Spatial Statistics 22 (2017) 180–195

Contents lists available at ScienceDirect

Spatial Statistics

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

Factor copula models for data with spatio-temporal dependence

Pavel Krupskii *, Marc G. Genton

CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

ARTICLE INFO

Article history:
Received 21 January 2017
Accepted 2 October 2017
Available online 10 October 2017

Keywords:
Copula
Heavy tails
Non-Gaussian random field
Spatial statistics
Tail asymmetry
Temporal dependence

ABSTRACT

We propose a new copula model for spatial data that are observed repeatedly in time. The model is based on the assumption that there exists a common factor that affects the measurements of a process in space and in time. Unlike models based on multivariate normality, our model can handle data with tail dependence and asymmetry. The likelihood for the proposed model can be obtained in a simple form and therefore parameter estimation is quite fast. Simulation from this model is straightforward and data can be predicted at any spatial location and time point. We use simulation studies to show different types of dependencies, both in space and in time, that can be generated by this model. We apply the proposed copula model to hourly wind data and compare its performance with some classical models for spatio-temporal data.

1. Introduction

Flexible and tractable models for data are often required in real-world applications, but building such models can be a challenging task if the data have complex structures. One example of such data is measurements of a process taken in space and in time, such as daily temperature measurements obtained at different weather stations or concentrations of a certain air pollutant measured by balloons launched from different locations. The dependence between two measurements that are made at different locations and at different times is usually weaker with a larger distance and time lag. Classical models for data with spatio-temporal dependence often assume multivariate normality with a spatio-temporal covariance matrix; see, for example, Gneiting (2002), Stein (2005) and Gneiting et al. (2007) for a review of covariance functions.

* Corresponding author.
E-mail addresses: pavel.krupskiy@kaust.edu.sa (P. Krupskii), marc.genton@kaust.edu.sa (M.G. Genton).

https://doi.org/10.1016/j.spasta.2017.10.001
2211-6753/

For non-Gaussian spatial data, Bárdossy (2006) introduced the chi-squared copula and Bárdossy and Li (2008) proposed a v-transformed copula. These copula models are obtained from a non-monotonic transformation of multivariate normal variables. They can handle dependence asymmetry but cannot be used for modeling data with tail dependence. Furthermore, the likelihood for these models is not tractable in high dimensions. To construct flexible distributions for spatial data and to do the interpolation, vine copulas can be used. Gräler (2014) used spatial vine copulas to model and interpolate data with very strong dependencies, and Erhardt et al. (2015) used C-vine copulas to model the spatial dependence structure locally. Parameters in their model can be estimated using the composite likelihood, and data can be interpolated at arbitrary spatial locations.

For data with spatio-temporal dependence, de Luna and Genton (2005) used vector autoregressive models with spatial structure for time-forward predictions in environmental applications, but these models are not computationally tractable if the innovation process is not Gaussian. Stroud et al. (2011) proposed a model for nonstationary spatio-temporal data in which the mean function at each time period is a locally-weighted mixture of linear regressions. The authors provided details for the Gaussian case but did not study the dependence properties of the proposed models in the general case. In practical applications, however, the multivariate normality assumption is not always suitable. For example, it would be unsuitable for data with strong joint dependence in the tails (i.e., when large/small values are simultaneously observed more often than predicted by the normal model), or for data with reflection asymmetry (i.e., when large values are simultaneously observed more often than small values, or vice versa). Fonseca and Steel (2011) introduced a model for spatio-temporal data that can handle heavy tails. However, the likelihood function in that model is not available in simple form, and it cannot handle dependence asymmetry. Schmidt et al. (2017) proposed a model for a skewed spatio-temporal process. Their model is based on the combination of Gaussian processes with purely spatial dependence structures and a purely temporal component. The joint density in that model is not possible to obtain in a simple form and it cannot handle data with tail dependence; see also the discussion by Genton and Hering (2017).

To overcome this problem, copulas can be used to construct flexible, multivariate distributions. A copula is a multivariate cumulative distribution function (cdf) with uniform $U(0, 1)$ marginals. Sklar (1959) showed that for any continuous d-dimensional cdf $F_{1,\ldots,d}$ with univariate marginals $F_1, \ldots, F_d$, there exists a unique copula $C_{1,\ldots,d}$ such that $F_{1,\ldots,d}(z_1, \ldots, z_d) = C_{1,\ldots,d}\{F_1(z_1), \ldots, F_d(z_d)\}$ for any $z_1, \ldots, z_d$. Copulas have been used in many different applications, such as modeling financial returns data (Patton, 2006; Krupskii and Joe, 2013), hydrology data (Genest and Favre, 2007) and others.

Recently, Krupskii et al. (in press) introduced a copula model for spatial data with replicates and without temporal dependence. The model is based on the process

$$W(\boldsymbol{s}) = Z(\boldsymbol{s}) + V_0, \quad \boldsymbol{s} \in \mathbb{R}^d,$$

where Z is a Gaussian process and V_0 is a common factor that does not depend on Z or location $\boldsymbol{s}$. In this paper, we propose an extension of this model that is based on the process W measured in space and in time:

$$W(\boldsymbol{s}, t) = Z(\boldsymbol{s}, t) + \alpha(\boldsymbol{s}, t)\mathcal{E}_{\mathcal{P}(t)}, \quad \boldsymbol{s} \in \mathbb{R}^d, t \in \mathbb{R}_+. \tag{1}$$

Here $Z(\boldsymbol{s}, t)$ is a Gaussian process in space and in time with zero mean, unit variance and covariance matrix $\boldsymbol{\Sigma_Z}$, $\alpha(\boldsymbol{s}, t)$ is a non-random function of space ($\boldsymbol{s}$) and time (t), $\mathcal{P}(t)$ is a Poisson process with intensity function $\Lambda(t)$ and $\mathcal{E}_t \sim_{i.i.d.} \mathrm{Exp}(1)$ are exponential factors that do not depend on $Z(\boldsymbol{s}, t)$ or on location $\boldsymbol{s}$.

The factors $\mathcal{E}_{\mathcal{P}(t)}$ allow for tail dependence for the copula corresponding to the joint distribution of the process $W(\boldsymbol{s}, t)$ measured at different spatial locations and at different time points. The intensity function, $\Lambda(t)$, of the Poisson process $\mathcal{P}(t)$ controls the rate of decay of dependence over time. The exponential distribution of $\mathcal{E}_{\mathcal{P}(t)}$ allows one to obtain the joint copula density in this model (1) in closed form so that the model parameters can be efficiently estimated using the maximum likelihood approach.

The rest of this paper is organized as follows. In Section 2 we define the model (1) for data observed at different spatial locations and time points and study its dependence properties based on the covariance function of $Z(\boldsymbol{s}, t)$ and the choice of $\alpha(\boldsymbol{s}, t)$ and $\Lambda(t)$. In Section 3 we generate