On the control of highly congested urban networks with intricate traffic patterns: a New York City case study

Carolina Osorio* Xiao Chen† Jingqin Gao‡ Mohamad Talas§ Michael Marsico¶

1 Introduction

This paper focuses on the design of traffic signals for highly congested urban networks. It is motivated by a case study of an intricate network within New York City, which we refer to as the Queensboro Bridge (QBB) network. Figure 1 displays a map of central Manhattan, the QBB area of interest is delimited by the black oval. Figure 2(a) displays a detailed map of the QBB area (left plot) and the corresponding network model (right plot).

The design of suitable signal plans, and more generally of suitable traffic management strategies, for this area is a major challenge due to the intricate traffic patterns, which arise for the following reasons.

Highly congested. The network is highly congested. Morning peak period vehicular traffic, which is the focus of this paper, is in the order of 12,000 trips per hour.

Uniformly congested. Congestion is homogenously distributed, i.e., it arises in all directions of travel. Additionally, all cross-streets are considered important. Hence, there is no clear hierarchy or priority between intersections or between roads.

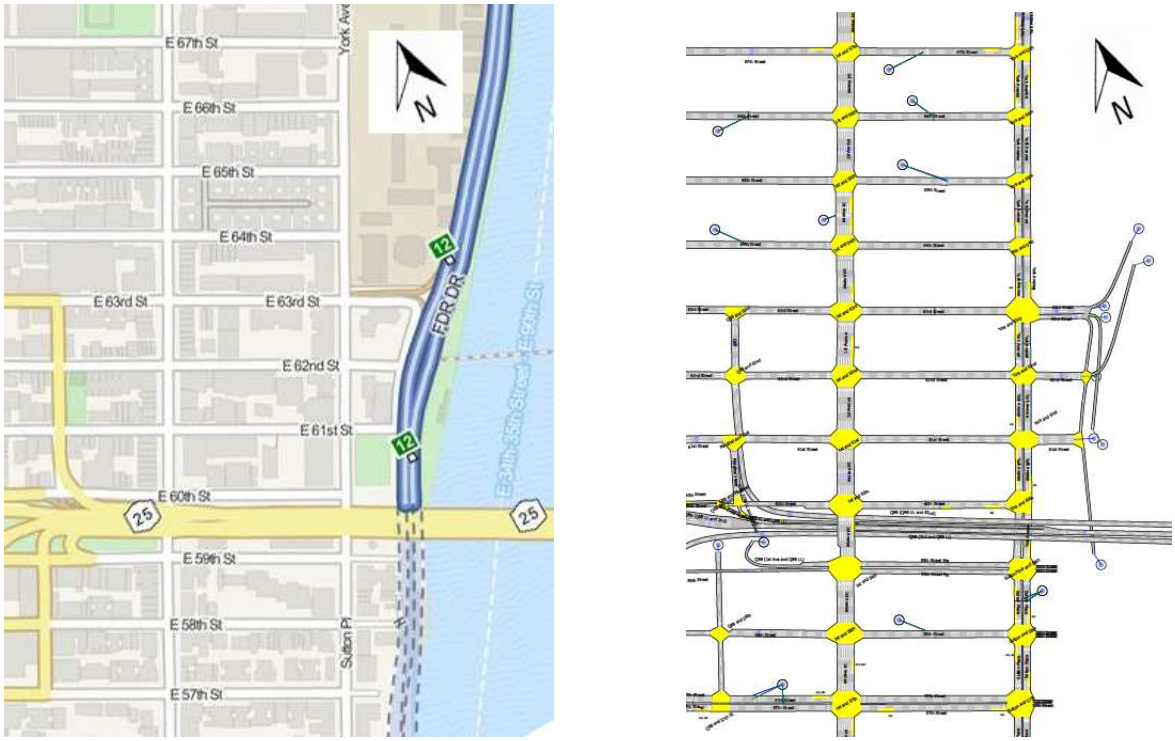
Queensboro Bridge access. This area includes the Ed Koch Queensboro Bridge, which connects Queens to Manhattan. This is the busiest New York City Department of Transportation (NYCDOT) bridge, with average weekday traffic in the order of 178,000 vehicles (NYCDOT; 2015).

*Massachusetts Institute of Technology, Cambridge, MA, USA
†School of Highway, Chang'an University, Xi'an, China
‡New York City Department of Transportation, New York, NY, USA
§New York City Department of Transportation, New York, NY, USA
¶New York City Department of Transportation, New York, NY, USA

Figure 1: Central Manhattan network with area of interest delimited by an oval (MapQuest.com, Inc; 2015).

During the morning peak hour 8-9am, which is the focus of this paper, an estimated 5770 vehicles cross the bridge in the Manhattan-bound direction (NYCDOT; 2012). The Queensboro Bridge also serves as a key transit corridor that carries 694 buses and approximately 16,000 bus passengers on a typical weekday. It is the busiest express bus route exiting Manhattan, connecting neighborhoods in Midtown Manhattan and eastern Brooklyn and Queens (NYCDOT; 2011). During peak period traffic, streets adjacent to the bridge are typically congested, and travel times along the bridge increase significantly (NYCDOT; 2010). The occurence of congestion along access/egress links to the bridge can have significant impacts on congestion both in Manhattan-side links and Queens-side links, and on transit service. Vehicular queues during peak periods often require traffic enforcement agents (TEAs) to prevent spillback and gridlock. This motivates the design of signal plans that mitigate the spatial propagation of congestion. The importance of this bridge for both Manhattan and Queens traffic indicates that even small local changes to traffic operations in this area can trigger significant larger-scale impacts.

The FDR Drive access. The FDR Drive, i.e., the Franklin D. Roosevelt East River Drive, is an



(a) (b)

Figure 2: Queensboro bridge network of interest: map (left plot) (MapQuest.com, Inc; 2015) and model (right plot).

important limited-access facility. It allows traffic to travel along Manhattan's east edge. It consists mostly of three lanes in each direction. Several of its exits and entrances are located within the study area (i.e., northbound exit at East 61st Street to NY 25 (Queensboro bridge), northbound entrance at East 62nd Street, southbound entrance/exit at East 63rd Street to NY 25 (Queensboro bridge)). Local streets near the FDR Drive (e.g. York Avenue at 63rd Street) often experience heavy turn volumes and high levels of congestion during peak hour.

Heavy pedestrian traffic. The network has intricate multi-modal traffic patterns (e.g., cars, trucks, buses, bikes and pedestrians). It has considerable pedestrian traffic. For instance, the pedestrian volume at the 1st Avenue intersection with 59^{th} Street can be of the order of 500 pedestrians per hour during morning peak period (NYCDEP; 2006). Hence, the design of traffic signals for vehicular traffic is constrained by non-negligeable pedestrian crossing times, and this at every intersection.

Short links. Most links of this network are short links. Such links are prone to the occurence of spillbacks, and more generally to a rapid spatial propagation of congestion.

Grid topology. The links are configured in a grid topology. This allows for complex traveler behavior, such as high-dimensional route choice alternatives. This makes the prediction of how traffic (i.e., travelers) will respond to changes in the traffic operations a greater challenge.

The combination of the above factors makes the design of signal control strategies a challenge. Traditional signal control strategies are difficult to tailor to the specific needs of such networks. For instance, they are designed based on low-resolution macroscopic traffic models which do not provide a detailed description of traveler behavior (e.g., route choice), of the underlying network supply (e.g., prevailing traffic operations), or of their interactions. They are designed such as to improve pre-determined objective functions (e.g., average link travel times, average number of stops), which may not be suitable for the network of interest.

For such networks, the most critical shortcoming of existing signal control methods is the lack of a detailed modeling or estimation of the between-link interactions, which trigger the spatial propagation of congestion. Most signal control strategies do not account for between-queue interactions. They are based on the use of vertical queues (also called point queues) that do not account for the spatial propagation of the vehicular queues, hence they do not describe phenomena such as spillbacks. They are appropriate for low to moderate levels of congestion. They are unsuitable for highly congested networks (Papageorgiou et al.; 2003; Abu-Lebdeh and Benekohal; 2003).

For congested networks, mitigating the spatial propagation of congestion, as well as the occurrence of spillbacks, is recognized as a major goal (Chow and Lo; 2007; Abu-Lebdeh and Benekohal; 2003). For instance, in the QBB network, the occurence of spillbacks can affect access/egress from the QBB, leading to significant impacts on Manhattan-Queens traffic throughput. The work of Geroliminis and Skabardoni (2011); Skabardonis and Geroliminis (2008) emphasize the importance of controlling queue spillbacks (or spillovers): "when spillovers occur, the travel delay can increase by 50%-100% for short distance links between successive intersections."

Several signal control methods based on queue-length information have been proposed. Examples include Abu-Lebdeh and Benekohal (2003); Diakaki et al. (2002); Michalopoulos and Stephanopoulos (1977a,b). They are all based on the use of macroscopic traffic models. The vast majority of the research in the field of macroscropic traffic modeling has focused on the development of link models, and, in the context of urban traffic, on the description of the within-link propagation of congestion. A description of the across-node (i.e., the between-link) interactions in an urban environment is limited (Lebacque and Khoshyaran; 2005; Tampère et al.; 2011; Flötteröd and Rohde; 2011; Corthout et al.; 2012). This is arguably due to the complexity of providing an analytical, let alone differentiable, description of node priorities. Existing approaches therefore embed a highly simplified description of between-link interactions, and are mostly based on non-differentiable expressions, which limits their use within traditional gradient-based optimization algorithms. Detailed and differentiable descriptions of between-link interactions are needed to account for the impact of spillbacks.

More recently, various signal control methods that rely on real-time queue-length estimates and are based on feedback control laws have been proposed (Gregoire et al.; 2014; Lioris et al.; 2014; Varaiya; 2013; Wongpiromsarn et al.; 2012). These algorithms are highly scalable, yet most urban networks currently do not have sensors deployed to provide accurate queue-length measurements (Papageorgiou and Varaiya; 2009). The extension of these algorithms to model between-queue interactions in the abscence of accurate queue-length estimates is a topic of ongoing research.

The challenges of developing analytical differentiable network models that provide a suitable description of between-link interactions while remaining sufficiently tractable for large-scale traffic control have led to an increased focus on the use of simulation-based models. For a review, see Barceló (2010). In particular, high-resolution simulators, known as microscopic simulators, can provide a detailed description of the intricate between-queue interactions. This is because these simulators represent individual vehicles, and they embed a detailed description of the network supply (e.g., prevailing traffic management strategies). Hence, they yield a high-resolution description of intricate local traffic phenomena, such as spillbacks. Nonetheless, the computational inefficiency of

these high-resolution models has mostly limited their use to what-if analysis (i.e., scenario-based analysis), such as in Bullock et al. (2004); Ben-Akiva et al. (2003). Their use within simulation-based optimization (SO) algorithms is limited (Osorio and Nanduri; 2015a,b; Osorio and Chong; 2015; Osorio and Selvam; 2015; Osorio and Bierlaire; 2013; Li et al.; 2010; Stevanovic et al.; 2009, 2008; Branke et al.; 2007; Yun and Park; 2006; Hale; 2005; Joshi et al.; 1995).

There is a need for SO algorithms that can identify solutions to transportation problems in a computationally efficient manner, i.e., algorithms that can identify points with improved performance within few simulation runs. Such algorithms would enable an efficient use of these inefficient simulators. They are of particular importance and relevance for transportation practitioners, who typically address the optimization problems under tight computational budgets (i.e., few simulation runs are carried out).

In our past work, we have proposed efficient SO algorithms. In order to achieve efficiency information from the simulator has been combined with information from analytical traffic models. The latter provide analytical and differentiable structural problem-specific information to the SO algorithms. For various types of signal control problems, we have shown that SO algorithms that embed analytical information from the macroscopic models outperform those that do not (Osorio and Nanduri; 2015a; Osorio and Chong; 2015; Osorio and Nanduri; 2015b; Chen et al.; 2012; Osorio and Bierlaire; 2013). In this paper, we study the tailor and use these SO algorithms to address signal control problems for networks with particularly intricate traffic patterns.

This paper compares the performance of two SO methodologies that differ only in the analytical traffic information provided to the SO algorithm. The first method uses an analytical traffic model, which has been used in our past SO work and which provides an analytical description of the between-link interactions. We formulate a second analytical traffic model that differs from the first only in the description of the between-link interactions. The second model does not account for between-queue interactions. The second SO method embeds this simplified traffic model. Both SO methods are used to address a signal control problem for morning peak period in the QBB network. The main contributions of this paper are the following.

SO methodologies. The main methodological question addressed in this paper is: what is the key information provided by the macroscopic models to the SO algorithms that ensures their computational efficiency? We show that it is necessary to provide an analytical description of between-link interactions in order to systematically and efficiently identify well performing signal plans.

The results of this paper therefore emphasize the need and importance of formulating analytical

and differentiable macroscopic models that both quantify these interactions and can be used for large-scale optimization. These results also highlight the need to develop methodologies that can improve our understanding of the relationship between spatial network dependencies and traffic operations.

The findings of this paper contribute to inform the design of the next-generation of SO algorithms. For instance, analytical between-queue dependency can be used to improve sampling strategies (e.g., sampling of model improvement points (Osorio and Bierlaire; 2013)) or to develop more efficient statistical comparisons of point performance (e.g., ranking and selection strategies).

Signal control. This paper contributes to the field of signal control. It provides evidence that detailed modeling or estimation of between-link interactions is critical for the control of urban networks with such intricate traffic patterns. Past simulation-based signal control studies have considered small-scale networks with a couple of arterials at most, and with simple (mostly linear) network topologies (Park and Kamarajugadda; 2009; Stevanovic et al.; 2008; Yun and Park; 2006). This paper considers a large-scale microscopic model of a grid-topology network with over 130 roads and over 30 signalized and non-signalized intersections. The signal control problem controls a set of 26 adjacent intersections, this is considered a complex and large-scale problem in the field of urban signal control.

The problem considered in this paper is a simulation-based non-convex constrained problem with a decision vector of dimension 64. This is considered a challenging problem in the field of SO.

State-of-practice. Current practice in the design of signal plans for this QBB area uses commercial software to design the signal plans. The software relies on simple traffic models. The derived signal plans are then embedded within these high-resolution simulators in order to provide a more detailed evaluation of their performance. Hence, the simulator is only used after the signal design process as an evaluation or validation tool. The results of this paper contribute to the state-of-practice by enabling the agency to systematically and efficiently use their high-resolution simulator within the signal design process. In other words, this work allows the agency to go one step further and to systematically use the simulation model within the signal optimization algorithm.

Manhattan case studies. In this paper, we control a set of 26 signalized intersections within an area of Manhattan. Other signal control strategies illustrated with Manahattan case studies

include a simulation analysis for a nine-intersection network (Spall and Chin; 1997) and both a simulation and an empirical analysis for a seven-intersection network (Rathi; 1988). To the best of our knowledge, this paper considers the largest and most intricate Manhattan simulation-based optimization problem addressed so far.

This paper is structured as follows. Section 2 presents the SO methodologies studied in this paper. Section 3 presents the signal control problem of interest. The QBB case study is presented in Section 4. The main conclusions are presented in Section 5.

2 Methodology

Section 2.1 gives a brief description of the types of urban traffic simulators considered in this paper. Section 2.2 presents the general SO methodology that is used in this paper. The methodology combines information from a simulation-based traffic model and an analytical traffic model. Section 2.3 presents the two analytical traffic models that are compared in this paper. Section 2.3.1 presents the model that has been used in past SO work. Section 2.3.2 formulates a simplified traffic model, which differs only in the description of the between-link interactions.

2.1 High-resolution traffic simulators

The high-resolution traffic models considered in this paper are stochastic microscopic urban traffic simulators. They describe demand at the scale of individual travelers or individual vehicles. For instance, they embed models that describe how travelers make travel decisions such as mode choice, route choice, or how drivers behave (e.g., car-following, lane-changing). These models account for the hetereogenity of traveler behavior. A given simulation run involves sampling a population of travelers, each with their own set of travel decisions. For instance, in the QBB case study of this paper, one simulation run samples approximately 12,000 travelers/vehicles distributed across a set of 55 origin-destination pairs. The en-route behavior of each traveler is then simulated by sampling from a variety of probabilistic behavioral models (e.g., car-following, lane-changing). These simulators also provide a detailed description of the underlying network supply (e.g., traffic management strategies).

2.2 Simulation-based optimization algorithm

We briefly describe the main ideas of the considered algorithm (Osorio and Bierlaire; 2013). The algorithm is given in Appendix A. For algorithmic details, we refer the reader to Osorio and Bierlaire (2013). The algorithm considers a broad family of SO problems. The problems have continuous

decision variables. They have a simulation-based objective function, i.e., the objective function is derived from the simulator, no closed-form analytical expression is available. The problems have general (e.g., nonconvex) constraints. Closed-form analytical and differentiable expressions are available for all constraints (i.e., the constraints are not simulation-based).

This family of problems can be formulated as follows.

$$\min_{x \in \Omega} f(x, z; p) \equiv E[G(x, z; p)], \tag{1}$$

where the purpose is to minimize the expected value of a given stochastic performance measure G, x denotes the deterministic continuous decision vector, z denotes other endogenous variables, and p denotes the deterministic exogenous parameters. For example, in a signal control problem, G can represent link or network queue-lengths, x can denote signal timing variables such as green times. The vector z can represent route choice decisions or signalized link flow capacities, while p accounts for network topology, lane attributes (length, grade, speed limits) or other exogenous prevailing traffic management strategies (e.g., lane-use priorities, pricing). The feasible region, Ω, consists of a set of general, typically nonconvex, deterministic, analytical and differentiable constraints. In signal control, these include lower bounds for green times and cycle time constraints.

High-resolution simulation models are computationally inefficient to evaluate. The simulation-based functions (e.g., f of Equation (1)) are not known in closed-form, they can only be estimated via simulation. In order to obtain an accurate estimate, numerous simulation replications are needed. Given their high-resolution, the estimates of network performance are typically nonlinear, often nonconvex, functions. Hence, Problem (1) is difficult to address. The algorithm of Osorio and Bierlaire (2013) has enabled to address such problems in a computationally efficient manner. It has been used to address various signal control problems (Osorio and Chong; 2015; Osorio and Nanduri; 2015a,b).

The algorithm is known as a metamodel simulation-based optimization (SO) algorithm. At iteration k of the algorithm, it considers an analytical approximation of the simulation-based objective function (f in Equation (1)), denoted $m_k(x; \beta_k)$, it then uses this approximation to solve a subproblem, which can be generally formulated as:

$$\min_{\Omega_k} m_k(x; \beta_k). \tag{2}$$

The analytical approximation, m_k, is known as a metamodel and is a function of the decision vector x and of an iteration-specific parameter vector β_k. At a given iteration k of the algorithm, a set of simulation observations are available and are used to fit the metamodel parameter vector, β_k, such that the metamodel approximates well the underlying simulation-based objective function f. As the

algorithm advances (i.e., as k increases), more simulation observations become available, and hence the accuracy of the metamodel is expected to increase.

In Osorio and Bierlaire (2013) the metamodel is formulated as follows:

$$m_k(x, y; \beta_k, q) = \beta_{k,1} f_A(x, y; q) + \phi(x; \tilde{\beta}_k), \tag{3}$$

where:

- ϕ denotes a general-purpose analytical and differentiable model. It is specified as a quadratic polynomial in x with diagonal second-derivative matrix. The parameter vector $\tilde{\beta}_k$ denotes the polynomial coefficients.

- $f_A(x, y; q)$ denotes the approximation of the simulation-based objective function f (of Equation (1)) provided by an analytical and differentiable traffic model, with endogenous traffic model variables y (e.g., link densities) and exogenous parameters q (e.g., total travel demand). In Equation (3), f_A is multiplied by the scalar $\beta_{k,1}$, which is the first element of the parameter vector β_k, i.e., $\beta_k = (\beta_{k,1}, \tilde{\beta}_k)$.

In the metamodel literature, ϕ is known as a functional model, while f_A is known as a physical model. The purpose of ϕ is to enable a general-purpose approximation of the objective function f, and the choice of a quadratic polynomial allows to guarantee asymptotic convergence of the algorithm. The purpose of f_A is to provide a problem-specific approximation of the simulation-based objective function f. Additionally, since this approximation is derived from a traffic model, it can provide the algorithm with a good and global (i.e., across the entire feasible region) approximation of the relationship between the decision vector and the objective function. The key to achieving computational efficiency lies in the function f_A. The latter allows the algorithm to identify points with good performance even when few simulation observations are available. Since it provides an approximation in the entire feasible region, it allows the algorithm to quickly (i.e., with few simulation observations) identify subregions with good performance.

Since the Subproblem (2) is solved at every iteration of the SO algorithm, it is critical to be able to solve it efficiently. In order to achieve this, the metamodel m_k should be a highly tractable and computationally efficient model. Thus, when using the SO algorithm of Osorio and Bierlaire (2013) to address a specific optimization problem, the key to achieving computational efficiency lies in the formulation of a computationally efficient expression for f_A.

Past signal control work has shown that the above metamodel (Equation (3)) outperforms both: (i) a metamodel that consists only of the general-purpose component (i.e., the metamodel consists only of ϕ), and (ii) a metamodel that consists only of the physical component (i.e., the metamodel

consists only of f_A) (Osorio and Nanduri; 2015a; Osorio and Chong; 2015; Osorio and Nanduri; 2015b; Chen et al.; 2012; Osorio and Bierlaire; 2013). Past studies have therefore shown the added value of providing the SO algorithm with analytical information from a macroscopic traffic model. This paper investigates what information provided by the macroscopic model is key for signal control. In particular, we show that it is key to provide the algorithm with an analytical differentiable description of between-link interactions.

2.3 Analytical traffic models

2.3.1 Analytical model with between-link interactions

The analytical traffic model used to derive f_A (of Equation (3)) is based on a combination of traffic flow theory ideas and queueing network theory ideas. Its formulation for general queueing networks is given in Osorio and Bierlaire (2009). Its formulation for urban road networks is given in Osorio (2010, Chap. 4). The main idea is that each lane of a road network is modeled as a finite (space) capacity queue. The model uses the queueing theoretic notion of *blocking* to describe the occurence and effects of vehicular spillbacks. Link (rather than network) models of vehicular traffic based on queueing theory include Osorio and Flötteröd (2015); Osorio et al. (2011); Heidemann (2001); Jain and Smith (1997); Heidemann (1996, 1994); Tanner (1962). The considered model is formulated as follows. The index i refers to a given queue.

γ_i	external arrival rate;
λ_i^{eff}	effective arrival rate;
μ_i	service rate;
$\hat{\mu}_i$	effective service rate;
ρ_i	traffic intensity;
ℓ_i	space capacity;
N_i	queue-length;
$P(N_i = \ell_i)$	probability of queue i being full;
p_{ij}	routing probability from queue i to queue j;
$\mathcal{D}_i$	set of downstream queues of queue i.

$$
\begin{cases}
\lambda_i^{\text{eff}} = \gamma_i(1 - P(N_i = \ell_i)) + \sum_j p_{ji}\lambda_j^{\text{eff}} & \text{(4a)} \\
P(N_i = \ell_i) = \dfrac{1-\rho_i}{1-\rho_i^{\ell_i+1}}\rho_i^{\ell_i} & \text{(4b)} \\
\rho_i = \dfrac{\lambda_i^{\text{eff}}}{(1 - P(N_i = \ell_i))\hat{\mu}_i} & \text{(4c)} \\
\dfrac{1}{\hat{\mu}_i} = \dfrac{1}{\mu_i} + \left(\sum_j p_{ij} P(N_j = \ell_j)\right)\left(\sum_{j\in\mathcal{D}_i} \dfrac{\lambda_j^{\text{eff}}}{\lambda_i^{\text{eff}}\hat{\mu}_j}\right). & \text{(4d)}
\end{cases}
$$

Equation (4a) is a flow conservation equation that defines the (expected) flow of a given queue i as the sum of the flow arising from outside the network (this external flow corresponds to the term γ_i, which is the rate at which vehicles start trips at queue i) and of the flow arising from upstream links (this is given by the summation term). Equation (4b) defines the probability that a queue is full. This probability is also known as the spillback probability or the blocking probability. It represents the probability that the underlying lane spills back. This expression is obtained by modeling the underlying lane as an M/M/1/ℓ queue (e.g., Bocharov et al.; 2004). This probability is a function of the lane's space capacity ℓ_i and the lane's traffic intensity ρ_i. The latter is defined by Equation (4c), and can be interpreted as the ratio of expected demand to expected supply. Note that the use of finite (space) capacity queues (i.e., assuming $\ell_i < \infty$) allows for any non-negative value of ρ_i. In particular, values larger than one are allowed. This means that highly congested scenarios where the expected demand exceeds the expected supply can be accounted for. This is particularly important when considering peak-period scenarios, like the one of this paper.

Equation (4d) describes the effective service rate of queue i, denoted $\hat{\mu}_i$, this notion is similar to an effective flow capacity term. The term *effective* relates to the fact that the flow capacity of a link can be affected by downstream traffic conditions (e.g., spillback of a downstream queue). The inverse of the effective service rate (i.e., $1/\hat{\mu}_i$) is known as the expected effective service time. It is defined as the sum of: (i) the expected service time (term $1/\mu_i$), where μ_i is an exogenous flow capacity estimated based on attributes of the underlying lane (e.g., maximum speed), and of (ii) the expected blocked time. The latter represents the expected additional time a vehicle spends in queue i while waiting for a spillback to dissipate at a downstream queue. In queueing theory, spillbacks are refered to as blocking, hence the term expected blocked time. The expected blocked time is approximated by the product of: (i) the probability that a downstream link spills back (summation in the first parenthesis), and of (ii) the expected unblocking time (summation of the second partenthesis). The expected unblocking time represents the expected time it takes for a spillback at a downstream queue

to dissipate.

The study of Equation (4d) is the main focus of this paper. This equation yields a differentiable description of between-link interactions. In other words, Equation (4d) describes the impact on queue i of spillbacks from its downstream queues. It is through this equation that the above model provides a description of the spatial propagation of congestion.

2.3.2 Analytical model without between-link interactions

In order to investigate the need for using an analytical description of between-link interactions, we formulate a second analytical model, which differs from that of Section 2.3.1 in that it does not account for the between-link (i.e., between-queue) interactions. It is formulated as follows.

$$\begin{cases} \lambda_i^{\text{eff}} = \gamma_i(1 - P(N_i = \ell_i)) + \sum_j p_{ji}\lambda_j^{\text{eff}} & \text{(5a)} \\ P(N_i = \ell_i) = \dfrac{1-\rho_i}{1-\rho_i^{\ell_i+1}}\rho_i^{\ell_i} & \text{(5b)} \\ \rho_i = \dfrac{\lambda_i^{\text{eff}}}{(1 - P(N_i = \ell_i))\mu_i}. & \text{(5c)} \end{cases}$$

Equations (5a) and (5b) are identical to Equations (4a) and (4b), respectively. Equation (5c) differs from Equation (4c) in that the expected supply is represented by μ rather than by $\hat{\mu}$. Recall that $\hat{\mu}$ is defined as the effective service rate. It differs from μ in that is accounts for the impact of downstream spillbacks on the links flow capacity. If for a given queue, the spillback probability of all of its downstream queues is zero, then $\hat{\mu}$ equals μ. Otherwise, $\hat{\mu} < \mu$. In other words, the model described by the System of Equations (5) accounts for the within-link congestion propagation (through Equation (5b)), yet does not account for the between-link propagation.

This paper compares the performance of two SO algorithms. Both consider a metamodel of the form Equation (3). They differ in the choice of their analytical traffic model. The first uses an analytical traffic model that accounts for the between-link interactions (the model is defined by (4)). The second uses an analytical traffic model that does not account for the between-link interactions (the model is defined by (5)). All other algorithmic details are identical for both algorithms.

3 Signal control problem

Reviews of traffic signal control terminology are given in Lin (2011); Osorio (2010, Appendix A); Papageorgiou et al. (2003). The current approach to signal control in the QBB area for peak period

traffic is the use of fixed-time signal plans. Fixed-time signal plans are also known as time-of-day or pre-timed plans. They are designed offline for a given time period (e.g., morning peak period) based on model forecasts and historical traffic patterns. Unlike adaptive or traffic-responsive signal plans, they do not vary with real-time changes in demand. Fixed-time signal plans are periodic plans with a period or cycle time of typically 90 seconds or 120 seconds for each intersection.

The decision variables of our problem are known as the green splits of the signal phases, i.e., the ratio of green times to cycle times of the phases. All other traditional signal plan variables (e.g., cycle times, offsets, stage structure) are assumed fixed. We consider a problem where the signals of all 26 intersections are determined jointly.

In order to formulate the signal control 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;
$\mathscr{I}$	set of intersection indices;
$\mathcal{L}$	set of link indices;
$\mathscr{P}_I(i)$	set of phase indices of intersection i.

The optimization problem is formulated as follows.

$$\min_x f(x) = \sum_{l \in \mathscr{L}} E[N_l(x, z; p)] \tag{6}$$

subject to

$$\sum_{j \in \mathscr{P}_I(i)} x(j) = \frac{c_i - d_i}{c_i}, \ \forall i \in \mathscr{I} \tag{7}$$

$$x \geq x_L, \tag{8}$$

where x is the decision vector, $f(x)$ represents the (unknown) simulation-based objective function, N_l is the number of vehicles on link l, z represents the vector of endogenous variables of the simulation model (e.g., link densities, link speeds), p represents the vector of exogenous parameters of the simulation model (e.g., origin-destination demand matrix, network topology). The objective function is the expected total number of vehicles in the network, i.e., it is the sum over all links of the expected number of vehicles on each link.

The constraints of this problem are traditional signal control constraints. Intersection i has a fixed (i.e., exogenous) cycle time, denoted c_i. Certain signal phases within the cycle (e.g., all-red

phases) are considered fixed. The sum of the durations of these fixed phases is refered to as the fixed cycle time and is denoted d_i. Hence, the left-hand side of Constraint (7) denotes the proportion of the cycle time that is endogenous (i.e., the proportion that can be allocated to the endogenous signal phases). Constraint (7) ensures that for each intersection the green times sum to the total available cycle time. Lower bounds are used to ensure a minimal duration for all green phases (Equation (8)). These are set to 5 seconds.

Flow capacity of the signalized lanes. For the lanes that are signal controlled, their corresponding flow capacity (denoted μ in the macroscopic model formulation) is endogenous. For all other lanes, the flow capacity is considered exogenous and is estimated via the use of national transportation standards (VSS; 1999). For a given signalized lane i, its flow capacity (i.e., its service rate) is related to its green time by:

$$\mu_i - \sum_{j \in \mathscr{P}_L(i)} x(j)s = e_i s, \tag{9}$$

where $x(j)$ is the green split (i.e., ratio of green time to cycle time) of signal phase j, s denotes the saturation flow, e_i denotes the ratio of fixed green time to cycle time of signalized queue i and $\mathscr{P}_L(i)$ denotes the set of endogenous signal phase indices of queue i. In this paper s is set to 1800 vehicles per hour.

Analytical approximation of the objective function. Recall from Equation (3) that the component f_A of the metamodel is an approximation of the simulation-based objective function (f in Equation (1)) derived by the analytical macroscopic model. The objective function considered in this paper (Equation (6)) represents the sum across links of the expected number of vehicles per link. Since the analytical traffic model represents each lane of a link as a set of queues. Then, f_A can be rewritten as the sum across queues of the expected number of vehicles per queue: $\sum_{i \in \mathscr{Q}} E[N_i]$, where $\mathscr{Q}$ represents the set of queues. A simple closed-form expression for f_A is given by:

$$\sum_{i \in \mathscr{Q}} E[N_i] = \sum_{i \in \mathscr{Q}} \rho_i \left(\frac{1}{1-\rho_i} - \frac{(\ell_i + 1)\rho_i^{\ell_i}}{1 - \rho_i^{\ell_i + 1}} \right). \tag{10}$$

The derivation of the expected queue-length for a given queue, $E[N_i]$, is presented in Osorio and Chong (2015, Appendix A).

	Microscopic (simulation-based)	Macroscopic (analytical)	
		System (4)	System (5)
SO-Spill	√	√	
SO-No-Spill	√		√

Table 1: Traffic models used by the compared methods.

4 Case study

In this section we consider the QBB area mapped in the left plot of Figure 2 and the corresponding simulation model displayed in the right plot. We address the signal control problem (6)-(8). We compare the performance of two SO approaches. Both consider the algorithm of Osorio and Bierlaire (2013). They differ only in the choice of the macroscopic model used as part of the metamodel. The first approach uses the model defined by the System of Equations (4). As described in Section 2.3.1 this model accounts for the impact of downstream spillbacks (or blockings) on a lane's flow capacity. We refer to this approach as the *SO-Spill* approach. The second approach uses the model defined by the System (5), which differs from (4) only in that it does not account for the impact of downstream spillbacks. We refer to this approach as the *SO-No-Spill* approach. Table 1 summarizes which traffic models are used by each of the two compared approaches.

4.1 QBB network

The network model (right plot of Figure 2) consists of 134 roads, 313 lanes. 27 signalized intersections and 5 non-signalized intersections. The signal control problem determines the signals of 26 of the 27 signalized intersections. We consider the 8am-9am morning peak period. The expected demand, represented by a static origin-destination matrix, consists of over 11,500 car trips and over 750 truck trips. The trips are distributed across a set of 55 origin-destination pairs. The calibrated micrscopic simulation model is developed with the Aimsun microscopic simulation software (TSS; 2013).

Prior to running the SO algorithms, the exogenous parameters of the macroscopic models (Section 2.3) are calibrated based on microscopic model outputs. Of the 313 lanes 29 did not have any simulated flow along them across the simulation replications run for calibration. The remaining 284 lanes are modeled by one queue each. This leads to a qeueueing network model with a total of 284 queues.

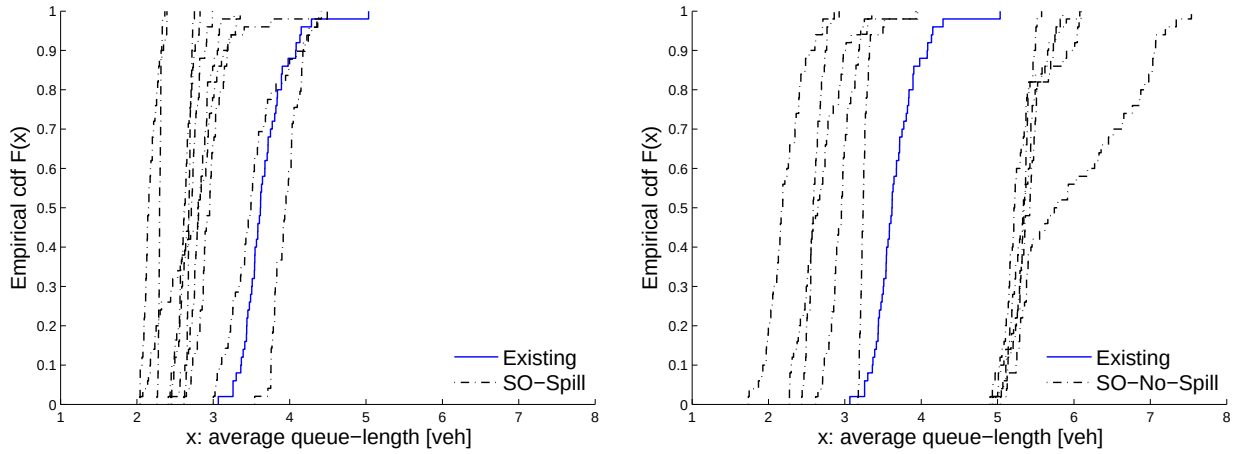
Across the 26 intersections of interest, there are a total of 64 endogenous signal phases (i.e., the decision vector is of dimension 64), which control the flow of a total of 120 queues. Each iteration

of the SO-Spill algorithm solves a trust region (TR) subproblem with 1604 endogenous variables, which consist of: 64 decision variables, 5 endogenous variables for each of the 284 queues and 120 endogenous service rate variables (μ) for each of the signalized queues. The TR subproblem is constrained by: 1136 nonlinear equality constraints (i.e., Equations (4b)-(4d) are implemented as 4 nonlinear equations for each of the 284 queues), 468 linear equality constraints (i.e., Equation (4a) for each of the 284 queues, Equation (9) for each of the 120 signalized queues, and Equation (7) for each decision variable), 1484 lower bound constraints (i.e., positivity constraints for the 5 queueing model endogenous variables for each of the 284 queues, and lower bounds for each of the 64 decision variables (Equation (8))), and 1 trust region nonlinear inequality constraint.

4.2 Comparison of SO-Spill with SO-No-Spill

This section compares the performance of the two SO algorithms: SO-Spill and SO-No-Spill. One run of a given SO algorithm involves calling the stochastic microscopic simulator, hence the outputs of an SO run are stochastic. Thus, in order to evaluate the performance of a given algorithm we run it 10 times with identical initial conditions. We initialize each run with the existing NYCDOT signal plan as the inital point. For a given algorithmic run, we allow for a total of 150 simulation calls, i.e., the simulation budget is set to 150. Each SO run yields as output a *proposed* signal plan. In summary, we run each SO algorithm 10 times to obtain 10 proposed signal plans. In order to evaluate the performance of a proposed signal plan, we run 50 simulation replications. For a given performance metric (e.g., average queue-length, throughput) we construct a cumulative distribution function (cdf) based on these 50 simulation observations. We then compare the distributions obtained from the different signal plans.

Figure 3(a) displays 11 curves. Each curve considers a given signal plan. Each curve displays the empirical cdf of the average queue-length of a given signal plan. The 10 dashed cdf's correspond to signal plans proposed by SO-Spill. The solid blue cdf corresponds to the existing NYCDOT signal plan. As described above, each curve is built based on 50 independent simulation observations. The x-axis is the average queue-length. This average is obtained as an average across all links of the average link queue-length. This average queue-length estimate is a scaled estimate of the objective function (Equation (6)), i.e. if we multiply the average queue-length by the total number of links in the network, then we obtain an estimate of the objective function. The y-axis is the cumulative probability. For example, consider a curve that passes through a given point $(x, y) = (3, 0.9)$. It can be interpreted as: for that signal plan 90% (i.e., a fraction of 0.9) of the observations have values smaller than 3. In other words, out of the 50 simulation observations of the average queue-lengths, 90% of them have values smaller than 3. Thus, for a given signal plan, the more its cdf is shifted to



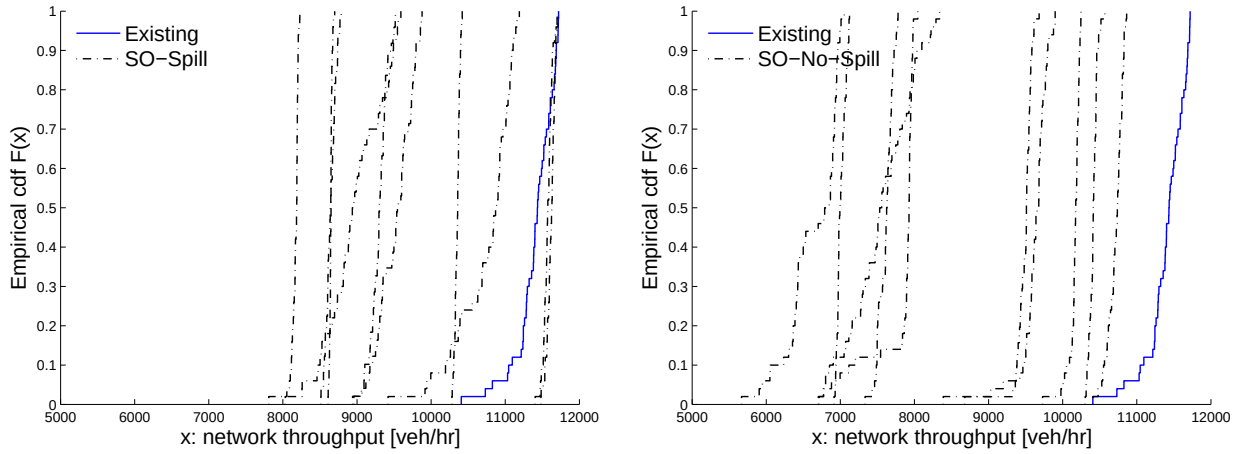
(a) (b)

Figure 3: Comparison of the average queue-lengths for the signal plans proposed by SO-Spill (left plot) and by SO-No-Spill (right plot).

the left, the higher the proportion of observations with low average queue-length observations, i.e., the better its performance. In Figure 3(a), 9 out of the 10 signal plans proposed by SO-Spill lead to smaller average queue-lengths than that of the existing signal plan. One plan performs worse than the existing plan.

Figure 3(b) also displays 11 curves: the 10 black dashed cdf curves correspond to the signal plans proposed by SO-No-Spill, the solid blue cdf corresponds to the existing NYCDOT signal plan. Figure 3(b) indicates that 5 out of the 10 plans proposed by SO-No-Spill have better performance than the existing plan, the remaining 5 plans have significantly worse performance than the existing plan. Figures 3(a) and 3(b) have the same x-axis range, and can be directly compared.

Based on NYCDOT guidance, the throughput of a given signal is also an important performance metric. Figure 4 analyses the performance of the proposed signal plans in terms of the network throughput. The latter is calculated as the number of finished trips by the end of the simulation period. For each plot of Figure 4, the x-axis is the network throughput, the y-axis is the cdf. Figure 4(a) (resp. Figure 4(b)) displays the throughput performance of the SO-Spill (resp. SO-No-Spill) signal plans. Since a higher throughput is indicative of better performance, the more these cdf curves are shifted to the right, the better the performance of the underlying signal plan. Figure 4(a) indicates that 2 out of the 10 signal plans proposed by SO-Spill lead to improved throughput compared to the existing signal plan, while the remaining 8 lead to lower throughput values. Figure 4(b) indicates that all 10 signal plans proposed by SO-No-Spill lead to lower throughput values when compared



(a) (b)

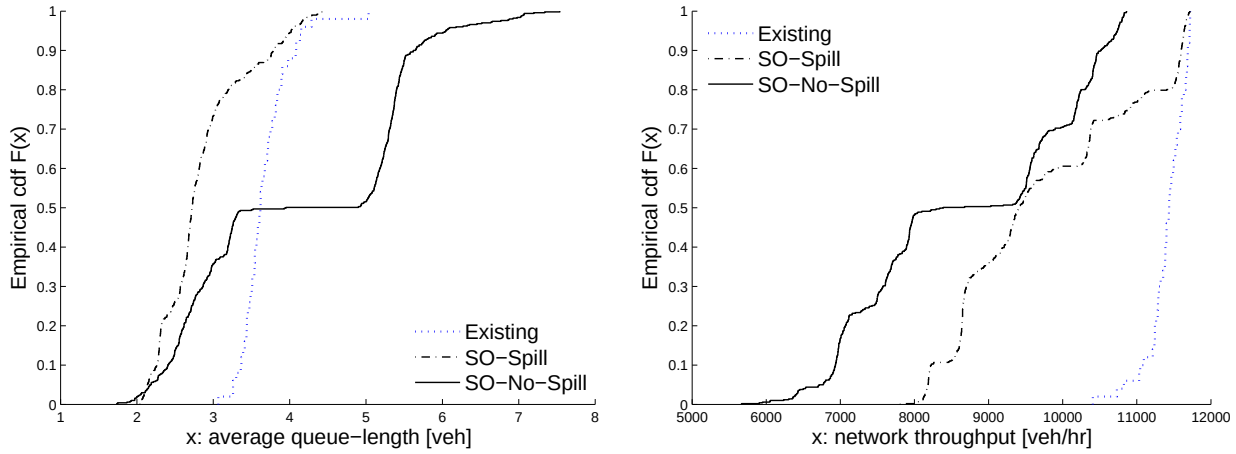
Figure 4: Comparison of the network throughput for the signal plans proposed by SO-Spill (left plot) and by SO-No-Spill (right plot).

to the existing signal plan. Figures 4(a) and 4(b) have the same x-axis range and can be directly compared.

The results of Figures 3(a) and 3(b) are aggregated in Figure 5(a). This figure displays the cdf of the existing NYCDOT plan (dotted blue curve). The results of all 10 signal plans proposed by SO-Spill are aggregated into a single cdf curve (dashed black curve). Those of SO-No-Spill are also aggregated into a single cdf curve (solid back curve). This figure illustrates that there is significant variability in the performance of the plans proposed by the SO-No-Spill algorithm. The signal plans proposed by SO-Spill lead to improved performance compared to the existing plan. Additionally, compared to the plans of SO-No-Spill, the SO-Spill plans have lower variability in performance.

We perform the same aggregate analysis for the network throughput. The results of Figures 4(a) and Figure 4(b) are aggregated in Figure 5(b). The variability of performance across signal plans is higher for the SO-No-Spill approach than for the SO-Spill approach. Nonetheless, for the throughput both approaches overall lead to signal plans with lower throughput values than the existing signal plan.

Based on criteria provided by NYCDOT, a proposed signal plan should improve the spatial propagation of congestion, yet should not deteriorate the network throughput. Hence, for a given SO algorithm, the "best" signal plan is selected by considering the plan with the lowest average queue-length that does not have worse network throughput than the existing NYCDOT plan. Based on this criteria none of the signal plans proposed by SO-No-Spill are suitable, since they all lead



(a) (b)

Figure 5: Aggregate comparison of the performance of the plans proposed by SO-Spill and SO-No-Spill: average queue-length (left plot) and network throughput (right plot).

to a deterioration of the network throughput. For the SO-No-Spill algorithm, we choose as "best" plan that with the lowest average queue-length, which happens to also be that with the highest average network throughput. For the selection of the "best" SO-Spill plan, we statistically test for throughput deterioration via paired t-tests. The tests assume a null hypothesis that states that the expected network throughput of the proposed signal plan equals that of the existing signal plan, and an alternative hypothesis that states that expected network throughput of the proposed plan is greater than that of the existing plan. Their significance level is set to 0.025.

A comparison of the performance of the "best" plans is displayed in Figure 6. Each plot of Figure 6 considers a specific performance measure. Figures 6(a)-6(d) consider, respectively, average link queue-length, network throughput, average trip travel time, and average lane spillback probability. The latter is calculated as the average across all lanes of the estimated lane spillback probability. For a given lane, its spillback probability is estimated as the proportion of time the lane is full. This proportion is derived from queue-length measurements obtained every 3 seconds. As discussed in Section 1, the spillback probability is an important performance metric for the QBB network. More generally, it is a suitable metric to quantify the spatial propagation of congestion.

Figure 6(a) indicates that the plans derived by both SO-Spill and SO-No-Spill lead to improved performance in terms of average queue-length compared to the existing plan. The SO-No-Spill plan outperforms the SO-Spill plan. In terms of network throughput, Figure 6(b) indicates that the SO-Spill plan outperforms the existing plan and the SO-No-Spill plan. The latter has worse performance

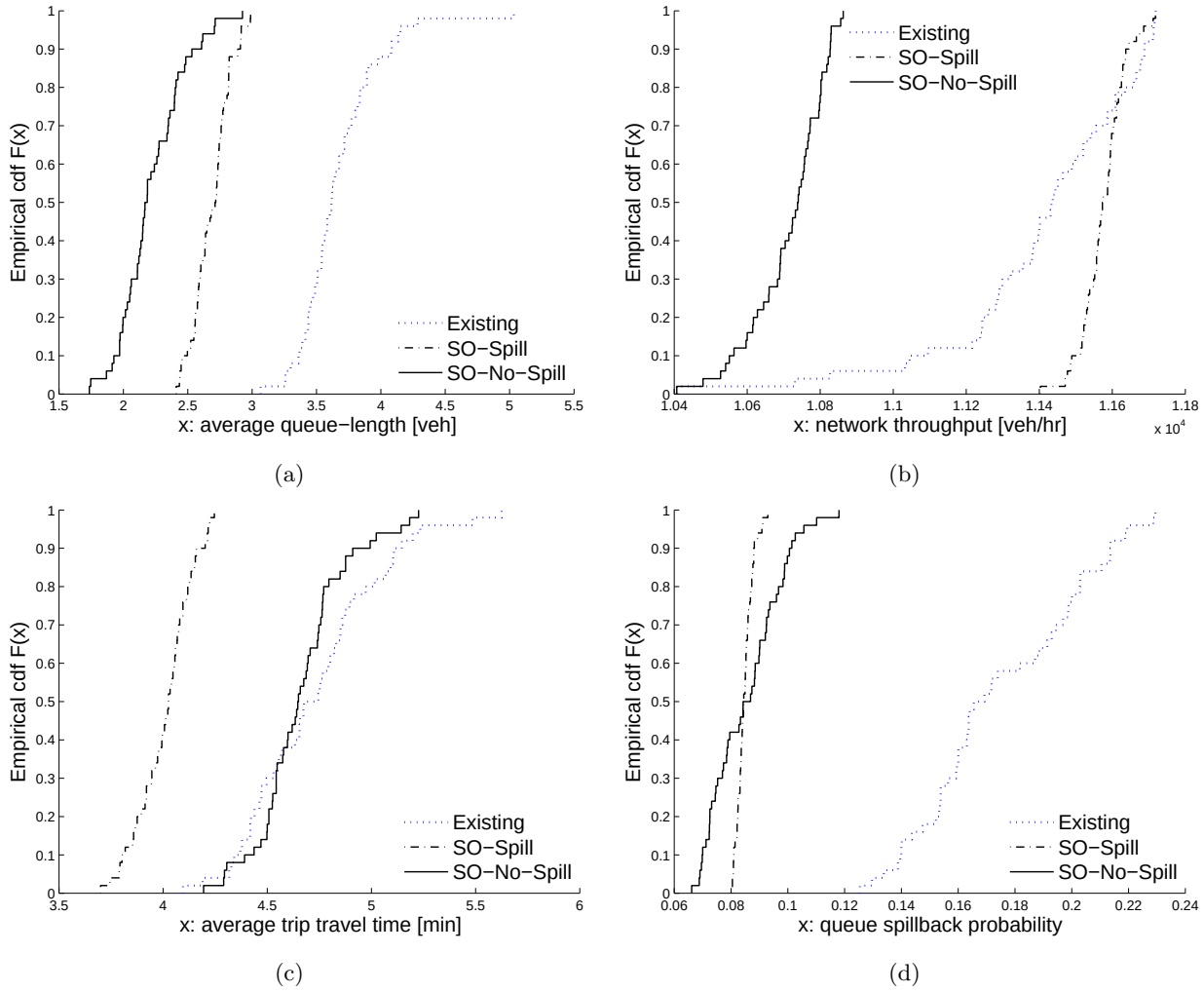


Figure 6: Comparison of the performance of the "best" plan proposed by SO-Spill and by SO-No-Spill.

than the existing plan. Figure 6(c) compares the performance in terms of average trip travel time. It indicates that the SO-Spill plan outperforms the existing and the SO-No-Spill plans. The latter has similar performance than the existing plan. In terms of average (across queues or lanes) spillback probabilities both SO plans have similar performance and outperform the existing plan. The SO-Spill plan has less variability than the SO-No-Spill plan.

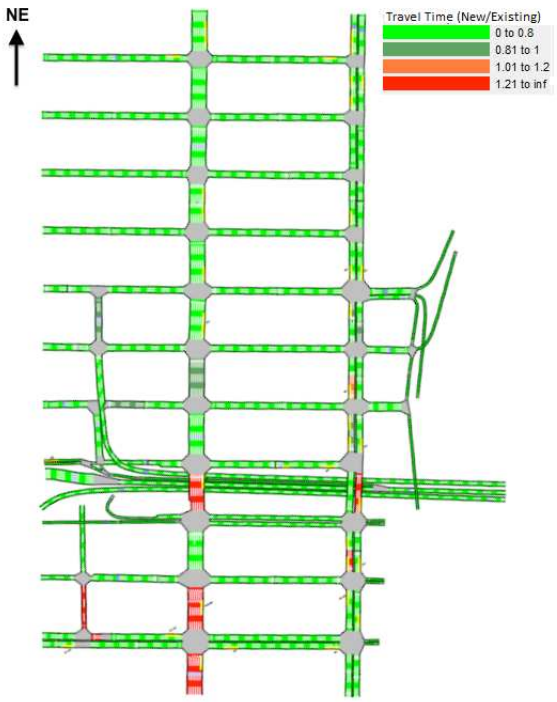
The results of this section show that the SO-Spill algorithm outperforms the SO-No-Spill algorithm. This indicates that there is a significant added value in providing the SO algorithm with an analytical description of the between-link interactions.

4.3 Spatial propagation of congestion

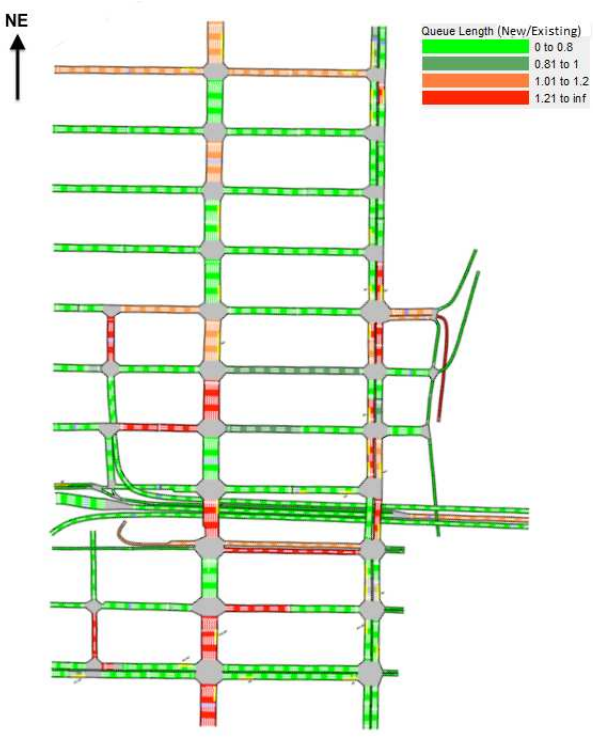
In this section, we analyse the impact of the SO-Spill plan on the spatial propagation of congestion. We carry out a more detailed comparison of the performance of the "best" SO-Spill plan and the existing plan. Figure 7(a) compares their performance across links. For a given signal plan (existing or SO-Spill), we estimate the average link travel time. This average is estimated as an average over both the entire simulation period (8am-9am) and over the 50 simulation replications. The links of Figure 7(a) are colored according to the ratio of average link travel time of the SO-Spill plan and the existing plan. Links are colored as followed: a decrease of average travel time (of the SO-Spill plan compared to the existing plan) of more than 20% corresponds to light green, a decrease within 0 and 20% corresponds to dark green, an increase between 0 and 20% corresponds to orange and an increase of more than 20% corresponds to red. Figure 7(a) shows that under the SO-Spill plan almost all links have an improvement in average link travel time of more than 20% compared to the existing signal plan.

Figure 7(b) is constructed just as Figure 7(a) but considers the average link queue-length. The majority of the links, and in particular almost all cross-streets, have significantly reduced queue-lengths. The SO-Spill plan leads to an improvement in average queue-lengths at the network level (the objective function) (as shown in Figure 6(a)) as well as an improvement at the link-level (as shown in Figure 7(b)). Hence, the SO-Spill plan achieves the initial goal of mitigating the spatial propoagation of congestion.

Let us now compare the signal plans in terms of the occurence of lane spillbacks. Both plots of Figure 8 compare for each lane the estimate of the spillback probability obtained under the existing signal plan (x-axis) and that obtained under the SO-Spill plan (y-axis). Figure 8(a) displays the estimates for all lanes of the network. The points that are under the diagonal (bottom right part of the plot), represent lanes with a reduced spillback probability under the SO-Spill plan. The spillback probability is reduced for 102 out of 284 lanes, it is increased for 30 lanes and remains equal for 152

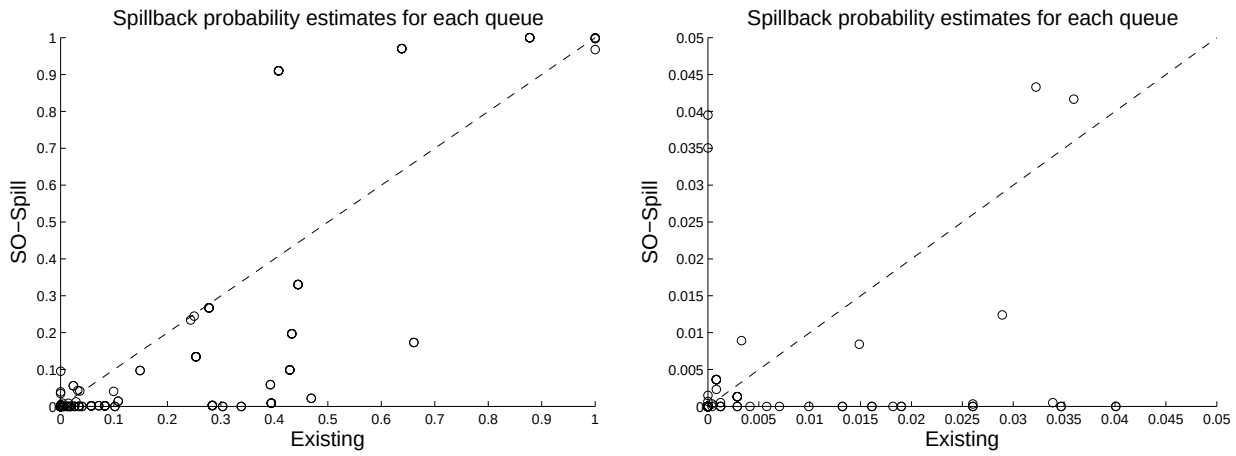


(a) Ratio of average link travel time



(b) Ratio of average link queue-length

Figure 7: Comparison of the spatial performance of the SO-Spill plan and the existing plan: ratio of average link travel times (left) and ratio of average link queue-lengths (right).



(a) (b)

Figure 8: Comparison of the lane spillback probabilities of the SO-Spill plan and of the existing plan.

lanes. The average difference of the spillback probability between the existing plan and the SO-Spill plan is 0.029, i.e., on average the SO-Spill plan leads to a 2.9% reduction in spillback occurence. Notice also that there are many lanes where the occurence of spillbacks is strictly positive under the existing plan yet is zero under the SO-Spill plan. This is the case for 10% of the lanes (29 out of the 284). This illustrates, once again, that the spatial propagation of congestion is mitigated under the SO-Spill plan. The design of such plans capable of mitigating the occurence of spillbacks is of great value for highly congested networks. Many points of Figure 8(a) are in the region $[0, 0.05] \times [0, 0.05]$. This region is displayed in more detail in Figure 8(b). The latter figure again indicates that for many lanes the spillback probability is set to zero under the SO-Spill plan.

Figure 9(a) compares for each lane the average queue-length under the existing plan (x-axis) and the SO-Spill plan (y-axis). Each point represents a lane. A more detailed plot of the range $[0, 5] \times [0, 5]$ is displayed in Figure 9(b). These figures indicate that overall the SO-Spill plan leads to a reduction at the lane-level of average queue-lengths. There is a reduction in 189 out of the 284 lanes. There is an increase for 88 lanes and no difference for 7 lanes. Figure 9(c) presents the cdf of the difference in average queue-lengths between the existing plan and the SO-Spill plan. For 24 of the 284 lanes, the difference is greater than 3 vehicles. For 9 it is greater than 5 vehicles. The largest difference is of the order of 12 vehicles, and is observed for 2 lanes.

At the network level, the SO-Spill plan leads to improvements of 26.8% in the average queue-length (2.69 versus 3.67 vehicles per link), 1.5% in the average network throughput (11576 versus 11405 vehicles), 24.2% in the average spillback probability (0.085 versus 0.112) and 13.3% in the

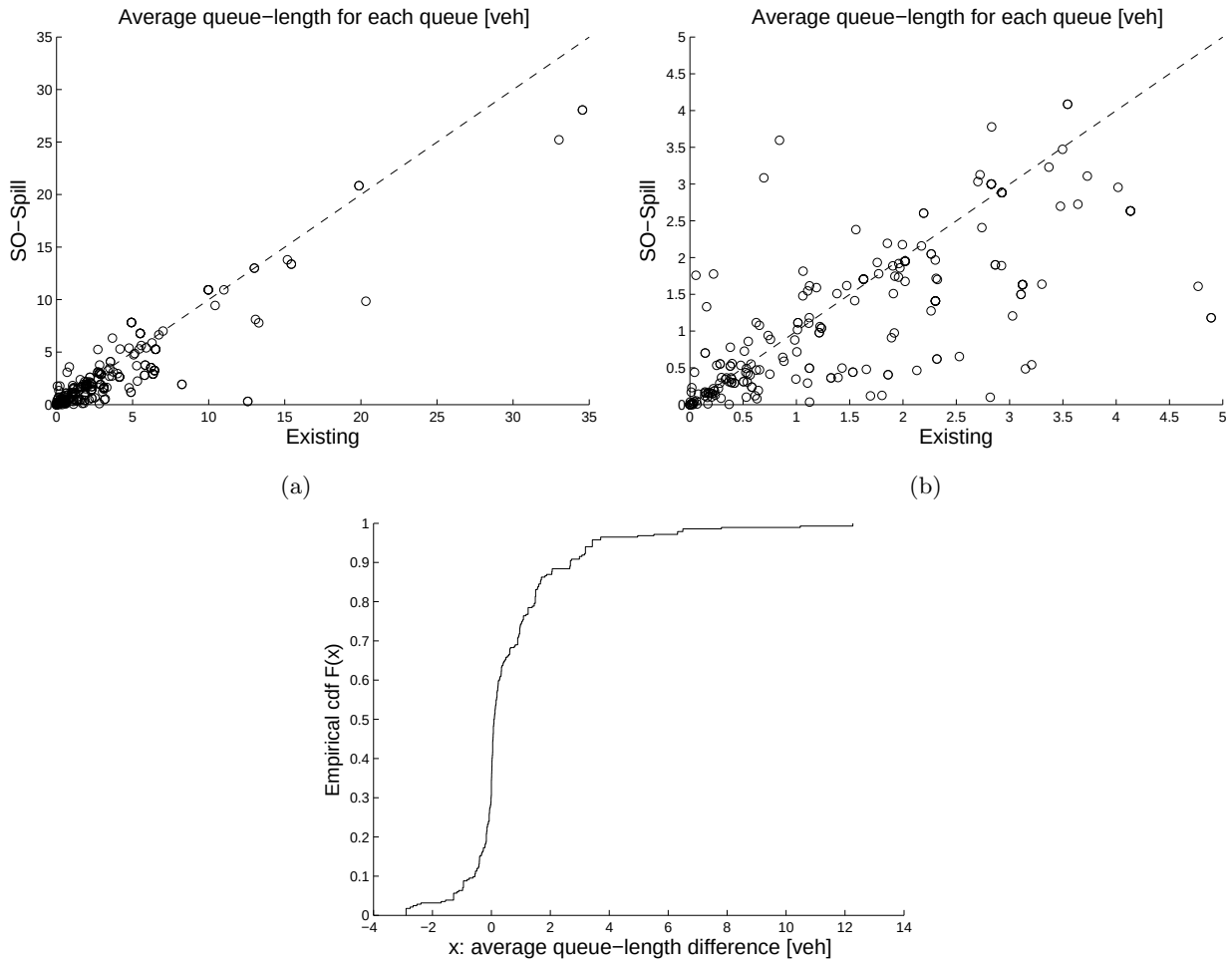


Figure 9: Comparison of the lane average queue-lengths of the SO-Spill plan and of the existing plan.

average trip travel time (4.1 versus 4.73 minutes).

The results of this section show that the plan proposed by the SO-Spill algorithm lead to significant improvements in performance compared to the existing plan, both at the network level and at the link level. The SO-Spill plan achieves the goal of limiting the spatial propagation of congestion within the complex QBB network.

5 Conclusions

This paper considers two SO algorithms that differ in the analytical information they provide to the algorithm. One provides analytical between-queue information, the other does not. The paper studies a signal control problem for a Manhattan network. The results show that for highly congested urban networks with intricate traffic patterns there is a significant added value in providing the algorithms with analytical between-queue information.

The development of analytical and differentiable traffic models that describe between-queue interactions is of interest to address a variety of network design and operations problems. Such information can be used at various steps of traditional SO algorithms (e.g., to define sampling distributions). Traditional analytical network models have focused on a detailed description of within-link vehicular interactions, there is limited analytical differentiable work that accounts for between-link interactions. Recent analytical work that addresses this challenge uses between-queue interaction information to approximate full path or full network distributions of the main performance measures (Osorio and Yamani; 2015; Flötteröd and Osorio; 2014; Osorio and Wang; 2012). In particular, to develop models that are suitable for large-scale network optimization, the work of Osorio and Yamani (2015); Osorio and Wang (2012) illustrate how link boundary conditions can be described aggregately, i.e., without the need of a detailed description of the within-link state. The complexity of such models is linear, rather than exponential, in the number of links in the network and is independent of the space capacity of the links. This makes them suitable for large-scale network optimization. The use of such detailed models to perform large-scale SO is a topic of ongoing research. The results of this fixed-time QBB analysis have also led to the extension of the SO-Spill algorithm to address traffic-responsive signal control problems (Chen et al.; 2015).

Acknowledgments

The work of C. Osorio was partially funded by National Science Foundation under Grant No.1351512. Any opinions, findings, and conclusions or recommendations expressed in this material are those of

the authors and do not necessarily reflect the views of the National Science Foundation. The work of X. Chen was partially funded by the Foundation for Science and Technology (FCT) through the MIT Portugal Program (PhD grant SFRH/BD/51296/2010) and by the European Regional Development Fund (FEDER) through the grant EMSURE - Energy and Mobility for Sustainable Regions (CENTRO-07-0224-FEDER-002004).

A SO algorithm

This SO algorithm is formulated in detail in Osorio and Bierlaire (2013) is based on the derivative-free trust region algorithm of Conn et al. (2009). The notation used is that of Osorio and Bierlaire (2013). The parameters of the algorithm are set according to the values in Osorio and Bierlaire (2013).

0. **Initialization.**

   Define for a given iteration k: $m_k(x, y; \alpha_k, \beta_k, q)$ as the metamodel (denoted hereafter as $m_k(x)$), x_k as the iterate, Δ_k as the trust region radius, $\nu_k = (\alpha_k, \beta_k)$ as the vector of parameters of m_k, n_k as the total number of simulation runs carried out up until and including iteration k, u_k as the number of successive trial points rejected, ε_k as the measure of stationarity (norm of the derivative of the Lagrangian function of the trust region (TR) subproblem with regards to the endogenous variables) evaluated at x_k.

   The constants $\eta_1, \gamma, \gamma_{inc}, \varepsilon_c, \bar{\tau}, \bar{d}, \bar{u}, \Delta_{max}$ are given such that: $0 < \eta_1 < 1$, $0 < \gamma < 1 < \gamma_{inc}$, $\varepsilon_c > 0$, $0 < \bar{\tau} < 1$, $0 < \bar{d} < \Delta_{max}$, $\bar{u} \in \mathbb{N}^*$. Set the total number of simulation runs permitted (across all points) n_{max}, this determines the computational budget. Set the number of simulation replications per point $\tilde{r}$ (here we use $\tilde{r} = 1$).

   Set $k = 0, n_0 = 1, u_0 = 0$. Determine x_0 and Δ_0 ($\Delta_0 \in (0, \Delta_{max}]$).

   Given the initial point x_0, compute $f_A(x_0)$ (analytical approximation of Equation (6)) and $\hat{f}(x_0)$ (simulated estimate of Equation (6)), fit an initial model m_0 (i.e., compute ν_0).

1. **Criticality step.** If $\varepsilon_k \leq \varepsilon_c$, then switch to *conservative mode*.

2. **Step calculation.** Compute a step s_k that reduces the model m_k and such that $x_k + s_k$ (the trial point) is in the trust region (i.e. approximately solve the TR subproblem).

3. **Acceptance of the trial point.** Compute $\hat{f}(x_k + s_k)$ and

$$\rho_k = \frac{\hat{f}(x_k) - \hat{f}(x_k + s_k)}{m_k(x_k) - m_k(x_k + s_k)}.$$

- If $\rho_k \geq \eta_1$, then accept the trial point: $x_{k+1} = x_k + s_k,\ u_k = 0$.
- Otherwise, reject the trial point: $x_{k+1} = x_k,\ u_k = u_k + 1$.

Include the new observation in the set of sampled points ($n_k = n_k + \tilde{r}$), and fit the new model m_{k+1}.

4. **Model improvement.** Compute $\tau_{k+1} = \frac{\|\nu_{k+1} - \nu_k\|}{\|\nu_k\|}$. If $\tau_{k+1} < \bar{\tau}$, then improve the model by simulating the performance of a new point x, which is uniformly drawn from the feasible space. Evaluate f_A and $\hat{f}$ at x. Include this new observation in the set of sampled points ($n_k = n_k + \tilde{r}$). Update m_{k+1}.

5. **Trust region radius update.**

$$\Delta_{k+1} = \begin{cases} \min\{\gamma_{inc}\Delta_k, \Delta_{max}\} & \text{if } \rho_k > \eta_1 \\ \max\{\gamma\Delta_k, \bar{d}\} & \text{if } \rho_k \leq \eta_1 \text{ and } u_k \geq \bar{u} \\ \Delta_k & \text{otherwise.} \end{cases}$$

If $\rho_k \leq \eta_1$ and $u_k \geq \bar{u}$, then set $u_k = 0$.

If $\Delta_{k+1} \leq \bar{d}$, then switch to *conservative mode*.

Set $n_{k+1} = n_k, u_{k+1} = u_k$, $k = k + 1$.

If $n_k < n_{max}$, then go to Step 1. Otherwise, stop.

References

Abu-Lebdeh, G. and Benekohal, R. (2003). Design and evaluation of dynamic traffic management strategies for congested conditions, *Transportation Research Part A* **37**(2): 109–127.

Barceló, J. (2010). *Fundamentals of traffic simulation*, Vol. 145 of *International Series in Operations Research and Management Science*, Springer, New York, USA.

Ben-Akiva, M., Cuneo, D., Hasan, M., Jha, M. and Yang, Q. (2003). Evaluation of freeway control using a microscopic simulation laboratory, *Transportation Research Part C* **11**(1): 29–50.

Bocharov, P. P., D'Apice, C., Pechinkin, A. V. and Salerno, S. (2004). *Queueing theory*, Modern Probability and Statistics, Brill Academic Publishers, Zeist, The Netherlands, chapter 3, pp. 96–98.

Branke, J., Goldate, P. and Prothmann, H. (2007). Actuated traffic signal optimization using evolutionary algorithms, *Proceedings of the 6th European Congress and Exhibition on Intelligent Transport Systems and Services*.

Bullock, D., Johnson, B., Wells, R. B., Kyte, M. and Li, Z. (2004). Hardware-in-the-loop simulation, *Transportation Research Part C* **12**(1): 73 – 89.

Chen, X., Osorio, C., Marsico, M., Talas, M., Gao, J. and Zhang, S. (2015). Simulation-based adaptive traffic signal control algorithm, *Transportation Research Board Annual Meeting*, Washington DC, USA.

Chen, X., Osorio, C. and Santos, B. F. (2012). A simulation-based approach to reliable signal control, *Proceedings of the International Symposium on Transportation Network Reliability (INSTR)*.

Chow, A. H. F. and Lo, H. K. (2007). Sensitivity analysis of signal control with physical queuing: Delay derivatives and an application, *Transportation Research Part B* **41**(4): 462–477.

Conn, A. R., Scheinberg, K. and Vicente, L. N. (2009). Global convergence of general derivative-free trust-region algorithms to first- and second-order critical points, *SIAM Journal on Optimization* **20**(1): 387–415.

Corthout, R., Flötteröd, G., Viti, F. and Tampere, C. (2012). Non-unique flows in macroscopic first-order intersection models, *Transportation Research Part B* **46**(3): 343–359.

Diakaki, C., Papageorgiou, M. and Aboudolas, K. (2002). A multivariable regulator approach to traffic-responsive network-wide signal control, *Control Engineering Practice* **10**: 183–195.

Flötteröd, G. and Rohde, J. (2011). Operational macroscopic modeling of complex urban intersections, *Transportation Research Part B* **45**(6): 903–922.

Flötteröd and Osorio, C. (2014). Stochastic analytical dynamic queueing network model with spillback, *International Symposium on Dynamic Traffic Assignment (DTA)*.

Geroliminis, N. and Skabardoni, A. (2011). Identification and analysis of queue spillovers in city street networks, *IEEE Transactions on Intelligent Transportation Systems* **12**(4): 1107 – 1115.

Gregoire, J., Qian, X., Frazzoli, E., de La Fortelle, A. and Wongpiromsarn, T. (2014). Capacity-aware back-pressure traffic signal control, *IEEE Trans. Control of Networked Systems*. forthcoming.

Hale, D. (2005). Traffic network study tool TRANSYT-7F, *Technical report*, McTrans Center in the University of Florida, Gainesville, Florida.

Heidemann, D. (1994). Queue length and delay distributions at traffic signals, *Transportation Research Part B* **28**(5): 377–389.

Heidemann, D. (1996). A queueing theory approach to speed-flow-density relationships, *Proceedings of the* 13^{th} *International Symposium on Transportation and Traffic Theory*, Lyon, France, pp. 103–118.

Heidemann, D. (2001). A queueing theory model of nonstationary traffic flow, *Transportation Science* **35**(4): 405–412.

Jain, R. and Smith, J. M. (1997). Modeling vehicular traffic flow using M/G/C/C state dependent queueing models, *Transportation Science* **31**(4): 324–336.

Joshi, S., Rathi, A. and Tew, J. (1995). An improved response surface methodology algorithm with an application to traffic signal optimization for urban networks, *in* C. Alexopoulos, K. Kang, W. R. Lilegdon and D. Goldsman (eds), *Proceedings of the 1995 Winter Simulation Conference*, pp. 1104–1109.

Lebacque, J. and Khoshyaran, M. (2005). First–order macroscopic traffic flow models: intersection modeling, network modeling, *in* H. Mahmassani (ed.), *Proceedings of the 16th International Symposium on Transportation and Traffic Theory*, Elsevier, Maryland, USA, pp. 365–386.

Li, P., Abbas, M., Pasupathy, R. and Head, L. (2010). Simulation-based optimization of maximum green setting under retrospective approximation framework, *Transportation Research Record* **2192**: 1–10.

Lin, S. (2011). *Efficient model predictive control for large-scale urban traffic networks*, PhD thesis, Delft University of Technology.

Lioris, J., Kurzhanskiy, A. A., Triantafyllos, D. and Varaiya, P. (2014). Control experiments for a network of signalized intersections using the '.Q' simulator, *IFAC/IEEE Workshop on Discrete Event Systems*, Cachan, France.

MapQuest.com, Inc (2015). New York City, NY, Scale undetermined; generated by Xiao Chen; using"MapQuest.com, Inc", `http://www.mapquest.com`. Accessed: 2015-01-02.

Michalopoulos, P. G. and Stephanopoulos, G. (1977a). Oversaturated signal systems with queue length constraintsi: Single intersection, *Transportation Research* **11**(6): 413 – 421.

Michalopoulos, P. G. and Stephanopoulos, G. (1977b). Oversaturated signal systems with queue length constraintsii: Systems of intersections, *Transportation Research* **11**(6): 423 – 428.

NYCDEP (2006). City Tunnel No. 3, Stage 2 Manhattan Leg Shaft 33B Project, http://www.nyc.gov/html/dep/pdf/shaft33b/4-10transit.pdf. Accessed: 2015-01-05.

NYCDOT (2010). Queensboro Bridge Bus Priority Study, http://www.nyc.gov/html/brt/downloads/pdf/201110_qbb_approach_summary.pdf. Accessed: 2015-01-05.

NYCDOT (2011). Queensboro bridge bus priority study, http://www.nyc.gov/html/brt/html/other/queensboro.shtml. Accessed: 2015-08-14.

NYCDOT (2012). 2012 New York City Bridge Traffic Volumes, http://www.nyc.gov/html/dot/downloads/pdf/bridge-traffic-report-2012.pdf. Accessed: 2015-03-1.

NYCDOT (2015). NYCDOT Bridges, http://www.nyc.gov/html/dot/html/infrastructure/bridges.shtml. Accessed: 2015-01-05.

Osorio, C. (2010). *Mitigating network congestion: analytical models, optimization methods and their applications*, PhD thesis, Ecole Polytechnique Fédérale de Lausanne.

Osorio, C. and Bierlaire, M. (2009). An analytic finite capacity queueing network model capturing the propagation of congestion and blocking, *European Journal of Operational Research* **196**(3): 996–1007.

Osorio, C. and Bierlaire, M. (2013). A simulation-based optimization framework for urban transportation problems, *Operations Research* **61**(6): 1333–1345.

Osorio, C. and Chong, L. (2015). A computationally efficient simulation-based optimization algorithm for large-scale urban transportation problems, *Transportation Science* **49**(3): 623–636.

Osorio, C. and Flötteröd, G. (2015). Capturing dependency among link boundaries in a stochastic network loading model, *Transportation Science* **49**(2): 420–431.

Osorio, C., Flötteröd, G. and Bierlaire, M. (2011). Dynamic network loading: a stochastic differentiable model that derives link state distributions, *Transportation Research Part B* **45**(9): 1410–1423.

Osorio, C. and Nanduri, K. (2015a). Energy-efficient urban traffic management: a microscopic simulation-based approach, *Transportation Science* **49**(3): 637–651.

Osorio, C. and Nanduri, K. (2015b). Urban transportation emissions mitigation: coupling high-resolution vehicular emissions and traffic models for traffic signal optimization, *Transportation Research Part B* . Forthcoming. Available at: http://web.mit.edu/osorioc/www/papers/osoNanEmissionsSO.pdf.

Osorio, C. and Selvam, K. (2015). Simulation-based optimization: achieving computational efficiency through the use of multiple simulators, *Transportation Science* . Forthcoming. Available at: http://web.mit.edu/osorioc/www/papers/osoSelMultiModel.pdf.

Osorio, C. and Wang, C. (2012). An analytical approximation of the joint distribution of queue-lengths in an urban network, *Procedia Social and Behavioral Sciences. Papers selected for the 15th meeting of the EURO Working Group on Transportation*, Vol. 54, pp. 917–925.

Osorio, C. and Yamani, J. (2015). Analytical and scalable analysis of transient tandem markovian finite capacity queueing networks, *Transportation Science* . Forthcoming. Available at: http://web.mit.edu/osorioc/www/papers/osoYamDynAggDisagg.pdf.

Papageorgiou, M., Diakaki, C., Dinopoulou, V., Kotsialos, A. and Wang, Y. (2003). Review of road traffic control strategies, *Proceedings of the IEEE* **91**(12): 2043–2067.

Papageorgiou, M. and Varaiya, P. (2009). Link vehicle-count - the missing measurement for traffic control, *in* A. Chassiakos (ed.), *Proceedings of the 12th IFAC Symposium on Control in Transportation Systems*, Redondo Beach CA, USA.

Park, B. B. and Kamarajugadda, A. (2009). Stochastic optimization for sustainable traffic signal control, *International Journal of Sustainable Transportation* **1**(3): 193–207.

Rathi, A. K. (1988). A control scheme for high traffic density sectors, *Transportation Research Part B* **22B**(2): 81–101.

Skabardonis, A. . and Geroliminis, N. (2008). Real-time monitoring and control on signalized arterials, *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations* **12**(2): 6474.

Spall, J. C. and Chin, D. C. (1997). Traffic-responsive signal timing for system-wide traffic control, *Transportation Research Part C* **5**(3-4): 153–163.

Stevanovic, A., Stevanovic, J., Zhang, K. and Batterman, S. (2009). Optimizing traffic control to reduce fuel consumption and vehicular emissions, *Transportation Research Record* **2128**: 105–113.

Stevanovic, J., Stevanovic, A., Martin, P. T. and Bauer, T. (2008). Stochastic optimization of traffic control and transit priority settings in VISSIM, *Transportation Research Part C* **16**(3): 332 – 349.

Tampère, C., Corthout, R., Cattrysse, D. and Immers, L. (2011). A generic class of first order node models for dynamic macroscopic simulations of traffic flows, *Transportation Research Part B* **45**(1): 289–309.

Tanner, J. C. (1962). A theoretical analysis of delays at an uncontrolled intersection, *Biometrika* **49**: 163–170.

TSS (2013). *AIMSUN 7 Dynamic Simulators User's Manual*, Transport Simulation Systems.

Varaiya, P. (2013). Max pressure control of a network of signalized intersections, *Transportation Research Part C* **36**: 177–195.

VSS (1999). *Norme Suisse SN 640023 Capacité, niveau de service, charges compatibles; carrefours avec feux de circulation*, Union des professionnels suisses de la route, VSS, Zurich.

Wongpiromsarn, T., Uthaicharoenpong, T., Wang, Y., Frazzoli, E. and Wang, D. W. (2012). Distributed traffic signal control for maximum network throughput, *IEEE Conference on Intelligent Transportation Systems*, p. 588595.

Yun, I. and Park, B. (2006). Application of stochastic optimization method for an urban corridor, *Proceedings of the Winter Simulation Conference*, pp. 1493–1499.