

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE MATEMÁTICA

Desempeño de la verosimilitud compuesta para datos espaciales

Memoria de Título presentada por:

Cloe Monserrat Romero Quiroz

como requisito parcial para optar al título de

Ingeniería Civil Matemática

Otorgado por la Universidad Técnica Federico Santa María

Profesor Guía

Dr. Ronny Vallejos Arriagada

Julio 2024.

TÍTULO DE LA MEMORIA:

Desempeño de la verosimilitud compuesta para datos espaciales

AUTOR:

CLOE MONSERRAT ROMERO QUIROZ

TRABAJO DE MEMORIA, presentado en cumplimiento parcial de los requisitos para el título de Ingeniería Civil Matemática de la Universidad Técnica Federico Santa María.

Dr. Profesor Guía

Dr. Revisor 1

Dr. Revisor 2

Valparaíso, Chile, Julio 2024.

AGRADECIMIENTOS

Me gustaría comenzar agradeciendo a todas las personitas que fueron parte de este proceso universitario, a todas esas personas con las que estudié y aprendí. Especialmente a mi grupo de amigos desde el inicio (Bruno, Dani, Maca y Monte) y a los amigos que hice con el tiempo (Franco y Sofía), sin ellos no hubiera sido lo mismo y no hubiera llegado tan lejos.

Gracias a mis padres por proporcionarme las herramientas y el sustento necesarios durante todos estos años, y por confiar en que lograría mis objetivos. A mi mejor amiga de la vida, Javiera, por siempre confiar en mí, incluso cuando yo no lo hacía, y por brindarme su apoyo incondicional en todo momento. A mi mejor amiga de la universidad, Dani, por siempre estar, ser la mejor partner y un ejemplo a seguir. A mi pareja, Bruno, quien estuvo a mi lado durante todo este proceso, especialmente en los momentos más difíciles, ayudándome, sosteniéndome y alentándome a seguir adelante.

Finalmente, gracias al Profesor Ronny Vallejos por aceptarme como su alumna memorista y confiar en este proceso, gracias a sus indicaciones y acompañamiento logré terminar esta etapa.

RESUMEN

En esta memoria se aborda el análisis de datos georeferenciados a gran escala, enfocado en los modelos de procesos Gaussianos espaciales y las metodologías necesarias para su análisis. La investigación se centra en la inferencia dentro de estos modelos, especialmente cuando se manejan grandes conjuntos de datos que presentan desafíos computacionales significativos debido al tamaño y complejidad de las matrices de covarianza involucradas.

Se exploran dos estrategias para mitigar la carga computacional, ambas basadas en el uso de la verosimilitud compuesta introducida por [Lindsay \(1988\)](#). Esta técnica permite capturar la dependencia espacial y disminuye el tamaño de las matrices involucradas, lo que reduce significativamente el tiempo de cálculo. La primera estrategia considera la relación de los bloques con sus bloques adyacentes utilizando pares de bloques, mientras que la segunda asume independencia entre los bloques y los trabaja individualmente.

Además, se analiza cómo afecta la manera en la que se seleccionan los bloques. Para ello, se estudian dos enfoques: el enfoque por filas (RW), en la cual los bloques se organizan en una cuadrícula regular, y el enfoque por columnas (CW), en el que los bloques están formados por múltiples columnas no adyacentes.

Se concluye que la estrategia que considera los pares de bloque proporciona mejores estimaciones, pero es menos eficiente computacionalmente que la que considera independencia. Para ambos enfoques, el método CW entrega mejores resultados.

El estudio incluye experimentos numéricos y análisis comparativos entre las distintas estrategias, además de una revisión de los conceptos fundamentales de la estadística.

ca espacial y los procesos estocásticos. Finalmente, se discuten las conclusiones y posibles trabajos futuros.

ABSTRACT

This work deals with the analysis of great-scale geo-referenced data, focusing on spatial Gaussian processes and the necessary methodologies for their analysis. The investigation centers itself on the inference of these models, specially when dealing with big sets of data, which presents significant computational challenges given the size and complexity of the covariance matrices involved.

Two strategies for the mitigation of the computational load are explored, both based on the use of composite likelihood introduced by [Lindsay \(1988\)](#). This technique allows the capture of spatial dependency and diminishes the size of the involved matrices, which significantly reduces the time of calculation. The first strategy considers the relation of the blocks with its adjacent blocks using block-pairs, while the second assumes independence between the blocks and treats them individually. Furthermore, the effects of the way the blocks are chosen is analyzed. To this end, two approaches are studied: the row-wise approach (RW), in which the blocks are arranged into a regular grid, and the column-wise approach (CW), in which every block is made from multiple non-adjacent columns.

We conclude that the strategy that considers the block-pairs renders the best estimators, but is less efficient than the one that considers independence. For both approaches, the CW method returns the best results.

The study includes numerical experiments and comparative analysis between the different strategies, as well as a revision of the fundamental concepts of the spatial statistic and the stochastic processes. Finally, the conclusions and possible future works are discussed.

CONTENIDO

AGRADECIMIENTOS	I
RESUMEN	III
ABSTRACT	V
1. PRELIMINARES Y BACKGROUND	1
1.1. Introducción	1
1.2. Tipos de datos en estadística espacial	7
1.3. Procesos estocásticos estacionarios	8
1.4. Modelo geoestadístico a utilizar	9
2. VEROSIMILITUD COMPUESTA POR BLOQUE	11
2.1. Verosimilitud compuesta por bloques (Eidsvik et al. (2014))	11
2.2. Verosimilitud por bloques (Acosta et al. (2021))	14
2.3. Dos enfoques para la selección de los bloques	15
2.3.1. Enfoques en una dimensión	15
2.3.2. Enfoques en una grilla bidimensional	16
3. INFERENCIA USANDO VEROSIMILITUD COMPUESTA POR BLOQUES	19
3.1. Propiedades del estimador de máxima verosimilitud compuesta	19
3.2. Cálculo de los estimadores de verosimilitud compuesta por bloques	23

4. EXPERIMENTOS NUMÉRICOS	25
4.1. Experimentos con el enfoque por filas (RW)	26
4.2. Experimentos con el enfoque por columnas (CW)	31
4.3. Comparación de enfoques	34
5. COMENTARIOS FINALES Y TRABAJOS A FUTURO	39
ANEXO 1: PRELIMINARES PROPIEDADES MATRICIALES	41
ANEXO 2: GRÁFICOS	43
ANEXO 3: TABLAS	48
BIBLIOGRAFÍA	51

Capítulo 1

PRELIMINARES Y BACKGROUND

1.1. Introducción

En los últimos años, se ha observado un crecimiento significativo en la magnitud y presencia de extensos conjuntos de datos científicos georeferenciados. Este aumento de tamaño se debe, en gran medida, a la proliferación de tecnologías como el GPS y la teledetección, además de la cada vez mayor capacidad de almacenamiento de bases de datos digitales. El interés creciente en la investigación climática ha situado estos conjuntos de datos en el centro de atención. Como consecuencia, se ha generado una creciente demanda de modelos estadísticos espaciales y metodologías más avanzadas para el análisis de esta información.

Los cálculos necesarios para llevar a cabo la inferencia y la predicción en los modelos de procesos Gaussianos espaciales, que constituyen el núcleo de la estadística espacial, presentan un desafío sustancial cuando se trata de conjuntos de datos a gran escala. Esto se debe a la compleja manipulación de matrices de covarianza de gran tamaño. En particular, el método de máxima verosimilitud es generalmente considerado el mejor método para estimar parámetros de modelos de covarianza, sin embargo, para campos aleatorios Gaussianos con una función de covarianza paramétrica, el cálculo

exacto de la verosimilitud requiere calcular inversas y determinantes de matrices de covarianza, y esta evaluación es lenta cuando el número de observaciones es sustancial.

Más precisamente, si consideramos $\{\mathbf{Y}(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}$ un campo aleatorio Gaussiano estacionario con media 0 y función de covarianza $\text{cov}(\mathbf{Y}(\mathbf{s}), \mathbf{Y}(\mathbf{s}')) = \sigma^2 \rho(\mathbf{s} - \mathbf{s}'; \boldsymbol{\phi})$, $\mathbf{s}, \mathbf{s}' \in \mathbb{R}^d$, $\sigma^2 > 0$, $\boldsymbol{\phi} \in \Phi \in \mathbb{R}^p$. El parámetro $\boldsymbol{\theta} = (\sigma, \boldsymbol{\phi})^\top$ puede ser estimado en base a un número finito de n observaciones $\mathbf{Y} = (\mathbf{Y}(\mathbf{s}_1), \dots, \mathbf{Y}(\mathbf{s}_n))^\top$.

La función de log-verosimilitud para un campo aleatorio Gaussiano, como el que se mencionó anteriormente, puede ser expresada como:

$$\ell(\mathbf{Y}; \boldsymbol{\theta}) = -\frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} \mathbf{Y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{Y}, \quad (1.1)$$

donde $\Sigma_{ij} = \sigma^2 \rho(\mathbf{s}_i - \mathbf{s}_j; \boldsymbol{\phi})$.

Entonces, una desventaja de las técnicas relacionadas con la verosimilitud es la eficiencia computacional de la función de verosimilitud. La evaluación de la Ecuación (1.1) requiere calcular la inversa y el determinante de una matriz $n \times n$ lo que, según los algoritmos más utilizados, requiere $O(n^3)$ pasos (Varin et al., 2011). Esto, para muchos conjuntos de datos con miles de observaciones significa un cálculo que requiere un tiempo prohibitivo. Por lo tanto, existe interés en encontrar aproximaciones a la función de verosimilitud que sean más eficientes de calcular, intentando al mismo tiempo no sacrificar tanta calidad de la estimación.

Se han propuesto varias estrategias para mitigar esta carga computacional. Por ejemplo la que fue propuesta por Vecchia (1988), que calculó densidades condicionales aproximadas a lo largo de una secuencia de puntos de muestreo ordenados arbitrariamente e ignorando las correlaciones de largo alcance. También, Stein et al. (2004) generalizaron la idea de Vecchia de varias maneras. Desarrollaron una variante del método para aproximar la función de verosimilitud restringida en lugar de la propia función de verosimilitud. Argumentaron que, en lugar de evaluar las densidades condicionales una observación cada vez, podría ser más eficiente hacerlo en bloques. Una estrategia alternativa es presentada por Furrer et al. (2006), y Shaby and Ruppert (2012) que utilizaron tapering a la matriz de covarianza, el cual consiste en

que la función de covarianza completa se multiplica por una función de correlación con soporte compacto, lo que da lugar a una matriz de covarianza sparse, esto que permite aprovechar algoritmos especializados. Otra metodología alternativa consiste en representar el proceso espacial en un subespacio de baja dimensión utilizando modelos de bajo rango (por ejemplo, [Banerjee et al. \(2008\)](#); [Cressie and Johannesson \(2008\)](#); [Stein \(2008\)](#)). Los campos aleatorios de Markov Gaussianos también son útiles para el cálculo rápido ([Aune et al., 2014](#); [Lindgren et al., 2011](#)).

Uno de los enfoques también tratados en la literatura, y que es foco de nuestro estudio, es el basado en la verosimilitud compuesta donde, de acuerdo a la definición propuesta por [Lindsay \(1988\)](#), sea $\mathbf{Y}$ un vector de variables aleatorias n-dimensional, con función de densidad de probabilidad $f(y; \boldsymbol{\theta})$ y $\boldsymbol{\theta} \in \Theta$ un vector p-dimensional de parámetros desconocidos. Definamos $\{\mathcal{A}_1, \dots, \mathcal{A}_K\}$ un conjunto de datos marginales o condicionales asociados a la verosimilitud $\mathcal{L}_k(\boldsymbol{\theta}; \mathcal{A}_k) \propto f(y \in \mathcal{A}_k; \boldsymbol{\theta})$, la verosimilitud compuesta se define como,

$$CL(\boldsymbol{\theta}; y) = \prod_{k=1}^K \mathcal{L}_k(\boldsymbol{\theta}; y)^{\omega_k}, \quad (1.2)$$

donde w_k son pesos no negativos que no dependen de $\boldsymbol{\theta}$ y con función de log-verosimilitud compuesta

$$\ell_{CL}(\boldsymbol{\theta}; y) = \sum_{k=1}^K \omega_k \log(\mathcal{L}_k(\boldsymbol{\theta}; y)) = \sum_{k=1}^K \omega_k \ell_k(\boldsymbol{\theta}; y). \quad (1.3)$$

En particular, en este estudio se trabajará la estrategia que considera los conjuntos $\mathcal{A}_k$ como bloques espaciales y la motivación detrás de esta elección es que captura gran parte de la dependencia espacial, al tiempo que integra el concepto de divide y vencerás de la verosimilitud compuesta, que reduce la complejidad computacional y facilita la computación paralela rápida. A diferencia de los métodos de base de bajo rango ([Banerjee et al., 2008](#); [Cressie and Johannesson, 2008](#); [Stein, 2008](#)); y los métodos de grilla incrustada (embedded lattice methods) ([Lindgren et al., 2011](#)), la estimación y predicción con la verosimilitud compuesta por bloques permite trabajar

directamente con procesos espaciales continuos de rango completo.

Siguiendo con las estrategias, [Curriero and Lele \(1999\)](#) exploraron un enfoque en el que estiman eficientemente semivariogramas, explotando la verosimilitud compuesta, introducida previamente, para la estimación de parámetros de modelos jerárquicos espaciales. Los mismos principios se aplican también al modelado de datos binarios en otro trabajo de [Heagerty and Lele \(1998\)](#). La idea principal propuesta es construir una función aproximada de log-verosimilitud añadiendo log-verosimilitudes marginales incluso cuando estos componentes no representan necesariamente réplicas independientes. Por otra parte, [Caragea and Smith \(2007\)](#) utilizaron promedios de bloques grandes o productos de bloques pequeños para construir estimadores. También utilizaron esquemas híbridos que combinaban sus métodos de bloques grandes y pequeños, y estudiaron sus propiedades asintóticas.

Recientemente, [Bevilacqua et al. \(2012\)](#) consideraron una verosimilitud compuesta ponderada basada en la diferencia de pares Gaussianos en el contexto espaciotemporal y [Eidsvik et al. \(2014\)](#) desarrollaron una verosimilitud compuesta por bloques Gaussianos por pares en la misma línea de [Caragea and Smith \(2007\)](#). Y además, [Bevilacqua et al. \(2016\)](#) utilizaron la verosimilitud compuesta para inferencia en campos aleatorios Gaussianos multivariados.

Como ejemplo de la problemática del uso de la máxima verosimilitud completa en conjuntos de datos grandes y como ésta se ve atenuada por el uso de la verosimilitud compuesta por bloque, [Acosta et al. \(2021\)](#) estudiaron una variante del tamaño de muestra efectivo (ESS) para datos espaciales de gran tamaño en el contexto de la verosimilitud por bloque. Utilizaron para la estimación del rango un modelo de covarianza de Matérn, con grillas de $n = 16k \times 8k$, con $k = 1, \dots, 8$ y tres configuraciones de bloques (2, 4 y 8 bloques idénticos). Observaron que el tiempo de ejecución de la verosimilitud completa incrementa sustancialmente luego de sobrepasar las 2000 observaciones, mientras que las opciones con verosimilitud compuesta se posicionaron en rangos de tiempo mucho menores (Figura 1.1). A partir de este estudio, podemos deducir que para imágenes de dimensión $n = 1000 \times 1000 = 1.000.000$ el tiempo de computación realmente sería prohibitivo.

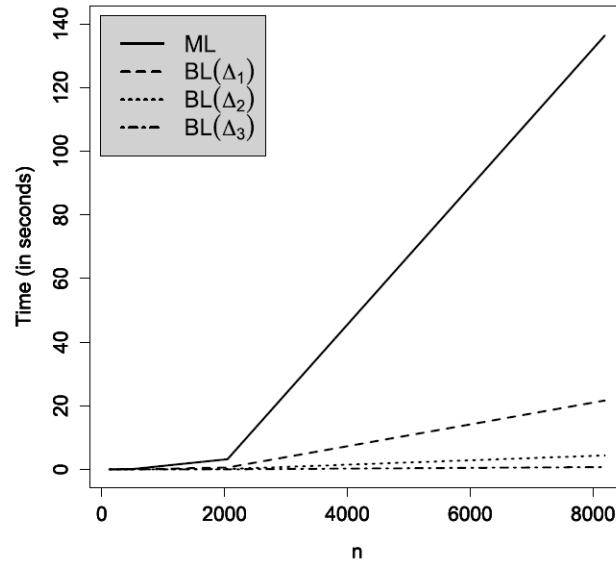


Figura 1.1: Gráfico obtenido desde el paper [Acosta et al. \(2021\)](#). Tiempo en segundos versus el tamaño de la muestra cuando se evalúa la función objetivo del método de máxima verosimilitud usual (ML) y la verosimilitud compuesta por bloque (BL). El método BL fue implementado con tres diferentes configuraciones: Δ_1 (dos bloques idénticos), Δ_2 (4 bloques idénticos) y Δ_3 (8 bloques idénticos).

En el trabajo de [Eidsvik et al. \(2014\)](#), se aprovechó la estructura de la verosimilitud compuesta por pares de bloques para utilizar la computación paralela, probando la estimación de parámetros en una unidad de procesamiento gráfico (GPU NVIDIA). Este enfoque resultó en un aumento significativo de la eficiencia computacional. Por lo tanto, resulta interesante explorar la verosimilitud compuesta por bloques utilizada en su estudio, junto con el algoritmo que proponen, en un entorno que no disponga de estas capacidades avanzadas (solo utilizando CPUs), y comparar su desempeño con otro enfoque de la verosimilitud compuesta, como lo es el utilizado por [Acosta et al. \(2021\)](#). En este último, no se utilizan pares de bloques, sino que se asume independencia entre todos los bloques. Por lo que, comparar estos dos métodos será el objetivo de este estudio.

Además, también se podrán a prueba los dos enfoques propuestos por [Mukerjee \(2023\)](#) para la forma de seleccionar los bloques. El primer enfoque, por filas (RW), agrupa los elementos más cercanos para conformar el bloque, mientras que el segundo,

por columnas (CW), considera que elementos distantes deberían ser parte del mismo bloque.

En lo que sigue del capítulo 1, se exponen los conceptos fundamentales sobre estadística espacial. Los métodos de verosimilitud compuesta por bloques que se van a comparar y los dos enfoques para la selección de los bloques son definidos en el capítulo 2. El método para la inferencia de parámetros se encuentra en el capítulo 3, al igual que el algoritmo con el que se estiman los parámetros. Los experimentos numéricos se muestran en el capítulo 4. Finalmente, las conclusiones y posibles trabajos futuros se discuten en el capítulo 5.

1.2. Tipos de datos en estadística espacial

En estadística espacial, es esencial tener en cuenta que las observaciones provienen de un proceso que varía en el espacio, representado como $\{\mathbf{Y}(\mathbf{s}) : \mathbf{s} \in D \subset \mathbb{R}^d\}$. Este proceso está asociado a un atributo $\mathbf{Y}(\mathbf{s})$ que estamos observando y a un dominio espacial D , que comúnmente es bidimensional ($D \subset \mathbb{R}^2$), aunque en ocasiones puede ser tridimensional, como en el caso de imágenes en tres dimensiones. Además, notemos que al considerar la dinámica espacial, se agrega una dimensión temporal. Por ejemplo, en el caso bidimensional, las observaciones se pueden representar como pares $(\mathbf{s}, t) \in \mathbb{R}^2 \times \mathbb{R}_+$.

La amplia gama de situaciones y aplicaciones que surgen en este contexto da paso a una variedad de tipos de datos espaciales, que se distinguen por las características del dominio D , por ejemplo, los que se presentan a continuación (Gaetan and Guyon, 2010; Schabenberger and Gotway, 2005):

- **Datos geoestadísticos:** El dominio D es un conjunto fijo y continuo, es decir, $\mathbf{Y}(\mathbf{s})$ puede ser observado en cualquier lugar de D y estos puntos son no-estocásticos. Es importante asociar la continuidad con el dominio, no con el atributo que está siendo medido. El valor de este atributo puede ser continuo o discreto. Por ejemplo, la temperatura del aire puede ser medido teóricamente en cualquier lugar, sin embargo, en la práctica la recolección no es tan exhaustiva.
- **Datos en grilla o areales:** El dominio es fijo y discreto, en otras palabras, no-aleatorio y contable. El número de ubicaciones puede ser infinito, pero lo que importa es que debe ser numerable. Aquí, los puntos $\mathbf{s}$ pueden no ser puntos en el espacio sino que regiones geográficas. Ejemplos; datos obtenidos de una zona censal o datos de teledetección notificados por píxeles.
- **Datos puntuales:** Aquí, el conjunto de lugares de observación $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, $\mathbf{x}_i \in D \subset \mathbb{R}^d$ es aleatorio, junto con el número $n = n(\mathbf{x})$ de lugares de observación; $\mathbf{x}$ es el resultado de un proceso espacial puntual observado en la ventana D . Se dice que el proceso $\mathbf{Y}$ está marcado si en cada $\mathbf{x}_i$ registramos

un atributo. Por ejemplo; la ubicación y profundidad de los cráteres lunares, el lugar de aparición de las malas hierbas y la intensidad de la luz en ese lugar pueden considerarse patrones de puntos marcados.

- **Datos espacio - temporales:** En este caso, es un proceso que varía en el espacio y tiempo, que se representa como $\{\mathbf{Y}(\mathbf{s}, t) : \mathbf{s} \in D(t) \subset \mathbb{R}^2, t \in T\}$, en donde la observación de un atributo $\mathbf{Y}$ en función del tiempo t y la ubicación espacial $\mathbf{s}$, denotado como $\mathbf{Y}(\mathbf{s}, t)$. El componente espacial en esta definición puede ser discreto o continuo, fijo o aleatorio, al igual que el componente temporal del proceso. Por ejemplo, la distribución de partículas en la atmósfera puede ser vista como un proceso espacio-temporal con dominio espacial continuo y tiempo continuo. En contraste, los datos sobre ingresos familiares disponibles por condados el 15 de abril de cada año constituyen un proceso de retícula espacial y un proceso de tiempo discreto.

1.3. Procesos estocásticos estacionarios

Algunos conceptos elementales para la comprensión del problema a tratar se presentan a continuación (Gaetan and Guyon, 2010; van Lieshout, 2019).

Consideremos $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad, D un conjunto de sitios y $(E, \mathcal{E})$ un espacio de medida, entonces se tienen las siguientes definiciones:

Definición 1.3.1. Proceso Estocástico: Un proceso estocástico tomando valores en E es una familia $X = \{X_s, s \in D\}$ de variables aleatorias definidas en $(\Omega, \mathcal{F}, \mathbb{P})$ y tomando valores en $(E, \mathcal{E})$. $(E, \mathcal{E})$ es llamado el espacio estado del proceso y D el conjunto (espacial) de sitios en donde el proceso está definido.

Definición 1.3.2. Proceso de segundo orden: X es un proceso de segundo orden (o campo aleatorio) si para todo $s \in D$, $\mathbb{E}(X_s^2) < \infty$. La media de X (que necesariamente existe) es la función $m : D \rightarrow \mathbb{R}$ definida por $m(s) = \mathbb{E}(X_s)$. La covarianza de X es la función $C : D \times D \rightarrow \mathbb{R}$ definido para todo s, t por $C(s, t) = \text{Cov}(X_s, X_t)$.

Definición 1.3.3. Proceso Gaussiano: X es un proceso Gaussiano (o campo aleatorio Gaussiano) en D si para todo subconjunto $\Lambda \subset D$ y secuencia de valores reales $a = (a_s, s \in \Lambda)$, $\sum_{s \in \Lambda} a_s X_s$ es una variable aleatoria Gaussiana.

Definición 1.3.4. Proceso débilmente estacionario: Un campo aleatorio X es débilmente estacionario si $\mathbb{E}(X_s^2) < \infty$ para todo $\mathbf{s} \in \mathbb{R}^d$, $\mathbb{E}(X_s) \equiv k$ es constante y $\text{Cov}(X_{\mathbf{s}_1}, X_{\mathbf{s}_2}) = C(\mathbf{s}_1 - \mathbf{s}_2)$ para alguna $C : \mathbb{R}^d \rightarrow \mathbb{R}$.

Definición 1.3.5. Proceso intrínsecamente estacionario: Un campo aleatorio X es intrínsecamente estacionario si $\mathbb{E}(X_s^2) < \infty$ para todo $\mathbf{s} \in \mathbb{R}^d$, $\mathbb{E}(X_s) \equiv k$ es constante y $\text{Var}(X_{\mathbf{s}_2} - X_{\mathbf{s}_1}) = f(\mathbf{s}_2 - \mathbf{s}_1)$ para alguna $f : \mathbb{R}^d \rightarrow \mathbb{R}$.

Definición 1.3.6. Proceso isotrópico: Un campo aleatorio X es llamado isotrópico si la función de covarianza depende exclusivamente de la distancia entre las ubicaciones espaciales, es decir, para cada $\mathbf{s}, \mathbf{s}' \in D$, $\text{Cov}(X_{\mathbf{s}_1}, X_{\mathbf{s}_2})$ sólo depende de $\|\mathbf{s}_1 - \mathbf{s}_2\|$, donde $\|\cdot\|$ denota la norma Euclidiana.

1.4. Modelo geoestadístico a utilizar

Para el modelo geoestadístico que vamos a utilizar, se va a asumir una ubicación $\mathbf{s} \in D \subset \mathbb{R}^2$, una variable de respuesta Gaussiana $Y(\mathbf{s})$ junto con un vector $p \times 1$ de variables explicativas espacialmente referenciadas $\mathbf{x}(\mathbf{s})$ que están asociados a través de un modelo de regresión espacial

$$Y(\mathbf{s}) = \mathbf{x}^\top(\mathbf{s})\boldsymbol{\beta} + w(\mathbf{s}) + \epsilon(\mathbf{s}), \quad (1.4)$$

donde $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ es el parámetro de regresión, y $\epsilon(\mathbf{s}) \sim N(0, \tau^2)$ es el error independiente. El residuo espacial $w(\mathbf{s})$ provee estructura de dependencia, capturando el efecto de las covariables no observadas con patrón espacial. Denotamos la colección de todas los parámetros de covarianza $\boldsymbol{\theta}$, los cuales incluyen el efecto nugget τ^2 . Algunos modelos de covarianza comunes para $\text{cov}(w(\mathbf{s}), w(\mathbf{s}')) = \mathbf{C}(\mathbf{s}, \mathbf{s}')$ son el exponencial, Matérn y Cauchy.

Asumimos que los datos están disponibles en n ubicaciones $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$, y se denota a la colección de datos como $\mathbf{Y} = (Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))^\top$. Entonces, $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, donde $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \mathbf{C} + \tau^2 \mathbf{I}_n$, con $\mathbf{C}(i, j) = \text{cov}(w(\mathbf{s}_i), w(\mathbf{s}_j))$. Además, la fila i -ésima de la matrix $\mathbf{X}$ contiene las variables explicatorias $\mathbf{x}^\top(\mathbf{s})$. Ignorando un escalar que no depende de $\boldsymbol{\beta}$ y $\boldsymbol{\theta}$, la log-verosimilitud es

$$\ell(\mathbf{Y}; \boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}). \quad (1.5)$$

Notando que $\boldsymbol{\Sigma}$ es $n \times n$, la dificultad con el método de máxima verosimilitud usual es evidente; evaluar la log-verosimilitud requiere computar $|\boldsymbol{\Sigma}|$ y una forma cuadrática que incluye $\boldsymbol{\Sigma}^{-1}$, ambas situaciones son intratables para valores grandes de n .

Capítulo 2

VEROSIMILITUD COMPUESTA POR BLOQUE

2.1. Verosimilitud compuesta por bloques (Eidsvik et al. (2014))

Eidsvik et al. (2014) en su artículo proponen la verosimilitud compuesta por bloque de la siguiente manera. Tomemos la región $D \in \mathbb{R}^d$, $d = 2$ o 3 , y dividámosla en M bloques $D_1, \dots, D_M$, con $\cup_k D_k = D$, $D_k \cap D_l = \emptyset$, para todos los pares de bloques k, l . Denotemos la respuesta en el bloque $k = 1, \dots, M$ como $\mathbf{Y}_k = \{\mathbf{Y}(\mathbf{s}_i); \mathbf{s}_i \in D_k\}$. Además, el número de sitios en el bloque k es n_k , con $\sum_k n_k = n$. Sea $\mathbf{Y}_{kl} = (\mathbf{Y}_k^\top, \mathbf{Y}_l^\top)^\top$ la colección de datos en los bloques k y l . A partir de la Ecuación (1.5), definimos la log-verosimilitud compuesta por bloques (CL) como

$$\ell_{CL}(\mathbf{Y}; \boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_{k=1}^{M-1} \sum_{l>k} \ell(\mathbf{Y}_{kl}; \boldsymbol{\beta}, \boldsymbol{\theta}) \quad (2.1)$$

$$= \sum_{k=1}^{M-1} \sum_{l>k} \left[-\frac{1}{2} \log |\boldsymbol{\Sigma}_{kl}| - \frac{1}{2} (\mathbf{Y}_{kl} - \mathbf{X}_{kl} \boldsymbol{\beta})^\top \boldsymbol{\Sigma}_{kl}^{-1} (\mathbf{Y}_{kl} - \mathbf{X}_{kl} \boldsymbol{\beta}) \right]. \quad (2.2)$$

En la ecuación anterior, $\mathbf{X}_{kl} = (\mathbf{X}_k^\top, \mathbf{X}_l^\top)^\top$ es la colección de todas las covariables en el bloque k y l , y Σ_{kl} es la matriz de covarianza de tamaño $(n_k + n_l) \times (n_k + n_l)$

$$\Sigma_{kl} = \begin{bmatrix} \Sigma_{kl}(1, 1) & \Sigma_{kl}(1, 2) \\ \Sigma_{kl}(2, 1) & \Sigma_{kl}(2, 2) \end{bmatrix}, \quad (2.3)$$

donde $\Sigma_{kl}(1, 1)$ es la matriz de covarianza de tamaño $n_k \times n_k$ de $\mathbf{Y}_k$, $\Sigma_{kl}(2, 2)$ es la matriz de covarianza de tamaño $n_l \times n_l$ de $\mathbf{Y}_l$, y $\Sigma_{kl}(1, 2) = \Sigma_{kl}^\top(2, 1)$ es la matriz de covarianza cruzada de tamaño $n_k \times n_l$ entre $\mathbf{Y}_k$ y $\mathbf{Y}_l$. Si $M = 1$ o 2 , la verosimilitud compuesta por bloques es la Ecuación (2.2) es igual a la verosimilitud completa de la Ecuación (1.5); si $M = n$ obtenemos la verosimilitud por pares. La verosimilitud por bloques es una elección natural para modelos espaciales, dado que el número de bloques M representa la compensación entre la eficiencia computacional y estadística. La expresión de la verosimilitud compuesta se puede simplificar si se omiten pares de bloques distantes en la Ecuación (2.2), asumiendo dependencia despreciable entre bloques que no son vecinos. Sea N_k que denote los vecinos del bloque k . La Figura 2.1 muestra un diseño de bloques regulares con $M = 5 \cdot 5 = 25$ bloques en un dominio en 2D. Ahí, bloques son vecinos si ellos comparten un borde en común. Los bordes del bloque 13 se muestran en la Figura 2.1.



Figura 2.1: Se divide el dominio espacial en bloques regulares. Cada bloque se comunica por pares con cada uno de sus vecinos, un bloque interior tiene ocho vecinos. El bloque con índice 13 tiene cuatro vecinos con índices superiores (líneas negras).

El conjunto de vecinos N_k puede ser dividido en una parte hacia adelante $N_k^{\rightarrow} = \{l > k\} \cap \{l \in N_k\}$ y una parte hacia atrás $N_k^{\leftarrow} = \{l < k\} \cap \{l \in N_k\}$. Estos dos conjuntos se muestran utilizando líneas negras y gris en los bordes en la Figura 2.1. Si sólo consideramos esos bloques de vecinos, la sumatoria en la Ecuación (2.2) evaluada sobre $l \in N_k^{\rightarrow}$ queda como

$$\begin{aligned} \ell_{CL}(\mathbf{Y}; \boldsymbol{\beta}, \boldsymbol{\theta}) &= \sum_{k=1}^{M-1} \sum_{l \in N_k^{\rightarrow}} \left[-\frac{1}{2} \log |\boldsymbol{\Sigma}_{kl}| - \frac{1}{2} (\mathbf{Y}_{kl} - \mathbf{X}_{kl} \boldsymbol{\beta})^{\top} \boldsymbol{\Sigma}_{kl}^{-1} (\mathbf{Y}_{kl} - \mathbf{X}_{kl} \boldsymbol{\beta}) \right] \\ &= \sum_j \left[-\frac{1}{2} \log |\boldsymbol{\Sigma}_j| - \frac{1}{2} (\mathbf{Y}_j - \mathbf{X}_j \boldsymbol{\beta})^{\top} \boldsymbol{\Sigma}_j^{-1} (\mathbf{Y}_j - \mathbf{X}_j \boldsymbol{\beta}) \right]. \end{aligned} \quad (2.4)$$

La notación abreviada con el índice j representa el conjunto de aristas $(k, l) | l \in N_k^{\rightarrow}$. Por ejemplo, en la Figura 2.1 (a), el conjunto de $j = (k, l)$'s es $(1, 2), (1, 6), (1, 7), (2, 3), (2, 6), \dots, (24, 25)$. La notación de bordes induce la correspondiente abreviatura para las variables de pares de bloques $\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}_{kl}$, $\mathbf{Y}_j = \mathbf{Y}_{kl}$, y $\mathbf{X}_j = \mathbf{X}_{kl}$ definidas en

las Ecuaciones (2.2) y (2.3).

Para datos en grillas regulares, y con un diseño de bloques regular, la distancia relativa entre puntos en los pares de bloques es la misma, generando matrices de covarianza Σ_j idénticas para todas las configuraciones iguales de pares de bloques vecinos $j = (k, l)$, bajo suposiciones de estacionalidad e isotropía en un campo aleatorio. En este caso, sólo hay que calcular unos pocos determinantes e inversas para cada evaluación de la verosimilitud compuesta por bloques.

2.2. Verosimilitud por bloques (Acosta et al. (2021))

El artículo de Acosta et al. (2021) se enfoca en explorar una variante del tamaño de muestra efectiva aplicada a datos espaciales de gran escala, específicamente dentro del marco de la verosimilitud por bloque. A diferencia de la verosimilitud compuesta propuesta por Eidsvik et al. (2014), la cual opera con pares de bloques $Y_{kl} = (Y_k, Y_l)$, Acosta et al. (2021) asume independencia entre todos los bloques. Esto implica que si dividimos la región D en M bloques $D_1, \dots, D_M$, con $\cup_k D_k = D$, $D_k \cap D_l = \emptyset$, para todos los pares de bloques k, l entonces podemos denotar el valor de las variables en el bloque $k = 1, \dots, M$ como $\mathbf{Y}_k = \{\mathbf{Y}(\mathbf{s}_i); \mathbf{s}_i \in D_k\}$. Además, el número de sitios en el bloque k es n_k , con $\sum_k n_k = n$. Basandose en la ecuación (1.5), la función de log-verosimilitud por bloques (BL) que se considera en su estudio es

$$\begin{aligned} \ell_{BL}(\mathbf{Y}; \boldsymbol{\beta}, \boldsymbol{\theta}) &= \sum_{k=1}^M \ell(\mathbf{Y}_k; \boldsymbol{\beta}, \boldsymbol{\theta}) \\ &= \sum_{k=1}^M \left[-\frac{1}{2} \log |\Sigma_k| - \frac{1}{2} (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta})^\top \Sigma_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta}) \right]. \end{aligned} \quad (2.5)$$

En la ecuación, $\mathbf{X}_k$ es la colección de todas las covariables en el bloque k , y Σ_k es la matriz de covarianza de tamaño $n_k \times n_k$ de $\mathbf{Y}_k$.

2.3. Dos enfoques para la selección de los bloques

En esta sección se explorarán dos enfoques posibles para la elección de los bloques, el enfoque por filas y por columnas, RW y CW por sus siglas en inglés, respectivamente. El primero es utilizado en [Acosta et al. \(2021\)](#), siendo el más natural, y el segundo es propuesto por [Mukerjee \(2023\)](#).

2.3.1. Enfoques en una dimensión

Sea n la cantidad de observaciones y $n = mb$, donde m y b son enteros positivos, y supongamos que m es la cantidad de bloques que queremos obtener de tamaño b . Dividir $\{1, \dots, n\}$ datos, según el enfoque RW, en m bloques de tamaño b , está dado por

$$B_u^{\text{row}} = \{(u-1)b + j : j = 1, \dots, b\}, \quad u = 1, \dots, m.$$

Se le llama enfoque por filas dado que los bloques corresponden a las filas de un arreglo $1, \dots, n$ ordenado de manera natural en una matriz de orden $m \times b$, por ejemplo, si $m = 3$ y $b = 5$, entonces los bloques están dados por las filas de la matriz

$$\begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \rightarrow B_1^{\text{row}} \\ 6 & 7 & 8 & 9 & 10 \rightarrow B_2^{\text{row}} \\ 11 & 12 & 13 & 14 & 15 \rightarrow B_3^{\text{row}} \end{array}$$

Por otra parte, en el enfoque por columnas, la partición de los datos en m bloques de tamaño b , está dada por

$$B_u^{\text{col}} = \{u + (j-1)m : j = 1, \dots, b\}, \quad u = 1, \dots, m.$$

Es llamado enfoque por columnas dado que esos bloques corresponden a las columnas de un arreglo $1, \dots, n$ ordenado de manera natural en una matriz de orden $b \times m$, por

ejemplo, si $m = 3$ y $b = 5$, entonces los bloques están dados por las filas de la matriz

$$\begin{array}{ccc}
 1 & 2 & 3 \\
 4 & 5 & 6 \\
 7 & 8 & 9 \\
 10 & 11 & 12 \\
 13 & 14 & 15 \\
 \downarrow & \downarrow & \downarrow \\
 B_1^{\text{col}} & B_2^{\text{col}} & B_3^{\text{col}}
 \end{array}$$

El enfoque RW parece más natural que el CW, en este último los puntos en cada bloque están mucho más esparcidos en vez de estar ordenados consecutivamente.

2.3.2. Enfoques en una grilla bidimensional

La misma noción de los enfoques en una dimensión puede ser extendida a dos dimensiones utilizando el producto cartesiano de conjuntos.

Sea una grilla de orden $n_1 \times n_2$ ($n_1, n_2 \geq 2$), con las filas y columnas indexadas en pares ordenados representando coordenadas de puntos espaciales en un plano. Sea I el conjunto de esos $n_1 n_2$ pares, y supongamos que $n_k = m_k b_k$, con $k = 1, 2$, donde m_1, b_1, m_2, b_2 son enteros positivos. Usando el producto Cartesiano, se divide I en $m = m_1 m_2$ bloques de tamaño $b = b_1 b_2$.

Los bloques RW en este caso serían

$$B_{u_1 u_2}^{\text{row}} = B_{u_1}^{(1)\text{row}} \times B_{u_2}^{(2)\text{row}} \quad (u_1 = 1, \dots, m_1; u_2 = 1, \dots, m_2),$$

mientras que los del CW quedarían como

$$B_{u_1 u_2}^{\text{col}} = B_{u_1}^{(1)\text{col}} \times B_{u_2}^{(2)\text{col}} \quad (u_1 = 1, \dots, m_1; u_2 = 1, \dots, m_2),$$

donde $\times$ es el producto Cartesiano, y para $u_k = 1, \dots, m_k$ ($k = 1, 2$),

$$B_{u_k}^{(k)\text{row}} = \{(u_k - 1)b_k + j : j = 1, \dots, b_k\}$$

y

$$B_{u_k}^{(k)\text{col}} = \{u_k + (j - 1)m_k : j = 1, \dots, b_k\}.$$

Por ejemplo, para el caso $n_1 \times n_2 = 16 \times 8$, $m = 4$ y $b = 32$, los bloques en la grilla para el RW, se verían como en la Figura 2.2 (a), mientras que los bloques para CW se verían como en la Figura 2.2 (b) donde los puntos de igual color pertenecen al mismo bloques.

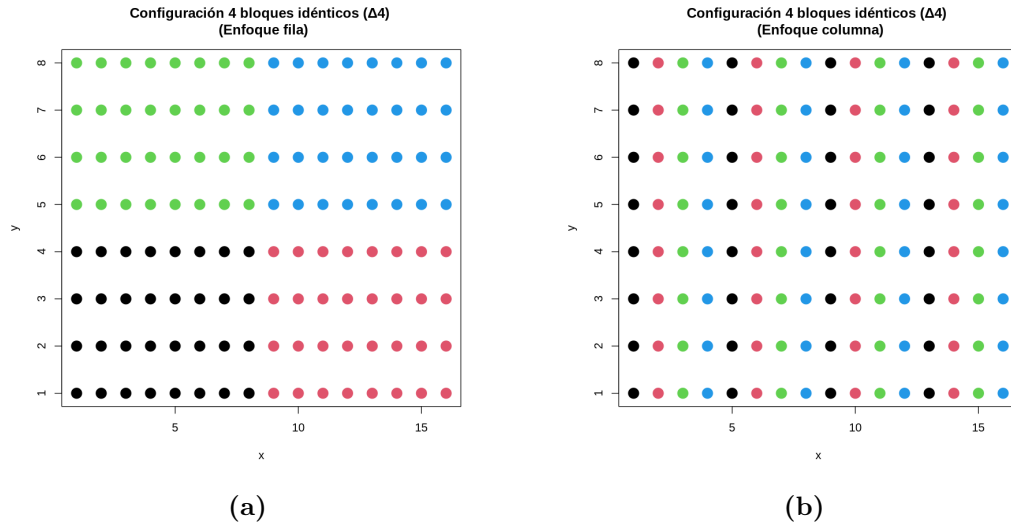


Figura 2.2: Grillas regulares divididas en 4 bloques de tamaño 32 según el enfoque por filas (a) y el enfoque por columnas (b), con $m = m_1 \times m_2 = 4 \times 1 = 4$ y $b = b_1 \times b_2 = 4 \times 8 = 32$.

En el caso del CW, la noción de vecino hacia adelante será el bloque a la derecha, es decir, $N_k^{\rightarrow} = \{k + 1\}$.

Con lo anterior claro, es posible saber la cantidad de matrices que debe invertir cada método con cada enfoque de configuración de bloques. En el caso de BL con RW, se deben invertir $m = m_1 m_2$ matrices de tamaño $b \times b$, al igual en el enfoque CW. Mientras que con la CL y enfoque RW, como se consideran pares de bloques, se

deben invertir $4(m_1 - 2)(m_2 - 1) + 5(m_2 - 1) + m_1 - 1$ matrices de tamaño $2b \times 2b$, mientras que con el enfoque CW se invierten m matrices de tamaño $2b \times 2b$. Esto ya nos da la intuición de que el método BL tardaría menos tiempo en ejecutarse que el CL, pues invierte menos matrices y de menor tamaño.

INFERENCIA USANDO VEROSIMILITUD COMPUESTA POR BLOQUES

3.1. Propiedades del estimador de máxima verosimilitud compuesta

Los estimadores de máxima verosimilitud compuesta para $\boldsymbol{\theta}$ y $\boldsymbol{\beta}$, están dados por

$$(\hat{\boldsymbol{\beta}}_{CL}, \hat{\boldsymbol{\theta}}_{CL}) = \operatorname{argmax}_{\boldsymbol{\beta}, \boldsymbol{\theta}} [\ell_{CL}(\mathbf{Y}; \boldsymbol{\beta}, \boldsymbol{\theta})],$$

donde $\ell_{CL}(\mathbf{Y}; \boldsymbol{\beta}, \boldsymbol{\theta})$ está definida en la Ecuación (2.4).

En general, los estimadores de máxima verosimilitud compuesta son conocidos por ser consistentes y asintóticamente normales bajo las mismas condiciones que los estimadores de máxima verosimilitud ([Lindsay \(1988\)](#)). En el caso de los procesos espaciales Guassianos, condiciones como las que se encuentran en [Mardia and Marshall \(1984\)](#) proporcionan las propiedades asintóticas deseadas para el estimador de Máxima Verosimilitud Compuesta resultante desde una perspectiva asintótica de dominio creciente (por ejemplo, [Curriero and Lele \(1999\)](#); [Varin \(2008\)](#)).

Para comenzar el análisis, definamos la función vectorial de score para la verosimilitud compuesta por bloques, como $\partial \ell_{CL}(\mathbf{Y}; \boldsymbol{\beta}, \boldsymbol{\theta}) / \partial \theta_r$, donde $\ell_{CL}(\mathbf{Y}; \boldsymbol{\beta}, \boldsymbol{\theta})$ está definida en la Ecuación (2.4) y $r = 1, \dots, R$, con R es la dimensión del parámetro $\boldsymbol{\theta}$. Para simplificar la notación, asumiremos aquí que $\mathbf{Y}$ es un campo Gaussiano de media cero. Derivando la Ecuación (2.4) con respecto a θ_r , el score puede ser expresado como

$$\frac{\partial \ell_{CL}(\mathbf{Y}; \boldsymbol{\theta})}{\partial \theta_r} = \sum_j \left[-\frac{1}{2} \frac{\partial \log |\boldsymbol{\Sigma}_j|}{\partial \theta_r} - \frac{1}{2} \frac{\partial (\mathbf{Y}_j^\top \boldsymbol{\Sigma}_j^{-1} \mathbf{Y}_j)}{\partial \theta_r} \right]. \quad (3.1)$$

Las derivadas del log determinante y la forma cuadrática son

$$\frac{\partial \log |\boldsymbol{\Sigma}_j|}{\partial \theta_r} = \text{traza} \left(\mathbf{Q}_j \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_r} \right), \quad \frac{\partial (\mathbf{Y}_j^\top \mathbf{Q}_j \mathbf{Y}_j)}{\partial \theta_r} = -\mathbf{Y}_j^\top \mathbf{Q}_j \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_r} \mathbf{Q}_j \mathbf{Y}_j.$$

Aquí, la matriz de precisión es $\mathbf{Q}_j = \boldsymbol{\Sigma}_j^{-1}$. La función de score se expresa como:

$$\frac{\partial \ell_{CL}(\mathbf{Y}; \boldsymbol{\theta})}{\partial \theta_r} = \sum_j \left[-\frac{1}{2} \text{traza} \left(\mathbf{Q}_j \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_r} \right) + \frac{1}{2} \mathbf{Y}_j^\top \mathbf{Q}_j \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_r} \mathbf{Q}_j \mathbf{Y}_j \right]. \quad (3.2)$$

Luego, usando las propiedades de las formas cuadráticas (A1.5), la esperanza de (3.1) es

$$E \left(\frac{\partial \ell_{CL}(\mathbf{Y}; \boldsymbol{\theta})}{\partial \theta_r} \right) = \sum_j \left[-\frac{1}{2} \text{traza} \left(\mathbf{Q}_j \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_r} \right) + \frac{1}{2} \text{traza} \left(\mathbf{Q}_j \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_r} \right) \right] = 0,$$

revelando que el score compuesto por bloques es una función de estimación insesgada para $\boldsymbol{\theta}$ para cualquier esquema de bloques. Esto también se obtiene directamente pues la Ecuación (2.4) es una suma de log-verosimilitudes.

Como es típico de los estimadores asintóticamente normales resultantes de funciones de estimación insesgadas, la covarianza asintótica de $\hat{\boldsymbol{\theta}}_{CL}$ bajo dominios crecientes asintóticos tiene forma de sándwich (Godambe (1960)), y entonces $\hat{\boldsymbol{\theta}}_{CL} \sim N(\boldsymbol{\theta}, \mathbf{G}^{-1})$ bajo condiciones apropiadas (Varin (2008)). Aquí, la matriz de información de Godambe es

$$\mathbf{G} = \mathbf{G}(\boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta})\mathbf{J}(\boldsymbol{\theta})^{-1}\mathbf{H}(\boldsymbol{\theta}),$$

donde

$$\mathbf{H}(\boldsymbol{\theta}) = -\mathbb{E}\left(\frac{\partial^2 \ell_{CL}(\mathbf{Y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}\right), \quad \mathbf{J}(\boldsymbol{\theta}) = \text{var}\left(\frac{\partial \ell_{CL}(\mathbf{Y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right),$$

son la matriz de sensibilidad y variabilidad, respectivamente.

Podemos notar que en el caso de la función de verosimilitud completa se tiene que $\mathbf{H}(\boldsymbol{\theta})\mathbf{J}(\boldsymbol{\theta})^{-1} = \mathbf{I}$, entonces $\mathbf{G}(\boldsymbol{\theta})$ es la Información de Fisher. Para la verosimilitud compuesta por bloques se pueden obtener las expresiones analíticas de $\mathbf{H}(\boldsymbol{\theta})$ y $\mathbf{J}(\boldsymbol{\theta})$. Para calcular la matriz $\mathbf{H}(\boldsymbol{\theta})$, usamos la propiedad de la traza (A1.4), en la Ecuación (3.2), obteniendo las segundas derivadas

$$\begin{aligned} \frac{\partial^2 \ell_{CL}(\mathbf{Y}; \boldsymbol{\theta})}{\partial \theta_r \partial \theta_s} &= \sum_j \left[-\frac{1}{2} \text{traza}(\mathbf{Q}_j \frac{\partial^2 \boldsymbol{\Sigma}_j}{\partial \theta_r \partial \theta_s}) + \frac{1}{2} \text{traza}(\mathbf{Q}_j \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_s} \mathbf{Q}_j \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_r}) \right. \\ &\quad - \frac{1}{2} \mathbf{Y}_j^\top \mathbf{Q}_j \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_s} \mathbf{Q}_j \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_r} \mathbf{Q}_j \mathbf{Y}_j + \frac{1}{2} \mathbf{Y}_j^\top \mathbf{Q}_j \frac{\partial^2 \boldsymbol{\Sigma}_j}{\partial \theta_s \partial \theta_r} \mathbf{Q}_j \mathbf{Y}_j \\ &\quad \left. - \frac{1}{2} \mathbf{Y}_j^\top \mathbf{Q}_j \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_r} \mathbf{Q}_j \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_s} \mathbf{Q}_j \mathbf{Y}_j \right]. \\ &= \sum_j \left[-\frac{1}{2} \text{traza}(\mathbf{Q}_j \frac{\partial^2 \boldsymbol{\Sigma}_j}{\partial \theta_r \partial \theta_s}) + \frac{1}{2} \text{traza}(\mathbf{Q}_j \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_s} \mathbf{Q}_j \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_r}) \right. \\ &\quad \left. - \mathbf{Y}_j^\top \mathbf{Q}_j \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_s} \mathbf{Q}_j \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_r} \mathbf{Q}_j \mathbf{Y}_j + \frac{1}{2} \mathbf{Y}_j^\top \mathbf{Q}_j \frac{\partial^2 \boldsymbol{\Sigma}_j}{\partial \theta_s \partial \theta_r} \mathbf{Q}_j \mathbf{Y}_j \right]. \end{aligned}$$

Entonces, por la fórmula de la esperanza de la forma cuadrática (A1.5), la esperanza de la matriz Hessiana se expresa como

$$\begin{aligned} \mathbb{E}\left(\frac{d^2 \ell_{CL}(\mathbf{Y}; \boldsymbol{\theta})}{d\theta_r d\theta_s}\right) &= \sum_j \left[-\frac{1}{2} \text{traza}(\mathbf{Q}_j \frac{d^2 \boldsymbol{\Sigma}_j}{d\theta_r d\theta_s}) + \frac{1}{2} \text{traza}(\mathbf{Q}_j \frac{d\boldsymbol{\Sigma}_j}{d\theta_s} \mathbf{Q}_j \frac{d\boldsymbol{\Sigma}_j}{d\theta_r}) \right. \\ &\quad \left. - \text{traza}(\mathbf{Q}_j \frac{d\boldsymbol{\Sigma}_j}{d\theta_s} \mathbf{Q}_j \frac{d\boldsymbol{\Sigma}_j}{d\theta_r}) + \frac{1}{2} \text{traza}(\mathbf{Q}_j \frac{d^2 \boldsymbol{\Sigma}_j}{d\theta_s d\theta_r}) \right] \\ &= -\sum_j \frac{1}{2} \text{traza}(\mathbf{Q}_j \frac{d\boldsymbol{\Sigma}_j}{d\theta_s} \mathbf{Q}_j \frac{d\boldsymbol{\Sigma}_j}{d\theta_r}). \end{aligned}$$

Por tanto, los elementos de $\mathbf{H}(\boldsymbol{\theta})$ se escriben como

$$\mathbf{H}_{rs}(\boldsymbol{\theta}) = -\mathbb{E} \left(\frac{d^2 l_{CL}(\mathbf{Y}; \boldsymbol{\theta})}{d\theta_r d\theta_s} \right) = \sum_j \frac{1}{2} \text{traza} \left(\mathbf{Q}_j \frac{d\boldsymbol{\Sigma}_j}{d\theta_s} \mathbf{Q}_j \frac{d\boldsymbol{\Sigma}_j}{d\theta_r} \right), \quad r, s = 1, \dots, R. \quad (3.3)$$

Seguimos ahora, con el cálculo de la matriz $\mathbf{J}$, es decir la varianza de la Ecuación de score (3.2). Para esto se necesita utilizar propiedades de la varianza (A1.6) y covarianza (A1.7) de una forma cuadrática $\mathbf{Y}^\top \mathbf{B} \mathbf{Y}$.

Si definimos $\mathbf{B}_{jr} = \mathbf{Q}_j \frac{\partial \boldsymbol{\Sigma}_j}{\partial \theta_r} \mathbf{Q}_j$, tenemos que la varianza del score queda como

$$\begin{aligned} \mathbf{J} &= \text{Var} \left(\frac{dl_{CL}}{d\boldsymbol{\theta}} \right) = \text{Var} \left(\sum_j \frac{1}{2} \mathbf{Y}_j^\top \mathbf{B}_{jr} \mathbf{Y}_j \right) = \sum_j \sum_{j'} \frac{1}{4} \text{Cov} (\mathbf{Y}_j^\top \mathbf{B}_{jr} \mathbf{Y}_j, \mathbf{Y}_{j'}^\top \mathbf{B}_{j'r} \mathbf{Y}_{j'}) \\ &= \frac{1}{4} \sum_j \text{Var} (\mathbf{Y}_j^\top \mathbf{B}_{jr} \mathbf{Y}_j) + \frac{1}{2} \sum_j \sum_{j' > j} \text{Cov} (\mathbf{Y}_j^\top \mathbf{B}_{jr} \mathbf{Y}_j, \mathbf{Y}_{j'}^\top \mathbf{B}_{j'r} \mathbf{Y}_{j'}) \\ &= \sum_j \frac{1}{2} \text{traza} \left(\mathbf{Q}_j \frac{d\boldsymbol{\Sigma}_j}{d\theta_s} \mathbf{Q}_j \frac{d\boldsymbol{\Sigma}_j}{d\theta_r} \right) + \frac{1}{2} \sum_j \sum_{j' > j} \text{Cov} (\mathbf{Y}_j^\top \mathbf{B}_{jr} \mathbf{Y}_j, \mathbf{Y}_{j'}^\top \mathbf{B}_{j'r} \mathbf{Y}_{j'}). \end{aligned} \quad (3.4)$$

Por otra parte, para la estimación de $\hat{\boldsymbol{\beta}}_{CL}$, se necesita reintroducir las covariables al modelo simplemente sustituyendo $(\mathbf{Y}_j - \mathbf{X}_j \hat{\boldsymbol{\beta}}_{CL})$ por $\mathbf{Y}_j$ en las expresiones. Para $\boldsymbol{\theta}$ fijo, el parámetro de regresión $\hat{\boldsymbol{\beta}}_{CL}$ puede ser obtenido analíticamente, derivando la Ecuación (2.4) en términos de $\boldsymbol{\beta}$, igualando a 0 y despejando su valor. De esta forma, se obtiene que

$$\hat{\boldsymbol{\beta}}_{CL} = \mathbf{A}^{-1} \mathbf{b}, \quad (3.5)$$

donde

$$\mathbf{A} = \sum_j \mathbf{X}_j^\top \mathbf{Q}_j \mathbf{X}_j, \quad \mathbf{b} = \sum_j \mathbf{X}_j^\top \mathbf{Q}_j \mathbf{Y}_j. \quad (3.6)$$

La matriz de covarianza de la distribución normal límite de $\hat{\boldsymbol{\beta}}_{CL}$ también tiene forma de sándwich, y se calcula desde la matriz Hessiana $\mathbf{H}(\boldsymbol{\beta}) = \mathbf{A}$ y la varianza de score correspondiente:

$$\begin{aligned}
\mathbf{J} &= \text{var} \left(\frac{\partial \ell_{CL}(\mathbf{Y}; \boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \right) = \sum_j \sum_{j'} \text{cov} (\mathbf{X}_j^\top \mathbf{Q}_j \mathbf{Y}_j, \mathbf{X}_{j'}^\top \mathbf{Q}_{j'} \mathbf{Y}_{j'}) \\
&= \sum_j \sum_{j'} \mathbf{X}_j^\top \mathbf{Q}_j \text{cov} (\mathbf{Y}_j, \mathbf{Y}_{j'}) \mathbf{Q}_{j'} \mathbf{X}_{j'}. \quad (3.7)
\end{aligned}$$

En la práctica, solo se suman los términos con bordes $j = (k, l)$ y $j' = (k', l')$ que tienen nodos comunes entre los bloques (k, l, k', l') .

3.2. Cálculo de los estimadores de verosimilitud compuesta por bloques

La optimización de la Verosimilitud Compuesta por Bloques procede iterativamente, resolviendo alternativamente para $\hat{\boldsymbol{\beta}}_{CL}$ dado $\hat{\boldsymbol{\theta}}_{CL}$, y para $\hat{\boldsymbol{\theta}}_{CL}$ dado $\hat{\boldsymbol{\beta}}_{CL}$. Mientras que cada optimización con respecto a los parámetros de regresión $\boldsymbol{\beta}$ puede expresarse analíticamente (véase la Ecuación (3.5)), la optimización para $\boldsymbol{\theta}$ debe realizarse numéricamente. Definimos un valor inicial $\boldsymbol{\theta}(0)$ y utilizamos el Método de descenso de Newton-Raphson con actualizaciones de la forma Fisher-scoring, es decir, actualizaciones de la siguiente manera

$$\boldsymbol{\theta}(b+1) = \boldsymbol{\theta}(b) - \mathbb{E} \left[\frac{\partial^2 \ell_{CL}(\mathbf{Y}; \boldsymbol{\beta}, \boldsymbol{\theta}(b))}{\partial \boldsymbol{\theta}^2} \right]^{-1} \frac{\partial \ell_{CL}(\mathbf{Y}; \boldsymbol{\beta}, \boldsymbol{\theta}(b))}{\partial \boldsymbol{\theta}} \quad (3.8)$$

El score $\partial \ell_{CL}(\mathbf{Y}; \boldsymbol{\beta}, \boldsymbol{\theta}(b))/\partial \boldsymbol{\theta}$ y la esperanza de la segunda derivada de la Verosimilitud Compuesta por Bloques puede ser derivada analíticamente. La convergencia típicamente ocurre al rededor de la quinto actualización del Fisher-scoring.

El algoritmo eficiente para las actualizaciones de Fisher-scoring, que propone [Eidsvik et al. \(2014\)](#), está dado en el [Algoritmo 1](#).

El algoritmo recorre cada par de vecinos de cada uno de los bloques, calcula $\boldsymbol{\Sigma}$ y realiza la factorización de Cholesky para facilitar el calculo de su inversa $\mathbf{Q}$. Luego, calcula el factor $\mathbf{Q}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ que será utilizado para el cálculo del score. A continuación, para cada parámetro de $\boldsymbol{\theta}$ calcula el factor $\mathbf{Q} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_r}$ que también es útil en el cálculo

del score. Posteriormente, en la línea 7, calcula el score siguiendo la Ecuación (3.2) derivada en la Sección (3.1). Finalmente, calcula la esperanza de la matriz Hessiana siguiendo la expresión de la Ecuación (3.3) obtenida también en la Sección (3.1). Así, el algoritmo retorna el gradiente de la logverosimilitud compuesta por bloque, la esperanza de la Hessiana de la misma y con ellos el cálculo de la actualización del paso de Fisher-Scoring de la Ecuación (3.8).

Algoritmo 1 Cálculo para una actualización de Fisher-scoring para Verosimilitud Compuesta por Bloques

Require: $\boldsymbol{\theta} = (\theta_1, \dots, \theta_R)^\top$, inicializar $u_r = 0$ y $\mathbf{H}_{rs} = 0$, $r = 1, \dots, R$, $s = r, \dots, R$.

```

1: for  $k = 1$  to  $M - 1$  do
2:   for  $l \in N_k^-$  do
3:     Construir y factorizar  $\boldsymbol{\Sigma}_{kl} = \mathbf{L}_{kl}\mathbf{L}_{kl}^\top$ .  $\mathbf{Q}_{kl} = \mathbf{L}_{kl}^{-\top}\mathbf{L}_{kl}^{-1}$ 
4:     Calcular  $\mathbf{q}_{kl} = \mathbf{Q}_{kl}(\mathbf{Y}_{kl} - \mathbf{X}_{kl}\boldsymbol{\beta})$ 
5:     for  $r = 1$  to  $R$  do
6:       Calcular  $\mathbf{W}_{klr} = \mathbf{Q}_{kl} \frac{\partial \boldsymbol{\Sigma}_{kl}}{\partial \theta_r}$ 
7:        $u_r \leftarrow u_r - \frac{1}{2} \text{traza}(\mathbf{W}_{klr}) + \frac{1}{2} \mathbf{q}_{kl}^\top \frac{\partial \boldsymbol{\Sigma}_{kl}}{\partial \theta_r} \mathbf{q}_{kl}$ 
8:       for  $s = r$  to  $R$  do
9:          $\mathbf{H}_{rs} \leftarrow \mathbf{H}_{rs} + \frac{1}{2} \text{traza}(\mathbf{W}_{klr} \mathbf{W}_{kls})$ 
10:      end for
11:    end for
12:  end for
13: end for

14: return  $\frac{\partial \ell_{CL}}{\partial \boldsymbol{\theta}} = (u_1, \dots, u_R)^\top$ ,  $-\mathbb{E} \left( \frac{\partial^2 \ell_{CL}}{\partial \theta_r \partial \theta_s} \right) = \begin{pmatrix} H_{11} & \cdots & H_{1R} \\ \vdots & \ddots & \vdots \\ H_{1R} & \cdots & H_{RR} \end{pmatrix}$ ,

 $\boldsymbol{\theta} = \boldsymbol{\theta} + \mathbf{H}^{-1}u.$ 

```

Capítulo 4

EXPERIMENTOS NUMÉRICOS

El objetivo de este capítulo es comparar el desempeño de la verosimilitud compuesta propuesta por [Eidsvik et al. \(2014\)](#) (CL), la verosimilitud compuesta por bloques trabajada por [Acosta et al. \(2021\)](#) (BL) y la verosimilitud usual o completa (ML). Además de comparar los dos enfoques para la configuración de bloques, el enfoque por filas (RW) y el enfoque por columnas (CW).

Las rutinas para realizar estos experimentos fueron desarrolladas en lenguaje R version 4.4.1 y ejecutadas en dos entorno virtuales, el primero con 2 vCPUs Intel Xeon a 2.2 GHz y 13 GiB de RAM, y el segundo con un procesador Intel Xeon de 40 núcleos a 2.20GHz y 64GiB de RAM.

El código para implementar el método BL fue amablemente proporcionado por el profesor Jonathan Acosta de la Facultad de Matemática de la Pontificia Universidad Católica de Chile. Mientras que la rutina de propia autoría para implementar CL, estará disponible en <https://github.com/CloeRomero/Project-Composite-likelihood>.

4.1. Experimentos con el enfoque por filas (RW)

En esta primera parte, nos enfocaremos en analizar el enfoque por filas y nos referiremos a los métodos como ML, BL y CL. Se generaron 500 campos aleatorios Gaussianos independientes de media 0, con forma de grilla regular de tamaño $n = 16k \times 8k$ con $k = 1, 2, 3$ (tamaños escogidos para mantener el calculo del método ML factible), y aplicamos el modelo de covarianza de Matérn con parámetro de suavidad $\nu = 1/2$, que resulta en la estructura exponencial $h \rightarrow \sigma^2 \exp(-\frac{h}{\phi})$. Para este modelo consideramos los parámetros $\sigma^2 = 1$ y $\phi = \{3, 9, 16\}$, de esta forma, estimamos $\theta = (\sigma^2, \phi)$.

Exploramos tres configuraciones de bloques: 2 bloques de igual tamaño ($\Delta 2$), 4 bloques de igual tamaño ($\Delta 4$) y 8 bloques de igual tamaño ($\Delta 8$). Por ejemplo, para el caso de $n = 16 \times 8 = 128$, las configuraciones se verían como en la Figura 4.1.

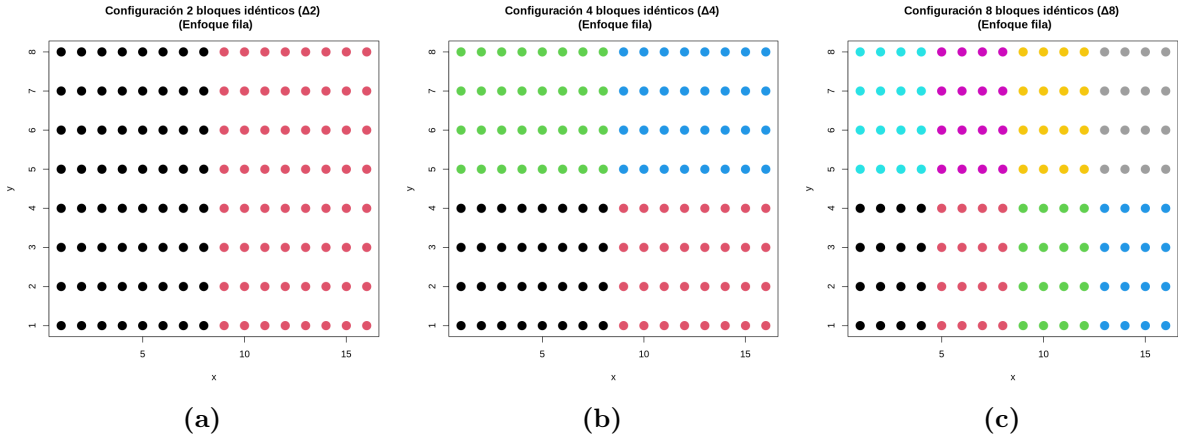


Figura 4.1: Configuraciones de bloques según el enfoque por filas para una grilla de tamaño $n = 16 \times 8 = 128$, con las configuraciones (a) $m = 2 \times 1 = 2$, $b = 8 \times 8 = 64$, (b) $m = 2 \times 2 = 4$, $b = 8 \times 4 = 32$ y (c) $m = 4 \times 2 = 8$, $b = 4 \times 4 = 16$.

Para el caso particular del $\theta = (1, 16)$, se consideró además con una grilla de $n = 16 \cdot 4 \times 8 \cdot 4 = 2048$ para observar el comportamiento en una grilla de mayor tamaño. Todas las estimaciones para $\theta = (1, 16)$ fueron realizadas en el segundo servidor.

Esto se muestra en la Figura 4.2, donde se evidencia que el método ML es el que más demora en su ejecución, seguido por el CL, siendo el BL el más rápido. Esto debido a que los métodos CL y BL, si bien invierten más matrices que el ML, son de menor tamaño lo que requiere menos esfuerzo computacional. El método CL toma más tiempo que el BL, porque requiere invertir una mayor cantidad de matrices de mayor tamaño para la misma configuración de bloques. Además, se observa que a medida que aumenta el número de bloques, los tiempos de ejecución de los métodos BL y CL disminuyen debido a que las dimensiones de las matrices se reducen.

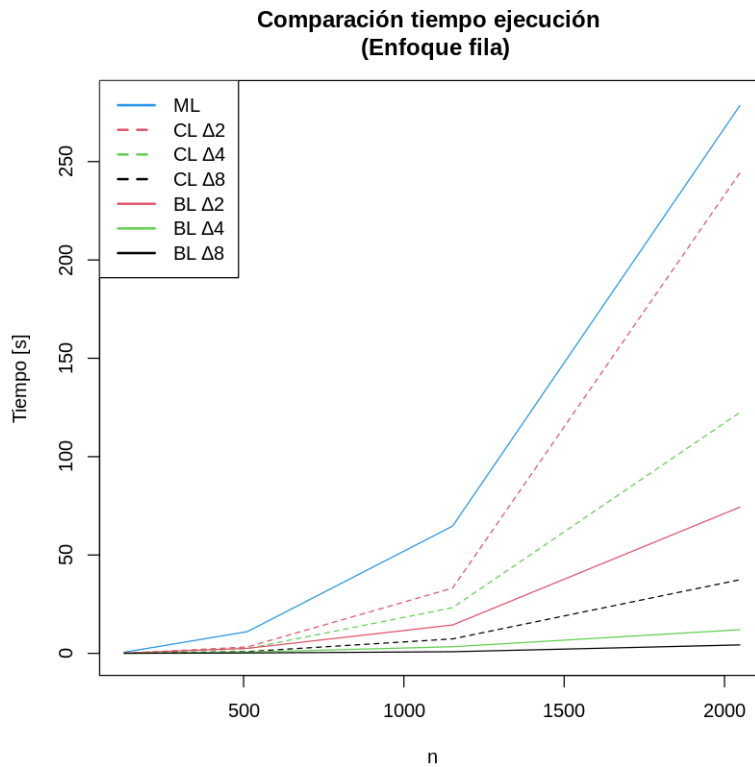


Figura 4.2: Tiempo de ejecución (en segundos) versus tamaño de muestra basado en 500 realizaciones independientes para los métodos ML, BL y CL, estimando $\theta = (1, 16)$. Los métodos BL y CL fueron implementados con tres diferentes configuraciones de bloques, según el enfoque por filas (RW): $\Delta 2$, $\Delta 4$ y $\Delta 8$.

Teniendo lo anterior en cuenta, notemos que, según lo que muestra la Figura 4.3, el método CL estima el parámetro ϕ con menor variabilidad que el método BL y muy cercana a la obtenida por el ML. Esto para todos los casos estudiados, tanto para los distintos tamaños de muestra, como distinta configuración de bloques y distintos ϕ . Esto ocasionado por el hecho de que el método CL utiliza para su cálculo más información que el BL. Al utilizar pares de bloques, invierte mayor cantidad de matrices y de mayor tamaño, por lo que demora más tiempo en ejecutarse, pero entrega mejores estimaciones. También notar que en el caso del BL, la variabilidad parece aumentar a medida que aumenta la cantidad de bloques, mientras que para el CL parece mantenerse más constante.

Lo anterior también es evidente al medir la eficiencia estadística mediante el Error Cuadrático Medio relativo (ECMr), que lo definimos como

$$\text{ECMr}_m(\theta_j) = \frac{\text{ECM}_{\text{ML}}(\theta_j)}{\text{ECM}_m(\theta_j)}, \quad j = 1, \dots, R, \quad m = CL, BL, BLc, CLc,$$

donde ECM_{ML} es el Error Cuadrático Medio de las estimaciones con el método ML, mientras que el ECM_m será el ECM de las estimaciones con los otros métodos. De esta forma, si observamos los ECMr en la Tabla 4.1, podemos notar que para $n = 1152$ (las tablas para $n = 128, 512$ se encuentran en el Anexo 3) el máximo rendimiento del método BL alcanza el 71,02 %, mientras que el del CL, cuando la configuración por bloques es $\Delta 2$, supera el 100 %, es decir, incluso genera mejores estimaciones que el ML, cuando utiliza toda la información disponible. Para las configuraciones $\Delta 4$ y $\Delta 8$, el rendimiento de CL no baja del 75 %.

De esta forma, obtenemos que ambos métodos, BL y CL, son una buena alternativa al ML, cuando se busca disminuir el tiempo de ejecución, sin embargo hay que tener en cuenta que el método más rápido no es el que entrega mejores estimaciones.

En el Anexo 2, se encuentran las Figuras A2.1 y A2.2, en donde se evidencian los mismos resultados para el tiempo de ejecución de los otros dos casos de θ y para la estimación de σ^2 .

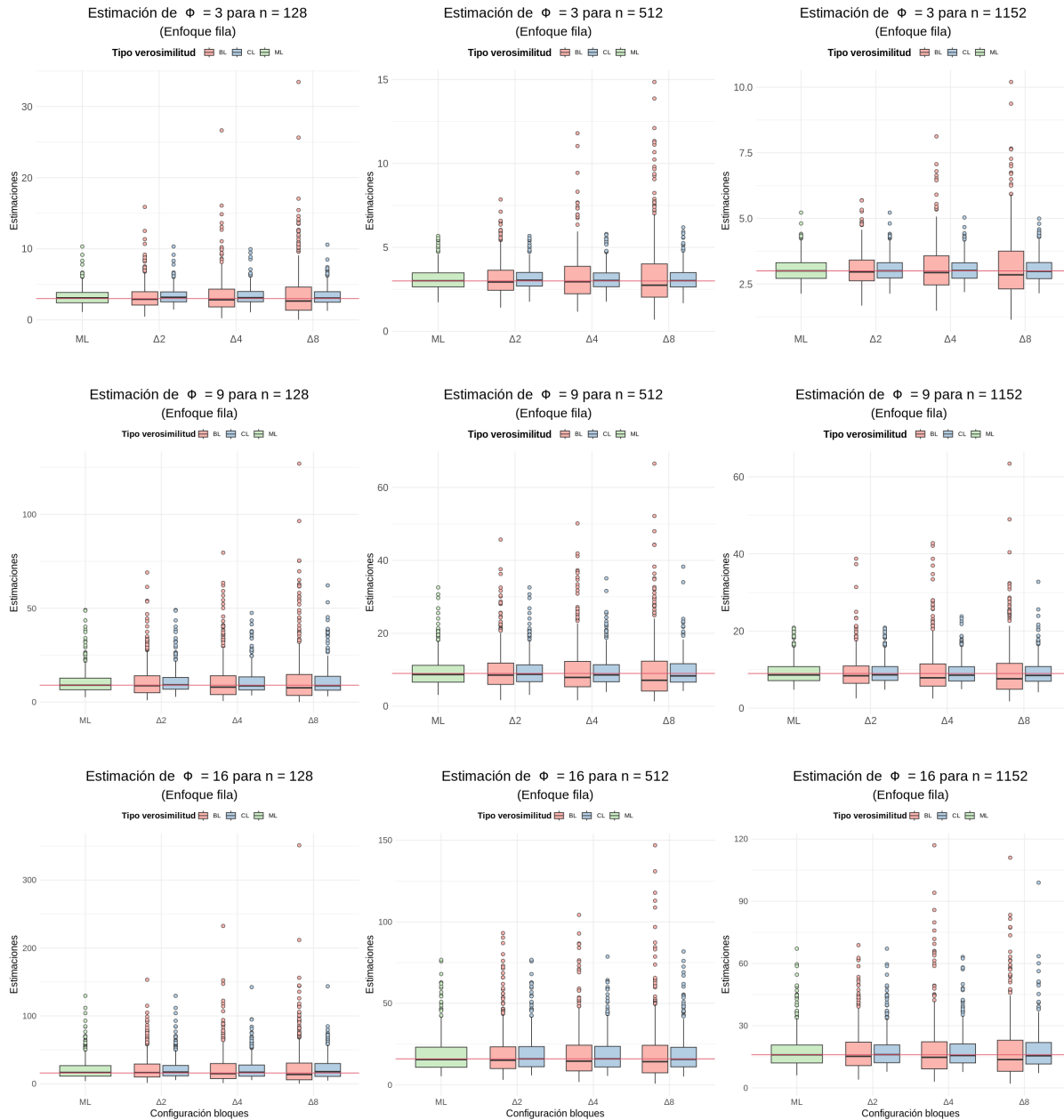


Figura 4.3: Estimación ML, BL y CL de ϕ basado en 500 realizaciones independientes con diferentes tamaños de muestra ($16 \times 8 = 128$, $32 \times 16 = 512$ y $48 \times 24 = 1152$) y tres diferentes ϕ (3, 9 y 16). Los métodos de verosimilitud compuesta y por bloques fueron implementados con tres diferentes configuraciones de bloques, según el enfoque por filas (RW): $\Delta 2$, $\Delta 4$ y $\Delta 8$.

$n = 1152$		$\theta = (1, 3)$		$\theta = (1, 9)$		$\theta = (1, 16)$	
	Config. bloques	σ^2	ϕ	σ^2	ϕ	σ^2	ϕ
ECMr _{CL}	$\Delta 2$	1,0069	1,0037	1,0046	1,0039	1,0002	1,0003
	$\Delta 4$	0,9983	0,9870	0,9452	0,9432	0,9838	0,9767
	$\Delta 8$	0,9547	0,9285	0,7704	0,7654	0,8977	0,8982
ECMr _{BL}	$\Delta 2$	0,5415	0,5315	0,4853	0,4786	0,7017	0,6850
	$\Delta 4$	0,2675	0,2664	0,2804	0,2719	0,4207	0,4059
	$\Delta 8$	0,1292	0,1293	0,2012	0,1960	0,3663	0,3618
ECMr _{CLc}	$\Delta 2$	1,0059	1,0019	1,0001	1,0001	0,9982	0,9974
	$\Delta 4$	1,0233	1,0291	1,0038	1,0059	1,0179	1,0225
	$\Delta 8$	0,9657	0,9562	0,9657	0,9562	0,9806	0,9806
ECMr _{BLc}	$\Delta 2$	0,9742	0,9346	0,9883	0,9774	0,9515	0,9554
	$\Delta 4$	0,8037	0,7115	0,8749	0,8447	0,9216	0,8800
	$\Delta 8$	0,5156	0,3802	0,7636	0,6477	0,8054	0,7324

Tabla 4.1: Errores Cuadráticos Medios relativos para las estimaciones de σ^2 y ϕ , cuando $n = 1152$, de los modelos BL y CL con los enfoques de configuración de bloques por fila y por columna (BLc y CLc).

4.2. Experimentos con el enfoque por columnas (CW)

En este caso, llamaremos a los métodos BLc y CLc, pues se emplearán con la configuración de bloques por columna, se seguirá usando el mismo modelo de covarianza y parámetros. De igual manera, se explorarán tres configuraciones de bloques: 2 bloques de igual tamaño ($\Delta 2$), 4 bloques de igual tamaño ($\Delta 4$) y 8 bloques de igual tamaño ($\Delta 8$). Como ejemplo, para el caso $n = 16 \times 8 = 128$, las configuraciones se verían como muestra la Figura 4.4.

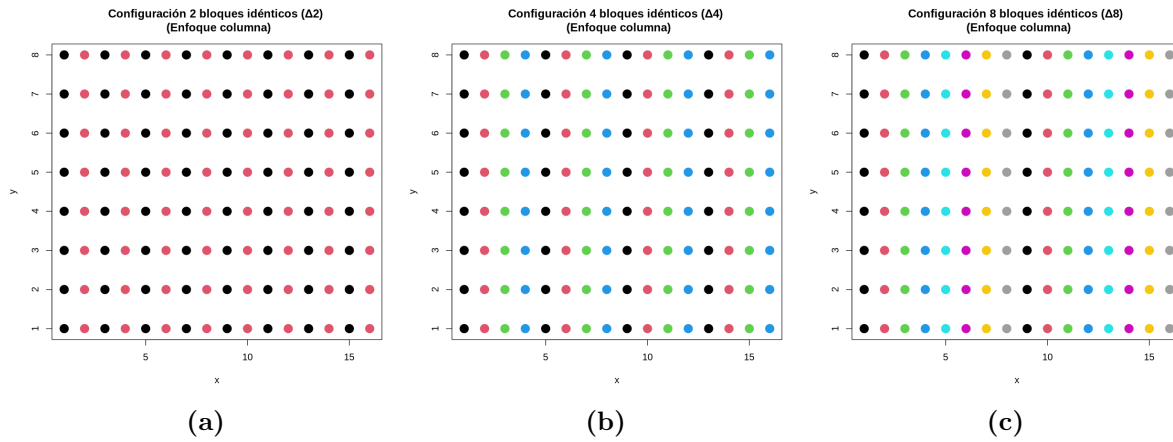


Figura 4.4: Configuraciones de bloques según el enfoque por filas para una grilla de tamaño $n = 16 \times 8 = 128$, con las configuraciones (a) $m = 2 \times 1 = 2$, $b = 8 \times 8 = 64$, (b) $m = 4 \times 1 = 4$, $b = 4 \times 8 = 32$ y (c) $m = 8 \times 1 = 8$, $b = 2 \times 8 = 16$.

Si observamos la Figura 4.5, podemos notar que en los casos en que ϕ toma el valor de 3 y 9 los tiempos de ejecución para las configuraciones $\Delta 4$ y $\Delta 8$ son similares, mientras que para $\Delta 2$, CLc es más rápido. Por otro lado, para $\phi = 16$, en todos los casos CLc demoró más tiempo que BLc.

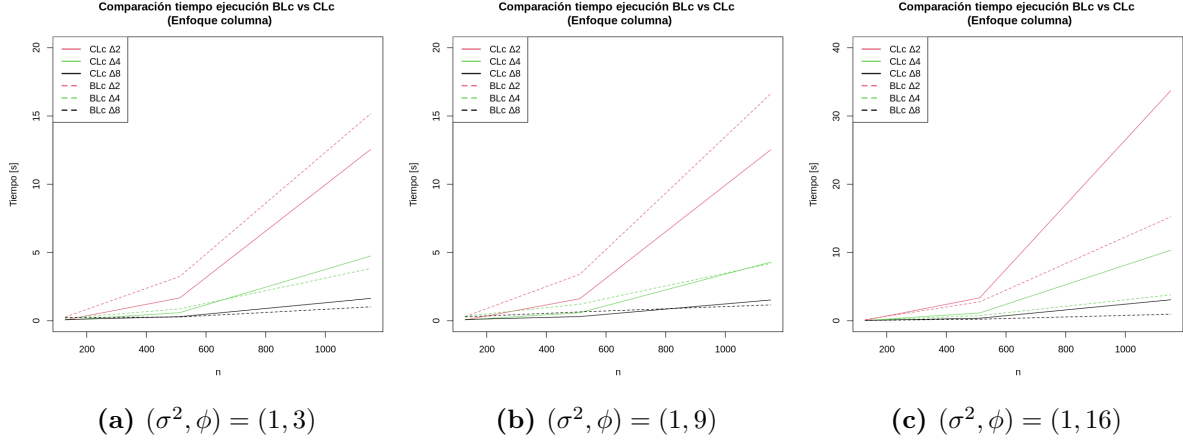


Figura 4.5: Tiempo de ejecución (en segundos) versus tamaño de muestra basado en 500 realizaciones independientes para tres diferentes ϕ (3, 9 y 16). Los métodos BLc y CLc fueron implementados con tres diferentes configuraciones de bloques, según el enfoque por columnas (CW): $\Delta 2$, $\Delta 4$ y $\Delta 8$.

En cuanto a la estimación de los parámetros, en la Figura 4.6 notamos que nuevamente que el método de verosimilitud compuesta, CLc en este caso, presenta estimaciones de ϕ con menor variabilidad que BLc. Sin embargo, esta reducción en la variabilidad parece ser menos pronunciada en comparación con el enfoque por filas, es más, cuando ϕ toma valores 9 y 16, con $n = 1152$, no se aprecia diferencia considerable. Si volvemos a revisar la Tabla 4.1, podemos observar que el desempeño del CLc en todos los casos, para $n = 1152$, alcanza el 95 % y en algunos, incluso supera el 100 %. Mientras, que el método BLc, en el mejor de los casos alcanza un 98,83 %, y en el peor un 38,02 % de rendimiento.

Los gráficos de la estimación de σ^2 , entregan información similar y se encuentran en la Figura A2.3 en el Anexo 2.

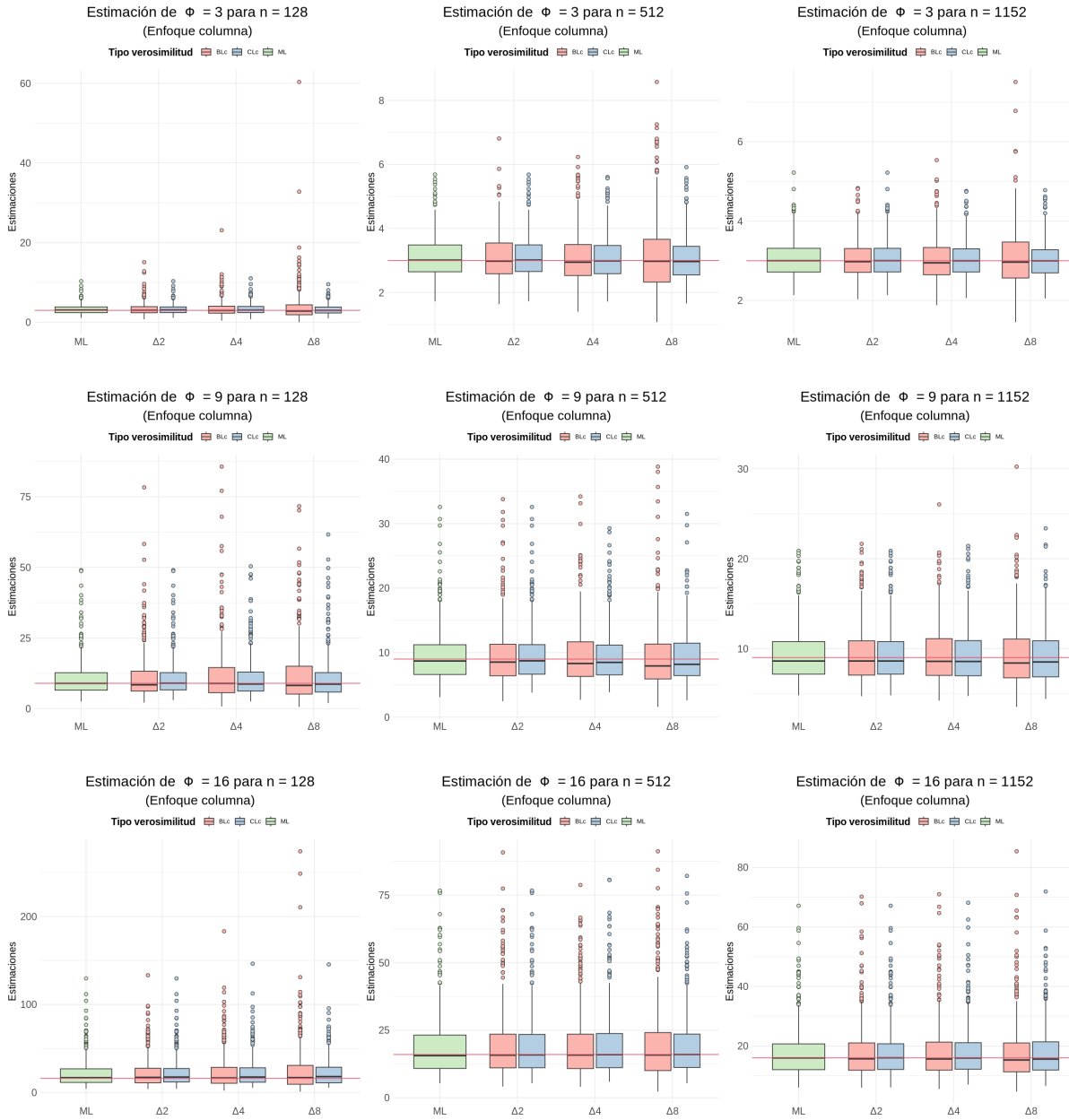


Figura 4.6: Estimación BLc y CLc de ϕ basado en 500 realizaciones independientes con diferentes tamaños de muestra ($16 \times 8 = 128$, $32 \times 16 = 512$ y $48 \times 24 = 1152$) y tres diferentes ϕ (3, 9 y 16). Los métodos fueron implementados con tres diferentes configuraciones de bloques, según el enfoque por columnas (CW): $\Delta 2$, $\Delta 4$ y $\Delta 8$.

4.3. Comparación de enfoques

A partir de los experimentos anteriores, podemos comparar los resultados para evaluar la incidencia de utilizar estos distintos enfoques en la estimación de los parámetros.

De la Figura 4.7, podemos concluir que el uso del enfoque por filas o por columnas en el método BL, no parece afectar en significativamente al tiempo de ejecución, esto porque el método invierte igual número de matrices y de igual tamaño, para ambos casos.

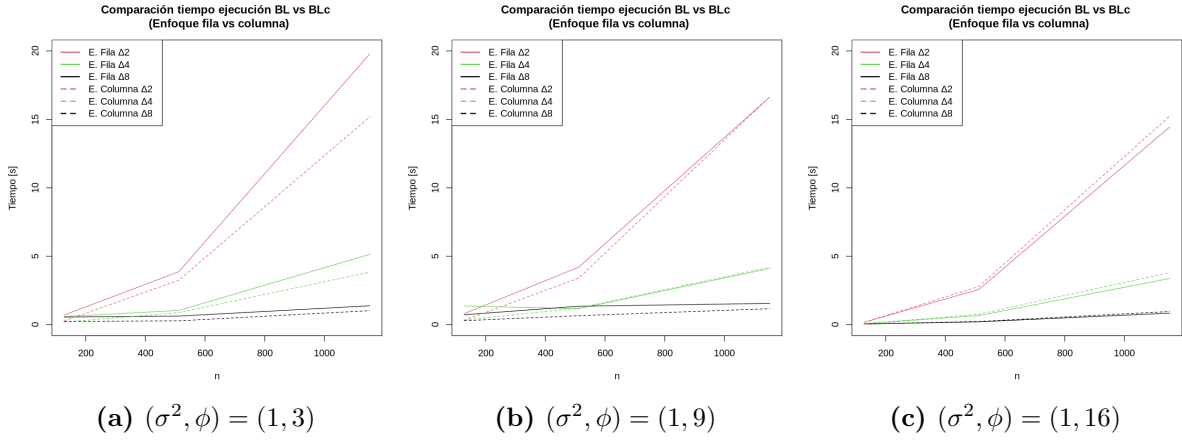


Figura 4.7: Tiempo de ejecución (en segundos) versus tamaño de muestra basado en 500 realizaciones independientes. Los métodos de BL y BLc fueron implementados con tres diferentes configuraciones de bloques, según los enfoque por filas y columnas: 2 bloques idénticos ($\Delta 2$), 4 bloques idénticos ($\Delta 4$) y 8 bloques idénticos ($\Delta 8$).

Por otro lado, para el caso de la verosimilitud compuesta (CL), sí se evidencia una reducción en el tiempo de ejecución cuando se utiliza el enfoque por columnas, tal como muestra la Figura 4.8. Esto debido a que, en el enfoque por columnas, cada bloque tiene sólo un vecino, por lo que se invierten menos matrices en comparación con el enfoque por filas, excepto en el caso $\Delta 2$, en donde invierten el mismo número de matrices.

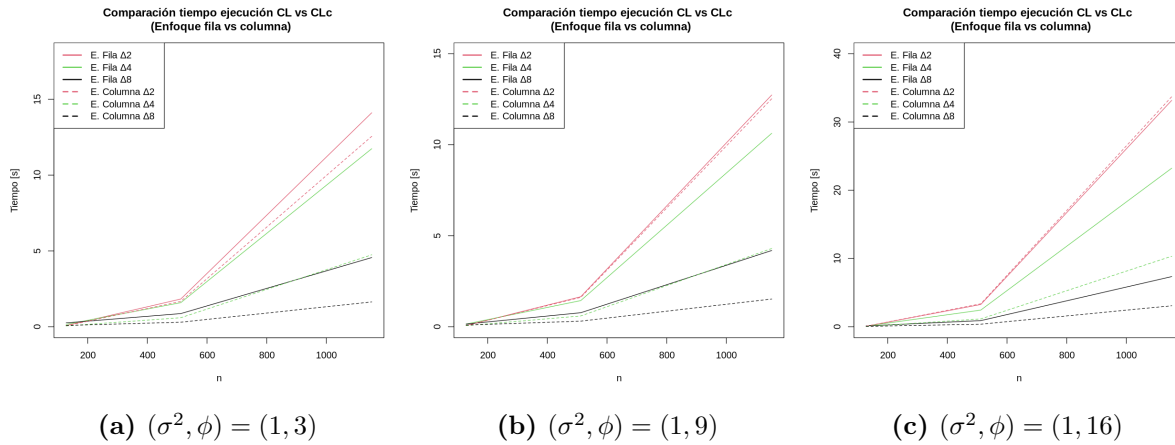


Figura 4.8: Tiempo de ejecución (en segundos) versus tamaño de muestra basado en 500 realizaciones independientes. Los métodos de CL y CLc fueron implementados con tres diferentes configuraciones de bloques, según los enfoque por filas y columnas: 2 bloques idénticos ($\Delta 2$), 4 bloques idénticos ($\Delta 4$) y 8 bloques idénticos ($\Delta 8$).

Al analizar las estimaciones de ϕ para los caso BL y BLc (Figura 4.9), es claro que la variabilidad es menor con el enfoque por columnas en cada uno de los casos, para cada ϕ y cada tamaño de muestra. Esto indica que este método es sensible al enfoque que se elija para configurar los bloques. La mejora de las estimaciones para este método también se evidencian en la Tabla 4.1, en donde se aprecia que el máximo rendimiento del BL fue un 70,17 %, mientras que el del BLc fue un 98,83 %, y el mínimo fue de un 12,92 % y 38,02 %, respectivamente, por lo que se evidencia una mejora sustancial en el rendimiento del BLc contra el BL.

En contraste, el método CL parece no ser tan sensible, pues la Figura 4.10 nos muestra que la variabilidad de las estimaciones no difiere significativamente entre los enfoques. Y la Tabla 4.1 lo confirma, ya que el rendimiento de CLc mejora, pero no tan drásticamente como lo hace el BLc.

Entonces, podemos deducir que el enfoque por columna es una mejor opción para estimar los parámetros con estos métodos, ya que como para el método CL no hay mucha diferencia en cuanto a la calidad de estimación, pero sí la hay en cuanto a tiempo de ejecución, es preferible el enfoque que demore menor en ejecutarse, es

decir, el enfoque por columnas. Por otro lado, para el método BL, aunque no hay una diferencia significativa en el tiempo de ejecución entre los enfoques, el enfoque por columnas proporciona estimaciones de mejor calidad, lo que lo hace preferible al enfoque por filas.

Los gráficos para la comparación de la estimación de σ^2 se encuentran en el [Anexo 2](#).

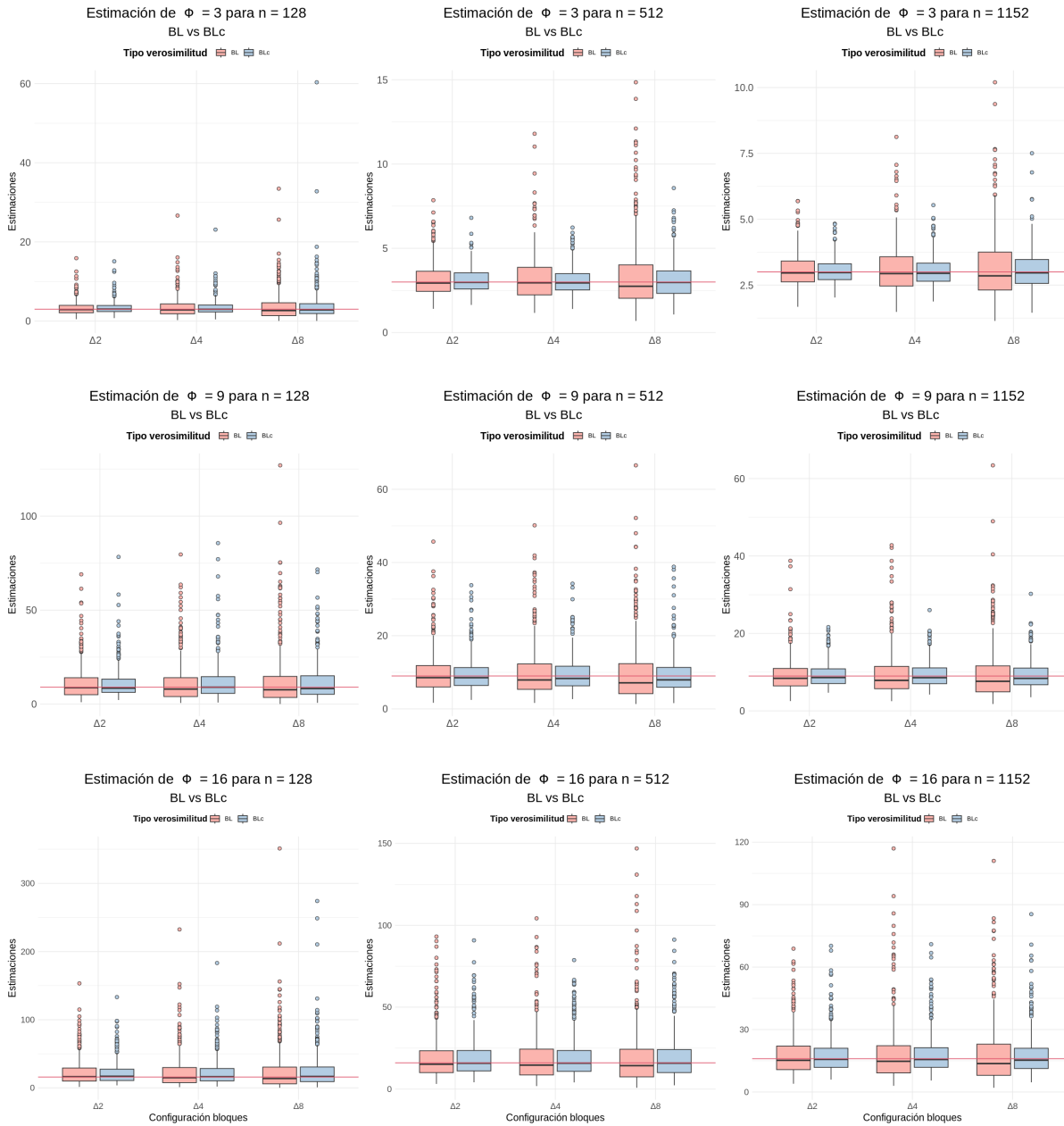


Figura 4.9: Estimación BL y BLc de ϕ basado en 500 realizaciones independientes con diferentes tamaños de muestra ($16 \times 8 = 128$, $32 \times 16 = 512$ y $48 \times 24 = 1152$) y tres diferentes ϕ (3, 9 y 16). Los métodos fueron implementados con tres diferentes configuraciones de bloques, según los enfoque por filas y columnas: 2 bloques idénticos ($\Delta 2$), 4 bloques idénticos ($\Delta 4$) y 8 bloques idénticos ($\Delta 8$).

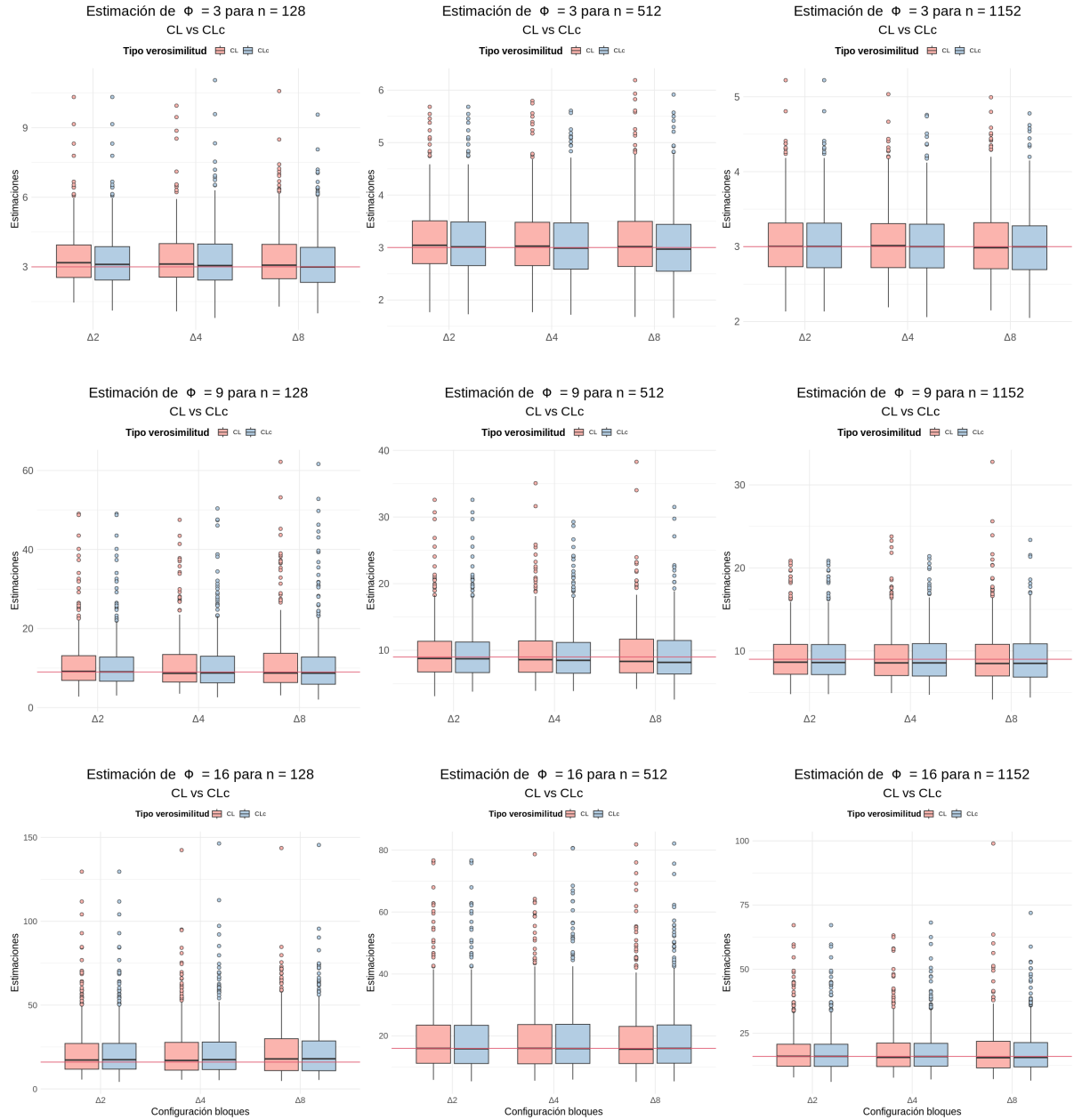


Figura 4.10: Estimación CL y CLc de ϕ basado en 500 realizaciones independientes con diferentes tamaños de muestra ($16 \times 8 = 128$, $32 \times 16 = 512$ y $48 \times 24 = 1152$) y tres diferentes ϕ (3, 9 y 16). Los métodos fueron implementados con tres diferentes configuraciones de bloques, según los enfoque por filas y columnas: 2 bloques idénticos ($\Delta 2$), 4 bloques idénticos ($\Delta 4$) y 8 bloques idénticos ($\Delta 8$).

COMENTARIOS FINALES Y TRABAJOS A FUTURO

Comparamos dos alternativas al método de máxima verosimilitud para datos espaciales y nuestro estudio arrojó que la mejor alternativa dependerá de qué ventaja se valore más, si la mejor calidad de la estimación o el menor tiempo de ejecución. La elección sería el método CLc o BLc, respectivamente. Además, confirmamos que la configuración de bloques por columnas beneficia las estimaciones de las verosimilitudes compuestas aquí estudiadas, ya sea en términos de tiempo de ejecución (CL), como en término de calidad de la estimación (BL).

Además, pudimos observar que la gran ventaja en tiempo de ejecución que presenta el método trabajado por [Eidsvik et al. \(2014\)](#) sí se debe principalmente a los avances tecnológicos de los que se apoya, ya que en un entorno menos especializado, su rendimiento, en cuanto a tiempo de ejecución, no superó al de la verosimilitud por bloques que asume la independencia entre todos los bloques.

Temas que quedaron inconclusos, que no se alcanzaron a tratar en este estudio son, primeramente, la prueba de estos métodos y enfoques a datos reales, no sólo simulaciones. También analizar otras estrategias para la selección de bloques, sobre todo en datos que no se encuentran en grillas regulares, que tampoco fueron estudiadas,

por ejemplo, teselaciones de Voronoi/Delaunay, como menciona [Eidsvik et al. \(2014\)](#). Buscar y probar configuraciones que no sólo mejore la calidad de la estimación si no también su eficiencia computacional.

También quedó por estudiar la predicción de observaciones en sitios no observados, probar si existe algún beneficio en la utilización del enfoque por columna para ese caso, aunque como menciona [Mukerjee \(2023\)](#), cuando tratamos con predicciones suele utilizarse información local, cercana al lugar no observado, por lo que quizás usar este enfoque no sea tan beneficioso, como sí lo fue para la estimación, ya que todas las observaciones, en general se rigen por los mismos parámetros de forma global.

Finalmente, nosotros estudiamos campos aleatorios Gaussianos con métodos diseñados para estos campos, interesante podría ser estudiar cómo se desempeñan estos métodos aplicados a otros tipos de datos, como del tipo t-Student ([Bevilacqua et al., 2020a](#)), Weibull ([Bevilacqua et al., 2020b](#)) o Poisson ([Morales-Navarrete et al., 2021](#)). Un estudio referente a esto podría ser el de [Caamaño et al. \(2022\)](#), donde estudia la verosimilitud compuesta con pesos basada en pares para campos aleatorios no-Gaussianos con una aplicación a campos aleatorios Tukey-hh.

ANEXO 1: PRELIMINARES

PROPIEDADES MATRICIALES

En esta sección, se expondrán las propiedades que resultarán de gran utilidad en los cálculos que se llevarán a cabo en las siguientes etapas de este proyecto.

Sean $\mathbf{A}, \mathbf{B}, \mathbf{X}(x)$ matrices $p \times p$, con $\mathbf{X}(x)$ siendo una matriz en función del parámetro θ , e $\mathbf{Y}$ un vector $p \times 1$ que no depende de θ , se tienen las siguientes propiedades:

- Derivadas matriciales:

- $$\frac{d(\mathbf{X}^{-1})}{d\theta} = -\mathbf{X}^{-1} \left(\frac{d\mathbf{X}}{d\theta} \right) \mathbf{X}^{-1}, \quad (\text{A1.1})$$

- $$\frac{d(\mathbf{Y}^t \mathbf{X} \mathbf{Y})}{d\theta} = \mathbf{Y}^t \frac{d\mathbf{X}}{d\theta} \mathbf{Y}, \quad (\text{A1.2})$$

- $$\frac{d \log |\mathbf{X}|}{d\theta} = \text{traza} \left(\mathbf{X}^{-1} \frac{d\mathbf{X}}{d\theta} \right), \quad (\text{A1.3})$$

- $$d \text{traza}(\mathbf{A} \mathbf{B}) = \text{traza}(d\mathbf{A} \mathbf{B}) + \text{traza}(\mathbf{A} d\mathbf{B}). \quad (\text{A1.4})$$

Sea ahora $\mathbf{Y}$ una colección de datos tal que $\mathbf{Y} \sim N_p(0, \mathbf{\Sigma})$, con $\mathbf{\Sigma}$ matriz de covarianza, se pueden observar las siguientes propiedades:

- Esperanza de una forma cuadrática:

•

$$\mathbb{E}(\mathbf{Y}^t \mathbf{B} \mathbf{Y}) = \text{traza}(\mathbf{B} \boldsymbol{\Sigma}). \quad (\text{A1.5})$$

■ Varianza y covarianza de una forma cuadrática:

•

$$\text{Var}(\mathbf{Y}^t \mathbf{B} \mathbf{Y}) = 2 \cdot \text{traza}(\mathbf{B} \boldsymbol{\Sigma} \mathbf{B} \boldsymbol{\Sigma}), \quad (\text{A1.6})$$

•

$$\text{Cov}(\mathbf{Y}^t \mathbf{A} \mathbf{Y}, \mathbf{Y}^t \mathbf{B} \mathbf{Y}) = 2 \cdot \text{traza}(\mathbf{A} \boldsymbol{\Sigma} \mathbf{B} \boldsymbol{\Sigma}). \quad (\text{A1.7})$$

ANEXO 2: GRÁFICOS

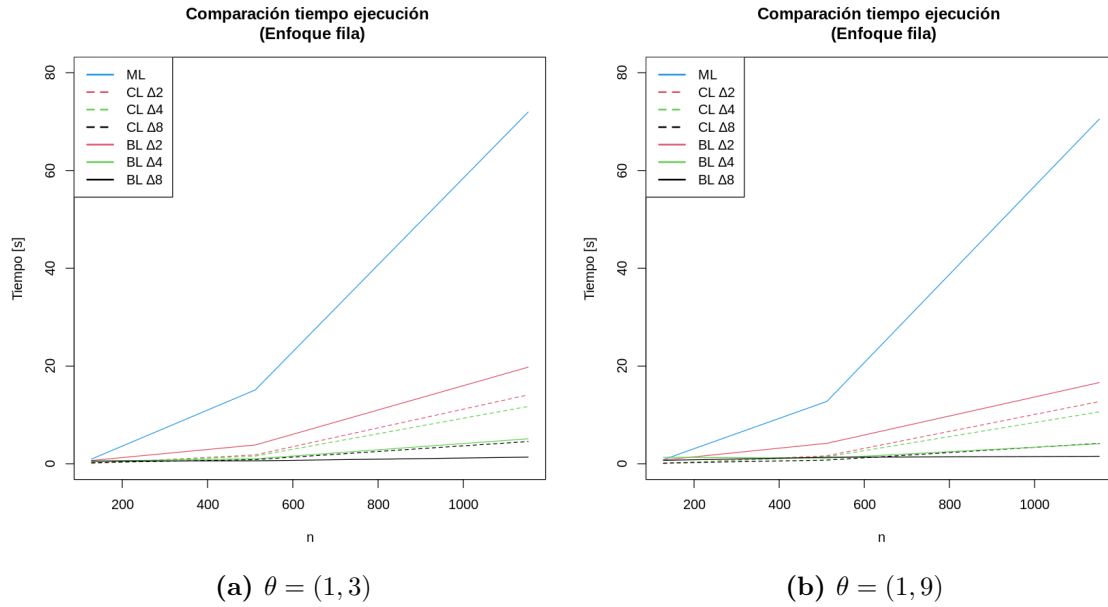


Figura A2.1: Tiempo de ejecución (en segundos) versus tamaño de muestra basado en 500 realizaciones independientes para los métodos ML, BL y CL, estimando (a) $\theta = (1, 3)$ y (b) $\theta = (1, 9)$. Los métodos BL y CL fueron implementados con tres diferentes configuraciones de bloques, según el enfoque por filas (RW): $\Delta 2$, $\Delta 4$ y $\Delta 8$.

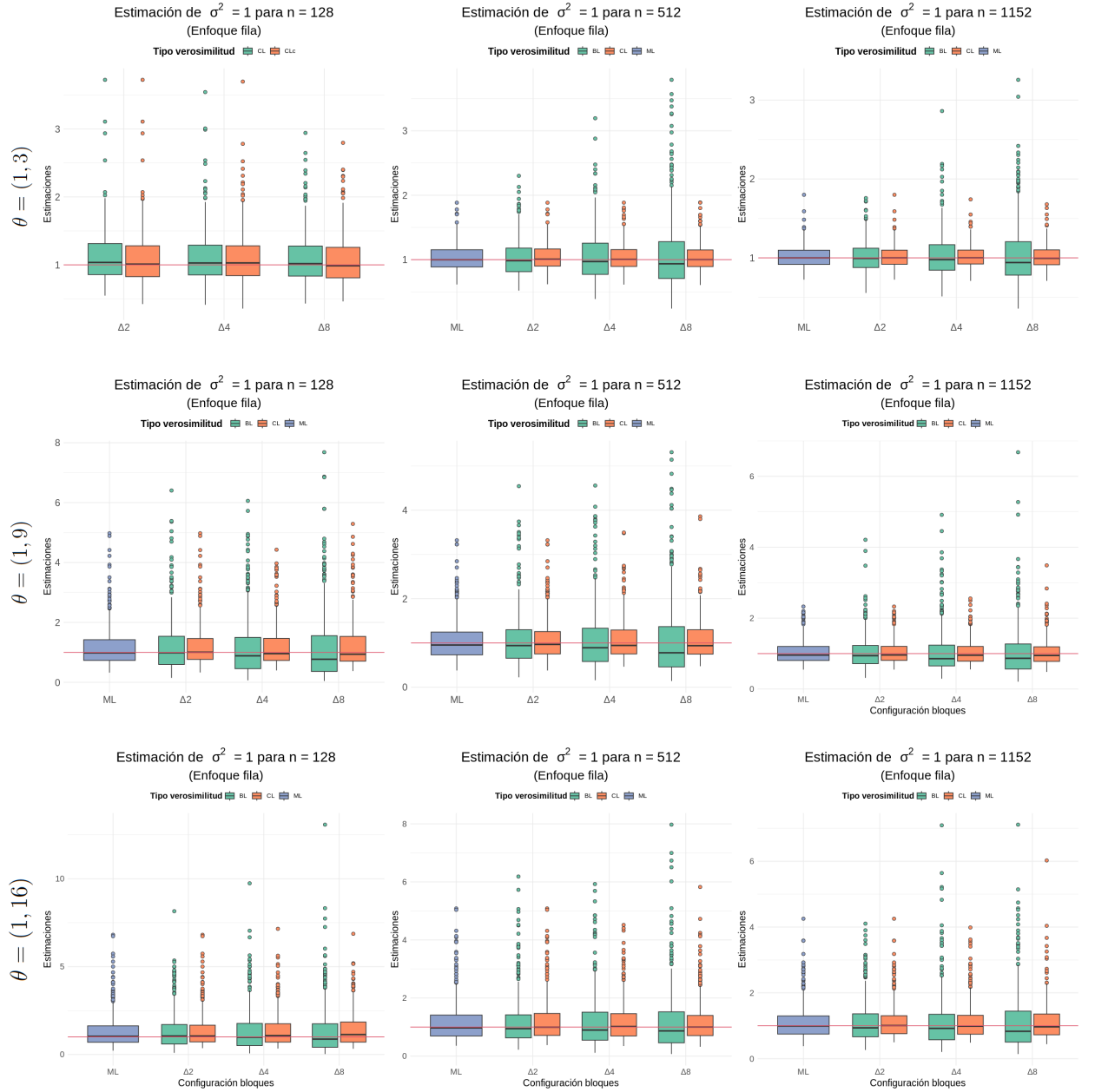


Figura A2.2: Estimación ML, BL y CL de σ^2 basado en 500 realizaciones independientes con diferentes tamaños de muestra ($16 \times 8 = 128$, $32 \times 16 = 512$ y $48 \times 24 = 1152$). Los métodos de verosimilitud compuesta y por bloques fueron implementados con tres diferentes configuraciones de bloques, según el enfoque por filas (RW): $\Delta 2$, $\Delta 4$ y $\Delta 8$.

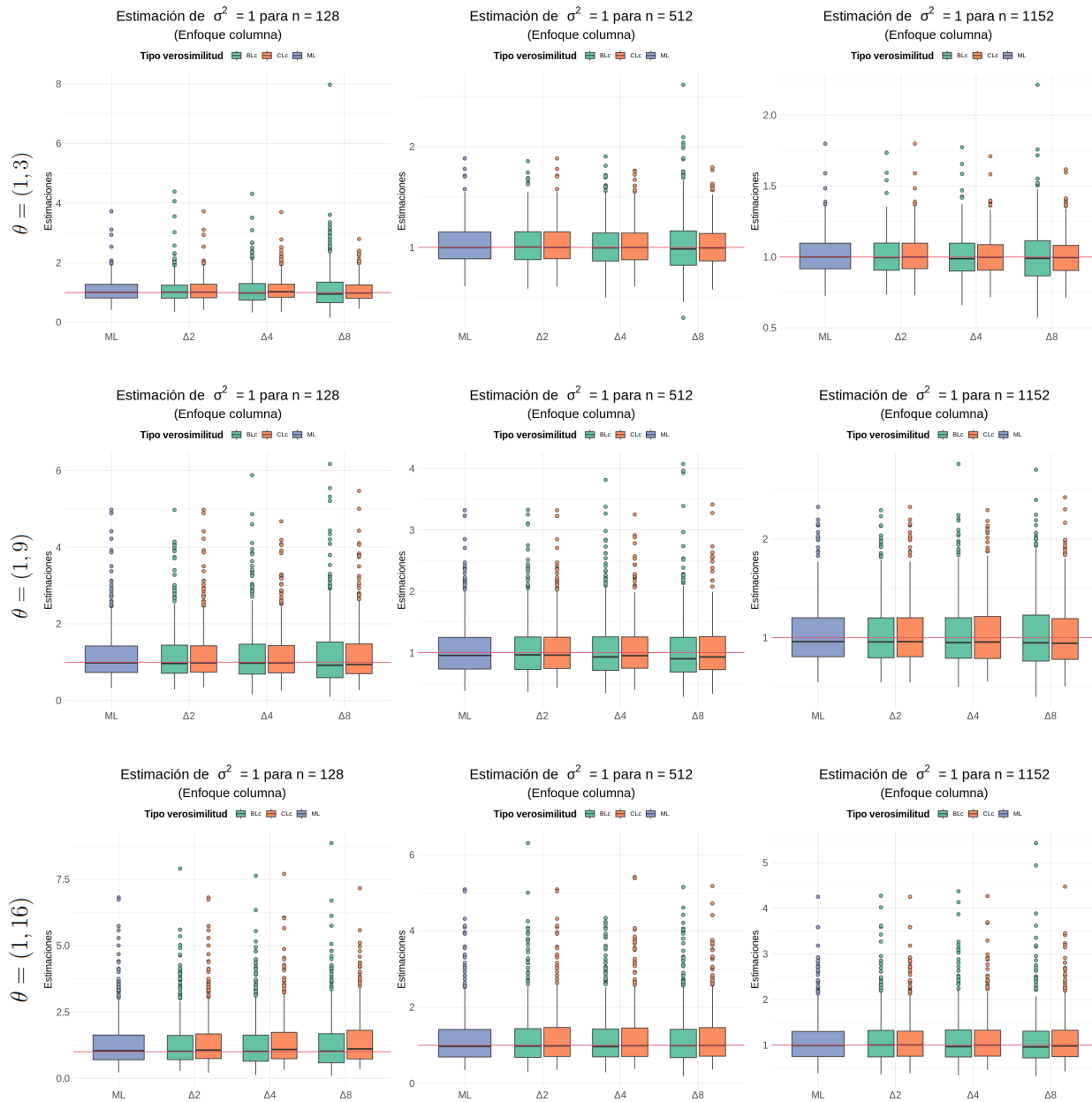


Figura A2.3: Estimación BLc y CLc de σ^2 basado en 500 realizaciones independientes con diferentes tamaños de muestra ($16 \times 8 = 128$, $32 \times 16 = 512$ y $48 \times 24 = 1152$). Los métodos fueron implementados con tres diferentes configuraciones de bloques, según el enfoque por columnas (CW): $\Delta 2$, $\Delta 4$ y $\Delta 8$.

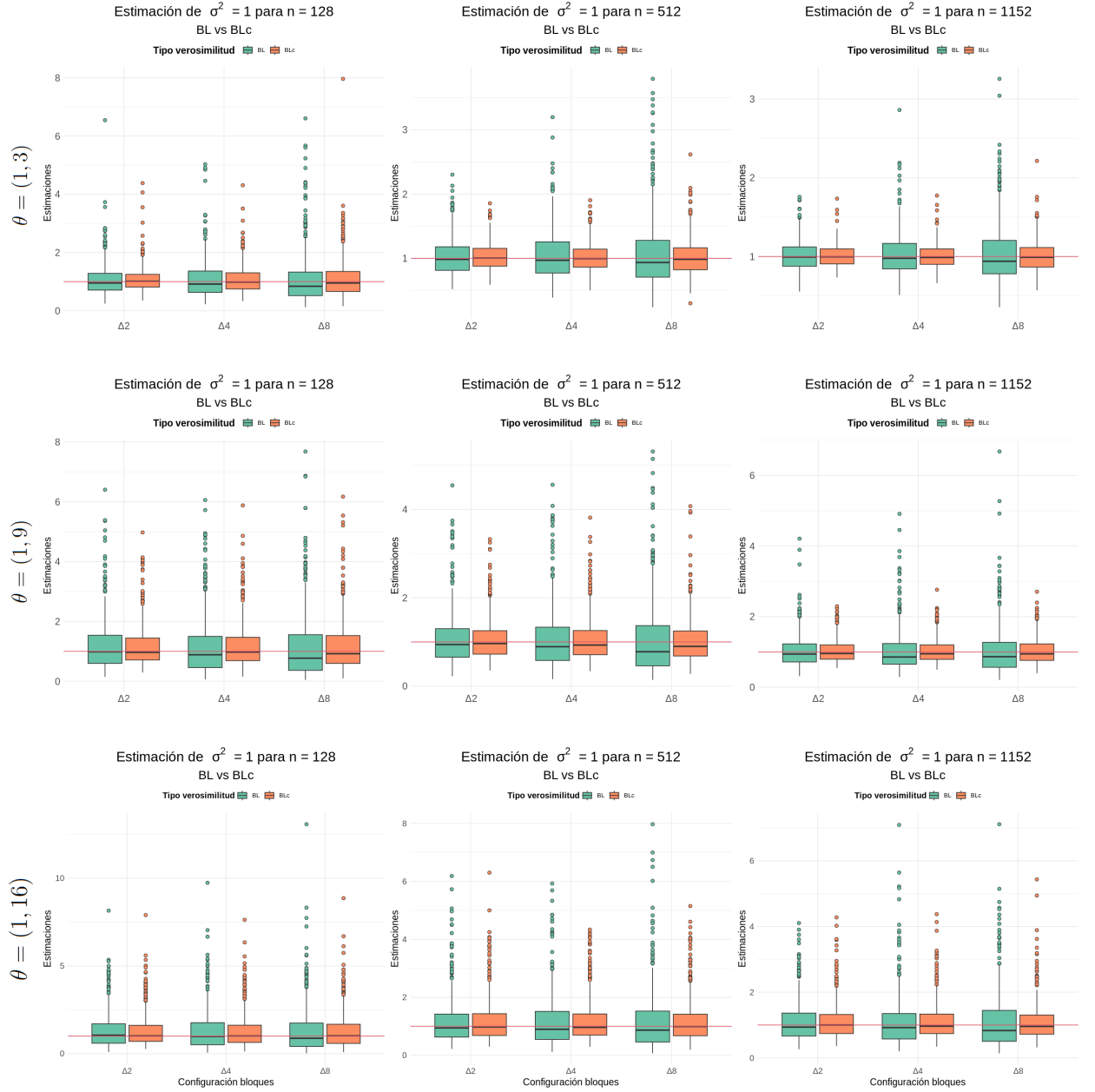


Figura A2.4: Estimación BL y BLc de σ^2 basado en 500 realizaciones independientes con diferentes tamaños de muestra ($16 \times 8 = 128$, $32 \times 16 = 512$ y $48 \times 24 = 1152$). Los métodos fueron implementados con tres diferentes configuraciones de bloques, según los enfoque por filas y columnas: 2 bloques idénticos ($\Delta 2$), 4 bloques idénticos ($\Delta 4$) y 8 bloques idénticos ($\Delta 8$).

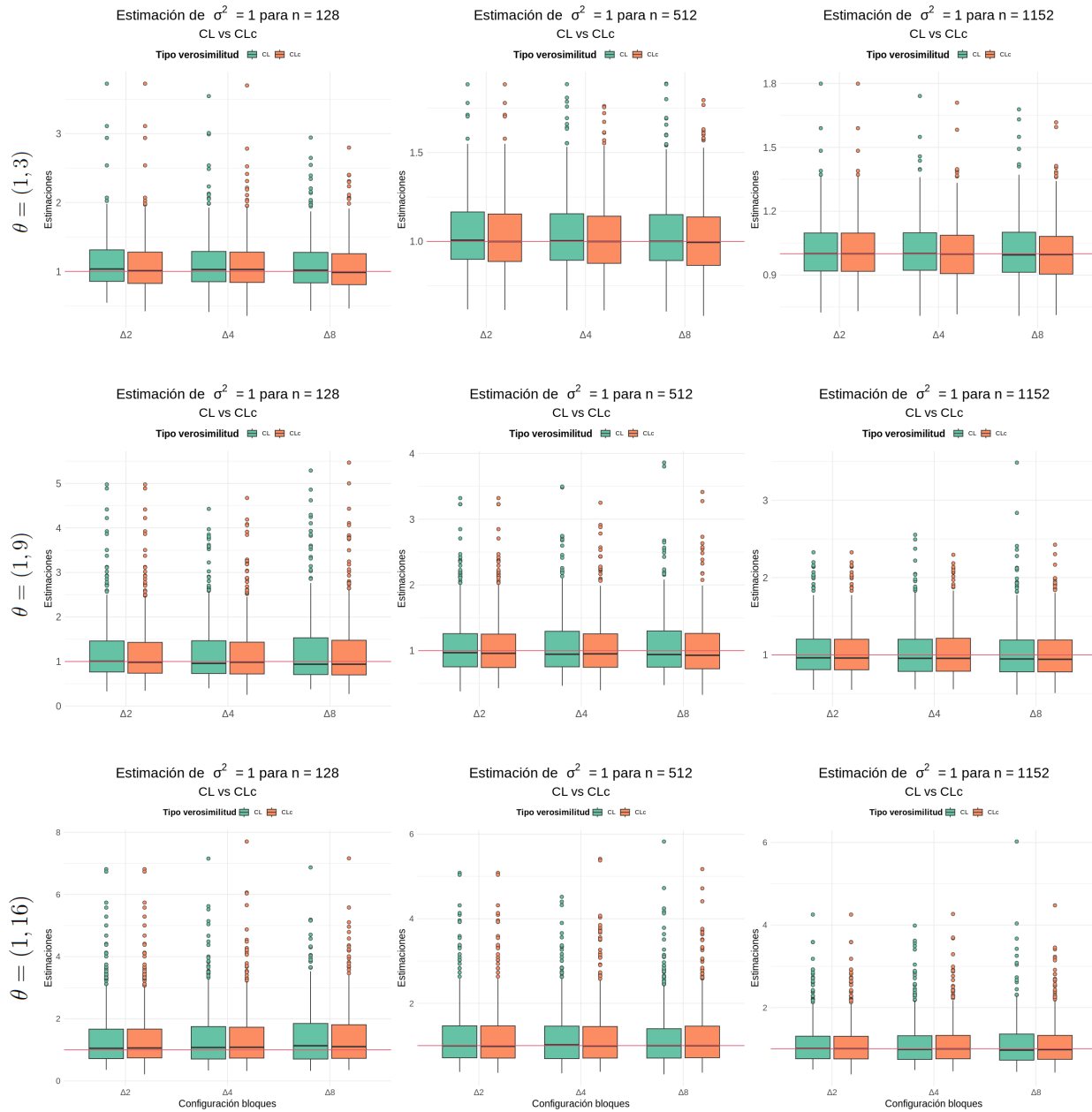


Figura A2.5: Estimación CL y CLc de σ^2 basado en 500 realizaciones independientes con diferentes tamaños de muestra ($16 \times 8 = 128$, $32 \times 16 = 512$ y $48 \times 24 = 1152$). Los métodos fueron implementados con tres diferentes configuraciones de bloques, según los enfoque por filas y columnas: 2 bloques idénticos ($\Delta 2$), 4 bloques idénticos ($\Delta 4$) y 8 bloques idénticos ($\Delta 8$).

ANEXO 3: TABLAS

$n = 128$	Config. bloques	$\theta = (1, 3)$		$\theta = (1, 9)$		$\theta = (1, 16)$	
		σ^2	ϕ	σ^2	ϕ	σ^2	ϕ
ECMr _{CL}	$\Delta 2$	0,9757	0,9856	0,9753	0,9748	1,0000	1,0000
	$\Delta 4$	0,9458	0,9516	1,0370	1,0121	1,0336	1,0018
	$\Delta 8$	0,9892	0,9212	1,0370	1,0121	0,9506	0,9093
ECMr _{BL}	$\Delta 2$	0,4305	0,4253	0,6030	0,5334	0,8467	0,7325
	$\Delta 4$	0,3206	0,2350	0,4872	0,3759	0,7300	0,5011
	$\Delta 8$	0,1796	1,2536	0,3972	0,2211	0,5412	0,2807
ECMr _{CLc}	$\Delta 2$	0,9971	1,0007	0,9964	0,9961	0,9640	0,9645
	$\Delta 4$	1,0016	0,9275	0,9868	0,9262	0,9631	0,9221
	$\Delta 8$	1,1021	1,0148	0,8446	0,7489	0,9710	0,9027
ECMr _{BLc}	$\Delta 2$	0,7493	0,5969	0,9184	0,6690	1,0327	0,9452
	$\Delta 4$	0,6400	0,3816	0,8049	0,4676	0,9586	0,6957
	$\Delta 8$	0,3118	0,0893	0,6229	0,4007	0,8432	0,3609

Tabla A3.1: Errores Cuadráticos Medios relativos para las estimaciones de σ^2 y ϕ , cuando $n = 128$, de los modelos BL y CL con los enfoques de configuración de bloques por fila y por columna (BLc y CLc).

$n = 512$		$\theta = (1, 3)$		$\theta = (1, 9)$		$\theta = (1, 16)$	
	Config. bloques	σ^2	ϕ	σ^2	ϕ	σ^2	ϕ
ECMr _{CL}	$\Delta 2$	1,0211	1,0250	1,0015	1,0019	1,0001	1,0001
	$\Delta 4$	0,9933	0,9733	1,0395	1,0579	1,1066	1,0740
	$\Delta 8$	0,9453	0,8960	1,0401	1,0695	0,9918	0,9868
ECMr _{BL}	$\Delta 2$	0,4970	0,4684	0,5713	0,5707	0,7042	0,6932
	$\Delta 4$	0,2705	0,2424	0,4177	0,3952	0,6432	0,5801
	$\Delta 8$	0,1397	0,1172	0,3032	0,2765	0,4733	0,4064
ECMr _{CLc}	$\Delta 2$	1,0020	1,0011	1,0026	1,0027	0,9854	0,9851
	$\Delta 4$	1,0006	0,9223	1,0648	1,0685	0,9574	0,9596
	$\Delta 8$	0,9876	0,8978	0,9741	0,9605	0,9741	0,9605
ECMr _{BLc}	$\Delta 2$	0,9454	0,8327	0,9877	0,9215	0,8921	0,9084
	$\Delta 4$	0,8243	0,6511	0,8780	0,8570	0,9105	0,8488
	$\Delta 8$	0,5021	0,3669	0,8270	0,7472	0,8496	0,6779

Tabla A3.2: Errores Cuadráticos Medios relativos para las estimaciones de σ^2 y ϕ , cuando $n = 512$, de los modelos BL y CL con los enfoques de configuración de bloques por fila y por columna (BLc y CLc).

Bibliografía

- Acosta, J., Alegría, A., Osorio, F., and Vallejos, R. (2021). Assessing the effective sample size for large spatial datasets: A block likelihood approach. *Computational Statistics & Data Analysis*, 162:107282.
- Aune, E., Simpson, D., and Eidsvik, J. (2014). Parameter estimation in high dimensional gaussian distributions. *Statistics and Computing*, 24:247–263.
- Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). Gaussian Predictive Process Models for Large Spatial Data Sets. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 70(4):825–848.
- Bevilacqua, M., Alegria, A., Velandia, D., and Porcu, E. (2016). Composite likelihood inference for multivariate gaussian random fields. *Journal of Agricultural, Biological, and Environmental Statistics*, 21(3):448–469. Publisher Copyright: © 2016, International Biometric Society.
- Bevilacqua, M., Caamaño, C., Arellano-Valle, R., and Morales-Oñate, V. (2020a). Non-gaussian geostatistical modeling using (skew) t processes. *Scandinavian Journal of Statistics*, 48.
- Bevilacqua, M., Caamaño, C., and Gaetan, C. (2020b). On modeling positive continuous data with spatio-temporal dependence. *Environmetrics*, 31.
- Bevilacqua, M., Gaetan, C., Mateu, J., and Porcu, E. (2012). Estimating space and space-time covariance functions for large data sets: A weighted composite like-

- likelihood approach. *Journal of the American Statistical Association*, 107(497):268–280.
- Caamaño, C., Bevilacqua, M., López, C., and Morales-Oñate, V. (2022). Nearest neighbours weighted composite likelihood based on pairs for (non-)gaussian massive spatial data with an application to tukey-hh random fields estimation.
- Caragea, P. C. and Smith, R. L. (2007). Asymptotic properties of computationally efficient alternative estimators for a class of multivariate normal models. *Journal of Multivariate Analysis*, 98(7):1417–1440.
- Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 70(1):209–226.
- Curriero, F. C. and Lele, S. (1999). A composite likelihood approach to semivariogram estimation. *Journal of Agricultural, Biological, and Environmental Statistics*, 4(1):9–28.
- Eidsvik, J., Shaby, B. A., Reich, B. J., Wheeler, M., and Niemi, J. (2014). Estimation and prediction in spatial models with block composite likelihoods. *Journal of Computational and Graphical Statistics*, 23(2):295–315.
- Furrer, R., Genton, M. G., and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15(3):502–523.
- Gaetan, C. and Guyon, X. (2010). *Spatial Statistics and Modeling*. Springer New York, NY.
- Godambe, V. P. (1960). An Optimum Property of Regular Maximum Likelihood Estimation. *The Annals of Mathematical Statistics*, 31(4):1208 – 1211.
- Heagerty, P. J. and Lele, S. R. (1998). A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association*, 93(443):1099–1111.

- Lindgren, F., Rue, H., and 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*, 73:423–498.
- Lindsay, B. (1988). Composite likelihood methods. *Contemporary Mathematics*, 80:221–239.
- Mardia, K. V. and Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71(1):135–146.
- Morales-Navarrete, D., Bevilacqua, M., Caamaño, C., and Castro, L. (2021). Modeling point referenced spatial count data: A poisson process approach.
- Mukerjee, R. (2023). Improving upon the effective sample size based on godambe information for block likelihood inference. *Computational Statistics*, 39:1–14.
- Schabenberger, O. and Gotway, C. A. (2005). *Statistical Methods for Spatial Data Analysis*, chapter 1. Chapman & Hall.
- Shaby, B. and Ruppert, D. (2012). Tapered covariance: Bayesian estimation and asymptotics. *Journal of Computational and Graphical Statistics*, 21(2):433–452.
- Stein, M. L. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Korean Statistical Society*, 37:3–10.
- Stein, M. L., Chi, Z., and Welty, L. J. (2004). Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 66(2):275–296.
- van Lieshout, M. (2019). *Theory of Spatial Statistics A Concise Introduction*. Chapman & Hall.
- Varin, C. (2008). On composite marginal likelihoods. *AStA Advances in Statistical Analysis*, 92(1):1–28.

- Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42.
- Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 50(2):297–312.