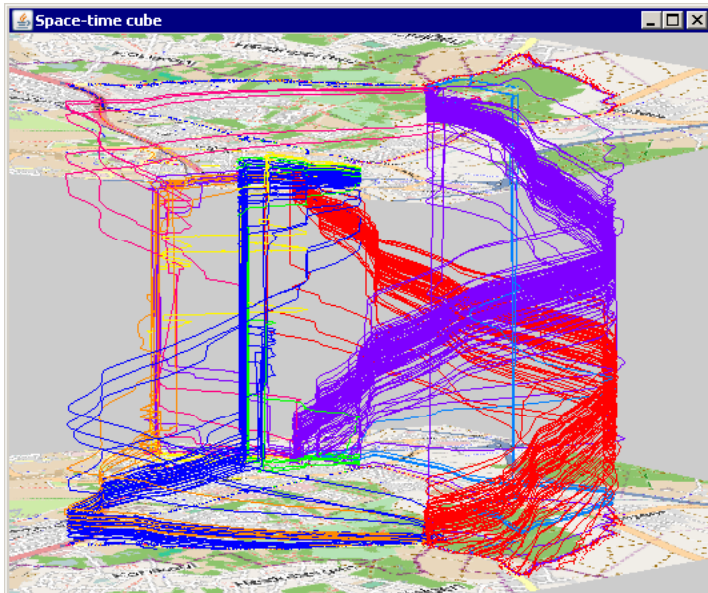
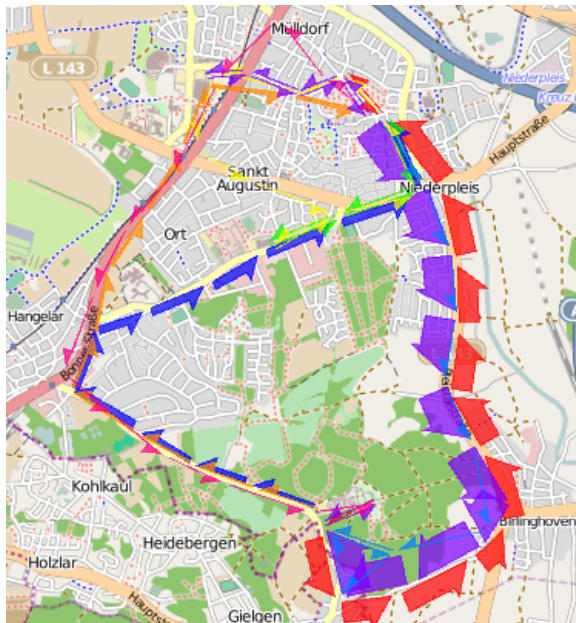

Visual Analytics of Movement



Gennady Andrienko & Natalia Andrienko

<http://geoanalytics.net>



Movement Data

Movement Data

Movement data is (typically) a temporal sequence of position records:

- {id,}x,y,t

Movement data are now collected in rapidly growing amounts owing to the development of tracking technologies:

- GPS, RFID, WiFi, ...

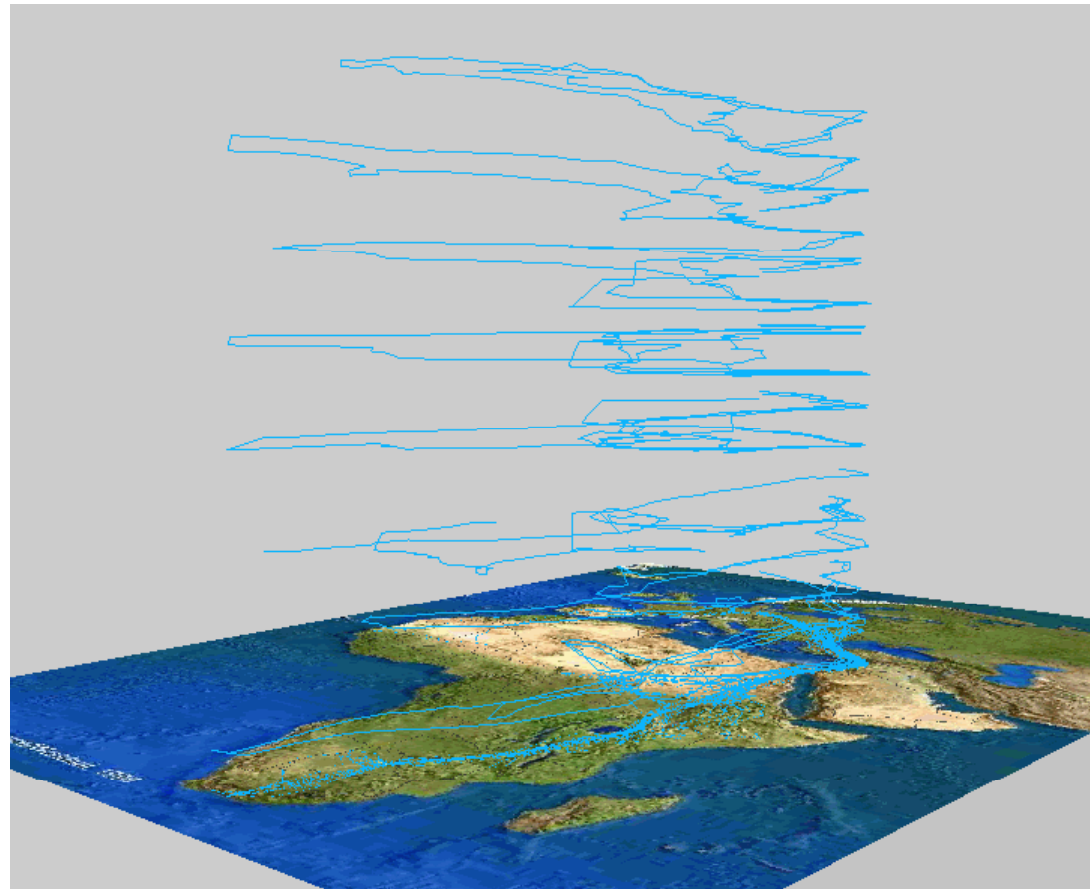
and also

- use of mobile phones, banking transactions, surveillance video (after transformation), sensor networks, ...

Examples of movement data: migration of white storks

Tracks of 35 storks during 8 years, about 2,000 positions

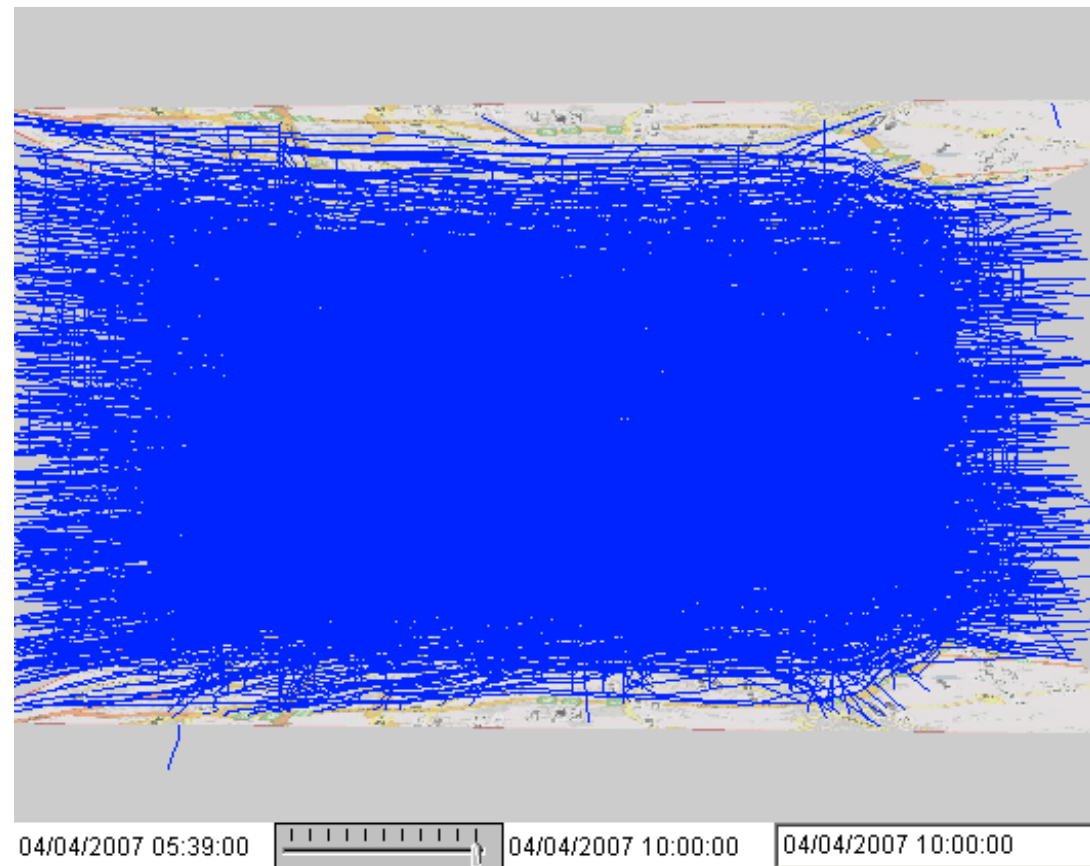
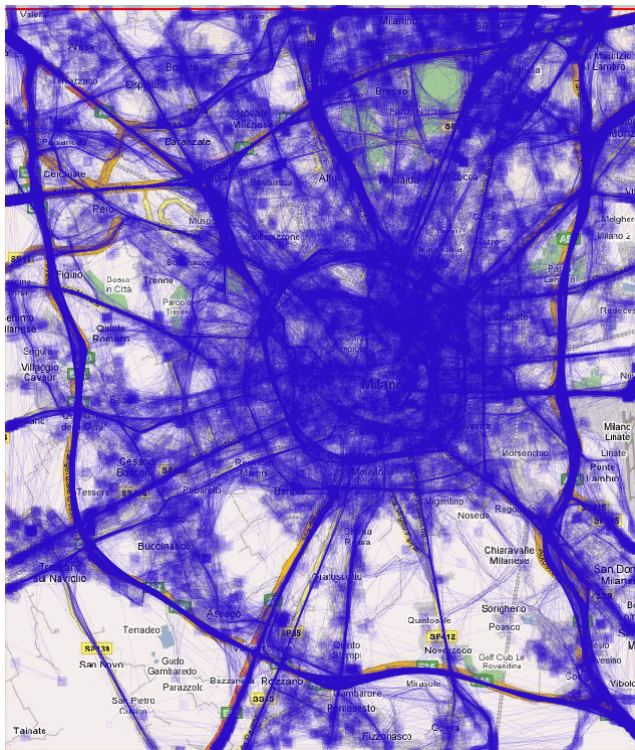
Animals



Examples of movement data: cars in Milan

- 2,075,216 position records of 17,241 cars during 1 week

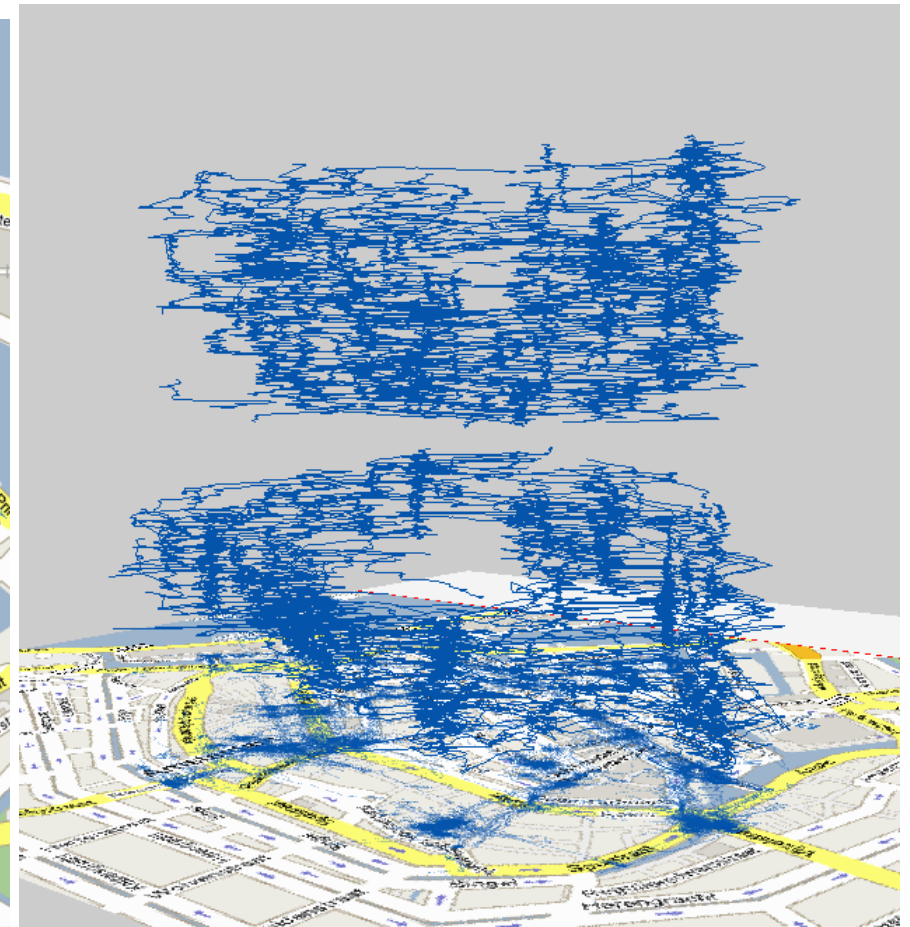
Vehicles;
network-constrained



Examples of movement data: children in Amsterdam

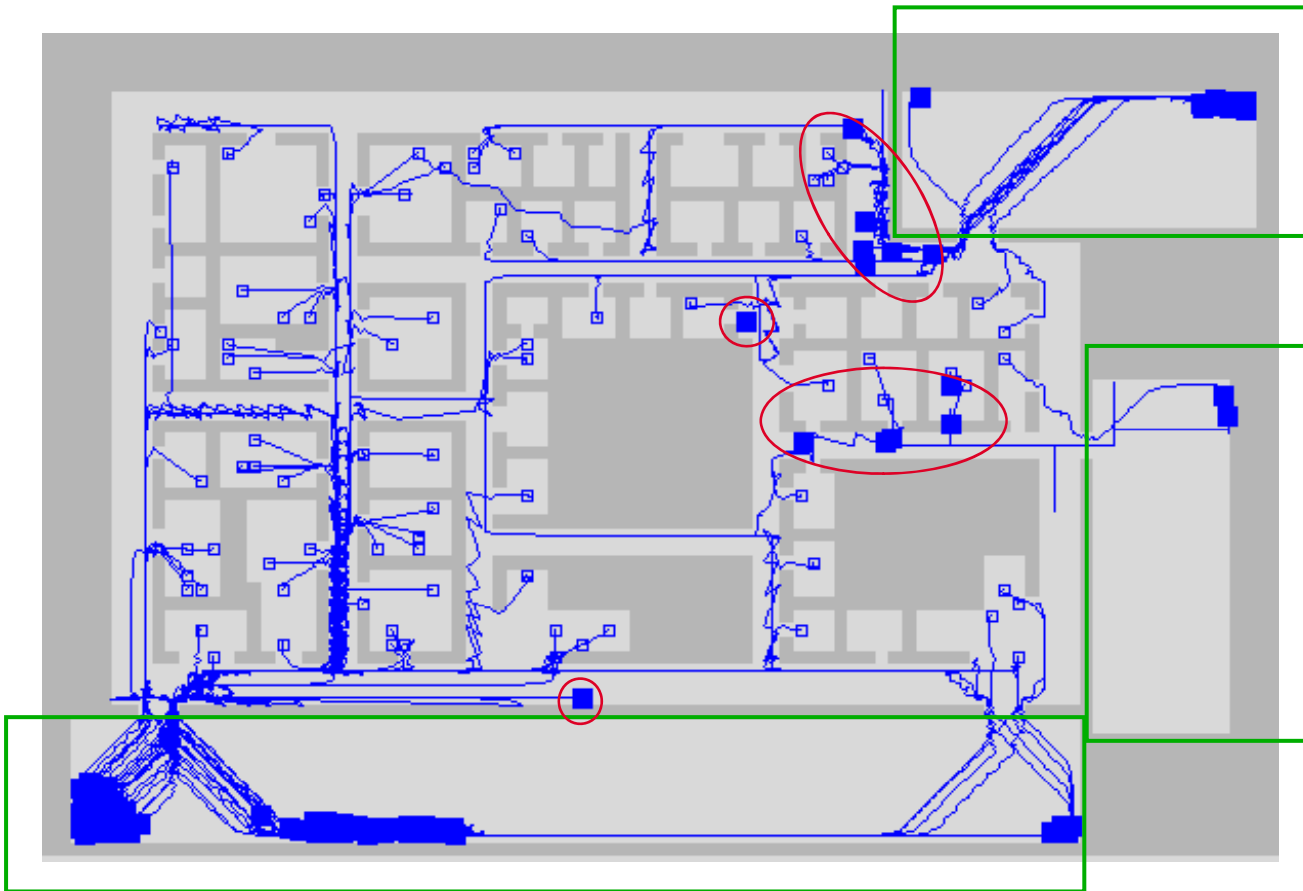
- GPS tracks of 303 schoolchildren playing an educational game in Amsterdam, about 57,000 points

Pedestrians

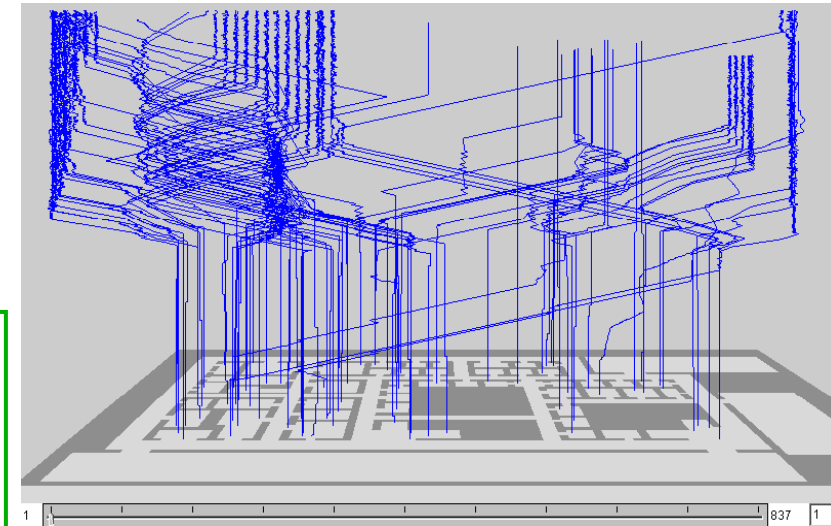


Examples of movement data: evacuation from a building

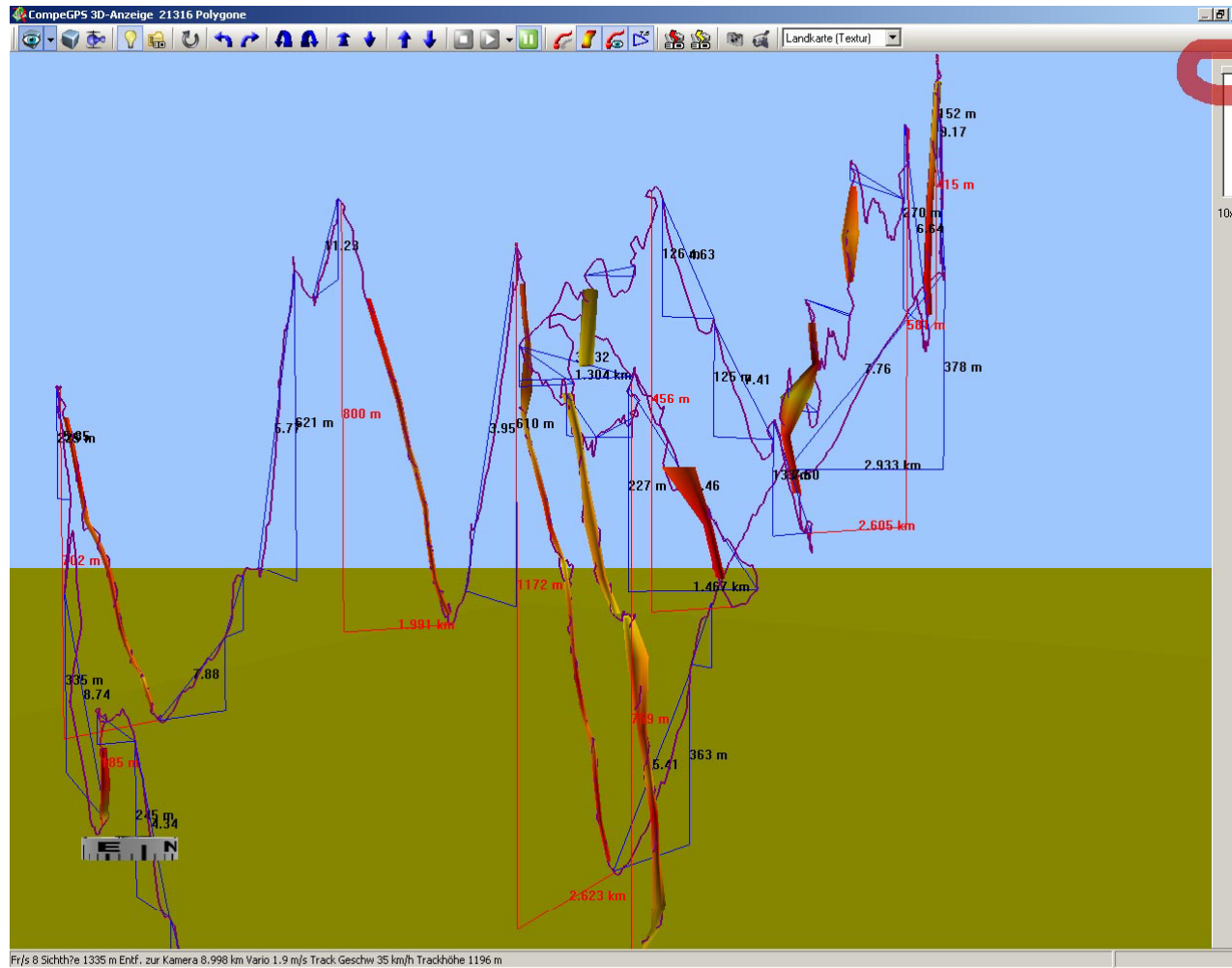
- Tracks of 82 people during evacuation, about 70,000 positions



Indoor movement;
simulated data



Examples of movement data: paragliding

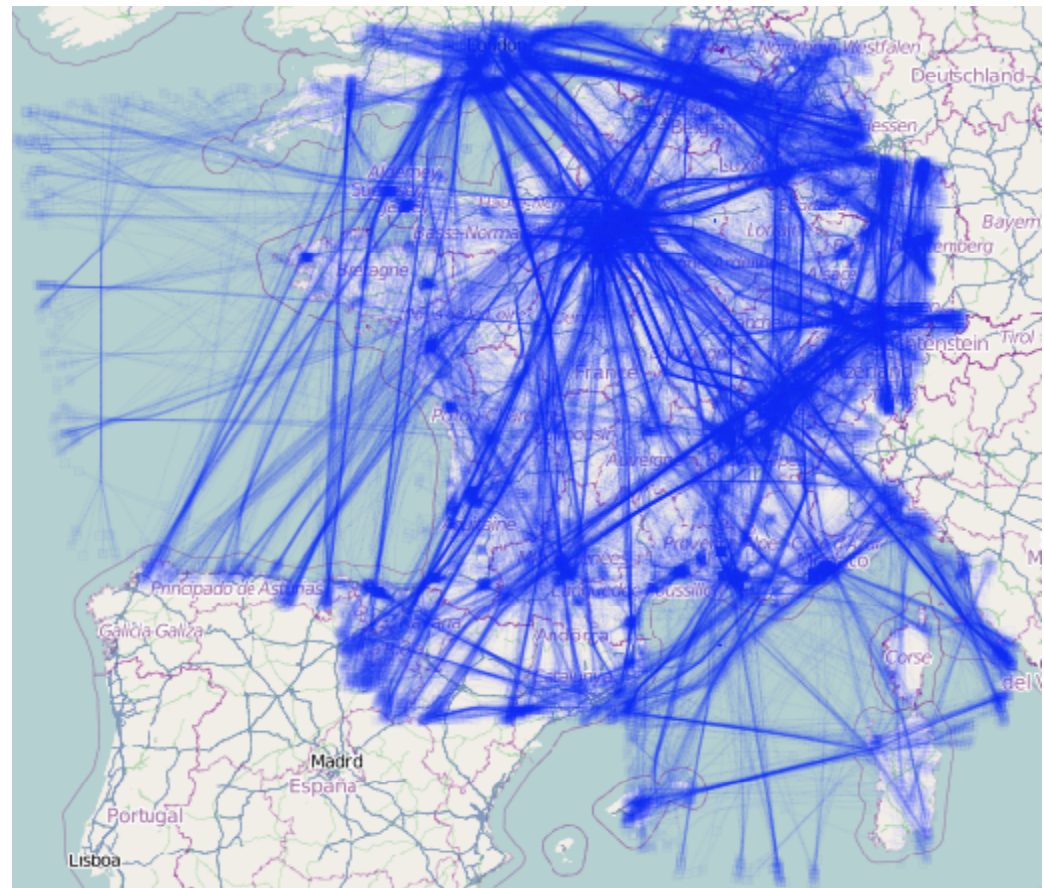


3D movement

Examples of movement data: air traffic

- 427,652 position records of 17,851 planes during 1 day
- Temporal resolution: 1-5 minutes

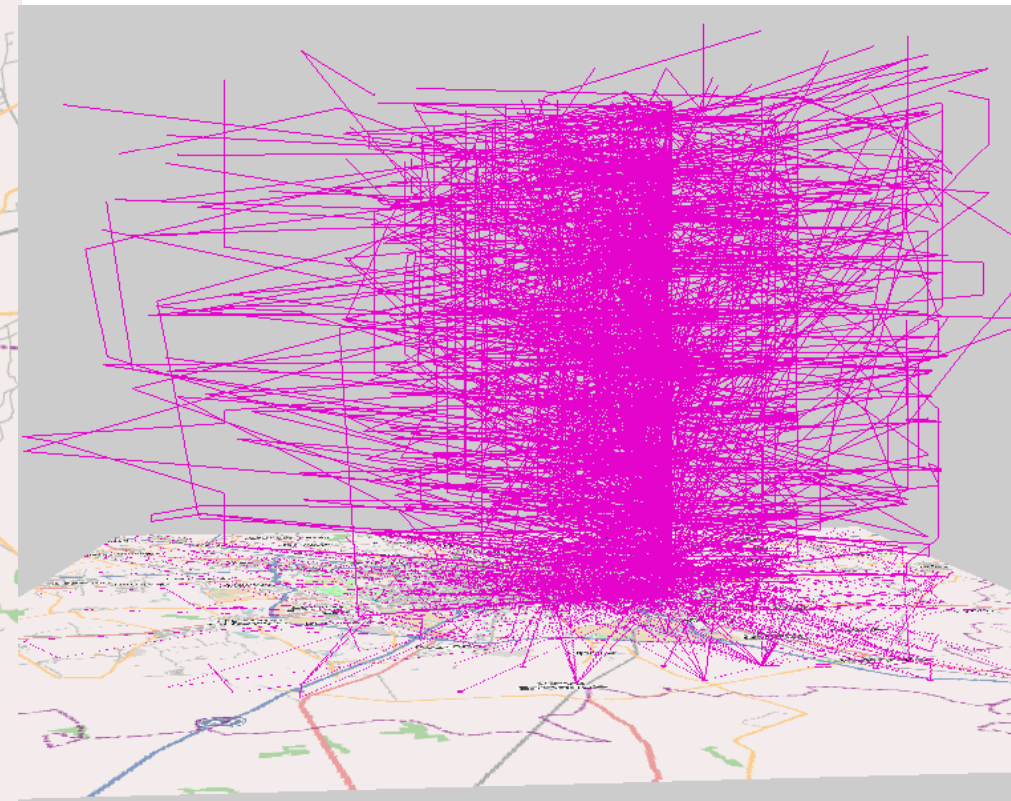
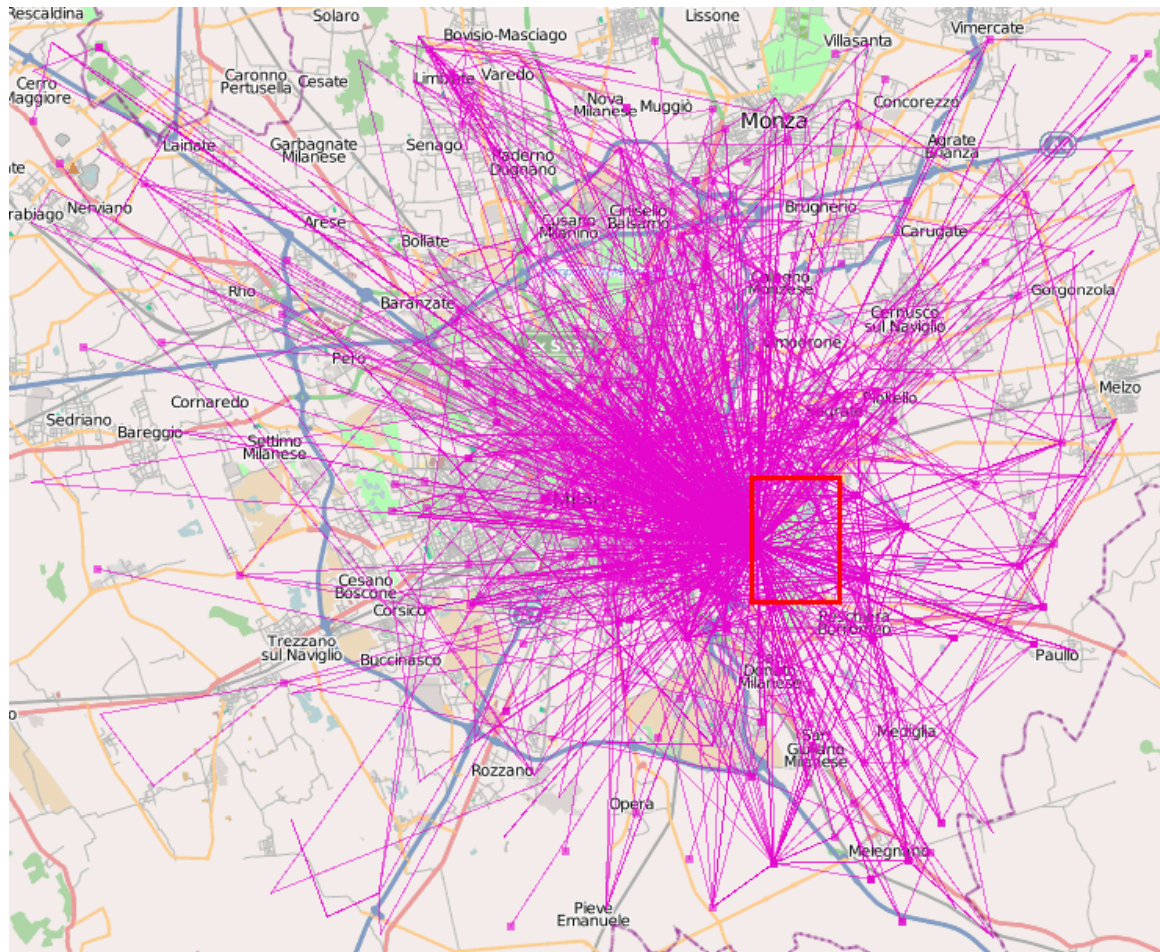
3D movement



Examples of movement data: calls of WIND customers in Milan

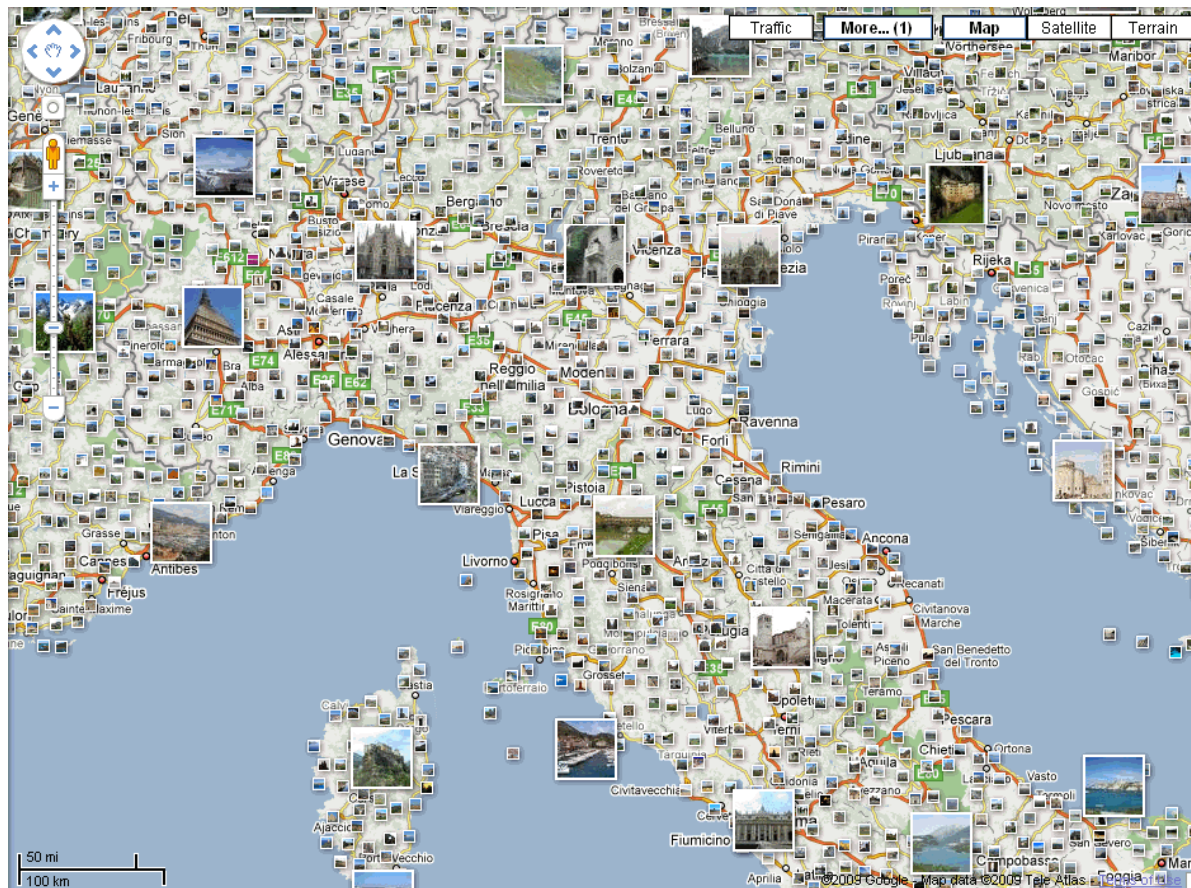
- 10 days; 550,000 customers; about 5,000,000 calls

Temporally sparse
position records



Examples of movement data: movement of flickr and panoramio photographers

- Millions of georeferenced photos are stored in panoramio.com (displayed in Google Earth); photos of the same user tell about his/her movement!

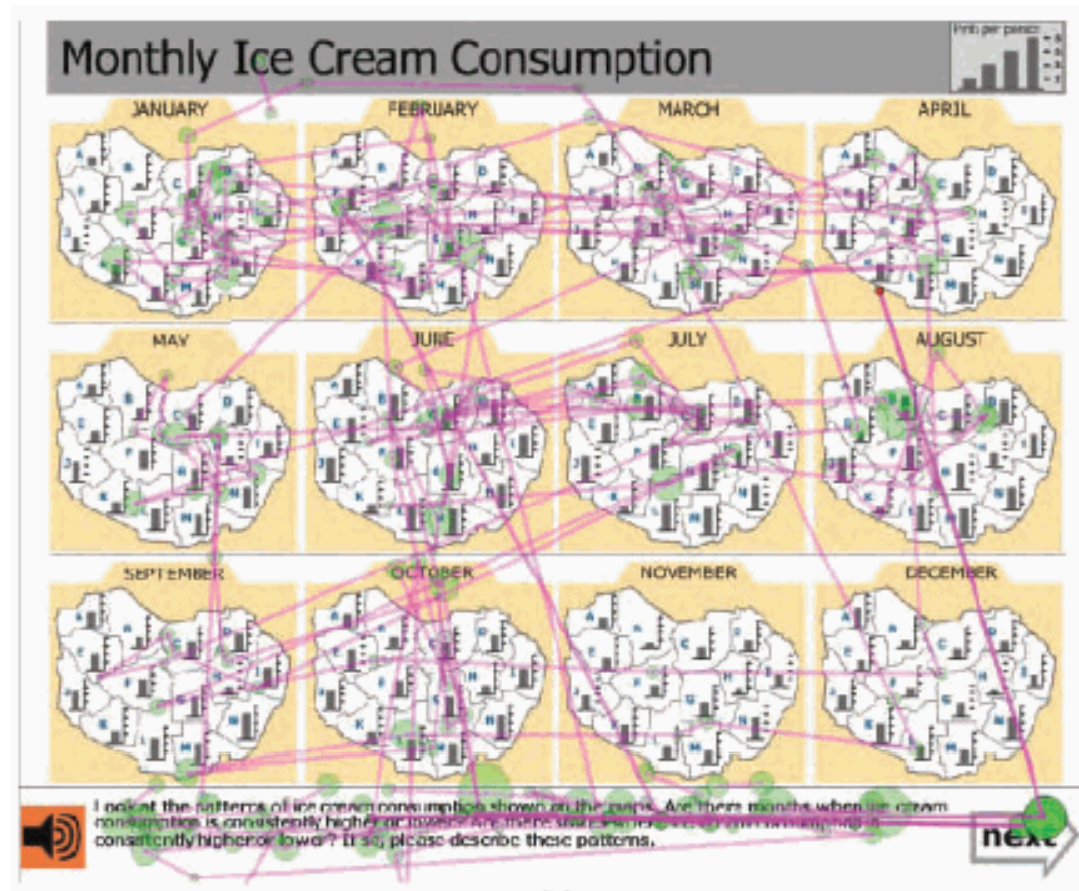


Temporally sparse
position records

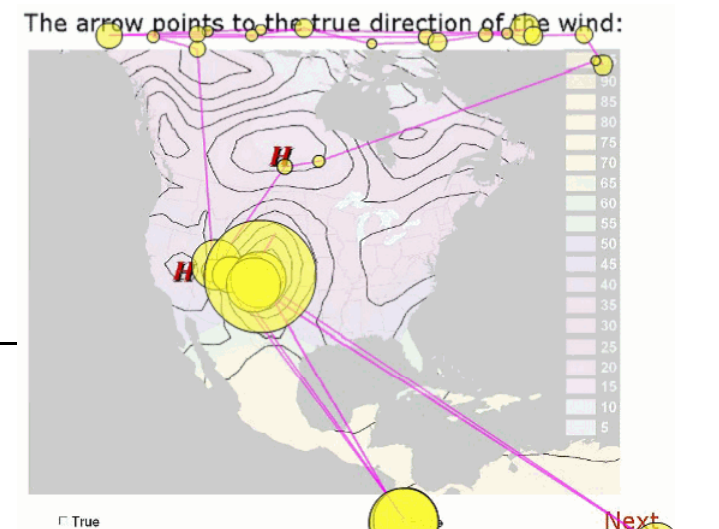
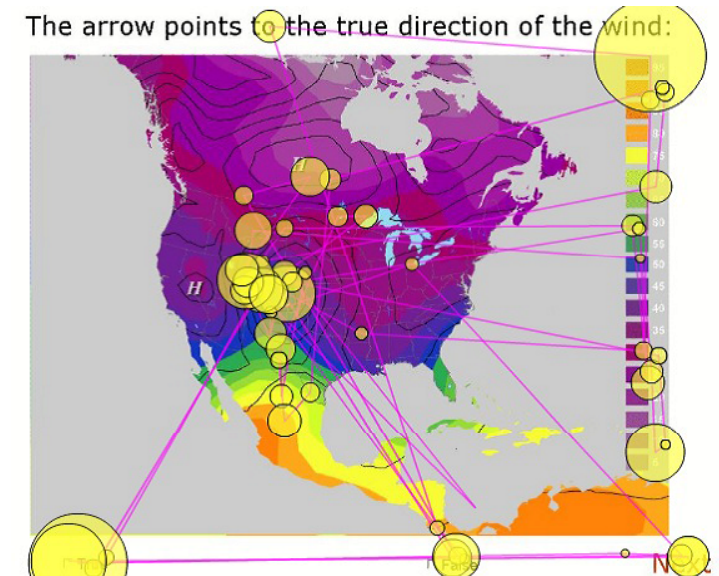
Web 2.0 data:
public access,
huge amounts

Examples of movement data: gaze movement

- Eye tracking is used in psychology for studying human attention, perception, and cognition



Non-geographic movement



Movement Data

Movement data: simple structure

- {id,}x,y,t

Complexities:

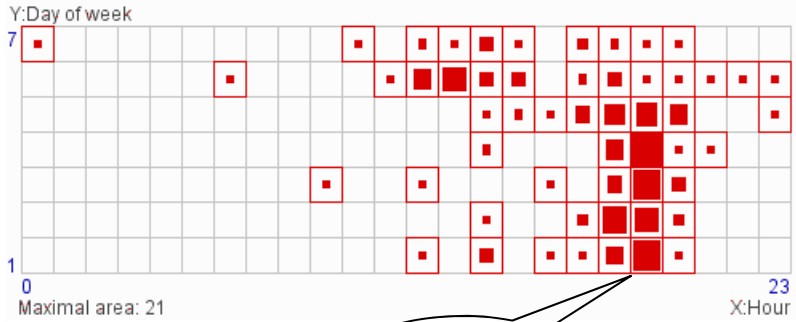
1. Huge amounts (number of moving entities, number of records)
2. Geographic space with its structure and complexity
3. Time, linear and also multiple nested and overlapping cycles
4. Data properties:
 - imprecision (errors in location, time, attributes)
 - irregular and/or sparse sampling
 - missing values
5. Large diversity of types of movement (constrained vs. unconstrained; smooth vs. abrupt; walking/wheeling/flying/sailing/swimming/jumping...)
6. Open-world, ill-defined problems

Exploration of Movement Data

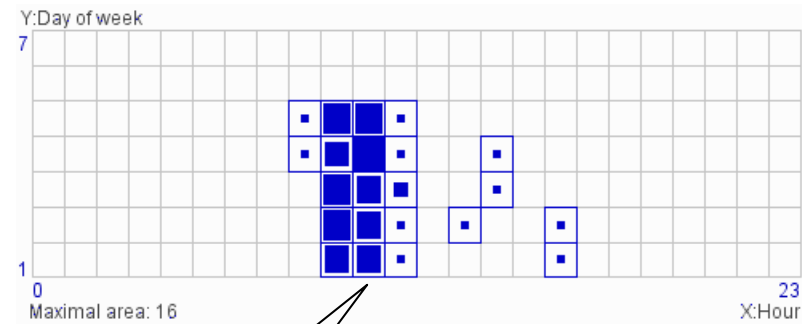
Movement data are usually semantically poor.

Semantics can be gained by means of exploratory analysis in which a human analyst uses his/her background knowledge and common sense.

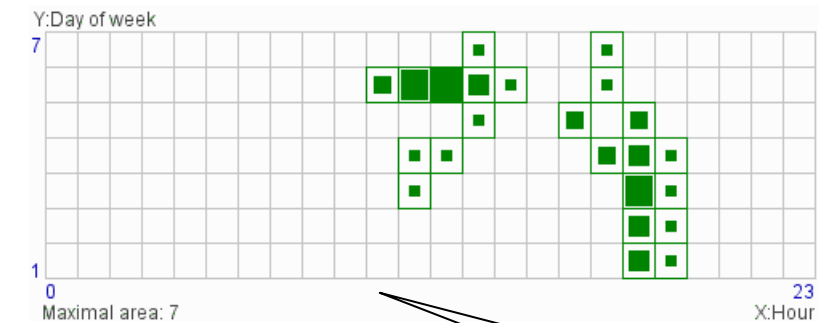
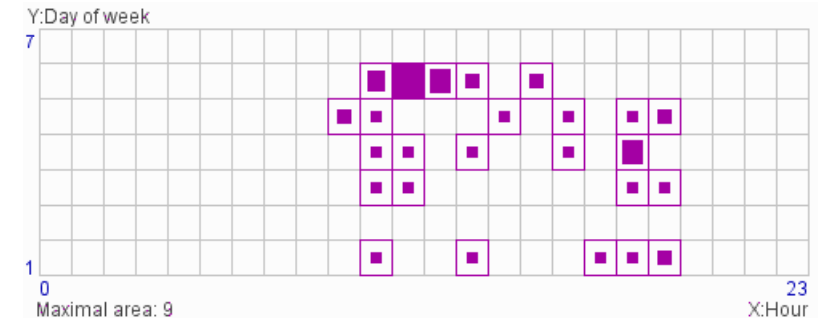
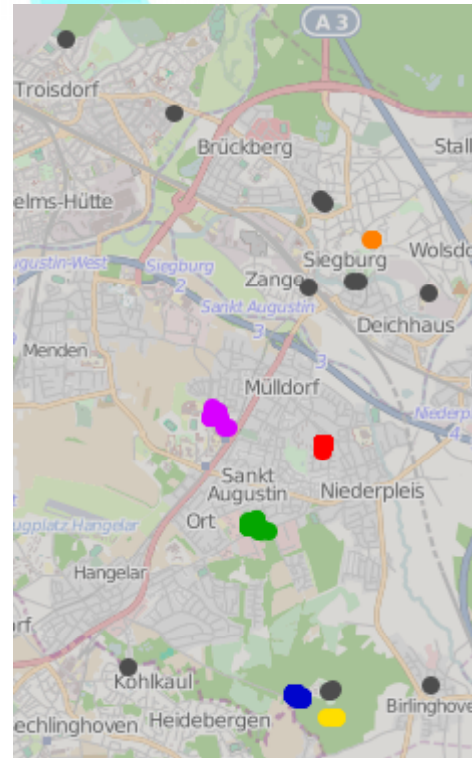
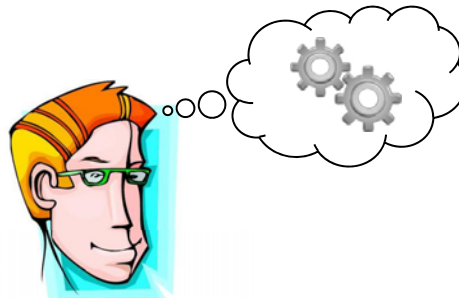
Spatial clusters of stops for ≥ 30 minutes



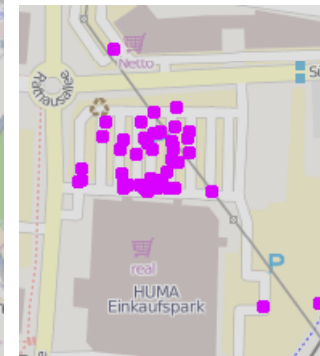
Home!



Work!

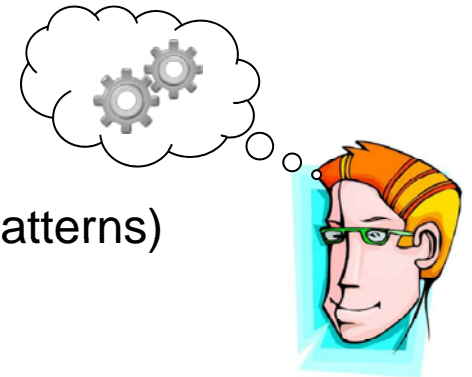


Shopping!



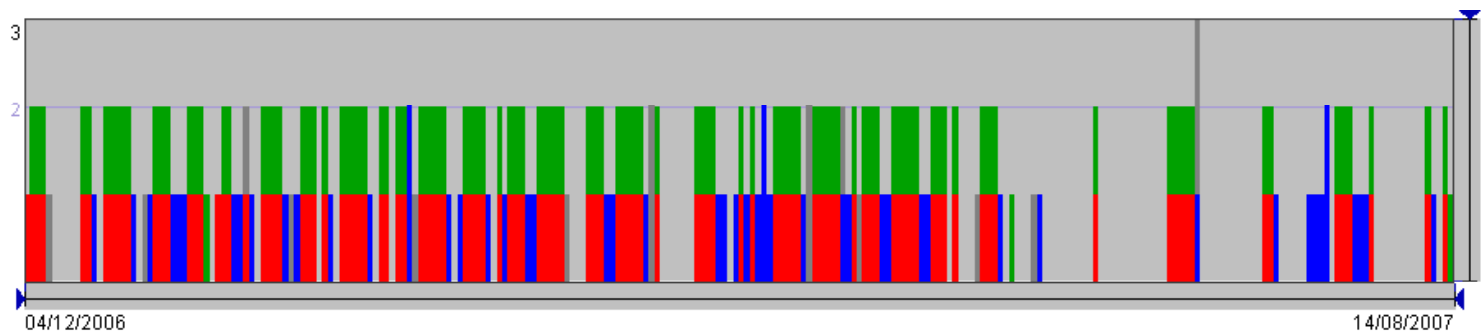
What was learned about the car owner:

- The places where the person lives, works, and shops and the places the person sometimes visits
- The durations of the stops; the times spent in the shopping areas
- The usual times of the trips and stops
- The typical routes and their distribution over time (daily and weekly patterns)



Other inferred personal information:

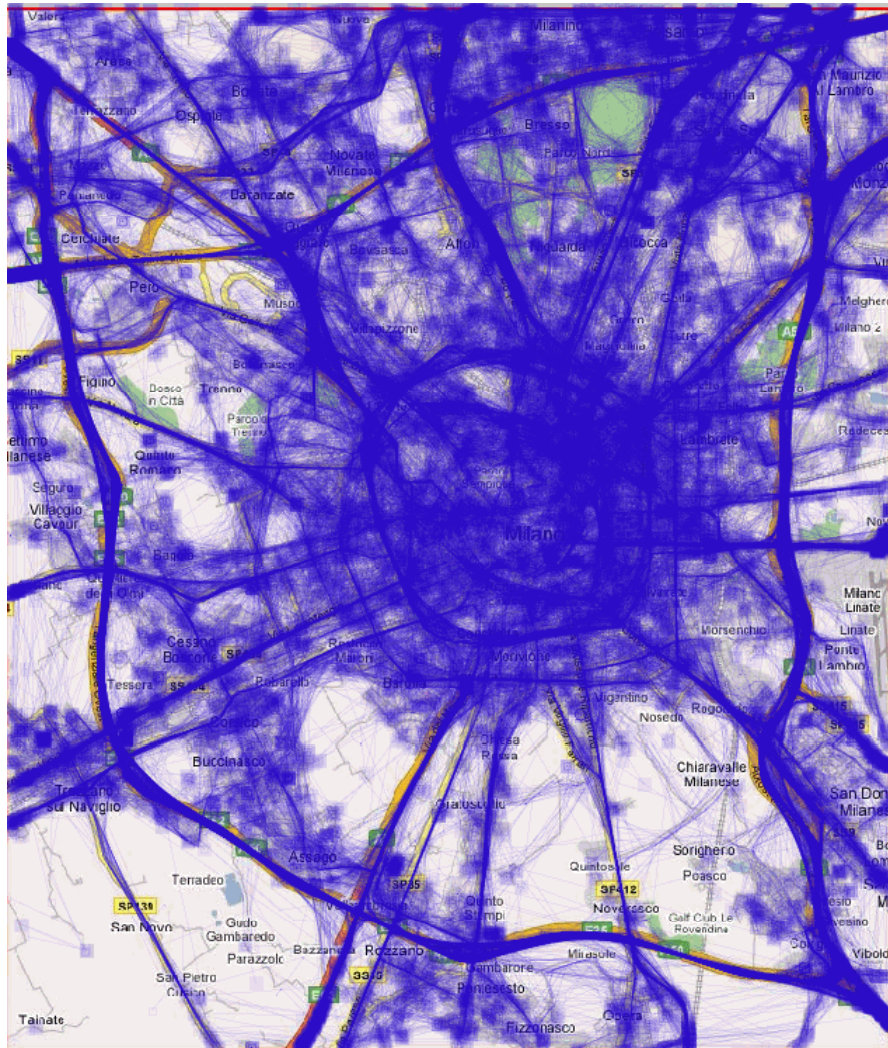
- The person has a flexible work schedule, has no small children, is often away or sick



*green: home to work
red: work to home
blue: home – home*

Aggregation of Movement Data

Example of visual analytics: analysis of city traffic



*E.g. Milan car movement data:
2,075,216 position records
of 17,241 cars during 1 week*

This is far too much for processing in RAM!

Approach:

- 1) aggregate the data in the database using the standard DBMS operations;
- 2) load the aggregates in RAM and visualise for enabling analysis

Spatial aggregation: by compartments (cells) of a territory division

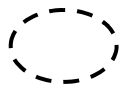
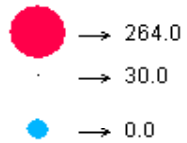
Temporal aggregation: by time intervals

- linear time
- cyclic time

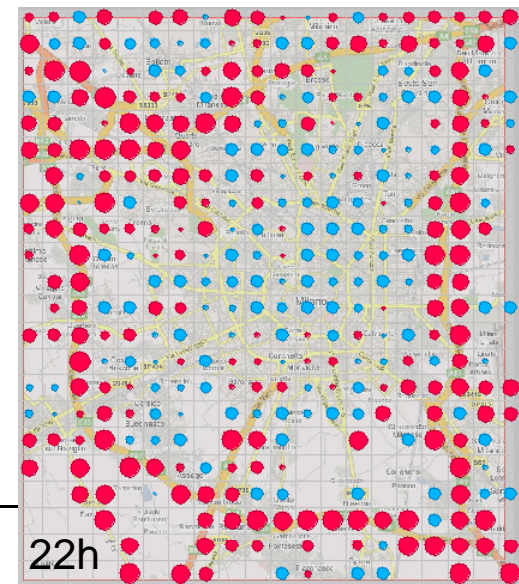
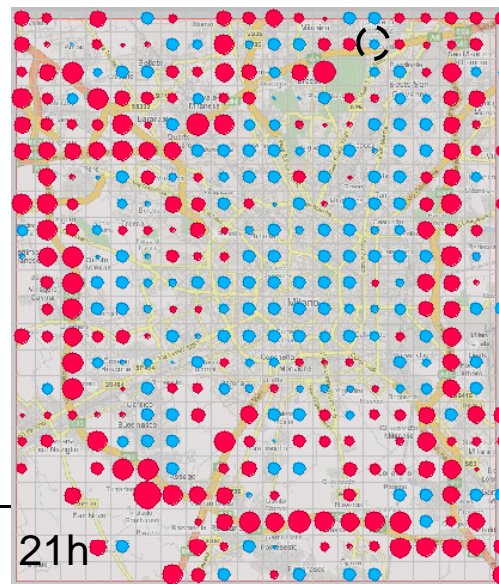
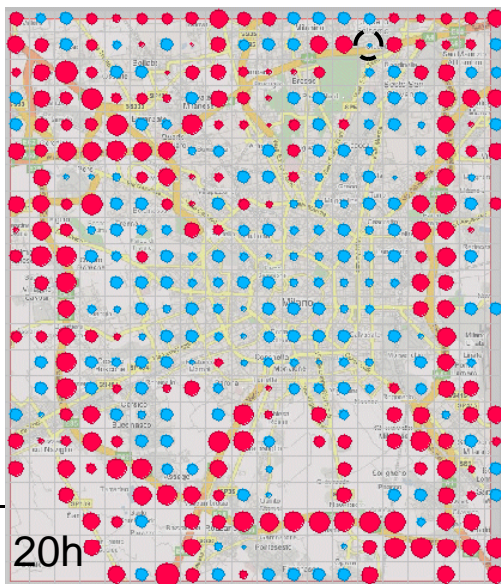
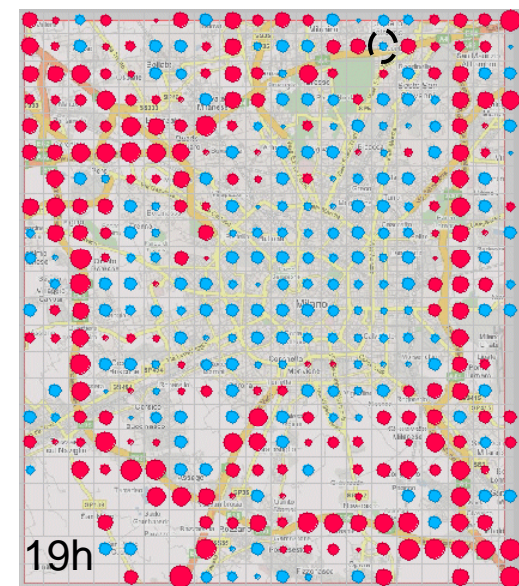
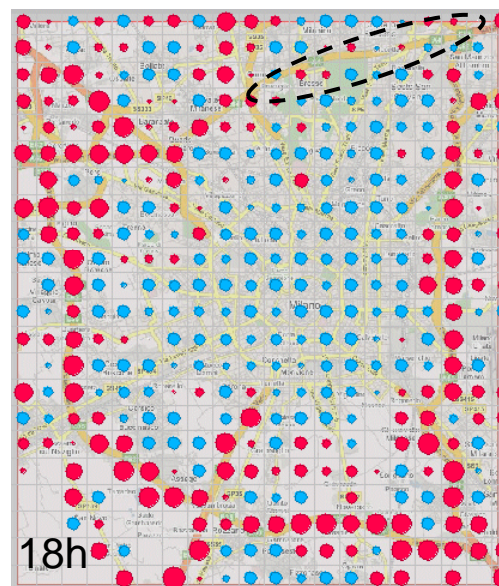
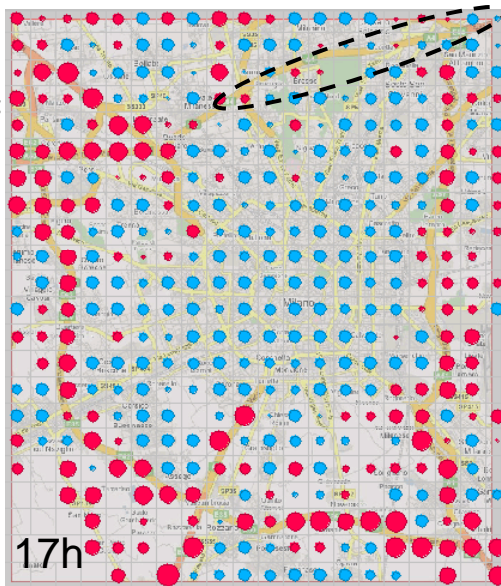
Median speeds by hourly intervals

Car data aggregated by grid
median speed (Day of week=1)

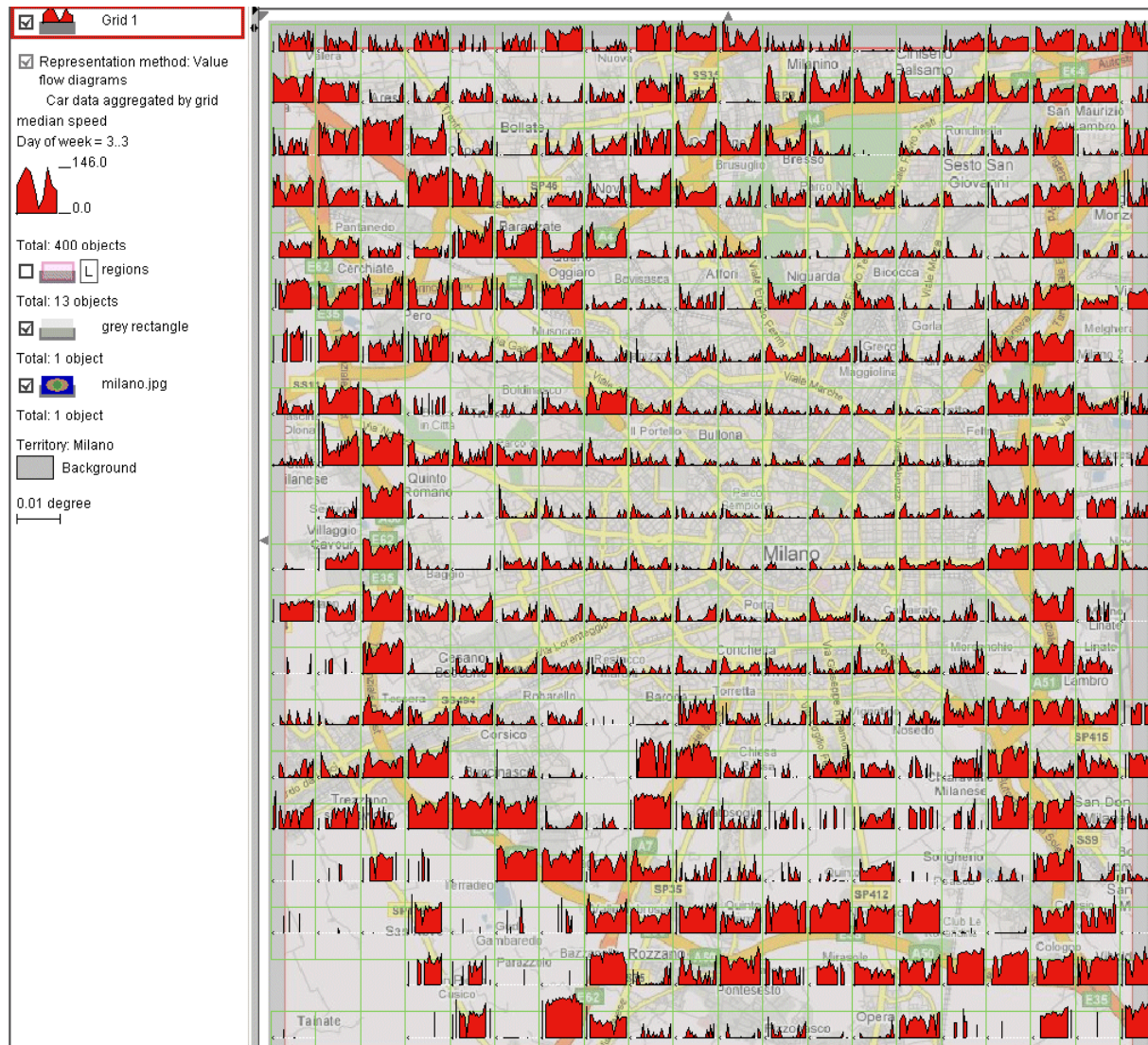
Circle area is proportional to value:



very low speeds
on a major belt
road



Hourly variation of the median speed in different places

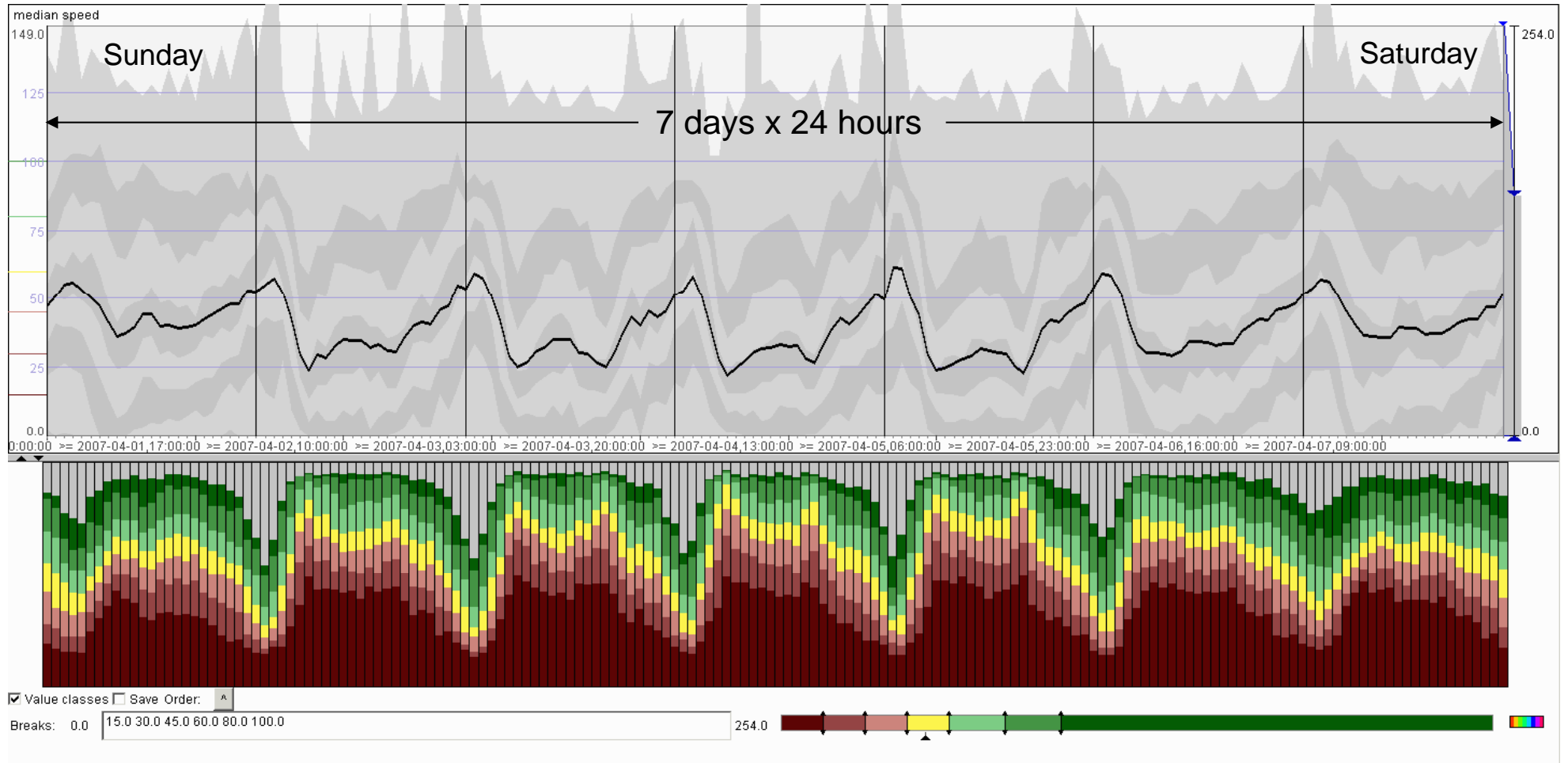


day: Wednesday



A frequent temporal pattern: significant drop of the speed in the morning and afternoon rush hours

Variation of the median speeds in all cells over the whole time period



Note the similarity of the daily patterns from Monday to Thursday and the difference of the Friday pattern

Daily and weekly cycles in the speed variation over the territory

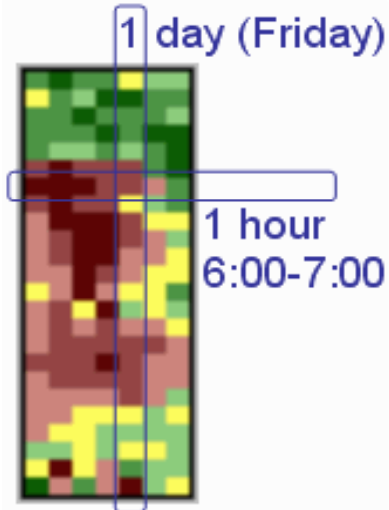
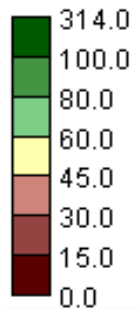
Car data aggregated by grid

Attribute: median speed

Parameters:

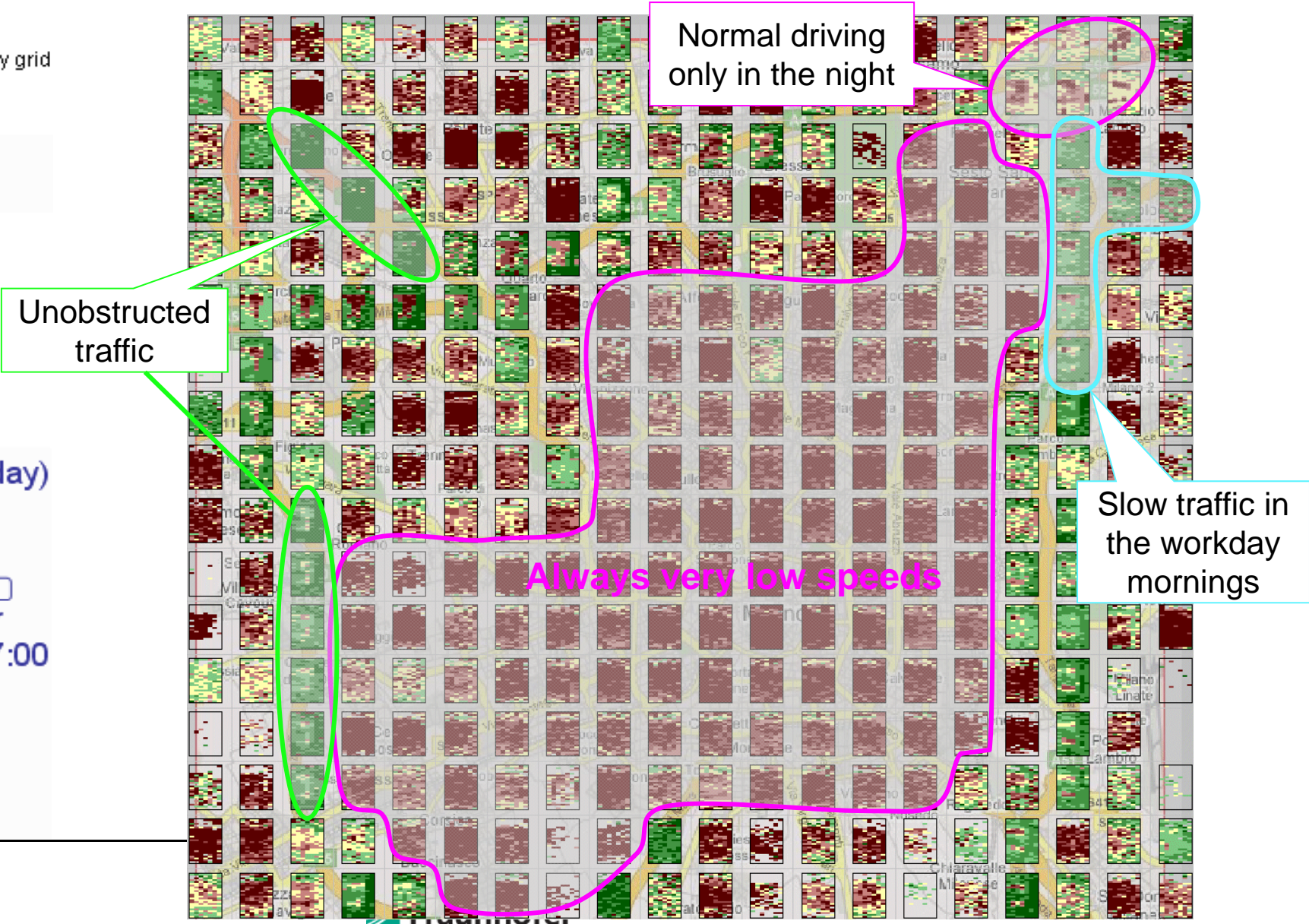
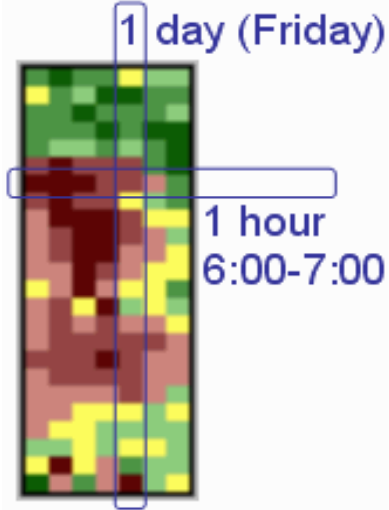
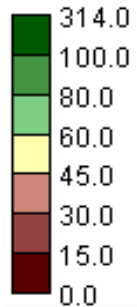
Hour: from 0 to 23

Day of week: from 1 to 7



Daily and weekly cycles in the speed variation over the territory

Car data aggregated by grid
Attribute: median speed
Parameters:
Hour: from 0 to 23
Day of week: from 1 to 7



Can computer help us in identifying these areas?

places

Aggregated data

times

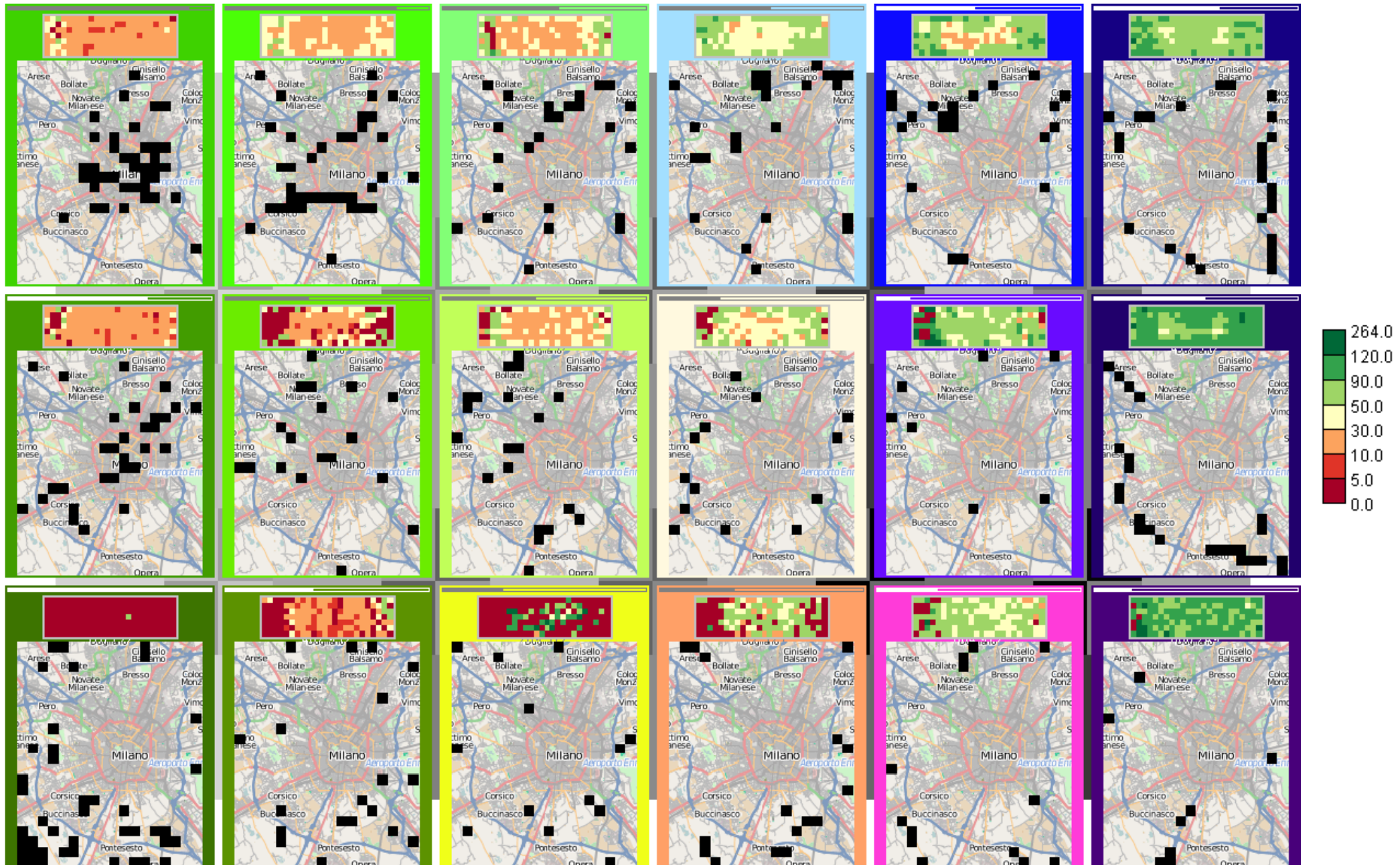
	N events by time intervals; Time interval (start)-01.1985	N events by time intervals; Time interval (start)-02.1985	N events by time intervals; Time interval (start)-03.1985	N events by time intervals; Time interval (start)-04.1985	N events by time intervals; Time interval (start)-05.1985	N events by time intervals; Time interval (start)-06.1985	N events by time intervals; Time interval (start)-07.1985	N events by time intervals; Time interval (start)-08.1985	N events by time intervals; Time interval (start)-09.1985	N events by time intervals; Time interval (start)-10.1985	N events by time intervals; Time interval (start)-11.1985	N events by time intervals; Time interval (start)-12.1985
ITF45	1	0	0	1	0	9	8	13	13	2	0	0
ITG28	0	0	0	0	0	0	0	0	0	0	0	0
ITG2A	0	0	0	0	0	0	0	0	0	0	0	0
ITD10	0	0	0	0	0	0	0	0	0	0	0	0
ITF61	39	1	0	21	4	8	165	411	215	67	6	2
ITG27	0	0	0	0	0	0	0	0	0	0	0	0
ITG2B	0	0	0	0	0	0	0	0	0	0	0	0
ITD33	0	0	0	4	1	1	0	2	2	9	1	0
ITD42	1	13	7	15	2	0	2	11	2	12	0	1
ITC44	0	0	0	3	0	0	0	0	0	0	0	1
ITD20	83	3	2	23	5	0	4	4	28	19	0	3
ITC14	0	0	0	0	0	0	0	0	0	0	0	0
ITD41	5	11	8	6	1	0	0	1	4	3	0	0
ITF62	0	0	0	1	0	0	14	22	10	0	0	1
ITG2C	0	0	0	0	0	0	0	0	0	0	0	0
ITC47	2	2	4	7	2	0	0	2	5	1	0	7
ITC42	4	2	2	15	6	0	1	1	0	2	1	6
ITC43	0	0	0	0	0	0	0	0	0	0	0	0
ITC41	1	0	1	19	9	1	0	0	0	2	0	0
ITD34	0	1	0	4	0	0	0	2	0	2	0	1
ITC46	0	0	4	19	5	0	0	3	6	9	0	2
ITF63	0	1	0	22	3	10	189	411	161	26	4	0
ITF64	3	0	0	0	0	0	0	0	0	0	0	0
ITD43	1	0	2	2	0	1	5	1	6	9	0	0
ITC20	1	0	5	15	1	1	2	0	5	2	0	3
ITD32	1	1	0	0	0	0	2	2	14	3	0	1
ITC12	0	0	1	35	4	0	0	0	3	2	0	7
ITC15	1	0	5	20	5	0	1	0	2	2	0	1
ITD35	0	0	0	0	0	1	0	0	0	2	0	0
ITD31	0	1	0	5	0	0	1	4	17	13	0	0
ITD44	4	4	6	5	1	1	2	4	4	9	0	0
ITC13	0	0	0	3	2	0	0	0	0	3	0	0
ITC45	0	0	0	2	0	0	0	0	1	0	0	0
ITD36	1	0	0	0	0	0	3	4	1	0	0	0
ITF65	1	0	0	0	0	0	0	0	5	0	0	0
ITG13	0	0	0	0	0	0	0	0	0	0	0	0
ITG12	0	0	0	0	0	0	0	0	0	0	0	0
ITG11	0	0	0	0	0	0	0	0	0	0	0	0
ITC11	0	2	17	34	11	0	0	4	8	6	3	18
ITC4A	0	0	0	0	0	0	0	0	0	0	0	0
ITC49	0	0	0	0	0	0	0	0	0	0	0	0
ITC4B	0	0	0	0	0	0	0	0	0	0	0	0
ITC48	0	0	1	7	0	0	0	4	0	0	0	0
ITC18	0	0	3	8	1	0	1	13	9	3	0	3
ITD37	0	0	0	1	0	1	0	0	0	0	0	0
ITD51	0	1	1	2	1	0	2	0	2	1	0	0

Apply clustering for grouping places by similarity of temporal dynamics

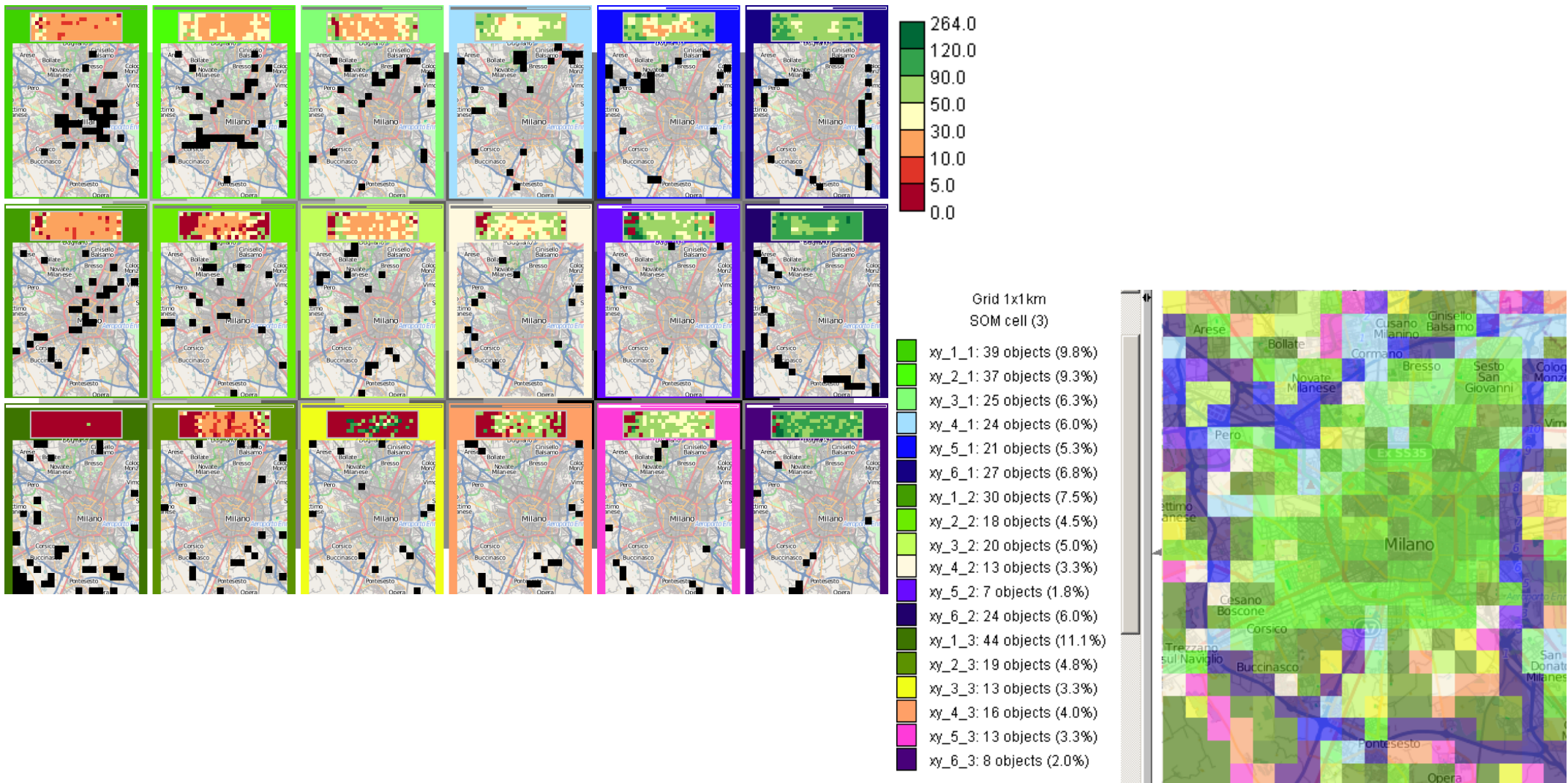
Self-Organizing Map (SOM)

- Self-Organizing Map (Kohonen 2001) is a neural network type vector projection and quantization algorithm.
- Input: a set of feature vectors, i.e. combinations of values of multiple attributes.
- By means of a competitive, iterative training process, a network of *prototype vectors* (or neurons, or cells) is trained (adjusted) to represent the input vector data.
- Output: a two-dimensional layout consisting of rectangular or hexagonal cells; each cell corresponds to one prototype vector and contains input vectors close to it.
- The network can be interpreted as a set of clusters and simultaneously as a two-dimensional projection of the data, which is convenient for visualization.

SOM result for dynamics of mean speeds in cells



SOM cell colours transmitted to map



Another use of clustering for aggregated data

Apply clustering for grouping time intervals by similarity of spatial distributions of values

places
↓
times →

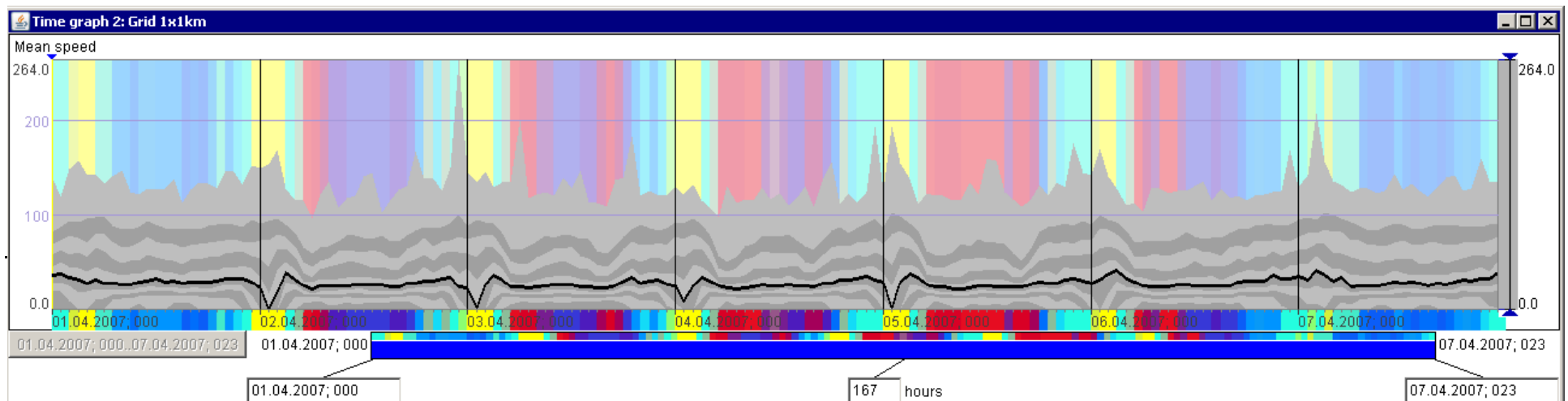
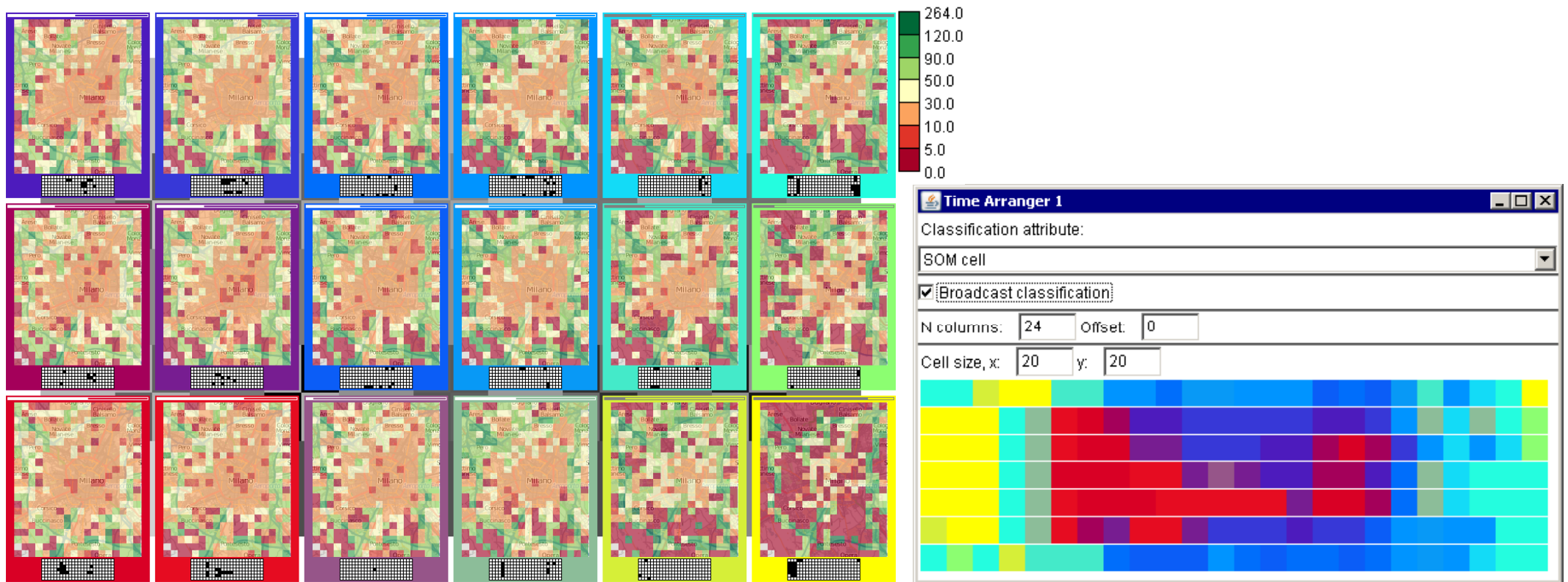
Aggregated data

	N events by time intervals; Time interval (start)=01.1985	N events by time intervals; Time interval (start)=02.1985	N events by time intervals; Time interval (start)=03.1985	N events by time intervals; Time interval (start)=04.1985	N events by time intervals; Time interval (start)=05.1985	N events by time intervals; Time interval (start)=06.1985	N events by time intervals; Time interval (start)=07.1985	N events by time intervals; Time interval (start)=08.1985	N events by time intervals; Time interval (start)=09.1985	N events by time intervals; Time interval (start)=10.1985	N events by time intervals; Time interval (start)=11.1985	N events by time intervals; Time interval (start)=12.1985
ITF45	1	0	0	1	0	9	8	13	13	2	0	0
ITG28	0	0	0	0	0	0	0	0	0	0	0	0
ITG2A	0	0	0	0	0	0	0	0	0	0	0	0
ITD10	0	0	0	0	0	0	0	0	0	0	0	0
ITF61	39	1	0	21	4	8	165	411	215	67	6	2
ITG27	0	0	0	0	0	0	0	0	0	0	0	0
ITG2B	0	0	0	0	0	0	0	0	0	0	0	0
ITD33	0	0	0	4	1	1	0	2	2	9	1	0
ITD42	1	13	7	15	2	0	2	11	2	12	0	1
ITC44	0	0	0	3	0	0	0	0	0	0	0	1
ITD20	83	3	2	23	5	4	4	28	19	0	3	0
ITC14	0	0	0	0	0	0	0	0	0	0	0	0
ITD41	5	11	8	6	1	0	0	1	4	3	0	0
ITF62	0	0	0	1	0	0	14	22	10	0	0	1
ITG2C	0	0	0	0	0	0	0	0	0	0	0	0
ITC47	2	2	4	7	2	0	0	2	5	1	0	7
ITC42	4	2	2	15	6	0	1	1	0	2	1	6
ITC43	0	0	0	0	0	0	0	0	0	0	0	0
ITC41	1	0	1	19	9	0	0	0	2	0	0	0
ITD34	0	1	0	4	0	0	0	2	0	2	0	1
ITC46	0	0	4	19	5	0	0	3	6	9	0	2
ITF63	0	1	0	22	3	10	189	411	161	26	4	0
ITF64	3	0	0	0	0	0	0	0	0	0	0	0
ITD43	1	0	2	2	0	1	5	1	6	9	0	0
ITC20	1	0	5	15	1	1	2	0	5	2	0	3
ITD32	1	1	0	0	0	0	2	2	14	3	0	1
ITC12	0	0	1	35	4	0	0	0	3	2	0	7
ITC15	1	0	5	20	5	0	1	0	2	2	0	1
ITD35	0	0	0	0	0	1	0	0	0	2	0	0
ITD31	0	1	0	5	0	0	1	4	17	13	0	0
ITD44	4	4	6	5	1	1	2	4	4	9	0	0
ITC13	0	0	0	3	2	0	0	0	0	3	0	0
ITC45	0	0	0	2	0	0	0	0	1	0	0	0
ITD38	1	0	0	0	0	0	3	4	1	0	0	0
ITF65	1	0	0	0	0	0	0	0	5	0	0	0
ITG13	0	0	0	0	0	0	0	0	0	0	0	0
ITG12	0	0	0	0	0	0	0	0	0	0	0	0
ITG11	0	0	0	0	0	0	0	0	0	0	0	0
ITC11	0	2	17	34	11	0	0	4	8	6	3	18
ITC4A	0	0	0	0	0	0	0	0	0	0	0	0
ITC49	0	0	0	0	0	0	0	0	0	0	0	0
ITC4B	0	0	0	0	0	0	0	0	0	0	0	0
ITC48	0	0	1	7	0	0	0	4	0	0	0	0
ITC18	0	0	3	8	1	0	1	13	9	3	0	3
ITD37	0	0	0	1	0	1	0	0	0	0	0	0
ITD51	0	1	1	2	1	0	2	0	2	1	0	0

SOM result for spatial distributions of mean speeds in hourly intervals



SOM cell colours transmitted to temporal displays



Exploration of Trajectories

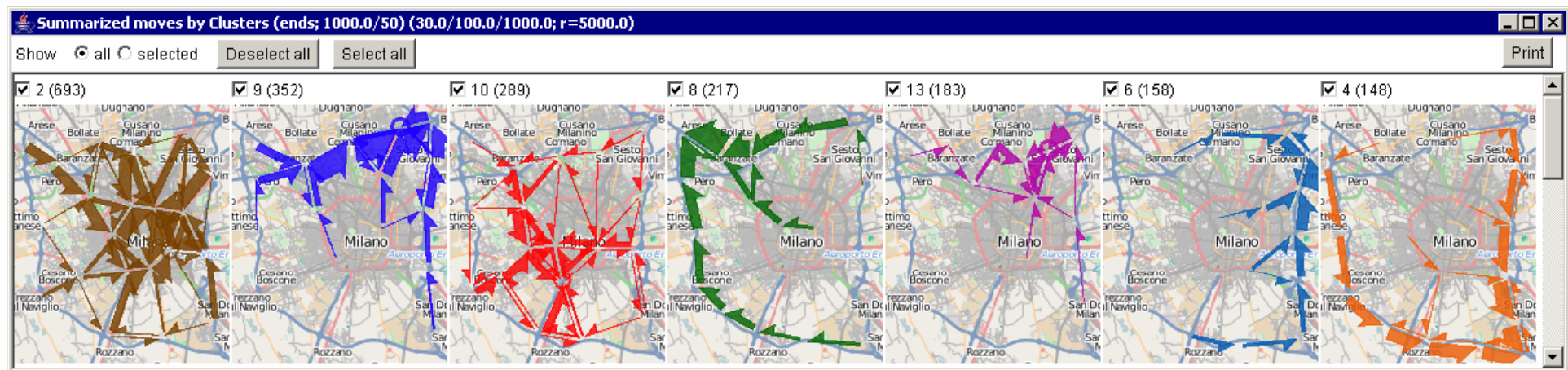
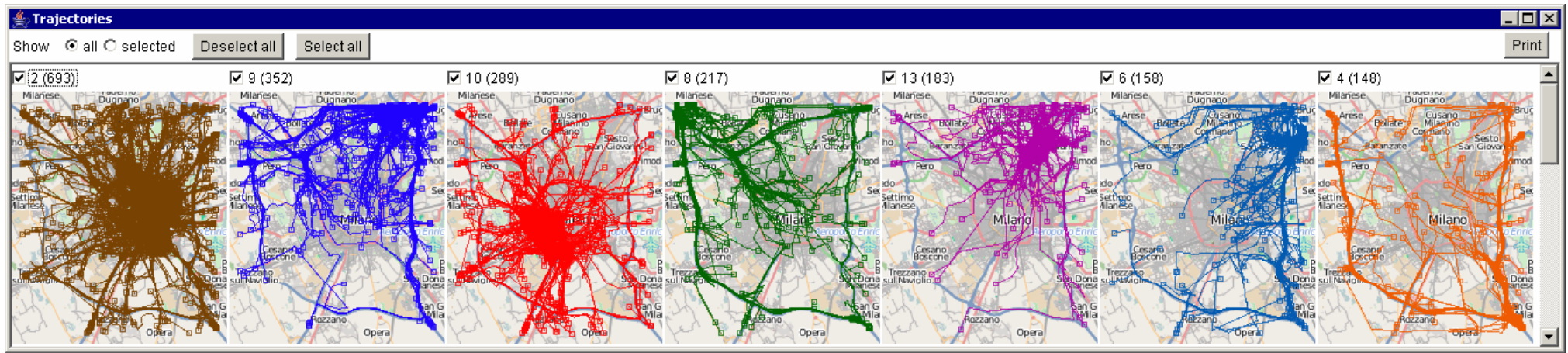
The standard spatio-temporal aggregation is not enough

- The aggregation allows us to deal with very large amounts of data
- It helps us to notice significant patterns and answer a number of questions about collective movements
- ... but we cannot retrieve information about *trips*: origins and destinations, routes, durations, speed dynamics, etc.

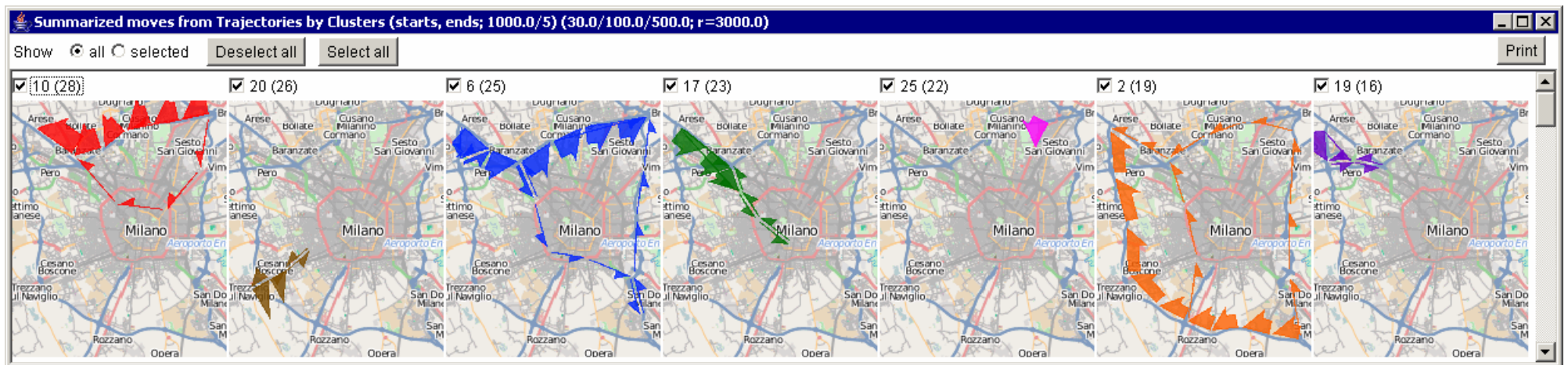
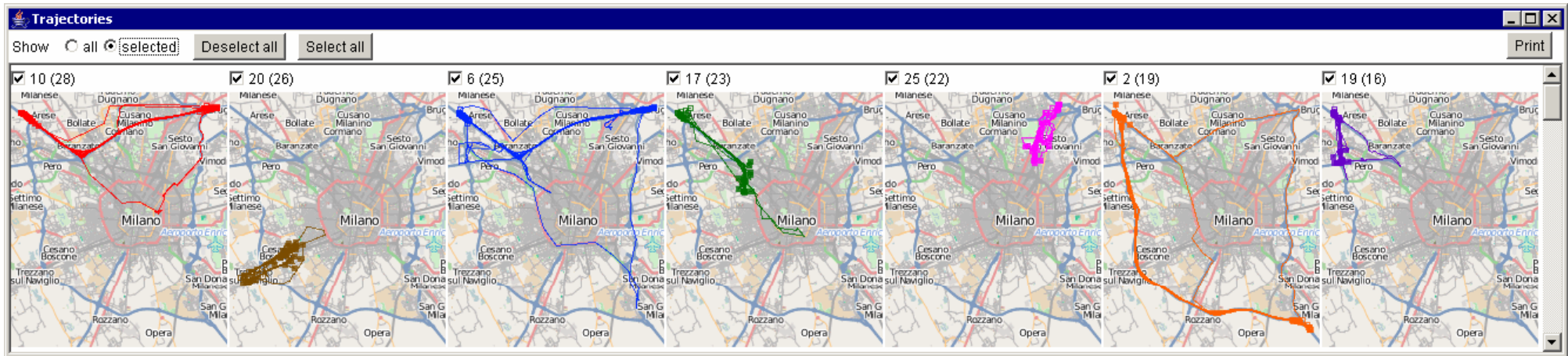
What can be done?

- Clustering techniques allow us to group trajectories by similarity
- Diverse aspects of similarity: origins, destinations, path lengths, durations, route shapes, dynamics
 - a library of different distance functions that can be combined through the progressive clustering approach
- We can examine and compare the groups instead of looking at every trajectory, which is unfeasible
- Groups of trajectories can be shown in a summarised form to reduce display clutter

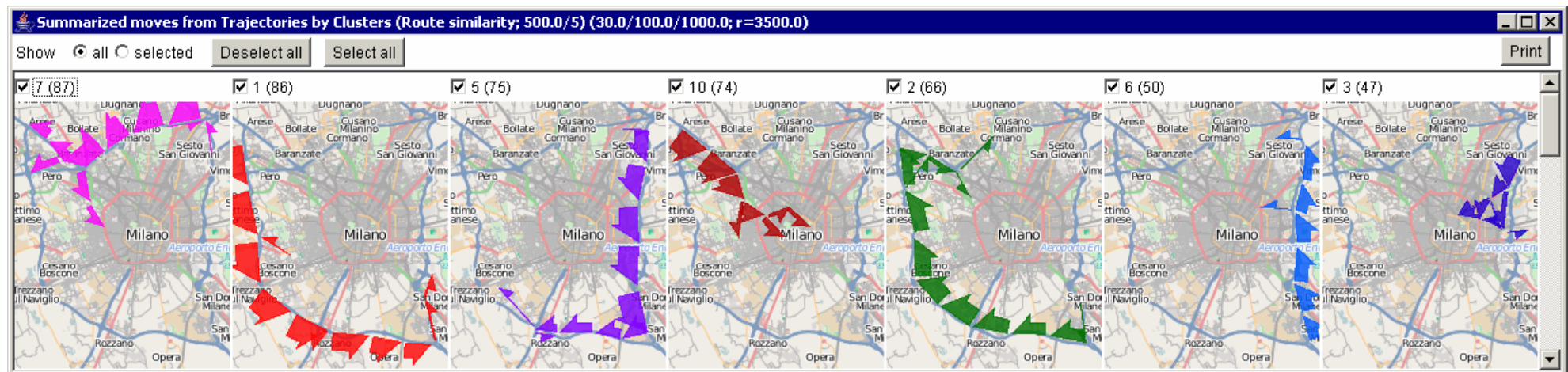
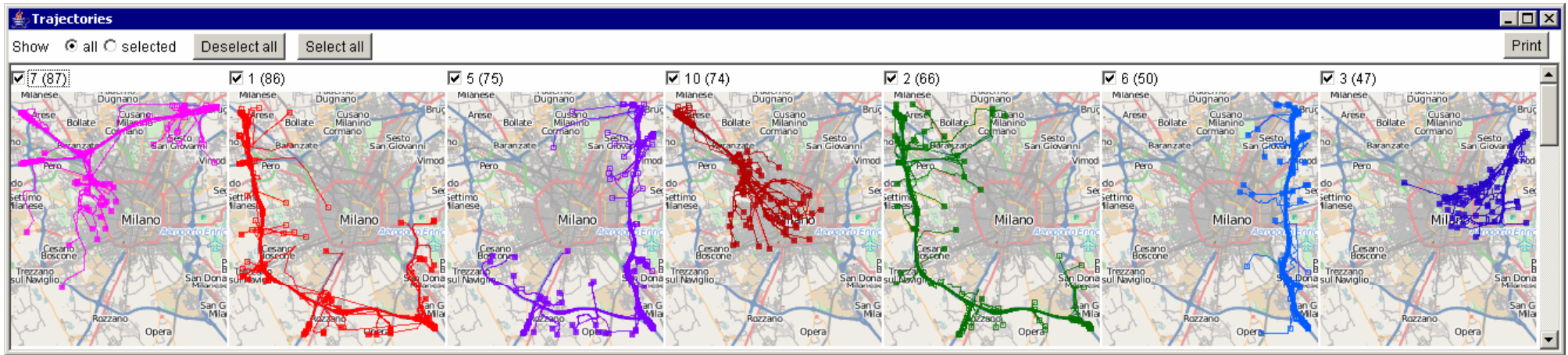
Example: clusters of trajectories by close destinations



Example: clusters of trajectories by common origins and destinations

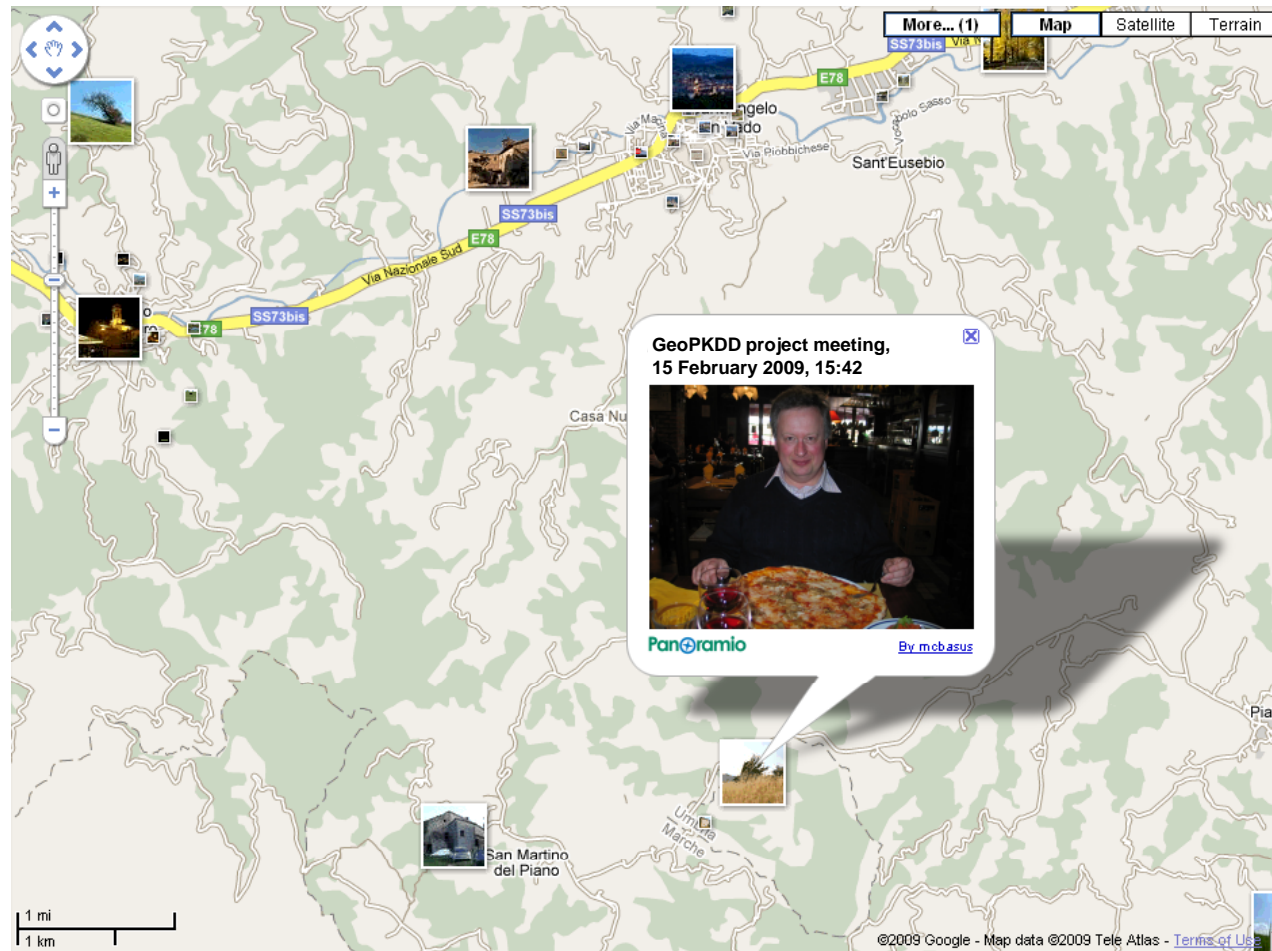


Example: clusters of trajectories by similar routes

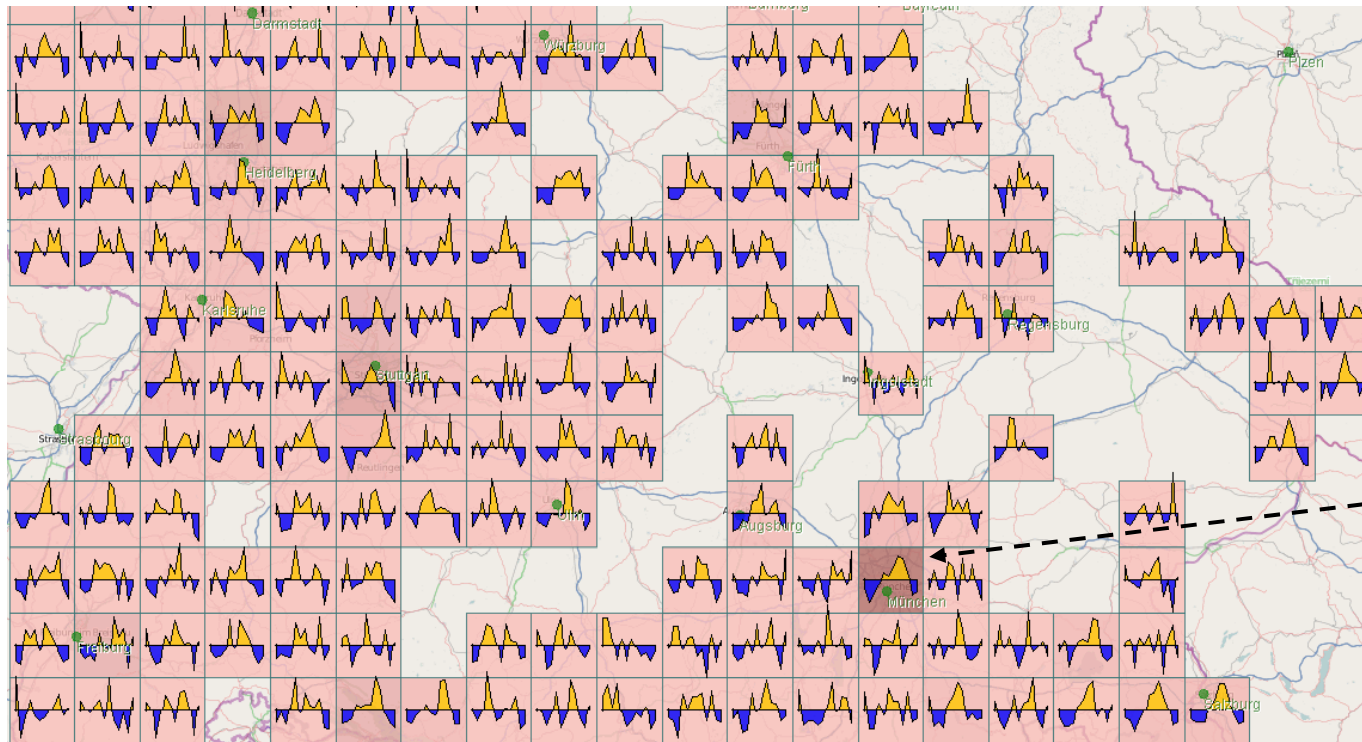


Exploration of Temporally Sparse Movement Data

Example of temporally sparse movement data (discontinuous trajectories): positions of photos in Panoramio or Flickr



Spatio-temporal aggregation of photo taking events



We can observe a seasonal variation of the frequency of taking photos in most of the places.

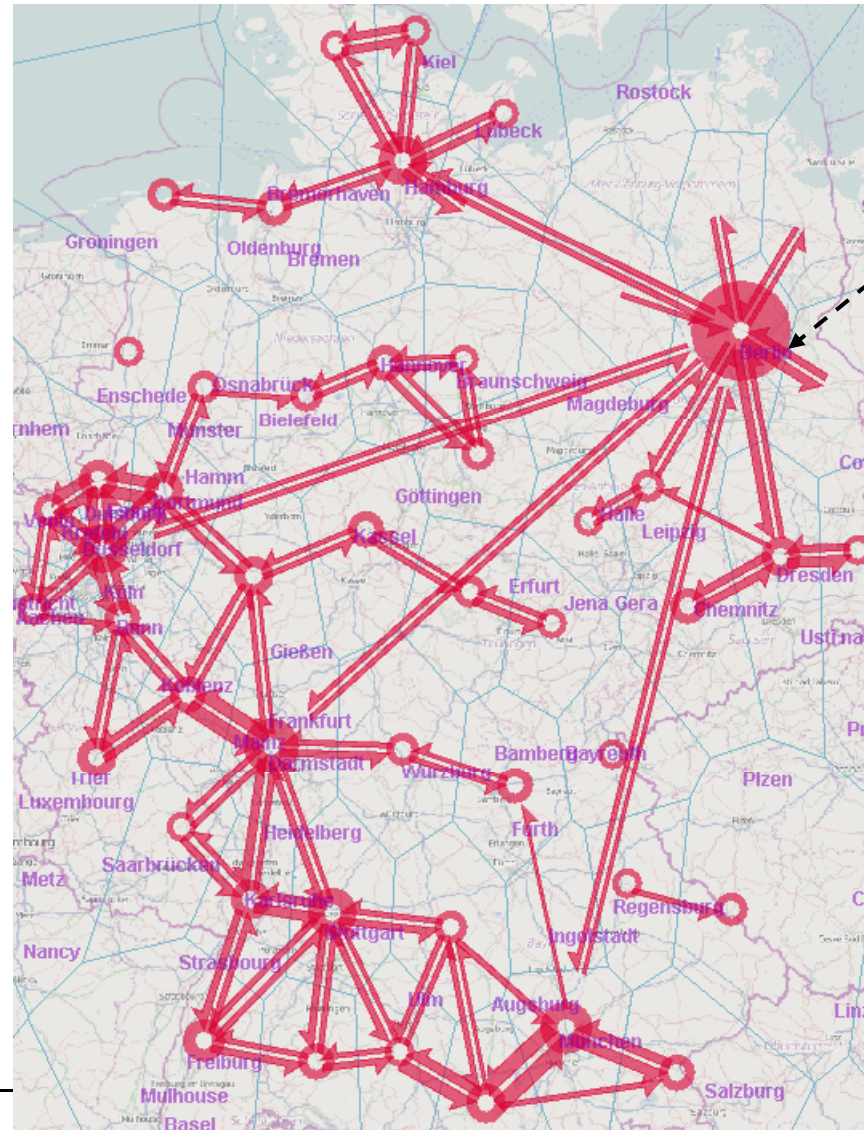
The area of Munich is the most frequently visited place on the south of Germany. The highest frequency of taking photos is in autumn, especially in September.

The data have been aggregated spatially by grid cells and temporally by months, irrespective of the years. The shading portrays the total number of different people who made photos in the cells. The cells visited by less than 100 people are hidden. The diagrams represent the yearly variation of the number of photos taken in the cells. The horizontal axis represents 12 months from January to December. The counts are transformed into normalized differences from the local mean values. The map fragment represents the south of Germany.

Analysis of movements (flows) at different spatial scales

Most movements occur within areas and between neighbouring areas. However, quite frequent are also distant moves between Berlin and the other biggest cities of Germany: Hamburg, Düsseldorf, Frankfurt, and Munich.

Very many trajectories of the photographers are within the area of Berlin



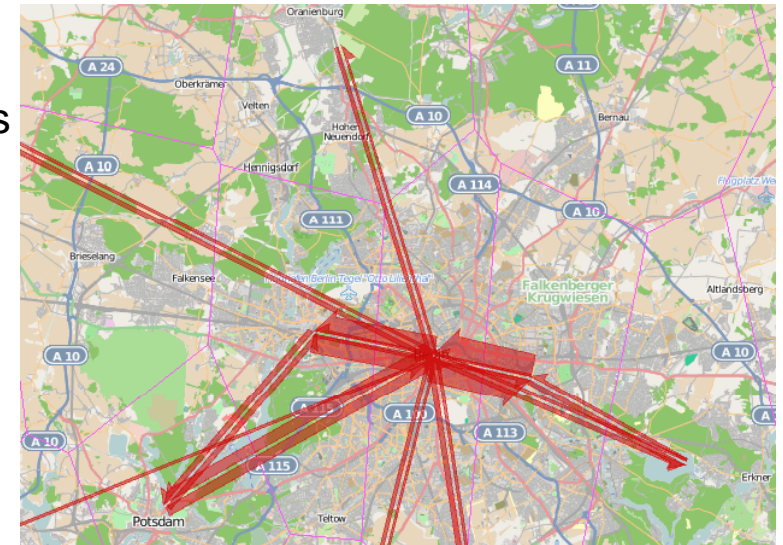
The data have been aggregated using the approach to movement summarisation on the basis of Voronoi polygons. The aggregate moves have been interactively filtered so that only the moves with the frequency 100 and more are visible.

Analysis of movements at different spatial scales

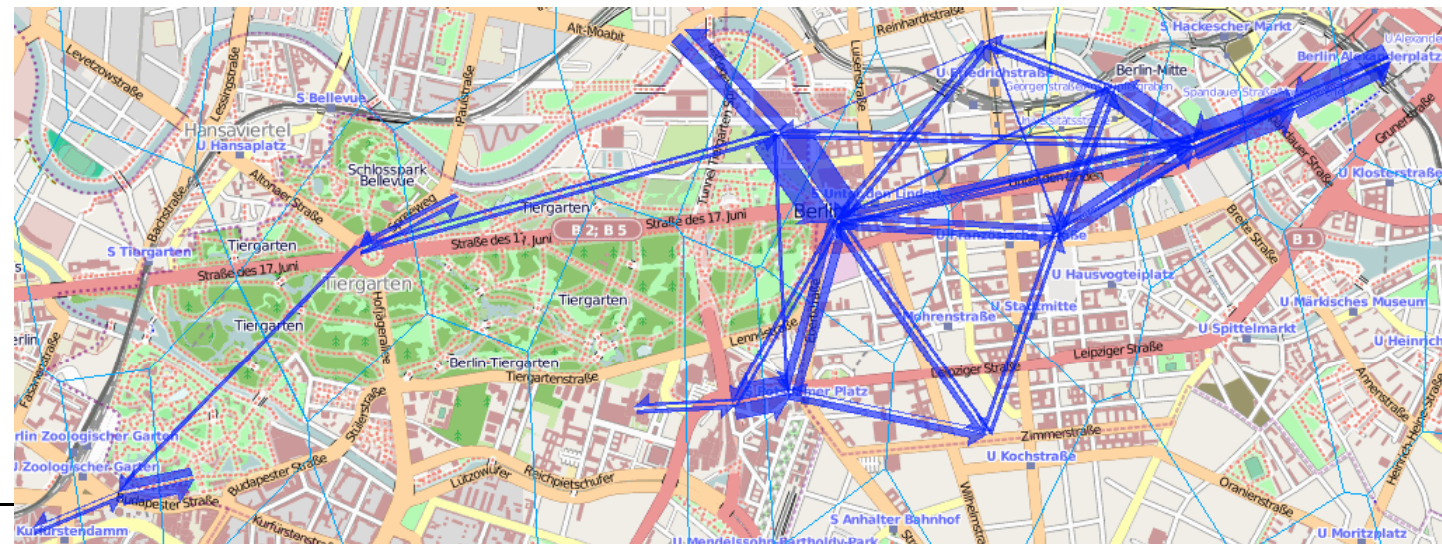
Movements in the western part of Germany



Movements in area of Berlin and surroundings



Movements in Berlin



Discovering Bits of Place Histories from People's Activity Traces

Not in living memory...

- Do you know the recent history of your place?
- Do you remember what happened in your place, for example, in March 2007?
- When did something important happen in your place (if any)?

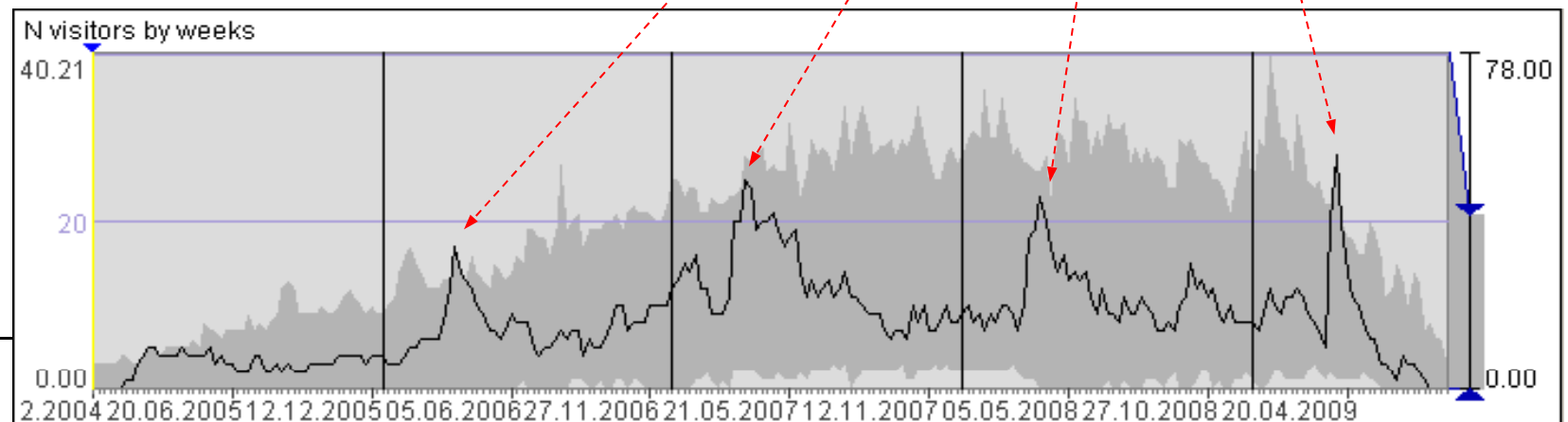
Reconstructing bits of history?

- If your memory is not perfect (mine is not), records of important events may help to reconstruct bits of history
 - Goodchild (2007) – citizens as sensors collecting valuable geographic information
- Publicly collected data contain evidences of events
 - flickr, twitter ...
- Databases of enterprises may be used for this purpose
 - Mobile phone companies



General idea

- At some time moments/periods more people than usual leave their traces in a place.
- This may be an indication of interesting events



Suitable data?

- Activity records are already in databases in structured form, a lot of data!
 - Person_ID, longitude, latitude, time, activity attributes
- Places are areas rather than points
- Definition of a place depends on the intended spatial scale of the analysis
 - The same is valid for time
- The amount of data does not fit to RAM and does not allow purely visual analysis

Methodology

- Suite of visual analytics tools for detecting events
 - Division of territory at the intended scale of analysis
 - Aggregation of data into time series for areas
 - Detecting events in time series, checking t-correlation
 - Interactive visual interpretation of the results
 - Of special interest (*why human judgment is needed*):
 - Periodicity in mostly non-periodic data
 - Non-periodicity in mostly periodic data
 - Any other regularity / irregularity
- Possibly, repeat analysis at another scale

Data examples

- Positions and timing of starts and ends of 2,956,738 phone calls in Milan (Italy) during 9 days
 - Provided by WIND
 - Stationary calls Vs. calls on move
 - Estimation of speed

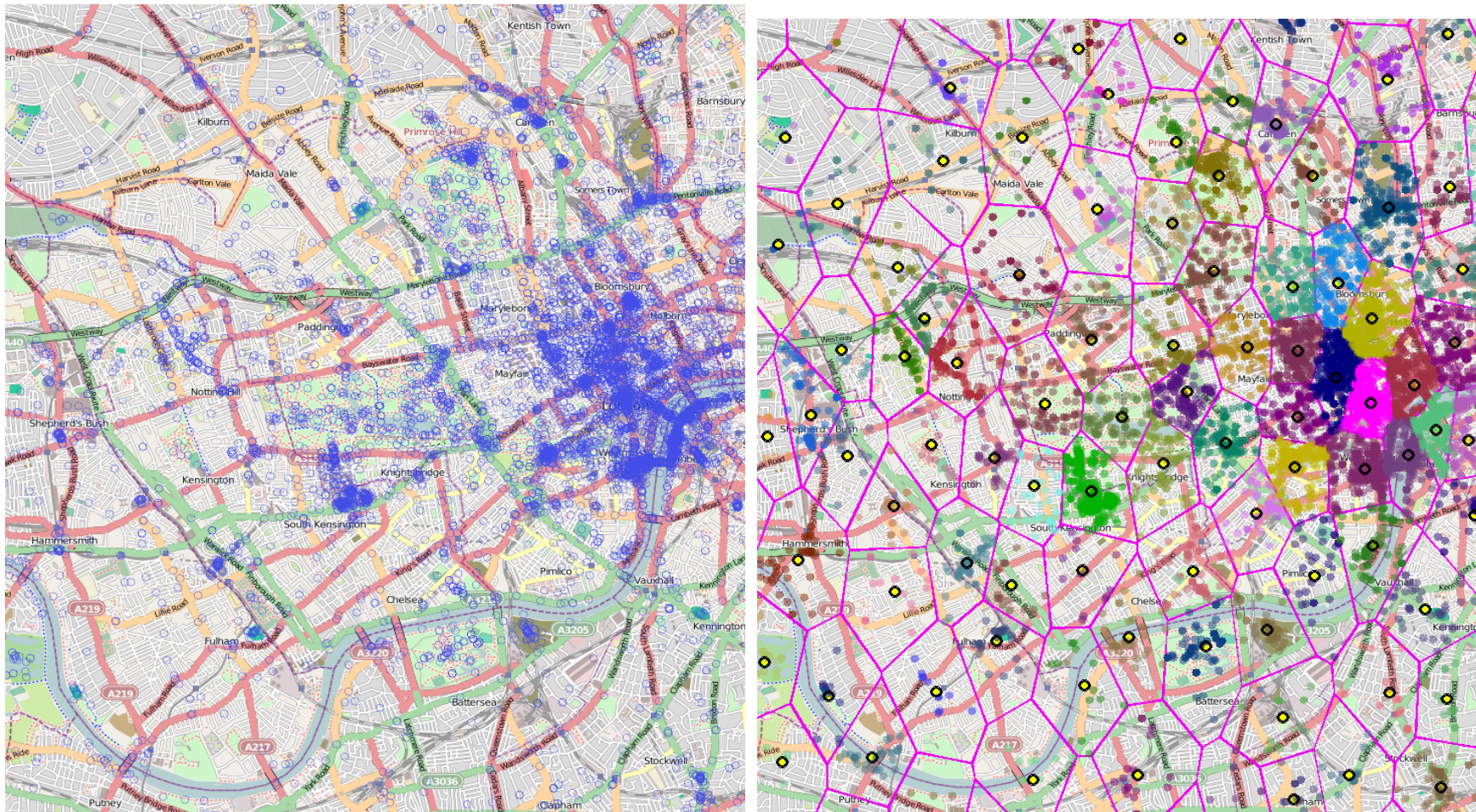


- Positions, time stamps, and titles of 8,686,034 photos in UK and Ireland during 5 years
 - Extracted from flickr.com by S.Kisilevich (Univ.Konstanz)



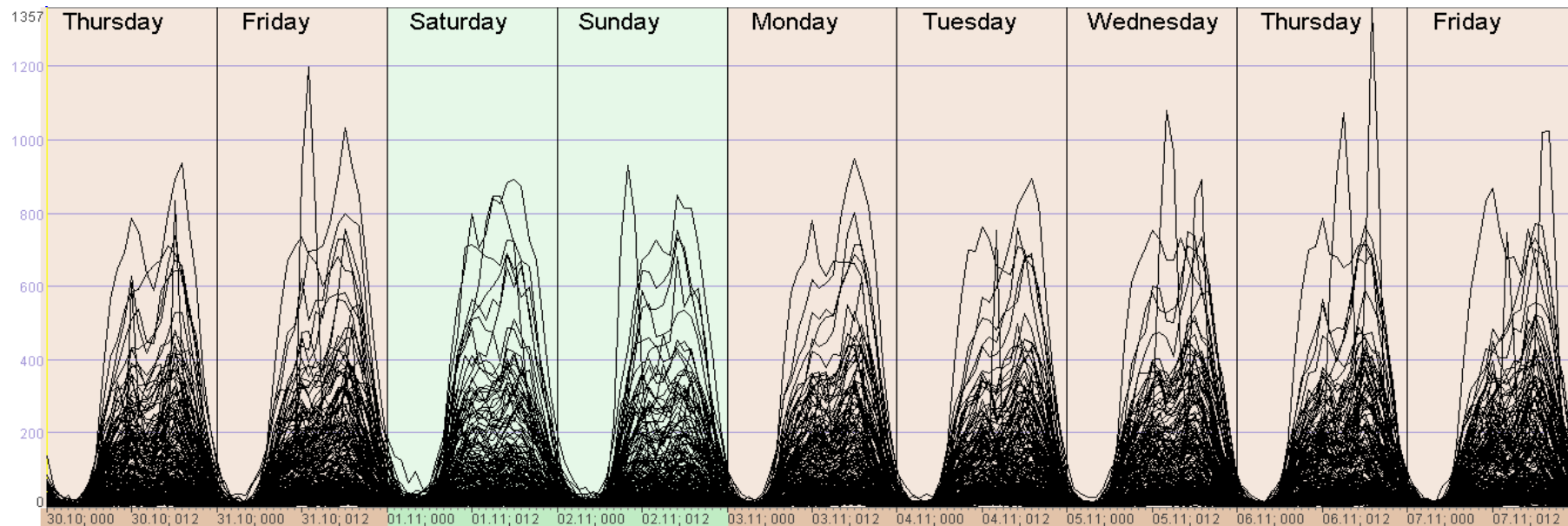
How we do it (1)

- Territory tessellation using space-bounded clustering of a sample



How we do it (2)

- **Spatio-temporal aggregation**
 - We use Oracle database
 - For given tessellation and selected time intervals, the system computes
 1. Number of different people who visited the areas in each interval
 2. Count of activities (e.g. calls, photos) that occurred in the areas in each interval



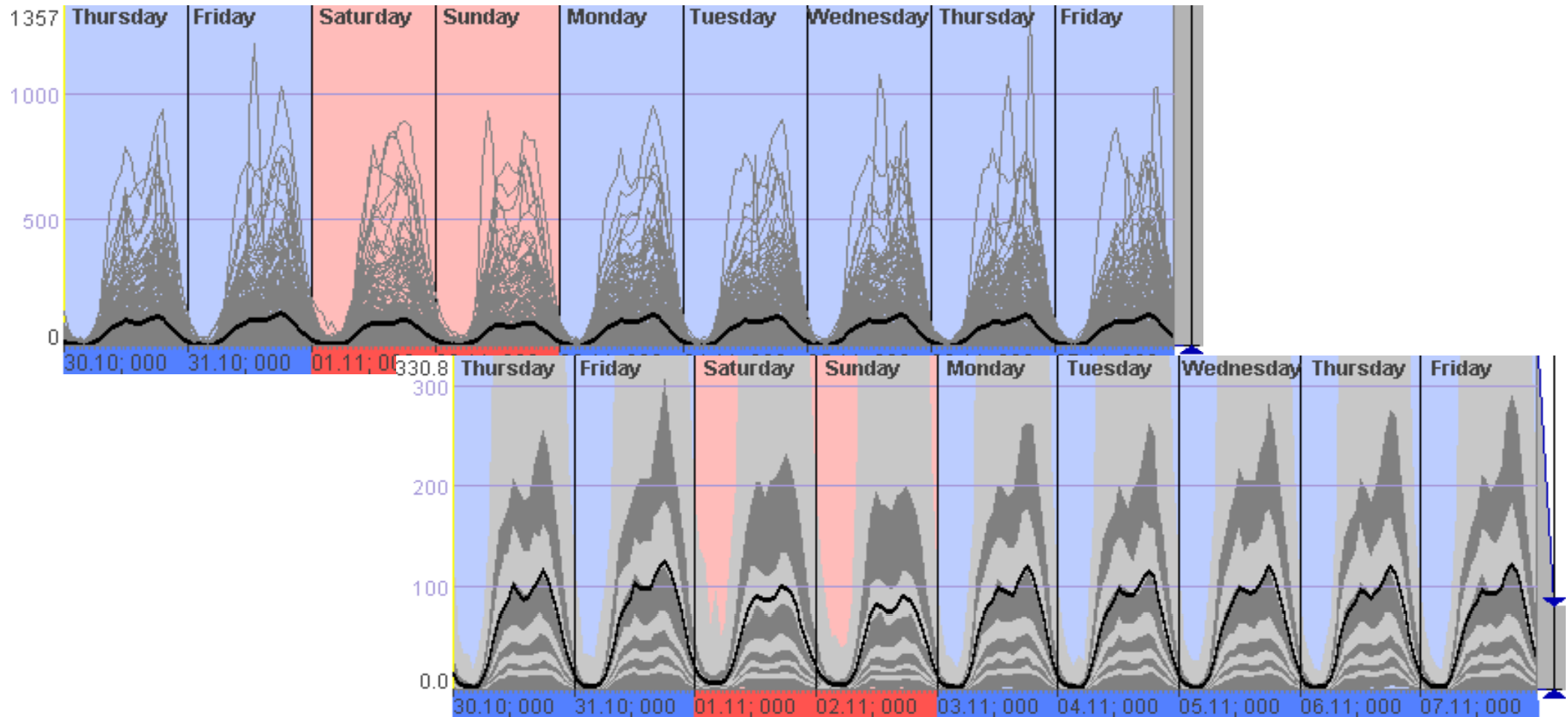
How we do it (3)

- **Time series analysis by statistical procedures**
 1. Periodicity (temporal correlation) detection:
max of the circular cross-correlation function of a time series and a synthetic test pattern generated for a chosen period
 2. Peak event detection:
identifying sudden increase (peaks) or decrease (pits) of values within the given time window;
aggregation of event attributes

Details in the IEEE VAST 2010 paper

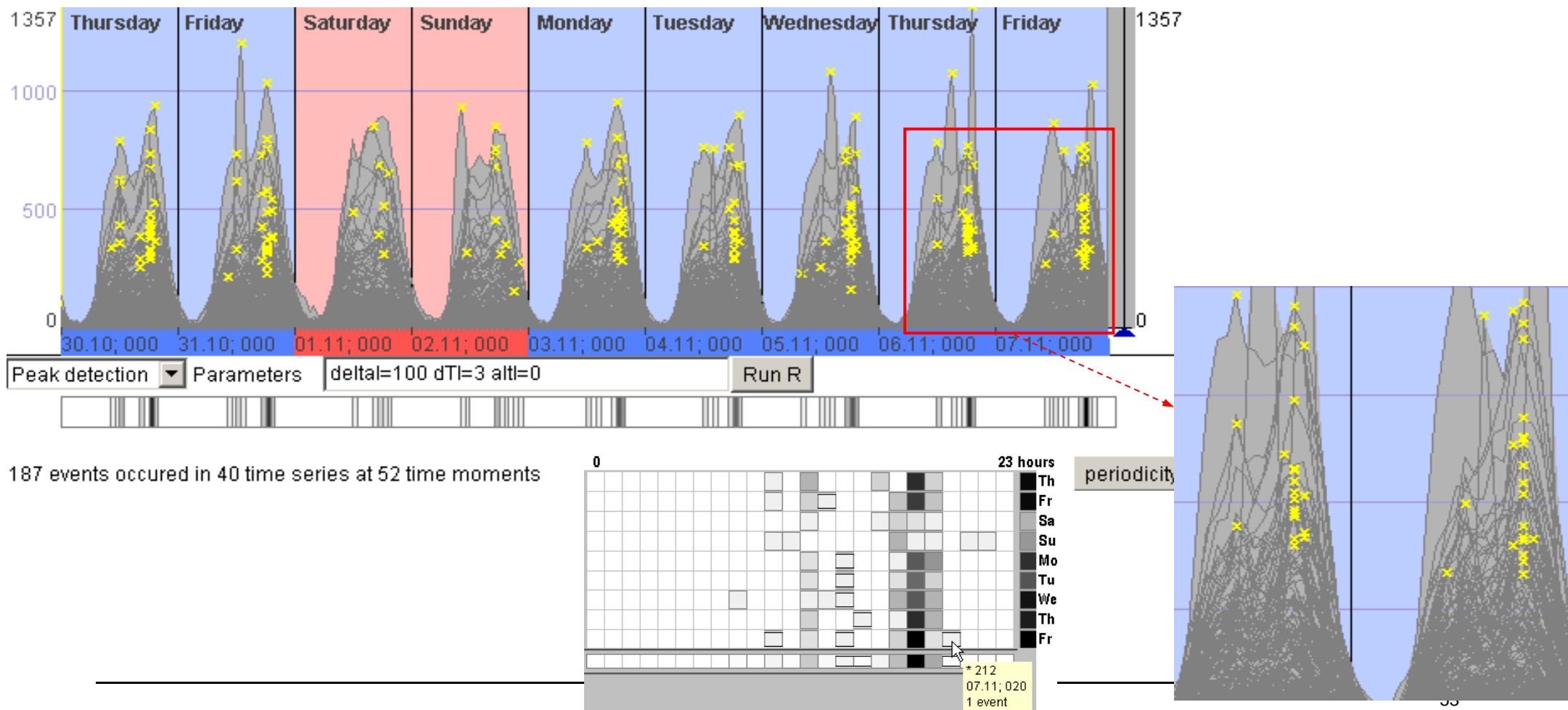
How we do it (4a)

- Interactive visual displays: time graph



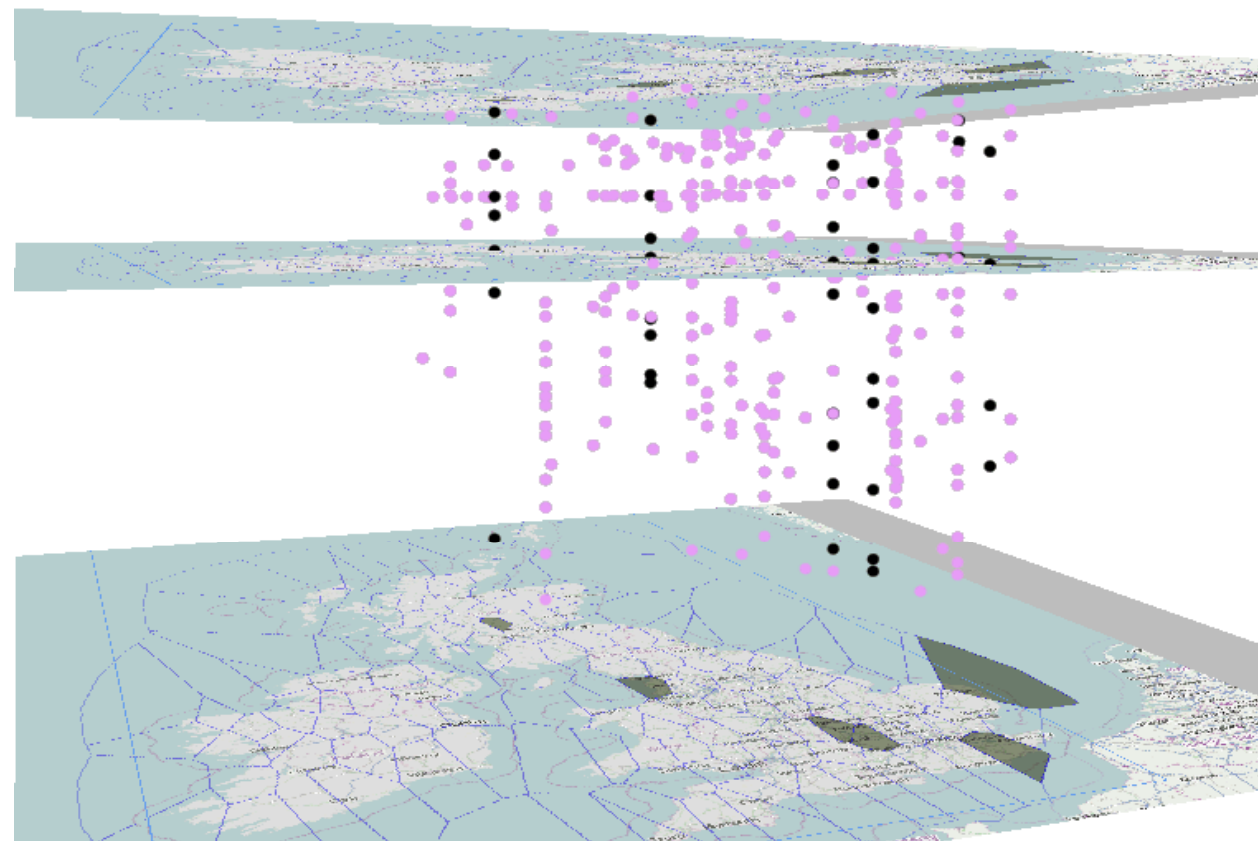
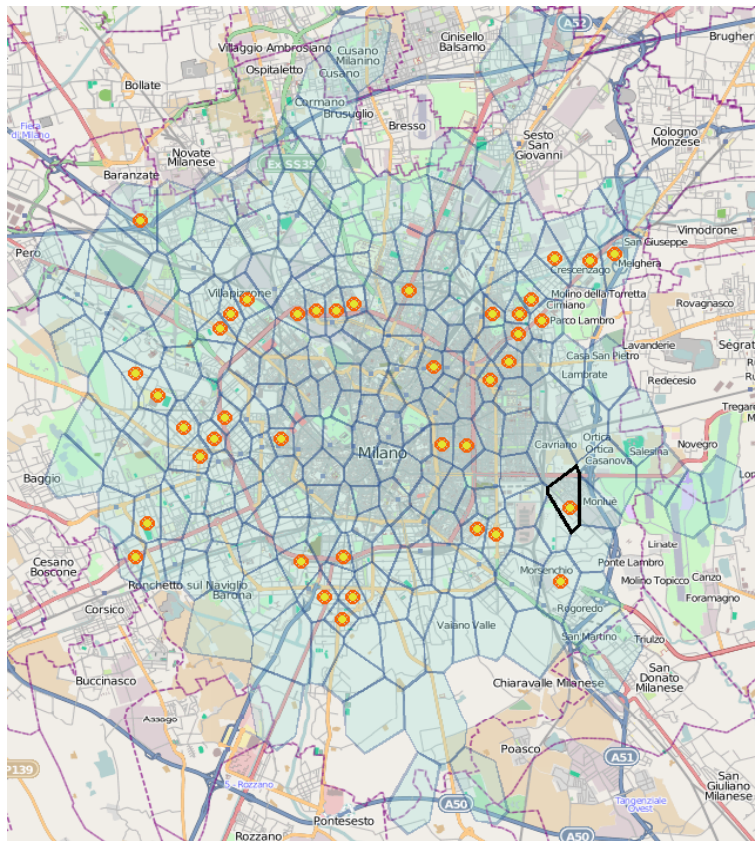
How we do it (4b)

- Interactive visual displays: event detection, event bar



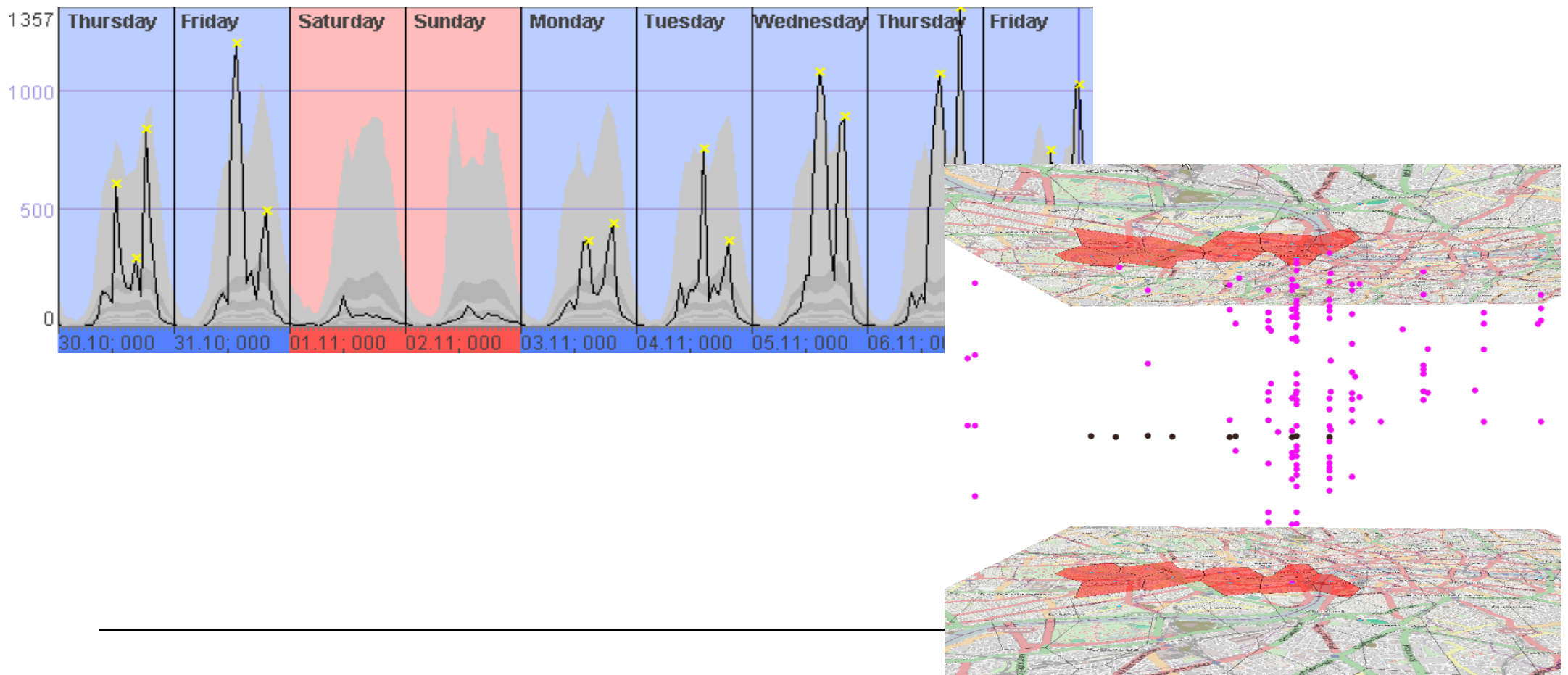
How we do it (4c)

- Interactive visual displays: map & space-time cube



How we do it (4d)

- Interactive visual displays: coordinated views
 - Filtering & highlighting by place, time, attributes



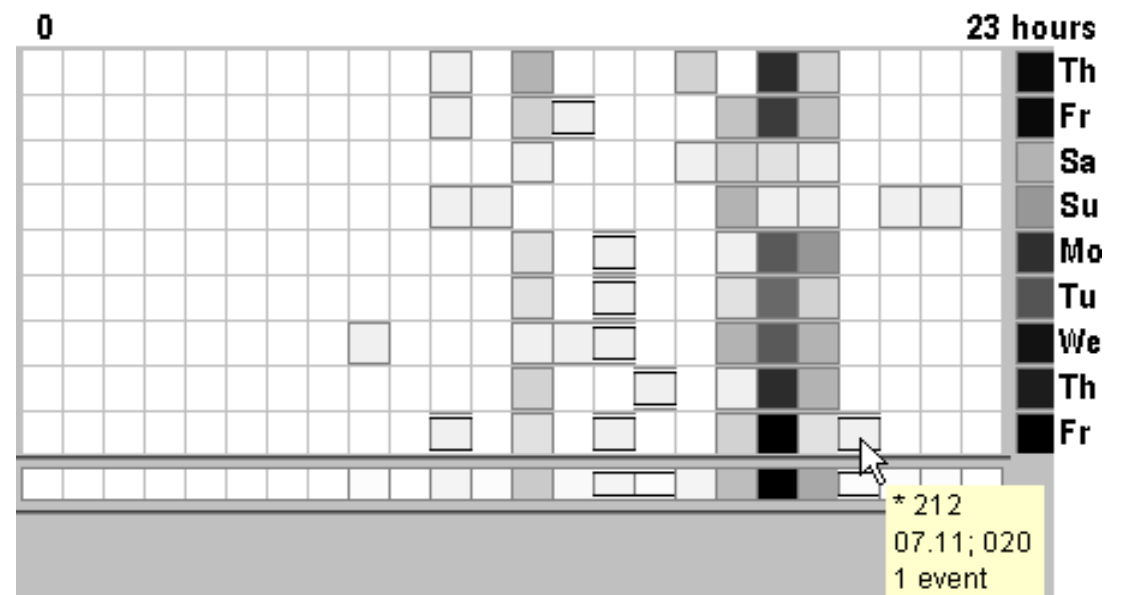
Case study 1: phone calls

- Positions and timing of starts and ends of 2,956,738 phone calls in Milan (Italy) during 9 days
 - Provided by WIND
 - Stationary calls Vs. calls on move
 - Estimation of speed



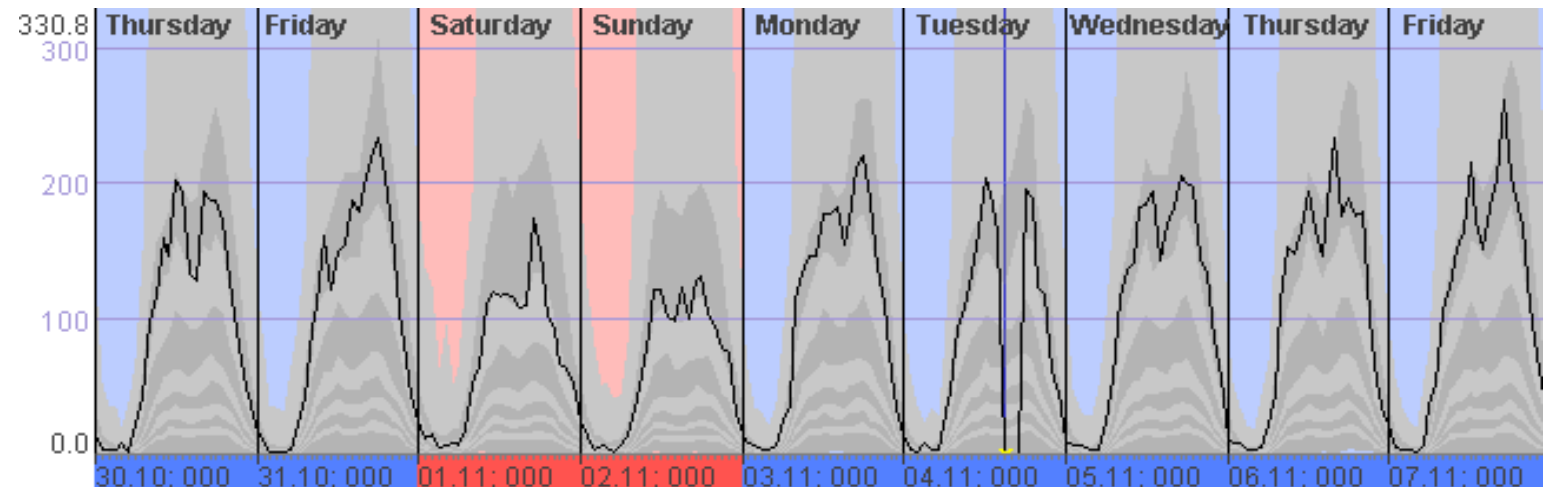
Findings: when peaks happen

- Peaks of calls happen at noon and in the evening, more at working days
 - Noon calls are mostly stationary (lunch breaks?)
 - Evening calls are mostly on the move
 - *“I am coming home, cook the pasta!”*

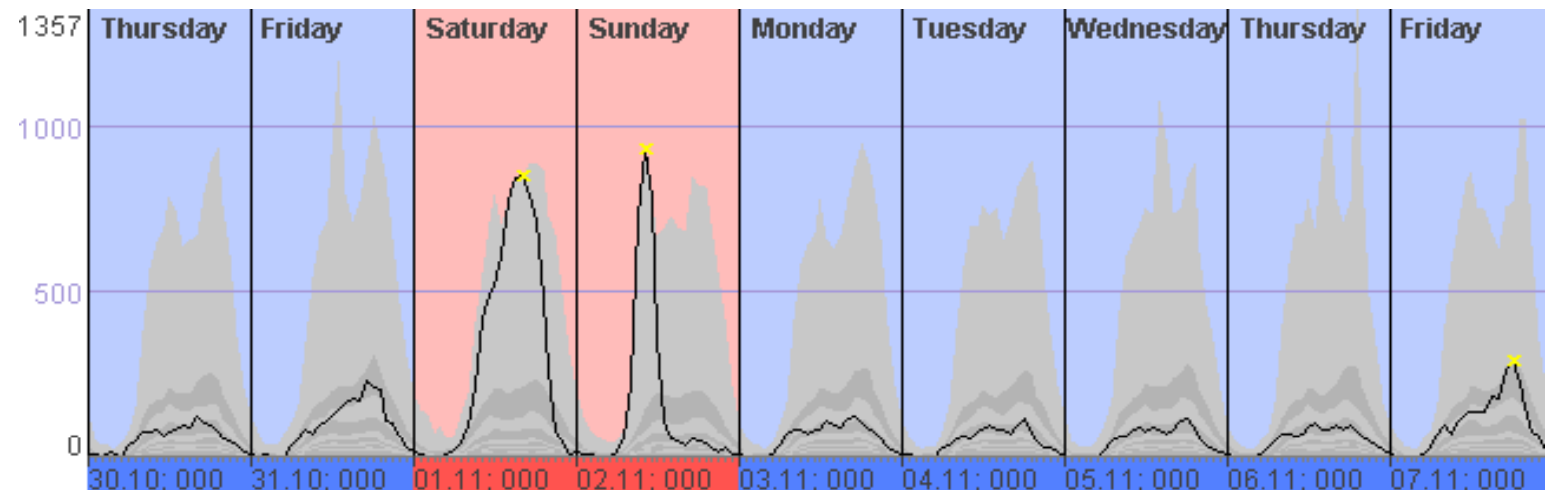


Findings: non-periodic peaks & pits

- Close to city center
(network maintenance)

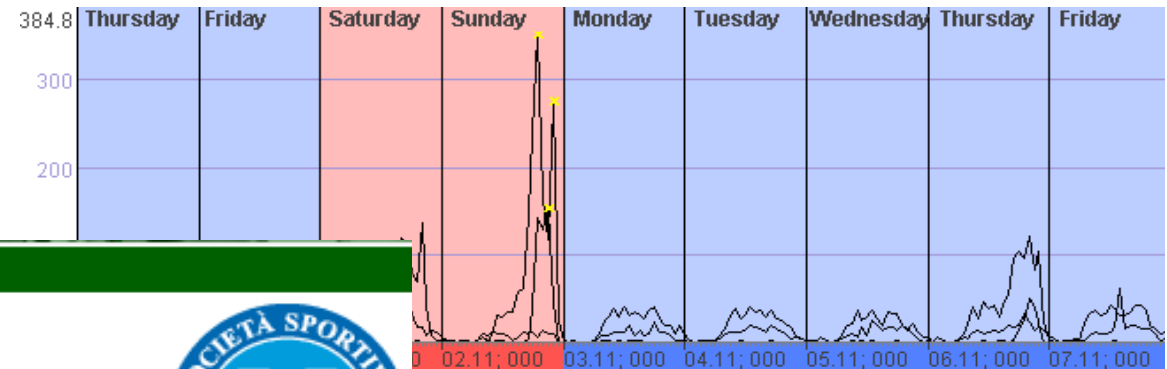


- Parking on North-East
(flea market)



Analysis at a different temporal scale

- Irregular peaks



SPIELSTATISTIK AC MAILAND - SSC NEAPEL 1:0 (0:0)



AC MAILAND - SSC NEAPEL

1:0 (0:0)

So 02.11.2008, 20:30 Uhr, Giuseppe Meazza, Mailand

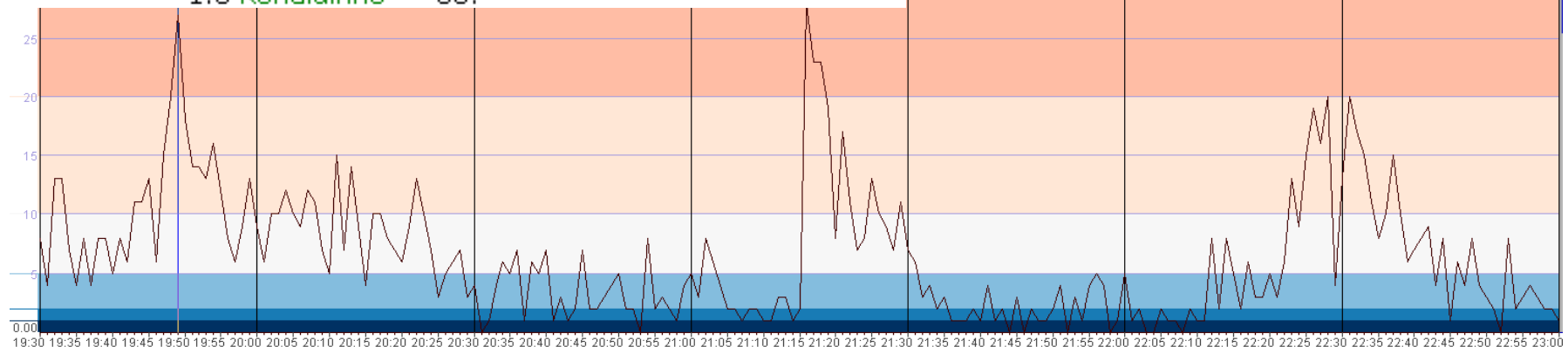
Serie A 2008/2009, 10. Spieltag

51.600 Zuschauer - Schiedsrichter: Gianluca Rocchi (Italien)



Tore:

1:0 Ronaldinho 86.



Fraunhofer
1st half



2nd half

Case study 2: flickr.com photos

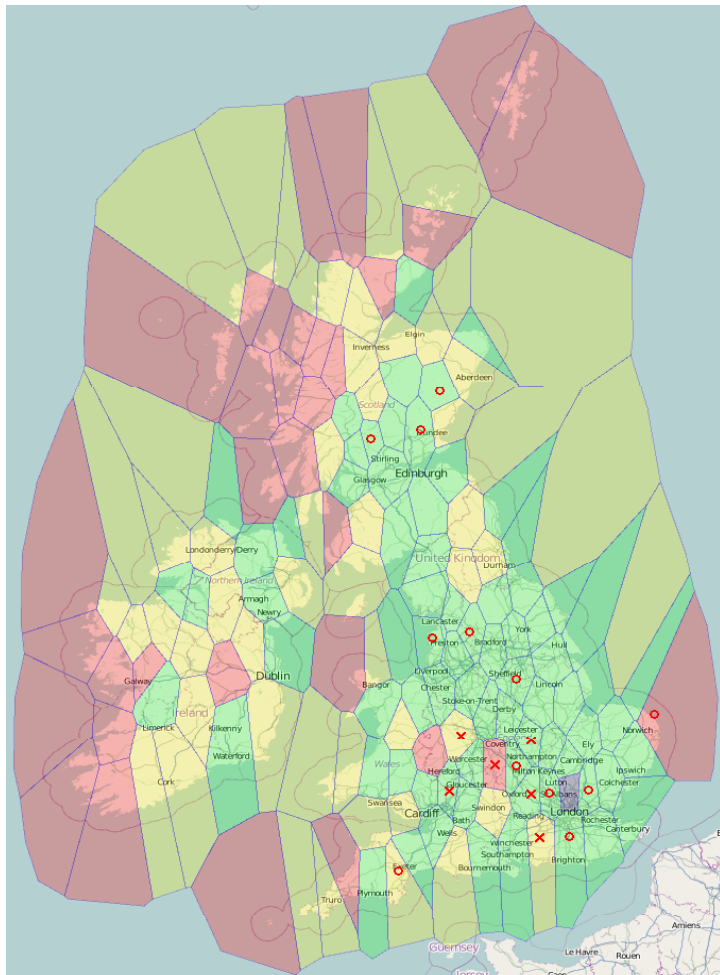
- Positions, time stamps, and titles of 8,686,034 photos in UK and Ireland during 5 years
 - Extracted from flickr.com



General patterns

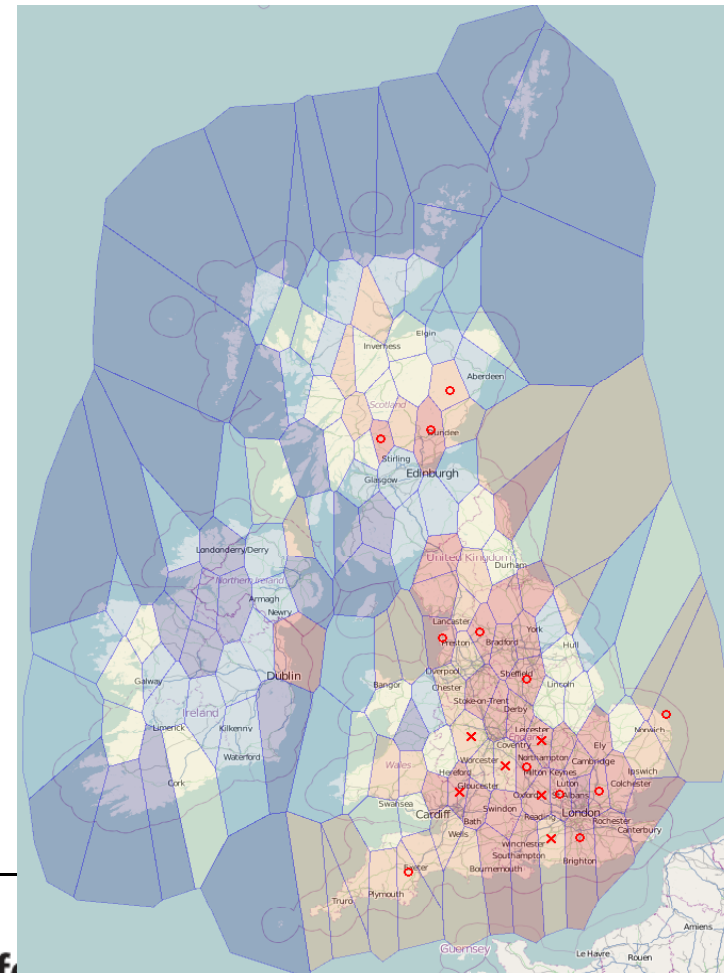
-  Non-periodic events
-  Periodic events

- Periodicity of time series



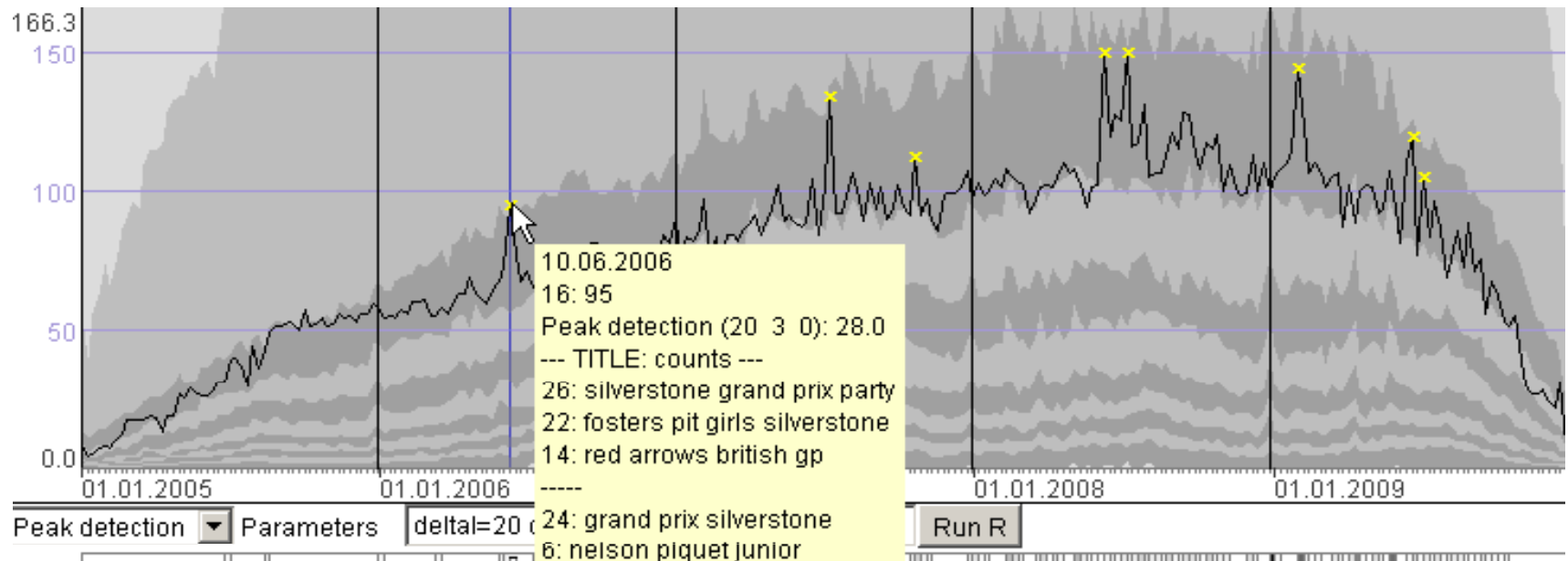
Green=low; Red=high

- Counts of photos



Blue=low; Red=high

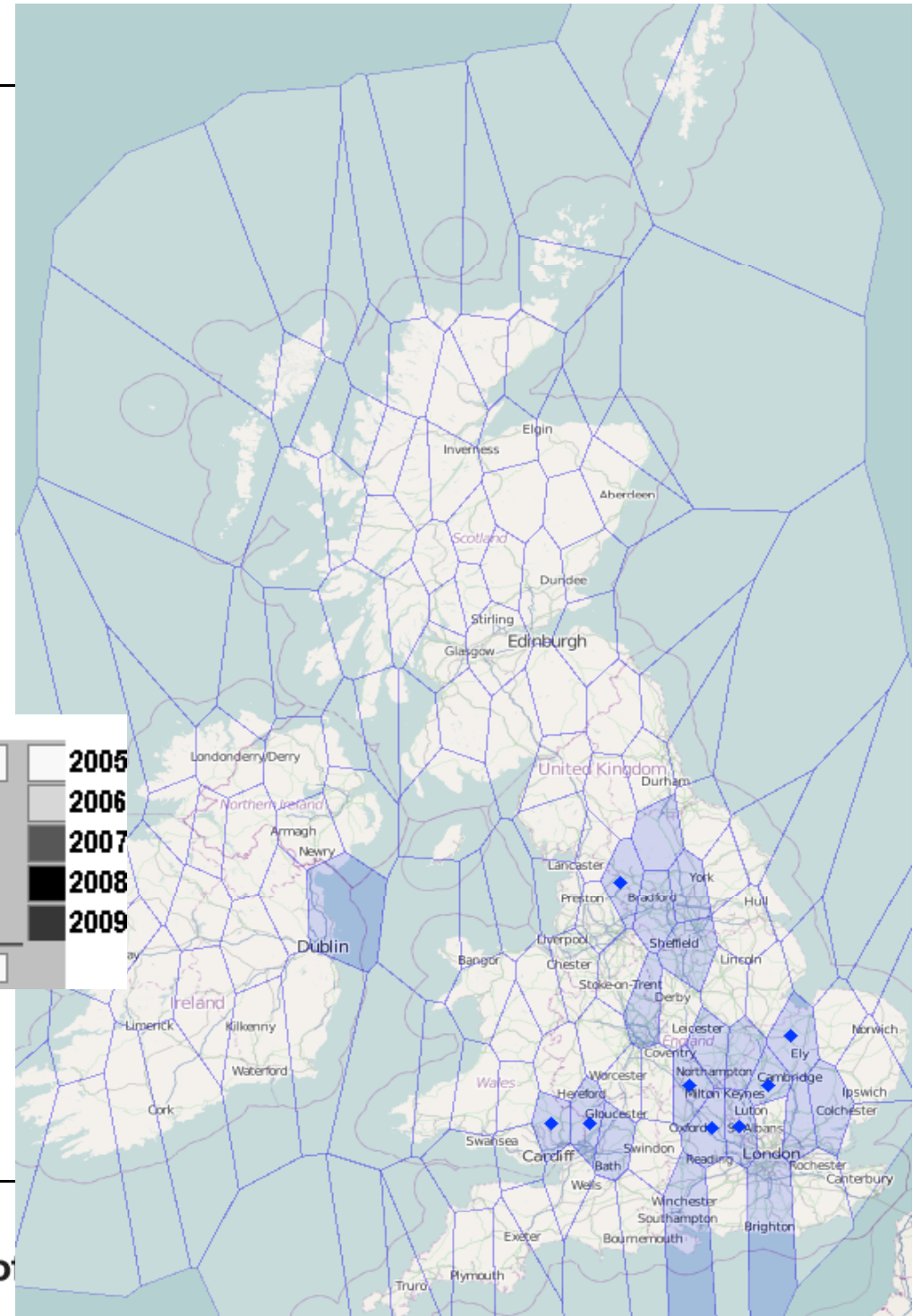
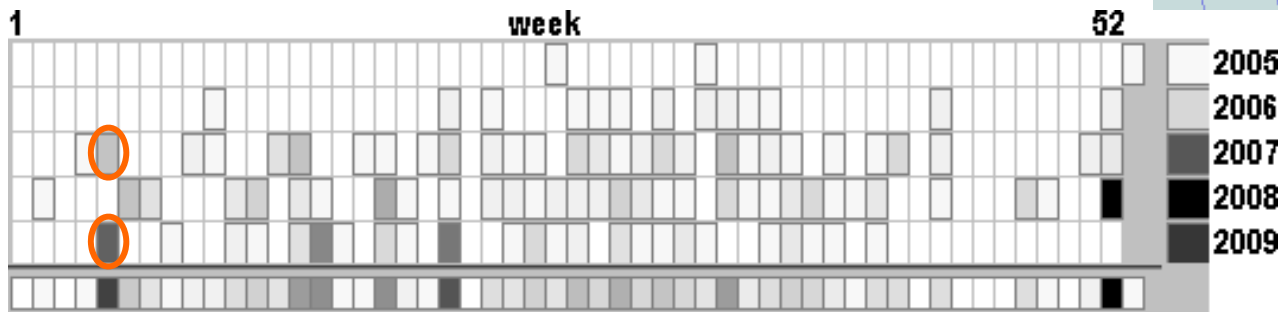
Examples of periodic events



- Silverstone Grand Prix
 - Royal International Air Tattoo
 - Glastonbury festival...
- Interpretation through summarization of photo titles

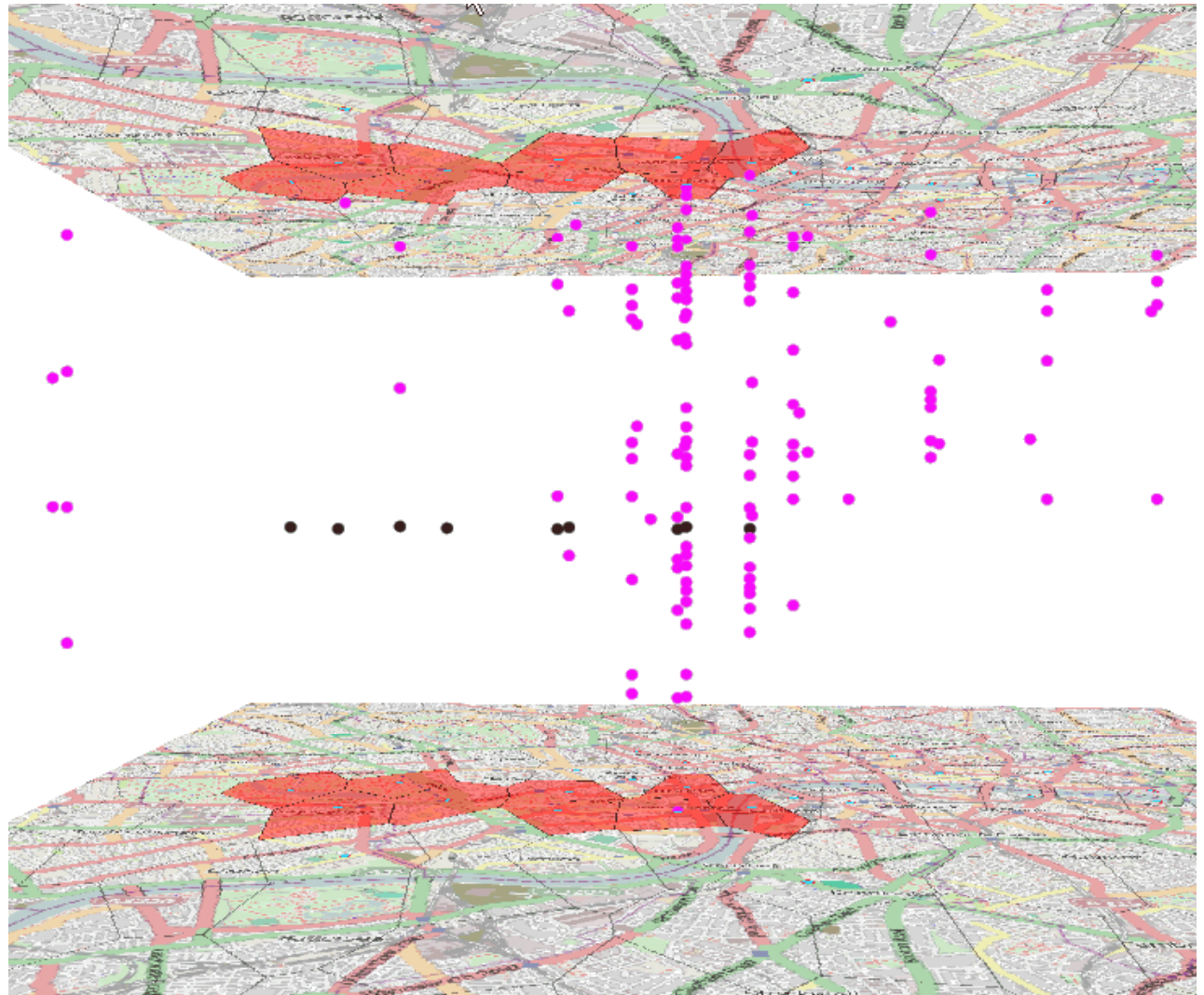
Irregular peaks

- Peaks in Feb 2009 and Feb 2007 with frequent “snow” in photo tags: exceptional snowfalls?



Analysis at a different spatial scale

- Tessellation of London area with finer resolution
- Prologue of “Tour de France”
London, July 2007



Summary

- Efficient data analysis
 - time for analyzing a previously unknown dataset vary from 30 to 60 minutes
- Flexible workflows.
User can arbitrary combine:
 - what → where + when
 - when → what + where
 - where → what + when
- Major issues for history reconstruction:
 - Spatial, temporal, and population coverage of the available data limits the applicability
 - Careful selection of suitable scales in space and time is required

Notes about privacy

- Availability of important but challenging spatio-temporal data sets (geospatial imagery, sensors, GPS and movement tracking, geo-tagging, flickr, wiki, ...)
 - Sometimes {unintentionally} breaking personal privacy

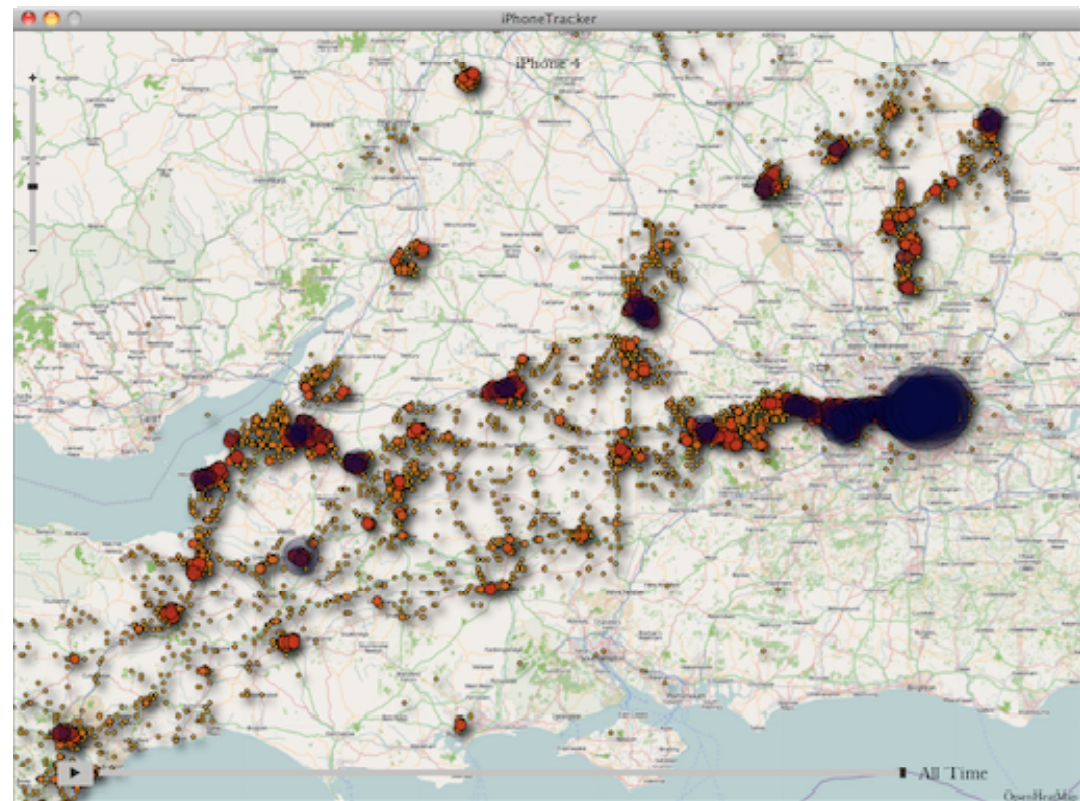
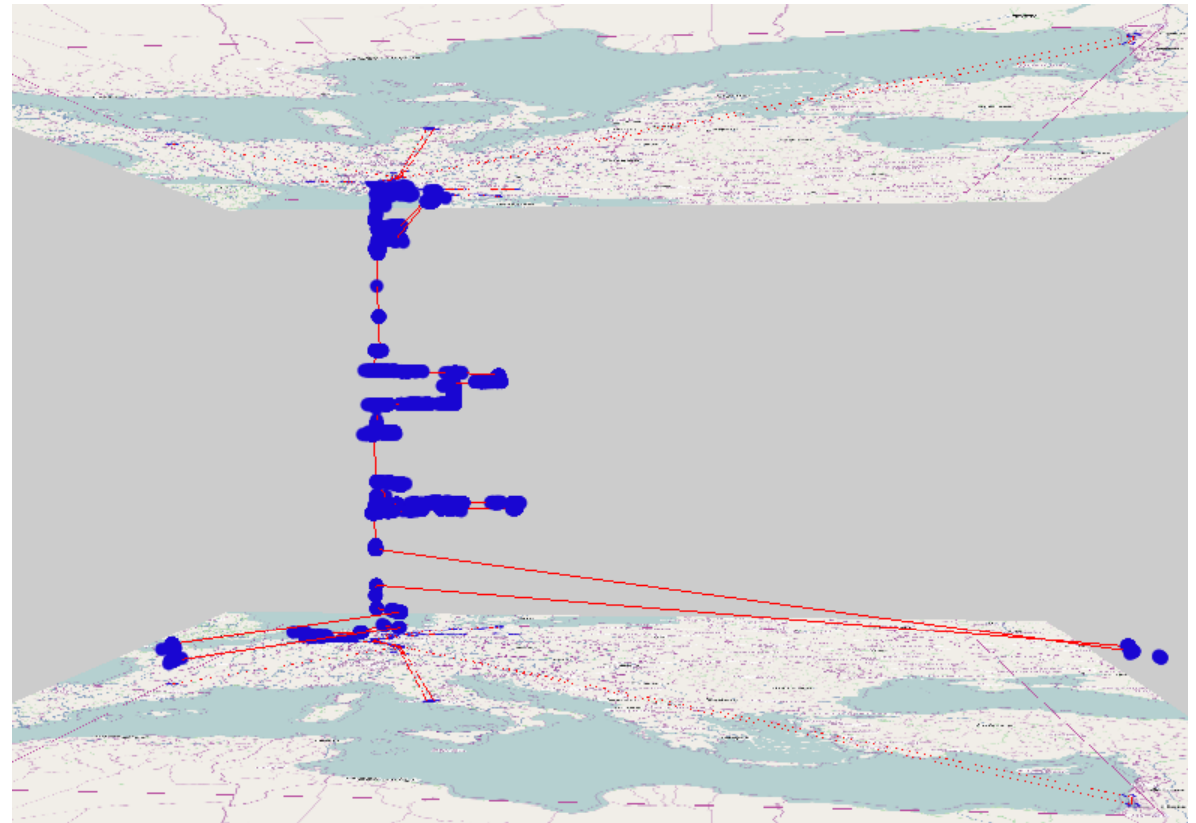
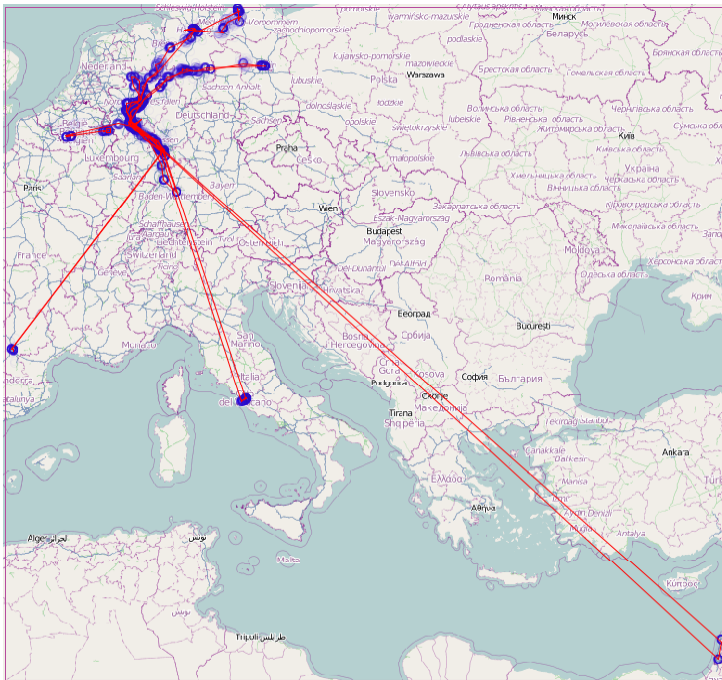


Image courtesy of <http://petewarden.github.com/iPhoneTracker/>

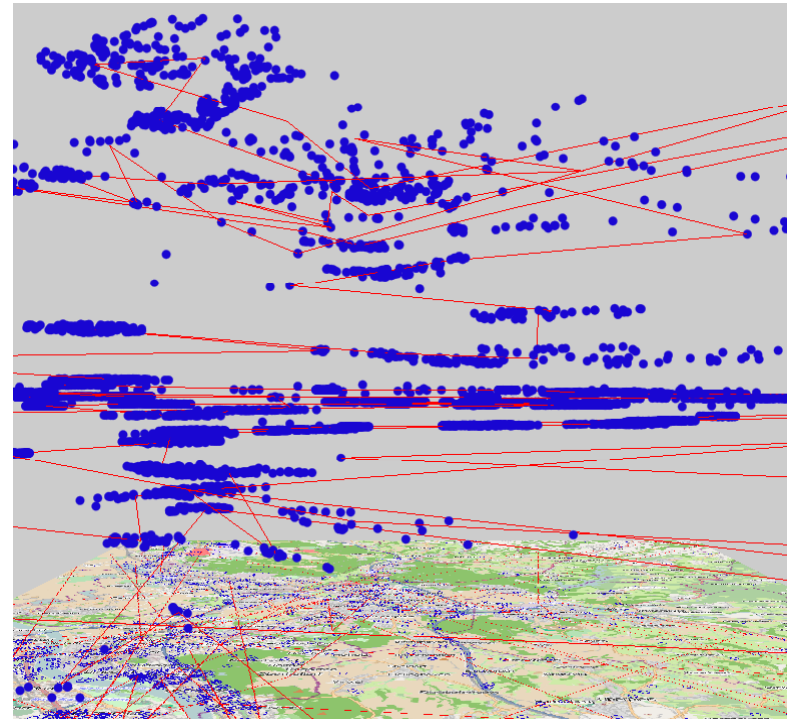
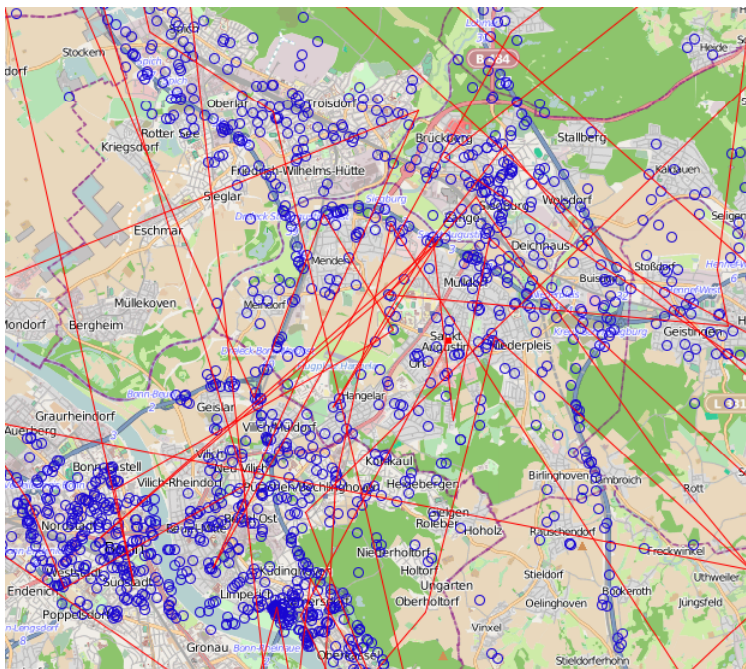
Are iPhone data really dangerous for {my} personal privacy?

- At large scale: definitely YES – show which cities have been visited and when



Are iPhone data really dangerous for {my} personal privacy?

- At small scale: to some extent only – some real mobility patterns are not visible



Conclusion: movement data

- Movement data link together space, time, and objects positioned in space and time.
- These linkages hold valuable information about
 - the moving objects
 - properties of space
 - properties of time
 - events and processes occurring in space and time
- To uncover various types of information hidden in movement data, the data must be considered from different perspectives, including
 - trajectories and movement behaviours of the objects
 - temporal variation of movement characteristics in places
 - temporal sequence of spatial situations (e.g. traffic situations)
 - flows (connections) among places
 - interactions among the objects
 - interactions between the objects and the environment (context)

Conclusion: Visual Analytics

- Visual Analytics aims at helping people in
 - distilling the relevant nuggets of information from large amounts of data
 - understanding the connections among relevant information
 - gaining insight from data
- Visual Analytics focuses on the division of labour between humans and machines
 - Goal: computational power amplifies human perceptual and cognitive capabilities
- Visual representations are the most effective means to convey information to human's mind and prompt human cognition and reasoning
- Visual Analytics combines interactive visualizations with automated analysis techniques such as
 - database processing
 - data mining algorithms
 - statistics
 - geographical analysis methods

What was NOT presented today

- Advanced visualizations
- Interactive and iterative clustering of trajectories using various similarity measures, out-of-memory clustering of very large trajectory collections
- Extraction and analysis of data derived from trajectories
 - Events
 - Spatial time series
- Analysis of movement data in context
- Integration with moving object database Secondo (R.Gueting) for advanced processing and pattern extraction



<http://geoanalytics.net>