



# **Discovering Bits of Place Histories from People's Activity Traces**

Gennady Andrienko, Natalia Andrienko,  
Martin Mladenov, Michael Mock, Christian Pölitz

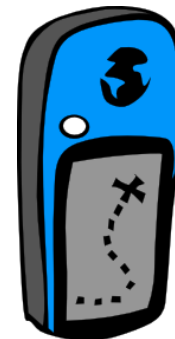
# Not in living memory...

- 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)?



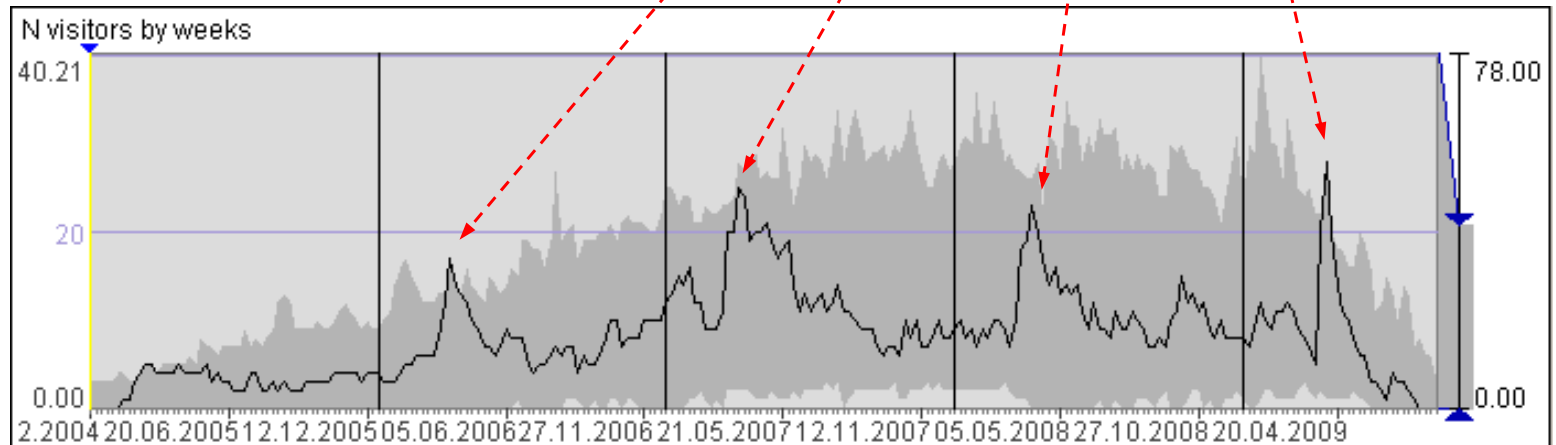
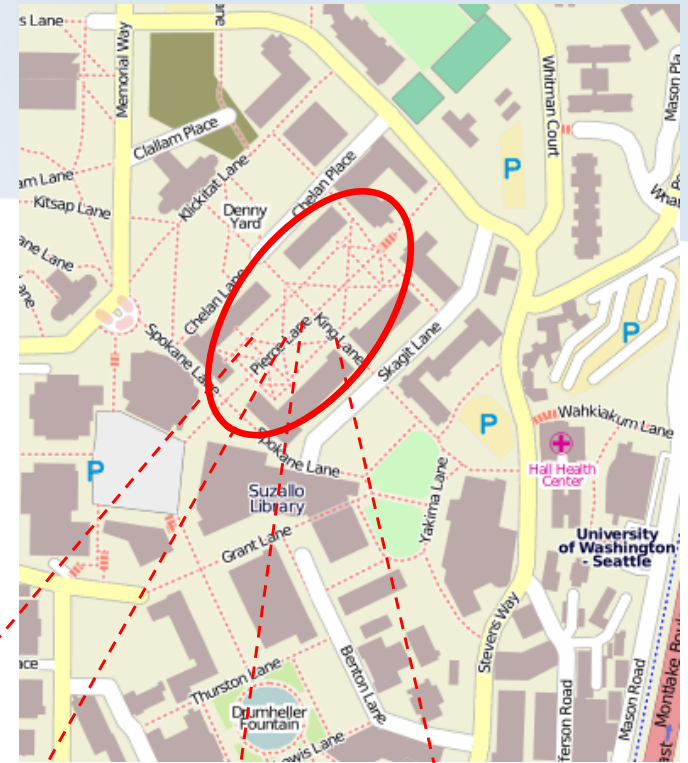
# Reconstructing bits of history?

- 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 (sorry, no InfoVis)



# 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



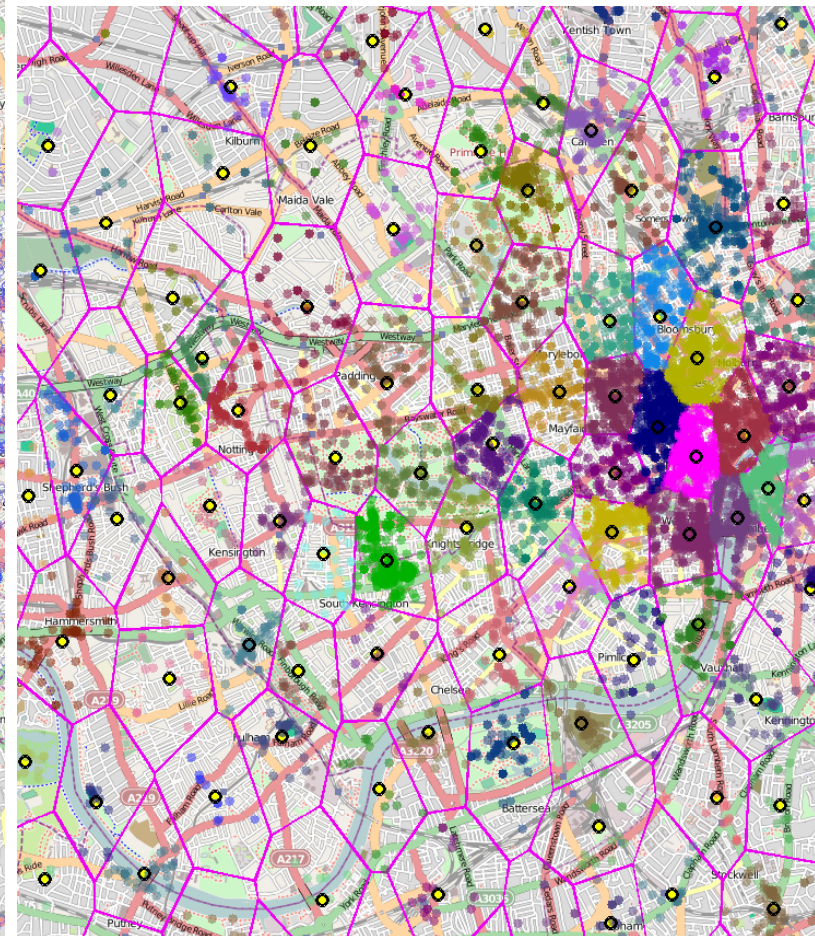
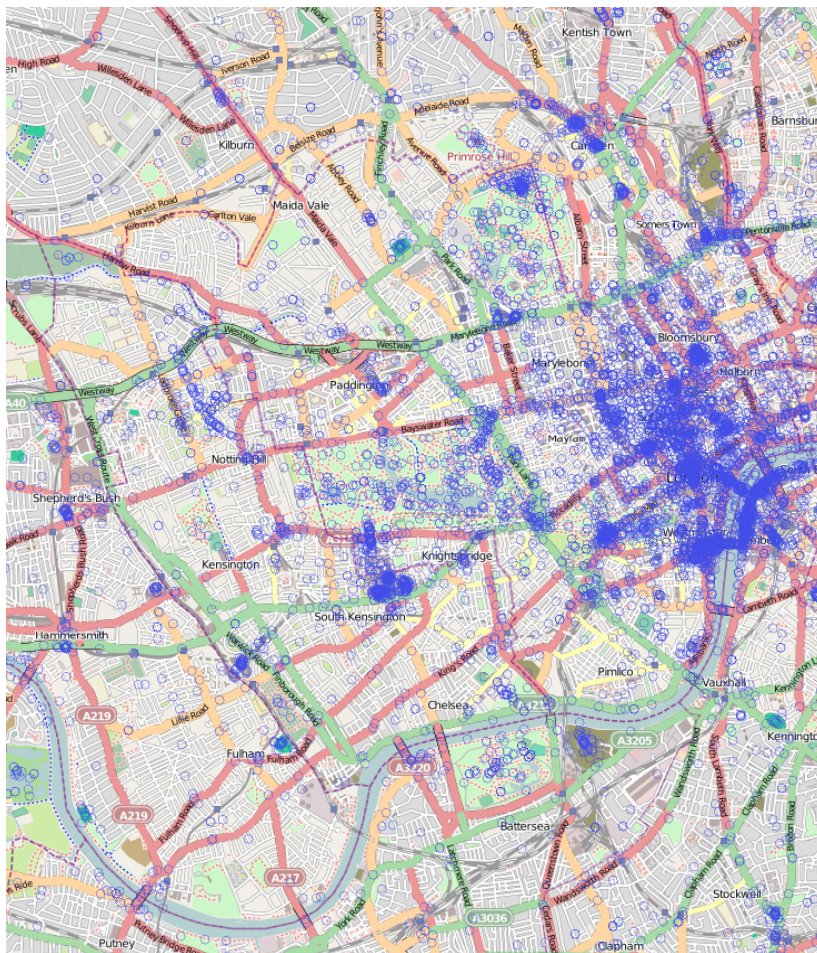
# Data examples

- 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)



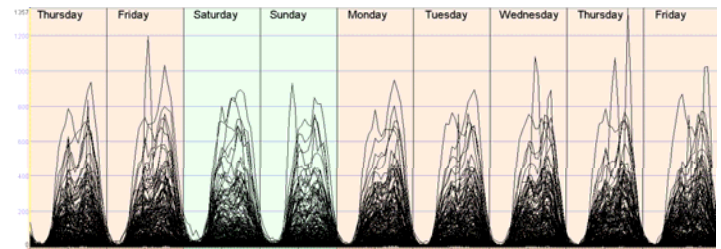
# How we do it (1)

- 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)

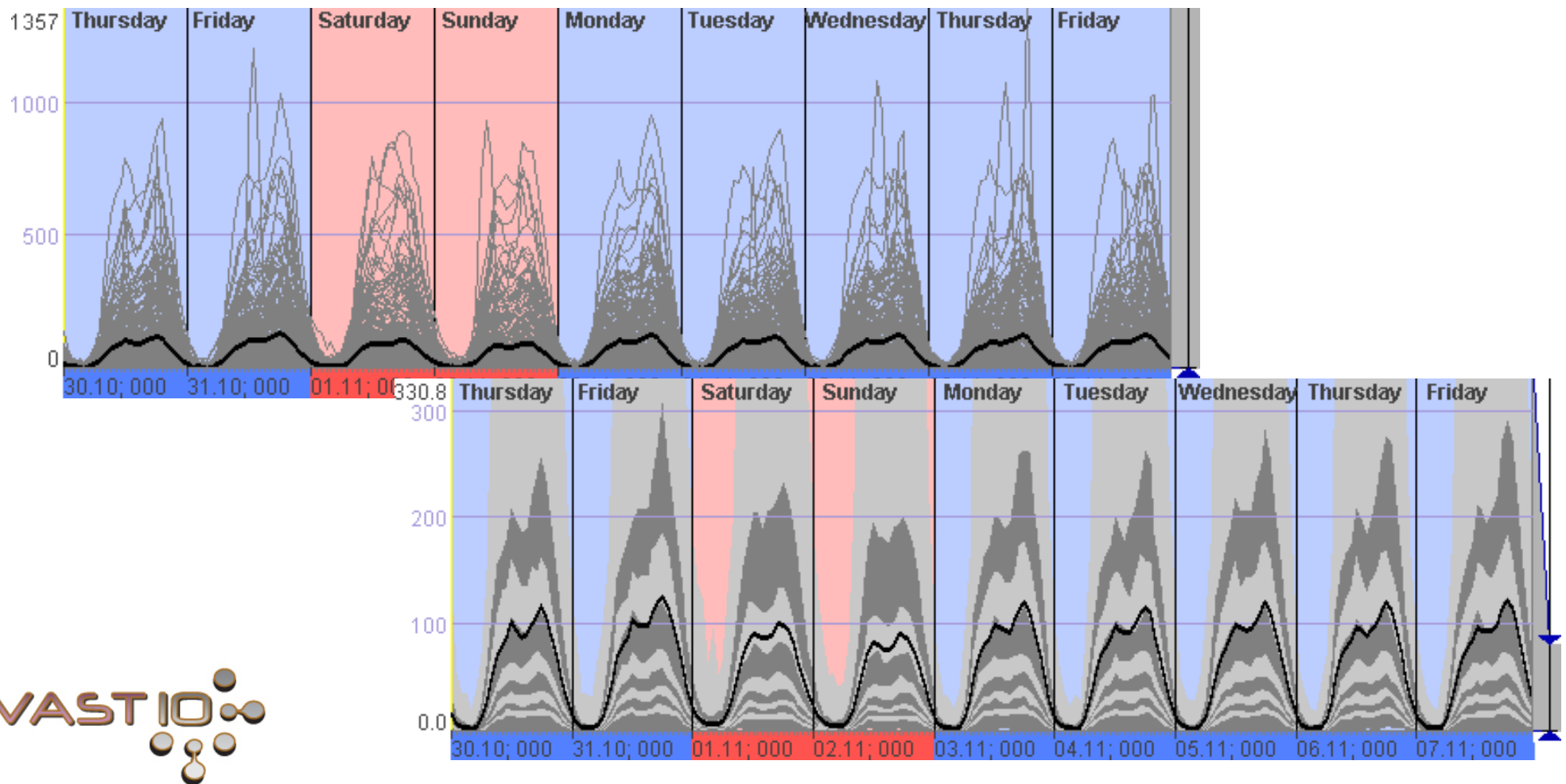
- **Time series analysis by statistical procedures**
  1. Periodicity (temporal correlation) detection:  
max of the circular cross-correlation function of a time series and a synthetic test pattern generated for a chosen period
  2. Peak event detection:  
identifying sudden increase (peaks) or decrease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

*Details in the paper*



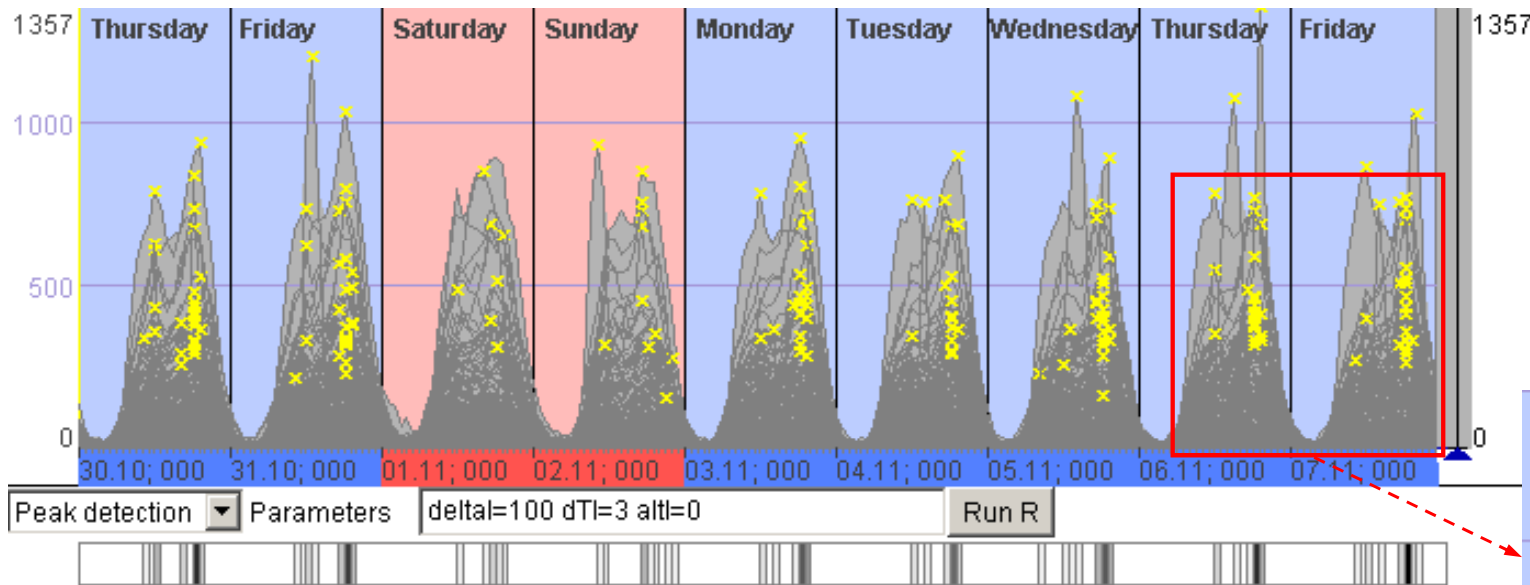
# How we do it (4a)

- **Interactive visual displays: time graph**

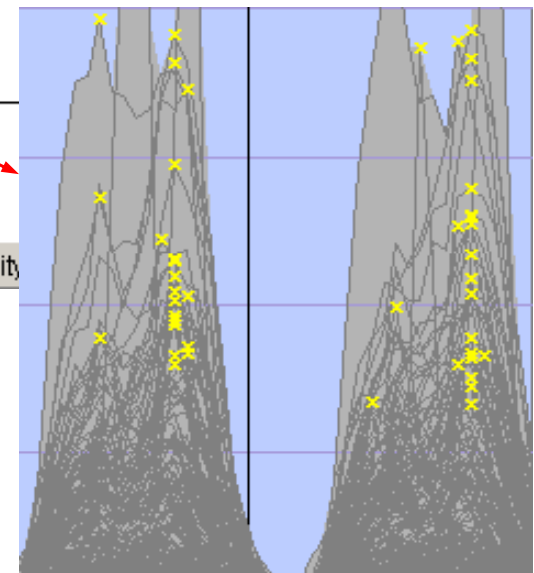
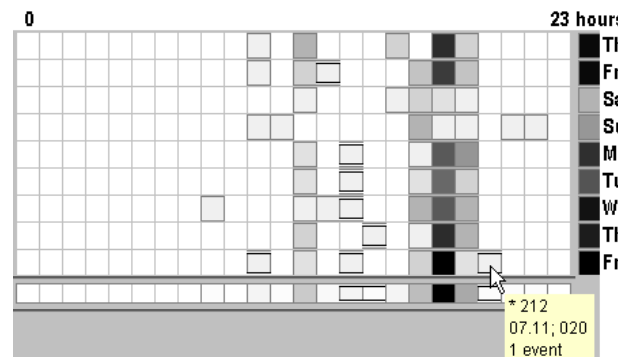


# How we do it (4b)

- Interactive visual displays: event detection, event bar

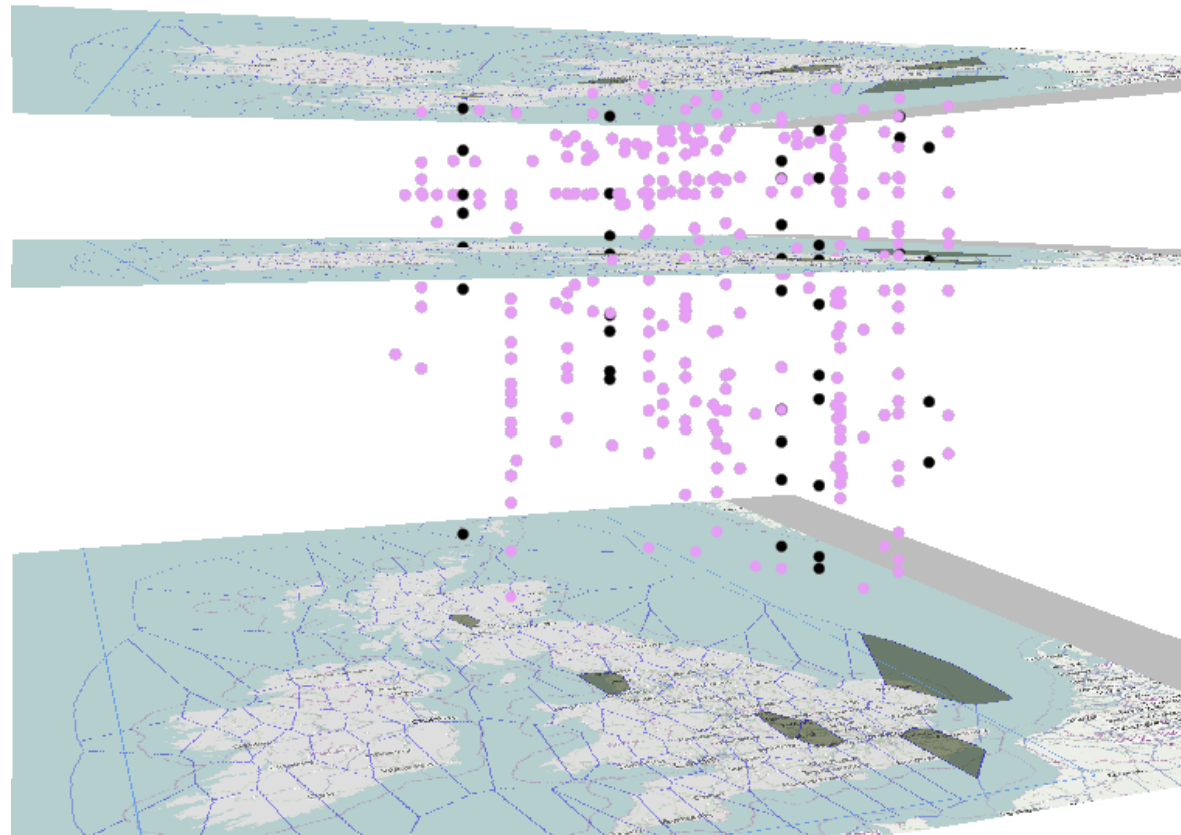
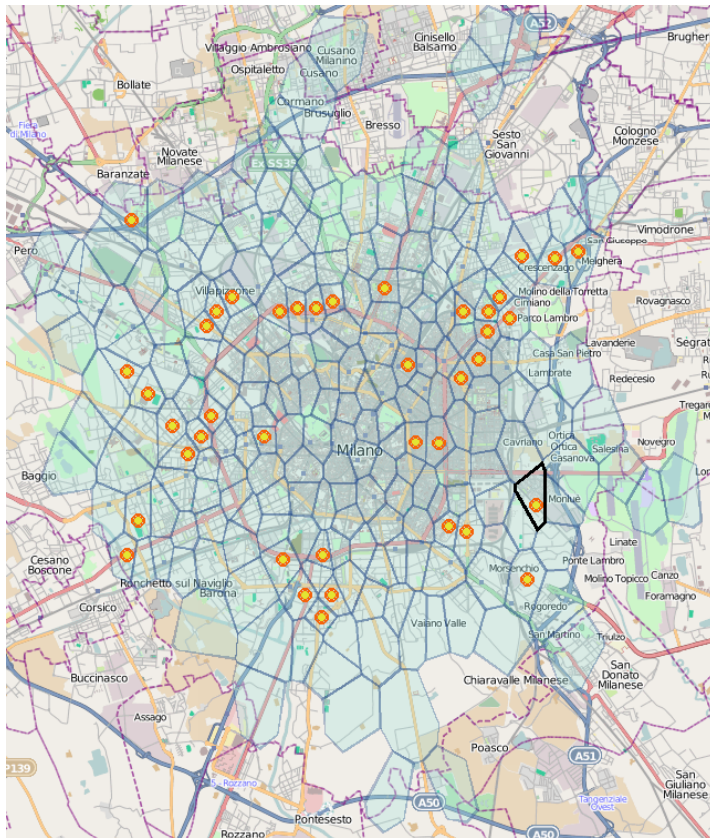


187 events occurred in 40 time series at 52 time moments



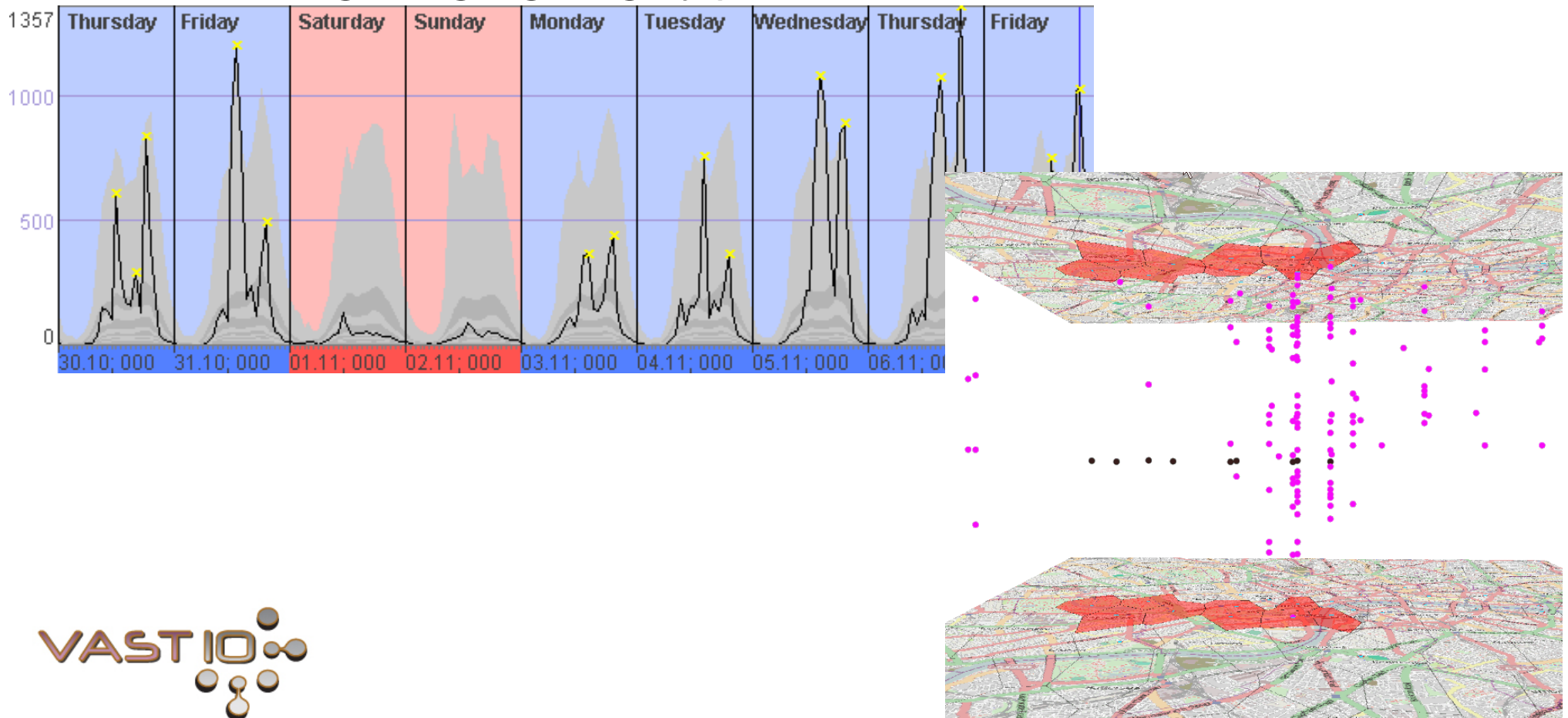
# How we do it (4c)

- **Interactive visual displays: map & space-time cube**



# How we do it (4d)

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



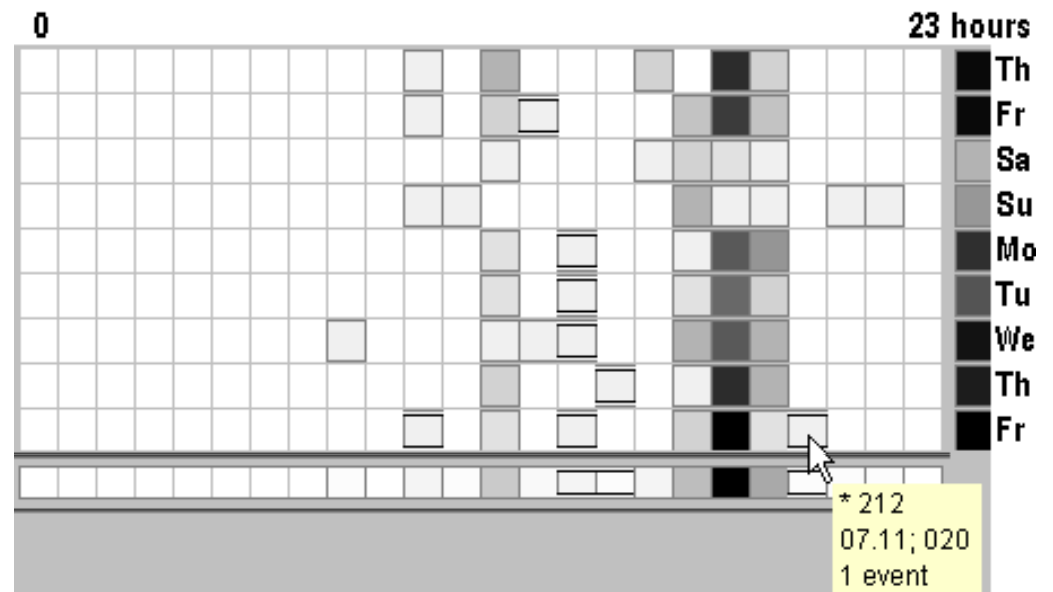
# Case study 1: phone calls

- 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



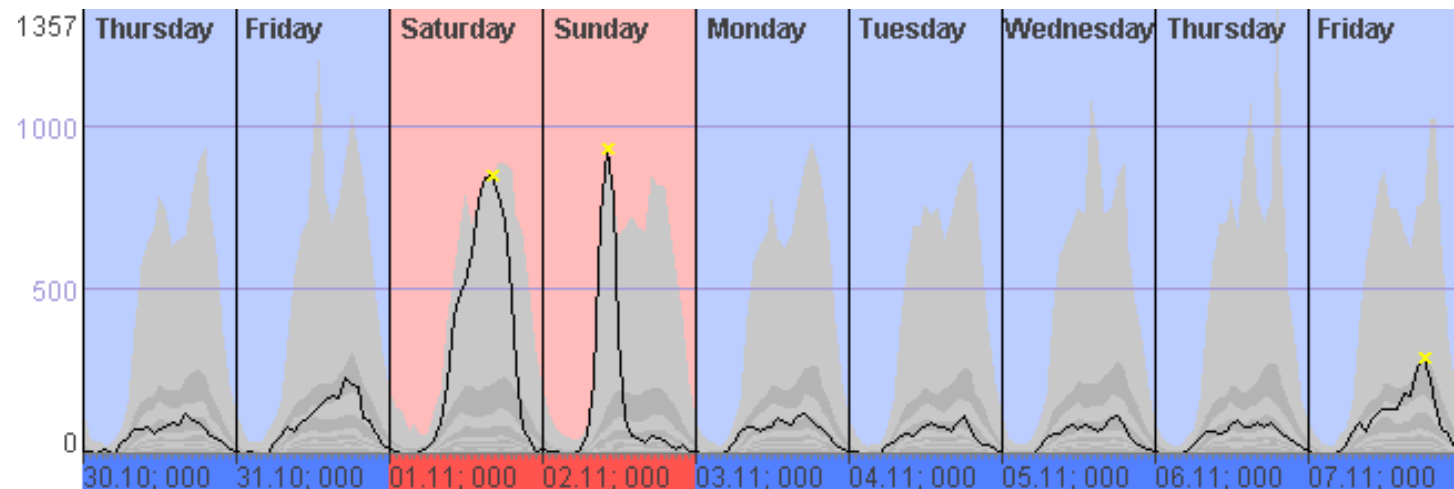
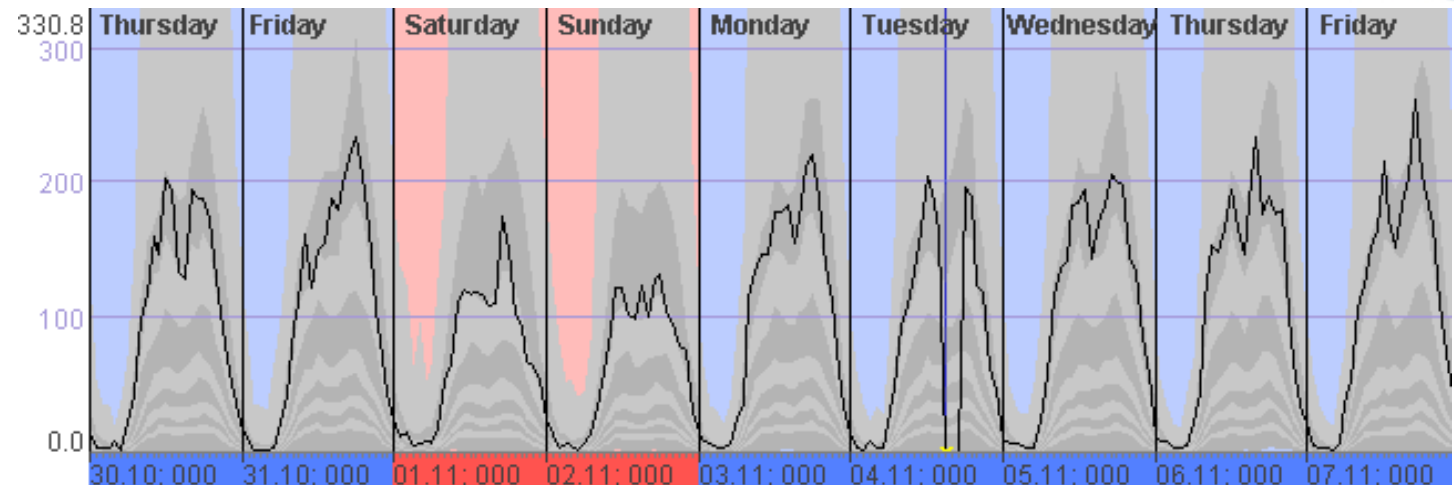
# Findings: when peaks happen

- 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!”*



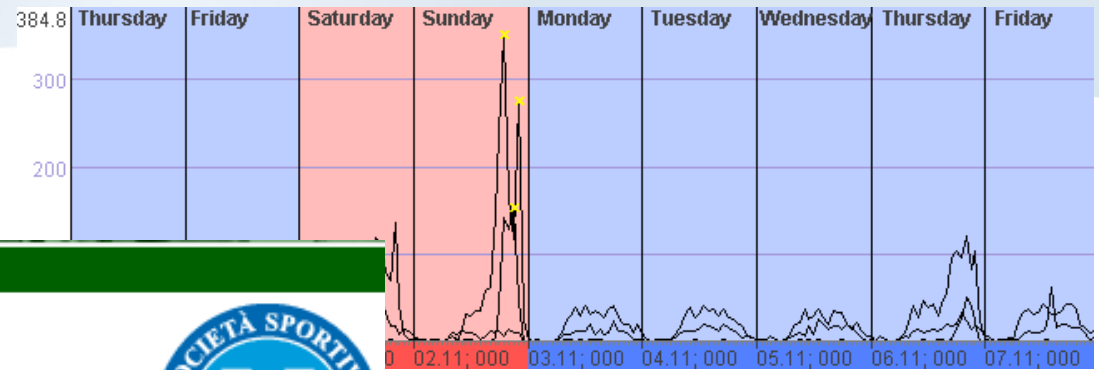
# Findings: non-periodic peaks & pits

- 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

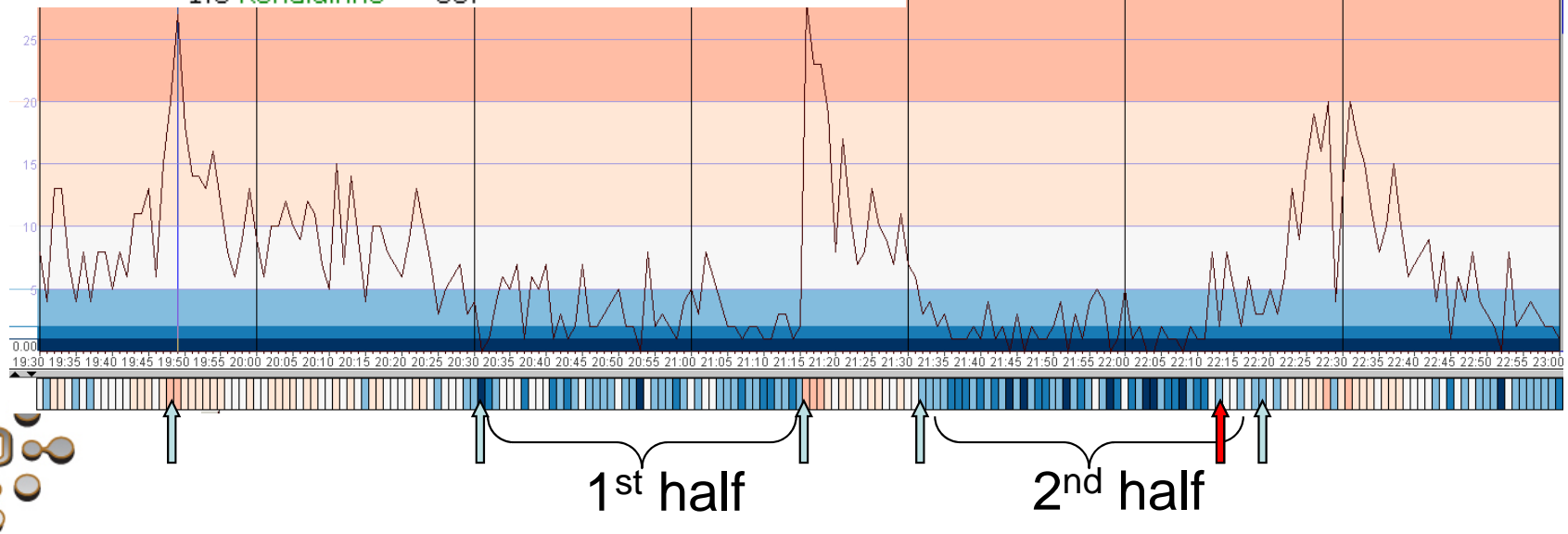
Serie A 2008/2009, 10. Spieltag

51.600 Zuschauer - Schiedsrichter: Gianluca Rocchi (Italien)



Tore:

1:0 Ronaldinho 86.





# Case study 2: flickr.com photos

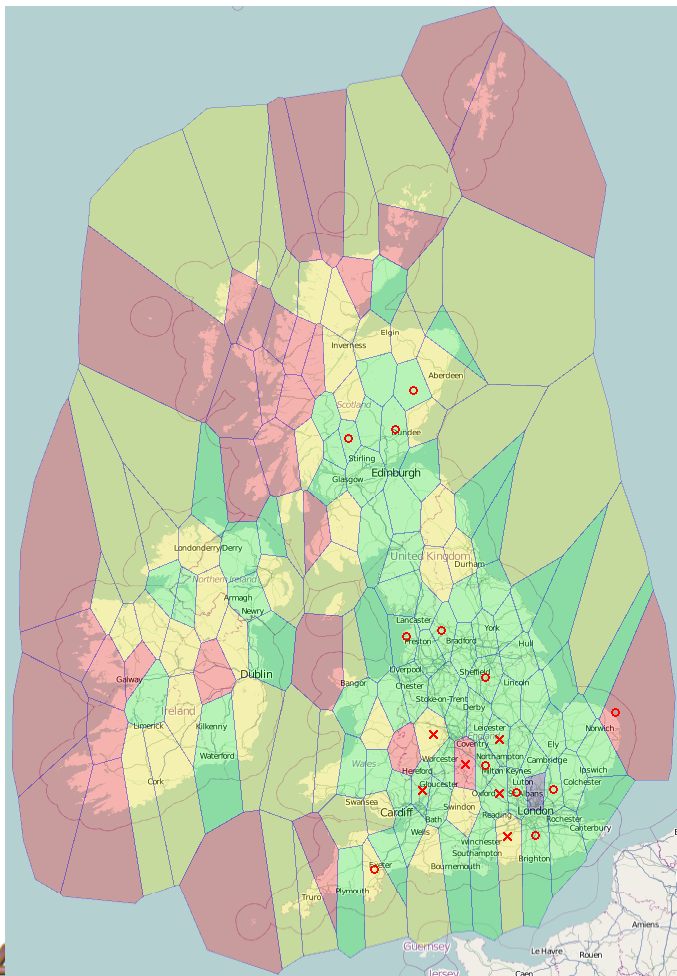
- Positions, time stamps, and titles of 8,686,034 photos in UK and Ireland during 5 years
  - Extracted from flickr.com



# General patterns

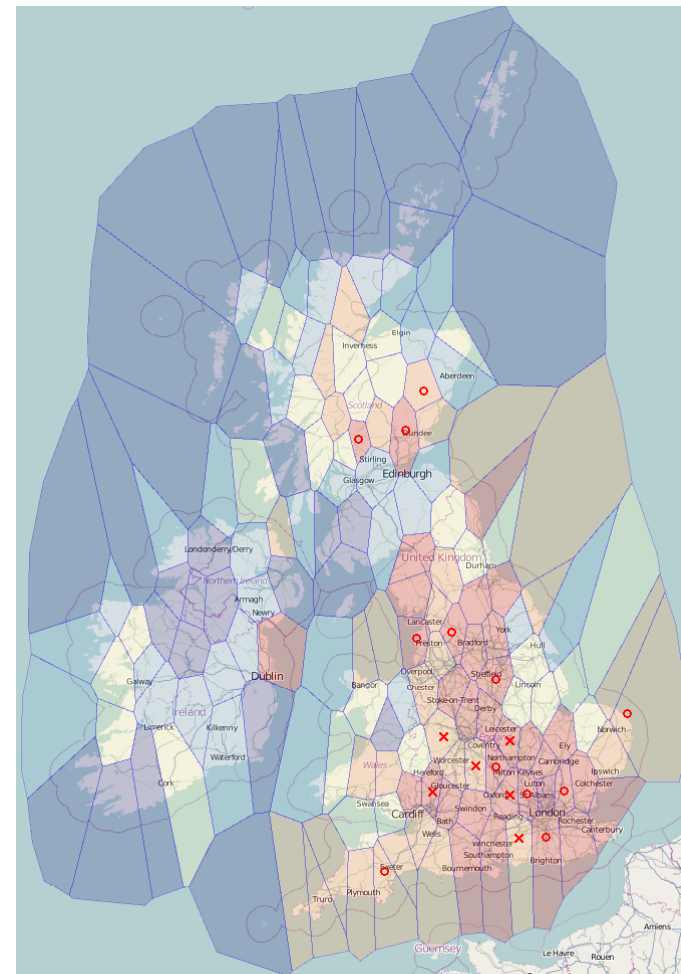
-  Non-periodic events
-  Periodic events

- Periodicity of time series



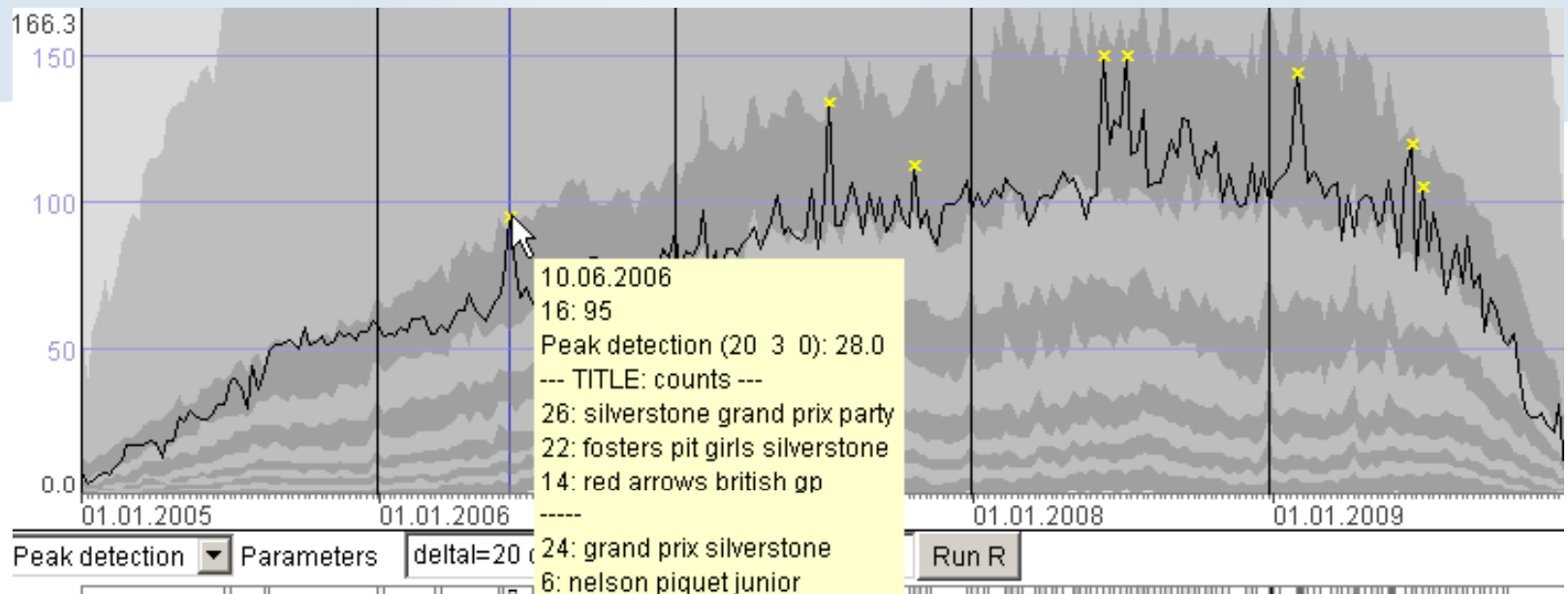
Green=low; Red=high

- Counts of photos



Blue=low; Red=high

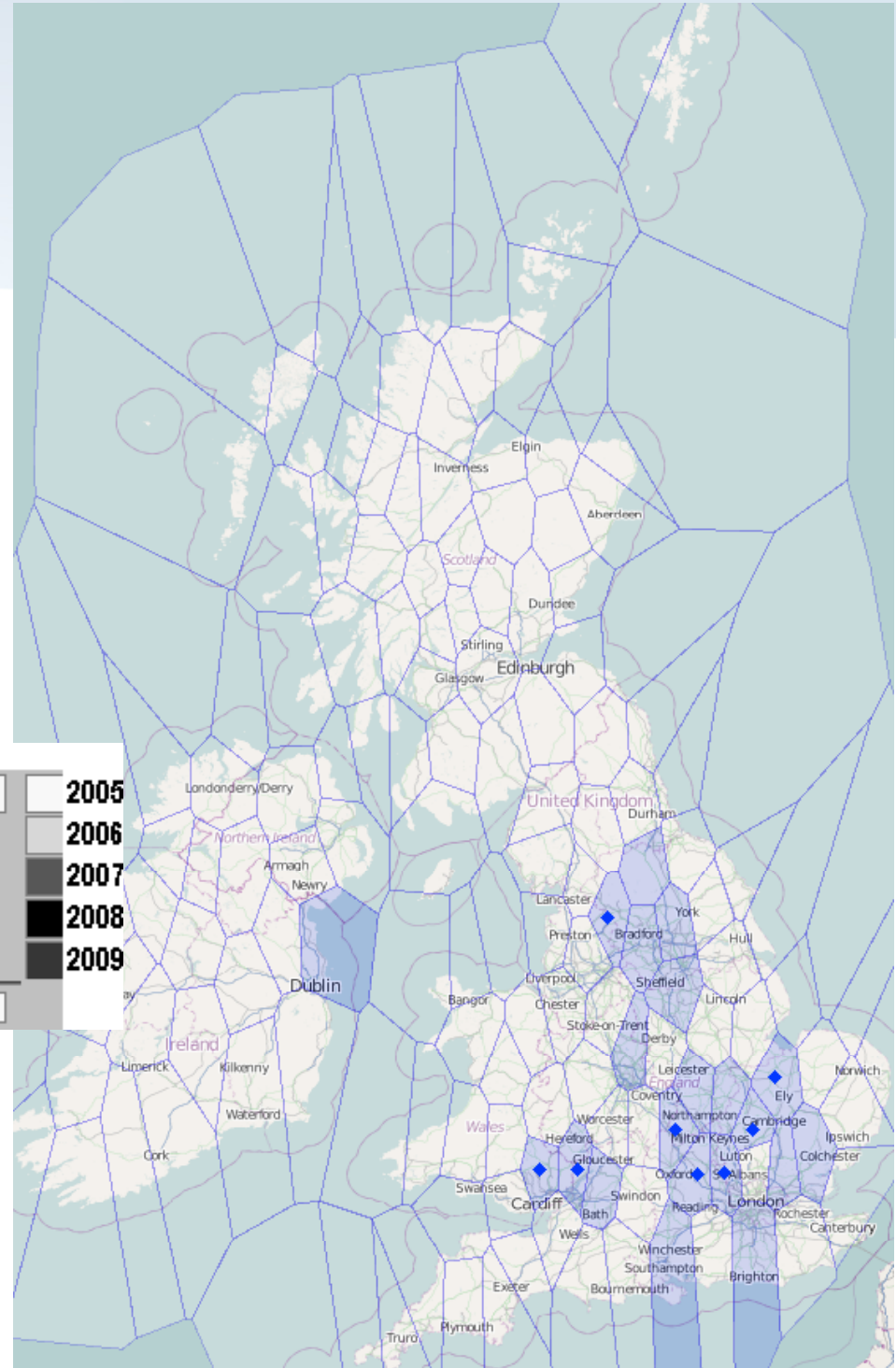
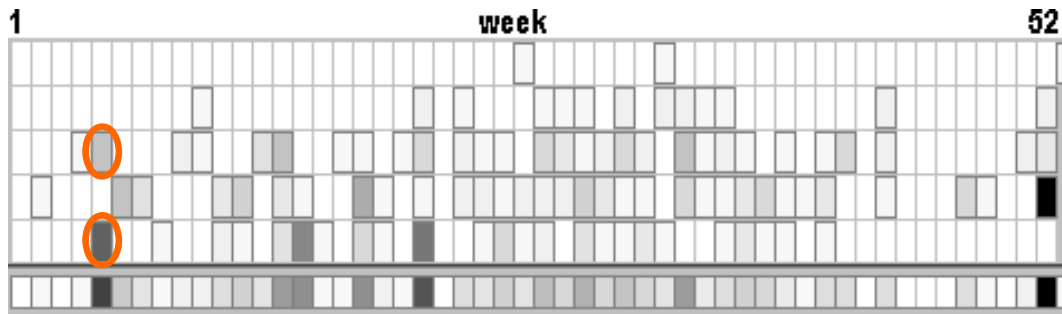
# Examples of periodic events



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

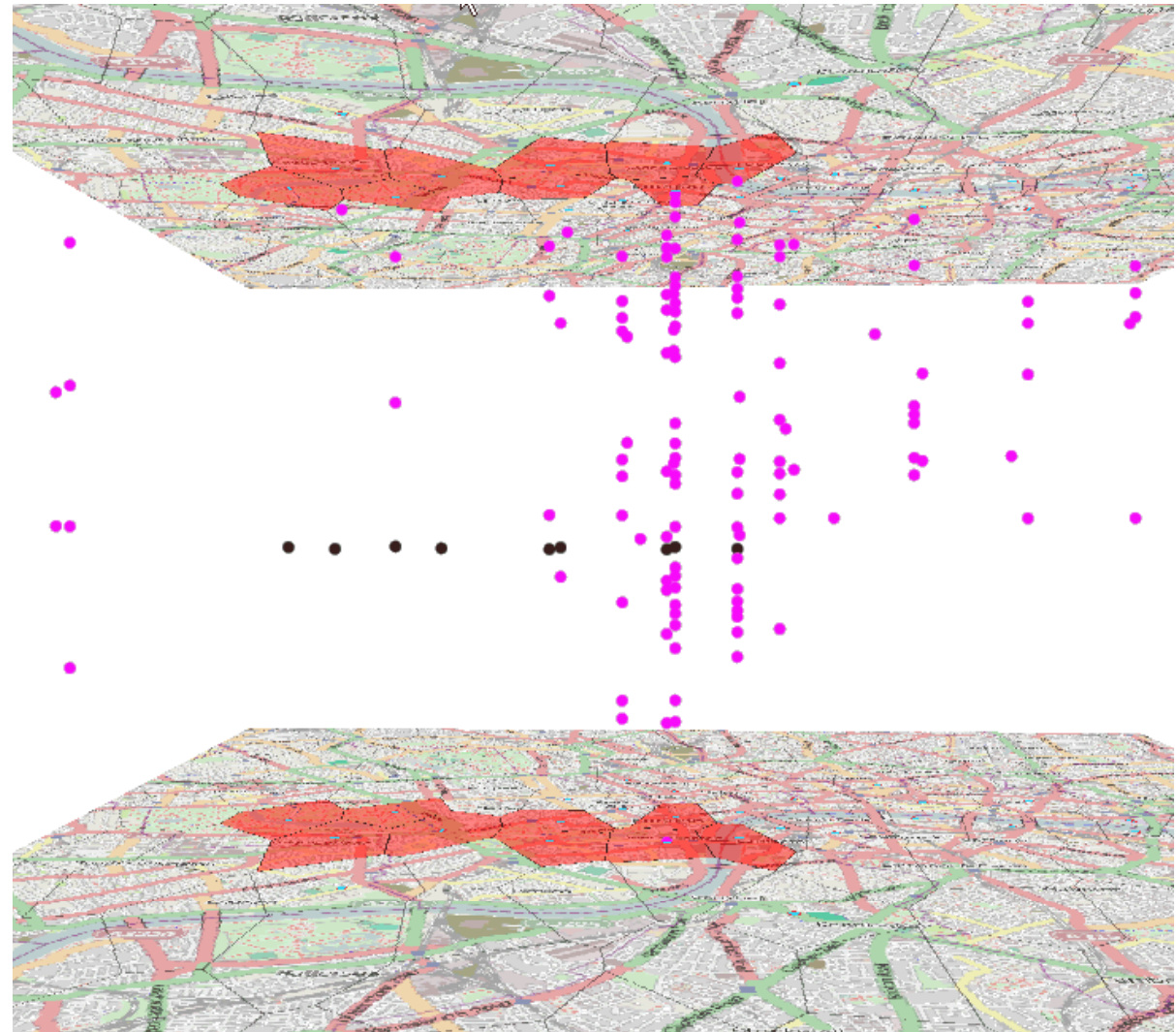
# Irregular peaks

- 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



# Demo

- Video...



# Conclusions (1)

- 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



# Conclusions (2)

- 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



# ToDo: enabling end users

The screenshot shows a web application interface for geoanalytics. The main area is a map of Germany, centered on the Cologne region. The interface includes several components:

- Navigation and Controls:** A sidebar on the left contains tabs for "Introduction", "Pictures", "Properties", and "Others". Below these are checkboxes for "Draw polygon for individual query", "Make use of the given database polygons", and "Disable triangle aggregation". There are also "Time Controller 1" and "Time Controller 2" sections with "Select an interval" options.
- Calendar:** A "Time Controller 3" section features a calendar for December 2007 and January 2008. The date "25" is highlighted in both months.
- Map and Legend:** The map shows various locations and roads. A legend/info panel on the right provides details for the selected date: "Title: null", "Person: 18438672@N00", "Date: Tue Dec 25 11:37:00 CET 2007", and "RawTags: brücke bw germany gottaged blackwhite europa europe cologne bn ponte nrw colonia brücke rhine bianconero germania northrhinewestphalia hoherollerbrücke sfidephotoamatori".
- Photo Gallery:** A horizontal strip of five photo thumbnails is displayed at the bottom. The first thumbnail shows a cathedral, and the second shows a bridge. The other three thumbnails show a seagull and a building.



- See at VAST 2011 ☺
- Visit us at <http://geoanalytics.net>