Spatial Statistics 18 (2016) 352–362

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

On the similarity analysis of spatial patterns

M.V. Alba-Fernández [a,*], F.J. Ariza-López [c], M. Dolores Jiménez-Gamero [b], J. Rodríguez-Avi [a]

[a] Departamento de Estadística e Investigación Operativa, Universidad de Jaén, Jaén, Spain
[b] Departamento de Estadística e Investigación Operativa, Universidad de Sevilla, Sevilla, Spain
[c] Departamento de Ingeniería Cartográfica, Geodésica y Fotogrametría, Universidad de Jaén, Jaén, Spain

ARTICLE INFO

Article history:
Received 16 January 2016
Accepted 25 July 2016
Available online 18 August 2016

Keywords:
Space-filling curve
Multinomial distribution
Simulation

ABSTRACT

The evaluation of the spatial similarity of two observed point patterns is an important issue in spatial data quality assessment. In this work we propose a formal procedure that takes advantage of the joint use of space-filling curves and the multinomial model in order to establish a statistical test to compare spatial point patterns. In this mix, the space-filling curves offer a mechanism to order the 2D, 3D or n-D space and the multinomial distribution the statistical approach for testing homogeneity. A simulation method is proposed in order to analyze the applied performance of this idea.

1. Introduction

The understanding of spatial point patterns is one of the major challenges of geographical analysis and is important in many sciences (e.g. biogeography, crop sciences, ecology, geology, etc.). The evaluation of the spatial similarity of two observed point patterns is an important issue in spatial data quality assessment (Ariza-López, 2002) where sampling is the way a point pattern is obtained. It is usual that an estimate of an attribute or property (e.g. concentration of a mineral, positional accuracy, presence, etc.), is derived from a sample of points under some spatial distributions (theoretical or observed), for instance, different geological structures, see Fig. 1, can determine presence/absence and concentration of a mineral. We think that, apart from any error or uncertainty consideration about an

* Corresponding author.
E-mail addresses: mvalba@ujaen.es (M.V. Alba-Fernández), fjariza@ujaen.es (F.J. Ariza-López), dolores@us.es (M.D. Jiménez-Gamero), jravi@ujaen.es (J. Rodríguez-Avi).

http://dx.doi.org/10.1016/j.spasta.2016.07.004
2211-6753/

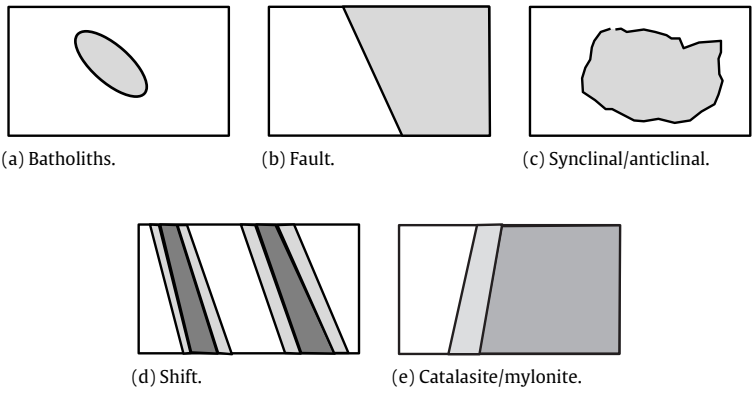

Fig. 1. Examples of some geological structures that can influence spatial patterns: (a) batholiths, (b) fault, (c) synclinal/anticlinal, (d) shift, (e) cataclasite/mylonite.

attribute or property estimation, we must confirm the underlying hypothesis about the location of points events, in our case: the spatial point distribution and its relation to an observed spatial point pattern or the similarity between two given point patterns obtained by sampling.

A spatial point pattern has been defined as a set of locations, irregularly distributed within a region of interest, which have been generated by stochastic mechanisms point process (Diggle, 1983). A fundamental property of a point process is its intensity (Perry et al., 2006). The simplest spatial point process is complete spatial randomness (CSR), where intensity remains constant, point events are uniformly distributed over space, and no interaction exists. Although there are many other point processes (e.g. aggregative, repulsive and marked), the CSR is one of the most extensively used processes. However, the methodology we are proposing is independent of the true point process, and it does not matter if it is known or unknown.

Two common methods in use to investigate discrete point events are the distance- and area-based tests. In the case of distance-based tests, we use whatever distance (e.g. Euclidean) between two events in order to determine a random, clustered, or uniform spatial pattern to the points (Bailey and Gatrell, 1995). On the other hand, in the case of area-based tests, we count the number of point events within a predefined spatial area (e.g. a quadrant, a census unit, etc.) (Andresen, 2009).

Our proposal is very different to other techniques such as Ripley's K function (Ripley, 1977). The K function characterizes point processes at many distance scales and allows the detection of different behaviors (e.g. random. clustering, inhibition) (Freeman and Ford, 2002); and it can be used to test different specific point patterns against predefined models—spatial point process (e.g. homogeneous Poisson process—complete spatial randomness, Matern hard-core process, Strauss process, etc.). For the correct application of these functions several assumptions must be accepted that are not necessarily fully met in reality, and if conditions are not met, the output may be incorrect (Bolibok, 2008).

In this setting, we develop an area-based test centered on the counting of positional events, which does not assume a theoretical model for the spatial distribution of the events (spatial distribution free). So it can be applied for the comparison of two spatial point patterns samples (e.g. two control test samples, two field works, presence at two different times, etc.), with independence of sample sizes.

We centered our work on the similarity of spatial pattern of points obtained from two samples. So the core of the question we address is how we can decide if two spatial patterns are similar or not. For this goal, we propose a formal procedure that allows us to determine if two samples of points come from the same spatial distribution. This method makes use of the space-filling curves as a tool to order the space together with the multinomial distribution obtained by counting the points in the resultant grid. This way, the similarity between sampled spatial distributions is equivalent to homogeneity of multinomial distributions. We use a simulation process in order to demonstrate the functionality of our proposal.