Spatial Statistics 17 (2016) 121–130

Contents lists available at ScienceDirect

Spatial Statistics

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

Likelihood analysis for a class of spatial geostatistical compositional models

Ana Beatriz Tozzo Martins [a], Wagner Hugo Bonat [b,*], Paulo Justiniano Ribeiro Jr. [b]

[a] Departament of Statistics, Maringá State University, Maringá, PR, Brazil

[b] Departament of Statistics, Paraná Federal University, Curitiba, PR, Brazil

ARTICLE INFO

Article history:
Received 19 July 2015
Accepted 15 June 2016
Available online 28 June 2016

Keywords:
Compositional data
Geostatistical models
Maximum likelihood
Regionalized composition

ABSTRACT

We propose a model-based geostatistical approach to deal with regionalized compositions. We combine the additive-log-ratio transformation with multivariate geostatistical models whose covariance matrix is adapted to take into account the correlation induced by the compositional structure. Such specification allows the usage of standard likelihood methods for parameters estimation. For spatial prediction we combined a back-transformation with the Gauss–Hermite method to approximate the conditional expectation of the compositions. We analyse particle size fractions of the top layer of a soil for agronomic purposes which are typically expressed as proportions of sand, clay and silt. Additionally a simulation study assesses the small sample properties of the maximum likelihood estimator.

1. Introduction

Compositional data are vectors of proportions, specifying fractions of a whole whose elements typically sum to one or 100%. Given the nature of this data, the direct application of usual statistical techniques based on the Gaussian multivariate distribution on the composition values is not suitable. As pointed out by Aitchison (1986), the constant sum constraints not only invalidate the assumption

* Corresponding author.
E-mail address: wbonat@ufpr.br (W.H. Bonat).

http://dx.doi.org/10.1016/j.spasta.2016.06.008

that our response variables are drawn from unbounded random processes, but also induce negative correlations between response variables.

Compositional data are frequent in earth sciences, such as, in mineralogy, agronomy, geochemistry and hydrology. In such applications, not rarely, compositions are recorded along with their spatial locations, and spatial patterns are of interest, characterizing what is called regionalized compositions (Pawlowsky, 1989). Models accounting for spatial patterns should account for both, the correlation induced by the composition structure and the spatial correlation at a suitable scale.

Practical analysis of compositional data is, in general, based on the seminal work of Aitchison (1982) and the comprehensive monograph by Aitchison (1986). Barceló-Vidal et al. (2001) presented the mathematical background of compositional data as equivalence classes. Further mathematical developments include the definitions of Euclidean vector space structure (Pawlowsky-Glahn and Egozcue, 2001) and isometric log ratio transformation (Egozcue et al., 2003). The R package `compositions` (van ver Boogaart and Tolosana-Delgado, 2006) provides a complete toolbox for analysis of compositional data including facilities to deal with regionalized compositions (van den Boogaart and Tolosana-Delgado, 2013).

The literature about regionalized compositions is concentrated around the contributions of Pawlowsky (1989), Pawlowsky and Burger (1992), Pawlowsky et al. (1995) and its applications (Odeh et al., 2003; Lark and Bishop, 2007). The monograph (Pawlowsky-Glahn and Olea, 2004) presented the state of the art for the analysis of regionalized compositional data in the year 2000. Tjelmeland and Lund (2003) proposed a model-based approach for the analysis of spatial compositional data under the Bayesian framework. Further developments and references can be found in Tolosana-Delgado (2006) and Pawlowsky-Glahn et al. (2015).

The approach adopted in Pawlowsky-Glahn and Olea (2004) can be summarized in three steps: (i) given a vector of B regionalized compositions apply the additive-log-ratio transformation (Aitchison, 1986). (ii) for the transformed vector use the orthodox cokriging approach (Wackernagel, 1998). (iii) adopt an unbiased back-transformation to predict the compositions back on the original compositional scale. Examples of this approach with emphasis on step (iii) can be found in Lark and Bishop (2007). Such approach includes the usage of traditional geostatistical techniques with parameter estimation based on the variogram and cross-variogram methods. Alternatively, a model-based geostatistical approach (Diggle et al., 1998) can be considered, allowing the adoption of likelihood based or Bayesian statistical methods for estimation and prediction, inheriting related properties of consistency, asymptotic normality and efficiency. The application of maximum likelihood inference in geostatistics offers several advantages: The method may focus on the parameters of interest (sill, range, anisotropy angle, etc.). The uncertainty of the estimates is easily assessed. The log-likelihood function may be used for model selection. The method is more efficient than others including the variogram and moments-based methods in mean square error terms. The method can also be used for optimum sample design. For more references discussing the advantages of maximum likelihood in the context of geostatistical models, see Pardo-Igúzquiza (1998), Stein (1999) and Diggle and Ribeiro Jr. (2007).

We adopt the model-based approach to deal with regionalized compositions. Following the aforementioned approach, we apply the additive-log-ratio transformation to obtain transformed response variables, for which we specify a common spatial component multivariate geostatistical model (Diggle and Ribeiro Jr., 2007). For estimation of the model parameters we adopt the maximum likelihood method. For spatial prediction, we adopt the approach proposed by Pawlowsky-Glahn and Olea (2004) combining a back-transformation and the Gauss–Hermite method to approximate the conditional expectation of the compositions. We also obtain simulations of the predictive distributions. Our approach produces predictions satisfying the required constant sum constraints and has interpretable parameters in the scale of the transformed response variables. We apply our model to analyse a data set about the distribution of mineral particles in the soil. We also present a simulation study to verify the small sample properties of the maximum likelihood estimator.

Section 2 presents the compositional geostatistical model along with the estimation and spatial prediction procedures. In Section 3 we apply the proposed model to analyse a real data set. Section 4 presents the simulation study. Finally, Section 5 provides some discussions and recommendations for future works. We provide the R code and data set in the supplementary material.