Spatial patterns analysis of environmental data using R

Authors

- **Carmen Vega**, University of Lausanne, *Switzerland*
- **Jean Golay**, University of Lausanne, *Switzerland*
- **Marj Tonini**, University of Lausanne, *Switzerland*
- **Mikhaïl Kanevski**, University of Lausanne, *Switzerland*

KEYWORDS : point process, spatial patterns, forest fire occurrence, K-function, Morisita index, fractal measures, multifractal, box-counting

Many environmental phenomena can be study as stochastic point processes where events are represented by their spatial locations (X, Ycoordinates) within a specified geographical region. In this regard, taking into account the spatial characteristics of the environmental data, exploratory spatial data analysis methods are used to discover patterns of spatial association (spatial clustering). Thus, clustering analysis can reveal information about the patterns of the underlying process and their relationship with the phenomenon under study.

The present paper presents some exploratory spatial data analysis tools implemented in R software to assess the spatial patterns of the forest fire distribution in Switzerland. In this research we considered 2,402 georeferenced forest fires ignition-points occurring from 1969 to 2008 in Canton Ticino. This Canton is located in the Southern Swiss Alps and it is the most fire-prone region in Switzerland. The applied clustering measures are Morisita index, fractal and multifractal dimensions (box-counting) and Ripley's *K*-function. These algorithms enable a global spatial structural

analysis describing the spatial degree of clustering of a point pattern [1] and defining whether the observed events occur randomly, in clusters or in some regular pattern [2].
The spatial variability of forest fires is a very complex process which is conditioned by an intermixture of human, topographic, meteorological and vegetation factors. To compute measures of clustering in complex-shape regions the concept of validity domain is applied to restrict the spatial dimensionality of the phenomenon on the mapping space. Within the validity domain is possible to generate spatially randomly distributed events which structure properties are well known. These properties can be compared to the real phenomena properties [1, 3] and the deviation between these measures (computed using real data and based on randomly generated patterns) can quantify the real clustering.

Each measure is described and executed for the raw data (forest fires in Canton Ticino database) and results are compared to the ones obtained from the reference patterns generated under the null hypothesis of spatial randomness embedded in its validity domain. This comparison enables estimating the degree of the deviation of the real data from the random patterns. It is shown that the relative results are very different from measures computed in regular geometrical spaces (usually considered in the literature) because they include empty spaces.

Computations were carried out using R free software for statistical computing and graphics [4]. R is a free software environment integrating facilities for data manipulation, calculation and graphical display. The R base can be extended via packages available through the Comprehensive R Archive Network (CRAN) which covers a very wide range of modern statistics. More specifically, the spatial point pattern analyses considered in the present study were supported by the spatstat package [5] and customised functions.

Morisita index

Morisita Index is a statistical index used to characterise quantitatively the clustering of point processes, in this case the forest fire events (ignition points). Calculation of Morisita Index consists on dividing the study area into identical Q quadrats of size D and counting the number of events ni falling within each single i [1]. Morisita index is computed as follows: $I_D = Q\ ((S\ n_i\ (n_i - 1))\ /\ N\ (N - 1))$, where N is the total number of points and I_D is Morisita index for a chosen quadrats of size D. In other words, this index measures how many times more likely it is to randomly select two points belonging to the same quadrat than it would be if the points were homogenously distributed [6]. Morisita index is first calculated for a relatively small quadrat's size which is then increased until it reaches a chosen value. If the studied point pattern is homogenously distributed on the studied area, every computed I_D fluctuates around the value of 1; while, if the points are clustered, the empty quadrats at small scales will increase the value of the index [1]. Morisita index computed for the forest fire dataset in Canton Ticino displays values above the unitary value for all considered quadrat's sizes which highlights the clustering of the fire distribution.

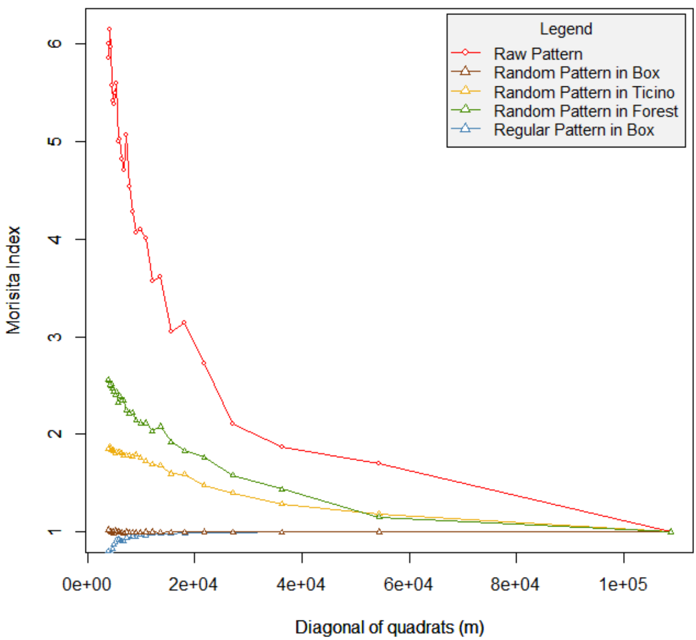

FIGURE 1

Box-counting

The box-counting method consists on superimpose a regular grid of boxes of length d on the region under study and count the number of boxes *N (d)* necessary to cover the whole dataset. This procedure is repeated using different values of *d*. The algorithm goes on until a minimum diameter *d* is reached. The number of occupied boxes increases with decreasing box size, leading to the following power-law relationship: $N(d) = d^{-df_{box}}$, where df_{box} is the fractal dimension measured with the box-counting method [3]. The fractal dimension is estimated as the slope of the linear regression fitting the data of the plot which relates log *(N (d))* to log *(d)*. If the observed events are homogenously/randomly distributed in the studied area the number of boxes of diameter *d* necessary to cover the whole dataset decreases as d^2. Consequently, the box-counting method enables the detection of clustering as a departure from a random situation [3]. The box-counting method diagram for the forest fires in Canton Ticino resulted in a df_{box} less than 2 indicating a high degree of clustering of the events distribution.

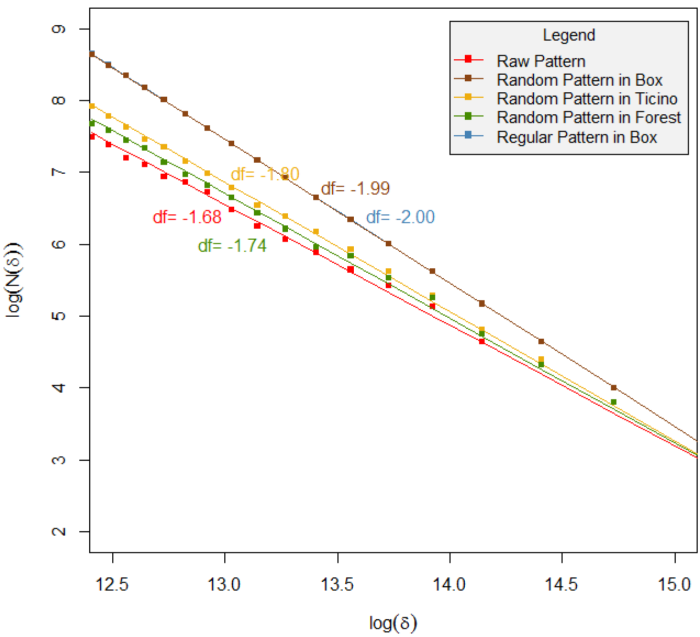

FIGURE 2

Multifractal characterisation

To describe the multi-scaling spatial structure of the forest fires distribution in Canton Ticino, a set of generalized q-dimensions, *Dq*, are measured through a generalization of the box-counting method. The generalized dimensions are defined in terms of the Rényi information of the *q*th order moment of the probability distribution. Let *N(d)* be the number of non-overlapping boxes of equal size *d* needed to cover the fractal (observed events) and $p_i(d)$ the mass probability function in the *i*th box. The generalized dimension *Dq* is computed through the parameter *q* by [7]: $Dq = (1 / (1 - q))$ ´ $lim_{d®0}$ $(log (S (p_i(d)^q)) / log (1 / d))$. The *Dq* diagram is obtained by calculating the generalized dimension varying *q*. When *q=0*, all nonempty boxes are equally weighted and *Dq* is equivalent to the box-counting dimension. For *q>0*, the mass within the boxes gradually gains more importance in the box contribution to the entropy value; therefore, the larger the mass values are in a box, the higher the weight of the box. Thus, *Dq* indicates the scaling of over-

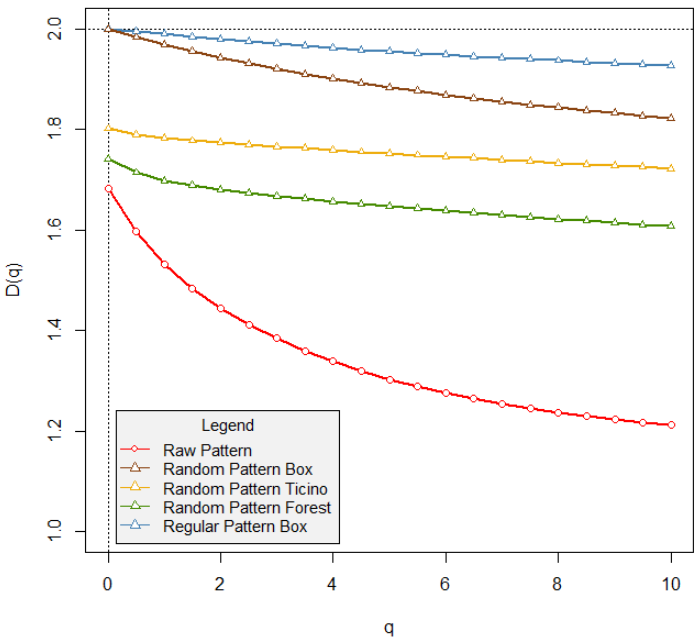

FIGURE 3

dense regions and strong singularities [7, 8]. The *Dq* for each *q*th moment exhibits a non-linear signature revealing the multifractal behaviour of the forest fires distribution in Canton Ticino. For the forest fire real data, *Dq* declines faster than for the generated random patterns, and the departure from these patterns reveals the highly clustering of the events distribution.

Ripley's K-function

The Ripley's *K*-function describes how the interaction between events varies through the space. For a spatial point process, the function is defined as: *K(r) = E(×) / Intensity*, where *E(×)* denotes the expected number of further events within a distance *r* of an arbitrary event [2, 9]. The analysis tests for complete spatial randomness (CSR) which implies no interactions among points; that is, the probability of the occurrence of fire at any point is independent of each other fire event. Subsequently, departure from CSR indicates clustering (aggregation) or dispersion (regularity or inhibition) of fire occurrences.

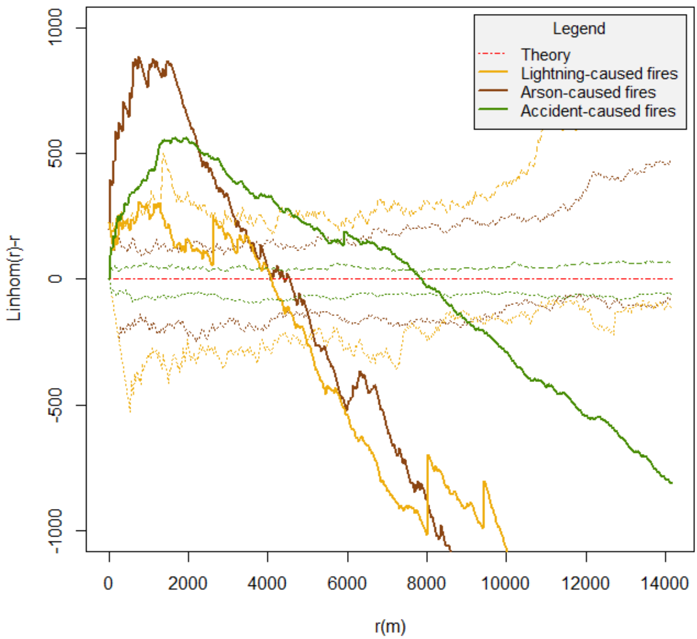

FIGURE 4

This is evaluated against a confidence interval constructed by performing n simulations of the events under the CSR hypothesis. This nonparametric spatial statistics tool can afterwards support nearest-neighbour analyses to detect areas with higher fire incidences. Results from this function revealed that human-caused fires are not randomly distributed presenting a clustering at scales up to 8km with a maximum degree of clustering at 2km.

Conclusions

Exploratory spatial data analysis tools were introduced to characterize the degree of clustering of the forest fire events in Canton Ticino. The concept of a validity domain was implemented. Each measure detected a quite higher degree of clustering of the forest fire network in the study area at every scale. This work demonstrates that in the case of complex regions relative values between measures are more important than absolute values. Further development of the research will consist in the custom implementation of the algorithms in the *R* environment to be functional for the analysis of multi-dimensional and multivariate point patterns.

[1] KANEVSKI M., MAIGNAN M. (2004). *Analysis and modelling of spatial environmental data*. Lausanne, EPFL Press.

[2] RIPLEY B.D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society.* Series B (Methodological), vol. 39 (2): 179-212.

[3] TUIA D., KANEVSKI M. (2008). *Environmental monitoring network characterization and clustering*. In “Advanced mapping of environmental data. Geostatistics, machine learning and bayesian maximum entropy”. Kanevski M. (Editor), iSTE and Wiley, 19-46.

[4] R DEVELOPMENT CORE TEAM. (2012). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.

[5] BADDELEY A., TURNER R. (2005). Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, vol. 12 (6): 1-42.

[6] HURLBERT S.H. (1990). Spatial Distribution of the Montane Unicorn. Oikos, vol. 58: 257-271.

[7] SEURONT L. (2010). *Fractals and multifractals in ecology and aquatic science.* Boca Raton, CRC Press.

[8] ADJALI I., APPLEY S. (2001). *The multifractal structure of the human population distribution*. In Modelling scale in geographical information science, edited by Tate N.J. and Atkinson P.M., West Sussex, John Wiley & Sons Ltd, 69-85.

[9] HERING A.S., BELL C.L., GENTON M.G. (2009). Modeling spatio-temporal wildfire ignition point patterns. *Environmental and Ecological Statistics*, vol. 16: 225-250.