Unveiling Land Use Dynamics: Insights from a Hierarchical Bayesian Spatio-Temporal Modelling of Compositional Data

Mario FIGUEIRA, Carmen GUARNER, David CONESA, Antonio LÓPEZ- QUÍLEZ, and Tamás KRISZTIN

Changes in land use patterns have significant environmental and socio-economic impacts, making it crucial for policymakers to understand their causes and consequences. This study, part of the European LAMASUS (Land Management for Sustainability) project, aims to support the EU's climate neutrality target by developing a governance model through collaboration among policymakers, land users, and researchers. We present a methodological synthesis for treating land use data using a Bayesian approach within spatial and spatio-temporal modelling frameworks. The study tackles the challenges of analysing land use changes, particularly the presence of zero values and computational issues with large datasets. It introduces joint model structures to address zeros and employs sequential inference and consensus methods for Big Data problems. Spatial downscaling models approximate smaller scales from aggregated data, circumventing high-resolution data complications. We explore Beta regression and Compositional Data Analysis (CoDa) for land use data, review relevant spatial and spatio-temporal models, and present strategies for handling zeros. The paper demonstrates the implementation of key models, downscaling techniques, and solutions to Big Data challenges with examples from simulated data and the LAMASUS project, providing a comprehensive framework for understanding and managing land use changes.

Key Words: Land use; Compositional data; Spatio-temporal models; Downscaling; Big Data; INLA.

1. INTRODUCTION

Changes in land use patterns have significant environmental and socio-economic impacts, making it crucial for policymakers to understand their causes and consequences (LeSage

M. Figueira (✉) · C. Guarner · D. Antonio · A. López-Quílez, Department of Statistics and Operations Research, University of Valencia, Carrer del Dr. Moliner, 50, 46100 Valencia, Burjassot, Spain
(E-mail: *Mario.Figueira@uv.es*).

T. Krisztin, International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, 2361 Laxenburg, Austria.

Journal of Agricultural, Biological, and Environmental Statistics
https://doi.org/10.1007/s13253-025-00678-6

Published online: 12 February 2025

2008; Chakir and Le Gallo 2013). Studies on land use are valuable for identifying the determinants of these changes, which affect biodiversity, water pollution, soil erosion, climate change, and economic and social welfare (Hersperger and Bürgi 2009; Chakir and Lungarska 2017; Moindjié et al. 2022; Bareille and Chakir 2024). These changes are driven by a combination of socio-economic factors, pedo-climatic conditions, and policy variables. Land use models are crucial for analysing these influences and their effects (Chakir and Parent 2009; Hersperger and Bürgi 2009; van Vliet et al. 2015). Moreover, the decision-making processes surrounding land use are complex, shaped by both local and global biophysical and socio-economic factors. Therefore, comprehensive knowledge of these influences aids in evaluating and formulating environmentally friendly public policies.

Various disciplines, including economics, statistics, geography, and land use science, have developed empirical land use modelling approaches using aggregate or individual data (Chakir 2009). However, many studies overlook spatial autocorrelation in modelling land use or use ad hoc methods (LeSage 2008; Chakir and Le Gallo 2013; Chakir and Le Gallo 2021; Moindjié et al. 2022), despite its prevalence in economic decisions. This underscores the need for more sophisticated spatial econometric models that account for spatial heterogeneity and interdependence to accurately measure and analyse land use patterns (LeSage 2008; James LeSage 2009; Elhorst 2013; Chakir and Le Gallo 2021).

This work is part of the European LAMASUS (Land Management for Sustainability) project. LAMASUS develops an innovative governance model through collaboration between policymakers, land users, and researchers. This Horizon Europe project aims to support the European Union's climate neutrality target by creating an open-access modelling toolbox for designing effective land use policies within the framework of the European Green Deal. In the context of the LAMASUS project, this paper presents a methodological synthesis for the treatment of land use data using a Bayesian approach and in the spatial and spatio-temporal modelling framework. This approach is integrated in conjunction with those that aim to provide a better understanding of the spatial dynamics of the processes that determine and drive land use changes. Enhanced understanding also enables the implementation of land use policies based on models that improve policy formulation.

LAMASUS faces different problems related to land use and land use change analysis. In this case, there are numerous land use categories, which are aggregated into a small set to facilitate the determination of the main drivers of land use at different scales, both administrative (NUTS scales) and high-resolution levels, that is, small spatial scales. From the aggregated spatial scales, to approximate a smaller scale, spatial downscaling models or disaggregation models are applied, avoiding dealing with high-resolution data that may present inconveniences related to the systematic presence of 0's or being large data banks, which would pose a Big Data problem. On the other hand, to work with the aggregated data with many disaggregated land use categories, joint model structures are proposed to deal with the systematic presence of 0's and 1's, while for problems that emerge from Big Data, a combined procedure of sequential inference and consensus is proposed.

In particular, the aim of this paper is to present different procedures and methods to analyse land use data in different contexts of spatial and spatio-temporal modelling. In this sense, Sect. 2 introduces land use data and its analysis by Beta regression, when only one land use category is available, or within the framework of Compositional Data Analysis (CoDa),

if multiple land use categories are available. In Sect. 3 we briefly review a selection of those spatial and spatio-temporal models that are of interest for assessing the spatial and spatio-temporal structure of land use data. Section 4 sets out the proposed methodology for dealing with the presence of 0's and 1's in land use data, both when a single category is available and when several categories are available. In Sect. 4 is briefly shown the implementation of the main spatial and spatio-temporal models for land use data. Section 5 illustrates how to deal with the presence of 0's and 1's in Beta regression and CoDa analysis. In Sect. 6, we present downscaling models, also known as disaggregation models, applied to land use and compositional data. Section 7 is related to Big Data in the framework of high-resolution land use data. Section 8 stands for some simulated data and real data from LAMASUS project. Finally, we conclude in Sect. 9.

2. LAND USE AND COMPOSITIONAL DATA

In this section, we focus on presenting the treatment of compositional data for the analysis of land use shares, providing a brief exposition of the theoretical framework for CoDa and the implementation of such analysis to address the drivers and underlying process in land use shares.

Land use data are usually presented as proportions of land use shares in a given area, which implies that land use shares, by definition, have a compositional nature (Pirzamanbein et al. 2020; Thomas-Agnan et al. 2021; Krisztin et al. 2022). Therefore, one approach to capture the variability of land use shares would be through Beta regression if we focus on a specific category, or through CoDa to analyse the joint variability of several land use categories, e.g. cropland, grassland, forest, urban and other. The joint analysis of the categories in the composition can be considered by different approaches: transforming the logratios (Aitchison 1986; Aitchison and Egozcue 2005; Greenacre et al. 2023) and modelling those logratios using multivariate Gaussian models, or using Dirichlet regression models (Connor and Mosimann 1969; Hijazi and Jernigan 2009) for the proportions without needing to transform them.

In general, compositional data consist of a set of *parts* that identifies the constituents of the composition $\{1, 2, \ldots, D\}$, while *components* $\{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_D\}$ are numerical proportions in which individual parts occur (Aitchison 1986). Therefore, CoDa can be defined by a matrix $\mathbf{Y}_{n\times D}$ of n observations times D parts of the composition. Each row of the matrix $\mathbf{Y}_{n\times D}$ satisfies the following closure condition $\sum_{d=1}^{D} \mathbf{Y}_{id} = 1$, and it is also assumed that each value in the composition falls within the interval $\mathbf{Y}_{id} \in (0, 1)$. The vector $\mathbf{y}_{1\times D}$ for the i-th observation, related to each row of $\mathbf{Y}_{n\times D}$, is called composition and it pertains to the simplex sample space. A simplex space with dimension d, denoted by $\mathbb{S}^d$, is defined as:

$$\mathbb{S}^D = \left\{ \mathbf{y} \in \mathbb{R}^D : y_d \in (0, 1), \sum_{d=1}^{D} y_d = 1 \right\}, \tag{1}$$

where due to the closure condition and the property $y_i \in (0, 1)$, the entire set of observations of the composition $\mathbf{Y}_{n\times D}$ does not follow the usual Euclidean geometry $\mathbb{R}^D$, but rather the

Aitchison geometry $\mathbb{S}^d$ with dimension $d = D - 1$ (Aitchison 1986; Greenacre et al. 2023). Additionally, it is also possible to define a *subcomposition* $\mathbf{y}_{1\times C}$ of $\mathbf{y}_{1\times D}$. This refers to a subset of C parts of the referring composition of D parts, where $C < D$ and the subcomposition is also subject to the closure condition $\sum_{i=1}^{C} y_i = 1$.

2.1. LOGRATIO TRANSFORMATIONS

In Aitchison's approach to compositional data analysis (Aitchison 1982, 1986), the focus was on distributional issues, such as finding a way to transform compositions into interval-scale multivariate vectors that could validly use the multivariate normal distribution. The proposed logistic transformations, which aimed to create "transformed-normal" models, relied on logarithmic transformations. These transformations allowed the use of standard unconstrained multivariate statistics applied to transformed data, with inferences translatable back into compositional statements (Aitchison 1986; Greenacre et al. 2023). However, this reliance on logarithms precluded zero data values, leading to the ongoing debate about zero replacement and treatment, a topic often overlooked in publications (Greenacre et al. 2023).

There are several proposals for logratio transformations to analyse CoDa, including those from Aitchison's original work (Aitchison 1982, 1986), such as *pairwise logratios* (LR), *additive logratios* (ALR), and *centred logratios* (CLR). Later proposals (Egozcue et al. 2003; Egozcue and Pawlowsky-Glahn 2006; Fiserová and Hron 2011) include *isometric logratios* (ILR) and *pivot logratios* (PLR). These transformations attempt to address various shortcomings of the alternative methods. However, each logratio transformation has its own strengths and weaknesses, making it impossible to propose a single transformation suitable for all CoDa datasets (Greenacre and Grunsky 2019; Greenacre et al. 2023).

In this work, we will focus on two of the logratios proposed by Aitchison (1986): ALR and CLR. The ALR transformation implies defining a subset of $d = D - 1$ logratios with the same denominator, called the *reference* (r) part. Therefore, the ALR with respect to a reference part of the composition is written as

$$\mathrm{ALR}(y_i \mid y_r) = \log\left(\frac{y_i}{y_r}\right),\ i \in \{1, ..., D\} : i \neq r, \tag{2}$$

where we define D ALR transformations, depending on the part chosen as reference. In this case, it is important to use a reference part that satisfies either statistical or meaningful objective. In fact, if the chosen reference part is almost constant, then the corresponding ALRs are approximately the logarithm of the parts, up to a nearly constant amount.

The CLR transformation is defined as the logratios of the parts with respect to their geometric mean $g(\mathbf{y}) = \prod_{i=1}^{D} y_i^{1/D}$:

$$\mathrm{CLR}(y_i) = \log\left(\frac{y_i}{g(\mathbf{y})}\right),\ i \in \{1, ..., D\}, \tag{3}$$

transformation that turns the sum-to-one constraint into a sum-to-zero constraint $\sum_{i=1}^{D} \mathrm{CLR}(y_i) = \sum_{i=1}^{D} \left[\log(y_i) - \log(g(\mathbf{y}))\right] = 0$. The usefulness of this transformation is that Euclidean

distances between the CLRs are identical to the distances using all possible logratios; in other words, the CLR is an isometric transformation.

Both logratio transformations allow us to model the logratios using multivariate Gaussian distributions $\mathbf{y}^* \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\mathbf{y}^*$ represents the logratios, either from CLR or ALR transformations. This multivariate model not only accounts for complex structures in the linear predictors, including spatial and spatio-temporal dependencies, but also captures the correlations between the logratios through the variance-covariance matrix $\boldsymbol{\Sigma}$ of the multivariate Gaussian distribution (Martínez-Minaya and Rue 2024).

2.2. DIRICHLET DISTRIBUTION

Dirichlet distribution is a generalisation of the Beta distribution for more than two proportions, and its probability density function is defined as

$$\pi(\mathbf{y} \mid \boldsymbol{\alpha}) = \frac{1}{\text{B}(\boldsymbol{\alpha})} \prod_{i=1}^{D} y_i^{\alpha_i - 1}, \tag{4}$$

where $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_D)$ is the vector of shape parameters $\alpha_i > 0 \ \forall i$, and $\text{B}(\boldsymbol{\alpha})$ is the Multinomial Beta function, being the normalising constant of the Dirichlet. The sum of the shape parameters $\alpha_0 = \sum_{i=1}^{D} \alpha_i$ is usually interpreted as a precision marginal parameter τ. In the Dirichlet distribution the sum-to-one constraint $\sum_{i=1} y_i = 1$ is also satisfied by the data vector $\mathbf{y} = (y_1, ..., y_D)$, $y_i \in (0, 1)$. Therefore, if $\mathbf{y} \sim \mathcal{D}(\boldsymbol{\alpha})$ denotes a variable that is Dirichlet distributed, the expected values are $\mathbb{E}(y_i) = \alpha_i / \alpha_0$, and the relation between the shape parameters and the linear predictor (η_i) is given by the log-link function $\log(\alpha_i) = \eta_i$. Thus, it is also possible to parameterise the Dirichlet distribution in terms of the mean as $\mu_i = \exp(\eta_i)/(\sum_{i=1}^{D} \exp(\eta_i))$.

The parameterization of the Dirichlet distribution in terms of the shape parameters (or the mean) linked to the linear predictor allows us to use complex structures, as in the INLA framework, by defining Latent Gaussian Models (LGM). These structures can include temporal components, spatial components such as the SPDE-FEM approach, or more complex spatio-temporal processes. The implementation of the Dirichlet distribution in `R-INLA` relies on methodology and R-package presented by (Martínez-Minaya et al. 2023). This provides the basis to implement complex spatial and spatio-temporal structures for Dirichlet distributed data inside the INLA framework.

3. SPATIAL AND SPATIO-TEMPORAL MODELS

The land use may present changes along space and time. Consequently, to assess the variability it is possible to construct spatio-temporal models that can provide a further description and insights (LeSage 2008; Chakir and Le Gallo 2021). In this section, we will provide a brief overview of the different structures that can be used in spatio-temporal modelling for land use data.

In most cases, land use data are provided as areas, i.e. values related to polygons that typically represent administrative boundaries. This areal data naturally arises when a fixed

domain is partitioned into a finite number of subregions where outcomes are aggregated. Therefore, an appropriate way to model this kind of spatial dependency is by building a structure of adjacency, or an adjacency matrix $\mathbf{W}$, between the different connected areas and setting a multivariate Gaussian density with a precision matrix based on a modulation of this adjacency structure (Banerjee et al. 2015), which is quite common in disease mapping as well (Martinez-Beneito and Botella-Rocamora 2019). This prior distribution defines the dependence between the connected areas and allows the assessment of spatial structures, from simpler ones like the Besag model (Besag 1974) or the Besag-York-Mollié model (Besag et al. 1991) to more complex multidimensional spatial structures (Martinez-Beneito et al. 2017). These models fall under the class of Conditional Autoregressive (CAR) models, which are highly popular and can be related to Simultaneous Autoregressive (SAR) models (Ver Hoef et al. 2018), with the latter being widely used in land use management and econometrics. For a review of these models, their properties, and applications, we refer to the works of MacNab (2022, 2023).

In spatial econometrics, the Spatial Error model (SEM), Spatial Lag model (SLM), and Spatial Durbin model (SDM) are commonly used to analyse spatial dependency (Elhorst 2013; Bivand et al. 2014). These models also define a spatial dependence structure for the explanatory covariates and the error term. The general framework that summarised these models is the general nesting spatial model equation (Elhorst 2013)

$$\begin{aligned} \mathbf{y} &= \delta\mathbf{W}\mathbf{y} + \alpha\mathbf{1}_N + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\theta} + \mathbf{u}, \\ \mathbf{u} &= \lambda\mathbf{W}\mathbf{u} + \varepsilon, \end{aligned} \tag{5}$$

where $\mathbf{1}_N$ denotes a vector $(N \times 1)$ of ones, α is a global intercept, $\mathbf{W}$ is an adjacency matrix of positive defined weights usually normalised such that each row sums to unity (LeSage 2008; James LeSage 2009), $\mathbf{W}\mathbf{y}$ represents the endogenous interaction effects of the response variables $\mathbf{y}$, $\mathbf{W}\mathbf{X}$ represents the exogenous interaction effects among the explanatory variables $\mathbf{X}$, $\mathbf{W}\mathbf{u}$ denotes the interaction among the spatial units, $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ represent unknown parameters. Finally, δ and λ are the spatial autoregressive parameter and spatial autocorrelation parameter, respectively. These models can be reassembled to fit within the INLA framework (Bivand et al. 2014; Gómez-Rubio et al. 2021) by fixing some parameters and performing their analysis outside INLA, using Monte Carlo (MC, Berild et al. (2022)) or Markov Chain Monte Carlo (MCMC, Gómez-Rubio and Rue (2018)).

The temporal terms can be incorporated into the linear predictor of the model in an additive way, i.e. as a new term evaluating a purely temporal trend. This can be synthesised into a model with one spatial component and one temporal component $\mathbf{y} = \delta\mathbf{W}\mathbf{y} + \alpha\mathbf{1}_N + \mathbf{X}\boldsymbol{\beta} + \mathbf{u}_s + \mathbf{u}_t + \varepsilon$, where $\mathbf{u}_s$ is the purely spatial component, $\mathbf{u}_t$ is the purely temporal term and ε is an independent and identically distributed error term. As an example, the temporal and spatial components can be integrated into a rewritten form of the SLM to fit within the INLA framework in the following way:

$$\mathbf{y} = (\mathbf{I} - \delta\mathbf{W})^{-1}\left[\mathbf{X}\boldsymbol{\beta} + \mathbf{u}_s + \mathbf{u}_t + \varepsilon\right] \tag{6}$$

where $\mathbf{u}_t \sim \text{MVN}(\mathbf{0}, \mathbf{Q}_t)$ stands for the purely temporal component, and $\mathbf{u}_s \sim \text{MVN}(\mathbf{0}, \mathbf{Q}_s)$ is the purely spatial term. In this case, $\mathbf{Q}_t^* = (\mathbf{I} - \delta\mathbf{W})^T \cdot \mathbf{Q}_t \cdot (\mathbf{I} - \delta\mathbf{W})$ and $\mathbf{Q}_s^* = (\mathbf{I} - \delta\mathbf{W})^T \cdot \mathbf{Q}_s \cdot (\mathbf{I} - \delta\mathbf{W})$ represents the rewritten temporal and spatial structure for the spatial error component. Meanwhile, the rewritten precision matrix for the error term is $\mathbf{Q}_\varepsilon^* = (\mathbf{I} - \delta\mathbf{W})^T \cdot (\mathbf{I} - \delta\mathbf{W})$. The model in this form can be implemented in the `R-INLA` software.

In addition to being able to evaluate additive spatio-temporal models, it is possible to consider different structures for spatial and temporal interaction components. Knorr-Held (2000) proposed four general structures to define spatio-temporal interaction models, in which the precision matrix of the spatio-temporal effect is constructed by means of the Kronecker product of the precision matrix of the spatial component by the precision matrix of the temporal component $\mathbf{Q}_{st} = \mathbf{Q}_s \otimes \mathbf{Q}_t$ (Clayton 1996). Four natural interactions arise from this formulation: (a) Type I interaction, this is defined as the Kronecker product of the precision matrices of an unstructured spatial $\mathbf{Q}_s = \tau_s\mathbf{I}$ and temporal effect $\mathbf{Q}_t = \tau_t\mathbf{I}$, where the structure for the spatio-temporal component is $\mathbf{Q}_{st} = \mathbf{Q}_s \otimes \mathbf{Q}_t = \tau_{st}\mathbf{I} \otimes \mathbf{I}$. This interaction makes sense when space and time have a discretised structure, otherwise this specification may not make sense. (b) Type II interaction assumes an unstructured spatial effect $\mathbf{Q}_s = \tau_s\mathbf{I}$ interacting with a structured spatial effect $\mathbf{Q}_t$, then the spatio-temporal precision matrix is $\mathbf{Q}_{st} = \mathbf{I} \otimes \mathbf{Q}_t$. (c) Type III interaction is the interaction of a structured spatial effect $\mathbf{Q}_s$ by an unstructured temporal effect $\mathbf{Q}_t = \tau_t\mathbf{I}$, obtaining that $\mathbf{Q}_{st} = \mathbf{Q}_s \otimes \mathbf{I}$. (d) Type IV interaction is the interaction between a structured spatial effect $\mathbf{Q}_s$ and a structured temporal effect $\mathbf{Q}_t$, which implies that the interacting spatio-temporal component is defined by $\mathbf{Q}_{st} = \mathbf{Q}_s \otimes \mathbf{Q}_t$.

The previously discussed precision structures for spatio-temporal effects are separable in the sense that they can be decomposed as the Kronecker product of a matrix associated purely with the spatial component and another matrix associated with the temporal component. However, this assumption can be artificial, as in real dynamic processes, the spatio-temporal dynamic component does not necessarily have to be separable in these terms. Therefore, an alternative is to construct non-separable structures, which can be defined based on a correspondence with stochastic partial differential equations (Lindgren et al. 2024) that express the dynamics of natural processes, or through a class of non-separable transformed multivariate Gaussian Markov random fields (Prates et al. 2022), which provide a straightforward way to construct the matrices and interpret the spatial, temporal and spatio-temporal parameters. In these models, the precision matrix cannot be decomposed into the Kronecker product of a spatial precision matrix and a temporal precision matrix, i.e. $\mathbf{Q}_{st} \neq \mathbf{Q}_s \otimes \mathbf{Q}_t$.

4. IMPLEMENTING SPATIAL CODA MODELS

Implementing spatial models in compositional data can be challenging due to the multivariate nature of the response variables. Each category or log-ratio of land use may have its own spatial or spatio-temporal component with a distinct set of hyperparameters. To address this complexity, rather than defining a separate component for each variable, two alternatives can be considered. One approach involves replicating some spatial or spatio-temporal components across different land use categories or log-ratios, resulting in identical

hyperparameters for these replicated components (Gómez-Rubio 2020). Another approach is to share spatial or spatio-temporal components across some categories, either scaled by parameters to be estimated or sharing the same value (Gómez-Rubio 2020).

This can be exemplified with the following toy model for three logratios $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3)$, where $\mathbf{Y} \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and the mean vector $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3)$ is modelled as:

$$\begin{aligned} \boldsymbol{\mu}_1 &= \beta_{10}\mathbf{1} + \mathbf{X}\boldsymbol{\beta}_{11} + \mathbf{u}_{1s}(\tau_s, \lambda_s) + \mathbf{u}_t(\tau_t), \\ \boldsymbol{\mu}_2 &= \beta_{20}\mathbf{1} + \mathbf{X}\boldsymbol{\beta}_{21} + \mathbf{u}_{2s}(\tau_s, \lambda_s) + \mathbf{u}_t(\tau_t), \\ \boldsymbol{\mu}_3 &= \beta_{30}\mathbf{1} + \mathbf{X}\boldsymbol{\beta}_{31} + \mathbf{u}_{3s}(\tau_s, \lambda_s) + \mathbf{u}_t(\tau_t), \end{aligned} \tag{7}$$

where $\beta_{\bullet 0}$ are the intercepts, $\boldsymbol{\beta}_{\bullet 1}$ represents the fixed effect vector related to the matrix of explanatory variables $\mathbf{X}$, $\mathbf{u}_{\bullet s}$ denotes the spatial effects, which are three different replicas from the same Leroux distribution (Leroux et al. 2000), given by $\mathbf{u} \sim \text{MVN}(\mathbf{0}, \mathbf{Q}_s) : \mathbf{Q}_s = \tau_s \cdot [\mathbf{I} + \lambda_s \cdot (\mathbf{D} - \mathbf{I} - \mathbf{W})]$, and $\mathbf{u}_t : u_{t,i} = u_{t,i-1} + \varepsilon_i(\tau_t)$ is a temporal component shared through the three linear predictors with the same values for each logratio. Finally, $\boldsymbol{\Sigma}$ is the variance-covariance matrix between of logratios, with $\boldsymbol{\Sigma}_{ii}^{-1} = \tau_i$ and $\boldsymbol{\Sigma}_{ij}^{-1} = \rho_{ij}/(\tau_i \cdot \tau_j)^{1/2}$, $i \neq j$.

In `R-INLA`, it is not possible to model multivariate models directly, as the method relies on the assumption of conditional independence of the data given the latent field and the hyperparameters. Therefore, the correlation structure for the multivariate normal observations cannot be evaluated in the likelihood; instead, it can be represented by a random effect with the desired structure in the linear predictor of the model. To illustrate this, suppose we have n observations of a D-dimensional multivariate Gaussian distribution. To evaluate the correlations of this multivariate distribution in `R-INLA`, we can construct a component on the latent field $\mathbf{u}$, such that $\mathbf{u} \sim \text{MVN}(\mathbf{0}, \mathbf{Q})$, whose precision matrix is defined as $\mathbf{Q}_{ii} = \tau_{ii}$ and $\mathbf{Q}_{ij} = \rho_{ij}/\sqrt{\tau_{ii}\tau_{jj}}$. Each realisation of this component produces a vector of D elements, so if this is replicated for each row of observations, we will have n replicas of this effect evaluated with the same hyperparameters $\{\boldsymbol{\tau}, \boldsymbol{\rho}\}$. This allows us to express the D-dimensional multivariate distribution in the likelihood with n components as a univariate distribution with $n \times D$ observations and fixed precision in the univariate Gaussian likelihood to a high value ($\log(\tau) \gg 1$). Therefore, in each term of the linear predictor we will have a realisation from this effect u corresponding to a category $d \in \{1, ..., D\}$ and an observation $i \in \{1, ..., n\}$.

Another way to express this is to consider that we have multivariate observations organised in rows for each multivariate observation and in columns for the different components of the variable distributed according to a multivariate distribution. This can be expressed by defining the model for each multivariate observation in the following way:

$$\mathbf{y}_{1\times D} = \boldsymbol{\mu}_{1\times D} + \boldsymbol{\varepsilon} \tag{8}$$

where $\boldsymbol{\varepsilon} \sim \text{MVN}(\mathbf{0}_{1\times D}, \boldsymbol{\Sigma}_{D\times D})$. If we incorporate this structure, $\boldsymbol{\varepsilon}$, into the latent field, then the model can be re-written as follows

$$y_i = \mu_i + \varepsilon_i \tag{9}$$

where $\varepsilon_i \sim \mathrm{N}(0, \tau)$, and τ is fixed at a high value. The latent field must then incorporate a multivariate structure with correlations that relate the components of the multivariate structure we are re-expressing. If we rearrange the matrix structure into a vector such that the indices for the multivariate components are represented as $(1_1, ..., n_1, ..., 1_d, ..., n_d, ..., 1_D, ..., n_D)$, then the multivariate component $\mathbf{u}$ must follow this indexing and can be incorporated into the latent field.

This demonstrates the flexibility to define various spatio-temporal structures for modelling compositional data, or more generally, multivariate data. It also highlights the combinatorial challenge of defining models within such a framework, where testing all possible combinations implies defining 2^D models for each component of the model, with D being the multivariate dimensionality. This issue suggests that constructing models in these contexts can benefit significantly from expert knowledge, which can help narrow down the feasible set of reasonable models. This approach avoids the need for extensive model selection procedures among numerous sets, which can become intractable with even a few components. Therefore, implementing spatial models in the context of compositional data requires establishing criteria based on expert knowledge or other relevant factors to limit the set of models to be evaluated. This is especially important given the various possibilities for defining spatio-temporal components that we have mentioned in the previous section.

5. DEALING WITH 0'S IN LAND USE DATA

In the first section, we presented various approaches to analyse land use data: beta models when focussing on a single category, and multivariate normal models for some transformation of logratios (additive logratios, centred logratios, isometric logratios) or Dirichlet models when dealing with multiple categories. These are all different approaches to CoDa analysis in general, and to land use data provided in proportions in particular. However, a significant challenge in dealing with compositional data across all these approaches is the presence of zeros and ones (Martín-Fernández and Thió-Henestrosa 2006; Martín-Fernández et al. 2011; Tsagris and Stewart 2018). Zeros can be problematic because they often indicate the absence of a component, complicating the application of standard statistical techniques that assume strictly positive data. Similarly, ones can signify the complete dominance of a single component (inducing zero values in the remaining components), which can distort correlations and overall analysis. Addressing these issues typically requires specific methods, such as zero-replacement strategies or transformations, to ensure that the compositional nature of the data is properly accounted for and that the analyses yield meaningful results.

In general, different types of zeros can be found in compositional data (CoDa): (i) rounded zeros or below-detection values, (ii) count zeros, and (iii) essential zeros. The first two types of zeros are associated with "false" zeros, where the structure of the experimental design, the sample size, or the sensitivity of the measuring instruments results in null values, even though the underlying phenomenon is not necessarily null. In contrast, the third type of zeros, also called structural or absolute zeros, assumes that the measurements indeed capture the real absence or null value of the phenomenon (Martín-Fernández et al. 2011). For the first two cases of rounded zeros and count zeros, replacement procedures or transformations

have been proposed (Rasmussen et al. 2020; Lubbe et al. 2021), while for essential zeros, likelihood modifications such as zero-inflated models or hierarchical structures can be used (Aitchison and Kay 2003; Tsagris and Stewart 2018; Tang et al. 2022).

In line with Aitchison and Kay (2003), a joint hierarchical model similar to Hurdle models (Mullahy 1986; Cameron and Trivedi 2013; Martínez-Minaya et al. 2018), which are widely used in the environmental and ecological sciences, can be proposed to deal with null values in likelihoods that do not allow them to be evaluated. This approach involves conditional modelling of the components according to whether each particular component has a zero or nonzero value. In order to apply a structure similar to that of a Hurdle model, there must be a correspondence in the modelling of the composition and subcompositions with that of the null and non-null values of the subcompositions. This implies that we can use it in the case of dealing with data distributed according to a Dirichlet or when we use the centred logratios, since it allows us to make the joint analysis of the zeros with the process that generates the logratios for the CLR, or the shape parameters in the case of the Dirichlet.

Let's assume that we have a matrix of n observations and D compositions, $\mathbf{Y}_{n\times D}$ where each element Y_{ij} is the value for the i-th observation and j-th composition. Then, we can define an incidence matrix $\mathbf{I}_{n\times D}$, that for each element I_{ij} we have $I_{ij} = 0 \iff Y_{ij} = 0$ and $I_{ij} = 1 \iff Y_{ij} \neq 0$. Each column of the incidence matrix will be the values that we will model according to a Bernoulli distribution, and in principle this modelling will be independent between the different compositions. This assumption of strict independence is not entirely accurate, since all components cannot be null at the same time. However, it is also not clear how this restriction impacts the correlations that may exist between the components of the different Bernoulli distributions we are considering. Therefore, we propose a simpler approach by assuming independence, or a weak dependence that would not significantly affect the estimates. Consequently, considering independence for the incidence matrix, the model for each composition integrating the vectors of incidence $\mathbf{i}_d = \mathbf{I}_{1:n,d}$ can be written as:

$$\begin{aligned} \mathbf{i}_1 &\sim \text{Ber}(\boldsymbol{\pi}_1) \\ \text{logit}(\boldsymbol{\pi}_1) &= \mathbf{X}\boldsymbol{\beta}_1 + \textstyle\sum_{k=1}^{K} f_{k,1}(\mathbf{z}_k) + \mathbf{u}_{st,1} \\ &\vdots \\ \mathbf{i}_D &\sim \text{Ber}(\boldsymbol{\pi}_D) \\ \text{logit}(\boldsymbol{\pi}_D) &= \mathbf{X}\boldsymbol{\beta}_D + \textstyle\sum_{k=1}^{K} f_{k,D}(\mathbf{z}_k) + \mathbf{u}_{st,D} \end{aligned} \tag{10}$$

where $\boldsymbol{\beta}_d$ is the vector of linear coefficients, $\mathbf{f}_d$ are nonlinear components and $\mathbf{u}_{st,d}$ the spatio-temporal components for the d-th component. While for the models for the values of the compositions, if we use the CLR we can write $\mathbf{Y}^* = \text{CLR}(\mathbf{Y})$, we have the following multivariate Gaussian model for the logratio of the compositions $\mathbf{Y}^* \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$. However, in this case, it is important to distinguish between rows where there is a component with a zero value, or where there is one component with a unit value and the rest are zeros, and rows where no values are zero, in order to analyse them analogously to how it is done with Hurdle models.

The multivariate model can be rewritten as D univariate Gaussian models in which the marginal precision is fixed at a very high value and the variance-covariance structure is

evaluated by a random effect as described in the previous section in order to implement these models in R-INLA:

$$\begin{aligned}
\mathbf{y}_1^* &\sim \mathrm{N}(\boldsymbol{\mu}_1, \tau^*) \\
\boldsymbol{\mu}_1 &= \mathbf{X}\boldsymbol{\beta}_1 + \textstyle\sum_{k=1}^{K} f_{k,1}(\mathbf{z}_k) + \mathbf{u}_{st,1} + \mathbf{u} \\
&\vdots \\
\mathbf{y}_D^* &\sim \mathrm{N}(\boldsymbol{\mu}_D, \tau^*) \\
\boldsymbol{\mu}_D &= \mathbf{X}\boldsymbol{\beta}_D + \textstyle\sum_{k=1}^{K} f_{k,D}(\mathbf{z}_k) + \mathbf{u}_{st,D} + \mathbf{u}
\end{aligned} \tag{11}$$

where τ^* is a fixed precision with a high value ($\log(\tau^*) \gg 1$), $\mathbf{u}$ is the random effect that accounts for the variance-covariance structure, $\mathbf{u} \sim \mathrm{MVN}(\mathbf{0}, \mathbf{Q})$, such that $\mathbf{Q} = \boldsymbol{\Sigma}^{-1}$. In addition, for each observation row with a zero value component, the CLR transformation is re-evaluated to exclude those null values. This procedure is the same whether we use the Dirichlet distribution or the Beta distribution, which is a simplification of the Dirichlet when only two categories are available.

6. DOWNSCALING MODELS

Empirical studies involving land use change often use aggregate data for regions, countries, or other geographic scales (Chakir 2009; Chakir and Le Gallo 2013). However, some land use studies have had access to individual parcel-level data for analysing water quality, the extent of urban sprawl, carbon sequestration costs, and habitat fragmentation (You and Wood 2006; Chakir 2009). The spatial pattern of land use is of particular interest and has been useful in identifying factors that drive changes in land use at the disaggregated level (Chakir 2009; An and Lee 2020). In a more general sense, this issue of aggregated, disaggregated, and different scales of data availability and analysis can be related to fusion models (Wang et al. 2018; Wang and Furrer 2021; Villejo et al. 2023), data misalignment (Moraga et al. 2017), change of support (Gelfand 2001; Bradley et al. 2016) and disaggregation modelling or downscaling (Nandi et al. 2023). In this general context, different issues arise concerning the availability at different scales of information concerning the response variable, as in data fusion models (Wang and Furrer 2021), or availability at different scales of relative information, as in downscaler models (You and Wood 2006; Berrocal et al. 2010). This implies extracting information from an aggregated scale to a disaggregated or continuous scale using an effect with a continuous spatial or spatio-temporal structure, structure that is commonly used in geostatistics to account for the data spatial dependence. This is possible to implement in INLA thanks to the definition of spatial models using the SPDE-FEM approach, which is integrated in the R-INLA package (Lindgren and Rue 2015; Bakka et al. 2018).

6.1. SPDE-FEM

The SPDE-FEM approach (which stands for *Stochastic Partial Differential Equations* and *Finite Element Methods*) allows for expressing the continuous spatial structure by approximating a Gaussian Field with a Gaussian Markov Random Field. This makes the inferential

process efficient by reducing computational costs. This approach was initially proposed in (Lindgren et al. 2011) by means of an approximate stochastic weak solution to the SPDE in Eq. (12), proving that a stationary SPDE solution has a Matérn covariance function and allowing to calculate directly its precision matrix, bypassing the need for resource-intensive inverse processes. Thus, formally we get that the structured spatial effect ξ with precision τ is related to a non-structured Gaussian random effect (white noise) through a filtering operator $(\kappa - \Delta)^{\alpha/2}$:

$$\begin{gathered}(\kappa^2 - \Delta)^{\alpha/2}\tau \cdot \xi(\mathbf{s}) = \mathcal{W}(\mathbf{s}), \\ \mathbf{s} \in \mathcal{D},\ \alpha = \nu + d/2,\ \kappa > 0,\ \nu > 0,\end{gathered} \tag{12}$$

where $\mathbf{s}$ are the locations; κ is a spatial scale parameter positive defined and related to ρ and ν through $\kappa = \sqrt{8\nu}/\rho$; α controls the smoothness of the performances, which by default is 2; τ regulates the variance, so it is also a parameter defined positive as $\tau^2 = \Gamma(\nu)\left[\Gamma(\alpha)(4\pi)^{d/2}\kappa^2\sigma^2\right]^{-1}$; and $\mathcal{D}$ is the spatial domain $\mathcal{D} \subset \mathbb{R}^d$, being d the Euclidean spatial dimension of such spatial domain. Additionally, $\Delta = \sum_i \partial^2/\partial s_i^2$ is the Laplacian operator and $\mathcal{W}$ denotes a spatial stochastic Gaussian process with unit variance. Therefore, the solution for $\xi(\mathbf{s}_i)$ leads to a Gaussian field with a covariance matrix defined by the Matérn function correlation $\mathcal{C}(h)$, where the covariance is the correlation multiplied by the marginal deviation:

$$\mathcal{C}(h) = \sigma^2 \cdot \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\sqrt{2\nu} \cdot \frac{h}{\rho}\right)^{\nu} \mathcal{K}_\nu\left(\sqrt{2\nu} \cdot \frac{h}{\rho}\right). \tag{13}$$

In the above equation, σ is the marginal standard deviation, ρ is the spatial range, $\mathcal{K}_\nu$ is a modified Bessel function of the second kind of order ν, where ν is a parameter of smoothness defined as $\nu = \alpha - d/2$. In INLA, $\alpha = 2$ is taken by default, and in our cases, the dimension of the analysis space is the plane ($d = 2$), resulting in a constant smoothness parameter value $\nu = 1$. Furthermore, under the SPDE approximation, the spatial range ρ is considered as the distance at which the correlation value is close to 0.1 (Lindgren et al. 2011).

This approach allows defining a spatial effect whose precision structure is sparse and approximates the solution of a continuous Gaussian field with a variance-covariance structure belonging to the Matérn family; where one particular case of this family is the exponential covariance function when $\nu = 1/2$. This spatial effect is defined by a Gaussian Markov Random Field (GMRF), $\mathbf{u} \sim \text{GMRF}(\mathbf{0}, \mathbf{Q})$, whose precision matrix for the standard case of $d = 2$ and $\nu = 1$ can be expressed as $\mathbf{Q} = \tau^2(\kappa^4\mathbf{C} + 2\kappa^2\mathbf{G} + \mathbf{G}\mathbf{C}^{-1}\mathbf{G})$. Here, $\mathbf{C}$ and $\mathbf{G}$ are two matrices that can be computed through the structure of a constrained refined Delaunay triangulation $\mathcal{T}$ by Finite Element Methods. The GMRF realisations at the node locations $\mathbf{s}_n$ can be projected to any other point within the mesh domain, usually to the observation locations $\mathbf{s}_o$, through a linear basis approximation synthesised in the matrix product of the projection matrix $\mathbf{A}(\mathbf{s}_n, \mathbf{s}_o)$ with the GMRF realisation at the mesh nodes $\mathbf{u}(\mathbf{s}_n)$, $\mathbf{u}(\mathbf{s}_o) = \mathbf{A}(\mathbf{s}_n, \mathbf{s}_o)\mathbf{u}(\mathbf{s}_n)$. This allows the spatial effect $\mathbf{u}(\mathbf{s}_o)$ to be rewritten as $\mathbf{u}(\mathbf{s}_o) \sim \text{GMRF}(\mathbf{0}, \mathbf{A}^T\mathbf{Q}\mathbf{A})$, which is clearly equivalent to $\mathbf{u}(\mathbf{s}_o) = \mathbf{A}(\mathbf{s}_n, \mathbf{s}_o)\mathbf{u}(\mathbf{s}_n)$.

6.2. DOWNSCALING MODEL WITH SPDE-FEM

The definition of a spatial effect whose structure is continuous in space and space-time allows for spatial and spatio-temporal downscaling. Let's assume we have the following structure for the linear predictors of the model of our data:

$$\begin{aligned} \eta_{i,1} &= \mathbf{X}\boldsymbol{\beta}_1 + \textstyle\sum_{k=1}^{K} f_{k,1}(z_i) + u_{st,1}(C_i, t_i) \\ &\vdots \\ \eta_{i,D} &= \mathbf{X}\boldsymbol{\beta}_D + \textstyle\sum_{k=1}^{K} f_{k,d}(z_i) + u_{st,D}(C_i, t_i) \end{aligned} \tag{14}$$

the downscaling is defined for the spatial structure when, for each area C_i, the spatially structured effect is calculated as follows:

$$u_{st}(C_i, t_i) = \frac{\int_{\mathbf{s}\in C_i} u_{st}(\mathbf{s}, t_i)\mathrm{d}\mathbf{s}}{|C_i|} \approx \frac{\sum_{i=1}^{n(\mathbf{s}_i \in C_i)} u_{st}(\mathbf{s}_i, t_i)}{n(\mathbf{s}_i \in C_i)} = \frac{\mathbf{A}(\mathbf{s}_n, s_i)\mathbf{u}_{st}(\mathbf{s}_n, t_i)}{n(\mathbf{s}_i \in C_i)} \tag{15}$$

where s_i are integration points, allowing the approximation of the integral in the region C_i by averaging the sum of the field at the integration points, and $n(s_i \in C_i)$ is the number of integration points in C_i; being the area of C_i determine as $|C_i| = \int_{\mathbf{s}\in C_i} \mathrm{d}\mathbf{s}$. If a spatio-temporal downscaling is implemented, then the integration would be performed in both dimensions, spatial s and temporal t, using integration points for each corresponding volume $C_i \times t_i$. Additionally, this procedure can also be applied to establish sub-models for those covariates with a spatial structure that are given in areas. These sub-models would involve the modelling of such covariates by incorporating the integral of the covariate scaled by a linear regression coefficient, analogous to how it is done in error models (Muff et al. 2015, 2017) in cases without downscaling. The linear predictors in Eq. (14) can be associated with either a Dirichlet distribution or a multivariate normal distribution related to some of the logratio transformations previously shown. If they correspond to a logratio transformation, it is necessary to re-write the linear predictors to incorporate the variance-covariance structure $\boldsymbol{\Sigma}$ of the multivariate normal distribution, as explained in previous sections.

The downscaling approach not only provides a way to obtain estimates on a continuous scale of the spatial effect but, due to its structure and implementation, also allows us to consider the same spatial effect for different spatial supports. This has several implications, as detailed in the initial references. One particular implication we would like to highlight is that when the spatial support varies over time-such as in the case of small areas for disease mapping or NUTS3-the use of a downscaled spatial or spatio-temporal effect ensures a coherent structure across the different supports, maintaining the spatial or spatio-temporal dependence of the data. Furthermore, this does not depend on the specific use of the SPDE-FEM approach, as it can be implemented with any other Gaussian field, such as spline and two-dimensional spline bases, or any other continuous field.

7. BIG DATA

Public smart card data, Wi-Fi access point data, wireless sensor networks, and data with spatial location information -such as social media data, mobile phone tracking, and other

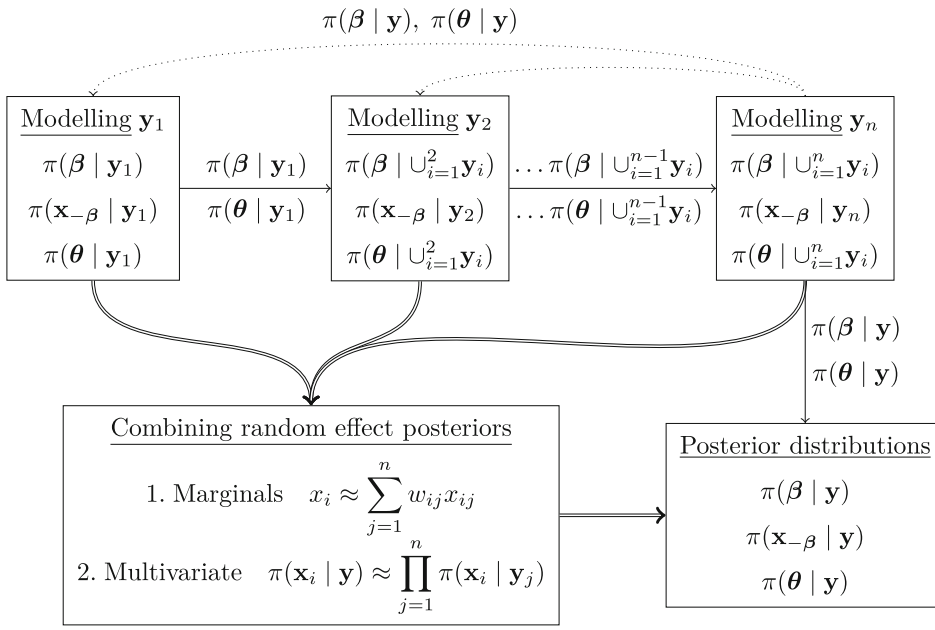

Figure 1. Scheme of the sequential consensus procedure. This approach assumes sequentially updating the fixed effects and hyperparameters, followed by performing a consensus for the random effects after these sequential updates (Figueira et al. 2024).

sensing information from Internet of Things devices- can provide useful ancillary data for LULC (Land Use and Land Cover) mapping. Compared to traditional geospatial data acquisition, these geospatial big data (GBD) are typically obtained at a lower cost and offer different coverages and better spatio-temporal resolutions (Liu et al. 2020; Zhang and Li 2022). They contain abundant human activity information, which can compensate for the lack of socio-economic data (Liu et al. 2020). By leveraging both Remote Sensing (RS) and GBD, it is possible to examine the physical and socio-economic characteristics of the urban land system (Martí et al. 2019). Despite the great potential of integrating RS and GBD for enhanced insights into urban land use, significant challenges remain in storing, managing, analysing, and visualising these data due to differences in spatial data quality (e.g. semantics, timestamps, and scale), technical formats, and data structures (Li et al. 2016). Therefore, a large amount of land use information is accessible, particularly with global and low-scale land use data, which gives rise to a big data problem.

This entails dealing with very heterogeneous and high dimensional data, particularly when data are available at global scale (Stanimirova et al. 2023) or when LULC data at regional scales are available at fine grid (You and Wood 2005; Chakir 2009). The joint analysis of data that are at different scales and whose spatial structure is in different supports, as well as combining different sources of information, means that modelling structures can be particularly complex (Figueira et al. 2023, 2024). In order to deal with these large databases and complex modelling structures, sequential inference procedures can be implemented in INLA, which reduces the computational cost. In particular, a sequential consensus inferential procedure can be established and implemented in `R-INLA` (Figueira et al. 2024).

This procedure, summarised in Fig. 1, involves the marginal updating of fixed effects and hyperparameters over a given partition of the data. The partition of the data can be done along the various sources of information available, such as different likelihoods or groups of likelihoods, or by leveraging the structure of the latent field (spatial, temporal, or spatio-temporal). Once the sequential inference is performed, the random effects information is combined using a consensus approach according to their (i) marginal or (ii) multivariate distributions. This approach also accommodates complex models with shared components between different likelihoods.

This procedure reduces computational burden in terms of memory and CPU usage by solving the dataset or the original model through partitions. While it provides good estimates for the latent field, the evaluation of hyperparameters may differ from that obtained by analysing the full model without partitioning. This is because only the marginal information is updated, and hyperparameters can exhibit non-negligible correlations in the posterior distribution.

8. EXAMPLES

In this section, several examples are provided to illustrate the implementation of the various methodological approaches described in the previous sections. These range from handling zeros and ones using adaptations of Hurdle models-widely used in environmental and ecological sciences-to methods for downscaling and procedures for managing large databases or complex models.

The first example demonstrates how to handle compositional data with 0's and 1's, using simulated data. The second example illustrates the results of a downscaling model applied to real land use data from the European LAMASUS project. Finally, the third example uses a large simulated dataset to showcase the effectiveness of the sequential consensus procedure in analysing extensive datasets and accurately estimating the underlying process.

8.1. DEALING WITH 0'S EXAMPLE

In this first example, we present two cases that illustrate how to handle the presence of zeros in CoDa. The first case uses data simulated from a Beta distribution to exemplify the simplest scenario, proving the Hurdle model for dealing with zeros. The second case also addresses CoDa, analysing 3 categories through a multivariate model of the CLR transformation.

In the Beta case, we simulate the data from the following hierarchical conditional model:

$$\begin{aligned} Z_i &\sim \text{Ber}(\pi_i), \\ \text{logit}(\pi_i) &= \beta_B + \alpha \cdot (\beta_0 + \beta_1 \cdot x_i + u_i), \\ Y_j \mid Z_j = 1 &\sim \text{Beta}(\mu_j, \phi), \\ \text{logit}(\mu_j) &= \beta_0 + \beta_1 \cdot x_j + u_{sj}, \end{aligned} \tag{16}$$

where Z_i are Bernoulli random variables explaining the process that generates the incidence vector (incidence matrix for CoDa), such that $z_i = 0 \iff y_i = 0$ and $z_1 \iff y_i \neq 0$.

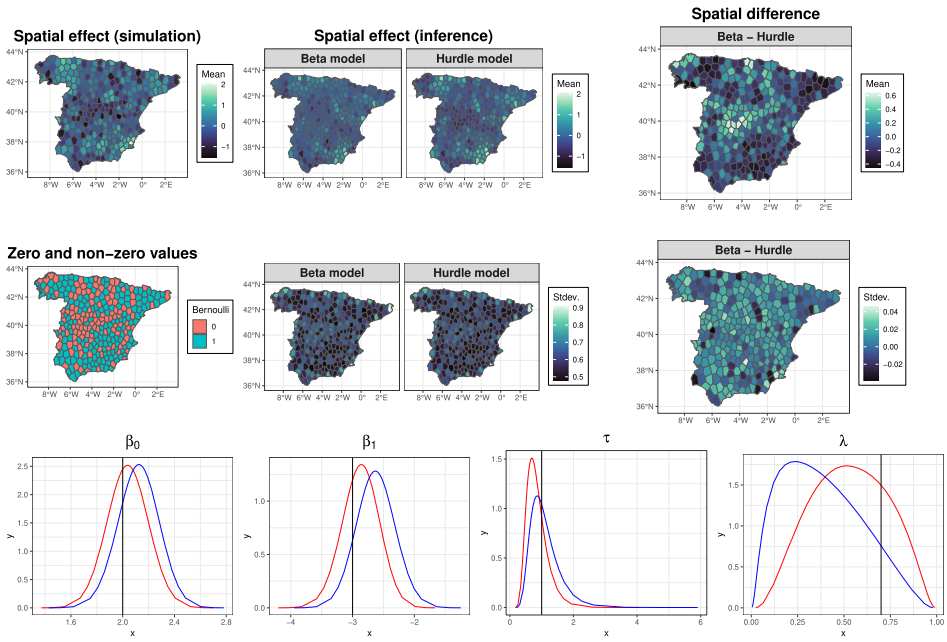

Figure 2. Simulated spatial effect, spatial distribution of zero and nonzero values, inferred spatial effect (mean and standard deviation), differences in the mean and standard deviation of the spatial structure between the Beta and Hurdle models, and posterior distributions of the fixed parameters and hyperparameters are presented. Distributions from the Hurdle model are shown in red, while those from the Beta model are shown in blue.

Y_j are Beta random variables conditioned on $Z_j = 1$, which implies that $Y_j \neq 0$. The latent field comprises an intercept β_0, a linear coefficient β_1 related to a covariate $\mathbf{x}$, and a spatial effect $\mathbf{u}_s$ following a Leroux structure. The spatial adjacency structure is built by partitioning mainland Spain with a Voronoi diagram, as shown in Fig. 2. This latent field is shared in the linear predictor of the Bernoulli process, scaled by α and including an intercept β_B that controls the overall proportion of 0's. Finally, ϕ stands for the precision of the Beta distribution.

Figure 2 shows the simulated spatial structure and the distribution of null and nonzero values, along with the inferential results of the beta model without incorporating the zeros and the Hurdle model incorporating the null values. It can be seen that, in general, the Hurdle model better captures the true values of the spatial effects where there are many zeros, as well as the posterior distributions of the model parameters. In the figure, we can see a slightly better identification of the $\boldsymbol{\beta}$ components, along with improved identification of the marginal precision τ and, in particular, the spatial dependence parameter λ of the Leroux distribution. To assess these improvements, we consider the DIC of the models and the RMSE for the spatial trend, as we have the true values of the spatial pattern from the simulation. RMSE is used to evaluate the improvement in identifying the spatial pattern according to the model. The DIC values for the models are -359 (Beta model) and -363 (Hurdle model), while the RMSE values are 0.56 for the Beta model and 0.51 for the Hurdle-like model.

The CoDa simulation is performed considering that only one ($\mathbf{Y}_1$) of the 3 categories ($\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3$) can present zero values. This model serves as a conceptual example of how it

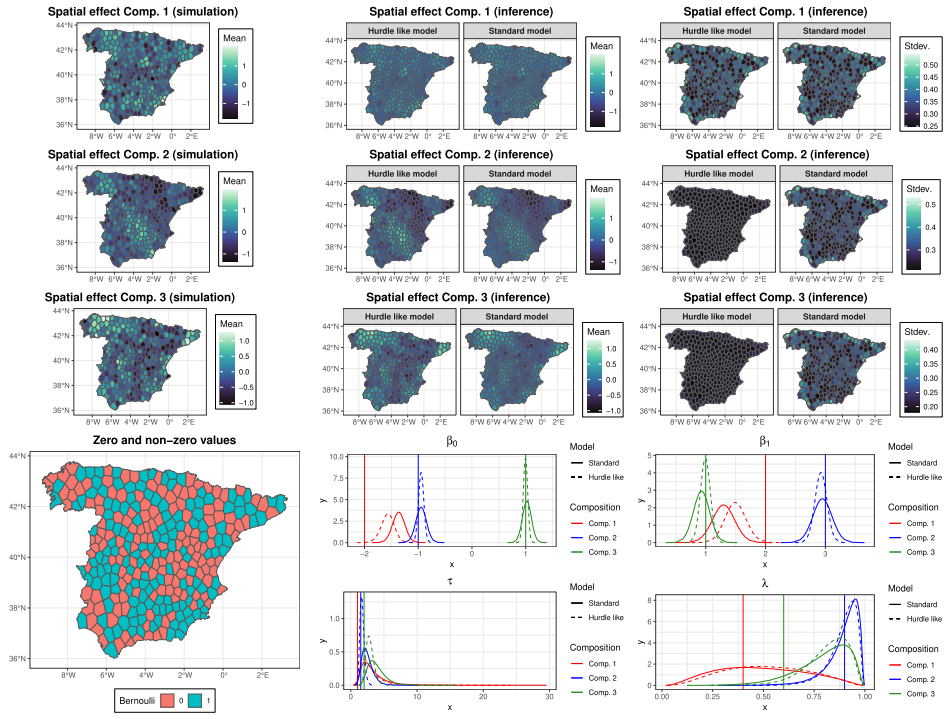

Figure 3. Simulated spatial effects, spatial distribution of the zero and nonzero values, and inferred spatial effects (mean and stdev.), along with the posterior distributions of the fixed parameters and hyperparameters.

can be implemented for CoDa, dealing with zeros and ones in several categories. Therefore, the model structure when no null values are present is as follows:

$$\begin{aligned} Z_{i1} &\sim \text{Ber}(\pi_i), \\ Y_{j1} \mid Z_{j1} = 1 &\sim \text{N}(\mu_{j1}, \tau^*), \\ Y_{j2} \mid Z_{j1} = 1 &\sim \text{N}(\mu_{j2}, \tau^*), \\ Y_{j3} \mid Z_{j1} = 1 &\sim \text{N}(\mu_{j3}, \tau^*), \end{aligned} \tag{17}$$

where the linear predictor for the CLRs are defined with the same structure as the one used for the Beta example with the addition of the $\mathbf{u}$ effect to take into account the correlation between the categories, $\eta_{id} = \beta_{0d} + \beta_{1d} x_i + u_{sid} + u_i$. However, when $Z_{i1} = 0$ the CLR change to the remaining two categories $(\mathbf{Y}_2, \mathbf{Y}_3)$ using the same underlying process in the linear predictors as before.

Figure 3 shows the simulated spatial effects and the distribution of the zero and nonzero values for the first composition (*Comp. 1*). The figure also shows the inferred spatial effects (mean and stdev.) for the different compositions, along with the distributions for the fixed parameters and hyperparameters. In the figure, we can see an improved identification of the $\boldsymbol{\beta}$ components, along with improved identification of the marginal precisions $\boldsymbol{\tau}$ and slight changes in the spatial dependence parameters $\boldsymbol{\lambda}$ of the Leroux distributions for the spatial effects.

In both cases we can observe an improvement in the estimation of the model by incorporating the Hurdle-type structure to integrate the analysis of the zeros. To assess this improvement, the DIC is also considered for the two models, the Hurdle-like model and the standard model. As a measure for evaluating the improvement in identifying the spatial pattern according to the model, RMSE has been used, given that we have the true values of the spatial effect from the simulation. The DIC between the models is: -4615 (Hurdle-like) and -2754 (standard model), while the RMSE is (0.50, 0.25, 0.20) for the Hurdle-like model and (0.54, 0.34, 0.31) for the standard model, for spatial effects of component 1, 2, and 3, respectively.

8.2. DOWNSCALING EXAMPLE: LAMASUS DATA

In this example, we present the implementation of the downscaling procedure for real land use data from the European LAMASUS project. In the context of the LAMASUS project, we aim to identify the components that determine or drive land use and land use change in Europe over the last decades. The data used in this example is spatially structured at the NUTS3 level between the years 2007 and 2018; where NUTS (Nomenclature of Territorial Units for Statistics) is the hierarchical system for dividing the economic territory of the European Union. The land use data analysed in this example is classified into five aggregated categories to avoid the presence of zero and unit values. These categories are: (i) cropland, (ii) grassland, (iii) forest, (iv) urban, and (v) other natural land. A wide range of explanatory variables are available to analyse the underlying processes that determine land use, including socio-economic and demographic variables such as GDP, GVA, and employment, as well as environmental variables like aspect and elevation.

We begin by focussing on the proportion of agricultural land use, specifically the cropland category. To model exclusively the proportion of cultivated land use, we employ a downscaling model using a Beta distribution. Specifically, we use a spatio-temporal hierarchical Bayesian model, as outlined as follows:

$$\begin{aligned} Y_{it} &\sim \text{Beta}\,(\mu_{it}, \phi) \\ \text{logit}\,(\mu_{it}) &= \boldsymbol{\beta} X_{it} + \textstyle\int_{\mathbf{s}\in C_i} \frac{u(\boldsymbol{s})\mathrm{d}\mathbf{s}}{|C_i|} + u_t, \end{aligned} \tag{18}$$

where Y_{it} represents the random variable for the proportion of cropland use in a specific region i and year t. The conditional distribution of Y_{it} follows a Beta distribution with mean μ_{it} and precision ϕ. The mean is linked through the logit function to the linear predictor. The covariates are represented by $\mathbf{X}_{it}$, and the vector of the linear regression coefficients are $\boldsymbol{\beta}$. The downscaling spatial effect is $u(\mathbf{s})$, while the temporal effect is modelled as a first-order autoregressive process $u_t = \phi_t u_{t-1} + \epsilon(\tau_t)$, where ϕ_t and τ_t are the temporal autoregressive parameter and the marginal precision of the temporal autoregressive effect, respectively.

The selection of covariates to be included in the model was performed by exploring all possible combinations of models without spatial and temporal effects. The combination with the lowest Watanabe-Akaike Information Criterion (WAIC; Watanabe (2013)) was

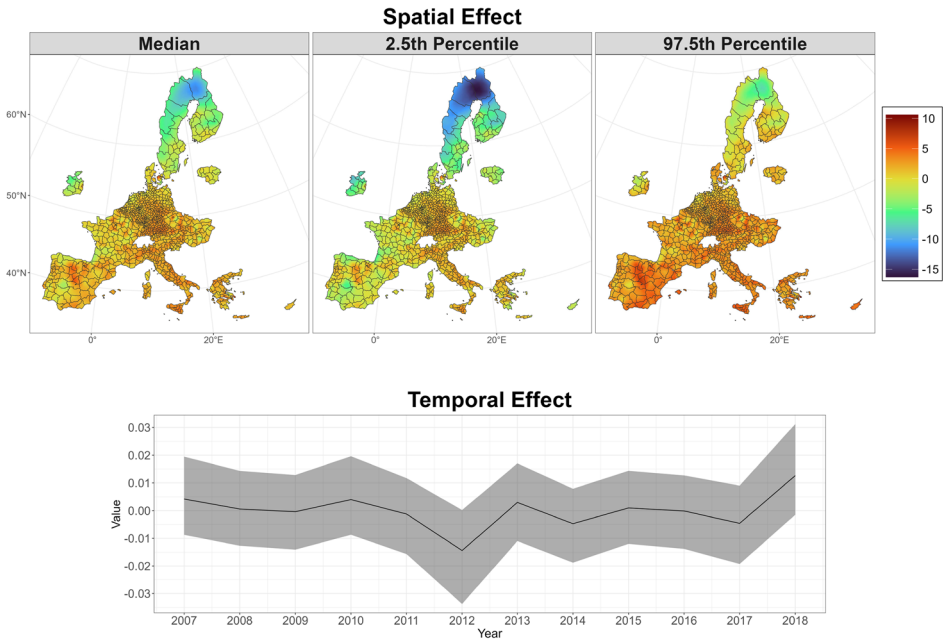

Figure 4. Cartographic representation of the continuous spatial effect and the temporal effect of the downscaling model.

selected. This means that for the 22 explanatory variables available, all possible inclusion and exclusion combinations were analysed, resulting in a total of 2^{22} models. The model with the lowest WAIC was chosen.

Some results are shown in Fig. 4. As the figure shows, the temporal effect is small. As for the spatial effect, it captures the variability not explained by the covariates. We can observe that, in certain areas, the spatial effect has a greater impact on the overall average effect than in other regions, where its impact is less pronounced. Regarding the covariates, some notable ones include decoupled payments (in euros) normalised by the total used agricultural area (in hectares), and less favoured area subsidies (in euros) normalised by the total used agricultural area (in hectares). Additionally, other relevant variables are the area of the region, elevation of the terrain, GVA, and total output (in euros).

The analysis for the 5 aggregated land use categories is performed using a multivariate Gaussian model with the ALR transformation of the categories. To select the covariates, we implemented a stepwise search algorithm driven by the WAIC of the models after a pre-selection where highly correlated ($\rho > 0.75$) variables were extracted. This algorithm involves the following steps: (i) a forward search among the covariates until no additional covariate can be included, then (ii) a backward step until no covariate can be deleted, and (iii) repeating steps (i) and (ii) until there is no change. The base model into which the algorithm is implemented encompasses a specific intercept, spatial, and temporal component for each ALR. The final model includes 9 explanatory covariates, and the model for the mean $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_4)$ of the multivariate Gaussian distribution $\text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is written as

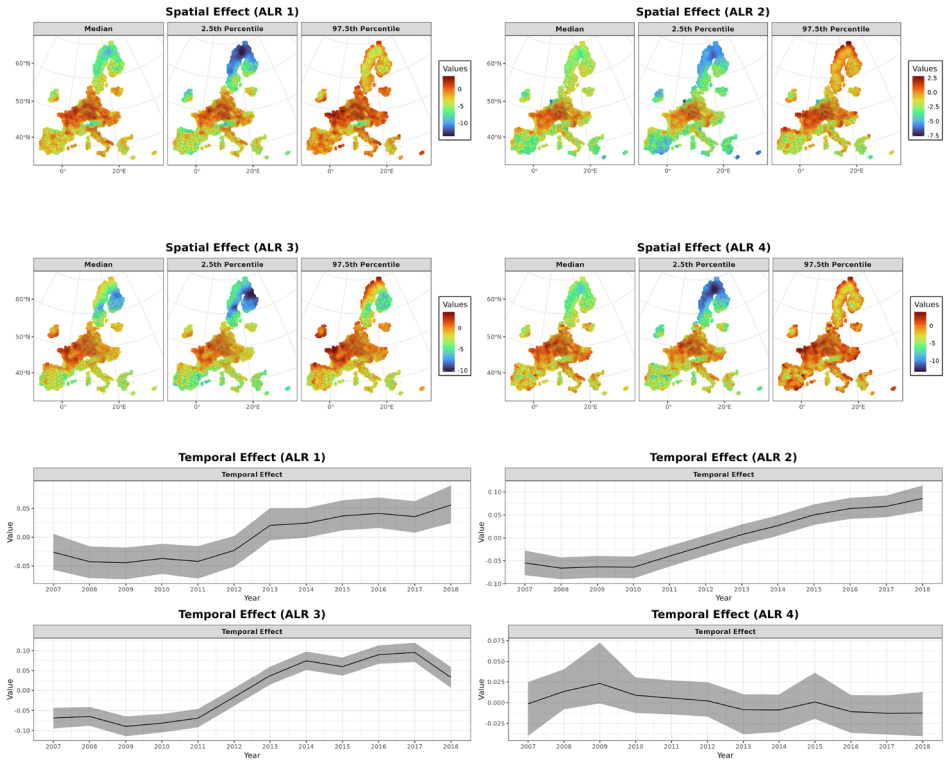

Figure 5. Cartographic representation of the continuous spatial effects and the temporal effects for the different ALRs of the downscaling model.

follows:

$$\begin{aligned}
\mu_{i1} &= \beta_{01} + \mathbf{X}_i\boldsymbol{\beta}_1 + u_{si1} + u_{ti1} + u_i, \\
\mu_{i2} &= \beta_{02} + \mathbf{X}_i\boldsymbol{\beta}_2 + u_{si2} + u_{ti2} + u_i, \\
\mu_{i3} &= \beta_{03} + \mathbf{X}_i\boldsymbol{\beta}_3 + u_{si3} + u_{ti3} + u_i, \\
\mu_{i4} &= \beta_{04} + \mathbf{X}_i\boldsymbol{\beta}_4 + u_{si4} + u_{ti4} + u_i,
\end{aligned} \tag{19}$$

where u_{sid} is the downscaled spatial effect, $u_{sid} = \int_{\mathbf{s}\in C_i} u_{sd}(\mathbf{s})\mathrm{d}\mathbf{s}/|C_i|$, and the component $\mathbf{u}$ accounts for the correlation $\boldsymbol{\Sigma}^{-1}$ between the categories, as explained in the corresponding section.

Figure 5 shows the downscaled continuous spatial effects for the different ALRs, together with the temporal effects. In this case, we can observe the scaling effect for the different ALRs, which implies a common spatial dependency structure over the years. This is combined with the temporal effect for the log-ratios, where the trend is similar for the first three, while the fourth shows a temporal effect with greater uncertainty and is less pronounced.

8.3. BIG DATA EXAMPLE

In this final example, we present the implementation of the algorithm summarised in Fig. 1 for evaluating large databases of compositional data. The purpose of the simulation

is not merely to demonstrate the results of the algorithm implementation for handling large databases, but also to showcase its versatility and its application to spatio-temporal downscaling models. This enables the processing of structured data in areas with spatial support that varies over the years.

The simulation is conducted for three categories evaluated using the ALR transformation, so the multivariate Gaussian structure will have a dimension of 2. The spatial structure is simulated using a continuous Gaussian field aggregated over three different spatial supports, each structured into areas. Specifically, the aggregation is based on three distinct areal structures, with 200, 250, and 300 areas, as shown in Fig. 6, making the downscaling model purely spatial. The Gaussian field is a separable spatio-temporal effect, defined by the following precision matrix: $\mathbf{Q}_{st} = \mathbf{Q}_s \otimes \mathbf{Q}_t$. The spatial precision matrix is defined using an SPDE in two dimensions, while the temporal part is determined by a first-order autoregressive structure with 600 (temporal) nodes. To construct the different spatial supports, the territory of mainland Spain has been partitioned using Voronoi diagrams.

The model used to simulate and infer the data is as follows:

$$\begin{aligned} \mu_{it1} &= \beta_{01} + \int_{\mathbf{s}\in C_i} u_1(\mathbf{s}, t)\mathbf{ds} + u_{it}, \\ \mu_{it2} &= \beta_{02} + \int_{\mathbf{s}\in C_i} u_2(\mathbf{s}, t)\mathbf{ds} + u_{it}, \end{aligned} \tag{20}$$

where β_{0d} are the intercepts, $\mathbf{u}_d(\mathbf{s}, t)$ are the spatio-temporal effects aggregated over the different spatial supports for each ALR. Finally, the effect $\mathbf{u}$ accounts for the correlation structure between the different ALRs, allowing for the implementation of the multivariate Gaussian distribution in `R-INLA`.

Figure 6 shows the mesh for the spatial effect and the different spatial supports in which the data was aggregated over the various years. It also presents the aggregated spatial effect using simulated data, as well as the desegregated spatial effects obtained from the inferential process of the two ALRs for the first temporal node. The data set was divided along the temporal scale, grouping into 6 temporal nodes, as larger grouping would exceed the server's memory capacity, which had 253 GB of RAM and 93 cores. Additionally, the complete analysis by the sequential consensus procedure for the entire dataset took 314.28 minutes.

9. CONCLUSIONS

In the field of land use science, the application of spatial or spatio-temporal models is essential for evaluating the key factors that modify and direct land use and land use changes. Assessing spatial patterns alongside temporal patterns provides a precise understanding of their distribution in space and their temporal evolution. To achieve this, approaches often use either large-scale aggregated data or small-scale disaggregated data. In the latter case, when small-scale data are available, two natural problems may arise: the presence of null values and computational complications when handling large databases. Conversely, when dealing with large-scale data, the issue of changing the spatial aggregation support over time may occur. This study presents methodologies to address the problem of null values, downscaling procedures for spatial or spatio-temporal structures, and big data challenges. These methodologies are illustrated with various examples to demonstrate their implementation.

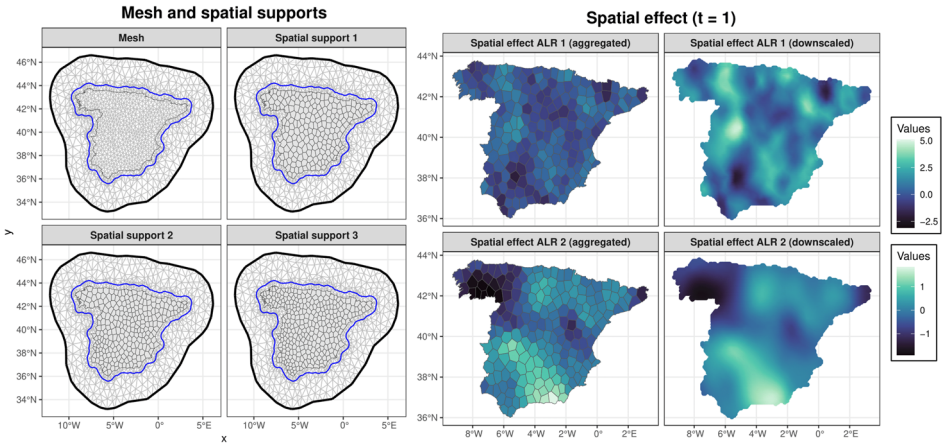

Figure 6. Mesh for spatial effects, spatial supports for data aggregation and outputs, comparing the aggregated simulated data with the downscaled spatial fields obtained from the inference for the first temporal node.

The results presented in the examples show the improvement when the presence of zeros is integrated into the analysis, along with the ability to perform dowscaling or disaggregation models using continuous Gaussian fields, specifically through the SPDE-FEM approach. Additionally, the implementation of a dowscaling model on a large dataset across different aggregation supports was shown. This highlights the possibility of encountering aggregated data in different spatial structures over the years, and how to deal with them in a Big Data context. This can be achieved by implementing the sequential consensus algorithm; otherwise, computational limitations would render the analysis unfeasible.

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

Data availability The link to the code and the data used can be found in the following GitHub repository: https://github.com/MarioFigueiraP/JABES_Unveiling_land_use_Dynamics.

Declarations

Conflict of interest The authors declare no conflict of interest.

[Received August 2024. Revised January 2025. Accepted January 2025.]

REFERENCES

Aitchison J (1982) The statistical analysis of compositional data. J R Stat Soc Ser B Stat Methodol 44:139–160. https://doi.org/10.1111/j.2517-6161.1982.tb01195.x

Aitchison J (1986) The statistical analysis of compositional data. Chapman & Hall, London

Aitchison J, Egozcue JJ (2005) Compositional data analysis: where are we and where should we be heading? Math Geol 37:829–850. https://doi.org/10.1007/s11004-005-7383-7

Aitchison J, Kay J (2003) Possible solution of some essential zero problems in compositional data analysis. In: Proceedings of CoDaWork'03, The 1st Compositional Data Analysis Workshop . University of Girona, Girona (Spain). http://ima.ud.es/Activitats/CoDaWork03/

An Y, Lee S (2020) A spatial disaggregation model to improve long-term land use forecasting with transport models based on zonal data. Appl Spat Anal Policy 13:187–208. https://doi.org/10.1007/s12061-019-09298-3

Bakka H, Rue H, Fuglstad G-A, Riebler A, Bolin D, Illian J, Krainski E, Simpson D, Lindgren F (2018) Spatial modeling with r-inla: a review. WIREs Comput Stat 10(6):1443. https://doi.org/10.1002/wics.1443

Banerjee S, Carlin BP, Gelfand AE (2014) Hierarchical modeling and analysis for spatial data. Chapman and Hall/CRC, London. https://doi.org/10.1201/b17115

Bareille F, Chakir R (2024) Structural identification of weather impacts on crop yields: disentangling agronomic from adaptation effects. Am J Agric Econ 106(3):989–1019. https://doi.org/10.1111/ajae.12420

Berild MO, Martino S, Gómez-Rubio V, Rue H (2022) Importance sampling with the integrated nested laplace approximation. J Comput Graph Stat 31:1225–1237. https://doi.org/10.1080/10618600.2022.2067551

Berrocal VJ, Gelfand AE, Holland DM (2010) A spatio-temporal downscaler for output from numerical models. J Agric Biol Environ Stat 15:176–197. https://doi.org/10.1007/s13253-009-0004-z

Besag J (1974) Spatial interaction and the statistical analysis of lattice systems. J R Stat Soc Series B (Methodological) 36:192–236

Besag J, York J, Mollié A (1991) Bayesian image restoration, with two applications in spatial statistics. Ann Inst Stat Math 43(1):1–20. https://doi.org/10.1007/BF00116466

Bivand RS, Gómez-Rubio V, Rue H (2014) Approximate Bayesian inference for spatial econometrics models. Spatial Stat 9:146–165. https://doi.org/10.1016/j.spasta.2014.01.002

Bradley JR, Wikle CK, Holan SH (2016) Bayesian spatial change of support for count-valued survey data with application to the american community survey. J Am Stat Assoc 111:472–487. https://doi.org/10.1080/01621459.2015.1117471

Cameron AC, Trivedi PK (2013) Regression analysis of count data. Cambridge University Press. https://doi.org/10.1017/CBO9781139013567

Chakir R (2009) Spatial downscaling of agricultural land-use data: an econometric approach using cross entropy. Land Econ 85(2):238–251. https://doi.org/10.3368/le.85.2.238

Chakir R, Le Gallo J (2013) Predicting land use allocation in France: a spatial panel data analysis. Ecol Econ 92:114–125. https://doi.org/10.1016/j.ecolecon.2012.04.009

Chakir R, Le Gallo J (2021) Spatial autocorrelation in econometric land use models: an overview. In: Daouia A, Ruiz-Gazen A (eds) Advances in contemporary statistics and econometrics: festschrift in honor of Christine Thomas-Agnan. Springer International Publishing, Cham, pp 339–362. https://doi.org/10.1007/978-3-030-73249-3_18

Chakir R, Lungarska A (2017) Agricultural rent in land-use models: comparison of frequently used proxies. Spat Econ Anal 12(2–3):279–303. https://doi.org/10.1080/17421772.2017.1273542

Chakir R, Parent O (2009) Determinants of land use changes: a spatial multinomial probit approach. Pap Reg Sci 88(2):327–345. https://doi.org/10.1111/j.1435-5957.2009.00239.x

Clayton D (1996) Generalized linear mixed models. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (eds.) Markov Chain Monte Carlo in Practice, pp. 275–301. Chapman and Hall, London . Chap. 16

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

Egozcue JJ, Pawlowsky-Glahn V (2006) Simplicial geometry for compositional data. Geol Soc, London, Special Publ 264:145–159. https://doi.org/10.1144/GSL.SP.2006.264.01.11

Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C (2003) Isometric logratio transformations for compositional data analysis. Math Geol 35:279–300. https://doi.org/10.1023/A:1023818214614

Elhorst JP (2014) Spatial econometrics: from cross-sectional data to spatial panels. Springer Berlin Heidelberg, Berlin, Heidelberg

Figueira M, Barber X, Conesa D, López-Quílez A, Martínez-Minaya J, Paradinas I, Pennino MG (2023) Bayesian feedback in the framework of ecological sciences . arXiv: 2305.17922

Figueira M, Conesa D, López-Quílez A, Paradinas I (2024) A computationally efficient procedure for combining ecological datasets by means of sequential consensus inference . arXiv: 2406.08174

Fiserová E, Hron K (2011) On the interpretation of orthonormal coordinates for compositional data. Math Geosci 43:455–468. https://doi.org/10.1007/s11004-011-9333-x

Gelfand AE (2001) On the change of support problem for spatio-temporal data. Biostatistics 2:31–45. https://doi.org/10.1093/biostatistics/2.1.31

Gómez-Rubio V (2020) Bayesian inference with INLA. Chapman and Hall/CRC, London. https://doi.org/10.1201/9781315175584

Gómez-Rubio V, Rue H (2018) Markov chain monte carlo with the integrated nested laplace approximation. Stat Comput 28:1033. https://doi.org/10.1007/s11222-017-9778-y

Gómez-Rubio V, Bivand RS, Rue H (2021) Estimating spatial econometrics models with integrated nested laplace approximation. Mathematics 9(17):2044. https://doi.org/10.3390/math9172044

Greenacre M, Grunsky E, Bacon-Shone J, Erb I, Quinn T (2023) Aitchison's compositional data analysis 40 years on: a reappraisal. Stat Sci 38(3):386–410. https://doi.org/10.1214/22-STS880

Greenacre M, Grunsky E (2019) The isometric logratio transformation in compositional data analysis: a practical evaluation. Proceedings of the 8th International Workshop on Compositional Data Analysis (CoDaWork2019): Terrassa, 3-8 June, 2019. Universitat Polit'ecnica de Catalunya-BarcelonaTECH, 2019, p. 4–10, 74–81

Hersperger AM, Bürgi M (2009) Going beyond landscape change description: quantifying the importance of driving forces of landscape change in a central Europe case study. Land Policy 26(3):640–648. https://doi.org/10.1016/j.landusepol.2008.08.015

Hijazi RH, Jernigan RW (2009) Modeling compositional data using Dirichlet regression models. J Appl Prob Stat 4:77–91

Knorr-Held L (2000) Bayesian modelling of inseparable space-time variation in disease risk. Stat Med 19(17–18):2555–2567

Krisztin T, Piribauer P, Wögerer M (2022) A spatial multinomial logit model for analysing urban expansion. Spat Econ Anal 17:223–244. https://doi.org/10.1080/17421772.2021.1933579

Leroux BG, Lei X, Breslow N (2000) Estimation of disease rates in small areas: a new mixed model for spatial dependence. In: Halloran ME, Berry D (eds) Statistical models in epidemiology, the environment, and clinical trials. Springer New York, New York, NY, pp 179–191. https://doi.org/10.1007/978-1-4612-1284-3_4

LeSage JP (2008) An introduction to spatial econometrics. Rev Econ Ind 123(4):19–44. https://doi.org/10.4000/rei.3887

LeSage J, Pace RK (2009) Introduction to spatial econometrics. Chapman and Hall/CRC, London. https://doi.org/10.1201/9781420064254

Li S, Dragicevic S, Castro FA, Sester M, Winter S, Coltekin A, Pettit C, Jiang B, Haworth J, Stein A, Cheng T (2016) Geospatial big data handling theory and methods: a review and research challenges. ISPRS J Photogramm Remote Sens 115:119–133. https://doi.org/10.1016/j.isprsjprs.2015.10.012

Lindgren F, Rue H (2015) Bayesian spatial modelling with r-inla. J Stat Softw 63(19):1–25. https://doi.org/10.18637/jss.v063.i19

Lindgren F, Rue H, Lindström J (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J R Stat Soc: Series B (Stat Methodol) 73(4):423–498

Lindgren F, Bakka H, Bolin D, Krainski E, Rue H (2024) A diffusion-based spatio-temporal extension of Gaussian Matérn fields. SORT-Stat Operat Res Transact 48(1):3–66. https://doi.org/10.57645/20.8080.02.13

Liu W, Wu W, Thakuriah P, Wang J (2020) The geography of human activity and land use: a big data approach. Cities 97:102523. https://doi.org/10.1016/j.cities.2019.102523

Lubbe S, Filzmoser P, Templ M (2021) Comparison of zero replacement strategies for compositional data with large numbers of zeros. Chemom Intell Lab Syst 210:104248. https://doi.org/10.1016/j.chemolab.2021.104248

MacNab YC (2022) Bayesian disease mapping: past, present, and future. Spatial Stat 50:100593. https://doi.org/10.1016/j.spasta.2022.100593

MacNab YC (2023) Revisiting gaussian markov random fields and bayesian disease mapping. Stat Methods Med Res 32:207–225. https://doi.org/10.1177/09622802221129040

Martí P, Serrano-Estrada L, Nolasco-Cirugeda A (2019) Social media data: challenges, opportunities and limitations in urban studies. Comput Environ Urban Syst 74:161–174. https://doi.org/10.1016/j.compenvurbsys.2018.11.001

Martinez-Beneito MA, Botella-Rocamora P (2019) Disease mapping: from foundations to multidimensional modeling, 1st edn. Chapman and Hall/CRC, New York

Martinez-Beneito MA, Botella-Rocamora P, Banerjee S (2017) Towards a multidimensional approach to Bayesian disease mapping. Bayesian Anal 12(1):239–259. https://doi.org/10.1214/16-BA995

Martínez-Minaya J, Rue H (2024) A flexible bayesian tool for coda mixed models: logistic-normal distribution with dirichlet covariance. Stat Comput 34:116. https://doi.org/10.1007/s11222-024-10427-3

Martínez-Minaya J, Lindgren F, López-Quílez A, Simpson D, Conesa D (2023) The integrated nested laplace approximation for fitting dirichlet regression models. J Comput Graph Stat 1–19:805. https://doi.org/10.1080/10618600.2022.2144330

Martínez-Minaya J, Cameletti M, Conesa D, Pennino MG (2018). Species distribution modeling: a statistical review with focus in spatio-temporal issues. https://doi.org/10.1007/s00477-018-1548-7

Martín-Fernández JA, Thió-Henestrosa S (2006) Rounded zeros: some practical aspects for compositional data. Geol Soc, London, Special Publ 264(1):191–201. https://doi.org/10.1144/GSL.SP.2006.264.01.14

Martín-Fernández JA, Palarea-Albaladejo J, Olea RA (2011) Dealing with Zeros. In: Pawlowsky-Glahn V, Buccianti A (eds) Compositional data analysis: theory and applications. Wiley, pp 43–58. https://doi.org/10.1002/9781119976462.ch4

Moindjié I-A, Pinsard C, Accatino F, Chakir R (2022) Interactions between ecosystem services and land use in France: a spatial statistical analysis. Front Environ Sci. https://doi.org/10.3389/fenvs.2022.954655

Moraga P, Cramb SM, Mengersen KL, Pagano M (2017) A geostatistical model for combined analysis of point-level and area-level data using inla and spde. Spatial Stat 21:27–41. https://doi.org/10.1016/j.spasta.2017.04.006

Muff S, Riebler A, Held L, Rue H, Saner P (2015) Bayesian analysis of measurement error models using integrated nested laplace approximations. J R Stat Soc. Series C (Appl Stat) 64:231–252

Muff S, Ott M, Braun J, Held L (2017) Bayesian two-component measurement error modelling for survival analysis using inla–a case study on cardiovascular disease mortality in switzerland. Comput Stat Data Anal 113:177–193. https://doi.org/10.1016/j.csda.2017.03.001

Mullahy J (1986) Specification and testing of some modified count data models. J Econo 33(3):341–365. https://doi.org/10.1016/0304-4076(86)90002-3

Nandi AK, Lucas TCD, Arambepola R, Gething P, Weiss DJ (2023) disaggregation : An r package for bayesian spatial disaggregation modeling. Journal of Statistical Software. https://doi.org/10.18637/jss.v106.i11

Pirzamanbein B, Poska A, Lindström J (2020) Bayesian reconstruction of past land cover from pollen data: model robustness and sensitivity to auxiliary variables. Earth Space Sci. https://doi.org/10.1029/2018EA000547

Prates MO, Azevedo DRM, MacNab YC, Willig MR (2022) Non-separable Spatio-temporal models via transformed multivariate gaussian Markov random fields. J R Stat Soc: Ser C: Appl Stat 71:1116–1136. https://doi.org/10.1111/rssc.12567

Rasmussen CL, Palarea-Albaladejo J, Johansson MS, Crowley P, Stevens ML, Gupta N, Karstad K, Holtermann A (2020) Zero problems with compositional data of physical behaviors: a comparison of three zero replacement methods. Int J Behav Nutr Phys Act 17(1):126. https://doi.org/10.1186/s12966-020-01029-z

Stanimirova R, Tarrio K, Turlej K, McAvoy K, Stonebrook S, Hu K-T, Arévalo P, Bullock EL, Zhang Y, Woodcock CE, Olofsson P, Zhu Z, Barber CP, Souza CM, Chen S, Wang JA, Mensah F, Calderón-Loor M, Hadjikakou M, Bryan BA, Graesser J, Beyene DL, Mutasha B, Siame S, Siampale A, Friedl MA (2023) A global land cover training dataset from 1984 to 2020. Scientific Data 10(1):879. https://doi.org/10.1038/s41597-023-02798-5

Tang M, Wu Q, Yang S, Tian G (2022) Dirichlet composition distribution for compositional data with zero components: an application to fluorescence in situ hybridization (FISH) detection of chromosome. Biom J 64:714–732. https://doi.org/10.1002/bimj.202000334

Thomas-Agnan C, Laurent T, Ruiz-Gazen A, Nguyen THA, Chakir R, Lungarska A (2021) In: Filzmoser, P., Hron, K., Martín-Fernández, J.A., Palarea-Albaladejo, J. (eds.) Spatial Simultaneous Autoregressive Models for Compositional Data: Application to Land Use, pp. 225–249. Springer, Cham . https://doi.org/10.1007/978-3-030-71175-7_12

Tsagris M, Stewart C (2018) A dirichlet regression model for compositional data with zeros. Lobachevskii J Math 39:398–412. https://doi.org/10.1134/S1995080218030198

van Vliet J, de Groot HLF, Rietveld P, Verburg PH (2015) Manifestations and underlying drivers of agricultural land use change in Europe. Landsc Urban Plan 133:24–36. https://doi.org/10.1016/j.landurbplan.2014.09.001

Ver Hoef JM, Hanks EM, Hooten MB (2018) On the relationship between conditional (CAR) and simultaneous (SAR) autoregressive models. Spatial Stat 25:68–85. https://doi.org/10.1016/j.spasta.2018.04.006

Villejo SJ, Illian JB, Swallow B (2023) Data fusion in a two-stage spatio-temporal model using the inla-spde approach. Spatial Stat 54:100744. https://doi.org/10.1016/j.spasta.2023.100744

Wang C, Furrer R (2021) Combining heterogeneous spatial datasets with process-based spatial fusion models: a unifying framework. Comput Stat Data Anal 161:107240. https://doi.org/10.1016/j.csda.2021.107240

Wang C, Puhan MA, Furrer R (2018) Generalized spatial fusion model framework for joint analysis of point and areal data. Spatial Stat 23:72–90. https://doi.org/10.1016/j.spasta.2017.11.006

Watanabe S (2013) A widely applicable Bayesian information criterion. J Mach Learn Res 14(1):867–897

You L, Wood S (2005) Assessing the spatial distribution of crop areas using a cross-entropy method. Int J Appl Earth Obs Geoinf 7:310–323. https://doi.org/10.1016/j.jag.2005.06.010

You L, Wood S (2006) An entropy approach to spatial disaggregation of agricultural production. Agric Syst 90:329–347. https://doi.org/10.1016/j.agsy.2006.01.008

Zhang C, Li X (2022) Land use and land cover mapping in the era of big data. Land 11:1692. https://doi.org/10.3390/land11101692