URBAN GEO BIG DATA FOR MOBILITY APPLICATIONS

POLITO LOCAL UNIT

Outline

1. Short recap
2. Mobility applications: urban public and private transportation
3. Available data
4. Processing, outputs and discussion of the results
5. Data and services deployment
6. Further development

Short recap

The mobility research project will take into account **IBA** (**Infomobility Based Applications**) and **LBS** (**Location Based Services**) mobility applications defining a methodological and operational approach (encompassing Geo Big Data in a possible crowdsourcing mode for the five cities where the local research units are based), splitted in different steps:

- Collection/harvesting of authoritative Geo Big Data and Open Source ones (vector and raster and textual if coming from social media), including crowdsourcing based data acquisition;
- Data quality control both in terms of geometric and thematic precision, accuracy, completeness, etc.;
- Implementation of a common data model;
- Deployment of a shared raw data geodatabase;
- Processing algorithms and techniques for added value information extraction from raw data;
- Data and Information publication in a dedicated environment based on OS Web GIS application.

Mobility applications: urban public and private transportation

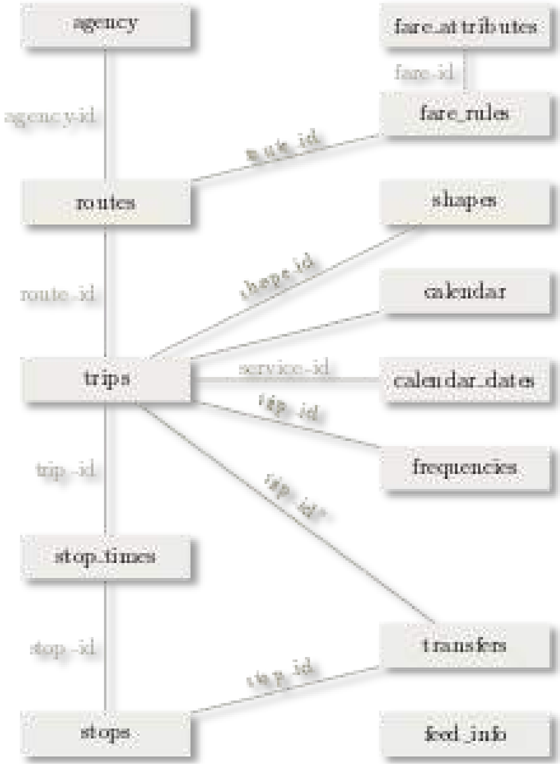
When taking into consideration mobility applications two different components should be ancompassed:

- **Public transportation** (**TPL**) refers to all the different mobility modes (road, rail, underground, air, water) where public interest, in terms of service deployment, is protected;
- **Private mobility** (**PM**) refers to different mobilty modes (mainly road and water in Italy) where transportation is related to individuals and should be regulated in order to guarantee the same rules and traffic conditions to everyone.

TPL, from a geomatics point of view, is approached using **geo data models encompassing all the necessary data able to return useful mobility information**, PM using data acquired by **sensor networks** (both on the ground or aerial) and if not available, by **On Board Units** (**OBU**), that are sensors mounted on board of vehicles acquiring similar data.

TPL Available data

For this component, geo data models encompassing all the necessary data able to return useful mobility information is the **General Transit Feed Specification** (**GTFS**) that defines a common format for public transportation schedules and associated geographic information

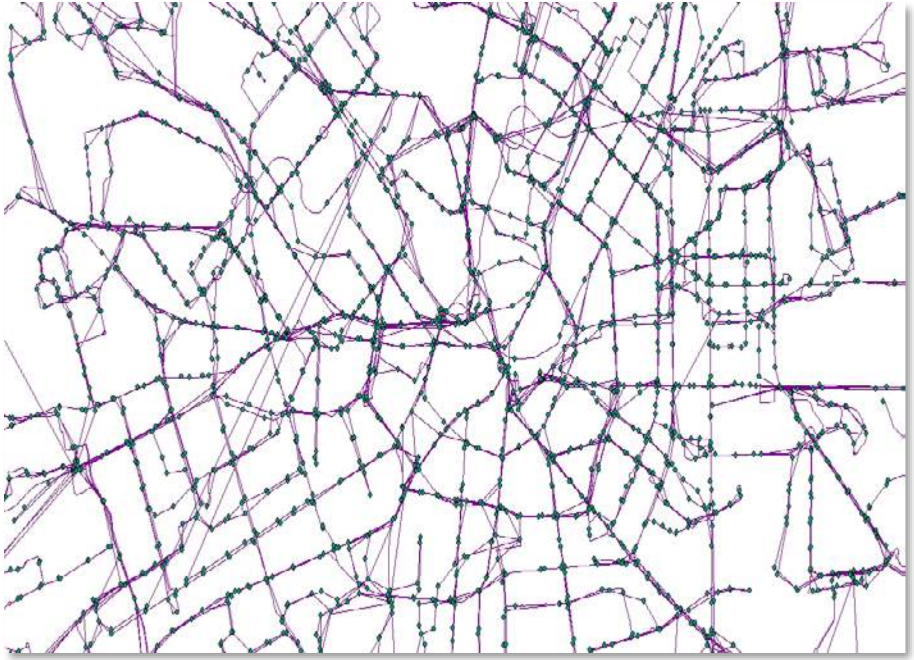


TPL Available data

GTFS format is composed by two different components:

1. **Geo layers** (routes and stops for any possible transportation mode), divided into line and point ones;

2. Attribute tables that can refer to different data associated to geo layers.



TPL Available data

GTFS format is composed by two different components:

1. Geo layers (routes and stops for any possible transportation mode), divided into line and point ones;

2. **Attribute tables** that can refer to different data associated to geo layers.

Filename	Required
agency.txt	Required
stops.txt	Required
routes.txt	Required
trips.txt	Required
stop_times.txt	Required
calendar.txt	Conditionally required
calendar_dates.txt	Conditionally required

TPL Available data

Milano

Geo layers available in WMS;

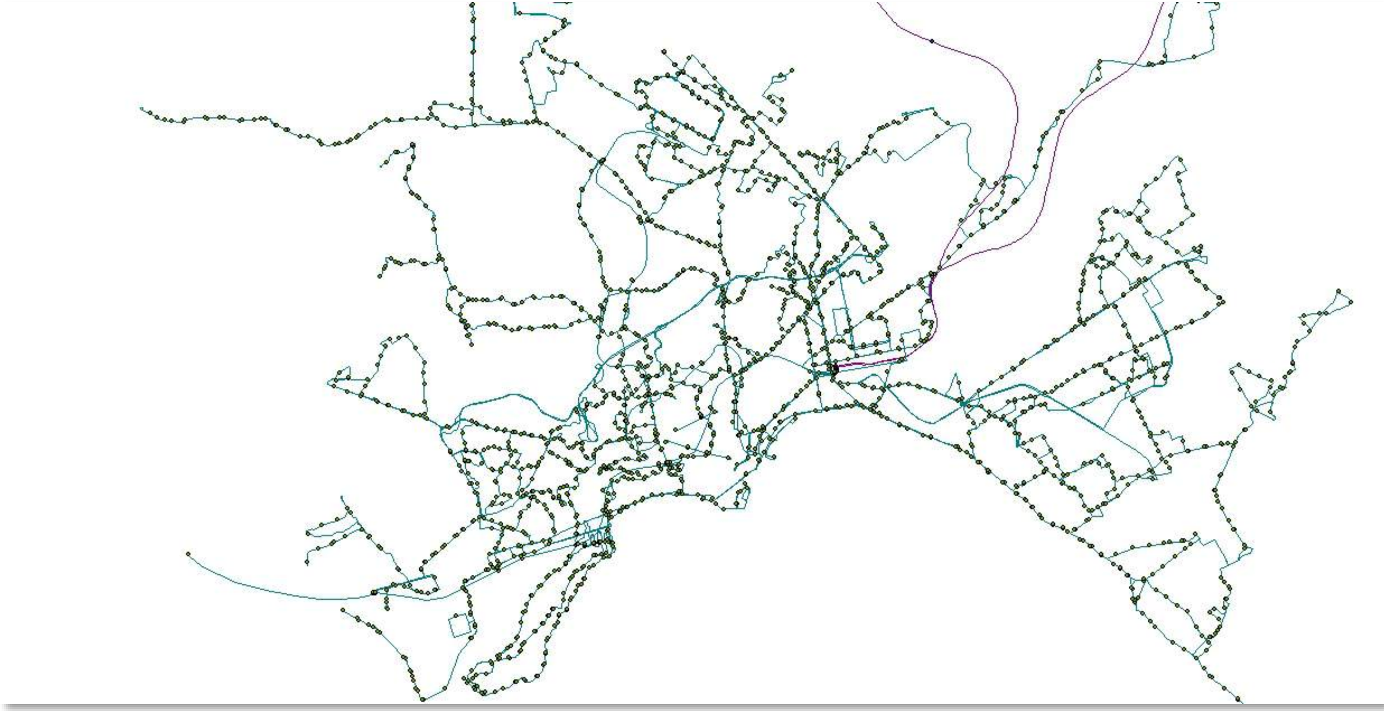
Attribute tables as ext. files

TPL Available data

Napoli

Geo layers available in WMS;

Attribute tables as ext. files



TPL Available data

Roma

Geo layers available in WMS;

Attribute tables as ext. files

TPL Available data

Torino

Geo layers available in WMS;

Attribute tables as ext. files

PM Available data

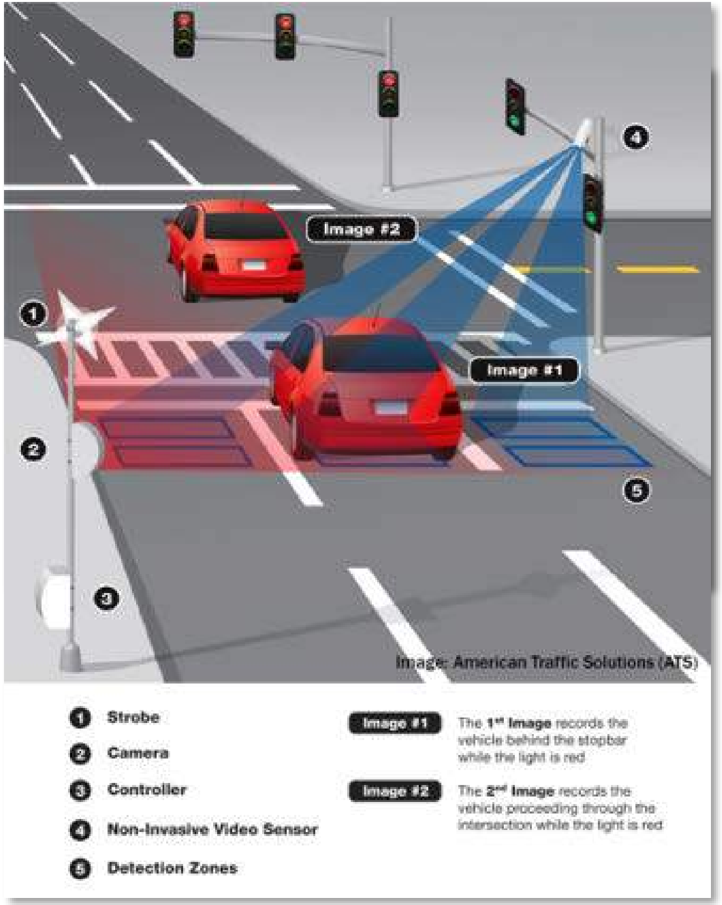
Population travel behaviour modelling is a fundamental process in transportation planning and in the management of urban transportation systems. It plays a crucial role in developing strategies that help **alleviate urban traffic congestion and support traffic management during special events and emergencies**.

Recently, **Floating Car Data (FCD)** has been used to monitor and model urban mobility; it is based on the collection of **georeferenced data**, by means of **GNSS receiver, inertial platforms, accelerometers and odometers**, regarding speed, direction of travel and time information from **on board unit (OBU)** in vehicles that are being driven.

PM Available data

This data collection technique is becoming more and more relevant for mobility domain applications, in order to overcome some specific issues:

1. **road network fixed sensors** (based on induction loops or aerial configuration) able to collect similar data are **not always sufficiently distributed** over a given area, and their installation and maintenance is rather costly;
2. **road network fixed sensors** acquire different type of data that, **not always, are characterized by a given homogeneity**, conveying to the impossibility to analyse, with standard procedures, this data over larger areas;
3. **road network fixed sensors** are usually installed over main roads, being **not able to analyse data also in collector's and local roads; that means, for example, the impossibility to route and/or to build up Origin/Destination** (O/D) matrix for consistent areas and/or for whole cities.



PM Available data

Due to the fact that an increasing number of vehicles are equipped with a "**black box**" that contains a GPS receiver, (typically fleets such as courier and freight services and private cars where this system allows insurance policy consistent savings), **acquired data can definitely help in trying to solve the previously mentioned issues**.

This is also possible mainly because **collected data**, being transmitted to a control centre by using mobile phone network and/or on-board radio unit, **are available in almost real time** for further processing, having a reduced latency time related to network downloading speed. Polito Local Unit, processed a significant amount of FCD data acquired over the cities part of the Prin project.

PM Available data

FCD data are acquired by **On-Board Unit** (**OBU**) mounted on board of vehicles, typically private cars linked to insurance policies, and trucks/vans managed in a fleet environment. One of the main information acquired by OBU is the position, acquired by means of a GPS receiver, using **both a temporal and speed sampling interval**: if longer distances are driven, shorter is the temporal sampling time, and vice versa. Every acquired record is composed by different fields:

- time stamps,
- latitude, longitude,
- speed, heading,
- Horizontal Dilution Of Precision (HDOP),
- engine status,
- vehicle type (private car/fleet).

Every single record is then assigned to a **unique ID representing a unique vehicle**. The field ID-DEVICE represents the identifier of the OBU from which the record is generated and can be used as a proxy for unique vehicle. Acquired data are then transmitted in almost real-time to a data centre via mobile network or radio connection.

PM Available data

FCD data have been acquired over the five different cities during the same week (7 days, considering a week characterized by the at least 5 sunny days). In particular data have been acquired from October the 5th to the 11th , 2018, with the total number of data as reported below:

city	count
Padova	801866
Napoli	8150826
Milano	3361692
Torino	4000656
Roma	12979442

Data processing

Below, the different steps carried out:

1. discuss and document the **FCD data model** and acquisition mode;
2. **conciliate GPS positions with available Open Source (OS) road networks**. Due to the intrinsic positional accuracy of the code coordinates GPS acquisition, planimetric position could be affected by positional errors (ranging from few centimetres to some meters) and, due to GPS multipath or shadowing effects, the coordinates could be affected also by gross errors;
3. **transfer mobility impedances**, intended as the relationship between road traveling time and traffic load pressure, **calculation**. Taking into account the data model defined at point 2), data are updated, assigning to every single arc of the road network travelling times and traffic load derived from the aggregation of all journeys (subdivided into private cars and fleet vehicles) and their timestamps;

Data processing

4. using the updated dataset, **different travel behaviour characteristics are analysed**:

   a) flows and velocity over every single arc, using the total number of vehicles associated to the road network;
   b) flows and velocity over every single arc
   c) traffic profiles, in different period of the day , over road network arcs and nodes;
   d) O/D spatial interactive matrix generation and updating.

Data processing

In the framework of the present research activity, **Open Transport Map**[1] (**OTM**), that is an open dataset based on OpenStreetMap[2] data and accessible in a scheme compatible to INSPIRE Transport Network, has been used to generate a reference road network dataset.

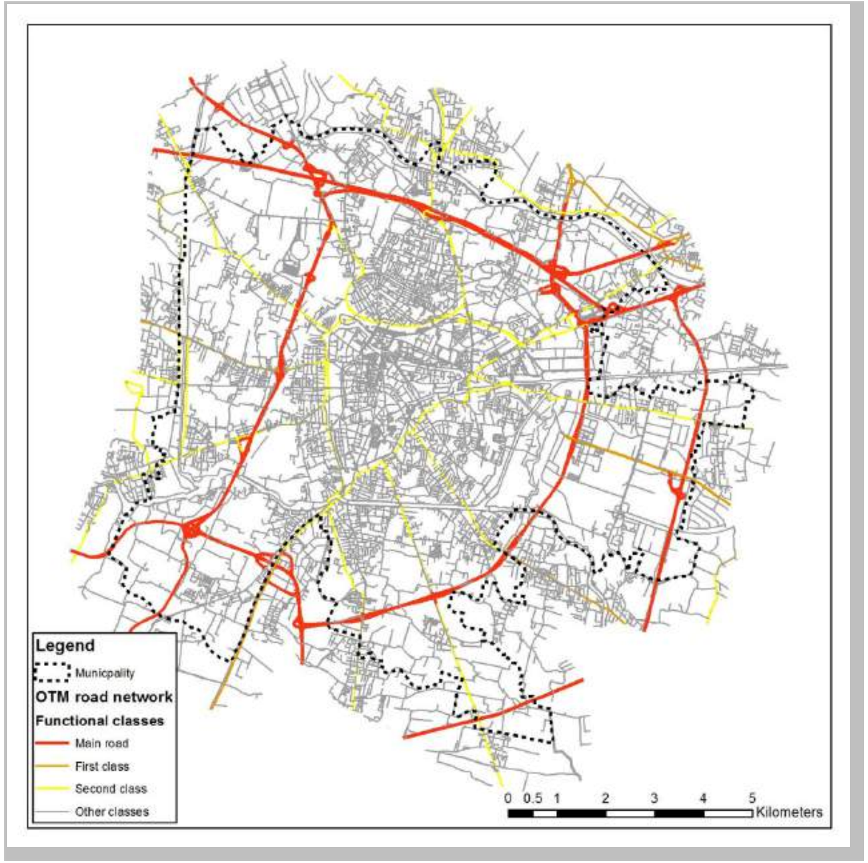
To process all the data related to the five cities, **two different approaches** have been implemented. The first one is related to general processing applied to all the cities involved, the second one to Torino specifically, in order to compare and calibrate FCD acquisitions with on site measurements, extending then to other cities where similar on site measurements are not available and/or not acquired.

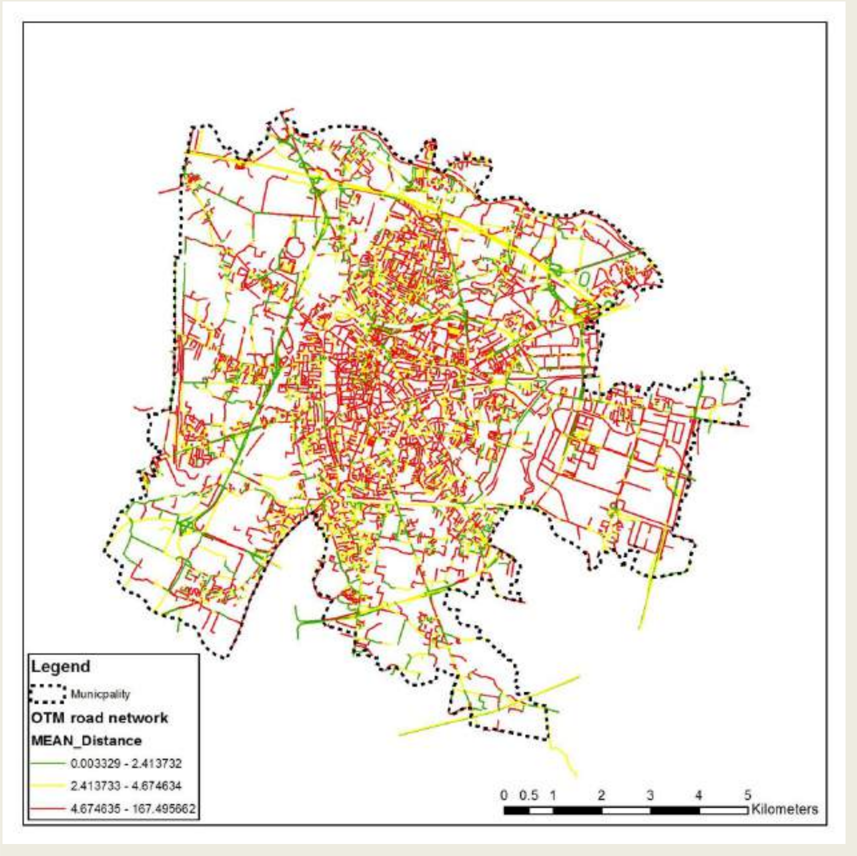
Data processing

Considering that the association of the FCD positions with digital maps of urban roads enables travel behaviour analysis, such as the estimation of speed and travel time of vehicles on different roads, for all the cities, the first two processing steps have been implemented:

1. **FCD** positions has been **uniquely assigned to a single OTM network** element by means of the identification of the nearest road element to the FCD position;
2. the **distance between the two elements has been calculated** and stored.

Data processing





Data processing

Analysing the statistical distribution of the distance between each FCD position and the nearest road element, standard deviation value (approx. 10 m) is considered **accurate enough** for the specific purposes of the analysis, i.e. to generate analysis to support mobility services at municipality level by mainly transforming single points to travel paths, minimising in this way the impact of outliers (e.g. an FCD points wrongly assigned to the incorrect travel direction is overcome by the possibility of generating correct path direction, exploiting **FCD timestamps**). Therefore, all FCD positions, including those with high positional errors, have been considered (Ravanelli et al, 2018; Pirotti et al., 2018).

Data processing

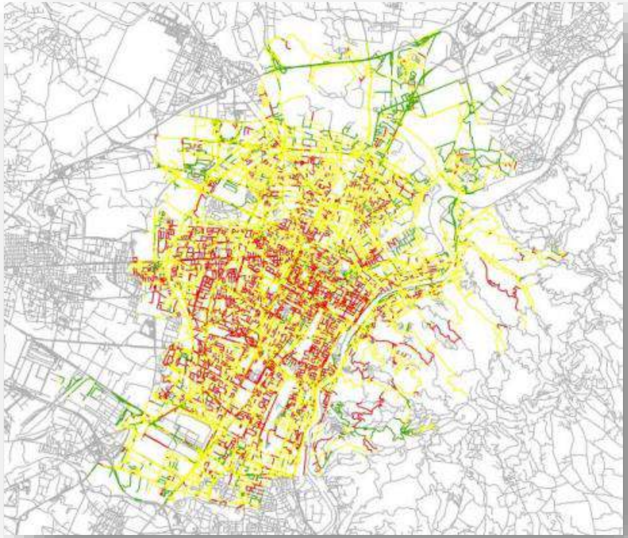
Having produced for all the five cities the reference data concerning OTM and FCD positions attributed to every single arc/node, from the methodological point of view, two different test should be carried on in order to consider FCD data statistically representativeness;

1. Investigate if there is **any correlation between FCD distance to arc (HDOP) and urban morphology**;
2. Determine if **FCD data** associated to every single arc/node are a **subsample of real measurement surveyed by fixed sensors**.

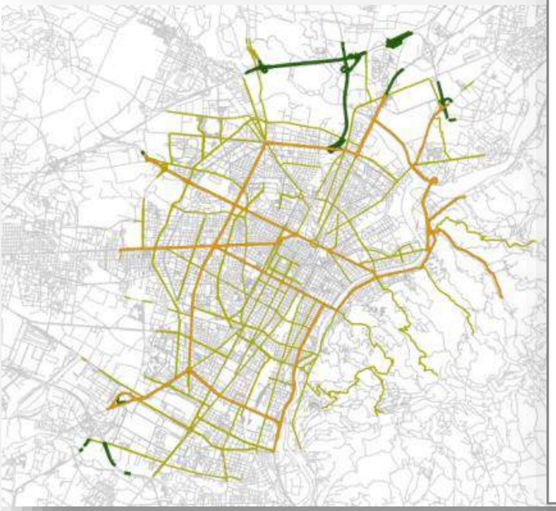
For the latter activities, **Torino use case** has been investigated having the availability of those measurements.

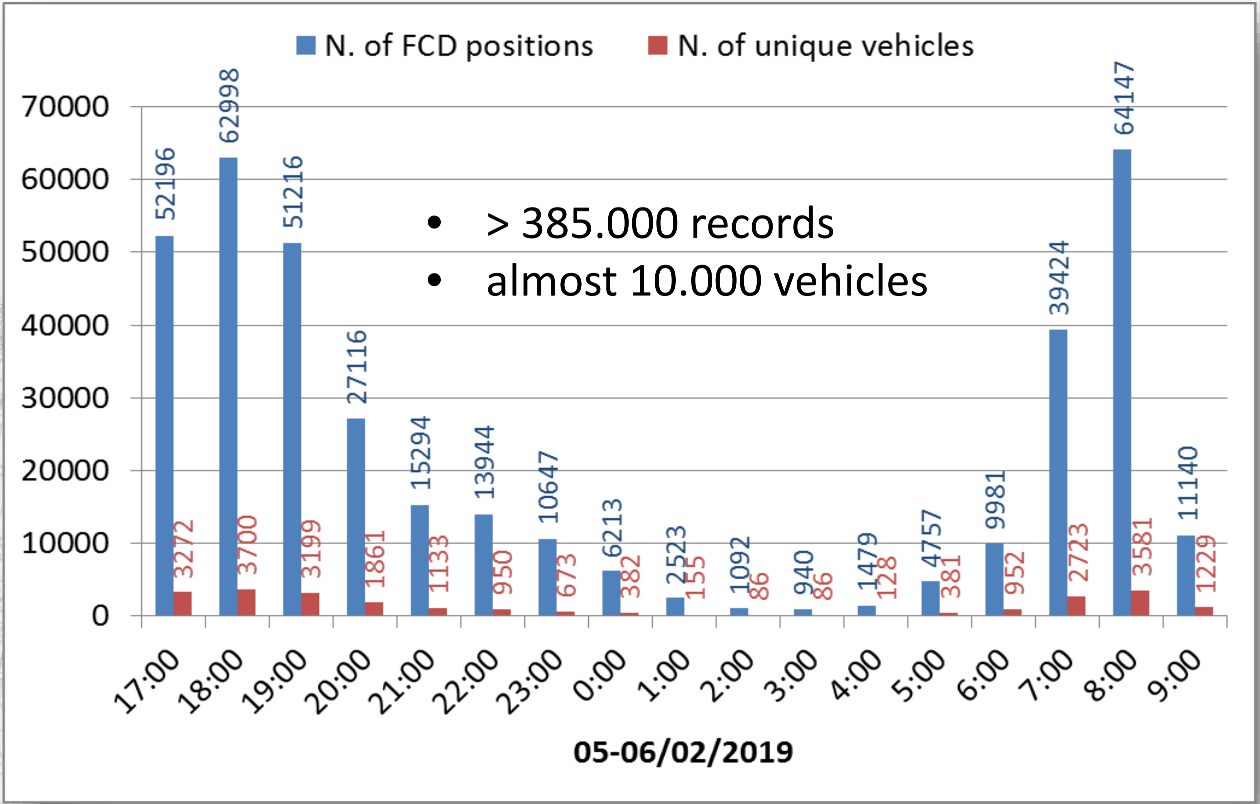
HDOP patterns

Horizontal Dilution Of Precision is affected by street canyons, in terms of GPS positioning accuracy decreasing caused by shadowing and multipath effects.
It is necessary to measure a possible correlation between this accuracy decreasing and building heights and density, in order to define any possible limit fo FCD usage.

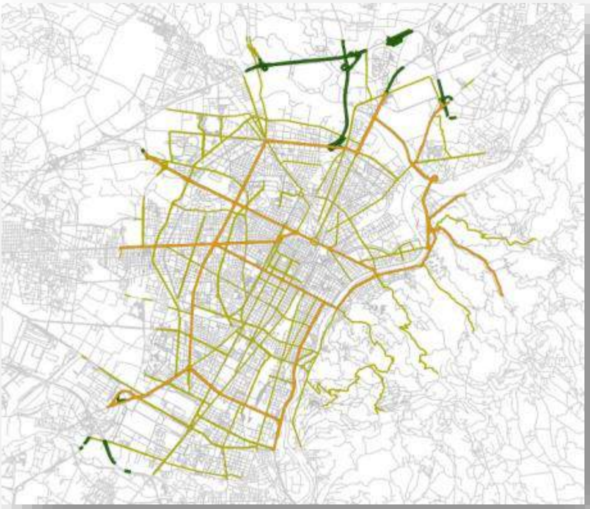


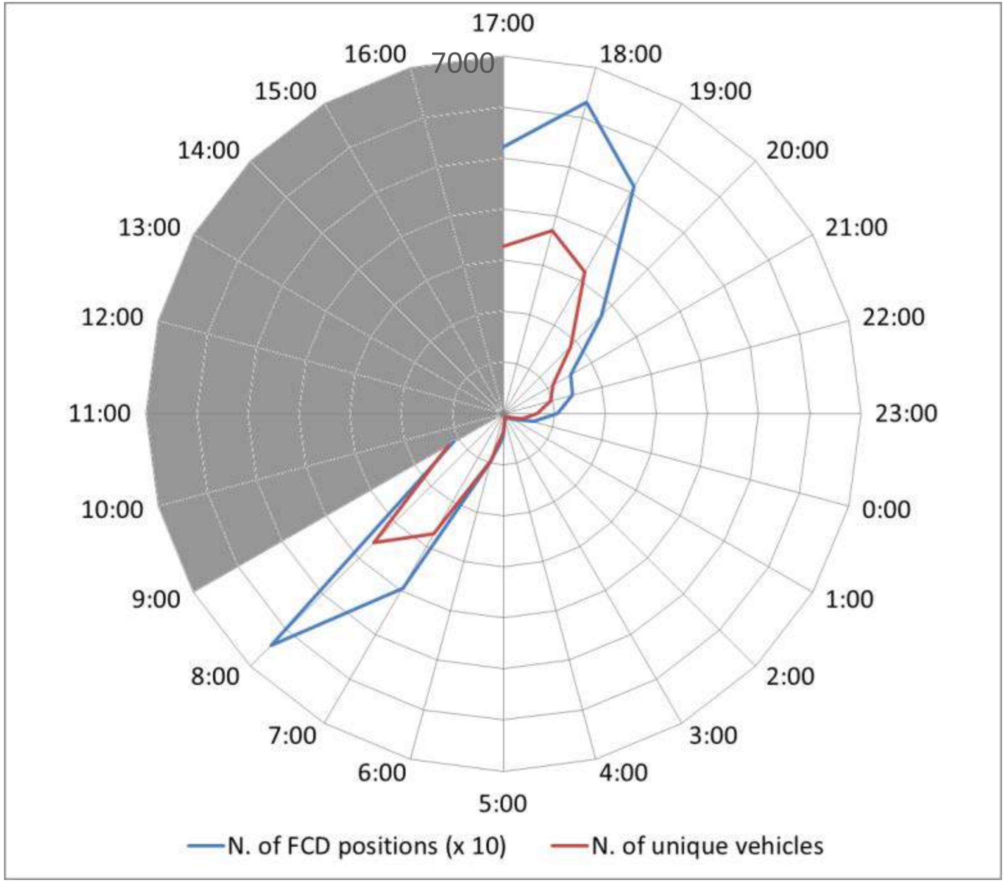
Flows analysis: Torino use case





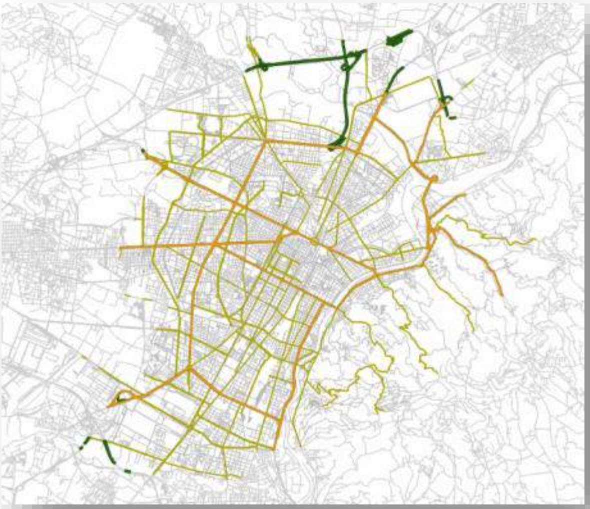
Flows analysis: Torino use case

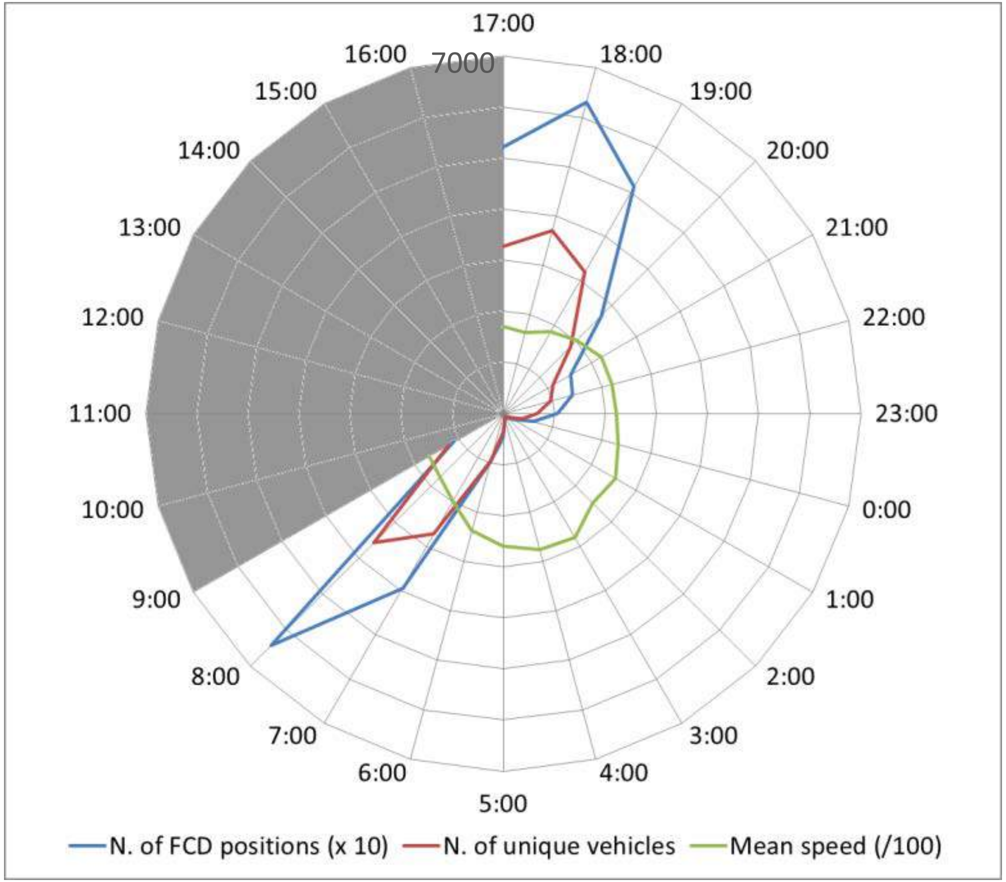




Fluxes analysis: Torino use case

URBAN-GEO BIG DATA





Comparison with traffic sensors

URBAN-GEO BIG DATA

Representativeness of the FCD sample

Freely available data from 124 sensors (out of the 3,400 installed in the municipality)

Comparison with traffic sensors

URBAN-GEO BIG DATA

Representativeness of the FCD sample

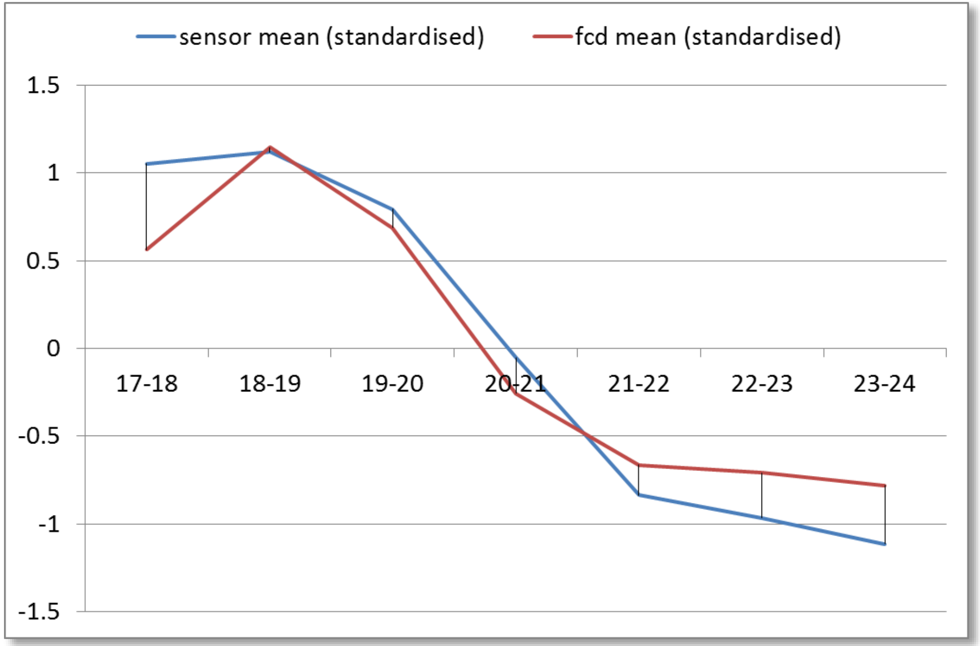
Freely available data from 124 sensors (out of the 3,400 installed in the municipality)

Time interval	N. of vehicles		
	Loop sensor	FCD	%
17:00-18:00	93841	1033	1.1
18-00-19:00	96121	1353	1.4
19:00-20:00	85507	1067	1.2
20:00-21:00	59688	506	0.8
21:00-22:00	36584	319	0.9
22:00-23:00	32866	242	0.7
23:00-24:00	27272	216	0.8
17:00-24:00	431879	4736	1.1

Comparison with traffic sensors

Representativeness of the FCD sample

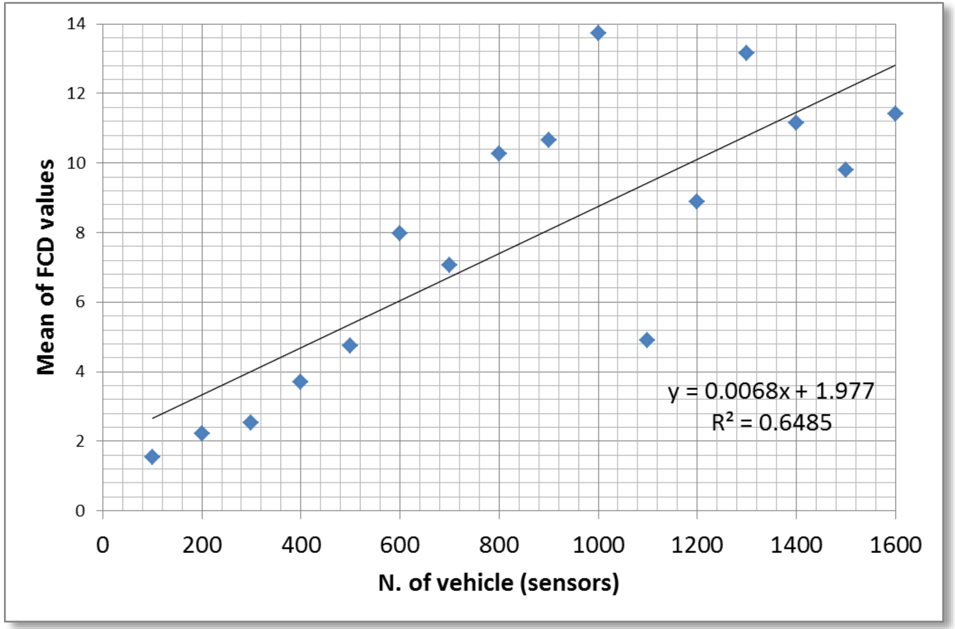
Freely available data from 124 sensors (out of the 3,400 installed in the municipality)



Comparison with traffic sensors

Freely available data from 124 sensors (out of the 3,400 installed in the municipality)

Representativeness of the FCD sample



Data processing: road functional classes

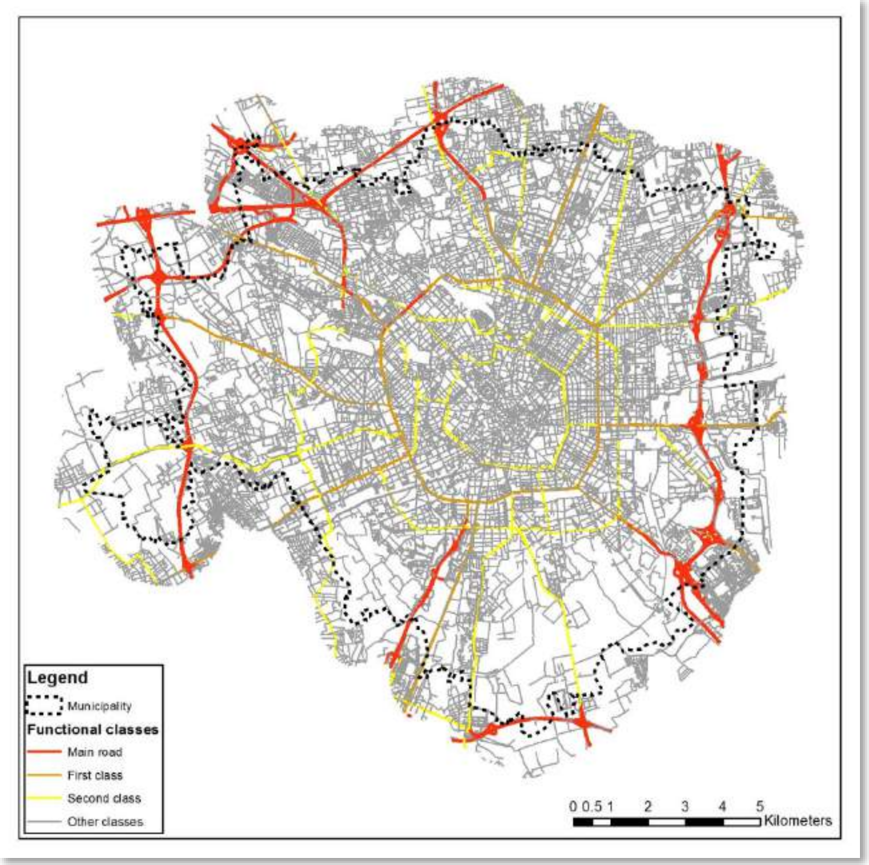
Road network elements classified on the base of the different functional classes

Data processing: road functional classes

URBAN-GEO BIG DATA

MILANO

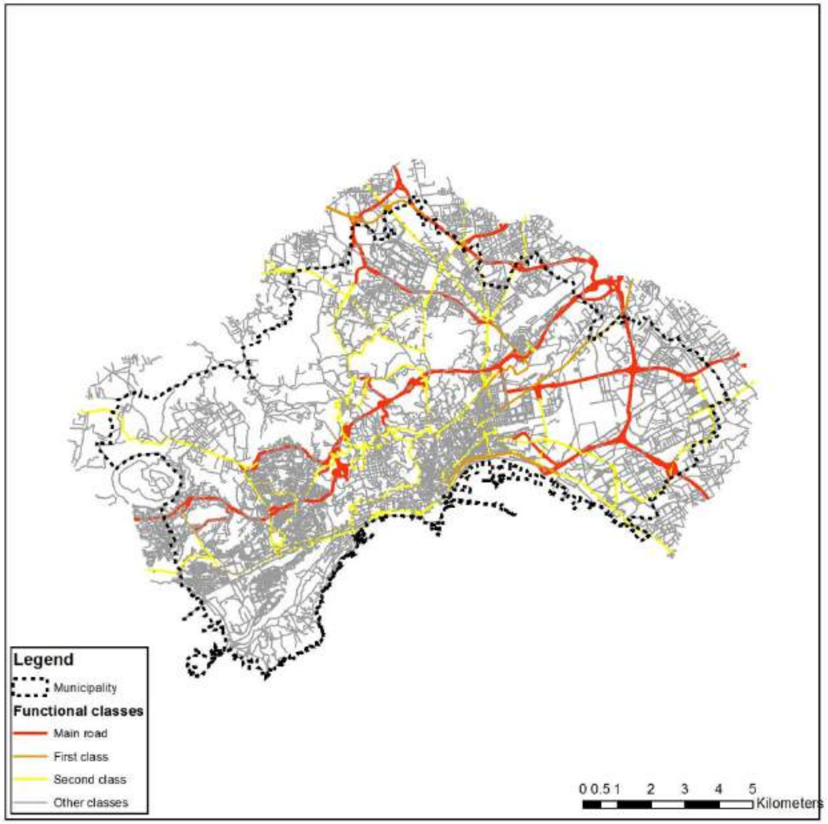
Road network elements classified on the base of the different functional classes



Data processing: road functional classes

NAPOLI

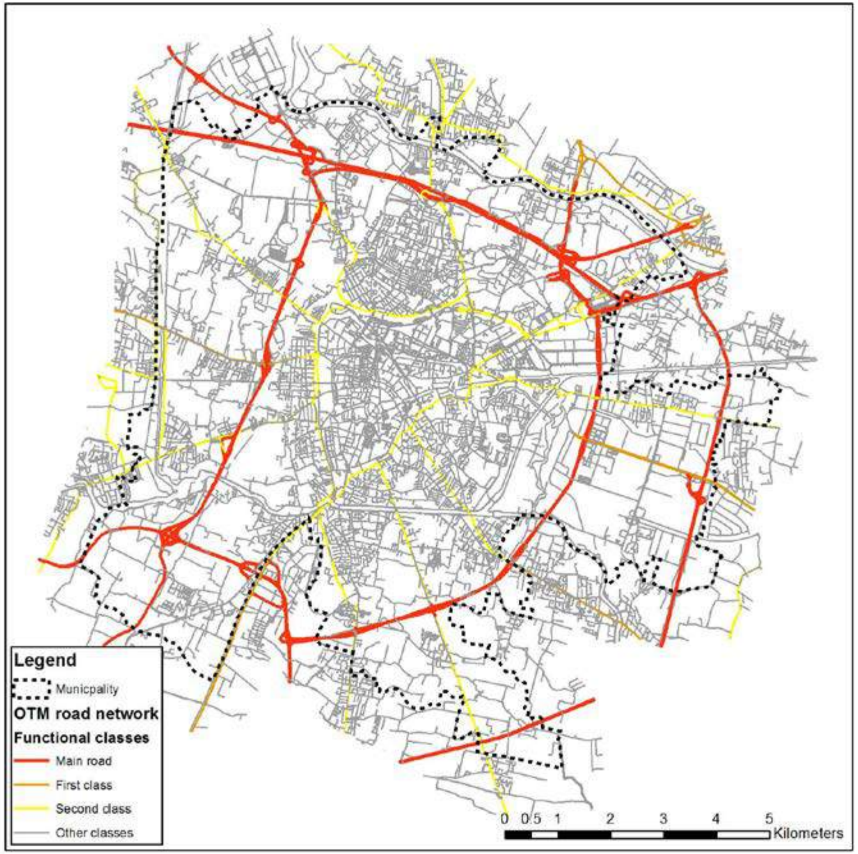
Road network elements classified on the base of the different functional classes



Data processing: road functional classes

PADOVA

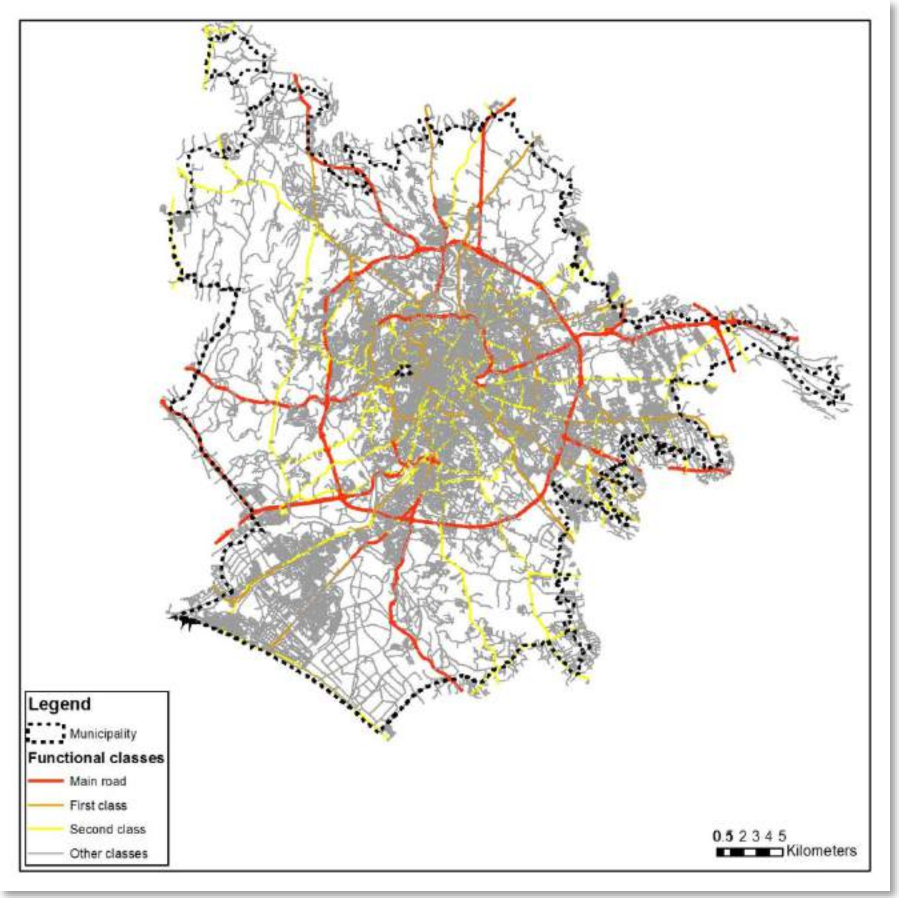
Road network elements classified on the base of the different functional classes



Data processing: road functional classes

ROMA

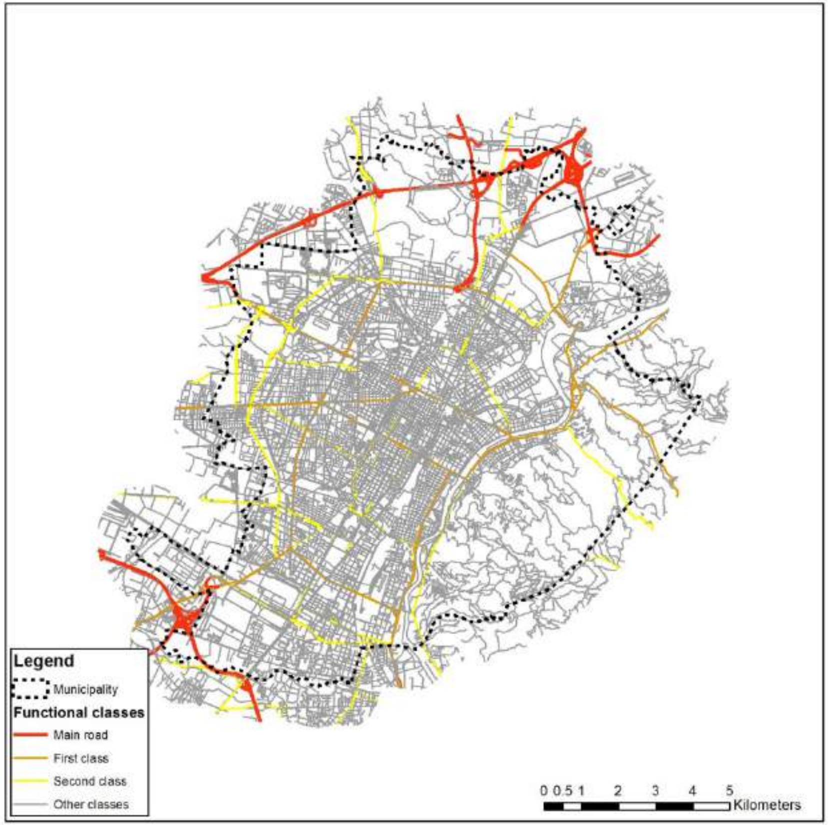
Road network elements classified on the base of the different functional classes



Data processing: road functional classes

TORINO

Road network elements classified on the base of the different functional classes



Data processing: FCD mean distance and HDOP

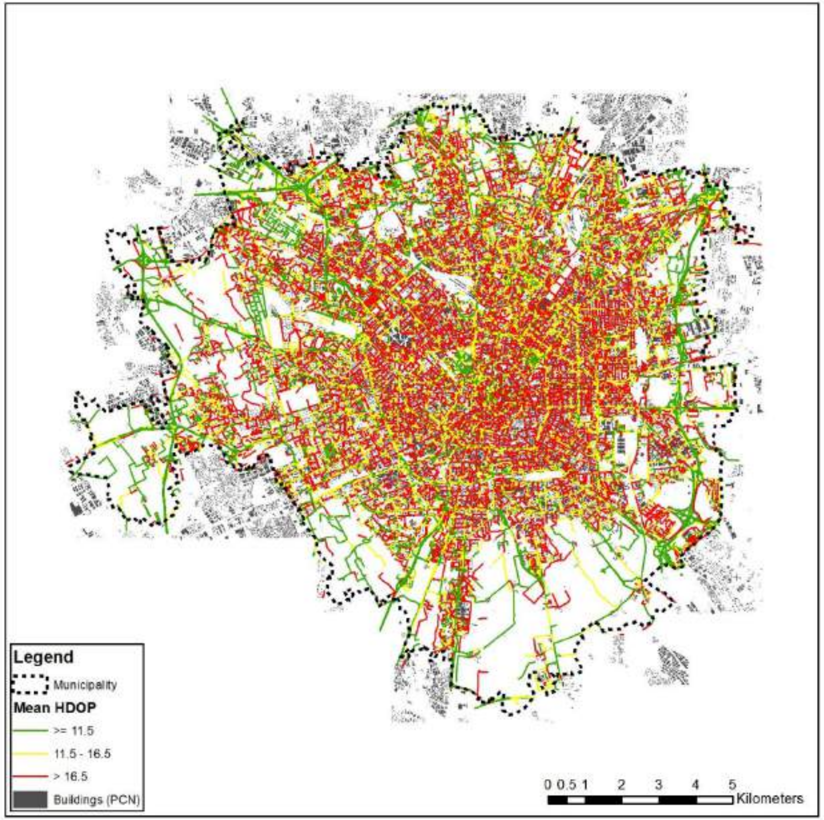
Mean distances and HDOPs calculated on the base of the spatial relationship between FCD location and every single arc/node of the road network

Data processing: FCD mean distance and HDOP

URBAN-GEO BIG DATA

MILANO

Mean distances and HDOPs calculated on the base of the spatial relationship between FCD location and every single arc/node of the road network

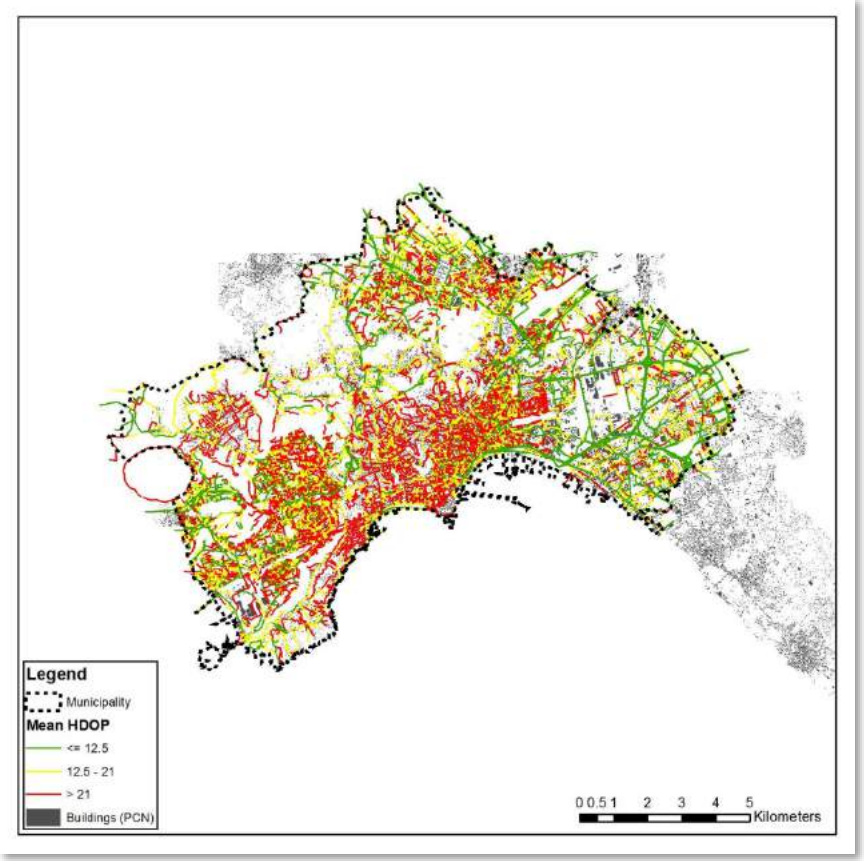


Data processing: FCD mean distance and HDOP

URBAN-GEO BIG DATA

NAPOLI

Mean distances and HDOPs calculated on the base of the spatial relationship between FCD location and every single arc/node of the road network

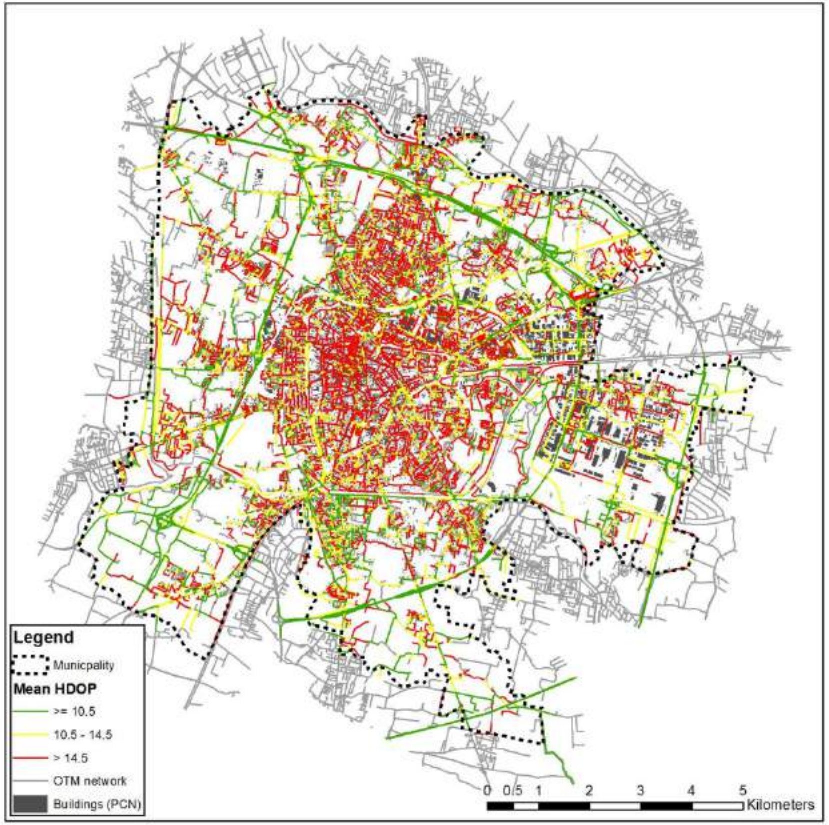


Data processing: FCD mean distance and HDOP

URBAN-GEO BIG DATA

PADOVA

Mean distances and HDOPs calculated on the base of the spatial relationship between FCD location and every single arc/node of the road network

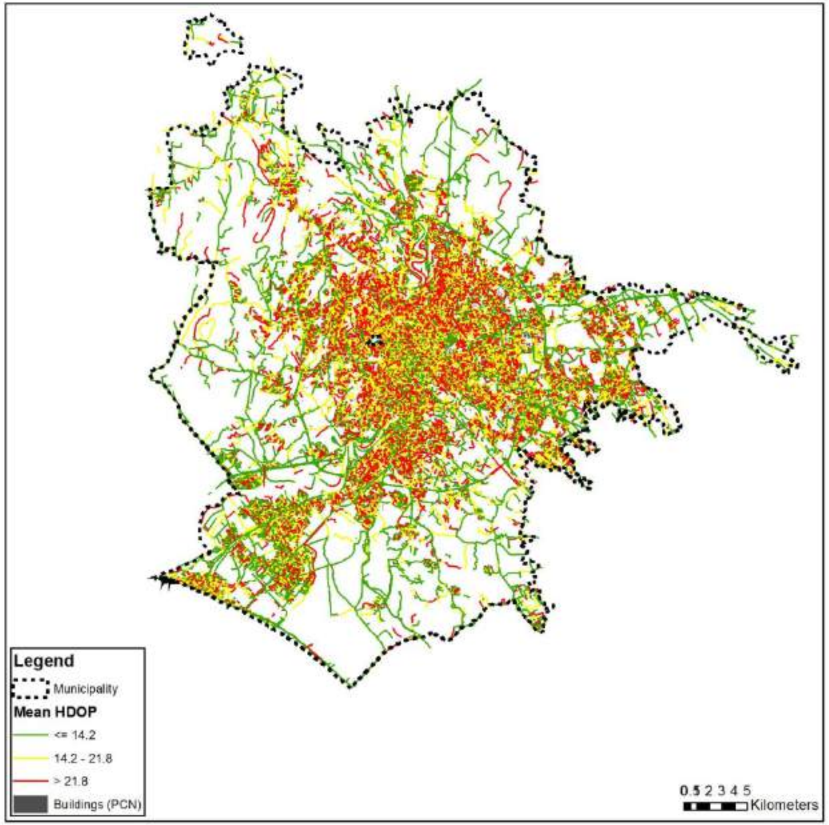


Data processing: FCD mean distance and HDOP

URBAN-GEO BIG DATA

ROMA

Mean distances and HDOPs calculated on the base of the spatial relationship between FCD location and every single arc/node of the road network



Data processing: FCD mean distance and HDOP

TORINO

Mean distances and HDOPs calculated on the base of the spatial relationship between FCD location and every single arc/node of the road network

Data processing: road network flows

Road network elements classified in function of the **number of** unique **vehicles** that travelled on each specific road element during the day (on hourly base)

Data processing: road network flows

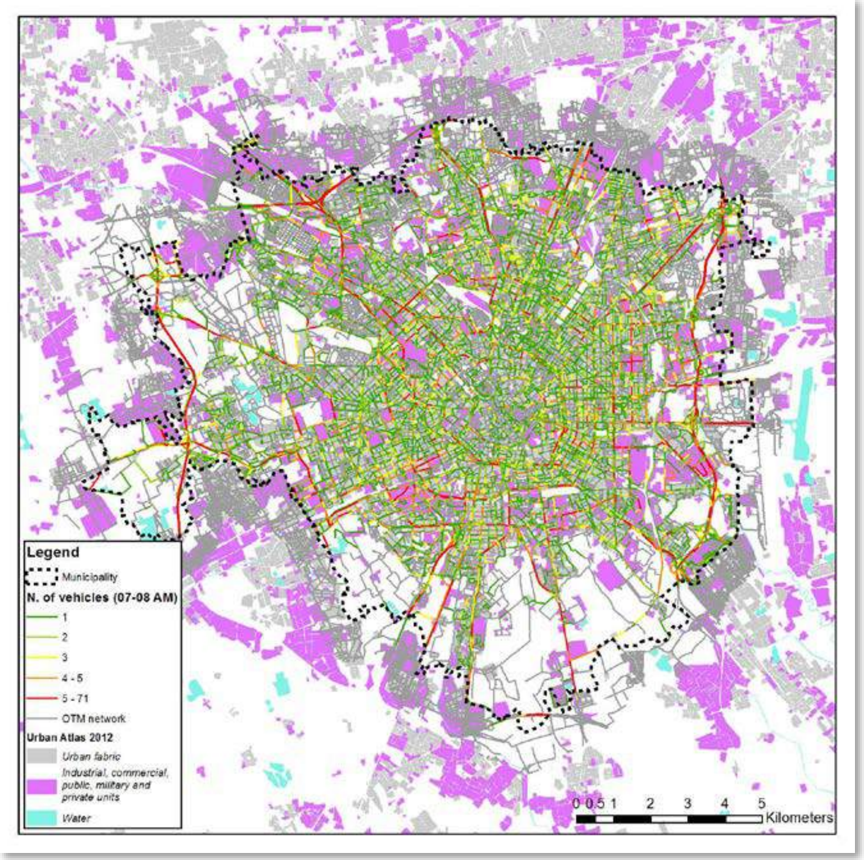
MILANO

Road network elements classified in function of the **number of** unique **vehicles** that travelled on each specific road element during the day (on hourly base)

Data processing: road network flows

MILANO

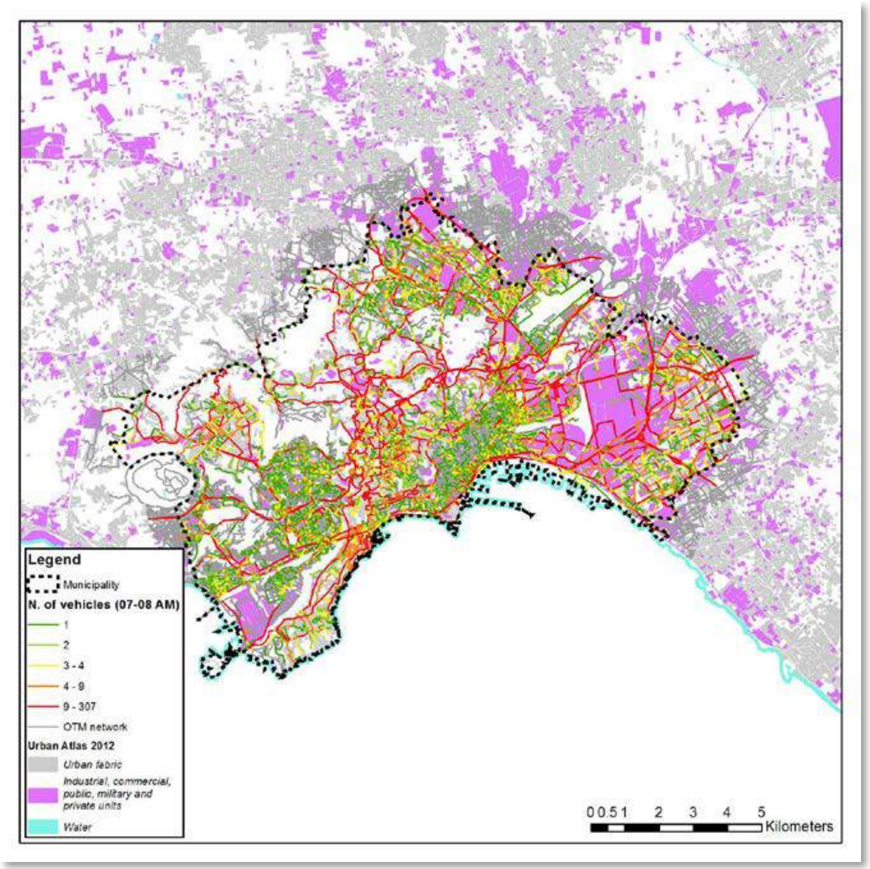
Road network elements classified in function of the **number of** unique **vehicles** that travelled on each specific road element during the day (on hourly base)



Data processing: road network flows

NAPOLI

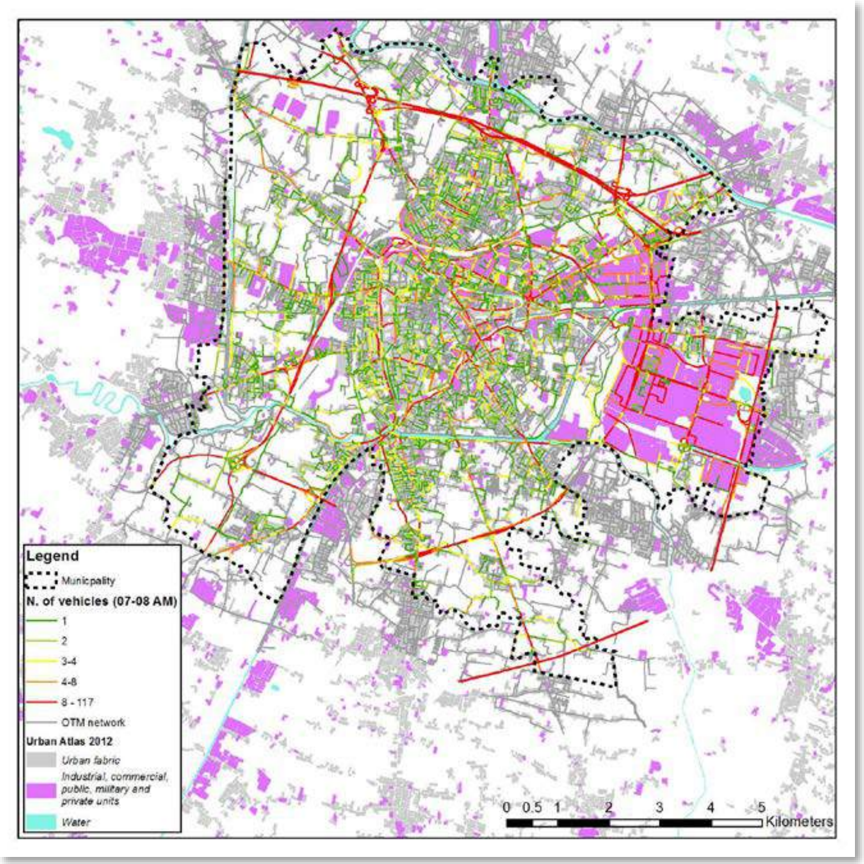
Road network elements classified in function of the **number of unique vehicles** that travelled on each specific road element during the day (on hourly base)



Data processing: road network flows

PADOVA

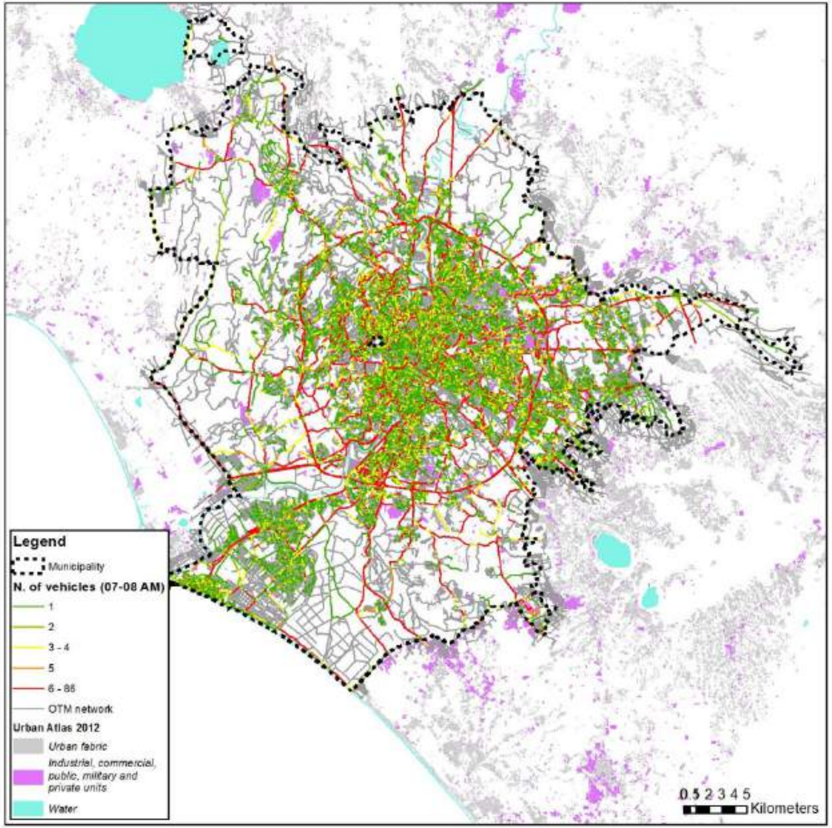
Road network elements classified in function of the **number of** unique **vehicles** that travelled on each specific road element during the day (on hourly base)



Data processing: road network flows

ROMA

Road network elements classified in function of the **number of** unique **vehicles** that travelled on each specific road element during the day (on hourly base)

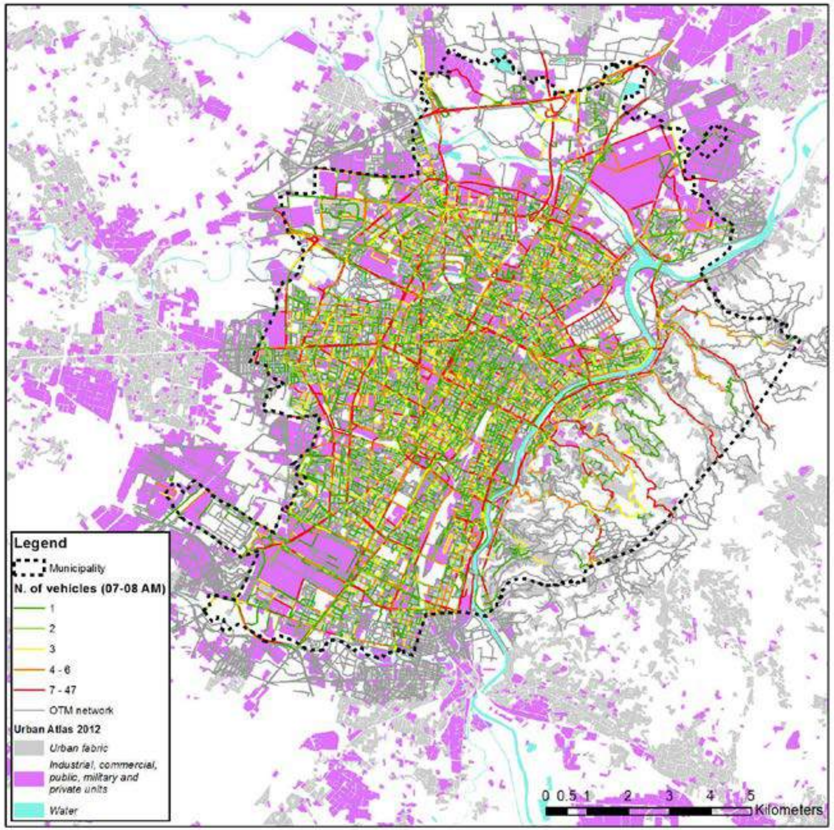


Data processing: road network flows

URBAN-GEO BIG DATA

TORINO

Road network elements classified in function of the **number of** unique **vehicles** that travelled on each specific road element during the day (on hourly base)



Data processing: road network speed

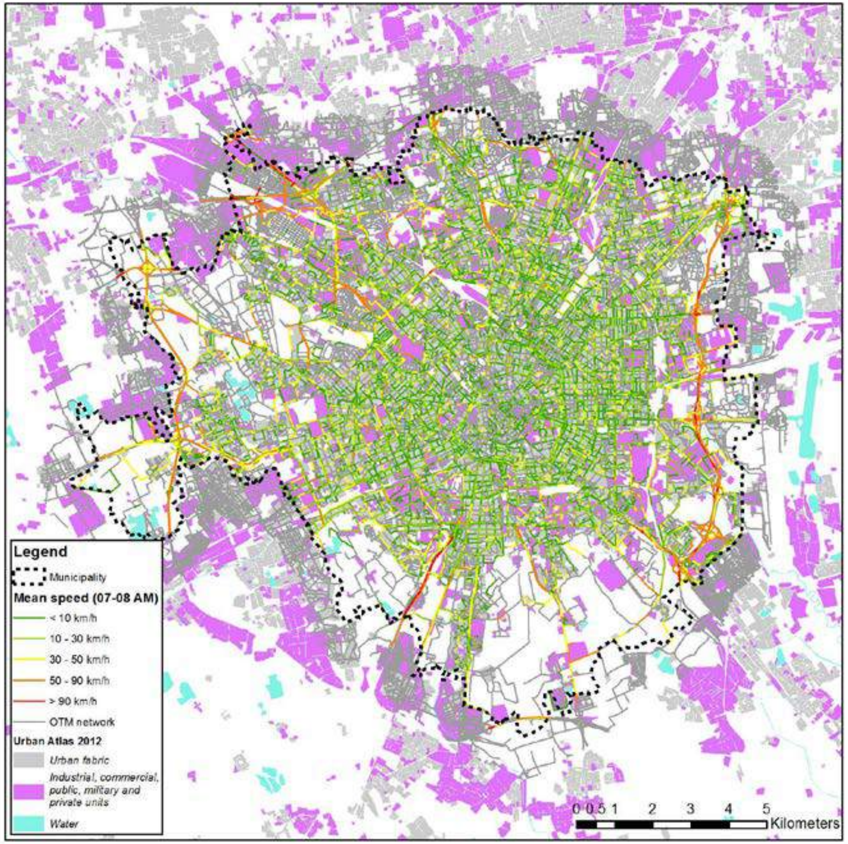
Road network elements classified in function of the **mean speed** on each specific road element during the day (on hourly base)

Data processing: road network speed

URBAN-GEO BIG DATA

MILANO

Road network elements classified in function of the **mean speed** on each specific road element during the day (on hourly base)

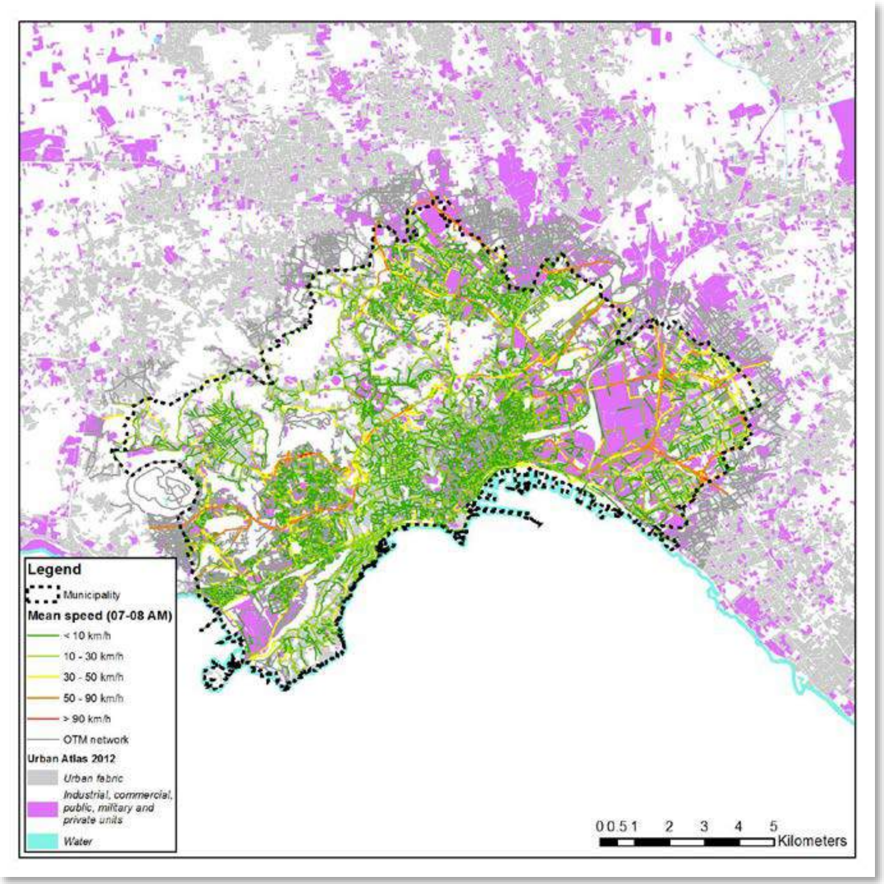


Data processing: road network speed

URBAN-GEO BIG DATA

NAPOLI

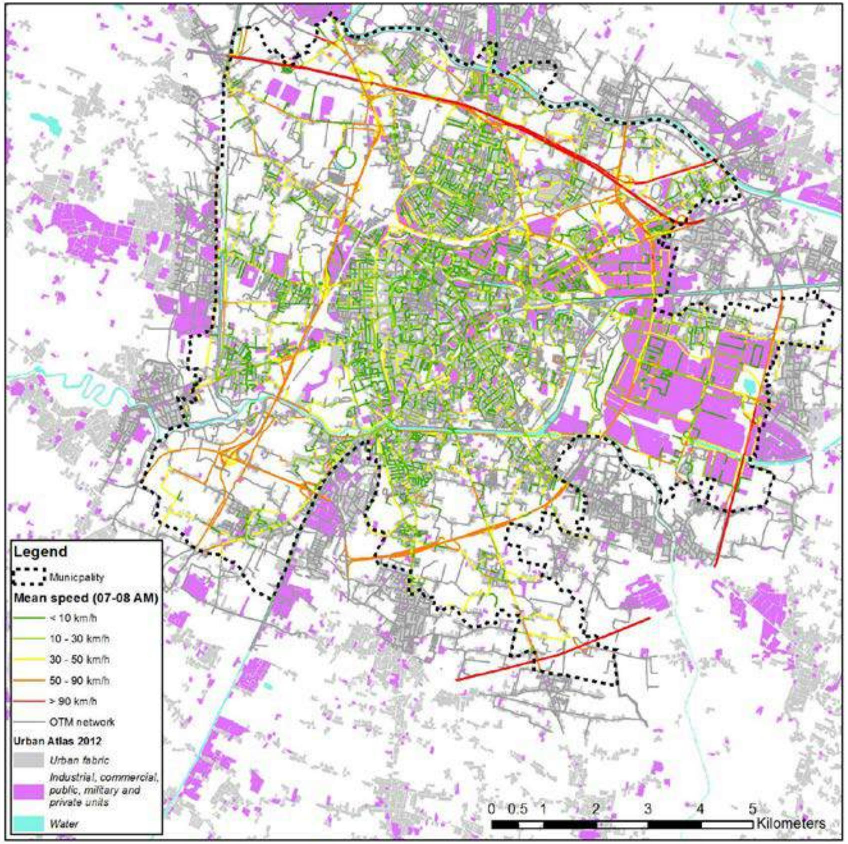
Road network elements classified in function of the **mean speed** on each specific road element during the day (on hourly base)



Data processing: road network speed

PADOVA

Road network elements classified in function of the **mean speed** on each specific road element during the day (on hourly base)

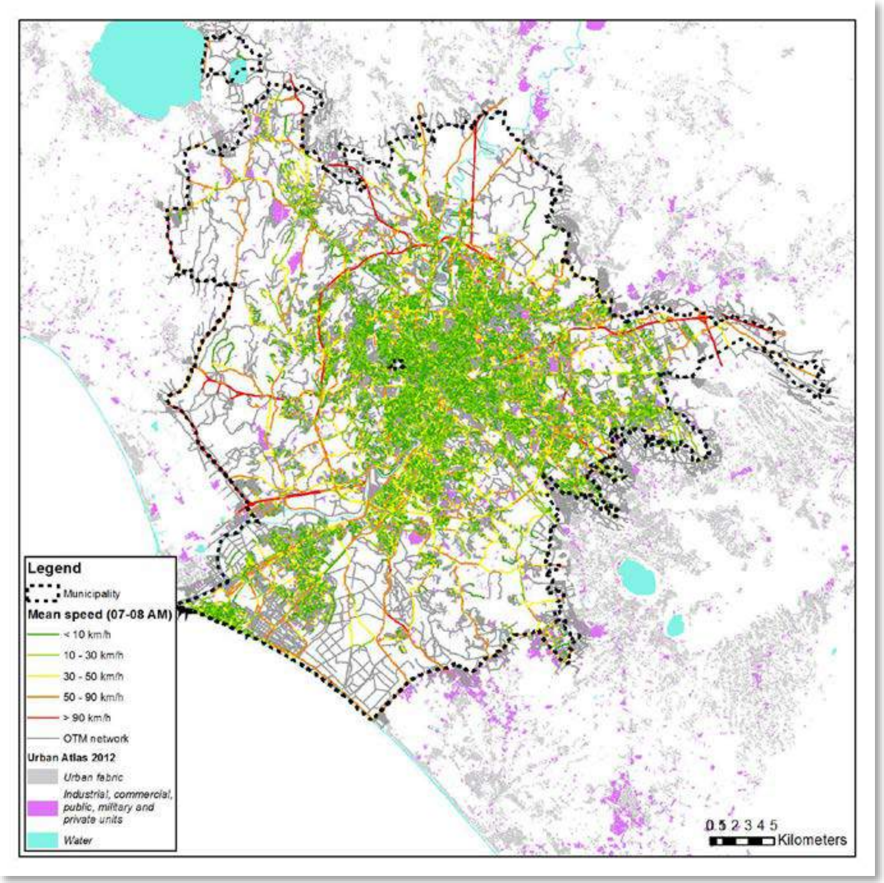


Data processing: road network speed

URBAN-GEO BIG DATA

ROMA

Road network elements classified in function of the **mean speed** on each specific road element during the day (on hourly base)

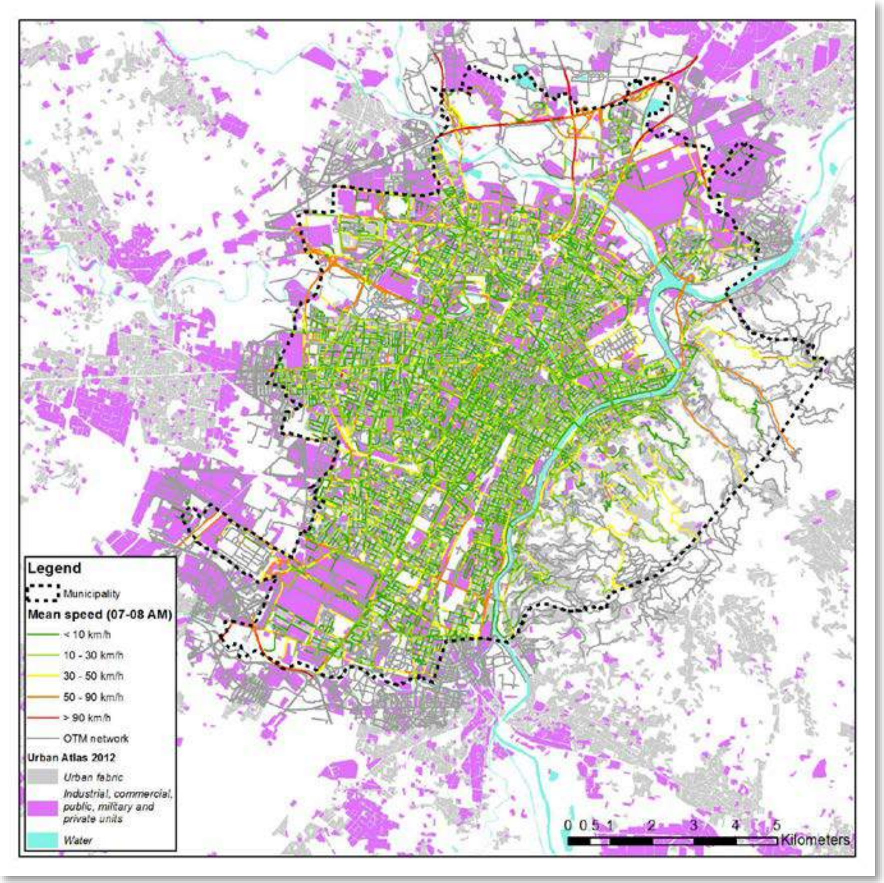


Data processing: road network speed

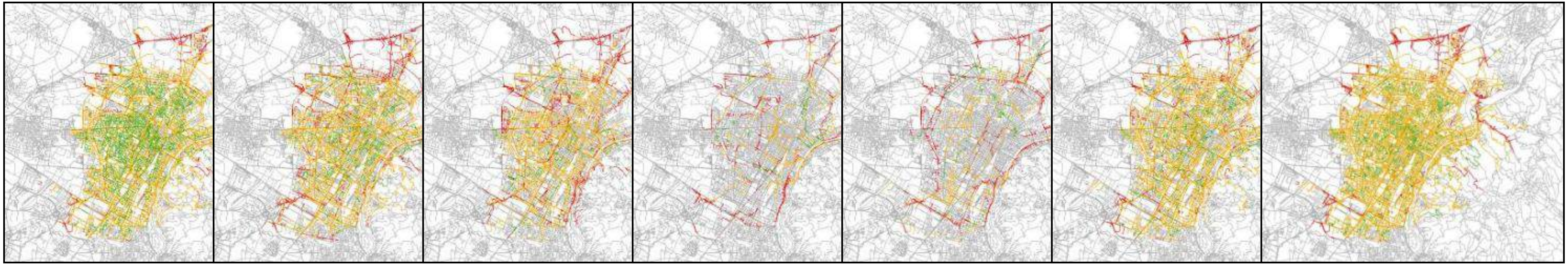
URBAN-GEO BIG DATA

TORINO

Road network elements classified in function of the **mean speed** on each specific road element during the day (on hourly base)



Mean speed/Travel times



The use of the timestamp associated to FCD data allows to calculate **mean speeds in the different moments of the day**, useful for estimating **dynamic travel times**

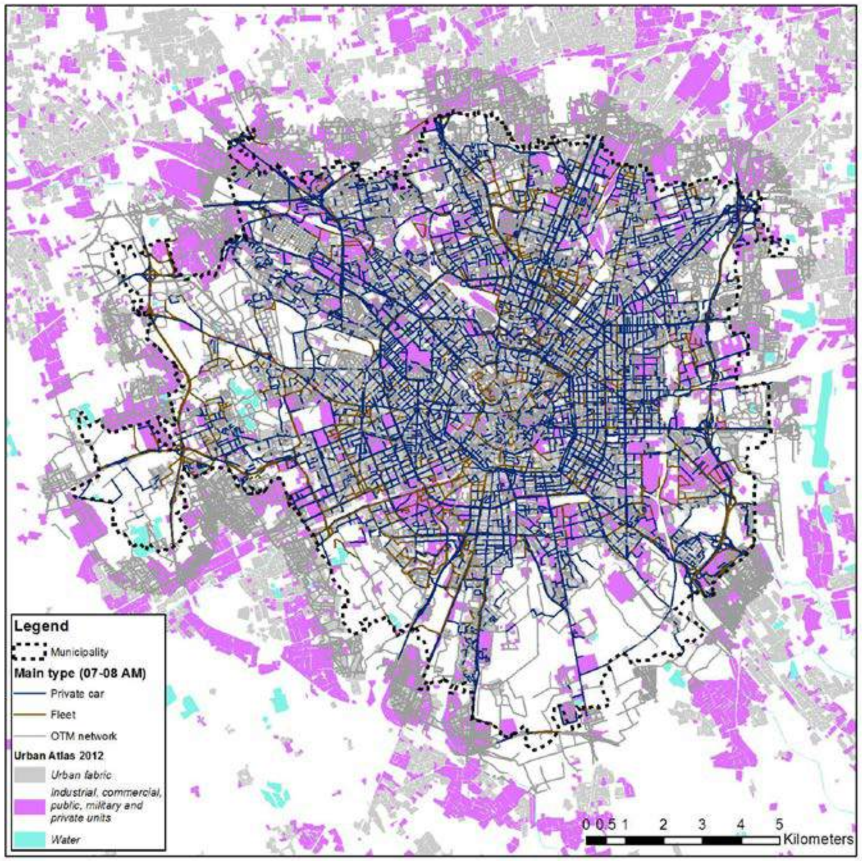
Data processing: predominant vehicle

Road network elements classified in function of predominant vehicle type (private or fleet) on each specific road element during the day (on hourly base)

Data processing: predominant vehicle

MILANO

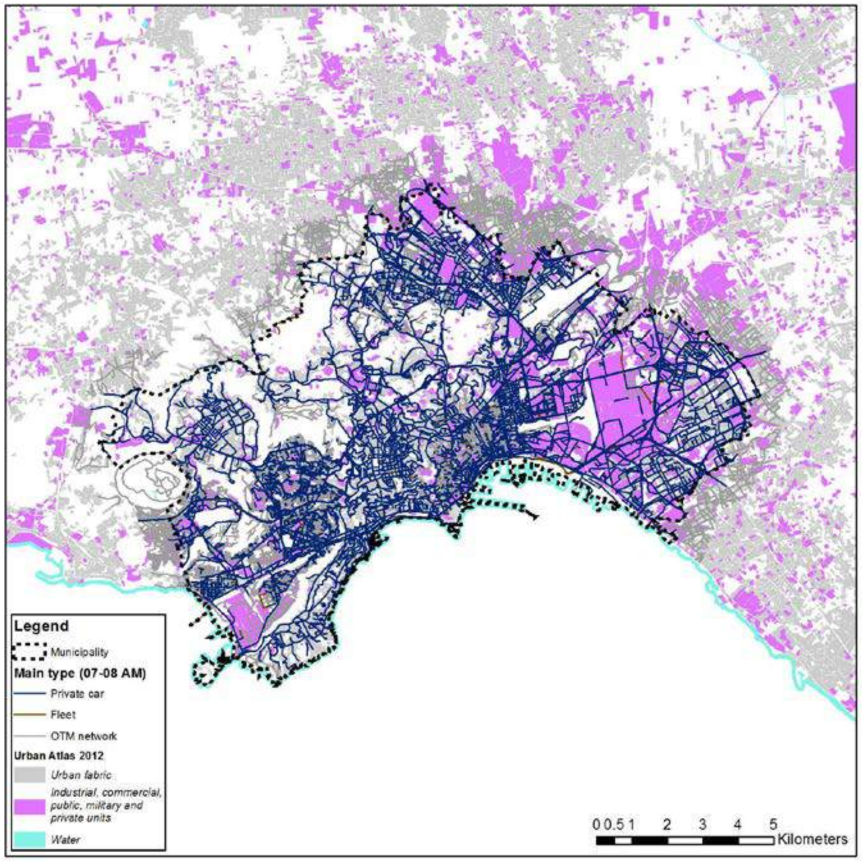
Road network elements classified in function of predominant vehicle type (private or fleet) on each specific road element during the day (on hourly base)



Data processing: predominant vehicle

NAPOLI

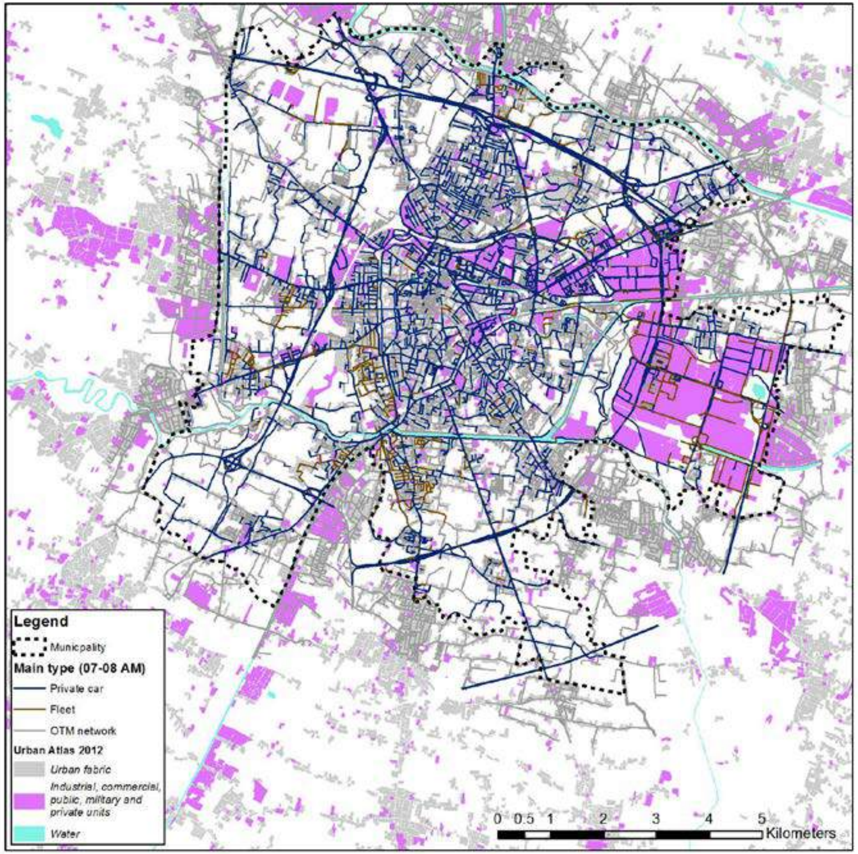
Road network elements classified in function of predominant vehicle type (private or fleet) on each specific road element during the day (on hourly base)



Data processing: predominant vehicle

PADOVA

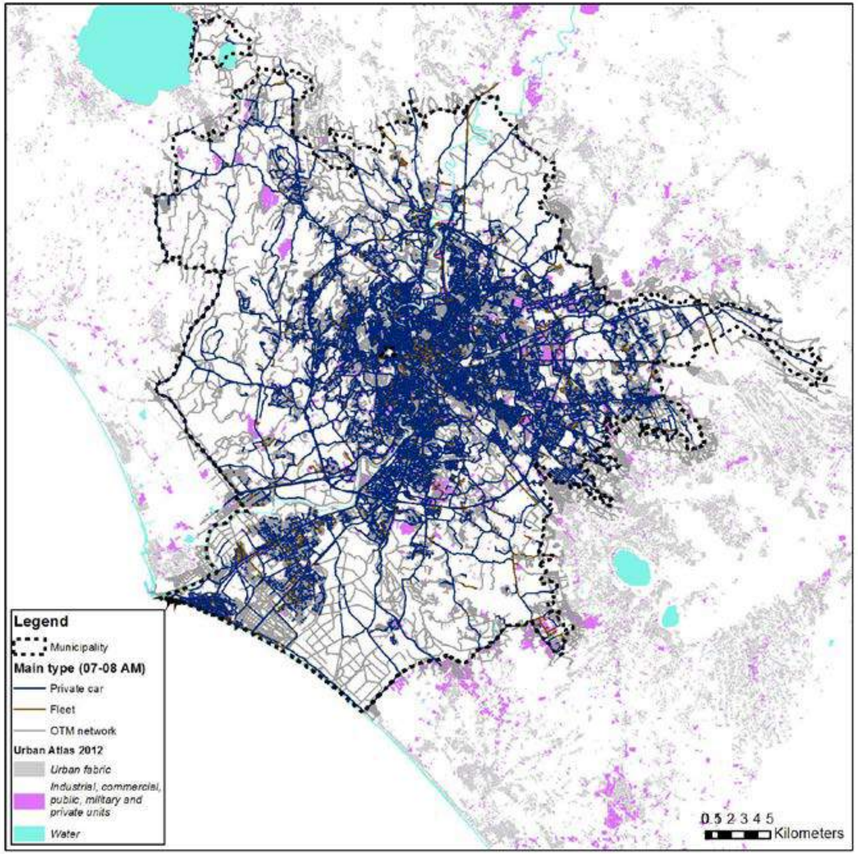
Road network elements classified in function of predominant vehicle type (private or fleet) on each specific road element during the day (on hourly base)



Data processing: predominant vehicle

ROMA

Road network elements classified in function of predominant vehicle type (private or fleet) on each specific road element during the day (on hourly base)

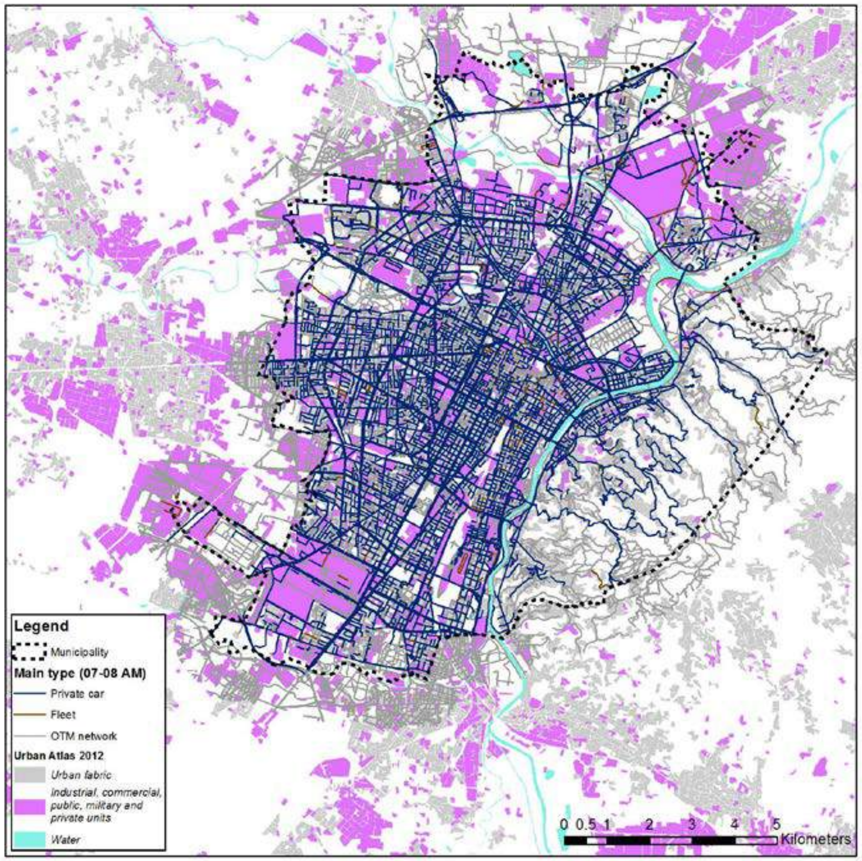


Data processing: predominant vehicle

URBAN-GEO BIG DATA

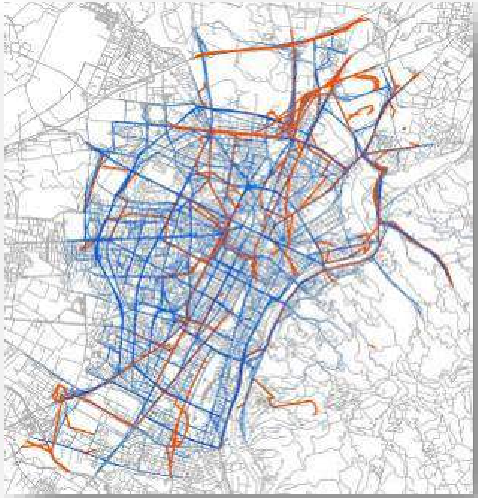
TORINO

Road network elements classified in function of predominant vehicle type (private or fleet) on each specific road element during the day (on hourly base)

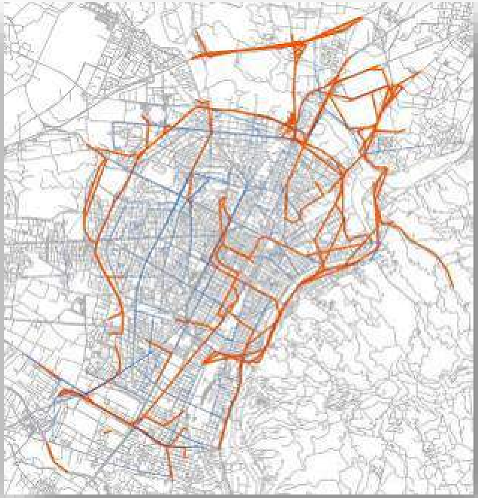


Data processing: predominant vehicle

Private cars vs. fleets



19:30 to 00:30



00:30 to 05:30

- Based on **attribute identifying the type of vehicle** (private car or fleet)
- dominance of **private cars** in **evening hours**
- dominance of **fleets** during the **night and early morning hours**

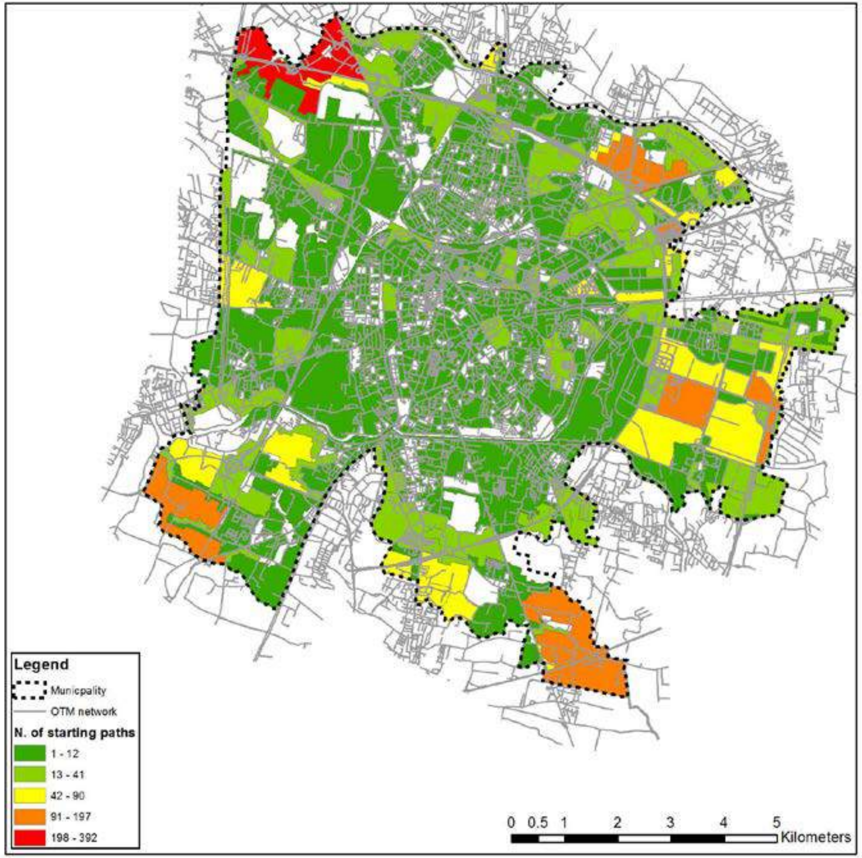
Data processing: O/D spatial matrix

Innovative Origin/Destination (O/D) spatial matrix based on path calculation (start, end and paths length)

Data processing: O/D spatial matrix

PADOVA

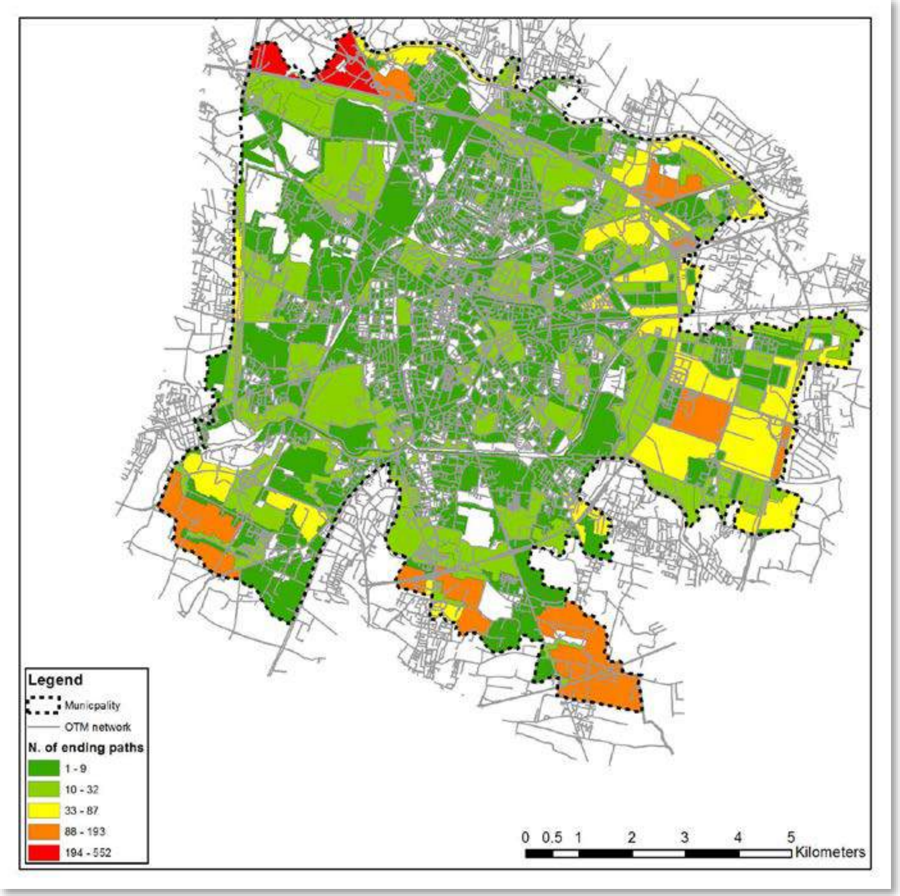
Innovative Origin/Destination (O/D) spatial matrix based on path calculation (**start**, end and paths length)



Data processing: O/D spatial matrix

PADOVA

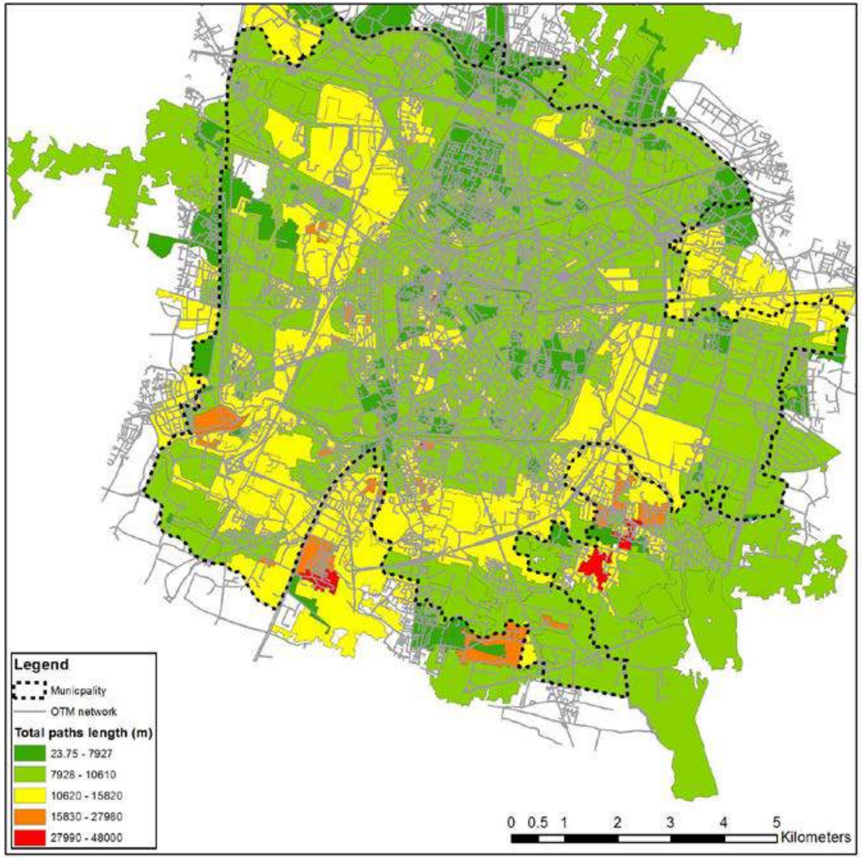
Innovative Origin/Destination (O/D) spatial matrix based on path calculation (start, **end** and paths length)



Data processing: O/D spatial matrix

PADOVA

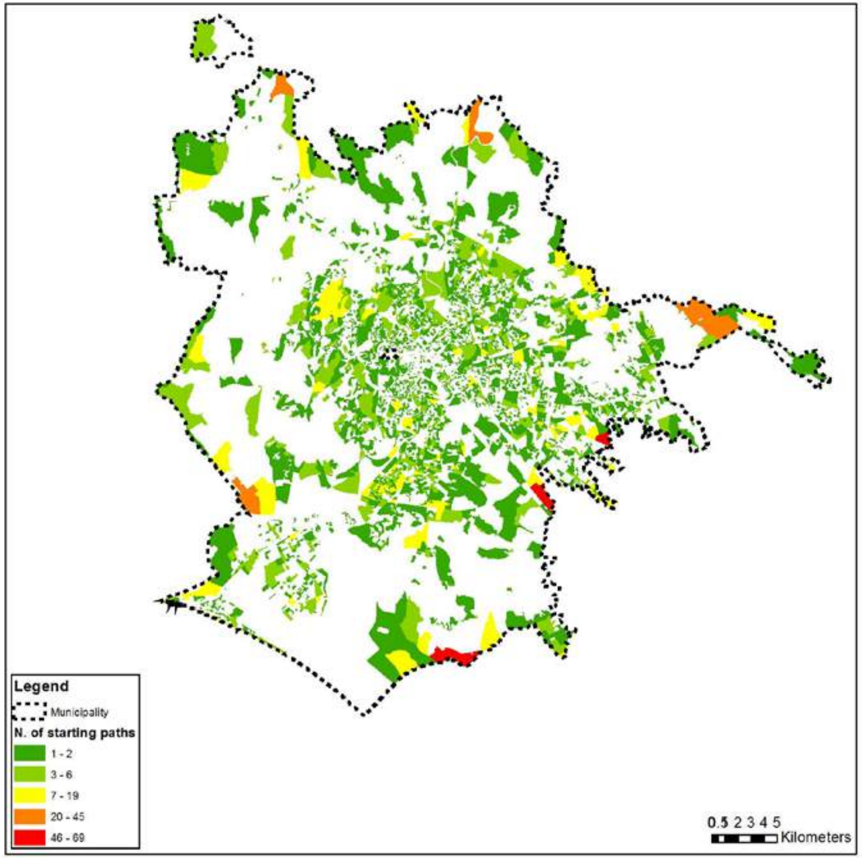
Innovative Origin/Destination (O/D) spatial matrix based on path calculation (start, **end** and paths length)



Data processing: O/D spatial matrix

ROMA

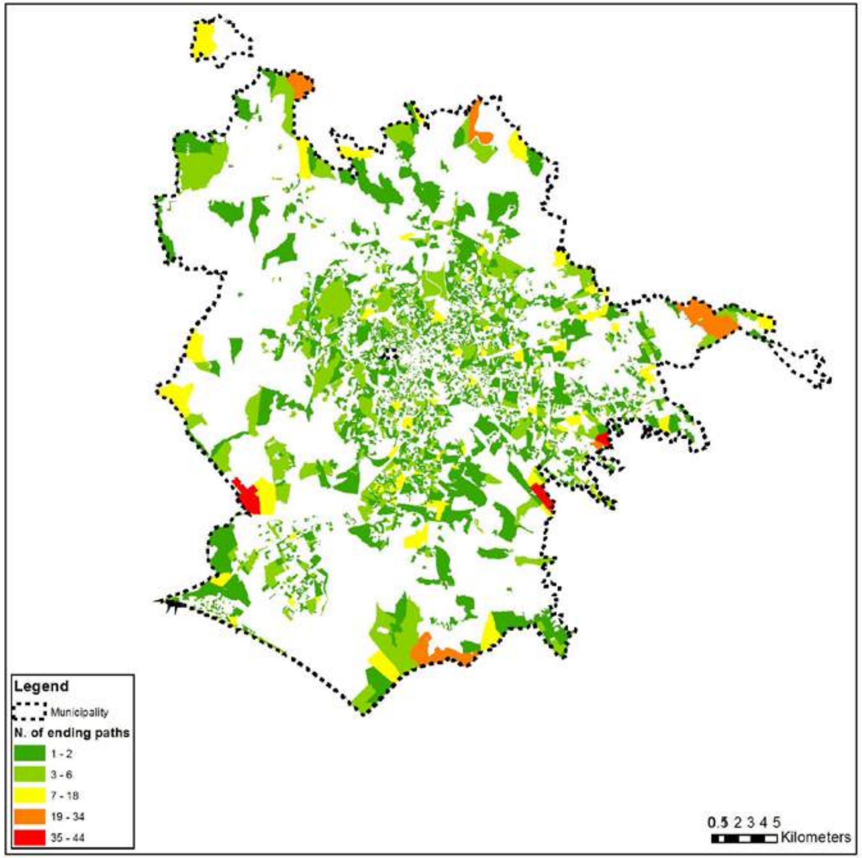
Innovative Origin/Destination (O/D) spatial matrix based on path calculation (**start**, end and paths length)



Data processing: O/D spatial matrix

ROMA

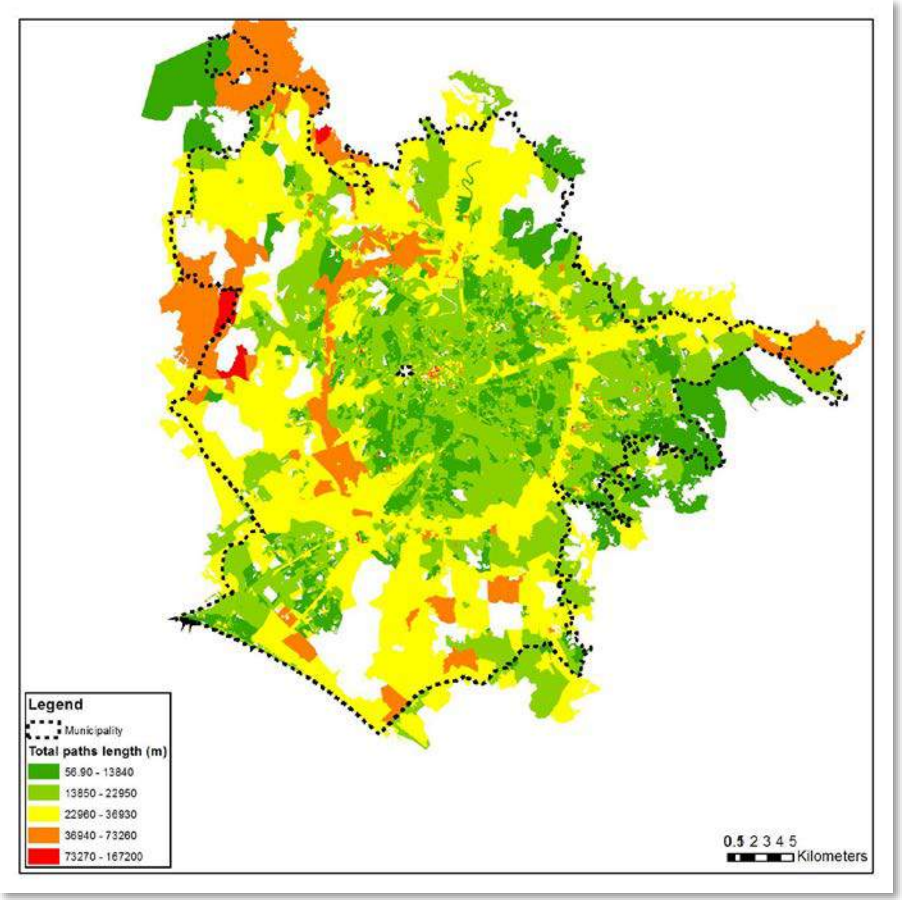
Innovative Origin/Destination (O/D) spatial matrix based on path calculation (start, **end** and paths length)



Data processing: O/D spatial matrix

ROMA

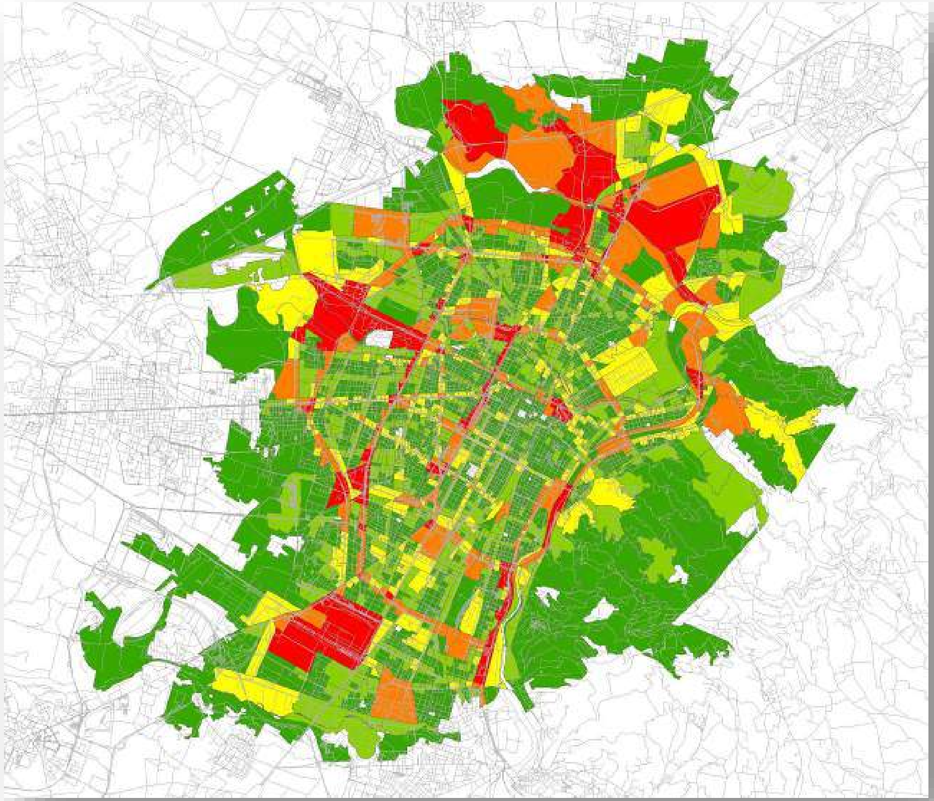
Innovative Origin/Destination (O/D) spatial matrix based on path calculation (start, end and **paths length**)



Data processing: O/D spatial matrix

TORINO

Innovative Origin/Destination (O/D) spatial matrix based on path calculation (start, end and **paths length**)



Data and services deployment

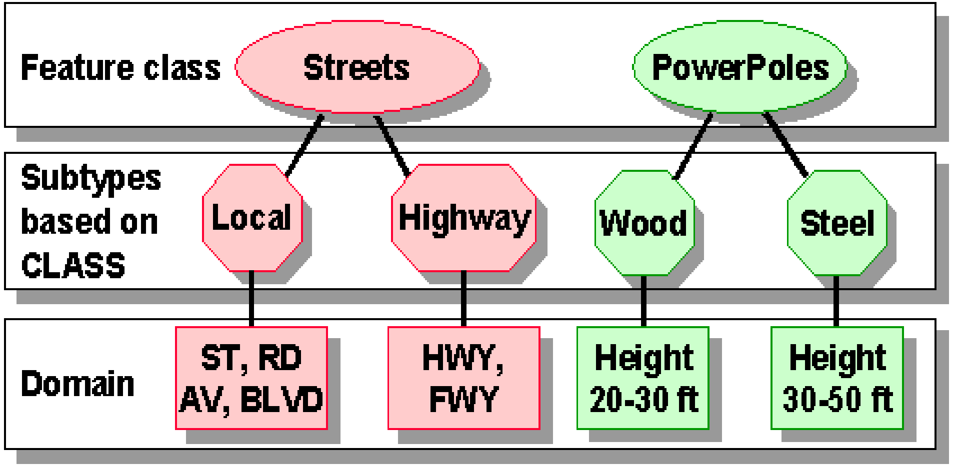
Public and private (OS Road network and associated FCD) data have been delivered via **WMS connection** having a thematization over the latter in terms of number of FCD point at **hourly interval**. All the other thematic layers (mean distances and HDOPs values, mean speed, predominant type) can be delivered in the same way building a multitemporal stack (24 intervals for 7 days) for every single data type.

As far as **O/D matrixes** are concerned, a completely new aggregation is created, thus, once finished the processing the simpliest way to deliver this data is to produce a thematic layer (multitemporal stack with 24 interval for 7 days) based on **census areas** (polygons), while travel paths lenght could be considered as a unique map updated weekly and/or montly/yearly.

Data and services deployment

Anyway, the first best is definitely a OS GeoDB format embedding:

- **Road network feature class** using **subtypes** (with defined **domain**) to categorize data;
- Create **connectivity rules** between other subtypes and feature classes to maintain the integrity of a network;
- Create **topology rules** between other subtypes and feature classes residing in a topology;
- Develop **relationship rules** between other subtypes (FCD time stamps and type), tables (FCD data), and feature classes (census area)



Further development

- Complete O/D processing using a HPC node;
- Insert all metadata in the system;
- Test data and service delivering in the two modalities (WMS vs GeoDB);
- Analyze FCD positioning accuracy for every city using ancillary data (e.g. 3D city models) to determine data positional accuracy and thus usability;
- Collaborate with the other Local Units in order to have direct interpretation of the data (especially travel paths representation);
- Possibly organize a side event (preferably in Roma) where presenting the project outcomes

URBAN GEO BIG DATA FOR MOBILITY APPLICATIONS

POLITO LOCAL UNIT