# Joint route guidance and demand management for multi-region traffic networks

C. Menelaou, S. Timotheou, P. Kolios, and C.G. Panayiotou

***Abstract*— Traffic congestion occurs as demand surpasses the available capacity of a road network, resulting to lower speeds and longer journey times. To effectively alleviate the problem, the number of vehicles should be maintained below the network's critical density; with route guidance constituting the primary control strategy to achieve this. However, the effectiveness of route guidance is limited in high-demand conditions.**

**In this work, we investigate a Model Predictive Control (MPC) framework that combines multi-regional route guidance with a novel demand management method. Route guidance is used to minimize the network's density imbalance while demand management is utilized to reduce the conditions that cause congestion. This can be achieved by manipulating vehicle routes (i.e., using route guidance) and/or by instructing a portion of the vehicles to wait at their origin before commencing their journey (demand management). Simulations are conducted to evaluate the performance of the proposed MPC optimization indicating the substantial improvements that can be achieved in traffic flow performance.**

## I. Introduction

Traffic congestion has become one of the most critical problems in modern city landscapes. Interestingly, urban congestion does not occur due to lack of capacity but, primarily, due to the absence of intelligent traffic management policies that can steer traffic away from hotspot regions. To achieve this, route guidance methods seek to reroute traffic flows towards alternative routes that reduce traffic imbalance across the road network. In doing so, the travel times are reduced and the user equilibrium is improved [1] and [2].

The appeal of route guidance methods is strengthened by the recent advancements in information and communication capabilities of onboard units which are now capable of providing real-time traffic state information to drivers and recommend alternatively routes to follow. Despite these great features, congestion still persist since the aforementioned routing solutions only focus on improving the user equilibrium which only slightly benefits the overall system operation [3].

The latter work also argues that routing solutions focused on improving the social optimum can substantially benefit the system since a slight decrease in the number of vehicles entering the network (i.e., demand) can substantially reduce travel times during congestion periods. Evidently, the number of vehicles that need to incur some waiting (at their origin) is significantly smaller than those actually benefiting as elaborated in [3]. Considering these insights, the integration of route guidance with intelligent demand management has become an attractive proposition with the potential to substantially curb the traffic congestion.

In this work, we propose a novel region-level model-predictive control scheme that integrates route guidance with demand management. Given the origin and destination pairs of the vehicular flows that request to navigate within the traffic network, the proposed scheme tries to find the alternative path that minimizes the destination arrival times. The proposed MPC scheme does not only suggest a route to follow, but also manages the external inflow rates, resulting to a congestion-free operation since a portion of the inflows are restricted at their origins (demand management). In this way, route guidance finds the optimum transfer flows across neighboring regions, while demand management regulates the external inflow rates as we have previously proposed in [4] and [5].

Hence, the main contribution of this work is the formulation of the joint route guidance and demand management problem and its solution using an approximate MPC scheme. The consideration of demand management in the formulation is shown to potentially eliminate the formation of cycles[1] that may result from other state-of-the-art route guidance solutions [6]. The reason is that cycles occur when there is a need to delay entrance within a region due to heavy congestion, something that can be avoided by restraining the vehicles from entering the network in the first place.

The remainder of this paper is organized as follows. Section II reviews the related work and indicates our contribution compared to the state-of-the-art. Section III presents the regional level system model and Section IV derives the mathematical formulation of the multi-region route guidance problem. Section V reformulates the problem as a Mixed Integer Linear Program (MILP),

This work has been supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 739551 (KIOS CoE) , the Government of the Republic of Cyprus through the Directorate General for European Programmes, Coordination and Development and through the Research Promotion Foundation (Project: CULTURE/BR-NE/0517/14).

C. Menelaou, S. Timotheou, P. Kolios and C.G. Panayiotou are with the KIOS Research and Innovation Center of Excellence, and the Department of Electrical and Computer Engineering, University of Cyprus, {cmenel02, timotheou.stelios, pkolios, christosp }@ucy.ac.cy

[1] Vehicles are flowing back and forth between neighboring regions.

while Section VI formulates the combination of the route guidance with the demand management method. Section VII includes simulation results demonstrating how the proposed MPC formulation outperforms the corresponding scheme with no demand management. Finally, Section VIII concludes this work and discusses future research directions.

## II. Background and related work

Initial research work on route guidance assumed detail microscopic models, where speed and position information of all vehicles is assumed to be known [1] and [7]. However, microscopic models are highly complex models that make the proposed solutions impractical, especially in large-scale networks [8]. Furthermore, in most cases, different microscopic models apply to different networks and driver behaviours, and hence a solution approach only applies to the particular scenario.

An alternative approach is to use a regional-level route guidance framework where the network is partitioned into smaller homogeneous [9] regions within which vehicles are responsible to follow a regional-level route as shown in [10] and [11]. To do that, the network fundamental diagram (NFD) is used, as it can offer low complexity modeling of large urban networks capturing the macroscopic relationship between the three main mobility parameters, i.e., speed, flow, and density. The NFD is composed of two distinct regimes, separated at the critical density point: 1) the free-flow regime where traffic flows at its maximum speed (free-flow speed) and 2) the congested regime is slowed down as congestion emerges. Therefore, utilizing the NFD dynamics, simple control schemes can be implemented without the need of acquiring extensive traffic information (e.g., the per-link densities) [12].

The work in [10] and [13] propose route guidance frameworks that correlate routing decisions with the NFD, in an effort to better spread the traffic load across a larger area of the network. However, these solutions are not able to cope well with heavy congestion; usually, such approaches aim to control restricted areas (e.g., the city center) so that performance improvements occur only for scenarios with relatively light traffic. This is due to the fact that in high demand, a load balancing method can only delay the emergence of congestion but not actually prevent it. The latter can only be achieved by sustaining the total number of vehicles in all regions below some critical values [8]. Recent attempts trying to only control the total number of vehicles result in travel time imbalances since traffic is not evenly distributed across the regions [14].

In light of this insight, the combination of route guidance with an intelligent demand management method has become an attractive method to tackle the two aforementioned problems since density will be curbed below certain values while minimizing the observed travel times.

Model predictive control (MPC) approaches are increasingly being employed to control traffic congestion, with the NFD serving as the prediction model. MPC has the ability to optimize the current states while considering future implications through the region's NFD model [15]. The work in [16] and [17] initially used a nonlinear MPC framework to control a free-way system and a two-region urban network, respectively. Similarly, more recent works in [11] and [14] attempt to use MPC to implement route guidance schemes based on the NFD dynamics. A hybrid MPC is presented in [18] for an urban region, equipped with time switching plans together with perimeter control where, the non-linear MPC problem is approximated to an MILP, showing the importance of the approximate model regarding the required computation times. Nonetheless, the integration of perimeter control with route guidance demonstrates its ability to postpone the emergence of congestion [6].

As elaborated above, in this work we develop an MPC framework to jointly solve the region-level route guidance and demand management problems in order to find the best alternative routing strategies which minimize the cumulative total time of all vehicles, where the total time of each vehicles accounts for both the waiting time outside the network and the travel time.

## III. System Model

### A. Traffic Flow Model

Let an urban area be partitioned into $R$ homogeneous[2] regions, denoted by $r \in \mathcal{R} = \{1, \ldots, R\}$. We assume that the regions are homogeneous [9] and the traffic dynamics within each region can be modeled using a triangular NFD as depicted in Fig 1 (a) [20]. As illustrated in Fig. 1 (b), the flow-density NFD is complemented by the fundamental relationship that the *intended outflow* $q_r(\rho_r(k))$ (veh/h) is equal to the product of density $\rho_r(k)$ (veh/km) and speed $u_r(\rho_r(k))$ (km/h) at each time-steps $k$, i.e., $q_r(\rho_r(k)) = \rho_r(k)u_r(k)$. Using the NFD one can define all important parameters of region $r$ such as the *jam density*, $\rho_r^J$, the *capacity* $q_r^C = \rho_r^C u_r^f$ where the region operates at its maximum outflow observed at the critical density, $\rho_r^C$, and the *backward congestion propagation speed* $w_r = q_r^C/(\rho_r^J - \rho_r^C)$ [21], where $u_r^f$ is the free-flow speed. In this work, we also assume that the distance traveled by a vehicle inside each region is independent of the origin-destination pair and the drivers' route choice, similar to [18].

Let sets $\mathcal{O} \subseteq \mathcal{R}$ and $\mathcal{D} \subseteq \mathcal{R}$ denote the regions considered as the origins and destinations of flows, respectively. Let also $\mathcal{J}_r^- \subseteq \mathcal{R}$ be the set of neighboring regions directly accessible from region $r \in \mathcal{R}$ (i.e., the immediately next region of $r \in \mathcal{R}$) and similarly let

[2]Each region is partitioned with respect to the homogeneous distribution of accumulated traffic as proposed in [19].

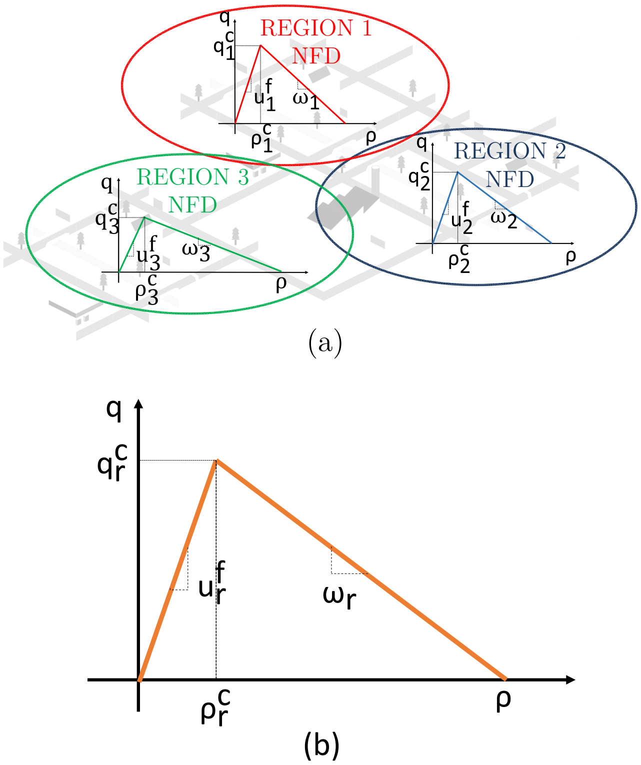



Fig. 1: (a) A three region traffic network where the outflow traffic dynamics are modeled with a dedicated NFD for each region. (b) The triangular flow-density NFD of region $r$ of an urban area.

$\mathcal{J}_r^+ = \mathcal{J}_r^- \cup \{r\}$, such that

$$\mathcal{J}_r = \begin{cases} \mathcal{J}_r^+, & \text{if } r \in \mathcal{D} \\ \mathcal{J}_r^-, & \text{otherwise.} \end{cases} \tag{1}$$

Let also $d_{od}(k)$ (veh) be the instantaneous external demand from $o \in \mathcal{O}$ to $d \in \mathcal{D}$, i.e. the number of *new* vehicles that request to enter region $o \in \mathcal{O}$ toward the destination region $d \in \mathcal{D}$ during time-step $k$. Let also $\tilde{d}_{od}(k)$ denote the admitted external demand that actually enters region $o \in \mathcal{O}$ towards $d \in \mathcal{D}$ during time-step $k$. Variable $\tilde{d}_{od}(k)$ is restricted by three factors:

1) The physical ability of the region to accommodate more vehicles.
2) The maximum possible demand that can physically enter region $o$ at time-step $k$ denoted by $D_{od}^{MAX}$.
3) Demand management which allows only a portion of the requested vehicles to enter the network; the remaining vehicles wait at their origins (outside the network) until they are admitted.

To keep track of the remaining flows to be served, $D_{od}(k)$ represents the cumulative external demand at time-step $k$ defined as

$$D_{od}(k+1) = D_{od}(k) - \tilde{d}_{od}(k) + d_{od}(k), k = 1, 2, \dots, \tag{2}$$

where $D_{od}(0) = 0$.

Let the variable $n_r(k)$ (veh) be the total number of vehicles of region $r \in \mathcal{R}$ at time-step $k$ and $n_{rd}(k)$ denote the number of vehicles in region $r \in \mathcal{R}$ destined to $d \in \mathcal{D}$ such that

$$n_r(k) = \sum_{d \in \mathcal{D}} n_{rd}(k). \tag{3}$$

Let also, the variable $n_{rjd}(k)$ denotes the number of vehicles moving from region $r \in \mathcal{R}$ to $d \in \mathcal{D}$ through the immediately neighbouring region $j \in \mathcal{J}_r$. Then, it is true that

$$n_{rd}(k) = \sum_{j \in \mathcal{J}_r} n_{rjd}(k). \tag{4}$$

Note that $n_{rjd}(k)$, $\{r, j\} \in \mathcal{D}$ determine the number of vehicles that exit the network from their destination, termed *exiting vehicles*.

In a similar fashion we can define the density variables

$$\rho_r(k) = \frac{n_r(k)}{L_r}, \tag{5a}$$

$$\rho_{rd}(k) = \frac{n_{rd}(k)}{L_r}, \tag{5b}$$

$$\rho_{rjd}(k) = \frac{n_{rjd}(k)}{L_r}, \tag{5c}$$

where parameter $L_r$ (km) denotes the total length of all roads of region $r$.

The *intended outflow* of each region $r \in \mathcal{R}$ is denoted by the function $q_r(\rho_r(k))$ (veh/h) which can be approximated using the asymmetric unimodal curve of the triangular NFD [21], shown in Fig. 1 (b), mathematically defined as

$$q_r(\rho_r(k)) = \begin{cases} \frac{q_r^C}{\rho_r^C} \rho_r(k), & \text{if } 0 \leq \rho_r(k) \leq \rho_r^C \\ w_r(\rho_r^J - \rho_r(k)), & \text{otherwise.} \end{cases} \tag{6}$$

We call $q_r(\rho_r(k))$ intended outflow, because it represents the amount of flow that region $r$ would transfer to its neighboring regions and/or the outside world, if no flow/storage capacity restrictions where applicable from other regions.

Accordingly, variables $q_{rd}(k)$ and $q_{rjd}(k)$ denote the *intended transfer flow* from region $r \in \mathcal{R}$ to destination region $d \in \mathcal{D}$ and the corresponding flow that passes through neighboring region $j \in J_r$, respectively, defined as

$$q_{rd}(k) = \sum_{j \in \mathcal{J}_r} q_{rjd}(k), \tag{7}$$

$$q_{rjd}(k) = \frac{\rho_{rjd}(k)}{\rho_r(k)} q_r(\rho_r(k)). \tag{8}$$

In fact, the intended transfer flow between neighboring regions $r \in \mathcal{R}$ and $j \in \mathcal{J}_r^-$ is restricted by their inter-boundary capacity, $C_{rj}(\rho_j(k))$, which is the maximum flow that can be exchanged between the two neighboring regions, for a specific value of $\rho_j(k)$. According to [6], $C_{rj}(\rho_j(k))$ can be defined as

$$C_{rj}(\rho_j(k)) = \begin{cases} C_{rj}^{\mathrm{MAX}}, \text{ if } \rho_j(k) \leq \alpha\rho_j^J, \\ \dfrac{C_{rj}^{\mathrm{MAX}}}{1-\alpha}(1-\dfrac{\rho_j(k)}{\rho_j^J}), \text{ otherwise}, \end{cases} \quad (9)$$

where $C_{rj}^{\mathrm{MAX}}$ is the maximum inter-boundary capacity and $\alpha\rho_j^J$ is the point where the inter-boundary capacity starts to decrease with $0 < \alpha < 1$. Considering Eq. (9) the value of $q_{rjd}(k)$ depends on the total number of vehicles in region $r \in \mathcal{R}$, while the transfer flow of neighboring region $j$ is analogous to its remaining storage capacity which also depends on the transfer flows from other regions $s \in \{\mathcal{J}_j - r\}$. Hence, the *actual transfer flow* from $r \in \mathcal{R}$ to $j \in \mathcal{J}_r$, denoted by the variable $\tilde{q}_{rjd}(k)$ and is defined as

$$\tilde{q}_{rjd}(k) = \min\left(q_{rjd}(k), C_{rj}(\rho_j(k))\frac{q_{rjd}(k)}{\sum_{y\in D} q_{rjy}(k)}\right). \quad (10)$$

Similar to [6] we omit the inter-boundary capacity constraints (9) and (10) from the prediction model used in the developed MPC optimization approach described in Section IV, as the effect of the critical capacity is significantly larger than that of the inter-boundary capacity. Furthermore, the work presented in [6] has extensively studied the sensitivity to changes of the inter-boundary capacity value, indicating that MPC schemes are insensitive to the inter-boundary capacities.

Taking the above into account, the dynamics of the number of vehicles in region $r \in \mathcal{R}$ with destination $d \in \mathcal{D}$, can be defined as

$$n_{rd}(k+1) = n_{rd}(k) + \tilde{d}_{rd}(k) + T_s \sum_{j\in\mathcal{J}_r} (q_{jrd}(k) - q_{rjd}(k)). \quad (11)$$

Finally, it is true that

$$q_{rjd}(k) = n_{jrd}(k)/T_s \quad (12)$$

where $T_s$ denotes the simulation time-step that governs the evolution of the regional dynamics (11).

## IV. Regional level route guidance control

In this section, we employ the regional model described in Section III to develop a mathematical formulation utilizing the NFD of each region to provide optimal route guidance.

### A. Objective function

Let variables $S^a(k)$ and $S^b(k)$ be the cumulative number of vehicles that request to enter the network and successfully arrive at their destination, respectively

$$S^a(k+1) = S^a(k) + \sum_{o\in\mathcal{O}}\sum_{d\in\mathcal{D}} d_{od}(k), \quad (13)$$

$$S^b(k+1) = S^b(k) + \sum_{d\in\mathcal{D}} n_{rjd}(k) \ \{r,j\} \in \mathcal{D}, \quad (14)$$

where $S^a(0) = 0$ and $S^b(0) = 0$.

Summing over all time-steps yields the cumulative total time of all vehicles $J_{CTT}$ (veh·h)

$$J_{CTT} = T_s \sum_k (S^a(k) - S^b(k)) \quad (15)$$

We formulate the problem using a Model Predictive Control framework with the control time-step equal to the simulation time-step. We consider that a new problem is resolved every $m$ time-steps. We also assume that the control and prediction horizons are equal to $mN_p$. Then, for the $l$-th MPC problem solution $l = 1, 2, \ldots$, we define the time horizon $\mathcal{K}_l = \{m(l-1)+1, \ldots, m((l-1)+N_p)\}$.

Under these considerations, we formulate the $l$-th problem of finding the optimal transfer flows $q_{rjd}(k)$ and admitted external flows $\tilde{d}_{od}(k)$ to minimize the total time as:

$$\min\ J_{CTT}^{MPC}(l) = T_s \sum_{k\in\mathcal{K}_l} (S^a(k) - S^b(k)) \quad (16a)$$

$$\begin{aligned} \text{s.t. } & \text{Traffic Dynamics (1)} - \text{(8) and (11)} - \text{(14)}, \\ & C_{rj}^{MAX} \geq \sum_{d\in\mathcal{D}} q_{rjd}(k), k \in \mathcal{K}_l,\ r \in \mathcal{R},\ j \in \mathcal{J}_r, \quad (16b) \\ & \tilde{d}_{od}(k) = \min\Big(((n_o^J - \sum_{d\in\mathcal{D}} d_{od}(k))/|\mathcal{D}|), d_{od}(k), \\ & \qquad D_{od}^{MAX}\Big),\ k \in \mathcal{K}_l, o \in \mathcal{O}, d \in \mathcal{D}, \quad (16c) \\ & 0 < \rho_r(k) \leq \rho_r^J, k \in \mathcal{K}_l, r \in \mathcal{R}, \quad (16d) \\ & S^a(0) = 0, S^b(0) = 0, \quad (16e) \\ & \text{Variables: } \rho_r(k),\ \tilde{d}_{od}(k),\ D_{rd}(k),\ q_{rdj}(k),\ q_{rd}(k), \\ & q_r(\rho_r(k)),\ S^a(k),\ S^b(k),\ n_r(k),\ n_{rd}(k),\ n_{rjd}(k) \end{aligned}$$

The mathematical optimization problem defined by the set of equations (16) is a non-convex Non-Linear Program (NLP) due to the presence of the min term in Eq. (16c), the non-affine function (6), and the product of variables in (8). In problem (16), constraints (1)-(8) and (11)-(14) define the traffic dynamics. The inter-boundary capacity of Eq. (9) is replaced by the constraint (16b) in an effort to not violate the physical limits of the inter-boundary capacity (i.e., $C_{rj}^{MAX}$). Furthermore, constraint (16c) allows all demand to enter unless it is physically restricted by the flow/storage capacity of the region. Constraint (16d) simply ensures that the density of each region is within physical limits, whereas (16e) is the initial condition for the cumulative variables $S^a(k)$ and $S^b(k)$.

The optimal transfer flows in problem (16) can be realized using local controllers located at the boundary of each region through traffic signal control as discussed in [14] and [18]. In addition, the fact that the control actions, i.e. the transfer flows $q_{rjd}(k)$ and the admitted external flows $\tilde{d}_{rd}$, take place only at the boundaries implies that the homogeneity of each region is not affected and their corresponding NFDs remain unchanged.

## V. MILP Reformulation

In this section we approximate problem (16) with a Mixed Integer Linear Program that can be solved to optimality using standard mathematical programming solvers. To achieve this we need to transform (16c), (6) and (8) into an MILP form.

Equality (16c) is associated with the minimum of three affine functions which can be handled by state-of-the-art MILP solvers (e.g., Gurobi [22]) with built-in functions that accurately model this operator using one binary variable for each affine function and appropriate MILP constraints [23].

To transform the product of variables in (8) into MILP constraints we consider an approximation approach. Combining (8) with the fact that $u_r(\rho_r(k)) = q_r(\rho_r(k))/\rho_r(k)$ yields

$$q_{rjd}(k) = \rho_{rjd}(k) u_r(\rho_r(k)) \tag{17}$$

which is comprised of variable $\rho_{rjd}(k)$ and $u_r(\rho_r(k))$ which is a nonlinear function of $\rho_r(k)$. Hence, to approximate (17) we consider segmentation of the density for the function $u_r(\rho_r(k))$.

Towards this direction, we introduce binary variables $b_r^h(k) = \{0,1\}$, $h \in \mathcal{H} = \{1,\dots,|\mathcal{H}|\}$, $r \in \mathcal{R}$ and $k \in \mathcal{K}_l$ which indicate whether $\rho_r(k) \in [\rho_r^{h-}, \rho_r^{h+})$, where $\rho_r^{h-}$ and $\rho_r^{h+}$ is the lower and upper bound of density segment $h$, as shown in Fig. 2. Note that the spacing is not uniform: $b_r^1(k)$ indicates whether region $r$ is in the free-flow so that $\rho_r^{1-} = 0$ and $\rho_r^{1+} = \rho_r^C$, while the rest indicate the corresponding segment in the congested state. Hence, it is true that

$$\sum_{h \in \mathcal{H}} b_r^h(k) = 1, \; r \in \mathcal{R}, \; k \in \mathcal{K}_l \tag{18}$$

By introducing continuous variables, $\rho_r^h(k) \in [0, \rho_r^J]$ and defining the constraints

$$\sum_{h \in \mathcal{H}} \rho_r^h(k) = \rho_r(k), \; r \in \mathcal{R}, \; k \in \mathcal{K}_l \tag{19}$$

$$b_r^h(k)\rho_r^{h^-} \le \rho_r^h(k) \le b_r^h(k)\rho_r^{h^+}, h \in \mathcal{H}, r \in \mathcal{R}, k \in \mathcal{K}_l \tag{20}$$

we ensure that for each time-step and region only one variable $\rho_r^h(k)$ is non-zero and equal to $\rho_r(k)$.

For example, consider the function $u_r(\rho_r(k))$ as presented in Fig. 2 with the following parameters: $\rho_r^J = 90$, $\rho_r^C = 30$, $q_r^c = 1800$ and $|\mathcal{H}| = 4$. We can separate $u_r(\rho_r(k))$ in the following four density segments: $[0, \rho_r^C]$, $(\rho_r^C, 50]$, $(50, 70]$ and $(70, 90]$. Now suppose that at time-step $k$, $\rho_r(k) = 40$ veh/km with $u_r^f = 60$ km/h and $u_r(40) = 37.5$ km/h. Hence, it is true that $b_r^2(k) = 1$ and $b_r^1(k) = b_r^3(k) = b_r^4(k) = 0$, as well as that $\rho_r^2(k) = \rho_r(k) = 40$ veh/km and $\rho_r^1(k) = \rho_r^3(k) = \rho_r^4(k) = 0$.

Having defined segments for the density, one can obtain $q_{rjd}^{h^-}(k)$ and $q_{rjd}^{h^+}(k)$, $h \in \mathcal{H}$, $k \in \mathcal{K}_l$, $r \in \mathcal{R}$, $j \in \mathcal{J}_r$, which provide lower and upper bounds on the transfer flows, defined as

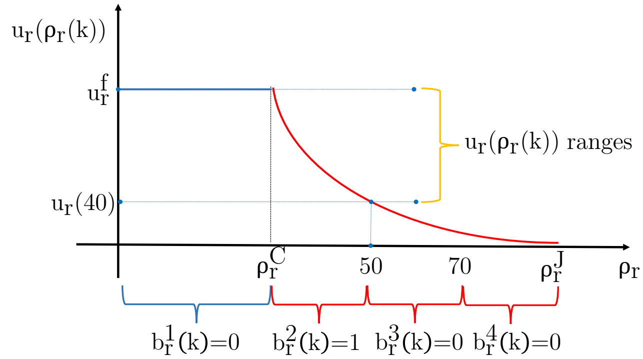



Fig. 2: The function $f(\rho_r(k)) = q_r(\rho_r(k))/\rho_r(k)$ that is produced when considering a triangular NFD.

$$q_{rjd}^{h^-}(k) = \rho_{rjd}(k) u_r(\rho_r^{h^+}) b_r^h(k), \tag{21}$$
$$q_{rjd}^{h^+}(k) = \rho_{rjd}(k) u_r(\rho_r^{h^-}) b_r^h(k), \tag{22}$$

where $u_r(\rho_r^{h^+})$ and $u_r(\rho_r^{h^-})$ are the corresponding lower and upper bounds on the speed for density segment $h$.

Eqs. (21) and (22) contain a continuous/binary variable product, so that each of these two equalities can be equivalently transformed to a set of MILP inequalities using the big “M” notation. Specifically, equalities (21) and (22) are equivalent to (23) and (24) comprised of four MILP constraints, respectively.

$$q_{rjd}^{h^-}(k) \le M b_r^h(k) \tag{23a}$$
$$q_{rjd}^{h^-}(k) \le \rho_{rjd}(k) u_r(\rho_r^{h^+}) \tag{23b}$$
$$q_{rjd}^{h^-}(k) \ge 0 \tag{23c}$$
$$q_{rjd}^{h^-}(k) \ge \rho_{rjd}(k) u_r(\rho_r^{h^+}) - (1 - b_r^h(k))M. \tag{23d}$$

$$q_{rjd}^{h^+}(k) \le M b_r^h(k) \tag{24a}$$
$$q_{rjd}^{h^+}(k) \le \rho_{rjd}(k) u_r(\rho_r^{h^+}) \tag{24b}$$
$$q_{rjd}^{h^+}(k) \ge 0 \tag{24c}$$
$$q_{rjd}^{h^+}(k) \ge \rho_{rjd}(k) u_r(\rho_r^{h^+}) - (1 - b_r^h(k))M. \tag{24d}$$

For both (23) and (24) it is true that $M = C_{rj}^{MAX}$ since the transfer flows are upper bounded by the maximum inter-boundary capacity. Combining (17), (23) and (24) one can obtain the following lower and upper bounds on $q_{rjd}(k)$:

$$\sum_{h \in \mathcal{H}} q_{rjd}^{h^-}(k) \le q_{rjd}(k) \le \sum_{h \in \mathcal{H}} q_{rjd}^{h^+}(k) \tag{25}$$

Hence, the larger the number of density segments $|\mathcal{H}|$ used to approximate $u_r(\rho_r)$ is, the tighter the bounds on the transfer flows will be.

Using the introduced variables $b_r^1(k)$ and $\rho_r^1(k)$ we can also reformulate $q_r(\rho_r(k))$ into the MILP equality

$$q_r(\rho_r(k)) = \Big(\frac{q_r^C}{\rho_r^C} + w_r\Big)\rho_r^1(k) + w_r \rho_r^J (1 - b_r^1(k)) - w_r \rho_r(k). \tag{26}$$

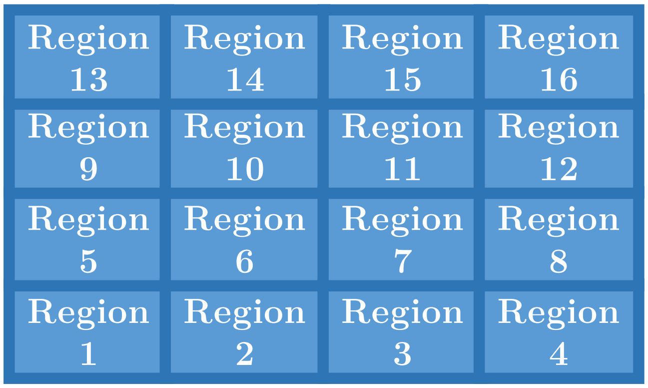



Fig. 3: Simulated urban area consisted from 16 regions.

Since we only provide bounds for the transfer flows, we consider the flow conservation equation within region $r \in \mathcal{R}$

$$\hat{q}_r(\rho_r(k)) = \sum_{j \in \mathcal{J}_r} \sum_{d \in \mathcal{D}} q_{rjd}(k) \tag{27}$$

to achieve better approximation to the final values of the flows through the developed MILP formulation.

To summarize, the problem presented in Eq. (16) can be reformulated into a MILP problem by replacing Eqs. (6) and (8) with Eqs. (18)-(20) and Eqs. (23)-(27). Hence, we get a system with linear inequality constraints with the introduction of additional continuous and binary variables.

## VI. Demand-management Formulation

To formulate the joint route guidance and demand management problem we need to replace Eq. (16c) with the following two equations,

$$\tilde{d}_{od}(k) \leq D_{od}^{MAX} \tag{28}$$

and

$$\tilde{d}_{od}(k) \leq D_{od}(k). \tag{29}$$

By doing this, the optimization will manage to select the amount of admitted external flows that minimize the objective function. Hence, a portion of the vehicles will remain at their origins if that option benefits the overall network performance. Besides, under the MPC framework, the optimization will select the volume of admitted external flows that optimize the near-future states. In other words, MPC will keep the densities of each region up to the values that maximize the actual outflow of each region, aiming at improving the overall network performance.

## VII. Simulation Results

In order to evaluate the performance of the proposed method, a Manhattan-style network topology (see Fig. 3) is considered consisting of 16 regions. The model presented in Eq. (11) is used in the simulations where each region is assume to have identical NFDs [20] with parameters: $\rho_r^C = 30$(veh/km), $\rho_r^J = 130$(veh/km), $L_r = 1$(km), $v_r^f = 60$ (km/h) and $q_r^C = 1800$ (veh/h). Note that, in the simulated network the inter-boundary capacity constraint is included as presented in Eq. (9) and (**??**), with $C_{rj}^{\text{MAX}} = 2000$ (veh/h) and $a = 0.25$. The prediction and control horizons are chosen as $mN_p = 20$ while $m = 5$ for both proposed schemes where the simulation sample time is $T_s = 60$s while the duration of whole simulation experiment is120 min. In this context the performance of the proposed MPC schemes is examined:

- **RG**: The route guidance with no demand management MPC scheme as described in Section V.
- **RGDM**: The joint route guidance and demand management MPC scheme as presented in Section VI.

Note that the comparison of RG and RGDM takes place under light, moderate and heavy demand conditions. In all examined scenarios we consider four regions acting as origins (1, 4, 11 and 16) and four as destinations (2, 8, 9 and 14). For all simulations we assume that the drivers are 100% compliant, the network is initially empty, and demand increases in three phases such that at the end of each phase we let a small period of time for the network to partially discharge. Finally, the formulated problems are constructed and solved using the Gurobi solver [22].

In the topmost part of Table I we depict the cumulative total time (CTT) of all vehicles compared with the total time that would take the vehicles to reach their destination assuming no congestion and no waiting at the origins. Furthermore, the lower part of Table I illustrates the average waiting time (AWT) of vehicles before commencing their trips. Under heavy demand, it is clearly indicated that the RGDM outperforms the RG approach as it results in lower total times. The cumulative total time of the RG grows exponentially as with larger demand congestion emerges due to the absence of congestion. On the contrary, RGDM maintains the density of each region close to the critical value and for this reason the observed travel times are close to the shortest path ones. Furthermore, it is interesting to observe that waiting time with the RG method due to congestion is almost identical to the enforced waiting with the RGDM method, but with significantly higher travel times. Finally, analysis of the derived routes from RGDM illustrated that cycles in the network are avoided as the vehicles prefer to wait at their origins instead to circulating within the network.

Fig. 4 illustrates the space-time diagram of density for the three demand scenarios considered. Comparing the three plots, it is evident that the use of demand management ensures that density is maintained around the critical values, even under moderate and heavy demand. Under the light demand scenario the performance of the two methods is almost identical.

Figs. 5 illustrate the cumulative number of vehicles that request to enter the network (generated) with the number of vehicles that have completed their trip (exiting vehicles) for the two methods. For the light demand

| | | Demand Scenarios | | |
|---|---|---|---|---|
| | | Light | Moderate | Heavy |
| CTT | RG | 3.84 | 4.61 | 8.11 |
| | RGDM | 3.84 | 3.96 | 4.58 |
| | Shortest Path | 3.84 | 3.82 | 3.81 |
| AWT | RG | 0 | 0.01 | 0.63 |
| | RGDM | 0 | 0.13 | 0.76 |

TABLE I: The cumulative total Time (CTT) and the Average Waiting Time (AWT) for different demand scenarios.

scenario both methods work equally well as no congestion occurs, but in both moderate and heavy scenarios, the RGDM outperforms the RG method as vehicles can be served with higher flow. Therefore, the RGDM method can offer significant total time reductions with enhanced network performance.

## VIII. Conclusions

This work integrates the multi-region route guidance framework with a demand management methodology that aims to prevent traffic congestion by allowing vehicles for a late departure. The performance evaluation confirms the usefulness of the proposed integration as it leads to substantial improvements in terms of network operation and overall travel time reductions compared to the ordinary route guidance framework.

Future research directions include the investigation of the robustness of the proposed demand management scheme against demand uncertainty and measurements noise with respect to the estimation of the actual density of each region. Furthermore, future work will analyse how driver compliance affects the performance of the proposed scheme and investigate new schemes that incentivize drivers to use demand management towards optimal system performance.

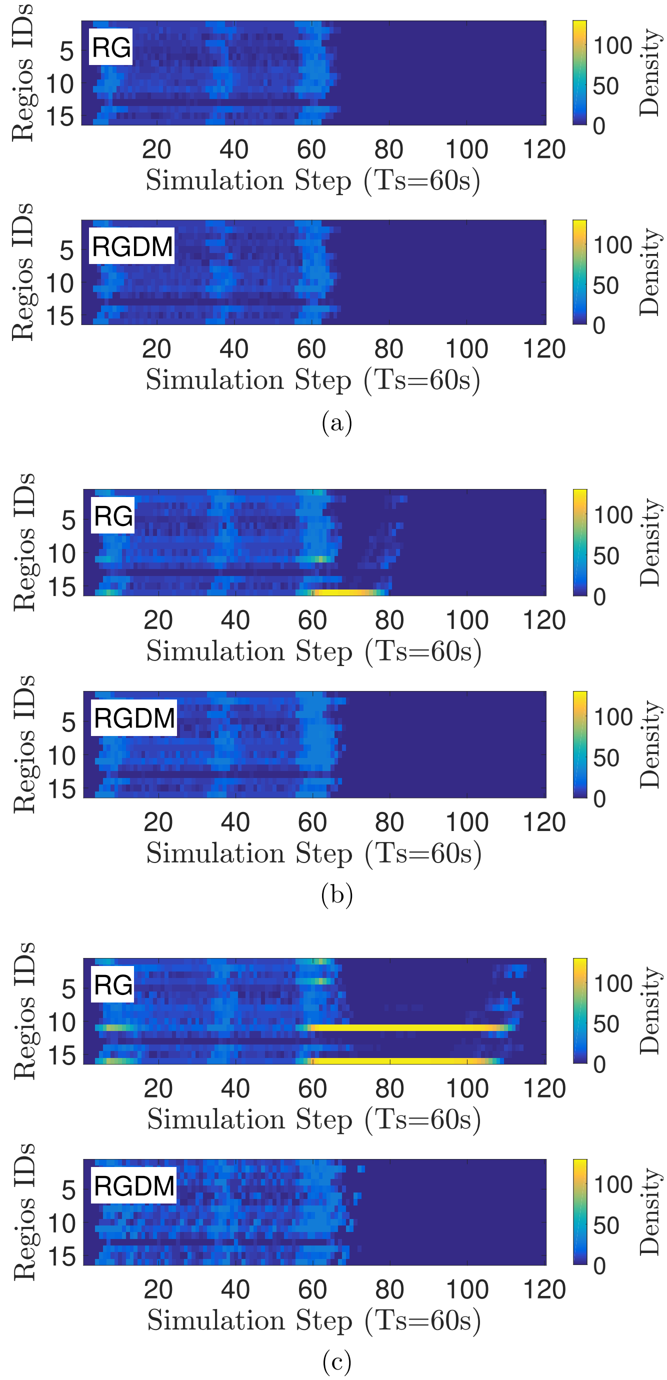



Fig. 4: The instantaneous density of each region observed at each simulation time-step considering (a) light, (b) moderate and (c) heavy demand scenarios.

## References

[1] M. Papageorgiou, "Dynamic modeling, assignment, and route guidance in traffic networks," *Transportation Research Part B: Methodological*, vol. 24, no. 6, pp. 471 – 495, 1990.

[2] H. Mahmassani, S. Peeta, T.-Y. Hu, and A. Ziliaskopoulos, "Dynamic traffic assignment with multiple user classes for real-time ATIS/ATMS applications," in *Proceedings of the Advanced Traffic Management Conference of the Federal Highway Administration*, 1993.

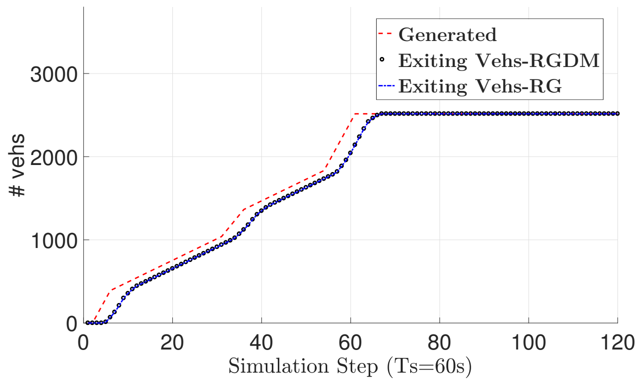
[3] S. Çolak, A. Lima, and M. C. González, "Understanding congested travel in urban areas," *Nature communications*, vol. 7, no. 10793, 03 2016.

[4] C. Menelaou, P. Kolios, S. Timotheou, C. Panayiotou, and M. Polycarpou, "Controlling road congestion via a low-complexity route reservation approach," *Transportation Research Part C: Emerging Technologies*, vol. 81, pp. 118–136, 2017.

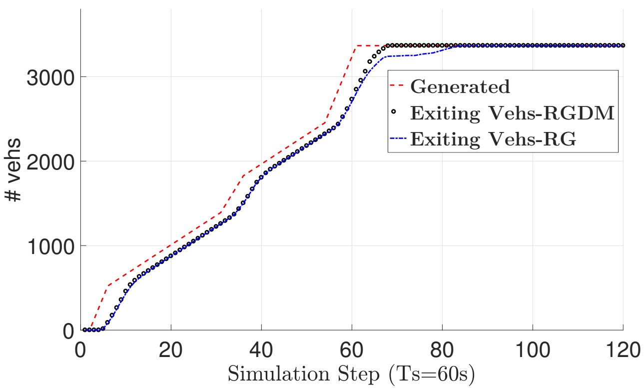
[5] C. Menelaou, S. Timotheou, P. Kolios, and C. G. Panayiotou, "Improved road usage through congestion-free route reservations," *Journal of the Transportation Research Board*, vol. 2621, pp. 71–80, 2017.

[6] I. I. Sirmatel and N. Geroliminis, "Economic model predictive control of large-scale urban road networks via perimeter control and regional route guidance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 4, pp. 1112–1121, April 2018.

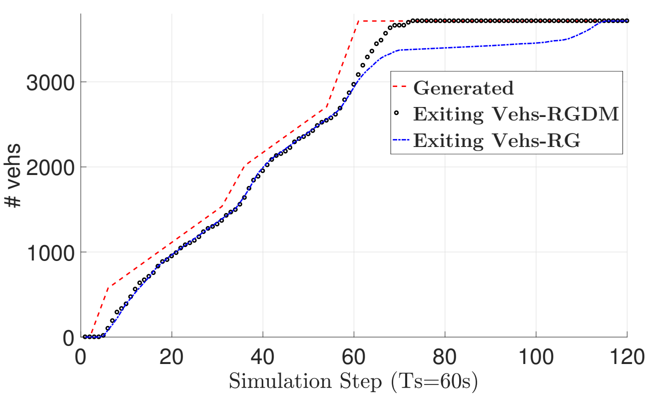
[7] A. K. Ziliaskopoulos, "A linear programming model for the single destination system optimum dynamic traffic assignment problem," *Transportation science*, vol. 34, no. 1, pp. 37–49, 2000.




(a)




(b)




(c)

Fig. 5: The cumulative summation of the vehicles number that: 1) request to enter the network (Generated) and 2) exit the network, considering (a) light, (b) moderate and (c) heavy load demand scenarios.


[8] C. F. Daganzo, “Urban gridlock: macroscopic modeling and mitigation approaches,” *Transportation Research Part B: Methodological*, vol. 41, no. 1, pp. 49–62, 2007.

[9] A. Mazloumian, N. Geroliminis, and D. Helbing, “The spatial variability of vehicle densities as determinant of urban network capacity,” *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 368, no. 1928, pp. 4627–4647, 2010.

[10] V. Knoop, S. Hoogendoorn, and J. Van Lint, “Routing strategies based on macroscopic fundamental diagram,” *Transportation Research Record: Journal of the Transportation Research Board*, no. 2315, pp. 1–10, 2012.

[11] M. Yildirimoglu, M. Ramezani, and N. Geroliminis, “Equilibrium analysis and route guidance in large-scale networks with mfd dynamics,” *Transportation Research Part C: Emerging Technologies*, vol. 59, pp. 404–420, 2015.

[12] Y. Ji and N. Geroliminis, “On the spatial partitioning of urban transportation networks,” *Transportation Research Part B: Methodological*, vol. 46, no. 10, pp. 1639–1656, 2012.

[13] M. Yildirimoglu and N. Geroliminis, “Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams,” *Transportation Research Part B: Methodological*, vol. 70, pp. 186–200, 2014.

[14] M. Hajiahmadi, V. L. Knoop, B. D. Schutter, and H. Hellendoorn, “Optimal dynamic route guidance: A model predictive approach using the macroscopic fundamental diagram,” in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, Oct 2013, pp. 1022–1028.

[15] E. F. Camacho and C. B. Alba, *Model predictive control*. Springer Science & Business Media, 2013.

[16] L. D. Baskar, B. D. Schutter, and J. Hellendoorn, “Hierarchical model-based predictive control for intelligent vehicle highway systems: Regional controllers,” in *13th International IEEE Conference on Intelligent Transportation Systems*, Sept 2010, pp. 249–254.

[17] N. Geroliminis, J. Haddad, and M. Ramezani, “Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: A model predictive approach,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 348–359, 2013.

[18] M. Hajiahmadi, J. Haddad, B. D. Schutter, and N. Geroliminis, “Optimal hybrid perimeter and switching plans control for urban traffic networks,” *IEEE Transactions on Control Systems Technology*, vol. 23, no. 2, pp. 464–478, March 2015.

[19] M. Saeedmanesh and N. Geroliminis, “Clustering of heterogeneous networks with directional flows based on “snake” similarities,” *Transportation Research Part B: Methodological*, vol. 91, pp. 250–269, 2016.

[20] C. F. Daganzo, “The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory,” *Transportation Research Part B: Methodological*, vol. 28, no. 4, pp. 269–287, 1994.

[21] L. Immers and S. Logghe, *Traffic flow theory*, Department of Civil engineering Section Traffic and Infrastructure, Belgium, 05 2003, course H 111.

[22] Gurobi Optimization Inc., “Gurobi Optimizer Reference Manual,” 2016. [Online]. Available: http://www.gurobi.com

[23] F. X. O. Suite, “MILP formulations and linearizations: A quick reference,” 2009.