ARTICLE IN PRESS

Spatial Statistics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Spatial Statistics

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

Modelling skewed spatial random fields through the spatial vine copula☆

Benedikt Gräler *

Institute for Geoinformatics, University of Münster, Heisenbergstr. 2, 48149 Münster, Germany

ARTICLE INFO

Article history:
Received 28 May 2013
Accepted 7 January 2014
Available online xxxx

Keywords:
Spatial copula
Multivariate extremes
Vine copula
Interpolation

ABSTRACT

Studying phenomena that follow a skewed distribution and entail an extremal behaviour is important in many disciplines. How to describe and model the dependence of skewed spatial random fields is still a challenging question. Especially when one is interested in interpolating a sample from a spatial random field that exhibits extreme events, classical geostatistical tools like kriging relying on the Gaussian assumption fail in reproducing the extremes. Originating from the multivariate extreme value theory partly driven by financial mathematics, copulas emerged in recent years being capable of describing different kinds of joint tail behaviours beyond the Gaussian realm. In this paper *spatial vine copulas* are introduced that are parametrized by distance and allow to include extremal behaviour of a spatial random field. The newly introduced distributions are fitted to the widely studied emergency and routine scenario data set from the spatial interpolation comparison 2004 (SIC2004). The presented spatial vine copula ranks within the top 5 approaches and is superior to all approaches in terms of the mean absolute error.

1. Introduction

Interpolation of spatial random fields is a common task in geostatistics. Simple approaches like inverse distance weighted predictions or the well known kriging procedures have routinely been

* Tel.: +49 251 8333082.
E-mail address: ben.graeler@uni-muenster.de.
URL: http://www.ifgi.de/graeler.

2211-6753/$ – see front matter © 2014 The Author. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.spasta.2014.01.001

applied for many years. However, when the underlying assumptions (i.e. Gaussianity) of these approaches are hard to be fulfilled, alternatives are needed. Copulas have been used in different but few applications in the domain of spatial statistics. Bárdossy (2006) was one of the first who applied copulas in a geostatistical context. Some recent advances incorporating copulas in this field have for instance been published by Kazianka and Pilz (2011, 2010a), Bárdossy (2011), Bárdossy and Pegram (2009) or Bárdossy and Li (2008). They use a comparatively small set of copula families to model spatial processes. Copulas describing the dependence structure of extremes can for instance be found in Grimaldi and Serinaldi (2006), Salvadori and De Michele (2013), Salvadori et al. (2011) or Kao and Govindaraju (2010). These applications typically investigate multivariate extremes without addressing spatial dependence.

The set of methods to model spatial data including extremes is diverse. The different approaches go beyond the field of geostatistics (e.g. Fournier and Furrer, 2005) and incorporate techniques such as neural networks (e.g. Timonin and Savelieva, 2005) or support vector machines (e.g. Pozdnoukhov, 2005) as presented in the spatial interpolation comparison 2004 (SIC2004: Dubois and Galmarini, 2005a). Typically studied spatial phenomena exhibiting extremes are for example radioactive radiation, as in SIC2004, rainfall data (Haberlandt, 2007) or air quality indicators (Horálek et al., 2007).

The advantage of the *spatial vine copula* approach presented in this paper is its flexibility in the selection of appropriate copula families through bivariate spatial copulas. Schepsmeier (2013) suggests an approach where the tree structure of the vine is derived through spatial distances, but the copula families do not change with distance. Another approach modelling several air-quality indicators across a set of stations is briefly introduced by Brechmann (2013) using a hierarchical Kendall copula.

The introduction of a bivariate spatial copula into a vine copula for interpolation has been described by Gräler and Pebesma (2011) and is extended in this paper. Convex combinations of bivariate copulas parametrized by distance are combined in a *vine copula* (also known as *pair-copula construction*: Aas et al., 2009; Bedford and Cooke, 2002) for a local neighbourhood. Adding marginal distributions to the spatial vine copula yields a full multivariate distribution describing a local spatially dependent distribution of the observed phenomenon.

In the following, we will assume a spatial random field $Z : \Omega \times \mathcal{S} \to \mathbb{R}$ defined over some spatial domain of interest $\mathcal{S}$ and probability space Ω. Typically, a sample $\mathbf{Z} = \big(z(s_1), \ldots, z(s_n)\big)$ has been observed at a set of distinct locations $s_1, \ldots, s_n \in \mathcal{S}$. Often, one is interested in modelling Z from the sample $\mathbf{Z}$ in order to predict $Z(s_0)$ at unobserved locations $s_0 \in \mathcal{S}$ or to simulate the spatial random field.

The remainder of this paper is organized as follows. The theoretical background of copulas, bivariate spatial copulas and vine copulas yielding the spatial vine copulas, which are the driving probabilistic tool in the applications, are addressed in the following section. A strategy to estimate a spatial vine copula is illustrated in Section 3. Section 4 discusses different uses of the multivariate distribution such as the possibility to predict values at unobserved locations or simulate from the spatial random field. An application is illustrated in Section 5 where we use the emergency and routine scenario data sets from the SIC2004 (Dubois and Galmarini, 2005a). Results are discussed in Section 6. Conclusions are drawn in Section 7.

2. Spatial vine copulas

Copulas describe the dependence between the margins of multivariate distributions. Sklar (1959) proofed that any multivariate distribution H can be split into its margins $F_1, \ldots, F_n$ and the copula C which couples the margins with a given dependence structure: $H(x_1, \ldots, x_n) = C\big(F_1(x_1), \ldots, F_n(x_n)\big)$. Many different families exist allowing for very different dependence structures. A copula can be imagined as a multivariate cumulative distribution function on the unit (hyper-) cube with uniform margins where its density reflects the strength of dependence between the margins. For further details we refer to the introductory book by Nelsen (2006).

Sklar's Theorem is true for any dimension $d \geq 2$, but we will at first only consider bivariate copulas $C : [0, 1]^2 \to [0, 1]$. The density of a copula (denoted as c) expresses the strength of dependence which changes over the range of the marginal distributions. The only copula exhibiting a