JOURNAL OF ADVANCED TRANSPORTATION
J. Adv. Transp. 2016; **50**:1046–1060
Published online 4 July 2016 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/atr.1389

A model-based demand-balancing control for dynamically divided multiple urban subnetworks

Shu Lin[1,2], Qing-Jie Kong[3]* and Qingming Huang[1,2]

[1]*School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing China*
[2]*Key Lab of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing China*
[3]*The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing China*

SUMMARY

Traffic control is an effective and efficient method for the problem of traffic congestion. It is necessary to design a high-level controller to regulate the network traffic demands, because traffic congestion is not only caused by the improper management of the traffic network but also to a great extent caused by excessive network traffic demands. Therefore, we design a demand-balance model predictive controller based on the macroscopic fundamental diagram-based multi-subnetwork model, which can optimize the network traffic mobility and the network traffic throughput by regulating the input traffic flows of the subnetworks. Because the transferring traffic flows among subnetworks are indirectly controlled and coordinated by the demand-balance model predictive controller, the subnetwork division can variate dynamically according to real traffic states, and a global optimality can be achieved for the entire traffic network. The simulation results show the effectiveness of the proposed controller in improving the network traffic throughput. Copyright © 2016 John Wiley & Sons, Ltd.

KEY WORDS: traffic network control; urban road network; macroscopic fundamental diagram; traffic flow equilibrium

1. INTRODUCTION

Traffic congestion produces environment pollution, reduces traveling efficiency, and thus causes economic losses. In order to keep the public roads used in a well-organized way by all the drivers, it is very important to adopt traffic control systems to manage transportation in a good manner. Well-designed traffic control and management strategies are effective methods for regulating traffic behaviors [1–5]. Therefore, to reduce the congestion degree in road networks, effective and efficient road network control policies are very important for a better utilization of existing transportation infrastructures.

A number of coordinated urban network control strategies have already been developed [6–10]. Fixed-time coordinated control strategies make control decisions off-line based on the traffic flow data collected and stored in the past. Traffic-responsive coordinated control strategies can in real time measure the traffic states in the network and adapt the control schemes according to the current measured traffic states. Model-based coordinated control strategies do not only introduce feedback control so as to adjust in real time the control decision according to the current detected traffic states but also predict into the future using a prediction model to make decisions good also in the long run [6, 11–18]. The structures for the coordinated control strategies can be centralized, distributed, or hierarchical. A centralized coordination control strategy optimizes the whole traffic network and

*Correspondence to: Qing-Jie Kong, The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. E-mail: kongqingjie@gmail.com

searches for a global optimal solution for the network. A distributed coordination control strategy allocates the control efforts to each local traffic controller and coordinates the local controllers through information exchange. A hierarchical coordination control strategy divides the overall complex control problem for large-scale systems into multiple levels; on each level, a specific control problem will be solved.

There are already well-known coordinated traffic-responsive control strategies for urban traffic networks. The Split Cycle Offset Optimization Technique (SCOOT) [8, 19] and Sydney Coordinated Adaptive Traffic System (SCATS) [20] are widely used in many big cities around the world [6], for example, SCOOT is used in Beijing, and SCATS is used in Shanghai. They are both dynamic traffic control strategies based on measured current traffic states in distributed, multi-level, hierarchical system structures. It has already been shown that these two systems work effectively in the real traffic world, but these two systems focus more on dynamic intersection controllers and local coordinations that consider only a few neighbor intersections. In the 1980s and 1990s, a number of model-based optimization control strategies based on simple traffic models emerged, for example, Optimization Policies for Adaptive Control (OPAC) [11], PRODYN [12], control of networks by optimization of switchovers (CRONOS) [13], RHODES [14], and Method for the Optimization of Traffic Signals In Online controlled Networks (MOTION) [21], which can forecast the future traffic behavior of the network based on models. With these forecasting models, the control strategies are able to make control decisions to guarantee better performance within an area of the traffic network in a near future. A real test was realized for OPAC in Reston, USA (16 intersections) [22], and for MOTION in the center of Köln-Deutz, Germany (12 intersections). However, the models used in these control approaches are mainly simple traffic models based on the traffic data measured by upstream detectors, which to some extent limit the performance for the future. Coordinated traffic-responsive control strategies that are able to avoid parts of the online computational complexity were also proposed. UTOPIA/SPOT [23] is a hierarchical system with simple local intersection controllers and a central controller for an area of urban networks. The central controller optimizes the control actions for the whole area based on the model of the network. The local controller makes the decision only based on local information, but with a penalty term to guarantee that the local decision is not too far from the central decision. Therefore, UTOPIA/SPOT avoids part of the online computational burden but results in suboptimal solutions. Traffic-responsive urban control (TUC) [18, 24] was proposed for controlling an urban traffic network based on the well-known simple store-and-forward model. TUC designs a feedback regulator off-line based on the store-and-forward model and online derives the traffic signals using a feedback control law by feeding it with real-time measured traffic states. Therefore, the TUC strategy reduces the online computational complexity significantly by moving the time-consuming optimization off-line. Compared with the fixed-time controller, TUC can reduce the total time spent by 20–54% for different scenarios [24]. TUC has been implemented in three cities—Chania, Greece (23 intersections); Southampton, UK (53 intersection); and Munich, Germany (25 intersections)—and has been proven to have good control effects [25]. Model predictive control (MPC)-based traffic network control strategies are proposed in real time, which optimize the overall traffic throughput from a global point of view but also cost more computational effort. In fact, traffic congestion is not only caused by the improper management of the traffic network but also to a great extent caused by excessive network traffic demands (traffic inputs).

Therefore, in this paper, we mainly focus on the demand control for large urban traffic networks, which is a high-level control algorithm through adjusting the traffic input allowed to enter the traffic network. We divide a large urban traffic network into subnetworks and design a demand-balance model predictive controller for these subnetworks by regulating the proportion of the input traffic demands on the boundary of the whole network. The demand-balance model predictive controller is designed based on a macroscopic fundamental diagram (MFD)-based subnetwork model, which is improved from the model proposed by Geroliminis *et al.* [26]. It indirectly coordinates the exchanging traffic flows among the subnetworks and automatically balances the allocation of the vehicles among the subnetworks, so as to maintain the maximum traffic mobility and to achieve the maximum traffic throughput. Because there is no control actions on the border of two adjacent subnetworks, it is possible to divide the subnetworks dynamically according to real traffic states.

The paper is organized as follows. In Section 2, we briefly review the background of MFD. Section 3 illustrates the content of MFD and its equilibriums, and then the MFD-based multi-subnetwork model is shown in Section 4. The demand-balancing model predictive controller is introduced in Section 5. Section 6 is a case study. Finally, Section 7 concludes the paper.

DOI: 10.1002/atr

2. BACKGROUND OF MACROSCOPIC FUNDAMENTAL DIAGRAMS

In 1935, a fundamental diagram (i.e., the flow–density relationship) is proven to exist for a single road by Greenshields *et al*. [27]. In 1969, the fundamental diagram for complex traffic networks was investigated by Godfrey *et al*. [28]. In recent years, convincing empirical results were obtained by Geroliminis and Daganzo [29], which illustrate that there also exists a fundamental diagram-like relationship for entire traffic networks. Thus, the fundamental diagram-like relationship between the network-aggregated vehicle occupancy and traffic flow is named as MFD, or also named as network fundamental diagram (NFD) of traffic networks. The MFDs show the characteristic between the network traffic accumulation (i.e., the number of vehicles in the network) and the network weighted average traffic flow. In the mean time, Daganzo and Geroliminis [30] theoretically analyzed the existence of MFD for networks with many intersections. In addition, Helbing [31] gave another version of analysis for the existence of NFD. In [32], the factors that influence the shape of MFDs are concluded as the structure of traffic networks, the heterogeneity of the networks, the traffic demands, and the control strategies. But, according to Geroliminis and Daganzo [30], the shape of an MFD is an existing characteristic of the network structure and is irrespective of traffic demands. However, the heterogeneity of the networks due to traffic congestion is found by Buisson and Ladier to be responsible to the hysteresis phenomena in MFDs [33].

Because the MFD provides the information on how the average network traffic flow changes with the network traffic density, it can illustrate the traffic mobility within a traffic network. In this regard, the MFD can be used as a performance indicator for the estimation of the mobility of signalized traffic networks under different control strategies [34–36]. Recently, control strategies have been proposed based on MFDs to regulate the aggregated network output flow and to improve the network mobility by adjusting the network inflow rate (i.e., adjusting the number of vehicles entering the network). Some of the control strategies are designed based on the partition considering about the homogeneity of the subnetworks [37–39]. Geroliminis *et al*. [26] proposed an MPC approach for a two-region urban network based on the subnetwork MFD models, and Haddad *et al*. [40] analyzed and proved the stability of the control approach for a two-region urban network. Ekbatani *et al*. [41, 42] proposed an easy-to-implement feedback control strategy for regulating the inflow rate of an urban network to increase the traffic mobility within the region, and in [43], the feedback-based gating control is applied at junctions located further upstream of the protected network to deal with the time delay. Lin *et al*. [44] designed a high-level traffic flow coordinator for a multi-level urban traffic network controller based on the MFD features of the subnetworks. An MFD-based subnetwork model predictive controller is proposed in [45], which can optimize and regulate the exchanging traffic flows among subnetworks to obtain an overall network performance. In [46], an optimal hybrid control problem for a multi-region MFD network is formulated as a mixed-integer nonlinear optimization problem to control the heterogeneous traffic network. In [47], a robust perimeter controller is designed for an urban region with the MFD representation including MFD uncertainty, and in [48], a robust constrained control of uncertain MFD networks is proposed to reformulate the nonlinear MFD model for multi-region MFD to a linear model with parameter uncertainties.

3. MACROSCOPIC FUNDAMENTAL DIAGRAM AND ITS EQUILIBRIA

The MFD illustrates an aggregated feature for traffic networks, which can describe how the aggregated network traffic flow changes with the aggregated network occupancy (or network accumulation) [29, 30], as Figure 1 shows. The network traffic accumulation $A(t)$ refers to the number of vehicles in a traffic network at a time instant t:

$$A(t) = \sum_{i \in L} a_i(t) \tag{1}$$

where $a_i(t)$ is the number of vehicles on link i at time instant t and L is the set of links in the traffic network. The network traffic accumulation can also be replaced by the average traffic density in the traffic network. The aggregated network traffic flow is actually the weighted network traffic

DOI: 10.1002/atr

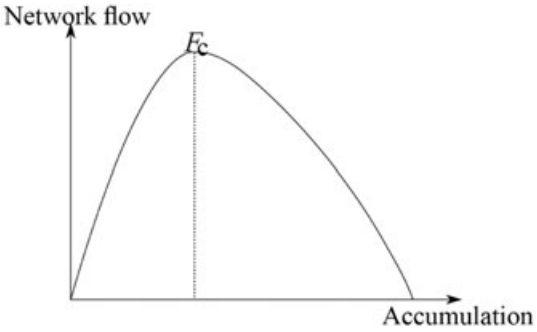

Figure 1. A well-defined traffic network macroscopic fundamental diagram.

flow (i.e., the space mean average flow), which represents the average traffic flow in a traffic network; that is,

$$F(t) = \frac{\sum_{i \in L} l_i \cdot f_i(t)}{\sum_{i \in L} l_i} \tag{2}$$

where $F(t)$ is the weighted average network flow rate, $f_i(t)$ is the traffic flow rate on link i at time instant t, and l_i is the length of link i.

However, a well-defined MFD curve in Figure 1 only exists for an ideal situation; that is, a well-defined MFD curve can only be realized when the traffic flows in a traffic network are always evenly scattered, that is, homogeneous [33]. When the traffic in a traffic network is heterogeneous, hysteresis will occur, especially in the saturated and over-saturated regions (i.e., right-hand-side region of the MFD curve). Therefore, the MFD curve expands to a region instead of a well-defined curve. During the elimination of spillbacks and gridlocks within the network, hysteresis phenomena appear in this region, because of the heterogeneity of the traffic flows in the network.

The traffic flows in a traffic network may reach an equilibrium on the MFD curve, that is, keeping a constant average network flow and a constant network accumulation. The equilibrium can be achieved, if the network output flow equals the network input flow, and the number of vehicles in the traffic network does not vary with time. Traffic flow equilibria exist on the MFD of a traffic network, when the average network input flow is proportional to the weighted average network flow; that is,

$$\overline{f}_{\text{in}}(t) = \frac{1}{r} F(t), \tag{3}$$

which is bounded by $\overline{f}_{\text{in}}(t) \in \left[0, \quad \frac{1}{r} F_{\text{c}}\right]$ and $\text{r} \in (0, 1]$; F_{c} is the critical average network flow on MFD.

If the traffic flows are scattered homogeneously in a traffic network, that is, always evenly scattered within the traffic network, then for any average network input flow $\overline{f}_{\text{in}}(t) \in \left[0, \quad \frac{1}{r} F_{\text{c}}\right]$, there exist two network flow equilibria on the MFD, $(A_{\text{l}}(t), F(t))$ and $(A_{\text{r}}(t), F(t))$, on each side of the MFD curve, which satisfy $\overline{f}_{\text{in}}(t) = \frac{1}{r} F(t)$. But, if the traffic flows are scattered heterogeneously in a traffic network, that is, unevenly scattered within the traffic network, then there exists only one network flow equilibrium on the MFD, $(A_{\text{l}}(t), F(t))$, on the left-hand side of the MFD curve, which satisfies $\overline{f}_{\text{in}}(t) = \frac{1}{r} F(t)$. For more details, please see [49].

4. MACROSCOPIC FUNDAMENTAL DIAGRAM-BASED MULTI-SUBNETWORK MODEL

Define a large urban traffic network with $\tilde{N}$ subnetworks; that is, the set of subnetworks is $\text{N} = \{1, 2, \ldots, \tilde{N}\}$. For each subnetwork, we assume that the subnetwork is simple enough to have a well-defined

DOI: 10.1002/atr

MFD. Therefore, we are able to establish a subnetwork model by mapping the network output flow through the number of vehicles in the network based on the MFD relationship. In each subnetwork, the vehicles are conceptually divided into two parts, the vehicles intending to stay inside the subnetwork and the vehicles intending to leave the subnetwork for neighboring subnetworks. These two parts are calculated separately in the evolution of the MFD-based subnetwork model. The subnetwork models are updated every time interval T_s.

Let us take subnetwork i for example (Figure 2) the number of vehicles in subnetwork i at time step k is separated by its traveling directions, that is, by the destination subnetworks for the traffic flows. We represent the number of vehicles in subnetwork i at time step k intending to stay in subnetwork i as $n_{ii}(k)$ and the number of vehicles in subnetwork i at time step k intending to go to subnetwork j as $n_{ij}(k)$, where $j \in N_i$ and N_i is the set of subnetworks in the neighborhood of subnetwork i. They are updated by the following equations:

$$\begin{aligned} n_{ii}(k+1) &= n_{ii}(k) + T_s \left[d_{ii}(k) + \sum_{j \in N_i} p_{ii}(k) \cdot M_{ji}(k) - M_{ii}(k) \right] \qquad i \in N; \\ n_{il}(k+1) &= n_{il}(k) + T_s \left[d_{il}(k) + \sum_{j \in N_i} p_{il}(k) \cdot M_{ji}(k) \right. \end{aligned} \tag{4}$$

$$\left. - M_{il}(k) \right] \qquad i \in N,\ l \in N_i \tag{5}$$

where similarly the traffic demand for subnetwork i is also expressed separately by $d_{il}(k)$ and $d_{ii}(k)$, the traffic demand for subnetwork i at time step k intending to go to subnetwork l ($l \in N_i$), and the traffic demand intending to stay in subnetwork i at time step k; $M_{ji}(k)$ is the traffic flow getting into subnetwork i from subnetwork j ($j \in N_i$ belongs to the neighborhood of subnetwork i) at time step k, and $M_{il}(k)$ is the traffic flow leaving subnetwork i for subnetwork l ($l \in N_i \cup i$, if $l = i$, it represents the traffic flow in subnetwork i intending to leave from subnetwork i at time step k); $p_{il}(k)$ represents the one-step transition probability of the traffic flow transferring from subnetwork i to subnetwork l at time step k (i, $l \in N$), which is the proportion of the traffic flow in the total network traffic flow intending to go to different destinations. Therefore, we have

$$\sum_{l \in N} p_{il}(k) = 1 \quad i \in N, \tag{6}$$

which illustrates that the sum of the one-step transition probabilities for the traffic flows in subnetwork i equals 1, that is, the traffic flows intending to stay in subnetwork i and to leave for the subnetworks in its neighborhood. Note that if subnetwork i and subnetwork l are physically not connected, then $p_{il}(k) = 0$. If subscript $i = l$, then the one-step transition probability $p_{ii}(k)$ stands for the proportion of the traffic flow in the network flow intending to keep on staying in subnetwork i at time step k. For all

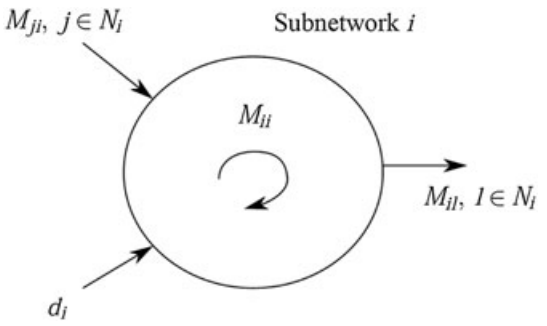

Figure 2. Illustration for the traffic flows of subnetwork i.

DOI: 10.1002/atr

subnetworks $i \in N$, the transition probabilities can be written in a transition matrix, which provides the transition probabilities among all subnetworks. The transition matrix has similar properties as Markov transition matrix, that is, $0 \le p_{ij}(k) \le 1$, $\sum_{j \in N} p_{ij}(k) = 1 (i \in N)$. However, the evolution of the traffic model is a non-strict Markov process, in which the current traffic states (number of vehicles of subnetworks) depend on the traffic states of the previous time step, the one-step transition probabilities, the network demand flows, and the subnetwork transferring flows.

The network traffic flow in subnetwork i can be obtained through the MFD function of the subnetwork, which provides the mapping relation between the network accumulation $n_i(k)$ and the network flow $M_i(k)$; that is,

$$M_i(k) = G_i(n_i(k)) \quad i \in N. \tag{7}$$

By multiplying the one-step transition probability, the transferring traffic flows can be obtained from the network traffic flow of subnetwork i as

$$M_{ii}(k) = p_{ii}(k) \cdot M_i(k) \quad i \in N; \tag{8}$$

$$M_{il}(k) = p_{il}(k) \cdot M_i(k) \quad i \in N,\ l \in N_i, \tag{9}$$

which means that the network traffic flow of subnetwork i is divided into the traffic flow intending to stay in the subnetwork and the traffic flows intending to leave for neighboring subnetworks.

The number of vehicles in subnetwork i, that is, the traffic accumulation of the subnetwork, can be updated by

$$n_i(k) = n_{ii}(k) + \sum_{l \in N_i} n_{il}(k) \quad i \in N, \tag{10}$$

which is the sum of the number of vehicles intending to stay in subnetwork i and the number of vehicles intending to leave subnetwork i for the neighboring subnetworks.

The one-step transition probabilities can be estimated by

$$p_{ii}(k) = \frac{n_{ii}(k)}{n_i(k)}, \quad p_{il}(k) = \frac{n_{il}(k)}{n_i(k)} \quad i \in N,\ l \in N_i \tag{11}$$

They can also be obtained through other methods, for example, O-D information analysis. In the previous definition, the transition probabilities are time dependent because of the variation of network traffic states (the number of vehicles in subnetworks). The transition probabilities will not become static, until the traffic flows in subnetworks reach their equilibria (i.e., steady states—the traffic flows and the number of vehicles stay constant).

Similarly, the traffic demand for subnetwork i can be calculated by

$$d_i(k) = d_{ii}(k) + \sum_{l \in N_i} d_{il}(k) \quad i \in N, \tag{12}$$

which is the sum of the traffic demand intending to stay in subnetwork i and the traffic demands intending to leave subnetwork i for the neighboring subnetworks. In addition, the transferring traffic flow getting into subnetwork i from neighboring subnetworks is

$$t_i(k) = \sum_{j \in N_i} M_{ji}(k) \quad i \in N. \tag{13}$$

Then, the input traffic flow for subnetwork i is

DOI: 10.1002/atr

$$f_{\text{in},i}(k) = d_i(k) + t_i(k) \quad i \in N \tag{14}$$

which contains the total traffic demand of subnetwork i and the total transferring traffic flow for subnetwork i from other subnetworks.

5. DEMAND-BALANCE MODEL PREDICTIVE CONTROL COORDINATING SUBNETWORKS

The subnetwork model described in Section 4 is adopted as the control model for predicting the future traffic states. The control time interval of the model predictive controller is defined as T_c, which is an integer multiple of the simulation time interval T_s of the subnetwork model; that is,

$$T_c = I \cdot T_s, \tag{15}$$

where I is an integer. The control time step is k_c, and the simulation time step is k_s. Let the prediction horizon be N_p, and then, the online optimization problem of MPC can be expressed as

$$\min_{\boldsymbol{r}(k_c)} J = \min_{\boldsymbol{r}(k_c)} J\left(\widehat{\mathbf{M}}(k_s), \mathbf{d}(k_s)\right) = -\sum_{k_s=1}^{N_p I} \sum_{i \in N} M_{ii}(k_s)$$

$$\begin{aligned} &\text{s.t.} \quad \text{Subnetwork model;} \\ &\mathbf{d}(ks) = \mathbf{r}(kc)^{\mathrm{T}} \cdot \mathbf{H}^{\mathrm{T}} \cdot \widetilde{\mathbf{d}}(ks) \\ &n_{ij}(k_s) \geq 0, \quad i,j \in N, k_s = 1, 2, \ldots, N_p I \\ &M_{ij}(k_s) \geq 0, \quad i,j \in N, k_s = 1, 2, \ldots, N_p I \\ &0 \leq r_i(k_c) \leq 1 \\ &i \in N, k_s = k_c, k_c + 1, \ldots, k_c + N_p - 1 \end{aligned} \tag{16}$$

where the control objective corresponds to the accumulated traffic flows that leave all the subnetworks and is subject to the model constraint, the control action constraint, the lower and upper bounds for the control variables, traffic accumulation in the subnetworks, and the output traffic flows. The optimization tends to minimize the objective function so as to maximize the accumulated traffic flows that leave all the subnetworks in the prediction horizon.

The objective function is calculated via the subnetwork model in Section 4. For the given prediction horizon N_p, the estimated traffic flows leaving all the subnetworks at time step k_s are expressed by the following vector:

$$\begin{aligned} \widehat{\mathbf{M}}(k_s + k|k_s) = [&\widehat{M}_{11}(k_s + k|k_s)\ \widehat{M}_{22}(k_s + k|k_s) \cdots \\ &\widehat{M}_{NN}(k_s + k|k_s)]^{\mathrm{T}}, \quad k = 1, 2, \ldots, N_p I \end{aligned} \tag{17}$$

where $\widehat{\mathbf{M}}(k_s + k|k_s)$ represents the traffic flow leaving all subnetworks at the kth simulation step in the future counted from the current simulation time step k_s and $\widehat{M}_{ii}(k_s + k|k_s)$ refers to the traffic flow leaving subnetwork i. The leaving traffic flows of all the subnetworks for the prediction horizon are then expressed by

$$\begin{aligned} \widehat{\mathbf{M}}(k_s) = [&\widehat{\mathbf{M}}(k_s + 1|k_s)\ \widehat{\mathbf{M}}(k_s + 2|k_s) \cdots \\ &\widehat{\mathbf{M}}(k_s + N_p I|k_s)]^{\mathrm{T}} \end{aligned} \tag{18}$$

Because the control variables are the proportion of input traffic demands that can enter the subnetworks, the input traffic demand $\mathbf{d}(k_s)$ in the subnetwork model (Section 4) is substituted by the supplied traffic demand $\widetilde{\mathbf{d}}(k_s)$ multiplied with the metering control action $\mathbf{r}(k_c)$, that is, $\mathbf{d}(k_s) =$

DOI: 10.1002/atr

$\mathbf{r}(k_c)^T \cdot \mathbf{H}^T \cdot \widetilde{\mathbf{d}}(k_s)$, where $\mathbf{H}^T$ is a coefficient matrix to make up the dimension difference between $\mathbf{r}(k_c)$ and $\widetilde{\mathbf{d}}(k_s)$. It is to say that the original supplied traffic demands for the subnetworks are regulated by the metering control actions on boundaries of the subnetworks. The input traffic demand of all the subnetworks for the prediction horizon at step k_s is defined as

$$\begin{aligned}\mathbf{d}(k_s+k|k_s) = [&d_1(k_s+k|k_s)\ d_2(k_s+k|k_s)\cdots \\ &\mathbf{d}_N(k_s+k|k_s)]^T, \quad k=0,1,\ldots,N_pI-1\end{aligned} \tag{19}$$

$$\begin{aligned}\mathbf{d}(k_s) = [&\mathbf{d}(k_s+0|k_s)\ \mathbf{d}(k_s+1|k_s)\cdots \\ &\mathbf{d}(k_s+N_pI-1|k_s)]^T\end{aligned} \tag{20}$$

Similarly, the supplied traffic demand of all the subnetworks for the future at step k_s is

$$\begin{aligned}\widetilde{\mathbf{d}}(k_s+k|k_s) = [&\widetilde{d}_1(k_s+k|k_s)\ \widetilde{d}_2(k_s+k|k_s)\cdots \\ &\widetilde{d}_N(k_s+k|k_s)]^T, \quad k=0,1,\ldots,N_pI-1\end{aligned} \tag{21}$$

$$\begin{aligned}\widetilde{\mathbf{d}}(k_s) = [&\widetilde{\mathbf{d}}(k_s+0|k_s)\ \widetilde{\mathbf{d}}(k_s+1|k_s)\cdots \\ &\widetilde{\mathbf{d}}(k_s+N_pI-1|k_s)]^T,\end{aligned} \tag{22}$$

and the future traffic control action of all the subnetworks at control step k_c is

$$\begin{aligned}\mathbf{r}(k_c+k|k_c = [&r_1(k_c+k|k_c)\ r_2(k_c+k|k_c) \\ &\cdots r_N(k_c+k|k_c)]^T, k=0,1,\ldots,N_p-1\end{aligned} \tag{23}$$

$$\begin{aligned}\mathbf{r}(k_c) = [&\mathbf{r}(k_c+0|k_c)\ \mathbf{r}(k_c+1|k_c)\cdots \\ &\mathbf{r}(k_c+N_p-1|k_c)]^T\end{aligned} \tag{24}$$

where $\mathbf{r}(k_c+k|k_c)$ denotes the control actions for all the subnetworks at the kth control step in the future counted from the current control time step k_c and $r_i(k_c+k|k_c)$ is the control action for subnetwork i.

By deriving the optimal control input $\mathbf{r}^*(k_c)$ from the aforementioned online optimization, the first sample of the optimal control sequence, that is,

$$\begin{aligned}\mathbf{r}^*(k_c) = [&\mathbf{r}^*(k_c+0|k_c)\ \mathbf{r}^*(k_c+1|k_c)\cdots \\ &\mathbf{r}^*(k_c+N_p-1|k_c)]^T\end{aligned} \tag{25}$$

where

$$\begin{aligned}\mathbf{r}^*(k_c+k|k_c) = [&r_1^*(k_c+k|k_c)\ r_2^*(k_c+k|k_c) \\ &\cdots r_N^*(k_c+k|k_c)]^T \quad k=0,1,\ldots,N_p-1\end{aligned} \tag{26}$$

is given back to the traffic network and is implemented. When arriving to the next control step k_c+1, the whole time horizon is shifted one step forward, and the optimization over the new prediction horizon starts over again, based on the prediction model that is fed with the measured traffic states, that is, the traffic accumulation of the subnetworks, obtained from the real traffic field. This rolling horizon scheme closes the control loop, enables the system to obtain feedback from the real traffic network, and

DOI: 10.1002/atr

makes the model predictive controller adaptive to the uncertainties and disturbances of the traffic environment [50, 51].

The objective function in (16) can be also substituted by

$$\min_{\boldsymbol{r}(k_c)} J = -\alpha \sum_{k_s=1}^{N_p I} \sum_{i \in N} M_{ii}(k_s) + \beta \sum_{k_c=1}^{N_p-1} \sum_{i \in N} [r_i(k_c) - r_i(k_c - 1)] \tag{27}$$

which means maximizing the total network output flow and minimizing the accumulative variations between the successive control actions. In the objective function, α and β are the coefficients, which represent the proportion between the total network output and the total control variation. Thus, we define $\alpha+\beta=1$ with $0\leq\alpha\leq1$ and $0\leq\beta\leq1$.

6. CASE STUDIES

The demand-balancing model predictive controller is designed to regulate the input traffic flows to an urban region with two subnetworks, so as to maintain the maximum throughput flow of the entire traffic network. The simulations are run for different scenarios to evaluate the control effect of the designed controller.

The traffic network has two subnetworks, S_1 and S_2, and they both have the same MFD. The MFD can be fitted by $G_i(n_i(t)) = 1.4877 \cdot 10^{-7} \cdot {n_i}^3 - 2.9815 \cdot 10^{-3} \cdot {n_i}^2 + 15.0912 \cdot n_i$, $i=1,\ 2$, $n_{1,\mathrm{cr}} = n_{2,\mathrm{cr}} = 3400$ [veh], $G_1(n_{1,\mathrm{cr}}) = G_2(n_{2,\mathrm{cr}}) = 2.268 \cdot 10^4$ [veh/hour], $n_{1,\mathrm{jam}} = n_{2,\mathrm{jam}} = 10\,000$ [veh]. The parameters for the designed demand-balancing model predictive controller are as follows: the simulation time interval is $T_s = 1$ [second], the control time interval is $T_c = 120$ [second], the prediction horizon is $N_p = 10$, and the control time horizon is $N_c = 2$. The objective function is first selected the same as in (16), that is, maximizing the total network output flows.

As Figure 3 shows, the results are provided for scenario A1, in which the initial accumulations for S_1 and S_2 are $n_1(0) = 1000$ [veh] and $n_2(0) = 500$ [veh], and the variations of the supplied traffic demands for the subnetworks are shown in Figure 3(a). The initial accumulations of two subnetworks are both on the left-hand side of the MFD curve, which means that the two subnetworks are initially uncongested. In Figure 3(f), the input traffic demands change with the supplied traffic demands, and

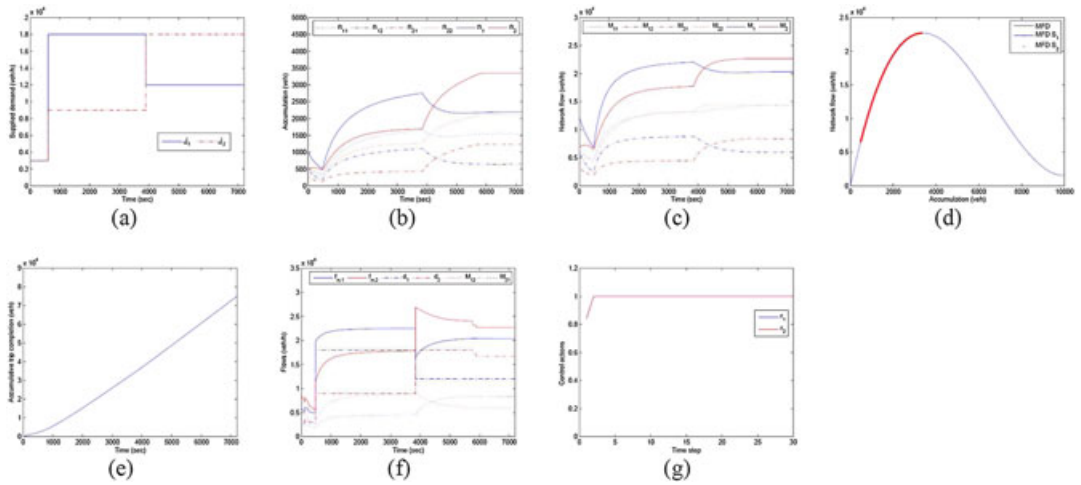

Figure 3. The demand-balancing model predictive control results for scenario A1. (a) The supplied traffic demands, (b) the evolution for network accumulations, (c) the evolution for network flows, (d) the subnetwork macroscopic fundamental diagrams (MFDs), (e) the accumulative trip completion, (f) the input subnetwork traffic flows, and (g) the control actions of the subnetworks.

DOI: 10.1002/atr

the input traffic flows for the subnetworks vary accordingly. To adapt to the input traffic flows, the network accumulations and the network average flows change accordingly and reach their equilibria in Figure 3(b and c), which is a validation for the conclusion in [49]. As Figure 3(g) shows, because the subnetworks are initially uncongested, the demand-balancing model predictive controller uses very little control effort to stabilize the two-subnetwork traffic system.

As Figure 4 shows, the results are provided for scenario A2, in which the supplied traffic demands for the subnetworks are the same as in scenario A1, as in Figure 3(a), but the initial accumulations for S_1 and S_2 are changed into $n_1(0) = 7000$ [veh] and $n_2(0) = 6000$ [veh]. The initial accumulations of two subnetworks are both on the right-hand side of the MFD curve, which means that the two subnetworks are initially congested. The evolution for network accumulations is shown in Figure 4(a), for subnetworks 1 and 2. It shows that the number of vehicles in the subnetworks drops rapidly from the congested region to the uncongested region and then increases and decreases because of the increment of the input subnetwork flows. This can be also seen in Figure 4(c); the traffic state trajectories of the subnetworks move from the congested part (right-hand side) of the MFDs to the uncongested part (left-hand side) of the MFDs. Similarly, in Figure 4(b), under the control of the demand-balancing model predictive controller, the network traffic flows of the two subnetworks first increase and then decrease, getting from the congested region to the uncongested region, and then they fluctuate in the uncongested region (i.e., the left-hand side of the MFDs) because of the variation of the input subnetwork flows. The control actions (i.e., the metering for the input traffic demands) are shown in Figure 4 (f), which illustrates that the input traffic demands for subnetworks are restricted at the beginning of the simulation, so as to allow the subnetworks to recover from the congested status. Although the subnetworks are initially congested, the demand-balancing model predictive controller is able to stabilize the subnetworks to their equilibria by regulating the proportion of the input network demands.

As Figure 5 shows, the results are provided for scenario B1, in which the initial accumulations for S_1 and S_2 are $n_1(0) = 7000$ [veh] and $n_2(0) = 8000$ [veh], and the supplied traffic demands for the subnetworks are constant as shown in Figure 5(a). The initial accumulations of the two subnetworks are both on the right-hand side of the MFD curve; that is, the two subnetworks are initially congested. In Figure 5(b), the network accumulation of subnetwork 1 reduces to the uncongested region, but the

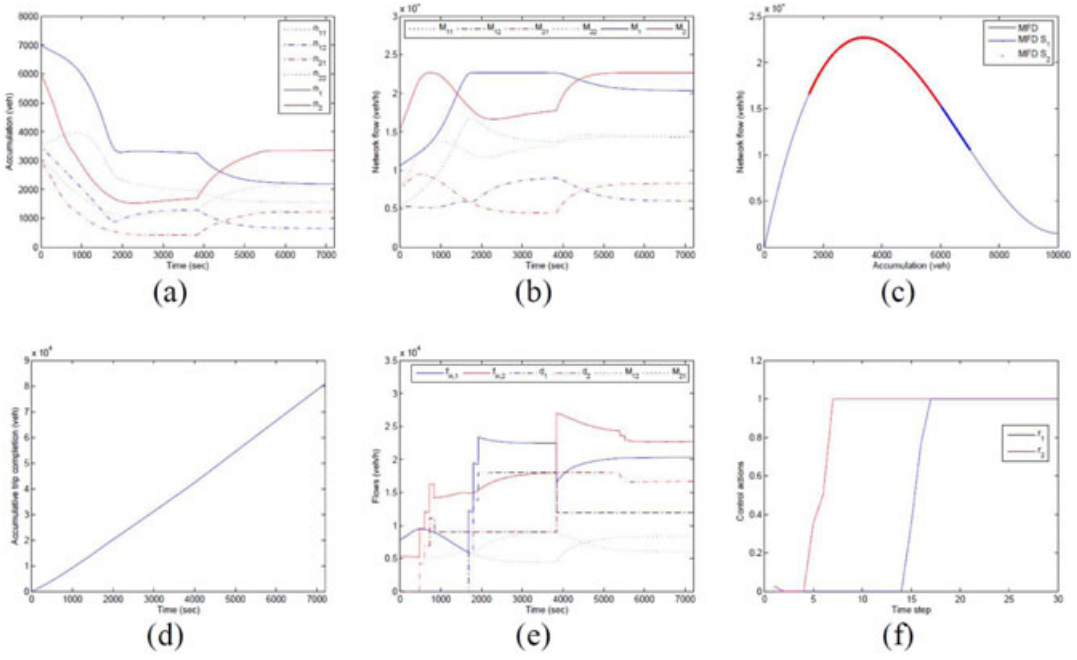

Figure 4. The demand-balancing model predictive control results for scenario A2. (a) The evolution for network accumulations, (b) the evolution for network flows, (c) the subnetwork macroscopic fundamental diagrams (MFDs), (d) the accumulative trip completion, (e) the input subnetwork traffic flows, and (f) the control actions of the subnetworks.

DOI: 10.1002/atr

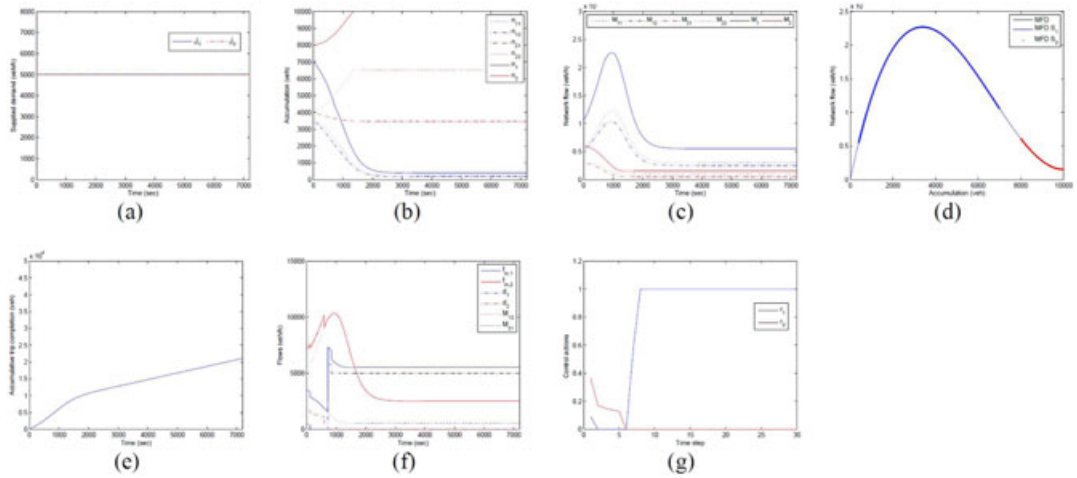

Figure 5. The demand-balancing model predictive control results for scenario B1. (a) The supplied traffic demands, (b) the evolution for network accumulations, (c) the evolution for network flows, (d) the subnetwork macroscopic fundamental diagrams (MFDs), (e) the accumulative trip completion, (f) the input subnetwork traffic flows, and (g) the control actions of the subnetworks.

network accumulation of subnetwork 2 deteriorates until it becomes totally congested. It means that the network controller sacrifices the performance of the more congested subnetwork S2, to guarantee the performance of S1 and thus to achieve a higher overall performance of the entire system. The same conclusion can also be drawn from Figure 5(c, d, and f); the demand-balancing model predictive controller can only take control actions to stabilize subnetwork 1 (adjusting the traffic states of subnetwork 1 to uncongested region and keeping its equilibrium) but cannot stabilize subnetwork 2 (the traffic in subnetwork 2 falls into complete congestion). To prevent subnetwork 2 from getting into congestion, we should satisfy $f_{\text{in},2} < f_{\text{out},2}$. According to (14), we have $M_{12} < F_2 = M_2$. Thus, we change the initial accumulation value to $n_2(0) = 6000$ [veh], so as to satisfy this condition and make the system sterilizable.

Therefore, for scenarios B2 and B3, the supplied traffic demands for the two subnetworks are the same as shown in Figure 5(a), but the initial accumulations for S_1 and S_2 are adapted to $n_1(0) = 7000$ [veh] and $n_2(0) = 6000$ [veh]. For the two scenarios, we adopt different control objectives for the demand-balancing model predictive controller, (16) for scenario B2 and (27) for scenario B3. In scenario B2, the controller tends to maximize the total network output flow. As Figure 6 shows, the two subnetworks can return to the uncongested region under the controller and stabilize on their equilibria. In scenario B3, the controller not only considers maximizing the total network output flow but also requires minimizing the variations of the control actions. As Figure 7 shows, the two subnetworks are also successfully stabilized by the controller. But it is obvious that the control actions become smoother as in Figure 7(f), which results in a smoother network input flows in Figure 7(e). By comparing the accumulative trip completion for scenarios B2 and B3 in Figure 7(d), the two controllers obtain almost the same results.

In Figure 8, the simulation setup is the same as in scenario A1, but with added disturbances in the supplied traffic demands. We add random errors in the supplied traffic demands in Figure 3 to simulate the disturbances in traffic demand estimation. The results show that the controller can adjust the control inputs to adapt to the uncertainties in supplied traffic demands, and the control actions for the two subnetworks can cooperate with each other to maintain the performance of the whole system.

By regulating the input traffic flows, the demand-balancing model predictive controller overcomes the drawback of the model predictive controller taking the subnetwork exchanging traffic flows as control measures, which is aroused by the manually created queues on the edges among subnetworks. It can utilize the capacity of the subnetworks to the maximum extent and keep the optimality of the derived control schemes. The demand-balancing controller implicitly regulates the transfer of traffic flows among subnetworks through the global optimization of the whole network, which enables it to adopt dynamic methods to divide the traffic network into subnetworks. A large network can be divided

DOI: 10.1002/atr

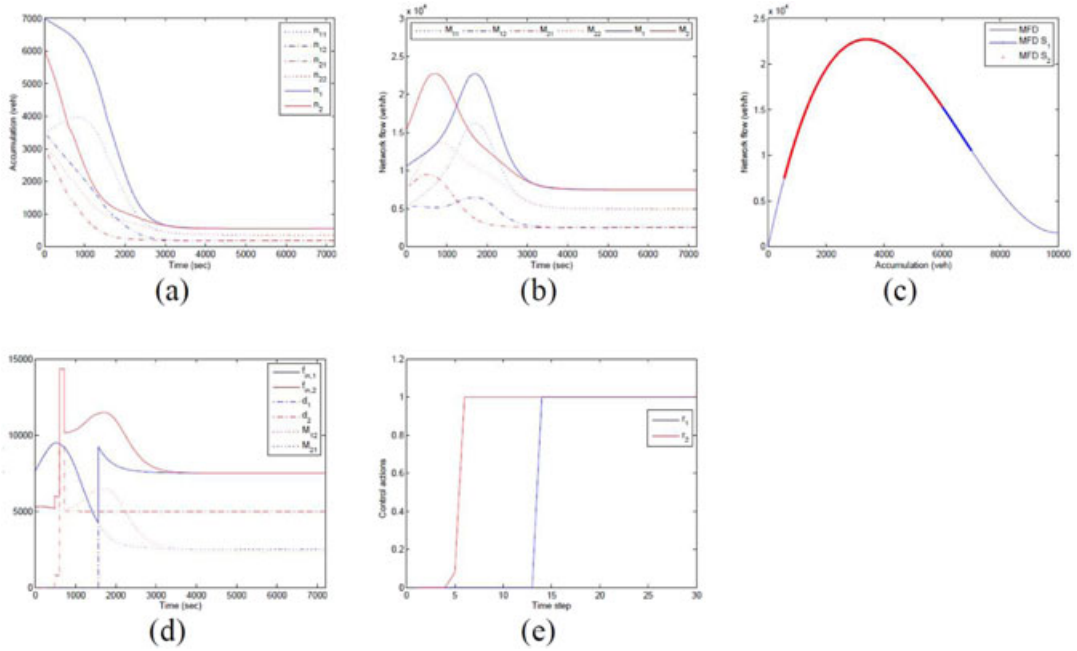

Figure 6. The demand-balancing model predictive control results for scenario B2. (a) The evolution for network accumulations, (b) the evolution for network flows, (c) the subnetwork macroscopic fundamental diagrams (MFDs), (d) the input subnetwork traffic flows, and (e) the control actions of the subnetworks.

into smaller subnetworks according to the physical structure of the network and the congestion degree of the roads as in [52].

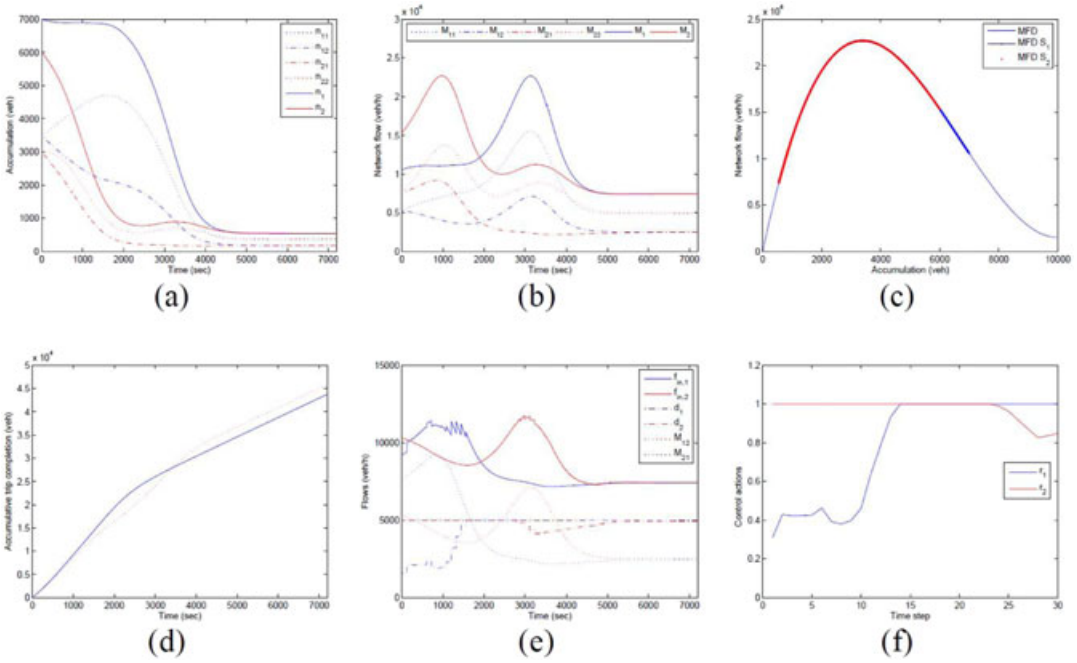

Figure 7. The demand-balancing model predictive control results for scenario B3. (a) The evolution for network accumulations, (b) the evolution for network flows, (c) the subnetwork macroscopic fundamental diagrams (MFDs), (d) the comparison for the accumulative trip completion, (e) the input subnetwork traffic flows, and (f) the control actions of the subnetworks.

DOI: 10.1002/atr

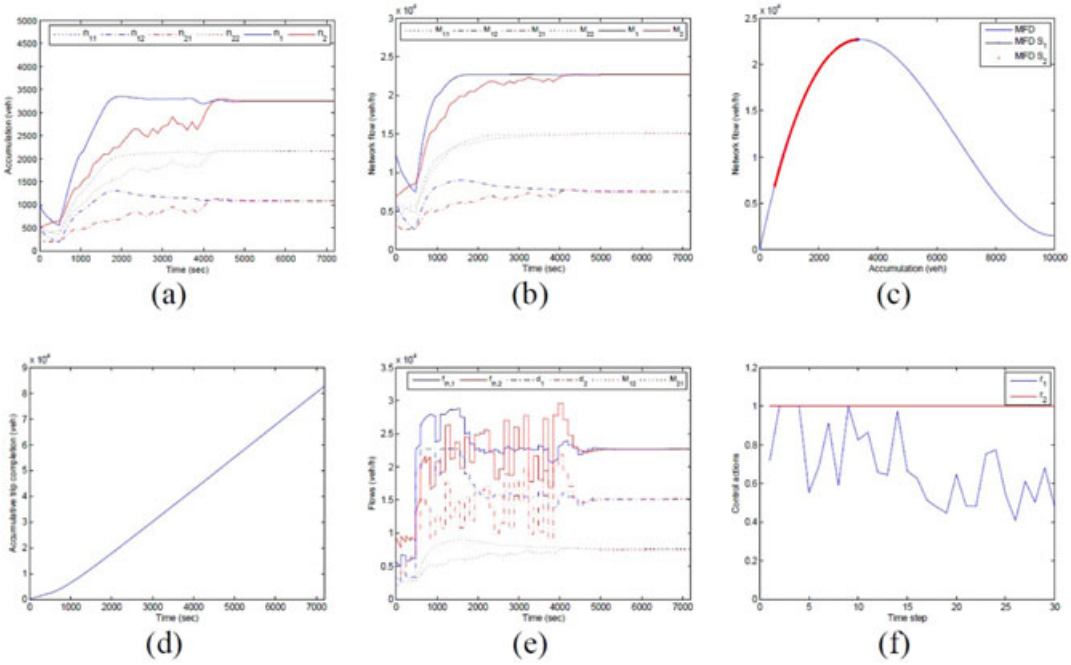

Figure 8. The demand-balancing model predictive control results for scenario A1 with demand disturbances. (a) The evolution for network accumulations, (b) the evolution for network flows, (c) the subnetwork macroscopic fundamental diagrams (MFDs), (d) the comparison for the accumulative trip completion, (e) the input subnetwork traffic flows, and (f) the control actions of the subnetworks.

7. CONCLUSIONS

On the basis of the MFD-based multi-subnetwork model, a high-level MPC strategy is proposed for adjusting the network input traffic demands, so as to balance the traffic densities of subnetworks and to maximize the network traffic throughput. The proposed demand-balancing model predictive controller can be adopted as a high-level controller for a hierarchical urban traffic control strategy.

The proposed demand-balancing model predictive controller is simulated on a traffic network with two subnetworks for different traffic scenarios. The simulation results show that the demand-balancing model predictive controller is able to stabilize the network traffic states of the subnetworks to their analytical equilibria. If the condition is satisfied, the demand-balancing model predictive controller can draw the subnetworks from the congested region to the uncongested region and can maintain the subnetwork traffic states around the maximum value of the network flow by regulating the network input flows. In addition, the demand-balancing model predictive controller indirectly controls the exchanging traffic flows among subnetworks; does not create queues on the boundary between subnetworks, which allows dynamic network division according to real-time varied traffic states; and maintains the global optimality for the entire traffic network.

In the future, multi-level MPC algorithms will be further investigated with the application of the demand-balancing MPC as the high-level controller, and experiments will be carried out in microscopic urban traffic simulation environments.

ACKNOWLEDGGEMENTS

The research is supported by the National Science Foundation of China (61203169, 61473288, 61433002, 61025011, and 61332016), Chinese International Cooperation Project of National Science Committee (grant no. 71361130012), the Beijing Natural Science Foundation (grant no. 4142055), the Early Career Development Award of SKLMCCS, and the European COST Action TU1102.

DOI: 10.1002/atr

REFERENCES

1. Hegyi A, De Schutter B Hellendoorn H. Model predictive control for optimal coordination of ramp metering and variable speed limits. *Transportation Research Part C: Emerging Technologies* 2005; **13**(3): 185–209.
2. Cong Z, De Schutter B Babuška R. Ant colony routing algorithm for freeway networks. *Transportation Research Part C* 2013; **37**: 1–19.
3. Yuan Y, Van Lint J, Wilson RE, van Wageningen-Kessels F Hoogendoorn SP. Real-time Lagrangian traffic state estimator for freeways. *IEEE Transactions on Intelligent Transportation Systems* 2012; **13**(1): 59–70.
4. Burger M, Van Den Berg M, Hegyi A, De Schutter B Hellendoorn J. Considerations for model-based traffic control. *Transportation Research Part C: Emerging Technologies* 2013; **35**: 1–19.
5. Li Z, Schonfeld P. Hybrid simulated annealing and genetic algorithm for optimizing arterial signal timings under oversaturated traffic conditions. *Journal of Advanced Transportation* 2015; **49**(1): 153–170.
6. Papageorgiou M, Diakaki C, Dinopoulou V, Kotsialos A Wang Y. Review of road traffic control strategies. *Proceeding of the IEEE* 2003; **91**(12): 2043–2067.
7. D. Hale, Traffic Network Study Tool, TRANSYT-7F, United States Version, McTrans Center in the University of Florida, 2005.
8. Robertson D, Bretherton R. Optimizing networks of traffic signals in real time—the SCOOT method. *IEEE Transactions on Vehicular Technology* 1991; **40**(1): 11–15.
9. Zhou H, Bouyekhf R Moudni AEL. Constrained h control of urban transportation network. *Journal of Advanced Transportation* 2015; **49**(3): 434–456.
10. Yang X, Lu YC Chang G-L. Exploratory analysis of an optimal variable speed control system for a recurrently congested freeway bottleneck. *Journal of Advanced Transportation* 2015; **49**(2): 195–209.
11. Gartner N. OPAC: A demand-responsive strategy for traffic signal control. Transportation Research Record. No. 906, 1983; pp. 75–81.
12. J. Farges, J. Henry, and J. Tufal, The PRODYN real-time traffic algorithm, in Proc. of the 4th IFAC Symp. Transportation Systems, 1983: 307–312.
13. F. Boillot, J. Blosseville, J. Lesort, V. Motyka, M. Papageorgiou, and S. Sellam, Optimal signal control of urban traffic networks, in Proc. of the 6th Conference on Road Traffic Monitoring and Control, London, England, 1992: 75–79.
14. Sen S, Head K. Controlled optimization of phases at an intersection. *Transportation Science* 1997; **31**(1): 5–17.
15. Dotoli M, Fanti MP Meloni C. A signal timing plan formulation for urban traffic control. *Control Engineering Practice* 2006; **14**(11): 1297–1311.
16. van den Berg M, Hegyi A, De Schutter B Hellendoorn J. Integrated traffic control for mixed urban and freeway networks: a model predictive control approach. *European Journal of Transport and Infrastructure Research* 2007; **7**(3): 223–250.
17. Hegyi A, De Schutter B Hellendoorn J. Model predictive control for optimal coordination of ramp metering and variable speed limits. *Transportation Research Part C: Emerging Technologies* 2005; **13**(3): 185–209.
18. Aboudolas K, Papageorgiou M Kosmatopoulos E. Store-and-forward based methods for the signal control problem in large-scale congested urban road networks. *Transportation Research Part C: Emerging Technologies* 2009; **17**(2): 163–174.
19. D. Bretherton, M. Bodger, and N. Baber, SCOOT—the future urban traffic control, in 12th IEEE International Conference on Road Transport Information and Control, London, UK, 2004, pp. 301–306.
20. P. Lowrie, The Sydney coordinated adaptive traffic system: principles, methodology, algorithms, in International Conference on Road Traffic Signalling, 1982, pp. 67–70.
21. C. Bielefeldt and F. Busch, MOTION—a new on-line traffic signal network control system, in 7th International Conference on Road Traffic Monitoring and Control, London, UK, 1994, pp. 55–59.
22. N. H. Gartner, F. J. Pooran, and C. M. Andrews, Implementation of the OPAC adaptive control strategy in a traffic signal network, in Proc. of the 2001 IEEE International Intelligent Transportation Systems Conference, Oakland, CA USA, 2001, pp. 195–200.
23. V. Mauro and C. Di Taranto, Utopia, in 2nd IFAC-IFIP-IFORS Symposium on Traffic Control and Transportation Systems, 1989, pp. 575–597.
24. Diakaki C, Papageorgiou M Aboudolas K. A multivariable regulator approach to traffic-responsive network-wide signal control. *Control Engineering Practice* 2002; **10**(2): 183–195.
25. Kosmatopoulos E, Papageorgiou M, Bielefeldt C, et al. International comparative field evaluation of a traffic-responsive signal control strategy in three cities. *Transportation Research Part A: Policy and Practice* 2006; **40** (5): 399–413.
26. Geroliminis N, Haddad J Ramezani M. Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: a model predictive approach. *IEEE Transactions on Intelligent Transportation Systems* 2012in press. DOI:10.1109/IITS.2012.2216877.
27. Greenshields B. A study of traffic capacity. *Highway Research Board Proceedings* 1935; **14**: 448–477.
28. Godfrey J. The mechanism of a road network. *Traffic Engineering and Control* 1969; **11**: 323–327.
29. Geroliminis N, Daganzo C. Existence of urban-scale macroscopic fundamental diagrams: some experimental findings. *Transportation Research Part B: Methodological* 2008; **42**(9): 759–770.
30. Daganzo C, Geroliminis N. An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B: Methodological* 2008; **42**(9): 771–781.

31. Helbing D. Derivation of a fundamental diagram for urban traffic flow. *The European Physical Journal B—Condensed Matter and Complex Systems* 2009; **70**(2): 229–241.
32. Ji Y, Daamen W, Hoogendoorn S, Hoogendoorn-Lanser S Qian X. Investigating the shape of the macroscopic fundamental diagram using simulation data. *Transportation Research Record: Journal of the Transportation Research Board* 2010; **2161**: 40–48.
33. Buisson C, Ladier C. Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. *Transportation Research Record: Journal of the Transportation Research Board* 2009; **2124**: 127–136.
34. Aboudolas K, Papageorgiou M, Kouvelas A Kosmatopoulos E. A rolling-horizon quadratic-programming approach to the signal control problem in large-scale congested urban road networks. *Transportation Research Part C: Emerging Technologies* 2010; **18**(5): 680–694.
35. Lin S, Zhou Z Xi Y. Model-based traffic congestion control in urban road networks: analysis of performance criteria. *Transportation Research Record* 2013; **2390**: 112–120.
36. Lin S, De Schutter B, Xi Y Hellendoorn H. Efficient network-wide model-based predictive control for urban traffic networks. *Transportation Research Part C: Emerging Technologies* Oct. 2012; **24**: 122–140.
37. Ji Y, Geroliminis N. On the spatial partitioning of urban transportation networks. *Transportation Research Part B: Methodological* 2012; **46**(10): 1639–1656.
38. Ji Y, Luo J Geroliminis N. Empirical observations of congestion propagation and dynamic partitioning with probe data for large-scale systems. *Transportation Research Record* 2014; **2422**(2): 1–11.
39. Mazloumian A, Geroliminis N Helbing D. The spatial variability of vehicle densities as determinant of urban network capacity. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 2010; **368**(1928): 4627–4647.
40. Haddad J, Geroliminis N. On the stability of traffic perimeter control in two-region urban cities. *Transportation Research Part B: Methodological* 2012; **46**(9): 1159–1176.
41. Keyvan-Ekbatani M, Kouvelas A, Papamichail I Papageorgiou M. Exploiting the fundamental diagram of urban networks for feedback-based gating. *Transportation Research Part B: Methodological* 2012; **46**(10): 1393–1403.
42. Ekbatani MK, Kouvelas A, Papamichail I Papageorgiou M. Congestion control in urban networks via feedback gating. *Procedia—Social and Behavioral Sciences* 2012; **48**(17): 1599–1610.
43. Keyvan-Ekbatani M, Papageorgiou M Knoop VL. Controller design for gating traffic control in presence of time-delay in urban road networks. *Transportation Research Part C: Emerging Technologies* 2015; **59**: 308–322.
44. S. Lin, T. Ling, and Y. Xi, Model predictive control for large-scale urban traffic networks with a multi-level hierarchy, in 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013), The Hague, The Netherlands, 2013, pp. 211–216.
45. Ramezani M, Haddad J Geroliminis N. Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control. *Transportation Research Part B: Methodological* 2015; **74**: 1–19.
46. Hajiahmadi M, Haddad J, De Schutter B Geroliminis N. Optimal hybrid perimeter and switching plans control for urban traffic networks. *IEEE Transactions on Control Systems Technology* 2015; **23**(2): 464–478.
47. Haddad J, Shraiber A. Robust perimeter control design for an urban region. *Transportation Research Part B: Methodological* 2014; **68**: 315–332.
48. Haddad J. Robust constrained control of uncertain macroscopic fundamental diagram networks. *Transportation Research Part C: Emerging Technologies* 2015; **59**: 323–339.
49. Lin S, Kong Q-J Huang Q. A simulation analysis on the existence of network traffic flow equilibria. *IEEE Transactions on Intelligent Transportation Systems* August 2014; **15**(4): 1706–1713.
50. Núñez A, Cortés C, Sáez D, De Schutter B Gendreau M. Multiobjective model predictive control for dynamic pickup and delivery problems. *Control Engineering Practice* 2014; **32**: 73–86.
51. Groot N, De Schutter B Hellendoorn H. Integrated model predictive traffic and emission control using a piecewise-affine approach. *IEEE Transactions on Intelligent Transportation Systems* 2013; **14**(2): 587–598.
52. Z. Zhou, S. Lin, and Y. Xi, A dynamic network partition method for heterogenous urban traffic networks, in 15th International IEEE Conference on Intelligent Transportation Systems (ITSC 2012), Alaska, USA, 2012, pp. 820–825.

DOI: 10.1002/atr