Optimizing distribution of metered traffic flow in perimeter control
Queue and delay balancing approaches

Keyvan-Ekbatani, Mehdi; Carlson, Rodrigo Castelan; Knoop, Victor L.; Papageorgiou, Markos

DOI
10.1016/j.conengprac.2021.104762

Publication date
2021

Document Version
Final published version

Published in
Control Engineering Practice

Control Engineering Practice 110 (2021) 104762

Contents lists available at ScienceDirect

Control Engineering Practice

journal homepage: www.elsevier.com/locate/conengprac

Optimizing distribution of metered traffic flow in perimeter control: Queue and delay balancing approaches

Mehdi Keyvan-Ekbatani [a,*], Rodrigo Castelan Carlson [b], Victor L. Knoop [c], Markos Papageorgiou [d]

[a] Complex Transport Systems Laboratory (CTSLAB), Department of Civil and Natural Resources Engineering, University of Canterbury, New Zealand
[b] Department of Automation and Systems, Technological Center, Federal University of Santa Catarina, Brazil
[c] Department of Transport and Planning, Delft University of Technology, The Netherlands
[d] Dynamic Systems and Simulation Laboratory, School of Production Engineering, Technical University of Crete, Greece

ARTICLE INFO

MSC:
00-01
99-00

Keywords:
Urban traffic control
Perimeter traffic flow control
Macroscopic or network fundamental diagram
Optimization

ABSTRACT

Perimeter traffic flow control based on the macroscopic or network fundamental diagram provides the opportunity of operating an urban traffic network at its capacity. Because perimeter control operates on the basis of restricting inflow via reduced green times at selected entry (gated) links, vehicles on those links may be subject to queuing and delay. The experienced delay or resulting queue lengths depend on the adopted policy for the distribution of the inflows and corresponding green times at the gated links. The chosen policy may have a significant impact on the traffic system under control. For example, managing queue lengths may reduce the interference with upstream traffic whereas the management of delays may improve users' perception with respect to equity and fairness. In this paper, an approach has been proposed to distribute the gated flow based on the queue lengths or experienced delay at the gated signalized junctions. This is in contrast to standard practice that distributes inflows proportionally to the gated links' saturation flows. Perimeter control is then evaluated in a microscopic simulator for a realistic traffic network and compared in three configurations against fixed-time: perimeter control without queue or delay management; perimeter control with relative queue balancing; and perimeter control with delay balancing. It has been found that managing the queues at the gated links not only improves the overall network performance but also reduces the possibility of queue propagation to the upstream junctions. This improves traffic flow outside the protected network by managing the queue propagation at the gated links and reducing the possibility of queue spill-back to upstream intersections. In addition, the results indicate that perimeter control with delay balancing has a similar performance as the case without queue or delay management being a suitable approach for flow distribution among the gated links. In the scenarios with perimeter control with either queue or delay balancing the gap between the ordered flow by the controller and the actual flow crossing the stop-line at the gated links reduced remarkably.

1. Introduction

Traffic control systems are widely used worldwide to enhance traffic conditions in and around metropolitan areas. Some traffic control systems intrinsically imply the creation of queues and the localized increase in time delays for some of the users as a mean for improving the overall system performance. This is the case, for example, of ramp metering (Papageorgiou & Kotsialos, 2002), where vehicles are queued and delayed at the on-ramp so as to keep the freeway operating at capacity. With the same objective, Mainstream Traffic Flow Control (MTFC) (Carlson et al., 2010) generates a so-called controlled congestion on the mainstream by holding back vehicles a few hundreds of meters upstream of a bottleneck. When the traffic control system affects different streams causing more localized delay and/or queues for one stream than for the other as, e.g., in integrated (Carlson et al., 2014) or coordinated (Kotsialos & Papageorgiou, 2004) freeway traffic flow control, the system may be perceived as unfair by the road users (Kotsialos & Papageorgiou, 2004), even if the overall system performance is improved.

The problem of dealing with experienced delay, travel time or waiting times from different traffic streams has been addressed in several works related to freeway traffic flow control. For example, balancing of waiting times was applied by Papamichail and Papageorgiou (2011) in the case of dual-branch metered on-ramps, where drivers may be able to observe the traffic conditions at the nearby queue. For the case

* Corresponding author.
E-mail addresses: mehdi.ekbatani@canterbury.ac.nz (M. Keyvan-Ekbatani), rodrigo.carlson@ufsc.br (R.C. Carlson), v.l.knoop@tudelft.nl (V.L. Knoop), markos@dssl.tuc.gr (M. Papageorgiou).

https://doi.org/10.1016/j.conengprac.2021.104762
Received 20 October 2020; Received in revised form 15 December 2020; Accepted 2 February 2021
Available online xxxx
0967-0661/

of MTFC of two merging motorways, Carlson et al. (2011) proposed a split-range-like feedback structure that balances delay on both motorways. Later on, the problem of balancing delays at all bottleneck approaches in the case of integrated ramp metering and MTFC was generalized by Iordanidou et al. (2017) overcoming some limitations of previous approaches (Carlson et al., 2014, 2011).

In the case of urban traffic, several traffic control algorithms have been proposed during the last decades, e.g., OPAC (Gartner, 1990), PRODYN (Henry et al., 1983), SCATS (Slavin et al., 2013), TUC (Diakaki et al., 2002) and other traffic management strategies such as the ones proposed by Guo and Harmati (2020), Han et al. (2016), van de Weg et al. (2019), and Zhou et al. (2013). As a matter of fact, a recent work showed that typical locally adaptive signal control schemes tend to have little to no effect on traffic at network level when congested due to downstream congestion and queue spillbacks (Gayah et al., 2014). Recent studies have shown that by limiting the inflow rate into the busiest parts of the urban networks (called perimeter control or gating), the congestion might be more efficiently mitigated. More specifically, the perimeter traffic flow control or gating attempts to limit the number or accumulation of vehicles inside of a given *protected network* (PN) by reducing the green times at selected *gated links* at its border, so as to increase overall network throughput. Although early applications of gating date from the 1960s (Wood, 1993), perimeter traffic flow control gained prominence some 50 years later after the proposition by Daganzo (2007) of this type of control based on the aggregated modeling of an urban network by a Macroscopic or Network Fundamental Diagram (MFD or NFD) boosted by the quasi-simultaneous verification of the diagram's existence with field data by Geroliminis and Daganzo (2008). Following this finding, there was a surge in NFD-based perimeter urban traffic flow control strategies for single regions (Haddad, 2017a; Keyvan-Ekbatani et al., 2012, 2015a), some considering expanding regions (Keyvan-Ekbatani et al., 2015b), or the presence of public transport (Ampountolas et al., 2017; Geroliminis et al., 2014), or the presence of freeways (Haddad et al., 2013), or the combination with other real-time urban traffic control strategies (Keyvan-Ekbatani et al., 2019), strategies for multiple regions (Aboudolas & Geroliminis, 2013; Geroliminis et al., 2013; Kouvelas et al., 2017), city-wide traffic control and the impacts of cordon queues (Ni & Cassidy, 2020), congestion pricing in a connected vehicle environment (Yang et al., 2019); and also model free perimeter control (Li & Hou, 2020; Ren et al., 2020). There are other streams of NFD applications which have been to a lesser extent in the spotlight, e.g., travel time reliability (Mahmassani et al., 2013), level of service and resilience of the network (Hoogendoorn et al., 2015), pedestrian dynamics (Hoogendoorn et al., 2017), traffic safety (Alsalhi et al., 2018), the effects of the public transport system on the NFD of corridors (Castrillon & Laval, 2018), and NFD for train traffic operations (Corman et al., 2019).

The reduction of the green times at the gated links aiming at reducing the inflow to the PN results in queued and delayed vehicles just outside the PN, at its border. Few works on this subject addressed the effect of the queued vehicles on the traffic conditions upstream of the gated links. Haddad (2017b) proposed a hybrid network structure to model the flow dynamics for two urban regions and the aggregated queue dynamics at the boundary of the regions. In an earlier work (Keyvan-Ekbatani et al., 2016), we have shown that the increased throughput achieved by the application of NFD-based feedback perimeter traffic flow control renders the gated link queues shorter than with fixed-time only. Further, we have shown that additional improvements in overall network delay may be obtained if the queue lengths at the gated links are managed so as to reduce the interference with upstream junctions. However, the balancing of gated link queues results in rather unbalanced delays that, as noted before, may be perceived unfair by the road users.

In this paper we propose a practical strategy for queue or delay balancing at the gated links for perimeter traffic flow control, following Iordanidou et al. (2017) and Keyvan-Ekbatani et al. (2016). More specifically, the feedback perimeter control strategy proposed by Keyvan-Ekbatani et al. (2012) is revisited and the inflow distribution is performed based on the management of queue lengths or delays at the gated links instead of in proportion to saturation flows as is the usual practice. Beyond the expected benefits in terms of performance and equity/fairness discussed earlier, the gap between controller decisions (ordered inflows) and effective result (actual inflows) is expected to reduce, thus increasing controller performance. Similar to Keyvan-Ekbatani et al. (2016), the problem is formulated as a continuous knapsack problem that is solved efficiently. However the optimization formulation by Keyvan-Ekbatani et al. (2016) in fact provided only quasi-balanced queues. The problem is reformulated here so as to provide real balanced relative queues at the gated links (Iordanidou et al., 2017). In addition, we extend the formulation to also allow for the balancing of delays at the gated links. The approach is unified so that the same formulation for the optimization problem is used via parameterization of the dynamic model. We also note that Keyvan-Ekbatani et al. (2016) obtained the queue measurements directly from the simulator, whereas here emulated loop detectors are used along with queue estimation via a Kalman filter (Vigos et al., 2008), rendering the results more realistic and closer to practice. Simulation results indicate that perimeter control with queue balancing improves the overall network delay while perimeter control with delay balancing has a similar performance as the case of perimeter control without queue or delay management.

In Section 2, we review the basics of NFD-based feedback perimeter control. In Section 3 we present the queue balancing and the delay balancing strategies, unifying both formulations. The network performance and the delay balancing strategy are evaluated in the sequel in Section 4. Finally, Section 5 concludes the paper.

2. NFD-based feedback perimeter control

Traffic management via perimeter control may be realized by the use of a feedback regulator designed after a control design model based on the concept of the NFD. In this section, we introduce some definitions related to the NFD and present the feedback regulator that delivers the total flow that should enter the PN. A discussion on how the total flow may be distributed among the gated links is also presented.

2.1. Network Fundamental Diagram (NFD)

As shown in Fig. 1, an NFD describes a relation between the network production (or average weighted flow) and accumulation (or average density) under homogeneous traffic conditions (Geroliminis & Daganzo, 2008). Unlike the ordinary fundamental diagram (FD) for highways, the NFD is reproducible and has a flat peak (with a range of critical density). Evaluating the NFD of a given network allows one to determine the maximum number of vehicles that should be maintained inside that network so as to have maximum throughput, i.e., operate at the diagram's critical region (indicated by orange in Fig. 1).

In this paper we follow Keyvan-Ekbatani et al. (2012) and represent the NFD alternatively as the relation between Total Traveled Distance, TTD (veh·km/h), and the Total Time Spent, TTS (veh·h/h). This form is suitable for use with the available traffic measurements. The TTD and TTS for a given network are derived via the following equations (Keyvan-Ekbatani et al., 2012):

$$TTS(k) = \sum_{z\in\mathbb{M}} \frac{T\cdot \tilde{N}_z(k)}{T} = \sum_{z\in\mathbb{M}} \tilde{N}_z(k) = \tilde{N}(k) \tag{1}$$

$$TTD(k) = \sum_{z\in\mathbb{M}} \frac{T\cdot q_z(k)\cdot L_z}{T} = \sum_{z\in\mathbb{M}} q_z(k)\cdot L_z \tag{2}$$

with $k = 0, 1, 2, \ldots$ the discrete time-step reflecting the corresponding traffic cycles with length T (h), $\mathbb{M}$ the set of network links equipped with detectors, and q_z (veh/h) and L_z (km), the measured flow and

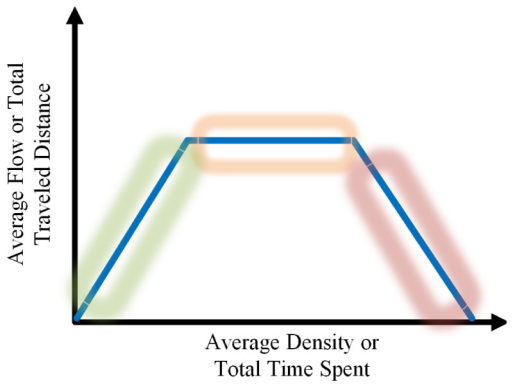


Fig. 1. Relationship between average flow and average density at the network level; referred as NFD in the literature.

length of link z, respectively. Note that in this case the TTS corresponds to the total number of vehicles $\tilde{N}$ estimated from all equipped links. The estimated number of vehicles in link z is given by

$$\tilde{N}_z(k) = L_z \cdot \frac{\mu_z}{100\lambda} \cdot o_z(k) \tag{3}$$

with o_z (%) and μ_z (lanes), the occupancy measurement and the number of lanes of link z, respectively, and λ (km) the average vehicle length.

2.2. Feedback perimeter traffic flow control

The dynamic system of the PN under perimeter traffic flow control can be modeled by a first-order model whose input is the total flow entering the network through the gated links, q_g, and the output is the total time spent, TTS. The model is linearized at around the (critical) TTS value for which the TTD is maximal. Thus, a discrete-time feedback Proportional–Integral (PI) regulator may be used to determine the amount of flow q_g that should enter the network through the gated links during the next time period (Keyvan-Ekbatani et al., 2012), and is given by

$$q_g(k) = q_g(k-1) - K_P\,[TTS(k) - TTS(k-1)] + K_I\left[T\hat{T}S - TTS(k)\right] \tag{4}$$

with K_P and K_I the proportional and integral gains, respectively, and $T\hat{T}S$ a set-point value within the NFD critical region for network throughput maximization. The TTS value used as the input of the regulator at every time step k is estimated from occupancy measurements via (3) and (1). The control objective is to keep the network traffic state in the orange and green zones of the NFD and avoid oversaturated traffic conditions, i.e., the red zone (see Fig. 1) during the peak period (i.e., excessive demand). More details about the design and implementation of (4) may be found in Keyvan-Ekbatani et al. (2012).

2.3. Flow distribution

The flow q_g resulting from (4) corresponds to the total flow that should enter the PN through the n gated links and must be appropriately distributed among them. The usual approach is to distribute the flow proportionally to each gated link according to its saturation flow as done, e.g., by Aboudolas and Geroliminis (2013), Keyvan-Ekbatani et al. (2012) and Kouvelas et al. (2017). This policy may even be integrated with a spill-back and low demand (at the gated links) alert algorithm for increased efficiency (Keyvan-Ekbatani et al., 2015b). Another possibility is distributing the total flow aiming at balancing the relative link queues as done by Papamichail and Papageorgiou (2011) for a dual branch freeway on-ramp and by Keyvan-Ekbatani et al. (2016) for perimeter traffic flow control. As discussed in the introduction, yet another possibility for distributing the total flow may aim at balancing the delays or waiting times at the gated links similar to what was done by Papamichail and Papageorgiou (2011) for a dual branch freeway on-ramp, by Carlson et al. (2011) for merging freeways, and by Iordanidou et al. (2017) for integrated freeway traffic flow control.

Whatever distribution method is pursued, the flows q_i to be implemented at each gated link i must satisfy

$$\sum_{i=1}^{n} q_i = q_g \tag{5}$$

for a successful perimeter control operation.

Because traffic lights are employed to regulate flows at the gated links, the distributed flows must be translated to green times. This translation depends on the links' saturation flow s_i, as well as the cycle length. The green time at gated link i can be computed as $g_i = (q_i \cdot T)/s_i$ (Keyvan-Ekbatani et al., 2019). Moreover, typical restrictions on the signal settings must be respected, thus the flow on a gated link may be subject to a lower bound $q_{\min,i}$ (veh/h), e.g., due to a specified minimum green time, and to an upper bound $q_{\max,i}$ (veh/h), e.g., due to the green time duration of the respective phase of the base signal plan, i.e.,

$$q_{\min,i} \le q_i \le q_{\max,i}. \tag{6}$$

It should be noted that the lower and upper flow bounds at the gated links imply respective bounds to be applied a posteriori to the result of (4), i.e.,

$$\sum_{i=1}^{n} q_{\min,i} \le q_g(k) \le \sum_{i=1}^{n} q_{\max,i}, \tag{7}$$

and the bounded value should be used as $q_g(k-1)$ in the next control step (cycle) to avoid the wind-up effect.

3. Delay and queue management

In this section we develop a delay balancing strategy for the gated links and review the queue management strategy. The optimization problem formulation presented by Keyvan-Ekbatani et al. (2016) that provided only quasi-balanced queues is revised to achieve real balanced queues. The reason for managing queues stems from the fact that the queues formed at gated links might exceed the link length and interfere with upstream junctions at some locations, while the link storage capacity (in number of vehicles) is underutilized at other locations. From that perspective, queue lengths could be taken into account as a mechanism for deciding on flow distribution (see Section 2.3) so as to make a better use of links storage space, thus reducing the effect on the traffic conditions at the surrounding areas and improving even more the overall network performance. Queue delays on the other hand, tend to affect negatively the driver perception of the system under operation (Zhang & Levinson, 2005). Therefore, delays could be taken into account for deciding on flow distribution so as to increase equity among drivers entering the PN and, thus, improve their perception about the benefits of the control system.

3.1. The queue model

Assuming there are no sinks or sources within the gated link, the time evolution of the queues N_i (veh) on each gated link i, $i = 1, 2, \dots, n$, may be derived from the equation of the conservation of the number of vehicles (Papamichail & Papageorgiou, 2011) as

$$N_i(k+1) = N_i(k) + T[d_i(k) - q_i(k)] \tag{8}$$

in which d_i (veh/h) is the flow that enters and q_i (veh/h) is the flow that leaves the gated link i during time period $[kT, (k+1)T)$.

3.2. Relative queue

A relative queue value is obtained dividing the current queue length $N_i(k)$ by the maximum queue length $N_{\max,i}$ (veh) defined for the respective gated link i, i.e.,

$$N_i^{\text{rel}}(k+1) = \frac{N_i(k)}{N_{\max,i}} + \frac{T[d_i(k) - q_i(k)]}{N_{\max,i}}. \tag{9}$$

We can rewrite (9) as

$$N_i^{\text{rel}}(k+1) = A_i^{\text{q}}(k) - B_i^{\text{q}}(k) q_i(k) \tag{10}$$

$$\text{with } A_i^{\text{q}}(k) = \frac{N_i(k) + T d_i(k)}{N_{\max,i}} \text{ and } B_i^{\text{q}}(k) = \frac{T}{N_{\max,i}}.$$

3.3. Delay

A simplified form for estimating the experienced delay $\tau_i(k)$ (h) at gated link i can be obtained if we divide (8) by the link inflow d_i (Papamichail & Papageorgiou, 2011):

$$\tau_i(k+1) = \frac{N_i(k)}{d_i(k)} + T\left[1 - \frac{q_i(k)}{d_i(k)}\right]. \tag{11}$$

Note that, since the value of d_i at $k+1$ is not available, the value at k is used under the assumption that the demand is constant over that period. This a reasonable assumption in traffic systems and shown to be effective in previous works (Iordanidou et al., 2017; Papamichail & Papageorgiou, 2011).

We can rewrite (11) as

$$\tau_i(k+1) = A_i^{\text{d}}(k) - B_i^{\text{d}}(k) q_i(k) \tag{12}$$

$$\text{with } A_i^{\text{d}}(k) = \frac{N_i(k)}{d_i(k)} + T \text{ and } B_i^{\text{d}}(k) = \frac{T}{d_i(k)}.$$

3.4. Relative queue or delay balancing

The ordered flow q_{g} may be distributed among all gated links aiming at balancing delays or relative link queues. The problem formulation leading either to balanced delays or to balanced relative queues is as follows:

$$\min \sum_{i=1}^{n} \frac{\left(A_i^m(k) - B_i^m(k) q_i(k)\right)^2}{B_i^m(k)} \tag{13}$$

$$\text{s.t.: (5) and (6),}$$

and $m \in \{\text{q}, \text{d}\}$ is chosen depending on whether we are balancing relative queues or delays. This formulation corresponds to the continuous quadratic knapsack problem.

It is straightforward to show through the application of the first-order optimality conditions that the solution of (13) provides exactly balanced delays or relative queues at the gated links as long as the resulting q_i are within the bounds given by (6). Otherwise, only the delays or relatives queues of the gated links with q_i within the bounds are exactly balanced while the others activate the corresponding bounds (6). Note that with balanced relative queue lengths, all queues will reach their respective maximum simultaneously.

If desired, weights can be assigned to each term of the summation in (13) in order to prioritize some gated links over the others (Iordanidou et al., 2017). The combination of queues and delays on the objective function does not seem to target any clear goal, therefore it was not investigated. We note that, if desired, inequality constraints could be added with upper bounds on the delays when queues are being balanced and vice-versa.

We solve the optimization problem (13) with an efficient semi-smooth Newton method, as proposed and implemented by Cominetti et al. (2014). The required number of iterations grows linearly with the number of gated links, i.e., it has time complexity $O(n)$. In fact, for the simulations in Section 4 the number of iterations was exactly n.

4. Simulation

For the evaluation of the delay balancing strategy proposed in Section 3, we simulate a realistic urban network model in the microscopic simulator AIMSUN (TSS - Transport Simulation Systems, 2016).

4.1. Network description

Fig. 2 (left) depicts the simulated network that corresponds to a portion of the city of Chania, Greece. The area within the dashed box is enlarged on the right side of Fig. 2 and corresponds to the city center of Chania. The shaded area is chosen as the PN. Note that, compared to the real network an extra junction has been added upstream of the gated link 2, to assess the impact of queue propagation to the upstream links.

The PN consists of some 80 junctions, 27 of which are signalized with traffic lights, and 165 links. Many of those links may experience over-saturated conditions in the peak-period. The traffic signals operate on the basis of fixed-time control plans with a cycle length of 90 s.

The demand is modeled by an O–D matrix that attempts to mimic the peak-period with realistic traffic conditions for a period of 4 h, including sinks and sources within the PN. A trapezoidal demand profile is produced which starts from low flow values, increases gradually to high flows that are sustained for a certain period before being reduced gradually to low flow values, until the network is virtually emptied by the end of the simulation.

4.2. Scenarios

We consider four different control scenarios: (i) *no-perimeter-control* (NPC), in which only fixed-time traffic control has been applied; (ii) *perimeter control without balancing* (PC), in which feedback perimeter traffic flow control has been implemented and the flow distribution is proportional to the links' saturation flows; (iii) *perimeter control with queue balancing* (PCQ), in which the feedback perimeter traffic flow control has been implemented and the flow distribution is obtained from the solution of the relative queue balancing problem (13) with $m = \text{q}$; and (iv) *perimeter control with delay balancing* (PCD), in which the feedback perimeter traffic flow control has been implemented and the flow distribution is obtained from the solution of the delay balancing problem (13) with $m = \text{d}$.

4.3. Simulation and control setup

Eight different signalized junctions have been chosen at the border of the PN for perimeter control, i.e., $n = 8$ junctions, with corresponding gated links identified by numbers 1–8 in Fig. 2 (right).

The regulator's gains are the same used by Keyvan-Ekbatani et al. (2012), i.e., $K_{\text{P}} = 20\ \text{h}^{-1}$ and $K_{\text{I}} = 5\ \text{h}^{-1}$. The maximum throughput (or TTD) for this network occurs in a TTS range of 500–750 veh·h/h (see Section 4.5.1 and Fig. 5(a)). Therefore, the set-point has been chosen as $\hat{TTS} = 600$ veh·h/h. For the estimation of the TTS to be used as the input of (4), occupancy measurements are collected at every cycle ($T = 90$ s) from loop detectors placed around the middle of every link in the PN. In addition, three loop detectors are placed at each one of the gated links for estimation of the link queues, one around the middle, one and the entrance and one at the exit. The regulator runs during the entire simulation period, however the calculated q_{g} value is implemented only if the real-time measured TTS is close to $\hat{TTS}$. During these periods we say that perimeter control is activated. In this case, the green splits of the fixed-time plans at the junctions with gated links are overridden by the control system, otherwise the fixed-time signal plans remain unchanged.

For the implementation of the delay and relative queue management and performance analysis, the inflows (d_i) and actual outflows (q_i) at the gated links were collected from loop detectors, and the

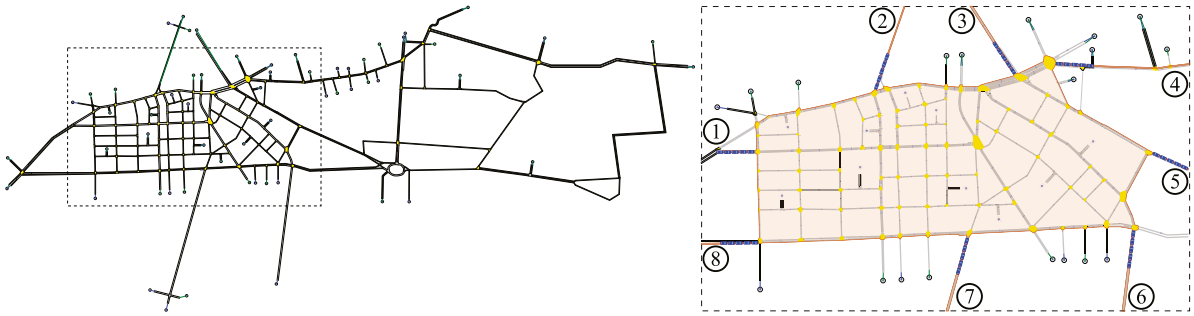


Fig. 2. Simulated Network (city of Chania, Greece) on the left; the area within the dashed box is enlarged in the right — the shaded area shows the protected network, and the eight gated links for perimeter urban traffic flow control are indicated.

occupancy (o_i) from the middle detector. Because of cycle to cycle variations, exponentially smoothed values d_i^{es} and q_i^{es} were used in place of d_i and q_i, i.e., $d_i^{\text{es}}(k) = \alpha d_i(k) + (1-\alpha)d_i^{\text{es}}(k-1)$ and $q_i^{\text{es}}(k) = \alpha q_i(k) + (1-\alpha)q_i^{\text{es}}(k-1)$, with $0 < \alpha < 1$ a smoothing factor. In this paper we use $\alpha = 0.5$, arbitrarily chosen. Then, the current queue lengths (N_i) were estimated via a practical Kalman Filter estimator as proposed by Vigos et al. (2008) instead of (8), since in a real setting cumulative errors result from (8) (and, consequently, from (12)). The filter gain was chosen as $K = 0.1$ (see, Vigos et al., 2008) without the need of further fine tuning. The queue estimates were used in (9) to obtain the estimates of the relative queues (N_i^{rel}) while, the delays (τ_i) were estimated through (12) using the estimated queues as inputs. Note that Keyvan-Ekbatani et al. (2016) collected these quantities directly from AIMSUN, not based on the emulated detectors as in this work. The maximum storage space of each gated link i, in number of vehicles, has been considered as the maximum queue length value ($N_{\max,i}$).

The dynamic traffic assignment option in AIMSUN has been activated, see (Keyvan-Ekbatani et al., 2012), in order to obtain a more realistic and homogeneous distribution of demand within the network.

4.4. Analysis

In this paper, the average delay (as provided by AIMSUN), expressed in s/km, is chosen as the metric of network performance (not only the PN). It indicates the increase of travel time compared to free flow conditions where traffic is uninterrupted by traffic lights. The total delay results from the multiplication of the delay with the total traveled distance. All results are based on traffic variables measured during a sample time of 90 s, i.e., equal to the cycle length. To account for the stochastic effects of the microscopic simulation, ten different replications (simulation runs with different seeds) were carried out for each scenario.

4.5. Results

The box plot displayed in Fig. 3 summarizes the overall network delay values (all replications) for the four scenarios (NPC, PC, PCQ, and PCD) in terms of minimum and maximum (top and bottom black bars), first quartile and the third quartile (top and bottom of the blue box), and median (red bar inside the box). As it can be observed in the figure, the average delay is remarkably reduced in the PC, PCQ and PCD scenarios (on average 33.8%, 41.0% and 39.0%, respectively).

The performance of the PCD scenario spans the performance of the PC and PCQ scenarios showing, thus, a higher variability of delay performance that hinders the obtained average improvement almost as high as PCQ. The PCQ scenario box (representing the interquartile range (IQR)) is slightly lower than the box for the other scenarios, because of less interference with upstream traffic due to the queue balancing. In the NPC scenario, the spread is rather large which is typical for heavily congested traffic.

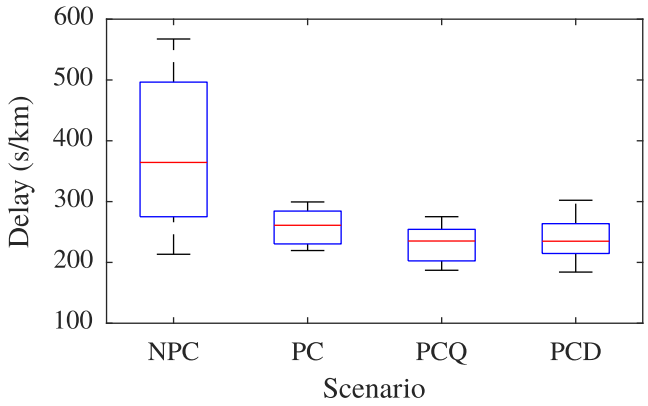


Fig. 3. Boxplot of the overall network delay for the four scenarios (all replications): no perimeter control (NPC), perimeter control without balancing (PC), perimeter control with relative queue balancing (PCQ), and perimeter control with delay balancing (PCD). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The better performance of the perimeter control scenarios can be further appreciated by the curves of cumulative Total Traveled Distance (TTD) in Fig. 4. It is clear from the figure that at around $t = 1.5$ h, when perimeter control is activated, the curves of the controlled scenarios depart from the NPC scenario indicating higher throughput. For the shown replication, PCQ and PCD outperform slightly the PC scenario as can be seen from 2.0 to 3.0 h.

An interesting simulation aspect revealed by this figure is the higher final cumulative TTD value for the NPC scenario. This is due to some vehicles that, after entering the protected network (PN), re-route to avoid congestion and leave the PN just to re-enter again through another gated link. A cumulative plot of inflows at the gated links (not shown) confirms a higher final cumulative inflow at the gated links for the NPC scenario as well.

4.5.1. Analysis of the NFDs

The NFDs of the PN for the four scenarios (all replications) are presented in Fig. 5. For a better distinction, the measurement points are displayed with black dots for the network loading (first 2 h) and with gray dots for the recovery period (last 2 h). The maximum network throughput (or TTD) of around 5400 veh·km/h can be observed to occur at a TTS range of 500–750 veh·h/h in the NFD of the NPC scenario in Fig. 5(a). Recall that the value of $T\hat{T}S$ was chosen within this range for maximum throughput.

Fig. 5(b) shows the NFD for the PC scenario. The regulator succeeded to maintain the TTS close to $T\hat{T}S$ during the perimeter control activation, showing that the approach is effective as expected (Keyvan-Ekbatani et al., 2012). Therefore, perimeter control contributes to keeping the production at a high level and, hence, the delays low (see Fig. 3). A typical hysteresis loop (for loading and unloading the

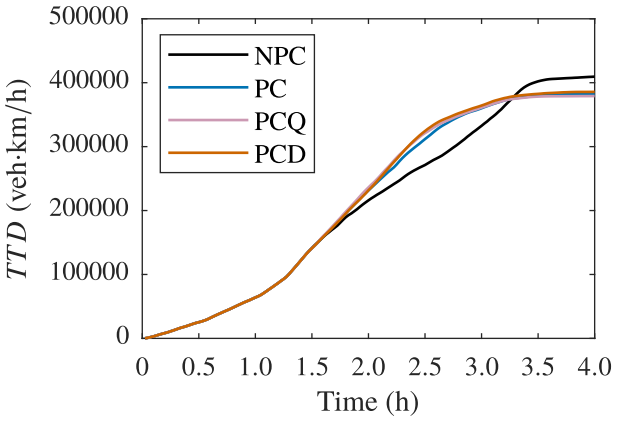


Fig. 4. Cumulative Total Traveled Distance (TTD) in PN for one replication.

network) is visible more clearly in this scenario (Gayah & Daganzo, 2011).

The NFD for the PCQ scenario is displayed in Fig. 5(c). As in the PC scenario (Fig. 5(b)), the regulator tries to keep the throughput at the maximum level. Unlike the NFD for PC, however, the critical accumulation ($\hat{TTS}$ = 600 veh·h/h) is exceeded in some replications. This is the result of the attempt of queue management to balance relative queues at the gated links. The resulting flow distribution may lead to higher inflow into the PN at junctions with higher demand and consequently create local congestion in some downstream junctions with short links. The combination of perimeter control with real-time control of the inner intersections could alleviate this problem. Another consequence is that a larger hysteresis and more scatter can be observed compared to the NFD for the PC scenario suggesting less homogeneous distribution of density in the PN.

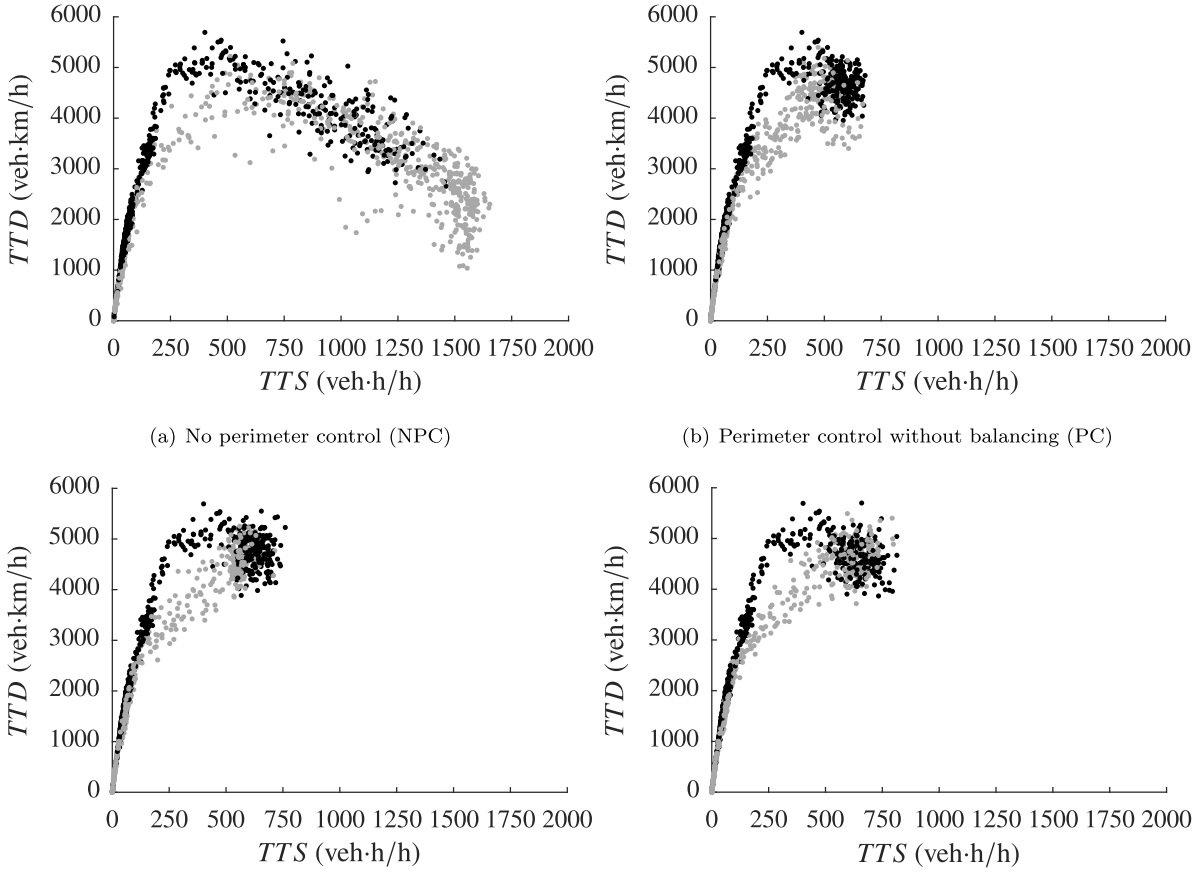
Finally, the NFD for the PCD scenario is displayed in Fig. 5(d). We see more scatter than the PC and PCQ scenarios, as well as more violations of the set-point.

4.5.2. Analysis of relative queues

For the analysis of the PCQ approach, four replications, one for each scenario, with closest delay values to the respective average network delay have been chosen. The time series of relative queues at the gated links are shown in Fig. 6. The time window in which the perimeter control is activated is shown by the vertical dashed-lines.

In Fig. 6(a), it can be seen that in the NPC scenario queues grow strongly at some of the gated links and last for about two hours (e.g., see the queues at gated links 2, 4, 5 and 7), with some queue spill-back. Thus, as also discussed by Keyvan-Ekbatani et al. (2016), even without perimeter control, queues are formed due to a combination of spill back effects from inside the network and excessive demand. On the other hand, on gated links with low demand, for instance gated links 6 and 8, queues are hardly noticeable for most of the time.

When applying perimeter control (PC), the inflow is restricted and queues are seen to further increase roughly at most gated links (except links 3 and 8, due to low demand), see Fig. 6(b). although more queues are seen to spill-back for short periods of time, the queues last much less time than in the NPC scenario (about half an hour to one hour less). Despite of the improvement due to perimeter control, the relative queues remain unbalanced and some gated links experience



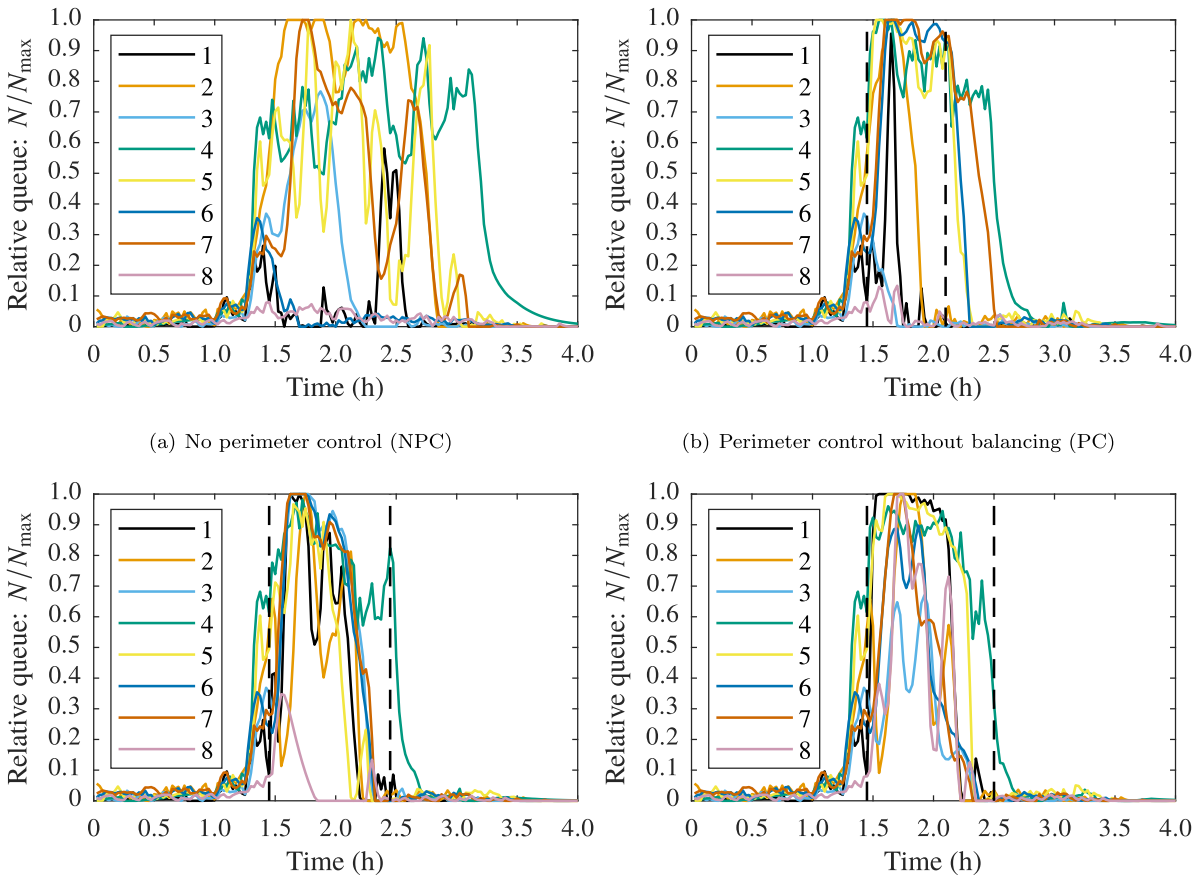
(a) No perimeter control (NPC)

(b) Perimeter control without balancing (PC)

(c) Perimeter control with relative queue balancing ('PCQ)

(d) Perimeter control with delay balancing (PCD)

Fig. 5. Network Fundamental Diagrams for the protected network (all replications).



(a) No perimeter control (NPC)

(b) Perimeter control without balancing (PC)

(c) Perimeter control with relative queue balancing (PCQ)

(d) Perimeter control with delay balancing (PCD)

Fig. 6. Relative queues at the gated links for one replication.

long queues even after switching off the perimeter control (see links 2, 4 and 7).

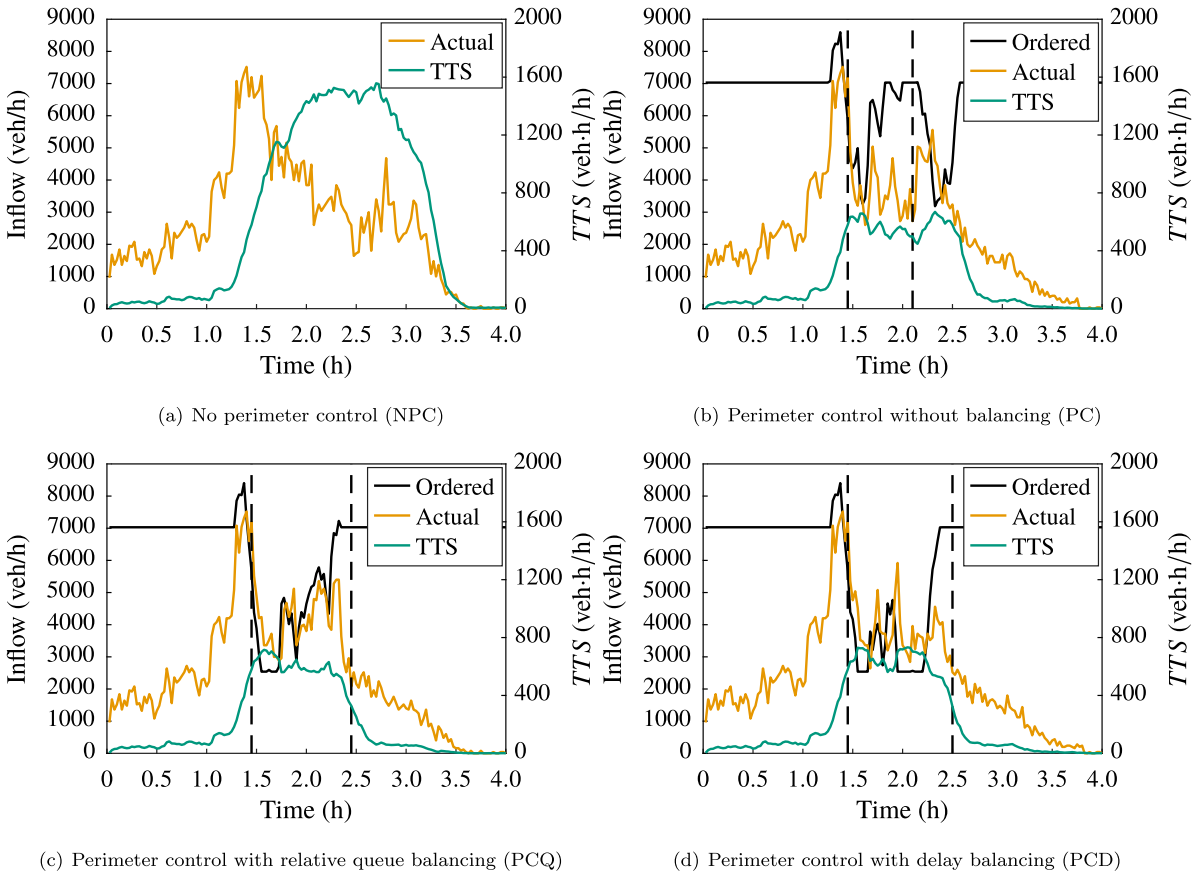
The comparison between Fig. 6(a) and (b) resolves an important general concern regarding perimeter control: perimeter control does not necessarily lead to longer queues (in space and time) than in the no perimeter control case. Throughput is higher with PC whereas with NPC links within the PN may be over-saturated and back-spilling. Consequently, lower flow can be served at the boundary links and longer queuing period may be experienced with NPC than with PC.

Fig. 6(c) depicts the results of applying perimeter control with queue balancing (PCQ). Although, the relative queue lengths are not perfectly equal due to different level of demand at the gated links, they are remarkably similar for more than half of the activation period and more than without the queue balancing. The duration of the queues is even shorter in this scenario (except link 4), with most queues dissolved at around 2.2 h. The spill-backs occur for even shorter periods of time.

The relative queues in the PCD scenario are shown in Fig. 6(d). Since the control objective in this scenario is balancing the experienced travel time and not the created queues, the queues seem less balanced than in the case of PCQ. Compared to the PCQ scenario, queues last for a longer period at high levels interfering more with upstream links.

4.5.3. Analysis of actual and ordered flows

Previous studies (Keyvan-Ekbatani et al., 2012, 2015a) revealed that distributing the flow ordered by the regulator (4) based on the saturation flow of the gated links may lead to underutilization of green times and, consequently, to discrepancies between ordered flow at the gated links and the realized or actual flow due to downstream oversaturation. Fig. 7 shows in orange the actual flow entering the PN through the gated links and in green the actual total time spent (TTS) for all four scenarios, whereas the black lines indicate the ordered flows by the regulator for the three controlled scenarios. In these controlled scenarios (Fig. 7(b)–(d)), the traffic conditions are identical as in the NPC scenario (Fig. 7(a)) up to around $t = 1.4$ h, when gating is switched on (after the first vertical dashed line), as *TTS* approaches its threshold value for perimeter control operation; and consequently the gating regulator orders low inflow values to maintain *TTS* around its set point. There is a clear gap between the ordered flow and actual flow during $t \in [1.7, 2.1]$ h in Fig. 7(b) due to the underutilization of the green times, as mentioned before. In fact, this was one of the motivations to carry out the current study and Keyvan-Ekbatani et al. (2016). Implementing the PCQ and PCD approaches led to a more efficient distribution of the ordered flow and consequently reduction of the gap between the ordered flow and actual flow (see Fig. 7(c) and (d)). Note that the increasing gap occurring between the actual and ordered flow after $t = 2.1$ h in the PCQ scenario is not harmful and is basically due to the gradual decrease of the demand profile. The slightly unbalanced queue values for some links in the PCQ scenario may be the result of local temporal blockages downstream of the gated links recognizable also in Fig. 7(c) during $t \in [1.5, 2]$ h. Increasing the capacity of the downstream junctions, by combining the perimeter control scheme with traffic-responsive control strategies within the PN (Keyvan-Ekbatani et al., 2019), might be fruitful to minimize the aforementioned gap as much as possible and, as mentioned earlier, reduce the effect this may have on links downstream of gated links with higher inflow. Note that there are time periods ($t \in [1.5, 1.7]$ h and $t \in [1.7, 2.3]$ h in Fig. 7(c) and (d)) that the orange line is above the black line. This occurs due to the high demand at some gated links that have interphases in their signal plans which are not controlled during the activation of perimeter control.



(a) No perimeter control (NPC)

(b) Perimeter control without balancing (PC)

(c) Perimeter control with relative queue balancing (PCQ)

(d) Perimeter control with delay balancing (PCD)

Fig. 7. Total ordered and actual inflows at the gated links and scenarios' total time spent (TTS) for one replication. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.5.4. Analysis of delay

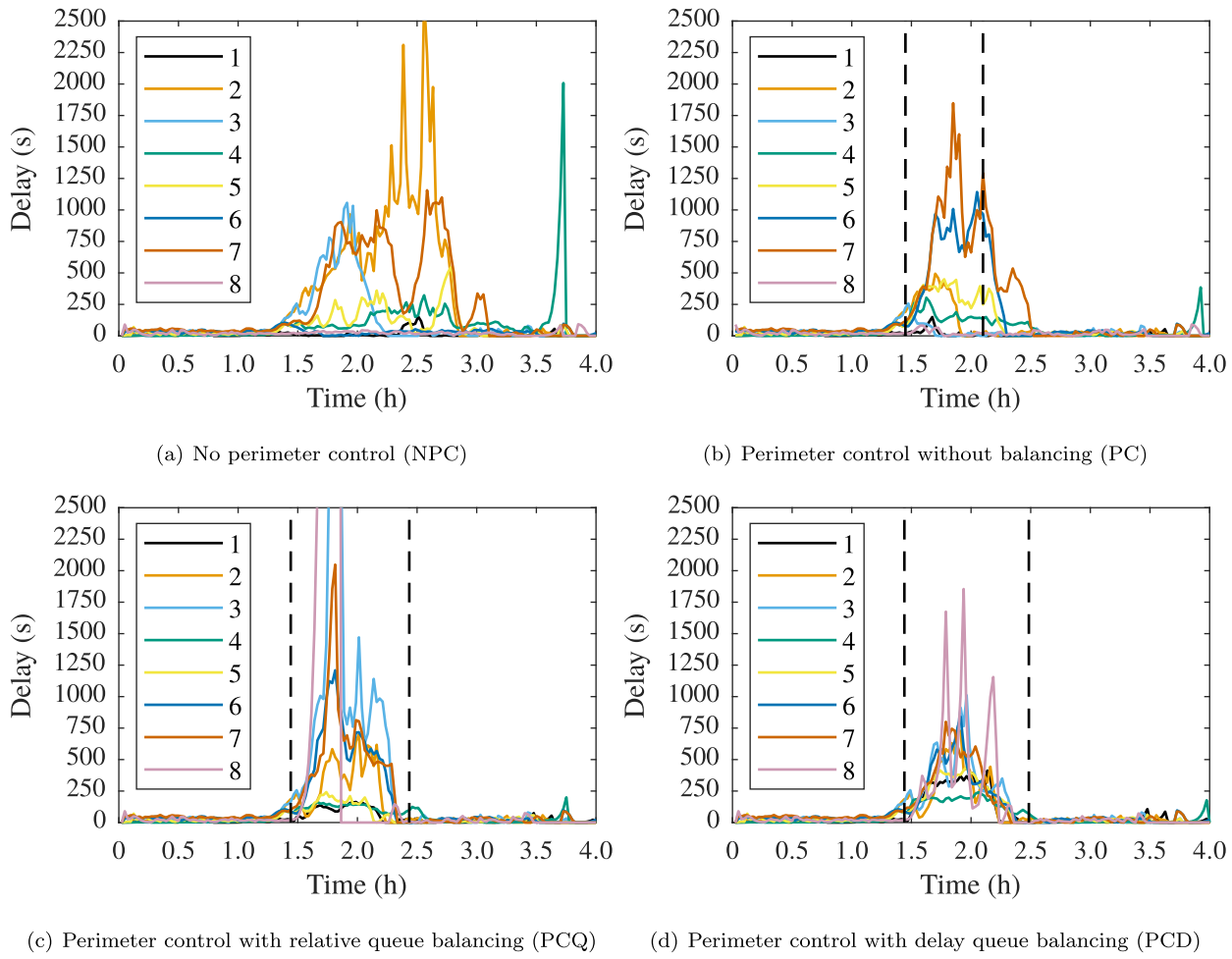
Fig. 8(a) depicts the estimated delay at the eight junctions indicated in Fig. 2 for the NPC scenario. During the peak period ($t \in [1.5, 3]$ h) the experienced delay at links 2 and 7 remain at high values (800–1000 s). Note that the delay surge in link 2 at around $t = 2.5$ h corresponds to a reduced inflow at the gated link as measured by the upstream detector. Indeed, the queue (Fig. 6(a)) exceeds the link length and covers the detector for most of the cycle time, affecting the accuracy of the estimation. The increased delays appearing by the end of the simulation are the result of low values of d_i (d_i^{es}) and could be filtered out. Comparing the delay values for different scenarios (e.g., Fig. 6(a) and (b)), interestingly, the highest delays have been experienced in the NPC scenario. As a result of the NFD-based perimeter control implementation (PC, PCQ and PCD), the traffic conditions within the PN improved and the network throughput increased significantly. Thus, lower delay values are also obtained at the gated links compared to the NPC scenario. As expected, the PCQ scenario, despite the perfectly balanced relative queues, suffers from unbalanced delays at the gated links. This can raise equity and fairness issues in case of real-life implementation. The high delay in link 8 and the spikes in links 3 and 7 within the time window 1.5 to 2.1 h are again the result of low inflows, either due to low demand or due to queues temporarily covering the detector. Anyhow, a higher delay for PCQ compared to PC and PCD is coherent during this period, since its inflows are lower during this period (hardly visible in Fig. 7, but clear on an cumulative plot not shown). The delay-based flow distribution algorithm, introduced in this paper, resulted in much closer (balanced) delays at the eight gated links (see Fig. 8(d)).

5. Conclusions

Monitoring and controlling the queue growth at the gated links is a crucial and challenging task for NFD-based perimeter control schemes. This paper focuses on proposing a knapsack-based optimal flow distribution to balance the experienced delays and relative queues at the gated links in a unified approach. In a comparative simulation study, queuing and delays at the gated links have been analyzed thoroughly. Four different simulation scenarios have been tested in microscopic simulation environment: (i) *no-perimeter-control* (NPC); (ii) *perimeter control without queue balancing* (PC); (iii) *perimeter control with queue balancing* (PCQ); and (iv) *perimeter control with delay balancing* (PCD).

The study revealed that even without perimeter control, the links at the periphery of the urban regions may experience long queuing and high delays due to over-saturated traffic conditions and queue spill-backs downstream. Applying the proposed delay management approach reduced the queuing period at the gated links compared to NPC scenario and consequently balanced the waiting times at the gated links. Applying the PCD approach reduced the overall network delay compared to NPC (similar to PCQ) while analogous overall performance has been achieved compared to PC with lower queue propagation probability at the gated links. This is indeed an important finding towards the real-field implementation of perimeter control strategies, assuring the efficiency of gating without negatively affecting the vicinity of the gated links. It has been also shown that the gap between the actual and ordered flow can be remarkably minimized in case of PCD. Note, this is an important consequence of applying PCD which avoids waste of green times at the gated links.

The proposed flow distribution approaches might be beneficial to be implemented in multi-region perimeter control strategies to reduce the negative impact of the queuing on the NFDs of the regions. Future



(a) No perimeter control (NPC)

(b) Perimeter control without balancing (PC)

(c) Perimeter control with relative queue balancing (PCQ)

(d) Perimeter control with delay queue balancing (PCD)

Fig. 8. Delay at the gated links for one replication.

research direction also includes integration of the recently proposed real-time deep-learning-based queue estimation approach by Lee et al. (2019) into the PCQ framework.

Note that the authors strived to make this paper as close as possible to a practical application by using a high-fidelity state-of-the-art simulator to simulate a model of a real traffic network. The network was calibrated with real data and uses realistic inputs. Real sensors have been emulated, and practical control strategies and estimators have been proposed and applied. A successful field implementation will depend on the engagement of vendors and public authorities, as well as adequate communication with road users.

The authors believe that the main technical issues for field implementation of perimeter control are well-advanced, and this paper contributes to addressing some of the remaining technical issues. On the other hand, field implementation of traffic control strategies and, in particular, of perimeter control needs to overcome some non-technical issues related to proprietary control infrastructure, public authority inertia and driver acceptance. The authors believe that, in view of the sensible societal gains expected from the implementation of perimeter control, the non-technical issues will be eventually overcome, as already happened in other domains of traffic control with a longer history.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

R. C. Carlson was supported by The Brazilian Agency for Higher Education (CAPES), under the project PrInt CAPES–UFSC “Automation 4.0”; and by the National Council for Scientific and Technological Development (CNPq). The authors would like to thank the three anonymous reviewers for their constructive feedback and input provided during the review process.

References

Aboudolas, K., & Geroliminis, N. (2013). Perimeter and boundary flow control in multi-reservoir heterogeneous networks. *Transportation Research, Part B (Methodological)*, *55*, 265–281. http://dx.doi.org/10.1016/j.trb.2013.07.003, URL http://linkinghub.elsevier.com/retrieve/pii/S0191261513001185.

Alsalhi, R., Dixit, V. V., & Gayah, V. V. (2018). On the existence of network macroscopic safety diagrams: Theory, simulation and empirical evidence. In X. Ma (Ed.), *PLoS One*, *13*(8), Article e0200541. http://dx.doi.org/10.1371/journal.pone.0200541, URL https://dx.plos.org/10.1371/journal.pone.0200541.

Ampountolas, K., Zheng, N., & Geroliminis, N. (2017). Macroscopic modelling and robust control of bi-modal multi-region urban road networks. *Transportation Research, Part B (Methodological)*, *104*, 616–637. http://dx.doi.org/10.1016/j.trb.2017.05.007, URL https://linkinghub.elsevier.com/retrieve/pii/S0191261515300370.

Carlson, R. C., Papamichail, I., & Papageorgiou, M. (2014). Integrated feedback ramp metering and mainstream traffic flow control on motorways using variable speed limits. *Transportation Research Part C (Emerging Technologies)*, *46*, 209–221. http://dx.doi.org/10.1016/j.trc.2014.05.017, URL https://linkinghub.elsevier.com/retrieve/pii/S0968090X14001727.

Carlson, R. C., Papamichail, I., Papageorgiou, M., & Messmer, A. (2010). Optimal motorway traffic flow control involving variable speed limits and ramp metering. *Transportation Science*, *44*(2), 238–253. http://dx.doi.org/10.1287/trsc.1090.0314, URL http://pubsonline.informs.org/doi/abs/10.1287/trsc.1090.0314.

Carlson, R. C., Ragias, A., Papamichail, I., & Papageorgiou, M. (2011). Mainstream traffic flow control of merging motorways using variable speed limits. In *2011 19th Mediterranean conference on control & automation* (pp. 674–681). Corfu, Greece: IEEE, http://dx.doi.org/10.1109/MED.2011.5983115, URL http://ieeexplore.ieee.org/document/5983115/.

Castrillon, F., & Laval, J. (2018). Impact of buses on the macroscopic fundamental diagram of homogeneous arterial corridors. *Transportmetrica B: Transport Dynamics*, *6*(4), 286–301. http://dx.doi.org/10.1080/21680566.2017.1314203, URL https://www.tandfonline.com/doi/full/10.1080/21680566.2017.1314203.

Cominetti, R., Mascarenhas, W. F., & Silva, P. J. S. (2014). A Newton's method for the continuous quadratic knapsack problem. *Mathematical Programming Computation*, *6*(2), 151–169. http://dx.doi.org/10.1007/s12532-014-0066-y, URL http://link.springer.com/10.1007/s12532-014-0066-y.

Corman, F., Henken, J., & Keyvan-Ekbatani, M. (2019). Macroscopic fundamental diagrams for train operations - are we there yet?. In *2019 6th International Conference on Models and Technologies for intelligent transportation systems* (pp. 1–8). Cracow, Poland: IEEE, http://dx.doi.org/10.1109/MTITS.2019.8883374, URL https://ieeexplore.ieee.org/document/8883374/.

Daganzo, C. F. (2007). Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Research, Part B (Methodological)*, *41*(1), 49–62. http://dx.doi.org/10.1016/j.trb.2006.03.001, URL https://linkinghub.elsevier.com/retrieve/pii/S0191261506000282.

Diakaki, C., Papageorgiou, M., & Aboudolas, K. (2002). A multivariable regulator approach to traffic-responsive network-wide signal control. *Control Engineering Practice*, *10*(2), 183–195. http://dx.doi.org/10.1016/S0967-0661(01)00121-6, URL http://linkinghub.elsevier.com/retrieve/pii/S0967066101001216.

Gartner, N. (1990). OPAC: Strategy for demand-responsive decentralized traffic signal control. *IFAC Proceedings Volumes*, *23*(2), 241–244. http://dx.doi.org/10.1016/S1474-6670(17)52677-4, URL https://linkinghub.elsevier.com/retrieve/pii/S1474667017526774.

Gayah, V. V., & Daganzo, C. F. (2011). Clockwise hysteresis loops in the macroscopic fundamental diagram: An effect of network instability. *Transportation Research, Part B (Methodological)*, *45*(4), 643–655. http://dx.doi.org/10.1016/j.trb.2010.11.006, URL https://linkinghub.elsevier.com/retrieve/pii/S0191261510001396.

Gayah, V. V., Gao, X. S., & Nagle, A. S. (2014). On the impacts of locally adaptive signal control on urban network stability and the macroscopic fundamental diagram. *Transportation Research, Part B (Methodological)*, *70*, 255–268. http://dx.doi.org/10.1016/j.trb.2014.09.010, URL https://linkinghub.elsevier.com/retrieve/pii/S0191261514001647.

Geroliminis, N., & Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research, Part B (Methodological)*, *42*(9), 759–770. http://dx.doi.org/10.1016/j.trb.2008.02.002, URL https://linkinghub.elsevier.com/retrieve/pii/S0191261508000180.

Geroliminis, N., Haddad, J., & Ramezani, M. (2013). Optimal perimeter control for two urban regions with macroscopic fundamental diagram: A model predictive approach. *IEEE Transactions on Intelligent Transportation Systems*, *14*(1), 348–359. http://dx.doi.org/10.1109/TITS.2012.2216877, URL http://ieeexplore.ieee.org/document/6353591/.

Geroliminis, N., Zheng, N., & Ampountolas, K. (2014). A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks. *Transportation Research Part C (Emerging Technologies)*, *42*, 168–181. http://dx.doi.org/10.1016/j.trc.2014.03.004, URL https://linkinghub.elsevier.com/retrieve/pii/S0968090X14000709.

Guo, J., & Harmati, I. (2020). Evaluating semi-cooperative Nash/Stackelberg Q-learning for traffic routes plan in a single intersection. *Control Engineering Practice*, *102*, Article 104525. http://dx.doi.org/10.1016/j.conengprac.2020.104525, URL https://linkinghub.elsevier.com/retrieve/pii/S0967066120301325.

Haddad, J. (2017). Optimal coupled and decoupled perimeter control in one-region cities. *Control Engineering Practice*, *61*, 134–148. http://dx.doi.org/10.1016/j.conengprac.2017.01.010, URL https://linkinghub.elsevier.com/retrieve/pii/S0967066117300102.

Haddad, J. (2017). Optimal perimeter control synthesis for two urban regions with aggregate boundary queue dynamics. *Transportation Research, Part B (Methodological)*, *96*, 1–25. http://dx.doi.org/10.1016/j.trb.2016.10.016, URL https://linkinghub.elsevier.com/retrieve/pii/S0191261516308025.

Haddad, J., Ramezani, M., & Geroliminis, N. (2013). Cooperative traffic control of a mixed network with two urban regions and a freeway. *Transportation Research, Part B (Methodological)*, *54*, 17–36. http://dx.doi.org/10.1016/j.trb.2013.03.007, URL https://linkinghub.elsevier.com/retrieve/pii/S0191261513000477.

Han, K., Liu, H., Gayah, V. V., Friesz, T. L., & Yao, T. (2016). A robust optimization approach for dynamic traffic signal control with emission considerations. *Transportation Research Part C (Emerging Technologies)*, *70*, 3–26. http://dx.doi.org/10.1016/j.trc.2015.04.001, URL https://linkinghub.elsevier.com/retrieve/pii/S0968090X15001345.

Henry, J., Farges, J., & Tuffal, J. (1983). The prodyn real time traffic algorithm. *IFAC Proceedings Volumes*, *16*(4), 305–310. http://dx.doi.org/10.1016/S1474-6670(17)62577-1, URL https://linkinghub.elsevier.com/retrieve/pii/S1474667017625771.

Hoogendoorn, S. P., Daamen, W., Knoop, V. L., Steenbakkers, J., & Sarvi, M. (2017). Macroscopic fundamental diagram for pedestrian networks: theory and applications. *Transportation Research Procedia*, *23*, 480–496. http://dx.doi.org/10.1016/j.trpro.2017.05.027, URL https://linkinghub.elsevier.com/retrieve/pii/S2352146517303058.

Hoogendoorn, S. P., Knoop, V. L., van Lint, H., & Vu, H. L. (2015). Applications of the generalized macroscopic fundamental diagram. In M. Chraibi, M. Boltes, A. Schadschneider, & A. Seyfried (Eds.), *Traffic and granular flow '13* (pp. 577–583). Cham: Springer International Publishing, http://dx.doi.org/10.1007/978-3-319-10629-8_65, URL http://link.springer.com/10.1007/978-3-319-10629-8_65.

Iordanidou, G.-R., Papamichail, I., Roncoli, C., & Papageorgiou, M. (2017). Feedback-based integrated motorway traffic flow control with delay balancing. *IEEE Transactions on Intelligent Transportation Systems*, *18*(9), 2319–2329. http://dx.doi.org/10.1109/TITS.2016.2636302, URL http://ieeexplore.ieee.org/document/7811261/.

Keyvan-Ekbatani, M., Carlson, R. C., Knoop, V. L., Hoogendoorn, S. P., & Papageorgiou, M. (2016). Queuing under perimeter control: Analysis and control strategy. In *2016 IEEE 19th international conference on intelligent transportation systems* (pp. 1502–1507). Rio de Janeiro, Brazil: IEEE, http://dx.doi.org/10.1109/ITSC.2016.7795756, URL http://ieeexplore.ieee.org/document/7795756/.

Keyvan-Ekbatani, M., Gao, X. S., Gayah, V. V., & Knoop, V. L. (2019). Traffic-responsive signals combined with perimeter control: investigating the benefits. *Transportmetrica B: Transport Dynamics*, *7*(1), 1402–1425. http://dx.doi.org/10.1080/21680566.2019.1630688, URL https://www.tandfonline.com/doi/full/10.1080/21680566.2019.1630688.

Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I., & Papageorgiou, M. (2012). Exploiting the fundamental diagram of urban networks for feedback-based gating. *Transportation Research, Part B (Methodological)*, *46*(10), 1393–1403. http://dx.doi.org/10.1016/j.trb.2012.06.008, URL https://linkinghub.elsevier.com/retrieve/pii/S0191261512000926.

Keyvan-Ekbatani, M., Papageorgiou, M., & Knoop, V. L. (2015). Controller design for gating traffic control in presence of time-delay in urban road networks. *Transportation Research Part C (Emerging Technologies)*, *59*, 308–322. http://dx.doi.org/10.1016/j.trc.2015.04.031, URL https://linkinghub.elsevier.com/retrieve/pii/S0968090X1500176X.

Keyvan-Ekbatani, M., Yildirimoglu, M., Geroliminis, N., & Papageorgiou, M. (2015). Multiple concentric gating traffic control in large-scale urban networks. *IEEE Transactions on Intelligent Transportation Systems*, *16*(4), 2141–2154. http://dx.doi.org/10.1109/TITS.2015.2399303, URL http://ieeexplore.ieee.org/document/7061487/.

Kotsialos, A., & Papageorgiou, M. (2004). Efficiency and equity properties of freeway network-wide ramp metering with AMOC. *Transportation Research Part C (Emerging Technologies)*, *12*(6), 401–420. http://dx.doi.org/10.1016/j.trc.2004.07.016, URL https://linkinghub.elsevier.com/retrieve/pii/S0968090X04000269.

Kouvelas, A., Saeedmanesh, M., & Geroliminis, N. (2017). Enhancing model-based feedback perimeter control with data-driven online adaptive optimization. *Transportation Research, Part B (Methodological)*, *96*, 26–45. http://dx.doi.org/10.1016/j.trb.2016.10.011, URL https://linkinghub.elsevier.com/retrieve/pii/S019126151630710X.

Lee, S., Xie, K., Ngoduy, D., & Keyvan-Ekbatani, M. (2019). An advanced deep learning approach to real-time estimation of lane-based queue lengths at a signalized junction. *Transportation Research Part C (Emerging Technologies)*, *109*, 117–136. http://dx.doi.org/10.1016/j.trc.2019.10.011, URL https://linkinghub.elsevier.com/retrieve/pii/S0968090X1830812X.

Li, D., & Hou, Z. (2020). Perimeter control of urban traffic networks based on model-free adaptive control. *IEEE Transactions on Intelligent Transportation Systems*, 1–13. http://dx.doi.org/10.1109/TITS.2020.2992337, URL https://ieeexplore.ieee.org/document/9108600/.

Mahmassani, H. S., Hou, T., & Saberi, M. (2013). Connecting networkwide travel time reliability and the network fundamental diagram of traffic flow. *Transportation Research Record: Journal of the Transportation Research Board*, *2391*(1), 80–91. http://dx.doi.org/10.3141/2391-08, URL http://journals.sagepub.com/doi/10.3141/2391-08.

Ni, W., & Cassidy, M. (2020). City-wide traffic control: Modeling impacts of cordon queues. *Transportation Research Part C (Emerging Technologies)*, *113*, 164–175. http://dx.doi.org/10.1016/j.trc.2019.04.024, URL https://linkinghub.elsevier.com/retrieve/pii/S0968090X1831708X.

Papageorgiou, M., & Kotsialos, A. (2002). Freeway ramp metering: an overview. *IEEE Transactions on Intelligent Transportation Systems*, *3*(4), 271–281. http://dx.doi.org/10.1109/TITS.2002.806803, URL http://ieeexplore.ieee.org/document/1166514/.

Papamichail, I., & Papageorgiou, M. (2011). Balancing of queues or waiting times on metered dual-branch on-ramps. *IEEE Transactions on Intelligent Transportation Systems*, *12*(2), 438–452. http://dx.doi.org/10.1109/TITS.2010.2093130, URL http://ieeexplore.ieee.org/document/5664791/.

Ren, Y., Hou, Z., Sirmatel, I. I., & Geroliminis, N. (2020). Data driven model free adaptive iterative learning perimeter control for large-scale urban road networks. *Transportation Research Part C (Emerging Technologies)*, *115*, Article 102618. http://dx.doi.org/10.1016/j.trc.2020.102618, URL https://linkinghub.elsevier.com/retrieve/pii/S0968090X19309842.

Slavin, C., Feng, W., Figliozzi, M., & Koonce, P. (2013). Statistical study of the impact of adaptive traffic signal control on traffic and transit performance. *Transportation Research Record: Journal of the Transportation Research Board*, *2366*(1), 117–126. http://dx.doi.org/10.3141/2356-14, URL http://journals.sagepub.com/doi/10.3141/2356-14.

TSS - Transport Simulation Systems (2016). AIMSUN dynamic simulator users manual v. 8.

Vigos, G., Papageorgiou, M., & Wang, Y. (2008). Real-time estimation of vehicle-count within signalized links. *Transportation Research Part C (Emerging Technologies)*, *16*(1), 18–35. http://dx.doi.org/10.1016/j.trc.2007.06.002, URL http://www.sciencedirect.com/science/article/pii/S0968090X07000381.

van de Weg, G. S., Keyvan-Ekbatani, M., Hegyi, A., & Hoogendoorn, S. P. (2019). Linear MPC-based urban traffic control using the link transmission model. *IEEE Transactions on Intelligent Transportation Systems*, 1–16. http://dx.doi.org/10.1109/TITS.2019.2938795, URL https://ieeexplore.ieee.org/document/8845756/.

Wood, K. (1993). Gating traffic into congested areas. In *IEE colloquium on electronics in managing the demand for road capacity* (pp. 11/1–11/2).

Yang, K., Menendez, M., & Zheng, N. (2019). Heterogeneity aware urban traffic control in a connected vehicle environment: A joint framework for congestion pricing and perimeter control. *Transportation Research Part C (Emerging Technologies)*, *105*, 439–455. http://dx.doi.org/10.1016/j.trc.2019.06.007, URL https://linkinghub.elsevier.com/retrieve/pii/S0968090X18316127.

Zhang, L., & Levinson, D. (2005). Balancing efficiency and equity of ramp meters. *Journal of Transportation Engineering*, *131*(6), 477–481. http://dx.doi.org/10.1061/(ASCE)0733-947X(2005)131:6(477), URL http://ascelibrary.org/doi/10.1061/%28ASCE%290733-947X%282005%29131%3A6%28477%29.

Zhou, H., Bouyekhf, R., & El Moudni, A. (2013). Modeling and entropy based control of urban transportation network. *Control Engineering Practice*, *21*(10), 1369–1376. http://dx.doi.org/10.1016/j.conengprac.2013.06.007, URL https://linkinghub.elsevier.com/retrieve/pii/S0967066113001111.