Transportation Research Part B 000 (2017) 1–22

Contents lists available at ScienceDirect

Transportation Research Part B

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

Macroscopic modelling and robust control of bi-modal multi-region urban road networks

Konstantinos Ampountolas[a,*], Nan Zheng[b,c], Nikolas Geroliminis[c]

[a] School of Engineering, University of Glasgow, G12 8QQ Glasgow, United Kingdom
[b] School of Transportation Science and Engineering, Beihang Univeristy, 100191 Beijing, China
[c] Urban Transport Systems Laboratory, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 2 September 2015
Revised 29 April 2017
Accepted 18 May 2017
Available online xxx

Keywords:
Macroscopic fundamental diagram (MFD)
Heterogeneous urban road networks
Perimeter and boundary flow control
Robust control
Convex optimisation

ABSTRACT

The paper concerns the integration of a bi-modal Macroscopic Fundamental Diagram (MFD) modelling for mixed traffic in a robust control framework for congested single- and multi-region urban networks. The bi-modal MFD relates the accumulation of cars and buses and the outflow (or circulating flow) in homogeneous (both in the spatial distribution of congestion and the spatial mode mixture) bi-modal traffic networks. We introduce the composition of traffic in the network as a parameter that affects the shape of the bi-modal MFD. A linear parameter varying model with uncertain parameter the vehicle composition approximates the original nonlinear system of aggregated dynamics when it is near the equilibrium point for single- and multi-region cities governed by bi-modal MFDs. This model aims at designing a robust perimeter and boundary flow controller for single- and multi-region networks that guarantees robust regulation and stability, and thus smooth and efficient operations, given that vehicle composition is a slow time-varying parameter. The control gain of the robust controller is calculated off-line using convex optimisation. To evaluate the proposed scheme, an extensive simulation-based study for single- and multi-region networks is carried out. To this end, the heterogeneous network of San Francisco where buses and cars share the same infrastructure is partitioned into two homogeneous regions with different modes of composition. The proposed robust control is compared with an optimised pre-timed signal plan and a single-region perimeter control strategy. Results show that the proposed robust control can significantly: (i) reduce the overall congestion in the network; (ii) improve the traffic performance of buses in terms of travel delays and schedule reliability, and; (iii) avoid queues and gridlocks on critical paths of the network.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Urban transport systems consist of multiple modes sharing and competing for the same road space including pedestrians, bicycles, cars, taxis, delivery vehicles and modes carrying more passengers, such as buses and trams. Realistic modelling and control of multimodal transport systems remain an important challenge, due to limited understanding of the dynamic interactions of the modes at the network level.

* Corresponding author.
E-mail addresses: Konstantinos.Ampountolas@glasgow.ac.uk (K. Ampountolas), Nan_Zheng@buaa.edu.cn (N. Zheng), Nikolas.Geroliminis@epfl.ch (N. Geroliminis).

http://dx.doi.org/10.1016/j.trb.2017.05.007
0191-2615/© 2017 Elsevier Ltd. All rights reserved.

Please cite this article as: K. Ampountolas et al., Macroscopic modelling and robust control of bi-modal multi-region urban road networks, Transportation Research Part B (2017), http://dx.doi.org/10.1016/j.trb.2017.05.007

The focus of this paper is on bi-modal urban networks consisting of private cars and public transport. We develop a robust perimeter and boundary flow control approach that adapts the traffic signal settings for mixed traffic networks. The objective is to consider the interactions between the two modes through an aggregated network level approach, as described by a bi-modal network macroscopic fundamental diagram. This is significantly different than most existing approaches for preferential treatment of buses that consider local level decisions (e.g. transit signal priority strategies). We are able to show that even if a city does not have dedicated space for buses (e.g. due to the lack of road space or the high cost of implementation) and traffic signals do not provide special priority, the overall performance of network is significantly improved for both modes in terms of efficiency (total delays) and reliability (bus scheduling). This work builds on recent findings for multimodal network modelling. Our analysis shows that for various demand conditions of bi-modal traffic, adaptive traffic lights at the perimeter of the network (single-region) and/or the boundary of neighbour regions (multi-region), which have much smaller implementation cost, can improve traffic conditions for all modes in the network; while any infrastructure changes have to be associated with system interference and higher costs. To the best of our knowledge, this is the first effort to develop efficient control schemes for networks with multiple modes of transport with the use of bi-modal MFDs.

Some recent developments optimise in real-time traffic signals at the local level to provide bus priority with an objective to maximise passenger flows (Christofa et al., 2013; 2016; Hu et al., 2015). Other works have investigated how to redistribute urban space between different modes to maximise passenger flows for idealised networks (Zheng and Geroliminis, 2013). Optimisation of passenger flows with special treatment of more efficient modes has been analysed at the local level or for arterial streets (e.g. dedicated bus lanes), but at the network level is quite premature, especially in the dynamic case. An aggregated modelling for multi-modal systems following the concept of an MFD can be an alternative if it unveils similar properties as in the single-mode case of vehicular traffic.

With respect to networks with car being the main mode of travel, it has been reported in many publications from empirical and simulated data for many networks that by spatially aggregating the highly scattered plots of flow versus density from individual links, the scatter almost disappeared and a well-defined MFD arises between space-mean flow and density (see e.g. Geroliminis and Daganzo, 2008; Buisson and Ladier, 2009; Ji et al., 2010; Mahmassani et al., 2013). The idea of an MFD with a critical accumulation belongs to Godfrey (1969) and reinitiated later in Mahmassani et al. (1987), Daganzo (2007), and other works. Despite these recent findings for the existence of MFDs in the dynamic case with low scatter, recent findings (see e.g. Geroliminis and Sun, 2011; Mazloumian et al., 2010; Gayah and Daganzo, 2011, and others) have identified the spatial distribution of vehicle density in the network as one of the important parameters that influence MFDs. The concept of an MFD can be applied for heterogeneous cities with multiple centres of congestion, if these cities can be partitioned into a small number of homogeneous clusters, as for example proposed in Ji and Geroliminis (2012) and Saeedmanesh and Geroliminis (2016). For a comprehensive review of MFD modelling and control for single-mode systems the reader could refer to Haddad et al. (2013) or Saberi et al. (2014).

Latest works extend the MFD from single-mode application to bi-modal, where cars and buses share the same infrastructure and look at passenger flow dynamics in addition to vehicular dynamics (Geroliminis et al., 2014; Chiabaut et al., 2014; Chiabaut, 2015). Gonzales and Daganzo (2012; 2013) provide an analysis of the morning and evening commute for aggregated dynamics for bi-modal networks that gives interesting insights for the interactions of the modes under congested and system optimum conditions. The existence of a mixed traffic, bi-modal (three-dimensional) MFD is investigated in Geroliminis et al. (2014), via micro-simulation studies for a large network with dynamic features. Remarkably, the extended versions of the MFD relate the accumulation (equivalent to traffic density) of cars and buses to vehicular and passenger flows, reflecting congestion dynamics and provide inspiration for investigating MFD control strategies for these networks. Empirical observations besides simulation studies are under current development. This is described later in more details.

The MFD concept can contribute to the development of simple perimeter flow control schemes to maximise throughput in single-region homogeneous networks (Daganzo, 2007; Keyvan-Ekbatani et al., 2012; Haddad and Shraiber, 2014) and multi-region heterogeneous networks (Geroliminis et al., 2013; Aboudolas and Geroliminis, 2013; Ramezani et al., 2015; Keyvan-Ekbatani et al., 2015; Kouvelas et al., 2017; Haddad, 2017; Haddad and Mirkin, 2017). The main idea of a perimeter control policy is to “meter” the input flow to the system and hold vehicles outside the controlled area if necessary, so that to avoid states in the congested regime of the MFD. A key advantage of this approach is that it does not require high computational effort if proxies of the critical accumulation are available and the current state of the network can be observed in real-time (e.g. with detector data). Stability analysis for the effect of demand and adaptive signal control in systems with MFD dynamics can be found in Aboudolas et al. (2010), Haddad and Geroliminis (2012), Gayah and Daganzo (2011), Gayah et al. (2014) and Zhang et al. (2013).

In this work, we deal with the perimeter flow control problem in single- and two-region bi-modal urban networks with the use of a bi-modal MFD, which is recently introduced by the authors in Geroliminis et al. (2014). When demand conditions are light the influence of the public transport stops to pick up and alight passengers to car traffic is quite small, but for severe congestion and high frequency in time and space of these stops, the network capacity can be influenced and interactions should be integrated. The bi-modal MFD relates the accumulation of cars and buses with the total circulating flow in the bimodal network. It was observed via simulation studies that the impact of each mode on the traffic flow in the network is different, i.e., each bus cannot be considered as equivalent to some number of passenger cars. In particular, cars are usually faster than buses (because of the bus stops) but if the percentage of buses in the overall accumulation is high, then the average speed of vehicles certainly differs from the one sustainable without the interference of buses. Thus

Please cite this article as: K. Ampountolas et al., Macroscopic modelling and robust control of bi-modal multi-region urban road networks, Transportation Research Part B (2017), http://dx.doi.org/10.1016/j.trb.2017.05.007

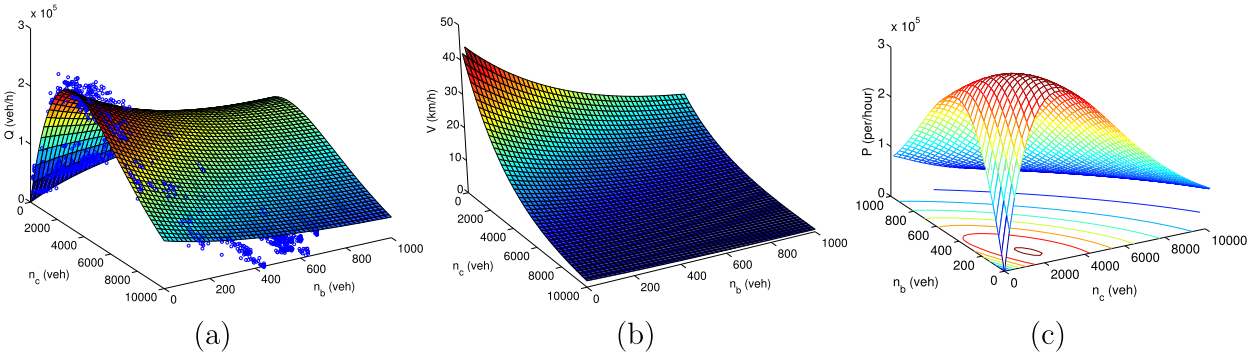

Fig. 1. (a) Approximated bi-modal vehicle MFD relating accumulation of cars and buses with output (circulating flow); (b) Approximated bi-modal vehicle MFD relating accumulation of cars and buses with space-mean speed; (c) Approximated bi-modal passenger MFD relating accumulation of cars and buses with passenger flow. (Source of bi-modal vehicle and passenger MFDs from Geroliminis et al., 2014.)

the maximum throughput (capacity) varies with the composition of traffic in the network. For this reason, we introduce the composition of traffic in the network as an uncertain parameter that affects the shape of the bi-modal MFD. Empirical studies should be a research priority to further investigate the conditions under which a bi-modal MFD holds for different types of networks.

While for the single-region problem the composition of traffic is assumed uniformly distributed over the network, and a single perimeter flow controller in the external boundary of the city is applied, we also notice that the centre of the network contains more buses than the periphery. Thus, in the two-region case, a modified partitioning approach (based on Ji and Geroliminis (2012)) is applied to cluster the network into two regions with different mode composition. A robust control approach is developed for both cases. To this end, the bi-modal MFD is used to describe the aggregated traffic dynamics in the bi-modal urban network. A linear system subject to uncertain time-varying vehicle composition is used as a basis for designing a Proportional–Integral (PI) robust perimeter flow controller. The control gain of the proposed scheme is calculated off-line using Linear Matrix Inequalities (LMIs) and Semi-Definite Programming (SDP). To evaluate the proposed scheme, a simulation-based comparison of single- and two-region robust perimeter flow control with an optimised pre-timed signal plan and a previously proposed single-region perimeter flow control strategy (of bang-bang type similar to the policy introduced by Daganzo (2007)) for an area of the Downtown San Francisco is carried out. Section 2 of the paper presents the methodological framework of the bi-modal MFD. Section 3 develops the dynamics of multi-region bi-modal urban networks govern by bi-modal MFDs to facilitate the design of the proposed robust control approach, which is described in Section 4. Section 5 presents the results and insights of the single- and two-region robust control via micro-simulation, while discussion follows in Section 6.

2. Macroscopic modelling of bi-modal multi-region urban networks

Consider a heterogeneous bi-modal urban network, where the traffic flow comprises two vehicle classes, i.e., passenger cars and buses. We assume that the heterogeneous network can be partitioned into N regions with homogeneous bi-modal traffic (Geroliminis et al., 2014). Let $n_i(t) = n_{i,c}(t) + n_{i,b}(t)$ be the accumulation of mixed traffic in region $i = 1, \ldots, N$, where $n_{i,c}(t)$ and $n_{i,b}(t)$ are the accumulation of cars and buses in each region i at time t, respectively. We assume that each region i has an extensive network of public transport lines with a varying range of service frequencies. The number of public lines and the service frequency can determine the composition of traffic $\delta_i(t) \in (0, 1)$ in each region i at time t, where $n_{i,c}(t) = \delta_i(t) n_i(t)$ and $n_{i,b}(t) = (1 - \delta_i(t)) n_i(t)$, $0 < \delta_i < 1$, $i = 1, \ldots, N$. The composition of traffic is assumed slowly time-varying. Let Q_i be the regional circulating flow (in vehicles per unit time per lane) as can be estimated with Edie's (1963) generalised definition of flow, i.e., weighted average of link flows with link lengths. If we further assume that the trip length in each region is constant and given by L_i, and l_i is the total network length (in lane-km), then the output and space-mean speed of the region i can be expressed in the steady-state as $O_i = Q_i l_i / L_i$ and $V_i = Q_i l_i / n_i$, respectively. The output O_i consists of the trips finishing within a region plus the trips that transfer to neighbour regions through the boundaries. In the case of mixed bi-modal traffic, the output flow is not only a function of the accumulation $n_i(t)$, but also of the composition of traffic $\delta_i(t)$ in each region i, i.e., $O_i = O(n_i(t), \delta_i(t))$, $i = 1, \ldots, N$. Geroliminis et al. (2014) considered the number of passengers occupying of each mode $m \in \{c, b\}$ and also obtained a passenger bi-modal MFD relating accumulation of cars and buses with passenger flow P.

Fig. 1(a)–(c) depict the shape of Q, V, and P (subscript i is omitted) in function of n_c and n_b, respectively, for a 2.5 miles square area of Downtown San Francisco shown in Fig. 3(a). Fig. 4(a) illustrates some representative public transport lines out of a total of 29 bus lines crossing the network modelled in micro-simulation. The shape of Q and P is the best-fit curve from an exponential family, see Geroliminis et al. (2014) for details. Fig. 1(a) confirms the existence of an MFD like-shape

Please cite this article as: K. Ampountolas et al., Macroscopic modelling and robust control of bi-modal multi-region urban road networks, Transportation Research Part B (2017), http://dx.doi.org/10.1016/j.trb.2017.05.007

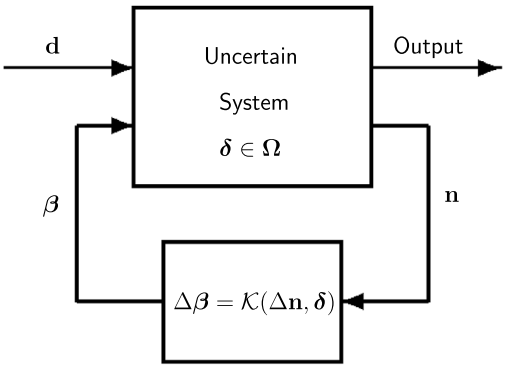

Fig. 2. Robust feedback control design block diagram.

for bi-modal mixed traffic, which shape depends on the accumulation of both cars and buses. As can be seen the circulating flow Q increases for small n_c and n_b and then decreases monotonically for values of n_c and n_b larger than some critical values, albeit with different slopes. Remarkably the slope of buses is higher that the slope of cars. This indicates that the effect of an additional bus in the network is higher than an additional car. A simple explanation (among others) is that buses make additional (to traffic) stops for passenger alighting and boarding, and thus negatively influencing traffic. Thus the maximum throughput varies with the composition of traffic δ in the network, which is a function of n_c and n_b. In other words, the composition of traffic in the network affects the shape of the bi-modal MFD.

Fig. 1(b) shows the derivation of network space-mean speed as a function of bus and car accumulations from the exponential function in Fig. 1(a). Note that speed is a convex function both with n_b and n_c. In particular, cars are usually faster than buses (because of bus stops) but if the percentage of buses in the overall accumulation is high, then the average speed of vehicles certainly differs from the one sustainable without the interference of buses. Note also that a non-linear shape for small accumulations is followed by an almost 2D planar shape. For the purposes of control we consider a linear approximation of speed. While this approximation might create some errors for the uncongested regime, it is close to exact around the critical values that maximise flows, which is the region that the controller will be designed.

Finally from Fig. 1(c) it can be seen that the shape of the bi-modal passenger MFD is significantly different from the one observed in Fig. 1(a) (bi-modal vehicle MFD). More precisely, for a given n_c passenger flow P first monotonically increases as n_b increases to a critical $\hat{n}_b$ and then decreases for $n_b > \hat{n}_b$. Thus, having buses in the network will significantly increase the efficiency of the system, but overloading the network with buses will eventually cause delays for all vehicles and reduce passenger throughput. Comparing Fig. 1(c) with Fig. 1(a) an interesting observation is that the passenger MFD maximises passenger throughput P at a non-zero accumulation of buses (observe the contour plot in Fig. 1(c)), while the vehicle MFD maximises vehicle flow for $\delta \approx 1$ (zero accumulation of buses). We will integrate this important observation later when designing our controller in Section 3. The dynamics of the bi-modal system will be presented there as well. The shape of Q, V, and P can be captured by different functions, e.g. quadratic or exponential. Analytical formulas for the critical accumulation $\hat{n}$ as a function of the composition of traffic δ can be developed as in Appendix A. Such a functional form is required in the development of the control framework.

3. Dynamics of multi-region bi-modal urban networks

Given the existence of a bi-modal MFD $Q_i(n_i(t), \delta_i(t))$ for each homogeneous region $i = 1, \ldots, N$, bi-modal regional traffic could be treated macroscopically as a single-region dynamic system with total accumulation n_i as a state variable and δ_i as a single parameter that affects the shape of the bi-modal MFD. Furthermore, we assume that the composition rate δ_i belongs to a polytopic compact set $\Omega_i = \left\{\delta_i(t) \,|\, \delta_{i,\min} \le \delta_i(t) \le \delta_{i,\max}, t \ge 0\right\}$, $i = 1, \ldots, N$, which is state independent, where $\delta_{i,\min}$ and $\delta_{i,\max}$ are the minimum and maximum composition of traffic in each region i (see Appendix B for details on the polytopic uncertainty). The set Ω_i can be easily specified for a given network partitioned in N regions from the number of public transport lines in each region i and their operational frequency or it can be directly observed with real-time data (e.g., from GPS equipped vehicles and loop detectors).

Let $q_{i,\text{in}}(t)$ and $q_{i,\text{out}}(t)$ be the inflow and outflow in region i at time t, respectively; $\mathcal{S}_i$ be a set with elements the origin regions whose outflow will go to destination region i. Also, let $d_i(t)$ be the uncontrolled traffic demand (disturbances) in region i at time t. Note that $d_i(t)$ includes both internal and external non-controlled inflows. The conservation equation for each neighbour region $i = 1, \ldots, N$ reads:

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

Please cite this article as: K. Ampountolas et al., Macroscopic modelling and robust control of bi-modal multi-region urban road networks, Transportation Research Part B (2017), http://dx.doi.org/10.1016/j.trb.2017.05.007

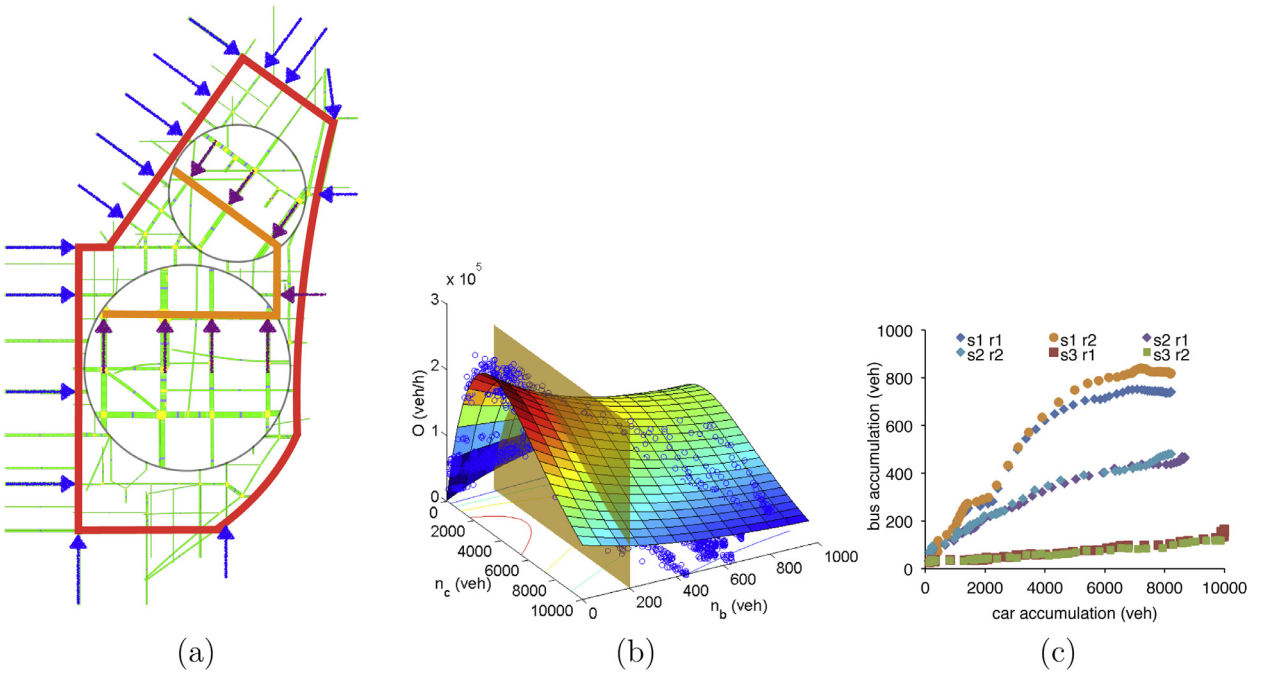

Fig. 3. (a) Snapshot of Downtown San Francisco; red colour indicates the external protected network area where single-region control is applied to 15 signalised junctions illustrated with blue arrows; orange colour indicates the internal protected network area where boundary control is applied to 8 signalised junctions illustrated with purple arrows; (b) The bi-modal MFD of Downtown San Francisco relating accumulation of cars and buses with circulating flow; the cross-section of the 3D surface for a constant accumulation of buses $n_b = 200$ veh demonstrating the typical dependence of flow with the composition of traffic in the network; (c) The relation of n_c and n_b (composition of traffic δ) for three different scenarios, each scenario ran for two replications. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Since the state $n_i(t)$ and composition of bi-modal traffic $\delta_i(t)$ evolve slowly with time t, we may assume that the flow $q_{i,\text{out}}(t)$ is given by the output $Q_i(n_i(t), \delta_i(t))$ (the bi-modal MFD), which is a function of the total accumulation $n_i(t)$ and composition $\delta_i(t)$. We assume that Q_i consists of the trips finishing within a region plus the trips that transfer to neighbour regions through the boundaries. The inflow to region $i = 1, \ldots, N$ is given by

$$q_{i,\text{in}}(t) = \beta_i(t) + \sum_{j \in \mathcal{S}_i} \gamma_{ji} Q_j(n_j(t), \delta_j(t)) \tag{2}$$

where $\beta_i(t)$ is the input flow to region i at time t (from the perimeter and neighbour regions), to be calculated by the controller (decision variable). The second term in (2) is the uncontrolled input flow to region i from its neighbour regions. The parameters γ_{ji} represent the fraction (splitting rates) of inbound boundary flow rate to region i, $j \in \mathcal{S}_i$, which is assumed constant and known (transfer flows). The parameters γ_{ji} can be readily specified for a given multi-region network from sensors (e.g. loop detectors) along the boundary of neighbour regions. Errors associated with the estimation of γ_{ji} may be accommodated within the uncontrolled demand d_i in (1).

Introducing (2) in (1) we obtain the following nonlinear state equation for each region i

$$\frac{dn_i(t)}{dt} = \beta_i(t) + \sum_{j \in \mathcal{S}_i} \gamma_{ji} Q_j(n_j(t), \delta_j(t)) - Q_i(n_i(t), \delta_i(t)) + d_i(t). \tag{3}$$

Both accumulation n_i and composition of traffic δ_i can be observed in real-time. Vehicle accumulation can be obtained with different types of sensors while buses are currently equipped with GPS devices capable of providing locational data at any given time. Note that the third term Q_i in (3) represents the output of region i. This consists of trip endings within region i, plus transfer flows to neighbour regions j belong to $\mathcal{S}_i$. These transfer flows might be controlled by $\beta_j(t)$, as there are inputs for regions j. Nevertheless, as we do not decompose accumulations based on destination this value cannot be accurately estimated. We consider this a reasonable approximation and we expect that the controllers β_i will be able to keep the system at the desired set points. Integrating more complex dynamics will create non-linearities that cannot be easily treated with this methodology. Alternatively, one could apply more complex approaches but the state observation might be more challenging (see e.g. Zheng and Geroliminis, 2013). While future work could shed more light in the theoretical basis of such approximation, the results highlight that the controller is capable of keeping the state of the system at the desired levels.

Given the existence of a bi-modal MFD $Q_i(n_i(t), \delta_i)$ with an optimum accumulation $\hat{n}_i$ at which maximum flow is reached for different δ_i, the nonlinear model (3) may be linearised around some set point $(\hat{n}_i, \hat{\beta}_i, \hat{d}_i)$. The set points $\hat{n}_i$ may be

Please cite this article as: K. Ampountolas et al., Macroscopic modelling and robust control of bi-modal multi-region urban road networks, Transportation Research Part B (2017), http://dx.doi.org/10.1016/j.trb.2017.05.007

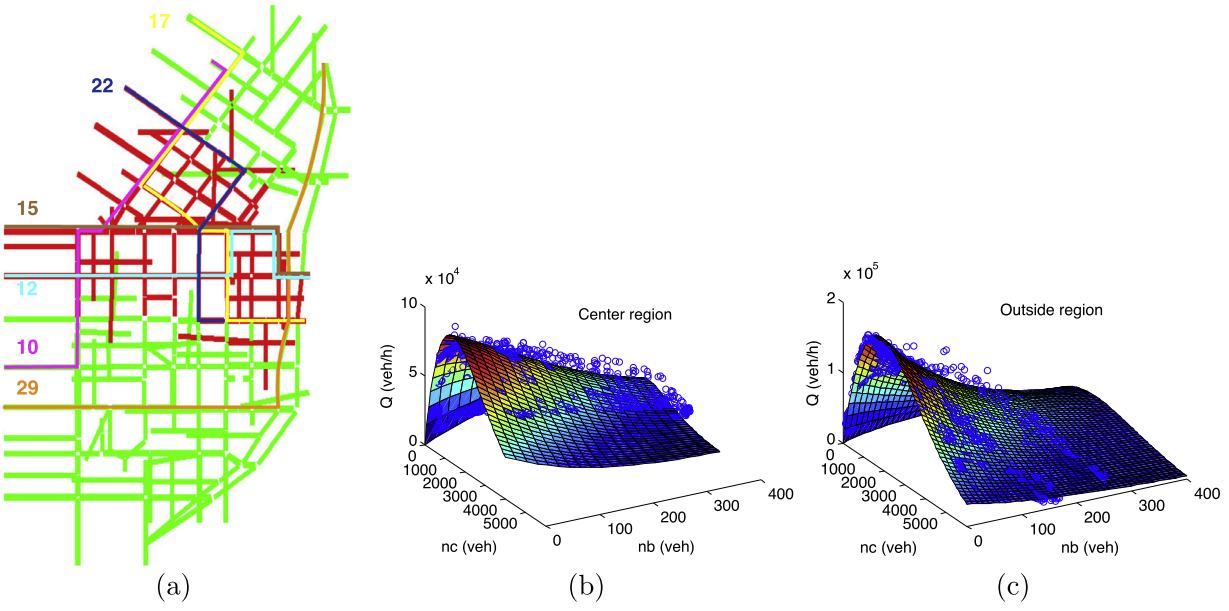

Fig. 4. (a) Snapshot of Downtown San Francisco and partitioning into two regions; red colour indicates the city centre; green colour indicates the rest of the network; numbers with different colours indicate major public transport lines in the network; (b, c) The bi-modal MFD of the centre and outside region relating accumulation of cars n_c and buses n_b with circulating flow, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

determined by fixing the composition rate δ_i as in analytical formulas (A.6) and (A.11), see Section Appendix A for details. The set points $\hat{\beta}_i$ can be derived from the inverse image of the corresponding bi-modal MFDs for given $\hat{n}_i$. Finally, the set-points $\hat{d}_i$ are usually determined via historical traffic data of a network. Denoting $\Delta x_i = x_i - \hat{x}_i$ analogously for all variables and considering fixed composition rate δ_i, the linearisation for each region $i = 1, \ldots, N$ yields (derivative of Q_i is taken with respect to n_i)[1]

$$\Delta \dot{n}_i(t) = \Delta \beta_i(t) + \sum_{j \in \mathcal{S}_i} \gamma_{ji} Q_j'(\hat{n}_j, \delta_j) \Delta n_j(t) - Q_i'(\hat{n}_i, \delta_i) \Delta n_i(t) + \Delta d_i(t). \tag{4}$$

Model (4) is Linear Parameter Varying (LPV) with uncertain parameters $\delta_i \in \Omega_i$. It approximates the original nonlinear system (3) when we are near the equilibrium point $(\hat{n}_i, \hat{\beta}_i, \hat{d}_i)$ about which the system was linearised. This LPV model will be used as a basis for robust control design in next section.

The basic dynamical system (4) can be extended to consider a broader class of inequality state and control constraints that are consistent with the physics of traffic. More precisely, input flow $\beta_i(t)$, $i = 1, \ldots, N$ is constrained as $\beta_{i,\min} \leq \beta_i(t) \leq \beta_{i,\max}$, where $\beta_{i,\min}$, $\beta_{i,\max}$ are the minimum and maximum permissible entrance flow of mixed traffic, respectively, and $\beta_{i,\min} > 0$ to avoid long queues and delays at the perimeter of the network. Moreover, the total accumulation $n_i(t)$, $i = 1, \ldots, N$ is constrained as $0 \leq n_i(t) \leq n_{i,\max}$, where $n_{i,\max}$ is the maximum accumulation in region $i = 1, \ldots, N$. It should be noted that our approach does not directly consider the control constraints (integrated indirectly in the implementation phase), while the state constraints are satisfied by appropriate selection of some weighting matrices in the design phase as explained in the next section.

4. Model uncertainty and robust control

The basic assumption in this work is that the bi-modal multi-region network has an extensive network of public transport lines with a varying range of service frequencies. Given any composition of traffic with rate $\delta_i \in \Omega_i$ in each region i, δ_i varies slowly with time, the bi-modal MFD in Fig. 1(a)[2] and the LPV system (4), we aim at designing a perimeter control strategy for the bi-modal network. This strategy minimises an upper bound of a worst-case cost criterion $\mathcal{L}(\mathbf{n}, \boldsymbol{\beta})$ given the uncertainty of the composition rate $\boldsymbol{\delta} \in \boldsymbol{\Omega}$ and an initial condition, where $\mathbf{n} \in \mathbb{R}^N, \boldsymbol{\beta} \in \mathbb{R}^N, \boldsymbol{\delta} \in \mathbb{R}^N$ are the state, control and parameter vectors, respectively (with elements n_i, β_i, δ_i, $i = 1, \ldots, N$); and $\boldsymbol{\Omega} = \cup_{i=1}^N \Omega_i$. In addressing this problem, there are several possibilities for generating the perimeter flow control input $\boldsymbol{\beta}(t)$. Fig. 2 depicts the typical block diagram for robust feedback control design. The system's output or performance is measured via suitable criteria $\mathcal{L}(\mathbf{n}, \boldsymbol{\beta})$, which would have a physical meaning such as total time spent or trip completion by all vehicles in the network over a time

[1] Notation: For functions $(\cdot)'$ indicates first-derivate. For matrices and vectors $(\cdot)^\mathsf{T}$ indicates transpose.
[2] A similar approach could also be pursued with the passenger MFD in Fig. 1(c).

Please cite this article as: K. Ampountolas et al., Macroscopic modelling and robust control of bi-modal multi-region urban road networks, Transportation Research Part B (2017), http://dx.doi.org/10.1016/j.trb.2017.05.007

horizon, throughput, etc. If both the state of the system $\mathbf{n}$ and uncertainty $\boldsymbol{\delta}$ are measurable in real-time, a state feedback control law $\Delta\boldsymbol{\beta}(t) = \mathcal{K}\big(\Delta\mathbf{n}(t), \boldsymbol{\delta}(t)\big)$, where $\mathcal{K}$ is a non-linear mapping, which is explicitly depends on the uncertainty can be implemented. This is the so-called gain-scheduled controller but its derivation and implementation is complex and out of the scope of this work. A simpler control law is that of constant state-feedback Proportional-Integral (PI) control given by

$$\Delta\boldsymbol{\beta}(t) = -\mathbf{K}\big[\Delta\mathbf{n}(t)^{\mathsf{T}} \;\; \mathbf{z}(t)^{\mathsf{T}}\big]^{\mathsf{T}} \tag{5}$$

where $\mathbf{z}(t) \in \mathbb{R}^p$ is an additional state vector and $\mathbf{K} \in \mathbb{R}^{N\times(N+p)}$ is a steady-state time-invariant control gain matrix. The vector $\mathbf{z}(t)$ computes the integral of the error signal (see (8) below), which is then used as a feedback term in (5) to provide zero steady-state error under persistent disturbances. In this case, the state of the system $\mathbf{n}$ (total accumulation in the bi-modal network) is measurable in real-time while the uncertain parameter $\boldsymbol{\delta} \in \boldsymbol{\Omega}$ is assumed known at design time and the control gain $\mathbf{K}$ is indirectly dependent on the uncertainty.

A suitable cost criterion for deriving the state-feedback controller (5) is given by the infinite horizon quadratic cost

$$\mathcal{L} = \frac{1}{2}\int_0^\infty \Big(\|\Delta\mathbf{n}(t)\|^2_{\mathbf{W}} + \|\Delta\boldsymbol{\beta}(t)\|^2_{\mathbf{R}} + \|\mathbf{z}(t)\|^2_{\mathbf{Z}}\Big)\, dt \tag{6}$$

where $\mathbf{W} \in \mathbb{R}^{N\times N}$, $\mathbf{R} \in \mathbb{R}^{N\times N}$ and $\mathbf{Z} \in \mathbb{R}^{p\times p}$ are diagonal weighting matrices that are positive semi-definite, positive definite and positive semi-definite, respectively. The weighting matrices can influence the magnitude of the state and control actions and their selection should be performed through a trial-and-error procedure during which the resulting control from (5) should be checked carefully in simulation. The selection of $\mathbf{W} \triangleq \mathrm{diag}(1/n_{i,\,\max})$, $i = 1, \ldots, N$, guarantees that overflow phenomena within each region of study would be avoided (see Aboudolas and Geroliminis (2013) and the design phase in Sections 5.2–5.3 for details), i.e. the state constraints should be directly satisfied to some extent. This criterion aims at maintaining the LPV system (4) to operate around the desired steady-state $(\hat{\mathbf{n}}, \hat{\boldsymbol{\beta}})$ for given $\boldsymbol{\delta} \in \boldsymbol{\Omega}$, while the system's throughput is maximised and steady-state error is minimised.

Applying the LPV model (4) to a bi-modal network partitioned in N regions the following state equation (in vector form) describes the evolution of the system in time (assuming $\Delta\mathbf{d}$ constant or slowly time-varying, i.e. $\Delta\mathbf{d}(t) \approx \mathbf{0}$)

$$\Delta\dot{\mathbf{n}}(t) = \mathbf{F}(\boldsymbol{\delta})\Delta\mathbf{n}(t) + \mathbf{G}\Delta\boldsymbol{\beta}(t), \quad \mathbf{n}(0) = \mathbf{n}_0 \tag{7}$$

where $\Delta\mathbf{n} \in \mathbb{R}^N$ and $\Delta\boldsymbol{\beta} \in \mathbb{R}^N$ are the state and control deviations vectors, respectively; and $\mathbf{F} \in \mathbb{R}^{N\times N}$ and $\mathbf{G} \in \mathbb{R}^{N\times N}$ are the parameter-varying state and control matrices, respectively. In particular, $\mathbf{F}$ is a square matrix with diagonal elements $F_{ii} = -Q_i'(\hat{n}_i, \delta_i)$ and off-diagonal elements $F_{ji} = \gamma_{ji}(t)Q_j'(\hat{n}_j, \delta_j)$ if $j \in \mathcal{S}_i$, and $F_{ji} = 0$ otherwise; $\mathbf{G}$ is an identity square matrix of dimension N.

To provide zero steady-state error under persistent disturbances, we augment the description of the original system (7) with a new state given by

$$\dot{\mathbf{z}}(t) = \mathbf{H}\Delta\mathbf{n}(t), \quad \mathbf{z}(0) = \mathbf{z}_0 \tag{8}$$

where $\mathbf{z} \in \mathbb{R}^p$ is the integral vector, $\mathbf{H} \in \mathbb{R}^{p\times N}$, and $p \le N$ must hold for control-theoretic reasons. The matrix $\mathbf{H}$ typically consists of 0's and 1's such that p components (or linear combinations of components) of accumulation of vehicles are integrated in (7).

Considering the cost criterion (6) and continuous-time system (7), (8), we obtain the following augmented state, control and weighting matrices

$$\mathbf{A}(\boldsymbol{\delta}) = \begin{bmatrix} \mathbf{F}(\boldsymbol{\delta}) & \mathbf{0} \\ \mathbf{H} & \mathbf{0} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} \mathbf{G} \\ \mathbf{0} \end{bmatrix} \quad \mathbf{S} = \begin{bmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z} \end{bmatrix}. \tag{9}$$

To calculate the time-invariant gain matrix $\mathbf{K}$ in (5), which depends only upon the augmented matrices $\mathbf{A}(\boldsymbol{\delta})$, $\mathbf{B}$, $\mathbf{S}$ and $\mathbf{R}$, the uncertain parameter vector $\boldsymbol{\delta} \in \boldsymbol{\Omega}$ is assumed known at design time and $\mathbf{A}(\boldsymbol{\delta})$ is parameterised over the polytopic uncertainty region $\boldsymbol{\Omega}$ (see Appendix B). In this way, the control gain $\mathbf{K}$ is indirectly dependent on the uncertainty.

The LPV state-feedback control problem (6)–(8) with associated matrices (9) can be formulated as a parameter dependent Linear Matrix Inequality (LMI) constraints problem (Becker and Packard, 1994) (see Appendix C). This problem can be solved using Semi-definite Programming (SDP) (Boyd et al., 1994) and efficient interior-point optimisation algorithms (Nesterov and Nemirovskii, 1994). The calculation of control gain $\mathbf{K}$ that minimises an upper bound of the worst-case infinite horizon quadratic cost (6) subject to the LPV system (7)–(8) can be effectuated via solution of the following SDP problem:

$$\begin{aligned} &\min_{\mathbf{P},\mathbf{K}} \quad \mathrm{trace}(\mathbf{P}) \\ &\text{subject to:} \\ &\big(\mathbf{A}(\boldsymbol{\delta}) + \mathbf{B}\mathbf{K}\big)^{\mathsf{T}}\mathbf{P} + \mathbf{P}\big(\mathbf{A}(\boldsymbol{\delta}) + \mathbf{B}\mathbf{K}\big) + \mathbf{S} + \mathbf{K}^{\mathsf{T}}\mathbf{R}\mathbf{K} \preceq \mathbf{0} \end{aligned} \tag{10}$$

where $\mathbf{P}$ is a positive definite matrix of appropriate dimension. Note that $\mathbf{n}^{\mathsf{T}}(0)\mathbf{P}\mathbf{n}(0)$, for all $\mathbf{n}(0) \neq \mathbf{0}$ with $\mathbf{P} \succ \mathbf{0}$ is an upper bound on the worst-case cost if it holds for all possible $\mathbf{A}(\boldsymbol{\delta})$ that satisfy the parameter dependent algebraic Riccati inequality above. This problem is nonconvex in $\mathbf{P}$ and $\mathbf{K}$ that leads to a Bilinear Matrix Inequality (BMI) but can be expressed

Please cite this article as: K. Ampountolas et al., Macroscopic modelling and robust control of bi-modal multi-region urban road networks, Transportation Research Part B (2017), http://dx.doi.org/10.1016/j.trb.2017.05.007

as a parameter dependent LMI by performing a congruence transformation with $\mathbf{Y} = \mathbf{P}^{-1}$, introducing $\mathbf{L} = \mathbf{KY}$, applying the Schur complement lemma, and solving the following convex SDP problem (see Appendix C):

$$\max_{\mathbf{Y},\mathbf{L}} \quad \text{trace}(\mathbf{Y})$$
$$\text{subject to:} \tag{11}$$
$$\begin{bmatrix} -\big(\mathbf{A}(\boldsymbol{\delta})\mathbf{Y} + \mathbf{BL}\big)^{\mathsf{T}} - \big(\mathbf{A}(\boldsymbol{\delta})\mathbf{Y} + \mathbf{BL}\big) & \mathbf{Y} & \mathbf{L}^{\mathsf{T}} \\ \mathbf{Y} & \mathbf{S}^{-1} & \mathbf{0} \\ \mathbf{L} & \mathbf{0} & \mathbf{R}^{-1} \end{bmatrix} \succeq \mathbf{0}.$$

This problem may be readily solved by public available software (e.g., CSDP Borchers (1999)) if the uncertain vector $\boldsymbol{\delta} \in \boldsymbol{\Omega}$ is known and $\mathbf{A}(\boldsymbol{\delta})$ in (11) is parameterised over the polytopic uncertainty region $\boldsymbol{\Omega}$ (see Appendix B). Further, the required computational effort is polynomial even for large-scale problems. After the solution of SDP problem (11), the control gain $\mathbf{K}$ can be recovered from $\mathbf{K} = \mathbf{LY}^{-1}$ ($\mathbf{Y}$ is invertible). 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 (5) with a given constant control gain $\mathbf{K}$ and state measurements $\mathbf{n}(t)$.

To satisfy the control constraints in Section 3, we next incorporate a saturation function in the controller (5) and simply saturate the input signal when it violates the constraint. This leads to the control law:

$$\boldsymbol{\beta}(t) = \text{sat}\Big\{\hat{\boldsymbol{\beta}} - \mathbf{K}\big[\Delta\mathbf{n}(t)^{\mathsf{T}} \;\; \mathbf{z}(t)^{\mathsf{T}}\big]^{\mathsf{T}}\Big\}, \tag{12}$$

where the saturation function is given by

$$\text{sat}\Big\{\beta_i(t)\Big\} \triangleq \begin{cases} \beta_{i,\min}, & \text{if } \beta_i(t) < \beta_{i,\min} \\ \beta_{i,\max}, & \text{if } \beta_i(t) > \beta_{i,\max} \\ \beta_i(t), & \text{otherwise.} \end{cases} \tag{13}$$

On the other hand, the state constraints can be satisfied to some extent by appropriate selection of the weighting matrix $\mathbf{S}$ in (6). The controller (12) is activated in real-time at a specific sample interval T (e.g. every 3–5 min) and only within specific time windows, where the system approaches or exceeds $\hat{\mathbf{n}}$. The obtained saturated $\boldsymbol{\beta}$ values (arriving flows) are then used to define the green times at a number of signalised junctions located along the boundary of neighbour regions or the perimeter of the network. To this end, the arriving flows are distributed to the corresponding junctions and converted to an entrance link green stage duration with respect to the saturation flow of the link, the number of lanes and the cycle time of the junctions (although different splitting policies may be employed). More advanced distribution strategies, as for example queue equalisation can be considered as future work.

5. Application and results

5.1. Network description

The test site is a 2.5 square mile area of Downtown San Francisco (see Fig. 3(a) for a snapshot modelled via the AIMSUN micro-simulator), including 100 junctions and 400 links with lengths varying from 400 ft to 1300 ft. The traffic signals are all multi-phase operating on a common cycle length of 90 s for the west boundary of the protected area and 60 s for the rest. Simulations were performed with time-dependent asymmetric Origin-Destination (OD) tables, starting from different initial compositions of bi-modal traffic. 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 of the test site. Driver adaptation creates MFDs with less hysteresis that represent better real-life conditions (Mahmassani et al., 2013). The initial profile for cars is based on real OD data while the profile for buses is determined by the number of public lines in the network and their operational frequency. Bus routes and frequencies for 29 bus lines in the network have been obtained from the San Francisco Municipal Transportation Agency (SFMTA). Higher demand scenarios are also analysed to generate various mode compositions. The simulation step for the microscopic simulation model of the test site, was set to 0.5 s. The simulation horizon for each scenario is 6 h (08:00–14:00) and pairs of data (n_c, n_b) were gathered every 5 min from the simulator to construct the bi-modal MFDs in Fig. 1.

The implementation of the proposed perimeter strategy to the test site corresponds to the design and application of (12). In the sequel, the proposed controller is designed for both single- and multi-region cases. In the single-region case the composition of traffic is assumed uniformly distributed over the network (see Fig. 3(a)), and a single perimeter controller to 15 signalised junctions (illustrated with blue arrows) in the external boundary of the city (in red colour) is applied. Note that the external boundary involves 24 signalised junctions in total. We also notice that the centre of the network contains more buses than the periphery. Thus, in a two-region approach, a partitioning scheme is applied to cluster the network in two regions with different mode composition (see Fig. 4(a)). In this case, a two-region robust perimeter controller in the perimeter and the boundaries of neighbour regions is applied. To this end, the arriving flows $\boldsymbol{\beta}$ resulting from (12) (operational constrained flows) are distributed to 8 (illustrated with purple arrows in Fig. 3(a)) and 15 signalised junctions (illustrated with

Please cite this article as: K. Ampountolas et al., Macroscopic modelling and robust control of bi-modal multi-region urban road networks, Transportation Research Part B (2017), http://dx.doi.org/10.1016/j.trb.2017.05.007

blue arrows in Fig. 3(a)) located at the perimeter and common boundary of the centre ($i = 1$) and outside ($i = 2$) regions, respectively. It should be noted that both cars and buses are subject to perimeter control at the external boundary shown in Fig. 3(a). Moreover, when buses are close to gated junction some priority is provided to enter the network by extending the arriving link green times and/or skipping stages. No other priority is provided to the buses once they are circulating in the network, as there is no dedicated space or lanes (i.e., mixed traffic). In both cases, the implementation of (12) in the network is effectuated by approximating the continuous space to discrete with sample interval $T = 180$ s (Apkarian, 1997).

Controlling the external boundary of a network will restrict vehicles from entering the network resulting in virtual queues. The delays for these vehicles are estimated as they do not have an option to change their routes and they are obliged to wait until they enter the network. If the majority of these trips are planning to travel in the central region and the control strategy does not affect their aggregated route patterns, then the anticipated benefits of the controller are correctly estimated. Nevertheless, if there are multiple destinations outside the simulated areas or if vehicles change their route choice to avoid the city centre, then indeed a more detailed analysis is necessary to better integrate these effects. Other studies have investigated the effect of perimeter control with MFDs in multi-region cases without controlling the external boundary of the network (Kouvelas et al., 2017) or with route choice integration (Yildirimoglu et al., 2015). Both studies consider car as the only mode of transport. Integrating bi-modal interactions should be a research priority.

5.2. Design of single-region robust perimeter flow control

To design the single-region controller ($N = 1$) the approximation of bi-modal MFD with linear and exponential type speed functions of Section Appendix A is utilised (see also Fig. 1(a)). In particular, the exponential approximation of the bi-modal MFD in Fig. 1(a) and its estimated parameters can be found in Geroliminis et al. (2014). This model is then used to specify the state matrices $\mathbf{F}(\boldsymbol{\delta})$ and $\mathbf{G}$ in (7) (or $\mathbf{A}(\boldsymbol{\delta})$ and $\mathbf{B}$ in (9)) after linearisation of (3) at $(\hat{n}, \hat{\beta})$ (see below for appropriate set-point values). Note that $\mathbf{G}$ is an identity square matrix of dimension N, i.e., $\mathbf{G} = 1$.

Fig. 3(b) depicts a cross-section (cutting plane) of the 3D surface in Fig. 1(a) for a constant accumulation of buses n_b corresponding to a specific number of public transport lines and service frequency in the network (property of the infrastructure), and thus to a slowly time-varying composition of traffic δ. This demonstrates the typical dependence of the flow with the composition of traffic in the network where the maximum flow for fixed δ (operational point) results from an one-variable derivative test of the logarithm of (A.9) with respect to n. In fact, the projection of the 3D surface on the cutting plane $n_b = 200$ veh provides a typical MFD relating the total accumulation n (where n_b is constant) with the outflow in the network. The shape and characteristics of this MFD for the Downtown San Francisco are similar to those found in Aboudolas and Geroliminis (2013) for single mode traffic with $\delta \triangleq 1$. Thus, for different composition of traffic values δ different controllers might be designed. Alternatively, the controller (12) can be designed by solving the SDP problem (11) to achieve robust regulation for all $\delta \in \Omega$, where $\Omega = \{\delta(t) \,|\, \delta_{\min} \leq \delta(t) \leq \delta_{\max}, t \geq 0\}$ (see Section 3 for details).

To determine the compact set Ω, simulations have been performed for different demand scenarios to generate various mode compositions with respect to SFMTA real data for the bus frequencies in the public transport lines. Fig. 3(c) depicts the results obtained for three different scenarios. It can be seen that the composition of traffic varies from $\delta_{\min} = 2\%$ to $\delta_{\max} = 15\%$, i.e. $\Omega = \{\delta(t) \,|\, 0.02 \leq \delta(t) \leq 0.15, t \geq 0\}$. To design the robust controller, the desired accumulation $\hat{n}$ is selected within the optimal range of the bi-modal MFD for maximum output with respect to Ω. More specifically, the value $\hat{n} = 2500$ veh (approximately 80% of 3000 veh, corresponding nominal arriving flow $\hat{\beta} = 93,540$ veh/h) is selected and the state matrix $\mathbf{A}(\delta) = A(\delta)$ in (11) is parametrised over Ω. The minimum and maximum permissible entrance flow of mixed traffic are given by $\beta_{\min} = 20,000$ veh/h and $\beta_{\max} = 120,000$ veh/h, respectively. The arriving flows (operational constrained flows) are distributed to 15 signalised junctions located at the perimeter of the network (illustrated with blue arrows in Fig. 3(a)). The selection of the weighting matrices $\mathbf{S} = \text{diag}(W, Z)$ and $\mathbf{R} = R$ (see (6) and (9)) should be performed through a trial-and-error procedure during which the resulting control from (11) should be checked carefully in simulation. In fact, high values of Z and R may result in very strong weighting and may lead to a nervous control behaviour, and possibly instability of the control system. Thus from a trial-and-error procedure, the weighting matrices in the cost criterion are chosen equal to $W = 1/n_{\max}$ ($n_{\max} = 10,000$ veh), $Z = 0.000001$, and $R = 0.0001$, respectively. The selection of $W = 1/n_{\max}$ guarantees that overflow phenomena within the region of study would be avoided and that regions of different sizes will be treated in a comparable way (see the first term in (6) and Aboudolas and Geroliminis (2013) for details). These values of the parameters above, were found to lead via the solution of problem (11) to control gain $K = 0.0667$ 1/h. Note that although the weighting matrices $\mathbf{S} = \text{diag}(W, Z)$ and $\mathbf{R} = R$ are chosen via a trial-and-error procedure, the designed control is robust with respect to any uncertainty in Ω.

5.3. Design of two-region robust perimeter and boundary control

Fig. 4(a) depicts a snapshot of the network and its partitioning into two regions ($N = 2$). The clustering builds on a modified partitioning algorithm proposed in Ji and Geroliminis (2012). The objective of partitioning networks with bi-modal traffic is to obtain small variance of traffic composition δ_i^{ζ} for all links ζ within each region $i = 1, 2$. As can be seen in Fig. 4(a) there exists a strong heterogeneity in the spatial distribution of δ_i^{ζ} (and, thus congestion), when comparing the mode composition between the city centre (red colour) and the rest of the network (green colour). Fig. 4(b)–(c) depict the resulting bi-modal vehicle MFDs of the two regions (centre and outside) relating accumulation of cars n_c and buses n_b with

Please cite this article as: K. Ampountolas et al., Macroscopic modelling and robust control of bi-modal multi-region urban road networks, Transportation Research Part B (2017), http://dx.doi.org/10.1016/j.trb.2017.05.007

circulating flow, after partitioning. As a first remark, Fig. 4(b)–(c) confirm the existence of bi-modal MFD like-shapes for the two urban regions (cf. with the bi-modal MFD of the whole network in Fig. 1(a)). Note that the one region is larger than the other and n_b of the centre region has a higher range of values, as this region is covered by more public transport lines. Clearly the composition of traffic δ_i in each region i affects the shape of the bi-modal MFD (cf. Fig. 4(b) with Fig. 4(c)). Details on the development of the two bi-modal MFDs can be found in Geroliminis et al. (2014).

For designing the controller in the case of $N = 2$ the bi-modal MFDs in Fig. 4(b)–(c) are used. Following the same procedure as in Section 5.2, the controller (12) can be designed by solving the SDP problem (11) to achieve robust regulation for all $\boldsymbol{\delta} = [\delta_1\ \delta_2]^{\mathsf{T}} \in \boldsymbol{\Omega} = \cup_{i=1}^{2} \Omega_i$, where $\Omega_i = \left\{\delta_i(t) \,|\, \delta_{i,\min} \le \delta_i(t) \le \delta_{i,\max}, t \ge 0\right\}$ for regions $i = 1, 2$. Simulation tests were revealed that the composition of traffic δ_i in each region i (centre $i = 1$ and outside $i = 2$) belongs to the set $\Omega_1 = \{\delta_1(t) \,|\, 0.02 \le \delta_1(t) \le 0.14\}$ and $\Omega_2 = \{\delta_2(t) \,|\, 0.01 \le \delta_2(t) \le 0.06\}$, respectively. The value $\gamma_{ji} = 1,\ \forall i, j = 1, 2$ is selected for the splitting rates. The desired accumulation $\hat{n}_i,\ i = 1, 2$ is selected within the optimal range of the corresponding bi-modal MFDs in Fig. 4(b) and (c) for maximum flow with respect to $\Omega_i,\ i = 1, 2$. More specifically, the values $\hat{n}_1 = 800$ veh (corresponds to a nominal arriving flow $\hat{\beta}_1 = 73,150$ veh/h) and $\hat{n}_2 = 1250$ veh (corresponds to a nominal arriving flow $\hat{\beta}_2 = 80,420$ veh/h) are selected and the state matrix $\mathbf{A}(\boldsymbol{\delta})$ is parametrised over $\boldsymbol{\Omega}$. Note that the summation $\hat{n}_1 + \hat{n}_2$ is smaller than the critical $\hat{n}$ of the single region controller, as the two regions reach congestion at different times with different compositions $\delta_i,\ i = 1, 2$. The minimum and maximum permissible entrance flow of mixed traffic of the centre ($i = 1$) and outside ($i = 2$) regions are given by $\beta_{1,\min} = 20,000$ veh/h and $\beta_{1,\max} = 80,000$ veh/h; $\beta_{2,\min} = 20,000$ veh/h and $\beta_{2,\max} = 100,000$ veh/h, respectively. The weighting matrices $\mathbf{S} = \operatorname{diag}(\mathbf{W}, \mathbf{Z})$ and $\mathbf{R}$ in the cost criterion (6) are chosen diagonal. More precisely, the diagonal elements of $\mathbf{W}$ are set equal to the inverses of the maximum accumulation of the corresponding regions, i.e. $W_{ii} = 1/n_{i,\max},\ i = 1, 2,$ where $n_{1,\max} = 3000$ veh and $n_{2,\max} = 4000$ veh. The diagonal elements of matrices $\mathbf{R}$ and $\mathbf{Z}$ were set equal to $R_{ii} = 0.001$ and $Z_{ii} = 0.0001,\ i = 1, 2,$ respectively. These values of weights, were found to lead via the solution of problem (11) to a control gain $\mathbf{K} \in \mathbb{R}^{2\times 2}$ that exhibits robust regulation and good performance.

5.4. Single-region robust control results

Fig. 5(a) and (b) depict the resulting MFD of five scenarios under pre-timed and single-region robust perimeter control cases. When perimeter control is applied, the network operates under efficient traffic conditions and states in the decreasing part of the MFD are not observed; under pre-timed control, the network becomes severely congested with states in the congested regime of the MFD. Moreover, the outflow is maintained to high values around the set point $\hat{\beta}$. We can also observe that the hysteresis formed in the offset period of congestion is reduced significantly. Fig. 5(c) depicts the resulting average performance (space-mean speed of buses and cars in km/h, average delay of buses and cars in min/km/veh, and stops of mixed traffic in number of stops/km/veh) for ten replications. Clearly the proposed perimeter control significantly increases the speed and decreases delays and number of stops in both modes of traffic. It also shows boxplots for the 25%, 50% and 75% percentiles. It is evident that the reliability of both modes have increased besides the decrease in all congestion performance measures. This highlights the robustness of the feedback controller to demand uncertainty and stochasticity in variables that are not utilised by the controller, like the route choice, the driver behaviour and the detailed bus routes.

Traffic conditions are identical for both control cases up to around 10:00. When perimeter control is switched on (as the total accumulation in the network reaches its set point), the perimeter strategy restricts the flow vehicles are allowed to enter the network to keep it from becoming congested. This results in a small drop in speed within the considered public transport lines for a short time period after perimeter control is activated. However, these temporary delays under perimeter control are proved beneficial for the total network circulating flow and the operation of the public transport lines, as can be seen from the speed profiles after 11:00. Thus, even if cars and buses are restricted in the perimeter of the network, they are able to reach their destinations faster than in the pre-time control case. Moreover, buses are able to travel with free flow traffic conditions inside the protected network and follow their normal time schedule. In particular, improvement is more significant for public transport lines that cover the centre of the network, especially the Market Avenue, one of the severely congested avenues in San Francisco, where largely covered by public transport lines 12, 15 and 17 (see Fig. 4(a)).

An analysis of the spatiotemporal dimension of traffic congestion in the central avenue (Market Avenue) of the network and its upstream links (southeast) can shed some light in the perimeter control actions within the transport public lines. The considered path includes the entire route for public lines 15 and 19 and six other public transport lines that overlap part of the path (see Fig. 4(a)), to investigate the interaction among conflicting public transport lines. To gather the bus trajectories that traverse this path, we simulate buses equipped with GPS-based trackers reporting location and speed every 3 s. Fig. 6(a) and (b) display the gathered bus trajectories for eight public transport lines (each with different colour) during the heart of the rush (11:00 to 13:00), when pre-timed control and single-region robust perimeter control are applied. Fig. 6(c), which illustrates the two-region robust control, is analysed later in the text. In these plots, the x-axis reflects the simulation time, while the y-axis reflects the one-dimensional distance travelled. Given that the studied network is a grid, the two-dimensional road distance is transformed into one-dimensional by calculating the Manhattan distance (ℓ_1-norm) between the GPS-reported location of a bus and the starting point of the path. The horizontal distance between consecutive bus trajectories with the same colour indicates the time headway between two buses servicing the same public transport line. The location of junctions and bus stops are also reported (see caption of Fig. 6 for details) to allow a better understanding of the stop-and-go phenomena within the public transport lines.

Please cite this article as: K. Ampountolas et al., Macroscopic modelling and robust control of bi-modal multi-region urban road networks, Transportation Research Part B (2017), http://dx.doi.org/10.1016/j.trb.2017.05.007

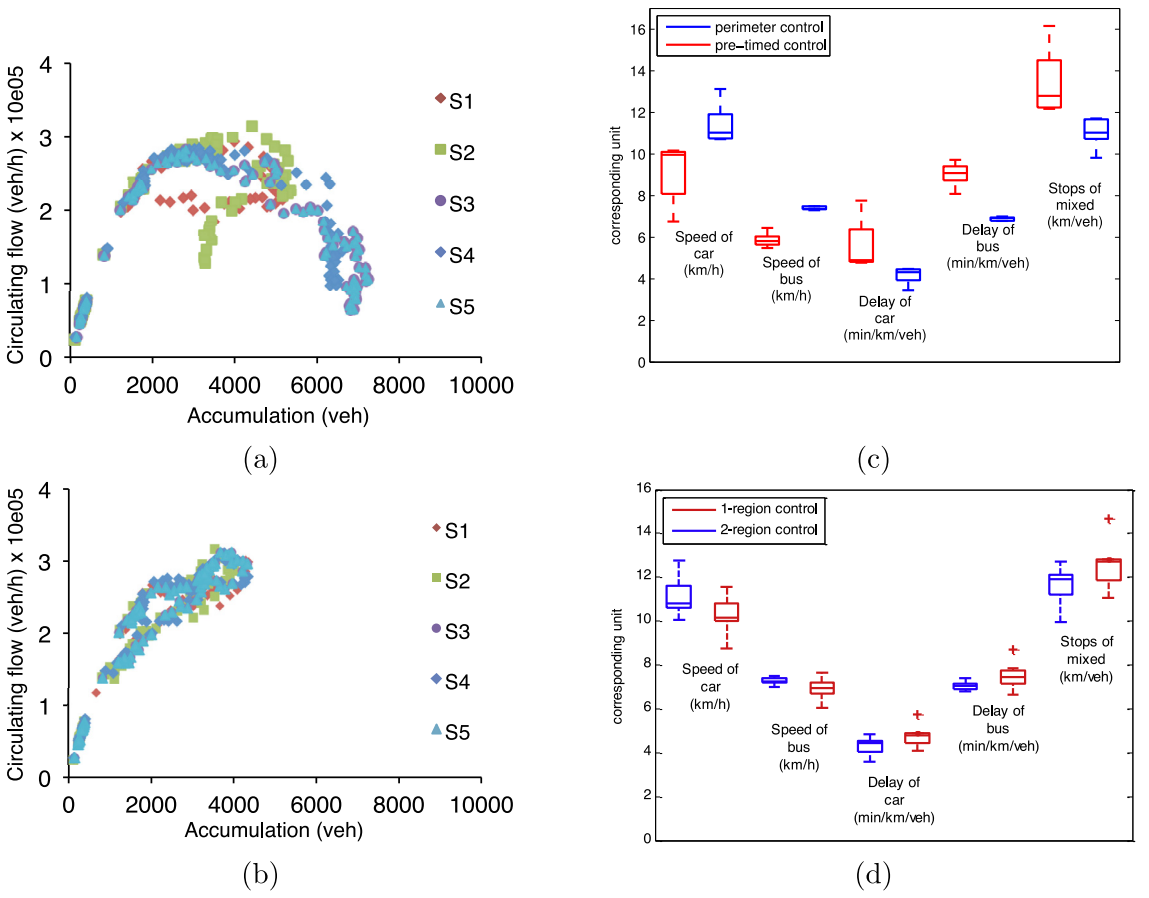

Fig. 5. (a) and (b): MFDs under pre-timed and single-region robust perimeter control cases, respectively (colors represent 5 different scenarios). Accumulation at *x*-axis indicates mixed bi-modal traffic; (c) and (d): minimum, maximum, and median values of different performance indices for the two modes of traffic under pre-timed control, single- and two-region robust perimeter control. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 6(a) and (b) underline the superiority of single-region robust perimeter flow control over pre-timed control to maintain public transport lines normal time schedule. As can be seen, traffic conditions are almost identical for both control cases from 11:00 to 11:20, as time goes on, in the pre-timed control case, buses entering their transport lines (upstream traffic) suffer increasing delays waiting other buses and cars in the centre of the network between 700 m and 1200 m (downstream traffic) to be served. Then traffic conditions are deteriorated in the centre of the network, link queues start spilling back and blocking upstream junctions; thus the entering traffic approximately matches the speed of the downstream traffic. This creates multiple backward moving shockwaves with negative speed that are illustrated with arrows in Fig. 6(a). Clearly when perimeter control is applied (cf. Fig. 6(b) with Fig. 6(a)) the network operates under free-flow traffic conditions and buses are able to follow their normal time schedule (with slight travel delays). More specifically, it can be seen that buses only experience delays between 11:50 and 12:30 at the same spatial distance. To further investigate what caused these delays, the traffic conditions in public transport lines 10 and 11 (among others) were carefully analysed. The inspection of different replications eventually shown that the delays are mainly caused by a sudden increase of left turn demand of cars and buses at a specific junction close to the protected network. Note that the existence of such cases can be possible under the perimeter control scheme, since we only control junctions at the perimeter of the network. In these cases, local control can be employed for smoothing the bus movements within regions.

5.5. Comparison between single-region and two-region robust control

The motivation and rationale behind the development of a two-region robust perimeter and boundary flow control scheme is attributed to the fact that the centre of the network contains more public transport lines than the periphery (especially the Market Avenue is covered by a high number of public transport lines, see Fig. 4(a)). Thus, the traffic composition δ^{ζ} for all links ζ within the centre and the rest of the network could observe high variances. The results presented in the sequel are based on the two-region clustering in Fig. 4 and the control design in Section 5.3. It is used to exploit and

Please cite this article as: K. Ampountolas et al., Macroscopic modelling and robust control of bi-modal multi-region urban road networks, Transportation Research Part B (2017), http://dx.doi.org/10.1016/j.trb.2017.05.007

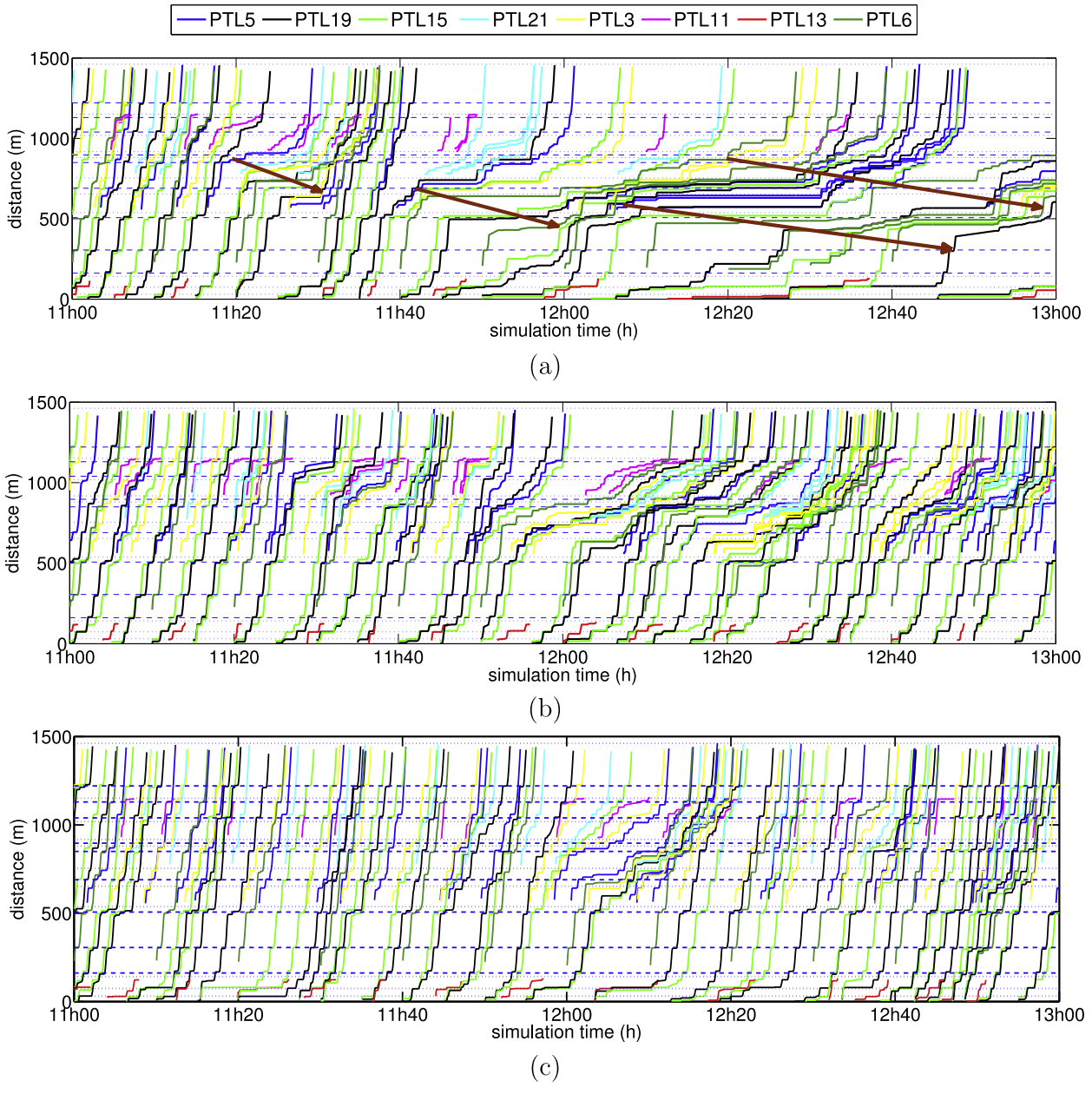

Fig. 6. Time-space diagrams for bus trajectories in eight public transport lines in the network during the heart of rush, under (a) pre-timed control, (b) single-region robust perimeter control, and (c) two-region robust perimeter and boundary control. Horizontal dotted lines indicate the location of junctions; horizontal dashed lines indicate the location of bus stops.

illustrate the benefits of two-region perimeter and boundary control over single-region (whole network) perimeter control. Single-region (whole network) robust perimeter control is designed according to Fig. 3 and Section 5.2.

Fig. 5(d) depicts the average performance (space-mean speed, delay, stops) and same percentiles as in Fig. 5(c) of each mode of traffic for ten replications of a demand scenario with strong heterogeneity in the spatial distribution of δ_i, $i = 1, 2$ for the centre and outside regions, respectively. It can be seen that the two-region control increases the speed and decreases delays and number of stops in both modes of traffic in average by 10%. It should be noted that for scenarios with small variability of δ_i the performance of the two-region over single-region control is slightly deteriorated, as expected.

To further exploit and illustrate the benefits of two-region perimeter and boundary control over single-region (whole network) perimeter control, Fig. 6(c) displays the gathered bus trajectories under two-region perimeter and boundary flow control. Under single-region robust control (see Fig. 6(b)), it can be seen that buses being served in the centre of the network between 700 m and 1000 m at 11:50–12.40 facing delays because of conflicting public transport lines that overlap

Please cite this article as: K. Ampountolas et al., Macroscopic modelling and robust control of bi-modal multi-region urban road networks, Transportation Research Part B (2017), http://dx.doi.org/10.1016/j.trb.2017.05.007

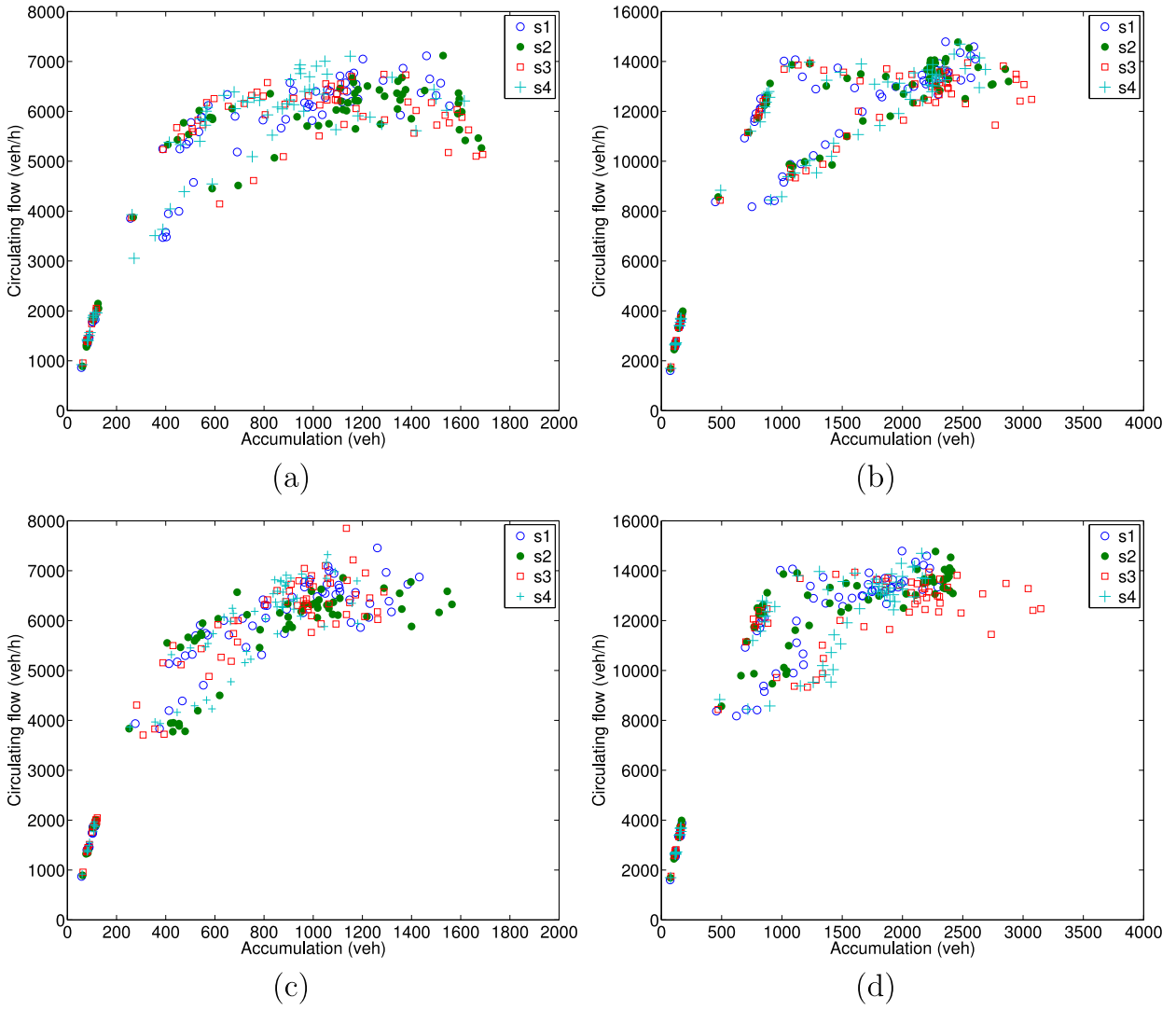

Fig. 7. Comparison of single- and two-region control: (a, b) MFDs under single-region control for the centre ($i = 1$) and outside ($i = 2$) regions; (c, d) MFDs under two-region control for the centre ($i = 1$) and outside ($i = 2$) regions. Accumulation at x-axis indicates mixed bi-modal traffic.

parts of the two regions and heterogeneity in the spatial distribution of δ_i, $i = 1, 2$. Comparing the bus trajectories with Fig. 6(b), two-region robust control is seen to be significantly better than single-region control. Thus, two-region control manages better the heterogeneity in the spatial distribution of vehicle composition and congestion. This is deemed to the actions of two-region control at the common boundary of the neighbour regions. On the other hand, it seems that the two region perimeter control causes delays at the perimeter of the network (spatial distance between 0 m and 100 m). However, this proves beneficial for the reliability of bus services and efficiency of the network. Efficiency and equity are partially competitive criteria, so in the next section we present some results that would shed some light on the bus service reliability and equity.

Fig. 7(a)–(d) depict the MFDs of the centre ($i = 1$) and outside ($i = 2$) regions resulting from the application of single- and two-region perimeter control. As a first remark, the diagrams indicate a hysteresis, i.e., a different path of measurement points when filling the network (onset of congestion) than when emptying (offset of congestion). It should be noted that DTA is active during the simulation (see Section 2), which allows for the drivers to choose their routes adaptively in response to traffic conditions and utilise less congested routes in the network. Given that hysteresis is more pronounced in the MFDs of the outside region (see Fig. 7(b) and (d)), we conjecture[3] that it might be attributed to the longer trip lengths occurred in the periphery of the network than the centre (see Fig. 4(a)) and the stronger heterogeneity in the spatial distribution of

[3] Here we exclude the well-known limitations of micro-simulation tools to emulate the human behaviour and adaptive drivers.

Please cite this article as: K. Ampountolas et al., Macroscopic modelling and robust control of bi-modal multi-region urban road networks, Transportation Research Part B (2017), http://dx.doi.org/10.1016/j.trb.2017.05.007

congestion and the bus lines. We can also observe that the hysteresis formed in the offset period of congestion is reduced significantly for both regions when two-region control is applied. Comparing Fig. 7(a) with Fig. 7(c) (centre region) and Fig. 7(b) with Fig. 7(d) (outside region), two-region control outperforms single-region control in terms of traffic states visited in the congested regimes, which underlines the utility of partitioning in the case that traffic composition indicates a high variability between neighbour regions.

5.6. Comparison between robust perimeter flow control and conventional control

In this section, we compare single- and two-region robust control with the bang-bang control approach proposed by Daganzo (2007), see Eq. (11) in Aboudolas and Geroliminis (2013). The purpose is to demonstrate the advantage of employing a robust control design approach in case of uncertain mode composition in the network. Bang-bang control is proven very efficient in case of single-region perimeter control, even though exhibits an oscillatory control behaviour, as demonstrated in Aboudolas and Geroliminis (2013). To this end, a second realistic scenario was employed with varying mode compositions based on SFMTA real data for the bus frequencies in the public transport lines. Four different control approaches applied and tested, namely optimised pre-timed control, single-region bang-bang perimeter control, single-region robust control, and two-region robust control. To design the single-region bang-bang strategy the values of $\hat{n}$, $\beta_{\min}$, and $\beta_{\max}$ are selected according to the analysis in Section 5.2. The bang-bang control is applied only to 15 junctions located at the perimeter of the test network (see Fig. 3(a)), as in single-region robust control case.

Fig. 8 displays the gathered bus trajectories for the four control approaches. Under pre-timed signal control and single-region bang-bang control (cf. Fig. 8(a) with Fig. 8(b)), it can be seen that buses being served in the centre of the network between 700 m and 1000 m after 11:30 facing significant delays. As expected, bang-bang control indicates better performance than the optimised pre-timed signal control. Nevertheless, it mainly shifts the active bottleneck a couple of intersections downstream, but still delays for the buses are unavoidable. Fig. 8(c) and Fig. 8(d) demonstrate that both single- and two-region robust control are more efficient than conventional bang-bang control, while two-region robust control is seen to be superior to all other strategies. Buses are slowing down during the heavy period, but they are able to traverse the most congested parts of the network with significantly lower stops and spillbacks very rarely occur (only a few horizontal lines that last for a couple of minutes).

Fig. 9 depicts the average speeds and average delays for the four control approaches. As can be seen pre-timed signal control and bang-bang control perform quite well from 11:00 to 11:30 but their efficiency is significantly deteriorate after 11.30 when demand input is at the highest values and congestion develops in the network. The ranking of the strategies with respect to these two criteria is in agreement with the findings of previous sections and the time-space diagrams in Fig. 8. In particular, at the end of the simulated congestion (13:00), the average speeds are (see Fig. 9(a)) 5 km/h for the pre-timed signal control, 8 km/h for the single-region bang-bang control, 10 km/h for the single-region robust control, and 12 km/h for the two-region robust control. As expected, the average delays per trip follow the opposite trend as shown in Fig. 9(b). As it will be explained in the next section, improvements are more significant if we look at detailed performance measures, as average speed also integrates buses moving at less congested parts of the network.

5.7. Level of bus service and equity

While the main goal of the control law (12) is to avoid states in the congested regime of the bi-modal MFD with respect to the variability of $\boldsymbol{\delta}$, it significantly improves the reliability of bus services, even if they share the same infrastructure with cars. To investigate the performance of the buses under the proposed control scheme, bus headway data is collected and headway distribution is chosen as a performance indicator of the bus service reliability. As the analysed public transport lines operate at high frequency, deviation from the ideal headway is more important than the scheduled arrival at bus stops.

Fig. 10(a)–(c) display the headway distributions of four selected public transport lines while Fig. 11(a) displays the normalised headway distribution under pre-timed control, single-region bang-bang control, and two-region robust control. The scheduled headway of each line is provided next to the line number, and the mean and the standard deviation of the actual headways are calculated (Fig. 10(a)–(c)). The four public transport lines are operating through the main streets of the network. The headways are collected for every bus stop during the peak period from 11:00 to 13:00, while the normalisation is calculated as $(h_i - s_i)/s_i$, where h_i and s_i are the real and scheduled headways of public transport line i, respectively. It can be observed that the mean headways are nearly close to the scheduled ones under two-region robust perimeter control (compare the mean with scheduled headways in Fig. 10(c)), and the deviation from schedule is much less than the one under pre-timed and single-region bang-bang control (Fig. 10(a) and (b)). The perimeter control also avoids significant delays, as there are hardly cases with headways higher than 20 min. Standard deviation is also two to three times smaller with the robust perimeter control.

The normalised bus headway distribution for all lines together provides additional insights for the reliability of the system as a whole. A larger concentration of values around 0 indicates a smaller deviation from the schedule and higher reliability (cf. the three subfigures in Fig. 11(a)). Note that under pre-timed control (see the first two subfigures in Fig. 11(a)) the distribution is positively skewed (skewness measures distribution asymmetry) and quite often the deviation from schedule is high, which results to higher waiting times for the passengers at bus stops. This indicates that the proposed robust perimeter control improves the performance of buses without providing bus signal priority or considering dedicated bus

Please cite this article as: K. Ampountolas et al., Macroscopic modelling and robust control of bi-modal multi-region urban road networks, Transportation Research Part B (2017), http://dx.doi.org/10.1016/j.trb.2017.05.007

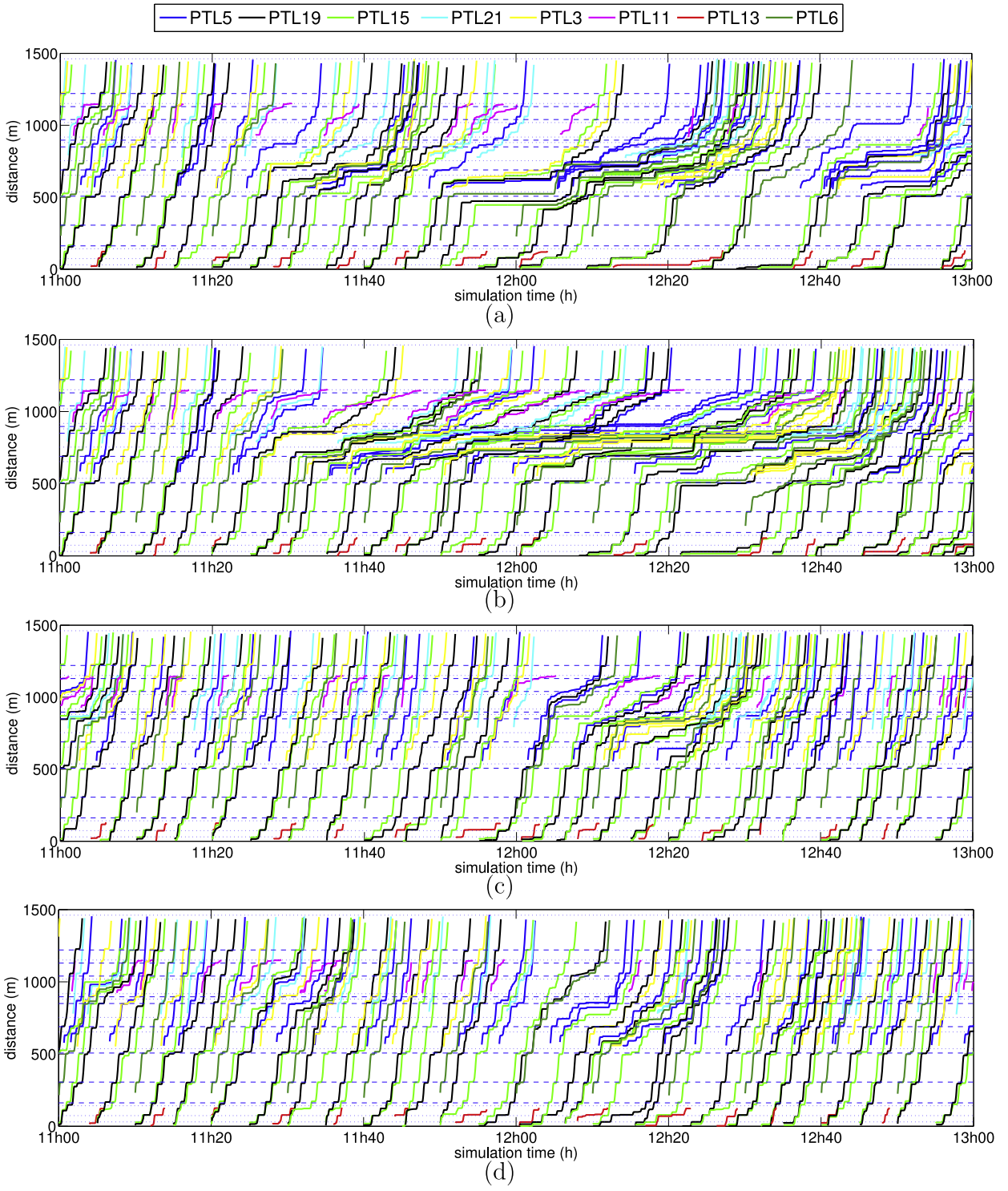

Fig. 8. Time-space diagrams for bus trajectories in eight public transport lines in the network during the heart of rush, under (a) pre-timed control; (b) bang-bang perimeter control; (c) single-region robust control; (d) two-region robust control.

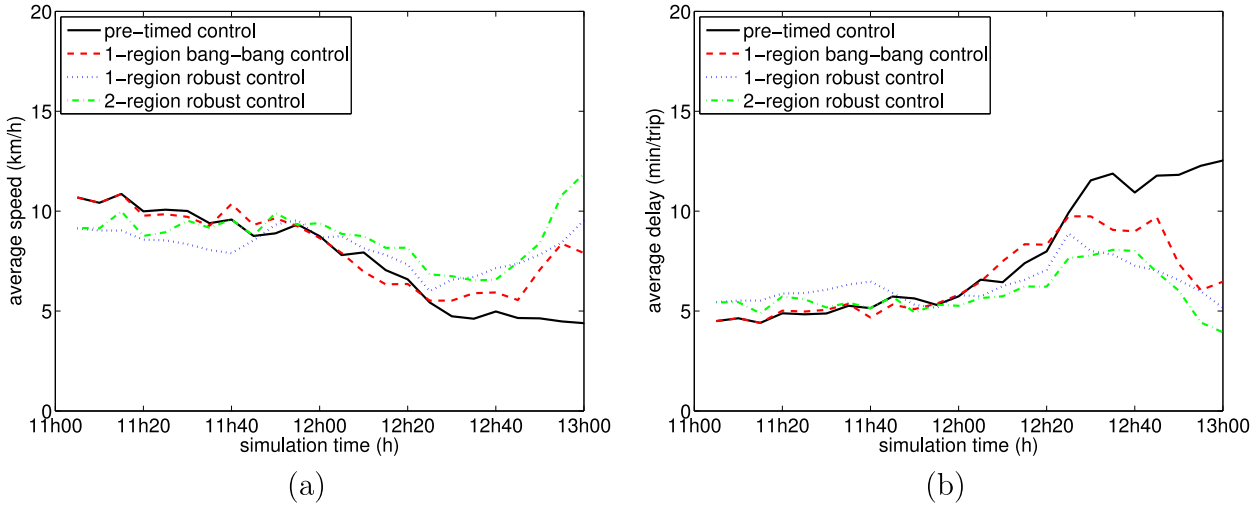

Fig. 9. (a) Average speeds and (b) average delays per trip for robust perimeter flow control and conventional bang-bang control.

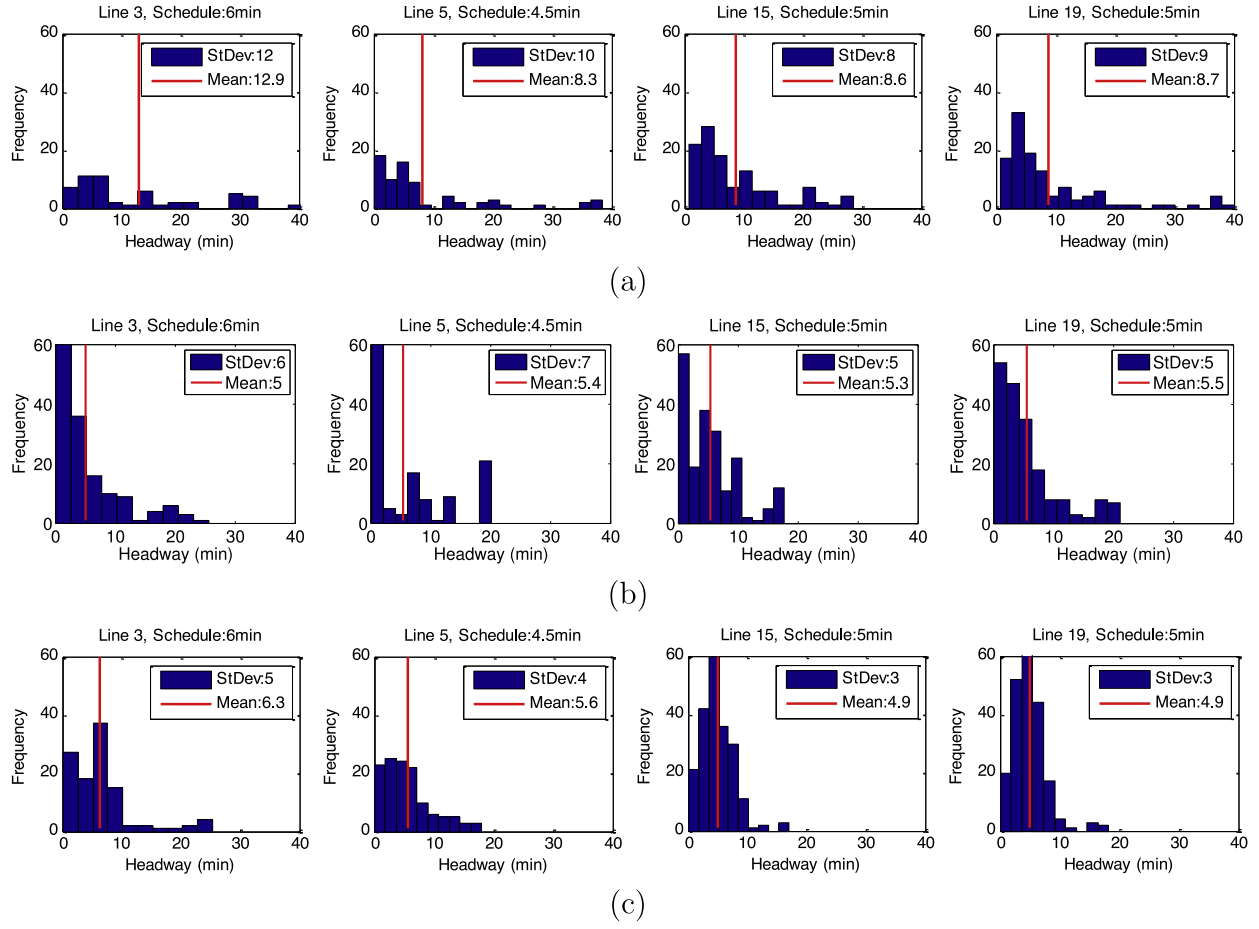

Fig. 10. Bus headway distributions and service irregularity of 4 selected public transport lines under: (a) pre-timed control, (b) single-region bang-bang control, (c) two-region robust control.

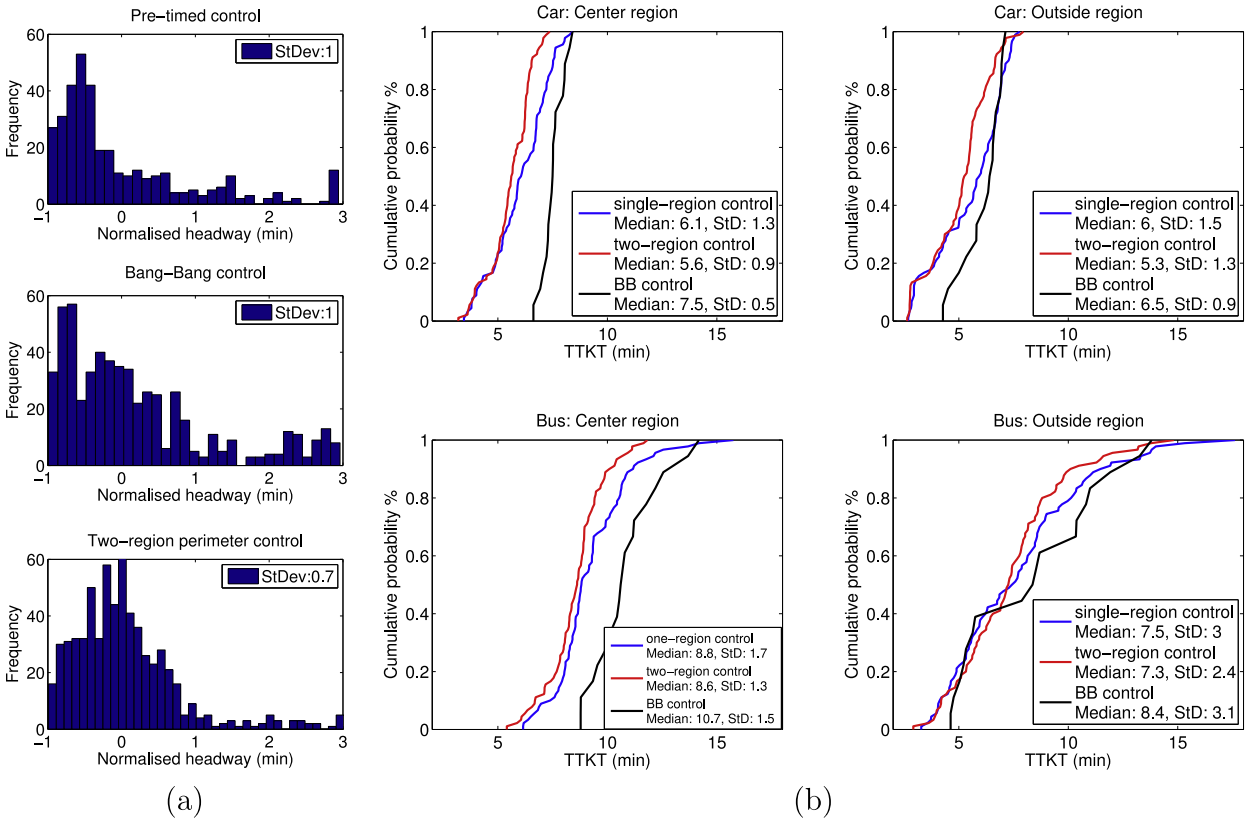

Fig. 11. (a) Normalised headway distributions for all public transport lines under pre-timed control, bang-bang perimeter flow control, and two-region perimeter and boundary flow control; (b) Cumulative distributions of TTKT for the two modes of traffic under various control schemes.

lanes. Thus such a control policy can influence the traditional way of public transport management, increase mobility and improve the reliability of bus services in congested bi-modal networks.

Let us now investigate the efficiency and equity properties of both modes of traffic under various control schemes. Travel Time per Kilometre-distance Travelled (TTKT) is chosen as performance indicator. Considering TTKT allows us to compare travel times of the same scale in case those travel times vary significantly due to different trip lengths. Different trip lengths are natural because the size of the two regions is uneven and the two modes of traffic generate different trips. Fig. 11(b) displays the cumulative probability distribution of the TTKT for each region and mode (cars and buses). TTKTs were calculated for 5 scenarios, every 5-min for a 2.5-h peak period. The three curves (with blue, red, and black colour) in each of the four subplots depict cumulative distributions of TTKT under single-region robust control, two-region robust control, and single-region bang-bang control, respectively. The median and the standard deviation of each distribution are displayed. As can be seen, the two-region robust control performs better than both single-region robust control and single-region bang-bang control. The median value of TTKT for cars improve by about 10% for both regions. For buses the improvement in terms of median is quite small (about 3%), but reliability increases (standard deviation is 20% smaller). Remarkably, the two-region control increases the reliability of the network as its cumulative curve of TTKT is less spreading. Finally, it is stimulating to observe that the two-region control provides fairly equal improvements to both regions, albeit the centre region seems to benefit slightly more than the outside region.

6. Discussion

In this paper, we addressed the problem of perimeter flow control in bi-modal multi-region urban networks. We presented a bi-modal MFD for mixed urban traffic and described the dynamics of cars and buses by an LPV system with bounded uncertainty for single- and multi-region cities. We then designed a PI controller for the LPV system that guarantees robust regulation and stability around a desired set point at the bi-modal MFD, while the system's throughput (for both travel modals) is maximised. We also considered extensions in case each region has different values of mode composition. We implemented the proposed controller in a simulation study in Downtown San Francisco. Results showed that the designed robust control can significantly: (i) reduce the overall congestion in the network, (ii) improve the traffic perfor-

Please cite this article as: K. Ampountolas et al., Macroscopic modelling and robust control of bi-modal multi-region urban road networks, Transportation Research Part B (2017), http://dx.doi.org/10.1016/j.trb.2017.05.007

mance of buses in terms of travel delays and schedule reliability, and (iii) avoid queues and gridlock on critical paths of the network.

While the main goal of the developed robust control strategy is to avoid states in the congested regime of the bi-modal MFD with respect to the variability of vehicle composition, it significantly improves the efficiency and reliability of bus services, even if they share the same infrastructure with cars. The proposed strategy could be easily integrated with more detailed control strategies that have local objectives. For instance, equalising queues and avoiding spill backs at the perimeter intersections or even within the partitioned regions (see e.g., the TUC strategy (Diakaki et al., 2002; 2003), store-and-forward paradigm (Aboudolas et al., 2009), max pressure control (Kouvelas et al., 2014), or optimising offsets for major routes of the city or for bimodal systems strategies as dynamic bus lanes, pre-signals for buses and local bus priority (see e.g., Skabardonis and Geroliminis, 2008; Guler and Cassidy, 2012; Guler and Menendez, 2014; Christofa et al., 2013; 2016). The fact that the upper level MFD type control is able to bring the system at less congested and more efficient states, it facilitates also the performance of local strategies, which would be unable to operate efficiently under heavily congested conditions. The reason is that local controllers might have conflicting objectives under heavy congestion, but these conflicts are not as severe under normal conditions, for example queues do not spill back from one intersection to the other. Further research is needed towards the system of systems control direction and how to further connect upper and lower level controllers.

On the modelling side, strong demand fluctuations either for cars or buses can create fast evolving transient states, spatial heterogeneity of congestion or route choice effects that influence the trip length distribution of vehicles in the network. As a result, the outflow in Eq. (1) might exhibit some variations or hysteresis (Mahmassani et al., 2013; Yildirimoglu and Geroliminis, 2014) and the outflow-MFD might not provide a good approximation as it is "memoryless," i.e. it ignores the history of the system. While such a model is firstly mentioned in Arnott (2013) and later described in Lamotte and Geroliminis (2016), its complexity makes the integration in control very challenging. Towards the same direction, another challenging topic is the theoretical analysis of the multi-region bi-modal nonlinear system with time varying delays, in order to develop a controller that takes into account the state-dependent delays and compare it with the one presented here. This is expected to shed more light on the importance of these delays under various demand and mode composition variations.

Acknowledgements

The first author would like to acknowledge support by the UK Economic and Social Research Council (ESRC) (grant number ES/L011921/1). The second and third authors would like to acknowledge support by the (European Research Council) ERC Starting Grant "METAFERW: Modeling and controlling traffic congestion and propagation in large-scale urban multimodal networks" (grant number 338205).

Appendix A. Analytical derivations for the critical vehicle accumulation

We provide in this subsection analytical derivations of the critical network accumulation that maximizes the flow as a function of the mode composition for different functional forms. This analysis is useful to determine in a rigorous way the set-points of the developed control strategies in Section 3.

For each class of vehicle we assume a linear-decreasing relationship between average speed and accumulation given by (subscript i is omitted)

$$V_c(n_c, n_b) = \alpha_c n_c + \beta_c n_b + \gamma_c \tag{A.1}$$

$$V_b(n_c, n_b) = \alpha_b n_c + \beta_b n_b + \gamma_b \tag{A.2}$$

where V_c and V_b are the space-mean speed of cars and buses, respectively, and α_m, β_m, γ_m, $m \in \{c, b\}$ are constant model parameters that can be estimated with real data for buses and cars (e.g., GPS equipped vehicles and loop detectors). The space-mean speed in the network is given by

$$V = \frac{V_c n_c + V_b n_b}{n_c + n_b} \tag{A.3}$$

where $n = n_c + n_b$ is the total accumulation in the network. For given (A.1)–(A.3) a city-scale bi-modal vehicle MFD for mixed traffic may be calculated as follows

$$\begin{aligned} Q(n_c, n_b) &= nV/L = (n_c V_c + n_b V_b)/L \\ &= \left(\alpha_c n_c^2 + \beta_b n_b^2 + (\alpha_b + \beta_c) n_c n_b + \gamma_c n_c + \gamma_b n_b\right)/L \end{aligned} \tag{A.4}$$

where L is the average trip length of all vehicles moving in the network. Consider now a composition of traffic $n_c = \delta n$ and $n_b = (1-\delta)n$, where $0 < \delta < 1$,[4] the bi-modal MFD in (A.4) can be rewritten as

$$Q(n, \delta) = \left(\left[\alpha_c \delta^2 + \beta_b (1-\delta)^2 + (\alpha_b + \beta_c)\delta(1-\delta)\right] n^2 + \left[\gamma_c \delta + \gamma_b (1-\delta)\right] n\right)/L. \tag{A.5}$$

[4] Since $0 < \delta < 1$ the accumulation of buses n_b is not a linear function of the accumulation of cars n_c, i.e. $n_b \neq n_c p$, $p \geq 1$.

In this model, the output is not only a function of the accumulation, but also of the composition of bi-modal traffic in the network $Q \triangleq Q(n, \delta)$.

Now if we want to calculate the critical accumulation n_{cr}, we may first fix the composition rate δ, and then take the one-variable derivative of (A.5) with respect to n and set it equal to zero

$$n_{cr} \triangleq \hat{n} = -\frac{1}{2}\frac{\gamma_c\delta + \gamma_b(1-\delta)}{\alpha_c\delta^2 + \beta_b(1-\delta)^2 + (\alpha_b + \beta_c)\delta(1-\delta)}. \tag{A.6}$$

This formula provides the critical accumulation for different values of the composition rate δ and constant model parameters α_m, β_m, γ_m, $m \in \{c, b\}$.

Let us also consider the case where an exponential-decreasing relationship between space-mean speed and accumulation is given by

$$V_c(n_c, n_b) = e_c \exp\left[-q_c(n_c, n_b)\right] \tag{A.7}$$

$$V_b(n_c, n_b) = e_b \exp\left[-q_b(n_c, n_b)\right] \tag{A.8}$$

where e_c and e_b are constant parameters and the exponent term $q_m(n_c, n_b)$, $m \in \{c, b\}$, is a polynomial function of n_c and n_b. Furthermore, we assume $e \triangleq e_c = e_b$ and $q \triangleq -q_c = -q_b$ for simplification and to be consistent with Geroliminis et al. (2014). Similarly to the derivation of (A.5) above, a more detailed bi-modal MFD for mixed traffic may be calculated as follows

$$Q(n, \delta) = en \exp\left[q(n, \delta)\right]/L \tag{A.9}$$

where

$$\begin{aligned} q(n, \delta) &\triangleq q(n_c, n_b) = \alpha n_c^2 + \beta n_b^2 + \gamma n_c n_b + \epsilon n_c + \theta n_b \\ &= \left[\alpha\delta^2 + \beta(1-\delta)^2 + \gamma\delta(1-\delta)\right]n^2 + \left[\epsilon\delta + \theta(1-\delta)\right]n \end{aligned} \tag{A.10}$$

if $q(n_c, n_b)$ is assumed quadratic and α, β, γ, ϵ, θ are constant model parameters. This second more detailed flow model is consistent with the exponential approximation proposed in Geroliminis et al. (2014).

In order to find the critical accumulation that maximises the 3D-MFD is convenient to work with the logarithm of (A.9). The logarithm is a continuous strictly increasing function over the range of (A.9), thus values that maximise $O(n, \delta)$ will also maximise its logarithm. In this case the critical accumulation for fixed composition rate δ is given by

$$n_{cr} \triangleq \hat{n} = \frac{-y \pm \sqrt{y^2 - 8x}}{4x} \tag{A.11}$$

where $x = \alpha\delta^2 + \beta(1-\delta)^2 + \gamma\delta(1-\delta)$ and $y = \epsilon\delta + \theta(1-\delta)$. A similar analysis could also be performed with the passenger MFD of Fig. 1(c).

Appendix B. Polytopic uncertainty

Consider a Linear Parameter-Varying (LPV) system

$$\mathbf{x}(k+1) = \mathbf{A}(\boldsymbol{\delta})\mathbf{x}(k) + \mathbf{B}(\boldsymbol{\delta})\mathbf{u}(k), \qquad k = 0, 1, \ldots \tag{B.1}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the state vector, $\mathbf{u} \in \mathbb{R}^m$ is the control vector, the matrices $\mathbf{A}(\cdot)$ and $\mathbf{B}(\cdot)$ are known functions of the parameter $\boldsymbol{\delta} \in \boldsymbol{\Omega}$, and $\boldsymbol{\Omega} \in \mathbb{R}^{n_\delta}$ is a compact set. Polytopic uncertainty may be described by the time-varying system

$$\mathbf{x}(k+1) = \mathbf{A}(k)\mathbf{x}(k) + \mathbf{B}(k)\mathbf{u}(k), \qquad k = 0, 1, \ldots \tag{B.2}$$

where

$$[\mathbf{A}(k)\ \mathbf{B}(k)] \in \boldsymbol{\Omega} = \text{Conv}\{[\mathbf{A}_1\ \mathbf{B}_1], [\mathbf{A}_2\ \mathbf{B}_2], \ldots, [\mathbf{A}_\mathcal{V}\ \mathbf{B}_\mathcal{V}]\} \tag{B.3}$$

with Conv$\{\cdot\}$ denoting the convex hull of a number of "extreme" known system vertices $[\mathbf{A}_i\, \mathbf{B}_i]$, $i = 1, \ldots, \mathcal{V}$, given by

$$[\mathbf{A}(k)\ \mathbf{B}(k)] = \sum_{i=1}^{\mathcal{V}} \lambda_i [\mathbf{A}_i\ \mathbf{B}_i], \quad \text{with} \quad \sum_{i=1}^{\mathcal{V}} \lambda_i = 1,\ \lambda_i \geq 0. \tag{B.4}$$

Now if the set $\boldsymbol{\Omega}$ is polytopic, and if $\mathbf{A}(\cdot)$ and $\mathbf{B}(\cdot)$ in (B.1) are affine functions of $\boldsymbol{\delta}$ given by

$$\mathbf{A}(\boldsymbol{\delta}) = \mathbf{A}_{\delta,0} + \sum_{i=1}^{n_\delta} \delta_i \mathbf{A}_{\delta,i} \quad \text{and} \quad \mathbf{B}(\boldsymbol{\delta}) = \mathbf{B}_{\delta,0} + \sum_{i=1}^{n_\delta} \delta_i \mathbf{B}_{\delta,i} \tag{B.5}$$

then the vertex pairs $[\mathbf{A}_i\, \mathbf{B}_i]$ in (B.3) are the mappings $[\mathbf{A}(\boldsymbol{\delta}^1)\mathbf{B}(\boldsymbol{\delta}^1)]$, $\left[\mathbf{A}(\boldsymbol{\delta}^2)\ \mathbf{B}(\boldsymbol{\delta}^2)\right], \ldots, \left[\mathbf{A}(\boldsymbol{\delta}^\mathcal{V})\ \mathbf{B}(\boldsymbol{\delta}^\mathcal{V})\right]$, with $\boldsymbol{\delta}^i$ being an enumeration of the $\mathcal{V}$ vertices of $\boldsymbol{\Omega}$ (Ziegler, 1995).

Please cite this article as: K. Ampountolas et al., Macroscopic modelling and robust control of bi-modal multi-region urban road networks, Transportation Research Part B (2017), http://dx.doi.org/10.1016/j.trb.2017.05.007

Appendix C. Linear matrix inequalities

The set of real matrices of dimension $m \times n$ are denoted by $\mathbb{R}^{m\times n}$. A real matrix $\mathbf{X}$ is symmetric if, and only if, $\mathbf{X} = \mathbf{X}^{\mathsf{T}}$. We denote $\mathbb{S}^n$ the set of real symmetric matrices in $\mathbb{R}^{n\times n}$, $\mathbb{S}^n = \left\{\mathbf{M} \mid \mathbf{M} = \mathbf{M}^{\mathsf{T}}, \mathbf{M} \in \mathbb{R}^{n\times n}\right\}$. A symmetric matrix $\mathbf{X}$ is called positive definite, positive semi-definite, negative definite, or negative semi-definite if $\mathbf{x}^{\mathsf{T}}\mathbf{X}\mathbf{x} > 0$, ≥ 0, < 0 or ≤ 0 for all nonzero vectors $\mathbf{x} \in \mathbb{R}^n$. For symmetric matrices, $\mathbf{X} \succ \mathbf{0}$, $\succeq \mathbf{0}$, $\prec \mathbf{0}$, $\preceq \mathbf{0}$ indicates that $\mathbf{X}$ is positive definite, positive semi-definite, negative definite, negative semi-definite, respectively. For square matrices trace($\mathbf{X}$) denotes the trace of $\mathbf{X}$.

Definition 1 (Linear Matrix Inequality)**.** A *Linear Matrix Inequality* (LMI) is a constraint of the form

$$\mathbf{F}(\mathbf{x}) \triangleq \mathbf{F}_0 + x_1\mathbf{F}_1 + \cdots + x_m\mathbf{F}_m \succeq \mathbf{0}, \tag{C.1}$$

where $\mathbf{F}_i \in \mathbb{S}^n$, $i = 0, 1, \ldots, m$, are given real symmetric matrices and where $\mathbf{x} \in \mathbb{R}^m$ is a vector of unknown real scalar *decision variables*.

A very important aspect of the LMI is that it defines a *convex set*. The inequality $\mathbf{F}(\mathbf{x}) \succeq \mathbf{0}$ (or $\succ \mathbf{0}$) means that $\mathbf{x}$ should render the symmetric matrix $\mathbf{F}(\mathbf{x})$ positive semi-definite (respectively, positive definite), that is, the maximum eigenvalue of $\mathbf{F}(\mathbf{x})$ should be nonnegative (respectively, positive). An LMI optimisation problem amounts to minimising a linear cost function over all $x_1, \ldots, x_m$ that satisfy constraints of the form (C.1). In other words, LMI optimisation problems are natural generalisations of linear programming problems in which inequalities are defined by the cone of positive definite matrices.

In control applications, LMIs arise with matrix variables rather than vector variables as for instance in problems (10) and (11) in Section 4. That is we consider inequalities of the general form $\mathbf{F}(\mathbf{X}) \succeq \mathbf{0}$, in which $\mathbf{X}$ is a matrix that belongs to an arbitrary finite-dimensional vector space $\mathcal{X}$ of matrices and where $\mathbf{F} : \mathcal{X} \to \mathbb{S}^n$ is an affine function. *Affine functions* assume the form $\mathbf{F}(\mathbf{x}) = \mathbf{F}_0 + \mathbf{F}_x(\mathbf{x})$, where $\mathbf{F}_0$ is fixed and where $\mathbf{F}_x$ is a linear map. Optimisation problems over symmetric semi-definite matrix constraints belong to the realm of *Semi-definite Programming* (SDP) and can be solved efficiently in polynomial time by interior-point algorithms (Nesterov and Nemirovskii, 1994).

In the sequel, we provide few important results needed in relaxing nonlinear matrix inequalities into LMIs. The interested reader is referred to Boyd et al. (1994) for details.

Definition 2 (Congruence Transformation)**.** Two matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{S}^n$ are said to be congruent if there exists a nonsingular matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$ such that $\mathbf{Y} = \mathbf{M}^{\mathsf{T}}\mathbf{X}\mathbf{M}$.

Proposition 1. *If* $\mathbf{X}$ *and* $\mathbf{Y}$ *are congruent then* $\mathbf{Y}\succ\mathbf{0}$ *if, and only if,* $\mathbf{X}\succ\mathbf{0}$.

Proof. If $\mathbf{X}\succ\mathbf{0}$ then $\mathbf{x}^{\mathsf{T}}\mathbf{X}\mathbf{x} > 0$, for all $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} \neq \mathbf{0}$. Since $\mathbf{X}$ and $\mathbf{Y}$ are congruent there exists $\mathbf{M}$ nonsingular such that $\mathbf{Y} = \mathbf{M}^{\mathsf{T}}\mathbf{X}\mathbf{M}$. Hence, using the fact that $\mathbf{M}$ is nonsingular, for all $\mathbf{x} \neq \mathbf{0}$, the vector $\mathbf{y} \triangleq \mathbf{M}^{-1}\mathbf{x} \neq \mathbf{0}$ and $\mathbf{X}\succ\mathbf{0}$ yield

$$\mathbf{x}^{\mathsf{T}}\mathbf{X}\mathbf{x} = \mathbf{y}^{\mathsf{T}}\mathbf{M}^{\mathsf{T}}\mathbf{X}\mathbf{M}\mathbf{y} = \mathbf{y}^{\mathsf{T}}\mathbf{Y}\mathbf{y} > 0$$

that is $\mathbf{Y}\succ\mathbf{0}$. □

The Schur complement lemma plays a fundamental role in converting nonlinear matrix inequalities into linear matrix inequalities.

Lemma 1 (Schur Complement)**.** *For all* $\mathbf{X} \in \mathbb{S}^n$, $\mathbf{Y} \in \mathbb{R}^{m\times n}$, $\mathbf{Z} \in \mathbb{S}^m$, *the following statements are equivalent:*

(a) $\mathbf{Z} \succ \mathbf{0}$, $\mathbf{X} - \mathbf{Y}^{\mathsf{T}}\mathbf{Z}^{-1}\mathbf{Y} \succ \mathbf{0}$
(b) $\begin{bmatrix}\mathbf{X} & \mathbf{Y}^{\mathsf{T}}\\ \mathbf{Y} & \mathbf{Z}\end{bmatrix} \succ \mathbf{0}$

(a) $\mathbf{Z} \succ 0$, $\mathbf{X} - \mathbf{Y}^{\mathsf{T}}\mathbf{Z}^{-1}\mathbf{Y} \succeq \mathbf{0}$
(b) $\mathbf{Z} \succ \mathbf{0}$, $\begin{bmatrix}\mathbf{X} & \mathbf{Y}^{\mathsf{T}}\\ \mathbf{Y} & \mathbf{Z}\end{bmatrix} \succeq \mathbf{0}$

Proof. Assume $\mathbf{Z} \succ \mathbf{0}$. The nonsingular matrix

$$\mathbf{M} = \begin{bmatrix}\mathbf{I} & \mathbf{0}\\ -\mathbf{Z}^{-1}\mathbf{Y} & \mathbf{I}\end{bmatrix}$$

establishes the congruence transformation

$$\mathbf{M}^{\mathsf{T}}\begin{bmatrix}\mathbf{X} & \mathbf{Y}^{\mathsf{T}}\\ \mathbf{Y} & \mathbf{Z}\end{bmatrix}\mathbf{M} = \begin{bmatrix}\mathbf{X} - \mathbf{Y}^{\mathsf{T}}\mathbf{Z}^{-1}\mathbf{Y} & \mathbf{0}\\ \mathbf{0} & \mathbf{Z}\end{bmatrix} \succ \mathbf{0}\ (\succeq \mathbf{0}).$$

□

References

Aboudolas, K., Geroliminis, N., 2013. Perimeter and boundary flow control in multi-reservoir heterogeneous networks. Transp. Res. Part B 55, 265–281.
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. Transp. Res. 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. Transp. Res. Part C 18 (5), 680–694.

Please cite this article as: K. Ampountolas et al., Macroscopic modelling and robust control of bi-modal multi-region urban road networks, Transportation Research Part B (2017), http://dx.doi.org/10.1016/j.trb.2017.05.007

Apkarian, P., 1997. On the discretization of LMI-synthesized linear parameter-varying controllers. Automatica 33 (4), 655–661.
Arnott, R., 2013. A bathtub model of downtown traffic congestion. J. Urban Econ. 76, 110–121.
Becker, G., Packard, A., 1994. Robust performance of linear parametrically varying systems using parametrically-dependent linear feedback. Syst. Control Lett. 23 (3), 205–215.
Borchers, B., 1999. CSDP, A C library for semidefinite programming. Optim. Methods Softw. 11, 613–623.
Boyd, S., El Ghaoui, L., Feron, E., Balakrishnan, V., 1994. Linear Matrix Inequalities in System and Control Theory. vol. 15 of Studies in Applied Mathematics. SIAM, Philadelphia, PA.
Buisson, C., Ladier, C., 2009. Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. Transp. Res. Rec. 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.), 13th International Symposium on Transportation and Traffic Theory. Pergamon, Lyon, France, pp. 697–711.
Chiabaut, N., 2015. Evaluation of a multimodal urban arterial: the passenger macroscopic fundamental diagram. Transp. Res. Part B 81 (2), 410–420.
Chiabaut, N., Xie, X., Leclercq, L., 2014. Performance analysis for different designs of a multimodal urban arterial. Transportmetrica B 2 (3), 229–245.
Christofa, E., Ampountolas, K., Skabardonis, A., 2016. Arterial traffic signal optimization: a person-based approach. Transp. Res. Part C 66, 27–47.
Christofa, E., Papamichail, I., Skabardonis, A., 2013. Person-based traffic responsive signal control optimization. IEEE Trans. Intell. Transp. Syst. 14 (3), 1278–1289.
Daganzo, C.F., 2007. Urban gridlock: macroscopic modeling and mitigation approaches. Transp. Res. Part B 41 (1), 49–62.
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. Transp. Res. Rec. 1856, 202–211.
Diakaki, C., Papageorgiou, M., Aboudolas, K., 2002. A multivariable regulator approach to traffic-responsive network-wide signal control. Control Eng. Pract. 10, 183–195.
Edie, L.C., 1963. Discussion of traffic stream measurements and definitions. In: Almond, J. (Ed.), 2nd International Symposium on the Theory of Traffic Flow. OECD, Paris, France, pp. 139–154.
Gayah, V.V., Daganzo, C.F., 2011. Clockwise hysteresis loops in the macroscopic fundamental diagram: an effect of network instability. Transp. Res. Part B 45 (4), 643–655.
Gayah, V.V., Gao, X.S., Nagle, A.S., 2014. On the impacts of locally adaptive signal control on urban network stability and the macroscopic fundamental diagram. Transp. Res. Part B 70, 255–268.
Geroliminis, N., Daganzo, C.F., 2008. Existence of urban-scale macroscopic fundamental diagrams: some experimental findings. Transp. Res. 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 Trans. Intell. Transp. Syst. 14 (1), 348–359.
Geroliminis, N., Sun, J., 2011. Properties of a well-defined macroscopic fundamental diagram for urban traffic. Transp. Res. Part B 45, 605–617.
Geroliminis, N., Zheng, N., Ampountolas, K., 2014. A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks. Transp. Res. Part C 42, 168–181.
Godfrey, J.W., 1969. The mechanism of a road network. Traffic Eng. Control 11 (7), 323–327.
Gonzales, E.J., Daganzo, C.F., 2012. Morning commute with competing modes and distributed demand: user equilibrium, system optimum, and pricing. Transp. Res. Part B 46 (10), 1519–1534.
Gonzales, E.J., Daganzo, C.F., 2013. The evening commute with cars and transit: duality results and user equilibrium for the combined morning and evening peaks. Transp. Res. Part B 57, 286–299.
Guler, S.I., Cassidy, M.J., 2012. Strategies for sharing bottleneck capacity among buses and cars. Transp. Res. Part B 46 (10), 1334–1345.
Guler, S.I., Menendez, M., 2014. Analytical formulation and empirical evaluation of pre-signals for bus priority. Transp. Res. Part B 64, 41–53.
Haddad, J., 2017. Optimal perimeter control synthesis for two urban regions with aggregate boundary queue dynamics. Transp. Res. Part B 96, 1–25.
Haddad, J., Geroliminis, N., 2012. On the stability of traffic perimeter control in two-region urban cities. Transp. Res. Part B 46 (9), 1159–1176.
Haddad, J., Mirkin, B., 2017. Coordinated distributed adaptive perimeter control for large-scale urban road networks. Transp. Res. Part C 77, 495–515.
Haddad, J., Ramezani, M., Geroliminis, N., 2013. Cooperative traffic control of a mixed network with two urban regions and a freeway. Transp. Res. Part B 54, 17–36.
Haddad, J., Shraiber, A., 2014. Robust perimeter control design for an urban region. Transp. Res. Part B 68, 315–332.
Hu, J., Park, B., Leeb, Y.-J., 2015. Coordinated transit signal priority supporting transit progression under connected vehicle technology. Transp. Res. Part C 55, 393–408.
Ji, Y., Daamen, W., Hoogendoorn, S., Hoogendoorn-Lanser, S., Qian, X., 2010. Macroscopic fundamental diagram: investigating its shape using simulation data. Transp. Res. Rec. 2161, 42–48.
Ji, Y., Geroliminis, N., 2012. On the spatial partitioning of urban transportation networks. Transp. Res. 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. Transp. Res. Part B 46 (10), 1393–1403.
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.
Kouvelas, A., Lioris, J., Fayazi, S., Varaiya, P., 2014. Maximum pressure controller for stabilizing queues in signalized arterial networks. Transp. Res. Rec. 2421, 133–141.
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.
Lamotte, R., Geroliminis, N., 2016. The morning commute in urban areas: Insights from theory and simulation. 95th Annual Meeting of the Transportation Research Board, Washington, D.C.
Mahmassani, H., Williams, J., Herman, R., 1987. Performance of urban traffic networks. In: Gartner, N., Wilson, N. (Eds.), 10th International Symposium on Transportation and Traffic Theory, Elsevier, Amsterdam, The Netherlands.
Mahmassani, H.S., Saberi, M., Zockaie, A., 2013. Urban network gridlock: theory, characteristics, and dynamics. Transp. Res. Part C 36, 480–497.
Mazloumian, A., Geroliminis, N., Helbing, D., 2010. The spatial variability of vehicle densities as determinant of urban network capacity. Philos. Trans. R. Soc. A 368 (1928), 4627–4647.
Nesterov, Y., Nemirovskii, A., 1994. Interior-point polynomial algorithms in convex programming. vol. 28 of Studies in Applied Mathematics. SIAM, Philadelphia, PA.
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.
Saberi, M., Mahmassani, H.S., Zockaie, A., 2014. Network capacity, traffic instability, and adaptive driving: findings from simulated urban network experiments. EURO J. Transp. Logist. 1–20. doi:10.1007/s13676-013-0040-2.
Saeedmanesh, M., Geroliminis, N., 2016. Clustering of heterogeneous networks with directional flows based on snake similarities. Transp. Res. Part B 91, 250–269.
Skabardonis, A., Geroliminis, N., 2008. Real-time monitoring and control on signalized arterials. J. Intell. Transp. Syst. Technol., Plann., Oper. 12 (2), 64–74.
Yildirimoglu, M., Geroliminis, N., 2014. Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams. Transp. Res. Part B 70, 186–200.

Please cite this article as: K. Ampountolas et al., Macroscopic modelling and robust control of bi-modal multi-region urban road networks, Transportation Research Part B (2017), http://dx.doi.org/10.1016/j.trb.2017.05.007

Yildirimoglu, M., Ramezani, M., Geroliminis, N., 2015. Equilibrium analysis and route guidance in large-scale networks with MFD dynamics. Transp. Res. Part C 59, 404–420.
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. Transp. Res. Part B 49, 1–23.
Zheng, N., Geroliminis, N., 2013. On the distribution of urban road space for multimodal congested networks. Transp. Res. Part B 57, 326–341.
Ziegler, G.M., 1995. Lectures on Polytopes. vol. 152 of Graduate Texts in Mathematics. Springer Verlag, Berlin.

Please cite this article as: K. Ampountolas et al., Macroscopic modelling and robust control of bi-modal multi-region urban road networks, Transportation Research Part B (2017), http://dx.doi.org/10.1016/j.trb.2017.05.007