

Optimal Perimeter Control for Two Urban Regions With Macroscopic Fundamental Diagrams: A Model Predictive Approach

Nikolas Geroliminis, Jack Haddad, and Mohsen Ramezani

Abstract—Recent analysis of empirical data from cities showed that a macroscopic fundamental diagram (MFD) of urban traffic provides for homogenous network regions a unimodal low-scatter relationship between network vehicle density and network space-mean flow. In this paper, the optimal perimeter control for two-region urban cities is formulated with the use of MFDs. The controllers operate on the border between the two regions and manipulate the percentages of flows that transfer between the two regions such that the number of trips that reach their destinations is maximized. The optimal perimeter control problem is solved by model predictive control, where the prediction model and the plant (reality) are formulated by MFDs. Examples are presented for different levels of congestion in the regions of the city and the robustness of the controller is tested for different sizes of error in the MFDs and different levels of noise in the traffic demand. Moreover, two methods for smoothing the control sequences are presented. Comparison results show that the performances of the model predictive control are significantly better than a “greedy” feedback control. The results in this paper can be extended to develop efficient hierarchical control strategies for heterogeneously congested cities.

Index Terms—Macroscopic fundamental diagrams (MFDs), model predictive control (MPC), perimeter control.

I. INTRODUCTION

EFFICIENT monitoring and traffic management of large-scale urban networks still remain a challenge for both traffic researchers and practitioners. A large urban network mainly consists of the following two elements: 1) urban links and, 2) signalized intersections. Modeling the traffic flow dynamics of each element in a large urban network with a large number of links and intersections is a complex task. We have to model the evolution of queues at each signalized intersection and account for the queue dynamic interactions between adjacent intersections, i.e., capturing the dynamics of propagation and

spillover of queues because of high demand. Even if this task is completed, a centralized control approach would be a very challenging task, not only because of the computational complexity but also because users might change their travel patterns (e.g., time of departure, route choice, and mode choice). Hence, instead of this micromodeling approach, the macroscopic fundamental diagram (MFD) aims at simplifying the micromodeling task of the urban network, where the collective traffic flow dynamics of subnetworks capture the main characteristics of traffic congestion, such as the evolution of space-mean flows and densities in different regions of the city. MFD can be utilized to introduce elegant control strategies to improve mobility and decrease delays in large urban networks, which local strategies cannot achieve.

The MFD of urban traffic provides for different network regions a unimodal low-scatter relationship between network vehicle density (veh/km) and network space-mean flow or outflow (veh/h) if congestion is roughly homogeneous in the region. Alternatively, the MFD links *accumulation*, which is defined as the number of vehicles in the region, and *trip completion flow*, which is defined as the output flow of the region. Network flow or trip completion flow increases with density or accumulation up to a critical point, while additional vehicles in the network cause strong reductions in the flow. The first theoretical proposition of such a physical model was developed in [1], while similar approaches were also initiated in [2] and [3]. The physical model of MFD was observed with dynamic features in congested urban networks in Yokohama, Japan, in [4]. This paper showed the following two important properties of MFD that can be utilized for management and control purposes: 1) some urban regions approximately exhibit an MFD, and 2) the shape of the MFD is not very sensitive to different demand patterns. Property 1 is important for monitoring purposes, because flow can easily be observed with different types of sensors, whereas outflow is more difficult. Property 2 is important for control purposes, because efficient active traffic management schemes can be developed without a detailed knowledge of origin–destination (OD) tables. Other investigations of MFD using empirical or simulated data can be found in [5]–[8] and other papers, whereas routing strategies that are based on MFD can be found in [9].

Recent studies [8], [10], [11] have shown that networks with heterogeneous distribution of density exhibit network flows that are smaller than those that approximately meet homogeneity conditions (low spatial variance of link density), particularly

Manuscript received February 13, 2012; revised June 21, 2012; accepted August 9, 2012. This work was supported in part by the Swiss National Science Foundation under Grant 200021-13016. The Associate Editor for this paper was B. De Schutter. (Corresponding author: Nikolas Geroliminis.)

The authors are with the Urban Transport Systems Laboratory, School of Architecture, Civil and Environmental Engineering, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland (e-mail: nikolas.geroliminis@epfl.ch; jack.haddad@epfl.ch; mohsen.ramezani@epfl.ch).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TITS.2012.2216877

for high network densities. Moreover, networks with a small variance of link densities have a *well-defined* MFD, i.e., low scatter of flows for the same densities. One possible solution for heterogeneous networks is that they might be partitioned into a number of more homogeneous regions with small variances of link densities, as each region will have a well-defined MFD. See [12] for more information on network partitioning. In [7] and [13], strong hysteresis phenomena in freeways that might not disappear after partitioning is shown. Nevertheless, the work in [12] showed that urban networks can be partitioned in a way that decreases the degree of heterogeneity within clusters. Partitions should not have a very small size, because the law of large numbers will not apply, and high scatter might exist in MFDs. In addition, a large number of partitions will not allow the development of simple control strategies as shown in this paper, because control might change the route choices, and detailed ODs might be needed.

Management and control of multiregion MFDs systems can improve urban mobility, prevent overcrowding, and relieve congestion in cities. The optimal control policy was derived for a single MFD system in [3]. The main logic behind this policy is that it aims at decreasing inflows in regions with high densities of destinations and points in the decreased part of an MFD and manage the accumulation to maintain the flow in the city at its maximum. However, in case of multiregion cities with multiple centers of congestion and/or attraction, control policies are more complicated and not well understood. For stability analysis of controlling two urban regions, see [14].

Because of the scatter in the MFD, mainly in the congested regime, errors are expected between the MFD model and the plant (reality). Therefore, an optimal open-loop control for the multiregion MFDs system would be a suboptimal solution compared with the optimal closed-loop control. The closed-loop control takes into account errors between the model and the plant by utilizing feedback-monitored information. Furthermore, the closed-loop control can tackle disturbances for which the model was not designed, e.g., noise in the traffic demand. The optimal closed-loop control is obtained by implementing the model predictive control (MPC) framework. A historical survey for industrial applications of MPC can be found in [15], while theoretical issues of MPC can be found in [16]–[20].

MPC is a receding horizon scheme, where, at each time step, an optimal open-loop of the problem with finite horizon is optimized; then, only the first controller is applied to the plant, and the procedure is again carried out. A receding horizon framework has been used for optimization in different traffic control problems, for example, ramp metering of freeway networks in [21] and [22], variable speed limits and route guidance for freeway networks in [23] and [24], signal control for large-scale urban networks in [25]–[27], and mixed urban and freeway networks in [28]. The open-loop optimization in the traffic MPC models, for example, in [23] and [24], uses a *direct simultaneous method* to transcribe it into a finite-dimensional nonlinear programming through the discretization of both control and state variables, whereas in [22] and [26], a feasible direction algorithm is utilized to solve the open-loop optimization problem. Overviews of different control applica-

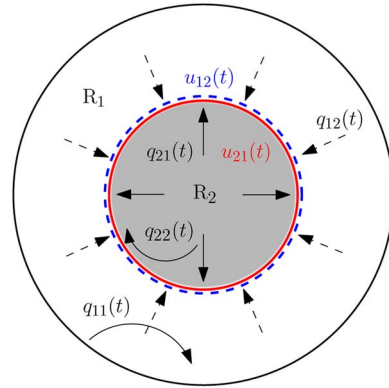


Fig. 1. Two-region MFDs system. Two regions R_1 and R_2 with four traffic demands $q_{11}(t)$, $q_{12}(t)$, $q_{21}(t)$, and $q_{22}(t)$ and two perimeter controllers $u_{12}(t)$ and $u_{21}(t)$.

tions in transportation problems can be found, e.g., in [29] and [30].

In this paper, the optimal perimeter control problem for two-region urban cities is formulated, where the dynamic equations are modeled according to their MFDs. Moreover, the optimal control solution is obtained by applying the MPC framework. The open-loop optimal control problem is solved using a *direct sequential method* that discretizes *only* the control variables with piecewise constant controls, whereas the state variables are continuous and integrated using the state-of-the-art methods for ordinary differential equation (ODE) solvers.

This paper is organized as follows. The control problem for a two-region MFDs system is presented in Section II. Then, in Section III, MPC is formulated, the parameters are tuned, and the control laws of a greedy controller (GC) are presented. Comparison results of case study examples are presented in Section IV, showing the performance differences between MPC and the GC. Finally, two different methods are introduced in Section V to smooth the control sequences.

II. TWO-REGION MACROSCOPIC FUNDAMENTAL DIAGRAMS SYSTEM

In this paper, a heterogeneous traffic network that can be partitioned into two homogeneous regions is considered. A traffic network for a two-region system is schematically shown in Fig. 1. Two regions R_i , $i = 1, 2$, where each region has a well-defined MFD are given as follows: 1) the periphery of the center R_1 , and 2) the city center R_2 . Note that the geographical relative position of these regions does not affect the dynamics of the problem; for example, it can be two regions next to each other. An endogenous traffic demand is defined as a flow in which its origin and destination are the same region, whereas the origin and destination of an exogenous traffic demand are not the same. For the two-region system, there are two endogenous traffic demands in R_1 , denoted by $q_{11}(t)$ (veh/s), and in R_2 , denoted by $q_{22}(t)$ (veh/s), and two exogenous traffic demands generated in R_1 and R_2 with destination to R_2 and R_1 , denoted by $q_{12}(t)$ and $q_{21}(t)$ (veh/s), respectively. Corresponding to the endogenous and exogenous traffic demands, four accumulation states are used to model the dynamic equations, $n_{ij}(t)$ (veh), $i, j = 1, 2$, where $n_{ij}(t)$ is the total number of

vehicles in R_i with destination to R_j at time t . Let us denote $n_i(t)$ (veh) as the accumulation or the total number of vehicles in R_i at time t , i.e., $n_i(t) = \sum_j n_{ij}(t)$.

MFD is defined by $G_i(n_i(t))$ (veh/s), which is the trip completion flow for region i at $n_i(t)$. The trip completion flow for region i is the sum of transfer flows, i.e., trips from i with destination j , $i \neq j$, plus the internal flow, i.e., trips from i with destination i . The transfer flow from i with destination to j is calculated corresponding to the ratio between accumulations, i.e., $M_{ij}(t) = n_{ij}(t)/n_i(t) \cdot G_i(n_i(t))$, $i \neq j$, whereas the internal flow from i with destination to i is calculated by $M_{ii}(t) = n_{ii}(t)/n_i(t) \cdot G_i(n_i(t))$. These relationships assume that the trip lengths for all trips within a region (internal or external) are similar, i.e., the distance traveled per vehicle inside a region is independent of the origin and destination of the trip. For a description of different cases, refer to [31], which will not alter the methodology. Simulation and empirical results [4] show that the shape of MFD can be approximated by a nonsymmetric unimodal curve skewed to the right, i.e., the critical density, which maximizes the network flow, is smaller than half the jammed density. Thus, we utilize a third-order function of $n_i(t)$, e.g., $G_i(n_i(t)) = a_i \cdot n_i(t)^3 + b_i \cdot n_i(t)^2 + c_i \cdot n_i(t)$, where a_i , b_i , and c_i are the estimated parameters.

In our formulated problem, the perimeter controllers, denoted by $u_{12}(t)$ and $u_{21}(t)$ (–), are introduced on the border between the two regions, as shown in Fig. 1, where the purpose is to control the transfer flows between the two regions such that the total number of vehicles that complete their trips and reach their destinations in the two-region MFDs system is maximized. Because the perimeter controllers exist only on the border between the two regions, the internal flows cannot be controlled or restricted, whereas the transfer flows are controlled by the controllers such that only a ratio transfers at time t . The perimeter controllers $u_{12}(t)$ and $u_{21}(t)$, where $0 \leq u_{12}(t) \leq 1$ and $0 \leq u_{21}(t) \leq 1$, are, respectively, the ratio of the transfer flow that transfers from R_1 to R_2 and from R_2 to R_1 at time t . It is also assumed that these controllers will not change the shape of the MFDs. Implementations of the controllers in real networks are discussed in the Conclusion of this paper.

The criterion is to maximize the output of the traffic network, i.e., the number of vehicles that complete their trips and reach their destinations. Therefore, the two-region MFDs control problem with four state variables is formulated as follows (similar to [32]):

$$J = \max_{u_{12}(t), u_{21}(t)} \int_{t_0}^{t_f} [M_{11}(t) + M_{22}(t)] dt \quad (1)$$

subject to

$$\frac{dn_{11}(t)}{dt} = q_{11}(t) + u_{21}(t) \cdot M_{21}(t) - M_{11}(t) \quad (2)$$

$$\frac{dn_{12}(t)}{dt} = q_{12}(t) - u_{12}(t) \cdot M_{12}(t) \quad (3)$$

$$\frac{dn_{21}(t)}{dt} = q_{21}(t) - u_{21}(t) \cdot M_{21}(t) \quad (4)$$

$$\frac{dn_{22}(t)}{dt} = q_{22}(t) + u_{12}(t) \cdot M_{12}(t) - M_{22}(t) \quad (5)$$

$$0 \leq n_{11}(t) + n_{12}(t) \quad (6)$$

$$0 \leq n_{21}(t) + n_{22}(t) \quad (7)$$

$$n_{11}(t) + n_{12}(t) \leq n_{1,\text{jammed}} \quad (8)$$

$$n_{21}(t) + n_{22}(t) \leq n_{2,\text{jammed}} \quad (9)$$

$$u_{\min} \leq u_{12}(t) \leq u_{\max} \quad (10)$$

$$u_{\min} \leq u_{21}(t) \leq u_{\max} \quad (11)$$

$$n_{11}(t_0) = n_{11,0}; n_{12}(t_0) = n_{12,0} \quad (12)$$

$$n_{21}(t_0) = n_{21,0}; n_{22}(t_0) = n_{22,0}$$

where t_f (s) is the final time, $n_{ij,0}$, $i, j = 1, 2$, are the initial accumulations at t_0 , $n_{1,\text{jammed}}$ and $n_{2,\text{jammed}}$ (veh) are the accumulations at the jammed density in R_1 and R_2 , respectively, and u_{\min} and u_{\max} are the lower and upper bounds for $u_{12}(t)$ and $u_{21}(t)$, respectively. Recall that $M_{ij}(t) = n_{ij}(t)/n_i(t) \cdot G_i(n_i(t))$, $i, j = 1, 2$. Equations (2)–(5) are the conservation of mass equations for $n_{ij}(t)$, whereas (6)–(9) are the lower and upper bound constraints on accumulations in R_1 and R_2 .

III. MODEL PREDICTIVE CONTROL FOR THE TWO-REGION MACROSCOPIC FUNDAMENTAL DIAGRAMS PROBLEM

The two-region MFDs problem (1)–(12) aims at finding the perimeter control inputs, i.e., ratios of transfer flows of R_1 and R_2 , that maximize the number of vehicles that complete their trips (reach their destinations). This problem is an optimal control problem with a nonlinear objective function (1) and dynamic equations (2)–(5), inequality state constraints (6)–(9), control constraints (10) and (11), and initial states (12). Moreover, errors are expected in the modeling because of the scatter in the MFDs, mainly in the congested regime and of the demand profile. Therefore, the optimal control problem is solved by applying the MPC approach, which can handle the state and control constraints and the errors in the MFDs modeling. Furthermore, MPC is a real-time implementable solution that can be utilized for real-time urban traffic applications.

MPC is a form of rolling horizon control in which the current control inputs are obtained by solving a finite-horizon open-loop optimal control problem at each time step, with a current state feedback from the plant being the initial state of the model, see Fig. 2. The open-loop optimization problem yields a sequence of optimal control inputs after several iterations of solving nonlinear programming, and the first control action in this sequence is applied to the plant, then the procedure is carried out again.

This scheme of feedback control, i.e., the feedback loop of states from the plant to the model as initial states for the optimization, can handle errors between the prediction model and the plant.

A. Two-Region MFDs Prediction Model and Optimization Problem

The MPC controller obtains the optimal control sequence for the current horizon by solving an optimization problem formulated with the prediction model; see the bottom of Fig. 2.

The prediction model used in the MPC scheme is formulated with (2)–(5). The dynamic equations predict the evolution of accumulations for the two regions with MFDs, given the initial

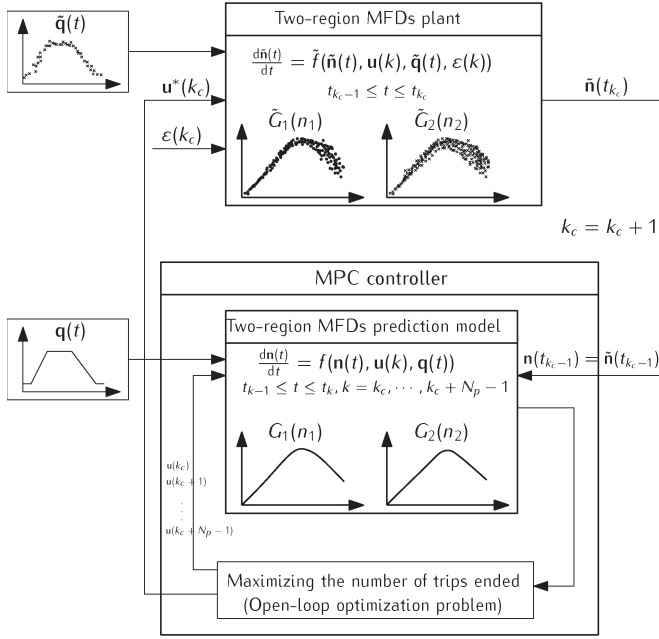


Fig. 2. MPC scheme for the two-region MFDs system.

accumulations and future values of perimeter control inputs and demand.

In this paper, we follow the direct methods for solving the optimization problem (other solution methods include dynamic programming and indirect methods). The direct methods are most commonly used because of their applicability and robustness, where their basic principle is “first discretize and then optimize”. These methods can handle inequality constraints and use state-of-the-art methods for nonlinear problem solvers.

The open-loop optimal control problem is solved using the direct sequential method, also referred to as single shooting or control vector parameterization in the literature, e.g., [33] and [34]. The direct sequential method transcribes the open-loop optimal control problem into a finite-dimensional nonlinear problem through the discretization of the control variables *only* with piecewise constant controls, whereas the ODEs are embedded in the nonlinear problem, i.e., numerical integration is used between the time steps. A schematic description of the direct sequential method is shown in Fig. 3. Note the continuous dynamics of the state variables $n_{ij}(t)$, $i, j = 1, 2$. Let N_p (–) be the finite-dimensional horizon, which starts from the current control step k_c . At each discrete time step k , $k_c \leq k \leq k_c + N_p - 1$, there are two perimeter control inputs, $u_{12}(k)$ and $u_{21}(k)$, which are assumed to be constant during the time period $t_{k-1} \leq t \leq t_k$. For online computational complexity, the number of control inputs that should be optimized are reduced to a horizon that is smaller than N_p , called the control horizon N_c , where $N_c \leq N_p$. The rest of the control variables, $u_{12}(k)$ and $u_{21}(k)$ for $k_c + N_c \leq k \leq k_c + N_p - 1$, are assumed to be equal to the control inputs at the end of the control horizon.

Following the direct sequential method, the control vector is discretized, and the two-region MFDs optimal control problem (1)–(12) is approximated by a finite-dimensional nonlinear programming problem in the piecewise constant control inputs. First, the equations of the prediction model (2)–(5) are rewritten

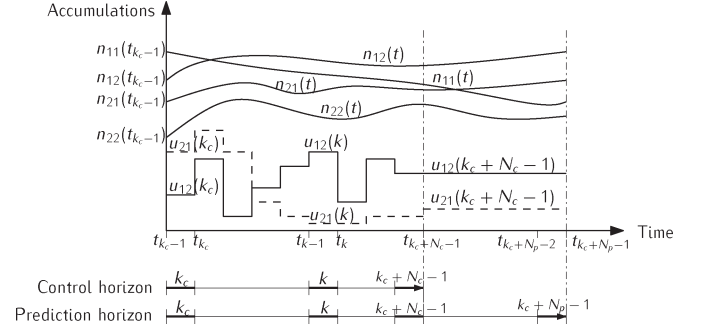


Fig. 3. Direct sequential method for solving the open-loop optimization problem.

in a compact form with *discrete control variables* at time step k_c with finite-dimensional N_p as follows:

$$\frac{d\mathbf{n}(t)}{dt} = f(\mathbf{n}(t), \mathbf{u}(k), \mathbf{q}(t)) \quad t_{k-1} \leq t \leq t_k \quad k = k_c, k_c + 1, \dots, k_c + N_p - 1 \quad (13)$$

where $\mathbf{n}(t) = [n_{11}(t), n_{12}(t), n_{21}(t), n_{22}(t)]^T$, $\mathbf{q}(t) = [q_{11}(t), q_{12}(t), q_{21}(t), q_{22}(t)]^T$, and $\mathbf{u}(k) = [u_{12}(k), u_{21}(k)]^T$. Then, the Lagrange form (1) is transferred into the Mayer form by introducing an additional state variable $z(t)$ and an additional differential equation $dz(t)/dt$. Moreover, the path constraints (6)–(9) must hold for all t (continuous variable), where $t_0 \leq t \leq t_f$; hence, the number of constraints would be infinite. However, several methods are efficient in dealing with path constraints in the sequential method, e.g., transcription as integral constraints. The optimization problem is now formulated as follows:

$$\min_{\mathbf{u}(k_c), \mathbf{u}(k_c+1), \dots, \mathbf{u}(k_c+N_p-1)} -z(t_{k_c+N_p-1}) \quad (14)$$

subject to

$$\frac{d\mathbf{n}(t)}{dt} = f(\mathbf{n}(t), \mathbf{u}(k), \mathbf{q}(t)) \quad (15)$$

$$\frac{dz(t)}{dt} = M_{11}(t) + M_{22}(t) \quad (16)$$

$$\mathbf{u}_{\min} \leq \mathbf{u}(k) \leq \mathbf{u}_{\max} \quad (17)$$

where $t_{k-1} \leq t \leq t_k \quad k = k_c, k_c + 1, \dots, k_c + N_p - 1$

$$\mathbf{u}(k) = \mathbf{u}(k_c + N_c - 1) \quad k = k_c + N_c, \dots, k_c + N_p - 1 \quad (18)$$

$$\sum_{k=k_c}^{k_c+N_p-1} \int_{t_{k-1}}^{t_k} \max\{0; -n_{11}(t) - n_{12}(t)\}^2 dt \leq \epsilon \quad (19)$$

$$\sum_{k=k_c}^{k_c+N_p-1} \int_{t_{k-1}}^{t_k} \max\{0; -n_{21}(t) - n_{22}(t)\}^2 dt \leq \epsilon \quad (20)$$

$$\sum_{k=k_c}^{k_c+N_p-1} \int_{t_{k-1}}^{t_k} \max\{0; n_{11}(t) + n_{12}(t) - n_{1,\text{jam}}\}^2 dt \leq \epsilon \quad (21)$$

$$\sum_{k=k_c}^{k_c+N_p-1} \int_{t_{k-1}}^{t_k} \max\{0; n_{21}(t) + n_{22}(t) - n_{2,\text{jam}}\}^2 dt \leq \epsilon \quad (22)$$

where $\mathbf{u}_{\min} = [u_{\min}, u_{\min}]^T$, and $\mathbf{u}_{\max} = [u_{\max}, u_{\max}]^T$. Note that the path constraints (6)–(9) are reformulated as integral constraints (19)–(22), respectively, with relaxation, where $\epsilon > 0$ is a small nonnegative constant.

B. Two-Region MFDs Plant

The dynamic equations of the two-region MFDs plant differ from the prediction model (2)–(5), because they include *errors* in the MFDs for both regions and *noise* in the traffic demand, see the top of Fig. 2 for an illustration. Hence, the evolutions of the accumulations over time are not the same for the prediction model and the plant, which are considered to have different magnitudes and profile (biased and unbiased). Thus, the plant and the model accumulation, MFD, and demand can significantly differ.

1) *Errors in MFDs*: Let us denote the MFDs with errors for R_1 and R_2 , by \tilde{G}_1 and \tilde{G}_2 , respectively. The errors in the MFDs result in errors in accumulations of the plant, which are denoted by $\tilde{n}_{ij}(t)$, $i, j = 1, 2$, to distinguish them from the accumulations of the prediction model $n_{ij}(t)$. Note that $\tilde{n}_i(t) = \sum_j \tilde{n}_{ij}(t)$. The variance of the MFD increases with the accumulations in the region, as described in [10]; see the top of Fig. 2. Reasons for this variance are asymmetric OD and route choices, which increase the heterogeneity in the distribution of congestion within a region. It is assumed that the variance of the MFD is uniformly distributed, where the error at step k is calculated at the time instant t_{k-1} as

$$\varepsilon(\tilde{n}_1(t_{k-1})) \sim U(-\alpha_1 \cdot \tilde{n}_1(t_{k-1}), \alpha_1 \cdot \tilde{n}_1(t_{k-1})) \quad (23)$$

$$\varepsilon(\tilde{n}_2(t_{k-1})) \sim U(-\alpha_2 \cdot \tilde{n}_2(t_{k-1}), \alpha_2 \cdot \tilde{n}_2(t_{k-1})) \quad (24)$$

where α_1 and α_2 (1/s) are given parameters. It is assumed that the errors $\varepsilon(\tilde{n}_1(t_{k-1}))$ and $\varepsilon(\tilde{n}_2(t_{k-1}))$ (veh/s) are constant during the time step $t_{k-1} \leq t \leq t_k$; therefore, the MFDs of the plant \tilde{G}_1 and \tilde{G}_2 for $t_{k-1} \leq t \leq t_k$ are

$$\tilde{G}_1(\tilde{n}_1(t)) = G_1(\tilde{n}_1(t)) + \varepsilon(\tilde{n}_1(t_{k-1})) \quad (25)$$

$$\tilde{G}_2(\tilde{n}_2(t)) = G_2(\tilde{n}_2(t)) + \varepsilon(\tilde{n}_2(t_{k-1})). \quad (26)$$

2) *Unbiased and Biased Noise in the Demand*: The following two different types of noise in the demand are considered: 1) unbiased noise with Gaussian distribution and 2) biased noise with a sudden jump in the demand profile for a time period. In both cases, let us denote the traffic demand $\mathbf{q}(t)$ with noise as $\tilde{\mathbf{q}}(t) = [\tilde{q}_{11}(t), \tilde{q}_{12}(t), \tilde{q}_{21}(t), \tilde{q}_{22}(t)]^T$. Unbiased demand noise represents random and recurrent variations of the demand from day to day because of travel patterns, whereas biased demand noises might represent cases of nonrecurrent events (e.g., special events or accidents).

The unbiased noise in the demand is assumed to have Gaussian distribution as follows:

$$\tilde{q}_{ij}(t) = \max(q_{ij}(t) + \mathcal{N}(0, \sigma_{ij}^2), 0) \quad (27)$$

where $i, j = 1, 2$, and σ_{ij}^2 (veh²/s²) is the variance for the traffic demand $q_{ij}(t)$.

Substituting the MFDs with errors (25)–(26) and the demand with noises $\tilde{\mathbf{q}}(t)$ in the dynamic equations (2)–(5), we get the two-region MFDs plant in a compact form (see also Fig. 2) as

$$\frac{d\tilde{\mathbf{n}}(t)}{dt} = \tilde{f}(\tilde{\mathbf{n}}(t), \mathbf{u}(k), \tilde{\mathbf{q}}(t), \varepsilon(k)) \quad (28)$$

where $\varepsilon(k) = [\varepsilon(\tilde{n}_1(t_{k-1})), \varepsilon(\tilde{n}_2(t_{k-1}))]^T$.

C. Greedy Controller (GC)

To investigate and estimate the performance of the MPC controller, comparison results are done with a GC for different levels and types of errors. GC is a state feedback control in which its policy is determined by the current accumulations $n_1(t)$ and $n_2(t)$. Let $n_{1,cr}$ and $n_{2,cr}$ (veh) be the accumulations that maximize G_1 and G_2 , respectively. GC is designed according to the following policy: if both regions are uncongested, i.e., $n_1(t) \leq n_{1,cr}$ and $n_2(t) \leq n_{2,cr}$, then both controllers should maximize the transfer flows, and therefore, $[u_{12}(t), u_{21}(t)] = [u_{\max}, u_{\max}]$. If one region is congested and the other region is uncongested, i.e., $n_1(t) \leq n_{1,cr}$ and $n_2(t) > n_{2,cr}$, or $n_1(t) > n_{1,cr}$ and $n_2(t) \leq n_{2,cr}$, then the controllers should minimize the transfer flow to the congested region and maximize the transfer flow to the uncongested region. If both regions are congested, i.e., $n_1(t) > n_{1,cr}$ and $n_2(t) > n_{2,cr}$, then the controllers should minimize the transfer flow to the “more congested” region and maximize the transfer flow to the “less congested” region; for example, if $n_1(t)/n_{1,jam} > n_2(t)/n_{2,jam}$, then R_1 is more congested than R_2 , and therefore, $[u_{12}(t), u_{21}(t)] = [u_{\max}, u_{\min}]$. The GC law is summarized in Table I.

D. Tuning the Prediction and Control Horizon Parameters

The performance of the MPC controller is affected by the prediction horizon N_p and the control horizon N_c . The prediction horizon N_p should be large enough such that the model can accurately predict the accumulations of the plant corresponding to the control inputs. Increasing the prediction horizon improves the performances of the MPC controller; however, a large N_p increases the optimization computing time, which may add some barriers for online implementation, i.e., the control actions cannot be implemented in the current step if the computing time that corresponds to a large N_p is larger than the time duration of the control time step. Similar considerations with regard to the tradeoff between computation complexity and results should accurately be done for the control horizon N_c .

The perimeter controllers can be actuated by signalized intersections that are placed in the border between the two regions of the urban network, i.e., the perimeter control sequences can be applied by choosing appropriate timing plans for the signalized intersections. The effect of perimeter control to the rest of the network and its MFDs will be discussed later. Let us assume that the signalized intersections have a fixed common cycle length, e.g., is equal to 60 (s). Then, the time duration of the time step k_c is set to be equal to the length of the cycle, i.e., $t_k - t_{k-1} = 60$. This duration is much larger than the time

TABLE I
GCS $u_{12}(t)$ AND $u_{21}(t)$ POLICY

$[u_{12}(t), u_{21}(t)]$	$n_1(t) \leq n_{1,cr}$	$n_1(t) > n_{1,cr}$
$n_2(t) \leq n_{2,cr}$	$[u_{max}, u_{max}]$	$[u_{max}, u_{min}]$
$n_2(t) > n_{2,cr}$	$[u_{min}, u_{max}]$	$[u_{max}, u_{min}]$ if $\frac{n_1(t)}{n_{1,jam}} > \frac{n_2(t)}{n_{2,jam}}$ $[u_{min}, u_{max}]$ otherwise

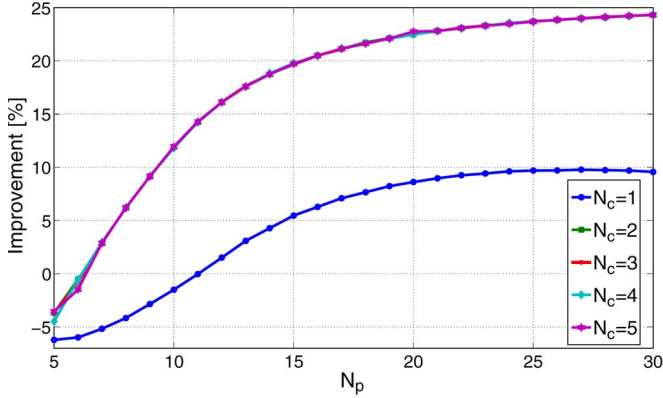


Fig. 4. Tuning parameters N_p and N_c for the MPC controller.

which is needed to solve the open-loop optimization problem (a few seconds).

Tuning the MPC parameters N_c and N_p is done for several examples, and similar results have been obtained. In the following discussion, the tuning analysis is presented only for one of the case study examples, as shown in Fig. 4. More information for the case study examples is presented in the next section.

The MPC parameters N_p and N_c are tuned according to the relative improvement of the trip completed corresponding to the MPC controller compared with the GC, i.e., improvement [%] = $(J_{MPC} - J_{GC})/J_{GC}$, where J is computed according to (1). As shown in Fig. 4, the improvement of the results decreases as the prediction horizon N_p increases; however, for $N_p \geq 20$, only a minor improvement is achieved. The MPC controller is less sensitive to the control horizon N_c , where $N_c \geq 2$ yields similar results for trip completion. It is shown that $N_c = 1$ yields a low MPC performance, because the open-loop optimization problem is more constrained. Therefore, the parameters are set as $N_p = 20$ and $N_c = 2$ for all subsequent case study examples. Note that, for a small prediction horizon $N_p < 6$, the MPC controller does not perform well compared with the GC.

IV. CASE STUDY EXAMPLES

In this section, results of several case study examples are presented to explore the features of the MPC controller under different conditions. The examples aim at examining the efficiency of the MPC controller in congested and uncongested regimes, which may vary with time because of variations in the demand and the MFDs. Hence, examples with different levels of demand and sizes of MFDs are presented. Furthermore, the robustness of the MPC and GC are examined by introducing

different uncertainties, i.e., different levels of error in MFD, and noises in the demand. For all examples presented in this paper, the selected MPC parameters are $N_p = 20$ and $N_c = 2$, the lower bound $u_{min} = 0.1$, and the upper bound $u_{max} = 0.9$. In examples 1–4, the MFDs for both regions are the same, where $G_i(n_i(t)) = (1.4877 \cdot 10^{-7} \cdot n_i^3 - 2.9815 \cdot 10^{-3} \cdot n_i^2 + 15.0912 \cdot n_i)/3600$, $i = 1, 2$, $n_{1,cr} = n_{2,cr} = 3400$ (veh), $G_1(n_{1,cr}) = G_2(n_{2,cr}) = 6.3$ (veh/s), and $n_{1,jam} = n_{2,jam} = 10000$ (veh). This shape is consistent with the MFD observed in Yokohama, see [4]. Note that the shape of the MFD is not predetermined in the problem formulation. In example 5, the MFD of R_1 is increased by 25%, as shown in Fig. 8(d), to test the control problem with two different MFDs.

In example 1, both regions R_1 and R_2 are initially congested, i.e., the initial accumulations $n_1(t_0) = 5400$ and $n_2(t_0) = 4000$ are in the decreasing part of the MFD, where $n_2(t_0)$ is 18% larger than $n_{cr,2}$, whereas $n_1(t_0)$ is 59% larger than $n_{cr,1}$, which means that R_2 operates close to capacity conditions. The time-varying demands shown in Fig. 5(d) simulate a morning peak hour with high demand $q_{12}(t)$ for trips from R_1 to R_2 , i.e., from the periphery to the city center. The evolution of accumulations over time $n_{ij}(t)$, $0 \leq t \leq 3600$, corresponding to the MPC controller, are presented in Fig. 5(a), while the evolutions presented in Fig. 5(b) correspond to the GC, see Table I. Note that, at the beginning of the control process, both the MPC controller and the GC decrease the total accumulation in R_1 , $n_1(t)$, and keep the total accumulation in R_2 , $n_2(t)$, unchanged. Afterward, the MPC controller tries to decrease $n_2(t)$ by changing $u_{21}(t)$ from 0.1 to 0.55 to let more vehicles enter R_1 . In contrast, the GC brings the two accumulations equal, i.e., $n_1(590) = n_2(590) = 4125$, and after that instance, both region accumulations increase together, while the *chattering* behavior occurs as a result of switching control between u_{min} and u_{max} according to Table I; note the saw lines of accumulations after $t = 600$ s.

The cumulative trip completion that corresponds to the MPC controller and the GC are shown in Fig. 5(c), while the control sequences $u_{12}(t)$ and $u_{21}(t)$ are shown in Fig. 5(e). The third-order function MFDs $G_1(n_1(t))$ and $G_2(n_2(t))$ coincide, as shown in Fig. 5(f), whereas the circle points are the calculated \tilde{G}_1 and \tilde{G}_2 , see (25) and (26). In Fig. 5(f), it is assumed that there are no errors in both MFDs, where $\alpha_1 = \alpha_2 = 0$, see (23) and (24).

The MPC performances for small ($\alpha_1 = \alpha_2 = 0.2$) and large ($\alpha_1 = \alpha_2 = 1$) errors in the MFDs are shown in Fig. 6(a) and (b), respectively. Comparison between the three levels of error, i.e., without errors in Fig. 5(e) and small and large errors in Fig. 6, shows that the control sequence $u_{21}(t)$ becomes less

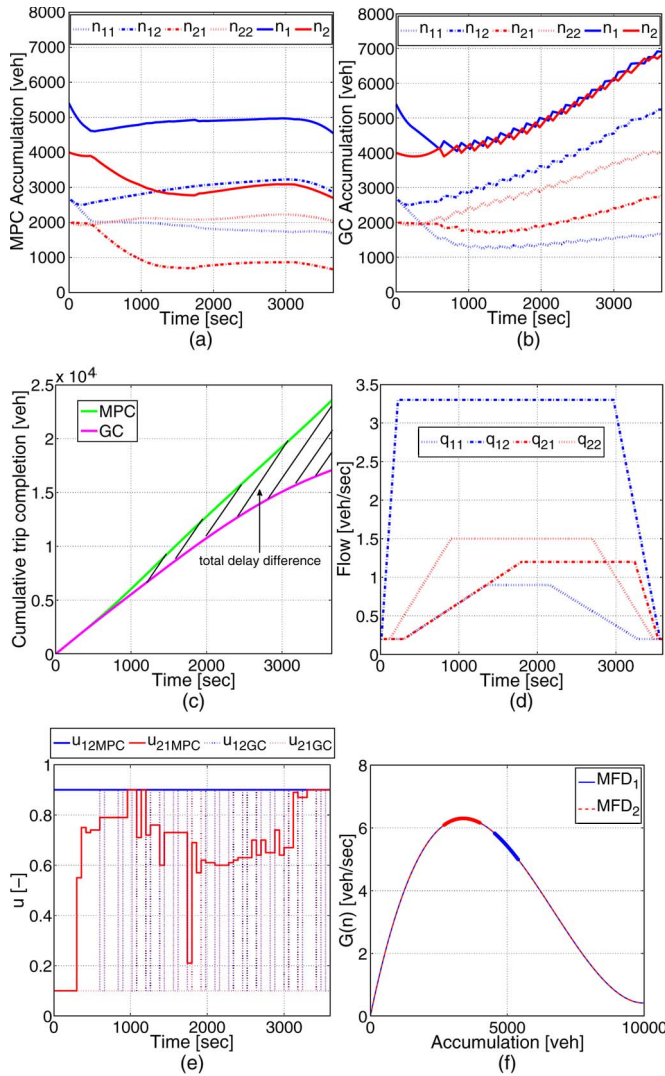


Fig. 5. Example 1. Regions R_1 and R_2 are initially congested (without errors in MFDs).

smooth when the errors in the MFD of the plant (reality) increase. Note that the performance of the MPC is not significantly affected when large errors in the MFD occur. Even if the macroscopic traffic congestion modeling is a rough approximation of urban networks, this methodology can still improve traffic conditions. Nevertheless, nonsmooth controllers might be difficult to apply in reality, create heterogeneities in the boundary, and scatter in the MFDs. We discuss smoothing issues in Section V.

In example 1, the demand is high such that, at the end of the control process, both regions are congested (with GC regions moving forward to face gridlock). The effect of demand on the MPC controller is scrutinized by examples 2 and 3. These examples have the same initial accumulations of example 1; however, the demands $q_{ij}(t)$ for examples 2 and 3 are proportionally decreased by 16% and 32%, respectively, compared with the demand for example 1 in Fig. 5(d). The performance of GC and MPC for different errors in demand and MFDs are summarized in Table II. The results shown in the table are an average over ten runs for small and large errors and for low

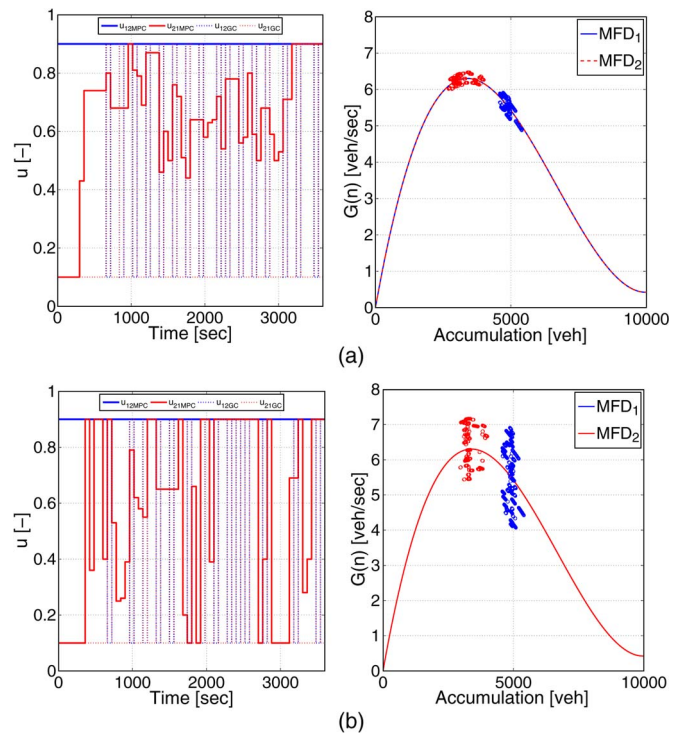


Fig. 6. Example 1. Small ($\alpha_1 = \alpha_2 = 0.2$) and large ($\alpha_1 = \alpha_2 = 1$) errors in MFDs.

and high noise. We show the cumulative trip completion by the end of the simulation and the difference in the total delays (veh · s) as this is expressed by the area between the MPC and GC cumulative trip completion curves in the figures, e.g., see Fig. 5(c). In example 1, there are 22%–24% savings in total, which represent, on the average, 5.5 savings (in minutes) per traveler trip.

Comparing examples 1, 2, and 3 (see Fig. 5(c) and Table II), we notice the advantage of the MPC controller compared to the GC according to the total delay (veh · s) that would be obtained if the GC is used instead of the MPC controller. The differences between the total delays are proportional to the congestion level, i.e., in an uncongested situation (the current accumulations are below the jammed accumulations, and the future demand is not going to change them), the performance of the GC is almost the same as the MPC controller, and the total delay difference is almost zero.

In example 4, in contrast to other examples, both regions R_1 and R_2 are initially uncongested, i.e., the initial accumulations are in the increasing part of the MFDs, whereas at the end of the control process, both regions are congested, as shown in Fig. 7, because of the high level of time-varying demand. The accumulation profiles for both the MPC controller and the GC show the same trend in the uncongested regime, whereas once the system reaches the critical accumulation point, a significant difference is shown between the MPC controller and the GC. It can be inferred that, when the regions become more congested, the difference between the MPC controller and the GC is more apparent.

In example 5, the region R_1 MFD and its corresponding internal demand $q_{11}(t)$, as shown in Fig. 5(d), are increased

TABLE II
TRIP COMPLETION THAT CORRESPONDS TO THE MPC CONTROLLER AND THE GC, AND THE TOTAL DELAY DIFFERENCE

Example 1	without errors ($\alpha_1 = \alpha_2 = 0$)			small errors ($\alpha_1 = \alpha_2 = 0.2$)			large errors ($\alpha_1 = \alpha_2 = 1$)		
	MPC (veh. $\cdot 10^3$)	GC (veh. $\cdot 10^3$)	MPC-GC (veh. $\cdot s \cdot 10^3$)	MPC	GC	MPC-GC	MPC	GC	MPC-GC
without noises	23.55	17.07	7791.6 (22.5%)	23.63	17.11	7883.5 (22.7%)	24.04	17.21	8370.0 (24.1%)
low noises	23.37	16.78	7864.3 (22.9%)	23.48	16.82	7997.3 (23.3%)	23.93	17.00	8379.5 (24.3%)
high noises	22.80	16.07	7892.6 (23.5%)	23.02	16.35	7931.9 (23.4%)	23.34	15.84	9146.5 (27.6%)
Example 2	without errors ($\alpha_1 = \alpha_2 = 0$)			small errors ($\alpha_1 = \alpha_2 = 0.2$)			large errors ($\alpha_1 = \alpha_2 = 1$)		
	MPC	GC	MPC-GC	MPC	GC	MPC-GC	MPC	GC	MPC-GC
without noises	24.11	20.49	6536.8 (17.3%)	24.12	20.51	6548.8 (17.3%)	24.13	20.56	6664.4 (17.6%)
low noises	24.15	20.43	6654.2 (17.7%)	24.15	20.47	6666.2 (17.7%)	24.17	20.56	6741.1 (17.9%)
high noises	24.29	20.28	6885.2 (18.4%)	24.30	20.34	6899.7 (18.4%)	24.33	20.45	6966.9 (18.6%)
Example 3	without errors ($\alpha_1 = \alpha_2 = 0$)			small errors ($\alpha_1 = \alpha_2 = 0.2$)			large errors ($\alpha_1 = \alpha_2 = 1$)		
	MPC	GC	MPC-GC	MPC	GC	MPC-GC	MPC	GC	MPC-GC
without noises	21.70	21.63	1789.3 (4.4%)	21.70	21.64	1736.1 (4.2%)	21.70	21.64	1836.7 (4.5%)
low noises	21.78	21.69	1923.3 (4.7%)	21.78	21.69	1985.8 (4.9%)	21.78	21.70	2051.6 (5.0%)
high noises	22.11	21.94	2466.7 (6.1%)	22.11	21.94	2490.4 (6.1%)	22.11	21.95	2578.9 (6.4%)
Example 4	without errors ($\alpha_1 = \alpha_2 = 0$)			small errors ($\alpha_1 = \alpha_2 = 0.2$)			large errors ($\alpha_1 = \alpha_2 = 1$)		
	MPC	GC	MPC-GC	MPC	GC	MPC-GC	MPC	GC	MPC-GC
without noises	24.40	22.72	1143.7 (2.6%)	24.40	22.78	1135.4 (2.6%)	24.43	22.96	1120.4 (2.5%)
low noises	24.35	22.61	1200.0 (2.7%)	24.36	22.66	1204.5 (2.7%)	24.39	22.85	1192.8 (2.7%)
high noises	24.25	22.26	1439.3 (3.3%)	24.29	22.31	1435.9 (3.3%)	24.28	21.85	2032.9 (4.7%)
Example 5	without errors ($\alpha_1 = \alpha_2 = 0$)			small errors ($\alpha_1 = \alpha_2 = 0.2$)			large errors ($\alpha_1 = \alpha_2 = 1$)		
	MPC	GC	MPC-GC	MPC	GC	MPC-GC	MPC	GC	MPC-GC
without noises	27.37	21.77	9596.6 (23.3%)	27.37	21.82	9576.3 (23.2%)	27.38	21.93	9615.9 (23.2%)
low noises	27.40	21.60	9784.1 (23.8%)	27.37	21.67	9739.1 (23.7%)	27.38	21.85	9705.4 (23.5%)
high noises	27.37	21.09	10249.1 (25.3%)	27.38	21.12	10282.5 (25.4%)	27.41	21.31	10284.0 (25.3%)

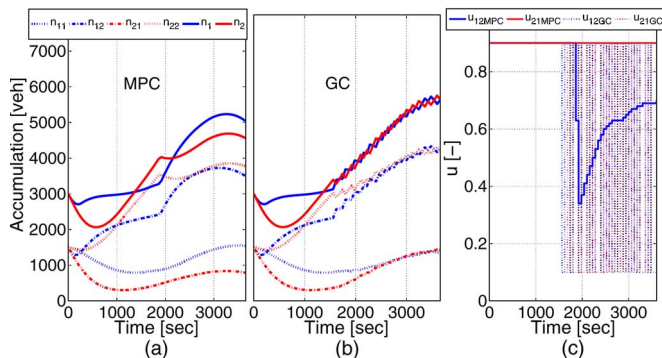


Fig. 7. Example 4. Regions R_1 and R_2 are initially uncongested and finally congested.

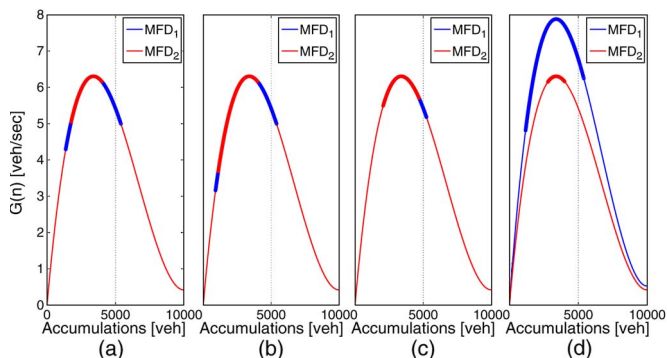


Fig. 8. MFDs for both regions without errors in the MFDs or noises in the demands for (a) example 2, (b) example 3, (c) example 4, and (d) example 5.

by 25%. The MFDs for this example are depicted in Fig. 8(d), whereas the MFDs for examples 2, 3, and 4 are shown in Fig. 8(a)–(c), respectively.

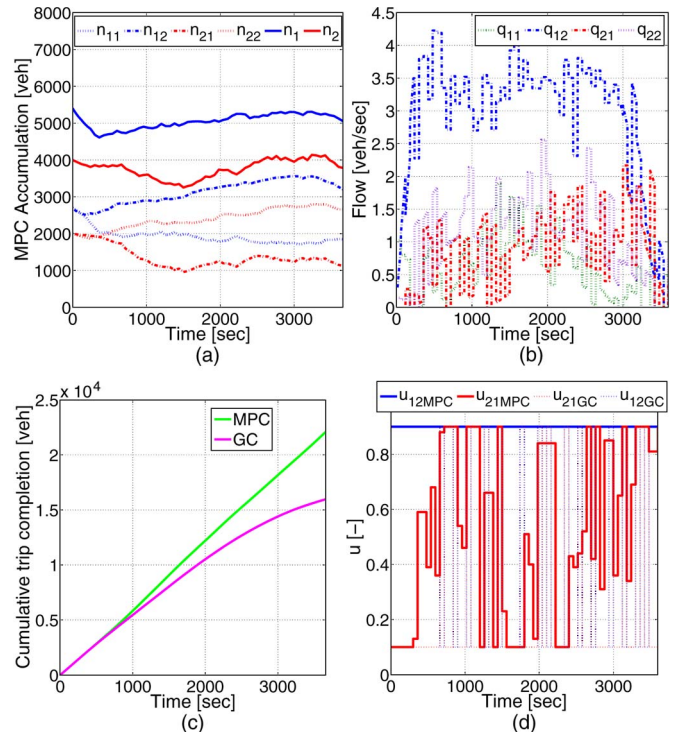


Fig. 9. Example 1. High unbiased noise in the demand.

In addition, the performance of the MPC controller that encounters unbiased and biased noise in the demand is investigated. Example 1 with high unbiased noise in the demand ($\sigma_{ij} = 0.5$, $i, j = 1, 2$), see (27), is illustrated in Fig. 9. The overall results of MPC remain similar; however, the corresponding applied MPC shows more fluctuations than the base

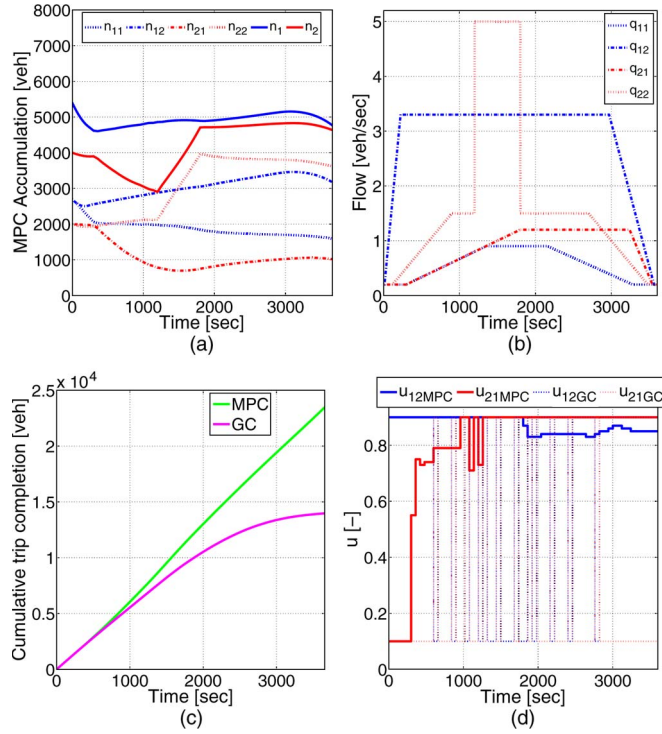


Fig. 10. Example 1. Biased noise in the demand.

example 1. The differences between the trip completion that corresponds to MPC and GC without errors, with small ($\alpha_1 = \alpha_2 = 0.2$) and large ($\alpha_1 = \alpha_2 = 1$) errors in the plant MFDs and without noise, and low ($\sigma_{ij} = 0.25$, $i, j = 1, 2$) and high ($\sigma_{ij} = 0.5$, $i, j = 1, 2$) unbiased noise in the demand profile are also summarized in Table II.

In Fig. 10, biased noise in the demand, which occurs at time instant 1200 (s) for a duration of 600 (s) is added to the base setup of example 1, see Fig. 10(b). The MPC controller does not receive the information of the sharp change in demand. As Fig. 2 shows, the input to the MPC controller is the demand without any noise, which can be estimated with traditional methods. Nevertheless, it results in a similar performance, whereas the GC makes both regions face gridlock (it can be inferred from the almost-horizontal ending part of the GC trip completion profile), see Fig. 10(c). Note that the MPC profile in Fig. 10(d) is identical to Fig. 5(e) for times before 1200 (s), and after that, with decreasing $u_{12}(t)$ from u_{\max} , the MPC can handle the unbiased sudden augmentation in the demand, which has a great impact on $n_2(t)$ accumulation, see Fig. 10(a). The results also show that the trip completion that corresponds to MPC are similar, comparing between unbiased noise [23.49 (veh · 10³)] and without noise in the demand (23.55 (veh · 10³), see Table II) for no errors in the MFDs case. However, large differences are obtained for GC as the trip completion decreased from 17.07 to 13.97 (veh · 10³), and the total delay difference increased from 7791.6 (veh · s · 10³) (22.5%) to 10684.3 (veh · s · 10³) (33.2%).

The results shown in this section imply that the MPC controller is superior for all examples with different levels of error in MFD and noise in the demand, biased or unbiased.

V. SMOOTHING CONTROL

In the previous section, the results of various examples have shown that the control values of two successive steps may significantly vary [e.g., see Fig. 9(d)], particularly in cases of demand variations that are more realistic. These large jumps make the implementation of control policies difficult for real cases, and they might jeopardize safety and increase the heterogeneity of congestion distribution, which can also result in highly scattered MFDs. Thus, we would like to limit the change (jump) in the control inputs and smooth the control sequences over the control process. In this section, the following two different methods are introduced to smooth the control sequences: 1) imposing the control constraints, and 2) modifying the objective function.

A. Constraints for Smoothing Control

One method for smoothing the control sequences that result from MPC is to impose smoothing control constraints to the optimal open-loop problem (14)–(22) over the control horizon N_c . The imposed smoothing control constraints that limit jumps in the control sequence up to u_{jump} (–) are given as follows:

$$|u_{12}(k) - u_{12}(k-1)| \leq u_{\text{jump}} \quad (29)$$

$$|u_{21}(k) - u_{21}(k-1)| \leq u_{\text{jump}} \quad (30)$$

for $k = k_c, k_c + 1, \dots, k_c + N_c - 1$, where u_{jump} is a given parameter, and $u_{12}(k_c - 1)$ and $u_{21}(k_c - 1)$ are the applied control inputs in the previous horizon.

In Section IV, we note that MPC for example 1 with high noise in the demand profile is very jumpy. To smooth the control sequences of MPC, we utilize the MPC formulation with the confining constraints on control inputs (29) and (30). The results of two values $u_{\text{jump}} = 0.1$ and $u_{\text{jump}} = 0.2$ are shown in Fig. 11(a) and (b), respectively. Both of them yield similar results to the unsmoothed example 1, see Fig. 9(a). However, the control sequences are smoother and show an identical trend. This test also reveals that the MPC formulation is robust to the selection of the u_{jump} value. It is not only a theoretical tool but it can also have direct applications in the field.

B. Modified Objective Function

The second method for smoothing the control sequences is done by introducing a tradeoff between the objective function, i.e., the maximum number of vehicles that complete their trips, and the sum of the square absolute difference between each two control sequences, for example, see [35]. Therefore, the objective function (14) in the optimal open-loop problem (14)–(22) is modified as follows:

$$\min_{\mathbf{u}(k_c), \dots, \mathbf{u}(k_c + N_p - 1)} \left\{ -z(t_{k_c + N_p - 1}) + \beta \sum_{k=k_c}^{k_c + N_c - 1} \left(|u_{12}(k) - u_{12}(k-1)|^2 + |u_{21}(k) - u_{21}(k-1)|^2 \right) \right\} \quad (31)$$

where β is the weight on control sequence changes.

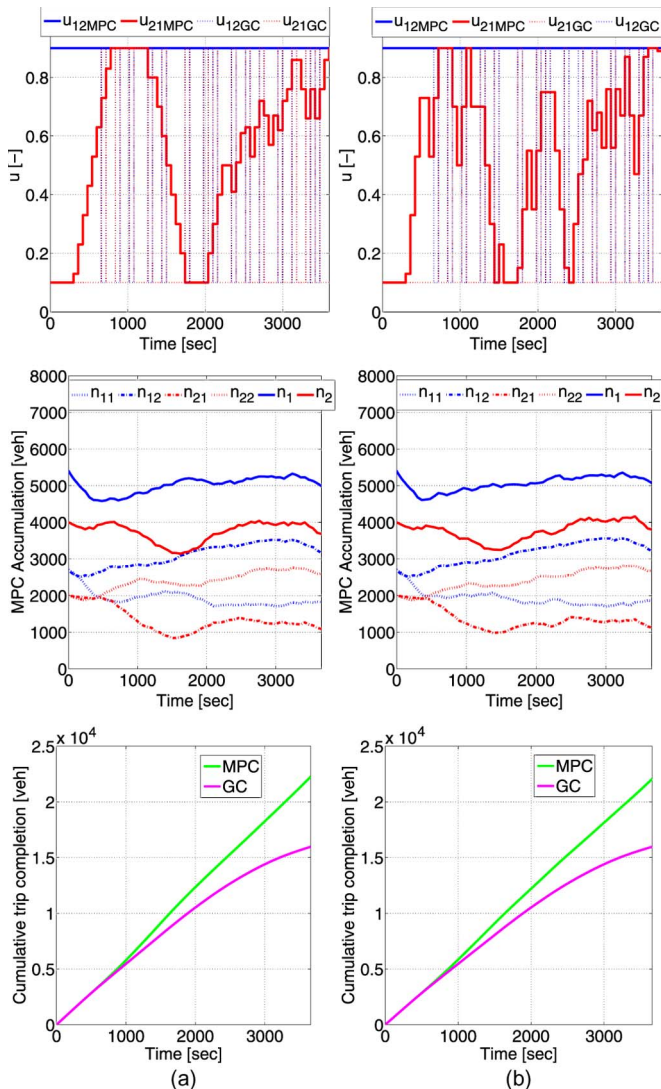


Fig. 11. Smoothing control by imposing constraints with (a) $u_{\text{jump}} = 0.1$ and (b) $u_{\text{jump}} = 0.2$.

The role of β in (31) and u_{jump} in (29) and (30) are similar in terms of making a balance between the desired smoothness and trip completion. Nevertheless, the calibration of u_{jump} is easier than β , because it represents a physical measure.

Similarly, the same example is chosen for smoothing by modifying the objective function. Fig. 12 shows the results, where each column corresponds to $\beta = 1, 10, 50,$ and 200 . It is apparent that the higher the β is, the smoother the control sequences become, and the lower the objective value becomes. For this specific example, $\beta = 10$ produces the best tradeoff between smoothness and optimality. Nevertheless, the value of β does not have a direct physical meaning (as u_{jump} does), and it might require additional calibration in the field. We tried to apply this approach in the other examples, and the appropriate value of β to obtain smooth control and efficient condition was varying. The first approach gave more robust results.

VI. DISCUSSION

The optimal perimeter control for two urban regions with MFD representation has been formulated and solved by im-

plementing the MPC scheme. Comparison results between MPC and GC show that MPC is superior for all numerical examples presented. These results can be of great importance to practitioners and city managers to unveil simple and robust signal timing planning that maximizes the network capacity and serves the maximum number of people under high-demand conditions. The results in this paper can be utilized to develop efficient hierarchical control strategies for heterogeneously congested cities. A network can be partitioned into homogeneous regions, and optimal control methodologies can identify the intertransfers between regions of a city to maximize the system output by utilizing the MPC developed in this paper.

A network with multiple homogenous regions (more than two regions) requires not only more state variables to model the dynamics but also a route choice model to be integrated in the model, because vehicles can travel from one region to another with different routes. When the number of homogeneous regions in the network becomes larger, the computational complexity and the time for solving the MPC problem increase, which might affect the real-time implementation of the proposed methodology. Improving the real-time implementation feasibility for networks with multiple regions should be a research priority. Note that the real-time implementation feasibility is not an issue for the two regions problem presented in this paper, because the computational time was significantly smaller than real time.

These policies can change the spatial distribution of congestion such that the network outflow increases. Given the estimated values in this paper, further analysis is needed to identify signal parameters in the individual regions of a city to smoothly move traffic at the desired flows without concentrating a large number of vehicles at the boundaries of the regions. This is a challenging task that requires knowledge on how the network flow for a region of a city changes as a function of topology, control, and level of congestion. By restricting access to congested cities, a city manager can significantly improve the system output, highlighting the importance of a reliable estimator of subnetwork/route capacity. Toward this direction, the work in [36] investigates the effect in the MFD, different degrees of variability in link lengths, and signal characteristics for different city topologies and signal structures. Nevertheless, the effect of perimeter control in the heterogeneity of density in each region (and the boundaries) cannot be investigated with the macroscopic plant of the specific paper. An application of these strategies in the field or in a microsimulation environment can shed more light toward this direction and identify the necessary local control schemes needed to successfully smooth boundary conditions. This should be a research priority.

Although there are vast contributions in traffic control problems for freeways through ramp metering, the area of control for large urban regions or mixed networks still remains a challenge. Recent findings ([7] and [13]) have shown that MFDs might not be a realistic representation for freeway systems; therefore, in case of mixed arterial-freeway networks, an MFD formulation for the arterial can be combined with a mesoscopic model for the freeway (e.g., a first- or second-order traffic flow

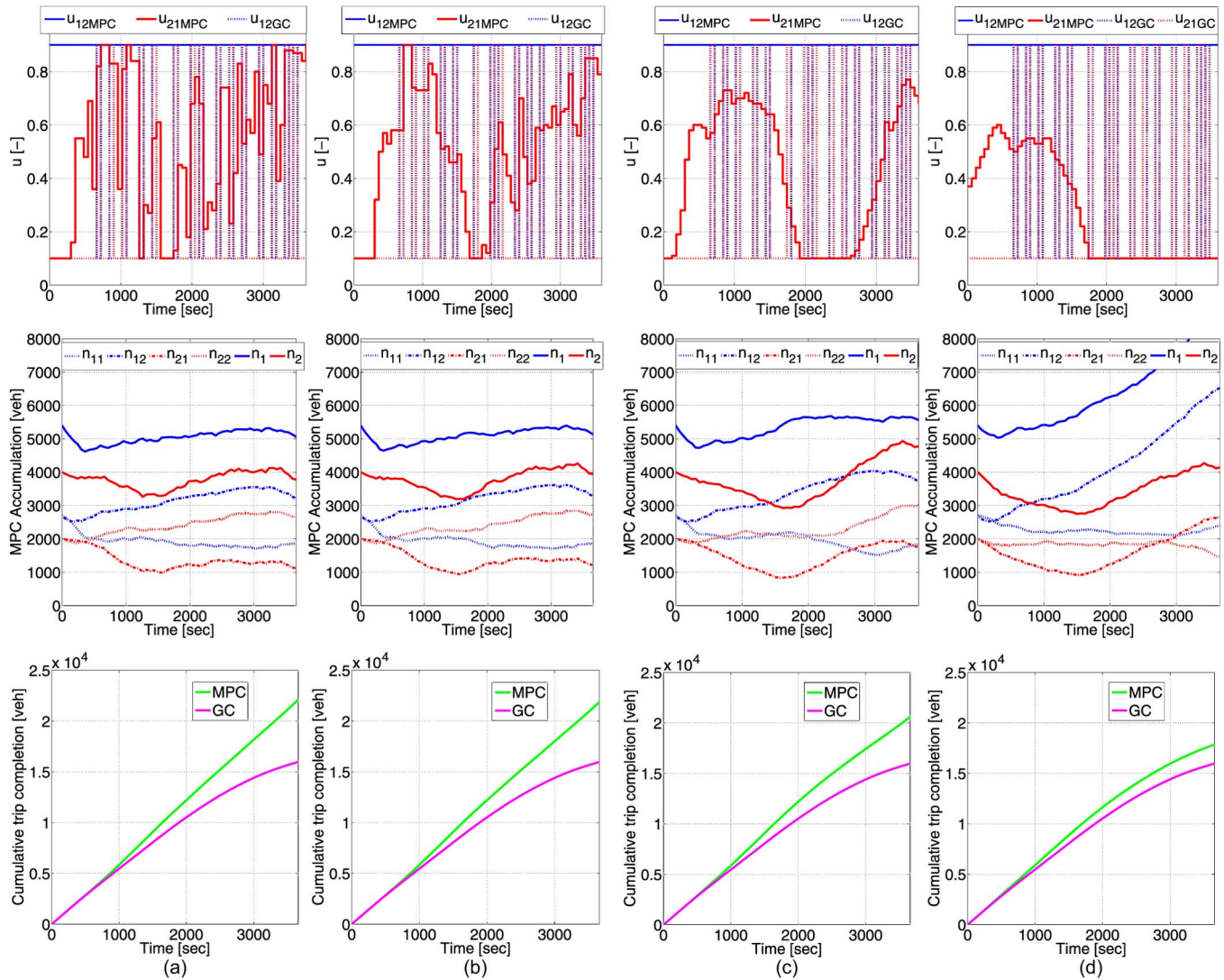


Fig. 12. Smoothing control by the modified objective function (31) with (a) $\beta = 1$, (b) $\beta = 10$, (c) $\beta = 50$, and (d) $\beta = 200$.

model) to describe the dynamics of the system and propose coordinated optimization schemes with ramp metering and perimeter control. Our research provides tools for shedding some light toward this direction.

AUTHOR CONTRIBUTIONS

All three authors equally contributed to all parts of the paper. Their names appear in alphabetical order.

REFERENCES

- [1] J. W. Godfrey, "The mechanism of a road network," *Traffic Eng. Control*, vol. 11, no. 7, pp. 323–327, Nov. 1969.
- [2] R. Herman and I. Prigogine, "A two-fluid approach to town traffic," *Science*, vol. 204, no. 4389, pp. 148–151, Apr. 1979.
- [3] C. F. Daganzo, "Urban gridlock: Macroscopic modeling and mitigation approaches," *Transp. Res. B: Methodol.*, vol. 41, no. 1, pp. 49–62, Jan. 2007.
- [4] N. Geroliminis and C. F. Daganzo, "Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings," *Transp. Res. B: Methodol.*, vol. 42, no. 9, pp. 759–770, Nov. 2008.
- [5] C. Buisson and C. Ladier, "Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams," *Transp. Res. Rec.*, vol. 2124, no. 1, pp. 127–136, 2009.

- [6] Y. Ji, W. Daamen, S. Hoogendoorn, S. Hoogendoorn-Lanser, and X. Qian, "Macroscopic fundamental diagram: Investigating its shape using simulation data," *Transp. Res. Rec.*, vol. 2161, pp. 42–48, 2010.
- [7] M. Saberi and H. Mahmassani, "Exploring properties of network-wide flow-density relations in a freeway network," in *Proc. Transp. Res. Board Annu. Meet.*, Washington, DC, 2012, pp. 1–21.
- [8] A. Mazlounian, N. Geroliminis, and D. Helbing, "The spatial variability of vehicle densities as determinant of urban network capacity," *Phil. Trans. R. Soc. A: Math., Phys. Eng. Sci.*, vol. 368, no. 1928, pp. 4627–4647, Oct. 2010.
- [9] V. L. Knoop, S. Hoogendoorn, and J. W. C. Van Lint, "Routing strategies based on the macroscopic fundamental diagram," in *Proc. Transp. Res. Board Annu. Meet.*, Washington, DC, 2012, pp. 1–18.
- [10] N. Geroliminis and J. Sun, "Properties of a well-defined macroscopic fundamental diagram for urban traffic," *Transp. Res. B: Methodol.*, vol. 45, no. 3, pp. 605–617, Mar. 2011.
- [11] C. F. Daganzo, V. V. Gayah, and E. J. Gonzales, "Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability," *Transp. Res. B: Methodol.*, vol. 45, no. 1, pp. 278–288, Jan. 2011.
- [12] Y. Ji and N. Geroliminis, "On the spatial partitioning of urban transportation networks," *Transp. Res. B*, vol. 46, no. 10, pp. 1639–1656, Dec. 2012.
- [13] N. Geroliminis and J. Sun, "Hysteresis phenomena of a macroscopic fundamental diagram in freeway networks," *Transp. Res. A, Policy Practice*, vol. 45, no. 9, pp. 966–979, Nov. 2011.
- [14] J. Haddad and N. Geroliminis, "On the stability of traffic perimeter control in two-region urban cities," *Transp. Res. B: Methodol.*, vol. 46, no. 9, pp. 1159–1176, Nov. 2012.

- [15] S. J. Qin and T. A. Badgwell, "A survey of industrial model predictive control technology," *Control Eng. Practice*, vol. 11, no. 7, pp. 733–764, Jul. 2003.
- [16] C. E. Garcia, D. M. Prett, and M. Morari, "Model predictive control: Theory and practice—A survey," *Automatica*, vol. 25, no. 3, pp. 335–348, May 1989.
- [17] E. F. Camacho and C. Bordons, *Model Predictive Control in the Process Industry*. Berlin, Germany: Springer-Verlag, 1995.
- [18] E. F. Camacho and C. Bordons, *Model Predictive Control*. Berlin, Germany: Springer-Verlag, 1999.
- [19] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert, "Constrained model predictive control: Stability and optimality," *Automatica*, vol. 36, no. 6, pp. 789–814, Jun. 2000.
- [20] J. M. Maciejowski, *Predictive Control With Constraints*. Harlow, U.K.: Prentice-Hall, 2002.
- [21] T. Bellemans, B. De Schutter, and B. D. Moor, "Model predictive control for ramp metering of motorway traffic: A case study," *Control Eng. Practice*, vol. 14, no. 7, pp. 757–767, Jul. 2006.
- [22] I. Papamichail, A. Kotsialos, I. Margonis, and M. Papageorgiou, "Coordinated ramp metering for freeway networks—A model-predictive hierarchical control approach," *Transp. Res. C: Emerging Technol.*, vol. 18, no. 3, pp. 311–331, Jun. 2010.
- [23] A. Hegyi, B. De Schutter, and J. Hellendoorn, "Optimal coordination of variable speed limits to suppress shock waves," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 1, pp. 102–112, Mar. 2005.
- [24] A. Hegyi, B. De Schutter, and H. Hellendoorn, "Model predictive control for optimal coordination of ramp metering and variable speed limits," *Transp. Res. C: Emerging Technol.*, vol. 13, no. 3, pp. 185–209, Jun. 2005.
- [25] N. H. Gartner, F. J. Pooran, and C. M. Andrews, "Optimized policies for adaptive control strategy in real-time traffic adaptive control systems, implementation and field testing," *Transp. Res. Rec.*, vol. 1811, pp. 148–156, 2002.
- [26] K. Aboudolas, M. Papageorgiou, A. Kouvelas, and E. Kosmatopoulos, "A rolling-horizon quadratic-programming approach to the signal control problem in large-scale congested urban road networks," *Transp. Res. C: Emerging Technol.*, vol. 18, no. 5, pp. 680–694, Oct. 2010.
- [27] S. Lin, B. De Schutter, Y. Xi, and H. Hellendoorn, "Fast model predictive control for urban road networks via MILP," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 3, pp. 846–856, Sep. 2011.
- [28] M. van den Berg, A. Hegyi, B. De Schutter, and J. Hellendoorn, "Integrated traffic control for mixed urban and freeway networks: A model predictive control approach," *Eur. J. Transp. Infrastruct. Res.*, vol. 7, no. 3, pp. 223–250, Sep. 2007.
- [29] C. Diakaki, M. Papageorgiou, and K. Aboudolas, "A multivariable regulator approach to traffic-responsive network-wide signal control," *Control Eng. Practice*, vol. 10, no. 2, pp. 183–195, Feb. 2002.
- [30] F.-Y. Wang, "Parallel control and management for intelligent transportation systems: Concepts, architectures, and applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 630–638, Sep. 2010.
- [31] N. Geroliminis, "Dynamics of peak hour and effect of parking for congested cities," presented at the Proc. Transp. Res. Board Annu. Meet., Washington, DC, 2009, 09-1685.
- [32] J. Haddad, M. Ramezani, and N. Geroliminis, "Model predictive perimeter control for urban areas with macroscopic fundamental diagrams," in *Proc. ACC*, Montréal, Canada, Jun. 2012, pp. 5757–5762.
- [33] K. L. Teo, C. J. Goh, and K. H. Wong, *A Unified Computational Approach to Optimal Control Problems*. New York: Longman, 1991.
- [34] J. T. Betts, *Practical Methods for Optimal Control and Estimation Using Nonlinear Programming*, 2nd ed. Philadelphia, PA: SIAM, 2010.
- [35] K. Aboudolas, M. Papageorgiou, and E. Kosmatopoulos, "Store-and-forward-based methods for the signal control problem in large-scale congested urban road networks," *Transp. Res. C: Emerging Technol.*, vol. 17, no. 2, pp. 163–174, Apr. 2009.
- [36] N. Geroliminis and B. Boyaci, "The effect of variability of urban systems characteristics in the network capacity," *Transp. Res. B*, vol. 46, no. 10, pp. 1607–1623, Dec. 2012.



Nikolas Geroliminis received the Diploma degree in civil engineering from the National Technical University of Athens, Athens, Greece, and the M.Sc. and Ph.D. degree in civil engineering from the University of California, Berkeley.

He is currently an Assistant Professor with the School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, where he is also the Head of the Urban Transport Systems Laboratory (LUTS). Before joining EPFL, he was an Assistant Professor with the Department of Civil Engineering, University of Minnesota, Twin Cities. He is a member of the Traffic Flow Theory Committee of the Transportation Research Board. He also serves on the editorial board of *Transportation Research—Part B, Transportation Letters*, and many international conference proceedings. His research interests include urban transportation systems, traffic flow theory and control, public transportation and logistics, optimization, and large-scale networks.



Jack Haddad received the B.Sc., M.Sc., and Ph.D. degrees in transportation engineering from the Technion—Israel Institute of Technology, Haifa, Israel, in 2003, 2006, and 2010, respectively.

In 2008, he was a visiting Ph.D. student with the Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands. He is currently a Postdoctoral Researcher with the Urban Transport Systems Laboratory (LUTS), School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. His research interests include the management and control of transport networks, traffic light control, optimal control theory, model predictive control, and hybrid systems.

Dr. Haddad is the recipient of the Technion Grant from the Office of the Executive Vice President for Research in 2008 for his Ph.D. research, the Lenget Award for Excellence in Teaching in 2007, and the Student Travel Grant from the Israeli Association for Automatic Control (IAAC) in 2010.



Mohsen Ramezani received the B.Sc. and M.Sc. degrees in electrical engineering (control systems) from the University of Tehran, Tehran, Iran, in 2008 and 2010, respectively. He is currently working toward the Ph.D. degree in the Urban Transport Systems Laboratory (LUTS), School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.

His research interests include intelligent transportation systems, traffic control, and traffic flow studies.

Mr. Ramezani is a member of the IEEE Intelligent Transportation Systems Society and the IEEE Control System Society.