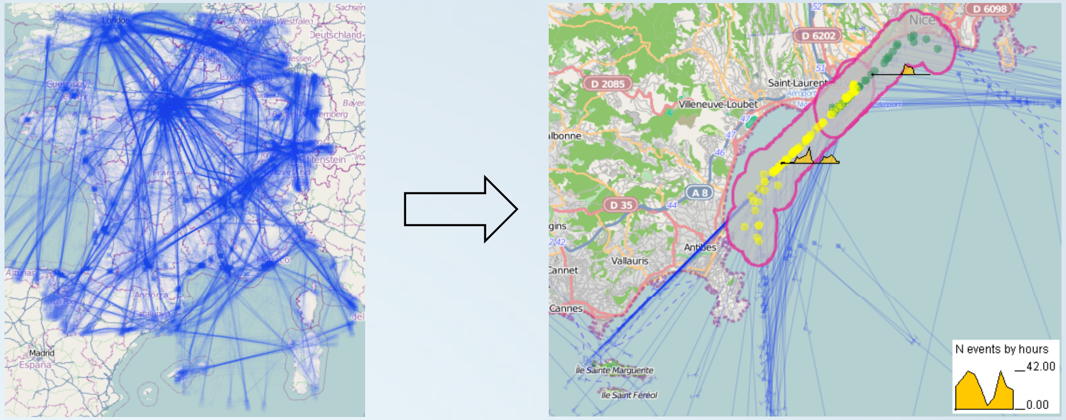
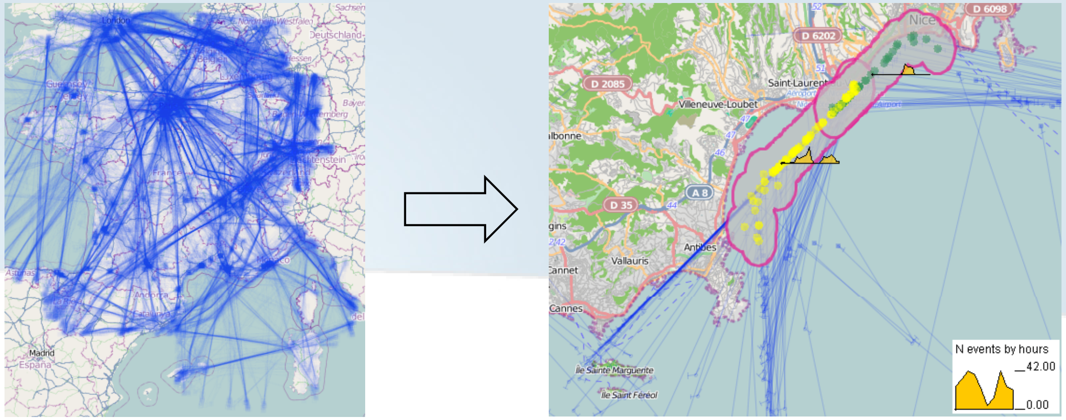


From Movement Tracks through Events to Places: Extracting and Characterizing Significant Places from Mobility Data

Gennady Andrienko, Natalia Andrienko,
Christophe Hurter, Salvatore Rinzivillo, Stefan Wrobel

http://geoanalytics.net/and

Why Visual Analytics



- Places are to be identified from data;
  - Unknown size and shape; places may overlap
- Huge amounts of imprecise and incomplete complex data
  - Moving objects trajectories of variable duration with varying sampling rate
- Human intelligence is needed to control and guide the analysis process
  - Scalable computations
  - Intelligent visualizations

General Analytical Procedure

- Extract relevant movement events
- Find significant places
  - Find dense spatial clusters of events, taking into account time and attributes
  - Surround clusters with spatial buffers
- Aggregate events by significant places
  - Events, trajectories → places + time series of attribute values
  - Trajectories → flows between places
    - Flow = vector <place 1, place 2> + time series of attribute values
- Analyze spatio-temporal aggregates

Examples of movement events (m-events)

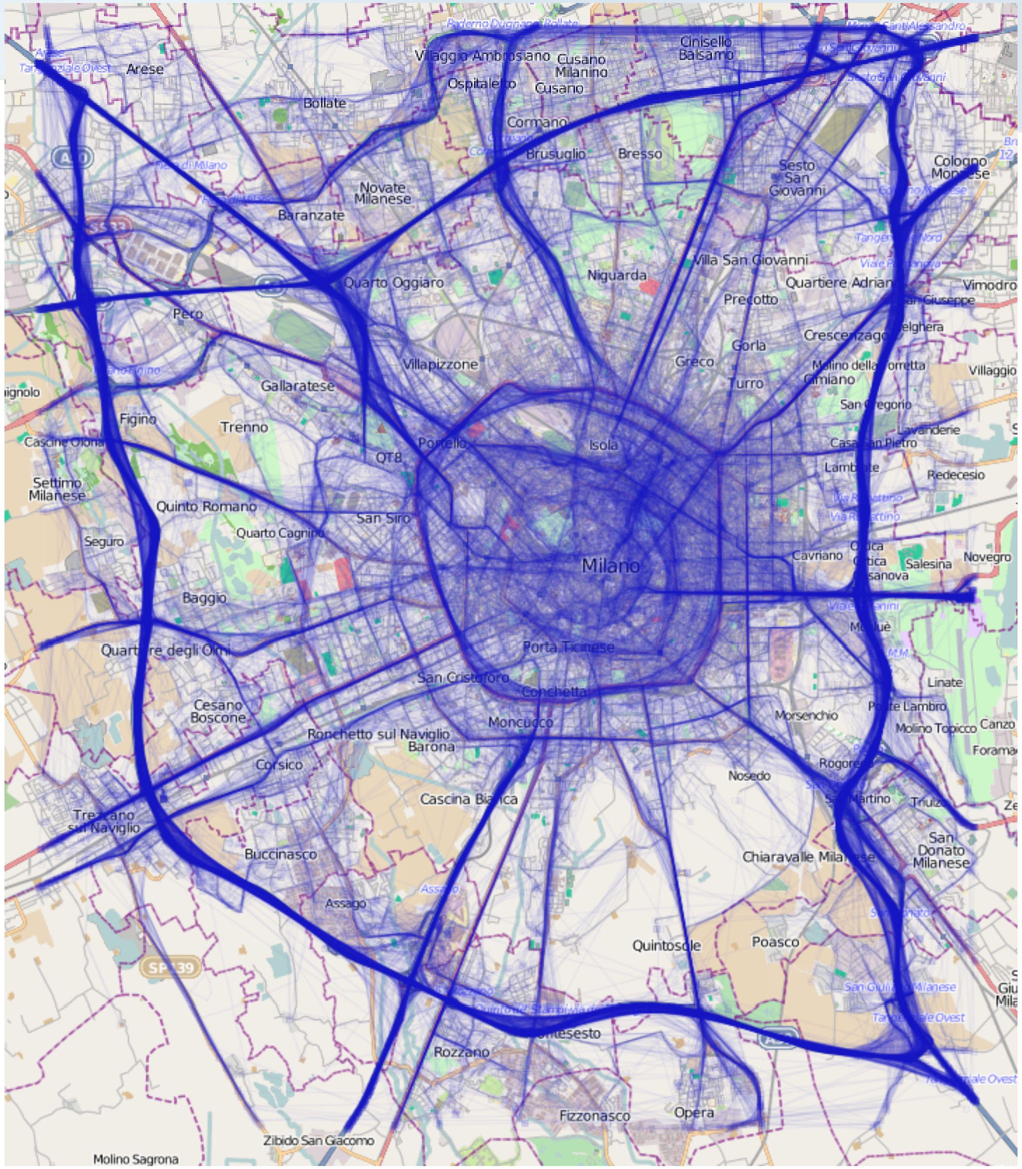
- Stop or low-speed driving
- Turn
- High acceleration
- Take-off / landing of an aircraft
- Meeting of two or more moving objects
- Driving late at night
- Stop at a place of interest
- Leaving stadium after a football game
- High heart rate {during jogging}

m-events are defined based on attributes

- Instant speed, travelled path in time window / from the beginning of the trip
- Heart rate, body temperature…
- Time of day, day of week of trajectory points
- Relationship to places, spatial objects, and events measured as
  - Spatial distance to n^{th} nearest place/object
  - Temporal distance to n^{th} nearest event
  - Neighborhood (counts of objects or events in given S,T,ST windows)

Example 1: detection and analysis of places of traffic congestions in Milan

- 8,206 GPS-tracks of cars in Milan, Italy; 235,448 points
- Collected during one day: Wednesday, the 4th of April, 2007
- Received from Comune di Milano (Municipality of Milan)



The trajectories are drawn on a map with 5% opacity

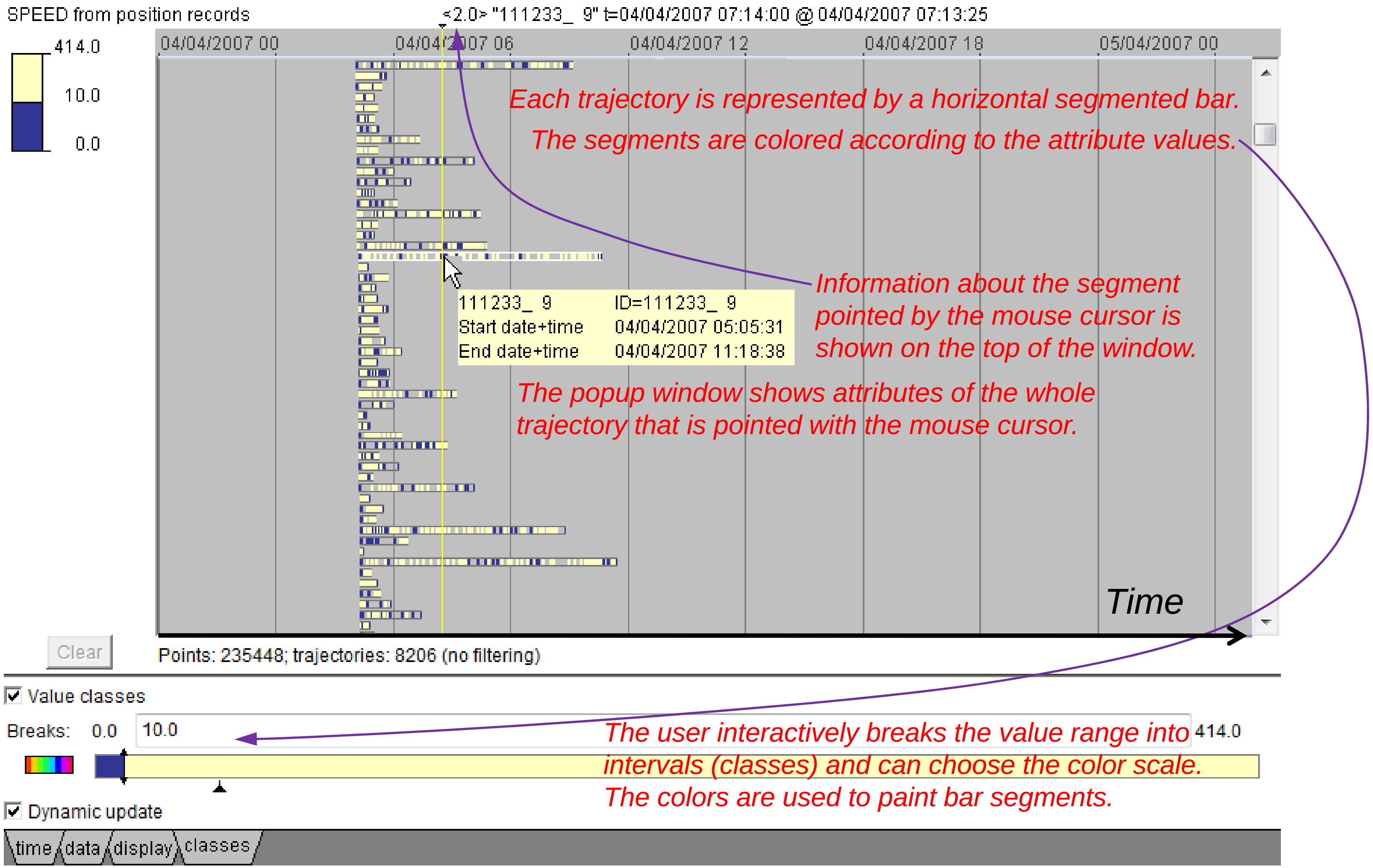
VAST 2011

Step 1: extraction of relevant events

1 2 3 4

- Relevant events = low speed events (e.g. speed <= 10 km/h)

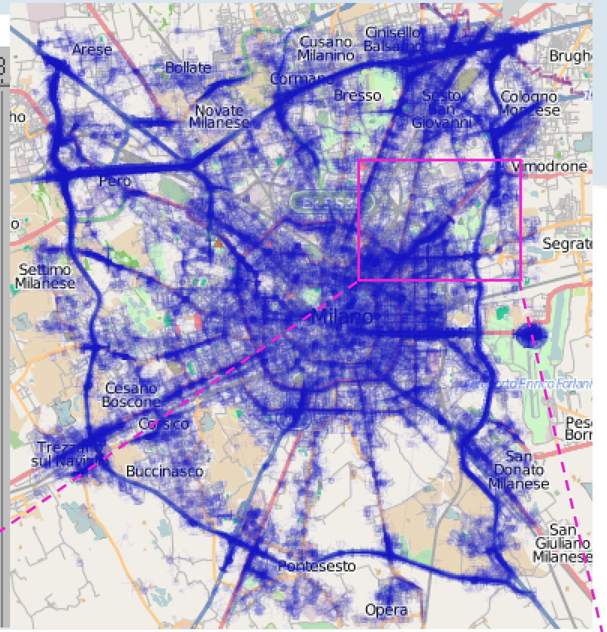
Extracting m-events from trajectories: interactive operations

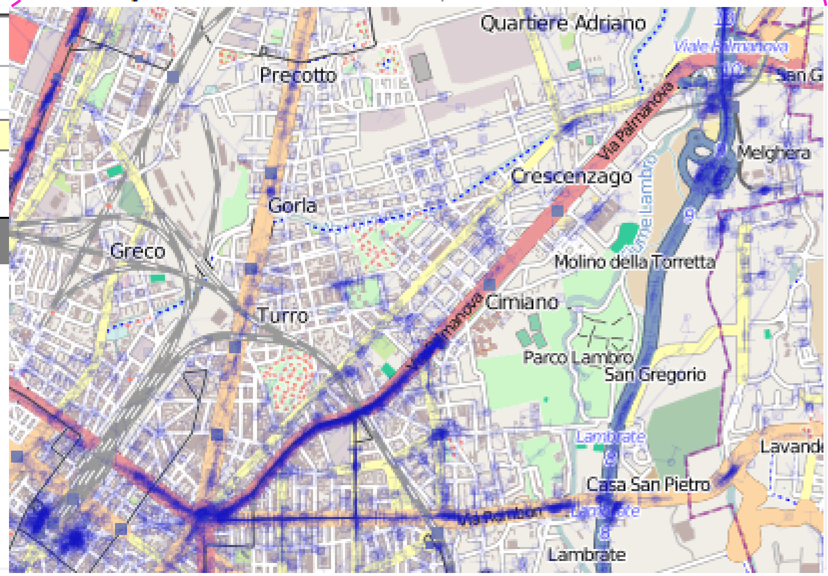


Extracting m-events from trajectories: interactive operations

The user clicks on a rectangle in the legend to switch off the respective interval or to switch it on again.

Only the bar segments representing values from the currently active interval(s) are shown.

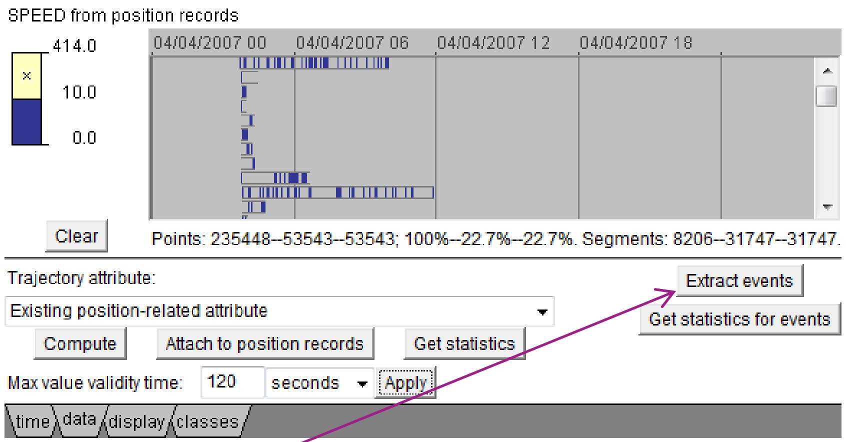




The map shows only the points and segments of the trajectories where the values of the dynamic attribute satisfy the filter.

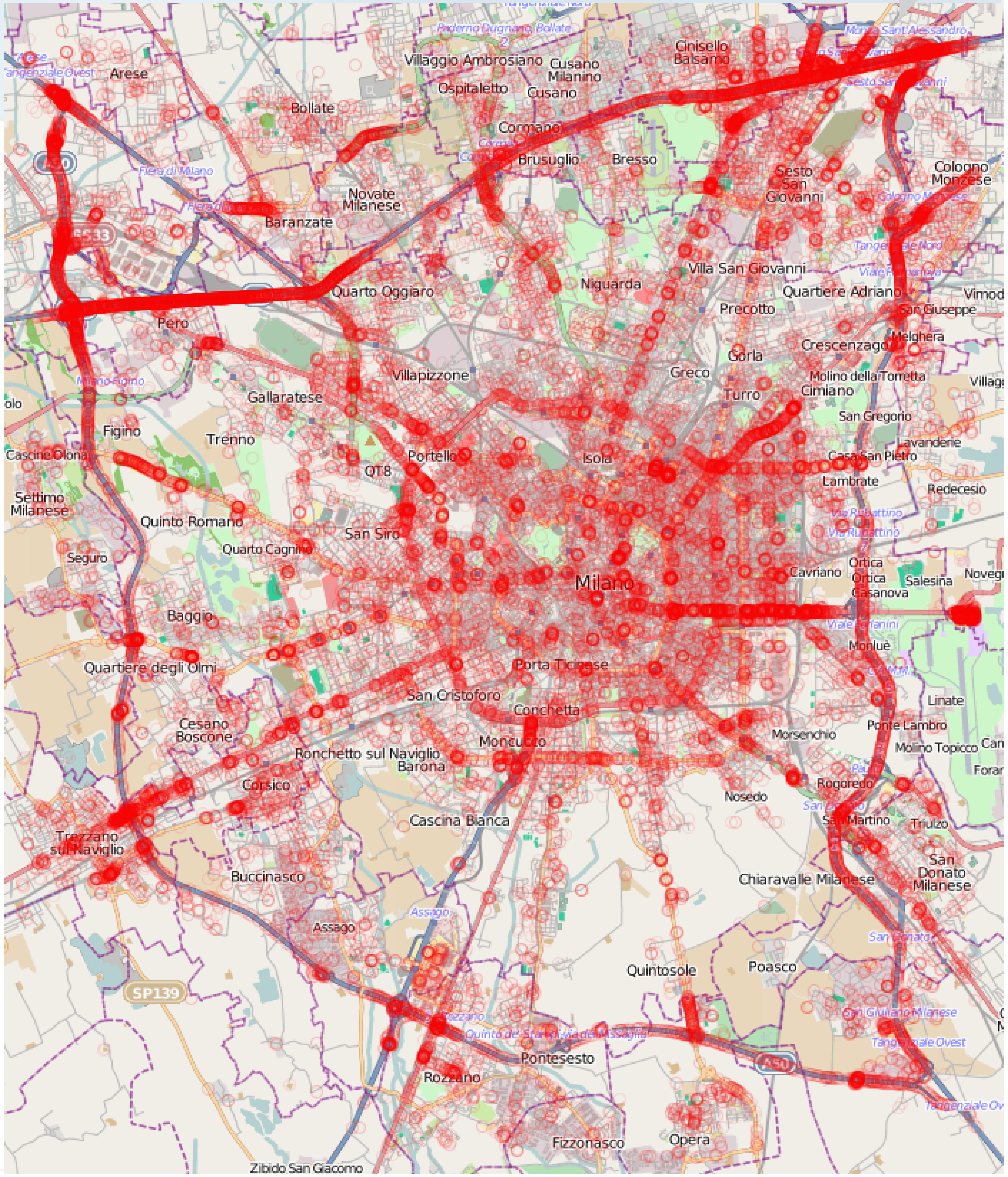
Here we see the points and segments where the speed was not more than 10 km/h.

Extraction of relevant movement events



This button in the "data" tab allows the user to extract m-events from the trajectories according to the current segment filter. The extracted events are organized in a new dataset consisting of points and multi-points with time references and attributes.

The map shows the extracted low speed events as an independent map layer. The m-events are represented by red hollow circles.



VAST 2011

Step 2: determination of the relevant places

1 2 3 4

- Relevant places = areas where traffic congestions occurred

*To distinguish low speed events caused by probable traffic congestions from occasional low speed events, we first find STD-clusters (**S**pace, **T**ime, **D**irection) of low speed events:*

many events close in space and time and having similar direction ⇒ traffic congestion

Clustering

- Density based clustering of points (Optics, DBScan) with appropriate distance function (similarity measure)

$$d_t(t_1, t_2) = \begin{cases} t_2^{start} - t_1^{end} & if\, t_1^{end} < t_2^{start} \\ t_1^{start} - t_2^{end} & if\, t_1^{start} > t_2^{end} \\ 0 & otherwise \end{cases} \quad (1)$$

- time

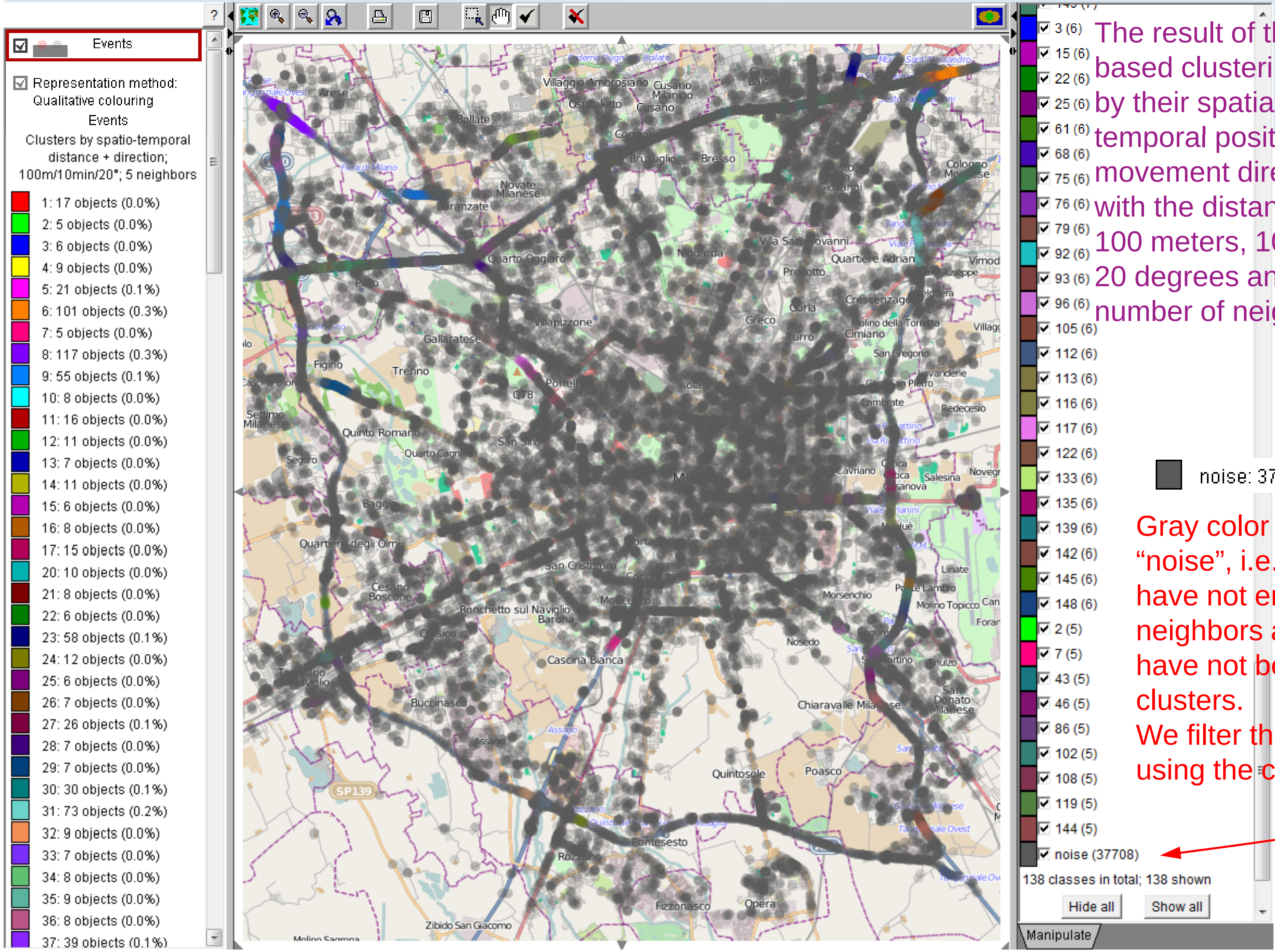
$$d(v_1, v_2, V) = \begin{cases} |v_1 - v_2|, & |v_1 - v_2| < V/2 \\ V - |v_1 - v_2|, & otherwise \end{cases} \quad (2)$$

- cyclic attrs

$$d = \begin{cases} \infty, & if\ (d_s > D_s)\ or\ \exists i\,|\,(d_i > D_i), \quad i = 0..n \\ \begin{cases} D_s * max\left(\frac{d_s}{D_s}, \frac{d_0}{D_0}, \ldots, \frac{d_n}{D_n}\right), & if\ (a) \\ D_s * \sqrt{\left(\frac{d_s}{D_s}\right)^2 + \sum_{i=0}^{n} \left(\frac{d_i}{D_i}\right)^2}, & if\ (b) \end{cases} \end{cases} \quad (3)$$

VAST 2011

STD-clustering of the m-events



The result of the density-based clustering of the events by their spatial positions, temporal positions, and movement directions (**STD**) with the distance thresholds 100 meters, 10 minutes, and 20 degrees and the minimum number of neighbors 5.

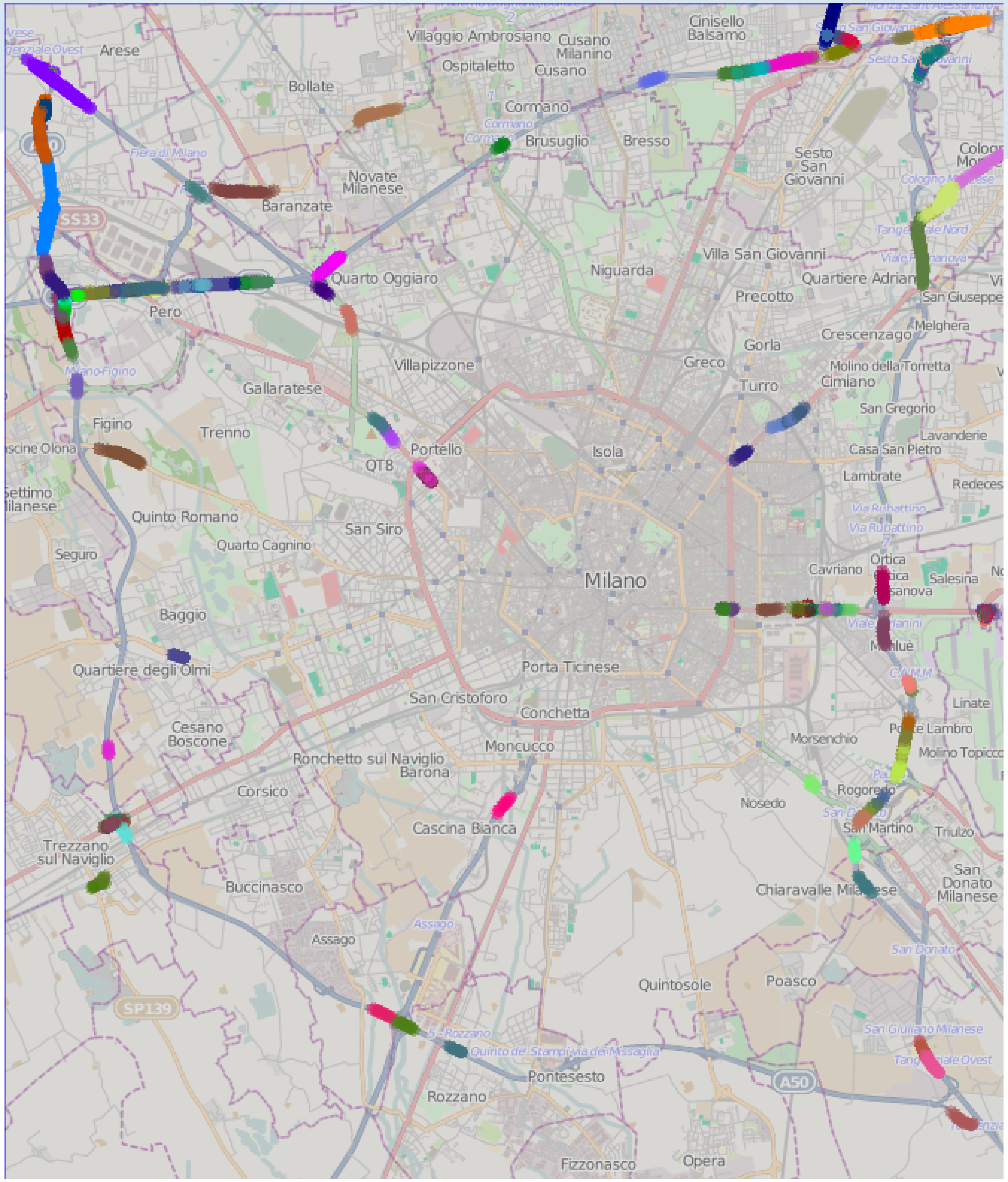
noise: 37708 objects (95.4%)

Gray color represents "noise", i.e., events that have not enough neighbors and therefore have not been put in clusters.
We filter the noise out using the checkbox.

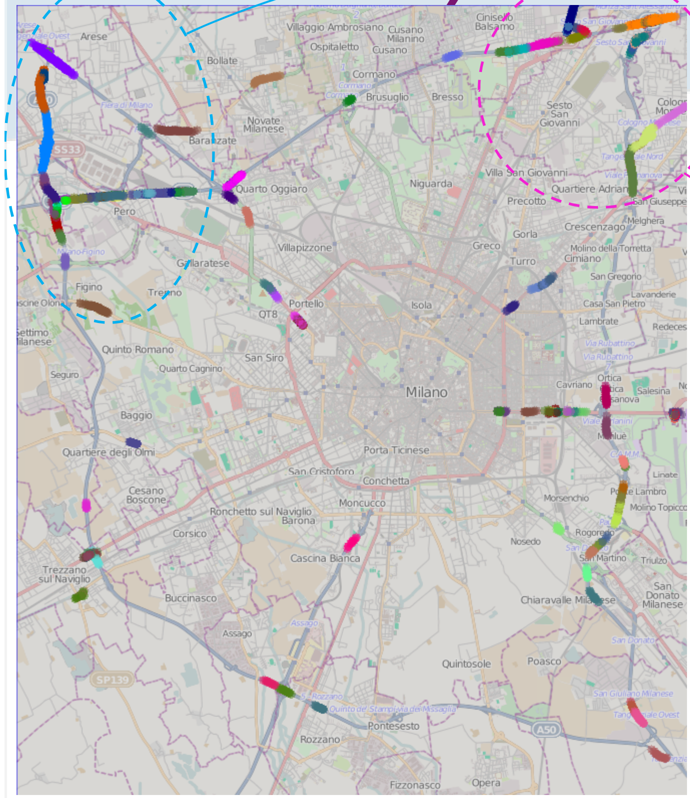
Iris, Descartes, CommonGIS, V-Analytics 1995-2010: Milan, Wed 4/4/2007, 10p5min1km

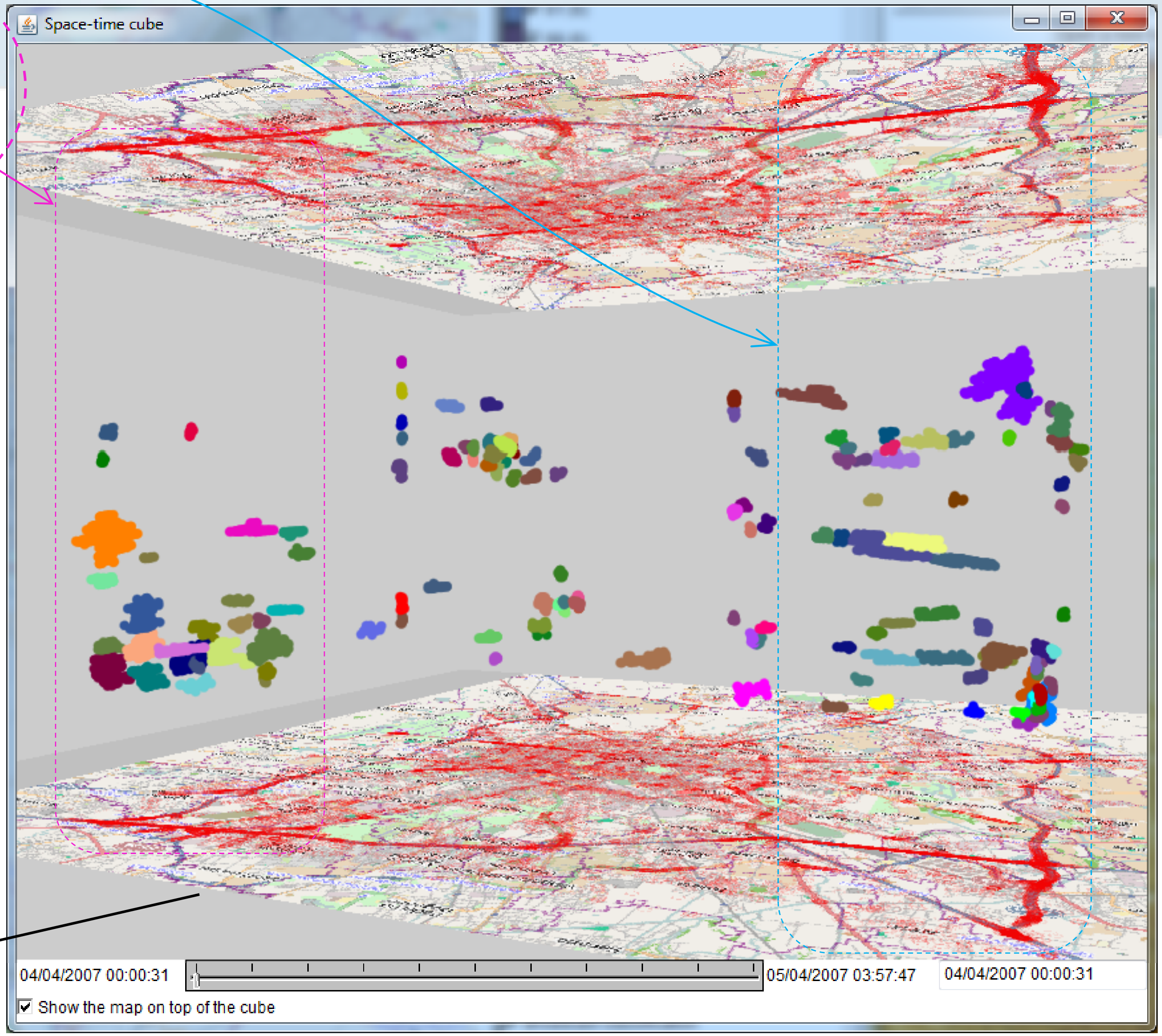
STD-clusters of m-events (noise excluded)

1 2 3 4



STD-clusters of m-events (noise excluded)



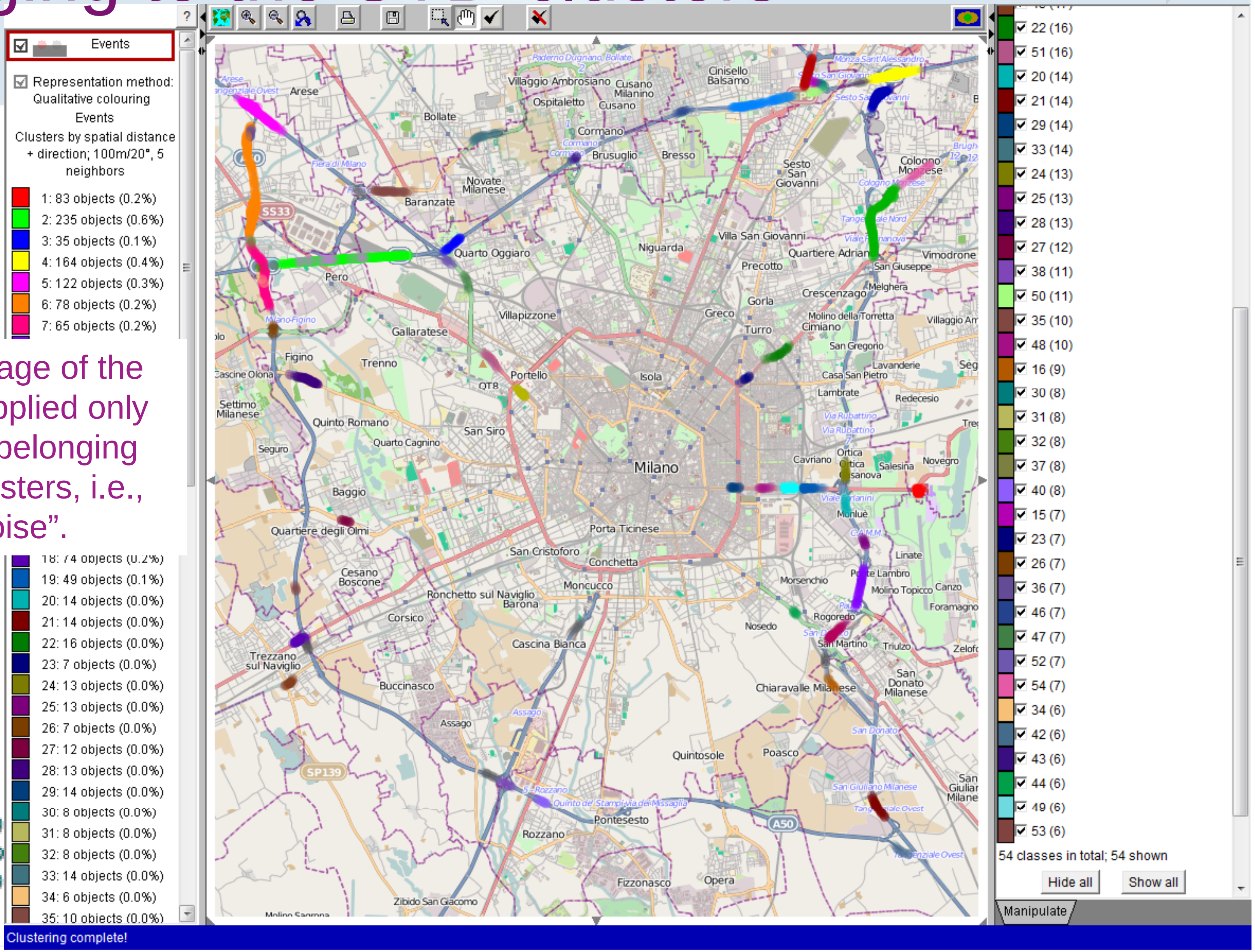


The space-time cube is viewed from the north.

North

VAST 2011

SD-clustering of the m-events belonging to the STD-clusters

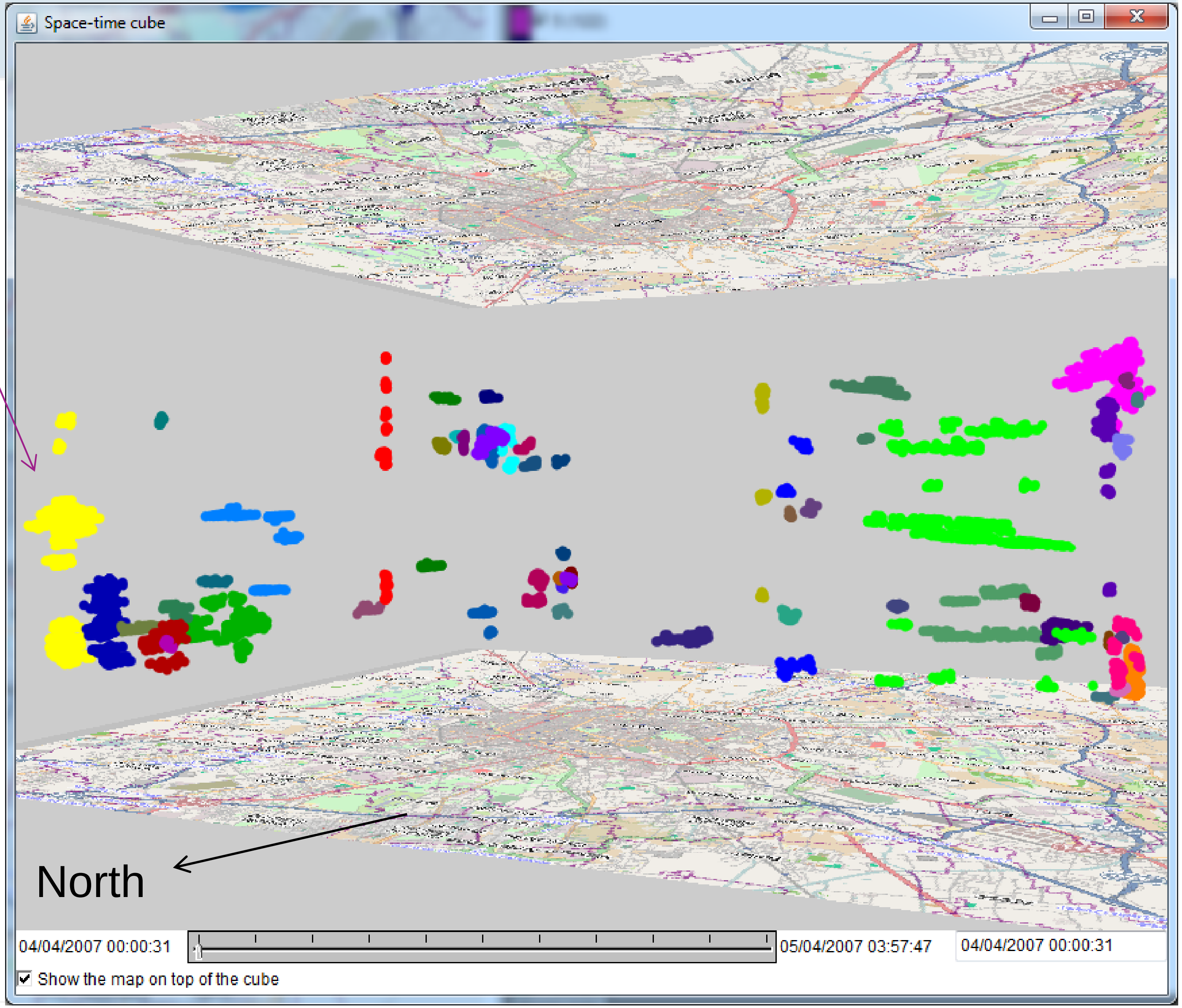


The second stage of the clustering is applied only to the objects belonging to the STD-clusters, i.e., without the "noise".

SD-clusters of the m-events

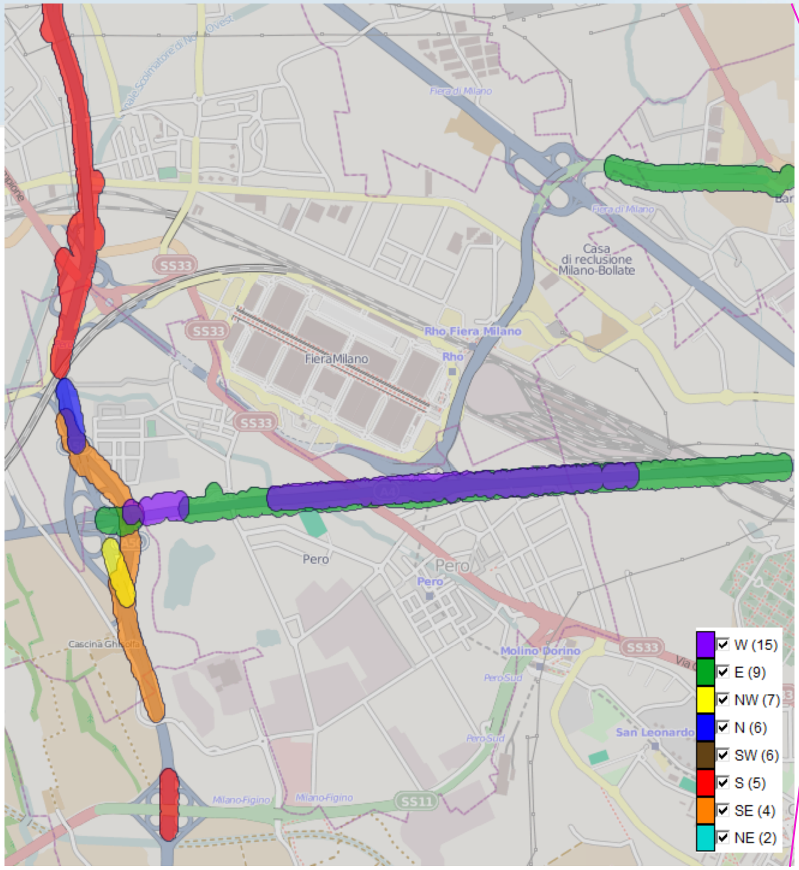
1 2 3 4

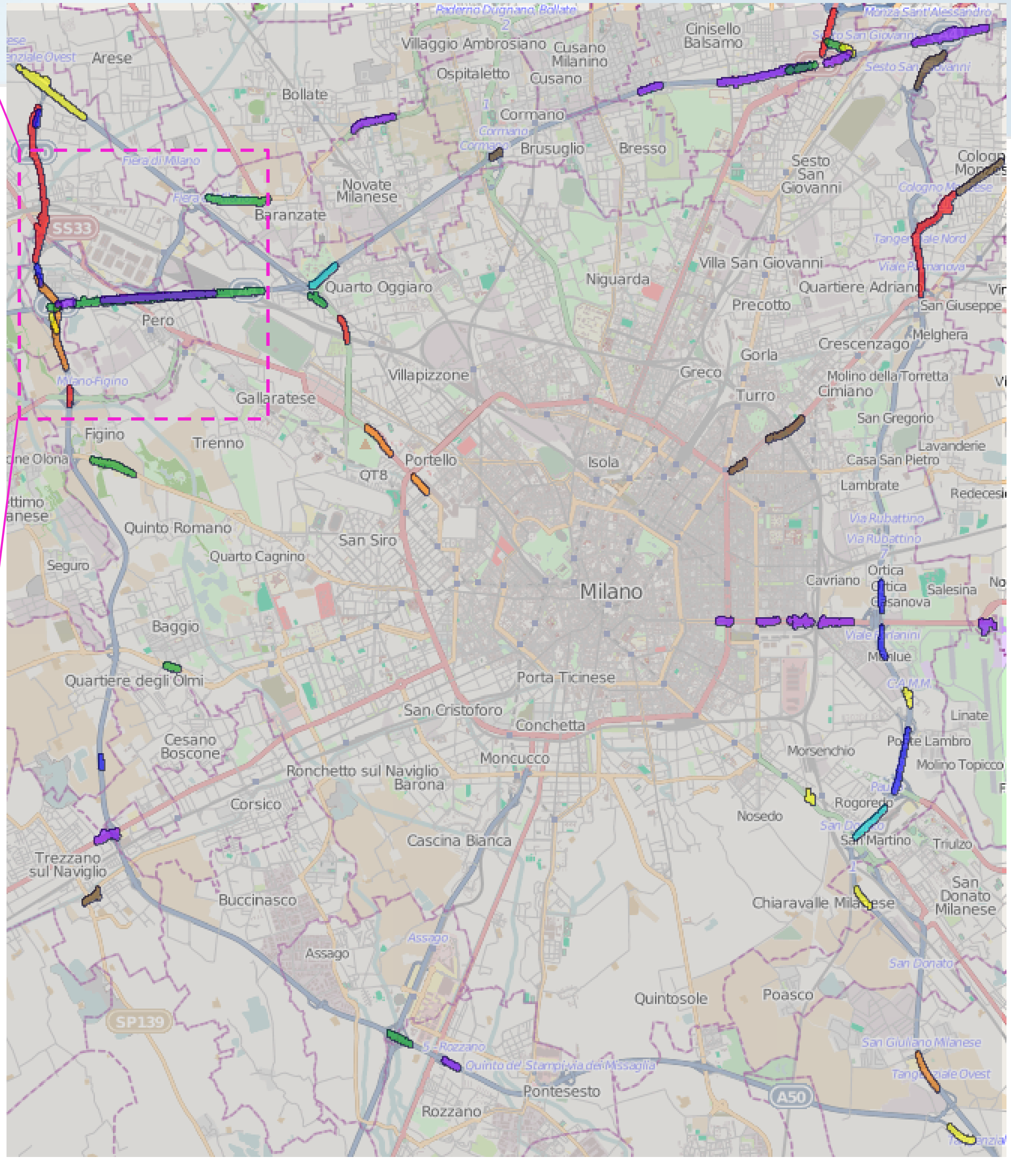
The SD-clustering has united STD-clusters that occurred in different times but overlap in space.



Spatial buffers around the clusters define the relevant places

1 2 3 4





The places are painted according to the prevailing movement directions of the respective events.

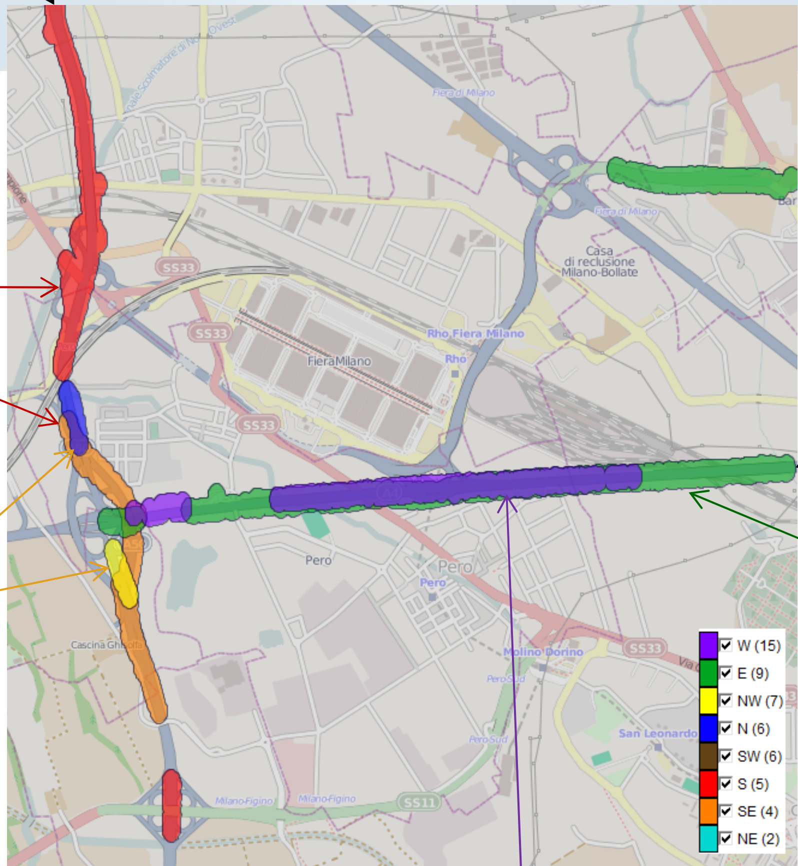
VAST 2011

1 2 3 4

Belt road north-south on the east of the city (A50)

Extended areas of congested traffic directed to the south and southeast

Smaller areas of obstructed movement directed to the north and northwest



Belt road west-east on the north of the city (A4)

Very long area of congested traffic directed to the east

Long area of congested movements directed to the west

VAST 2011

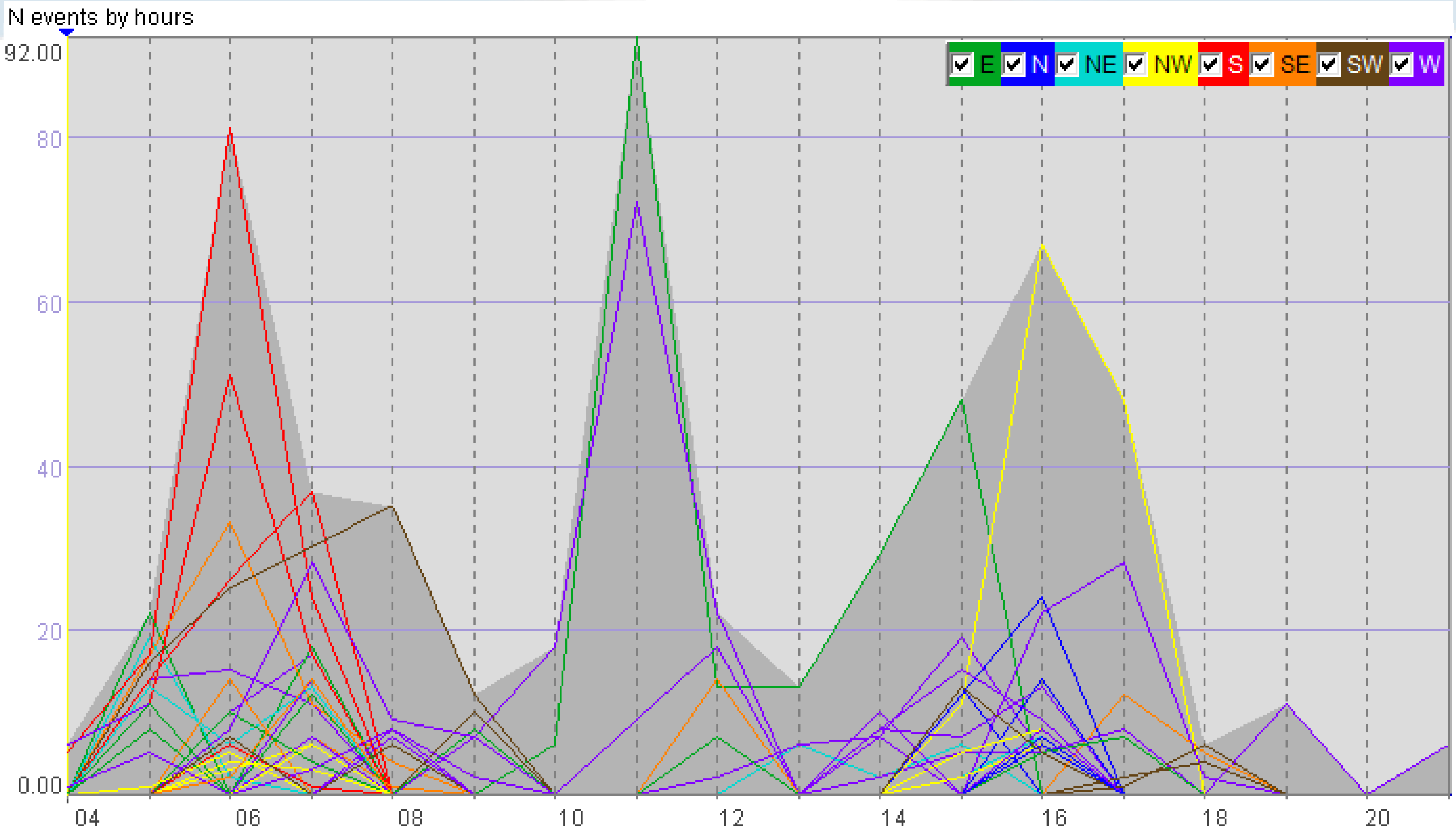
Step 3: spatio-temporal aggregation of the low speed events

- The m-events are aggregated by the areas (spatial buffers) and time intervals, e.g., hourly.
- Each area receives one or more time series of the aggregate attributes, e.g., event counts.

Step 4: exploration of the aggregated data

Time graph: counts of the low speed events by the places and hourly intervals

1 2 3 4

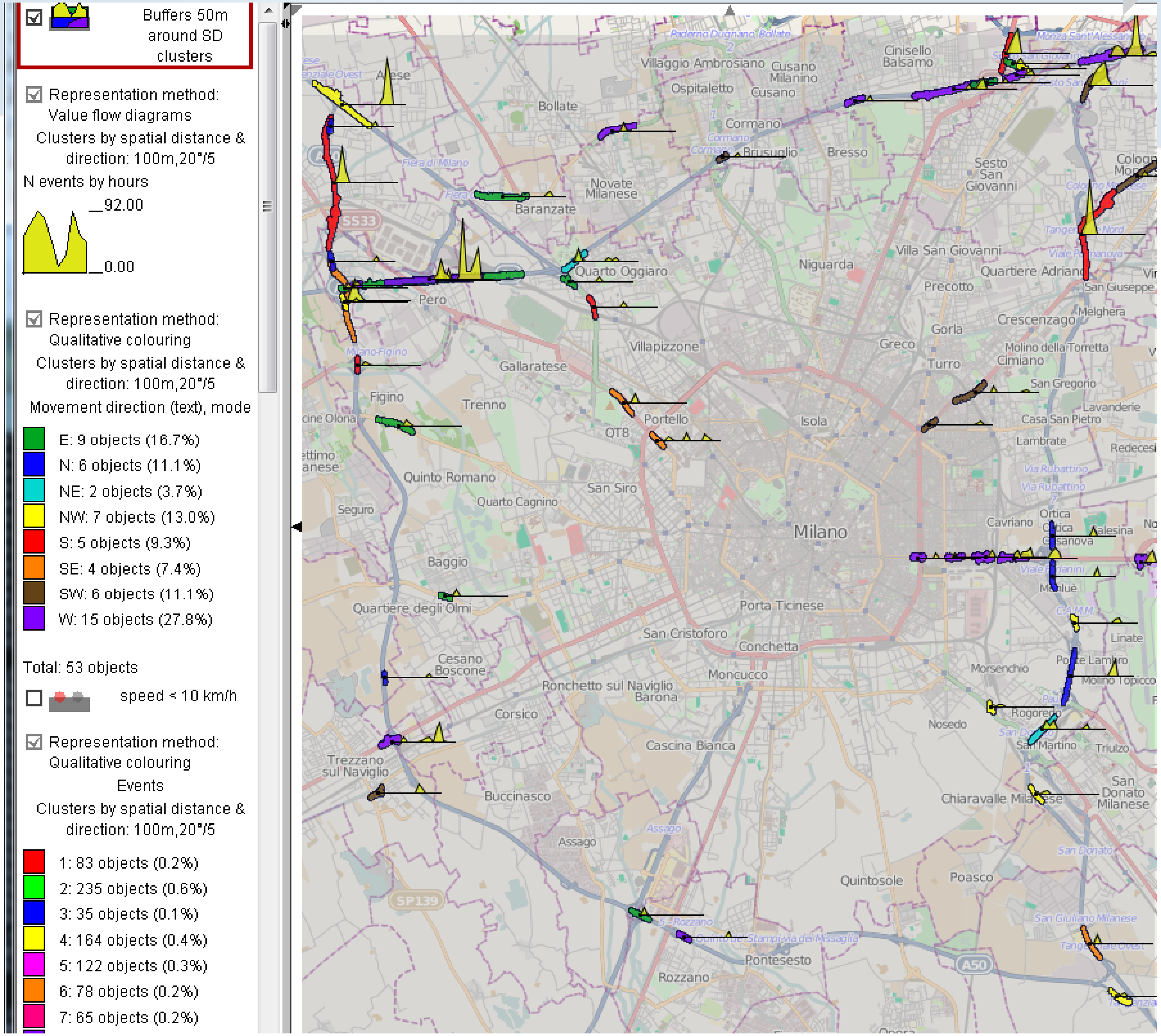


VAST 2011

The temporal variation profiles on a map

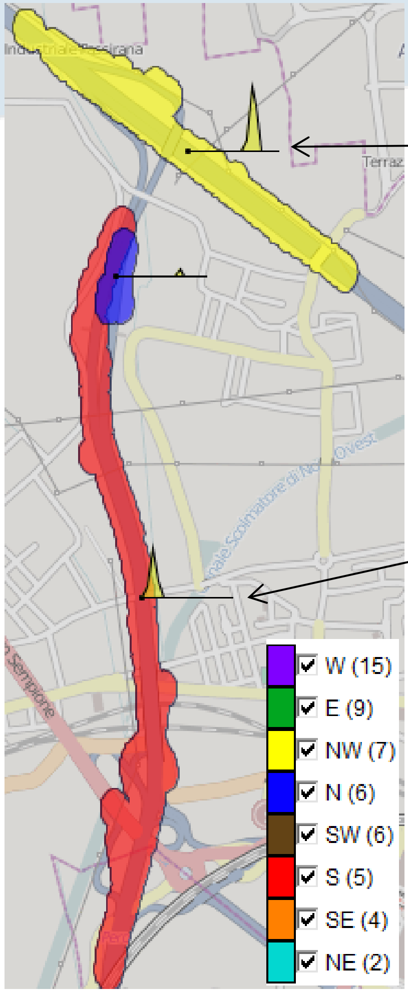
1 2 3 4

The temporal diagrams show the variation of the attribute value (vertical dimension) over time (horizontal dimension).



VAST 2011

Map fragment (northwest) enlarged



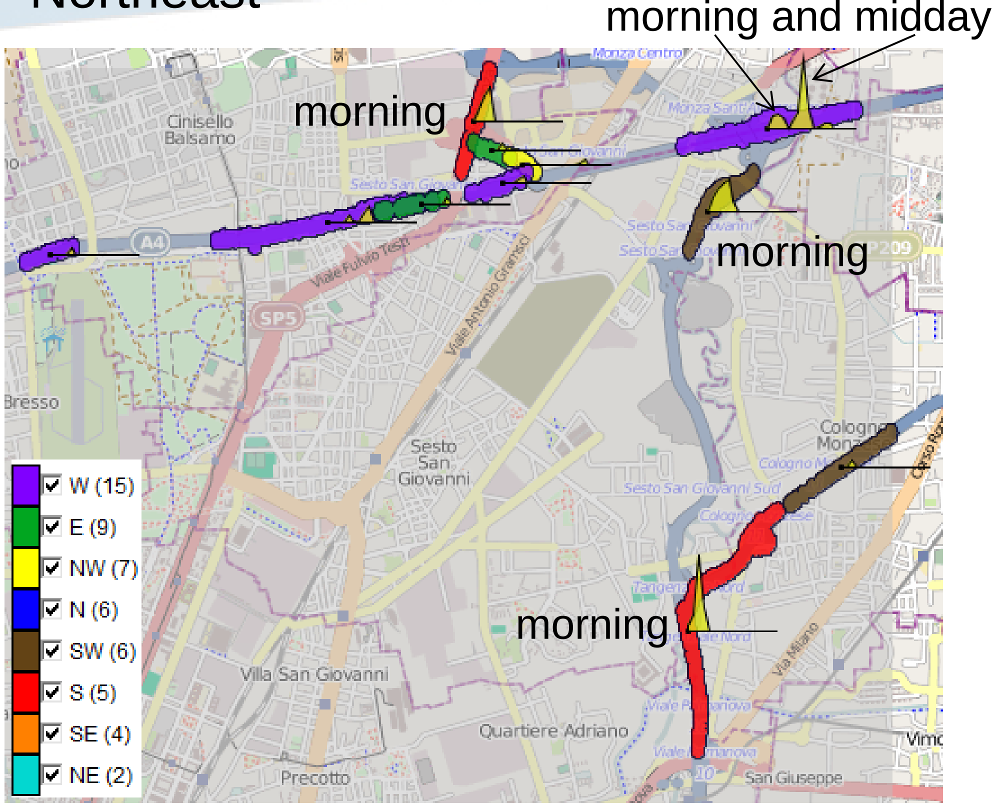
Congested traffic in the afternoon in the direction out of the city (northwest)

Congested traffic in the morning in the direction to the south

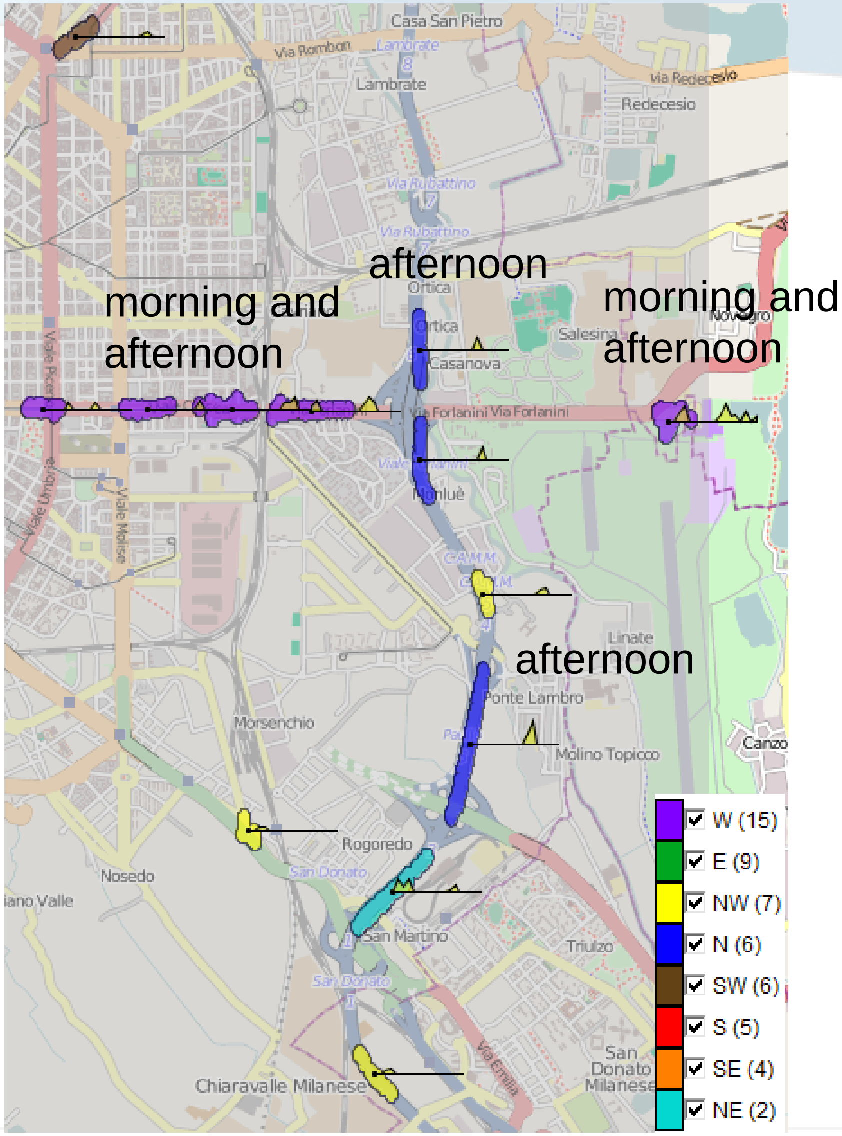
Map fragments enlarged

1 2 3 4

Northeast



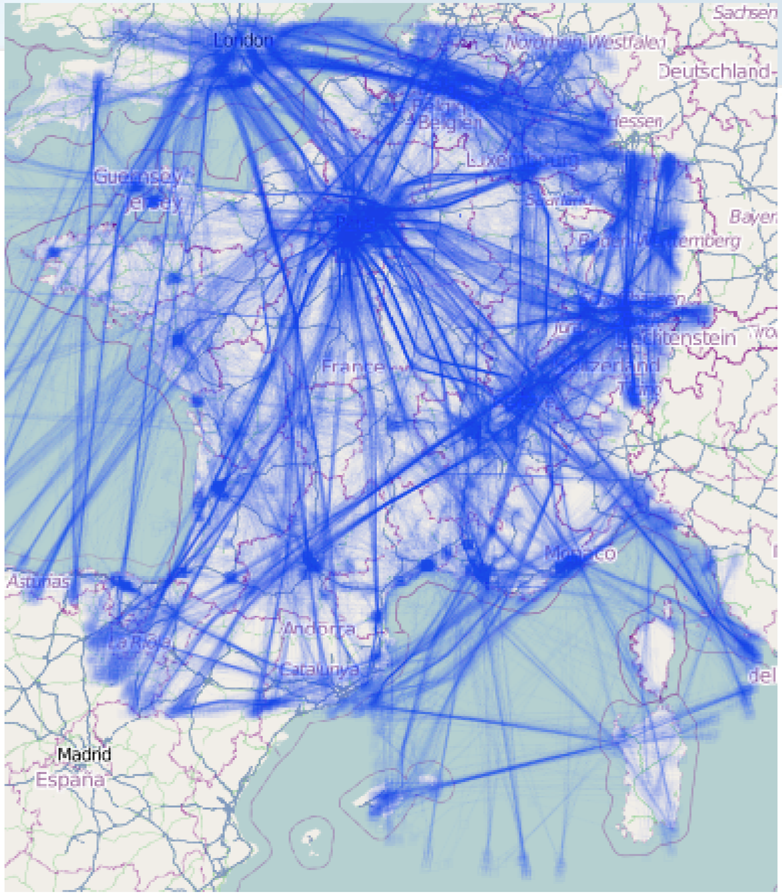
East



VAST 2011

Example 2: investigation of spatio-temporal patterns of air traffic in France

- 17,851 trajectories; 427,651 records collected by radars
- Friday, the 22nd of February, 2008
- Received from ENAC Toulouse, France



The trajectories are drawn on a map with 1% opacity

Step 1a: extraction of relevant events

1 2 3 4

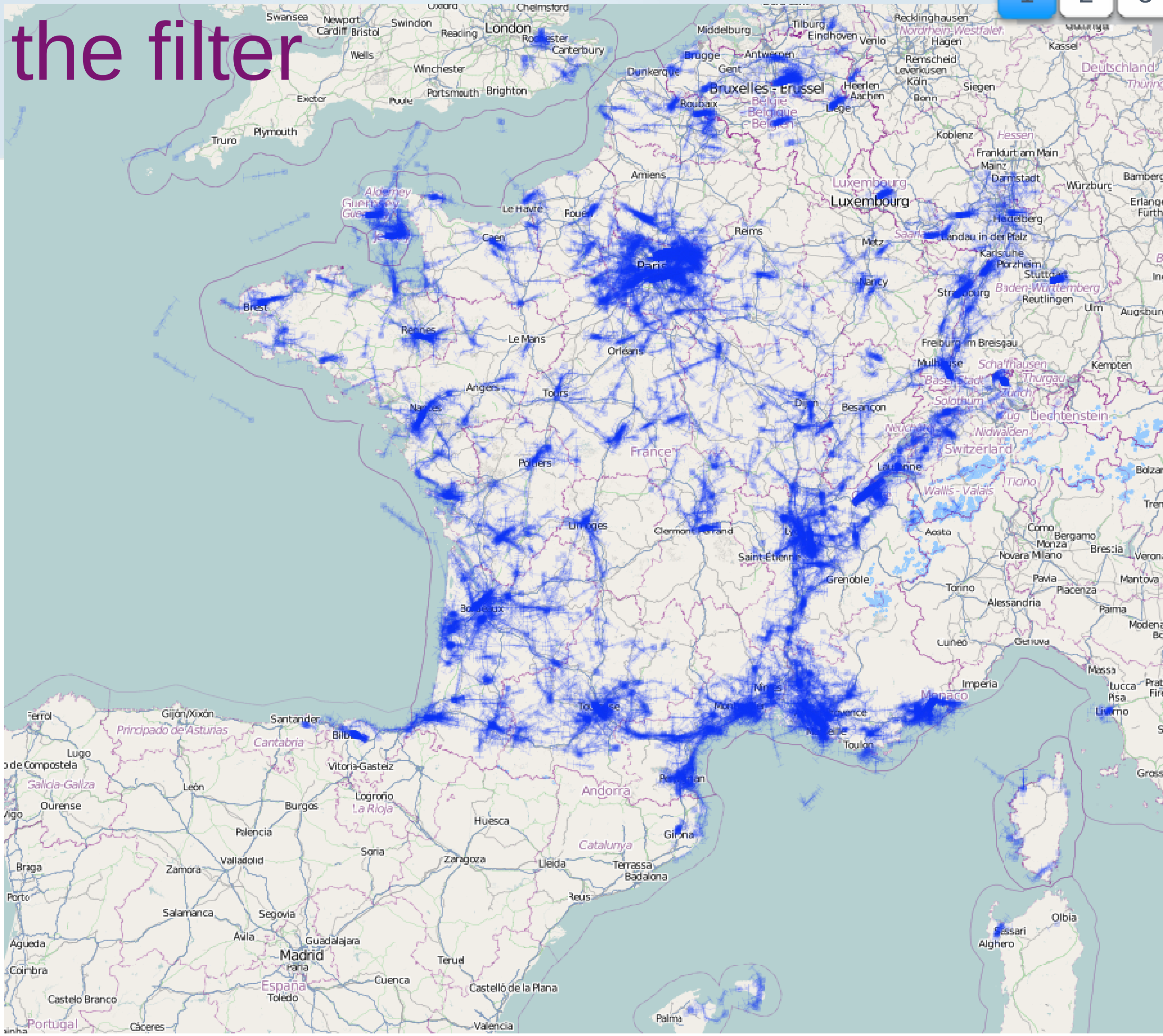
- Relevant events = landings
  - Altitude <= 1km in the last 5 minutes of a trajectory
    - In case of multiple points, take the last one

**** complex condition involving an attribute and a temporal relation*

VAST 2011

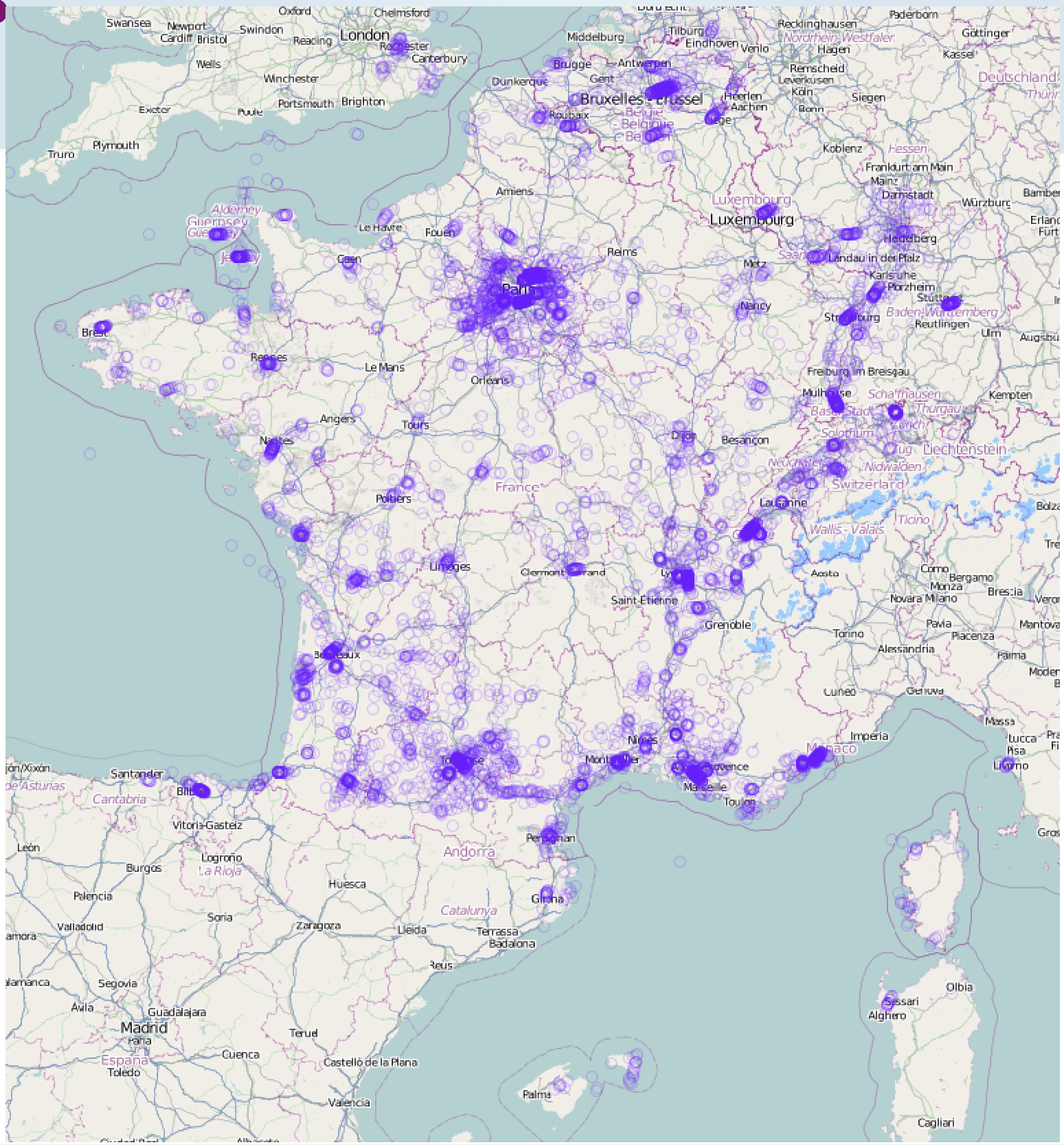
Map with trajectory segments satisfying the filter

1 2 3 4



Extracted events of probable landings

1 2 3 4



Step 2a: determination of the relevant places

1 2 3 4

- Relevant places = areas where the landings occur
  - Determined by means of SD-clustering using thresholds 1 km and 30 degrees; minimum number of neighbors 5

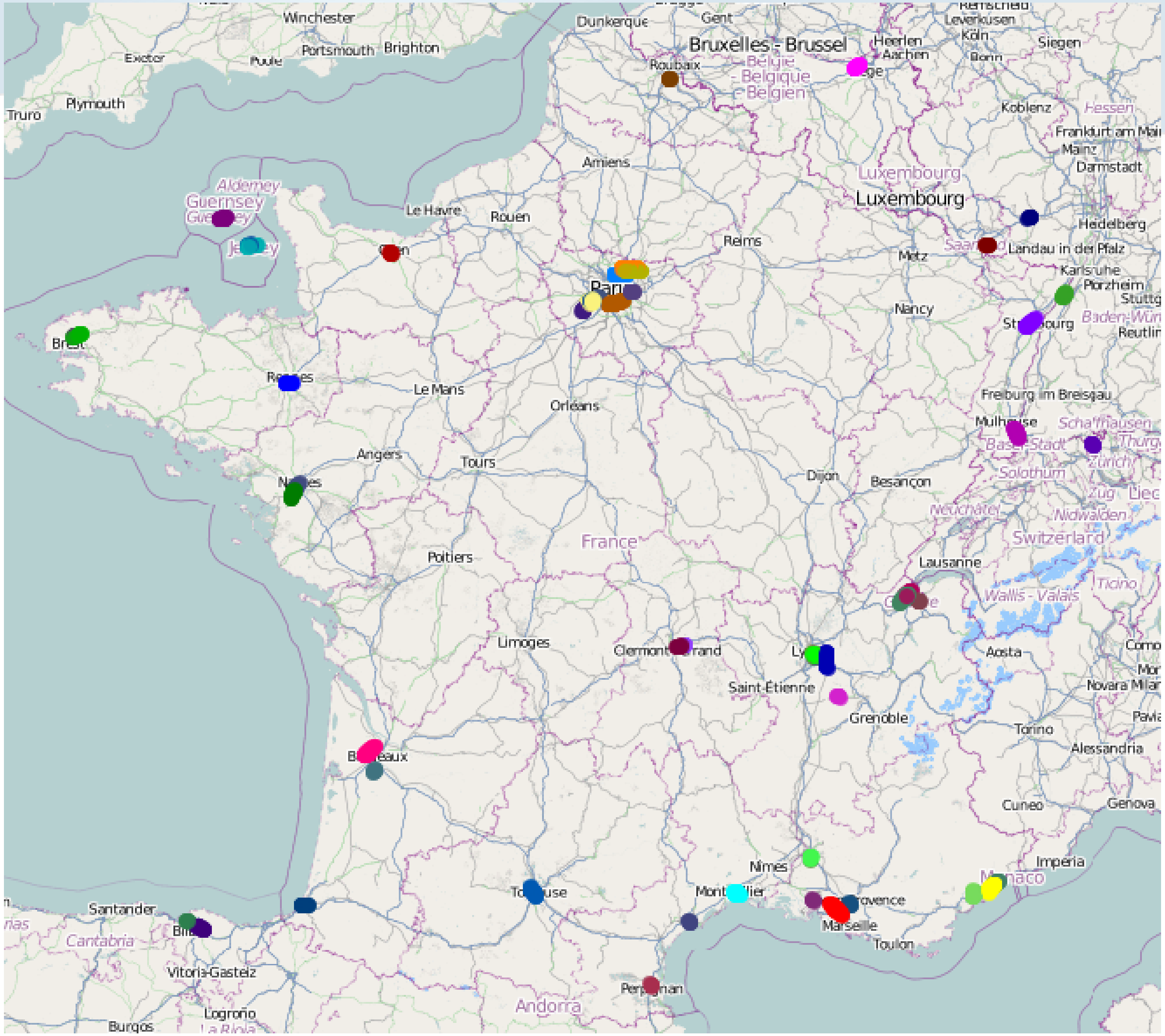
VAST 2011

SD-clusters of the landing events on a map

1 2 3 4

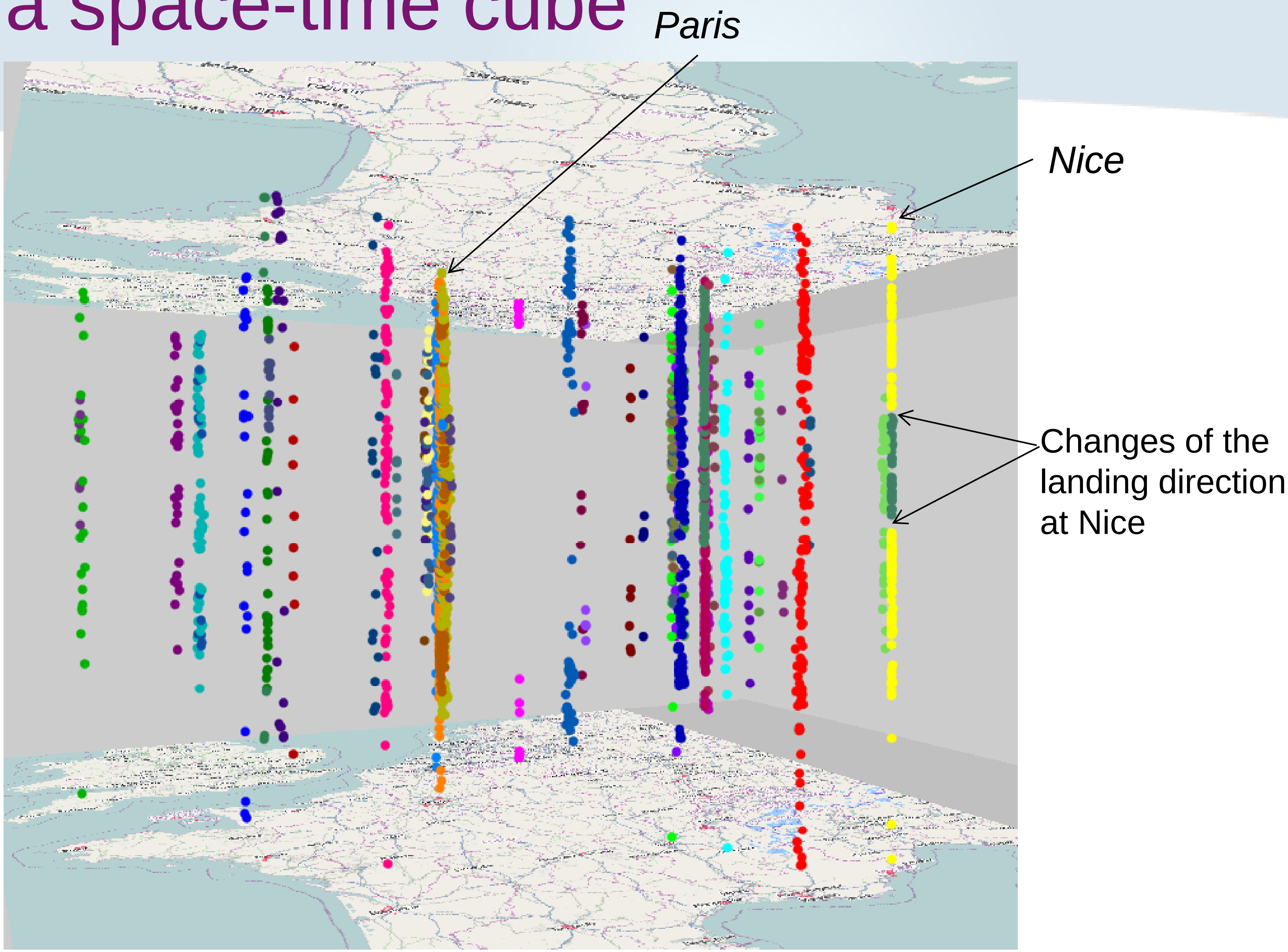
Clusters by
spatial positions (S)
and directions (D).

The noise has been
hidden.



VAST 2011

SD-clusters of the landing events in a space-time cube



Step 3a: spatio-temporal aggregation of the landing events

1 2 3 4

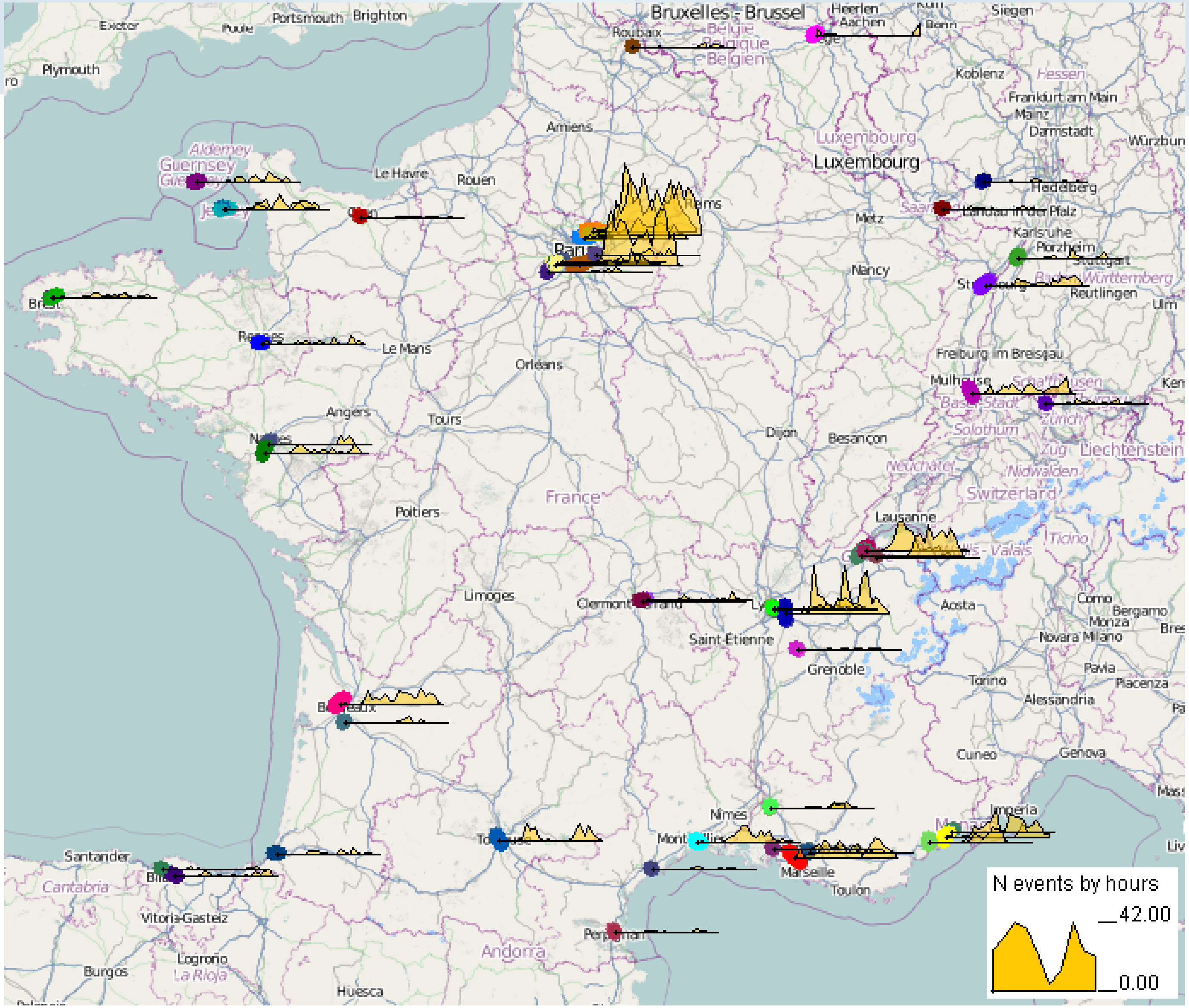
- The m-events are aggregated by the areas (spatial buffers around the SD-clusters) and hourly time intervals.
- Each area receives one or more time series of the aggregate attributes, e.g., event counts.

Step 4a: exploration of the aggregated event data

1 2 3 4

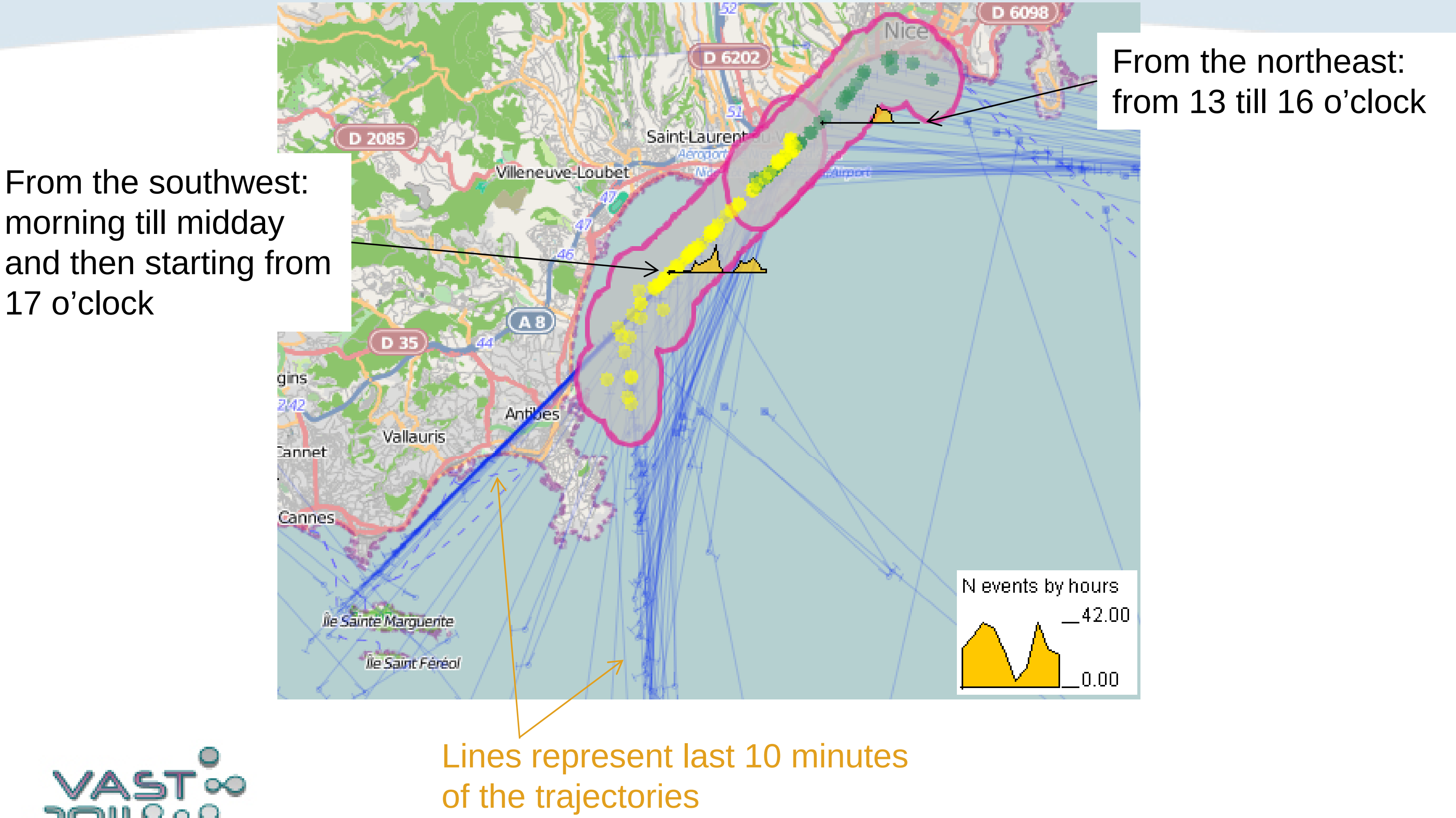
The dynamics of the landings by the places

1 2 3 4



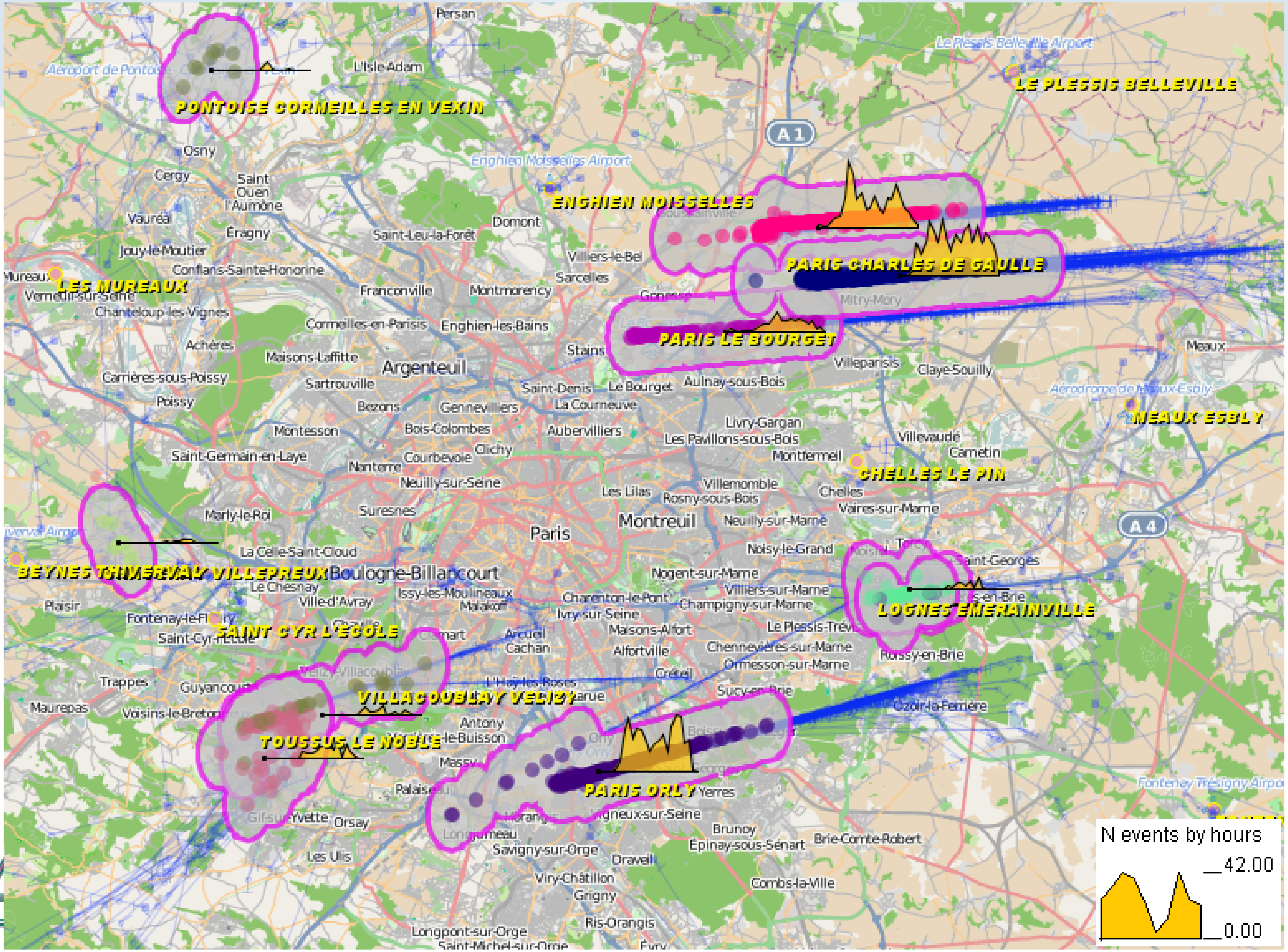
Temporal profiles for the two landing directions at Nice

1 2 3 4



Temporal profiles of the landings in Paris

1 2 3 4



Step 1b: extraction of relevant events

- Relevant events = landings (extracted previously) + takeoffs *

  * Altitude <= 1km in the first 5 minutes of a trajectory

    – In case of multiple points, take the first one

Step 2b: determination of the relevant places

- Relevant places = airport areas = areas (spatial buffers) around spatial clusters of takeoff and landing events
  - Stage 1: SD-clustering of the takeoff events using thresholds 1 km and 30 degrees; minimum number of neighbors 5
    - exclude the noise
  - Stage 2: S-clustering of the takeoff and landing events taken together (noise excluded)

Step 3b: spatio-temporal aggregation of the trajectories

1 2 3 4

- Select the trajectories having both takeoff and landing events.
- For each pair of places <place 1, place 2>, find all trajectories that start in place 1 and end in place 2.
- Compute statistics (e.g. counts) by time intervals (e.g. hourly) and whole time
- Result: vector <place 1, place 2> + time series of aggregate attributes + totals (for the whole time)

VAST 2011

Step 4b: exploration of the aggregated movement data

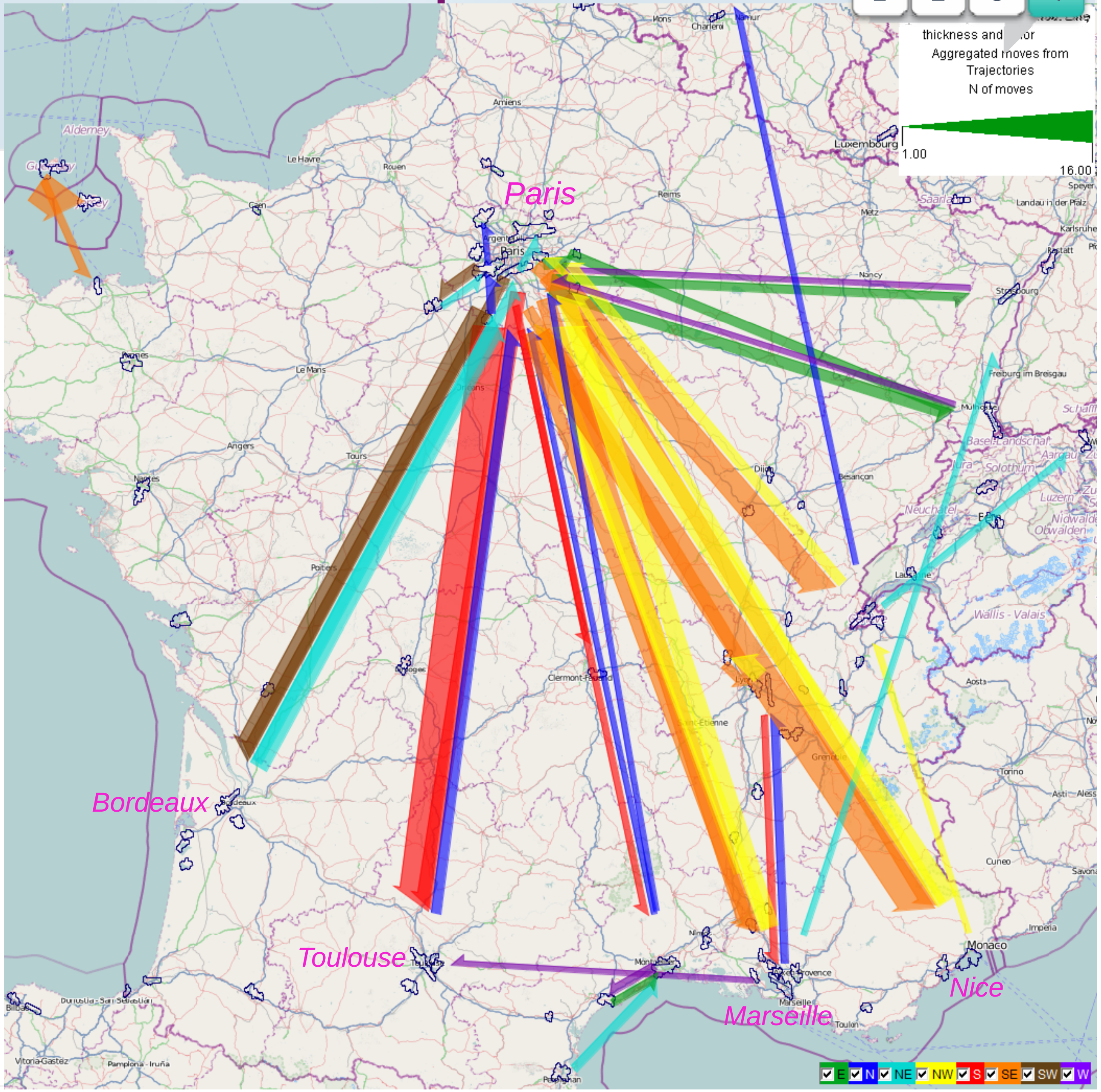
1 2 3 4

Major flows between airports

1 2 3 4

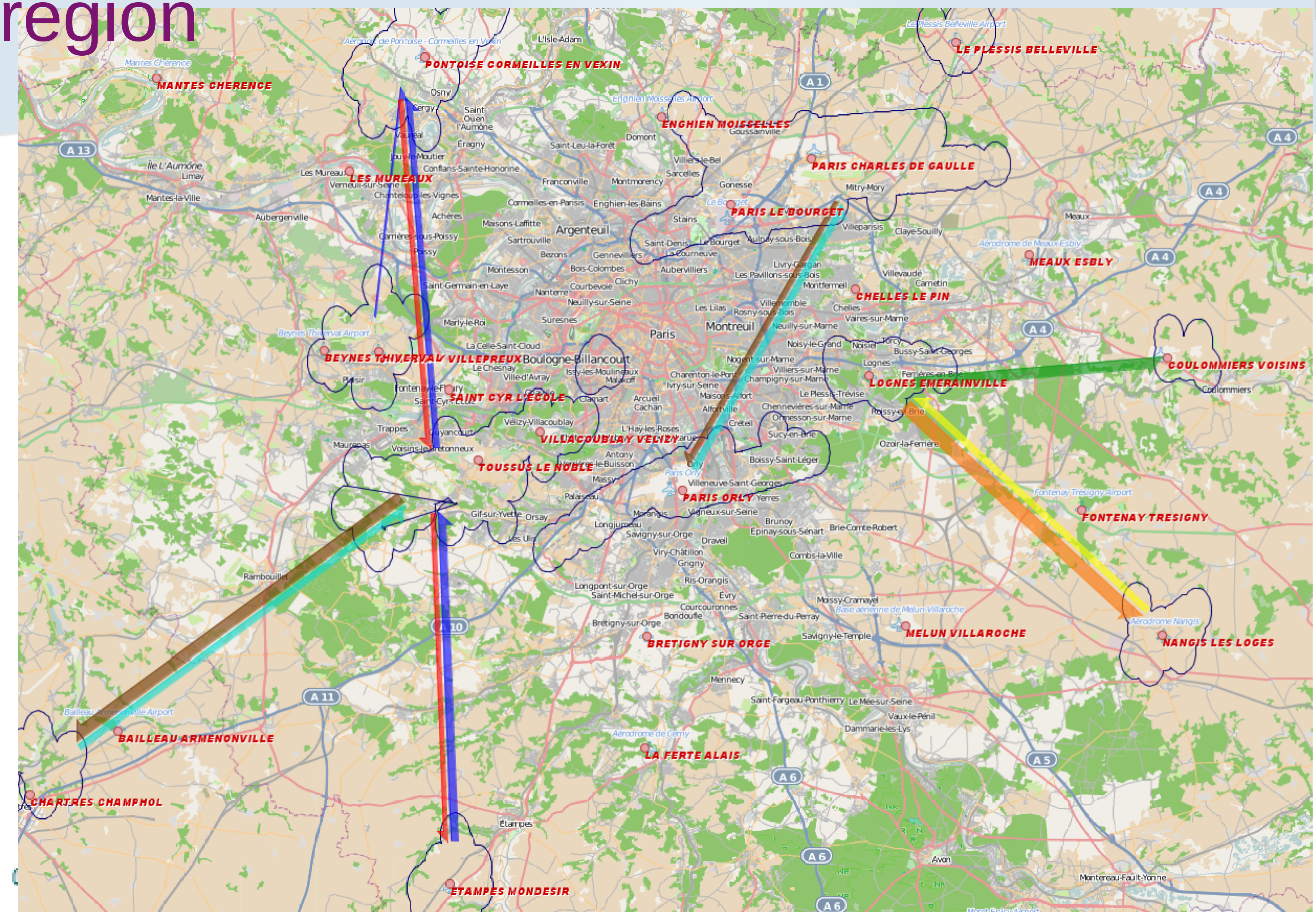
The arrows represent aggregate moves (flows) between areas.
The thickness is proportional to the total number of flights.
The colors represent the directions.
Minor flows (less than 5 flights) have been hidden.

A radial structure of the air traffic is visible (Paris ↔ periphery).
Different airports in Paris may be used for flights to/from the same city (e.g., Bordeaux, Toulouse, Marseille, Nice)



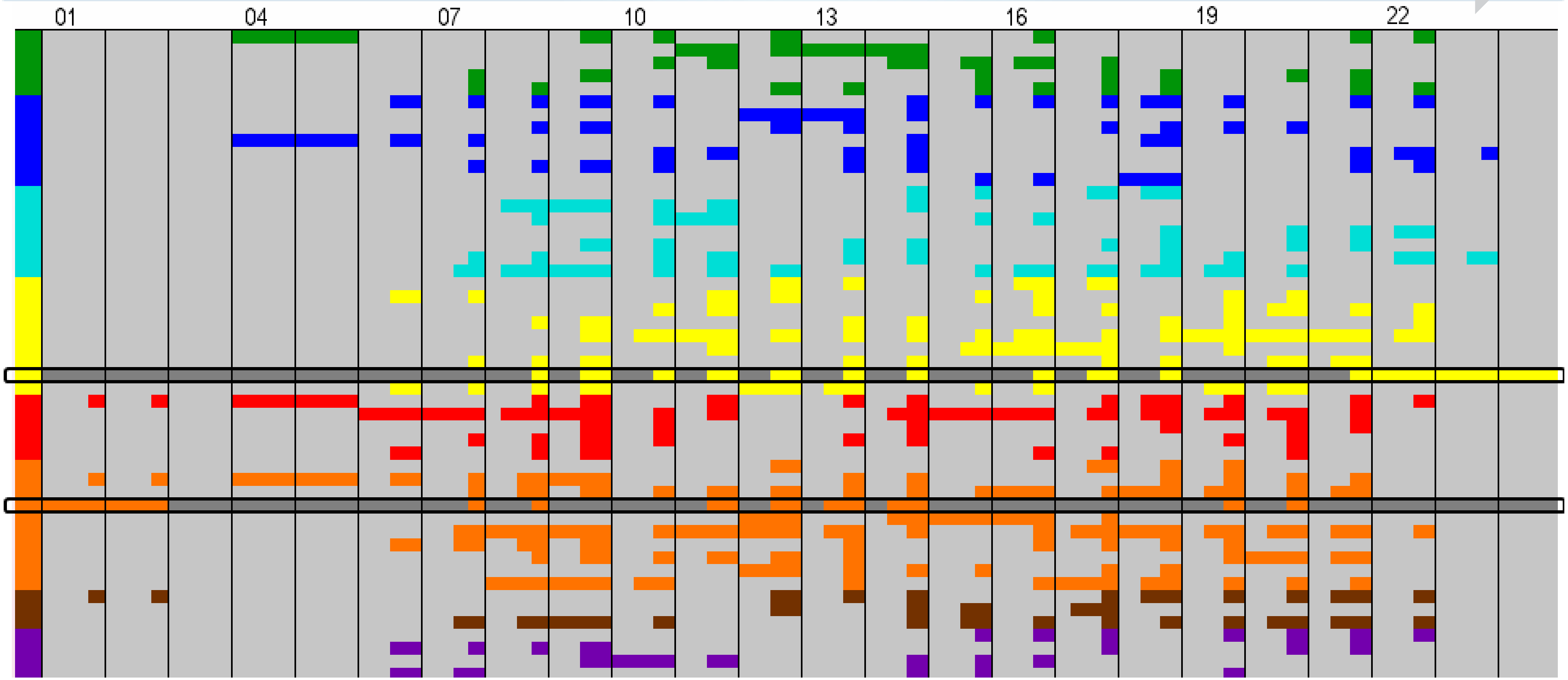
VAST 2011

Short-distance flows in the Paris region



Counts of the flights by hourly intervals

1 2 3 4

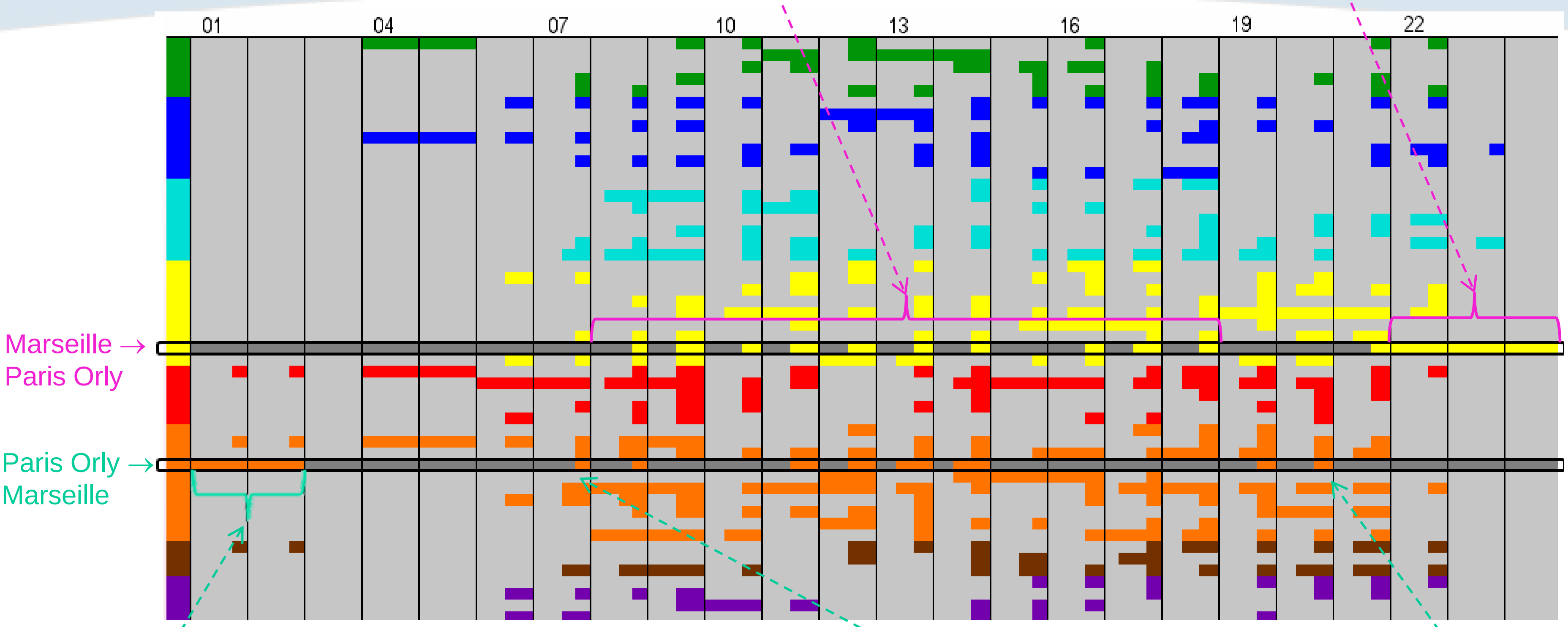


The rows correspond to the aggregate flows, i.e., pairs of places. The columns correspond to the time intervals used for the aggregation. The flow counts are represented by proportional lengths of the colored segments. The colors represent the flight directions. The rows corresponding to the connections Marseille → Paris Orly (yellow) and Paris Orly → Marseille (orange) are highlighted.

Connections Marseille ↔ Paris Orly

Three flights every hour from 22h to 24h

One or two flights every hour between 08h and 18h except 15h



Three flights per hour at 01h and 02h

One flight per hour in intervals 07h, 08h, from 11h to14h, 19h, 20h

VAST 2011

Conclusions:
generic and scalable analytical procedure

1. Extract relevant movement events from trajectories
2. Find and delineate significant places
   a) Find dense spatial clusters (S-clusters) of the events
      - Possibly, also by time (T), direction (D) and/or other attributes*
      - Possibly, 2-stage clustering: ST $\rightarrow$ S, STD $\rightarrow$ SD, SD $\rightarrow$ S, etc.

        ** appropriate similarity measures*

   b) Surround the clusters by spatial buffers or convex hulls
3. Aggregate the events and/or trajectories by the significant places and time intervals
4. Analyze the spatio-temporal aggregates

All steps could be scaled up by database processing

VAST 2011

http://geoanalytics.net/and