Spatial Statistics & R

Our recent developments

C. Vega Orozco, J. Golay, M. Tonini, M. Kanevski

Center for Research on Terrestrial Environment
Faculty of Geosciences and Environment
University of Lausanne

June 26, 2013

UNIL | Université de Lausanne
Center for Research on Terrestrial Environment

FNSNF
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

Outline

Introduction

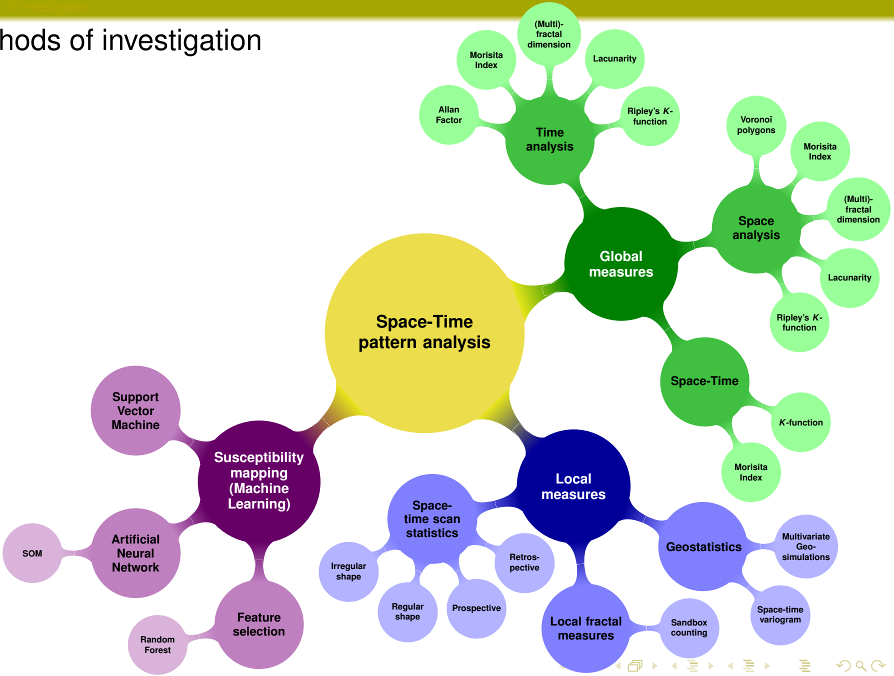
Spatial Statistics

- It concerns the analysis of how spatial objects and its attributes are distributed in space and space-time.
- Three main branches of spatial statistics are commonly distinguished: geostatistics, lattice statistics and spatial point pattern analysis.
- **Spatial point pattern analysis:** distribution of point objects in space and analysis and modelling of the associated spatial pattern.

Motivation

- Many environmental, socio-economic and other data can be treated as stochastic point processes.
- Frequently, such events exhibit scaling behaviour indicating clustering of their spatial point pattern.

Methods of investigation



Outline

Point Processes

- The events are represented by **points** (geographical coordinates) and **marks** (attributes).
- It aims at analysing the geometrical structure of patterns formed by objects that are distributed in 1-, 2- or 3-dimensional space.
- Analyses concern the degree of clustering of the points and the spatial scale at which these operate.
- It provides information on the geometrical properties of the point structure and on underlying processes that have caused the patterns.
- Methods:
  - Topological measures
  - Statistical measures: the Morisita index
  - (Multi)Fractal measures: the box-counting method and the generalized Rényi dimensions

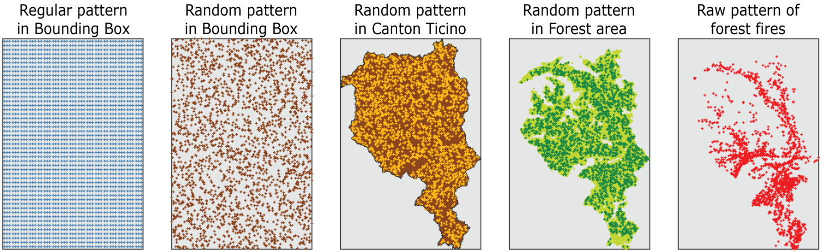
Validity Domains & Simulations

The concept of ***Validity Domain*** (VD) is applied to take into account the natural constraints of the geographical space.

Within the VD, it is possible to simulate random point patterns (e.g. CSR or other patterns with known properties).

These patterns are compared to the real phenomena allowing quantifying the real clustering.

Example of simulations in VD:

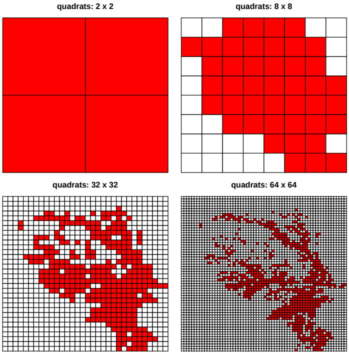


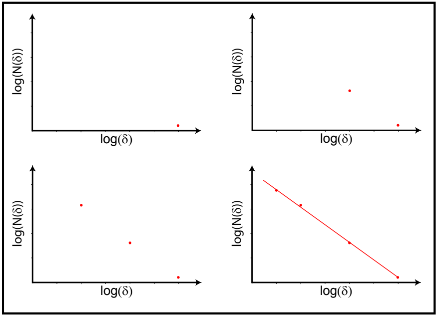
Outline

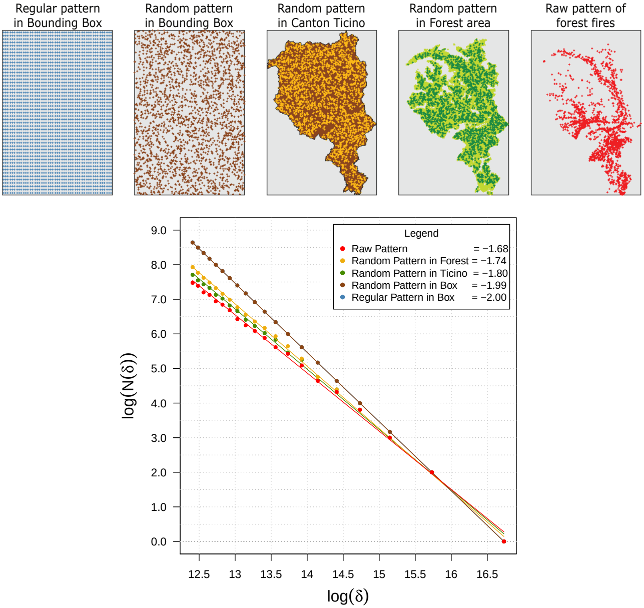
Fractal Dimension: the box-counting method

A regular grid of boxes of size δ is superimposed on the study area and the number of boxes, $N(\delta)$, necessary to cover the point set is counted.
Then, the size of the boxes is reduced and the number of boxes is counted again. The algorithm goes on until a minimum δ size is reach.
For a fractal point pattern, the scales (δ) and the number of boxes ($N(\delta)$) follow a power law:

$$N(\delta) \propto \delta^{-df_{box}}$$



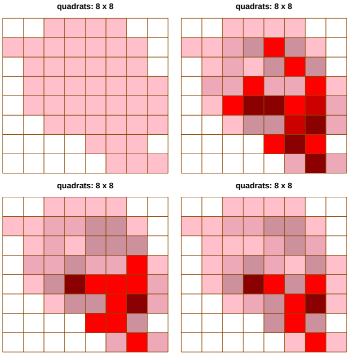


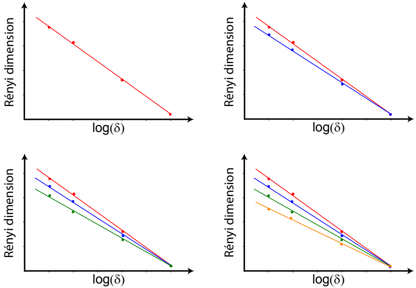

Regular pattern in Bounding Box
Random pattern in Bounding Box
Random pattern in Canton Ticino
Random pattern in Forest area
Raw pattern of forest fires
Legend
Raw Pattern = −1.68
Random Pattern in Forest = −1.74
Random Pattern in Ticino = −1.80
Random Pattern in Box = −1.99
Regular Pattern in Box = −2.00
log(N(δ))
log(δ)
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
12.5
13
13.5
14
14.5
15
15.5
16
16.5

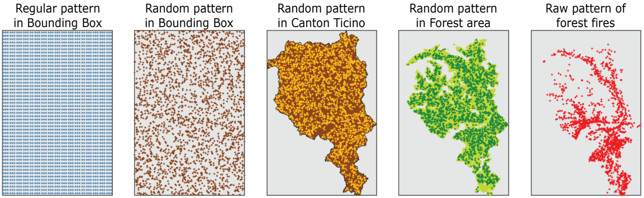
Multifracal measures: the generalized Rényi dimensions

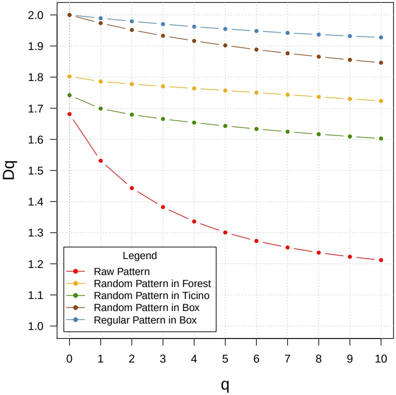
A regular grid of boxes of size δ is superimposed on the study area. Let $N(\delta)$ be the number of non-overlapping boxes of size δ needed to cover the fractal and $p_i(\delta)$ the mass probability function within the ith box.
The generalized Rényi dimensions, D_q, is computed through the parameter q:

$$D_q = \frac{1}{(1-q)} \lim_{\delta \to 0} \frac{\log(\sum_{i=1}^{N(\delta)} p_i(\delta)^q)}{\log(1/\delta)}$$






Regular pattern in Bounding Box
Random pattern in Bounding Box
Random pattern in Canton Ticino
Random pattern in Forest area
Raw pattern of forest fires

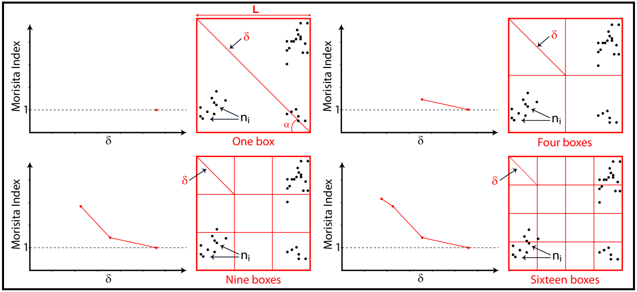

2.0
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
Dq
0
1
2
3
4
5
6
7
8
9
10
q
Legend
Raw Pattern
Random Pattern in Forest
Random Pattern in Ticino
Random Pattern in Box
Regular Pattern in Box

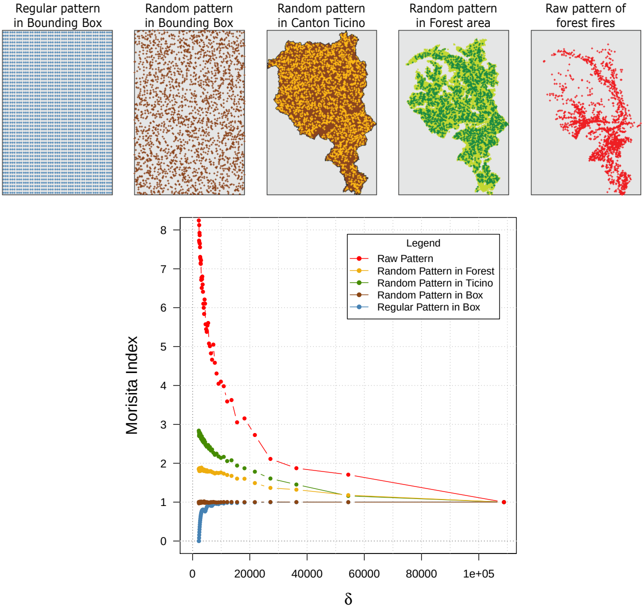
The Morisita index

It measures how many times more likely it is to randomly select two points belonging to the same box than it would be if the points were randomly distributed.

The Morisita index, I_δ, for a chosen box size δ is computed as follows:

$$I_\delta = Q \frac{\sum_{i=1}^{Q} n_i(n_i - 1)}{N(N-1)}$$




Regular pattern in Bounding Box
Random pattern in Bounding Box
Random pattern in Canton Ticino
Random pattern in Forest area
Raw pattern of forest fires
Morisita Index
δ
0
1
2
3
4
5
6
7
8
0
20000
40000
60000
80000
1e+05
Legend
Raw Pattern
Random Pattern in Forest
Random Pattern in Ticino
Random Pattern in Box
Regular Pattern in Box

Outline

Conclusions

- The clustering measures allowed characterizing the spatial pattern of complex point distributions such as environmental data.
- Simulations within validity domains allowed quantifying the real clustering. No need for statistical significance tests.
- R software allows creating new functions and developing and customising existing functions.
- Developed functions:
  - Morisita index and k-Morisita index
  - Multifractal dimensions: generalized Rényi dimensions
  - Multifractal dimensions: multifractal singularity spectrum
  - Box-counting fractal dimension
  - Allan factor
  - Envelopes for CSR simulation patterns and permutations

Thank you for your attention!

Acknowledgements

This research was partly supported by the Swiss NFS project No. 200021-140658, "Analysis and modelling of space-time patterns in complex regions".