POLITECNICO
MILANO 1863

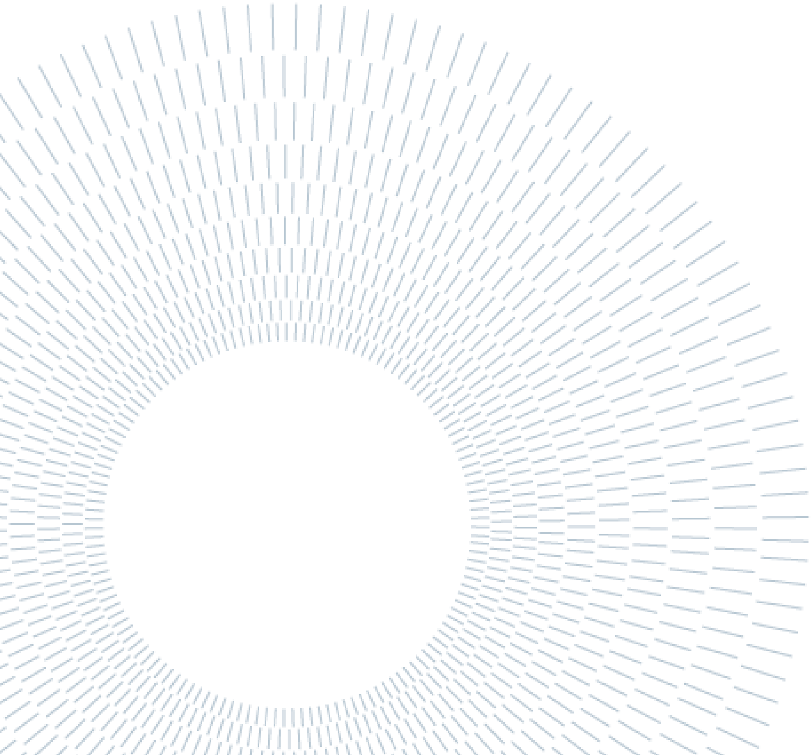
School of Civil, Environmental and Land Management Engineering
Master’s Degree in Civil Engineering – Transport Infrastructures

TomTom Data Applications for the Assessment of Tactical Urbanism Interventions: The Case of Bologna

Author: Marco Pozzoni
Student ID: 968906

Advisor: Prof. Roberto Maja
Co-advisor: Dr. Giulia Ceccarelli

Academic Year 2022/2023



Abstract

This work aims to evaluate how a temporary school square in the neighbourhood of Bolognina, in the city of Bologna, impacted on vehicular patterns through exploiting TomTom Floating Car Data (FCD) from before and after the intervention, which was implemented under the principles of tactical urbanism approach.

After statistical validation of available datasets through two-tailed paired Student's t-tests, trend analyses have been performed on sample sizes and speed-related values to detect global variations in the first place, and more thoroughly among clusters of road segments based on graph-calculated distance from the intervention site. Results suggest that traffic flows have been relocated from segments directly interested by the intervention, where a sharp decrease has been registered (-23.9%), towards adjacent streets or segments in a buffer area, which have recorded a moderate increase (+5.7% and +2.4% respectively), so the phenomenon of traffic evaporation did not take place as opposed to more widespread tactical urbanism interventions described in literature.

OD matrices per 15-minute time fractions over the three selected peak time slots have been extracted through a data driven procedure proposed by Krishnakumari et al. (2019) and adapted to the case under examination, in order to obtain reliable input data for a future development of traffic microsimulation models. The extraction method is based on least squares optimization problems solving systems of linear equations representing OD flows assigned to observed link, after selecting a set of $\bar{k}$ shortest paths through a Path Size Logit (PSL) model.

Even though the availability of large amount of data could not overcome typical underdetermination of the problem, due to the key issue of data dependence among traffic counts, the validation of retrieved matrices returned good results in terms of correlation between observed and estimated link flows. In the few cases where the quality of correlation fell, underlying causes have been investigated and the influence of outliers, amplified by the high fragmentation of the provided road graph, might represent the core problem.

Key-words: tactical urbanism, OD matrix estimation, Floating Car Data

Abstract in italiano

Questo lavoro punta a valutare come l'implementazione di una piazza scolastica temporanea, secondo i principi dell'urbanistica tattica, abbia impattato sugli schemi di traffico del quartiere Bolognina, nella città di Bologna, attraverso l'utilizzo di Floating Car Data (FCD) forniti da TomTom raccolti prima e dopo l'implementazione.

Dopo la validazione statistica dei dati attraverso t test accoppiati a due code, sono state svolte delle analisi sull'andamento dei flussi e dei parametri relativi alle velocità per individuare variazioni globali in primis e più approfonditamente tra classi di segmenti stradali formate sulla base della distanza sul grafo dal sito dell'intervento. I risultati mostrano che i flussi di traffico sono migrati dalle strade coinvolte dall'intervento, dove si è registrata una netta flessione (-23.9%), verso le strade adiacenti o nell'area a margine, che hanno segnato un moderato incremento (+5.7% e +2.4% rispettivamente), quindi il fenomeno di "evaporazione del traffico" non è stato riscontrato a differenza di esempi di interventi più estesi descritti in letteratura.

Matrici OD in frazioni di 15 minuti per ciascuna delle tre fasce orarie di punta selezionate sono state ricavate tramite una procedura basata sui dati proposta da Krishnakumari et al. (2019) e adattata al caso in questione, al fine di ottenere dati di input affidabili per lo sviluppo futuro di modelli microsimulativi di traffico. Il metodo di stima si basa sulla risoluzione mediante ottimizzazioni ai minimi quadrati di sistemi di equazioni lineari che rappresentano flussi OD assegnati ai segmenti osservati, dopo aver selezionato un insieme di $\bar{k}$ percorsi di costo minimo attraverso un modello Logit con fattore di sovrapposizione dei percorsi.

Sebbene l'ampia disponibilità di dati non abbia potuto superare la sottodeterminazione tipica del problema, a causa della questione chiave della dipendenza tra conteggi di traffico, la validazione delle matrici ricavate ha restituito buoni risultati in termini di correlazione tra flussi osservati e stimati. Nei pochi casi in cui la qualità della correlazione si è ridotta, sono state indagate le ragioni di fondo e l’influenza di valori anomali, amplificati dall’elevata frammentazione del grafo stradale di base, potrebbe averne rappresentato il problema centrale.

Parole chiave: urbanistica tattica, stima di matrici OD, Floating Car Data

Acknowledgments

This research has been carried out during a curricular internship at Fondazione Transform Transport ETS (Milan, Italy), the non-profit research foundation launched by Systematica Srl on March 2022 and focused on innovation in mobility and transport planning.

The author thanks the team of citiEU Consultancy LTD (Guildford, United Kingdom) for their fruitful collaboration and for sharing data.

The analysed data were treated according to the General Data Protection Regulation (EU, 2016/679).

This research received no specific grant from any funding agency in the public, commercial and not-for-profit sectors.

The author would like to thank Prof. Roberto Maja, Dr. Giulia Ceccarelli and Dr. Andrea Gorrini for their supervision and collaboration.

The results of this research have been submitted for oral presentation to the European Transport Conference: Pozzoni, M., Ceccarelli G., Gorrini, A., Manenti, L., Sanfilippo, L., Brignone, A. (2023, submitted). TomTom Data Applications for the Assessment of Tactical Urbanism Interventions: The Case of Bologna. In: *51th European Transport Conference 2023 (ETC 2023),* 6-8 September 2023, Milan (Italy).

Table of Contents

Abstract iii

Abstract in italiano v

Acknowledgments vii

Table of Contents ix

List of Figures xi

List of Tables xiii

1 Introduction 1

2 General framework 3

3 Literature review 7

- 3.1. Tactical urbanism 7
- 3.2. Floating Car Data (FCD) 8
- 3.3. OD matrix estimation 9

4 Enabling data and methodology 11

- 4.1. TomTom data 11
- 4.2. Trend statistical analysis 16
- 4.3. OD matrices estimation procedure 18
  - 4.3.1. Workflow 18
  - 4.3.2. Representativity of sample data 19
  - 4.3.3. Zoning 21
  - 4.3.4. Generation and attraction 23
  - 4.3.5. Shortest paths calculation 24
  - 4.3.6. Route choice 25
  - 4.3.7. Link traffic counts selection 26
  - 4.3.8. Equation system solving 29
  - 4.3.9. Validation of results 30

5 Results 33

- 5.1. Statistical analysis 33

5.1.1. T-tests 33
5.1.2. Trend analysis 37
5.2. OD matrices estimation 48
6 Discussion 57
7 Conclusion and future developments 61
Bibliography 65

List of Figures

Figure 1. Aerial view of the temporary school square in via Procaccini, Bologna (photo by Margherita Caprilli) 4

Figure 2. Selected area for TomTom query 15

Figure 3. FRC classification 15

Figure 4. Buffer segments classification 17

Figure 5. Workflow of OD matrix estimation procedure 19

Figure 6. Network size reduction 22

Figure 7. Origin-Destination nodes 23

Figure 8. Average speed on weekdays for FRC 2-3 segments - comparison *ex-ante/ex-post* 37

Figure 9. Average speed on weekdays for FRC 4-6 segments - comparison *ex-ante/ex-post* 38

Figure 10. Average speed 85th percentile on weekdays for FRC 2-3 segments - comparison *ex-ante/ex-post* 39

Figure 11. Average speed 85th percentile on weekdays for FRC 4-6 segments - comparison *ex-ante/ex-post* 39

Figure 12. Average traffic flow on weekdays on intervention streets - comparison *ex-ante/ex-post* 40

Figure 13. Average traffic flow on weekdays on adjacent streets - comparison *ex-ante/ex-post* 41

Figure 14. Average traffic flow on weekdays on buffer streets - comparison *ex-ante/ex-post* 41

Figure 15. Average traffic flow on weekdays on control streets - comparison *ex-ante/ex-post* 42

Figure 16. Average traffic flow variation adjusted over control segments fluctuations 43

Figure 17. Average traffic flow on weekdays - comparison *ex-ante/ex-post* 44

Figure 18. Relative variation in average traffic flow - time slot 7:00-9:00 AM.... 45

Figure 19. Relative variation in average traffic flow - time slot 1:00-3:00 PM 45

Figure 20. Relative variation in average traffic flow - time slot 5:00-7:00 PM 46

Figure 21. Absolute variation in average traffic flow - time slot 7:00-9:00 AM .. 47

Figure 22. Absolute variation in average traffic flow - time slot 1:00-3:00 PM... 47

Figure 23. Absolute variation in average traffic flow - time slot 5:00-7:00 PM... 48

Figure 24. Contraction of OD matrix layout 49

Figure 25. $R2$ values for *ex-ante* scenario 52

Figure 26. $R2$ values for *ex-post* scenario 52

Figure 27. Linear regression of discarded flows - *ex-ante,* weekdays, 7:15-7:30 AM time fraction.. 54

Figure 28. Linear regression of discarded flows - *ex-ante,* weekdays, 5:45-6:00 PM time fraction.. 54

Figure 29. Absolute variation of OD flows between *ex-ante* and *ex-post* scenarios 55

List of Tables

Table 1. Functional Road Classes (FRC) description 12

Table 2. Expansion coefficients ($Cexp,k$) calculation 20

Table 3. Origin-Destination zones 22

Table 4. Results of t-tests for average speed values 33

Table 5. Results of t-tests for sample size values 34

Table 6. Results of t-tests performed on speed 85th percentile values 35

Table 7. Average traffic flow variation relative to control segments 43

Table 8. Number m of selected linear equations per each time fraction 50

Table 9. Coefficients of determination R-squared ($R2$) 51

Table 10. Theil's inequality coefficients for low-$R2$ regressions 53

1 Introduction

The purpose of this master thesis work is to assess how a punctual tactical urbanism intervention changed mobility patterns in the neighbourhood road network in which it was implemented, by analysing data collected through innovative techniques such as GPS-based Floating Car Data (FCD). In particular, the goal is to extract reliable Origin-Destination matrices as inputs for traffic demand models through elaborating available FCD datasets.

Tactical urbanism is an innovative urban design approach that enables public authorities and administrations to temporarily requalify deteriorated or misused urban spots to be regenerated and reinstated into their original purposes. Such approach is characterized by limited financial resources usage and streamlined bureaucratic procedures and requires the involvement of the public, especially locals or usual users of the area, who can have their say during the project design and the later assessment phases. In the recent years, the practice of tactical urbanism has been recently increasingly adopted in urban settings throughout the world, as we can count many examples in Milan (Italy), São Paulo (Brazil), Bogotá (Colombia), Quito (Ecuador), Mumbai (India), Istanbul (Turkey), Addis Ababa (Ethiopia) (GCDI, 2022).

While the impact on pedestrian accessibility of such reconverted spaces has been recorded from previous experiences as good, literature has recently raised the question on how it can affect the surroundings in terms of transport network performance. There are some questions still to be thoroughly addressed by academic research like:

- To what extent can a punctual intervention affect the surrounding road network?
- What happens in the streets directly affected? And what to the buffer area? How large is such area?

Furthermore, more and more innovative technologies are being employed to elaborate input data for traffic demand models: while the traditional method of collecting OD matrices to run simulations has always been either household surveys, travel interviews or traffic counts backed by fixed sensors, nowadays research found out that the impressive amount of data passively recorded by devices embedded in vehicles as well as mobile phones or even wearable devices could be of use when estimating mobility patterns, both boosting coverage of data samples and enhancing predictions by solving the problem of underdetermination, which usually affects estimations. The research questions regarding the field of traffic modelling that this paper aims to answer are:

- Can a GPS-based FCD dataset overcome underdetermination problems when estimating OD matrices through recorded link traffic counts and average speeds/travel times? If yes, how?
- Which procedure is suitable to obtain such matrix without any ground truth OD data? Will it provide a reliable matrix?

After a focus on the case of study in the following Section 2 and a literature review based on previous assessments of tactical urbanism interventions and on the state-of-the-art for the estimation of OD matrices through Floating Car Data in Section 3, available data and the selected methodology to be used are presented in Section 4. Results of trend statistical analysis of link supply patterns and estimated OD matrices are displayed in Section 5; discussions on results and limitations of adopted procedures in Section 6; conclusions and future developments in Section 7.

2 General framework

This research work follows a stream of mobility-related studies conducted by Transform Transport[1]. They addressed the topic of walkability for children, meant as a metric to evaluate friendliness of walking and spending time in urban areas for young users aged 5-13. First, a study was conducted to assess levels of pedestrian accessibility for services devoted to children and their needs in the city of Bologna through GIS analysis of available georeferenced data at macro scale (Abdelfattah et al., 2021) and through Space Syntax at a more detailed scale focusing on the Bolognina district (Gorrini et al., 2023). Second, following the implementation enacted by the municipality of Bologna and promoted by Foundation of Urban Innovation of a temporary school square for young students in via Camillo Procaccini, Bologna close to a secondary school, pedestrian and vehicular flow patterns were monitored on a micro scale via video analytics backed by machine learning techniques, providing temporal and spatial analyses of recorded data in order to support the iterative design process typical of tactical urbanism approach (Ceccarelli et al., 2023).

The intervention under examination consists of a reconversion of an unregulated parking space located in the crossroads among Via Camillo Procaccini, Via Andrea da Faenza and Via Antonio di Vincenzo in the neighbourhood of Bolognina, into a temporary school square: 300 square metres have been pedestrianised, equipped with street furniture (benches, seating spots and flower boxes) and highlighted through colourful paint. The aim of the project in which the intervention takes part is to ensure young students augmented safety and autonomy in their home-school-home journeys and to provide a space for socializing and waiting together before the school opens. It was implemented in March 2022 after three weeks of works and inaugurated on April 2nd, 2022: as a street experiment and according to tactical urbanism principles, it lasted approximately 12 months, during which usage patterns were monitored in order to understand if the intervention had reached expected goals and to later define a permanent configuration of this urban space.

---

[1] Fondazione Transform Transport ETS (Milan, Italy) is a non-profit research foundation launched by Systematica Srl in March 2022 and focused on innovation in mobility and transport planning. It provides innovative, inclusive, and sustainable mobility solutions for shaping the future of cities worldwide in line with the UN's SDG 11 (Sustainable cities and communities) – see: https://transformtransport.org/

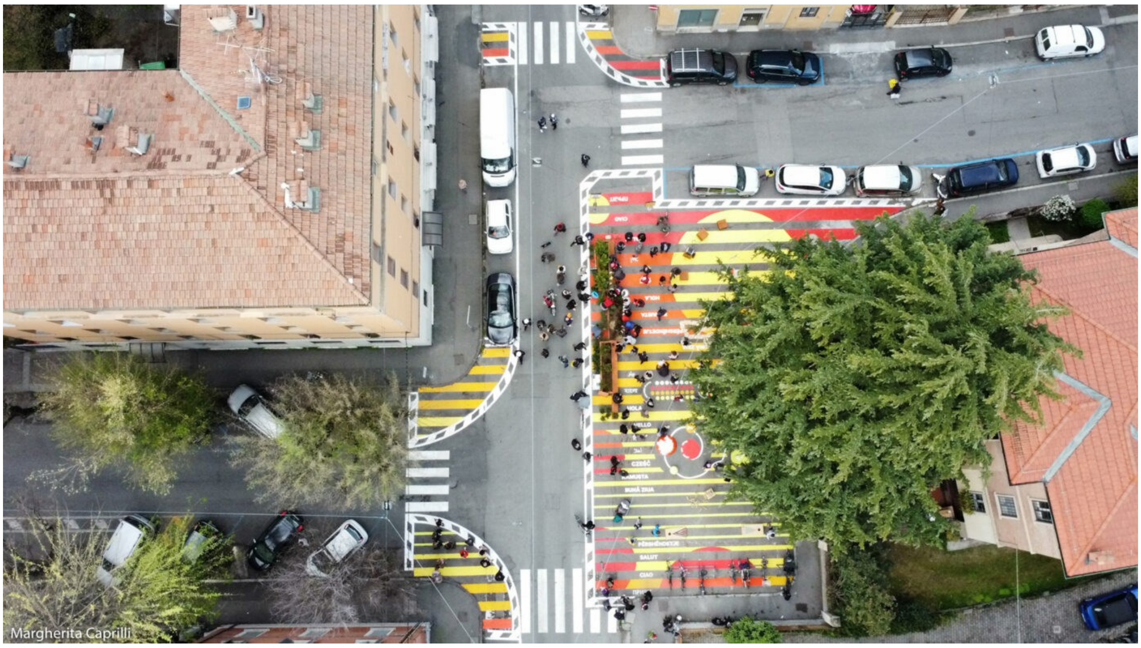


Figure 1. Aerial view of the temporary school square in via Procaccini, Bologna (photo by Margherita Caprilli)

This current study aims to widen the assessment perspective from the punctual observations at the intervention spot to a larger buffer area surrounding the temporary school square. The main focus is set on vehicular traffic variations at neighbourhood scale, checking if vehicle patterns in the district have been disrupted, remained untouched or even improved in terms of congestion by the new road allocation scheme.

To that end, via a GPS-based dataset containing aggregate traffic information for peak daily time slots, private car mobility trends will be analysed at macro scale via different clustering of road segments, comparing data from before and after the implementation of the square.

Then, data will be elaborated to extract OD matrices which, once fed to a microsimulation software in a future step of this research, will serve as inputs for further assessments on streets and intersections in the neighbourhood, extracting useful metrics to detect variations in vehicular patterns, congestion problems if present or definitive approval of the intervention from road traffic management point of view.

3 Literature review

3.1. Tactical urbanism

As stated in the introduction, tactical urbanism approach has been increasingly promoted by public administrations to boost the implementation of rapid and effective changes in the urban setting under their competence. Such street experiments aim at establishing the idea that streets should belong to people rather than traffic (Bertolini, 2020) by physically reducing road space capacity for vehicles in favour of pedestrian and soft mobility. This practice allegedly leads to *traffic evaporation*, referring to the reduction in traffic flows as a consequence of capacity-limiting interventions on the transport network: such phenomenon can be intended as the opposite of induced traffic, which results from an expansion of road capacity (e.g., opening of a new road, widening of the road via new lanes). While literature focused a lot on induced traffic implications on mobility patterns variations, so far very few researchers have investigated traffic evaporation potentially caused by reallocation scheme interventions, despite the current and popular need to switch to more sustainable modal choices than private car. There are studies on "disappearing" traffic caused by road closures particularly on physiological bottlenecks of road networks e.g., bridges (Hunt et al., 2002) and tunnels (Tennøy et al., 2021), either planned or not: overall, common results from these studies suggest that a decrease in traffic flows should be expected when implementing road space reallocation schemes or, more generally, interventions that reduce street capacity, with traffic behaviour change in the new scheme area to be proportional to the level of disruption to the network. More, even though capacity of the network is downgraded, road congestion tends to be less severe than conventional traffic models would suggest. However, because of the poor research effort involved as of today, correlation among studies remains difficult due to different geographical contexts and types of interventions.

Research also lacks work exploring impacts on a wider perspective on the network, so not only in the local intervention area but on a meso-scaled point of view. A valuable exception to this is Nello-Deakin (2022), which exploited datasets of traffic counts provided by permanent sensors to evaluate the extent of traffic evaporation following the extensive implementation program of multiple tactical urbanism interventions in the Eixample district of the city of Barcelona, Spain. The study investigated not only streets directly interested by interventions but also assessed traffic levels in a buffer area within 500 metres, leading to confute the assertion that traffic would simply gather onto more convenient paths and congest roads in a limited buffer area, as results suggest that overall traffic has diminished, with a very low relative increase in the streets adjacent to interventions.

3.2. Floating Car Data (FCD)

Besides traditional methods, lately more and more traffic data sources and collection techniques have been either improved or invented, each of them featuring their own advantages and disadvantages according to the usage and the collection purpose.

The aim of this research effort is to progressively outdate traditional techniques such as travel surveys (household, on-board etc.), which can be time- and money-consuming other than poorly accurate for low-sampled trips.

As comprehensively portrayed by Leduc (2008), road traffic data collection methods can be distinguished based on whether measurements are performed by sensors – or people, in the case of manual counts – located along the roadside or by vehicles themselves acting as moving sensors for the road network. This second cluster is known as Floating Car Data (FCD) and works by collecting real-time traffic data through consecutive positions of equipped vehicles via mobile phones or GPS devices, along with complementary data such as instantaneous speed and direction of travel. By providing high quality and cost-effective data, as there is no need of implementation of hardware in situ but only on vehicles, FCD is a promising alternative to existing technologies for road data collection and it is also crucial in the development and functioning of Intelligent Transport Systems (ITS), which mostly rely on precise real-time information on traffic conditions through the network. It is to remind, though, that such raw data do not explicitly provide information to calibrate or validate estimates on behavioural choice, which travel surveys can usually offer.

Based on the connectivity option, FCD can be GPS-based or reliant on cellular phone networks: while GPS provides a 10 times better precision but suffers lack of vehicles equipped with suitable devices, cellular based technology compensates low accuracy with the large number of sample size corresponding to a wide coverage.

Lately, Floating Car Data has been used as ground data for research studies addressing mobility tasks and issues, such as detecting and analysing urban patterns from GPS traces (Necula, 2015), estimating traffic delays and network speeds from low-frequency GPS taxi traces (Deng et al., 2015), detecting traffic congestion and incidents (D'Andrea et al., 2017), or estimating or updating OD matrices, which will be furtherly addressed in the next subsection.

3.3. OD matrix estimation

Origin-Destination matrix estimation can be a challenging requirement prior to transport network simulations: in the framework of 4-stage traffic models, it can be thought as the inverse procedure of traffic assignment step (Cascetta, 1998): while the latter loads the network with flows determined according to estimated or observed traffic demand and route choice models, OD matrix estimation goes the other way round focusing on estimating path flows (which once aggregated represent OD pair flows) based on available records of link flows in the reference period. Such data is usually available through travel surveys or traffic counts in specific sections of the transport network under examination, either manual or via automated sensors (magnetic spires, cameras etc.).

Traditional methods of vehicular demand estimation through traffic count data recorded on links of the network require ground truth OD data: such process is usually an update of an outdated Origin-Destination matrix via traffic counts collected in the reference period, with data held in the old matrix being attributed a level of confidence in relation to their age.

Estimation of OD matrices from scratch is a challenge from mathematical standpoint, as errors affecting route choice models and observations can lead to undetermined problems (no existing matrix capable of representing actual observed flows); moreover, despite the wider coverage with respect to interviews or surveys, traffic count campaigns, either manual or via fixed automated sensors, usually provide flow information only on a limited number of links of the network, resulting in underdetermined equation systems whose solution need to be estimated through regression procedures such as constrained generalized least squares or maximum likelihood algorithms. Furthermore, the absence of an outdated OD matrix as ground data for the estimation process strongly affects Origin-Destination pairs that are not covered by any of the planned traffic counts.

Due to the current availability of large streams of passively collected data, data-driven methods have started to be developed and validated by academic research as an efficient alternative to traditional OD estimation methodologies. For instance, in Bonnel et al. (2018) and later in Fekih et al. (2021), data from mobile phone networks were used to estimate OD matrices in the Rhône Alpes region in France. Demissie et al. (2022) analysed GPS trajectory data over one year to estimate origin-destination flows of trucking vehicles within the province of Alberta, Canada. Ge et al. (2016) used aggregated data of GPS traces to avoid privacy issues and implemented a sequential updater based on maximum entropy principle to update an outdated matrix. Tolouei et al. (2016) validated such methods by comparing matrices obtained through roadside interviews together with trip-end and gravity models and through the application of mobile phone data, stating that trip matrices developed through mobile data were as accurate as the ones estimated through conventional models if refined and adjusted to remove intrinsic biases and limitations; more, the advantage of larger sample sizes allowed mobile data to estimate in a more consistent manner trips where no roadside observed data were available. Krishnakumari et al. (2019) proposed a method applicable in presence of 3D supply patterns only (sample size and speed values per each time period) on all segments of the network, consisting of a large equation system to be integrated with principal component analysis in case of severely underdetermined systems (typical of larger networks) and featuring fundamental assumptions regarding route choice, e.g., cutting off the number of considered paths and assigning a proportionality coefficient to each path in the OD pair-specific set calculated through a route choice model.

Overall, according to the nature and the aggregation level of available datasets, literature provides appropriate methods and solutions depending on the purpose with which data can be elaborated.

4 Enabling data and methodology

4.1. TomTom data

The available dataset employed to carry out analyses and elaboration in this research work were provided by TomTom as Traffic Stats API Area Analysis. Such application programming interface is based on the collection of real-time Floating Car Data anonymously sent by GPS enabled devices to TomTom servers in exchange of accurate on-trip routing and alerts on traffic conditions. Elaboration of collected data allows to return valuable insights into traffic levels on the road network through time. Queried datasets for this specific API consist of aggregate information on the links of the selected network for a requested reference period: in particular, for each segment in the selected area of the road map elaborated by TomTom the following data are available per each fraction of each time slot:

- sample size
- speed
  - average
  - harmonic average
  - median
  - standard deviation
  - percentiles (5th – 95th)
- travel time
  - median
  - average
  - standard deviation

Besides data visualization through interactive dashboards, datasets can be downloaded as JSON, GEOJSON or shapefiles to be treated via data analysis programming languages or elaborated through GIS.

Segments of the road network in the selected area are clustered by TomTom according to the character of service they are supposed to provide as roads. Functional Road Classes (FRCs) are designed to categorize segments based on their functional importance, as this classification defines the role that any particular road or street plays in carrying the flow of trips through the road network. In Table 1 FCRs are outlined, as reported by TomTom Developers (2022):

Table 1. Functional Road Classes (FRC) description

FRC	Denomination	Description
0	Motorways; Freeways; Major Roads	All roads that are officially assigned as motorways.
1	Major Roads less important than Motorways	All roads of high importance, but not officially assigned as motorways, that are part of a connection used for international and national traffic and transport.
2	Other Major Roads	All roads used to travel between different neighbouring regions of a country.
3	Secondary Roads	All roads used to travel between different parts of the same region.
4	Local Connecting Roads	All roads making all settlements accessible or making parts (north, south, east, west, and central) of a settlement accessible.
5	Local Roads of High Importance	All local roads that are the main connections in a settlement. These are the roads where important through traffic is possible e.g.,: • arterial roads within suburban areas, industrial areas or residential areas • a rural road, which has the sole function of connecting to a national park or important tourist attraction
6	Local Roads	All roads used to travel within a part of a settlement or roads of minor connecting importance in a rural area.

7	Local Roads of Minor Importance	All roads that only have a destination function, e.g., dead-end roads, roads inside a living area, alleys: narrow roads between buildings, in a park or garden.
8	Other Roads	All other roads that are less important for a navigation system: • a path: a road that is too small to be driven by a passenger car • stairs • pedestrian tunnels • pedestrian bridges • alleys that are too small to be driven by a passenger car

For the case of study, a query to TomTom servers was sent asking for the following data:

- periods of time: September and October 2021 (*ex-ante*) – September and October 2022 (*ex-post*)
- date ranges: weekdays (Mondays to Fridays) – weekends (Saturdays, Sundays)
- daily time slots: 3 two-hour slots split into 15-minute fractions
    - 7:00 - 9:00 AM → school entrance
    - 1:00 - 3:00 PM → school exit
    - 5:00 - 7:00 PM → spontaneous usage of the square
- area: Bolognina district, Bologna (Figure 2)
- FRC: 1 to 6

Provided sample size is intended as the total number of vehicles registered in the segment through all the days of the sampling period, while each supplied value of speed or travel time is calculated as arithmetical average (unless stated otherwise) over each daily value in the selected date range and specific time fraction.

The choice of periods of time has been made based on the following assumptions:

- in the months of September and October schools are open, fundamental requirement given that the intervention is expected to impact also school-related traffic;
- a two-month dataset is enough to use as representative for mean traffic values;
- as the implementation of the temporary square happened in March, *ex-post* data are collected sufficiently later than the implementation of the temporary square, which is advisable to consider as several pieces of research literature (Goodwin et al., 1998; Hunt et al., 2002) encountered unusual congestion in the days straight after the road closure.

The choice for daily time slots and time fractions has been made in accordance with daily slots and granularity used for video analytics observations on the same intervention and related analyses performed in Ceccarelli et al. (2023).

Functional Road Classes have been subset due to budgetary constraints from 1 to 6; however, this is not as impacting as limitation since no motorway segment is included in the selected area and FRC 7 and FRC 8 road segments would not improve the dataset and the consequent evaluations with added benefit.

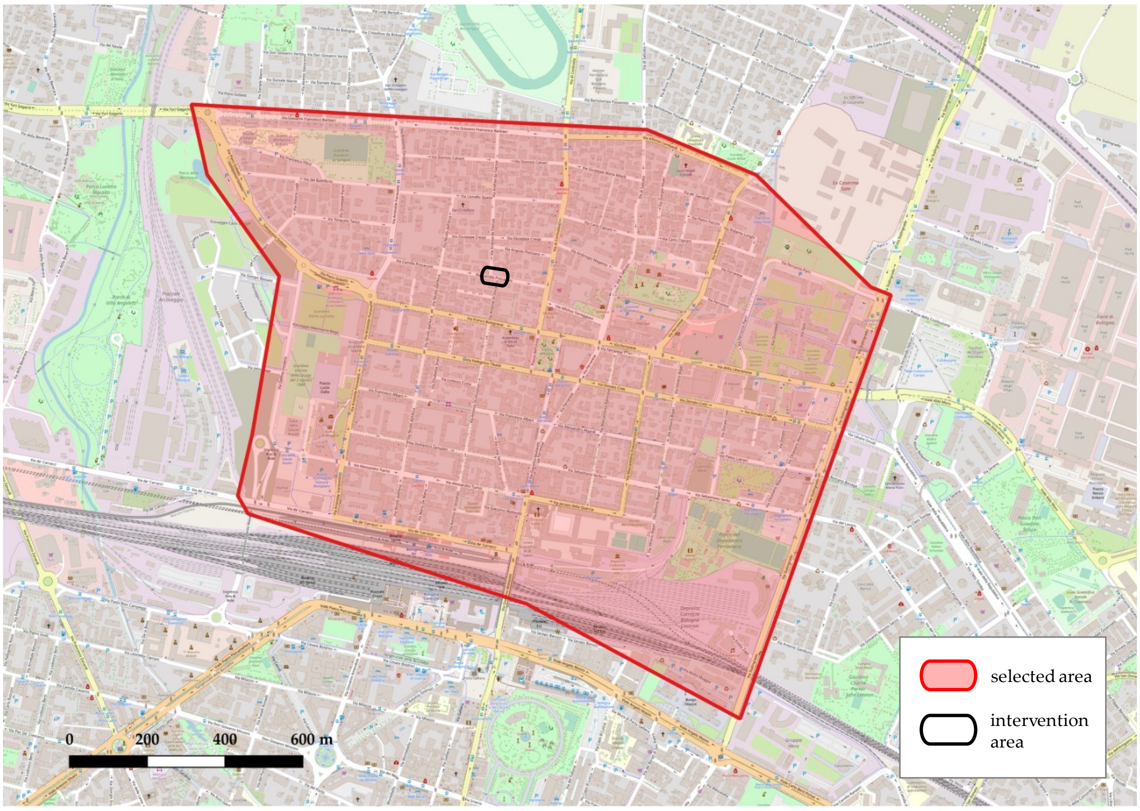

selected area
intervention area
0
200
400
600 m

Figure 2. Selected area for TomTom query

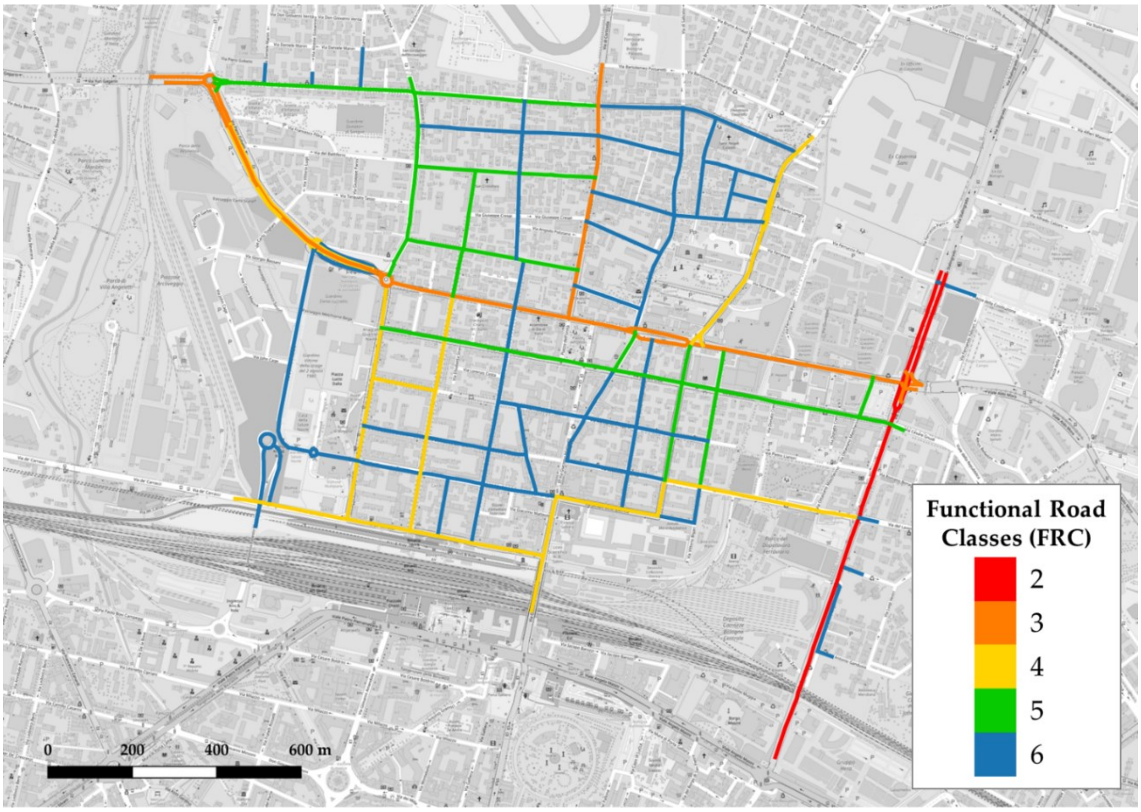

Functional Road Classes (FRC)
2
3
4
5
6
0
200
400
600 m

Figure 3. FRC classification

4.2. Trend statistical analysis

The goal of this step is to analyse the requested datasets under a macroscopic point of view on the whole network before and then by clusters identified by proper attributes of road segments in order to detect the magnitude of variations (if there are any) in traffic-related parameters such as sample sizes and speed-related values.

First, to assess significant differences between distinct clusters of available data, a set of Student's t-tests is performed. In this case, two-tailed paired (or dependent samples) t-tests have been used.

From the logical perspective, t-tests are based on the rejection of a null hypothesis, formulated at the start of the test: in the case of paired t-tests, in order to ascertain whether there is a substantial difference between the two datsets, the null hypothesis postulates no difference between the means of the two sets of samples. The key result of null hypothesis significance testing is the p-value, standing as the probability of obtaining results as compatible as the observed ones given that the null hypothesis is true. In other words, p-values determine whether difference between observation sample sets under the null hypothesis is due to randomness, intrinsic in the sampling process, or not. More in detail, after setting a significance level α, intended as the probability of rejecting the null hypothesis:

- if $p > \alpha$, empirical evidence is not strong enough to reject the initial hypothesis;
- if $p < \alpha$, observed data are statistically significant, so the null hypothesis is rejected.

One of the main assumptions to be made in order to run Student's t-tests is to assume equal variances between the data samples.

At this early stage of the study, comparisons intend to assess first whether sampled data are a good representation of a real and coherent situation, by evaluating consistent differences between values of speed and sample size recorded for different date ranges (weekdays vs. weekends) and for different FCRs, and then to assess a significant difference between values retrieved for *ex-ante* and *ex-post* traffic conditions.

Once data are statistically validated, a set of trend analyses is developed to detect any variations in sample sizes and speed-related values; in order to do so, an additional classification of segments is established. According to Nello-Deakin (2022), roads in the network are clustered on the basis of their proximity to the site of intervention. In particular:

- the intervention streets are segments directly interested by the modification;
- adjacent streets are close to the intervention and represent the best alternatives to avoid the site of intervention;
- streets in the buffer area are no farther than 400 meters on road graph distance;
- all the other roads in the network are classified as control segments.

In Figure 4, the selected road network is shown according to the classification above mentioned.

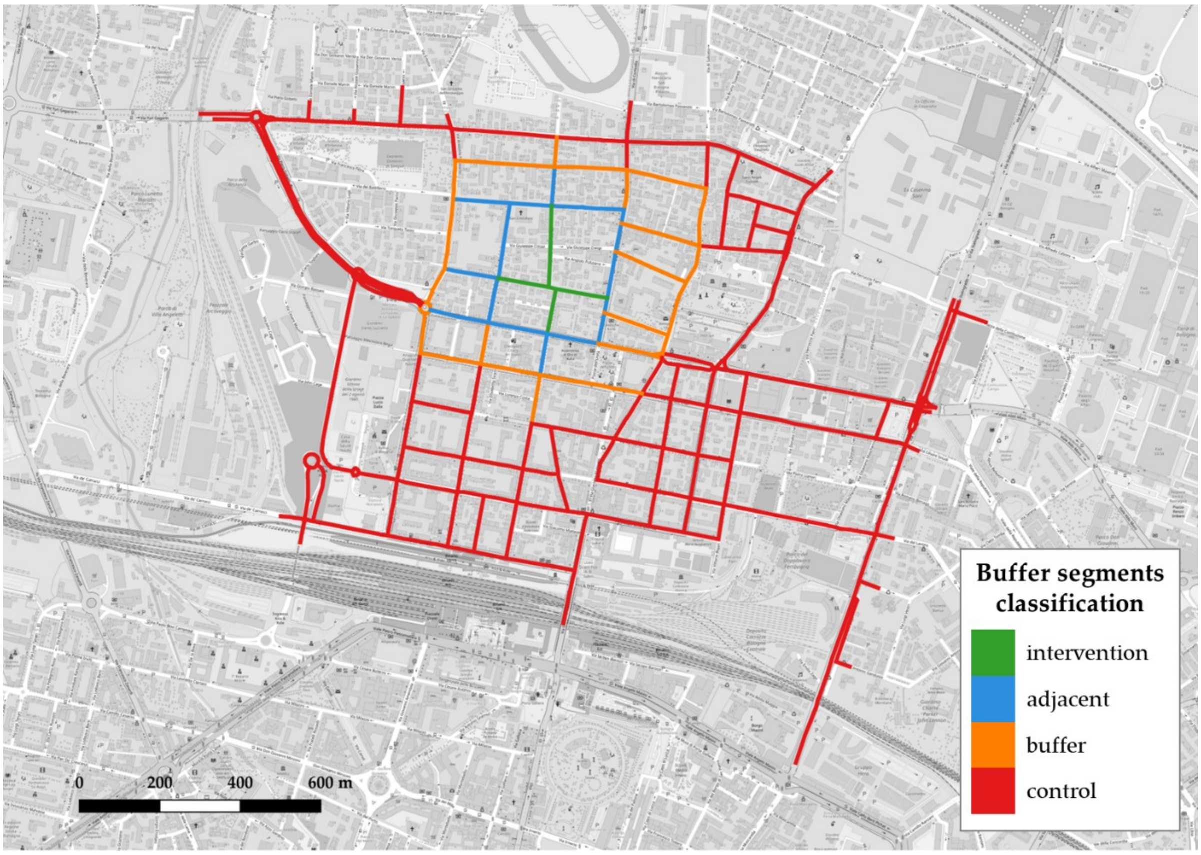


Figure 4. Buffer segments classification

4.3. OD matrices estimation procedure

The aim of this step is to estimate Origin-Destination matrices which fit the supply patterns extracted from the dataset with the same granularity of available data (one every 15 minutes). The procedure that will be used is the one proposed in Krishnakumari et al. (2019).

In this paper, the researchers presented an algorithm that allows to obtain OD matrices from 3D link supply patterns, intended as values of speed as well as the key figure of transiting vehicles over each time fraction. The algorithm intends to bypass the usual iterative loading of the road network until convergence. It is fundamentally based on the assumptions of cutting off the number of possible shortest paths per OD pair and of attributing a proportional share of the whole pair flow to each path in the previously selected set according to a Path Size Logit (PSL) model using observed travel times as key explanatory variable.

In order to exploit and elaborate available data, a script in R environment has been developed, implementing all procedural steps described ahead, except for the resolution of equation systems and results visualization, elaborated through Excel.

4.3.1. Workflow

As already mentioned before, estimation of OD matrices can be thought of as the inverse process of traffic assignment, being the fourth and last step in the framework of 4-stage traffic models.

For an area of study subdivided into n zones, Origin-Destination matrices describing traffic demand on the road network that connects such zones are tables made of n^2 cells. Under the hypothesis of no intrazonal trips, the problem of OD matrix estimation features $n^2 - n$ unknowns, provided that zones generate and attract trips at the same time.

Specifically for each time fraction k, linear equations eligible to solve the problem are identified through traffic counts y_k^m, which can be seen as link flows derived from the combination of the assignment matrix $\boldsymbol{A_k}$ and the vector $\boldsymbol{x_k}$ containing OD pairs:

$$A_k x_k = y_k \tag{4.1}$$

representing a stream of k systems of m linear equations in $n^2 - n$ unknowns can be either underdetermined, full-rank (in rare cases) or overdetermined: either way, solutions for each time fractions will be extracted solving an optimization problem due to inconsistency in the datasets.

In Figure 5, a flow chart summarizing the procedure from raw data to validated OD matrices is displayed:

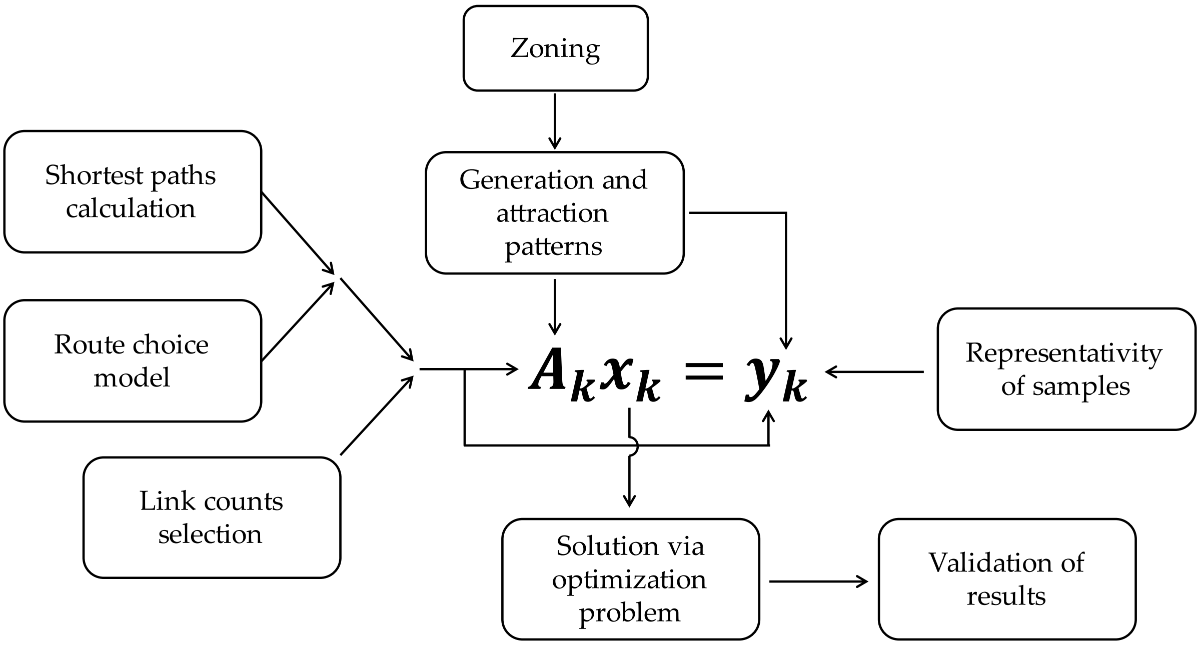

Zoning
Shortest paths calculation
Generation and attraction patterns
Route choice model
$A_k x_k = y_k$
Representativity of samples
Link counts selection
Solution via optimization problem
Validation of results

Figure 5. Workflow of OD matrix estimation procedure

4.3.2. Representativity of sample data

Representativity of input data for transport models are of evident importance, as transportation planners and mobility experts rely on the outcomes of these studies to advise decision makers regarding possible implications for traffic and mobility matters and issues.

Unfortunately, no official data are retrievable about coverage of TomTom data samples, so it is necessary to extract a consistent expansion coefficient per each time fraction k to be applied to available TomTom data. Previously elaborated mean traffic counts from video analytics have been used as comparison to calculate expansion coefficients. As k time fractions match in both studies, it is possible to expand daily average counts by using the ratio between corresponding sample sizes, so that a $C_{exp,k}$ for each period k (distinguishing between weekday and weekend) is calculated:

$$C_{exp,k} = \frac{TomTom\ sample\ size}{Video\ analytics\ counts} \tag{4.2}$$

In the next table, available counts from both datasets and the calculated expansion coefficient $C_{exp,k}$ are listed:

Table 2. Expansion coefficients ($C_{exp,k}$) calculation

Time fractions *k*	**Weekdays**			**Weekends**		
	Video analytics	**TomTom**	$C_{exp,k}$ **[%]**	**Video analytics**	**TomTom**	$C_{exp,k}$ **[%]**
7:00 - 7:15 am	7.87	0.350	4.45	3	0.125	4.17
7:15 - 7:30 am	12	0.650	5.42	2.67	0.25	9.37
7:30 - 7:45 am	13.07	0.625	4.78	4	0	0.00
7:45 - 8:00 am	18.93	0.450	2.38	4.5	0.375	8.33
8:00 - 8:15 am	25	0.850	3.40	5	0.3125	6.25
8:15 - 8:30 am	20.13	0.600	2.98	7.5	0.3125	4.17
8:30 - 8:45 am	16.73	0.500	2.99	7.25	0.5625	7.76
8:45 - 9:00 am	20.47	1.000	4.89	8	0.375	4.69
1:00 - 1:15 pm	12.5	0.475	3.80	11.75	0.3125	2.66
1:15 - 1:30 pm	13.71	0.650	4.74	14	0.5625	4.02
1:30 - 1:45 pm	14.85	0.900	6.06	10.5	0.5	4.76
1:45 - 2:00 pm	14.64	0.575	3.93	9.25	0.625	6.76
2:00 - 2:15 pm	18.85	0.725	3.84	12.5	0.375	3.00
2:15 - 2:30 pm	15.78	0.675	4.28	12	0.625	5.21
2:30 - 2:45 pm	15.85	0.750	4.73	10.75	0.5625	5.23
2:45 - 3:00 pm	13.21	0.550	4.16	11.5	0.4375	3.80
5:00 - 5:15 pm	16.85	0.975	5.79	15.25	0.75	4.92
5:15 - 5:30 pm	17.85	0.850	4.76	14.75	0.75	5.08
5:30 - 5:45 pm	17.45	0.925	5.30	15.75	0.625	3.97
5:45 - 6:00 pm	18.42	0.950	5.16	15.25	0.8125	5.33
6:00 - 6:15 pm	20.08	0.875	4.36	16.25	0.4375	2.69
6:15 - 6:30 pm	21.77	1.375	6.32	18	0.5	2.78
6:30 - 6:45 pm	20.46	1.025	5.01	16	0.875	5.47
6:45 - 7:00 pm	20.85	1.000	4.80	10.75	1.125	10.47

Mean		***4.51***		***5.04***
SD		***0.95***		***2.31***

On average, C_{exp} for weekdays is 4.51% (with standard deviation equal to 0.95%), while C_{exp} for weekends is 5.04% (with SD equal to 2.31%). While 4.51% corresponds to a value endorsed by other examples in literature using such type of data and features a low value of standard deviation, as weekends' standard deviation is relatively high, a consistent correlation between TomTom data and video analytics data is not possible, so the decision is not to use data collected on weekends, but to focus on weekdays data only for both trend analyses of vehicular patterns and the estimation of reliable OD matrices.

4.3.3. Zoning

Zoning is an essential requirement for the later application of the procedure: in this case, as the network is at neighbourhood scale, origin and destination zones are identified through access and exit streets, i.e., in correspondence with the intersections of the trimming polygon of the selected network (Figure 2) and the road graph of the city of Bologna. Such choice of zoning finds itself to be useful for the purpose of micro-simulating the road network once OD matrices are obtained, also avoiding that sample size values on minor local roads could be discarded in the elaboration.

Since the available network extension is much larger than the area which is supposedly affected by the tactical urbanism intervention and in order to ease the OD matrix estimation process by limiting the number of origin and destination zones, a reduction of the segments in the network is performed: via Stalingrado, on the eastern side of the selected area, is removed along with all the eastward afferent segments, as well as the tunnel denominated Asse Nord-Sud and the connection with via de' Carracci, at the southwestern corner of the network. In Figure 6, removed segments are highlighted in red colour.

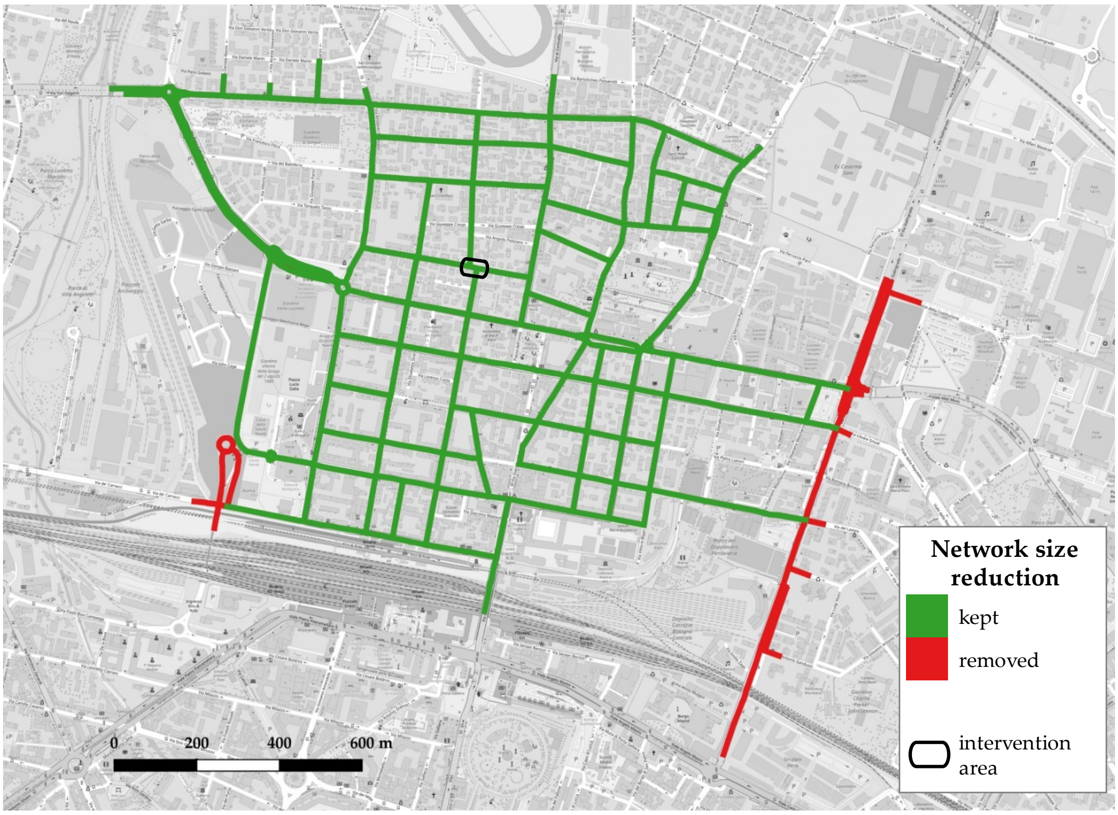

Network size reduction
kept
removed
intervention area
0
200
400
600 m

Figure 6. Network size reduction

In Table 3, the list of origin and destination zones of the edited road network and respective nodes denomination is displayed:

Table 3. Origin-Destination zones

Node id	Access/exit segment	O	D
382	Via de’ Carracci	•	•
304	Via Giacomo Matteotti	•	•
58	Via Ferrarese	•	•
37	Via di Corticella	•	•
370	Via Aristotile Fioravanti	•	•
5	Via Ezio Cesarini	•	
854	Via Alceste Giovannini		•
2	Via Daniele Manin	•	
735	Via Yuri Gagarin		•
496	Via Yuri Gagarin	•	
105	Via della Liberazione	•	
870	Via Donato Creti		•
896	Via Sebastiano Serlio		•

In Figure 7, a map showing the network highlights the position of origin and destination nodes:

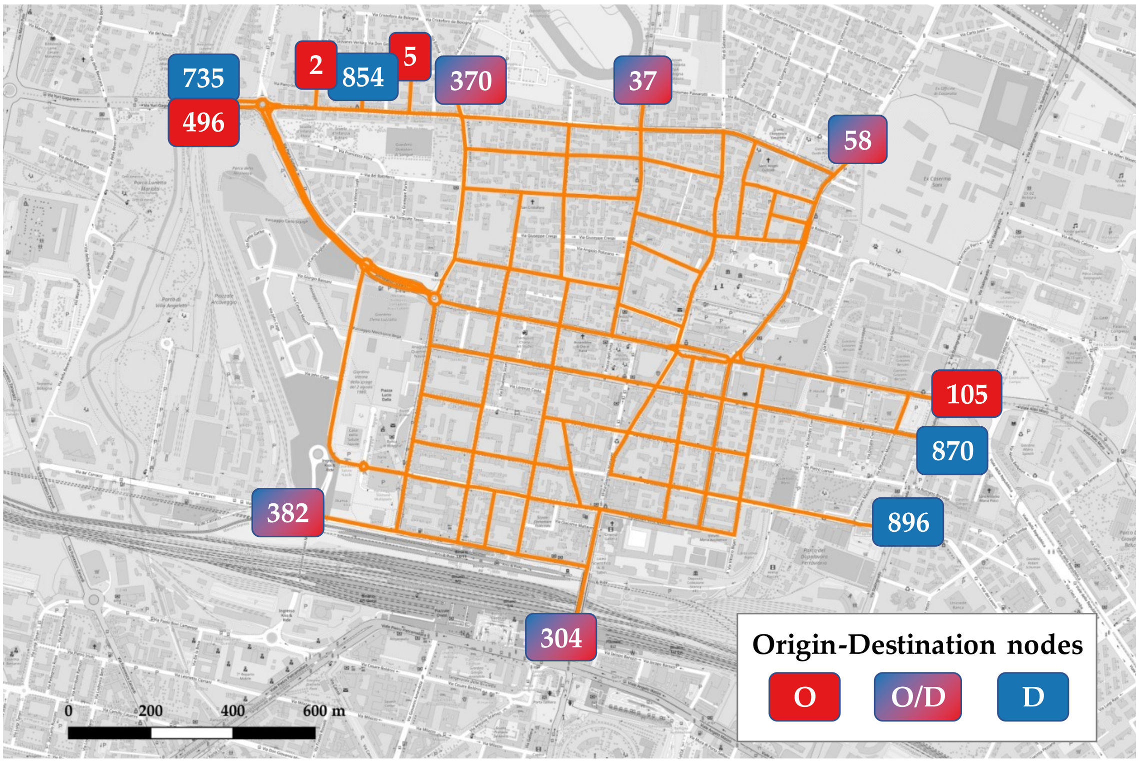

735
496
2
854
5
370
37
58
105
870
896
382
304
0
200
400
600 m
Origin-Destination nodes
O
O/D
D

Figure 7. Origin-Destination nodes

4.3.4. Generation and attraction

A set of input data of generated and attracted trip patterns in each time fraction of 15 minutes is necessary for the estimation procedure: as availability of sample size data interests each link on the selected network, generation and attraction patterns for OD zones are easily identified with corresponding sample data associated to access/exit segments. These data enforce two fundamental constraints for OD flows:

$$\sum_j x_{ijk} = P_{ik} \tag{4.3}$$

$$\sum_i x_{ijk} = A_{jk} \tag{4.4}$$

being x_{ijk} the demand for OD pair $i \rightarrow j$ of trips departing in time fraction k;

P_{ik} the sum of outbound OD flows from origin node i during time fraction k;

A_{jk} the sum of inbound OD flows to destination node j during time fraction k.

4.3.5. Shortest paths calculation

Calculation of shortest paths between origin and destination nodes through a weighted graph is a well-known, computationally demanding challenge in this kind of studies. For a portion of urban road network as the one under examination, characterized by a grid layout, calculating all simple paths (i.e., paths without loops and that do not contain the same node more than once), attributing costs according to chosen explanatory variables and selecting a set containing the most convenient paths, i.e., the ones featuring the minimum costs, can take days for a decent 64-bit calculator, so it is necessary to pursue a faster way.

For each OD pair, a limited set of possible paths is calculated: the chosen algorithm for this step is Yen's algorithm, proposed in Yen (1971), which computes the first $\bar{k}$ shortest loopless paths for a determined pair of nodes in an oriented graph with non-negative costs attributed to edges. It can employ any effective algorithm for the calculation of the shortest path, then proceeding to compute $\bar{k} - 1$ best deviations with the same algorithm used beforehand.

The underlying logic of the algorithm is as follows: once the first shortest path is computed, for each node i belonging to this path from the origin node, the best alternative way to the destination node is calculated by previously removing the edge $(i, i + 1)$ from the graph; all alternatives are compared regarding the total cost and the one featuring the lowest value is selected as best alternative; the algorithm continues to the next iteration until all $\bar{k}$ shortest paths for the selected OD pair are sorted.

For the case study, the algorithm was implemented in the R environment by the library *yenpathy*, basically transposing the algorithm script from C++; the shortest path computation algorithm adopted by the function in R is Dijkstra's.

A sensitivity test has been performed with the aim of assessing determinedness of equation systems resulted from the procedure, which stated that the most suitable number of possible shortest paths is $\bar{k} = 4$, given the extent of the network and the number of origin and destination zones resulted from the zoning process. Such choice of $\bar{k}$ would lead overall to slightly underdetermined problems, optimizing the trade-off between severe underdetermination resulting in inconsistent solutions and dispersion of flow on paths whose route choice proportionality is unrealistic.

4.3.6. Route choice

To attribute route choice proportionality to each path in the calculated set, a Path Size Logit model is used: it consists of a Multinomial Logit with travel time as key explanatory variable, with the addition of penalties for paths that share segments with others in the same set. The main reason for penalizing paths containing shared links is the consideration that the user tends not to recognize one path as a valid alternative if most of the segments are shared between two different choices.

As presented in Prato (2009), the following PSL model is assumed:

$$p_{ijk}^{n} = \frac{e^{(V_n + \beta^{PS} \cdot \ln PS_n)}}{\sum_{r_{ijk}^{n} \in P_{ijk}} e^{(V_r + \beta^{PS} \cdot \ln PS_r)}} \tag{4.5}$$

with p_{ijk}^{n} route choice proportional factor of path n;

$V_n = -C_n = -{TT_n}^k$;

β^{PS} path size parameter to be estimated;

PS_n path size factor of path n.

The disutility function V_n expresses how costs negatively affect attractiveness of the choice of path n through explanatory variables, travel time ${TT_n}^k$ only in this example. In the occurrence of segments featuring null travel time for the time fraction k, indicating that no vehicle has been recorded on the segment, the algorithm automatically converts the attribute value for travel time into 1000 minutes to exclude such segments from the computation of convenient routes. Lacking useful data for calibration (usually provided by interviews), β^{PS} is assumed equal to 1, accordingly to the formulation of the Logit model proposed by Krishnakumari et al. (2019). Path sizes are calculated through the original formulation proposed by Ben-Akiva et al. (1999):

$$PS_n = \sum_{a \in r_{ijk}^n} \frac{L_a}{L_n} \frac{1}{\sum_{r_{ijk}^n \in P_{ijk}} \delta_{an}} \tag{4.6}$$

where L_a is the length of link a;

L_n is the length of path n;

δ_{an} is an integer variable expressing the link-path incidence, i.e., the number of paths in the set in which link a features.

It follows that path flows are directly related to OD flows by each proportionality factor:

$$x_{ijk}^n = p_{ijk}^n x_{ijk} \qquad \forall\, i, j, k; n = 1 \dots N_{ijk}^* \tag{4.7}$$

4.3.7. Link traffic counts selection

Once production and attraction patterns are determined and path sets together with relative choice proportionality are assumed as described beforehand, an additional and essential constraint on OD flows is provided by link traffic counts $\tilde{y}_k = C_{exp,k} y_k$, previously increased by expansion coefficients calculated in section 4.3.1, which represent the main piece of information for traffic demand.

A link traffic count provides information on OD pairs in the form of these equations:

$$\widetilde{y_k^a} = \sum_{r_{ijk}^n \in P_k^a} x_{ijk}^n = \sum_{r_{ijk}^n \in P_k^a} p_{ijk}^n x_{ijk} \tag{4.8}$$

saying that the expanded traffic count on link a in time fraction k is the sum of the path flows x_{ijk}^n corresponding to all the paths contained in the set $P_k^a = \{r_{ijk}^n \mid a \in r_{ijk}^n\}$.

If the whole set of traffic counts were to be used, the linear system for the determination of OD pairs for each time fraction k would feature equations (4.3), (4.4) and (4.8):

$$\begin{cases} x_{i1k} + \cdots + x_{ijk} + \cdots = P_{ik} \\ x_{1jk} + \cdots + x_{ijk} + \cdots = A_{jk} \\ p_{11k}^{n} x_{11k} + \cdots + p_{ijk}^{n} x_{ijk} + \cdots = \widetilde{y_k^a} \end{cases} \tag{4.9}$$

However, not every configuration of traffic counts can lead to a satisfying solution of the OD estimation problem. For instance, in (4.9) coefficient matrix $\boldsymbol{A_k}$ is surely a singular matrix because of dependence among link traffic counts. As stated in Ortúzar and Willesden (2011) and Espitia Echeverría (2018), traffic counting suffers from two fundamental problems:

- data inconsistency
  Traffic counts are affected from multiple errors caused by intrinsic flaws in the measurement procedure, leading to inconsistent flows attributed to each link; this means that even if the equation system is full rank, there might be no matrix able to satisfy each value of the observed flow vector; it is thus necessary to consider a vector of errors affecting the set of observed flows.
- data dependence
  Traffic is modelled over the road network under the hypothesis of flow continuity at nodes, meaning that difference between inbound and outbound flows must be null; thus, flow on one of the segments related to each node is deductible as linear combination of the others, leading to linearly dependent equations which would not bring any additional information.

Usually, traffic counts campaigns with limited availability of sensors priorly consider such problems and outline a configuration of counting spots which leads to an independent dataset. As, for the case under examination, link counts are available for each segment for each time fractions in the daily time slots, a criterion for selecting a limited number of independent equations must be assumed.

Yang et al. (1998) proposed four basic rules for the selection of optimal locations for link counters:

1. **OD pair coverage**: a minimum share of demand flow per OD pair should be observed;
2. **Maximum flow fraction**: for each OD pair, the link with the greatest demand flow fraction between the pair should be observed;

3. **Maximum coverage**: in the case of limitations of number of available counters, the greatest number possible of OD pairs is to be observed;
4. **Link independence**: traffic counts should not be linearly dependent.

On the basis of such rules, localization methods for traffic counters are developed. Most methods presented in literature proceed from the assumption that a prior OD matrix is available so, by assigning this particular matrix to the network, it is possible to obtain information over used paths between OD pairs and link flows, which constitute input data for localization methods.

In this case, no prior matrix is available: however, information on paths is in fact available because of the assumptions formulated in sections 4.3.5 and 4.3.6 (implications of such premises on the accuracy of the results will be discussed later), so the same methods could be applied to the current case.

The proposed method works as follows:

1. equations (4.3) and (4.4) are chosen as first equations of the linear system, being “privileged” traffic counts providing information of production and attraction patterns; though, it is to remark that one equation must be discarded because it is a linear combination of all the others in the set;
2. for each link in the network, excluding access/exit segments, which have been already considered in the previous step, it is computed how many OD pairs have at least one path transiting through the link itself;
3. starting from the link with the highest number of transiting OD, the corresponding matrix row is added, and a rank check is performed: if the rank of the matrix at this step is lower than the rank of the matrix with the added row, the row is declared linearly independent and the corresponding equation is definitely added to the solving equation system.

The procedure is repeated for all the available traffic counts which, in the current time fraction, have recorded a value greater than 0; all discarded rows because linearly dependent are placed in a separate container, which will be used to validate resulting OD pair flows once calculated.

In summary, this algorithm seeks to represent as many OD pairs as possible while ensuring that no redundant information is considered, solving the issue of data dependence. Of course, the number of considered possible shortest paths is relevant to the determinedness of the resulting equation system: by increasing the number $\bar{k}$ of considered paths between OD pairs, flows would be dispersed over unrealistically convenient routes, so the benefit of handling more determined equation systems allegedly leading to more consistent solutions would be counterproductive, if not detrimental for the accuracy of final outcomes.

4.3.8. Equation system solving

At the end of the previous step, a system of linear equations is obtained for each time fraction k:

$$\boldsymbol{A_k x_k} = \boldsymbol{\tilde{y}_k} \tag{4.10}$$

in which $\boldsymbol{A_k x_k}$ represent the vector of estimated link flows $\boldsymbol{\hat{y}_k}$ and $\boldsymbol{\tilde{y}_k}$ stands for the vector of observed flows multiplied by the corresponding expansion coefficient; in order to avoid meaningless solutions featuring negative OD flows, the vector of unknows, consisting of OD flows in time fraction k, has a non-negativity constraint: $\boldsymbol{x_k} \geq 0$.

This equation system can be either underdetermined, full-rank or overdetermined, as the coefficient matrix may not be a square matrix. As a result of the choice of $\bar{k}$, equation systems both from *ex-ante* and *ex-post* situations are slightly underdetermined, with a few cases of full-rank condition.

Nevertheless, the problem raised before about data inconsistency is yet to be confronted. Assumptions on paths and route choice also affect accuracy of the solution.

To address this issue, it is necessary to consider a vector of errors summarizing measurement errors affecting observed flows, assumption errors and flaws introduced with the proposed model:

$$\boldsymbol{A_k x_k} = \boldsymbol{\tilde{y}_k} + \boldsymbol{\varepsilon_k} \tag{4.11}$$

With the introduction of vector $\boldsymbol{\varepsilon_k}$, independently from the determinedness of each linear system, the solution has to be searched through solving an optimization problem: the objective function $Z(\boldsymbol{x})$ (4.12) to be minimized is the norm of vector of errors, which from (4.11) it corresponds to the difference vector between estimated flows and observed flows:

$$Z(\boldsymbol{x_k}) = \|\boldsymbol{\varepsilon_k}\|_2 = \|\boldsymbol{A_k x_k} - \widetilde{\boldsymbol{y}}_{\boldsymbol{k}}\|_2 = \|\widehat{\boldsymbol{y}}_{\boldsymbol{k}} - \widetilde{\boldsymbol{y}}_{\boldsymbol{k}}\|_2 \tag{4.12}$$

To solve this optimization problem, the nonlinear Generalized Reduced Gradient (GRG) embedded in the Excel plugin Solve is employed. It consists of an iterative algorithm which calculates at each step the gradient of the objective function: starting from a trial solution $\boldsymbol{x}_0$, it computes numerically partial derivatives to form the gradient and takes a further step in the desired direction, e.g., for minimization problems the step in the most negative direction, in order to reach the optimal solution as fast as possible, which is considered to be the step in which gradient is the null vector. It is indicated for nonlinear "smooth" problems, so suitable for functions represented by regular hypersurfaces featuring continuous gradients in every direction.

As it needs a first trial solution as starting point, it could be tricked into recognizing local optima as global solutions: in order to partially overcome this limitation, it is possible to select from Solve plugin settings the option *Multistart*, which reduces the risk of running into solutions situated in local optima by repeating the procedure starting from different trail solutions, but obviously increasing computational time.

4.3.9. Validation of results

Once the Solve plugin has reached a solution, a definitive vector $\boldsymbol{x_k}$ containing OD flows is extracted and after proper rearrangements, OD matrices are available with the correct table layout.

To validate the process, linear regression of considered and discarded link flows between estimations and observations are performed. The aim of this step is to assess the level of correlation between the two sets of link flows, in order to evaluate the accuracy of the determination of OD flows.

Linear regression basically returns a linear function $y = mx + y_0$ which interpolates two paired sets of values from variables or, like in this case, the observed values of a variable $\tilde{y}$ and the corresponding estimations $\hat{y}$. As a matter of fact, the two sets can be related through the linear function considering also an error term ε:

$$\hat{y} = m\tilde{y} + y_0 + \varepsilon \tag{4.13}$$

The coefficient of multiple correlation, indicated by R, is defined as Pearson's correlation coefficient between the predicted and the observed values of a variable; in other words, it is a measure of how good a variable can be predicted through models based on linear relations of a set of other variables. The key parameter is the coefficient of determination R^2 (R-squared): it is the proportion of how common variance between independent and dependent variables is, so it is used in regression models to evaluate how much of the variance of one variable is explained in the other variable's variance. R^2 may vary between 0 and 1: a high value ($R^2 > 0.7$) means that variations in one variable mostly depend on the variations in the other variable; a low value ($R^2 < 0.3$) shows that only a small portion of values in one variable is explained by the other, so the two sets are not well correlated.

Correlation coefficient r often overestimates relationship between variables, especially when employed samples are not numerous, while coefficient of determination tends to be more accurate. Furthermore, R^2 is usually well defined no matter the nature of the independent variable, if random or fixed.

In case R^2 drops under the threshold of 0.7, Theil's inequality coefficient U will be computed; although primarily used to evaluate economic inequality, it was proposed by Toledo et al. (2004) to statistically validate traffic simulation models. A useful property of this coefficient is that it can overcome the effect of outliers (Barceló et al., 2010), as opposed to usually adopted RMS estimators. U is calculated as follows:

$$U = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - \tilde{y}_i)^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i)^2} + \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\tilde{y}_i)^2}} \tag{4.14}$$

with n being the number of elements in both the observed and estimated data;

$\hat{y}_i$ the estimations;

$\tilde{y}_i$ the observations.

U is bounded between the values of 0 and 1: $U = 0$ means that measurements and estimations fit perfectly; $U = 1$ depicts an unacceptable fit. As suggested by

Barceló et al. (2010), predicted series featuring $U > 0.2$ should be disregarded because of their inaccurate character.

Theil's inequality coefficient can be decomposed in three proportions: U_M (bias), U_S (variance) and U_C (covariance). Their expressions are listed below:

$$U_M = \frac{\left(\bar{\hat{y}}_\imath - \bar{\tilde{y}}_\imath\right)^2}{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - \tilde{y}_i)^2} \tag{4.15}$$

$$U_S = \frac{(\hat{\sigma} - \tilde{\sigma})^2}{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - \tilde{y}_i)^2} \tag{4.16}$$

$$U_C = \frac{2(1 - r)\hat{\sigma}\tilde{\sigma}}{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - \tilde{y}_i)^2} \tag{4.17}$$

where $\bar{\hat{y}}_\imath$ and $\bar{\tilde{y}}_\imath$ are means, $\hat{\sigma}$ and $\tilde{\sigma}$ standard deviations of estimations and observations respectively; r the Pearson's correlation coefficient between the two sets.

In particular, U_M gives information about the systematic error, U_S tells how well the variability is replicated by the model while U_C informs about the non-systematic error. Because of how they are constructed, the sum of all three proportions must return 1: the indication of a good fit lies in U_M and U_S being as small as they can get (in return, U_C should be as close to 1 as possible).

5 Results

The procedure described in the previous section has been implemented using both RStudio and Microsoft Excel; in the current section obtained results will be displayed and commented.

5.1. Statistical analysis

5.1.1. T-tests

In the first place, t-tests have been performed comparing datasets by date ranges (weekdays and weekends) and groups of Functional Road Classes (FRC 2-3 and FRC 4-6). In the tables below, the investigated values of the datasets (average speed, sample size and speed 85th percentile) have been condensed in peak time slots through arithmetical average among all the segments belonging to the cluster. The tables feature mean and standard deviation of both *ex-ante* and *ex-post* values, p-values resulted from the statistical test and the difference in means ΔM.

Table 4. Results of t-tests for average speed values

Average speed	***ex-ante***		***ex-post***		**p-value**	**ΔM**	
	M	***SD***	***M***	***SD***			
weekdays	30.82	9.77	29.74	9.29	7.06E-29	-1.08***	*7-9am*
	31.55	9.62	30.77	9.17	2.04E-19	-0.77***	*1-3pm*
	28.13	9.59	27.24	9.11	2.78E-18	-0.89***	*5-7pm*
weekends	34.40	11.49	33.31	11.15	3.07E-12	-1.09***	*7-9am*
	32.58	10.06	31.73	9.67	6.03E-15	-0.86***	*1-3pm*
	30.55	9.65	29.07	9.13	6.80E-25	-1.49***	*5-7pm*

FRC 2-3	37.48	10.40	36.73	9.79	2.10E-11	-0.76***	*7-9am*
	36.61	9.38	36.14	9.03	2.55E-07	-0.48***	*1-3pm*
	32.84	9.20	32.18	9.20	7.60E-08	-0.66***	*5-7pm*
FRC 4-6	30.38	10.25	29.15	9.81	1.53E-23	-1.24***	*7-9am*
	29.98	9.35	29.01	8.75	6.09E-26	-0.97***	*1-3pm*
	27.74	9.27	26.31	8.54	4.50E-35	-1.43***	*5-7pm*
weekdays FRC 2-3	34.29	9.62	33.64	9.30	9.43E-09	-0.65***	*7-9am*
	35.75	9.38	35.18	8.90	1.11E-05	-0.57***	*1-3pm*
	31.01	9.63	30.51	9.35	1.06E-03	-0.50**	*5-7pm*
weekdays FRC 4-6	29.23	9.43	27.96	8.73	1.73E-22	-1.27***	*7-9am*
	29.62	9.10	28.75	8.58	3.57E-15	-0.87***	*1-3pm*
	26.81	8.60	25.74	8.60	3.59E-16	-1.07***	*5-7pm*
weekends FRC 2-3	40.68	10.20	39.82	9.29	1.03E-05	-0.86***	*7-9am*
	37.48	9.32	37.09	9.08	3.69E-03	-0.38**	*1-3pm*
	34.68	9.43	33.85	8.75	2.04E-05	-0.83***	*5-7pm*
weekends FRC 4-6	31.53	10.90	30.34	10.67	1.22E-08	-1.20***	*7-9am*
	30.34	9.59	29.27	8.92	3.73E-13	-1.07***	*1-3pm*
	28.67	9.17	26.88	8.45	4.63E-21	-1.79***	*5-7pm*

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 5. Results of t-tests for sample size values

Sample size	***ex-ante***		***ex-post***		**p**	**ΔM**	
	M	***SD***	***M***	***SD***			
weekdays	68.46	66.58	72.62	68.70	4.60E-55	4.16***	*7-9am*
	61.52	60.89	64.88	64.01	2.85E-51	3.36***	*1-3pm*
	85.05	77.51	86.12	77.18	9.73E-10	1.07***	*5-7pm*
weekends	22.38	23.58	25.38	25.68	3.41E-86	3.00***	*7-9am*
	51.78	57.68	56.07	60.01	7.13E-83	4.28***	*1-3pm*
	69.95	75.79	72.39	73.65	5.35E-21	2.44***	*5-7pm*

FRC 2-3	86.03	70.39	92.47	71.29	1.29E-62	6.44***	*7-9am*
	108.91	70.02	116.67	71.76	1.57E-91	7.76***	*1-3pm*
	148.84	80.84	149.88	80.84	3.50E-03	1.04**	*5-7pm*
FRC 4-6	26.83	32.30	29.10	33.88	4.32E-74	2.27***	*7-9am*
	32.73	33.08	34.74	34.24	3.64E-62	2.01***	*1-3pm*
	44.84	44.87	46.92	45.18	1.39E-38	2.09***	*5-7pm*
weekdays FRC 2-3	129.38	73.37	137.12	73.16	1.20E-28	7.74***	*7-9am*
	115.74	68.58	123.86	70.84	2.00E-45	8.12***	*1-3pm*
	157.91	77.75	159.86	76.07	3.96E-06	1.95***	*5-7pm*
weekdays FRC 4-6	40.57	39.09	43.10	40.73	3.42E-38	2.52***	*7-9am*
	36.70	36.00	37.88	36.85	4.11E-21	1.18***	*1-3pm*
	51.69	49.11	52.36	49.11	5.36E-05	0.67***	*5-7pm*
weekends FRC 2-3	42.68	27.80	47.82	28.80	2.71E-61	5.14***	*7-9am*
	102.08	70.95	109.49	72.12	2.57E-48	7.40***	*1-3pm*
	139.78	89.24	139.90	84.35	8.24E-01	0.13	*5-7pm*
weekends FRC 4-6	13.08	13.52	15.10	15.70	9.00E-39	2.02***	*7-9am*
	28.76	29.39	31.61	31.13	1.04E-44	2.85***	*1-3pm*
	37.98	38.51	41.48	40.21	1.29E-39	3.50***	*5-7pm*

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 6. Results of t-tests performed on speed 85th percentile values

Speed – 85th percentile	*ex-ante*		*ex-post*		**p**	**ΔM**	
	M	*SD*	*M*	*SD*			
weekdays	41.98	11.93	40.41	10.79	1.14E-14	-1.57***	*7-9am*
	42.78	11.69	41.72	10.67	3.40E-09	-1.06***	*1-3pm*
	39.10	11.95	37.67	10.66	3.21E-12	-1.43***	*5-7pm*
weekends	45.83	14.56	44.02	13.71	1.71E-12	-1.80***	*7-9am*
	44.04	12.78	42.44	11.55	3.45E-11	-1.59***	*1-3pm*
	41.18	12.04	39.65	10.80	9.39E-10	-1.53***	*5-7pm*

FRC 2-3	50.28	11.46	49.14	10.59	1.66E-11	-1.14***	*7-9am*
	48.92	10.43	48.22	9.75	2.52E-06	-0.70***	*1-3pm*
	44.54	10.19	43.94	10.19	7.06E-05	-0.60***	*5-7pm*
FRC 4-6	40.99	13.28	39.05	11.97	6.94E-18	-1.94***	*7-9am*
	40.89	12.21	39.27	10.57	7.45E-15	-1.62***	*1-3pm*
	38.13	12.04	36.24	10.15	5.18E-17	-1.89***	*5-7pm*
weekdays FRC 2-3	46.39	10.55	45.61	10.23	9.51E-11	-0.78***	*7-9am*
	47.84	10.46	47.25	9.73	5.41E-04	-0.59***	*1-3pm*
	42.69	10.93	42.14	10.44	1.48E-03	-0.55**	*5-7pm*
weekdays FRC 4-6	39.97	12.00	38.03	10.19	2.82E-11	-1.94***	*7-9am*
	40.46	11.50	39.18	10.12	3.18E-07	-1.28***	*1-3pm*
	37.46	10.13	35.62	10.13	1.85E-10	-1.84***	*5-7pm*
weekends FRC 2-3	54.18	11.03	52.67	9.76	1.93E-06	-1.51***	*7-9am*
	50.00	10.30	49.19	9.70	9.22E-04	-0.81***	*1-3pm*
	46.39	10.39	45.74	9.63	8.78E-03	-0.65**	*5-7pm*
weekends FRC 4-6	42.00	14.39	40.07	13.45	1.68E-08	-1.94***	*7-9am*
	41.31	12.88	39.35	11.01	4.01E-09	-1.95***	*1-3pm*
	38.79	12.01	36.86	10.15	2.27E-08	-1.93***	*5-7pm*

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

The null hypothesis stating that difference in means is zero is rejected in nearly all the tests that have been performed, under the level of probabilistic significance level $\alpha = 0.001$. A few tests have returned p-values slightly greater than 0.001 but under the threshold value of $\alpha = 0.05$. The t-test comparing sample sizes in *ex-ante* and *ex-post* datasets for FRC 2-3 segments at weekends in the time slot 5:00 - 7:00 PM returned a very high p-value (0.82), so for this specific test the null hypothesis could not be rejected under any reasonable level of probabilistic significance.

On the whole, the available datasets which have been compared from a statistical significance point of view through t-tests and later by analysing trends have demonstrated significant differences, so worthwhile to be investigated for mobility patterns variations. However, since from calculating expansion coefficients it was noticed that weekends data could present issues regarding the representativity of real traffic flow data, only values recorded on weekdays have been exploited in the next sections.

Being given the green light from t-tests to analyse trends in traffic-related parameters, speed values have been analysed between the FRCs clusters, while sample sizes have been treated under the buffer segments classification (intervention, adjacent, buffer and control).

In the following pages, histograms representing flow trends per each time fractions both for *ex-ante* and *ex-post* situations are reported.

5.1.2. Trend analysis

5.1.2.1. Speed values

In Figure 8 and Figure 9, results from trend analyses of average speed on weekdays are shown:

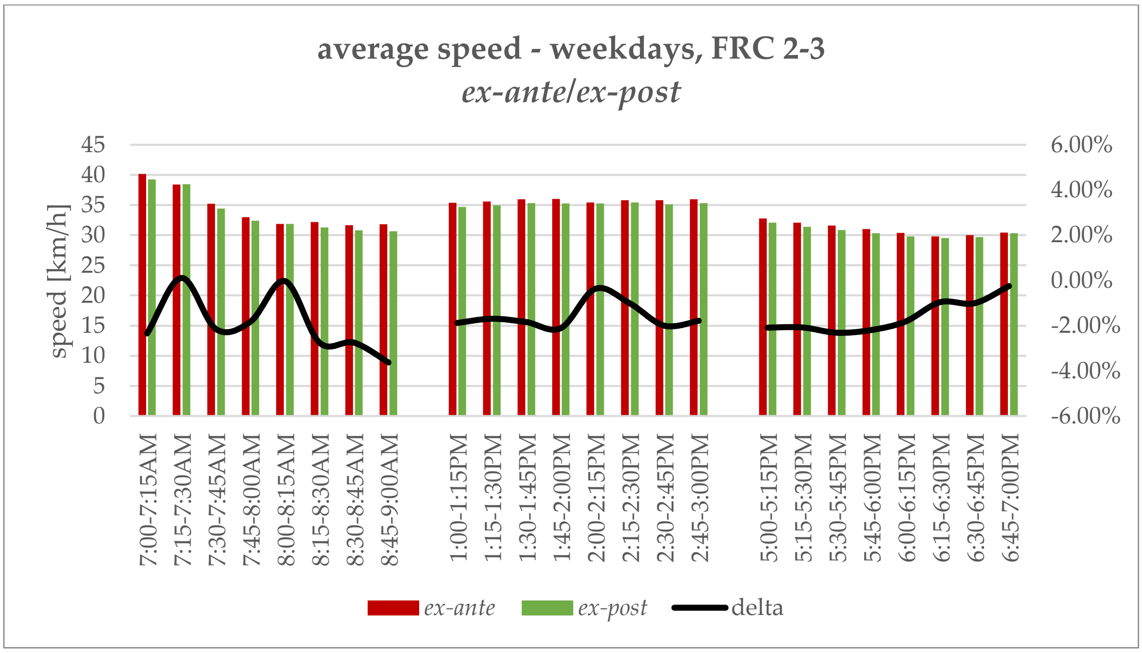


Figure 8. Average speed on weekdays for FRC 2-3 segments - comparison *ex-ante/ex-post*

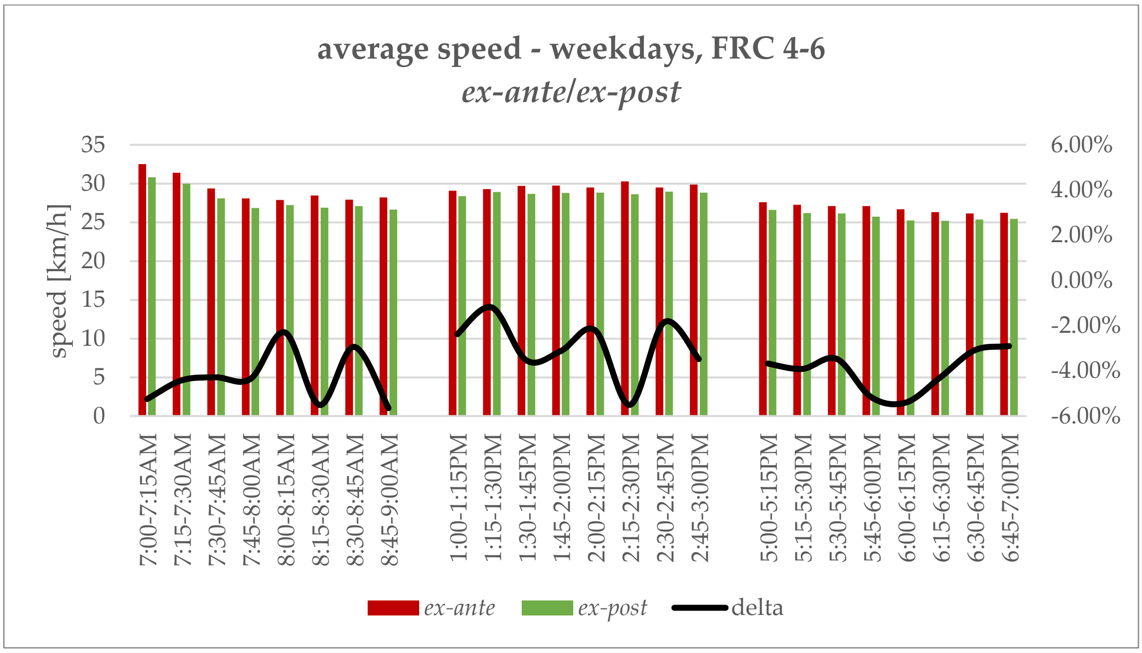


Figure 9. Average speed on weekdays for FRC 4-6 segments - comparison *ex-ante/ex-post*

When comparing *ex-ante* and *ex-post* situations, the plots outline a slight decrease in the average value of speed both for FRC 2-3 and FRC 4-6 segments of approximately -1% and -3% respectively.

Nonetheless, as average speed returns a global representation of traffic behaviour and does not provide useful indications about the range of speed variation in the sample as well as the number of vehicles which it is referring to, it is advisable to perform a critical analysis on 85th percentile speed values, widely recognized as a good referral for speed limit imposition since it is indicative for a safe driving behaviour (Ministero dei Trasporti, 2006). Plots in Figure 10 and Figure 11 show 85th percentile speed values trends:

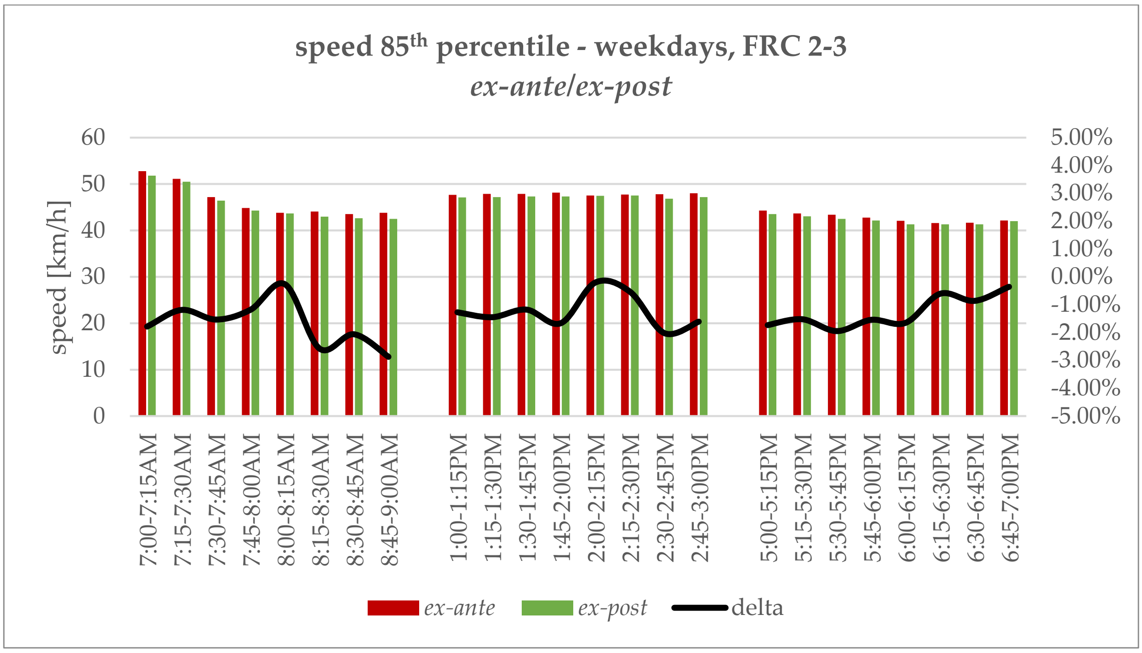


Figure 10. Average speed 85th percentile on weekdays for FRC 2-3 segments - comparison *ex-ante/ex-post*

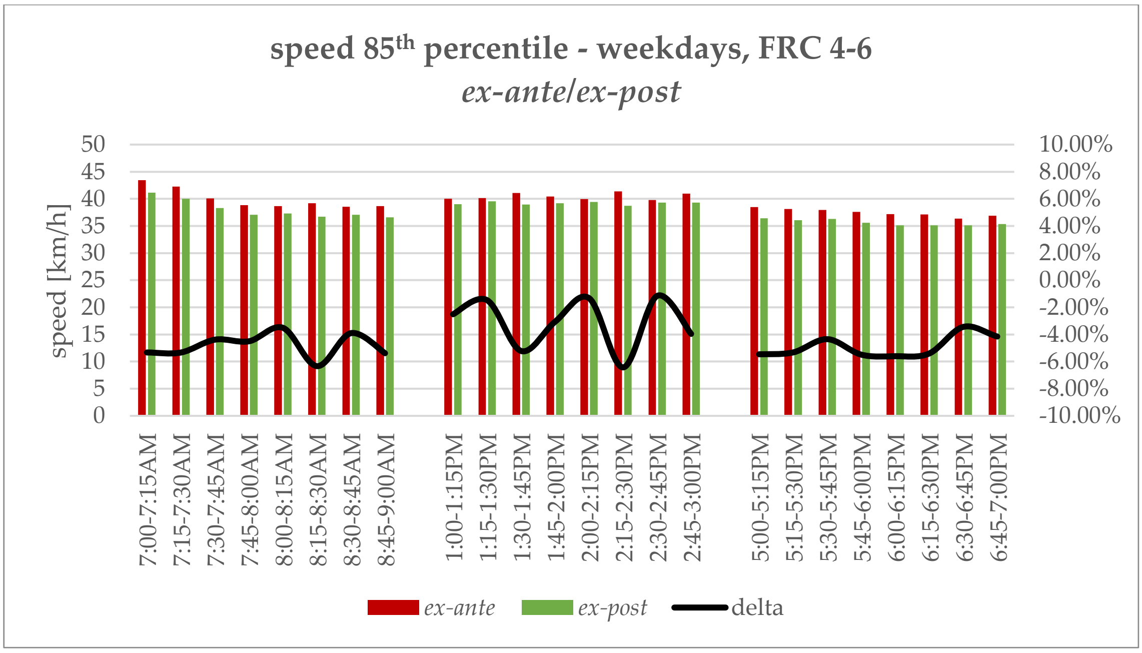


Figure 11. Average speed 85th percentile on weekdays for FRC 4-6 segments - comparison *ex-ante/ex-post*

Speed 85th percentiles follow the same trend as average speed: this is presumably due to strong similarities in the distributions of speed values over time fractions.

Referring to recommended procedures for imposing speed limit and to the new trend of implementing 30 km/h zones in urban areas, as the representative lower bound for 85th speed percentile for FRC 4-6 segments extracted from the trend analysis is 35 km/h, the extensive application of the speed limit of 30 km/h zones would be too restrictive in terms of acceptance from the users, with the risk of general incompliance, let alone on FRC 2-3 streets, whose 85th percentile of speed value is 45 km/h on average.

5.1.2.2. Sample size values

In order to assess variations in traffic flow values, the investigated parameter is the sample size adapted to actual values through expansion coefficients in Table 2 to represent real traffic flow, averaged over the entire cluster. For this analysis, road segments have been clustered according to Nello-Deakin (2022) classification regarding the position of the intervention that supposedly triggered the mobility patterns to change.

In the next figures, plots displaying results from the trend analyses are presented. In Figure 12, average traffic flow on weekends for intervention streets is displayed:

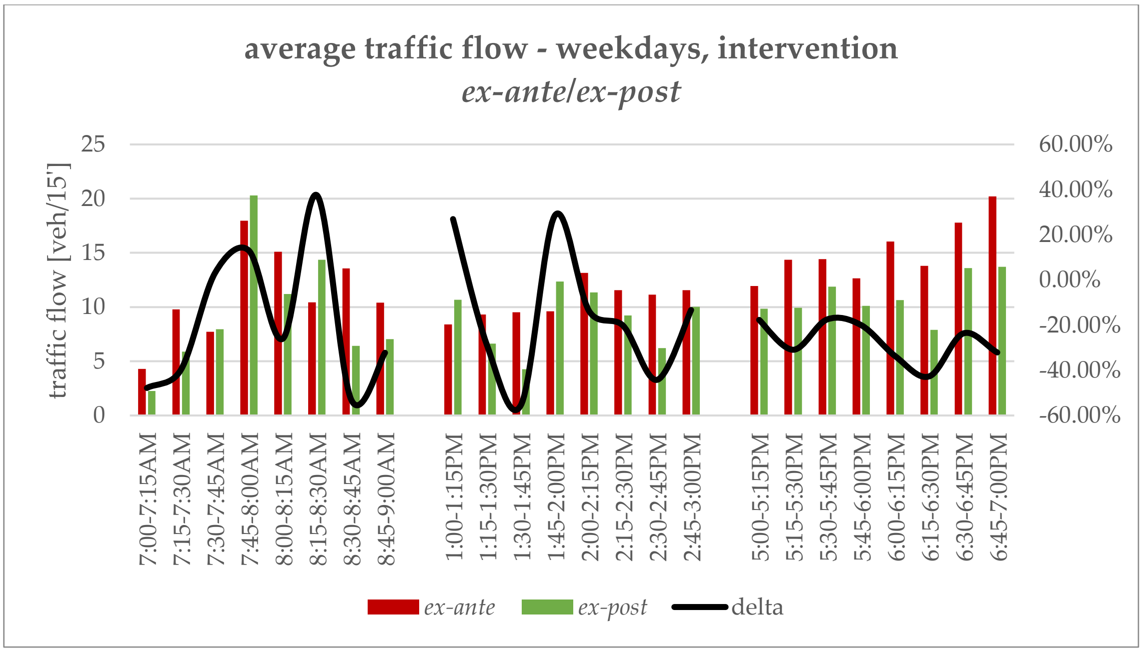


Figure 12. Average traffic flow on weekdays on intervention streets - comparison *ex-ante/ex-post*

Intervention streets variation trend is characterized by considerable peaks mostly due to little flows interesting such segments, making it more susceptible to high relative variability. From Figure 12 it is clearly visible that in time slot 5:00-7:00 PM traffic flow experienced a reduction of approximately -30%, while the other time slots do not display as clear reduction in flows as the evening slot, being those subject to high fluctuations.

Following plots show average traffic flow trends for adjacent (Figure 13), buffer (Figure 14) and control segment class (Figure 15).

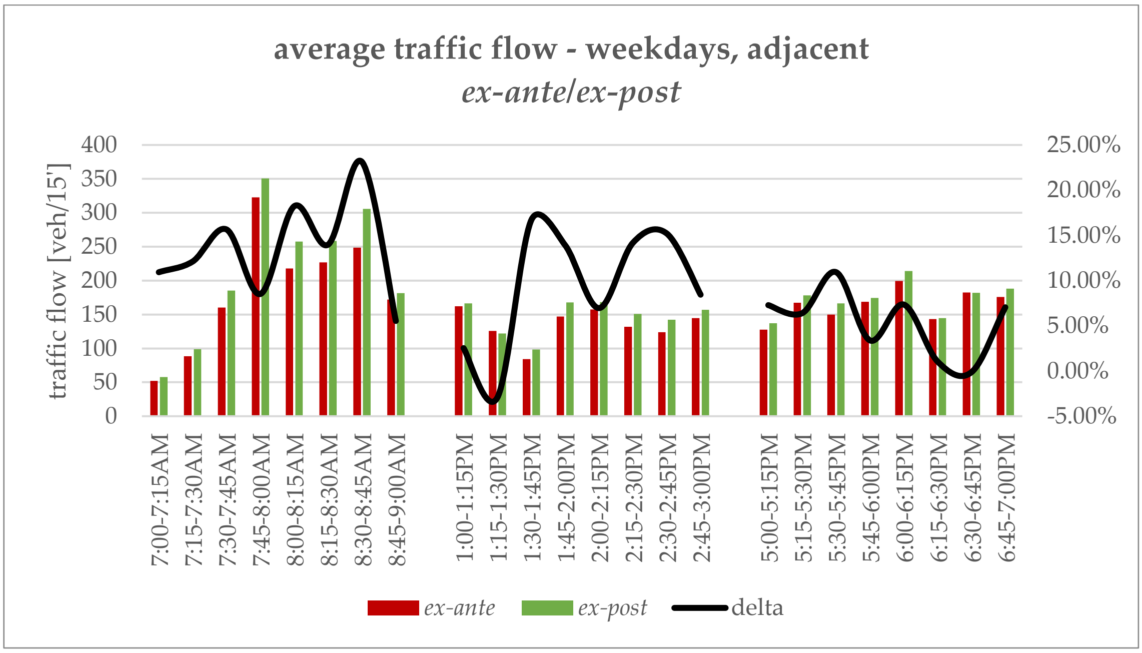


Figure 13. Average traffic flow on weekdays on adjacent streets - comparison *ex-ante/ex-post*

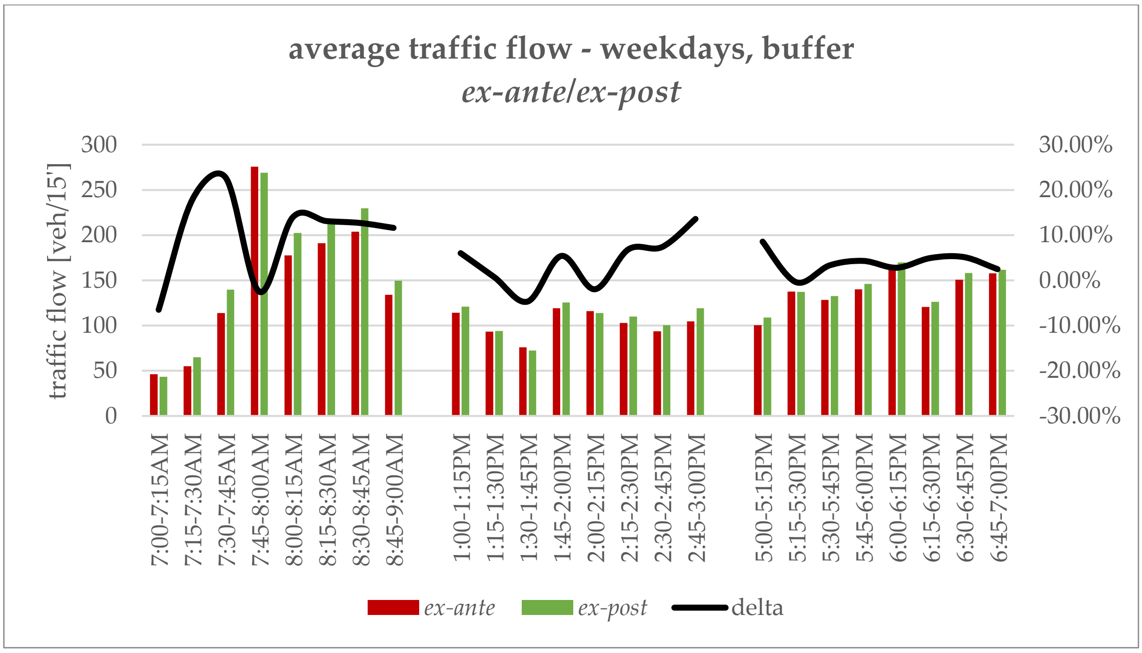


Figure 14. Average traffic flow on weekdays on buffer streets - comparison *ex-ante/ex-post*

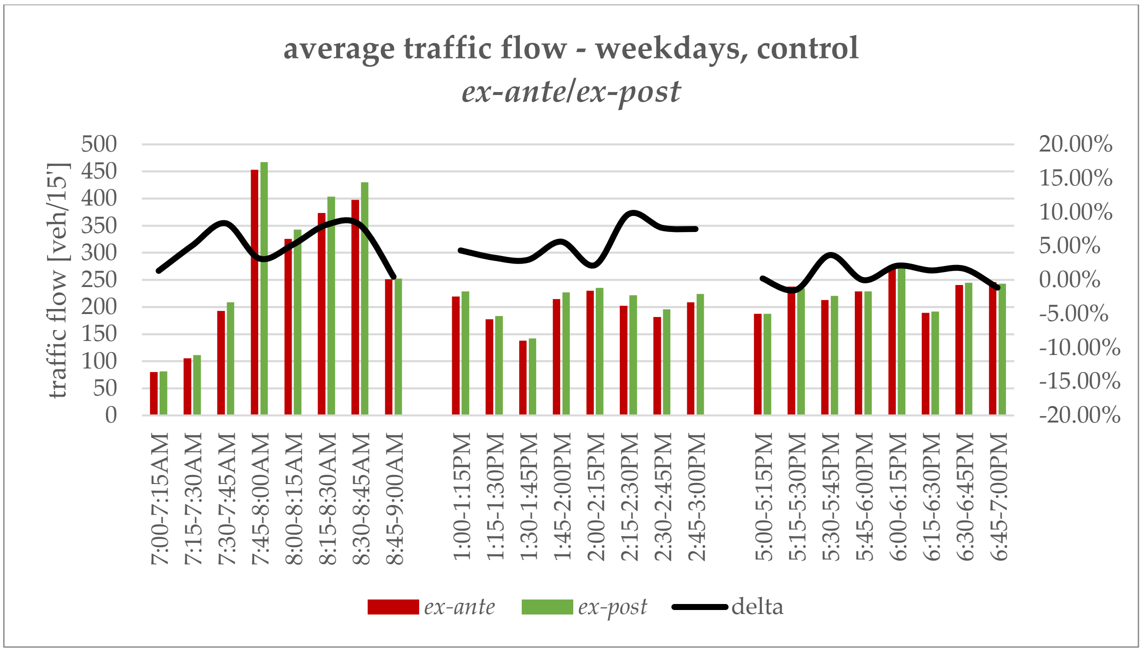


Figure 15. Average traffic flow on weekdays on control streets - comparison *ex-ante/ex-post*

From the plots before shown, it is clear that, whilst traffic flow appears to be decreased for intervention streets between *ex-ante* and *ex-post* scenarios, the other clusters in the buffer classification seem to be in countertrend, with positive delta values in most of time fractions.

Of course, traffic levels in a neighbourhood may have changed because of other factors than the temporary school square implementation within a whole year. In order to understand if and how traffic patterns have changed because of the intervention, in Figure 16 the idea is to adjust buffer classes variations over control segments fluctuations, with the aim of cancelling the noise sourced from other events that may have affected other zones of the selected road network. Adjusted relative variations have been computed as the difference between delta values of the buffer classes and the ones for the control class in the same time fraction k:

$$\Delta_{(i,a,b),k}^{adj} = \Delta_{(i,a,b),k} - \Delta_{c,k} \tag{5.1}$$

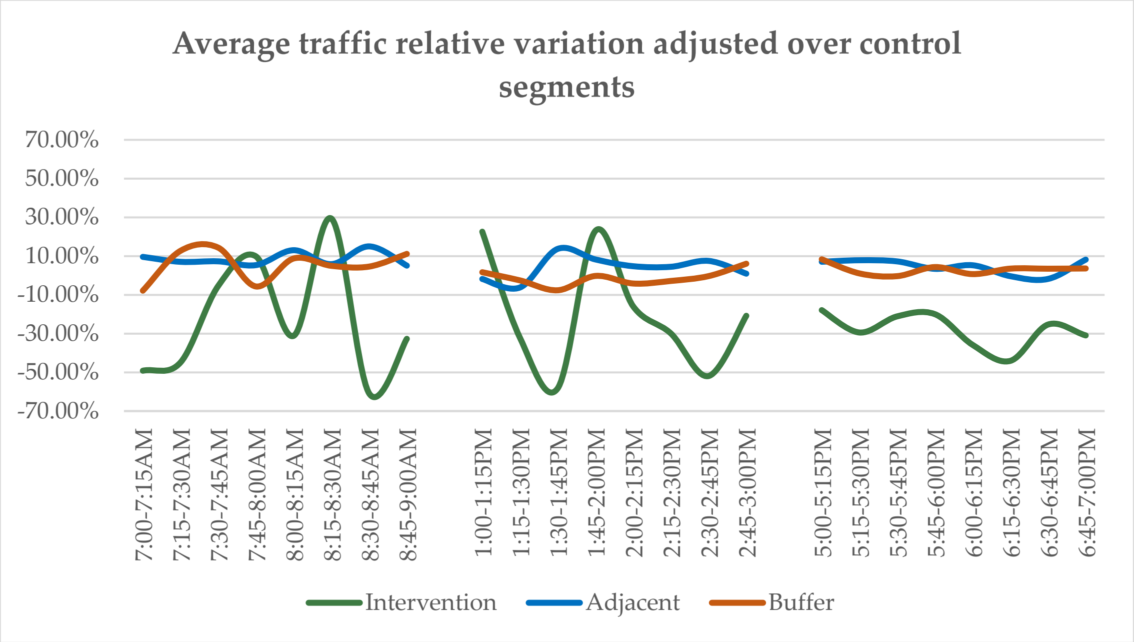


Figure 16. Average traffic flow variation adjusted over control segments fluctuations

Average indexes of traffic flow variation are presented in Table 7, consisting of mean values per time slot of relative fluctuations for average traffic flows:

Table 7. Average traffic flow variation relative to control segments

Buffer segment class	7:00 - 9:00 AM	1:00 - 3:00 PM	5:00 - 7:00 PM	*Average*
Intervention	-23.11%	-20.42%	-28.08%	-23.87%
Adjacent	+8.53%	+3.96%	+4.61%	+5.70%
Buffer	+5.38%	-1.25%	+3.09%	+2.41%

From both Figure 16 and Table 7, it is possible to outline findings relatively to the impact that the tactical urbanism intervention may have had on the road network:

- the intervention streets, despite presenting considerable peaks due to little flow so more sensitive to variations, averagely faced a consistent reduction, coherently with both the nature and the main purpose of a capacity-limiting, pedestrian-friendly road intervention;
- segments in the adjacent and buffer zones have recorded, albeit lower both in relative and in absolute terms with respect to intervention

segments, a positive relative variation in the matter of traffic flow with respect to the general fluctuations calculated over control segments.

These figures suggest that the implementation of the new temporary school square has moved traffic patterns from the intervention area to alternative paths which deviate from the former most convenient routes using adjacent streets. A further level of interpretation of such data may conclude from available evidence that traffic has disappeared from segments interested by the new implementation, but not from the entire road network (corresponding trend analysis for weekdays on the whole road network is displayed in Figure 17), so the phenomenon of *traffic evaporation*, encountered in other, more extensive street experiments (Nello-Deakin, 2022), did not occur here, pointing out that in the case under examination a punctual tactical urbanism could be disruptive to traffic patterns, even if slightly, but did not succeed in diminishing traffic flow in the neighbourhood network.

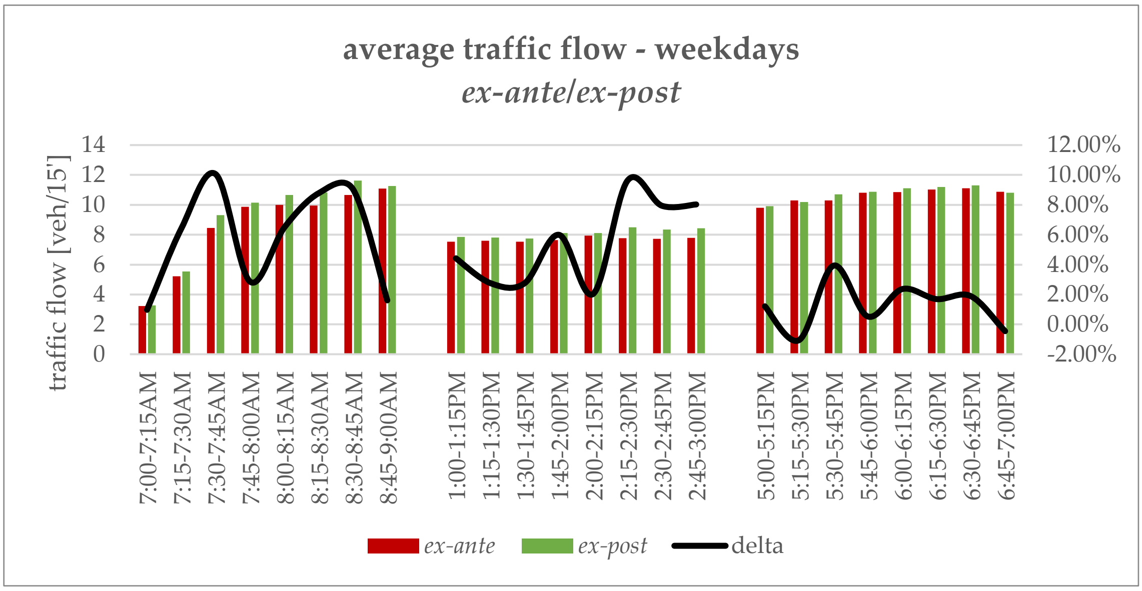


Figure 17. Average traffic flow on weekdays - comparison *ex-ante/ex-post*

Sticking to relative fluctuations in traffic flows, network maps have been elaborated through GIS for each time slot, representing percentage changes between *ex-ante* and *ex-post* conditions per each segment.

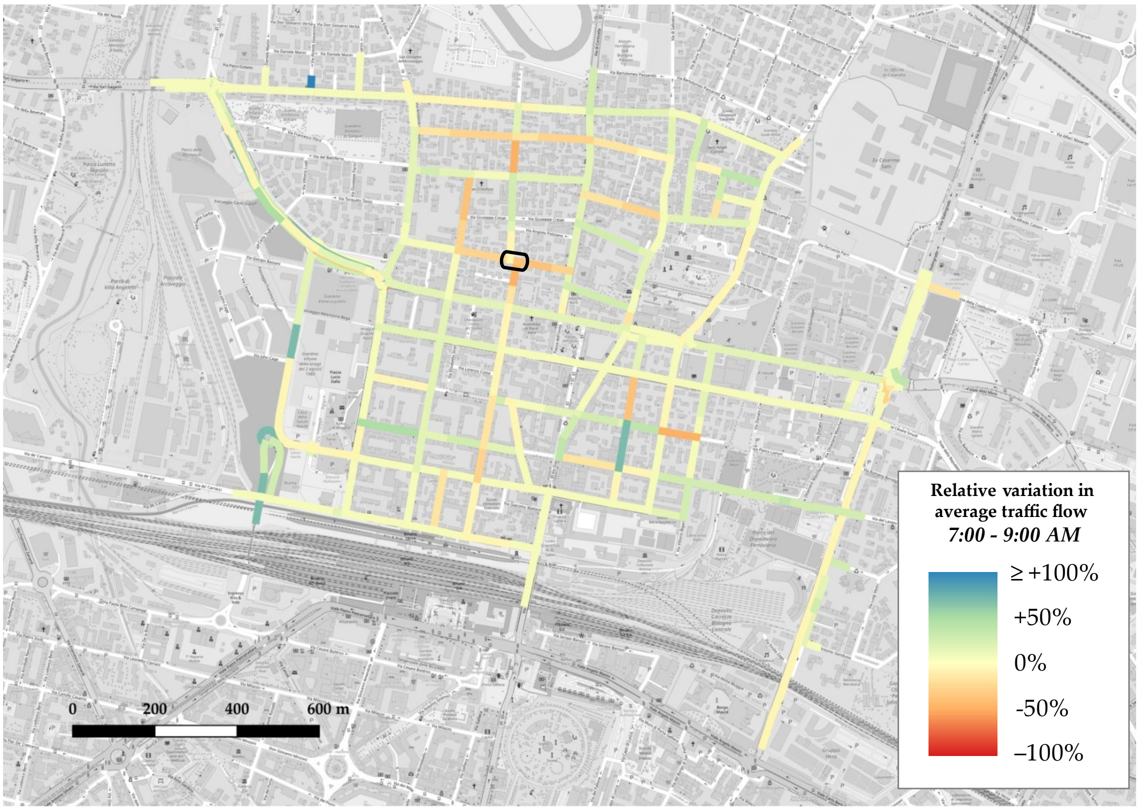

0
200
400
600 m
Relative variation in average traffic flow
7:00 - 9:00 AM
≥ +100%
+50%
0%
-50%
–100%

Figure 18. Relative variation in average traffic flow - time slot 7:00-9:00 AM

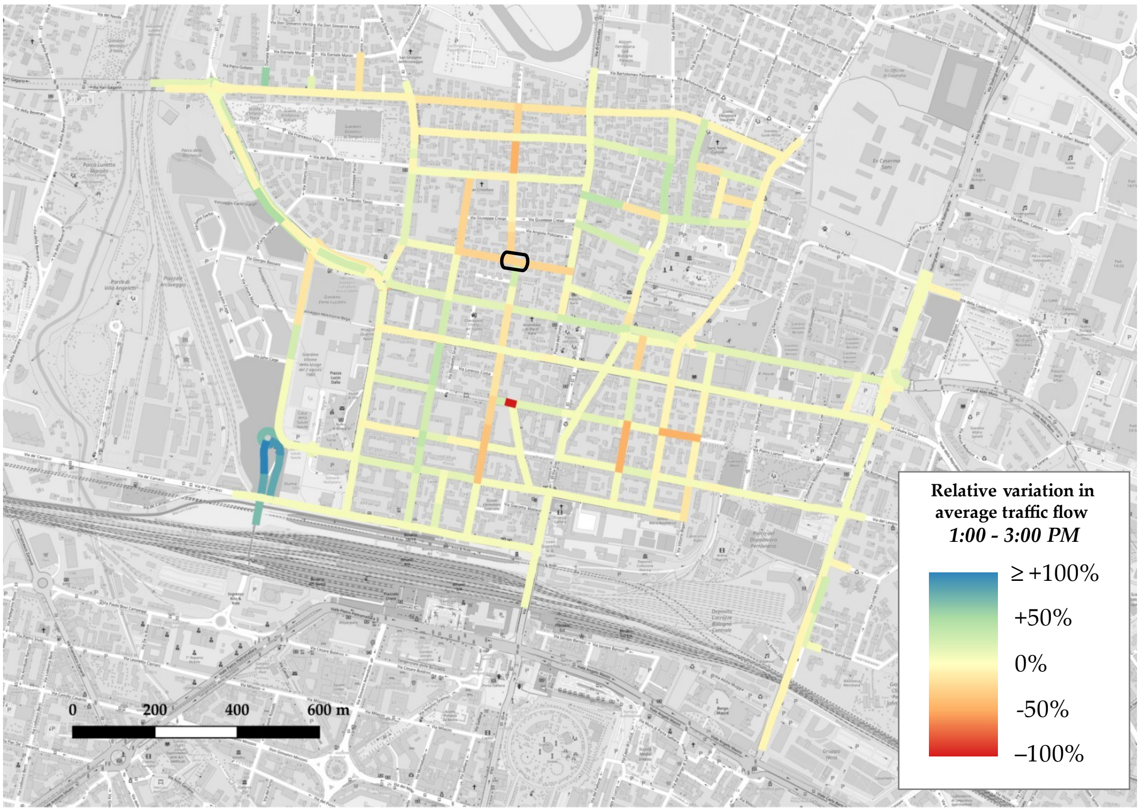

0
200
400
600 m
Relative variation in average traffic flow
1:00 - 3:00 PM
≥ +100%
+50%
0%
-50%
–100%

Figure 19. Relative variation in average traffic flow - time slot 1:00-3:00 PM

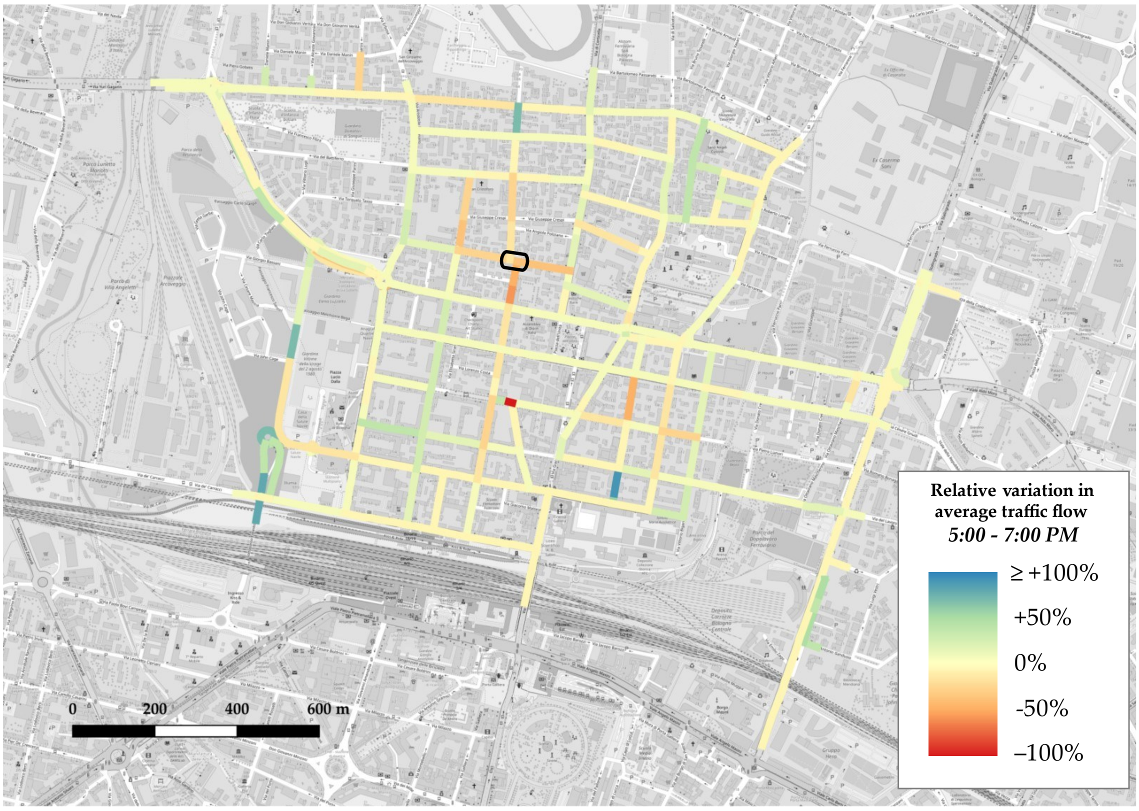


Figure 20. Relative variation in average traffic flow - time slot 5:00-7:00 PM

GIS maps confirm what commented earlier about the relocation of traffic patterns from segments directly affected by the intervention to adjacent and buffer streets: it is to underline once again that, in spite of a quite consistent relative reduction in intervention streets, highlighted in the maps through stronger warm colours, surrounding segments have recorded a relatively slight increase in traffic flow, simple because of already considerable traffic flows interesting such segments.

To better understand which roads may have experienced stronger impacts and to validate considerations so far expressed, network maps have been elaborated through GIS, representing absolute variations in each time slot between *ex-ante* and *ex-post* conditions per each segment.

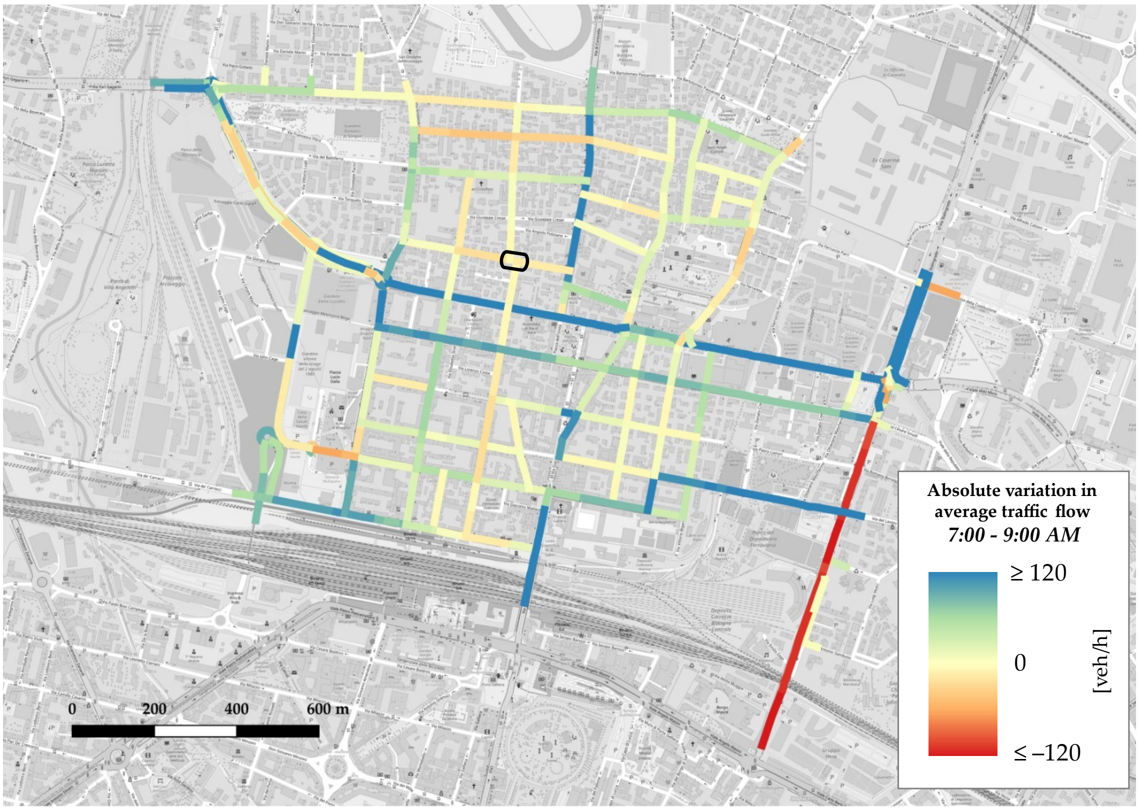

0
200
400
600 m
Absolute variation in average traffic flow
7:00 - 9:00 AM
≥ 120
0
≤ –120
[veh/h]

Figure 21. Absolute variation in average traffic flow - time slot 7:00-9:00 AM

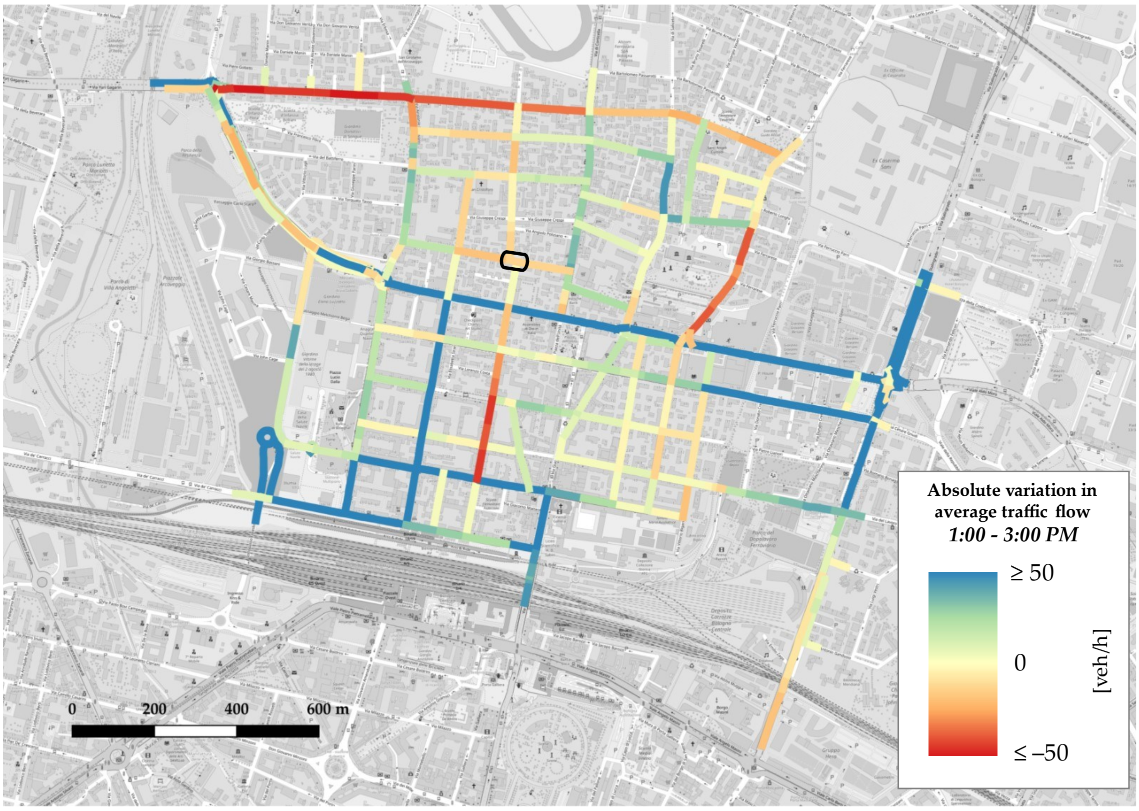

0
200
400
600 m
Absolute variation in average traffic flow
1:00 - 3:00 PM
≥ 50
0
≤ –50
[veh/h]

Figure 22. Absolute variation in average traffic flow - time slot 1:00-3:00 PM

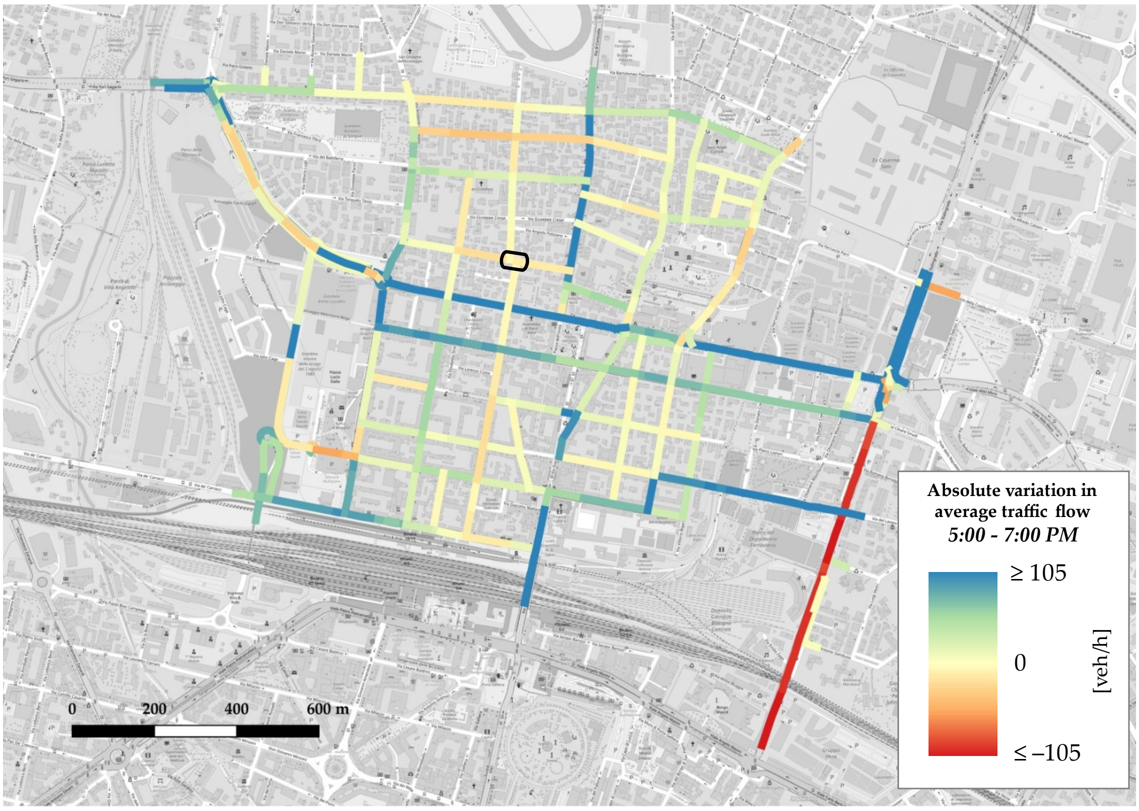


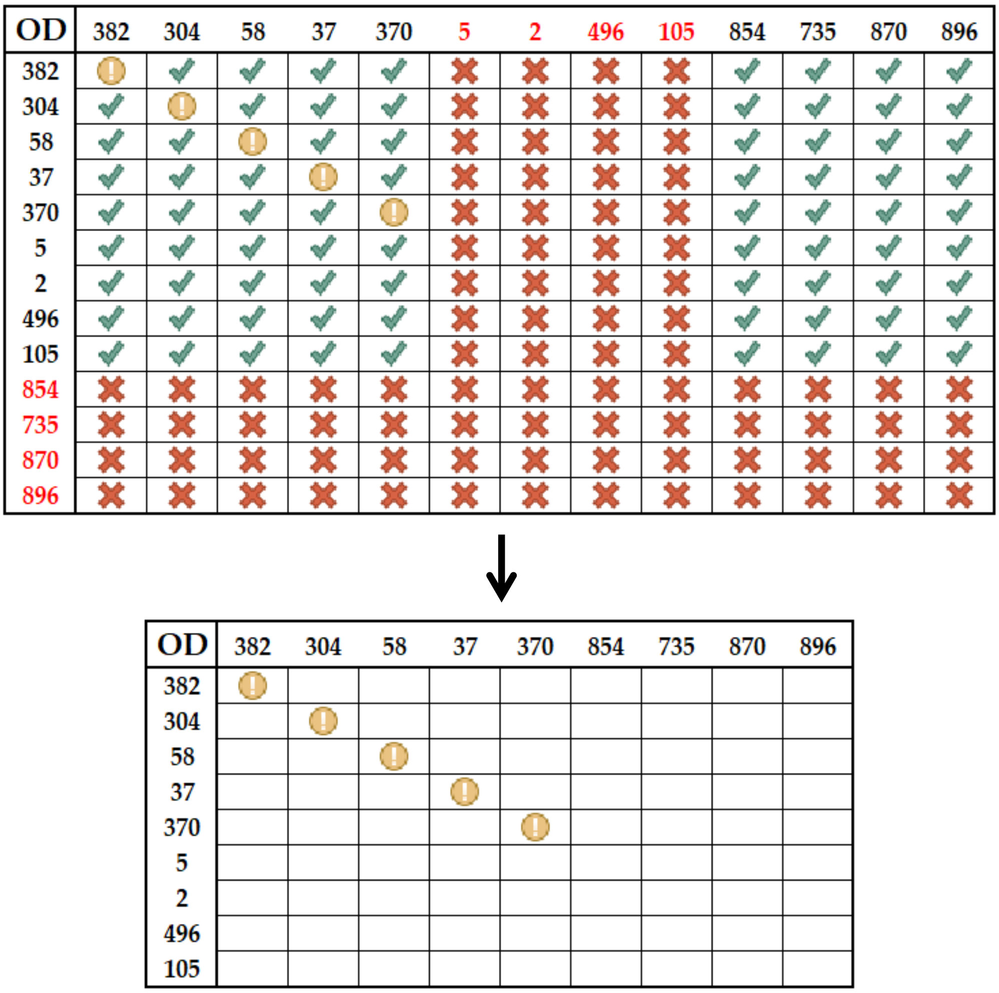
Figure 23. Absolute variation in average traffic flow - time slot 5:00-7:00 PM

Referring to absolute variation maps elaborated through GIS and supported by global indexes displayed in Table 7, a clear difference in vehicular patterns emerges focusing on the surroundings of the new temporary square: intervention segments have recorded a slight negative variation in every addressed time slot; adjoining and neighbouring segments such as via di Corticella, representing the major axis which accesses the neighbourhood from its northern side, via Franco Bolognese and via Pellegrino Tibaldi, core axes for the east-west through-traffic phenomenon, figuring either in the adjacent or buffer class, have been affected by an appreciable increase in traffic flow. These considerations stand as hint of vehicular pattern changes towards a worsening of traffic conditions in the road segments surrounding the intervention area.

5.2. OD matrices estimation

The described procedure has been applied to retrieved datasets containing traffic-related values recorded during weekdays for each time fraction *k* to obtain OD matrices, both for *ex-ante* and *ex-post* traffic conditions; data collected on weekends have been discarded for the reasons commented in section 4.3.2.

According to the list of origin and destination nodes resulted from the zoning process, the number of unknowns for each equation system, which corresponds to the number of empty cells of Origin-Destination matrices, is 76; other cells of each $n \times n$ matrix are automatically set to the value of zero because no path could be established between the nodes (due to postulated absence of intrazonal trips and for the presence of origin or destination only nodes). Because of this, henceforth it will be used a contracted OD matrix layout, which excludes rows representing only destination nodes and columns only for origin nodes:



OD	382	304	58	37	370	5	2	496	105	854	735	870	896
382	!	✓	✓	✓	✓	✗	✗	✗	✗	✓	✓	✓	✓
304	✓	!	✓	✓	✓	✗	✗	✗	✗	✓	✓	✓	✓
58	✓	✓	!	✓	✓	✗	✗	✗	✗	✓	✓	✓	✓
37	✓	✓	✓	!	✓	✗	✗	✗	✗	✓	✓	✓	✓
370	✓	✓	✓	✓	!	✗	✗	✗	✗	✓	✓	✓	✓
5	✓	✓	✓	✓	✓	✗	✗	✗	✗	✓	✓	✓	✓
2	✓	✓	✓	✓	✓	✗	✗	✗	✗	✓	✓	✓	✓
496	✓	✓	✓	✓	✓	✗	✗	✗	✗	✓	✓	✓	✓
105	✓	✓	✓	✓	✓	✗	✗	✗	✗	✓	✓	✓	✓
854	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
735	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
870	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
896	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗

↓

OD	382	304	58	37	370	854	735	870	896
382	!								
304		!							
58			!						
37				!					
370					!				
5									
2									
496									
105									

Figure 24. Contraction of OD matrix layout

Through the process of link counts selection, m linear equations representing counted segments have been chosen to build non-singular coefficient matrices $\boldsymbol{A}_k$

and corresponding vectors of observed traffic flows $\tilde{\boldsymbol{y}}_k$: in Table 8, m is reported for each case.

Table 8. Number m of selected linear equations per each time fraction

$\boldsymbol{k}$	$\boldsymbol{m}$ **linear equations**	
	ex-ante	***ex-post***
1	73	62*
2	58*	74
3	75	73
4	76	74
5	74	75
6	75	75
7	75	74
8	75	74
9	65*	75
10	74	75
11	74	76
12	74	74
13	65*	74
14	75	75
15	63*	74
16	74	74
17	69*	75
18	75	74
19	75	75
20	74	75
21	74	74
22	75	75
23	75	75
24	73	75

* $m < 70$

Following the choice of cutting off the number of possible paths between OD pairs into $\bar{k} = 4$ most convenient routes in terms of travel times, nearly the majority of equation systems used in the algorithm have resulted slightly underdetermined, with only a couple of exceptions of full-rank conditions. In a few cases, the number of equations has dropped under the threshold of 70; however, this seems not to have affected the results in terms of correlation between estimations and observed flows.

OD matrices for each 15-minute fraction k have been extracted through the optimization problem which minimizes the norm of the difference vector between estimated and observed link flows, described in section 4.3.8. Linear regressions between observations and estimations have been established in two separate sets, the first consisting of flows selected to take part in the coefficient matrix, the second containing discarded flows used for the validation of the method: hence, it is possible to compute coefficients of determination R^2 for each linear regression and to evaluate the accuracy of the estimation process. In Table

9, numerical values of R^2 per each time fraction, both for *ex-ante* and *ex-post* situations are shown:

Table 9. Coefficients of determination R-squared (R^2)

k	Time fraction	Coefficients of determination R^2			
		ex-ante		*ex-post*	
		Selected	**Discarded**	**Selected**	**Discarded**
1	7:00 - 7:15 AM	0.909	0.747	0.885	0.718
2	7:15 - 7:30 AM	0.928	0.568*	0.827	0.587*
3	7:30 - 7:45 AM	0.852	0.431*	0.886	0.568*
4	7:45 - 8:00 AM	0.859	0.794	0.869	0.808
5	8:00 - 8:15 AM	0.911	0.825	0.873	0.816
6	8:15 - 8:30 AM	0.912	0.813	0.909	0.822
7	8:30 - 8:45 AM	0.904	0.738	0.877	0.777
8	8:45 - 9:00 AM	0.867	0.773	0.895	0.825
9	1:00 - 1:15 PM	0.856	0.730	0.916	0.799
10	1:15 - 1:30 PM	0.933	0.835	0.898	0.826
11	1:30 - 1:45 PM	0.911	0.841	0.899	0.826
12	1:45 - 2:00 PM	0.900	0.809	0.929	0.814
13	2:00 - 2:15 PM	0.937	0.816	0.927	0.822
14	2:15 - 2:30 PM	0.911	0.819	0.927	0.851
15	2:30 - 2:45 PM	0.922	0.635*	0.910	0.821
16	2:45 - 3:00 PM	0.901	0.781	0.910	0.842
17	5:00 - 5:15 PM	0.921	0.824	0.847	0.790
18	5:15 - 5:30 PM	0.881	0.729	0.884	0.778
19	5:30 - 5:45 PM	0.883	0.774	0.886	0.800
20	5:45 - 6:00 PM	0.844	0.660*	0.833	0.752
21	6:00 - 6:15 PM	0.901	0.801	0.875	0.764
22	6:15 - 6:30 PM	0.861	0.784	0.872	0.810
23	6:30 - 6:45 PM	0.855	0.738	0.874	0.795
24	6:45 - 7:00 PM	0.884	0.686*	0.845	0.783

* $R^2 < 0.7$

From the coefficients of determination calculated and presented in the previous table, correlation between estimated and observed values of link flows once the OD matrix has been elaborated is defined as strong, since $R^2 > 0.7$ in every regression for selected flows and almost any correlation for validating the model through discarded flows. Figure 25 and Figure 26 show plotted series for R^2 values along with high positive correlation threshold $R^2 = 0.7$, represented by the horizontal green line.

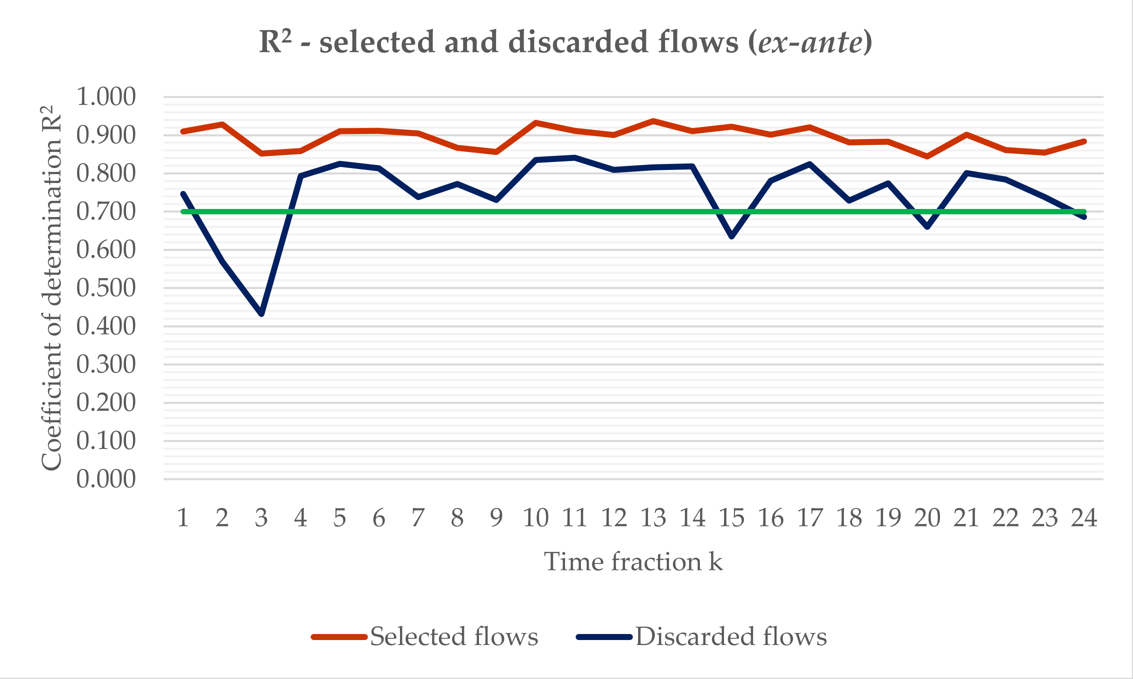


Figure 25. R^2 values for *ex-ante* scenario

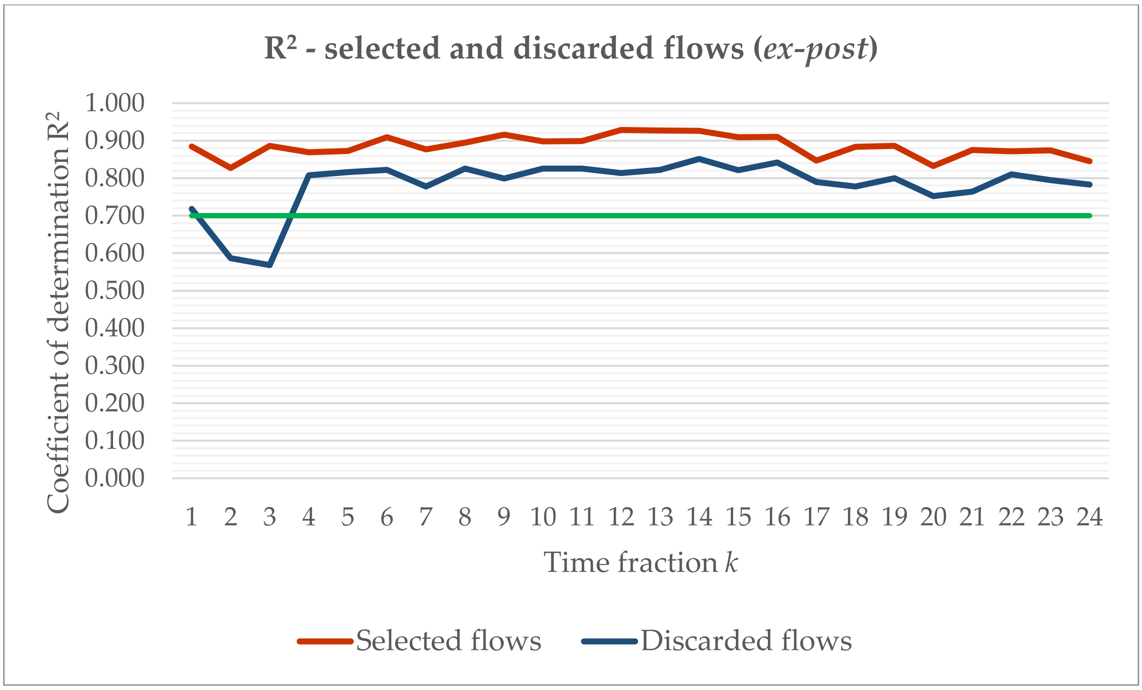


Figure 26. R^2 values for *ex-post* scenario

Feeling necessary to further inspect regressions whose R-squared is below the acceptance threshold, it is clever to investigate the main factors that could affect the quality of the correlation, therefore the size of correlation coefficients r (Goodwin et al, 2006):

a) variability in the data;
b) matching degree for shapes for distributions;
c) nonlinearity;
d) outliers;
e) sample characteristics;
f) measurements errors.

Notwithstanding that f) is clear and present for every set, could and surely is affecting correlation quality, the other factors could be investigated to look for the downgrading element.

For each of the "bad performing" correlations, Theil's inequality coefficient is computed, and later decomposed into its three indicators which should be able to explain (or at least exclude) which is the problem.

Table 10. Theil's inequality coefficients for low-R^2 regressions

	Theil's inequality coefficient U						
	ex-ante, k					***ex-post, k***	
	2	**3**	**15**	**20**	**24**	**2**	**3**
U	0,2110	0,2699	0,1913	0,1843	0,1768	0,1997	0,2091
U_M	0,0008	0,0119	0,0110	0,0090	0,0038	0,0001	0,0004
U_S	0,0015	0,0549	0,0007	0,0009	0,0070	0,0002	0,0078
U_C	0,9977	0,9332	0,9883	0,9901	0,9892	0,9997	0,9918

From the reported values, regressions featuring lower R^2, and for this reason addressed here, present values of U close to the threshold of 0.2. By checking inequality proportions values, it seems that no issue is encountered relatively to systematic errors or variability, since the inequality proportions relative to bias and variance stick to small values, certifying the quality of the prediction. Since these statistics do not directly address the effect of outliers, through a qualitative check on linear regression scatterplots it is possible to highlight considerable incidence of isolated points (actually streams of points, due to the fragmentation of the underlying road graph). In Figure 27 and Figure 28, examples are shown.

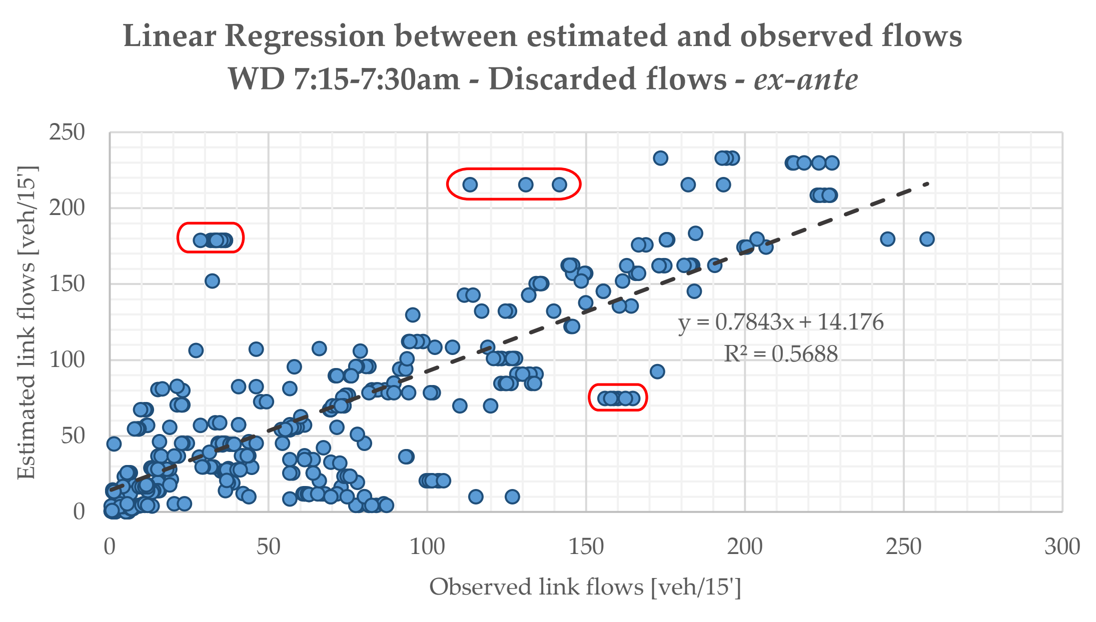


Figure 27. Linear regression of discarded flows - *ex-ante*, weekdays, 7:15-7:30 AM time fraction

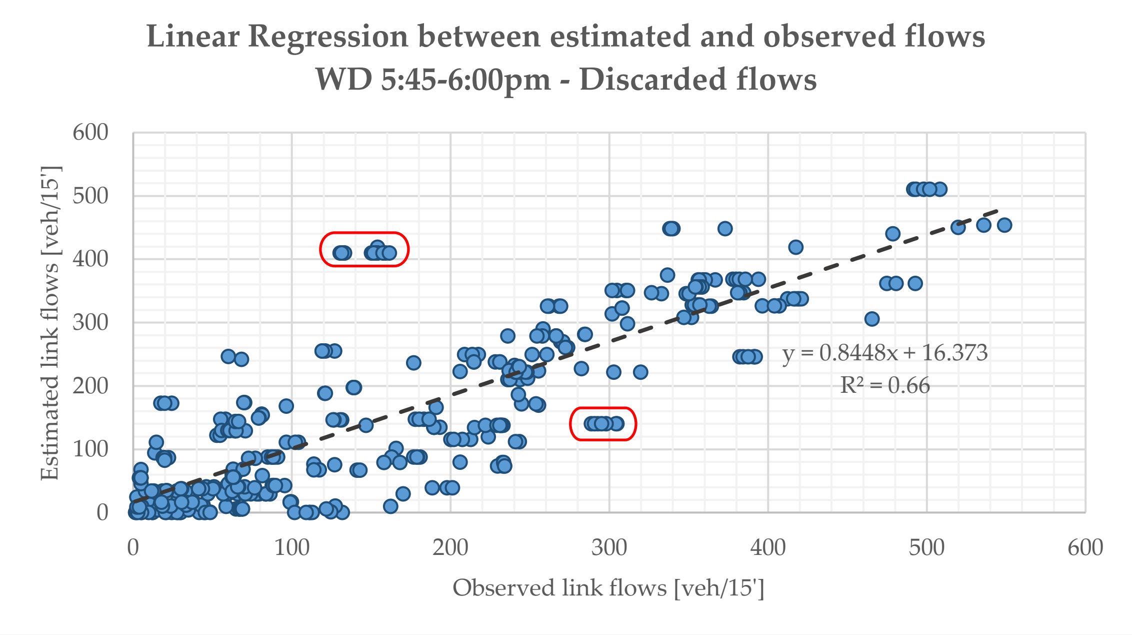


Figure 28. Linear regression of discarded flows - *ex-ante*, weekdays, 5:45-6:00 PM time fraction

In Figure 29, absolute deviation of each OD flow is represented in three different chromatic matrices, disaggregated per time slot.

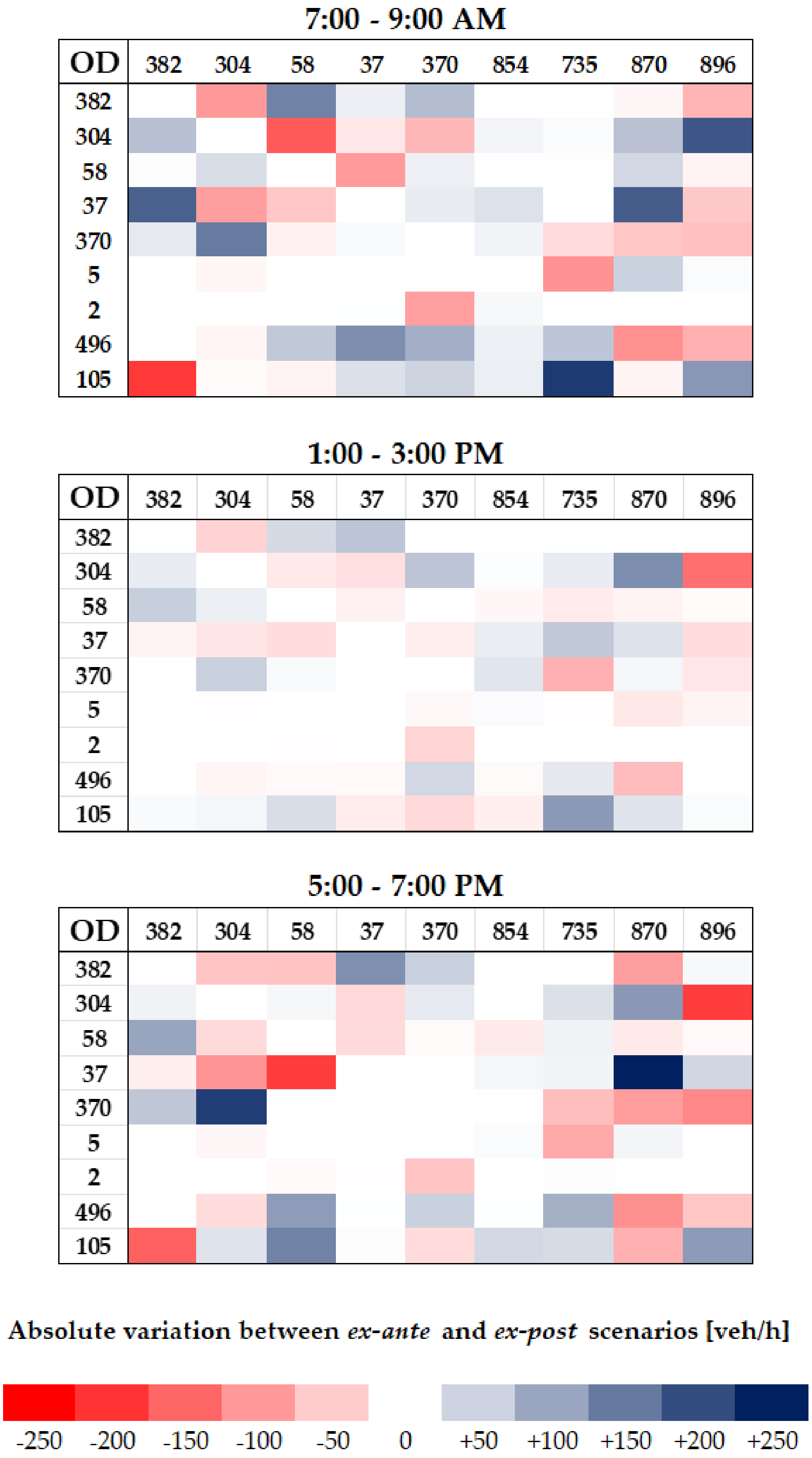


Figure 29. Absolute variation of OD flows between *ex-ante* and *ex-post* scenarios

From the previous figure, some quick but effective remarks both on accuracy of estimated OD matrices and on the general trend of OD flows can be made:

- in all three matrices, from the lightest shade to the darkest one, blue cells depicting increments are more than red cells, which denote decrements; this stands in consistency with trend analysis for average traffic flow on weekends suggested (Figure 17), which displayed a general rise in each time slot;
- darkest colours appear in correspondence of OD pairs whose source and/or sink nodes represent accesses or exits from the network through major axes, in accordance with network absolute variation maps (Figure 21 to Figure 23) which reported increases on east-west through-traffic routes and main accesses to the neighbourhood from both north and south;
- time slot 1:00-3:00 PM features lighter colours, sign that absolute variations during lunchtime are lower with respect to the morning and evening slots; this could either suggest a less disrupted pattern behaviour or, more probably, be a consequence of the fact that, as it is not one of the two peak time slots by definition, it reportedly carries less traffic flow, so in absolute terms also variations tend to be less evident.

Further validation of estimated OD matrices will be attainable once the obtained OD demand is fed to micro simulation software that assigns 15-minute matrices back to the road network (the process could be done both statically and dynamically) and elaborates network maps displaying estimated flows on each segment to be compared with observations.

6 Discussion

Overall, the applied procedure worked, allowing both to make meaningful considerations about traffic flow variations over the neighbourhood network and to extract reliable OD matrices for each time fraction k, but there are some limitations that need to be addressed, to highlight the implications on the final results.

First, representativity of sampled data is a fundamental step since the main application of the employed procedure is representing real traffic values in order to support decision-making on actual scenarios. To faithfully expand sample sizes, proved expansion coefficients based on actual coverage of accessible sensors should be employed; if not available, it is necessary at least to perform a comparison between flow observations extracted from the set as representative for the sample and traffic counts covering the total flow in the same counting spots and in the same period of the GPS-based sampling process. In any cases, the two datasets to be compared need to be a robust estimate of what they represent, so that the expansion coefficient calculation returns reliable multipliers. Nonetheless, as observations systematically contain a considerable number of errors, expansion coefficients will likely amplify such errors.

Then, assumptions have been made through the procedure in order to smoothen the most complicated steps, but this may come at a price in accuracy in the final results:

- the choice of cutting off the set of shortest paths to a determined number $\bar{k}$ may exclude paths that are significant and actually used by the OD pair demand; such approximation redistributes discarded path flows on the ones already included in the set, so the final outcome could unrealistically concentrate flows in few convenient segments according to the assumed cost functions; however, this assumption is needed as the listing of all paths for every considered Origin-

Destination pair is computationally way too demanding and it does not benefit from a practical point of view.

- Under the assumption of a common $\bar{k}$ for each OD pair, through sensitivity tests it is possible to choose the most suitable value of such parameter, balancing the credibility of the assumption related to the size of the network under examination (a too large $\bar{k}$ can be counterproductive and also unrepresentative, given the fact that the common user does consider a limited set of paths as alternatives, especially at neighbourhood scale) and the determination of the mathematical problem, in order to obtain consistent outcomes; it could be intelligent to choose an OD-pair-specific $\bar{k}$ to select the appropriate number of possible paths in relation to effectively convenient alternatives for the considered trip.
- Path Size Logit (PSL) model for the calculation of route choice probability only considered experienced travel time as explanatory variable, though enriched through the path size factor: other traffic-related variables can be considered, for instance speed limits imposed on segments (with a tight tie with 30 km/h zones question), or presence of signalized intersections, which may distort the perception of the user on the convenience of the path.

Extraction of 15-minute OD matrices overall returned reliable input data for further implementation of simulation models, whose accuracy have been validated through regression parameters. Nevertheless, there are some time fractions that recorded low R^2 values in the validation via discarded flows. As suggested from evidence in the last section, these weak results in correlation must not be imputed to the determinedness of equation systems (as already said, the number m of equations could not always explain low values for R-squared) but they may find an explication in the proposed criterion of selection of link traffic counts. As outlined before, the logic was to select segments which could give the richest information by containing as much OD pairs as possible while being independent from already considered equations, in order to avoid the mentioned data dependence issue. The main assumption that led to this proposal is that, even though no prior matrix was available, the two fundamental assumptions regarding path sets cut-off and route choice probability provided information about used paths as if a prior matrix were at hand. However, as the underlying logic is based on assumptions, the correlation between actual paths and predicted ones could have been affected by biases, and so could have the criterion.

It is though worth understanding why overall OD flows have fitted also discarded link flows after being assigned to the road network but visible inaccuracies have been registered in some time fractions. After considering goodness-of-fit measures such as Theil's inequality coefficients and its related proportions, which excluded systematic errors as well as ability to reproduce variability, and by qualitative assessment of linear regression scatterplots, encountered inaccuracies could be explained through the detrimental effect of outliers on such statistical validation tools. In particular, due to the high fragmentation of the road graph provided by TomTom, which negatively affects data inconsistency, if the model assigns an outlying estimation to a road segment, in the linear regression the street is represented by a number of fragments which emphasize the negative outlying effects. To overcome this problem, a possible solution could be the harmonization of the road graph together with segments attributes.

In the end, intrinsic in this method there might be strong biases due to the fact that data have been recorded from vehicles equipped with navigation systems. The construction of set of paths, as already said, is led once again by the assumption of the user choosing among a selection of best paths, fits with the ground functioning of navigation systems, as every device informs the user about the best paths along with some convenient real-time alternatives together with predicted travel times. This assumption, though, clashes with the nature of systematic trips, which are not guided by any navigation system, but derive from habits of the systematic traveller. This could be a major limitation, because the algorithm tends to concentrate flows on best paths, discarding actually used segments, and cannot be verified in absence of ground truth data.

7 Conclusion and future developments

This work aimed at assessing the effects on private vehicular patterns in the neighbourhood of Bolognina, in the city of Bologna, following the implementation in March 2022 of a temporary square dedicated to young students in via Camillo Procaccini, under the principles of tactical urbanism approach. It made use of TomTom Floating Car Data collected over the months of September and October 2021 (before the intervention, *ex-ante*) and 2022 (after the intervention, *ex-post*), consisting of traffic counts, average speed values and travel times for each segment of the selected road network over three daily time slots and per time fractions of 15 minutes.

After validating datasets and data clustering through hypothesis tests on statistical significance, trend analyses have shown that traffic flows experienced a slight global increase within a year, with a corresponding mild increase in speed-related values.

More in detail, for what concerns traffic flows, under the classification of segments in accordance to graph distance from the intervention site, streets directly interested by the intervention recorded a considerable decrease (-23.9%), with adjacent streets and segments suffering a more contained but positive increase (+5.70% and +2.41%, respectively). This suggests that the phenomenon of *traffic evaporation*, observed in findings related to more widespread implementations of tactical urbanism interventions, did not occur in this situation, which exhibited a more impacting traffic flow relocation on alternative paths.

For trend analyses related to speed values, some data-led considerations can be made regarding the scheduled extensive application of 30 km/h speed limits in the whole neighbourhood: as recorded 85th percentiles of speed values show higher values than the projected speed limit, the planned implementation is not backed by widely recognized methodologies.

On the whole, extraction of OD matrices per 15-minute time fractions through data-driven procedures returned reliable results, which have been validated through coefficients of determination for regressions between observed and estimated link flows, also thanks to the large availability of data discarded from the equation system construction due to data dependence.

Nonetheless, despite the big number of observations, almost every equation system adopted to extract matrices resulted, even if slightly, underdetermined: this is because many observations were dependent on the others, due to high fragmentation of the segments in the network (each one with their own supply pattern of traffic-related information), to the grid layout of the road network and to the cut-off of the number of possible paths per OD pair.

Future developments of this work could include:

- an improvement of the adopted procedure through harmonization of the road graph and segments attributes in order to overcome issues such as encountered outliers or considerable data inconsistency;
- microsimulations of the road network exploiting as demand inputs the extracted OD matrices, both as further validation of the extracting procedure and as a more thorough assessment of the impact the tactical urbanism intervention had on the private mobility patterns via extraction of traffic-related metrics (queue lengths at intersections, Level of Service);
- vehicular network performance-based evaluation of further developments of the design of the temporary school square with the already retrieved input data.

Bibliography

Abdelfattah, L., Boni, G., Carnevalini, G., Gorrini, A., Messa, F., & Presicce, D. (2021). A user-centric approach to the 15-minute city. Examining children's walkability in Bologna. *Proceedings of the 57th ISOCARP World Planning Congress 2021* (ISOCARP 2021), (p. 582-591). Doha (Qatar).

Asuero, A. G., Sayago, A., & González, A. G. (2006). The Correlation Coefficient: An Overview. In *Critical Reviews in Analytical Chemistry* (Vol. 36, Issue 1, pp. 41–59). Informa UK Limited. https://doi.org/10.1080/10408340500526766

Barceló, J., Montero, L., Marqués, L., & Carmona, C. (2010). Travel Time Forecasting and Dynamic Origin-Destination Estimation for Freeways Based on Bluetooth Traffic Monitoring. In *Transportation Research Record: Journal of the Transportation Research Board* (Vol. 2175, Issue 1, pp. 19–27). SAGE Publications. https://doi.org/10.3141/2175-03

Ben-Akiva, M. & Bierlaire, M. (1999) Discrete Choice Methods and Their Applications to Short Term Travel Decisions. In: Hall, M., Ed., *Handbook of Transportation Science*, Kluwer Academic Publishers, USA, 5-33.

Bertolini, L. (2020). From "streets for traffic" to "streets for people": can street experiments transform urban mobility? In *Transport Reviews* (Vol. 40, Issue 6, pp. 734–753). https://doi.org/10.1080/01441647.2020.1761907

Cascetta, E. (1998). *Teoria e metodi dell'ingegneria dei sistemi di trasporto*. UTET.

Ceccarelli, G., Messa, F., Gorrini, A., Presicce, D., Choubassi, R. (2023, *submitted*). Deep Learning Analytics for Pre-post Evaluation of Tactical Urbanism Interventions: The Case of Bologna. *Journal of Public Space*.

D'Andrea, E., & Marcelloni, F. (2017). Detection of traffic congestion and incidents from GPS trace analysis. In *Expert Systems with Applications* (Vol. 73, pp. 43–56). Elsevier BV. https://doi.org/10.1016/j.eswa.2016.12.018

Deng, B., Denman, S., Zachariadis, V., & Jin, Y. (2015). Estimating traffic delays and network speeds from low-frequency GPS taxis traces for urban transport modelling. *European Journal of Transport and Infrastructure Research*, Vol 15 No 4 (2015). https://doi.org/10.18757/EJTIR.2015.15.4.3102

Espitia Echeverría, E. A. (2018). *Analisi dell'influenza della localizzazione di sensori di rilevamento di traffico nell'aggiornamento di matrici OD - caso di studio: Torino*. Master's thesis, Politecnico di Milano.

Global Design Cities Initiative (2022). *How to Implement Street Transformations. A Focus on Pop-up and Interim Road Safety Projects*. Available at: https://globaldesigningcities.org/publication/how-to-implement-street-transformations/

Goodwin, L. D., & Leech, N. L. (2006). Understanding Correlation: Factors That Affect the Size of r. In *The Journal of Experimental Education* (Vol. 74, Issue 3, pp. 249–266). Informa UK Limited. https://doi.org/10.3200/jexe.74.3.249-266

Goodwin, P., C. Hass-Klau, & S. Cairns. (1998). Evidence on the Effects of Road Capacity Reductions on Traffic Levels. *Traffic Engineering and Control*, June 1998, pp. 348–354.

Gorrini, A., Presicce, D., Messa, F., Choubassi, R. (2023). Walkability for Children in Bologna: Beyond the 15-minute City Framework. *Journal of Urban Mobility*, 3, 1-10. https://doi.org/10.1016/j.urbmob.2023.100052

Hunt, J. D., Brownlee, A. T., & Stefan, K. J. (2002). Responses to Centre Street Bridge Closure: Where the "Disappearing" Travelers Went. In *Transportation Research Record: Journal of the Transportation Research Board* (Vol. 1807, Issue 1, pp. 51–58). SAGE Publications. https://doi.org/10.3141/1807-07

Krishnakumari, P., Lint, H. van, Djukic, T., & Cats, O. (2019). A data driven method for OD matrix estimation. In *Transportation Research Procedia* (Vol. 38, pp. 139–159). Elsevier BV. https://doi.org/10.1016/j.trpro.2019.05.009

Leduc, G. (2008). Road Traffic Data: Collection Methods and Applications. *Working Papers on Energy, Transport and Climate Change*, 1(55)

Ministero dei Trasporti (2006). *II° direttiva sulla corretta ed uniforme applicazione delle norme del codice della strada in materia di segnaletica e criteri per l'installazione e la manutenzione*.

Necula, E. (2015). Analyzing Traffic Patterns on Street Segments Based on GPS Data Using R. In *Transportation Research Procedia* (Vol. 10, pp. 276–285). Elsevier BV. https://doi.org/10.1016/j.trpro.2015.09.077

Nello-Deakin, S. (2022). Exploring traffic evaporation: Findings from tactical urbanism interventions in Barcelona. In *Case Studies on Transport Policy* (Vol. 10, Issue 4, pp. 2430–2442). Elsevier BV. https://doi.org/10.1016/j.cstp.2022.11.003

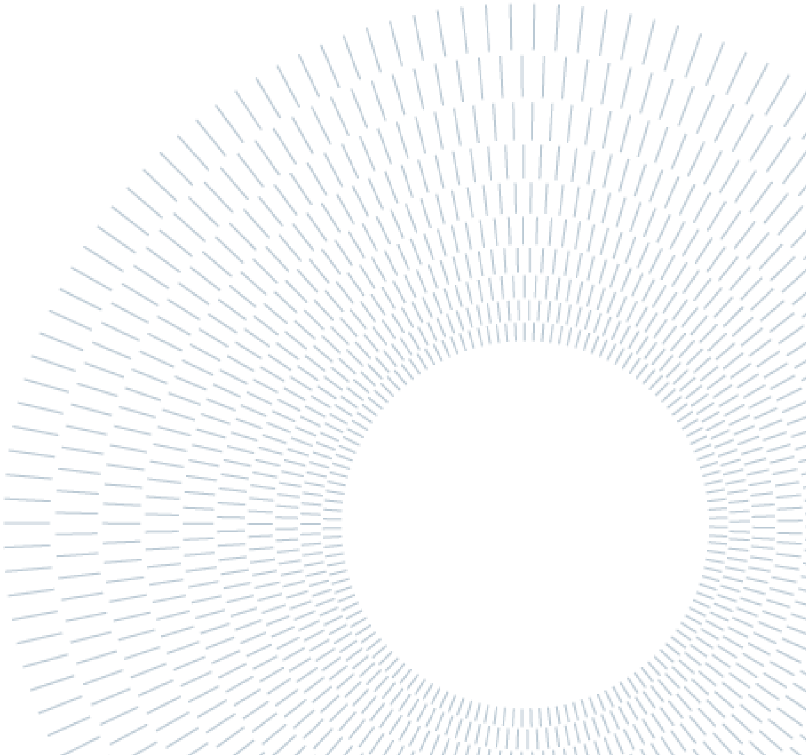
Ortúzar, J., & Willumsen, L. (2011). *Modelling Transport* (4th). United Kingdom: John Wiley & Sons.

Prato, C. G. (2009). Route choice modeling: past, present and future research directions. In *Journal of Choice Modelling* (Vol. 2, Issue 1, pp. 65–100). Elsevier BV. https://doi.org/10.1016/s1755-5345(13)70005-8

Tennøy, A., & Hagen, O. H. (2021). Urban main road capacity reduction: Adaptations, effects and consequences. In *Transportation Research Part D: Transport and Environment* (Vol. 96, p. 102848). Elsevier BV. https://doi.org/10.1016/j.trd.2021.102848

Toledo, T., & Koutsopoulos, H. N. (2004). Statistical Validation of Traffic Simulation Models. In *Transportation Research Record: Journal of the Transportation Research Board* (Vol. 1876, Issue 1, pp. 142–150). SAGE Publications. https://doi.org/10.3141/1876-15

TomTom Developers (2022, August 25). *Area Analysis | Traffic Stats*. https://developer.tomtom.com/traffic-stats/documentation/api/area-analysis

Yen, J. Y. (1971). Finding the K Shortest Loopless Paths in a Network. In *Management Science* (Vol. 17, Issue 11, pp. 712–716). Institute for Operations Research and the Management Sciences (INFORMS). https://doi.org/10.1287/mnsc.17.11.712