Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
A flexible special case of the CSN for spatial modeling and prediction

José Ulises Márquez-Urbina \textsuperscript{a,b}, Graciela González-Farías \textsuperscript{c,*}

\textsuperscript{a} Centro de Investigación en Matemáticas Unidad Monterrey, Av. Alianza Centro No. 502, Parque de Investigación e Innovación Tecnológica (PIIT), Apodaca, 66628, N.L., Mexico
\textsuperscript{b} Consejo Nacional de Ciencia y Tecnología, Av. Insurgentes Sur 1582, Col. Crédito Constructor, Benito Juárez, 03940, CDMX, Mexico
\textsuperscript{c} Centro de Investigación en Matemáticas, Jalisco s/n, Valenciana, Guanajuato, 36023, GTO, México

\textbf{ARTICLE INFO}

\textbf{Article history:}
Received 3 April 2021
Received in revised form 5 August 2021
Accepted 18 November 2021
Available online 26 November 2021

\textbf{Keywords:}
Flexible closed-skew normal
Profile predictive likelihood
COVID-19

\textbf{ABSTRACT}

We introduce a parsimonious, flexible subclass of the closed-skew normal (CSN) distribution that produces valid stationary spatial models. We derive and prove some relevant properties for this subfamily; in particular, we show that it is identifiable, closed under marginalization and conditioning and that a null correlation implies independence. Based on the subclass, we propose a discrete spatial model and its continuous version. We discuss why these random fields constitute valid models, and additionally, we discuss least-squares estimators for the models under the subclass. We propose to perform predictions on the model using the profile predictive likelihood; we discuss how to construct prediction regions and intervals. To compare the model against its Gaussian counterpart and show that the numerical likelihood estimators are well-behaved, we present a simulation study. Finally, we use the model to study a heuristic COVID-19 mortality risk index; we evaluate the model’s performance through 10-fold cross-validation. The risk index model is compared with a baseline Gaussian model.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

The data collected in many scientific studies and applications exhibit skewness and a differentiated behavior at the distribution tails. In recent years, to have spatial models capable of reproducing...
such properties, there has been an interest in generalizing the well-studied Gaussian random field. Proposals in this direction intend, in different ways, to construct models with finite-dimensional distributions belonging to tractable families more extensive than the Gaussian class. In this context, since they share multiple statistical properties with the Gaussian distribution while introducing skewness, the different skew-normal families constitute attractive options. Among these skew families, the multivariate closed skew-normal (CSN) distribution is of particular interest because it is closed under some essential statistical operations.

The CSN probability distribution was introduced in a series of papers (Gupta et al., 2004; González-Farías et al., 2004a) as an extension of Azzalini’s skew-normal distribution. The CSN distribution can exhibit skewness in all its coordinates and, as pointed before, shares several advantages for statistical modeling with the normal distribution. Among its more useful properties, which explain its name, the CSN family is closed under singular and non-singular linear transformations, marginalization, conditioning, and the addition and concatenation of independent CSN random vectors. These closure properties and its ability to introduce skewness on all coordinates distinguish the CSN family from Azzalini’s distribution. However, to satisfy these properties, the CSN requires more parameters than Azzalini’s skew-normal distribution.

The CSNdistribution was first derived using a hidden truncation argument; more precisely, by using a simple multivariate extension of the model in Copas and Li (1997). Such a genesis, which shares in some degree with the skew-normal, is an essential characteristic of the CSN distribution and provides it with interpretability. As some authors have shown (see Arnold and Beaver, 2002), hidden truncation always introduces skewness into a distribution.

The CSN distribution has been previously employed in the context of spatial models. Allard and Naveau (2007) propose a class of random fields defined in terms of the multivariate CSN, a moment-based procedure to estimate its parameters, and discuss a spatial prediction procedure. Karimi and Mohammadzadeh (2011) define a discrete closed-skew Gaussian random field and propose a Bayesian prediction method employing CSN distributions; they apply their model and prediction method to strain data. Karimi and Mohammadzadeh (2012) define a spatial regression model with CSN correlated errors and employ a Bayesian approach to model and predict missing observations; their model is applied to CO data from Tehran, Iran. Rimstad and Omre (2014) extend (Allard and Naveau, 2007) by using a grid representation and a similar parameterization; they estimate the model by maximum likelihood and use a Bayesian approach to predict elastic variables in seismic data from the North Sea. In all these references and many others, they do not study the involved processes' formal existence and stationarity. For practical purposes, it is true that many calculations exhibit a good fit to the data, but their replication outside the sample cannot be guaranteed.

Minozzo and Ferracuti (2012) noted that some spatial models reported in the literature with CSN and skew-normal finite-dimensional distributions are ill-defined, including some of the models discussed above. They arrive at this conclusion by noticing that the finite-dimensional distributions in these stochastic processes do not correctly characterize them. Minozzo and Ferracuti (2012) also present some alternatives for hierarchical models that could be properly characterized by the finite-dimensional distributions. Mahmoudian (2018) expands Minozzo and Ferracuti’s ideas and details a formal condition that guarantees that spatial random fields with CSN and normal-skew finite-dimensional distributions are marginal consistent in the sense of Kolmogorov’s extension theorem (Billingsley, 1995). Furthermore, Mahmoudian (2018) proposes some valid skew distributions to construct spatial random fields that satisfy the marginal consistency; most of them do not correspond to the models mentioned above.

Other skew-distributions have been employed to define spatial or spatio-temporal models. A recent example of this is the publication by Tagle et al. (2019). We do not further discuss them because they are not central to this work. In such cases, the validity and existence of such models is also discussed by Minozzo and Ferracuti (2012) and Mahmoudian (2018), but for the most part they remain unclear.

In the present work, we introduce a particular flexible case of the CSN distribution that produces valid stationary spatial random fields and whose parameters are interpretable. This parameterization was first presented by the second author in 2006 at Chicago during the event Multivariate Methods in Envirometrics. In allusion to this subclass’s flexibility, we name it the FS-CSN family.
This subclass shares several important features with the normal distribution; in particular, we show that this subclass is identifiable, closed under marginalization and conditioning, and that a null correlation implies independence. Besides, we specify the distribution that a random vector with the FS-CSN distribution follows after applying a linear transformation. The FS-CSN subclass is similar to one discussed by Mahmoudian (2018); it differs in the type of square root matrix employed for the parameterization and that, in our case, the mean and covariance matrix are the parameters in the subclass. Mahmoudian (2018) does not discuss any of the properties proved in this work nor does he use them for analyzing data.

Using the FS-CSN, we propose a spatial model for a response variable defined over a lattice and discuss the profile likelihood function’s approach to predict the response variable at unobserved sites. An important property satisfied by the model is that its first two moments correspond with the observations’ first two moments. This property permits interpreting the mean and variance analogous to the Gaussian case and applying all the usual tools employed in analyzing spatial data. As we later note in our discussion, when this identification fails as it happens in various CSN models in the literature, it is unclear the validity of using, for instance, variograms to gain information about the variables under study. We also specify the continuous version of the spatial model mentioned before. We discuss the validity of both models and other properties of the models; in particular, we note that they satisfy Kolmogorov’s extension theorem marginal consistency (Billingsley, 1995).

Furthermore, we propose to perform prediction in the model’s framework using the profile predictive likelihood (Bjornstad, 1990). The profile predictive likelihood (PPL) approach was introduced by Mathiasen (1979); it uses the information in the joint distribution of the data and the observations to perform prediction. We also address how to construct prediction regions and intervals based on the PPL. We do a simulation study to assess the likelihood estimation for the proposed model and empirically compare it with the estimates of its Gaussian counterpart when the data is skewed. We conclude by applying the discrete spatial model introduced in this work to understand a heuristic mortality risk measure for COVID-19 in the Northeast of Mexico. We model this risk measure at a municipal level using a set of variables containing information about access to healthcare services, the prevalence of comorbidities in the population, and the region’s socioeconomic status. For comparison, a Gaussian model is also fitted. The performance of both models and their PPL-predictions is studied via 10-fold cross-validation. In this application, we show that the prediction produces good results and that the model also describes well the observed sites. In an appendix, we also study the least-squares estimators and residuals for linear models with FS-CSN errors; we deduce the distribution of the generalized least-squares estimator and its corresponding residuals. The discussion of all these results is presented in the respective sections.

The paper is organized as follows. Section 2 presents a summary of the properties of the CSN class relevant to this work. In Section 3, we introduce the FS-CSN. Section 4 presents the spatial models considered in this work and their respective properties. Section 5 summarizes some elements of the PPL and explains how to determine prediction regions. Section 6 contains the simulation study, and Section 7 includes the model’s application to the heuristic risk measure for COVID-19. Section 8 presents the conclusion of this work. The paper also includes four appendices. Appendix A presents the proofs for the properties of the FS-CSN subclass. Appendix B discusses the generalized least-squares estimator for linear models with FS-CSN errors. Finally, Appendix C briefly presents a motivation for the risk index considered, while Appendix D includes the tables produced in this work.

2. Preliminaries

The closed-skew normal (CSN) distribution generalizes the usual normal model by adding skewness into it. In this section, we discuss some characteristics of this distribution that are employed in this work. We recommend (González-Farías et al., 2004b) for a comprehensive report of the CSN distribution’s features.

The CSN distribution arises naturally as a consequence of a hidden truncation process in a population normally distributed. Such truncation occurs, for example, when we select the data by only recording a value when a condition is satisfied. For the CSN class’s deduction, the hidden
Truncation process is represented by conditioning a normal random vector on a set of latent variables subject to an explicitly given restriction (namely, \( Z \geq 0 \)). This hidden truncation and conditioning argument represents a multivariate extension of the model for missing data \( (Y \) is observed only if \( Z > 0 \)) or comparative trials (a subject is allocated to a treatment or other depending on if \( Z > 0 \) or \( Z \leq 0 \), respectively) introduced by Copas and Li (1997). This hidden truncation argument generates the following probability density function.

Let \( \phi_p(\cdot; \mu, \Sigma) \) and \( \Phi_p(\cdot; \mu, \Sigma) \) denote the pdf and the cdf, respectively, of a \( p \)-dimensional normal random vector with mean \( \mu \) and covariance matrix \( \Sigma \). The \( p \)-dimensional multivariate pdf of the CSN distribution is given by

\[
f(w) = \phi_p(w; \mu, \Sigma) \frac{\Phi_q(D(w - \mu); v, \Delta)}{\Phi_q(0; v, \Delta + D\Sigma D')} \quad w \in \mathbb{R}^p,
\]

where \( q \in \mathbb{N}, \mu \in \mathbb{R}^p, v \in \mathbb{R}^q, \Sigma \) is a \( p \times p \) positive definite scale matrix, \( D \) is a \( q \times p \) matrix, and \( \Delta \) is a \( q \times q \) positive definite scale matrix. We will denote the distribution associated to (1) by \( \text{CSN}_p(q; \mu, \Sigma, D, v, \Delta) \).

Notice that the normalization in the pdf of the FS-CSN distribution is not, in general, \( 1/2^m \) unless we consider particular parameters. In the literature, there are representations of multivariate skew-distributions with normalization constants \( 1/2^m \); however, in some cases, the pdf does not integrate one because the parametric space does not include the whole hyperplane involved in the parameterization. As a matter of fact, this is the reason why the CSN was developed (Gupta et al., 2004).

In the distribution \( \text{CSN}_p(q; \mu, \Sigma, D, v, \Delta) \), the parameters \( \mu \) and \( \Sigma \) correspond to location and scale parameters, respectively. The dimension \( q \) can be interpreted as the degree of skewness freedom in the distribution. The parameters \( D, v, \) and \( \Delta \) constitute skewness parameters; they influence the skewness size, but the matrix \( D \) also controls its direction. The normal and skew-normal distributions are particular cases of the CSN distribution: taking \( q = 0 \) or \( D = 0 \) produces a Gaussian distribution, while \( q = 1 \) and \( v = 0 \) reduce the CSN to the skew-normal class.

The CSN distribution has a marginal representation (Domínguez-Molina et al., 2007) which we employ since it allows us to work the computational aspects of the CSN distribution. Let \( N_p(\mu_0, \Sigma_0) \) denote an \( n \)-dimensional multivariate normal distribution with mean \( \mu_0 \in \mathbb{R}^n \) and covariance matrix \( \Sigma_0 \) and let \( T_{N_0}(\mu_0, \Sigma_0; v_0) \) denote a \( N_n(\mu_0, \Sigma_0) \) distribution truncated below \( v_0 \in \mathbb{R}^n \). Besides, let \( v \sim N_p(0, I_p), \ u \sim TN_q(0, \Delta + D\Sigma D'; v), I_p \) be the \( p \times p \) identity matrix, and \( u \) be independent of \( v \). Then, the distribution of

\[
y = \mu + (\Sigma^{-1} + D'\Delta^{-1}D)^{-1/2} v + SD' (\Delta + D\Sigma D')^{-1} u
\]

is \( \text{CSN}_p(q; \mu, \Sigma, D, v, \Delta) \). Besides, as shown in González-Farías et al. (2004a), the moment generating function (mgf) corresponding to the \( \text{CSN}_p(q; \mu, \Sigma, D, v, \Delta) \) distribution is given by

\[
M(t) = \frac{\Phi_q(D\Sigma t; v, \Delta + D\Sigma D')}{\Phi_q(0; v, \Delta + D\Sigma D')} e^{t'\mu} + \frac{1}{2} t'\Sigma t, \quad t \in \mathbb{R}^p.
\]

The CSN distribution has many properties that are relevant for statistical modeling. For mentioning some, we have that the CSN distribution is closed under all types of linear transformations, disregarding its rank condition; closed under marginalization and conditioning; closed under concatenation of independent CSN random vectors; and closed under convolutions. It is important to note that all these statistical operations commute with the hidden truncation process. In other words, the distribution resulting from applying the corresponding statistical operation on the normal random vector and then do the hidden truncation process yields precisely the same distribution obtained if we do the operation after the truncation process. The exact formulas and results appear in Díaz-García and González-Farías (2008), Domínguez-Molina et al. (2003), González-Farías et al. (2004a).

Due to its utility for this work, we specify the precise result for the closure under linear transformations. Namely, if \( X \sim \text{CSN}_p(q; \mu, \Sigma, D, v, \Delta) \), \( l \leq p \) is a natural number, \( A \) is an \( l \times p \) matrix of rank \( l \), and \( c \) is an \( l \)-dimensional vector, then

\[
Y = AX + c \sim \text{CSN}_{l,q}(\mu_Y, \Sigma_Y, D_Y, v_Y, \Delta_Y).
\]
where
\[
\mu_Y = A\mu + c, \quad \Sigma_Y = A\Sigma A', \quad D_Y = D\Sigma A'\Sigma_Y^{-1}, \quad v_Y = v,
\]
\[
\Delta_Y = \Delta + D\Sigma D' - D\Sigma A'\Sigma_Y^{-1}A\Sigma D'.
\]
If \( A \) is an arbitrary matrix, the formula (4) remains valid if \( \Sigma_Y^{-1} \) is replaced with a symmetric generalized inverse of \( \Sigma_Y \) (Díaz-García and González-Farías, 2008).

3. A flexible subclass of the closed-skew normal distribution

According to Mahmoudian (2018), because of the form of their finite-dimensional distributions, the existence of some of the skew-gaussian random fields in the literature is not guaranteed. In the following, we introduce a particular subclass of the CSN distribution that produces valid spatial models in Mahmoudian’s sense; we called it FS-CSN. This subclass shares some useful properties with the multivariate normal distribution; in particular, it is identifiable, closed under marginalization and conditioning, and within the subfamily, null correlation implies independence. Besides, it is possible to obtain the distribution for linear transformations of FS-CSN random vectors.

Another nice characteristic of the FS-CSN subclass is that it is parameterized in terms of its first two moments and a parameter controlling the skewness, making them easy to interpret and manipulate. Thanks to these properties, as we review in Section 5, it is possible to obtain valid, interpretable spatial models that can be used in conjunction with the usual tools to analyze spatial data.

The section is divided into two subsections. The first subsection introduces the FS-CSN subclass, while the second subsection presents some of its properties.

3.1. Motivation and definition of the FS-CSN subclass

Let \( \lambda \in \mathbb{R}, n \in \mathbb{N}, \) and let \( I_n \) denote the \( n \times n \) identity matrix. Assume that \( X_i \sim \text{CSN}_{1,1}(0, 1, \lambda, 0, 1), \) \( i = 1, \ldots, n, \) are independent. Thus, \( X \equiv (X_1, \ldots, X_n)' \sim \text{CSN}_{n,n}(0, I_n, \lambda I_n, 0, I_n). \) Let \( b = (2/\pi)^{1/2} \) and \( \delta = \lambda(1 + \lambda^2)^{-1/2}. \) Using the representation (2), we obtain that
\[
E(X_i) = b\delta, \quad \text{Var}(X_i) = 1 - b^2\delta^2, \quad i = 1, 2, \ldots, n.
\]

Then, the random variable \( Y \) resulting from the classical standardization
\[
Y \equiv \frac{1}{\sqrt{1 - b^2\delta^2}}(X - b\delta I_n), \quad 1_n = (1, 1, \ldots, 1)' \in \mathbb{R}^n,
\]
satisfies
\[
E(Y) = 0, \quad \text{Var}(Y) = I_n.
\]

Let \( \Sigma_n \) be a \( n \times n \) positive definite matrix. Since \( \Sigma_n \) is positive definite, Cholesky factorization theorem implies that there exists a unique lower triangular matrix \( \Sigma^c_n \) with positive diagonal entries such that \( \Sigma_n = \Sigma^c_n(\Sigma^c_n)' \). Let \( \tau = (1 - b^2\delta^2)^{-1/2}, \mu \in \mathbb{R}^n, \sigma > 0, \) and define
\[
Z_n := \mu + \sigma \Sigma^c_n Y.
\]

From relation (4) we obtain
\[
Z_n \sim \text{CSN}_{n,n}(\mu - b\delta\sigma\tau \Sigma^c_n 1_n, \sigma^2\tau^2 \Sigma_n, \frac{\lambda}{\sigma\tau}(\Sigma^c_n)^{-1}, 0, I_n).
\]

Besides,
\[
E(Z_n) = \mu + \sigma \Sigma^c_n E(Y) = \mu, \quad \text{Var}(Z_n) = \sigma^2 \Sigma^c_n \text{Var}(Y)(\Sigma^c_n)' = \sigma^2 \Sigma_n.
\]

Remark 3.1. If \( \Sigma \) is a positive definite matrix, \( \Sigma^c \) will denote the unique lower triangular matrix with positive diagonal entries such that \( \Sigma = \Sigma^c(\Sigma^c)' \). As we discussed before, we know that such a matrix exists and is unique, thanks to the Cholesky decomposition theorem. Besides, \( \Sigma^{-c} \) will denote the inverse of \( \Sigma^c \). We will refer to \( \Sigma^c \) as the Cholesky factor associated with \( \Sigma \).
Under the conditions considered for its construction, the distribution in (6) defines a subclass of the multivariate closed-skew normal distribution. Such a subclass is entirely determined by its first two moments \((\mu, \sigma^2 \Sigma_n)\), the skewness parameter \(\lambda\), and the factor \(\lambda\). For simplicity, we will denote this subclass as FS-CSN\(_n(\mu, \Sigma_n, \sigma, \lambda)\).

**Definiton 3.2.** Let \(\lambda \in \mathbb{R}, \sigma > 0, \mu \in \mathbb{R}^n\) and \(\Sigma_n\) be a \(n \times n\) positive definite matrix. Define \(b = (2/\pi)^{1/2}, \delta \equiv \delta(\lambda) = \lambda(1 + \lambda^2)^{-1/2},\) and \(\tau \equiv \tau(\lambda) = (1 - b^2 \delta^2)^{-1/2}\). For an \(n\)-dimensional random vector \(Z_n\), we write \(Z_n \sim \text{FS-CSN}(\mu, \Sigma_n, \sigma, \lambda)\) if

\[
Z_n \sim \text{CSN}_n(\mu - b\delta \sigma \tau \Sigma_n^{-1} \mathbf{1}_n, \sigma^2 \tau^2 \Sigma_n, \frac{\lambda}{\sigma \tau} \Sigma_n^{-c}, \mathbf{0}, I_n).
\]

Besides, we say that an \(n\)-dimensional random vector \(Z_n\) belongs to the subclass FS-CSN if \(Z_n \sim \text{FS-CSN}(\mu, \Sigma_n, \sigma, \lambda)\) for some parameters \((\mu, \Sigma_n, \sigma, \lambda)\) satisfying the previous conditions.

**Remark 3.3.** For simplicity, when working with the FS-CSN subclass, we refrain from making explicit the dependence on \(\lambda\) of the quantities \(\delta\) and \(\tau\). To avoid confusions, we only employ these Greek letters in the FS-CSN distribution context. Also, in the following paragraphs, we always take \(b = (2/\pi)^{1/2}\).

As noted before, for \(Z_n \sim \text{FS-CSN}(\mu, \Sigma_n, \sigma, \lambda)\), we have that

\[
E(Z_n) = \mu, \quad \text{Var}(Z_n) = \sigma^2 \Sigma_n.
\]

Besides, as we show in Proposition 3.4, the parameter \(\lambda\) induces skewness into the model. For \(\lambda = 0\), the FS-CSN class reduces to the multivariate normal distribution.

Substituting the corresponding parameters in (1) and doing some simplifications, we obtain that the pdf associated with the distribution FS-CSN\(_n(\mu, \Sigma_n, \sigma, \lambda)\) is

\[
f(\mathbf{z}) = 2^n \phi_n(\mathbf{z}; \mu - b\delta \sigma \tau \Sigma_n^{-1} \mathbf{1}_n, \sigma^2 \tau^2 \Sigma_n) \phi_n(\frac{\lambda}{\sigma \tau} \Sigma_n^{c-1}(\mathbf{z} - \mu) + b\delta \lambda \mathbf{1}_n; \mathbf{0}, I_n),
\]

for all \(\mathbf{z} \in \mathbb{R}^n\). Furthermore, a direct consequence of formula (3) is that the moment generating function of the distribution FS-CSN\(_n(\mu, \Sigma_n, \sigma, \lambda)\) is given by

\[
M(\mathbf{t}) = 2^n \phi_n(\lambda \sigma \tau (\Sigma_n^{c-1}) \mathbf{t}; 0, (1 + \lambda^2)^{1/2}) e^{t^\top \mu - b\delta \sigma \tau^\top \Sigma_n^{c-1} \mathbf{1}_n + \frac{1}{2} \sigma^2 \tau^2 \Sigma_n \mathbf{t}}, \quad \mathbf{t} \in \mathbb{R}^n.
\]

### 3.2. Some properties of the FS-CSN subclass

In the following, we state some properties for the FS-CSN subfamily. The corresponding proofs are detailed in Appendix A. We start with expressions for the FS-CSN class’s skewness and kurtosis.

**Proposition 3.4.** Let \(Z_1 \sim \text{FS-CSN}_1(\mu, 1, \sigma, \lambda)\). Then, the skewness \(\gamma_1\) and the excess kurtosis \(\gamma_2\) of \(Z_1\) are given by

\[
\gamma_1 = b(2b^2 - 1)\delta^3 \tau^3,
\]

\[
\gamma_2 = \tau^4(-3b^4 \delta^4 - 6b^2 \delta^2 + 4b^2 \delta^2 + 3) - 3.
\]

The choice \(\Sigma_1 = 1\) is not restrictive because, in such an instance, the variance can be controlled by the parameter \(\sigma\). Besides, notice that \(\gamma_1\) and \(\gamma_2\) do not depend on the parameters \(\mu, \Sigma, \) and \(\sigma\).

The FS-CSN subclass is closed under marginalization, with the marginal distribution only depending on its corresponding parameters and dimension.

**Proposition 3.5.** Let \(Z_n \sim \text{FS-CSN}_n(\mu, \Sigma_n, \sigma, \lambda)\) and let \(Z_n\) be partitioned into \(Z_n = (W_1, W_2)\), where \(W_1\) has dimension \(1 \leq k < n\). Besides, assume that \(\mu\) and \(\Sigma_n\) are partitioned into

\[
\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma_n = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},
\]

where \(\mu_1 \in \mathbb{R}^k\) is a vector and \(\Sigma_{11}\) is a \(k \times k\) matrix. Then, the random vector \(W_1\) satisfies \(W_1 \sim \text{FS-CSN}_n(\mu_1, \Sigma_{11}, \sigma, \lambda)\).
Remark 3.6. Proposition 1 in Mahmoudian (2018) states that a skew-normal distribution satisfies Kolmogorov’s marginal consistency property if, and only if, it is closed under marginalization, and the marginal distribution only depends on its corresponding parameters and dimension. Thus, a direct implication of Proposition 3.5 is that the FS-CSN subclass satisfies Kolmogorov’s extension theorem’s marginal consistency property.

We also have that the FS-CSN subclass is closed under conditioning. As we discuss in Section 5, this property is handy to determine confidence regions.

Proposition 3.7. Let \( Z_n \sim \text{FS-CSN}_n(\mu, \Sigma_n, \sigma, \lambda) \). Assume that \( Z_n, \mu, \) and \( \Sigma_n \) are partitioned as in Proposition 3.5. Besides, define

\[
\begin{align*}
\nu_1 &= \mu_1 - b\delta \sigma \tau \Sigma_n^{12} \mu_k, \\
\nu_2 &= \mu_2 - b\delta \sigma \tau [\Sigma_n^{12} \mu_n]^{n-k},
\end{align*}
\]

where, for \( \mathbf{x} \in \mathbb{R}^n, [\mathbf{x}]^k = (x_{n-k+1}, \ldots, x_n) \) denotes the projection on the last \( k \) components of \( \mathbf{x} \). Then, the pdf \( f_{W_2|W_1} \) of the random vector \( W_2|W_1 = w_1 \) is given by

\[
f_{W_2|W_1}(w_2|w_1) = 2^{n-k}f_{n-k}(w_2; \mu_{2|1}, \sigma^2 \Sigma_{2|1}) \phi_{n-k} \left( \frac{\lambda}{\sigma \tau} \Sigma_{2|1}^{-\kappa}(w_2 - \mu_{2|1}); 0, I_n \right)
\]

for \( w_1 \in \mathbb{R}^k \) and \( w_2 \in \mathbb{R}^{n-k} \), where \( \mu_{2|1} = \nu_2 + \Sigma_{21} \Sigma_{11}^{-1}(w_1 - \nu_1) \) and \( \Sigma_{2|1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \).

As we show in the following proposition, within the FS-CSN class, null-correlation implies independence.

Proposition 3.8. Let \( Z_n \sim \text{FS-CSN}_n(\mu, \Sigma_n, \sigma, \lambda) \) and let \( Z_n \) be partitioned into \( Z_n = (W_1, W_2) \), where \( W_1 \) has dimension \( 1 \leq k \leq n \). Also, assume that \( \mu \) and \( \Sigma_n \) are partitioned as indicated in Eq. (10). The FS-CSN random vectors \( W_1 \) and \( W_2 \) are independent if and only if \( \text{Cov}(W_1, W_2) = \Sigma_{12} = \Sigma_{21} = 0 \).

The FS-CSN distribution is not closed under linear transformations. However, formula (4) implies that if we apply an affine transformation to an FS-CSN random vector, we obtain a distribution very similar to the FS-CSN subfamily.

Proposition 3.9. Let \( Z_n \sim \text{FS-CSN}_n(\mu, \Sigma_n, \sigma, \lambda) \). Assume that \( B \) is an \( l \times n \) matrix and \( e \in \mathbb{R}^n \) is a deterministic vector. Then,

\[
W \equiv BZ_n + e \sim \text{CSN}_n, (\mu_W, \Sigma_W, D_W, 0, \Delta_W),
\]

with

\[
\begin{align*}
\mu_W &= B\mu + e - b\delta \sigma \tau B \Sigma_n^{12} \mu_n, \\
\Sigma_W &= \sigma^2 \tau^2 B \Sigma_n B', \\
D_W &= \frac{\lambda}{\sigma \tau} (B \Sigma_n^C)'^+, \\
\Delta_W &= I_n + \lambda^2 (I_n - (B \Sigma_n^C)'^+) B \Sigma_n^C,
\end{align*}
\]

where \( C^+ \) denote the Moore–Penrose pseudoinverse of \( C \). Besides, if \( B \) has rank \( l \), \( \Delta_W = I_n \).

Proposition 3.9 can be employed to determine the least-squares estimators for linear models. In Appendix B, we have included a discussion about least-squares estimators when the response variable follows an FS-CSN distribution for those interested in more detail.

Including restrictions on \( \Sigma_n \), we can make the FS-CSN class identifiable. This property is particularly relevant to define and estimate models.

Proposition 3.10. Assume that \( \Sigma_n \) is a correlation matrix. Then, the subclass \( \text{FS-CSN}_n(\mu, \Sigma_n, \sigma, \lambda) \) is identifiable.

There may be other properties useful for modeling that the FS-CSN distribution might satisfy. They will be addressed in other investigations. We conclude the section with the following remark.
Remark 3.11. The FS-CSN distribution has an essential property for linear modeling. Namely, if we assume that a response variable follows an FS-CSN distribution, then the errors inherit the distribution. The CSN distribution also has this attribute, but in the case of the FS-CSN, the covariance matrix is inherited intact. This property, which might seem simple, is not valid for many skew distributions, like Azzalini’s distribution.

4. FS-CSN spatial models

In this section, we introduce two spatial random fields based on the FS-CSN subclass. The first model, defined on a finite grid, expresses a response variable as a linear combination of a set of covariables. The second spatial random field is the continuous version of the first model. As we later discuss, both models are based on a stationary FS-CSN random field.

An important property satisfied by both models, overlooked and failed by many spatial skew models in the literature, is that the model’s first two moments correspond with the observations’ first two moments. Such property is very relevant because it allows an interpretation of the mean and variance analogous to the Gaussian case. Thanks to this, despite the non-null skewness, it is possible to apply and interpret all the usual tools employed in analyzing spatial data, like the variograms. It is important to remark that this does not occur when the model’s moments do not correspond with the data’s moments; for instance, when the covariance does not coincide with the data’s covariance, it is invalid to use variograms in the usual manner. The models proposed are based on a correlation model with exponential decay. As we mention, this assumption is not fundamental since different correlation models can be used.

4.1. A finite spatial model

In this subsection, we introduce a spatial model over a finite grid. Assume that $\mathcal{D} \subseteq \mathbb{R}^d$ is a finite collection of spatial sites, not necessarily spatially regular, and write $\mathcal{D} = \{s_1, \ldots, s_n\}$, $n \in \mathbb{N}$. Let $Y(s)$ denote a real-valued response variable at the site $s \in \mathcal{D}$. We model the random vector $Y \equiv (Y(s_1), \ldots, Y(s_n))$ through the linear regression with correlated errors

$$ Y = X\beta + \epsilon, \quad \epsilon \equiv (\epsilon(s_1), \ldots, \epsilon(s_n))' \sim \text{FS-CSN}_n(0, \Sigma, \sigma, \lambda), \quad (11) $$

where $X = (x(s_1)', \ldots, x(s_n)')'$ is a full column rank $n \times p$ matrix, $x(s) \in \mathbb{R}^p$ is a (row) vector of explanatory variables at site $s$, $\beta$ is a vector of regression coefficients, and the parameters $(\Sigma, \sigma, \lambda)$ satisfy the conditions in specified in Definition 3.2.

Following the discussion from Section 3, the parameter $\lambda \in \mathbb{R}$ corresponds to a skewness parameter, while parameters $(\beta, \Sigma, \sigma)$ satisfy $E(Y) = X\beta$ and $\text{Var}(Y) = \sigma^2 \Sigma$. To obtain an identifiable model, besides the previous assumptions, we are going to assume that $\Sigma = (\Sigma_{ij})$ is the correlation matrix of the random vector $Y$.

The random variable $\epsilon(s)$ represents the error associated to the site $s \in \mathcal{D}$. Thus, the matrix $\Sigma$ corresponds to the correlation between the errors. We are going to assume that the errors $\epsilon$ have an exponential correlation structure; that is, given two sites $s_i$ and $s_j$, we have that

$$ \text{Corr}(\epsilon(s_i), \epsilon(s_j)) = \Sigma_{ij} = \exp \left( -\rho |s_i - s_j| \right), \quad (12) $$

where $\rho > 0$ is a correlation parameter and $|\cdot| : \mathbb{R}^2 \rightarrow \mathbb{R}^+ \cup \{0\}$ is a distance function. It is well-known that $\rho$ produces valid correlation matrices.

Specification (12) is an element of a comprehensive class of correlation structures; in general, we could take any correlation structure

$$ \Sigma_{ij} = h \left( |s_i - s_j|, \rho \right) $$

for a positive definite function $h : \mathbb{R}^+ \cup \{0\} \times D \rightarrow [-1, 1]$ defined for an open set $D \subset \mathbb{R}^m$, $m \in \mathbb{N}$, such that it is continuous with respect to $\rho$ and $h(0, \rho) = 1$.

Since $\mathcal{D}$ is finite, the existence of the stochastic process $\{Y(s), s \in \mathcal{D}\}$ is straightforward. To guarantee the existence of the continuous generalization of the random field defined by (11) requires more elements, as we discuss in Section 4.2.
4.2. A continuous version

In the following, we will define a particular spatial stochastic process that is a continuous version of the model defined in (11). Assume that $D \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$, is a finite domain. Consider a stochastic process $Z$ defined over the domain $D$. We are going to assume that the finite dimensional distributions of $Z$ can be modeled as in (11); that is, for any collection of sites $s_1, \ldots, s_n \in D$, we assume that

$$Z(s_1, \ldots, s_n) \equiv (Z(s_1), \ldots, Z(s_n))' \sim \text{FS-CSN}_n(X \beta, \Sigma_n, \sigma, \lambda), \quad (13)$$

where $X = (x(s_1), \ldots, x(s_n))'$ is a full column rank $n \times p$ matrix, $x(s) \in \mathbb{R}^p$ is a (row) vector of explanatory variables at site $s$, $\beta' \in \mathbb{R}^p$ is a (row) vector of regression coefficients, $\lambda \in \mathbb{R}$ controls the skewness, $\sigma > 0$ and $\Sigma_n = (\Sigma_{ij})$ satisfies $\Sigma_{ij} = \exp(-\rho |s_i - s_j|)$. Since the first two moments of $Z(s_1, \ldots, s_n)$ correspond to the data’s first two moments, we can say that $Z$ is a valid spatial model. Besides, notice that the implicit errors model in (13) is stationary.

In general, the formal existence of a given random field is not straightforward. It is necessary to analyze the finite-dimensional distributions of the random field to verify such an existence. Namely, according to Kolmogorov’s Extension Theorem (Billingsley, 1995), the finite-dimensional distributions must satisfy two conditions: symmetry under permutations and marginal consistency. The symmetry property states that the distribution should be invariant under permutations of coordinates of variables. The consistency condition indicates that the marginal distributions should be equal to the higher-dimension distributions after the extra dimensions are integrated out. Mahmoudian (2018) shows that the consistency condition is not satisfied by many of the skew random fields in the literature. Proposition 1 of Mahmoudian (2018) provides a characterization of the marginal consistency property for the skew-normal distributions. In particular, Mahmoudian’s result and Proposition 3.5 imply that the FS-CSN subclass satisfies Kolmogorov’s extension theorem’s marginal consistency property, making the distribution valid in a sense defined by Mahmoudian (2018).

4.3. Some comments on the models’ properties

In the previous spatial models, the skewness parameter does not affect the mean and the variance. This attribute allows for a better estimation of the correlation. The simulation results discussed later empirically confirm this aspect: the model (11) estimates better the correlation than a normal distribution in the likelihood framework.

As mentioned before, the identification of the parameters in the distribution of the errors with the data’s first two moments is an important attribute. In particular, thanks to this, the errors model in (11) and (13) is stationary. Another advantage of having $(\Sigma_n, \sigma)$ as parameters is that it allows the use of typical correlation models and the corresponding tools to analyze dependencies. If $(\Sigma_n, \sigma)$ would not correspond to the correlation matrix, as it happens in some spatial models in the literature, it is unclear how to include the information a variogram could provide. Finally, since $\lambda$ does not affect the mean and the variance, it is possible to introduce skewness without affecting the regression’s interpretation.

Besides, in this case, $X \beta$ represents the mean of the response variable, which is not true for most of the skew-normal regression models in the literature.

5. The profile predictive likelihood

In the study of spatial models, one of the main interests is to predict the value of a response variable at an unobserved site. In this work, we use the profile predictive likelihood (PPL) function for this purpose (Bjørnstad, 1990; Mathiasen, 1979).

Consider a random vector $(Y, Z)$ with pdf $f_\theta(y, z)$, where $\theta$ is an unknown parameter vector and $Y$ and $Z$ are random vectors. Assume that $Y = y$ is the observed data and that our primary goal is to predict the unobserved value $z$ of $Z$. Since $z$ is the “parameter” of interest, $\theta$ can be considered a nuisance parameter. According to Berger and Wolpert’s likelihood principle for
prediction (see Bjornstad, 1990), all the information about \((z, \theta)\) is comprised in the joint likelihood function

\[ l_y(z, \theta) = f_\theta(y, z). \]

Following this principle, a reasonable way to perform prediction for \(z\) is eliminating the nuisance parameter \(\theta\) in the likelihood \(l_y\); that is, to determine a likelihood for \(z\) by removing \(\theta\) from \(l_y\). Those types of likelihoods are named predictive likelihoods.

There are various forms of removing \(\theta\) from \(l_y\), each resulting in different predictive likelihoods; Bjornstad (1990) discusses several of them. In this work, we are interested in the so-called profile predictive likelihood, which follows the same reasoning behind the usual profile likelihood: the profile predictive likelihood \(L_p\) for \((y, z)\) is defined as

\[ L_p(z | y) = l_y(z, \hat{\theta}_z) = \sup_{\theta} f_\theta(y, z). \] (14)

In this context, a natural predictor for the unobserved value of \(Z\) is the value of \(z\) that maximizes (14). This predictor is called the maximum likelihood predictor, and we denote it as \(\hat{z}_{\text{mlp}}\). In Section 7, we employ this estimator to perform prediction on risk data for COVID-19.

Notice that \(L_p\) can be normalized as a function of \(z\) to be a probability density. However, determining the corresponding normalization constant can be complicated. When \(L_p\) is normalized, an alternative predictor for \(Z\) is given by the mean of \(L_p\). This predictor is called the predictive expectation of \(Z\) and is denoted by \(E_p(Z)\). Assuming that \(L_p\) is normalized, Bjornstad (1990) suggests the construction of predictive regions \(P_p\) as

\[ P_p = \{z : L_p(z | y) \geq k_\alpha\}, \] (15)

where \(k_\alpha\) is determined such that

\[ \int_{P_p} L(z | y) dz = 1 - \alpha. \] (16)

Kreutz et al. (2013) propose to construct prediction intervals for the profile predictive likelihood using Wilks’ Theorem and statistic. Following such approach, one can construct prediction regions through

\[ \left\{z_0 : \sup_{z, \theta} \left[ \log(l_y(z, \theta)) - \log(L_p(z_0, \theta)) \right] \leq \chi^2_{p, 1-\alpha} \right\}, \] (17)

where \(p\) is the dimension of \(z\). Another form to construct prediction regions is given by conditioning the joint distribution \(f_\theta(y, z)\). Namely, if \((\hat{\theta}, \hat{z}_{\text{mlp}}) = \arg \max_{(\theta, z)} f_\theta(y, z)\) and \(L_p\) is substituted by \(f_\theta(z | y)\) in (15)–(16), \(P_p\) constitutes a reasonable prediction region.

In general, the previous approaches to determine predictive regions require considerable computational resources for dimensions higher than one. Besides, visualizing and interpreting regions for dimensions higher than three is implausible. In the present article, to expedite computations and visualize regions, we only consider prediction intervals.

In this work our interest resides in the case where \(f_\theta\) is the pdf of the FS-CSN distribution of model (11). In that instance, the parameters in the distribution are effectively \(\theta = (\beta, \rho, \lambda, \sigma)\). In such case, optimizing (14) needs to be done numerically. Besides, thanks to Proposition 3.7, for the FS-CSN subfamily it is possible to determine confidence regions parametrically using \(f_\theta(z | y)\).

6. A simulation study

This section presents two simulation studies to assess the likelihood estimation in the FS-CSN model (11) when the data is generated from such a model. Our objective is to empirically compare the likelihood estimators of the FS-CSN model (11) and its Gaussian counterpart. For this, we generate \(n = 25,50,100,200,300,400,500\) random sites \(\{s_1, \ldots, s_n\}\), a vector of random covariates for each of the sites, and \(10^3\) repetitions of a response variable following model (11).
In the first study, we simulate the coordinates of the site’s centroids \( \{(X_i, Y_i)\} \) independently and identically distributed, with \( X_i, Y_i \sim \text{Uniform}(0, 5000) \), \( i = 1, \ldots, n \). The vector of random covariates \( \mathbf{W'} = (W_{1i}, \ldots, W_{ni}) \) is generated independently through \( W_{ii} \sim N_i(\mu = 5, \sigma = 2) \), \( i = 1, \ldots, n \). Given \( \{(X_i, Y_i)\} \) and \( \mathbf{W} \), we simulate a response variable \( \mathbf{R} = (R_1, \ldots, R_n) \) as

\[
\mathbf{R} \equiv (R_1, \ldots, R_n)' \sim \text{FS-CSN}([\beta_0 + \beta_1 \mathbf{W'}], \Sigma_n, \sigma, \lambda),
\]

with \( \beta_0 = 10, \beta_1 = 2, \sigma = 10^{0.5} \approx 3.162, \lambda = 2.5, \rho = 10^{-0.3} \approx 0.501 \) and

\[
\Sigma_{ij} = \exp \left(-\rho \left| (X_i, Y_i) - (X_j, Y_j) \right| \right), \quad \Sigma_n = (\Sigma_{ij}),
\]

where \( |\cdot| \) denotes the Euclidean norm. We generate 1000 independent repetitions of \( \mathbf{R} \); we estimate the parameters in (18) and its Gaussian counterpart (i.e., \( \lambda = 0 \)) for each of these repetitions via maximum likelihood. After cleaning the results, we study the performance of the estimators of both models via the bias, variance (Var), and mean square error (MSE) criteria. Table D.1 contains the results for these criteria, including their corresponding Monte Carlo standard errors (MCSE). Table D.2 contains the quartiles and the mean of the estimated parameters.

The second simulation study is generated analogously but with some differences. In this case, the number of centroids is \( n = 50, 100, 200 \), the centroids’ coordinates are distributed Uniform(0, 500), and the skewness parameter is \( \lambda = 7 \). For the second simulation, we also study the performance of the estimators in the FS-CSN and Gaussian models. Table D.3 shows the performance of the estimators in the second study according to the bias, Var, and MSE criteria. Table D.4 presents their means and quartiles.

In Table D.1, we can see that for the first simulation, the parameters \( \beta_0, \beta_1, \) and \( \sigma \) are well-behaved in both models. We can also observe that the bias, Var, and MSE associated with the FS-CSN estimates of \( \lambda \) become smaller when the number of sites increases. The case of \( \rho \) is the most interesting. For \( n \leq 100 \), the normal estimates of \( \rho \) exhibit less bias than the FS-CSN estimates; however, for \( n \geq 200 \), the FS-CSN estimates of \( \rho \) show the smallest bias. Besides, the FS-CSN estimates of \( \rho \) consistently exhibit smaller Var and MSE than in the Gaussian model. Finally, notice that for \( n \geq 100 \), the normal case displays more significant variances; in Table D.2, we can see that this is caused by some unusually large values of \( \rho \) in the normal case. As we explain later, we filtered some extremely large values for \( \rho \) in the Gaussian case; as we can see, this is not enough to control large deviations, that is, extremely large numerical values for estimating this parameter. This situation does not occur, with the same sample, with the FS-CSN distribution.

For the second simulation, Table D.3 shows that the parameters \( \beta_0, \beta_1, \) and \( \sigma \) are again well-behaved in both models; the same is true for the parameter \( \lambda \) in the FS-CSN model. In the second simulation, for all \( n \), the FS-CSN model estimates better \( \rho \) in all the metrics. As corroborated by Table D.4, this is partly explained by some unexpectedly large estimates for \( \rho \) in the normal case. The different behavior for the estimates of \( \rho \) among simulations might be caused by the difference in the average distances between sites. In general, it seems that under skewness, the FS-CSN model is better for estimating the correlation parameter \( \rho \), particularly for short distances or more than 200 sites. However, further investigation is required to confirm this affirmation. In the application, we see a similar result.

The simulations and the numerical optimization of the log-likelihood functions are performed in R. The FS-CSN distribution simulation is based on the marginal representation (2). The log-likelihood function’s optimization is done with the package \texttt{optimx} (Nash and Varadhan, 2011; Nash, 2014). It is also possible to employ R’s package \texttt{csn} to take advantage of some of its functions (Pavlyuk and Girtcius, 2015).

For cleaning the data, we check the optimality of the numerical estimators via two subgradient optimality conditions. The inspection of these conditions is performed by the optimization function of the \texttt{optimx} package. The first condition (kkt1) indicates that the solution reported has an approximately null gradient; the second condition (kkt2) indicates that the Hessian is positive-definite. In the normal case, it is worth noting that condition kkt2 was never satisfied for the default settings; consequently, we only employ condition kkt1 to determine convergence to an optimal in this case. Besides, also in the normal case, we eliminate some estimations that produced extremely large values of \( \rho \).
7. Application: Predicting the values of a heuristic COVID-19 mortality risk index

In this section, the spatial model described in Section 4 is applied to unveil the spatial structure in a COVID-19 mortality risk index. The data employed in the study comprises information at the municipal level about Mexico’s northeastern states and consists of data relevant for the surveillance of the COVID-19 pandemic. Our main objective is to show the use of the FS-CSN subclass in an interesting and current example and to contrast the results produced by the distribution with those obtained by its Gaussian counterpart. A second objective is to understand how the severity of the COVID-19 pandemic is related to space and a set of covariables. The results obtained are discussed, and some relevant figures are presented. We compare the quality of the fitted models against each other and the observed data, notably their PPL-predictions.

7.1. About the data and the mortality risk index

The data considered for this study consist of information at the municipal level about the COVID-19 cases and deaths, the gross domestic product (GDP), the prevalence of the main COVID-19 comorbidities, the healthcare services available, the total population, and the population’s percentage that is older than 65 years of age. The data about comorbidities contain the prevalence of diabetes, heart diseases, cancer, and lung diseases, while the data about available healthcare services consist of the total number of doctor’s offices, pharmacies, public and private hospitals, and hospital beds.

Most of this data is publicly available, although not necessarily located in the same repositories. The only exception is the GDP data; however, the Mexican National Institute of Statistics and Geography (INEGI) commented that this information would be made public at a later date. Besides, the study’s data only consist of information about Mexico’s northeastern states collected up to November 26th, 2020. In total, there are 295 municipalities in these states. The data was shared through the COVID-19 surveillance system that the Mexican federal government implemented.

From the data about COVID-19 cases and deaths, it is derived a heuristic index that intends to compare the risk of death due to COVID-19 that the municipalities are exposed to at a particular time. Specifically, the risk index $R_i$ at municipality $i$ is defined as

$$R_i = \sigma_P^{-1} (P_i - \mu_P),$$

where

$$P_i = \ln\left(\frac{(\text{Total population in } i)^2}{\text{Total COVID-19 deaths reported in } i}\right),$$

and $\mu_P$ and $\sigma_P$ represent the mean and the standard deviation of all the quantities $P_i$ in the municipalities in the studied region. The standardization is considered for improving the index’s ability of comparison. The values of this index approximately lie in the interval $(−2.25, 2.75)$. In this context, higher positive values mean higher risk. The risk index also provides a crude measure to compare the resources required to attend the COVID-19 pandemic at the municipal level. In Appendix C, we explain the motivation of this risk index.

Since the index relies on reported deaths, places with no deaths are not considered. Besides, as we have concerns about the data’s reliability at places with few inhabitants, municipalities with less than 3000 inhabitants are not considered. There are 45 municipalities with less than 3000 inhabitants, from which 21 reported COVID-19 deaths by the time the data was collected. The municipalities with reported COVID-19 losses are distributed in the following manner: 11 municipalities have one reported death, 5 municipalities have two deaths, 2 municipalities report three deaths, 2 municipalities count four deaths, and one municipality reports five deaths. It is relevant to mention that we fitted a complete data model including all the municipalities with reported deaths. The results obtained were good; however, our concerns about the data made us decide not consider municipalities with very few people. Such restriction on the data induces skewness in it. Thus, only $n = 224$ municipalities are finally considered. For improving the fitting
of the model, we also considered the Mexican municipalities bordering the studied region; however, to report the results, we only use the 224 municipalities.

The data described above is complemented with the coordinates of the geographic centroids for the municipalities. The Mexican COVID-19 surveillance system also provided this data.

7.2. Modeling the risk index

Naturally, one can expect that the spread of an infectious disease and its consequences are associated with the movement within and between populations. Thus, a reasonable assumption is that the municipality’s risk index is related to the municipalities located nearby. Besides, it is possible to argue that the risk of dying of COVID-19 is somehow related to age, access to good healthcare services, and the comorbidities suffered. Thus, we decide to model the risk index \( R_i \) at municipality \( i \) through model (11), employing as covariates the data described in the previous section.

Looking at the structure in the data, we decided to make some further preprocessing to exploit the available information effectively. Essentially, we used principal component analysis (PCA) to reduce the dimensions in the healthcare services information and the comorbidities data. Due to the heterogeneity among states, the PCA-components are computed for every state. We used the first two components in both datasets; these two PCA-components explain, in both cases, more than 90% of the variance at all the states. The sparse structure with many zeros in the healthcare services dataset reinforced the convenience of applying PCA.

Thus, we model the risk index \( R_i \) at the municipality \( i \) with the FS-CSN regression with correlated errors given by

\[
R_i = \beta_0 + \beta_1 G_i + \beta_2 C_{i1} + \beta_3 C_{i2} + \beta_4 O_i + \beta_5 S_{i1} + \beta_6 S_{i2} + \epsilon_i \quad i = 1, \ldots, n, \tag{19}
\]

where \( G_i \) corresponds to the GDP for \( i \), \( O_i \) is the percentage of the population in \( i \) that is older than 60 years, \((S_{i1}, S_{i2})\) and \((C_{i1}, C_{i2})\) are, respectively, the first two PCA-components for the healthcare services and comorbidities data, \((\epsilon_1, \ldots, \epsilon_n) \sim \text{FS-CSN}_n(0, \Sigma_n, \sigma, \lambda)\) and

\[
\text{Corr}(\epsilon_i, \epsilon_j) = \Sigma_{ij} = C(d_G(C_i, C_j)), \quad C(t) = \begin{cases} \sigma^2 \exp(-\rho \cdot t) & \text{if } t > 0, \\ \kappa^2 + \sigma^2 & \text{if } t = 0 \end{cases}, \tag{20}
\]

where \( C_i \) denotes the geographical centroid of the municipality \( i \), and \( d_G \) corresponds to the geodesic distance measured in hundreds of kilometers. The quantities \( G_i, C_{ij}, O_i, S_{ij} \) are taken standardized. Notice that model (19) is not exactly a particular case of the model (11). Specification (20), where we allow \( C(0) \neq \sigma^2 \), indicates that \( \Sigma_n \) is not a correlation matrix. This small change is not very significant for the results stated in Section 3, but it adds flexibility to the modeling framework (11) by allowing for discontinuities at the origin of the covariance function. This discontinuity is known as a nugget effect.

For comparison, we also consider the Gaussian counterpart of the model (19). That is, we consider the version of the model (19) with \( \lambda = 0 \) fixed. Both versions of the model (19) are fitted using likelihood estimation. The models are optimized numerically in \( R \) using the log-likelihood function and the method ‘nlminb’ implemented in the package optimx (Nash and Varadhan, 2011; Nash, 2014). To ensure that a global maximum is obtained, different sensible initial values are tried. The numerical optimization has a nice behavior, converging almost always to the same parameter values for all the initial conditions.

To evaluate the models’ performance and the quality of their PPL-predictions, we conduct 10-fold cross-validation on the model (19) and its Gaussian counterpart. Table D.5 contains the cross-validation means and quartiles of the parameters estimated for both models. The value \( \lambda > 0 \) is expected in the FS-CSN model because the data is not symmetric and because we dropped municipalities with less than 3000 inhabitants. In the FS-CSN model, the mean value estimates \( \sigma = 0.553, \kappa = 0.344 \) indicate that the variance of the components is approximately 0.402, which is approximately the same for the normal model (0.465). Despite the standardization of the covariables, some of the regression coefficients are small in both models; though, the inclusion
of those variables improves a little bit the predictions. In this regard, it is relevant to notice that the most significant covariables for modeling the risk index are the GDP and proportion of the population over 65 years old. The variables in the model that measure the access to local healthcare services and the community's health are much less significant. A plausible interpretation is that people in small communities move to bigger towns and cities to treat their illnesses if they have enough money, making less important the local healthcare services; and that the mentioned comorbidities are evenly present in the northeast of Mexico that their effect becomes less significative in the risk index.

The mean value for the correlation parameter \( \rho = 2.052 \) in the FS-CSN model indicates that the effective range (the distance where the variogram reaches 95% of its maximum) is around 146 km, which differs from the approximately 94 km indicated by the range estimated by adjusting an exponential variogram model to the risk data’s sample variogram. In the normal model, the mean value for the correlation parameter \( \rho = 25.219 \) induces a more extreme difference. That value indicates an effective range of around 12 km. An extreme value in the estimation of \( \rho \) in the normal model causes the large value for the mean of \( \rho \); this instability also appears in the simulations where the normal model estimated extreme values for \( \rho \). However, it is worth mentioning that, since the estimate for \( \lambda \) is relatively small, one could expect that the estimates for \( \rho \) be, in general, similar in both models; the values in Table D.5 corroborate this. Fig. 1 presents an exponential variogram model fitting to the risk data’s sample variogram computed using R’s package gstat (Pebesma, 2004; Gräler et al., 2016). The plot suggests that the correlation model with exponential decay is appropriate to describe the risk index’s correlations. The nugget and the effective range adjusted to the variogram model are 0.434 and approximately 94 km, respectively. Fig. 1 is very relevant since it shows that a model with exponential decay is appropriate to describe the correlation between two municipalities’ risk indices.

Table D.6 contains the mean absolute error (MAE), and the root mean squared error (RMSE) for the FS-CSN and normal models in each cross-validation fold. It also presents the same errors for the PPL-predictions of both models and the mean length of the prediction intervals (MLPI). The prediction intervals are based on (17). For the FS-CSN model fitting, the averages of the MAE and RMSE errors are 0.416 and 0.612, respectively, while the averages of the MAE and RMSE errors in the normal model are 0.420 and 0.615, respectively. The averages of the MAE and RMSE errors for the prediction in the FS-CSN model are 0.460 and 0.575, respectively; the averages of the MAE and RMSE errors for the prediction in the Gaussian model are 0.463 and 0.570, respectively. The average of the MLPI is 2.159 for the FS-CSN model and 2.226 for the normal model. These results suggest that the performance of both models is in average equivalent.
Fig. 2 shows maps with the absolute prediction error for both models and each municipality in Mexico’s northeast. The municipalities that appear in gray correspond to municipalities with no deaths reported (up to November 26th, 2020) or less than 3000 inhabitants. To produce these maps, the risk index $R_i$ is divided by 2.75 to obtain values between $(-1, 1)$ and have a more intuitive visualization. Thus, the absolute prediction error takes values between $[0, \infty)$; and, a value of 1 indicates an error of 2.7 between the prediction and the observation. It is worth clarifying that all the errors are smaller than 1. The left map shows the absolute prediction error for the FS-CSN model, while the map on the right shows the absolute prediction error for the normal model. In both maps, most of the municipalities are in black, which indicates that the models show a reasonable adjustment, reflecting that the MAE and the RMSE errors are small. The maps also visually indicate that both models have a similar performance. A larger $\lambda$ would produce a more substantial difference between the FS-CSN and Gaussian model.

For producing the maps of Fig. 2, we do not include the information of two municipalities: Mezquital, Dgo. and Guadalupe y Calvo (GyC), Chih. Mezquital and GyC are isolated and impoverished municipalities with considerable ethnic populations; besides, GyC has a public security problem due to organized crime. These conditions make them implausible to be described under model (19). It is relevant to mention that Mezquital and GyC’s information is employed in the rest of the estimation and prediction process for both models.

Until this point, the results show that both models – the FS-CSN and Gaussian models – have a very similar performance in prediction. It remains unclear which is the better model. Since the Gaussian model is a particular case of the FS-CSN model, one way to answer this question is to test, within the FS-CSN subclass, the hypotheses $H_0 : \lambda = 0$ vs $H_1 : \lambda \neq 0$. For this task, it is possible to use the generalized likelihood ratio test (GLRT) and the approximation given by Wilks’ theorem. For each fold, the $p$-value produced by the GLRT is smaller than $5 \cdot 10^{-3}$. Thus, after a Bonferroni correction, the null hypothesis $H_0$ is rejected with a 95% confidence in all folds. Thus, we have strong evidence that the FS-CSN model is more appropriate than the Gaussian counterpart.

The results described above indicate that both models seem appropriate to describe the risk data up to a certain degree. However, since there is evidence that the skewed model is more appropriate than the Gaussian model to adjust the data and because, according to the simulation experiments, the skewed model seems to estimate better the correlation parameter, we conclude that the FS-CSN is a better option than the Gaussian model for the modeling this particular data. On the other hand, the FS-CSN contains the normal case, it could be that the skewness is not that significant, but since the normal distribution belongs to the FS-CSN subclass, we think that this particular class is robust to work in both cases, with skewness or without skewness. Thus, model (19) provides a tool supporting the decision-making and the management of the pandemic. According to the
above discussion, the GDP and the proportion of people older than 65 are the value that affects the risk index the most. The effect of these two variables is inverse; that is, their signs are different, which indicates that they somehow compensate for each other. The other variables seem to have a marginal effect; however, they reduce the prediction error. They correspond to the health of the population and the available local health services. These results are in accordance with patterns found in other studies around the world.

8. Conclusion

This work proposes the FS-CSN subclass as an option for spatial modeling and prediction. As we showed, the FS-CSN subclass shares several important features with the Gaussian distribution. Specifically, the FS-CSN is identifiable, closed under marginalization and conditioning, and a null correlation implies independence. Nevertheless, the class has some advantages over the usual Gaussian model; in particular, it allows higher skewness and kurtosis and seems to estimate the correlation structure under skewness better.

The FS-CSN distribution exhibits good attributes for its practical use on linear models with a spatial covariance structure. For instance, a significant attribute of the FS-CSN spatial model introduced here is its parameterization in terms of the four interpretable parameters \((\beta, \Sigma_n, \sigma, \lambda)\). The parameter \(\beta\) is the regression vector, and it provides the model’s mean through \(X\beta\). If the model’s mean is not \(X\beta\), the validity (interpretation) of the model is unclear. The parameters \((\Sigma_n, \sigma)\) represent the correlation and the variance of the data, respectively; thanks to this, the respective errors model is stationary. Another advantage of having as parameters to \((\Sigma_n, \sigma)\) is that it allows the use of typical correlation models and the corresponding tools to analyze dependencies. If \(\Sigma_n\) would not correspond to the correlation matrix, as it happens in some spatial models in the literature, it is unclear how to include the information a variogram could provide. The parameter \(\lambda\) controls the skewness and kurtosis exhibited by the model without affecting the other parameters, making it possible to introduce skewness without affecting the regression’s interpretation. Thanks to all these properties and the identifiability of the FS-CSN distribution, we could say that the proposed spatial model is valid.

On the other hand, for the FS-CSN model, the closure and form of the marginal distribution imply the model’s marginal consistency as in the Kolmogorov extension theorem, while the closure of the conditional distributions allows determining confidence regions parametrically. The property that null correlation implies independence assists in the interpretations of results and modeling. Besides, the FS-CSN allows determining the exact distribution of the residuals of the spatial models with FS-CSN errors.

The FS-CSN subclass is similar to one discussed by Mahmoudian (2018); however, it differs from our proposal in that we use Cholesky decomposition instead of the square root matrix and that, in our parameterization, the mean and the covariance matrix are the parameters of the subclass. These changes cause differences in the properties of the distribution and the spatial models that it produces.

We finalize with an application related to a risk index for COVID-19. The spatial model proposed in this work produces good forecasts for the risk index at the municipal level; the model also displays the ability to recognize outliers caused by entirely different variables that dominate those particular sites. The presentation of this application intends to illustrate, in an interesting and current situation, the use of the FS-CSN distribution in spatial modeling and prediction and compare it with its Gaussian counterpart. However, the main objective of the present work is to introduce the FS-CSN class and some of its main properties. To conclude, we remark that in all the numerical examples studied in the article, the estimation and the prediction in the FS-CSN subclass are pretty satisfactory, including the simulation study results.

During the process of this work, we realized that some aspects need to be addressed in a more detailed form in a future investigation. Perhaps, one of the most interesting is to determine proper ways to deal with the outliers in spatial modeling for skewed distributions. In general, it is unclear how to differentiate between observations within the model and outliers when the skewness is concentrated on one side of the distribution. It would be helpful to determine when an observation
is atypical. There are several references that study this problem, e.g., (Kou et al., 2007; Yamanishi et al., 2004; Ramaswamy et al., 2000). However, most of them seem to suggest that the best strategy is to remove the possible outliers from the modeling, a proposal which in our opinion, may be inadequate in the context of skewed distributions.

Acknowledgments

We acknowledge Armando Domínguez (UAS) for some early discussions that lead to the flexible subfamily’s current form. We also thank Pablo López (CentroGEO) and Raúl Sierra (CONABIO) for sharing some of the data that allowed the analysis of the COVID-19 mortality risk index. We appreciate Israel Martínez’s (Lancaster University) valuable help in manipulating the server where the simulations were run. We acknowledge Oscar Sánchez-Siordia (CentroGEO) for some comments and discussions regarding the risk index. We are also very thankful to Miguel Bedolla (Alamo Colleges NVC) for his careful reading and comments to improve this work. Graciela González-Farías thanks the partial support through the project CONACYT-CBA1-S-43514.

Appendix A. Proofs

In this Appendix, we present the proofs of the results stated in Section 2.

**Proof Proposition 3.4.** The formulas can be obtained by differentiating the corresponding moment generating function and making some lengthy algebraic manipulations. □

**Proof Proposition 3.5.** Notice that the Cholesky factor $\Sigma_n^c$ can be partitioned into

$$
\Sigma_n^c = \begin{pmatrix}
\tilde{\Sigma}_{11}^c & 0 \\
\tilde{\Sigma}_{21}^c & \tilde{\Sigma}_{22}^c
\end{pmatrix},
$$

(A.1)

The submatrices in (A.1) not necessarily correspond to the Cholesky factors of the submatrices of $\Sigma_n$; however, as we note in the following, $\tilde{\Sigma}_{11}^c$ indeed does correspond to the Cholesky factor of $\Sigma_{11}^c$. Thanks to partition (10), we can write

$$
\begin{pmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{pmatrix} = \Sigma_n = \Sigma_n^c(\Sigma_n^c)' = \begin{pmatrix}
\tilde{\Sigma}_{11}^c(\tilde{\Sigma}_{11}^c)' & \tilde{\Sigma}_{12}^c(\tilde{\Sigma}_{11}^c)' \\
\tilde{\Sigma}_{21}^c(\tilde{\Sigma}_{11}^c)' & \tilde{\Sigma}_{22}^c(\tilde{\Sigma}_{11}^c)'
\end{pmatrix},
$$

which implies that $\Sigma_{11} = \tilde{\Sigma}_{11}^c(\tilde{\Sigma}_{11}^c)'$. Since $\Sigma_{11}$ is positive definite and $\tilde{\Sigma}_{11}^c$ is a lower triangular matrix with positive elements in the diagonal, the uniqueness of the Cholesky factor $\Sigma_{11}^c$ implies that necessarily $\tilde{\Sigma}_{11}^c = \Sigma_{11}^c$.

Let $Y$ be defined as in (5) and let $Y$ be partitioned into $Y = (Y_1', Y_2')'$, where the dimensions of $Y_1$ and $Y_2$ are $k$ and $n-k$, respectively. Since $\Sigma_n^c$ is a lower triangular matrix, the first $k$ coordinates of $\Sigma_n^c Y$ correspond to the coordinates in $\tilde{\Sigma}_{11}^c Y_1 = \Sigma_{11}^c Y_1$. Then, we can write

$$
W_1 \overset{d}{=} \mu_1 + \sigma \Sigma_{11}^c Y_1.
$$

Considering that $\Sigma_{11}^c$ is positive definite (positive elements in the diagonal), relation (4) implies that $W_1$ is distributed according to

$$
W_1 \sim \text{CSN}_{k,k}(\mu_1 - b\delta\tau \Sigma_{11}^c 1_k, \sigma^2\tau^2 \Sigma_{11}^c, \frac{\lambda}{\sigma\tau} \Sigma_{11}^c, 0, I_k).
$$

Since $\Sigma_{11}^c$ is the Cholesky factor of $\Sigma_{11}$, this implies the desired result. □

We require the following auxiliary lemma to prove Proposition 3.7.
Lemma A.1. Let $\Sigma_n$ be a positive definite matrix. Assume that $\Sigma_n$ is partitioned as in Proposition 3.5. Then, the matrix $\Sigma_n^{-c}$ can be partitioned into

$$
\Sigma_n^{-c} = \begin{pmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{pmatrix},
$$

(\text{A.2})

with $P_{11}$ a $k \times k$ matrix. Besides, we have that necessarily $P_{11} = \Sigma_{11}^{-c}$, $P_{22} = \Sigma_{22}^{-c}$, and $P_{21} = -\Sigma_{21}^{-c} \Sigma_{21}^{-1} \Sigma_{11}^{-1}$, where $\Sigma_{21} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$.

**Proof.** Assume that $\Sigma_n^{-c}$ is partitioned as in Eq. (A.1). The triangular partition (A.2) is appropriate because the inverse matrix of a lower triangular matrix is also a lower triangular matrix. This proves the first part of this lemma.

On the other hand, since $\Sigma_n^{-1}$ denotes the inverse of $\Sigma_n$, we have that $\Sigma_n^{-1} \Sigma_n^{-c} = I_n = \Sigma_n^{-c} \Sigma_n^{-1}$. Expanding by blocks the matrix multiplications $\Sigma_n^{-c} \Sigma_n^{-c}$ and $\Sigma_n^{-1} \Sigma_n^{-c}$ and equating the upper right blocks, we obtain that $P_{11} \Sigma_{11}^{-1} = I_k = \Sigma_{11}^{-1} P_{11}$, which implies that $P_{11} = \Sigma_{11}^{-1}$. In the proof of Proposition 3.5, we show that $\Sigma_{11}^{-1} = \Sigma_{11}^{-c}$. Therefore, $P_{11} = \Sigma_{11}^{-c} = \Sigma_{11}^{-1}$.

Now, we are going to prove that $P_{22} = \Sigma_{22}^{-c}$. Applying the block matrix inversion formula to $\Sigma_n$, we obtain that the lower right block of $\Sigma_n^{-1}$ is given by $(\Sigma_{21})^{-1}$. Since $\Sigma_n^{-1} = (\Sigma_n^{-c})^{-1} = (\Sigma_n^{-c} \Sigma_n^{-1})^{-1} = (\Sigma_n^{-c} \Sigma_n^{-c})^{-1}$, expanding by blocks the matrix product $(\Sigma_n^{-c} \Sigma_n^{-c})^{-1}$, we find that necessarily $P_{22}^{-1} P_{22} = (\Sigma_{22}^{-c} \Sigma_{22}^{-c})^{-1}$. Therefore, $P_{22}^{-1} P_{22} = \Sigma_{22}^{-c}$. Notice that $P_{22}^{-1}$ is also inverse triangular. Also, because $\Sigma_n^{-1}$ has positive diagonal elements, $\Sigma_n^{-c}$ has positive diagonal elements; this implies that $P_{22}^{-1}$ and $P_{22}^{-1}$ also have positive diagonal elements. Since $P_{22}^{-1}$ is lower triangular with positive elements in the diagonal and $P_{22}^{-1} (P_{22}^{-1})' = \Sigma_{22}^{-1}$, necessarily $P_{22} = \Sigma_{22}^{-c}$. This implies that $P_{22} = \Sigma_{22}^{-1}$.

It only remains to prove that $P_{21} = -\Sigma_{21}^{-c} \Sigma_{21}^{-1}$. Again expanding by blocks both sides of the equation $\Sigma_n^{-1} = (\Sigma_n^{-c} \Sigma_n^{-1})^{-1}$ and equating the upper right blocks, we obtain that $P_{22}^{-1} P_{21} = -\Sigma_{21}^{-c} \Sigma_{21}^{-1} \Sigma_{21}^{-1}$. Since we have already proved that $P_{22} = \Sigma_{22}^{-c}$, we have

$$
(\Sigma_{21}^{-1})' P_{22}^{-1} P_{21} = -\Sigma_{21}^{-c} \Sigma_{21}^{-1} \Sigma_{21}^{-1} = - (\Sigma_{21}^{-c} \Sigma_{21}^{-1})' \Sigma_{21}^{-1} = \Sigma_{21}^{-1} \Sigma_{21}^{-1},
$$

which implies that $P_{21} = -\Sigma_{21}^{-c} \Sigma_{21}^{-1}$. This concludes the proof. □

**Proof Proposition 3.7.** Let $z = (w_1, w_2) \in \mathbb{R}^n$, with $w_1 \in \mathbb{R}^k$ and $w_2 \in \mathbb{R}^{n-k}$. We prove the desired result by computing $f_{W_2|w_1} = f_{z_2|z_1}$. Assume that $\Sigma_n^{-c}$ is partitioned as in (A.2), with $P_{11}$ a $k \times k$ matrix. Thanks to Lemma A.1, we know that this triangular partition is appropriate and that $P_{11} = \Sigma_{11}^{-c}$, $P_{22} = \Sigma_{22}^{-c}$ and $P_{21} = -\Sigma_{21}^{-c} \Sigma_{21}^{-1}$. Notice that $(z - \mu - b\delta \sigma \tau \Sigma_n^{-1} 1_n) = ((w_1 - v_1), (w_2 - v_2))'$ and that

$$
\Sigma_n^{-c} (z - \mu - b\delta \sigma \tau \Sigma_n^{-1} 1_n) = ((P_{11}(w_1 - v_1))', (P_{21}(w_1 - v_1) + P_{22}(w_2 - v_2)))'.
$$

Using these equations, the properties of independence and conditioning of the multivariate normal distribution imply that we can write

$$
\frac{\phi_n(\frac{\lambda}{\sigma \tau} \Sigma_n^{-c} (z - \mu - b\delta \sigma \tau \Sigma_n^{-1} 1_n); 0, I_n)}{\phi_k(\frac{\lambda}{\sigma \tau} \Sigma_n^{-1} (w_1 - v_1); 0, I_k)} = \frac{\phi_n(\lambda \Sigma_{21}^{-1} (w_2 - \mu)_{21}; 0, I_{n-k})}{\phi_k(\frac{\lambda}{\sigma \tau} \Sigma_{11}^{-1} (w_1 - v_1); 0, I_k)}
$$

(\text{A.3})

and

$$
\frac{\phi_n(z; \mu - b\delta \sigma \tau \Sigma_n^{-1} 1_n, \sigma^2 \tau^2 \Sigma_n)}{\phi_n(w_1; v_1, \sigma^2 \tau^2 \Sigma_{11})} = \frac{\phi_n(w_2; 2\mu_{21}, \sigma^2 \tau^2 \Sigma_{21})}{\phi_n(w_1; v_1, \sigma^2 \tau^2 \Sigma_{11})}.
$$

(\text{A.4})

Thus, thanks to Proposition 3.5 and Eqs. (A.3) and (A.4), we have that

$$
f_{W_2|w_1}(w_2|w_1) = \frac{f_{z_2|z_1}(w_1, w_2)}{f_{z_1}(w_1)}
$$

18
\[ 2^{-n-k} \phi_n(\mathbf{z}; \mu - b\delta \sigma \tau \Sigma_n^c \mathbf{1}_n, \sigma^2 \tau^2 \Sigma_n) \Phi_n \left( \frac{\lambda}{\sigma \tau} \Sigma_n^c (\mathbf{z} - \mu - b\delta \sigma \tau \Sigma_n^c \mathbf{1}_n); \mathbf{0}, I_n \right) \]

\[ = 2^{-n-k} \phi_{n-k}(\mathbf{w}_2; \mu_{2|1}, \sigma^2 \tau^2 \Sigma_{2|1}) \Phi_{n-k} \left( \frac{\lambda}{\sigma \tau} \Sigma_{2|1}^c (\mathbf{w}_2 - \mu_{2|1}); \mathbf{0}, I_{n-k} \right). \]

Therefore, \( \mathbf{W}_2 | \mathbf{W}_1 = \mathbf{w}_1 \sim \text{FS-CSN}_{n-k}(\mu_{2|1}, \Sigma_{2|1}, \sigma, \lambda). \)

**Proof Proposition 3.8.** We only prove the sufficient condition; the necessary condition is a well-known fact. Our proof is based on the moment generating function. Assume that \( \text{Cov}(\mathbf{W}_1, \mathbf{W}_2) = \Sigma_{12} = \Sigma_{21} = 0. \) Let \( \mathbf{t} \equiv (t_1', t_2')' \in \mathbb{R}^n, \) where \( t_1 \in \mathbb{R}^k \) and \( t_2 \in \mathbb{R}^{n-k}. \) Notice that \( \Sigma_{12} = \Sigma_{21} = 0 \) implies that the Cholesky factor \( \Sigma_n^c \) satisfies

\[ \Sigma_n^c = \begin{pmatrix} \Sigma_{11}^c & 0 \\ 0 & \Sigma_{22}^c \end{pmatrix}, \]

where \( \Sigma_{11}^c \) and \( \Sigma_{22}^c \) are the Cholesky factors associated with \( \Sigma_{11} \) and \( \Sigma_{22}, \) respectively. Thus,

\[ \phi_n(\lambda \sigma \tau (\Sigma_n^c)' \mathbf{t}; \mathbf{0}, (1 + \lambda^2) I_n) = \phi_n(\lambda \sigma \tau (t_1', t_2')' \Sigma_{11}^c, t_2'; \mathbf{0}, (1 + \lambda^2) I_k) \]

and

\[ t' \Sigma_n^c \mathbf{1}_n = (t_1', t_2') \begin{pmatrix} \Sigma_{11}^c & 0 \\ 0 & \Sigma_{22}^c \end{pmatrix} \begin{pmatrix} 1_k \\ 1_{n-k} \end{pmatrix} = t_1' \Sigma_{11}^c \mathbf{1}_k + t_2' \Sigma_{22}^c \mathbf{1}_{n-k}. \]

Therefore, since \( \Sigma_{12} = \Sigma_{21} = 0, \) the mgf of \( Z_n \) is given by

\[ M_{Z_n}(\mathbf{t}) = 2^k \phi_n(\lambda \sigma \tau (\Sigma_{11}^c)' \mathbf{t}_1; \mathbf{0}, (1 + \lambda^2) I_k) e^{t'_1 \mu_1 + b\delta \sigma \tau t_1' \Sigma_{11}^c \mathbf{1}_k + \frac{1}{2} t'_1 \Sigma_{11}^c t_1} \]

\[ \times 2^{n-k} \phi_{n-k}(\lambda \sigma \tau (\Sigma_{22}^c)' \mathbf{t}_2; \mathbf{0}, (1 + \lambda^2) I_{n-k}) e^{t'_2 \mu_{n-k} + b\delta \sigma \tau t_2' \Sigma_{22}^c \mathbf{1}_{n-k} + \frac{1}{2} t'_2 \Sigma_{22}^c t_2}. \]

On the other hand, Eq. (9) and Proposition 3.5 imply that the moment generating functions of \( \mathbf{W}_1 \) and \( \mathbf{W}_2 \) are given by

\[ M_{\mathbf{W}_1}(\mathbf{t}_1) = 2^k \phi_n(\lambda \sigma \tau (\Sigma_{11}^c)' \mathbf{t}_1; \mathbf{0}, (1 + \lambda^2) I_k) e^{t'_1 \mu_1 + b\delta \sigma \tau t_1' \Sigma_{11}^c \mathbf{1}_k + \frac{1}{2} t'_1 \Sigma_{11}^c t_1}, \]

\[ M_{\mathbf{W}_2}(\mathbf{t}_2) = 2^{n-k} \phi_{n-k}(\lambda \sigma \tau (\Sigma_{22}^c)' \mathbf{t}_2; \mathbf{0}, (1 + \lambda^2) I_{n-k}) e^{t'_2 \mu_{n-k} + b\delta \sigma \tau t_2' \Sigma_{22}^c \mathbf{1}_{n-k} + \frac{1}{2} t'_2 \Sigma_{22}^c t_2}. \]

Using this information in Eq. (A.5), we obtain that \( M_{Z_n}(\mathbf{t}) = M_{\mathbf{W}_1}(\mathbf{t}_1) M_{\mathbf{W}_2}(\mathbf{t}_2), \) which implies that \( \mathbf{W}_1 \) is independent of \( \mathbf{W}_2. \) This concludes the proof. \( \square \)

**Proof Proposition 3.9.** As we have noted before, if \( \Sigma_Y^{-1} \) is replaced by the pseudoinverse \( \Sigma_Y^+, \) formula (4) remains valid for any arbitrary matrix \( \mathbf{A}. \) By definition, \( Z_n \) follows a CSN distribution with

\[ \mu_{Z_n} = \mu - b\delta \sigma \tau \Sigma_n^c \mathbf{1}_n, \quad \Sigma_{Z_n} = \sigma^2 \tau^2 \Sigma_n, \quad D_{Z_n} = \frac{\lambda}{\sigma \tau} \Sigma_n^{-c}, \quad v_{Z_n} = \mathbf{0}, \quad \Delta_{Z_n} = \mathbf{I}_n. \]

Thus, the relation (4) implies

\[ \mu_{\mathbf{W}} = B \mu + \mathbf{e} - b\delta \sigma \tau B \Sigma_n^c \mathbf{1}_n, \quad \Sigma_{\mathbf{W}} = \sigma^2 \tau^2 B \Sigma_n B', \quad v_{\mathbf{W}} = \mathbf{0}, \]

\[ D_{\mathbf{W}} = \lambda \sigma \tau \Sigma_n^{-c} \Sigma_n B' \Sigma_n^+ = \lambda \sigma \tau (B \Sigma_n^c)' \Sigma_n^+, \]

\[ \Delta_{\mathbf{W}} = \mathbf{I}_n + \lambda^2 \mathbf{I}_n - \lambda^2 \sigma^2 \tau^2 (B \Sigma_n^c)' \Sigma_n^+ B \Sigma_n (\Sigma_n^c)' = (1 + \lambda^2) \mathbf{I}_n - \lambda^2 \sigma^2 \tau^2 (B \Sigma_n^c)' \Sigma_n^+ B \Sigma_n^c. \]

From the properties of the pseudoinverse, we obtain that

\[ \sigma^2 \tau^2 \Sigma_n^+ = (B \Sigma_n B')^+ = (B \Sigma_n^c (\Sigma_n^c)' B')^+ = (B \Sigma_n^c)' (B \Sigma_n^c)' = ((B \Sigma_n^c)')^+ (B \Sigma_n^c)^+. \]
J.U. Márquez-Urbina and G. González-Farías

Spatial Statistics 47 (2022) 100556

and

$$D_W = \lambda \sigma \tau (B' \Sigma_n^{C})' \Sigma_W^{+} = \frac{\lambda}{\sigma \tau} (B' \Sigma_n^{C})' ((B' \Sigma_n^{C})')^+ (B' \Sigma_n^{C})^+ = \frac{\lambda}{\sigma \tau} (B' \Sigma_n^{C})^+,$$

$$\Delta_W = (1 + \lambda^2)I_n - \lambda^2 \sigma^2 \tau^2 (B' \Sigma_n^{C})' \Sigma_W^{+} B \Sigma_n^{C}
= (1 + \lambda^2)I_n - \lambda^2 (B \Sigma_n^{C})^+ B \Sigma_n^{C}
= I_n + \lambda^2 (I_n - (B \Sigma_n^{C})^+ B \Sigma_n^{C}).$$

This concludes the first part of our proof.

When \(\text{rank}(B) = 1\), we have that \(\text{rank}(B \Sigma_n^{C}) = \text{rank}(B) = 1\). This implies that the columns of \(B \Sigma_n^{C}\) are linearly independent. Therefore, \((B \Sigma_n^{C})^+ B \Sigma_n^{C} = I_n\), which ends the proof. \(\square\)

**Proof Proposition 3.10.** Let \(\Sigma = (\Sigma_{ij})\) and \(\Sigma = (\Sigma_{ij})\) be \(n \times n\) correlation matrices. Let \(Z\) and \(\tilde{Z}\) random vectors such that \(Z \sim \text{FS-CSN}_{n}(\mu, \Sigma, \sigma, \lambda)\) and \(\tilde{Z} \sim \text{FS-CSN}_{n}(\bar{\mu}, \bar{\Sigma}, \bar{\sigma}, \bar{\lambda})\). Let \(f_z\) and \(f_{\tilde{z}}\) denote the pdf of \(Z\) and \(\tilde{Z}\), respectively. Assume that \(f_z = f_{\tilde{z}}\). We will show that \((\mu, \Sigma, \sigma, \lambda) = (\bar{\mu}, \bar{\Sigma}, \bar{\sigma}, \bar{\lambda})\).

Thanks to the equations in (7), we have that

$$\mu = E(Z) = E(\tilde{Z}) = \bar{\mu}, \quad \sigma^2 \Sigma = \text{Var}(Z) = \text{Var}(\tilde{Z}) = \bar{\sigma}^2 \bar{\Sigma}.$$  

(A.6)

Since \(\Sigma\) and \(\Sigma\) are correlation matrices, \(\Sigma_{ii} = 1 = \tilde{\Sigma}_{ii}\) for \(i = 1, \ldots, n\). Thus, the equations in (A.6) imply that \(\sigma = \bar{\sigma}\) and \(\Sigma = \bar{\Sigma}\).

It only remains to prove that \(\lambda = \bar{\lambda}\). Consider the standardizations \(Y = \sigma^{-1} \Sigma_n^{-c} (Z - \mu)\) and \(\tilde{Y} = \bar{\sigma}^{-1} \bar{\Sigma}_n^{-c} (\tilde{Z} - \bar{\mu})\). Let \(Y_1\) and \(\tilde{Y}_1\) be the first coordinates of the random vectors \(Y\) and \(\tilde{Y}\), respectively. Besides, let \(\gamma_1(Y_1)\) and \(\gamma_1(\tilde{Y}_1)\) denote the skewness of the random variable \(Y_1\) and \(\tilde{Y}_1\), respectively. Since \(f_z = f_{\tilde{z}}\), we have that \(\gamma_1(Y_1) = \gamma_1(\tilde{Y}_1)\). Then, according to Proposition 3.4

$$\quad (2b^2 - 1)b\beta^3 \tau^3 = \gamma_1(Y_1) = \gamma_1(\tilde{Y}_1) = (2b^2 - 1)b\beta^3 \tilde{\beta}^3$$

$$\iff \delta \tau = \tilde{\delta} \tilde{\tau} \iff \delta = \frac{\tilde{\delta}}{\sqrt{1 - b^2 \beta^2}} = \frac{\tilde{\delta}}{\sqrt{1 - b^2 \tilde{\beta}^2}},$$

where \(\tau = (1 - b^2 \beta^2)^{-1/2}, \tilde{\tau} = (1 - b^2 \tilde{\beta}^2)^{-1/2}, \delta = \lambda(1 + \lambda^2)^{-1/2}\) and \(\tilde{\delta} = \bar{\lambda}(1 + \bar{\lambda}^2)^{-1/2}\). Doing some algebraic operations, we obtain \(\delta = \tilde{\delta}\), which implies that \(\lambda = \bar{\lambda}\). This ends our proof. \(\square\)

**Appendix B. Generalized least-squares estimators**

Consider the regression model

$$Y = X\beta + \epsilon, \quad \epsilon \sim \text{FS-CSN}_{n}(0, \Sigma_n, \sigma, \lambda)$$

(B.1)

where \(X\) is a (fixed) deterministic full column rank \(n \times p\) matrix, \(\beta \in \mathbb{R}^p\) is a vector of regression coefficients, \(\Sigma_n\) is a \(n \times n\) (arbitrary) correlation matrix, \(\sigma > 0\), and \(\lambda \in \mathbb{R}\). Due to its simplicity, the classical weighted least-squares (WLS) estimator constitutes a practical way to estimate \(\beta\). This estimator, obtained as the ordinary least-squares estimator of the transformation of (B.1) into a model with uncorrelated errors, is given by

$$\hat{\beta}_{\text{WLS}} = (X' (\Sigma_n^{-c})' \Sigma_n^{-c} X)^{-1} (X' (\Sigma_n^{-c})' \Sigma_n^{-c} Y).$$

Proposition 3.9 implies that

$$\hat{\beta}_{\text{WLS}} \sim \text{CSN}_{n,p}(\mu_{\text{WLS}}, \Sigma_{\text{WLS}}, D_{\text{WLS}}, 0, \Delta_{\text{WLS}}),$$

where

$$\mu_{\text{WLS}} = \beta - b\delta \tau (X' (\Sigma_n^{-c})' \Sigma_n^{-c} X)^{-1} X' (\Sigma_n^{-c})' \Sigma_n^{-c} I_n, \quad \Sigma_{\text{WLS}} = \sigma^2 \tau^2 (X' (\Sigma_n^{-c})' (\Sigma_n^{-c}) X)^{-1},$$

$$D_{\text{WLS}} = \frac{\lambda}{\sigma \tau} \Sigma_n^{-c} X, \quad \Delta_{\text{WLS}} = I_n + \lambda^2 |I_n - H_W|.$$
and $H_W = \Sigma_n^{-c}X(X'\Sigma_n^{-c}X)^{-1}X'(\Sigma_n^{-c})'$. Besides, the residual vector $r_{WLS} \equiv Y - X\hat{\beta}_{WLS}$ satisfies

$$r_{WLS} \sim \text{CSN}_{n,n}(-b\delta\sigma\tau(I_n - H_W)1_n, \sigma^2\tau^2(I_n - H_W), \frac{\lambda}{\sigma\tau}(I_n - H_W), 0, I_n + \lambda^2H_W).$$

The WLS procedure does not provide estimates for the error distribution parameters. However, in the spatial setting (11), there are methods to estimate such parameters when the error is FS-CSN. For instance, the variance and skewness parameter $(\sigma^2, \lambda)$ can be estimated using sample estimators for the variance and the skewness of $Y_i$, respectively. Besides, the correlation matrix $\Sigma_n$ can be estimated by fitting the variogram model corresponding to (12) to the sample variogram.

When the errors in (B.1) are independent, i.e. $\epsilon \sim \text{FS-CSN}(0, I_n, \sigma, \lambda)$, the WLS estimator $\hat{\beta}_{WLS}$ reduces to the ordinary least-squares (OLS) estimator $\hat{\beta}_{OLS} = (X'X)^{-1}(X'Y)$. In such a case,

$$\hat{\beta}_{OLS} \sim \text{CSN}_{n,n}(\beta - b\delta\sigma\tau(X'X)^{-1}X'1_n, \sigma^2\tau^2(X'X)^{-1}, \frac{\lambda}{\sigma\tau}X, I_n + \lambda^2I_n),$$

where $H = X(X'X)^{-1}X'$. The residuals vector $r_{OLS} \equiv Y - X\hat{\beta}_{OLS}$ satisfies

$$r_{OLS} \sim \text{CSN}_{n,n}(-b\delta\sigma\tau(I_n - H)1_n, \sigma^2\tau^2(I_n - H), \frac{\lambda}{\sigma\tau}(I_n - H), 0, I_n + \lambda^2H).$$

Notice that the mean of the residuals $r_{OLS}$ is 0, and its variance depends on $I_n - H$, as in the usual case of the residuals of regression on normal assumptions. In both cases, the residuals follow a singular distribution and are not independent by the same factor.

### Appendix C. Motivation of the risk index

The motivation of the risk index is relatively simple. Let $M_i$ be the number of deaths caused by COVID-19 during a fixed period at municipality $i$. And, let $P_i$ denote the total population in $i$. The quantity $q_i = M_i / P_i$ may be thought of as the probability that an inhabitant of $i$ dies from COVID-19. More precisely, the quantity $q_i$ represents the probability that an inhabitant of $i$ dies from COVID-19 in a period of the same order that the time it took to accumulate the deaths $M_i$ in $i$. Thus, the distribution of the time (units) that it takes for an individual from $i$ to die is Geometric($q_i$). In this way, the mean time that it takes for the total population in $i$ to die is $P_i / q_i = P_i^2 / M_i$. Thus, $T_i = P_i^2 / M_i$ is a proxy for the duration of the pandemic in $i$ if the conditions remain unchanged. Therefore, larger $T_i$ means that the pandemic will last more time; thus, higher $R_i$ values could be interpreted as higher risk.

### Appendix D. Tables

See Tables D.1–D.6.

#### Table D.1

Bias, Var, and MSE for the 1st simulation study.

| $n$ = 25 | FS-CSN | Normal |
|----------|--------|-------|
|          | $\lambda$ | $\sigma$ | $\rho$ | $\beta_0$ | $\beta_1$ | $\sigma$ | $\rho$ | $\beta_0$ | $\beta_1$ |
| Bias     | -0.439 | -0.179 | -0.446 | -0.034 | 0.013 | -0.194 | -0.101 | -0.047 | 0.016 |
| MCSE Bias | 0.016 | 0.000 | 0.000 | 0.004 | 0.000 | 0.000 | 0.006 | 0.005 | 0.000 |
| Var      | 10.001 | 0.231 | 0.016 | 2.741 | 0.122 | 0.230 | 3.858 | 2.886 | 0.128 |
| MCSE Var | 0.893 | 0.014 | 0.003 | 0.167 | 0.007 | 0.014 | 1.668 | 0.172 | 0.007 |
| MSE      | 0.035 | 0.001 | 0.000 | 0.007 | 0.000 | 0.001 | 0.066 | 0.007 | 0.000 |
| MCSE MSE | 0.035 | 0.001 | 0.000 | 0.007 | 0.000 | 0.001 | 0.066 | 0.007 | 0.000 |

(continued on next page)
The table presents the bias, the variance, and the mean square error produced by the parameters estimated from the first simulation study. It also includes the corresponding Monte Carlo standard errors.

Table D.2
Means and quartiles of the 1st simulation study.

|       | FS-CSN                  | Normal                  |
|-------|------------------------|-------------------------|
|       | λ          | σ      | ρ    | β₂₀   | β₁₀   | σ      | ρ    | β₂₀   | β₁₀   |
| n = 25| Min.     | -11.747 | 1.245 | 0.001 | 4.951 | 0.906 | 1.245 | 0.002 | 4.878 | 0.820 |
|       | 1st Q    | -0.002  | 2.667 | 0.009 | 8.712 | 1.778 | 2.648 | 0.009 | 8.776 | 1.773 |
|       | Median   | 1.737   | 2.961 | 0.018 | 9.920 | 2.026 | 2.957 | 0.018 | 9.908 | 2.028 |
|       | Mean     | 2.061   | 2.983 | 0.055 | 9.966 | 2.013 | 2.969 | 0.400 | 9.953 | 2.016 |
|       | 3rd Q    | 3.353   | 3.273 | 0.036 | 11.104| 2.255 | 3.265 | 0.056 | 11.059| 2.265 |
|       | Max.     | 15.267  | 4.685 | 0.720 | 16.097| 3.097 | 4.752 | 29.289| 15.916| 3.167 |

|       | FS-CSN                  | Normal                  |
|-------|------------------------|-------------------------|
| n = 50| Min.         | -4.557  | 2.305 | 0.003 | 5.850 | 1.218 | 2.303 | 0.003 | 5.874 | 1.192 |
|       | 1st Q        | 1.696   | 2.862 | 0.015 | 9.282 | 1.841 | 2.856 | 0.015 | 9.227 | 1.822 |
|       | Median       | 2.600   | 3.079 | 0.029 | 10.062| 1.991 | 3.062 | 0.032 | 10.094| 1.990 |
|       | Mean         | 3.021   | 3.091 | 0.068 | 10.046| 1.992 | 3.079 | 0.168 | 10.050| 1.993 |

(continued on next page)
The table contains the mean and the quartiles, including the minimum and maximum value, of the parameters estimated in the first simulation study.

Table D.3
Bias, Var, and MSE for the 2nd simulation study.

| n = 50 | Bias | 0.112 | −0.162 | −0.048 | −0.002 | 0.002 | −0.174 | 1.088 | −0.021 | 0.006 |
|        | MCSE Bias | 0.040 | 0.000 | 0.002 | 0.003 | 0.000 | 0.000 | 0.025 | 0.004 | 0.000 |
|        | Var | 15.257 | 0.116 | 0.063 | 1.000 | 0.045 | 0.125 | 9.326 | 1.458 | 0.070 |
|        | MCSE Var | 1.386 | 0.008 | 0.158 | 0.071 | 0.003 | 0.009 | 1.417 | 0.104 | 0.005 |
|        | MSE | 0.072 | 0.000 | 0.008 | 0.004 | 0.000 | 0.000 | 0.008 | 0.005 | 0.000 |
|        | MCSE MSE | 0.072 | 0.000 | 0.008 | 0.004 | 0.000 | 0.000 | 0.008 | 0.005 | 0.000 |
| n = 100 | Bias | 1.112 | −0.064 | −0.070 | −0.031 | 0.003 | −0.073 | 1.233 | −0.030 | 0.004 |
|        | MCSE Bias | 0.035 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 0.029 | 0.002 | 0.000 |
|        | Var | 13.447 | 0.057 | 0.363 | 0.424 | 0.015 | 0.065 | 10.986 | 0.755 | 0.029 |
|        | MCSE Var | 1.293 | 0.004 | 0.111 | 0.030 | 0.001 | 0.005 | 2.592 | 0.050 | 0.002 |
|        | MSE | 0.082 | 0.000 | 0.005 | 0.002 | 0.000 | 0.000 | 0.150 | 0.003 | 0.000 |
|        | MCSE MSE | 0.082 | 0.000 | 0.005 | 0.002 | 0.000 | 0.000 | 0.150 | 0.003 | 0.000 |

(continued on next page)
### Table D.3 (continued)

| FS-CSN          | Normal         |
|-----------------|----------------|
| $\lambda$, $\sigma$, $\rho$, $\beta_0$, $\beta_1$ | $\sigma$, $\rho$, $\beta_0$, $\beta_1$ |
| $n = 200$       |                |
| Bias            | -0.003         |
| MCSE Bias       | 0.000          |
| Var             | 0.159          |
| MCSE Var        | 0.033          |
| MSE             | 0.000          |
| MCSE MSE        | 0.000          |
| $n = 200$       |                |
| Bias            | -0.031         |
| MCSE Bias       | 0.000          |
| Var             | 0.033          |
| MCSE Var        | 0.015          |
| MSE             | 0.002          |
| MCSE MSE        | 0.000          |

The table presents the bias, the variance, and the mean square error produced by the parameters estimated from the second simulation study. It also includes the corresponding Monte Carlo standard errors.

### Table D.4

Means and quartiles of the 2nd simulation study.

| FS-CSN          | Normal         |
|-----------------|----------------|
| $\lambda$, $\sigma$, $\rho$, $\kappa$, $\beta_0$, $\beta_1$ | $\sigma$, $\rho$, $\beta_0$, $\beta_1$ |
| $n = 50$        |                |
| Min.            | -0.400         |
| 1st Q           | 4.404          |
| Median          | 6.244          |
| 3rd Q           | 9.215          |
| Max.            | 22.967         |
| $n = 50$        |                |
| Min.            | 0.208          |
| 1st Q           | 5.614          |
| Median          | 7.441          |
| 3rd Q           | 9.906          |
| Max.            | 21.031         |
| $n = 100$       |                |
| Min.            | -0.002         |
| 1st Q           | 6.049          |
| Median          | 7.393          |
| 3rd Q           | 9.269          |
| Max.            | 17.960         |
| $n = 200$       |                |
| Min.            | -0.002         |
| 1st Q           | 6.049          |
| Median          | 7.393          |
| 3rd Q           | 9.269          |
| Max.            | 17.960         |

The table contains the mean and the quartiles, including the minimum and maximum value, of the parameters estimated in the second simulation study.

### Table D.5

Parameters estimation.

| FS-CSN          | Normal         |
|-----------------|----------------|
| $\lambda$, $\sigma$, $\rho$, $\kappa$, $\beta_0$, $\beta_1$, $\beta_2$, $\beta_3$, $\beta_4$, $\beta_5$, $\beta_6$ | $\sigma$, $\rho$, $\beta_0$, $\beta_1$, $\beta_2$, $\beta_3$, $\beta_4$, $\beta_5$, $\beta_6$ |
| $n = 200$       |                |
| Min.            | 2.094          |
| 1st Q           | 2.203          |
| Median          | 2.250          |
| Mean            | 2.273          |
| 3rd Q           | 2.334          |
| Max.            | 2.449          |
| $n = 200$       |                |
| Min.            | 0.531          |
| 1st Q           | 0.574          |
| Median          | 0.574          |
| Mean            | 0.574          |
| 3rd Q           | 0.579          |
| Max.            | 0.607          |

The table presents the cross-validation means and quartiles of the parameters estimated for the models of the risk index $R_i$. It includes both the FS-CSN and normal models.
### Table D.6
MAE and RSME errors and the MLPI.

|   | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| FS-CSN MAE model | 0.423 | 0.404 | 0.419 | 0.411 | 0.418 | 0.427 | 0.411 | 0.408 | 0.422 | 0.418 |
| FS-CSN RSME model | 0.618 | 0.600 | 0.617 | 0.608 | 0.612 | 0.622 | 0.605 | 0.601 | 0.615 | 0.617 |
| FS-CSN MAE prediction | 0.455 | 0.557 | 0.373 | 0.602 | 0.344 | 0.415 | 0.444 | 0.513 | 0.550 | 0.343 |
| FS-CSN RSME prediction | 0.579 | 0.665 | 0.524 | 0.733 | 0.414 | 0.541 | 0.584 | 0.615 | 0.691 | 0.406 |
| MLPI | 2.138 | 2.124 | 2.273 | 2.235 | 2.168 | 2.154 | 2.138 | 2.142 | 2.022 | 2.200 |

|   | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Normal MAE model | 0.427 | 0.409 | 0.425 | 0.406 | 0.424 | 0.432 | 0.416 | 0.413 | 0.428 | 0.424 |
| Normal RSME model | 0.623 | 0.603 | 0.623 | 0.600 | 0.618 | 0.627 | 0.611 | 0.605 | 0.620 | 0.621 |
| Normal MAE prediction | 0.408 | 0.557 | 0.310 | 0.735 | 0.378 | 0.348 | 0.431 | 0.521 | 0.536 | 0.405 |
| Normal RSME prediction | 0.520 | 0.673 | 0.454 | 0.895 | 0.457 | 0.457 | 0.537 | 0.583 | 0.643 | 0.484 |
| MLPI | 2.200 | 2.188 | 2.253 | 2.577 | 2.183 | 2.128 | 2.191 | 2.129 | 2.206 | 2.200 |

For each of the folds in the cross-validation, the table presents the MAE and the RSME of the fitting of both models and their PPL-predictions. It also contains the mean length of the prediction intervals (MLPI) in each of the folds.

### References

Allard, D., Naveau, P., 2007. A new spatial skew-normal random field model. Comm. Statist. Theory Methods 36 (9), 1821–1834.

Arnold, B.C., Beaver, R.J., 2002. Skewed multivariate models related to hidden truncation and/or selective reporting. Test 11 (1), 7–54.

Billingsley, P., 1995. Probability and Measure. In: Wiley Series in Probability and Statistics, Wiley.

Bjornstad, J.F., 1990. Predictive likelihood: A review. Statist. Sci. 5 (2), 242–254.

Copas, J.B., Li, H., 1997. Inference for non-random samples. J. R. Stat. Soc. Ser. B Stat. Methodol. 59 (1), 55–95.

Díaz-García, J.A., González-Farías, G., 2008. Singular extended skew-elliptical distributions. J. Korean Stat. Soc. 37 (4), 385–392.

Domínguez-Molina, J.A., González-Farías, G., Gupta, A.K., 2003. The multivariate closed skew normal distribution. Technical Report 03–12, Department of Mathematics and Statistics, Bowling Green State University.

Domínguez-Molina, J.A., González-Farías, G., Ramos-Quiroga, R., Gupta, A.K., 2007. A matrix variate closed skew-normal distribution with applications to stochastic frontier analysis. Comm. Statist. Theory Methods 36 (9), 1691–1703.

González-Farías, G., Domínguez-Molina, J.A., Gupta, A.K., 2004a. Additive properties of skew normal random vectors. J. Statist. Plann. Inferenc. 126 (2), 521–534.

González-Farías, G., Domínguez-Molina, J.A., Gupta, A.K., 2004b. The closed skew-normal distribution. In: Genton, M.G. (Ed.), Skew-Elliptical Distributions and their Applications: A Journey beyond Normality. Chapman and Hall/CRC, pp. 25–42.

Gräler, B., Pebesma, E., Heuvelink, G., 2016. Spatio-temporal interpolation using gstat. R J. 8 (1), 204–218.

Gupta, A.K., González-Farías, G., Domínguez-Molina, J.A., 2004. A multivariate skew normal distribution. J. Multivariate Anal. 89 (1), 181–190.

Karimi, O., Mohammadzadeh, M., 2011. Bayesian spatial prediction for discrete closed skew Gaussian random field. Math. Geosci. 43 (5), 565–582.

Karimi, O., Mohammadzadeh, M., 2012. Bayesian spatial regression models with closed skew normal correlated errors and missing observations. Statist. Papers 53 (1), 205–218.

Kou, Y., Lu, C.-T., Dos Santos, R.F., 2007. Spatial outlier detection: A graph-based approach. In: 19th IEEE International Conference on Tools with Artificial Intelligence(IC TAI 2007). Vol. 1, pp. 281–288. http://dx.doi.org/10.1109/ICTAI.2007.139.

Kreutz, C., Raue, A., Kaschek, D., Timmer, J., 2013. Profile likelihood in systems biology. FEBS J. 280 (11), 2564–2571.

Mahmoudian, B., 2018. On the existence of some skew-Gaussian random field models. Statist. Probab. Lett. 137, 331–335.

Mathisen, P.E., 1979. Prediction functions. Scand. J. Stat. 6 (1), 1–21.

Minozzo, M., Ferracuti, L., 2012. On the existence of some skew-stationary processes. Chilean J. Stat. 3 (2), 157–170.

Nash, J.C., 2014. On best practice optimization methods in R. J. Stat. Softw. 60 (2), 1–14.

Nash, J.C., Varadhan, R., 2011. Unifying optimization algorithms to aid software system users: optimx for R. J. Stat. Softw. 43 (9), 1–14.

Pavlyuk, D., Girtcius, E., 2015. Csn: Closed skew-normal distribution. URL https://CRAN.R-project.org/package=csn, R Package Version 1.1.3.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosci. 30 (7), 683–691.

Ramgawat, S., Rastogi, R., Shim, K., 2000. Efficient algorithms for mining outliers from large data sets. In: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. pp. 427–438.

Rimstad, K., Omre, H., 2014. Skew-Gaussian random fields. Spatial Stat. 10, 43–62.

Tagle, F., Castruccio, S., Crippa, P., Genton, M.G., 2019. A non-Gaussian spatio-temporal model for daily wind speeds based on a multi-variate skew-t distribution. J. Time Series Anal. 40 (3), 312–326.

Yamanishi, K., Takeuchi, J.-I., Williams, G., Milne, P., 2004. On-line unsupervised outlier detection using finite mixtures with discount learning algorithms. Data Min. Knowl. Discov. 8 (3), 275–300.