Statistics and Computing (2024) 34:116
https://doi.org/10.1007/s11222-024-10427-3

ORIGINAL PAPER

A flexible Bayesian tool for CoDa mixed models: logistic-normal distribution with Dirichlet covariance

Joaquín Martínez-Minaya[1] · **Haavard Rue**[2]

Received: 28 August 2023 / Accepted: 14 March 2024

Abstract
Compositional Data Analysis (CoDa) has gained popularity in recent years. This type of data consists of values from disjoint categories that sum up to a constant. Both Dirichlet regression and logistic-normal regression have become popular as CoDa analysis methods. However, fitting this kind of multivariate models presents challenges, especially when structured random effects are included in the model, such as temporal or spatial effects. To overcome these challenges, we propose the logistic-normal Dirichlet Model (LNDM). We seamlessly incorporate this approach into the **R-INLA** package, facilitating model fitting and model prediction within the framework of Latent Gaussian Models. Moreover, we explore metrics like Deviance Information Criteria, Watanabe Akaike information criterion, and cross-validation measure conditional predictive ordinate for model selection in **R-INLA** for CoDa. Illustrating LNDM through two simulated examples and with an ecological case study on *Arabidopsis thaliana* in the Iberian Peninsula, we underscore its potential as an effective tool for managing CoDa and large CoDa databases.

Keywords CoDa · Dirichlet · INLA · Spatial

1 Introduction

Compositional Data analysis is an increasingly popular topic for understanding processes that consist in values that correspond to disjoint categories, the sum of which is a constant. Those values are usually proportions or percentages, and in such cases the constant is 1 or 100. The data generated from these processes are widely known as Compositional Data (CoDa). For the sake of simplicity and without loss of generality, from now on, we assume the constant to be 1. Connor and Mosimann (1969) proposed Dirichlet regression to deal with CoDa. Since then, several studies have been conducted using this technique, and most of them have proved that it is a very valuable tool for modelling CoDa, see for example Hijazi and Jernigan (2009) and Pirzamanbein et al. (2020).

There are other approaches to CoDa analysis. Aitchison (1986) presented an unified theory, developing a range of methods based on the idea that "information in compositional vectors is concerned with relative, not absolute magnitudes". With this statement, the notion of ratios among proportions emerged and the concept of log-ratios arose as the preferred method for dealing with CoDa. Modelling CoDa using logistic-normal gained ground, and the bases of CoDa were established.

A vast body of literature exists on the subject of applying these methods using both Dirichlet regression and logistic-normal regression in different fields, including Ecology (Kobal et al. 2017; Douma and Weedon 2019), Geology (Buccianti and Grunsky 2014; Engle and Rowan 2014), Genomics (Tsilimigras and Fodor 2016; Shi et al. 2016; Washburne et al. 2017; Creus Martí et al. 2022), Environmental Sciences (Aguilera et al. 2021; Mota-Bertran et al. 2022) or Medicine (Dumuid et al. 2018; Fairclough et al. 2018).

Nevertheless, one of the biggest problems encountered when dealing with CoDa models is performing inference. To do so, different approaches have been proposed; in particu-

✉ Joaquín Martínez-Minaya
jmarmin@eio.upv.es

Haavard Rue
haavard.rue@kaust.edu.sa

[1] Multivariate Statistical Engineering Research Group, Department of Applied Statistics and Operational Research, and Quality, Universitat Politècnica de València, Camí de Vera, SN 46022 Valencia, Spain

[2] Computer, Electrical and Mathematical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

Published online: 16 April 2024

lar, many R-packages have been implemented not only from the frequentist perspective (Cribari-Neto and Zeileis 2010; Templ et al. 2011; Maier 2014), but also from the Bayesian paradigm. R-packages such as **BayesX** (Klein et al. 2015), **Stan** (Sennhenn-Reulen 2018), **BUGS** (van der Merwe 2018) and **R-JAGS** (Plummer 2016) have tools for dealing with CoDa. These Bayesian packages are mainly based on Markov chain Monte Carlo (MCMC) methods, which construct a Markov chain whose stationary distribution converges to the posterior distribution. However, the computational cost of MCMC can be high. Moreover, the integrated nested Laplace approximation (INLA) methodology (Rue et al. 2009), which is mainly intended for approximating the posterior distribution using the Laplace integration method, has become an alternative to MCMC guaranteeing a higher computational speed for Latent Gaussian Models (LGMs). With the incorporation of new techniques from Bayesian variational inference (Niekerk and Rue 2021; Van Niekerk et al. 2023) and the optimisation of the computation, which improves its parallel performance (Gaedke-Merzhäuser et al. 2023), a new era is emerging in the INLA software. Hence, incorporating a tool for dealing with CoDa would be a convenient way to tackle the large CoDa databases sometimes encountered.

Nonetheless, in **R-INLA**, it is still a challenge to fit models when we deal with a multivariate likelihood such as the ones defined in simplex of dimension $D(\mathbb{S}^D)$. There are some approximations for the Dirichlet likelihood that involve converting the original Dirichlet observations into Gaussian pseudo-observations conditioned to the linear predictor (Martínez-Minaya et al. 2023) or just converting a CoDa multivariate response into coordinates using the isometric log-ratio transformation (Mota-Bertran et al. 2022) and fitting them in an independent way. However, there is no unified way to fit these models inside **R-INLA** and take advantage of all its facilities.

In this paper we present the logistic-normal Dirichlet model (LNDM), which mainly uses logistic-normal distribution with Dirichlet covariance through the additive log-ratio transformation as likelihood. This allows us to integrate it within the **R-INLA** package in a very simple way. Thus, we benefit from all the other features of **R-INLA** for model fitting, model selection and predictions within the framework of LGMs. Additionally, we present how measures such the Deviance Information Criteria (Spiegelhalter et al. 2002, DIC), the Watanabe Akaike information criterion (Watanabe and Opper 2010; Gelman et al. 2014, WAIC), or the cross-validation measure conditional predictive ordinate (CPO) for evaluating the predictive capacity (Pettit 1990; Roos and Held 2011) are computed in **R-INLA** for dealing with CoDa. To show how the method works, two simulate examples and a real example in the field of Ecology were implemented. In the last part, we conducted a spatial analysis of the plant *Arabidopsis thaliana* on the Iberian Peninsula.

The paper is then divided into 7 more sections. Section 2 introduces CoDa, the distributions that can be defined in $\mathbb{S}^D$, and their equivalence. Section 3 presents some fundamentals of the INLA methodology. Section 4 is devoted to introducing the logistic-normal regression with Dirichlet covariance. In Sect. 5, we introduce spatial models as well as model selection measures in CoDa. Section 6 focuses on presenting a simulated spatial study. In Sect. 7, we provide a real application of this method and, finally, Sect. 8 concludes and discusses future avenues of research.

2 CoDa background

This section is devoted to introducing some preliminary concepts for a better understanding of CoDa. In particular, we present some basic and formal definitions of the two main distributions employed when we deal with CoDa.

2.1 CoDa: Definitions

Let $\boldsymbol{y}_{D\times 1}$ be a vector that satisfies $\sum_{d=1}^{D} y_d = 1$, and $0 < y_d < 1, d = 1, \ldots, D$. This vector is called a composition, and it pertains to the simplex sample space. The simplex of dimension D, denoted by $\mathbb{S}^D$, is defined as:

$$\mathbb{S}^D = \left\{ \boldsymbol{y} \in \mathbb{R}^D \mid 0 < y_d < 1; \ \sum_{d=1}^{D} y_d = 1 \right\}. \tag{1}$$

As in the ordinary real Euclidean space, there is a geometry defined in $\mathbb{S}^D$. It does not follow the usual Euclidean geometry, and it was introduced by Pawlowsky-Glahn and Egozcue (2001) and Egozcue et al. (2003). It is called Aitchison geometry. The definitions of perturbation and powering are sufficient to obtain a vector space of compositions and the usual properties such as commutativity, associativity and distributivity hold. With the definition of the Aitchison inner product, the Aitchison norm and the Aitchison distance, an Euclidean linear vector space is obtained (Pawlowsky-Glahn and Egozcue 2001).

Following the fundamentals proposed by Aitchison (1986), log-ratios play an important role in CoDa analysis. They can be constructed in different ways, including centered log-ratio, isometric log-ratio or additive log-ratio, among others (Egozcue et al. 2012). In this work, we focus on the well-known additive log-ratio transformation because of its straightforward interpretation (Greenacre et al. 2023), and due to its being a one-to-one mapping from $\mathbb{S}^D$ to $\mathbb{R}^{D-1}$. It is defined as:

$$z_{(D-1)\times 1} = alr(\boldsymbol{y}) := \left[\log\left(\frac{y_1}{y_D}\right), \ldots, \log\left(\frac{y_{D-1}}{y_D}\right)\right], \quad (2)$$

where D is the reference category. In Greenacre et al. (2023), the authors depicted some criteria to select the reference category. They recommended choosing the one whose logarithm has low variance as a reference, and avoiding taking a reference with low relative abundances across samples. The new variables generated are called alr-coordinates. The inverse alr, also called alr^{-1} is

$$alr^{-1}(\boldsymbol{z}) = \left[\frac{\exp(z_1)}{1+\sum_{d=1}^{D-1}\exp(z_d)}, \ldots, \frac{\exp(z_{D-1})}{1+\sum_{d=1}^{D-1}\exp(z_d)}, \frac{1}{1+\sum_{d=1}^{D-1}\exp(z_d)}\right].$$

In addition to Aitchison geometry, several probability distributions have also been characterised in $\mathbb{S}^D$ (Figueras et al. 2003), although here we focus on the normal distribution on the simplex or logistic-normal distribution, and the Dirichlet distribution.

2.2 Logistic-normal distribution and Dirichlet distribution

Logistic-normal distribution was defined by Aitchison and Shen (1980) and it was studied in depth in Aitchison (1986). A D random vector $\boldsymbol{y}$ is said to have a logistic-normal distribution $\mathcal{LN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, or alternatively a normal distribution on $\mathbb{S}^D$, if any of its vector of log-ratio coordinates has a joint $(D-1)$-variate normal distribution. This definition can be adapted straight to a CoDa response using alr-coordinates, as:

$$\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{LN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \iff alr(\boldsymbol{y}) \mid \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (3)$$

$\boldsymbol{\mu}$ being a $D-1$ dimensional vector and $\boldsymbol{\Sigma}$ a $(D-1)\times(D-1)$ covariance matrix. Alternatively, the Dirichlet distribution was introduced in Connor and Mosimann (1969), and it is the generalisation of the widely known beta distribution. A D random vector $\boldsymbol{y}$ is said to have a Dirichlet distribution $\mathcal{D}(\boldsymbol{\alpha})$, if it has the following probability density:

$$p(\boldsymbol{y} \mid \boldsymbol{\alpha}) = \frac{1}{\mathrm{B}(\boldsymbol{\alpha})}\prod_{d=1}^{D} y_d^{\alpha_d - 1}, \quad (4)$$

$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_D)$ being the vector of shape parameters for each category, $\alpha_D > 0\ \forall d$, $y_d \in (0,1)$, $\sum_{d=1}^{D} y_d = 1$, and $\mathrm{B}(\boldsymbol{\alpha})$ the multinomial Beta function, which serves as the normalising constant. The multinomial Beta function is defined as $\mathrm{B}(\boldsymbol{\alpha}) = \prod_{d=1}^{D}\Gamma(\alpha_d)/\Gamma(\sum_{d=1}^{D}\alpha_d)$. The sum of all α's, $\alpha_0 = \sum_{d=1}^{D}\alpha_c$, is usually interpreted as a precision parameter. The Beta distribution is the particular case when $D = 2$. In addition, each variable is marginally Beta distributed with $\alpha = \alpha_d$ and $\beta = \alpha_0 - \alpha_d$. If $\boldsymbol{y} \sim \mathcal{D}(\boldsymbol{\alpha})$, the expected values are $\mathrm{E}(y_d) = \alpha_d/\alpha_0$, the variances are $\mathrm{Var}(y_d) = [\alpha_c(\alpha_0 - \alpha_d)]/[\alpha_0^2(\alpha_0+1)]$ and the covariances are $\mathrm{Cov}(y_d, y_{d'}) = -\alpha_d\alpha_{d'}/[\alpha_0^2(\alpha_0+1)]$.

2.3 Relation between the two distributions

As pointed out in Aitchison (1986, 126–129), the logistic-normal and the Dirichlet distribution are separate in the sense that they are never exactly equal for any choice of parameters. However, through the Kullback–Leibler divergence (KL), which measures by how much the approximation q misses the target p, the Dirichlet distribution can be approached with the logistic-normal distribution. The solution to the minimisation of the KL:

$$K(p,q) = \int_{\mathcal{S}^D} p(\boldsymbol{y} \mid \boldsymbol{\alpha}) \log\left(\frac{p(\boldsymbol{y} \mid \boldsymbol{\alpha})}{q(\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})}\right) d\boldsymbol{y}, \quad (5)$$

where $p(\boldsymbol{y} \mid \boldsymbol{\alpha})$ represents the density function of the Dirichlet, and $q(\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$, the logistic-normal density function, is minimised by:

$$\begin{aligned} \boldsymbol{\mu} &= \boldsymbol{E}\left[\log\left(\tfrac{y_1}{y_D}\right), \ldots, \log\left(\tfrac{y_{D-1}}{y_D}\right)\right] = \boldsymbol{E}\left[alr(\boldsymbol{y})\right], \\ \boldsymbol{\Sigma} &= \boldsymbol{Var}\left[\log\left(\tfrac{y_1}{y_D}\right), \ldots, \log\left(\tfrac{y_{D-1}}{y_D}\right)\right] = \boldsymbol{Var}\left[alr(\boldsymbol{y})\right], \end{aligned} \quad (6)$$

and the solution can be written in terms of the digamma ϕ and trigamma ϕ' functions as:

$$\begin{aligned} \mu_d &= \phi(\alpha_d) - \phi(\alpha_D), \quad d = 1, \ldots, D-1, \\ \Sigma_{dd} &= \phi'(\alpha_d) + \phi'(\alpha_D), \quad d = 1, \ldots, D-1, \\ \Sigma_{dk} &= \phi'(\alpha_D), \quad d \neq k. \end{aligned} \quad (7)$$

This approach plays an important role in this paper, as it constitutes the basis for defining logistic-normal regression with Dirichlet covariance. But first we introduce the model framework in which this likelihood is included, that is, Latent Gaussian Models (LGMs, Rue et al. 2009).

3 LGMs and INLA

The popularity of INLA lies in the fact that it allows fast approximate inference for LGMs. Furthermore, the INLA software is experiencing a new era, facilitated by the integration of novel techniques from Bayesian variational inference (Niekerk and Rue 2021; Van Niekerk et al. 2023) and enhanced computation optimization, leading to improved

parallel performance (Gaedke-Merzhäuser et al. 2023). This section is devoted to briefly introducing the structure of LGMs and how INLA makes inference and prediction with the new advances in INLA.

3.1 LGMs

In Van Niekerk et al. (2023) a new formulation of INLA is presented. So, we follow it to introduce the notions of INLA. LGMs can be seen as three-stage hierarchical Bayesian models in which observations $\boldsymbol{y}_{N\times 1}$ can be assumed to be conditionally independent given a latent Gaussian random field $\boldsymbol{X}$ and hyperparameters $\boldsymbol{\theta}_1$

$$\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\theta}_1 \sim \prod_{n=1}^{N} p(y_n \mid \boldsymbol{X}, \boldsymbol{\theta}_1). \tag{8}$$

The versatility of the model class is related to the specification of the latent Gaussian field:

$$\boldsymbol{X} \mid \boldsymbol{\theta}_2 \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}^{-1}(\boldsymbol{\theta}_2)), \tag{9}$$

which includes all the latent (non-observable) components of interest, such as fixed effects and random terms, describing the process underlying the data. The hyperparameters $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$ control the latent Gaussian field and/or the likelihood for the data.

Additionally, the LGMs are a class generalising the large number of related variants of additive and generalised models. If $\boldsymbol{\eta}_{N\times 1}$ is a column vector representing the linear predictor, then different effects can be added to it:

$$\boldsymbol{\eta}_{N\times 1} = \boldsymbol{X}\boldsymbol{\beta} + \sum_{l=1}^{L} f_l(\boldsymbol{u}_l) \tag{10}$$

where $\boldsymbol{X}$ is the design matrix for the fixed part (including the first column of 1 s if intercepts are added to the model), and $\boldsymbol{\beta}_{(M+1)\times 1}$ is a column vector for the linear effects of $\boldsymbol{X}$ on $\boldsymbol{\eta}$. $\{\boldsymbol{f}\}$ are unknown functions of $\boldsymbol{U}$. This formulation can be seen as any model where each one of the $f^l(.)$ terms can be written in matrix form as $\boldsymbol{A}_l\boldsymbol{u}_l$. So, expression (10) can be rewritten as $\boldsymbol{\eta} = \boldsymbol{A}\boldsymbol{X}$, with $\boldsymbol{A}$ a sparse design matrix that links the linear predictors to the latent field.

When we do inference, the aim is to estimate $\boldsymbol{X}_{(M+1+L)\times 1} = \{\boldsymbol{\beta}, \boldsymbol{f}\}$, which represents the set of unobserved latent variables (latent field). If a Gaussian prior is assumed for $\boldsymbol{\beta}$ and $\boldsymbol{f}$, the joint prior distribution of $\boldsymbol{X}$ is Gaussian. This yields the latent field $\boldsymbol{X}$ in the hierarchical LGM formulation. The vector of hyperparameters $\boldsymbol{\theta}$ contain the non-Gaussian parameters of the likelihood and the model components. These parameters commonly include variance, scale or correlation parameters.

In most cases, the latent field in addition to be Gaussian, is also a Gaussian Markov random field (GMRF, Rue and Held 2005). A GMRF is a multivariate Gaussian random variable with additional conditional independence properties: x_j and x_j' are conditionally independent given the remaining elements if and only if the (i, j) entry of the precision matrix is 0. Implementation of INLA method use this property to speed up computation.

3.2 INLA

The main idea of the INLA approach is to approximate the posteriors of interest: the marginal posteriors for the latent field, $p(\mathcal{X}_m \mid \boldsymbol{y})$, and the marginal posteriors for the hyperparameters, $p(\theta_k \mid \boldsymbol{y})$. With the modern formulation (Van Niekerk et al. 2023), the main enhancement is that the latent field is not augmented with the 'noisy' linear predictors. Then, the joint density of the latent field, hyperparameters and the data is derived as:

$$p(\boldsymbol{X}, \boldsymbol{\theta} \mid \boldsymbol{y}) \propto p(\boldsymbol{\theta})p(\boldsymbol{X} \mid \boldsymbol{\theta}) \prod_{n=1}^{N} p(y_n \mid (\boldsymbol{A}\boldsymbol{X})_n, \boldsymbol{\theta}). \tag{11}$$

Thus, the initial step in approaching the posterior distributions involves determining the mode and the Hessian at the mode of $\tilde{p}(\boldsymbol{\theta} \mid \boldsymbol{y})$:

$$\tilde{p}(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto \left.\frac{p(\boldsymbol{X}, \boldsymbol{\theta} \mid \boldsymbol{y})}{p_G(\boldsymbol{X} \mid \boldsymbol{\theta}, \boldsymbol{y})}\right|_{\boldsymbol{X}=\boldsymbol{\mu}(\boldsymbol{\theta})}. \tag{12}$$

being $p_G(\boldsymbol{X} \mid \boldsymbol{\theta}, \boldsymbol{y})$ the Gaussian approximation to $p(\boldsymbol{X} \mid \boldsymbol{\theta}, \boldsymbol{y})$ computed as depicted in Van Niekerk et al. (2023):

$$\boldsymbol{X} \mid \boldsymbol{\theta}, \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{Q}_{\boldsymbol{X}}^{-1}(\boldsymbol{\theta})). \tag{13}$$

The subsequent step involves obtaining the conditional posterior distributions of the elements in $\boldsymbol{X}$. To achieve this, it suffices to perform integration $\boldsymbol{\theta}$ out from (13) using T integration points θ_t and area weights δ_t defined by some numerical integration scheme:

$$\begin{aligned}\tilde{p}(\mathcal{X}_m \mid \boldsymbol{y}) &= \int p_G(\mathcal{X}_m \mid \boldsymbol{\theta}, \boldsymbol{y}) d\boldsymbol{\theta} \\ &\approx \sum_{t=1}^{T} p_G(\mathcal{X}_m \mid \theta_t, \boldsymbol{y})\tilde{p}(\theta_t \mid \boldsymbol{y})\delta_t.\end{aligned} \tag{14}$$

Finally, the recent proposed Variational Bayes correction to Gaussian means by Niekerk and Rue (2021) is used to efficiently calculate an improved mean for the marginal posterior of the latent field. All this methodology can be used through R with the **R-INLA** package. For more details about **R-INLA** we refer the reader to Blangiardo and Cameletti (2015), Zuur

et al. (2017), Wang et al. (2018), Krainski et al. (2018), Moraga (2019), Gómez-Rubio (2020), Van Niekerk et al. (2023), where practical examples and code guidelines are provided.

4 INLA for fitting logistic-normal regression with Dirichlet covariance

This part of the paper focuses on presenting our approximation for fitting CoDa.

4.1 Bayesian logistic-normal regression with Dirichlet covariance

To define the likelihood we need the logistic-normal distribution and the structure of the variance–covariance matrix presented in Eq. (7).

Definition 1 $\boldsymbol{y} \in \mathbb{S}^D$ follows a logistic-normal distribution with Dirichlet covariance $\mathcal{LND}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if and only if $alr(\boldsymbol{y}) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and:

$$\begin{aligned} \Sigma_{dd} &= \sigma_d^2 + \gamma, \quad d = 1, \ldots, D-1, \\ \Sigma_{dk} &= \gamma, d \neq k, \end{aligned}$$

where $\sigma_d^2 + \gamma$ represents the variance of each log-ratio and γ is the covariance between log-ratios.

From now on, we will refer to $\mathcal{ND}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as the multivariate normal with Dirichlet covariance structure, as depicted in Definition 1. Let $\boldsymbol{y}$ be a multivariate random variable such as $\boldsymbol{y} \sim \mathcal{LND}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, which by definition is equivalent to $alr(\boldsymbol{y}) \sim \mathcal{ND}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Because of its easy interpretability in terms of log-ratios with the reference category, we focus on modelling $alr(\boldsymbol{y})$ as a $\mathcal{ND}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Let $\boldsymbol{\mu}^{(d)}_{N\times 1}$, a column vector representing the linear predictor for the nth observation in the dth alr-coordinate, and $\boldsymbol{X}^{(d)}$ with dimension $N \times (M^{(d)}+1)$, $d = 1, \ldots, D-1$, the design matrix, which can be different for each dth alr-coordinate; in other words, each alr-coordinate can be explained by different covariates. Let $\boldsymbol{f}^{(d)}$ be a set of $L^{(d)}$ unknown functions of $\boldsymbol{U}$ that also can vary depending on the alr-coordinate. For the sake of simplicity, and without loss of generality, we assume $M^{(d)} = M$ and $L^{(d)} = L$, fixing the number of covariates and the number of functions as the same in each linear predictor. Finally, we define $\boldsymbol{\beta}^{(d)}_{(M+1)\times 1}$ a $M+1$-dimensional column vector that contains the parameters corresponding to the fixed effects including the intercept.

Then, the logistic-normal Dirichlet model (LNDM) can be expressed as follows:

$$alr(\boldsymbol{y}) \sim \mathcal{ND}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{15}$$

$$\boldsymbol{\mu}^{(d)} = \boldsymbol{X}\boldsymbol{\beta}^{(d)} + \sum_{l=1}^{L} \boldsymbol{f}_l^{(d)}(u_l), \tag{16}$$

being $\boldsymbol{X} = \{\boldsymbol{\beta}^{(d)}, \boldsymbol{f}^{(d)}; d = 1, \ldots, D-1\}$ the latent field, $\boldsymbol{\theta}_1 = \{\sigma_d^2, \gamma : d = 1, \ldots, D-1\}$ the hyperparameters corresponding to the likelihood, and $\boldsymbol{\theta}_2$ the hyperparameters corresponding to the functions f.

4.2 LNDM in R-INLA

R-INLA has been implemented in the sense that each data item is linked to one element of the Gaussian field. Although in this new INLA era, this condition disappears (Van Niekerk et al. 2023), it is still a challenge to fit models with multivariate likelihoods. Some approximations exist for Multinomial likelihood using the Poisson–Laplace trick (Baker 1994), or the Dirichlet likelihood converting the original Dirichlet observations into Gaussian pseudo-observations conditioned to the linear predictor (Martínez-Minaya et al. 2023). In our case, the main challenge is to estimate the variance-covariance matrix of the $\mathcal{ND}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, in particular, $p(\gamma \mid \boldsymbol{y})$. To do so, we adopt the strategy of modelling each alr-coordinate as if we were modelling multiple likelihoods (Krainski et al. 2018), and the covariance hyperparameter is estimated using independent random effects through the following well-known proposition.

Proposition 1 *Let z_d, $d = 1, \ldots, D-1$ be independent Gaussian random variables with different mean μ_d variances σ_d^2, and $u \sim \mathcal{N}(0, \gamma)$. Then, the multivariate random variable $\boldsymbol{y}$, defined as:*

$$\begin{aligned} y_1 &= z_1 + u, \\ y_2 &= z_2 + u, \\ \vdots &= \vdots \\ y_{D-1} &= z_{D-1} + u, \end{aligned} \tag{17}$$

follows a multivariate Gaussian with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ whose elements are:

$$\begin{aligned} \Sigma_{dd} &= \sigma_d^2 + \gamma, \ d = 1, \ldots, D-1, \\ \Sigma_{dj} &= \gamma, d \neq j. \end{aligned}$$

This proposition is simple but powerful, as with independent Gaussian distributions and a shared random effect between predictors, $p(\gamma \mid \boldsymbol{y})$ can be easily estimated. So, this structure fits perfectly in the context of LGMs. Thus, to estimate LNDM in **R-INLA**, we only need to add an individual shared random effect between linear predictors corresponding to the different alr-coordinates.

4.3 A simulated example

In this section, we exemplify, using a simulated scenario, the process of fitting CoDa using **R-INLA**. To elucidate, we initiate with a simplistic case featuring solely three categories

and one covariate. We presuppose that the impact of this covariate differs for each predictor. Subsequently, we designate this model as a Type II model. The model structure with which we operate in this example is:

$$alr(\boldsymbol{Y}) \sim \mathcal{ND}((\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}), \boldsymbol{\Sigma}), \tag{18}$$

$$\boldsymbol{\mu}^{(d)} = \boldsymbol{X}\boldsymbol{\beta}^{(d)}, \tag{19}$$

where $\boldsymbol{X}_{N\times 2}$ is a matrix with ones in the first column and values of the covariate simulated from a Uniform distribution between -0.5 and 0.5. Four different parameters compose the model, and they form the latent field: $\boldsymbol{X} = \{\beta_0^{(1)}, \beta_0^{(2)}, \beta_1^{(1)}, \beta_1^{(2)}\}$. Moreover, three different hyperparameters are included in the model and they form the set of hyperparameters $\boldsymbol{\theta} = \{\sigma_1^2, \sigma_2^2, \gamma\}$.

4.3.1 Data simulation

In this part of the manuscript, we present an example of how simulation can be conducted. First at all, we define the values of the hyperparameters and we compute the correlation matrix in $\boldsymbol{\Sigma}$. $N = 1000$, $D = 3$, $\sigma_1^2 = 0.5$, $\sigma_2^2 = 0.4$ and $\gamma = 0.1$ are the choosen values for the simulation.

```
R> D <- 3
R> N <- 1000
R> sigma2 <- c(0.5, 0.4)
R> cov_param <- 0.1
R> sigma_diag <- sqrt(sigma2 + cov_param)
```

Correlation matrix can also be easily computed. This matrix is formed for $((D-1)^2 - (D-1))/2$ values out of the diagonal.

```
R> rho <- diag(1/sigma_diag)
    diag(1/sigma_diag)
R> diag(rho) <- 1
```

Next step is simulating the covariate.

```
R> x = runif(N)-0.5
```

Subsequently, with fixed betas, $\beta_0^{(1)} = -1$, $\beta_1^{(1)} = 1$, $\beta_0^{(2)} = -1$, $\beta_1^{(2)} = 2$, we construct the values for the two linear predictors.

```
R> betas = matrix(c(-1, 1,
         -1, 2), nrow = D-1, byrow = TRUE)
R> X <- data.frame(1, x)
R> lin.pred <- X
```

Simulating from a multivariate Gaussian with the structure previously constructed is the next step. And with it, we obtain the alr-coordinates.

```
R> Sigma <- matrix(sigma_diag, ncol = 1)
    matrix(sigma_diag, nrow = 1)
R> Sigma <- Sigma*rho
R> lin.pred
  apply(., 1, function(z)
    MASS::mvrnorm( n  = 1,
              mu = z,
              Sigma = Sigma))
  t(.)-> alry
```

Finally, we move to the simplex assuming the third category the reference one. the output is a matrix with the response variable summing their rows up to one. We create a data.frame in order to keep the CoDa, the alr-coordinates and the covariate x. In Fig. 1, CoDa generated and alr-coordinates have been depicted.

```
R> y.simplex <- compositions::alrInv(alry)
R> data <- data.frame(alry, y.simplex, x)
```

4.3.2 Preparing data for being introduced in R-INLA

In this section, the most labor-intensive step is preparing the database to be input into **R-INLA**. To do this, we make use of structures like `inla.stack`. In this structure, we need to include the multiresponse variable, where we incorporate different alr-coordinates. Additionally, we input the covariates, indicating which alr-coordinate they affect, along with an index that assist us in introducing the shared random effect for estimating the hyperparameter γ. So, we start defining such index.

```
R> id.z <- 1:dim(alry)[1]
```

Posteriorly, we extent the dataset for constructing the multivariate response which is a matrix with dimension $(N \times (D-1)) \times (D-1)$, being the first column formed for the first alr-coordinate en N first rows, and NAs in the rest; the second column formed by the second alr-coordinate in the positions $(N+1)$:(2N), and NAs in the rest, and so on.

```
R> data_ext <- data
  tidyr::pivot_longer(., cols = all_of
      (paste0("alry.", 1:(D-1))),
       names_to  = "y.names",
        values_to = "y.resp")
  .[order(ordered(.$y.names)),]
R> data_ext$y.names <- ordered(data_
  ext$y.names)

R> names_y <- paste0("alry.", 1:(D-1))

R> 1:length(names_y)
  lapply(., function(i){
    data_ext
      dplyr::filter(y.names == names_y
     [i]) -> data_comp_i
```

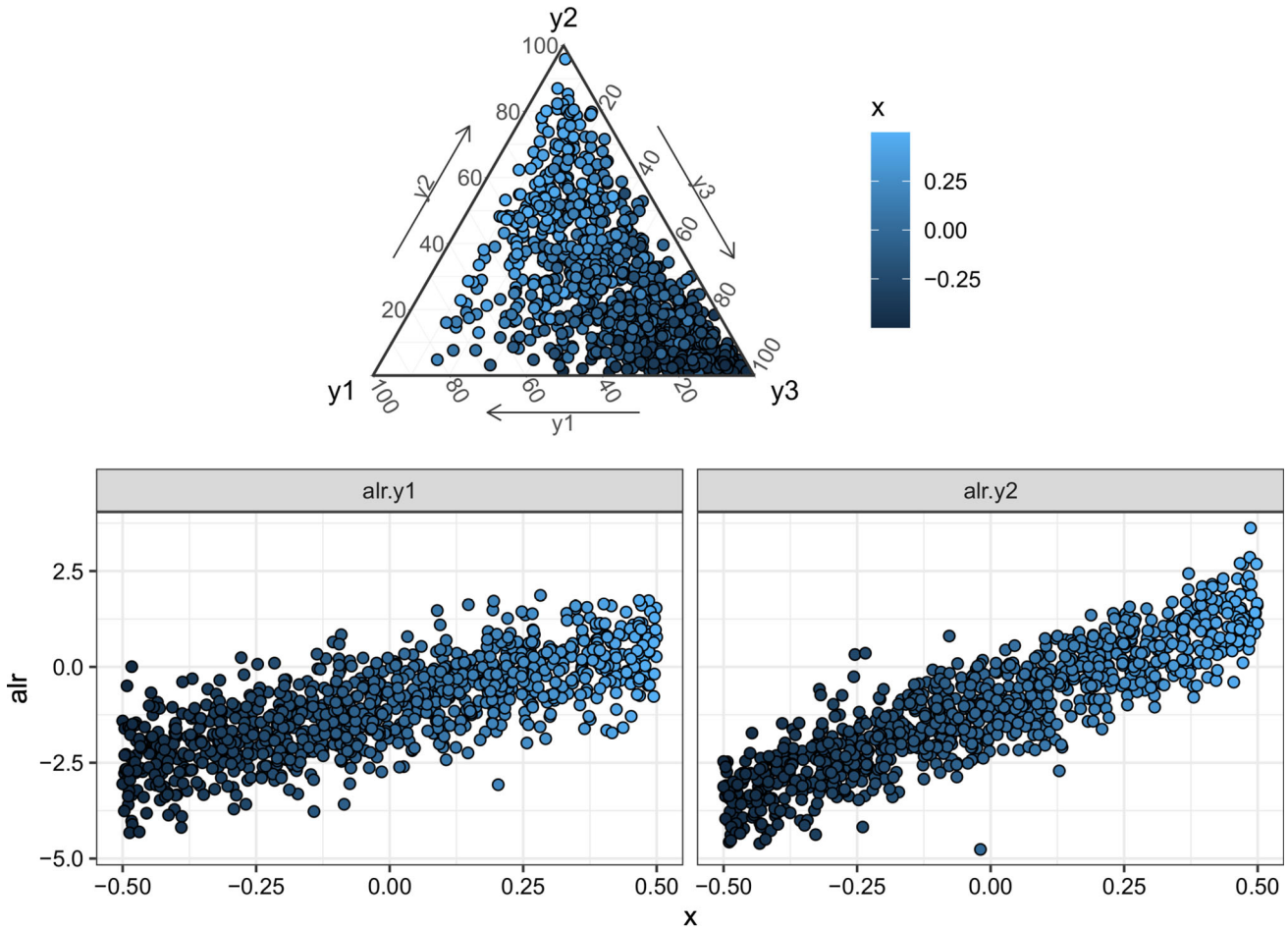

Fig. 1 Top: CoDa simulated represented in the simplex. Bottom: alr-coordiantes in terms of the generated covariate x

```
    #Response
    y_alr <- matrix(ncol = names_y
        length(.), nrow = dim(data_comp
         _i)[1])
    y_alr[, i] <- data_comp_i$y.resp
  }) -> y.resp

R> 1:length(names_y)
  lapply(., function(i){
    y_aux <- data_ext
      dplyr::select(y.resp, y.names)
      dplyr::filter(y.names == names
        _y[i])
      dplyr::select(y.resp)
      as.matrix(.)
    aux_vec <- rep(NA, (D-1))
    aux_vec[i] <- 1
    kronecker(aux_vec, y_aux)
  }) -> y_list

R> y_tot <- do.call(cbind, y_list)
```

```
R> y_tot


R>        [,1] [,2]
R> [1,] -1.580   NA
R> [2,] -1.345   NA
R> [3,] -1.735   NA
R> [4,] -1.012   NA
R> [5,] -0.584   NA
R> [6,] -0.041   NA
```

In the model, covariates are included as random effects with big variance. So, we need the values of the covariates, and also, an index indicating to which alr-coordinate it belongs.

```
R> variables <- c("intercept", data
                  dplyr::select(starts_
                    with("x"))
                  colnames(.))
R> id.names <- paste0("id.", variables)
R> id.variables <- rep(data_ext$y.names
                       as.numeric(.),
```

```
                    length(variables))
  matrix(., ncol = length(variables),
    byrow = FALSE)
R> colnames(id.variables) <- id.names
```

Finally, we create the `inla.stack` for estimation, and we are ready for fitting the model.

```
R> stk.est <- inla.stack(data    =
  list(resp = y_tot),
   A       = list(1),
   effects = list(cbind(data_ext
   dplyr::select(starts_with("x")),
  data_ext
  dplyr::select(starts_with("id.z")),
  id.variables,
  intercept = 1)),
  tag     = 'est')
R> colnames(id.variables) <- id.names
```

4.3.3 Fitting the model

For fitting the model, it is required to define priors for the parameters and hyperparameters. Prior considered for the parameters are the default ones used in **R-INLA**. However, PC-priors (Simpson et al. 2017) are considered for the standard deviations and the root square of the covariance parameter γ, in particular, PC-prior$(1, 0.01)$ were used for σ_1, σ_2 and $\sqrt{\gamma}$. So, the required formula to be introduced in **R-INLA** was:

```
R> list_prior <- rep(list(list(prior =
   "pc.prec",
    param = c(1, 0.01))), D-1)

R> formula.typeII <- resp ~ -1 +
  f(id.intercept, intercept,
    model   = "iid",
    initial = log(1/1000),
    fixed   = TRUE) +
  f(id.x, x,
    model   = "iid",
    initial = log(1/1000),
    fixed   = TRUE) +
  f(id.z,
    model = "iid",
    hyper = list(prec = list(prior =
  "pc.prec",
                  param = c(1, 0.01))),
  constr = TRUE)
```

and the call to **R-INLA**:

```
model.typeII <- inla(formula.typeII,
family          = rep("gaussian", D-1),
data            = inla.stack.data
(stk.est),
control.compute = list(config = TRUE),
control.predictor = list(A = inla.stack
.A(stk.est),
   compute = TRUE),
control.family = list_prior,
verbose = FALSE)
```

In Figs. 2 and 3, marginal posterior distributions jointly with the simulated value are depicted showing that we were able to recover the original value.

5 Spatial LNDM and model selection

Once the LNDM is defined, a particular focus lies on how more intricate structures within the linear predictor can be accommodated within the **R-INLA** framework. Furthermore, another issue pertains to model selection. Hence, this section is dedicated to spatial LNDMs and the utilization of measures such as Deviance Information Criteria (DIC), Watanabe Akaike information criterion (WAIC), and LCPO for model selection.

5.1 Spatial LNDMs

Of particular interest are the LNDMs in the spatial context. The analysis of the spatial process refers to the analysis of data collected in space. Space can be indexed over a discrete domain or a continuous one. So, spatial statistics is traditionally divided into three main areas depending on the type of problem and data: lattice data, Geostatistics and point patterns. For a review of models of different types of spatial data, see Haining and Haining (2003) and Cressie and Wikle (2015). When a spatial effect has to be included in the model, it is common to formulate mixed-effects regression models in which the linear predictor is made up of a trend plus a spatial variation, the spatial effect being modelled with correlation random effects and matching perfectly the structure presented in Eq. (16).

R-INLA provides many options when implementing Gaussian latent spatial effects (Gómez-Rubio 2020), including intrinsic conditional autoregressive models (iCAR) or conditional autoregressive models (CAR) for areal data (Besag et al. 1991) or spatial effect with Matérn covariance function for continuous processes (Lindgren et al. 2011). In this manuscript, we focus in the last, but it can be easily applicable to other latent Gaussian effects.

The Matérn covariance function is one of the most widely used in Geostatistics due to its flexibility. Although initially it could not be directly incorporated into the **R-INLA** structure,

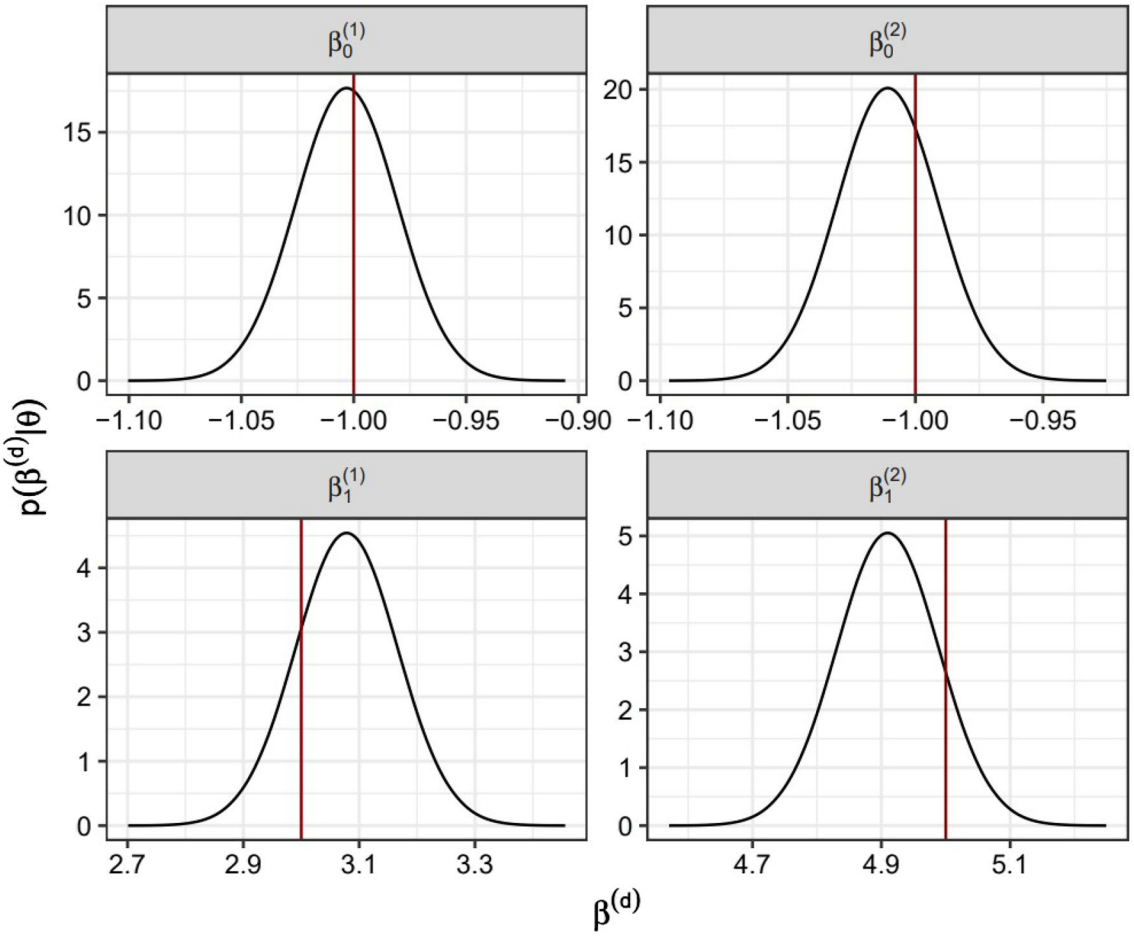

Fig. 2 Marginals posterior distributions for the fixed effects. Vertical lines represent the real values

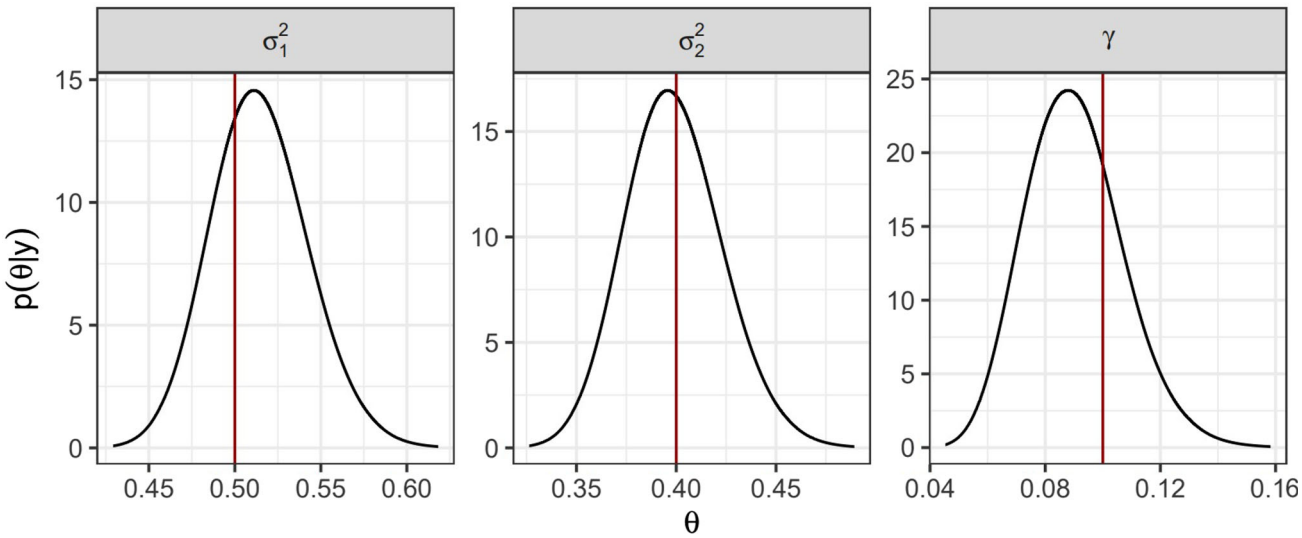

Fig. 3 Marginals posterior distributions for the hyperparameters. Vertical lines represent the real values

in Lindgren et al. (2011) introduced a solution through the SPDE module, approximating the spatial latent effect with a Matérn function as a solution to a stochastic partial differential equation using the finite element method (FEM). Since then, this methodology has been applied in numerous scientific articles across different areas (Martínez-Minaya et al. 2018).

These effects can be easily included in the LNDM. As we are adopting a multiple likelihood modelling strategy, we make use of the features that **R-INLA** provides for fitting multiple likelihoods in a jointly way. The `copy` command is intended share random effects, i.e., to use the same latent effect in different linear predictors. It also, allows to share exactly the same latent effect but adding a proportionality hyperparmeter. The `replicate` feature provides a way to add different random effects per linear predictor sharing the same hyperparameters. For details about its implementation, we refer the reader to the website https://www.r-inla.org/ and books by Krainski et al. (2018) and Gómez-Rubio (2020).

Applying these principles and emphasizing both fixed effects and continuous spatial random effects, the examples presented in this paper follow a systematic framework that leads to the development of eight distinct model types. Then, the model structure employed for the remainder of the paper is as follows:

$$alr(\boldsymbol{Y}) \sim \mathcal{ND}((\boldsymbol{\mu}^{(1)}, \ldots, \boldsymbol{\mu}^{(d)}), \boldsymbol{\Sigma}), \tag{20}$$

$$\boldsymbol{\mu}^{(d)} = \boldsymbol{X}\boldsymbol{\beta}^{(d)} + \boldsymbol{\omega}^{(d)}, d = 1, \ldots, D-1, \tag{21}$$

$\boldsymbol{\mu}^{(d)} = (\mu_1^{(d)}, \ldots, \mu_N^{(d)})$ being the different linear predictor for the nth observation in the dth alr-coordinate, and $\boldsymbol{X}_{N\times(M+1)}$ the design matrix, containing 1s in the first column if intercepts are considered in the model. $\boldsymbol{\omega}^{(d)}$ represents the spatial random effect with Matérn covariance for each dth alr-coordinate, $\boldsymbol{\omega}^{(d)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}^{-1}(\sigma_{\boldsymbol{\omega}}, \phi))$, depending on the standard deviation of the spatial effect $\sigma_{\boldsymbol{\omega}}$ and its range ϕ. $\boldsymbol{\beta}^{(d)}_{(M+1)\times 1}$ is the parameter vector corresponding to the fixed effects. The latent field is composed of the parameters corresponding to the fixed effects and the realisations of the random field.

$$\boldsymbol{X} = \{\boldsymbol{\beta}^{(d)}, \boldsymbol{\omega}^{(d)} : d = 1, \ldots, (D-1)\}.$$

In contrast, $\boldsymbol{\theta}_1 = \{\sigma_d^2, \gamma : d = 1, \ldots, (D-1)\}$ are the hyperparameters corresponding to the likelihood, and $\boldsymbol{\theta}_2 = \{\sigma_{\boldsymbol{\omega}}, \phi\}$ are the hyperparameters corresponding to the spatial random effect. Together they form the field of hyperparameters. Gaussian priors are usually assigned for the fixed effects and PC-priors for the hyperparameters (Simpson et al. 2017).

Based on the model structure defined in Eq. (21), **R-INLA** offers flexibility by allowing us to introduce fixed effects and random effects in different ways with the features previously explained. For the fixed effects, two different assumptions between parameters of the different alr-coordinates are plausible. The first is under the assumption that the effect of the m-covariate is the same for the different alr-coordinates, i.e. they are sharing the same parameter for fixed effects: $\beta_m^{(d)} = \beta_m^{(k)}, d \neq k$ and $d, k = 1, \ldots, (D-1), m = 0, \ldots M$. We denote it by β_m. For the second, we consider that the effect of the m-covariate could be different for each alr-coordinate. Note that this one is more general, as it includes the case where the effects are equal and also the case where we do not have the same covariates in each linear predictor. We denote them by $\beta_m^{(d)}$.

With regard to the random effects, we distinguish three different cases. The first one considers that the spatial random field is the same for all the linear predictors, i.e. $\boldsymbol{\omega}^{(d)} = \boldsymbol{\omega}^{(k)}$, $d \neq k$ and $d, k = 1, \ldots, (D-1)$. They share exactly the same spatial term. So, we denote it by $\boldsymbol{\omega}$ as it is not dependent on the alr-coordinates predictor. The second case is under the assumption that the spatial fields are proportional, in other words, $\boldsymbol{\omega}^{(d)} = \alpha^{(d)}\boldsymbol{\omega}^{(k)}$, $d \neq k$ and $d, k = 1, \ldots, (D-1)$. We denoted it by $\boldsymbol{\omega}^{(*d)}$. Finally, the third case states that the realisation of the spatial random effect is different for each linear predictor. However, they share the same hyperparameters, i.e. $\boldsymbol{\omega}^{(d)} \neq \boldsymbol{\omega}^{(k)}$, $d \neq k$, and $d, k = 1, \ldots, (D-1)$, where $\boldsymbol{\omega}^{(d)} \sim \mathcal{N}(0, \boldsymbol{Q}^{-1}(\sigma_{\boldsymbol{\omega}}, \phi))$. We denote it by $\boldsymbol{\omega}^{(d)}$.

By combining fixed and random terms, we reach eight different structures for the linear predictors (See Table 1 for details about the latent field and hyperparameters):

- Type I: share the same parameters for fixed effects, and do not include spatial random effects.
- Type II: have different parameters for fixed effects, and do not include spatial random effects.
- Type III: share the same parameters for fixed effects, and share the same spatial effect.
- Type IV: have different parameters for fixed effects, and share the same spatial effect.
- Type V: share the same parameters for fixed effects, and the spatial effects between linear predictors are proportional. Realisations of the spatial field are the same, but a proportionality hyperparameter is added in two of the three linear predictors.
- Type VI: have different parameters for fixed effects, and the spatial effects between linear predictors are proportional. Realisations of the spatial field are the same, but a proportionality hyperparameter is added in two of the three linear predictors.
- Type VII: share the same parameters for fixed effects, and different realisations of the spatial effect for each linear predictor. Although realisations of random effects are different, they share the same hyperparameters.
- Type VIII: have different parameters for fixed effects, and different realisations of the spatial effect for each

linear predictor. Although realisations of random effects are different, they share the same hyperparameters.

5.2 Model selection and validation

Regarding the model selection process, sometimes there are a large number of models resulting from all the possible combinations of covariates, and combining them with the possible latent effects that can be incorporated increases the number of possibilities exponentially. **R-INLA** has proved to be fast enough to compute huge numbers of models as well as different measures to make the model selection process feasible. Such measures include Deviance Information Criteria (Spiegelhalter et al. 2002, DIC), defined as a hierarchical modelling generalisation of the Akaike information criterion (AIC); Watanabe Akaike information criterion (Watanabe and Opper 2010; Gelman et al. 2014, WAIC), which is the sum of two components: one quantifying the model fit and the other evaluating the model complexity; or the cross-validation measure conditional predictive ordinate (CPO) for evaluating the predictive capacity and its log-score (Pettit 1990; Roos and Held 2011, LCPO). The models with the lowest values of DIC, WAIC or LCPO have preference over the rest.

However, **R-INLA** is programmed to handle univariate likelihoods, and the variability added with the inclusion of the new random effect is not being considered when the calculation of the deviance is computed. This affects the computation of the DIC and WAIC. So, an additional process is needed to calculate DIC and WAIC when the response variable follows a multivariate normal distribution. This process must be able to incorporate the elements that are off the diagonal of the variance–covariance matrix. To achieve this, a post-processing of the model is performed for obtaining samples of the jointly posterior distributions using the feature `inla.posterior.sample` function, and the likelihood of the multivariate normal distribution is calculated. The remaining calculations for DIC are done following the formula defined in Spiegelhalter et al. (2002), meanwhile, WAIC is computed following the formula in Watanabe and Opper (2010). These two ways have been implemented in two different functions in `R`. The functions are called `DIC.mult` and `WAIC.mult` and are available in the repository https://github.com/jmartinez-minaya/INLAcomp.

The same does not apply to the CPO, as it is based on the posterior predictive distribution. In Appendix A, there is a proof of why the CPO is not affected by the approach we propose here. However, we believe that the CPO cannot be calculated in the same way when dealing with CoDa, and therefore, we propose a new definition.

5.2.1 CPO

In the context of CoDa cross-validation process, excluding a category from a CoDa point may not make sense, as we know that CoDa have a constraint: their sum must be 1. This implies that the remaining categories provide valuable information about the category we are excluding. One might think that working in the log-ratio coordinates could alleviate this issue, but that is not the case. The reference category is present in all the log-ratios, and thus we encounter a similar situation. At that point, the remaining log-ratio coordinates provide information about the category we have removed during cross-validation. In this manner, the concept of friendship emerges. Consequently, we can assert that the first alr-coordinate of individual n is friend of the second alr-coordinate of individual n, and is thereby contributing information. Hence, in order to conduct cross-validation for individual n and alr-coordinate d, it is necessary to exclude the values from all alr-coordinates pertaining to that individual. Accordingly, we can define the CPO for the nth data point and dth alr-coordinate as:

$$\text{CPO}_n^{(d)} = \int p(alr(\boldsymbol{y})_n^{(d)} \mid \boldsymbol{X}, \boldsymbol{\theta}) \, p(\boldsymbol{X}, \boldsymbol{\theta} \mid alr(\boldsymbol{y})_{-n}^{\bullet}) \, d\boldsymbol{X} d\boldsymbol{\theta}, \tag{22}$$

being $alr(\boldsymbol{y})_n^{(d)}$ the observed vector for the n-data point and the d alr-coordinate, and $alr(\boldsymbol{y})_{-n}^{\bullet}$ represents the observed data in alr-coordinates ($N-1$ data points with $D-1$ components for data point) excluding the n data point with its corresponding $D-1$ alr-coordinates. We then easily compute the log-score (Gneiting and Raftery 2007) as:

$$\text{LCPO} = -\frac{1}{N \cdot (D-1)} \sum_{d=1}^{D-1} \sum_{n=1}^{N} \log\left(\text{CPO}_n^{(d)}\right). \tag{23}$$

6 Continuos spatial data: a simulation study

The goals of this simulation are twofold. Firstly, we seek to assess the reliability of model selection criteria previously presented. As we have pointed out, these metrics play a crucial role in identifying the model that best represents the underlying process. Secondly, we aim to demonstrate capability of **R-INLA** to accurately recover the initial parameters.

6.1 Simulated data

We conducted a simulation of a spatial LNDM Type VIII renowned for its high flexibility as the fixed effects vary by linear predictor, and spatial effects realizations differ accordingly. The simulation involved one covariate, simulated from

Table 1 Different structures included in the model in an additive way with their corresponding latent field and the hyperparameters to be estimated

Models	Predictor	Latent field ($\boldsymbol{X}$)	Hyperparameters ($\boldsymbol{\theta}$)
Type I	$\boldsymbol{X\beta}$	$\{\beta_0, \ldots, \beta_M\}$	$\{\sigma_d^2, \gamma\}$
Type II	$\boldsymbol{X\beta}^{(d)}$	$\{\beta_0^{(d)}, \ldots, \beta_M^{(d)}\}$	$\{\sigma_d^2, \gamma\}$
Type III	$\boldsymbol{X\beta} + \boldsymbol{\omega}$	$\{\beta_0, \ldots, \beta_M, \omega_1, \ldots, \omega_N\}$	$\{\sigma_d^2, \gamma, \sigma_{\boldsymbol{\omega}}, \phi\}$
Type IV	$\boldsymbol{X\beta}^{(d)} + \boldsymbol{\omega}$	$\{\beta_0^{(d)}, \ldots, \beta_M^{(d)}, \omega_1, \ldots, \omega_N\}$	$\{\sigma_d^2, \gamma, \sigma_{\boldsymbol{\omega}}, \phi\}$
Type V	$\boldsymbol{X\beta} + \boldsymbol{\omega}^{*(d)}$	$\{\beta_0, \ldots, \beta_M, \omega_1, \ldots, \omega_N\}$	$\{\sigma_d^2, \gamma, \sigma_{\boldsymbol{\omega}}, \phi, \alpha^{(1)}, \ldots, \alpha^{(D-2)}\}$
Type VI	$\boldsymbol{X\beta}^{(d)} + \boldsymbol{\omega}^{*(d)}$	$\{\beta_0^{(d)}, \ldots, \beta_M^{(d)}, \omega_1, \ldots, \omega_N\}$	$\{\sigma_d^2, \gamma, \sigma_{\boldsymbol{\omega}}, \phi, \alpha^{(1)}, \ldots, \alpha^{(D-2)}\}$
Type VII	$\boldsymbol{X\beta} + \boldsymbol{\omega}^{(d)}$	$\{\beta_0, \ldots, \beta_M, \omega_1^{(d)}, \ldots, \omega_N^{(d)}\}$	$\{\sigma_d^2, \gamma, \sigma_{\boldsymbol{\omega}}, \phi\}$
Type VIII	$\boldsymbol{X\beta}^{(d)} + \boldsymbol{\omega}^{(d)}$	$\{\beta_0^{(d)}, \ldots, \beta_M^{(d)}, \omega_1^{(d)}, \ldots, \omega_N^{(d)}\}$	$\{\sigma_d^2, \gamma, \sigma_{\boldsymbol{\omega}}, \phi\}$

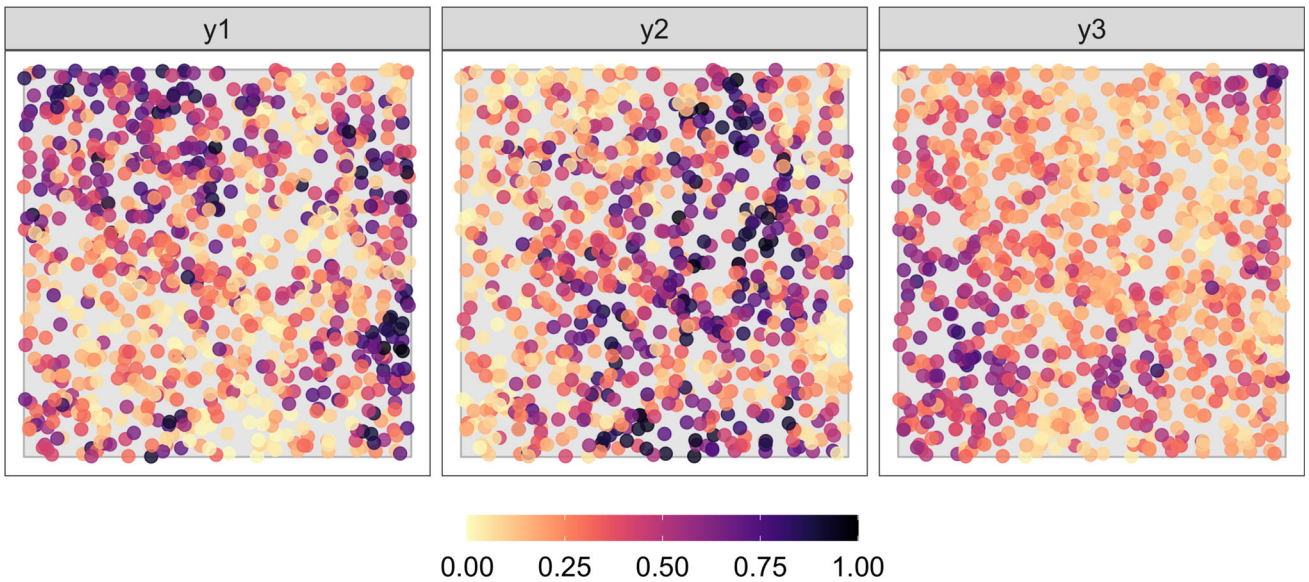

Fig. 4 CoDa simulated. Proportion per category

a Uniform distribution between -0.5 and 0.5; two different realizations of a Matérn field in the square space [0,10]$\times$ [0,10] with range $\phi = 4$ and $\sigma_\omega = 1$ (See Fig. 6); one thousand observations ($N = 1000$) and three dimensions ($D = 3$). Given that $D = 3$, applying the alr transformation yields two linear predictors. In the context of Type VIII and considering we simulated only one covariate, we are tasked with estimating two parameters, denoted as $\beta_1^{(1)}$ and $\beta_1^{(2)}$. These parameters were pre-set to specific values: -2.27 and -2.3 respectively. Turning our attention to the likelihood hyperparameters, we encounter two variance hyperparameters σ_1^2 and σ_2^2 and one covariance parameter γ. For this simulation, these hyperparameters were fixed at predetermined values 0.32, 0.59 and 0.1. Resulting data simulation is depicted in Fig. 4 and the alr-coordinates using the third category as reference are displayed in Fig. 5. We selected the third category as reference as it was the one whose logarithm had the lowest variance.

6.2 Model selection

The simulation originates from the Type VIII model, and we sought to fit alternative model types (refer to Table 1). Subsequently, we computed the DIC, WAIC, and LCPO for each model. Results are depicted in Table 2. Upon analysis, it is evident that, in all three cases, the Type VIII model consistently exhibits the best fit to our simulated data. This conclusion is supported by consistently smaller values across all three evaluation metrics.

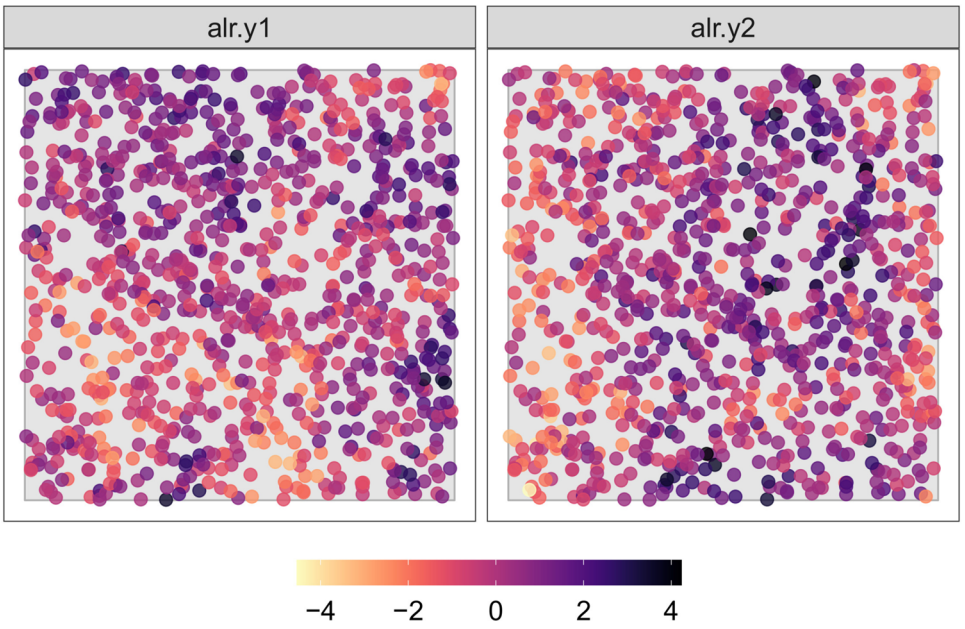

Fig. 5 Additive log-ratio transformation of CoDa using the third category as the reference one

Table 2 LNDMs with their corresponding DIC, WAIC and LCPO

Models	Predictor	DIC	WAIC	LCPO
Type I	$\boldsymbol{X\beta}$	6822.373	6822.132	1.704
Type II	$\boldsymbol{X\beta}^{(d)}$	6265.526	6265.444	1.571
Type III	$\boldsymbol{X\beta}+\boldsymbol{\omega}$	6221.710	6206.081	1.546
Type IV	$\boldsymbol{X\beta}^{(d)}+\boldsymbol{\omega}$	5585.697	5569.343	1.386
Type V	$\boldsymbol{X\beta}+\boldsymbol{\omega}^{*(d)}$	6068.076	6072.944	1.517
Type VI	$\boldsymbol{X\beta}^{(d)}+\boldsymbol{\omega}^{*(d)}$	5397.224	5405.861	1.303
Type VII	$\boldsymbol{X\beta}+\boldsymbol{\omega}^{(d)}$	5772.775	5782.855	1.446
Type VIII	$\boldsymbol{X\beta}^{(d)}+\boldsymbol{\omega}^{(d)}$	**4694.686**	**4711.903**	**1.176**

In bold, the best model, whose DIC, WAIC, and LCPO values are the lowest

6.3 Parameters recovery

As previously discussed, the optimal model is the Type VIII model. This model comprises: 2 parameters corresponding to fixed effects, $\beta_1^{(1)}$ and $\beta_1^{(2)}$, and the realizations of the spatial random effects which form the latent Gaussian field ($\boldsymbol{X}$); 3 hyperparameters related to likelihood σ_1^2, σ_2^2 and γ, and 2 hyperparameters associated with spatial random effects which forms the set of hyperparameters ($\boldsymbol{\theta}$).

The 95% credible interval of the parameter $\beta_1^{(1)}$ is [2.103, 2.4] with a median value of 2.251. In contrast, for the parameter $\beta_1^{(2)}$, the 95% credible interval is $[-2.469, -2.086]$ with a median value of -2.277. Comparing these intervals with the true parameter values, -2.27 and 2.3 respectively, we conclude that estimation is accurate enough. A similar pattern emerges for the latent fields with Matérn covariance matrices. In Fig. 6, we depict the original spatial latent fields alongside the medians and estimated 95% credible intervals. Once again, we observe a reliable estimation. Finally, we examine the behavior of the hyperparameters. In Fig. 7, the posterior distributions of the hyperparameters are illustrated jointly with the true values. Once more, the estimations align well with the actual values. From these findings, we can conclude that the method is proficient in recovering the true parameter values effectively.

7 The case of *Arabidopsis thaliana*

This section is devoted to showing an application of continuous spatial LNDMs in a real setting.

7.1 The data and the model

We worked with a collection of 301 accessions of the annual plant *Arabidopsis thaliana* on the Iberian Peninsula. For each accession, the probability of belonging to each of the 4 genetic clusters (GC) inferred in Martínez-Minaya et al. (2019), namely, GC1, GC2, GC3 and GC4, were available (Fig. 8), their sum total being 1. We were interested in estimating the probability of membership, which in this particular context can be thought of as the habitat suitability for each genetic cluster. To do so, we employed LNDMs including climate covariates and spatial terms in the linear predictor. In particular, two bioclimatic variables were used to define the climatic part: annual mean temperature (*BIO*1) and annual precipitation (*BIO*12). The complete dataset was downloaded from the repository Martínez-Minaya et al.

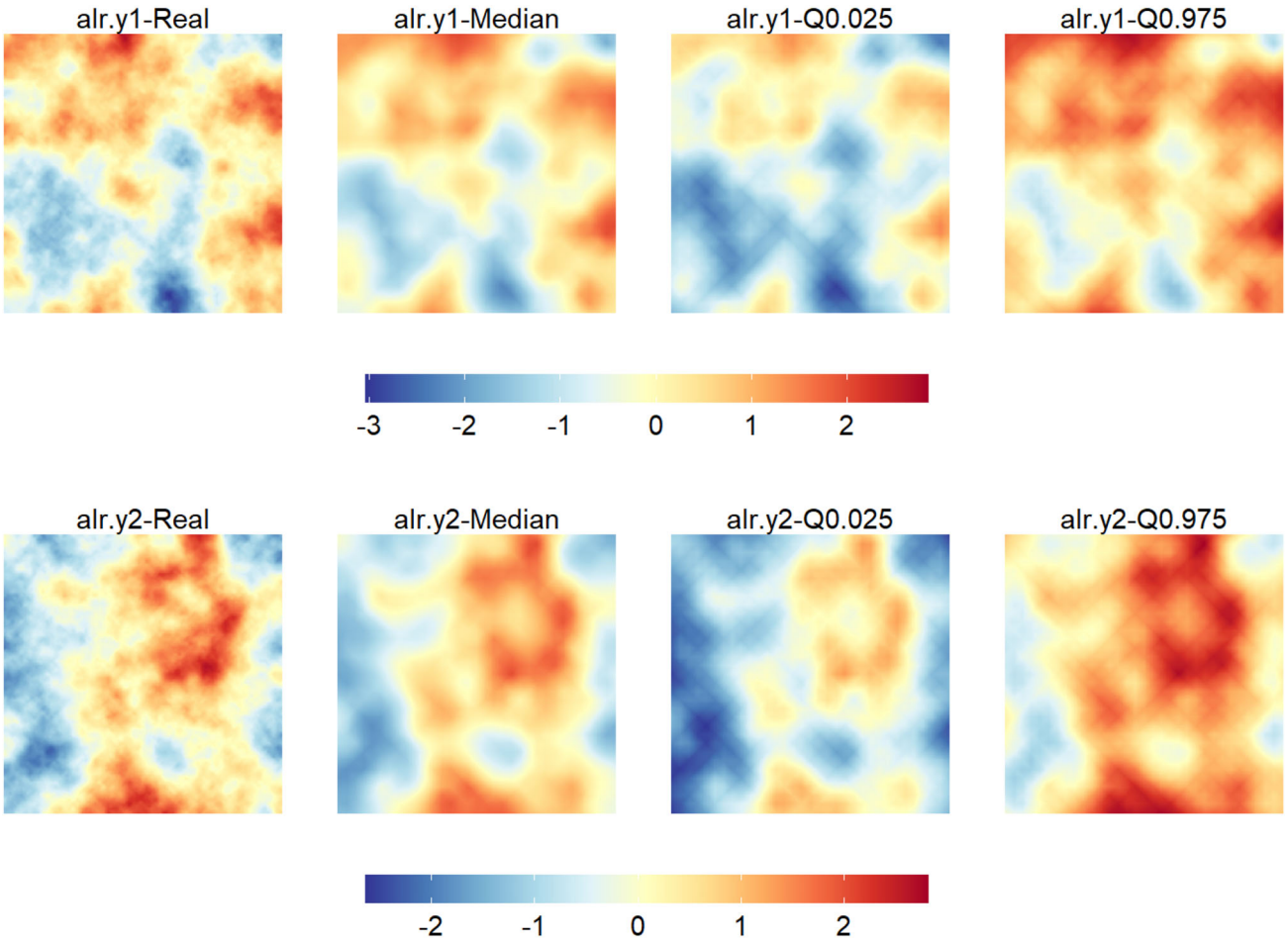

Fig. 6 Real values for the latent fields with Matérn covariance matrix used in the simulation. Median and 95% credible intervals for the the estimated field

(2019). Climate covariates were scaled before conducting the analysis.

As mentioned, four categories were employed in this problem: GC1, GC2, GC3 and GC4. So, we dealt with proportions in $\mathbb{S}^4$. To produce the LNDM, we selected GC4 as the reference category because it was the one whose logarithm had the lowest variance. We were thus dealing with a three dimensional $\mathcal{NM}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The transformed data is shown in Fig. 9.

7.2 Model selection, model fitting and prediction

Model selection was conducted including the intercept and also the two climatic covariates combining them with the spatial effects for the different structures presented in Table 1. 8 models were fitted and the DIC, WAIC and LCPO were computed (Table 3).

In view of the results in the model selection, and based on DIC and WAIC, we observed that the one with type VIII structure seemed to be the best at representing the process of interest. On the contrary, the LCPO indicates that the best model features a Type VI structure. However, as the difference is just 0.019, we proceeded with the model Type VIII for making the computation of the posterior distributions and also for making the predictions. Then, **R-INLA** allowed us to compute the posterior distribution for the fixed effects (Fig. 10) in each alr-coordinate. As we have argued in favour of alr, it is easy to interpret in terms of ratios.

If we focus on the covariate $BIO1$ (annual mean temperature), we observed that in presence of $BIO12$, it is relevant with a probability of 0.972 for the coefficient to be lower than 0 in the the first alr-coordinate, 0.99 for the second one, and 0.99 for the third. Therefore, in all three cases, we shall presume the covariate to be relevant and proceed to interpret the coefficients (Fig. 10). We observed that the ratio between the probability of belonging to GC1 and the probability of belonging to GC4 reduces by approximately 20% when the scaled covariate annual mean temperature increased by one unit. For the case of the ratio between the probability

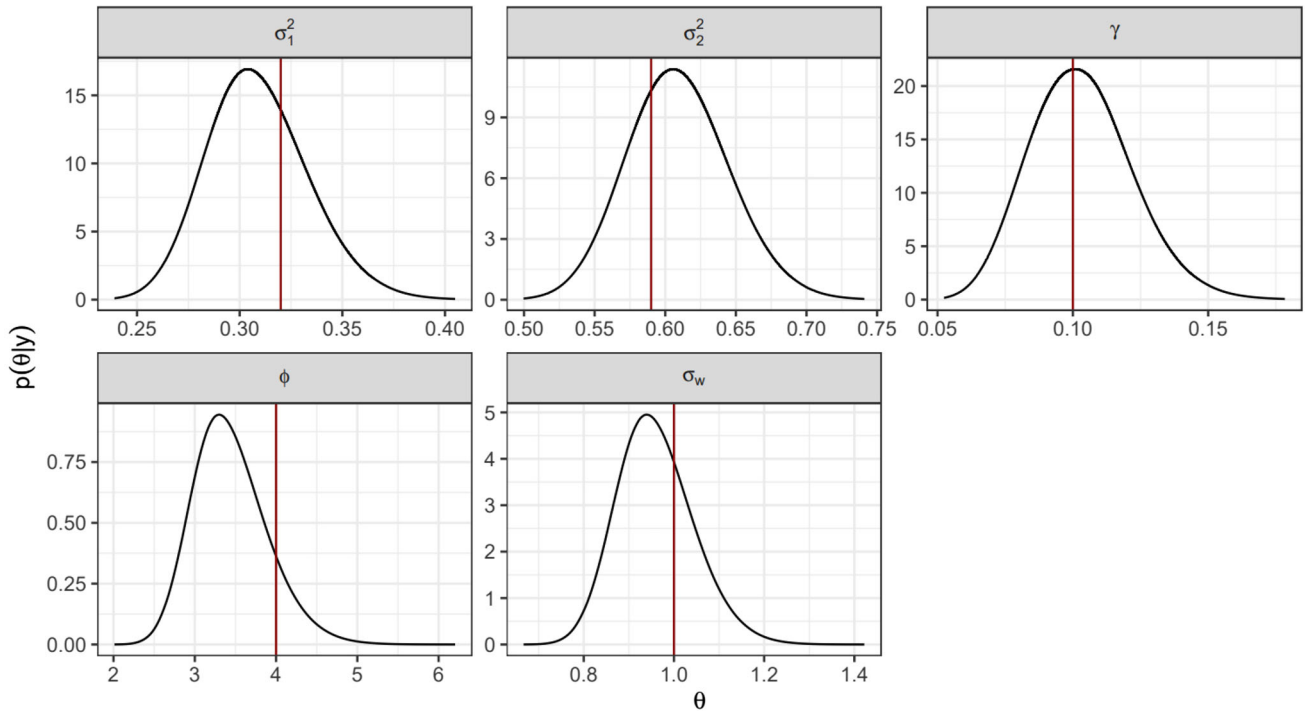

Fig. 7 Marginals posterior distributions for the hyperparameters. Vertical lines represent the real values

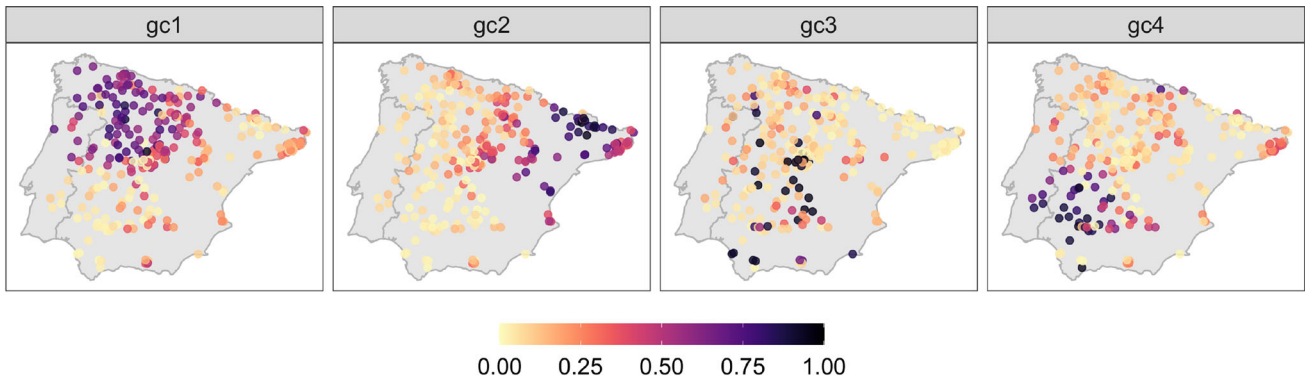

Fig. 8 Probability of membership of GC1, GC2, GC3 and GC4 on the Iberian Peninsula

of belonging to GC2 and GC4, it decreased by 32% when the scaled covariate annual mean temperature increased by one unit. Finally, the ratio between the probability of belonging to GC3 and GC4 decreased by 50% when the covariate annual mean temperature increased by one unit.

If we focus on the covariate present in the model $BIO12$ (annual precipitation), we noted that in presence of $BIO1$, it is relevant with a probability of 0.72 for the coefficient to be lower than 0 in the the first alr-coordinate. Not happen the same for the second and third alr-coordinate, as the probability to be lower than 0 are 0.43 and 0.46 respectively. As a result, we assume the covariate's relevance in the first alr-coordinate and we proceed to interpret its coefficient (Fig. 10). The ratio between the probability of belonging to GC1 and the probability of belonging to GC4 decreases by approximately 6% when the scaled covariate $BIO12$ increased by one unit and $BIO1$ remains constant.

With the method implemented here, we are able to make predictions not only on the alr-coordinates scale (Fig. 11), but also on the original scale (Fig. 12). If we focus on Fig. 11, we observe how in the north-west of Spain the ratio between the probability of belonging to GC1 and GC4 reached 12,

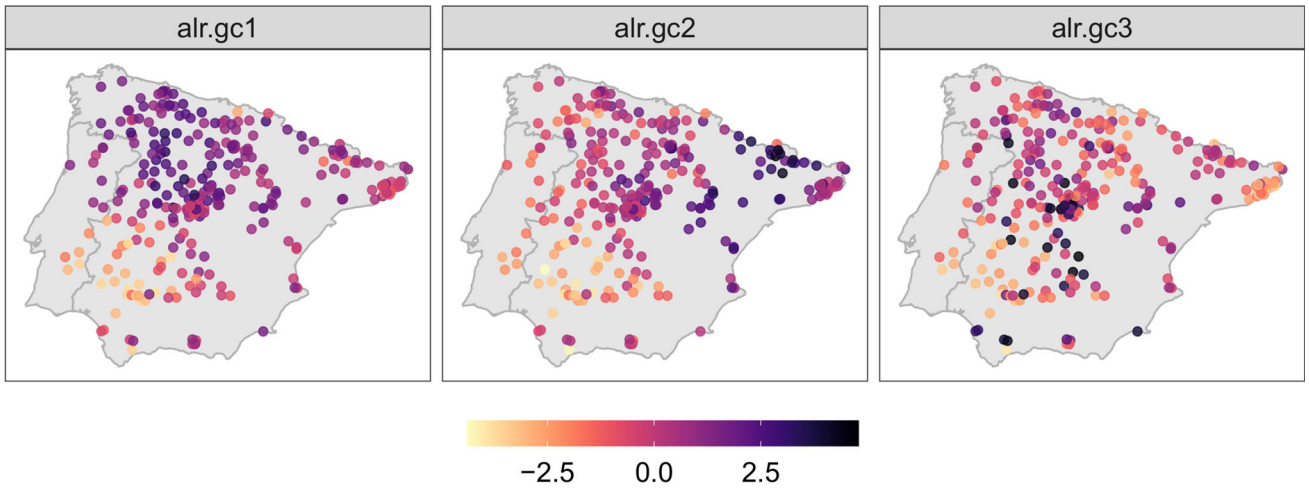

Fig. 9 Additive log-ratio transformation of the proportion of GC1, GC2, GC3 and GC4 on the Iberian Peninsula, using GC4 as the reference category

Table 3 LNDMs with their corresponding DIC, WAIC and LCPO

Models	Predictor	DIC	WAIC	LCPO
Type I	$\boldsymbol{X\beta}$	3353.890	3353.786	1.894
Type II	$\boldsymbol{X\beta}^{(d)}$	3294.560	3295.630	1.867
Type III	$\boldsymbol{X\beta}+\boldsymbol{\omega}$	3202.248	3208.578	1.786
Type IV	$\boldsymbol{X\beta}^{(d)}+\boldsymbol{\omega}$	3146.029	3154.323	1.758
Type V	$\boldsymbol{X\beta}+\boldsymbol{\omega}^{*(d)}$	3060.331	3060.484	1.470
Type VI	$\boldsymbol{X\beta}^{(d)}+\boldsymbol{\omega}^{*(d)}$	3004.190	3005.256	**1.383**
Type VII	$\boldsymbol{X\beta}+\boldsymbol{\omega}^{(d)}$	2752.965	2759.735	1.416
Type VIII	$\boldsymbol{X\beta}^{(d)}+\boldsymbol{\omega}^{(d)}$	**2741.096**	**2750.654**	1.402

meaning that at those points the probability of belonging to GC1 is 12 times greater than the probability of belonging to GC4. Something similar happened in the north-east of the Iberian Peninsula, where the probability of belonging to GC2 is 12 times greater than the probability of belonging to GC4. The case of the third alr-coordinate seems a bit different, and the greatest difference between the probability of belonging to GC3 and GC4 is found in the centre of the Iberian Peninsula.

Finally, it is accessible to compute marginal posterior distribution of the hyperparameters and, consequently, the covariance parameter between the alr-coordinates (Fig. 13).

8 Conclusions and future work

CoDa are becoming more and more common, especially in the context of genomics, and require increasingly powerful computational tools to be analysed. Thus, we believe that finding a way to include a likelihood that can deal with CoDa

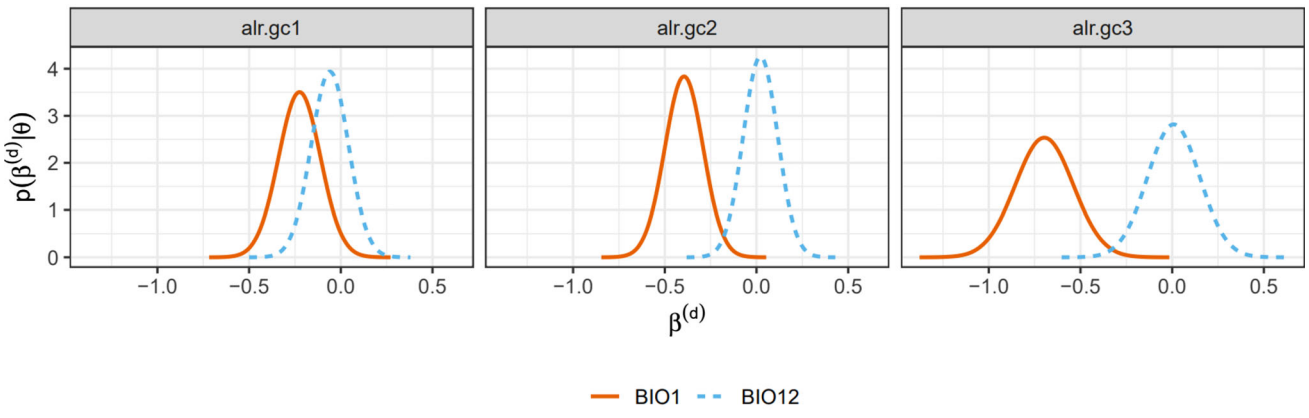

Fig. 10 Marginal posterior distribution for the parameters corresponding to the fixed effects or each of the alr-coordinates: $BIO2$ and $BIO12$

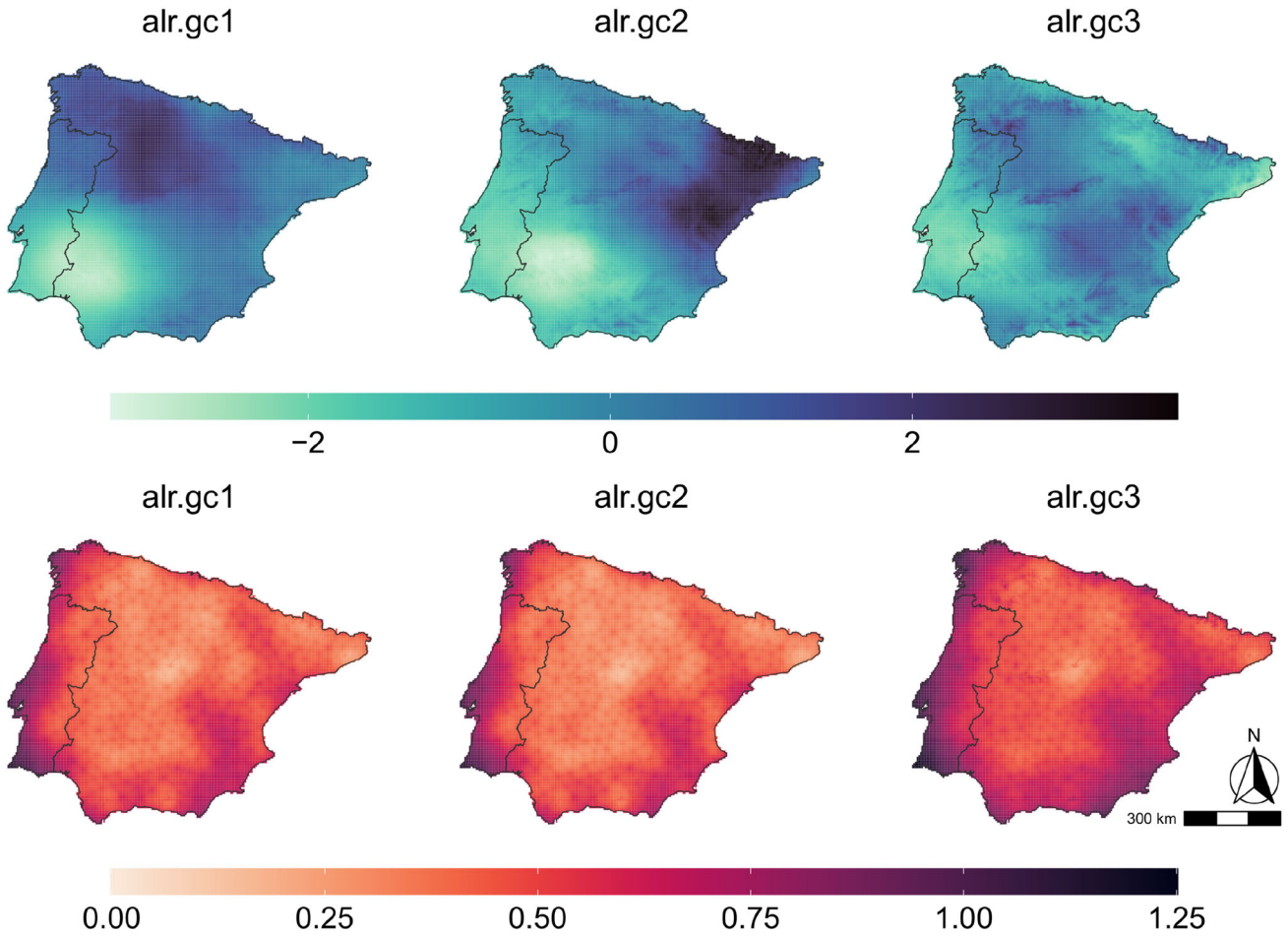

Fig. 11 Mean and standard deviation of the posterior predictive distribution for the alr-coordinates

in the context of LGMs can facilitate inference and predictions. That is why in this manuscript, we have introduced a different way to make inference on Bayesian CoDa analysis. By doing so, we attempt to include it in the context of LGMs, thereby making the range of possibilities that **R-INLA** offers available to the logistic-normal distribution with Dirichlet covariance likelihood.

The main idea underlying the proposed method is to approximate the multivariate likelihoods with univariate ones sharing an independent random effect that can be fitted by **R-INLA**, in particular, Gaussian likelihoods. This idea is similar to the one proposed for modelling Multinomial likelihood in **R-INLA**, where using the Poisson trick (Baker 1994) to reparameterise the model we need to fit independent Poisson observations, or the one proposed in (Martínez-Minaya et al. 2023) to approximate Dirichlet likelihoods using conditionally independent Gaussians. Simpson et al. (2016) also used a similar strategy, constructing a Poisson approximation to the true log-Gaussian Cox process likelihood and making it possible to carry out inference on a regular lattice over the observation window by counting the number of points in each cell. But this work does not intend to be a substitute for the **dirinla** package (Martínez-Minaya et al. 2023) or for the Bayesian ilr approach (Mota-Bertran et al. 2022): it is simply a viable alternative when dealing with CoDa that allows the estimation and prediction of very complex models in the context of CoDa. Furthermore, functions are provided for the computation of DIC and WAIC within the framework of **R-INLA**, accompanied by the definition of the CPO for CoDa.

We have reported an example in the field of Ecology, showing the potential of **R-INLA** when continuous spatial effects can be added in the linear predictor. We have exploited the options that **R-INLA** has available using tools in the context of multiple likelihoods, such as `copy` or `replicate` (Gómez-Rubio 2020). With them, our aim was to show practitioners the number of models that can be fitted in this context. Although here we have focused mainly on spatial processes, this tool can be easily applied in other contexts:

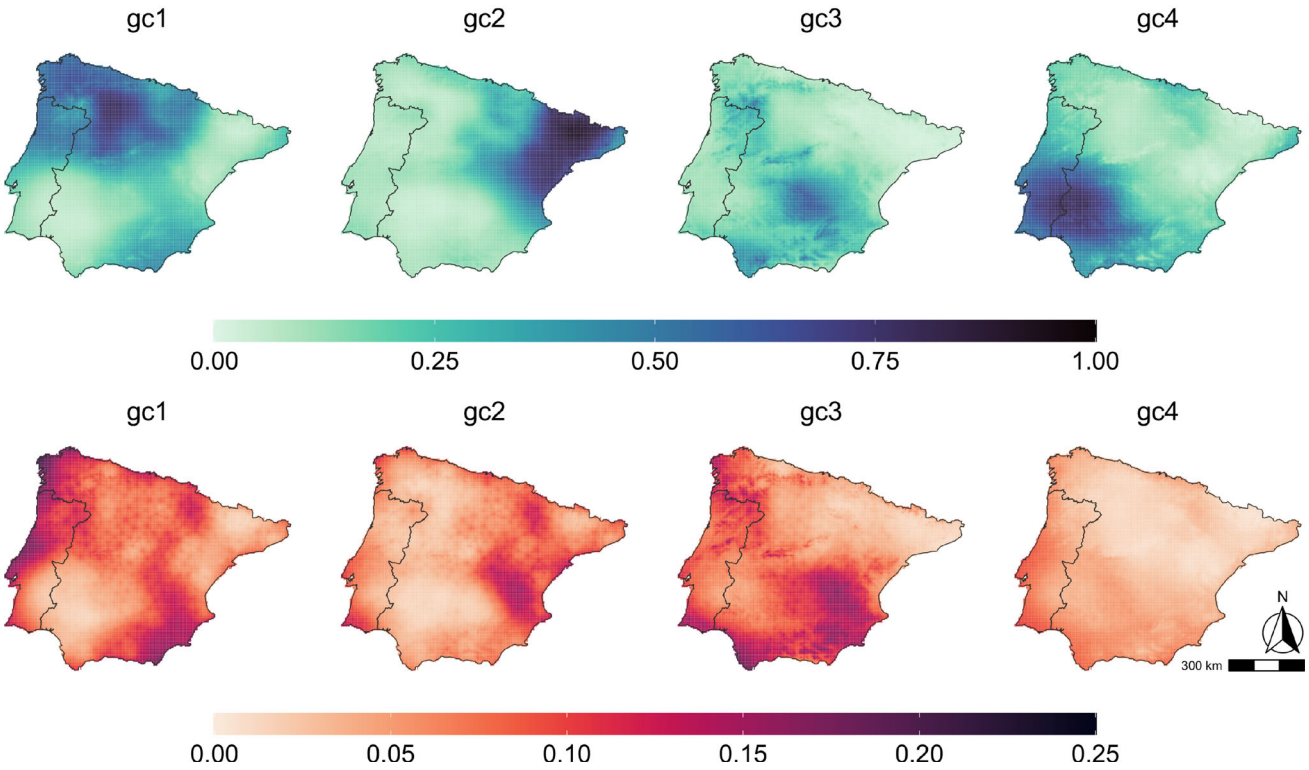

Fig. 12 Mean and standard deviation of the posterior predictive distribution for the probability of belonging to GC1, GC2, GC3 and GC4

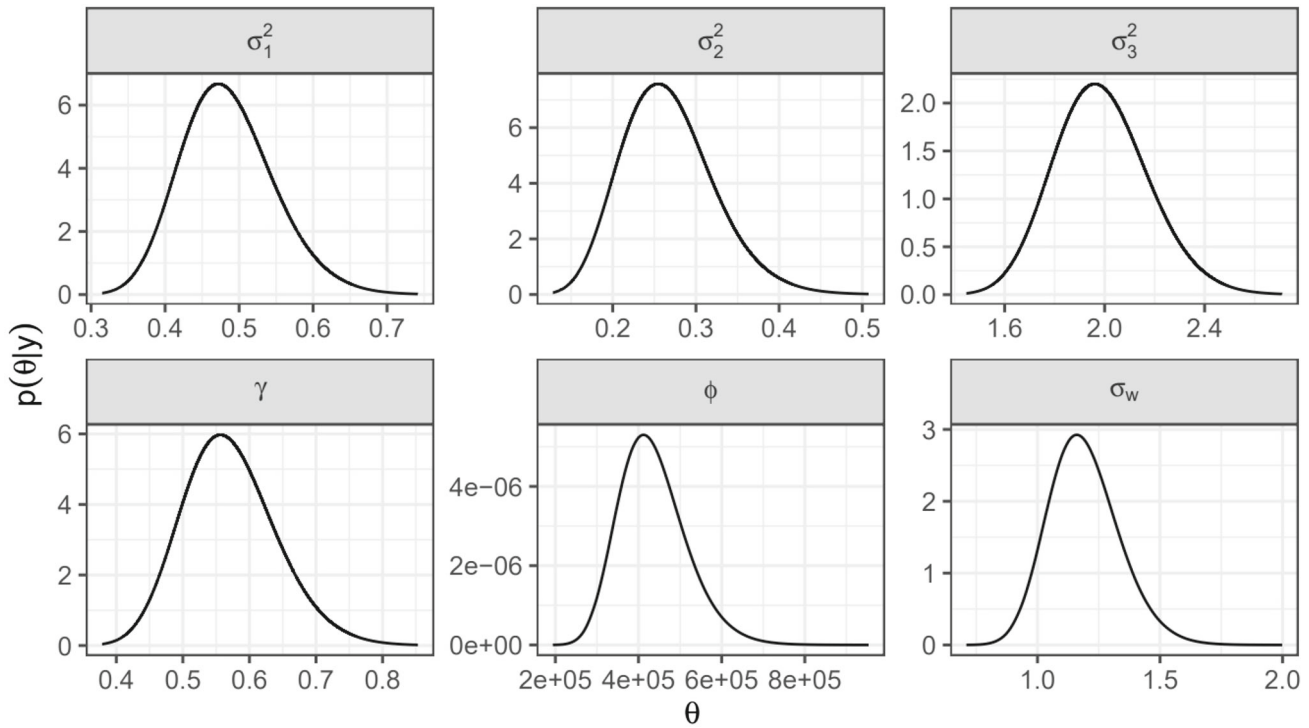

Fig. 13 Marginal posterior distribution for the hyperparameters of the model

temporal, spatiotemporal, etc., as long as we exprees the model in the context of LGMs.

Supplementary information

Code: The functions are stored in a R-package call INLAComp, it is on https://github.com/jmartinez-minaya/INLAcomp. The results shown in the paper are stored in https://jmartinez-minaya.github.io/supplementary.html.

Acknowledgements Joaquín Martínez-Minaya gratefully acknowledges the Ministry of Science, Innovation and Universities (Spain) for research project PID2020-115882RB-I00. Joaquín Martínez-Minaya also acknowledges for Funding for open access charge: CRUE-Universitat Politècnica de València.

Author Contributions JM-M and HR developed the methodology. JM-M developed models and performed simulations. JM-M wrote the main manuscript. JM-M and HR reviewed the manuscript.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Declarations

Conflict of interest The authors declare no Conflict of interest.

Appendix A: CPO computation in R-INLA

To verify that the CPO is not affected when fitting the model, it is enough to simplify the problem to the calculation of the posterior predictive distribution for the following two models:

MODEL I:

$$\begin{aligned} &y_i \sim \mathcal{N}(\mu_i, \sigma_y^2),\ i = 1, \dots N, \\ &\mu_i = \beta_0, \\ &\beta_0 \sim \mathcal{N}(0, \sigma_0^2), \end{aligned} \tag{A1}$$

MODEL II:

$$\begin{aligned} &y_i \sim \mathcal{N}(\mu_i, \sigma_{\boldsymbol{\epsilon}}^2), \\ &\mu_i = \beta_0 + \omega_i, \\ &\beta_0 \sim \mathcal{N}(0, \sigma_0^2) \\ &\omega_i \sim \mathcal{N}(0, \sigma_\omega^2). \end{aligned} \tag{A2}$$

Let assume for simplicity that $\sigma_y^2, \sigma_{\boldsymbol{\epsilon}}^2, \sigma_\omega^2$ and σ_0^2 are fixed numbers. Both models are equivalent, and $\sigma_{\boldsymbol{\epsilon}}^2 + \sigma_\omega^2 = \sigma_y^2$. However, as we have pointed out, in **R-INLA**, an additional process is required for computing DIC and WAIC. Nevertheless, it is not necessary for CPO, let's see why.

Proposition 2 *Let $y_i, i = 1, \dots, N$ independent realisations of a Gaussian distribution with mean μ_i and variance σ_y^2. The expressions (A1) and (A2) reflect two different ways of representing the process, although both models are equivalent. Thus, the CPO of the Model I and Model II are the same.*

Proof For proving that both CPOs are equal, it is enough to show that the posterior predictive distribution of both models is the same.

We start with a general linear mixed model following the expression in Eq. (10)

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{A_\omega \omega} + \boldsymbol{\epsilon}, \tag{A3}$$

being $\boldsymbol{X}$ and $\boldsymbol{A_\omega}$ design matrices, $\boldsymbol{\beta}$, a vector of fixed effects which follows a multivariate Gaussian prior distribution with mean $\boldsymbol{m}$ and covariance matrix $\boldsymbol{M}$, and $\boldsymbol{\omega}$ a vector of random effects which follows a multivariate Gaussian prior distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{G}$. The covariance matrices for $\boldsymbol{\omega}$ and $\boldsymbol{\epsilon}$ are assumed to be non singular, and positive definite, and $\boldsymbol{\omega}$ and $\boldsymbol{\epsilon}$ are independent.

Following Fahrmeir et al. (2013), if the covariance structures $\boldsymbol{G}$ and $\boldsymbol{R}$ are known, and $\boldsymbol{C} = (\boldsymbol{X}, \boldsymbol{U})$, $\boldsymbol{B} = \begin{pmatrix} \boldsymbol{M}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{G}^{-1} \end{pmatrix}$, $\tilde{\boldsymbol{m}} = \begin{pmatrix} \boldsymbol{M}^{-1}\boldsymbol{m} \\ \mathbf{0} \end{pmatrix}$, then the posterior distribution is multivariate Gaussian with the the following expectation and Covariance matrix.

$$E((\boldsymbol{\beta}, \boldsymbol{\gamma}) \mid \boldsymbol{y}) = (\boldsymbol{C}'\boldsymbol{R}^{-1}\boldsymbol{C} + \boldsymbol{B})^{-1} \left(\tilde{\boldsymbol{m}} + \boldsymbol{C}'\boldsymbol{R}^{-1}\boldsymbol{y}\right) \tag{A4}$$

$$\boldsymbol{Cov}((\boldsymbol{\beta}, \boldsymbol{\gamma}) \mid \boldsymbol{y}) = \left(\boldsymbol{C}'\boldsymbol{R}^{-1}\boldsymbol{C} + \boldsymbol{B}\right)^{-1} \tag{A5}$$

Model I:

For model I, depicted in Eq. (A1), $\boldsymbol{R}$ is a diagonal matrix in $\mathbb{R}^{n\times n}$ whose elements in the diagonal are σ_y^2. As we do not have random effects $\boldsymbol{C} = \boldsymbol{X}$, which is a column matrix in $\mathbb{R}^{N\times 1}$ whose elements are 1. $\tilde{\boldsymbol{m}}$ is a column matrix in $\mathbb{R}^{1\times 1}$ whose elements are 0, and finally $\boldsymbol{B} = \boldsymbol{M}^{-1}$ in $\mathbb{R}^{1\times 1}$, whose

element is $\frac{1}{\sigma_0^2}$. Then, $\beta_0 \mid \boldsymbol{y} \sim \mathcal{N}(\mu_{\beta_0}, \sigma^2_{\beta_0})$, being:

$$\mu_{\beta_0} = \boldsymbol{E}(\beta_0 \mid \boldsymbol{y}) = \frac{1}{\frac{N}{\sigma_y^2} + \frac{1}{\sigma_0^2}} \frac{N\overline{\boldsymbol{y}}}{\sigma_y^2} \tag{A6}$$

$$\sigma^2_{\beta_0} = \boldsymbol{Var}(\beta_0 \mid \boldsymbol{y}) = \frac{1}{\frac{N}{\sigma_y^2} + \frac{1}{\sigma_0^2}} \tag{A7}$$

The posterior predictive distribution for a new observation y'

$$p\left(y' \mid \boldsymbol{y}\right) = \int p\left(y' \mid \beta_0\right) \cdot p\left(\beta_0 \mid \boldsymbol{y}\right) d\beta_0 \tag{A8}$$

is Gaussian with mean μ_{β_0} and variance $\sigma^2_{\beta_0} + \sigma_y^2$.

Model II:

Regarding model II, depicted in Eq. (A2), $\boldsymbol{R}$ is also diagonal matrix in $\mathbb{R}^{N\times N}$ whose elements in the diagonal are $\sigma^2_{\boldsymbol{\epsilon}}$. $\boldsymbol{V}$, again is a column matrix in $\mathbb{R}^{N\times 1}$ whose elements are 1, and $\boldsymbol{U}$ is an identity matrix in $\mathbb{R}^{N\times 1}$. Then $\boldsymbol{C} = (\boldsymbol{V}, \boldsymbol{U})$. $\tilde{\boldsymbol{m}}$ is a column matrix in $\mathbb{R}^{(N+1)\times 1}$ whose elements are 0. Finally $\boldsymbol{B}$ is a diagonal matrix in $\mathbb{R}^{(N+1)\times(N+1)}$, whose first element of the diagonal is $\frac{1}{\sigma_0^2}$ and the rest are $\frac{1}{\sigma^2_{\boldsymbol{\omega}}}$.

Computing the joint posterior distribution for $\beta_0, \boldsymbol{\omega}$, we obtain that it follows a multivariate Gaussian with:

$$\begin{aligned}
&\boldsymbol{E}(\beta_0, \boldsymbol{\omega} \mid \boldsymbol{y}) \\
&= \boldsymbol{Cov}(\beta_0, \boldsymbol{\omega} \mid \boldsymbol{y}) \begin{pmatrix} \frac{1}{\sigma^2_{\boldsymbol{\epsilon}}} & \frac{1}{\sigma^2_{\boldsymbol{\epsilon}}} & \frac{1}{\sigma^2_{\boldsymbol{\epsilon}}} & \cdots & \frac{1}{\sigma^2_{\boldsymbol{\epsilon}}} \\ \frac{1}{\sigma^2_{\boldsymbol{\epsilon}}} & 0 & 0 & \ldots & 0 \\ 0 & \frac{1}{\sigma^2_{\boldsymbol{\epsilon}}} & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \frac{1}{\sigma^2_{\boldsymbol{\epsilon}}} \end{pmatrix} \boldsymbol{y}
\end{aligned} \tag{A9}$$

$$\begin{aligned}
&\boldsymbol{Cov}(\beta_0, \boldsymbol{\omega} \mid \boldsymbol{y}) \\
&= \begin{pmatrix} \frac{N}{\sigma^2_{\boldsymbol{\epsilon}}} + \frac{1}{\sigma_0^2} & \frac{1}{\sigma^2_{\boldsymbol{\epsilon}}} & \frac{1}{\sigma^2_{\boldsymbol{\epsilon}}} & \cdots & \frac{1}{\sigma^2_{\boldsymbol{\epsilon}}} \\ \frac{1}{\sigma^2_{\boldsymbol{\epsilon}}} & \frac{1}{\sigma^2_{\boldsymbol{\epsilon}}} + \frac{1}{\sigma^2_{\boldsymbol{\omega}}} & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sigma^2_{\boldsymbol{\epsilon}}} & 0 & 0 & \ldots & \frac{1}{\sigma^2_{\boldsymbol{\epsilon}}} + \frac{1}{\sigma^2_{\boldsymbol{\omega}}} \end{pmatrix}^{-1}
\end{aligned} \tag{A10}$$

The posterior predictive distribution for a new observation y' with mean μ' can be computed as:

$$p\left(y' \mid \boldsymbol{y}\right) = \int p\left(y' \mid \mu'\right) \cdot p\left(\mu' \mid \boldsymbol{y}\right) d\mu', \tag{A11}$$

being $p(\mu' \mid \boldsymbol{y}) = \int p\left(\mu' \mid \beta_0, \sigma^2_{\boldsymbol{\omega}}\right) \cdot p\left(\beta_0 \mid \boldsymbol{y}\right) d\beta_0$. Clearly, it is Gaussian with mean μ_{β_0} and variance $\sigma^2_{\beta_0} + \sigma^2_{\boldsymbol{\omega}}$. Note that $\sigma^2_{\beta_0}$ is the variance of the posterior marginal of β_0. This corresponds to the first element of $\boldsymbol{Cov}(\beta_0, \boldsymbol{\omega} \mid \boldsymbol{y})$, which is $\frac{1}{\frac{N}{\sigma^2_{\boldsymbol{\omega}}} + \frac{1}{\sigma_0^2}}$. Something similar happens with μ_{β_0}, the first element of the resulting matrix $\boldsymbol{E}(\beta_0, \boldsymbol{\omega} \mid \boldsymbol{y})$, which is $\frac{1}{\frac{N}{\sigma^2_{\boldsymbol{\epsilon}}+\sigma^2_{\boldsymbol{\omega}}} + \frac{1}{\sigma_0^2}} \frac{N\overline{\boldsymbol{y}}}{\sigma^2_{\boldsymbol{\epsilon}}+\sigma^2_{\boldsymbol{\omega}}}$

Finally, and coming back to Eq. (A11), we obtain that the posterior predictive distribution of $y' \mid \boldsymbol{y}$ is Gaussian, with mean μ_{β_0} and variance $\sigma^2_{\beta_0} + \sigma^2_{\boldsymbol{\omega}} + \sigma^2_{\boldsymbol{\epsilon}}$.

As a consequence, the two models have the same posterior predictive distributions, and then CPO is equal for both. □

References

Aguilera, A., Bautista, F., Gutiérrez-Ruiz, M., Ceniceros-Gómez, A.E., Cejudo, R., Goguitchaichvili, A.: Heavy metal pollution of street dust in the largest city of Mexico, sources and health risk assessment. Environ. Monit. Assess. **193**(4), 1–16 (2021). https://doi.org/10.1007/s10661-021-09344-z

Aitchison, J.: The Statistical Analysis of Compositional Data. Chapman and Hall London, London (1986)

Aitchison, J., Shen, S.M.: Logistic-normal distributions: some properties and uses. Biometrika **67**(2), 261–272 (1980)

Baker, S.G.: The multinomial-Poisson transformation. J. R. Stat. Soc. Ser. D (Stat.) **43**(4), 495–504 (1994)

Besag, J., York, J., Mollié, A.: Bayesian image restoration, with two applications in spatial statistics. Ann. Inst. Stat. Math. **43**(1), 1–20 (1991)

Blangiardo, M., Cameletti, M.: Spatial and spatio-temporal Bayesian models with R-INLA. Wiley, New Jersey (2015)

Buccianti, A., Grunsky, E.: Compositional data analysis in geochemistry: Are we sure to see what really occurs during natural processes? J. Geochem. Explor. **141**, 1–5 (2014). https://doi.org/10.1016/j.gexplo.2014.03.022

Connor, R.J., Mosimann, J.E.: Concepts of independence for proportions with a generalization of the Dirichlet distribution. J. Am. Stat. Assoc. **64**(325), 194–206 (1969). https://doi.org/10.1080/01621459.1969.10500963

Cressie, N., Wikle, C.K.: Statistics for Spatio-Temporal Data. Wiley, New Jersey (2015)

Creus Martí, I., Moya, A., Santonja, F.: Bayesian hierarchical compositional models for analysing longitudinal abundance data from microbiome studies. Complexity **2022** (2022) https://doi.org/10.1155/2022/4907527

Cribari-Neto, F., Zeileis, A.: Beta regression in R. J. Stat. Softw. **34**(2) (2010)

Douma, J.C., Weedon, J.T.: Analysing continuous proportions in Ecology and Evolution: A practical introduction to beta and Dirichlet regression. Methods Ecol. Evol. **10**(9), 1412–1430 (2019). https://doi.org/10.1111/2041-210X.13234

Dumuid, D., Stanford, T.E., Martin-Fernández, J.-A., Pedišić, Ž, Maher, C.A., Lewis, L.K., Hron, K., Katzmarzyk, P.T., Chaput, J.-P., Fogelholm, M., et al.: Compositional data analysis for physical activity, sedentary time and sleep research. Stat. Methods Med. Res. **27**(12), 3726–3738 (2018). https://doi.org/10.1177/09622802177108

Egozcue, J.J., Daunis-I-Estadella, J., Pawlowsky-Glahn, V., Hron, K., Filzmoser, P.: Simplicial regression. Norm. Model. (2012)

Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barceló-Vidal, C.: Isometric logratio transformations for compositional data analysis. Math. Geol. **35**(3), 279–300 (2003)

Engle, M.A., Rowan, E.L.: Geochemical evolution of produced waters from hydraulic fracturing of the Marcellus Shale, Northern Appalachian basin: a multivariate compositional data analysis

approach. Int. J. Coal Geol. **126**, 45–56 (2014). https://doi.org/10.1016/j.coal.2013.11.010
Fahrmeir, L., Kneib, T., Lang, S., Marx, B., Fahrmeir, L., Kneib, T., Lang, S., Marx, B.: Regression models. In: Methods and Applications. Springer, New York (2013)
Fairclough, S.J., Dumuid, D., Mackintosh, K.A., Stone, G., Dagger, R., Stratton, G., Davies, I., Boddy, L.M.: Adiposity, fitness, health-related quality of life and the reallocation of time between children's school day activity behaviours: a compositional data analysis. Prev. Med. Rep. **11**, 254–261 (2018). https://doi.org/10.1016/j.pmedr.2018.07.011
Figueras, G., Pawlowsky-Glahn, V., Vidal, C., et al.: Distributions on the simplex (2003)
Gaedke-Merzhäuser, L., Niekerk, J., Schenk, O., Rue, H.: Parallelized integrated nested Laplace approximations for fast Bayesian inference. Stat. Comput. **33**(1), 25 (2023)
Gelman, A., Hwang, J., Vehtari, A.: Understanding predictive information criteria for Bayesian models. Stat. Comput. **24**(6), 997–1016 (2014)
Gneiting, T., Raftery, A.E.: Strictly proper scoring rules, prediction, and estimation. J. Am. Stat. Assoc. **102**(477), 359–378 (2007)
Gómez-Rubio, V.: Bayesian inference with INLA. CRC Press, Boca Raton (2020)
Greenacre, M., Grunsky, E., Bacon-Shone, J., Erb, I., Quinn, T.: Aitchison's compositional data analysis 40 years on: a reappraisal. Stat. Sci. (2023). https://doi.org/10.1214/22-STS880
Haining, R.P., Haining, R.: Spatial Data Analysis: Theory and Practice. Cambridge University Press, Cambridge (2003)
Hijazi, R.H., Jernigan, R.W.: Modelling compositional data using Dirichlet regression models. J. Appl. Probab. Stat. **4**(1), 77–91 (2009)
Klein, N., Kneib, T., Klasen, S., Lang, S.: Bayesian structured additive distributional regression for multivariate responses. J. R. Stat. Soc.: Ser. C (Appl. Stat.) **64**(4), 569–591 (2015)
Kobal, M., Kastelec, D., Eler, K.: Temporal changes of forest species composition studied by compositional data approach. Forest-Biogeosci For. **10**(4), 729–738 (2017). https://doi.org/10.3832/ifor2187-010
Krainski, E.T., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., Rue, H.: Advanced spatial modeling with Stochastic partial differential equations Using R and INLA. CRC Press, Boca Raton (2018)
Lindgren, F., Rue, H., Lindström, J.: An explicit link between gaussian fields and gaussian Markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **73**(4), 423–498 (2011)
Maier, M.J.: DirichletReg: Dirichlet regression for compositional data in R (2014)
Martínez-Minaya, J., Conesa, D., Fortin, M.-J., Alonso-Blanco, C., Picó, F.X., Marcer, A.: A hierarchical Bayesian beta regression approach to study the effects of geographic genetic structure and spatial autocorrelation on species distribution range shifts. https://doi.org/10.5281/zenodo.2552025
Martínez-Minaya, J., Lindgren, F., López-Quílez, A., Simpson, D., Conesa, D.: The integrated nested Laplace approximation for fitting Dirichlet regression models. J. Comput. Graph. Stat. (2023). https://doi.org/10.1080/10618600.2022.2144330
Martínez-Minaya, J., Cameletti, M., Conesa, D., Pennino, M.G.: Species distribution modeling: a statistical review with focus in spatio-temporal issues. Stoch. Environ. Res. Risk Assess. **32**, 3227–3244 (2018)
Martínez-Minaya, J., Conesa, D., Fortin, M.-J., Alonso-Blanco, C., Picó, F.X., Marcer, A.: A hierarchical Bayesian beta regression approach to study the effects of geographical genetic structure and spatial autocorrelation on species distribution range shifts. Mol. Ecol. Resour. **19**(4), 929–943 (2019). https://doi.org/10.1111/1755-0998.13024
Merwe, S.: A method for Bayesian regression modelling of composition data. arXiv:1801.02954 (2018)
Moraga, P.: Geospatial health data: modeling and visualization with R-INLA and shiny. CRC Press, Boca Raton (2019)
Mota-Bertran, A., Saez, M., Coenders, G.: Compositional and Bayesian inference analysis of the concentrations of air pollutants in Catalonia, Spain. Environ. Res. **204**, 112388 (2022). https://doi.org/10.1016/j.envres.2021.112388
Niekerk, J., Rue, H.: Correcting the Laplace method with variational Bayes. arXiv:2111.12945 (2021)
Pawlowsky-Glahn, V., Egozcue, J.J.: Geometric approach to statistical analysis on the simplex. Stoch. Environ. Res. Risk Assess. **15**(5), 384–398 (2001)
Pettit, L.: The conditional predictive ordinate for the normal distribution. J. R. Stat. Soc.: Ser. B (Methodol.) **52**(1), 175–184 (1990)
Pirzamanbein, B., Poska, A., Lindström, J.: Bayesian reconstruction of past land cover from pollen data: Model robustness and sensitivity to auxiliary variables. Earth Space Sci. **7**(1), e2018EA00057 (2020). https://doi.org/10.1029/2018EA000547
Plummer, M.: Rjags: Bayesian Graphical Models Using MCMC. In: R package version 4–6 (2016). https://CRAN.R-project.org/package=rjags
Roos, M., Held, L.: Sensitivity analysis in Bayesian generalized linear mixed models for binary data. Bayesian Anal. **6**(2), 259–278 (2011)
Rue, H., Held, L.: Gaussian Markov Random Fields: Theory and Applications. Chapman & Hall, New York (2005)
Rue, H., Martino, S., Chopin, N.: Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J R Stat Soc Ser B (Stat Methodol) **71**(2), 319–392 (2009)
Sennhenn-Reulen, H.: Bayesian Regression for a Dirichlet distributed response using Stan. arXiv:1808.06399 (2018)
Shi, P., Zhang, A., Li, H., et al.: Regression analysis for microbiome compositional data. Ann. App. Stat. **10**(2), 1019–1040 (2016). https://doi.org/10.1214/16-AOAS928
Simpson, D., Rue, H., Riebler, A., Martins, T.G., Sørbye, S.H.: Penalising model component complexity: a principled, practical approach to constructing priors. Stat. Sci. **32**(1), 1–28 (2017). https://doi.org/10.1214/16-STS576
Simpson, D., Illian, J.B., Lindgren, F., Sørbye, S.H., Rue, H.: Going off grid: Computationally efficient inference for log-Gaussian Cox processes. Biometrika **103**(1), 49–70 (2016)
Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A.: Bayesian measures of model complexity and fit. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **64**(4), 583–639 (2002)
Templ, M., Hron, K., Filzmoser, P.: RobCompositions: an R-package for Robust statistical analysis of compositional data, pp. 341–355. John Wiley and Sons, New Jersey (2011)
Tsilimigras, M.C., Fodor, A.A.: Compositional data analysis of the microbiome: fundamentals, tools, and challenges. Ann. Epidemiol. **26**(5), 330–335 (2016). https://doi.org/10.1016/j.annepidem.2016.03.002
Van Niekerk, J., Krainski, E., Rustand, D., Rue, H.: A new avenue for Bayesian inference with INLA. Comput. Stat. Data Anal. **181**, 107692 (2023)
Wang, X., Ryan, Y.Y., Faraway, J.J.: Bayesian Regression Modeling with INLA. Chapman and Hall/CRC, London (2018)
Washburne, A.D., Silverman, J.D., Leff, J.W., Bennett, D.J., Darcy, J.L., Mukherjee, S., Fierer, N., David, L.A.: Phylogenetic factorization of compositional data yields lineage-level associations in microbiome datasets. PeerJ **5**, 2969 (2017). https://doi.org/10.7717/peerj.2969

Watanabe, S., Opper, M.: Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. J. Mach. Learn. Res. **11**(12) (2010)

Zuur, A.F., Ieno, E.N., Saveliev, A.A.: Beginner's guide to spatial, temporal, and spatial-temporal ecological data analysis with R-INLA. Highland Statistics Ltd, Newburgh (2017)