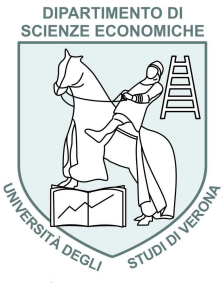

Working Paper Series
Department of Economics
University of Verona

Monte Carlo likelihood inference in multivariate model-based geostatistics

Marco Minozzo, Clarissa Ferrari

WP Number: 33 November 2012

ISSN: 2036-2919 (paper), 2036-4679 (online)

Monte Carlo Likelihood Inference in Multivariate Model-Based Geostatistics

Marco Minozzo and Clarissa Ferrari

Abstract Though in the last decade many works have appeared in the literature dealing with model-based extensions of the classical (univariate) geostatistical mapping methodology based on linear Kriging, very few authors have concentrated, mainly for the inferential problems they pose, on model-based extensions of classical multivariate geostatistical techniques like the linear model of coregionalization, or the related 'factorial kriging analysis'. Nevertheless, in presence of multivariate spatial non-Gaussian data, in particular count data, as in many environmental applications, the use of these classical techniques can lead to incorrect predictions about the underling factors. To overcome this problem, here we discuss a hierarchical geostatistical factor model that extends, following a model-based geostatistical approach, the classical geostatistical proportional covariance model. For this model we investigated likelihood-based inferential procedures based on the Monte Carlo EM algorithm and on Monte Carlo likelihood. In particular, we discuss some of their theoretical properties and report some simulation studies performed to investigate their sampling distributions.

Key words: Cokriging, generalized linear mixed models, linear model of coregionalization, Monte Carlo EM, spatial factor model, spatial prediction

Marco Minozzo
Department of Economics, University of Verona, Via dell'Artigliere 19, IT-37129 Verona, Italy, e-mail: marco.minozzo@univr.it

Clarissa Ferrari
Department of Economics, University of Verona, Via dell'Artigliere 19, IT-37129 Verona, Italy, e-mail: claclafer@gmail.com

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 (Chilès and Delfiner, 1999). 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) (see also Diggle and Ribeiro, 2007) and somehow extending the works of Zhang (2007) and 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 (similarly to the work of Christensen (2004) in the univariate case) on the parameters of the model by exploiting the capabilities of the Monte Carlo EM (MCEM) algorithm (see Wei and Tanner, 1990) and of Monte Carlo likelihood. Lastly, in Section 4 we end with some discussion of open problems and further work.

2 The Modeling Framework

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 $\{Y_i(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^2\}$, $i = 1,\ldots,m$. 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 $\{Z_i(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^2\}$, $i = 1,\ldots,m$, are mean zero joint stationary latent 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, that $Y_i(\mathbf{x})|Z_i(\mathbf{x}) \sim f_i(y; M_i(\mathbf{x}))$, 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 the latent part of the model, we adopt the following structure. For the m joint stationary Gaussian processes $\{Z_i(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^2\}$, let us assume the linear factor model

$$Z_i(\mathbf{x}) = \sum_{p=1}^{P} a_{ip} F_p(\mathbf{x}) + \xi_i(\mathbf{x}), \tag{3}$$

where a_{ip} are $m \times P$ coefficients, $F_p(\mathbf{x})$, $p = 1, \ldots, P$, 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 $\{F_p(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^2\}$ and $\{\xi_i(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^2\}$ 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.

As far as the latent processes $Z_i(\mathbf{x})$, $i = 1, \ldots, m$, are concerned, we have that

$$\begin{aligned}
\mathrm{Cov}[Z_i(\mathbf{x}), Z_j(\mathbf{x}+\mathbf{h})] &= \mathrm{E}\left[\left(\sum_{p=1}^{P} a_{ip} F_p(\mathbf{x}) + \xi_i(\mathbf{x})\right) \cdot \left(\sum_{p=1}^{P} a_{jp} F_p(\mathbf{x}+\mathbf{h}) + \xi_j(\mathbf{x}+\mathbf{h})\right)\right] \\
&= \mathrm{E}\left[\sum_{p=1}^{P} a_{ip} a_{jp} F_p(\mathbf{x}) F_p(\mathbf{x}+\mathbf{h}) + \xi_i(\mathbf{x}) \xi_j(\mathbf{x}+\mathbf{h})\right] \\
&= \sum_{p=1}^{P} a_{ip} a_{jp}\, \mathrm{E}[F_p(\mathbf{x}) F_p(\mathbf{x}+\mathbf{h})] + \mathrm{E}[\xi_i(\mathbf{x}) \xi_j(\mathbf{x}+\mathbf{h})],
\end{aligned}$$

that is, we have the covariance functions $\mathrm{Cov}[Z_i(\mathbf{x}), Z_i(\mathbf{x}+\mathbf{h})] = \sum_{p=1}^{P} a_{ip}^2\ \rho(\mathbf{h}) + \psi_i\, \rho(\mathbf{h})$, and the cross-covariance functions $\mathrm{Cov}[Z_i(\mathbf{x}), Z_j(\mathbf{x}+\mathbf{h})] = \sum_{p=1}^{P} a_{ip} a_{jp} \rho(\mathbf{h})$, taking, respectively, $i = j$ and $i \neq j$. Of course, taking $\mathbf{h} = 0$, we also have that $\mathrm{Var}[Z_i(\mathbf{x})] = \sum_{p=1}^{P} a_{ip}^2 + \psi_i$, and $\mathrm{Cov}[Z_i(\mathbf{x}), Z_j(\mathbf{x})] = \sum_{p=1}^{P} a_{ip} a_{jp}$. Notice that the assumption that all common and unique factors share the same spatial autocorrelation structure, implies a specific autocorrelation structure also for the m joint stationary Gaussian processes $Z_i(\mathbf{x})$. Indeed, it implies that the covariance and cross-covariance functions of the processes $Z_i(\mathbf{x})$, which determine the (conditional) mean of the observable $Y_i(\mathbf{x})$, are all proportional to the spatial autocorrelation function $\rho(\mathbf{h})$. In turn, this implies a specific autocorrelation structure also for the $Y_i(\mathbf{x})$.

To completely specify the model, we need to detail the conditional distributions $f_i(y; M_i(\mathbf{x}))$ and the link functions $h_i(\cdot)$, that may depend, in general, on the index i. In the following we will briefly discuss some possibilities that could be considered in the presence of non-Gaussian data.

2.1 Poisson distribution

For a given i, if $y_i(\mathbf{x}_k)$ are count data, we might assume that, for any given $\mathbf{x}$, these are conditionally Poisson distributed, that is, that

$$f_i(y;M_i(\mathbf{x})) = \exp\{-M_i(\mathbf{x})\}(M_i(\mathbf{x}))^y/y!, \quad y = 0,1,2,\ldots, \tag{4}$$

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 $\ln(M_i(\mathbf{x})) = \beta_i + Z_i(\mathbf{x})$. As shown in Minozzo and Fruttini (2004) (see also Minozzo and Ferrari, 2012), in this case the process $Y_i(\mathbf{x})$ is both strongly and weakly stationary. In fact, defining $\varsigma_i^2 = \mathrm{Var}[Z_i(\mathbf{x})]$, $\tau_i^* = \exp\{\beta_i + \varsigma_i^2/2\}$ and $\tau_i^2(\mathbf{x}) = \mathrm{Var}[Y_i(\mathbf{x})|Z_i(\mathbf{x})]$, with some algebra we obtain:

$$\begin{aligned}
\mu_i &= \mathrm{E}[Y_i(\mathbf{x})] = \mathrm{E}[\mathrm{E}[Y_i(\mathbf{x})|Z_i(\mathbf{x})]] = \mathrm{E}[\exp\{\beta_i + Z_i(\mathbf{x})\}] = \tau_i^*; \\
\sigma_i^2 &= \mathrm{Var}[Y_i(\mathbf{x})] = \mathrm{E}[\tau_i^2(\mathbf{x})] + \mathrm{Var}[M_i(\mathbf{x})] = \tau_i^* + (\tau_i^*)^2[\exp\{\varsigma_i^2\} - 1]; \\
C_{ii}(\mathbf{h}) &= \mathrm{Cov}[Y_i(\mathbf{x}), Y_i(\mathbf{x}+\mathbf{h})] = (\tau_i^*)^2[\exp\{\varsigma_i^2\rho(\mathbf{h})\} - 1], \quad \mathbf{h} \neq 0.
\end{aligned} \tag{5}$$

It is easy to check that, in the case in which $\rho(\mathbf{h}) = \rho(-\mathbf{h})$, we have $C_{ii}(-\mathbf{h}) = C_{ii}(\mathbf{h})$. Moreover, $C_{ii}(\infty) = 0$ and $C_{ii}(0) = \sigma_i^2 \neq C_{ii}(0^+)$, that is, the covariance function $C_{ii}(\mathbf{h})$ is discontinuous at the origin.

As for the variogram of the observable $Y_i(\mathbf{x})$ we have, for $\mathbf{h} \neq 0$,

$$\begin{aligned}
\gamma_{ii}(\mathbf{h}) &= \frac{1}{2}\mathrm{Var}\big[Y_i(\mathbf{x}+\mathbf{h}) - Y_i(\mathbf{x})\big] = \frac{1}{2}\mathrm{E}\big[(Y_i(\mathbf{x}+\mathbf{h}) - Y_i(\mathbf{x}))^2\big] \\
&= \tau_i^* + (\tau_i^*)^2\big[\exp\big\{\varsigma_i^2\big\} - \exp\big\{\varsigma_i^2\rho(\mathbf{h})\big\}\big].
\end{aligned} \tag{6}$$

Let us note that in this case the variogram of $Y_i(\mathbf{x})$ is always equal to zero at the origin with nugget $\gamma_{ii}(0^+) = \tau_i^*$ that equals the mean μ_i and it is, in general, strictly greater than 0. It is also easy to check that the sill of the variogram satisfies $\gamma_{ii}(\infty) = C_{ii}(0) = \sigma_i^2$. Using standard techniques, also covariance functions and cross-variograms, between observable $Y_i(\mathbf{x})$ and $Y_j(\mathbf{x})$ having conditional Poisson distribution, can be obtained (see, for instance, Minozzo and Fruttini, 2004).

2.2 Gamma distribution

Alternatively, for a given i, for continuous positive data, 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

$$f_i(y;M_i(\mathbf{x})) = (y^{\nu-1}/\Gamma(\nu))\exp\{-y\nu/M_i(\mathbf{x})\}(\nu/M_i(\mathbf{x}))^\nu, \quad y > 0. \tag{7}$$

Here the 'shape' parameter ν is constant for $\mathbf{x} \in \mathbb{R}^2$, whereas the 'scale' parameter b varies over $\mathbb{R}^2$ depending on the conditional expectation $M_i(\mathbf{x})$.

Even in this case it is possible to show that $Y_i(\mathbf{x})$ is second order stationary. Indeed, $\mu_i = \mathrm{E}[Y_i(\mathbf{x})] = \tau_i^*$,

$$\sigma_i^2 = \mathrm{Var}[Y_i(\mathbf{x})] = \frac{1}{\nu} e^{-2\beta_i} (\tau_i^*)^4 + (\tau_i^*)^2 [\exp\{\varsigma_i^2\} - 1],$$

and $C_{ii}(\mathbf{h}) = \mathrm{Cov}[Y_i(\mathbf{x}), Y_i(\mathbf{x}+\mathbf{h})]$, for $\mathbf{h} \neq 0$, is the same as in the case of the Poisson distribution. Also, for the variogram of the observable $Y_i(\mathbf{x})$ we have that, for $\mathbf{h} \neq 0$,

$$\gamma_{ii}(\mathbf{h}) = \frac{1}{\nu} e^{-2\beta_i} (\tau_i^*)^4 + (\tau_i^*)^2 \left[\exp\left\{\varsigma_i^2\right\} - \exp\left\{\varsigma_i^2 \rho(\mathbf{h})\right\} \right]. \tag{8}$$

It is immediate to see that whereas the nugget is equal to $\nu^{-1} e^{-2\beta_i} (\tau_i^*)^4$, the second addendum is the same as in the Poisson case. Let us remark that covariance functions and cross-variograms, between $Y_i(\mathbf{x})$ and $Y_j(\mathbf{x})$ having both conditional Gamma distribution, or having one Gamma and one Poisson conditional distribution, can still be obtained using standard techniques (see, for instance, Minozzo and Ferrari, 2012).

2.3 Binomial distribution

Another possibility for the conditional distribution of the observed data, that may be useful in the case of discrete data, is given by the Binomial distribution. In this case we might assume that for a given i and for any given $\mathbf{x}$, conditionally on $Z_i(\mathbf{x})$, the random variable $Y_i(\mathbf{x})$ has a Binomial distribution, that is,

$$Y_i(\mathbf{x}) \mid Z_i(\mathbf{x}) \sim f_i(y; M_i(\mathbf{x})) = \binom{n}{y} \left(\frac{M_i(\mathbf{x})}{n} \right)^y \left(1 - \frac{M_i(\mathbf{x})}{n} \right)^{n-y},$$

where $y = 0, 1, \ldots, n$. Here, we can take as link function the usual logit function $h_i(M_i(\mathbf{x})) = \ln(M_i(\mathbf{x})/(n - M_i(\mathbf{x}))$. Let us note that for this distribution, the parameter n is constant for every $\mathbf{x} \in \mathbb{R}^2$, whereas the ratio $M_i(\mathbf{x})/n$, that in this case can be interpreted as a probability, varies over $\mathbb{R}^2$ and depends on the conditional expectation $M_i(\mathbf{x})$.

Similarly to the case of the previous distributions, even here it is possible to show that the process $Y_i(\mathbf{x})$ is both strongly and weakly stationary, and also to derive analytic expressions for the covariance function and the variogram.

2.4 Skew-Normal distribution

In the case in which, for a given i, we assume that $f_i(y;M_i(\mathbf{x}))$ is Skew-Normal and $h_i(M_i(\mathbf{x}))$ is a translation by a constant, maybe because we are in presence of continuous skew data, it is still easy to verify that the process $\{Y_i(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^2\}$ is second-order, and also strongly, stationary. In this particular case, we can write

$$Y_i(\mathbf{x}) = \beta_i + Z_i(\mathbf{x}) + \omega_i S_i(\mathbf{x}), \tag{9}$$

where $\omega_i \in \mathbb{R}^+$ and $S_i(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^2$, are mutually independently distributed as (univariate) Skew-Normals $\mathrm{SN}(0,1,\alpha_i)$ such that, for every $\mathbf{x} \in \mathbb{R}^2$, the density of $S_i(\mathbf{x})$ is given by $f_{S_i}(s) = 2\phi(s;1)\Phi(\alpha_i s)$, for $-\infty < s < \infty$, where $\alpha_i \in \mathbb{R}$ is a skewness parameter, $\phi(\,\cdot\,;\sigma^2)$ is the 1-dimensional Normal density function with zero mean and variance σ^2, and $\Phi(\cdot)$ is the scalar N(0,1) distribution function. In assuming (9), we have implicitly chosen the link function $h_i(M_i(\mathbf{x})) = M_i(\mathbf{x}) - \omega_i(2/\pi)^{1/2}\alpha_i/(1+\alpha_i^2)^{1/2}$. Thus, we have that, for every $\mathbf{x} \in \mathbb{R}^2$, $Y_i(\mathbf{x})|Z_i(\mathbf{x}) \sim \mathrm{SN}(\beta_i + Z_i(\mathbf{x}), \omega_i^2, \alpha_i)$.

To derive the (stationary) autocorrelation structure of the process $\{Y_i(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^2\}$, consider first that (see Minozzo and Ferracuti, 2012), for every $\mathbf{x} \in \mathbb{R}^2$,

$$\mathrm{E}\left[Y_i(\mathbf{x})\middle|Z_i(\mathbf{x})\right] = \beta_i + Y_i(\mathbf{x}) + \omega_i\left(\frac{2}{\pi}\right)^{1/2}\frac{\alpha_i}{(1+\alpha_i^2)^{1/2}},$$
$$\mathrm{Var}\left[Y_i(\mathbf{x})\middle|Z_i(\mathbf{x})\right] = \omega_i^2\left[1 - \frac{2}{\pi}\frac{\alpha_i^2}{(1+\alpha_i^2)}\right].$$

Then, with some algebra, we can derive the autocovariance function and the variogram of the process $\{Y_i(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^2\}$. For instance, for $\mathbf{h} \neq 0$, the variogram is

$$\gamma_{ii}(\mathbf{h}) = \omega_i^2\left[1 - \frac{2}{\pi}\frac{\alpha_i^2}{(1+\alpha_i^2)}\right] + \varsigma_i^2\big(1 - \rho(\mathbf{h})\big). \tag{10}$$

This variogram is discontinuous at zero, that is, $\gamma_{ii}(0) \neq \gamma_{ii}(0^+)$, and we have

$$\gamma_{ii}(0^+) = \omega_i^2\left[1 - \frac{2}{\pi}\frac{\alpha_i^2}{(1+\alpha_i^2)}\right], \qquad \lim_{||\mathbf{h}||\to\infty}\gamma_{ii}(\mathbf{h}) = \omega_i^2\left[1 - \frac{2}{\pi}\frac{\alpha_i^2}{(1+\alpha_i^2)}\right] + \varsigma_i^2.$$

On the other hand, for the process (9) it is easy to verify that, for $\mathbf{h} \neq 0$, the autocovariance function is given by $\mathrm{Cov}\big[Y_i(\mathbf{x}+\mathbf{h}), Y_i(\mathbf{x})\big] = \varsigma_i^2\rho(\mathbf{h})$.

Although the random variables $Y_i(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^2$, are conditionally distributed as Skew-Normals, and also (see, for instance, Azzalini, 2005) marginally distributed as $\mathrm{SN}(\beta_i, \varsigma_i^2 + \omega_i^2, \alpha_i\omega_i/\sqrt{\varsigma_i^2(1+\alpha_i^2)+\omega_i^2})$, such that, for every $\mathbf{x} \in \mathbb{R}^2$, the marginal density of $Y_i(\mathbf{x})$ is given by, for $-\infty < y < \infty$,

$$f_{Y_i}(y) = 2\phi(y - \beta_i; \varsigma_i^2 + \omega_i^2)\Phi\left(\frac{\alpha_i\omega_i}{\sqrt{\varsigma_i^2(1+\alpha_i^2)+\omega_i^2}}\frac{1}{\sqrt{\varsigma_i^2+\omega_i^2}}(y-\beta_i)\right),$$

the other (multivariate) finite-dimensional marginal distributions of the process $\{Y_i(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^2\}$ are not Skew-Normal. However, it is possible to show (see Minozzo and Ferracuti, 2012) that they are Closed Skew-Normal, according to the definition of González-Farías et al. (2004). The knowledge of the marginal distributions of the process $Y_i(\mathbf{x})$ is particularly important for applications since (under the hypothesis of stationarity) in many circumstances they are the only observable distributions of the model.

3 Monte Carlo likelihood inference

Assuming that the number P of latent common factors and the spatial autocorrelation function $\rho(\mathbf{h})$ have already been chosen (as well as the additional parameters of the conditional distributions $f_i(y;M_i(\mathbf{x}))$), the model depends on the parameter vector $\theta = (\beta, \mathbf{A}, \psi)$, where $\beta = (\beta_1, \ldots, \beta_m)^T$, $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_m)^T$, with $\mathbf{a}_i = (a_{i1}, \ldots, a_{iP})$, for $i = 1, \ldots, m$, and $\psi = (\psi_1, \ldots, \psi_m)^T$. Let us note that, as the classical linear factor model, also our model is not identifiable. However, the only indeterminacy stays in a rotation of the matrix $\mathbf{A}$. To perform likelihood inference on the parameter vector θ we would need to maximize, with respect to θ, 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, we have to resort to some simulation techniques like the Monte Carlo EM or the Monte Carlo likelihood.

Our implementation of the MCEM algorithm proceeds as follows. Let us define $\xi = (\xi_1, \ldots, \xi_m)$ where $\xi_i = (\xi_i(\mathbf{x}_1), \ldots, \xi_i(\mathbf{x}_K))^T$, $i = 1, \ldots, m$, and $\mathbf{F} = (\mathbf{F}_1, \ldots, \mathbf{F}_P)$ where $\mathbf{F}_p = (F_p(\mathbf{x}_1), \ldots, F_p(\mathbf{x}_K))^T$, $p = 1, \ldots, P$, and let $f(\mathbf{y}, \xi, \mathbf{F}; \theta)$ 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 θ_{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 $(\xi^{(r)}, \mathbf{F}^{(r)})$, $r = 1, \ldots, R_s$, from the (filtered) conditional distribution $f(\xi, \mathbf{F} | \mathbf{y}; \theta_{s-1})$;
E step – compute $Q_s(\theta, \theta_{s-1}) = (1/R_s) \sum_{r=1}^{R_s} \ln f(\mathbf{y}, \xi^{(r)}, \mathbf{F}^{(r)}; \theta)$;
M step – take as the new guess θ_s the value of θ which maximizes $Q_s(\theta, \theta_{s-1})$.

In our modeling framework, the S-step of the MCEM algorithm can be implemented through Markov chain Monte Carlo (MCMC) techniques, whereas the M-step typically requires the use of numerical procedures.

When the matrix $\mathbf{A}$ is known, and the conditional distributions $f_i(y;M_i(\mathbf{x}))$ are, for instance, Poisson, Gamma or Binomial, it is possible to show that the complete log-likelihood belongs to the curved exponential family and so, by choosing an appropriate increasing sequence R_s, that the algorithm converges to the maximum likelihood estimate (see Fort and Moulines, 2003). On the other hand, when

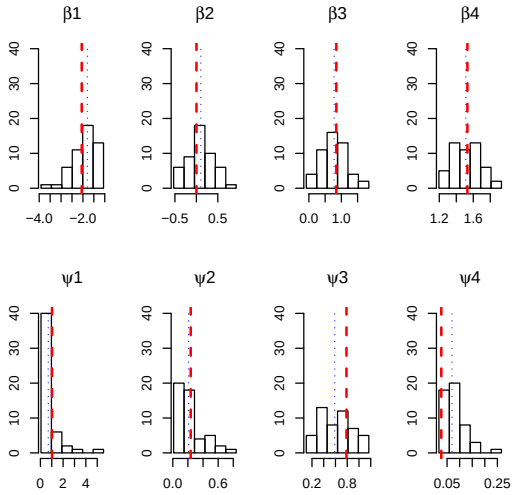

Fig. 1 Histograms of the simulated marginal distributions of the MCEM estimator for the 8 parameters of a model with $m = 4$ Poisson observable variables and $P = 1$ common factor (**A** is known), obtained by running the algorithm over 50 simulated data sets. Dashed lines are the true parameter values; dotted lines are the empirical arithmetic means of the distributions.

the matrix **A** is unknown, the complete log-likelihood does not belong any more to the curved exponential family and theoretical convergence properties are not available. In this case, and assuming, for instance, $f_i(y; M_i(\mathbf{x}))$ to be Poisson, Gamma or Binomial, it is possible to show that the complete log-likelihood to be maximized in the M-step of the MCEM algorithm is concave and so admits just one local maximum. Although this does not guarantee by itself the convergence of the algorithm to some local maximum (of the marginal likelihood), it allows a straightforward computational implementation of the M-step of the algorithm. Despite the lack of theoretical results on the sampling properties of the MCEM estimator, either in the case in which **A** is known or unknown, we found, 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 of a model with $m = 4$ observable Poisson variables, $P = 1$ common factor, and **A** known.

The computational burden of the MCEM algorithm can be substantially reduced by adopting an importance weighting in the stochastic step, avoiding to obtain a new sample at each iteration. After having initialized the number of samples R, the initial guess for the parameters to be estimated $\theta_0 = (\beta_0, \mathbf{A}_0, \psi_0)$, and having generated R samples $(\xi^{(r)}, \mathbf{F}^{(r)})$, $r = 1, \ldots, R$, from the (filtered) conditional distribution $f(\xi, \mathbf{F}|\mathbf{y}; \theta_0)$, and assuming that, after the $(s-1)$th iteration, the current guess for the parameters is given by θ_{s-1}, the sth iteration of the importance sampling MCEM algorithm runs as follows:

IS step – compute the importance weights, for $r = 1, \ldots, R$,

$$w_r = \frac{f(\mathbf{y}|\xi^{(r)}, \mathbf{F}^{(r)}; \theta_{s-1}) \cdot f(\xi^{(r)}; \theta_{s-1})}{f(\mathbf{y}|\xi^{(r)}, \mathbf{F}^{(r)}; \theta_0) \cdot f(\xi^{(r)}; \theta_0)};$$

E step – compute $Q_s(\theta, \theta_{s-1}) = \sum_{r=1}^{R} w_r \ln f(\mathbf{y}, \xi^{(r)}, \mathbf{F}^{(r)}; \theta) / \sum_{r=1}^{R} w_r$;
M step – take as the new guess θ_s the value of θ which maximizes $Q_s(\theta, \theta_{s-1})$.

As before, the maximization involved in the M-step can be carried out by resorting to some numerical procedures.

Alternatively to the MCEM algorithm, to estimate the parameters of the model we might resort to Monte Carlo likelihood. For instance, assuming $\mathbf{A}$ as known, the likelihood can be written as

$$\begin{aligned} L(\beta, \psi) &= f(\mathbf{y}; \beta, \psi) = \int f(\mathbf{y}, \xi, \mathbf{F}; \beta, \psi) \, \mathrm{d}\xi \mathrm{d}\mathbf{F} \\ &= \int \frac{f(\mathbf{y}|\xi, \mathbf{F}; \beta) f(\xi; \psi) f(\mathbf{F})}{f(\mathbf{y}|\xi, \mathbf{F}; \tilde{\beta}) f(\xi; \tilde{\psi}) f(\mathbf{F})} f(\xi, \mathbf{F}|\mathbf{y}; \tilde{\beta}, \tilde{\psi}) f(\mathbf{y}; \tilde{\beta}, \tilde{\psi}) \, \mathrm{d}\xi \mathrm{d}\mathbf{F} \\ &= f(\mathbf{y}; \tilde{\beta}, \tilde{\psi}) \int \frac{f(\mathbf{y}|\xi, \mathbf{F}; \beta) f(\xi; \psi)}{f(\mathbf{y}|\xi, \mathbf{F}; \tilde{\beta}) f(\xi; \tilde{\psi})} f(\xi, \mathbf{F}|\mathbf{y}; \tilde{\beta}, \tilde{\psi}) \, \mathrm{d}\xi \mathrm{d}\mathbf{F} \\ &= f(\mathbf{y}; \tilde{\beta}, \tilde{\psi}) \; \tilde{\mathrm{E}} \left[\frac{f(\mathbf{y}|\xi, \mathbf{F}; \beta) f(\xi; \psi)}{f(\mathbf{y}|\xi, \mathbf{F}; \tilde{\beta}) f(\xi; \tilde{\psi})} \middle| \mathbf{y} \right], \end{aligned} \tag{11}$$

where the expectation is with respect to $f(\xi, \mathbf{F}|\mathbf{y}; \tilde{\beta}, \tilde{\psi})$. To obtain a simulated likelihood surface, we just need to approximate this last expectation by simulating samples from the conditional density $f(\xi, \mathbf{F}|\mathbf{y}; \tilde{\beta}, \tilde{\psi})$, for some appropriate choice of $\tilde{\beta}$ and $\tilde{\psi}$ ($f(\mathbf{y}; \tilde{\beta}, \tilde{\psi})$ is constant with respect to β and ψ). In this case, it is possible to show the hypoconvergence of the Monte Carlo log-likelihood to the exact log-likelihood by following Theorem 2 of Geyer (1994).

4 Conclusions

In this paper we have proposed a modeling framework for the analysis of multivariate spatial data that cannot be assumed Gaussian. Our hierarchical geostatistical factor model can be considered as an extension to non-Gaussian data of the classical and widely used geostatistical proportional covariance model. Basically, we extended this classical model, in two ways: by embedding it in a hierarchical structure, and by introducing unique factors in the latent part of the model. In this way, we managed to obtain a model for non-Gaussian data flexible enough to account for variables with different distributions and variability. Future generalizations could involve the introduction of more spatial scales as in the classical linear model of coregionalization. These would allow to obtain a more flexible spatial autocorrelation structure in which the latent processes $Z_i(\mathbf{x})$, which are responsible for the

level of the observable $Y_i(\mathbf{x})$, are not constrained to have proportional covariance and cross-covariance functions.

About the estimation of the parameters of the model, we considered some computationally intensive likelihood based procedures like the MCEM algorithm and Monte Carlo likelihood. From a computation perspective, although we tested the feasibility of these estimation procedures for some moderate sample sizes, and for reasonably simple models, in more complex situations the computational burden could increase considerably. To solve this problem, future work might consider the development of ad hoc techniques. Let us finally note that in this work we did not tackle important inferential issues such as model validation and model uncertainty, and that we investigated the sampling distributions of the MCEM estimator only through Monte Carlo simulation studies.

Acknowledgements We gratefully acknowledge funding from the Italian Ministry of Education, University and Research (MIUR) through PRIN 2008 project 2008MRFM2H, and Polo Scientifico Didattico 'Studi sull'Impresa' (Vicenza) of the University of Verona.

References

1. Azzalini, A.: The skew-normal distribution and related multivariate families. Scand. J Statist. **32**, 159–188 (2005)
2. Chilès, J.P., Delfiner, P.: Geostatistics — Modeling Spatial Uncertainty. Wiley (1999)
3. Christensen, O.F.: Monte Carlo maximum likelihood in model-based geostatistics. J. Comput. Graph. Stat. **13**, 702–718 (2004)
4. Diggle, P.J., Moyeed, R.A., Tawn, J.A.: Model-based geostatistics (with discussion). Appl. Stat. **47**, 299–350 (1998)
5. Diggle, P.J., Ribeiro Jr., P.J.: Model-Based Geostatistics. Springer (2007)
6. Fort, G., Moulines, E.: Convergence of the Monte Carlo expectation maximization for curved exponential families. Ann. Stat. **31**, 1220–1259 (2003)
7. Geyer, C.J.: On the convergence of Monte Carlo maximum likelihood calculations. J. Roy. Statist. Soc. Ser. B **56**, 261–274 (1994)
8. González-Farías, G., Domínguez-Molina, J.A., Gupta, A.K.: The closed skew-normal distribution. In: Genton, M.G. (ed.) Skew-Elliptical Distributions and Their Applications — A Journey Beyond Normality, pp. 25–42. Chapman & Hall/CRC (2004)
9. Minozzo, M., Ferracuti, L.: On the existence of some skew-normal stationary processes. Chilean Journal of Statistics **3**, 157–170 (2012)
10. Minozzo, M., Ferrari, C.: Multivariate geostatistical mapping of radioactive contamination in the Maddalena Archipelago (Sardinia, Italy). AStA Adv. Stat. Anal. (2012) doi: 10.1007/s10182-012-0201-x
11. Minozzo, M., Fruttini, D.: Loglinear spatial factor analysis: an application to diabetes mellitus complications. Environmetrics **15**, 423–434 (2004)
12. Wei, G.C.G., Tanner, M.A.: A Monte Carlo implementation of the EM algorithm and poor man's data augmentation algorithm. J. Am. Stat. Assoc. **85**, 699–704 (1990)
13. Zhang, H.: Maximum-likelihood estimation for multivariate spatial linear coregionalization models. Environmetrics **18**, 125–139 (2007)
14. Zhu, J., Eickhoff, J.C., Yan, P.: Generalized linear latent variable models for repeated measures of spatially correlated multivariate data. Biometrics **61**, 674–683 (2005)