considers a metamodel $m_k(x, y; \alpha_k, \beta_k, q)$, an iterate $x_k$ (point considered to have best performance so far) and a trust region (TR) radius $\Delta_k$, and solves the TR subproblem to derive a trial point (i.e., a signal plan with potentially improved performance). The TR subproblem is formulated as follows:

$$
\min_{x, y} m_k = \alpha_k Q(x, y; q) + \phi(x; \beta_k) \quad (B1)
$$

subject to

$$
\sum_{j \in J_k} x(j) \leq c_i, \quad \forall i \in J, \quad (B2)
$$

$$
h(x, y; q) = 0, \quad (B3)
$$

$$
\mu_i - \sum_{j \in J_k} x(j) = c_i s, \quad \forall i \in I, \quad (B4)
$$
where $s$ denotes the saturation flow rate, $e_i$ is the ratio of fixed green time to cycle time of signalized lane $\ell', \ell'$ denotes the set of indices of the signalized lanes, $P_i(\ell)$ denotes the set of phase indices of lane $\ell$, and $h$ is the analytical traffic model, which is given by the System of Equations (A1).

The TR subproblem differs from the signal control problem given in §3.1 as follows. The TR subproblem approximates the objective functions by the metamodel at iteration $k$, $m_k$.

It includes the following additional constraints:

- Inequality (B5). This is known as the trust region constraint.
- Equation (B3). The function $h$ of Equation (B3) represents the queuing network model given above by the System of Equations (A1).
- Equation (B4). This equation relates the green splits of a phase to the flow capacity of the underlying signalized lanes (i.e., the service rate of the queues).
- Equation (B6). The endogenous variables of the queuing model are subject to positivity constraints.

### References


Yu L (1994) A mathematical programming based approach to macroscopic traffic assignment in a dynamic network with queues. Technical report, Department of Civil Engineering, Queen’s University, Kingston, Ontario, Canada.