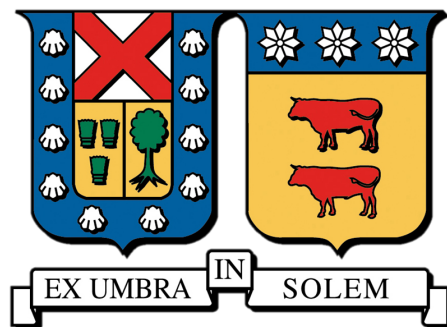




UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

DEPARTAMENTO DE MATEMÁTICA
VALPARAÍSO - CHILE

Effective Sample size for Spatial General linear Models

Tesis presentada por:

Benjamín Bravo Carrasco

*Como requisito para optar al título profesional de Ingeniería Civil
Matemática - Magister en Ciencias de la ingeniería civil Matemática*

Director de Tesis:

Dr. Ronny Vallejos

Diciembre, 2025



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): Memoria o trabajo de título ☒ Tesis de Postgrado

Título del trabajo:

Effective Sample Size for Spatial General linear Models

Nombre del candidato(a):

Benjamín Fernando Bravo Carrasco

Carrera / Grado:

Ingeniería Civil Matemática / Magister en Ciencias Mención Matemática

Campus: Casa Central Departamento: Matemática

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Ronny Vallejos Arriagada, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente

DEJO CONSTANCIA que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

X El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por (**marcar una opción**):

6 meses 12 meses 2 años 3 años 5 años 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: __02/09/2026__ Firma: _____



Estudiante o Candidato(a):

Fecha: 15/05/2026 **Firma:** Benjamin Bravo C

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

TÍTULO DE LA TESIS:

Effective Sample Size for Spatial Generalized linear Models.

AUTOR:

Benjamín Bravo Carrasco

TRABAJO DE TESIS, presentado en cumplimiento parcial de los requisitos para el título de Ingeniería Civil Matemática de la Universidad, y Magister en Ciencias de la ingeniería Matemática Técnica Federico Santa María.

COMISIÓN DE TESIS:

Integrantes

Firma

Dr. Ronny Vallejos

Universidad Técnica Federico Santa María, Chile.

Dr. Felipe Osorio

Dr. Nixon Jerez

Universidad Técnica Federico Santa María, Chile.

Dr. Jonathan Acosta

Pontificia Universidad Católica de Chile, Chile.

Valparaíso, Mayo 2026.

RESUMEN

Este manuscrito propone una nueva definición del tamaño de muestra efectivo (ESS) para Modelos Lineales Mixtos Generalizados Espaciales (SGLMMs), extendiendo las definiciones existentes las cuales únicamente han sido desarrolladas para procesos espaciales gaussianos, a distribuciones con respuesta del tipo conteo, y binaria. Entre ellas se encuentran la distribución Poisson, binomial negativa y Bernoulli. Esta definición se basa en utilizar la aproximación de Laplace a la verosimilitud marginal, y considera tanto la estructura de correlación espacial como la distribución de la respuesta. Se establecieron las siguientes propiedades teóricas: una cota inferior, condiciones suficientes para que $ESS \leq n$, y una expresión cerrada para matrices de correlación circulantes. Se realizaron estudios de simulación para evaluar la precisión de la aproximación de Laplace frente a estimaciones basadas en TMB, encontrando una buena concordancia entre las distribuciones estudiadas, y la estimación basada en INLA bajo estructuras de correlación Matérn, donde se observó que el desempeño depende de los parámetros de suavidad y rango. Además, realizamos un experimento de dominio creciente para investigar si la consistencia establecida para modelos gaussianos se extiende al marco de SGLMM. Por último, se realizó un experimento con datos reales de conteo espacial de malezas.

ABSTRACT

This work proposes a new effective sample size definition for Spatial Generalized Linear Mixed Models (SGLMMs), extending existing definitions developed for Gaussian spatial processes to accommodate count and binary response distributions, including Poisson, Negative Binomial and Bernoulli. The definition is based on a Laplace approximation to the marginal likelihood and accounts for both the spatial correlation structure and the response distribution. We establish theoretical properties, including a lower bound, sufficient conditions for $ESS \leq n$, and a closed-form expression for circulant correlation matrices. Simulation studies assess the accuracy of the Laplace approximation against TMB-based estimates, finding good agreement across response families, and evaluate INLA-based estimation under Matérn correlation structures, where performance depends on the smoothness and range parameters. An increasing domain experiment investigates whether consistency properties established for Gaussian models extend to the SGLMM setting. The methodology is illustrated on a real spatial weed count dataset.

AGRADECIMIENTOS

En primer lugar, quiero agradecer a mi director de tesis, Ronny Vallejos, por su gran guía, paciencia y dedicación a lo largo de este trabajo. Su gran disposición a responder todas mis dudas ha sido fundamental para que completara este escrito. También quiero agradecer al profesor Felipe Osorio por su valiosa participación en las discusiones que enriquecieron esta tesis.

A mis padres, Cristián y Sandra, a mi hermano Sebastián y a mi cuñada Valentina, por su apoyo incondicional durante todo este proceso. Sin lugar a dudas, sin ellos no podría haber llegado hasta aquí.

Este trabajo fue financiado por el Fondo Nacional de Desarrollo Científico y Tecnológico, proyecto FONDECYT N° 1230012.

Finalmente, agradezco a mis amigos por los momentos compartidos y por hacer más llevadera mi etapa universitaria.

A Cristián y Sandra.

Contents

Resumen	i
Abstract	ii
Agradecimientos	iii
1 Motivation	1
2 Preliminaries and Background	5
2.1 Spatial Random Fields	5
2.1.1 Gaussian random fields and covariance	6
2.2 Effective Sample Size	9
2.2.1 Effective Sample Size for Constant Mean Models	9
2.2.2 Effective Sample Size for Spatial Regression Models	10
2.3 Spatial Generalized Linear Mixed Models	11
2.3.1 From Linear Models to Generalized Linear Models	11
2.3.2 The Exponential Family	12
2.3.3 Generalized Linear Models	13
2.3.4 Incorporating Dependence: Mixed Models	14
2.3.5 Generalized Linear Mixed Models	15
2.3.6 Spatial Generalized Linear Mixed Models	16
2.4 Computational Methods for SGLMMs	17

2.4.1	The Computational Challenge	17
2.4.2	Laplace Approximation	18
2.4.3	Integrated Nested Laplace Aproximation (INLA)	20
2.4.4	Template Model Builder (TMB)	21
3	ESS Derivation	23
3.1	Constructing the ESS	23
3.2	Theoretical properties	25
3.3	Exploratory Analysis of ESS Behavior	28
4	Experiments	31
4.1	Numerical Experiments	31
4.1.1	Experiment 1	31
4.1.2	Experiment 2	39
4.1.3	Experiment 3	47
4.2	Real Data Experiment	52
5	Discussion and Future Work	58
A		61
A.1	ESS Derivation	61
	Bibliografía	64

Chapter 1

Motivation

In areas as diverse as district-level disease records, species richness per plot, or pest infestation in crops, spatial count data are generated. These data present a particular complexity: the value observed at one location is not independent of what occurs in its surroundings. Environmental factors, dispersal processes, or socioeconomic characteristics operate at spatial scales, creating similarity among nearby observations. This phenomenon of spatial dependence, modeled through spatial autocorrelation, is not only an observable pattern in the data but also carries direct and measurable consequences for statistical inference: as nearby observations become more similar, they contribute redundant information, reducing the effective sample size below the number of collected observations.

This reduction is not merely a theoretical concern. Spatial autocorrelation in georeferenced data creates a systematic mismatch between the number of collected observations and the actual amount of information contained in those observations. The greater the spatial dependence, the more severe this mismatch becomes, ultimately undermining the reliability of standard inferential procedures that assume independence. Quantifying this loss of effective information is therefore essential — yet it remains a largely unresolved problem.

Indeed, the quantification of latent information in spatial count data has received relatively little attention in the statistical literature. Unlike the case of Fisher information and other measures that have been extensively studied for independent data and even for spatial Gaussian processes, the case of spatially correlated counts presents particularities that complicate the analysis. The combination of discrete observations with

spatial dependence patterns introduces both conceptual and computational challenges for estimating and interpreting how much real information these data contain.

This phenomenon can be intuitively understood as follows: if two locations are very close in space, their observed values tend to be very similar, which implies that each additional observation contributes less new information about the phenomenon. In other words, nearby observations are partially redundant, whereas observations farther apart carry more independent information about the phenomenon of interest. Ignoring this redundancy of information has consequences for statistical inference. [Dutilleul \(1993\)](#) showed through simulations that, in correlation analyses with spatial data, N observations with moderate autocorrelation ($\rho = 0.2$) may contain information equivalent to only $N/4$ independent observations.

To quantify this reduction in information, the concept of Effective Sample Size (ESS) has been proposed, defined as the equivalent number of independent observations associated with a spatial sample of size n ([Cressie, 1993](#)). In the frequentist context, the earliest definitions of ESS were linked to procedures for adjusting statistical tests. [Clifford et al. \(1989\)](#) and [Dutilleul \(1993\)](#) defined ESS as the number of effective degrees of freedom required to modify t and F tests that assess correlation between spatial processes. Later, [Vallejos y Osorio \(2014\)](#) formalized ESS in terms of Fisher information for Gaussian models with constant mean:

$$ESS = \mathbf{1}^\top \mathbf{R}(\boldsymbol{\theta})^{-1} \mathbf{1}$$

where $\mathbf{R}(\boldsymbol{\theta})$ is the spatial correlation matrix and $\mathbf{1}$ is a vector of ones of length n . This methodology has subsequently been extended to various contexts. [Acosta y Vallejos \(2018\)](#) generalized it for spatial linear regression $ESS = \text{tr}(\mathbf{X}^\top \mathbf{R}(\boldsymbol{\theta})^{-1} \mathbf{X})/p$, where $\mathbf{X}$ is an $n \times p$ design matrix, also establishing asymptotic properties of the maximum likelihood estimator of ESS under an increasing-domain framework. Additional developments can be found in [Acosta et al. \(2018\)](#) and [Acosta et al. \(2021\)](#). More recently, [Alegría \(2023\)](#) extended the concept of ESS to spatio-temporal processes with separable covariance functions, [Mukerjee \(2024\)](#) explored new ways of incorporating blocks in a pairwise-likelihood framework, [Ferrer y Vallejos \(2025\)](#) characterized conditions on the correlation matrix that guarantee $ESS \leq n$ and [Alegría et al. \(2026\)](#) extended the definition of ESS for functional random fields.

The definitions mentioned above have been developed in a context in which the model assumes a Gaussian-type response. Because the data we will investigate in this write is

related to spatial counts, inherently discrete, this approach loses validity. Hence a count distribution such as the Binomial (when a fixed number of individuals is available), the Poisson, or the Negative Binomial (when over-dispersion is present) would be more appropriate.

To properly describe models with discrete response, we consider models that fall within the family of generalized linear models (GLMs) (McCullagh y Nelder, 1989). This type of model provides the ability to handle distributions other than Gaussian by incorporating link functions and variance structures consistent with the nature of the phenomenon. However, these models assume independence among observations, a key assumption that cannot be overlooked within the framework of spatial statistics.

This creates the need to extend the GLM framework to accommodate spatial dependence while preserving distributional flexibility. We initially explored the approach of Gotway y Stroup (1997), which addresses the problem using quasi-likelihood approximations and estimation based on estimating equations with working correlation matrices. Although this method produces consistent estimators of β , it treats the spatial structure as a nuisance parameter rather than as a quantity of interest, since the working correlation matrices are used to facilitate estimation but not for direct interpretation. For our objective, which includes not only obtaining valid estimators for β but also characterizing and quantifying the structure of spatial dependence, this limitation becomes critical. Hence, we adopt the framework of Spatial Generalized Linear Mixed Models (SGLMMs hereafter, Diggle *et al.* (1998)).

Yet, using the SGLMMs framework has its own unique challenges. The main issue is that the marginal likelihood

$$L(\Psi; \mathbf{y}) = \int_{\mathbb{R}^n} \prod_{i=1}^n f(y_i|u_i; \beta) \phi_n(\mathbf{u}; \mathbf{0}, \mathbf{R}(\psi)) d\mathbf{u}, \quad (1.1)$$

where $f(y_i|u_i; \beta)$ is the conditional distribution of the response given the latent effect u_i , and $\phi_n(\mathbf{u}; \mathbf{0}, \mathbf{R}(\psi))$ is the n -dimensional Gaussian density of the latent field $\mathbf{u}$ with mean $\mathbf{0}$ and covariance matrix $\mathbf{R}(\psi)$, involves a high-dimensional integral that is analytically intractable. So to compute this integral, we approximate it using the Laplace approximation. Further details will be provided in Section 3.1. Applying this approximation yields the following expression

$$ESS = n \cdot \frac{\text{tr}(\mathbf{X}^\top (\mathbf{W}^{-1} + \mathbf{R}(\psi))^{-1} \mathbf{X})}{\text{tr}(\mathbf{X}^\top (\mathbf{W}^{-1} + \mathbf{I})^{-1} \mathbf{X})}, \quad (1.2)$$

where $\mathbf{X}$ is the $n \times p$ design matrix of spatial covariates, $\mathbf{W}$ is a diagonal matrix of working weights determined by both the response distribution and the link function, and $\mathbf{R}(\boldsymbol{\psi})$ is the spatial correlation matrix governed by the parameter $\boldsymbol{\psi}$. This expression suffers a scale problem discussed in Section 3.1, which we address by developing a new scale-invariant definition of the ESS that, for most correlation structures commonly used in spatial statistics, decreases as correlation increases. To the best of our knowledge, this is the first ESS definition explicitly accounting for both the spatial correlation structure and the response distribution in the SGLMMs framework.

In addition to the formula for the ESS in SGLMMs we investigate how different response distributions affect the ESS under various correlation structures. We also establish several theoretical properties of the proposed ESS, namely a lower bound $ESS > 1$, sufficient conditions under which $ESS \leq n$, where n is the total number of observations, and an explicit closed-form result for circulant correlation matrices. We conduct numerical experiments to validate the approximation accuracy against TMB-based estimates across Bernoulli, Negative Binomial, and Poisson responses; simulation studies to assess INLA-based ESS estimation under Matérn correlation structures for Poisson, Negative Binomial, and Gamma families; an increasing domain experiment to empirically investigate whether the consistency properties established by Acosta y Vallejos (2018) extend to the SGLMM framework; and a real data application to weed count data from the Bjertorp farm dataset (Guillot *et al.*, 2009).

This manuscript is organized as follows. Section 2 presents the theoretical background and preliminaries, covering spatial random fields, the Effective Sample Size, Spatial Generalized Linear Mixed Models, and the computational methods employed. Section 3 describes the construction of the ESS for SGLMMs and its theoretical properties. Section 4 presents the experimental results, including numerical, simulation, and real data experiments. Section 5 discusses the findings, outlines limitations, and identifies directions for future research. The detailed derivation of the ESS formula is deferred to the Appendix (A) to maintain the flow of the main text. All code and data used in this work are publicly available: the implementation can be found at <https://github.com/benjaminbravoo/essforspatialglm>, and the Bjertorp dataset is accessible through the georob R package (Guillot *et al.*, 2009).

Chapter 2

Preliminaries and Background

The central challenge in defining an effective sample size for SGLMMs lies in properly accounting for the reduction of information induced by spatial dependence while accommodating non-Gaussian responses and hierarchical random effects. This chapter addresses this challenge by developing the necessary theoretical framework in stages.

We begin by introducing spatial random fields and Gaussian covariance structures, which formalize the notion of spatial dependence. We then present existing definitions of the Effective Sample Size for Gaussian settings, both for constant mean and regression models. Next, we build the modeling framework from generalized linear models through mixed models to their spatial extensions, arriving at the SGLMM. The chapter concludes by discussing the computational methods required for inference in these models, namely the Laplace approximation, INLA, and TMB, thereby setting the stage for our proposed methodology.

2.1 Spatial Random Fields

To properly define the ESS and the SGLMMs, we first establish the theoretical foundation underpinning both constructs.

In many applications, observed data arise from phenomena that vary continuously across space—temperature fields, pollution concentrations, or disease prevalence do not exist solely at measurement locations. Instead, the phenomena takes values everywhere in the study region, and our observations only represent a little sample from the continuous spatial field.

This raises a fundamental question: How can we mathematically represent a quantity that varies randomly across space? The answer follows from a natural generalization of random variables. Whereas a standard random variable takes a random value that becomes fixed upon observation, we extend this concept spatially by defining a random variable $Z(\mathbf{s})$ at each location $\mathbf{s}$ in the study region, representing the uncertain value of the phenomenon at that location.

Defining individual random variables at each location is insufficient; the crucial step is to model their interdependence. We achieve this by treating the entire collection $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ as a unified mathematical object—a stochastic process indexed by spatial location. When the index set is a spatial domain $D \subset \mathbf{R}^d$, where d is the dimension of the space (typically $d = 2$ for georeferenced data), we refer to such processes as spatial processes or random fields.

This conceptual framework leads to the following definitions.

Definition 2.1.1 (Random Variable). *Let $(\Omega, \mathcal{F}, P)$ be a probability space. A random variable is a measurable function $X : \Omega \rightarrow \mathbb{R}$ such that for every Borel set $B \in \mathcal{B}(\mathbb{R})$, the pre-image $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ belongs to $\mathcal{F}$.*

Definition 2.1.2 (Stochastic Process). *A stochastic process $\{Z(t) : t \in T\}$ is a collection of random variables indexed by a set T . For each fixed $t \in T$, $Z(t)$ is a random variable. When $T \subset \mathbb{R}^d$ for $d \geq 1$, the process is called a spatial process or random field.*

Definition 2.1.3 (Random Field). *A random field is a stochastic process $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ where $D \subseteq \mathbf{R}^d$ is a spatial domain. For spatial data analysis, we typically consider $d = 2$ for data on a geographical surface (e.g., latitude, longitude), and $d = 3$ for spatial-temporal processes or three-dimensional spatial data.*

2.1.1 Gaussian random fields and covariance

In principle, random fields can follow arbitrary distributions. In practice, however, Gaussian processes dominate spatial statistics. The reason is simple: they balance modeling flexibility with computational feasibility.

Definition 2.1.4 (Gaussian Random Field). *A random field $\{Z(\mathbf{s}) : \mathbf{s} \in D \subseteq \mathbb{R}^d\}$ is called a Gaussian random field if for any finite collection of locations $\mathbf{s}_1, \dots, \mathbf{s}_n \in D$, the vector $(Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))'$ follows a multivariate normal distribution.*

As in the univariate case, the Gaussian random field distribution is fully specified by two functions: the mean function $\mu(\mathbf{s}) = \mathbb{E}[Z(\mathbf{s})]$ and the covariance function $C(\mathbf{s}, \mathbf{s}') = \text{Cov}(Z(\mathbf{s}), Z(\mathbf{s}'))$.

Although this characterization permits great flexibility, the models we focus on in this work are those that satisfy stationarity and isotropy.

Definition 2.1.5 (Weak Stationarity). *A random field $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ is called weakly stationary (or second-order stationary) if the following conditions hold:*

(i) *The mean function is constant: $\mathbb{E}[Z(\mathbf{s})] = \mu$ for all $\mathbf{s} \in D$*

(ii) *The covariance depends only on the separation vector:*

$$\text{Cov}(Z(\mathbf{s}), Z(\mathbf{s}')) = C(\mathbf{h}),$$

where $\mathbf{h} = \mathbf{s} - \mathbf{s}'$.

Definition 2.1.6 (Isotropy). *A stationary random field $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ is called isotropic if the covariance function depends only on the Euclidean distance between locations:*

$$\text{Cov}(Z(\mathbf{s}), Z(\mathbf{s}')) = C(\|\mathbf{s} - \mathbf{s}'\|) = C(h),$$

where $h = \|\mathbf{h}\| = \|\mathbf{s} - \mathbf{s}'\|$.

An important insight of these two properties is rotational invariance. Given that the correlation only depends on distance, if two locations are separated by the same Euclidean distance, they will exhibit identical correlation regardless of their relative orientation in space.

Stationary and isotropy are satisfied by numerous correlation structures, the most popular choices in spatial statistics include:

1. **Exponential:**

$$C(h; \tau^2, \sigma^2, \phi) = \begin{cases} \tau^2 + \sigma^2, & \text{if } h = 0, \\ \sigma^2 \exp\left\{-\frac{h}{\phi}\right\}, & \text{if } h > 0, \end{cases}$$

2. **Gaussian:**

$$C(h; \tau^2, \sigma^2, \phi) = \begin{cases} \tau^2 + \sigma^2, & \text{if } h = 0, \\ \sigma^2 \exp\left\{-\left(\frac{h}{\phi}\right)^2\right\}, & \text{if } h > 0, \end{cases}$$

3. Spherical:

$$C(h; \tau^2, \sigma^2, \phi) = \begin{cases} \tau^2 + \sigma^2, & \text{if } h = 0, \\ \sigma^2 \left[1 - \frac{3h}{2\phi} + \frac{h^3}{2\phi^3} \right], & \text{if } 0 < h < \phi, \\ 0, & \text{if } h \geq \phi, \end{cases}$$

In these parameterizations, $\tau^2 \geq 0$ is the nugget effect, representing micro-scale variation or measurement error; $\sigma^2 > 0$ is the partial sill, representing the spatially structured variance; $\tau^2 + \sigma^2$ is the sill, the total process variance; and $\phi > 0$ is the range parameter, controlling the rate of correlation decay with distance.

Among the correlation functions available, the Matérn correlation function has become the gold standard in spatial statistics due to its theoretical properties and practical flexibility. Introduced by [Matérn \(1960\)](#) and popularized by [Handcock y Stein \(1993\)](#), this family provides explicit control over the smoothness of the underlying spatial process through a shape parameter ν . Moreover, it encompasses many classical correlation functions—including the exponential and Gaussian—as special or limiting cases, making it a highly versatile modeling tool.

Definition 2.1.7. *The Matérn correlation function defined as*

$$\rho(h; \phi, \nu) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{h}{\phi} \right)^{\nu} K_{\nu} \left(\frac{h}{\phi} \right)$$

where $d \geq 0$ is the distance between locations, $\phi > 0$ is the range parameter controlling the rate of correlation decay, $\nu > 0$ is the smoothness parameter, $\Gamma(\cdot)$ denotes the gamma function, and $K_{\nu}(\cdot)$ is the modified Bessel function of the second kind. The parameter ν directly relates to the differentiability of the process: a spatial process with Matérn covariance is k times mean-square differentiable if and only if $\nu > k$. Important special cases include the exponential ($\nu = 0.5$) and, in the limit as $\nu \rightarrow \infty$, the Gaussian correlation function.

So, as we established before, throughout this work, we impose three key assumptions on the spatial random effect $Z(\mathbf{s})$: it follows a Gaussian random field ensuring tractable inference; it is stationary, meaning $\mathbb{E}[U(\mathbf{s})] = 0$ and $\text{Cov}(Z(\mathbf{s}), Z(\mathbf{s}')) = C(\mathbf{s} - \mathbf{s}')$; and it is isotropic, so that $C(\mathbf{s} - \mathbf{s}') = C(\|\mathbf{s} - \mathbf{s}'\|)$, meaning the covariance depends solely on Euclidean distance.

2.2 Effective Sample Size

The first definition to the Effective Sample Size was given by [Cressie \(1993\)](#), which states that, the ESS, is the number of equivalent independent observations associated with a spatial sample of size n .

The phenomenon that spatial correlation in georeferenced data produces redundancy is well documented in the spatial statistics literature; with the degree of redundancy directly proportional to the strength of the autocorrelation ([Cressie, 1993](#); [Haining, 1990](#); [Schabenberger y Gotway, 2005](#)). This arises because spatially close observations are not independent; instead, they share information through their spatial dependence structure ([Griffith, 2005](#); [Vallejos y Osorio, 2014](#)). Ignoring this redundance can lead to overly narrow credible intervals, overconfident posterior inference, and misleading assessments of posterior uncertainty in subsequent spatial modeling.

Since the ESS compute the reduction of effective information from a spatial sample due to spatial correlation, it directly depends on the covariance matrix. Thus, estimation uncertainty in the covariance parameters propagates to the ESS, with direct implications of posterior precision.

2.2.1 Effective Sample Size for Constant Mean Models

[Vallejos y Osorio \(2014\)](#) provided a rigorous framework for the ESS by establishing formal definitions, proving key properties, and deriving practical implications. Their contributions span several important settings including single mean spatial processes under various correlation structures, double mean spatial processes, and elliptically contoured distributions. The general definition is as follows.

Definition 2.2.1. *Let $\{Y(\mathbf{s}) : \mathbf{s} \in D \subset \mathbb{R}^r\}$ be a random field such that for $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n \in D$, the vector $\mathbf{Y} = [Y(\mathbf{s}_1), Y(\mathbf{s}_2), \dots, Y(\mathbf{s}_n)]^\top \sim N(\mu \mathbf{1}, \mathbf{R}(\boldsymbol{\theta}))$, where $\mathbf{R}(\boldsymbol{\theta})$ is a non-singular $n \times n$ correlation matrix and $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^k$ is a vector of unknown parameters. The quantity*

$$ESS = ESS(n, \mathbf{R}(\boldsymbol{\theta}), r) = \mathbf{1}^\top \mathbf{R}(\boldsymbol{\theta})^{-1} \mathbf{1} \quad (2.1)$$

is called the ESS of $\mathbf{Y}$, where $\mathbf{1}$ is an $n \times 1$ vector of ones.

The ESS proposed by the authors satisfy some desirable and intuitive properties than any measure of the sample size should satisfy:

- (i) $\text{ESS} = n$, when there is no spatial correlation.
- (ii) $\text{ESS} = 1$, for a perfect correlation between all possible pairs of variables.
- (iii) ESS is increasing in n .
- (iv) ESS is decreasing as the correlation increases.

2.2.2 Effective Sample Size for Spatial Regression Models

Although Definition 2.2.1 is well-suited for Single mean spatial processes, many applications often require modeling the mean as a function of spatially referenced covariates. This motivated [Acosta y Vallejos \(2018\)](#) the development of a more general definition that accommodates the concept to the spatial regression framework, where the mean structure is given by $\mathbb{E}[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$ rather than $\mu\mathbf{1}$.

We can notice that, the definition of the ESS in the Single mean spatial process: $\text{ESS} = \mathbf{1}^\top \mathbf{R}(\boldsymbol{\theta})^{-1} \mathbf{1}$ is just the Fisher information quantity with respect to μ . If we consider a regression of a response $\mathbf{Y}$ given the predictors $\mathbf{x}_j = (x_j(s_1), \dots, x_j(s_p))$, $j = 1, \dots, p$ such that $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_p)$ and

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\beta}$ is a vector parameter of size p , the Fisher information for $\boldsymbol{\beta}$ in this case is

$$\mathcal{I}(\boldsymbol{\beta}) = \mathbf{X}^\top \mathbf{R}(\boldsymbol{\theta})^{-1} \mathbf{X}.$$

Given this, and the fact that one of the desirable properties of the ESS is that it lies in the interval $[1, n]$, a natural definition would be

$$\text{ESS} = n \cdot \frac{\text{tr}(\mathbf{X}^\top \mathbf{R}(\boldsymbol{\theta})^{-1} \mathbf{X})}{\text{tr}(\mathbf{X}^\top \mathbf{X})}, \quad (2.2)$$

where $\text{tr}(\cdot)$ is the trace operator. This formulation exhibits an undesirable scale-dependency: for $\mathbf{X} = (\mathbf{1}, \mathbf{x}_2)$ and $\widetilde{\mathbf{X}} = (\mathbf{1}, k\mathbf{x}_2)$ with $k \in \mathbb{R} \setminus \{0\}$, the respective effective sample sizes satisfy $n_1^* = n_2^*$ if and only if $k^2 = 1$, meaning that arbitrary rescaling of covariates would artificially alter the ESS.

To address this issue, [Acosta y Vallejos \(2018\)](#) normalize the design matrix by requiring that $\mathbf{x}_j^\top \mathbf{x}_j = n$ for all $j = 2, \dots, p$. Under this constraint, $\text{tr}(\mathbf{X}^\top \mathbf{X}) = np$, which gives the following scale-invariant definition.

Definition 2.2.2 (Effective Sample Size for Spatial Regression Models). *Consider a spatial regression model of the form*

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (2.3)$$

where $\mathbf{Y} = (Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))^\top$, $\mathbf{X} = (\mathbf{1}, \mathbf{x}_2, \dots, \mathbf{x}_p)$ is an $n \times p$ design matrix with $\mathbf{x}_j^\top \mathbf{x}_j = n$ for all $j = 2, \dots, p$, $\boldsymbol{\beta} \in \mathbb{R}^p$ is the vector of regression coefficients, and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{R}(\boldsymbol{\theta}))$, where $\mathbf{R}(\boldsymbol{\theta})$ is a non-singular correlation matrix with $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^q$.

The **effective sample size** of $\mathbf{Y}$ is defined as

$$n^* = \frac{\text{tr}(\mathbf{X}^\top \mathbf{R}(\boldsymbol{\theta})^{-1} \mathbf{X})}{p} = \frac{1}{p} \sum_{j=1}^p \mathbf{x}_j^\top \mathbf{R}(\boldsymbol{\theta})^{-1} \mathbf{x}_j. \quad (2.4)$$

This definition is consistent with that of [Vallejos y Osorio \(2014\)](#) in the sense that when $X = \mathbf{1}$ (constant mean model), the ESS reduces to their original formulation: $\text{ESS} = \mathbf{1}^\top \mathbf{R}(\boldsymbol{\theta})^{-1} \mathbf{1}$. Moreover, in the independence case, that is $\mathbf{R}(\boldsymbol{\theta}) = I$, the ESS equals the actual sample size, $\text{ESS} = n$.

Likewise in the original definition, this formulation satisfies $\text{ESS} \geq 1$ for any correlation structure, and given the conditions established by [Ferrer y Vallejos \(2025\)](#) on $R(\boldsymbol{\theta})$ and $\mathbf{X}$ we can ensure $\text{ESS} \leq n$, thus bounding the effective sample size within the interval $[1, n]$ under appropriate restrictions, as will be further discussed in Section 3.

2.3 Spatial Generalized Linear Mixed Models

Statistical models provide a fundamental framework for understanding complex phenomena by explaining the variability of a response variable in terms of predictor variables. In spatial contexts, this relationship becomes more intricate due to the presence of spatial correlation among observations, which violates the independence assumption underlying many classical modeling approaches.

2.3.1 From Linear Models to Generalized Linear Models

One of the simplest, and widely studied model is the Linear model. Its keys assumptions include (i) the observations are independent of each other, (ii) the mean equals a linear combinations of the predictors, (iii) homoscedasticity: the variance of the response is constant for every observation, (iv) the error are normally distributed.

The GLM extend this framework by relaxing assumptions (ii)-(iv). Specifically, GLMs replace these assumptions with: (ii') a monotone transformation of the mean equals a linear combination of the predictors, (iii') the variance is a function of the mean of each observation, (iv') the distribution of the observations is a member of the exponential family.

These generalizations enable GLMs to handle discrete, non-negative, and bounded responses that arise frequently in practice. To formalize this framework, we first introduce the exponential family of distributions.

2.3.2 The Exponential Family

Definition 2.3.1 (Exponential Family). *A family of distributions $P_{\alpha, \phi}$ is called an exponential family if the probability density function (or mass function) of a random vector $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ can be expressed as*

$$f(\mathbf{y}; \boldsymbol{\theta}, \phi) = \exp [\phi \{ \mathbf{y}^\top \boldsymbol{\theta} - b(\boldsymbol{\theta}) - s(\mathbf{y}, \phi) \}] \quad (2.5)$$

where $b(\cdot)$ and $s(\cdot, \cdot)$ are appropriate functions, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)^\top \in \mathcal{A} \subset \mathbb{R}^n$ represents the natural parameter vector, and ϕ is the dispersion parameter. When $\mathbf{Y}$ has density (2.5), we say that $\mathbf{Y}$ follows an exponential family and write $\mathbf{Y} \sim EF(\boldsymbol{\theta}, \phi)$.

Some examples of distribution are the following

- Normal Distribution

$$f(y; \theta, \phi) = \exp \left[\phi \left\{ y\theta - \frac{\theta^2}{2} - \frac{y^2\phi}{2} + \frac{1}{2} \log \left(\frac{\phi}{2\pi} \right) \right\} \right]. \quad (2.6)$$

- Poisson Distribution

$$f(y; \theta, \phi) = \exp [\phi \{ y\theta - e^\theta - \log(y!) \}]. \quad (2.7)$$

- Binomial Distribution

$$f(y; \theta, \phi) = \exp \left[\phi \left\{ y\theta - n \log(1 + e^\theta) + \log \binom{n}{y} \right\} \right]. \quad (2.8)$$

- Gamma Distribution

$$f(y; \theta, \phi) = \exp \left[\phi \left\{ y\theta + \log(-\theta) - \log(y) + \frac{\log(\Gamma(\phi)) - \log(\phi)}{\phi} \right\} \right]. \quad (2.9)$$

- Negative Binomial Distribution

$$f(y; \theta, \phi) = \exp \left[\phi \left\{ y\theta - r \log(1 + e^\theta) + \log \binom{y+r-1}{y} \right\} \right]. \quad (2.10)$$

One of the characteristic of the exponential family is that the distribution can be fully characterised using the mean of the sample in the following way.

Proposition 2.3.1. *If $\mathbf{Y} \sim EF(\boldsymbol{\theta}, \phi)$, then the following properties hold:*

$$\mathbb{E}(\mathbf{Y}) = \frac{\partial b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \boldsymbol{\mu}, \quad \text{Cov}(\mathbf{Y}) = \phi^{-1} \frac{\partial^2 b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \phi^{-1} V(\boldsymbol{\mu})$$

where $\boldsymbol{\mu} = \mathbb{E}(\mathbf{Y})$ is the mean vector and $V(\boldsymbol{\mu})$ is the variance function. Moreover,

$$\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\alpha}^\top} = \mathbf{V}(\boldsymbol{\mu}), \quad \frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\mu}^\top} = \mathbf{V}(\boldsymbol{\mu})^{-1},$$

Some examples of exponential families are the normal, poisson, gamma, negative binomial, exponential, and others.

2.3.3 Generalized Linear Models

Definition 2.3.2 (Generalized Linear Models). *Consider $Y_1, \dots, Y_n$ independent random variables, each with probability density function*

$$f(y_i; \theta_i, \phi) = \exp [\phi \{y_i \theta_i - b(\theta_i)\} + c(y_i, \phi)] \quad (2.11)$$

where

$$\mathbb{E}(Y_i) = \mu_i = b'(\theta_i), \quad \text{Var}(Y_i) = \phi^{-1} V(\mu_i)$$

with $\mathbf{V}(\boldsymbol{\mu}) = \frac{d\boldsymbol{\mu}}{d\boldsymbol{\theta}}$ denoting the variance function and $\phi^{-1} > 0$ the dispersion parameter.

A Generalized Linear Model (GLM) is defined by (2.11) together with

$$g(\mu_i) = \eta_i, \quad \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta},$$

where η_i represents the linear predictor, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ with $p < n$ is the regression coefficient vector, and $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ denotes a vector of explanatory variables. The function $g(\cdot)$ is monotonic and differentiable, and is called the link function.

There are many link functions that can be used in GLMs, the choice depends on the exponential family and the nature of the response variable; a particularly important case is the canonical link function, defined by $g(\mu_i) = \theta_i$, where θ_i is the natural parameter. Common example includes:

- **Normal:** $g(\boldsymbol{\mu}) = \boldsymbol{\mu}$ (identity link)
- **Bernoulli/Binomial:** $g(\mu_i) = \log\left(\frac{\mu_i}{1 - \mu_i}\right) = \text{logit}(\mu_i)$ (logit link)
- **Poisson:** $g(\boldsymbol{\mu}) = \log(\boldsymbol{\mu})$ (log link)

Other commonly used link functions include:

- **Probit link:** $g(\boldsymbol{\mu}) = \Phi^{-1}(\boldsymbol{\mu})$, where Φ is the standard normal CDF
- **Complementary log-log link:** $g(\boldsymbol{\mu}) = \log(-\log(1 - \boldsymbol{\mu}))$
- **Aranda-Ordaz link:** $g(\mu_i) = \log\left(\frac{(1 - \mu_i)^{-\rho} - 1}{\rho}\right)$, for $\mu \in (0, 1)$ and $\rho > 0$

2.3.4 Incorporating Dependence: Mixed Models

Although GLMs provide a flexible framework for modeling a wide variety of response variables through the exponential family, a key assumption underlying their formulation is that the observations are independent. In many real applications this assumption is violated due to the presence of hierarchical structures, repeated measurements, clustering effects, or spatial correlation. Ignoring such dependence typically leads to underestimated variances, invalid inference, and biased predictions

To overcome this limitation, [Breslow y Clayton \(1993\)](#) introduced the framework of GLMMs. These models extend GLMs by incorporating random effects into the linear predictor, allowing the model to account for correlation structures such as grouping, clustering, repeated measurements, or spatial dependence. In this way, GLMMs combine the distributional flexibility of GLMs with the hierarchical structure of Mixed Models.

The Linear Mixed Models is a model for data that violates the assumption of independent observations required by classical linear models. Its often used for grouped data like schools, hospital, or plots of land, and longitudinal data experiments where the data are recorded over time.

Linear Mixed Models belong to the class of hierarchical models, which introduce an additional layer to the modeling framework by inducing conditional independence among observations. In these models, the response variable is modeled as if its variability arises from two distinct sources: a population-level component, represented by fixed effects

that capture systematic patterns common across all units, and a unit-specific component, represented by random effects that account for individual deviations from the overall structure. This hierarchical specification formalizes the dependence structure by allowing each observational unit to possess its own deviation from the population mean, while still borrowing strength from other units through a shared random effects distribution. Consequently, observations are independent conditional on the random effects, yet marginally correlated due to the shared stochastic component.

Definition 2.3.3 (Linear Mixed Model (LMM)). *Let $Y_1, \dots, Y_n$ be random variables and let $\mathbf{u} = (u_1, \dots, u_q)^\top$ denote a vector of random effects. Conditional on $\mathbf{u}$, the observations $Y_1, \dots, Y_n$ are assumed to be independent and satisfy*

$$Y_i \mid \mathbf{u} = \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \mathbf{u} + \varepsilon_i, \quad i = 1, \dots, n,$$

where $\varepsilon_i \sim N(0, \sigma^2)$ are independent errors. In matrix notation, this can be written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}.$$

The hierarchical structure is completed by assuming $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, and $\mathbf{u} \perp \boldsymbol{\varepsilon}$, so that marginally $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}\mathbf{G}\mathbf{Z}^\top + \sigma^2 \mathbf{I}_n)$, meaning observations sharing the same random-effect components are correlated.

2.3.5 Generalized Linear Mixed Models

As in LMMs, GLMMs assume conditional independence of responses given the random effects. The distinguishing feature is the incorporation of a link function connecting the linear predictor to the conditional mean, thereby extending the framework to handle non-Gaussian responses while maintaining the hierarchical structure that captures dependence.

Definition 2.3.4 (Generalized Linear Mixed Model (GLMM)). *Let $Y_1, \dots, Y_n$ be conditionally independent random variables given a vector of random effects $\mathbf{u} = (u_1, \dots, u_q)^\top$. Conditional on $\mathbf{u}$, each Y_i has density (or mass function)*

$$f(y_i \mid \mathbf{u}; \theta_i, \phi) = \exp\{\phi [y_i \theta_i - b(\theta_i)] + c(y_i, \phi)\}, \quad (2.12)$$

with

$$\mathbb{E}(Y_i \mid \mathbf{u}) = \mu_i = b'(\theta_i), \quad \text{Var}(Y_i \mid \mathbf{u}) = \phi^{-1} V(\mu_i),$$

where $V(\mu_i)$ denotes the variance function of the exponential family.

A Generalized Linear Mixed Model (GLMM) is defined by (2.12) together with the structural model

$$g(\mu_i) = \eta_i, \quad \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \mathbf{u},$$

where: η_i is the linear predictor; $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ is the vector of fixed-effect parameters; $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ is the vector of fixed-effect covariates; $\mathbf{z}_i$ is the vector associated with random effects; $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$ is a vector of random effects with covariance matrix $\mathbf{G}$.

The function $g(\cdot)$ is a monotonic and differentiable link function connecting the conditional mean μ_i to both fixed and random components of the linear predictor. Marginally, the Y_i are correlated through the distribution of $\mathbf{u}$.

2.3.6 Spatial Generalized Linear Mixed Models

Standard GLMMs can accommodate various correlation structures through random effects, but these are often designed for specific dependence that occurs within predefined groups or clusters. Nonetheless, in spatial data, this assumption is often inadequate, as one of the fundamental principles is the Tobler's First Law of Geography, which states that «everything is related to everything else, but near things are more related than distant things». That is, the correlation between observation should depend on their spatial proximity rather than on a grouping structure.

To capture the spatial dependence, the random effects in SGLMMs are modeled with a spatial covariance structure where correlation decreases as a function of distance. Recall from Section 2.1 that several parametric families satisfy the stationary and isotropy conditions required for spatial modeling. Commonly used families include: Gaussian, exponential, spherical and Matérn.

With all of these components in mind—the hierarchical structure from mixed models, link functions for non-Gaussian responses, and spatial correlation structures—we are now in a position to formally define Spatial Generalized Linear Mixed Models.

Definition 2.3.5 (Spatial Generalized Linear Mixed Model). *Let $\mathcal{D} \subset \mathbb{R}^d$ be a spatial region of interest. For the i -th observation, let $\mathbf{s}_i \in \mathcal{D}$ denote its spatial location, $Y(\mathbf{s}_i)$ the observed response, and $\mathbf{x}(\mathbf{s}_i) \in \mathbb{R}^p$ a vector of co-variates. A Spatial Generalized*

Linear Mixed Model (SGLMM), as introduced by [Diggle et al. \(1998\)](#), specifies that the conditional mean of the response given the spatial random effect $\mathbf{u}(\mathbf{s}_i)$ is

$$\mathbb{E}[y(\mathbf{s}_i) \mid \mathbf{u}(\mathbf{s}_i)] = \mu(\mathbf{s}_i) = g^{-1}(\mathbf{x}(\mathbf{s}_i)^\top \boldsymbol{\beta} + \mathbf{u}(\mathbf{s}_i)), \quad i = 1, \dots, n \quad (2.13)$$

where $g(\cdot)$ is a monotonic and differentiable link function, $\boldsymbol{\beta} \in \mathbb{R}^p$ is the fixed effects parameter vector, and $\mathbf{u}(\cdot) = \{\mathbf{u}(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$ is a second-order stationary Gaussian random field with $\mathbb{E}[\mathbf{u}(\mathbf{s})] = 0$ and spatial covariance function

$$\text{Cov}(\mathbf{u}(\mathbf{s}_i), \mathbf{u}(\mathbf{s}_j)) = \sigma^2 \rho(d_{ij}; \boldsymbol{\psi}), \quad (2.14)$$

where $\sigma^2 > 0$, $\rho(\cdot; \boldsymbol{\psi})$ is a valid spatial correlation function with parameters $\boldsymbol{\psi}$, and $d_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$. Conditional on the spatial random effects $\mathbf{u} = (\mathbf{u}(\mathbf{s}_1), \dots, \mathbf{u}(\mathbf{s}_n))^\top$, the responses are mutually independent with density

$$f(y_i \mid \mathbf{u}(\mathbf{s}_i); \boldsymbol{\beta}, \phi) = \exp \{ \phi[y_i \theta_i - b(\theta_i)] + c(y_i, \phi) \}, \quad (2.15)$$

belonging to an exponential family, where $\theta_i = \theta(\mu(\mathbf{s}_i))$ is the natural parameter and ϕ is the dispersion parameter. The marginal likelihood is given by

$$L(\boldsymbol{\Psi}; \mathbf{y}) = \int_{\mathbb{R}^n} \prod_{i=1}^n f(y_i \mid \mathbf{u}(\mathbf{s}_i); \boldsymbol{\beta}, \phi) \cdot \phi_n(\mathbf{u}; \mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\psi})) d\mathbf{u}, \quad (2.16)$$

where $\boldsymbol{\Psi} = (\boldsymbol{\beta}, \phi, \sigma^2, \boldsymbol{\psi})$ and $\phi_n(\cdot; \mathbf{0}, \boldsymbol{\Sigma})$ is the multivariate normal density with covariance matrix $\boldsymbol{\Sigma}$ having entries $[\boldsymbol{\Sigma}]_{ij} = \sigma^2 \rho(d_{ij}; \boldsymbol{\psi})$.

The SGLMM framework is mainly used to model spatially referenced non-Gaussian data. Examples include modeling disease prevalence in epidemiology [Diggle y Giorgi \(2019\)](#), species abundance in ecology [Paradinas et al. \(2023\)](#); [Doser et al. \(2024\)](#), pollution exceedance in environmental monitoring ([Diggle et al., 1998](#)), and agricultural phenomena. In each case, SGLMM simultaneously account for the discrete or non-Gaussian nature of the response and the spatial correlation induced by unmeasured geographical co-variates, providing more accurate inference than models that ignore spatial structure.

2.4 Computational Methods for SGLMMs

2.4.1 The Computational Challenge

SGLMMs face a fundamental computational challenge that has been extensively studied in the literature. While Definition 2.3.5 establishes that $Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n)$ are condi-

tionally independent given the spatial random effects $\mathbf{u}$, inference remains challenging because these random effects are latent variables. Although the conditional density in Equation (2.15) is tractable, practical inference requires the marginal likelihood. As we have seen in Equation (2.16), obtaining the likelihood requires integrating the product of n conditional densities over the n -dimensional Gaussian random effect. For non-Gaussian response distributions, this integral has no analytical solution, rendering exact maximum likelihood estimation computationally infeasible for even moderately sized datasets.

To overcome this computational challenge, several methodological approaches have been proposed since the advent of MCMC methods in the early 1990s. These can be broadly classified into two categories: sampling-based methods and approximation-based methods. Sampling methods such as MCMC provide asymptotically exact inference, but can be prohibitively slow for large spatial datasets. Approximation methods, which include variational inference and Laplace approximation, trade some accuracy for substantial computational gains, and have gained prominence with the development of modern computational frameworks such as Integrated Nested Laplace Approximation (INLA) (Rue *et al.*, 2009) and Template Model Builder (TMB) (Kristensen *et al.*, 2016), both of which leverage the Laplace approximation to achieve efficient inference for latent Gaussian models.

2.4.2 Laplace Approximation

In modern Bayesian statistics, the Laplace approximation has emerged as a fundamental computational tool, particularly for hierarchical models. While the basic method was developed by Laplace in the 18th century for asymptotic analysis, Tierney y Kadane (1986) formalized its application to Bayesian inference, demonstrating its effectiveness for approximating intractable marginal posterior distributions in hierarchical models where latent variables must be integrated out.

The fundamental idea is conceptually straightforward: approximate the target distribution with a normal distribution that matches its peak location and local curvature, thereby replacing intractable integrals with tractable Gaussian calculations.

Definition 2.4.1 (Laplace Approximation). *Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a twice-differentiable function with a unique global maximum at $\hat{\boldsymbol{\theta}}$, and let $H = \nabla^2 f(\hat{\boldsymbol{\theta}})$ denote the Hessian*

matrix at the mode, assumed to be negative definite. The Laplace approximation to the integral

$$I = \int_{\mathbb{R}^d} \exp\{f(\boldsymbol{\theta})\} d\boldsymbol{\theta},$$

is given by

$$I \approx \exp\{f(\hat{\boldsymbol{\theta}})\} (2\pi)^{d/2} |-\mathbf{H}|^{-1/2},$$

where d is the dimension of $\boldsymbol{\theta}$.

The approximation is derived by replacing $f(\boldsymbol{\theta})$ with its second-order Taylor expansion around the mode $\hat{\boldsymbol{\theta}}$:

$$f(\boldsymbol{\theta}) \approx f(\hat{\boldsymbol{\theta}}) + \nabla f(\hat{\boldsymbol{\theta}})^\top (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + \frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^\top \mathbf{H} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}).$$

Since $\hat{\boldsymbol{\theta}}$ is a critical point, the gradient term $\nabla f(\hat{\boldsymbol{\theta}})$ vanishes, leaving

$$f(\boldsymbol{\theta}) \approx f(\hat{\boldsymbol{\theta}}) + \frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' H (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}).$$

Substituting this approximation into the original integral yields

$$I \approx \exp\{f(\hat{\boldsymbol{\theta}})\} \int_{\mathbb{R}^d} \exp\left\{\frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' H (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right\} d\boldsymbol{\theta}.$$

The integral on the right is a standard Gaussian integral with covariance $-\mathbf{H}^{-1}$, which evaluates to $(2\pi)^{d/2} |-\mathbf{H}|^{-1/2}$, yielding the stated result.

In Bayesian inference, when $f(\boldsymbol{\theta}) = \log p(\boldsymbol{\theta}|\mathbf{y})$ is the log-posterior density, the Laplace approximation yields

$$p(\boldsymbol{\theta}|\mathbf{y}) \approx \mathcal{N}(\hat{\boldsymbol{\theta}}, (-H)^{-1}),$$

where $\hat{\boldsymbol{\theta}}$ is the posterior mode and $(-H)^{-1}$ is the approximate posterior covariance matrix.

Despite the computational advantages, the Laplace approximation has important limitations, because this approximation assumes that the condition distribution of the random effects $\mathbf{u}$ given the data and parameter is approximately Gaussian. This assumption may be violated when

- (i) The sample size is small relative to the spatial domain.
- (ii) Observations are highly sparse or contain many extreme values

- (iii) Spatial correlation is extremely strong (very large range parameter) or the nugget effect is negligible, potentially leading to numerical instability in matrix inversions.

However for SGLMMs with moderate to large datasets, the Laplace approximation is generally accurate. This is because the latent field has a Gaussian prior, and the posterior, being proportional to the product of this Gaussian prior and the likelihood, remains approximately Gaussian when sufficient data are available. The spatial correlation structure induces borrowing of strength across neighboring locations, which smooths the posterior distribution and makes it more amenable to Gaussian approximation.

2.4.3 Integrated Nested Laplace Approximation (INLA)

Even though the Laplace approximation provides a powerful framework distributions, its direct application to complex hierarchical models requires careful implementation. [Rue *et al.* \(2009\)](#) developed Integrated Nested Laplace Approximation (INLA) to efficiently handle a broad class of hierarchical models known as Latent Gaussian Models (LGMs). This class includes SGLMMs as a prominent special case, where exponential family observations depend on an unobserved Gaussian spatial random field. INLA has revolutionized spatial statistics by providing fast, accurate Bayesian inference without the computational burden of MCMC methods.

At the first level, the observed data $\mathbf{y} = (y_1, \dots, y_n)$ follow distributions from the exponential family¹, conditional on a latent Gaussian field $\mathbf{u} = (u_1, \dots, u_n)$ and hyperparameters $\boldsymbol{\theta}$. The second level specifies that $\mathbf{u}$ follows a Gaussian distribution with mean depending on covariates and precision matrix depending on $\boldsymbol{\theta}$. Finally in the third level we assign a prior distribution $\pi(\boldsymbol{\theta})$ to the hyperparameter. This structure can be mathematically expressed as:

$$\begin{aligned} y_i | u_i, \boldsymbol{\theta} &\sim p(y_i | u_i, \boldsymbol{\theta}), \quad i = 1, \dots, n, \\ \mathbf{u} | \boldsymbol{\theta} &\sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \mathbf{Q}^{-1}(\boldsymbol{\theta})), \\ \boldsymbol{\theta} &\sim \pi(\boldsymbol{\theta}), \end{aligned}$$

where $\mathbf{Q}(\boldsymbol{\theta})$ is the precision matrix (inverse covariance).

¹While the LGM framework can accommodate certain non-exponential family likelihoods, the exponential family covers the vast majority of practical applications, including all standard GLM families.

The key innovation of INLA lies in its nested approximation strategy, that exploits the hierarchical structure of LGMs, avoiding the approximation of the full posterior $p(\mathbf{u}, \theta | \mathbf{y})$. INLA lies in three main steps.

Step 1: Approximating the hyperparameter posterior. First of all, INLA approximates the marginal posterior of the hyperparameter using a Laplace approximation.

$$\tilde{p}(\theta | \mathbf{y}) \propto \frac{p(\mathbf{u}, \theta, \mathbf{y})}{p_G(\mathbf{u} | \theta, \mathbf{y})} \Big|_{\mathbf{u}=\mathbf{u}^*(\theta)}$$

where $\mathbf{u}^*(\theta)$ is the mode of the conditional distribution $p(\mathbf{u} | \theta, \mathbf{y})$ and $p_G(\mathbf{u} | \theta, \mathbf{y})$ is a Gaussian approximation to this conditional. This approximation is evaluated on a carefully selected grid of hyperparameter values.

Step 2: Approximating the latent field conditionals. For each hyperparameter configuration θ in the grid, INLA approximates the conditional posterior $p(x_i | \theta, \mathbf{y})$ for each latent component using another Laplace approximation (or variants such as the simplified Laplace or Gaussian approximation).

Step 3: Numerical integration. The marginal posteriors of interest are obtained by integrating over the hyperparameter space using numerical quadrature:

$$\tilde{p}(u_i | \mathbf{y}) = \int \tilde{p}(u_i | \theta, \mathbf{y}) \tilde{p}(\theta | \mathbf{y}) d\theta \approx \sum_k \tilde{p}(u_i | \theta_k, \mathbf{y}) \tilde{p}(\theta_k | \mathbf{y}) \Delta_k,$$

where the sum is over the grid points and Δ_k are appropriate integration weights.

The main appeal of INLA is computational: unlike MCMC, which can take hours for large spatial datasets or fine-resolution domains, INLA produces deterministic results in seconds or minutes by exploiting the sparse precision matrix structure inherent to spatial models. The accuracy of the approximation relies on the latent field posterior being approximately Gaussian, a condition that typically holds in spatial models with sufficient data.

2.4.4 Template Model Builder (TMB)

Template Model Builder ([Kristensen *et al.*, 2016](#)) is a general-purpose framework for maximum likelihood estimation in latent-variable and hierarchical models. While it

employs the Laplace approximation like INLA, TMB offers greater modeling flexibility by allowing users to specify arbitrary likelihood functions through C++ template code. This flexibility comes at the cost of more explicit implementation but enables models that do not conform to INLA’s Latent Gaussian Model structure. TMB achieves computational efficiency comparable to INLA through automatic differentiation (AD) and optimized numerical algorithms.

The computational efficiency of TMB derives from automatic differentiation, which computes exact derivatives by constructing a computational graph of elementary operations and applying the chain rule systematically. The graph can be traversed forward (inputs to outputs) or reverse (outputs to inputs). Critically, for scalar functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$, reverse mode computes the full gradient in $O(1)$ function evaluations independent of n , compared to $O(n)$ for forward mode or numerical methods—a game-changing efficiency for high-dimensional optimization.

TMB implements the Laplace approximation through a nested optimization workflow. For a model with parameters $\boldsymbol{\beta}$ and latent effects $\boldsymbol{u}$ with joint negative log-likelihood $f(\boldsymbol{u}, \boldsymbol{\beta})$, TMB iteratively:

- (i) Finds the conditional mode $\hat{\boldsymbol{u}}(\boldsymbol{\beta})$ using AD-accelerated Newton-Raphson.
- (ii) Computes the Laplace-approximated marginal likelihood $L^*(\boldsymbol{\beta}) = f(\hat{\boldsymbol{u}}(\boldsymbol{\beta}), \boldsymbol{\beta}) - \frac{1}{2} \log |\boldsymbol{H}(\boldsymbol{\beta})|$ where $\boldsymbol{H}(\boldsymbol{\beta}) = \nabla_{uu}^2 f$.
- (iii) Optimizes $L^*(\boldsymbol{\beta})$ using AD-computed gradients $\nabla_{\boldsymbol{\beta}} L^*$.

A key advantage is TMB’s exploitation of sparsity. Spatial models typically exhibit sparse Hessians since each location depends only on nearby neighbors. TMB constructs specialized computational tapes for f , ∇f , and the sparse $\nabla_{uu}^2 f$, avoiding unnecessary computations and achieving computational efficiency rivaling INLA.

Additionally, TMB automatically computes the observed information matrix $\mathcal{J}(\boldsymbol{\theta}) = -\nabla^2 \log L(\boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}}$. This Hessian is provided exactly as a by product of optimization, eliminating the need for numerical approximation.

Chapter 3

ESS Derivation

In this chapter, we present a general expression for quantifying the ESS within the SGLMM framework, an empirical verification of this expression using TMB, and several properties that support the validity of our ESS measure.

3.1 Constructing the ESS

As reviewed in Section 2.2, the ESS is a measure that compute the number of equivalent independent observations associated with a spatial sample of size n . Following the same principle as the previous definitions of the ESS (Definitions 2.2.1 and 2.2.2), we use the Fisher information quantity with respect to $\boldsymbol{\beta}$ to obtain a measure of effective sample size. A complication arises in this setting, since the likelihood in SGLMMs involves integrating the product of n conditional densities, and the observed information matrix is defined by.

$$\mathcal{I}(\boldsymbol{\beta}) = -\frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T},$$

where $\ell(\boldsymbol{\beta})$ corresponds to the log-likelihood function. This function cannot be computed analytically due to the high-dimensional derivatives through the integral in (2.16). To overcome this difficulty, we employ the Laplace approximation to the marginal likelihood, which provides a tractable expression for the log-likelihood and its derivatives. The details of this approximation are provided in Appendix A. The approximation yields

$$\ell(\boldsymbol{\beta}) \approx \frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{H}| + \sum_{i=1}^n \log f_i(y_i | \hat{u}_i; \boldsymbol{\beta}) + \log \phi_n(\hat{\mathbf{u}}; \mathbf{0}, \mathbf{R}(\boldsymbol{\psi})),$$

where $\boldsymbol{\psi}$ denotes the vector of spatial covariance parameters, $f_i(y_i|u_i; \boldsymbol{\beta})$ is the conditional density of y_i given the latent process value u_i , $\phi_n(\mathbf{u}; \mathbf{0}, \mathbf{R}(\boldsymbol{\psi}))$ is the multivariate Gaussian density of the latent process, $\hat{\mathbf{u}} = (\hat{u}_1, \dots, \hat{u}_n)^\top$ is the mode of

$$L(\mathbf{u}; \boldsymbol{\Psi}) := \sum_{i=1}^n \log f_i(y_i|u_i; \boldsymbol{\beta}) + \log \phi_n(\mathbf{u}; \mathbf{0}, \mathbf{R}(\boldsymbol{\psi})),$$

where $\boldsymbol{\Psi} = (\boldsymbol{\beta}^\top, \boldsymbol{\psi}^\top)^\top$ is the full parameter vector, and $\mathbf{H} = - \left. \frac{\partial^2 L(\mathbf{u}; \boldsymbol{\Psi})}{\partial \mathbf{u} \partial \mathbf{u}^\top} \right|_{\mathbf{u}=\hat{\mathbf{u}}}$ is the negative Hessian evaluated at the mode $\hat{\mathbf{u}}$.

Differentiating the log-likelihood approximation with respect to $\boldsymbol{\beta}$ yields:

$$\frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = \frac{\partial^2 L(\mathbf{u}; \boldsymbol{\Psi})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} + \frac{\partial^2 L(\mathbf{u}; \boldsymbol{\Psi})}{\partial \boldsymbol{\beta} \partial \mathbf{u}^\top} \cdot \mathbf{H}^{-1} \cdot \frac{\partial^2 L(\mathbf{u}; \boldsymbol{\Psi})}{\partial \mathbf{u} \partial \boldsymbol{\beta}^\top} + O(n^{-1}). \quad (3.1)$$

Taking the negative sign, evaluating at the mode $\hat{\mathbf{u}}$, neglecting higher-order terms, and applying the Sherman-Morrison-Woodbury identity, the observed information matrix is approximated by:

$$\mathcal{I}(\boldsymbol{\beta}) \approx \mathbf{X}^\top (\mathbf{W}^{-1} + \mathbf{R}(\boldsymbol{\psi}))^{-1} \mathbf{X}. \quad (3.2)$$

where $\mathbf{W} = \text{diag}(w_1, \dots, w_n)$, $w_i = \frac{1}{V(\mu_i)[g'(\mu_i)]^2}$ and similarly to Equation (2.2), we can now define the ESS as

$$ESS = n \cdot \frac{\text{tr}(\mathbf{X}^\top (\mathbf{W}^{-1} + \mathbf{R}(\boldsymbol{\psi}))^{-1} \mathbf{X})}{\text{tr}(\mathbf{X}^\top (\mathbf{W}^{-1} + \mathbf{I})^{-1} \mathbf{X})} \quad (3.3)$$

Notice that as in Equation (2.2), this formulations suffers from an undesirable scale dependency. Consider $\mathbf{X} = (\mathbf{1}, \mathbf{x}_2)$ and $\widetilde{\mathbf{X}} = (\mathbf{1}, k\mathbf{x}_2)$, with $k \in \mathbb{Z} \setminus \{0\}$, the ESS for $\mathbf{X}$ and $\widetilde{\mathbf{X}}$ respectively are:

$$ESS_{\mathbf{X}} = n \cdot \frac{\mathbf{1}^\top (\mathbf{W}^{-1} + \mathbf{R}(\boldsymbol{\psi}))^{-1} \mathbf{1} + \mathbf{x}_2^\top (\mathbf{W}^{-1} + \mathbf{R}(\boldsymbol{\psi}))^{-1} \mathbf{x}_2}{\mathbf{1}^\top (\mathbf{W}^{-1} + \mathbf{I})^{-1} \mathbf{1} + \mathbf{x}_2^\top (\mathbf{W}^{-1} + \mathbf{I})^{-1} \mathbf{x}_2},$$

$$ESS_{\widetilde{\mathbf{X}}} = n \cdot \frac{\mathbf{1}^\top (\mathbf{W}^{-1} + \mathbf{R}(\boldsymbol{\psi}))^{-1} \mathbf{1} + k^2 \mathbf{x}_2^\top (\mathbf{W}^{-1} + \mathbf{R}(\boldsymbol{\psi}))^{-1} \mathbf{x}_2}{\mathbf{1}^\top (\mathbf{W}^{-1} + \mathbf{I})^{-1} \mathbf{1} + k^2 \mathbf{x}_2^\top (\mathbf{W}^{-1} + \mathbf{I})^{-1} \mathbf{x}_2}.$$

Clearly, $ESS_{\mathbf{X}} \neq ESS_{\widetilde{\mathbf{X}}}$ unless $k = \pm 1$, showing that the measure is affected by the scale of the covariates.

To address the scale dependency issue, in the remainder of this work the ESS is defined using a normalized design matrix $\mathbf{X} = (\mathbf{1}, \mathbf{x}_2, \dots, \mathbf{x}_p)$ where the columns are normalized such that $\mathbf{x}_j^\top \mathbf{x}_j = n$ for all $j = 2, \dots, p$, with $\mathbf{x}_j = (x_j(\mathbf{s}_1), x_j(\mathbf{s}_2), \dots, x_j(\mathbf{s}_n))^\top$. Under this normalization, the ESS for SGLMMs is given by.

Definition 3.1.1 (ESS for SGLMMs). *For the SGLMM defined in (2.3.5) with normalized design matrix $\mathbf{X}$, the effective sample size is defined as:*

$$ESS = n \cdot \frac{\text{tr}(\mathbf{X}^\top (\mathbf{W}^{-1} + \mathbf{R}(\boldsymbol{\psi}))^{-1} \mathbf{X})}{\text{tr}(\mathbf{X}^\top (\mathbf{W}^{-1} + \mathbf{I})^{-1} \mathbf{X})}. \quad (3.4)$$

This definition captures the effective information in a SGLMM by comparing the precision under spatial correlation to a baseline reference. If $\mathbf{R}(\boldsymbol{\psi}) = \mathbf{I}$ (no spatial correlation), then $ESS = n$, as expected.

Example 3.1.1. *Consider the SGLMM defined in (2.3.5) with correlation function $\mathbf{R}(\psi) = (1 - \psi)\mathbf{I} + \psi\mathbf{1}\mathbf{1}^\top$ and $-1/(n - 1) < \psi < 1$. Then*

$$ESS = n \cdot \frac{\mathbf{d}^\top \mathbf{r} - \frac{\psi \|\mathbf{X}^\top \mathbf{d}\|^2}{1 + \psi \mathbf{1}^\top \mathbf{d}}}{\mathbf{e}^\top \mathbf{r}}, \quad (3.5)$$

where $\mathbf{d} = (w_1/(1 + (1 - \psi)w_1), \dots, w_n/(1 + (1 - \psi)w_n))$, $r_i = \|\mathbf{x}_i\|^2 = \sum_j X_{ij}^2$, (square norm of every row), and $\mathbf{e} = (w_1/(1 + w_1), \dots, w_n/(1 + w_n))$.

In particular, let $\mathbf{X} = (\mathbf{1}, \mathbf{x}_2)$ with $\mathbf{x}_2$ centred ($\mathbf{1}^\top \mathbf{x}_2 = 0$) and $\|\mathbf{x}_2\|^2 = s$, and let $\mathbf{W} = w\mathbf{I}$. Writing

$$d = \frac{w}{1 + (1 - \psi)w}, \quad e = \frac{w}{1 + w},$$

equation (3.5) simplifies to

$$ESS(\psi) = \frac{n d}{e} \cdot \frac{\frac{n}{1 + \psi n d} + s}{n + s} = \frac{n(1 + w)}{1 + (1 - \psi)w} \cdot \frac{\frac{n}{1 + \psi n d} + s}{n + s}. \quad (3.6)$$

If, in addition, $\mathbf{x}_2$ is normalised so that $s = n$, this becomes

$$ESS(\psi) = \frac{n(1 + w)}{2(1 + (1 - \psi)w)} \left(1 + \frac{1}{1 + \psi n d} \right). \quad (3.7)$$

3.2 Theoretical properties

We now establish key theoretical propositions of the ESS for SGLMMs.

Proposition 3.2.1. *Let $\mathbf{R}(\boldsymbol{\psi})$, $\boldsymbol{\psi} \in \mathbb{R}^q$, be a non-singular correlation matrix and let $\mathbf{X}$ be an $n \times p$ matrix with $n \geq 2$ as in Definition 3.1.1. Then $ESS > 1$.*

Proof. From Definition 3.1.1, $\text{ESS} > 1$ if and only if $\text{tr}(\mathbf{X}^\top \mathbf{F} \mathbf{X}) > 0$, where

$$\mathbf{F} := (\mathbf{W}^{-1} + \mathbf{R}(\boldsymbol{\psi}))^{-1} - \frac{1}{n} (\mathbf{W}^{-1} + \mathbf{I}_n)^{-1}. \quad (3.8)$$

Hence it suffices to show that $\mathbf{F}$ is positive definite. Setting $\mathbf{A} = \mathbf{W}^{-1} + \mathbf{R}(\boldsymbol{\psi})$ and $\mathbf{C} = \mathbf{W}^{-1} + \mathbf{I}_n$, both positive definite, we have

$$\mathbf{F} \succ \mathbf{0} \iff \mathbf{A}^{-1} \succ (n\mathbf{C})^{-1} \iff n\mathbf{C} \succ \mathbf{A},$$

where the last equivalence follows from the anti monotonicity of the matrix inverse on positive definite matrices. Expanding,

$$n\mathbf{C} - \mathbf{A} = n(\mathbf{W}^{-1} + \mathbf{I}_n) - (\mathbf{W}^{-1} + \mathbf{R}(\boldsymbol{\psi})) = (n-1)\mathbf{W}^{-1} + (n\mathbf{I}_n - \mathbf{R}(\boldsymbol{\psi})).$$

Since $\mathbf{R}(\boldsymbol{\psi})$ is a non-singular correlation matrix, its eigenvalues satisfy $\lambda_i > 0$ and $\sum_{i=1}^n \lambda_i = \text{tr}(\mathbf{R}(\boldsymbol{\psi})) = n$, which forces $\lambda_{\max}(\mathbf{R}(\boldsymbol{\psi})) < n$. Therefore $n\mathbf{I}_n - \mathbf{R}(\boldsymbol{\psi}) \succ \mathbf{0}$. Moreover, $(n-1)\mathbf{W}^{-1} \succ \mathbf{0}$ since $n \geq 2$ and all weights are positive. As a sum of two positive definite matrices, $n\mathbf{C} - \mathbf{A} \succ \mathbf{0}$, which establishes that $\mathbf{F}$ is positive definite and hence $n^* > 1$. $\square$

Proposition 3.2.2. *Let $\mathbf{A} := \mathbf{I} \otimes (\mathbf{W}^{-1} + \mathbf{R}(\boldsymbol{\theta})) - \frac{\text{vec}(\mathbf{X})\text{vec}(\mathbf{X})^\top}{\text{vec}(\mathbf{X})^\top (\mathbf{I} \otimes (\mathbf{W}^{-1} + \mathbf{I})^{-1}) \text{vec}(\mathbf{X})}$. Then $\text{ESS} \leq n$, if and only if $\mathbf{A}$ is nonnegative definite.*

Proof. Set $\mathbf{H} := \mathbf{I}_p \otimes (\mathbf{W}^{-1} + \mathbf{R}(\boldsymbol{\theta}))$, $\mathbf{u} := \text{vec}(\mathbf{X})$, and $p(\mathbf{X}) := \mathbf{u}^\top (\mathbf{I}_p \otimes (\mathbf{W}^{-1} + \mathbf{I}_n)^{-1}) \mathbf{u}$, so that $\mathbf{A} = \mathbf{H} - \mathbf{u}\mathbf{u}^\top / p(\mathbf{X})$. For any $\mathbf{w} \in \mathbb{R}^{np}$,

$$\begin{aligned} \mathbf{w}^\top \mathbf{A} \mathbf{w} &= \mathbf{w}^\top \mathbf{H} \mathbf{w} - \frac{(\mathbf{w}^\top \mathbf{u})^2}{p(\mathbf{X})} \\ &\geq (\mathbf{w}^\top \mathbf{H} \mathbf{w}) \left(1 - \frac{\mathbf{u}^\top \mathbf{H}^{-1} \mathbf{u}}{p(\mathbf{X})} \right) \quad (\text{Cauchy-Schwarz}) \\ &= (\mathbf{w}^\top \mathbf{H} \mathbf{w}) \left(1 - \frac{\text{tr}(\mathbf{X}^\top (\mathbf{W}^{-1} + \mathbf{R})^{-1} \mathbf{X})}{\text{tr}(\mathbf{X}^\top (\mathbf{W}^{-1} + \mathbf{I}_n)^{-1} \mathbf{X})} \right) \\ &= (\mathbf{w}^\top \mathbf{H} \mathbf{w}) \left(1 - \frac{\text{ESS}}{n} \right). \end{aligned}$$

Thus $\mathbf{A}$ is nonnegative definite whenever $\text{ESS} \leq n$. Conversely, if $\text{ESS} > n$, taking $\mathbf{w} = \mathbf{H}^{-1} \mathbf{u}$ achieves equality in Cauchy-Schwarz and gives $\mathbf{w}^\top \mathbf{A} \mathbf{w} < 0$, so $\mathbf{A}$ is not nonnegative definite. $\square$

Proposition 3.2.2 adapts the general result from [Ferrer y Vallejos \(2025\)](#) to the SGLMM framework. The condition involves only the inversion of a diagonal matrix ($\mathbf{W}^{-1}$), making it computationally straightforward. Following [Ferrer y Vallejos \(2025\)](#), we also have an equivalent determinant-based characterization:

Proposition 3.2.3. $ESS \leq n$, if and only if, $\det(\mathbf{A}) \geq 0$.

Proof. By the matrix determinant lemma, $\det(\mathbf{A}) = \det(\mathbf{H})(1 - \mathbf{u}^\top \mathbf{H}^{-1} \mathbf{u}/p(\mathbf{X})) = \det(\mathbf{H})(1 - ESS/n)$. Since $\mathbf{H}$ is positive definite, $\det(\mathbf{H}) > 0$, and the result follows from Proposition 3.2.2. $\square$

Let's see the following example using Circulant matrices.

Definition 3.2.1. A symmetric circulant correlation matrix $\mathbf{R} \in \mathbb{R}^{n \times n}$ is determined by its first row $(1, \rho_1, \rho_2, \dots, \rho_{\lfloor n/2 \rfloor}, \dots, \rho_2, \rho_1)$, where $R_{ij} = \rho_{|i-j| \bmod n}$ with $\rho_0 = 1$ and $\rho_k = \rho_{n-k}$ for symmetry.

The eigenvalues of $\mathbf{R}$ are given by

$$\lambda_k = 1 + 2 \sum_{j=1}^{\lfloor (n-1)/2 \rfloor} \rho_j \cos\left(\frac{2\pi jk}{n}\right) + \mathbb{1}_{n \text{ even}} \cdot (-1)^k \rho_{n/2}, \quad k = 0, 1, \dots, n-1, \quad (3.9)$$

with corresponding eigenvectors given by the columns of the DFT matrix $\mathbf{F}$, where $F_{jk} = n^{-1/2} \exp(2\pi ijk/n)$.

For the intercept-only model $\mathbf{X} = \mathbf{1}_n$, we obtain the following characterization.

Proposition 3.2.4. Let $\mathbf{R}$ be a circulant correlation matrix with eigenvalues $\{\lambda_k\}_{k=0}^{n-1}$ as in (3.9), and let $\mathbf{W} = w\mathbf{I}$. In the intercept-only SGLMM,

$$ESS \leq n \iff \lambda_0 \geq 1$$

Proof. Since $\mathbf{X} = \mathbf{1}_n$ and $\mathbf{W} = w\mathbf{I}$, the matrix $\mathbf{A}$ simplifies to

$$\mathbf{A} = w^{-1}\mathbf{I} + \mathbf{R} - \frac{\mathbf{1}_n \mathbf{1}_n^\top}{\mathbf{1}_n^\top (w^{-1}\mathbf{I} + \mathbf{I})^{-1} \mathbf{1}_n} = w^{-1}\mathbf{I} + \mathbf{R} - \frac{w^{-1} + 1}{n} \mathbf{1}_n \mathbf{1}_n^\top.$$

Using the spectral decomposition $\mathbf{R} = \mathbf{F}^* \mathbf{\Lambda} \mathbf{F}$ and noting that $\mathbf{1}_n = \sqrt{n} \mathbf{f}_0$ where $\mathbf{f}_0$ is the first column of $\mathbf{F}^*$, the matrix $\mathbf{A}$ is unitarily equivalent to a diagonal matrix with entries $w^{-1} + \lambda_k$ for $k \geq 1$, and $w^{-1} + \lambda_0 - (w^{-1} + 1) = \lambda_0 - 1$ for $k = 0$. Thus $\mathbf{A} \succeq 0$ if and only if $\lambda_0 \geq 1$.

Given that $\mathbf{R}$ is a correlation matrix, $\lambda_k \geq 0$ for all k , so $w^{-1} + \lambda_k > 0$ for $k \geq 1$ is automatic. Thus $\mathbf{A} \succeq 0$ iff $\lambda_0 \geq 1$ $\square$

Since $\text{tr}(\mathbf{R}) = n$ and $\lambda_0 = \sum_{j=0}^{n-1} \rho_j = 1 + \sum_{j=1}^{n-1} \rho_j$, the condition $\lambda_0 \geq 1$ is equivalent to requiring that the average off-diagonal correlation be non-negative. This holds for all positively correlated structures but can fail when sufficiently strong negative correlations are present.

A particularly tractable case is the first-order periodic autoregressive structure.

Example 3.2.1 (Periodic AR(1)). *Let $\mathbf{R}$ be the circulant matrix with $\psi_k = \psi^{\min(k, n-k)}$ for $\psi \in (0, 1)$ and $n \geq 2$. Then*

$$\lambda_0 = \sum_{j=0}^{n-1} \psi^{\min(j, n-j)} > 1$$

since every summand is positive, and consequently $\text{ESS} \leq n$ in the intercept-only model regardless of w .

Explicitly, $\lambda_0 = (1+\psi)(1-\psi^{n/2})/(1-\psi)$ when n is even, and $\lambda_0 = (1+\psi-2\psi^{(n+1)/2})/(1-\psi)$ when n is odd.

3.3 Exploratory Analysis of ESS Behavior

As we can see in 3.1.1, the ESS depends on four components: (i) the co-variates, (ii) the model response, (iii) the link function, and (iv) the covariance matrix.

Different models and covariance structures lead to different ESS behavior. So we explore this, to see if the non-decreasing pattern found by [Acosta y Vallejos \(2018\)](#) remains in this expression. We consider 3 different response distributions: Gaussian, Binomial and Poisson, in addition with 2 different correlation structures: simple intraclass described by $\mathbf{R}(\rho) = (1 - \rho)\mathbf{I} + \rho\mathbf{J}$ with $\mathbf{J} = \mathbf{1}\mathbf{1}^\top$, $-1/(n-1) < \rho < 1$, and Toeplitz structure, given by

$$R(\rho)_{ij} = \begin{cases} \rho^{|i-j|}, & \text{if } i \neq j, \\ 1, & \text{otherwise,} \end{cases}$$

for $|\rho| < 1$.

For the design matrix, we fix $\mathbf{X} = (\mathbf{1}, \mathbf{x}_2)$, where $\mathbf{1}$ is the intercept and $\mathbf{x}_2$ is a random vector with entries uniformly distributed on $[0, 1]$.

The results (see image 3.1) shows how the ESS changes its behavior depending on the response. We can see that the model with Normal and Binomial response shows

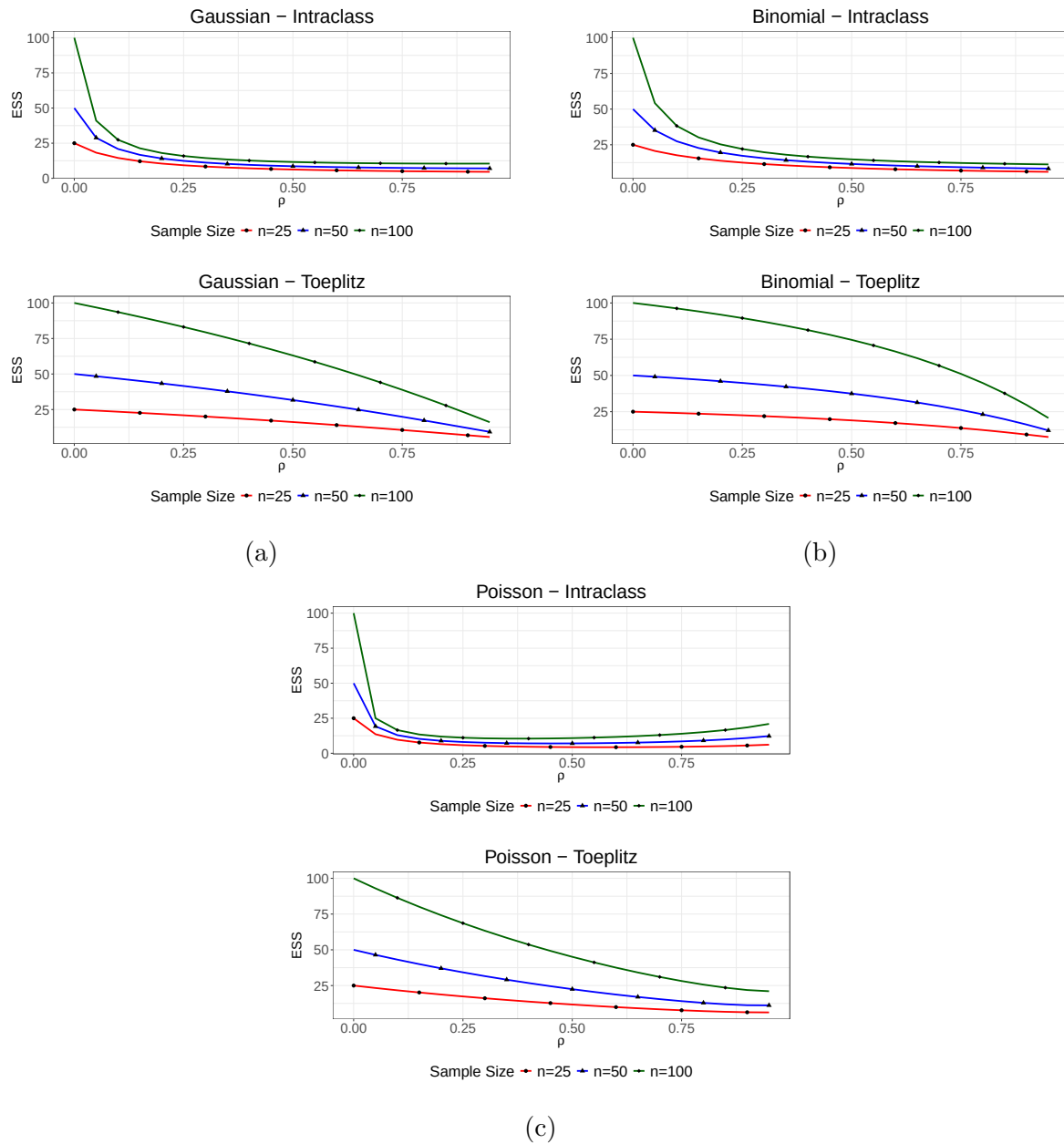


Figure 3.1: ESS as a function of ρ for three response distributions — Gaussian (a), Binomial (b), and Poisson (c) — under intraclass and Toeplitz correlation structures.

a decreasing pattern with both correlation structures, as expected from the constant mean case. In contrast the model with Poisson response shows a non decreasing pattern, mimicking the behavior seen by [Acosta y Vallejos \(2018\)](#) with the regression model.

Chapter 4

Experiments

4.1 Numerical Experiments

In this section, we present three complementary experiments. The first experiment is designed to show that the approximation involved during the ESS construction are accurate for finite samples. The second is designed to show the computational feasibility of INLA-based estimation and explores the behaviour of ESS under different spatial configurations. The third investigates empirically whether the increasing domain asymptotic properties of the ESS established by [Acosta y Vallejos \(2018\)](#) for spatial regression models extend to the SGLMM framework, or whether the evidence suggests otherwise.

These experiments were performed in R. The first experiment uses the TMB framework ([Kristensen *et al.*, 2016](#)), while the second and third employ the INLA package ([Rue *et al.*, 2009](#)) via the SPDE approach ([Lindgren *et al.*, 2011](#)). All code is available at <https://github.com/benjaminbravoo/essforspatialglm>.

4.1.1 Experiment 1

The observed information matrix approximation derived in Equation (3.2) relies on two key assumptions: (i) the validity of the Laplace approximation to the marginal likelihood, and (ii) the negligibility of $O(n^{-1})$ terms in the second derivative. While (i) can be justified under the conditions given in Section 2.4.2, assumption (ii) holds

only asymptotically. Therefore, the accuracy of this approximation in finite samples requires an empirical verification.

To verify this, we leverage TMB's automatic differentiation capabilities (described in section 2.4.4) to compute the empirical observed information matrix. Specifically we conduct a Monte Carlo experiment with M replicates, where we approximate the expected mean observed information matrix as:

$$\hat{\mathcal{I}}_{\text{TMB}}(\boldsymbol{\beta}) \approx \frac{1}{M} \sum_{m=1}^M \hat{\mathcal{I}}^{(m)}(\boldsymbol{\beta}),$$

where $\hat{\mathcal{I}}^{(m)}(\boldsymbol{\beta})$ is the observed information matrix from the m -th simulated dataset. We then compute ESS_{TMB} using the same ratio structure as in equation (3.3), but replacing the analytical observed information matrix with $\hat{\mathcal{I}}_{\text{TMB}}(\boldsymbol{\beta})$.

Two commonly encountered response settings in SGLMMs were considered: binary and count data.

For the binary outcomes we use a Bernoulli response

$$\begin{aligned} Y_i | u_i &\sim \text{Bernoulli}(\pi_i), \quad i = 1, \dots, n. \\ \text{logit}(\pi_i) &= \mathbf{x}_i^\top \boldsymbol{\beta} + u_i, \\ \mathbf{u} &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}(\rho)), \quad \rho \in \mathbb{R}. \end{aligned}$$

and for count outcomes, we considered Negative Binomial response:

$$\begin{aligned} Y_i | u_i &\sim \text{NegBin}(\mu_i, r = 3), \quad i = 1, \dots, n. \\ \log(\mu_i) &= \mathbf{x}_i^\top \boldsymbol{\beta} + u_i, \\ \mathbf{u} &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}(\rho)), \quad \rho \in \mathbb{R}. \end{aligned}$$

and Poisson response:

$$\begin{aligned} Y_i | u_i &\sim \text{Poisson}(\lambda_i), \quad i = 1, \dots, n. \\ \log(\lambda_i) &= \mathbf{x}_i^\top \boldsymbol{\beta} + u_i, \\ \mathbf{u} &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}(\rho)), \quad \rho \in \mathbb{R}. \end{aligned}$$

Here $\mathbf{x}_i^\top$ is the i -th row of $\mathbf{X}$ and $\mathbf{R}(\rho)$ is the spatial correlation matrix. We consider two simple correlation structures: Intraclass (exchangeable) structure: $\mathbf{R}(\rho) = (1 - \rho)\mathbf{I} + \rho\mathbf{J}$ where $\mathbf{J} = \mathbf{1}\mathbf{1}^\top$ and $-1/(n - 1) < \rho < 1$. This implies that $R_{ij}(\rho) = \rho$ for $i \neq j$ and

$R_{ii}(\rho) = 1$, and Toeplitz structure: $R_{ij}(\rho) = \rho^{|i-j|}$ for $i \neq j$ and $R_{ii}(\rho) = 1$, where $|\rho| < 1$.

To evaluate the ESS estimator under these response families and correlation structures, we performed $M = 1000$ independent replicates per configuration. Every replication was constructed in the following way

Step 1. Sample the latent spatial process:

$$\mathbf{u}^{(m)} \sim \mathcal{N}(0, \mathbf{R}(\rho)),$$

Step 2. Obtain the linear predictor.

$$\eta^{(m)} = \mathbf{X}\boldsymbol{\beta}_{\text{true}} + \mathbf{u}^{(m)},$$

where $\boldsymbol{\beta}_{\text{true}}$ and $\mathbf{X}$ remains fixed among all iterations.

Step 3. Get the EF, parameter via inverse links function for each response family:

- Binomial: $\pi_i^{(m)} = \text{logit}^{-1}(\eta_i^{(m)}) = \frac{\exp(\eta_i^{(m)})}{1 + \exp(\eta_i^{(m)})}$.
- Negative Binomial: $\mu_i^{(m)} = \exp(\eta_i^{(m)})$.
- Poisson: $\lambda_i^{(m)} = \exp(\eta_i^{(m)})$.

Step 4. Generate responses:

- Binomial: $y_i^{(m)} \sim \text{Bernoulli}(\pi_i^{(m)})$.
- Negative Binomial: $y_i^{(m)} \sim \text{NegBin}(\mu_i^{(m)}, r = 3)$.
- Poisson: $y_i^{(m)} \sim \text{Poisson}(\lambda_i^{(m)})$.

and for each one, we compute the observed information matrix at the true $\boldsymbol{\beta}$ parameter, using TMB (Section 2.4.4).

$$\hat{\mathcal{I}}^{(m)}(\boldsymbol{\beta}_{\text{true}}) = - \left. \frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \right|_{\boldsymbol{\beta} = \boldsymbol{\beta}_{\text{true}}}.$$

Finally, we estimate the mean observed information as the mean of the observed information matrices over all M runs.

$$\hat{\mathcal{I}}_{\text{TMB}}(\boldsymbol{\beta}_{\text{true}}) = \frac{1}{M} \sum_{m=1}^M \hat{\mathcal{I}}^{(m)}(\boldsymbol{\beta}_{\text{true}}) \xrightarrow{M \rightarrow \infty} \mathbb{E}_{\mathbf{Y}} \left[- \frac{\partial^2 \ell(\boldsymbol{\beta}_{\text{true}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \right].$$

Using this approximation, we construct a consistent estimator for the ESS: the quotient of $\widehat{\mathcal{I}}_{\text{TMB}}$ using $\mathbf{R} = \mathbf{R}(\rho)$ over the one computed with $\mathbf{R} = \mathbf{I}$.

Note that, although we evaluate the Hessian at fixed $\boldsymbol{\beta}_{\text{true}}$, the observed information varies across replications (it depends on the observed data $\mathbf{y}^{(m)}$). Averaging over replications is therefore necessary.

For each configuration, $n = 100$ observations were generated with $\boldsymbol{\beta} = (1, 0.25)^\top$ and a design matrix $\mathbf{X}$ whose first column is an intercept and whose second column is drawn from a Uniform(0, 1) distribution. The analytical and TMB-based ESS estimates were then compared using two ratios

$$\text{ESS Ratio} = \frac{\text{ESS}_{\text{analytical}}}{\text{ESS}_{\text{TMB}}},$$

and the relative error

$$\text{Relative Error} = \frac{|\text{ESS}_{\text{analytical}} - \text{ESS}_{\text{TMB}}|}{\text{ESS}_{\text{TMB}}} \times 100.$$

Figures 4.1, 4.2, 4.3, 4.4, 4.5 and 4.6 compare the ESS obtained from our approximation with that obtained through TMB. Each figure shows a comparison between analytical and TMB-based ESS results, followed by the absolute relative error, the ESS ratio and a scatterplot between them.

Tables 4.1, 4.2 and 4.3 show the detailed numerical results for all evaluated scenarios. TMB-based estimates are reported along with standard errors in parentheses. The standard error for ESS_{TMB} was obtained using the delta method. Specifically,

$$\sigma_{\text{ESS}_{\text{TMB}}} = \text{ESS} \times \sqrt{\left(\frac{\sigma_{\bar{T}_{\rho^*}}}{\bar{T}_{\rho^*}}\right)^2 + \left(\frac{\sigma_{\bar{T}_0}}{\bar{T}_0}\right)^2},$$

where $\bar{T}_{\rho^*}$ corresponds to $\text{tr}(\widehat{\mathcal{I}}_{\text{TMB}}(\boldsymbol{\beta}))$ when $\rho = \rho^*$.

Table 4.1 shows notable differences in behavior between the two correlation structures for the binomial model. Overall the two correlations structures performs well. Intraclass correlation structure reveals a relative error consistently ranging from approximately 1% to 4% across all values of ρ and Toeplitz correlation structure shows lower relative errors increasing monotonically with ρ up to 3%.

In Table 4.2, we see a completely different phenomenon between the two correlation structures. Here, in contrast to the binomial case, intraclass structure shows non-monotonic behavior, remaining below 5% for all $\rho \leq 0.30$ and reaching a minimum

ρ	Intraclass					Toeplitz				
	$\bar{T}_\rho$	ESS (TMB)	ESS (Ana.)	Ratio	Rel. Err.%	Trace	ESS (TMB)	ESS (Ana.)	Ratio	Rel. Err.%
0.00	19.126 (0.165)	100.000 (1.217)	100.000	1.000	0.00	19.126 (0.165)	100.000 (1.217)	100.000	1.000	0.00
0.05	11.545 (0.352)	60.364 (1.911)	59.106	0.979	2.08	18.838 (0.172)	98.492 (1.237)	98.494	1.000	0.002
0.10	8.499 (0.334)	44.438 (1.788)	43.070	0.969	3.08	18.557 (0.186)	97.027 (1.283)	96.880	0.998	0.15
0.15	6.846 (0.308)	35.793 (1.638)	34.525	0.965	3.54	18.229 (0.210)	95.309 (1.371)	95.145	0.998	0.17
0.20	5.801 (0.302)	30.333 (1.599)	29.228	0.964	3.64	17.882 (0.215)	93.496 (1.381)	93.275	0.998	0.24
0.25	5.084 (0.299)	26.584 (1.582)	25.631	0.964	3.59	17.498 (0.232)	91.486 (1.446)	91.253	0.997	0.26
0.30	4.567 (0.306)	23.881 (1.614)	23.034	0.965	3.55	17.098 (0.256)	89.397 (1.545)	89.060	0.996	0.38
0.35	4.168 (0.309)	21.793 (1.627)	21.077	0.967	3.28	16.642 (0.272)	87.013 (1.606)	86.674	0.996	0.39
0.40	3.866 (0.322)	20.212 (1.693)	19.554	0.967	3.26	16.165 (0.286)	84.518 (1.664)	84.067	0.995	0.53
0.45	3.616 (0.336)	18.904 (1.765)	18.337	0.970	3.00	15.635 (0.307)	81.746 (1.753)	81.206	0.993	0.66
0.50	3.416 (0.360)	17.858 (1.886)	17.347	0.971	2.86	15.046 (0.320)	78.666 (1.805)	78.054	0.992	0.78
0.55	3.245 (0.371)	16.968 (1.947)	16.528	0.974	2.60	14.403 (0.338)	75.306 (1.880)	74.563	0.990	0.99
0.60	3.110 (0.385)	16.262 (2.016)	15.841	0.974	2.59	13.681 (0.345)	71.532 (1.907)	70.676	0.988	1.20
0.65	2.984 (0.401)	15.601 (2.099)	15.259	0.978	2.19	12.855 (0.350)	67.212 (1.920)	66.323	0.987	1.32
0.70	2.882 (0.421)	15.067 (2.203)	14.762	0.980	2.02	11.963 (0.359)	62.546 (1.951)	61.416	0.982	1.81
0.75	2.836 (0.423)	14.827 (2.213)	14.335	0.967	3.32	10.872 (0.381)	56.842 (2.050)	55.846	0.982	1.75
0.80	2.720 (0.460)	14.220 (2.408)	13.964	0.982	1.79	9.684 (0.365)	50.633 (1.957)	49.472	0.977	2.29
0.85	2.657 (0.475)	13.894 (2.484)	13.642	0.982	1.81	8.278 (0.366)	43.283 (1.947)	42.122	0.973	2.68
0.90	2.600 (0.493)	13.595 (2.580)	13.361	0.983	1.72	6.624 (0.347)	34.633 (1.837)	33.584	0.970	3.03
0.95	2.550 (0.512)	13.331 (2.677)	13.116	0.984	1.62	4.674 (0.355)	24.437 (1.867)	23.649	0.968	3.22

Table 4.1: ESS comparison between TMB and analytical approximation for the Binomial model. Standard deviations across simulation replicates are shown in parentheses.

ρ	Intraclass					Toeplitz				
	$\bar{T}_\rho$	ESS (TMB)	ESS (Ana.)	Ratio	Rel. Err.%	Trace	ESS (TMB)	ESS (Ana.)	Ratio	Rel. Err.%
0.00	73.793 (1.785)	100.000 (3.420)	100.000	1.000	0.00	73.793 (1.785)	100.000 (3.420)	100.000	1.000	0.00
0.05	23.138 (0.497)	31.355 (1.014)	28.373	0.905	9.51	70.028 (1.703)	94.897 (3.255)	93.983	0.990	0.96
0.10	15.434 (0.451)	20.916 (0.793)	19.002	0.909	9.15	66.265 (1.609)	89.798 (3.077)	88.177	0.982	1.81
0.15	12.392 (0.495)	16.793 (0.784)	15.445	0.920	8.03	62.655 (1.410)	84.906 (2.805)	82.574	0.973	2.75
0.20	10.843 (0.544)	14.694 (0.818)	13.666	0.930	6.99	59.004 (1.349)	79.958 (2.661)	77.166	0.965	3.49
0.25	9.942 (0.630)	13.473 (0.913)	12.673	0.941	5.94	55.437 (1.229)	75.124 (2.465)	71.946	0.958	4.23
0.30	9.404 (0.728)	12.744 (1.033)	12.103	0.950	5.03	51.923 (1.152)	70.363 (2.309)	66.909	0.951	4.91
0.35	9.066 (0.824)	12.285 (1.155)	11.794	0.960	4.00	48.509 (1.038)	65.737 (2.122)	62.051	0.944	5.61
0.40	8.892 (0.924)	12.050 (1.286)	11.666	0.968	3.19	45.099 (0.977)	61.115 (1.984)	57.365	0.939	6.14
0.45	8.832 (1.043)	11.969 (1.443)	11.672	0.975	2.48	41.761 (0.912)	56.592 (1.844)	52.851	0.934	6.61
0.50	8.837 (1.172)	11.975 (1.614)	11.788	0.984	1.56	38.473 (0.865)	52.136 (1.721)	48.504	0.930	6.97
0.55	8.939 (1.314)	12.114 (1.804)	12.000	0.991	0.94	35.260 (0.801)	47.782 (1.585)	44.325	0.928	7.24
0.60	9.109 (1.479)	12.344 (2.027)	12.303	0.997	0.34	32.122 (0.781)	43.530 (1.492)	40.314	0.926	7.39
0.65	9.351 (1.658)	12.672 (2.268)	12.695	1.002	0.18	29.120 (0.737)	39.462 (1.381)	36.474	0.924	7.57
0.70	9.642 (1.907)	13.066 (2.604)	13.181	1.009	0.88	26.058 (0.755)	35.312 (1.332)	32.812	0.929	7.08
0.75	9.771 (2.235)	13.240 (3.046)	13.769	1.040	3.99	23.236 (0.785)	31.488 (1.309)	29.337	0.932	6.83
0.80	10.429 (2.543)	14.133 (3.463)	14.473	1.024	2.40	20.384 (0.878)	27.623 (1.364)	26.070	0.944	5.62
0.85	10.977 (2.909)	14.875 (3.958)	15.312	1.029	2.94	17.777 (0.991)	24.090 (1.464)	23.046	0.957	4.33
0.90	11.672 (3.442)	15.817 (4.680)	16.313	1.031	3.13	15.304 (1.252)	20.740 (1.770)	20.337	0.981	1.94
0.95	12.552 (4.099)	17.009 (5.570)	17.514	1.030	2.97	13.202 (1.876)	17.890 (2.579)	18.143	1.014	1.41

Table 4.2: ESS comparison between TMB and analytical approximation for Negative Binomial model. Standard deviations across simulation replicates are shown in parentheses.

ρ	Intraclass					Toeplitz				
	$\bar{T}_\rho$	ESS (TMB)	ESS (Ana.)	Ratio	Rel. Err.%	Trace	ESS (TMB)	ESS (Ana.)	Ratio	Rel. Err.%
0.00	97.832 (1.764)	100.000 (2.550)	100.000	1.000	0.00	97.832 (1.764)	100.000 (2.550)	100.000	1.000	0.00
0.05	25.973 (0.570)	26.549 (0.754)	26.113	0.984	1.64	91.295 (1.688)	93.318 (2.410)	93.143	0.998	0.19
0.10	17.502 (0.605)	17.890 (0.698)	17.618	0.985	1.52	85.027 (1.557)	86.912 (2.233)	86.656	0.997	0.29
0.15	14.369 (0.726)	14.687 (0.788)	14.501	0.987	1.27	79.175 (1.452)	80.930 (2.081)	80.516	0.995	0.51
0.20	12.875 (0.862)	13.161 (0.913)	12.999	0.988	1.22	73.522 (1.374)	75.152 (1.951)	74.701	0.994	0.60
0.25	12.096 (1.017)	12.364 (1.063)	12.211	0.988	1.24	68.164 (1.263)	69.675 (1.801)	69.194	0.993	0.69
0.30	11.700 (1.222)	11.959 (1.267)	11.812	0.988	1.23	63.086 (1.251)	64.484 (1.728)	63.978	0.992	0.78
0.35	11.555 (1.445)	11.811 (1.493)	11.660	0.987	1.28	58.283 (1.127)	59.575 (1.575)	59.040	0.991	0.90
0.40	11.605 (1.675)	11.862 (1.726)	11.684	0.985	1.50	53.730 (1.067)	54.921 (1.473)	54.368	0.990	1.01
0.45	11.829 (1.949)	12.091 (2.004)	11.850	0.980	1.99	49.385 (1.033)	50.480 (1.394)	49.954	0.990	1.04
0.50	12.156 (2.294)	12.425 (2.356)	12.140	0.977	2.30	45.313 (1.036)	46.317 (1.349)	45.791	0.989	1.14
0.55	12.645 (2.688)	12.925 (2.757)	12.548	0.971	2.92	41.542 (1.026)	42.463 (1.298)	41.875	0.986	1.38
0.60	13.309 (3.172)	13.604 (3.251)	13.078	0.961	3.87	37.967 (1.091)	38.809 (1.317)	38.207	0.984	1.55
0.65	14.169 (3.787)	14.483 (3.879)	13.741	0.949	5.12	34.836 (1.151)	35.608 (1.341)	34.791	0.977	2.29
0.70	15.188 (4.624)	15.524 (4.735)	14.557	0.938	6.23	31.744 (1.347)	32.447 (1.496)	31.637	0.975	2.50
0.75	15.885 (5.739)	16.237 (5.874)	15.556	0.958	4.20	29.328 (1.637)	29.978 (1.758)	28.765	0.960	4.05
0.80	18.173 (7.308)	18.576 (7.478)	16.782	0.903	9.66	27.001 (2.109)	27.600 (2.212)	26.209	0.950	5.04
0.85	20.529 (9.496)	20.984 (9.714)	18.300	0.872	12.79	25.529 (2.964)	26.094 (3.066)	24.034	0.921	7.90
0.90	23.889 (13.285)	24.418 (13.587)	20.209	0.828	17.24	24.873 (4.595)	25.425 (4.720)	22.367	0.880	12.03
0.95	29.682 (21.094)	30.340 (21.569)	22.659	0.747	25.32	26.060 (9.078)	26.638 (9.291)	21.550	0.809	19.10

Table 4.3: ESS comparison between TMB and analytical approximation for Poisson model. Standard deviations across simulation replicates are shown in parentheses.

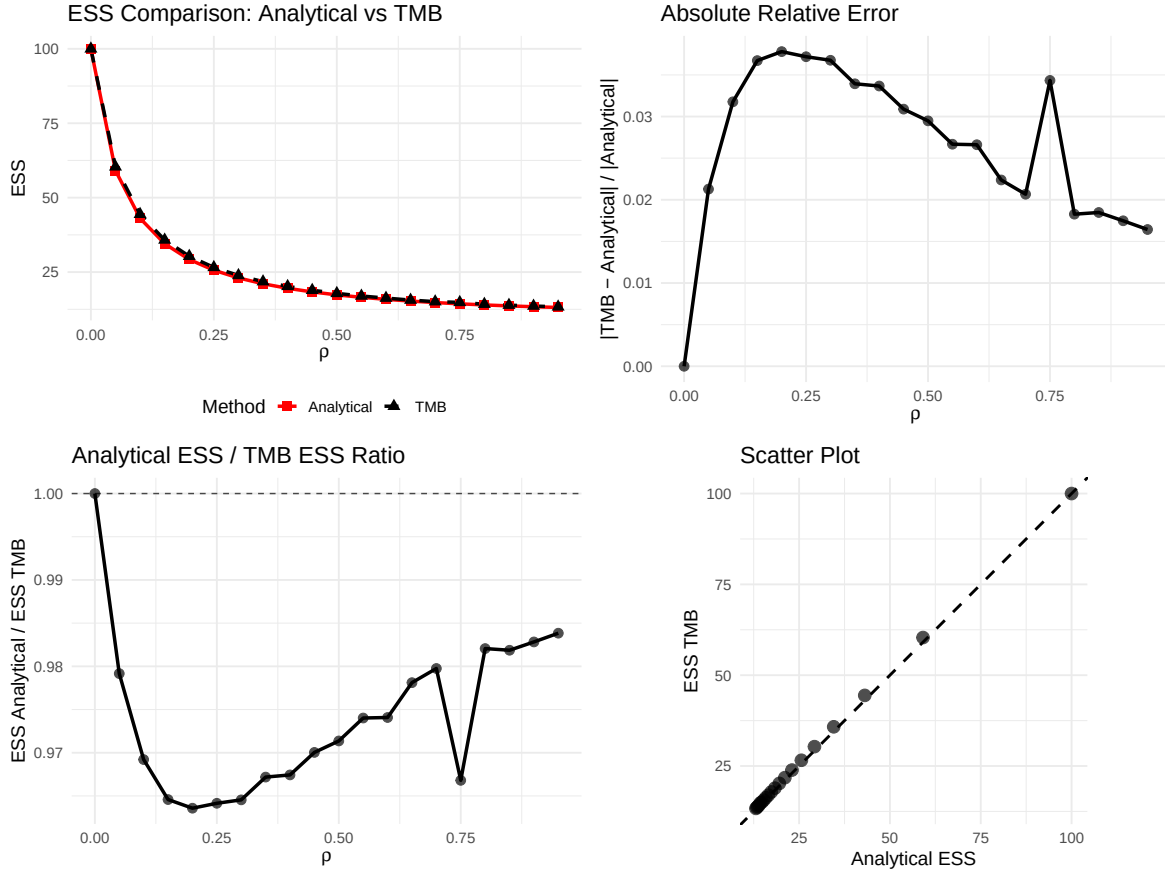


Figure 4.1: Binomial Intraclass: analytical vs TMB-based ESS.

of approximately 0.2% at $\rho = 0.65$ before increasing slightly at higher correlations. For the Toeplitz structure, we see that the relative errors increase monotonically until $\rho = 0.55$ —hitting around 7% and then start dropping to 1% at $\rho = 0.95$.

The Poisson model shown in Table 4.3 behaves well in most cases. For both correlation structures the errors stay low when ρ is small, more precisely, errors stay below 3% up to $\rho = 0.55$. For the intraclass structure, relative errors increase gradually from $\rho = 0.55$ to $\rho = 0.75$, reaching approximately 6%. However, for high values of ρ , errors spike sharply, hitting over 25% at $\rho = 0.95$. Toeplitz structure does better: errors are under 3% for $\rho \leq 0.7$ and only increase substantially when $\rho \geq 0.9$, getting to around 19% at $\rho = 0.95$.

Overall, these results show excellent agreement between both approaches across the vast majority of tested scenarios. For both the Binomial and Negative Binomial models, the approximation maintains high accuracy throughout the entire range of correlations considered. However, in the Poisson model under extremely high correlations, the

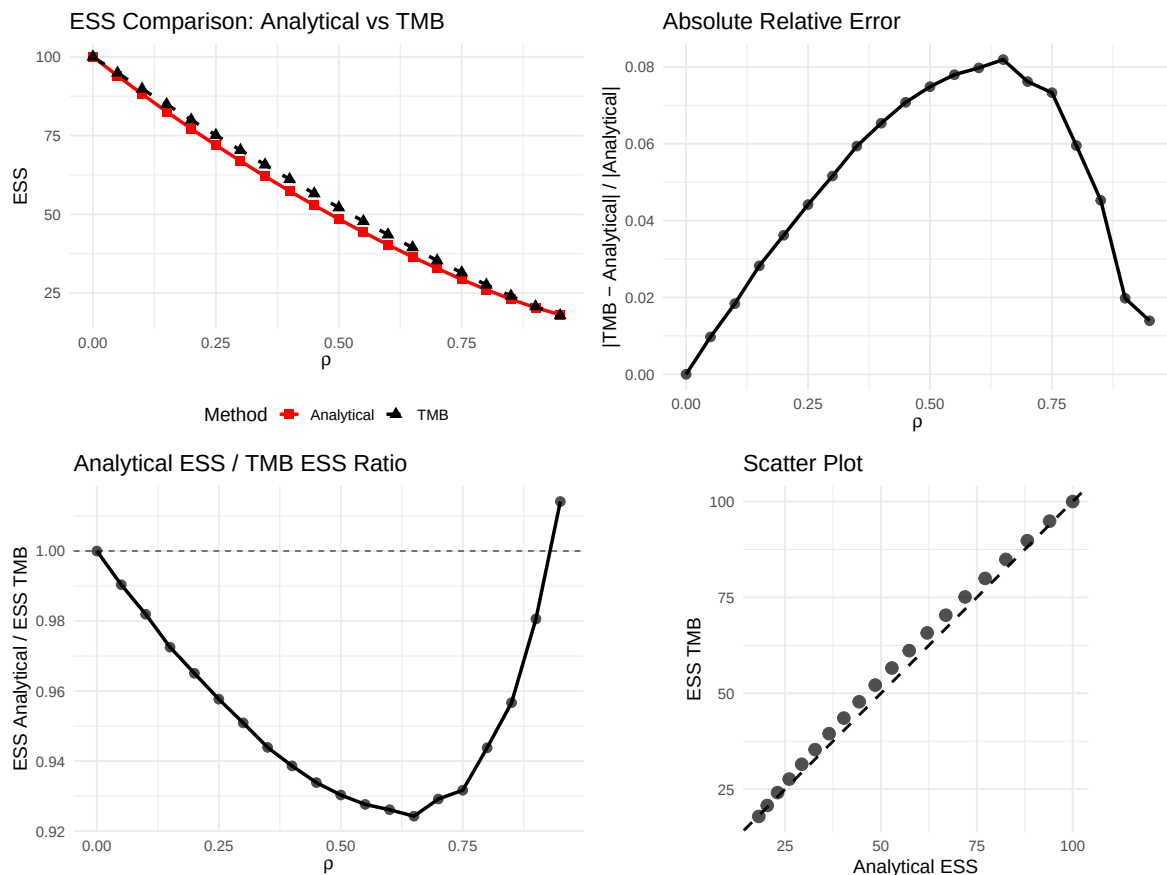


Figure 4.2: Binomial Toeplitz: analytical vs TMB-based ESS.

approximation begins to underestimate the loss of information, with relative errors increasing substantially.

In practical terms, these findings strongly support the use of our analytic ESS formula, as the observed limitations are confined to extreme correlation values $\rho > 0.85$ that are uncommon in most applied settings.

4.1.2 Experiment 2

Although INLA's computational efficiency makes it preferable over MCMC (Rue *et al.*, 2009; Lindgren *et al.*, 2011), the reliance on multiple approximation layers may introduce errors in derived quantities like ESS, making simulation-based validation necessary. Therefore, we conducted an extensive simulation study, generating spatial data from Poisson, Negative Binomial and Gamma models. Specifically, we pursue two objectives: (i) validating that INLA provides sufficiently accurate ESS estimates

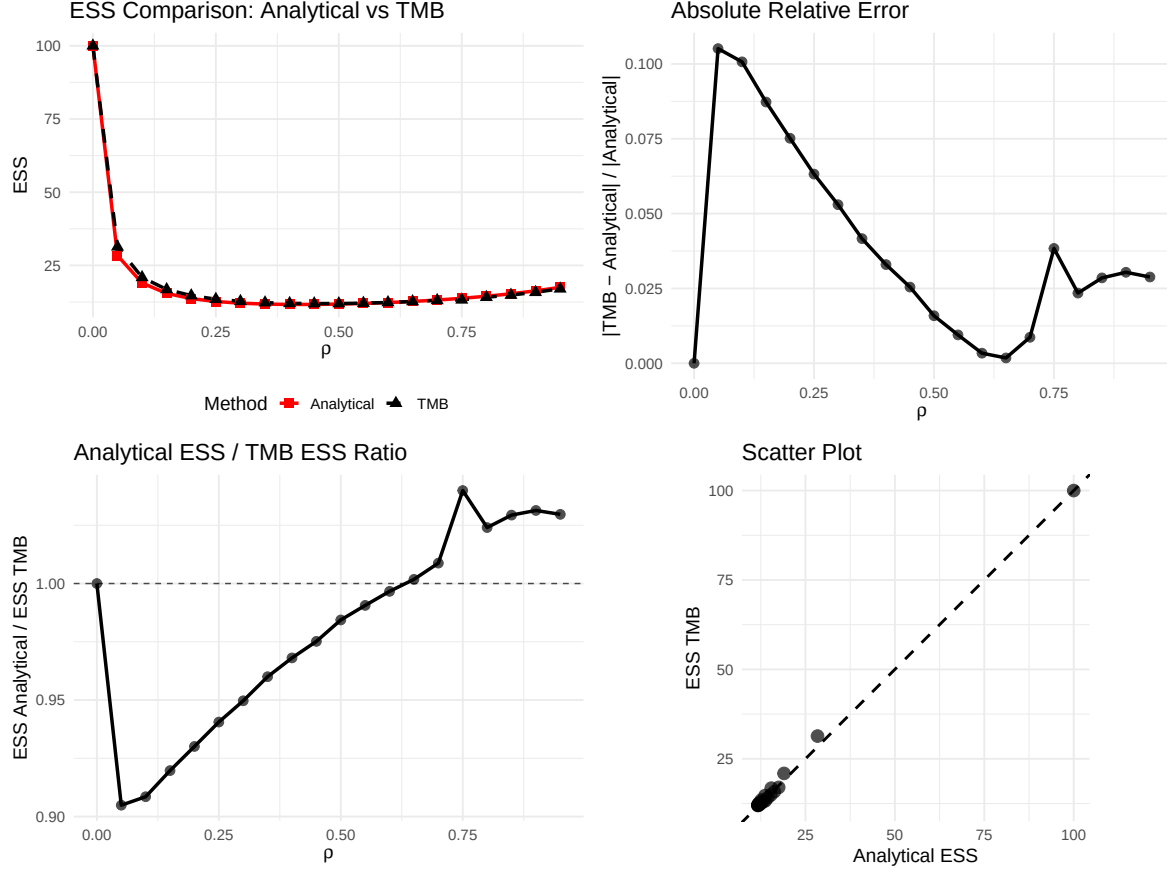


Figure 4.3: Negative Binomial Intraclass: analytical vs TMB-based ESS.

across different likelihood families and spatial configurations, and (ii) quantifying the approximation error introduced by INLA's analytical approximations.

Data were generated on a regular 16×16 grid (256 points) over the unit square $\mathcal{D} = [0, 1] \times [0, 1]$. Let $\mathcal{S} = \{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ denote this set of spatial locations, where $\mathbf{s}_i = (s_{i,x}, s_{i,y}) \in \mathcal{D}$ with $n = 256$. We used a hierarchical spatial log-Gaussian model to generate the data

$$\begin{aligned} \mathbf{w} &\sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{R}(\phi, \nu)), \\ \boldsymbol{\eta} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \\ \boldsymbol{\mu} &= \exp(\boldsymbol{\eta}), \\ Y_i &\sim EF(\mu_i; \theta), \end{aligned}$$

where $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^\top$, $\mathbf{u} = (u(s_1), \dots, u(s_n))^\top$ is the spatial random effect with Matérn correlation matrix $\mathbf{R}(\phi, \nu)$ as given in Definition 2.1.7, $\mathbf{X}$ is an $n \times 2$ design matrix constructed with an intercept column and a spatial covariate $s_{i,x} \cdot s_{i,y}$, $\boldsymbol{\beta} =$

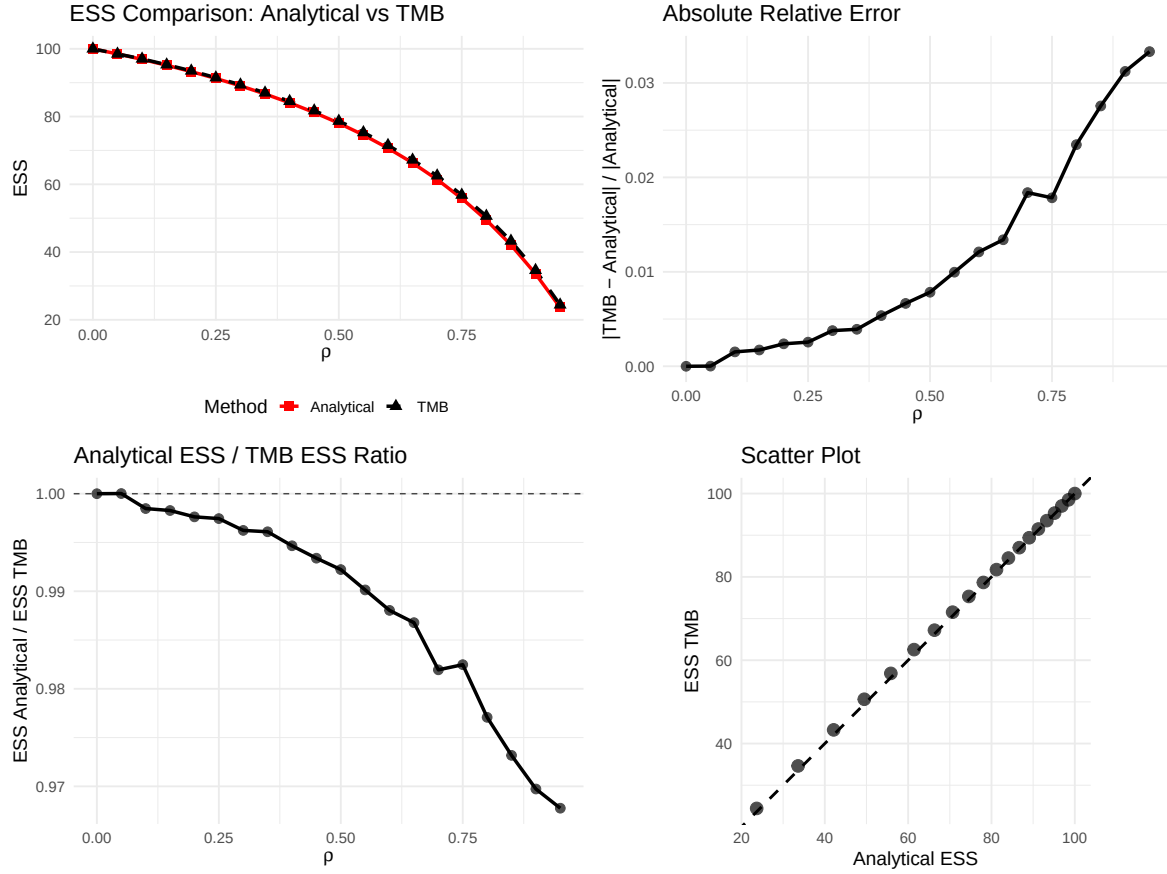


Figure 4.4: Negative Binomial Toeplitz: analytical vs TMB-based ESS.

$(\beta_0, \beta_1)^\top$ are regression coefficients, and $EF(\cdot; \theta)$ denotes the likelihood family with additional parameter $\theta \in \mathbb{R}^+$ when applicable (Definition 2.5).

Across all scenarios, we fixed the regression coefficients at $\beta = (2, 0.1)^\top$ and the spatial variance at $\sigma^2 = 1$. The nugget effect was set to zero, $\tau^2 = 0$, as the SPDE approximation in INLA cannot reliably identify nugget and spatial variance components jointly without exact replication at observation locations.

In the case of Negative Binomial and Gamma, we fixed the dispersion parameter, $r = 5$ and $\alpha_\gamma = 2$ respectively, at their true values during INLA model fitting. This isolates the approximation error in ESS and provides a best-case scenario for INLA performance: if ESS estimation is inaccurate even with known dispersion parameters, this indicates fundamental issues with spatial parameter approximation or ESS calculation.

The main experimental factors are the smoothness parameter ν and the spatial scale parameter ϕ . We considered three values for $\nu \in \{0.5, 0.75, 1\}$ and for each one, we defined four values of ϕ based on the practical range $r = \sqrt{8\nu}\phi$, which represents

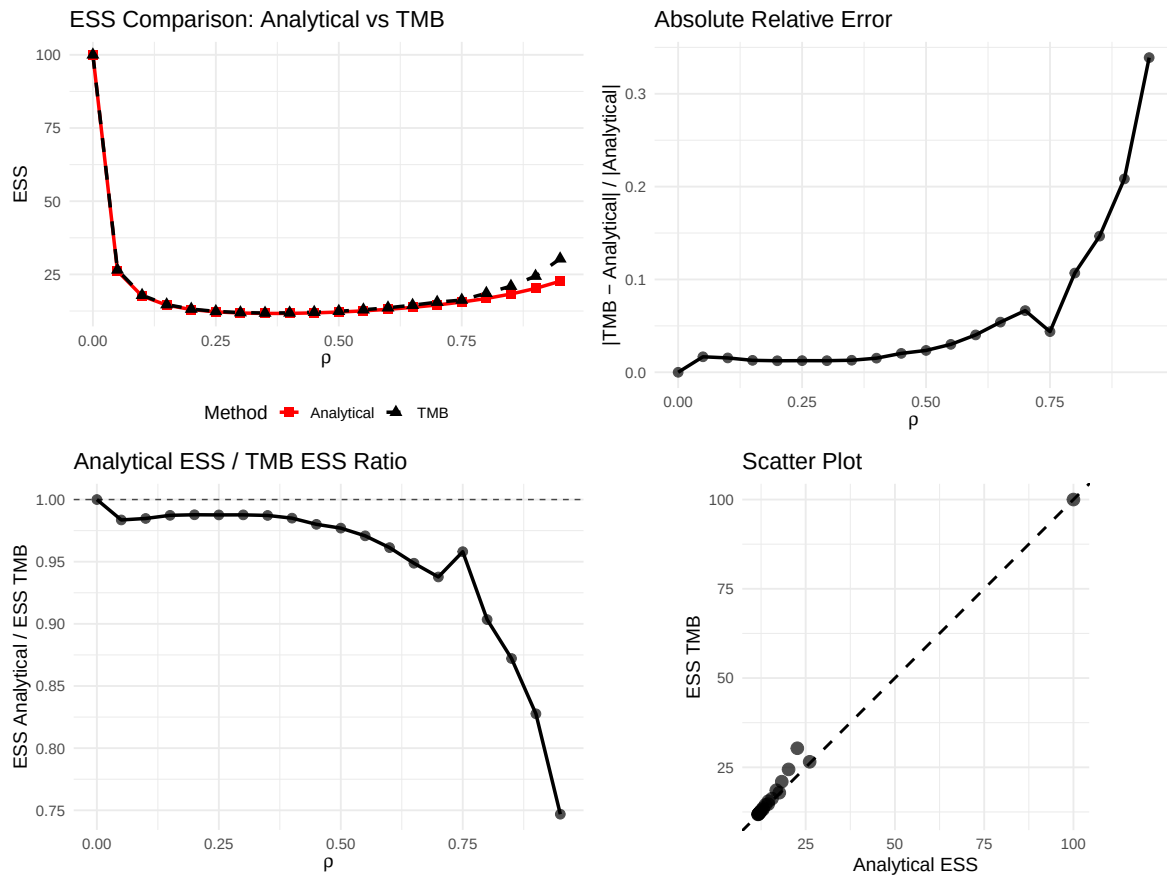


Figure 4.5: Poisson Intraclass: analytical vs TMB-based ESS.

the effective distance of spatial correlation. The corresponding values of ϕ for each combination of ν are shown in Table 4.4, yielding 12 spatial configurations per family and 36 distinct scenarios in total, each replicated $M = 500$ times with independent random seeds.

 Table 4.4: Values of ϕ for each ν to achieve target practical ranges.

ν	Practical Range (% of domain)			
	15%	25%	35%	45%
0.5	0.075	0.125	0.175	0.225
0.75	0.053	0.088	0.124	0.159
1	0.043	0.072	0.101	0.130

With respect to the INLA configuration, we constructed the mesh adapting its resolution to the true practical range r of each configuration. Specifically, the inner maximum edge length was set to $r/5$, the outer maximum edge to $r/2.5$, and the cutoff distance

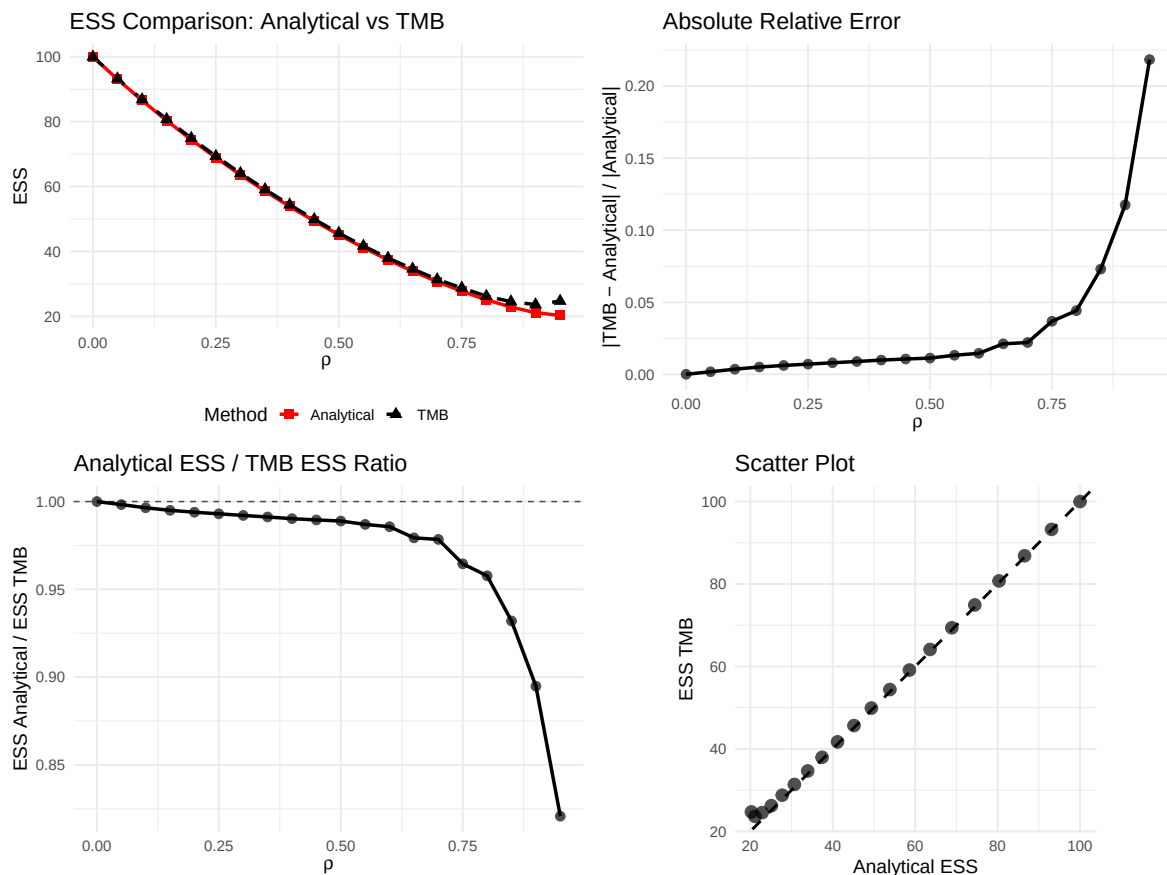


Figure 4.6: Poisson Toeplitz: analytical vs TMB-based ESS.

to $r/15$, with offset values of 0.1 and 0.2 for the inner and outer extensions respectively. The spatial hyperparameters were assigned PC priors with $P(r < 0.3) = 0.5$ for the range and $P(\sigma > 1) = 0.5$ for the marginal standard deviation, in order to evaluate INLA performance under favorable prior specification.

All computations were performed in R (version 4.4.3) using the INLA package for model fitting.¹

For each scenario, the true ESS was computed using the known data-generating parameters and compared against the INLA-based estimate. Let n_i^* and $\hat{n}_i^*$ denote the true and estimated ESS for replicate i , respectively, and let n^* and $\hat{n}^*$ denote their averages over the B replicates. Estimation accuracy was quantified via the bias $\text{Bias} = \hat{n}^* - n^*$,

¹Software: R 4.4.3, INLA 25.04.16, MASS 7.3-65. Hardware: AMD Ryzen 5 5600H (6 cores/12 threads), 16GB RAM. Platform: Windows 10 x64. Total computation time: approximately 10 hours

the root mean squared error

$$\text{RMSE} = \sqrt{\frac{1}{M} \sum_{i=1}^M (\hat{n}_i^* - n_i^*)^2},$$

the mean relative error

$$\text{MRE} = \frac{1}{M} \sum_{i=1}^M \frac{|\hat{n}_i^* - n_i^*|}{n_i^*},$$

the mean absolute error

$$\frac{1}{M} \sum_{i=1}^M |\hat{n}_i^* - n_i^*|,$$

and coverage C_{30} , defined as the proportion of replicates for which

$$\frac{|\hat{n}_i^* - n_i^*|}{n_i^*} < 0.3.$$

Table 4.5 presents the mean parameter estimates and ESS across all replications. The regression coefficients β_0 and β_1 are well recovered across all scenarios, with estimates centered on their true values; covariance parameter, however, show larger estimation errors: σ^2 is overestimated by 3–5% for $\nu = 0.5$, growing progressively to 30% for $\nu = 1.0$. The range parameter $\hat{\phi}$ is also overestimated across all scenarios, with bias increasing from 6–25% at $\nu = 0.5$ to 13–43% at $\nu = 1.0$, and variability rising substantially with the range (SD from ≈ 0.01 to ≈ 0.09). The true n^* decreases rapidly with ϕ and stabilizes for larger ranges, with little variability across replicates ($SD < 1.1$). The estimated $\hat{n}^*$ mirrors this pattern but with systematic underestimation and much larger variability (SD between 1.9 and 12.7).

Table 4.6 presents the precision metrics for the estimated ESS. The MAE ranges from 2.33 to 13.30 observations over true ESS values between 11 and 49, corresponding to relative errors of 19–29%. The bias is consistently negative across all scenarios, indicating that INLA systematically underestimates the ESS. C_{30} exceeds 80% when smoothness is low and ranges are large, but falls to around 50% for smoother fields with intermediate ranges. The underestimation is consistent across all three families, suggesting a systematic rather than random source of error.

Overall, INLA-based ESS estimation performs satisfactorily under low smoothness and moderate spatial dependence, where coverage exceeds 80% and relative errors remain below 25%. For smoother fields or stronger spatial dependence, precision deteriorates, though the systematic underestimation observed across all scenarios provides a consistent and predictable picture of INLA’s behavior in this context.

ν	ϕ	Poisson					Neg. Binomial					Gamma							
		$\hat{\phi}$	$\hat{\sigma}^2$	$\hat{\beta}_0$	$\hat{\beta}_1$	n^*	$\widehat{n^*}$	$\hat{\phi}$	$\hat{\sigma}^2$	$\hat{\beta}_0$	$\hat{\beta}_1$	n^*	$\widehat{n^*}$	$\hat{\phi}$	$\hat{\sigma}^2$	$\hat{\beta}_0$	$\hat{\beta}_1$	n^*	$\widehat{n^*}$
0.5	0.075	0.0862 (0.0152)	0.9703 (0.1614)	1.9799 (0.2155)	0.1119 (0.2146)	49.01 (0.97)	42.10 (11.37)	0.0893 (0.0177)	0.9714 (0.1848)	1.9817 (0.2242)	0.1081 (0.2198)	46.70 (0.78)	38.86 (12.32)	0.0913 (0.0205)	0.9487 (0.1874)	2.0001 (0.2214)	0.1141 (0.2251)	46.99 (1.06)	38.45 (12.69)
	0.125	0.1401 (0.0289)	0.9549 (0.2217)	2.0229 (0.3142)	0.0730 (0.2793)	24.22 (0.78)	22.29 (6.12)	0.1477 (0.0361)	0.9515 (0.2407)	2.0261 (0.3134)	0.0743 (0.2855)	23.00 (0.62)	20.52 (7.27)	0.1547 (0.0435)	0.9608 (0.2544)	2.0336 (0.3236)	0.0804 (0.2915)	23.16 (0.84)	20.13 (7.74)
	0.175	0.2030 (0.0447)	0.9711 (0.2561)	1.9765 (0.3916)	0.1122 (0.3078)	16.24 (0.76)	14.71 (3.67)	0.2116 (0.0560)	0.9786 (0.2820)	1.9769 (0.4021)	0.1153 (0.3191)	15.42 (0.61)	13.89 (4.54)	0.2186 (0.0710)	0.9716 (0.3210)	1.9865 (0.4067)	0.1135 (0.3242)	15.55 (0.89)	14.06 (4.95)
	0.225	0.2387 (0.0598)	0.9655 (0.3206)	1.9677 (0.4268)	0.1286 (0.3441)	12.54 (0.62)	12.78 (3.48)	0.2535 (0.0784)	0.9676 (0.3527)	1.9724 (0.4339)	0.1302 (0.3501)	11.93 (0.50)	11.95 (3.75)	0.2560 (0.0933)	0.9798 (0.3802)	1.9708 (0.4296)	0.1351 (0.3534)	12.02 (0.70)	12.37 (4.69)
	0.061	0.0710 (0.0105)	0.9716 (0.1719)	1.9838 (0.2258)	0.1018 (0.2119)	48.61 (0.94)	40.05 (8.89)	0.0730 (0.0127)	0.9666 (0.1988)	1.9834 (0.2270)	0.1040 (0.2143)	46.32 (0.76)	37.21 (9.73)	0.0748 (0.0154)	0.9486 (0.1933)	2.0023 (0.2293)	0.1037 (0.2213)	46.61 (1.02)	37.16 (12.07)
0.75	0.102	0.1280 (0.0244)	1.0174 (0.2866)	2.0084 (0.3241)	0.0805 (0.2897)	23.37 (0.77)	18.37 (4.19)	0.1298 (0.0283)	0.9991 (0.2872)	2.0139 (0.3234)	0.0800 (0.2877)	22.22 (0.61)	17.57 (4.78)	0.1318 (0.0331)	0.9959 (0.2876)	2.0217 (0.3232)	0.0837 (0.2912)	22.38 (0.83)	17.78 (5.74)
	0.143	0.1976 (0.0427)	1.1273 (0.3705)	1.9743 (0.3989)	0.1124 (0.3148)	15.47 (0.74)	11.69 (2.52)	0.1987 (0.0526)	1.1035 (0.4016)	1.9751 (0.4094)	0.1107 (0.3321)	14.72 (0.60)	11.43 (3.07)	0.1956 (0.0573)	1.0661 (0.3840)	1.9794 (0.4127)	0.1176 (0.3261)	14.86 (0.87)	11.98 (3.54)
	0.184	0.2413 (0.0602)	1.1486 (0.4727)	1.9687 (0.4329)	0.1281 (0.3698)	11.86 (0.60)	10.00 (2.08)	0.2434 (0.0781)	1.1128 (0.4795)	1.9766 (0.4369)	0.1248 (0.3684)	11.31 (0.49)	9.89 (2.64)	0.2307 (0.0803)	1.0673 (0.4668)	1.9717 (0.4429)	0.1338 (0.3860)	11.41 (0.70)	10.59 (3.09)
	0.053	0.0601 (0.0080)	0.9887 (0.1795)	1.9829 (0.2184)	0.0994 (0.2092)	47.98 (0.95)	40.71 (8.38)	0.0615 (0.0100)	0.9740 (0.1985)	1.9909 (0.2226)	0.0922 (0.2136)	45.71 (0.76)	38.10 (9.28)	0.0625 (0.0121)	0.9646 (0.2057)	2.0015 (0.2237)	0.1003 (0.2160)	46.00 (1.02)	38.21 (11.06)
	0.088	0.1130 (0.0201)	1.0771 (0.2992)	1.9993 (0.3252)	0.0965 (0.2805)	23.06 (0.74)	17.49 (3.85)	0.1126 (0.0240)	1.0544 (0.2985)	2.0011 (0.3321)	0.0935 (0.2815)	21.94 (0.59)	17.09 (4.50)	0.1122 (0.0263)	1.0322 (0.3045)	2.0052 (0.3300)	0.1034 (0.2869)	22.10 (0.80)	17.61 (5.11)
1	0.124	0.1775 (0.0367)	1.2066 (0.4027)	1.9794 (0.4049)	0.1047 (0.3315)	15.10 (0.71)	11.00 (2.29)	0.1664 (0.0403)	1.1414 (0.4129)	1.9746 (0.4168)	0.1093 (0.3368)	14.40 (0.57)	11.43 (2.85)	0.1628 (0.0415)	1.0827 (0.4019)	1.9796 (0.4286)	0.1126 (0.3497)	14.54 (0.83)	11.91 (3.20)
	0.159	0.2266 (0.0608)	1.2965 (0.6111)	1.9608 (0.4399)	0.1394 (0.3768)	11.59 (0.60)	9.25 (1.86)	0.2085 (0.0599)	1.1700 (0.5389)	1.9716 (0.4530)	0.1318 (0.3935)	11.10 (0.49)	9.58 (2.23)	0.1999 (0.0659)	1.1160 (0.5156)	1.9694 (0.4493)	0.1382 (0.3909)	11.20 (0.71)	10.20 (2.72)

Table 4.5: Mean parameter estimates and ESS over $M = 500$ replications for Poisson, Negative Binomial, and Gamma families across smoothness ν and spatial range configurations ϕ . Standard deviations in parentheses.

ν	ϕ	Poisson						Neg. Binomial						Gamma					
		MAE	MRE%	RMSE	Bias	Bias%	C ₃₀	MAE	MRE%	RMSE	Bias	Bias%	C ₃₀	MAE	MRE%	RMSE	Bias	Bias%	C ₃₀
0.5	0.075	11.35	23.11	13.42	-6.91	-14.03	68.8	12.46	26.66	14.63	-7.84	-16.75	58.4	12.91	27.46	15.38	-8.54	-18.09	58.4
	0.125	5.18	21.33	6.44	-1.94	-7.93	74.2	6.02	26.10	7.68	-2.48	-10.75	64.4	6.88	29.66	8.40	-3.03	-12.91	54.2
	0.175	3.30	20.29	4.04	-1.53	-9.24	79.8	3.76	24.39	4.81	-1.53	-9.82	65.8	4.27	27.47	5.25	-1.49	-9.28	60.0
	0.225	2.42	19.31	3.46	0.23	1.92	83.4	2.78	23.37	3.75	0.02	0.29	75.4	3.10	25.95	4.68	0.35	2.99	71.6
0.75	0.061	10.75	22.07	12.44	-8.56	-17.54	71.2	11.61	25.03	13.43	-9.11	-19.59	62.6	13.30	28.50	15.45	-9.45	-20.17	53.8
	0.102	5.70	24.32	6.54	-5.00	-21.35	62.6	5.82	26.13	6.71	-4.65	-20.84	58.6	6.36	28.33	7.40	-4.59	-20.41	55.2
	0.143	4.16	26.81	4.57	-3.78	-24.31	56.6	4.02	27.28	4.54	-3.29	-22.20	54.4	4.05	27.21	4.64	-2.88	-19.15	56.0
	0.184	2.45	20.65	2.78	-1.86	-15.65	76.6	2.55	22.55	3.02	-1.42	-12.41	70.8	2.58	22.64	3.20	-0.82	-7.05	71.6
1	0.053	9.66	20.12	11.14	-7.27	-15.11	79.0	10.41	22.75	12.07	-7.62	-16.61	69.6	11.45	24.87	13.66	-7.79	-16.84	64.6
	0.088	6.00	25.98	6.76	-5.57	-24.12	59.8	5.79	26.31	6.65	-4.85	-22.05	59.2	5.87	26.48	6.85	-4.49	-20.19	57.6
	0.124	4.32	28.51	4.70	-4.10	-27.11	48.8	3.67	25.36	4.16	-2.98	-20.54	60.0	3.64	24.93	4.21	-2.63	-17.87	63.0
	0.159	2.66	22.92	2.96	-2.34	-20.21	67.8	2.34	21.05	2.70	-1.51	-13.53	75.6	2.33	20.81	2.88	-1.00	-8.78	77.0

Table 4.6: Accuracy metrics across Poisson, Negative Binomial, and Gamma distributions. MRE and Bias are reported as percentages

4.1.3 Experiment 3

This experiment focuses on the asymptotic theory underlying the ESS established by [Acosta y Vallejos \(2018\)](#). The theory relies on the increasing domain framework of [Mardia y Marshall \(1984\)](#), which guarantees consistency and asymptotic normality of the maximum likelihood estimator of ESS as the spatial domain expands while the minimum distance between observation sites remains bounded away from zero. Nevertheless, this theoretical framework was developed under Gaussian assumptions with frequentist inference, and its extension to SGLMMs with Bayesian inference is not straightforward, given that the regularity conditions of Theorem 3 in [Acosta y Vallejos \(2018\)](#) are not directly verifiable in this context. Therefore we investigate empirically whether consistency-like behavior of the ESS estimator extends to the SGLMM framework under an increasing domain scheme, focusing on a Poisson response with a regular grid design as a representative setting.

Main analysis ($\nu = 0.5$)

The experimental design was structured around a sequence of expanding regular grids, each defined as an $r \times r$ lattice with fixed unit spacing $\gamma = 1$ between adjacent sites so that the spatial domain $\mathcal{D}_r = [0, r - 1]^2$ expands as r increases. Specifically, for all pairs of distinct sites $s_i, s_j \in \mathcal{D}_r$, we have $\|s_i - s_j\| \geq \gamma = 1$, which satisfies the domain condition of [Mardia y Marshall \(1984\)](#).

The data-generating process follows the same hierarchical structure as Experiment 2:

$$\begin{aligned} \mathbf{w} &\sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{R}(\phi, \nu)), \\ \boldsymbol{\eta} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \\ \boldsymbol{\mu} &= \exp(\boldsymbol{\eta}), \\ Y_i &\sim \text{Poisson}(\mu_i), \end{aligned}$$

where the spatial random effect $\mathbf{u}$ has a Matérn correlation structure with fixed parameters $\phi = 1$, $\nu = 0.5$, $\sigma^2 = 1$, and $\tau^2 = 0$. The design matrix $\mathbf{X}$ includes an intercept and the spatial covariate $s_{i,x} \cdot s_{i,y}$, with regression coefficients $\boldsymbol{\beta} = (2, 0.1)^\top$. We selected the Poisson family to isolate the effect of the increasing domain scheme from the influence of the likelihood family, which was already examined in Experiment 2.

For each grid size, $M = 500$ independent replicates were generated for $r \in \{8, 16, 32, 64, 128, 256\}$. For larger grid sizes ($r \geq 64$), construction and storage of the dense covariance, correlation, and spectral decomposition matrices becomes memory-prohibitive, collectively requiring over 13 GB of RAM for $r = 128$ alone, which combined with the INLA fitting procedure exceeds the available computational resources. Therefore, spatial random fields were generated using the circulant embedding method via Fast Fourier Transform (FFT) (Wood y Chan, 1994), which significantly reduces computational cost while producing statistically equivalent samples.

Each replicate was fitted using INLA with the SPDE approach, with PC priors centered at the true parameter values, and the true ESS was computed for each replicate using the known data-generating parameters, providing a benchmark for comparison.

The SPDE mesh was constructed using a fixed `max.edge` parameter equal to half the practical range, `max.edge` = $\sqrt{8\nu} \phi/2$, which remains constant across all grid sizes, ensuring constant local approximation quality. As the domain expands, the total number of mesh nodes grows to maintain this constant local resolution over a larger area, with the ratio of nodes to observations summarized in Table 4.7

Table 4.7: INLA mesh nodes relative to observation points for each grid size r .

r	$n = r^2$	Mesh nodes	Ratio
8	64	112	1.750
16	256	336	1.312
32	1024	1168	1.141
64	4096	4368	1.066
128	16384	16912	1.032
256	65536	66576	1.016

We tested consistency using the percentage bias and the root mean squared error, where n_i^* and $\hat{n}_i^*$ denote the true and estimated ESS for replicate i respectively, and n^* and $\hat{n}^*$ their respective averages over the M replicates:

$$\text{Bias}(\%) = \frac{\hat{n}^* - n^*}{n^*} \times 100, \quad \text{RMSE} = \sqrt{\frac{1}{M} \sum_{i=1}^M (\hat{n}_i^* - n_i^*)^2},$$

along with the mean ratio $\hat{n}^*/n^*$. Distributional convergence toward normality was additionally examined via the Kolmogorov-Smirnov distance between the standardized

empirical distribution of $\hat{n}^*$ and the standard normal, as well as a graphical assessment using empirical distribution function (ECDF) and QQ-plots.

It should be noted that the mesh resolution used in this experiment, with `max.edge` = $r/2$, is coarser than the commonly recommended resolution of approximately $r/5$ for accurate SPDE approximations. This choice was imposed by computational constraints: finer meshes for large grid sizes exceeded available memory and computation time. As a consequence, the absolute estimation errors observed in this experiment are likely influenced by mesh approximation error in addition to statistical estimation error.

Table 4.8: Simulation results for Experiment 3. All results are based on $M = 500$ replicates.

r	n	n^*	$\hat{n}^*$	Bias (%)	Ratio	KS distance
8	64	18.49	31.80	72.00	1.723	0.056
16	256	58.18	92.48	58.95	1.590	0.081
32	1024	204.58	274.67	34.26	1.343	0.047
64	4096	765.49	932.95	21.88	1.219	0.028
128	16384	2961.15	3479.93	17.52	1.175	0.033
256	65536	11644.00	13405.78	15.13	1.151	0.023

Table 4.8 presents the results across all grid sizes. The percentage bias decreases monotonically from 72.00% at $r = 8$ to 15.13% at $r = 256$, and the mean ratio decreases progressively from 1.723 to 1.151, showing a consistency-like behaviour. Kolmogorov-Smirnov distance decreases steadily from $r = 32$ onward, reaching 0.023 at $r = 256$, suggesting a convergence to the standard normal distribution, which is further supported by the ECDF as QQ-plots in Figure 4.7.

Computation times per replicate are summarized in Figure 4.8, growing from under 5 seconds at $r = 8$ to over 1,000 seconds at $r = 256$, following an approximately $O(n^{0.87})$ growth with $n = r^2$. Extrapolating this trend, a single replicate at $r = 512$ would require approximately 0.7 hours, making $M = 500$ replicates prohibitively expensive within the available computational resources.

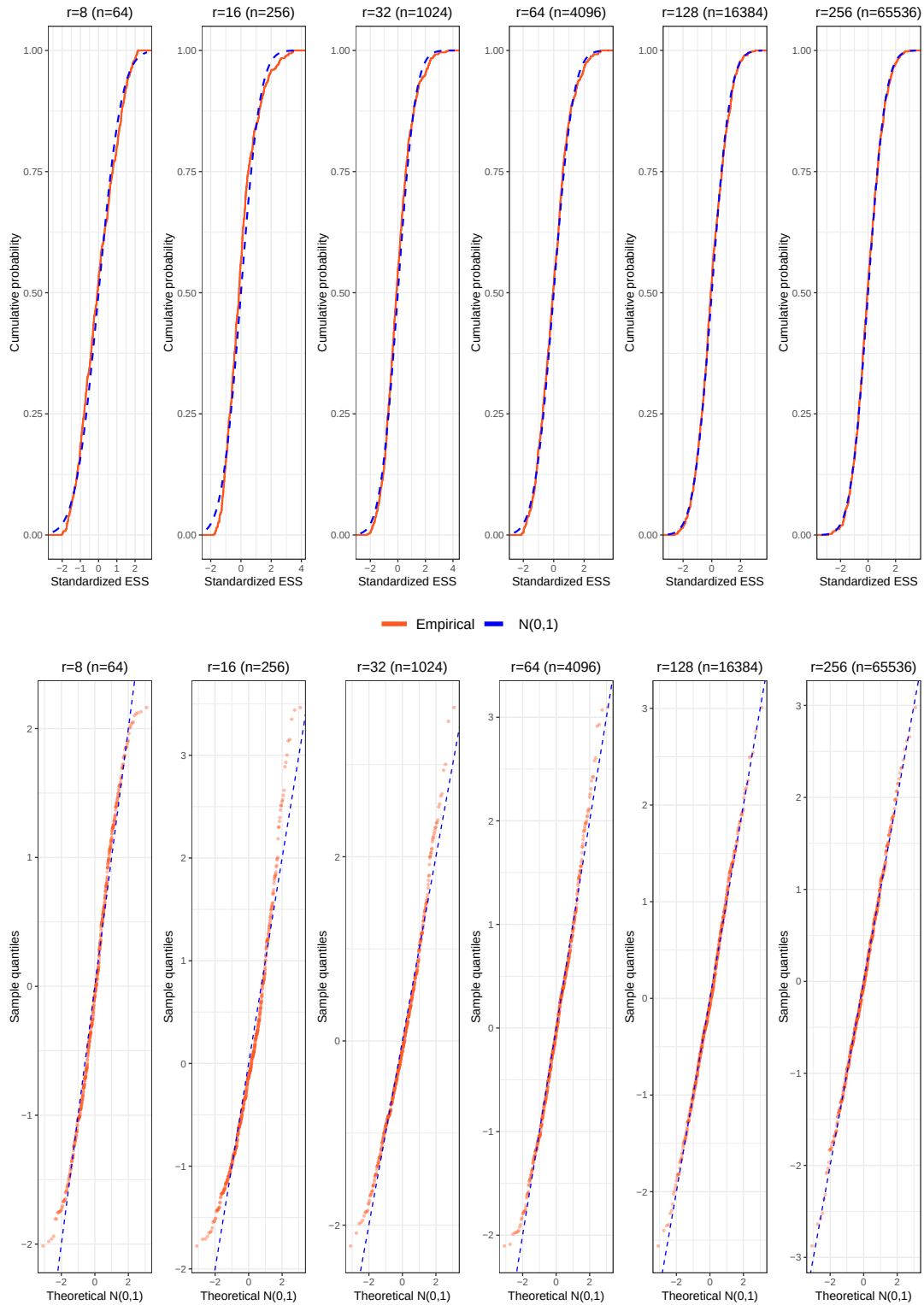


Figure 4.7: Empirical cumulative distribution functions (ECDF, top) and Q-Q plots (bottom) of the standardized ESS, both compared against the standard normal distribution (dashed line).

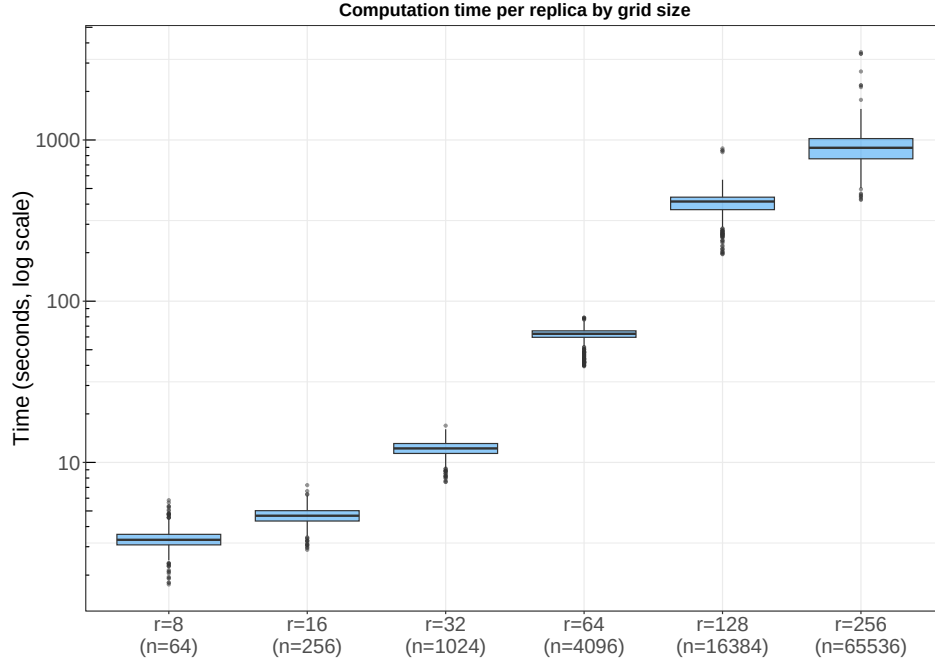


Figure 4.8: Distribution of computation times per replicate across grid sizes $r \in \{8, 16, 32, 64, 128, 256\}$ for $\nu = 0.5$ with $M = 500$ replicates. The vertical axis is on a logarithmic scale.

Sensitivity to the smoothness parameter

The preceding analysis was conducted for $\nu = 0.5$ with $M = 500$ replicates. Extending this to other values of ν with the same number of replicates is computationally prohibitive, as illustrated in Figure 4.8. Therefore, the following exploratory analysis uses $M = 100$ replicates to examine whether consistency-like behavior extends to smoother fields.

We thus replicated the same experimental setup for $\nu \in \{0.60, 0.70, 0.80, 0.90, 1.0\}$ with $M = 100$ replicates and $r \in \{8, 16, 32, 64, 128\}$, with the addition of a hybrid estimator $\tilde{n}^*$ for each replicate, computed using the linear predictor $\hat{\eta}$ estimated by INLA but the true correlation parameters ϕ and ν , so that any discrepancy from n^* reflects only the error attributable to the estimation of the mean structure rather than to $\hat{\phi}$.

Table 4.9 summarizes mean ratio for each configuration. For the smallest values of ν , the ratio $\hat{n}^*/n^*$, exhibits a monotonic decrease toward 1 within the explored domains. As ν increases, the ratio crosses 1 and then moves away from 1 as the domain grows, with this departure becoming more pronounced for larger ν . In contrast, the hybrid ratio $\tilde{n}^*/n^*$ remains close to 1 in almost all cases, falling at most to 0.98, suggesting

that the bias in $\hat{n}^*$ is largely driven by the estimation of ϕ .

This behavior is consistent with smoother fields requiring larger domains to reliably identify ϕ , suggesting that the domain sizes required for consistency-like behavior grow with ν . However, it is not clear from these results alone whether the ratio eventually returns toward 1 for larger domains, or whether the bias in $\hat{\phi}$ persists asymptotically. To investigate this, we selected $\nu = 0.63$ as an intermediate value where preliminary results indicated the ratio would cross 1 near $r = 64$, allowing us to track its subsequent behavior in larger domains. Using the same setup with $M = 100$ replicates, we extended the grid sizes to $r \in \{8, 16, 32, 64, 128, 160, 190, 220, 256\}$.

The results, presented in Table 4.10, confirm the anticipated pattern: the ratio $\hat{n}^*/n^*$ crosses 1 at $r \approx 64$ and continues to decrease as the domain grows, reaching a minimum of 0.923 at $r = 190$. Beyond this point, the ratio increases to 0.935 at $r \geq 220$, suggesting that the estimator has passed its worst-case bias and is beginning to recover toward 1. In contrast, the hybrid ratio $\tilde{n}^*/n^*$, which uses the true correlation parameters, remains close to 1 across all domain sizes, never falling below 0.988.

Comparing with the results for $\nu = 0.5$, at $r = 256$ the ratio for $\nu = 0.63$ (0.935) is actually closer to 1 than for $\nu = 0.5$ (1.151), with the key difference that $\nu = 0.5$ approaches 1 from above while $\nu = 0.63$ approaches it from below. Taken together, the results for both values of ν suggest that consistency-like behavior is not restricted to $\nu = 0.5$, but that the rate of convergence and the shape of the trajectory toward 1 depend on the smoothness of the underlying field.

4.2 Real Data Experiment

The dataset used for this example was collected at the Bjertorp farm located 58.26° N–13.13° E in the south-west of Sweden. It corresponds to the Bjertorp dataset, an agricultural dataset, which consist of exact weed counts (ranging from 4 to 300 plants per frame) in frames of 0.5 m $\times$ 0.75 m taken at 100 sites and estimates derived from pictures that were taken exactly over these frames (ranging from 7 to 222 plants) (Guillot *et al.*, 2009). For this application, we use only the exact weed counts, as they represent ground truth measurements without the detection errors inherent in image-based methods.

This dataset, illustrates a common real-world application of the SGLMMs: spatial

ν	r	n^*	$\hat{n}^*$	$\tilde{n}^*$	$\hat{n}^*/n^*$	$\tilde{n}^*/n^*$	$\hat{\phi}$	$\hat{\sigma}^2$	$\hat{\beta}_0$	$\hat{\beta}_1$
0.60	8	16.39 (0.49)	26.43 (11.00)	16.29 (0.41)	1.614	0.994	0.786 (0.374)	1.560 (0.546)	2.069 (0.351)	0.087 (0.354)
	16	50.63 (0.91)	64.13 (22.02)	50.20 (0.90)	1.267	0.991	0.904 (0.209)	1.163 (0.144)	1.980 (0.208)	0.113 (0.195)
	32	175.89 (1.97)	198.63 (36.64)	174.00 (1.72)	1.130	0.989	0.939 (0.105)	1.037 (0.063)	2.004 (0.100)	0.092 (0.104)
	64	653.18 (3.74)	677.85 (64.71)	645.98 (3.40)	1.038	0.988	0.977 (0.053)	0.997 (0.029)	2.001 (0.068)	0.102 (0.071)
	128	2518.92 (7.47)	2488.99 (130.87)	2489.88 (6.72)	0.988	0.988	1.001 (0.029)	0.981 (0.015)	1.994 (0.030)	0.105 (0.031)
0.70	8	14.86 (0.48)	22.48 (10.18)	14.76 (0.42)	1.516	0.994	0.838 (0.365)	1.343 (0.437)	2.076 (0.358)	0.088 (0.353)
	16	44.93 (0.88)	49.57 (15.20)	44.56 (0.85)	1.103	0.991	0.980 (0.198)	1.049 (0.138)	1.986 (0.218)	0.107 (0.200)
	32	154.30 (1.89)	153.06 (24.41)	152.66 (1.63)	0.992	0.989	1.011 (0.098)	0.959 (0.068)	2.004 (0.107)	0.093 (0.110)
	64	569.50 (3.58)	520.17 (42.48)	563.17 (3.17)	0.913	0.989	1.048 (0.048)	0.932 (0.031)	2.001 (0.072)	0.104 (0.075)
	128	2189.46 (7.15)	1917.15 (92.11)	2164.16 (6.16)	0.876	0.988	1.068 (0.028)	0.923 (0.015)	1.994 (0.033)	0.106 (0.033)
0.80	8	13.66 (0.47)	18.76 (8.79)	13.57 (0.41)	1.376	0.994	0.931 (0.441)	1.197 (0.363)	2.079 (0.384)	0.080 (0.379)
	16	40.50 (0.86)	40.37 (11.32)	40.14 (0.84)	0.996	0.991	1.041 (0.191)	0.995 (0.147)	1.976 (0.231)	0.122 (0.205)
	32	137.57 (1.83)	124.16 (20.12)	136.15 (1.63)	0.903	0.991	1.069 (0.100)	0.922 (0.071)	2.006 (0.112)	0.090 (0.116)
	64	504.81 (3.44)	424.00 (33.31)	499.31 (2.98)	0.840	0.990	1.097 (0.048)	0.902 (0.036)	2.002 (0.076)	0.103 (0.078)
	128	1935.13 (6.85)	1565.86 (66.46)	1913.06 (6.17)	0.809	0.989	1.113 (0.025)	0.898 (0.017)	1.994 (0.035)	0.106 (0.036)
0.90	8	12.70 (0.46)	16.25 (7.35)	12.63 (0.41)	1.284	0.996	0.970 (0.404)	1.120 (0.362)	2.079 (0.396)	0.082 (0.392)
	16	36.95 (0.84)	34.56 (8.20)	36.67 (0.84)	0.935	0.992	1.079 (0.179)	0.979 (0.166)	1.978 (0.244)	0.113 (0.216)
	32	124.25 (1.76)	107.20 (15.50)	122.96 (1.58)	0.863	0.990	1.096 (0.098)	0.903 (0.085)	2.006 (0.120)	0.092 (0.125)
	64	453.42 (3.30)	363.07 (24.49)	448.56 (2.98)	0.801	0.989	1.125 (0.042)	0.894 (0.041)	2.000 (0.080)	0.106 (0.083)
	128	1733.30 (6.58)	1330.42 (56.54)	1713.99 (5.80)	0.768	0.990	1.144 (0.026)	0.891 (0.020)	1.994 (0.037)	0.105 (0.037)
1.0	8	11.91 (0.45)	14.84 (7.05)	11.86 (0.42)	1.250	0.996	1.011 (0.439)	1.082 (0.378)	2.072 (0.410)	0.095 (0.396)
	16	34.06 (0.82)	30.29 (6.69)	33.80 (0.82)	0.890	0.992	1.113 (0.171)	0.962 (0.175)	1.982 (0.256)	0.111 (0.227)
	32	113.41 (1.70)	93.98 (12.06)	112.26 (1.52)	0.829	0.990	1.120 (0.087)	0.895 (0.084)	2.007 (0.123)	0.091 (0.129)
	64	411.68 (3.18)	322.18 (21.74)	407.35 (2.80)	0.783	0.989	1.140 (0.043)	0.889 (0.042)	2.002 (0.084)	0.105 (0.086)
	128	1569.50 (6.33)	1170.27 (42.91)	1552.51 (5.69)	0.746	0.989	1.162 (0.023)	0.892 (0.021)	1.994 (0.038)	0.105 (0.039)

Table 4.9: Sensitivity analysis results for $\nu \in \{0.60, 0.70, 0.80, 0.90, 1.0\}$ with $M = 100$ replicates. Standard deviations in parentheses.

ν	r	$\hat{n}^*/n^*$	$\tilde{n}^*/n^*$	Bias% $\hat{n}^*$	Bias% $\tilde{n}^*$	$\hat{\phi}$	$\hat{\sigma}^2$	$\hat{\beta}_0$	$\hat{\beta}_1$
0.63	8	1.636	0.994	63.37	-0.59	0.780 (0.359)	1.514 (0.526)	2.077 (0.355)	0.083 (0.357)
	16	1.201	0.991	20.07	-0.89	0.929 (0.197)	1.116 (0.139)	1.981 (0.208)	0.113 (0.191)
	32	1.073	0.989	7.24	-1.06	0.966 (0.097)	1.008 (0.063)	2.004 (0.101)	0.092 (0.107)
	64	0.994	0.989	-0.56	-1.12	1.000 (0.050)	0.972 (0.029)	2.000 (0.069)	0.105 (0.072)
	128	0.952	0.988	-4.85	-1.162	1.022 (0.027)	0.957 (0.015)	2.004 (0.036)	0.096 (0.037)
	160	0.929	0.988	-7.06	-1.176	1.034 (0.021)	0.951 (0.012)	2.001 (0.025)	0.100 (0.024)
	190	0.923	0.988	-7.75	-1.179	1.038 (0.016)	0.950 (0.011)	2.001 (0.022)	0.098 (0.022)
	220	0.935	0.988	-6.52	-1.179	1.030 (0.014)	0.950 (0.009)	1.999 (0.021)	0.100 (0.021)
	256	0.935	0.988	-6.51	-1.179	1.030 (0.011)	0.952 (0.007)	1.998 (0.017)	0.103 (0.018)

Table 4.10: Simulation results for $\nu = 0.63$ with $M = 100$ replicates. Standard deviations in parentheses.

count data with non Gaussian distribution. The strong spatial correlation documented by the authors, with an effective range of approximately 50m, motivates a formal quantification of the effective sample size, which is the focus of the present analysis. This dataset exhibits a structured spatial sampling design, with locations following an approximately regular grid pattern across most of the field, complemented by an area of higher sampling density with irregular spacing (see Figure 4.9). The spatial distribution of the weed counts and the semivariogram associated is shown in Figures 4.10 and 4.11.

For this dataset, [Guillot *et al.* \(2009\)](#) compared two different approaches within the model-based geostatistics framework of [Diggle *et al.* \(1998\)](#). Specifically a transformed Gaussian Cox process and a transformed Gaussian random field, each implemented under full Bayes via MCMC and plug-in estimation. In contrast, our analysis adopts a Poisson log-normal model fitted via INLA.

Let $Y(\mathbf{s}_i)$ denote the weed count at location $\mathbf{s}_i$, $i = 1, 2, \dots, 100$, the model used has the form:

$$Y(\mathbf{s}_i) \mid \eta(\mathbf{s}_i) \sim \text{Poisson}(\mu(s_i))$$

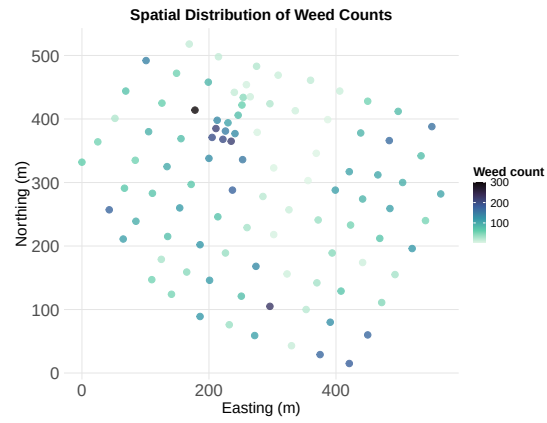


Figure 4.9: Spatial distribution of the Bjertop dataset.

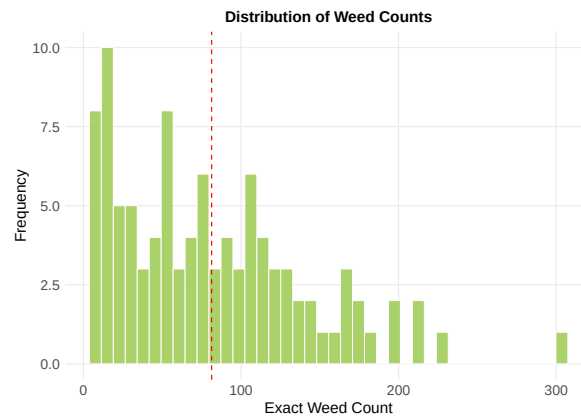


Figure 4.10: Histogram of the Bjertop dataset.

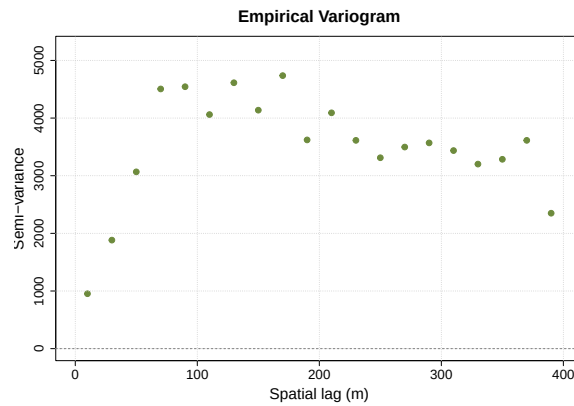


Figure 4.11: Semivariogram of the Bjertop dataset.

where $\mu(\mathbf{s}_i) = \exp(\eta(\mathbf{s}_i))$ and the linear predictor is given by:

$$\eta(\mathbf{s}_i) = \beta_0 + \mathbf{u}(\mathbf{s}_i)$$

with β_0 the intercept and $\mathbf{u}(\mathbf{s})$ a zero-mean Gaussian random field (GRF). Given that we implemented the model using INLA via R-INLA package the spatial field $\mathbf{u}(s)$ was approximated using the Stochastic Partial Differential Equation (SPDE) approach (Lindgren *et al.*, 2011). We compare two different covariance functions (i) $\nu = 0.5$: Exponential covariance. (ii) $\nu = 1$ default INLA-SPDE specification.

Table 4.11 shows the model comparison results. The exponential covariance model showed better fit across all criteria (lower DIC, WAIC, and RMSE) and was selected for subsequent ESS calculations.

Parameter	$\nu = 0.5$	$\nu = 1.0$
β_0	4.080 (0.369)	4.049 (0.284)
$\log(\kappa)$	-1.851 (0.087)	-3.730 (0.140)
$\log(\tau)$	1.915 (0.298)	2.550 (0.215)
Range (m)	176.9 (53.7)	129.8 (28.1)
σ^2	0.0700	0.8431
DIC	787.39	787.93
WAIC	772.27	773.47
RMSE	1.83	1.95
Correlation	0.9997	0.9996

Table 4.11: Parameter estimates and model fit comparison for the two Matérn smoothness parameters. Posterior means with standard deviations in parentheses (where applicable). Lower DIC, WAIC, and RMSE indicate better fit.

In order to quantify the uncertainty in the spatial covariance parameters when calculating the ESS, we drew 10,000 samples from the joint posterior distribution of model parameters (Rue *et al.*, 2013; Martins *et al.*, 2013). For each posterior sample, we reconstructed the spatial correlation matrix and computed the corresponding ESS value. This approach yields to a posterior distribution for ESS, from which we computed the posterior mean and 95% credible intervals.

The posterior distribution is shown in Figure 4.12. The posterior mean ESS was 10.9 (95% CI: 4.6, 21.7), representing 10.9% of the nominal sample size. This reduction is directly attributable to the spatial clustering of observations within the estimated

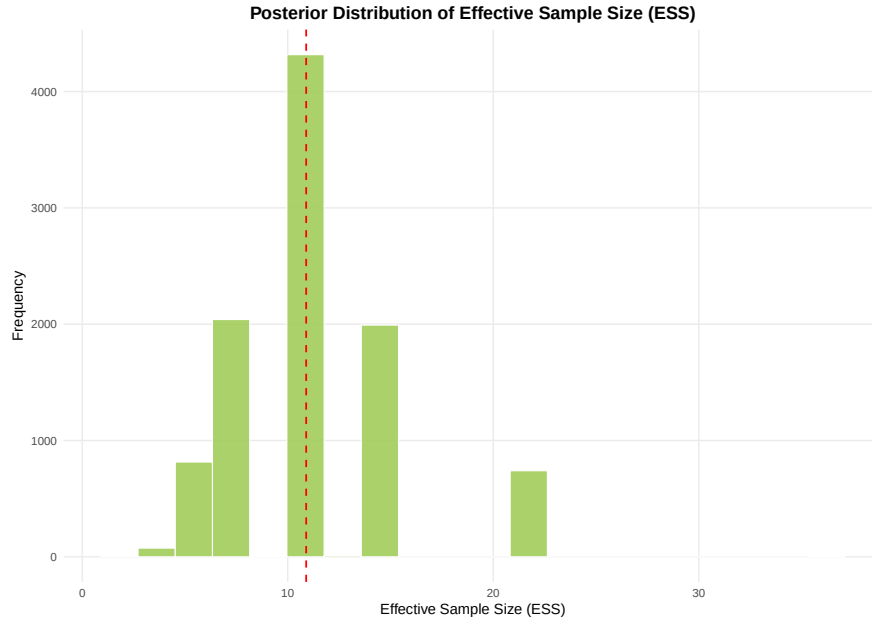


Figure 4.12: Posterior distribution of Effective Sample Size (ESS) obtained from 10,000 posterior samples. The red dashed line indicates the posterior mean ($ESS = 10.9$).

range of 177 m. Since many observation pairs are separated by distances shorter than this range, they share redundant information about the underlying weed intensity field. To the best of our knowledge, this is the first formal quantification of the ESS for this dataset, as previous analysis focused on spatial prediction accuracy rather than information content on the sample.

Chapter 5

Discussion and Future Work

Extending the ESS framework to SGLMMs introduces two challenges that do not arise in previously studied settings. The first is the analytical intractability of the marginal likelihood, which forces reliance on approximations. The second is the practical difficulty of estimating the spatial covariance parameters on which the ESS depends.

Both of these problems were addressed in the previous chapter. Experiment 1 shows that the approximations used to overcome the analytical intractability of the marginal likelihood are accurate across a wide range of realistic scenarios, suggesting that this first challenge does not pose a fundamental obstacle in practice. The second challenge, however, proves more difficult: the estimation of spatial parameters, shown in Experiment 2, reveals that the systematic overestimation of ϕ under a fixed spatial domain propagates directly into the ESS through $\mathbf{R}(\phi)$, regardless of the likelihood family employed.

The source of the estimation error in Experiment 2 is not fully understood. It cannot be attributed to the ESS formula itself, nor to variability in the true ESS across replications. One possible explanation is the joint non-identifiability of ϕ and σ^2 under a fixed spatial domain (Zhang, 2004): since both parameters jointly determine the covariance structure $\sigma^2 \mathbf{R}(\phi)$, their separate estimation is inherently difficult. However, this explanation alone is not fully satisfactory: under a comparable setting — a 16×16 grid with $\nu = 0.5$ — Acosta y Vallejos (2018) report substantially lower estimation errors despite facing the same identifiability constraints. Whether the remaining discrepancy is attributable to the inferential method, the approximation quality of the SPDE mesh, or the inherently greater complexity of Bayesian inference in the SGLMM

setting cannot be determined from the experiments presented here. To address these limitations, we suggest to replicate the simulation study using standard MCMC with exact matern covariance. This would eliminate both the Laplace approximation and the SPDE mesh simultaneously, at the expense of a higher computational cost, which may be prohibitive for the spatial configurations considered here.

This uncertainty in the estimation of ϕ is further examined in Experiment 3, where the hybrid ratio ($\tilde{n}^*/n^*$), computed using the true correlation parameters but the estimated mean structure, remains close to 1 across all values of ν and domain sizes. This confirms that the bias in $\hat{n}^*$ is driven by the estimation of ϕ rather than by the estimation of the mean structure, and that the ESS calculation itself is correct given accurate parameter estimates.

Whether the asymptotic properties of the ESS estimator established by [Acosta y Vallejos \(2018\)](#) extend to the SGLMMs framework is addressed empirically in Experiment 3 through an increasing domain scheme. For $\nu = 0.5$, the results are consistent with consistency-like behavior, as the bias decreases monotonically with domain size, the mean ratio converges progressively toward 1. The distribution of ESS also converges toward normality for $r \geq 128$, as illustrated by Figure 4.7. For $\nu = 0.63$, a non-monotonic pattern is observed: the ratio crosses 1, reaches a minimum, and then partially recovers, with the recovery statistically confirmed by a one-sided t -test ($p = 0.001$). Both patterns are consistent with consistency-like behavior under increasing domain, converging toward 1 from opposite sides.

For smoother fields ($\nu \in \{0.70, 0.80, 0.90, 1.0\}$), the ratio moves away from 1 as the domain grows, without showing recovery within the explored domains. Whether this reflects a genuine difference between these values of ν and the cases where consistency-like behavior was observed, or simply a consequence of requiring larger domains, a finer mesh, or both, cannot be determined from the available data and represents an important direction for future research.

This work has several limitations worth acknowledging. First, the ESS can become unstable under extreme spatial correlation values, as evidenced by the Poisson model at high ψ in Experiment 1. This behavior has also been reported by [Acosta y Vallejos \(2018\)](#) in the Gaussian case. Second, the INLA-based estimation explored in Experiment 2 was conducted under a fixed spatial domain, which inherently limits the identifiability of spatial parameters and likely contributes to the systematic underestimation observed. Furthermore, the mesh resolution employed in Experiment 3 was

constrained by computational resources, which may have inflated the estimation errors beyond what would be observed under finer approximations. Finally, all experiments considered regular or quasi-regular spatial grids, leaving the behavior of the ESS under clustered, preferential, or adaptive sampling designs unexamined.

These results point to several additional questions that merit further investigation. On the theoretical side, the non-monotone behavior of the ESS under Poisson and potentially other count-data likelihoods deserves a formal characterization. Although Proposition 3.2.2 provides a matrix condition for $\text{ESS} \leq n$, a complete description of the parameter regimes yielding monotone versus non-monotone behavior remains open. Similarly, a formal extension of the asymptotic results of [Acosta y Vallejos \(2018\)](#) to SGLMMs would require establishing consistency and asymptotic normality of the ESS estimator under conditions that accommodate the Laplace approximation and non-Gaussian likelihoods. From a design perspective, the behavior of the ESS under clustered, preferential, or adaptive sampling designs is practically relevant and largely unexplored, and constitutes a natural direction for future work.

Taken together, these contributions establish a foundation for applying the ESS framework to SGLMMs. The ESS was shown to exceed 1 under general conditions and to admit a tractable upper bound characterization, while the empirical experiments demonstrated its practical computability within the INLA-SPDE framework. The results also highlight the central role of spatial parameter estimation in determining the quality of ESS-based inference, pointing to a broader challenge in spatial statistics: disentangling the contribution of the correlation structure from that of the observation model in non-Gaussian settings. Addressing this challenge represents a natural direction for future work, with implications that extend beyond the ESS to the broader problem of model-based inference in complex spatial models.

Appendix A

A.1 ESS Derivation

We have the following marginal likelihood for SGLMMs presented in equation (2.16).

$$\mathcal{L}(\Psi; \mathbf{y}) = \int_{\mathbb{R}^n} \prod_{i=1}^n f(y_i | u(\mathbf{s}_i); \boldsymbol{\beta}, \phi) \cdot \phi_n(\mathbf{u}; \mathbf{0}, \mathbf{R}(\boldsymbol{\psi})) d\mathbf{u}.$$

If we use the Laplace approximation (Section 2.4.2) and apply logarithm, we get the following

$$\ell(\Psi) = \log(\mathcal{L}(\Psi; \mathbf{y})) = -\frac{1}{2} \log |\mathbf{H}| + \frac{n}{2} \log(2\pi) + L(\Psi, \hat{\mathbf{u}}),$$

where $\mathbf{H} = -\frac{\partial^2 L(\mathbf{u}, \Psi)}{\partial \mathbf{u} \partial \mathbf{u}^\top} \Big|_{\mathbf{u}=\hat{\mathbf{u}}}$, with $\hat{\mathbf{u}}$ the mode of L with respect to $\mathbf{u}$, and $L(\Psi, \mathbf{u}) = \sum_{i=1}^n \log(f(y_i | u(\mathbf{s}_i); \boldsymbol{\beta}, \phi)) + \log(\phi_n(\mathbf{u}; \mathbf{0}, \mathbf{R}(\boldsymbol{\psi})))$, whose Laplace conditions are

$$\frac{\partial L(\hat{\mathbf{u}}, \hat{\Psi})}{\partial \mathbf{u}} = \mathbf{0} \quad \text{and} \quad \frac{\partial^2 L(\hat{\mathbf{u}}, \hat{\Psi})}{\partial \mathbf{u} \partial \mathbf{u}^\top} < 0. \quad (\text{A.1})$$

Our goal is to obtain the Fisher information matrix for $\boldsymbol{\beta}$. To this end, we first compute the derivatives of $\ell(\Psi)$ with respect to $\boldsymbol{\beta}$. By the chain rule,

$$\frac{\partial \ell(\Psi)}{\partial \boldsymbol{\beta}} = \underbrace{\frac{\partial L(\mathbf{u}, \Psi)}{\partial \mathbf{u}} \Big|_{\mathbf{u}=\hat{\mathbf{u}}}}_{=\mathbf{0} \text{ (A.1)}} \cdot \frac{\partial \hat{\mathbf{u}}}{\partial \boldsymbol{\beta}} + \frac{\partial L(\mathbf{u}, \Psi)}{\partial \boldsymbol{\beta}} \Big|_{\mathbf{u}=\hat{\mathbf{u}}} - \frac{1}{2} \text{tr} \left(\mathbf{H}^{-1} \frac{\partial \mathbf{H}}{\partial \boldsymbol{\beta}} \right). \quad (\text{A.2})$$

For the second derivative, we get

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= \underbrace{\frac{\partial^2 L(\mathbf{u}, \Psi)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \Big|_{\mathbf{u}=\hat{\mathbf{u}}}}_{(a)} + \underbrace{\frac{\partial^2 L(\mathbf{u}, \Psi)}{\partial \boldsymbol{\beta} \partial \mathbf{u}^\top} \Big|_{\mathbf{u}=\hat{\mathbf{u}}}}_{(b)} \cdot \underbrace{\mathbf{H}^{-1}}_{(c)} \cdot \underbrace{\frac{\partial^2 L(\mathbf{u}, \Psi)}{\partial \mathbf{u} \partial \boldsymbol{\beta}^\top} \Big|_{\mathbf{u}=\hat{\mathbf{u}}}}_{(d)} \\ &\quad - \underbrace{\frac{1}{2} \frac{\partial}{\partial \boldsymbol{\beta}^\top} \left[\text{tr} \left(\mathbf{H}^{-1} \frac{\partial \mathbf{H}}{\partial \boldsymbol{\beta}} \right) \right]}_{(e)}. \end{aligned} \quad (\text{A.3})$$

Term (e) involves third-order derivatives of $L(\mathbf{u}, \Psi)$ with respect to $\mathbf{u}$. Under the Laplace approximation, the posterior of $\mathbf{u}$ is assumed to be approximately Gaussian near $\hat{\mathbf{u}}$, making these terms negligible. We therefore neglect (e) and derive each of the remaining terms, beginning with (c).

Recall that $\mathbf{H} = -\frac{\partial^2 L(\mathbf{u}, \Psi)}{\partial \mathbf{u} \partial \mathbf{u}^\top} \Big|_{\mathbf{u}=\hat{\mathbf{u}}}$. For the exponential family, we have $b'(\boldsymbol{\theta}_i) = \boldsymbol{\mu}_i$ and $V(\boldsymbol{\mu}_i) = b''(\boldsymbol{\theta}_i)$. From the first derivative,

$$\begin{aligned} \frac{\partial L(\mathbf{u}, \Psi)}{\partial \mathbf{u}_i} &= \frac{\partial}{\partial \mathbf{u}_i} \left[\sum_{j=1}^n (y_j \boldsymbol{\theta}_j - b(\boldsymbol{\theta}_j)) + \log(\phi_n(\mathbf{u}; \mathbf{0}, \mathbf{R}(\boldsymbol{\psi})) \right] \\ &= (y_i - b'(\boldsymbol{\theta}_i)) \cdot \frac{\partial \boldsymbol{\theta}_i}{\partial \boldsymbol{\mu}_i} \cdot \frac{\partial \boldsymbol{\mu}_i}{\partial \mathbf{u}_i} - [\mathbf{R}^{-1} \mathbf{u}]_i \\ &= (y_i - \boldsymbol{\mu}_i) \cdot \frac{1}{V(\boldsymbol{\mu}_i)} \cdot \frac{1}{g'(\boldsymbol{\mu}_i)} - [\mathbf{R}^{-1} \mathbf{u}]_i, \end{aligned} \quad (\text{A.4})$$

where we assume $a(\phi) = 1$, which holds for the Bernoulli and Poisson families. Differentiating again,

$$\frac{\partial^2 L(\mathbf{u}, \Psi)}{\partial \mathbf{u}_i \partial \mathbf{u}_j} = \begin{cases} -\frac{1}{V(\boldsymbol{\mu}_i)[g'(\boldsymbol{\mu}_i)]^2} + (y_i - \boldsymbol{\mu}_i) \frac{\partial}{\partial \mathbf{u}_i} \left[\frac{1}{V(\boldsymbol{\mu}_i) g'(\boldsymbol{\mu}_i)} \right] - [\mathbf{R}^{-1}]_{ii} & \text{if } i = j, \\ -[\mathbf{R}^{-1}]_{ij} & \text{if } i \neq j. \end{cases}$$

Recognizing $\mathbf{W} = \text{diag}\left(\frac{1}{V(\boldsymbol{\mu}_i)[g'(\boldsymbol{\mu}_i)]^2}\right)$ as the GLM weight matrix, and noting that $(y_i - \boldsymbol{\mu}_i) \approx 0$ at $\mathbf{u} = \hat{\mathbf{u}}$, we obtain

$$\mathbf{H} \approx \mathbf{W} + \mathbf{R}^{-1}. \quad (\text{A.5})$$

For term (d), we differentiate (A.4) with respect to $\boldsymbol{\beta}^\top$:

$$\begin{aligned} \frac{\partial^2 L(\mathbf{u}, \Psi)}{\partial \mathbf{u}_i \partial \boldsymbol{\beta}^\top} &= \frac{\partial}{\partial \boldsymbol{\beta}^\top} \left[(y_i - \boldsymbol{\mu}_i) \cdot \frac{1}{V(\boldsymbol{\mu}_i) g'(\boldsymbol{\mu}_i)} - [\mathbf{R}^{-1} \mathbf{u}]_i \right] \\ &= -\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}^\top} \cdot \frac{1}{V(\boldsymbol{\mu}_i) g'(\boldsymbol{\mu}_i)} + (y_i - \boldsymbol{\mu}_i) \cdot \frac{\partial}{\partial \boldsymbol{\beta}^\top} \left[\frac{1}{V(\boldsymbol{\mu}_i) g'(\boldsymbol{\mu}_i)} \right] \\ &= -\frac{1}{g'(\boldsymbol{\mu}_i)} \mathbf{x}_i^\top \cdot \frac{1}{V(\boldsymbol{\mu}_i) g'(\boldsymbol{\mu}_i)} + (y_i - \boldsymbol{\mu}_i) \cdot \frac{\partial}{\partial \boldsymbol{\beta}^\top} \left[\frac{1}{V(\boldsymbol{\mu}_i) g'(\boldsymbol{\mu}_i)} \right] \\ &= -\frac{\mathbf{x}_i^\top}{V(\boldsymbol{\mu}_i)[g'(\boldsymbol{\mu}_i)]^2} + (y_i - \boldsymbol{\mu}_i) \cdot \frac{\partial}{\partial \boldsymbol{\beta}^\top} \left[\frac{1}{V(\boldsymbol{\mu}_i) g'(\boldsymbol{\mu}_i)} \right], \end{aligned} \quad (\text{A.6})$$

where we used $\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}^\top} = \frac{1}{g'(\boldsymbol{\mu}_i)} \mathbf{x}_i^\top$, since $g(\boldsymbol{\mu}_i) = \boldsymbol{\eta}_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{u}_i$. Evaluating at $\mathbf{u} = \hat{\mathbf{u}}$, where $(y_i - \boldsymbol{\mu}_i) \approx 0$, we obtain in matrix form

$$\frac{\partial^2 L(\mathbf{u}, \Psi)}{\partial \mathbf{u} \partial \boldsymbol{\beta}^\top} \Big|_{\mathbf{u}=\hat{\mathbf{u}}} \approx -\mathbf{W} \mathbf{X}, \quad (\text{A.7})$$

where $\mathbf{X}$ is the $n \times p$ matrix with rows $\mathbf{x}_i^\top$. Note that $(b) = (d)^\top$, i.e., $\frac{\partial^2 L}{\partial \boldsymbol{\beta} \partial \mathbf{u}^\top} \Big|_{\mathbf{u}=\hat{\mathbf{u}}} \approx -\mathbf{X}^\top \mathbf{W}$.

Finally, for term (a):

$$\begin{aligned}
\frac{\partial^2 L(\mathbf{u}, \boldsymbol{\Psi})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= \frac{\partial}{\partial \boldsymbol{\beta}^\top} \sum_{i=1}^n \left[(y_i - \boldsymbol{\mu}_i) \cdot \frac{1}{V(\boldsymbol{\mu}_i) g'(\boldsymbol{\mu}_i)} \right] \cdot \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}^\top} \\
&= \sum_{i=1}^n \left[-\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \cdot \frac{1}{V(\boldsymbol{\mu}_i) g'(\boldsymbol{\mu}_i)} + (y_i - \boldsymbol{\mu}_i) \cdot \frac{\partial}{\partial \boldsymbol{\beta}^\top} \left[\frac{1}{V(\boldsymbol{\mu}_i) g'(\boldsymbol{\mu}_i)} \right] \right] \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}^\top} \\
&= \sum_{i=1}^n \left[-\frac{\mathbf{x}_i}{g'(\boldsymbol{\mu}_i)} \cdot \frac{1}{V(\boldsymbol{\mu}_i) g'(\boldsymbol{\mu}_i)} + (y_i - \boldsymbol{\mu}_i)(\dots) \right] \frac{\mathbf{x}_i^\top}{g'(\boldsymbol{\mu}_i)} \\
&= -\sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^\top}{V(\boldsymbol{\mu}_i) [g'(\boldsymbol{\mu}_i)]^2} + \text{terms involving } (y_i - \boldsymbol{\mu}_i). \tag{A.8}
\end{aligned}$$

Evaluating at $\mathbf{u} = \hat{\mathbf{u}}$, where $(y_i - \boldsymbol{\mu}_i) \approx 0$,

$$\frac{\partial^2 L(\mathbf{u}, \boldsymbol{\Psi})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \Big|_{\mathbf{u}=\hat{\mathbf{u}}} \approx -\mathbf{X}^\top \mathbf{W} \mathbf{X}. \tag{A.9}$$

Substituting (a)–(d) into the second derivative of ℓ , we obtain

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \approx -\mathbf{X}^\top \mathbf{W} \mathbf{X} + \mathbf{X}^\top \mathbf{W} (\mathbf{W} + \mathbf{R}^{-1})^{-1} \mathbf{W} \mathbf{X},$$

and applying the Woodbury identity $A - A(A + B)^{-1}A = (A^{-1} + B^{-1})^{-1}$ with $A = \mathbf{W}$ and $B = \mathbf{R}^{-1}$, this simplifies to

$$\mathbf{I}(\boldsymbol{\beta}) \approx \mathbf{X}^\top (\mathbf{W}^{-1} + \mathbf{R})^{-1} \mathbf{X}, \tag{A.10}$$

an approximation to the observed information matrix of the Laplace-approximated log-likelihood $\ell(\boldsymbol{\Psi})$, completing the derivation of (3.2). The exact Fisher information for GLMMs with canonical link is $\mathbf{X}^\top \text{cov}[\mathbf{y} - E\{E(\mathbf{y}|\mathbf{u})|\mathbf{y}\}]\mathbf{X}$ (Wand, 2007), which involves intractable conditional moments. We therefore adopt (A.10) as a tractable surrogate.

Bibliography

- Acosta, J., Alegría, A., Osorio, F., y Vallejos, R. (2021). Assessing the effective sample size for large spatial datasets: A block likelihood approach. *Computational Statistics & Data Analysis*, 162:107282.
- Acosta, J. y Vallejos, R. (2018). Effective sample size for spatial regression models. *Electronic Journal of Statistics*, 12(2):3147–3180.
- Acosta, J., Vallejos, R., y Griffith, D. (2018). On the effective geographic sample size. *Journal of Statistical Computation and Simulation*, 88(10):1958–1975.
- Alegría, A. (2023). Effective sample size for georeferenced and temporally evolving data. *Spatial Statistics*, 53:100721.
- Alegría, A., Gómez, J., Mateu, J., y Vallejos, R. (2026). Effective sample size for functional spatial data.
- Breslow, N. E. y Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421):9–25.
- Clifford, P., Richardson, S., y Hémon, D. (1989). Assessing the significance of the correlation between two spatial processes. *Biometrics*, 45(1):123–134.
- Cressie, N. (1993). *Statistics for Spatial Data*. Wiley, New York.
- Diggle, P. J. y Giorgi, E. (2019). *Model-based Geostatistics for Global Public Health: Methods and Applications*. CRC Press, Boca Raton.
- Diggle, P. J., Tawn, J. A., y Moyeed, R. A. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3):299–350.

- Doser, J. W., Finley, A. O., Saunders, S. P., Kery, M., Weed, A. S., y Zipkin, E. F. (2024). spAbundance: An R package for single-species and multi-species spatially explicit abundance models. *Methods in Ecology and Evolution*, 15:1024–1033.
- Dutilleul, P. (1993). Modifying the t test for assessing the correlation between two spatial processes. *Biometrics*, 49:305–314.
- Ferrer, C. y Vallejos, R. (2025). Is the effective sample size always less than n ? a spatial regression approach. *Statistics & Probability Letters*, 218:110309.
- Gotway, C. A. y Stroup, W. W. (1997). A generalized linear model approach to spatial data analysis and prediction. *Journal of Agricultural, Biological and Environmental Statistics*, 2(2):157–178.
- Griffith, D. A. (2005). Effective geographic sample size in the presence of spatial autocorrelation. *Annals of the Association of American Geographers*, 95(4):740–760.
- Guillot, G., Lorén, N., y Rudemo, M. (2009). Spatial prediction of weed intensities from exact count data and image-based estimates. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*.
- Haining, R. (1990). *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge University Press, Cambridge.
- Handcock, M. S. y Stein, M. L. (1993). A Bayesian analysis of kriging. *Technometrics*, 35(4):403–410.
- Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., y Bell, B. M. (2016). Tmb: Automatic differentiation and laplace approximation. *Journal of Statistical Software*, 70(5):1–21.
- Lindgren, F., Rue, H., y Lindström, J. (2011). An explicit link between gaussian fields and gaussian markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(4):423–498.
- Mardia, K. V. y Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71(1):135–146.

- Martins, T. G., Simpson, D., Lindgren, F., y Rue, H. (2013). Bayesian computing with inla: New features. *Computational Statistics & Data Analysis*, 67:68–83.
- Matérn, B. (1960). *Spatial Variation*. Swedish National Institute for Forestry Research, Stockholm. Republished as Lecture Notes in Statistics, Vol. 36, Springer, 1986.
- McCullagh, P. y Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall, London, 2 edición.
- Mukerjee, R. (2024). Improving upon the effective sample size based on Godambe information for block likelihood inference. *Computational Statistics*, 39:891–904.
- Paradinas, I., Illian, J., y Smout, S. (2023). Understanding spatial effects in species distribution models. *PLOS ONE*, 18(5):e0285463.
- Rue, H., Martino, S., y Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(2):319–392.
- Rue, H., Martino, S., Lindgren, F., Simpson, D., y Riebler, A. (2013). R-inla: Approximate bayesian inference using integrated nested laplace approximations. <http://www.r-inla.org/>.
- Schabenberger, O. y Gotway, C. A. (2005). *Statistical Methods for Spatial Data Analysis*. Chapman & Hall/CRC, Boca Raton.
- Tierney, L. y Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86.
- Vallejos, R. y Osorio, F. (2014). Effective sample size of spatial process models. *Spatial Statistics*.
- Wand, M. P. (2007). Fisher information for generalised linear mixed models. *Journal of Multivariate Analysis*, 98(7):1412–1416.
- Wood, A. T. A. y Chan, G. (1994). Simulation of stationary gaussian processes in $[0, 1]^d$. *Journal of Computational and Graphical Statistics*, 3(4):409–432.
- Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465):250–261.