Efficient Max-Pressure Traffic Signal Control for Large-Scale Congested Urban Networks

Dimitrios Tsitsokas

Anastasios Kouvelas

Nikolas Geroliminis

École Polytechnique Fédérale de Lausanne **September 2021**

STRC | **21st Swiss Transport Research Conference**
Monte Verità / Ascona, September 12 – 14, 2021

École Polytechnique Fédérale de Lausanne

Efficient Max-Pressure Traffic Signal Control for Large-Scale Congested Urban Networks

Dimitrios Tsitsokas, Nikolas Geroliminis
Urban Transport Systems Laboratory
École Polytechnique Fédérale de Lausanne
GC C2 390, Station 18, 1015 Lausanne, Switzerland
phone: +41-21-69-32481
fax: +41-21-69-32431
{dimitrios.tsitsokas, nikolas.geroliminis}@epfl.ch

Anastasios Kouvelas
Institute for Transport Planning and Systems
ETH Zurich
HIL F 37.2, Stefano-Franscini-Platz 5, 8093 Zurich, Switzerland
phone: +41-44-633 66 95
fax: +41-44-633 66 95
anastasios.kouvelas@ivt.baug.ethz.ch

September 2021

Abstract

Traffic responsive signal control systems bear high potential in reducing delays in congested networks due to their ability of dynamically adjusting right-of-way assignment among conflicting movements, based on real-time traffic measurements. In this work, we focus on distributed traffic signal control for large-scale networks based on the existing Max-Pressure controller, which has been shown to stabilize queues and maximize throughput in congested conditions. Max-Pressure constitutes a feedback control law that tries to balance queues around an intersection by updating green times between signal stages as a function of current queue measurements. Nevertheless, its increased infrastructure requirements impose high implementation costs. Our objective is to investigate how network performance changes when controller is installed only in subsets, (instead of all) of network nodes, while exploring strategies of identifying the most critical nodes. A modified version of Store-and-Forward traffic model is used to emulate spatio-temporal traffic evolution in a large-scale network and evaluate system performance for different controller layouts. Firstly, we observe significant improvement in terms of total delay and network MFD production when Max-Pressure control is applied. More than 85% of the improvement observed when controlling all network nodes can be achieved by controlling only 25% of properly selected nodes, thus reducing implementation costs to one fourth. Further research is needed in order to optimize node selection for the Max-Pressure layout, through evaluation of node impact to network performance. Moreover, investigating the potential of further gains via combining Max-Pressure with centralized control strategies, e.g. perimeter control, is a promising research direction.

Keywords

Decentralized/Distributed Control; Adaptive Traffic Signal Control; Max-Pressure; Back-Pressure; Dynamic Traffic Simulation; Store-and-Forward (SaF); Large-Scale Networks

Introduction

Continuing urbanization and population growth supported by the modern way of living unavoidably intensify congestion problems in metropolitan regions, as a result of the increasing difference between travel demand and network capacity, which leads to excessive delays during peak periods. Given that increasing network capacity space-wise is often unrealistic in city centers, reaching the maximum operating capacity by efficient traffic distribution while preventing excessive demand from causing instability and gridlocks is crucial for maintaining mobility in over-saturated conditions. Traffic-responsive signal control systems carry significant potential in reaching this objective due to their ability of dynamically adjusting green/red light splits between competing approaches based on the prevailing traffic conditions, by following an intuitive control logic. In contrast to static (fixed-time) signal control strategies, which function according to a pre-determined time-periodic schedule designed to serve a specific demand scenario, traffic-responsive control systems are more capable in efficiently handling over-saturated conditions due to their agility in updating traffic signal plans of isolated or coordinated intersections according to actual traffic information provided in real-time, thus achieving more efficient green time utilization.

Several types of traffic-responsive signal control systems have been proposed and applied in the field, following different optimization methods, as well as modeling and control approaches. Some of the most commonly used are SCOOT (Hunt *et al.,* 1981), OPAC (Gartner, 1983), PRODYN (Henry *et al.,* 1984), SCATS (Lowrie, 1990), UTOPIA (Mauro and Di Taranto, 1990), and the more recent ones RHODES (Mirchandani and Head, 2001) and TUC (Diakaki *et al.,* 2002). These systems employ a centralized control logic, in the sense that traffic information from the entire controlled region is required to be transferred to a central processing unit, where the respective control algorithm will process it and determine the control actions that need to be taken and communicated back to every single intersection. This centralized architecture, although it constitutes the state-of-the-art in the field of adaptive signal control due to higher number of degrees of freedom, it is characterized by low applicability. This is mainly due to its requirements in communication and computing infrastructure, which imposes a high installation, maintenance and operational cost, and also to its relatively low scalability, in the sense that it is difficult to install gradually or expand an existing system to cover a larger region. On the contrary, decentralized approaches, based on local control of isolated or coupled intersections, have been developed and continue to attract interest due not only to their significantly lower infrastructure requirements and scalability, but also to their high potential in improving the network performance even in over-saturated conditions. An overview of the different aspects of centralized and decentralized control strategies is found in Chow *et al.* (2020).

This work focuses on Max-Pressure (MP) decentralized feedback controller (also referred to as Back-Pressure), which is applied to isolated intersections. It is based on a simple algorithm that modifies the right-of-way assignment between competing approaches based on feedback information of queues forming in roads upstream and downstream of the intersection, in a periodic cyclic process. MP, initially proposed for packet scheduling in wireless communication networks by Tassiulas and Ephremides (1990), it was formulated as a signalized intersection controller through the works of Varaiya (2013a), Varaiya (2013b), Wongpiromsarn *et al.* (2012), Zhang *et al.* (2012) and was theoretically proven by Varaiya (2013a) to stabilize demand and lead to maximum network throughput under specific assumptions. This stability guarantee, together with MP's independence from any a priori knowledge of traffic demand and its decentralized nature, render MP a promising and easily applicable control strategy for signalized networks facing increased congestion. This potential has motivated a significant amount of research focusing on MP controller, with some recent examples being the works of Gregoire *et al.* (2014), Kouvelas *et al.* (2014), Le *et al.* (2015), Manolis *et al.* (2018), Mercader *et al.* (2020). Through these works, several improved versions of the initial formulation have been proposed in an effort to improve its performance (e.g. using normalized pressure in Gregoire *et al.* (2014) or strict cyclic phase policy in Le *et al.* (2015) and Kouvelas *et al.* (2014)) or overpass difficulties related to requirements of turning-ratio knowledge (Le *et al.* (2015)), the assumption of incapacitated queues or the queue monitoring method. In Mercader *et al.* (2020), pressure is calculated by using travel time estimation instead of queue length, in an attempt of relaxing the need for expensive queue measuring equipment.

Nevertheless, under heavily congested conditions and fast growing queues, local traffic signal control might not operate as successfully. Perimeter control (see for example Kouvelas *et al.* (2017) or Sirmatel *et al.* (2021)) can prevent over-saturation in regions of the city experiencing high demand. Investigating the limits of traffic responsive local strategies is still a challenge. Moreover, little research interest has been expressed regarding layout efficiency of MP-based decentralized systems, with most works assuming MP installation and operation to all network intersections. Even though MP does not require the highly costly infrastructure of centralized systems, it still needs the equipment to measure queues and estimate turning movements inside all incoming and outgoing roads of every intersection in order to function properly. In large networks, installing this type of equipment in all intersections might be too expensive and is probably inefficient, since there is a high chance that not all intersections are congested enough or experience highly imbalanced queues, so that they the can fully benefit from an MP controller and at the same time considerably improve the global network performance.

In this work, we propose and evaluate a classification method for selecting the critical intersections to include in a network-wide MP controller layout while aiming at a high benefit-to-cost ratio. We construct and evaluate different MP layouts for a large-scale traffic network of a real city. A modified version of the Store-and-Forward paradigm (see Aboudolas *et al.* (2009)), with characteristics of the S-Model (see Lin *et al.* (2011)) is used to emulate traffic in the network. Its mathematical structure enables proper modeling of spill-backs and allows dynamic updates of traffic signal settings (e.g. see Tsitsokas *et al.* (2021)). MP controller is considered to operate at different subsets of network nodes. We evaluate two MP node layouts resulting from the described classification process, containing only a fraction of the network nodes, and compare their traffic performance to the fixed-time control case, the typical case of holistic implementation of MP to all network nodes, and a set of randomly constructed MP node layouts containing different fractions of network nodes. We perform a detailed analysis of the network behavior for all above scenarios, and observe significant performance improvement in all cases of MP control. In parallel, results show that a properly selected fraction of network nodes with MP controller is capable of leading to comparable performance gain as the much more costly case of MP controller installed to all network nodes.

The rest of this paper is structured as follows. At first, the theoretical framework of the MP controller is briefly introduced in the next section, together with the required notation explanation. The problem of specifying feasible updated green times from pressures while respecting several types of constraints is then formulated as a quadratic program and the characteristics of the traffic simulation model are briefly discussed. Afterwards, the experiment setup is described and the methodological framework of node classification and selection is presented. Simulation results and analysis follow, together with an analysis of the system's behaviour for the different scenarios, compared to the fixed-time control case. Finally, the paper is concluded by summarizing the main findings and discussing possible future research.

1 Methodological Framework

In this section, the theoretical and mathematical description of the distinct elements used in this work are presented. Firstly, the MP decentralized controller for signalized network intersections is described in detail and the necessary notation is introduced. The process for recalculation of green time duration for all phases of the signalized intersection based on the computed pressures while respecting several constraints that ensure the applicability (e.g. integer green times), stability (e.g. gradual changes between consecutive cycles) and maximum waiting time for all vehicles (e.g. minimum green time per phase) is formulated as a quadratic programming problem. Afterwards, an intuitive method to identify potentially critical intersections for installing MP controllers is introduced. Finally, we briefly discuss the main properties of the utilized traffic simulation model.

1.1 The Max-Pressure Controller

The MP traffic controller as presented by Varaiya (2013a) is a feedback-based signal control algorithm that has been proven to stabilize queues in a controllable network, resulting in maximum throughput, on condition that a set of simplifying assumptions apply. It is also distributed and scalable, in the sense that it requires no communication with the rest of the network and can be introduced in an isolated intersection independently of what happens to the rest of the network nodes. The controller does not require any knowledge of the actual or expected traffic demand, in contrast to other model predictive control approaches. Its operation only requires real-time queue measurements in the links around the intersection and knowledge of turning ratios of all alternative approaches that traverse the intersection. Several different versions of the controller have been introduced. The present version is similar to the one in Kouvelas *et al.* (2014). The detailed description of MP algorithm together with the necessary notation is given below.

The traffic network is represented as a directed graph (N, Z) consisting of a set links $z \in Z$ and a set of nodes $n \in N$. At any signalized intersection n, I_n and O_n denote the set of incoming and outgoing links, respectively. The cycle time C_n and offset, which enables coordination with the neighboring intersections, are pre-defined (or calculated online by a different algorithm) and are not modified by MP. Intersection n is controlled on the basis of a pre-timed signal plan (including the fixed total lost time L_n), which defines the sequence, configuration and initial timing of a fixed number of phases that belong to set F_n. During activation of each phase $j \in F_n$, a set of non-conflicting approaches v_j (i.e. connections between pairs of incoming-outgoing links of the node) get right-of-way (green light) simultaneously. The saturation flow of every link z, denoted as S_z, refers to the maximum possible flow that can be transferred to downstream links, depending on link and intersection geometry. The turning ratio of an approach between links $i - w$, where $i \in I_n, w \in O_n$ is denoted as $\beta_{i,w}$ and refers to the fraction of the outflow of upstream link i that will move to downstream link w. The present version of MP assumes that turning ratios are known to the controller. However, it has been shown that control effectiveness is not deteriorated if turning ratios are estimated (see Le *et al.* (2015)). By definition, the following relation stands for every node n,

$$\sum_{j \in F_n} g_{n,j}(k_n) + L_n = (\text{or} \leq)\ C_n \tag{1}$$

where $k_n = 1, 2, \ldots$ is the discrete-time control cycle index, and $g_{n,j}$ denotes the green time duration of phase j at cycle k_n. The inequality may apply in cases where long all-red phases are imposed for any reason (e.g. gating).

The version of MP controller employed in this work assumes that phase sequencing is given as input (e.g. from pre-timed scheduling) and does not change during the control. This means that during every cycle, all phases will be activated for a minimum time in the same ordered sequence. As a result, the following constraint also applies for every node n,

$$g_{n,j}(k_n) \geq g_{n,j,\min},\; j \in F_n \tag{2}$$

where, $g_{n,j,\min}$ is the minimum green time required for phase j of node n, which often matches the required amount of time for the respective pedestrian movements. The control variables of this problem are $g_{n,j}(k)$ and correspond to the effective green time of every stage $j \in F_n$ of intersection $n \in N$. The network state is depicted by the number of vehicles $x_z(k)$ that are waiting to be served in every link z for every control cycle k (i.e. at the start of cycle $[kC_n, (k+1)C_n]$). Assuming that real-time measurements or estimates of the queue lengths (states) and turning ratios of all controlled intersections are available, the pressure $p_z(k)$ of every incoming link $z \in I_n$ of node n for the control cycle k is computed as

$$p_z(k) = \left[\frac{x_z(k)}{c_z} - \sum_{w \in O_n} \frac{\beta_{z,w} x_w(k)}{c_w} \right] S_z, \quad z \in I_n \tag{3}$$

In equation (3), x_z and c_z denote the current number of vehicles and the storage capacity (maximum number of vehicles) of link z, respectively. The normalization of queues by dividing by the storage capacity aims at considering the link length in the calculation, so that pressure of a shorter link is higher than that of a longer one with the same number of vehicles in it. This means that pressure contains the probability of a queue to spill-back in the following cycle. Pressures are calculated at the end of every cycle based on the latest queue measurements, which consist the state feedback variables. Then, pressures are used by the controller for the signal settings update. It should be noted that the above formulation assumes that flow transfer is always possible between z and all $w \in O_n$ links of node n. Otherwise, the second term includes only downstream links for which flow transfer is allowed from z. Also, x_z is not necessarily the instantaneous queue length of z at the end of the cycle. In different MP versions, it may also refer to the mean or the maximum observed queue length, during the previous cycle. In this work, a preliminary analysis showed that mean values of queue lengths are generating better results and this version was adopted. By looking at the two terms, one can notice that essentially, the pressure depicts the difference between upstream and downstream link occupancies. Therefore, higher pressure indicates higher potential in traffic production, i.e. several vehicles waiting to be served and enough available space in downstream links to receive them. Low or close to zero pressure indicates lower need for right-of-way, either due to small queue upstream, or due to lack of space (large queue) downstream. We should note that negative pressures are meaningless here, so the constraint $p_z(k) \geq 0$ must always hold.

Equation 3 is applied for all $z \in I_n$ and pressures for all incoming links of node n are computed. Then, the pressure corresponding to each stage j is defined as the sum of the pressures of all links belonging to the stage, as follows.

$$P_{n,j}(k) = \max\left\{0, \sum_{z \in v_j} p_z(k)\right\}, \quad j \in F_n \tag{4}$$

This metric is then used as weight for the distribution of the total available green time between the competing stages of the intersection.

1.2 Green time calculation

After pressure values $P_{n,j}$ are available for every phase $j \in F_n$ of intersection n, the total amount of effective green time,

$$G_n = C_n - L_n = \sum_{j \in F_n} g^{\star}_{n,j}, \quad n \in N \tag{5}$$

is distributed to the phases of node n according to pressure values. In the above equation (5), $g^{\star}_{n,j}$ denotes the green time of phase j at intersection n that is calculated by static fixed-time analysis, using any of the standard algorithms. It holds that $g^{\star}_{n,j} \geq g_{n,j,\min}, \forall j \in F_n$.

There are several different approaches that have been proposed regarding green time calculation, some of which also include phase activation based on pressures. In this version, since phases are activated in a strictly defined and non-changing order with a guaranteed minimum green time, green duration is assigned to phases proportionately to the computed pressures, as follows

$$\tilde{g}_{n,j}(k) = \frac{P_j(k)}{\sum_{i \in F_n} P_i(k)} G_n, \quad j \in F_n \tag{6}$$

Eq. (6) provides the raw green times calculated according to MP control logic. However, these values cannot be included directly to the intersection signal plan, because it must first be guaranteed that they comply with a set of necessary constraints. Therefore, an additional step is added in the process, whose objective is to translate Max Pressure outputs of eq. (6) to practically applicable green times $G_{i,j}$. This is done by solving online, for every node n and control cycle k, the following optimization problem (also addressed in Diakaki *et al.* (2002)):

$$
\begin{aligned}
\underset{G_{n,j}}{\text{minimize}} \quad & \sum_{j \in F_n} (\tilde{g}_{n,j} - G_{n,j})^2 \\
\text{subject to} \quad & \sum_{j \in F_n} G_{n,j} + L_n = C_n \\
& G_{n,j} \geq g_{n,j,\min}, \;\; j \in F_n \\
& \left| G_{n,j} - G_{n,j}^{p} \right| \leq g_{n,j}^{R} \\
& G_{n,j} \in \mathbb{Z}^{+} \\
& \forall j \in F_n
\end{aligned}
\tag{7}
$$

According to the above formulation, the applicable green times for every phase $G_{i,j}$, $j \in F_n$, should be as close to the non-feasible regulator-defined greens $\tilde{g}_{n,j}$ as possible, while satisfying a set of constraints. The first constraint states that eq. (1) must always hold, therefore the sum of the updated feasible green times plus the total lost time L_n should be equal to cycle C_n. The second constraint ensures that all phases get a predefined minimum green duration $g_{n,j,\min}$. In order to avoid potential instability of the system due to large changes in the signal timing happening too fast, we impose a threshold to maximum absolute change of every phase duration between consecutive cycles. This is expressed in the third constraint, where $G_{n,j}^{p}$ denotes the applied green times of the previous cycle and $g_{n,j}^{R}$ is the maximum allowed change of the duration of phase j between consecutive cycles. Finally, feasible green times must belong to the positive integers set. This type of integer quadratic-programming problem can easily be solved by any commercial solver fast enough to allow online solution for every control cycle. The solution of this optimization problem, i.e. variables $G_{n,j}$, $\forall j \in F_n$, are the new feasible phase durations for node n which will be applied in the next control cycle. The above process is repeated at the end of the cycle for every controlled intersection, regardless of what is happening to the rest of the network. The controller only requires real-time queue information of the adjacent intersections and respective turning ratios and the algorithm is executed once per cycle.

1.3 Intersection selection

The purpose of this study is to investigate the effectiveness of a network-wide application of decentralized MP control when only a fraction of the network signalized intersections are controlled, which would drastically decrease the implementation cost of this strategy. For this purpose, we describe here a classification method to assess the importance of each node in terms of MP control. Thinking that MP aims at stabilizing and balancing the queues around an intersection, an intuitive hypothesis is that controlling intersections experiencing high levels of congestion and high variance of queues forming at surrounding links would benefit more of an MP controller. On this basis, we propose the following node selection process:

Step 1: A full simulation run of the network is executed for the case of the fixed-time signal control settings and peak-period results are extracted.

Step 2: For every network node n and for the peak-period, the number of control cycles during which n is congested is counted. A node n is considered congested at control cycle k if the mean queue of any incoming link $z \in I_n$ is higher than 80% of the links' storage capacity, i.e. if:

$$\frac{1}{t_k} \sum_{i=(k-1)t_k+1}^{kt_k} x_z(i) \geq 0.8c_z \tag{8}$$

for any $z \in I_n$, where t_k is the number of discrete time-steps of the traffic model that compose one cycle, k is the cycle index and c_z the storage capacity of link z.

Step 3: Only nodes that are congested for more than 85% of the defined peak period are candidates for MP controller. The rest are removed from the selection pool.

Step 4: For every remaining node n, we calculate two metrics: m_1^n which expresses the average congestion experienced by node n during peak-period, computed as

$$m_1^n = \frac{1}{t_p} \frac{1}{\|I_n\|} \sum_{i \in T_P} \sum_{z \in I_n} \frac{x_z(i)}{c_z} \tag{9}$$

where T_P is the set of time-step indices corresponding to the peak period and t_p is the size of this set, i.e. $t_p = \|T_P\|$, and m_2^n, which expresses the average variance of the queues forming around a node during peak-period, computed by

$$m_2^n = \frac{1}{t_p} \sum_{i \in T_P} \text{var}(X_z^n(i)) \tag{10}$$

where $X_z^n(i) = \{x_z(i)/c_z | \forall z \in I_n\}$ is the set of normalized queues of all incoming links z of node n at time step i.

Step 5: Finally, we create the set of MP controlled nodes by selecting all nodes out of the selection pool, for which $m_1^n > M_1$ and $m_2^n > M_2$, where M_1 and M_2 are case-specific thresholds. These thresholds will determine the number of nodes that will be selected, so they can be defined according to the approximate number of nodes that we wish to include in the MP control layout. Short variations around the selected values do not significantly influence the results.

1.4 Traffic simulation model

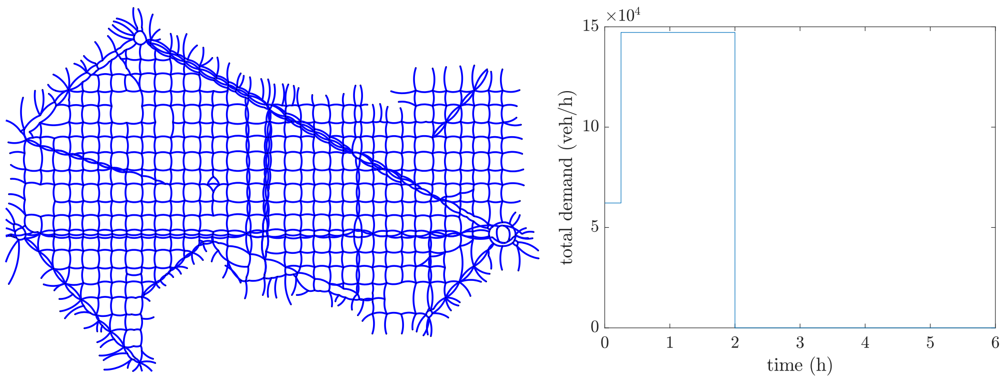
In order to assess the effectiveness of the MP controller in several different topological layouts of a large-scale network, a modified version of the mesoscopic queue-based Store-and-Forward (SaF) paradigm (see Aboudolas *et al.* (2009)), with enhanced structural properties derived by S-Model (see Lin *et al.* (2011)), is utilized as our traffic simulator. The exact version of the traffic model used in this work is described in details in Tsitsokas *et al.* (2021) and detailed presentation of its mathematical formulation is omitted. Instead, in order to facilitate understanding of this work, a brief qualitative description of the model's properties is provided in this section.

In accordance with the notation of MP controller, the traffic model monitors the state of the network, represented as a directed graph (N, Z), by updating the number of vehicles (considered as a continuous variable) inside every link z, $x_z(k)$, according to a time-discretized flow conservation equation. Vehicle flow is generated in the network according to a dynamic Origin-Destination demand matrix while routing of the flow is dictated by turning ratios. Link inflows include the sum of transit flows coming from all upstream links plus the newly generated demand of the link. Outflows include transit flows to all downstream links and trip endings inside the link. Backwards propagation of congestion is properly modeled, by replacing the initial assumption of point-queues of SaF by utilizing a more accurate transit flow calculation method. More specifically, the model imposes zero transfer flow for the following time-step if the receiving link is already congested (queue at the current time step is close to link's storage capacity). Therefore, in case of congestion, queue spill-backs are properly captured by the model.

The inherent inaccuracy in estimating travel time and real queue length of links, owed to the dimensionless nature of the initial SaF version has been decreased by integrating the structure of S-Model (Lin *et al.* (2011)), where every link queue x_z consists of two distinct queues, m_z and w_z, representing vehicle flows moving and waiting at the intersection at the end of each link, respectively. Transit flow vehicles that enter a new link firstly join the moving part of the link, where they are assumed to move with free flow speed. They transfer to the queuing part after a number of time steps determined by the position of the queue end at the moment when they entered the link. A binary function of the current time step dictates whether an approach of an intersection currently gets green light, according to the current signal plan. If an approach gets red light, this function will set the current transit flow (outflow) of this approach to zero for this time-step, regardless of the state of the queues upstream or downstream.

Figure 1: (a) Map of the modeled network of Barcelona city center; (b) Dynamic profile of total generating travel demand.

(a) (b)



By performing all calculations described by the equations of the model for a number of time steps, we get a complete traffic simulation for the specific demand scenario. By knowing the queues of every link of the network at every discrete time-step (state variables m_z and w_z, $\forall z \in Z$), we can calculate the total travel time of all vehicles together with other traffic performance measures, as we will see below.

2 Simulation experiments and Result Analysis

In order to evaluate the impact of the number and spatial distribution of nodes controlled by an MP controller in the global performance of the network, we create and evaluate a set of different MP node layouts by performing simulation experiments. The case-study network is introduced and the process of creating the MP scenarios is explained. Afterwards,the experimental setup is presented and simulation results for all evaluated cases are discussed.

2.1 Network description

The case-study network utilised in this study is a replica of part of the city center of Barcelona, Spain, which is shown in Figure 1(a). A well-calibrated model of the network in Aimsun microsimulation environment with realistic Origin-Destination demand pattern and fixed-time traffic signal plan of all controlled intersections was used as input. The network includes 1570 links and 933 nodes, out of which 565 represent signalized intersections with signal control cycles ranging from 90 to 100 sec. The dynamic profile of the total generating demand that was considered for all simulation experiments is depicted in figure 1b, and refers to a 6-hour simulation with a 15-minute warm-up period, representing the morning peak period.

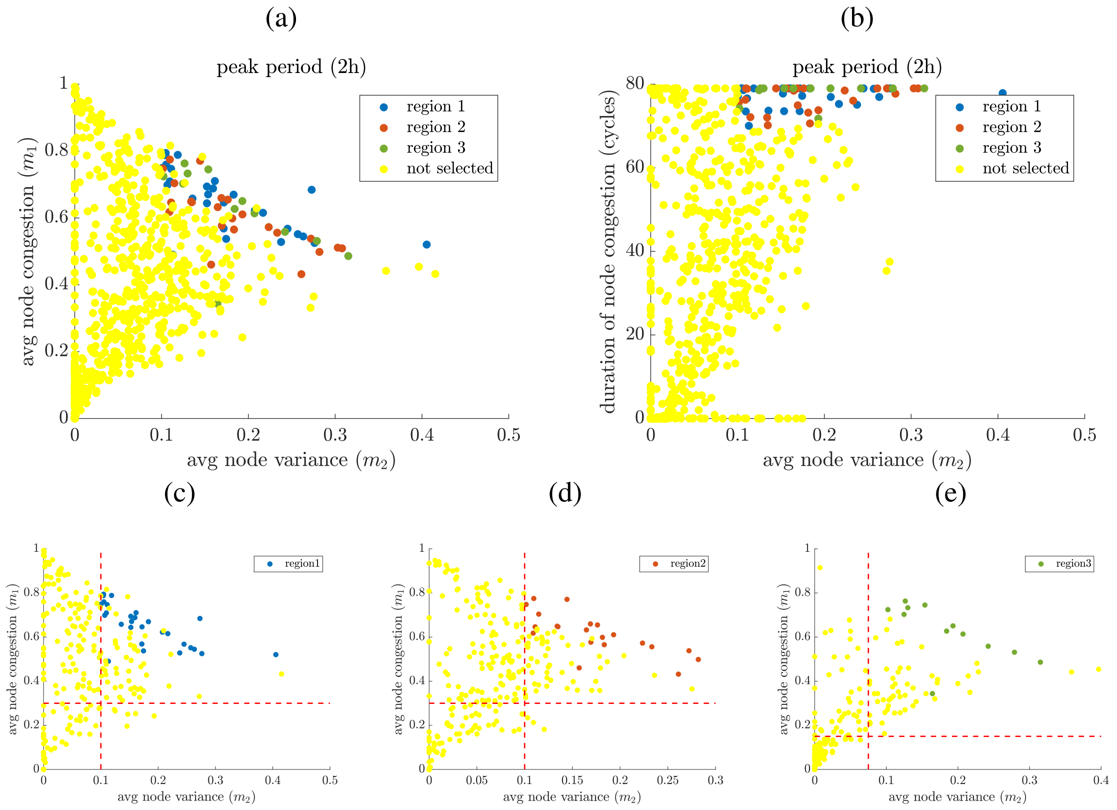
2.2 Max-Pressure scenarios

We construct here the set of cases that we evaluate in this work. The case of fixed-time, static signal control, labelled as 'FTC' is used as reference, for evaluating the network performance improvement of all MP control scenarios. In this case, no traffic-responsive control is applied and the pre-timed signal settings are used. The second scenario is the typical case that is often considered, which includes global application of MP control in all eligible network intersections and is labelled as 'MP-0'.

In order to evaluate the proposed selection process, we create two scenarios, labeled as 'MP-S1' and 'MP-S2' according to the procedure explained in the previous section. Simulation results for 'FTC' case are used as basis for the selection process. Since queue variance is highly correlated with traffic heterogeneity in the wider region, before setting variance threshold M_2, we divide the network to three homogeneously congested regions, by using the clustering method described in Saeedmanesh and Geroliminis (2016). Then, for every region we set different congestion and variance thresholds, M_1 and M_2. The values used as thresholds are chosen through a trial-and-error process, according to the aimed number of nodes. 'MP-S1' and 'MP-S2' are subsets including approximately 10 % and 25 % of all signalized network nodes. For 'MP-S1', M_1 is set as 0.3, 0.3 and 0.15 and M_2 is set to 0.1, 0.1 and 0.075, for regions 1, 2 and 3, respectively. For 'MP-S2', M_1 is set as 0.2, 0.25 and 0.1 and M_2 is set to 0.08, 0.08 and 0.05, for regions 1,2 and 3, respectively, The peak-period chosen lasts 2-hours (0.5 to 2.5 h of simulation time) and for both scenarios we exclude nodes that remained congested for less than 70 out of 80 cycles of 90 sec included in the 2-hour period.

Figure 3 graphically describes the selection process for scenario 'MP-S1', where each dot corresponds to a signalized network intersection. Figures 3(a) and (b) refer to the entire network and show the relation between congestion and queue variance (M_1 vs M_2) and duration of congestion and queue variance (M_2), respectively, for all network nodes; (c)–(e) depict the same information as (a) but separately for every homogeneous region, while red dotted lines represent the imposed thresholds. The blue, red and green dots represent selected nodes from regions 1, 2 and 3, respectively, while yellow dots are filtered-out, either because they fall below thresholds or because they did not fulfill the requirement of Step 2 of the process. Figure 5 provides a spatial visualization of the selected nodes in the cases of 'MP-S1' and 'MP-S2'.

Figure 3: Classification of signalized nodes according to the proposed method based on the results of the 'FTC' scenario.(a) Average congestion vs. queue variance for all network nodes; (b) Number of congested cycles during the 2-hour peak period vs. node queue variance; (c)–(e) Average congestion vs. queue variance per region. Dotted lines depict the congestion-variance thresholds imposed per region for the MP-S1 control scenario.



These scenarios are evaluated and compared to 'FTC' and 'MP-0', as well as a set of randomly constructed layouts, for which MP regulator is randomly assigned to node subsets corresponding to 10%, 25%, 50 % and 75% of total network nodes. For each of these categories, simulation is performed for 10 randomly constructed node subsets. The randomly constructed layouts are used as reference, in order to estimate the real improvement achieved through the proposed targeted selection process.

2.3 Experiment setup

At first, preliminary microsimulation analysis is performed in Aimsun for the fixed-time control scenario, for the purpose of extracting time-dependent turning ratios for all approaches. This set is used as input for the traffic model described in the previous section, that is used for all

Figure 5: Plan of the studied network with 3-region clustering and spatial visualization of selected intersections for MP control: (a) Scenario 'MP-S1'; (b) Scenario 'MP-S2'.

(a) Scenario 'MP-S1' (b) Scenario 'MP-S2'.

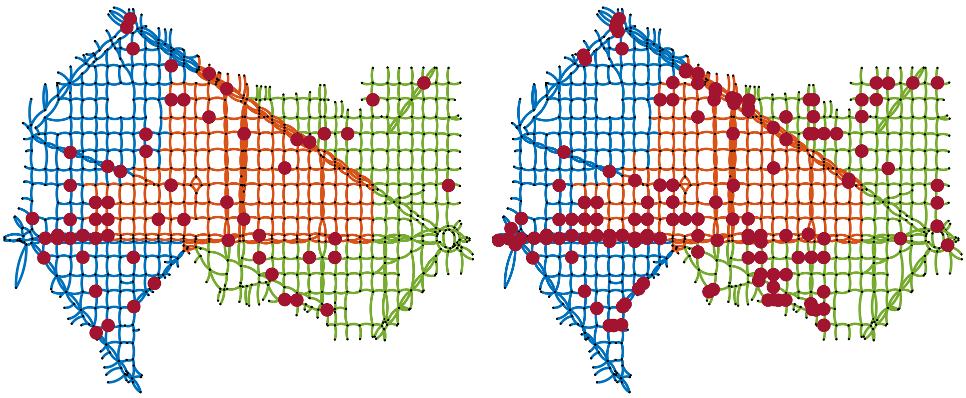
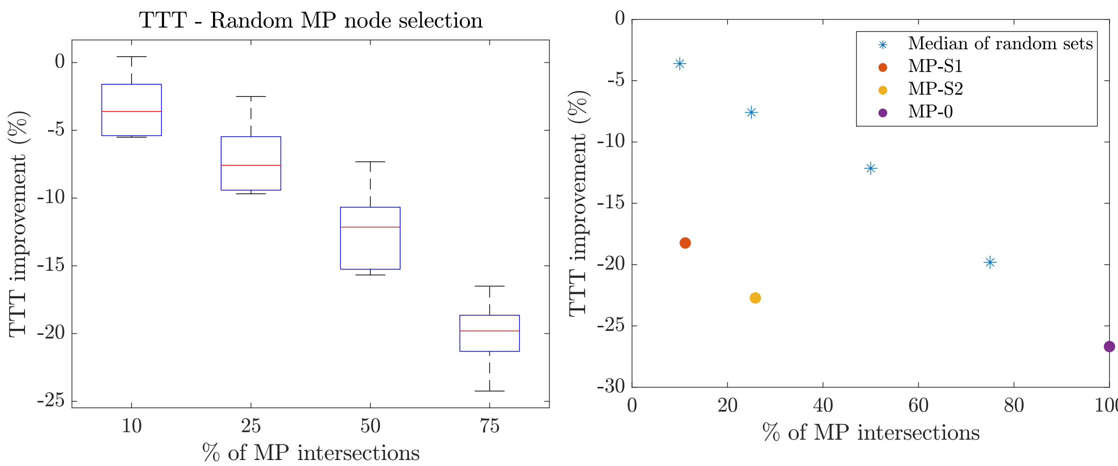


Figure 7: Comparison of Total Travel Time improvement with respect to the fixed-time control (FTC) scenario. (a) Randomly selected control nodes and (b) Median of random scenarios with 'MP-S1' and 'MP-S2', where control nodes are selected based on congestion and queue variance criteria.

(a) (b)



simulation experiments presented here. For every evaluated scenario, simulation is run for 6 hours, in time steps of 1 second. Free-flow speed of vehicles in the moving part of links, v_{ff} is set to 25 km/h and average vehicle length for capacity calculations is set to 5 meters. For all MP controlled intersections, the controller modifies only phases lasting longer than 7 seconds in the pre-timed scheduling, which was considered as minimum green time per phase, $g_{n,j,\min}$. Maximum allowed fluctuation of green times between consecutive cycles, $g^R_{n,j}$ is set to 5 seconds, for every modified phase of any intersection. All performance measures presented in this work are calculated based on the state variables provided by our simulation model. For the MP control,

Table 1: Performance measures of different MP layout scenarios. Percentile improvements with respect to 'FTC' scenario are shown in parentheses.

Performance Measure	FTC	MP-0		MP-S1		MP-S2	
Total Travel Time (VHT)	226837	166313	(–26.3%)	185473	(–18.2%)	175305	(–22.7%)
Total VQ time (VHT)	135986	77929	(–42.7%)	93002	(–31.6%)	84369	(–38.0%)
Mean VQ (veh)	22663	12988	(–42.7%)	15500	(–31.6%)	14061	(–38.0%)
SMS (km/h)	2.3	4.2	(+81.2%)	3.3	(+41.5%)	4.3	(+84.9%)

the queue values given as inputs to the controller are the real queue values calculated by the simulation model, thus assuming that the controller receives perfect information of the network traffic state.

2.4 Simulation results

In this section, simulation results from all tested layout scenarios of MP control system are summarized and compared to the reference case of fixed-time control ('FTC'). Figure 7(a) depicts the network performance for all scenarios of MP control where the set of controlled intersections is randomly selected. The vertical axis refers to the Total Travel Time (TTT) improvement with respect to the reference scenario of fixed-time control. Achieved percentile improvement in TTT for 10 randomly selected node subsets is shown by boxplots, one for every percentage of network nodes controlled. Figure 7(b) presents the median improvements of the random selection cases, together with the achieved improvement of the scenarios 'MP-S1' and 'MP-S2', where controlled nodes are selected based on congestion and queue variance characteristics, as described in the previous section, and the one of scenario 'MP-0', where all network signalized intersections receive a controller. Firstly, we observe that MP control improves network performance in all tested cases, with the highest improvement achieved by 'MP-0' scenario, in which all intersections receive a MP regulator. For this case, the achieved delay saving is more than 26 %. Moreover, we notice that significant improvement in travel time is achieved even when the controller is set to only a small fraction of intersections, and grows higher as more nodes are controlled. However, the achieved performance improvement is considerably higher in the cases of 'MP-S1' and 'MP-S2' scenarios, compared to cases with random selection of equal number of intersections. As a matter of fact, 'MP-S1' scenario only includes around 10% of the network intersections and leads to more than 18% travel time savings, while the best case of randomly built scenario containing approximately the same number of intersections (10%) only achieves 6% time savings compared to the fixed-time scenario. In other words, 75% of the network nodes randomly selected reach the same performance with the case of smart selection of only 10% of nodes. The same observation is made for 'MP-S2' scenario, which achieves 22% time savings while controlling only 25% of the network intersections, when the best random scenario with equal number of controlled nodes achieves less than 10% time savings.

A more detailed analysis based on simulation results is performed for the three best performing scenarios, 'MP-0', 'MP-S1' and 'MP-S2', and results are contrasted to those of the fixed-time control scenario, 'FTC'. Table 1 summarizes the performance of the four scenarios in terms of total travel time, time spend waiting in Virtual Queues (outside the network) due to congestion and length of Virtual Queues, as well as Space Mean Speed (SMS). As already discussed, MP decentralized control is successful in significantly reducing travel time in over-saturated conditions in all cases. More importantly, we notice that approximately 70% of the highest observed improvement ('MP-0') is possible to achieve by controlling only 10% of the network intersections ('MP-S1'), therefore with only one tenth of the initial cost. A large part of the overall delay savings result from the observed reduction of waiting time in queues, which in all cases drops by more than 31%. Speed is also significantly improved for all cases, even though the reported values are very low, because they include the long waiting time spent in virtual queues outside of the network.

Figure 9 is a collection of subfigures that graphically depict and contrast the observed system state together with various performance indicators, during the simulation experiments of the three MP control scenarios 'MP-0','MP-S1','MP-S2', and the one corresponding to fixed-time control 'FTC'. Figure 9a shows the evolution of the observed relation between accumulation and travel production in the network, that is often depicted in Macroscopic Fundamental Diagrams (MFD). We observe that in all cases of MP control, the capacity of the network is increased with respect to the fixed-time scenario, resulting in smaller peak accumulation for the same demand scenario. This is a particularly interesting finding as it indicates that a decentralized MP system can increase the system serving capacity in over-saturated conditions without drastically modifying the existing infrastructure. One possible explanation is that MP regulator assigns more green time to critical links where queues are close to capacity, thus reducing the probability of spill-backs that can rapidly propagate backwards and lead to local gridlocks. Therefore, waiting time in virtual queues is decreased as network service capacity increases. However, the controller does not impede the capacity drop caused by the accumulation of congestion in the network.

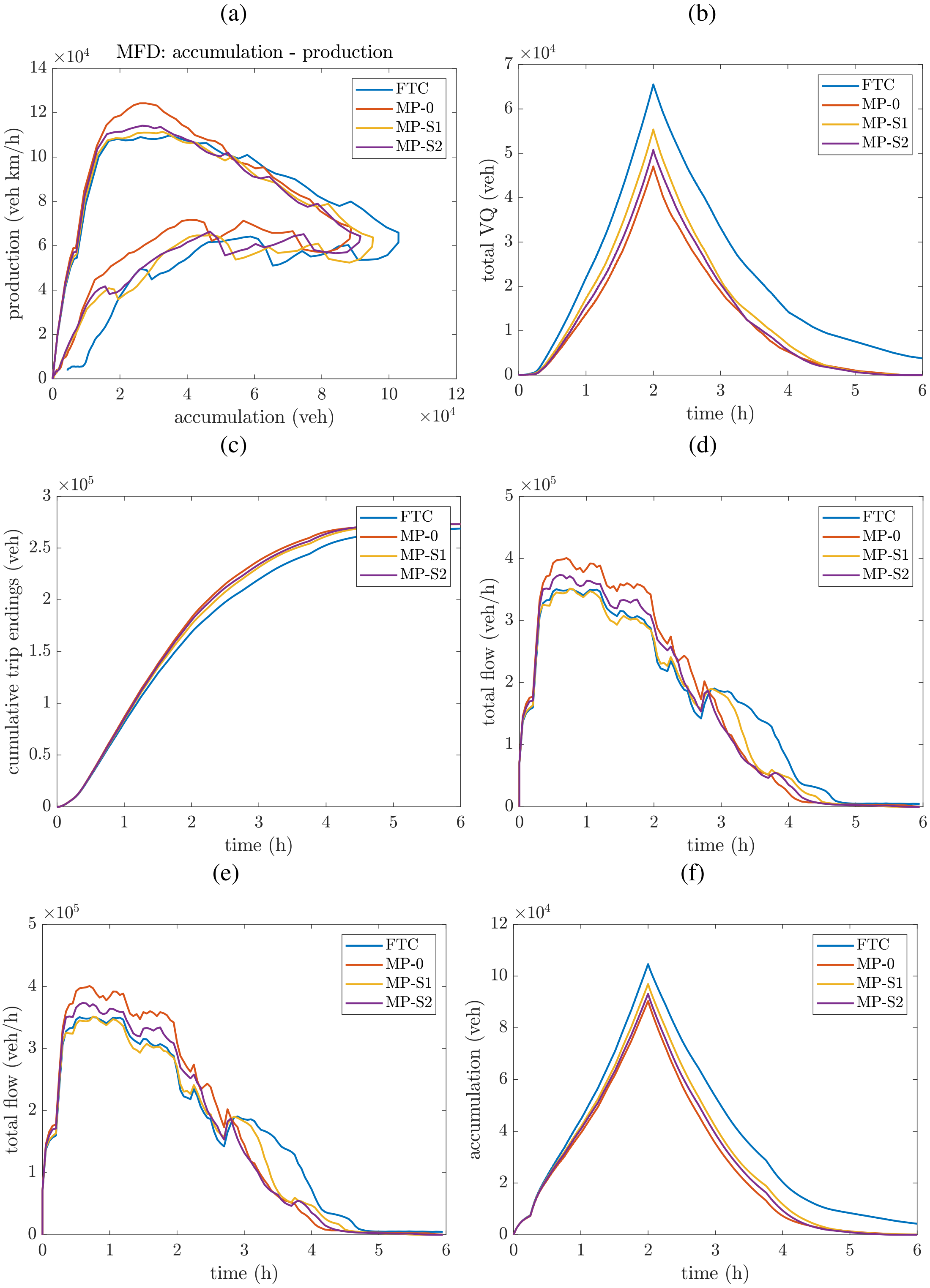
Figure 9(b) shows the evolution of the sum of vehicles waiting in virtual queues over time, where we notice a significant reduction in all cases of MP control, which is correlated to the increased network capacity. Figure 9(c) depicts the cumulative trip endings for every case, where we notice that all MP controlled scenarios are improved with respect to trip-ending rate. Figure 9(d) shows the evolution of the standard deviation of all queues of the network over time, where we observe a slight increase while congestion builds but a significant drop soon after the peak is reached, that grows during the unloading of the network. This indicates that MP controller helps reducing heterogeneity in traffic distribution under congested conditions with respect to fixed-time control. This property of the MP controller could be particularly useful in MFD-based control strategies, where traffic homogeneity is important for the effectiveness of the control decisions. Finally, Figures 9(e) and (f) depict the network flow and accumulation, respectively, over time, confirming the improved performance in the cases of MP control.

3 Discussion

In this work, a decentralized traffic-responsive signal control framework based on the existing MP controller is discussed and applied in a large-scale network instance through a dynamic queue-based traffic simulation model. The focus is on examining the potential benefits of the system in terms of network traffic performance when local MP regulators are applied to only a subset of the network intersections, as opposed to most existing approaches where all network nodes are controlled, in an effort to increase field applicability by decreasing the installation and maintenance cost related to the required queue monitoring infrastructure. A method of classification of network nodes, in terms of experienced congestion and variance of adjacent queues, is also presented, in the scope of assessing the importance of each node for receiving a MP controller, as part of a network control layout. High node importance indicates higher impact of the local control to the global network performance. The results indicate that significant improvement in terms of total delay, waiting time and space mean speed can be achieved via MP control. While performance improvement increases with the number of controlled nodes, more than 85% of the maximum observed total travel time improvement is also achieved when only 25% of properly selected, critical nodes receive MP controller, thus with only one fourth of the cost. This result proves the proposed node selection method effective.

Moreover, the demonstrated ability of the controller in reducing traffic heterogeneity in the network in congested conditions, compared to the fixed-time signal control, motivates research in efficient ways of combining decentralized MP control with central control systems, such as MFD-based perimeter control, that can highly benefit from this property, while potentially achieving higher performance gains by protecting congestion-prone regions from reaching over-saturated states. Future work should include the combination of MP node selection with perimeter control. A recent study by Keyvan-Ekbatani *et al.* (2019) combined an adaptive

Figure 9: Comparison of simulation results of the three best performing MP control scenarios, 'MP-0', 'MP-S1' and 'MP-S2', vs. the fixed-time control case 'FTC'.



local controller in all nodes together with perimeter control for a single region and showed important improvements. Investigating how a novel critical node selection method coupled with perimeter control for multi-region systems could improve traffic performance should be a research priority.

4 References

Aboudolas, K., M. Papageorgiou and E. Kosmatopoulos (2009) Store-and-forward based methods for the signal control problem in large-scale congested urban road networks, *Transportation Research Part C: Emerging Technologies*, **17** (2) 163–174.

Chow, A. H., R. Sha and S. Li (2020) Centralised and decentralised signal timing optimisation approaches for network traffic control, *Transportation Research Part C: Emerging Technologies*, **113**, 108–123.

Diakaki, C., M. Papageorgiou and K. Aboudolas (2002) A multivariable regulator approach to traffic-responsive network-wide signal control, *Control Engineering Practice*, **10** (2) 183–195.

Gartner, N. H. (1983) *OPAC: A demand-responsive strategy for traffic signal control*, no. 906.

Gregoire, J., X. Qian, E. Frazzoli, A. De La Fortelle and T. Wongpiromsarn (2014) Capacity-aware backpressure traffic signal control, *IEEE Transactions on Control of Network Systems*, **2** (2) 164–173.

Henry, J.-J., J. L. Farges and J. Tuffal (1984) The prodyn real time traffic algorithm, in *Control in Transportation Systems*, 305–310, Elsevier.

Hunt, P., D. Robertson, R. Bretherton and R. Winton (1981) Scoot-a traffic responsive method of coordinating signals, *Technical Report*.

Keyvan-Ekbatani, M., X. Gao, V. V. Gayah and V. L. Knoop (2019) Traffic-responsive signals combined with perimeter control: investigating the benefits, *Transportmetrica B: Transport Dynamics*, **7** (1) 1402–1425.

Kouvelas, A., J. Lioris, S. A. Fayazi and P. Varaiya (2014) Maximum pressure controller for stabilizing queues in signalized arterial networks, *Transportation Research Record*, **2421** (1) 133–141.

Kouvelas, A., M. Saeedmanesh and N. Geroliminis (2017) Enhancing model-based feedback perimeter control with data-driven online adaptive optimization, *Transportation Research Part B: Methodological*, **96**, 26–45.

Le, T., P. Kovács, N. Walton, H. L. Vu, L. L. Andrew and S. S. Hoogendoorn (2015) Decentralized signal control for urban road networks, *Transportation Research Part C: Emerging Technologies*, **58**, 431–450.

Lin, S., B. De Schutter, Y. Xi and H. Hellendoorn (2011) Fast model predictive control for urban road networks via milp, *IEEE Transactions on Intelligent Transportation Systems*, **12** (3) 846–856.

Lowrie, P. (1990) Scats, sydney co-ordinated adaptive traffic system: A traffic responsive method of controlling urban traffic.

Manolis, D., T. Pappa, C. Diakaki, I. Papamichail and M. Papageorgiou (2018) Centralised versus decentralised signal control of large-scale urban road networks in real time: a simulation study, *IET Intelligent Transport Systems*, **12** (8) 891–900.

Mauro, V. and C. Di Taranto (1990) Utopia, *IFAC Proceedings Volumes*, **23** (2) 245–252.

Mercader, P., W. Uwayid and J. Haddad (2020) Max-pressure traffic controller based on travel times: An experimental analysis, *Transportation Research Part C: Emerging Technologies*, **110**, 275–290.

Mirchandani, P. and L. Head (2001) A real-time traffic signal control system: architecture, algorithms, and analysis, *Transportation Research Part C: Emerging Technologies*, **9** (6) 415–432.

Saeedmanesh, M. and N. Geroliminis (2016) Clustering of heterogeneous networks with directional flows based on âsnakeâ similarities, *Transportation Research Part B: Methodological*, **91**, 250–269.

Sirmatel, I. I., D. Tsitsokas, A. Kouvelas and N. Geroliminis (2021) Modeling, estimation, and control in large-scale urban road networks with remaining travel distance dynamics, *Transportation Research Part C: Emerging Technologies*, **128**, 103157.

Tassiulas, L. and A. Ephremides (1990) Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks, paper presented at the *29th IEEE Conference on Decision and Control*, 2130–2132.

Tsitsokas, D., A. Kouvelas and N. Geroliminis (2021) Modeling and optimization of dedicated bus lanes space allocation in large networks with dynamic congestion, *Transportation Research Part C: Emerging Technologies*, **127**, 103082.

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

Varaiya, P. (2013b) The max-pressure controller for arbitrary networks of signalized intersections, in *Advances in Dynamic Network Modeling in Complex Transportation Systems*, 27–66, Springer.

Wongpiromsarn, T., T. Uthaicharoenpong, Y. Wang, E. Frazzoli and D. Wang (2012) Distributed traffic signal control for maximum network throughput, paper presented at the *2012 15th international IEEE conference on intelligent transportation systems*, 588–595.

Zhang, R., Z. Li, C. Feng and S. Jiang (2012) Traffic routing guidance algorithm based on backpressure with a trade-off between user satisfaction and traffic load, paper presented at the *2012 IEEE Vehicular Technology Conference (VTC Fall)*, 1–5.