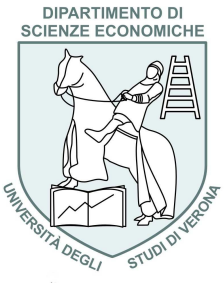

Working Paper Series
Department of Economics
University of Verona

A hierarchical geostatistical factor model for multivariate Poisson count data

Marco Minozzo

WP Number: 22 December 2011

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

Submitted to the Annals of Applied Statistics

A HIERARCHICAL GEOSTATISTICAL FACTOR MODEL FOR MULTIVARIATE POISSON COUNT DATA

By Marco Minozzo[*‡] and Clarissa Ferrari[†]

University of Verona

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 recover some of the optimality properties enjoyed by the predictions supplied by these techniques in the case of Gaussian data, in particular unbiasedness and minimum mean square error, we propose a hierarchical geostatistical factor model extending, following a model-based geostatistical approach, the classical geostatistical proportional covariance model. For this model we investigate a likelihood-based inferential procedure adopting the Monte Carlo EM algorithm and show its sampling performances mainly through some thorough simulation studies. Besides, we also show how mapping of the latent factors can be efficiently carried out exploiting a linearity property of the predictions.

1. Introduction. In recent years, the analysis of multivariate spatial data sets has received more and more attention. This is mainly due to the increasing possibility to gather large amount of georeferenced data. For instance, in many environmental applications, data on more than one variable

[*]Marco Minozzo is Associate Professor of Statistics at the Department of Economics, of the University of Verona.

[†]Clarissa Ferrari is Post Doc Researcher at the Department of Economics of the University of Verona.

[‡]This work was partially supported by a grant from the Italian Ministry of Education, University and Research (MIUR) through PRIN 2008 project 2008MRFM2H. The authors express thanks to Alessandro Ludovisi, Piera Pandolfi and Maria I. Taticchi, Department of Animal Biology and Ecology, University of Perugia, for access to the plankton data.

AMS 2000 subject classifications: Primary 62H11, 62M30, 62P12; secondary 65C05, 62H25

Keywords and phrases: Generalized linear mixed model, Markov chain Monte Carlo, Model-based geostatistics, Monte Carlo EM, Multivariate geostatistics, Spatial prediction

are frequently gathered at different spatial locations over a region of interest. Traditionally, multivariate geostatistics has been the reference discipline for this kind of problems and the standard techniques for the analysis of these data have been cokriging, the linear model of coregionalization (LMC), or the simpler proportional covariance model, otherwise known as intrinsic correlation model, and the related 'factorial kriging analysis' (Matheron 1982). These models have provided the generalization to geostatistics of the classical linear factor model and of principal component analysis. A crucial issue of these generalizations has been the coherent modeling of the multivariate spatial autocorrelation structure using admissible covariance functions (see for some detail Grzebyk and Wackernagel 1994; and Wackernagel 1998).

Though this body of classical geostatistical tools has been developed without making any distributional assumption, and considering only first and second order properties, that is, considering only covariance and variogram functions, many authors have also studied these models under the assumption of Gaussianity. For instance, interesting developments of the LMC in a Bayesian framework can be found in Schmidt and Gelfand (2003) and in Gelfand, Schmidt, Banerjee and Sirmans (2004), whereas a maximum likelihood analysis via the expectation-maximization (EM) algorithm has been carried out by Zhang (2007). More recently, Banerjee, Gelfand, Finley and Sang (2008) have developed a flexible model for Gaussian data based on the decomposition by basic functions and kernel convolutions of a spatial predictive process.

Now, as far as prediction of the underlying factors is concerned, in the case in which the observed data cannot be assumed Gaussian, the classical LMC, or better the associated factorial cockriging predictions, and the above Gaussian developments, are not appropriate any more, that is, they do not provide, in general, optimal (with minimum mean square prediction error) unbiased predictions, since their linear predictors do not coincide any more with the desirable (non necessary linear in the data) conditional expectations. To cope with similar problems, in the univariate case, a model-based approach to geostatistics for non-Gaussian data based on generalized linear mixed models has been advanced by Diggle, Moyeed and Tawn (1998) in a Bayesian inferential framework using Markov chain Monte Carlo (MCMC) techniques (see also Cressie 2001; Banerjee, Carlin and Gelfand 2004; Diggle and Ribeiro 2007). Subsequently, for these models, Zhang (2002) also derived maximum likelihood estimates by using a Monte Carlo version of the EM gradient algorithm, whereas Christensen (2004) studied the use of Monte Carlo maximum likelihood.

Following these non-Gaussian geostatistical developments, some authors

also proposed extensions of classical multivariate geostatistical models, in particular of the LMC, to non-Gaussian data. Some of the first attempts in this direction are due to Wang and Wall (2003), where a common factor model has been developed in a Bayesian framework, Minozzo and Fruttini (2004), where an extension of the proportional covariance model has been proposed following a hierarchical model-based approach, and to Zhu, Eickhoff and Yan (2005), where maximum likelihood inference using a Monte Carlo EM algorithm (MCEM) is proposed for a spatio-temporal extension of Wang and Wall's work. Let us mention that other approaches for the modeling of multivariate geostatistical data, have been proposed by Reich and Fuentes (2007), who considered a framework in which they used a stick-breaking prior in a semiparametric Bayesian context, by Christensen and Amemiya (2002), with a framework in which no distributional assumptions are made, and by Bailey and Krzanowski (2000) and Calder (2007) who considered various extensions of the classical LMC.

In this work, adopting a model-based geostatistical approach (see Diggle et al. 1998) we consider an extension of the classical geostatistical proportional covariance model for Poisson count data. In particular, our model extends the proportional covariance model by considering, in addition to common factors, also specific factors, as in the classical linear factor model, and by embedding the resulting linear factor structure in a hierarchical construction to allow for observable Poisson variables. In this way we manage to obtain a quite flexible model that allows to accommodate count variables with different variability. For the estimation of the parameters of our model, since the integration of the complete likelihood is not analytically feasible, and so neither direct maximization of the likelihood, nor the standard EM algorithm are practically possible, we develop an MCEM algorithm to carry out maximum likelihood inference (see, for instance, Wei and Tanner 1990; Booth and Hobert 1999; McLachlan and Krishnan 2008). Then, for this algorithm, we discuss in some detail its maximization step and show that, under some conditions, it converges (almost surely) to the maximum likelihood estimate. Moreover, we show through some extensive simulation studies that it also have acceptable sampling properties. Besides, inspired by the results of Zhang (2002), derived in a univariate model-based geostatistical context, we also show that our model own a nice linearity property (similarly to linear Kriging) that allows to diminish considerably the computational burden required by the mapping of the latent factors. Indeed, this linearity property of the conditional expectation of the latent factors allows to avoid to carry out an MCMC run at each node of a mapping grid and to exploit the predictions at the sampling sites to obtain the prediction at any other site.

The paper is organized as follows. In Section 2.1 we present our hierarchical geostatistical factor model whereas in Section 2.2 we derive the autocorrelation functions and the variograms for the observable data. In Section 3 we formalize our inferential procedure based on the MCEM algorithm; some details are given in the Appendices B and C. Besides, in Section 3.3 and in Appendix D we discuss an efficient way to build maps of the latent factors. Simulation results are presented in Section 4 whereas an environmental application on biological count data is discussed in Section 5. Some final remarks conclude the paper. All the appendices are available as Supplementary Material.

2. The hierarchical geostatistical factor model.

2.1. *The model.* In the following we extend the ideas put forward by Diggle et al. (1998) in the univariate case for the modeling of non-Gaussian spatial data, to the multivariate case by introducing a hierarchical (generalized linear mixed) model in which the latent structure is given by a Gaussian spatial factor model. Let $y_i(\mathbf{x}_k)$, $i = 1, \ldots, m$, $k = 1, \ldots, K$, be a set of georeferenced 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, we assume that, for any given i and $\mathbf{x}$, 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 $h_i(\cdot)$. Specifically, we assume that the data are conditionally Poisson distributed, that is, that, for any i,

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

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 $\ln(M_i(\mathbf{x})) = \beta_i + Z_i(\mathbf{x})$.

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

$$Z_i(\mathbf{x}) = \sum_{p=1}^{P} a_{ip}\, F_p(\mathbf{x}) + \xi_i(\mathbf{x}), \tag{4}$$

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 $Z_i(\mathbf{x})$, 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. The latent factors $F_p(\mathbf{x})$ and $\xi_i(\mathbf{x})$ are assumed to be stationary Gaussian processes with $\mathrm{E}[F_p(\mathbf{x})] = \mathrm{E}[\xi_i(\mathbf{x})] = 0$ and covariance functions

$$\mathrm{Cov}[F_p(\mathbf{x}), F_q(\mathbf{x}+\mathbf{h})] = \mathrm{E}[F_p(\mathbf{x}) \cdot F_q(\mathbf{x}+\mathbf{h})] = \begin{cases} \rho(\mathbf{h}), & p = q, \\ 0, & p \neq q, \end{cases} \tag{5}$$

$$\mathrm{Cov}[\xi_i(\mathbf{x}), \xi_j(\mathbf{x}+\mathbf{h})] = \mathrm{E}[\xi_i(\mathbf{x}) \cdot \xi_j(\mathbf{x}+\mathbf{h})] = \begin{cases} \psi_i \rho(\mathbf{h}), & i = j, \\ 0, & i \neq j, \end{cases} \tag{6}$$

where $\mathbf{h} \in \mathbb{R}^2$ and $\rho(\mathbf{h})$ is a real spatial correlation function common to all factors with $\rho(0) = 1$ and $\rho(\mathbf{h}) \to 0$, as $||\mathbf{h}|| \to \infty$, while 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, that is,

$$\mathrm{Cov}[F_p(\mathbf{x}), \xi_i(\mathbf{x}+\mathbf{h})] = 0, \tag{7}$$

for all p and i. Figure 1 shows, at every spatial location $\mathbf{x}$, the directed acyclic independence graph representing the conditional independence structure of the hierarchical geostatistical factor model.

Let us note that (see Appendix A in Supplementary Material) the latent processes $Z_i(\mathbf{x})$, $i = 1, \ldots, m$, have 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 cross-covariance functions $\mathrm{Cov}[Z_i(\mathbf{x}), Z_j(\mathbf{x}+\mathbf{h})] = \sum_{p=1}^{P} a_{ip} a_{jp} \rho(\mathbf{h})$. Of course, taking $\mathbf{h} = 0$, we see that $\mathrm{Var}[Z_i(\mathbf{x})] = \sum_{p=1}^{P} a_{ip}^2 + \psi_i$, and that $\mathrm{Cov}[Z_i(\mathbf{x}), Z_j(\mathbf{x})] = \sum_{p=1}^{P} a_{ip} a_{jp}$.

For a given choice of the number P of common factors and of the autocorrelation function $\rho(\mathbf{h})$, the model depends on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{A}, \boldsymbol{\psi})$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)^T$, $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_m)^T$, with $\mathbf{a}_i = (a_{i1}, \ldots, a_{iP})$, for $i = 1, \ldots, m$, and $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_m)^T$.

Let us remark that, as the classical factor model, our model is not identifiable. Indeed, for any orthogonal $P \times P$ matrix $\mathbf{G}$ with elements g_{qp}

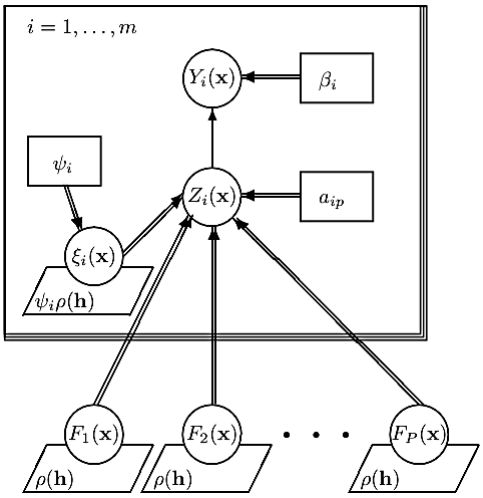

FIG 1. *Hierarchical geostatistical factor model. Directed acyclic independence graph representing the dependence structure among the elements of the hierarchical geostatistical factor model at any given spatial location* $\mathbf{x}$. *'Floors' below factors* $F_p(\mathbf{x})$ *and* $\xi_i(\mathbf{x})$ *contains the corresponding spatial autocovariance functions.*

$(q, p = 1, \ldots, P)$, such that $\mathbf{G}\mathbf{G}^T = \mathbf{G}^T\mathbf{G} = \mathbf{I}$, we have that the spatial processes $F_q^{(G)}(\mathbf{x}) = \sum_{p=1}^{P} g_{qp}F_p(\mathbf{x})$, $q = 1, \ldots, P$, are still Gaussian with mean zero, variance one, and covariance function equal to (5), so that they are not distinguishable from the common factors $F_p(\mathbf{x})$, $p = 1, \ldots, P$. However, this is the only indeterminacy in the model and we will see in the following how we can deal with it.

2.2. *Variograms and cross-variograms.* Under the assumptions of the previous section, it is easy to see that the random functions $Y_i(\mathbf{x})$, $i = 1, \ldots, m$, are second-order stationary. In fact, defining $\varsigma_i^2 = \mathrm{Var}[Z_i(\mathbf{x})]$ and $\tau_i^* = \exp\{\beta_i + \varsigma_i^2/2\}$, and considering that, for every $\mathbf{x} \in \mathbb{R}^2$, $\mathrm{E}[\exp\{Z_i(\mathbf{x})\}] = \exp\{\varsigma_i^2/2\}$ and $\mathrm{E}[\exp\{2Z_i(\mathbf{x})\}] = \exp\{2\varsigma_i^2\}$, and also that under the Poisson assumption $\mathrm{Var}[Y_i(\mathbf{x})|Z_i(\mathbf{x})] = M_i(\mathbf{x})$, with some algebra we have that

$$\begin{aligned} \mu_i &= \mathrm{E}[Y_i(\mathbf{x})] = \mathrm{E}[\mathrm{E}[Y_i(\mathbf{x})|Z_i(\mathbf{x})]] = \mathrm{E}[\exp\{\beta_i + Z_i(\mathbf{x})\}] = \tau_i^*, \\ \sigma_i^2 &= \mathrm{Var}[Y_i(\mathbf{x})] = \mathrm{E}[M_i(\mathbf{x})] + \mathrm{Var}[M_i(\mathbf{x})] = \tau_i^* + (\tau_i^*)^2[\exp\{\varsigma_i^2\} - 1], \end{aligned}$$

and

$$C_{ii}(\mathbf{h}) = \mathrm{Cov}[Y_i(\mathbf{x}), Y_i(\mathbf{x}+\mathbf{h})] = (\tau_i^*)^2[\exp\{\varsigma_i^2\rho(\mathbf{h})\} - 1], \quad \mathbf{h} \neq 0. \tag{8}$$

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.

Let us note that, if $h_i(M_i(\mathbf{x})) = \mathbf{d}_i^T(\mathbf{x}) \cdot \boldsymbol{\beta}^{(i)} + Z_i(\mathbf{x})$, where $\mathbf{d}_i(\mathbf{x})$ is a vector of covariates depending on $\mathbf{x}$, and $\boldsymbol{\beta}^{(i)}$ is a vector of parameters, then the random functions $Y_i(\mathbf{x})$ would not be stationary any more.

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

$$\gamma_{ii}(\mathbf{h}) = \frac{1}{2}\operatorname{Var}\big[Y_i(\mathbf{x}+\mathbf{h}) - Y_i(\mathbf{x})\big] = \tau_i^* + (\tau_i^*)^2\big[\exp\{\varsigma_i^2\} - \exp\{\varsigma_i^2\rho(\mathbf{h})\}\big], \tag{9}$$

(see Appendix A in Supplementary Material). Let us note that the variograms of the observable $Y_i(\mathbf{x})$ is always equal to zero at the origin and have nugget $\gamma_{ii}(0^+) = \tau_i^*$ that equals the mean μ_i and is, in general, strictly greater than 0. It is also easy to check that the sill of the variogram satisfies $\gamma_{ii}(\infty) = C_{ii}(0) = \sigma_i^2$.

On the other hand, the random functions $Y_i(\mathbf{x})$, $i = 1, \ldots, m$, are also jointly second order stationary. For any two random functions $Y_i(\mathbf{x})$ and $Y_j(\mathbf{x})$, with $i \neq j$, considering that $\mathrm{E}[\exp\{Z_i(\mathbf{x}) + Z_j(\mathbf{x}+\mathbf{h})\}] = \exp\{\varsigma_i^2/2 + \varsigma_j^2/2 + \varsigma_{ij}\rho(\mathbf{h})\}$, where $\varsigma_{ij} = \operatorname{Cov}[Z_i(\mathbf{x}), Z_j(\mathbf{x})]$, we have that

$$C_{ij}(\mathbf{h}) = \operatorname{Cov}[Y_i(\mathbf{x}), Y_j(\mathbf{x}+\mathbf{h})] = \tau_i^*\tau_j^*[\exp\{\varsigma_{ij}\rho(\mathbf{h})\} - 1]. \tag{10}$$

Note that $C_{ij}(\mathbf{h}) = C_{ji}(\mathbf{h})$ and that, if $\rho(\mathbf{h}) = \rho(-\mathbf{h})$, we have $C_{ij}(-\mathbf{h}) = C_{ij}(\mathbf{h})$.

For the cross-variogram between the observable $Y_i(\mathbf{x})$ and $Y_j(\mathbf{x})$, with $i \neq j$, we have,

$$\begin{aligned} \gamma_{ij}(\mathbf{h}) &= \frac{1}{2}\operatorname{Cov}\big[(Y_i(\mathbf{x}+\mathbf{h}) - Y_i(\mathbf{x})), (Y_j(\mathbf{x}+\mathbf{h}) - Y_j(\mathbf{x}))\big] \\ &= \tau_i^*\tau_j^*\big[\exp\{\varsigma_{ij}\} - \exp\{\varsigma_{ij}\rho(\mathbf{h})\}\big], \end{aligned} \tag{11}$$

(see Appendix A in Supplementary Material). Notice that for the cross-variogram we have $\gamma_{ij}(0) = \gamma_{ij}(0^+) = 0$ and $\gamma_{ij}(\infty) = \tau_i^*\tau_j^*[\exp\{\varsigma_{ij}\} - 1] = C_{ij}(0)$, which can be smaller or greater than zero, depending on the sign of ς_{ij}. The theoretical variogram $\gamma_{ii}(\mathbf{h})$ and the theoretical cross-variogram $\gamma_{ij}(\mathbf{h})$ are depicted in Figure 2.

3. Prediction and estimation.

3.1. *Prediction.* One of the main reasons for adopting a factor model stems from the possibility to simplify a multivariate data set by detecting one or more common latent factors responsible for the dependencies in the observations from the phenomenon under study. From this standpoint it is of paramount importance to be able to obtain predictions of the latent

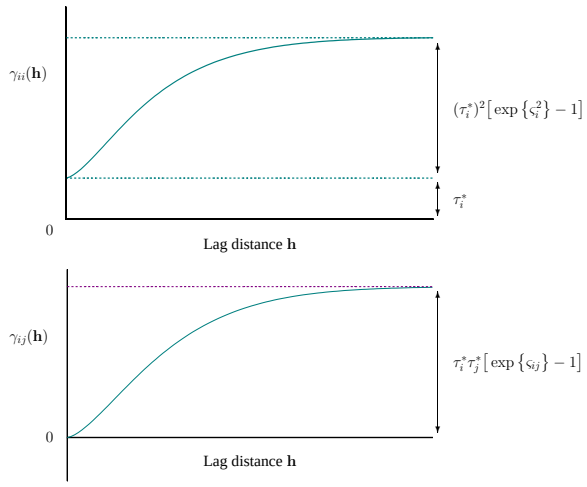

FIG 2. *Theoretical variogram* $\gamma_{ii}(\mathbf{h})$ *and cross-variogram* $\gamma_{ij}(\mathbf{h})$ *of the observable* $Y_i(\mathbf{x})$, $i = 1, \ldots, m$, *assuming an autocorrelation function* $\rho(\mathbf{h}) = \exp\{-((\alpha||\mathbf{h}||)^{\delta})\}$. *Various shapes can be obtained by appropriately choosing values for the parameters* β_i, ς_i^2 *and* ς_{ij}, *and for the parameters* α *and* δ *of the autocorrelation function.*

factors $F_p(\mathbf{x})$ and $\xi_i(\mathbf{x})$ at a given set of spatial locations. To describe the problem, let $\mathbf{x}_0 \in \mathbb{R}^2$, be a given spatial location, and let $F_p(\mathbf{x})$ be a common factor of interest. Assuming that the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{A}, \boldsymbol{\psi})$, of the model is known, the optimum (at minimum mean square error) unbiased prediction of the factor $F_p(\mathbf{x})$ at the spatial location $\mathbf{x}_0$ is provided by the conditional expectation of $F_p(\mathbf{x}_0)$, given the data $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_m)$, where $\mathbf{y}_i = (y_i(\mathbf{x}_1), \ldots, y_i(\mathbf{x}_K))^T$, $i = 1, \ldots, m$, that is, by

$$\mathrm{E}[F_p(\mathbf{x}_0)|\mathbf{y}] = \int_{\mathbb{R}} u\, f_{F_p(\mathbf{x}_0)}(u|\mathbf{y};\boldsymbol{\theta})\, \mathrm{d}u, \tag{12}$$

where $f_{F_p(\mathbf{x}_0)}(u|\mathbf{y};\boldsymbol{\theta})$ (that we sometimes write as $f(F_p(\mathbf{x}_0)|\mathbf{y};\boldsymbol{\theta})$) is the conditional distribution of the common factor given the data. On the other hand, a measure of accuracy of this prediction is given by the conditional variance $\mathrm{Var}[F_p(\mathbf{x}_0)|\mathbf{y}]$. Although these conditional moments are difficult to evaluate analytically and are not available in close form, rather accurate numerical evaluations of their values can be obtained by simulation using MCMC methods. In particular, the integral in (12) can be evaluated by resorting to the Metropolis-Hastings (M-H) algorithm to approximate the conditional distribution $f_{F_p(\mathbf{x}_0)}(u|\mathbf{y};\boldsymbol{\theta})$.

The basis of the M-H algorithm lies in the simulation of a Markov chain whose equilibrium distribution is the required target probability function

that we need to approximate (see, for instance, Smith and Roberts (1993)). A generic M-H iteration, that is, a generic step of the chain, involves sampling a candidate value t^* for the next value of the chain from a proposal distribution $q(t^*|t^{(i)})$, where $t^{(i)}$ is the current value of the chain. Then, the Markov chain moves from the current value $t^{(i)}$ to the candidate value t^* with acceptance probability $\Delta(t^{(i)}, t^*) = \pi(t^*) \cdot q(t^{(i)}|t^*)/(\pi(t^{(i)}) \cdot q(t^*|t^{(i)}))$; and remains at the current state $t^{(i)}$ with probability $1 - \Delta(t^{(i)}, t^*)$.

In our case, taking $f(F_p(\mathbf{x}_0)|\mathbf{y};\boldsymbol{\theta})$ as our target distribution to approximate, and considering that it can be rewritten as

$$f(F_p(\mathbf{x}_0)|\mathbf{y};\boldsymbol{\theta}) = \frac{f(F_p(\mathbf{x}_0),\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})} \propto f(\mathbf{y}|F_p(\mathbf{x}_0);\boldsymbol{\theta})\ f(F_p(\mathbf{x}_0);\boldsymbol{\theta}), \tag{13}$$

and also assuming some proposal distribution $q(F_p^*(\mathbf{x}_0)|F_p^{(i)}(\mathbf{x}_0))$, the acceptance probability is given by

$$\Delta(F_p^{(i)}(\mathbf{x}_0), F_p^*(\mathbf{x}_0)) = \frac{f(\mathbf{y}|F_p^*(\mathbf{x}_0);\boldsymbol{\theta})\ f(F_p^*(\mathbf{x}_0);\boldsymbol{\theta})\ q(F_p^{(i)}(\mathbf{x}_0)|F_p^*(\mathbf{x}_0))}{f(\mathbf{y}|F_p^{(i)}(\mathbf{x}_0);\boldsymbol{\theta})\ f(F_p^{(i)}(\mathbf{x}_0);\boldsymbol{\theta})\ q(F_p^*(\mathbf{x}_0)|F_p^{(i)}(\mathbf{x}_0))}. \tag{14}$$

Analogously, for the prediction of the unique factor $\xi_i(\mathbf{x})$, we would like to obtain the conditional expectation

$$\mathrm{E}[\xi_i(\mathbf{x}_0)|\mathbf{y}] = \int_{\mathbb{R}} v\, f_{\xi_i(\mathbf{x}_0)}(v|\mathbf{y};\boldsymbol{\theta})\, \mathrm{d}v, \tag{15}$$

where $f_{\xi_i(\mathbf{x}_0)}(v|\mathbf{y};\boldsymbol{\theta})$ (that we sometimes write as $f(\xi_i(\mathbf{x}_0)|\mathbf{y};\boldsymbol{\theta})$) is the conditional distribution of the unique factor given the data, and the conditional variance $\mathrm{Var}[\xi_i(\mathbf{x}_0)|\mathbf{y}]$. Thus, considering that we can write

$$f(\xi_i(\mathbf{x}_0)|\mathbf{y};\boldsymbol{\theta}) = \frac{f(\xi_i(\mathbf{x}_0),\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})} \propto f(\mathbf{y}|\xi_i(\mathbf{x}_0);\boldsymbol{\theta})\ f(\xi_i(\mathbf{x}_0);\boldsymbol{\theta}), \tag{16}$$

and assuming a proposal distribution $q(\xi_i^*(\mathbf{x}_0)|\xi_i^{(i)}(\mathbf{x}_0))$, the acceptance probability is given by

$$\Delta(\xi_i^{(i)}(\mathbf{x}_0), \xi_i^*(\mathbf{x}_0)) = \frac{f(\mathbf{y}|\xi_i^*(\mathbf{x}_0);\boldsymbol{\theta})\ f(\xi_i^*(\mathbf{x}_0);\boldsymbol{\theta})\ q(\xi_i^{(i)}(\mathbf{x}_0)|\xi_i^*(\mathbf{x}_0))}{f(\mathbf{y}|\xi_i^{(i)}(\mathbf{x}_0);\boldsymbol{\theta})\ f(\xi_i^{(i)}(\mathbf{x}_0);\boldsymbol{\theta})\ q(\xi_i^*(\mathbf{x}_0)|\xi_i^{(i)}(\mathbf{x}_0))}. \tag{17}$$

Once predictions of the factors $F_p(\mathbf{x})$ and $\xi_i(\mathbf{x})$ have been obtained, predictions of functionals of interest could also be obtained from these.

3.2. *Estimation via the Monte Carlo EM algorithm.* Assuming that the number P of common factors and the spatial autocorrelation function $\rho(\mathbf{h})$ have already been chosen, adopting a non-Bayesian inferential framework, likelihood inference would require the maximization, over the parametric space of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{A}, \boldsymbol{\psi})$, of the likelihood based on the marginal density function of the observations $\mathbf{y}$. However, since this marginal density is not available, and since the integration required in the E-step of the EM algorithm (Dempster, Laird and Rubin 1977) would not be easy, we have to resort to some stochastic versions of the EM algorithm, like the stochastic EM (StEM) algorithm or the MCEM algorithm (see, for instance, Wei and Tanner 1990; Nielsen 2000; Kuhn and Lavielle 2004; McLachlan and Krishnan 2008).

As far as the identifiability of the parameters of the model is concerned, in the following discussion we will not impose any constrain on the coefficients of the $m \times P$ matrix $\mathbf{A}$. Indeed the EM algorithm is guaranteed to increase monotonically the marginal likelihood and it is not affected by the non-identifiability of the model. Defining $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m)$ where $\boldsymbol{\xi}_i = (\xi_i(\mathbf{x}_1), \ldots, \xi_i(\mathbf{x}_K))^T$, $i = 1, \ldots, m$, and $\mathbf{F} = (\mathbf{F}_1, \ldots, \mathbf{F}_P)$ where $\mathbf{F}_p = (F_p(\mathbf{x}_1), \ldots, F_p(\mathbf{x}_K))^T$, $p = 1, \ldots, P$, the complete log-likelihood based on the joint distribution $f(\mathbf{y}, \boldsymbol{\xi}, \mathbf{F}; \boldsymbol{\theta})$, accounting also for the unobservable factors, can be written, by exploiting the conditional independence properties of the model, as

$$
\begin{aligned}
(18)\; l_c(\boldsymbol{\theta}) &= \ln f(\mathbf{y}, \boldsymbol{\xi}, \mathbf{F}; \boldsymbol{\theta}) = \ln(f(\mathbf{y}|\boldsymbol{\xi}, \mathbf{F}; \boldsymbol{\theta})\, f(\boldsymbol{\xi}, \mathbf{F}; \boldsymbol{\theta})) \\
&= \ln\left(\left[\prod_{i=1}^{m}\prod_{k=1}^{K} f_i(y_{ik}; M_{ik})\right] \cdot f(\boldsymbol{\xi}; \boldsymbol{\psi}) \cdot f(\mathbf{F})\right) \\
&= \ln\left(\left[\prod_{i=1}^{m}\prod_{k=1}^{K} \frac{\exp(-M_{ik})(M_{ik})^{y_{ik}}}{y_{ik}!}\right] \cdot \left[\prod_{i=1}^{m} f(\boldsymbol{\xi}_i)\right] \cdot \left[\prod_{p=1}^{P} f(\mathbf{F}_p)\right]\right),
\end{aligned}
$$

where, $f(\boldsymbol{\xi}_i)$, $i = 1, \ldots, m$, and $f(\mathbf{F}_p)$, $p = 1, \ldots, P$, are K-multivariate Normals, and, for ease of notation, we put $y_{ik} = y_i(\mathbf{x}_k)$ and $M_{ik} = M_i(\mathbf{x}_k)$. Let us note that M_{ik}, $i = 1, \ldots, m$, and $k = 1, \ldots, K$, depend on the parameters $\boldsymbol{\beta}$ and $\mathbf{A}$, whereas $f(\boldsymbol{\xi}_i)$, $i = 1, \ldots, m$, depend on the parameter $\boldsymbol{\psi}$. On the other hand, likelihood inference requires the maximization of the marginal log-likelihood $l(\boldsymbol{\theta}) = \ln f(\mathbf{y}; \boldsymbol{\theta})$ based on the marginal density of the data, that is, on the integral, with respect to the latent factors,

$$
(19) \qquad f(\mathbf{y}; \boldsymbol{\theta}) = \int_{\boldsymbol{\Xi}} \int_{\boldsymbol{\digamma}} f(\mathbf{y}, \boldsymbol{\xi}, \mathbf{F}; \boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\xi} \mathrm{d}\mathbf{F},
$$

where $\boldsymbol{\Xi} = \mathbb{R}^{K \mathrm{x} m}$ and $\boldsymbol{\digamma} = \mathbb{R}^{K \mathrm{x} P}$.

As already noted, since this integral cannot be solved analytically, to maximize the log-likelihood $l(\boldsymbol{\theta})$ to obtain the maximum likelihood estimate of $\boldsymbol{\theta}$ we will resort to the MCEM algorithm. Our implementation of the algorithm proceeds iteratively as follows. 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 $(\boldsymbol{\xi}^{(r)}, \mathbf{F}^{(r)})$, $r = 1,\ldots, R_s$, from the (filtered) conditional distribution $f(\boldsymbol{\xi}, \mathbf{F}|\mathbf{y}; \boldsymbol{\theta}_{s-1})$;

E step – compute

$$Q_s(\boldsymbol{\theta}, \boldsymbol{\theta}_{s-1}) = \frac{1}{R_s} \sum_{r=1}^{R_s} \ln f(\mathbf{y}, \boldsymbol{\xi}^{(r)}, \mathbf{F}^{(r)}; \boldsymbol{\theta}), \tag{20}$$

as an approximation of the conditional expectation

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{s-1}) &= \mathrm{E}_{\boldsymbol{\theta}_{s-1}}[\ln f(\mathbf{y}, \boldsymbol{\xi}, \mathbf{F}; \boldsymbol{\theta})|\mathbf{y}] \\ &= \int_{\boldsymbol{\Xi}} \int_{F} (\ln f(\mathbf{y}, \boldsymbol{\xi}, \mathbf{F}; \boldsymbol{\theta})) f(\boldsymbol{\xi}, \mathbf{F}|\mathbf{y}; \boldsymbol{\theta}_{s-1}) \, \mathrm{d}\boldsymbol{\xi} \mathrm{d}\mathbf{F}; \end{aligned}$$

M step – take as the new guess $\boldsymbol{\theta}_s$ the value of $\boldsymbol{\theta}$ which maximizes $Q_s(\boldsymbol{\theta}, \boldsymbol{\theta}_{s-1})$.

Choosing $R_s = 1$ this procedure yields the StEM algorithm, whereas for R_s very large it approximates the deterministic EM algorithm. Moreover, a simulated annealing version could be obtained by choosing a slowly increasing sequence $R_s \to \infty$, as $s \to \infty$, (see Lavielle and Moulines 1997; Fort and Moulines 2003).

The evaluation of $Q_s(\boldsymbol{\theta}, \boldsymbol{\theta}_{s-1})$ in (20) requires samples $(\boldsymbol{\xi}^{(r)}, \mathbf{F}^{(r)})$, $r = 1,\ldots, R_s$, from the conditional distribution $f(\boldsymbol{\xi}, \mathbf{F}|\mathbf{y}; \boldsymbol{\theta}_{s-1}) = f(\mathbf{y}, \boldsymbol{\xi}, \mathbf{F}; \boldsymbol{\theta}_{s-1})/f(\mathbf{y}; \boldsymbol{\theta}_{s-1})$, proportional to $f(\mathbf{y}|\boldsymbol{\xi}, \mathbf{F}; \boldsymbol{\theta}_{s-1})\, f(\boldsymbol{\xi}, \mathbf{F}; \boldsymbol{\theta}_{s-1})$, which can be generated, for instance, via MCMC methods. To this end, the M-H algorithm can be applied in steps to subsets of $(\boldsymbol{\xi}, \mathbf{F})$ (see, for instance, Smith and Robert 1993). In particular, we can proceed on the subsets $(\boldsymbol{\xi})$ and $(\mathbf{F})$ applying the M-H algorithm alternatively to the target distributions $f(\boldsymbol{\xi}|\mathbf{F}, \mathbf{y}; \boldsymbol{\theta}_{s-1}) \propto f(\mathbf{y}|\boldsymbol{\xi}, \mathbf{F}; \boldsymbol{\theta}_{s-1})\, f(\boldsymbol{\xi}; \boldsymbol{\theta}_{s-1})$ and $f(\mathbf{F}|\boldsymbol{\xi}, \mathbf{y}; \boldsymbol{\theta}_{s-1}) \propto f(\mathbf{y}|\boldsymbol{\xi}, \mathbf{F}; \boldsymbol{\theta}_{s-1})\, f(\mathbf{F}; \boldsymbol{\theta}_{s-1})$.

For the M-step, in case of the StEM algorithm, that is, for $R_s = 1$, we have to maximize the complete log-likelihood $l_c(\boldsymbol{\theta}) = Q_s(\boldsymbol{\theta}, \boldsymbol{\theta}_{s-1}) = \ln f(\mathbf{y}, \boldsymbol{\xi}^{(1)}, \mathbf{F}^{(1)}; \boldsymbol{\theta})$. With some algebra, this can be conveniently decomposed as

$$l_c(\boldsymbol{\theta}) = \sum_{i=1}^{m} l_i(\beta_i, \mathbf{a}_i) + \sum_{i=1}^{m} \ln f(\boldsymbol{\xi}_i^{(1)}; \psi_i) + \sum_{p=1}^{P} \ln f(\mathbf{F}_p^{(1)}), \tag{21}$$

where

$$l_i(\beta_i, \mathbf{a}_i) = \sum_{k=1}^{K} \big[- M_{ik} + y_{ik} \ln(M_{ik}) - \ln(y_{ik}!) \big],$$

$$\ln f(\boldsymbol{\xi}_i^{(1)}; \psi_i) = -\frac{K}{2} \ln(2\pi\psi_i) \, - \frac{1}{2} \ln |\mathbf{P}_\rho| \, - \frac{1}{2\psi_i} (\boldsymbol{\xi}_i^{(1)})^T \mathbf{P}_\rho^{-1} \boldsymbol{\xi}_i^{(1)},$$

$$\ln f(\mathbf{F}_p^{(1)}) = -\frac{K}{2} \ln(2\pi) \, - \frac{1}{2} \ln |\mathbf{P}_\rho| \, - \frac{1}{2} (\mathbf{F}_p^{(1)})^T \mathbf{P}_\rho^{-1} \mathbf{F}_p^{(1)},$$

and $\mathbf{P}_\rho$ is the $K \times K$ matrix with elements $\rho(\mathbf{x}_l - \mathbf{x}_k)$, which depend on the autocorrelation function $\rho(\mathbf{h})$.

Proceeding in steps, the values of the parameters β_i and $\mathbf{a}_i$ maximizing $l_c(\boldsymbol{\theta})$, for every $i = 1, \ldots, m$, can be obtained by considering the terms $l_i(\beta_i, \mathbf{a}_i)$. Analytic maximization leads to a system of $P + 1$ nonlinear equations which can be solved numerically. The maximization of $l_c(\boldsymbol{\theta})$ with respect to the parameters ψ_i, instead, can be accomplished by considering the terms $\ln f(\boldsymbol{\xi}_i^{(1)}; \psi_i)$, separately for every $i = 1, \ldots, m$. Equating to zero the first derivatives of $\ln f(\boldsymbol{\xi}_i^{(1)}; \psi_i)$ with respect to ψ_i leads to the point of maximum $\psi_i = (1/K)(\boldsymbol{\xi}_i^{(1)})^T \mathbf{P}_\rho^{-1} \boldsymbol{\xi}_i^{(1)}$.

In the case of the MCEM algorithm, that is, with $R_s > 1$, the M-step of the algorithm requires the maximization of the 'average' log-likelihood

(22)
$$\sum_{r=1}^{R_s} l_c^{(r)}(\boldsymbol{\theta}) = \sum_{i=1}^{m} \sum_{r=1}^{R_s} l_i^{(r)}(\beta_i, \mathbf{a}_i) + \sum_{i=1}^{m} \sum_{r=1}^{R_s} \ln f(\boldsymbol{\xi}_i^{(r)}; \psi_i) + \sum_{p=1}^{P} \sum_{r=1}^{R_s} \ln f(\mathbf{F}_p^{(r)}),$$

which can be accomplished by proceeding similarly to the case of the StEM algorithm. Typically, in the case in which R_s is small, the MCEM algorithm does not converge pointwise and estimates of the parameters have to be obtained by considering some summary statistic (usually the sample mean) of the values $\boldsymbol{\theta}_s$, $s = 1, 2, \ldots$, whereas for R_s large (or R_s increasing) the algorithm should converge to a local maximum of the likelihood (see for some results Fort and Moulines 2003). In particular, in the case in which the matrix $\mathbf{A}$ is known, and the unknown parameter vector to be estimated is given by $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\psi})$, the model satisfies (see Appendix B in Supplementary Material) the conditions of Fort and Moulines (2003) for the almost sure convergence of their stable MCEM algorithm. Thus, the structure of $l_c(\boldsymbol{\theta})$, when $\mathbf{A}$ is known, guarantees the almost sure convergence of the stable MCEM algorithm to the maximum of the marginal log-likelihood $l(\boldsymbol{\theta})$ based on the marginal density of the data given in (19), that is, to the maximum likelihood estimate. However, this result does not imply, in particular (weak or strong) consistency of the estimator implicitly defined by the

MCEM algorithm, to the true parameter value. On the other hand, when $\mathbf{A}$ is unknown and we have to estimate $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{A}, \boldsymbol{\psi})$, we show in Appendix C in Supplementary Material that the 'average' complete log-likelihood (22) to be maximized in the M-step of the MCEM algorithm (with $R_s > 1$) is concave with a unique local and global maximum. Although this does not guarantee convergence of the algorithm, it is of paramount importance in the computational implementation of the M-step of the MCEM algorithm. We will show in Section 4 that even in the case in which $\mathbf{A}$ is unknown, the MCEM algorithm does provide quite sensible estimates.

Once estimates of the parameters have been obtained, being the model non-identifiable, different rotations of the estimate of the matrix $\mathbf{A}$ may be considered to allow for diverse interpretations of the underlying common factors.

3.3. *Mapping.* Assuming to know, or as known (if estimated from data) the parameters of the model, in Section 3.1 we saw how predictions of the underlying factors, at any given spatial location, could be carried out by MCMC methods. However, for a given latent factor under investigation, it may be of interest to obtain not just a prediction at a given single location, but a map of predictions over a large set of spatial locations identified by the nodes of a given grid. In the case of a univariate hierarchical geostatistical model, Zhang (2002) showed how it was possible to avoid to carry out an MCMC run at each node of the grid by using a linearity property similar to linear Kriging for the predictions of the underlying Gaussian process.

Interestingly, analogous linearity properties hold also for our model and they will allow to considerably diminish the computational burden required by the mapping of the latent factors. In fact, once predictions at the K sampling sites have been obtained, maybe through an MCMC run, prediction at any other location can be provided by exploiting these MCMC predictions (and this MCMC sample), without performing any additional Monte Carlo simulation. For any given set of sampling sites $\mathbf{x}_1, \ldots, \mathbf{x}_K$, consider a prediction site $\mathbf{x}_0 \in \mathbb{R}^2$, such that $\mathbf{x}_0 \neq \mathbf{x}_k$, $k = 1, \ldots, K$. Then, as shown in Appendix D in Supplementary Material, from the conditional independence properties of our model we have that $f(Z_i(\mathbf{x}_0)|\mathbf{z}, \mathbf{y}) = f(Z_i(\mathbf{x}_0)|\mathbf{z})$, for any $i = 1, \ldots, m$, where $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_m)$, and $\mathbf{z}_i = (Z_i(\mathbf{x}_1), \ldots, Z_i(\mathbf{x}_K))^T$, $i = 1, \ldots, m$. Thus, since $Z_i(\mathbf{x})$, $i = 1, \ldots, m$, are Gaussian processes, from the linearity of the conditional expectations of the multivariate Gaussian distribution, we can write

$$\mathrm{E}[Z_i(\mathbf{x}_0)|\mathbf{z}, \mathbf{y}] = \mathrm{E}[Z_i(\mathbf{x}_0)|\mathbf{z}] = \sum_{i=1}^{m} \sum_{k=1}^{K} c_{ik}\, Z_i(\mathbf{x}_k),$$

for some appropriate coefficients c_{ik}, $i = 1, \ldots, m$, $k = 1, \ldots, K$, which depend on $\mathbf{x}_0$, and in particular

$$\mathrm{E}[Z_i(\mathbf{x}_0)|\mathbf{y}] = \mathrm{E}[\mathrm{E}[Z_i(\mathbf{x}_0)|\mathbf{z}, \mathbf{y}]|\mathbf{y}] = \sum_{i=1}^{m} \sum_{k=1}^{K} c_{ik}\ \mathrm{E}[Z_i(\mathbf{x}_k)|\mathbf{y}]. \tag{23}$$

Moreover, as far as the common factors are concerned, considering (see Appendix D in Supplementary Material) that $f(F_p(\mathbf{x}_0)|\mathbf{F}, \boldsymbol{\xi}, \mathbf{y}) = f(F_p(\mathbf{x}_0)|\mathbf{F}_p)$, for any $p = 1, \ldots, P$, and considering also that $F_p(\mathbf{x})$, $p = 1, \ldots, P$, are Gaussian processes, we can also write

$$\mathrm{E}[F_p(\mathbf{x}_0)|\mathbf{F}, \boldsymbol{\xi}, \mathbf{y}] = \mathrm{E}[F_p(\mathbf{x}_0)|\mathbf{F}_p] = \sum_{k=1}^{K} c_k\, F_p(\mathbf{x}_k),$$

for some appropriate coefficients c_k, $k = 1, \ldots, K$, depending on $\mathbf{x}_0$, and then

$$\mathrm{E}[F_p(\mathbf{x}_0)|\mathbf{y}] = \mathrm{E}[\mathrm{E}[F_p(\mathbf{x}_0)|\mathbf{F}, \boldsymbol{\xi}, \mathbf{y}]|\mathbf{y}] = \sum_{k=1}^{K} c_k\ \mathrm{E}[F_p(\mathbf{x}_k)|\mathbf{y}]. \tag{24}$$

This last expression allows us to provide predictions $\mathrm{E}[F_p(\mathbf{x}_0)|\mathbf{y}]$, at any generic location $\mathbf{x}_0$, using the predictions $\mathrm{E}[F_p(\mathbf{x}_k)|\mathbf{y}]$ at the sampling sites $\mathbf{x}_k$, $k = 1, \ldots, K$, which in turn can be numerically evaluated with a single MCMC run.

Let us note that the coefficients c_k in (24) can easily be obtained by considering the $(K+1) \times (K+1)$ variance covariance matrix of the multivariate Gaussian distribution of $(\mathbf{F}_p^T, F_p(\mathbf{x}_0))^T$. Indeed, for a given spatial autocorrelation function $\rho(\mathbf{h})$, for which we assume, for simplicity, that $\rho(\mathbf{h}) = \rho(-\mathbf{h})$, we know that

$$\mathrm{E}[F_p(\mathbf{x}_0)|\mathbf{F}_p] = \Sigma_{F_p(\mathbf{x}_0),\mathbf{F}_p} \cdot \Sigma_{\mathbf{F}_p}^{-1} \cdot \mathbf{F}_p,$$

where $\Sigma_{F_p(\mathbf{x}_0),\mathbf{F}_p}$ is the $1 \times K$ matrix $[\rho(\mathbf{x}_0 - \mathbf{x}_1) \cdots \rho(\mathbf{x}_0 - \mathbf{x}_K)]$ and $\Sigma_{\mathbf{F}_p}^{-1}$ is the inverse of the $K \times K$ matrix

$$\begin{bmatrix} 1 & \rho(\mathbf{x}_1 - \mathbf{x}_2) & \cdots & \rho(\mathbf{x}_1 - \mathbf{x}_K) \\ \rho(\mathbf{x}_2 - \mathbf{x}_1) & 1 & \cdots & \rho(\mathbf{x}_2 - \mathbf{x}_K) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(\mathbf{x}_K - \mathbf{x}_1) & \rho(\mathbf{x}_K - \mathbf{x}_2) & \cdots & 1 \end{bmatrix}.$$

In addition, to evaluate the accuracy of the predictions of a common factor $F_p(\mathbf{x})$, for some $p = 1, \ldots, P$, a map of $\mathrm{Var}[F_p(\mathbf{x}_0)|\mathbf{y}]$, for each $\mathbf{x}_0$ of a given

set of sites, can be obtained in the same fashion as before by considering that, since $f(F_p(\mathbf{x}_0)|\mathbf{F}, \boldsymbol{\xi}, \mathbf{y}) = f(F_p(\mathbf{x}_0)|\mathbf{F}_p)$, we also have that

$$\mathrm{Var}[F_p(\mathbf{x}_0)|\mathbf{F}, \boldsymbol{\xi}, \mathbf{y}] = \mathrm{Var}[F_p(\mathbf{x}_0)|\mathbf{F}_p] = 1 - \Sigma_{F_p(\mathbf{x}_0),\mathbf{F}_p} \cdot \Sigma_{\mathbf{F}_p}^{-1} \cdot \Sigma_{\mathbf{F}_p,F_p(\mathbf{x}_0)},$$

where $\Sigma_{\mathbf{F}_p,F_p(\mathbf{x}_0)}$ is the $K \times 1$ matrix $[\rho(\mathbf{x}_1 - \mathbf{x}_0) \cdots \rho(\mathbf{x}_K - \mathbf{x}_0)]^T$. So, exploiting standard properties of the variance, we can write

$$\begin{aligned}
\mathrm{Var}[F_p(\mathbf{x}_0)|\mathbf{y}] &= \mathrm{E}[\mathrm{Var}[F_p(\mathbf{x}_0)|\mathbf{F}, \boldsymbol{\xi}, \mathbf{y}]|\mathbf{y}] + \mathrm{Var}[\mathrm{E}[F_p(\mathbf{x}_0)|\mathbf{F}, \boldsymbol{\xi}, \mathbf{y}]|\mathbf{y}] \\
&= \mathrm{E}\left[1 - \Sigma_{F_p(\mathbf{x}_0),\mathbf{F}_p} \cdot \Sigma_{\mathbf{F}_p}^{-1} \cdot \Sigma_{\mathbf{F}_p,F_p(\mathbf{x}_0)} \middle| \mathbf{y}\right] + \mathrm{Var}\left[\sum_{k=1}^{K} c_k\, F_p(\mathbf{x}_k) \middle| \mathbf{y}\right] \\
&= 1 - \Sigma_{F_p(\mathbf{x}_0),\mathbf{F}_p} \cdot \Sigma_{\mathbf{F}_p}^{-1} \cdot \Sigma_{\mathbf{F}_p,F_p(\mathbf{x}_0)} + \sum_{k=1}^{K} c_k^2\, \mathrm{Var}[F_p(\mathbf{x}_k)|\mathbf{y}] \\
&\quad + \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j \neq i}}^{K} c_i c_j\, \mathrm{Cov}[F_p(\mathbf{x}_i), F_p(\mathbf{x}_j)|\mathbf{y}], \qquad (25)
\end{aligned}$$

which can be numerically evaluated by considering that $\mathrm{Var}[F_p(\mathbf{x}_k)|\mathbf{y}]$, for $k = 1, \ldots, K$, and $\mathrm{Cov}[F_p(\mathbf{x}_i), F_p(\mathbf{x}_j)|\mathbf{y}]$, for $i \neq j$, can be approximated using an MCMC sample form the conditional distribution of $\mathbf{F}_p$ given $\mathbf{y}$.

4. Simulation study. We present here some Monte Carlo simulation studies run to test the performances of the MCEM estimation algorithm and of the MCMC mapping procedure for our spatial factor model. All the computational procedures have been implemented in the R language (R Development core Team 2008). The MCMC samples needed for the S-step of the estimation algorithm and for the mapping of the latent factors have been obtained with the help of the OpenBUGS software (Lunn, Thomas, Best and Spiegelhalter 2009) using the R package R2WinBUGS which provides an easy interface between R and OpenBUGS.

In order to evaluate the sensitivity of the estimation algorithm to the initial values, we carried out some simulation experiments in which we run the MCEM algorithm starting from different sets of initial guesses for the parameter values. In Figure 3 we report the results of an experiment where we assumed a model with $m = 4$ and $P = 1$, with $\mathbf{A}$ known, and a spatial autocorrelation function $\rho(\mathbf{h}) = \exp((-\alpha||\mathbf{h}||)^{\delta})$ with $\alpha = 0.001$ and $\delta = 1.5$. Assuming $\boldsymbol{\beta} = (-2.054, 0.002, 0.954, 1.531)^T$ and $\boldsymbol{\psi} = (1.025, 0.237, 0.587, 0.027)^T$, we run the MCEM algorithm (considering 500 MCMC samples, of which 200 burn-in, at each of the 500 iterations) starting from four different sets of initial guesses for $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ and considering

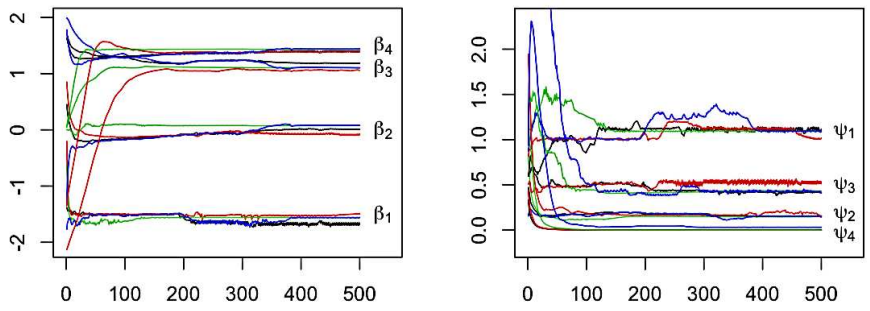

FIG 3. *Convergence diagnostic for the MCEM estimation algorithm. Four runs of the MCEM algorithm, of length 500, corresponding to four different sets of initial values, distinguished by different colours:* $\boldsymbol{\beta}_{black} = (-1.32, 0.57, 1.75, 1.67)$, $\boldsymbol{\psi}_{black} = (0.6, 0.6, 0.40.4)$*;* $\boldsymbol{\beta}_{red} = (0, 1, -2, 1)$, $\boldsymbol{\psi}_{red} = (0.5, 0.5, 0.5, 0.5)$*;* $\boldsymbol{\beta}_{green} = (0, 0, 0, 0)$, $\boldsymbol{\psi}_{green} = (1, 1, 1, 1)$*;* $\boldsymbol{\beta}_{blue} = (-2, -2, 2, 2)$, $\boldsymbol{\psi}_{blue} = (1, 0.2, 5, 0.8)$*. We considered a model with* $m = 4$ *and one common factor.*

a simulated realization from the model on an 8×8 grid (with nodes 1 km apart). As we can see, after a reasonable number of iterations the algorithm is quite insensitive to the initial guesses.

In the case in which the matrix $\mathbf{A}$ is fixed, we know that the MCEM algorithm converges to the maximum likelihood estimate, whereas in the case in which $\mathbf{A}$ is unknown we lack of theoretical convergence results. To show that the MCEM algorithm does provide sensible estimates even in the case in which $\mathbf{A}$ is unknown, we report here two Monte Carlo experiments, considering a model with $m = 4$ and $P = 1$, and a model with $m = 6$ and $P = 2$, run with the intention to simulate the sampling distribution of the estimator. For each of the two models, assuming a given set of parameter values, we simulated 50 data sets on an 8×8 grid (with nodes 1 km apart) assuming an autocorrelation function $\rho(\mathbf{h}) = \exp((-\alpha||\mathbf{h}||)^{\delta})$, where $\alpha = 0.001$ and $\delta = 1.5$. Then, for each of the two experiments, we run the MCEM estimation algorithm 50 times, one for each simulated data set. For each run of the MCEM algorithm, initial guesses for the parameter values were provided by some preliminary estimates obtained by following an estimation procedure based on the method of moments (see, for instance, Minozzo and Fruttini 2004). Also, for each run of the algorithm, we considered 500 iterations, each involving in the S-step, depending on the model, 100 MCMC samples (after 50 samples of burn-in) for the model with $m = 4$ and $P = 1$, and 200 MCMC samples (after 300 samples of burn-in) for the model with $m = 6$ and $P = 2$. The choice to keep both the MCMC sample

size and the number of MCEM iterations fixed have been made on the basis of some preliminary Monte Carlo investigations. Indeed, as far as the number of MCEM iterations is concerned, many convergence criteria and stopping rules have appeared in the literature. For instance, in order to reduce the chances of a premature stop, Booth and Hobert (1999) suggested to stop the algorithm when the inequality $\max_l[(\theta_{s+1}^{(l)} - \theta_s^{(l)})/(\theta_s^{(l)} + \epsilon_1)] < \epsilon_2$ (where $\theta_s^{(l)}$, for $l = 1, \ldots, L$, is the guess for the lth component of the parameter vector $\boldsymbol{\theta}$ after the sth iteration, and ϵ_1 and ϵ_2 are two predefined thresholds) is satisfied for three successive iterations. This criterion should reduce the chances to stop the algorithm for the closeness of two successive guesses, not due to convergence, but due to the random perturbation inherent in the stochastic sampling in the S-step of the MCEM algorithm. With the objective to overcome this problem, more than one stopping rule are often adopted. Here, to somehow minimize the computational load, we decided, on the basis of some preliminary simulations in which we implemented some of the common stopping rules present in the literature, for a fixed number of iterations (and for a fixed MCMC sample size) trying to achieve a reasonable estimation accuracy.

Assuming a model with $m = 4$ and $P = 1$ and $\boldsymbol{\beta} = (0.002, -2.054, 0.845, 1.531)^T$, $\boldsymbol{\psi} = (0.237, 1.025, 0.787, 0.027)^T$ and $\mathbf{A} = (0.854, 0.123, -0.987, 0.548)^T$, in Figure 4 we give the histograms of the simulated marginal distributions of the MCEM estimator obtained by running the algorithm with the constrain $a_{11} > 0$ in the M-step to guarantee identifiability (see for instance Anderson and Rubin 1956). Here, for each of the 50 runs, estimates are obtained by considering the arithmetic mean over the last 200 iterations. As we can see, the simulated distributions are all reasonably centered around the true parameter values (dashed lines). These distributions provide evidence that, not only the algorithm seems to converge (hopefully to the maximum likelihood estimate), but also that the algorithm seems to have reasonably nice sampling properties. Let us note that the variability of the distributions can basically be ascribed to two different sources: the sampling variability of the simulated data sets; and the Monte Carlo error of the estimation algorithm that depends on the numbers of iterations and on the size of the MCMC sample considered in the S-step.

As for the second simulation experiment with a model with $m = 6$ and $P = 2$, we assumed for $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ the parameter values given in Table 1, and $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_6)^T$, with $\mathbf{a}_1 = (1.5, 0)$, $\mathbf{a}_2 = (1.9, 1.8)$, $\mathbf{a}_3 = (-1.7, -1.1)$, $\mathbf{a}_4 = (1.8, 1.4)$, $\mathbf{a}_5 = (-1.9, 1.6)$, $\mathbf{a}_6 = (1.8, -1.4)$. To guarantee identifiability of the model, we run the MCEM algorithm with the constrains $a_{11} > 0$, $a_{12} = 0$ and $a_{22} > 0$ in the M-step. In Table 1 we supply some

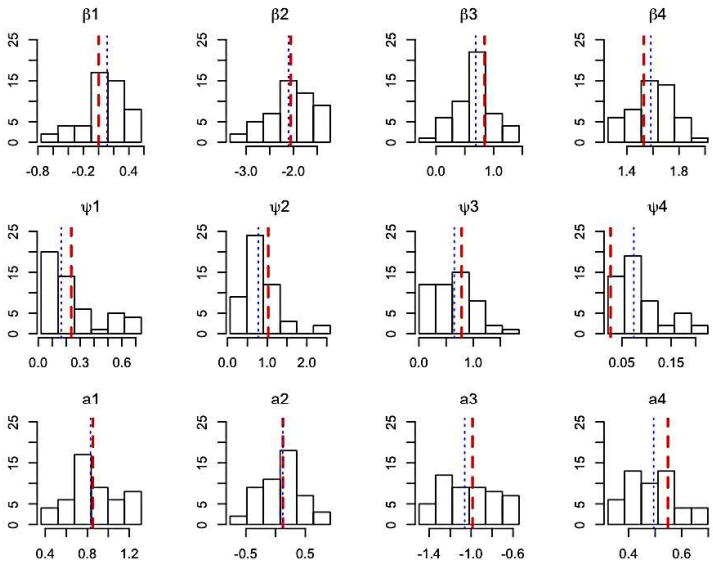

FIG 4. *Marginal distributions of the MCEM estimator. Histograms of the simulated marginal distributions of the MCEM estimator for the 12 parameters of a model with $m = 4$ and one common factor, obtained by running the algorithm over 50 simulated data sets. Dashed lines are the true parameter values; dotted lines are the empirical arithmetic means of the distributions.*

TABLE 1

Sampling statistics of the MCEM estimator. Means and percentiles of the approximated marginal distributions of the parameters obtained by running the MCEM algorithm over 50 simulated data sets from a model with $m = 6$ and two common factors.

Parameter	True value	Mean	Percentiles ($p_{0.025}$, $p_{0.975}$)
β_1	-2.054	-1.817	(-3.783, -0.638)
β_2	0.002	0.503	(-1.143, 1.848)
β_3	0.845	0.446	(-0.714, 1.517)
β_4	1.531	2.045	(1.116, 3.041)
β_5	-1.265	-1.335	(-3.426, 0.349)
β_6	1.156	1.296	(-0.209, 2.296)
ψ_1	1.025	0.710	(0.044, 2.245)
ψ_2	0.237	0.156	(0.012, 0.404)
ψ_3	0.787	0.552	(0.111, 1.095)
ψ_4	0.027	0.069	(0.005, 0.182)
ψ_5	1.581	1.183	(0.053, 3.415)
ψ_6	0.582	0.382	(0.015, 1.697)

summary statistics for the simulated marginal distributions of the estimates of the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ obtained by considering, for each of the 50 runs, the arithmetic mean over the last 50 iterations. For all twelve parameters, the true values stay inside the intervals delimited by the two percentiles. Figure 5 show instead (on the same graph) the six simulated bivariate distributions of the estimates of the 11 standardized factor loadings. Here, instead of considering the distributions of the coefficients a_{ip}, we considered, to allow for an easier interpretation, the distributions of the standardized factor loadings defined as $\lambda_{ip} = \text{Corr}[Z_i(\mathbf{x}), F_p(\mathbf{x})] = \text{Cov}[Z_i(\mathbf{x}), F_p(\mathbf{x})]/\sqrt{\text{Var}[Z_i(\mathbf{x})]} = a_{ip}/\sqrt{\varsigma_i^2}$, which are constrained to stay inside the circle of radius one centered at the origin. For our simulation experiment, the true values of these loadings corresponding to the chosen set of parameter values can be organized in the matrix $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_6)^T$, where $\boldsymbol{\lambda}_1 = (0.829, 0)$, $\boldsymbol{\lambda}_2 = (0.714, 0.676)$, $\boldsymbol{\lambda}_3 = (-0.769, -0.498)$, $\boldsymbol{\lambda}_4 = (0.787, 0.612)$, $\boldsymbol{\lambda}_5 = (-0.682, 0.575)$, $\boldsymbol{\lambda}_6 = (0.746, -0.582)$. The sampling distributions obtained show that the MCEM algorithm seems to provide reasonably accurate estimates even in the case in which $\mathbf{A}$ is unknown. (The R code for the two simulation experiments is available in *Code.zip* in Supplementary Material.)

Finally, to illustrate the prediction and mapping capabilities of our model, we report a simulation experiment in which we reconstructed a realization of the unobservable common factor $F_1(\mathbf{x})$ in a model with $m = 4$ and $P = 1$. To this aim, we first simulated, for a given set of parameter values, a realization

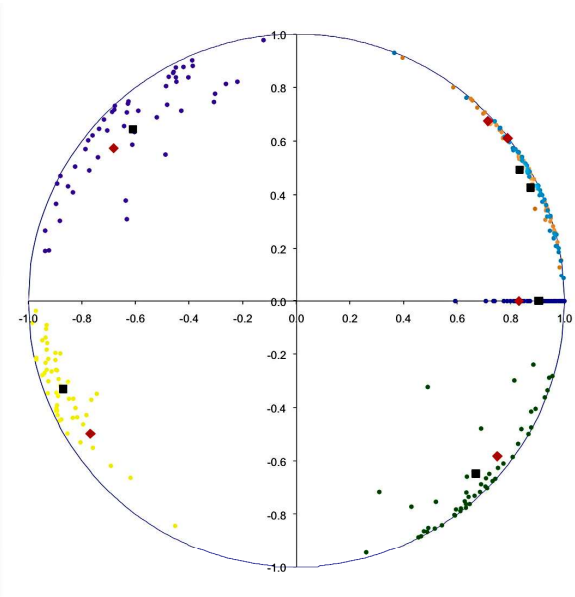

FIG 5. *Sampling distribution of the MCEM estimator of the standardized factor loadings. Monte Carlo bivariate distributions of the MCEM estimator of the 11 loadings of an identifiable model with $m = 6$ and two common factors (the six bivariate distributions are distinguished by different colours). Red rhombi correspond to the true values of the loadings (that is, to rows of the true loading matrix $\mathbf{\Lambda}$); black squares are the empirical means of the distributions obtained by running the MCEM algorithm over 50 simulated data sets from a model with $m = 6$ and two common factors. The first and the second common factors are represented by the horizontal and vertical axes respectively.*

from the model, that is, a realization of the underlying factors and of the observable data, on a 20×20 grid (with nodes 1 km apart). Then, on the basis (only) of the realized observed data, and using the true set of parameter values, we predicted the underlying common factor, that is, we simulated by MCMC the conditional expectation $\mathrm{E}[F_1(\mathbf{x}_0)|\mathbf{y}]$ given in (12), for each node $\mathbf{x}_0$ of the 20×20 grid, following the procedure outlined in Section 3.1 (the R code is available in *Code.zip* in Supplementary Material). Figure 6 shows the predictions for the common factor obtained using MCMC samples of size 500 (following 500 samples of burn-in), together with the (true) simulated realization. The comparison of the predictions (bottom panel) with the (true and unobserved) simulated realization (top panel) reveals a reasonably good agreement between the two maps, indicating the possibility of reconstructing the underlying common factor from the observed data.

5. Plankton data. Let us now consider our model for the analysis of a multivariate spatial data set which cannot be assumed to be Gaussian. The data set consists of counts of different plankton species found in Lake Trasimeno which is the largest lake on the Italian peninsula south of the Po River with a surface area of 128 km^2 (see Ludovisi, Minozzo, Pandolfi and Taticchi 2005). In order to ensure a comprehensive coverage of all plankton biotope, the lake was parceled out in 120 squares (1 km side on the WGS84-UTM grid) and 13 sampling parcels were randomly extracted. For each extracted parcel, three sampling sites were located in correspondence with the vertices of an equilateral triangle (50 m side) centered in the grid square. Overall, a total of 39 sampling sites were selected. On these sites, after water samples had been collected and using a measurement methodology based on stereomicroscopy, roughly 60 plankton species were found and classified. Here, according to the main contribution in term of biomass (volumetric amount of animal for species) observed, we concentrate on three taxa: two zooplankton species, *Cyclops* and *Diaphanosoma brachiurum*, and one phytoplankton species, *Euglena viridis*. Our interest lies in the modeling of these count data with the objective to highlight the correlation among the species via the detection of a common latent factor. From the analysis of the empirical variograms (see Figure 7) of the $m = 3$ count variables $y_1(\mathbf{x}_k)$, $y_2(\mathbf{x}_k)$ and $y_3(\mathbf{x}_k)$, for $k = 1, \ldots, K$, where $K = 39$, we fixed for the latent factors the spatial autocorrelation function $\rho(\mathbf{h}) = \exp((-\alpha||\mathbf{h}||)^\delta)$, where $\alpha = 0.0008$ and $\delta = 0.75$.

Then, assuming just one ($P = 1$) common factor for the three count variables, we proceeded to estimate the nine parameters of a model with $m = 3$ and $P = 1$ with the MCEM algorithm: namely the intercepts β_1, β_2, β_3;

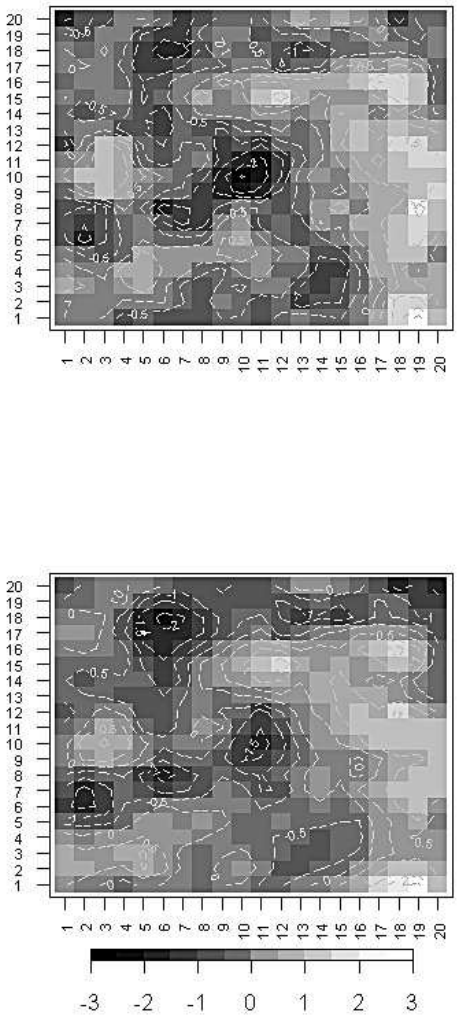

FIG 6. *Filtering of the common spatial factor. Map of predictions (bottom) of the latent common spatial factor for a model with* $m = 4$ *and one common factor obtained by MCMC simulation, assuming the parameters as known. In the top panel it is shown the true (simulated) common factor with which we generated the Poisson data on a grid of dimension* 20×20.

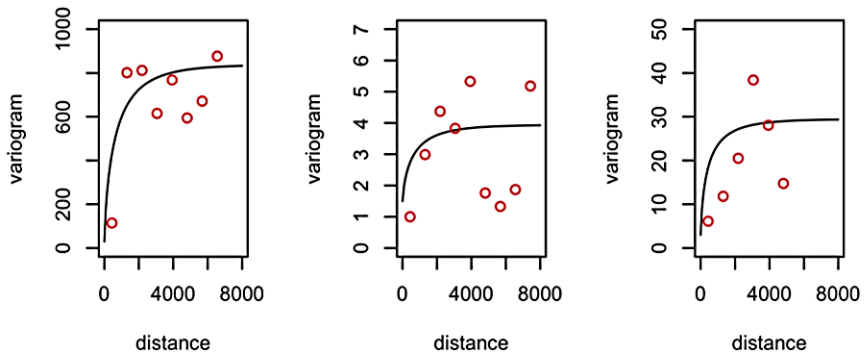

FIG 7. *Variograms for the plankton count data. Empirical (circles) and theoretical (full line) variograms $\gamma_{ii}(\mathbf{h})$ for the three (Cyclops, Diaphanosoma and Euglena) plankton count data. The theoretical variograms $\gamma_{ii}(\mathbf{h})$ are obtained by substituting the MCEM parameter estimates in the theoretical formula.*

the variances ψ_1, ψ_2, ψ_3; and the coefficients a_{11}, a_{21}, a_{31}. Differently from the simulation study, here we did not impose any constrain, to guarantee identifiability, on the coefficients a_{ip}, and considered an increasing MCMC sample size in the S-step of the algorithm. In particular, we increased the MCMC sample size following an harmonic progression in which the increments are relatively large during the first iterations and much smaller in the latter ones; actually, we started with an MCMC sample size of 80 and finished, after 350 MCEM iterations, with an MCMC sample size of 3304. As noted in Wei and Tanner (1990), it is often inefficient to start the MCEM algorithm with a large sample size in the S-step since it is likely that the maximum likelihood estimate is far from the MCEM guesses during the first iterations. These increments allow us to obtain very small fluctuations in the last iterations of the MCEM algorithm. The MCEM estimates obtained by averaging over the last iterations are given by $\boldsymbol{\beta} = (2.801, 0.030, 0.217)^T$, $\boldsymbol{\psi} = (0.474, 0.001, 0.682)^T$ and $\mathbf{A} = (-0.784, -0.980, -0.910)^T$. The standardized factor loadings $\lambda_{i1} = a_{i1}/\sqrt{\text{Var}[Z_i(\mathbf{x})]}$ representing the correlation between the underlying $Z_i(\mathbf{x})$ and the common factor $F_1(\mathbf{x})$ are given by $\boldsymbol{\Lambda} = (0.311, 0.976, 0.836)^T$. The sign and the value of these loadings suggest that the common spatial factor is responsible for most of the variability and correlation in the three plankton counts. This is quite reasonable having in mind the particular environment of the lake and how the three species are biologically interrelated. Figure 7 shows (full lines) the theoretical variograms $\gamma_{11}(\mathbf{h})$, $\gamma_{22}(\mathbf{h})$ and $\gamma_{33}(\mathbf{h})$ obtained by substituting in (9) the MCEM

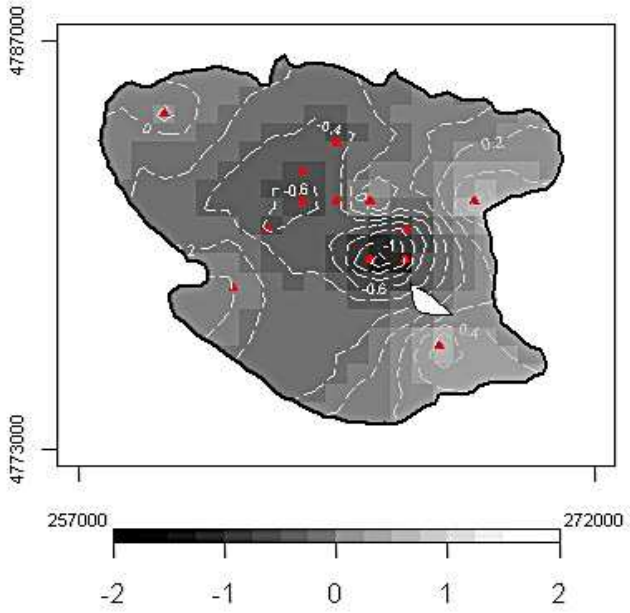

FIG 8. *Map of the common factor for the plankton data. Map of predictions of the common factor responsible for the plankton count data from Lake Trasimeno; predictions have been obtained for each node of a* 25×23 *grid covering the lake. Red triangles are the 13 sampling stations where are located the 39 sampling sites (three for each station). Contour lines are depicted in light gray.*

parameter estimates.

Assuming these estimates as the true parameter values, predictions of the latent factors can be obtained via MCMC. In Figure 8 we show the map of predictions of the common latent factor $F_1(\mathbf{x})$ obtained over a 25×23 regular grid covering the whole lake area. The spatial distribution of the common factor is coherent with previous studies (see, for instance, Ludovisi et al. 2005) and basically confirms the concentration of these plankton species in the center of the lake.

6. Conclusions. In this paper we developed a hierarchical geostatistical factor model for multivariate spatial data as an alternative to the classical and widely used geostatistical LMC which should not be used in the case of non-Gaussian, in particular count data. Basically, we extended the LMC, or better, the proportional covariance 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 count data flexible enough to account for variables with different variability. To estimate the parameters of the model we investigated a computationally intensive likelihood based procedure exploiting the capabilities of the MCEM algorithm. Our investigations would suggest both the computational feasibility of the chosen inferential procedure and its ability to provide quite reasonable estimates.

Let us finally remark that generalizations of the model presented here could be developed. For instance, to account for other kind of variable, other than count data, different distributions for the observed data and different link functions could be considered. Other generalizations could involve the introduction of more spatial scales as in the classical LMC. From a computation perspective, we last notice that, although we showed the feasibility of the MCEM estimation algorithm for a small or moderate number of sites, and for reasonably simple models, in more complex situations the computational burden could increase considerably. To solve this problem, future work might consider the use of importance sampling or the development of ad hoc techniques to diminish the computational cost.

SUPPLEMENTARY MATERIAL

Supplement A: Appendices
(http://lib.stat.cmu.edu/aoas/???/???; .pdf). This document contains: Appendix A with algebraic details about the correlation structure of the model and derivations of the variograms and cross-variograms; Appendix B with the discussion of the convergence of the MCEM algorithm when the loadings are known; Appendix C with the discussion of the maximization step of the MCEM algorithm when the loadings are unknown; Appendix D with the derivation of the conditional distribution of the latent factors.

Supplement B: Code
(http://lib.stat.cmu.edu/aoas/???/???; .zip). This zip file contains the scripts of the estimation procedure based on the MCEM algorithm for the models with $m = 4$ and $P = 1$, and with $m = 6$ and $P = 2$, and the script for obtaining the prediction map of the common factor for the model with $m = 4$ and $P = 1$.

References.

[1] Anderson, T. W., and Rubin, H. (1956), "Statistical Inference in Factor Analysis," *Third Berkeley Symp. Math. Statist. and Prob.*, 5, 111–150.

[2] Bailey, T. C., and Krzanowski, W. J. (2000), "Extensions to Spatial Factor Methods With an Illustration in Geochemistry," *Mathematical Geology*, 32, 657–682.

[3] Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, London: Chapman & Hall/CRC.

[4] Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008), "Gaussian Predictive Process Models for Large Spatial Data Sets," *Journal of the Royal Statistical Society*, Ser. B, 70, 825–848.

[5] Booth, J. G., and Hobert, J. P. (1999), "Maximizing Generalized Linear Mixed Model Likelihoods With an Automated Monte Carlo EM Algorithm," *Journal of the Royal Statistical Society*, Ser. B, 61, 165–285.

[6] Calder, C. A. (2007), "Dynamic Factor Process Convolutio Models for Multivariate Space-Time Data With Application to Air Quality Assessment," *Environmental and Ecological Statistics*, 14, 229–247.

[7] Christensen, O. F. (2004), "Monte Carlo maximum likelihood in model-based geostatistics," *Journal of Computational and Graphical Statistics*, 13, 702–718.

[8] Christensen, W., and Amemiya, Y. (2002), "Latent Variable Analysis of Multivariate Spatial Data," *Journal of the American Statistical Association*, 97, 302–317.

[9] Cressie, N. (2001), "Geostatistical Methods for Mapping Environmental Exposures," in *Spatial Epidemiology: Methods and Applications*, eds. P. Elliott et al., Oxford: Oxford University Press, pp. 185–205.

[10] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood From Incomplete Data Via the EM Algorithm," *Journal of the Royal Statistical Society*, Ser. B, 39, 1–38.

[11] Diggle, P. J., Moyeed, R. A., and Tawn, J. A. (1998), "Model-Based Geostatistics" (with discussion), *Applied Statistics*, 47, 299–350.

[12] Diggle, P. J., and Ribeiro, P. J. (2007), *Model-Based Geostatistics*, Springer.

[13] Fort, G., and Moulines, E. (2003), "Convergence of the Monte Carlo Expectation Maximization for Curved Exponential Families," *The Annals of Statistics*, 31, 1220–1259.

[14] Gelfand, A., Schmidt, A., Banerjee, S., and Sirmans, C. (2004), "Nonstationary Multivariate Process Modeling through Spatially Varying Coregionalization," *Test*, 13, 263–312.

[15] Grzebyk, M., and Wackernagel, H. (1994), "Multivariate Analysis and Spatial/Temporal Scales: Real and Complex Models," in *Proceedings of the XVII International Biometrics Conference*, Hamilton, Ontario, pp. 19–33.

[16] Kuhn, E., and Lavielle, M. (2004), "Coupling a Stochastic Approximation Version of EM With an MCMC Procedure," *ESAIM: Probability and Statistics*, 8, 115–131.

[17] Lavielle, M., and Moulines, E. (1997), "A Simulated Annealing Version of the EM Algorithm for Non-Gaussian Deconvolutions," *Statistics and Computing*, 7, 229–236.

[18] Ludovisi, A., Minozzo, M., Pandolfi, P., and Taticchi, M. I. (2005), "Modelling the Horizontal Spatial Structure of Planktonic Community in Lake Trasimeno (Umbria, Italy) Using Multivariate Geostatistical Methods," *Ecological Modelling*, 181, 247–262.

[19] Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N. (2009), "The BUGS Project: Evolution, Critique and Future Directions" (with discussion), *Statistics in Medicine*, 28, 3049–3082.

[20] Matheron, G. (1982), "Pour une Analyse Krigeante des Données Régionalisées," Publication N-732, Centre de Géostatistique, Fontainebleau, France.

[21] McLachlan, G. J., and Krishnan, T. (2008), *The EM Algorithm and Extensions* (2nd ed.), New York: Wiley.

[22] Minozzo, M., and Fruttini, D. (2004), "Loglinear Spatial Factor Analysis: An Application to Diabetes Mellitus Complications," *Environmetrics*, 15, 423-434.

[23] Nielsen, S. F. (2000), "The Stochastic EM Algorithm: Estimation and Asymptotic

Results," *Bernoulli*, 6, 457–489.

[24] R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*, Wien: R Foundation for Statistical Computing.

[25] Reich, B. J., and Fuentes, M. (2007), "A Multivariate Semiparametric Bayesian Spatial Modeling Framework for Hurricane Surface Wind Fields," *The Annals of Applied Statistics*, 1, 249–264.

[26] Schmidt, A. M., and Gelfand, A. E. (2003), "A Bayesian Coregionalization Approach for Multivariate Pollutant Data," *Journal of Geophysical Research - Atmosphere*, 108, D24, 8783.

[27] Smith, A. F. M., and Roberts, G. O. (1993), "Bayesian Computation via Gibbs Sampler and Related Markov Chain Monte Carlo Methods (with discussion)," *Journal of the Royal Statistical Society*, Ser. B., 55, 3–23.

[28] Wackernagel, H. (1998), *Multivariate Geostatistics*, Berlin: Springer-Verlag.

[29] Wang, F., and Wall, M. M. (2003), "Generalized Common Spatial Factor Model," *Biostatistics*, 4, 569–582.

[30] Wei, G. C. G., and Tanner, M. A. (1990), "A Monte Carlo Implementation of the EM Algorithm and Poor Man's Data Augmentation Algoritm," *Journal of the American Statistical Association*, 85, 699–704.

[31] Zhang, H. (2002), "On Estimation and Prediction for Spatial Generalized Linear Mixed Models," *Biometrics*, 58, 129–136.

[32] Zhang, H. (2007), "Maximum-Likelihood Estimation for Multivariate Spatial Linear Coregionalization Models," *Environmetrics*, 18, 125–139.

[33] Zhu, J., Eickhoff, J. C., and Yan, P. (2005), "Generalized Linear Latent Variable Models for Repeated Measures of Spatially Correlated Multivariate Data," *Biometrics*, 61, 674–683.

Department of Economics
University of Verona
Via dell'Artigliere 19
37129 Verona, Italy
E-mail: marco.minozzo@univr.it
clarissa.ferrari@univr.it
URL: http://www.univr.it