A linear formulation for model predictive perimeter traffic control in cities

Conference Paper

Author(s):
Kouvelas, Anastasios; Saeedmanesh, Mohammadreza; Geroliminis, Nikolas

Publication date:
2017-07

Permanent link:
https://doi.org/10.3929/ethz-b-000276000

Originally published in:
IFAC-PapersOnLine 50(1), https://doi.org/10.1016/j.ifacol.2017.08.1411

A Linear Formulation for Model Predictive Perimeter Traffic Control in Cities ⋆

Anastasios Kouvelas, Mohammadreza Saeedmanesh, Nikolas Geroliminis

Urban Transport Systems Laboratory
School of Architecture, Civil & Environmental Engineering
École Polytechnique Fédérale de Lausanne
CH-1015 Lausanne, Switzerland (e-mails: {tasos.kouvelas, mohammadreza.saeedmanesh, nikolas.geroliminis}@epfl.ch)

Abstract: An alternative approach for real-time network-wide traffic control in cities that has recently gained a lot of interest is perimeter flow control. The basic concept of such an approach is to partition heterogeneous cities into a small number of homogeneous regions (zones) and apply perimeter control to the inter-regional flows along the boundaries between regions. The transferring flows are controlled at the traffic intersections located at the borders between regions, so as to distribute the congestion in an optimal way and minimize the total delay of the system. The focus of the current work is to study three aspects that are not covered in the perimeter control literature, which are: (a) the treatment of some model parameters that are not measurable in real life implementations and can affect the performance of the controller (e.g. advanced online estimation schemes can be developed for this purpose), (b) integration of appropriate external demand information that has been considered system disturbance in the derivation of feedback control laws in previous works, and (c) mathematical formulation of the original nonlinear problem in a linear form, so that optimal control can be applied in a (rolling horizon) model predictive concept. This work presents the mathematical analysis of the optimal control problem, as well as the approximations and simplifications that are assumed in order to derive the formulation of a linear optimization problem. Preliminary simulation results for the case of a macroscopic environment (plant) are presented, in order to demonstrate the efficiency of the proposed approach. Results for the closed-loop model predictive control scheme are presented for the nonlinear case, which is used as "benchmark", as well as the linear case.

Keywords: Model predictive control; nonlinear optimisation; linear approximation; urban perimeter control.

1. INTRODUCTION

Traffic congestion is a major problem for urban environments and modern metropolitan areas. Most cities around the world have been persistently becoming denser and wider over the last decades and the problem of urban traffic management is steadily gaining momentum due to its economic, social and environmental impact. Many efforts have been carried out to optimize signal settings during the peak hours, where networks face serious congestion problems and the performance of the infrastructure degrades significantly. The state-of-practice strategies fail to deal efficiently with oversaturated conditions (i.e. queue spillbacks and partial gridlocks), as they are either designed by use of simplified models that do not accurately replicate some traffic flow phenomena (e.g. propagation of congestion), or based on application-specific heuristics. An alternative approach for real-time network-wide traffic control that has recently gained a lot of interest is perimeter flow control (or gating). The basic concept of such an approach is to partition heterogeneous cities into a small number of homogeneous regions (zones) and apply perimeter control to the inter-regional flows along the boundaries between regions. The inter-transferring flows are controlled at the traffic intersections located at the borders between regions, so as to distribute the congestion in an optimal way and minimize the total delay of the system (an alternative objective could maximize the total throughput).

Perimeter flow control can be viewed as a high-level regional control scheme and might be combined with other strategies (e.g. local, distributed or coordinated controllers) in a hierarchical control framework; this topic has gained a lot of attraction in the research community lately. For a recent review on this research direction the reader is referred to Keyvan-Ekbatani et al. (2012); Yildirimoglu et al. (2015). The original model for the dynamics of the multi-region process (plant) is highly nonlinear and the modelling tool that is utilized is the Macroscopic Fundamental Diagram (MFD) (see e.g. Ramezani et al. (2015)). MFD provides a concave, low-scatter relationship

⋆ This research has been supported by the ERC (European Research Council) Starting Grant "METAFERW: Modelling and controlling traffic congestion and propagation in large-scale urban multi-modal networks" (Grant #338205).

between network vehicle accumulations [*veh*] or density [*veh/km*] and network circulating flow [*veh/h*] or production [*veh·km*] for every region of the system which is homogeneously congested. Clustering techniques can help partition a network into regions with small variance of link densities (see e.g. Saeedmanesh and Geroliminis (2016)). Recently, there has been some effort to simplify this model; the authors in Hajiahmadi et al. (2015) proposed a hybrid model for formulating this problem and applied model predictive control (MPC) in conjunction with switching plans control. A pre-defined set of MFDs is introduced (which correspond to regional switching plans), and some affine approximations are made in order to reduce the complexity of the derived hybrid MPC approach.

In the current work we devise a linear program (LP) formulation to solve the finite-time optimal perimeter flow control problem. Moreover, we investigate two more aspects of the problem that have not been covered in the literature: (a) the treatment of some model parameters that are not easily measurable with loop detectors data in real life implementations, and can affect the performance of the controller; especially the regional route choice parameters and the origin-destination information, and (b) integration of realistic external demand information that has been considered system disturbance in the derivation of multivariable feedback control laws in previous works (Kouvelas et al. (2015)). Finally, the derived LP is solved in a rolling time horizon using a feedback MPC framework, and the control decisions are applied to the nonlinear plant for evaluation. The efficiency of the control decisions is compared to a "benchmark" case, in which the nonlinear MPC problem is solved using state-of-the-art nonlinear numerical solvers. Simulation results for the case of a macroscopic model (plant) are presented. Note that this benchmark approach is more challenging for application in real life due to computational requirements, and most importantly, the lack of detailed data (i.e. high resolution vehicle trajectories data is needed). Essentially, real-world data availability can restrain the methodologies that can be applied, and, consequently, the real-time applicability of the benchmark approach is deemed cumbersome.

2. AGGREGATED DYNAMICS FOR A PARTITIONED CITY

Consider an urban network partitioned in N homogeneous regions with well-defined MFDs (Figure 1). The index $i \in \mathcal{N} = \{1, 2, \ldots, N\}$ denotes the region of the system, $n_i(t)$ the total accumulation (number of vehicles) in region i, and $n_{ij}(t)$ the number of vehicles in region i with final destination region $j \in \mathcal{N}$, at a given time t. Let $\mathcal{N}_i$ be the set of all regions that are directly reachable from the borders of region i, i.e. adjacent regions to region i. The discrete time MFD dynamics of the N-region system can be described by the following first order difference equations

$$n_{ii}(k_\mathrm{p}+1) = n_{ii}(k_\mathrm{p}) + T_\mathrm{p}\Bigg(q_{ii}(k_\mathrm{p}) - M_{ii}(k_\mathrm{p}) - \sum_{h\in\mathcal{N}_i} M_{ii}^h(k_\mathrm{p}) + \sum_{h\in\mathcal{N}_i} M_{hi}^i(k_\mathrm{p})\Bigg) \quad (1)$$

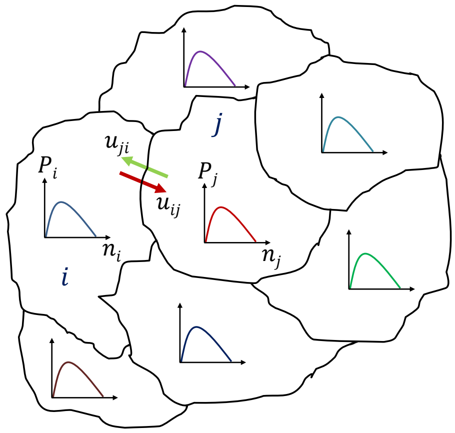

Fig. 1. The network of a city partitioned into a multi-region MFD system.

$$n_{ij}(k_\mathrm{p}+1) = n_{ij}(k_\mathrm{p}) + T_\mathrm{p}\Bigg(q_{ij}(k_\mathrm{p}) - \sum_{h\in\mathcal{N}_i} M_{ij}^h(k_\mathrm{p}) + \sum_{h\in\mathcal{N}_i} M_{hj}^i(k_\mathrm{p})\Bigg) \quad (2)$$

where $i, j \in \mathcal{N}$, $i \neq j$; $k_\mathrm{p} = 0, 1, \ldots, K_\mathrm{p} - 1$ is the model discrete time index, T_p [*sec*] the sample time period of the model (i.e. time $t = k_\mathrm{p} T_\mathrm{p}$). The exogenous variables $q_{ij}(k_\mathrm{p})$ [*veh/sec*] denote the (uncontrollable) traffic flow demand that is generated in region i, at time step k_p, with final destination in region j (i.e. $q_{ii}(k_\mathrm{p})$ is the demand generated in region i that has final destination in region i). The variables $M_{ij}^h(k_\mathrm{p})$ [*veh/sec*] denote the transfer flows from region i to region h, that have final destination region j, while $M_{ii}(k_\mathrm{p})$ [*veh/sec*] is the internal trip completion rate of region i (without going through another region). Consequently, the total accumulation $n_i(k_\mathrm{p})$ for region i can be computed by $n_i(k_\mathrm{p}) = \sum_{j\in\mathcal{N}} n_{ij}(k_\mathrm{p})$.

We assume that for each region i there exists a production MFD between accumulation $n_i(k_\mathrm{p})$ and total production, $P_i(n_i(k_\mathrm{p}))$ [*veh·m/sec*], which describes the performance of the system in an aggregated way. This MFD can be estimated using measurements from loop detectors and/or GPS trajectories. Transfer flows $M_{ij}^h(k_\mathrm{p})$ and internal trip completion rates $M_{ii}(k_\mathrm{p})$ are calculated according to the corresponding production MFD of the region, and proportionally to the accumulations $n_{ij}(k_\mathrm{p})$ as follows

$$M_{ii}(k_\mathrm{p}) = \theta_{ii}(k_\mathrm{p}) \frac{n_{ii}(k_\mathrm{p})}{n_i(k_\mathrm{p})} \frac{P_i(n_i(k_\mathrm{p}))}{L_i} \quad (3)$$

$$M_{ij}^h(k_\mathrm{p}) = \min\left\{C_{ih}(n_h(k_\mathrm{p})), u_{ih}(k_\mathrm{c})\theta_{ij}^h(k_\mathrm{p}) \frac{n_{ij}(k_\mathrm{p})}{n_i(k_\mathrm{p})} \frac{P_i(n_i(k_\mathrm{p}))}{L_i}\right\} \quad (4)$$

where $i, j \in \mathcal{N}$, $h \in \mathcal{N}_i$; $k_\mathrm{c} = \lfloor k_\mathrm{p}/N_\mathrm{c} \rfloor$, with N_c some positive integer, is the control discrete time index (i.e. the control cycle is always a multiple of the plant sample time T_p), and L_i [*m*] is the average trip length for region i, which is assumed to be independent of time and destination, internal (inside region i) or external (to some other region j). The parameters $\theta_{ii}(k_\mathrm{p})$, $\theta^h_{ij}(k_\mathrm{p})$ reflect the route choice of drivers and are assumed to be exogenous (i.e. can be constant or time varying and they are provided by another methodology). The transfer flows $M^h_{ij}(k_\mathrm{p})$ are the minimum between the sending flow from region i (which only depends on the accumulation of the region), and the receiving capacity $C_{ih}(n_h(k_\mathrm{p}))$ [*veh/sec*] of region h. This flow capacity is a piecewise function of the accumulation $n_h(k_\mathrm{p})$ (usually modelled with two pieces, one constant value and a decreasing curve), and is introduced to prevent overflow phenomena within the regions, i.e. each region i has a maximum accumulation $n_{i,\max}$ such that

$$0 \leq n_i(k_\mathrm{p}) \leq n_{i,\max},\ \forall\ i \in \mathcal{N}. \tag{5}$$

If $n_i(k_\mathrm{p}) = n_{i,\max}$, the region reaches gridlock and all the inflows along the periphery are restricted. Finally, the control variables $u_{ih}(k_\mathrm{c})$, $\forall\ i \in \mathcal{N}$, $h \in \mathcal{N}_i$ denote the fraction of the flow that is allowed to transfer from region i to region h for the time interval $[(N_\mathrm{c}k_\mathrm{c})T_\mathrm{p}, (N_\mathrm{c}k_\mathrm{c} + N_\mathrm{c} - 1)T_\mathrm{p}]$. Physical constraints are applied to the values of the control variables as follows

$$0 \leq u_{ih}(k_\mathrm{c}) \leq 1,\ \forall\ i \in \mathcal{N}, h \in \mathcal{N}_i \tag{6}$$

but also – depending on the implementation – operational constraints of the following form might apply, i.e.,

$$|u_{ih}(k_\mathrm{c}) - u_{ih}(k_\mathrm{c} - 1)| \leq u^\mathrm{R}_{ih},\ \forall\ i \in \mathcal{N}, h \in \mathcal{N}_i. \tag{7}$$

Equations (1)–(4) are a discretized version of equations presented in Ramezani et al. (2015) and represent the traffic dynamics of an N-region urban network considering the heterogeneity effect and integrating an aggregated routing model. Note, that these equations allow the drivers to choose any arbitrary sequence of regions as their route and their path can cross region boundaries multiple times.

2.1 Nonlinear model predictive control (NMPC)

The MFD dynamics described in the previous section derive a nonlinear model that has been used in other works (Geroliminis et al. (2013); Ramezani et al. (2015)) to apply nonlinear model predictive control (NMPC). Here, we solve the same problem again to obtain a benchmark, which is then used as a comparison to the LP formulation presented later. In order to have a well defined problem and without loss of generality – since this is a nonlinear MPC problem – the following assumptions are made for formulating the problem:

- the quantities $q_{ii}(k)$, $q_{ij}(k)$ and $\theta_{ii}(k)$, $\theta^h_{ij}(k)$ are considered exogenous variables that can be measured or given by another algorithm beforehand,
- as in many similar works, the capacity constraint $C_{ih}(n_h(k_\mathrm{p}))$ in (4) is dropped, since from a control viewpoint it is not necessary; the control actions will not allow the system to operate in states close to gridlock, and this constraint is never activated inside the NMPC.

Given these two reasonable assumptions the nonlinear optimal control problem for a horizon of N_p model steps is defined as follows

$$\underset{\substack{n_{ii}(k),\ n_{ij}(k),\\ u_{ih}(\kappa)}}{\text{maximize}} \sum_{k=k_\mathrm{p}}^{k_\mathrm{p}+N_\mathrm{p}-1} \sum_{i\in\mathcal{N}} L_i \left[M_{ii}(k) + u_{ih}(\kappa)M^h_{ij}(k)\right] \tag{8}$$

subject to

$$\text{equations (1), (2), (3), (4), (5), (6), (7)}$$

$$M^h_{ij}(k) = u_{ih}(\kappa)\theta^h_{ij}(k)\frac{n_{ij}(k)}{n_i(k)}\frac{P_i(n_i(k))}{L_i} \tag{9}$$

$$n_i(k) = \sum_{j\in\mathcal{N}} n_{ij}(k) \tag{10}$$

$$k = k_\mathrm{p}, k_\mathrm{p} + 1, \ldots, k_\mathrm{p} + N_\mathrm{p} - 1,\ \kappa = \lfloor k/N_\mathrm{c} \rfloor \tag{11}$$

$$\forall\ i, j \in \mathcal{N},\ h \in \mathcal{N}_i$$

The objective function (8) tries to maximize the total production of the system for a horizon of N_p model steps (or $N_\mathrm{p}T_\mathrm{p}$ seconds). This problem can be solved in reasonable time by use of advanced nonlinear optimization toolboxes (e.g. `ipopt`[1]), and in our case serves as a benchmark for the results reported later. Note that if we want to compare the results with the LP approach described later, the objective function should be the same in both cases in order to have a fair comparison.

2.2 Linearising the problem

In the current work we derive a linear approximation of the model described in the previous section, and we formulate a linear MPC problem that can be utilized for real-time control purposes. In order to linearise the dynamic equations we assume the following simplifications and approximations:

- We introduce the model parameters for the accumulation proportions, i.e. $\alpha_{ii}(k) = n_{ii}(k)/n_i(k)$ and $\alpha_{ij}(k) = n_{ij}(k)/n_i(k)$, $i \in \mathcal{N}$, $j \in \mathcal{N}_i$. One approach, that is implemented here, is to get feedback for the values of these parameters every time that we roll the horizon. They can be estimated in real-time from measurements (e.g. using extended Kalman filter or maximum likelihood approximation) and then be kept constant for all the optimisation horizon. Note that the parameters can be time varying but they need to be exogenous signals for the MPC framework.
- New "dummy" control variables $u_{ii}(k)$ are introduced $\forall\ k = k_\mathrm{p}, k_\mathrm{p}+1, \ldots, k_\mathrm{p}+N_\mathrm{p}-1$, that restrict the trip completion rates at every region i. Although these variables are not reasonable from a physical point of view, they are required in order for the problem to be linear. A conjecture is that the solution of MPC will always result in $u_{ii}(k) = 1$, $\forall\ i \in \mathcal{N}$, $\forall\ k = k_\mathrm{p}, k_\mathrm{p} + 1, \ldots, k_\mathrm{p} + N_\mathrm{p} - 1$, but this needs to be validated through results.
- Most importantly, we introduce new decision variables

  $$f_{ii}(k) = u_{ii}(k)G_i(n_i(k))\theta_{ii}(k)\alpha_{ii}(k) \tag{12}$$

  $$f_{ih}(k) = u_{ih}(k)G_i(n_i(k))\sum_{j\in\mathcal{N}}\theta^h_{ij}(k)\alpha_{ij}(k) \tag{13}$$

  $$\forall\ i, j \in \mathcal{N},\ h \in \mathcal{N}_i$$

  that help linearise the equations. In (12)–(13) the variables $\theta_{ii}(k)$, $\theta^h_{ij}(k)$, $\alpha_{ii}(k)$, $\alpha_{ij}(k)$ are considered time varying exogenous signals and as a result the nonlinearity of the problem comes from the product of the control inputs $u_{ih}(k)$ with the MFD functions.

[1] `http://www.i2c2.aut.ac.nz/Wiki/OPTI/index.php/Solvers/IPOPT`

- To overcome this, we approximate the MFDs of the regions with piecewise affine (PWA) functions $G_i(n_i(k))$ that form a convex set (see e.g. Figure 3 for a case study with 4 regions). Each MFD can be approximated with $l = 1, 2, \ldots, N_i$ affine functions, and we denote as $G_i^l(n_i(k))$ each affine function l.

  In conclusion, the control variables have the property of being bounded (i.e. $u_{ih}(k) \in [0, 1]$, $\forall\ i \in \mathcal{N},\ h \in \mathcal{N}_i,\ k = k_\mathrm{p}, k_\mathrm{p}+1, \ldots, k_\mathrm{p}+N_\mathrm{p}-1$) and the MFDs that can be approximated by PWA functions. As a result, we are looking for an optimal solution within a convex set, and in this particular case the product can be linearised by introducing the new variables $f_{ii}(k)$, $f_{ih}(k)$ (see Gomes and Horowitz (2006) for some theoretical analysis of a similar convexification in a ramp metering control problem). Moreover, PWA approximations is a popular technique to reduce complexity and attempt to linearise nonlinear systems (see e.g. Xu et al. (2016)); alternatively someone could use robust control to solve another version of this problem, by introducing uncertain parameters (see e.g. Haddad (2015)). Finally, once the optimal solution is computed, there is a unique transformation between the new variables and the original control variables $u_{ii}(k)$, $u_{ih}(k)$. This is a modelling trick that allows us to simplify the problem without loosing any accuracy in the dynamics.

3. LINEAR MODEL PREDICTIVE CONTROL (LMPC)

The assumptions outlined above are reasonable approximations/simplifications of the nonlinear model in order to derive a linear formulation that can be used for online MPC. In this section we formulate a linear model predictive control (LMPC) problem, which does not keep track of the origin-destination information of vehicles. This dynamic model has less online data requirements (as it carries lower level of information, i.e. only state and demand trajectories d_i, instead of d_{ij}), but under certain optimization horizons can provide similar optimal solutions for the control variables (see Section 4).

3.1 LMPC without OD information

Moving one step forward with our approximation, the new model does not need to keep track of the OD information (aggregated information about each region can be sufficient for control purposes). Hence, by adding all the states n_{ij} and n_{ii} for each region i, we get a linear model that does not consider OD data, but only aggregated demands in the region level. In that case, the derived LP that approximates the original system and can be solved online is as follows

$$\underset{\substack{n_i(k),\ f_{ii}(k),\\ f_{ih}(k)}}{\text{maximize}} \sum_{k=k_\mathrm{p}}^{k_\mathrm{p}+N_\mathrm{p}-1} \sum_{i\in\mathcal{N}} L_i \left[f_{ii}(k) + f_{ih}(k) \right] \tag{14}$$

subject to

$$n_i(k+1) = n_i(k) + T_\mathrm{p} \Bigg(q_i(k) - f_{ii}(k) - \sum_{h\in\mathcal{N}_i} f_{ih}(k) + \sum_{h\in\mathcal{N}_i} f_{hi}(k) \Bigg) \tag{15}$$

$$0 \le f_{ii}(k) \le \theta_{ii}(k)\alpha_{ii}(k) G_i^l(n_i(k)) \tag{16}$$

$$0 \le f_{ih}(k) \le G_i^l(n_i(k)) \sum_{j\in\mathcal{N}} \theta_{ij}^h(k)\alpha_{ij}(k) \tag{17}$$

$$0 \le n_i(k) \le n_{i,\max} \tag{18}$$

$$k = k_\mathrm{p}, k_\mathrm{p}+1, \ldots, k_\mathrm{p}+N_\mathrm{p}-1 \tag{19}$$

$$\forall\ i, j \in \mathcal{N},\ h \in \mathcal{N}_i,\ l = 1, 2, \ldots, N_i$$

where the objective function (14) again represents the total production of the system. All the constraints of this problem are linear, and, as a consequence, the computational requirements are quite low, even for a network with many regions and large prediction horizons. Note that constraint (7) can also be applied for the first step of the LMPC without breaking the linearity, by using the following linear inequalities

$$f_{ih}(k_\mathrm{p}) \le \left(u_{ih}^{\mathrm{PR}} + u_{ih}^{\mathrm{R}} \right) G_i^{\mathrm{P}}(n_i(k_\mathrm{p})) \sum_{j\in\mathcal{N}} \theta_{ij}^h(k_\mathrm{p})\alpha_{ij}(k_\mathrm{p}) \tag{20}$$

$$f_{ih}(k_\mathrm{p}) \ge \left(u_{ih}^{\mathrm{PR}} - u_{ih}^{\mathrm{R}} \right) G_i^{\mathrm{P}}(n_i(k_\mathrm{p})) \sum_{j\in\mathcal{N}} \theta_{ij}^h(k_\mathrm{p})\alpha_{ij}(k_\mathrm{p}) \tag{21}$$

$$\forall\ i, j \in \mathcal{N},\ h \in \mathcal{N}_i$$

where u_{ij}^{R} are user defined bounds, u_{ij}^{PR} are the control commands applied to the plant in the previous control cycle, and $G_i^{\mathrm{P}}(n_i(k_\mathrm{p}))$ are the affine functions that the accumulations $n_i(k_\mathrm{p})$ of each region i belong to (this can be easily found as $n_i(k_\mathrm{p})$ are known). Besides, only the first control decision is applied to the nonlinear plant and then we roll the horizon.

4. PROOF OF CONCEPT

This section presents some simulation results obtained for the described methodology. The simulation model (plant) is the nonlinear model presented (1)–(5). The test case network is a replica of the network used in Kouvelas et al. (2017) and corresponds to a part of the CBD of Barcelona in Spain (Figure 2(a)). The network is partitioned into 4 regions in order to apply perimeter control (Figure 2(b)). Figure 3 presents the MFDs of the 4 regions from data obtained from a microsimulation model in Kouvelas et al. (2017). The red lines present the PWA approximation of the MFDs with the affine functions utilised by the LP approximation. The approximation is pretty accurate and this relaxation should not cause any problems in the optimization procedure. In that respect, the nonlinear model and PWA approximation are almost identical; note also, that this approximation can be done with many more lines than presented here without hardening significantly the computations of the LP solver (6 pieces are shown for each MFD in Figure 3, while in the LMPC 30 pieces have been used for better accuracy).

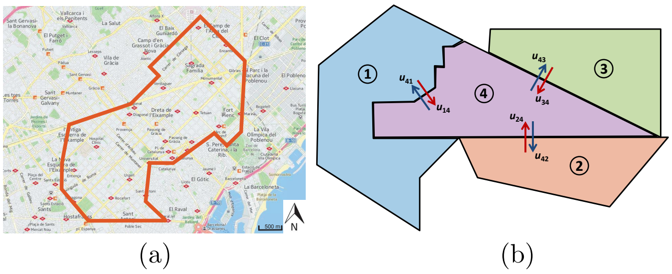

Fig. 2. Test case network: (a) map of Barcelona CBD; (b) partitioning into 4 homogeneous regions and control variables.

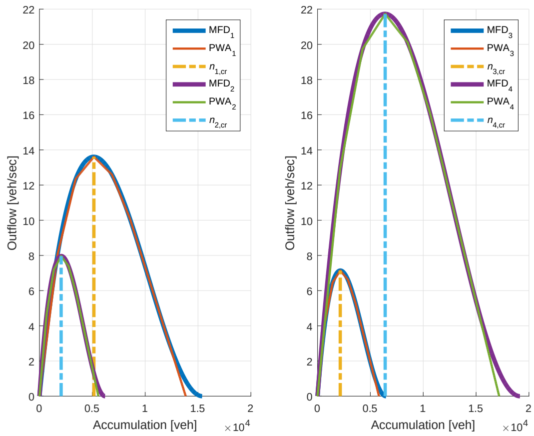

Fig. 3. MFDs for the 4 regions of the case study network (blue); piecewise affine approximation of the MFDs to be used in linear MPC (red).

First, we present a comparison of the plant (1)–(5) and the linear model presented earlier (12)–(13), (15), (18) for the no control (NC) case (i.e. $u_{ih}(k_c) = 1$, $\forall\ i \in \mathcal{N}$, $h \in \mathcal{N}_i$, $k_c = \lfloor k_p/N_c \rfloor$), which will also be our base scenario. Figure 4 presents a demand scenario (i.e. generated vehicles per time unit for all the simulation horizon) for the case study with 4 regions (e.g. 4×4 OD matrix). For this demand, Figure 5 displays the trajectories of accumulations n_{ij} for the plant and for the linear model, which actually correspond to the estimation of n_{ij} used within the LMPC framework (using n_i and α_{ij}). The prediction horizon for the linear model is 21 times higher than the sample time of the plant (e.g. $N_p = 21$). The trajectories of the accumulations n_{ij} demonstrate that the linear model can be used to approximate the original nonlinear plant (given small prediction horizons and feedback of model parameters α_{ij}). This model is a quite accurate representation of the original system, thus is appropriate to be used for the LMPC framework.

Finally, Figure 6 displays the results for the regional accumulations when the two approaches are applied to control the transferring flows between regions. The objective of the controller is to maximize the production in the network and the parameters used for this simulation are $T_p = 20sec$, $N_c = 3$, $N_p = 21$, $u_{ij}^{R} = 0.2$, $N_i = 30$, $\forall\, i, j \in \mathcal{N}$. It is clear that the controller improves the traffic states of the system, and the area between the blue and the dashed black lines (or dashed green lines) corresponds to the total delay improvement in the system. These results are quite promising, as it is demonstrated that we can

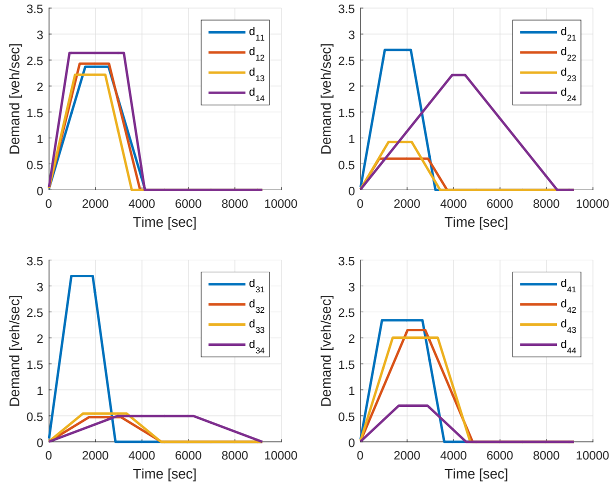

Fig. 4. Traffic demand for the four regions and all simulation horizon (4×4 OD matrix, i refers to origin and j to destination).

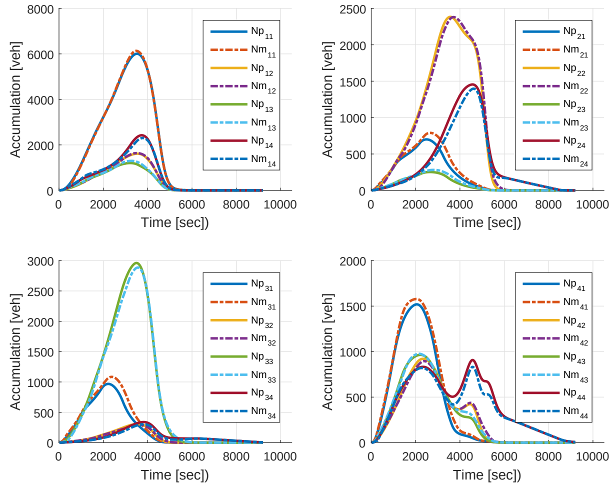

Fig. 5. Vehicle accumulations n_{ij} for the plant (solid lines) and the linear model (dashed lines) when applied for 21 steps of prediction ($N_p = 21$) and the NC case.

achieve the same level of improvement by using the linear approximation of the model described in section 3.1, which also does not require OD information. Table 1 presents the performance evaluation of the two approaches. All simulated scenarios start and finish with an empty network and serve all the demand (79143 vehicles) [2]. The total travelled distance (TTD) in all scenarios is equal to 169.55×10^3 veh·km and the total travelled time (TTT) by all vehicles in the network is computed by

$$\mathrm{TTT} = \sum_{k=0}^{K_p-1} \sum_{i \in \mathcal{N}} n_i(k) \tag{22}$$

and presented in Table 1 for each case. Both NMPC and LMPC achieve a substantial improvement in terms of network mean speed (NMS=TTD/TTS) and TTT (around 20–30%).Their performance is quite similar, however the LMPC approach has less data requirements and it is a LP formulation that guarantees convergence and optimality.

[2] Note that for very small accumulation values at the beginning and end of simulation NC is applied in all cases.

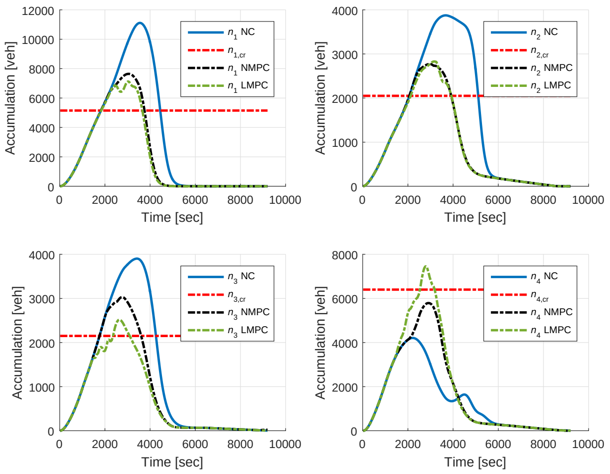

Fig. 6. Vehicle accumulations n_i for the 4 regions for NC case (blue lines), NMPC (dashed black lines), and LMPC (dashed green lines) when applied for 21 steps of prediction ($N_\mathrm{p} = 21$).

Table 1. Performance of different approaches.

Criteria	NC	NMPC	(%)	LMPC	(%)
TTS (veh·h)$\times 10^3$	17.73	13.76	-22.39	13.58	-23.41
NMS (km/h)	9.57	12.32	28.74	12.49	30.51

5. CONCLUSION

A LP formulation is derived for solving the perimeter control problem in multi-region cities. The originally non-linear system is relaxed and approximated by a simplified linear model that under certain assumptions can track the behaviour of the multi-region system. The new model requires less information in terms of real-time measurements (e.g. traffic states, OD demands), and, because it is formulated as a LP, it guarantees optimality and fast convergence of the solver. The simulation results need to be further investigated in order to assess the goodness of the solution obtained through the linear MPC to the benchmark, which consists of the nonlinear MPC problem with full information about demands and state feedback measurements. Note that a field implementation of the benchmark approach would require real-time measurements (or estimates) of n_{ij} and d_{ij}, while the LMPC version requires only region-based measurements (i.e. n_i and d_i) and not detailed origin-destination data.

Future work will deal with the development of a more solid methodology for estimating the model parameters α_{ij} (e.g. online extended Kalman filter), as the values of these parameters are crucial for the optimization horizon. These parameters can be estimated at every control cycle and then considered constant for all the optimization horizon of the MPC and this could not be sufficient for the improvement of the system. Simple estimation/prediction techniques can be used to enhance the knowledge for this parameters and help the convex problem to track the nonlinear dynamics in a better way. Investigations about different convex objective functions for the MPC is also another research topic. The proposed methodology needs to be evaluated for different realistic objective functions and demand profiles. Finally, another research direction is to use perimeter control as a first-level controller in cities (as it deals with zone interactions) and develop a second-level of distributed control (e.g. Kouvelas et al. (2014)) for optimizing locally. The combination of the two provides a hierarchical control scheme that could potentially be more efficient in alleviating traffic congestion in cities, but this needs also to be further investigated.

ACKNOWLEDGEMENTS

The authors are grateful to Işik İlber Sırmatel for the valuable comments during the development of this work.

REFERENCES

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.

Gomes, G. and Horowitz, R. (2006). Optimal freeway ramp metering using the asymmetric cell transmission model. *Transportation Research Part C*, 14(4), 244–262.

Haddad, J. (2015). Robust constrained control of uncertain macroscopic fundamental diagram networks. *Transportation Research Part C*, 59, 323–339.

Hajiahmadi, M., Haddad, J., De Schutter, B., and Geroliminis, N. (2015). Optimal hybrid perimeter and switching plans control for urban traffic networks. *IEEE Transactions on Control Systems Technology*, 23(2), 464–478.

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*, 46(10), 1393–1403.

Kouvelas, A., Lioris, J., Fayazi, S.A., and Varaiya, P. (2014). Maximum pressure controller for stabilizing queues in signalized arterial networks. *Transportation Research Record*, (2421), 133–141.

Kouvelas, A., Saeedmanesh, M., and Geroliminis, N. (2015). Feedback perimeter control for heterogeneous urban networks using adaptive optimization. In *18th IEEE International Conference on Intelligent Transportation Systems*, 882–887.

Kouvelas, A., Saeedmanesh, M., and Geroliminis, N. (2017). Enhancing model-based feedback perimeter control with data-driven online adaptive optimization. *Transportation Research Part B*, 96, 26–45.

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

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

Xu, J., van den Boom, T., De Schutter, B., et al. (2016). Model predictive control for continuous piecewise affine systems using optimistic optimization. In *American Control Conference, 2016*, 4482–4487.

Yildirimoglu, M., Ramezani, M., and Geroliminis, N. (2015). Equilibrium analysis and route guidance in large-scale networks with MFD dynamics. *Transportation Research Part C*, 59, 404–420.