On the calibration of time-varying trip length distributions for the aggregated traffic models

Sérgio F. A. Batista\textsuperscript{*1}, Ludovic Leclercq\textsuperscript{†1} and Monica Menendez\textsuperscript{‡2,3}

\textsuperscript{1}Univ. Lyon, IFSTTAR, ENTPE, LICIT, F-69675, Lyon, France
\textsuperscript{2}Division of Engineering, New York University Abu Dhabi, United Arab Emirates
\textsuperscript{3}Tandon School of Engineering, New York University, USA

February 28, 2019

Words count: 2989 words (excluding references)

Abstract

One of the key questions for Macroscopic Fundamental Diagram traffic models lies in the calibration of the trip lengths, i.e. the travel distances of vehicles in the regions. Few studies in the literature have attempted to propose methodological frameworks to calibrate the trip lengths. Batista et al. (2019) propose a framework to explicitly calculate trip length distributions. However, they do not consider the influence of the traffic conditions on the trip lengths. In this paper, we propose to extend their methodological framework to determine time-varying trip length distributions according to the changes in the traffic conditions.

1 Introduction

Aggregated traffic models at the city network level were early introduced by Godfrey (1969) and later reconsidered by Daganzo (2007) and Geroliminis & Daganzo (2008). These traffic models, although designed for urban areas require the partition of the city network (Figure 1 (a)) into regions (Figure 1 (b)) (see e.g., Saeedmanesh & Geroliminis 2016, 2017, Lopez et al. 2017, Casadei et al. 2018, Ambühl et al. 2019), where the traffic conditions are approximately homogeneous. Figure 1 (c) shows the regional network that corresponds to the city network partitioning. Let $X$ be the set of regions that define the regional network. In each region, the traffic
states are regulated by a Macroscopic Fundamental Diagram (MFD). An MFD is a relationship between the average circulating flow of vehicles and the average density. The traffic dynamics for a single region \( r \in X \) are governed by a conservation equation, where the vehicles’ accumulation \( n_r(t) \) depends on the balance between the inflow \( Q_{in,r}(t) \) and the outflow \( Q_{out,r}(t) \):

\[
\frac{dn_r}{dt} = Q_{in,r}(t) - Q_{out,r}(t), \quad t > 0
\]  

There are two kinds of MFD models that can be distinguished in the literature, depending on the assumptions made on \( Q_{out,r}(t) \): the accumulation-based model (Daganzo, 2007; Geroliminis & Daganzo, 2008); and the trip-based model (Arnott, 2013; Lamotte & Geroliminis, 2016; Mariotte et al., 2017; Mariotte & Leclercq, 2018).

One of the main components of MFD models is the setting of the vehicles’ trip lengths, i.e. the travel distances in the regions. Most applications of the MFD models have been designed for testing control algorithms and strategies, where the authors consider a constant average travel distance for all vehicles in the same region (see e.g., Daganzo, 2007; Keyvan-Ekbatani et al., 2012; Ekbatani et al., 2013; Haddad, 2017; Zhong et al., 2017; Yang et al., 2018). Aboudolas & Geroliminis (2013) and Kouvelas et al. (2017) tested perimeter control strategies in real city networks, but the authors also consider a constant average travel distance for all vehicles traveling in the same region. Up to now, little attention has been paid in the literature to the challenging task of the trip lengths calibration for MFD models applications.

Figure 2 depicts the scale-up of trips in the city network into paths in the regional network as well as the associated challenges. One can observe that the green and blue trips cross a different sequence of regions, following the definition of the city network partitioning. This ordered sequence of crossed regions by a trip is called regional path. The green and blue trips have different travel distances inside each crossed region. This leads to different travel distances associated to the same regional path. They are then characterized by distributions of trip lengths inside each crossed region. An example is depicted in Figure 2 for the green regional path inside the gray region. In fact, the trip length distribution associated to the green regional path in the gray region contains the information of the plausible travel
distances of the green trips in the city network. The question is how to properly
calibrate the trip lengths distributions based on the information of trips in the city
network. Yildirimoglu & Geroliminis (2014) proposed a methodological framework
to implicitly estimate time-varying average trip lengths. Recently, Batista et al.
(2018) and Batista et al. (2019) went one step further and proposed a methodology
to explicitly calculate distributions of trip lengths for the calibration of dynamic
MFD models. The methodology is based on trip patterns in the city network and
different levels of information from the regional network. Since the true trip patterns
in the city network are unknown and change over time, the authors constructed a
virtual set of trips. For this, they randomly sampled a large number of origin-
destination (od) pairs in the city network and calculated the shortest-paths in
terms of distance for each pair. These virtual trips were then filtered according to
different levels of information from the regional network. The latter ranged from
no information about the previous and next regions to be traveled by the trips, to
their associated regional path. The first level of information calculates trip length
distributions considering the travel distances of all trips that cross one region. It
assigns a common average trip length for all vehicles traveling in the same region,
independent of their regional path. The most detailed level only considers trips
that define the same regional path. This allows to determine a different trip length
distribution for all regional paths that cross the same region. The authors showed
that the first level of information is not able to capture the trip length variability
of all regional paths crossing the same region. Moreover, they also showed that
the trip lengths calibration clearly influences the modeled traffic dynamics in the
regions. The authors concluded that filtering the trips by their associated regional
path to explicitly calculate the trip length distributions should be considered.

Yildirimoglu & Geroliminis (2014) and Leclercq et al. (2015) showed that
the vehicles’ trip lengths are influenced by the traffic states. In this paper, we
revisit the methodological framework proposed by Batista et al. (2018) and Batista
et al. (2019) to explicitly determine trip length distributions for the calibration of
MFD models. We propose to extend this framework for time-varying trip length
distributions. In Sect. 2, we review in more detail the methodological framework
proposed by Batista et al. (2018) and Batista et al. (2019). We also discuss the
extension of this methodological framework to determine time-varying trip length
distributions. In Sect. 3, we test the proposed extension on the 6th district Lyon
network, that is divided into four regions. We discuss some preliminary results for
two regional paths. In Sect. 4, we outline the conclusions of this paper.

2 Methodological framework

One of the key questions for the application of MFD-based models is the calibration
of the trip lengths. We start this section by briefly introducing the methodological
framework (Sect. 2.1) proposed by Batista et al. (2018) and Batista et al. (2019)
to calculate static trip length distributions considering the most detailed level of
information. In Sect. 2.2, we discuss the extension of this methodological framework
to determine time-varying trip length distributions.
Fig. 2 – The scale up of trips into regional paths, that are characterized by trip length distributions.
2.1 Static trip length distributions

Batista et al. (2018) and Batista et al. (2019) proposed a methodological framework to explicitly calculate trip length distributions. The methodology utilizes a representative set of trips in the city network as well as its partitioning. Let $G(N, A)$ be a graph that represents the city network, where $N$ is the set of all nodes and $A$ is the set of all links. To partition the city network, one can use one of the different techniques described in the literature (see e.g., Saeedmanesh & Geroliminis, 2016, 2017, Lopez et al., 2017, Casadei et al., 2018, Ambühl et al., 2019). Since the full set of true trips in the city network is not known and very difficult to estimate, the authors propose to construct a set of virtual trips by randomly sampling several origin-destination nodes in the city network and then calculate the shortest-path in distance for each pair. Let $\Gamma$ be this set of virtual trips. The trip length distributions are determined based on four levels of information from the regional network. In this paper, we focus on the most detailed level that filters the virtual trips by their associated regional path. Generically, a regional path $p$ is defined as:

$$p = (p_1, \ldots, p_m, \ldots, p_R), \forall m = 1, \ldots, R \land m \in X$$

(2)

where $R$ is the number of regions that define $p$, $p_1$ and $p_R$ are the Origin (O) and Destination (D) regions, respectively. In this paper, we distinguish between internal and regional paths (Batista & Leclercq, 2018). Figure 3 depicts these differences. A regional path crosses an ordered sequence of different regions. An internal path is defined by virtual trips that travel only within the same region.

The set of trip lengths $L^p_r$ of regional path $p$ in region $r$ is (Batista et al., 2018, 2019):

$$L^p_r = \{\delta^p_{rk} l_{rk}\}, \forall k \in \Gamma$$

(3)

where $l_{rk}$ is the length of virtual trip $k$ that occurs in region $r$; and $\delta^p_{rk}$ is a binary variable that equals 1 if virtual trip $k$ defines regional path $p$, or 0 otherwise.

The average trip length $\bar{L}^p_r$ of regional path $p$ in region $r$ is:

$$\bar{L}^p_r = \frac{\sum_k \delta^p_{rk} l_{rk}}{\sum_k \delta^p_{rk}}, \forall k \in \Gamma$$

(4)

This level of information has two limitations (Batista et al., 2019). First, the average trip length $\bar{L}^p_r$ (see Eq. 4) is strongly influenced when the set size $L^p_r$ is small, i.e. there is a low number of virtual trips associated to the regional path $p$. 

![Fig. 3 – (a) Internal path. (b) Regional path.](image-url)
Let \( N_p \) be the number of virtual trips associated to regional path \( p \). Second, the trip length distributions for the Origin and Destination regions of regional path \( p \) depend on the spatial distribution of the sampled od pairs inside these regions.

### 2.2 Time-varying trip length distributions

The set of trip lengths \( L_p \) (see Eq. 3) is calculated from virtual trips that represent shortest-paths in terms of distance, independent of the traffic conditions in the regions. However, Yildirimoglu & Geroliminis (2014) showed that the trip lengths depend on the traffic conditions and consequently change over time. In this paper, we propose to extend the methodological framework described in the previous section to calculate time-varying trip length distributions for MFD traffic models.

Generically, the set of trip lengths \( L_p \) depends on the traffic conditions:

\[
L_p = f(v_1(n_1), \ldots, v_m(n_m)), \forall m \in X
\]

where \( v_m(n_m) \) is the speed-MFD that regulates the traffic conditions inside a generic region \( m \in X \). For a generic regional path \( p \), \( L_p \) is a multi-dimensional function where the number of variables depend on the number of regions that define \( X \), with vehicles' accumulation \( n_m \).

The speed-MFD of a region \( r \in X \) ranges between 0 and the free-flow speed \( v^{ff}_r \), i.e. \( v_r(n_r) \in [0, v^{ff}_r(n_r)] \). The first step of the proposed extension consists in building the multi-dimensional numerical grid, by discretizing the speed \( v_r(n_r) \) of each region \( r \in X \):

\[
\omega_r(n_r) = \{v^1_r(n_r), \ldots, v^{N_r}_r\}, \forall r \in X
\]

where \( N_r \) is the total number of speed samples for region \( r \); and \( \omega_r(n_r) \) is the set of discretized speeds for region \( r \).

For each regional OD pair, we sample a set of origin and destination nodes in the city network that are in the Origin and Destination regions, respectively. Note that this set is fixed for each OD pair. Then, for each point of the multi-dimensional numerical grid, we calculate the time-dependent shortest path connecting each od pair in the city network using the classical Dijkstra algorithm. The travel time of link \( a \) is:

\[
t_a = \frac{l_a \delta_{ar}}{v^p_r(n_r)}, \forall a \in A \land \forall h = 1, \ldots, N_r
\]

where \( l_a \) is the length of link \( a \); and \( A \) is the set of links that define the city network.

We filter the time-dependent virtual trips following the regional path they define and calculate the average trip lengths \( \bar{T}_p^r \) (see Eq. 4). Let \( \Omega^{OD} \) represent the set of all regional paths gathered that connect the regional OD pair. This step allows to determine the \( \bar{T}_p^r \) values in the multi-dimensional grid. There are two challenges to implement this methodology. First, we need to sample a significant number of od pairs in the city network for each regional OD pair. Second, the calculation of the time-dependent virtual trips for each point in the multi-dimensional numerical grid requires a large computational burden. To bypass these challenges, we make use of the Latin hypercube technique (Tang, 1993) for sampling the od pairs. This allows to gather a representative subset of all possible connections between the origin and destination nodes. In one hand, using this approach significantly reduces the
computational cost required. On the other hand, it also solves the question of the
sample size of od pairs we should consider, i.e. the number of od pairs that we
should sample to calculate the virtual trips set (Batista et al., 2019).

Suppose that \( \hat{L}_p^r \) is the average trip length that we aim to estimate given
a set \( v^*_r, \forall r \in X \) of average speeds in the regional network. That is \( \hat{L}_p^r = f(v^*_1(n_1), \ldots, v^*_m(n_m)), \forall m \in X \). To estimate \( \hat{L}_p^r \), we first find the location of
\( v^*_r, \forall r \in X \) in the numerical grid \( \omega_r(n_r) \) (see Eq. 6). We then gather the \( 2N_r \) closest points as well as the corresponding calculated average trip lengths \( L_p^r \). We fit a
multi-dimensional linear regression model defined as:

\[
\begin{align*}
T_p^r &= \alpha_0 + \sum_{i \in X} \alpha_i v_i + \sum_{i \neq j} \alpha_{ij} v_i v_j, \forall r \in p \land \forall p \in \Omega^{OD} \tag{8}
\end{align*}
\]

where \( v_i \) are the predictors and represent the speed samples of region \( i \) previously
gathered from \( \omega_r(n_r) \); \( \alpha_0, \alpha_i \) and \( \alpha_{ij} \) are the regression coefficients to be deter-
mined; and \( \Omega^{OD} \) is the set of all regional paths connecting the regional OD pair.

The average trip length \( \hat{L}_p^r \) is then estimated as:

\[
\hat{L}_p^r = \alpha_0 + \sum_{i \in X} \alpha_i v_i^* + \sum_{i \neq j} \alpha_{ij} v_i^* v_j^*, \forall r \in p \land \forall p \in \Omega^{OD} \tag{9}
\]

3 Results and discussion

We now discuss the implementation of the methodological framework introduced in
the previous section. The test network is the 6th district of Lyon (France) depicted
in Fig. 4 (a). It is composed by 757 links and 431 nodes. This city network is
partitioned into four regions, for which we fitted the production-MFD and speed-
MFD functions depicted in Fig. 4 (b) and Fig. 4 (c) respectively. We assume a
bi-parabolic shape for the MFD (one parabola for the increasing and one for the
decreasing part of the MFD, with a first derivative equal to zero for the critical
accumulation that maximizes production). The free-flow speeds for regions 1 to 4
are \( v_{1f} = 5.2 \) (m/s), \( v_{2f} = 6.4 \) (m/s), \( v_{3f} = 6.2 \) (m/s) and \( v_{4f} = 5.8 \) (m/s).
Fig. 5 – Left: Evolution of the number of time-dependent virtual trips $N_p$ associated with regional path $p = \{1\}$ as function of the speeds $v_2$ and $v_3$. Right: same but for the average trip length $L_1$ (in m).
Fig. 6 – Evolution of the number of time-dependent virtual trips $N_p$ associated with regional path $p = \{124\}$ as well as of the average trip lengths (in m) for regions $1 (L_1)$, $2 (L_2)$ and $4 (L_4)$ as function of the speeds $v_2$ and $v_3$. 
We first analyze how the average trip lengths $T_p^r$ are influenced by the traffic states (i.e. their relation with the regional mean speeds $\bar{v}_r(n_r), \forall r \in X$) for an internal path $p = \{1\}$ and a regional path $p = \{124\}$. To construct the multidimensional numerical grid, we discretize the speed-MFDs (see Fig. 4 (c)) of the four regions as $\omega_1(n_1) \in [0.4 : \delta v : 5.2], \omega_2(n_2) \in [0.4 : \delta v : 6.4], \omega_3(n_3) \in [0.2 : \delta v : 6.2]$ and $\omega_4(n_4) \in [0.2 : \delta v : 5.8]$, where $\delta v = 0.4$ (m/s).

Fig. 3 depicts the evolution of the number of time-dependent virtual trips $(N_p)$ as well as of the average trip length $T_p^1 = \{1\}$ as function of the speeds $\bar{v}_2$ and $\bar{v}_3$. We show the results for four values of $\bar{v}_1 = 5.2, 2.8, 1.6, 0.8$ (m/s). We observe that under free-flow conditions of regions 1, 2 and 3, the number of time-dependent virtual trips that define regional path $p = \{1\}$ is $N_p \sim 1800$. A similar trend is observed when the three regions are congested (i.e. when $\bar{v}_1$, $\bar{v}_2$ and $\bar{v}_3$ are low). The average trip lengths $L_1$ are also not influenced by the traffic states because similar time-dependent virtual trips set are obtained for the previous two scenarios. As region 1 becomes more congested, while the adjacent regions 2 and 3 are still in free-flow conditions, the number of time-dependent virtual trips $N_p$ associated with the internal path $p = \{1\}$ decrease. For lower values of $\bar{v}_1$, the travel times of the city network links of region 1 increase. When the speeds in region 2 and 3 are in free-flow conditions, the city network links travel times for these two regions will be inferior than the ones of region 1. This means that the time-dependent virtual trips will detour to the links in regions 2 and 3 and then define other regional paths, such as for example $p = \{1231\}$ and $p = \{1321\}$. The set of time-dependent virtual change and the average trip length $T_p^1 = \{1\}$ is then dependent on the traffic conditions (i.e. on the observed mean speeds in the regions).

Fig. 6 also depicts the evolution of $N_p$ as well as of the average trip lengths for regions 1 ($L_1$), 2 ($L_2$) and 4 ($L_4$) as function of $\bar{v}_2$ and $\bar{v}_3$, for regional path $p = \{124\}$. One can observe similar trends as in the case of regional path $p = \{1\}$. For low $\bar{v}_2$ values, the link travel times of region 2 increase. When region 3 is in free-flow conditions, the time-dependent virtual trips detour from region 2 to 3 and therefore define a different regional path $p = \{134\}$. This reduces the $N_p$ associated to regional path $p = \{124\}$ and influences the average trip lengths. When both regions 2 and 3 are in free-flow conditions, the time-dependent virtual trips are more probable to cross the border of region 3 when their origin nodes are closer to this region. However, when $\bar{v}_3$ is low, the travel times of the links in this region increase and the time-dependent virtual trips detour to region 2. This increases the average trip length in region 1 as observed in Fig. 6. This is also true for the destination region 4.

We also estimate the average trip lengths for regional paths $p = \{1\}$ and $p = \{124\}$ (see Eq. 8 to Eq. 9), considering different set of $\bar{v}_1$, $\bar{v}_2$ and $\bar{v}_3$. We have also determined the average trip lengths for these sets of speeds, based on the calculation of the time-varying virtual trips. The results are listed in Table 1 for regional path $p\{1\}$ and in Table 2 for regional path $p\{124\}$. We observe that the estimated trip lengths show a good agreement with the calculated ones based on the time-varying virtual trips. The exceptions happen for regional path $p\{124\}$ and low values of $\bar{v}_2$ and $\bar{v}_3$, i.e. when the average trip lengths are more sensible to $N_p$. One solution is to consider a smaller $\delta v$ to construct the numerical grid for smaller $\bar{v}_r, \forall r \in X$. 

10
Tab. 1 – Estimated $\hat{L}_1$ and calculated average trip lengths $L_1$ (in m) for regional path $p = \{1\}$ and different values of $\bar{v}_1$, $\bar{v}_2$ and $\bar{v}_3$ (in m/s).

<table>
<thead>
<tr>
<th>$\bar{v}_1$</th>
<th>$\bar{v}_2$</th>
<th>$\bar{v}_3$</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$L_1$</td>
</tr>
<tr>
<td>5.10</td>
<td>0.60</td>
<td>1.15</td>
<td>407</td>
</tr>
<tr>
<td>2.50</td>
<td>5.80</td>
<td>6.13</td>
<td>369</td>
</tr>
<tr>
<td>3.40</td>
<td>2.15</td>
<td>4.36</td>
<td>407</td>
</tr>
<tr>
<td>4.33</td>
<td>3.77</td>
<td>5.20</td>
<td>407</td>
</tr>
<tr>
<td>1.05</td>
<td>5.40</td>
<td>2.40</td>
<td>324</td>
</tr>
<tr>
<td>0.66</td>
<td>0.50</td>
<td>5.95</td>
<td>406</td>
</tr>
<tr>
<td>0.45</td>
<td>0.60</td>
<td>0.88</td>
<td>407</td>
</tr>
</tbody>
</table>

Tab. 2 – Same as in Table 1 but for $p = \{124\}$ and different values of $\bar{v}_2$ and $\bar{v}_3$.

<table>
<thead>
<tr>
<th>$\bar{v}_2$</th>
<th>$\bar{v}_3$</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$L_1$</td>
</tr>
<tr>
<td>0.60</td>
<td>1.15</td>
<td>200</td>
</tr>
<tr>
<td>5.80</td>
<td>6.13</td>
<td>208</td>
</tr>
<tr>
<td>2.15</td>
<td>4.36</td>
<td>214</td>
</tr>
<tr>
<td>3.77</td>
<td>5.20</td>
<td>210</td>
</tr>
<tr>
<td>5.40</td>
<td>2.40</td>
<td>209</td>
</tr>
<tr>
<td>0.50</td>
<td>5.95</td>
<td>134</td>
</tr>
</tbody>
</table>
4 Outline

In this paper, we propose an extension of the methodological framework proposed by [Batista et al. (2019)] to determine time-varying trip lengths for the calibration of aggregated traffic models. We show how the traffic conditions in the regions influence the average trip lengths. We also show that the proposed methodology yields good estimation results for the average trip lengths. We plan to pursue this study for other regional paths and analyze the trip lengths estimation for larger city network partitioned into a larger number of regions.

Acknowledgments

This project is supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement No 646592 - MAGuM project).

References


