



# Analysis of the Floating Car Data of Turin Public Transportation system: first results

Roberta Ravanelli, Mattia Crespi

University of Rome "La Sapienza"  
Department of Civil, Constructional and Environmental Engineering  
Geodesy and Geomatics Division

Rome, 22<sup>nd</sup> January 2018

PRIN PROJECT: URBAN GEOmatics for Bulk Information Generation, Data Assessment and  
Technology Awareness



# Floating Car Data (FCD)

- ▶ We started to analyse the **Floating Car Data (FCD)** of **Turin Public Transportation** system, carried out by the GTT company
- ▶ The **data** were acquired by every vehicle of the fleet through its On Board Unit (OBU) in the **month** of **April 2017**, with a **variable time interval** (difference of several seconds)
- ▶ The **data** are provided in the **CSV** format and include the **geographical coordinates** along with a **set of attributes** (vehicle code, line code, turn, timestamp, ecc.)
- ▶ The original file is very **heavy** (2.19 GB) and it was converted in a **database** through a Python script based on the `sqlite3` and `pandas` libraries

# Database generation

About 30·000·000 records!

DB Browser for SQLite - B:\PRIN\_GEO\_BIG\_DATA\Data\_floating\_car\_data\_Boccardo\scd\csc\_database\_formatted.db

File Edit View Help

New Database Open Database Write Changes Revert Changes

Database Structure Browse Data Edit Fragmas Execute SQL

Table: fcd\_table

	index	linea	turno(?)	date	mezze	lat	lon
	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	1	36	5	2017-04-28 21:05:09.000000	802	45.073677062...	7.5964450836...
2	2	64	1	2017-04-28 11:10:02.000000	3041	45.064193725...	7.6750035285...
3	3	51	2	2017-04-27 08:54:40.000000	977	45.119415283...	7.7108950614...
4	4	6	3	2017-04-26 13:41:13.000000	6027	45.073696136...	7.6814966201...
5	5	44	3	2017-04-12 13:47:58.000000	9017	45.066818237...	7.5776481628...
6	6	58	3	2017-04-10 07:18:26.000000	2620	45.038261413...	7.6190347671...
7	7	5	5	2017-04-09 08:49:13.000000	1039	45.028236389...	7.6017150878...
8	8	81	2	2017-04-08 09:20:20.000000	1254	44.994644165...	7.7242064476...
9	9	11	17	2017-04-06 11:56:30.000000	948	45.124114990...	7.6440901756...
10	10	16CS	10	2017-04-24 19:16:25.000000	2857	45.072139739...	7.6556334495...
11	11	585B	22	2017-04-25 20:18:18.000000	2785	45.060665130...	7.6614084243...
12	12	63	6	2017-04-18 10:46:06.000000	2769	45.011482238...	7.6365866661...
13	13	725B	22	2017-04-15 19:34:18.000000	1000	45.095348358...	7.6690135002...
14	14	67	2	2017-04-13 11:31:45.000000	3007	45.004108428...	7.6849350929...
15	15	57	8	2017-04-12 22:27:30.000000	855	45.067314147...	7.6714982986...
16	16	3	23	2017-04-07 08:51:16.000000	5012	45.099411010...	7.6488001014...
17	17	13	8	2017-04-06 09:30:10.000000	2857	45.076423643...	7.6698732376...
18	18	5	8	2017-04-05 17:17:22.000000	800	45.056266284...	7.6644783020...

1 - 38 of 30215120

Go to: 1



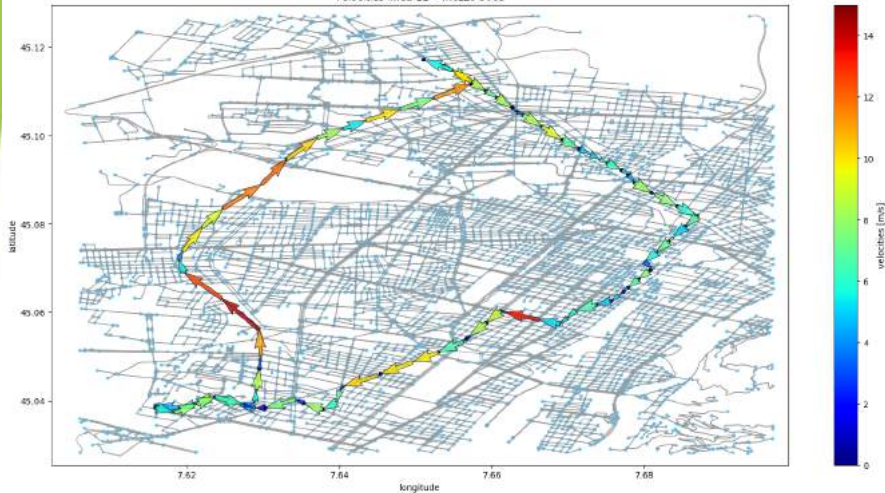
# Velocity analysis

- ▶ The **data** were **organized** for **lines**, then for **vehicles** and finally they were **chronologically ordered**
- ▶ For every line of the transportation network:
  - ▶ the *Vincenty* formula was used to compute the **planimetric displacement**  $\Delta s$  between **two positions** of the specific vehicle in two **consecutive time moments**
  - ▶ the **velocities** were computed as  $v = \frac{\Delta s}{\Delta t}$
- ▶ 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

```
73 def boh(drive_network_graph, start_lon, start_lat, delta_lon, delta_lat, velocities, soglia
74         fig_from_function, ax_from_function = ox.plot_graph(drive_network_graph, close = False
75         nz = mcolors.Normalize(vmin = soglia_inf, vmax = soglia_sup)
76         #plt.gca().set_aspect('equal', adjustable='box')
77         plt.quiver( start_lon, # start x
78                   start_lat, # start y
79                   delta_lon, # delta x
80                   delta_lat, # delta y
81                   angles='xy', # 'xy': arrows point from (x,y) to (x+dx, y+dy). Use this for
82                   scale=1, # piú è grande, + le frecce sono corte Number of data units per a
83                   scale_units='xy', # usando le scale units, non è piú necessario alterare l
84                   color=cm.jet(nz(velocities)), # color = velocities
85                   zorder = 5, #piú è alto, piú il plot è in primo piano
86                   edgecolor='k', # colore bordo freccia
87                   linewidth=.7,
88                   alpha=0.8) # trasparenza
```

# Example of computed velocities

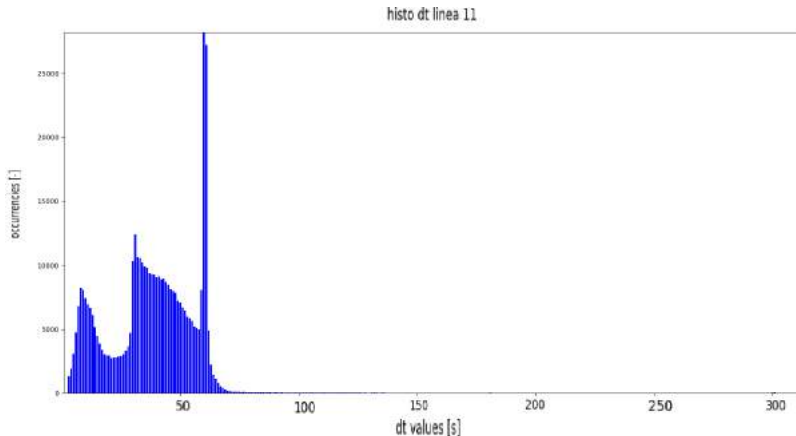
Velocities linea 11 - mezzo 3063



# Outlier removal

Before proceeding with the time analysis, the **outliers** were **removed** by eliminating all the records:

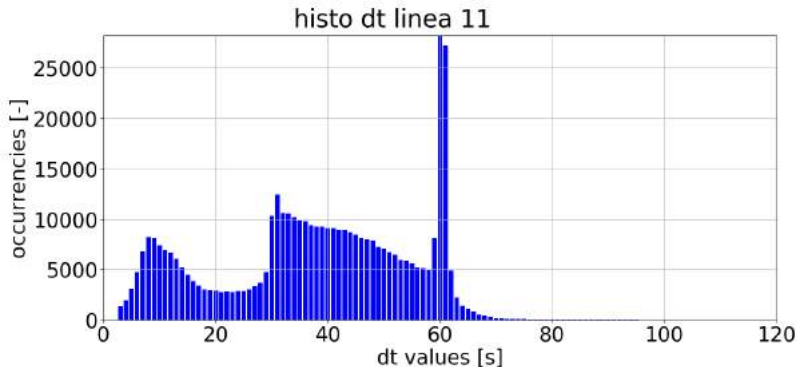
1. whose  $\Delta t$  are higher than 99.5<sup>th</sup> percentile and lower than 0.5<sup>th</sup> (statistically not significant)
2. characterized by a velocity higher than 5 times the mean



# Outlier removal

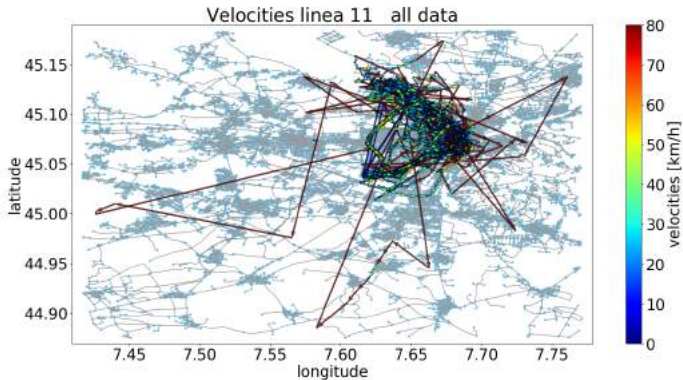
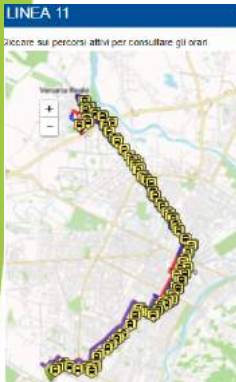
Before proceeding with the time analysis, the **outliers** were **removed** by eliminating all the records:

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



# Line 11: velocities

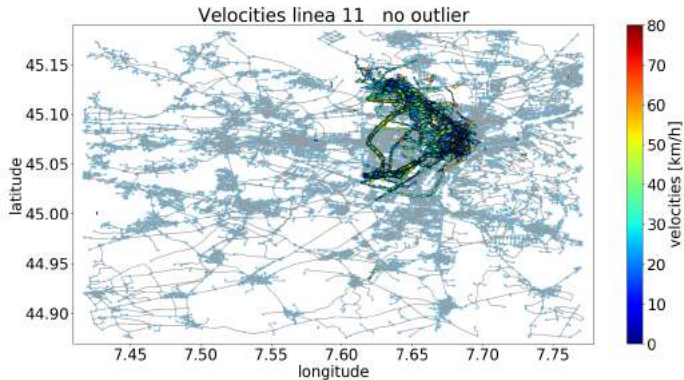
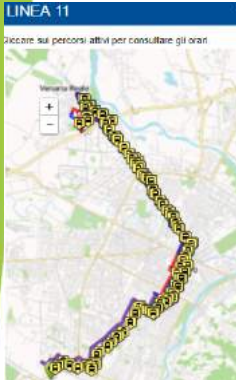
After the **outlier removal**, the reconstructed path follows more closely the actual line route: the **longest arrows**, probably due to the bus routes from and to the depot, are **eliminated**





# Line 11: velocities

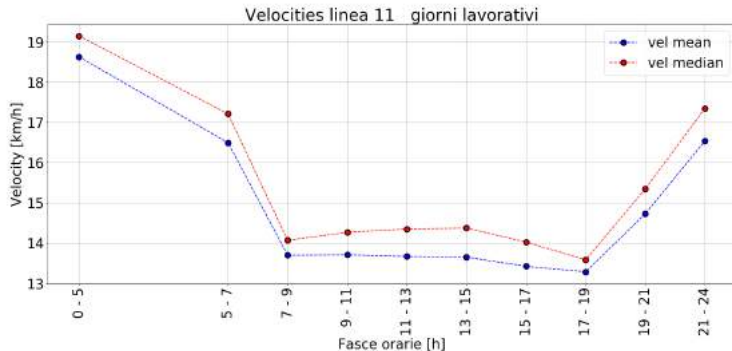
After the **outlier removal**, the reconstructed path follows more closely the actual line route: the **longest arrows**, probably due to the bus routes from and to the depot, are **eliminated**



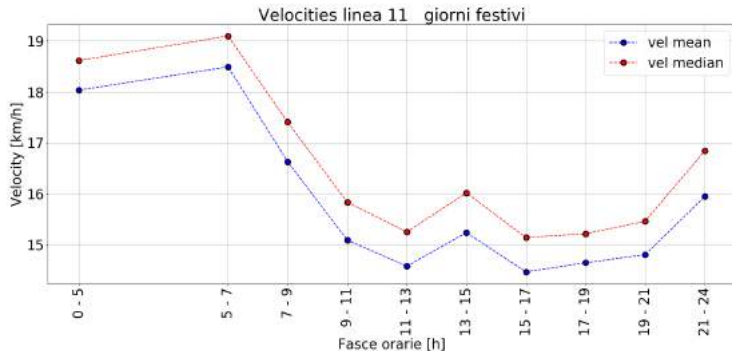
# Temporal analysis

- ▶ Once the outliers were removed, a **temporal analysis** was performed
- ▶ The data were divided into **working** and **weekend days**, considering the following **time slots**:
  - ▶ 0 - 5
  - ▶ 5 - 7
  - ▶ 7 - 9
  - ▶ 9 - 11
  - ▶ 11 - 13
  - ▶ 13 - 15
  - ▶ 15 - 17
  - ▶ 17 - 19
  - ▶ 19 - 21
  - ▶ 21 - 24

# Line 11: time slot velocities in working days

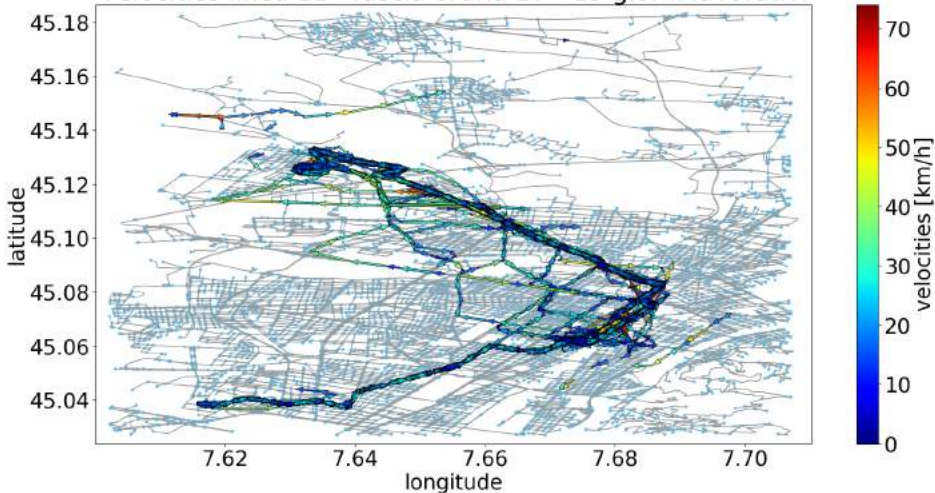


# Line 11: time slot velocities in weekend days



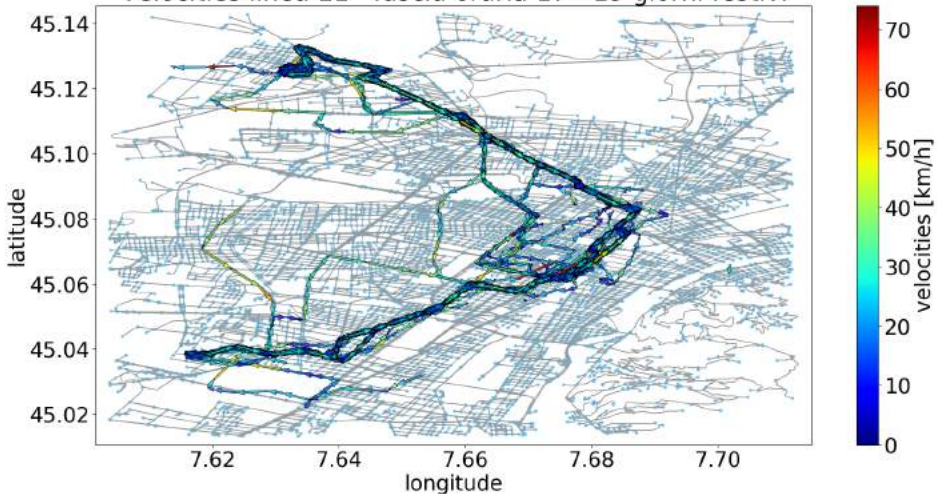
# Line 11: velocities in the time slot 17 - 19

Velocities linea 11 fascia oraria 17 - 19 giorni lavorativi



# Line 11: velocities in the time slot 17 - 19

Velocities linea 11 fascia oraria 17 - 19 giorni festivi

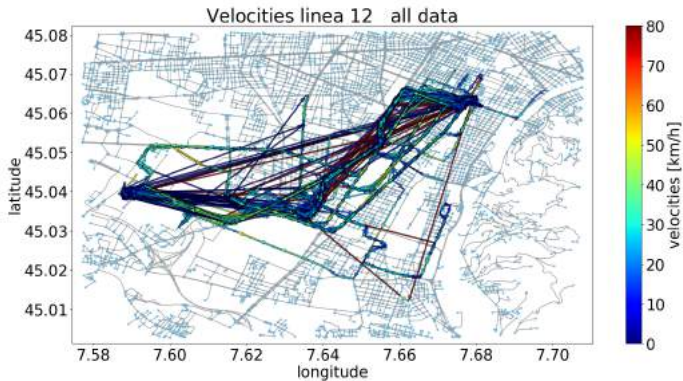
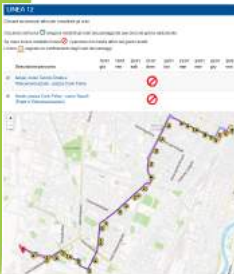


# Considerations

It can be noticed that:

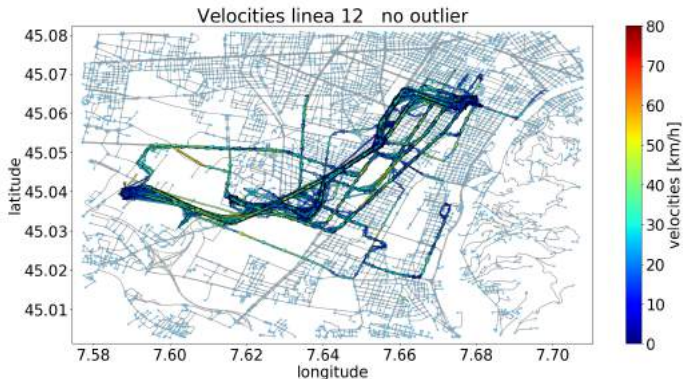
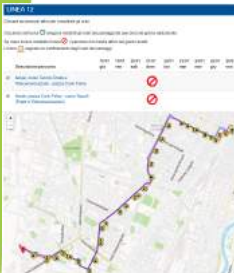
- ▶ the **highest velocities** occur at **night** and in **late evening**, with a **local peak** shortly after the **lunch hour**
- ▶ the **lowest velocities** occur during the **peak hours**, in correspondence of the **office entrance** and **exit hours**
- ▶ the differences between **working** and **weekend days** are more **evident** in the **peak hour** time slots
- ▶ during **0-5** and **21-24** time slots the **difference** is **small**, since in these hours the **traffic** level is significantly lower also in the **working days**

# Line 12: velocities

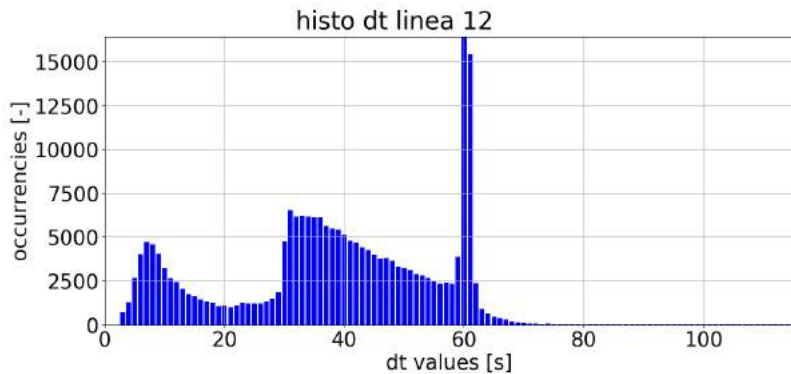




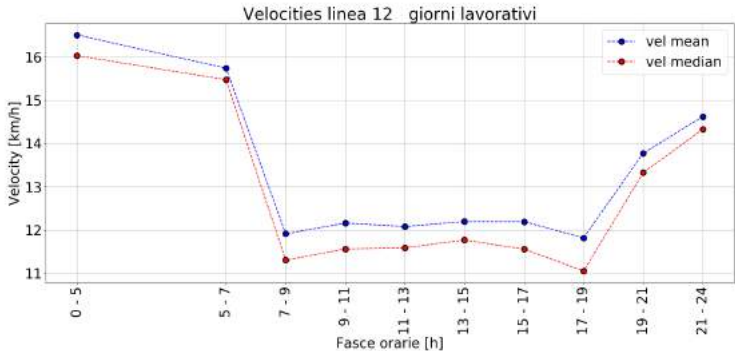
# Line 12: velocities



## Line 12: $\Delta t$ histo

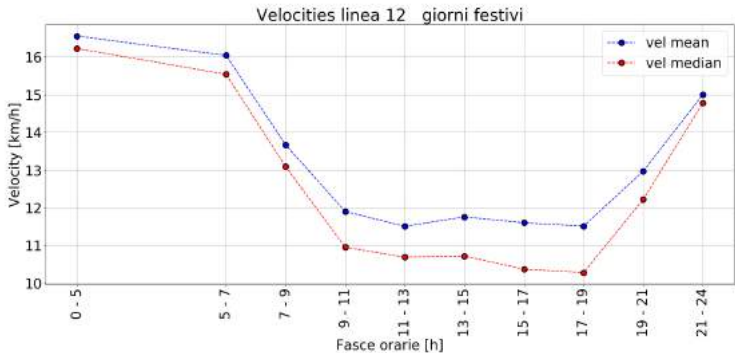


# Line 12: time slot velocities in working days



# Line 12: time slot velocities in weekend days

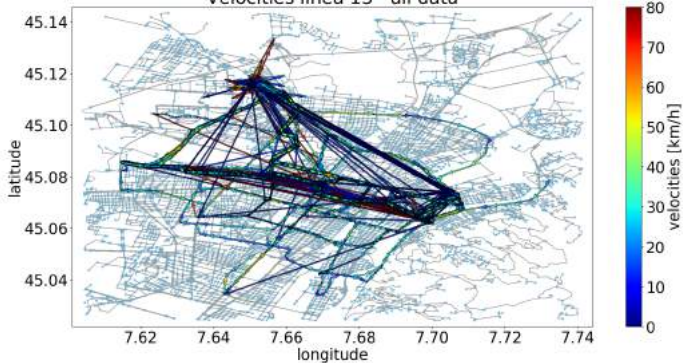
The line is not active on Sundays



# Line 13: velocities

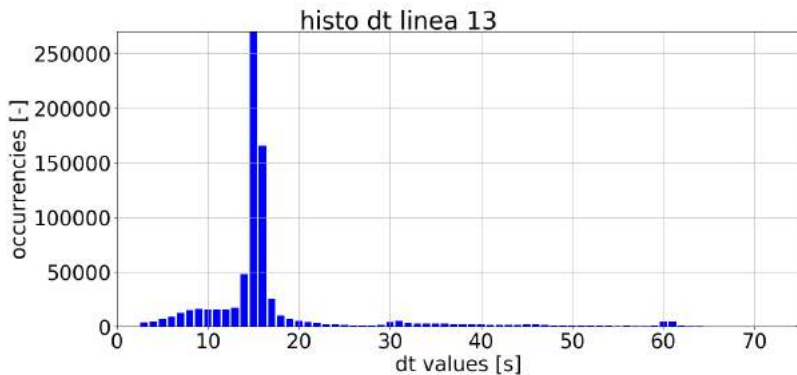


Velocities linea 13 all data

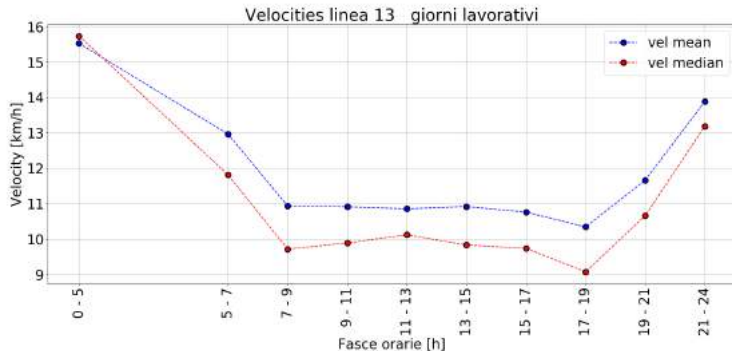




## Line 13: $\Delta t$ histo



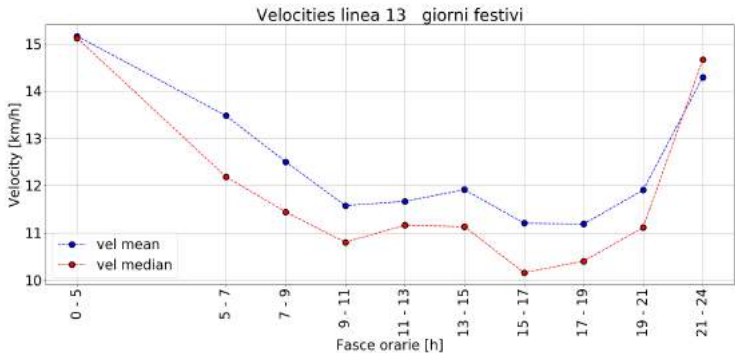
# Linea 13: time slot velocities in working days





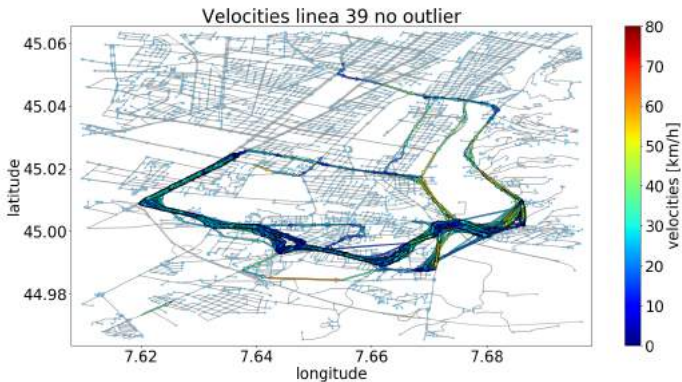
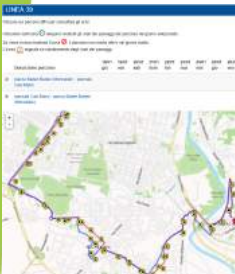
# Linea 13: time slot velocities in weekend days

The line is not active on Sundays

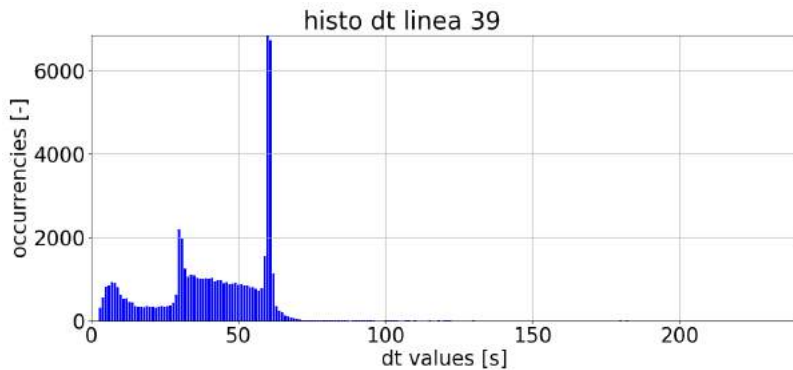




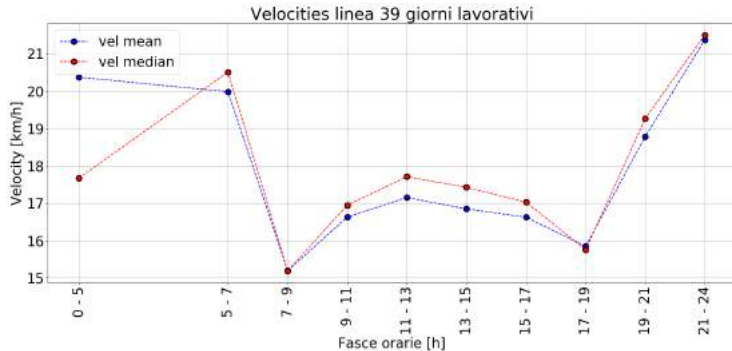
# Line 39: velocities



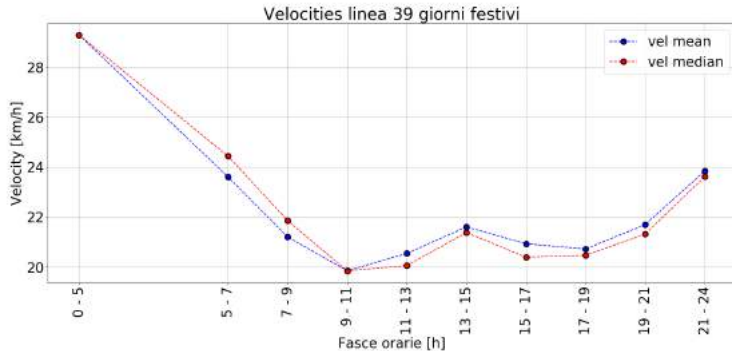
## Line 39: $\Delta t$ histo



# Line 39: time slot velocities in working days

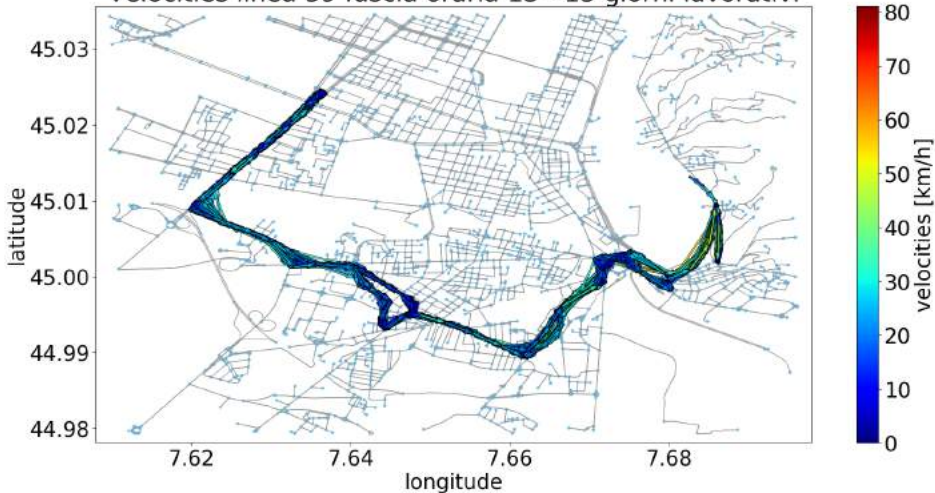


# Linea 39: time slot velocities in weekend days



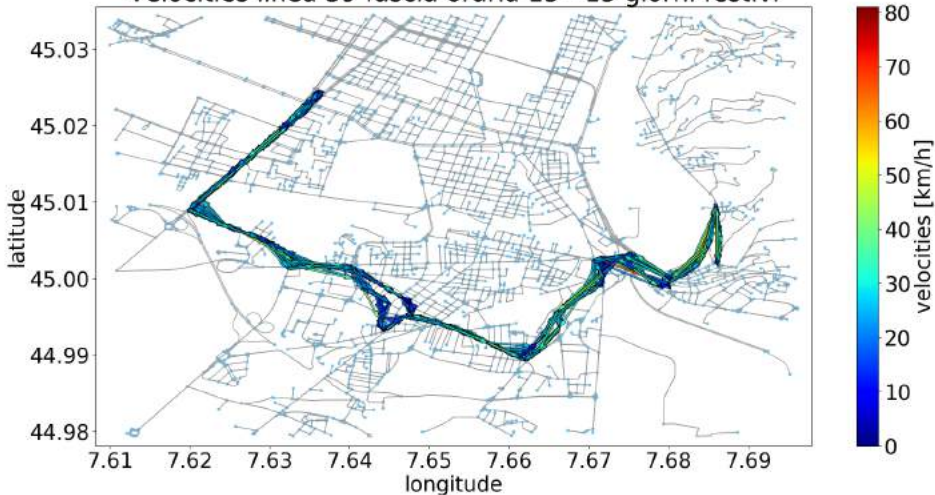
# Line 39: velocities in the time slot 13 - 15

Velocities linea 39 fascia oraria 13 - 15 giorni lavorativi



# Line 39: velocities in the time slot 13 - 15

Velocities linea 39 fascia oraria 13 - 15 giorni festivi





# Open issues

- ▶ To deepen the temporal analysis
- ▶ Presence of **velocities not referable** to the **actual path** of the **lines**
- ▶ To **assign** the **velocities** to the **line network topology**



**Necessity** to obtain **graphs** describing the **path** of every **line**

**Thank you for your kind attention!**

