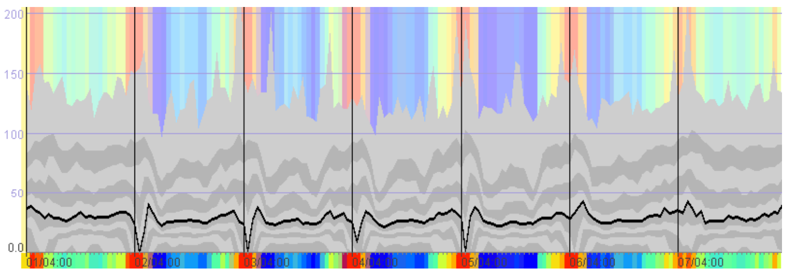
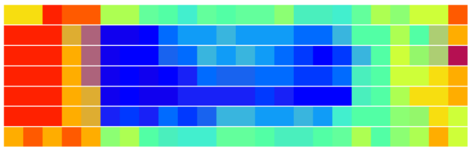
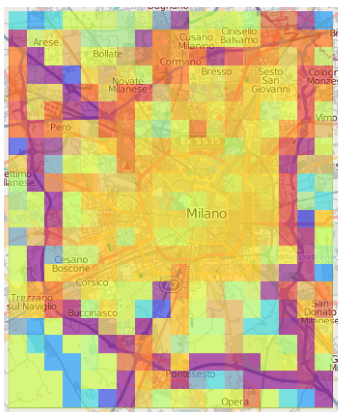
Space-in-Time and Time-in-Space Self-Organizing Maps for Exploring Spatiotemporal Patterns







Gennady Andrienko & Natalia Andrienko
http://geoanalytics.net
Fraunhofer
IAIS

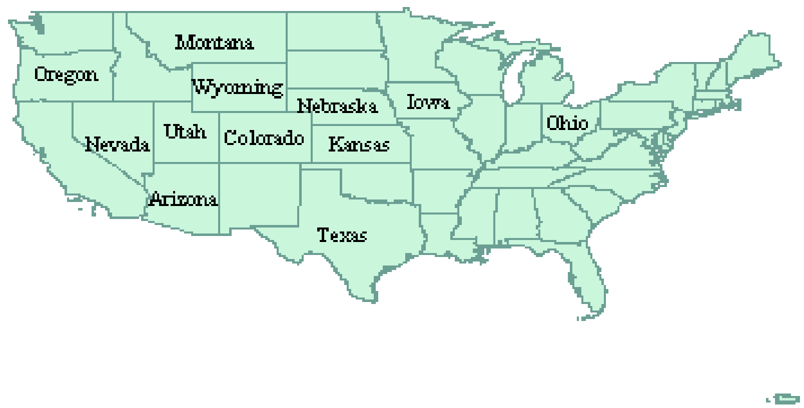
Sebastian Bremm, Tobias Schreck, Tatiana von Landesberger
TU Darmstadt
Peter Bak, Daniel Keim
Univ. Konstanz

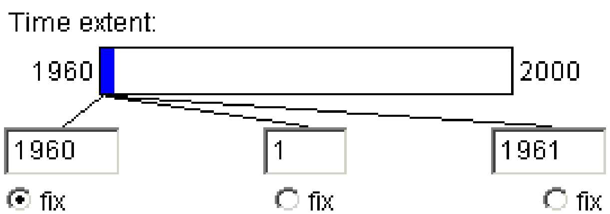
DFG Priority Research Program on Visual Analytics, SPP 1335

10 June 2010, presentation at EuroVis
Bordeaux, FR

Spatial Time Series: Data Structure

Spatial references: states of the USA



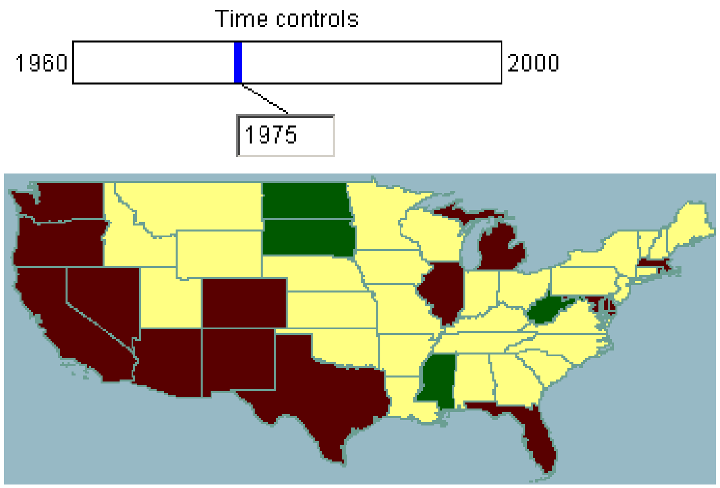


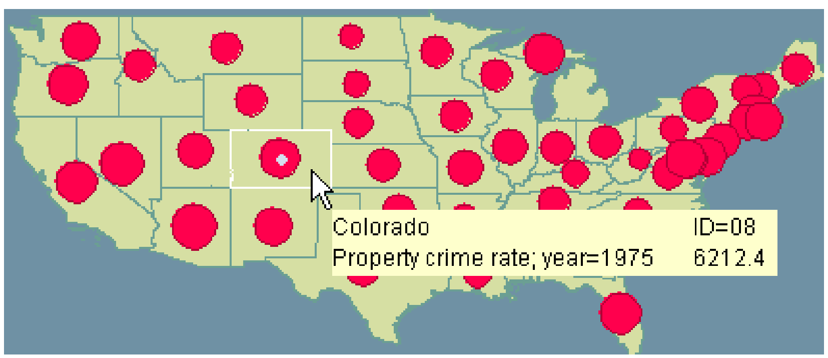
Temporal references: years from 1960 till 2000 (41)

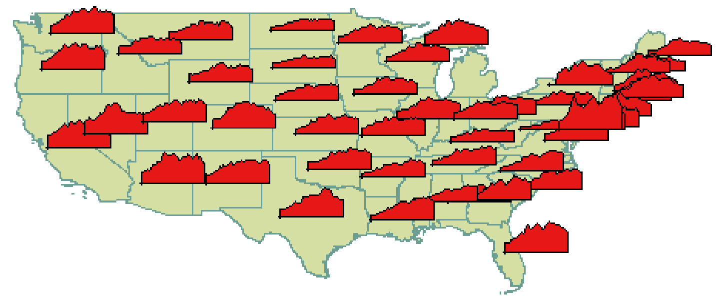
Attributes: population + various crime rates

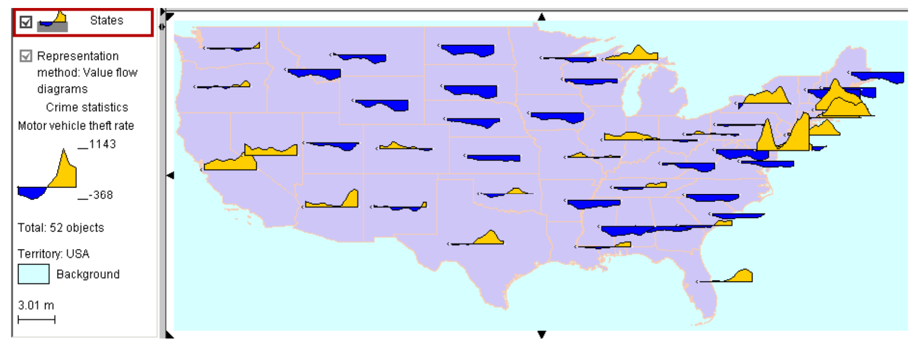
Population	Index offense rate	Violent Crime rate	Murder and nonnegligent manslaughter rate	Forcible rape rate	Robbery rate	Aggravated assault rate	Property crime rate	Burglary rate	Larceny-theft rate	Motor vehicle theft rate

Spatial Visualizations: animated maps, diagram maps

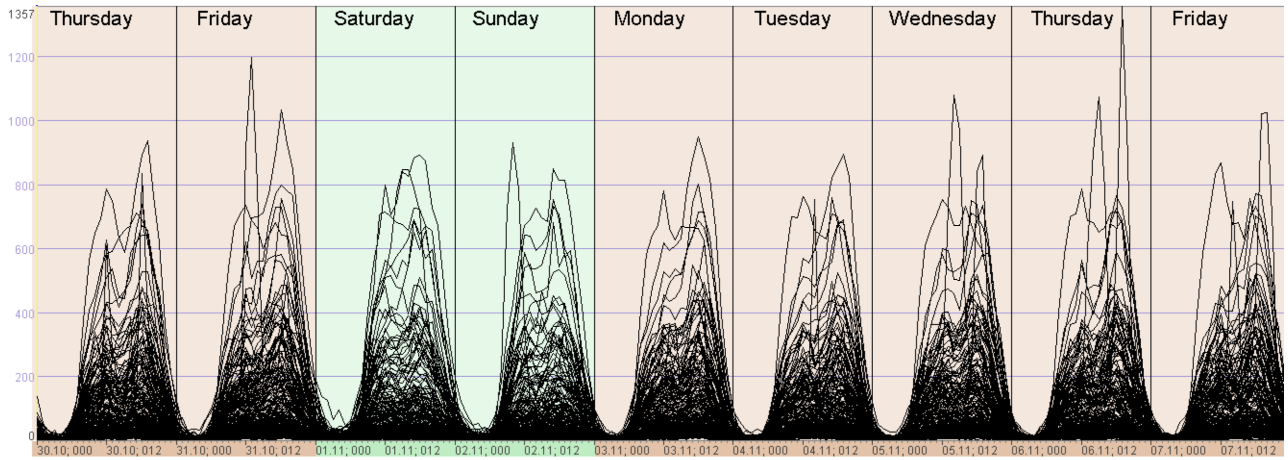


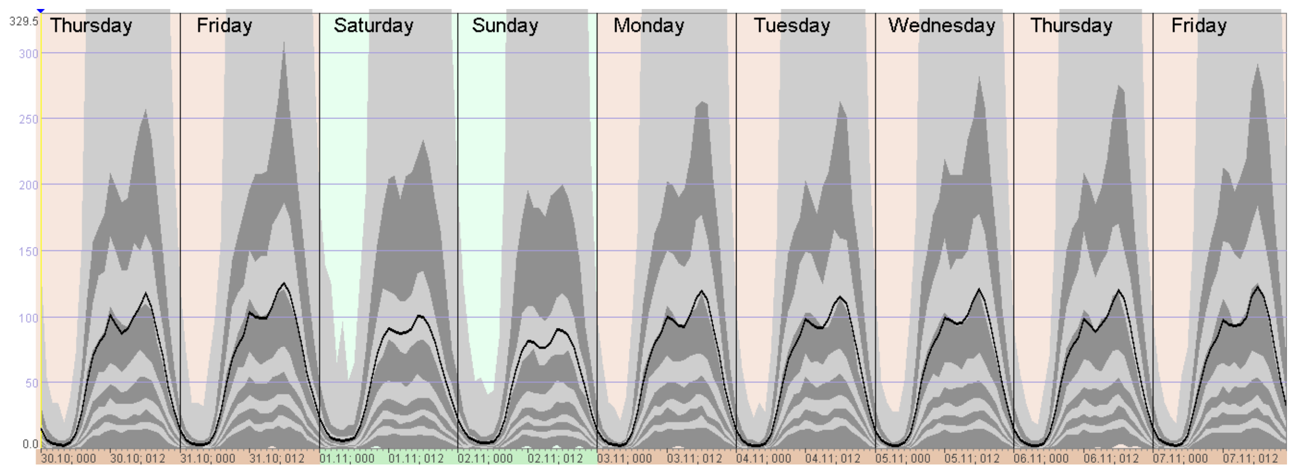






Temporal visualizations





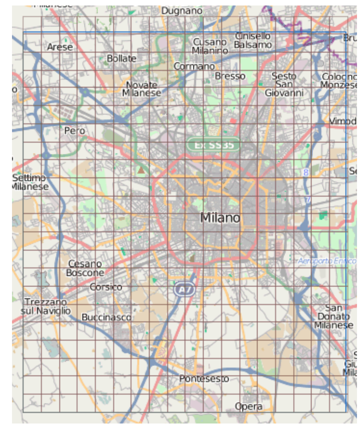
Scalability problem

- What if we have
  - Multiple attributes
  - Many places
  - Long time series
- Interactive visualization is not sufficient

- We need grouping in space and time => Clustering

Data sets

- Cars in Milan, Italy
  - 175,890 trajectories of 17,241 cars over 7 days
  - 2,075,216 records <id,x,y,t,speed>
  - Aggregated over 18x22 1km^2 rectangular regions and 7x24 hourly intervals



- USA crime statistics
  - 7 crime attributes
  - 52 states
  - 41 years

Approach

	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	0	1	4	3	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

Approach

2) Group time intervals by similarity of spatial situations:
clustering "Space in Time"

1) Group places by similarity of temporal dynamics:

clustering "Time in Space"

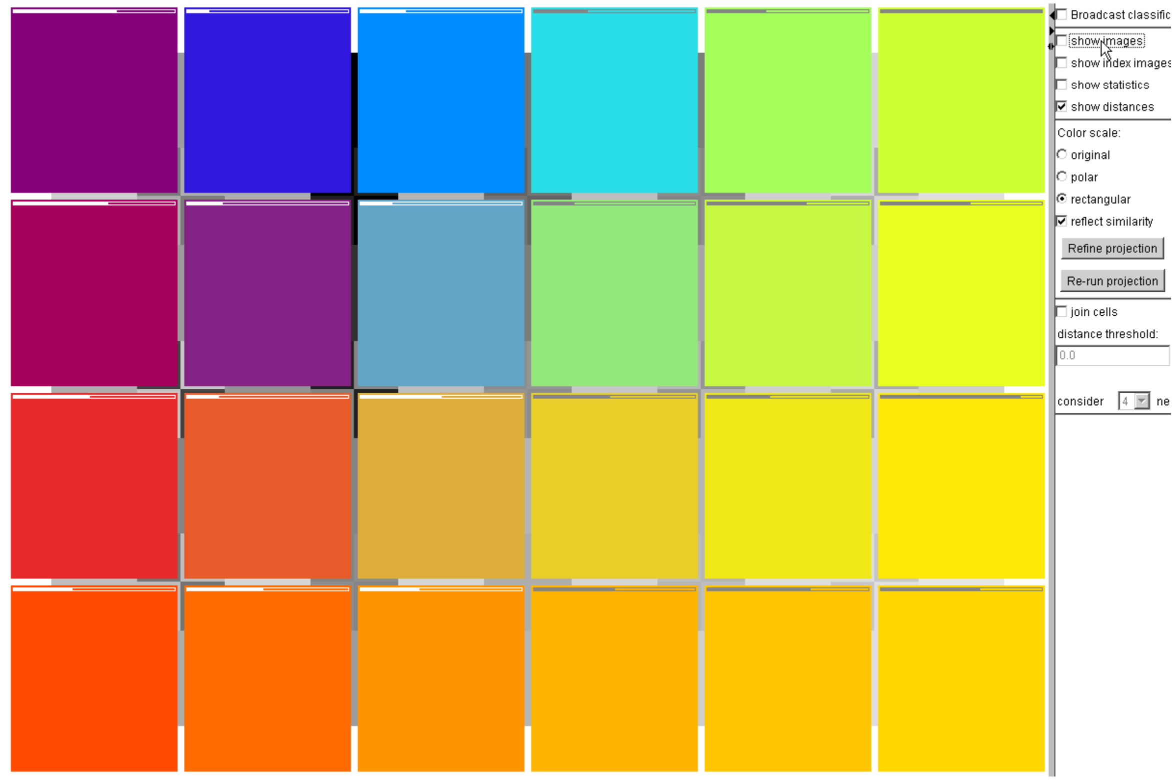
	N events by time intervals; Time interval (start)=01.1985	N events by time intervals; Time interval (start)=02.1985	N events by time intervals; Time interval (start)=03.1985	N events by time intervals; Time interval (start)=04.1985	N events by time intervals; Time interval (start)=05.1985	N events by 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	0	1	4	3	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

SOM

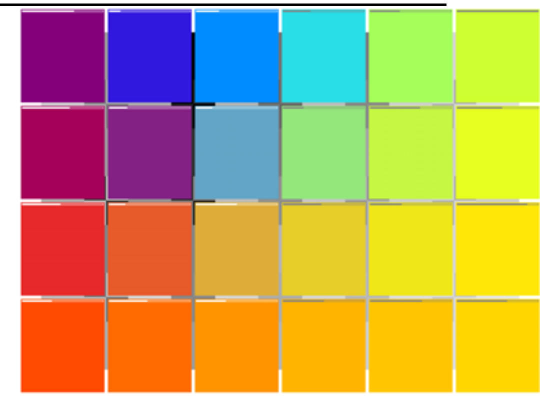
- Self-Organizing Map (Kohonen 2001) is a neural network type vector projection and quantization algorithm.
- By means of a competitive, iterative training process, a network of prototype vectors (or neurons, or cells) is trained (adjusted) to the input vector data.
- The output of the algorithm is a network of vectors that is approximately topology preserving w.r.t. the input data.
- The network can be interpreted as a set of clusters and simultaneously as a map to lay out the input data elements (e.g., in the nearest neighbor sense w.r.t. the prototypes).
- Typically, two-dimensional rectangular or hexagonal prototype vector networks are assumed.
- The capability of SOM to arrange input data in a regular network structure provides good opportunities for visualization.

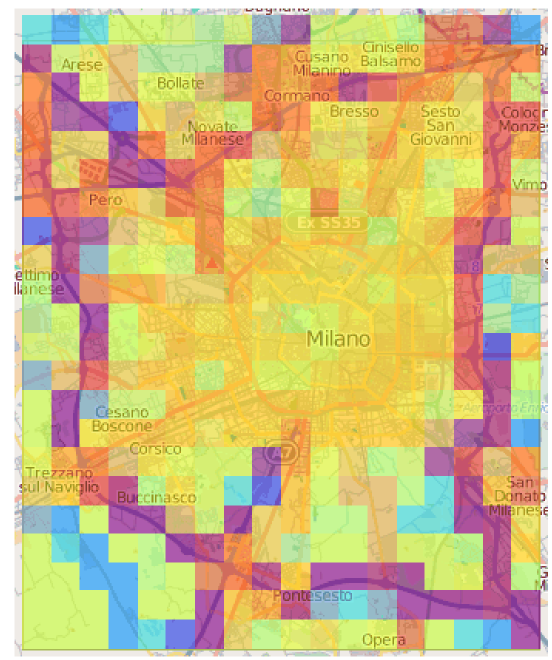
Space-in-Time and Time-in-Space SOMs: visualization

1. Bars on top of a cell show number of objects inside
2. Shading of borders between cells reflects similarity of features
3. Similarity of colors also reflects similarity of the features
4. Colors are projected on other displays:
   maps (if grouping places)
   and
   time graphs (if grouping time intervals)



Time-in-Space SOM of driving speeds

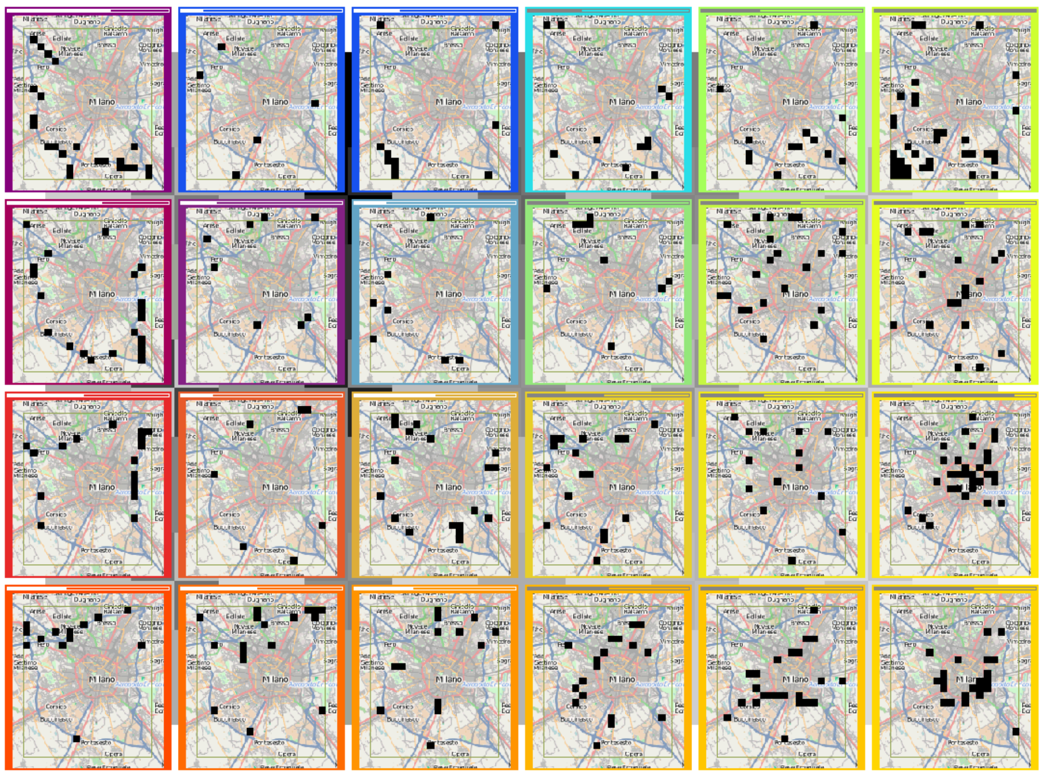


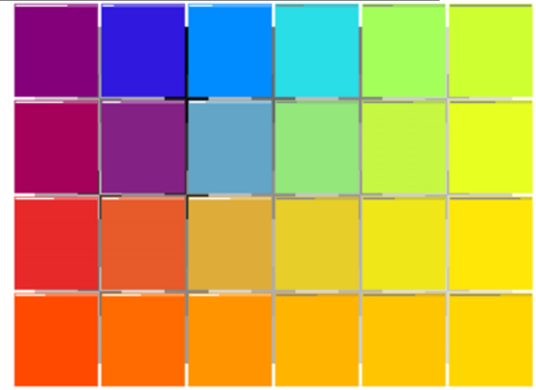


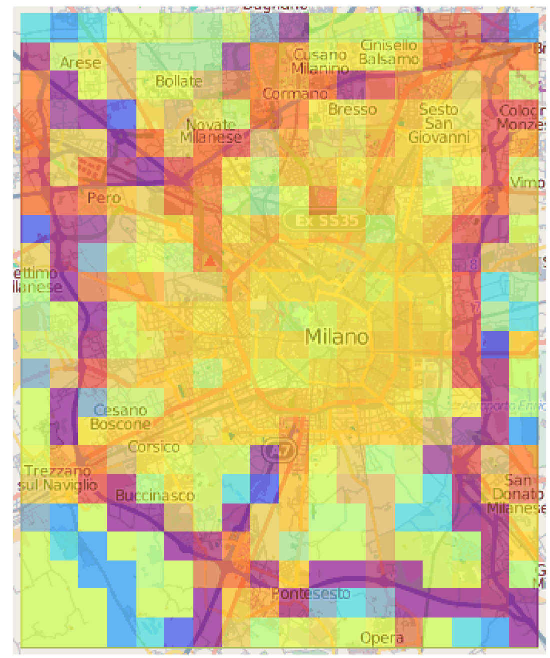
Time-in-Space SOM of driving speeds

Inside cells:

- index images (what is grouped)



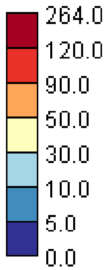


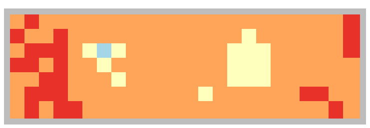


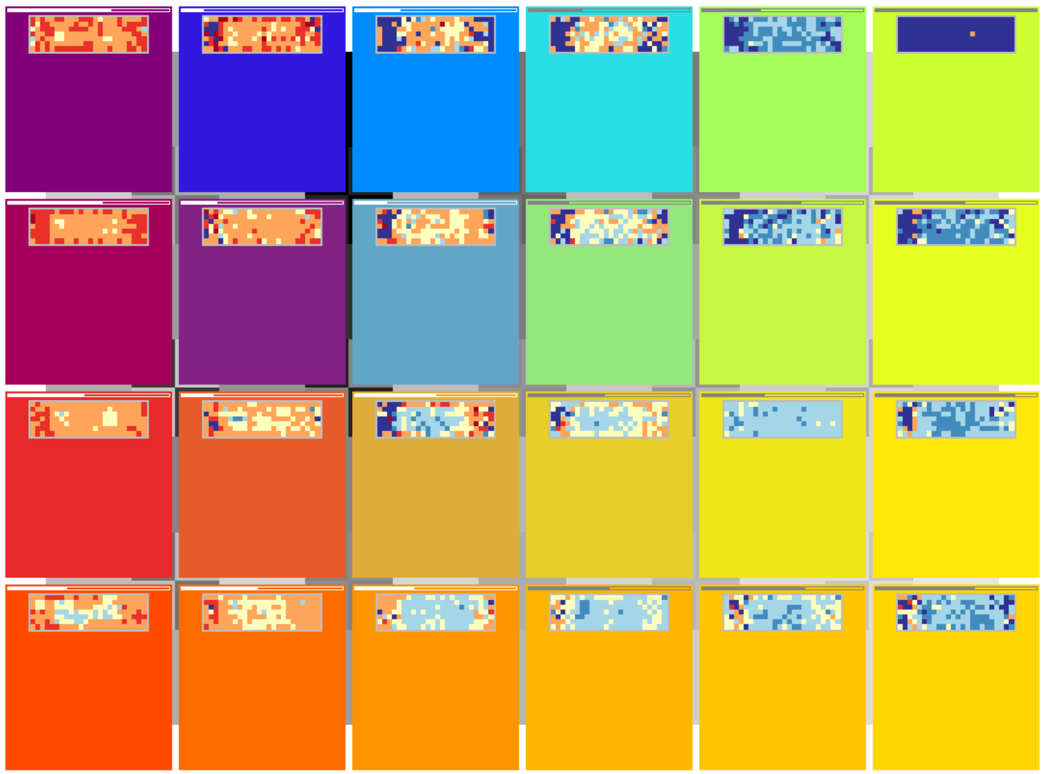
Time-in-Space SOM of driving speeds

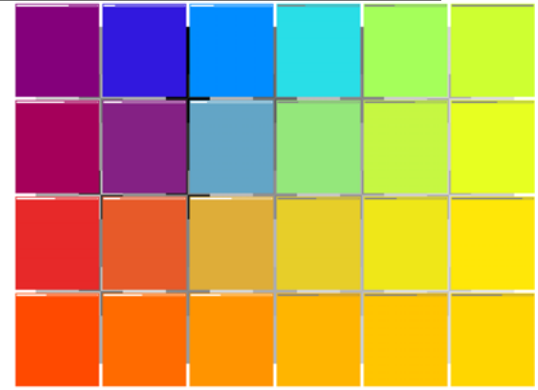
Inside cells:

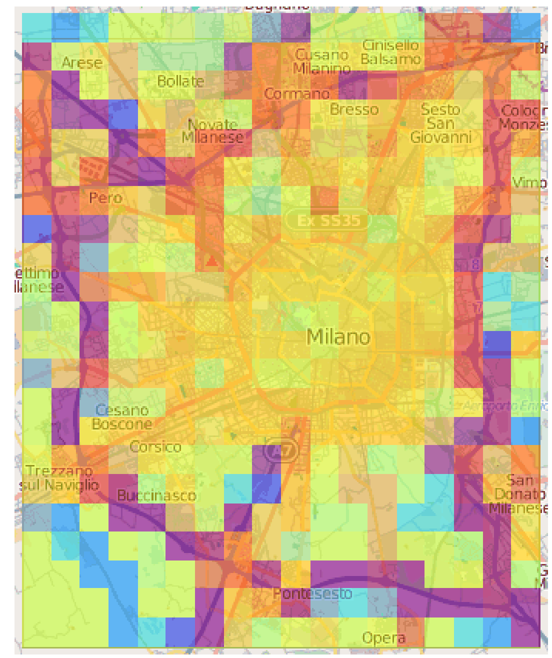
- feature images (what are the features)







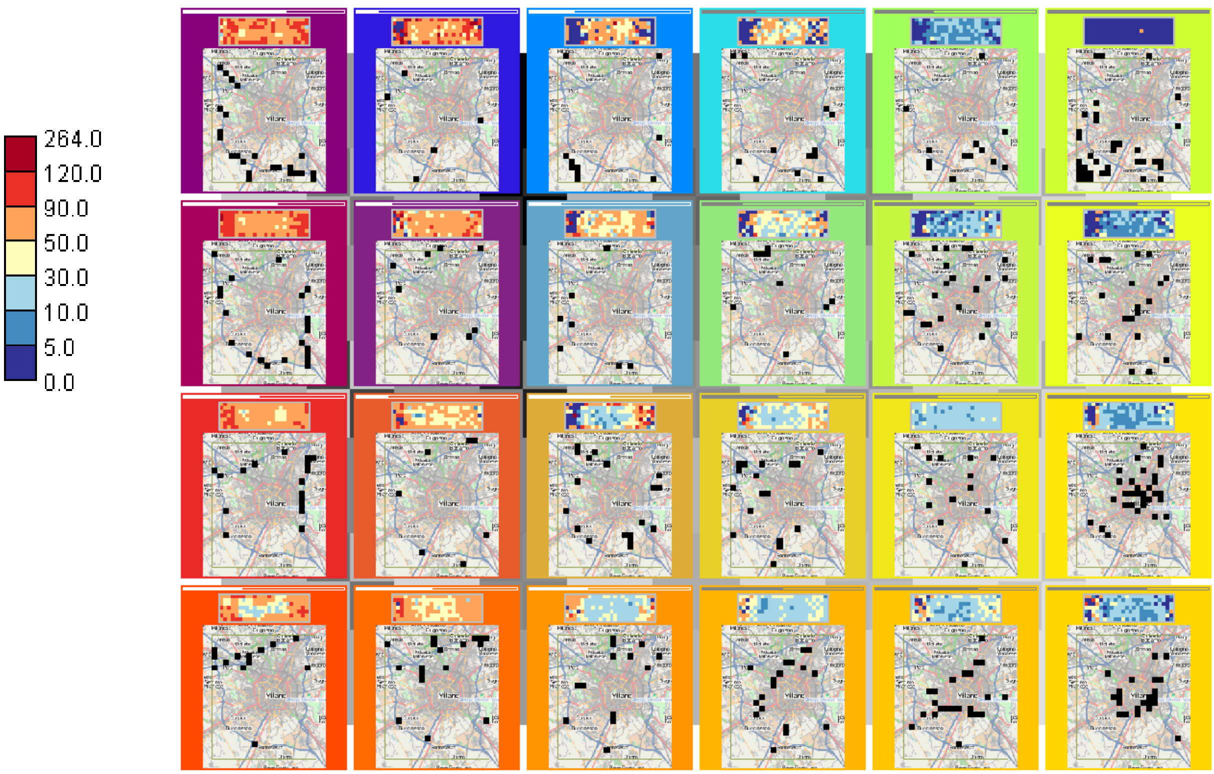


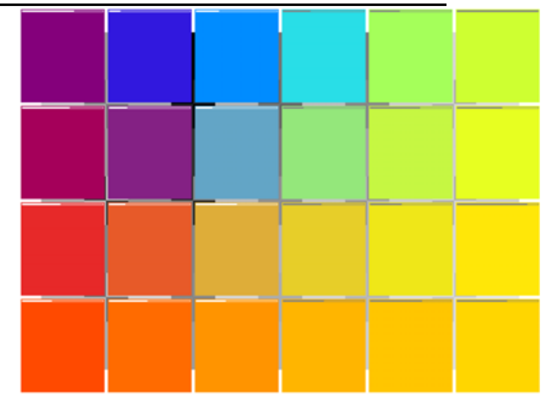


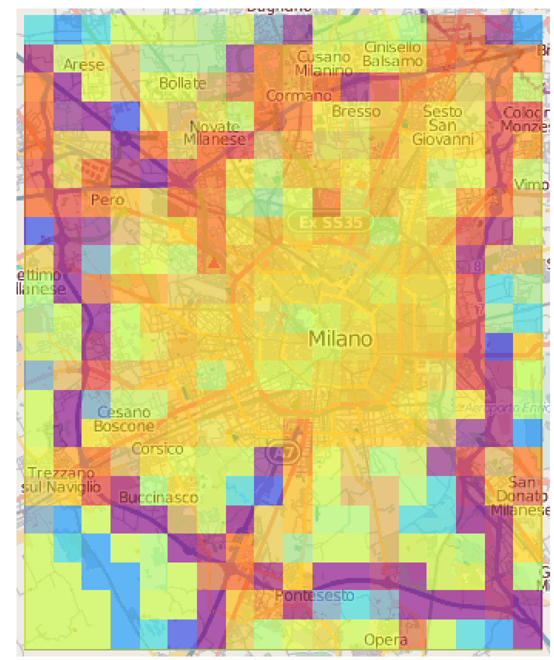
Time-in-Space SOM of driving speeds

Inside cells:

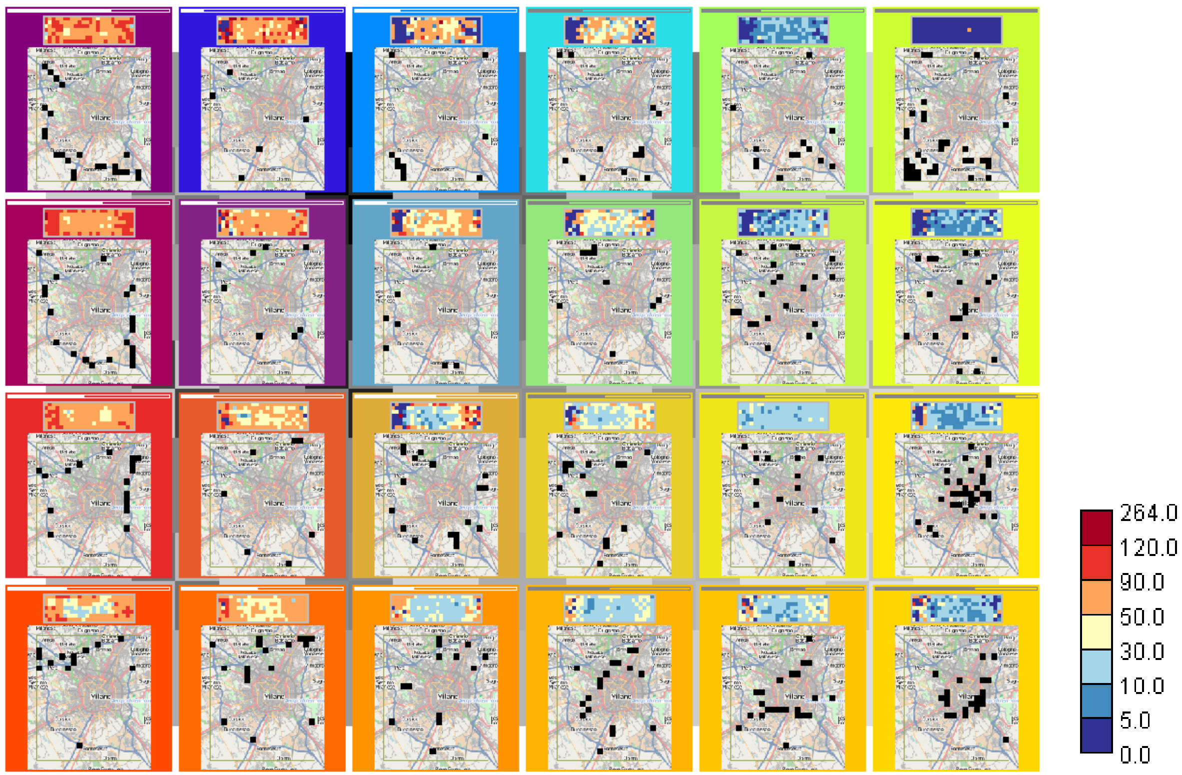
- index images (what is grouped)
- feature images (what are the features)



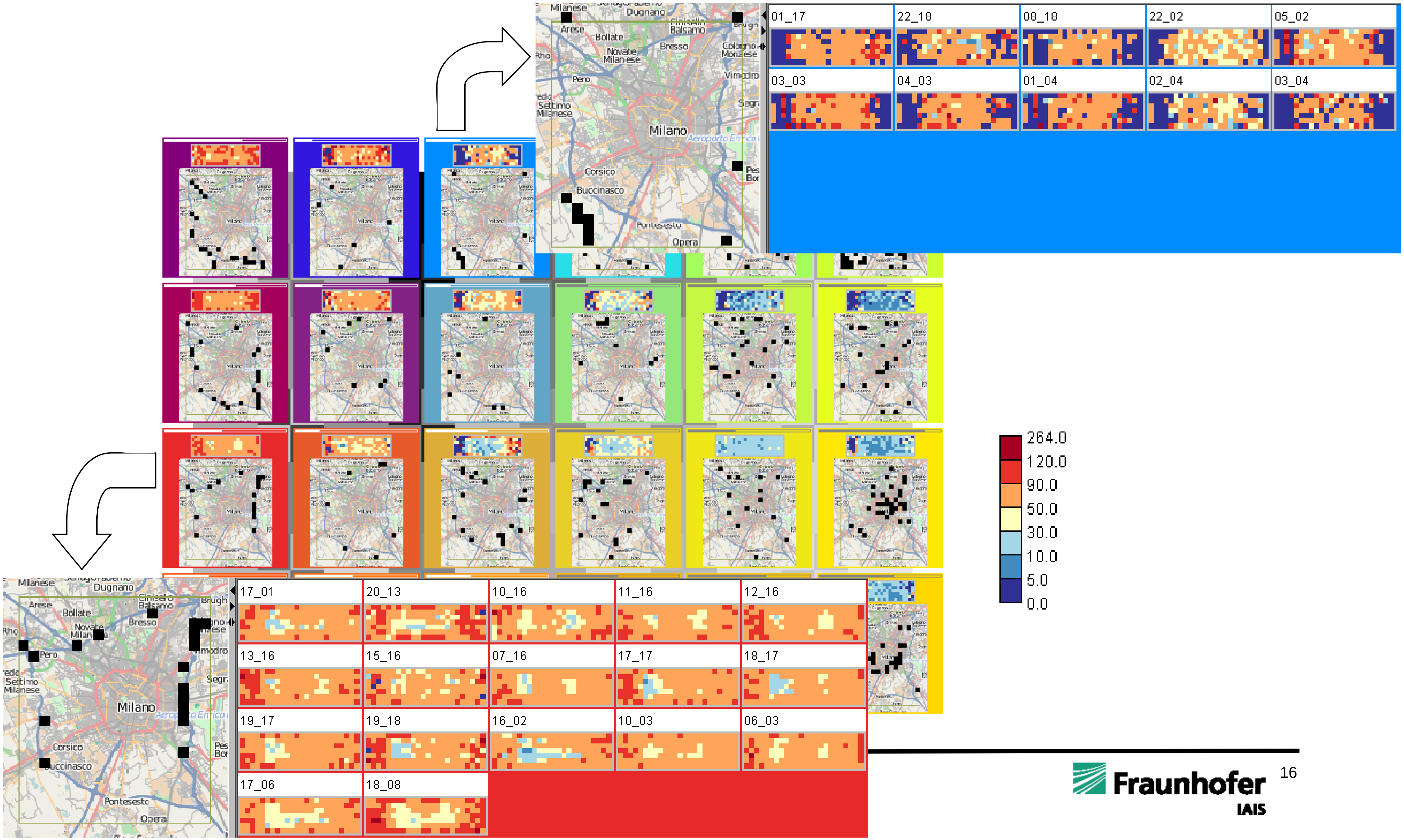




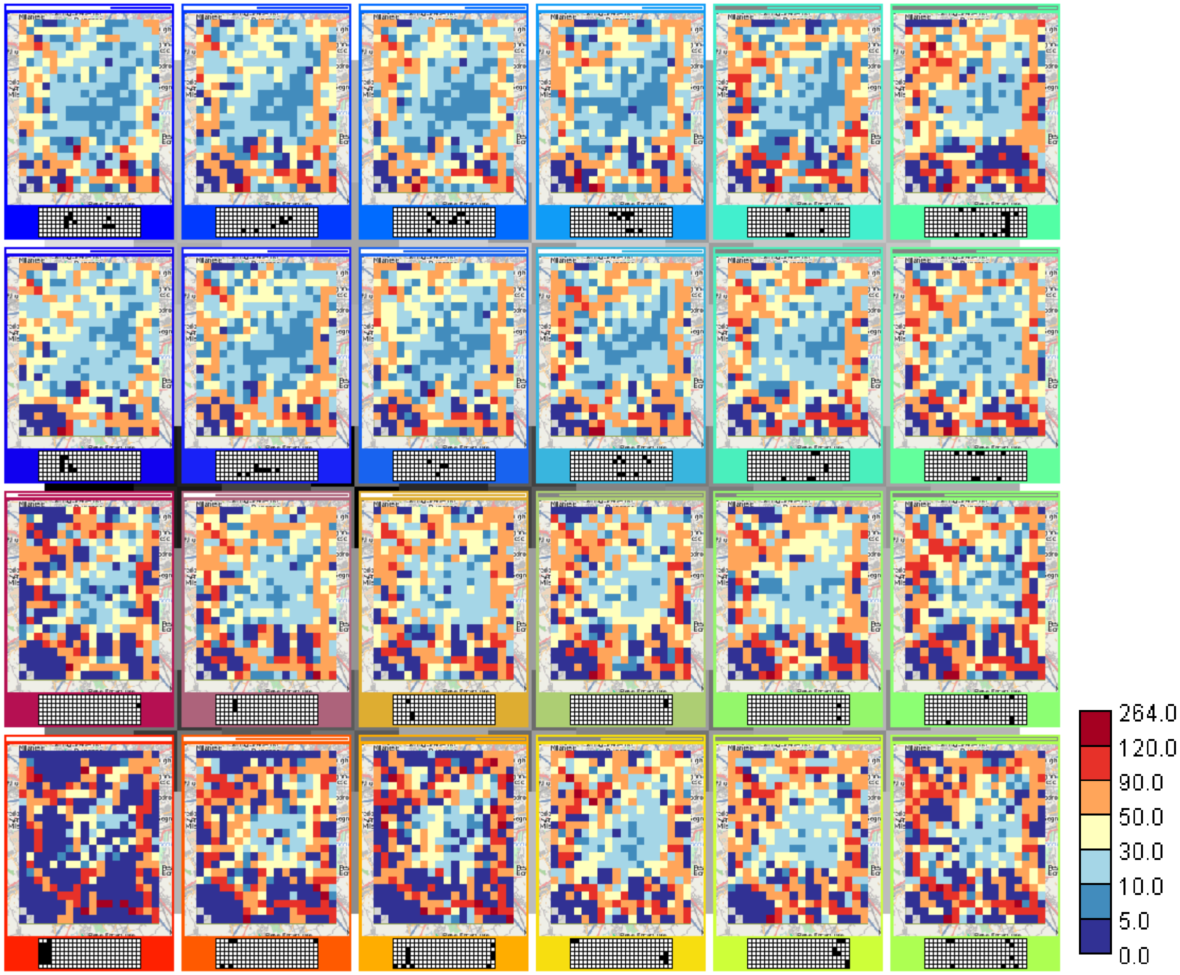
Time-in-Space SOM: details



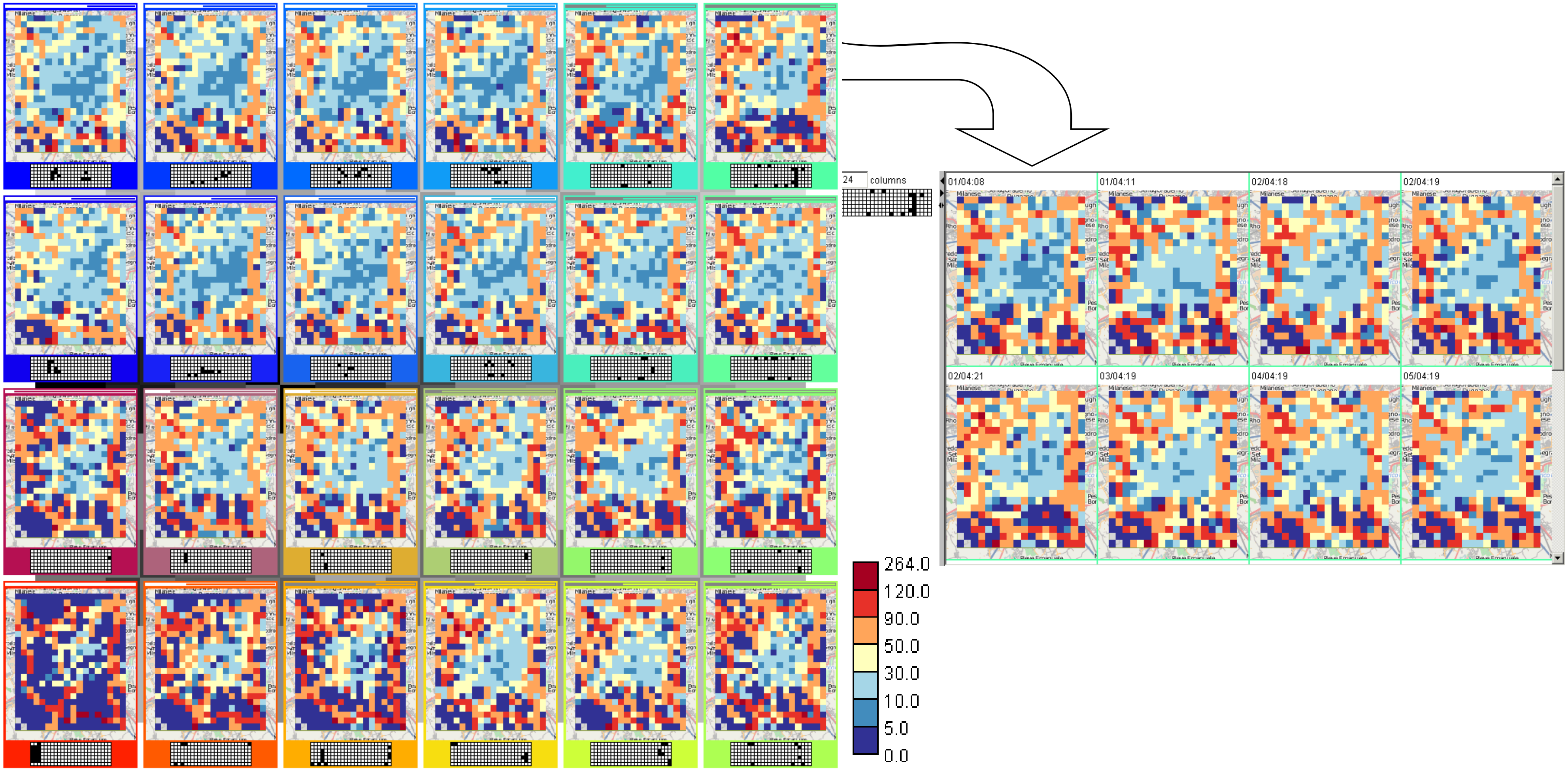
Time-in-Space SOM: details



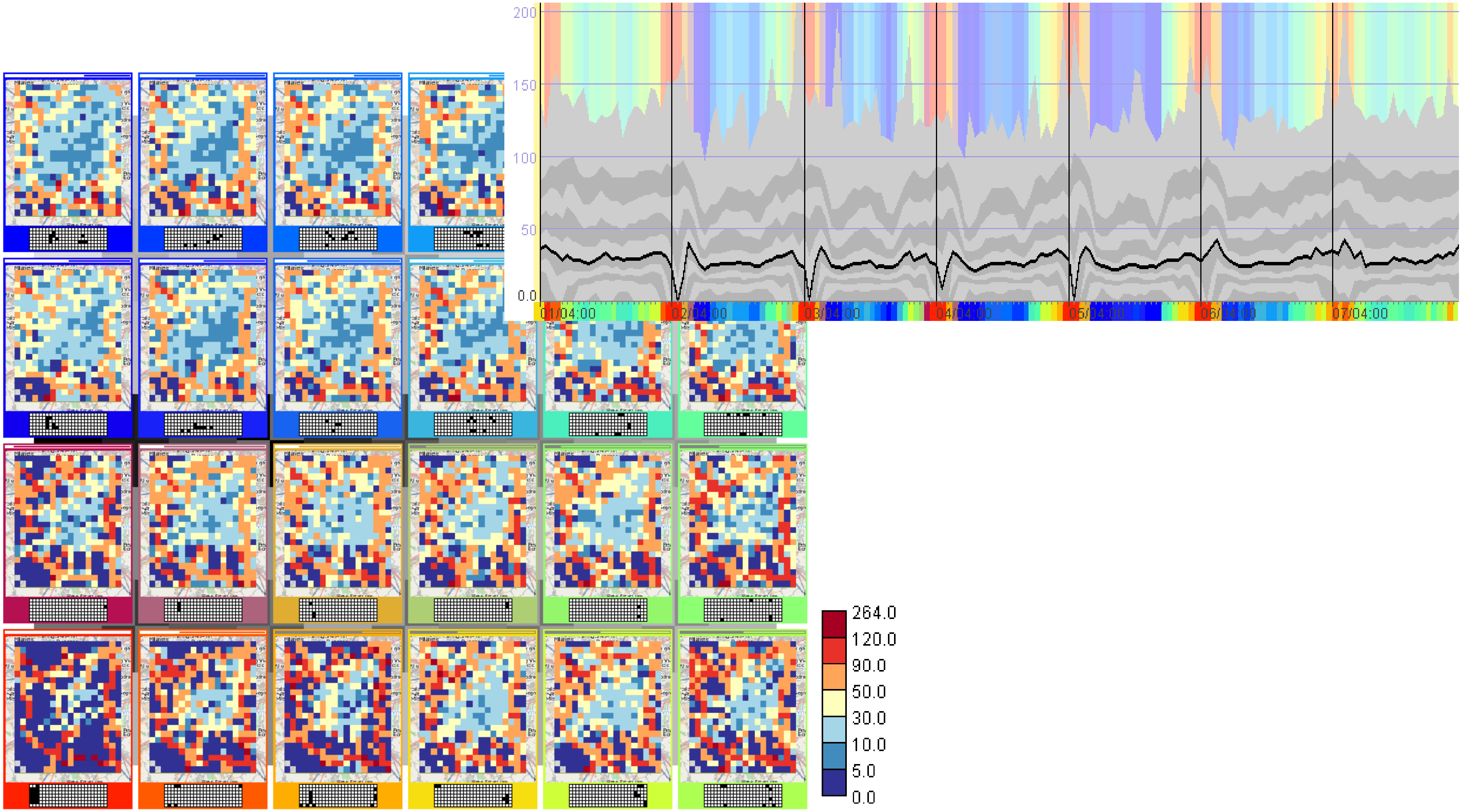
Space-in-time SOM (index images)



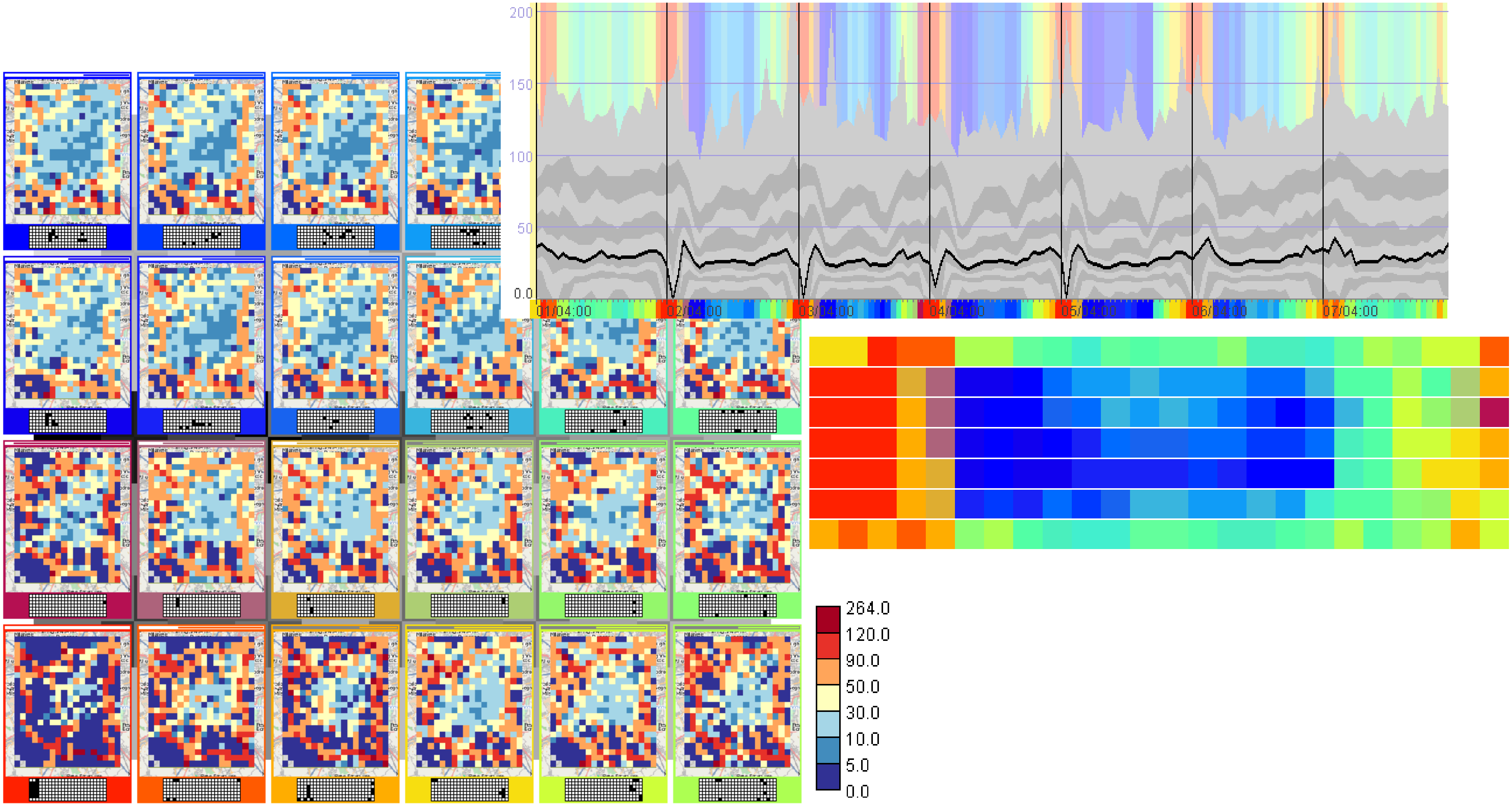
Space-in-time SOM (index & feature images)



Space-in-time SOM (colours of time intervals)



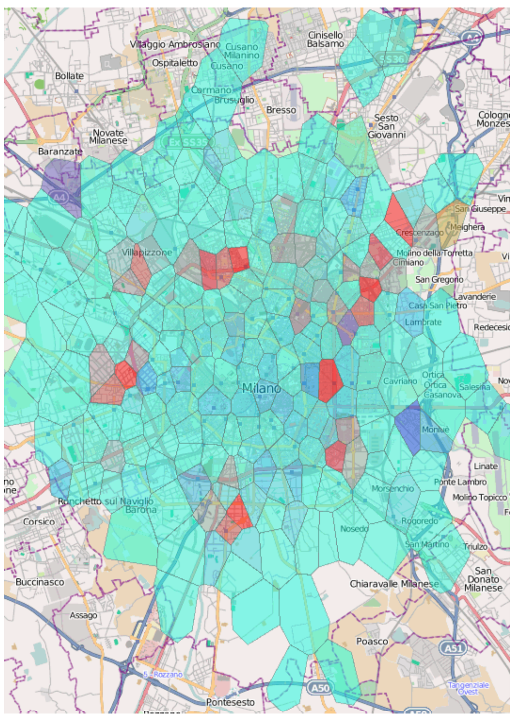
Space-in-time SOM (colours of time intervals)

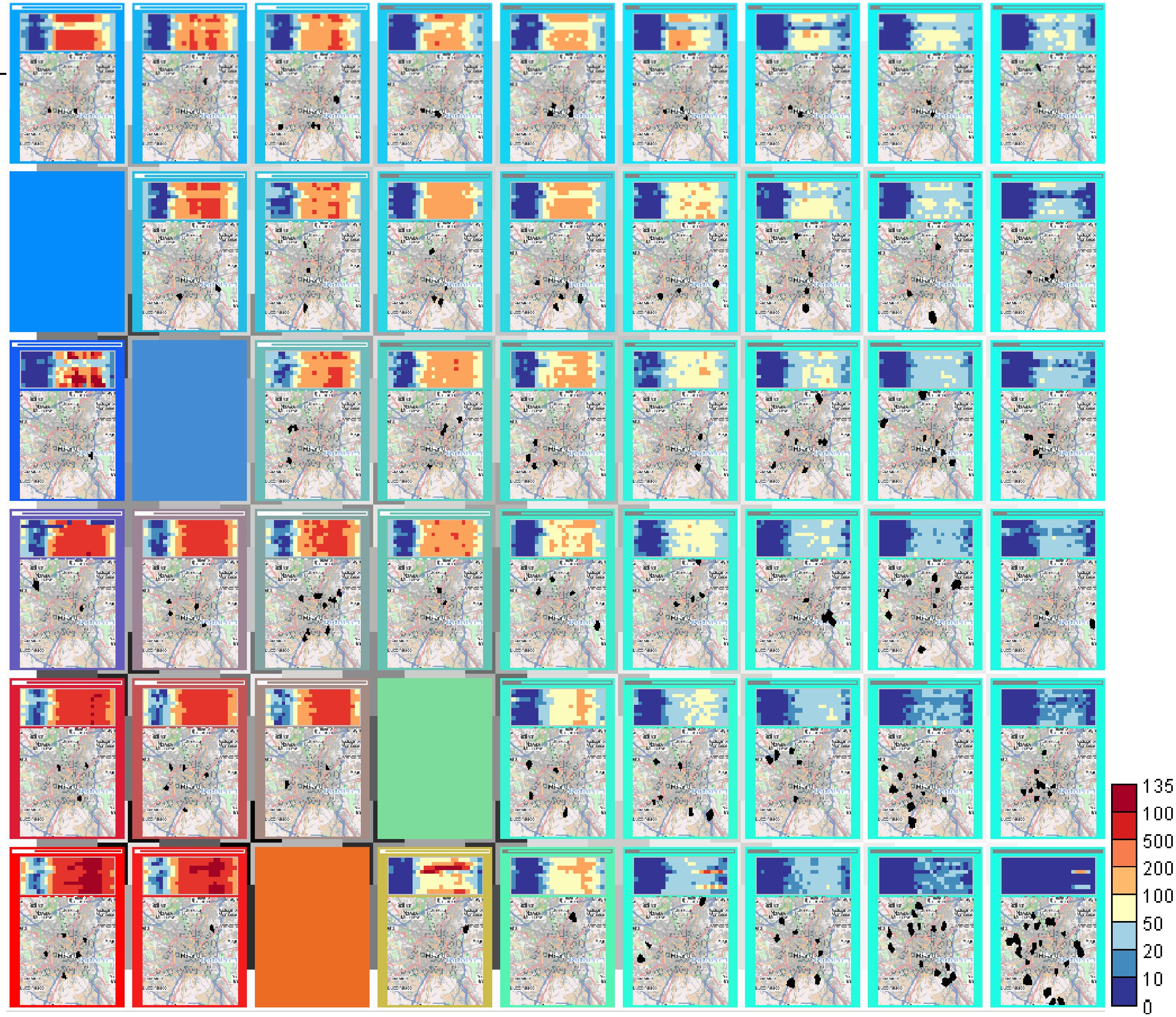


Detect the expected

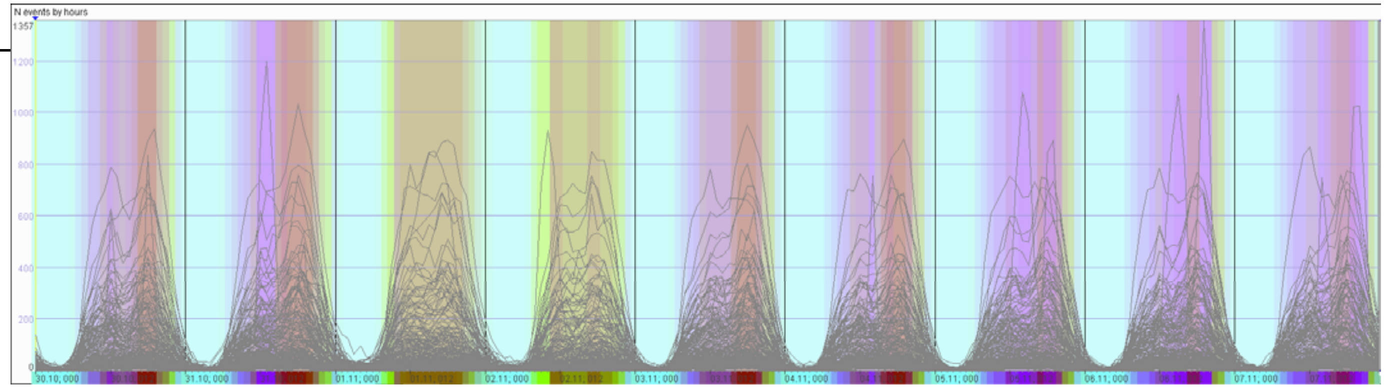
- One more data set about Milan:
  mobile phone calls for 9 days aggregated by hours and regions
- We expect high periodicity and clear regionalization of the calling activity

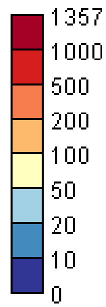
Time-in-Space SOM of mobile phone calls

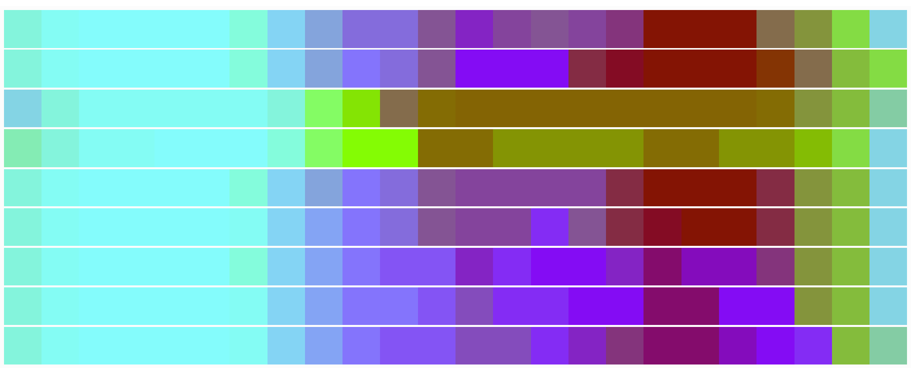


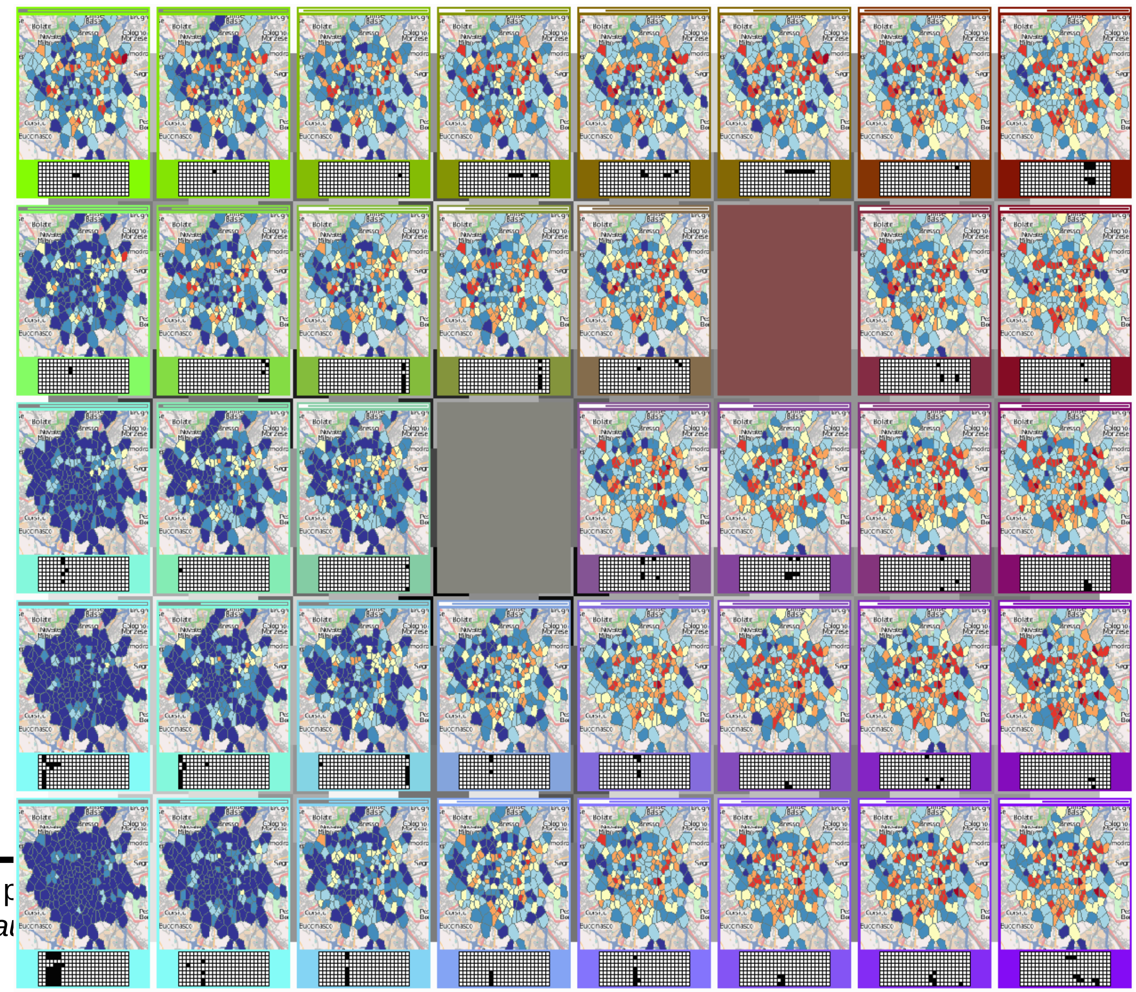


Space-in-Time SOM of mobile phone calls









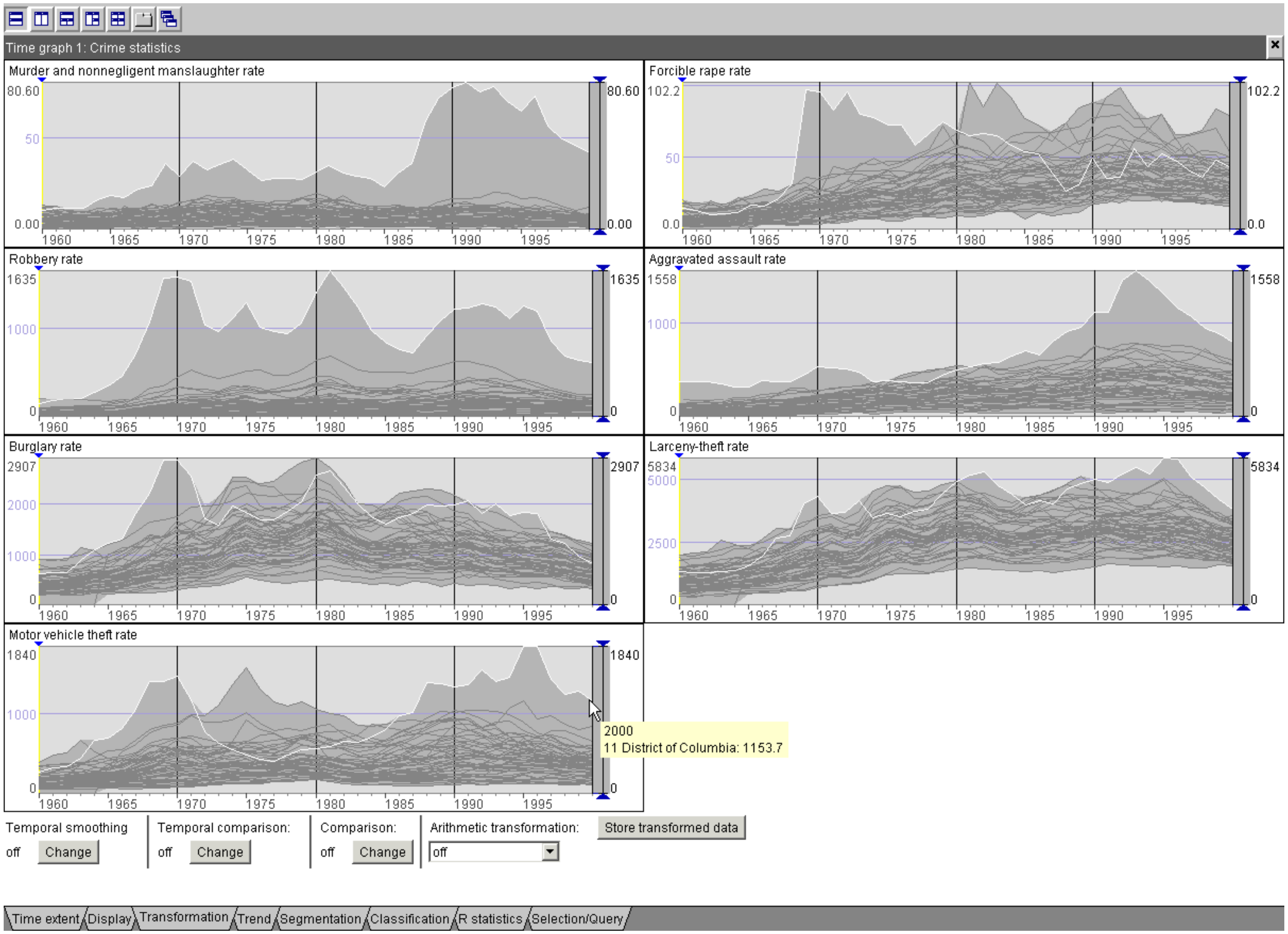
Discover the unexpected

- USA crime statistics
  - 7 crime attributes
  - 52 states
  - 41 years

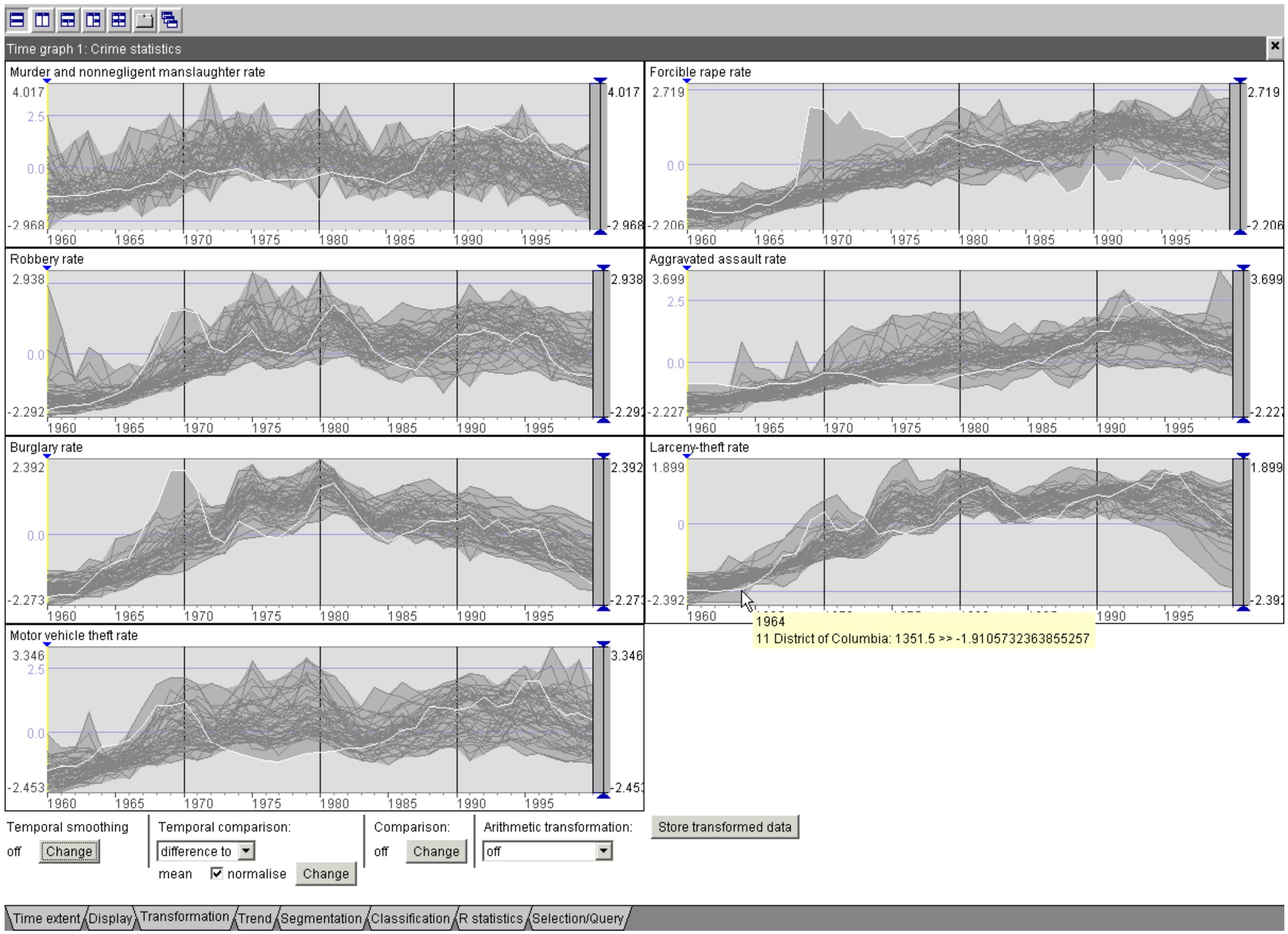
Problems:

- Different ranges of attributes
- Outliers

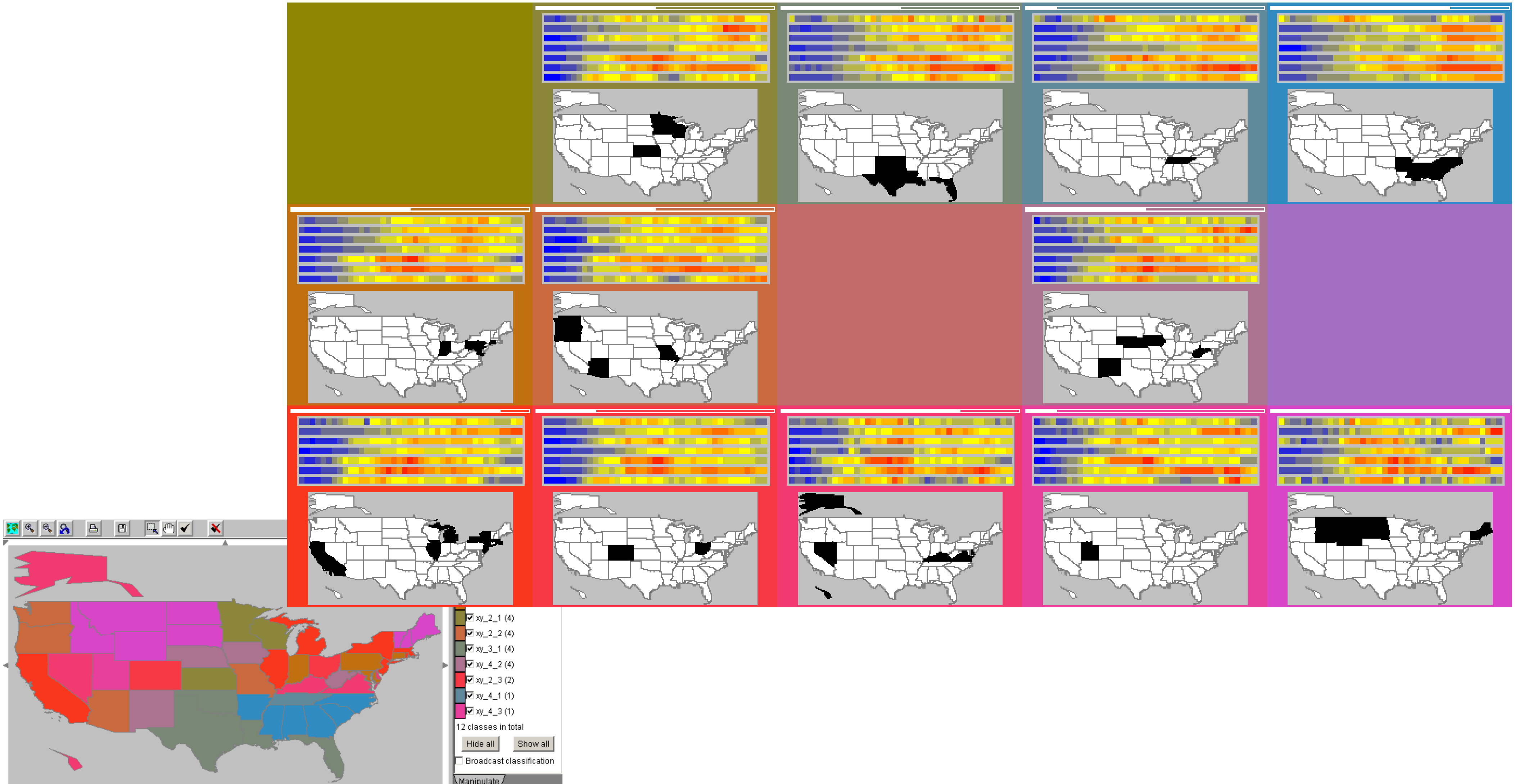
Crime attributes



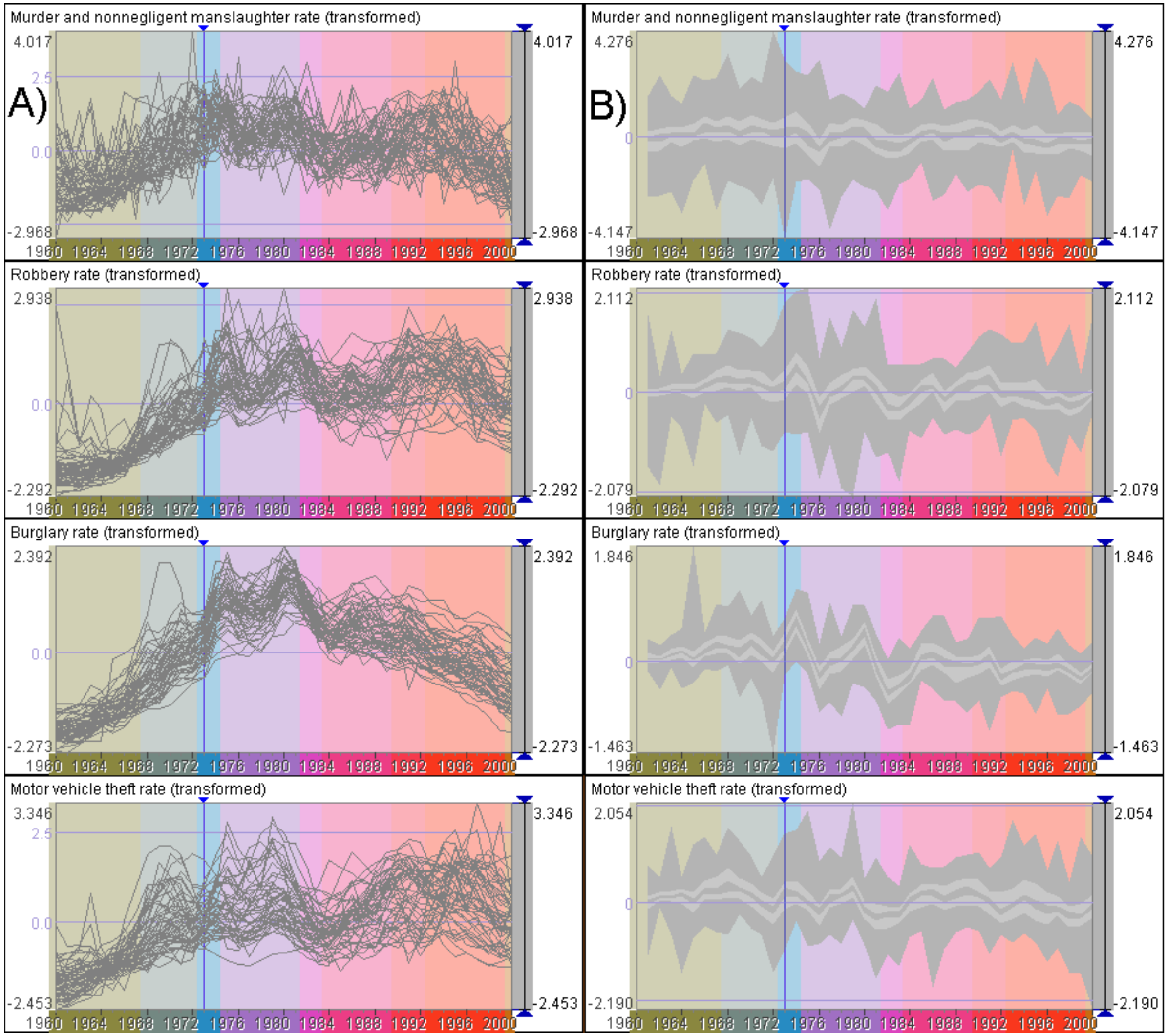
Crime attributes: normalized



Similarity of states by crime dynamics



Similarity of time periods by situations



Conclusions

- Interactive and animated maps and graphs are not sufficient for analyzing large and complex space-time data. Visual methods need to be augmented by computations.
- With space-in-time and time-in-space SOMs we consider data from two different perspectives:
  - places grouped by similar attribute dynamics
  - time intervals grouped by similar spatial distributions of attribute values
- Colours of the groups reflect their similarity
- Case studies demonstrate the value of the approach
  - for detecting the expected
  - for discovering the unexpected
- Current work: non-tabular data, other clustering/projection methods

Live demo is possible