ARTICLE IN PRESS

Spatial Statistics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Spatial Statistics

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

Adapted kriging to predict the intensity of partially observed point process data

Edith Gabriel [a,b,d,*], Florent Bonneu [a], Pascal Monestiez [b,d], Joël Chadœuf [c]

[a] *Avignon University, LMA EA 2151, 84000 Avignon, France*
[b] *INRA, BioSP, 84000 Avignon, France*
[c] *INRA, Statistics, GAFL, UR 1052, 84000 Avignon, France*
[d] *Zone Atelier Plaine & Val de Sèvre, CEBC-CNRS, 79360 Villiers-en-Bois, France*

ARTICLE INFO

Article history:
Received 15 December 2015
Accepted 25 May 2016
Available online xxxx

Keywords:
Intensity estimation
Point process
Prediction
Spatial statistics

ABSTRACT

We consider a stationary and isotropic spatial point process whose realization is observed within a large window. We assume it to be driven by a stationary random field U. In order to predict the local intensity of the point process, $\lambda(x|U)$, we propose to define the first- and second-order characteristics of a random field, defined as the regularized counting process, from the ones of the point process and to interpolate the local intensity by using a kriging adapted to the regularized process.

1. Introduction

In many projects the study window is too large to extensively map the intensity of the point process of interest since observation methods may be available at a much smaller scale only. That is for instance the case when studying the spatial repartition of a bird species at a national scale, while the observations are made in windows of few hectares. The intensity must then be estimated from data issued out of samples spread in the study window, and hence, from a partial realization of the point process in this window.

In the following, we consider a stationary and isotropic point process, Φ, which we assume to be driven by a stationary random field, U. We define the local intensity of Φ by its intensity conditional to

* Corresponding author at: Avignon University, LMA EA 2151, 84000 Avignon, France.
E-mail address: edith.gabriel@univ-avignon.fr (E. Gabriel).

http://dx.doi.org/10.1016/j.spasta.2016.05.004
2211-6753/

Please cite this article in press as: Gabriel, E., et al., Adapted kriging to predict the intensity of partially observed point process data. Spatial Statistics (2016), http://dx.doi.org/10.1016/j.spasta.2016.05.004

the random field U. We denote it $\lambda(x|U)$. A simple example of such a process Φ is the Thomas process which is a Poisson cluster process where the cluster centers (parents) are assumed to be Poisson and the offsprings are normally distributed around the parent point. This process is stationary and the local intensity corresponds to the intensity of the inhomogeneous Poisson process of offsprings, i.e. the conditional intensity given the parent process. We will refer to the estimation of the local intensity when we want to the know it at point locations lying in the observation window of the point process, and to its prediction when point locations are outside the observation window.

Usually, when estimating a non-constant intensity, we observe the full point pattern within a window and we want to know its local changes over a given mesh. This issue has been addressed in several ways: kernel smoothing, see Silverman (1986) and Guan (2008) in presence of covariates, and van Lieshout (2012) for a general class of weight function estimators that encompasses both kernel and tessellation based estimators; or parametric methods; see for instance (Illian et al., 2008) for a review. A recurrent and remaining question in these approaches is which bandwidth/mesh should we use? This has been addressed by using cross-validation (Härdle, 1991) or double kernel (Devroye, 1989).

In contrast to the previous methods which look at the intensity changes inside the observation window, our main interest lies in predicting the intensity outside the observation window, all the more when it is not connected as it frequently happens when sampling in plant ecology. To predict the intensity we could use (Tscheschel and Stoyan, 2006)'s reconstruction method based on the first- and second-order characteristics of the point process. Once the empirical point pattern predicted within a given window, one can get the intensity by kernel smoothing. As it is a simulation-based method, it requires long computation times, especially when the prediction window is large and/or the point process is complex. As alternative method, few authors model the point pattern by a point process with the intensity driven by a stationary random field. In Diggle and Ribeiro (2007) and Diggle et al. (2013), the approach is heavily based on a complete modeling and considers a log-Gaussian model. The parameter estimation, the intensity estimation and its prediction outside the observation window are obtained using a Bayesian framework. The method developed in Monestiez et al. (2006) and Bellier et al. (2013) is close to classical geostatistics. Basically, it consists of counting the number of points within some grid cells, computing the related empirical variogram and theoretically relating it to the one obtained from the random field driving the intensity. Then, the variogram is fitted and kriging is used to predict the intensity. Its advantage is that the estimation is only based on its first- and second-order moments so that the model does not need to be fully specified. While this approach requires less hypotheses, the model remains constrained within the class of Cox processes. Moreover, the mesh size is arbitrary defined. van Lieshout and Baddeley (2001) developed, for a wider class of parametric models, a Bayesian approach for extrapolating and interpolating clustered point patterns.

Here, we propose to interpolate the local intensity by an adapted kriging, where the kriging weights depend on the local structure of the point process. Hence, our method uses all the data to locally predict at a given point, which it is not the case of most of kernel methods. It also uses the information at a fine scale of the point process, which it is not the case in geostatistical approaches. Furthermore, it does not require a specific model but only (an estimation of) the first- and second-order characteristics of the point process.

In Section 2 we define the regularized process as a random field of point counts on grid cells and we link up the mean and variogram of this random field to the intensity and pair correlation function of the point process. The kriging weights, the related interpolator and its properties are presented in Section 3 as well as the optimal mesh of the interpolation grid. In Section 4 we use our kriging interpolator to estimate and predict the intensity of Montagu's Harriers' nest locations in a region of France. In Section 5, we discuss the influence of the mesh and the rate and shape of unobserved areas on the statistical properties of our kriging interpolator from numerical results.

2. Linking up characteristics of two theories

2.1. About geostatistics

For any real valued random field $Z(x)$, $x \in \mathbb{R}^2$, the first-order characteristic is the mean value function: $\mathbb{E}[Z(x)] = m(x)$ and the second-order characteristics are classically described in

Please cite this article in press as: Gabriel, E., et al., Adapted kriging to predict the intensity of partially observed point process data. Spatial Statistics (2016), http://dx.doi.org/10.1016/j.spasta.2016.05.004