A hierarchical Bayesian logit model for spatial multivariate choice data

Oyama, Yuki; Murakami, Daisuke; Krueger, Rico

Published in:
Journal of Choice Modelling

Link to article, DOI:
10.1016/j.jocm.2024.100503

Publication date:
2024

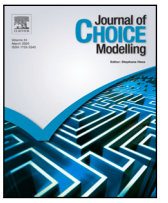
Document Version
Publisher's PDF, also known as Version of record

Journal of Choice Modelling 52 (2024) 100503

Contents lists available at ScienceDirect

Journal of Choice Modelling

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



A hierarchical Bayesian logit model for spatial multivariate choice data

Yuki Oyama [a,*], Daisuke Murakami [b], Rico Krueger [c]

[a] Department of Civil Engineering, The University of Tokyo, Tokyo, Japan
[b] Department of Statistical Data Science, The Institute of Statistical Mathematics, Tokyo, Japan
[c] Department of Technology, Management and Economics, Technical University of Denmark, Kongens Lyngby, Denmark

ARTICLE INFO

Dataset link: https://github.com/yuki-oyama/bayes-sl

Keywords:
Spatial choice data
Spatial autoregressive model
Mixed logit
Bayesian estimation
Pólya-Gamma data augmentation
Neighborhood perception

ABSTRACT

Spatial perceptions mediate human–environment interaction, and understanding spatial perceptions of humans can play a key role in the planning of activities. This study aims to analyze spatial multivariate binary choice data representing if an individual perceives a spatial unit to belong to a certain category (*e.g.*, her neighborhood or set of potential activity places). To reasonably analyze such data, we present a spatial autoregressive mixed logit (SAR-MXL) model that accounts for both inter-individual heterogeneity and spatial dependence. We rely on the Bayesian approach for posterior inference of model parameters, where Pólya-Gamma data augmentation (PG-DA) is adopted to address the non-conjugacy of the logit kernel. The PG-DA technique eliminates the need for the Metropolis–Hastings step during the Markov Chain Monte Carlo process and allows for fast and efficient posterior inference. The high efficiency of the Bayesian SAR-MXL model is demonstrated through a numerical experiment. The proposed framework is applied to street-based neighborhood perception data, and we empirically analyzed the factors associated with the street perception probability of individuals. The result suggests a clear improvement of the model fit by incorporating spatial dependence and random parameters.

1. Introduction

Spatial perceptions mediate human–environment interaction, understanding spatial perceptions of humans plays a key role in the planning of activities in cities (Lynch, 1960; Downs and Stea, 1973; Golledge et al., 2000), e.g., how people perceive the boundary of a neighborhood or how people compose the set of potential activity places for daily living or travel. These spatial perceptions of individuals can be described as spatial multivariate binary choice responses, typically represented by a two-dimensional vector $\boldsymbol{y} = [y_{ni}]$ where y_{ni} takes one if individual $n \in \{1, \ldots, N\}$ perceives spatial unit $i \in \{1, \ldots, S\}$ and zero otherwise. For spatial perception analysis, spatial units are often defined in a way that is not consistent with how individuals perceive the division of space. As a result, spatial multivariate choice data may exhibit spatial dependence due to systematic correlation in unobserved factors. It is essential to capture spatial dependence and to evaluate not only the direct effects but also potential spillover effects on individuals' perception resulting from changes in the attributes of spatial units. In addition, because perception is subjective, sensitivity to changes in attributes is likely to vary across individuals. The analysis of spatial multivariate choice data thus requires the consideration of (1) spatial dependence and (2) inter-individual heterogeneity.

* Corresponding author.
E-mail address: oyama@bin.t.u-tokyo.ac.jp (Y. Oyama).

https://doi.org/10.1016/j.jocm.2024.100503
Received 11 December 2023; Received in revised form 16 April 2024; Accepted 28 May 2024
Available online 7 June 2024

Several advanced spatial choice models that capture unobserved heterogeneity and spatial dependence are presented in the literature (e.g., Bhat and Guo, 2004; Miyamoto et al., 2004; Sener et al., 2011). However, most of them deal with a single response per each individual and cannot directly be used for the analysis of spatial multivariate choice data. Bhat (2011)'s maximum approximate composite marginal likelihood (MACML) method can be extended to deal with spatial multivariate choice data, as in Bhat (2015) and Mondal and Bhat (2022). However, the MACML method is mainly designed for models whose full likelihood function is infeasible to evaluate in closed form, such as multinomial probit (MNP) models, while our study deals with a closed-form logit model that has a tractable likelihood function.

Moreover, the maximum likelihood (ML) approach, including MACML (Bhat, 2011; Mondal and Bhat, 2022) and maximum simulated likelihood (MSL) estimation (e.g., Train, 2009), is the predominant estimation strategy for the existing spatial choice models. While the ML approach estimates only population parameters, the Bayesian approach specifies the complete conditional posterior distributions for all parameters in the model. The Bayesian approach demonstrates comparable stability to different sample sizes (Calabrese and Elkink, 2014) and high extensibility (LeSage, 1997, 2000) over other estimation methods in spatial econometrics, including ML (McMillen, 1992), generalized method of moments (GMM) (Pinkse and Slade, 1998), linearized GMM (LGMM) (Klier and McMillen, 2008), and two-step GMM procedures (Piras and Sarrias, 2023). Focusing on this flexible extensibility, this study incorporates random parameters into the Bayesian spatial autoregressive (SAR) logit model in a straightforward manner.

However, the Bayesian approach can be computationally expensive and inefficient in terms of effective sample size for the spatial logit model. Due to the non-conjugacy, i.e., the unavailability of conjugate prior for the logit kernel, the posterior sampling of the model parameters during the Markov Chain Monte Carlo (MCMC) process needs the help of the Metropolis–Hastings (MH) algorithm which is slower than direct sampling and can involve autocorrelated samples leading to a low effective sample size. To address this problem, Polson et al. (2013) recently presents the Pólya-Gamma data augmentation (PG-DA) technique for binary logit models. PG-DA augments the Gibbs sampler by introducing an additional PG distributed latent variable, which ensures conjugate updates and thus does not need an inefficient MH algorithm within the MCMC method.

In this study, we present a hierarchical Bayesian SAR mixed logit (SAR-MXL) model to accommodate inter-individual heterogeneity and spatial dependence in spatial multivariate binary choice data. To address the non-conjugacy of the model, we rely on PG-DA-based Bayesian estimation (Polson et al., 2013). While the PG-DA technique is successfully incorporated into spatial binary choice data (Krisztin and Piribauer, 2021; Krisztin et al., 2022) and spatial count data (Bansal et al., 2021; Buddhavarapu et al., 2016, 2021; Gurumurthy et al., 2022; Krueger et al., 2020a), we add to the literature by deriving the conjugate posterior distributions for fixed and individual-specific random parameters in the multivariate binary choice setting where both unobserved taste heterogeneity and spatial dependence are considered. The efficiency of this hierarchical Bayesian MCMC with PG-DA for the SAR-MXL model will be compared against the MH algorithm through numerical experiments. It should be noted that while our Bayesian SAR-MXL model can be used to analyze traditional multivariate choice data (e.g., Edwards and Allenby, 2003) as well, its application is limited to binary choice responses due to the effectiveness of PG-DA in binary choice settings (Bansal et al., 2019).

Using the proposed SAR-MXL framework, we empirically analyze the subjective neighborhood perceptions of individuals. Neighborhood perception is studied in the field of behavioral geography and urban studies (Coulton et al., 2001; Smith et al., 2010). Guimpert and Hurtubia (2018) and Oyama (2024) decompose neighborhood perception into binary choice responses of individuals perceiving spatial units such as meshes or street segments. By estimating binary logit models, these studies quantitatively analyze factors forming the perceived neighborhoods of individual respondents. Guimpert and Hurtubia (2018) also capture unobserved heterogeneity in distance sensitivity. However, these previous analyses treat the spatial units independently and evaluate only the direct effects of changes in policy variables. The definition of spatial perception units may be inconsistent with how individuals perceive the division of space. Therefore, treating spatial units independently ignores the correlation in unobserved factors and leads to bias in model accuracy and policy evaluation. The present study contributes to the literature by considering spatial dependence as well as inter-individual heterogeneity using the SAR-MXL model, which allows for the improvement in prediction performance and the analysis of direct and indirect impacts of changes in policy variables.

Specifically, we use the data indicating if an individual perceives a street segment to belong to a specific neighborhood, which was collected by Oyama (2024) through on-street surveys in two urban neighborhoods in Tokyo. Given that the streets compose a network, we consider the spatial dependence using different network-based spatial weight matrices. This application reveals the formative factors of respondents' subjective spatial city image and the direct/indirect impacts of changes in the attributes of streets, which can help policies related to the image building of the neighborhood.

The remainder of the paper is structured as follows. Section 2 presents the Bayesian SAR-MXL model. Section 3 describes the Bayesian estimation strategy. Section 4 shows a numerical experiment, and Section 5 presents the empirical application to spatial perception analysis. Section 6 discusses the results. Finally, Section 7 concludes the study.

2. Model formulation

2.1. Spatial autoregressive mixed logit model

This study analyzes spatial multivariate binary choice data, where multiple responses to spatial units of an individual are observed. Let $y_{ni} \in \{0,1\}$ be a binary response of an individual $n \in \{1,\dots,N\}$ for a spatial unit $i \in \{1,\dots,S\}$. We model the probability of y_{ni} being one based on the binary logit framework:

$$P(y_{ni}=1) = \frac{\exp(v_{ni})}{1+\exp(v_{ni})} \tag{2.1}$$

with the utility function $u_{ni} = v_{ni} + \eta_{ni}$ of space i perceived by individual n, whereby v_{ni} is the systematic component and η_{ni} is the error component following an independently and identically distributed (i.i.d.) Gumbel distribution with unit scale.

To reasonably analyze such subjectively perceived spatial data, we introduce the assumptions of spatial dependence and inter-individual heterogeneity. This results in the formulation of a *spatial autoregressive mixed logit* (SAR-MXL) model. The utility v_{ni} is then defined as

$$v_{ni} = \boldsymbol{\alpha}^\top \boldsymbol{x}_{ni}^F + \boldsymbol{\beta}_n^\top \boldsymbol{x}_{ni}^R + \rho \sum_{j \neq i} w_{ij} v_{nj} + \epsilon_{ni} \tag{2.2}$$

wherein we distinguish observed attributes of individual n and spatial unit i between $\boldsymbol{x}_{ni}^F \in \mathbb{R}^{K_F}$ and $\boldsymbol{x}_{ni}^R \in \mathbb{R}^{K_R}$ that pertain the fixed taste parameters $\boldsymbol{\alpha} \in \mathbb{R}^{K_F}$ and random taste parameters $\boldsymbol{\beta}_n \in \mathbb{R}^{K_R}$, respectively. The latter captures the inter-personal heterogeneity in the perception. A nonnegative and row-stochastic spatial weight matrix $\mathbf{W} = [w_{ij}]$ with zero-diagonals contains information on the $S \times S$ linkages between the spatial units in the sample, and ρ is a scalar parameter[1] measuring the strength of spatial autocorrelation. ϵ_{ni} is a disturbance term following an i.i.d. Gaussian distribution with zero mean and unit variance (LeSage et al., 2011; Krisztin and Piribauer, 2021).

The reduced form of (2.2) shows the spatial dependence of the utilities more clearly:

$$\boldsymbol{v}_n = \mathbf{A}^{-1}(\mathbf{X}_n^F \boldsymbol{\alpha} + \mathbf{X}_n^R \boldsymbol{\beta}_n + \boldsymbol{\epsilon}_n) = \mathbf{A}^{-1}(\mathbf{X}_n^F \boldsymbol{\alpha} + \mathbf{X}_n^R \boldsymbol{\beta}_n) + \mathbf{A}^{-1}\boldsymbol{\epsilon}_n, \tag{2.3}$$

where

$$\mathbf{A} = \mathbf{I}_S - \rho \mathbf{W}, \tag{2.4}$$

and $\mathbf{I}_S$ is an $S \times S$ identity matrix. $\mathbf{X}_n^F = (\boldsymbol{x}_{ni}^F, \boldsymbol{x}_{n2}^F, \dots, \boldsymbol{x}_{nS}^F)$ and $\mathbf{X}_n^R = (\boldsymbol{x}_{ni}^R, \boldsymbol{x}_{n2}^R, \dots, \boldsymbol{x}_{nS}^R)$ are $S \times K_F$ and $S \times K_R$ matrices, respectively. In this SAR-MXL model specification, the utility v_{ni} of individual n for spatial unit i depends on the characteristics of other spatial units $j \neq i$, and the spatial multiplier $\mathbf{A}^{-1} = (\mathbf{I} - \rho\mathbf{W})^{-1} = \sum_{r=0}^{\infty} \rho^r \mathbf{W}^r$ captures the spatial dependence.[2] The disturbance $\boldsymbol{\epsilon}_n = (\epsilon_{n1}, \epsilon_{n2}, \dots, \epsilon_{nS})$ captures the spatial dependence in the errors, whereas $\boldsymbol{\eta}_n = (\eta_{n1}, \eta_{n2}, \dots, \eta_{nS})$ associated with the logit formulation (2.1) are spatially independent.

The model parameters to be estimated are $\boldsymbol{\alpha}$, $\boldsymbol{\beta}_{1:N}$, and ρ, where $\boldsymbol{\beta}_{1:N} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_N)$ is an array of the random parameters over individuals. The likelihood function of the SAR-MXL model is

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}_{1:N}, \rho) = \prod_{n=1}^{N} \prod_{i=1}^{S} \frac{\exp(v_{ni})^{y_{ni}}}{1 + \exp(v_{ni})}. \tag{2.5}$$

2.2. Effect size

In the SAR framework, the change in an attribute at a spatial unit can have both *direct* and *indirect* impacts on the probability change. The indirect impacts describe the spillover effects of an attribute change. Let $\boldsymbol{\Lambda}_{n,k}$ denote a $S \times S$ matrix whose element $e_{ij}^{n,k}$ represents the effect size of kth variable x_{njk} for individual n and spatial unit j on the probability $P(y_{ni} = 1)$ of the same individual n perceiving spatial unit i. We define the *direct impact* (DI), *indirect impact* (II), and *total impact* (TI) as follows:

$$\mathrm{DI}_{n,k} = \frac{1}{S} \mathbf{1}_S \mathrm{diag}(\boldsymbol{\Lambda}_{n,k}) \tag{2.6}$$

$$\mathrm{TI}_{n,k} = \frac{1}{S} \mathbf{1}_S^\top \boldsymbol{\Lambda}_{n,k} \mathbf{1}_S \tag{2.7}$$

$$\mathrm{II}_{n,k} = \mathrm{TI}_{n,k} - \mathrm{DI}_{n,k} \tag{2.8}$$

where $\mathbf{1}_S$ is a $S \times 1$ vector of ones. Then, the average impacts across individuals are

$$\mathrm{ADI}_k = \frac{1}{N} \sum_{n=1}^{N} \mathrm{DI}_{n,k} \tag{2.9}$$

$$\mathrm{AII}_k = \frac{1}{N} \sum_{n=1}^{N} \mathrm{II}_{n,k}. \tag{2.10}$$

In this study, $e_{ij}^{n,k}$ can be either elasticity or marginal effect. Elasticity $\mathrm{EL}_{ij}^{n,k}$ measures the percentage change in probability $P_{ni} \equiv P(y_{ni} = 1)$ with respect to a given percentage change in attribute x_{njk}, and marginal effect $\mathrm{ME}_{ij}^{n,k}$ measures the absolute change in P_{ni} with respect to a unit change x_{njk}. When focusing on an attribute x_{njk} that pertains to a random parameter $\beta_{n,k}$, these effect sizes are defined as:

$$\mathrm{EL}_{ij}^{n,k} = \frac{\partial P_{ni}}{\partial x_{njk}} \frac{x_{njk}}{P_{ni}} = (1 - P_{ni}) x_{njk} \bar{a}_{ij} \beta_{n,k} \tag{2.11}$$

[1] This study focuses on inter-individual heterogeneity in taste parameters, and the spatial pattern is assumed for simplicity to be common among individuals. However, the spatial parameter ρ can be extended to be heterogeneous across individuals similarly to LeSage and Chih (2018).

[2] For the stability of Eq. (2.3), ρ can take on feasible values in the range $1/\lambda_{\min} < \rho < 1/\lambda_{\max}$, where $\lambda_{\min}$ and $\lambda_{\max}$ are the minimum and maximum eigenvalues of the spatial matrix $\mathbf{W}$. This condition is the same as the normal SAR model that does not consider individual heterogeneity (Anselin and Florax, 1995).

$$\mathrm{ME}_{ij}^{n,k} = \frac{\partial P_{ni}}{\partial x_{njk}} = P_{ni}(1 - P_{ni})\bar{a}_{ij}\beta_{n,k} \tag{2.12}$$

where $\bar{a}_{ij}$ is the element of ith row and jth column of matrix $\mathbf{A}^{-1}$. We calculate elasticities for continuous variables and marginal effects for dummy variables.

3. Bayesian estimation

3.1. Bayesian specification

Following the standard MXL model formulation, we assume the random (individual-specific) parameter $\boldsymbol{\beta}_n$ to be distributed according to a multivariate normal, i.e., $\boldsymbol{\beta}_n \sim \mathcal{N}(\boldsymbol{\zeta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\zeta}$ is a K_R-dimensional mean vector and $\boldsymbol{\Sigma}$ is a $K_R \times K_R$ covariance matrix.[3] The remaining parameters, $\boldsymbol{\alpha}, \boldsymbol{\zeta}, \boldsymbol{\Sigma}$ and ρ are treated in a fully-Bayesian manner. The fixed parameters $\boldsymbol{\alpha}$ and the mean vector $\boldsymbol{\zeta}$ are assumed to follow multivariate normal distributions. As for the covariance matrix $\boldsymbol{\Sigma}$, we employ Huang's half-t prior (Huang and Wand, 2013), which Akinc and Vandebroek (2018) show outperforms the inverse Wishart prior that is often employed in fully Bayesian MXL models (Train, 2009) in terms of parameter recovery. We also employ a beta distribution for the prior of spatial parameter ρ, following the literature (LeSage and Pace, 2009).

The generative process of the fully Bayesian SAR-MXL model is summarized as follows:

- $\rho | a_\rho, b_\rho \sim \mathrm{Beta}(a_\rho, b_\rho)$
- $\boldsymbol{\alpha} | \lambda_0, \Xi_0 \sim \mathcal{N}(\lambda_0, \Xi_0)$
- $\boldsymbol{\zeta} | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0 \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$
- $a_k | A_k \sim \mathrm{Gamma}(\frac{1}{2}, \frac{1}{A_k^2})$, $k = 1, \ldots, K_R$
- $\boldsymbol{\Sigma} | \nu, \boldsymbol{a} \sim \mathrm{IW}(\nu + K_R - 1, 2\nu \mathrm{diag}(\boldsymbol{a}))$
- $\boldsymbol{\beta}_n | \boldsymbol{\zeta}, \boldsymbol{\Sigma} \sim \mathcal{N}(\boldsymbol{\zeta}, \boldsymbol{\Sigma})$
- $y_{ni} | \boldsymbol{\alpha}, \boldsymbol{\beta}_n, \rho, \mathbf{X}_n \sim \mathrm{Logit}(\boldsymbol{\alpha}, \boldsymbol{\beta}_n, \rho, \mathbf{X}_n)$

We have known hyperparameters $\{a_\rho, b_\rho, \lambda_0, \Xi_0, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \nu, A_{1:K}\}$ and a collection of model parameters $\theta = (\boldsymbol{\alpha}, \boldsymbol{\zeta}, \boldsymbol{\Sigma}, \boldsymbol{a}, \boldsymbol{\beta}_{1:N}, \rho)$ whose posterior distribution to be estimated.

3.2. Pólya-Gamma data augmentation

The main challenge of the standard Gibbs sampler for posterior inference with the logit model is the unavailability of a conjugate prior for the kernel, which requires a computationally inefficient MH method. To address this challenge, we use a Pólya-Gamma data augmentation (PG-DA) estimation strategy proposed by Polson et al. (2013). PG-DA is the state-of-the-art technique to handle non-conjugacy in MCMC estimation of binary logit models by introducing a PG-distributed latent variable.

Following Polson et al. (2013), the Laplace transform of the PG density is given by

$$\frac{\exp(v_{ni})^a}{(1 + \exp(v_{ni}))^b} = 2^{-b} \exp(\kappa v_{ni}) \int_0^\infty \exp\left(-\frac{\omega_{ni} v_{ni}^2}{2}\right) p(\omega_{ni}) d\omega_{ni} \tag{3.1}$$

where $\kappa = a - b/2$, and ω_{ni} is a PG distributed random variable with scale b and location parameter zero, i.e., $p(\omega_{ni}) \sim \mathcal{PG}(b, 0)$.

Note that this integral identity does not rely on numerical approximation and can be estimated by sampling from the conditional posterior PG distribution. Using the result in Polson et al. (2013), the conditional posterior for $\boldsymbol{\omega}$ also takes the form of a PG distribution:

$$p(\boldsymbol{\omega}_n | \boldsymbol{\alpha}, \boldsymbol{\beta}_n, \rho, \boldsymbol{y}_n) = \mathcal{PG}\left(1, \mathbf{A}^{-1}(\mathbf{X}_n^F \boldsymbol{\alpha} + \mathbf{X}_n^R \boldsymbol{\beta}_n + \boldsymbol{\epsilon}_n)\right) \tag{3.2}$$

3.3. Conditional posterior distributions

We notice that when $a = y_{ni}$ and $b = 1$, Eq. (3.1) coincides with the likelihood of the logit model (2.1):

$$p(y_{ni} | \boldsymbol{\beta}_n, \boldsymbol{\alpha}, \rho) = \frac{\exp(v_{ni})^{y_{ni}}}{1 + \exp(v_{ni})} = \frac{1}{2} \exp(\kappa_{ni} v_{ni}) \int_0^\infty \exp\left(-\frac{\omega_{ni} v_{ni}^2}{2}\right) p(\omega_{ni}) d\omega_{ni}, \tag{3.3}$$

and the conditional distribution of y_{ni} on data point ω_{ni} is

$$p(y_{ni} | \boldsymbol{\beta}_n, \boldsymbol{\alpha}, \rho, \boldsymbol{\omega}_{ni}) = \frac{1}{2} \exp\left(\kappa_{ni} v_{ni} - \frac{\omega_{ni} v_{ni}^2}{2}\right). \tag{3.4}$$

[3] Semi-parametric representation of unobserved heterogeneity can be a more flexible alternative to the parametric distribution (e.g., Krueger et al., 2020b), but we leave its integration to future work.

Using this result, the conditional posterior density of $\boldsymbol{\alpha}$ is given by

$$p(\boldsymbol{\alpha}|\boldsymbol{y}_{1:N}, \boldsymbol{\beta}_{1:N}, \rho, \boldsymbol{\omega}_{1:N}) \propto p(\boldsymbol{\alpha}|\lambda_0, \Xi_0) \prod_{n=1}^{N} \prod_{i=1}^{S} p(y_{ni}|\boldsymbol{\beta}_n, \boldsymbol{\alpha}, \rho, \boldsymbol{\omega}_{ni})$$

$$\propto \exp\left(-\frac{1}{2}(\boldsymbol{\alpha} - \lambda_0)' \Xi_0 (\boldsymbol{\alpha} - \lambda_0)\right) \prod_{n=1}^{N} \exp\left(\kappa_n \mathbf{A}^{-1} \mathbf{X}_n^F \boldsymbol{\alpha} - \frac{\boldsymbol{\omega}_n \boldsymbol{v}_n^2}{2}\right), \tag{3.5}$$

and similarly, the conditional posterior density of $\boldsymbol{\beta}_n$ is given by

$$p(\boldsymbol{\beta}_n|\boldsymbol{y}_n, \boldsymbol{\alpha}, \rho, \boldsymbol{\omega}_n) \propto p(\boldsymbol{\beta}_n|\zeta, \Sigma) \prod_{i=1}^{S} p(y_{ni}|\boldsymbol{\beta}_n, \boldsymbol{\alpha}, \rho, \boldsymbol{\omega}_{ni}) \tag{3.6}$$

$$\propto \exp\left(-\frac{1}{2}(\boldsymbol{\beta}_n - \zeta)' \Sigma (\boldsymbol{\beta}_n - \zeta)\right) \exp\left(\kappa_n \mathbf{A}^{-1} \mathbf{X}_n^R \boldsymbol{\beta}_n - \frac{\boldsymbol{\omega}_n \boldsymbol{v}_n^2}{2}\right). \tag{3.7}$$

As a result, the conditional posterior distributions of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_n$ reduce to the following Gaussian distributions:

$$\boldsymbol{\alpha}|\boldsymbol{y}_{1:N}, \boldsymbol{\beta}_{1:N}, \rho, \boldsymbol{\omega}_{1:N} \sim \mathcal{N}(\boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\alpha) \tag{3.8}$$

$$\boldsymbol{\beta}_n|\boldsymbol{y}_{1:N}, \boldsymbol{\alpha}, \rho, \boldsymbol{\omega}_n \sim \mathcal{N}(\boldsymbol{\mu}_{\beta_n}, \boldsymbol{\Sigma}_{\beta_n}) \tag{3.9}$$

where

$$\boldsymbol{\mu}_\alpha = \boldsymbol{\Sigma}_\alpha \left(\Xi_0^{-1} \lambda_0 + \sum_{n=1}^{N} (\mathbf{A}^{-1} \mathbf{X}_n^F)^\top \boldsymbol{\Omega}_n (\boldsymbol{z}_n - \mathbf{A}^{-1} \mathbf{X}_n^R \boldsymbol{\beta}_n - \mathbf{A}^{-1} \boldsymbol{\epsilon}_n)\right) \tag{3.10}$$

$$\boldsymbol{\Sigma}_\alpha = \left(\Xi_0^{-1} + \sum_{n=1}^{N} (\mathbf{A}^{-1} \mathbf{X}_n^F)^\top \boldsymbol{\Omega}_n (\mathbf{A}^{-1} \mathbf{X}_n^F)\right)^{-1} \tag{3.11}$$

$$\boldsymbol{\mu}_{\beta_n} = \boldsymbol{\Sigma}_{\beta_n} \left(\Sigma^{-1} \zeta + (\mathbf{A}^{-1} \mathbf{X}_n^R)^\top \boldsymbol{\Omega}_n (\boldsymbol{z}_n - \mathbf{A}^{-1} \mathbf{X}_n^F \boldsymbol{\alpha} - \mathbf{A}^{-1} \boldsymbol{\epsilon}_n)\right) \tag{3.12}$$

$$\boldsymbol{\Sigma}_{\beta_n} = \left(\Sigma^{-1} + (\mathbf{A}^{-1} \mathbf{X}_n^R)^\top \boldsymbol{\Omega}_n (\mathbf{A}^{-1} \mathbf{X}_n^R)\right)^{-1} \tag{3.13}$$

$$\boldsymbol{\Omega}_n = \mathrm{diag}(\omega_{n1}, \ldots, \omega_{nS}) \tag{3.14}$$

$$\boldsymbol{z}_n = (\kappa_{n1}/\omega_{n1}, \ldots, \kappa_{nS}/\omega_{nS})^\top \tag{3.15}$$

Note that in the above derivation, our study combines the mixed logit case (Bansal et al., 2019) and the spatial binary logit case (Krisztin and Piribauer, 2021).

3.4. Augmented Gibbs sampler

The joint distribution of the model factorizes as

$$p(\boldsymbol{y}_{1:N}, \boldsymbol{\theta}) = p(\rho) p(\boldsymbol{\alpha}|\lambda_0, \Xi_0) p(\boldsymbol{\zeta}|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) p(\boldsymbol{\Sigma}|\boldsymbol{a}, \nu) \prod_{k=1}^{K_R} p(a_k|A_k) \prod_{n=1}^{N} p(\boldsymbol{\beta}_n|\boldsymbol{\zeta}, \boldsymbol{\Sigma}). \tag{3.16}$$

Then, the PG-augmented Gibbs sampler algorithm for the SAR-MXL model can be summarized as in Algorithm 1.

Algorithm 1 Pseudo-code of the Pólya-Gamma augmented Gibbs sampler for the SAR-MXL model

for (iteration 1 **to** M) **do**
  Update $\boldsymbol{a}$ by sampling $a_k \sim \mathrm{Gamma}(\frac{\nu + K_R}{2}, \frac{1}{A_k^2} + \nu(\Sigma^{-1})_{kk})$, $\forall k$
  Update $\boldsymbol{\Sigma}$ by sampling $\boldsymbol{\Sigma} \sim \mathrm{IW}(\nu + N + K_R - 1, 2\nu \mathrm{diag}(\boldsymbol{a}) + \sum_{n=1}^{N} (\boldsymbol{\beta}_n - \boldsymbol{\zeta})(\boldsymbol{\beta}_n - \boldsymbol{\zeta})^\top)$
  Update $\boldsymbol{\zeta}$ by sampling $\boldsymbol{\zeta} \sim \mathcal{N}(\frac{1}{N} \sum_{n=1}^{N} \boldsymbol{\beta}_n, \frac{\boldsymbol{\Sigma}}{N})$
  for $n = 1$ **to** N **do**
    Update $\boldsymbol{\omega}_n$ by sampling from the PG distribution of (3.2)
    Update $\boldsymbol{\beta}_n$ by sampling from the Gaussian distribution of (3.9)
  end for
  Update $\boldsymbol{\alpha}$ by sampling from the Gaussian distribution of (3.8)
  Update ρ using a griddy Gibbs sampler (Ritter and Tanner, 1992)
end for

This algorithm is executed for a sufficiently large number of iterations M, whereby the first M_0 iterations are discarded for burn-in. After burn-in, every fth draw is retained as thinning, resulting in $R = (M - M_0)/f$ draws retained as realizations from the posterior.

Note that because the conditional posterior for ρ does not have a conjugate distribution, a griddy Gibbs sampler (Ritter and Tanner, 1992) is used to sample from the conditional posterior, based on

$$p(\rho|\boldsymbol{y}_{1:N},\boldsymbol{\alpha},\boldsymbol{\beta}_{1:N}) = \frac{p(\rho)p(\boldsymbol{y}_{1:N}|\rho,\boldsymbol{\alpha},\boldsymbol{\beta}_{1:N})}{\int_{-1}^{1} p(\rho)p(\boldsymbol{y}_{1:N}|\rho,\boldsymbol{\alpha},\boldsymbol{\beta}_{1:N})d\rho} \approx \frac{p(\rho)p(\boldsymbol{y}_{1:N}|\rho,\boldsymbol{\alpha},\boldsymbol{\beta}_{1:N})}{\sum_{g=1}^{G} p(\rho_g)p(\boldsymbol{y}_{1:N}|\rho_g,\boldsymbol{\alpha},\boldsymbol{\beta}_{1:N})} \tag{3.17}$$

where ρ_g is the gth grid (discretized) point, i.e., $\rho_g = -1 + 2g/(G+1)$.[4]

4. Numerical experiment

This section presents numerical experiments to demonstrate the performance of the proposed method. The experiments use simulated observations and aim to assess (1) the reproducibility of the estimates parameters and effect sizes and (2) the efficiency of the method in comparison with the MH method that is usually adopted in MCMC for the logit model. The reproducibility is tested under different strengths ρ of spatial dependence. The efficiency is examined under different ρ as well as different numbers S of spatial units.

Note that our data involves two dimensions of spatial unit i and individual n. This is not typical for spatial statistics as spatial data does not generally include the information of individuals. Therefore, the existing methods or packages for the estimation of spatial models cannot be used for our framework, and we implemented the proposed model by writing our own Python code.

4.1. Simulation setup

The observations y_{ni} for the experiment are randomly sampled from the binomial distribution with the probability of the SAR-MXL model:

$$y_{ni} \sim \text{Bin}\left(1, \frac{\exp(v_{ni})}{1+\exp(v_{ni})}\right) \tag{4.1}$$

where $\boldsymbol{v}_n$ is defined by (2.2).

In this experiment, we consider 50 individuals ($N = 50$) and 200 spatial units ($S = 200$) in the default setting. We considered two parameters with fixed effects ($K_F = 2$) and two random parameters ($K_R = 2$), and their true values were set as follows:

- $\boldsymbol{\alpha}^* = (1.0, -1.0)$
- $\boldsymbol{\zeta}^* = (1.0, -1.0)$
- $\boldsymbol{\Sigma}^* = \begin{pmatrix} 0.5 & -0.2 \\ -0.2 & 0.5 \end{pmatrix}$

The explanatory variables $\mathbf{X}$ were sampled from the standard normal distribution and then normalized. The spatial weight matrix $\mathbf{W}$ was constructed using 4-nearest neighbors based on a spatial location pattern randomly generated from a uniform distribution $\mathcal{U}(0,1)$. During the MCMC estimation, we iterated Algorithm 1 for 10,000 times and the last 5000 draws were retained without thinning. As for the griddy Gibbs sampler for the conditional posterior for ρ, we used 100 grid points.

We test different strengths ρ of spatial dependence and examine how the estimation results and efficiency are affected. The efficiency is also evaluated for different numbers S of spatial units to discuss the method's applicability to different data sizes.

4.2. Results

4.2.1. Estimation results

Tables 1 and 2 report the parameter estimation results. Overall, the true parameter values were well reproduced, including the parameters of the random parameter distributions, regardless of the strength of spatial dependence in the generated observations. Even for the extreme cases ($\rho^* = 0.9$ and $\rho^* = -0.9$), the autocorrelation parameter ρ was estimated with small biases.

The estimates of the effect sizes are reported in Table 3, wherein the true values of the effect sizes are obtained by substituting the true parameter values $\boldsymbol{\alpha}^*$ and $\boldsymbol{\beta}^*_{1:N}$ into Eqs. (2.11) and (2.12). We tested the cases with $\rho^* \in \{0.0, 0.4, 0.8\}$ in this experiment. As well as the parameters, the present PG-DA method accurately estimated both the direct and indirect impacts of changes in the attributes. The sizes of indirect impacts increased as the strength ρ of spatial dependence increased.

These results show the stability of our Bayesian estimation method with respect to spatial dependence. However, the estimates were slightly biased when ρ^* is positive and large, compared to when ρ^* is small or negative. A possible explanation for this is that, with a high and positive autocorrelation, the observations can exhibit similar spatial patterns and the effective sample size (ESS) can be reduced.

[4] For the efficient computation, we compute and store $\mathbf{A}$ and its inverse $\mathbf{A}^{-1}$ for every grid point ρ_g before entering the MCMC loop.

Table 1
Means of posteriors with $\rho^* \in [0.0, 0.9]$.

	True	$\rho*$									
		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{\rho}$		−0.042	0.090	0.175	0.276	0.367	0.460	0.577	0.683	0.765	0.881
$\hat{\alpha}_1$	1	1.002	1.036	0.956	0.994	1.007	1.036	1.048	0.985	1.047	1.059
$\hat{\alpha}_2$	−1	−0.967	−1.036	−0.988	−0.924	−0.982	−1.053	−0.985	−1.023	−1.096	−1.120
$\hat{\zeta}_1$	1	1.048	0.918	1.002	0.985	0.982	1.100	1.142	0.983	1.095	1.106
$\hat{\zeta}_2$	−1	−0.946	−1.102	−1.066	−1.132	−0.960	−1.141	−1.029	−1.126	−1.021	−0.919
$\hat{\Sigma}_{11}$	0.5	0.482	0.542	0.489	0.574	0.497	0.588	0.508	0.502	0.538	0.546
$\hat{\Sigma}_{22}$	0.5	0.505	0.629	0.549	0.500	0.449	0.525	0.615	0.637	0.724	0.739
$\hat{\Sigma}_{12}$	−0.2	−0.228	−0.139	−0.261	−0.198	−0.197	−0.207	−0.144	−0.283	−0.191	−0.324

Table 2
Means of posteriors with $\rho^* \in [-0.9, -0.1]$.

	True	$\rho*$								
		−0.9	−0.8	−0.7	−0.6	−0.5	−0.4	−0.3	−0.2	−0.1
$\hat{\rho}$		−0.917	−0.812	−0.727	−0.624	−0.518	−0.384	−0.326	−0.225	−0.102
$\hat{\alpha}_1$	1	0.980	0.971	0.992	0.988	0.955	0.991	0.953	0.993	0.988
$\hat{\alpha}_2$	−1	−0.985	−1.000	−0.985	−0.955	−0.976	−0.968	−0.985	−0.995	−1.031
$\hat{\zeta}_1$	1	1.079	1.028	1.030	0.970	0.826	1.062	0.836	1.081	1.242
$\hat{\zeta}_2$	−1	−1.043	−0.864	−0.787	−1.030	−0.992	−0.989	−0.888	−0.969	−1.076
$\hat{\Sigma}_{11}$	0.5	0.415	0.420	0.516	0.418	0.475	0.335	0.550	0.731	0.519
$\hat{\Sigma}_{22}$	0.5	0.501	0.467	0.500	0.739	0.443	0.391	0.495	0.403	0.458
$\hat{\Sigma}_{12}$	−0.2	−0.173	−0.244	−0.264	−0.232	−0.150	−0.055	−0.155	−0.282	−0.134

Table 3
Mean of average direct and indirect impacts estimates with $\rho^* \in \{0.0, 0.4, 0.8\}$.

Attribute	Effect	$\rho^* = 0.0$		$\rho^* = 0.4$		$\rho^* = 0.8$	
		True	Est.	True	Est.	True	Est.
Elasticities							
x_1^F	Direct	−0.149	−0.154	−0.146	−0.129	−0.174	−0.179
	Indirect	0.000	0.000	−0.012	−0.011	−0.154	−0.132
x_2^F	Direct	−0.152	−0.141	−0.150	−0.150	−0.173	−0.187
	Indirect	0.000	0.000	−0.011	−0.011	−0.156	−0.142
x_1^R	Direct	−0.173	−0.178	−0.223	−0.235	−0.237	−0.250
	Indirect	0.000	0.000	−0.014	−0.015	−0.215	−0.196
x_2^R	Direct	−0.184	−0.189	−0.222	−0.212	−0.310	−0.342
	Indirect	0.000	0.000	−0.015	−0.014	−0.268	−0.251
Marginal effects							
x_1^F	Direct	0.148	0.150	0.143	0.135	0.120	0.121
	Indirect	0.000	−0.004	0.083	0.078	0.304	0.275
x_2^F	Direct	−0.148	−0.143	−0.143	−0.144	−0.120	−0.124
	Indirect	0.000	0.003	−0.083	−0.083	−0.304	−0.282
x_1^R	Direct	0.132	0.134	0.148	0.154	0.119	0.122
	Indirect	0.000	−0.003	0.086	0.088	0.303	0.276
x_2^R	Direct	−0.133	−0.131	−0.131	−0.130	−0.141	−0.148
	Indirect	0.000	0.003	−0.076	−0.075	−0.359	−0.336

4.2.2. *Efficiency*

We then performed the comparison between PG-DA and MH methods in terms of efficiency under different strengths ρ of spatial dependence, which Polson et al. (2013) do not consider. The proposal distribution was given by a zero-mean Gaussian distribution with variance manually adjusted so that the average acceptance rate becomes roughly 23.4%, known as the asymptotic optimal rate (Roberts and Rosenthal, 2001). For the comparison, we measured ESS as well as the Monte Carlo standard error (MCSE), which is defined by the standard deviation of the sampled ρ values divided by the ESS. A larger ESS indicates less autocorrelation in the chains, suggesting better/faster mixing. A smaller MCSE indicates better accuracy.

Table 4 summarizes the ESS and MCSE values when $\rho^* \in \{0.0, 0.4, 0.8\}$. In all the cases, PG-DA attained larger ESS values and smaller MCSE values. Fig. 1 compares the autocorrelation plot for the MCMC chains of ρ. The plot confirms that PG-DA significantly reduced autocorrelation and improved ESS. This result is consistent with Polson et al. (2013). In other words, the efficiency of PG-DA was preserved even when the ρ parameter was newly introduced for modeling spatial autocorrelation.

To assess the modeling accuracy for smaller samples, ESS and MCSE of ρ are additionally evaluated with $S \in \{25, 50, 100\}$ assuming $\rho^* = 0.4$. The result is summarized in Table 5, wherein the result in the case with $S = 200$ is the same as in Table 4. The posterior mean of MH is upwardly biased, and the bias tends to increase as S decreases. In contrast, the posterior means of PG-DA have similar values across different S values. PG-DA seems to be more accurate for small samples. In addition, the ESS of PG-DA

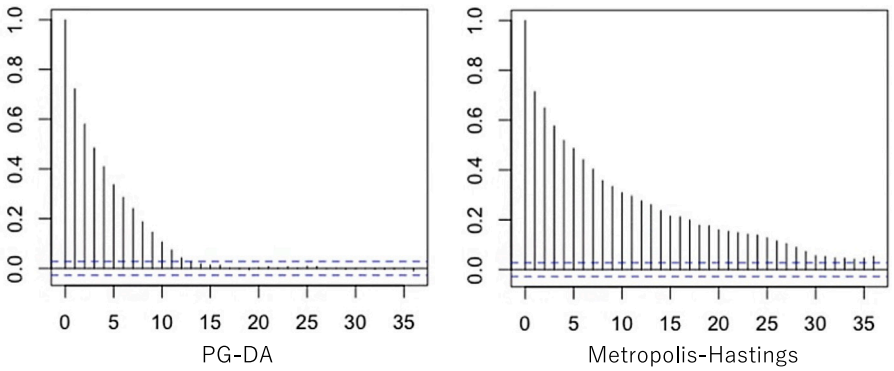


Fig. 1. Autocorrelation plot of the MCMC chain for ρ with $\rho^* = 0.4$. The blue dashed lines represent the 95% confidence interval for no autocorrelation.

Table 4
ESS and MCSE of ρ with $\rho^* \in \{0.0, 0.4, 0.8\}$.

	$\rho^* = 0.0$		$\rho^* = 0.4$		$\rho^* = 0.8$	
	PG-DA	MH	PG-DA	MH	PG-DA	MH
ESS	1310.3	634.4	541.8	311.0	221.1	68.63
MCSE ($\times 10^{-4}$)	8.07	11.29	10.12	11.11	8.67	10.09

Table 5
Posterior mean, ESS, and MCSE of ρ for different $S \in \{25, 50, 100, 200\}$, with $\rho^* = 0.4$.

	$S = 25$		$S = 50$		$S = 100$		$S = 200$	
	PG-DA	MH	PG-DA	MH	PG-DA	MH	PG-DA	MH
$\hat{\rho}$	0.378	0.509	0.373	0.432	0.376	0.410	0.367	0.369
ESS	666.6	283.7	641.3	304.3	581.6	282.4	541.8	311.0
MCSE ($\times 10^{-4}$)	20.77	23.79	16.67	18.99	11.80	14.42	10.12	11.11

was always larger than MH, suggesting the efficiency of the sampling. Although the MCSE of PG-DA increased with a decrease in S, the values were smaller than MH. The accuracy and fast mixing of PG-DA were confirmed for small samples.

5. Empirical analysis

In this section, we apply the proposed Bayesian SAR-MXL model to an empirical analysis of neighborhood perception, using the data collected by Oyama (2024) in two Tokyo neighborhoods. The neighborhood perception of an individual is represented by spatial multivariate choice data as in Guimpert and Hurtubia (2018). We aim to capture inter-individual heterogeneity and spatial dependence in such responses.

5.1. Data and variables

The data used in this study is the subjective cognitive areas of two neighborhoods in Tokyo, Kiyosumi-Shirakawa and Kiba. The data was collected through an on-street survey by Oyama (2024), where respondents drew a cognitive map as a polygon on the base map, relying on their experiences or prior information on the neighborhoods. The cognitive maps were converted into street-based perception data $y_{ni} \in \{0, 1\}$, indicating whether the respondent n perceived the street i to belong to the neighborhood ($y_{ni} = 1$) or not ($y_{ni} = 0$). The sample contains responses from 25 residents and 25 visitors for each neighborhood, resulting in 100 respondents in total. The target networks of Kiyosumi-Shirakawa and Kiba contain 258 and 473 streets, respectively (Fig. 2). Thus, the number of observations is 12,900 for Kiyosumi-Shirakawa and 23,650 for Kiba, giving a total of 36,550.

As for the spatial attributes of the streets, we mainly focus on (1) the distance from the reference point, (2) boundaries, and (3) streetscape features as in Oyama (2024). The reference point for each neighborhood is indicated by the star-marked point in Fig. 2, which was defined as the closest location to the metro station named after the neighborhood (Kiyosumi-Shirakawa Station of the Hanzomon Line, and Kiba Station of the Tozai Line) among the points along the main park (Kiyosumi Garden & Park for Kiyosumi-Shirakawa and Kiba Park for Kiba). The distance is defined as the network shortest path distance. As for boundaries, we used three dummy variables indicating if a street is inside the boundaries of the river, avenue, and administrative district (Fig. 2). The landscape features were extracted as pixel values from the street photos. The definitions and statistics of the variables are summarized in Appendix A.

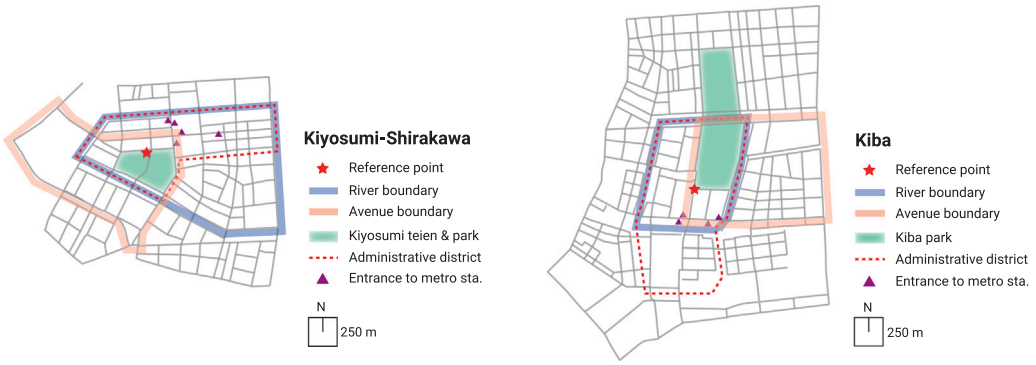


Fig. 2. Street networks of the two neighborhoods.
Source: Extracted from Oyama (2024).

In addition, we consider observed heterogeneity captured by the personal characteristics of the respondents, including age, gender, and familiarity with the neighborhood. Familiarity is measured by the years of residence for residents and by the visit frequency for visitors.

Note that the reader who is interested in the details of the data is referred to Oyama (2024) who provides a detailed data description and some aggregation results.

5.2. Network-based weight matrix

Since the spatial units of our data are streets that compose a network, we defined the spatial weight matrix **W** based on the network and the shortest path distance on it. The following weight matrices were tested:

1. **Adjacency**: $w_{ij} = 1$ only when i is adjacent to j
2. **Angle**: $w_{ij} = a_{ij}/180$ where $a_{ij} \in [0, 180)$ is the angle between streets i and j and $a_{ij} = 0$ if i is not adjacent to j
3. K-**nearest**: $w_{ij} = 1$ if i is one of Kth nearest neighbors to j
4. **Within**-D: $w_{ij} = 1$ if i is within distance D from j

In addition to adjacency and distance-based neighbor matrices, we considered the angle-based matrix with the hypothesis that two neighboring streets with a small angle difference can be recognized as being the same continual street and their utilities are positively correlated with each other.

Note that although distance-based matrices, such as power or exponential, are often used in spatial statistics, they were not considered in this study because they performed very poorly for the networks in the preliminary tests.

5.3. Goodness-of-fit metrics

Given the observations $\{y_{ni}\}$ and the retained sample from conditional posterior $\theta = (\theta^1, \ldots, \theta^R)$ where R is the number of the retained realizations, the following three metrics were used to measure goodness-of-fit of the models:

1. LPPD (log pointwise predictive density) (Gelman et al., 2014):

$$\mathrm{LPPD} = \sum_n \sum_i \log \int p(y_{ni}|\theta_n) p(\theta_n) d\theta_n \approx \sum_n \sum_i \left(\frac{1}{R} \sum_{r=1}^{R} \log p(y_{ni}|\theta_n^r) \right) \tag{5.1}$$

   LPPD corresponds to the logarithm of the pointwise likelihood integrated over the posterior distribution of the relevant model parameters and can be approximated with the scores with the posterior draws.

2. RSME (root mean square error):

$$\mathrm{RSME} = \frac{1}{NS} \sum_n \sum_s \left(\frac{1}{R} \sum_{r=1}^{R} \sqrt{\left(p(y_{ni}|\theta_n^r) - y_{ni} \right)^2} \right) \tag{5.2}$$

   RMSE measures the discrepancy between the predicted and observed street perceptions.

3. ACC (accuracy):

$$\mathrm{ACC} = 1 - \frac{1}{NS} \sum_n \sum_s \left(\frac{1}{R} \sum_{r=1}^{R} |\delta(p(y_{ni}|\theta_n^r) > 0.5) - y_{ni}| \right) \tag{5.3}$$

Table 6
Model comparison result.

	Kiyosumi-Shirakawa				Kiba			
	Logit	SAR	MXL	SAR-MXL	Logit	SAR	MXL	SAR-MXL
LPPD	−4795.689	−4755.895	−2652.493	**−2375.310**	−10 300.790	−10 194.871	−7520.648	**−6715.309**
RSME	0.239	0.235	0.129	**0.112**	0.283	0.278	0.204	**0.178**
ACC	0.837	0.840	0.917	**0.929**	0.802	0.804	0.854	**0.874**

Table 7
Comparison of the SAR-MXL model with different network weight matrices.

	Kiyosumi-Shirakawa						
	Adjacency	Angle	Nearest ($K = 3$)	Nearest ($K = 5$)	Nearest ($K = 10$)	Within (≤250 m)	Within (≤500 m)
LPPD	−2375.310	−2464.288	**−2358.741**	−2575.323	−2616.436	−2592.515	−2602.237
RSME	0.112	0.116	**0.111**	0.123	0.126	0.124	0.126
ACC	**0.928**	0.925	**0.928**	0.921	0.919	0.921	0.919
$\hat{\rho}$ (mean)	0.739	0.666	0.734	0.414	0.254	0.327	0.187
	Kiba						
	Adjacency	Angle	Nearest ($K = 3$)	Nearest ($K = 5$)	Nearest ($K = 10$)	Within (≤250 m)	Within (≤500 m)
LPPD	−6715.309	−6810.303	−6769.014	−6764.419	**−6529.790**	−7048.425	−7285.278
RSME	0.178	0.181	0.180	0.179	**0.172**	0.188	0.195
ACC	0.874	0.872	0.874	0.875	**0.880**	0.870	0.864
$\hat{\rho}$ (mean)	0.760	0.734	0.740	0.779	0.805	0.652	0.489

where $\delta(x > 0.5)$ takes one if x is greater than 0.5 and zero otherwise. ACC is the percentage of the model predicting the accurate perception on the basis of maximum a posteriori probability (MAP).

In a Bayesian context, as defined above, the metrics can be obtained by evaluating the scores at the posterior samples of the model parameters.

5.4. Results

We performed the MCMC estimation for each neighborhood with 50,000 iterations whereby the first 25,000 iterations were discarded as burn-in and every fifth draw was retained to reduce the level of autocorrelation. We used 100 grid points for the griddy Gibbs sampler.

Model fit We compared the following four models (where SD stands for spatial dependence and IH for inter-individual heterogeneity):

1. **Logit**: without SD and IH,
2. **SAR**: with SD but without IH,
3. **MXL**: without SD but with IH,
4. **SAR-MXL**: with both SD and IH,

where we used a simple adjacency matrix for **W** of SAR and SAR-MXL models. Table 6 reports the result. Because smaller LPPD and RSME and greater ACC indicate better fit, the result suggests that the capture of both spatial dependence and inter-individual heterogeneity clearly improved the model fit in both neighborhoods. We thus focus on the SAR-MXL model hereafter.

Then, we compared the different network-based spatial weight matrices **W**, as defined in Section 5.2. Table 7 reports the comparison result. The simple adjacency matrix overall performed well. The angle-based adjacency matrix did not outperform the simple one. While the 3-nearest neighbors performed best for Kiyosumi-Shirakawa, the 10-nearest neighbors performed best for Kiba. This implies that the perception utilities of streets in Kiba more widely depend on each other than those in Kiyosumi-Shirakawa.

Table 7 also reports the estimated posterior mean of spatial parameter $\hat{\rho}$ for each case. Regardless of the definition of **W**, it was estimated as a positive value, indicating the positive spatial correlation between the utilities of streets spatially close to each other.

Estimation results Finally, we report the estimation results of the SAR-MXL models. We used the weight matrix that performed best for each neighborhood (3-nearest in Kiyosumi-Shirakawa and 10-nearest in Kiba).

To examine the convergence of the Gibbs sampler for the SAR-MXL models, we report Gelman–Rubin diagnostic (Gelman and Rubin, 1992). Specifically, we ran two MCMC chains for the same model and calculated the potential scale reduction factor (PSRF) of each model parameter. Fig. 3 shows that the PSRF values of all parameters for both neighborhoods are below 1.1, which is the generally recommended threshold to indicate the convergence (Gelman and Rubin, 1992).

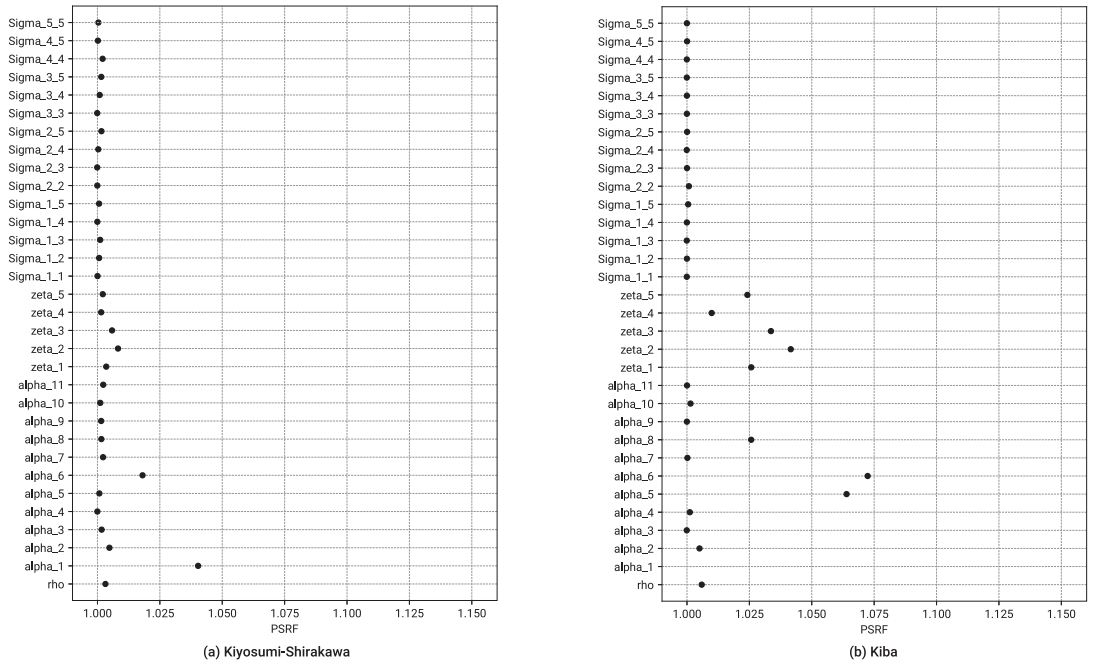


Fig. 3. Potential scale reduction factors (PSRF) of posteriors.

Table 8 reports the mean, standard deviation, and credible interval of the posterior of each estimated parameter for the SAR-MXL model,[5] and Table 9 reports the direct and indirect impacts of the changes in the policy variables. According to the formulations in Section 2.2, we calculated elasticities for continuous variables and marginal effects for dummy variables, using the retained sample from the conditional posterior.

The main findings from the estimation results are fourfold. Firstly, the spatial parameter was estimated as a large positive value for both neighborhoods. This indicates a high spatial dependence between the utilities of streets that are spatially close to each other. Since we used a SAR specification, the utility of a street depends not only on its attributes but also on the attributes of the neighboring streets. The estimated indirect impacts indicate nontrivial spillover effects of attribute changes of a street on its neighboring streets.

Secondly, the marginal effects together with the positive signs of the coefficients for the inner-boundaries attributes indicate that the boundaries played as edges of the neighborhood perception of the individuals. Therefore, mitigating boundaries can be an effective way to expand the extent of neighborhood perception. The results suggest that the river boundary is the strongest edge of perception in both neighborhoods. It has a particularly large effect in Kiyosumi-Shirakawa, where on average a 8.6% change in perception probability is suggested between inside and outside the river boundary, while the change is 3.7% in Kiba. It also has a clear indirect marginal effect on surrounding streets, 18.6% change in Kiyosumi-Shirakawa and 13.1% in Kiba, respectively. The avenue boundary has a nontrivial impact on street perception probability in Kiba, but its impact is less than that of the river boundary in both neighborhoods.

Thirdly, as expected, the distance is negatively associated with the perception probability. A 1% increase in distance from the reference point of a street decreases on average the perception probability of the street by 2.99% in Kiyosumi-Shirakawa and 1.86% in Kiba, respectively. We also captured observed heterogeneity by the interaction terms of distance with personal characteristics and the direction in which a street is located from the reference point. In Kiba, people 40 years or older are more sensitive to distance, implying that their perceived neighborhood tend to be smaller than those of younger people. Residents who have lived in the neighborhoods for less than 10 years tend to have larger perceived neighborhoods than those who have lived longer, possibly reflecting the ambiguity of their neighborhood perception.

The sensitivity to distance also highly depends on the direction. A positive coefficient value for a direction interaction term mitigates the resistance to distance, where the west direction was treated as the reference. In Kiyosumi-Shirakawa, the streets in the south are the least sensitive to distance, followed by the east, north, and west. There is a wide river on the west side of Kiyosumi-Shirakawa, and the cognitive area hardly extended to the west because of the river. In Kiba, the streets in the north are the least sensitive to distance, followed by the south, east, and west. This result captures the effect of the extent of Kiba Park; indeed, most respondents drew their cognitive maps so that the entire area of the park is included (Fig. 2).

[5] The off-diagonal elements of the estimated covariance matrix $\hat{\boldsymbol{\Sigma}}$ are reported in Appendix B.

Table 8
Estimation result of the SAR-MXL models.

	Kiyosumi-Shirakawa				Kiba			
	mean	std. dev.	2.50%	97.50%	Mean	std. dev.	2.50%	97.50%
Spatial parameter								
$\hat{\rho}$	0.734	0.022	0.683	0.762	0.805	0.007	0.802	0.822
Fixed parameters								
Intercept	−0.022	0.272	−0.478	0.534	0.393	0.095	0.239	0.612
Boundaries								
Inner river	1.272	0.102	1.088	1.480	0.376	0.037	0.303	0.447
Inner avenue	0.047	0.057	−0.065	0.158	0.162	0.028	0.107	0.216
Inner admin. district	0.230	0.047	0.135	0.322	0.146	0.036	0.075	0.215
Distance × demographic chars.								
Live less than 10 yrs	0.347	0.202	−0.133	0.720	0.202	0.082	0.057	0.380
Visit once a month or less	−0.065	0.176	−0.427	0.277	0.151	0.095	−0.037	0.330
Female	0.068	0.155	−0.247	0.368	0.048	0.074	−0.097	0.189
40 yrs or older	−0.025	0.165	−0.344	0.292	−0.165	0.090	−0.331	0.018
Distance × direction								
East	1.130	0.116	0.914	1.368	0.360	0.041	0.281	0.441
North	0.301	0.164	−0.027	0.622	1.301	0.044	1.204	1.378
South	1.617	0.118	1.385	1.848	0.837	0.051	0.739	0.939
Random parameters								
Mean								
Distance	−4.093	0.325	−4.778	−3.497	−2.121	0.092	−2.301	−1.934
Green pixel	0.028	0.006	0.017	0.040	0.033	0.002	0.028	0.037
Building pixel	−0.003	0.004	−0.010	0.003	0.002	0.001	−0.001	0.004
Road pixel	0.013	0.005	0.004	0.021	−0.006	0.002	−0.010	−0.002
Sky pixel	0.000	0.005	−0.011	0.010	0.013	0.003	0.007	0.019
Std. dev.								
Distance	1.542	0.148	1.267	1.846	1.067	0.054	0.963	1.175
Green pixel	0.038	0.005	0.028	0.047	0.035	0.002	0.031	0.040
Building pixel	0.032	0.003	0.026	0.038	0.025	0.001	0.023	0.028
Road pixel	0.043	0.003	0.037	0.050	0.039	0.002	0.035	0.042
Sky pixel	0.043	0.005	0.034	0.053	0.043	0.003	0.037	0.050

Table 9
Estimated average impacts of changes in attributes.

	Kiyosumi-Shirakawa				Kiba			
	Mean	std. dev.	2.50%	97.50%	Mean	std. dev.	2.50%	97.50%
Marginal effects								
Inner river								
Direct	0.087	0.011	0.068	0.111	0.037	0.004	0.029	0.046
Indirect	0.186	0.021	0.148	0.231	0.133	0.017	0.102	0.168
Inner avenue								
Direct	0.003	0.004	−0.004	0.011	0.016	0.003	0.010	0.022
Indirect	0.007	0.008	−0.009	0.024	0.057	0.011	0.036	0.081
Elasticities								
Distance								
Direct	−3.026	0.409	−3.876	−2.277	−1.836	0.217	−2.274	−1.425
Indirect	−6.255	0.940	−8.203	−4.518	−6.315	0.758	−7.857	−4.885
Green pixel								
Direct	0.139	0.122	−0.099	0.380	0.201	0.075	0.056	0.348
Indirect	0.315	0.281	−0.214	0.894	0.767	0.291	0.214	1.357
Building pixel								
Direct	−0.284	0.466	−1.199	0.629	−0.091	0.222	−0.527	0.341
Indirect	−0.593	1.037	−2.646	1.431	−0.288	0.826	−1.909	1.331
Road pixel								
Direct	0.376	0.592	−0.779	1.541	−0.282	0.343	−0.958	0.385
Indirect	0.815	1.328	−1.784	3.433	−0.986	1.255	−3.483	1.436
Sky pixel								
Direct	−0.030	0.250	−0.519	0.459	0.102	0.158	−0.208	0.409
Indirect	−0.058	0.559	−1.156	1.046	0.374	0.564	−0.725	1.482

Finally, we evaluated the extent to which the visual landscape features of the streets affect their perception probability in each neighborhood. In both neighborhoods, the pixel values had random policy effects among individuals as the credible intervals of their effect sizes included both negative and positive values. Only the Green pixel value in Kiba showed clear positive effects. On average, a 1% increase in the Green pixel of a street increases the perception probability of the street by 0.20% as a direct impact

Table 10
Definition and statistics of explanatory variables. See Oyama (2024) for more details.

Variable	Definition	Unit	Kiyosumi-Shirakawa				Kiba			
			Mean	Std.	Min	Max	Mean	Std.	Min	Max
Distance	Network shortest path distance from the reference point to street	Kilometer	0.73	0.33	0	1.66	0.93	0.39	0	1.96
Boundaries										
Inner river	Equals 1 if the street is inside the river boundary, 0 otherwise.	Dummy	0.46	0.50	0	1	0.10	0.29	0	1
Inner avenue	Equals 1 if the street is inside the avenue boundary, 0 otherwise.	Dummy	0.22	0.41	0	1	0.18	0.39	0	1
Inner administrative district	Equals 1 if the street is inside the administrative district of the neighborhood, 0 otherwise.	Dummy	0.27	0.45	0	1	0.16	0.36	0	1
Streetscape features										
Building pixel	Building pixel value of street image	Pixel	6.50	7.61	0	38	41.05	14.52	1	72
Road pixel	Road pixel value of street image	Pixel	42.33	12.17	7	70	36.65	7.09	3	55
Sky pixel	Sky pixel value of street image	Pixel	37.53	6.08	16	50	10.64	6.01	0	46
Green pixel	Green pixel value of street image	Pixel	11.22	5.65	0	28	8.54	9.84	0	52
Other pixel	Other pixel value of street image, treated as reference.	Pixel	2.41	3.49	0	27	3.12	4.82	0	48
Directions										
East	Ratio of east direction in which the street is located	Ratio	0.36	0.35	0	0.997	0.25	0.32	0	0.994
North	Ratio of north direction in which the street is located	Ratio	0.15	0.29	0	0.990	0.35	0.37	0	0.999
South	Ratio of south direction in which the street is located	Ratio	0.32	0.30	0	0.960	0.19	0.27	0	0.954
West	Ratio of west direction in which the street is located, treated as reference	Ratio	0.17	0.29	0	0.998	0.21	0.29	0	0.995
Demographic characteristics										
40 yrs or older	Equals 1 if the respondent is 40 years or older, 0 otherwise.	Dummy	0.42	0.49	0	1	0.44	0.50	0	1
Female	Equals 1 if the respondent is female, 0 otherwise.	Dummy	0.58	0.49	0	1	0.54	0.50	0	1
Live less than 10 yrs	Equals 1 if the respondent is a resident who lived in the neighborhood for less than 10 years.	Dummy	0.34	0.47	0	1	0.30	0.46	0	1
Visit once a month or less	Equals 1 if the respondent is a visitor who visits the neighborhood once a month or less.	Dummy	0.18	0.38	0	1	0.26	0.44	0	1

and those of the surrounding streets by 0.73% as an indirect (spillover) impact. This result indicates that street greenery forms the spatial image of the Kiba neighborhood and increasing it can reinforce such an image.

6. Discussion

The numerical experiments in Section 4 demonstrated the efficiency of the PG-DA over the MH method in the Bayesian spatial choice model. While Polson et al. (2013) has shown its efficiency in basic binary logit and negative binomial model settings, this study additionally confirmed the efficiency under different strengths of spatial dependence. The MH method particularly suffered from low ESS when spatial dependence was strong. Given that strong spatial dependence was found in the empirical analysis in Section 5, our results well supported the effectiveness of PG-DA-based Bayesian estimation in spatial choice modeling.

The empirical results in Section 5 were generally consistent with those of Oyama (2024) who estimated a simple binary logit framework. However, the key difference to the previous study is that we captured both spatial dependence and inter-individual heterogeneity and significantly improved the model fit by using the novel SAR-MXL framework. We found strong spatial dependence among the utilities of street segments for individuals to perceive them within the neighborhood. Also, the incorporation of random parameters captured unobserved heterogeneity among individuals and made the observed heterogeneity somewhat unclear. This implies that gender, age and familiarity may not be enough to explain inter-individual heterogeneity in our neighborhood perception data.

The empirical analysis found clear direct and indirect (spillover) effects of river boundaries on the perception probability of streets in both neighborhoods. Moreover, by using the pixel values extracted from street images, we found the nonnegligible effects of streetscape greenery and buildings to reinforce the neighborhood perception in Kiba. These results show the potential of the proposed SAR-MXL framework in policy analysis. Yet, it should be noted that the effect sizes were the means across all observations with the size of $N \times S$, which might have made the discussion on effect sizes unclear. One possible way to analyze the effect sizes more specifically is to focus on several representative streets in terms of policy or location.

7. Conclusion

This study proposed a hierarchical Bayesian SAR-MXL model to analyze multivariate binary choice data. Based on the seminal work of Polson et al. (2013), we addressed the non-conjugacy issue of the logit framework and derived the conjugate posterior

Table 11
Off-diagonal elements of the estimated covariance matrix for empirical analysis.

		Kiyosumi-Shirakawa				Kiba			
		Mean	std. dev.	2.50%	97.50%	Mean	std. dev.	2.50%	97.50%
Distance	Green pixel	0.0110	0.0126	−0.0139	0.0366	0.0136	0.0071	0.0007	0.0291
Distance	Building pixel	0.0016	0.0096	−0.0182	0.0207	−0.0064	0.0047	−0.0165	0.0018
Distance	Road pixel	−0.0294	0.0137	−0.0600	−0.0065	−0.0146	0.0074	−0.0310	−0.0015
Distance	Sky pixel	−0.0162	0.0139	−0.0457	0.0087	−0.0200	0.0089	−0.0404	−0.0048
Green pixel	Building pixel	−0.0001	0.0003	−0.0006	0.0005	0.0003	0.0002	0.0001	0.0007
Green pixel	Road pixel	−0.0005	0.0004	−0.0014	0.0001	−0.0011	0.0003	−0.0018	−0.0006
Green pixel	Sky pixel	−0.0001	0.0004	−0.0009	0.0006	0.0002	0.0003	−0.0004	0.0008
Building pixel	Road pixel	−0.0010	0.0003	−0.0018	−0.0005	−0.0007	0.0002	−0.0011	−0.0004
Building pixel	Sky pixel	0.0004	0.0003	−0.0001	0.0011	0.0004	0.0002	0.0000	0.0008
Road pixel	Sky pixel	−0.0008	0.0004	−0.0017	0.0000	−0.0006	0.0003	−0.0013	0.0000

distributions of both fixed and individual-specific parameters. The high efficiency of the framework was demonstrated by the numerical experiment.

In the application, we successfully captured the heterogeneity and spatial dependence on the extent the spatial attributes affect the perception probability of streets. As such, the Bayesian SAR-MXL model is a flexible framework to analyze subjective spatial perception and can be applied to other contexts. One may use the SAR-MXL model as a choice set formation model for a destination choice model. Since the choice set definition is crucial in destination choice modeling, the integration of our model may improve its predictability.

Finally, we would like to mention the limitations and potential extensions of this study. First, our framework is designed for binary choice responses. Although the extension to mixed MNL specification is not straightforward (Bansal et al., 2019), in the spatial statistics context, Krisztin et al. (2022) successfully applied the PG-DA strategy to a spatial MNL model and analyzed urban expansion. It would be appealing to extend our framework to an MNL case. Also, we made some assumptions that could be relaxed in future works. For example, LeSage and Chih (2018) show the incorporation of heterogeneous spatial parameter ρ into a Bayesian spatial model, and Krueger et al. (2020b) examine semi-parametric representations of unobserved heterogeneity in hierarchical Bayesian logit models. The integration of these techniques into our framework could extend its applicability.

CRediT authorship contribution statement

Yuki Oyama: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Daisuke Murakami:** Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis. **Rico Krueger:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis.

Declaration of competing interest

The authors declare no conflicts of interest associated with this manuscript.

Data availability

Source code and synthetic data are available at the author's webpage (https://github.com/yuki-oyama/bayes-sl).

Acknowledgment

This work was supported by JSPS, Japan KAKENHI Grants 23K26280 and 23H01586. The authors thank two anonymous reviewers for their constructive feedback to improve the quality of the paper.

Appendix A. Explanatory variables for empirical analysis

See Table 10.

Appendix B. Off-diagonal elements of covariance matrix

See Table 11.

References

Akinc, D., Vandebroek, M., 2018. Bayesian estimation of mixed logit models: Selecting an appropriate prior for the covariance matrix. J. Choice Model. 29, 133–151.

Anselin, L., Florax, R.J., 1995. Small sample properties of tests for spatial dependence in regression models: Some further results. In: New Directions in Spatial Econometrics. Springer, pp. 21–74.

Bansal, P., Krueger, R., Bierlaire, M., Daziano, R.A., Rashidi, T.H., 2019. Pólygamma data augmentation to address non-conjugacy in the Bayesian estimation of mixed multinomial logit models. arXiv preprint arXiv:1904.07688.

Bansal, P., Krueger, R., Graham, D.J., 2021. Fast Bayesian estimation of spatial count data models. Comput. Stat. Data Anal. 157, 107152.

Bhat, C.R., 2011. The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. Transp. Res. B 45 (7), 923–939.

Bhat, C., 2015. A new spatial (social) interaction discrete choice model accommodating for unobserved effects due to endogenous network formation. Transportation 42, 879–914.

Bhat, C.R., Guo, J., 2004. A mixed spatially correlated logit model: formulation and application to residential choice modeling. Transp. Res. B 38 (2), 147–168.

Buddhavarapu, P., Bansal, P., Prozzi, J.A., 2021. A new spatial count data model with time-varying parameters. Transp. Res. B 150, 566–586.

Buddhavarapu, P., Scott, J.G., Prozzi, J.A., 2016. Modeling unobserved heterogeneity using finite mixture random parameters for spatially correlated discrete count data. Transp. Res. B 91, 492–510.

Calabrese, R., Elkink, J.A., 2014. Estimators of binary spatial autoregressive models: A Monte Carlo study. J. Reg. Sci. 54 (4), 664–687.

Coulton, C.J., Korbin, J., Chan, T., Su, M., 2001. Mapping residents' perceptions of neighborhood boundaries: a methodological note. Am. J. Community Psychol. 29 (2), 371–383.

Downs, R.M., Stea, D., 1973. Cognitive maps and spatial behavior: Process and products. In: Dodge, M., Kitchin, R., Perkins, C. (Eds.), The Map Reader: Theories of Mapping Practice and Cartographic Representation. Wiley, US, pp. 312–317.

Edwards, Y.D., Allenby, G.M., 2003. Multivariate analysis of multiple response data. J. Market. Res. 40 (3), 321–334.

Gelman, A., Hwang, J., Vehtari, A., 2014. Understanding predictive information criteria for Bayesian models. Stat. Comput. 24 (6), 997–1016.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 7 (4), 457–472.

Golledge, R.G., Jacobson, R.D., Kitchin, R., Blades, M., 2000. Cognitive maps, spatial abilities, and human wayfinding. Geogr. Rev. Japan. Ser. B 73 (2), 93–104.

Guimpert, I., Hurtubia, R., 2018. Measuring, understanding and modelling the walking neighborhood as a function of built environment and socioeconomic variables. J. Transp. Geogr. 71, 32–44.

Gurumurthy, K.M., Bansal, P., Kockelman, K.M., Li, Z., 2022. Modelling animal-vehicle collision counts across large networks using a bayesian hierarchical model with time-varying parameters. Anal. Methods Accid. Res. 36, 100231.

Huang, A., Wand, M.P., 2013. Simple marginally noninformative prior distributions for covariance matrices. Bayesian Anal. 8 (2), 439–452.

Klier, T., McMillen, D.P., 2008. Clustering of auto supplier plants in the United States: generalized method of moments spatial logit for large samples. J. Bus. Econom. Statist. 26 (4), 460–471.

Krisztin, T., Piribauer, P., 2021. A Bayesian spatial autoregressive logit model with an empirical application to European regional FDI flows. Empir. Econ. 61 (1), 231–257.

Krisztin, T., Piribauer, P., Wögerer, M., 2022. A spatial multinomial logit model for analysing urban expansion. Spatial Econ. Anal. 17 (2), 223–244.

Krueger, R., Bansal, P., Buddhavarapu, P., 2020a. A new spatial count data model with Bayesian additive regression trees for accident hot spot identification. Accid. Anal. Prev. 144, 105623.

Krueger, R., Rashidi, T.H., Vij, A., 2020b. A Dirichlet process mixture model of discrete choice: Comparisons and a case study on preferences for shared automated vehicles. J. Choice Model. 36, 100229.

LeSage, J.P., 1997. Bayesian estimation of spatial autoregressive models. Int. Reg. Sci. Rev. 20 (1–2), 113–129.

LeSage, J.P., 2000. Bayesian estimation of limited dependent variable spatial autoregressive models. Geogr. Anal. 32 (1), 19–35.

LeSage, J.P., Chih, Y.-Y., 2018. A Bayesian spatial panel model with heterogeneous coefficients. Reg. Sci. Urban Econ. 72, 58–73.

LeSage, J.P., Kelley Pace, R., Lam, N., Campanella, R., Liu, X., 2011. New orleans business recovery in the aftermath of hurricane katrina. J. R. Stat. Soc. Ser. A 174 (4), 1007–1027.

LeSage, J.P., Pace, R.K., 2009. Spatial econometric models. In: Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications. Springer, pp. 355–376.

Lynch, K., 1960. The Image of the City. MIT Press.

McMillen, D.P., 1992. Probit with spatial autocorrelation. J. Reg. Sci. 32 (3), 335–348.

Miyamoto, K., Vichiensan, V., Shimomura, N., Páez, A., 2004. Discrete choice model with structuralized spatial effects for location analysis. Transp. Res. Rec. 1898 (1), 183–190.

Mondal, A., Bhat, C.R., 2022. A spatial rank-ordered probit model with an application to travel mode choice. Transp. Res. B 155, 374–393.

Oyama, Y., 2024. Spatial city image and its formative factors: A street-based neighborhood cognition analysis. Cities 149, 104898.

Pinkse, J., Slade, M.E., 1998. Contracting in space: An application of spatial statistics to discrete-choice models. J. Econometrics 85 (1), 125–154.

Piras, G., Sarrias, M., 2023. One or two-step? Evaluating GMM efficiency for spatial binary probit models. J. Choice Model. 48, 100432.

Polson, N.G., Scott, J.G., Windle, J., 2013. Bayesian inference for logistic models using Pólya–Gamma latent variables. J. Am. Stat. Assoc. 108 (504), 1339–1349.

Ritter, C., Tanner, M.A., 1992. Facilitating the Gibbs sampler: the Gibbs stopper and the griddy-Gibbs sampler. J. Amer. Statist. Assoc. 87 (419), 861–868.

Roberts, G.O., Rosenthal, J.S., 2001. Optimal scaling for various Metropolis-Hastings algorithms. Stat. Sci. 16 (4), 351–367.

Sener, I.N., Pendyala, R.M., Bhat, C.R., 2011. Accommodating spatial correlation across choice alternatives in discrete choice models: an application to modeling residential location choice behavior. J. Transp. Geogr. 19 (2), 294–303.

Smith, G., Gidlow, C., Davey, R., Foster, C., 2010. What is my walking neighbourhood? A pilot study of english adults' definitions of their local walking neighbourhoods. Int. J. Behav. Nutr. Phys. Activity 7 (1), 1–8.

Train, K.E., 2009. Discrete Choice Methods with Simulation. Cambridge University Press.