Spatial Statistics 16 (2016) 118–136

Contents lists available at ScienceDirect

Spatial Statistics

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

Testing the random field model hypothesis for random marked closed sets

Antonín Koubek [a,*], Zbyněk Pawlas [a], Tim Brereton [b], Björn Kriesche [b], Volker Schmidt [b]

[a] Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University in Prague, Sokolovská 83, 18675 Praha 8, Czech Republic

[b] Institute of Stochastics, Ulm University, 89069 Ulm, Germany

ARTICLE INFO

Article history:
Received 13 October 2015
Accepted 1 March 2016
Available online 11 March 2016

Keywords:
Random marked closed set
Random field model
Envelope test
Subsampling
Monte Carlo test
Mark-weighted K-function

ABSTRACT

When developing statistical models, it is of fundamental importance to decide if the various components are independent of one another, preferably using a formal statistical test. Non-parametric versions of such tests are particularly useful, as they do not require extensive *a priori* knowledge about the underlying models. In this paper, we develop such tests for random marked closed sets, which have many applications in spatial statistics. More precisely, we investigate two approaches to testing if the marks are independent of the closed set. Both approaches are based on second-order characteristics of random marked closed sets. The first approach uses a global rank envelope test based on the mark-weighted K-function. The second approach uses an asymptotic test developed for marked point processes. We carry out extensive simulation studies to assess the performance of these tests, demonstrating that the global rank envelope test is a better choice. Finally, we apply this test to two real world data sets.

* Corresponding author.
E-mail address: koubekan@karlin.mff.cuni.cz (A. Koubek).

http://dx.doi.org/10.1016/j.spasta.2016.03.001

1. Introduction

The notion of a *random marked closed set* (RMCS) is defined in Ballani et al. (2012) as a random upper semi-continuous function on a random domain (random closed set). Many settings in spatial statistics are naturally described by RMCSs. For example, the closed set could represent forest areas and the mark function could express the local stand density, or the closed set could represent bodies of water and the mark function could represent water quality. In Šedivý et al. (2013), RMCSs are used to investigate metallic materials, with the grain boundary network being represented by a random set marked by the disorientation angles of the grain boundaries.

Marked point processes are special examples of RMCSs. They are used to model spatial data that consist of measurements at irregularly scattered spatial locations; see, e.g., Chiu et al. (2013), Cressie (2015) and Illian et al. (2008). The locations are treated as a realization of a spatial point process (which is a particular example of random closed set) and the associated measurements are treated as the corresponding marks. In practice, the points and the marks are often correlated. For instance, in forestry data, the points (representing tree locations) and marks (representing tree size) could be correlated; see, e.g., the examples in Stoyan and Wälder (2000). A critical task when developing a statistical model of a marked point process is to determine if the values of the marks are independent of the locations of the corresponding points.

The simplest model of a marked point process is an independently marked (or randomly labeled) point process, where the marks are i.i.d. and independent of the points. In order to model correlated marks, Karr (1986) considered marks generated by a random field that is independent of the points. Such marking is called *geostatistical marking* and the corresponding marked point process model is called a *random field model*; cf. Illian et al. (2008, Section 5.1.3) and Schlather et al. (2004). Several methods have been proposed for testing the hypothesis that a marked point process has independent marks; see Illian et al. (2008, Section 7.5.2) and Myllymäki et al. (2015). Tests have also been proposed for determining if a stationary marked point process follows the random field model; see Guan (2006), Guan and Afshartous (2007), Illian et al. (2008, Section 7.5.3), Schlather et al. (2004) and Zhang (2014).

A natural generalization of a marked point process is to consider the labeling of random closed sets. These are studied in Molchanov (1984), where the connected components of the set are labeled by nominal marks. Labeled random closed sets are also studied in Nott and Wilson (2000), where they are used to model multi-phase data, representing the presence of different sulphides in an ore sample. In Ayala and Simó (1995), where the random closed set only has two possible marks (representing degenerated and normal nerve fiber) a test is introduced to determine if the marks are mutually independent and independent of the set.

The notion of a RMCS further generalizes both marked point processes and labeled random closed sets. A fundamental question when using a RMCS to model data is whether the marks are independent of the domain. This situation is an analogue of the random field model for marked point processes. If the random field model hypothesis is satisfied, then the statistical analysis of the RMCS is simplified considerably, since the two components may be investigated separately. If it is not satisfied, then care must be taken to adequately model the nature of the dependency.

In general, when working with data that is assumed to come from a RMCS, little is known *a priori* about the random marks and the random closed set. Indeed, an important application of a test of the random field hypothesis is to determine whether the random marks and random closed set can be studied and modeled separately of one another. Thus, it is important to develop tests that do not rely on detailed knowledge about the models from which the random marks and random closed set come. For this reason, in this paper, we develop two non-parametric tests of the random field model hypothesis. Both approaches are based on second-order characteristics of RMCSs. The first test is a global rank envelope test (see Myllymäki et al., in press) based on the mark-weighted K-function. In order to apply this test, a set of test points needs to be carefully chosen. The second test overlays the random set with an independent point process and then uses a test developed for marked point processes in Guan and Afshartous (2007). We carry out simulations to assess the performance of these tests. These simulations suggest that the global rank envelope test performs best. Finally, we apply our methodology to real world radar data on adjusted hourly precipitation in Germany and to microscopic image data showing the nanostructure of a thin film organic semiconductor.