Contents lists available at ScienceDirect

## Transportation Research Part B

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

# Analytical derivation of the optimal traffic signal timing: Minimizing delay variability and spillback probability for undersaturated intersections

## Reza Mohajerpoor<sup>a</sup>, Meead Saberi<sup>b</sup>, Mohsen Ramezani<sup>a,\*</sup>

<sup>a</sup> School of Civil Engineering, The University of Sydney, Sydney, Australia <sup>b</sup> School of Civil and Environmental Engineering, University of New South Wales, Sydney, Australia

## ARTICLE INFO

Article history: Received 11 January 2018 Revised 7 November 2018 Accepted 10 November 2018

Keywords: Delay variability Spillback Kinematic wave theory Travel time reliability Optimal signal timing Undersaturated intersection

## ABSTRACT

Individual vehicles experience a large variability of delay at signalized intersections in urban networks. Existing traffic signal optimization frameworks often overlook the implications of delay variability and spillback for design and analysis of signal timing plans. This paper presents an analytical solution based on the shockwave theory to estimate delay variability at an undersaturated intersection. We also propose a new optimal signal timing formulation that minimizes the delay variability and probability of spillback in addition to total delay. Several algorithms are proposed to attain the global optimal signal timing for prefixed (given) and dynamic (optimal) cycle length control strategies. Illustrative microsimulation and numerical studies demonstrate the effectiveness of the proposed formulated models and signal optimization algorithms.

© 2018 Elsevier Ltd. All rights reserved.

## 1. Introduction

Signal optimization at isolated intersections has been widely studied (see e.g. Wardrop, 1952; Webster, 1958; Gazis, 1964; AKÇELIK, 1981). The seminal work of Webster (1958) proposed an analytical formula based on queuing theory for the optimal cycle length at an isolated undersaturated intersection. In another seminal work, Gazis (1964) proposed an optimization framework based on a semi-graphical approach that was also formulated using Pontryagin's minimum principle for oversaturated intersections. Nevertheless, the introduction of the *kinematic shockwave* theory (also called LWR or shockwave theory) Lighthill and Whitham (1955) and Richards (1956) provided an alternative method for modelling and analyzing the traffic flow on arterial roads (see e.g. Dion et al., 2004; Skabardonis and Geroliminis, 2008; Liu et al., 2009; Cheng et al., 2011; Ramezani and Geroliminis, 2015). Estimation of delay at intersections using the shockwave theory was first presented in Michalopoulos and Pisharody (1981). Dion et al. (2004) present a comprehensive literature review on different approaches to model delay including deterministic queuing algorithms (Haddad et al., 2010; Aboudolas et al., 2010; Ioslovich et al., 2011; Varaiya, 2013a, 2013b; Haddad et al., 2014), shockwave based models (Michalopoulos et al., 1981; Ban et al., 2009; Zheng et al., 2017; Wada et al., 2017), and microscopic simulation based models (Dowling et al., 2004).

The literature on the optimization of traffic signals can be classified into studies that focus on optimizing (i) a single isolated intersection (Lee et al., 2017; Wang et al., 2017; Yu et al., 2017; Yu et al., 2018), (ii) multiple coordinated intersec-

\* Corresponding author.

https://doi.org/10.1016/j.trb.2018.11.004 0191-2615/© 2018 Elsevier Ltd. All rights reserved.







E-mail address: mohsen.ramezani@sydney.edu.au (M. Ramezani).

Nomenclature						
q <sub>i</sub> q <sub>i</sub> k <sub>i</sub> k <sub>i</sub> k <sub>i</sub>	The arrival flow of Approach <i>i</i> [veh/unit time] The saturation flow (or capacity) of Approach <i>i</i> [veh/unit time] The arrival traffic density of Approach <i>i</i> [veh/unit distance] The saturation (critical) traffic density of Approach <i>i</i> [veh/unit distance]					
k <sup>j</sup> R <sub>i</sub> , G <sub>i</sub> C L <sub>i</sub>	The jam density of Approach <i>i</i> [veh/unit distance] Respectively, the red time and green time of Approach <i>i</i> during a cycle [unit time] The cycle time [unit time] The loss time of Approach <i>i</i> [unit time]					
$\Delta_i$ $x_i^j \text{ and } t_i^j$ $\hat{i} \in \Omega$ $\delta(t)$ $H(t)$	The length of Approach <i>i</i> [unit length] The position and the time of queue clearance at Approach <i>i</i> , respectively The complement of <i>i</i> in set $\Omega$ The Dirac delta function which is infinity at $t = 0$ and zero otherwise The Heaviside step function which is zero for $t \le 0$ and one otherwise					

tions (Dinopoulou et al., 2006; Kosmatopoulos et al., 2007; Ramezani et al., 2017; Wada et al., 2017), and (iii) network-wide signal control (e.g. Diakaki et al., 2002; Keyvan-Ekbatani et al., 2015; Kouvelas et al., 2017; Ramezani et al., 2015; Aalipour et al., 2018). The main outcome of network-wide signal control strategies (e.g. perimeter control) is to manage the congestion level of different regions of the network such that each region remains uncongested. The perimeter control is based on macroscopic and large-scale models and is expected to enable undersaturated traffic conditions at the intersections. Thus to complement the perimeter control, this paper studies local-level (i.e. intersection) signal control while offering a new focus on reliability measures.

Haddad et al. (2010, 2014) proposed a discrete-event max-plus approach to model the traffic flow at a two-way isolated undersaturated intersection and considered a weighted sum of queue lengths at the intersection as the optimization criterion. In addition, the above theory is applied in loslovich et al. (2011) to optimize an oversaturated intersection assuming a continuous-time model as in Gazis (1964). Moreover, Varaiya (2013b) proposed the concept of max-pressure and applied it to the *store and forward* queuing model to stabilize an arbitrary network of intersections under uncertain demands and turning ratios. This theory has been recently applied to control isolated intersections in Lee et al. (2017).

Despite the vast literature on signal timing optimization, very few studies have considered delay variability and spillback avoidance. The literature on the investigation of delay variability has mostly concentrated on the mathematical or numerical approaches to estimate the delay variability at a single approach considering stochastic arrival/departure flow rates (Zheng et al., 2017; Fu and Hellinga, 2000; Chen et al., 2016). Zheng et al. (2017) proposed a LWR theory based approach to mathematically model the delay distribution at a single approach of two adjacent signalized intersections addressing deterministic and stochastic upstream flow rates and signal coordination effects. Very recently, Wada et al. (2017) proposed a two-level mixed-integer linear programming optimization algorithm based on the variational LWR model of traffic flow to adjust the timings and offsets of coordinated traffic signals for deterministic traffic conditions and their stochastic extensions.

Spillback avoidance has been the subject of a few studies in the literature (see e.g. Ramezani et al., 2017; Liu and Chang, 2011; Li and Zhang, 2014). Ramezani et al. (2017) proposed a queue spillback avoidance control framework for an arterial based on the link partitioning approach. Liu and Chang (2011) proposed a heuristic genetic algorithm-based signal optimization framework with the priority given to minimizing the possibility of spillback occurrence. Moreover, Li and Zhang (2014) attempted to minimize the probability of spillback using a linear program with an objective representing the spillback's risk, assuming a prefixed cycle length. Those works employ the queuing theory for estimating the queue length at the intersection. Nevertheless, studies based on the store and forward queuing model do not provide full spatial and temporal characteristics of queuing dynamics. This is particularly essential to accurately analyze the possibility of spillback, since when the queue at an approach exceeds the detector's location, models based on the queuing theory face observability problems (Liu et al., 2009; Cheng et al., 2011).

This paper investigates the traffic signal control problem using the shockwave theory. We present an analytical model to estimate the delay variability of queued vehicles, and the queue lengths at the intersection. Using the model, we formulate the optimal cycle time and the green phase allocations at the intersection considering constant flow rates and taking loss times into account. Furthermore, prefixed (given) cycle length scenarios are also studied. We optimize the signal timing by minimizing the following objective functions: (i) delay variability, and (ii) probability of spillback. The delay distribution at the intersection is established, and the variance of delay is introduced as a criterion for increasing the performance reliability of the signal control. Furthermore, undersaturation condition and spillback avoidance are formulated to define the optimization constraints. Accordingly, each problem is formulated as a convex program, and descriptive algorithms are proposed to achieve the analytical *global optimal* signal setting for various optimization scenarios.

The rest of the paper is organized as follows. Preliminaries of shockwave theory and delay estimation are given in Section 2. The opimization algorithms for dynamic and prefixed cycle length strategies are given in Section 3, where several frameworks for minimizing the delay variability and probability of spillback objectives are proposed in Sections 3.2 and



**Fig. 1.** (a) The time-space, and (b) the fundamental diagrams of each approach of a two-way undersaturated intersection. The generated shockwaves are shown by thick lines. The minimum effective green time at each approach to maintain the undersaturation property is depicted, and the effect of loss times in enlarging the vehicle delay at each approach is highlighted. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.3, respectively. Comprehensive microsimulation and numerical studies are conducted in Section 4 to highlight the importance and efficacy of various proposed signal optimization algorithms. The summary and future directions are provided in Section 5.

#### 2. Preliminaries

Variable  $t_i \in [0, t_i^j]$ ,  $i = \{1, 2\}$  is defined as the time that a vehicle joins the queue at Approach *i*. We use the time-space diagram (TSD) that contains the trajectory of every vehicle entering and exiting the system, and the traffic flow fundamental diagram (FD). The TSD and FD are related through the definition of shockwave (Lighthill and Whitham, 1955). Fig. 1 depicts the TSD, FD, and shockwaves at a two-phase homogeneous and undersaturated intersection.

To derive the analytical formulations of the paper, we assume: [A] infinite acceleration; [B] the arrival and discharge flows and speeds of each approach are constant and known during the cycle time; [C] a triangular FD (Fig. 1(b)); [D] a two-phase intersection; and [E] the loss times at the intersection are known and constant. For simplicity in notations we first study a two-phase intersection, then extend the algorithms to multi-phase intersections. Note that the arrival flow rates can vary cycle by cycle and Assumption [B] is not a limiting assumption for an isolated intersection. However, the assumption may not be valid for closely spaced intersections when the spillback is imminent Haddad and Mahalel (2014). This motivates us to also investigate the spillback and design the signal timings to minimize the probability of spillback. In the microsimulation case study (Section 4.1), the above assumptions are relaxed, which enables to scrutinize the effect of the assumptions on the theoretical outcomes.

From Fig. 1 (assuming zero loss times), the queue clearance point can be formulated as follows:

$$t_i^{\mathbf{J}} = \beta_i (R_i + L_i), \tag{1}$$

$$x_i^j = \frac{q_i^a}{k_i^j - k_i^a} t_i^j,\tag{2}$$

where

$$eta_i riangleq rac{k_i^c ig(k_i^j-k_i^{\mathfrak{a}}ig)}{k_i^j ig(k_i^c-k_i^{\mathfrak{a}}ig)} > 1.$$

Formulas (1) and (2) are obtained based on the fact that the speed of arrival and discharge shockwaves at Approach *i* are the slopes of lines  $A_i J_i$  and  $C_i J_i$  in the FD depicted in Fig. 1b, respectively.



Fig. 2. (a) The description of the notations and auxiliary axes used in the proof of Theorem 1 on the TSD; (b)-(d) The delay PDFs at each approach and the whole intersection.

The delay experienced by each individual vehicle can be formulated based on the time and position of the vehicle when joining the queue ( $t_i$  and  $x_i$ ) from Fig. 1. Given that a vehicle that joins the back of the queue at time  $t_i$  and position  $x_i$  discharges the queue at time  $\tilde{t}_i > t_i$ , simple trigonometric calculations based on the shockwave theory can be carried out, together with previously derived formulas (1) and (2) to obtain the delay experienced by the vehicle from formulas below:

$$D_i(t_i) = -\frac{t_i}{\beta_i} + R_i + L_i \qquad \forall t_i \in [0, t_i^j].$$
(3a)

$$D_{i}(x_{i}) = -\frac{k_{i}^{j} - k_{i}^{a}}{\beta_{i}q_{i}^{a}}x_{i} + R_{i} + L_{i} \qquad \forall x_{i} \in [0, x_{i}^{j}].$$
(3b)

If the cycle length is prefixed (given and equal to *C*), the decision variable in the optimization framework is either  $R_1$  or  $R_2$ . Let us assume hereafter that  $R_1$  is the decision variable of interest. Therefore, (1)–(3) can be rewritten, substituting  $R_2 = C - R_1$ .

## 3. Signal optimization

We formulate the optimization problem using two alternatives objective functions: (i) the variance of delay distribution and (ii) the probability of spillback, taking into account a spillback avoidance constraint. The first objective is a measure to model and minimize travel time variability at the intersection.

#### 3.1. Formulating the constraints

Given the red time  $R_i$ , to discharge the queue at Approach *i* the minimum green time of the approach must fulfill the following inequality, which is obtained after some trigonometric calculations based on Fig. 1 :

$$G_{i}^{\min} + L_{i} = R_{\hat{i}}^{\min} \ge t_{i}^{j} - R_{i} + \frac{k_{i}^{a}}{q_{i}^{a}} x_{i}^{j} \quad i = \{1, 2\},$$
(4)

where  $\hat{i} \in \{1, 2\}$  is the complement of i in the set  $\{1, 2\}$  (e.g. if i = 1, then  $\hat{i} = 2$ ). Substituting (1) and (2) into (4) it is obtained that

$$R_{\hat{i}} - \eta_i (R_i + L_i) - L_i \ge 0. \tag{5}$$

where

$$\eta_i \triangleq rac{k_i^{\mathrm{a}}}{k_i^{\mathrm{c}} - k_i^{\mathrm{a}}} > 0 \qquad i = \{1, 2\}.$$

In addition, it is well-known and can be shown from the queuing theory that the necessary and sufficient condition for keeping an intersection undersaturated is (Gazis, 1964):

$$\sum_{i=1,2} \frac{q_i^a}{q_i^c} + \frac{L}{C} \le 1.$$
(6)

As such, for an intersection with bounded loss times controlled by a dynamic cycle length, the necessary and sufficient condition for undersaturation is expressed as

$$\sum_{i=1,2} \frac{q_i^a}{q_i^c} < 1.$$
(7)

Taking the speed-flow relationship in the FD into account, Condition (7) is equivalent to  $\eta_1\eta_2 < 1$ .

To avoid the spillback phenomena at Approach *i*, the position of the back of the queue at the clearance time, i.e.  $x_i^{j}$ , should be less than or equal the length of the approach. This constraint can be expressed as the following inequality :

$$R_i + L_i - k_i^j \left(\frac{1}{q_i^a} - \frac{1}{q_i^c}\right) \Delta_i \le 0,$$
(8)

where  $\Delta_i$  is the length of Approach  $i \in \{1, 2\}$ .

Under the prefixed cycle length control strategy, constraints (5) and (8) are reformulated as follows (replace  $R_2 = C - R_1$ ):

$$(1+\eta_1)R_1+\eta_1L_1+L_1-C \le 0 \tag{9a}$$

 $(1+\eta_2)R_1 - \eta_2(C+L_2) - L_2 \ge 0 \tag{9b}$ 

$$R_1 + L_1 - k_1^{j} \left( \frac{1}{q_1^a} - \frac{1}{q_1^c} \right) \Delta_1 \le 0$$
(10a)

$$R_1 - L_2 + k_2^{j} \left(\frac{1}{q_2^a} - \frac{1}{q_2^c}\right) \Delta_2 - C \ge 0$$
(10b)

Considering constant and bounded free flow speed of vehicles, there should be at least a minimum amount of time for every vehicle as well as pedestrian to pass through the intersection that is represented by

$$R_i^{\min} \le R_i. \tag{11}$$

## 3.2. Minimizing the delay variability

#### 3.2.1. Optimized cycle length

The distribution of delay can provide crucial information about the reliability of the signal control approach. A traffic signal control that minimizes this objective provides candid green times to all approaches. As such, vehicles on minor approaches will experience more equitable delays compared to the case when the total delay at the intersection is minimized. In light of that, the forthcoming theorem summarises our contributions on this problem using the shockwave theory. It is emphasized that the *continuous* model of the distribution of delay is studied, which is a more convenient layout with respect to the discrete model for optimization purposes.

**Theorem 1.** The following statements hold for a two-phase undersaturated intersection wherein Assumptions [A–E] are satisfied:



**Fig. 3.** The feasibility area (shaded) and the solution domains (bold solid lines) of the delay variability optimization problem assuming no loss time at the intersection: (a,b) Case (1); (c) Case (2). The green dashed line represents  $R_1 - R_2 = 0$ ; and  $R_i^{max} = k_i^j \left(\frac{1}{q_i^2} - \frac{1}{q_i^2}\right) \Delta_i$ . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(a) The delay probability density function (PDF) of Approach i,  $i = \{1, 2\}$ , can be expressed as

$$f_{D_i}(d) = \frac{n_i}{Cq_i^a} \delta(d) + \theta_i (H(d) - H(d - R_i - L_i)),$$
(12)

where  $\theta_i = \frac{\beta_i k_i^j}{C(k_i^j - k_i^3)}$ ,  $\delta(\cdot)$  is the Dirac delta function,  $d \ge 0$  is the delay experienced by a vehicle, and  $\tilde{n}_i$  is the estimated number of undelayed vehicles at Approach i:

$$\tilde{n}_i = Cq_i^a - q_i^a \varrho_i \beta_i (R_i + L_i), \tag{13}$$

where  $Q_i \triangleq \frac{k_i^j}{k_i^j - k_i^a}$ .

(b) Without loss of generality if Approach *i* is the dominant approach and thus  $R_{\hat{i}} + L_{\hat{i}} \ge R_i + L_i$ , then for every  $d \ge 0$  the delay PDF at the whole intersection can be expressed as

$$f_D(d) = \zeta_1 \delta(d) + \zeta_2 (H(d) - H(d - R_i - L_i)) + \zeta_3 (H(d - R_i - L_i) - H(d - R_{\hat{i}} - L_i)),$$
(14)

where

$$\zeta_{1} = \frac{\dot{n}_{1} + \dot{n}_{2}}{C(q_{1}^{a} + q_{2}^{a})},$$
  

$$\zeta_{2} = \sum_{i=1}^{2} \frac{\beta_{i}\varrho_{i}q_{i}^{a}}{C(q_{1}^{a} + q_{2}^{a})},$$
  

$$\zeta_{3} = \frac{\beta_{i}\varrho_{i}q_{i}^{a}}{C(q_{1}^{a} + q_{2}^{a})}.$$

(c) The variance of delay at the intersection assuming that Approach i is the dominant approach is

$$D_{\text{var}} = \sigma^2 [f_D] = \zeta_1 \bar{D}_{\mu}^2 + \frac{1}{3} (\zeta_2 - \zeta_3) \left( R_i + L_i - \bar{D}_{\mu} \right)^3 + \frac{1}{3} \zeta_2 \bar{D}_{\mu}^3 + \frac{1}{3} \zeta_3 \left( R_{\hat{i}} + L_{\hat{i}} - \bar{D}_{\mu} \right)^3, \tag{15}$$

where  $\bar{D}_{\mu}$  is the expected value of delay that is obtained from the following formula:

$$D_{\mu} = \mathbb{E}[f_D] = 0.5\zeta_2(R_i + L_i)^2 + 0.5\zeta_3((R_i + L_i)^2 - (R_i + L_i)^2).$$
(16)

**Proof.** The proof is given in Appendix A.

Given that cycle time is the summation of the decision variables  $(R_1 \text{ and } R_2)$ ,  $\zeta_1$ ,  $\zeta_2$ ,  $\zeta_3$ , and  $\bar{D}_{\mu}$  are nonlinear. Hence,  $D_{\text{var}}$  is also a nonlinear function of  $R_i$ ,  $i = \{1, 2\}$ . Furthermore, investigating the convexity of the variance function (see Figs. 10 and 11 in Section 4) confirms that  $D_{\text{var}}$  is a *non-convex* function of  $R_i$ , thus, the infimum of the objective function is *not* necessarily on the border of the feasibility set. This makes the analytical calculation of the optimal solution cumbersome. Nevertheless, it can be shown after carrying out comprehensive mathematical manipulations that given  $R_{\hat{i}} + L_{\hat{i}} \ge R_i + L_i$ ,  $D_{\text{var}}(R_1, R_2)$  is strictly increasing with respect to  $R_{\hat{i}}$  and it is a convex function of  $R_{\hat{i}}$ . A proof of this statement is given in Appendix B. Hence, the optimal cycle length and red time allocations at the intersection based on the minimization of the variance function is on the line  $R_{\hat{i}} = R_{\hat{i}}^{\min}$  in the  $(R_1, R_2)$  plane.

Algorithm 1 provides an analytical optimization framework assuming zero loss times for simplicity in the explanations. Algorithm 6 given in Appendix C is the extension of the algorithm addressing the loss times at the intersection.

Algorithm 1 Signal optimization for minimizing the delay variability at the intersection assuming optimized cycle length.

- **I.** If inequality (6) is realized, there exists a global optimal solution and one may proceed to the next step. Otherwise, the intersection is over-saturated and the algorithm terminates with no solution.
- **II.** Calculate  $\eta_1$  and  $\eta_2$  according to their definitions and identify the category of the optimization: (1) if  $\eta_{\hat{i}} \ge 1$ ,  $i, \hat{i} \in \{1, 2\}$ , proceed to Step III; and (2) if  $\eta_1 < 1$  and  $\eta_2 < 1$ , jump to Step IV. **III.** It can be assumed that  $R_{\hat{i}} \le R_i$ . If  $\eta_{\hat{i}}R_{\hat{i}}^{\min} \ge R_i^{\min}$ , then the global optimal solution is  $(R_{\hat{i}}^{\min}, \eta_{\hat{i}}R_{\hat{i}}^{\min})$  (the order may change

III. It can be assumed that  $R_{\hat{i}} \le R_{\hat{i}}$ . If  $\eta_{\hat{i}}R_{\hat{i}}^{\text{intri}} \ge R_{\hat{i}}^{\text{intri}}$ , then the global optimal solution is  $(R_{\hat{i}}^{\text{intri}}, \eta_{\hat{i}}R_{\hat{i}}^{\text{intri}})$  (the order may change depending on the value of  $\hat{i}$ ). The optimal point is demonstrated as the blue bold points in Fig. 3(a,b). Otherwise, the following optimization problem should be solved using a numerical approach such as the method of Lagrange multipliers:

$$\begin{array}{l} \underset{R_{i} = R_{i}^{\min}, R_{i}}{\min} \quad D_{var}(R_{\hat{i}}) \\ \eta_{i}R_{\hat{i}} - R_{i}^{\min} \leq 0, \\ -R_{\hat{i}} + R_{i}^{\min} \leq 0, \\ -R_{\hat{i}} + \eta_{i}R_{i}^{\min} \leq 0, \\ R_{\hat{i}} - k_{\hat{i}}^{j} \left(\frac{1}{q_{i}^{2}} - \frac{1}{q_{i}^{2}}\right) \Delta_{\hat{i}} \leq 0, \\ R_{i}^{\min} - k_{i}^{j} \left(\frac{1}{q_{i}^{2}} - \frac{1}{q_{i}^{2}}\right) \Delta_{\hat{i}} \leq 0. \end{array}$$

$$(O1-1)$$

The solution lies in the feasible closed set with  $R_i = R_i^{\min}$ , as demonstrated in the upper Fig. 3(a,b).

**IV.** Let us assume that  $R_i^{\min} \ge R_i^{\min}$ . The optimal solution can be found after solving the following problem (see Fig. 3(c)):

$\begin{array}{ll} \underset{R_i,R_i=R_i^{\min}}{\min} & D_{\text{var}}(R_i) \end{array}$	
$\eta_{\hat{i}} R_{\hat{i}}^{\min} - R_i \le 0,$	
$-R_i + R_i^{\min} \le 0,$	(01.2)
$R_i - R_{\hat{i}}^{\min} \leq 0,$	(01-2)
$R_i-k_i^{ m j}igg(rac{1}{q_i^{ m a}}-rac{1}{q_i^{ m c}}igg)\Delta_i\leq 0,$	
$R_{\widehat{i}}^{\min}-k_{\widehat{i}}^{\mathrm{j}}igg(rac{1}{q_{\widehat{i}}^{2}}-rac{1}{q_{\widehat{i}}^{2}}igg)\Delta_{\widehat{i}}\leq0.$	

#### 3.2.2. Prefixed cycle length

After substituting  $R_2 = C - R_1$  into (15), the following formula are obtained for the variance of delay:

$$D_{\text{varc}} = \begin{cases} (a.) \ \zeta_{1c}\bar{D}_{\mu_{c}}^{2} + \frac{1}{3}(\zeta_{2c} - \zeta_{3c})(R_{1} + L_{1} - \bar{D}_{\mu_{c}})^{3} + \frac{1}{3}\zeta_{3c}(C + L_{2} - R_{1} - \bar{D}_{\mu_{c}})^{3} + \frac{1}{3}\zeta_{2c}\bar{D}_{\mu_{c}}^{3} & R_{1} \leq \frac{C}{2} \\ (b.) \ \zeta_{1c}\bar{D}_{\mu_{c}}^{2} + \frac{1}{2}(\zeta_{2c} - \tilde{\zeta}_{3c})(C + L_{2} - R_{1} - \bar{D}_{\mu_{c}})^{3} + \frac{1}{2}\tilde{\zeta}_{3c}(R_{1} + L_{1} - \bar{D}_{\mu_{c}})^{3} + \frac{1}{2}\zeta_{2c}\bar{D}_{\mu_{c}}^{3} & R_{1} \geq \frac{C}{2} \end{cases}$$
(17)

where  $\bar{D}_{\mu_c} \triangleq 0.5\beta_1\varrho_1q_1^a(R_1+L_1)^2/N_T + 0.5\beta_2\varrho_2q_2^a(C-R_1+L_2)^2/N_T$ ;

 $\zeta_{1c} \triangleq 1 - q_1^a \varrho_1 \beta_1 (R_1 + L_1) / N_T - q_2^a \varrho_2 \beta_2 (C - R_1) / N_T$ ,  $\zeta_{2c} \triangleq \sum_{i=1}^2 q_i^a \varrho_i \beta_i / N_T$ ,  $\zeta_{3c} \triangleq \beta_2 \varrho_2 q_2^a / N_T$ ,  $\tilde{\zeta}_{3c} \triangleq \beta_1 \varrho_1 q_1^a / N_T$ , and  $N_T \triangleq C(q_1^a + q_2^a)$ . The optimization problem for minimizing the delay variability at the intersection can thus be formulated as

(03)

$$\begin{array}{ll} \underset{R_1}{\mininitial} & D_{varc}(R_1) \\ \text{subject to:} \\ \tilde{R}_1^{\min} \leq R_1 \leq \tilde{R}_1^{\max} \end{array}$$

Investigating (17) indicates that  $D_{varc}$  is a continuous 6th degree polynomial function of  $R_1$ , which is non-convex. Hence, its derivative  $\dot{D}_{varc} = \partial D_{varc}/\partial R_1$  is a 5th order polynomial. Therefore,  $\dot{D}_{varc} = 0$  would have at most 5 different admissible solutions  $R^*_{varc} \in [\tilde{R}_1^{\min}, \tilde{R}_1^{\max}]$ . This observation results in Algorithm 2 for seeking the analytical global optimal solution of (O3) if  $\tilde{R}_1^{\min} \leq \tilde{R}_1^{\max}$ .

#### **Algorithm 2** Signal optimization for minimizing the delay variability at the intersection assuming prefixed cycle length.

- **I.** Calculate  $\eta_1$  and  $\eta_2$  according to their definitions and identify the category of the optimization: (1) if  $\eta_{\hat{i}} \ge 1$ ,  $\hat{i} \in \{1, 2\}$ , proceed to Step II; and (2) if  $\eta_1 < 1$  and  $\eta_2 < 1$ , jump to Step III.
- **II.** If  $\eta_1 \ge 1$  ( $\eta_2 \ge 1$ ), it can be assumed that  $R_1 \le C/2$  ( $R_1 \ge C/2$ ) and (17.a) ((17.b)) can be selected as the objective function. Then, the optimal solution can be found from the set of critical points { $\tilde{R}_{1nin}^{min}, R_{varc}^*, \tilde{R}_{1n}^{max}$ }.
- **III.** The optimization problem O3 should be solved once assuming (17.a) as the objective function and changing the constraint to  $\tilde{R}_1^{\min} \le R_1 \le C/2$ , and once considering (17.a) and constraint  $C/2 \le R_1 \le \tilde{R}_1^{\max}$  to this aim. The global optimal solution can then be selected as the one that returns a smaller delay variability.

## 3.3. Minimizing the probability of spillback

Signal optimization for minimizing the probability of spillback occurrence can be better scrutinized using the shockwave theory instead of the queuing theory. Spillover phenomena at an approach can happen due to various reasons including high arrival flow rates, or short link lengths. This is highly undesirable as spillback can propagate through the network and eventually result in gridlock (Mahmassani et al., 2013; Geroliminis and Skabardonis, 2011). Hence, strong attention should be devoted to minimizing the probability of spillover at a signalized intersection.

#### 3.3.1. Optimized cycle length

An appropriate objective function to this aim is the reciprocal of the left-hand-side of (8):

$$D_{\rm sb} = \sum_{i=1,2} \frac{\epsilon_i}{k_i^j \left(\frac{1}{q_i^3} - \frac{1}{q_i^c}\right) \Delta_i - R_i - L_i},\tag{18}$$

where  $\epsilon_i > 0$ ,  $\sum_{i=1,2} \epsilon_i = 1$ , are weighting gains. The function returns small positive values when the probability of spillback is low, and very large values as the back of the queue approaches to the end of the link. Hence, the optimization problem is reconfigured to  $(i, \hat{i} \in \{1, 2\})$ :

$$\begin{array}{l} \underset{R_1,R_2}{\text{minimize}} \quad D_{\text{sb}}(R_1,R_2) \\ \text{Subject to (5), (8), and (11)} \end{array}$$
(04)

An alternative objective function for minimizing the possibility of spillback can be defined as

$$D_{\rm sb_2} = \sum_{i=1,2} \frac{\epsilon_i (R_i + L_i)}{k_i^j \Delta_i \left(\frac{1}{q_i^3} - \frac{1}{q_i^c}\right)},\tag{19}$$

which is equivalent to  $\sum_{i=1,2} \epsilon_i x_i^j / \Delta_i$ . However,  $D_{sb}$  responds more strongly and abruptly to situations where the spillback probability increases to critical levels. This point is demonstrated in the simulation studies in Section 4.

Program (O4) is convex, and both  $D_{sb}$  and  $D_{sb_2}$  are strictly increasing functions of the optimization variables. It can be analytically shown using the method of Lagrange multipliers that the optimal solution lies on the border of the feasibility set, demonstrated in Fig. 4(a–c) for an intersection without loss time and in Fig. 4(d–g) for an intersection with loss time. In other words, the optimal solution is attained from two non-zero values of the Lagrange multipliers which correspond to two constraints that can also be identified graphically (Fig. 4). Therefore, four outcomes for the global solution of program (O4) are possible depending on the location of the point  $D^* \triangleq (R_1^{\min}, R_2^{\min})$  in the solution space. The solution can be sought from Algorithm 3. The point  $R^*$  defined in the algorithm is the intersection of the hyper lines associated with the constraints (5) that are depicted in Fig. 4.

It is evident that increase in loss time would increase the probability of spillback at the intersection. Fig. 4 demonstrates the effect of loss time on the feasibility region of the intersection. Furthermore, inequality (8), i.e. the spillback avoidance constraint, implies an upper-bound for the red time at each approach (see Fig. 4).



**Fig. 4.** The feasibility regions and possible optimal solution outcomes of the optimized cycle length control: (a-c) intersection without loss time and (d-g) intersection with loss times. The black dot point in each figure indicates the optimal signal setting, and  $R_i^{max} = k_i^2 \left(\frac{1}{q_i^2} - \frac{1}{d_i^2}\right) \Delta_i - L_i$ .

**Algorithm 3** Global optimal signal timing  $(R_1^{opt}, R_2^{opt})$  for minimizing the probability of spillback at the intersection assuming optimized cycle length.

- 1.  $R^{\star} \triangleq \left(\frac{\eta_2(\eta_1+1)}{1-\eta_1\eta_2}L_1 + \frac{\eta_2+1}{1-\eta_1\eta_2}L_2, \frac{\eta_1(\eta_2+1)}{1-\eta_1\eta_2}L_2 + \frac{\eta_1+1}{1-\eta_1\eta_2}L_1\right)$ , if the point  $D^{\star}$  lies on the left and below of the point  $R^{\star}$  (the bold point in Fig. 4(d)).
- 2.  $(R_1^{\min}, L_1 + \eta_1(R_1^{\min} + L_1))$ , if the point  $D^*$  lies on the right of the point  $R^*$  and right of the hyper line  $\eta_1(R_1 + L_1) R_2 + L_1 = 0$  (the dark bold point in Fig. 4(e)).
- 3.  $(L_2 + \eta_2(R_2^{\min} + L_2), R_2^{\min})$ , if the point  $D^*$  lies on the top of the point  $R^*$  and left of the hyper line  $-R_1 + L_2 + \eta_2$  $(R_2 + L_2) = 0$  (the dark bold point in Fig. 4(f)).
- 4.  $D^*$ , if the point  $D^*$  lies on the top of the point  $R^*$  and between the hyper lines  $-R_2 + L_1 + \eta_1(R_1 + L_1) = 0$  and  $\eta_2(R_2 + L_2) R_1 + L_2 = 0$  (the dark bold point in Fig. 4(g)).

Note that total delay at the intersection can be formulated as

$$D_T = \sum_{i=1,2} \int_0^{x_i^i} k_i^j D_i(x_i) dx_i$$
  
=  $\sum_{i=1,2} \gamma_i (R_i + L_i)^2.$  (20)

where  $\gamma_i \triangleq 0.5\beta_i \frac{q_i^a}{k_i^j - k_i^a} k_i^j$ , which is in agreement with the Webster's deterministic formula. Since  $D_T$  is also a convex function of the signal timing variables, the feasibility region for minimizing the total delay is as demonstrated in Fig. 4 and the analytical optimal solution can be obtained from Algorithm 3.

## 3.3.2. Prefixed cycle length

After replacing  $R_2$  by  $C - R_1$  in (18) and (19), objective functions  $D_{sb}$  and  $D_{sb_2}$  are respectively modified to:

$$D_{\rm sbc} = \frac{\epsilon_1}{\delta_1 - R_1} + \frac{\epsilon_2}{\delta_2 + R_1},\tag{21}$$



Fig. 5. The feasibility set of the solution space of the prefixed cycle length control scenario.

$$D_{\rm sbc_2} = \frac{\epsilon_1 (R_1 + L_1)}{k_1^j \Delta_1 \left(\frac{1}{q_1^a} - \frac{1}{q_1^c}\right)} + \frac{\epsilon_2 (C - R_1 + L_2)}{k_2^j \Delta_2 \left(\frac{1}{q_2^a} - \frac{1}{q_2^c}\right)},\tag{22}$$

where  $\delta_1 \triangleq k_1^j \left(\frac{1}{q_1^3} - \frac{1}{q_1^c}\right) \Delta_1 - L_1$  and  $\delta_2 \triangleq k_2^j \left(\frac{1}{q_2^3} - \frac{1}{q_2^c}\right) \Delta_2 - C - L_2$ . Moreover, the undersaturation and spillback avoidance constraints respectively follow (9) and (10) that imply lower-bounds and upper-bounds of  $R_1$ . Let us define  $\tilde{R}_1^{\min} \triangleq \max\{R_1^{\min}; (\eta_2(C+L_2)+L_2)/(1+\eta_2); C+L_2-k_2^j \left(\frac{1}{q_2^3} - \frac{1}{q_2^c}\right) \Delta_2\}$  and  $\tilde{R}_1^{\max} \triangleq \min\{R_1^{\max}; (C-\eta_1L_1-L_1)/(1+\eta_1); -L_1+k_1^j \left(\frac{1}{q_1^3} - \frac{1}{q_1^c}\right) \Delta_1\}$ . Accordingly, the optimization problem can be summarized as

$$\begin{array}{l} \underset{R_{1}}{\min \text{minimize}} \quad D_{\text{sbc}}(R_{1}) \left( \text{or } D_{\text{sbc}_{2}}(R_{1}) \right) \\ \text{subject to:} \\ \tilde{R}_{1}^{\min} \leq R_{1} \leq \tilde{R}_{1}^{\max} \end{array}$$
(05)

The feasible set of the optimization problem (05) is highlighted using a bold line in Fig. 5. Note that the necessary and sufficient condition for the existence of a solution for problem (05) is  $\tilde{R}_1^{\min} \le \tilde{R}_1^{\max}$ . The undersaturation constraint (6) and the spillback avoidance condition (8) are embedded in  $\tilde{R}_1^{\min} \le \tilde{R}_1^{\max}$  respectively through  $1/(1 + \eta_2)(\eta_2(C + L_2) + L_2) \le 1/(1 + \eta_1)(C - \eta_1 L_1 - L_1)$  and  $C + L_2 - k_2^{i}\left(\frac{1}{2} - \frac{1}{2}\right)\Delta_2 \le -L_1 + k_2^{i}\left(\frac{1}{2} - \frac{1}{2}\right)\Delta_1$ .

 $1/(1 + \eta_1)(C - \eta_1 L_1 - L_1)$  and  $C + L_2 - k_2^j \left(\frac{1}{q_2^3} - \frac{1}{q_2^c}\right) \Delta_2 \le -L_1 + k_1^j \left(\frac{1}{q_1^3} - \frac{1}{q_1^c}\right) \Delta_1$ . It is clear that  $D_{sbc2}(\cdot)$  is monotonic (i.e. strictly increasing or decreasing), and thus it is convex. Moreover, the convexity of  $D_{sbc2}(\cdot)$  can also be confirmed after taking the second derivative of it. Hence, if  $\tilde{R}_1^{\min} \le \tilde{R}_1^{\max}$ , using the method of Lagrange multipliers the analytical global optimal solution of (O5) can be obtained using Algorithm 4.

**Algorithm 4** Global optimal signal timing  $R_1^{\text{opt}}$  for minimizing the probability of spillback at the intersection with prefixed cycle length.

(23)

i. If  $D_{\rm sbc}$  is the objective function, the solution is:

1.  $R_{sh}^{\star}$ , which is the positive solution of

$$(\epsilon_1 - \epsilon_2)R_1^2 + 2(\epsilon_1\delta_2 + \epsilon_2\delta_1)R_1 + \epsilon_1\delta_2^2 - \epsilon_2\delta_1^2 = 0,$$

if 
$$\tilde{R}_1^{\min} \leq R_{ch}^{\star} \leq \tilde{R}_1^{\max}$$

2.  $\tilde{R}_1^{\min}$ , if  $R_{sb}^{\star} \leq \tilde{R}_1^{\min}$ .

3. 
$$\tilde{R}_1^{\max}$$
, if  $R_{ab}^{\star} \geq \tilde{R}_1^{\max}$ .

ii. If  $D_{\text{sbc}_2}$  is the objective function, the solution is:

1.  $\tilde{R}_{1}^{\min}$  if  $\tilde{\delta} \triangleq \frac{\epsilon_{1}}{\delta_{1}+L_{1}} - \frac{\epsilon_{2}}{\delta_{2}+C+L_{2}} > 0.$ 2.  $\tilde{R}_{1}^{\max}$  if  $\tilde{\delta} < 0.$ 3. Any  $R_{1} \in [\tilde{R}_{1}^{\min}, \tilde{R}_{1}^{\max}]$  if  $\tilde{\delta} = 0.$  Another objective could be primarily minimizing the total delay, while minimizing the probability of spillback. Since  $D_{sb}$  returns small values when the probability of spillback is minor and large values as it increases, a weighted sum of the total delay and spillback probability objective functions, i.e.  $D_{Tsbc} = \epsilon_3 D_{Tc} + \epsilon_4 D_{sbc}$  (with  $\epsilon_i > 0$ ,  $\sum_{i=3,4} \epsilon_i = 1$ ), can be chosen for the optimization, where  $D_{Tc} = \gamma_1 (R_1 + L_1)^2 + \gamma_2 (C + L_2 - R_1)^2$  formulates the total delay at the intersection. Alternatively,  $D_{sbc_2}$  could be employed in lieu of  $D_{sbc}$  in  $D_{Tsbc}$  to define  $D_{Tsbc_2}$ . After recalling the method of Lagrange multipliers, the analytical global optimal solution of (05) with  $D_{Tsbc}$  as the ob-

After recalling the method of Lagrange multipliers, the analytical global optimal solution of (O5) with  $D_{\text{Tsbc}}$  as the objective function is obtained following Algorithm 5. Moreover, if  $D_{\text{Tsbc}_2}$  is chosen as the objective function as a substitute of  $D_{\text{Tsbc}}$ , then  $R^*_{\text{Tsbc}}$  in Algorithm 5 should be replaced by

$$R_{\text{Tsbc}_2}^{\star} \triangleq \epsilon_3 R_1^{\star} + \epsilon_4 \delta/2.$$

**Algorithm 5** Global optimal signal timing  $R_1^{opt}$  for minimizing the weighted sum of the total delay and the probability of spillback at the intersection, represented as  $D_{Tsbc}$ , assuming prefixed cycle length.

1.  $R^{\star}_{Tshc}$ , which is the positive solution of

 $\begin{aligned} \epsilon_4(\epsilon_1 - \epsilon_2)R_1^2 + 2(\epsilon_4(\epsilon_1\delta_2 + \epsilon_2\delta_1) + \epsilon_3(\gamma_1 + \gamma_2))R_1 + \epsilon_4(\epsilon_1\delta_2^2 - \epsilon_2\delta_1^2) + \epsilon_3(\gamma_1L_1 - \gamma_2(C + L_2)) &= 0, \\ \text{provided that } \tilde{R}_1^{\min} \leq R_{1\text{sbc}}^\star \leq \tilde{R}_1^{\max}. \end{aligned}$ 2.  $\tilde{R}_1^{\min}$ , if  $R_{1\text{sbc}}^\star \leq \tilde{R}_1^{\min}$ .
3.  $\tilde{R}_1^{\max}$ , if  $R_{1\text{sbc}}^\star \geq \tilde{R}_1^{\max}$ .

For further analysis, let us assume that  $\epsilon_1 = \epsilon_2$ , and  $L_1 = L_2$ . Then the optimal red time to minimize the probability of spillback  $D_{\text{sbc}}$  is  $R_{\text{sb}}^* = 0.5(\delta_1 - \delta_2)$ , and  $\tilde{\delta} > 0(< 0)$  if and only if  $\delta_1 - \delta_2 < C(> C)$ . In other words, if  $\delta_1 - \delta_2 < C(> C)$  then taking  $D_{\text{sbc}}$  as the objective function, the optimal red time lies in  $\tilde{R}_1^{\min} \le R_{\text{sb}}^* < 0.5C$  ( $0.5C < R_{\text{sb}}^* \le \tilde{R}_1^{\max}$ ), and taking  $D_{\text{sbc}_2}$  as the objective function, the optimal red time is  $\tilde{R}_1^{\min}$  ( $\tilde{R}_1^{\max}$ ). Recalling that  $\delta_i$  is proportional to the link length  $\Delta_i$ , the term  $\delta_1 - \delta_2$  represents the relative link length difference of the two approaches. As such, if for instance Approach 1 has a relatively shorter link length, then it receives a larger green time to reduce the possibility of spillback using either criteria. However,  $D_{\text{sbc}_2}$  as the objective function devotes the highest possible green time to the approach, ignoring other factors that can have impact on the possibility of the spillback phenomenon.

In addition, if  $\delta_1 - \delta_2 = C$  that results in  $\bar{\delta} = 0$ , the objective  $D_{\text{sbc}_2}$  does not provide an optimal solution. On the other hand, minimizing  $D_{\text{sbc}}$  indicates that the optimal red time at each approach is exactly 0.5C, regardless of any symmetrical conditions at the approaches (e.g. arrival flow and density discrepancies).

**Remark 1.** As a critical comparison with the *relaxed cyclic discrete-event max-plus* (R-CDMP) signal optimization method proposed in Haddad et al. (2010, 2014), the following points are remarkable:

- The R-CDMP method is a optimized cycle length control approach minimizing an objective that is a function of the back of the queues in lieu of the green times at the intersection.
- The method is based on the queuing theory. Although queuing theory is not capable of providing an accurate estimation for  $x_i^j$ , the closest objective function to the one used in Haddad et al. (2010, 2014) is

$$J = \sum_{i=1}^{2} \omega_i x_i^j = \sum_{i=1}^{2} \omega_i q_i^a \beta_i / (k_i^j - k_i^a) R_i$$

The function is a strictly increasing function of its variables, hence, the global optimal solution for the optimization problem is unique and can be achieved from Algorithm 3. Indeed, three possible outcomes for the solution is expectable, depending on the minimum green times at each approach of the intersection (see Fig. 4(a-c)).

Hence, the results of R-CDMP algorithm can be achieved from Algorithm 4, with more accuracy in estimating the queue lengths coming from using the shockwave theory.

**Remark 2.** It can be shown that due to Assumption [C] the delay variability objective function (15) and total delay (20) are functions of arrival flows and saturation flows of the critical movements at the intersection, but not the jam densities. However, the probability of spillback objective functions (18) and (19) are in addition functions of jam densities and links lengths. This helps to increase the sensitivity of the probability of spillback objectives to the link lengths and capacity of the links in storing vehicles.



**Fig. 6.** (a) Time-space diagram of the shockwave generated at Phase  $m \ge i$  of the intersection during a cycle. The number of vehicles that experience a delay between  $R_{i-1}$  and  $R_i$  in this phase can be obtained using  $h_{i-1,i}^m$ . (b) Probability distribution of delay at an intersection with 4 phases assuming  $R_1 + L_1 \le \cdots \le R_4 + L_4$ .

#### 3.4. Extension of the algorithms to multiple phases

To show how we can address a more general and practical situation, this section is devoted to the extension of the proposed signal optimization algorithms to multiple phases. Let us assume that the intersection has in general  $N \ge 2$  phases and N critical lanes. It can be shown that the time and position of queue clearance at the critical approach i can be calculated from (1) and (2), respectively. Moreover, the spillback avoidance constraints (8) remain the same with  $i = \{1, \dots, N\}$ . However, from (4) the undersaturation constraints is modified noting that  $G_i + L_i = C - R_i = \sum_{i \neq i} R_i$ :

$$\sum_{j \neq i} R_j - \eta_i (R_i + L_i) - L_i \ge 0, \quad i, j \in \{1, \cdots, N\}.$$
(24)

In addition, it is clear that the necessary and sufficient condition of the undersaturation of the intersection is  $\sum_{i=1,\dots,N} \frac{q_i^2}{q_i^c} + \frac{L}{c} \leq 1$ .

Under the prefixed cycle length control scheme, we consider parameters  $R_1, \dots, R_{N-1}$  as the independent variables and  $R_N = C - \sum_{j \neq N} R_j$ . Accordingly, the undersaturation and spillback avoidance constraints can be modified to suit the fixed-cycle length control strategy.

#### 3.4.1. Minimizing the delay variability

Without loss of generality, let us assume that  $R_1 + L_1 \le R_2 + L_2 \le \cdots \le R_N + L_N$ , i.e. the first critical approach is the most dominant approach and the *N*'th critical approach is the least dominant one (the higher  $q_i^a/q_i^c$ , the more dominant the approach *i* is). Assuming constant and uniform arrival and departure flow rates, it is clear that the delay distribution at the intersection resembles Fig. 6(b) (for an intersection with 4 phases). Accordingly, the probability density function, expected delay, and variance of delay can be calculated as an extension of Theorem 1, that is illustrated in the following corollary.

**Corollary 1.** The following statements hold for an N-phase undersaturated intersection, if Assumptions [A–E] are fulfilled:

(a) The delay PDF of Approach  $i \in \{1, \dots, N\}$  can be expressed as

$$f_{D_i}(d) = \frac{\tilde{n}_i}{Cq_i^a}\delta(d) + \theta_i(H(d) - H(d - R_i - L_i)),$$
(25)

where  $\tilde{n}_i = Cq_i^a - q_i^a \varrho_i \beta_i (R_i + L_i)$ , and  $\theta_i$  and  $\varrho_i$  are as defined in Theorem 1.

(b) The delay PDF at the whole intersection can be written as (recall that  $d \ge 0$  is the delay experienced by a vehicle)

$$f_D(d) = \hat{\zeta}_0 \delta(d) + \sum_{i=1}^N \hat{\zeta}_i (H(d - R_{i-1} - L_{i-1}) - H(d - R_i - L_i)),$$
(26)

where  $R_0 = L_0 = 0$  and

$$\begin{aligned} \hat{\zeta}_0 &= \sum_{j=1}^N \frac{\tilde{n}_j}{Cq_j^{\mathrm{a}}}, \\ \hat{\zeta}_i &= \frac{\sum_{j=i}^N \beta_j \varrho_j q_j^{\mathrm{a}}}{C \sum_{j=1}^N q_j^{\mathrm{a}}} \quad i \in \{1, \cdots, N\}, \end{aligned}$$

and  $\beta_i$  and  $\varrho_i$  are defined in (1) and (13), respectively.

(c) The variance of delay at the intersection can be calculated from the following formula:

$$D_{\text{var}} = \hat{\zeta}_0 \bar{D}_E^2 + \frac{1}{3} \hat{\zeta}_1 \bar{D}_E^3 + \frac{1}{3} \sum_{i=1}^{N-1} (\hat{\zeta}_i - \hat{\zeta}_{i+1}) (R_i + L_i - \bar{D}_E)^3 + \frac{1}{3} \hat{\zeta}_N (R_N + L_N - \bar{D}_E)^3,$$
(27)

where  $\bar{D}_{E}$  is the expected value of delay at the intersection that is obtained from the formula below

$$\bar{D}_E = 0.5 \sum_{j=1}^{N} \hat{\zeta}_j \Big( (R_j + L_j)^2 - (R_{j-1} + L_{j-1})^2 \Big).$$
<sup>(28)</sup>

**Proof.** The corollary is proved in Appendix D.

Now that the delay variability at the intersection is obtained, the following nonlinear optimization program can be solved to obtain the associated optimum signal timings:

$$\begin{array}{l} \underset{R_{1},\cdots,R_{N}}{\text{minimize}} \quad D_{\text{var}}(R_{1},\cdots,R_{N}) \\ \text{Subject to (25), (8), and (11)} \end{array}$$
(06)

#### 3.4.2. Minimizing the probability of spillback

Objective functions  $D_{sb}$  and  $D_{sb_2}$  can be straightforwardly extended to multiple phases as

$$D_{\rm sb} = \sum_{i=1,\dots,N} \frac{\epsilon_i}{k_i^j \left(\frac{1}{q_i^3} - \frac{1}{q_i^c}\right) \Delta_i - R_i - L_i},\tag{29}$$

$$D_{\rm sb_2} = \sum_{i=1,\cdots,N} \frac{\epsilon_i (R_i + L_i)}{k_i^j \Delta_i \left(\frac{1}{q_i^3} - \frac{1}{q_i^c}\right)},\tag{30}$$

where  $\epsilon_i > 0$ ,  $\sum_{i=1,\dots,N} \epsilon_i = 1$ . Accordingly, the optimal signal timings ensuring the intersection's undersaturation, spillback avoidance, and minimum green times can be obtained solving the following optimization problem:

$$\begin{array}{l} \underset{R_{1}, \dots, R_{N}}{\text{minimize}} \quad D_{\text{sb}}(R_{1}, \dots, R_{N}) (\text{or } D_{\text{sb}_{2}}) \\ \text{Subject to (25), (8), and (11)} \end{array}$$
(07)

**Remark 3.** Note that the probability of spillback ( $D_{sb}$  and  $D_{sb_2}$ ) and total delay objective functions for a multi-phase undersaturated intersection are convex, thus their optimization will efficiently result in their global optimal solutions. However, the delay variability ( $D_{var}$ ) objective function (27) is not convex. Therefore, using nonlinear optimization programs do not necessarily result in the global optimal solution. However, if the initial condition for optimization is selected appropriately, achieving the global optimal solution can be expected, since  $D_{var}$  is a smooth and continuous function of signal timings. As we do not expect a huge difference between the optimal solution minimizing the total delay and delay variability, a suitable starting point for optimizing the delay variability is the optimal signal timing for the total delay minimization problem.

#### 4. Numerical examples

This section provides a microsimulation case study and two numerical case studies to evaluate the proposed formulas and signal optimization algorithms. Case study 1 is a microsimulation study aiming at validating the theoretical models developed in the paper regarding the delay variability at the intersection. Case study 2 studies the optimization algorithms following the dynamic cycle length signal setting, and case study 3 is devoted to the proposed signal optimization frameworks for the prefixed cycle length control scenarios.

#### 4.1. Case study 1: microsimulation study

Microsimulation is an effective tool in validating the proposed models for calculating the distribution and variance of vehicles delays as well as the total delay formula at an intersection. This study is crucial to capture the effects of the stochasticities of the arrival and departure flow rates, as well as acceleration/deceleration of vehicles on the proposed models



**Fig. 7.** Comparison of the total delays at the hypothetical intersection obtained from the microsimulation model and theoretical model (20). (a) Total delays are obtained for the intersection by fixing the signal timings at  $R_1 = R_2 = 60$  [s] and  $L_1 = L_2 = 4$  [s], and changing the arrival flow rates  $q_2^a = 1/3q_1^a \in [300, 700]$  [veh/h]. (b) Total delays are obtained for the intersection given fixed arrival flow rates  $q_2^a = 1/3q_1^a = 700$  [veh/h] and diversified signal timings  $R_1 = R_2 \in [60, 90]$  [s].

that have been established based on some simplifying assumptions (Assumptions [A-D]) including deterministic flows and infinite acceleration/deceleration of vehicles.

As such, a microsimulation study was conducted in Aimsun environment Aimsun (2014) on a hypothetical two-leg intersection with two approaches. The major and minor approaches had three and one lanes, respectively, and the average capacity flows and densities were measured to be  $q_2^c = 1/3q_1^c = 1800$  [veh/h] and  $k_2^c = 1/3k_1^c = 90$  [veh/km]. The arrival and departure flow rates were stochastic and the program takes the dynamics of every vehicle into consideration to replicate a real-world traffic flow. Simulation time-step was 0.9 s and to obtain a valid average response, the results for every experiment were aggregated over 10 replications that each took 30 min duration.

Firstly, the total delay experienced by vehicles were examined under two cases to make sure the two models (microsimulation and theoretical) comply under different arrival flows and signal settings: (a) fixing signal timings  $R_1 = R_2 = 60$  [s] and  $L_1 = L_2 = 4$  [s] and changing the arrival flow rates  $q_2^a = 1/3q_1^a \in [300, 700]$  [veh/h], and (b) fixing the arrival flows  $q_2^a = 1/3q_1^a = 700$  [veh/h] and changing the signal settings  $R_1 = R_2 \in [60, 90]$  [s]. The comparison of the results obtained from microsimulation and the total delay formula (20) in Fig. 7 suggests an excellent agreement between the two models.

Secondly, the distribution of vehicles delays were measured from the microsimulation model. To this aim, arrival flow rates, loss times and cycle time were fixed at  $q_1^a = 1200$  [veh/h] and  $q_2^a = 500$  [veh/h],  $L_1 = L_2 = 5$  [s] and C = 120 [s], respectively. Microsimulation experiments were conducted for various green times for the major and minor approaches. The uniform delay PDF at the intersection for each experiment is demonstrated in Fig. 8, where the results are compared against the PDFs obtained from our theoretical model (12). The delay PDFs of the vehicles in the microsimulation study are obtained following the same terminology of the theoretical model, wherein equal probability for delays larger than zero and less than the maximum delay experienced by vehicles on the approach that receives a lower red time (let us call it  $d_1^M$ ) is assumed. Similarly, an equal probability for delays larger than  $d_1^M$  and lower than the maximum delay experienced by vehicles ( $d_2^M > d_1^M$ ) is assumed. In other words, three bins are considered in deriving the delay PDFs: (i) d = 0, (ii)  $0 < d \le d_1^M$  and (iii)  $d_1^M < d \le d_2^M$ . Moreover, standard deviations (STDs) of delays obtained for various microsimulation experiment are compared against STDs obtained from our model (square root of (15)) in Fig. 9.

It can be observed from Fig. 8 that ignoring the top left sub-figure (associated with  $G_1 = 75$  [s]) the maximum delay experienced by vehicles on each approach in the microsimulation studies are consistently slightly higher than their expected value that is the effective red time ( $R_i + L_i$ ) of the approach. This difference corresponds to acceleration/deceleration effects and the time required for each vehicle to pass the intersection and enter the next road segment, as well as variability (stochasticity) in arrival flow rates. The same reasoning applies to the observation that the model (12) consistently slightly overestimates (by about 10%) the number of non-stopped vehicles (microsimulation results in the figure. However, the figure confirms that  $\zeta_3$  values obtained from the theory and microsimulation are almost equal.

In addition, according to Fig. 9 STDs of delays obtained from the theoretical model and microsimulation are close and in agreement for experiments with  $G_2 \ge 40$  [s]. However, for the first experiment with  $G_2 = 35$  [s], i.e. when the minor approach experiences its largest red time, there is a significant difference between the two models. This observation can be justified by the fact that the minor approach has a higher per lane arrival rate, and taking the stochasticity of the arrival rate this green time does not provide a sufficient margin for all the queued vehicles to pass the intersection. Therefore, occasionally there might be a vehicle that has to stop for the next cycle. This phenomena violates the undersaturation of the intersection and results in larger errors in the theoretical model results. This reasoning can be further justified from Fig. 8,



**Fig. 8.** Probability distribution of vehicles delays at the intersection for various signal timings given fixed cycle length C = 120 [s] and arrival flow rates  $q_1^a = 1200$  [veh/h] and  $q_2^a = 500$  [veh/h] obtained from the microsimulation model and theoretical model (12). It can be seen that the number of undelayed vehicles obtained from microsimulation is consistently lower than the theoretical model  $\tilde{n}_1 + \tilde{n}_2$ . Moreover, the maximum delays experienced by vehicles on each approach are generally slightly higher than the effective red times of the approach. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 9.** Standard deviation of delay at the intersection obtained from the microsimulation model and the theoretical model (square root of (15)). The cycle time was fixed at C = 120 [s] and the arrival flow rates were constant  $q_1^a = 1200$  [veh/h] and  $q_2^a = 500$  [veh/h]. The results are obtained for various signal timings  $R_1 \in [45, 85]$  [s]. Note that at  $R_1 = 45$  [s] there is a higher disagreement between the microsimulation and theoretical results due to formation of occasional residual overflow queues at the minor approach in the microsimulations.

where it can be seen that the difference between the maximum delays obtained from the two models in this experiment is much larger than the other experiments.

#### 4.2. Case study 2: Optimized cycle length signal settings

Here, we consider three scenarios with varying arrival flow rates at an intersection with fixed parameters given in case study 1. In scenario (i), Approach 2 is dominant with  $q_1^a = 300$  [veh/h] and  $q_2^a = 1100$  [veh/h]. In scenario (ii), Approach 1 is dominant with  $q_1^a = 1000$  [veh/h] and  $q_2^a = 200$  [veh/h]. In scenario (iii), there is no significantly dominant approach with  $q_1^a = 700$  [veh/h] and  $q_2^a = 600$  [veh/h].

Assuming 20 s of minimum red time at each approach and  $L_1 = L_2 = 10$  [sec], the feasibility region for each scenario is shown in Fig. 10. The link length of each approach is 210 [m]. The optimal signal setting, the queue clearance point, and the total delay for each scenario are summarized in Table 1, where the results are compared with the prefixed cycle length



(a) Scenario (i): Approach 2 is dominant



(b) Scenario (ii): Approach 1 is dominant



(c) Scenario (iii): no dominant approach

**Fig. 10.** Left: the feasibility regions and the  $D_{var}$  contour map in the  $(R_1, R_2)$  space, and Right: delay variability in 3D space for various scenarios. Filled circles demonstrate the point  $R^*$  (in black) and the optimal point minimizing the total delay (in orange). Minimum red times are shown by dotted pink lines. The red dashed lines and blue lines in the left hand side figures demonstrate the boundaries of the undersaturation condition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

control strategy adapting the Webster's deterministic formula to obtain the cycle length. In each scenario the optimized cycle length control algorithm (Algorithm 3) is significantly more efficient than the prefixed cycle length control strategy in reducing the total delay and queue length at the intersection.

It can be seen from Table 1 that the prefixed cycle length control strategy cannot handle Scenario (i), which is due to spillback at Approach 2. Accordingly, total delay and the probability of spillback considerably increase when the optimization is based on the cycle length attained from the Webster's formula. Note that shorter cycle lengths than those obtained from Algorithm 3 result in the oversaturation of the intersection. It demonstrates that applying the proposed optimized signal optimization algorithms in lieu of the conventional Webster's formula, mitigate congestion by reducing the total delay and the probability of spillback at the intersection.

#### Table 1

Optimal signal settings for minimizing the total delay at the intersection in case study 2, and the resulted back of the queues  $(x_1^i, x_2^i)$  and total delay (TD). The results from optimized cycle length control algorithm (Algorithm 3) are compared against the prefixed cycle length strategy.  $C_w$  is the given cycle length obtained from the Webster's deterministic formula, and  $C^{opt} = R_1^{opt} + R_2^{opt}$ . n.a. = not applicable.

Scenario	Optimal cycle le		Prefixed cycle length					
	$(R_1^{\text{opt}}, R_2^{\text{opt}})$ [s]	$(x_1^j, x_2^j) \ [m]$	TD [veh.s]	$R_1^{\text{opt}}$ [s]	$(x_1^{j}, x_2^{j})$ [m]	$C_w/C^{opt} [s]/[s]$	TD [veh.s]	
	$L_1 = L_2 = 10$ [s]							
(i)	(65,25)	(46.4,140.3)	762	n.a.	n.a.	n.a.	n.a.	
(ii)	(20,47.5)	(89.3,21.2)	384	21.7	(96.7,37.2)	105/67.5	588	
(iii)	(34,38)	(77.3,67.9)	594	54.2	(123.3,128.1)	126/72	1494	
	$L_1 = L_2 = 5$ [s]							
(i)	(44.3,20)	(31.6,112.2)	366	70	(50,112.2)	90/64.3	528	
(ii)	(20,36.25)	(89.3,16.2)	246	20	(89.3,17.9)	60/56.2	258	
(iii)	(20,20.9)	(45.4,37.3)	186	31.1	(70.6,73.1)	72/40.9	468	

#### Table 2

The optimal signal setting and queue clearance point  $(x_1^j, x_2^j)$  for minimizing the delay variability at the intersection in case study 2 (employing Algorithm 6), and for minimizing the total delay (Algorithm 3). Moreover, TD = Total Delay, and STD = Standard Deviation of delay at the intersection.

Objective function	Delay variability				Total delay and probability of spillback			
	$(R_1^{\text{opt}}, R_2^{\text{opt}})$ [s]	$(x_1^j,x_2^j)\ [m]$	TD [veh.s]	STD [veh.s]	$(R_1^{opt}, R_2^{opt})$ [s]	$(x_1^j,x_2^j)\ [m]$	TD [veh.s]	STD [veh.s]
Scenario	$(R_1^{\min}, R_2^{\min}) = (20, 20)[s]$							
(i)	(44.3,20)	(31.6,122.2)	367	10.4	(44.3,20)	(31.6,122.2)	367	10.4
(ii)	(20,36.2)	(89.3,16.2)	248.5	8.7	(20,36.2)	(89.3,16.2)	248.5	8.7
(iii)	(20,20.9)	(45.4,37.3)	183.3	7.5	(20,20.9)	(45.4,37.3)	183.3	7.5
Scenario	$(R_{1}^{\min}, R_{2}^{\min}) = (40, 10)[s]$							
(i)	(40,17.1)	(28.6,96.2)	293.8	9.5	(40,14)	(28.6,78.2)	243.1	9.8
(ii)	(40,61.2)	(178.6,27.3)	770	14.9	(40,61.2)	(178.6,27.3)	770	14.9
(iii)	(40,33.7)	(90.9,60.2)	509.8	13.3	(40,33.6)	(90.9,60.1)	508.8	13.3
Scenario	$(R_{1}^{\min}, R_{2}^{\min}) = (10, 40)[5]$							
(i)	(75.7,40)	(54.1,224.5)	$1.1 \times 10^3$	17.3	(75.7,40)	(54.1,224.5)	$1.1 \times 10^3$	17.3
(ii)	(19.2,40)	(85.9,17.9)	247.1	9.5	(10.6,40)	(47.4,17.9)	139.6	9.7
(iii)	(29.5,40)	(67.1,71.4)	442.7	12.6	(27.5,40)	(62.5,71.4)	421.2	12.8

To scrutinize the effect of loss times in the optimal signal settings, the study was repeated for an intersection with shorter  $L_1$  and  $L_2$  equal to 5 [s]. The results are reported in Table 1, where it can be observed that smaller loss times drastically reduces the total delay, as well as the queue lengths at the intersection.

Assuming  $L_1 = L_2 = 5$  [s], the optimal dynamic signal timings for minimizing the delay variability obtained from Algorithm 6 are summarized in Table 2. We examine three sets of minimum red times for each scenario. When  $R_1^{\min} = R_2^{\min} = 20$  [s], the optimal signal setting for minimizing the total delay and delay variability are similar. However, this is not the case for Scenario (i) when  $(R_1^{\min}, R_2^{\min}) = (40, 10)$  [s], Scenario (ii) when  $(R_1^{\min}, R_2^{\min}) = (10, 40)$  [s], and Scenario (iii) when either of the two latter minimum red times are initiated. In either of these cases, minimizing the delay variability results in a larger red time at one of the approaches, thus larger cycle length and larger total delay at the intersection. However, these changes result in less difference in the red times at the two approaches, and thus more equitable waiting times for drivers.

The intersection total delay variability is plotted in Fig. 10 both in 3D and in terms of isoline contours for different studied scenarios. The delay variability is also depicted in Fig. 11 in terms of various  $R_i$  and fixed  $R_{\hat{i}}$ ,  $i = \{1, 2\}$ . When  $R_i \ge R_{\hat{i}}$ , the variance of delay is strictly increasing with respect to  $R_i$ , but can be increasing or decreasing with respect to  $R_{\hat{i}}$ .

## 4.3. Case study 3: Prefixed cycle length signal optimization

In this case study, the cycle length is adopted as the Webster's deterministic formula, and various prefixed cycle length optimization algorithms are evaluated. We assume  $q_1^c = 1800$  [veh/h],  $q_2^c = 1400$  [veh/h],  $k_1^c = 30$  [veh/km],  $k_2^c = 20$  [veh/km], and  $k_1^j = k_2^j = 140$  [veh/km]. Analogous to the previous case study, three scenarios for the arrival flow rates and the links lengths are considered. In scenario (i), Approach 1 is dominant with  $q_1^a = 1000$  [veh/h],  $q_2^a = 200$  [veh/h],  $\Delta_1 = 200$  [m], and  $\Delta_2 = 50$  [m]. In scenario (ii), Approach 2 is dominant with  $q_1^a = 400$  [veh/h],  $q_2^a = 900$  [veh/h],  $\Delta_1 = 200$  [m], and  $\Delta_2 = 450$  [m]. In scenario (iii), there is no significantly dominant approach with  $q_1^a = 700$  [veh/h],  $q_2^a = 600$  [veh/h],  $\Delta_1 = 300$  [m], and  $\Delta_2 = 240$  [m].

Let us assume that  $R_1^{\min} = R_2^{\min} = 15$  [s],  $L_1 = L_2 = 10$  [s], and  $\epsilon_i = 0.5$ ,  $i = \{1, \dots, 4\}$ . Table 3 summarizes and compares the calculated optimal cycle length from Webster's formula, together with the feasibility region bounds, optimal red time

#### Table 3

Results of prefixed cycle length optimization problem for various scenarios in case study 3. Objective functions  $D_{\text{Tc}}$ ,  $D_{\text{sbc}}$  and  $D_{\text{sbc2}}$ ,  $D_{\text{Tsb}}$  and  $D_{\text{tsb2}}$ , and  $D_{\text{varc}}$  represent the total delay, probability of spillback, mixed total delay and spillback probability, and the delay variability at the intersection, respectively. Moreover, TD = Total Delay, SD = Standard Deviation of delay, PoS1 = Probability of Spillback (21), and PoS2 = Probability of Spillback (22) at the intersection.

Scenario	Boundaries							
		C [s]	$\widetilde{R}_1^{ ext{mi}}$	<sup>n</sup> [s]	$\tilde{R}_1^{\max}[s]$			
(i)		116.1	20	5.6	34.8			
(11)		259.4	1/2	9.4	186			
		191.7	92	2.2	107.2			
Objective								
Scenario	$R_1^{\text{opt}}[s]$	$(x_1^{J}, x_2^{J})[m]$	TD [veh.s]	SD [veh.s]	PoS1 [1/h]	PoS2		
(i)	26.6	(118.7,41.4)	738.8	21.2	430.1	0.87		
(ii)	186	(189.8,367.1)	$5.2 \times 10^{3}$	47.1	$\infty$	0.96		
(111)	92.2	(209.5,207.4)	$3.4 \times 10^{3}$	32.4	379.8	0.86		
Objective			D <sub>sbc</sub> (defined	in (21))				
Scenario	$R_1^{\text{opt}}[s]$	$(x_1^{j}, x_2^{j})[m]$	TD [veh.s]	SD [veh.s]	PoS1 [1/h]	PoS2		
(i)	26.6	(118.7,41.4)	738.8	21.2	430.1	0.87		
(ii)	182.7	(186.4,383.5)	$5.3  imes 10^3$	45.6	$1.1 \times 10^3$	0.97		
(iii)	104.3	(237,182.2)	$3.5 \times 10^{3}$	32.9	203.1	0.86		
Objective			$D_{\rm sbc2}$ (defined	in (22))				
Scenario	$R_1^{opt}[s]$	$(x_1^j,x_2^j)[m]$	TD [veh.s]	SD [veh.s]	PoS1 [1/h]	PoS2		
(i)	26.6	(118.7,41.4)	738.8	21.2	430.1	0.87		
(ii)	186	(189.8,367.1)	$5.2  imes 10^3$	47.1	$\infty$	0.96		
(iii)	107.2	(243.6,176.2)	$3.5 \times 10^{3}$	33.4	208.6	0.85		
Objective			$D_{\mathrm{Tsb}}$					
Scenario	$R_1^{\text{opt}}[s]$	$(x_1^j,x_2^j)[m]$	TD [veh.s]	SD [veh.s]	PoS1 [1/h]	PoS2		
(i)	26.6	(118.7,41.4)	738.8	21.2	430.1	0.87		
(ii)	179.4	(183.1,400)	$5.3 imes10^3$	44.2	$7.2  imes 10^{16}$	0.98		
(iii)	92.2	(209.5,207.4)	$3.4 \times 10^{3}$	32.4	379.8	0.86		
Objective	D <sub>Tsb2</sub>							
Scenario	$R_1^{opt}[s]$	$(x_1^j,x_2^j)[m]$	TD [veh.s]	SD [veh.s]	PoS1 [1/h]	PoS2		
(i)	34.8	(155.4,37.6)	897.1	19.3	$\infty$	0.9		
(ii)	179.4	(183.1,400)	$5.4  imes 10^3$	44.2	$7.2  imes 10^{16}$	0.98		
(iii)	92.2	(209.5,207.4)	$3.4  imes 10^3$	32.4	379.8	0.86		
Objective	D <sub>varc</sub> (defined in (17))							
Scenario	$R_1^{\text{opt}}[\text{sec}]$	$(x_1^j, x_2^j)[m]$	TD [veh.sec]	SD [veh.sec]	PoS1 [1/hr]	PoS2		
(i)	34.8	(155.4,37.6)	897.1	19.3	$5.5  imes 10^{10}$	0.9		
(ii)	179.4	(183.1,400)	$5.4 imes10^3$	44.2	$2.7 imes10^9$	0.98		
(iii)	95.6	(217.2,200.3)	$3.4  imes 10^3$	32.4	267.3	0.86		

signal at approach 1 estimated using various criteria proposed in Algorithms 2, 3, 5, and 6, and the back of the queues at the queue clearance time (i.e.  $x_i^j$ ,  $i = \{1, 2\}$ ). The delay variability is plotted for  $R_1 \in [\tilde{R}_1^{\min}, \tilde{R}_1^{\max}]$  for each scenario in Fig. 12 (a, c, and e). In this particular example, the variance function  $D_{varc}(R_1)$  is convex over the feasibility region in each scenario, such that in Scenario (i) and (ii) it is strictly decreasing and increasing, respectively. This is confirmed by the numerical solution given in Table 3. Moreover, it is observed that the optimal signal setting that minimizes the total delay is different from the optimal  $R_1$  that minimizes the variance of total delay in each scenario. Interestingly, the optimal signal timing that minimizes the delay variability creates longer queues at the dominant approach and tries to fairly distribute the green times between the conflicting approaches. The latter observation comes from the fact that the optimal  $R_1$  attained from minimizing  $D_{varc}$  is always closer to C/2.

In addition, the comparison of the optimal results from the two proposed objective functions for minimizing the probability of spillback indicates that the optimum red time for minimizing the second objective (i.e.  $D_{sbc2}$ ), which is a linear function of  $R_1$ , is always on the boundary of the feasibility region. This is not the case for the results obtained from the first objective (i.e.  $D_{sbc}$ ), where the possibility of spillback is more realistically captured. Applying the first spillback probability objective in the mixed total delay and spillback probability minimization criterion has lower influence on the optimal solution for minimizing the total delay. To provide further insights into this analysis,  $D_{sbc}$  and  $D_{sbc2}$  are plotted against  $R_1 \in [\tilde{R}_1^{\min}, \tilde{R}_1^{\max}]$  in Fig. 12 (b, d, and f) for various scenarios.



Fig. 11. The variation of delay variability with respect to red times in three studied scenarios of case study 2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

## 5. Summary and future works

The paper has presented an analytical framework based on the shockwave theory to formulate and minimize the delay variability at an undersaturated intersection. A closed-form formula for the variance of delay at the intersection has been obtained that is employed for travel time reliability optimization using convex optimization techniques. The analytical global optimal signal settings for minimizing various objectives comprising the delay variability and the probability of spillback have been achieved ensuring the spillback avoidance, and adapting prefixed (given) and optimized (dynamic) cycle length control strategies. Comprehensive microsimulation and numerical studies emphasized the effectiveness of the proposed signal optimization algorithms, as well as the superiority of dynamic cycle length over given cycle length control scenarios. Moreover, it has been shown that minimum delay variability results in a more equitable distribution of delays.

The theoretical findings of the paper can be further studied for signal optimization based on various criteria for *over-saturated* intersections. Furthermore, the proposed signal optimization frameworks can be extended to multiple coordinated signals to improve the travel time reliability on arterials and even the network level. Moreover, addressing stochastic arrival flows at the intersection is crucial, particularly when the intersection is operating near saturation conditions.

## Acknowledgment

The authors appreciate the reviewers for their constructive feedback in improving the paper's presentation and content.

## Appendix A. Proof of Theorem 1

(a) There are  $\tilde{n}_i \approx Cq_i^a - k_i^j x_i^j$  vehicles that do not stop at Approach *i*. Every other vehicle joins the queue and experiences a delay with a value in  $(0, R_i + L_i]$  (see Fig. 1(a)). Hence, the PDF of each approach can be approximated as an impulse function at d = 0 and a pulse function of length  $R_i + L_i$  at d > 0, as explained in (12) and depicted in Fig. 2(b-c). Moreover,



**Fig. 12.** (a,c,e) The variance of delay with respect to  $R_1 \in [\tilde{R}_1^{\min}, \tilde{R}_1^{\max}]$ , and (b,d,f) The comparison of the studied spillback probability objective functions for different scenarios of case study 3.

the coefficient of the impulse function  $\delta(t)$  should represent the probability that a vehicle at Approach *i* does not stop, which depends on  $\tilde{n}_i$ . Parameter  $\theta_i$  is obtained from the fact that the integration of the PDF function must be 1.

(b) The PDF at the whole intersection can be obtained from the superposition of the PDFs of the two approaches (see Fig. 2(d)) and (14). The term  $\zeta_1$ , i.e. the probability that a vehicle do not stop at the intersection, follows from the fact that the total number of vehicles and non-stopped vehicles in a cycle are  $C(q_1^a + q_2^a)$  and  $\tilde{n}_1 + \tilde{n}_2$ , respectively. Now, if one assumes  $R_{\hat{i}} + L_{\hat{i}} \ge R_i + L_i$  the possibility that a vehicle experience a delay between 0 and  $R_i + L_i$  is (see Fig. 2(a) for the representation of  $\Delta t_{h1}$ ):

$$\Pr[0 < d \le R_i] = \frac{q_i^a \varrho_i t_i^j + q_i^a \varrho_i(t_i^j - \Delta t_{h1})}{C(q_i^a + q_2^a)}$$

Furthermore, the probability of vehicle delays between  $R_i + L_i$  and  $R_i + L_i$  is:

$$\Pr[R_i < d \le R_{\hat{i}}] = \frac{q_{\hat{i}}^a \varrho_{\hat{i}} k_{\hat{i}}^j \Delta t_{h1}}{C(q_1^a + q_2^a)}.$$

The term  $\Delta t_{h1}$  can be obtained from h, where h is the queue length for vehicles that experience larger delays than  $R_i$ . To obtain h, let us define the speeds of the arrival and departure shockwaves at Approach  $i = \{1, 2\}$  as  $u_i$  and  $w_i$ , respectively. From the FDs the absolute values of the speeds of the shockwaves are

$$\tan(u_i) = \frac{q_i^{\mathrm{a}}}{k_i^{\mathrm{i}} - k_i^{\mathrm{a}}}, \quad \tan(w_i) = \frac{q_i^{\mathrm{c}}}{k_i^{\mathrm{i}} - k_i^{\mathrm{c}}}.$$

Thereafter, from Fig. 2(a) we have  $h/\Delta t_{h1} = \tan(u_{\hat{i}}), h/\Delta t_{h2} = \tan(w_{\hat{i}}), R_{\hat{i}} + L_{\hat{i}} + \Delta t_{h2} = R_i + L_i + \Delta t_{h1}$ , and

$$R_{\hat{i}} + L_{\hat{i}} - R_i - L_i = h\left(\frac{1}{\tan(u_{\hat{i}})} - \frac{1}{\tan(w_{\hat{i}})}\right)$$

Accordingly,  $\Delta t_{h1}$  is obtained as

$$\Delta t_{h1} = \beta_{\hat{i}} (R_{\hat{i}} + L_{\hat{i}} - R_i - L_i)$$

Now, recalling  $\zeta_2(R_i + L_i) = \Pr[0 < d \le R_i + L_i]$  and  $\zeta_3(R_i + L_i - R_i - L_i) = \Pr[R_i + L_i < d \le R_i + L_i]$ , the expressions of  $\zeta_i$ ,  $i = \{2, 3\}$  can be obtained as demonstrated in the theorem.

(c) This part of the theorem can be proved after direct substitutions into the definitions of the expected value and variance of  $f_D(d)$ .

## Appendix B. Proving the convexity of $D_{var}(R_1, R_2)$ with respect to the larger red time

**Proposition 1.** Let us assume  $R_{\hat{i}} \ge R_i$ ,  $\hat{i} \in \{1, 2\}$ . Then the variance function  $D_{var}(R_1, R_2)$  is a convex function of  $R_{\hat{i}}$ .

**Proof.** Without loss of generality and for simplicity in notations we assume  $R_2 \ge R_1$  and  $L_1 = L_2 = 0$ . The variance function  $D_{\text{var}}(R_1, R_2)$ , (15), can be formulated as follows:

$$D_{\text{var}}(R_1, R_2) = \left(1 - \chi_{11} \frac{R_1}{R_1 + R_2} - \chi_{12} \frac{R_2}{R_1 + R_2}\right) \left(\chi_{41} \frac{R_1^2}{R_1 + R_2} + \chi_{42} \frac{R_2^2}{R_1 + R_2}\right)^2 + \frac{1}{3} \frac{\chi_{21} - \chi_{31}}{R_1 + R_2} \left(R_1 - \chi_{41} \frac{R_1^2}{R_1 + R_2} - \chi_{42} \frac{R_2^2}{R_1 + R_2}\right)^3 + \frac{1}{3} \frac{\chi_{21}}{R_1 + R_2} \left(\chi_{41} \frac{R_1^2}{R_1 + R_2} + \chi_{42} \frac{R_2^2}{R_1 + R_2}\right)^3 + \frac{1}{3} \frac{\chi_{31}}{R_1 + R_2} \left(R_2 - \chi_{41} \frac{R_1^2}{R_1 + R_2} - \chi_{42} \frac{R_2^2}{R_1 + R_2}\right)^3,$$
(B.1)

where  $\chi_{11} \triangleq \frac{\beta_1 \varrho_1 q_1^4}{q_1^3 + q_2^2}$ ,  $\chi_{12} \triangleq \frac{\beta_2 \varrho_2 q_2^3}{q_1^3 + q_2^2}$ ,  $\chi_{41} \triangleq \frac{0.5 q_1^3 \beta_1 \varrho_1}{q_1^3 + q_2^2}$ ,  $\chi_{42} \triangleq \frac{0.5 q_2^3 \beta_2 \varrho_2}{q_1^3 + q_2^3}$ ,  $\chi_{21} \triangleq \sum_{i=1,2} \frac{\beta_i q_i^3 \varrho_i}{q_1^3 + q_2^2}$ , and  $\chi_{31} \triangleq \frac{\beta_2 q_2^3 \varrho_2}{q_1^3 + q_2^3}$ . Now, let us define  $x = R_2/R_1 \ge 1$ . Accordingly, (B.1) can be written as

$$D_{\text{var}}(x) = C_6 \frac{x^6}{(1+x)^4} + C_5 \frac{x^5}{(1+x)^4} + C_4 \frac{x^4}{(1+x)^4} + C_3 \frac{x^3}{(1+x)^4} + C_2 \frac{x^2}{(1+x)^4} + C_1 \frac{x}{(1+x)^4} + C_0 \frac{1}{(1+x)^4},$$
(B.2)

where  $C_i > 0$ ,  $i = \{0, \dots, 6\}$  are real constant scalars defined as

$$\begin{split} &C_{6} = \chi_{42}^{3} \chi_{21} + 3\chi_{42}^{2} \frac{1 - \chi_{12}}{R_{1}} - \chi_{42}^{3} (\chi_{21} - \chi_{31}) - \chi_{31} (\chi_{42} - 1)^{3}, \\ &C_{5} = 3\chi_{42}^{2} \frac{2 - \chi_{11} - \chi_{12}}{R_{1}} + \chi_{31} ((\chi_{42} - 1)^{2} + (2\chi_{42} - 2)(\chi_{42} - 1)) + 3\chi_{42}^{2} (\chi_{21} - \chi_{31}), \\ &C_{4} = 3\chi_{42}^{2} \frac{1 - \chi_{11}}{R_{1}} - (\chi_{21} - \chi_{31}) (2\chi_{42} + \chi_{42} (2\chi_{42} (\chi_{41} - 1) + 1) + \chi_{42}^{2} (\chi_{41} - 1))) \\ &- \chi_{31} (2\chi_{42} + (\chi_{42} - 1)(2\chi_{41} (\chi_{42} - 1) + 1) + \chi_{41} (\chi_{42} - 1)^{2} - 2) \\ &+ 6\chi_{41} \chi_{42} \frac{1 - \chi_{12}}{R_{1}} + 3\chi_{41} \chi_{42}^{2} \chi_{21}, \\ &C_{3} = \chi_{31} (4\chi_{41} (\chi_{42} - 1) + \chi_{41} (2\chi_{42} - 2) + 1) + 6\chi_{41} \chi_{42} \frac{2 - \chi_{11} - \chi_{12}}{R_{1}} \\ &+ (\chi_{21} - \chi_{31}) (4\chi_{42} (\chi_{41} - 1) + \chi_{42} (2\chi_{41} - 2) + 1), \\ &C_{2} = 3\chi_{41}^{2} \frac{1 - \chi_{12}}{R_{1}} - \chi_{31} (2\chi_{41} + \chi_{41} (2\chi_{41} (\chi_{42} - 1) + 1) + \chi_{42} (\chi_{42} - 1))) \\ &- (\chi_{21} - \chi_{31}) (2\chi_{41} + (\chi_{41} - 1) (2\chi_{42} (\chi_{41} - 1) + 1) + \chi_{42} (\chi_{41} - 1))^{2} - 2) \\ &+ 6\chi_{41} \chi_{42} \frac{1 - \chi_{12}}{R_{1}} - \chi_{31} (2\chi_{41} + \chi_{41} (2\chi_{41} (\chi_{42} - 1) + 1) + \chi_{42} (\chi_{41} - 1))^{2} - 2) \\ &+ 6\chi_{41} \chi_{42} \frac{1 - \chi_{11}}{R_{1}} + 3\chi_{41}^{2} \chi_{31} + (\chi_{21} - \chi_{31}) ((\chi_{41} - 1)^{2} + (2\chi_{41} - 2) (\chi_{41} - 1))), \\ &C_{0} = \chi_{41}^{3} \chi_{21} - (\chi_{21} - \chi_{31}) (\chi_{41} - 1)^{3} - \chi_{41}^{3} \chi_{31} + 3\chi_{41}^{2} \frac{1 - \chi_{11}}{R_{1}} . \end{split}$$

After differentiating (B.2) twice with respect to x, it follows that

$$\frac{\partial^2}{\partial x^2} D_{\text{var}}(x) = \frac{2\sum_{i=0}^6 \hat{C}_i x^i}{(1+x)^6},$$
(B.3)

where  $\hat{C}_i > 0$ ,  $i = \{0, \dots, 6\}$  are constant scalars defined as:  $\hat{C}_0 = 10C_0 - 4C_1 + C_2$ ;  $\hat{C}_1 = 6C_1 - 6C_2 + 3C_3$ ;  $\hat{C}_2 = 3C_2 - 6C_3 + 6C_4$ ;  $\hat{C}_3 = C_3 - 4C_4 + 10C_5$ ;  $\hat{C}_4 = 15C_6$ ;  $\hat{C}_5 = 6C_6$ ; and  $\hat{C}_6 = C_6$ . Therefore, according to Boyd and Vandenberghe (2004) due to the positive-definiteness of the second derivative of  $D_{var}(x)$ , the variance function is convex with respect to  $R_2$ .

## Appendix C. Signal optimization for minimizing the delay variability at the intersection with loss times

In the sequel, an algorithm is proposed to obtain the analytical global optimal signal timings for minimizing the delay variability at the intersection with loss times.

## Algorithm 6

- **I.** If Inequality (6) is realized, there exists a global optimal solution and one may proceed to the next step. Otherwise, the intersection is oversaturated and the algorithm terminates with no solution.
- **II.** If  $R^* \ge D^*$ , then the optimal solution is  $R^*$ .
- **III.** Determine if (i)  $\eta_i \ge 1$  (i = 1 or 2), then proceed to Step IV, or (ii)  $\eta_1 < 1$  and  $\eta_2 < 1$ , then proceed to Step V.
- **IV.** It can be deduced that  $R_i \ge R_i$ . Thereafter,
  - 1. If  $\eta_i R_i^{\min} + (\eta_i + 1)L_i \ge R_i^{\min}$ , then the optimal solution is  $(R_i^{\min}, \eta_i R_i^{\min} + (\eta_i + 1)L_i)$ .
  - 2. Otherwise, the solution lies on the line  $R_{\hat{i}} = R_{\hat{i}}^{\min}$  and the optimal  $R_i$  is obtained from solving the following program:

$$\begin{array}{l} \underset{R_{i},R_{i}=R_{i}^{\min}}{\min} \quad D_{var}(R_{i}) \\ \text{Subject to:} \\ -R_{i} + R_{i}^{\min} \leq 0, \\ -R_{i} + \eta_{i}R_{i}^{\min} + (1 + \eta_{i})L_{i} \leq 0, \\ R_{i} - \frac{1}{\eta_{i}}R_{i}^{\min} + \frac{1 + \eta_{i}}{\eta_{i}}L_{i} \leq 0, \\ R_{i} + L_{i} - k_{i}^{j} \left(\frac{1}{q_{i}^{3}} - \frac{1}{q_{i}^{c}}\right)\Delta_{i} \leq 0, \\ R_{i}^{\min} + L_{i}^{2} - k_{i}^{j} \left(\frac{1}{q_{i}^{3}} - \frac{1}{q_{i}^{c}}\right)\Delta_{i} \leq 0. \\ \end{bmatrix}$$
(02-1)

**V.** Let us assume that  $R_i^{\min} \leq R_i^{\min}$ .

 $R_{\hat{i}}^{m}$ 

- 1. If  $R_{\hat{i}}^{\star} \geq R_{\hat{i}}^{\min}$ , then the optimal solution is  $(R_{\hat{i}}^{\min}, \eta_i R_{\hat{i}}^{\min} + (\eta_i + 1)L_i)$ .
- 2. If  $D^* > R^*$ , one of the following conditions arise:
  - **a.** If  $R_i^{\min} \ge \frac{1}{\eta_i} R_i^{\min} \frac{1+\eta_i}{\eta_i} L_i$ , the optimal solution is  $(R_i^{\min}, \eta_i R_i^{\min} + (\eta_i + 1)L_i)$ .
  - **b.** Otherwise, the solution lies on the line  $R_{\hat{i}} = R_{\hat{i}}^{\min}$  and the optimal  $R_i$  is attained from solving the following optimization problem:

$$\begin{array}{l} \underset{R_{i},R_{i}=R_{i}^{\min}}{\min} \quad D_{var}(R_{i}) \\ \text{Subject to:} \\ & -R_{i}+R_{i}^{\min} \leq 0, \\ & R_{i}-R_{i}^{\min} \leq 0, \\ -R_{i}+\eta_{i}R_{i}^{\min}+(1+\eta_{i})L_{i} \leq 0, \\ & R_{i}-\frac{1}{\eta_{i}}R_{i}^{\min}+\frac{1+\eta_{i}}{\eta_{i}}L_{i} \leq 0, \\ & R_{i}-L_{i}-k_{i}^{j}\left(\frac{1}{q_{i}^{2}}-\frac{1}{q_{i}^{2}}\right)\Delta_{i} \leq 0, \\ & \\ & \lim_{i}+L_{i}^{2}-k_{i}^{j}\left(\frac{1}{q_{i}^{2}}-\frac{1}{q_{i}^{2}}\right)\Delta_{i} \leq 0. \\ & \\ \end{array}$$
(02-2)

- 3. Otherwise, we have one of the following conditions:
  - **a.** If  $R_{\hat{i}}^{\min} \leq \eta_{\hat{i}} R_{\hat{i}}^{\min} + (1 + \eta_{\hat{i}}) L_{\hat{i}}$ , then the optimal solution is  $(\eta_{\hat{i}} R_{\hat{i}}^{\min} + (\eta_{\hat{i}} + 1) L_{\hat{i}}, R_{\hat{i}}^{\min})$ .
  - **b.** Elseways, the optimal solution can be sought solving the optimization program (O2-2) noting that the constraint  $R_i \frac{1}{\eta_i} R_i^{\min} + \frac{1+\eta_i}{\eta_i} L_i \le 0$  is already satisfied.

#### Appendix D. Proof of Corollary 1

The proof follows the same line of the proof of Theorem 1. Part (a) indicates that the delay PDF of Phase i is composed of a Dirac function associated with vehicles that do not stop at the intersection, and a uniform distribution function associated with delayed vehicles that experience delays between 0 and  $R_i + L_i$ . Note that the number of vehicles from Phase *i* that do not stop at the intersection during a cycle is simply  $\tilde{n}_i = Cq_i^a - k_i^j x_i^j$  and the probability that a vehicle stops at the intersection is  $k_i^j x_i^j / (Cq_i^a)$ . The delay PDF at the whole intersection follows from the following facts:

- i. The probability that a vehicle does not experience delay at the intersection is  $\sum_{i=1}^{N} \tilde{n}_i / \sum_{i=1}^{N} Cq_i^a$  which results in the definition of  $\hat{\zeta}_0$ .
- ii.  $\Pr[R_{i-1} + L_{i-1} \le d \le R_i + L_i] = \frac{\sum_{j=1}^{N} q_j^2 \varrho_j \beta_j}{\sum_{j=1}^{N} Cq_j^2} (R_i + L_i R_{i-1} L_{i-1}) = \hat{\zeta}_i (R_i + L_i R_{i-1} L_{i-1}).$  The expression can be deduced from Fig. 6 (a), in a way that the number of vehicles in Phase  $m \ge i$  that experience delay between  $R_{i-1} + L_{i-1}$  and

 $R_i + L_i$  is equal to  $k_m^j h_{i-1,i}^m$ , where  $h_{i-1,i}^m$  represents the queue length associated with those vehicles. Simple trigonometric manipulations result in an expression for  $h_{i-1,i}^m$  as  $h_{i-1,i}^m = (R_i + L_i - R_{i-1} - L_{i-1})/(R_m + L_m) x_m^j$ , which itself results in the expression above.

Finally, expected value of delay and variance of delay can be calculated directly from the expression of delay PDF (25).

#### References

- Aalipour, A., Kebriaei, H., Ramezani, M., 2018. Analytical optimal solution of perimeter traffic flow control based on mfd dynamics: a pontryagin's maximum principle approach. IEEE Trans. Intell. Transp. Syst. 1-11.
- Aboudolas, K., Papageorgiou, M., Kouvelas, A., Kosmatopoulos, E., 2010. A rolling-horizon quadratic-programming approach to the signal control problem in large-scale congested urban road networks. Transp. Res. Part C 18 (5), 680-694.
- Aimsun, T., 2014. Aimsun User Manual.

AKÇELIK, R., 1981. Traffic Signals: Capacity and Timing Analysis. Report ARR NO. 123. Australian Road Research Board.

- Ban, X., Herring, R., Hao, P., Bayen, A., 2009. Delay pattern estimation for signalized intersections using sampled travel times. Transp. Res. Rec. 2130, 109-119. Boyd, S., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press.
- Chen, P., Sun, J., Qi, H., 2016. Estimation of delay variability at signalized intersections for urban arterial performance evaluation. J. Intell. Transp. Syst. 00-00.
- Cheng, Y., Qin, X., Jin, J., Ran, B., Anderson, J., 2011. Cycle-by-cycle queue length estimation for signalized intersections using sampled trajectory data. Transp. Res. Rec. (2257) 87-94.
- Diakaki, C., Papageorgiou, M., Aboudolas, K., 2002. A multivariable regulator approach to traffic-responsive network-wide signal control. Control Eng. Pract. 10 (2), 183–195.

Dinopoulou, V., Diakaki, C., Papageorgiou, M., 2006. Applications of the urban traffic control strategy tuc. Eur. J. Oper. Res. 175 (3), 1652-1665.

Dion, F., Rakha, H., Kang, Y.-S., 2004. Comparison of delay estimates at under-saturated and over-saturated pre-timed signalized intersections. Transp. Res. Part B 38 (2), 99-122.

- Dowling, R., Skabardonis, A., Halkias, J., McHale, G., Zammit, G., 2004. Guidelines for calibration of microsimulation models: framework and applications. Transp. Res. Rec. 1876, 1–9, doi:10.3141/1876-01.
- Fu, L., Hellinga, B., 2000. Delay variability at signalized intersections. Transp. Res. Rec. 1710, 215–221. doi:10.3141/1710-25.

Gazis, D.C., 1964. Optimum control of a system of oversaturated intersections. Oper. Res. 12 (6), 815-831.

- Geroliminis, N., Skabardonis, A., 2011. Identification and analysis of queue spillovers in city street networks. IEEE Trans. Intell. Transp. Syst. 12 (4), 1107-1115. Haddad, J., Mahalel, D., 2014. Offset effects on the capacity of paired signalised intersections during oversaturated conditions. Transportmetrica A 10 (8), 740-758
- Haddad, J., Mahalel, D., Ioslovich, I., Gutman, P.-O., 2014. Constrained optimal steady-state control for isolated traffic intersections. Control Theory Technol. 12 (1), 84–94.
- Haddad, J., Schutter, B.D., Mahalel, D., Ioslovich, I., Gutman, P.O., 2010. Optimal steady-state control for isolated traffic intersections. IEEE Trans. Autom. Control 55 (11), 2612-2617.
- loslovich, I., Haddad, J., Gutman, P.-O., Mahalel, D., 2011. Optimal traffic control synthesis for an isolated intersection. Control Eng. Pract. 19 (8), 900-911.
- Keyvan-Ekbatani, M., Yildirimoglu, M., Geroliminis, N., Papageorgiou, M., 2015. Multiple concentric gating traffic control in large-scale urban networks. IEEE Trans. Intell. Transp. Syst. 16 (4), 2141-2154.
- Kosmatopoulos, E.B., Papageorgiou, M., Vakouli, A., Kouvelas, A., 2007. Adaptive fine-tuning of nonlinear control systems with application to the urban traffic control strategy tuc. IEEE Trans. Control Syst. Technol. 15 (6), 991-1002.
- Kouvelas, A., Saeedmanesh, M., Geroliminis, N., 2017. Enhancing model-based feedback perimeter control with data-driven online adaptive optimization. Transp. Res. Part B 96, 26-45.
- Lee, S., Wong, S.C., Varaiya, P., 2017. Group-based hierarchical adaptive traffic-signal control part i: formulation. Transp. Res. Part B 105, 1-18.

Li, J., Zhang, H., 2014. Coupled linear programming approach for decentralized control of urban traffic. Transp. Res. Rec. 2439, 83–93.

- Lighthill, M.J., Whitham, G.B., 1955. On kinematic waves. ii. a theory of traffic flow on long crowded roads. Proc. R. Soc. Lond. A 229 (1178), 317-345.
- Liu, H.X., Wu, X., Ma, W., Hu, H., 2009. Real-time queue length estimation for congested signalized intersections. Transp. Res. Part C 17 (4), 412-427.

Liu, Y., Chang, G.-L., 2011. An arterial signal optimization model for intersections experiencing queue spillback and lane blockage. Transp. Res. Part C 19 (1), 130-144.

Mahmassani, H.S., Saberi, M., Zockaie, A., 2013. Urban network gridlock: theory, characteristics, and dynamics. Transp. Res. Part C 36, 480-497.

Michalopoulos, P.G., Pisharody, V.B., 1981. Derivation of delays based on improved macroscopic traffic models. Transp. Res. Part B 15 (5), 299-317.

Michalopoulos, P.G., Stephanopoulos, G., Stephanopoulos, G., 1981. An application of shock wave theory to traffic signal control. Transportation Research Part B: Methodological 15 (1), 35-51.

- Ramezani, M., Geroliminis, N., 2015. Queue profile estimation in congested urban networks with probe data. Comput. Aided Civ. Infrastruct. Eng. 30 (6), 414-432.
- Ramezani, M., Haddad, J., Geroliminis, N., 2015. Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control. Transp. Res. Part B 74, 1-19.

Ramezani, M., de Lamberterie, N., Skabardonis, A., Geroliminis, N., 2017. A link partitioning approach for real-time control of queue spillbacks on congested arterials. Transportmetrica B 5 (2), 181–194.

Richards, P.I., 1956. Shock waves on the highway. Oper. Res. 4 (1), 42–51.

Skabardonis, A., Geroliminis, N., 2008. Real-time monitoring and control on signalized arterials. J. Intell. Transp. Syst. 12 (2), 64-74.

Varaiya, P., 2013. Max pressure control of a network of signalized intersections. Transp. Res. Part C 36, 177-195.

Varaiya, P., 2013. The Max-Pressure Controller for Arbitrary Networks of Signalized Intersections. Springer New York, New York, NY, pp. 27-66.

Wada, K., Usui, K., Takigawa, T., Kuwahara, M., 2017. An optimization modeling of coordinated traffic signal control based on the variational theory and its stochastic extension. Transp. Res. Part B.

Wang, X.B., Cao, X., Wang, C., 2017. Dynamic optimal real-time algorithm for signals (doras): case of isolated roadway intersections. Transportation Research Part B: Methodological 106, 433–446.

Wardrop, J.G., 1952. Road paper, some theoretical aspects of road traffic research. Proc. Inst. Civ. Eng. 1 (3), 325-362.

Webster, F.V., 1958. Traffic Signal Settings. Report.

Yu, C., Feng, Y., Liu, H.X., Ma, W., Yang, X., 2018. Integrated optimization of traffic signals and vehicle trajectories at isolated urban intersections. Transp. Res. Part B 112, 89-112.

Yu, C., Ma, W., Han, K., Yang, X., 2017. Optimization of vehicle and pedestrian signals at isolated intersections. Transp. Res. Part B 98, 135–153.

Zheng, F., Zuylen, H.v., Liu, X., 2017. A methodological framework of travel time distribution estimation for urban signalized arterial roads. Transp. Sci.