

# ANALYSIS OF THE FLOATING CAR DATA OF TURIN PUBLIC TRANSPORTATION SYSTEM: FIRST RESULTS

R. Ravanelli<sup>1,\*</sup>; M. Crespi<sup>1</sup>

<sup>1</sup> Geodesy and Geomatics Division, DICEA - University of Rome "La Sapienza", Rome, Italy  
<roberta.ravanelli, mattia.crespi>@uniroma1.it

Commission IV, WG IV/4

**KEY WORDS:** Geo Big Data, Floating Car Data, Trajectory Mining, Impedance Map Generation, Network Topology

## ABSTRACT:

Global Navigation Satellite System (GNSS) sensors represent nowadays a mature technology, low-cost and efficient, to collect large spatio-temporal datasets (Geo Big Data) of vehicle movements in urban environments.

Anyway, to extract the mobility information from such Floating Car Data (FCD), specific analysis methodologies are required. In this work, the first attempts to analyse the FCD of the Turin Public Transportation system are presented. Specifically, a preliminary methodology was implemented, in view of an automatic and possible real-time impedance map generation. The FCD acquired by all the vehicles of the Gruppo Torinese Trasporti (GTT) company in the month of April 2017 were thus processed to compute their velocities and a visualization approach based on Osmnx library was adopted.

Furthermore, a preliminary temporal analysis was carried out, showing higher velocities in weekend days and not peak hours, as could be expected. Finally, a method to assign the velocities to the line network topology was developed and some tests carried out.

## 1. INTRODUCTION

The largest part of movements in an urban environment is constrained to the road network. Nowadays, thanks to the recent development of navigation technologies, Global Navigation Satellite System (GNSS) sensors constitute a low-cost and efficient tool to collect such movement information, especially if compared with more traditional traffic monitoring methods like loop detectors or automatic plate number recognition (Yang and Gidófalvi, 2018).

GNSS sensors indeed acquire from time to time, even at high rate (e.g. 1 second or more), the position of the tracked object, so that its continuous movement is recorded as a trajectory containing a sequence of sampled points, obviously corrupted by noise (Yang and Gidófalvi, 2018) due to the pretty well-known problems encountered by GNSS in urban environment (obstructions, multipath, etc.). In the field of transportation, GNSS data collected from vehicles are frequently referred as Floating Car Data (FCD), and these Geo Big Data are recognized of more and more high value, since they contain required key information for estimating traffic impedance maps, possibly in real-time.

Anyway, this relevant objective can be only pursued provided that the GNSS trajectories are mapped to the road network. This work is precisely included in this background: the aim is to develop a reliable methodology able to perform the preliminary analyses needed for computing the impedance map. Specifically, the FCD of Turin (Italy) Public Transportation system were investigated, addressing the issues related to the management and visualization of such a huge amount of data. Furthermore, preliminary tests for projecting the raw data to the route lines were also successfully performed.

In detail, Section 2 illustrates the methodology developed to manage such a Geo Big Data dataset, while Section 3 focuses on the analysis of the first results and its related problems; in the end, in Section 4, some conclusions are drawn and future prospects are outlined.

## 2. METHODOLOGY

The FCD, here analysed off-line to set up a general methodology for the future real-time determination of roads impedance, were acquired in the month of April 2017 by the On Board Units (OBU) installed on the vehicles of the Gruppo Torinese Trasporti (GTT) company. The data, unfortunately characterized by a variable acquisition rate (from few seconds to tens of seconds), were provided in CSV format and they include the pairs of WGS84 geographical coordinates (longitude, latitude) along with a set of attributes such as the vehicle ID, the line ID and the timestamp.

The original file (2.19 GB) was converted in a database containing more than 30 million records in order to easily manage such a huge amount of information. The data were firstly organized into lines, then for vehicles and finally they were ordered according to the timestamps. In this way, for every line of the transportation network, it was possible to use the Vincenty formula (Vincenty, 1975) for computing the 2D displacements ( $\Delta s$ ) between two positions of the specific vehicle in two consecutive epochs ( $\Delta t$ ) starting from geographic coordinates. Then the velocities were easily calculated as  $v = \frac{\Delta s}{\Delta t}$ .

As shown in Figure 1, the computed velocities were represented as arrows and plotted on top of the Turin drive network graph, automatically downloaded from Open Street Map through the OSMnx Python library (Boeing, 2017).

Anyway, since the final goal is to compute the impedance maps of the roads served by the GTT bus lines, it is necessary to remove

\*Corresponding author

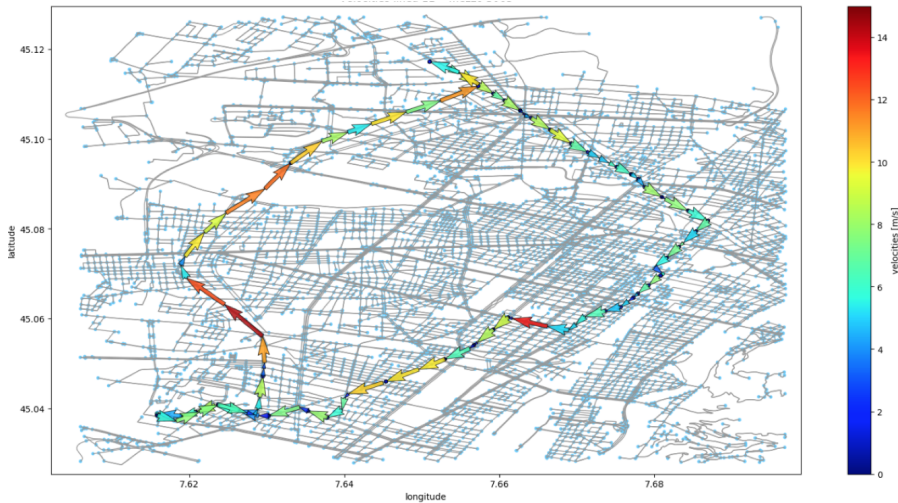
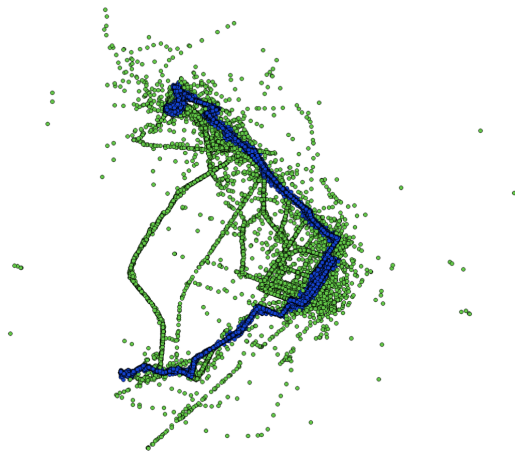


Figure 1. Velocity representation in the case of the vehicle 3063 of line 11.



(a) Line 11



(b) Line 39

Figure 2. Case studies for two bus lines in Turin (Italy, April 2017): blue dots are the positions actually recorded along the regular bus line path, green dots are the outliers.

all the spurious data/outliers in order to consider exclusively the velocities really representative of the situation of a certain route section (tree) at a certain epoch (Figure 2).

Then, before proceeding with the temporal analysis, the outliers were eliminated by removing all the records characterized by:

1.  $\Delta t$  higher than 99.5<sup>th</sup> percentile and lower than 0.5<sup>th</sup> (statistically not significant);
2. a velocity higher than 5 times the mean.

In this way, the values located in the tail of the  $\Delta t$  histograms are cut out (see Figure 3 and Figure 8) and the reconstructed path follows more closely the real line routes (Figure 2): the longest arrows, probably due to the paths of the vehicles from and to the depot, are discarded (see Figure 4, Figure 5, Figure 6, Figure 7). Anyway, it is still possible to observe the presence of velocities not referable to the actual path of the lines, so that some further investigations about the use of Local Outlier Factor (LOF) algorithm (Breunig et al., 2000) were carried out in (Pirrotti et al., 2018 - *under review*) and are not covered in this work.

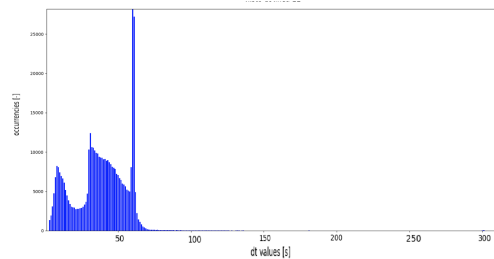


Figure 3. Histogram of  $\Delta t$  in the case of line 11 before the outlier removal.

Once the most of outliers were removed, a temporal analysis was performed by organizing the FCD into working and weekend days and considering the following time intervals during the day: 0 - 5; 5 - 7; 7 - 9; 9 - 11; 11 - 13; 13 - 15; 15 - 17; 17 - 19; 19 - 21; 21 - 24.

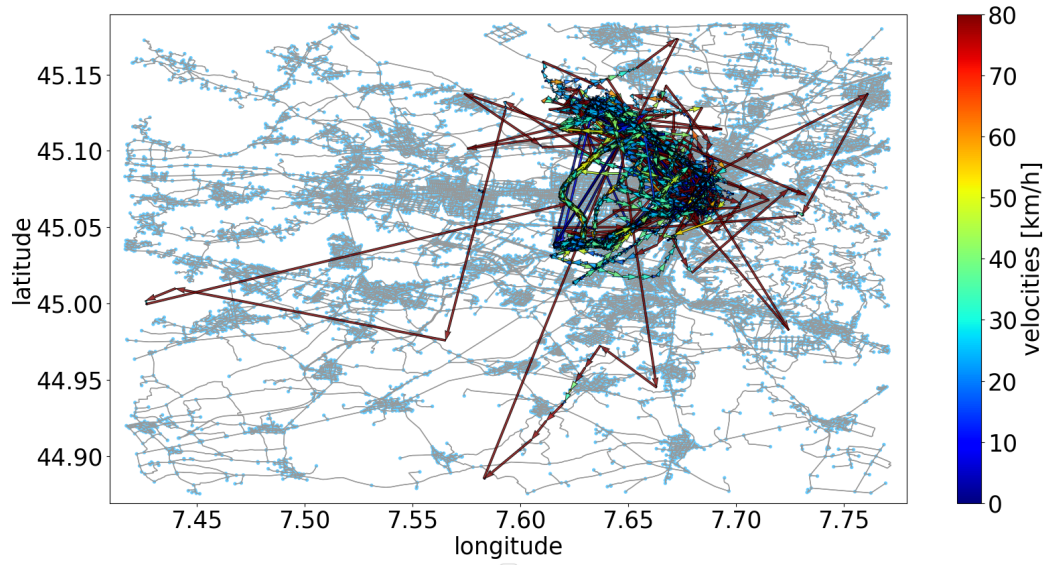


Figure 4. Velocities in the case of line 11 before the outlier removal.

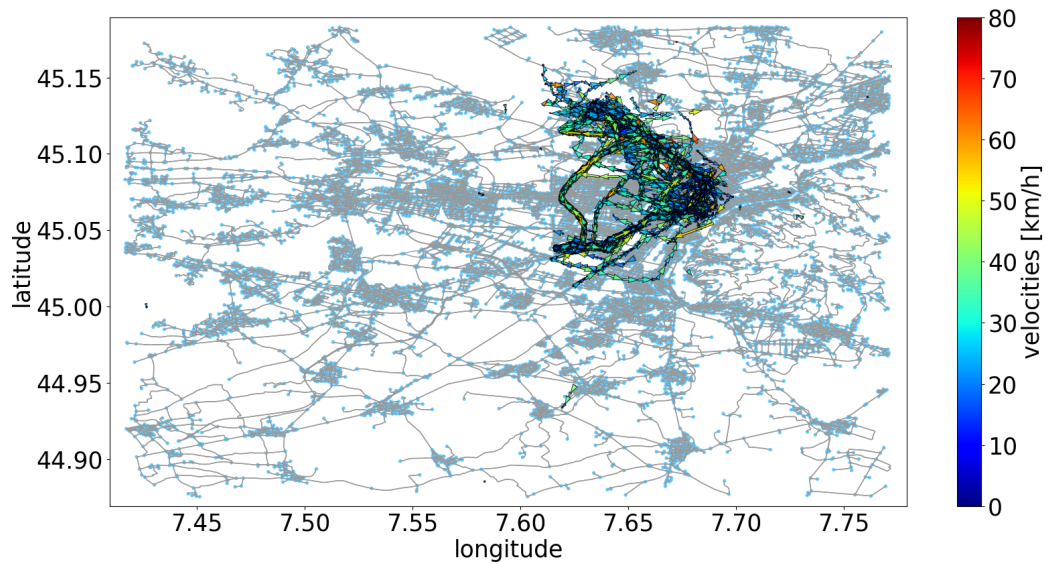


Figure 5. Velocities in the case of line 11 after the outlier removal.

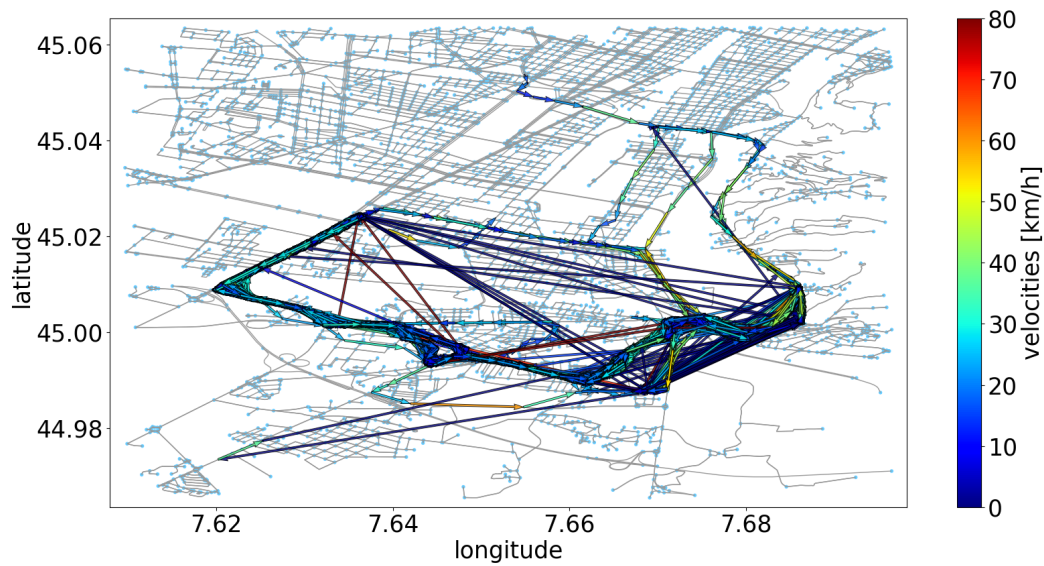


Figure 6. Velocities in the case of line 39 before the outlier removal.

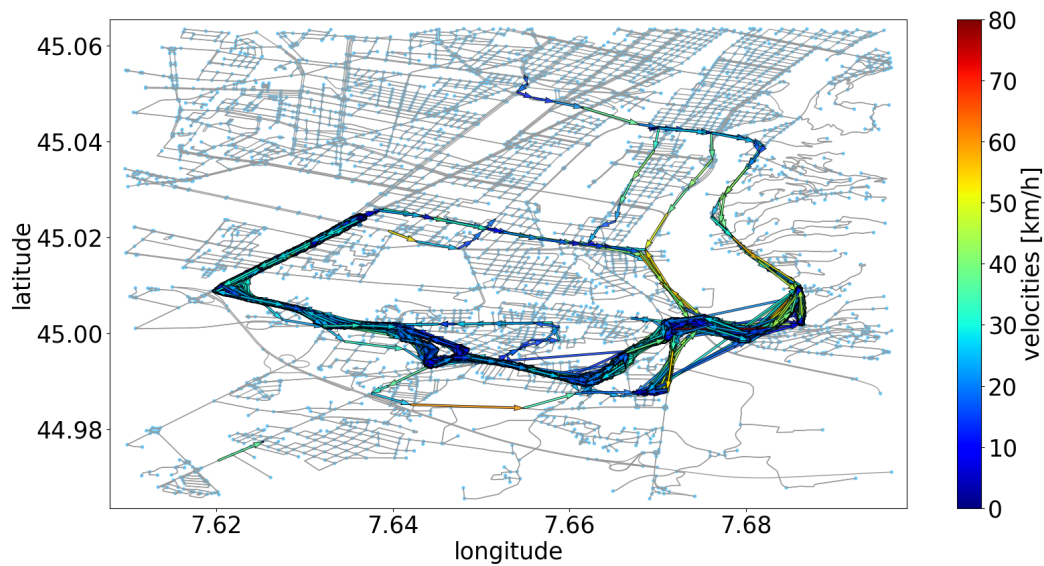


Figure 7. Velocities in the case of line 39 after the outlier removal.

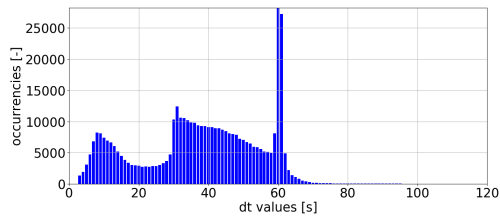


Figure 8. Histogram of  $\Delta t$  in the case of line 11 after the outlier removal.

### 3. RESULTS

The results for the line 11 are reported in Figure 9 and Figure 10 which show how the highest velocities occur at night and in late evening, with a local peak shortly after the lunch hour, for both working and weekend days. Furthermore, how it could be expected, the lowest velocities are recorded during the peak hours (7 - 9, 17 - 19). Finally, the differences between working and weekend days are more evident in the peak hours, whereas during the 0 - 5 and 21 - 24 time intervals the difference is smaller, since in these hours the traffic level is significantly lower also in the working days. The same behaviour can be observed also in the case of the line 39 (Figure 11 and Figure 12).

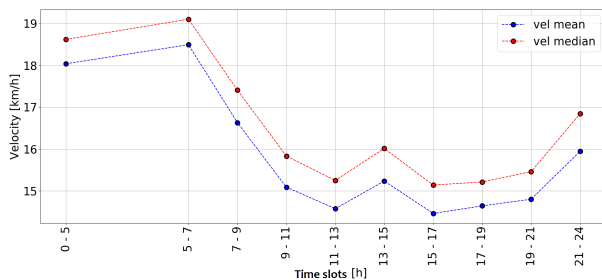


Figure 9. Line 11: time slot velocities in weekend days.

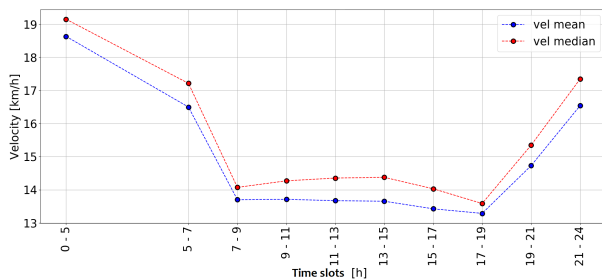


Figure 10. Line 11: time slot velocities in working days.

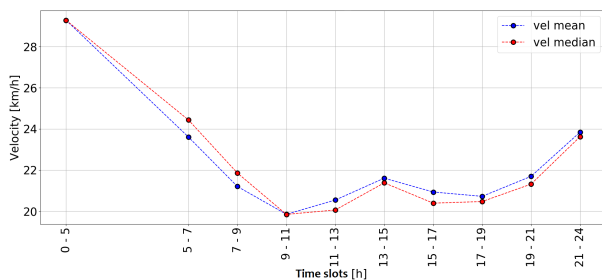


Figure 11. Line 39: time slot velocities in weekend days.

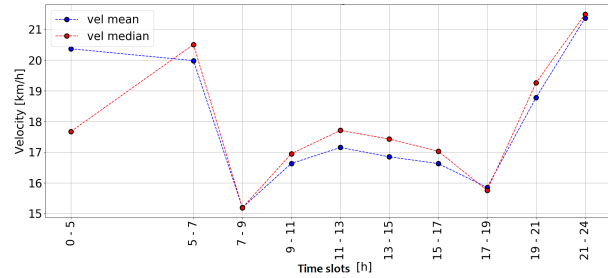


Figure 12. Line 39: time slot velocities in working days.

Moreover, in order to map the GNSS trajectories to the road network, it is necessary to assign the velocities to the line network topology. For this purpose, some investigations were performed and a preliminary strategy was implemented: for every point of the FCD, the closest tree of the specific line network is searched, as Figure 14 shows. In this way, the point can be assigned to the selected tree and thus projected on the route network. At the same time, though, this strategy can cause some issues when the point to assign to the line network is located in a segment in which the distance between two (or more) trees is comparable to the GNSS measurement errors (Figure 13).

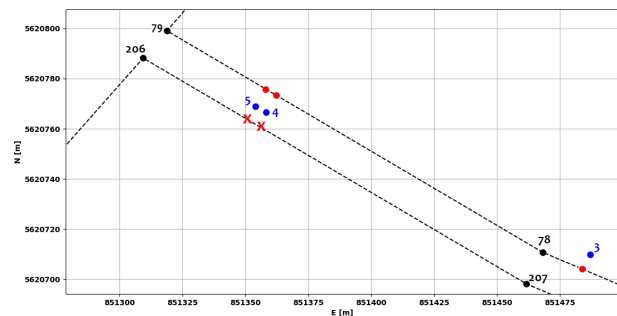


Figure 13. Network projection strategy: in black the line network, in blue the original FCD points, in red the corresponding projections within the line network: the cross markers represent the wrong projections, the dot markers represent the right projections.

A possible solution is to consider the cardinality information contained in the line network and the temporal information contained in the FCD by selecting the tree that is closest to the previous selected tree. For example, it is rather improbable that the FCD points 4 and 5 in Figure 13 may be assigned to the tree 206-207, since the vehicle was located in the tree 77-78 few moments before.

### 4. CONCLUSIONS AND PROSPECTS

A first strategy to analyse the FCD of the Turin Public Transportation system was implemented, in view of an automatic and possible real-time impedance map generation.

A huge amount of FCD were processed to compute the velocities of the vehicles and a visualization approach based on Osmnx library was adopted. Furthermore, a preliminary temporal analysis was carried out, showing higher velocities in weekend days and not peak hours, as could be expected.

Finally, a method to assign the velocities to the line network topology was developed and successfully tested. Further tests,

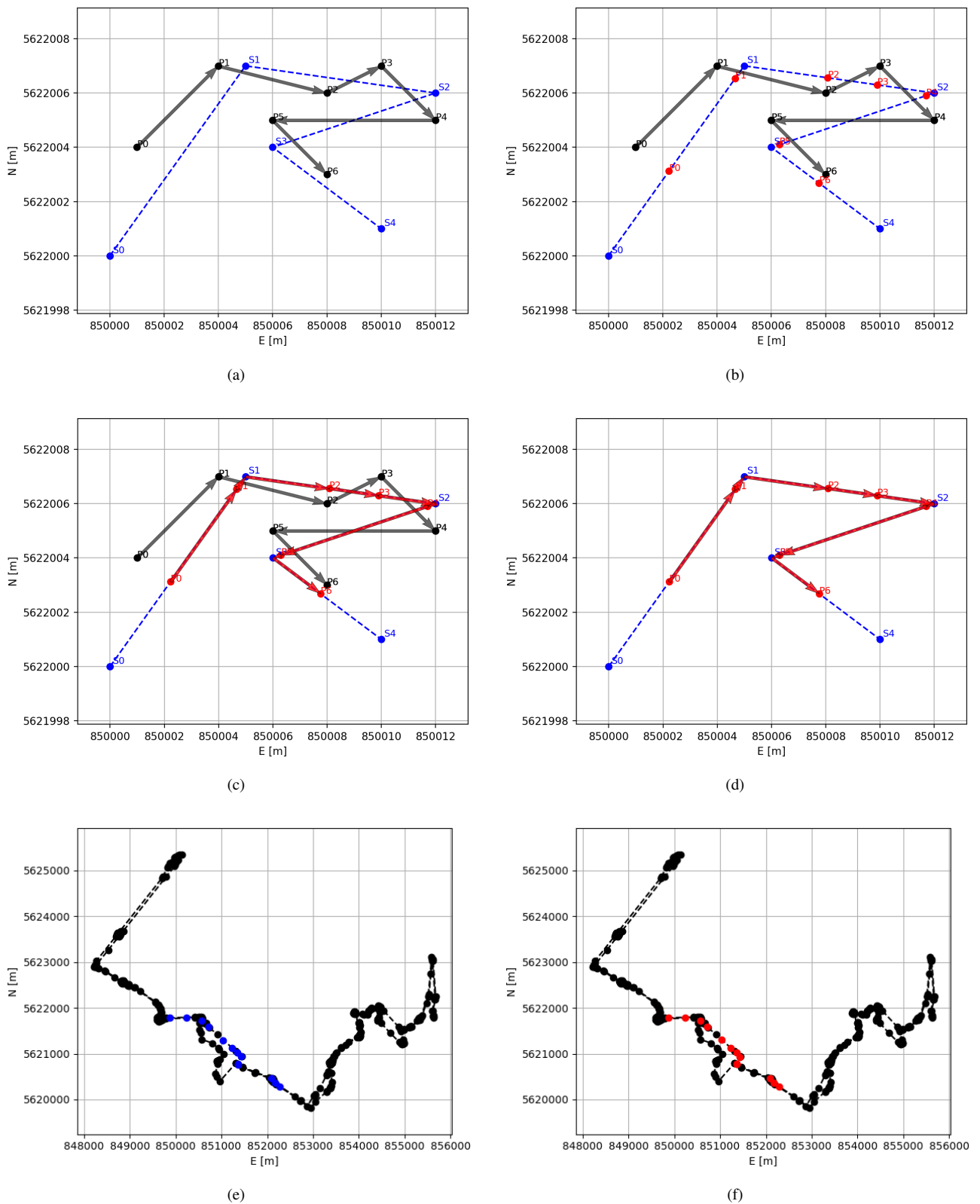


Figure 14. (a) The projection of FCD points within the line network: the black dots represent the original FCD and the black arrows the computed velocities; in blue the stops of the line route. (b) The red dots are the projection of the FCD within the line route. (c) and (d) the red arrows are the projected velocities. (e) Original FCD for vehicle 363 in service for line 39. (f) Projected FCD for vehicle 363 in service for line 39.

which will be developed on the available data, are however needed to check the effective reliability and real-time feasibility of the designed methodology.

#### ACKNOWLEDGEMENTS

This work was supported by URBAN-GEO BIG DATA, a Project of National Interest (PRIN) funded by the Italian Ministry of Education, University and Research (MIUR) id. 20159CNLW8. Furthermore, the authors desire to thank Prof. Piero Boccardo of the Polytechnic University of Turin for providing the FCD.

#### REFERENCES

- Boeing, G., 2017. Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems* 65, pp. 126–139.
- Breunig, M. M., Kriegel, H.-P., Ng, R. T. and Sander, J., 2000. Lof: identifying density-based local outliers. In: *ACM sigmod record*, Vol. 29number 2, ACM, pp. 93–104.
- Pirotti, F., Ravanelli, R., Fissore, F. and Masiero, A., 2018 - *under review*. Implementation and assessment of two outlier detection methods over large spatial point clouds. *Open Geospatial Data, Software and Standards Editorial Office*.
- Vincenty, T., 1975. Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey review* 23(176), pp. 88–93.
- Yang, C. and Gidófalvi, G., 2018. Mining and visual exploration of closed contiguous sequential patterns in trajectories. *International Journal of Geographical Information Science* 32(7), pp. 1282–1304.