Accepted Manuscript

Geostatistical estimation and prediction for censored responses

José A. Ordoñez, Dipankar Bandyopadhyay, Victor H. Lachos, Celso R.B. Cabral

PII: S2211-6753(17)30043-X
DOI: https://doi.org/10.1016/j.spasta.2017.12.001
Reference: SPASTA 277

To appear in: *Spatial Statistics*

Received date : 13 February 2017
Accepted date : 2 December 2017

Please cite this article as: Ordoñez J.A., Bandyopadhyay D., Lachos V.H., Cabral C.R.B., Geostatistical estimation and prediction for censored responses. *Spatial Statistics* (2017), https://doi.org/10.1016/j.spasta.2017.12.001

Geostatistical estimation and prediction for censored responses

José A. Ordoñez[a], Dipankar Bandyopadhyay[b], Victor H. Lachos[c] *, Celso R. B. Cabral[d]

[a] Department of Statistics, Campinas State University, Campinas, São Paulo, Brazil

[b] Department of Biostatistics, Virginia Commonwealth University, Richmond, VA, U.S.A.

[c] Department of Statistics, University of Connecticut, Storrs, CT 06269, U.S.A.*

[d] Department of Statistics, Federal University of Amazonas, Manaus, Brazil

Abstract

Spatially-referenced geostatistical responses that are collected in environmental sciences research are often subject to detection limits, where the measures are not fully quantifiable. This leads to censoring (left, right, interval, etc), and various ad hoc statistical methods (such as choosing arbitrary detection limits, or data augmentation) are routinely employed during subsequent statistical analysis for inference and prediction. However, inference may be imprecise and sensitive to the assumptions and approximations involved in those arbitrary choices. To circumvent this, we propose an *exact* maximum likelihood estimation framework of the fixed effects and variance components and related prediction via a novel application of the Stochastic Approximation of the Expectation Maximization (SAEM) algorithm, allowing for easy and elegant estimation of model parameters under censoring. Both simulation studies and application to a real dataset on arsenic concentration collected by the Michigan Department of Environmental Quality demonstrate the advantages of our method over the available naïve techniques in terms of finite sample properties of the estimates, prediction, and robustness. The proposed methods can be implemented using the `R` package `CensSpatial`.

Key words: Censored geostatistical data, Kriging, Limit of detection (LOD), SAEM algorithm.

1. Introduction

Geostatistical data modeling has now virtually permeated all areas of epidemiology, hydrology, agriculture, environmental science, demographic studies, etc. Here, the prime objective is to account for the spatial correlation among the observations collected at various locations, and also to predict the values of interest for non-sampled sites. For most applications, the data are assumed to be fully observed. However, in many situations, the measured (spatial) responses are censored, i.e., the measurements are subjected to some upper or lower detection limits (depending on the measuring device), below or above which they

*Address for correspondence: Víctor Hugo Lachos Dávila, Department of Statistics, University of Connecticut, 215 Glenbrook Rd. U-4120, Storrs, CT 06269-4120; E-mail: hlachos@uconn.edu