Spatial Statistics 14 (2015) 472–490

Contents lists available at ScienceDirect

Spatial Statistics

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

Semi-parametric pairwise inference methods in spatial models based on copulas

Jean-François Quessy [a,*], Louis-Paul Rivest [b], Marie-Hélène Toupin [b]

[a] *Département de Mathématiques et d'Informatique, Université du Québec à Trois-Rivières, Trois-Rivières, Canada*
[b] *Département de Mathématiques et Statistique, Université Laval, Québec, Canada*

A R T I C L E I N F O

Article history:
Received 5 August 2014
Accepted 24 August 2015
Available online 28 September 2015

Keywords:
Copula
Pairwise likelihood
Rank-based methods
Spatial dependence
Spatial interpolation

A B S T R A C T

In this paper, semi-parametric models based on copulas are considered for the modeling of stationary and isotropic spatial random fields. To this end, a general family of multivariate distributions is introduced in which the dependence structure between any finite sets of locations is modeled via a copula and where the strength of the relationships between any two locations is controlled by a link function. Because the density of most of the multivariate spatial copulas is untractable, it is proposed that inferential procedures for these models be based on a pairwise approach taking into account the bivariate densities only. Specifically, a rank-based estimation procedure using the so-called pairwise likelihood is proposed and a semi-parametric spatial interpolation method for the prediction at un-sampled locations is developed; both methods are investigated with the help of simulated spatial random fields. The usefulness of the newly introduced tools is illustrated on the Meuse river data.

* Corresponding author.
E-mail addresses: jean-francois.quessy@uqtr.ca (J.-F. Quessy), louis-paul.rivest@mat.ulaval.ca (L.-P. Rivest), marie-helene.toupin.1@ulaval.ca (M.-H. Toupin).

http://dx.doi.org/10.1016/j.spasta.2015.08.002
2211-6753/

1. Introduction

The modeling of spatial data is an important topic that has retained a lot of attention in the scientific literature. Such data occur in various areas including geostatics, meteorology and astronomy. The mathematical basis of most of the statistical procedures is the assumption of an isotropic random field with continuous spatial index (usually defined on $\mathbb{R}^2$ or $\mathbb{R}^3$) for which information is available for a finite set of locations only. Under this setup, spatial dependence between stations is typically modeled by using the popular variogram and the related covariance function. For more details, see for example Ripley (1981) and Cressie (1993).

An approach that gained in popularity recently is the modeling of spatial dependence using copulas. Briefly stated, a copula captures all the dependence features in a multidimensional random vector. Similarly as in *standard* multivariate modeling, the use of copulas in a spatial framework is very attractive for two main reasons: on one part, it allows for the construction of flexible models appearing as interesting alternatives to traditional ones, and on the other hand, dependence can be interpreted without having to specify the marginal distribution functions.

The first work that considers using copulas as an alternative to variogram-based methods is apparently due to Bàrdossy (2006); in this paper, copula-based models are built for the finite distributions associated to a random field. Following that idea, spatial interpolation using the full conditional distribution was performed by Bárdossy and Li (2008), Li et al. (2011) and Kazianka and Pilz (2010); this latter work investigates the case of discrete marginal distributions. Some works were most concerned with parameter estimation, for example Kazianka and Pilz (2011) who proposed a Bayesian approach and Bai et al. (2014) using pairwise likelihood in the context of spatially clustered data. More prediction-oriented procedures based on vine copulas have been proposed recently by Gräler and Pebesma (2011), Erhardt et al. (2014a,b) and Gräler (2014). In all these references, the marginal distributions are specified up to unknown parameters.

The main goal of this work is to adopt a less restrictive semi-parametric approach where the marginal distributions are unspecified. Such a framework where marginals are seen as infinite-dimensional nuisance parameters calls for the use of the ranks instead of the original observations. Another feature of the strategy taken here is the systematic use of a pairwise approach for inference; in other words, the statistical procedures that are proposed make use of the information provided by the pairs of locations only. This has the advantage of avoiding the computation of the often untractable multivariate distribution. The specific goals of this article are to (i) describe a general framework for spatial copula models, (ii) investigate the performance of rank-based pairwise likelihood estimators and (iii) perform spatial interpolation at un-sampled locations.

The paper is organized as follows. A general framework for spatial random fields based on copulas is introduced in Section 2. Some special cases including the Normal and non-central chi-square spatial copulas are described in Section 3. The rank-based pairwise likelihood estimator is described in Section 4 and its sample properties are investigated via simulations. The semi-parametric pairwise interpolation method is introduced in Section 5 and its behavior in small samples is studied as well. An illustration showing the usefulness of the introduced tools is given in Section 6 on the Meuse river data. The paper ends with a short discussion, in Section 7. All the technical details are relegated to Appendix A, and additional simulation results are presented in a *Supplementary material* section (see Appendix B).

2. A general copula-based framework for spatial random fields

2.1. General setup

Let $\{Z(\mathbf{x}) \mid \mathbf{x} \in S\}$ be a continuous random field defined on a subset S of $\mathbb{R}^m$. Let also $\{\mathbf{x}_i \in S,\ i = 1, \ldots, n\}$ be a lattice, *i.e.* a collection of locations in S. For example, it can be a set of meteorological stations, or locations where soil samples have been taken. The random field evaluated at these locations is noted $Z_1, \ldots, Z_n$, where $Z_i = Z(\mathbf{x}_i),\ i \in \{1, \ldots, n\}$. Further assume that the random field is stationary, which means that the marginal distributions are the same at each location. In other words, there is a distribution function F_Z such that $\mathrm{P}(Z_i \le z) = F_Z(z)$ for each $i \in \{1, \ldots, n\}$.