A model predictive perimeter control with real-time partitions

Shang Jiang * **Mehdi Keyvan-Ekbatani** ** **Dong Ngoduy** ***

* *Department of Civil and Natural Resource Engineering, University of Canterbury, Christchurch, New Zealand (e-mail: shang.jiang@pg.canterbury.ac.nz)*
** *Department of Civil and Natural Resource Engineering, University of Canterbury, Christchurch, New Zealand (corresponding author e-mail: mehdi.ekbatani@canterbury.ac.nz)*
*** *Department of Civil Engineering, Monash University, Melbourne, Australia (e-mail: Dong.Ngoduy@monash.edu)*

Abstract: Previous studies through simulation and empirical data have shown that a Network Macroscopic Fundamental Diagram (NMFD) exists and can be used for designing network optimal perimeter control strategies. These control strategies rely on well defined NMFDs, which highly depend on the homogeneity of the traffic condition in the network. However, it is known that traffic dynamics change drastically during the day in different zones in a large-scale network, and different control strategies might lead to heterogeneous traffic distribution across the urban network. One potential direction is re-partitioning the network to maintain the well defined NMFDs. However, re-partitioning the network changes each sub network's size, such that it makes the well-defined NMFDs unpredictable. This paper provides a model predictive control-based optimization approach for perimeter control using real-time partitioning to avoid this problem and utilize re-partitioning techniques. Results show that the proposed method can be used in a heterogeneous network to improve control performance by redistributing accumulations via re-partitioning over time. Our results, which are compared to no control and the traditional model predictive control, yield that the proposed method is superior to the others.

Keywords: Perimeter control; Macroscopic fundamental diagram; Network fundamental diagram; Network partitioning; Model predictive control

1. INTRODUCTION

Modeling microscopic traffic dynamics in an urban traffic network with a large number of intersections and road links is a challenging task. In order to build a realistic model, not only the computational burden needs to be considered, but also the queues, disturbances, drivers' behavior, and the interplay between neighboring intersections are required in detail. On the contrary, the NMFD-based approaches have shown beneficial for macroscopic modeling of traffic dynamics and, ultimately, mitigating congestion in large-scale networks (Keyvan-Ekbatani et al., 2019). NMFD is particularly useful for designing perimeter control strategies, resulting from the property of NMFDs that the shape of NMFDs is relatively stable to a small fluctuation of traffic demand patterns. In reality, road users' daily trip schedule shows a fixed pattern in which their known places, such as the workplace, home, a favorite shopping center, are usually their common origins and destinations (Zong et al., 2019). This stable trip pattern provides the essential presupposition of the utilization of well defined NMFDs to design network level control strategies to maintain the throughput of a network at its maximum level by limiting incoming redundancy vehicles. However, note that only networks with a homogeneous traffic state have a well defined NMFD. Pre-partitioning is one of the possible solutions. The logic behind the pre-partitioning is to divide the network into some sub networks with relatively homogeneous traffic state to construct a multi-region system such that each sub network in the system has a well defined NMFD. Then, control strategies can be developed based on the multi-region system.

Compared to the control strategies for single-region networks, see Daganzo (2007) and Keyvan-Ekbatani et al. (2012) for example, pre-partitioning is more important for control strategies for multiple-region system. Geroliminis et al. (2013) designed the initial control strategies for multi-region system. Then Ramezani et al. (2015) improved the strategy via hierarchical structure to come up with a higher level perimeter control upon a lower level feedback control. Then, the multi-region system model predictive control strategy is linearized by Kouvelas et al. (2017). Zhou et al. (2016) designed hierarchical MPC gating regulation by which the implementation of MPC operates on both network and link levels. The improvement is also observed in Sirmatel and Geroliminis (2017), which aggregated route guidance into the multi-region system perimeter control. Recently, Haddad and Zheng (2018) developed an adaptive approach for a multi-region system with a delay based model to be better suitable for traffic wave propagation and travel time evolution. Sirmatel and

Geroliminis (2019) increased the availability of the model predictive control strategy for the multi-region system by giving a solution of the estimation problem of initial states for the prediction model in the model predictive control strategy for the multi-region system.

The multi-region NMFDs system's stability might be affected by the undesired queues generated from the control strategies by which the queues increase the variance of link density in the pre-partitioned regions. The shape of the NMFD would be changed. Even though some advanced queue management, see Keyvan-Ekbatani et al. (2021) for example, can be applied to multi-reservoir system, the unstable shape problem of the NMFD is still suspending. To mitigate the influence of the queues, real-time partitioning could also be taken into account.

Partitioning can be applied by some classic clustering techniques (e.g. k-mean, hierarchical clustering and spectral clustering). For example, Ji and Geroliminis (2012) proposed a multi-step approach that utilizes N-cut strategy. Also, one can implement a hierarchical concept based on a series of connectivity constraints. Algorithms begin with every single quantity or the whole set of databases. Then iterative merging or splitting algorithms can be employed, which has been achieved within some methods, see Guo (2008) and An et al. (2017) for example. The partitioning improvement can also be observed in Gu and Saberi (2019) and Saeedmanesh and Geroliminis (2016), as it is a continuously developing topic. However, real-time partitioning changes each sub network's size such that it makes the well-defined NMFDs unpredictable. This causes the difficulty in which real-time partitions can hardly be applied in network level control strategies.

This study takes both control and real-time partitioning into account and forges a model predictive control with a partitioning method to improve control performance. This method is then tested on a multi-region NMFDs system and compared versus *no control* and *traditional MPC* scenarios to present the importance of real-time partitions used in traffic flow perimeter control. The remainder of the paper is organized as follows: section II presents the applied methodology in this study, including the network transmission model, the partitioning method, and the control objective with constraints. Then, the paper discusses the simulation results obtained from the implementation of multi-region scenarios. In the end, conclusions have been addressed.

2. METHODOLOGY

2.1 Multi-cell NMFDs system

A heterogeneous urban network can be pre-partitioned into several relatively homogeneous sub networks or cells, as illustrated in Fig.1. The cell i is included in the set of all cells $\mathbf{R}$. Assume the sub urban network i has a well-defined NMFD $G_i(n_i(t))$, where $n_i(t)$ is the accumulation in cell i. For the multi-region system, $d_{ii}(t)$ denotes endogenous traffic demands in i, while exogenous traffic demands from origin of i with destination to one of the cell j is d_{ij}, $i, j \in \mathbf{R}$. Then, the accumulation $n_i(t)$ in i can be calculated as: $n_i(t) = \sum_{j\in\mathbf{R}} n_{ij}(t)$, where $n_{ij}(t)$ is the accumulation in i with destination to j. Let $h \in \mathbf{N}_i$ ($i \in \mathbf{R}$) be one of the cells in the set of all the neighboring cells $\mathbf{N}_i$.

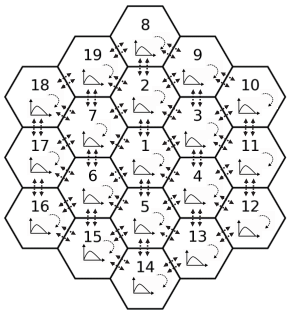


Fig. 1. Multi-cell system

The perimeter controllers $u_{ih}(t)$ and $u_{hi}(t)$ are designed to limit the transfer flows between cells in order to maximize the total travel distance of all vehicles that emerged in the multi-region NMFDs system. They allow ratio transfers across boundaries at time t (e.g. $u_{ih}(t)$ denotes the transfer ratio gating the transfer flow from i to h). The traffic dynamics of the multi-region NMFDs system as in Ramezani et al. (2015) can be written as follows:

$$\forall i \in \mathbf{R},\ \frac{dn_{ii}(t)}{dt} = d_{ii}(t) - M_{ii}(t) + \sum_{h\in\mathbf{N}_i} u_{hi}(t)\hat{M}_{hii}(t) \tag{1}$$

$$\forall i, j \in \mathbf{R}, i \neq j,\ \frac{dn_{ij}(t)}{dt} = d_{ij}(t) - \sum_{h\in\mathbf{N}_i} u_{ih}(t)\hat{M}_{ihj}(t) + \sum_{h\in\mathbf{N}_i, h\neq j} u_{hi}(t)\hat{M}_{hij}(t) \tag{2}$$

where $\hat{M}_{ihj}(t)$ signifies the effective transfer flow from i with destination to j via the neighboring downstream cell h, which is restricted by boundary capacity $C_{ih}(n_h(t))$ between i and h, similar to $\hat{M}_{hij}(t)$ and $\hat{M}_{hii}(t)$. $M_{ii}(t)$ is the completion trips in i at time t. The boundary capacity is affected by both the serve capacity of the roads connecting i and h and also constrained by the supply of h. Considering the reality, the road serve capacity is relatively constant such that the boundary capacity can be assumed as a function of $n_h(t)$. Therefore, the $\hat{M}_{ihj}(t)$ is described as:

$$\hat{M}_{ihj}(t) = min[M_{ihj}(t), S_{ih}(n_h(t))\frac{n_{ij}(t)\theta_{ihj}(t)}{\sum_{k\in\mathbf{R}} n_{ik}(t)\theta_{ihk}(t)}, C_{ih}\frac{n_{ij}(t)\theta_{ihj}(t)}{\sum_{k\in\mathbf{R}} n_{ik}(t)\theta_{ihk}(t)}] \tag{3}$$

where $\theta_{ihj}(t)$ is the splitting ratio of flow via h from origin i with destination to j, $\sum_{h\in\mathbf{N}_i} \theta_{ihj}(t) = 1$, the $S_{ih}^{\max}$ is the supply constraint, which is designed by:

$$S_{ih}(n_h(t)) = \omega_h(n_h^{\text{jam}} - n_h(t))\cdot \frac{\sum_{k\in\mathbf{R}} n_{ihk}(t)}{\sum_{m\in\mathbf{N}_h, k\in\mathbf{R}} n_{mhk}(t) + n_{hh}(t)} \tag{4}$$

where m is one of the neighboring cell in the set of all its neighboring cells $\mathbf{N}_h$, while the transfer users from

m with a certain destination k through h at time t is denoted by $\sum_{k\in\mathbf{R}} n_{ihk}(t)$. There is also boundary capacity C_{ih}, specifying the service capability of the connecting roads between the two cells. The boundary capacity can be defined as:

$$C_{ih} = \sum_{l\in\mathbf{L}_{ih}} a_l C_l \tag{5}$$

where l is a link in $\mathbf{L}_{ih}$, which denotes the set of all the road links from i to h; a_l is a variable defining the number of lanes on link l; the C_l is the lane capacity of link l, note that there is an assumption of existing the same capacities for all lanes on one link. $G_i(n_i(t))$ is designed as a trapezoid NMFD because in an urban network the capacity generally occurs in a range of accumulations. Then, as suggested in Mariotte and Leclercq (2019), the external flow to a downstream reservoir is only restricted by the supply function of the downstream reservoir, whereas the internal flow follows the NMFD. The internal and external flows can be formulated as:

$$M_i^{\mathrm{NMFD}}(t) = \begin{cases} G_i(n_i(t)) & \text{for internal flow} \\ O_i(n_i(t)) & \text{for external flow} \end{cases} \tag{6}$$

The internal flow $G_i(n_i(t))$ can be shaped as the complete trapezoid NMFD, while the outflow of the external flow of high accumulations is only restricted by the downstream supply. Thus, $G_i(n_i(t))$ and $O_i(n_i(t))$ can be written as:

$$G_i(n_i(t)) = \begin{cases} \upsilon_i n_i(t) & \text{for } 0 \le n_i(t) < n_i^{a_0} \\ C_i & \text{for } n_i^{a_0} \le n_i(t) \le n_i^{b_0} \\ \omega_i(n_i^{\mathrm{jam}} - n_i(t)) & \text{for } n_i^{b_0} < n_i(t) \le n_i^{\mathrm{jam}} \end{cases} \tag{7}$$

and

$$O_i(n_i(t)) = \begin{cases} \upsilon_i n_i(t) & \text{for } 0 \le n_i(t) < n_i^{a_0} \\ C_i & \text{otherwise} \end{cases} \tag{8}$$

where $i \in \mathbf{R}$; n_i denotes the number of vehicles in i, it follows physical constraints of the empty and gridlock or jam situations ($0 \le n_i \le n_i^{\mathrm{jam}}$), the υ_i and ω_i are NMFD parameters, C_i denotes the maximum flow of the NMFD, $n_i^{a_0}$ and $n_i^{b_0}$ signify the lower and upper bounds of critical accumulation, while n_i^{jam} is the gridlock or jam accumulation. Then, the transfer flow $M_{ihj}(t)$ can be obtained from the external flow function as:

$$M_{ihj}(t) = \theta_{ihj}(t)\frac{n_{ij}(t)}{n_i(t)} O_i(n_i(t)) \tag{9}$$

while the internal flow function is:

$$M_{ii}(t) = \frac{n_{ii}(t)}{n_i(t)} G_i(n_i(t)) \tag{10}$$

2.2 Partitioning

To explicitly describe the method, we redefine 'cells' as the sub networks resulting from pre-partitioning and 'reservoirs' as the sub networks resulting from the real-time partitioning. By applying pre-partitioning, the entire network of hosts is divided into a fixed number and size of cells. These cells would remain constant throughout the whole period. Static partitioning might be implemented to divide a large-scale network into a multi-cell NMFDs system. There will be no discussion of this step in this paper since the focus will be on applying real-time partitioning for control purposes based on a multi-cell NMFDs system. It is assumed that the system can be well built. Then, real-time partitioning produces a certain number of reservoirs. A reservoir consists of single or multiple cells. The traffic dynamic between these cells follows the network transmission model that we have presented above. However, the control ratio should only be applied to the border between reservoirs.

To implement the partitioning, we convert the multi-cell network into a graph $G_{\mathrm{p}}(\mathbf{V}, \mathbf{E})$ in which a cell i is regarded as a vertex $v \in \mathbf{V}$ with a number of directed edges ($e \in \mathbf{E}$), that are geographically connected to the neighboring cells. Initially, we search for the highest accumulation vertex v_{seed} at time t, and set it as the seed vertex. Secondly, the seed iteratively grows by merging the most similar neighboring vertex to form a gradually increasing vector $\bar{v}$. If the length of the vector $\bar{v}$ is $l_{\bar{v}}$ and the accumulation in each $v \in \bar{v}$ is $n_v(t)$, the similarity between the $\bar{v}$ and a neighboring vertex is calculated as:

$$\forall v \in \mathbf{N}_v,\ s_v = \left(\frac{\sum_{x\in\bar{v}} n_x(t)}{l_{\bar{v}}} - n_v\right)^2 \tag{11}$$

where $\mathbf{N}_v$ is the set of all vertex neighboring $\bar{v}$, $n_x(t)$ is the accumulation of a certain vertex $x \in \bar{v}$ at time t. The vertex with minimum s_v is sequentially added into $\bar{v}$. The process would be repeated until $l_{\bar{v}}$ equal to the cardinality of $\mathbf{V}$, where the vector $\bar{v}$ can represent the sequence of the relationship among cells by considering both accumulation and geographic information. By cutting the vector, the whole network can be partitioned into a number of reservoirs. Let the cutting variable $p \in \{1, 2, 3...l_{\bar{v}} - 1\}$ denote the cutting point after the p-th vertex in $\bar{v}$, for example, if $p_1 = 3$ and $p_2 = 5$, the first cutting point is located between the third vertex and the fourth vertex in $\bar{v}$ and the second cutting point is between fifth and sixth vertex, such that the first 3 vertex are in reservoir 1, and the vertex 4 and vertex 5 are deployed in reservoir 2, and the rest are in reservoir 3. By this definition, we can partition the entire network using a number of variables $\{p_1, p_2, ...p_{l_{\bar{v}}-1}\}$. The partitions will update at each partitioning interval T_{p}.

2.3 Model predictive control with partitioning

Let the $\boldsymbol{u}^{\mathrm{d}}$ signify the vector of control ratio between reservoirs during the prediction horizon N_{P}. The control horizon is defined as N_{C}, which is smaller than N_{P}. To discretize the problem, it is assume that the control ratio will be constant for ΔT s), and the manipulation happens at every interval T. Thus, the objective function can be written as:

$$\boldsymbol{J} = \max_{\boldsymbol{u}^{\mathrm{d}}, \boldsymbol{p}} \sum_{T=0}^{N_{\mathrm{P}}-1} \sum_{i\in\mathbf{R}} M_{ii}(T) \tag{12}$$

subject to:

$$\boldsymbol{n}(0) = \hat{\boldsymbol{n}}(t_{\mathrm{c}}) \tag{13}$$

$$\forall 0 \le T \le N_{\mathrm{P}} - 1, \boldsymbol{n}(T+1) = f\Big(\boldsymbol{n}(T), \boldsymbol{d}(T), \mathbf{u}^{\mathrm{d}}(T), \Theta(T), P(T)\Big) \tag{14}$$

$$\forall i \in \mathbf{R},\ 0 \le \sum_{j\in\mathbf{R}} n_{ij}(T) \le n_i^{\mathrm{jam}} \tag{15}$$

$$P(T) = \begin{cases} \hat{P}(t_c) & \text{for } T = 0 \wedge T \neq T_p \\ f_p(\boldsymbol{p}(T), \boldsymbol{n}(T)) & \text{for } T = T_p \wedge \\ & T_p \leq \dfrac{t_c \ mod \ (T_p \Delta T)}{\Delta T} + N_C \\ P(T-1) & \text{otherwise} \end{cases} \tag{16}$$

$$\boldsymbol{u}^d(0) = \hat{\boldsymbol{u}}^d(t_c) \tag{17}$$

$$\boldsymbol{u}^{d+}(T) + \boldsymbol{u}^{d-}(T) = 1 \tag{18}$$

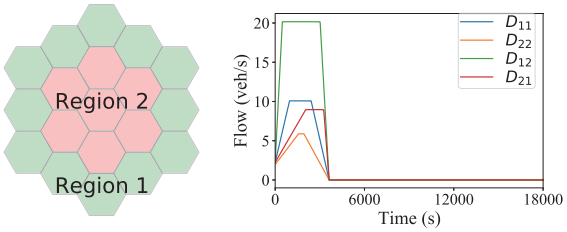
$$\forall 0 \leq T \leq N_P - 1, \ \boldsymbol{u}_{min} \leq \boldsymbol{u}^d(T) \leq \boldsymbol{u}_{max} \tag{19}$$

$$\forall N_C + 1 \leq T \leq N_P - 1, \ \boldsymbol{u}^d(T) = \boldsymbol{u}^d(T-1) \tag{20}$$

where the $\boldsymbol{u}^d(0)$ is the initial control ratio by the randomly selected first guess $\hat{\boldsymbol{u}}^d(t_c)$; $\boldsymbol{n}(0)$ is the initial accumulation states, which is measurement or estimation $\hat{\boldsymbol{n}}(t_c)$ from the plant; the accumulations, demands, control ratio and splitting ratio vectors at interval T are indicated by $\boldsymbol{n}(T)$, $\boldsymbol{d}(T)$, $\mathbf{u}^d(T)$ and $\Theta(T)$, respectively; $f(.)$ describes the dynamic of the multi-cell NMFD system. To find the optimal partitions, the cutting points $\boldsymbol{p}$ are also considered as variables for optimization. $P(T)$ is the partitions at interval T. It is initially obtained by partitions observation $\hat{P}(t_c)$ at time t_c, and recalculated at time T_p, which represents the decision making time of partitioning. T_p should follow the partitioning plan of the plant. Due to partitioning period being usually a few times more than control interval, the partitioning would not always be applied at the first step of the prediction. Therefore, in case of $T_p > (t_c \ mod \ (T_p \Delta T))/\Delta T + N_C$, there is no necessity of re-partitioning. $f_p(.)$ is the partitioning method introduced in 2.2. The mixed integer non-linear optimization problem can be solved via sequential quadratic programming, e.g. APOPT solver (Hedengren et al., 2012). The control ratios at time interval T should be bounded by the maximum and minimum limitation, which are represented by two vectors $\boldsymbol{u}_{max}$ and $\boldsymbol{u}_{min}$ respectively. Due to the physical meaning of control ratio (green phase over cycle), each ratio should be between 0 and 1. In terms of fairness, the upper bound is designed as a ratio less than 1, and the lower bound is larger than 0 in order to allow at least a certain number of vehicles from a certain direction to cross the intersection. There exist signalized intersections scattering along partition boundaries. The boundary control can be applied by utilizing the existing signals. Such that the transfer flows from opposite directions at time T ($\boldsymbol{u}^{d+}(T)$ and $\hat{\boldsymbol{u}}^{d-}(T)$) go via the same intersection system. In other words, the opposite directions control ratios are restricted by each other since the sum of green time for different directions at one intersection should equal to or less than the total cycle. Hence, the sum of all opposite direction control ratios at interval T is set equivalent to 1, being represented by the constraint $\boldsymbol{u}^{d+}(T) + \hat{\boldsymbol{u}}^{d-}(T) = 1$.

3. NUMERICAL TEST

This section presents a numerical test example to explore the characteristics of the proposed model predictive control with the partitioning scheme. This example shows the investigation of the effect of no control (NC), model predictive control using fixed partitions (MPC), and model predictive control with real-time partitioning (MPC-P). Note that the scenarios, such as some cases in Geroliminis et al. (2013) are not thoroughly tested since the main contribution of this study is modeling the aggregation of real-time partitioning with a perimeter control strategy. This study emphasizes the performance of real-time partitioning strategy in control compared to fixed partitions.



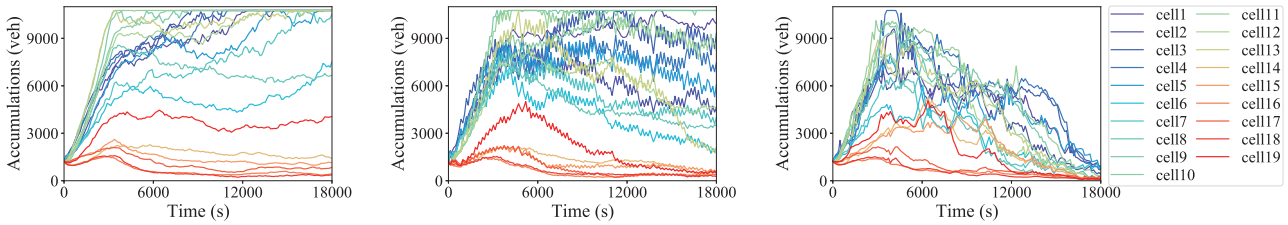
(a) Two regions network (b) Demand profile

Fig. 2. Network layout and demand profile

Table 1. NMFD parameters

Parameter	Value	Units
v_i	10.57	h^{-1}
ω_i	3.84	h^{-1}
n_i^{jam}	10762	veh
$n_i^{a_0}$	1736	veh
$n_i^{b_0}$	5986	veh
C_i	18341	veh/h

The testbed consists of 19 cells, which is shown in Fig.2(a) by two regions to designate the periphery (12 cells) and the city center (7 cells) of an urban network. The NMFD of the cells has been assumed to be the same, and the values in Table 1 are derived from the realistic NMFD of Christchurch, New Zealand (for further details on the network description and characteristics see Johari et al. (2020)). In the numerical test, the initial accumulation in each cell is identical ($n_i(0) = 1232$). To simulate the peak period traffic, the total demand from region 1 to region 2 (D_{12}) is designed larger than the others, including the opposite demand flows from region 2 to region 1 (D_{21}), the internal demand in region 1 (D_{11}) and region 2 (D_{22}), see profile in Fig.2(b). We also assume a largely biased distribution of transfer demand to exhibit the importance of partitioning (cell 8, cell 9, cell 10, cell 811, cell 12, and cell 13 contain 90% transfer demands in region 1). Random recurrent noises are added to the demand for simulating the stochasticity of the demand in the plant and assuming errors in the NMFDs in the plant according to the method mentioned in Geroliminis et al. (2013). For the whole test, the selected MPC controller and partitioning variables are as follows: the prediction horizon $N_p = 20$, the control horizon $N_C = 2$, the control interval is 120 s, the re-partitioning period is 600 s, the lower bound and upper bound of control ratio are 0.1 and 0.9 respectively, the ratio for non-controlled boundaries is fixed as 0.5. The splitting ratio is fixed and calculated by the shortest path method initially. Note that due to the space limitation, we will not discuss the influence of rerouting techniques in this paper.



(a) NC (b) MPC (c) MPC-P

Fig. 3. Accumulation profile of each cell: (a) no control; (b) model predictive control with fixed partitions; (c) model predictive control with real-time partitions

The total boundary capacity of the observed cell is 186900 (veh/h).

We compared three scenarios that are essential to exhibit the importance of partitioning using in control strategy: (i) no control (NC) i.e. only fixed control ratio have applied to all boundaries; (ii) model predictive control with fixed partitions where the MPC controller is only implemented on the boundary between region 1 and region 2; (iii) model predictive control with partitions; the partitioned reservoirs can be updated in each re-partitioning period via a single objective function that finds the optimal control ratio and the optimal partitions simultaneously.

Fig.3 illustrates the accumulations $n_i(t)$ evolution over simulation duration for the three strategies. Initially, the accumulations in all cells increase in all cases. Then, the traffic state in most cells for no control case goes quickly to gridlock or jam, see Fig.3(a). The MPC case shows a better circulation in which the traffic situation in a few cells can reach the gridlock or jam accumulation, while there is no gridlock or jam observed after 5000 s for the MPC-P case. The mean completion trips for MPC-P within the simulation duration is 10.38 (veh/s), which is approximately 20% and 40% higher than MPC and NC (note that the completion trips decrease because of the congestion release during the last 3000 seconds. Thus the effective performance is even higher). Both the shapes in these figures and the total values demonstrate that although both MPC and MPC-P are effective in mitigating congestion and improve traffic situation, the MPC-P performs superiorly.

A more clear investigation is presented in Fig.4, which exhibits the completion trips ($\sum_{i\in\mathbf{R}} M_{ii}(t)$) over the simulation duration. Utilizing the control strategies can vastly increase the entire traffic performance, see MPC and MPC-P versus NC. Obviously, the MPC-P is better than the MPC in terms of completion trips. Note that the completion trips for MPC-P decrease after 16000 s, which does not mean the performance deteriorates. This is caused by faster mitigation of the congestion and higher throughput, which results in an empty network at the end of the simulation.

To understand why the strategies perform differently, we plot Fig.5. This figure shows the accumulation standard deviation (STD) versus time. With the uneven demand increasing, the heterogeneity for all three cases grows sharply. Then, the proposed MPC-P approach tries to reverse the tendency and bring the whole network back

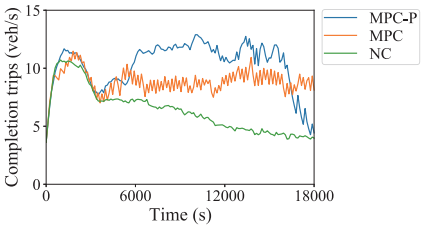


Fig. 4. Completion trips

to more homogeneous states. As a result, the network undergoes redistribution, avoiding gridlock or jam and restoring flows to its potential maximum. In contrast, there is less effect on accumulation redistribution for MPC and NC case. Therefore, in these scenarios after a certain cell reaches the congested traffic state, its upstream cells quickly become congested. Consequently, the congestion might further propagate to the whole network.

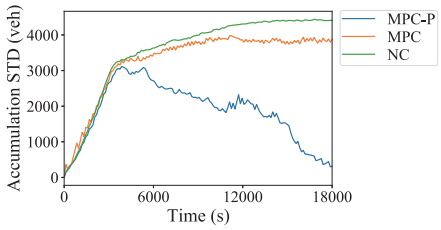


Fig. 5. Accumulation homogeneity

Fig.6 lists the partitions over the simulation duration. Since there is only one partitioning decision parameter set in the objective function, the entire network is bi-partitioned for each partitioning step. The computational cost for each prediction averages 16 seconds, small compared to the signal cycle of 120 seconds. In reality, a longer duration (e.g. half an hour) of the partitioning step can be considered. Thus, the proposed method can be feasible for dealing with the congestion in heterogeneous traffic states from both performance and efficient computational perspectives.

4. CONCLUSIONS

This study presents a model predictive control structure that integrates perimeter control scheme and real-time partitioning to improve the mobility in the heterogeneous traffic networks. The optimization makes decisions by finding suitable control and partitioning variables to optimize

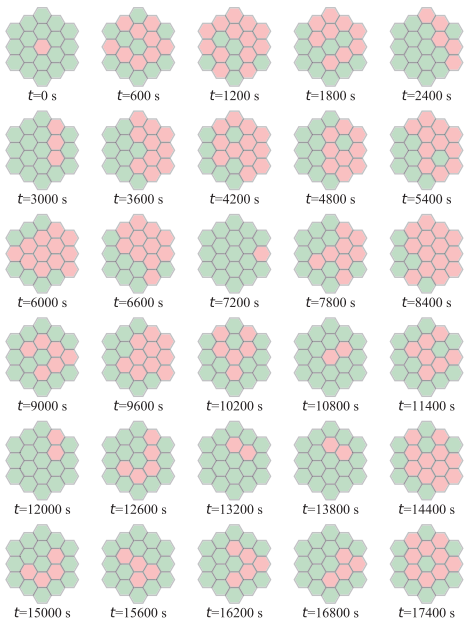


Fig. 6. Real-time partitioning during perimeter control

the total trip completion rate. The perimeter control has been applied to the boundary between reservoirs to limit transfer flows to keep the total traffic performance as its maximum. The applied partitioning method enables reservoirs to be updated over time without changing the well-defined NMFDs in each cell to overcome the difficulty of combining real-time partitioning and perimeter control. The numerical test shows an efficient performance (i.e. 20% and 40% higher than conventional MPC and no control) of the proposed MPC-P method under the simulated peak traffic situation with uneven demand pattern. The accumulation homogeneity investigation also exhibits that the performance improvement comes from increasing the accumulation homogeneity across the entire network.

REFERENCES

An, K., Chiu, Y.C., Hu, X., and Chen, X. (2017). A network partitioning algorithmic approach for macroscopic fundamental diagram-based hierarchical traffic network management. *IEEE Transactions on Intelligent Transportation Systems*, 19(4), 1130–1139.

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

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

Gu, Z. and Saberi, M. (2019). A bi-partitioning approach to congestion pattern recognition in a congested monocentric city. *Transportation Research Part C: Emerging Technologies*, 109, 305–320.

Guo, D. (2008). Regionalization with dynamically constrained agglomerative clustering and partitioning (redcap). *International Journal of Geographical Information Science*, 22(7), 801–823.

Haddad, J. and Zheng, Z. (2018). Adaptive perimeter control for multi-region accumulation-based models with state delays. *Transportation Research Part B: Methodological*.

Hedengren, J., Mojica, J., Cole, W., and Edgar, T. (2012). Apopt: Minlp solver for differential and algebraic systems with benchmark testing. In *Proceedings of the INFORMS National Meeting, Phoenix, AZ, USA*, volume 1417, 47.

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

Johari, M., Keyvan-Ekbatani, M., and Ngoduy, D. (2020). Impacts of bus stop location and berth number on urban network traffic performance. *IET Intelligent Transport Systems*.

Keyvan-Ekbatani, M., Carlson, R.C., Knoop, V.L., and Papageorgiou, M. (2021). Optimizing distribution of metered traffic flow in perimeter control: Queue and delay balancing approaches. *Control Engineering Practice*, 110, 104762.

Keyvan-Ekbatani, M., Gao, X., Gayah, V.V., and Knoop, V.L. (2019). Traffic-responsive signals combined with perimeter control: investigating the benefits. *Transportmetrica B: Transport Dynamics*, 7(1), 1402–1425.

Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I., and Papageorgiou, M. (2012). Exploiting the fundamental diagram of urban networks for feedback-based gating. *Transportation Research Part B: Methodological*, 46(10), 1393–1403.

Kouvelas, A., Saeedmanesh, M., and Geroliminis, N. (2017). A linear formulation for model predictive perimeter traffic control in cities. *IFAC-PapersOnLine*, 50(1), 8543–8548.

Mariotte, G. and Leclercq, L. (2019). Flow exchanges in multi-reservoir systems with spillbacks. *Transportation Research Part B: Methodological*, 122, 327–349.

Ramezani, M., Haddad, J., and Geroliminis, N. (2015). Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control. *Transportation Research Part B: Methodological*, 74, 1–19.

Saeedmanesh, M. and Geroliminis, N. (2016). Clustering of heterogeneous networks with directional flows based on "snake" similarities. *Transportation Research Part B: Methodological*, 91, 250–269.

Sirmatel, I.I. and Geroliminis, N. (2017). Economic model predictive control of large-scale urban road networks via perimeter control and regional route guidance. *IEEE Transactions on Intelligent Transportation Systems*, 19(4), 1112–1121.

Sirmatel, I.I. and Geroliminis, N. (2019). Moving horizon demand and state estimation for model predictive perimeter control of large-scale urban networks. In *2019 18th European Control Conference (ECC)*, 3650–3655. IEEE.

Zhou, Z., De Schutter, B., Lin, S., and Xi, Y. (2016). Two-level hierarchical model-based predictive control for large-scale urban traffic networks. *IEEE Transactions on Control Systems Technology*, 25(2), 496–508.

Zong, F., Tian, Y., He, Y., Tang, J., and Lv, J. (2019). Trip destination prediction based on multi-day gps data. *Physica A: Statistical Mechanics and its Applications*, 515, 258–269.