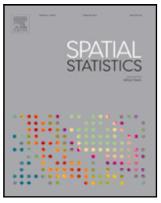
Spatial Statistics 67 (2025) 100893

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta



A spatial autoregressive graphical model

Sjoerd Hermes [a,b,*], Joost van Heerwaarden [a,b], Pariya Behrouzi [a]

[a] Mathematical and Statistical Methods, Wageningen University, The Netherlands
[b] Plant Production Systems, Wageningen University, The Netherlands

ARTICLE INFO

Keywords:
Graphical models
Spatial autoregressive models
Asymmetric effects
Intercropping

ABSTRACT

Within the statistical literature, a significant gap exists in methods capable of modelling asymmetric multivariate spatial effects that elucidate the relationships underlying complex spatial phenomena. For such a phenomenon, observations at any location are expected to arise from a combination of within- and between-location effects, where the latter exhibit asymmetry. This asymmetry is represented by heterogeneous spatial effects between locations pertaining to two different categories, that is, a feature inherent to each location in the data, such that based on the feature label, asymmetric spatial relations are postulated between neighbouring locations with different labels. Our novel approach synergises the principles of multivariate spatial autoregressive models and the Gaussian graphical model. This synergy enables us to effectively address the gap by accommodating asymmetric spatial relations, overcoming the usual constraints in spatial analyses. However, the resulting flexibility comes at a cost: the spatial effects are not identifiable without either prior knowledge of the underlying phenomenon or additional parameter restrictions. Using a Bayesian-estimation framework, the model performance is assessed in a simulation study. We apply the model on intercropping data, where spatial effects between different crops are unlikely to be symmetric, in order to illustrate the usage of the proposed methodology. An R package containing the proposed methodology can be found on https://CRAN.R-project.org/package=SAGM.

1. Introduction

Gaussian graphical models are statistical learning techniques used to make inference on conditional dependence relationships within a set of variables arising from a multivariate normal distribution (Lauritzen, 1996). Methodological developments have expanded the use of the Gaussian graphical model beyond non-temporal to temporal data (see for relevant research during the last 5 years Barigozzi and Brownlees, 2019; Paci and Consonni, 2020; Yang and Peng, 2020; Dallakyan et al., 2022). These developments allow for improved analyses of complex, multivariate temporal datasets. However, barely any attention has been given to the development of graphical models pertaining to the higher dimensional cousin of time: space. Where dynamic models deal with temporal variation, assuming data close in time (Yang and Peng, 2020), there are clear parallels with the spatial analogue, for which data close together in space can be expected to be similar due to potential presence of location-specific latent factors affecting the distributions of the data that pertain to the location.

One example of spatial dependence between different variables can be found in intercropping trials. Intercropping is the cultivation of multiple crop species in a single field. Certain traits of the crops, such as yield, are not only dependent on within-crop processes, but also depend on the effects of traits expressed by different crops growing on neighbouring plots (subsections of fields

* Corresponding author.
E-mail address: sjoerd.hermes@wur.nl (S. Hermes).

https://doi.org/10.1016/j.spasta.2025.100893
Received 17 July 2024; Received in revised form 31 January 2025; Accepted 5 March 2025
Available online 15 April 2025

that contain multiple plants of the same crop). These effects can be either positive (complementarity and facilitation) or negative (competition) (Bourke et al., 2021) with respect to biomass production and yield. Knowledge of such effects is invaluable in selecting intercropping combinations with the goal of yield-maximisation (Bourke et al., 2021). Contemporary research on intercropping systems has shown that these spatial effects exhibit asymmetry (Huang et al., 2017; Gao et al., 2021), that is, the effect of variables pertaining to crop c_1 on variables pertaining to crop c_2 is not the same as the effect of those same variables pertaining to crop c_2 on those of crop c_1.

Despite a growing interest in the design of productive intercropping systems (Federer, 2012), there has been little methodological development around the identification of the kind of multi-trait and multi-species interactions that would determine which crops should ideally be combined (Brooker et al., 2015). Given that intercropping data typically consist of observations on locations in a finite domain, two classes of models seem appropriate: the class of conditional autoregressive (CAR) models (Besag et al., 1991; Leroux et al., 2000), and the class of spatial autoregressive (SAR) models (Ord, 1975). Thanks to the Markov property, where any observation is independent of its non-neighbours given its neighbours, the CAR models have a more local dependence structure than the SAR models. Instead, the SAR models exhibit what is known as spillover effects, whereby a phenomenon at some location not only influences directly neighbouring locations, but whose influence permeates towards non-neighbouring locations. In this article, we study a type of SAR model, as we expect these spillovers to be present for intercropping systems and various agricultural processes in general (Long, 1998; Li et al., 2007; Griffin, 2010). A second reason as to why we focus on the class of SAR models is that it enables us to construct a type of graphical model that is easy to interpret, and forms the spatial analogue to the popular vector autoregressive graphical model (Abegaz and Wit, 2013; Paci and Consonni, 2020). The relationship between the class of CAR models and graphical models has been established by Dobra et al. (2011), Dobra (2016) and Liang (2019), which explicitly account for spatial autocorrelation by developing Bayesian models that construct two graphs: a neighbourhood graph where the vertices indicate different regions and the presence of an edge indicates whether the regions share a border (are neighbours), and a conditional dependence graph that shows which variables are independent given all other variables. However, this approach has no way of inferring the spatial effects of one variable of crop c_1 on another of crop c_2, which is of interest to researchers wanting to evaluate, for example, the impact of applying fertiliser on one plot on the growth of plants in surrounding plots. This inability to model heterogeneous multivariate spatial effects across multiple crops, where the heterogeneity arises from differing spatial effects per crop, whilst simultaneously capturing the complex dependence relations that occur within plots is not unique to CAR models, as this is also not possible with existing SAR models.

This article proposes a new statistical methodology: the spatial autoregressive graphical model. The methodological novelty arises from the method's capacity to learn multivariate asymmetric between-location effects, subject to various identifiability constraints, combined with the capacity of illustrating complex within-location effects through a conditional independence structure, whereby the between- and within-location effects are illustrated by means of a graph, thereby facilitating interpretability. The need for such a method extends beyond intercropping applications. Any phenomenon in which asymmetric spatial effects between two categories can be postulated provides a potential application. The categories are features inherent to each location in the data, such that based on the label of that feature, asymmetric spatial relations can occur between neighbouring locations with different feature labels. One potential area of application is the field of epidemiology, where researchers might want to evaluate the existence and extend of asymmetric effects in rural–urban disease transmissions and driving variables (Ferrari et al., 2010). Alternatively, for a more timely example, whether there is asymmetry among the inter-country transmission rates of COVID-19 between neighbouring countries, where the countries can be categorised in strict anti-COVID policy – assuming that the movement of people between countries remains possible – versus lax anti-COVID policy (Keita, 2020). A last example is within economics or sociology, where observed values on variables such as crime, unemployment rate, income per capita, population size and congestion within a city are not only a result of within-city processes, but are also affected by what happens in neighbouring cities (Goetzke, 2008). The asymmetry arises, for example, due to different spatial relations between cities in the USA with a democratic versus a republican mayor, which in turn might imply different policies being put into practice that affect the aforementioned variables.

We propose a spatial autoregressive graphical model that fills the methodological gap of methods than can capture asymmetric between-location spatial effects together with within-location effects. Our proposed spatial autoregressive graphical model builds on the recently proposed multivariate spatial autoregressive (MSAR) model by Yang and Lee (2017), who extended the univariate spatial autoregressive model to the multivariate setting. The asymmetry of spatial effects between two different categories, e.g. crops in intercropping, is captured through a straightforward manipulation of the spatial weight matrices of the MSAR model, whilst the within-location dependencies share some of the properties of the Gaussian graphical model; namely that the within-location effects are derived from an underlying network consisting of conditional dependencies, which, in turn, offer a parsimonious representation of the complex within-location dependencies as well as being easily interpretable for researchers. The proposed method has attractive asymmetric between-location independence relationships, that are not present in other spatial autoregressive models. Using this approach, we can identify positive and negative interaction effects between crops that optimise collective performance, thereby selecting combinations of genotypes or crops that are promising in intercropping situations. However, the resulting flexibility comes at a cost: the spatial effects are not identifiable without either prior knowledge of the underlying phenomenon or additional parameter restrictions.

Section 2 introduces and formalises the methodological framework, discusses some relevant properties of the model and delves into the identifiability issue. Bayesian inference is discussed in Section 3. Section 4 presents an elaborate simulation study, where the performance of the newly proposed method is evaluated on simulated spatial data. An application of the new method on real intercropping data, illustrating the usage of the proposed methodology, is given in Section 5. To facilitate researchers with the analysis of their own spatial data, an R package was developed. This package can be found at https://CRAN.R-project.org/package=SAGM.

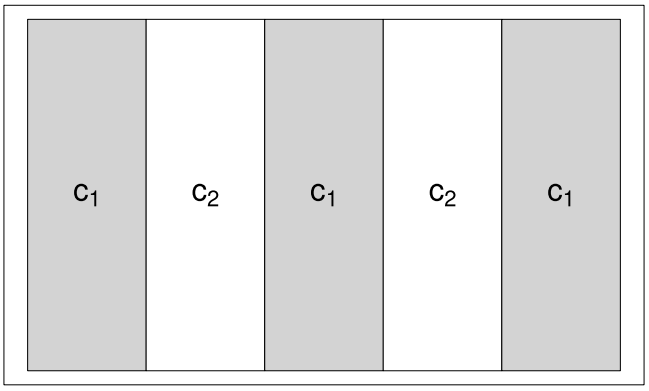


Fig. 1. Toy example with two different categories (e.g. crops): c_1 and c_2. Each rectangle represents a location.

2. Spatial autoregressive graphical model

In what follows, we use lower case symbols a to denote a scalar, $\boldsymbol{a}$ to denote vectors and $\boldsymbol{A}$ to denote matrices. Moreover, let $\boldsymbol{A}^T$ denote the transpose of $\boldsymbol{A}$ and $\boldsymbol{A}^{-T}$ the transpose of inverse matrix $\boldsymbol{A}^{-1}$. Furthermore, let c_1 and c_2 denote the two categories present in the data, then, for any data matrix $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ consisting of p variables with measurements across n locations or observations, $\boldsymbol{X}_{\{c_1,c_2\}}$ is used to denote a subset (rows) of $\boldsymbol{X}$ containing only those observations in $\boldsymbol{X}$ which pertain to categories c_1 and c_2. The Kronecker product is denoted by $\otimes$, whilst $\text{vec}(\boldsymbol{A})$ represents the vectorisation operation, which stacks the columns of $\boldsymbol{A}$, resulting in a vector. Let $\boldsymbol{I}_n$ denote an $n\times n$ identity matrix. Conversely, we use $\boldsymbol{O}_n$ to denote a $n\times n$ matrix containing only zeroes. For categories c_1 and c_2 let $\mathcal{N}_{c_1}$ denote the set of categories neighboured to category c_1. For a given graph $\mathcal{G}$, let $(i—j)$ denote an undirected edge between i and j and let $(i \to j)$ denote a directed edge from i to j. We use $\mathcal{A} \perp\!\!\!\perp \mathcal{B}$ to denote the independence between sets of variables $\mathcal{A}$ and $\mathcal{B}$ and $\mathcal{A} \perp\!\!\!\perp \mathcal{B}|\mathcal{S}$ to denote the conditional independence of $\mathcal{A}$ and $\mathcal{B}$ given another set of variables $\mathcal{S}$.

The proposed methodology requires some background knowledge about the MSAR model. Assume data arising from a stochastic process $\{X(l_i) : i = 1, \dots, n\}$, with $\bigcup_{i=1}^{n} l_i = \mathcal{D}$ and $l_i \cap l_{i'} = \emptyset$ for all $i \neq i'$. Ergo: the data are lattice data, where the lattice $\mathcal{D}$ is a non-random and finite domain consisting of locations $l_1, \dots, l_n$. The MSAR (Yang and Lee, 2017) is given by the following equation

$$\boldsymbol{X} = \boldsymbol{W}\boldsymbol{X}\boldsymbol{\Psi} + \boldsymbol{E}, \tag{1}$$

with $\boldsymbol{X} \in \mathbb{R}^{n\times p}, \boldsymbol{W} \in [0,1]^{n\times n}$, $\boldsymbol{\Psi} \in \mathbb{R}^{p\times p}$ and $\boldsymbol{E} \in \mathbb{R}^{n\times p}$ and where the row vectors of $\boldsymbol{E} = (\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n)^T, \boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{ip}), 1 \leq i \leq n$, are assumed to be i.i.d. normally distributed with mean vector $\boldsymbol{0}$ and positive definite covariance matrix $\boldsymbol{\Sigma}_E$. $\boldsymbol{W}$ is a known spatial weight matrix and spatial effect matrix $\boldsymbol{\Psi}$ indicates how variables in one location affect the value of variables in neighbouring locations and vice versa. For the MSAR model, the spatial effects captured by $\boldsymbol{\Psi}$ remain the same across all locations in $\mathcal{D}$. This is not realistic in the case of intercropping data, where spatial effects vary across locations due to the different crops (Huang et al., 2017; Gao et al., 2021). More general, whenever locations in spatial data can be assigned a category, such that spatial effects between locations pertaining to two different categories are expected to be asymmetric, the model proposed in (1) is unsatisfactory. This is because imposing a single spatial effect matrix is unable to account for the complexity of the data generating mechanism, thereby resulting in meaningless parameter estimates, see the online supplementary material. We illustrate this categorisation of locations and the proposed method by means of a toy example consisting of two categories: c_1 and c_2, which is shown in Fig. 1. In this toy example, we expect asymmetric spatial effects to occur between neighbours with a differing category label, motivating the need for different spatial effect matrices when the data contain two categories.

2.1. Accounting for asymmetric spatial effects

To account for the asymmetry in spatial effects found in Fig. 1, the model in Eq. (1) requires additional flexibility. This is done by setting

$$\boldsymbol{X} = \boldsymbol{W}_{c_2,c_1}\boldsymbol{X}\boldsymbol{\Psi}_{c_2,c_1} + \boldsymbol{W}_{c_1,c_2}\boldsymbol{X}\boldsymbol{\Psi}_{c_1,c_2} + \boldsymbol{E}, \tag{2}$$

where $\boldsymbol{\Psi}_{c_2,c_1} \in \mathbb{R}^{p\times p}$ contains the spatial effect of each variable of category c_2 on each variable of category c_1 and $\boldsymbol{\Psi}_{c_1,c_2} \in \mathbb{R}^{p\times p}$ contains the spatial effect of each variable of category c_1 on each variable of category c_2. Spatial weight matrices $\boldsymbol{W}_{c_2,c_1} \in [0,1]^{n\times n}$ and $\boldsymbol{W}_{c_1,c_2} \in [0,1]^{n\times n}$ are chosen such that only the matrix elements pertaining to directly neighbouring locations of respectively category c_1 and c_2 containing respectively category c_2 and c_1 are nonzero.

The locations found in spatial data are not necessarily homogeneous: they might exhibit differences with respect to the number of units (plants) per location, the location size, latent characteristics and other properties (Kelejian and Prucha, 2010). In such cases, assuming that $\text{Var}\left[\text{vec}\left(\boldsymbol{\varepsilon}_i\right)\right] = \boldsymbol{\Sigma}_E$ for all i can be overly restrictive. Instead, modifying the method found in Chan (2020), we let $\text{vec}(\boldsymbol{E}) \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_E \otimes \boldsymbol{\Omega}_E)$ for diagonal $n \times n$ covariance matrix $\boldsymbol{\Omega}_E$, such that we can account for this heteroskedasticity. If, for

example, we expect the variance of $\boldsymbol{E}$ to decrease with the number of plants per location, then we can a-priori define the diagonal elements of $\boldsymbol{\Omega}_E$ to be $1/r_i$, where r_i denotes the number of plants for location i, or estimate these quantities in a Bayesian fashion, whilst ensuring that they scale in r_i. Even though this research direction is not pursued further here, we hope that this suggestion is useful to researchers dealing with heteroskedastic data.

Given their importance for the proposed method, the spatial weight matrices merit a formal definition.

Definition 1 (*Spatial Weight Matrix Structure*)**.** Given two neighbouring locations with categories c_1 and c_2, the adjacency matrix $\boldsymbol{A}_{c_2,c_1}$, isolating the observations of category c_2 with neighbour category c_1, is of the following form:

$$a_{c_2,c_1,ij}\begin{cases}1, & \text{if location } i \text{ contains category } c_1 \text{ and location } j \text{ contains category } c_2 \text{ and}\\ & \text{locations } i \text{ and } j \text{ are neighbours}\\ 0, & \text{otherwise,}\end{cases}$$

where $a_{c_2,c_1,ij}$ denotes the i,jth element of $\boldsymbol{A}_{c_2,c_1}$. Subsequently, $w_{c_2,c_1,ij} = \frac{a_{c_2,c_1,ij}}{\sum_{j'=1}^{n} a_{c_2,c_1,ij'}}$.

Thus, following Fig. 1, we obtain the following weight matrices

$$\boldsymbol{W}_{c_2,c_1} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0.5 & 0 & 0.5 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad \boldsymbol{W}_{c_1,c_2} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0\\ 0.5 & 0 & 0.5 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0.5 & 0 & 0.5\\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3}$$

To ensure that the parameters can be estimated, spatial autoregressive models typically impose restrictions on the values that the spatial autoregressive parameter can take (c.f. Ord, 1975; Yang and Lee, 2017). Our method is no different in this regard.

Assumption 1 (*Stability Condition*)**.** To estimate the spatial effects, we require that $\det\left(\boldsymbol{I}_{np} - \sum_{k=1}^{2} \boldsymbol{\Psi}_k^T \otimes \boldsymbol{W}_k\right) > 0$, which is guaranteed whenever $\lambda_{\min}\left(\sum_{k=1}^{2} \boldsymbol{\Psi}_k^T \otimes \boldsymbol{W}_k\right) > -1$ and $\lambda_{\max}\left(\sum_{k=1}^{2} \boldsymbol{\Psi}_k^T \otimes \boldsymbol{W}_k\right) < 1$, where $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ are the minimum and maximum eigenvalues respectively.

In addition to the spatial effects contained in the $\boldsymbol{\Psi}_k$, $k \in \{1,2\}$, where k is used to denote the category ordering, our interest also lies in the within-location effects, contained in $\boldsymbol{\Theta}_E = \boldsymbol{\Sigma}_E^{-1}$, the precision matrix of $\boldsymbol{E}$. In order to obtain these within-location effects, a so-called spatial filter is required. Let $\boldsymbol{\Psi} = \{\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2\}$, the model specification in (2) allows for the construction of such a filter $\boldsymbol{R}(\boldsymbol{\Psi})\text{vec}(\boldsymbol{X}) = \text{vec}(\boldsymbol{E})$, where $\boldsymbol{R}(\boldsymbol{\Psi}) = \left(\boldsymbol{I}_{np} - \sum_{k=1}^{2} \boldsymbol{\Psi}_k^T \otimes \boldsymbol{W}_k\right)$ is the spatial filter matrix, filtering out the spatial effects on the observations as $\text{vec}(\boldsymbol{X}) - \left(\sum_{k=1}^{2} \boldsymbol{\Psi}_k^T \otimes \boldsymbol{W}_k\right)\text{vec}(\boldsymbol{X}) = \text{vec}(\boldsymbol{E})$, resulting in the within-location data, assuming that all spatial dependencies are accurately captured by the $\boldsymbol{W}_k$ (Getis, 1990; Millo, 2014). This result is straightforwardly obtained from the data generating process, $\text{vec}(\boldsymbol{X}) = \boldsymbol{R}(\boldsymbol{\Psi})^{-1}\text{vec}(\boldsymbol{E})$, which in turn is obtained by vectorising (2).

On a related note, as $\boldsymbol{X}$ is a linear combination of Gaussian variables, $\boldsymbol{X}$ itself is Gaussian with $\mathbb{E}[\text{vec}(\boldsymbol{X})] = \boldsymbol{0}$ and $\text{Var}[\text{vec}(\boldsymbol{X})] = \boldsymbol{R}(\boldsymbol{\Psi})^{-1}(\boldsymbol{\Sigma}_E \otimes \boldsymbol{I}_n)\boldsymbol{R}(\boldsymbol{\Psi})^{-T}$, see the online supplementary material. Consequently, $\text{vec}(\boldsymbol{X})$ is distributed as $\text{vec}(\boldsymbol{X}) \sim N_{np}\left(\boldsymbol{0}, \boldsymbol{R}(\boldsymbol{\Psi})^{-1}(\boldsymbol{\Sigma}_E \otimes \boldsymbol{I}_n)\boldsymbol{R}(\boldsymbol{\Psi})^{-T}\right)$, shortened as $\text{vec}(\boldsymbol{X}) \sim N_{np}(\boldsymbol{0}, \boldsymbol{\Sigma}_X)$. The resulting $\boldsymbol{\Sigma}_X = \boldsymbol{\Theta}_X^{-1}$ matrix is a $np \times np$ block matrix, where each block contains the variances or covariances within and between categories. $\boldsymbol{\Theta}_E$ consists of the original conditional dependences within a location, prior to being affected by any spatial processes. The $\boldsymbol{\Psi}_k$ can cause 0-elements in $\boldsymbol{\Theta}_E$ (e.g. the relation between variables i and j) to become nonzero in $\boldsymbol{\Theta}_X$ for locations $1, \ldots, n$. Therefore, evaluating the within-location elements of $\boldsymbol{\Theta}_X$ is inadequate when interest lies in the original conditional dependences within a location. In essence, in addition to the $\boldsymbol{\Psi}_k$, our interest lies in the original underlying conditional dependencies found in $\boldsymbol{\Theta}_E$, even though we only have data after the occurrence of spatial effects. By separating these effects, the original within-location dependencies can be obtained. As such, there is an equivalence between $\boldsymbol{\Theta}_E$, containing the within-location interactions, and the precision matrix corresponding to the time series chain graph (Dahlhaus and Eichler, 2003; Abegaz and Wit, 2013), which contains the contemporaneous interactions, in addition to establishing the connection to the Gaussian graphical model.

2.2. The spatial chain graph and its properties

Gaussian graphical models are multivariate statistical models that use graphs $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ to represent the full conditional dependence structure between variables represented by a set of vertices $\mathcal{V} = \{1, 2, \ldots, p\}$ through the use of a set of undirected edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$, and depend on a $n \times p$ data matrix $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p)$, $\boldsymbol{X}_j = (X_{1j}, X_{2j}, \ldots, X_{nj})^T, j = 1, \ldots, p$, where the n row vectors in $\boldsymbol{X}$ are independent and identically distributed according to $N_p(\boldsymbol{0}, \boldsymbol{\Theta}^{-1})$. $\boldsymbol{\Theta}$ contains the scaled partial correlations, where the partial correlations are given by: $\rho_{ij} = -\frac{\theta_{ij}}{\sqrt{\theta_{ii}\theta_{jj}}}$. Therefore, $(i,j) \notin \mathcal{E} \Leftrightarrow \theta_{ij} = 0$.

The spatial autoregressive graphical model is a special type of Gaussian graphical model consisting of both directed and undirected edges. This graphical model captures the complex dependency structure of the data in a parsimonious and well-interpretable manner. To formalise this notion, a few definitions are introduced.

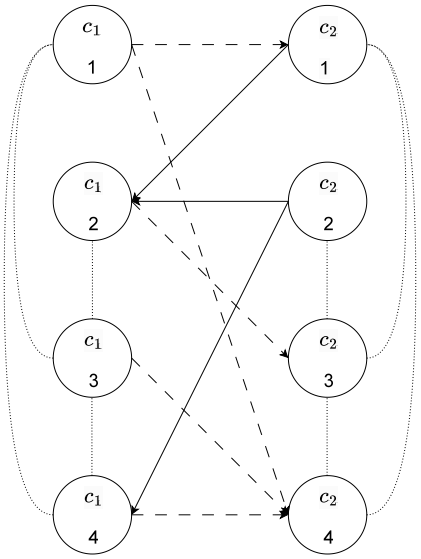


Fig. 2. Spatial chain graph, as derived from $\boldsymbol{\Theta}_E$ (.....), $\boldsymbol{\Psi}_{c_1,c_2}$ (– – – →) and $\boldsymbol{\Psi}_{c_2,c_1}$ (←——) from Eq. (2). The graph represents the data generating process underlying observations across two categories, c_1 and c_2, with four variables.

Definition 2 (*Within-location Effect Graph*)**.** The within-location graph corresponding to $\boldsymbol{X}$ is an undirected graph $\mathcal{G}_w = (\mathcal{V}, \mathcal{E}_w)$ with edge set $\mathcal{E}_w$ such that

$$(i—j) \notin \mathcal{E}_w \Leftrightarrow \theta_{Eij} = \theta_{Eji} = 0, \quad 1 \le i \neq j \le p.$$

Definition 3 (*Between-location Effect Graph*)**.** The between-location effect graph corresponding to $\boldsymbol{X}$ is a directed graph $\mathcal{G}_{\not\perp:c_1,c_2} = (\mathcal{V}_{c_1} \cup \mathcal{V}_{c_2}, \mathcal{E}_{\not\perp:c_1,c_2} \cup \mathcal{E}_{\not\perp:c_2,c_1}) = \mathcal{G}_{\not\perp:c_2,c_1}$, where $c_2 \in \mathcal{N}_{c_1}$ and with edge set $\mathcal{E}_{\not\perp:c_1,c_2} \cup \mathcal{E}_{\not\perp:c_2,c_1}$ such that

$$(i \to j) \notin \mathcal{E}_{\not\perp:c_1,c_2} \Leftrightarrow \psi_{c_1,c_2,ij} = 0, \quad (j \to i) \notin \mathcal{E}_{\not\perp:c_1,c_2} \Leftrightarrow \psi_{c_1,c_2,ji} = 0,$$
$$(i \to j) \notin \mathcal{E}_{\not\perp:c_2,c_1} \Leftrightarrow \psi_{c_2,c_1,ij} = 0, \quad (j \to i) \notin \mathcal{E}_{\not\perp:c_2,c_1} \Leftrightarrow \psi_{c_2,c_1,ji} = 0.$$

Therefore, $\mathcal{E}_{\not\perp:c_1,c_2}$ represents the set of all directed edges from category c_1 to category c_2 and $\mathcal{E}_{\not\perp:c_2,c_1}$ represents the set of all directed edges from category c_2 to category c_1.

From Definition 2 we note that θ_{Eij} is the scaled partial correlation between variables i and j within any location and Definition 3 shows that $\psi_{c_1,c_2,ij}$ is the spatial effect of variable i of category c_1 on the value of variable j of category c_2. Directed edges in the method proposed by Dahlhaus and Eichler (2003) reflect Granger causality. This is, however, not the case for the method that we propose, because almost all spatial orderings are multi-dimensional, and one can move back and forth between locations.

By combining Definitions 2 and 3, a spatial chain graph is obtained that contains the full between- and within-location dependency structure.

Definition 4 (*Spatial Chain Graph*)**.** The spatial chain graph corresponding to $\boldsymbol{X}$ is a partially directed graph $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s) = (\mathcal{V}_{c_1} \cup \mathcal{V}_{c_2}, \mathcal{E}_{\not\perp:c_1,c_2} \cup \mathcal{E}_{\not\perp:c_2,c_1} \cup \mathcal{E}_w)$ with edge set $\mathcal{E}_{\not\perp:c_1,c_2} \cup \mathcal{E}_{\not\perp:c_2,c_1} \cup \mathcal{E}_w$ as defined in Definitions 2 and 3 respectively. Thus, for $c_2 \in \mathcal{N}_{c_1}$, there exists a directed edge from $\mathcal{V}_{c_1,i}$ to $\mathcal{V}_{c_2,j}$ if and only if $(i \to j) \in \mathcal{E}_{\not\perp:c_1,c_2}, 1 \le i \neq j \le p$. Moreover, for any category c_1, there exists an undirected edge between $\mathcal{V}_{c_1,i}$ and $\mathcal{V}_{c_1,j}$ if and only if $(i—j) \in \mathcal{E}_w$.

The spatial chain graph introduced in Definition 4 consists of two identical vertex columns, where each column reflects the p variables for c_1 and c_2 respectively. As an example, consider Eq. (2) pertaining to the toy example illustrated in Fig. 1. Suppose that $p = 4$ and that we have obtained the following parameter estimates

$$\hat{\boldsymbol{\Psi}}_{c_1,c_2} = \begin{pmatrix} \psi_{11} & 0 & 0 & \psi_{14} \\ 0 & 0 & \psi_{23} & 0 \\ 0 & 0 & 0 & \psi_{34} \\ 0 & 0 & 0 & \psi_{44} \end{pmatrix} \quad \hat{\boldsymbol{\Psi}}_{c_2,c_1} = \begin{pmatrix} 0 & \psi_{12} & 0 & 0 \\ 0 & \psi_{22} & 0 & \psi_{24} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \hat{\boldsymbol{\Theta}}_E = \begin{pmatrix} \theta_{11} & 0 & \theta_{13} & \theta_{14} \\ 0 & \theta_{22} & \theta_{23} & 0 \\ \theta_{31} & \theta_{32} & \theta_{33} & \theta_{34} \\ \theta_{41} & 0 & \theta_{43} & \theta_{44} \end{pmatrix}.$$

where we dropped the category subscripts of the matrix elements for improved readability.

To illustrate the properties of $\boldsymbol{\Theta}_E$ and the different $\boldsymbol{\Psi}_k$, consider Fig. 2. One of the conditional independence relations that exist within each plot (prior to being affected by the spatial effects) found in the example is the one where variables 1 and 2 are

conditionally independent. However, between neighbouring locations, there exist direct spatial effects between these variables. By "direct spatial effects", we refer to any effect from one plot to its first-order neighbours, as will be illustrated shortly.

Note that, whilst the graph presented in Definition 4 reduces a complex asymmetric spatial process into a relatively simple figure, the underlying model does not allow for predictions on new locations. However, when exogenous data exist, out of sample prediction is possible. Details are provided in the online supplementary material.

With the spatial chain graph defined, we are now ready to discuss some of its basic properties. First, we illustrate a within-category type of dependency. Stated formally, let $\mathcal{G}_{\mathcal{S}}$ be a spatial chain graph according to Definition 4 corresponding to $\boldsymbol{X}$ with mutually disjoint sets $\mathcal{A}, \mathcal{B}, \mathcal{S} \subset \{1, \dots, p\}$. Then, if, for a given c, $\mathcal{S}$ separates $\mathcal{A}$ and $\mathcal{B}$ in $\mathcal{G}_{\mathcal{S}}$, we have that $\boldsymbol{E}_{\mathcal{A}} \perp\!\!\!\perp \boldsymbol{E}_{\mathcal{B}} | \boldsymbol{E}_{\mathcal{S}}$. This result follows from the global Markov property (Lauritzen, 1996, page 38), which we can use as the spatially filtered data $\boldsymbol{\epsilon}_i$ are independently distributed as $N(\boldsymbol{0}, \boldsymbol{\Theta}_E^{-1}), i = 1, \dots, n$, where information pertaining to within-category relations for any category c can be found in $\boldsymbol{\Theta}_E$. As this property holds for any category in the spatial chain graph, the local and pairwise Markov properties hold as well (Lauritzen, 1996, page 38). This result shows that prior to any spatial effects taking place, or by filtering these spatial effects out, at each location, if in spatial chain graph $\mathcal{G}_{\mathcal{S}}$, two sets of variables are separated by some third, that is, there exists no edge between the first two sets without going through the third, then those two former sets are conditionally independent on the third set at each location. For this result to hold, the separation in the graph should be on the variables for a single category. Or in other words, all three sets of variables are in the same vertex column of $\mathcal{G}_{\mathcal{S}}$. Separating two sets of variables for two different categories does not result in conditional independence relations. This result is useful because it sheds light on underlying relationships at each location, which are otherwise hidden due to the influence of spatial effects on the observations.

In line with the typical spatial autoregressive models that do not exhibit independence between non-first-order neighbours, we show that the proposed method does not either. Let the data generating process be $\text{vec}(\boldsymbol{X}) = \left(\boldsymbol{I}_{np} - \sum_{k=1}^{2} \boldsymbol{\Psi}_k^T \otimes \boldsymbol{W}_k\right)^{-1} \text{vec}(\boldsymbol{E})$. The inverse term is what causes dependencies between a location and its non-first-order neighbours through the residuals as $\left(\boldsymbol{I}_{np} - \sum_{k=1}^{2} \boldsymbol{\Psi}_k^T \otimes \boldsymbol{W}_k\right)^{-1} \text{vec}(\boldsymbol{E}) = \left[\boldsymbol{I}_{np} + \sum_{q=1}^{\infty} \left(\sum_{k=1}^{2} \boldsymbol{\Psi}_k^T \otimes \boldsymbol{W}_k\right)^{q}\right] \text{vec}(\boldsymbol{E})$, dependencies between a location and its non-first-order neighbours exist. Naturally, the first order term $\left(\sum_{k=1}^{2} \boldsymbol{\Psi}_k^T \otimes \boldsymbol{W}_k\right)^{1}$ results in a non-zero matrix, but so does the first series of higher order terms, causing spillovers to non-neighbouring locations of the residual term $\boldsymbol{E}$. However, we have that $\left(\sum_{k=1}^{2} \boldsymbol{\Psi}_k^T \otimes \boldsymbol{W}_k\right)^{q} \to \boldsymbol{O}_{np}$ as $q \to \infty$, with $q \in \mathbb{N}^+$, due to the restriction imposed by Assumption 1. This result implies that the strength of the spatial effect decreases as the neighbour order increases. The spillovers occur only within sequences of adjacent locations, not between sequences of adjacent locations. From an applied perspective, the spillovers make sense: whilst most of the interactions between-locations occur between those that are adjacent to one another, it is impossible to guarantee complete isolation from non-adjacent locations. Moreover, the further away locations are, the less strong the spatial effect.

If both spatial effect matrices are zero-matrices, the following result holds: let $\mathcal{G}_{\mathcal{S}}$ be a spatial chain graph according to Definition 4 corresponding to $\boldsymbol{X}$ with $\mathcal{A} \subset \{1, \dots, p\}$ and $\mathcal{B} \subset \{1, \dots, p\}$, such that $\boldsymbol{X}_{\{c_1\}} \cup \boldsymbol{X}_{\{c_2\}} = \boldsymbol{X}$. Then, if $c_2 \in \mathcal{N}_{c_1}$ and $\boldsymbol{\Psi}_{c_2,c_1} = \boldsymbol{\Psi}_{c_1,c_2} = \boldsymbol{O}_p$ we have that $\boldsymbol{X}_{\{c_1\},\mathcal{A}} \perp\!\!\!\perp \boldsymbol{X}_{\{c_2\},\mathcal{B}}$. This result describes the situation in which the data $\boldsymbol{X}$ consists of two categories: c_1 and c_2. If both the $\boldsymbol{\Psi}_k = \boldsymbol{O}_p$, then any subset of variables in the data pertaining to c_1 is independent of any subset of variables in the data pertaining to c_2. Whilst this presents a special case, if it occurs, no spatial modelling is required, and simpler models can be applied on the data. A proof of this result is provided in the online supplementary material.

2.3. Identifiability

The popular adage "there is no such thing as a free lunch", holds true even in statistics. Given that the proposed model is highly parameterised, identification is a concern. Here, with identifiability we mean that if $\boldsymbol{\Psi} \neq \boldsymbol{\Psi}'$, then $L(\boldsymbol{\Psi}; \boldsymbol{X}) \neq L(\boldsymbol{\Psi}'; \boldsymbol{X})$, for all $\boldsymbol{X}$, and where $L(\cdot)$ denotes the likelihood. The proposed model has an identifiability problem. As a consequence, estimation becomes non-trivial. This is not something that is often considered in the Bayesian setting, as informative priors have the potential to circumvent such identification issues (Gelfand and Sahu, 1999). For the proposed method, we allow for both informative priors and parameter restrictions such that model identification, and in turn Bayesian learning, becomes possible. In the proposed method, we have that the model is identifiable for $\boldsymbol{\Theta}_E$, but not for the various $\boldsymbol{\Psi}_k$, without informative priors or additional parameter restrictions. This is shown in the online supplementary material. To this end, we first introduce a general identifiability condition that uses known or well-informed effects to estimate the unknown effects, followed by special cases of this condition. The general identifiability condition is given by

Condition 1 (*Prior Condition*). *To identify the spatial effects, we require that sufficient knowledge on a subset of the elements of the $\boldsymbol{\Psi}_k$ exists.*

Condition 1 is quite abstract and general. Nevertheless a sufficient condition for identifying the model is when $p^2/2$ elements in each $\boldsymbol{\Psi}_k$ are known. For many scenarios, this is unrealistic, and this condition can be relaxed in the case where knowledge on the spatial effect is not absolute, but instead represents a well-informed opinion. To elaborate on this, if the prior on a (almost) known or well-informed spatial effect is a $N(\mu, \tau)$ prior, μ should reflect existing knowledge and $\tau > 0$ should be small. The more certainty exists on the spatial effect, the smaller τ becomes, until at the limit no estimation is required and the known effect enters the model as a constant. Alternatively, sometimes it is only known whether some spatial effect is positive or negative. This can be reflected by imposing a prior on that spatial effect that restricts it to be either positive or negative respectively. By imposing concentrated priors

on the known or well-informed effects, we can restrict the parameter space, such that only the true parameter value maximises the likelihood. The more restricted the parameter space becomes, the less possible target values the spatial effect can take, until eventually the model is identified by a unique value for the spatial effect.

Whilst fixing $p^2/2$ elements in the $\boldsymbol{\Psi}_k$ results in an identifiable model, when using concentrated priors based on well-informed effects, it is a-priori not known how much knowledge the model requires to be identifiable. It does not suffice to run multiple Markov chains and conduct diagnostic checks, as non-convergence of the Markov chains can hint at identifiability issues, as is done by Eberly and Carlin (2000), but it can also hint at other problems underlying the MCMC algorithm. Therefore, it remains a difficult problem in practice to evaluate how informative the priors need to be in order to identify the parameters of interest. In any case, it is true that imposing more informative, and a greater number of, informative priors is beneficial to parameter estimation, as is noted by Eberly and Carlin (2000). We found that for the proposed model, on simulated data, using normal priors centred around the true value with a standard deviation of less than or equal to 0.25 resolves the identifiability problem.

We note that Condition 1 is a rather typical one within the context of Bayesian statistics whereby knowledge is translated into priors to improve statistical inference. Although in this case, prior knowledge is a requirement for identifiability. No such assumption is needed if the same set of spatial effects is assumed for all locations, that is, by assuming a single $\boldsymbol{\Psi}$, as in Eq. (1). However, this would amount to abandoning the proposed method's unique ability to model asymmetric spatial effects and will therefore not be considered here. Whilst Condition 1 appears to be very strong, the applicability of this assumption is primarily intended for spatial systems that contains many "knowns", such as physiological-based models. Combining the knowledge contained within such models with the proposed method opens up a novel line of research: the analysis of asymmetric spatial relations, whose importance cannot be understated. An example for the combined usage of physiological-based and statistical models can be found in Jahangirlou et al. (2023). In our proposed method, researchers can move beyond the approach specified in Jahangirlou et al. (2023) by directly incorporating the physiological knowledge found in crop growth models by means of concentrated priors, instead of using physiological knowledge as the basis for choosing a statistical method.

As an alternative to Condition 1, we introduce the following special cases

Condition 2 (*Structural Condition*). *To identify the spatial effects, we impose the following restrictions on the various* $\boldsymbol{\Psi}_k$ *(i)* $\psi_{kii} = 0$ *and (ii)* $\boldsymbol{\Psi}_k = \boldsymbol{\Psi}_k^T$ *or (iii) the* $\boldsymbol{\Psi}_k$ *are triangular matrices.*

In the online supplementary material we show numerically that imposing restrictions (i) and (ii) results in identifiable $\boldsymbol{\Psi}_k$. Note that it is possible to mix (ii) and (iii) for different k. The assumption that the $\boldsymbol{\Psi}_k$ are triangular, restriction (iii), seems overly strong. However, when considering our intercropping example, triangular $\boldsymbol{\Psi}_k$ can be expected in the case of one-directional relationships such as the effect of a production variable (e.g. yield) on a growth related variable (e.g. root length), which are typically expected to be zero, whilst the reverse relationship need not be zero. When data contains several of such variables, restriction (iii) appears to be appropriate. Another similar example presents itself in data consisting of production (e.g. yield), environmental (e.g. temperature) and management (e.g. fertiliser) variables. Whilst management and environmental variables might affect production variables, it is known a-priori that the reverse direction contains no effect, warranting the use of restriction (iii).

3. Spatial chain graph learning

A Bayesian framework for the spatial autoregressive model consists of two choices: the prior choice for the autoregressive parameters and the prior choice for the within-location precision matrix. In fact, an additional consideration for the autoregressive parameters needs to be made: whether or not sparsity-inducing priors are imposed. Sparsity – having a majority of zero coefficients – has seen a surge of interest over the last few years, including in the spatial autoregressive literature (Pfarrhofer and Piribauer, 2019), and has various advantages over priors that result in non-sparse estimates, such as selecting only relevant variables in high-dimensional situations ($n \ll p$) for improved predictive accuracy, increased interpretability by forcing certain coefficients to zero and faster parameter estimation. To ensure that the proposed method is generalisable, we allow for both non-sparse and sparse spatial autoregressive priors. For the former. we choose the normal prior (LeSage and Chih, 2018) and for the latter the normal-gamma prior. The normal-gamma prior has shown good performance under conditions of (intermediate) sparsity (Huber and Feldkircher, 2019; Kastner and Huber, 2020).

Even though elaborate prior structures are considered for the residual covariance matrix in the aforementioned articles, they are not applicable to the proposed model, as they do not lend themselves to structure learning of the within-location effect graph, as the graph space remains unexplored. To learn the graph structure of a graphical model, researchers typically impose the conjugate G-Wishart prior on the precision matrix and assume a uniform or truncated Poisson on the graph space (Dawid and Lauritzen, 1993; Roverato, 2002; Mohammadi and Wit, 2015). However, unless a decomposable graph is assumed, which is atypical, this prior requires computationally intensive evaluation of a normalising constant, which hinders generalisability to big or high-dimensional data. For this reason, Wang (2012) introduced the Bayesian graphical lasso; an efficient framework to perform Bayesian inference for the graphical lasso. An improvement of the Bayesian graphical lasso has recently been proposed by Li et al. (2019): the graphical horseshoe prior. This prior has shown excellent performance in graph structure learning, whilst remaining computationally efficient and will be our prior choice on $\boldsymbol{\Theta}_E$. Even though the horseshoe prior is known to be conservative in terms of variable selection (van der Pas et al., 2017), i.e. few true zero parameters in the model are falsely selected, whilst some of the true non-zero parameters are not selected, Li et al. (2019) mitigated the number of false negatives by using the 50% credible interval for variable selection. These priors are implemented in an efficient Gibbs sampling algorithm, that allows researchers to estimate the parameters corresponding

to the spatial chain graph for small, moderate or large p in a reasonable amount of time. We will only illustrate the Gibbs sampler for the normal-gamma prior on the autoregressive parameters and not for the normal prior, as its implementation is trivial. Nonetheless, we will illustrate the performance of all priors in Section 4.

The first step of our Bayesian estimation procedure is the likelihood $L(\boldsymbol{\Psi}, \boldsymbol{\Theta}_E; \boldsymbol{X}) = \prod_{i=1}^{n} P(\boldsymbol{x}_i|\boldsymbol{\Psi}, \boldsymbol{\Theta}_E)$, which under the assumption of multivariate normality of the residuals in (2), is given by

$$\begin{aligned}&(2\pi)^{-\frac{np}{2}} \det(\boldsymbol{\Theta}_E)^{\frac{n}{2}} \det[\boldsymbol{R}(\boldsymbol{\Psi})] \exp\left\{-\frac{1}{2}[\boldsymbol{R}(\boldsymbol{\Psi})\mathrm{vec}(\boldsymbol{X})]^T \left(\boldsymbol{\Theta}_E \otimes \boldsymbol{I}_n\right)[\boldsymbol{R}(\boldsymbol{\Psi})\mathrm{vec}(\boldsymbol{X})]\right\} \\ &\propto \det(\boldsymbol{\Theta}_E)^{\frac{n}{2}} \det[\boldsymbol{R}(\boldsymbol{\Psi})] \exp\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{S}\boldsymbol{\Theta}_E\right)\right], \end{aligned} \tag{4}$$

with $\boldsymbol{S} = \left(\boldsymbol{X} - \sum_{k=1}^{2} \boldsymbol{W}_k \boldsymbol{X} \boldsymbol{\Psi}_k\right)^T \left(\boldsymbol{X} - \sum_{k=1}^{2} \boldsymbol{W}_k \boldsymbol{X} \boldsymbol{\Psi}_k\right)$. A-priori we assume independence of $\boldsymbol{\Theta}_E$ and the $\boldsymbol{\Psi}_k$. Due to the formulation of the likelihood in (4), the block sampling algorithm of Li et al. (2019) is straightforwardly applied on the estimation of $\boldsymbol{\Theta}_E$ when holding $\boldsymbol{\Psi}_1$ and $\boldsymbol{\Psi}_2$ constant. Horseshoe priors are imposed on the upper-triangular elements of the precision matrix (due to the symmetry of $\boldsymbol{\Theta}_E$), and a uniform prior on the diagonal elements. This results in the following element-wise hierarchical structure

$$\begin{aligned} \theta_{Eii} &\propto 1, \\ \theta_{Eij:i<j} &\sim N(0, \lambda_{ij}^2 \xi^2), \\ \lambda_{ij:i<j} &\sim C^+(0,1), \\ \xi &\sim C^+(0,1), \end{aligned}$$

where the global ξ and local λ_{ij} hyperparameters follow a half-Cauchy $C^+(0,1)$ distribution. The posterior on $\boldsymbol{\Theta}_E$ then has the following form

$$P_{\boldsymbol{X}}(\boldsymbol{\Theta}_E|\boldsymbol{\Lambda}, \xi) \propto \det(\boldsymbol{\Theta}_E)^{\frac{n}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{S}\boldsymbol{\Theta}_E\right)\right] \prod_{i<j} \exp\left(-\frac{\theta_{Eij}^2}{2\lambda_{ij}^2 \xi^2}\right) \mathbb{I}_{\boldsymbol{\Theta}_E \in \mathcal{M}_p^+},$$

where $\mathbb{I}$ is an indicator function, $\boldsymbol{\Lambda} = (\lambda_{ij}^2)$ and $\mathcal{M}_p^+$ is the space of $p \times p$ positive definite matrices. Sampling from this posterior is not straightforward, and Li et al. (2019) propose a modified version of the efficient block Gibbs sampler introduced in Wang (2012), whereby one column and row of $\boldsymbol{\Theta}_E$ are updated at a time.

Continuing with the autoregressive parameters, the normal-gamma global–local shrinkage prior leads to the following hierarchical structure

$$\begin{aligned} \psi_{kij}|\alpha_{kij} &\sim N(0, 2\omega^{-2}\alpha_{kij}), \\ \alpha_{kij} &\sim \mathrm{G}(\kappa, \kappa), \\ \omega^2 &\sim \mathrm{G}(b_0, b_1), \end{aligned}$$

for $k \in \{1,2\}$ and $1 \leq i, j \leq p$, with local shrinkage parameter α_{kij}, global shrinkage parameter ω^2 and hyperparameters κ, b_0 and b_1 specified by the researcher. Both the local and global shrinkage parameters follow a Gamma distribution. Due to the hierarchical nature of these priors, the conditional posteriors are similar to those of Huber and Feldkircher (2019) and Pfarrhofer and Piribauer (2019), as the likelihood is not involved, except for the lowest level (that of the ψ_{kij}), for which no known distribution exists in our case given the presence of the $\det[\boldsymbol{R}(\boldsymbol{\Psi})]$ term. Therefore, we use a Metropolis–Hastings (MH) step in the Gibbs sampler to sample from the full conditionals of ψ_{kij}. The conditional posteriors of the shrinkage parameters are

$$\begin{aligned} P(\alpha_{kij}|\psi_{kij}, \omega) &\sim \mathrm{GIG}\left(\kappa - \frac{1}{2}, \psi_{kij}^2, \kappa\omega^2\right), \\ P(\omega^2|\alpha_{111}, \dots, \alpha_{2pp}) &\sim \mathrm{G}\left(b_0 + \kappa 2p^2, b_1 + \frac{\kappa}{2}\sum_{k=1}^{2}\sum_{i=1}^{p}\sum_{j=1}^{p}\alpha_{kij}\right), \end{aligned}$$

such that the local shrinkage parameters follow a generalised inverse Gaussian (GIG) distribution and the global shrinkage parameters follow a Gamma distribution. As stated above, to obtain the ψ_{kij}, we use a MH step within the Gibbs algorithm, where a new value ψ'_{kij} is proposed $\psi'_{kij} \sim N(\psi_{kij}^{(t-1)}, \zeta)$, for $1 \leq t \leq$ number of iterations, with ζ being a tuning parameter to increase efficiency of the MH step. To avoid numerical underflow, we take the natural logarithm of the acceptance rate: $\log(r) = \log\left[P\left(\boldsymbol{X}|\psi'_{1ij}, \boldsymbol{\Theta}_E^{(t)}, \boldsymbol{\Psi}_{1-ij}^{(t-1)}, \boldsymbol{\Psi}_2^{(t-1)}\right)\right] + \log\left[P\left(\psi'_{1ij}\right)\right] - \log\left[P\left(\boldsymbol{X}|\psi_{1ij}^{(t-1)}, \boldsymbol{\Theta}_E^{(t)}, \boldsymbol{\Psi}_{1-ij}^{(t-1)}, \boldsymbol{\Psi}_2^{(t-1)}\right)\right] - \log\left[P\left(\psi_{1ij}^{(t-1)}\right)\right]$. Then, sample $u \sim U(0,1)$, and if $\log(u) \leq \min\{0, \log(r)\}$ we set $\psi_{kij}^{(t)} = \psi'_{kij}$, and set $\psi_{kij}^{(t)} = \psi_{kij}^{(t-1)}$ otherwise. The hyperparameters in the acceptance step are omitted and we set $k = 1$ to simplify the notation. Note that due to the separable nature of the priors on the individual spatial effect components, combined with the restrictions imposed by Conditions 1 and 2, we end up with $2p^2$ MH steps (parameters) for Condition 1 and $p(p+1)$ and $2p^2$ steps for restrictions (ii) and (iii) of Condition 2 respectively within each iteration of the Gibbs sampling procedure. To guarantee the stability condition imposed on the spatial effects, see Assumption 1, we redraw values for ψ'_{kij} until the condition is met. The variable selection for the proposed method follows that of Li et al. (2019), where, if the 50% posterior credible interval for any element does not contain zero, that element is selected. Details on the full Bayesian estimation procedure are provided in the online supplementary material.

4. Simulation study

To illustrate the performance of the proposed methodology on data of different dimensionality and network structure, a simulation study was conducted. Whilst, to the best of our knowledge, alternative models do not exist, the relevance of this study is found in its capacity to shed light on the adequacy of Bayesian learning of the true parameters underlying the proposed method.

The three network structures that underlie $\boldsymbol{\Theta}_E$ for the simulation study are random, scale-free and star. Whilst our main interest is illustrating to what extend the Gibbs method can estimate the $\boldsymbol{\Psi}_k$, multiple networks underlying $\boldsymbol{\Theta}_E$ were chosen to represent various types of within-location dependence relations, thereby illustrating the versatility of the proposed method.

The data is simulated by first constructing $\boldsymbol{\Theta}_E$ according to the chosen network type, with an edge probability of 0.2 in the case of the random network, followed by independently drawing the residual vectors $\boldsymbol{\epsilon}_i \sim N_p(\mathbf{0}, \boldsymbol{\Theta}_E^{-1})$ for $i = 1, \dots, n$. Under assumptions of sparsity, i.e. the simulations where normal-gamma priors are used, for each $k \in \{1, 2\}$ we draw $\lfloor 3p(p-1)/8 \rfloor$ elements from a $U(-1, 1)$ distribution, and randomly assign these to spatial effects ψ_{kij}, whilst setting all other $\lceil p(p-1)/8 \rceil$ elements to zero. These elements are taken from the triangular $\boldsymbol{\Psi}_k$, which are made symmetric afterwards, if the restriction requires this (see Condition 2). When normal priors are used, we draw spatial effects $\psi_{kij} \sim U(-1, 1)$ for all k, i, j. For both priors, the spatial effects are subsequently transformed to adhere to restriction (i) by setting $\psi_{kii} = 0$ as well as either restriction (ii) or restriction (iii) from Condition 2. Finally, we set $\text{vec}(\boldsymbol{X}) = (\boldsymbol{I}_{np} - \sum_{k=1}^{2} \boldsymbol{\Psi}_k^T \otimes \boldsymbol{W}_k)^{-1} \text{vec}(\boldsymbol{E})$, where the $\boldsymbol{W}_k$ are equivalent to those found in (3), when generalised to larger values of n.

For each network structure, 20 spatial autoregressive graphical models are fitted for data consisting of two categories, with $n \in \{25, 50, 100\}$ and $p \in \{4, 8, 20\}$. We showcase model performance using restrictions (i), (ii) and (iii) of Condition 2, where we set all $\boldsymbol{\Psi}_k$ to be upper triangular for models adhering to restriction (iii). We do not evaluate the performance of the method under Condition 1, as it is a generalisation of restriction (iii) found in Condition 2. The dimensionality of simulated data is consistent with the data commonly found in intercropping designs ($p = 4$ and $p = 8$). Nevertheless, the proposed method can handle high(er)-dimensional data, as shown by the simulations with $p = 20$. For each parameter combination, we use the first 1000 iterations as burn-in from a total of 2000 iterations. Inference is based on the 1000 posterior draws for all parameters, where we use the posterior mean as a point estimate to compute the discrepancy measures.

The discrepancy between the true and the estimated parameters are measured using the Frobenius norm: $\|\boldsymbol{\Theta}_E^0 - \hat{\boldsymbol{\Theta}}_E\|_F$ for $\boldsymbol{\Theta}_E$, and the F_1 score: $2\frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$ and the root mean square error: $\sqrt{\frac{1}{2p^2} \sum_{k=1}^{2} \sum_{i=1}^{p} \sum_{j=1}^{p} (\psi_{kij}^0 - \hat{\psi}_{kij})^2}$ for the $\boldsymbol{\Psi}_k$, where the discrepancy measures are averaged over the 20 fitted models. For both the Frobenius norm in $[0, \infty)$ and the root mean square error in $[0, \infty)$, lower values are indicative of better model performance, whilst the opposite holds true for the F_1 score in $[0, 1]$. The F_1 score is computed using the same edge selection method found in Li et al. (2019). That is, if the 50% posterior credible interval for any element in the $\boldsymbol{\Psi}_k$ does not contain zero, that element is considered a discovery, and vice versa. The results for the simulation studies can be found in Table 1 where normal priors are implemented with hyperparameters $\mu = 0, \sigma = 1$ and normal-gamma priors are implemented with hyperparameters $b_0 = b_1 = 0.01$ and $\kappa = 0.1$ (Huber and Feldkircher, 2019; Pfarrhofer and Piribauer, 2019). Convergence diagnostics of the simulations are shown in the online supplementary material. In addition, the Supplementary data contain an evaluation of some model-misspecified simulations, simulations of the MSAR model introduced by Eq. (1) when asymmetry is present in the data generating mechanism, and the computation time of the proposed method.

As observed in Table 1, for all network types, prior assumptions, identifiability restrictions and number of parameters, the discrepancy measures for posterior mean point estimates reduce as the sample size of the data grows. For low p, the model is able to provide accurate point estimates. However, as p increases, even for higher values of n, the discrepancy between the point estimate and the true value is still substantial. However, due to the high number of parameters that need to be estimated, this is no surprise. Whilst the estimation accuracy of the method appears invariant to the underlying network structure, it is not invariant to the identifiability restriction. This result is not particularly remarkable, considering that under the triangular restriction, the lower triangular estimates of the $\boldsymbol{\Psi}_k$ contribute next to nothing to the discrepancy measures, whilst under a symmetric restriction these elements contribute equally to the discrepancy measures as the upper triangular elements do. Finally, both priors show satisfactory and similar performance.

5. An application on an intercropping system

We have primarily motivated the present model through an intercropping framework, begging a demonstration on actual intercropping data. The included example consists of a paired-crop type design of Belgian endive and beetroot, where each plot has a single neighbouring plot of the other species. The categories in the proposed method are therefore given by the two different crops. Belgian endive can be found to the right of beetroot four times and five times to the left, resulting in a total of 18 observations. The design is shown in Fig. 3, which also contains the variable values for the plots. The variables in this dataset are dried plant weight, fresh plant weight, (aboveground) plant height and number of plants, where the first three variables are averaged across all plants within a plot. Instead of including two different weight variables, we use the fresh plant weight to construct the plant root length based on the average ratio of root weight/total weight for the crops. The different locations represent plots containing multiple plants of a single crop. Given the presence of borders between each pair of neighbouring plots, no spatial spillovers are expected to occur between non-neighbouring plots. Therefore, each pair of Belgian endive and beetroot is independent of all other pairs. The data used in this study was generated through the strategic investment theme 'Biodiversity-positive Food Systems' of Wageningen University & Research (2022).

Table 1
Simulation results for the random, scale-free and star networks, using both priors for both symmetric and triangular parameter restrictions. FN stands for Frobenius norm, F_1 stands for F_1 score and RMSE stands for root mean square error. The discrepancy measures are averaged across 20 fitted models for each parameter combination and rounded to 2 decimals. Standard errors are provided between parentheses.

	Normal prior on Ψ				Normal-gamma prior on Ψ					
Random network	Symmetric restriction		Triangular restriction		Symmetric restriction			Triangular restriction		
n, p	FN	RMSE	FN	RMSE	FN	F_1	RMSE	FN	F_1	RMSE
25, 4	2.60 (0.62)	0.28 (0.03)	2.51 (0.57)	0.22 (0.02)	2.06 (0.38)	0.74 (0.03)	0.36 (0.04)	2.32 (0.53)	0.78 (0.01)	0.25 (0.02)
50, 4	0.89 (0.12)	0.18 (0.02)	0.75 (0.06)	0.15 (0.01)	0.91 (0.16)	0.77 (0.02)	0.25 (0.03)	0.97 (0.16)	0.79 (0.01)	0.16 (0.01)
100, 4	0.89 (0.17)	0.15 (0.02)	0.70 (0.10)	0.11 (0.01)	1.14 (0.57)	0.78 (0.01)	0.18 (0.02)	0.66 (0.12)	0.80 (0.01)	0.12 (0.01)
25, 8	9.03 (1.21)	0.31 (0.02)	11.17 (3.80)	0.27 (0.01)	20.88 (7.23)	0.79 (0.01)	0.37 (0.02)	11.99 (3.29)	0.82 (0.01)	0.35 (0.02)
50, 8	6.34 (1.91)	0.23 (0.02)	3.35 (1.00)	0.19 (0.01)	12.42 (5.70)	0.81 (0.01)	0.25 (0.02)	4.12 (0.89)	0.83 (0.01)	0.22 (0.01)
100, 8	4.10 (1.12)	0.21 (0.02)	2.29 (1.01)	0.13 (0.00)	8.86 (5.24)	0.83 (0.01)	0.14 (0.01)	1.86 (0.55)	0.84 (0.00)	0.13 (0.01)
25, 20	659.08 (228.26)	0.40 (0.01)	1081.98 (124.89)	0.43 (0.01)	179.81 (63.65)	0.80 (0.00)	0.48 (0.02)	214.65 (78.69)	0.84 (0.00)	0.49 (0.02)
50, 20	265.35 (103.37)	0.31 (0.01)	747.85 (86.70)	0.36 (0.01)	89.12 (41.37)	0.81 (0.01)	0.35 (0.01)	154.50 (52.56)	0.85 (0.00)	0.37 (0.02)
100, 20	115.76 (68.14)	0.240 (0.01)	356.32 (41.19)	0.26 (0.01)	36.55 (18.79)	0.83 (0.00)	0.23 (0.01)	64.72 (21.43)	0.86 (0.00)	0.23 (0.01)
Scale-free network										
25, 4	2.38 (0.34)	0.28 (0.03)	2.21 (0.39)	0.23 (0.01)	3.33 (0.76)	0.77 (0.02)	0.35 (0.03)	4.05 (1.39)	0.77 (0.02)	0.23 (0.01)
50, 4	1.26 (0.21)	0.23 (0.02)	1.09 (0.13)	0.16 (0.01)	1.69 (0.46)	0.78 (0.02)	0.23 (0.03)	1.70 (0.42)	0.79 (0.01)	0.19 (0.01)
100, 4	1.06 (0.20)	0.15 (0.02)	0.74 (0.10)	0.12 (0.01)	1.66 (0.84)	0.79 (0.02)	0.16 (0.02)	0.95 (0.12)	0.78 (0.01)	0.12 (0.01)
25, 8	10.89 (1.85)	0.31 (0.02)	12.34 (2.84)	0.29 (0.01)	10.37 (2.04)	0.80 (0.01)	0.40 (0.03)	11.47 (3.17)	0.82 (0.01)	0.38 (0.02)
50, 8	6.02 (1.45)	0.25 (0.02)	5.03 (1.50)	0.19 (0.01)	6.40 (1.79)	0.81 (0.01)	0.25 (0.02)	4.96 (0.93)	0.84 (0.00)	0.23 (0.01)
100, 8	4.90 (1.97)	0.19 (0.02)	1.61 (0.32)	0.13 (0.00)	3.66 (0.83)	0.83 (0.01)	0.18 (0.02)	2.11 (0.47)	0.85 (0.00)	0.15 (0.01)
25, 20	1081.72 (446.70)	0.40 (0.01)	1390.24 (188.14)	0.46 (0.01)	127.72 (24.95)	0.79 (0.00)	0.49 (0.02)	113.56 (0.00)	0.83 (0.01)	0.45 (0.02)
50, 20	682.91 (311.92)	0.34 (0.01)	733.92 (106.53)	0.35 (0.01)	74.38 (12.09)	0.81 (0.00)	0.40 (0.02)	69.45 (11.96)	0.84 (0.00)	0.31 (0.01)
100, 20	282.84 (186.90)	0.23 (0.01)	412.56 (46.80)	0.28 (0.01)	51.85 (8.17)	0.82 (0.00)	0.26 (0.01)	43.90 (7.64)	0.85 (0.00)	0.22 (0.01)
Star network										
25, 4	2.36 (0.23)	0.30 (0.03)	2.98 (0.67)	0.22 (0.01)	3.67 (0.79)	0.74 (0.02)	0.31 (0.04)	3.05 (0.58)	0.78 (0.02)	0.26 (0.02)
50, 4	1.24 (0.20)	0.20 (0.03)	1.28 (0.24)	0.15 (0.01)	1.94 (0.58)	0.78 (0.01)	0.20 (0.02)	1.39 (0.25)	0.79 (0.01)	0.18 (0.01)
100, 4	0.95 (0.11)	0.15 (0.01)	0.98 (0.14)	0.11 (0.01)	1.68 (0.85)	0.78 (0.02)	0.17 (0.02)	0.85 (0.09)	0.80 (0.01)	0.11 (0.01)
25, 8	12.01 (3.26)	0.31 (0.02)	13.52 (3.64)	0.29 (0.01)	14.22 (3.49)	0.79 (0.01)	0.43 (0.03)	15.80 (4.64)	0.83 (0.01)	0.34 (0.02)
50, 8	5.70 (1.44)	0.23 (0.01)	4.95 (1.45)	0.19 (0.01)	7.06 (2.11)	0.82 (0.01)	0.25 (0.02)	9.57 (5.04)	0.84 (0.00)	0.21 (0.01)
100, 8	3.39 (0.66)	0.19 (0.02)	2.70 (0.76)	0.13 (0.00)	6.73 (2.48)	0.83 (0.01)	0.20 (0.02)	5.86 (2.41)	0.85 (0.00)	0.17 (0.01)
25, 20	596.11 (250.65)	0.40 (0.01)	1729.98 (260.41)	0.44 (0.01)	127.62 (30.98)	0.80 (0.00)	0.49 (0.02)	216.45 (50.60)	0.85 (0.00)	0.51 (0.02)
50, 20	243.76 (113.58)	0.27 (0.02)	621.94 (174.77)	0.32 (0.01)	85.40 (21.56)	0.81 (0.00)	0.37 (0.01)	113.69 (27.83)	0.86 (0.01)	0.35 (0.01)
100, 20	144.80 (55.62)	0.23 (0.01)	299.62 (84.37)	0.25 (0.01)	40.13 (10.12)	0.82 (0.01)	0.21 (0.01)	67.49 (18.26)	0.87 (0.01)	0.18 (0.01)

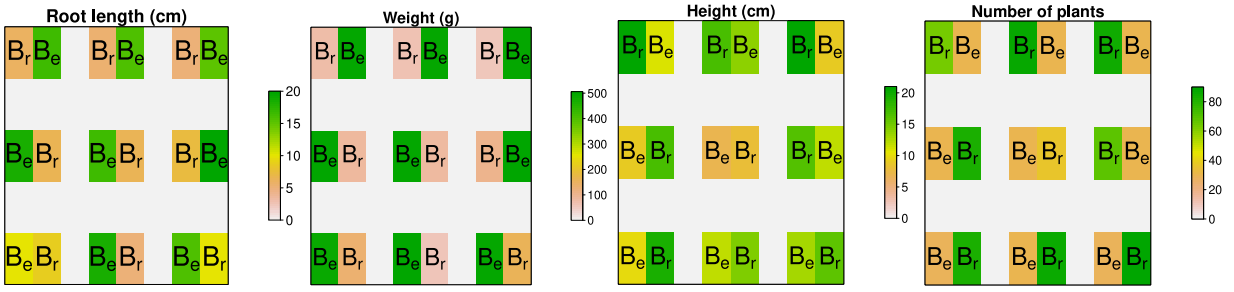


Fig. 3. Paired crop intercropping design with two different crops: Belgian endive (B_e) and beetroot (B_r). Each (coloured) rectangle represents a plot containing multiple plants of the same crop. Variable values for the root length, weight, height and number of plants are illustrated by colours. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Whilst the plots containing Belgian endive typically contain more biomass in terms of weight than those containing beetroot, the number of plants per plot, as well as the average plant height tend to be higher for beetroot.

We use the following equation to represent the stochastic model underlying the observed data

$$\boldsymbol{X} = \boldsymbol{W}_{B_r,B_e}\boldsymbol{X}\boldsymbol{\Psi}_{B_r,B_e} + \boldsymbol{W}_{B_e,B_r}\boldsymbol{X}\boldsymbol{\Psi}_{B_e,B_r} + \boldsymbol{E}, \quad \boldsymbol{\varepsilon}_i \sim N_p(\boldsymbol{0}, \boldsymbol{\Theta}_E^{-1}), \quad i = 1, \ldots, 18,$$

where $\boldsymbol{W}_{B_r,B_e}, \boldsymbol{W}_{B_e,B_r} \in [0,1]^{18\times 18}$ are of the following form

$$\boldsymbol{W}_{B_r,B_e} = \begin{pmatrix} 0 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 0 & & \vdots & \vdots \\ 0 & 0 & 1 & 0 & \ddots & & \\ \vdots & \vdots & & \ddots & \ddots & & \\ 0 & 0 & \ldots & & & 0 & 0 \\ 0 & 0 & \ldots & & & 1 & 0 \end{pmatrix} \quad \boldsymbol{W}_{B_e,B_r} = \begin{pmatrix} 0 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 1 & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ddots & & \\ \vdots & \vdots & & \ddots & \ddots & & \\ 0 & 0 & \ldots & & & 0 & 1 \\ 0 & 0 & \ldots & & & 0 & 0 \end{pmatrix}.$$

In order to adhere to the identifiability restrictions of Condition 1, this application makes use of informative normal priors for the spatial effects, whereby knowledge obtained from research is translated into priors. Based on Czaban et al. (2023), who evaluated an intercropping system consisting of the same two crops as the present analysis, we expect positive effects of a crop's root length on the root length, weight and height of neighbouring crops, due to the enhanced nutrient uptake by longer roots, which might be transported to neighbouring crops. Conversely, we expect that the number of plants has a negative effect on all variables of the other

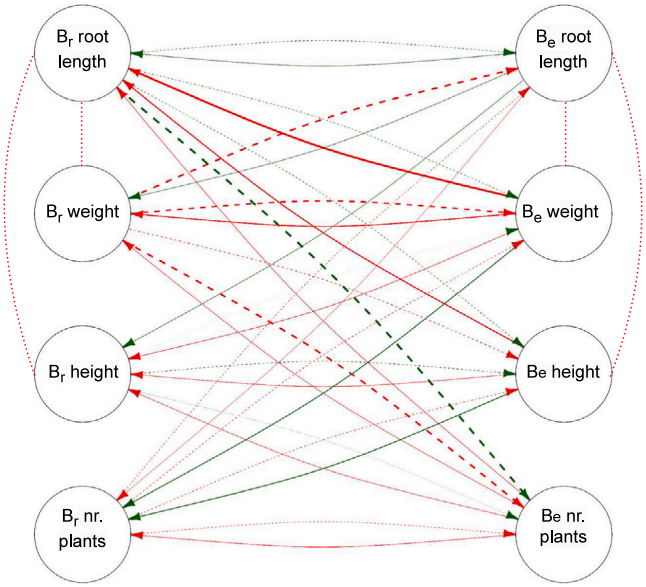


Fig. 4. Spatial chain graph, based on the posterior means from $\boldsymbol{\Theta}_E$ (.....), $\boldsymbol{\Psi}_{B_r,B_e}$ (– – – →) and $\boldsymbol{\Psi}_{B_e,B_r}$ (←——), with positive (——) and negative effects (——), where the edge width is indicative of the effect size. The graph represents the data generating process underlying observations across two different crops, beetroot (B_r) and Belgian endive (B_e), with four variables.

crop due to the arising competition effects (Ren et al., 2016). The last informative priors that are imposed are normal priors with a negative mean relating to the spatial effect of Belgian endive height and weight on beetroot height (Coutinho et al., 2017). This results in a total of 16 known effects, leaving 16 spatial effects on which normal-gamma priors are imposed. The informative priors have means of 1 and −1 for respectively positive and negative effects, with a standard deviation of 0.01, reflecting our strong belief in non-negligible effect sizes for the relations mentioned in the literature. As an aside, we should state that using a single precision matrix to model the within-crop effects is not too strong an assumption for this application. The elements of the precision matrix reflect the effects that occur within the plants, i.e. trait to trait effects, prior to the spatial or competition effects from neighbouring plots containing different plants. As such these are typical allometric relationships, where, up to a scaling factor, no strong difference between the crops is expected (Niklas, 2004; Westgeest et al., 2024). In practice, when observing two different crops, the data is standardised, thereby reducing the impact of scale differences between the crops. We run the Gibbs sampling algorithm with a burnin of 50 000, for 100 000 total iterations. The estimated spatial chain graph is shown in Fig. 4, whilst diagnostic plots for a subset of the within- and between-location effects are given in Fig. 5. Note that, prior to our analysis of the results, we run the Henze–Zirkler test for multivariate normality (Henze and Zirkler, 1990), using the posterior means of $\boldsymbol{\Psi}_{B_r,B_e}$ and $\boldsymbol{\Psi}_{B_e,B_r}$ to confirm that the filtered residuals $\boldsymbol{E}$ are indeed multivariate normally distributed (test statistic $W \approx 0.70$ and $p > 0.1$).

Even though normal-gamma priors were imposed on the unknown spatial effects, the estimated spatial chain graph is very dense, see Fig. 4, most notably for the between-category effects for which only two effects were estimated to be 0. Conversely, the within-category effects exhibit a substantial amount of sparsity. In addition, even though the informative priors were by and large the same for the two spatial effect matrices, the estimated effects illustrate the need for asymmetry. Moving on to the interpretation of the spatial chain graph, provided that the informative priors accurately reflect the underlying model, unusual negative within-category relationships between root length and weight, as well as between root length and height, are observed. A potential explanation for these negative effects is that longer roots do not necessarily lead to an increased nutrient uptake, or a nutrient uptake that encourages plant growth, as plants already acquire more nutrients evidenced from the (imposed) positive relationship between a plant's root length and the weight/height of neighbouring crops. Therefore, whenever a plant grows longer roots, which requires resources, valuable nutrients are wasted that could have gone into additional plant growth. A second within-category effect of note is the conditional independence between the number of plants and the other variables, showing that within plots, increasing the planting density has no negative effects on plant growth, despite the increased competition between plants for resources. Whereas the between-category effects of root length on height and weight were more or less fixed by informative priors, the effect of a crop's root length on a neighbouring crop's number of plants was not. This effect was positive for the root length of beetroot on the number of Belgian endive plants, but negative for the root length of Belgian endive on the number of beetroot plants.

In order to gain more insight into the exact nature of these synergistic and competitive effects, a second, more thorough analysis is needed, consisting of data with more variables that influence plant production, such as biological characteristics of the plant, as well as soil properties and management techniques to advance intercropping research.

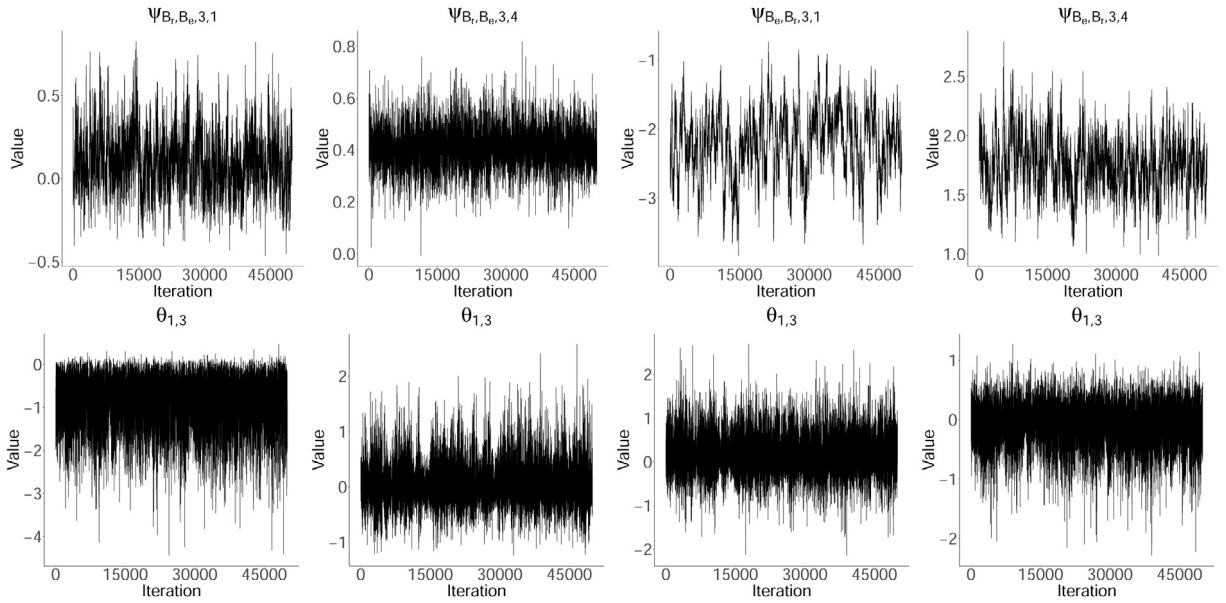


Fig. 5. MCMC trace plots. The notation $\psi_{B_r,B_e,i,j}$ found at the figure titles reflects the spatial relations from variable i of beetroot on variable j of Belgian endive, whereas $\theta_{i,j}$ reflects the relation between variables i and j within a given location. The trace plots contain 50 000 draws from the posterior distributions of the within- and between-location effects respectively.

6. Conclusion and discussion

In this contribution, we propose a new statistical methodology that is able to infer multivariate symmetric within-location and asymmetric between-location spatial effects. In essence, the proposed model can be seen as a fusion, and extension, of two different models: the multivariate spatial autoregressive model and the Gaussian graphical model. By exploiting the flexibility offered by the spatial weight matrices, asymmetry in spatial effects is accounted for. As this results in a highly parameterised model, identifiability restrictions are imposed on the spatial effect matrices. The inferred local and spatial effects are represented by means of a graphical framework, facilitating interpretability. Extensive simulations for a variety of network structures demonstrate that the proposed method is able to effectively reconstruct both the within- and between-location effects.

A downside of the method is the restrictiveness of the identifiability conditions. These conditions rely on prior knowledge of the (potential absence of) spatial relations underlying the phenomenon of interest. The question that arises is that if such prior knowledge exists, why use this method to begin with? The answer is that the proposed method provides numerical values and uncertainty pertaining to the values of these effects, which goes above and beyond typical prior knowledge. In addition, the method only requires knowledge on a subset of the spatial processes underlying any observed phenomenon. The incorporation of prior knowledge in statistical methods can lead to discoveries that are otherwise unattainable, i.e. the full spatial dependency process, and is part of the Bayesian philosophy of statistical inference. Moreover, researchers are not necessarily restricted to the set of identifiability restrictions proposed in this article. There exist other possible identification restrictions, which might be more suited to particular research questions or settings. On a final note, the proposed methodology challenges practitioners to translate knowledge into informative priors. An example whereby existing knowledge and statistics can be combined with statistics constitutes the usage of physiological-based – such as crop growth models – from which part of the known physiological relations can be translated into priors and applied on a more complex spatial scenario with asymmetric relations, using the proposed method.

One recommendation for future research is to allow for a different precision matrix for each category. In some cases the assumption of stable within-location effects might be overly restrictive. On the flip side, removing such a restriction would greatly increase the computation time of the method, making it unattractive for datasets consisting of more observations and variables. In addition, it would result in additional identifiability issues. Related to this is the absence of a left–right distinction in the spatial effects, as illustrated by the application. Such a distinction could be prove to be especially valuable for statistical analyses of intercropping systems, as the influence of weather-related variables on crops can change spatial effects, depending on the whether crops are located left or right of their neighbours, and therefore merits additional investigation. Whilst this article focused on data consisting of two categories, in practice, more than two categories may be present in the data. Even though the proposed method cannot handle such data without yet another set of identifiability constraints, there might exist practical relevance in such an extension. Another avenue for future research is in methods that speed up computation of the proposed method, such that data consisting of many observations and variables can be accommodated for. Computing determinants of large matrices in every iteration of the Gibbs algorithm is computationally expensive, or even prohibitive for large data. Whilst (approximate) methods have been developed for symmetric matrices, heterogeneous spatial effects, or multivariate spatial effects, no existing method substantially

reduced the computation time of determinants of the form introduced by the method proposed in this article. A final recommendation is to extend the proposed method with a temporal component, as many complex processes evolve over time, which might also result in different sets of within- and between-location effects at each time point.

Acknowledgements

We would like to thank the editor Alfred Stein and the anonymous referees for their insightful comments which helped improve our paper. In addition, we thank Anna Edlinger, Jan Peter van der Hoeve, Peter Bourke and Piter Bijma (Wageningen University & Research) for sharing the data. Generation of the data was supported by the Dutch Ministry of Agriculture, Nature and Food Quality (KB44-001-001).

Appendix A. Supplementary data

The supplementary material contains the proofs and derivations referenced in Section 2, out of sample prediction, identifiability results, the algorithm of the Bayesian estimation framework, an illustration of convergence of the Gibbs algorithm, additional simulation results and an evaluation of the computation time. The R package SAGM is available at https://CRAN.R-project.org/package=SAGM.

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spasta.2025.100893.

References

Abegaz, F., Wit, E., 2013. Sparse time series chain graphical models for reconstructing genetic networks. Biostatistics 14 (3), 586–599.
Barigozzi, M., Brownlees, C., 2019. Nets: Network estimation for time series. J. Appl. Econometrics 34 (3), 347–364.
Besag, J., York, J., Mollié, A., 1991. Bayesian image restoration, with two applications in spatial statistics. Ann. Inst. Statist. Math. 43, 1–20.
Bourke, P.M., Evers, J.B., Bijma, P., van Apeldoorn, D.F., Smulders, M.J., Kuyper, T.W., Mommer, L., Bonnema, G., 2021. Breeding beyond monoculture: putting the "intercrop" into crops. Front. Plant Sci. 12, 734167.
Brooker, R.W., Bennett, A.E., Cong, W.-F., Daniell, T.J., George, T.S., Hallett, P.D., Hawes, C., Iannetta, P.P., Jones, H.G., Karley, A.J., et al., 2015. Improving intercropping: a synthesis of research in agronomy, plant physiology and ecology. New Phytol. 206 (1), 107–117.
Chan, J.C., 2020. Large Bayesian VARs: A flexible Kronecker error covariance structure. J. Bus. Econom. Statist. 38 (1), 68–79.
Coutinho, P.W.R., Oliveira, P.S.R.d., Echer, M.d.M., Cadorin, D.A., Vanelli, J., 2017. Establishment of intercropping of beet and chicory depending on soil management1. Rev. Ciência Agron. 48, 674–682.
Czaban, W., Han, E., Lund, O.S., Stokholm, M.S., Jensen, S.M., Thorup-Kristensen, K., 2023. The enhancing effect of intercropping sugar beet with chicory on the deep root growth and nutrient uptake. Agric. Ecosyst. Environ. 347, 108360.
Dahlhaus, R., Eichler, M., 2003. Causality and graphical models in time series analysis. Oxford Statist. Sci. Ser. 115–137.
Dallakyan, A., Kim, R., Pourahmadi, M., 2022. Time series graphical lasso and sparse VAR estimation. Comput. Statist. Data Anal. 176, 107557.
Dawid, A.P., Lauritzen, S.L., 1993. Hyper Markov laws in the statistical analysis of decomposable graphical models. Ann. Statist. 1272–1317.
Dobra, A., 2016. Graphical modeling of spatial health data. In: Handbook of Spatial Epidemiology. Chapman and Hall/CRC, pp. 593–612.
Dobra, A., Lenkoski, A., Rodriguez, A., 2011. Bayesian inference for general Gaussian graphical models with application to multivariate lattice data. J. Amer. Statist. Assoc. 106 (496), 1418–1433.
Eberly, L.E., Carlin, B.P., 2000. Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. Stat. Med. 19 (17–18), 2279–2294.
Federer, W.T., 2012. Statistical Design and Analysis for Intercropping Experiments: Volume 1: Two Crops. Springer Science & Business Media.
Ferrari, M.J., Djibo, A., Grais, R.F., Bharti, N., Grenfell, B.T., Bjornstad, O.N., 2010. Rural–urban gradient in seasonal forcing of measles transmission in Niger. Proc. R. Soc. B: Biol. Sci. 277 (1695), 2775–2782.
Gao, D., Pan, X., Khashi u Rahman, M., Zhou, X., Wu, F., 2021. Common mycorrhizal networks benefit to the asymmetric interspecific facilitation via K exchange in an agricultural intercropping system. Biol. Fertil. Soils 57 (7), 959–971.
Gelfand, A.E., Sahu, S.K., 1999. Identifiability, improper priors, and Gibbs sampling for generalized linear models. J. Amer. Statist. Assoc. 94 (445), 247–253.
Getis, A., 1990. Screening for spatial dependence in regression analysis. In: Papers of the Regional Science Association. Vol. 69, Springer, pp. 69–81.
Goetzke, F., 2008. Network effects in public transit use: evidence from a spatially autoregressive mode choice model for New York. Urban Stud. 45 (2), 407–417.
Griffin, T.W., 2010. The spatial analysis of yield data. Geostat. Appl. Precis. Agric. 89–116.
Henze, N., Zirkler, B., 1990. A class of invariant consistent tests for multivariate normality. Comm. Statist. Theory Methods 19 (10), 3595–3617.
Huang, C., Liu, Q., Gou, F., Li, X., Zhang, C., van der Werf, W., Zhang, F., 2017. Plant growth patterns in a tripartite strip relay intercrop are shaped by asymmetric aboveground competition. Field Crop. Res. 201, 41–51.
Huber, F., Feldkircher, M., 2019. Adaptive shrinkage in Bayesian vector autoregressive models. J. Bus. Econom. Statist. 37 (1), 27–39.
Jahangirlou, M.R., Morel, J., Akbari, G.A., Alahdadi, I., Soufizadeh, S., Parsons, D., 2023. Combined use of APSIM and logistic regression models to predict the quality characteristics of maize grain. Eur. J. Agron. 142, 126629.
Kastner, G., Huber, F., 2020. Sparse Bayesian vector autoregressions in huge dimensions. J. Forecast. 39 (7), 1142–1165.
Keita, S., 2020. Air passenger mobility, travel restrictions, and the transmission of the covid-19 pandemic between countries. Covid Econ. 9, 77–96.
Kelejian, H.H., Prucha, I.R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. J. Econometrics 157 (1), 53–67.
Lauritzen, S.L., 1996. Graphical Models, vol. 17, Clarendon Press.
Leroux, B.G., Lei, X., Breslow, N., 2000. Estimation of disease rates in small areas: a new mixed model for spatial dependence. In: Statistical Models in Epidemiology, the Environment, and Clinical Trials. Springer, pp. 179–191.
LeSage, J.P., Chih, Y.-Y., 2018. A Bayesian spatial panel model with heterogeneous coefficients. Reg. Sci. Urban Econ. 72, 58–73.
Li, H., Calder, C.A., Cressie, N., 2007. Beyond Moran's I: testing for spatial dependence based on the spatial autoregressive model. Geogr. Anal. 39 (4), 357–375.
Li, Y., Craig, B.A., Bhadra, A., 2019. The graphical horseshoe estimator for inverse covariance matrices. J. Comput. Graph. Statist. 28 (3), 747–757.
Liang, Y., 2019. Graph-based multivariate conditional autoregressive models. Stat. Theory Relat. Fields 3 (2), 158–169.
Long, D.S., 1998. Spatial autoregression modeling of site-specific wheat yield. Geoderma 85 (2–3), 181–197.
Millo, G., 2014. Maximum likelihood estimation of spatially and serially correlated panels with random effects. Comput. Statist. Data Anal. 71, 914–933.
Mohammadi, A., Wit, E., 2015. Bayesian structure learning in sparse Gaussian graphical models. Bayesian Anal. 10 (1), 109–138.
Niklas, K.J., 2004. Plant allometry: is there a grand unifying theory? Biol. Rev. 79 (4), 871–889.

Ord, K., 1975. Estimation methods for models of spatial interaction. J. Amer. Statist. Assoc. 70 (349), 120–126.
Paci, L., Consonni, G., 2020. Structural learning of contemporaneous dependencies in graphical VAR models. Comput. Statist. Data Anal. 144, 106880.
Pfarrhofer, M., Piribauer, P., 2019. Flexible shrinkage in high-dimensional Bayesian spatial autoregressive models. Spat. Stat. 29, 109–128.
Ren, Y., Liu, J., Wang, Z., Zhang, S., 2016. Planting density and sowing proportions of maize–soybean intercrops affected competitive interactions and water-use efficiencies on the Loess Plateau, China. Eur. J. Agron. 72, 70–79.
Roverato, A., 2002. Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. Scand. J. Stat. 29 (3), 391–411.
van der Pas, S., Szabó, B., van der Vaart, A., 2017. Uncertainty quantification for the horseshoe (with discussion). Bayesian Anal. 12 (4), 1221–1274.
Wageningen University & Research, 2022. Biodiversity-positive food systems. www.wur.eu/bpfs, Last accessed on 2024-01-30.
Wang, H., 2012. Bayesian graphical lasso models and efficient posterior computation. Bayesian Anal. 7 (4), 867–886.
Westgeest, A.J., Vasseur, F., Enquist, B.J., Milla, R., Gómez-Fernández, A., Pot, D., Vile, D., Violle, C., 2024. An allometry perspective on crops. New Phytol..
Yang, K., Lee, L.-f., 2017. Identification and QML estimation of multivariate and simultaneous equations spatial autoregressive models. J. Econometrics 196 (1), 196–214.
Yang, J., Peng, J., 2020. Estimating time-varying graphical models. J. Comput. Graph. Statist. 29 (1), 191–202.