A Latent Variable Approach to Modelling Multivariate Geostatistical Skew-Normal Data

Luca Bagnato and Marco Minozzo

Abstract

In this paper we propose a spatial latent factor model to deal with multivariate geostatistical skew-normal data. In this model we assume that the unobserved latent structure, responsible for the correlation among different variables as well as for the spatial autocorrelation among different sites is Gaussian, and that the observed variables are skew-normal. For this model we provide some of its properties like its spatial autocorrelation structure and its finite dimensional marginal distributions. Estimation of the unknown parameters of the model is carried out by employing a Monte Carlo Expectation Maximization algorithm, whereas prediction at unobserved sites is performed by using closed form formulas and Markov chain Monte Carlo algorithms. Simulation studies have been performed to evaluate the soundness of the proposed procedures.

L. Bagnato (✉)
Università Cattolica del Sacro Cuore, Milano, Italy
e-mail: luca.bagnato@unicatt.it

M. Minozzo (✉)
Dipartimento di Scienze Economiche, Università degli Studi di Verona, Via dell'Artigliere 19, 37129 Verona, Italy
e-mail: marco.minozzo@univr.it

1   Introduction

Although a large variety of spatial data sets (on radioactive contamination, rainfalls, winds, etc.) contain measurements with a considerable amount of skewness, its modelling still remains an issue. For instance, with regard to radiological monitoring, in [5], disregarding any physically-based modelling approach, it is argued on the necessity of developing mapping algorithms for emergency detection taking into consideration the skewness in the data. A boost to these developments came from the Spatial Interpolation Comparison (SIC) 2004 (see [11]) in which, whereas the routine scenario could easily be modelled using a Gaussian random field, the emergency scenario, which mimics an accidental release of radioactivity, needs to be modelled taking properly into account that, due to the presence of extreme measurements, the data are positively skewed. Just to cite a few works, to deal with skewed measurements coming from radioactive monitoring, [18] and [15] propose copula-based geostatistical approaches, whereas [9] argues that the structuring of extreme values can be faced in a coherent manner by using the class of Hermitian isofactorial models. Moreover, [4] proposes a Gaussian anamorphosis transformation to deal with skewed data coming from contaminated facilities, and [19] argues in favor of a Bayesian approach pointing out that both the Gaussian copula and the non-Gaussian χ^2-copula models are inappropriate to model strongly skewed radioactivity measurements. Other works dealing with skewed radiological measurements are [27], which is concerned with the estimation of the variogram and the development of optimal sampling plans, [7], which proposes a dynamic spatial Bayesian model for non-Gaussian measurements from radioactivity deposition, as well as the works in [22, 28] and [32]. On the other hand, a general approach developed to cope with some types of univariate non-Gaussian spatial data (including skew data) has been proposed in [8] by defining a family of transformed Gaussian random fields that provides an alternative to trans-Gaussian kriging.

Whereas in the univariate case, that is, in presence of just one regionalized variable, spatial modelling and prediction have been extensively studied for different types of non-Gaussian data, in particular skew data, in a multivariate non-Gaussian context only a limited number of works have been published. Among these, [26] and [25] extend to multivariate geostatistical non-Gaussian data the modelling approach of [10], whereas [6] proposes a hierarchical Bayesian approach to model Gaussian, count, and ordinal variables, by designing a Gibbs sampler with Metropolis-Hastings steps. Other works dealing with multivariate spatial data are those in [33], which explores the use of the Bayesian Maximum Entropy approach in presence of both continuous and categorical regionalized variables, and in [31], which uses Markov chain Monte Carlo methods for the Bayesian modelling of multivariate counts.

In this paper, to model skewness in a multivariate (that is, in presence of more than one regionalized variable) geostatistical context, we propose an alternative approach based on the use of the skew-normal distribution. Our modelling approach, which extends some of the ideas in [24] (see also [35]), is based on the skew-normal distribution [2, 3] and on a latent Gaussian factor structure. Just to give some examples, this approach might prove useful in the modelling of the radiological data in [16] or the data related to the Fukushima disaster (data are available from TEPCO at http://www.tepco.co.jp) where more than one radiological measurement has been collected for each sampling site. Apart from providing a much greater flexibility with respect to the traditional Gaussian random fields, it is possible to show that our model has all its finite-dimensional marginal distributions belonging to the family of the closed skew-normal distribution [13, 14]. It must be mentioned that the modelling construction proposed here is substantially different from some of the most popular constructions based on the skew-normal distribution that have recently appeared in the literature to model univariate skewed spatial data, like those, for instance, of [1, 20] and [17] (for a critical discussion on these constructions see [24]).

The paper is organized as follows. The model and its properties are presented in Sect. 2 and in Sect. 3, respectively. In Sect. 4 we present the estimation and prediction procedures and some simulation results, and in section "Conclusions" we make some final comments. More technical results are presented in the Appendix.

2   A Multivariate Closed Skew-Normal Geostatistical Model

In the following we define a model for geostatistical multivariate skewed data exploiting the ideas in [24] and in [25], by building the model on an unobserved latent Gaussian spatial factor structure. 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 $\mathbf{x}_k$. Each of these m measured variables can be viewed as a partial realization of a particular stochastic process $Y_i(\mathbf{x})$, $i = 1, \ldots, m$, $\mathbf{x} \in \mathbb{R}^2$. We assume that these stochastic processes are given by

$$Y_i(\mathbf{x}) = \beta_i + Z_i(\mathbf{x}) + \omega_i S_i(\mathbf{x}), \quad i = 1, \ldots, m, \tag{1}$$

where β_i and ω_i are unknown constants, representing, respectively, an intercept and a scale parameter, and $Z_i(\mathbf{x})$ and $S_i(\mathbf{x})$ are latent processes. In particular, for every $i = 1, \ldots, m$, $Z_i(\mathbf{x})$ is a mean zero stationary Gaussian process, whereas for every $i = 1, \ldots, m$, and for each $\mathbf{x} \in \mathbb{R}^2$, $S_i(\mathbf{x})$ is an independent random variable distributed as a skew-normal [2], that is, $S_i(\mathbf{x}) \sim SN(0, 1, \alpha_i)$, which means that, for every $\mathbf{x} \in \mathbb{R}^2$, the density of $S_i(\mathbf{x})$ is given by $f_{S_i}(s) = 2\phi_1(s; 1)\Phi(\alpha_i s)$, for $-\infty < s < \infty$, where $\alpha_i \in \mathbb{R}$, $\phi_1(\cdot; 1)$ is the scalar normal density function with zero mean and unit variance, and $\Phi(\cdot)$ is the scalar N(0, 1) distribution function.

Let us note that, for each $i = 1, \ldots, m$, and for every $\mathbf{x} \in \mathbb{R}^2$, conditionally on $Z_i(\mathbf{x})$, the random variable $Y_i(\mathbf{x})$ has a skew-normal distribution, that is,

$$Y_i(\mathbf{x}) \,|\, Z_i(\mathbf{x}) \sim SN\left(\beta_i + Z_i(\mathbf{x}), \omega_i^2, \alpha_i\right), \tag{2}$$

which means that we can write its density as

$$f(y_i(\mathbf{x})|z_i(\mathbf{x})) = 2\,\phi_1\left(y_i(\mathbf{x}) - \beta_i - z_i(\mathbf{x}); \omega_i^2\right) \Phi\left(\frac{\alpha_i}{\omega_i}\left(y_i(\mathbf{x}) - \beta_i - z_i(\mathbf{x})\right)\right),$$

where $\phi_1(\,\cdot\,; \sigma^2)$ is the scalar normal density function with zero mean and positive variance σ^2. Moreover, for each $i = 1, \ldots, m$, and for every $\mathbf{x} \in \mathbb{R}^2$, the (scalar) random variable $Y_i(\mathbf{x})$ has a (marginal) skew-normal distribution, that is,

$$Y_i(\mathbf{x}) \sim SN\left(\beta_i, \varsigma_i^2 + \omega_i^2, \alpha_i \omega_i \,\bigg/ \sqrt{\varsigma_i^2\left(1 + \alpha_i^2\right) + \omega_i^2}\right), \tag{3}$$

where $\varsigma_i^2 = \mathrm{Var}\left[Z_i(\mathbf{x})\right]$.

A similar result holds also for the other marginal distributions of the process. Indeed, with some algebra it is possible to show that all finite dimensional marginal distributions of the (weakly and strongly stationary) multivariate spatial process $(Y_1(\mathbf{x}), \ldots, Y_m(\mathbf{x}))^T$, for $\mathbf{x} \in \mathbb{R}^2$, are closed skew-normal (CSN). This implies, for instance, that, for each $i = 1, \ldots, m$, the univariate spatial process $Y_i(\mathbf{x})$, for $\mathbf{x} \in \mathbb{R}^2$, has all its finite-dimensional marginal distributions belonging to the CSN family (see the Appendix), and that, for any fixed spatial location $\mathbf{x} \in \mathbb{R}^2$, the random vector $(Y_1(\mathbf{x}), \ldots, Y_m(\mathbf{x}))^T$ has a multivariate CSN distribution [13, 14]. In principle, these results make the approach very appealing since they allow, due to the stationarity of the processes, to empirically check some of the distributional properties of the model. For instance, for a given set of observations, the empirical distribution of $y_i(\mathbf{x}_k)$, $k = 1, \ldots, K$, for any given $i = 1, \ldots, m$, can be compared with the marginal skew-normal distribution in (3).

For the latent part of the model, that is, for the stationary Gaussian processes $Z_i(\mathbf{x})$, $i = 1, \ldots, m$, we assume that

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

where a_{ip} are $m \times P$ real coefficients, and $F_p(\mathbf{x})$, $p = 1, \ldots, P$, are $P \leq m$ non-observable spatial processes (*common factors*) responsible for the cross-correlations in the model. The processes $F_p(\mathbf{x})$, $p = 1, \ldots, P$, are assumed zero mean, stationary, and Gaussian with covariance function

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

where $\mathbf{h} \in \mathbb{R}^2$ and $\rho(\mathbf{h})$ is a real spatial autocorrelation function common to all factors with $\rho(\mathbf{0}) = 1$ and $\rho(\mathbf{h}) \to 0$, as $\|\mathbf{h}\| \to \infty$. Similarly to the classical linear factor model, this latent linear structure is responsible for a specific correlation structure among the processes $Z_i(\mathbf{x})$. In particular, for each $i = 1, \dots, m$, the covariance functions are given by $\mathrm{Cov}\left[Z_i(\mathbf{x}), Z_i(\mathbf{x}+\mathbf{h})\right] = \sum_{p=1}^{P} a_{ip}^2 \rho(\mathbf{h})$, whereas the cross-covariance functions are given by $\mathrm{Cov}\left[Z_i(\mathbf{x}), Z_j(\mathbf{x}+\mathbf{h})\right] = \sum_{p=1}^{P} a_{ip} a_{jp} \rho(\mathbf{h})$. Taking $\mathbf{h} = \mathbf{0}$, we find that $\mathrm{Var}\left[Z_i(\mathbf{x})\right] = \sum_{p=1}^{P} a_{ip}^2$ and $\mathrm{Cov}\left[Z_i(\mathbf{x}), Z_j(\mathbf{x})\right] = \sum_{p=1}^{P} a_{ip} a_{jp}$.

3   Variograms and Cross-Variograms

Let us consider here the correlation structure of the observable processes, induced by the latent factor model. For the observable stochastic processes $Y_i(\mathbf{x}), i = 1, \dots, m$, we can show that

$$\mathrm{E}\left[Y_i(\mathbf{x})\right] = \beta_i + \omega_i \delta_i \left(\frac{2}{\pi}\right)^{\frac{1}{2}}, \qquad \mathrm{Var}\left[Y_i(\mathbf{x})\right] = \varsigma_i^2 + \omega_i^2 \left[1 - \frac{2}{\pi}\delta_i^2\right],$$

where $\delta_i = \alpha_i / \sqrt{1 + \alpha_i^2}$, and, for $\mathbf{h} \neq \mathbf{0}$,

$$C_{ii}(\mathbf{h}) = \mathrm{Cov}\left[Y_i(\mathbf{x}), Y_i(\mathbf{x}+\mathbf{h})\right] = \varsigma_i^2 \rho(\mathbf{h}). \tag{5}$$

Note that if $\rho(\mathbf{h}) = \rho(-\mathbf{h})$ we have that $C_{ii}(\mathbf{h}) = C_{ii}(-\mathbf{h})$. Furthermore, $C_{ii}(\infty) = 0$ and $C_{ii}(\mathbf{0}) \neq C_{ii}(\mathbf{0}^+) = \varsigma_i^2$, that is, the covariance function $C_{ii}(\mathbf{h})$ is discontinuous at the origin.

On the other hand, for $\mathbf{h} \neq \mathbf{0}$, the variogram of the observable $Y_i(\mathbf{x})$ takes the form

$$\gamma_{ii}(\mathbf{h}) = \frac{1}{2}\mathrm{Var}\left[Y_i(\mathbf{x}+\mathbf{h}) - Y_i(\mathbf{x})\right] = \omega_i^2 \left[1 - \frac{2}{\pi}\delta_i^2\right] + \varsigma_i^2 \left[1 - \rho(\mathbf{h})\right], \tag{6}$$

which is, similarly to the covariance function, discontinuous in zero. In fact, we have that $\gamma_{ii}(\mathbf{0}) = 0$ and $\gamma_{ii}(\mathbf{0}^+) = \omega_i^2[1 - (2/\pi)\delta_i^2]$. Note that $\gamma_{ii}(\infty) = C_{ii}(\mathbf{0})$. To visually asses Formula (6), Fig. 1 shows the form taken by the variogram $\gamma_{ii}(\mathbf{h})$
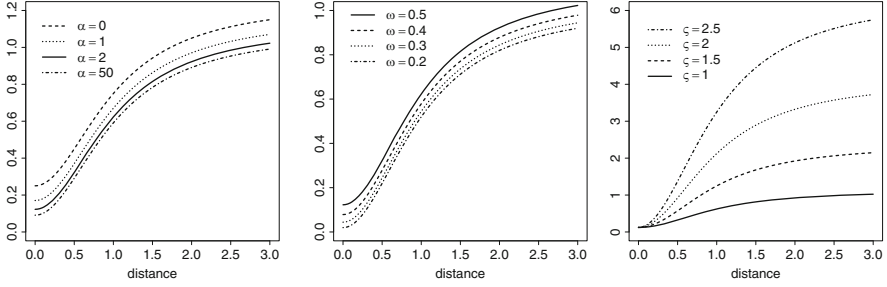
Fig. 1 The graphs show the shape of the theoretical variogram $\gamma_{ii}(\mathbf{h})$ given in Formula (6), for a Cauchy autocorrelation function with both parameters equal to 1, and for different values of the other parameters: (*left*) $\omega = 0.5$, $\varsigma = 1$; (*middle*) $\alpha = 2$, $\varsigma = 1$; (*right*) $\alpha = 2$, $\omega = 0.5$. The *solid line* in the three graphs corresponds to the same set of parameter values. The *line* in the first graph corresponding to $\alpha = 0$ gives the variogram in the case of a Gaussian process

for different values of the parameters, in the case of a Cauchy spatial autocorrelation function $\rho(\mathbf{h}) = \left[1 + (\|\mathbf{h}\|/\gamma)^2\right]^{-\eta}$, with $\gamma = 1$ and $\eta = 1$. As we can see, the nugget of the variogram decreases for decreasing values of ω and for values of the skewness parameter α departing from zero.

For any two stochastic processes $Y_i(\mathbf{x})$ and $Y_j(\mathbf{x})$, with $i \neq j$, it is easy to show that

$$C_{ij}(\mathbf{h}) = \mathrm{Cov}\left[Y_i(\mathbf{x}), Y_j(\mathbf{x}+\mathbf{h})\right] = \mathrm{Cov}\left[Z_i(\mathbf{x}), Z_j(\mathbf{x}+\mathbf{h})\right] = \varsigma_{ij}\rho(\mathbf{h}), \quad (7)$$

where $\varsigma_{ij} = \sum_{p=1}^{P} a_{ip}a_{jp} = \mathrm{Cov}\left[Z_i(\mathbf{x}), Z_j(\mathbf{x})\right]$. Note that $C_{ij}(\mathbf{h}) = C_{ji}(\mathbf{h})$ and that if $\rho(\mathbf{h}) = \rho(-\mathbf{h})$, then $C_{ij}(\mathbf{h}) = C_{ij}(-\mathbf{h})$.

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

$$\gamma_{ij}(\mathbf{h}) = \frac{1}{2}\mathrm{Cov}\left[Y_i(\mathbf{x}+\mathbf{h}) - Y_i(\mathbf{x}), Y_j(\mathbf{x}+\mathbf{h}) - Y_j(\mathbf{x})\right] = \varsigma_{ij}\left[1 - \rho(\mathbf{h})\right].$$
$$(8)$$

4 Estimation and Prediction

Assuming to know the number P of common factors and the spatial autocorrelation function $\rho(\mathbf{h})$, the model depends on the parameter vector $\boldsymbol{\vartheta}^* = (\boldsymbol{\beta}, \mathbf{A}, \boldsymbol{\omega}, \boldsymbol{\alpha})$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)^T$, $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_m)^T$ with $\mathbf{a}_i = (a_{i1}, \ldots, a_{iP})^T$, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m)^T$, and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)^T$. Note that, similarly to the classical factor model, our model is not identifiable. Indeed, there are two groups of orthogonal transformations of the matrix $\mathbf{A}$, given by permutation matrices and by some special reflection matrices, that leave the model unchanged [30]. However, this is the only indeterminacy in the model and can easily be faced.

In the following, we will further assume to know the parameters $\boldsymbol{\omega}$ and $\boldsymbol{\alpha}$. In this case, by resorting to Markov chain Monte Carlo (MCMC), and in particular to the Metropolis-Hasting algorithm, a likelihood based estimation procedure for the parameter $\boldsymbol{\vartheta} = (\boldsymbol{\beta}, \mathbf{A})$ can be developed by exploiting the Monte Carlo Expectation Maximization (MCEM) algorithm. Let $\mathbf{F} = (\mathbf{F}_1, \ldots, \mathbf{F}_P)^T$, where $\mathbf{F}_p = (F_p(\mathbf{x}_1), \ldots, F_p(\mathbf{x}_K))^T$, $p = 1, \ldots, P$, and let $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_m)^T$, where $\mathbf{y}_i = (y_i(\mathbf{x}_1), \ldots, y_i(\mathbf{x}_K))^T$, $i = 1, \ldots, m$. Whereas the marginal log-likelihood $l(\boldsymbol{\vartheta}) = \ln f(\mathbf{y}; \boldsymbol{\vartheta})$ is not available due to the presence of multidimensional integrals in the derivation of the marginal density $f(\mathbf{y}; \boldsymbol{\vartheta})$, the complete log-likelihood based on the joint distribution $f(\mathbf{y}, \mathbf{F}; \boldsymbol{\vartheta})$ is easily given by

$$
\begin{aligned}
l_c(\boldsymbol{\vartheta}) &= \ln f(\mathbf{y}, \mathbf{F}; \boldsymbol{\vartheta}) = \ln \left(f(\mathbf{y}|\mathbf{F}; \boldsymbol{\vartheta}) \cdot f(\mathbf{F}) \right) \\
&= \ln \left\{ \left(\prod_{i=1}^{m} \prod_{k=1}^{K} f(y_{ik}; Z_{ik}, \beta_i) \right) \cdot f(\mathbf{F}) \right\} \\
&= \ln \left\{ \left(\prod_{i=1}^{m} \prod_{k=1}^{K} 2 \, \phi_1\left(y_{ik} - \beta_i - Z_{ik}; \omega_i^2\right) \Phi\left(\frac{\alpha_i}{\omega_i}(y_{ik} - \beta_i - Z_{ik})\right) \right) \cdot \left(\prod_{p=1}^{P} f(\mathbf{F}_p) \right) \right\},
\end{aligned}
$$
(9)

where $y_{ik} = y_i(\mathbf{x}_k)$ and $Z_{ik} = Z_i(\mathbf{x}_k)$. In this situation, the marginal log-likelihood $l(\boldsymbol{\vartheta}) = \ln f(\mathbf{y}; \boldsymbol{\vartheta})$ can be maximized by resorting to the Monte Carlo Expectation Maximization (MCEM) algorithm (see, for instance, [23] and [12]).

At the sth iteration, the MCEM algorithm involves three steps: S-step, E-step and M-step. In the first step (S-step), R_s samples $\mathbf{F}^{(r)}$, $r = 1, \ldots, R_s$, are drawn from the (filtered) conditional distribution $f(\mathbf{F}|\mathbf{y}; \boldsymbol{\vartheta}_{s-1})$, where $\boldsymbol{\vartheta}_{s-1}$ is the guess of the parameter $\boldsymbol{\vartheta}$ after the $(s-1)$th iteration. These samples can be collected by using some Markov chain Monte Carlo (MCMC) procedure based on the Metropolis-Hustings algorithm. In the second step (E-step) the following approximation of the conditional expectation of the complete log-likelihood is computed

$$
Q_s(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_{s-1}) = \hat{\mathrm{E}}\left[\ln f(\mathbf{y}, \mathbf{F}; \boldsymbol{\vartheta})|\mathbf{y}\right] = \frac{1}{R_s} \sum_{r=1}^{R_s} \ln f\left(\mathbf{y}, \mathbf{F}^{(r)}; \boldsymbol{\vartheta}\right).
$$

The last step (M-step) supplies as the new guess $\boldsymbol{\vartheta}_s$ the value of $\boldsymbol{\vartheta}$ which maximizes $Q_s(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_{s-1})$.

Although convergence results for this algorithm are not available, it is nevertheless possible to show that the "average" complete likelihood which is maximized in the M-step of the MCEM algorithm is concave and admits a unique local (and global) maximum. This result allows to safely implement standard numerical maximization techniques.

Assuming as known all parameters of the model, prediction of the observable processes $Y_i(\mathbf{x})$ at an unobserved spatial location (or at an unobserved set of spatial locations) can be carried out either by exploiting some of the properties of the CSN distribution, or by implementing some MCMC algorithm. On the other hand, for the prediction of the unobserved common factors $F_p(\mathbf{x})$, we need to resort to MCMC algorithms. In the case in which we are interested in predicting a common factor on a large set of spatial locations (maybe on a grid), instead of carring out an MCMC run at each spatial location, we can carry out an MCMC run only at the sampling points (that is, only at those points for which we gathered observations), and then exploit a linear property similar to Kriging, and also similar to that found by [34] in a univariate framework, to obtain predictions at all other spatial locations.

To asses the goodness of the MCEM estimation procedure we performed some simulation studies. To give some examples, in Fig. 2 we show the results of some simulation analyses. For these analyses we considered $m = 2$ and $P = 1$, that is, two observable variables and one latent common factor $F(\mathbf{x})$. In the first two simulation experiments we considered a powered exponential (stable) spatial autocorrelation function $\rho(\mathbf{h}) = \exp\left[-(\gamma \|\mathbf{h}\|)^\eta\right]$, with $\gamma = 10^{-5}$ and $\eta = 1.5$, whereas in the last two experiments we considered a Cauchy autocorrelation function with $\gamma = 7{,}000$ and $\eta = 1$. For any given set of parameter values $\boldsymbol{\vartheta}^*$ and a given spatial autocorrelation function $\rho(\mathbf{h})$, we simulated 50 realizations from the model over $K = 25$ equally spaced fixed sampling points located on the nodes of a grid. For each simulated realization, we run the MCEM estimation algorithm, assuming as unknown only the parameters a_{11}, a_{21}, β_1 and β_2. Each time, we considered 800 iterations of the MCEM algorithm, and at each step of the algorithm we considered 800 MCMC samples (of which 400 burn-in). As shown in Fig. 2, despite some possible distortion (which could be due to the modest sample size), the sampling distributions look quite reasonable. However, though our simulation experiments gave us reassuring results, we feel that more efforts should be made to fully investigate the theoretical inferential properties of the proposed inferential procedure.

As far as the computational load of our estimation procedure is concerned, implementing our algorithm with the help of the OpenBUGS software [21] using the package R2WinBUGS in R [29], and using standard commercial personal computers, the computing times are still demanding. Just to give an example, with 25 observations on a grid simulated assuming the powered exponential autocorrelation function and the value of the parameters used to obtain the simulated distributions in the second row of Fig. 2, one iteration of the MCEM algorithm (with an MCMC sample size of 800) took 41 s. Increasing the size of the grid to 49 observations, the computing time increases to 102 s. Let us note that much of the time is needed for the maximization step of the MCEM algorithm. In the former case, the time needed to generate the MCMC sample was less than 1 second, whereas the time needed by the maximization step was 40 s. Thus, to obtain one MCEM estimate, using 800 iterations of the MCEM, takes more than 9 h, and to obtain a simulated distribution, based on 50 replicates, of the MCEM estimator (that is, one row of Fig. 2) takes several days.
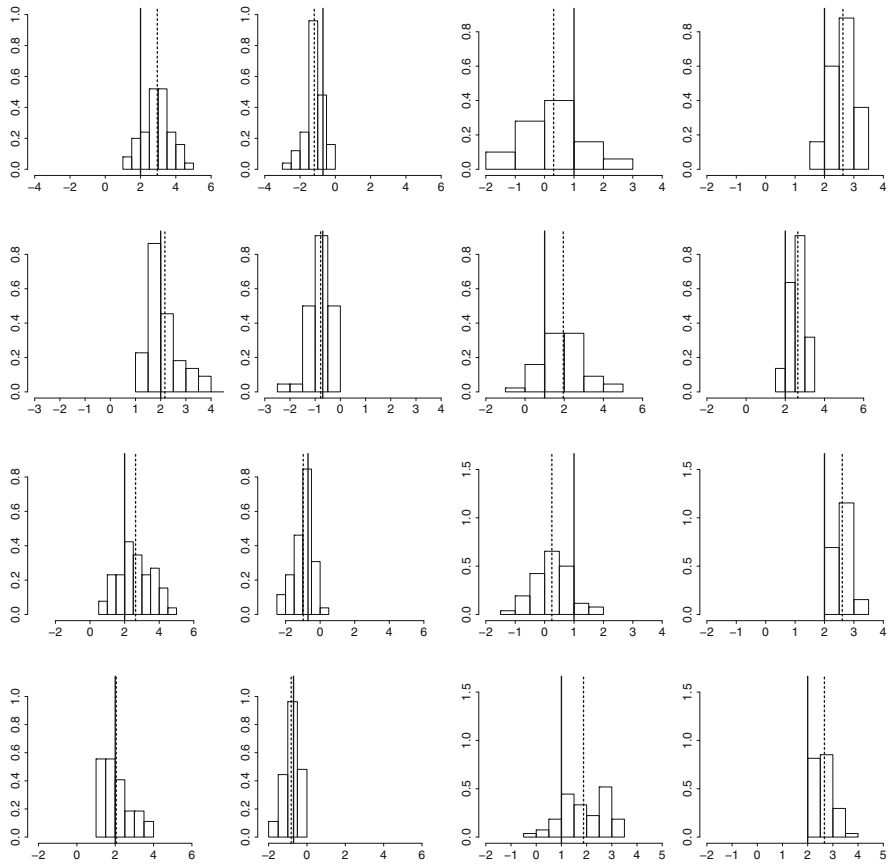
Fig. 2 The histograms show the simulated univariate marginal sampling distributions of the MCEM estimator of the parameters a_{11}, a_{21}, β_1 and β_2 (*from left to right*) in a model with $m = 2$ and $P = 1$ obtained in four simulation experiments (*from top to bottom*). The *vertical solid lines* represent the true parameter values, whereas the *vertical dashed lines* represent the empirical means over the 50 simulated realizations. For the spatial autocorrelation function $\rho(\mathbf{h})$ we chose a powered exponential model with $\gamma = 0.00001$ and $\eta = 1.5$ in the first two simulation experiments (*first two rows*), and a Cauchy model with $\gamma = 7,000$ and $\eta = 1$ in the last two simulation experiments (*last two rows*). The parameters α_1 and α_2 were fixed equal to: -1 and 1 (*first row*); 2 and 2 (*second row*); -1 and 1 (*third row*); 2 and 2 (*fourth row*). For all four simulation experiments, the other parameters where equal to: $a_{11} = 2$, $a_{21} = -0.7$, $\beta_1 = 1$, $\beta_2 = 2$, $\omega_1 = 1$, $\omega_2 = 1$

Conclusion

In this work we have proposed and studied a model for the analysis of multivariate geostatistical data showing some degree of skewness. Our geostatistical model based on latent factors can be considered as an extension to skewed non-Gaussian data of the classical geostatistical proportional covariance model.

By framing our model in a hierarchical context, that is, by extending to the multivariate case the model-based geostatistical approach in [10], it would be possible to extend the present work to deal with regionalized variables of different kind. Instead of assuming that the conditional distributions of $Y_i(\mathbf{x})$ given $Z_i(\mathbf{x})$ are all skew-normal, we might assume, for different values of $i = 1, \ldots, m$, that they are of different type. For instance, [25] considers a model in which some of the (conditional) distributions, of the observable regionalized variables, are Poisson whereas some others are Gamma. In this way, we could obtain a model for non-Gaussian data flexible enough to account for observable regionalized variables showing different departures from normality.

On the other hand, a generalization in a different direction might involve the introduction of more spatial scales as in the classical linear model of coregionalization. This would supply a more flexible spatial autocorrelation structure in which the latent processes $Z_i(\mathbf{x})$, which are behind the level of the observable regionalized variables $Y_i(\mathbf{x})$, are not constrained to have proportional covariance and cross-covariance functions. However, the high level of complexity of this generalization would require a large amount of data to be detected and would pose serious inferential problems.

As regard to the model presented in this work, we presented a computationally intensive likelihood based inferential procedure, exploiting the capabilities of the MCEM algorithm. It must be noted that with this procedure we estimated just some of the parameters of the model, assuming the others as known. In particular, we assumed as known the parameters $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m)^T$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)^T$ that characterize the shape of the skew-normal (conditional) distributions. In this way we avoided many of the well known inferential problems posed by the estimation of the parameters of the skew-normal distribution. Although in this work we did not discuss any inferential procedure for these parameters, these can nevertheless be calibrated comparing the theoretical marginal distributions and the theoretical variograms with the corresponding empirical counterparts. From a computational perspective, although we checked the feasibility of our estimation procedure for reasonable sample sizes and for different parameter values, it must be remarked that in more complex situations the computational burthen might increase considerably.

Appendix

In this appendix we report some distributional results regarding the observable processes $Y_i(\mathbf{x})$. Let us first recall some definitions. Following, for instance, [2], we say that a random vector $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ has an *extended skew-normal distribution* with parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\alpha}$ and τ, and we write $\mathbf{Y} \sim \mathrm{ESN}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \tau)$, if it has probability density function of the form

$$f(\mathbf{y}) = \phi_n(\mathbf{y} - \boldsymbol{\mu}; \boldsymbol{\Sigma}) \cdot \Phi(\alpha_0 + \boldsymbol{\alpha}^T \mathbf{D}^{-1}(\mathbf{y} - \boldsymbol{\mu})) / \Phi(\tau), \quad \text{for} \quad \mathbf{y} \in \mathbb{R}^n, \qquad (10)$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$ is a vector of location parameters, $\phi_n(\cdot; \boldsymbol{\Sigma})$ is the n-dimensional normal density function with zero mean vector and (positive-definite) variance-covariance matrix $\boldsymbol{\Sigma}$ having elements σ_{ij}, $\Phi(\cdot)$ is the scalar $N(0,1)$ distribution function, $\mathbf{D} = \mathrm{diag}(\sigma_{11}, \ldots, \sigma_{nn})^{1/2}$ is the diagonal matrix formed with the standard deviations of the scale matrix $\boldsymbol{\Sigma}$, $\boldsymbol{\alpha} \in \mathbb{R}^n$ is a vector of skewness parameters, and $\tau \in \mathbb{R}$ is an additional parameter. Moreover, $\alpha_0 = \tau(1 + \boldsymbol{\alpha}^T \mathbf{R} \boldsymbol{\alpha})^{1/2}$ where $\mathbf{R}$ is the correlation matrix associated to $\boldsymbol{\Sigma}$, that is, $\mathbf{R} = \mathbf{D}^{-1} \boldsymbol{\Sigma} \mathbf{D}^{-1}$. Clearly, this distribution extends the multivariate normal distribution through the parameter vector $\boldsymbol{\alpha}$, and for $\boldsymbol{\alpha} = 0$ it reduces to the latter. When $\tau = 0$, also $\alpha_0 = 0$ and (10) reduces to

$$f(\mathbf{y}) = 2 \cdot \phi_n(\mathbf{y} - \boldsymbol{\mu}; \boldsymbol{\Sigma}) \cdot \Phi(\boldsymbol{\alpha}^T \mathbf{D}^{-1}(\mathbf{y} - \boldsymbol{\mu})), \quad \text{for} \quad \mathbf{y} \in \mathbb{R}^n. \qquad (11)$$

In this case we simply say that $\mathbf{Y}$ has a *skew-normal distribution* and we write, more concisely, $\mathbf{Y} \sim \mathrm{SN}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$.

According to [13] and [14], we say that the n-dimensional random vector $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ has a multivariate *closed skew-normal distribution*, and we write $\mathbf{Y} \sim \mathrm{CSN}_{n,m}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}_c, \boldsymbol{\nu}, \boldsymbol{\Delta})$, if it has probability density function of the form

$$f(\mathbf{y}) = \frac{1}{\Phi_m(0; \boldsymbol{\nu}, \boldsymbol{\Delta} + \mathbf{D}_c^T \boldsymbol{\Sigma} \mathbf{D}_c)} \cdot \phi_n(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \cdot \Phi_m(\mathbf{D}_c^T(\mathbf{y} - \boldsymbol{\mu}); \boldsymbol{\nu}, \boldsymbol{\Delta}), \quad \text{for} \quad \mathbf{y} \in \mathbb{R}^n, \qquad (12)$$

where: m is an integer greater than 0; $\boldsymbol{\mu} \in \mathbb{R}^n$; $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ is a positive-definite matrix; $\mathbf{D}_c \in \mathbb{R}^{n \times m}$ is an $n \times m$ matrix; $\boldsymbol{\nu} \in \mathbb{R}^m$ is a vector; $\boldsymbol{\Delta} \in \mathbb{R}^{m \times m}$ is a positive-definite matrix; and $\phi_n(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\Phi_n(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ are the probability density function and the cumulative distribution function, respectively, of the n-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$.

Though, as we have already noticed, the multivariate finite-dimensional marginal distributions of the multivariate spatial process $(Y_1(\mathbf{x}), \ldots, Y_m(\mathbf{x}))^T$, for $\mathbf{x} \in \mathbb{R}^2$, are not skew-normal (in the sense of [2]), it is possible to show that they are closed skew-normal, according to the definition of [13]. This implies that, for any given $i = 1, \ldots, m$, each univariate spatial process $Y_i(\mathbf{x})$ has all its finite-dimensional marginal distributions that are closed skew-normal. To see this (see also [24]), consider n spatial locations $\mathbf{x}_1, \ldots, \mathbf{x}_n$, and the corresponding n-dimensional random vector $\mathbf{Y} = (Y_i(\mathbf{x}_1), \ldots, Y_i(\mathbf{x}_n))^T$. Recalling that for any given $\mathbf{x} \in \mathbb{R}^2$ we can write $Y_i(\mathbf{x}) = \beta_i + Z_i(\mathbf{x}) + \omega_i S_i(\mathbf{x})$, the vector $\mathbf{Y}$ can be written as $\mathbf{Y} = \beta_i \mathbf{1}_n + \mathbf{Z} + \mathbf{D}_\omega \mathbf{S} = \mathbf{W} + \mathbf{V}$, where $\mathbf{W} = \beta_i \mathbf{1}_n + \mathbf{Z}$, $\mathbf{V} = \mathbf{D}_\omega \mathbf{S}$, $\mathbf{Z} = (Z_i(\mathbf{x}_1), \ldots, Z_i(\mathbf{x}_n))^T$, $\mathbf{S} = (S_i(\mathbf{x}_1), \ldots, S_i(\mathbf{x}_n))^T$ and $\mathbf{D}_\omega$ is the $n \times n$ diagonal matrix with ω_i on the diagonal. Now, since $S_i(\mathbf{x})$, for $\mathbf{x} \in \mathbb{R}^2$, are independently and identically distributed as $\text{CSN}_{1,1}(0, 1, \alpha_i, 0, 1)$, according to Theorem 3 of [14], we have that $\mathbf{S} \sim \text{CSN}_{n,n}(0, \mathbf{I}_n, \mathbf{D}_\alpha, 0, \mathbf{I}_n)$, where $\mathbf{D}_\alpha$ is the $n \times n$ diagonal matrix with α_i on the diagonal. On the other hand, since $\mathbf{Z}$ follows a multivariate normal distribution with mean 0 and covariance matrix $\mathbf{\Sigma}_Z$ with entries given by $\text{Cov}[Z_i(\mathbf{x}), Z_i(\mathbf{x} + \mathbf{h})] = \varsigma_i^2 \rho(\mathbf{h})$, we also have that $\mathbf{Z} \sim \text{CSN}_{n,1}(0, \mathbf{\Sigma}_Z, 0, 0, 1)$. Moreover, being $\mathbf{W}$ distributed as a multivariate normal with mean $\beta_i \mathbf{1}_n$ and covariance matrix $\mathbf{\Sigma}_Z$, we can write that $\mathbf{W} \sim \text{CSN}_{n,1}(\beta_i \mathbf{1}_n, \mathbf{\Sigma}_Z, 0, 0, 1)$, and using Theorem 1 of [14] we can also write that $\mathbf{V} \sim \text{CSN}_{n,n}(0, \mathbf{D}_{\omega^2}, \mathbf{D}_{\alpha/\omega}, 0, \mathbf{I}_n)$, where $\mathbf{D}_{\omega^2}$ is the $n \times n$ diagonal matrix with ω_i^2 on the diagonal, and $\mathbf{D}_{\alpha/\omega}$ is the $n \times n$ diagonal matrix with α_i/ω_i on the diagonal. Thus, considering that $\mathbf{Y} = \mathbf{W} + \mathbf{V}$, we can conclude, using Theorem 4 of [14], that $\mathbf{Y} \sim \text{CSN}_{n,n+1}(\beta_i \mathbf{1}_n, \mathbf{\Sigma}_Z + \omega_i^2 \mathbf{I}_n, \mathbf{D}^*, 0, \mathbf{\Delta}^*)$, for some matrices $\mathbf{D}^*$ and $\mathbf{\Delta}^*$.

References

1. Allard, D., Naveau, P.: A new spatial skew-normal random field model. Commun. Stat. Theory **36**, 1821–1834 (2007)
2. Azzalini, A.: The skew-normal distribution and related multivariate families. Scand. J. Stat. **32**, 159–188 (2005)
3. Azzalini, A., Dalla Valle, A.: The multivariate skew-normal distribution. Biometrika **83**, 715–726 (1996)
4. Bechler, A., Romary, T., Jeannée, N., Desnoyers, Y.: Geostatistical sampling optimization of contaminated facilities. Stoch. Env. Res. Risk A. **27**, 1967–1974 (2013)
5. Brenning, A., Dubois, G.: Towards generic real-time mapping algorithms for environmental monitoring and emergency detection. Stoch. Env. Res. Risk A. **22**, 601–611 (2008)
6. Chagneau, P., Mortier, F., Picard, N., Bacro, J.-N.: A hierarchical Bayesian model for spatial prediction of multivariate non-Gaussian random fields. Biometrics **67**, 97–105 (2011)
7. De, S., Faria, Á.E.: Dynamic spatial Bayesian models for radioactivity deposition. J. Time Ser. Anal. **32**, 607–617 (2011)
8. De Oliveira, V., Kedem, B., Short, D.A.: Bayesian prediction of transformed Gaussian random fields. J. Am. Stat. Assoc. **92** 1422–1433 (1997)

9. Desnoyers, Y., Chilès, J.-P., Dubot, D., Jeannée, N., Idasiak, J.-M.: Geostatistics for radiological evaluation: study of structuring of extreme values. Stoch. Env. Res. Risk A. **25**, 1031–1037 (2011)
10. Diggle, P.J., Moyeed, R.A., Tawn, J.A.: Model-based geostatistics (with discussion). Appl. Stat. **47**, 299–350 (1998)
11. Dubois, G., Galmarini, S.: Spatial interpolation comparison (SIC) 2004: introduction to the exercise and overview of results. In: Dubois, G. (ed.) Automatic Mapping Algorithms for Routine and Emergency Monitoring Data - Spatial Interpolation Comparison 2004, Office for Official Publication of the European Communities (2005)
12. Fort, G., Moulines, E.: Convergence of the Monte Carlo expectation maximization for curved exponential families. Ann. Stat. **31**, 1220–1259 (2003)
13. 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, London (2004)
14. González-Farías, G., Domínguez-Molina, J.A., Gupta, A.K.: Additive properties of skew-normal random vectors. J. Stat. Plan. Infer. **126**, 521–534 (2004)
15. Gräler, B.: Modelling skewed spatial random fields through the spatial vine copula. Spat. Stat. (2014). http://dx.doi.org/10.1016/j.spasta.2014.01.001
16. Herranz, M., Romero, L.M., Idoeta, R., Olondo, C., Valiño, F., Legarda, F.: Inventory and vertical migration of ^{90}Sr fallout and ^{137}Cs/^{90}Sr ratio in Spanish mainland soils. J. Env. Radioact. **102**, 987–994 (2011)
17. Hosseini, F., Eidsvik, J., Mohammadzadeh, M.: Approximate Bayesian inference in spatial GLMM with skew normal latent variables. Comput. Stat. Data Anal. **55**, 1791–1806 (2011)
18. Kazianka, H., Pilz, J.: Copula-based geostatistical modeling of continuous and discrete data including covariates. Stoch. Env. Res. Risk A. **24**, 661–673 (2010)
19. Kazianka, H., Pilz, J.: Bayesian spatial modeling and interpolation using copulas. Comput. Geosci. **37**, 310–319 (2011)
20. Kim, H.-M., Mallick, B.K.: A Bayesian prediction using the skew Gaussian distribution. J. Stat. Plan. Infer. **120**, 85–101 (2004)
21. Lunn, D., Spiegelhalter, D., Thomas, A., Best, N.: The BUGS project: evolution, critique and future directions. Stat. Med. **28**, 3049–3067 (2009)
22. Maglione, D.S., Diblasi, A.,M.: Exploring a valid model for the variogram of an isotropic spatial process. Stoch. Env. Res. Risk A. **18**, 366–376 (2004)
23. McLachlan, G.J., Krishnan, T.: The EM Algorithm and Extensions, 2nd edn. Wiley, New York (2007)
24. Minozzo, M., Ferracuti, L.: On the existence of some skew-normal stationary processes. Chil. J. Stat. **3**, 157–170 (2012)
25. Minozzo, M., Ferrari, C.: Multivariate geostatistical mapping of radioactive contamination in the Maddalena Archipelago (Sardinia, Italy). AStA Adv. Stat. Anal. **97**, 195–213 (2013)
26. Minozzo, M., Fruttini, D.: Loglinear spatial factor analysis: an application to diabetes mellitus complications. Environmetrics **15**, 423–434 (2004)
27. Oliver, M.A., Badr, I.: Determining the spatial scale of variation in soil radon concentration. Math. Geol. **27**, 893–922 (1995)
28. Pilz, J., Spöck, G.: Why do we need and how should we implement Bayesian kriging methods. Stoch. Env. Res. Risk A. **22**, 621–632 (2008)
29. R Development Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0 (2012)
30. Ren, Q., Banerjee, S.: Hierarchical factor models for large spatially misaligned data: a low-rank predictive process approach. Biometrics **69**, 19–30 (2013)

31. Schmidt, A.M., Rodriguez, M.A.: Modelling multivariate counts varying continuously in space. Bayesian Stat. **9** (2011). doi:10.1093/acprof:oso/9780199694587.003.0020
32. Spöck, G.: Spatial sampling design with skew distributions: the special case of trans-Gaussian kriging. Ninth International Geostatistical Congress, Oslo, Norway, 11–15 June 2012, 20 pages
33. Wibrin, M., Bogaert, P., Fasbender, D.: Combining categorical and continuous spatial information within the Bayesian maximum entropy paradigm. Stoch. Env. Res. Risk A. **20**, 423–433 (2006)
34. Zhang, H.: On estimation and prediction for spatial generalized linear mixed models. Biometrics **58**, 129–136 (2002)
35. Zhang, H., El-Shaarawi, A.: On spatial skew-Gaussian processes and applications. Environmetrics **21**, 33–47 (2010)