Shrinkage in Space

Spillovers in a Bayesian Hierarchical Model

Nikolas Kuschnig*

Preliminary Draft

July 4, 2023

In this paper, I present a modelling approach to jointly investigate connectivity between observations and its consequences — spillover effects. The approach is fully Bayesian, and uses hierarchical shrinkage priors to flexibly provide regularization where needed and let the data speak where it is possible. I make the prior information that is embodied in the restrictive assumptions of previous spatial models explicit, and loosen them by estimating connectivity parameters. For effective estimation, I develop efficient sampling procedures and a Gaussian process approximation to evaluate Jacobian determinants.

Keywords: network, spatial, spillovers, peer effects, network formation, neighborhood

*Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria; correspondence to nikolas.kuschnig@wu.ac.at. Thanks to Jesús Crespo Cuaresma, Sylvia Frühwirth-Schnatter, Angelo Mele, Bryan Graham, Lukas Vashold, Fabian Siuda, Frank Schorfheide, Francis X. Diebold, Marko Mlikota, and participants of the Central European University's 2022 PhD workshop as well as internal seminars at the Vienna University of Economics and Business for helpful comments. Part of the research presented here was undertaken at Johns Hopkins University. The author gratefully acknowledges financial support from the 'Graf Hardegg Stiftung', the Austrian Economic Association (NOeG), and the Austrian national bank (OeNB anniversary fund, project No. 18799).

1. Introduction

Economic theory frequently suggests connections between individual units of observation. Connectivity and the resulting spillover effects between units lie at the heart of pressing research questions in microeconomics (Ambrus et al., 2014; Chyn and Katz, 2021; Ioannides and Datcher Loury, 2004; Weidmann and Deming, 2021) and macroeconomics (Acemoglu et al., 2016; Crespo Cuaresma et al., 2019; Gofman, 2017; Rose, 2004). Good answers to these questions and deeper insights can be important for education policy (Board and Meyer-ter Vehn, 2021; Lin, 2010; Mele, 2020), labor market outcomes (Beaman, 2012; Hensvik and Skans, 2016; Munshi, 2003), supply chain management (Acemoglu et al., 2012; Atalay et al., 2014, 2011; Kranton and Minehart, 2001), innovation (Bloom et al., 2013; Ductor et al., 2014; König et al., 2019; Newman, 2001). However, empirical studies of spillover effects abstract from the nature of connections and modelling approaches are limited. Current econometric methods for analyzing spillover effects generally rely on holding the structure of connectivity fixed, thus obscuring uncertainties and potentially distorting results.

This paper introduces an integrated model to jointly analyze both the structure and consequences of connectivity between units. I propose a Bayesian hierarchical approach to comprehensively address both issues. Suitable shrinkage priors allow for more nuanced assumptions; prior information is flexibly imposed where available and needed, while important aspects of the model are freed up and learned from the data. The resulting model facilitates the explicit treatment of connectivity structures and spillovers in a layered framework. As I demonstrate with an application to deforestation spillovers from croplands in the Brazilian Amazon, this model avoids bias from misspecified connectivities, better reflecting reality and the surrounding uncertainties.

The main contribution of this paper is a framework for jointly estimating spillover/peer effects and connectivity structures, i.e., spatial weight or adjacency matrices. A number of earlier studies pursue similar goals in the spatial econometric (Debarsy and LeSage, 2020; Lam and Souza, 2020; Qu and Lee, 2015; Zhang and Yu, 2018) and the network econometric (Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2016; Johnsson and Moon, 2021) literature. These studies generally adopt model averaging approaches, and are limited to few discrete candidates for the connectivity structure. The hierarchical setup that I propose in this paper eliminates this constraint. It provides a flexible and nuanced way of modelling and estimating more general forms of connectivity, conveying a more comprehensive picture of spillover effects. For this approach to work in practice,

there are three main obstacles to overcome.

The first challenge lies in the forms of the connectivity structure themselves. Here, I provide a unified framework for common forms of connectivity used in both the spatial and network econometric literatures. I characterize connectivity as a weighted digraph, and develop a functional approximation that nests many common connectivity structures. This approximation places individual units in a metric space, augmented with a notion of potential, and highlights the strong simplifying assumptions made in the literature. Lastly, I derive conclusive bounds for the autoregressive parameter of overall connectivity strength, ensuring non-singularity and stationarity of the model. This allows for more flexible models that don't impose a row-stochastic form a priori, and, as I demonstrate, avoid bias from this potential misspecification of connectivity.

The second challenge is related to the curse of dimensionality. I address these using suitable prior distributions that can induce regularization. With few exceptions (e.g. Lam and Souza, 2020), the notion of regularization is arguably neglected in the literature, and Uniform priors are most commonly used (e.g. Debarsy and LeSage, 2020). I show how established priors are limited in their flexibility, impose strong information in unsuspected dimensions, and often cannot accommodate the nuanced prior information that is available. I introduce priors that can reflect this prior information, e.g., location-based priors for the structural parameters in the connectivity function. For the autoregressive parameter of overall connectivity strength, I propose a Beta-Gamma mixture prior that can accommodate flexible shapes and provides sensible regularization without distorting estimates. These priors facilitate the efficient estimation of connectivity parameters that have previously been fixed.

The third challenge is of a technical nature and concerns estimation of the model, and the computations involved. Posterior inference relies on Markov chain Monte Carlo (MCMC) sampling or variational methods, and the interdependence that results from connectivity can make even simple models computationally prohibitive. Models with an autoregressive term, in particular, rely on the evaluation of a costly Jacobian determinant. In this paper, I focus on full posterior inference, and develop efficient sampling schemes for the hierarchical and structural parameters, which facilitate the straightforward estimation of extensible models. For such models, the Jacobian determinant features at least one additional dimension, making established procedures obsolete. In order to still allow for rapid and accurate estimation, I develop an adaptive Gaussian process approximation. These optimizations allow for full Bayesian inference in spatial econometric models that can be extended with little overhead.

I apply this modelling approach by revisiting the impacts of agriculture on deforestation in Mato Grosso, Brazil (following Kuschnig et al., 2021). Mato Grosso lies at the intersection of the Amazon rainforest, the Cerrado savannah, and the Pantanal wetlands, and is a historical deforestation hotspot and a modern agricultural powerhouse (Macedo et al., 2012). The dynamics of deforestation in this region are particularly insightful due to the presence of these biomes. Different deforestation interventions[1] across them induce leakage, where deforestation in the Amazon is displaced to other biomes (Villoria et al., 2022), and spillover effects, where, for instance, pastures are replaced by croplands and displaced into formerly forested areas (Kuschnig et al., 2021). In this setting, connectivity between observations plays a crucial role, and geographical and theoretical information help inform its structures. As I will demonstrate, considerable uncertainty remains, and the proposed hierarchical approach yields deeper and more robust insights.

The remainder of this paper is structured as follows.

2. Methods

In this section, I (i) introduce the framework, (ii) describe a hierarchical modelling approach for the structure of connectivity, (iii) and derive conclusive results for the normalization of the connectivity that allow for the identification of the overall connectivity strength.

2.1. Framework

Consider a set of *agents* $\mathscr{A} = \{1, \dots, N\}$; for each agent, we observe some *response* $Y \in \mathbb{R}$ and a vector of *characteristics* $X \in \mathbb{R}^P$ a total of T times.[2] We want to learn about the relationship between the response and the characteristics, but have reason to suspect that agents are not independent of each other — they are connected in a *network*.

We can formalize this network using the *graph* $\mathscr{G} = \{\mathscr{A}, \mathscr{E}\}$, where $\mathscr{E}$ is a set of connections (or edges) between agents. Specifically, we represent the network using a *weighted digraph*. This means that edges are *directed* from one agent to another, that

[1] Examples include the Brazilian Forest Code, which requires (inter alia) the maintenance of a certain percentage of natural vegetation that differs across biomes (Garrett et al., 2021), and the Soy Moratorium, which bans soy from previously deforested areas in the Amazon from the supply chains of participants (Gibbs et al., 2015), and different land tenure regimes (Carrero et al., 2022).

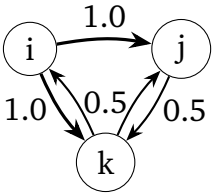
[2] For simplicity, we will generally assume that $T = 1$.

is, connections need not be reciprocal. Moreover, the graph is equipped with a *weight* function that measures the strength of all possible connections. That is

$$w : \mathscr{A} \times \mathscr{A} \mapsto \mathbb{R}^+,$$

which satisfies $w(i,i) = 0$, meaning that there are no self-links, and $w(i,j) \geqslant 0$ for all $i, j \in \mathscr{A}$, where equality holds if and only if the edge $\{i,j\} \notin \mathscr{E}$.

We can represent the graph $\mathscr{G}$ with its $N \times N$ *adjacency matrix* $\mathbf{G}$. Its entries correspond to the weight function, meaning that $g_{ij} = w(i,j)$ and consequently $g_{ii} = 0$ for all $i, j \in \mathscr{A}$. An agent i has the *neighbor* j if $g_{ij} > 0$ and is a neighbor to j if $g_{ji} > 0$. For the purpose of identification, it is useful to work with a *normalized* adjacency matrix. For this purpose, we introduce the matrix $\mathbf{W}$, which is normalized such that its spectral radius is unity.



In this framework, we are interested in learning about

$$Y = f(\mathscr{G}, X) + \varepsilon.$$

Departing from any economist's favorite, linear model, an agent's response may additionally depend on the network in terms of the responses and the characteristics of their neighbors. An extension that allows us to account for these network effects is

$$\mathbf{y} = \lambda \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\theta} + \mathbf{e}, \tag{1}$$

where we introduce *network lags* of the responses and characteristics, i.e., weighted averages of the neighbors' values, on the right-hand side.

This specification resembles the standard models in the network- and spatial-econometric literatures,[3] and captures network effects in a parsimonious manner. The autoregressive term is of particular note, as it gives rise to a *network filter*. This becomes apparent when expressing Equation 1 as

$$\begin{aligned} \mathbf{z} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\theta} + \mathbf{e}, \\ \mathbf{y} &= (\mathbf{I} - \lambda \mathbf{W})^{-1}\mathbf{z}, \end{aligned}$$

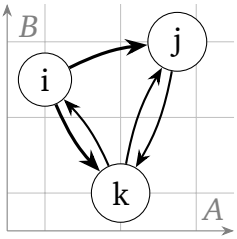
[3] Specifically, it resembles the *linear-in-means* and the *spatial Durbin* models. These models generally normalize $\mathbf{W}$ to be *row-stochastic*, meaning that the network lags yield the unweighted average of connected agents. In these literatures, the network itself is assumed to be known up to a few candidates.

where $\mathbf{z}$ is a latent variable, and $\mathbf{S} = (\mathbf{I} - \lambda\mathbf{W})$ can be understood as a link function.

2.2. A hierarchical model for $\mathscr{G}$

For a comprehensive treatment of the model expressed in Equation 1, we want to model the normalized adjacency matrix and the network $\mathscr{G}$ that gives rise to it. A straightforward approach would be to consider single connections, i.e., off-diagonal entries of the adjacency matrix. However, at $\mathcal{O}(N^2)$ unknown entries, this approach suffers from the curse of dimensionality in most settings. There is little information that can be learned from the data, unless the network is static over time and we can observe the agents repeatedly. We can address this issue by directly imposing shrinkage on the unknown connections, or by imposing structure on the model of our network.

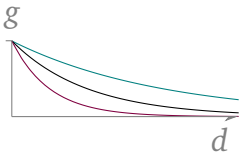
Assume that we can locate our agents in some metric space $(\mathscr{P}, d)$. Then, a natural way to think of connectivity is in terms of the distance. Let $\mathbf{p}_i$ denote the (latent) coordinates of agent i, then we could model $\mathscr{G}$ using, for instance,



$$g_{ij} = \exp\left\{-d\left(\mathbf{p}_i, \mathbf{p}_j\right)\right\}.$$

This setup may be an appealing approximation for some networks, but it suffers from two major drawbacks — a lack of *sparsity* and the imposition of *symmetry*. Standard distances are symmetric and don't tend to diverge to infinity, constraining their utility for modelling sparse, directed networks. We can alleviate these concerns with slight adaptations.

In the first step, we introduce a parameter that controls the speed of decay,



$$g_{ij} = \exp\left\{-\delta_i \times d\left(\mathbf{p}_i, \mathbf{p}_j\right)\right\}.$$

This setup allows us to account for asymmetries, and facilitates the sparsification of connections. An agent with strong and far-reaching influence on others may, in addition to a central location, experience less distance-decay. This is captured by smaller parameter values, such as $\delta_i = 0.5$. Meanwhile, most agents may only have local influence on others, as expressed by a higher degree of distance-decay. Larger parameter values, such as $\delta_i = 2$, restrain spillovers to be localized, and can effectively allow for sparsification. For instance, we may want to limit connections to a ball around agents by including a minimum threshold for positive connectivity.

In the second step, we address the number of connections by modelling them in the

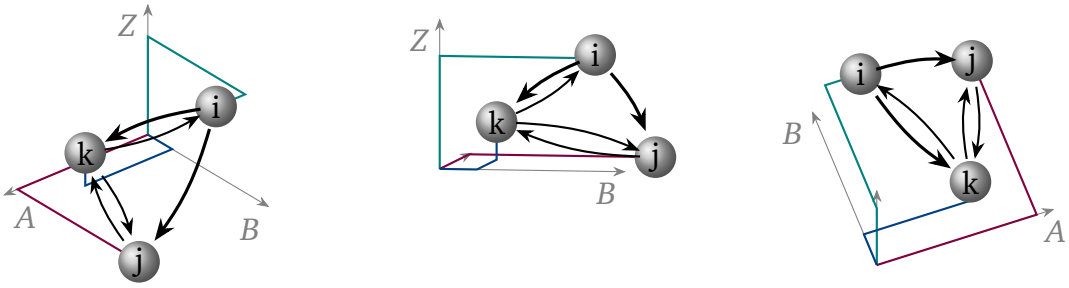


FIGURE 1: Illustration of a three-agent network from three angles. Agents are located in a 2D space with the coordinates A, B, and their connectivity is mediated by the potential, Z.

style of a k-nearest neighbors algorithm. Let $\mathbf{d}_i$ be a vector of distances from i, and d_{ij} be a short-hand for the distance from agent i to j. Then we extend the model using an indicator function, $\mathbf{1}$, as

$$g_{ij} = \exp\left\{-\delta_i \times \mathrm{d}_{ij}\right\} \times \mathbf{1}\left[\mathrm{d}_{ij} \leqslant \kappa\left(k_i \mid \mathbf{d}_i\right)\right],$$

where the function $\kappa(k_i)$ yields the k_i'th lowest distance from i. The parameter k_i controls the number of connections of an agent, and can be intuitively understood as a follower count.

Models of $\mathscr{G}$ that follow this pattern can be expressed as

$$g_{ij} = f\left(k_i, \delta_i, \mathbf{p}_i, \mathbf{p}_j \mid (\mathscr{P}, d)\right), \tag{2}$$

and can serve as flexible approximations. The central assumption places the agents in a metric space; the constraints of this representation are alleviated by introducing the measures of *popularity*, δ_i and k_i. The effects of this popularity, or potential, are illustrated in Figure 1. Further flexibility can be gained by considering generalized notions of metric spaces that suit the hypothesized underlying network.[4]

Networks constructed from Equation 2 generalize a number of commonly used network specifications, which can be obtained by restraining the $\mathcal{O}(N)$ unknown parameters. A standard example in a spatial setting would, for instance, use spherical distances between physical locations and a distance-decay specification. In a social network setting, coordinates may refer to homophilic characteristics, with distances induced by an intrinsic metric on their divergence. The implied metric space may be pruned via a k-nearest specification and could be augmented with popularity measures to capture

[4]We can interpret our graph $\mathscr{G}$ as a generalized metric space on the set of agents, with a quasimetric that is induced by the smallest weight of any directed path (with distance ∞ if no such path exists).

the structures of familiar social networks.

2.3. Identification, and normalization

To identify the strength of network effects, as captured by λ and $\boldsymbol{\theta}$, we need to normalize the adjacency matrix in some way. In addition, we require the link function, $\mathbf{S}\left(\lambda, k_i, \delta_i, \mathbf{p}_i, \mathbf{p}_j\right)$, to be invertible and stationary. As I will show below, normalization with the inverse spectral radius of the adjacency matrix, $\rho\left(\mathbf{G}\right)^{-1}$, addresses these concerns.

Theorem 1. *Let* $\mathbf{I}$ *denote the identity matrix, and* α *be a real scalar. Then* $(\mathbf{I} - \alpha\mathbf{G})$ *is* non-singular *for* $\alpha \in \left(\omega_{\min}^{-1}, \omega_{\max}^{-1}\right)$*, where* $\omega_{\min}$ *and* $\omega_{\max}$ *are the minimum and maximum real eigenvalues of* $\mathbf{G}$*.*

Proof. This statement is true if $\alpha \times \omega_i \neq 1$ for all i, which we will show directly. For $\omega_i = 0$, this is trivially the case; we need to show it for all $\omega_i \neq 0$. Notice that $\text{trace}(\mathbf{G}) = 0$, which (combined with $\omega_i \neq 0$) implies that $\omega_{\min} < 0$ and $\omega_{\max} > 0$. In order to show our result, we have two requirements. For positive eigenvalues we need to show that $\alpha < \omega_i^{-1}$, and for negative ones that $\alpha > \omega_i^{-1}$. The result follows from knowing that $\omega_{\min}^{-1} < \alpha < \omega_{\max}^{-1}$. □

Non-singularity ensures the existence of the filter, and the continuity of its determinant. Next, we are concerned with its stationarity.[5] We can guarantee stationarity of the filter if its Neumann series is convergent, or equivalently, if the spectral radius of $\mathbf{S}$ is less than unity. The Perron-Frobenius theorem gives us the following result.

Corollary 1. *The spectral radius of* $(\mathbf{I} - \alpha\mathbf{G})$ *is less than unity for* $\alpha \in \left(-\omega_{\max}^{-1}, \omega_{\max}^{-1}\right)$.

Proof. This statement is true if $\rho\left(\mathbf{G}\right) = \omega_{\max}$, which we will show directly. If $\mathbf{G}$ is irreducible (or, equivalently, if $\mathcal{G}$ is strongly connected) the result follows directly from the Perron-Frobenius theorem. Otherwise, the result follows from knowing that $\mathbf{G}$ is either the null matrix, or is, up to an isomorphism, composed of null and irreducible submatrices. □

As can be seen, conclusive bounds for λ are determined by the spectral radius of $\mathbf{G}$, which coincides with its maximum real eigenvalue. This result is good news for the

[5] In the network context, stationarity can be understood as the notion that the (absolute) connectivity should be decreasing with the order of neighbors, i.e., units are more connected to their direct neighbors than to their neighbors' neighbors.

literature, where the standard domain, $\lambda \in (-1, 1)$, is supported by the standard choice of row-stochastic normalization for $\mathbf{W}$. At the same time, the spectral radius, $\rho(\mathbf{G})$, emerges as an attractive, structure-preserving, alternative normalization factor.[6]

3. Priors and estimation

Effective estimation of models in the form of Equation 1 and Equation 2 is aided by a degree of regularization. In the Bayesian framework, this is readily achieved via priors, which provide a suitable framework for introducing external information. Below, I present suitable priors for unknown parameters in the models considered. Afterwards, I discuss their estimation using Markov chain Monte Carlo (MCMC) methods.

3.1. Hierarchical shrinkage priors

In the wider econometric literature, the overall connectivity strength λ is the only parameter considered to be unknown, and plays a central role when analyzing connectivity. With a standardized connectivity matrix, the domain $(-1, 1)$ guarantees an invertible and stationary network filter. Bayesian approaches generally assign the parameter a Uniform prior; a useful generalization (first proposed by LeSage and Parent, 2007) is the Beta prior

$$p(\bar{\lambda}) \sim \text{Be}(1 + \tau, 1 + \tau),$$

where $\tau \geq 0$, and we use $\bar{\lambda} = (\lambda + 1)/2$, scaled to live on $(0, 1)$, for simplicity. Here, $\text{Be}(a, b)$ denotes the density of a Beta distribution with shapes a and b, i.e.

$$\text{Be}(x \mid a, b) = \frac{x^{a-1}(1-x)^{b-1}}{\text{Beta}(a, b)}.$$

For $\tau = 0$, the prior is uniform over all values. With this in mind, the prior parameter τ can be understood as excess support for the origin — but, as we will see, this

[6] For large connectivity matrices, the prospect of computing its spectral radius may appear daunting, especially when that matrix is mutable. While direct methods for determining eigenvalues can be computationally prohibitive at a complexity of $\mathcal{O}(N^3)$, iterative methods, such as the Lanczos method or Arnoldi iteration, can provide a remedy. These methods allow us to only compute the required largest eigenvalue, converge to an exact result at a general complexity of $\mathcal{O}(N^2)$, and particularly benefit from sparsity, which is a common feature of larger networks and a desirable property of approximations.

interpretation is misleadingly narrow.

Beyond the Uniform prior, we sometimes encounter $\tau = 0.01$ in the literature, which results in a fairly flat prior that places slightly higher weight at the origin. On one hand, this reflects a clear orientation on flat priors that are uninformative with respect to certain values. On the other hand, however, this is driven by necessity due to the Beta distribution's undesirable properties. These undesirable properties are moderate peaks in density, and excessive drop-offs towards the tails, as illustrated in Figure 2. There, we can see three Beta densities, placing increasing mass at the origin. Even for $\tau = 100$, the peak remains moderate, while the density at the tails (e.g. at 0.9) becomes miniscule, and the credible support of the prior incredibly narrow.

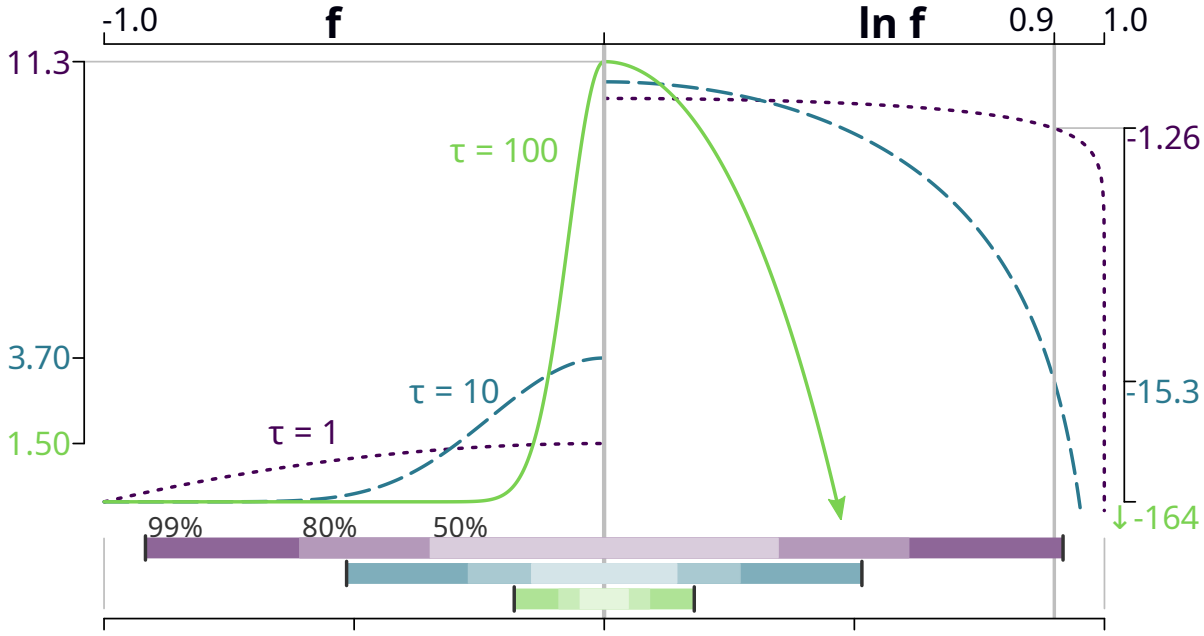


FIGURE 2: Density (left), log-density (right), and 99%, 80%, and 50% credible intervals (bottom) for λ with a Beta prior, i.e. $p(\bar{\lambda}) \sim \text{Be}(1+\tau, 1+\tau)$, with $\tau \in \{1, 10, 100\}$. More informative priors (with increasing τ) lead to narrow credible intervals, with values in the tail (e.g. 0.9, compare the right panel) receiving infinitesimal prior support.

With any prior distribution, we want to express the prior information that is available to us without distorting insights that we can obtain from the data. For our prior for λ, we thus want to incorporate information and avoid potential distortions. First, consider the often implicit, and sometimes explicit preference for parsimony, i.e. $\lambda = 0$. Without decent support for non-zero values, it appears reasonable to skip the superfluous complexity of connectivity. This potentially special role of zero highlights the distorting information induced by the Uniform prior. While this flat prior indicates no preference

for any values, it implicitly prefers large values, essentially imposing connectivity on the model. For instance, we have $p(|\lambda| > 0.1) = 9 \cdot p(|\lambda| \leq 0.1)$, i.e., .

This motivates our departing point, which is the the following mixture prior

$$p(\bar{\lambda}) \sim \begin{cases} \text{Be}(1+\tau_0, 1+\tau_0), & \text{if } \gamma = 1, \\ \text{Be}(1+\tau_1, 1+\tau_1), & \text{if } \gamma = 0, \end{cases}$$

where $\tau_0 \ll \tau_1$ are shape parameters, and γ is an indicator. This prior essentially represents a variable selection procedure for λ. For $\gamma = 0$, the sharp spike at zero that is induced by τ_1 leads to a collapse to the linear model. For $\gamma = 1$, we have a comparatively flat prior for λ, mirroring standard setups. This small adaptation to a spike-and-slab prior takes us in the right direction conceptually, but arguably remains too rigid (except, perhaps, for panels with time-specific connectivity). Next, we consider further sources of prior information, and develop a practical prior that can accommodate our prior convictions.

The second source of prior information we want to reflect concerns the model specification — in particular, the boundaries of λ. Parameters that lie at their boundary commonly cause issues for statistical models (Chernoff, 1954; Self and Liang, 1987; Chen and Liang, 2010), and λ is no exception. Numerical instability from the filter approaching singularity is an obvious example, but not the most damning. Instead, a major issue are pathological solutions[7] that arise, not from the phenomenon under investigation, but the peculiar structure of the model (see, for example, Angrist, 2014; Halleck Vega and Elhorst, 2015). In the face of these technical and theoretical caveats, it is sensible to regularize λ and limit the support at the boundaries by imposing a suitable prior.

I propose the following continuous mixture of Beta distributions as a prior, which can coalesce the above sources of prior information in a straightforward way.

$$p(\bar{\lambda}) \sim \text{Be}(1+\tau, 1+\tau), \qquad p(\tau) \sim \text{Ga}(a, b), \tag{3}$$

where the mixing density is a Gamma distribution with shape a and rate b, which has proven to be useful for this purpose in similar settings (see Park and Casella, 2008; Griffin and Brown, 2010). This mixture affords us considerable flexibility for prior elicitation, allowing us to act upon the prior information discussed above.

The Beta-Gamma mixture can accommodate a wide range of shapes, as exemplified

[7] One example for a model considered here, is the perfect fit from $\lambda \to -1$, $\delta \to 0$, and $\alpha = \sum_{i=1}^{n} y_i$.

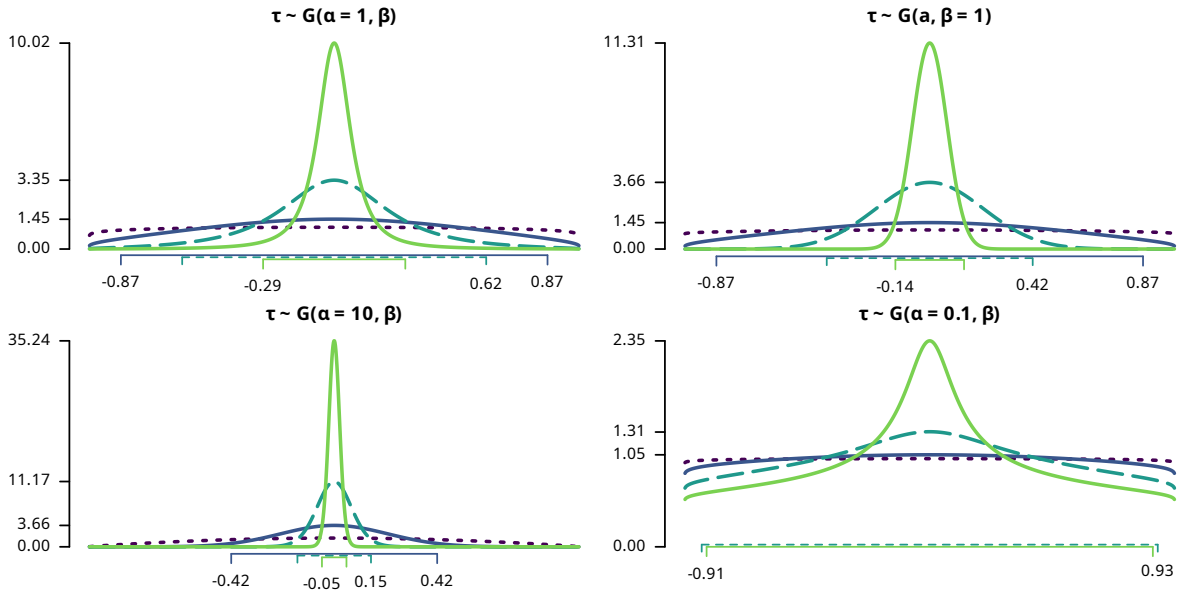


FIGURE 3: Density (left), log-density (right), and 99%, 80%, and 50% credible intervals (bottom) with a Beta-Gamma prior.

in Figure 3. We can place considerable mass at the origin, while also accommodating values in the tails. If we constrain the prior to the special case of an Exponential ($a = 1$, top-left) and orient ourselves on Figure 2, we can clearly see that the mixture prior yields more pronounced peaks with wider credible support and without excessive drop-offs. By varying both parameters, we are able to flexibly induce fine-tuned priors, essentially without overhead.

The proposed Beta-Gamma shrinkage prior changes the prior specification from an issue of choosing specific values, to one concerned with parsimony and regularization. Increased weight at the origin means that the prior does not induce spillover effects per se, while the tail behavior allows us to provide regularization without limiting support to narrow regions a priori. The result is a flexible prior that can better express many prior convictions. Nonetheless, there may be reservations to even weakly informative priors.[8] While they may not always be a necessity, it is important to keep in mind that we introduce this structure to free up the previously implicit and infinitely informative priors that are fixed structural parameters.

The structure of the network is determined by unknown parameters concerning (1) the locations, (2) the speed of distance-decay, and (3) the number of neighbors. For

[8] Ignore, for this example, that the standard flat priors are heavily informative in terms of size.

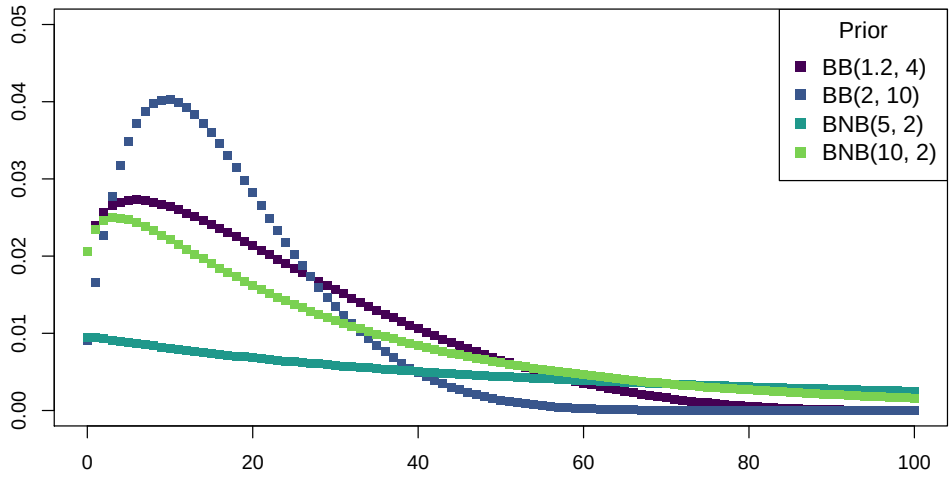


FIGURE 4: Visualization of different prior setups for k, with $n = 100$.

each, I present suitable prior distributions that allow for more flexible and credible models.

The first candidate for more explicit treatment is the number of neighbors in a k-nearest neighbor specification. Limited variations to the parameter are commonly considered for robustness checks, and have previously been addressed within a model averaging framework (see Lesage and Fischer, 2008; Debarsy and LeSage, 2020; Zhang and Yu, 2018, for instance). While the discrete parameter lends itself to model averaging, we will take it one step further and treat k itself in a fully Bayesian way. For this, we can consider it as the result of N trials that determine whether any two units are neighbors or not. Such a trial can be modelled with a Beta-binomial distribution, i.e.

$$p(k) \sim \text{BB}(N, a, b),$$

where N is the number of trials (i.e. potential neighbors), a and b describe the probability of success, and $\text{BB}(N, a, b)$ denotes the density of a Beta-binomial distribution, i.e.

$$\text{BB}(x \,|\, N, a, b) = \binom{N}{x} \frac{\text{Beta}(a + x, b + N - x)}{\text{Beta}(a, b)}.$$

With this prior, there is no need to constrain our model to a selection of values. We can open the full model space by setting $N = n - 1$, retain sparsity in connections by choosing $a < b$ appropriately, and allow for efficient estimation using Markov chain Monte Carlo methods.

In the literature, there is a strong preference for few neighbors, expressed in empirical

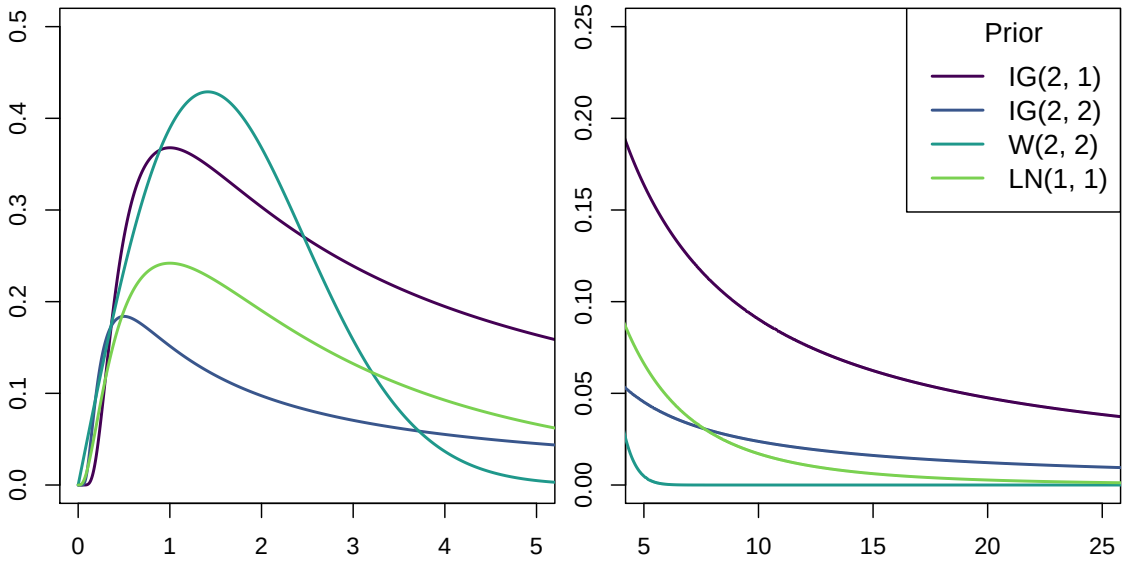


FIGURE 5: Visualization of different prior setups for δ. The left panel highlights behavior near the focal value of $\delta = 1$; the right panel shows the tail behavior.

work and motivated by practical and theoretical reasons. In Figure 4, we can see that the Beta-Binomial prior allows for nuanced priors that reflect this preference. However, it mirrors the underlying Beta distribution, in that it experiences excessive drop-off in the tails. If this aspect is considered restrictive, a Beta-negative-Binomial prior (visualized) can present an even less informative alternative.

Next, we consider the distance-decay parameter δ. Earlier works that are limited to local lags show that there is impactful uncertainty around this parameter (Halleck Vega and Elhorst, 2015; Kuschnig, 2022). Standard model averaging approaches provide no remedy due to the continuous nature of δ. With our fully Bayesian approach, however, we merely need a sensible prior for the parameter. A useful option is

$$\delta \sim \text{Ga}^{-1}(a, b).$$

The inverse-Gamma distribution is flexible enough to accommodate general prior conceptions. For our parameter, an important benefit is that it avoids placing weight at and near zero values, where every unit is equally connected to every other unit. If these small values at the left tail are a problem, a log-Normal prior can be a useful alternative. When strong and explicit prior information is available, the Weibull distribution can be another useful alternative. For a visualization of selected priors, see Figure 5.

Regarding specific values for δ, the literature is not particularly informative. While

suitable values depend on the distances involved, $\delta = 1$ can serve as an anchor for prior elicitation. Below it, i.e. for $\delta \ll 1$, connections between neighbors are weighted more equally, with less regard to the distance. For $\delta \gg 1$, by contrast, only the closest neighbors retain relevance as connectivity levels off faster than the distance.

3.2. Sampling-based estimation

We consider the model in Equations 1 and 2, where we constrain the connectivity to $f(\lambda, \delta)$ to illustrate. In this section, we describe a Markov chain Monte Carlo (MCMC) approach to obtain full posteriors of this setup.

First, note that we can readily obtain posterior draws of $\left(\boldsymbol{\beta}, \sigma^2\right)$ conditional on (λ, δ, τ) using the approach by Makalic and Schmidt (2015). Next, we draw from the conditional posterior of λ, and then τ. Finally, we draw from the conditional posterior δ, and repeat — giving us the procedure in Figure 6.

0. Set starting values for $\lambda, \delta, \tau, \sigma^2$.
1. Draw from the conditional posteriors of the nested linear model,
   a) $p\left(\boldsymbol{\beta} \,|\, \mathbf{y}, \lambda, \delta, \tau, \sigma^2\right)$,
   b) $p\left(\sigma^2 \,|\, \mathbf{y}, \lambda, \delta, \tau, \boldsymbol{\beta}\right)$.
2. a) Draw from $p\left(\lambda \,|\, \mathbf{y}, \boldsymbol{\beta}, \sigma^2, \delta, \tau\right)$, and
   b) $p\left(\tau \,|\, \mathbf{y}, \lambda, \cdot\right)$.
3. Draw from $p\left(\delta \,|\, \mathbf{y}, \boldsymbol{\beta}, \sigma^2, \delta\right)$.
4. Go to the first step until enough draws are obtained.

FIGURE 6: Stylized algorithm for sampling from the model.

For the first step, we can rely on standard techniques, such as Gibbs sampling, by conditioning on the connectivity parameters. In the second and third steps, the conditional posteriors of λ, τ, and δ have no well-known form, and we must use another approach.

For λ, we can use a Metropolis-Hastings step to draw from its conditional posterior,

$$p(\lambda \,|\, \mathbf{y}, \tau, \cdot) \propto |\mathbf{S}(\lambda, \delta)| \exp\left\{-\frac{1}{2\sigma^2}\left(\mathbf{S}(\lambda, \delta)\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{S}(\lambda, \delta)\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)\right\} p(\lambda \,|\, \tau).$$

The conditional posterior of τ can be expressed as

$$\begin{aligned} p(\tau \,|\, \mathbf{y}, \lambda, \cdot) &\propto \lambda^{\tau} \, (1-\lambda)^{\tau} \, \tau^{a-1} \exp^{-\tau b}, \\ &\propto \tau^{a-1} \exp^{-\tau \left[b - \log\left(\lambda - \lambda^2\right) \right]}, \end{aligned}$$

which is the kernel of a Gamma density, which we can directly draw from using a Gibbs step. This means that our hierarchical prior setup for λ imposes essentially no overhead over conventional specifications.

Lastly, another Metropolis-Hastings step allows us to draw from

$$p(\delta \,|\, \mathbf{y}, \cdot) \propto |\mathbf{S}(\lambda, \delta)| \exp\left\{ -\frac{1}{2\sigma^2} \left(\mathbf{S}(\lambda, \delta)\mathbf{y} - \mathbf{X}\boldsymbol{\beta} \right)' \left(\mathbf{S}(\lambda, \delta)\mathbf{y} - \mathbf{X}\boldsymbol{\beta} \right) \right\} p(\delta).$$

With the exception of τ, these sampling steps are well-known, and all of them are conceptually straightforward. However, they pose one major computational challenge — that is, the determinant of the $N \times N$ Jacobian matrix $\mathbf{S}(\lambda, \delta)$.

3.3. Evaluating the Jacobian determinant

The likelihood, and hence the posterior, of our model involves a Jacobian determinant, which poses a central computational constraint for estimation (Bivand et al., 2013). In standard models, we can use a spectral decomposition of the fixed connectivity matrix $\mathbf{W}$ to compute the determinant with the eigenvalue method, using

$$\ln |\mathbf{I} - \lambda \mathbf{W}| = \sum_{i=1}^{n} \ln \left(1 - \lambda \omega_i\right).$$

There are other approaches for large matrices, e.g. based on the lower-upper decomposition, spline approximations, or algebraic results that make use of special connectivity structures (see Bivand et al., 2013). However, all of these approaches rely on the connectivity structure in $\mathbf{W}$ being fixed, and would thus present a potentially insurmountable computational challenge for more flexible models.

In order to still allow for rapid estimation using MCMC, we introduce the following Gaussian process approximation

$$|\mathbf{S}(\lambda, \delta, \dots)| \approx \text{GP}\left(\mu(\lambda, \delta, \dots), \boldsymbol{\Sigma}(\lambda, \delta, \dots)\right),$$

This allows us to approximate the Jacobian determinant with high accuracy (cf. the

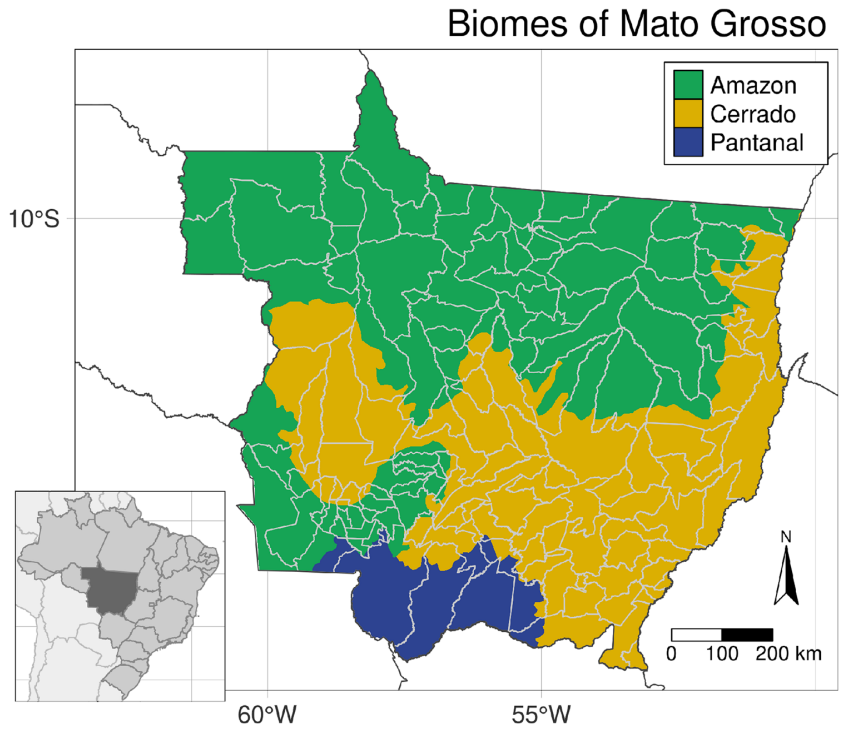


FIGURE 7: Biomes and location of Mato Grosso (taken from Kuschnig et al., 2021, under CC-BY 4.0).

Supplementary Material). Essentially, we compute the eigenvalues for a grid of values (of δ, etc.), use those to determine $|\mathbf{S}(\lambda, \delta, \dots)|$ using the eigenvalue method, and fit these training samples using Gaussian process regression. This approach provides a quantification of uncertainty, and allows for retraining if the sampler moves to values far from the grid. For our approach, we rely on a constant mean, μ, and a Gaussian kernel for Σ, but other options are available. Notably, Gaussian processes are widespread in the field of spatial statistics, and have a parallel in spline regression, which can be used to approximate one-dimensional Jacobian determinants.

4. An application to Amazon deforestation

In this section, I apply the proposed modelling approach to investigate the impacts of agriculture on deforestation in Mato Grosso, Brazil. First, I briefly introduce the model and data used. Then, I present results on the impacts of croplands that are obtained using increasingly flexible connectivity models. Finally, present and discuss the structure of the underlying networks.

4.1. Model & data

The analysis uses the original panel dataset of 141 municipalities in the state of Mato Grosso from 2006–2017 ($N = 1,692$). The model is a variant of Equation 1 that suppresses local spillovers ($\boldsymbol{\theta} = \mathbf{0}$) and includes time- and municipality-fixed effects, i.e.

$$\mathbf{y}_t = \lambda \mathbf{W}(\cdot)\mathbf{y}_t + \mathbf{X}_{t-1}\boldsymbol{\beta} + \boldsymbol{\mu} + \boldsymbol{\psi}_t + \boldsymbol{\varepsilon}_t.$$

The dependent variable $\mathbf{y}_t$ is the yearly change of forest per area, and the regressors are the (lagged) shares of forest, pasture, and cropland area, population density, cattle density (per pasture), soy yields in Brazilian real per harvested area, and an indicator for the incidence of particularly dry months. These potential confounders allow us to isolate the impact of agriculture on deforestation, and were chosen based on theoretical considerations and to be in-line with earlier studies (see Busch and Ferretti-Gallon, 2017). For a more detailed discussion, see Kuschnig et al. (2021).

For the result of interest, we will focus on the partial effect of croplands, which is given by

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}_{\text{crops}}} = \mathbf{S}(\lambda, \cdot)^{-1} \left(\mathbf{I}\beta_{\text{crops}} \right),$$

and depends on a connectivity-induced multiplier effect. For the structure of connectivity, we consider an exponential distance-decay function based on the centroids of municipalities. We start by fixing $\mathbf{W}$ and gradually allow for more flexibility. First, we free up δ, which controls the speed of distance-decay, using weakly informative Gamma prior. Next, we model the locations of observations using a Dirichlet prior, i.e.

$$\mathbf{p}_i \sim \text{Dirichlet}\left(\boldsymbol{\alpha}_i\right),$$

where the prior parameters are set such that the prior mean places observations at their centroids. The uncertainty around this prior location is allowed to be relatively large.

4.2. Results

In Figure 8, we can see the average partial effect of croplands (direct and indirect) of our model of choice, where δ, the speed of distance-decay, is modelled explicitly. We can see that a naive specification that fixes $\delta = 2$ underestimates both the impact of

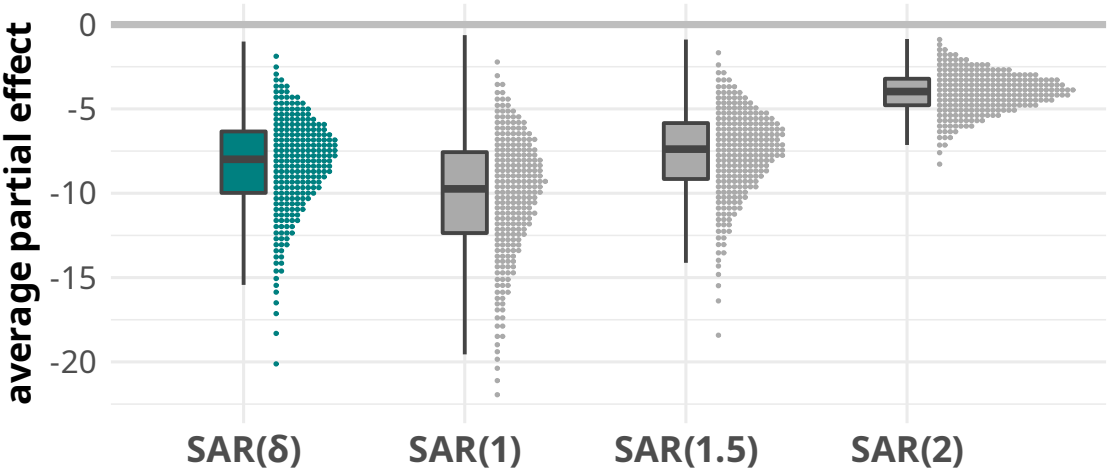


FIGURE 8: Visualization of the average partial effect of croplands on deforestation.

croplands, and the uncertainty surrounding it. For $\delta = 1$, both the impact and the uncertainty is overestimated instead. If the parameter is fixed close to its posterior mean, i.e. $\delta = 1.5$, as one might do when adopting an empirical Bayes approach, the uncertainty is underestimated.

In Figure 9, we can see the connectivity of municipalities as measured by their eigenvector centrality. On the left, we use municipality centroids with estimated δ as locations, while we additionally estimate the locations on the right. We can see that centroids lead to strong connections in the center-most municipalities, where the centroids tend to overlap due to irregular shapes. When estimating positions, we find that tightly connected observations cluster together, featuring stronger connectivity, but with very high degrees of distance-decay. These clusters of municipalities are much more spread out, and can better reflect the patterns present in the data.

5. Conclusion

In this paper, I introduced a comprehensive framework for modelling connectivity between units of observations, and the resulting spillover effects. I used a Bayesian hierarchical approach that is readily extensible, allows for regularization that helps in freeing up parameters, and natively conveys uncertainty. For this purpose, I developed a graph-based framework to unify network and spatial econometric connectivities, and derived suitable bounds to guarantee non-singularity and stationarity. I introduced the flexible Beta-Gamma mixture prior for the parameter of overall connectivity strength

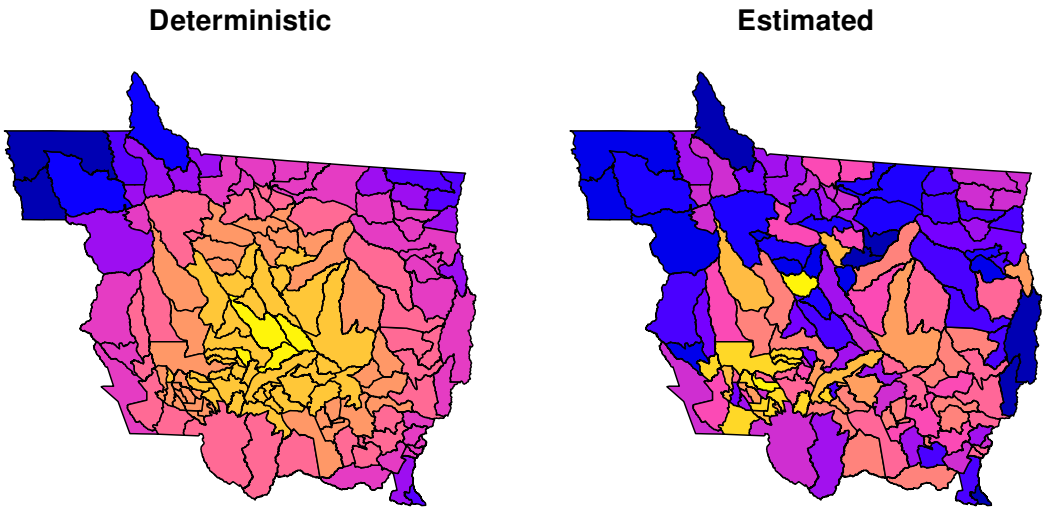


FIGURE 9: Visualization of the centrality of municipalities with deterministic and estimated locations. A weakly informative prior on the (latent) positions allows tightly connected observations to form small clusters.

in autoregressive models, and discussed suitable priors for parameters that determine the connectivity structure. I presented an efficient sampling approach that is readily extensible and allows for full posterior inference. To allow for efficient sampling with a number of free connectivity parameters, I proposed a Gaussian process approximation for the Jacobian determinant in autoregressive models. The result is a flexible and extensible framework for connectivity models that works in general settings, and was demonstrated in an application to Amazon deforestation.

References

ACEMOGLU, D., U. AKCIGIT, AND W. R. KERR (2016): "Innovation network," *Proceedings of the National Academy of Sciences*, 113, 11483–11488, 10.1073/pnas.1613559113.

ACEMOGLU, D., V. M. CARVALHO, A. OZDAGLAR, AND A. TAHBAZ-SALEHI (2012): "The network origins of aggregate fluctuations," *Econometrica*, 80, 1977–2016, 10.3982/ECTA9623.

AMBRUS, A., M. MOBIUS, AND A. SZEIDL (2014): "Consumption risk-sharing in social networks," *American Economic Review*, 104, 149–82, 10.1257/aer.104.1.149.

ANGRIST, J. D. (2014): "The perils of peer effects," *Labour Economics*, 30, 98–108, 10.1016/j.labeco.2014.05.008.

Atalay, E., A. Hortaçsu, J. Roberts, and C. Syverson (2011): "Network structure of production," *Proceedings of the National Academy of Sciences*, 108, 5199–5202, 10.1073/pnas.1015564108.

Atalay, E., A. Hortaçsu, and C. Syverson (2014): "Vertical integration and input flows," *American Economic Review*, 104, 1120–48, 10.1257/aer.104.4.1120.

Barrios, T., R. Diamond, G. W. Imbens, and M. Kolesár (2012): "Clustering, spatial correlations, and randomization inference," *Journal of the American Statistical Association*, 107, 578–591, 10.1080/01621459.2012.682524.

Beaman, L. A. (2012): "Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the U.S.," *Review of Economic Studies*, 79, 128–161, 10.1093/restud/rdr017.

Bhattacharya, A., D. Pati, N. Pillai, and D. Dunson (2015): "Dirichlet-Laplace priors for optimal shrinkage," *Journal of the American Statistical Association*, 110, 1479–1490, 10.1080/01621459.2014.960967.

Bivand, R., J. Hauke, and T. Kossowski (2013): "Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods," *Geographical Analysis*, 45, 150–179, 10.1111/gean.12008.

Bloom, N., M. Schankerman, and J. Van Reenen (2013): "Identifying technology spillovers and product market rivalry," *Econometrica*, 81, 1347–1393, 10.3982/ECTA9466.

Board, S., and M. Meyer-ter Vehn (2021): "Learning dynamics in social networks," *Econometrica*, 89, 2601–2635, 10.3982/ECTA18659.

Busch, J., and K. Ferretti-Gallon (2017): "What drives deforestation and what stops it? A meta-analysis," *Review of Environmental Economics and Policy*, 11, 3–23, 10.1093/reep/rew013.

Carrero, G. C., R. T. Walker, C. S. Simmons, and P. M. Fearnside (2022): "Land grabbing in the Brazilian Amazon: Stealing public land with government approval," *Land Use Policy*, 120, 106133, 10.1016/j.landusepol.2022.106133.

Carvalho, C. M., N. G. Polson, and J. G. Scott (2010): "The horseshoe estimator for sparse signals," *Biometrika*, 97, 465–480, 10.1093/biomet/asq017.

Chen, Y., and K.-Y. Liang (2010): "On the asymptotic behaviour of the pseudolikelihood ratio test statistic with boundary problems," *Biometrika*, 97, 603–620, 10.1093/biomet/asq031.

Chernoff, H. (1954): "On the distribution of the likelihood ratio," *The Annals of Mathematical Statistics*, 573–578, 10.1214/aoms/1177728725.

Chyn, E., and L. F. Katz (2021): "Neighborhoods matter: Assessing the evidence for place effects," *Journal of Economic Perspectives*, 35, 197–222, 10.1257/jep.35.4.197.

Corrado, L., and B. Fingleton (2012): "Where is the economics in spatial econometrics?" *Journal of Regional Science*, 52, 210–239, 10.1111/j.1467-9787.2011.00726.x.

Crespo Cuaresma, J., G. Doppelhofer, M. Feldkircher, and F. Huber (2019): "Spillovers from US monetary policy: Evidence from a time varying parameter global vector autoregressive model," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182, 831–861, 10.1111/rssa.12439.

Datta, J., and Jayanta. K. Ghosh (2013): "Asymptotic properties of Bayes risk for the Horseshoe prior," *Bayesian Analysis*, 8, 111–132, 10.1214/13-BA805.

Debarsy, N., and J. P. LeSage (2020): "Bayesian model averaging for spatial autoregressive models based on convex combinations of different types of connectivity matrices," *Journal of Business & Economic Statistics*, 1–12, 10.1080/07350015.2020.1840993.

Ductor, L., M. Fafchamps, S. Goyal, and M. J. van der Leij (2014): "Social networks and research output," *Review of Economics and Statistics*, 96, 936–948, 10.1162/REST_a_00430.

Garrett, R. D., F. Cammelli, J. Ferreira, S. A. Levy, J. Valentim, and I. Vieira (2021): "Forests and sustainable development in the Brazilian Amazon: History, trends, and future prospects," *Annual Review of Environment and Resources*, 46, 625–652, 10.1146/annurev-environ-012220-010228.

Gibbs, H. K., L. Rausch, J. Munger et al. (2015): "Brazil's soy moratorium," *Science*, 347, 377–378, 10.1126/science.aaa0181.

Gofman, M. (2017): "Efficiency and stability of a financial architecture with too-interconnected-to-fail institutions," *Journal of Financial Economics*, 124, 113–146, 10.1016/j.jfineco.2016.12.009.

Goldsmith-Pinkham, P., and G. W. Imbens (2013): "Social networks and the identification of peer effects," *Journal of Business & Economic Statistics*, 31, 253–264, 10.1080/07350015.2013.801251.

Griffin, J. E., and P. J. Brown (2010): "Inference with Normal-Gamma prior distributions in regression problems," *Bayesian Analysis*, 5, 171–188, 10.1214/10-BA507.

Halleck Vega, S., and J. P. Elhorst (2015): "The SLX model," *Journal of Regional Science*, 55, 339–363, 10.1111/jors.12188.

Hensvik, L., and O. N. Skans (2016): "Social networks, employee selection, and labor market outcomes," *Journal of Labor Economics*, 34, 825–867, 10.1086/686253.

Hsieh, C.-S., and L. F. Lee (2016): "A social interactions model with endogenous friendship formation and selectivity," *Journal of Applied Econometrics*, 31, 301–319, 10.1002/jae.2426.

Ioannides, Y. M., and L. Datcher Loury (2004): "Job information networks, neighborhood effects, and inequality," *Journal of Economic Literature*, 42, 1056–1093, 10.1257/0022051043004595.

Johnsson, I., and H. R. Moon (2021): "Estimation of peer effects in endogenous social networks: Control function approach," *Review of Economics and Statistics*, 103, 328–345, 10.1162/rest_a_00870.

König, M. D., X. Liu, and Y. Zenou (2019): "R&D networks: Theory, empirics, and policy implications," *Review of Economics and Statistics*, 101, 476–491, 10.1162/rest_a_00762.

Kranton, R. E., and D. F. Minehart (2001): "A theory of buyer-seller networks," *American Economic Review*, 91, 485–508, 10.1257/aer.91.3.485.

Kuschnig, N. (2022): "Bayesian spatial econometrics: A software architecture," *Journal of Spatial Econometrics*, 3, 6–25, 10.1007/s43071-022-00023-w.

Kuschnig, N., J. Crespo Cuaresma, T. Krisztin, and S. Giljum (2021): "Spillover effects in agriculture drive deforestation in Mato Grosso, Brazil," *Scientific Reports*, 11, 1–9, 10.1038/s41598-021-00861-y.

Lam, C., and P. C. L. Souza (2020): "Estimation and selection of spatial weight matrix in a spatial lag model," *Journal of Business & Economic Statistics*, 38, 693–710, 10.1080/07350015.2019.1569526.

Lesage, J. P., and M. M. Fischer (2008): "Spatial growth regressions: Model specification, estimation and interpretation," *Spatial Economic Analysis*, 3, 275–304, 10.1080/17421770802353758.

LeSage, J. P., and R. K. Pace (2014): "The biggest myth in spatial econometrics," *Econometrics*, 2, 217–249, 10.3390/econometrics2040217.

LeSage, J. P., and O. Parent (2007): "Bayesian model averaging for spatial econometric models," *Geographical Analysis*, 39, 241–267, 10.1111/j.1538-4632.2007.00703.x.

LeSage, J., and R. K. Pace (2009): *Introduction to spatial econometrics*: Chapman and Hall/CRC, 10.1201/9781420064254.

Lin, X. (2010): "Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables," *Journal of Labor Economics*, 28, 825–860, 10.1086/653506.

Macedo, M. N., R. S. DeFries, D. C. Morton, C. M. Stickler, G. L. Galford, and Y. E. Shimabukuro (2012): "Decoupling of deforestation and soy production in the Southern Amazon during the late 2000s," *Proceedings of the National Academy of Sciences*, 109, 1341–1346, 10.1073/pnas.1111374109.

Makalic, E., and D. F. Schmidt (2015): "A simple sampler for the horseshoe estimator," *IEEE Signal Processing Letters*, 23, 179–182, 10.1109/LSP.2015.2503725.

Mele, A. (2020): "Does school desegregation promote diverse interactions? an equilibrium model of segregation within schools," *American Economic Journal: Economic Policy*, 12, 228–57, 10.1257/pol.20170604.

Munshi, K. (2003): "Networks in the modern economy: Mexican migrants in the U.S. labor market," *Quarterly Journal of Economics*, 118, 549–599, 10.1162/003355303321675455.

Newman, M. E. J. (2001): "The structure of scientific collaboration networks," *Proceedings of the National Academy of Sciences*, 98, 404–409, 10.1073/pnas.98.2.404.

Park, T., and G. Casella (2008): "The Bayesian lasso," *Journal of the American Statistical Association*, 103, 681–686, 10.1198/016214508000000337.

van der Pas, S. L., B. J. K. Kleijn, and A. W. van der Vaart (2014): "The horseshoe estimator: Posterior concentration around nearly black vectors," *Electronic Journal of Statistics*, 8, 2585–2618, 10.1214/14-EJS962.

van der Pas, S., B. Szabó, and A. van der Vaart (2017): "Uncertainty quantification for the Horseshoe (with discussion)," *Bayesian Analysis*, 12, 1221–1274, 10.1214/17-BA1065.

Qu, X., and L. Lee (2015): "Estimating a spatial autoregressive model with an endogenous spatial weight matrix," *Journal of Econometrics*, 184, 209–232, 10.1016/j.jeconom.2014.08.008.

Rose, A. K. (2004): "Do we really know that the WTO increases trade?" *American Economic Review*, 94, 98–114, 10.1257/000282804322970724.

Self, S. G., and K.-Y. Liang (1987): "Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions," *Journal of the American Statistical Association*, 82, 605–610, 10.1080/01621459.1987.10478472.

Villoria, N., R. Garrett, F. Gollnow, and K. Carlson (2022): "Leakage does not fully offset soy supply-chain efforts to reduce deforestation in Brazil," *Nature Communications*, 13, 1–10, 10.1038/s41467-022-33213-z.

Weidmann, B., and D. J. Deming (2021): "Team players: How social skills improve team performance," *Econometrica*, 89, 2637–2657, 10.3982/ECTA18461.

Zhang, X., and J. Yu (2018): "Spatial weights matrix selection and model averaging for spatial autoregressive models," *Journal of Econometrics*, 203, 1–18, 10.1016/j.jeconom.2017.05.021.

A. Supplementary information

Intuition and interpretation of spillover effects

The general spatial model gives rise to three spatial effects that are visualized in Figure A1. There may be local spillover effects, affecting direct neighbors, which arise from the lagged explanatories, $\mathbf{WX}\boldsymbol{\theta}$. The autoregressive term, $\lambda\mathbf{Wy}$, allows the responses of all units to be influenced by the response's of their neighbors. The observed equilibrium results from these global spillover effects, which spread across all connected units. Lastly, the error term may exhibit spatial autocorrelation due to $\rho\mathbf{We}$.

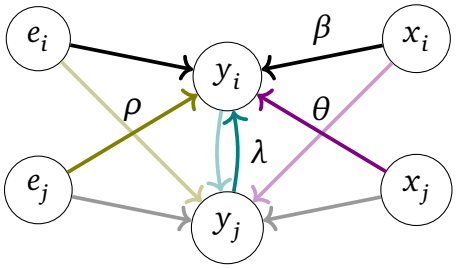


FIGURE A1: Illustration of a unit i receiving direct effects from itself, and indirect effects from other units $j \neq i$ in a general model.

To illustrate this further, consider a field experiment investigating the effects of certain treatments on crop yields, as stylized in Figure A2. If we apply fertilizer to field 'C2', we may find local spillover effects to neighboring fields. This effect may occur directly ($x_i \rightarrow y_j$), e.g., when roots surpass field boundaries, or via a third mediator ($x_i \rightarrow z_j \rightarrow y_j$) when the treatment is dispersed and affects the (unmeasured) 'treatment' of neighbors, e.g. when fertilizer is washed away by rain. Yields themselves are behind global spillovers, which may occur, e.g., if wheat grows too large for its stalks and knocks others over, or if there is a form of communication to prompt or dissuade growth.[9] We will encounter spatial autocorrelation if there are third factors that feature a spatial pattern. This could be a hedge blocking sunlight and wind, the topography, or a plethora of other factors.

Interpretation

The interpretation of spillover effects in the general spatial econometric model is not entirely straightforward, and depends on the spatial lags involved. Firstly, the lag of the error term does not directly convey the presence of spillovers, but implies correlation

[9] Such mechanisms can be beneficial for the fitness of a plant. Trees that grow too densely undermine their overall fitness, while a crop that is flowering successfully can indicate favorable conditions to others.

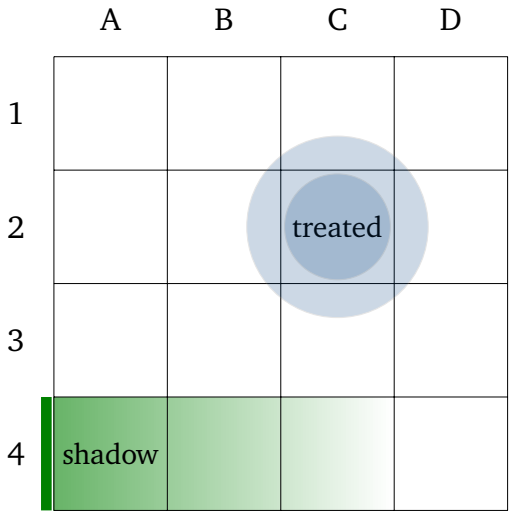


FIGURE A2: Stylised illustration of a grid of fields for analysing the effects of fertiliser on crop yield. The field in 'C2' is treated (e.g. with fertiliser), but spillover effects of the treatment (e.g. via dissolution) impact neighbouring fields. Meanwhile, a hedge to the left of field 'A4' causes correlation between it and its neighbours by throwing shade.

across observations (see Barrios et al., 2012, for a comparison with clustering). Next, the spatially lagged explanatory variables represent 'local' spillovers from the weighted characteristics of direct neighbors (see Halleck Vega and Elhorst, 2015, for further discussion). Lastly, the endogenous lag of the response variable represents 'global' spillovers across all neighbors. It is arguably the most widely used, but also the most controversial term (see Angrist, 2014; Corrado and Fingleton, 2012), and is best understood as the reflection of an equilibrium response.

In the standard spatial econometric model, the free parameters can only be interpreted conditional on the connectivity structure. This circumstance is a common source of errors and misconceptions (LeSage and Pace, 2014; Kuschnig, 2022), and holds similarly for the more complex models considered in this paper. Most of these issues stem from a misinterpretation of the partial effects. For the model in Equation 1, the partial effects of a variable j are given by

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}_k} = \mathbf{S}(\lambda \mid \mathbf{W})^{-1} \begin{pmatrix} \beta_k & w_{12}\theta_j & \dots & w_{1N}\theta_k \\ w_{21}\theta_k & \beta_k & \dots & w_{2N}\theta_k \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1}\theta_k & w_{N2}\theta_k & \dots & \beta_k \end{pmatrix}. \tag{4}$$

This matrix of partial effects is not straightforward to interpret, and summary measures are used instead. An obvious example is the average partial effect, for which we divide the sum of all elements in Equation 4 by n. LeSage and Pace (2009) consider two summary measures that offer insights into the spatial structure — the average direct effect (considering only diagonal elements), and the average indirect effect (considering

only off-diagonal elements).

Horseshoe prior for the nested linear model

For the nested linear model, there is a variety of established priors to choose from, including the standard Normal Inverse-Gamma prior or a number of shrinkage priors (e.g. Bhattacharya et al., 2015; Griffin and Brown, 2010; Park and Casella, 2008, etc.). We will consider the Horseshoe prior by Carvalho et al. (2010) for its proven performance and efficient sampling (see Makalic and Schmidt, 2015), and an inverse Gamma prior for the variance. That is

$$\begin{aligned} \beta_i \mid \iota_i \sim \mathrm{N}(0, \iota_i), \qquad \iota_i \mid \zeta \sim \mathrm{C}^+(0, \zeta), \qquad \zeta \mid s \sim \mathrm{C}^+(0, \varsigma), \\ \sigma^2 \sim \mathrm{G}^{-1}(a, b), \end{aligned}$$

where ι_i are coefficient-specific hyperparameters with common scale t, and $\mathrm{C}^+(0, a)$ denotes the density of the half-Cauchy distribution with scale a, given by

$$\mathrm{C}^+(x \mid 0, a) = \begin{cases} \frac{2}{\pi a}\left(1 + \frac{x^2}{a^2}\right)^{-1}, & \text{if } x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

The Horseshoe prior offers flexible shrinkage, accommodating sparsity via sharp spikes at the origin without distorting estimates, thanks to its heavy tails. The parameters facilitate global regularization, but also allow for local shrinkage of individual coefficients. As a result, the Horseshoe is a standard choice in the literature (also see Datta and Ghosh, 2013; van der Pas et al., 2014, 2017), allowing us to focus on connectivity-related parameters.

B. Supplementary Material

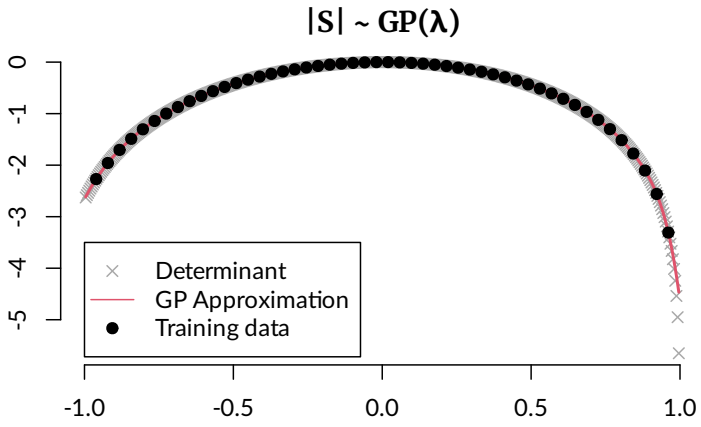


FIGURE B3: One-dimensional Gaussian process approximation to the Jacobian determinant $|\mathbf{S}(\lambda)|$ (on the vertical axis) using 50 training samples.

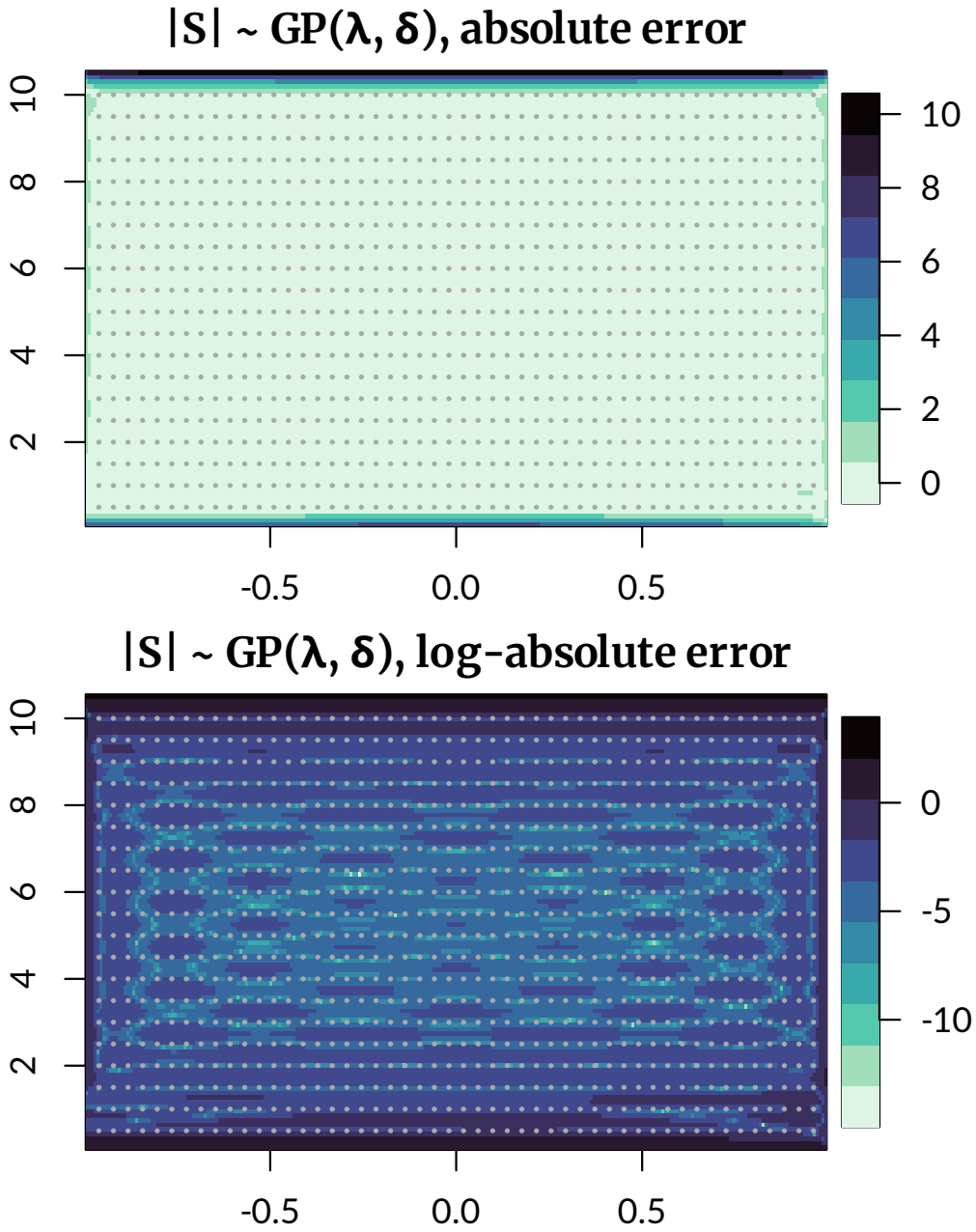


FIGURE B4: Absolute and log-absolute error of the Gaussian process approximation for the (two-dimensional) Jacobian determinant $|\mathbf{S}(\lambda, \delta)|$ using a 50×20 training grid. Distances are between $n = 100$ locations with coordinates sampled from a Uniform distribution. Gray dots indicate training samples.