Spatial Statistics 14 (2015) 382–399

Contents lists available at ScienceDirect

Spatial Statistics

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

Transformed Gaussian Markov random fields and spatial modeling of species abundance

Marcos O. Prates [a,*], Dipak K. Dey [b,c,d], Michael R. Willig [c,e], Jun Yan [b,c,d]

[a] Department of Statistics, Universidade Federal de Minas Gerais, Avenida Antôni Carlos, 6627, Belo Horizonte, Brazil

[b] Department of Statistics, University of Connecticut, Storrs, CT 06269, USA

[c] Center for Environmental Sciences & Engineering, University of Connecticut, Storrs, CT 06269, USA

[d] Institute for Public Health Research, University of Connecticut Health Center, East Hartford, CT 06108, USA

[e] Department of Ecology & Evolutionary Biology, University of Connecticut, Storrs, CT 06269, USA

ARTICLE INFO

Article history:
Received 1 August 2014
Accepted 11 July 2015
Available online 29 July 2015

Keywords:
Bayesian inference
Beta field
Gamma field
Gaussian copula
Generalized linear mixed model

ABSTRACT

Gaussian random field and Gaussian Markov random field have been widely used to accommodate spatial dependence under the generalized linear mixed models framework. To model spatial count and spatial binary data, we present a class of transformed Gaussian Markov random fields, constructed by transforming the margins of a Gaussian Markov random field to desired marginal distributions that accommodate asymmetry and heavy tail, as needed in many empirical circumstances. The Gaussian copula that characterizes the dependence structure facilitates inferences and applications in modeling spatial dependence. This construction leads to new models such as gamma or beta Markov fields with Gaussian copulas, that are used to model Poisson intensities or Bernoulli rates in hierarchical spatial analyses. The method is naturally implemented in a Bayesian framework. To illustrate our methodology, abundances of variety of gastropod species were collected as counts or presence versus absence from a network of spatial locations in the Luquillo Mountains of Puerto Rico. Gastropods are of considerable ecological importance in terrestrial ecosystems because of their species richness, abundances, and critical roles in ecosystem processes such as decomposition and nutrient cy-

* Corresponding author.
E-mail address: marcosop@est.ufmg.br (M.O. Prates).

http://dx.doi.org/10.1016/j.spasta.2015.07.004

cling. The new models outperform the traditional models based on Bayesian model comparison with conditional predictive ordinate. The validity of Bayesian inferences and model selection were assessed through simulation studies for both spatial Poisson regression and spatial Bernoulli regression.

1. Introduction

Spatial count or binary data are generally analyzed with a generalized linear mixed model (GLMM), where spatial dependence is captured by Gaussian random field (GRF) effects (e.g., Breslow and Clayton, 1993). When data are point-referenced or geostatistical, and prediction at unobserved sites is of main concern, Diggle et al. (1998) extended the kriging method to the spatial GLMM (SGLMM) with GRF random effects to predict the surface of the spatial random effects. Under this scheme, Christensen and Waagepetersen (2002) developed predictions for the count of weeds at unobserved sites over a region. For lattice or areal data, as is the case in our application, a Markov dependence with an appropriate neighborhood structure is often imposed on the GRF random effects, which offers both intuitive interpretation and computational advantages. A Gaussian Markov random field (GMRF) is represented by an undirected graph, and is more naturally defined through its precision matrix. The (i, j)th entry of the precision matrix is nonzero if and only if i and j are connected in the graph (Rue and Held, 2005). GLMMs with random effects of GMRF have been used in many fields. Because of the public concerns regarding global change and public health, recent applications have surged in environmental sciences (e.g., Wikle et al., 1998; Rue et al., 2004) and epidemiology (e.g., Besag et al., 1991; Schmid and Held, 2004).

We propose a hierarchical spatial generalized linear model (GLM) that is subtly different from the GLMM with GRF random effects. At the first level, the observed data are independent Poisson or Bernoulli variables given the Poisson intensities or Bernoulli rates. At the second level, the Poisson intensities or Bernoulli rates are modeled by a transformed GRF (TGRF) such that the marginal distributions are of any desired form. Similarly, a transformed GMRF (TGMRF) can be defined if the GRF is a GMRF, and the Markov property is retained regardless of the transformations. With gamma or beta margins, this leads to gamma fields or beta fields for modeling Poisson intensities or Bernoulli rates, respectively. Our specification offers new avenues to construct hierarchical spatial GLMs and a fresh look at common SGLMMs with GRF random effects. Clearly, the new framework will facilitate the definition of an adequate marginal distribution for the mean parameters that is not necessarily a simple task in the conditional modeling framework. Moreover, the dependence structure is kept unchanged in the TGMRF because of the use of the Gaussian copula and, therefore, the interpretation of the β parameters are kept unchanged. A limitation of the new methodology in comparison to the traditional conditional approach is that, although it can be done, extension of the model to include more random effects, e.g. temporal effects, is not as trivial as it is in additive models. Inferences are conducted in the Bayesian framework with a general purpose, easy-to-implement Gibbs sampling algorithm.

The essence of TGRF or TGMRF is the Gaussian copula (Nelsen, 2006; Song, 2000; Masarotto and Varin, 2012), which has been used under other terminologies in various contexts. For multivariate data, it is equivalent to the Gaussian copula regression model (Pitt et al., 2006), where the response vector may be a combination of discrete and continuous variables. Under a graphical model framework it is similar to the copula Gaussian graphical model of Dobra and Lenkoski (2011), where the dependence structure determined by the precision matrix is of specific interest. In some fields such as hydrology, it is named as meta-Gaussian distribution (e.g., Guillot and Lebel, 1999; Schaake et al., 2007). In geostatistics with point-referenced data, it is called the anamorphosis Gaussian field (Chilès and Delfiner, 1999) or Gaussian copula model (Bárdossy, 2006; Kazianka and Pilz, 2010), where the main interest of these models has being interpolation and prediction at unmeasured locations. For this setup, a Matlab toolbox implementation is available (Kazianka, 2013). Most geostatistical applications