Likelihood Inference in Multivariate Model-Based Geostatistics [1]

Clarissa Ferrari, Marco Minozzo
University of Verona, marco.minozzo@univr.it

Abstract: Multivariate model-based geostatistics refers to the extension of classical multivariate geostatistical techniques, in particular the classical linear model of coregionalization, to the case of non-Gaussian data. Extensions of this kind are still limited in the statistical literature, mainly for the inferential problems they pose, and almost invariably inference is carried out in a Bayesian context. In this work we present some new results on likelihood inference for the unknown parameters of a hierarchical geostatistical factor model. In particular, we show the implementation of some Monte Carlo EM algorithms and discuss their performances, in particular their sampling distributions, mainly through some simulation studies.

Keywords: Cokriging, Generalized linear mixed model, Markov chain Monte Carlo, Monte Carlo EM, Multivariate geostatistics.

1 Introduction

The classical linear model of coregionalization, or its simpler counterpart, the proportional covariance model, otherwise known as intrinsic correlation model, and the related 'factorial kriging analysis' have become standard tools in many areas of application for the analysis of multivariate spatial data. However, in presence of non-Gaussian data, in particular count or skew data, the use of these geostatistical instruments can lead to misleading predictions and to erroneous conclusions about the underling factors. To cope with these situations, following the proposal put forward in the univariate case by Diggle et al. (1998), and somehow extending the works of Zhang (2007) and of Zhu et al. (2005), we propose in Section 2 a hierarchical multivariate spatial model, built upon a generalization of the classical geostatistical proportional covariance model. Adopting a non-Bayesian inferential framework, and assuming that the number of underlying common factors and their spatial autocorrelation structure are known, in Section 3 we show how to carry out likelihood inference on the parameters of the model by exploiting the capabilities of Markov chain Monte Carlo (MCMC) and Monte Carlo EM (MCEM) algorithms.

[1] We gratefully acknowledge funding from the Italian Ministry of Education, University and Research (MIUR) through PRIN 2008 project 2008MRFM2H.

2 Multivariate Model-Based Geostatistics

Let us consider the following hierarchical extension of the classical geostatistical linear model of coregionalization. Let $y_i(\mathbf{x}_k)$, $i = 1, \ldots, m$, $k = 1, \ldots, K$, be a set of geo-referenced data measurements relative to m regionalized variables, gathered at K spatial locations $\mathbf{x}_k$. These m regionalized variables are seen as a partial realization of a set of m random functions $Y_i(\mathbf{x})$, $i = 1, \ldots, m$, $\mathbf{x} \in \mathbb{R}^2$. For these functions we assume, for any $\mathbf{x}$, and for $i \neq j$,

$$Y_i(\mathbf{x}) \perp\!\!\!\perp Y_j(\mathbf{x})|Z_i(\mathbf{x}) \quad \text{and} \quad Y_i(\mathbf{x}) \perp\!\!\!\perp Z_j(\mathbf{x})|Z_i(\mathbf{x}), \tag{1}$$

and, for $\mathbf{x}' \neq \mathbf{x}''$, and $i, j = 1, \ldots, m$,

$$Y_i(\mathbf{x}') \perp\!\!\!\perp Y_j(\mathbf{x}'')|Z_i(\mathbf{x}') \quad \text{and} \quad Y_i(\mathbf{x}') \perp\!\!\!\perp Z_j(\mathbf{x}'')|Z_i(\mathbf{x}'), \tag{2}$$

where $Z_i(\mathbf{x})$, $i = 1, \ldots, m$, $\mathbf{x} \in \mathbb{R}^2$, are mean zero joint stationary Gaussian processes.

Moreover, for any given i and $\mathbf{x}$, we assume that, conditionally on $Z_i(\mathbf{x})$, the random variables $Y_i(\mathbf{x})$ have conditional distributions $f_i(y; M_i(\mathbf{x}))$, that is, $Y_i(\mathbf{x})|Z_i(\mathbf{x}) \sim f_i(y; M_i(\mathbf{x}))$, specified by the conditional expectations $M_i(\mathbf{x}) = \mathrm{E}[Y_i(\mathbf{x})|Z_i(\mathbf{x})]$, and that $h_i(M_i(\mathbf{x})) = \beta_i + Z_i(\mathbf{x})$, for some parameters β_i and some known link functions $h_i(\cdot)$. For instance, we might assume that for some or all i, and for any given $\mathbf{x}$, the data are conditionally Poisson distributed, that is, that

$$f_i(y; M_i(\mathbf{x})) = \exp\{-M_i(\mathbf{x})\}(M_i(\mathbf{x}))^y/y!, \qquad y = 0, 1, 2, \ldots, \tag{3}$$

and that the linear predictor $\beta_i + Z_i(\mathbf{x})$ is related to the conditional mean $M_i(\mathbf{x})$ through a logarithmic link function so that $\ln(M_i(\mathbf{x})) = \beta_i + Z_i(\mathbf{x})$. On the other hand, for the rest of the i, we might assume that, for any given $\mathbf{x}$, conditionally on $Z_i(\mathbf{x})$, the random variables $Y_i(\mathbf{x})$ are Gamma distributed with conditional expectations $M_i(\mathbf{x}) = \mathrm{E}\big[Y_i(\mathbf{x})\big|Z_i(\mathbf{x})\big] = \exp\big\{\beta_i + Z_i(\mathbf{x})\big\} = \nu b$, (here again $h_i(\cdot) = \ln(\cdot)$) and conditional variances $\mathrm{Var}\big[Y_i(\mathbf{x})\big|Z_i(\mathbf{x})\big] = \nu b^2 = \nu^{-1}\exp\big\{2\beta_i + 2Z_i(\mathbf{x})\big\} = \nu^{-1}(M_i(\mathbf{x}))^2$, where $\nu > 0$ and $b > 0$ are parameters, that is, we might assume

$$f_i(y; M_i(\mathbf{x})) = (y^{\nu-1}/\Gamma(\nu))\exp\{-y\nu/M_i(\mathbf{x})\}(\nu/M_i(\mathbf{x}))^{\nu}, \qquad y > 0. \tag{4}$$

Here the 'shape' parameter ν is constant for $\mathbf{x} \in \mathbb{R}$, whereas the 'scale' parameter b varies over $\mathbb{R}$ depending on the conditional expectation $M_i(\mathbf{x})$. In addition to the Poisson or Gamma distributions, other discrete or continuous distributions could be considered to account for particular set of data.

For the latent part of the model, we adopt the following structure. For the m joint stationary Gaussian processes $Z_i(\mathbf{x})$, let us assume the linear factor model

$$Z_i(\mathbf{x}) = \sum_{p=1}^{P} a_{ip}F_p(\mathbf{x}) + \xi_i(\mathbf{x}), \tag{5}$$

where a_{ip} are $m \times P$ coefficients, $F_p(\mathbf{x})$, $p = 1, \dots, P$, are $P \leq m$ non-observable spatial components (common factors) responsible for the cross correlation between the variables $Z_i(\mathbf{x})$, and $\xi_i(\mathbf{x})$ are non-observable spatial components (unique factors) responsible for the residual autocorrelation in the $Z_i(\mathbf{x})$ unexplained by the common factors. We assume that $F_p(\mathbf{x})$ and $\xi_i(\mathbf{x})$ are mean zero stationary Gaussian processes with covariance functions $\mathrm{Cov}\big[F_p(\mathbf{x}), F_p(\mathbf{x}+\mathbf{h})\big] = \rho(\mathbf{h})$, and $\mathrm{Cov}\big[\xi_i(\mathbf{x}), \xi_i(\mathbf{x}+\mathbf{h})\big] = \psi_i \rho(\mathbf{h})$, where $\mathbf{h} \in \mathbb{R}^2$, $\rho(\mathbf{h})$ is a real spatial autocorrelation function common to all factors such that $\rho(0) = 1$ and $\rho(\mathbf{h}) \to 0$, as $||\mathbf{h}|| \to \infty$, and ψ_i are non-negative real parameters. We also assume that the processes $F_p(\mathbf{x})$ and $\xi_i(\mathbf{x})$ have all cross-covariances identically equal to zero.

Assuming that the number P of latent common factors and that the spatial autocorrelation function $\rho(\mathbf{h})$ have already been chosen, the model depends on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{A}, \boldsymbol{\psi})$, where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)^T$, $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_m)^T$, with $\mathbf{a}_i = (a_{i1}, \dots, a_{iP})$, for $i = 1, \dots, m$, and $\boldsymbol{\psi} = (\psi_1, \dots, \psi_m)^T$. Let us note that, as the classical linear factor model, our model is not identifiable. However, the only indeterminacy stays in a rotation of the matrix $\mathbf{A}$.

3 Likelihood inference via MCEM

Adopting a non-Bayesian inferential framework, likelihood inference on the parameters of the model would require the maximization, with respect to $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{A}, \boldsymbol{\psi})$, of the likelihood based on the marginal density function of the observations $y_i(\mathbf{x}_k)$. However, since this marginal density is not available, and since the integration required in the E-step of the EM algorithm would not be easy, here, to maximize the log-likelihood, we will resort to the MCEM algorithm (see Wei and Tanner 1990).

Our implementation of the algorithm proceeds as follows. Let us define $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_m)$ where $\boldsymbol{\xi}_i = (\xi_i(\mathbf{x}_1), \dots, \xi_i(\mathbf{x}_K))^T$, $i = 1, \dots, m$, and $\mathbf{F} = (\mathbf{F}_1, \dots, \mathbf{F}_P)$ where $\mathbf{F}_p = (F_p(\mathbf{x}_1), \dots, F_p(\mathbf{x}_K))^T$, $p = 1, \dots, P$, and let $f(\mathbf{y}, \boldsymbol{\xi}, \mathbf{F}; \boldsymbol{\theta})$ be the joint distribution of the model, that is, the complete log-likelihood, accounting also for the unobserved factors. Assuming that the current guess for the parameters after the $(s-1)$th iteration is given by $\boldsymbol{\theta}_{s-1}$, and that R_s is a fixed positive integer, the sth iteration of the MCEM algorithm involves the following three steps (stochastic, expectation, maximization):

- *S step* – draw R_s samples $(\boldsymbol{\xi}^{(r)}, \mathbf{F}^{(r)})$, $r = 1, \dots, R_s$, from the (filtered) conditional distribution $f(\boldsymbol{\xi}, \mathbf{F} | \mathbf{y}; \boldsymbol{\theta}_{s-1})$;
- *E step* – compute $Q_s(\boldsymbol{\theta}, \boldsymbol{\theta}_{s-1}) = (1/R_s) \sum_{r=1}^{R_s} \ln f(\mathbf{y}, \boldsymbol{\xi}^{(r)}, \mathbf{F}^{(r)}; \boldsymbol{\theta})$;
- *M step* – take as the new guess $\boldsymbol{\theta}_s$ the value of $\boldsymbol{\theta}$ which maximizes $Q_s(\boldsymbol{\theta}, \boldsymbol{\theta}_{s-1})$.

With R_s very large this procedure approximates the EM algorithm, whereas a simulated annealing version could be obtained by choosing an increasing sequence $R_s \to \infty$, as $s \to \infty$, (see, for instance, Fort and Moulines 2003). The S-step of the

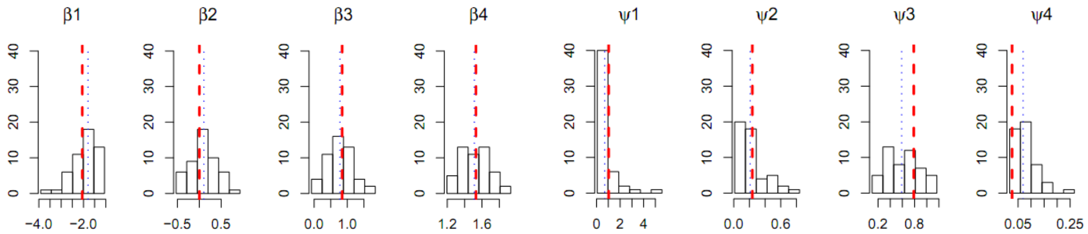

Figure 1: Histograms of the simulated marginal distributions of the MCEM estimator for the 8 parameters of a model with $m = 4$ and one common factor, obtained by running the algorithm over 50 simulated data sets. Dashed lines are the true parameter values; dotted lines are the empirical arithmetic means of the distributions.

algorithm can be dealt with through importance sampling or MCMC techniques, whereas the M-step typically requires the use of numerical routines.

When the matrix $\mathbf{A}$ is known, the complete log-likelihood belongs to the curved exponential family and by choosing an appropriate increasing sequence R_s the algorithm converges to the maximum likelihood estimate (Fort and Moulines 2003). On the other hand, when the matrix $\mathbf{A}$ is unknown, the complete likelihood does not belong any more to the curved exponential family and theoretical convergence properties are not available. However, we show, either in the case in which $\mathbf{A}$ is known or unknown, through some extensive simulation studies, that the MCEM algorithm provides estimates with quite reasonable sampling distributions. For instance, Figure 1 shows the simulated distributions of the MCEM estimator in the case in which $P = 1$ and $\mathbf{A}$ is known.

References

Diggle P. J., Moyeed R. A., Tawn J. A. (1998) Model-based geostatistics (with discussion), *Applied Statistics*, 47, 299–350.

Fort G., Moulines E. (2003) Convergence of the Monte Carlo expectation maximization for curved exponential families, *The Annals of Statistics*, 31, 1220–1259.

Wei G. C. G., Tanner M. A. (1990) A Monte Carlo implementation of the EM algorithm and poor man's data augmentation algorithm. *Journal of the American Statistical Association*, 85. 699–704.

Zhang H. (2007) Maximum-likelihood estimation for multivariate spatial linear coregionalization models, *Environmetrics*, 18, 125–139.

Zhu J., Eickhoff J. C., Yan P. (2005) Generalized linear latent variable models for repeated measures of spatially correlated multivariate data, *Biometrics*, 61, 674–683.