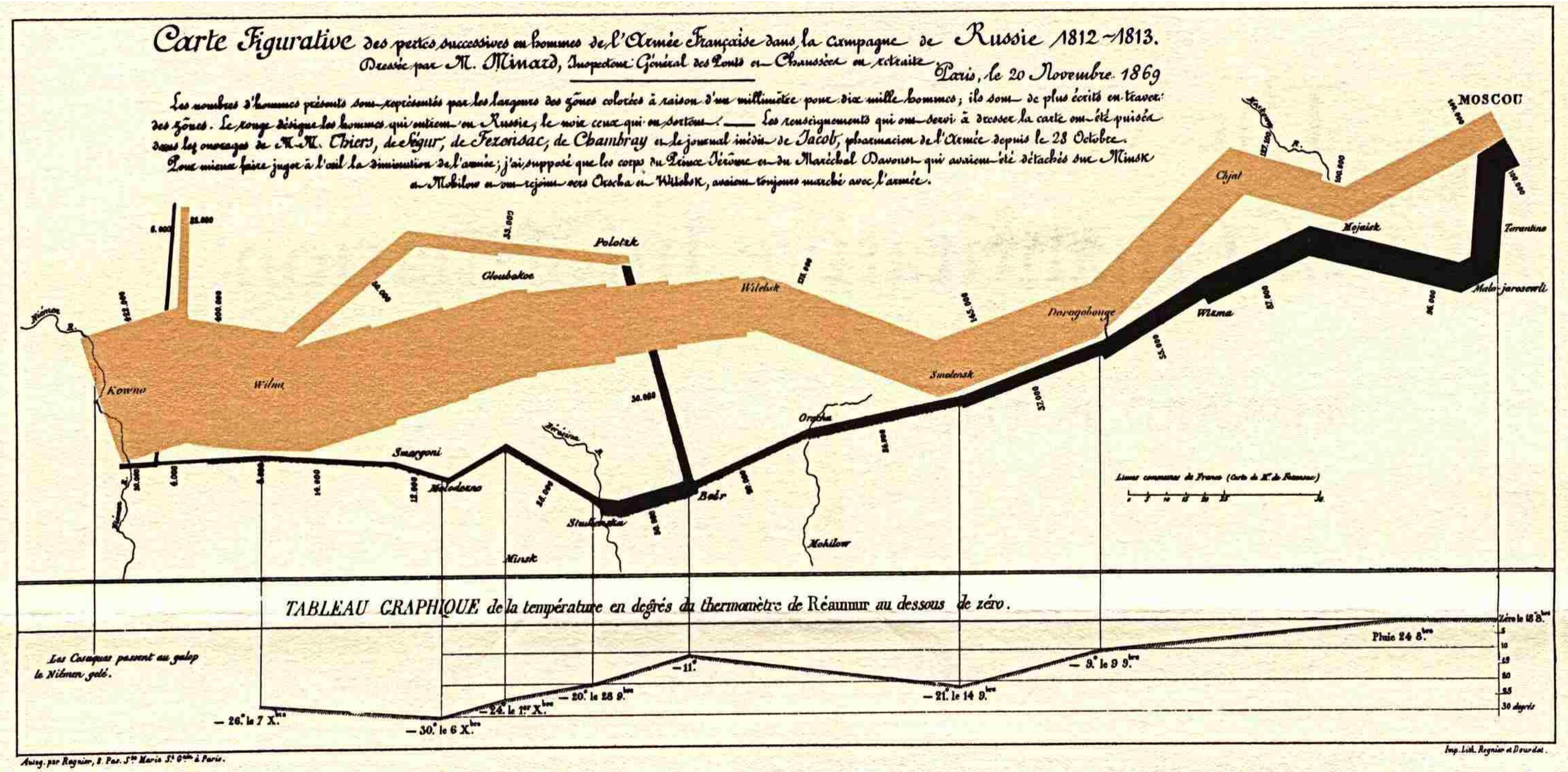
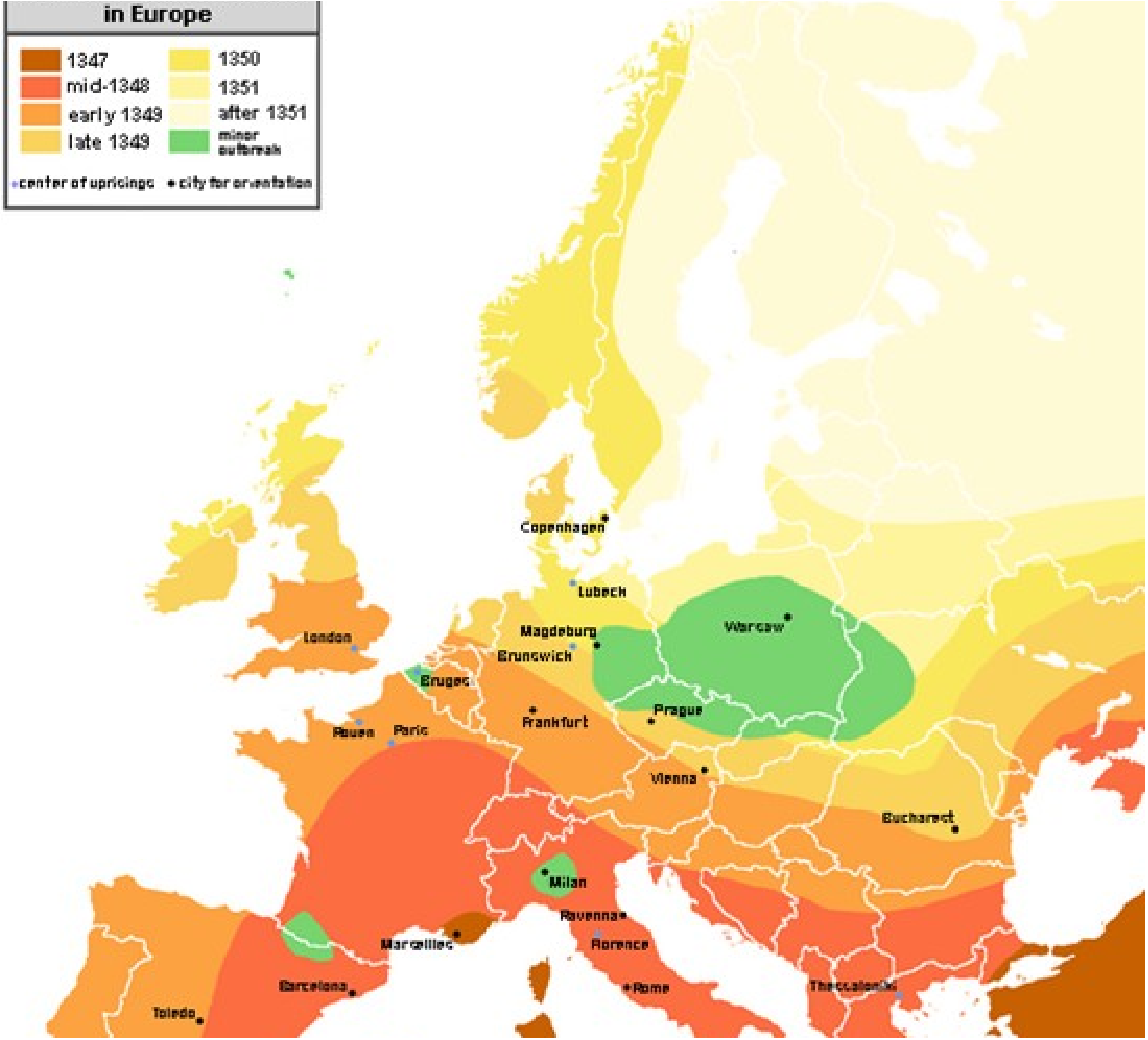
Mobility Data Mining

Mobility data Analysis Foundations

Understanding Human Mobility: a long path



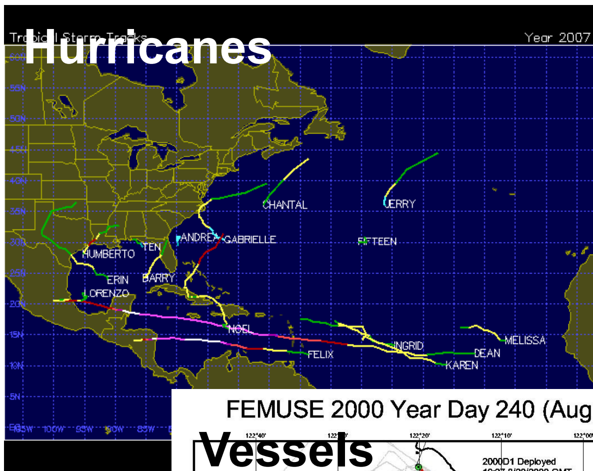
Charles Minard. “Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813”, 1869.

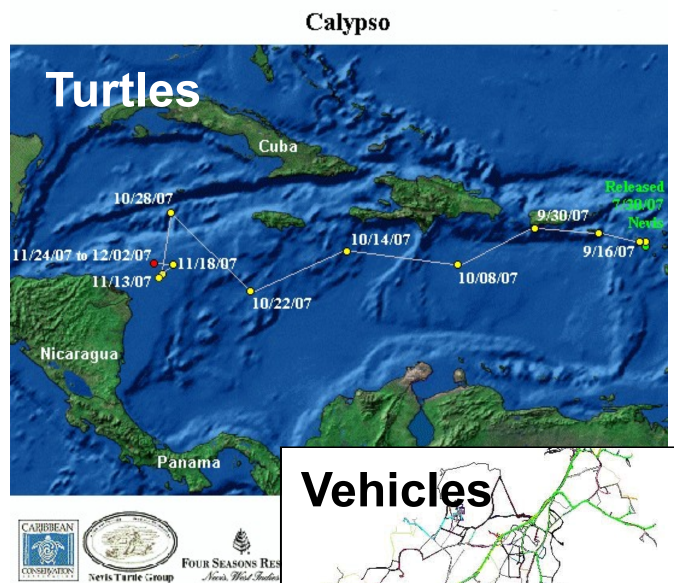


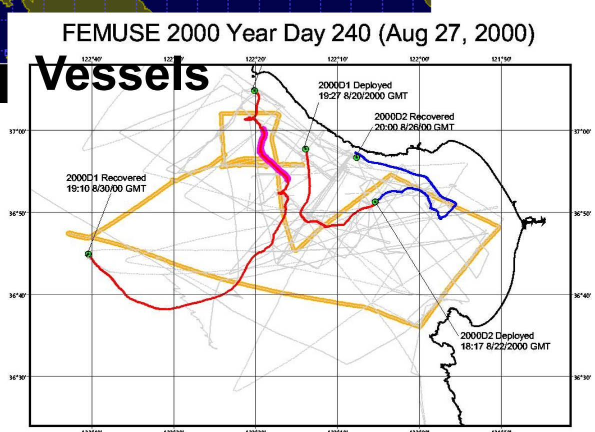
Spread of the Black Death in Europe (1346–53)

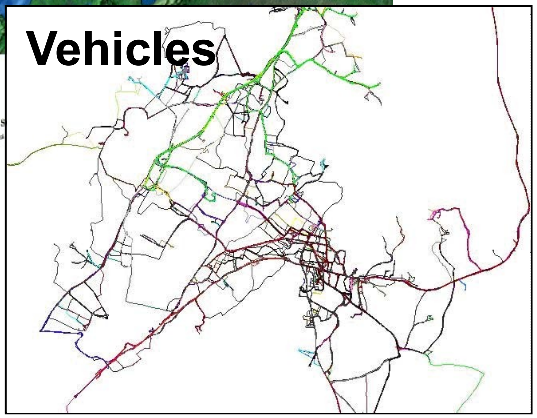
Moving Object Data

- Several domains:





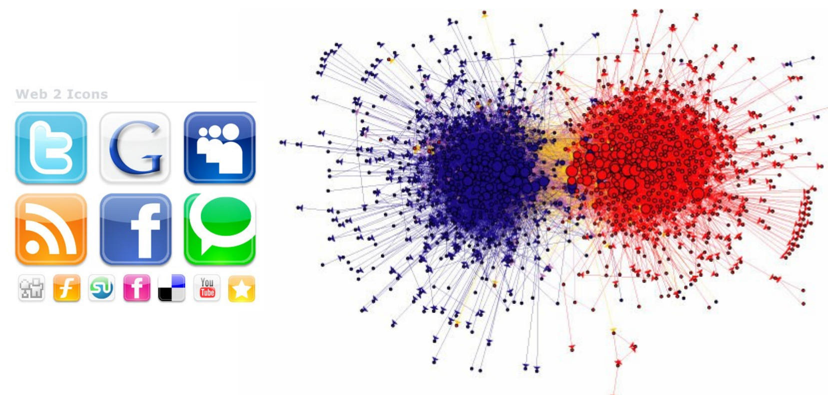




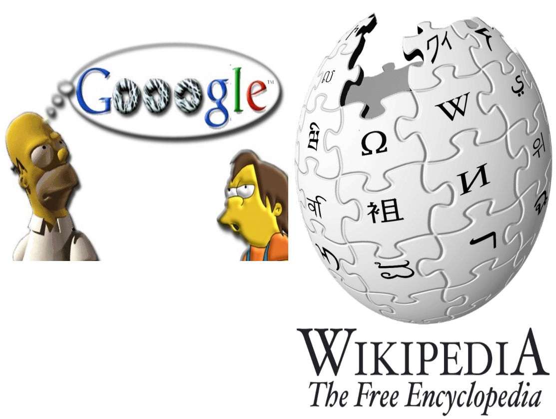
The novelty : BIG DATA

What we buy

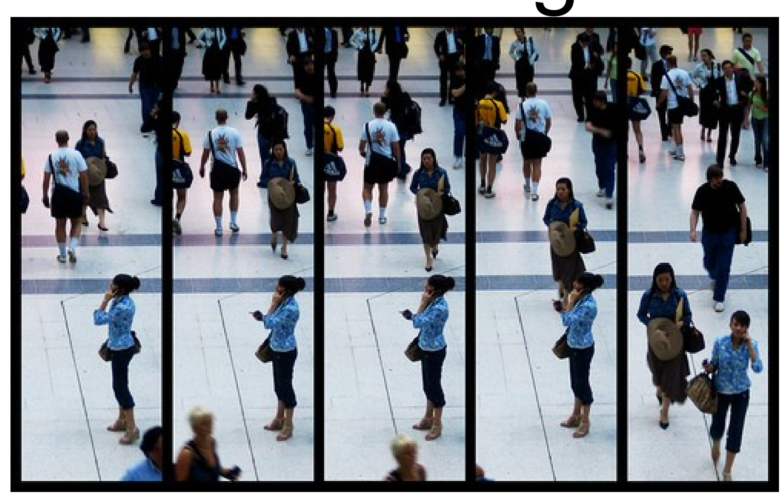
Whom we interact with



What we search for



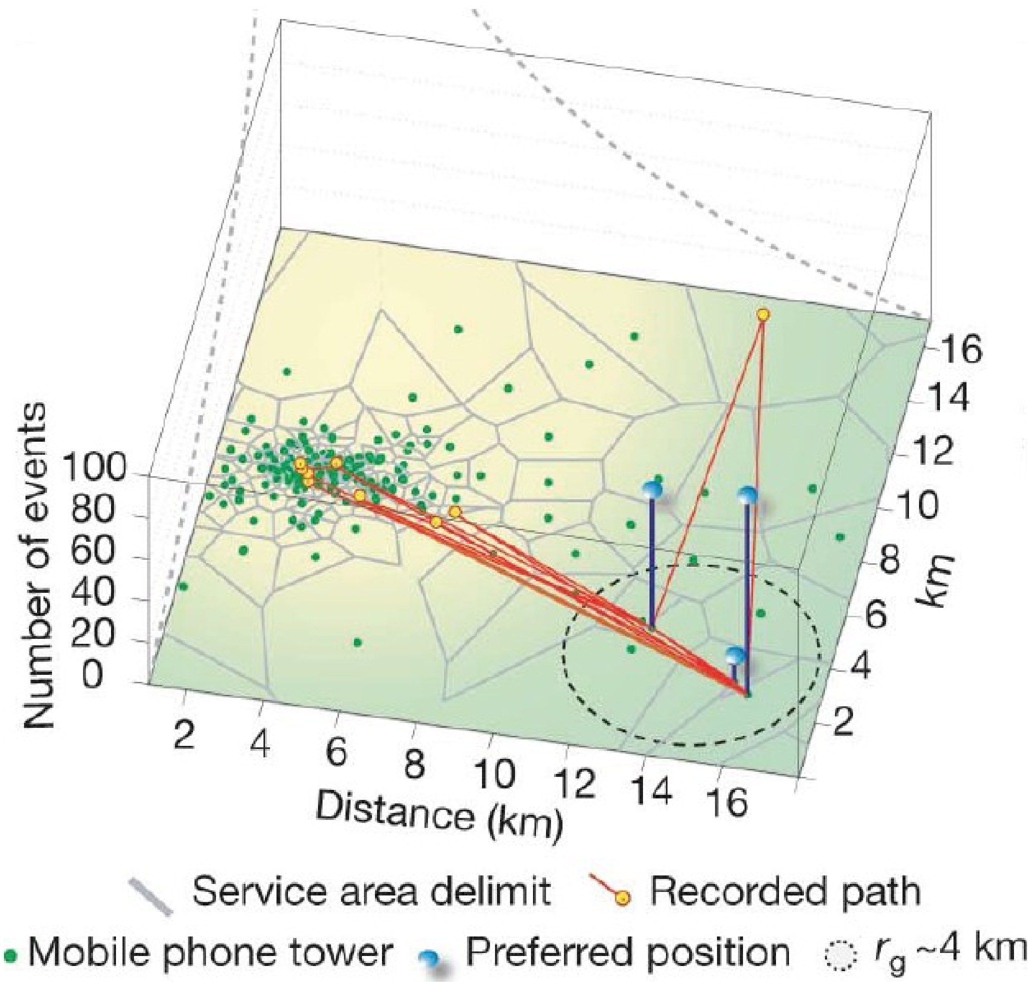
Where we go



Why Mining Moving Object Data?

- Large diffusion of mobile devices, mobile services and location-based services

Country-wide mobile phone data



when you call

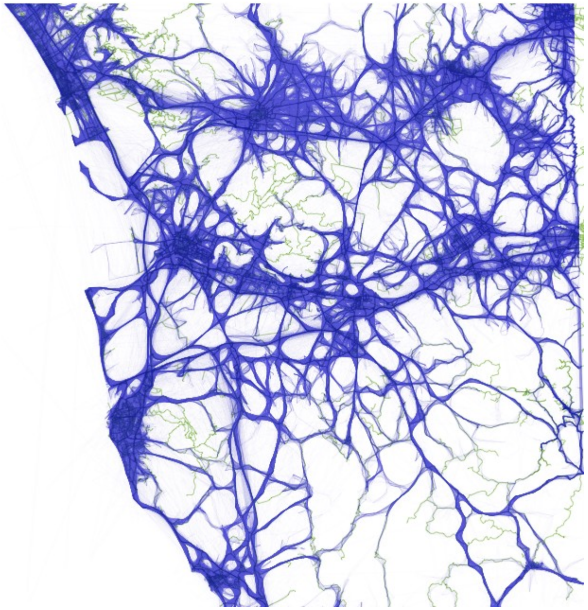
where you call

who you call

GPS tracks

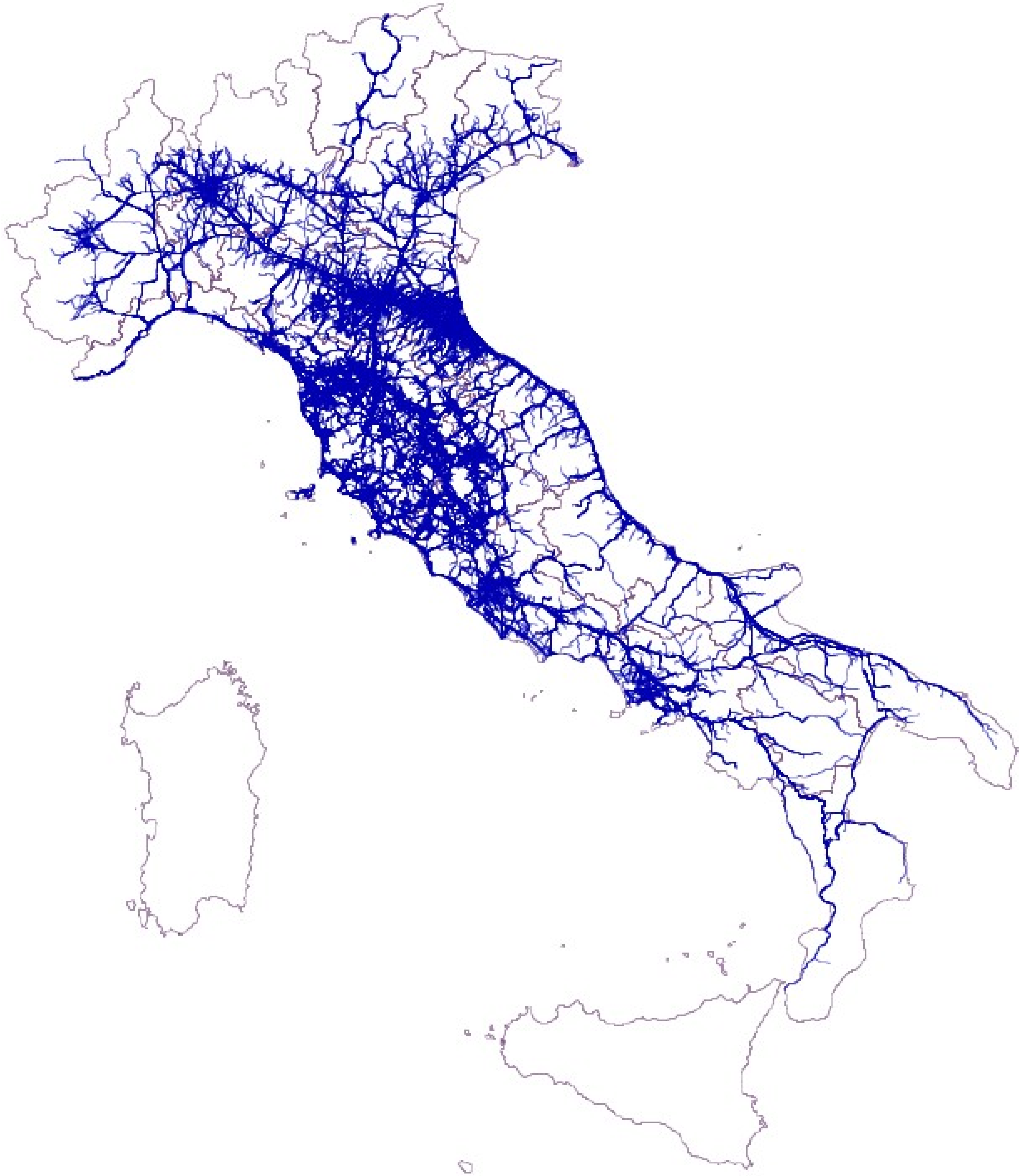
Onboard navigation devices send GPS tracks to central servers

```
Ide;Time;Lat;Lon;Height;Course;Speed;PDOP;State;NSat
...
8;22/03/07 08:51:52;50.777132;7.205580; 67.6;345.4;21.817;3.8;1808;4
8;22/03/07 08:51:56;50.777352;7.205435; 68.4;35.6;14.223;3.8;1808;4
8;22/03/07 08:51:59;50.777415;7.205543; 68.3;112.7;25.298;3.8;1808;4
8;22/03/07 08:52:03;50.777317;7.205877; 68.8;119.8;32.447;3.8;1808;4
8;22/03/07 08:52:06;50.777185;7.206202; 68.1;124.1;30.058;3.8;1808;4
8;22/03/07 08:52:09;50.777057;7.206522; 67.9;117.7;34.003;3.8;1808;4
8;22/03/07 08:52:12;50.776925;7.206858; 66.9;117.5;37.151;3.8;1808;4
8;22/03/07 08:52:15;50.776813;7.207263; 67.0;99.2;39.188;3.8;1808;4
8;22/03/07 08:52:18;50.776780;7.207745; 68.8;90.6;41.170;3.8;1808;4
8;22/03/07 08:52:21;50.776803;7.208262; 71.1;82.0;35.058;3.8;1808;4
8;22/03/07 08:52:24;50.776832;7.208682; 68.6;117.1;11.371;3.8;1808;4
...
```

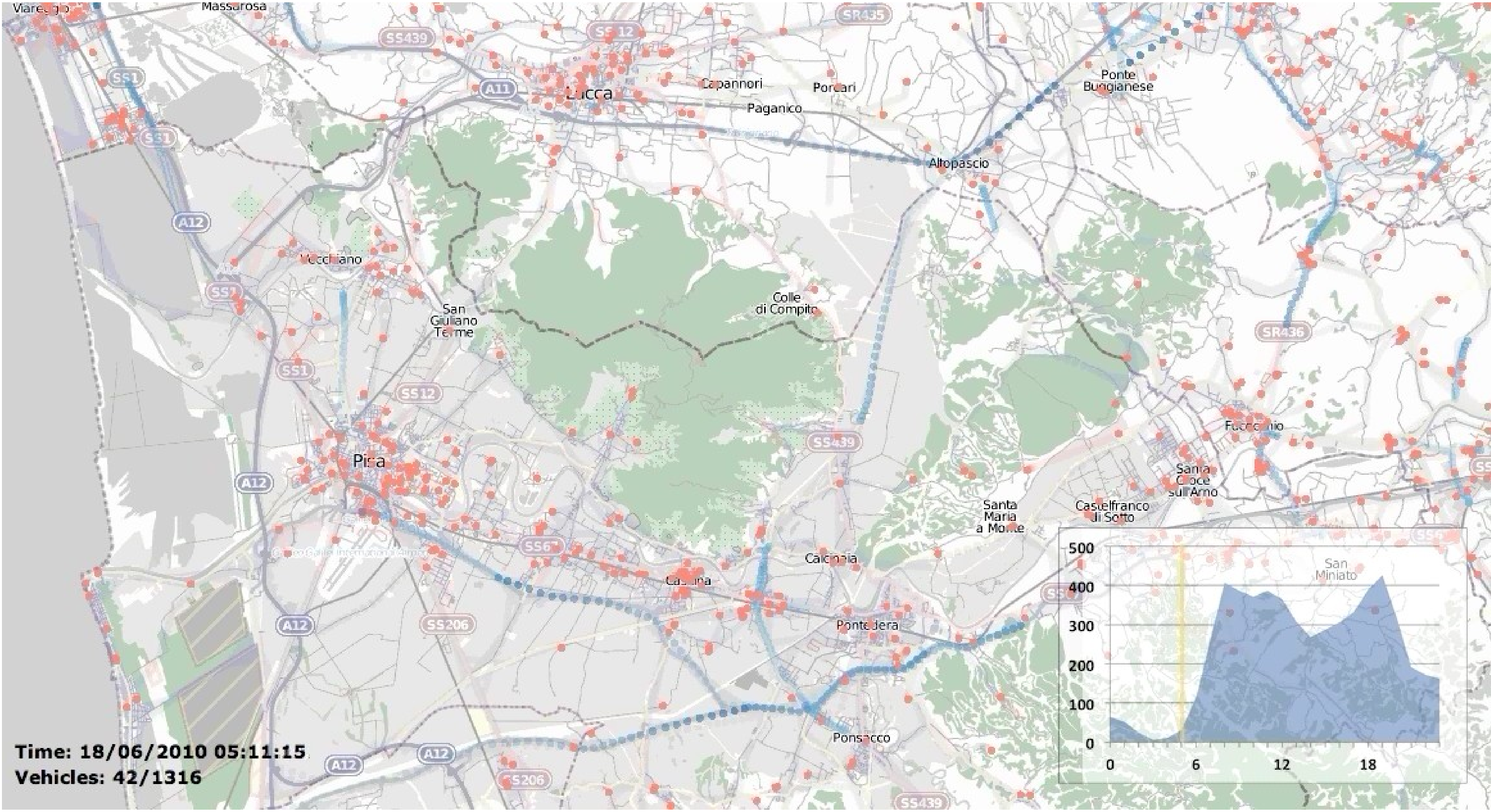


Sampling rate from few secs to 1-2 minutes

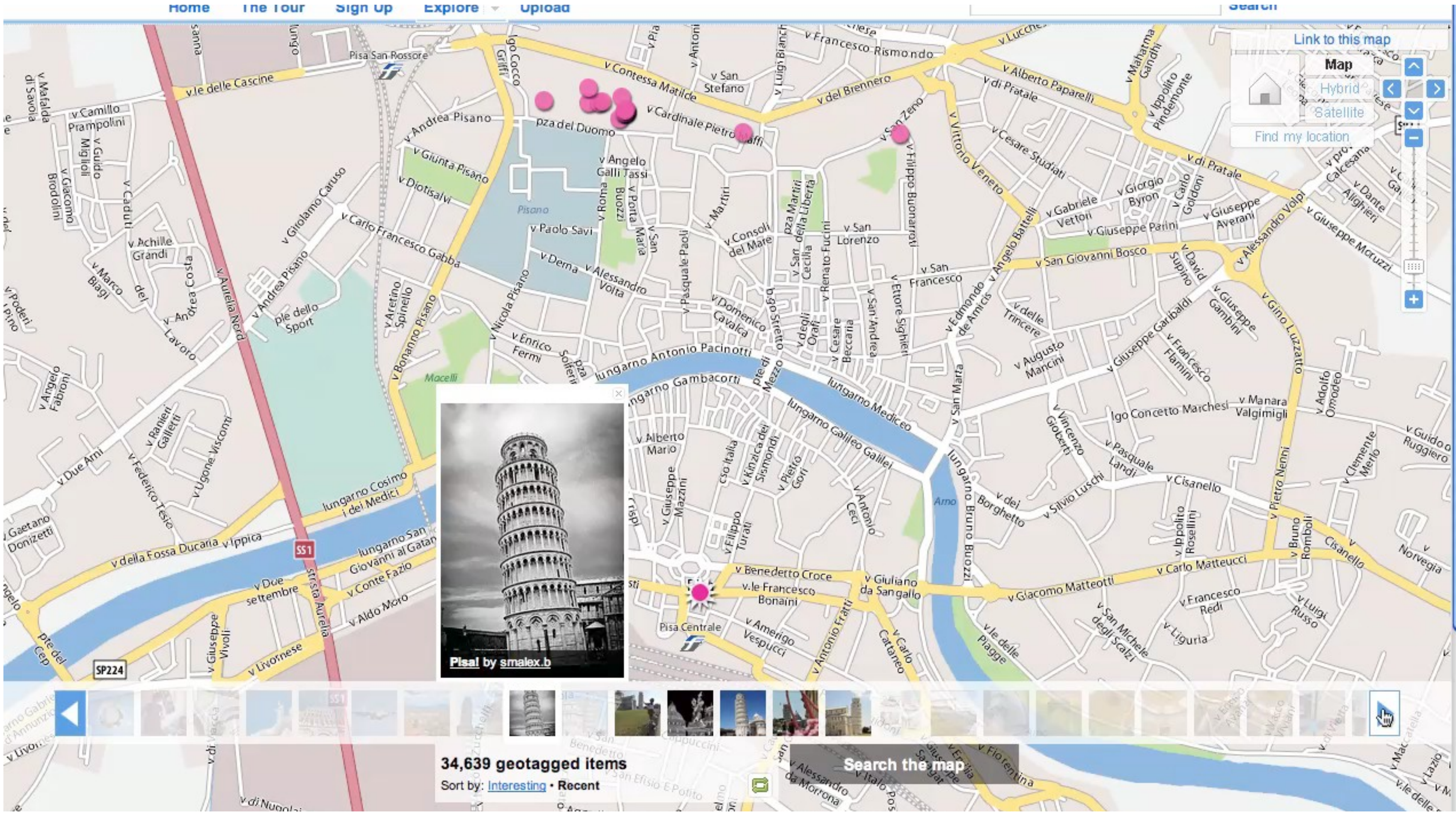
Spatial precision ~ 10 m



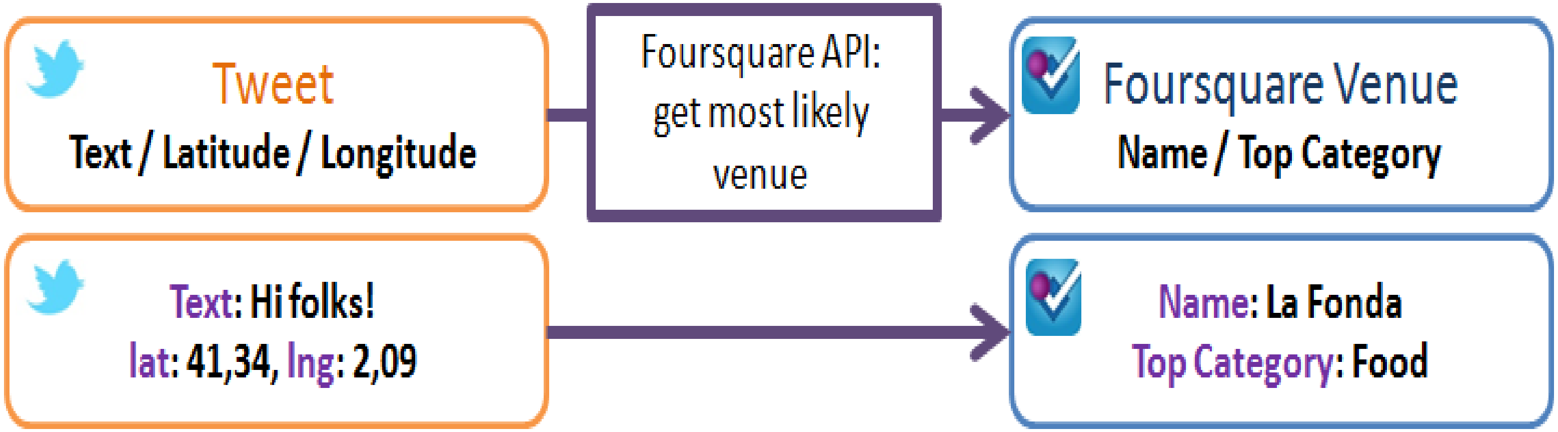
Urban Mobility Complexity: vehicles



Social networks



Twitter


Tweet
Text / Latitude / Longitude
Foursquare API: get most likely venue
Foursquare Venue
Name / Top Category
Text: Hi folks!
lat: 41,34, lng: 2,09
Name: La Fonda
Top Category: Food

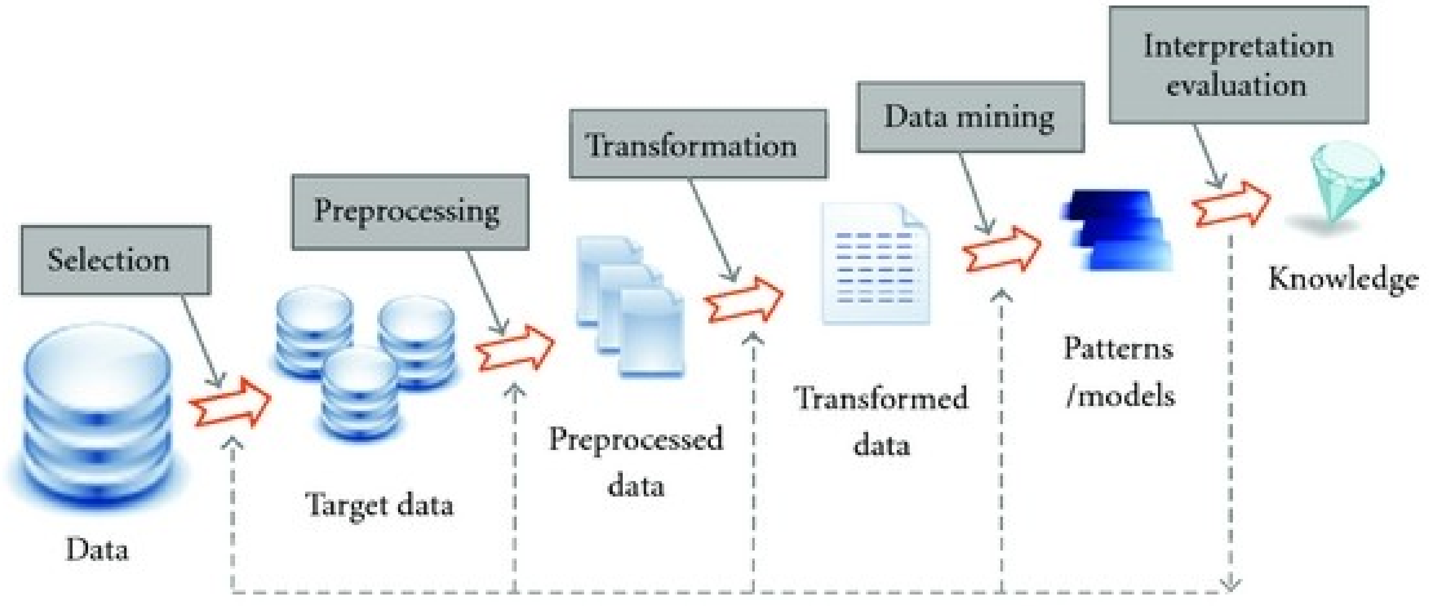
Research Impacts

- Moving object and trajectory data mining has many important, real-world applications driven by the real need
  - Ecological analysis (e.g., animal scientists)
  - Weather forecast
  - Traffic control
  - Location-based services
  - Homeland security (*e.g.*, border monitoring)
  - Law enforcement (*e.g.*, video surveillance)
  - ...

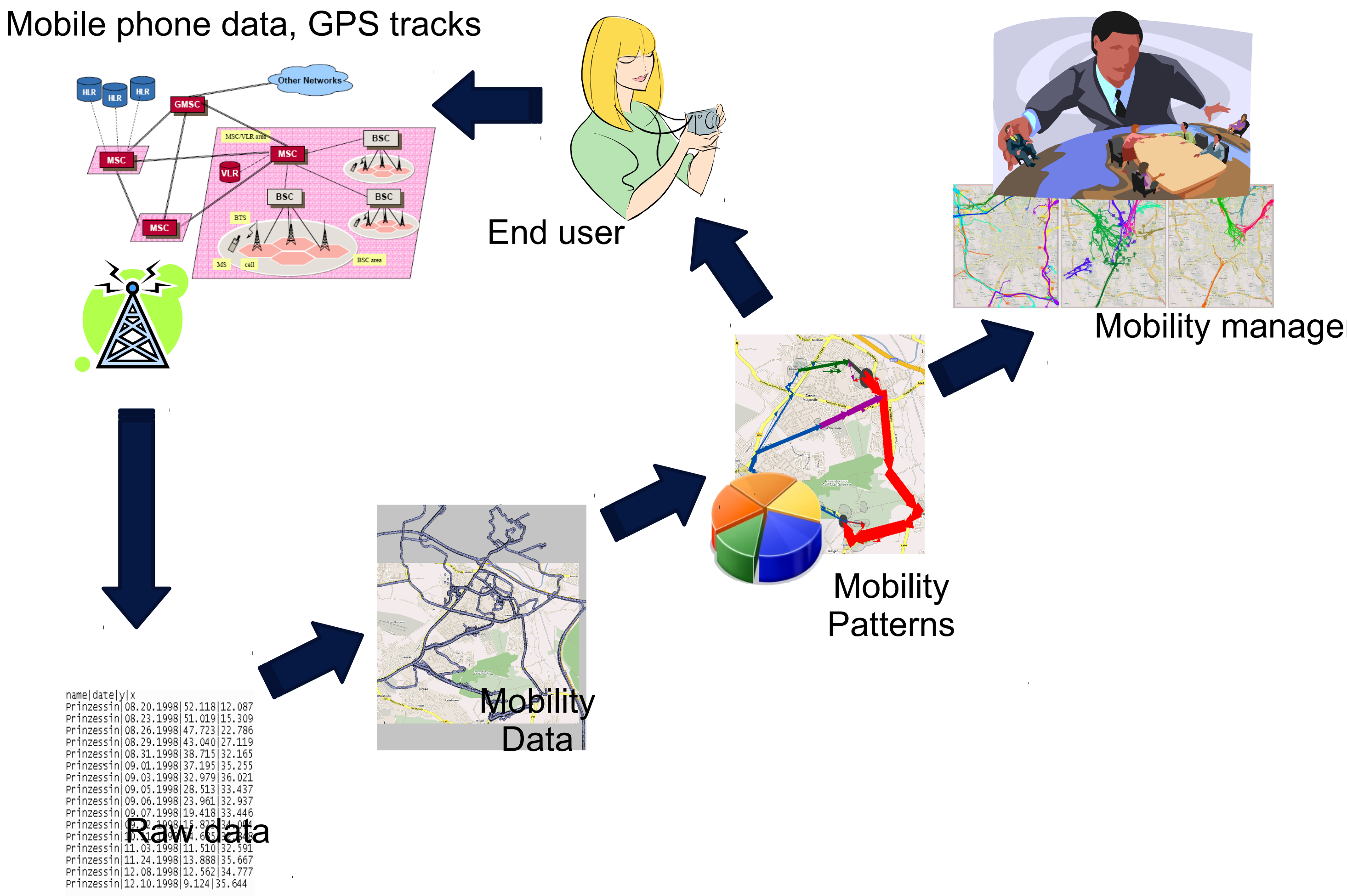
Complexity of the Moving Object Data

- Uncertainty
  - Sampling rate could be inconstant: From every few seconds transmitting a signal to every few days transmitting one
  - Data can be sparse: A recorded location every 3 days
- Noise
  - Erroneous points (e.g., a point in the ocean)
- Background
  - Cars follow underlying road network
  - Animals movements relate to mountains, lakes, ...
- Movement interactions
  - Affected by nearby moving objects

Knowledge Discovery process



The KDD process for Mobility Data

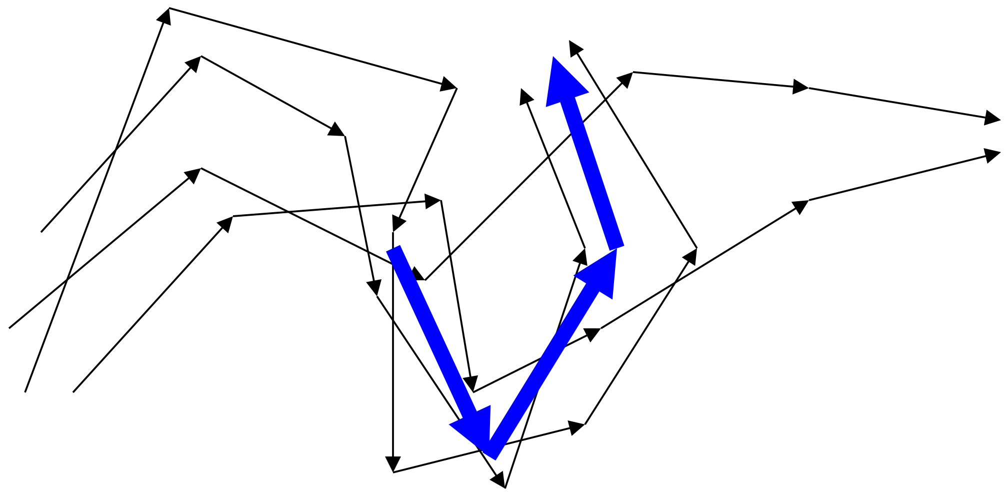


Data mining …

- … is about finding models that emerge directly from the data
  - Data-driven vs hypothesis-driven analysis
- Local models
  - **Patterns**: find groups of items/events that frequently co-occur in the data
- Global models
  - **Clustering**: find a natural partition of the data into groups of similar objects
  - **Classification**: find a function that predicts the value of a specified variable given the values of the others

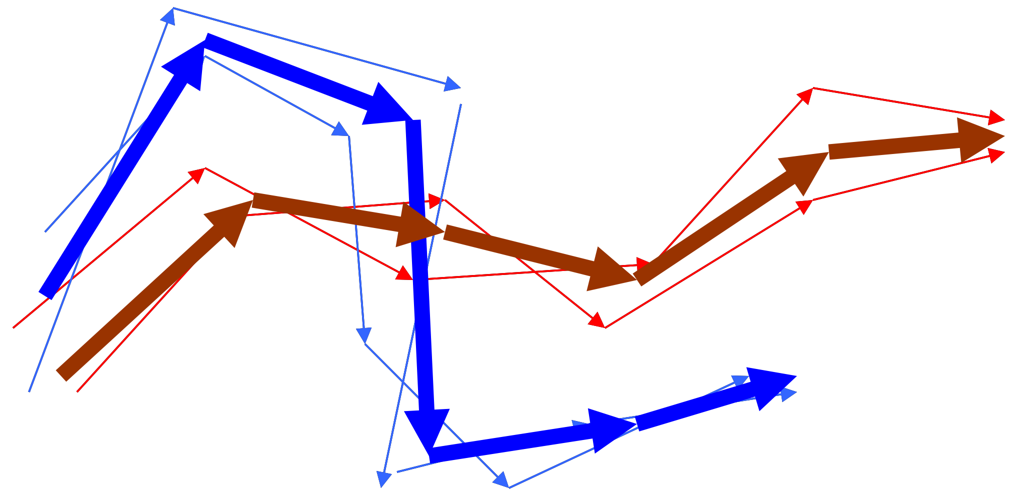
Trajectory patterns

- Discover frequently followed itineraries



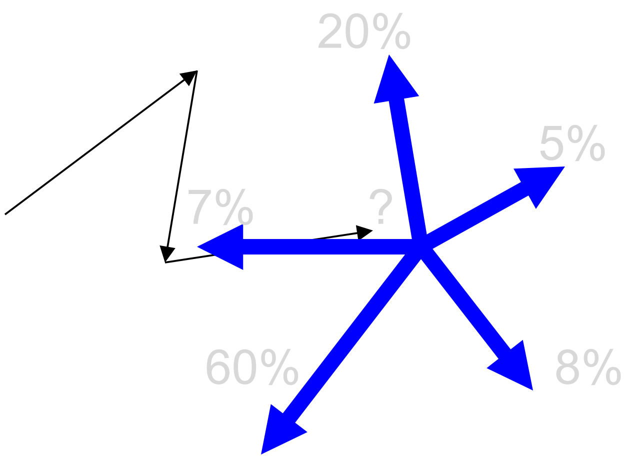
Trajectory Clustering

- Group together similar trajectories
- For each group produce a summary



Trajectory classification and prediction

- Extract behaviour rules from history
- Use rules to predict behaviour of future users



GPS processing and statistics

Raw GPS Data

ID $\underbrace{}^{\text{Timestamp}}$ Latitude Longitude $\underbrace{}^{\text{Others (optional)}}$

946826,14/06/10 14:08:54,43964259,10216895,0,0,1,0,0

457380,13/06/10 22:05:27,43682201,10408320,0,0,3,0,0

457380,13/06/10 22:06:00,43682688,10408501,10,10,3,1,33

457380,13/06/10 22:06:34,43683609,10409146,14,24,3,1,115

457380,13/06/10 22:07:09,43685653,10410117,52,18,3,1,241

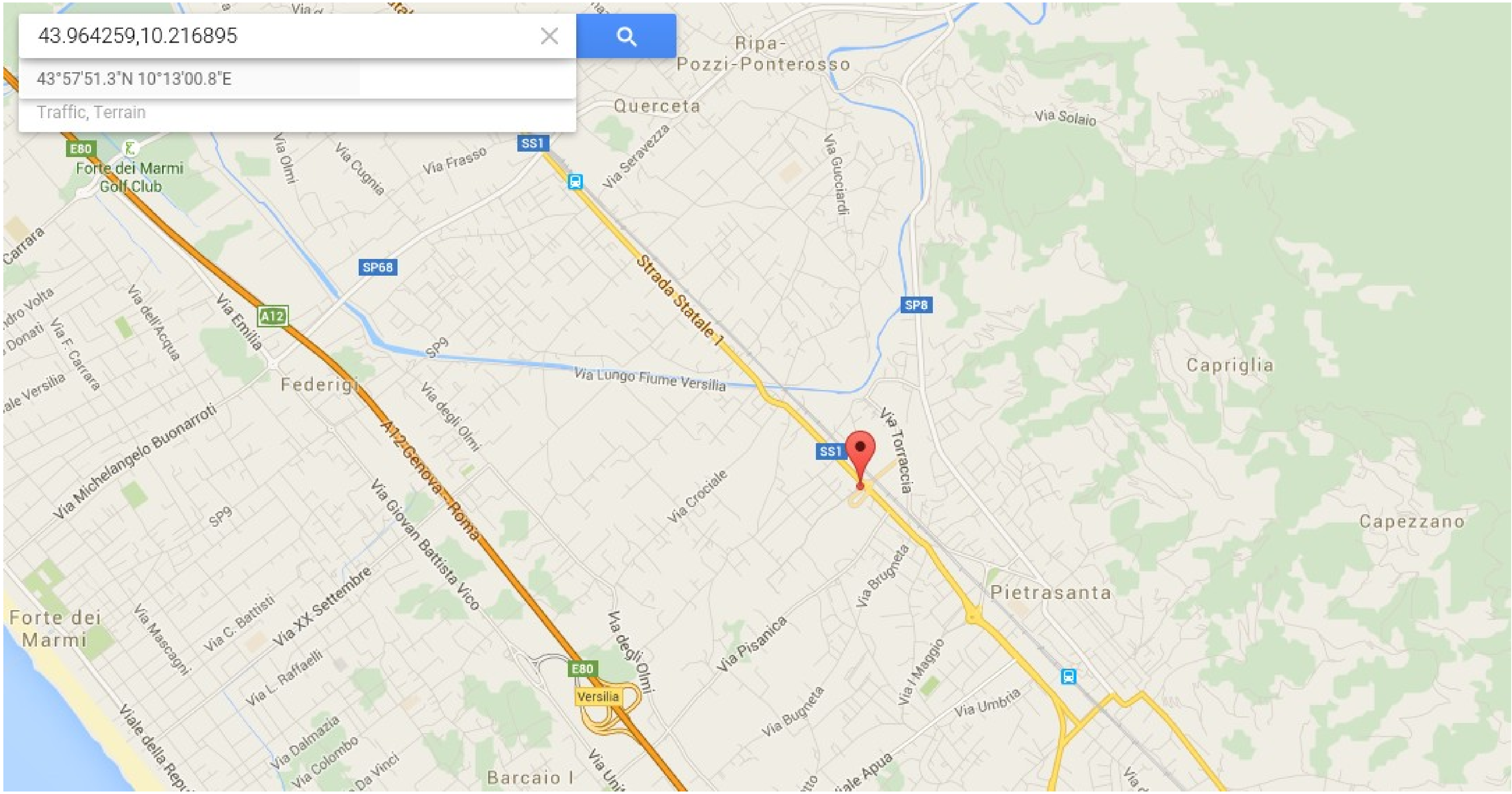
457380,13/06/10 22:07:43,43689775,10412032,50,18,3,1,484

457380,13/06/10 22:08:19,43692906,10413910,32,356,3,1,401

457380,13/06/10 22:08:53,43690801,10415016,60,126,3,1,279

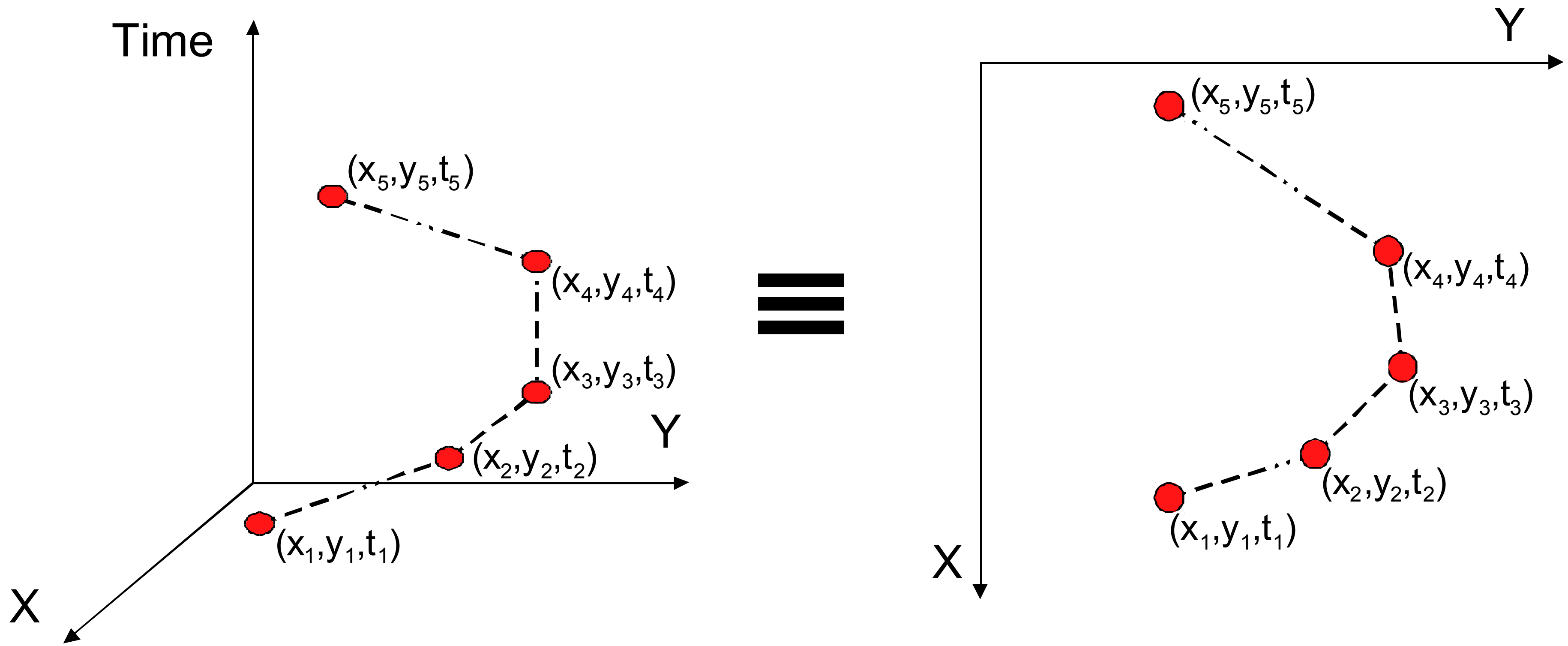
Sample point on the map

946826,14/06/10 14:08:54,43964259,10216895,0,0,1,0,0



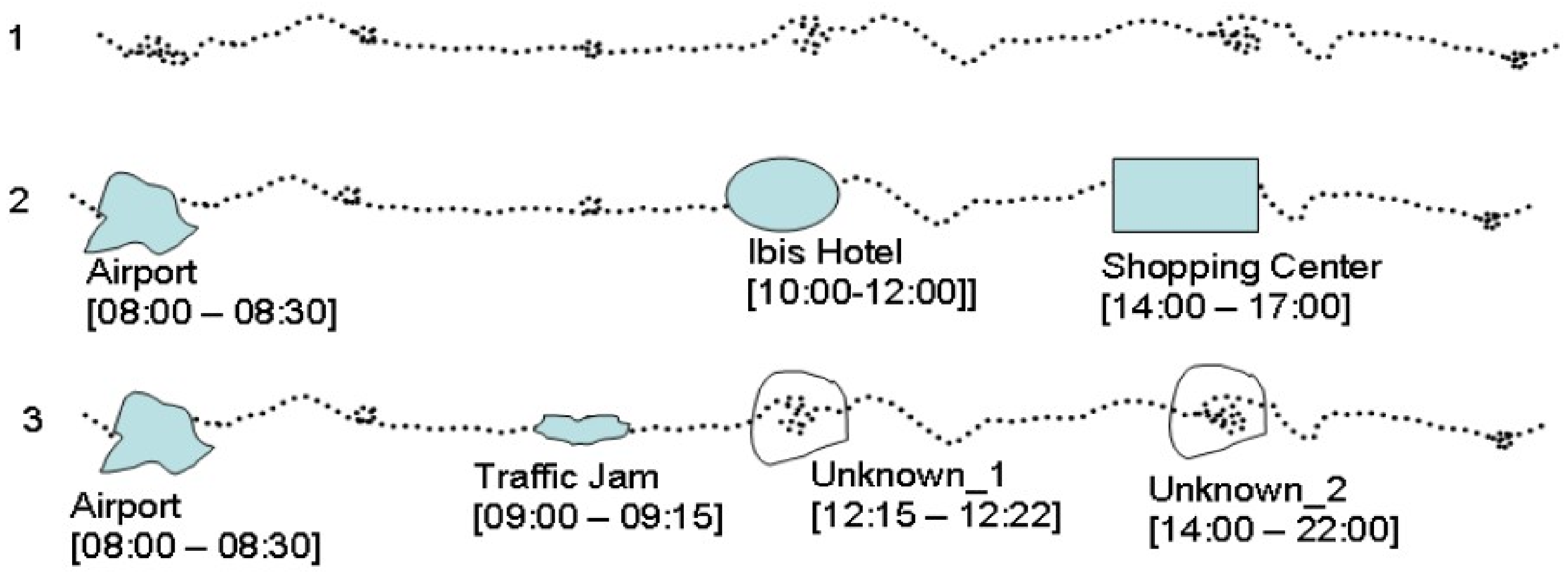
Trajectory data

- Mobility of an object is described by a set of trips
- Each trip is a trajectory, i.e. a sequence of time-stamped locations



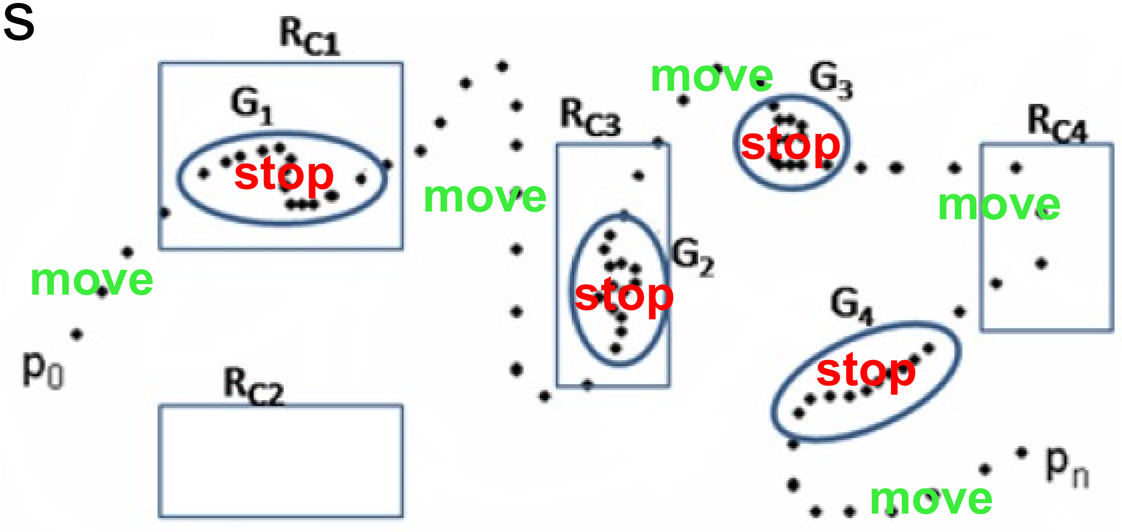
Trajectory reconstruction

- Raw data forms a continuous stream of points
- How to cut it into stops and trips?
  - Example on smart phone traces:



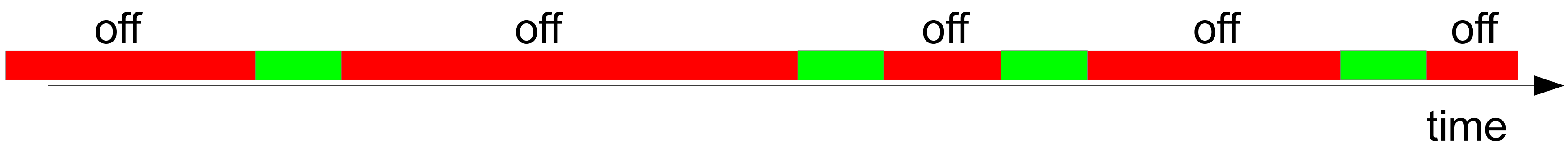
Trajectory reconstruction

- General criteria based on speed
  - If it moves very little (threshold Th_S) over a significant time interval (threshold Th_T) then it is practically a stop
  - Trajectory (trip) = contiguous sequence of points between two stops

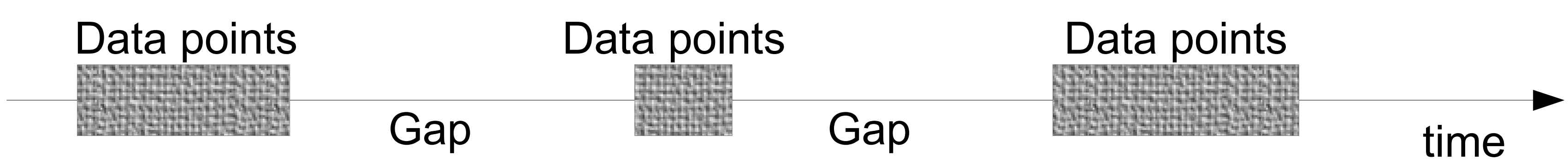


Trajectory reconstruction

- Special cases, easier to treat
  - Stop explicitly in the data: e.g. engine status on/off
    - Simply "cut" trajectories on status transitions

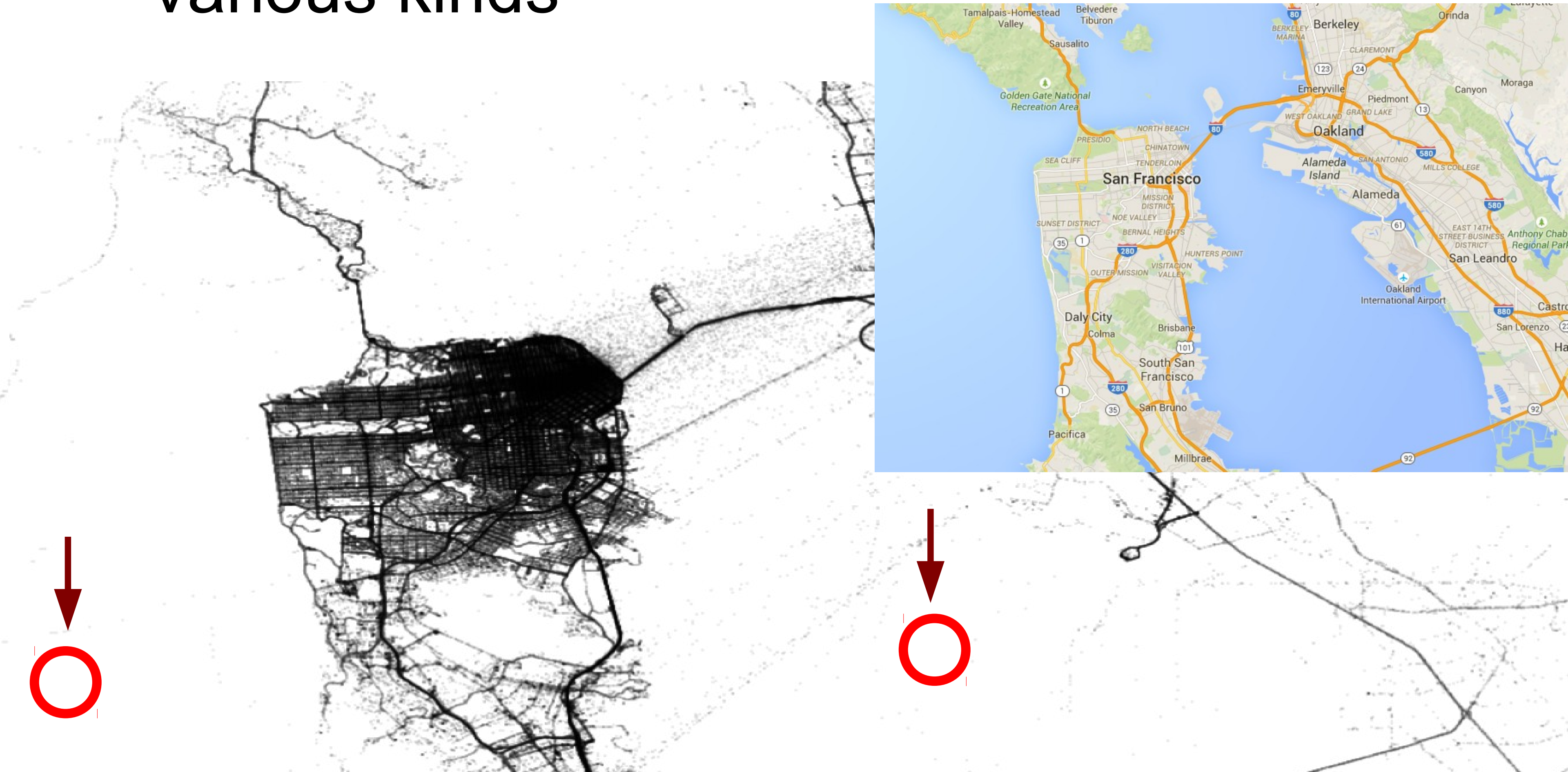


  - Device is off during stops:
    - Typical of cars data
    - A stop results in a time gap in the data
    - Exceptions: short stops might remain undetected



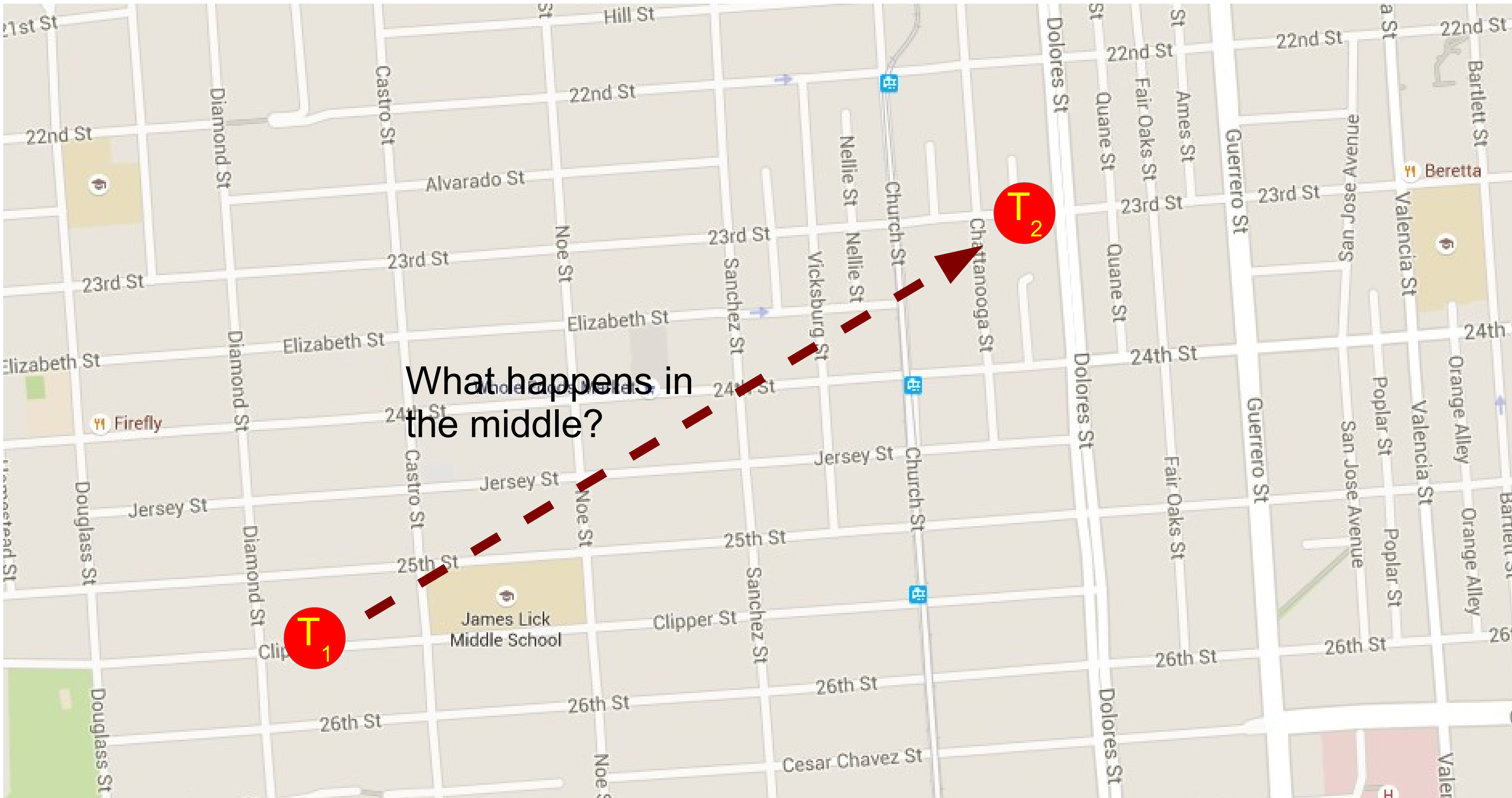
Outliers / noise

- Single points might contain errors of various kinds



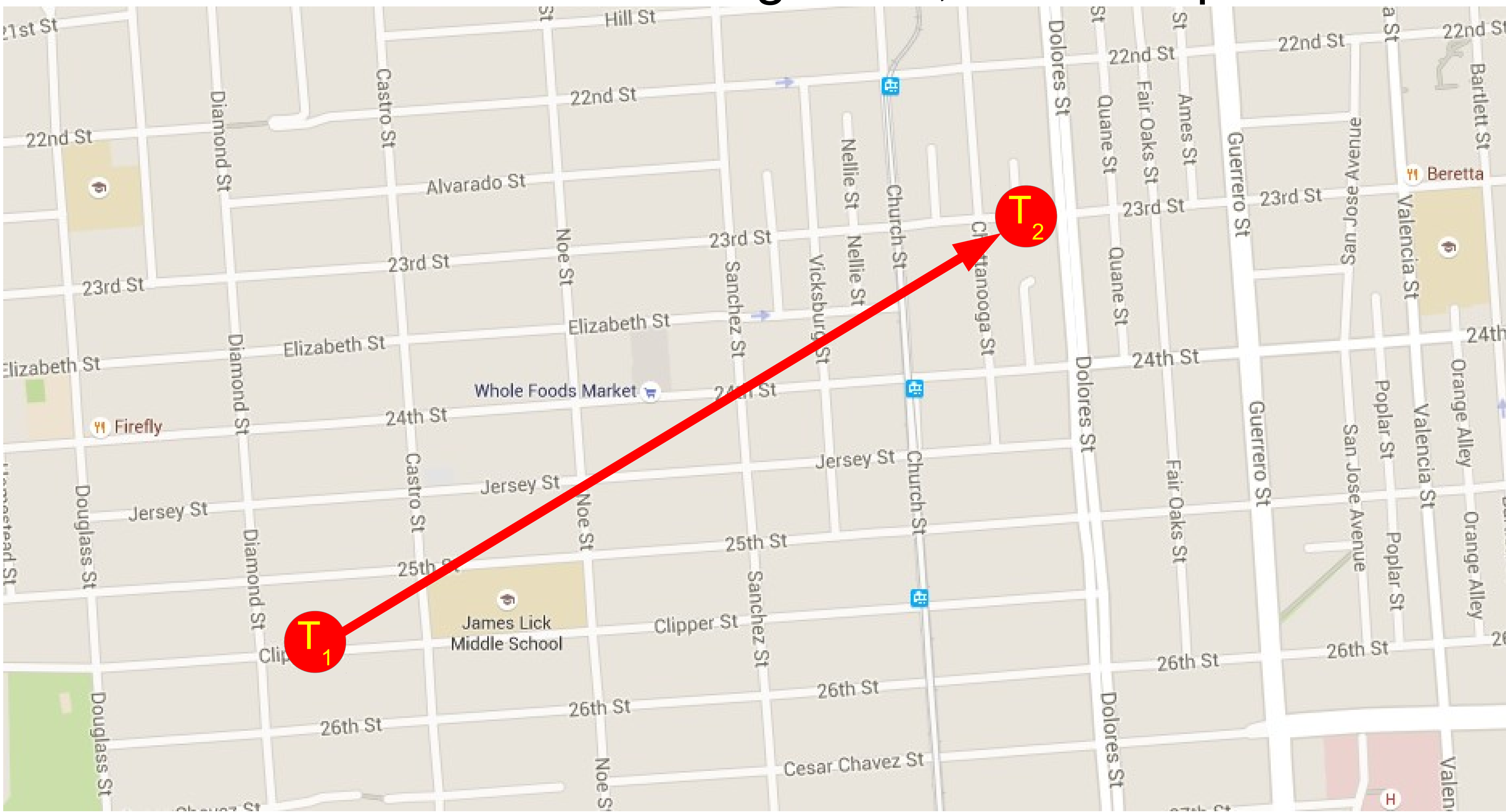
Gaps

- Sometimes the space/time gap between consecutive points is significant



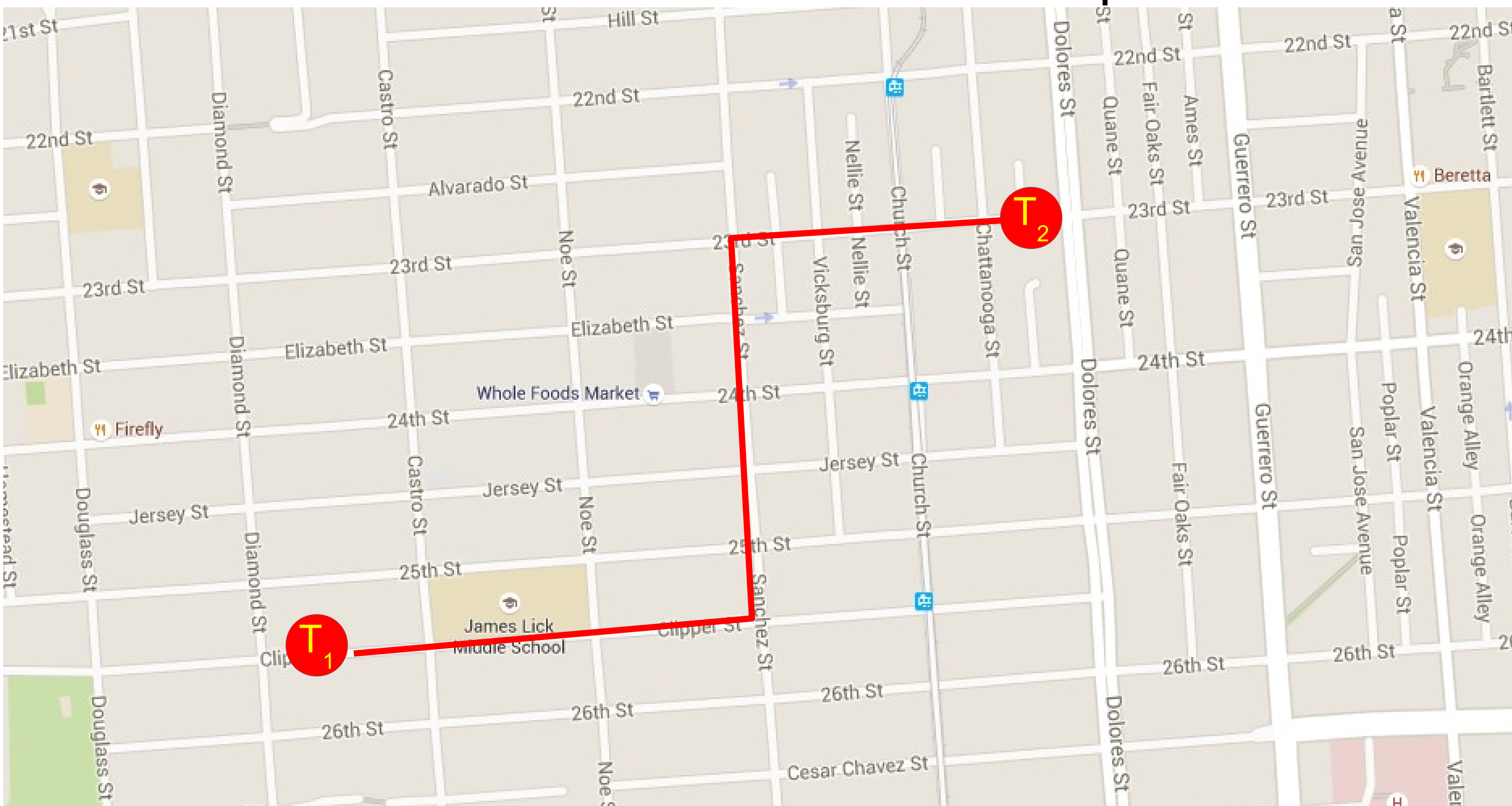
Free vs. constrained movement

- Typical solutions:
  - Free movement => straight line, uniform speed

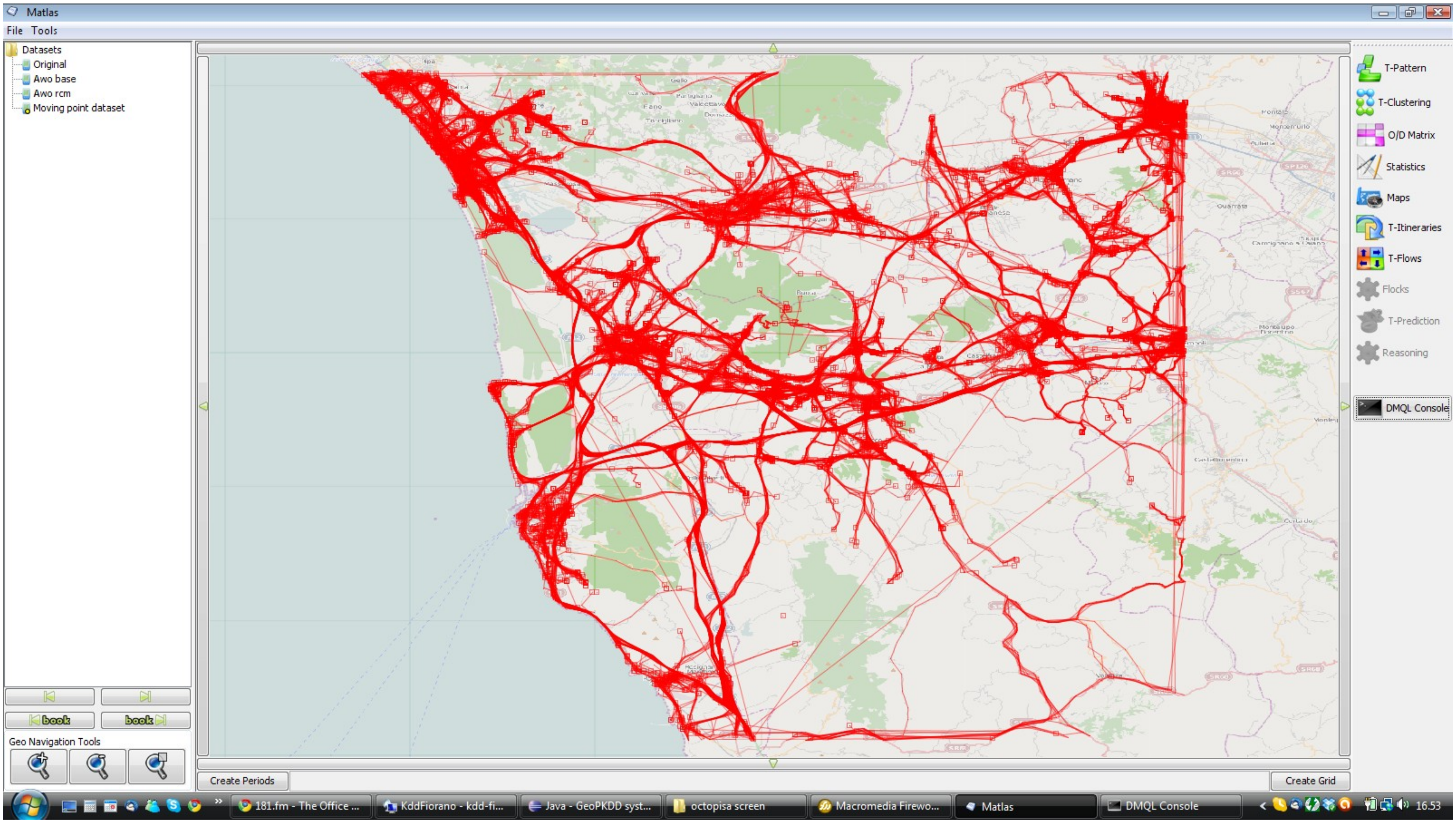


Free vs. constrained movement

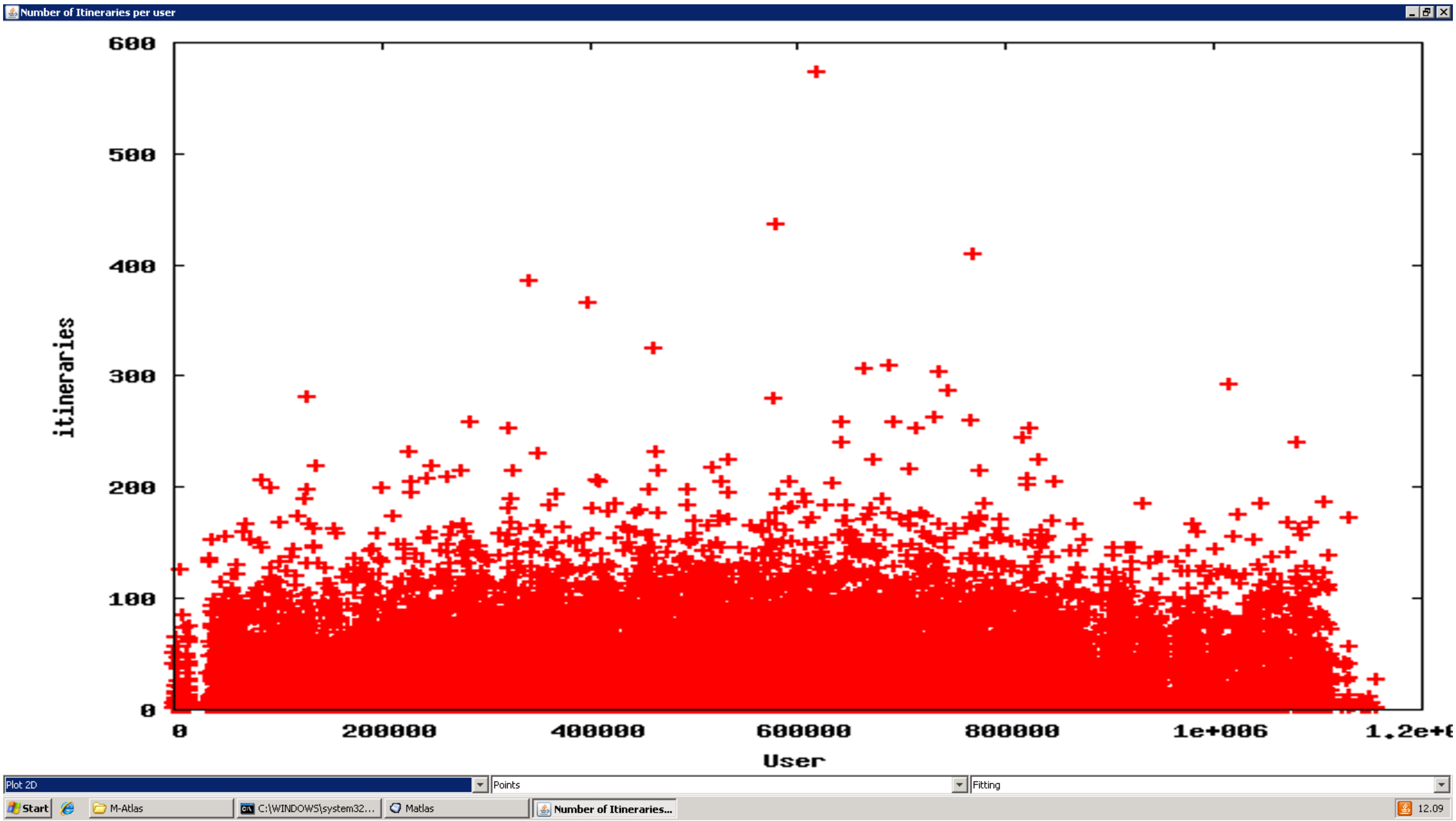
- Typical solutions:
  - Constrained movement => shortest path



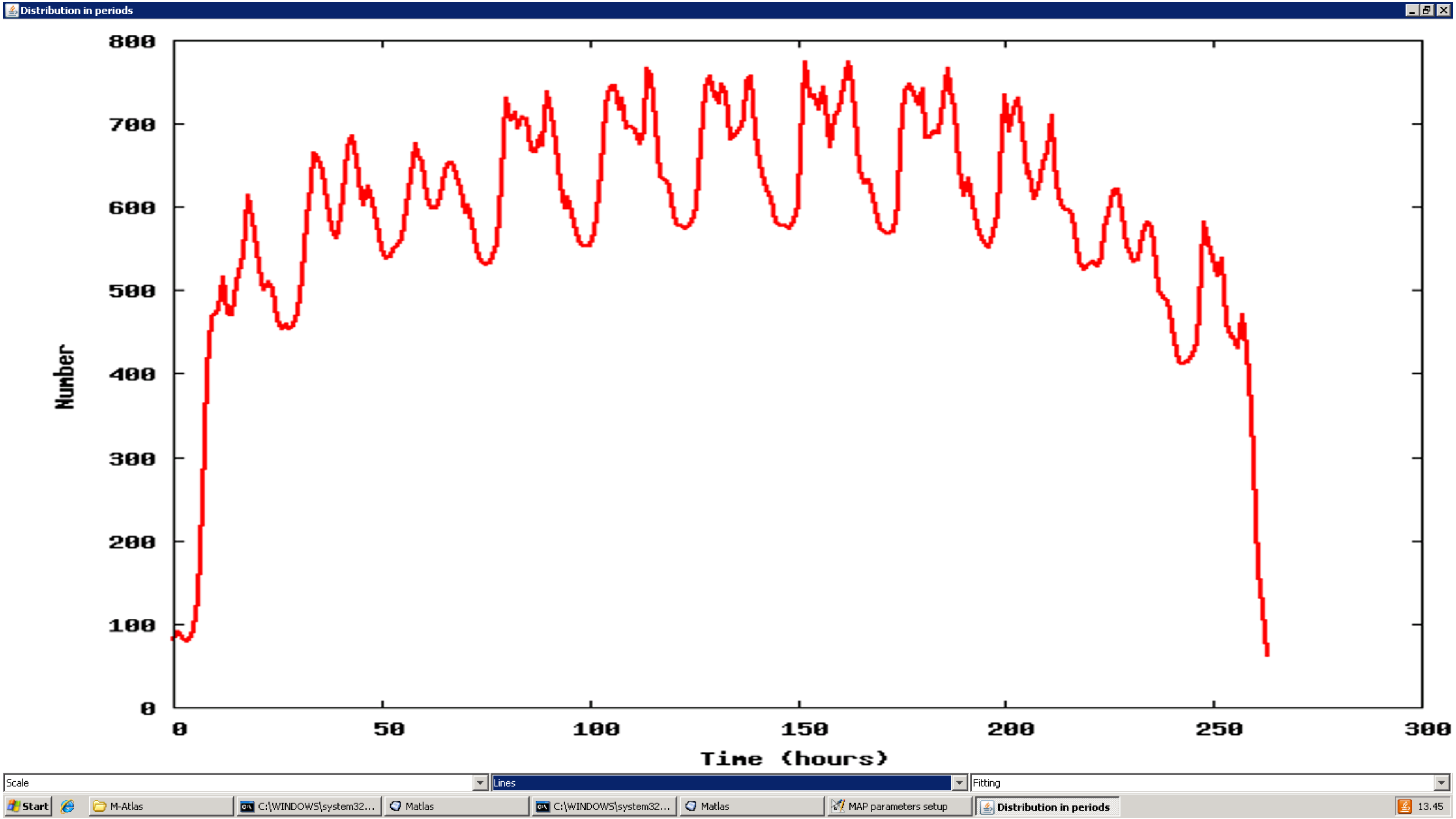
A Dataset (2/7 → 12/7)



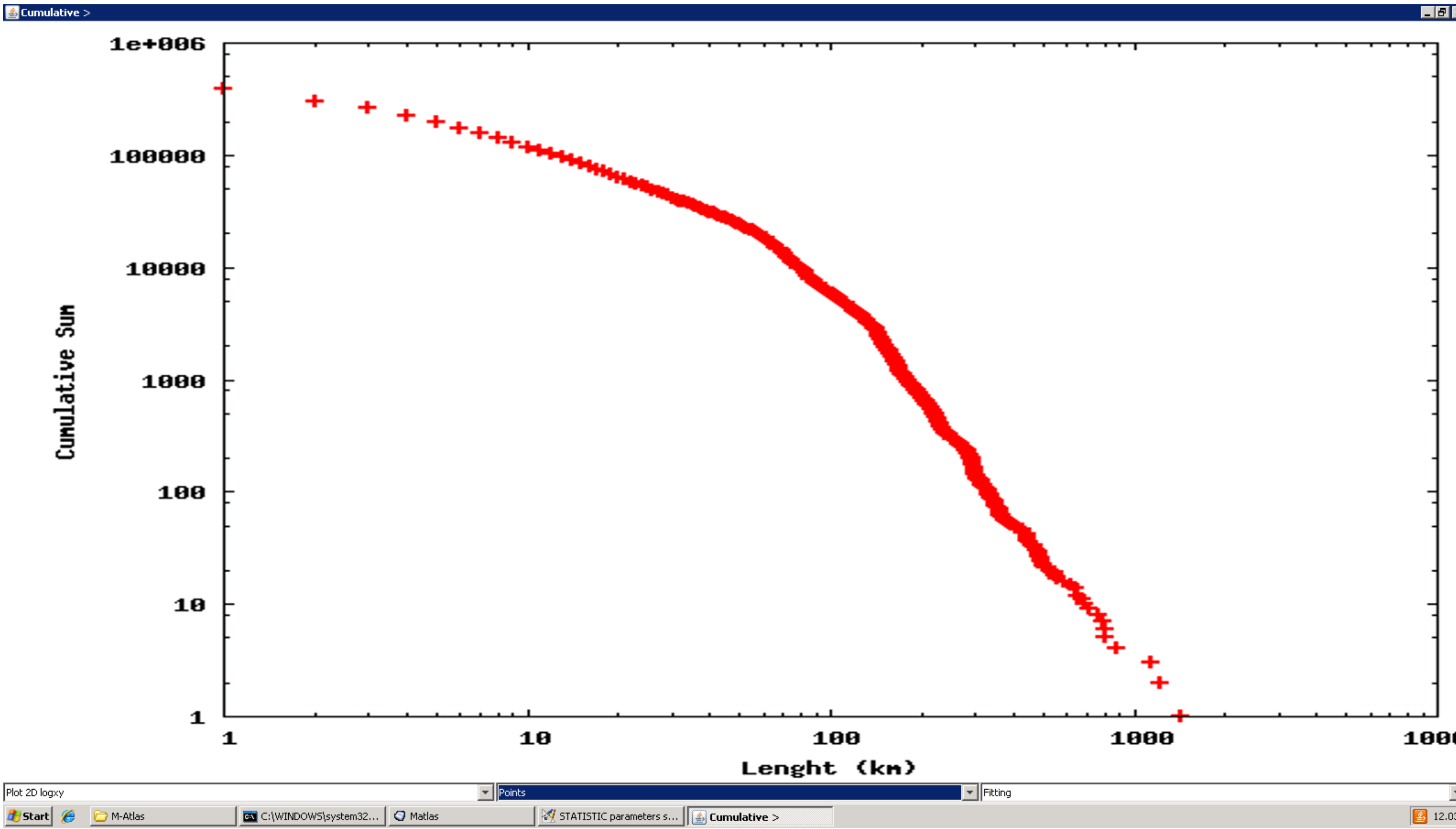
Number of trajectories per User



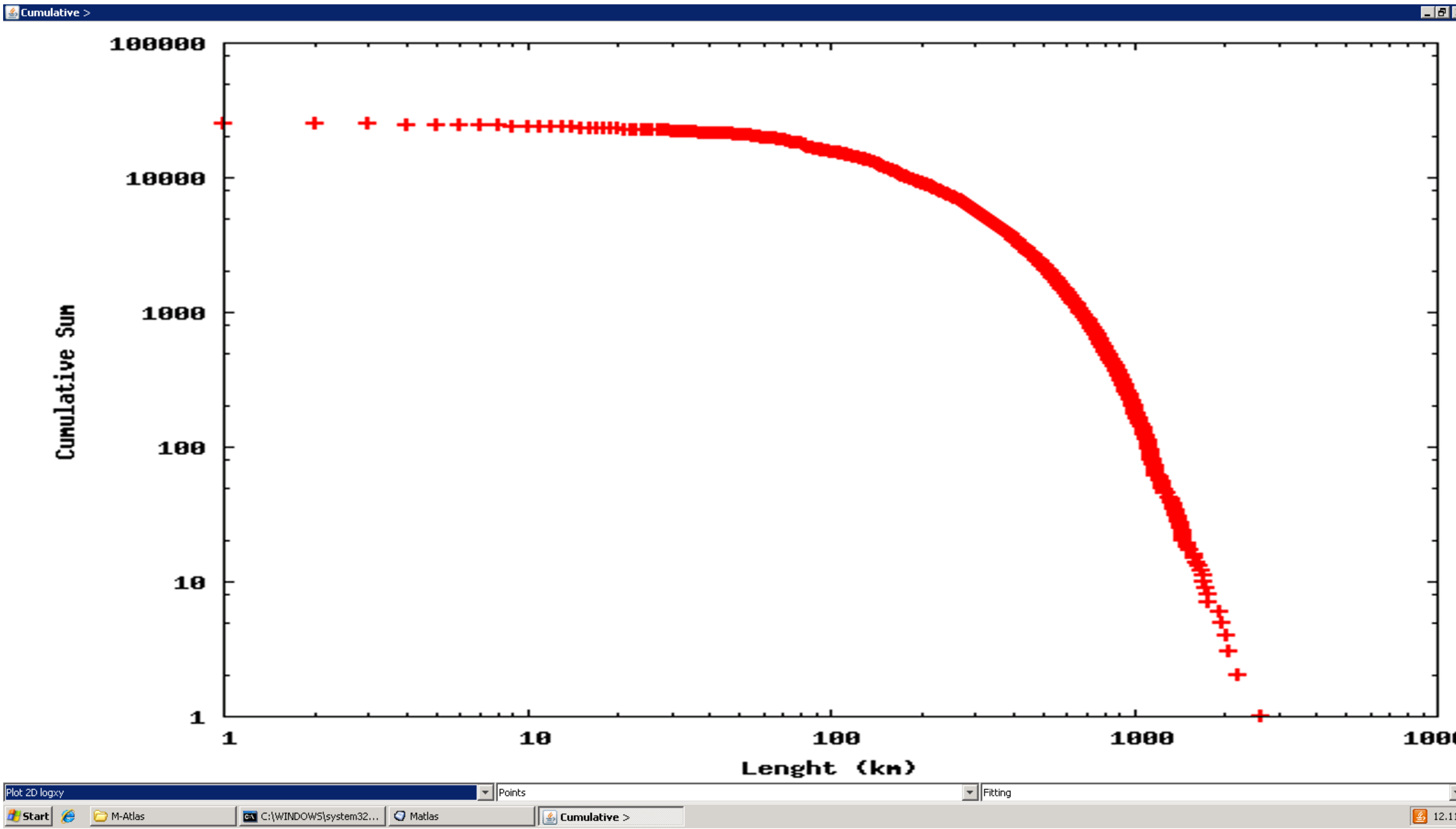
Distribution in periods (hours)



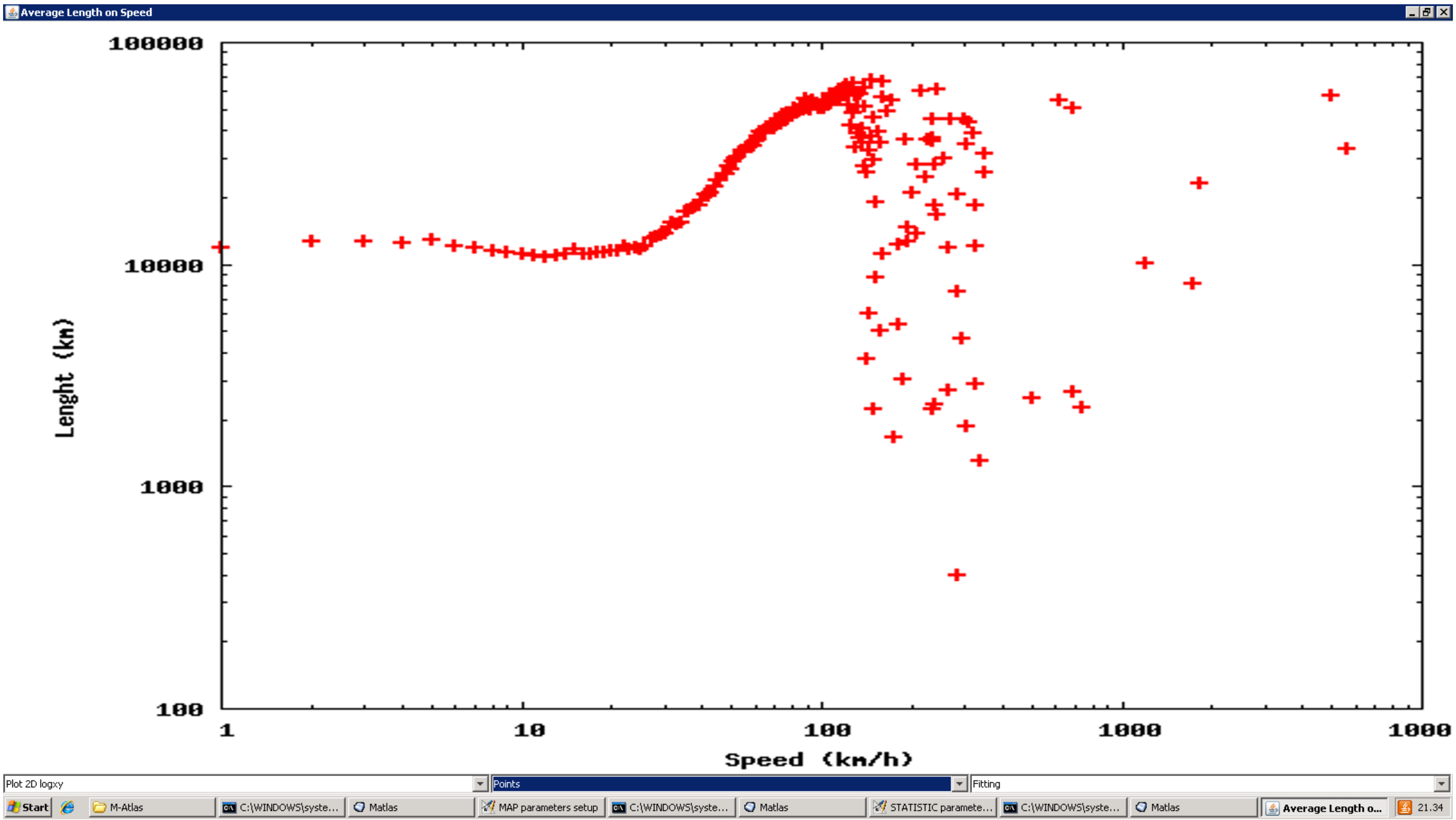
Distribution of lengths (Cumulative)



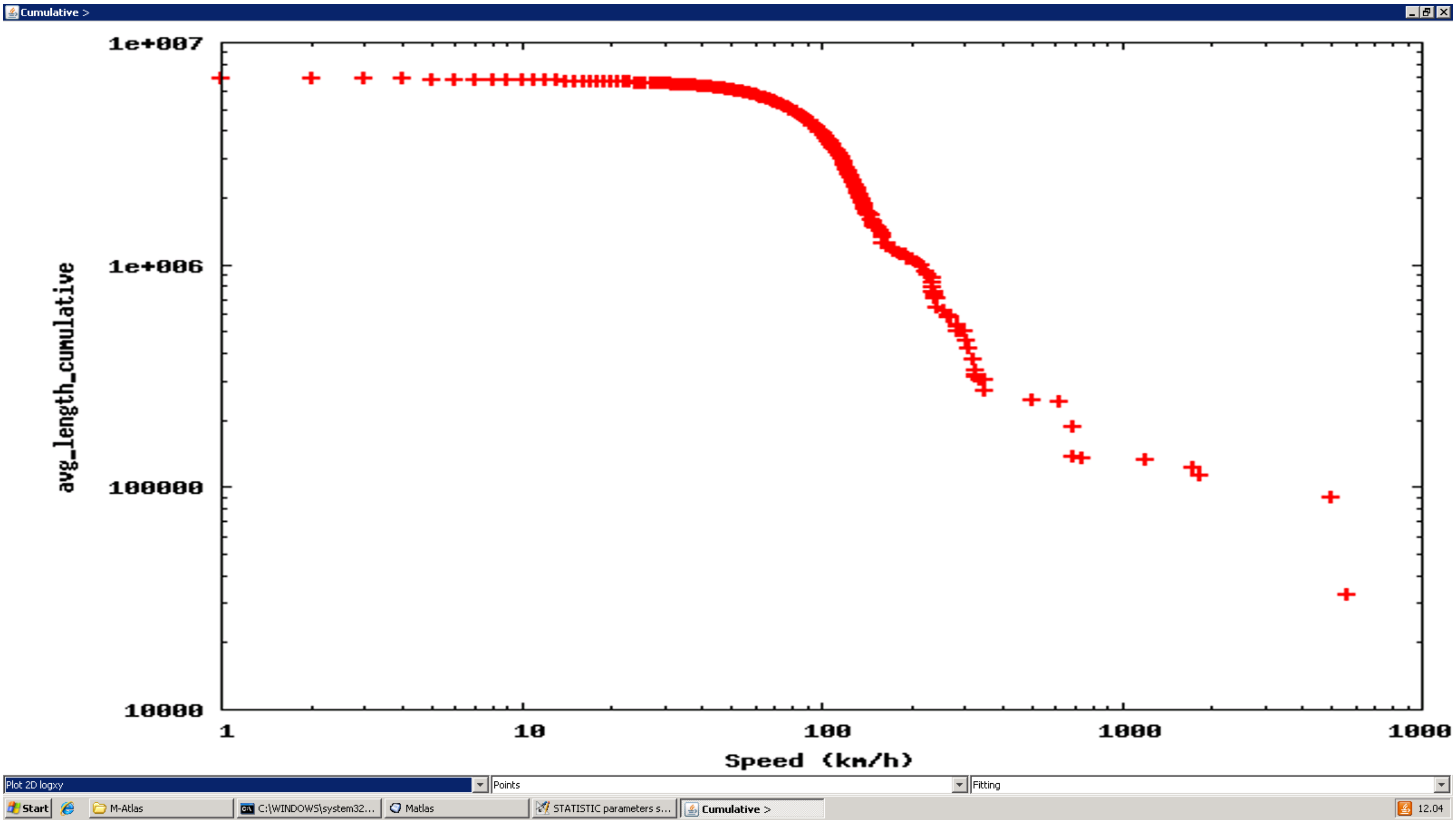
Distribution of Lengths per User (Cumulative)



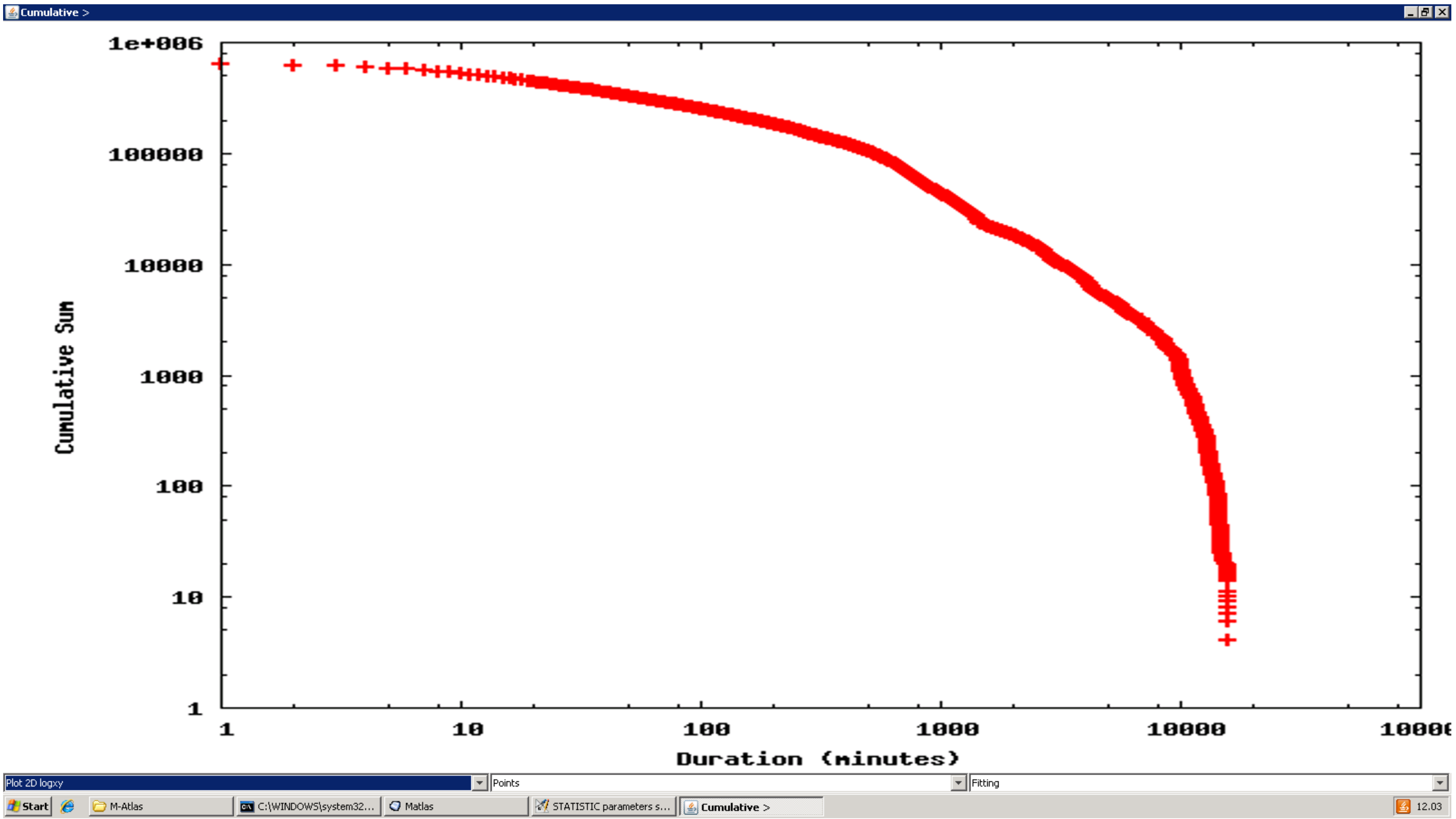
Average length on speed



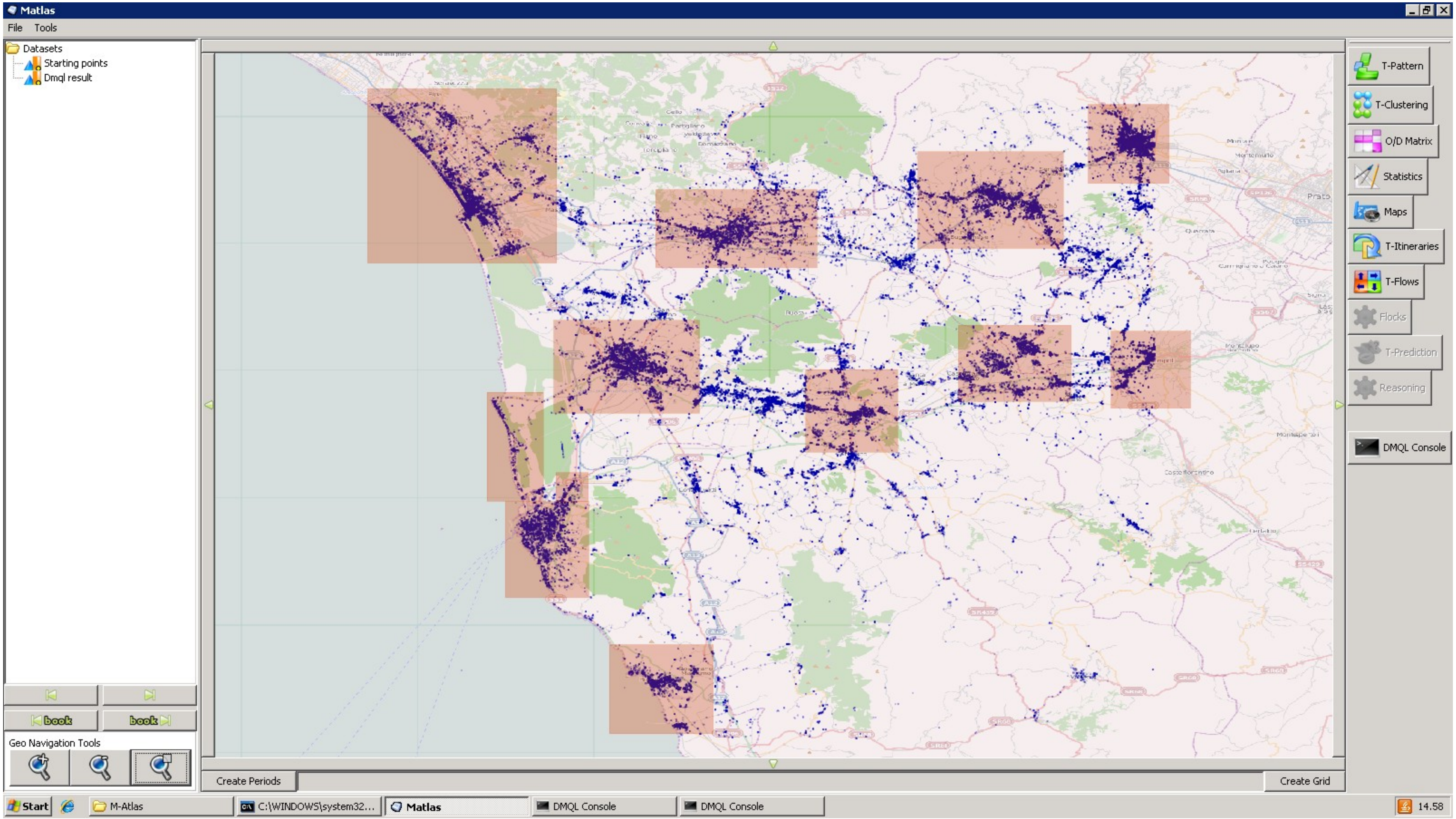
Average length on speed (Cumulative)



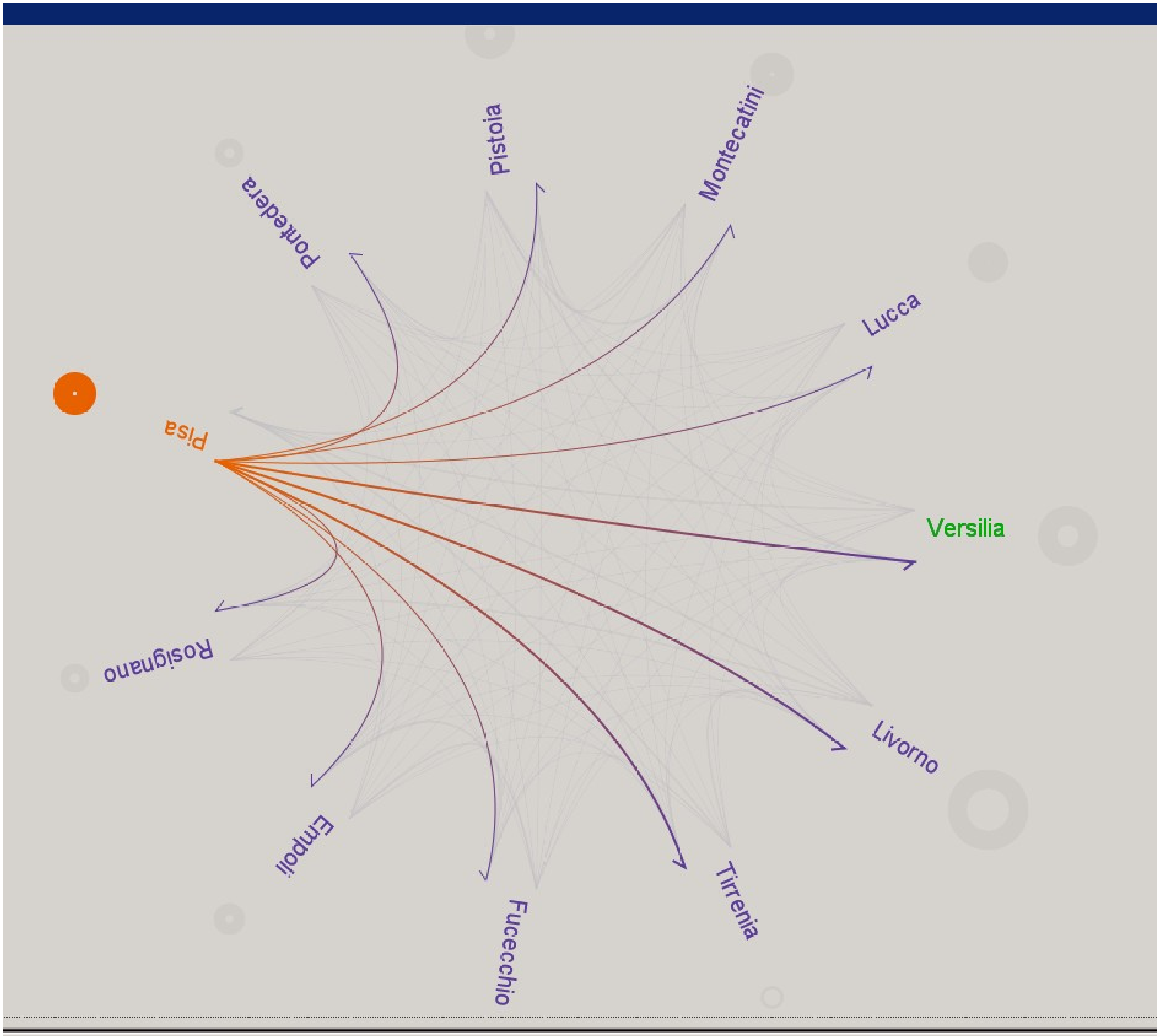
Distribution of Durations (Cumulative)



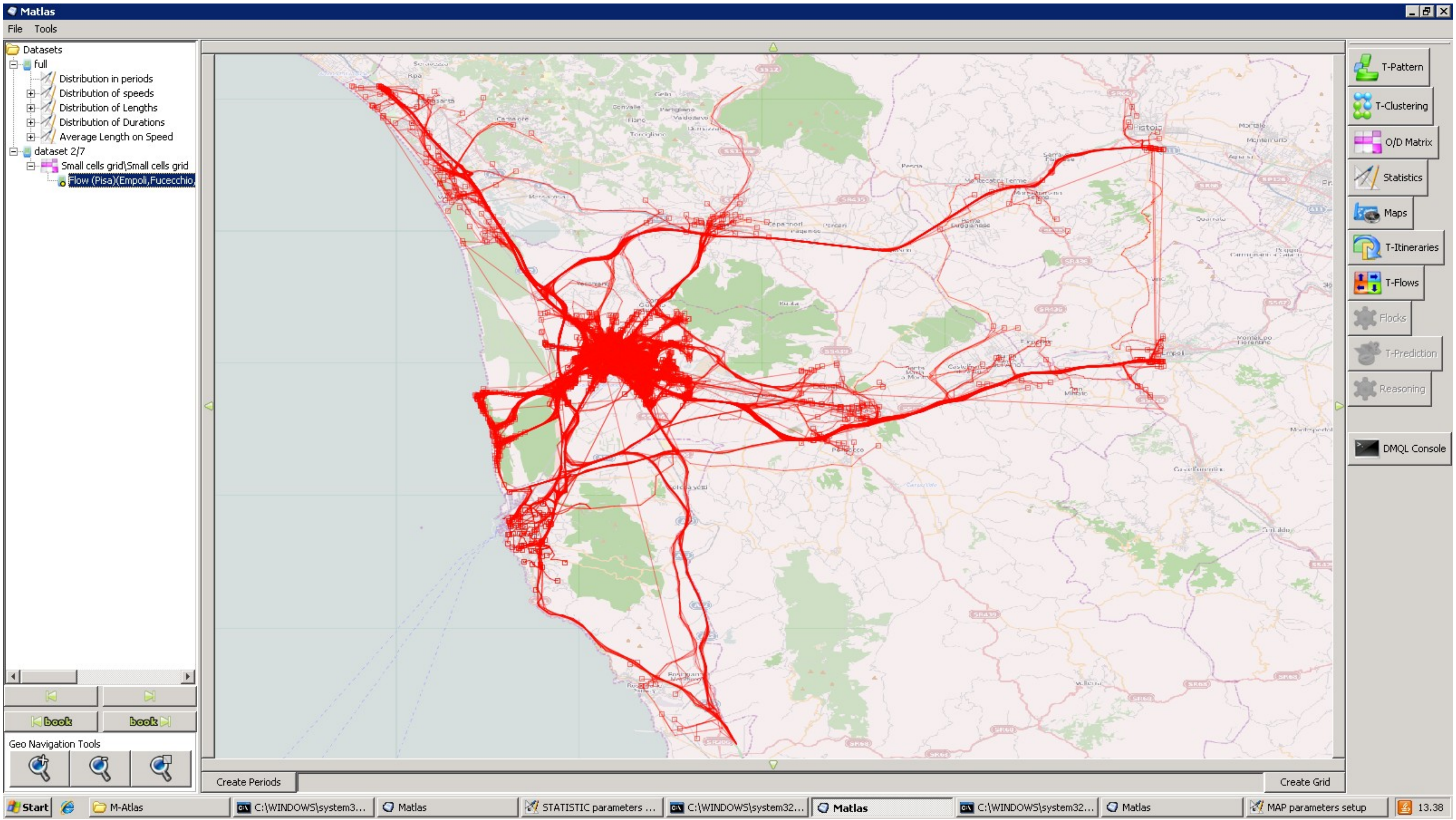
Cities (Approximation)



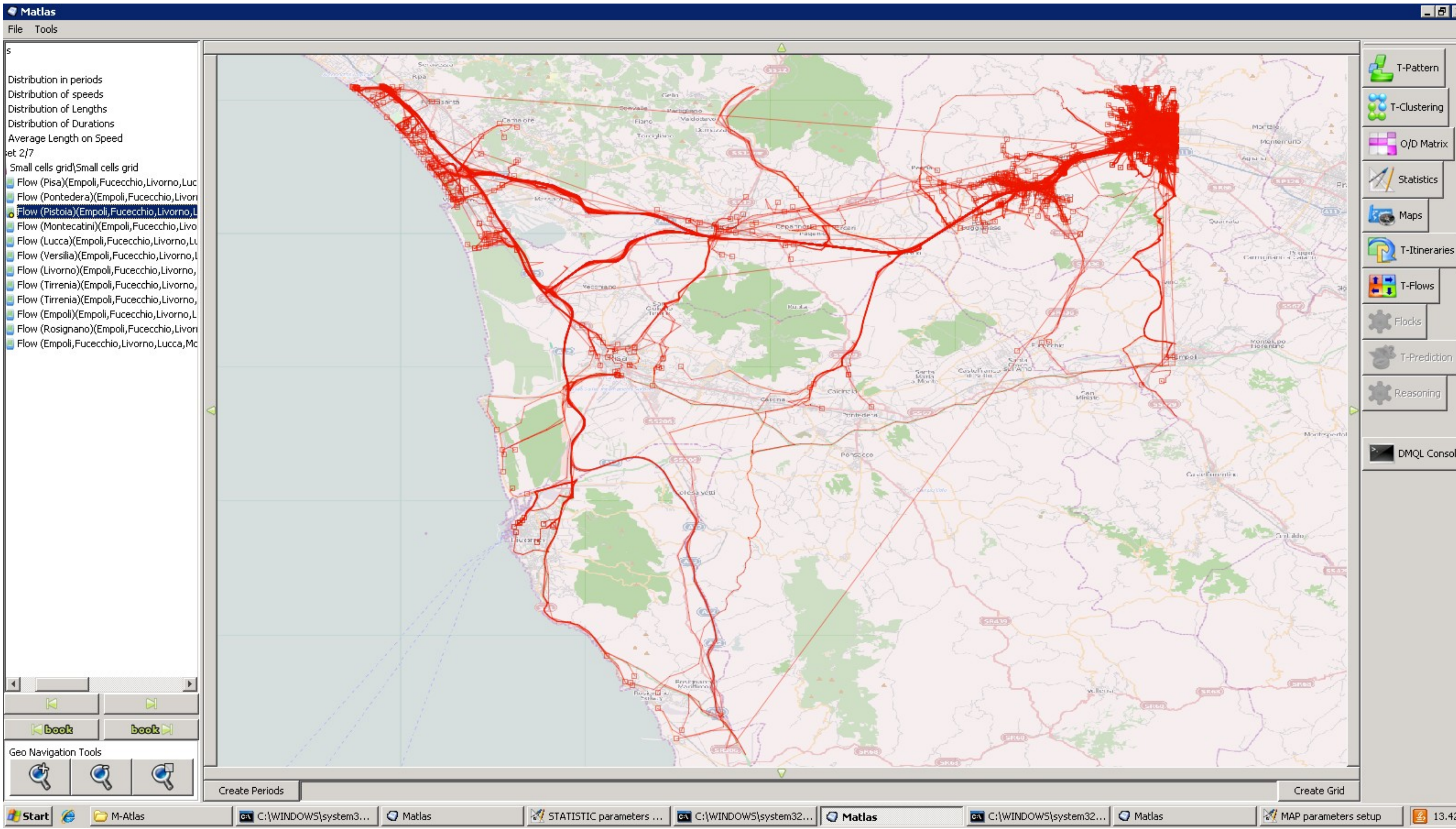
OD Matrix (Cities ←→ Cities)



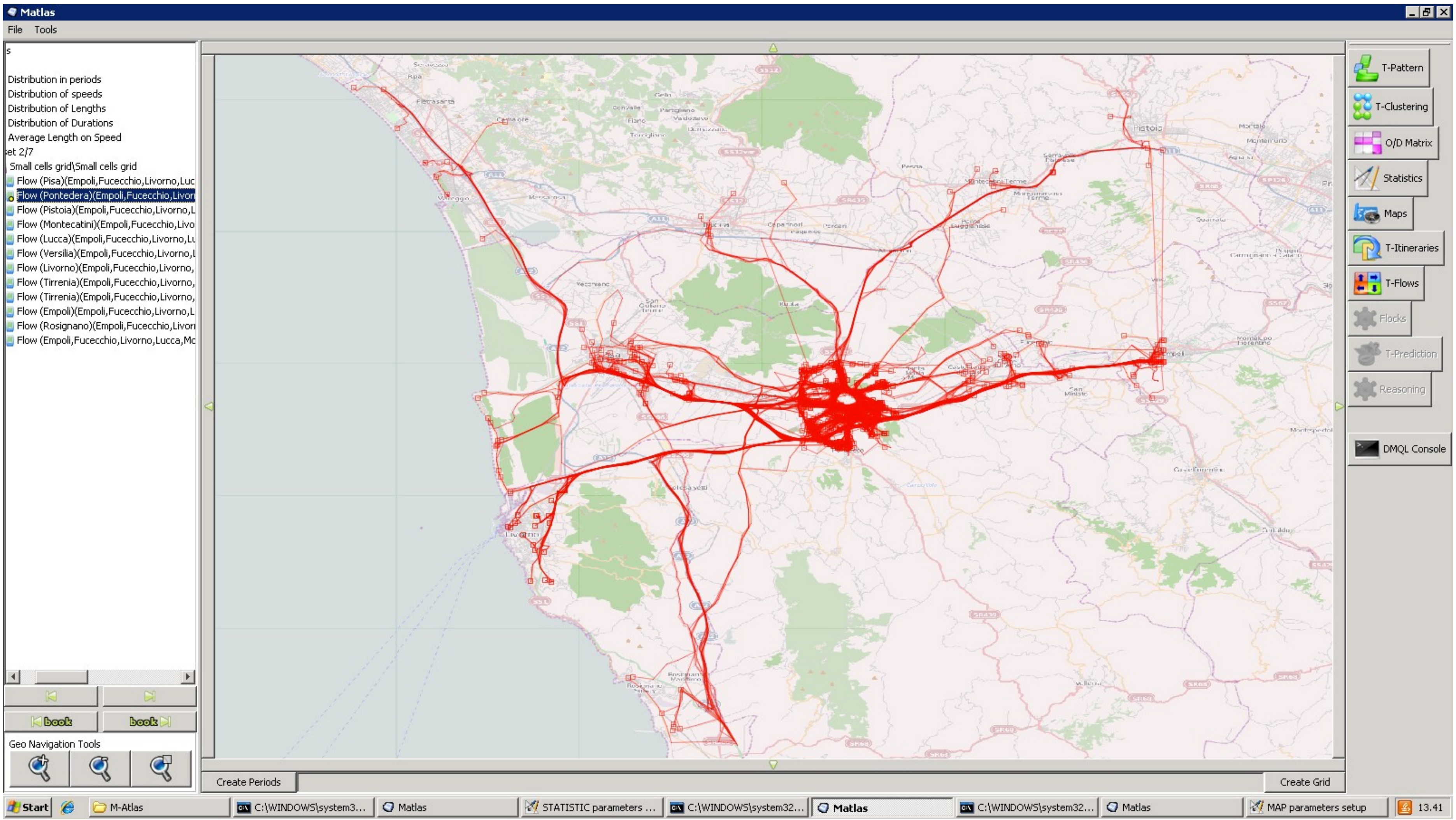
Flow from Pisa



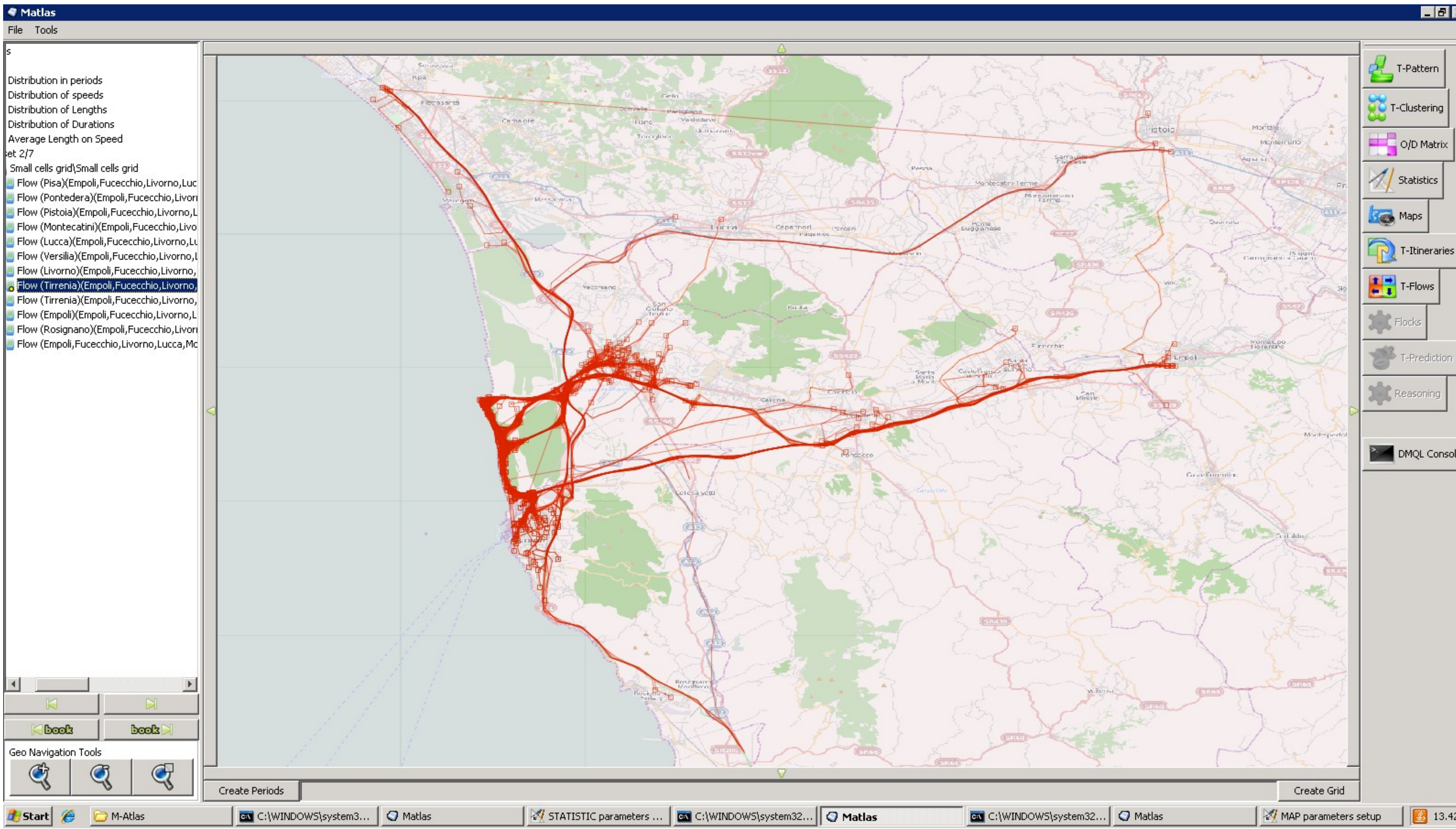
Flow From Pistoia



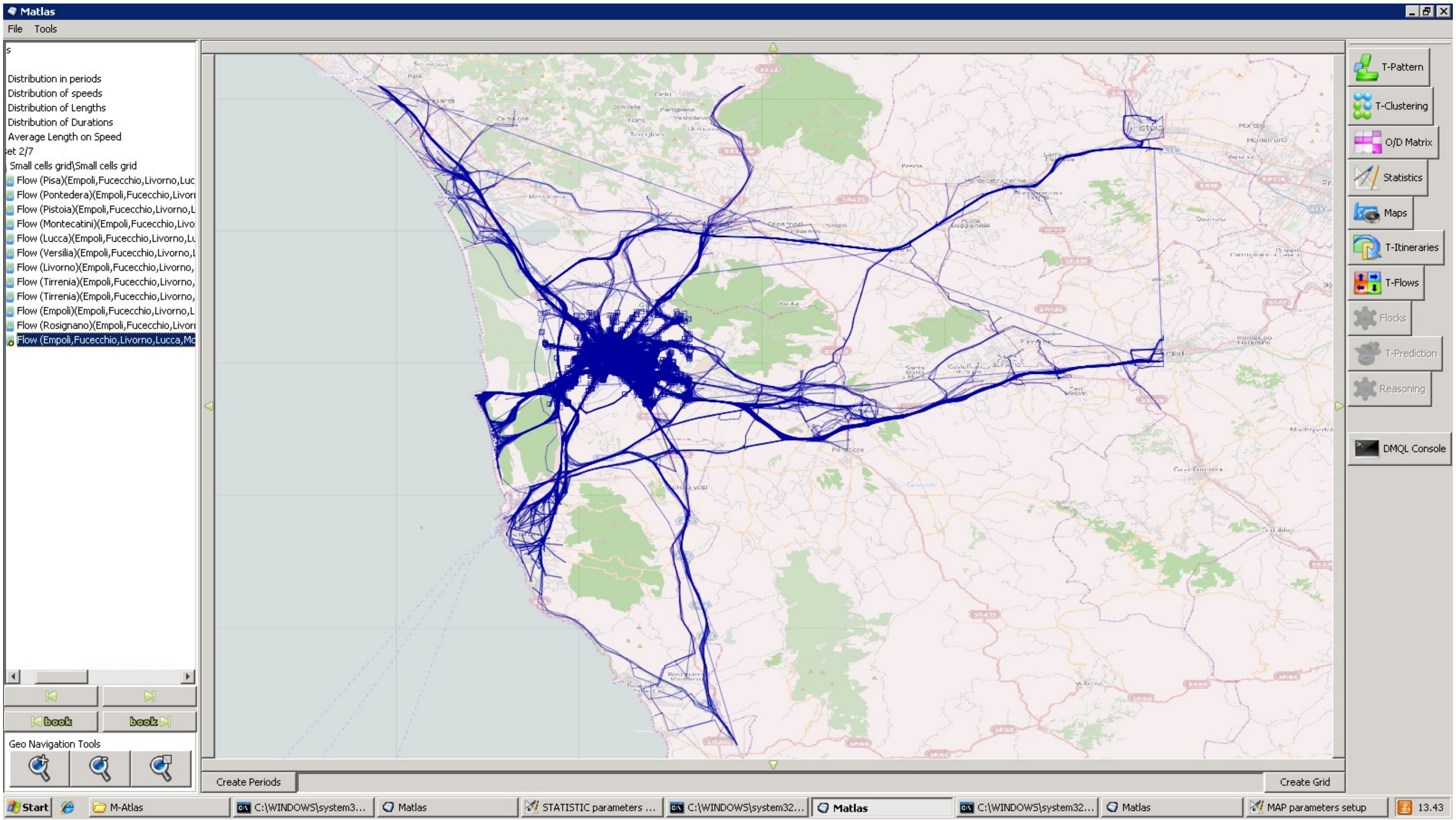
Flow From Pontedera



Flow From Tirrenia



Flow To Pisa


Matlas
File Tools
Distribution in periods
Distribution of speeds
Distribution of Lengths
Distribution of Durations
Average Length on Speed
Small cells grid\Small cells grid
Flow (Pisa)(Empoli,Fucecchio,Livorno,Luc
Flow (Pontedera)(Empoli,Fucecchio,Livorn
Flow (Pistoia)(Empoli,Fucecchio,Livorno,L
Flow (Montecatini)(Empoli,Fucecchio,Livo
Flow (Lucca)(Empoli,Fucecchio,Livorno,Lu
Flow (Versilia)(Empoli,Fucecchio,Livorno,L
Flow (Livorno)(Empoli,Fucecchio,Livorno,
Flow (Tirrenia)(Empoli,Fucecchio,Livorno,
Flow (Tirrenia)(Empoli,Fucecchio,Livorno,
Flow (Empoli)(Empoli,Fucecchio,Livorno,L
Flow (Rosignano)(Empoli,Fucecchio,Livorn
Flow (Empoli,Fucecchio,Livorno,Lucca,Mc
Geo Navigation Tools
T-Pattern
T-Clustering
O/D Matrix
Statistics
Maps
T-Itineraries
T-Flows
Flocks
T-Prediction
Reasoning
DMQL Console
Create Periods
Create Grid
Start
M-Atlas
Matlas
STATISTIC parameters ...
MAP parameters setup
13.43