Perimeter and boundary flow control in multi-reservoir heterogeneous networks

Konstantinos Aboudolas *, Nikolas Geroliminis

Urban Transport Systems Laboratory, School of Architecture, Civil & Environmental Engineering, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

A B S T R A C T

In this paper, we macroscopically describe the traffic dynamics in heterogeneous transportation urban networks by utilizing the Macroscopic Fundamental Diagram (MFD), a widely observed relation between network-wide space-mean flow and density of vehicles. A generic mathematical model for multi-reservoir networks with well-defined MFDs for each reservoir is presented first. Then, two modeling variations lead to two alternative optimal control methodologies for the design of perimeter and boundary flow control strategies that aim at distributing the accumulation in each reservoir as homogeneously as possible, and maintaining the rate of vehicles that are allowed to enter each reservoir around a desired point, while the system's throughput is maximized. Based on the two control methodologies, perimeter and boundary control actions may be computed in real-time through a linear multivariable feedback regulator or a linear multivariable integral feedback regulator. Perimeter control occurs at the periphery of the network while boundary control occurs at the inter-transfers between neighborhood reservoirs. To this end, the heterogeneous network of San Francisco is partitioned into three homogeneous reservoirs and the proposed feedback regulators are compared with a pre-timed signal plan and a single-reservoir perimeter control strategy. Finally, the impact of the perimeter and boundary control actions is demonstrated via simulation by the use of the corresponding MFDs and other performance measures. A key advantage of the proposed approach is that it does not require high computational effort and future demand data if the current state of each reservoir can be observed with loop detector data.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Realistic modeling and efficient control of heterogeneous transportation networks remain a big challenge, due to the high unpredictability of choices of travelers (in terms of route, time of departure and mode of travel), the uncertainty in their reactions to the control, the spatiotemporal propagation of congestion, and the lack of coordinated actions coupled with the limited infrastructure available. While there is a vast literature of congestion dynamics, control and spreading in one-dimensional traffic systems with a single mode of traffic, most of the analysis at the network level is based on simplistic models or simulation, which require a large number of input parameters (sometimes unobservable with existing data) and cannot be solved in real-time. Still congestion governance in large-scale systems is currently fragmented and uncoordinated with respect to optimizing the goals of travel efficiency and equity for multiple entities. Understanding these

* Corresponding author. Tel.: +41 (0)21 69 32481.
E-mail addresses: konstantinos.ampountolas@epfl.ch (K. Aboudolas), nikolas.geroliminis@epfl.ch (N. Geroliminis).

interactions for complex and congested cities is a big challenge, which will allow revisiting, redesigning and integrating smarter traffic management approaches to generate more sustainable cities.

With respect to traffic signal control, many methodologies have been developed, but still a major challenge is the deployment of advanced and efficient traffic control strategies in heterogeneous large-scale networks, with particular focus on addressing traffic congestion and propagation phenomena. Widely used strategies like SCOOT (Hunt et al., 1982) and SCATS (Lowrie, 1982), although applicable to large-scale networks, are less efficient under oversaturated traffic conditions with long queues and spillbacks. However, recently ad hoc gating schemes (engineering solutions) have been incorporated in these systems to resolve local spill-over situations (Bretherton et al., 2003; Luk and Green, 2010). Other advanced traffic-responsive strategies (Gartner et al., 2001, 2005) use complex optimization algorithms, which do not permit a real-time network-wide application. A practicable work to address oversaturated traffic conditions was the recently developed feedback control strategy TUC (Diakaki et al., 2002; Diakaki et al., 2003; Aboudolas et al., 2009; Kouvelas et al., 2011). TUC attempts to minimize the risk of oversaturation and spillback of link queues by minimizing and balancing the links' relative occupancies. Furthermore, TUC also includes a local gating feature to protect downstream links from overload in the sense of limiting the entrance in a link when close to overload. However, these policies might be suboptimal or delayed reactive for heterogeneous networks with multiple centers of congestion and heavily directional demand flows.

An alternative avenue to real-time network-wide traffic signal control for urban networks is a hierarchical two-level approach, where at the first level perimeter and boundary flow control between different regions of the network advances the aggregated performance, while at the second level a more detailed control can be applied to smooth traffic movements within these regions (e.g. TUC). The physical tool to advance in a systematic way this research is the Macroscopic Fundamental Diagram (MFD) of urban traffic, which provides for network regions under specific regularity conditions (mainly homogeneity in the spatial distribution of congestion and the network topology), a unimodal, low-scatter relationship between network vehicle accumulations n (veh) and network outflow (veh/h), as shown in Fig. 1(a). The idea of an MFD with an optimum (critical) accumulation $\tilde{n}$ at which capacity is reached (maximum circulating flow or throughput) belongs to Godfrey (1969), but the empirical verification of its existence with dynamic features is recent (Geroliminis and Daganzo, 2008). Given also the linear relationship between network outflow and circulating (or space-mean) flow (due to time-invariant regional trip length), MFD can also be expressed as space-mean flow vs. accumulation. Both expressions are utilized in this paper due to their similarity. Circulating flow can be directly measured by loop detectors while outflow requires a wide deployment of GPS. This property is important for modeling purposes as details in individual links are not needed to describe the congestion level of cities and its dynamics. It can also be utilized to introduce simple perimeter flow control policies to improve mobility in homogeneous networks (Daganzo, 2007; Keyvan-Ekbatani et al., 2012; Geroliminis et al., 2013). The general idea of a perimeter flow control policy is to "meter" the input flow to the system and to hold vehicles outside the controlled area if necessary. A key advantage of this approach is that it does not require high computational effort if proxies for $\tilde{n}$ are available (e.g. critical accumulation, critical average occupancy or critical density) and the current state of the network n can be observed with loop detector data in real-time (see Fig. 1(a)). A drawback of this approach is that it creates queues blocking the urban roads outside the controlled area. Alternatively, route choice or dynamic pricing models can be directly incorporated in the perimeter flow control problem to avoid long queues and delays at the perimeter of the controlled area (see e.g. Haddad et al., 2013; Knoop et al., 2012; Geroliminis and Levinson, 2009).
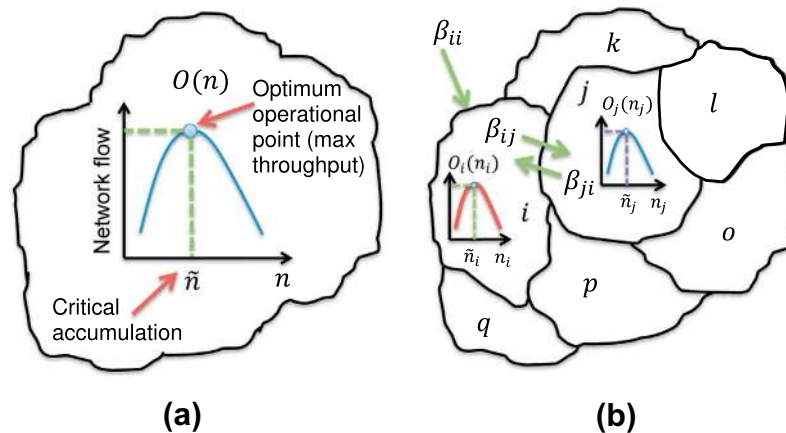

Fig. 1. A network modeled as (a) single-reservoir system and (b) multi-reservoir system. In (a), the macroscopic fundamental diagram $O(n)$ defines a unimodal, low-scatter relation between network vehicle accumulations n (veh) and network outflow or output (veh/h) for all road sections. The maximum outflow in the network may be observed over a range of accumulation-values that is close to a critical accumulation $\tilde{n}$. In (b), each reservoir i exhibits a macroscopic fundamental diagram $O_i(n_i)$ where n_i is the regional accumulation; each destination reservoir i is reachable (from the perimeter or boundary) from a number of origin reservoirs, which defines the set $\mathcal{S}_i$, e.g. $\mathcal{S}_i = \{i,j,k,p,q\}$ ($\mathcal{S}_i$ includes i since the destination reservoir i is reachable from the perimeter) and $\mathcal{S}_j = \{i,k,l,o,p,q\}$.

Despite these findings for the existence of an MFD with low scatter, these curves should not be a universal law. Recent works (Mazloumian et al., 2010; Geroliminis and Sun, 2011; Daganzo et al., 2011; Knoop et al., 2012; Saberi and Mahmassani, 2012) have identified the spatial distribution of vehicle density in the network as one of the key components that affect the scatter of an MFD and its shape. They observed that the average network flow is consistently higher when link density variance is low for the same network density, but higher densities can create points below an MFD when they are heterogeneously distributed. Other investigations of empirical and simulated studies for network level traffic patterns can be found in Buisson and Ladier (2009), Aboudolas et al. (2010), Ji et al. (2010), Gayah and Daganzo (2011a), Wu et al. (2011), Mahmassani et al. (2013a), Mahmassani et al. (2013), Zhang et al. (2013) and elsewhere.

These results are of great importance because the concept of an MFD can be applied for heterogeneously loaded networks with multiple centers of congestion, if these networks can be partitioned into a small number of homogeneous reservoirs (regions). The objectives of partitioning are to obtain (i) small variance of link densities within a reservoir, which increases the network flow for the same average density and (ii) spatial compactness of each reservoir which makes feasible the application of perimeter and boundary flow control (Ji and Geroliminis, 2012). The objective is to partition a heterogeneous network into homogeneous reservoirs with small variance of link densities and well-defined MFDs, as shown in Fig. 1(b). On the other hand, single-reservoir perimeter flow control (Daganzo, 2007; Keyvan-Ekbatani et al., 2012) may enhance an uneven distribution of vehicles in different parts of the network (for example due to asymmetric route choices and origin–destination matrices), and, as a consequence, may invalidate the homogeneity assumption of traffic loads and degrade the total network throughput. Thus, in this work we put some effort to deal with the important issues of efficiency, heterogeneity and equity in perimeter flow control. In particular, for a given partition of a heterogeneous network into some homogeneous regions and corresponding MFDs (see Fig. 1(b)) with a critical (sweet spot) accumulation that maximizes the regional circulating flow (outflow or trip completion rate), we develop perimeter and boundary flow control strategies to improve mobility in heterogeneous networks. In this approach, perimeter flow control occurs at the periphery of the network while boundary flow control occurs at the inter-transfers between neighborhood reservoirs.

More specifically, a generic mathematical model of an N-reservoir network with well-defined MFDs for each reservoir is presented first. Two modeling variations lead to two alternative optimal control methodologies for the design of perimeter and boundary flow control strategies that aim at distributing the accumulation in each reservoir as homogeneously as possible, and maintaining the rate of vehicles that are allowed to enter each reservoir around a desired point, while the system's throughput is maximized. Based on the two control methodologies, perimeter and boundary control actions may be computed in real-time through a linear multivariable feedback regulator or a linear multivariable integral feedback regulator. To this end, the heterogeneous network of the Downtown of San Francisco is partitioned into three homogeneous regions that exhibit well-defined MFDs. These MFDs are then used to design and compare the two feedback regulators with a pre-timed signal control plan and a single-reservoir perimeter control strategy. Finally, the impact of the perimeter and boundary control actions to the three reservoirs and the whole network is demonstrated via simulation by the use of the corresponding MFDs and other performance measures, under a number of different demand scenarios.

2. Dynamics for heterogeneous networks partitioned in N reservoirs

Consider a heterogeneous network partitioned in N reservoirs (Fig. 1(b)). Denote by $i = 1, \ldots, N$ a reservoir in the system, and let $n_i(t)$ be the accumulation of vehicles in reservoir i at time t; $n_{i,\max}$ be the maximum accumulation of vehicles in reservoir i. We assume that for each reservoir $i = 1, \ldots, N$ there exists an MFD, $O_i(n_i(t))$, between accumulation n_i and output O_i (number of trips exiting reservoir i per unit time either because they finished their trip or because they move to another reservoir), which describes the behavior of the system when it evolves slowly with time t.

Let $q_{i,\text{in}}(t)$ and $q_{i,\text{out}}(t)$ be the inflow and outflow in reservoir i at time t, respectively; $\mathcal{S}_i$ be the set of origin reservoirs whose outflow will go to destination reservoir i (including reservoir i in case that reservoir i is reachable from the perimeter, see Fig. 1(b)). Also, let $d_i(t)$ be the uncontrolled traffic demand (disturbances) in reservoir i at time t. Note that $d_i(t)$ includes both internal (off-street parking for taxis and pockets for private vehicles) and external non-controlled inflows. The conservation equation for each reservoir $i = 1, \ldots, N$ reads:

$$\frac{dn_i(t)}{dt} = q_{i,\text{in}}(t) - q_{i,\text{out}}(t) + d_i(t) \tag{1}$$

Since the system of each reservoir evolves slowly with time t, we may assume that the outflow $q_{i,\text{out}}(t)$ is given by the output $O_i(n_i(t))$ (the MFD), which is a function of the accumulation $n_i(t)$, where output $O_i(n_i(t))$ is the sum of the exit flows from reservoir i to reservoir j, plus the internal output (internal trip completion rates at i). If i and j are two reservoirs sharing a common boundary, we denote by β_{ji} $(j \neq i)$ the fraction of the flow rate in reservoir j that are allowed to enter reservoir i and by β_{ii} the fraction of the flow rate in the perimeter of the network allowed to enter reservoir i (see Fig. 1(b)). The inflow to reservoir i is given by

$$q_{i,\text{in}}(t) = \sum_{j \in \mathcal{S}_i} \beta_{ji}(t - \tau_{ji}) O_j(n_j(t)) \tag{2}$$

where $\beta_{ji}(t - \tau_{ji})$ are the input variables from reservoir j to reservoir i at time t, to be calculated by the perimeter and boundary controller, and τ_{ji} is the travel time needed for vehicles to approach reservoir i from origin reservoir j. Given that we

assume no knowledge of generated demand from areas outside of the entire network (external perimeter), the input flow from the perimeter to reservoir i is considered proportional to the outflow of region i, $O_i(n_i)$. While this simplification might not be always the case, the value of β_{ii} will be consistent with the physical properties of the system when the controller is active, i.e. for large input flow and/or congested conditions (see also (4) below). Without loss of generality, we assume that $\tau_{ji} = 0$, i.e., vehicles can immediately get access to the receiving reservoirs when exiting from the sending reservoirs of the network. This assumption can be readily removed if needed by introducing additional auxiliary variables (e.g. see Chapter 2 in Åström and Wittenmark (1996)).

Additionally, $\beta_{ji}(t)$ is constrained as follows

$$\beta_{ji,\min} \leqslant \beta_{ji}(t) \leqslant \beta_{ji,\max} \tag{3}$$

where $\beta_{ji,\min}$, $\beta_{ji,\max}$ are the minimum and maximum permissible entrance rate of vehicles, respectively, and $\beta_{ji,\min} > 0$ to avoid long queues and delays at the perimeter of the network and the boundary of neighborhood reservoirs. Moreover, the following constraints are introduced to prevent overflow phenomena within the reservoirs

$$\sum_{i=1}^{N} (\beta_{ji}(t) + \varepsilon_i) \leqslant 1, \quad \forall j = 1, \ldots, N \tag{4}$$

where $\varepsilon_i > 0$ is a portion of uncontrolled flow that enters reservoir i. Finally the accumulation $n_i(t)$ cannot be higher than the maximum accumulation $n_{i,\max}$ for each reservoir i

$$0 \leqslant n_i(t) \leqslant n_{i,\max}, \quad \forall i = 1, \ldots, N \tag{5}$$

Introducing (2) in (1) we obtain the following non-linear state equation

$$\frac{dn_i(t)}{dt} = \sum_{j \in \mathcal{S}_i} \beta_{ji}(t) O_j(n_j(t)) - O_i(n_i(t)) + d_i(t) \tag{6}$$

Given the existence of MFDs $O_i(n_i(t))$ with an optimum (critical) accumulation $\tilde{n}_i$ at which capacity is reached for each reservoir $i = 1, \ldots, N$ (see Fig. 1(b)), the non-linear model (6) may be linearized around some set point $\hat{n}_i$, $\hat{\beta}_{ji}$, and $\hat{d}_i$ that satisfies the steady state version of (6), given by

$$0 = \sum_{j \in \mathcal{S}_i} \hat{\beta}_{ji}(t) O_j(\hat{n}_j(t)) - O_i(\hat{n}_i(t)) + \hat{d}_i(t) \tag{7}$$

Denoting $\Delta x = x - \hat{x}$ analogously for all variables and assuming first-order Taylor approximation, the linearization yields

$$\Delta \dot{n}_i(t) = \sum_{j \in \mathcal{S}_i} \Delta \beta_{ji}(t) O_j(\hat{n}_j(t)) + \sum_{j \in \mathcal{S}_i} \hat{\beta}_{ji}(t) \Delta n_j(t) O'_j(\hat{n}_j(t)) - \Delta n_i(t) O'_i(\hat{n}_i(t)) + \Delta d_i(t) \tag{8}$$

The linear system (8) approximates the original non-linear system (6) when we are near the equilibrium point about which the system was linearized. In our case, this equilibrium point should be close to the critical accumulation $\tilde{n}_i$ for each reservoir $i = 1, \ldots, N$, where the individual reservoirs' output is maximized.

Applying (8) to a network partitioned in N reservoirs the following state equation (in vector form) describes the evolution of the system in time

$$\Delta \dot{\mathbf{n}}(t) = \mathbf{F} \Delta \mathbf{n}(t) + \mathbf{G} \Delta \boldsymbol{\beta}(t) + \mathbf{H} \Delta \mathbf{d}(t) \tag{9}$$

where $\Delta \mathbf{n} \in \mathbb{R}^N$ is the state deviations vector of $\Delta n_i = n_i - \hat{n}_i$ for each reservoir $i = 1, \ldots, N$; $\Delta \boldsymbol{\beta} \in \mathbb{R}^M$ is the control deviations vector of $\Delta \beta_{ji} = \beta_{ji} - \hat{\beta}_{ji}$, $\forall i = 1, \ldots, N$, $j \in \mathcal{S}_i$; $\Delta \mathbf{d} \in \mathbb{R}^N$ is the demand deviations vector of $\Delta d_i = d_i - \hat{d}_i$ for each reservoir $i = 1, \ldots, N$; and $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{H}$ are the state, control, and demand matrices, respectively. In particular, $\mathbf{F} \in \mathbb{R}^{N \times N}$ is a square matrix with diagonal elements $F_{ii} = -(1 - \hat{\beta}_{ii}) O'_i(\hat{n}_i(t))$ if $i \in \mathcal{S}_i$, and $F_{ii} = -O'_i(\hat{n}_i(t))$ otherwise, and off-diagonal elements $F_{ji} = \hat{\beta}_{ji}(t) O'_j(\hat{n}_j(t))$ if $j \in \mathcal{S}_i$, and $F_{ji} = 0$ otherwise; $\mathbf{G} \in \mathbb{R}^{N \times M}$ is a rectangular matrix, where $M \leqslant N^2$ (depends on the network partition and the set $\mathcal{S}_i$, $i = 1, \ldots, N$) with elements $G_{ji} = O_j(\hat{n}_j(t))$ if the origin reservoir j is reachable from the destination reservoir i, and $G_{ji} = 0$ otherwise; $\mathbf{H}$ is an identity square matrix of dimension N (see Appendix A.1 for more details). It should be noted that each reservoir $i = 1, \ldots, N$ is equipped with (at least) one boundary controller $\beta_{ij}, j \in \mathcal{S}_i$ and it might be equipped with one perimeter controller β_{ii} (depends on the network partition and the set $\mathcal{S}_i$, $i = 1, \ldots, N$) thus the number of control variables M is greater than the number of state variables N for any network partition and the linear system (9) of a multi-reservoir network is completely controllable.

The continuous-time linear state system (9) of the multi-reservoir network may be directly translated in discrete-time, using Euler first-order time discretization with sample time T, as follows

$$\Delta \mathbf{n}(k+1) = \mathbf{A} \Delta \mathbf{n}(k) + \mathbf{B} \Delta \boldsymbol{\beta}(k) + \Delta \mathbf{d}(k) \tag{10}$$

where k is the discrete time index, and $\mathbf{A} = e^{\mathbf{F}T} \approx (\mathbf{I} + \frac{1}{2}\mathbf{A}T)(\mathbf{I} - \frac{1}{2}\mathbf{A}T)^{-1}$, $\mathbf{B} = \mathbf{F}^{-1}(\mathbf{A} - \mathbf{I})\mathbf{G}$ (if $\mathbf{F}$ is non-singular) are the state and control matrices of the corresponding discrete-time system. This discrete-time linear model (10) will be used as a basis for feedback control design in the subsequent sections (see Appendix A).

3. Multivariable feedback regulators for perimeter and boundary flow control

The linear control theory offers a number of methods and theoretical results for feedback regulator design in a systematic and efficient way. Multivariable feedback regulators have been applied in the transport area mainly for coordinated ramp metering (Papageorgiou et al., 1990) and traffic signal control (Diakaki et al., 2002; Diakaki et al., 2003; Aboudolas et al., 2009). In the sequel, we present two alternative optimal control methods for the design of feedback perimeter and boundary flow control strategies for multi-region and heterogeneously loaded networks. The first methodology is a multivariable feedback regulator derived through the formulation of the problem as a Linear-Quadratic (LQ) optimal control problem. The second methodology obtained through the formulation of the problem as a Linear-Quadratic-Integral (LQI) optimal control problem, which provides zero steady-state error under persistent disturbances and eliminates the need of set values $\hat{\beta}_{ji}$.

3.1. Perimeter and boundary flow control objectives

In the case of a single-reservoir system (Fig. 1(a)) which exhibits an MFD, a suitable control objective is to minimize the total time that vehicles spend in the system including both time waiting to enter and time traveling in the network. It is known that the corresponding optimal policy is to allow as many vehicles to enter the network as possible without allowing the accumulation to reach states in the congested regime. This policy can be formalized as follows Daganzo (2007): when the network operating in the uncongested regime ($n < \tilde{n}$), vehicles are allowed to enter the perimeter of the network as quickly as they arrive with respect to the critical accumulation $\tilde{n}$; once accumulation reaches $\tilde{n}$ (i.e. $n \geqslant \tilde{n}$) entrance to the network is limited to the minimum entrance flow. It is well-known that this policy corresponds to the so-called "bang-bang control" (BBC) given by

$$q_{\text{in}}(k) = \begin{cases} q_{\max} & \text{if } n(k) < \tilde{n} \text{ and } n(k+1) < \tilde{n} \\ q_{\min} & \text{else} \end{cases} \tag{11}$$

where $q_{\min}$ and $q_{\max}$ are the minimum and maximum entrance flow, respectively. Bang-bang control works well when the system under consideration has relatively slow dynamics, but tends to oscillate between the extremes $q_{\min}$ and $q_{\max}$.

In the case of a multi-reservoir system (Fig. 1(b)), however, a single-reservoir bang-bang policy (11) may induce uneven distribution of vehicles in the reservoirs, and, as a consequence, may invalidate the homogeneity assumption of traffic loads within the reservoirs and degrade the total network throughput and efficiency. As it is demonstrated later in the paper, the critical accumulation $\tilde{n}$ and the maximum output $O(\tilde{n})$ of a network modeled as a single-reservoir system can be different from the critical accumulation $\tilde{n}_i, i = 1, \ldots, N$ and the maximum output $O_i(\tilde{n}_i), i = 1, \ldots, N$ of the same network partitioned in N reservoirs, i.e. $\tilde{n}$ is not necessarily equal to $\sum_{i=1}^{N} \tilde{n}_i$, as the different regions might not reach the critical values simultaneously. Moreover, the time each of the reservoirs reaches the congested regime is very different.

With these observations at hand, a suitable control objective for a multi-reservoir system aims at: (I) distributing the accumulation of vehicles n_i in each reservoir i as homogeneously as possible over time and the network reservoirs, and (II) maintaining the rate of vehicles β_{ji} that are allowed to enter each reservoir around a set (desired) point $\hat{\beta}_{ji}$ while the system's throughput is maximized. A possible way to act in the sense of point (I) is to equalize the distribution of the relative accumulation of vehicles $n_i/n_{i,\max}$ despite inhomogeneous time and space distribution of arrival flows. Requirement (II) is taken by setting the desired point $\hat{\beta}_{ji}$ be equal to the rate of vehicles correspond to output $O_j(\hat{n}_j), i = 1, \ldots, N, j \in \mathcal{S}_i$.

The specification of set points $\hat{n}_i$ (and corresponding $\hat{\beta}_{ji}$) for monocentric networks with well-defined destination attractions is easy, while heterogeneous networks with multiple regions of attraction would require a non-trivial choice of $\hat{n}_i$. Physically speaking, if a control approach can keep all regions below or close to the critical accumulation of each MFD, $\tilde{n}_i$ that maximizes the regional outflow, then the problem is well resolved. A challenge, which will be investigated in the future, is the dynamic partitioning of a network and the dynamic choice of $\hat{n}_i$ as a functions of the level of congestion in each region, $n_i(k)$ and the distribution of destinations across the network. For example if heavily directional flows from the periphery of a network pass through a small region to enter the center, the set point for the small region should be smaller than the set point of the periphery. In case it is not possible to keep $n_i(k) < \tilde{n}_i, \forall i = 1, \ldots, N$, a controller can be designed with $\hat{n}_i$ deviating by little (e.g. 10–20%) from the critical accumulation $\tilde{n}_i$, in such a way to prevent congestion from the reservoir with the highest density of destinations.

3.2. Multivariable feedback regulator

A first approach towards feedback perimeter and boundary control based on the dynamics for a network partitioned in N reservoirs in (10) and the control objective mentioned in the previous section is derived as follows. We consider the following quadratic cost criterion that expresses the control objectives in mathematical terms:

$$\mathcal{L}(\boldsymbol{\beta}) = \frac{1}{2} \sum_{k=0}^{\infty} \left(\|\Delta\mathbf{n}(k)\|_{\mathbf{Q}}^2 + \|\Delta\boldsymbol{\beta}(k)\|_{\mathbf{R}}^2 \right) \tag{12}$$

where $\mathbf{Q}$ and $\mathbf{R}$ are diagonal weighting matrices that are positive semi-definite and positive definite, respectively. The first term in (12) is responsible for minimization and balancing of the relative accumulation of vehicles $n_i/n_{i,\max}$ in each reservoir

i (objective (I)). To this end, the diagonal elements of $\mathbf{Q}$ are set equal to the inverses of the maximum accumulation of the corresponding reservoirs (see Diakaki et al., 2002; Aboudolas et al., 2009 for details). The second term in (12) is responsible for objective (II) in Section 3.1 and the choice of the weighting matrix $\mathbf{R} = r\mathbf{I}$ can influence the magnitude of the control actions.

Minimization of the performance criterion in (12) subject to (10) (assuming $\Delta\mathbf{d}(k) = \mathbf{0}$) leads to the LQ multivariable feedback regulator (see Appendix A.2)

$$\boldsymbol{\beta}(k) = \hat{\boldsymbol{\beta}} - \mathbf{K}[\mathbf{n}(k) - \hat{\mathbf{n}}] \tag{13}$$

where matrix $\mathbf{K}$ is the steady-state solution of the corresponding Riccati equation, which depends only upon the problem matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{Q}$, and $\mathbf{R}$. Note that the corresponding discrete-time linear system (10) is completely controllable and reachable and as a consequence a dead-beat gain $\mathbf{K}$ can be off-line calculated for a low value of the scalar weight r, i.e. regulator (13) bringing the system (10) to steady state in (at most) N steps.

3.3. Multivariable integral feedback regulator

The basic approach in integral feedback control is to create a state within the controller that computes the integral of the error signal, which is then used as a feedback term to provide zero steady-state error. We do this by augmenting the description of the original system (10) with a new state given by

$$\mathbf{z}(k + 1) = \mathbf{z}(k) + \mathbf{Y}\Delta\mathbf{n}(k) \tag{14}$$

where $\mathbf{z} \in \mathbb{R}^p$ is the integral vector, $\mathbf{Y} \in \mathbb{R}^{p \times N}$, and $p \leqslant M$ must hold for control-theoretic reasons (see Appendix A.1). The matrix $\mathbf{Y}$ typically consists of 0's and 1's such that p components (or linear combinations of components) of accumulation of vehicles are integrated in (10). The augmented discrete-time system (10), (14) can be written in compact form as

$$\Delta\tilde{\mathbf{n}}(k + 1) = \widetilde{\mathbf{A}}\Delta\tilde{\mathbf{n}}(k) + \widetilde{\mathbf{B}}\Delta\boldsymbol{\beta}(k) + \widetilde{\mathbf{H}}\Delta\mathbf{d}(k) \tag{15}$$

where $\tilde{\mathbf{n}}(k) = [\mathbf{n}(k) \quad \mathbf{z}(k)]^{\tau}$ is the augmented state vector, and $\widetilde{\mathbf{A}}$, $\widetilde{\mathbf{B}}$, $\widetilde{\mathbf{H}}$ are the augmented state, control, and demand matrices, respectively (see Appendix A.3 for the structure of the augmented matrices). For deriving the integral feedback regulator, the control goal is to minimize the augmented quadratic criterion

$$\mathcal{L}(\boldsymbol{\beta}) = \frac{1}{2}\sum_{k=0}^{\infty}\left(\|\Delta\mathbf{n}(k)\|_{\mathbf{Q}}^2 + \|\Delta\boldsymbol{\beta}(k)\|_{\mathbf{R}}^2 + \|\mathbf{z}(k)\|_{\mathbf{S}}^2\right) \tag{16}$$

where $\mathbf{S}$ is an additional positive semi-definite diagonal weighting matrix. Similarly to the LQ cost criterion (12), the first term in (16) is responsible for objective (I) in Section 3.1, i.e. minimization and balancing of the relative accumulation of vehicles $n_i/n_{i,\max}$ in each reservoir i. To this end, the diagonal elements of $\mathbf{Q}$ are set equal to the inverses of the maximum accumulation of the corresponding reservoirs (see Section 3.2). The second term is responsible for objective (II), while the third term corresponds to the magnitude of the error signal. The choice of the weighting matrices $\mathbf{R} = r\mathbf{I}$ and $\mathbf{S} = s\mathbf{I}$, where r, s are positive scalars, is performed via a trial-and-error procedure so as to achieve a satisfactory control behavior (i.e. non-oscillatory behavior, good quantitative and qualitative performance) for a given multi-reservoir network. The trial-and-error procedure can be conducted by designing the controller with different (r, s)-values (different $\mathbf{R}$ and $\mathbf{S}$ matrices, see Appendix A for more details) and assessing the results for representative scenarios of demand.

Minimization of the performance criterion (16) subject to (15) (assuming $\Delta\mathbf{d}(k) = \mathbf{0}$) leads to the LQI multivariable feedback regulator (see Appendix A.3)

$$\Delta\boldsymbol{\beta}(k) = -\widetilde{\mathbf{K}}\begin{bmatrix}\Delta\mathbf{n}(k)\\ \mathbf{z}(k)\end{bmatrix} \tag{17}$$

where $\widetilde{\mathbf{K}}$ is the steady-state solution of the corresponding Riccati equation. Decomposing $\widetilde{\mathbf{K}} = [\mathbf{K}_1 \quad \mathbf{K}_2]$, we get the final multivariable integral feedback regulator (see Appendix A.3)

$$\boldsymbol{\beta}(k) = \boldsymbol{\beta}(k-1) - \mathbf{K}_p[\mathbf{n}(k) - \mathbf{n}(k-1)] - \mathbf{K}_I[\mathbf{n}(k) - \hat{\mathbf{n}}] \tag{18}$$

where $\mathbf{K}_p = \mathbf{K}_1 - \mathbf{K}_2\mathbf{Y}$ and $\mathbf{K}_I = \mathbf{K}_2\mathbf{Y}$ are the proportional and integral gains, respectively. The calculation of $\widetilde{\mathbf{K}}$ via solution of the discrete-time Riccati equation, is straightforward and the required computational effort is low even for large-scale networks partitioned in many reservoirs. Moreover, this computational effort is required only off-line, while on-line (i.e. in real-time) the calculations are limited to the execution of (18) with a given constant control matrix $\widetilde{\mathbf{K}}$ and state measurements $\mathbf{n}(k)$. Finally, it is well-known in linear control theory that the integral multivariable regulator (18) leads automatically to $\mathbf{n} = \hat{\mathbf{n}}$ under steady state conditions (disturbance rejection), i.e. when the multi-reservoir system (15) evolves slowly with time (cf. our assumption in Section 2) and $\Delta\mathbf{d}(k)$ is constant or slowly time-varying. This feature is particularly useful since the implementation of the ordered value $\boldsymbol{\beta}(k)$ in real-life networks is biased due to infrastructure limitations, as we will see later (see Sections 3.4 and 4).

3.4. Constraints and implementation issues

We conclude this section with some remarks pertaining to the control and state constraints of a multi-reservoir system and to the implementation of the multivariable feedback perimeter and boundary flow control in real-time.

A potential disadvantage of the linear-quadratic theory is that it does not allow for direct consideration of the inequality constraints (3)–(5). In this work, the control constraints (3) are imposed after application of the feedback regulators (13) or (18) as we will see later. Regarding the state constraints (5), one may see that the balancing of the relative accumulation of vehicles $(n_i - \hat{n}_i)^2/n_{i,max}$ via the control objective (12) or (16) reduces the risk of a reservoir to reach the congested regime in an indirect way. Finally, the overflow constraints (4) can be satisfied by appropriate selection of $\beta_{ji,min}$, $\beta_{ji,max}$. Alternatively, one can solve the same problem as a Model-Predictive perimeter Control (MPC) problem including all constraints by using the current state (current estimates of the accumulation in each reservoir) of the traffic system as the initial state $\mathbf{n}(0)$ as well as predicted demand flows $\mathbf{d}(k)$ over the a finite-time horizon (Geroliminis et al., 2013). However, MPC requires that a (quadratic or non-linear) optimization problem be solved and future demands be predicted in real-time, and thus more effort is needed for online use. On the other hand, if demand flow predictions are available (i.e. $\Delta \mathbf{d}(k) \neq \mathbf{0}$ in Sections 3.2 and 3.3) cheap and efficient feedforward control can be used to improve the response of the system under uncertainty (Papageorgiou, 1996).

The feedback regulators (13) or (18) are activated in real-time at each control interval T and only within specific time windows (e.g. by use of two thresholds $n_{i,act}$ and $n_{i,stop}$[1]), based on the current accumulation $\mathbf{n}(k)$, to calculate the fraction of flow rate $\boldsymbol{\beta}(k)$ to be allowed to enter each reservoir and transfer between neighborhood reservoirs. The required real-time information on the vehicle accumulation $\mathbf{n}(k)$ can be directly obtained via loop detector time-occupancy measurements. The loop detectors may be placed anywhere within the link, but the estimation is most accurate for detector locations around the upstream middle of the link. While occupancy measures do not provide an unbiased estimator of $\mathbf{n}(k)$ for all levels of congestion, recent work in queue length and travel time estimation (see e.g. Geroliminis and Skabardonis, 2011; Ban et al., 2011) can be integrated in the above framework for a more accurate estimation of $\mathbf{n}(k)$ with different types of sensors. Recent studies (Buisson and Ladier, 2009; Courbon and Leclercq, 2011) have shown that the location of loop detectors can affect the shape of the estimated MFD and the value of critical accumulation, as the occupancy value is representative in the proximity of the detector and not for the whole link. Nevertheless, a well-defined $\hat{\mathbf{n}}$ is obtained for a given location of detectors and it can be utilized in the design of the regulators. Thus, an accurate estimation of $\mathbf{n}(k)$ is not expected to improve the performance of the perimeter controller but a more careful consideration can be a future research direction. Note also that loop detectors do not directly measure network outflow, but the network circulating flow. Nevertheless, given the roughly linear relationship between the two (see e.g. Geroliminis and Daganzo, 2008) the estimated $\mathbf{n}(k)$ is proper for the design.

After the application of the feedback regulators (13) or (18), if the ordered value $\boldsymbol{\beta}(k)$ violates the operational constraints (3), it should be adjusted to become feasible, i.e. truncated to $[\boldsymbol{\beta}_{min}, \boldsymbol{\beta}_{max}]$. Moreover, the values of $\boldsymbol{\beta}(k-1)$ used on the right-hand side of (18), should be the bounded values of the previous time step (i.e. after the application of constraints (3)) to avoid possible wind-up phenomena in the regulator. The obtained $\beta_{ji}(k)$ values are then converted to arriving flows $q_{ji}(k)$ (by multiplying $\beta_{ji}(k)$ by $O_j(n_j(k))$) and used to define the green periods of the signalized intersections located at the boundary of neighborhood reservoirs or the perimeter of the network. To this end, the latter flows are equally distributed to the corresponding intersections and converted to an entrance link green stage duration with respect to the saturation flow of the link and the cycle time of the intersections. More specifically, traffic signals at the perimeter of the network and the boundary of neighborhood reservoirs that belong to the set I_{ji} are operated on the basis of a number of fixed stages and cycle-time C_{ji} that is always equal (or equivalent fraction) of the perimeter control period T. For given I_{ji}, $q_{ji}(k)$ and C_{ji}, the implemented entrance link green may be calculated from $G_{ji} = (q_{ji}(k)C_{ji})/(S_{ji}|I_{ji}|)$, where S_{ji} (in veh/h) is the entrance links' saturation flow, typically equal to $\lambda \cdot 1800$ veh/h, where λ denotes the number of entrance link lanes. In case that the ordered value for implementation is very different than the actual one (due to infrastructure limitations, queue spillbacks or wasted green), straightforward techniques can be applied to overcome this deficiency (e.g. queue equalization or increase of entrance link green in low demand intersections).

4. Implementation

4.1. Network description and simulation setup

The test site is a 2.5 square mile area of Downtown San Francisco (Financial District and South of Market Area), including about 100 intersections and 400 links with lengths varying from 400 to 1300 feet (Fig. 2(a)). The number of lanes for through traffic varies from 2 to 5 lanes and the free flow speed is 30 miles per hour. Traffic signals are all multiphase fixed-time operating on a common cycle length of 90 s for the west boundary of the area (The Embarcadero) and 60 s for the rest.

For the simulation tests, the test area of Downtown San Francisco is modeled via the AIMSUN microscopic simulator and typical loop-detectors have been installed around the middle of each network link, according to Fig. 2(b). The simulation step for the microscopic simulation model of the test site, was set to 0.5 s. For the application of the proposed perimeter and

---

[1] These thresholds should be selected lower than the set accumulation $\hat{n}_i, i = 1, \dots, N$ to avoid possible oscillatory behavior of the regulators.
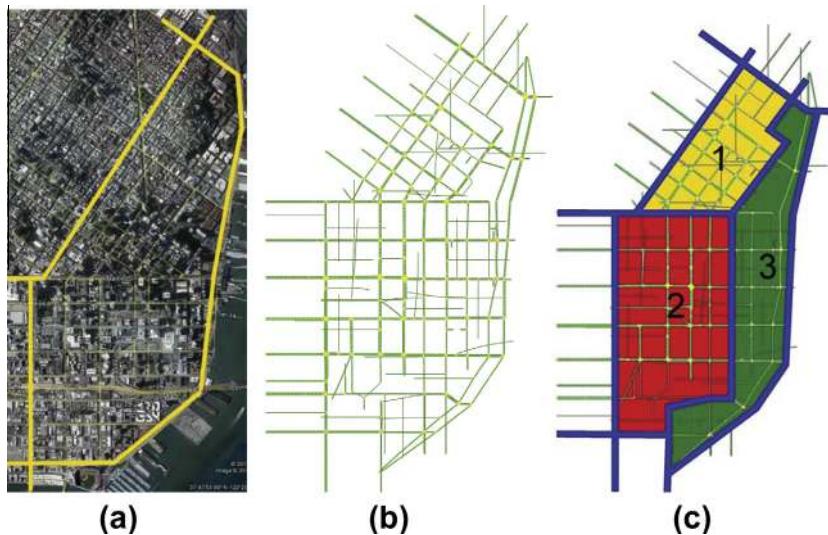
Fig. 2. The test site of Downtown San Francisco: (a) real network; (b) simulation model; (c) partitioning of the network into 3 reservoirs.

boundary control strategies, the test site is partitioned into three homogeneous reservoirs ($N = 3$) with small variances of link densities (Ji and Geroliminis, 2012), according to Fig. 2(c). The three reservoirs are separated by blue[2] lines in Fig. 2(b), and consist of 112 (yellow colored area), 128 (red colored area), and 147 links (green colored area), respectively.

Initially, to derive and investigate the shape of the MFDs of the three reservoirs, simulations are performed with a field-applied, fixed-time signal control plan. To account for stochastic effects of the simulator, ten replications (with different seeds) were carried out for a 4-h (8:00–12:00) time-dependent scenario with strong demand. During this scenario the network is filled and severe congestion is faced for 2 h with many link queues spilling back into upstream links. Based on the derived MFDs the BBC controller (11) and the proposed feedback regulators FPC-LQ (13) and FPC-LQI (18) are designed. Additionally, two 6.5 h (9:00–14:30) scenarios based on real origin–destination (OD) data were defined in order to compare the aforementioned perimeter control strategies (BBC, FPC-LQ and FPC-LQI) with the no control case under different traffic conditions. To simulate somewhat adaptive drivers and account for drivers' route choice effects in the OD scenarios, the Dynamic Traffic Assignment (DTA) module (C-Logit route choice model (Cascetta et al., 1996)) is activated every 3 min, a time interval that is consistent with the average trip length in the test area of San Francisco. All strategies are applied every $T = 180$ s, a control interval that is twofold or threefold to the cycle length of all the considered intersections. Finally, all strategies' decisions are modified to satisfy the constraints (3). These decisions are then forwarded to $|I_{ij}| = 25$ signalized intersections located the perimeter (15 intersections) or the boundary of neighborhood reservoirs (10 intersections) of the test network for application, i.e. by modifying the green duration of the phases where perimeter and boundary arriving flows are involved, as described in Section 3.4. Note that when the single-reservoir bang-bang strategy (11) is applied only intersections located at the perimeter of the test network are modified. Previous work (Daganzo and Geroliminis, 2008; Aboudolas et al., 2010; Geroliminis and Boyaci, 2012; Zhang et al., 2013) have shown that traffic-responsive signal control strategies and different signal settings can change the shape of the MFD and consequently the critical accumulations. Nevertheless, in our experiments we only control the traffic signal at the perimeter of the network and the boundaries of the three reservoirs and we observe that the critical accumulations do not differ noticeably in the no control and perimeter flow control cases.

4.2. Macroscopic fundamental diagrams and heterogeneity

Fig. 3(a) displays the MFD resulting for the considered demand scenario and ten replications (R1 to R10), each with different seed. This figure plots the throughput-load relationship (veh/h vs. veh) in the network for the whole simulation time period (total network flow) as estimated by the loop detectors. Each measurement point in the diagram corresponds to 180 s. As a first remark, Fig. 3(a) confirms the existence of an MFD for the test area of Downtown San Francisco with moderate scatter across different replications. It can be seen that the maximum throughput values (around $30 \cdot 10^4$ veh/h) in Fig. 3(a) occur in an accumulation range from 4000 to 6000 vehs. If the accumulation is allowed to increase to values of $n > 6000$ veh, then the network becomes severely congested with states in the regime of the MFD where, the throughput decreases with accumulation (negative slope) and the system can lead to network-wide gridlock. For a single-reservoir system in order to prevent this throughput degradation, the accumulation n should be maintained in the mentioned observed range (close to the

---

[2] For interpretation of color in Fig. 2, the reader is referred to the web version of this article.
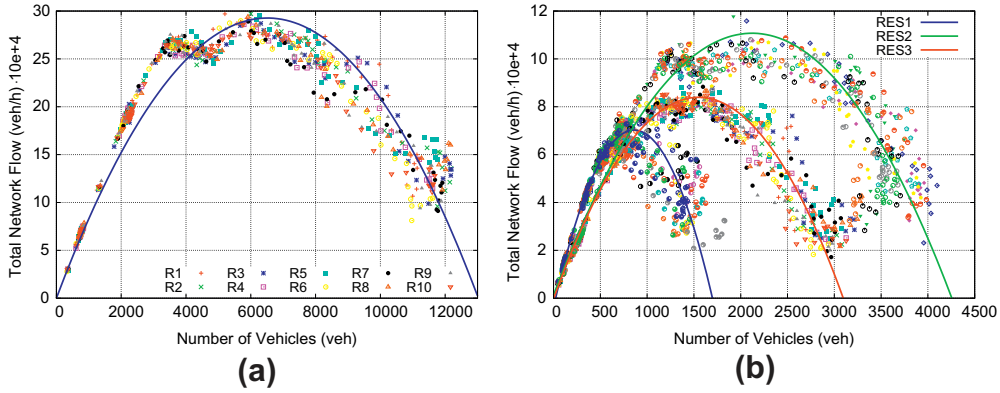
Fig. 3. Macroscopic Fundamental Diagrams for the: (a) whole network; (b) three reservoirs.

critical accumulation $\tilde{n} \approx 6000$ veh) during the heart of the rush while the system's throughput is maximized (Daganzo, 2007).

Fig. 3(b) displays the MFDs of the three reservoirs resulting for the considered demand scenario and ten replications. It can be seen that all three reservoirs experience MFD with quite moderate scatter across different replications. There is a clear distinction between congested and uncongested regime for all reservoirs. Nevertheless, an interesting observation is that the time each of the reservoirs reaches the congested regime is very different. The reservoir 3 (red curve) reaches congestion at time 10:30 (for an accumulation $n_3 \approx 1500$ veh) and then it propagates in the reservoirs 2 (green curve) and 1 (blue curve), at time 10:45 (for $n_2 \approx 2000$ veh) and 11:00 (for $n_1 \approx 750$ veh), respectively. This propagation of congestion would not be observable by looking at the unified MFD in Fig. 3(a), which reaches the congestion at time 10:45 for an accumulation $n \approx 6000$ veh. It also postpones the activation of the controller for reservoir 3 and treats all reservoirs equivalently. This establishes our heterogeneity presumption stated in Section 3.1. If a single-reservoir perimeter control is applied, then a uniform strategy will restrict input in all 3 reservoirs, while at that time each of them is in a different regime of each own MFD (uncongested for 1, congested for 2 and at critical for 3). Definitely, this strategy will be suboptimal as each reservoir should be treated differently. Note as well that the maximum achievable throughput (outflow) is different for each reservoir $i = 1$, 2,3 (around $7 \cdot 10^4$ veh/h, $11 \cdot 10^4$ veh/h, and $8 \cdot 10^4$ veh/h, respectively) and occur in accumulation ranges [500, 1000] veh, [1100, 2250] veh, and [1000, 1700] veh, respectively. The difference in maximum flow levels and congested regimes imply corresponding differences of the highest accumulation of vehicles (load) $n_{i,\max}$ that is reached by each reservoir $i = 1$, 2, 3, which is applied in the cost criteria (12) and (16) via the weighting matrix **Q**.

4.3. Design of the perimeter flow control strategies

We now perform the implementation and comparison of the proposed strategies FPC-LQ, FPC-LQI with the BBC strategy corresponding to the design and application of the feedback regulators (13), (18) and the bang-bang controller (11). For the design of the bang-bang controller (11) the value $\hat{n} = 5500$ (around 90% of the critical accumulation $\tilde{n}$) is selected for the critical accumulation according to the analysis in Section 4.2 (see Fig. 3(a)). For the proposed partitioning, each reservoir is reachable from the perimeter and the boundary, i.e. $\mathcal{S}_i = \{1, 2, 3\}$, $\forall i = 1, 2, 3$ (see Fig. 2(c)). Thus for the design of the proposed strategies, each reservoir i is equipped with one perimeter controller β_{ii} and two boundary controllers $\beta_{ji}, j \neq i$, and the control vector is given by $\boldsymbol{\beta} = [\beta_{11} \quad \beta_{21} \quad \beta_{31} \quad \beta_{12} \quad \beta_{22} \quad \beta_{32} \quad \beta_{13} \quad \beta_{23} \quad \beta_{33}]^{\tau}$. The state vector $\mathbf{n}(k)$ includes the accumulation of vehicles for each reservoir i and is given by $\mathbf{n} = [n_1 \quad n_2 \quad n_3]^{\tau}$.

The set (desired) accumulation $\hat{n}_i$ for each reservoir i is selected within the optimal range of the corresponding MFD for maximum output, given the analysis in Section 4.2. More specifically, the following values $\hat{n}_1 = 600$ veh, $\hat{n}_2 = 1250$ veh, and $\hat{n}_3 = 1100$ veh are selected for the current implementation (see Fig. 3(b)). The desired flow rate of the control inputs are based on the corresponding $O_i(\hat{n}_i)$ values for each reservoir i and given by $\hat{\boldsymbol{\beta}} = [0.3 \quad 0.2 \quad 0.25 \quad 0.35 \quad 0.25 0.35 0.3 0.2 0.25]^{\tau}$. Finally, the minimum and maximum permissible rates are chosen to satisfy the overflow constraints (4) and given by $\boldsymbol{\beta}_{\min} = \mathbf{0.1}^{\tau}$ and $\boldsymbol{\beta}_{max} = [0.6 \quad 0.4 \quad 0.5 \quad 0.7 \quad 0.5 \quad 0.7 \quad 0.6 \quad 0.4 \quad 0.5]^{\tau}$, respectively.

For the derivation of the gain matrix $\mathbf{K} \in \mathbb{R}^{9 \times 3}$ in (13) it suffices to specify the state matrices $\mathbf{A} \in \mathbb{R}^{3 \times 3}$, $\mathbf{B} \in \mathbb{R}^{3 \times 9}$, and the weighting matrices $\mathbf{Q} \in \mathbb{R}^{3 \times 3}$, $\mathbf{R} \in \mathbb{R}^{9 \times 9}$. Accordingly, for the derivation of the gain matrix $\widetilde{\mathbf{K}} \in \mathbb{R}^{9 \times 6}$ in (17) it suffices to specify the state matrices $\widetilde{\mathbf{A}} \in \mathbb{R}^{6 \times 6}$, $\widetilde{\mathbf{B}} \in \mathbb{R}^{6 \times 9}$, and the weighting matrices $\mathbf{Q} \in \mathbb{R}^{3 \times 3}$, $\widetilde{\mathbf{R}} = \mathbf{R} \in \mathbb{R}^{9 \times 9}$, $\mathbf{S} \in \mathbb{R}^{3 \times 3}$. All state matrices are developed for the particular network on the basis of the selected set (desired) point $\hat{\mathbf{n}}$, $\hat{\boldsymbol{\beta}}$, the matrix $\mathbf{Y} = \mathbf{I}_3$ (only for (18)), and the linearization according to (8) and (10). The weighting matrices (**Q**, **R**, and **S**) in the quadratic cost criteria (12), (16) are chosen diagonal. More precisely, the diagonal elements of **Q** are set equal to the inverses of the maximum accumulation of the corresponding reservoirs, i.e. $Q_{ii} = 1/n_{i,\max}$, $i = 1, 2, 3$ (see Sections 3.1, 3.2, and 3.3). For the LQ cost criterion (12), the diagonal elements of matrix **R** were set equal to $r = 0.00001$ while for the LQI cost criterion (16) the diagonal

elements of matrices **R** and **S** were set equal to $r = 0.005$ and $s = 0.0001$, respectively. These low values of the scalar weights r and r, s were found to lead to dead-beat gains **K** and $\widetilde{\mathbf{K}}$ (not shown) that exhibit good performance. Note that each row of the gain matrices **K** and $\widetilde{\mathbf{K}}$ contains the non-zero elements (weights) of the corresponding reservoirs, which highlights that the accumulation of all reservoirs contributes in the ordered flow in control laws (13) and (18), respectively. Thus interactions between reservoirs are taken into consideration.

5. Results and insights

In the sequel, we present simulation results for the proposed perimeter control strategies that are obtained by applying a non-adaptive demand scenario (Scenario 1, based on pre-specified turning movements at intersections) and two OD demand profiles (Scenarios 2 and 3). In the OD scenarios the DTA module is activated in the microsimulator and (some of) the drivers choose their routes adaptively in response to traffic conditions. Results include the most important traffic performance indices.

5.1. Non-adaptive demand scenarios

The simulation results for non-adaptive demand are summarized in (i) Fig. 4 that graphically describes in details the evolution of congestion for each reservoir under no control and FPC-LQ control and (ii) Table 1 that presents different performance indices (average of all replications).

Fig. 4(a) and (b) display the MFDs of the three reservoirs resulting for the considered non-adaptive demand scenario and one replication when no control and perimeter control are applied, respectively. Clearly, when perimeter control is applied, the three reservoirs remain semi-congested and only a few states observed in the congested regime; under no control, the network becomes severely congested with states in the congested regime of the corresponding MFDs. Note that, in absence of perimeter control, the output at the end of the simulation is around $2 \cdot 10^4$ veh/h for all reservoirs.

To further illustrate the perimeter strategy actions, the flow of the three reservoirs for one replication are depicted in Fig. 4(c) and (d). Traffic conditions are identical for both control cases up to around 9:15 am, when perimeter control is switched on (due to reservoir 3), as accumulation n_i reaches its set point $\hat{n}_i$ for each reservoir i (see Fig. 4(a) and (b)), albeit at different times (reservoirs 2 and 1 are switched on at time 10:30 and 10:50, respectively), the perimeter strategy (a) restricts the rate vehicles are allowed to enter the network to keep it from becoming congested, and (b) manages the inter-transfers between the reservoirs to respect homogeneity in the network reservoirs over time. Thus, output is maintained at high levels that is close to the target points $O(\hat{n}_i)$ for each reservoir i (corresponding to $\hat{\beta}$), i.e. around $6 \cdot 10^4$ veh/h, $8.5 \cdot 10^4$ veh/h, and $7 \cdot 10^4$ veh/h (see Fig. 4(d)), respectively, in contrast to the no control case (see Fig. 4(c)). Remarkably, the accumulation n_i of each reservoirs i is not exactly maintained to $\hat{n}_i$ (see Fig. 4(b)) due to the selection of the weights $Q_{ii} = 1/n_{i,\ max}$ in the cost criteria (12) and (16).

Fig. 4(e), f show the throughput (total flow of all links in the network) and the virtual waiting queues for the same replication, respectively. Fig. 4(e) indicates that the perimeter strategy maintains the overall throughput to high values (via appropriate actions within the reservoirs) during the heart of the rush (after 10:30), compared to the no control case, even if it involves slightly longer waiting queues at the origins of the network (Fig. 4(f)). Note that the virtual queue size decreases with time and the difference would diminish if the network was allowed to serve all remaining vehicles.

Table 1 displays the obtained results in terms of the performance indices (PI) *Total Travel Time* (TTT), *Total Distance Traveled* (TDT), *Space-mean Speed*, and *Total output* (total number of vehicles that exit the network) during the whole scenario for the feedback perimeter control (FPC) and No Control (NC) cases.[3] This table also displays the number of vehicles within network links inside the three reservoirs and the virtual waiting queues (in veh) that have been stored at the origin links of the network at the end of simulation, *Vehicles Inside* and *Vehicles Waiting Out – Virtual Queue* (VQ), respectively, because the network is not empty at the end of the simulation. Thus, space-mean speed and TTT were also calculated taking into account the virtual queues. Note that by directly extracting performance measures of space-mean speed or TTT from the simulator will not include vehicles waiting outside the network (virtual queues), which will consistently underestimate the time spent of gated/controlled vehicles.[4] The total number of simulated (generated) vehicles in the network is the sum of the three last rows in Table 1.

As can be seen in Table 1, the perimeter control strategy leads to an improvement of the evaluation criteria compared to no control for the whole network, albeit by different percentages. More specifically, when perimeter control is applied, TTT and space-mean speed are improved in average by 11.7% (5.7% with virtual queues) and 15.4% (8% with virtual queues), respectively, compared to no control. In contrast, the higher virtual waiting queue in perimeter control (8.6% compared to no control) indicates that the control action creates temporary queues at the perimeter of the network (β_{ii} controllers). However, this proves propitious for the total network throughput and the traffic state inside the three reservoirs as we will see later. Clearly, a high virtual waiting queue in perimeter control (compared to no control) is the price to pay for this particular improvement. Finally, the lower number of vehicles within the network links (Vehicles Inside) at the end of simulation (25%

---

[3] The FPC indices were obtained as follows: each strategy FPC-LQ, FPC-LQI was run for ten replications (with different seeds) and the average score of each index was counted; then, the total score obtained by FPC was calculated as the average of FPC-LQ and FPC-LQI.

[4] While the virtual queue and delay estimation outside the external perimeter provide a proxy for the additional delays of vehicles before entering the protected network, future work should analyze the effect in larger networks where the external zone is also simulated in detail.
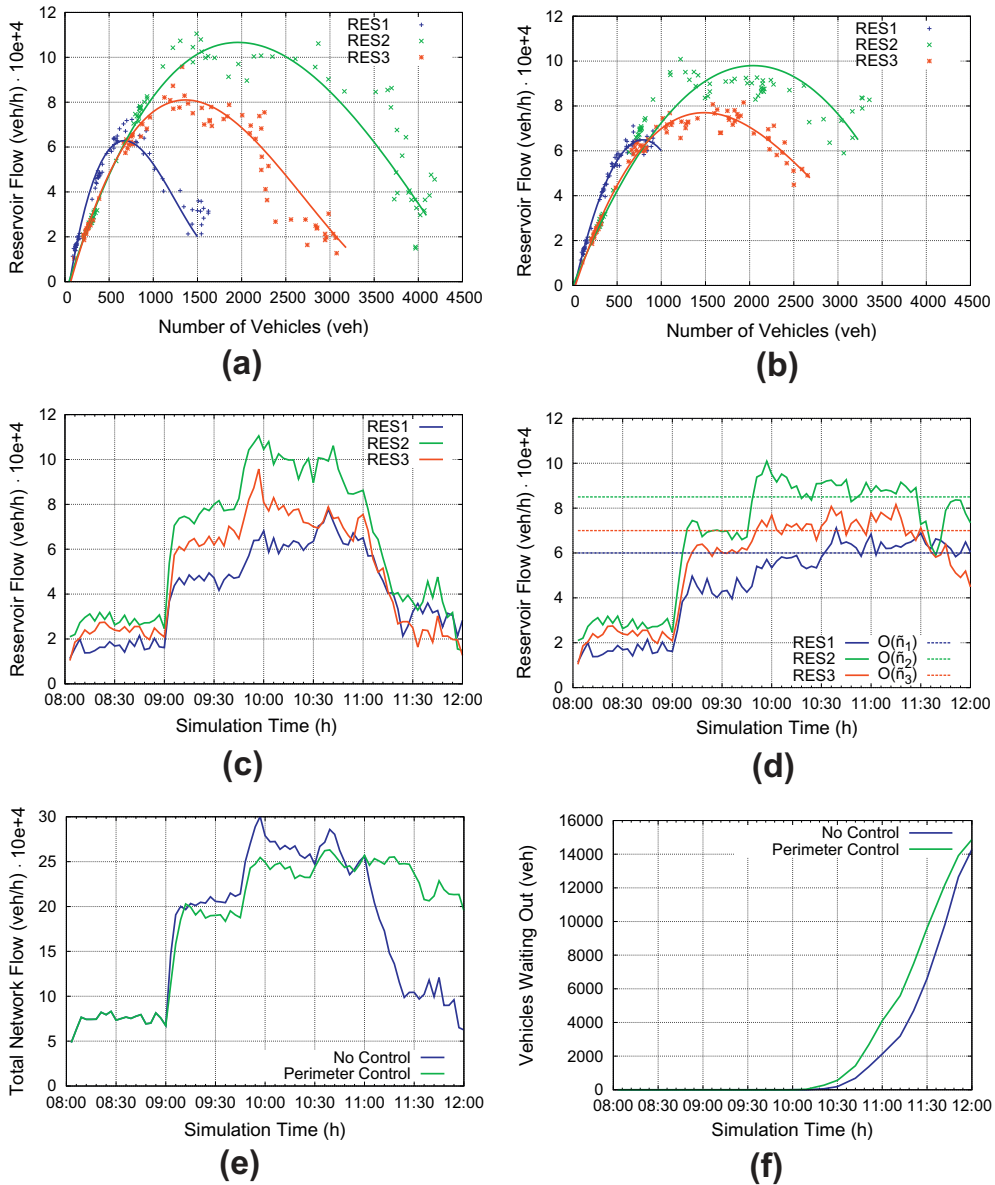
Fig. 4. Simulation results for one replication of the non-adaptive demand Scenario 1: (a) MFDs of the three reservoirs with NC; (b) MFDs of the three reservoirs with FPC-LQ. The continuous curves in (a) and (b) represent the best fit of data to a third-degree polynomial; (c) Reservoir flow over time with NC; (d) Reservoir flow over time with FPC-LQ; (e) Total network flow over time; (f) Vehicles waiting out of the network over time.

Table 1
PIs of the FPC strategy vs. NC for the non-adaptive demand scenario.

Evaluation criteria	No control	FPC	Av. improvement (%)	Units
Total Distance Travelled (TDT)	107,404	109,375	1.8	km
Total Travel Time without VQ (TTT)	18,460	16,296	11.7	veh-h
Total Travel Time with VQ (TTT VQ)	21,317	20,105	5.7	veh-h
Space-mean Speed without VQ = TDT/TTT	5.82	6.71	15.4	km/h
Space-mean Speed with VQ = TDT/TTT VQ	5.04	5.44	8.0	km/h
Vehicles Inside	11,702	8773	25.0	vehs
Vehicles Waiting Out (Virtual Queue)	13,757	14,935	8.6	vehs
Total Output	55,076	56,057	1.8	vehs

compared to no control) indicates that the control action does not create queues that spill back to upstream intersections at the boundary of neighborhood reservoirs (β_{ji}, $i \neq j$, controllers). The simulation ends with a high demand and congestion in

the network. The improvements of Table 1 are expected to be much higher if the simulation is extended to allow vehicles inside the network to reach their destinations, as we will see in Section 5.2.

5.2. Simulation results for adaptive drivers and hysteresis loops

In this section two different OD profiles (Scenarios 2 and 3) are analyzed with 10% higher demand in Scenario 2 than 3. The same performance indices are gathered as in Scenario 1. These scenarios also include an offset of congestion to highlight the additional improvements of FPC. Both feedback perimeter control strategies produce similar performance indices and minor differences in the evolution of congestion and only one of the two (FPC-LQI) is described in details. It is clear physical and quantitative evidence that traffic is improved.

Fig. 5 compares the MFDs for each of the three reservoirs and for the whole network for no control (NC) with a pre-timed signal plan, single-reservoir perimeter control (BBC) and FPC-LQI for one of the replications. Remarkably, the diagrams indicate a hysteresis, i.e. a different path of measurement points when filling the network (onset) than when emptying
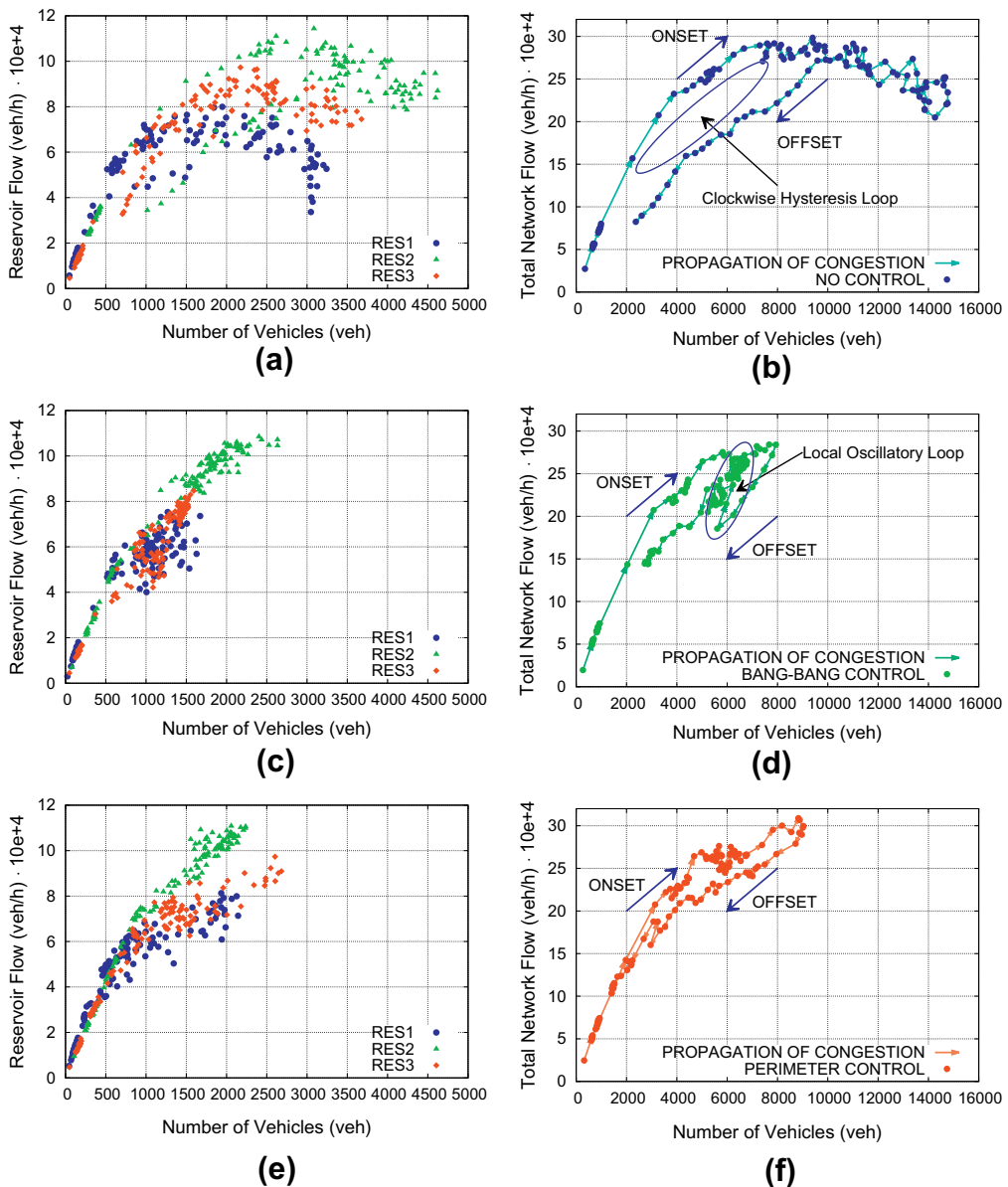

Fig. 5. Simulation results for one replication of the adaptive OD scenario: (a), (c), (e) MFDs of the three reservoirs under NC, BBC, and FPC-LQI, respectively; (b), (d), (f) MFD of the whole network under NC, BBC, and FPC-LQI, respectively.
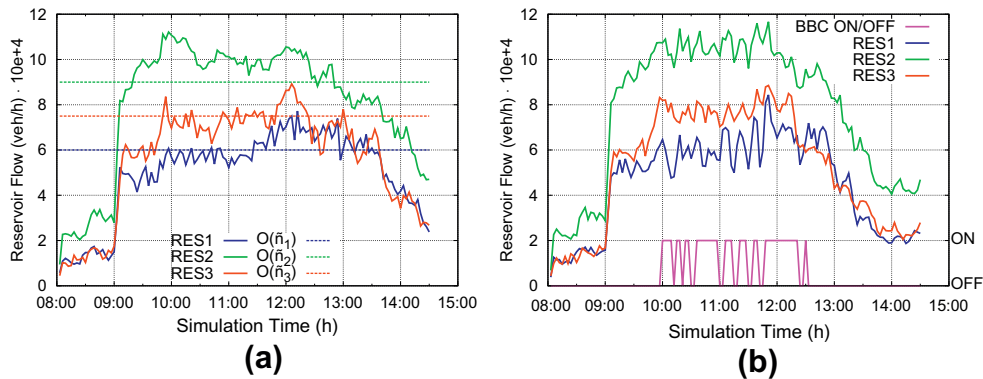
Fig. 6. Comparison of the FPC-LQI with BBC for one replication.

the network (offset). Additionally, the shape of the MFDs for the three reservoirs and the whole network in NC case (Fig. 5(a) and (b)) are shifted (horizontally) to the right, when compared with the MFD shape in the non-adaptive demand case (Fig. 3(a) and (b)). This is because the activation of the DTA allows for the drivers to choose their routes adaptively in response to traffic conditions and utilize less congested routes in the network. As a result, the level of homogeneity within each reservoir improves and network flow increases (especially for congested conditions). This is in accordance with previous observations, see e.g. Mazloumian et al. (2010) for the effect of variance, Gayah and Daganzo (2011b) and Mahmassani et al. (2013) for the efficiency of simple networks with somewhat adaptive drivers and the effect of DTA, respectively. In the NC control case all three reservoirs experience significant congestion with states in the decreasing part of the MFD (Fig. 5(a)). The different distribution of congestion in the onset and the offset creates a strong clock-wise hysteresis loop (Fig. 5(b)), which results in a strong drop in network flow even for states below the critical accumulation. Note a 30% capacity drop for accumulation n = 6500 veh.

Both BBC and FPC strategies succeed to improve the flow of traffic and avoid states in the congested regimes of MFDs (Fig. 5(c) and (e)). Nevertheless, the FPC-LQI succeeds better improvements compared to BBC from multiple perspectives. The higher degree of homogeneity among reservoirs in FPC (Fig. 5(e)) creates a higher performance both in the onset and offset of congestion and significantly decreases the hysteresis (Fig. 5(f)) observed both at the NC and BBC case (Fig. 5(b) and (d), respectively). Note the local oscillatory loop which creates a capacity loss around 15% and the strong fluctuation of flow. The reasons for hysteresis in the BBC case are that despite that the total network accumulation is around the critical value (around 6000 veh), individual reservoirs exhibit oscillations due to the bang-bang policy and due to no consideration of individual treatment of each reservoir (input flow is distributed equally among all traffic signals in the perimeter, as described in Section 3.4). Note also the significantly higher scatter of MFD for reservoir 1, even in the uncongested regime (Fig. 5(c)).

This underlines that appropriate designed perimeter control strategies for multi-reservoir systems might prove beneficial in ameliorating deficiencies associated with single-reservoir systems (e.g. propagation of congestion). The main reasons are that (i) the controllers for each reservoir are activated at different times, (ii) the cost criteria (12) and (16) allow a more homogeneous distribution of accumulations among reservoirs (compare Fig. 5(c) and (e)), (iii) input flows at each external perimeter at treated differently and (iv) transfer flows across reservoirs are also controlled to respect homogeneity in loads and the network reservoirs over time. These properties results to a higher obtained maximum flow (as can be seen by comparing Fig. 5(d) with Fig. 5(f)). These happens because reservoir 2 accumulation is retained around 2300 veh for FPC (compared to 2700 veh for BBC) without significant capacity loss, while the other 2 reservoirs obtain higher accumulations that result to higher flows (compare Fig. 5(c) with Fig. 5(e)).

A further analysis of the properties of the two strategies (FPC vs. BBC) can shed more light in the oscillatory behavior and higher hysteresis of BBC. Fig. 6 depicts the flow of the three reservoirs for one replication. As a first remark, Fig. 6(b) confirms the oscillatory behavior of the BBC strategy since the flow values within the three reservoirs exhibit high-frequency variations over time (compared to FPC). More specifically, the BBC strategy is activated after 10:00 am, as accumulation n reaches its set point $\hat{n} = 5500$ and then deactivated/activated several times up to around 12:30 pm. On the other hand, the FPC strategy is activated around 9:30 am due to reservoir 2 and 15 min later for reservoirs 3 and 1, as accumulation n_i reaches its set point $\hat{n}_i$ for each reservoir i, albeit at different time. Note that in Scenario 1, control in reservoir 3 is firstly activated. Thus, the FPC strategy will automatically identify when and where to activate the controllers and this depends on the traffic conditions, even if demand profile is unknown.

Tables 2 and 3 display the obtained results for the two OD scenarios and all considered strategies.[5] It can be seen that the FPC strategy leads to an improvement of the evaluation criteria compared to NC and BBC for the whole network. More

---

[5] Given that FPC-LQ and FPC-LQI have similar behavior with insignificant changes in all performance measures (difference less than 2%), we present only the average improvement in the tables from all simulation runs.

Table 2
PIs of the FPC strategy vs. NC for the adaptive OD Scenario 2.

Evaluation criteria	No control	FPC	Av. improvement (%)	Units
Delay Time	691	385	44	sec/km
Number of Stops	27.47	18.02	34	#/veh/km
Total Distance Travelled (TDT)	95,801	101,840	6	km
Total Travel Time (TTT)	74.21	49.11	34	sec $\times 10^6$
Space-mean Speed = TDT/TTT	4.65	7.47	61	km/h
Mean Virtual Queue Length	4717	4530	4	vehs
Vehicles Waiting Out	9835	2194	78	vehs
Total Output	94,744	105,995	12	vehs

Table 3
PIs of the FPC strategy vs. BBC for the adaptive OD Scenario 3.

Evaluation criteria	BBC	FPC	Av. improvement (%)	Units
DelayTime	352	308	12	sec/km
Number of Stops	17.06	15.29	10	#/veh/km
Total Distance Travelled (TDT)	105,884	103,222	3	km
Total Travel Time (TTT)	47.90	43.13	10	sec $\times 10^6$
Space-mean Speed = TDT/TTT	7.96	8.62	8	km/h
Mean Virtual Queue Length	2895	1890	35	vehs
Vehicles Waiting Out	546	57	90	vehs
Total Output	108,440	108,765	0	vehs

specifically, when FPC is applied, TTT, Delay, and Space-mean Speed are improved in average by 34% (12%), 44% (10%), and 61% (8%), respectively, compared to NC (BBC). Regarding the Mean Virtual Queue Length, it can be seen that FPC creates temporary queues at the perimeter of the network (β_{ii} controllers) but these queues are lower than NC and BBC due to activation of the DTA and the oscillatory behavior of the BBC. However, these temporary queues in FPC and BBC is proved beneficial for the total network throughput.[6] Thus, even if vehicles are restricted in the perimeter of the network, they are able to reach their destinations faster than in the no control case ("slower is faster" effect, see Helbing and Mazloumian, 2009). Finally, the significantly lower number of vehicles within the network links (Vehicles Waiting Out) at the end of simulation indicates that the FPC control action does not create queues that spill back to upstream intersections at the boundary of neighborhood reservoirs. These results demonstrate the superiority of multi-reservoir feedback perimeter control (FPC) over single-reservoir perimeter control (BBC) in heterogeneously loaded networks (cf. Fig. 3 and the analysis in Section 4.2).

6. Discussion

In this paper, we addressed the problem of perimeter control for congested networks partitioned in reservoirs. This can be of great importance towards the development of generic, elegant, and efficient perimeter control strategies that appropriately account for the spatial and temporal heterogeneity of congestion between the reservoirs. First, by exploiting the properties of the MFD, we described the dynamics of the rush hour in case of multi-reservoir networks that are not uniformly congested. Motivated by the need to distribute the accumulation of vehicles in each reservoir as homogeneously as possible and maintain the rate of vehicles that are allowed to enter each reservoir around a desired point while the system's throughput is maximized, we then stated our control objective. In order to provide solutions that can be implemented in real time, we introduced two control strategies for determining the perimeter and boundary controllers, namely multivariable feedback regulator and integral feedback regulator. A key advantage of our approach is that it does not require high computational effort and future demand data if the state can be observed.

The impact of the perimeter control actions for each reservoir and the whole network was demonstrated by use of the corresponding MFDs and performance indices in a simulation application for a congested downtown area. The proposed strategy was demonstrated to preserve high network performance and equity in the heterogeneous test network and significantly reduce the hysteresis loops in the MFD.

These findings are of great importance for the traffic engineering community because the concept of an MFD (a) can be applied for heterogeneously loaded large-scale networks with multiple centers of congestion, if these networks can be partitioned into a small number of homogeneous regions, and (b) can be used towards the development of efficient perimeter and boundary flow control strategies.

---

[6] The effect of perimeter control might be even higher if outside areas are included in the simulation due to traffic assignment (somewhat adaptive drivers) that would distribute the gated flow (outside the external perimeter) more homogeneously along alternative routes.

Current extensions of this work are: (a) real-time estimation of the critical accumulation as a function of the level of congestion in each region and the distribution of destinations across the city, and (b) dynamic partitioning and control of heterogeneously congested networks. A better understanding of the role played by network topology in the spatiotemporal propagation of congestion and in the scatter and the hysteresis of the MFD using real data should also be a research priority. A field implementation is under preparation in Australia.

Appendix A. Linear control theory

A.1. Linearization of non-linear systems

Consider a non-linear, dynamic process described by the discrete-time state equation

$$\mathbf{x}(k+1) = \mathbf{f}[\mathbf{x}(k), \mathbf{u}(k), \mathbf{d}(k)] \tag{A.1}$$

and output equation

$$\mathbf{y}(k) = \mathbf{g}[\mathbf{x}(k)] \tag{A.2}$$

where $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{u} \in \mathbb{R}^m$, $\mathbf{d} \in \mathbb{R}^q$, and $\mathbf{y} \in \mathbb{R}^p$ are the state, control, disturbance, and output vectors, respectively; $\mathbf{f} \in \mathbb{R}^n$ and $\mathbf{g} \in \mathbb{R}^p$ are non-linear vector functions that describe the dynamic process and output, respectively. Assume existence of a desired steady-state $(\hat{\mathbf{x}}, \hat{\mathbf{u}}, \hat{\mathbf{d}}, \hat{\mathbf{y}})$. Linearization of (A.1), (A.2) around the desired steady-state yields

$$\Delta\mathbf{x}(k+1) = \mathbf{A}\Delta\mathbf{x}(k) + \mathbf{B}\Delta\mathbf{u}(k) + \mathbf{H}\Delta\mathbf{d}(k) \tag{A.3}$$
$$\Delta\mathbf{y}(k) = \mathbf{Y}\Delta\mathbf{x}(k) \tag{A.4}$$

where $\Delta\mathbf{x} = \mathbf{x} - \hat{\mathbf{x}}$, $\Delta\mathbf{u} = \mathbf{u} - \hat{\mathbf{u}}$, $\Delta\mathbf{d} = \mathbf{d} - \hat{\mathbf{d}}$ and $\Delta\mathbf{y} = \mathbf{y} - \hat{\mathbf{y}}$ are the linearized state, control, disturbance and output vectors; $\mathbf{A} = \partial\mathbf{f}/\partial\mathbf{x}|_{\hat{\mathbf{x}}}$, $\mathbf{B} = \partial\mathbf{f}/\partial\mathbf{u}|_{\hat{\mathbf{u}}}$, $\mathbf{H} = \partial\mathbf{f}/\partial\mathbf{d}|_{\hat{\mathbf{d}}}$ and $\mathbf{Y} = \partial\mathbf{g}/\partial\mathbf{x}|_{\hat{\mathbf{x}}}$ are the state, control, disturbance and output matrices, respectively. We assume that $\Delta\mathbf{d}(k) = \mathbf{0}$ and $[\mathbf{A}, \mathbf{B}]$ is completely controllable. Moreover, we augment the original state Eq. (A.1) by the integral equation

$$\mathbf{z}(k+1) = \mathbf{z}(k) - \hat{\mathbf{y}} + \mathbf{g}[\mathbf{x}(k)] \tag{A.5}$$

where $\mathbf{z} \in \mathbb{R}^p$ are additional state variables. Linearization of (A.5) with $\hat{\mathbf{z}} = \mathbf{0}$ yields

$$\mathbf{z}(k+1) = \mathbf{z}(k) + \mathbf{Y}\Delta\mathbf{x}(k) \tag{A.6}$$

The difference Eq. (A.6) computes the integral of the output signal (error), which can be used as a feedback term to provide zero steady-state error (see Appendix A.3). Finally, we assume that the matrix $\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{Y} & \mathbf{0} \end{bmatrix}$ has range $n + p$. A necessary condition for this assumption to hold is that the number of output variables should be less than the number of control variables, i.e. $p \leqslant m$.

A.2. Linear-Quadratic (LQ) control

Let us consider the quadratic cost criterion

$$\mathcal{L}(\mathbf{u}) = \frac{1}{2}\sum_{k=0}^{\infty}\left(\|\Delta\mathbf{x}(k)\|_{\mathbf{Q}}^2 + \|\Delta\mathbf{u}(k)\|_{\mathbf{R}}^2\right) \tag{A.7}$$

where $\mathbf{Q}$ and $\mathbf{R}$ are diagonal weighting matrices that are positive semi-definite and positive definite, respectively. Minimization of (A.7) subject to (A.3) ($\Delta\mathbf{d}(k) = \mathbf{0}$) leads to the feedback law (Papageorgiou, 1996)

$$\mathbf{u}(k) = \hat{\mathbf{u}} - \mathbf{K}[\mathbf{x}(k) - \hat{\mathbf{x}}] \tag{A.8}$$

where the gain matrix $\mathbf{K} \in \mathbb{R}^{m \times n}$ is calculated (for given matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{Q}$, and $\mathbf{R}$) from

$$\mathbf{K}(k) = [\mathbf{B}^\tau\mathbf{P}(k+1)\mathbf{B} + \mathbf{R}]^{-1}\mathbf{B}^\tau\mathbf{P}(k+1)\mathbf{A} \tag{A.9}$$

and the matrix $\mathbf{P} \in \mathbb{R}^{n \times n}$ is the solution of the Riccati difference equation

$$\mathbf{P}(k) = \mathbf{A}^\tau\mathbf{P}(k+1)\mathbf{A} - \mathbf{K}^\tau(k)\mathbf{B}^\tau\mathbf{P}(k+1)\mathbf{A} + \mathbf{Q} \tag{A.10}$$

with the terminal condition $\mathbf{P}(K') = \mathbf{I}$, where K' is the optimization time horizon. Starting from this terminal condition, (A.9) and (A.10) may be executed backwards in time to obtain $\mathbf{K}(k)$, $k = K' - 1, K' - 2, \ldots, 0$. The gain matrix $\mathbf{K}(k)$, resulting from the solution of (A.9) and (A.10), is generally time-variant. However, if the time horizon K' is sufficient long $\mathbf{K}(k)$ converges towards a time-invariant gain matrix $\mathbf{K}$ to be used in (A.8) (see Papageorgiou (1996) or Åström and Wittenmark (1996) for more details).

A.3. Linear-Quadratic Integral (LQI) control

Let us now consider the augmented quadratic cost criterion

$$\mathcal{L}(\mathbf{u}) = \frac{1}{2}\sum_{k=0}^{\infty}\left(\|\Delta\mathbf{x}(k)\|_{\mathbf{Q}}^2 + \|\Delta\mathbf{u}(k)\|_{\mathbf{R}}^2 + \|\mathbf{z}(k)\|_{\mathbf{S}}^2\right) \tag{A.11}$$

where $\mathbf{S}$ is an additional positive semi-definite diagonal weighting matrix. Considering the augmented cost criterion (A.11) and discrete-time system (A.3), (A.6) we obtain the following augmented state, control, disturbance, and weighting matrices

$$\widetilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{Y} & \mathbf{I} \end{bmatrix} \quad \widetilde{\mathbf{B}} = \begin{bmatrix} \mathbf{B} \\ \mathbf{0} \end{bmatrix} \quad \widetilde{\mathbf{H}} = \begin{bmatrix} \mathbf{H} \\ \mathbf{0} \end{bmatrix} \quad \widetilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{S} \end{bmatrix} \quad \widetilde{\mathbf{R}} = \mathbf{R} \tag{A.12}$$

Minimization of (A.11) subject to (A.3), (A.6) (assuming $\Delta\mathbf{d}(k) = \mathbf{0}$) leads to the feedback law (Papageorgiou, 1996)

$$\Delta\mathbf{u}(k) = -\widetilde{\mathbf{K}}\begin{bmatrix} \Delta\mathbf{x}(k) \\ \mathbf{z}(k) \end{bmatrix} \tag{A.13}$$

where $\widetilde{\mathbf{K}} \in \mathbb{R}^{m\times(n+p)}$ is the steady-state solution of the corresponding Riccati equation. Decomposing $\widetilde{\mathbf{K}} = [\mathbf{K}_1 \quad \mathbf{K}_2]$ the following is obtained from (A.13)

$$\mathbf{u}(k) = \hat{\mathbf{u}} - \mathbf{K}_1\Delta\mathbf{x}(k) - \mathbf{K}_2\mathbf{z}(k) \tag{A.14}$$

Subtracting (A.14) at $k-1$ from (A.14) at k and considering (A.6), we get after some algebra the final integral feedback law

$$\mathbf{u}(k) = \mathbf{u}(k-1) - \mathbf{K}_p[\mathbf{x}(k) - \mathbf{x}(k-1)] - \mathbf{K}_I[\mathbf{x}(k) - \hat{\mathbf{x}}] \tag{A.15}$$

where $\mathbf{K}_p = \mathbf{K}_1 - \mathbf{K}_2\mathbf{Y}$ and $\mathbf{K}_I = \mathbf{K}_2\mathbf{Y}$ are the proportional and integral gains, respectively. The time-invariant gain matrix $\widetilde{\mathbf{K}}$, which depends only upon the augmented matrices $\widetilde{\mathbf{A}}$, $\widetilde{\mathbf{B}}$, $\widetilde{\mathbf{Q}}$, and $\widetilde{\mathbf{R}}$ may be calculated analogously to $\mathbf{K}$ in Appendix A.2 (see (A.9) and (A.10)) through the backward integration of the augmented Riccati matrix $\widetilde{\mathbf{P}}(k)$ starting from the terminal condition $\widetilde{\mathbf{P}}(K') = \mathbf{I}$ until convergence.

References

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

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. Transportation Research Part C 18 (5), 680–694.

Åström, K., Wittenmark, B., 1996. Computer-Controlled Systems: Theory and Design, Prentice Hall Information and System Sciences Series third ed. Prentice Hall, Englewood Cliffs, NJ.

Ban, X.J., Hao, P., Sun, Z., 2011. Real time queue length estimation for signalized intersections using travel times from mobile sensors. Transportation Research Part C 19 (6), 1133–1156.

Bretherton, D., Bowen, G., Wood, K., 2003. Effective urban traffic management and control: recent developments in SCOOT. In: Proceedings of the 82nd Annual Meeting of the Transportation Research Board, Washington, DC.

Buisson, C., Ladier, C., 2009. Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. Transportation Research Record 2124, 127–136.

Cascetta, E., Nuzzolo, A., Russo, F., Vitetta, A., 1996. A modified logit route choice model overcoming path overlapping problems: specification and some calibration results for interurban networks. In: Lesort, J. (Ed.), Proceedings from the Thirteenth International Symposium on Transportation and Traffic Theory. Pergamon, Lyon, France, pp. 697–711.

Courbon, T., Leclercq, L., 2011. Cross-comparison of macroscopic fundamental diagram estimation methods. Procedia-Social and Behavioral Sciences 20 (0), 417–426.

Daganzo, C.F., 2007. Urban gridlock: macroscopic modeling and mitigation approaches. Transportation Research Part B 41 (1), 49–62.

Daganzo, C.F., Gayah, V.V., Gonzales, E.J., 2011. Macroscopic relations of urban traffic variables: bifurcations, multivaluedness and instability. Transportation Research Part B 45 (1), 278–288.

Daganzo, C.F., Geroliminis, N., 2008. An analytical approximation for the macroscopic fundamental diagram of urban traffic. Transportation Research Part B 42 (9), 771–781.

Diakaki, C., Dinopoulou, V., Aboudolas, K., Papageorgiou, M., Ben-Shabat, E., Seider, E., Leibov, A., 2003. Extensions and new applications of the traffic-responsive urban control strategy: coordinated signal control for urban networks. Transportation Research Record 1856, 202–211.

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

Gartner, N.H., Pooran, F.J., Andrews, C.M., 2001. Implementation of the OPAC adaptive control strategy in a traffic signal network. In: Proceedings of the 2001 IEEE Intelligent Transportation Systems Conference, Oakland, CA, USA.

Gayah, V.V., Daganzo, C.F., 2011a. Clockwise hysteresis loops in the macroscopic fundamental diagram: an effect of network instability. Transportation Research Part B 45 (4), 643–655.

Gayah, V.V., Daganzo, C.F., 2011b. Exploring the effect of turning maneuvers and route choice on a simple network. Transportation Research Record 2249, 15–19.

Geroliminis, N., Boyaci, B., 2012. The effect of variability of urban systems characteristics in the network capacity. Transportation Research Part B 46 (10), 1607–1623.

Geroliminis, N., Daganzo, C.F., 2008. Existence of urban-scale macroscopic fundamental diagrams: some experimental findings. Transportation Research Part B 42 (9), 759–770.

Geroliminis, N., Haddad, J., Ramezani, M., 2013. Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: a model predictive approach. IEEE Transactions on Intelligent Transportation Systems 14 (1), 348–359.

Geroliminis, N., Levinson, D.M., 2009. Cordon pricing consistent with the physics of overcrowding. In: Lam, W.H.K., Wong, S.C., Lo, H.K. (Eds.), Transportation and Traffic Theory 2009. Springer-Verlag, US, pp. 219–240.

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

Geroliminis, N., Sun, J., 2011. Properties of a well-defined macroscopic fundamental diagram for urban traffic. Transportation Research Part B 45, 605–617.

Godfrey, J.W., 1969. The mechanism of a road network. Traffic Engineering and Control 11 (7), 323–327.

Haddad, J., Geroliminis, N., Ramezani, M., 2013. Cooperative traffic control of a mixed network with two urban regions and a freeway. Transportation Research Part B 54, 17–36.

Helbing, D., Mazloumian, A., 2009. Operation regimes and slower-is-faster effect in the control of traffic intersections. The European Physical Journal B 70 (2), 257–274.

Hunt, P.B., Robertson, D.I., Bretherton, R.D., Royle, M.C., 1982. The SCOOT on-line traffic signal optimization technique. Traffic Engineering and Control 23, 190–192.

Ji, Y., Daamen, W., Hoogendoorn, S., Hoogendoorn-Lanser, S., Qian, X., 2010. Macroscopic fundamental diagram: investigating its shape using simulation data. Transportation Research Record 2161, 42–48.

Ji, Y., Geroliminis, N., 2012. On the spatial partitioning of urban transportation networks. Transportation Research Part B 46 (10), 1639–1656.

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 46 (10), 1393–1403.

Knoop, V.L., Hoogendoorn, S.P., Van Lint, J.W.C., 2012. Routing strategies based on the macroscopic fundamental diagram. Transportation Research Record 2315, 1–10.

Kouvelas, A., Aboudolas, K., Papageorgiou, M., Kosmatopoulos, E., 2011. A hybrid strategy for real-time traffic signal control of urban road networks. IEEE Transactions on Intelligent Transportation Systems 12 (3), 884–894.

Lowrie, P.R., 1982. SCATS: the Sydney co-ordinated adaptive traffic system – Principles, methodology, algorithms. In: Proceedings of the IEE International Conference on Road Traffic Signalling. London, England, pp. 67–70.

Luk, J., Green, D., 2010. Balancing traffic density in a signalized network. Austroads Research Report AP-R369/10.

Mahmassani, H.S., Hou, T., Saberi, M., 2013a. Connecting network-wide travel time reliability and the network fundamental diagram of traffic flow. In: Proceedings of the 92nd Annual Meeting of the Transportation Research Board, Washington, DC.

Mahmassani, H.S., Saberi, M., Ali Zockaie K., 2013. Urban Network Gridlock: Theory, Characteristics, and Dynamics. Procedia - Social and Behavioral Sciences 80, 79–98.

Mazloumian, A., Geroliminis, N., Helbing, D., 2010. The spatial variability of vehicle densities as determinant of urban network capacity. Philosophical Transactions of the Royal Society A 368 (1928), 4627–4647.

Mirchandani, P., Wang, F.-Y., 2005. RHODES to intelligent transportation systems. IEEE Intelligent Systems 20 (1), 10–15.

Papageorgiou, M., 1996. Optimierung, 2nd ed. Oldenbourg, Munich, Germany.

Papageorgiou, M., Blosseville, J.-M., Hadj-Salem, H., 1990. Modelling and real-time control of traffic flow on the southern part of Boulevard Périphérique in Paris Part II: Coordinated on-ramp metering. Transportation Research Part A 24 (5), 361–370.

Saberi, M., Mahmassani, H.S., 2012. Exploring the properties of network-wide flow-density relations in a freeway network. Transportation Research Record 2315, 153–163.

Wu, X., Liu, H.X., Geroliminis, N., 2011. An empirical analysis on the arterial fundamental diagram. Transportation Research Part B 45 (1), 255–266.

Zhang, L., Garoni, T.M., de Gier, J., 2013. A comparative study of macroscopic fundamental diagrams of arterial road networks governed by adaptive traffic signal systems. Transportation Research Part B 49 (0), 1–23.