Identification and characterizing of the prevailing paths on a urban network for MFD-based applications

Sergio Batista, Manon Seppecher, Ludovic Leclercq

Identification and characterizing of the prevailing paths on a urban network for MFD-based applications

S. F. A. Batista[a,b,*], Manon Seppecher[b,c], Ludovic Leclercq[b,**]

[a]*Division of Engineering, New York University Abu Dhabi, Saadiyat Marina District PO Box 129188 - Abu Dhabi, United Arab Emirates*
[b]*Univ. Gustave Eiffel, Univ. Lyon, ENTPE, LICIT, F-69518, Lyon, France*
[c]*CITEPA, Paris, France*

Abstract

One of the main challenges for multi-regional application of the aggregated traffic models based on the Macroscopic Fundamental Diagram, lies in the identification and characterization of the most prevailing paths chosen by drivers. In this paper, we propose a methodological framework, based on two distinct methods, to determine these prevailing paths. The first method requires the information about travel patterns in the urban network as well as the information about the city network partitioning. The second method is more parsimonious, and consists on the direct calculation of shortest-cost paths on the aggregated network. For this, we propose four impedance functions that utilize topological features of the urban network and its partitioning. We test the performance of this methodological framework for determining the most prevailing paths on a network representing the metropolitan area of Lyon (France). We consider a set of real trajectories (i.e. GPS data) of drivers in this network as a benchmark. We show that the proposed methods are able to identify the most prevailing paths as the ones chosen by drivers, as evidenced by a large similarity value between the sets of paths. Based on a maximum likelihood estimation, we also show that the Weibull distribution is the one that better reproduces the functional form of the network-wide distribution of travel distances. However, the characterization of the functional form of such distributions characteristic to each region defining a path is not trivial, and depends on the complex topological features of the urban network concerning the definition of its partitioning. We also show that the Euclidean distance metrics provides good estimates of the average travel distances. Interestingly, we also show that the most prevailing paths are not necessarily the ones that have the lowest average travel distances.

Keywords: Prevailing paths, Urban network, Regional Choice set, Trips, MFD models.

Highlights

- We distinguish between internal and regional paths.
- We propose two methods to determine paths on regional networks.
- We investigate if these methods are able to identify the prevailing paths chosen by drivers.
- We investigate the functional form of the network-wide trip length distribution.
- We investigate the characterization of the travel distances of paths.

1. Introduction

Traffic congestion remains a problem in large metropolitan areas worldwide. One attractive tool, for studying and designing innovative strategies to alleviate congestion, is the aggregated traffic models based on the Macroscopic Fundamental Diagram. The pioneering works on these kind of traffic models are Godfrey (1969), Herman and Prigogine

*Corresponding author. Tel. : +971 26 28 76 98, sergio.batista@nyu.edu
**Corresponding author. Tel. : +33 (0) 4 72 04 77 16, ludovic.leclercq@univ-eiffel.fr

(1979), Mahmassani et al. (1984) and Vickrey (2020). However, they only attracted more attention after the works of Daganzo (2007) and Geroliminis and Daganzo (2008). In the past decade, the MFD-based traffic models (Mariotte et al., 2017; Mariotte and Leclercq, 2019; Jin, 2020) have been used in a broad spectrum of applications, ranging from the test and design of control strategies (e.g. Aboudolas and Geroliminis, 2013; Geroliminis et al., 2013; Ekbatani et al., 2013; Ramezani et al., 2015; Haddad, 2017; Haddad and Mirkin, 2017; Zhong et al., 2017; Kouvelas et al., 2017; Yang et al., 2018; Haddad and Zheng, 2018; Mohajerpoor et al., 2019; Haitao et al., 2019; Sirmatel and Geroliminis, 2019), to perimeter control implementations with route guidance (e.g. Yildirimoglu et al., 2015, 2018; Sirmatel and Geroliminis, 2018; Ingole et al., 2020b), traffic management (Yildirimoglu and Geroliminis, 2014; Laval et al., 2018; Batista and Leclercq, 2019), pricing schemes (e.g. Zheng et al., 2016; Gu et al., 2018; Yang et al., 2019; Zheng and Geroliminis, 2020), multi-modal transportation networks (e.g. Loder et al., 2017; Loder et al., 2019; Paipuri and Leclercq, 2020a,b), and studying the effects network hysteresis during network loading (e.g. Leclercq and Paipuri, 2020), the effects of parking on urban traffic dynamics (e.g. Cao and Menendez, 2015; Leclercq et al., 2017; Cao et al., 2019), environmental applications (Amirgholy et al., 2017; Ingole et al., 2020a; Saedi et al., 2020), the dynamic modeling and control of taxi services (e.g. Ramezani and Nourinejad, 2018), or ride-sourcing services in multi-modal networks (e.g. Wei et al., 2020; Beojone and Geroliminis, 2020).

The aggregated traffic models require the partitioning of the city network into regions, where the traffic conditions are approximately homogeneous, i.e. all vehicles traveling in the same region experience similar mean speeds that are defined by the Macroscopic Fundamental Diagram (MFD). It defines the relation between the average circulating flow in a region and its accumulation of vehicles at a given time instant. The partition of the city network can be done using any technique described in the literature (Saeedmanesh and Geroliminis, 2016, 2017; Lopez et al., 2017; Casadei et al., 2018; Ambühl et al., 2019). This allows to define the regional network, where the connections between adjacent regions depend on the allowed travel directions in the city network. We note that the nodes of the city network serve as the borders between adjacent regions. Let $G(E, X)$ be the graph that defines the regional network, with E edges and X nodes representing by the regions. In this paper, we assume that the regions obtained from the partitioning, are well-defined, compact and fully connected. The main challenge lies on the differences between trips in the city network, and paths on the regional network. Fig. 1 depicts the scaling of trips into paths in the regional network, following the definition of the partitioning. One can observe that the four trips in the city network cross a different sequence of regions. The two green trips are then associated to a different path on the regional network, than the blue trip as well as the purple one. In this paper, we distinguish between two kind of paths on regional networks:

- *Regional path:* is the ordered sequence of regions crossed by the trips from their Origin (O) to their Destination (D) regions. Two examples are represented by the green and blue paths in Fig. 1.
- *Internal path:* represents internal trajectories of vehicles inside the same region. One example is represented by the purple path in Fig. 1.

This distinction between internal and regional paths is of crucial importance. In the application of MFD-based traffic models, regional paths carry exchange flows between adjacent regions. Internal paths carry internal flows. In fact, internal circulating flows can act as active bottlenecks (Mariotte and Leclercq, 2019), that block the traversing flows, traveling on regional paths, at the borders of a given region. This plays a crucial role on how congestion propagates throughout the network, and therefore on the system's performance.

Generically, we define a path p on the regional network as:

$$p = \{O\} \bigcup_{i=1,\dots,|p|-1} \{\{r \cdot \delta_{rp}, \forall r \in \Lambda_i\} \backslash \{0\}\} \tag{1}$$

where δ_{rp} is a binary variable that equals 1 if region r is part of path p, or 0 otherwise; Λ_i is the set of adjacent regions to the region listed in the i-th position of the path p; and $|.|$ represents the total number of regions defining path p. To better understand how this mathematical definition works, we showcase its application to define the sequence of regions of the blue and purple regional paths depicted in Fig. 1. We can observe that the purple path is an internal path. The second condition in Eq. 1 is an empty set, i.e. $\{\{r \cdot \delta_{rp}, \forall r \in \Lambda_5\} \backslash \{0\}\} = \emptyset$, since $|p| = 1$. The purple path is then defined as $p = \{5\} \bigcup \{\emptyset\} = \{5\}$. We now discuss the case of the blue regional path. In this case, we iteratively add the next adjacent regions to be traveled until the Destination region is reached. First, the set of adjacent

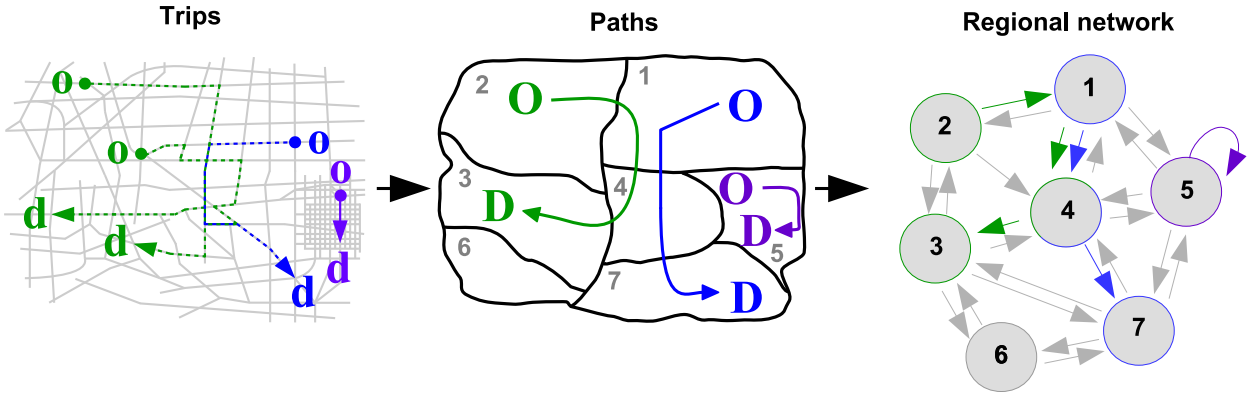


Figure 1: The scaling-up of trips in the city network to paths in the regional network.

regions to the Origin region 1 is $\Lambda_1 = \{2, 4, 5\}$. The next adjacent region to be traveled is region 4, then the set $\{\{r \cdot \delta_{rp}, \forall r \in \Lambda_1\} \backslash \{0\}\} = \{4\}$. Second, the set of adjacent regions to region 4 is $\Lambda_4 = \{1, 2, 3, 5, 7\}$. The next region to be traveled is region 7, and then the set $\{\{r \cdot \delta_{rp}, \forall r \in \Lambda_4\} \backslash \{0\}\} = \{7\}$, which is the Destination region. The blue regional path is then defined as $p = \{1\} \bigcup \{4\} \bigcup \{7\} = \{147\}$.

The main difference between trips and regional paths lies in the characterization of travel distances. Fig. 1 shows two green trips that are linked to the same regional path. These trips travel on a sequence of links with a fixed physical length, from their origin to destination nodes. But, one can observe that these two trips have different travel distances inside each region they cross. Regional paths are then characterized by trip length distributions (TLD) (Batista et al., 2019), instead of a fixed physical length as it is the case of trips in the city network. On the other hand, the total number of trips linked to a regional path defines its prevalence (or significance) level. This means that the most prevailing regional path connecting one OD pair, is the one with the largest number of trips associated. As one example, the green and blue paths have two and one trips associated, respectively. Therefore, the green regional paths is more prevalent. Yildirimoglu and Geroliminis (2014) and Batista et al. (2019) propose different methodological frameworks for estimating the TLD to characterize regional paths. Yildirimoglu and Geroliminis (2014) discusses a methodology for estimating implicit traffic-dependent TLD, and regional paths. Batista et al. (2019) goes one step further. The authors discuss a methodology to determine explicit TLD based on a set of trips and different levels of information regarding the sequence of regions crossed by the trips. In both studies, the regional paths are the result of scaling-up the trips according to the definition of the regions.

Most of the MFD-based applications discussed in the literature require the use of paths on regional networks. For example, in the paper of Daganzo (2007), the case study is one hypothetical region, where vehicles travel an average travel distance of 1000 meters on an implicit internal path (according to our definition). While, in Yildirimoglu and Geroliminis (2014) and Batista et al. (2019), the authors consider a regional network with multiple paths. Recently, Mariotte et al. (2020) discusses the calibration and validation of the MFD traffic models on the city of Lyon, which is partitioned into several regions. The estimated accumulations from the MFD models are compared with real data from loop and probe detectors. The authors show that the correct identification and characterization of paths on regional networks is one of the main critical elements for an accurate prediction of the accumulations in the regions.

In this paper, we propose two dedicated methods for determining the paths on regional networks. The first method is based on the aggregation of a set of trips in the city network into paths on the regional network, following the definition of the partitioning. For this, we propose to construct a set of virtual trips by randomly sample trips in city network. We differentiate between two variants of this method, where the sampling might or not be done concerning the definition of the partitioning. This set of virtual trips is useful when real trip patterns in the city network are not available. The second method is more parsimonious and consists in directly determining the paths on the regional graph. We set different impedance functions based only on topological features of the network, to determine the shortest-cost paths. For this, we consider the exchange flow capacity between borders of adjacent regions, the Euclidean distance or an

average shortest-distance between borders of regions and/or their centroids. We test these methods on the network of Lyon metropolitan area (France). We analyze the performance of both methods against simulated data as well as real data gathered from Global Positioning System (GPS) trajectories. We investigate the following hypothesis/research questions:

1. Which is the best approach to sample virtual set of trips for determining paths?
2. Is the second method able to efficiently estimate similar paths as the ones determined from a simulated and a real dataset of trips?
3. Is the assumption to calculate a set of virtual-trips in distance a good proxy to capture the paths chosen by drivers?
4. How does the partitioning of the city network influences the performance of both methods to estimate the paths chosen by drivers?
5. What is the functional form of the network-wide TLD?
6. Is there a general functional form for the TLD of the regions defining a regional path?
7. Do the Euclidean distance and the average shortest distance between borders and centroids, provide a good approximation of the travel distances in the regions or the total travel distance of a path?
8. How different are the travel distances of paths, determined based on a set of shortest-trips in distance and based on a set of real data?
9. Is there a relationship between significance level of paths and their travel distances?

The remainder of this paper is organized as follows. In Sect. 2, we provide a brief literature review about the methods used to determine trips in the city network. We also introduce the two methods for determining the regional paths. We then discuss the advantages and disadvantages of each method. In Sect. 3, we analyze the application of both methods for determining regional paths on a large city network. We also analyze the performance of both methods against both simulated and real data. In Sect. 4, we discuss the characterization of regional paths through the TLD. In Sect. 5, we outline the main conclusions of this paper. In Sect. 6, we provide a critical assessment about both the proposed methodological framework stressing its main advantages and limitations, and the results of this paper.

2. Methodological framework

In this section, we start by providing a brief literature review about methods used for determining trips on a city network. We also discuss the different challenges related to determining trips in city networks in comparison to paths on the regional network. We then introduce the methodological framework of the two different methods for determining paths on regional networks, and discuss the main advantages and limitations of each one.

In Table A.1, we summarize the notation used in this paper.

2.1. Literature review on choice set generation methods on city networks

The modeling of drivers' trip choices consists in determining a set of routes in the city network that travelers might choose. The goal is to determine the route choice set $\Omega^{od}, \forall(o,d) \in \Xi$ for all origin (o) and destination (d) nodes of the city network. Let Ξ be the set of all od pairs of the city network, and $G(A,Z)$ the city network graph where A and Z represent the set of links and nodes, respectively. In this paper, we use lowercase letters for referring to od pairs in the city network. While, the capital OD refer to Origin-Destination regions in the regional network. We also denote W as the set of all OD pairs of the regional network.

The simplest and most commonly used approach in the literature for determining Ω^{od}, is the Dijkstra algorithm. It can be used to compute the K-shortest trips (Eppstein, 1998; Hadjiconstantinou and Christofides, 1999) that minimize the total cost (e.g. travel distance and/or travel time), without considering travelers' preferences. To name a few examples, the travelers might have specific preferences for choosing highways or main roads, or to avoid traffic lights or traffic jams. The computed trips show, in general, a high degree of similarity, differing from each other only on small detours. As an alternative, van der Zijpp and Catalano (2005) discusses an algorithm that computes the K-constrained shortest trips. The idea is to only compute trips that satisfy a pre-defined set of constraints. Azevedo et al. (1993) discusses a different approach that consists of a route search and then a link elimination. First, the algorithm computes the trip with the minimal travel cost and adds it to Ω^{od}. Then, it eliminates from a few to all links of the

computed trip, from the city network graph. This process is repeated until there are no more trips connecting the od pair. The question is how to properly set the elimination rule. Instead, de la Barra et al. (1993) proposes to increase the link costs that define the computed trips. This process is iteratively repeated until two similar trips are found. The method has a bad performance when the link costs are either low, and the same trip is repeatedly identified, or high and less attractive trips are computed. Ben-Akiva et al. (1984) proposes a method that determines trips based on labels corresponding to travelers' preferences. However, the good performance of this methods relies on the proper setting of the drivers' preferences set (Ramming, 2002; Prato and Bekhor, 2006). Prato and Bekhor (2006) proposes a branch-and-bound algorithm to explicitly solve a constrained route-enumeration problem. While this technique improves the heterogeneity of the choice set, the computational costs strongly depend on the number of computed trips. The simulation approach consists in simulating the generalized link costs from probability distributions (see e.g. Nielsen, 1997; Ramming, 2002; Nielsen et al., 2002; Bierlaire and Frejinger, 2005; Prato and Bekhor, 2006; Bliemer et al., 2007), and then perform a shortest-trip search. The process is repeated until the number of desired routes is reached. Instead, the doubly stochastic approach (Nielsen, 2000) simulates both the generalized link costs and link attributes. The computed routes are filtered according to a set of preference constraints of drivers. More recently, Flötteröd and Bierlaire (2013) proposed a methodology based on the Metropolis-Hastings algorithm for sampling trips. As stressed by the authors, this method is computationally expensive since it may require the calculation of several shortest-trips. Prato (2009) provides a literature review about choice set generation models.

The generation of the choice set $\Omega^{od}, \forall (o,d) \in \Xi$ is extremely challenging. First, there are a large number of possible routes connecting each od pair, in addition to the large number of od pairs in the city network. In general, the choice sets are calculated based on shortest-trip calculations. This is computational costly for large-scale city networks. Second, many of the previous methods yield a choice set composed by trips with a large level of correlation, i.e. trips with a large fraction of overlapping links. The third challenge lies on the appraisal of travelers' preferences. Zhou et al. (2014) shows that travelers do not necessarily choose the shortest-trips. The question is how to determine a choice set of routes that correspond to the travelers' preferences.

2.2. The determination of paths on regional networks

In this paper, we focus on the calculation of paths on the regional network. In the follows, we propose two methods for determining paths on regional networks. The first method relies on exhaustive calculations of shortest-trips in the city network. The second method relies on the calculation of paths directly on the regional network, considering different settings of aggregated impedance functions, that are established based on topological features of the network as well as its partitioning.

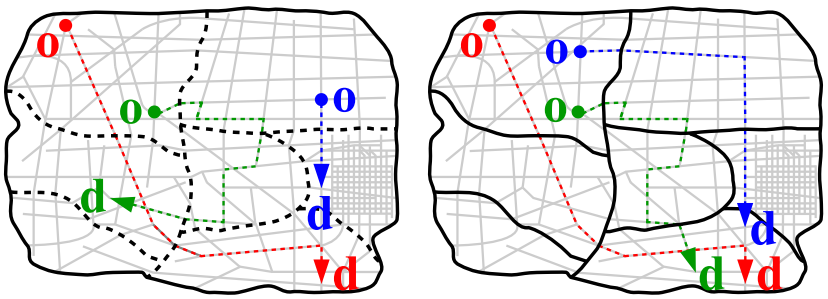
2.2.1. Method 1

This method is somehow similar to traditional choice set generation models for city networks, as discussed in the previous section. These models rely on shortest-trip calculations. It consists in the scaling-up of a set of trips in the city network, following the definition of its partitioning. In the ideal scenario, one can use vehicles trajectories gathered from GPS trajectories. For example, Paipuri et al. (2020) calibrates regional paths and their travel distances based on mobile phone data gathered for the city of Dallas (USA). The authors filter the trips according to the specific sequence of traveled distances for identifying and characterizing the paths on the regional network. However, the main challenge is here to evaluate if this set of trips is representative of the full trip patterns in the city network. On the other hand, the information about real trips patterns is usually unknown. As an alternative, one can construct a set of virtual trips (Batista et al., 2019). This is achieved by randomly sample N_{od} od pairs in the city network (i.e. all nodes are equally probable of being sampled), and then determine the shortest-trip in distance connecting each one. Let Θ be the set of virtual trips.

There are two alternatives for performing the sampling of the od pairs in the city network (see Fig. 2):

- *Variant 1.1 (V1.1):* the sampling of the od pairs is performed independently of the city network partitioning (Fig. 2 (a)).
- *Variant 1.2 (V1.2):* for each regional OD pair, we do the sampling of $N_{od}/|W|$ od pairs in the city network, where $|.|$ represents the total number of regional OD pairs, inside the specific Origin and Destination regions (Fig. 2 (b)).

We then determine the shortest-trips in distance for all sampled od pairs. The regional paths are determined by scaling-up these trips following the definition of the city network partitioning. We emphasize that this scaling set is valid for both a set of real trajectories and a set of virtual trips, calculated as previously discussed. For each regional choice set $\Omega^{OD}, \forall (O, D) \in W$, we consider the K most representative regional paths.



(a) (b)

Figure 2: Sampling of od pairs: (a) independent of the city network partitioning (*Variant 1.1*); and (b) focusing on each regional OD pair (*Variant 1.2*).

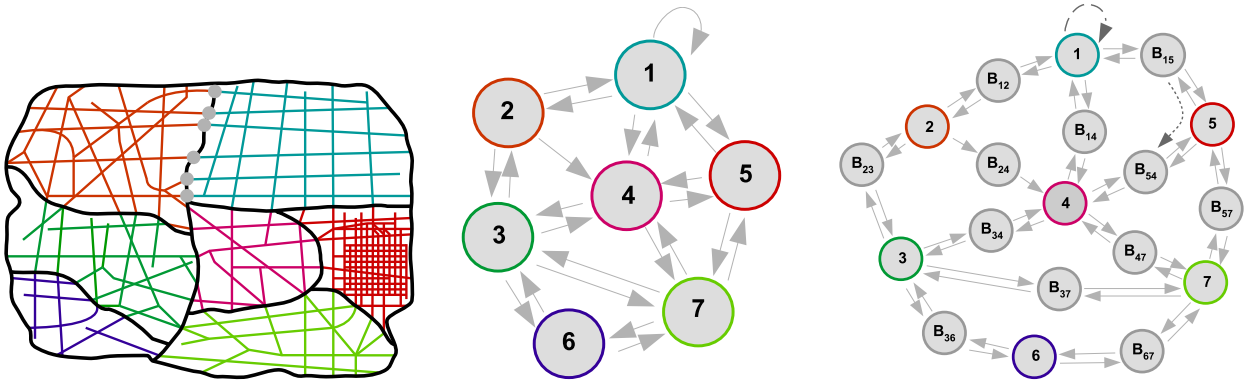
In this paper, we focus on the calculation of static regional paths, i.e. we do not consider the effects that traffic dynamics have on the path selection by drivers. This will be subject to future research.

2.2.2. *Method 2*

This method consists in directly determining the paths on an aggregated graph, that is gathered based on topological features of the city network and its partition. It determines the internal and regional paths differently. In this paper, we filter the paths that cross more than one time the same region. This is to avoid potential problematic paths for the application of the MFD-based traffic models (Mariotte and Leclercq, 2019), such as paths that travel on the same region more than one time. Then, when the Origin and Destination regions match, we consider that there is only one possible internal path to be traveled. While, for determining the regional paths, we distinguish between two different variants of this graph.

In the first variant, we consider a regional graph. Fig. 3 (a) shows one example of a partitioned city network. We define ρ_{ij} as the set of the city network nodes that are located at the partition (or border) between two generic adjacent regions i and j. Fig. 3 (b) depicts the regional graph, where the nodes represent the regions. The question is how to define the gray edges connecting the nodes. These connections are dictated by the allowed travel directions of the nodes located at the borders between adjacent regions. For example, it is possible to travel in both directions between the yellow and blue regions. However, it is only possible to travel from the blue to the green region. This happens because the two border nodes only allow to travel in this direction. One possibility for determining the regional choice set $\Omega^{OD}, \forall (O, D) \in W$ is to calculate all regional paths, but even on a regional network this might lead to large computational costs. The computational cost increases exponentially with the number of regions. Instead, we determine only the K-shortest paths on the regional graph for reducing the length of the regional choice set $\Omega^{OD}, \forall (O, D) \in W$. To determine the shortest-paths on the regional graph, we assign the edge costs based only on topological features of the city network and its partitioning. We consider two cost functions for assigning the edge costs:

- *Variant 2.1:* We propose a capacity-oriented cost function, that accounts for the flow capacity (q^c_{af}) of each lane f of the incoming link a to each border node listed in ρ_{ij}, and that allows to travel from the generic region i to



(a) (b) (c)

Figure 3: (a) Partitioning of the city network. (b) Regional graph representing the regional network. (c) Variant of the regional graph, including the borders between adjacent regions as nodes.

j. The edge cost C_{ij} is determined as:

$$C_{ij} = \sum_{a} \sum_{f=1}^{N_{lanes}} \frac{1}{q^c_{af} \delta_{aij}}, \forall a \in A \wedge \forall i \in X \wedge \forall j \in \Lambda \tag{2}$$

where N_{lanes} is the total number of lanes of each incoming link a to each border node listed in ρ_{ij}; Λ is the set of adjacent regions to region i; and δ_{aij} is a binary variable that equals 1 if link a allows to travel from region i to region j. We consider a standard value of $q^c_{af} = 1800$ [veh/h], which is a fair assumption for a single link. We recall the reader that the border B_{ij} between two adjacent regions i and j, is located at nodes (i.e. intersections) of the city network. This value may be an overestimation of the real capacity of a given border node (or intersection), if traffic lights or stop signs are present. In this paper, we do not focus on the dynamic influence of the traffic signals or signal stops, and rather focus only on topological features of the network (i.e. static regional paths). Moreover, we consider a similar flow capacity q^c_{af} for each lane of each link. Then, from Eq. 1, the edge cost C_{ij} only depends on the total number of lanes of all incoming links to the border B_{ij} between two adjacent regions i and j. But, we emphasize that the edge cost defined in Eq. 1 is general. It can also account for the effects of traffic lights or stop signs by properly calibrating the flow capacity q^c_{af} per lane.

- *Variant 2.2:* We propose to define the edge costs (C_{ij}) based on the Euclidian distance between the centroids of two generic adjacent regions i and j:

$$C_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, \forall i \in X \wedge j \in \Lambda \wedge i \neq j \tag{3}$$

where $(x_i, y_i))$ and $(x_j, y_j))$ represent the cartesian coordinates of the centroids of regions i and j, respectively. The closest city network node to the geometric center of a generic region $i \in X$ represents its centroid.

In the second approach, we consider a more detailed definition of the regional graph, where the nodes represent the regions as well as the borders B_{ij} between two adjacent regions i and j. Fig. 3 (c) depicts one example of the more detailed regional graph. It has a total of $R + N_{borders}$ nodes, where $N_{borders}$ represents the total number of borders between adjacent regions of the network, and R is the total number of regions defining the regional network. While, the maximum number of edges is given by $\sum_{r \in X} N_{borders}(N_{borders} + 1) + R$. This value is only a maximum indicator

since the edges between the nodes depend on the allowed travel directions in the city network. For example, the topology of the city network only allows to travel from the blue to the green regions (Fig. 3 (a)). Then, it is only allowed to travel from the blue region 2 to the border node B_{24} and then to region 4, in the graph as depicted in Fig. 3 (c). There are three kind of edges in this variant of the regional graph:

i. the internal edges, that for simplicity of the illustration, we just show one example represented by the dashed gray line on the top of region 1;
ii. the edges connecting the centroids of the regions to the borders B_{ij}, and vice-versa. These are shown in Fig. 3 (c) by the full gray lines between the borders B_{ij} and the centroids of the regions.
iii. the edges connecting two adjacent regions i and j to region m, i.e. the connection between B_{ij} to B_{jm}. For simplicity of the illustration purposes, we just show one example of this kind of edges in Fig. 3 (c), that is represented by the dotted gray line between the borders B_{15} and B_{54}, traveling through region 5.

This induces differences on the regional path construction on both variants of the regional graphs depicted in Fig. 3 (b-c). Consider as an example the regional path $p = \{154\}$. In the case of the regional graph shown in Fig. 3 (b), this path would only contain the edges connecting regions 1 to 5 and then 5 to 4. While, in the variant of the regional graph shown in Fig. 3 (c), this path contains the edge connecting region 1 to the border B_{15}, then the edges connecting the border B_{15} to B_{54} passing through region 5, and finally the edge connecting the border B_{54} to region 4. Once this sequence is identified, we determine the regional path following the regions traveled.

The regional choice set $\Omega^{OD}, \forall (O, D) \in W$ is also determined by the computation of the K-shortest paths in this graph. We also consider two different ways of assigning the edge costs:

- *Variant 2.3:* We propose to use the average travel distances between the centroid node of region i and the border nodes with the adjacent region j, and vice-versa; and the average travel distance to cross region m, by traveling from the border with adjacent region i to the border with adjacent region j. We define $L_{i,\rho_{ij}} = \{l_k\}$ as the set of trip lengths between the centroid node of region i and all border nodes listed in ρ_{ij}. Similarly, we define $L_{\rho_{ij},j} = \{l_k\}$ as the set of trip lengths between all border nodes listed in ρ_{ij} and the centroid node of region j. These sets are determined by the computation of all possible shortest-trips in distance between the centroid node of region i and all border nodes listed in ρ_{ij}, and vice-versa. For the edges connecting two borders, we define $L_{\rho_{im},\rho_{mj}} = \{l_k\}$ as the set of trip lengths connecting all border nodes between region m and adjacent region i (i.e. all nodes listed in ρ_{im}), to all border nodes between region m and adjacent region j (i.e. all nodes listed in ρ_{mj}). We also determine these sets by calculating all possible shortest-trips in distance that connect all border nodes listed in ρ_{im} to the ones listed in ρ_{mj}. From these sets of trip lengths, we determine the average travel distances, and update the edge costs as:

$$C_{ij} = \begin{cases} \overline{L}_{i,\rho_{ij}} & \text{if travel from } i \text{ to border } \rho_{ij} \\ \overline{L}_{\rho_{ij},j} & \text{if travel from border } \rho_{ij} \text{ to } j \\ \overline{L}_{\rho_{im},\rho_{mj}} & \text{if region } m \text{ is crossed, when traveling from regions } i \text{ to } j. \end{cases} , \forall i \in X \wedge \forall j \in \Lambda \wedge i \neq j \quad (4)$$

- *Variant 2.4:* We propose to use as a metric, the Euclidian distance between the centroid node of region i and the centroid node n_{ij} of the border B_{ij}; between the same n_{ij} and the centroid of region j; and between the centroid nodes n_{im} and n_{mj} of the borders B_{im} and B_{jm}, respectively, when region m is crossed. We also assign, by default, a unitary cost for internal edges. The edge costs are then updated as:

$$C_{ij} = \begin{cases} \overline{L}_{i,n_{ij}} & \text{if travel from centroid node of region } i \text{ to } n_{ij} \\ \overline{L}_{n_{ij},j} & \text{if travel from } n_{ij} \text{ to centroid node of region } j \\ \overline{L}_{n_{im},n_{mj}} & \text{if region } m \text{ is crossed, when traveling from regions } i \text{ to } j. \end{cases} , \forall i \in X \wedge \forall j \in \Lambda \wedge i \neq j \quad (5)$$

2.3. A comparative analysis between both methods

We briefly describe in this section, the advantages and limitations of each of the two methods. The first method is computationally expensive since it requires the computation of a set of virtual trips, i.e. including the calculation of several shortest-trips in distance, if real data is not available. Moreover, the main challenge of this method lies on the optimal calibration of N_{od}. The computation of a large set of virtual trips on large city networks can become unfeasible. The main advantage of this method lies on the fact that it allows to directly determine the TLD of regional paths, as well as their level of significance, for the application of the MFD-based traffic models (Batista et al., 2019). In the next section, we discuss how to properly calibrate N_{od}.

The second method is more parsimonious and computes the K-paths with much lower computational costs. The scaling of the city into the regional network allows to have a tractable number of regions, reducing the complexity of the problem and the computational power required for the calculation of paths. The calibration of the edge costs of the regional graph only relies on topological features of the city network as well as of its partition. But, despite being computationally lighter and tractable, this method does not allow to determine the TLD. One may use the Euclidian distance as a proxy to characterize the trip lengths of the paths. However, the Euclidian distance is not necessarily representative of the TLD determined from the set of virtual trips, that captures the topology of the city network, and which are recognized to play an important role in MFD-based applications (Batista et al., 2019). In the next section, we investigate the similarities of the paths calculated by both methods. The question is if this method finds similar paths as the most prevailing ones found by the first method.

3. Regional paths and choice sets analysis

In this section, we discuss the implementation of the two methods previously introduced for calculating paths on regional networks. We start by introducing the test network. We then show how to properly calibrate the set of virtual trips required for Method 1. We also analyze and discuss the similarities between the choice sets calculated through the different variants of Methods 1 and 2.

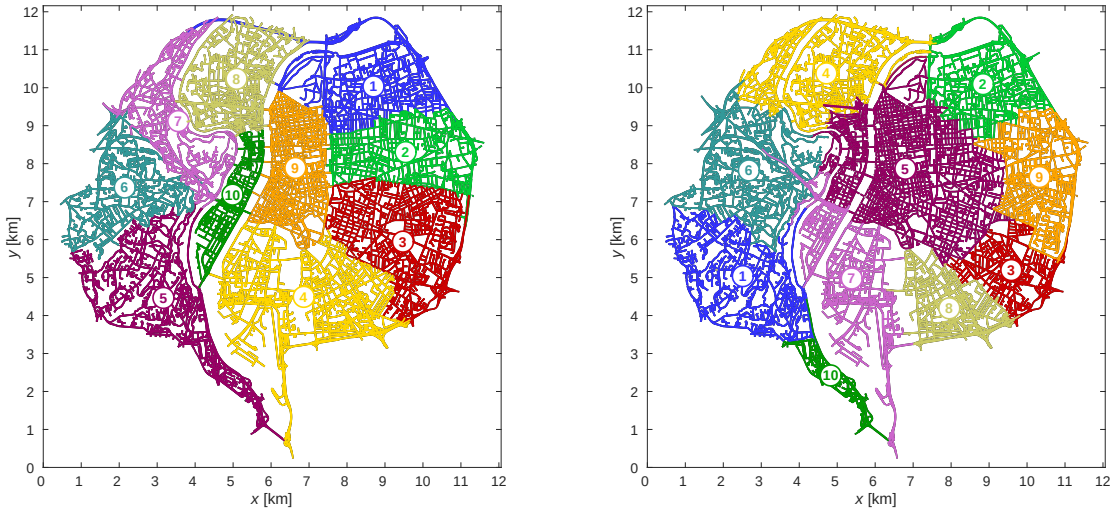
3.1. Test scenario and network definition

The test network is depicted in Fig. 4, and corresponds to the metropolitan area of Lyon (France). The network has 19697 nodes and 19967 links. It is partitioned into 10 regions, based on administrative regions defined by the municipality. In this paper, we consider two different definitions of the network partitioning as depicted in Fig. 4 (a) (Partitioning 1) and (b) (Partitioning 2).

We construct a set of virtual trips Θ, where each node in the city network is a possible origin or destination of a travel. We consider all possible combinations of origin and destination nodes in the city network, for defining Θ. This yields a total of 108,529,021 virtual trips, i.e. $N^{od} = 108,529,021$. Of course, constructing such a large set of virtual trips requires large computational resources. In the case of the Lyon network, these calculations took approximately 10 days. The data reduction also took approximately one week. In this paper, we determined this full set of virtual trips to set it as the reference and for the purpose of our analysis. In the next section, we discuss in more detail the calibration of the virtual set of trips.

The full set of paths Φ is determined by scaling-up all of these virtual trips listed in Θ, according to the city network partitioning. We obtain a total of 2423 and 3355 paths, concerning Partitioning 1 (Fig. 4 (a)) and 2 (Fig. 4 (b)), respectively. We filter the paths that cross more than one time the same region. This allows to avoid meaningful paths for the application of the aggregated multi-regional MFD models. We are left with a total of 689 paths determined from 91,342,632 valid trips, for Partitioning 1 (Fig. 4 (a)). While for Partitioning 2 (Fig. 4 (b)), we have a total of 564 paths determined from 87,799,074 valid trips. This means that there are approximately $\sim 84\%$ and 81% of valid trips for Partitioning 1 and 2, respectively. We define Φ_1 as this set of valid paths.

We also define a set ζ that contains all real trips of drivers in the city network. We refer hereafter to this set as the real data, that comes from a GPS data set provided by an European navigation system operation. We consider the full month of March 2018 in our analysis. We would like to emphasize that most of the GPS trajectories in our dataset correspond to trips of drivers that are located in our study area. These trips represent the morning and evening commute of people in the city center. The trajectories of drivers are map-matched with the Lyon network depicted in Fig. 4. The real data set covers a larger area than the Lyon metropolitan area. Each entry on the original



(a) (b)

Figure 4: Metropolitan network of Lyon (France), divided into 10 regions and considering two distinct definitions of the partitioning as shown in panels: (a) Partitioning 1; (b) Partitioning 2.

database characterizes the passage of a given vehicles in one specific link of the city network, that defines its trajectory. Each entry contains the information about the vehicle label, the identification of the link traveled, a timestamp at the entrance of the link, an estimation of the travel speed, and the coverage percentage of the link. By relating the coverage information with the link geometry, we compute the distance traveled by each vehicle on each link. We also need to do a pre-processing of the data for detecting potential static phases from the individual GPS tracks, and then parse them into separate trips. We have filtered from our analysis, the trips that have an origin and/or destination outside the city network depicted in Fig. 4, as well as the trips that are excessively fragmented. Additionally, we have also filtered out redundant vehicles from our analysis. We are left with a set ζ containing 101,729 trips. These trips are scaled-up concerning the two definitions of the city network partitioning. We are left with a total of 2470 and 2226 paths regarding Partitioning 1 (Fig. 4 (a)) and 2 (Fig. 4 (b)), respectively. We have then filtered the trips that cross more than one time the same region. This leads to a final set of 86,114 and 87,694 trips, yielding 473 and 335 valid paths, for Partitioning 1 (Fig. 4 (a)) and 2 (Fig. 4 (b)), respectively. We define Φ_3 as the set of all regional paths gathered from the real trajectories of drivers listed in ζ.

3.2. Calibration of Method 1

The application of Method 1 requires the proper definition of a set of virtual trips Θ, that should provide a good city network coverage. In order words, the question relies on finding the value of N_{od} that provides an optimal city network coverage, where the calculation of the regional paths will no longer be influenced by the bias introduced by the random sampling of od pairs. We define a criterion that estimates the percentage of the city network links ($N_{cov}^{links}(N_{od})$) that are traveled by virtual trips in the set Θ, with length N_{od}. Mathematically, we determine $N_{cov}^{links}(N_{od})$ as:

$$N_{cov}^{links}(N_{od}) = \sum_{i=1}^{N_{links}} \frac{N_{used}^{links}(N_{od})}{N_{links}} \tag{6}$$

where $N_{used}^{links}(N_{od})$ is the total number of links traveled by the trips listed in Θ; and N_{links} is the total number of links defining the city network, i.e. 19967 for the Lyon network depicted in Fig. 4.

In this analysis, we focus on Partitioning 1 (see Fig. 4 (a)). We consider ten different settings of N_{od}: $1{\times}10^2$, $5{\times}10^2$, $1{\times}10^3$, $5{\times}10^3$, $1{\times}10^4$, $5{\times}10^4$, $1{\times}10^5$, $5{\times}10^5$, $1{\times}10^6$ and $1{,}085{\times}10^9$. Note that, $N_{od} = 1,085.10^9$ represents all possible combinations of od pairs of nodes in the city network, i.e. the full enumeration of possible virtual trips. For each setting of N_{od}, we run fifteen trials to determine the set of virtual trips Θ, and then derive the city network coverage using Eq. 6. In the V1.1, we sample the od pairs of nodes independently of the city network partitioning. In V1.2, for each setting of N_{od}, we sample $N_{od}/100$ od pairs of nodes in the city network that are located inside each of the possible 100 Origin and Destination pairs of regions. Fig. 5 depicts the $N_{used}^{links}(N_{od})$ distributions determined based on all fifteen trials for each setting of N_{od}, and both variants of Method 1. The red dots represent the outliers of the distributions. Larger values of N_{od} ensure a better coverage of the city network, and reduce the variability of the regional paths determined for the different regional OD pairs.

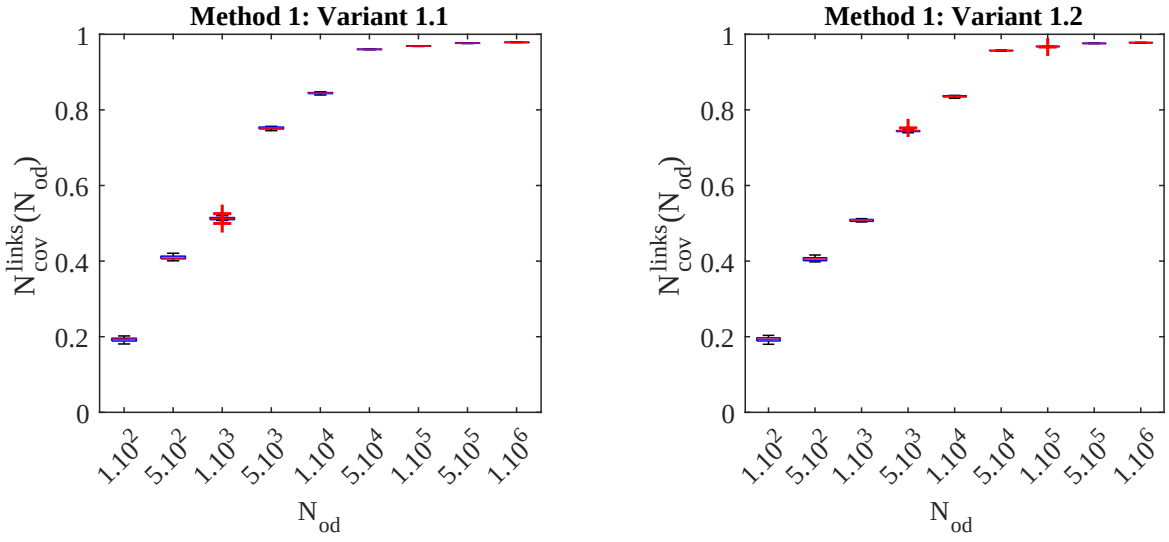


Figure 5: Evolution of $N_{cov}^{nodes}(N_{od})$ as function of N_{od}, for the two variants of Method 1.

Table B.1 and Table B.2 show how the regional choice sets Ω^{OD}, for several OD pairs, vary with the increase of N_{od}. We consider a total of eight OD pairs: 42; 24; 75; 57; 25; 52; 37; and 73. This choice allows to investigate simple tests (i.e. OD pairs where the regions are close, such as regions 2-4 and 5-7), as well as harder tests of longer paths (i.e. Origin and Destination regions which might lead to paths with more than four regions, like the regions 2-5 and 3-7). The results are listed for both variants of Method 1, and only for one trial for each of the N_{od} values. The regional paths are listed from the first to the third most prevailing one, for both regional OD pairs. We recall the reader that the total number of virtual trips linked to a regional path defines its level of significance.

We first analyze the regional choice sets determined for V1.1. For a low $N^{od} = 100$, we only have one regional path connecting the OD pairs 42 (i.e. $p = \{4, 9, 2\}$), 24 (i.e. $p = \{2, 3, 4\}$), 57 (i.e. $p = \{5, 6, 7\}$) and 37 (i.e. $p = \{3, 4, 9, 10, 7\}$). However, there are no paths for the remaining OD pairs. As N^{od} increases, more paths are found connecting several of the OD pairs. While N^{od} is still low, i.e. when it still does not yet ensure a good city network coverage as depicted in Fig. 5, we can find the same regional paths, but listed with a different level of significance. This is observed for the OD pairs 42, 24, 75 and 57. One example is the regional choice set Ω^{42} determined for $N^{od} = 1 \times 10^3$ and $N^{od} = 5 \times 10^3$. In the first case, we find $p = \{4, 3, 2\}$ and $p = \{4, 9, 3, 2\}$ as the first and second most prevailing regional paths, respectively. While, in the second case, we also find these two regional paths but with different levels of significance. For $N^{od} > 1 \times 10^5$, we find similar Ω^{75} independently of N^{od}, where the regional paths are listed with the same level of significance. This is verified when we have ensured a good city network coverage, and the calculation of the regional choice set $\Omega^{OD}, \forall (O, D) \in W$ becomes independent of N^{od}. While, the other OD

pairs require a larger N^{od}. For example, regions 2-5 and 3-7 are located on opposite directions of the network. We emphasize to the reader that the city of Lyon is crossed by the rivers Saône and Rhône, which merge at the bottom of region 10 (see Fig. 4 (a)). These two rivers act as natural borders between regions, for the partitioning depicted in Fig. 4 (a). To travel between regions 2-5 and 3-7, one has to cross at least one of the two rivers. There are only a few bridges that allow to cross the rivers, which naturally increases the difficulty on finding the regional choice set Ω^{OD} for the OD pairs 25, 52, 37 and 73. This highlights a strong limitation of the variant V1.1, where trips are sampled independently of the city network partitioning, thus not ensuring that all paths are found for harder test cases. Another intriguing case are the OD pairs 42 and 24. For OD 42, we are able to find the same paths with the same level of significance for $N^{od} > 1 \times 10^5$. However, the OD 24 requires almost full enumeration to find the three most prevailing paths. This happens because, and as previously discussed, variant V1.1 samples trips independently of the city network partitioning, not ensuring that all paths are found. The complex topological features of the city network (e.g. the presence of several one-way streets) concerning its partitioning plays an important role. This is a limitation of variant V1.1, putting in evidence the importance of constructing the set of virtual trips concerning the definition of the partitioning (i.e. V1.2). In fact, for variant V1.2, we are able to find all three paths of the regional choice set Ω^{24}, for a lower $N^{od} = 5 \times 10^5$. This is also a general trend for the other OD pairs, where a lower city network coverage is required such that Ω^{OD} becomes independent of N^{od}.

Overall, the sampling of od pairs of nodes in the city network accounting for the definition of the partition, allows to determine similar regional choice sets Ω^{OD} for a lower N^{od}, than in the case of variant V1.1. This answers our first conjecture in the Introduction, showing that variant V1.2 is a better approach for determining the set of virtual trips.

3.3. Calculation of paths in regional networks: a comparative analysis between the methods

In this section, we analyze the similarity between the sets of paths determined by Method 2 and Method 1 against each other, as well as against the set of paths gathered from real data, i.e. GPS trajectories. We consider the two definitions of the partitioning of the city network as depicted in Fig. 4. We focus our analysis on the regional choice sets encompassing the three most significant paths.

We apply Method 2 as follows. In the case where the Origin and Destination regions match, the regional choice set only consists of one internal path. When the Origin and Destination regions are different, we determine the regional choice set Ω^{OD} for all $(O, D) \in W$, based on the three shortest-cost paths. The path costs are updated for the four variants of Method 2, following Equation 2 to Equation 5 (see Sect. 2.2.2 for more details). We define $\phi_{V2.j}, j = 1, \ldots, 4$, as the set of calculated paths for the $j-$th Variant of Method 2. This set can be mathematically expressed as:

$$\phi_{V2.j} = \left\{ \bigcup_{(O,D)\in W} \Omega^{OD} \right\}, j = 1, \ldots, 4 \tag{7}$$

In the case of Method 1, we focus on the set of virtual trips with $N_{od} = 1,085 \times 10^9$, where Φ_1 is this set of all paths. We define $\phi_1 \in \Phi_1$ as the set containing the three most prevailing paths for each regional $(O, D) \in W$. This set can be mathematically defined in a similar way as $\phi_{V2.j}, j = 1, \ldots, 4$. We also define ϕ_3 as the set containing the three most prevailing paths for each regional OD pair, determined from the real GPS trajectories. Mathematically, we can define ϕ_3 in a similar way as ϕ_1, where in this case, the regional choice set Ω^{OD} is defined from paths gathered from the real trajectories of drivers.

Our analysis between the sets ϕ_1 and ϕ_3 against $\phi_{V2.j}, j = 1, \ldots, 4$, is based on a similarity and strict similarity criteria. The similarity criterion only evaluates how alike are the sets of paths, without looking into their significance rank. The strict similarity criterion counts the number of paths that are found in both sets with the same significance level. However, we would like to reinforce that the most important aspect is that our Methods 1 and 2 are able to determine similar sets of paths, also in comparison with the real GPS trajectories.

We determine the similarity criterion $\alpha^{\phi_i \phi_{V2.j}}$ as:

$$\alpha^{\phi_i \phi_{V2.j}} = \frac{|\phi_i \cap \phi_{V2.j}|}{|\phi_{V2.j}|}, i = 1, 3 \wedge j = 1, \ldots, 4 \tag{8}$$

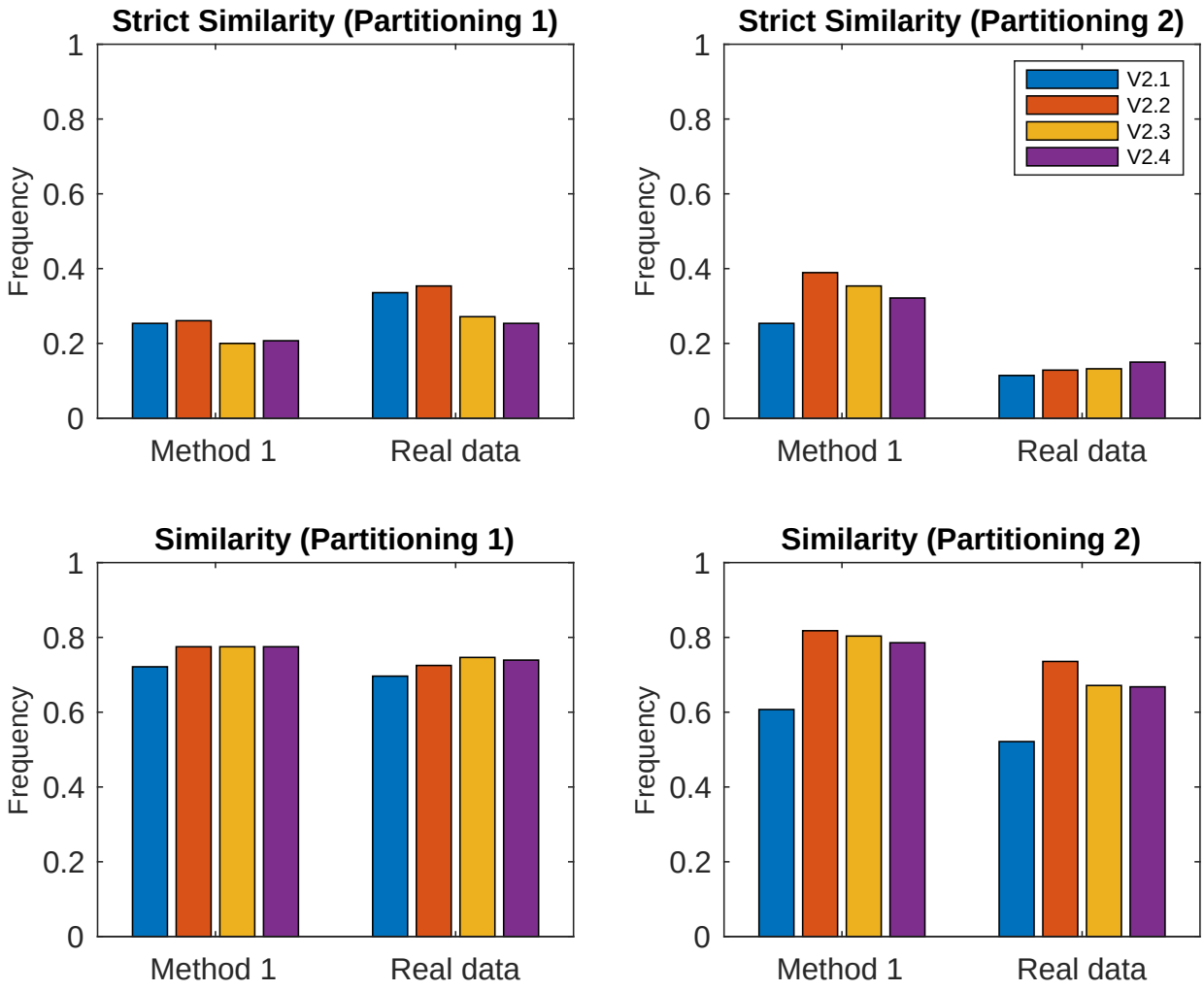


Figure 6: Similarity and strict similarity criteria between the four variants of Method 2 and Method 1, and the real data. The left panels show the results for Partitioning 1, while the right panels show the results for Partitioning 2.

where |.| represents the length of the set.

The strict similarity criterion $\beta^{\phi_i \phi_{V2.j}}$ is:

$$\beta^{\phi_i \phi_{V2.j}} = \frac{\sum_{(O,D) \in W} \sum_p \delta_p^{OD}}{|\phi_{V2.j}|}, i = 1, 3 \wedge j = 1, \ldots, 4 \tag{9}$$

where δ_p^{OD} is a binary variable that equals 1 if path p that connects the regional OD pair is ranked with a similar level of significance in $\phi_i, i = 1, 3$, and ϕ_2, or 0 otherwise.

Fig. 6 depicts the similarity and strict similarity results of the four variants of Method 2 (i.e. $\phi_{V2.j}, j = 1, \ldots, 4$), compared to Method 1 (i.e. ϕ_1) as well as the real data (i.e. ϕ_3). The left panels depict the results for Partitioning 1, while the right panels show the results for Partitioning 2. We observe that the similarity values range between ~ 70-80%, for Partitioning 1. The performance of these three variants are similar to some extent. However, the similarity values drop to ~50-60% for V2.1 and Partitioning 2. The similarity values for the variants V2.2 to V2.4 are maintained around ~80% when compared to the simulated data, i.e. Method 1. While, they slightly decrease

for the variants V2.3 and V2.4 when compared to the real data. On the other hand, we also observe that Method 2 provides low strict similarity values of ~10-30% for Partitioning 1 and ~10-40% for Partitioning 2, when compared to the simulated (Method 1) and the real data. This means that the shortest-cost paths, where the costs are updated according to Equation 2 to Equation 5 (i.e. V2.1 to V2.4), do not correspond in general to the most prevailing regional paths determined by Method 1, or from the real data. Despite this fact, we emphasize that the most important factor is the ability of Method 2 in finding similar paths as both Method 1 and the real data. This is enhanced by the large similarity values, where the Euclidean distance between the centroids of the regions (i.e. variant V2.2) is the best proxy to calibrate Method 2, and determine paths. Moreover, variant V2.2 requires a simpler definition of the regional network (see Fig. 3 (b)) and is also based on a simpler Euclidean distance metrics between the centroids of the regions. This answers our second conjecture listed in the Introduction.

We also analyze the similarity and strict similarities between the sets ϕ_1 (Method 1) and ϕ_3 (real data). The similarity results are 80% and 80% for Partitioning 1 and 2, respectively. While, the strict similarity results are 19% and 28% for Partitioning 1 and 2, respectively. The large similarity values show that the assumption of shortest-trips in distance to calibrate Method 1 represent a good proxy for determining the actual paths chosen by drivers. It is known that drivers do not necessarily choose the shortest-trips in distance for their travels in the city network (Zhou et al., 2014). However, these trips represent similar paths in the regional network as the ones determined by Method 1. On the other hand, the low strict similarity values can be explained by the different weights associated to each path, as both set of trips represent different travel patterns in the city network. Overall, these results answer our third conjecture listed in the Introduction. The set of virtual trips in distance is a good proxy to determine the most prevailing paths chosen by drivers.

We also analyze the influence of the partitioning on the similarity and strict similarity results between both methods 1 and 2 as well as the real data. The regions of Partitioning 2 (Fig. 4 (b)) are more compact than in the case of Partitioning 1 (Fig. 4 (a)). We observe similar similarity values between Method 1 and 2 as well as between Methods 1 and 2 and the real data. While, there are some differences in terms of the strict similarity. It is in general larger for the four variants of Method 2 concerning Method 1, for the case of Partitioning 2 when compared to Partitioning 1. However, these values are much lower when looking at Method 2 concerning the real data, for Partitioning 2. The strict similarity is also larger for Partitioning 2, when we compare Method 1 and the real data. These results answer our fourth conjecture listed in the Introduction. The definition of the partitioning does not appear to have an influence on the similarity results between the sets, as it does on the strict similarity. Therefore, we conclude that the definition of the partition does not influence the performance of both Methods 1 and 2 for identifying the most prevailing paths chosen by drivers.

We also would like to briefly comment on the computational costs of Methods 1 and 2. Method 1 requires the knowledge about a set of trip patterns in the city network, which can come from real or simulated data. If one utilizes a set of real GPS data patterns, there is the need to do a curation of the raw data, which includes the map-matching of the data as well as the elimination of invalid trips (i.e. trips with missing information about the sequence of traveled links, unknown origin and/or destinations, etc.). In our paper, the reduction of the raw GPS data, for the whole month of March, to the final valid set of trips, took approximately two months. On the other hand, the construction of a set of virtual trips can also be a cumbersome task, as discussed in the previous section. However, we focus our attention on the computation of the full enumeration of trips, which took approximately one month of computations. Then, we need to reduce this data by filtering the trips concerning the sequence of traveled regions, following the definition of the city network partitioning. In our case, this process took approximately 2 weeks, to loop over all trips and determine their corresponding path on the regional network as well as its travel distances. Method 2 is much lighter in terms of computational power required. Once we have the definition of the partitioning, we can determine the regional network. We then determine shortest-cost paths directly on the regional graph, where the edge costs are assigned based on simple metric that utilize topological features of the city network as well as its partitioning. Once we have gathered the definition of the regional graph and set the edge costs, the calculation of the shortest-cost paths is almost instantaneous. In our test case, this takes approximately less than 2 minutes. The results discussed in this section, show the efficiency of Method 2 to determine the most prevailing paths as the ones determined from both simulated and real trip patterns in the city network.

4. Characterization of paths on regional networks

In this section, we discuss the characterization of paths through their TLD. We first investigate the characterization of an internal path for the whole metropolitan area of Lyon, using the set of virtual trips as well as the real data. We then investigate the characterization of regional paths, when the city network is partitioned into regions. Lastly, we investigate the relationship between the level of significance of a path and its average travel distance.

4.1. Trip Length Distribution of the whole network

In this section, we focus on the characterization of an internal path for the whole metropolitan area of Lyon, depicted in Fig. 4, using both the simulated (i.e. the set of virtual trips) and real data. The full city network is considered as a single region, where all vehicles travel on the same internal path. Fig. 7 depicts the trip length distributions, TLDs, for this internal path determined from the set of virtual trips and from the real data. We would like to emphasize that these trip length distributions are unimodal. While the simulated data shows to be a good approximation of a skewed normal distribution, the functional form of the TLD determined from real data is more difficult to infer. The real trips of drivers gathered from GPS traces favor longer trips than the shortest-one in distance. For example, the observed peaks at ~ 2, 4 and 7 kms, in the TLD of the real data are due to trips of drivers on the ring road that goes around the city of Lyon. These trips correspond to a large fraction of the dataset.

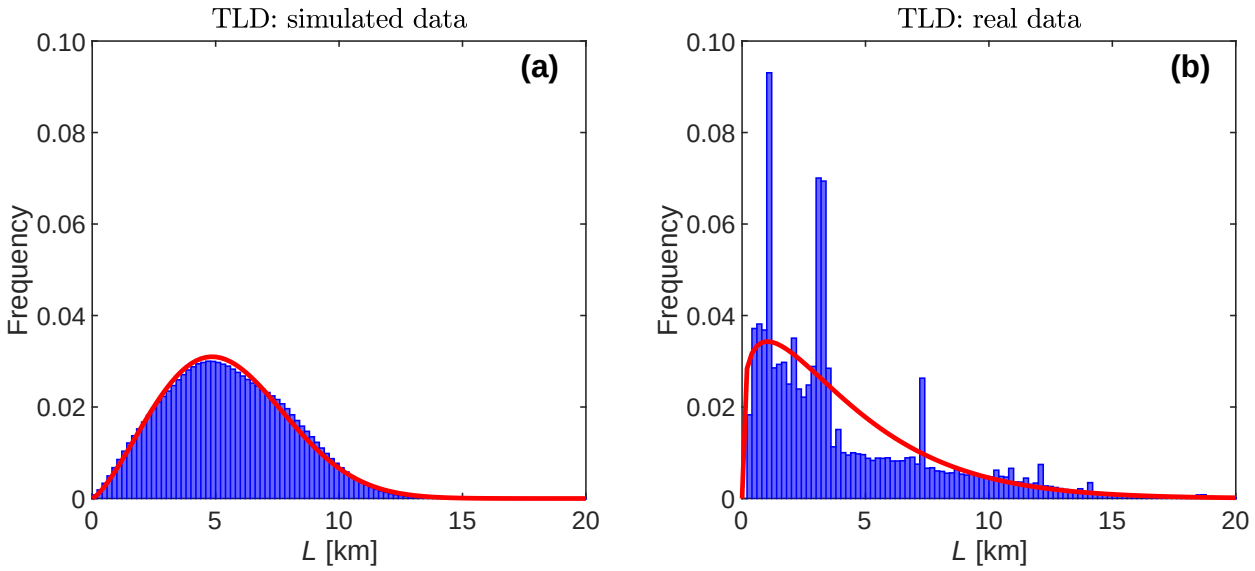


Figure 7: Unimodal trip length distribution (TLD) of the internal path, determined from: (a) the simulated data (i.e. set of virtual trips); (b) the real data. The red solid line represents the fitted Weibull distributions.

The functional form of the TLD has been subject of debate in many seminal works in the literature. Many of these studies focus on the home-based trips since they are an important part of our daily routine. The challenge lies in finding the best statistical distribution to fit the empirical data. Some authors utilize a Gamma or Weibull distributions (Ortuzar and Willumsen, 2011; Yang et al., 2013; Moeckel, 2017), while others focus on a log-normal (Katsis et al., 2014) or a negative exponential distributions (Yang et al., 2013). In the case of MFD-based applications, some authors consider constant TLD (e.g. Arnott, 2013; Mariotte et al., 2017), or a either an implicit (e.g. Daganzo, 2007) or explicit (e.g. Vickrey, 2020) negative exponential distribution. Lamotte and Geroliminis (2016) showed that the trip-based and accumulation-based MFD traffic models predict similar traffic dynamics if the TLD follow a negative exponential. In this paper, we focus on these unimodal four distributions, and investigate which is the best one that fits the TLDs determined from the simulated and real data (see Fig. 7). We utilize a maximum likelihood estimation (MLE) as a measure to estimate the parameters of these four distributions by maximizing their likelihood function. The fitted parameters and the values of the maximum log-likelihood functions of these four distributions are listed in Table 1.

The maximum log-likelihood estimation shows that the Weibull distribution is the one that better fits both TLDs. The fitted Weibull distributions are depicted by the red solid lines in Fig. 7. It is interesting to observe that the Weibull distribution is the one that better reproduces the functional form of both TLD, determined from the simulated and real data, despite both histograms being quite alike.

Distribution	Parameters of the log-likelihood estimation	Simulation Data	Real data
Gamma	k_{Ga}	15	3.9
	θ [m]	2855	1400
	$log(\mathcal{L}(k_{Ga}, \theta))$	-6.2581×10^6	-1.0033×10^9
NE	λ_{NE} [m]	2.41×10^{-4}	1.83×10^{-4}
	$log(\mathcal{L}(\lambda_{NE}))$	-6.2838×10^6	-1.0416×10^9
Log-normal	μ [m]	7.95	8.47
	σ^2 [m]	0.9159	0.5744
	$log(\mathcal{L}(\mu, \sigma^2))$	-5.3212×10^{12}	-4.1175×10^{15}
We	λ_{We} [m]	4433	6174
	k_{We} [m]	1.2	2.3
	$log(\mathcal{L}(\lambda_{We}, k_{We}))$	-7.359×10^5	-1.450×10^8

Table 1: Fitted parameters of the Gamma (Ga), Negative Exponential (NE), log-normal and Weibull (We) distributions, for both the simulated and real data. The values of the maximum log-likelihood functions are also listed.

These results answer our fifth conjecture listed in the Introduction. The best functional form to represent the network-wide TLD is a Weibull distribution, based on the maximum likelihood estimation results.

4.2. Trip Length Distributions and regional paths

In this section, we focus on the characterization of paths on regional networks. First, we discuss a major difference between regional paths. For this, we focus on the regional paths $p = \{1, 9, 4\}$ and $p = \{4, 9, 1\}$, and investigate the differences between their trip length distributions, based on Partitioning 1. By definition, these paths are distinct from each other since they cross a different ordered sequence of regions. For example, region 1 is the Origin for path $p = \{1, 9, 4\}$. But, it represents the destination region for path $p = \{4, 9, 1\}$. The main question is how alike are the trip length distributions that characterize each region of these two regional paths. Fig. 8 depicts the trip length distributions for these two regional paths, determined from the simulated data for Partitioning 1. The horizontal black dashed lines represent the average trip lengths of each region. We can observe that the average trip length for regions 1 and 4 are close for both regional paths, however, the shape of the trip length distribution is completely different. The trip length distributions of Origin and Destination regions of paths are influenced by: (i) the spatial distribution of origin and destination nodes of the city network inside these regions; and (ii) the city network topology. The case of the intermediate regions of paths, such as region 9 of $p = \{1, 9, 4\}$ and $p = \{4, 9, 1\}$, is different. The trip length distributions of these regions are only influenced by the city network topology, to go from one border to another. This is true only because the TLD is determined from a set of shortest-trips in distance (i.e. static trips). As evidenced by the trip length distribution peak of region 9 and $p = \{1, 9, 4\}$, the preferential sequences of links to go from regions 1 to 4 have similar travel distances. However, the preferential sequence of links to go from regions 4 to 1 have different travel distances, as shown by the two predominant peaks. This happens thanks to the allowed travel directions of the city network links, to go from one the border with region 1 to the one with region 4, and vice-versa. These results prove that the ordered sequence of regions to go from an Origin to a Destination region matter, agreeing with our definition of regional path. On the other hand, these results clearly reinforce the fact that one should not make any prior assumptions about the functional form of the TLDs of regional paths, as they can differ from region to region and among paths. This answers our sixth conjecture listed in the Introduction. A more detailed analysis about the time-dependence of the TLD of a regional path is left for future work.

We also investigate how different are the trip length distributions determined from the simulated data, i.e. the set of virtual trips, and the real data gathered from the GPS trajectories. For this purpose, we focus on regional

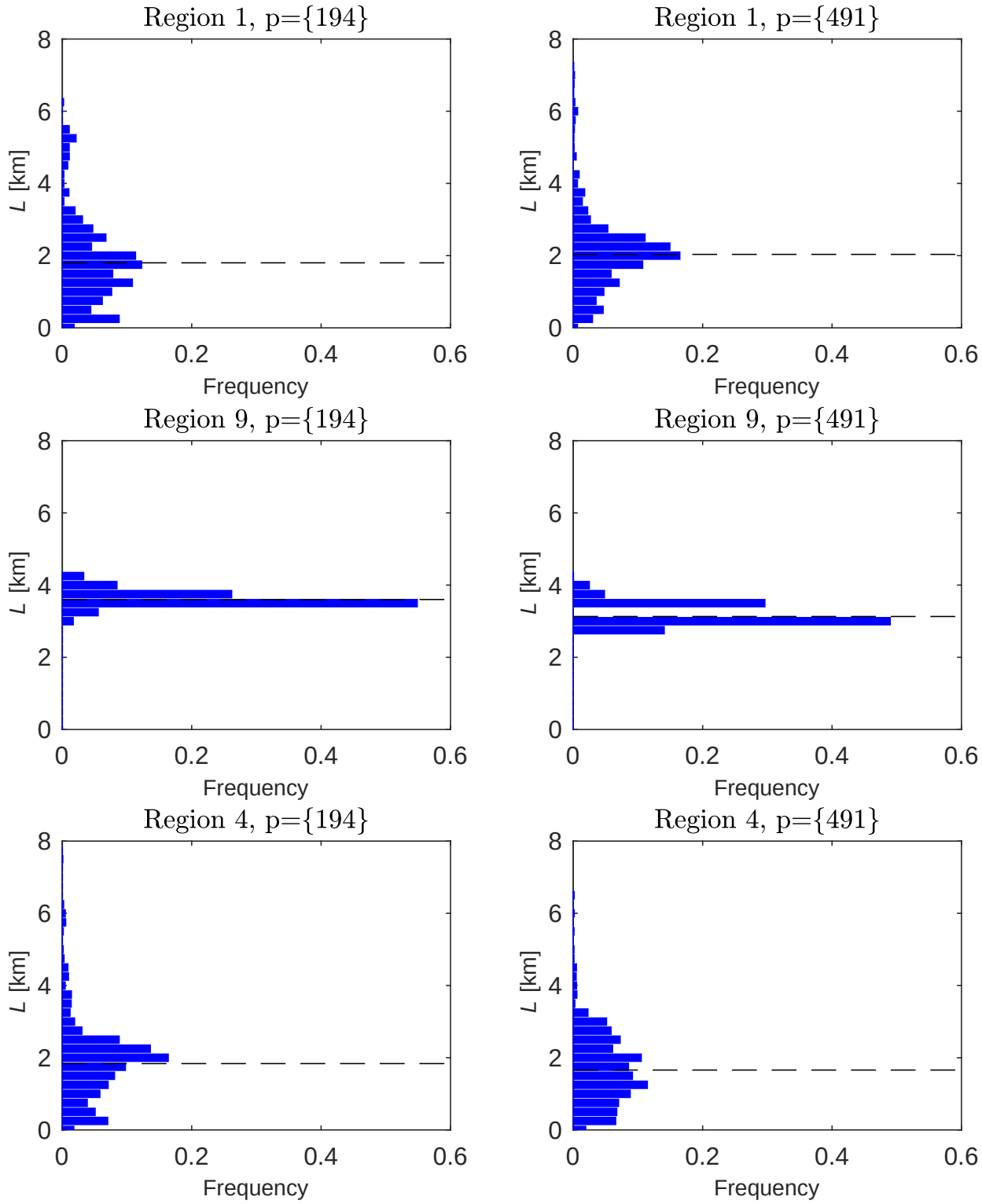


Figure 8: Trip length distributions for the regional paths $p = \{1, 9, 4\}$ (left panel) and $p = \{4, 9, 1\}$ (right panel). The horizontal black dashed lines represent the average trip length of each distribution. These distributions are determined from the simulated data, i.e. the set of virtual trips, for Partitioning 1.

path $p = \{4, 3, 2\}$ and Partitioning 1. Fig. 9 depicts the trip length distributions for the regions of this regional path, determined from the simulated and real data. The average travel distances calibrated for the variants V2.3 and V2.4 are also depicted in the figure. We note that V2.2 is not considered in this analysis, since the travel distance corresponds to the Euclidean distance between the centroids of the regions, and the distance traveled in each region is unknown. The horizontal black dashed lines also represent the average of the distributions. The horizontal red dashed lines represent the average shortest travel distance between the centroid node of region 4 and the border with region 3 B_{43}, between the borders B_{43} and B_{32}, and between the border B_{32} and the centroid node of region 2. The horizontal cyan dashed lines represent the Euclidean distances between the centroid node of region 4 and the centroid node n_{43} of the border B_{43}, between the centroids n_{43} and n_{32} of the borders B_{43} and B_{32}, respectively, and between the centroid node n_{32} and the one of region 2. As one can observe, the average trip lengths and the shape of the distributions are different. There are two factors that may explain these differences. First, drivers do not necessarily choose the shortest-trips in distance (e.g. Zhou et al., 2014), to travel from their origin to the destination node in the city network. Second, the uniform sampling of origin-destination pairs of nodes in the city network might not yield a representative set of the real travel patterns of drivers. We can also observe that the shortest average travel distance between centroid nodes of regions and their borders and between borders (i.e. the variant V2.3), seems to be a good proxy of the average travel distances of regions 3 and 2, for both the simulated and real data cases. This fact is also true for the Euclidean distance between the centroid nodes of the regions and the ones of the borders, as well as between the centroid nodes of borders (i.e. the variant V2.4).

In the follow-up of this analysis, we investigate how the calculated average trip lengths for the regions, change between the common paths listed in Φ_1 (i.e. set of paths gathered from the simulated data SD) and in Φ_3 (i.e. set of paths gathered from the real GPS trajectories). In this analysis, we focus on both Partitioning 1 and 2. We define $\overline{L}_{rp}$ as the average travel distance in a generic region r that defines a generic regional path p. In this analysis, we focus on the Origin, Intermediate and Destination regions of the paths separately. We determine the relative differences ϵ between the average travel distances $\overline{L}_{rp}^{\Phi_1}$ and $\overline{L}_{rp}^{\Phi_3}$ of the common regional paths in the sets Φ_1 and Φ_3:

$$\epsilon = \frac{\overline{L}_{rp}^{\Phi_1} - \overline{L}_{rp}^{\Phi_3}}{\overline{L}_{rp}^{\Phi_3}} \times 100\%, \forall r \in p \wedge \forall p \in (\Phi_1 \cap \Phi_3) \tag{10}$$

Note that this equation is valid to determine ϵ for the Origin, Intermediate and Destination regions of paths.

We also do a similar analysis between the trip lengths determined for the common paths in $\phi_{V2.3}$ and $\phi_{V2.4}$, i.e. the sets of paths gathered through the variants V2.3 and V2.4, and Φ_1 and Φ_3. The question here is if the shortest average trip lengths (V2.3) and the Euclidean distance (V2.4) proxies provide similar average travel distances in the regions, as the ones determined from the simulated and real data, and for the common regional paths. In a similar way as in Eq. 10, we also determine the relative differences ϵ between the travel distances as:

$$\epsilon = \frac{\overline{L}_{rp}^{V2.j} - \overline{L}_{rp}^{\Phi_i}}{\overline{L}_{rp}^{\Phi_i}} \times 100\%, \forall r \in p \wedge \forall p \in (\Phi_1 \cap \Phi_3) \wedge \forall i = 1, 3 \wedge \forall j = 3, 4 \tag{11}$$

where $\overline{L}_{rp}^{V2.j}$ is the trip length determined for the generic region r of regional path p, by the j-*th* variant of Method 2. In this case, we focus on the variants V2.3 and V2.4.

Fig. 10 depicts the box-and-whisker diagrams of the relative differences ϵ for the Origin, Intermediate and Destination regions of the common paths found in the different combination of sets: Φ_1 and Φ_3 (case SD-GPS); Φ_1 and $\phi_{V2.3}$ (case SD-V2.3); Φ_3 and $\phi_{V2.3}$ (case GPS-V2.3); Φ_1 and $\phi_{V2.4}$ (case SD-V2.4); and Φ_3 and $\phi_{V2.4}$ (case GPS-V2.4). The red dots represent the outliers of the distributions. The left panels represent the results for Partitioning 1. While, the right panels depict the results for Partitioning 2.

We first analyze the relative differences of the average trip lengths $\overline{L}_{rp}$ of the common regional paths in the sets Φ_1 and Φ_3, i.e. case SD-GPS. The median of the box-and-whisker diagrams of ϵ are close to 0, meaning that $\overline{L}_{rp}^{\Phi_3} \approx \overline{L}_{rp}^{\Phi_1}$. This is observed for both Partitioning 1 and 2. We also observe that the distribution is slightly skewed towards positive values. This means that $\overline{L}_{rp}^{\Phi_3} \gtrsim \overline{L}_{rp}^{\Phi_1}$. The latter is explained by the difference in the travel patterns between

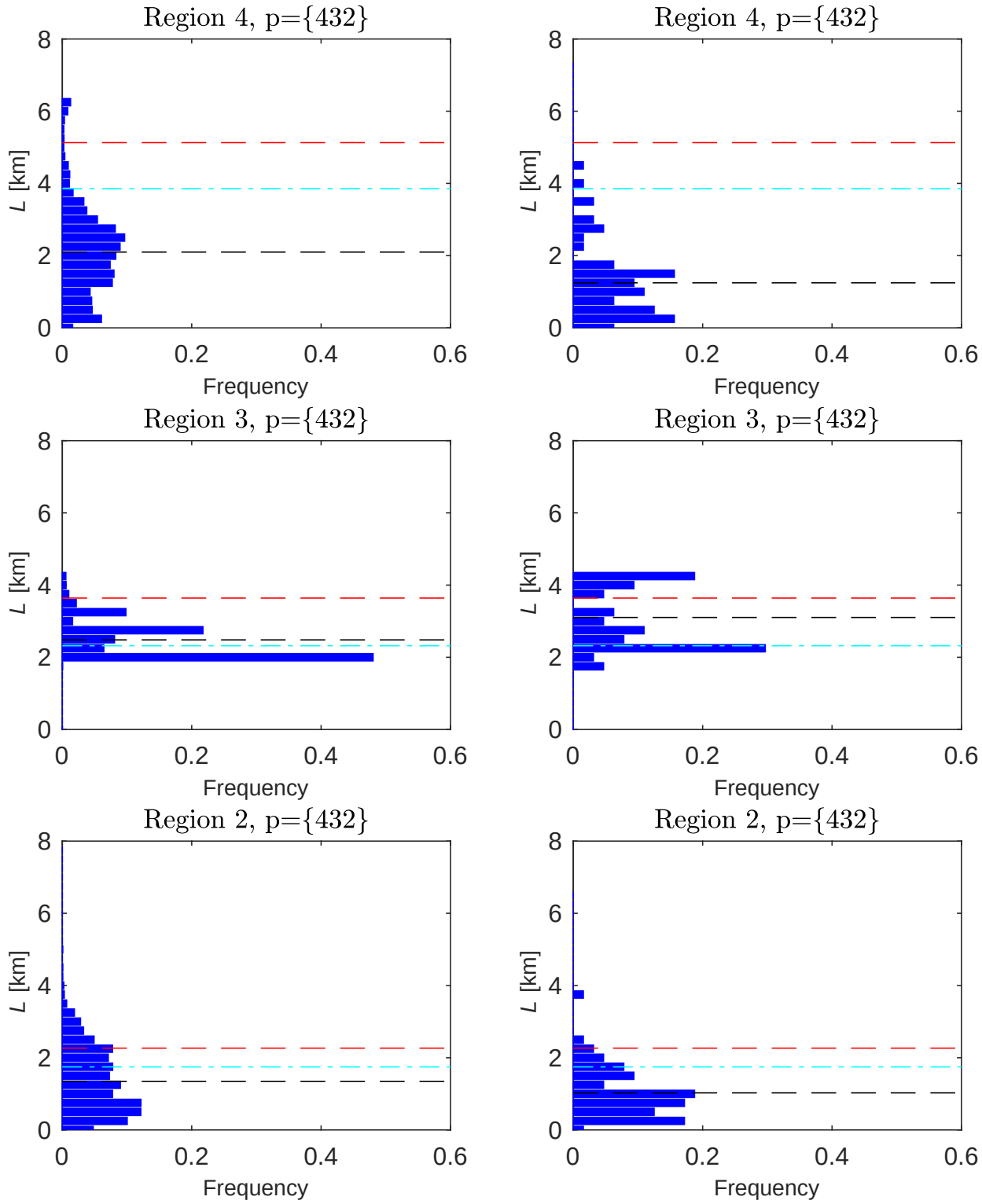


Figure 9: Trip length distributions for the regional path $p = \{4, 3, 2\}$, determined from the simulated data (left panel) as well as from the real data (right panel). The horizontal black dashed lines represent the average trip length of each distribution. The horizontal red and cyan dashed lines represent the average travel distances determined for the variants V2.3 and V2.4.

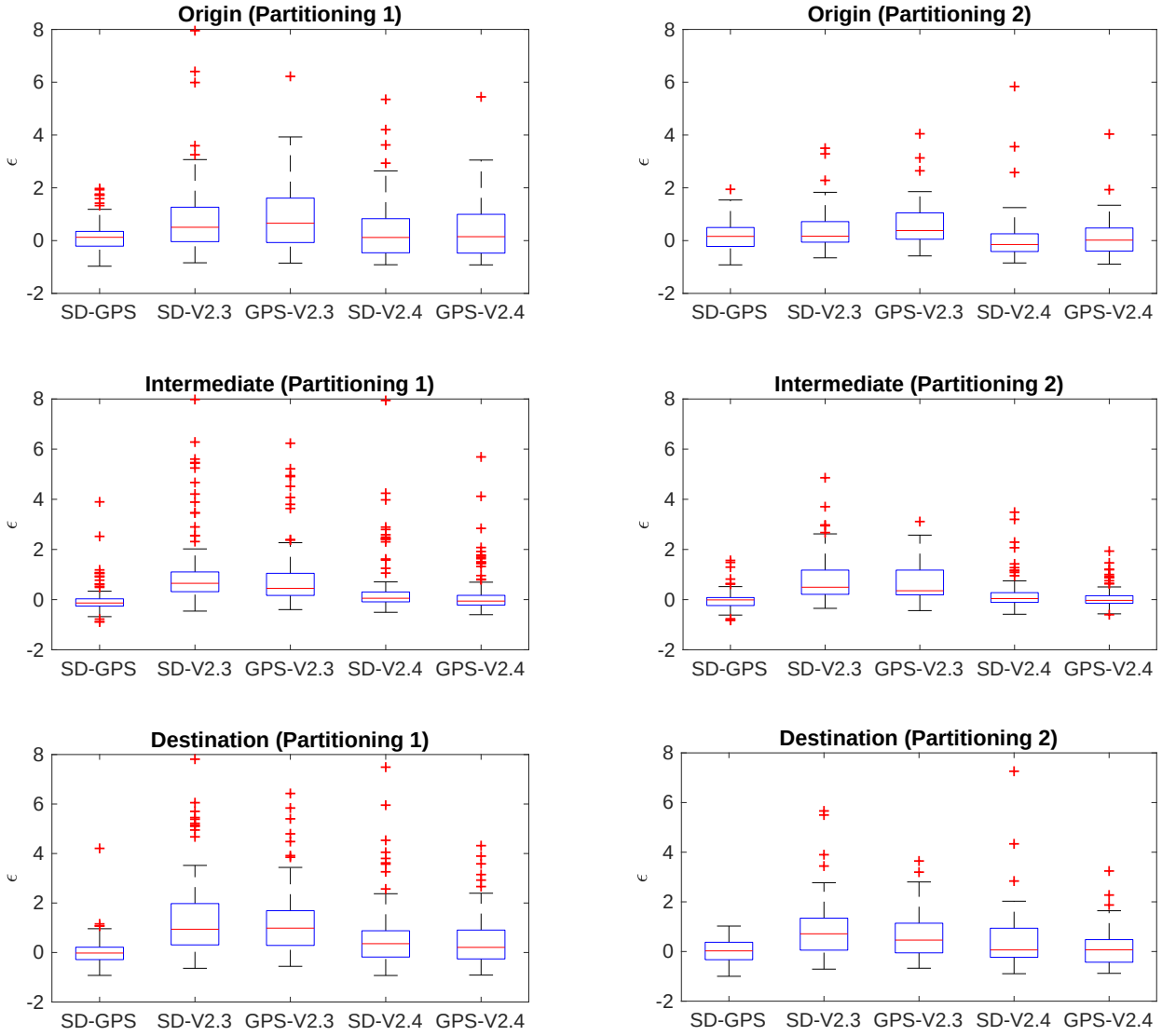


Figure 10: Box-and-whisker diagrams of the relative differences ϵ for the Origin, Intermediate and Destination regions of the common regional paths in the sets Φ_1 (SD), Φ_3 (GPS), $\phi_{V2.3}$ (V2.3) and $\phi_{V2.4}$ (V2.4).*Left:* Partitioning 1. *Right:* Partitioning 2.

the set of virtual trips and the real ones actually chosen by drivers, which tend to be longer than the shortest-trip in distance between their origin-destination pair. In particular, in the data-set of the real trajectories, drivers show a clear preference to take the ring road that bypasses the city of Lyon, therefore traveling longer distances. Recently, Yang et al. (2018) has proved the existence of a universal law for the trip detour ratios, i.e. a measure of how much a driver detours its trip concerning the shortest-route. It is possible to utilize this tool to calibrate the real travel distance of drivers, based on the information on the set of real trips, and therefore correct for this bias observed in Fig. 10. Therefore, if the information about a real set of trip patterns in the network is not available, we can utilize the set of virtual trips to: (i) identify the most prevailing paths chosen by drivers, as discussed in Section 3, with a large level of similarity; and (ii) have a good estimation of the traveled distances through the universal law proxy discussed by Yang et al. (2018).

The Euclidean distance used in V2.4 represents a good proxy for estimating mean trip lengths for the Intermediate regions. This is true for both Partitioning 1 and 2. For the SD-V2.4 case, the box-and-whisker diagram is narrow and

the median is close to 0%, despite the presence of several outliers. This means Euclidean distance proxy provides a good approximation to the average trip lengths estimated from the simulated data, i.e. $\overline{L}_{rp}^{\phi_{V2.4}} \approx \overline{L}_{rp}^{\Phi_1}$. In the GPS-V2.4 case, the interquartile range of the box-and-whisker diagram is larger than in the previous SD-V2.4 case, despite the median being also close to 0%. This means that the Euclidean distance is a better proxy for estimating trip lengths closer to the ones determined based on the simulated data. On the other hand, the box-and-whisker diagrams for the SD-V2.3 and GPS-V2.3 cases, show larger median values and interquartile ranges than the ones of the V2.4 case. This shows that the Euclidean distance is also a better proxy than the average shortest-trip used in V2.3, for estimating trip lengths closer to the ones determined from the simulated and real GPS data.

For the Origin and Destination regions, the medians of the box-and-whisker diagrams are in general larger than 0%. This is observed for both Partitioning 1 and 2. The interquartile ranges are also larger than in the case of the Intermediate regions. However, the interquartile ranges are, in general, slightly lower for Partitioning 2 than 1. In the cases of the Origin and Destination regions, the relative differences ϵ increase since the calculation of the trip lengths $\overline{L}_{rp}$ for V2.3 and V2.4 is insensitive to the spatial distribution of the origin-destination nodes inside the Origin and Destination regions, respectively. We recall the reader that in the case of V2.3, we determine the trip lengths based on the calculation of shortest-trips in distance between the centroid node of the Origin and Destination regions and their border nodes, and between the border nodes of Intermediate regions of paths. While, in the case of V2.4, we determine the travel distances based on the Euclidean distance between the centroid node of the Origin and Destination regions and the centroid node of their borders, and between the centroid nodes of the borders of Intermediate regions of paths. For Partitioning 1, the interquartile ranges of the box-and-whisker diagrams of V2.4 are in general smaller than the ones of V2.3. These differences are smaller for Partitioning 2. The medians of these diagrams for V2.3 are also closer to 0% than the ones of V2.3.

Overall, the results of this section partially answer the seventh and eight conjectures listed in the Introduction. First, we show that the Euclidean distance is a better proxy for estimating trip lengths in the regions, that are closer to the ones determined from the simulated and real GPS trajectories data for the regional paths. Second, the average travel distances determined from the real data are slightly superior than the ones determined from the simulated data.

4.3. Analysis of the total travel distances of paths

In this section, we investigate the relative differences of average travel distances of paths. As in the previous section, we focus on both Partitioning 1 and 2. First, we focus on the common regional paths determined from the simulated data (i.e. listed on Φ_1) and the real GPS trajectories data (i.e. listed on Φ_3). Let $\overline{L}_p$ be the average travel distance of a generic regional path p, and is determined as:

$$\overline{L}_p = \sum_{r \in p} \overline{L}_{rp} \delta_{rp}, \forall p \in (\Phi_i \cup \phi_{V2.j}) \wedge \forall i = 1, 3 \wedge \forall j = 3, 4 \tag{12}$$

The average travel distance in a generic region r that defines a generic path p is determined based on the set of virtual or real trips, or using any of the edge cost functions defined in Equation 4 to Equation 5. In this analysis, we discard the internal paths as Method 2 does not assign an edge cost. We recall that, in this method, the internal paths are assigned by definition when the Origin and Destination regions are the same. Furthermore, we note that the variant V2.2 only allows to determine a travel distance for the entire regional path, since we determine the Euclidean distance between the centroid nodes of the regions. For this reason, we also include this variant in our analysis in this section. In this case, we determine the average travel distance $\overline{L}_p$ as:

$$\overline{L}_p = \sum_{r=1}^{|p|-1} C_{p_r, p_{r+1}}, \forall p \in \phi_{V2.2} \tag{13}$$

where $|.|$ represents the total number of regions defining regional path p; and $C_{p_r, p_{r+1}}$ represents the Euclidean distance between the centroid nodes of regions p_r and p_{r+1} that define the regional path p (see Equation 4).

Similarly to the previous section, we also determine the relative differences ϵ as:

$$\epsilon = \frac{\overline{L}_p^{\Phi_1} - \overline{L}_p^{\Phi_3}}{\overline{L}_p^{\Phi_3}} \times 100\%, \forall p \in (\Phi_1 \cap \Phi_3) \tag{14}$$

where $\overline{L}_p^{\Phi_1}$ and $\overline{L}_p^{\Phi_3}$ are the average travel distances of path p, determined from the simulated and real GPS trajectories data, respectively.

We also investigate which proxy used to calibrate Method 2 is the best for determining average travel distances $\overline{L}_p$, that are close to the ones derived simulated data and real GPS trajectories data, for the regional paths. We then also determine the relative differences ϵ as:

$$\epsilon = \frac{\overline{L}_p^{V2.j} - \overline{L}_p^{\Phi_i}}{\overline{L}_p^{\Phi_i}} \times 100\%, \forall r \in p \wedge \forall p \in (\Phi_1 \cap \Phi_3) \wedge \forall i = 1, 3 \wedge \forall j = 3, 4 \tag{15}$$

where $\overline{L}_p^{V2.j}$ is the travel distance determined based on Equation 3 to Equation 5, for each of the j-*th* variant of Method 2 considered; and $\overline{L}_p^{\Phi_i}$ is the average travel distance determined for regional path p based on either the simulated or real GPS trajectories data.

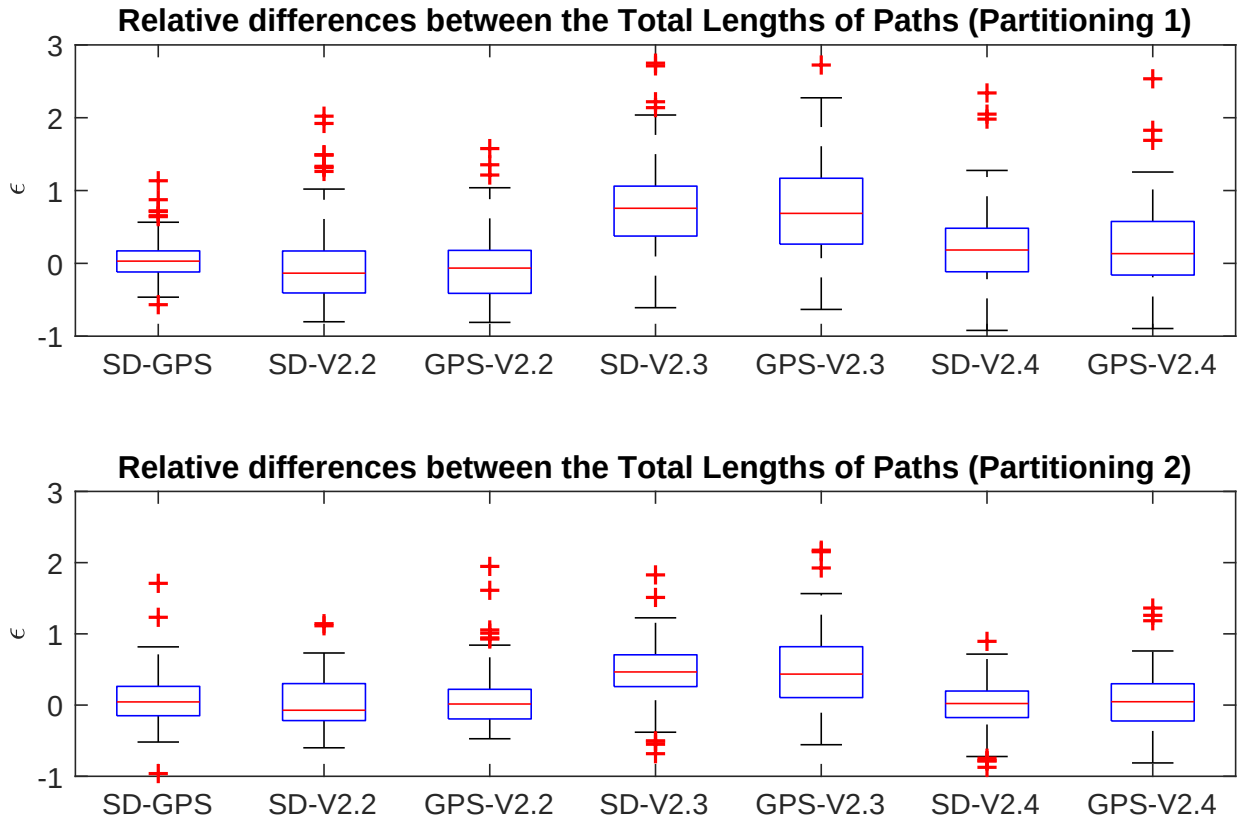


Figure 11: Box-and-whisker diagrams of the relative differences ϵ for the common regional paths in the sets Φ_1 (SD), Φ_3 (GPS), $\phi_{V2.2}$ (V2.2), $\phi_{V2.3}$ (V2.3) and $\phi_{V2.4}$ (V2.4). *Top:* Partitioning 1. *Bottom:* Partitioning 2.

Fig. 11 depicts the box-and-whisker diagrams of the relative differences ϵ for common paths found in the different combination of sets: Φ_1 and Φ_3 (case SD-GPS); Φ_1 and $\phi_{V2.2}$ (case SD-V2.2); Φ_3 and $\phi_{V2.2}$ (case GPS-V2.2); Φ_1 and $\phi_{V2.3}$ (case SD-V2.3); Φ_3 and $\phi_{V2.3}$ (case GPS-V2.3); Φ_1 and $\phi_{V2.4}$ (case SD-V2.4); and Φ_3 and $\phi_{V2.4}$ (case GPS-V2.4). The red dots represent the outliers of the distributions. The top panel depict the results for Partitioning 1, while the bottom panel shows the results for Partitioning 2. One can observe that the average travel distances $\overline{L}_p$ determined from the simulated data, slightly underestimates the real average distances traveled by drivers on the paths. This happens because in the dataset of trajectories, there is a clear preference of drivers to take the ring road around the city of Lyon, and therefore travel longer distances than the shortest one between their origin-destination points in the city networks. This is then reflected on the actual distances traveled on the regional paths. On the other hand, one can also observe that the Euclidean distance used to calibrate the variant V2.4, is the best proxy for determining average travel distances for the regional paths with respect to the ones determined from both the simulated and real GPS trajectories data. The medians are both close to 0%, and the interquartile range of the box-and-whisked diagrams are in general narrower than in the case of the other variants.

The results of this section complete the answers to the the seventh and eight conjectures listed in the Introduction. First, the average travel distances of paths determined from the real data is slightly superior than the ones determined from the simulated data. Second, the Euclidean distance proxy used to calibrate variant V2.4 provides a closer estimation of the average travel distance of paths, than the other two variants.

4.4. Significance level of paths and their travel distance

In this section, we investigate the relationship between the level of significance of a regional path and its average travel distance. The question is do the most prevailing regional paths, connecting one OD pair, have the lowest average travel distances. We recall the reader that the significance level of a regional path is related to the number of trips, in the city network, it has associated. The most prevailing regional path connecting one regional OD pair, is the one that has the largest number of trips associated. In this section we focus on Partitioning 1. The test network depicted in Fig. 4 (a) has 10 regions, which yields a total of 100 possible regional OD pairs. For each regional OD pair, we rank the existent regional paths according to their level of significance. We note that for each regional OD pair, we have only one most prevailing regional path. This is also true for the second and third most prevailing paths. We determine the fractions of the first, second and third most prevailing paths that have the first, second and third lowest average travel distances. We also determine the same fraction for the less prevailing regional paths. This analysis is done for both the simulated and real data. Fig. 12 depicts the bar plot that summarizes the fraction of the $i-$th most prevailing (MS) regional paths, and their average travel distance (ATD). The colored bars depict the fraction of paths that are labeled as the $i-$th most prevailing ones with respect to their average travel distance. As an example, the blue percentage of the first colored bar in Fig. 12 (a) shows that $\sim 40\%$ of the paths with the lowest travel distances are also labeled as the most prevailing ones. This means that out of 100 most prevailing regional paths in our test scenario, only ~ 40 % have the lowest average travel distance.

We first analyze the results obtained from the simulated data. We observe that $\sim 40\%$ of the paths with the lowest average travel distances, are also the most prevailing ones. The surprising fact is that also $\sim 40\%$ of the most prevailing paths have the longest average travel distances. While, only $\sim 15\%$ and $\sim 5\%$ of the most prevailing regional paths have the second and third lowest average travel distances, connecting the OD pairs. In general, the less prevailing regional paths regional paths also have the longer average travel distances. The results gathered from the real data also show similar trends, where only $\sim 25\%$ of the most prevailing regional paths are also the ones with the lower average travel distance. We also observe the similar trend that paths with longer average travel distances are also less prevailing. These results answer our ninth conjecture listed in the Introduction. Overall, the most prevailing regional paths, i.e. the ones that have the largest number of trips associated, are not necessarily the ones have the lowest average travel distances.

5. Conclusions

In this paper, we focused on the identification and characterization of the most prevailing paths on regional networks. We distinguish between internal paths, i.e. a path that only travels inside one region, and regional paths, i.e. a path that crosses an ordered sequence of regions. We propose two dedicated methods to determine paths on regional

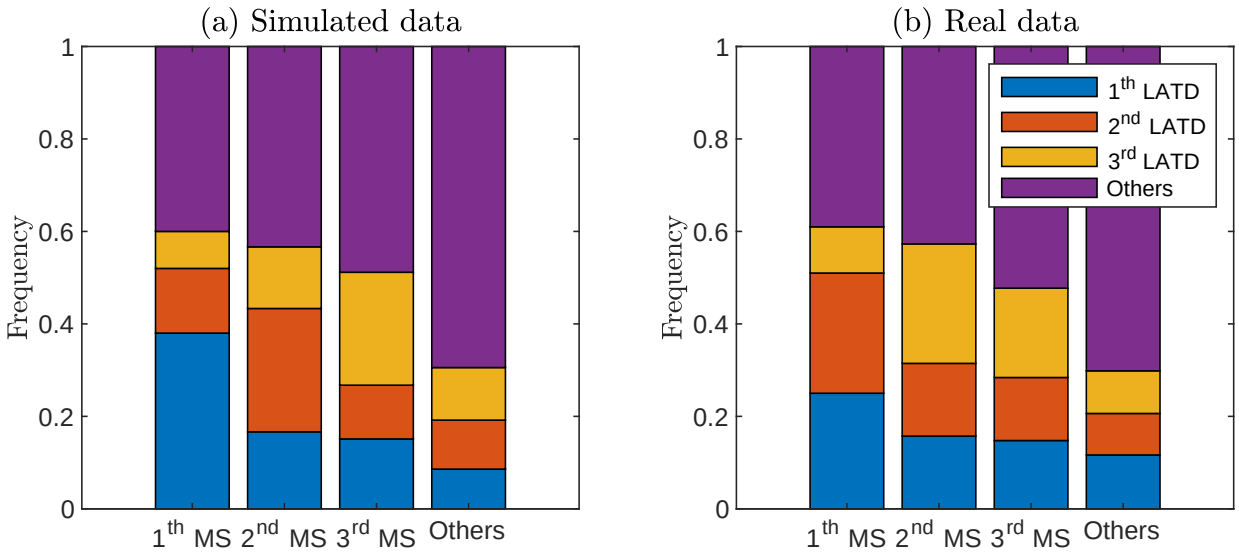


Figure 12: Bar graph of the fraction that relate the level of significance of regional paths with their average travel distance. The results are depicted for the (a) simulated data and (b) real data. The first three bars represent the three most prevailing regional paths, while the fourth bar represents the other less prevailing regional paths. The colors of the bars are related with the average travel distance of the paths. The color bars refer to the paths with the lowest (blue), second lowest (orange), third lowest (yellow) and longer (purple) travel distances. The acronym LATD stands for lowest average travel distance.

networks. The first method requires the complex calibration of a set of virtual trips, that are then scaled-up according to the city network partitioning. We discuss two variants of this method, where the sampling of the origin-destination pairs of nodes can be performed independently or not of the definition of the city network partitioning. The second method is more parsimonious and computationally lighter. It consists in directly determining the shortest-cost paths on the regional network. The challenge lies on the calibration of the edge costs of the regional network. We propose four variants of impedance cost functions, where we consider the following proxies: (i) the border flow capacity; (ii) the Euclidian distance between the centroids of the regions; and (iii) the average shortest distance or (iv) the Euclidian distance between the centroids of the regions and their borders.

We test the implementation of these two methods for determining the most prevailing paths on a city network that consists of the metropolitan area of Lyon (France). We have considered two different definitions of the partitioning. The test results answered our nine conjectures listed in the Introduction.

1. We show that sampling trips in the city network concerning the definition of its partitioning is a better approach to construct a set of virtual trips.
2. The second method is able to determine the most prevailing paths from both the simulated and real data, with a large level of similarity, being the variant V2.2 (i.e. Euclidean distance between the centroids of the regions) the best one to calibrate this method to determine paths.
3. The set of virtual trips in distance is a good proxy to determine the most prevailing paths chosen by drivers.
4. The definition of the partitioning does significantly influence the performance of Methods 1 and 2 to determine the real paths chosen by drivers, as evidenced by the large similarity levels of the sets.
5. The functional form of the network-wide TLD is better reproduced by a Weibull distribution. This is true for both the simulated and real data. The scale λ_{We} and shape k_{We} parameters of the Weibull distribution fitted for the trip length distribution from the simulated data are 4443 [m] and 1.2 [m], respectively. While for the trip length distribution from the real data are $\lambda_{We} = 6174$ [m] and shape $k_{We} = 2.3$ [m]. Note that, these distributions are unimodal.
6. We show that the complex topological features of the city network strongly influence the shape of the TLDs of the regions defining a path. This strengths the fact the one should not make any prior assumptions regarding the

functional form of these TLD for MFD-based applications.

7. The Euclidean distance proxy utilized to calibrated the variant 2.4 is a better proxy to determine closer average travel distances as the ones determined from both the real and simulated data.
8. The average travel distances determined from the real data are slightly superior than the ones determined from the simulated data. This is true at both the region and path levels.
9. We also show that the most prevailing paths are not necessarily the ones that have the lowest average travel distance.

We stress that these results might be specific to the test network. The performance of both methods for identifying the most prevailing paths might change for a different network. However, the performance of both methods is not significantly influenced by the two studied definitions of the partitioning. Note that, we assume that the definition of the partitioning yields well-defined, fully connected and compact regions. The testing of the performance of these two methods to determine the most prevailing paths on other city networks is needed.

6. Discussion

In this paper, we discuss a methodological framework to determine the most prevailing paths chosen by drivers. This methodology is valid for any city network, given the definition of its partitioning. Method 2 is computationally lighter than Method 1, and is able to identify the most prevailing paths with a large level of similarity. It presents a promising tool to determine paths for applications of accumulation-based MFD models. The proxies utilized to calibrate Method 2, also provide good estimation of the average travel distances in the regions as well as of paths. Method 1 requires the knowledge of a set of trips in the city network. One possibility is to utilize a set of real GPS trajectories. However, it requires a complex task of map-matching the trajectories in the network, filtering invalid trips (e.g. with unknown or uncertain origins and/or destinations, or missing information about the sequence of traveled links which happens when two links are very close). In this paper, we discuss one alternative by constructing a set of shortest-trips in distance. This requires a complex calibration, which is unrealistic for large city networks. Moreover, the sampling of trips is completely random or naive. This means that all origin-destination pairs in the city network are equally probable of being selected. In reality, the selection of the origin-destination pairs should rely on how the demand is distributed over the network, i.e. some pairs should be more likely of being selected. This limitation presents an interesting line of future research, where we plan to develop a more robust and intelligent sampling methodology of the origin-destination pairs of nodes in the city network. The first challenge is to identify a subset of representative od pairs od nodes from the set of all possible travels in the city network. The second challenge is to link this prediction to real trip patterns of drivers.

In this paper, we focused on the calculation of static paths, i.e. we do not consider how changes in the traffic dynamics in the regions influence the calculation of paths. For example, the inclusion of traffic lights and stop signs in the calculation of paths by both methods would influence their level of significance and therefore their ranking on the choice set. In the case of Method 1, this would clearly influence the calculation of the set of virtual trips. The inclusion of traffic lights and stop signs lead to the emergence of congestion patterns and spill-back effects, which lead drivers to detour their trips. This means that the calculation of the set of virtual trips would be based on the calculation of shortest-trips in time instead of distance. This would probably lead to the emergence of new prevailing paths in the choice set. The ranking of paths determined by both variants V1.1 and V1.2 would be different. In the case of Method 2, this would not influence the proxy based on the Euclidean distance (which corresponds to the variants V2.2 and V2.4). However, the flow capacity of each lane q^c_{af} changes, and the paths determined by variant V2.1 might change concerning the evolution of the traffic dynamics in the regions. The same is true for the calculation of the shortest average travel distances between the centroids and the borders as well as between borders. This proxy is utilized to calibrated variant V2.3. The time-dependence of paths in regional networks presents another interesting line for future research.

One of the most important highlights of this paper regards the functional form of the TLD. We showed that a Weibull distribution is the one that better produces the network-wide TLD determined from both the simulated and real data. However, the characterization of the functional form of the TLD of the regions that define each path is much more complex. In fact, the functional form of such TLD depend on the complex topological features of the city network concerning the definition of its partitioning. Gaussian Mixture Models present a promising tool to characterize these trip length distributions. This is especially true to account for dynamic effects such as congestion,

which leads vehicles to detour their trips (Yang et al., 2018) and therefore influencing the functional form of the trip length distributions. This presents another interesting line for future research.

Acknowledgements

The authors thank the anonymous reviewers for their critical assessment of our paper as well as for their comments and suggestions that have much improved this paper.

S. F. A. Batista and L. Leclercq acknowledge funding by the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation program (grant agreement No 646592 - MAGnUM project).

Authors contributions

S. F. A. Batista contributed to the conceptualization, methodology, validation and results analysis, and writing of the original draft of the paper. L. Leclercq contributed to the conceptualization, methodology and review & editing of the paper. Manon Seppecher did the data curation of the GPS trajectories, and participated in the review & editing of the paper. All the authors have approved the final version of this paper submitted to publication.

Appendices

A. Table of notations

Table A.1 summarizes the notation used in this paper.

Table A.1: Nomenclature used in this paper.

City network:	
o	Origin node.
d	Destination node.
Ω^{od}	Route choice set.
Ξ	Set of all origin-destination pairs of nodes of the city network.
A	Set of links defining the city network.
a	Generic link of the city network.
l_a	Length of link a.
Z	Set of nodes defining the city network.
N_{od}	Total number of origin-destination pairs sampled in the city network.
Θ	Set of virtual trips.
ρ_{ij}	Set of the city network nodes that are located at the partition (or border) between two generic adjacent regions i and j.
q^c_{af}	Flow capacity of each lane f of the incoming link a to each border node.
N_{lanes}	Total number of lanes of each incoming link a to each border node listed in ρ_{ij}.
δ_{aij}	Binary variable that equals 1 if link a allows to travel from region i to region j.
$(x_i, y_i))$	Cartesian coordinates of the centroid node of a generic region i.
ζ	Set of all real trips traveled by drivers, determined from GPS trajectories.
$N^{links}_{used}(N_{od})$	Total number of links traveled by the trips listed in Θ.
N_{links}	Total number of links defining the city network.
$N^{links}_{cov}(N_{od})$	Percentage of the city network links that are traveled by virtual trips in the set Θ.
Regional network:	
O	Origin region.
D	Destination region.
E	Edges of the regional graph.
X	Set of regions that define the regional network (or set of nodes of the regional graph).
p	Regional path.
r	Region.
L_{rp}	Trip length distribution of a generic region that defines a generic path p.
$\overline{L}_{rp}$	Average travel distance in a generic region r that defines a generic regional path p.
$\overline{L}_p$	Average travel distance of a generic regional path p.
Ω^{OD}	Regional choice set.
W	Set of regional OD pairs.
δ_{rp}	Binary variable that equals 1 if region r defines path p.
C_{ij}	Edge cost of the regional network graph.
Λ	Set of adjacent regions to a generic region i.
B_{ij}	Border between two generic adjacent regions i and j.
$N_{borders}$	Total number of borders between adjacent regions of the network.
R	Total number of regions in X.
l_k	Trip length or travel distance of a virtual trip k.

Continued on next page

Table A.1 – *Continued from previous page*

$L_{i,\rho_{ij}}$	Set of trip lengths between the centroid node of region i and all border nodes listed in ρ_{ij}.
$L_{\rho_{ij},j}$	Set of trip lengths between all border nodes listed in ρ_{ij} and the centroid node of region j.
$L_{\rho_{im},\rho_{mj}}$	Set of trip lengths connecting all border nodes between region m and adjacent region i, to all border nodes between region m and adjacent region j.
n_{ij}	Centroid node of the the border B_{ij} between two generic regions i and j.
$\overline{L}_{i,n_{ij}}$	Euclidian distance between the centroid node of generic region i and the centroid node n_{ij} of the border B_{ij}.
$\overline{L}_{n_{ij},j}$	Euclidean distance between n_{ij} and the centroid node of generic region j.
$\overline{L}_{n_{im},n_{mj}}$	Centroid nodes n_{im} and n_{mj} of the borders B_{im} and B_{jm}, respectively, when region m is crossed.
Φ_1	Set of all regional paths gathered from the simulated data.
ϕ_1	Set containing the three most prevailing regional paths for each regional $(O, D) \in W$, determined from the simulation data.
Φ_3	Set of all regional paths gathered from the real trajectories of drivers.
ϕ_3	Set containing the three most prevailing regional paths for each regional OD pair, determined from the real GPS trajectories.
$\phi_{V2.j}$, $j = 1, \ldots, 4$	Full set of calculated paths for the $j-$th variant of Method 2.
$\alpha^{\Phi_i \phi_{V2.j}}$	Similarity criterion between the sets Φ_i and $\phi_{V2.j}$.
$\beta^{\Phi_i \phi_{V2.j}}$	Strict similarity criterion between the sets Φ_i and $\phi_{V2.j}$.
δ_p^{OD}	Binary variable that equals 1 if regional path p that connects the regional OD pair is ranked with a similar level of significance in Φ_i, $i = 1, 3$, and ϕ_2, or 0 otherwise.
ϵ	Relative differences between trip lengths determined for the same generic regional path, from different sets.
Parameters of the statistical distributions:	
l_i	Sample/observation of the trip length distribution L_{rp}.
I	Length of the set L_{rp}.
Γ	Gamma function.
k_{Ga}	Shape parameter of the Gamma distribution.
θ	Scale parameter of the Gamma distribution.
λ_{NE}	Rate or inverse parameter scale of the negative exponential distribution.
μ	Mean of the log-normal distribution.
σ^2	Standard deviation of the log-normal distribution.
λ_{We}	Shape parameter of the Weibull distribution.
k_{We}	Scale parameter of the Weibull distribution.

B. Table of paths determined for Method 1

O	D	Method 1: V1.1				
		N_{od}				
		1.10^2	5.10^2	1.10^3	5.10^3	1.10^4
4	2	4,9,2	4,3,2	4,3,2	4,9,3,2	4,3,2
		~	~	4,9,3,2	4,3,2	4,9,3,2
		~	~	~	4,9,2	~
2	4	2,3,4	2,3,4	2,3,4	2,3,4	2,3,4
		~	~	~	~	~
		~	~	~	~	~
7	5	~	7,5	7,5	7,5	7,5
		~	7,6,5	7,6,5	7,6,5	7,6,5
		~	~	~	~	~
5	7	5,6,7	5,7	~	5,6,7	5,7
		~	5,6,7	~	~	5,4,9,10,7
		~	~	~	~	~
2	5	~	~	2,9,10,7,6,5	2,9,10,5	~
		~	~	~	2,9,10,7,6,5	~
		~	~	~	2,9,10,7,5	~
5	2	~	5,10,9,2	~	5,6,7,10,9,2	5,6,7,10,9,2
		~	5,7,10,9,2	~	5,6,7,10,9,1,2	5,6,7,10,9,1,2
		~	5,6,7,10,9,2	~	5,10,9,2	~
3	7	3,4,9,10,7	3,9,10,7	~	3,9,10,7	3,9,10,7
		~	3,4,9,10,7	~	~	3,2,9,8,7
		~	3,2,9,8,7	~	~	~
7	3	~	7,10,9,4,3	~	7,10,9,4,3	7,10,9,3
		~	7,10,9,3	~	7,8,10,9,3	7,10,9,2,3
		~	7,10,9,2,3	~	7,10,9,3	7,10,9,4,3

O	D	N_{od}				
		5.10^4	1.10^5	5.10^5	1.10^6	$1{,}085.10^9$
4	2	4,3,2	4,3,2	4,3,2	4,3,2	4,3,2
		4,9,2	4,9,2	4,9,3,2	4,9,3,2	4,9,3,2
		4,9,3,2	4,9,3,2	4,9,2	4,9,2	4,9,2
2	4	2,3,4	2,3,4	2,3,4	2,3,4	2,3,4
		~	~	~	2,9,4	2,9,4
		~	~	~	~	2,1,9,4
7	5	7,5	7,5	7,5	7,5	7,5
		7,6,5	7,6,5	7,6,5	7,6,5	7,6,5
		7,8,10,5	7,10,5	7,10,5	7,10,5	7,10,5
5	7	5,7	~	5,6,7	5,6,7	5,6,7
		5,6,7	~	5,7	5,7	5,7
		~	~	5,10,9,1,7	5,10,9,1,7	5,10,9,1,7
2	5	2,9,10,7,6,5	~	2,9,10,5	2,9,10,7,6,5	2,9,10,7,6,5
		~	~	2,9,10,7,6,5	2,9,10,5	2,9,10,5
		~	~	2,3,4,5	2,3,4,5	2,3,4,5
5	2	5,6,7,10,9,2	5,6,7,10,9,2	5,6,7,10,9,2	5,10,9,2	5,10,9,2
		5,10,9,1,2	~	5,10,9,2	5,6,7,10,9,2	5,6,7,10,9,2
		~	~	5,6,7,10,9,1,2	5,10,9,1,2	5,10,9,1,2
3	7	3,9,10,7	~	3,9,10,7	3,9,10,7	3,9,10,7
		3,4,9,10,7	~	3,4,9,10,7	3,4,9,10,7	3,4,9,10,7
		3,2,9,8,7	~	3,2,9,8,7	3,2,9,8,7	3,2,9,8,7
7	3	7,8,10,9,3	7,10,9,3	7,10,9,3	7,8,10,9,2,3	7,8,10,9,2,3
		7,10,9,3	7,10,9,4,3	7,8,10,9,3	7,10,9,3	7,10,9,3
		7,10,9,4,3	7,10,9,2,3	7,10,9,4,3	7,10,9,4,3	7,10,9,4,3

Table B.1: Evolution of the regional choice set for the OD pairs 42, 24, 75, 57, 25, 52, 37 and 73, as function of the network coverage (N_{od}). The results are listed for the two variants of Method 1. The regional paths are sorted by their level of significance.

O	D	Method 1: V1.2				
		N_{od}				
		1.10^2	5.10^2	1.10^3	5.10^3	1.10^4
4	2	4,9,3,2 ~ ~	4,3,2 4,9,2 4,9,3,2	4,3,2 4,9,2 4,9,3,2	4,3,2 4,9,3,2 4,9,2	4,3,2 4,9,3,2 4,9,2
2	4	2,3,4 ~ ~	2,3,4 ~ ~	2,3,4 ~ ~	2,3,4 2,1,9,4 ~	2,3,4 ~ ~
7	5	7,6,5 ~ ~	7,5 7,6,5 7,10,9,4,5	7,5 7,6,5 ~	7,5 7,6,5 7,8,10,5	7,5 7,6,5 7,10,5
5	7	5,7 ~ ~	5,6,7 5,7 ~	5,7 5,6,7 ~	5,6,7 5,7 ~	5,6,7 5,7 ~
2	5	~ ~ ~	2,9,10,5 2,9,10,7,6,5 2,9,10,7,5	2,9,10,5 2,9,10,7,6,5 2,9,10,7,5	2,9,10,7,6,5 2,9,10,5 2,3,4,5	2,9,10,7,6,5 2,9,10,5 2,3,4,10,5
5	2	5,10,9,2 ~ ~	5,6,7,10,9,1,2 5,6,7,10,9,2 ~	5,6,7,10,9,2 5,10,9,2 5,10,9,1,2	5,10,9,2 5,6,7,10,9,2 5,10,9,1,2	5,10,9,2 5,6,7,10,9,2 5,6,7,10,9,1,2
3	7	3,2,9,8,7 ~ ~	3,9,10,7 3,4,9,10,7 3,2,9,8,7	3,9,10,7 3,4,9,10,7 ~	3,9,10,7 3,4,9,10,7 3,2,9,8,7	3,9,10,7 3,4,9,10,7 3,2,9,8,7
7	3	7,10,9,4,3 ~ ~	7,10,9,3 7,10,9,4,3 7,8,10,9,3	7,10,9,4,3 7,8,10,9,2,3 7,8,10,9,3	7,10,9,3 7,8,10,9,2,3 7,8,10,9,3	7,8,10,9,2,3 7,10,9,4,3 7,8,10,9,3

O	D	N_{od}				
		5.10^4	1.10^5	5.10^5	1.10^6	$1{,}085.10^9$
4	2	4,3,2 4,9,3,2 4,9,2	4,3,2 4,9,3,2 4,9,2	4,3,2 4,9,3,2 4,9,2	4,3,2 4,9,3,2 4,9,2	4,3,2 4,9,3,2 4,9,2
2	4	2,3,4 2,9,4 ~	2,3,4 2,9,4 ~	2,3,4 2,9,4 2,9,1,4	2,3,4 2,9,4 2,9,1,4	2,3,4 2,9,4 2,9,1,4
7	5	7,5 7,6,5 7,10,5	7,5 7,6,5 7,10,5	7,5 7,6,5 7,10,5	7,5 7,6,5 7,10,5	7,5 7,6,5 7,10,5
5	7	5,6,7 5,7 5,10,9,1,7	5,6,7 5,7 5,10,9,1,7	5,6,7 5,7 5,4,9,10,7	5,6,7 5,7 5,4,9,10,7	5,6,7 5,7 5,10,9,1,7
2	5	2,9,10,7,6,5 2,9,10,5 2,9,10,7,5	2,9,10,7,6,5 2,9,10,5 2,3,4,5	2,9,10,7,6,5 2,9,10,5 2,3,4,5	2,9,10,7,6,5 2,9,10,5 2,3,4,5	2,9,10,7,6,5 2,9,10,5 2,3,4,5
5	2	5,6,7,10,9,2 5,10,9,2 5,10,9,1,2	5,10,9,2 5,6,7,10,9,2 5,10,9,1,2	5,10,9,2 5,6,7,10,9,2 5,10,9,1,2	5,10,9,2 5,6,7,10,9,2 5,10,9,1,2	5,10,9,2 5,6,7,10,9,2 5,10,9,1,2
3	7	3,9,10,7 3,4,9,10,7 3,2,9,8,7	3,9,10,7 3,4,9,10,7 3,2,9,8,7	3,9,10,7 3,4,9,10,7 3,2,9,8,7	3,9,10,7 3,4,9,10,7 3,2,9,8,7	3,9,10,7 3,4,9,10,7 3,2,9,8,7
7	3	7,8,10,9,2,3 7,10,9,4,3 7,8,10,9,3	7,8,10,9,2,3 7,10,9,3 7,10,9,4,3	7,8,10,9,2,3 7,10,9,3 7,10,9,4,3	7,8,10,9,2,3 7,10,9,3 7,10,9,4,3	7,8,10,9,2,3 7,10,9,3 7,10,9,4,3

Table B.2: Same as in Table B.1, but for V1.2 of Method 1.

References

Aboudolas, K., Geroliminis, N., 2013. Perimeter and boundary flow control in multi-reservoir heterogeneous networks. Transportation Research Part B: Methodological 55, 265–281.
URL https://dx.doi.org/10.1016/j.trb.2013.07.003

Ambühl, L., Loder, A., Zheng, N., Axhausen, K. W., Menendez, M., 2019. Approximative network partitioning for mfds from stationary sensor data. Transportation Research Record.
URL https://dx.doi.org/10.1177/0361198119843264

Amirgholy, M., Shahabi, M., Gao, H. O., 2017. Optimal design of sustainable transit systems in congested urban networks: A macroscopic approach. Transportation Research Part E: Logistics and Transportation Review 103, 261 – 285.
URL http://www.sciencedirect.com/science/article/pii/S1366554516309061

Arnott, R., 2013. A bathtub model of downtown traffic congestion. Journal of Urban Economics 76, 110–121.
URL https://dx.doi.org/10.1016/j.jue.2013.01.001

Azevedo, J., Santos Costa, M., Silvestre Madeira, J., Vieira Martins, E., 1993. An algorithm for the ranking of shortest paths. European Journal of Operational Research 69, 97–106.
URL https://dx.doi.org/10.1016/0377-2217(93)90095-5

Batista, S. F. A., Leclercq, L., 2019. Regional dynamic traffic assignment framework for mfd multi-regions models. Transportation Science 53, 1563–1590.
URL https://dx.doi.org/10.1287/trsc.2019.0921

Batista, S. F. A., Leclercq, L., Geroliminis, N., 2019. Estimation of regional trip length distributions for the calibration of the aggregated network traffic models. Transportation Research Part B: Methodological 122, 192–217.
URL https://dx.doi.org/10.1016/j.trb.2019.02.009

Ben-Akiva, M., Bergman, M. J., Daly, A., Ramaswamy, V., 1984. Modeling interurban route choice behaviour. In: Proceedings of the 9th International Symposium on Transportation and Traffic Theory. Utrecht, The Netherlands.

Beojone, C. V., Geroliminis, N., 2020. On the inefficiency of ride-sourcing services towards urban congestion.
URL https://arxiv.org/abs/2007.00980

Bierlaire, M., Frejinger, E., 2005. Route choice models with subpath components. In: Proceedings of the 5^{th} Swiss Transport Research Conference. Ascona, Switzerland.

Bliemer, M. C. J., Bovy, P. H. L., Li, H., 2007. Some properties and implications of stochastically generated route choice sets. In: Proceedings of the 6^{th} Tristan Conference. Pukhet, Thailand.

Cao, J., Menendez, M., 2015. System dynamics of urban traffic based on its parking-related-states. Transportation Research Part B: Methodological 81, 718 – 736, iSTTT 21 for the year 2015.
URL https://dx.doi.org/10.1016/j.trb.2015.07.018

Cao, J., Menendez, M., Waraich, R., 2019. Impacts of the urban parking system on cruising traffic and policy development: the case of zurich downtown area, switzerland. Transportation 46, 883908.
URL https://doi.org/10.1007/s11116-017-9832-9

Casadei, G., Bertrand, V., Gouin, B., Canudas-de-Wit, C., 2018. Aggregation and travel time calculation over large scale traffic networks: An empiric study on the grenoble city. Transportation Research Part C: Emerging Technologies 95, 713–730.
URL https://dx.doi.org/10.1016/j.trc.2018.07.033

Daganzo, C., 2007. Urban gridlock: Macroscopic modeling and mitigation approaches. Transportation Research Part B: Methodological 41, 49–62.
URL https://dx.doi.org/10.1016/j.trb.2006.03.001

de la Barra, T., Perez, B., Anez, J., 1993. Multidimensional path search and assignment. In: Proceedings of the 21^{st} PTRC Summer Annual Meeting. Manchester, England.

Ekbatani, M., Papageorgiou, M., Papamichail, I., 2013. Urban congestion gating control based on reduced operational network fundamental diagrams. Transportation Research Part C: Emerging Technologies 33, 74–87.
URL https://dx.doi.org/10.1016/j.trc.2013.04.010

Eppstein, D., 1998. Finding k shortest paths. Journal of the Society for Industrial and Applied Mathematics 28 (2), 652–673.
URL https://dx.doi.org/10.1137/S0097539795290477

Flötteröd, G., Bierlaire, 2013. Metropolis-hastings sampling of paths. Transportation Research Part B 48, 53–66.
URL https://dx.doi.org/10.1016/j.trb.2012.11.002

Geroliminis, N., Daganzo, C., 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. Transportation Research Part B: Methodological 42, 759–770.
URL https://dx.doi.org/10.1016/j.trb.2008.02.002

Geroliminis, N., Haddad, J., Ramezani, M., 2013. Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: a model predictive approach. IEEE Transactions on Intelligent Transportation Systems 14, 348–359.
URL https://dx.doi.org/10.1109/TITS.2012.2216877

Godfrey, J. W., 1969. The mechanism of a road network. Traffic Engineering and Control 11, 323–327.
URL https://trid.trb.org/view.aspx?id=117139

Gu, Z., Shafiei, S., Liu, Z., Saberi, M., 2018. Optimal distance- and time-dependent area-based pricing with the network fundamental diagram. Transportation Research Part C: Emerging Technologies 95, 1 – 28.
URL https://doi.org/10.1016/j.trc.2018.07.004

Haddad, J., 2017. Optimal perimeter control synthesis for two urban regions with aggregate boundary queue dynamics. Transportation Research Part B: Methodological 96, 1–25.
URL https://dx.doi.org/10.1016/j.trb.2016.10.016

Haddad, J., Mirkin, B., 2017. Coordinated distributed adaptive perimeter control for large-scale urban road networks. Transportation Research Part

C: Emerging Technologies 77, 495–515.
URL https://dx.doi.org/10.1016/j.trc.2016.12.002
Haddad, J., Zheng, Z., 2018. Adaptive perimeter control for multi-region accumulation-based models with state delays. Transportation Research Part B: Methodological, 1–21.
URL https://dx.doi.org/10.1016/j.trb.2018.05.019
Hadjiconstantinou, E., Christofides, N., 1999. An efficient implementation of an algorithm for finding k-shortest paths. Networks 34, 88–101.
URL https://dx.doi.org/10.1002/(SICI)1097-0037(199909)34:2<88::AID-NET2>3.0.CO;2-1
Haitao, H., Yang, K., Liang, H., Menendez, M., Guler, S. I., 2019. Providing public transport priority in the perimeter of urban networks: A bimodal strategy. Transportation Research Part C: Emerging Technologies 107, 171 – 192.
URL https://dx.doi.org/10.1016/j.trc.2019.08.004
Herman, R., Prigogine, I., 1979. A two-fluid approach to town traffic. Science 204, 148–151.
URL https://dx.doi.org/10.1126/science.204.4389.148
Ingole, D., Mariotte, G., Leclercq, L., 2020a. Minimizing network-wide emissions by optimal routing through inner-city gating. Transportation Research Part D: Transport and Environment 86, 102411.
URL https://dx.doi.org/10.1016/j.trd.2020.102411
Ingole, D., Mariotte, G., Leclercq, L., 2020b. Perimeter gating control and citywide dynamic user equilibrium: A macroscopic modeling framework. Transportation Research Part C: Emerging Technologies 111, 22 – 49.
URL https://dx.doi.org/10.1016/j.trc.2019.11.016
Jin, W.-L., 2020. Generalized bathtub model of network trip flows. Transportation Research Part B: Methodological 136, 138 – 157.
URL https://doi.org/10.1016/j.trb.2020.04.002
Katsis, P., Papageorgiou, T., Ntziachristos, L., 2014. Modelling the trip length distribution impact on the co2 emissions of electrified vehicles. Energy and Power 4 (1A), 57–64.
URL https://dx.doi.org/10.5923/s.ep.201401.05
Kouvelas, A., Saeedmanesh, M., Geroliminis, N., 2017. Enhancing model-based feedback perimeter control with data-driven online adaptive optimization. Transportation Research Part B: Methodological 96, 26–45.
URL https://dx.doi.org/10.1016/j.trb.2016.10.011
Lamotte, R., Geroliminis, N., 2016. The morning commute in urban areas: Insights from theory and simulation. In: Transportation Research Board 95^{th} Annual Meeting. pp. 16–2003.
Laval, J. A., Leclercq, L., Chiabaut, N., 2018. Minimal parameter formulations of the dynamic user equilibrium using macroscopic urban models: Freeway vs city streets revisited. Transportation Research Part B: Methodological 117, 676 – 686, tRB:ISTTT-22.
URL https://dx.doi.org/10.1016/j.trb.2017.08.027
Leclercq, L., Paipuri, M., 2020. Macroscopic traffic dynamics under fast-varying demand. Transportation Science 53, 1501–1799.
URL https://dx.doi.org/10.1287/trsc.2019.0908
Leclercq, L., Sénécat, A., Mariotte, G., 2017. Dynamic macroscopic simulation of on-street parking search: A trip-based approach. Transportation Research Part B: Methodological 101, 268–282.
URL https://dx.doi.org/10.1016/j.trb.2017.04.004
Loder, A., Ambühl, L., Menendez, M., Axhausen, K. W., 2017. Empirics of multi-modal traffic networks using the 3d macroscopic fundamental diagram. Transportation Research Part C: Emerging Technologies 82, 88–101.
URL https://dx.doi.org/10.1016/j.trc.2017.06.009
Loder, A., Dakic, I., Bressan, L., Ambühl, L., Bliemer, M. C., Menendez, M., Axhausen, K. W., 2019. Capturing network properties with a functional form for the multi-modal macroscopic fundamental diagram. Transportation Research Part B: Methodological 129, 1 – 19.
URL https://dx.doi.org/10.1016/j.trb.2019.09.004
Lopez, C., Leclercq, L., Krishnakumari, P., Chiabaut, N., van Lint, H., 2017. Revealing the day-to-day regularity of urban congestion patterns with 3d speed maps. Scientific Reports 7, 1–11.
URL https://dx.doi.org/10.1038/s41598-017-14237-8
Mahmassani, H., Williams, J. C., Herman, R., 1984. Investigation of network-level traffic flow relationships: Some simulation results. Transportation Research Record: Journal of the Transportation Research Board 971, 121–130.
URL https://dx.doi.org/10.3141/2315-16
Mariotte, G., Leclercq, L., 2019. Flow exchanges in multi-reservoir systems with spillbacks. Transportation Research Part B: Methodological 122, 327 – 349.
URL https://dx.doi.org/10.1016/j.trb.2019.02.014
Mariotte, G., Leclercq, L., Batista, S., Krug, J., Paipuri, M., 2020. Calibration and validation of multi-reservoir mfd models: A case study in lyon. Transportation Research Part B: Methodological 136, 62 – 86.
URL https://dx.doi.org/10.1016/j.trb.2020.03.006
Mariotte, G., Leclercq, L., Laval, J. A., 2017. Macroscopic urban dynamics: Analytical and numerical comparisons of existing models. Transportation Research Part B 101, 245–267.
URL https://dx.doi.org/10.1016/j.trb.2017.04.002
Moeckel, R., 2017. Constraints in household relocation: Modeling land-use/transport interactions that respect time and monetary budgets. Journal of Transport and Land Use 10 (1), 211–228.
URL https://dx.doi.org/10.5198/jtlu.2015.810
Mohajerpoor, R., Saberi, M., Vu, H. L., Garoni, T. M., Ramezani, M., 2019. H_∞ robust perimeter flow control in urban networks with partial information feedback. Transportation Research Part B: Methodological.
URL https://dx.doi.org/10.1016/j.trb.2019.03.010
Nielsen, O. A., 1997. On the distributions of the stochastic components in sue (stochastic user equilibrium) traffic assignment models. In: Transportation planning methods: proceedings of seminar held at the European Transport Forum Annual Meeting, Brunel University, England 1-5

September 1997. pp. 77–93.
Nielsen, O. A., 2000. A stochastic transit assignment model considering differences in passengers utility functions. Transportation Research Part B 34 (5), 377–402.
URL https://dx.doi.org/10.1016/S0191-2615(99)00029-6
Nielsen, O. A., Daly, A., Frederiksen, R. D., 2002. A stochastic route choice model for car travellers in the copenhagen region. Networks and Spatial Economics 2, 327–346.
URL https://dx.doi.org/10.1023/A:102089542
Ortuzar, J. D., Willumsen, L. G., 2011. Modelling transport, 4th Edition. Chichester, England: John Wiley Sons.
Paipuri, M., Leclercq, L., 2020a. Bi-modal macroscopic traffic dynamics in a single region. Transportation Research Part B: Methodological 133, 257 – 290.
URL https://dx.doi.org/10.1016/j.trb.2020.01.007
Paipuri, M., Leclercq, L., 2020b. Empirical validation of bimodal mfd models. Frontiers in Future Transportation.
URL https://dx.doi.org/10.3389/ffutr.2020.00001
Paipuri, M., Xu, Y., González, M. C., Leclercq, L., 2020. Estimating mfds, trip lengths and path flow distributions in a multi-region setting using mobile phone data. Transportation Research Part C: Emerging Technologies 118, 102709.
URL http://www.sciencedirect.com/science/article/pii/S0968090X20306240
Prato, C. G., 2009. Route choice modelling: past, present and future research directions. Journal of Choice Modelling 2, 65–100.
URL https://dx.doi.org/10.1016/S1755-5345(13)70005-8
Prato, C. G., Bekhor, S., 2006. Applying branch and bound techniques to route choice set generation. Transportation Research Record, 19–28.
URL https://dx.doi.org/10.3141/1985-03
Ramezani, M., Haddad, J., Geroliminis, N., 2015. Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control. Transportation Research Part B 74, 1–19.
URL https://dx.doi.org/10.1016/j.trb.2014.12.010
Ramezani, M., Nourinejad, M., 2018. Dynamic modeling and control of taxi services in large-scale urban networks: A macroscopic approach. Transportation Research Part C: Emerging Technologies 94, 203 – 219, iSTTT22.
URL https://dx.doi.org/10.1016/j.trc.2017.08.011
Ramming, M., 2002. Network Knowledge and Route Choice, PhD thesis. Massachusetts Institute of Technology.
Saedi, R., Verma, R., Zockaie, A., Ghamami, M., Gates, T. J., 2020. Comparison of support vector and non-linear regression models for estimating large-scale vehicular emissions, incorporating network-wide fundamental diagram for heterogeneous vehicles. Transportation Research Record 2674 (5), 70–84.
URL https://dx.doi.org/10.1177/0361198120914304
Saeedmanesh, M., Geroliminis, N., 2016. Clustering of heterogeneous networks with directional flows based on "snake" similarities. Transportation Research Part B: Methodological 91, 250–269.
URL https://dx.doi.org/10.1016/j.trb.2016.05.008
Saeedmanesh, M., Geroliminis, N., 2017. Dynamic clustering and propagation of congestion in heterogeneously congested urban traffic networks. Transportation Research Procedia 23, 962–979.
URL https://dx.doi.org/10.1016/j.trb.2017.08.021
Sirmatel, I. I., Geroliminis, N., 2018. Economic model predictive control of large-scale urban road networks via perimeter control and regional route guidance. IEEE Transactions on Intelligent Transportation Systems 19, 1112–1121.
URL https://dx.doi.org/10.1109/TITS.2017.2716541
Sirmatel, I. I., Geroliminis, N., 2019. Nonlinear moving horizon estimation for large-scale urban road networks. IEEE Transactions on Intelligent Transportation Systems, 1–12.
URL https://dx.doi.org/10.1109/TITS.2019.2946324
van der Zijpp, N. J., Catalano, S. F., 2005. Path enumeration by finding the constrained k-shortest paths. Transportation Research Part B: Methodological 39, 545–563.
URL https://dx.doi.org/10.1016/j.trb.2004.07.004
Vickrey, W., 2020. Congestion in midtown manhattan in relation to marginal cost pricing. Economics of Transportation 21, 100152.
URL https://dx.doi.org/10.1016/j.ecotra.2019.100152
Wei, B., Saberi, M., Zhang, F., Liu, W., Waller, S. T., 2020. Modeling and managing ridesharing in a multi-modal network with an aggregate traffic representation: A doubly dynamical approach. Transportation Research Part C: Emerging Technologies 117, 102670.
URL https://dx.doi.org/10.1016/j.trc.2020.102670
Yang, F., Jin, P. J., Wan, X., Li, R., Ran, B., 2013. Dynamic origin-destination travel demand estimation using location-based social networking data. In: 92nd Annual Meeting Transportation Research Board. Washington DC, USA.
Yang, H., Ke, J., Ye, J., 2018. A universal distribution law of network detour ratios. Transportation Research Part C: Emerging Technologies 96, 22 – 37.
URL https://doi.org/10.1016/j.trc.2018.09.012
Yang, K., Menendez, M., Zheng, N., 2019. Heterogeneity aware urban traffic control in a connected vehicle environment: A joint framework for congestion pricing and perimeter control. Transportation Research Part C: Emerging Technologies 105, 439 – 455.
URL https://dx.doi.org/10.1016/j.trc.2019.06.007
Yang, K., Zheng, N., Menendez, M., 2018. Multi-scale perimeter control approach in a connected-vehicle environment. Transportation Research Part C: Emerging Technologies 94, 32–49.
URL https://dx.doi.org/10.1016/j.trc.2017.08.014
Yildirimoglu, M., Geroliminis, N., 2014. Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams. Transportation Research Part B: Methodological 70, 186–200.
URL https://dx.doi.org/10.1016/j.trb.2014.09.002

Yildirimoglu, M., Ramezani, M., Geroliminis, N., 2015. Equilibrium analysis and route guidance in large-scale networks with mfd dynamics. Transportation Research Part C: Emerging Technologies 59, 404 – 420, special Issue on International Symposium on Transportation and Traffic Theory.
URL https://dx.doi.org/10.1016/j.trc.2015.05.009

Yildirimoglu, M., Sirmatel, I. I., Geroliminis, N., 2018. Hierarchical control of heterogeneous large-scale urban road networks via path assignment and regional route guidance. Transportation Research Part B: Methodological 118, 106–123.
URL https://dx.doi.org/10.1016/j.trb.2018.10.007

Zheng, N., Geroliminis, N., 2020. Area-based equitable pricing strategies for multimodal urban networks with heterogeneous users. Transportation Research Part A: Policy and Practice 136, 357 – 374.
URL https://dx.doi.org/10.1016/j.tra.2020.04.009

Zheng, N., Rérat, G., Geroliminis, N., 2016. Time-dependent area-based pricing for multimodal systems with heterogeneous users in an agent-based environment. Transportation Research Part C: Emerging Technologies 62, 133–148.
URL https://dx.doi.org/10.1016/j.trc.2015.10.015

Zhong, R., Chen, C., Huang, Y., Sumalee, A., Lam, W., Xu, D., 2017. Robust perimeter control for two urban regions with macroscopic fundamental diagrams: A control-lyapunov function approach. Transportation Research Procedia 23, 922–941.
URL https://dx.doi.org/10.3141/2493-09

Zhou, L., Zhong, S., Ma, S., Jia, N., 2014. Prospect theory based estimation of driver's risk attitudes in route choice behaviors. Accident Analysis and Prevention 73, 1–11.
URL https://dx.doi.org/10.1016/j.aap.2014.08.004