Integrating heterogeneity dynamics in aggregated urban network modeling and control

Mohsen Ramezani, EPFL/LUTS
Jack Haddad, Technion/T-SMART
Nikolas Geroliminis, EPFL/LUTS

Conference paper STRC 2014
Integrating the dynamics of heterogeneity in aggregated network modeling and control

Mohsen Ramezani
EPFL, LUTS
Lausanne, Switzerland
Phone: +41 21 6932484
email: mohsen.ramezani@epfl.ch

Jack Haddad
Technion, T-SMART
Haifa, Israel
Phone: +972 778871742
email: jh@technion.ac.il

Nikolas Geroliminis
EPFL, LUTS
Lausanne, Switzerland
Phone: +41 21 6932481
email: nikolas.geroliminis@epfl.ch

May 2014

Abstract

Real traffic data and simulation analysis reveal that for some urban networks a well-defined Macroscopic Fundamental Diagram (MFD) exists, which provides a unimodal and low-scatter relationship between the network vehicle density and outflow. Recent studies demonstrate that link density heterogeneity plays a significant role in the shape and scatter level of MFD. Evidently, a more homogeneous network in terms of link density can result in higher network outflow, which implies a network performance improvement. In this study, we introduce two macroscopic models, region and subregion-based MFDs, that model the density heterogeneity in urban networks in order to anticipate the heterogeneity effect. We study the dynamics of heterogeneity and how they can affect the accuracy and scatter of a multi-region MFD model. We also introduce a perimeter flow control by integrating the MFD modeling. The perimeter controller operates on the border between two urban regions, and manipulates the percentages of flows that transfer between the two regions such that the network delay is minimized. The optimal perimeter control problem can be solved by a model predictive control approach, where the prediction model is the aggregated region-based MFD and the plant (reality) is formulated by the subregion-based MFDs, which is a more detailed model.

The perimeter control alleviates the traffic congestion directly by manipulating the perimeter transfer flows of regions while indirectly tries to maximize the outflow of each region by making the regions more homogeneous. The results of this research can provide an insight on the dynamics of heterogeneity in urban network traffic control.

Keywords

Macroscopic fundamental diagram (MFD) – Heterogeneity modeling – Perimeter Control
1. Introduction

Efficient traffic control and management of large-scale transportation networks still remain a challenge both for traffic researchers and practitioners. Unlike microscopic approaches that usually utilize disaggregate traffic flow models, as behavior of each vehicle is modeled in detail, e.g. car following and lane changing model, in this paper, we follow the macroscopic (network level) approach utilizing the macroscopic fundamental diagram (MFD) that aims to simplify the micro-modeling task of the urban network where the collective traffic flow dynamics of sub-networks capture the main characteristics of traffic congestion, such as the evolution of space-mean flows and densities in different regions [1].

The MFD provides a unimodal, low-scatter relationship between network vehicle density (veh/km) and network space-mean flow or outflow (veh/hr) for different network regions, if congestion is roughly homogeneous in the region. Alternatively, the MFD links accumulation, defined as the number of vehicles in the region, and trip completion flow, defined as the output flow of the region. Recently, the macroscopic (network) traffic modeling has intensively attracted the traffic flow community. The physical model of MFD was initially proposed by [2] and observed with dynamic features in congested urban network in Yokohama by [1], and investigated using empirical or simulated data by [3-8] and others.

Studies [5, 6, 9, 10, 11] have shown that networks with heterogeneous distribution of density exhibit network flows smaller than those that approximately meet homogeneity conditions (low spatial variance of link density), especially for high network densities. Networks with small variance of link densities have a well-defined MFD, i.e. low scatter of flows for the same densities. The same studies observed that the average network flow is consistently higher when link density variance is low for the same network density, but higher densities can create points below an MFD when they are heterogeneously distributed. Following these findings the concept of an MFD can be applied for heterogeneously loaded cities with multiple centers of congestion, if these cities can be partitioned in a small number of homogeneous clusters. Recent work [12] created clustering algorithms for heterogeneous transportation networks with an objective to obtain small variance of link densities within a cluster. Recently, in agreement with previous publications in heterogeneity, [10] proposed and calibrated with simulated data an MFD where the effect of heterogeneity decreases the MFD output with a functional relationship.

The MFD can be utilized to introduce elegant real-time control strategies to improve mobility and decrease delays in large urban networks, that local ones are unable to succeed, see pioneer works in [13-15]. Perimeter control strategies, i.e. manipulating the transfer flows at the perimeter border of the urban region, utilizing the concept of the MFD have been introduced
for single-region cities in [13, 16], and for multi-region cities in [14, 15, 17]. Moreover, route guidance strategies with the utilization of MFD have been studied in [18] for grid networks.

In [20] different control strategies with different levels of coordination have been introduced for metropolitan transportation networks that have a hierarchical structure which consists of freeways and urban roads. Previous works [8, 21, 22] have shown that traffic-responsive signal control strategies and different signal settings can change the shape of the MFD and the critical accumulations. While in both works [15, 20], we don’t explicitly model the effect of link heterogeneity, in this paper we aim at introducing a new model that not only integrates the link heterogeneity, but also a route choice model between paths through subregions. We also integrate the effect of subregion receiving capacity. Recently [19] introduced combination of perimeter control for the boundary with switching timing plans for each region. The different timing plans of individual intersections result in different shapes of MFDs.

The control problems in previous works, e.g. [15, 20], have been solved by the model predictive control (MPC) approach. It was shown that this control approach can handle different levels of error in traffic demand and noise in MFDs shape. Nevertheless, the model and the plant in the MPC framework were inherently similar, but the errors in demand and the MFD distinguish between the two. The objectives of this paper are two-fold, in modeling and the control aspects. First, we would like to further investigate the relation between the heterogeneity and the MFD. With respect to modeling, we plan to study the dynamics of heterogeneity and how it can affect accuracy and scatter of a multi-region MFD model, which consists of variables that can be obtained with existing sensor technology. We also plan to further understand under what conditions route choice and control can affect the scatter of an MFD and the trip length in different regions of a city. With respect to control, our objective is to integrate the dynamics of heterogeneity in the optimization framework and design perimeter control strategies that can decrease congestion heterogeneity and increase system performance. While these questions are challenging, this paper sheds some light towards this direction.
2. Modeling density heterogeneity in urban regions with perimeter control

In this paper, we introduce two aggregated models with an objective to integrate the dynamics of heterogeneity in a network: (i) a region-based model considers networks partitioned into several regions that are split by perimeter controllers, and (ii) a subregion-based model, which is a more detailed model, that considers networks that not only partitioned into regions and split by perimeter controllers, but also each region is partitioned into subregions. The region-based model is an extended model of that introduced in [15], as the heterogeneity dynamics is integrated in the regional MFDs, while the subregion-based model is a new model that (a) describes the evolution of subregion accumulations, (b) integrates the heterogeneity dynamics in the subregion MFDs, (c) integrates a route choice model, and (d) models the effect of receiving (or boundary) capacity of subregion destination. Fig. 1 depicts a schematic urban network with (part of) internal and transfer flows for region I and subregions i, j, r in the (i) region- and (ii) subregion-based models, respectively. All the related variables are introduced later in details.

2.1 Region-based model

Let us assume that an urban network is partitioned into \( R \) regions, \( \mathcal{R} = \{1, 2, \ldots, R\} \). Let \( Q_{IJ}(t) \) (veh/s) be the traffic demand flow generated in region \( I \) with direct destination to region \( J \), and \( J \in \mathcal{H}_I \), where \( \mathcal{H}_I \) is the set of regions that are directly reachable from region \( I \). Let \( N_{IJ}(t) \) (veh) be the accumulation in region \( I \) with direct region destination \( J; I, J \in \mathcal{R} \) and \( J \in \mathcal{H}_I \), and \( N_I(t) \) (veh) be the total accumulation in region \( I \).

The total production \( P_I(t) \) (veh.km travelled per unit time) in each region is a function of the regional accumulation and its variance across all links in the region, as has been reported in [5, 9, 10, 11]. The trip completion flow for region \( I \) is the sum of transfer flows, i.e. trips from \( I \) with direct destination \( J, J \in \mathcal{H}_I \), plus the internal flow, i.e. trips from \( I \) with direct destination \( I \). The transfer flow from \( I \) with destination to \( J \) is denoted by \( M_{IJ}(t) \) (veh/s), while \( M_{II}(t) \) denotes the internal flow from \( I \) with destination to \( I \). They are calculated corresponding to the ratio between accumulations as follows

\[
M_{II}(t) = \frac{N_{II}(t)}{N_I(t)} \cdot \frac{P_I(N_I(t), \sigma(N_I(t)))}{L_{II}(t)} \tag{1a}
\]

\[
M_{IJ}(t) = \frac{N_{IJ}(t)}{N_I(t)} \cdot \frac{P_I(N_I(t), \sigma(N_I(t)))}{L_{IJ}(t)} \tag{1b}
\]
where $P_I(.)$ (veh/s.m) is the MFD production (the total distance travelled) for region $I$ at $N_I(t)$, $L_I(t)$ (m) is the average trip length for trips in region $I$, $L_{IJ}(t)$ (m) is the average trip length for trips from region $I$ to $J$, and $\sigma$ which models the link density heterogeneity in space for an urban region.

It is assumed that between each two regions $I$ and $J, J \in \mathcal{H}_I$, exist perimeter controllers $U_{IJ}(t)$ and $U_{JI}(t)$, and $0 \leq U_{IJ}(t), U_{JI}(t) \leq 1$ that constrain the transfer flows from $I$ to $J$ and from $J$ to $I$, respectively. The mass conservation equations of an $R$-region MFDs system are as follows:

$$\dot{N}_{II}(t) = Q_{II}(t) - M_{II}(t) + \sum_{j \in \mathcal{H}_I} U_{JI}(t) \cdot M_{JI}(t) \quad (2)$$

$$\dot{N}_{IJ}(t) = Q_{IJ}(t) - \sum_{j \in \mathcal{H}_I} U_{IJ}(t) \cdot M_{IJ}(t) \quad (3)$$

These equations are a generalized ($R$ regions instead of two) equations presented in [15], and with integrated heterogeneity. Note that route choice modeling is not integrated in the region-based dynamic equations and this model is not aware that travelers make route choice decisions when conditions change. Note also that it is assumed that drivers are not allowed to cross a boundary more than once, e.g. a trip from region $I$ to $I$ by crossing region $J$ is not considered. This will change dynamic equations (2) and (3) and is under ongoing research.
2.2 Subregion-based model

The subregion-based model is a more detailed model as it is assumed that the urban region can be modeled as a collection of several smaller urban areas, called subregions which still contain a significant number of links to be described by a low-scatter MFD. Each subregion has different accumulations evolution capturing the heterogeneity in link density for the urban region. This modeling approach will give us the opportunity to investigate more rigorously several assumptions in the MFD literature that have been empirically observed, e.g. trip length in a region is about constant, if and how route choice, perimeter control and O-D affect the heterogeneity and the distribution of congestion. These are interesting research questions that have been raised by many researchers and it is not clear yet under what network conditions an MFD provides a decent representation of network performance.

Let us consider region \( I \in \mathcal{R} \) which is heterogeneous in space link density and consists of subregions, as schematically shown in Fig. 1. We denote \( \mathcal{SR} \) as the set of all subregions in the urban network, while \( \mathcal{SR}_I \) is the set of subregions that belongs to the region \( I \). Let \( q_{ij}(t) \) (veh/s) be the demand from subregion \( i \) to subregion \( j \), \( n_{ij}(t) \) (veh) be the accumulation in subregion \( i \) with final subregion destination \( j \), and \( n_i(t) \) (veh) be the total accumulation in subregion \( i \). The MFD production for subregion \( i \), denoted by \( p_i(t) \) (veh/s.m), is the total distance traveled for subregion \( i \) at \( n_i(t) \), which is equal to the sum of the transfer and internal flows multiplied by the average trip length in subregion \( i \), \( l_i(t) \) (m).

Let \( m_{ij}^h(t) \) (veh/s) be the transfer flow from subregion \( i \) with final subregion destination \( j \), through the immediate next subregion \( h \in \mathcal{H}_i \), where \( \mathcal{H}_i \) is the set of subregions that are directly reachable from subregion \( i \). The transfer flow is calculated corresponding to the ratio between subregion accumulations, i.e. \( m_{ij}^h(t) = \theta_{ij}^h(t) . n_{ij}(t) / n_i(t) . p_i(n_i(t)) / l_i(t) \), where \( \theta_{ij}^h(t) \) (-) is the flow percentage of the total transfer flows from subregion \( i \) to destination \( j \) that passes immediately through subregion \( h \). Note that a route choice model is integrated in the subregion-based model. The \( \theta_{ij}^h(t) \) are calculated by a logit model according to the travel times through the two current best shortest paths, between each subregion origin and destination (O-D) through all \( h \), which are calculated utilizing Dijkstra’s algorithm. The travel time for each path is calculated by summing travel times through subregions, where each subregion travel time is calculated as the fraction between the distance travelled inside the subregion (through its center) and its average speed \( v_i(t) \) (m/s) calculated from the subregion MFD in the beginning of the trip, i.e. \( v_i(t) = p_i(n_i(t)) / n_i(t) \). Trip length within subregion \( i \) is assumed to be independent of origin, destination and route choice, which is consistent with the field data in [1] and the assumptions made for the region-based models of previous publications [15, 20].
The internal flow from subregion \( i \) with destination to subregion \( i \), denoted by \( m_{ii}(t) \) (veh/s), is calculated by \( m_{ii}(t) = n_{ii}(t) / n_i(t) \cdot p_i(n_i(t)) / l_i(t) \). Note that the region-based model implicitly assumed that the internal regional trips never leave the region and also external trips cross the boundary between the regions only once. The route choice of subregion-based model meets these assumptions. The subregion-based model also integrates the effect of receiving capacity of the destination subregion. In other words, flow transferring into a subregion might be restricted since accumulation at subregion destination is such high that there is not enough space to fully receive the inter transfer flows.

Finally, the transfer flows might be controlled by subregion perimeter controllers on the border between subregions, e.g. \( 0 \leq u_{ih}(t) \leq 1 \) denotes the perimeter control input between subregions \( i \) and \( h \). The mass conservation equations for the subregions are as follows

\[
\dot{n}_{ii}(t) = q_{ii}(t) - m_{ii}(t) + \sum_{h \in \mathcal{H}_i} u_{hi}(t) \cdot \tilde{m}_{hi}(t) \tag{4}
\]

\[
\dot{n}_{ij}(t) = q_{ij}(t) - \sum_{h \in \mathcal{H}_i} u_{ih}(t) \cdot \tilde{m}_{ij}(t) + \sum_{h \in \mathcal{H}_i, h \neq j} u_{hi}(t) \cdot \tilde{m}_{hj}(t) \quad \forall j \in \mathcal{H}_i \tag{5}
\]

\[
\dot{n}_{ir}(t) = q_{ir}(t) - \sum_{h \in \mathcal{H}_i} u_{ih}(t) \cdot \tilde{m}_{ir}(t) + \sum_{h \in \mathcal{H}_i} u_{hi}(t) \cdot \tilde{m}_{hr}(t) \quad i \neq r; \forall r \notin \mathcal{H}_i \tag{6}
\]

Note that our intension is not to control inter transfers between any two subregions, but only in the boundaries of the region-based model. In this way we will keep the computational effort small and we will not rely on information which is difficult to be obtained with real data. Nevertheless, as stated before the more detailed model will shed light on the dynamics of heterogeneity and how it can affect the performance of an MFD region-based model, which consists of variables that can be obtained with existing sensors more accurately.
3. **Optimal Perimeter control for heterogeneous networks**

The aim of optimal perimeter control for heterogeneous networks is to minimize the network delay, defined as the integral of the network accumulation with respect to time, by manipulating the perimeter controllers. We utilize the model predictive control (MPC) framework to solve the optimal control problem. The reader can refer to [15, 20, 23, 24, 25, 26, 27] for different application of MPC in traffic control problems.

Both models, the subregion- and region-based models, are utilized in the MPC framework. The subregion-based model describes the traffic flow dynamics in reality (MPC-plant), while the region-based model is utilized to calculate the optimal control inputs in the optimization loop (MPC-model). Recall that the subregion-based model describes in more details the mass conservation dynamics based on sub-regional MFDs that also integrates the constraints on the transfer flows by the receiving capacity, while the region-based model is the MPC-model that is suitable for performing tractable optimization. Note that both models integrate heterogeneity effect, one at the regional level, while the other at the sub-regional level.

The MPC controller determines the optimal control inputs in a receding horizon manner, meaning that at each time step an objective function is optimized over a prediction horizon of \( K_p \) steps and a sequence of optimal control inputs are derived. Then the first sample of the control inputs is applied to the system and the procedure is carried out again with a shifted horizon. The closed-loop optimal control scheme in the MPC framework takes into account not only the errors between the plant and the model, but also the disturbances, e.g. variations in the expected demands that might affect the system.

The optimal control problem is directly formulated in the MPC framework as follows:

\[
\min_{u_{1j}, u_{fj}} \sum_{l \in \mathcal{K}, f \in \mathcal{K}_f} \sum_{k=0}^{k_p-1} N_{ij}(k_c) + N_{ij}(k_c) \quad (7)
\]

subject to (1)-(3). The problem (7) is a nonlinear optimization problem and it can be solved using nonlinear optimization algorithms.
4. Case Study

In this section, we present a case study example to explore the characteristics of the proposed region-based and subregion-based models along with the MPC control scheme. Note that a main contribution of this paper is developing two different models with different scales of aggregation and utilize them in the MPC framework as the prediction model and the plant, in contrast to [15, 20] in which the dynamics of model and plant in the MPC frameworks were inherently similar, but the demand prediction errors and the MFD noisy scatter distinguish between the model and the plant. The case study consists of two regions, designating the periphery and city center of an urban network, each comprises of 12 and 7 subregions, respectively as schematically shown in Fig. 2(a). Without loss of generality, we assume every subregion has the same MFD (production) consistent with the MFD (production) observed in Yokohama, and consequently, the well-defined relationship between mean and STD of subregion link occupancy exists and the sub-regional average trip length is constant and known. Note that the region average trip lengths are varying as the model evolves.

The exogenous sub-regional time varying demand is simulating one hour of morning peak followed by half an hour of low demand to clear the network (see Fig. 2(b)) while region 1 generates most of the demand towards region 2 which as the central business district attracts trips. We compare the MPC controller with the no control case where there is no restriction on the perimeter transfer flows. The selected MPC controller parameters are as follows, the prediction horizon $K_p = 20$, the control horizon $N_c = 2$, the control lower bound $U_{\text{min}} = 0.1$, and the upper bound $U_{\text{max}} = 0.9$.

In the numerical example, all subregions are initially uncongested, i.e. the initial region accumulations are $N_1(0) = 29000$ (veh) and $N_2(0) = 19000$ (veh). The evolution of subregion accumulations $n_i(t)$ over 1.5 hours of simulation are illustrated in Fig. 2(f) for the MPC controller and the case without control in Fig. 2(d). In addition, Fig. 2(e) and 2(c) show respectively the region accumulations $N_{ij}(t)$ for the MPC and no control cases revealing that with no control, region 2, i.e. city center, faces the gridlock while region 1 is underutilized, whereas MPC controller is effective to manage the morning rush hour by manipulating the perimeter controllers to restrict the inflow from the periphery to the city center.

Fig. 2(h) shows the MPC control sequences. At the beginning of the control process, the MPC controller does not restrict inter flow transfers since both regions are uncongested. Afterwards at $t = 780$ (s) as all the region 2 subregions reaches to their critical accumulation, MPC controller attempts to regulate the region 2 accumulation at the capacity by changing $U_{12}$ from $U_{\text{max}}$ to $U_{\text{min}}$, since without any restriction region 2 will face the gridlock (see the subregion accumulation with no control in 2(d)). Accordingly, accumulations of subregions in region 1
increase comparing with the no control case. The situation remains invariant till \( t = 3660 \) (s), then because of the decrease in the demand, regions accumulation shifts towards the uncongested state. Thus, the MPC controller gradually allows more vehicles to enter to the city center to minimize the total delay in the network. Ultimately, the total delay for MPC control case is \( 3.88 \times 10^8 \) (veh.s) and for no control case is \( 4.24 \times 10^8 \) (veh.s) that indicates 8.5% improvement. We expect that with lower (higher) level of total demand this improvement will decrease (increase).

It is apparent in Fig. 2(d) that the accumulation of each subregion cannot exceed the jam accumulation. In addition, this example shows a realistic traffic phenomenon that the central part of the city center first becomes congested, i.e. subregion 19, and then the congestion propagates across its neighbor subregions. This can be captured by considering the receiving capacity of subregions in the subregion-based model. Moreover because of topological symmetry and demand similarity, there are two groups of subregions in region 1, the ones with odd number and the others with even numbers in Fig. 2(a). The difference is that the subregions that have more border with region 2 become more congested. The same applies for region 2 where all the subregions have similar trend of accumulation whereas subregion 19 has different dynamics because of its different topological property. Finally to investigate the accordance of the region-based model with the subregion-based model, Fig. 2(g) depicts the region accumulations of the region-based model that comparison between them and Fig. 2(e) which shows region accumulations of the subregion-based model, reveals that the region-based model given assumed observable variables, e.g. \( N_I, N_U, L_I, \) and \( L_U \), accurately model the dynamics of subregion-based model.
Figure 2  Case study results. For detailed description see the text.
5. Conclusion and future work

This paper has presented two urban traffic models based on the MFD at different levels of spatial aggregations to model the dynamics of density heterogeneity. A heterogeneous urban region can be partitioned into homogeneous subregions as the detailed model aims at modeling the accumulation dynamics of subregions, while the dynamics of urban regions are modeled in an aggregated manner.

We utilize the subregion- and region-based model as the plant and the model in the MPC framework to formulate the optimal perimeter control for urban regions. The results in this paper can be utilized to develop efficient hierarchical control strategies for heterogeneously congested cities. Towards this direction, a future research would be to integrate variable perimeter control inputs for each subregion in the region boundary to actively control the density heterogeneity. Another research direction is related to model route choice with experienced travel time estimation and identification of equilibrium conditions with an MFD concept.
6. References


