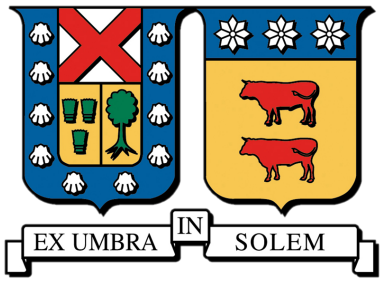


Universidad Técnica Federico Santa María

Verosimilitud por bloques para campos aleatorios Gaussianos-*envueltos*

Autor:
Joaquín Aguirre

2026-1

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: Verosimiltud por bloques para campos aleatorios Gaussianos-envueltos

Nombre del candidato(a): Joaquín Antonio Aguirre Adaros

Carrera / Grado: Magíster en ciencias mención matemática

Campus: Casa Central. **Departamento:** Departamento de matemáticas (DMAT)

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Francisco Andrés Cuevas Pacheco, 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)

☒ 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: 07/08/2026 **Firma:** Francisco

Estudiante o Candidato(a):

Fecha: 04/08/2026 **Firma:** JAntonio

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.

Índice

1. Introducción **5**

2. Preliminares **7**

2.1. Variables aleatorias direccionales 7
2.2. Función de densidad circular 8
2.3. Función característica y momentos 8
2.4. Métricas circulares 9
2.5. Extensión al caso multivariado 10
2.6. Correlación circular 12
2.7. Distribuciones circualres 15
2.8. Distribuciones *proyectadas* 15
2.8.1. Construcción general 15
2.8.2. Caso particular: *proyectada Normal* 16
2.9. Distribuciones *intrínsecas* 18
2.9.1. Construcción general 18
2.9.2. Caso particular: Distribución von Mises 18
2.10. Distribuciones *envueltas* 20
2.10.1. Construcción general 20
2.10.2. Caso particular: Distribución Gaussiana-*envuelta* 20
2.10.3. Ventajas de la distribución Gaussiana-*envuelta* 22
2.10.4. Aproximaciones computacionales mediante truncamiento 23
2.11. Estadística espacial 26
2.11.1. Motivación 26
2.11.2. Definición 26
2.11.3. Campos aleatorios Gaussianos 27

3. Campos aleatorios Gaussianos-*envueltos* **29**

3.1. Motivación 29
3.2. Definición 29
3.3. Propiedades de campos aleatorios Gaussianos-*envueltos* 29

3.3.1. Variograma . . . 30
3.3.2. Función característica y momentos trigonométricos . . . 30
3.4. Correlación circular . . . 30

4. Verosimilitud compuesta **32**
4.1. Inferencia de verosimilitud compuesta . . . 33
4.1.1. Matriz de información de Godambe . . . 33
4.1.2. Resultados Asintóticos . . . 34
4.1.3. Insesgado y consistente . . . 34
4.2. Formas de verosimilitud compuesta . . . 36
4.2.1. Verosimilitud por parejas . . . 36
4.2.2. Verosimilitud por bloques . . . 37
4.3. Algoritmo K-means para la formación de bloques . . . 38

5. Estimación distribucional de los estimadores CL mediante Bootstrap paramétrico **40**

6. Propuesta **42**
6.1. Verosimilitud compuesta por parejas para campos Gaussianos-*envueltos* . . . 42
6.2. Verosimilitud compuesta por bloques para campos Gaussianos-*envueltos* . . . 43
6.3. Casos extremos y papel del tamaño de bloque . . . 44
6.4. Modelo latente y parametrización direccional . . . 45

7. Simulaciones **46**
7.1. Diseño del estudio de simulación . . . 46
7.1.1. Configuración espacial y parámetros . . . 46
7.1.2. Configuración de los métodos de estimación . . . 47
7.1.3. Especificaciones . . . 48
7.1.4. Observaciones . . . 56

8. Aplicación **57**
8.1. Análisis descriptivo . . . 57
8.2. Ajuste . . . 61

9. Conclusiones **63**

10.Trabajos futuros **66**

11.Referencias **67**

1. Introducción

En el análisis estadístico clásico, los datos continuos suelen modelarse como realizaciones de variables definidas sobre toda la recta real, como es el caso de la distribución normal. Sin embargo, en numerosas aplicaciones las observaciones poseen una naturaleza intrínsecamente periódica, de modo que valores que difieren en un múltiplo de 2π (o de $360°$) representan en realidad el mismo valor o estado. En este contexto, resulta más apropiado considerar que el soporte de la variable aleatoria es el círculo unitario (en lugar de la recta), lo que da lugar al estudio de los llamados *datos direccionales*.

El término *dato direccional* alude a la representación de un vector en el círculo unitario mediante un ángulo. Por ello, son ejemplos típicos la dirección de la mirada de un sujeto, el rumbo de un barco, la dirección del viento registrada en una estación meteorológica o la orientación de animales marcados con collares GPS. En todos estos casos, valores cercanos a 0 y a 2π representan direcciones prácticamente idénticas, de modo que las herramientas desarrolladas para datos reales estándar pueden resultar engañosas si no se tiene en cuenta esta estructura periódica. Estas consideraciones motivan el desarrollo de modelos y métodos específicos para datos circulares y, en general, para datos definidos sobre variedades como la circunferencia o la esfera.

Una manera natural de construir variables aleatorias angulares consiste en aplicar la operación módulo 2π a variables aleatorias continuas definidas sobre la recta real, como la normal. De esta idea surge la distribución Gaussiana-*envuelta*, ampliamente utilizada en estadística direccional por las propiedades que hereda de la normal y por su extensión natural al caso multivariado.

Otra clase de datos particularmente relevante es la de los datos georreferenciados, que surgen cuando resulta esencial incorporar la estructura espacial en el análisis (por ejemplo, al estudiar precios de viviendas en un vecindario o valores de píxeles en una imagen). En aplicaciones como direcciones de mareas o vientos, además de la componente espacial aparece una componente direccional, por lo que se requiere desarrollar herramientas específicas para datos direccionales georreferenciados. Esto motiva el uso de campos aleatorios Gaussianos-*envueltos* y, en particular, el problema central de esta tesis: proponer una alternativa de inferencia que evite las dificultades computacionales de métodos tradicionales como la verosimilitud completa, la cual resulta impracticable para tamaños muestrales moderados o grandes.

Nuestro enfoque para abordar este problema se basa en una verosimilitud compuesta por bloques. Esta construcción puede verse como una extensión natural del enfoque de verosimilitud compuesta por parejas desarrollado por Alegría et al. [2016] para campos Gaussianos-*envueltos* espacio–temporales, en el que las contribuciones de la verosimilitud se construyen a partir de pares de observaciones, como sugiere el nombre. Aquí, en cambio, consideramos bloques de datos de mayor dimensión, con el objetivo de capturar estructuras de dependencia

más ricas y, al mismo tiempo, mejorar la eficiencia estadística de los estimadores.

Adicionalmente, nuestro enfoque busca mantener un buen desempeño computacional, de modo que la metodología resulte viable en la práctica para una amplia gama de aplicaciones. Esto puede ser más difícil de lograr en métodos que dependen de manera intensiva de simulaciones de Monte Carlo, debido al elevado costo computacional que suelen implicar.

Desde una perspectiva bayesiana, Jona-Lasinio et al. [2012] proponen un modelo jerárquico para este tipo de datos que, como es habitual en estadística bayesiana, requiere del uso de algoritmos MCMC (Markov Chain Monte Carlo) para llevar a cabo la inferencia. Dado el costo computacional asociado a estos procedimientos, no consideraremos este enfoque en nuestro análisis comparativo.

Otra alternativa natural es el uso del algoritmo EM. Por ejemplo, Fisher and Lee [1994] emplean este enfoque en el contexto de series de tiempo Gaussiana-*envueltas*, y en principio podría adaptarse una estrategia similar para campos Gaussianos-*envueltos*. Sin embargo, en dicho algoritmo aparecen términos integrales de difícil tratamiento analítico, que requerirían aproximarse mediante simulaciones de Monte Carlo intensivas. Por esta razón, esta segunda alternativa tampoco será considerada en nuestro estudio.

La memoria se organiza de la siguiente manera. Tras la introducción, en la sección *Preliminares* se presentan las herramientas básicas de estadística direccional: datos direccionales, variables aleatorias direccionales, distribuciones direccionales, la distribución Gaussiana-*envuelta* y las principales ventajas de esta última como modelo para datos angulares.

A continuación, se introduce el marco de la estadística espacial y se estudian campos aleatorios Gaussianos-*envueltos*, que constituyen el objeto matemático principal de esta memoria. Luego, se define la verosimilitud compuesta y se revisa el método bootstrap, completando así el marco teórico necesario para situar nuestro trabajo.

El núcleo de la tesis contiene la propuesta metodológica, el estudio de simulación, una aplicación a datos reales y las conclusiones. Finalmente, se recopilan posibles líneas de trabajo futuro y se presentan las referencias bibliográficas utilizadas.

2. Preliminares

2.1. Variables aleatorias direccionales

La **estadística direccional** es la rama de la estadística que se ocupa del análisis de datos cuya naturaleza es *angular* o *direccional*, es decir, datos que representan direcciones u orientaciones en el espacio. Típicamente, esto incluye datos definidos sobre la circunferencia unitaria (datos *circulares*, en S^1) o, para mayores dimensiones, es usual usar representaciones en la esfera unitaria p-dimensional S^p, la cual se define como $S^p = \{\mathbf{x} \in \mathbb{R}^{p+1} : \|\mathbf{x}\| = 1\}$. No obstante, nosotros en este escrito estaremos particularmente interesados en representaciones en el toro p-dimensional dado por $\mathbb{T}^p = [0, 2\pi)^p \cong S^1 \times \cdots \times S^1$, esto debido a que es coherente con respecto a la estructura espacial que propondremos más adelante.

Un rasgo característico de los datos direccionales es su periodicidad intrínseca: por ejemplo, los ángulos 0° y 360° representan la *misma* dirección. Consecuentemente, las operaciones estadísticas habituales para datos lineales (en $\mathbb{R}$) requieren adaptaciones. En particular, el cálculo de promedios debe redefinirse para respetar la geometría circular. Por ejemplo, el promedio de 2° y 358° *no* debe calcularse como $(2 + 358)/2 = 180°$, pues 180° no refleja la dirección media real (que en este caso es 0°). En la práctica se utiliza la *media circular*, definida a partir de la suma vectorial de los vectores unitarios correspondientes a cada ángulo.

La estadística direccional tiene usos en variedad de campos aplicados. En meteorología, por ejemplo, se analizan direcciones del viento [Carta et al., 2008], en particular, este caso va a ser el que vamos a usar para más adelante en la etapa de aplicar nuestra nueva metodología, analizar este tipo de datos puede ayudar a describir, interpretar y predecir distintos fenómenos del área como las tormentas o también puede servir para ubicar de forma informada y estratégica turbinas eólicas. Otros posibles ejemplos son en geología, donde se estudian epicentros [Chang, 1993], los cuales pueden representarse como puntos en la esfera y también en biología se estudian trayectorias de animales migratorios [Schmidt-Koenig, 1965]. Estas aplicaciones motivan la necesidad de una teoría estadística que respete la geometría direccional y permita conducir análisis coherentes con respecto a esta estructura intrínseca de los datos.

Para introducirnos a este tema, lo primero que tenemos que definir es qué es una variable aleatoria direccional. Definimos una **variable aleatoria direccional** como aquella cuya realización pertenece a una variedad compacta con estructura circular, esférica o toroidal. En el caso más simple, una **variable aleatoria circular** Θ toma valores en el círculo unitario S^1, que puede identificarse con el intervalo $[0, 2\pi)$ mediante la parametrización angular, o equivalentemente con el subconjunto $\{(\cos\theta, \sin\theta) : \theta \in [0, 2\pi)\} \subset \mathbb{R}^2$, ver en Mardia and Jupp [2000], Jammalamadaka and SenGupta [2001].

Los datos direccionales o circulares representan direcciones de algún fenómeno y, por tanto, suelen expresarse mediante un ángulo $\theta \in [0, 2\pi)$. Alternativamente, pueden representarse como puntos en el círculo unitario mediante la transformación $\theta \mapsto (\cos\theta, \sin\theta)$, obteniendo así su *vector direccional* correspondiente. Una propiedad fundamental de estos datos es su

periodicidad o **modularidad**: θ y $\theta+2k\pi$ representan la misma dirección para todo $k \in \mathbb{Z}$. Esta identificación modifica sustancialmente la noción de proximidad respecto a la usual en $\mathbb{R}$

2.2. Función de densidad circular

Una variable aleatoria circular Θ la podemos caracterizar por su función de densidad de probabilidad $f(\theta)$, definida en $\mathbb{R}$. Esta función de densidad cumple con lo que uno esperaría de una función de densidad tradicional, es decir, ser no negativa en todo su dominio e integrar 1, sin embargo, por su periocidad, es extendible a todo $\mathbb{R}$ a través de $\tilde{f}$ tal que $\tilde{f}(x+2k\pi) = f(x) \quad \forall x \in [\alpha, \alpha+2\pi) \quad \wedge \forall k \in \mathbb{Z}$. Gracias a esto, podemos usar la extensión $\tilde{f}$ y luego restringirnos al intervalo de largo 2π más conveniente, por tanto, sin perdida de generalidad, asumiremos que estaremos en el intervalo $[0, 2\pi)$ a menos que se especifique lo contrario

2.3. Función característica y momentos

Una consecuencia directa de la periodicidad es que la función característica $\varphi_\Theta(t) = \mathbb{E}[e^{it\Theta}]$ satisface

$$\varphi_\Theta(t) = \mathbb{E}[e^{it\Theta}] = \mathbb{E}[e^{it(\Theta+2\pi)}] = e^{2\pi it}\varphi_\Theta(t).$$

Esto implica que $\varphi_\Theta(t)(1 - e^{2\pi it}) = 0$, por lo que $\varphi_\Theta(t)$ debe ser no nula únicamente cuando $e^{2\pi it} = 1$, es decir, cuando $t \in \mathbb{Z}$. Por tanto, la función característica de una variable circular solo es válida cuando $t \in \mathbb{Z}$.

En estadística direccional, los momentos usuales son reemplazados por **momentos trigonométricos**. El p-ésimo momento trigonométrico se define como

$$\alpha_p = \mathbb{E}[\cos(p\Theta)] \quad \text{y} \quad \beta_p = \mathbb{E}[\sin(p\Theta)],$$

o de forma compleja, $\mu_p = \mathbb{E}[e^{ip\Theta}] = \alpha_p + i\beta_p$. En particular, el primer momento trigonométrico

$$\mu_1 = \mathbb{E}[e^{i\Theta}] = \int_0^{2\pi} e^{i\theta} f(\theta)\,\mathrm{d}\theta = \bar{C} + i\bar{S},$$

donde

$$\bar{C} = \int_0^{2\pi} \cos(\theta) f(\theta)\,\mathrm{d}\theta \quad \text{y} \quad \bar{S} = \int_0^{2\pi} \sin(\theta) f(\theta)\,\mathrm{d}\theta,$$

permite definir la **dirección media** $\bar{\theta}$ mediante

$$\cos(\bar{\theta}) = \frac{\bar{C}}{\bar{R}} \quad \text{y} \quad \sin(\bar{\theta}) = \frac{\bar{S}}{\bar{R}},$$

donde $\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2} = |\mu_1|$ es la **longitud media resultante**. El ángulo $\bar{\theta}$ se recupera mediante

$$\bar{\theta} = \operatorname{atan}(\bar{S}, \bar{C}),$$

donde atan es la función arcotangente de dos argumentos que preserva el cuadrante correcto[1] [Mardia and Jupp, 2000].

La longitud media resultante $\bar{R} \in [0, 1]$ mide la concentración de la distribución alrededor de la dirección media: $\bar{R} \approx 1$ indica una fuerte concentración, mientras que $\bar{R} \approx 0$ indica una dispersión uniforme [Fisher, 1993]. La **varianza circular** se define como $V = 1 - \bar{R}$, y la **desviación estándar circular** como $v = \sqrt{-2 \ln \bar{R}}$ [Mardia and Jupp, 2000].

2.4. Métricas circulares

Como los datos direccionales poseen propiedades modulares, observaciones cercanas a 0 y cercanas a 2π son en realidad cercanas entre sí. Sin embargo, las nociones usuales de distancia en $\mathbb{R}$ no capturan este efecto, pudiendo inducir diferencia o lejanía en datos realmente similares o cercanos. Por ello, se emplean métricas que respetan la estructura y la propiedad modular de los datos [Jammalamadaka and SenGupta, 2001].

Dos métricas circulares comúnmente utilizadas son:

- **Distancia angular**:

$$d_{\text{ang}}(\alpha, \beta) = \min\big(|\alpha - \beta|,\, 2\pi - |\alpha - \beta|\big) = \pi - \big|\pi - |\alpha - \beta|\big|.$$

  Esta métrica mide la longitud del arco más corto entre dos puntos en el círculo.

- **Distancia cordal**:

$$d_{\text{cor}}(\alpha, \beta) = 1 - \cos(\alpha - \beta) = 2 \sin^2\left(\frac{\alpha - \beta}{2}\right).$$

  Esta métrica corresponde a la mitad del cuadrado de la distancia euclidiana entre los vectores unitarios $(\cos\alpha, \sin\alpha)$ y $(\cos\beta, \sin\beta)$ en $\mathbb{R}^2$. Por tanto, esta métrica depende de la longitud de la cuerda que une ambos puntos del círculo unitario

En la Figura 1 se ilustran gráficamente las diferencias entre la distancia angular (a lo largo del arco) y la distancia cordal (a través de la cuerda) entre dos puntos sobre la circunferencia unitaria.

Estas métricas logran capturar el comportamiento modular o periódico en lugar del comportamiento lineal al que estamos acostumbrados; por lo tanto, podemos usar estas funciones como una verdadera noción de cercanía.

[1] Para $y, x \in \mathbb{R}$, $\text{atan}(y, x)$ se define como el ángulo $\theta \in (-\pi, \pi]$ tal que $\cos\theta = x/\sqrt{x^2 + y^2}$ y $\sin\theta = y/\sqrt{x^2 + y^2}$. En particular, si $x > 0$ entonces $\text{atan}(y, x) = \arctan(y/x)$, mientras que si $x < 0$ se ajusta el valor de $\arctan(y/x)$ sumando o restando π para ubicar θ en el cuadrante correcto.

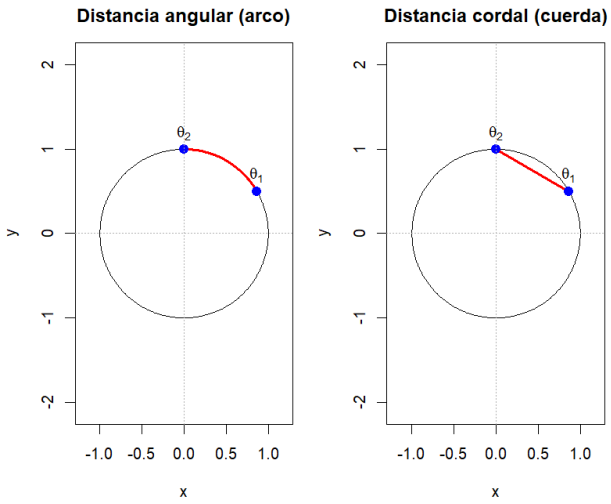


Figura 1: Ilustración de la distancia angular (arco) y la distancia cordal (cuerda) entre dos direcciones sobre la circunferencia unitaria.

2.5. Extensión al caso multivariado

Los conceptos anteriores se extienden de manera natural al caso de variables aleatorias **direccionales multivariadas** o **toroidales**. En este contexto, consideramos un vector aleatorio $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_p)$ cuya componente i-ésima es un ángulo, esto es, $\Theta_i \in [0, 2\pi)$ para todo $i = 1, \ldots, p$. El espacio muestral de $\boldsymbol{\Theta}$ es el **toro p-dimensional**

$$\mathbb{T}^p = [0, 2\pi)^p \cong S^1 \times \cdots \times S^1,$$

que se obtiene como producto cartesiano de p circunferencias unitarias.

Para visualizar, el toro bidimensional $\mathbb{T}^2$ puede representarse como la superficie de una dona en $\mathbb{R}^3$. En esta representación, uno de los ángulos (por ejemplo, θ_1) parametriza la circunferencia principal del toro, mientras que el otro ángulo (θ_2) parametriza la circunferencia del "tubo" que gira alrededor de dicha circunferencia principal, como se ilustra en la Figura 2.

Cada realización $(\theta_1, \ldots, \theta_p)$ del vector aleatorio puede interpretarse como un punto en esta variedad, o alternativamente, mediante su representación vectorial en $\mathbb{R}^{2p}$ como

$$\big(\cos\theta_1, \sin\theta_1, \ldots, \cos\theta_p, \sin\theta_p\big) \in S^1 \times \cdots \times S^1 \subset \mathbb{R}^{2p}.$$

De forma análoga, la función de densidad de un vector aleatorio $f(\theta_1, \ldots, \theta_p)$ sobre $\mathbb{T}^p$ debe satisfacer la no negatividad, integrar 1 y tiene una extensión natural $\tilde{f} = f(\theta_1 + 2k_1\pi, \ldots, \theta_p + 2k_p\pi)$ para todo $(k_1, \ldots, k_p) \in \mathbb{Z}^p$:

Los momentos trigonométricos multivariados siguen la misma lógica que en el caso univariado. Sea $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_p)$ un vector aleatorio direccional p-dimensional. Para cada componente $j \in \{1, \ldots, p\}$, los momentos marginales de primer orden se definen como

$$\bar{C}_j = \mathbb{E}[\cos\Theta_j] \quad \text{y} \quad \bar{S}_j = \mathbb{E}[\sin\Theta_j],$$

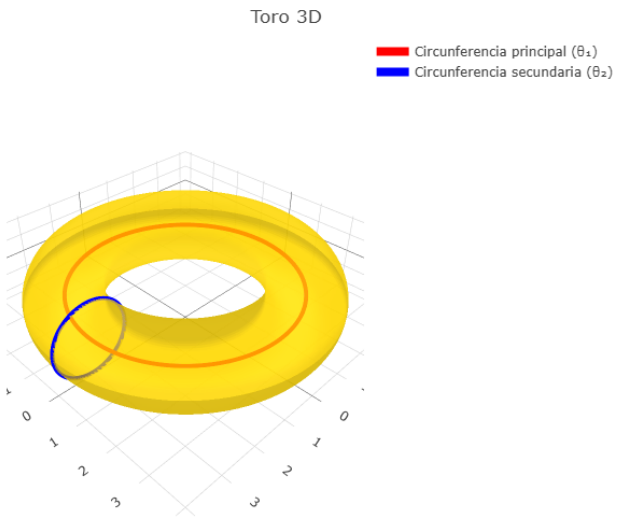


Figura 2: Representación del toro bidimensional $\mathbb{T}^2$ como superficie en $\mathbb{R}^3$, con un punto identificado por sus coordenadas angulares (θ_1, θ_2).

y permiten definir la dirección media $\bar{\theta}_j$ y la longitud media resultante $\bar{R}_j = \sqrt{\bar{C}_j^2 + \bar{S}_j^2}$ para cada coordenada angular. Sin embargo, en el caso multivariado es fundamental considerar también la estructura de dependencia entre las componentes angulares, lo que puede caracterizarse mediante momentos cruzados de la forma

$$\mathbb{E}\big[\cos(r_1\Theta_1 + \cdots + r_p\Theta_p)\big]$$

para distintos vectores de enteros $(r_1, \ldots, r_p) \in \mathbb{Z}^p$ [Mardia and Jupp, 2000, Jammalamadaka and SenGupta, 2001].

Definir posibles métricas en el toro $\mathbb{T}^p$ también requieren cierto cuidado. Una opción natural es la llamada *métrica producto*, que para dos puntos $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)$ y $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ se define como

$$d(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sqrt{\sum_{j=1}^{p} d_{\text{circ}}(\alpha_j, \beta_j)^2},$$

donde d_{circ} denota cualquiera de las métricas circulares introducidas anteriormente (angular o cordal). Esta construcción respeta la estructura de producto del toro y mantiene nuestra noción de distancia que introdujimos para el caso circular. Como caso particular, si usamos la métrica producto con la distancia angular, obtenemos lo que se conocería como la distancia geodésica [Lee, 2018].

El estudio de variables toroidales es relevante en aplicaciones donde múltiples ángulos se miden simultáneamente, como en la caracterización de conformaciones moleculares mediante ángulos diedros en bioinformática estructural, en el análisis de ritmos circadianos multivariados en cronobiología, y en problemas de sincronización de osciladores en física y neurociencia [Mardia and Jupp, 2000, Ley and Verdebout, 2017].

2.6. Correlación circular

Las nociones de dependencia naturalmente, difieren para el contexto circular también debido a la naturaleza distinta de los datos, por tanto es importante definir correctamente correlación en el sentido circular. Para esto vamos a introducir el coeficiente de Jammalamadaka-Sarma, que será fundamental para nuestro análisis posterior.

Veamos las limitaciones de la correlación lineal en este contexto.

Sean $(\alpha_1, \beta_1), \ldots, (\alpha_n, \beta_n)$ observaciones de un par de variables circulares (A, B) con función de densidad conjunta $f(\alpha, \beta)$ definida en el toro $[0, 2\pi) \times [0, 2\pi)$. Denotemos por μ_α y μ_β las direcciones medias de las variables marginales respectivas. Entonces, si usáramos el coeficiente de correlación lineal de Pearson

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{\mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big]}{\sqrt{\mathbb{E}\big[(X - \mathbb{E}[X])^2\big]\,\mathbb{E}\big[(Y - \mathbb{E}[Y])^2\big]}},$$

lo cual sería lo natural en un contexto lineal. Esto conduce a varias problemáticas [Jammalamadaka and SenGupta, 2001], en particular, la definición no respeta el comportamiento modular de las variables, por ejemplo, si la media de ambas variables fuera $\mu = 0{,}01$ rad, entonces un valor como 6,27 rad inflaría las diferencias $\alpha_i - \mu_\alpha$ y $\beta_i - \mu_\beta$ casi al máximo posible, cuando en realidad sabemos intuitivamente que son datos muy cercanos. El efecto sobre ρ sería completamente contrario al real, haciendo que la interpretación del coeficiente sea inconsistente. Esto motiva a introducir un coeficiente de correlación específicamente diseñado para el caso circular.

Para superar estas limitaciones, diversos investigadores han propuesto coeficientes de correlación específicamente diseñados para datos circulares. El más ampliamente utilizado es el coeficiente de Jammalamadaka-Sarma [Jammalamadaka and Sarma, 1988], el cual introduciremos en esta sección y usaremos en este trabajo.

Sean A y B variables aleatorias circulares con direcciones medias μ_α y μ_β respectivamente. El coeficiente de correlación circular de Jammalamadaka-Sarma se define como:

$$\rho_C(\mathrm{A}, \mathrm{B}) = \frac{\mathbb{E}[\sin(\mathrm{A} - \mu_\alpha)\sin(\mathrm{B} - \mu_\beta)]}{\sqrt{\mathbb{E}[\sin^2(\mathrm{A} - \mu_\alpha)]\,\mathbb{E}[\sin^2(\mathrm{B} - \mu_\beta)]}}.$$

Esta definición puede interpretarse como la correlación de Pearson aplicada a las desviaciones angulares desde las direcciones medias, medidas mediante la función seno [Jammalamadaka and SenGupta, 2001]. La elección de la función seno no es arbitraria: es la manera natural de medir desviaciones angulares con signo en el círculo, siendo análoga a $x - \mu$ en la recta.

Por lo tanto, la versión muestral se ve como

$$\hat{\rho}_C = \frac{\sum_{i=1}^{n} \sin(\alpha_i - \bar{\alpha})\sin(\beta_i - \bar{\beta})}{\sqrt{\sum_{i=1}^{n} \sin^2(\alpha_i - \bar{\alpha})\ \sum_{i=1}^{n} \sin^2(\beta_i - \bar{\beta})}}.$$

Para una muestra $\{(\alpha_i, \beta_i)\}_{i=1}^n$, donde $\bar{\alpha}$ y $\bar{\beta}$ son las direcciones medias muestrales, definidas como:

$$\bar{\alpha} = \operatorname{atan}\left(\sum_{i=1}^{n} \sin \alpha_i, \sum_{i=1}^{n} \cos \alpha_i\right), \quad \bar{\beta} = \operatorname{atan}\left(\sum_{i=1}^{n} \sin \beta_i, \sum_{i=1}^{n} \cos \beta_i\right).$$

Propiedades del coeficiente. Revisamos que esta definición cumple con propiedades deseadas de un coeficiente de correlación análogas para el caso lineal [Jammalamadaka and SenGupta, 2001, Fisher, 1993].

De la misma definición se puede verificar de manera trivial que se cumplen las cotas de $-1 \leq \rho_C \leq 1$, además de la simetría $\rho_C(A, B) = \rho_C(B, A)$ y que respeta el signo, es decir, $\rho_C(-A, B) = -\rho_C(A, B)$ dado por la propiedad del seno.

Otro comportamiento requerido ajustado para este contexto es que sea invariante con respecto a rotaciones, es decir, $\forall c \in [0, 2\pi)$ se tiene

$$\rho_C(A + c, B) = \rho_C(A, B) = \rho_C(A, B + c)$$

,

esto es fundamental, dado que la relación no debe depender del punto de referencia utilizado. La deducción de esto se hace simplemente usando la identidad del seno de la suma y simplificando.

Aunque el coeficiente de Jammalamadaka-Sarma es el más utilizado debido a su simplicidad y propiedades deseables, existen algunas alternativas en la literatura, donde según el contexto, podrían ser más adecuadas.

Coeficiente de Fisher-Lee. Fisher and Lee [1983] propusieron un coeficiente basado en momentos trigonométricos complejos:

$$\rho_{FL}(\mathrm{A}, \mathrm{B}) = \frac{|\mathbb{E}[e^{i(\mathrm{A}-\mu_\alpha)} e^{-i(\mathrm{B}-\mu_\beta)}]|}{\sqrt{(1 - |\mathbb{E}[e^{2i(\mathrm{A}-\mu_\alpha)}]|)(1 - |\mathbb{E}[e^{-2i(\mathrm{B}-\mu_\beta)}]|)}}.$$

que tiene propiedades similares pero puede ser más natural en contextos donde se trabaja con representaciones complejas de datos circulares, como en procesamiento de señales o análisis de Fourier.

Coeficiente basado en rangos. Para mayor robustez ante valores atípicos o desviaciones de supuestos distribucionales, Mardia and Jupp [2000] discuten coeficientes basados en rangos circulares, análogos al coeficiente de correlación de Spearman en el caso lineal. Estos coeficientes son particularmente útiles cuando la distribución subyacente es desconocida o se sospecha la presencia de observaciones anómalas.

Coeficiente T de Jupp-Mardia. Para datos en la esfera $\mathbb{S}^2$ (caso tridimensional), Jupp and Mardia [1980] propusieron el coeficiente T, que generaliza estas ideas a dimensiones superiores. Este coeficiente tiene aplicaciones importantes en geología (orientaciones de estratos rocosos), cristalografía (orientaciones cristalinas), y palemagnetismo.

2.7. Distribuciones circualres

En esta sección presentamos los tres casos más famosos y estudiados de distribuciones en el círculo, donde mostraremos sus definiciones, derivaciones y propiedades.

2.8. Distribuciones *proyectadas*

2.8.1. Construcción general

La familia de distribuciones *proyectadas* se construye considerando una variable aleatoria $\mathbf{X} \sim F$ latente, definida en $\mathbb{R}^2$ con $\mathbb{P}(\mathbf{X} = \mathbf{0}) = 0$, luego definimos $\mathbf{U}$ como la variable latente normalizada, es decir,

$$\mathbf{U} = \frac{\mathbf{X}}{\|\mathbf{X}\|}.$$

.

Suele ser útil usar la parametrización en coordenadas polares para derivar de forma más directa densidades de esta familia. Concretamente, expresar $\mathbf{X} = (R, \Theta)$, donde el Jacobiano de la transformación es conocido, r, luego con eso se puede reescribir la densidad bivariada de forma práctica y bastaría con marginalizar sobre R, o en otras palabras, integrar sobre r, que toma valores de 0 a ∞ [Mardia and Jupp, 2000].

Podemos notar que la extensión natural al caso multivariado es hacia la hyperesfera dado que la definición sería completamente análoga, usando variables latentes p-variadas, sin embargo, esta formulación daría con datos esféricos, lo cual no es compatible con nuestro estudio dado que necesitamos de datos toroidales.

La forma en que se ha propuesto en la literatura formular datos toroidales de carácter *proyectados* a sido en usar un vector aleatorio $\mathbf{X}$ $2p$-variado, agrupado por pares de forma $(\mathbf{X_1}^\top, \cdots, \mathbf{X_p}^\top)^\top$, donde

$$\mathbf{X_i} = (X_{i1}, X_{i2})^\top \quad \forall i \in \{1, \cdots, p\}.$$

A cada par se le obtiene su proyección asociada a como se definió en el caso bivariado. Esta formulación permite definir dependencia asociada a la matriz de covarianza del vector aleatorio latente [Kato et al., 2025].

Una propiedad potente de las distribuciones *proyectadas* es que permiten modelar unimodalidad, multimodalidad, simetría y asimetría [Wang and Gelfand, 2013]. Esto implica que pueden ser utilizadas para variedad de datos reales. No obstante, formulaciones *proyectadas* suelen tener otras complejidades, como por ejemplo de identificabilidad, en particular, un vector aleatorio latente $\mathbf{X}$ puede derivar la misma distribución *proyectada* que su versión amplificada $\lambda\mathbf{X}$, con λ escalar no nulo [Wang and Gelfand, 2013].

2.8.2. Caso particular: *proyectada Normal*

El caso más interesante de esta familia es la distribución *Normal proyectada*, estudiada principalmente en Wang and Gelfand [2013], aunque en Mardia and Jupp [2000] se introdujeron varias propiedades previamente.

Variable Normal proyectada Siguiendo la construcción previamente explicada, usamos $\mathbf{X} \sim N_2(\boldsymbol{\mu}, \Sigma)$ como una variable aleatoria Gaussiana bivariada y definimos

$$U = \frac{\mathbf{X}}{\|\mathbf{X}\|} \sim PN_2(\boldsymbol{\mu}, \Sigma),$$

donde al parametrizar X en su forma polar, identificando $\mathbf{v}(\theta) = (\cos\theta, \sin\theta)^\top$ y marginalizando por r se puede obtener la función de densidad que dieron en Mardia and Jupp [2000], la cual al mantener la notación que usan ellos, toma la forma de

$$f(\theta; \boldsymbol{\mu}, \Sigma) = \frac{\phi_2\big(\mathbf{v}(\theta);\, \mathbf{0}, \Sigma\big) + |\Sigma|^{-1/2}\, D(\theta)\, \Phi\big(D(\theta)\big)\, \phi\left(\dfrac{\boldsymbol{\mu} \wedge \mathbf{v}(\theta)}{|\Sigma|^{1/2} \left(\mathbf{v}(\theta)^\top \Sigma^{-1} \mathbf{v}(\theta)\right)^{1/2}} \right)}{\mathbf{v}(\theta)^\top \Sigma^{-1} \mathbf{v}(\theta)},$$

donde

$$D(\theta) = \frac{\boldsymbol{\mu}^\top \Sigma^{-1} \mathbf{v}(\theta)}{\left(\mathbf{v}(\theta)^\top \Sigma^{-1} \mathbf{v}(\theta)\right)^{1/2}}, \qquad \boldsymbol{\mu} \wedge \mathbf{v}(\theta) = \mu_1 \sin\theta - \mu_2 \cos\theta,$$

$\phi_2(\cdot;\, \mathbf{0}, \Sigma)$ denota la función de densidad de $N_2(\mathbf{0}, \Sigma)$, y ϕ, Φ son la densidad y la función de distribución acumulada de $N(0,1)$, respectivamente [Mardia and Jupp, 2000].

Vector Normal proyectado. Para datos angulares con soporte en el toro $\mathbb{T}^p = [0, 2\pi)^p$ seguimos la formulación de proyectar por pares como se usa en Wang and Gelfand [2013], Mastrantonio [2018], Kato et al. [2025], donde definimos

$$\mathbf{X} = (\mathbf{X}_1^\top, \ldots, \mathbf{X}_p^\top)^\top \sim N_{2p}(\boldsymbol{\mu}_w, \Sigma_w), \qquad \mathbf{X}_i = (X_{i1}, X_{i2})^\top \in \mathbb{R}^2,$$

y para cada $i = 1, \ldots, p$,

$$\Theta_i = \operatorname{atan}\left(\frac{X_{i2}}{X_{i1}}\right) \in [0, 2\pi), \qquad R_i = \|\mathbf{X}_i\| \in \mathbb{R}^+,$$

. El vector $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_p)^\top$ se distribuye como una normal proyectada multivariada sobre el hipertoro,

$$\boldsymbol{\Theta} \sim PN_p(\boldsymbol{\mu}_x, \Sigma_x),$$

con parámetros identificados a través del vector aleatorio Gaussiano latente. Notamos que las medias y bloques diagonales 2×2 de Σ_x controlan la forma marginal de cada Θ_i, mientras que los bloques cruzados $\Sigma_{x,ij}$ $(i \neq j)$ inducen la dependencia. Podemos notar que cada

marginal efectivamente distribuye normal proyectada por construcción dado que marginal de normal es también normal.

Para este caso, no se tiene una expresión en forma cerrada para la densidad marginal de $\boldsymbol{\Theta}$ cuando $p \geq 2$. Por tanto, se suele trabajar con la densidad del vector aleatorio $2p$-variado en forma polar, usando así una formulación aumentada de $\boldsymbol{\Theta}$ como $(\boldsymbol{\Theta}, \mathbf{R})$:

$$f(\boldsymbol{\theta}, \mathbf{r}; \boldsymbol{\mu}_x, \Sigma_x) = \left(\prod_{i=1}^{p} r_i \right) \phi_{2p}(\mathbf{x} \mid \boldsymbol{\mu}_x, \Sigma_x), \qquad \mathbf{x} = \mathbf{x}(\boldsymbol{\theta}, \mathbf{r}),$$

donde los R_i se tratan como variables latentes y $\phi_{2p}(\cdot \mid \boldsymbol{\mu}_w, \Sigma_w)$ denota la densidad de la $N_{2p}(\boldsymbol{\mu}_w, \Sigma_w)$. Esto es útil debido a que si se formula a través de estadística bayesiana, podemos usar algoritmos tipo Gibbs sampling para atacar el problema como propone Mastrantonio [2018].

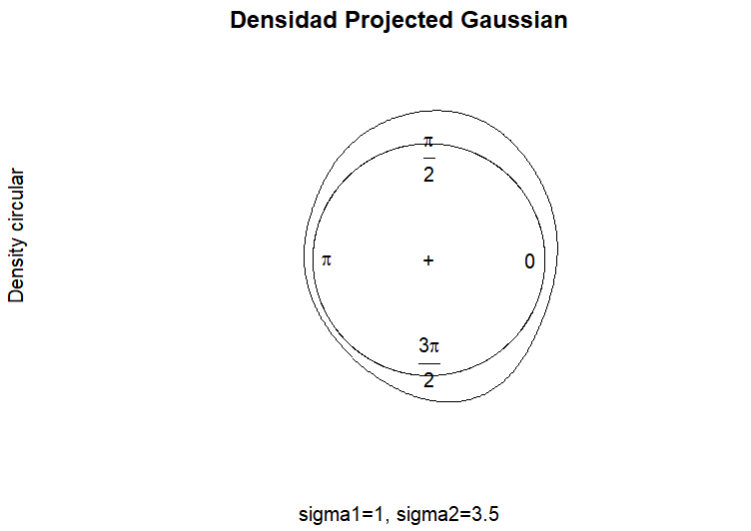


Figura 3: Función de densidad de la distribución *proyectada Normal* univariada con diferentes parámetros. Se observa la capacidad de la distribución para modelar asimetría.

Como se muestra en la Figura 3, la distribución puede modelar asimetría.

2.9. Distribuciones *intrínsecas*

2.9.1. Construcción general

La familia de distribuciones *intrínsecas* se llaman así debido a que partimos de la base que los datos ya viven dentro de la variedad asociada, lo cual se traduce en usar una distribución auxiliar y condicionarla a estar en la variedad. Para una variable circular *intrínseca* habría que usar $\mathbf{X} \sim F$ y condicionarlo sobre el círculo unitario, ie., La distribución direccional *intrínseca* correspondiente se define como

$$\mathbb{P}_{\mathbb{S}^1}(\mathbf{X}) = \mathbb{P}(\mathbf{X} \mid \|\mathbf{X}\| = 1).$$

Para el caso multivariado, notamos que la extensión a la hyperesfera es trivial dado que bastaría usar un $\mathbf{X}$ p-variado y condicionarlo con respecto a la norma [Mardia and Jupp, 2000], sin embargo, nosotros requerimos de datos toroidales.

Para definir los datos *intrínsecos* en el toro, necesitamos usar un vector latente $\mathbf{X}$ $2p$-variado, agrupado por pares de forma $(\mathbf{X_1}^\top, \cdots, \mathbf{X_p}^\top)^\top$, donde

$$\mathbf{X_i} = (X_{i1}, X_{i2})^\top \quad \forall i \in \{1, \cdots, p\}.$$

Luego definimos la densidad como

$$\mathbb{P}_{\mathbb{T}^p}(\mathbf{X}) = \mathbb{P}(\mathbf{X} \mid \|\mathbf{X_i}\| = 1 \quad \forall i \in \{1, \cdots, p\}).$$

2.9.2. Caso particular: Distribución von Mises

La distribución von Mises se suele decir que es el análogo circular de la distribución normal debido a similitud de propiedades y pertenece específicamente a la familia de distribuciones *intrínsecas* Mardia and Jupp [2000].

Variable von Mises. La distribución von Mises se obtiene condicionando una distribución normal bivariada con media μ e isotrópica, es decir, matriz de covarianza $\sigma^2 I_2$. El condicionamiento es naturalmente con respecto al círculo unitario y tras reparametrización en coordenadas polares, su función de densidad toma la forma

$$f(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(\theta - \mu)\}, \quad \theta \in [0, 2\pi),$$

donde $\mu \in [0, 2\pi)$ es el parámetro de dirección media, $\kappa \geq 0$ es el parámetro de concentración, e $I_0(\cdot)$ denota la función de Bessel modificada de primera clase de orden cero [Mardia and Jupp, 2000].

Esta distribución se suele denotar por $vM(\mu, \kappa)$ en la literatura. Como podríamos intuir del comportamiento de un parámetro de concentración, para $\kappa = 0$ se reduce a la distribución uniforme circular, en cambio cuando $\kappa \to \infty$, converge a una masa puntual en μ. La distribución von Mises es unimodal y simétrica alrededor de μ.

Vector von Mises. Existen varias extensiones multivariadas para la von Mises, sin embargo, nosotros nos vamos a enfocar en el caso toroidal propuesto por Mardia et al. [2008], llamado distribución von Mises multivariada (MVM)

$$f(\boldsymbol{\theta};\, \boldsymbol{\mu}, \boldsymbol{\kappa}, \Lambda) \propto \exp\left\{\boldsymbol{\kappa}^\top \cos(\boldsymbol{\theta} - \boldsymbol{\mu}) + \sin(\boldsymbol{\theta} - \boldsymbol{\mu})^\top \Lambda \sin(\boldsymbol{\theta} - \boldsymbol{\mu})\right\},$$

donde $\boldsymbol{\mu} \in \mathbb{T}^p$ son las direcciones medias, $\boldsymbol{\kappa} \in \mathbb{R}^p_+$ es el vector de concentraciones marginales, y $\Lambda \in \mathbb{R}^{p\times p}$ es una matriz simétrica con diagonal nula que codifica las interacciones entre componentes, con las operaciones trigonométricas aplicadas componente a componente. Esta formulación se construye a través de una normal $2p$-variada $\mathbf{X}$. De forma análoga a como vimos para la normal proyectada, se agrupa de a 2 formando $(\mathbf{X}_1, \ldots, \mathbf{X}_p)^\top$ y se restringe al hypertoro condicionando por $\|\mathbf{X}_i\| = 1$ para todo $i \in \{1, \ldots, p\}$, esto bajo el supuesto de que cada bloque diagonal 2×2 de la matriz de covarianza asociada es isotrópico. Esto cumple claramente con el criterio de ser intrínseca. Existe una formulación más general llamada mGvM de Navarro et al. [2016].

Algunas propiedades importantes de estas formulaciones de von Mises son que pertenecen a la familia exponencial, que las distribuciones condicionales completas $\Theta_i \mid \boldsymbol{\Theta}_{-i}$ son von Mises univariadas [Mardia et al., 2008], y que son óptimas en el sentido de maximizar la entropía sujeta a los momentos trigonométricos de primer orden fijados [Navarro et al., 2016].

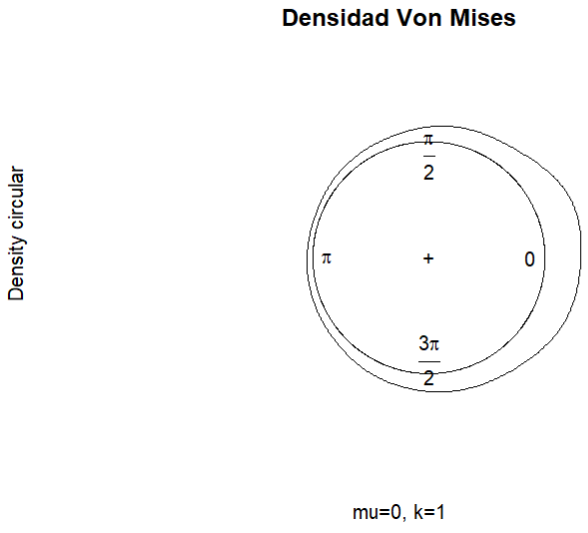


Figura 4: Función de densidad de la distribución von Mises univariada para $\mu = \pi$ y diferentes valores del parámetro de concentración κ. Mayor concentración se observa para valores grandes de κ.

2.10. Distribuciones *envueltas*

2.10.1. Construcción general

La familia de distribuciones *envueltas* se construye a partir de una distribución definida en la recta real mediante una operación de "envoltura" sobre el círculo unitario, la cual se puede interpretar simplemente como una transformación modular. Para una variable aleatoria X definida en $\mathbb{R}$ con distribución F, definimos la variable aleatoria circular

$$\Theta = X \text{ mód } 2\pi,$$

donde la operación módulo mapea $\mathbb{R}$ sobre $[0, 2\pi)$ [Mardia and Jupp, 2000].

.

2.10.2. Caso particular: Distribución Gaussiana-*envuelta*

La distribución Gaussiana-*envuelta* es el ejemplo fundamental de esta familia y ha sido estudiada extensivamente desde el trabajo pionero de Mardia [1975].

Variable Gaussiana-envuelta. Sea $X \sim N(\mu, \sigma^2)$ una variable aleatoria Gaussiana en $\mathbb{R}$. Definimos

$$\Theta = X \text{ mód } 2\pi \sim WN(\mu, \sigma^2),$$

donde WN denota la distribución Gaussiana-*envuelta* con parámetro de media μ (reducido módulo 2π sin pérdida de generalidad) y de escala $\sigma^2 > 0$.

Utilizando la representación con variable latente $X = \Theta + 2\pi K$, con lo cual podemos deducir la función de densidad mediante la marginalización de K, quedando

$$f(\theta; \mu, \sigma^2) = \sum_{k=-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\theta - \mu + 2\pi k)^2}{2\sigma^2}\right\}, \quad \theta \in [0, 2\pi).$$

La suma se puede interpretar como las contribuciones de todas las vueltas alrededor del círculo [Mardia and Jupp, 2000].

Para fines computacionales, esta serie infinita converge rápidamente y puede truncarse. Alternativamente, usando series de Fourier, la densidad puede expresarse como

$$f(\theta; \mu, \sigma^2) = \frac{1}{2\pi}\left[1 + 2\sum_{p=1}^{\infty} \exp\left\{-\frac{p^2\sigma^2}{2}\right\} \cos(p(\theta - \mu))\right],$$

lo cual puede ser una alternativa más eficiente computacionalmente bajo ciertos contextos [Kurz et al., 2014].

Los momentos trigonométricos son

$$\mathbb{E}[e^{ip\Theta}] = \exp\left\{ip\mu - \frac{p^2\sigma^2}{2}\right\},$$

esto permite calcular fácilmente la dirección media y otros estadísticos circulares.

Vector Gaussiano envuelto. A diferencia de los otros tipos de distribuciones presentadas, la extensión al caso toroidal procede de manera natural y directa. Sea $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ un vector aleatorio Gaussiano en $\mathbb{R}^p$. Definimos

$$\boldsymbol{\Theta} = \mathbf{X} \text{ mód } 2\pi \sim WN_p(\boldsymbol{\mu}, \Sigma),$$

donde la operación módulo se aplica componente a componente, resultando en $\boldsymbol{\Theta} \in [0, 2\pi)^p$.

Equivalentemente, usando la representación con variables latentes,

$$\mathbf{X} = \boldsymbol{\Theta} + 2\pi\mathbf{K},$$

donde $\mathbf{K} = (K_1, \ldots, K_p)^\top \in \mathbb{Z}^p$ representa el vector de números de vueltas en cada componente, y $0 \leq \Theta_i < 2\pi$ para todo $i \in \{1, \ldots, p\}$.

La función de densidad conjunta es

$$f(\boldsymbol{\theta}; \boldsymbol{\mu}, \Sigma) = \sum_{\mathbf{k} \in \mathbb{Z}^p} \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu} + 2\pi\mathbf{k})^\top \Sigma^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu} + 2\pi\mathbf{k})\right\},$$

para $\boldsymbol{\theta} \in [0, 2\pi)^p$ [Mardia and Jupp, 2000].

La matriz de covarianza Σ captura las dependencias entre las componentes antes de la envoltura, lo cual induce los patrones de correlación circular del modelo. Los momentos trigonométricos multivariados se calculan mediante

$$\mathbb{E}[e^{i\mathbf{p}^\top \boldsymbol{\Theta}}] = \exp\left\{i\mathbf{p}^\top \boldsymbol{\mu} - \frac{1}{2}\mathbf{p}^\top \Sigma \mathbf{p}\right\},$$

para cualquier vector de enteros $\mathbf{p} \in \mathbb{Z}^p$.

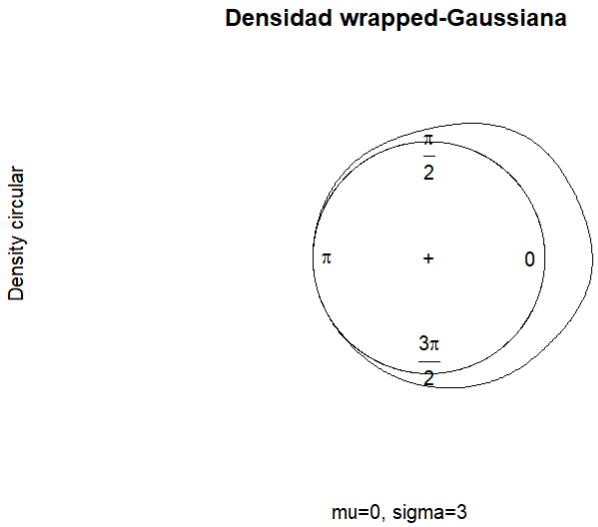


Figura 5: Función de densidad de la distribución Gaussiana-*envuelta* univariada para diferentes valores de σ^2. Para σ^2 pequeño, la distribución se aproxima a una von Mises; para σ^2 grande, se aproxima a la uniforme circular.

2.10.3. Ventajas de la distribución Gaussiana-*envuelta*

Como se adelantó en la introducción, nuestro objeto de estudio se centra en la distribución Gaussiana-*envuelta*. Naturalmente, existen ciertas ventajas en proponer un modelo Gausssiano-envuelto por sobre otros y en esta sección fundamentamos las ventajas posibles, especialmente pensando en el caso que propondremos a futuro.

Lo primero, siguiendo la línea para datos toroidales, usar la extensión multivariada de la Gaussiana-envuelta es natural y directa, por lo que es única, en cambio, tanto para el caso de la normal proyectada o la von Mises, las extensiones para datos toroidales tienen más de alguna propuesta, afectando principalemente las nociones de dependencia. Por ejemplo para el caso de la von Mises se tiene el modelo seno [Mardia et al., 2008] y el modelo coseno [Mardia et al., 2007], ambas con propiedades distintas. Para el caso de la normal proyectada, se tiene las extensiones propuestas en Mastrantonio [2018] y Kato et al. [2025], donde Mastrantonio [2018] es un caso más general pero la interpretación de la dependencia puede ser más ambigua.

Otra ventaja que tiene la Gaussiana-envuelta sobre la formulación de normales proyectadas es la no identificabilidad por escala, esto naturalmente se hereda de forma matricial al caso multivariado, lo cual es claramente un comportamiento no deseado, en cambio, la Gaussiana-envuelta basta con restringir la media al intervalo $[0, 2\pi)$ para evitar problemas de ese tipo.

Con respecto a la von Mises, existen ciertos trabajos que indican que, por comportamientos cualitativos, puede usarse para modelar los mismos tipos de datos que una Gaussiana-envuelta y vice-versa. Además, en la práctica suele ser difícil encontrar una distinguibilidad estadística entre los modelos, especialmente para tamaños muestrales no grandes y/o ciertas

configuraciones de parámetros, por lo que la elección de un modelo u otro suele justificarse principalmente por facilidad de implementación, inferencia y costos computacionales asociados. Los resultados y trabajos sobre esto están basados principalmente en el caso univariado pero teorizamos que las conclusiones se pueden extrapolar al caso multivariado.

Particularmente en Mardia and Jupp [2000] dictan que cuando una $WN(\mu, \sigma^2)$ tiene σ^2 pequeño, entonces $vM(\mu, \kappa)$ con $\kappa \approx 1/\sigma^2$ es una buena aproximación, usando como argumento el teorema del límite central en el círculo. En el trabajo de Collett and Lewis [1981] se usa la razón de verosimilitudes para poder determinar distinguibilidad estadística entre ambos modelos, donde concluyen que para muestras pequeñas son prácticamente indistinguibles, para muestras medianas se obtiene distinguibilidad en casos de concentración κ altos o bajos y para tamaños muestrales más grandes, los estadísticos obtienen cierto poder para detectar diferencias. En Pewsey et al. [2005] se refinó los resultados anteriores con un estudio más robusto y concluyen que los umbrales de tamaño muestral para poder distinguir entre modelos son incluso más grandes.

Los tres modelos para el caso toroidal tienen complicaciones computacionales. Para el caso de la normal proyectada, no hay forma cerrada y como vimos previamente, lo usual sigue expresarlo a través de forma aumentada y formular el problema a través de estadística bayesiana usando algoritmos de muestreo para la inferencia, en particular en Mastrantonio [2018] usan Gibb sampling, esto implica una complejidad computacional importante dado a la necesidad de un esquema de muestreo exhaustivo. Para el caso de la von Mises, tenemos que para datos toroidales, las constantes normalizadoras no tienen forma cerrada y por tanto la verosimilitud tampoco, induciendo complejidad computacional para estimación de máxima verosimilitud. Para la Gaussiana envuelta se tiene forma cerrada, sin embargo, la densidad se tiene que aproximar mediante un truncamineto de la serie [Kurz et al., 2014].

Otra propiedad interesante de la Gaussiana-envuelta es que es cerrada bajo convolución; eso implica que la suma de $WN(\mu_1, \sigma_1^2) + WN(\mu_2, \sigma_2^2) \sim WN(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ según independencia. Esta propiedad implica una ventaja de modelación en una variedad de escenarios, como por ejemplo, series temporales.

Finalmente, otra propiedad interesante de la Gaussiana-envuelta es la conexión directa de la correlación lineal con respecto a la correlación circular. Basándonos en el coeficiente Jammalamadaka-Sarma, se tiene que hay una relación monótona entre ρ_C y ρ_{lin}, con $\rho_{lin} = \sigma_{\alpha\beta}/(\sigma_\alpha \sigma_\beta)$, siendo σ_α^2, σ_β^2 y $\sigma_{\alpha\beta}$ de la matriz de covarianza Σ del Gaussiano latente. De hecho para valores pequeños de σ_α^2 o de σ_β^2 se tiene que $\rho_C \approx \rho_{lin}$ [Jammalamadaka and SenGupta, 2001]

2.10.4. Aproximaciones computacionales mediante truncamiento

En esta sección vamos a revisar las estrategias para el truncamiento en la Gaussiana-envuelta que se han propuesto en la literatura. Este paso es fundamental para la inferencia dado que la densidad al ser una serie, no es viable de implementar directamente en el computador pero también es vital perder el mínimo de información en el proceso.

Para el caso univariado, la densidad puede expresarse en dos representaciones equivalen-

tes, cada una más eficiente según valores de σ^2:

Caso de concentración alta (σ^2 pequeño). Para valores relativamente pequeños de σ^2 , se utiliza la representación directa

$$f(\theta;\mu,\sigma^2) = \sum_{k=-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\theta-\mu+2\pi k)^2}{2\sigma^2}\right\}, \quad \theta \in [0,2\pi),$$

donde se realiza un truncamiento simétrico $k \in \{-M,\ldots,M\}$ con $M \in \mathbb{N}$ suficientemente grande.

La razón de esta eficiencia radica en que las contribuciones de términos con $|k|$ grande decaen exponencialmente rápido cuando σ^2 es pequeño. Hornik and Grün [2013] demuestran que el error de truncamiento es del orden $O(\exp\{-M^2/(2\sigma^2)\})$, garantizando convergencia rápida.

Caso de concentración baja (σ^2 grande). Para valores mayores de σ^2 (típicamente $\sigma^2 \geq 1$), es preferible utilizar la representación de Fourier

$$f(\theta;\mu,\sigma^2) = \frac{1}{2\pi}\left[1 + 2\sum_{p=1}^{\infty} \exp\left\{-\frac{p^2\sigma^2}{2}\right\}\cos(p(\theta-\mu))\right],$$

donde el truncamiento toma la forma $p \in \{1,\ldots,N\}$ con $N \in \mathbb{N}$.

En este régimen, los coeficientes de Fourier decaen exponencialmente con p^2, proporcionando convergencia aún más rápida que en el caso anterior [Jammalamadaka and SenGupta, 2001].

Con esto nos bastaría encontrar criterios de selección para ese punto de truncamiento, Kurz et al. [2014] abordan eso usando criterios cuantitativos que permiten obtener una precisión deseada sin tener que sobrecargar en costos computacionales. Para esto usan los dos casos según σ^2 para expresar la densidad, en el caso de la representación directa se propone usar

$$M = \left\lceil \sqrt{2\sigma^2 \ln\left(\frac{1}{\varepsilon\sqrt{2\pi\sigma^2}}\right)}/(2\pi) \right\rceil,$$

donde ε es la tolerancia de error deseada.

En cambio para la representación alternativa, proponen usar

$$N = \left\lceil \sqrt{2\ln(2/\varepsilon)/\sigma^2} \right\rceil.$$

Estos criterios han sido validados empíricamente y se utilizan ampliamente en implementaciones computacionales modernas [Pewsey et al., 2013].

Para la Gaussiana-*envuelta* multivariada el truncamiento debe realizarse sobre $\mathbb{Z}^p$. Jona-Lasinio et al. [2012] proponen una estrategia basada en los valores propios de Σ, más precisamente, usan

$$f(\boldsymbol{\theta};\boldsymbol{\mu},\Sigma) \approx \sum_{\mathbf{k}\in\mathscr{K}_M} \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu}+2\pi\mathbf{k})^\top\Sigma^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu}+2\pi\mathbf{k})\right\},$$

donde $\mathscr{K}_M = \{\mathbf{k}\in\mathbb{Z}^p : \|\mathbf{k}\|_\infty \leq M\}$ o, más eficientemente, usando una región elipsoidal basada en la métrica de Mahalanobis inducida por Σ^{-1}.

Esta aproximación permite evaluar la densidad con precisión controlada incluso en dimensiones moderadas ($p \leq 10$), lo cual es crucial para aplicaciones de modelado espacial [Wang and Gelfand, 2020a].

2.11. Estadística espacial

La estadística espacial es una rama de la estadística que se ocupa del análisis de datos georeferenciados. La ubicación espacial de las observaciones se suele usar para determinar la estructura de dependencia Cressie [1993], Diggle and Ribeiro [2007].

2.11.1. Motivación

Los datos espaciales aparecen naturalmente en diversas disciplinas científicas: en ciencias ambientales para modelar la distribución de contaminantes Gelfand et al. [2010], en agricultura de precisión para optimizar el rendimiento de cultivos Oliver and Webster [2010], en epidemiología para estudiar la propagación de enfermedades Elliott et al. [2000], y en minería para estimar la distribución de minerales en el subsuelo Chilès and Delfiner [2012].

La característica principal de los datos espaciales es la presencia de *autocorrelación espacial*, lo cual busca emular el fenómeno descrito por la primera ley de la geografía de Tobler: "todas las cosas están relacionadas entre sí, pero las cosas más próximas en el espacio tienen una relación mayor que las distantes" Tobler [1970]. Esta dependencia espacial viola el supuesto de independencia de las observaciones presente en métodos estadísticos clásicos, requiriendo enfoques metodológicos especializados.

2.11.2. Definición

Sea $\mathscr{D} \subset \mathbb{R}^d$ un dominio espacial, típicamente con $d = 2$ o $d = 3$. Un *campo aleatorio* es una colección de variables aleatorias $\{Y(\mathbf{s}) : \mathbf{s} \in \mathscr{D}\}$ indexadas por la ubicación espacial $\mathbf{s}$.

Para cualquier conjunto finito de ubicaciones $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \subset \mathscr{D}$, el vector aleatorio $\mathbf{Y} = (Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n))^\top$ tiene una distribución conjunta que caracteriza el proceso.

Existen varias formas de caracterizar el campo aleatorio, donde suelen usarse como supuestos que facilitan la inferencia en la práctica.

La primera que nombraremos son los **campos aleatorios estrictamente estacionarios**, esto se define si para cualquier conjunto de ubicaciones $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ y cualquier vector de desplazamiento $\mathbf{h}$, la distribución conjunta de $(Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n))$ es idéntica a la de $(Y(\mathbf{s}_1 + \mathbf{h}), \ldots, Y(\mathbf{s}_n + \mathbf{h}))$ Cressie [1993].

Una versión más relajada de lo anterior son los **campos aleatorios débilmente estacionarios**, los cuales se caracterizan por

$$\mathbb{E}[Y(\mathbf{s})] = \mu \quad \forall \mathbf{s} \in \mathscr{D}, \tag{1}$$

$$\mathrm{Cov}(Y(\mathbf{s}), Y(\mathbf{s}')) = C(\mathbf{s} - \mathbf{s}') \quad \forall \mathbf{s}, \mathbf{s}' \in \mathscr{D}, \tag{2}$$

donde $C(\cdot)$ es la función de covarianza que depende solo del vector de separación $\mathbf{h} = \mathbf{s} - \mathbf{s}'$ Stein [1999].

La última caracterización que presentaremos son los **campos aleatorios isotrópicos**, lo cual quiere decir que la función de covarianza asociada depende solamente de la distancia

euclidiana en $\mathscr{D}$, es decir,

$$\text{Cov}(Y(\mathbf{s}), Y(\mathbf{s}')) = C(\|\mathbf{s} - \mathbf{s}'\|) \quad \forall \mathbf{s}, \mathbf{s}' \in \mathscr{D}.$$

Asumiremos a partir de ahora que todo campo aleatorio será débilmente estacionario e isotrópico.

Usualmente se suele asumir la estructura de covarianza, la cual se especifica mediante una familia paramétrica. Funciones de covarianza válidas deben garantizar que $\mathbf{\Sigma}$ sea definida positiva. Ejemplos clásicos incluyen:

- **Exponencial**: $C(\mathbf{h}) = \sigma^2 \exp\left(-\frac{\|\mathbf{h}\|}{\phi}\right)$
- **Gaussiana**: $C(\mathbf{h}) = \sigma^2 \exp\left(-\frac{\|\mathbf{h}\|^2}{\phi^2}\right)$
- **Matérn**: $C(\mathbf{h}) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}\|\mathbf{h}\|}{\phi}\right)^{\nu} K_\nu \left(\frac{\sqrt{2\nu}\|\mathbf{h}\|}{\phi}\right)$

donde σ^2 es la varianza parcial (sill), ϕ el parámetro de rango, ν el parámetro de suavidad, $\Gamma(\cdot)$ la función gamma, y $K_\nu(\cdot)$ la función de Bessel modificada de segundo tipo Stein [1999], Chilès and Delfiner [2012].

Una definición importante dentro de este contexto es el *semivariograma*, el cual se expresa mediante

$$\gamma(\mathbf{h}) = \frac{1}{2}\text{Var}(Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})) = C(\mathbf{0}) - C(\mathbf{h}),$$

donde $\gamma(\mathbf{h})$ cuantifica la disimilitud esperada entre observaciones separadas por $\mathbf{h}$ [Matheron, 1963].

Esto caracteriza el campo aleatorio y suele usarse por temas de interpretabilidad, análisis exploratorios o incluso inferencia.

2.11.3. Campos aleatorios Gaussianos

El campo aleatorio Gaussiano es el ejemplo por excelencia para estos tipos de modelos debido a que están relacionados a la distribución normal, definiéndose a través de que cada colección finita de variables aleatorias sigue una distribución normal multivariada.

De forma más concreta tenemos que el campo aleatorio $\{Y(\mathbf{s}) : \mathbf{s} \in \mathscr{D}\}$ se llama Gaussiano si para cualquier $n \in N$ y ubicaciones $\mathbf{s}_1, \ldots, \mathbf{s}_n \in \mathscr{D}$:

$$\mathbf{Y} = (Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n))^\top \sim \mathscr{N}_n(\boldsymbol{\mu}, \mathbf{\Sigma}),$$

donde $\mu_i = \mathbb{E}[Y(\mathbf{s}_i)]$ y $\Sigma_{ij} = \text{Cov}(\|s_i - s_j\|)$.

Los campos aleatorios Gaussianos son altamente utilizados, especialmente porque heredan varias propiedades de la normal que son útiles y facilitan la modelación,por ejemplo, las distribuciones condicionales y marginales son normales también, por lo que hay tratabilidad analítica [Stein, 1999]. Otra propiedad importante es que está full especificado por la media μ y la función de covarianza $C(\|\mathbf{h}\|)$. También algo que se hereda de la normal es que se maximiza entropía dentro de las distribuciones con la misma media y covarianza especificada [Jaynes, 1957]

3. Campos aleatorios Gaussianos-*envueltos*

3.1. Motivación

En muchas aplicaciones prácticas, los fenómenos espaciales de interés son inherentemente direccionales o angulares. Ejemplos incluyen la dirección del viento en estaciones meteorológicas Breckling [1989], la orientación de fracturas en formaciones geológicas Fisher [1993], patrones de movimiento de animales en estudios ecológicos Lagona et al. [2015], y orientaciones de fibras en materiales compuestos Hernández-Stumpfhauser et al. [2017]. Estas variables toman valores en el círculo $\mathbb{S}^1 = [0, 2\pi)$, donde la periodicidad es una característica fundamental: un ángulo de 0° es idéntico a 360°.

La modelación espacial de datos direccionales presenta desafíos únicos que no pueden abordarse directamente con métodos de estadística espacial clásica. Por ejemplo, al promediar direcciones de 5° y 355°, el promedio aritmético de 180° es completamente erróneo, mientras que el promedio circular correcto es 0°. Esto ilustra que las técnicas deben adaptarse.

Los campos aleatorios Gaussianos-*envueltos* proporcionan un marco riguroso y computacionalmente tratable para modelar datos direccionales espacialmente correlacionados Hernández-Stumpfhauser et al. [2017], Wang and Gelfand [2020b]. Estos procesos heredan propiedades convenientes de los campos Gaussianos mientras respetan la geometría circular del espacio muestral Núñez-Antonio and Gutiérrez-Peña [2018].

3.2. Definición

Sea $\{Y(\mathbf{s}) : \mathbf{s} \in \mathscr{D} \subset \mathbb{R}^d\}$ un campo aleatorio Gaussiano con media $\mu(\mathbf{s})$ y función de covarianza $C(\|h\|)$. El campo aleatorio Gaussiano-*envuelto* $\{\Theta(\mathbf{s}) : \mathbf{s} \in \mathscr{D}\}$ se define mediante la operación módulo como

$$\Theta(\mathbf{s}) = Y(\mathbf{s}) \mod 2\pi,$$

donde la operación módulo se aplica por coordenadas [Mardia and Jupp, 2000].

Para cualquier conjunto finito de ubicaciones $\mathbf{s}_1, \ldots, \mathbf{s}_n \in \mathscr{D}$, el vector $\boldsymbol{\Theta} = (\Theta(\mathbf{s}_1), \ldots, \Theta(\mathbf{s}_n))^\top$ sigue una distribución Gaussiana-*envuelta* multivariada con función de densidad

$$f(\boldsymbol{\theta}) = \sum_{\mathbf{k} \in \mathbb{Z}^n} \phi_n(\boldsymbol{\theta} + 2\pi\mathbf{k}; \boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

donde $\phi_n(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denota la densidad Gaussiana multivariada con vector de medias $\boldsymbol{\mu}$ y matriz de covarianzas $\boldsymbol{\Sigma}$, y $\mathbf{k} = (k_1, \ldots, k_n)^\top$ es un vector de enteros.

3.3. Propiedades de campos aleatorios Gaussianos-*envueltos*

El campo aleatorio Gaussiano-*envuelto* hereda naturalmente varias estructuras del campo Gaussiano latente subyacente, aunque estas deben interpretarse cuidadosamente o ajustarlas apropiadamente para contexto circular.

3.3.1. Variograma

El semivariograma clásico, definido como $\gamma(\mathbf{h}) = \frac{1}{2}\text{Var}(Y(\mathbf{s}+\mathbf{h}) - Y(\mathbf{s}))$, no es directamente aplicable a datos circulares debido a la ambigüedad en la definición de diferencias angulares Mardia and Jupp [2000].

Para datos direccionales, se han propuesto variogramas circulares que respetan la geometría del círculo. Una definición natural utiliza la distancia angular Núñez-Antonio and Gutiérrez-Peña [2018]:

$$\gamma_c(\mathbf{h}) = \frac{1}{2}\mathbb{E}\left[d_c(\Theta(\mathbf{s}+\mathbf{h}), \Theta(\mathbf{s}))^2\right],$$

donde $d_c(\theta_1, \theta_2) = \pi - |\pi - |\theta_1 - \theta_2||$ es la distancia angular más corta en el círculo.

Alternativamente, para procesos Gaussianos-*envueltos* estacionarios, el variograma puede expresarse en términos del proceso latente. Si $Y(\mathbf{s})$ tiene covarianza $C(\mathbf{h})$, entonces Wang and Gelfand [2020b]:

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}) = \sigma^2 - C(\mathbf{h}),$$

proporciona información sobre la estructura de dependencia espacial del proceso latente que induce el comportamiento del proceso circular.

3.3.2. Función característica y momentos trigonométricos

Una propiedad fundamental de la distribución Gaussiana-*envuelta* es su función característica, que para el caso univariado con parámetros μ y σ^2 está dada por Mardia [1975]:

$$\varphi_p(\Theta) = \mathbb{E}[e^{ip\Theta}] = e^{ip\mu - p^2\sigma^2/2}, \quad p \in \mathbb{Z}.$$

Para el caso espacial multivariado, los momentos trigonométricos conjuntos preservan la estructura de dependencia Jammalamadaka and SenGupta [2001]. El momento trigonométrico conjunto de primer orden está dado por:

$$\mathbb{E}[e^{i\Theta(\mathbf{s})}] = e^{i\mu - \sigma^2/2},$$

el cual es constante bajo estacionariedad.

El momento trigonométrico conjunto de segundo orden entre dos ubicaciones es:

$$\mathbb{E}[e^{i(\Theta(\mathbf{s})+\Theta(\mathbf{s}'))}] = e^{2i\mu - \sigma^2 - C(\mathbf{s}-\mathbf{s}')},$$

lo que revela cómo la función de covarianza del proceso latente modula la dependencia circular.

3.4. Correlación circular

Siguiendo lo anterior, hemos definido los campos aleatorios Gaussianos-*envueltos* a través de campos aleatorios Gaussianos latentes; por consiguiente, definiremos la correlación circular

a partir de la correlación usual del campo latente. Para variables direccionales bivariadas no existe una única manera de definir medidas de correlación; Jammalamadaka and SenGupta [2001] presenta una discusión extensa del tema en el capítulo 8.

Nuestra correlación circular que vamos a escoger es la definida por Jammalamadaka and Sarma [1988], la cual cumple con varias propiedades deseables haciéndola así particularmente interesante para modelación espacial. Esta correlación se basa en la correlación del proceso latente y, por tanto, tiene una interpretación natural en términos de su estructura.

Definición 1 (Correlación Circular de Jammalamadaka-Sarma)**.** *Para dos variables direccionales* Θ_1 *y* Θ_2 *provenientes de un proceso Gaussiano*-envuelto *bivariado con vector de medias* $\boldsymbol{\mu} = (\mu, \mu)^\top$ *y matriz de covarianzas*

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{pmatrix},$$

la correlación circular se define como Jammalamadaka and Sarma [1988]:

$$\rho_c(\Theta_1, \Theta_2) = \frac{\sinh(\rho\sigma^2)}{\sinh(\sigma^2)},$$

donde $\sinh(\cdot)$ *denota la función seno hiperbólico.*

Esta medida de correlación posee las propiedades adecuadas de un coeficiente de correlación, como vimos previamente, además de tener una expresión cerrada para este contexto.

Entonces, para un campo aleatorio Gaussiano-*envuelto* estacionario con función de covarianza válida $C(\mathbf{h}) = \sigma^2\rho(\mathbf{h})$, donde $\rho(\mathbf{h})$ es la función de correlación del proceso latente, la función de correlación circular espacial toma la forma:

$$\rho_c(\mathbf{h}) = \frac{\sinh(\sigma^2\rho(\mathbf{h}))}{\sinh(\sigma^2)}.$$

Esto implica que la autocorrelación espacial circular hereda propiedades de la autocorrelación espacial lineal, donde también es isotrópica y cumple con comportamientos límites de $\rho_C(0) = 1$ correlación perfecta en misma ubicación.

De manera más concreta, podemos ver cómo se comporta la correlación circular comparada con la correlación lineal del proceso latente mediante la Figura 6 para distintas configuraciones de parámetros para un modelo Matérn.

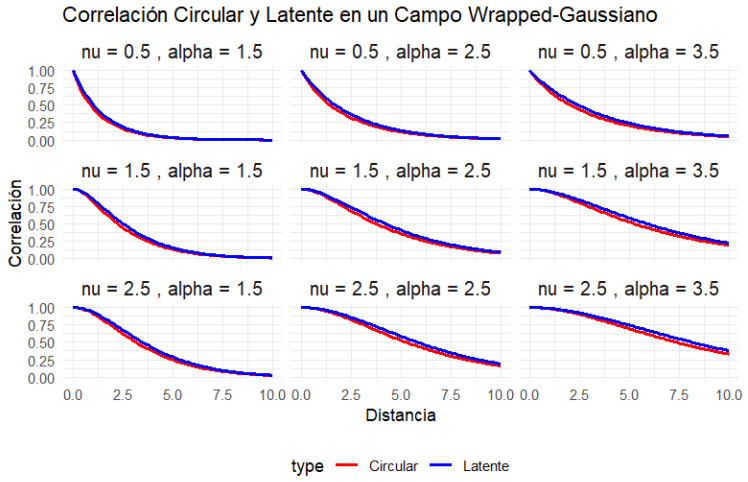


Figura 6: Comparación entre correlación lineal del proceso Gaussiano latente (ρ) y correlación circular (ρ_c) para diferentes modelos Matérn

4. Verosimilitud compuesta

La motivación de la verosimilitud compuesta es el buscar una alternativa basada en verosimilitud razonable y con propiedades interesantes a la hora de que el cálculo de la verosimilitud usual no se puede llevar en la práctica ya sea debido a la lentitud o por imposibilidad del problema de optimización dado los costos computacionales asociados.

Esta técnica se basa en aproximar la verosimilitud clásica usando combinaciones de verosimilitudes restringidas a subconjuntos de la muestra. Esto es una propuesta razonable con respecto al sentido del problema dado que usa la información de la muestra como verosimilitud, pero de manera restringida y la combina, por lo que se podría decir que es un enfoque del estilo "dividir y conquistar", lo que es altamente usado para algoritmos en informática.[Cormen et al., 2009, Cap. 4]

Al segmentar los datos, bajamos la dimensionalidad del problema, lo cual ayuda a evitar posibles integraciones de dimensión alta o cálculos matriciales de gran tamaño. Estas problemáticas pueden ser recurrentes dentro del contexto de máxima verosimilitud dado que es común que densidades se definan a través de matrices inversas o por integrales, de hecho el caso Gaussiano cae en la primera problemática, por tanto los costos computacionales suelen mejorar bajo esta perspectiva.

La verosimilitud compuesta ha tomado harto interés dentro del contexto de estadística espacial debido a que la complejidad de sus modelos suelen crecer muy rápido con respecto al tamaño de la muestra, de hecho, a pesar de que la máxima verosimilitud suele ser el "gold standar", esta complejidad o infactibilidad numérica provoca buscar otras formas de llevar a cabo la estimación. Una forma clásica de estimación en estadística espacial es mínimos cuadrados usando el variograma [Cressie, 1993], lo cual suele ser bastante más eficiente computacionalmente.

Otro posible beneficio de este enfoque es que en algunos casos se pueden evitar cálculos

difíciles mediante obtener analíticamente las densidades en una expresión conveniente computacionalmente. Esto de hecho es lo que se hace en el trabajo de Alegría [2024], donde usan la expresión de las densidades más pequeñas sin inversión de matrices u otras operaciones matriciales, lo cual lleva a una mayor eficiencia en código.

De manera más concreta, definimos la verosimilitud compuesta o CL, como su sigla en inglés, de la forma

$$CL(\theta) = \sum_{i=1}^{K} w_i \, l_i(\theta),$$

donde $l_i(\theta)$ corresponde a la log-verosimilitud del segmento de datos i-ésimo y w_i es un peso escalar con respecto a la log-verosimilitud del segmento i-ésimo, el cual se usa para mejorar eficiencia y/o también reducir costos computacionales.

Podemos notar que esto corresponde a una log-verosimilitud ponderada bajo el supuesto de independencia entre segmentos de datos, lo cual por lo general no es cierto y, por tanto, CL suele definir un problema mal especificado; sin embargo, este estimador tiene ciertas propiedades interesantes.

4.1. Inferencia de verosimilitud compuesta

4.1.1. Matriz de información de Godambe

Como dijimos anteriormente, el estimador de verosimilitud compuesta es tal que maximiza nuestra función de verosimilitud compuesta respectiva, es decir, $\widehat{\boldsymbol{\phi}} = \text{argmax}_{\boldsymbol{\phi}} \text{CL}(\boldsymbol{\phi}; \boldsymbol{x})$. Asumiendo suficiente regularidad de CL, definimos $u(\phi; x) = \nabla_\phi CL(\phi; x)$ y también las matrices de sensibilidad y variabilidad como

$$H(\phi) = E_\phi \left\{ -\nabla_\phi u(\phi; Y) \right\} = \int \left\{ -\nabla_\phi u(\phi; y) \right\} f(y; \phi) dy$$

y

$$J(\phi) = \text{var}_\phi \left\{ u(\phi; Y) \right\},$$

respectivamente. Recordemos que la verosimilitud compuesta está mal especificada debido al supuesto de independencia entre las verosimilitudes que forman nuestra función objetivo, lo que implica que no es realmente una función de verosimilitud, bajo esto, H y J difieren, siendo esto un hecho importante dado que en el contexto de verosimilitud clásica, estas coinciden.

Con esto, la matriz de información de Fisher necesita ser sustituida por la matriz de información de Godambe (Godambe, 1960) dada por

$$G(\phi) = H(\phi)J(\phi)^{-1}H(\phi),$$

también conocida como la matriz de información tipo "sandwich".

4.1.2. Resultados Asintóticos

En el caso de n observaciones independientes e idénticamente distribuidas $Y_1, \ldots, Y_n$ del modelo $f(y;\phi)$ en $\mathbb{R}^m$, y $n \to \infty$ con m fijo, se tiene un resultado presentado en Varin et al. [2011b].

bajo condiciones de regularidad en las densidades logarítmicas componentes tenemos un teorema central del límite para la estadística de puntuación de verosimilitud compuesta, que lleva al resultado que el estimador de máxima verosimilitud compuesta, $\hat{\theta}_{CL}$, está asintóticamente distribuido de manera normal:

$$\sqrt{n}(\hat{\theta}_{CL} - \theta) \xrightarrow{d} N_p\{0, G^{-1}(\theta)\},$$

donde $N_p(\mu, \Sigma)$ es la distribución normal p-dimensional con media y varianza como se indicó, y $G(\theta)$ es la matriz de información de Godambe en una sola observación.

Estos resultados asintóticos son muy importantes dado que sirven para cuantificar la incertidumbre de la estimación, además de que permite hacer test de hipótesis, regiones de confianza, etc.

4.1.3. Insesgado y consistente

Se satisfacen ciertas propiedades importantes que son que la función de estimación es insesgada y que el estimador es Fischer-consistente. Recordaremos las definiciones de estos conceptos, primero partamos con lo que es ser un estimador insesgado.

Sea $\{P_\theta : \theta \in \Theta\}$ una familia de distribuciones de probabilidad y sea $g(\theta; Y)$ una función medible en la observación Y y suficientemente diferenciable en el parámetro θ. Llamaremos $g(\theta; Y)$ una *función de estimación* para θ si, para cada muestra observada Y, la ecuación

$$g(\theta; Y) = 0$$

puede utilizarse para definir un estimador de θ (por ejemplo, como una raíz de dicha ecuación). Decimos que $g(\theta; Y)$ es una *función de estimación insesgada* si, para todo $\theta \in \Theta$,

$$\mathbb{E}_\theta\big[g(\theta; Y)\big] \;=\; 0.$$

En otras palabras, la esperanza de la función de estimación se anula en el verdadero valor del parámetro, lo que garantiza que la ecuación $g(\theta; Y) = 0$ es, en promedio, correcta en el modelo verdadero [Godambe, 1960].

Ahora definiremos la consistencia de Fischer [Heritier et al., 2009]. Sea $\{P_\theta : \theta \in \Theta\}$ una familia de distribuciones de probabilidad indexadas por el espacio paramétrico y sea T un funcional que asigna a cada distribución F un valor en el espacio de parámetros, es decir $T : \mathscr{F} \to \Theta$. Decimos que T es *consistente en el sentido de Fisher* (o *Fisher-consistente*) para el parámetro θ si, para todo $\theta \in \Theta$,

$$T(F_\theta) = \theta,$$

donde F_θ denota la función de distribución asociada a P_θ.

En términos de un estimador muestral, si $\hat{\theta}_n = T(\hat{F}_n)$ donde $\hat{F}_n$ es la distribución empírica de una muestra de tamaño n, la consistencia de Fisher exige que el funcional límite T recupere exactamente el parámetro verdadero cuando se evalúa en la distribución poblacional F_θ, esto es, $T(F_\theta) = \theta$.

En palabras más simples, esto quiere decir que aplicar el estimador a la distribución de probabilidad real, llevaría a obtener el θ real que la define.

La demostración que CL usa una función de estimación insesgada se presenta a continuación:

Sea $\{P_\theta : \theta \in \Theta\}$ una familia de modelos y considérese una verosimilitud compuesta de la forma

$$CL(\theta; y) = \prod_{k \in \mathscr{K}} f_k(y; \theta)^{w_k},$$

donde $f_k(y; \theta)$ denota una densidad (marginal o condicional) asociada al componente k y $w_k > 0$ son pesos. La log-verosimilitud compuesta es

$$\ell_c(\theta; y) = \log CL(\theta; y) = \sum_{k \in \mathscr{K}} w_k \log f_k(y; \theta),$$

y el score compuesto viene dado por

$$U_c(\theta; y) = \frac{\partial}{\partial \theta} \ell_c(\theta; y) = \sum_{k \in \mathscr{K}} w_k \, U_k(\theta; y),$$

con $U_k(\theta; y) = \partial \log f_k(y; \theta)/\partial\theta$ el score del componente k.

Bajo condiciones de regularidad estándar y suponiendo que el modelo está bien especificado con parámetro verdadero θ_0, cada score parcial satisface

$$\mathbb{E}_{\theta_0}\big[U_k(\theta_0; Y)\big] = 0,$$

puesto que

$$\mathbb{E}_{\theta_0}[U_k(\theta_0; Y)] = \int \frac{\partial}{\partial \theta} \log f_k(y; \theta_0) \, f_k(y; \theta_0) \, \mathrm{d}y = \frac{\partial}{\partial \theta} \int f_k(y; \theta) \, \mathrm{d}y \bigg|_{\theta=\theta_0} = \frac{\partial}{\partial \theta} 1 = 0.$$

Por linealidad de la esperanza se obtiene inmediatamente

$$\mathbb{E}_{\theta_0}\big[U_c(\theta_0; Y)\big] = \sum_{k \in \mathscr{K}} w_k \, \mathbb{E}_{\theta_0}[U_k(\theta_0; Y)] = 0.$$

Así, $U_c(\theta; Y)$ define una función de estimación insesgada y el estimador de verosimilitud compuesta, definido como solución de $U_c(\theta; Y) = 0$, es consistente y asintóticamente normal bajo supuestos regulares.

Para la demostración de ser Fischer consistente, [Varin et al., 2011b] lo aborda mediante un argumento basado en la divergencia de Kullback-Leibler compuesta.

4.2. Formas de verosimilitud compuesta

Para esta sección vamos a asumir que estamos dentro del contexto de estadística espacial, donde todos nuestros datos van a estar indexados a una locación en el espacio $s_i \in \mathbb{R}^n$, lo cual nos permite heredar del espacio su sentido de proximidad usando la métrica que hereda, como podría ser la distancia Euclidiana por ejemplo. ϕ lo denotaremos como el vector de parámetros asociado de aquí en adelante.

4.2.1. Verosimilitud por parejas

La verosimilitud por parejas es una de las formas más utilizadas de verosimilitud compuesta para inferencia en modelos espaciales. En la práctica, suele ser una alternativa atractiva por el *trade-off* entre eficiencia estadística y costo computacional, ya que requiere únicamente verosimilitudes bivariadas, evitando el manejo directo de matrices de covarianza de gran dimensión. Además, su formulación es relativamente sencilla y se adapta con facilidad a distintos tipos de datos y estructuras de dependencia.

Como su nombre sugiere, la verosimilitud por parejas (pairwise likelihood, PL) se construye a partir de todas (o de un subconjunto) de las parejas de observaciones en la muestra $X(s_1), \ldots, X(s_n)$. Denotando por $\ell_{ij}(\phi)$ el logaritmo de la verosimilitud bivariada asociada al par $(X(s_i), X(s_j))$ bajo el modelo parametrizado por ϕ, la verosimilitud compuesta por parejas se define como

$$\mathrm{PL}(\phi) \;=\; \sum_{i<j} w_{ij}\, \ell_{ij}(\phi),$$

donde $\{w_{ij}\}$ son pesos que permiten seleccionar y ponderar las parejas incluidas. En el contexto espacial, es común escoger $w_{ij} \in \{0, 1\}$ dependiendo de una cota de distancia $d_s > 0$; por ejemplo, si $d(s_i, s_j) < d_s$ se fija $w_{ij} = 1$ y, en caso contrario, $w_{ij} = 0$. De este modo, sólo contribuyen aquellas parejas suficientemente cercanas en el espacio, lo que reduce el costo computacional y limita la contribución de pares con interacción prácticamente nula.

Este tipo de construcción ha sido ampliamente utilizado en estadística espacial para campos Gaussianos y modelos relacionados. Por ejemplo, Varin et al. [2011a] discuten el uso de verosimilitudes compuestas, incluyendo la verosimilitud por parejas, como herramienta general para modelos de dependencia. En el contexto de procesos Gaussianos espaciales, Eidsvik et al. [2014a] emplean verosimilitudes compuestas para ajustar modelos jerárquicos de gran dimensión, mientras que Bevilacqua et al. [2012] y Joe and Xu [1996] utilizan enfoques de verosimilitud por parejas para covarianzas no estacionarias y modelos espaciales flexibles. Estos trabajos ilustran cómo la verosimilitud por parejas permite realizar inferencia de forma

eficiente, especialmente en contextos donde la verosimilitud completa resulta inviable por los costos computacionales.

4.2.2. Verosimilitud por bloques

La verosimilitud por bloques es una extensión natural de la verosimilitud por parejas, la cual sigue dentro del contexto de verosimilitud compuesta. La idea básica consiste en particionar la muestra espacial en L bloques $B_1, \ldots, B_L$, de modo que cada bloque *clusters* de observaciones cercanas en el espacio. A continuación, se consideran las distribuciones conjuntas asociadas a pares de bloques, que capturan una dependencia de orden superior a la bivariada, pero manteniendo matrices de covarianza de dimensión moderada.

Sea $\mathbf{X} = (X(\mathbf{s}_1), \ldots, X(\mathbf{s}_n))^\top$ el vector de observaciones y denotemos por $\mathbf{X}_i$ el subvector correspondiente al bloque B_i, $i = 1, \ldots, L$. Para cada pareja de bloques (i, j), definimos $\mathbf{X}_{ij} = (\mathbf{X}_i^\top, \mathbf{X}_j^\top)^\top$ de dimensión n_{ij} y su verosimilitud conjunta $\ell_{ij}(\mathbf{x}_{ij}; \boldsymbol{\phi})$ bajo el modelo paramétrico considerado. La *verosimilitud compuesta por bloques* se define entonces como

$$\mathrm{BCL}(\boldsymbol{\phi}; \mathbf{x}) = \sum_{1 \le i < j \le L} w_{ij}\, \ell_{ij}(\boldsymbol{\phi}; \mathbf{x}_{ij}), \tag{3}$$

donde w_{ij} son los pesos que determinan qué parejas de bloques contribuyen a la función objetivo, análogo al caso de parejas y $\ell_{ij}(\boldsymbol{\phi}; \mathbf{x}_{ij})$ es la log-verosimilitud del par de bloques (i, j).

Para la elección de los pesos w_{ij} como 0 o 1, se utiliza un criterio en función de una medida de proximidad entre bloques, por ejemplo, la distancia entre sus centroides. Esta elección reduce de manera importante el costo computacional. [Caragea and Smith, 2007, Eidsvik et al., 2014b].

El estimador de verosimilitud por bloques de $\boldsymbol{\phi}$ se obtiene como

$$\widehat{\boldsymbol{\phi}}_{\mathrm{BCL}} = \arg\max_{\boldsymbol{\phi}} \mathrm{BCL}(\boldsymbol{\phi}; \mathbf{x}).$$

En términos de compromiso entre eficiencia estadística y costo computacional, el tamaño de los bloques juega un rol central: al aumentar el número de observaciones dentro de cada bloque, las densidades $f_{n_{ij}}$ contienen más información sobre la estructura de dependencia subyacente, lo que en general mejora la eficiencia de los estimadores, aunque a costa de trabajar con matrices de covarianza de mayor dimensión. En el extremo opuesto, si cada bloque contiene una única observación, se recupera la verosimilitud compuesta por parejas como caso particular [Bevilacqua et al., 2012].

Este enfoque ha sido aplicado de manera exitosa en distintos contextos de estadística espacial. Por ejemplo, Caragea and Smith [2007] estudian aproximaciones tipo bloque para campos Gaussianos en grillas regulares, mientras que Eidsvik et al. [2014b] desarrollan un esquema de verosimilitud compuesta por bloques para la estimación y predicción en grandes conjuntos de datos espaciales georreferenciados. Más recientemente, Mohammadian Mosammam and Mohammadi [2019] emplean verosimilitud compuesta por bloques para estimar

funciones de covarianza espaciales, y Acosta et al. [2021] usan un enfoque de verosimilitud por bloques para estudiar el tamaño de muestra efectivo en grandes bases de datos espaciales.

En el contexto específico de este trabajo, la estructura general en (3) se especificará más adelante al caso de campos Gaussianos-*envueltos*.

4.3. Algoritmo K-means para la formación de bloques

Nuestro algoritmo de generación de bloques se basa en K-means, donde usaremos como entrada los sitios de realización del campo aleatorio.

Este algoritmo permite controlar el número de *clusters* que se quieren formar; sin embargo, no controla de manera directa el tamaño de los bloques. Aun así, es posible influir indirectamente en dicho tamaño: con un mayor número de *clusters* se esperan bloques más pequeños, mientras que con un menor número de *clusters* se esperan bloques más grandes. Para evitar ambigüedades, es recomendable elegir K como un porcentaje del número total de sitios.

La formulación del algoritmo K-means surge del objetivo de separar los datos en K grupos, con K previamente fijado. Denotemos los *clusters* como $C_1, \cdots, C_K$ y las observaciones como $\mathbf{x}_1, \cdots, \mathbf{x}_n$, con n el tamaño muestral. Entonces, los *clusters* cumplen con

- $$\bigcup_{i=1}^{K} C_i = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$$
- $$C_i \cap C_j = \emptyset, \quad \forall i \neq j.$$

es decir, los grupos abordan todos los datos pero no se solapan.

Para determinar estos *clusters* buscamos que sean coherentes, en el sentido de que los datos dentro de cada *cluster* se parezcan entre sí; equivalentemente, que dentro de cada *cluster* haya poca variabilidad, lo cual se cuantifica mediante una función W.

Por lo tanto se tiene un problema de optimización de la forma

$$\min_{C_1,\cdots,C_K} \sum_{i=1}^{K} W(C_i)$$

donde así penalizamos la variabilidad en cada grupo. Una elección clásica para este W es

$$W(C_k) = \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2,$$

siendo $|C_k|$ la cardinalidad del cluster k-ésimo y x_{ij} es la coordenada j-ésima del dato i-ésimo, por lo que que es la suma de las distancias euclidianas de todas las parejas dentro del cluster ponderado por el inverso de su cardinalidad.

Finalmente, para formar los clusters, el algoritmo de forma iterativa calcula los promedios o çentroides"de los clusters previos y luego se hace una nueva asignación según a que centroide está más cercano el dato i-ésimo. Este algoritmo necesita inicializarse de manera aleatoria y además asegura convergencia a un mínimo local del problema. James et al. [2021]

Un ejemplo de la aplicación de este algoritmo para la generación de bloques dentro del contexto de datos espaciales, donde se utiliza las ubicaciones en el plano para determinación se puede ver en Figura 7

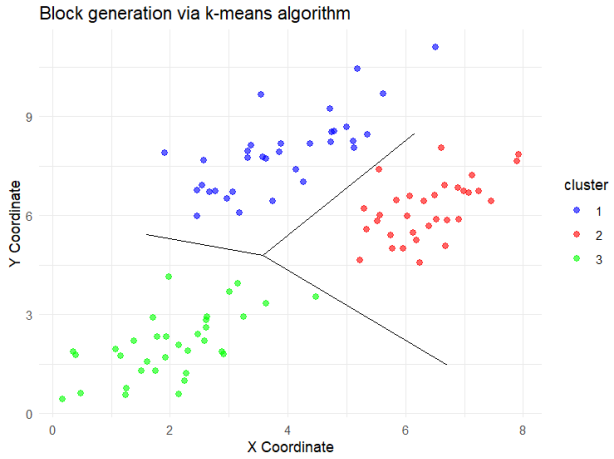


Figura 7: Generación de bloques vía k-means

5. Estimación distribucional de los estimadores CL mediante Bootstrap paramétrico

Para los resultados asintóticos vistos en la sección anterior, se tiene que para obtener la distribución límite, se necesita la computación de la matriz de información de Godambe, a su vez de también la matriz de sensibilidad J y la matriz Hessiana H por consecuencia. La implicación de estos cálculos puede traer un desafío computacional importante a medida que el tamaño de muestra n crezca, o también al trabajar con distribuciones con funciones de densidad bastante complejas.

En particular, para una distribución Gaussiana-*envuelta*, la función de densidad es sumamente compleja. Por ello, calcular la matriz Hessiana y la matriz de sensibilidad se vuelve inviable, lo que impide obtener de forma explícita la distribución asintótica. Sin embargo, mediante técnicas de bootstrap podemos estudiar el comportamiento de los estimadores, como se menciona en Efron and Tibshirani [1986].

Primero revisamos cómo funciona el método bootstrap. El bootstrap es un algoritmo que se basa en simular nuevas muestras a partir de la distribución F de los datos observados.

Por lo general se utiliza la distribución empírica, ya que no se conoce la distribución verdadera. En nuestro caso, no es necesario un bootstrap no paramétrico, dado que asumimos un modelo paramétrico asociado a un campo aleatorio Gaussiano-*envuelto*.

En consecuencia, podemos simular desde el campo aleatorio Gaussiano-*envuelto* considerado e insertar nuestras estimaciones $\hat{\alpha}, \hat{\sigma}, \hat{\mu}$. De este modo obtenemos un remuestreo de campos aleatorios Gaussianos-*envueltos* de forma efectiva y coherente.

Por lo tanto, Consideremos un campo aleatorio $Y(s)$, donde $s \in D$ representa la ubicación espacial en el dominio D, y las observaciones del campo en n ubicaciones espaciales $s_1, s_2, \ldots, s_n$ son $Y(s_1), Y(s_2), \ldots, Y(s_n)$.

Para generar B muestras bootstrap del campo aleatorio $Y(s)$, denotamos las muestras remuestreadas como:

$$Y_b^*(s_i), \quad b = 1, 2, \ldots, B, \quad i = 1, 2, \ldots, n$$

Donde:

- $Y_b^*(s_i)$ es la b-ésima muestra bootstrap del valor del campo aleatorio en la ubicación s_i,
- B es el número total de replicaciones bootstrap,
- n es el número de ubicaciones espaciales observadas.

Por lo tanto, el conjunto completo de B campos remuestreados se puede expresar como:

$$\{Y_b^*(s_i) : b = 1, 2, \ldots, B, \quad i = 1, 2, \ldots, n\}$$

De aquí podemos aplicar el método de estimación de parámetros escogido y aplicarlo para cada sampleo de campo aleatorio, obteniendo así muestras de nuestros estadísticos de estimación de la forma

$$\{\hat{\phi}_b = (\hat{\alpha}_b, \hat{\sigma}_b, \hat{\mu})_b : \quad b = 1, \cdots, B\}$$

Con esta muestra podemos obtener estimaciones de la media y la covarianza del estimador

$$\hat{\mu}_B^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\phi}_b$$

$$\hat{\Sigma}_B^* = \frac{1}{B-1}\sum_{b=1}^{B}(\hat{\phi}_b - \hat{\mu}_B^*)(\hat{\phi}_b - \hat{\mu}_B^*)^\top$$

También tenemos que para B grande la distribución empírica es razonablemente similar a la distribución real, por lo que cobra sentido y es interesante el calcular regiones de confianza para nuestro estimador.

La fórmula general para una región de confianza elipsoidal es:

$$(\hat{\mu}_B^* - \phi)^\top(\hat{\Sigma}_B^*)^{-1}(\hat{\mu}_B^* - \phi) \leq \chi^2_{p,1-\alpha}$$

Donde $\chi^2_{p,1-\alpha}$ es el percentil de la distribución χ^2 con $p = 3$ grados de libertad y nivel de confianza $1 - \alpha$.

6. Propuesta

En esta tesis proponemos un nuevo enfoque de inferencia para campos aleatorios Gaussianos-*envueltos* basado en verosimilitudes compuestas por bloques. El objetivo principal es obtener estimadores de los parámetros del campo latente que no dependan de simulaciones de Monte Carlo intensivas, como ocurre en enfoques ya existentes.

Por un lado, Fisher and Lee [1994] desarrollan métodos de inferencia para series de tiempo circulares usando algoritmos EM, lo cual requiere de simulaciones para aproximar integrales que surgen naturalmente del algoritmo. Por otro lado, Jona-Lasinio et al. [2012] proponen un enfoque jerárquico bayesiano para campos Gaussianos-*envueltos* espaciales, donde la inferencia se realiza mediante algoritmos MCMC, con un costo computacional elevado y dependiente de simulaciones. En contraste, nuestro objetivo es construir estimadores basados en verosimilitudes compuestas que sólo requieran evaluar densidades en dimensión moderada, evitando así procedimientos iterativos basados en Monte Carlo.

Como punto de referencia, consideramos el trabajo de Alegría et al. [2016], donde se propone un esquema de verosimilitud compuesta por parejas para campos Gaussianos-*envueltos* espacio–temporales. Definimos tal método como nuestro *benchmark* de comparación. Sobre la base de este trabajo, proponemos una extensión mediante verosimilitudes compuestas por bloques, que permite incorporar información de órdenes superiores a la bivariada con la idea de mejorar la eficiencia estadística sin alcanzar el costos computacional muy altos.

En toda la tesis trabajaremos con función de covarianza de tipo Matérn,

$$C(\mathbf{h};\alpha,\sigma^2) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\|\mathbf{h}\|}{\alpha}\right)^{\nu} K_\nu\left(\frac{\|\mathbf{h}\|}{\alpha}\right),$$

donde $\boldsymbol{\phi} = (\alpha, \sigma^2)^\top$ es el vector de parámetros de correlación y varianza respectivamente, y $\nu > 0$ es un parámetro de suavidad que asumiremos fijado a priori. Esta elección es estándar en estadística espacial y permite interpretar los parámetros en términos de rango y variabilidad. En el contexto direccional, varios de estos parámetros admiten una interpretación natural en términos de concentración angular, como se detalla más adelante.

6.1. Verosimilitud compuesta por parejas para campos Gaussianos-*envueltos*

Sea $\{X(\mathbf{s}) : \mathbf{s} \in D\}$ un campo aleatorio Gaussiano real con media constante μ y covarianza $C(\cdot;\boldsymbol{\phi})$ de tipo Matérn. Las observaciones direccionales se definen como

$$Y(\mathbf{s}) = X(\mathbf{s}) \text{ mód } 2\pi \in [0, 2\pi),$$

de modo que el vector de datos en n ubicaciones $\mathbf{s}_1, \ldots, \mathbf{s}_n$ viene dado por $\mathbf{Y} = (Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n))^\top$.

Para un conjunto de observaciones $\mathbf{y} \in [0, 2\pi)^n$, la densidad Gaussiana-*envuelta* n-dimensional puede escribirse como

$$f_n(\mathbf{y};\boldsymbol{\mu},\Sigma) = \sum_{\mathbf{k}\in\mathbb{Z}^n} g_n(\mathbf{y} - \boldsymbol{\mu} + 2\pi\mathbf{k};\Sigma), \qquad 0 \le y_i < 2\pi, \quad i = 1, \ldots, n,$$

donde $g_n(\cdot;\Sigma)$ denota la densidad normal n-dimensional con media cero y matriz de covarianza Σ, y $\boldsymbol{\mu} = \mu\mathbf{1}_n$.

La verosimilitud compuesta por parejas se construye a partir de las densidades bivariadas asociadas a cada par de la muestra. Para $1 \leq i < j \leq n$, denotamos por

$$\mathbf{Y}_{ij} = (Y(\mathbf{s}_i), Y(\mathbf{s}_j))^\top$$

y por $f_2(\cdot;\boldsymbol{\mu}_{ij}, \Sigma_{ij}(\boldsymbol{\phi}))$ la densidad Gaussiana-*envuelta* bivariada correspondiente, donde $\boldsymbol{\mu}_{ij} = (\mu,\mu)^\top$ y $\Sigma_{ij}(\boldsymbol{\phi})$ es la submatriz 2×2 de la matriz de covarianza latente. La verosimilitud compuesta por parejas (en versión logarítmica) se define como

$$\mathrm{PL}(\boldsymbol{\theta};\mathbf{y}) = \sum_{1\leq i<j\leq n} w_{ij} \ln f_2\big(\mathbf{y}_{ij};\boldsymbol{\mu}_{ij},\Sigma_{ij}(\boldsymbol{\phi})\big), \tag{4}$$

donde $\mathbf{y}_{ij}$ son las observaciones en el par (i,j), $\boldsymbol{\theta} = (\boldsymbol{\phi},\mu)^\top$ es el vector de parámetros, y los pesos $w_{ij} \in \{0,1\}$ permiten seleccionar únicamente aquellas parejas de ubicaciones que se consideran suficientemente cercanas.

Desde el punto de vista computacional, la expresión de f_2 involucra una suma en $\mathbb{Z}^2$, la cual se aproxima en la práctica mediante un truncamiento adecuado del conjunto de índices, reduciendo el esfuerzo de cálculo sin comprometer la precisión de las verosimilitudes pequeñas. El método de Alegría et al. [2016] se basa precisamente en maximizar (4) para obtener estimadores de $\boldsymbol{\theta}$, constituyendo nuestro punto de partida.

6.2. Verosimilitud compuesta por bloques para campos Gaussianos-*envueltos*

Como alternativa y extensión de (4), proponemos emplear una verosimilitud compuesta por bloques. Para ello, particionamos el conjunto de ubicaciones $\{\mathbf{s}_1,\ldots,\mathbf{s}_n\}$ en L bloques disjuntos $B_1,\ldots,B_L$, de manera que cada bloque contenga observaciones que sean cercanas en el espacio. Denotamos por $\mathbf{Y}_i$ el subvector de observaciones en el bloque B_i y por

$$\mathbf{Y}_{ij} = (\mathbf{Y}_i^\top, \mathbf{Y}_j^\top)^\top$$

el vector formado por la concatenación de los bloques i y j, de dimensión n_{ij}. La densidad Gaussiana-*envuelta* correspondiente viene dada por

$$f_{n_{ij}}(\mathbf{y}_{ij};\boldsymbol{\mu}_{ij},\Sigma_{ij}(\boldsymbol{\phi})) = \sum_{\mathbf{k}\in\mathbb{Z}^{n_{ij}}} g_{n_{ij}}(\mathbf{y}_{ij} - \boldsymbol{\mu}_{ij} + 2\pi\mathbf{k};\Sigma_{ij}(\boldsymbol{\phi})),$$

donde $\boldsymbol{\mu}_{ij}$ es el vector de medias (constante) asociado a los dos bloques y $\Sigma_{ij}(\boldsymbol{\phi})$ es la submatriz de covarianza del campo latente restringida a los sitios de $B_i \cup B_j$.

La verosimilitud compuesta por bloques propuesta se define mediante

$$\mathrm{BCL}(\boldsymbol{\theta};\mathbf{y}) = \sum_{1\leq i<j\leq L} w_{ij} \log f_{n_{ij}}\big(\mathbf{y}_{ij};\boldsymbol{\mu}_{ij},\Sigma_{ij}(\boldsymbol{\phi})\big), \tag{5}$$

donde los pesos $w_{ij} \in \{0,1\}$ determinan qué parejas de bloques contribuyen a la función objetivo. En la práctica, la suma sobre $\mathbf{k} \in \mathbb{Z}^{n_{ij}}$ en $f_{n_{ij}}$ se aproxima mediante un conjunto truncado $M_{ij} \subset \mathbb{Z}^{n_{ij}}$, de forma análoga al caso bivariado pero permitiendo bloques de mayor dimensión. El estimador de verosimilitud compuesta por bloques se define como

$$\widehat{\boldsymbol{\theta}}_{\mathrm{BCL}} = \arg\max_{\boldsymbol{\theta}} \mathrm{BCL}(\boldsymbol{\theta}; \mathbf{y}).$$

Este enfoque se puede interpretar como una extensión de la verosimilitud compuesta por parejas, en el sentido de que los bloques permiten capturar dependencias de orden superior, manteniendo un costo computacional intermedio entre la verosimilitud por parejas y la verosimilitud clásica completa.

6.3. Casos extremos y papel del tamaño de bloque

La verosimilitud por bloques (5) permite recuperar, como casos extremos, tanto la verosimilitud compuesta por parejas como la verosimilitud clásica.

- **Bloques de tamaño uno.** Si cada observación individual define un bloque, es decir, $B_i = \{\mathbf{s}_i\}$ para $i = 1, \ldots, n$, entonces la combinación de dos bloques (i, j) coincide con el par de observaciones $\mathbf{Y}_{ij} = (Y(\mathbf{s}_i), Y(\mathbf{s}_j))^\top$. En este caso, $n_{ij} = 2$ y la expresión (5) se reduce exactamente a la verosimilitud compuesta por parejas (4).

- **Un único bloque global.** En el extremo opuesto, si consideramos todos los datos como un único bloque $B_1 = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ y fijamos $L = 1$, la suma en (5) desaparece y sólo queda un término que coincide con la log-verosimilitud completa del campo Gaussiano-*envuelto*, basada en la densidad $f_n(\cdot; \boldsymbol{\mu}, \Sigma(\boldsymbol{\phi}))$. En este caso, el enfoque deja de ser compuesto y recuperamos la verosimilitud clásica.

Para tamaños de bloque intermedios (por ejemplo, bloques de tamaño 2, 3 o 4), la función objetivo (5) se sitúa naturalmente entre estos dos extremos, tanto en términos de información estadística como de costo computacional. A medida que el tamaño de los bloques aumenta, se espera una mayor eficiencia estadística, en el sentido de que la verosimilitud compuesta se aproxima más a la verosimilitud completa, aunque a costa de trabajar con matrices de covarianza de mayor dimensión y con sumas en conjuntos de indexación de mayor cardinalidad. Por el contrario, bloques muy pequeños (cercanos al caso de parejas) ofrecen menor costo computacional, pero incorporan menos información de orden superior.

En la práctica, el uso de bloques de tamaño moderado (por ejemplo, de 2 o 3 observaciones) puede ofrecer un equilibrio atractivo: se obtiene una mejora sobre la verosimilitud por parejas en términos de sesgo y varianza de los estimadores, sin llegar a costos associados cercanos a la verosimilitud clásica.

6.4. Modelo latente y parametrización direccional

Finalmente, discutimos la parametrización de los parámetros del campo latente en el contexto direccional. Dado que la densidad Gaussiana-*envuelta* es invariante frente a desplazamientos de 2π en la media, resulta natural trabajar con una versión circular de ésta. En efecto, si definimos $\tilde{\mu} = \mu$ mód 2π, se tiene

$$f_n(\mathbf{y}; \boldsymbol{\mu}, \Sigma) = f_n(\mathbf{y}; \tilde{\boldsymbol{\mu}}, \Sigma),$$

donde $\boldsymbol{\mu} = \mu \mathbf{1}_n$ y $\tilde{\boldsymbol{\mu}} = \tilde{\mu}\, \mathbf{1}_n$. En consecuencia, la función objetivo (ya sea (4) o (5)) es invariante al reemplazar μ por cualquier representante equivalente módulo 2π.

Con el fin de realizar la optimización sin restricciones, parametrizamos la media circular a través de una transformación desde la recta real,

$$\tilde{\mu} = 2 \arctan(\mu) + \pi,$$

de modo que para todo $\mu \in \mathbb{R}$ se obtiene $\tilde{\mu} \in (0, 2\pi)$s. Durante la inferencia se trabaja sobre $\mu \in \mathbb{R}$, mientras que la interpretación final se realiza en términos de la media circular $\tilde{\mu}$.

En cuanto a la varianza del campo latente, resulta útil interpretarla en términos de una medida de concentración angular. Para una normal *envuelta* univariada con varianza σ^2, la concentración direccional puede definirse como

$$c = \exp\left(-\frac{\sigma^2}{2}\right) \in (0, 1],$$

de modo que valores pequeños de σ^2 se asocian a altas concentraciones (alrededor de la media circular), mientras que valores grandes de σ^2 producen distribuciones casi uniformes en $[0, 2\pi)$. Esta reparametrización respeta el comportamiento esperado de una medida de concentración y permite interpretar los resultados en términos direccionales de manera más directa.

En capítulos posteriores se presentará un estudio de simulación extensivs para evaluar el desempeño del estimador basado en verosimilitud compuesta por bloques, comparándolo con el enfoque de verosimilitud por parejas de Alegría et al. [2016]. Este estudio permitirá analizar el efecto del tamaño de bloque y la elección de los pesos espaciales en el conjunto de eficiencia estadística y costo computacional.

7. Simulaciones

7.1. Diseño del estudio de simulación

Para comparar nuestra metodología basada en bloques con el método de verosimilitud por parejas, diseñamos un estudio de simulación exhaustivo que evalúa el rendimiento de ambos enfoques en diversos escenarios. El estudio considera diferentes configuraciones de parámetros y estructuras de covarianza, permitiendo una comparación justa y sistemática.

7.1.1. Configuración espacial y parámetros

El estudio se desarrolla en el cuadrado unitario $[0,1]^2$, donde generamos grillas irregulares con un tamaño muestral fijo de $n = 100$ ubicaciones. Para el campo aleatorio Gaussiano latente, fijamos el parámetro de media $\mu = \pi$ en todos los casos. Con respecto a los parámetros de covarianza tenemos:

- **Función de covarianza**: Utilizamos la familia Matérn con tres valores de suavidad:
  - $\nu = 0{,}5$
  - $\nu = 1{,}5$
  - $\nu = 2{,}5$
- **Parámetro de rango**:
  - $\alpha = 0{,}10$
  - $\alpha = 0{,}15$
  - $\alpha = 0{,}20$
- **Parámetro de varianza**: Fijamos $\sigma^2 = 1$

La parametrización de las funciones de covarianza se realiza de modo que el parámetro α represente la distancia mínima a la cual la correlación entre dos sitios cae por debajo de 0,05. Esta interpretación práctica define α como la distancia más allá de la cual dos sitios dejan de estar significativamente correlacionados.

Las funciones de covarianza específicas son:

- **Matérn $\nu = 0{,}5$ (Exponencial)**

$$C(h) = \sigma^2 \exp\left(-\frac{3h}{\alpha}\right)$$

- **Matérn $\nu = 1{,}5$**

$$C(h) = \sigma^2 \exp\left(-\frac{4{,}7619h}{\alpha}\right)\left(1 + \frac{4{,}7619h}{\alpha}\right)$$

- **Matérn** $\nu = 2{,}5$

$$C(h) = \sigma^2 \exp\left(-\frac{5{,}91865h}{\alpha}\right)\left(1 + \frac{5{,}91865h}{\alpha} + \frac{(5{,}91865h)^2}{3\alpha^2}\right)$$

7.1.2. Configuración de los métodos de estimación

Para el método de verosimilitud por parejas, utilizamos la implementación estándar sin modificaciones adicionales.

Para los métodos basados en bloques, consideramos cinco configuraciones diferentes que representan porcentajes variables de los datos:

- **Modelo BLOCK 50**: 50 clusters (50 % de los datos)
- **Modelo BLOCK 60**: 60 clusters (60 % de los datos)
- **Modelo BLOCK 70**: 70 clusters (70 % de los datos)
- **Modelo BLOCK 80**: 80 clusters (80 % de los datos)
- **Modelo BLOCK 90**: 90 clusters (90 % de los datos)

El criterio de proximidad para la formación de bloques se varia mediante los umbrales $ds = 0{,}10, 0{,}15, 0{,}20$, que representan la distancia máxima permitida entre centros de clusters del algoritmo k-means para que dos bloques se consideren en la función objetivo.

Para mantener la viabilidad computacional mientras se obtienen aproximaciones significativas, utilizamos el criterio de truncación propuesto por Jona-Lasinio et al. [2012], con $k \in \{-1, 0, 1\}^{n_{\text{block}}}$. Según Jona-Lasinio et al. [2012], esta truncación es adecuada cuando $\sigma^2 \leq 2$, lo cual se cumple en nuestro caso al fijar $\sigma^2 = 1$, asegurando así que las aproximaciones mantengan precisión estadística sin incrementar excesivamente el costo computacional.

Cada escenario se diferencia en tener alguna de las 3 funciones de covarianza propuestas y en el valor del parámetro de rango asociado dentro de los 3 propuestos. Para cada uno de estos escenarios (9 casos) definimos simular 200 campos aleatorios Gaussianos-envueltos.

Luego para cada simulación, se aplican los distintos métodos, donde cada método se define según si es por parejas o bloques, por el umbral ds para los pesos (3 casos) y por la cantidad de bloques si corresponde. En resumen son 18 métodologías de estimación distintas.

Este diseño resulta en un total de $32,400$ escenarios evaluados, proporcionando una base sólida para la comparación estadística de los métodos.

La evaluación del rendimiento se basa en:

- Error absoluto medio (MAE) para σ^2, α y μ
- Tiempo promedio de ejecución

Resultados de las simulaciones

Los resultados del estudio se presentan mediante tablas comparativas que resumen las métricas de rendimiento promedio para todos los escenarios considerados, organizadas por modelo de estimación. Para facilitar la interpretación visual, se implementa un esquema de codificación por colores donde:

- Los **mejores valores** (menores valores de MAE o tiempo de ejecución) se destacan en color azul
- Los **peores valores** se resaltan en color rojo

Este formato permite identificar rápidamente los modelos con mejor rendimiento para cada métrica y configuración específica.

7.1.3. Especificaciones

El código completo utilizado para generar las simulaciones, implementar los algoritmos de estimación y producir los resultados presentados está disponible públicamente en el repositorio de GitHub:

`https://github.com/Joaquin-Aguirre-Adaros/Block-Likelihood-Wrapped-Gaussian-Fields`

Cuadro 1: Especificaciones técnicas de la instancia computacional

Componente	**Especificación**
Proveedor	vast.ai
CPU	Intel Xeon E7-8890 v4 (190 cores)
GPU	2x NVIDIA RTX 3060
VRAM	24 GB
Almacenamiento	300 GB SSD
Versión de R	4.3.2
Paquetes principales	mvtnorm, parallel

Para manejar la intensa carga computacional de este estudio (32,400 escenarios de estimación), se empleó una instancia en la nube proporcionada por vast.ai (Tabla 1), que permitió ejecutar las simulaciones en paralelo, acelerando así el tiempo de cómputo global.

La optimización de las funciones de verosimilitud se realizó mediante el algoritmo Nelder-Mead Nelder and Mead [1965], implementado en la función `optim()` del paquete `stats` de R R Core Team [2023]. La configuración específica utilizada fue:

```
optim(inic, pcl_block, method = "Nelder-Mead",
      control = list(reltol = 1e-16, maxit = 1e4, trace = FALSE))
```

Esta configuración proporciona una tolerancia relativa muy estricta y se permiten suficientes iteraciones posibles para asegurar que al algoritmo convergió y que los valores obtenidos tengan el menor ruido posible de error numérico asociado.

El algoritmo Nelder-Mead fue seleccionado por su robustez para problemas de optimización no lineal sin restricciones y su capacidad para manejar funciones objetivo no diferenciables Lagarias et al. [1998].

Cuadro 2: Resultados del estudio de simulación para modelo PAREJAS

ds	Cov.	Config.	MAE_{σ^2}	MAE_{α}	MAE_{μ}	Tiempo (s)
0.10	1	1	**0.1421**	**0.0288**	**0.1287**	75.99
		2	**0.1638**	0.0403	**0.1558**	62.46
		3	**0.1803**	**0.0502**	**0.1781**	30.70
	2	1	**0.1557**	0.0169	0.1368	83.83
		2	0.1567	0.0216	0.1517	56.60
		3	**0.2090**	**0.0313**	**0.1760**	32.71
	3	1	0.1501	**0.0151**	**0.1360**	88.44
		2	**0.1764**	**0.0205**	**0.1541**	54.24
		3	**0.2058**	**0.0308**	0.1906	35.04
0.15	1	1	**0.1327**	**0.0268**	0.1225	166.63
		2	**0.1454**	0.0377	0.1457	117.72
		3	0.1665	0.0458	**0.1669**	65.77
	2	1	0.1387	0.0168	**0.1314**	152.91
		2	0.1452	**0.0215**	0.1385	128.00
		3	0.1929	**0.0313**	**0.1700**	**66.60**
	3	1	0.1316	**0.0144**	**0.1327**	151.04
		2	**0.1629**	**0.0202**	0.1439	**124.03**
		3	**0.1924**	**0.0294**	0.1776	**69.21**
0.20	1	1	**0.1256**	**0.0264**	0.1179	280.03
		2	**0.1377**	0.0364	**0.1399**	204.18
		3	0.1645	0.0467	0.1627	**110.48**
	2	1	**0.1332**	0.0165	**0.1270**	**277.74**
		2	**0.1448**	**0.0212**	**0.1371**	**191.24**
		3	**0.1871**	**0.0317**	**0.1701**	**115.62**
	3	1	**0.1275**	0.0144	0.1298	**271.61**
		2	0.1583	0.0202	**0.1411**	**210.41**
		3	**0.1798**	**0.0291**	**0.1776**	**113.18**

Nota: MAE_θ denota el Error Absoluto Medio para el parámetro θ, calculado como $\text{MAE}_\theta = \frac{1}{n}\sum_{i=1}^{n}|\hat{\theta}_i - \theta_{\text{true}}|$. Tiempo promedio de estimación en segundos. Config. 1: $\sigma^2 = 1$, $\alpha = 0{,}10$, $\mu = \pi$; Config. 2: $\sigma^2 = 1$, $\alpha = 0{,}15$, $\mu = \pi$; Config. 3: $\sigma^2 = 1$, $\alpha = 0{,}20$, $\mu = \pi$. Cov. = Estructura de covarianza (1, 2 o 3).

Cuadro 3: Resultados del estudio de simulación para modelo BLOCK (nb = 90)

ds	Cov.	Config.	MAE_{σ^2}	MAE_{α}	MAE_{μ}	Tiempo (s)
0.10	1	1	0.1485	0.0317	0.1312	**182.03**
		2	0.1725	**0.0390**	0.1603	**136.84**
		3	0.1839	**0.0502**	0.1834	**75.41**
	2	1	0.1573	0.0170	**0.1345**	**195.72**
		2	0.1577	0.0194	**0.1510**	**128.28**
		3	0.2114	0.0298	0.1826	**82.61**
	3	1	**0.1478**	0.0142	0.1389	**195.90**
		2	0.1860	0.0196	0.1552	**152.07**
		3	0.2093	0.0284	**0.1852**	**99.08**
0.15	1	1	0.1338	0.0310	**0.1188**	**320.19**
		2	0.1492	0.0364	**0.1450**	**221.15**
		3	**0.1625**	**0.0445**	0.1676	**131.58**
	2	1	**0.1366**	**0.0180**	0.1263	**314.41**
		2	0.1429	0.0192	0.1372	**229.19**
		3	0.1920	0.0286	0.1706	**137.50**
	3	1	**0.1279**	0.0138	0.1344	**333.43**
		2	0.1659	0.0190	**0.1372**	**237.26**
		3	0.1914	0.0261	**0.1713**	**148.47**
0.20	1	1	0.1269	**0.0338**	**0.1139**	**513.84**
		2	0.1404	**0.0399**	0.1389	**323.23**
		3	**0.1600**	0.0473	0.1632	**197.26**
	2	1	0.1311	**0.0204**	0.1231	**483.30**
		2	0.1406	0.0211	0.1353	**331.33**
		3	0.1847	0.0294	0.1680	**200.16**
	3	1	0.1246	**0.0159**	**0.1300**	**494.36**
		2	**0.1604**	**0.0210**	**0.1333**	**336.38**
		3	0.1774	0.0260	0.1704	207.27

Nota: MAE_{θ} denota el Error Absoluto Medio para el parámetro θ, calculado como $\text{MAE}_{\theta} = \frac{1}{n}\sum_{i=1}^{n}|\hat{\theta}_i - \theta_{\text{true}}|$. Tiempo promedio de estimación en segundos. Config. 1: $\sigma^2 = 1$, $\alpha = 0{,}10$, $\mu = \pi$; Config. 2: $\sigma^2 = 1$, $\alpha = 0{,}15$, $\mu = \pi$; Config. 3: $\sigma^2 = 1$, $\alpha = 0{,}20$, $\mu = \pi$. Cov. = Estructura de covarianza (1, 2 o 3).

Cuadro 4: Resultados del estudio de simulación para modelo BLOCK (nb = 80)

ds	Cov.	Config.	MAE_{σ^2}	MAE_{α}	MAE_{μ}	Tiempo (s)
0.10	1	1	0.1546	0.0309	0.1419	121.83
		2	0.1790	0.0395	0.1707	85.36
		3	0.1971	0.0509	0.1929	53.38
	2	1	0.1682	0.0166	0.1349	123.62
		2	**0.1535**	**0.0191**	0.1594	108.94
		3	0.2151	0.0282	0.1920	77.80
	3	1	0.1537	**0.0132**	0.1449	147.67
		2	0.1865	0.0186	0.1634	134.65
		3	0.2176	0.0265	0.1967	86.33
0.15	1	1	0.1357	**0.0311**	0.1243	221.36
		2	0.1486	0.0372	0.1457	165.81
		3	0.1691	0.0453	0.1698	98.13
	2	1	0.1406	0.0169	0.1245	244.92
		2	**0.1372**	**0.0181**	**0.1358**	165.33
		3	**0.1892**	0.0263	0.1728	111.88
	3	1	0.1288	0.0129	0.1334	243.63
		2	0.1631	0.0176	0.1426	204.06
		3	0.1902	0.0244	0.1748	139.75
0.20	1	1	0.1274	0.0333	0.1181	371.57
		2	0.1401	0.0379	**0.1370**	267.84
		3	0.1612	**0.0478**	0.1611	172.73
	2	1	0.1321	0.0176	0.1189	374.30
		2	0.1325	0.0194	0.1295	284.59
		3	0.1790	0.0271	0.1668	192.48
	3	1	**0.1206**	0.0144	0.1285	420.16
		2	0.1534	0.0182	0.1363	296.16
		3	0.1740	0.0240	**0.1688**	198.65

Nota: MAE_θ denota el Error Absoluto Medio para el parámetro θ, calculado como $\text{MAE}_\theta = \frac{1}{n}\sum_{i=1}^{n}|\hat{\theta}_i - \theta_{\text{true}}|$. Tiempo promedio de estimación en segundos. Config. 1: $\sigma^2 = 1$, $\alpha = 0{,}10$, $\mu = \pi$; Config. 2: $\sigma^2 = 1$, $\alpha = 0{,}15$, $\mu = \pi$; Config. 3: $\sigma^2 = 1$, $\alpha = 0{,}20$, $\mu = \pi$. Cov. = Estructura de covarianza (1, 2 o 3).

Cuadro 5: Resultados del estudio de simulación para modelo BLOCK (nb = 70)

ds	Cov.	Config.	MAE_{σ^2}	MAE_{α}	MAE_{μ}	Tiempo (s)
0.10	1	1	0.1655	0.0315	0.1561	102.75
		2	0.1914	0.0401	0.1863	66.39
		3	0.2110	0.0509	0.2165	42.74
	2	1	0.1813	**0.0163**	0.1466	113.78
		2	0.1666	0.0200	0.1864	91.09
		3	0.2236	**0.0280**	0.2064	62.42
	3	1	0.1700	0.0133	0.1612	138.25
		2	0.2043	**0.0181**	0.1819	97.65
		3	0.2229	0.0255	0.2256	65.00
0.15	1	1	0.1377	0.0310	0.1250	177.11
		2	0.1525	**0.0362**	0.1474	121.12
		3	0.1728	0.0458	0.1796	77.38
	2	1	0.1428	0.0160	**0.1230**	189.36
		2	0.1376	0.0185	0.1458	169.77
		3	0.1894	0.0251	0.1720	96.60
	3	1	0.1313	0.0126	0.1347	212.42
		2	0.1658	0.0164	0.1458	187.70
		3	**0.1869**	0.0228	0.1836	108.96
0.20	1	1	0.1288	0.0325	0.1176	328.77
		2	0.1443	0.0367	0.1349	224.28
		3	0.1617	0.0473	0.1641	137.82
	2	1	0.1326	0.0164	**0.1174**	330.11
		2	0.1319	0.0193	0.1327	243.00
		3	**0.1738**	0.0255	0.1610	140.30
	3	1	0.1214	0.0139	0.1276	339.01
		2	0.1543	0.0170	0.1344	263.60
		3	0.1700	0.0226	0.1712	169.49

Nota: MAE_θ denota el Error Absoluto Medio para el parámetro θ, calculado como $\text{MAE}_\theta = \frac{1}{n}\sum_{i=1}^{n}|\hat{\theta}_i - \theta_{\text{true}}|$. Tiempo promedio de estimación en segundos. Config. 1: $\sigma^2 = 1$, $\alpha = 0{,}10$, $\mu = \pi$; Config. 2: $\sigma^2 = 1$, $\alpha = 0{,}15$, $\mu = \pi$; Config. 3: $\sigma^2 = 1$, $\alpha = 0{,}20$, $\mu = \pi$. Cov. = Estructura de covarianza (1, 2 o 3).

Cuadro 6: Resultados del estudio de simulación para modelo BLOCK (nb = 60)

ds	Cov.	Config.	MAE_{σ^2}	MAE_{α}	MAE_{μ}	Tiempo (s)
0.10	1	1	**0.1888**	0.0319	**0.1742**	68.54
		2	**0.2127**	0.0454	**0.2033**	47.64
		3	**0.2357**	**0.0584**	**0.2352**	29.34
	2	1	**0.2001**	**0.0178**	**0.1673**	77.37
		2	0.1863	0.0228	**0.1963**	65.08
		3	0.2446	0.0291	**0.2320**	46.13
	3	1	0.1777	0.0144	**0.1703**	100.29
		2	**0.2293**	0.0199	**0.2020**	79.20
		3	**0.2509**	0.0268	**0.2526**	46.79
0.15	1	1	0.1446	0.0300	**0.1314**	141.00
		2	**0.1574**	**0.0378**	**0.1536**	124.84
		3	0.1847	**0.0485**	**0.1872**	63.58
	2	1	**0.1460**	0.0157	0.1295	157.22
		2	0.1446	0.0197	**0.1473**	141.52
		3	0.1969	**0.0238**	**0.1832**	85.87
	3	1	0.1351	0.0124	0.1373	180.09
		2	**0.1740**	**0.0158**	**0.1529**	156.66
		3	0.1889	0.0217	0.1927	98.61
0.20	1	1	0.1282	0.0308	**0.1202**	326.56
		2	**0.1463**	0.0375	0.1361	283.36
		3	**0.1676**	0.0477	**0.1650**	159.60
	2	1	0.1328	**0.0153**	0.1182	340.22
		2	0.1326	0.0193	0.1303	279.40
		3	0.1756	0.0237	0.1642	187.77
	3	1	0.1207	0.0129	0.1279	349.99
		2	0.1585	0.0161	0.1409	333.08
		3	0.1708	0.0209	0.1734	**221.74**

Nota: MAE_θ denota el Error Absoluto Medio para el parámetro θ, calculado como $\text{MAE}_\theta = \frac{1}{n}\sum_{i=1}^{n}|\hat{\theta}_i - \theta_{\text{true}}|$. Tiempo promedio de estimación en segundos. Config. 1: $\sigma^2 = 1$, $\alpha = 0{,}10$, $\mu = \pi$; Config. 2: $\sigma^2 = 1$, $\alpha = 0{,}15$, $\mu = \pi$; Config. 3: $\sigma^2 = 1$, $\alpha = 0{,}20$, $\mu = \pi$. Cov. = Estructura de covarianza (1, 2 o 3).

Cuadro 7: Resultados del estudio de simulación para modelo BLOCK (nb = 50)

ds	Cov.	Config.	MAE$_{\sigma^2}$	MAE$_{\alpha}$	MAE$_{\mu}$	Tiempo (s)
0.10	1	1	0.1861	**0.0322**	0.1577	**44.44**
		2	0.2048	**0.0470**	0.1919	**32.44**
		3	0.2300	0.0571	0.2159	**19.47**
	2	1	0.1878	**0.0178**	0.1580	**57.19**
		2	**0.1869**	**0.0238**	0.1916	**46.10**
		3	**0.2464**	0.0307	0.2264	**31.27**
	3	1	**0.1835**	0.0144	0.1651	**69.82**
		2	0.2115	0.0201	0.1948	**50.25**
		3	0.2462	**0.0252**	0.2376	**32.96**
0.15	1	1	**0.1469**	0.0295	0.1286	**113.97**
		2	0.1554	0.0377	0.1454	**72.50**
		3	**0.1864**	0.0461	0.1775	**44.72**
	2	1	0.1437	**0.0155**	0.1248	**118.37**
		2	**0.1505**	0.0197	0.1437	**99.96**
		3	**0.1991**	0.0245	0.1814	66.67
	3	1	**0.1368**	**0.0120**	**0.1394**	**132.96**
		2	0.1731	**0.0158**	0.1499	134.98
		3	0.1887	**0.0204**	**0.1956**	75.03
0.20	1	1	**0.1294**	0.0289	0.1171	**242.70**
		2	0.1441	**0.0346**	**0.1325**	**193.11**
		3	0.1616	**0.0430**	0.1580	116.79
	2	1	**0.1259**	**0.0153**	0.1175	291.88
		2	**0.1315**	**0.0188**	**0.1283**	233.88
		3	0.1754	**0.0232**	**0.1605**	156.44
	3	1	0.1236	**0.0118**	**0.1258**	336.71
		2	**0.1510**	**0.0155**	0.1369	278.39
		3	**0.1670**	**0.0196**	0.1761	187.95

Nota: MAE$_\theta$ denota el Error Absoluto Medio para el parámetro θ, calculado como MAE$_\theta = \frac{1}{n}\sum_{i=1}^{n}|\hat{\theta}_i - \theta_{\text{true}}|$. Tiempo promedio de estimación en segundos. Config. 1: $\sigma^2 = 1$, $\alpha = 0{,}10$, $\mu = \pi$; Config. 2: $\sigma^2 = 1$, $\alpha = 0{,}15$, $\mu = \pi$; Config. 3: $\sigma^2 = 1$, $\alpha = 0{,}20$, $\mu = \pi$. Cov. = Estructura de covarianza (1, 2 o 3).

7.1.4. Observaciones

El análisis de los resultados revela patrones generales que permiten comprender el comportamiento de los modelos y determinar el método de estimación más adecuado según el escenario. Se destacan las siguientes observaciones:

- **Efecto del umbral (ds)**:
  - Umbrales más altos (ej. ds=0.20) generalmente mejoran la precisión de las estimaciones pero incrementan sustancialmente el tiempo de convergencia.
  - Para umbrales bajos (ds=0.10), el método de parejas y los modelos de bloques con mayor número de clusters (nb=80, 90) muestran mejor rendimiento en precisión.
  - A medida que aumenta el umbral, los métodos basados en bloques mejoran su rendimiento, especialmente aquellos con menor número de clusters (nb=50, 60).
- **Efecto de la estructura de covarianza**:
  - Estructuras de covarianza más suaves (Cov. 2 y 3) inducen mayores errores de estimación en comparación con la estructura exponencial (Cov. 1).
  - Para la estructura de covarianza 1 (exponencial), el método de verosimilitud por parejas presenta excelente rendimiento en todos los umbrales.
  - Para estructuras de covarianza 2 y 3 con umbrales medianos o altos (ds=0.15, 0.20), los métodos de bloques generalmente proporcionan mejores resultados.
- **Efecto del número de bloques**:
  - El número óptimo de bloques depende del contexto, especialmente según umbral elegido y rango asociado.
  - Los modelos con mayor número de clusters (nb=80, 90) suelen ser los más lentos computacionalmente.
  - El modelo de bloques con 50 clusters y umbral ds=0.20 muestra el mejor rendimiento general, especialmente para las estructuras de covarianza 2 y 3.
- **Comparación entre métodos**:
  - El método de parejas generalmente es más rápido que los de bloques, excepto para casos con menor número de bloques (nb=50) y umbral pequeño (ds=0.10), donde los bloques son más rápidos pero con menor precisión.
  - Los métodos de bloques muestran una tendencia a estimar mejor el parámetro de rango (α), incluso en umbrales pequeños donde su desempeño general es inferior.
- **Comportamiento de parámetros**:
  - Rangos más grandes (α=0.20) inducen una convergencia más rápida en todos los modelos.
  - El error en la estimación del rango tiende a aumentar con estructuras de covarianza más complejas y con valores de rango más grandes.

8. Aplicación

Los códigos utilizados para esta sección se pueden encontrar en la parte de `application` dentro del repositorio de este trabajo: `https://github.com/Joaquin-Aguirre-Adaros/Block-Likelihood-Wrapped-Gaussian-Fields`.

Para esta sección se utilizó un conjunto de datos disponible públicamente a través del paquete de R `rWind` [Fernández-López et al., 2021a, Fernández-López and Schliep, 2019]. En particular, se emplea el objeto de ejemplo `wind.data`, el cual se carga en R mediante:

```
library(rWind)
data(wind.data)
```

El objeto `wind.data` corresponde a un ejemplo de datos de viento obtenido con la función `wind.dl` para la Península Ibérica el 12 de febrero de 2015 a las 00:00 (UTC), y contiene una lista con un `data.frame` de 651 observaciones y 7 variables: tiempo (UTC), latitud, longitud, las componentes horizontales del viento (`ugrd10m` y `vgrd10m`, en $\mathrm{m\,s^{-1}}$), además de dirección (`dir`) y velocidad (`speed`) [Fernández-López et al., 2021b].

La función `wind.dl` descarga datos del Global Forecast System (GFS) de la National Weather Service (NOAA/NCEP), con resolución espacial de 0.5° (aprox. 50 km) y viento reportado a 10 m sobre la superficie [Fernández-López et al., 2021c].

El analizar este tipo de datos tiene un sentido más allá de ajustar un modelo estadístico. El viento es un fenómeno que afecta varios procesos, como por ejemplo las turbinas eólicas, también pueden influir en el transporte como para veleros o aviones y también tiene implicaciones meteorológicas, por tanto modelar estos tipos de datos de forma correcta puede llevar a análisis, conclusiones y decisiones en variedad donde proponer un modelo y una técnica de inferecnia adecuada es fundamental.

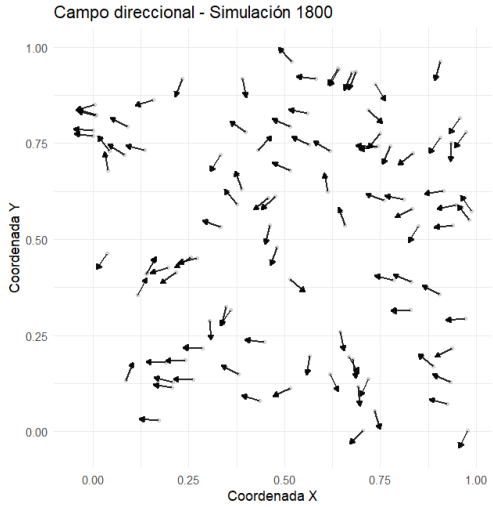
El enfoque que se usó es restringir los datos a un subespacio donde se pueda suponer, mediante simple inspección, que hay un comportamiento aleatorio intrínseco y que usar un modelo de campo Gaussiano-*envuelto* puede ser razonable. Luego se transforman los datos al cuadrado $[0,1] \times [0,1]$ para poder comparar con resultados obtenidos en el estudio de simulación, aplicando estimación por parejas y bloques, junto con su desviación estándar dada por bootstrap.

8.1. Análisis descriptivo

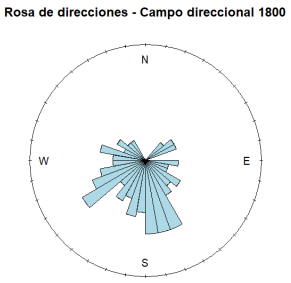
En esta sección utilizaremos un caso representativo de nuestro estudio de simulación con el fin de ilustrar patrones esperables bajo un campo Gaussiano-*envuelto*. La idea es contar con una referencia empírica para interpretar el comportamiento de los datos reales y, en particular, evaluar si resulta razonable asumir que el proceso observado puede ser descrito mediante un modelo Gaussiano-*envuelto*. Este paso es útil para reducir el riesgo de mala especificación al contrastar, al menos tratando de hacer un match cualitativo entre los datos observados y el modelo propuesto.

Al graficar un escenario puntual de la simulación (Figura 8), se aprecia que el campo direccional presenta una dirección predominante. La mayor parte de las observaciones se concentra en torno a una dirección principal, mientras que direcciones aproximadamente opuestas son poco frecuentes. Sin embargo, el campo no es determinista ya que aparecen desviaciones locales y comportamientos más erráticos, consistentes con un proceso de carácter aleatorio, donde ocasionalmente pueden observarse direcciones que no "calzan" perfectamente con sus vecindades más cercanas. Por lo tanto, no se espera un patrón suave como el de un modelo puramente determinista, como podría ser la solución de una ecuación diferencial.

Esta impresión se refuerza con el gráfico de rosas. La masa principal se ubica alrededor de la dirección predominante, y la probabilidad asignada a direcciones opuestas es pequeña (Figura 8). En consecuencia, para evaluar coherencia cualitativa, se debiese poder apreciar cierto grados de aleatoriedad local pero con una sola moda de manera concentrada.



(a) Campo direccional simulado (caso seleccionado).



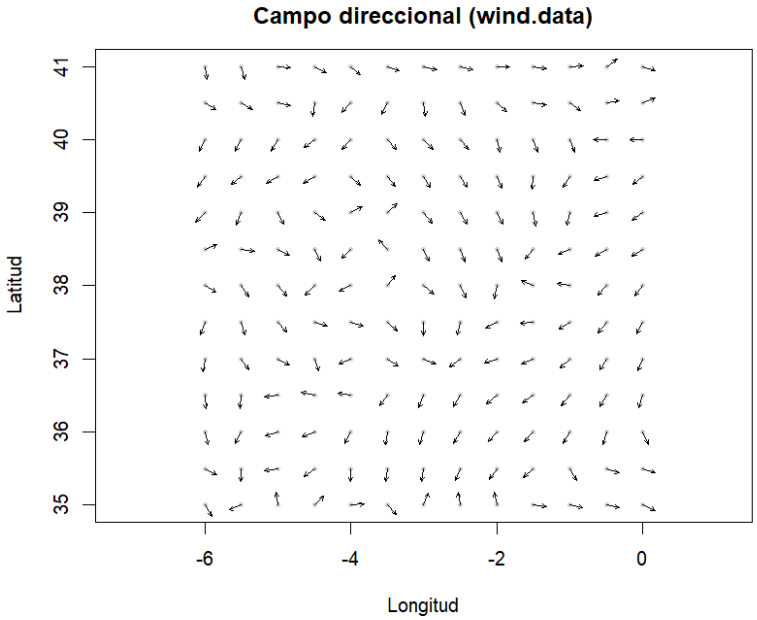
(b) Gráfico de rosas para el mismo caso.

Figura 8: Diagnóstico descriptivo en un caso de simulación Gaussiano-*envuelto*. Se observa una dirección predominante con variabilidad local y baja frecuencia de direcciones opuestas.

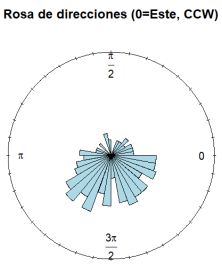
Procedemos a analizar un subconjunto espacial del campo de viento, restringiendo la región a longitudes entre -6 y 0 y latitudes entre 35 y 41. Sobre este subconjunto calculamos estadísticas descriptivas circulares de la dirección, y complementamos el análisis con una visualización del campo direccional en el plano y su correspondiente gráfico de rosas (Figura 9).

En conjunto, la Figura 9 y la Tabla 8 sugieren dos rasgos relevantes. Primero, el campo presenta variabilidad local apreciable (no es un patrón puramente determinista), lo que es consistente con una naturaleza estocástica. No obstante, existe una dirección dominante, la media direccional se ubica cerca de 271,64° y el gráfico de rosas evidencia una concentración marcada en torno a esa dirección, con baja frecuencia de direcciones opuestas.

Estas características son coherentes con el supuesto de un proceso direccional generado por la *envolvoltura* de un campo gaussiano latente, donde se es capaz de producir una



(a) Dirección del viento en la región de estudio (subconjunto espacial).



(b) Distribución angular de las direcciones: gráfico de rosas.

Figura 9: Diagnóstico descriptivo para las direcciones del viento en el subconjunto lon $\in [-6, 0]$ y lat $\in [35, 41]$. El campo exhibe variabilidad local, junto con una dirección predominante reflejada por la concentración angular en el gráfico de rosas.

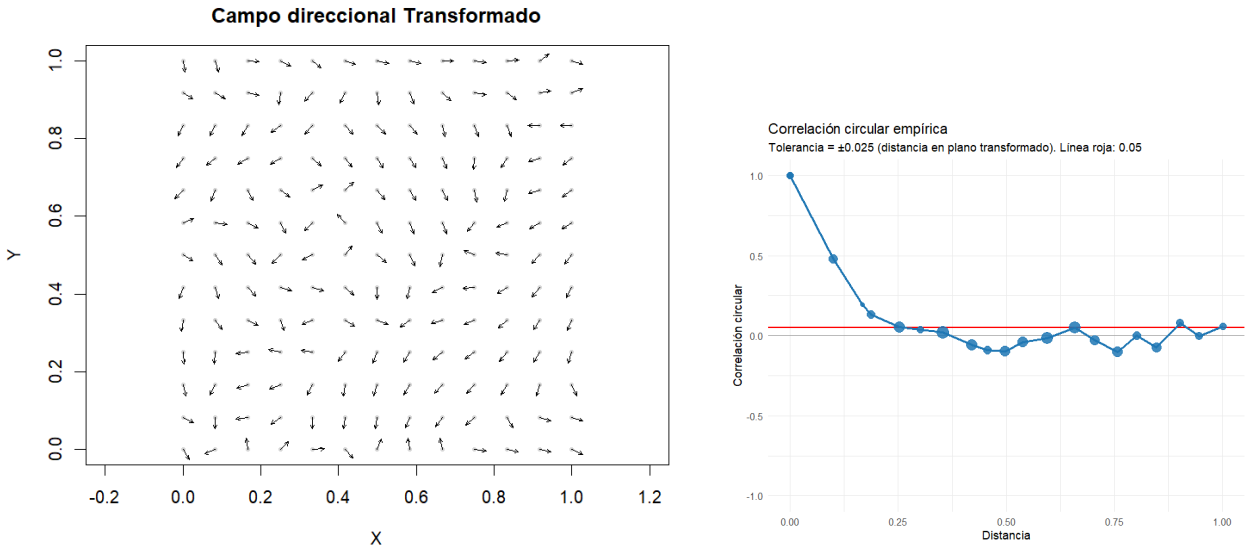
Cuadro 8: Estadísticas descriptivas circulares de la dirección del viento (subconjunto lon $\in [-6, 0]$, lat $\in [35, 41]$).

Métrica	**Valor**	**Interpretación**
Media direccional $\hat{\mu}$ (rad)	4.7411	Dirección promedio en radianes
Media direccional $\hat{\mu}$ (deg)	271,64°	Equivalente en grados (referencia $[0, 360)$)
Desviación direccional $\widehat{\text{sd}}_c$ (rad)	1.1187	Dispersión angular alrededor de $\hat{\mu}$
Concentración $\hat{\kappa}$	0.5348	1 = alta concentración, 0 = baja concentración

distribución angular unimodal y concentrada, sin eliminar la variabilidad estocástica local observada. Por lo tanto, desde un punto de vista descriptivo, resulta razonable modelar las direcciones del viento en esta región mediante un campo Gaussiano-*envuelto*.

Como complemento al análisis exploratorio, resulta informativo estudiar la correlación circular empírica en función de la distancia. Este diagnóstico cumple con evaluar de manera cualitativa si la dependencia observada es compatible con un comportamiento aproximadamente isotrópico y también con obtener una noción sobre el rango de interacción intrínseco que determina la distancia máxima donde sigue habiendo una dependencia significativa.

Para hacer este análisis comparable con la configuración del estudio de simulación, se considera el dominio transformado al cuadrado $[0,1] \times [0,1]$. El campo direccional resultante tras la transformación se muestra en la Figura 10a. A partir de estas coordenadas reescaladas se calcula la correlación circular empírica por bandas de distancia, obteniendo la curva de la Figura 10b.



(a) Campo direccional en el dominio reescalado $[0,1] \times [0,1]$.

(b) Correlación circular empírica en función de la distancia (escala transformada).

Figura 10: Diagnósticos exploratorios para evaluar dependencia espacial e isotropía a partir de direcciones de viento.

La Figura 10b muestra un patrón decreciente hasta cruzar el límite de significancia. Este comportamiento es el que uno esperaría bajo una estructura de covarianza isotrópica. En contraste, si la dependencia variara fuertemente con la dirección (anisotropía marcada), sería razonable esperar fluctuaciones en la curva al mezclar, para una misma distancia, pares de puntos orientados en direcciones espaciales distintas.

Por lo tanto, este diagnóstico entrega evidencia cualitativa a favor de asumir isotropía, además sugiere un rango de interacción razonable, donde no se tendría un rango mayor que la distancia máxima en $\mathscr{D}$, lo cual sería en este caso $\sqrt{2}$. Rangos muy altos podrían significar también estar en un contexto razonablemente determinístico, sugeriendo también presencia

de aleatoriedad. Esto análisis también ayuda a escoger umbrales ds de forma informada.

8.2. Ajuste

Para el ajuste en datos reales se consideraron dos estrategias de verosimilitud compuesta, la versión por parejas y el modelo por bloques. En este último caso, se fijó el número de bloques como un 50 % del número de ubicaciones, lo que se traduce en $B = 85$ clústeres obtenidos mediante `k-means`. En ambos enfoques se ajustaron tres especificaciones de covarianza Matérn, manteniendo fijo el parámetro de suavidad en $\nu \in \{0{,}5, 1{,}5, 2{,}5\}$ (los mismos valores utilizados en el estudio de simulación), y estimando el resto de parámetros del modelo.

Con el objetivo de hacer el ajuste directamente comparable con el diseño de simulación, se trabajó en el dominio reescalado $[0, 1] \times [0, 1]$. Para ello, a cada ubicación $u = (\text{lon}, \text{lat})^\top$ se le aplicó la transformación

$$T(u) \;=\; \frac{u + (6, -35)^\top}{6} \;=\; \left(\frac{\text{lon} + 6}{6}, \frac{\text{lat} - 35}{6}\right)^\top .$$

Esta transformación se justifica bajo el supuesto de isotropía, donde el parámetro de rango α podría verse afectado, sin embargo, según su interpretación, bastaría multiplicarlo por 6 para recuperar la noción en la escala original, donde sigue siendo consitente en términos de interpretación.

A partir del análisis descriptivo (correlación circular en función de la distancia), se observó una dependencia espacial persistente, compatible con rangos relativamente altos en la escala $[0, 1]^2$. Esto sugiere utilizar un umbral de vecindad del orden de $d_s = 0{,}2$ para el modelo por bloques, de modo consistente con la evidencia obtenida en el estudio de simulación.

Los resultados del ajuste se resumen en la Tabla 9. Para cada modelo, el mejor valor de ν se selecciona por mayor $\widehat{\log CL}$ (en negrita). Los valores reportados corresponden a los resultados finales de optimización para cada ν en ambos esquemas de verosimilitud compuesta.

En términos de interpretación, ambos esquemas entregan estimaciones coherentes con el diagnóstico descriptivo, donde se recupera una dirección media cercana a la dirección predominante observada en los datos, y un parámetro de rango/escala alto en la escala transformada, compatible con dependencia espacial persistente. Además, en ambos modelos el criterio de selección favorece $\nu = 2{,}5$, lo cual sugiere que una covarianza Matérn más suave captura mejor la estructura de dependencia presente en este subconjunto.

En conjunto, la estabilidad de las estimaciones entre especificaciones y la consistencia con los patrones observados (concentración angular alrededor de una dirección dominante, pero con variabilidad local no determinista) respaldan que el enfoque Gaussiano-*envuelto* constituye una aproximación razonable para describir estas direcciones de viento en la región analizada.

Como análisis final, estimamos la incertidumbre de los parámetros ajustados mediante *bootstrap*. Evitando así cálculos explícitos de la matriz de información de Godambe, lo cual

Modelo	**Parámetro**	$\nu = 0{,}5$	$\nu = 1{,}5$	$\nu = 2{,}5$
Parejas	$\hat{\mu}$ (rad)	4.692468	4.691406	**4.691016**
	$\hat{\sigma}^2$	1.158077	1.160910	**1.161757**
	$\hat{\alpha}$	0.342536	0.259953	**0.237929**
	$\widehat{\log CL}$	-4111.137437	-4105.657298	**-4104.786966**
Bloques ($B = 85$)	$\hat{\mu}$ (rad)	4.748250	4.750303	**4.750812**
	$\hat{\sigma}^2$	1.273225	1.275910	**1.266743**
	$\hat{\alpha}$	0.428918	0.274006	**0.238509**
	$\widehat{\log CL}$	-1672.213263	-1663.959803	**-1663.451150**

Cuadro 9: Ajuste por verosimilitud compuesta bajo covarianza Matérn con suavidad ν fija. En cada modelo, el mejor ν corresponde al mayor valor de $\widehat{\log CL}$ (en negrita). *Nota:* los valores de $\widehat{\log CL}$ no son comparables entre el modelo por parejas y el modelo por bloques, pues corresponden a construcciones compuestas distintas; sólo deben compararse *dentro* de cada modelo.

es intratable. Aplicar esta metodología nos ayuda también a determinar que propuesta está más *confiada* en sus estimaciones, cuantificando de cierta forma que tan preciso creen ser.

En la Tabla 10 se reportan las desviaciones estándar bootstrap para los parámetros σ^2, α y μ, tanto para el modelo por parejas como para el modelo por bloques.

Modelo	$\text{sd}(\hat{\sigma}^2)$	$\text{sd}(\hat{\alpha})$	$\text{sd}(\hat{\mu})$
Parejas	0.212	0.029	0.224
Bloques	0.235	0.023	0.232

Cuadro 10: Desviaciones estándar estimadas por *bootstrap* para los estimadores bajo el modelo por parejas y el modelo por bloques.

Los resultados del bootstrap son coherentes con respecto al estudio de simulación. Por un lado, el modelo por parejas presenta menor variabilidad al estimar la varianza σ^2 y la media direccional μ, sugiriendo estimaciones más estables para estos parámetros. Por otro lado, el modelo por bloques obtiene menor desviación estándar para el parámetro de rango α, lo que indica una mayor precisión relativa en la caracterización de la dependencia espacial.

Estos cálculos de incertidumbres apoyan el hecho que la metodología por bloques tiene mayor facilidad para capturar la estructura de dependencia espacial, sin embargo, la por parejas parece ser más precisa con respecto al resto de parámetros.

9. Conclusiones

La verosimilitud compuesta por parejas es en la práctica un estándar de referencia en estadística espacial debido a su equilibrio favorable entre información estadística y costo computacional. De hecho, incluso en el caso gaussiano donde la verosimilitud completa es tratable para tamaños moderados y existe mayor flexibilidad para proponer construcciones compuestas alternativas el enfoque por parejas suele emplearse como *benchmark*. En el contexto de campos Gaussianos-*envueltos*, nuestros resultados confirman que esta no es la excepción, el método por parejas destacó por su rapidez y por un desempeño competitivo en precisión en una fracción importante de los escenarios simulados. Sin embargo, tanto el estudio de simulación como el caso aplicado a datos reales muestran que existen situaciones en las que la verosimilitud por bloques puede ser una alternativa preferible.

Una primera comparación entre los métodos es su sensibilidad a la elección del umbral de distancia. El método por parejas es considerablemente menos sensible que el método por bloques, incluso en escenarios con rangos relativamente bajos. Esta diferencia tiene la explicación directa del efecto de agrupar observaciones, la función objetivo por bloques se construye a partir de pares de bloques, en lugar de pares de puntos, reduciendo drásticamente el número de términos disponibles. Además, se impone un umbral de vecindad muy restrictivo, el conjunto de pares de bloques que contribuyen a la verosimilitud compuesta puede volverse demasiado pequeño, deteriorando la información efectiva utilizada para la estimación.

Una segunda fuente de sensibilidad proviene de la distorsión que introduce el agrupamiento sobre la nocíon de cercanía. Dos ubicaciones que, a nivel puntual, están suficientemente próximas como para contribuir en el enfoque por parejas pueden terminar asignadas a clústeres distintos cuyos centroides no queden dentro del umbral seleccionado. En ese caso, interacciones relevantes se excluyen en la construcción por bloques. Por ello, tras el agrupamiento, el umbral de vecindad debe escogerse de manera más conservadora para mitigar este efecto.

La Figura 11 ilustra un caso extremo de esta situación donde afecta la distorsión de distancias al agrupar. El combinar un agrupamiento desfavorable con un umbral demasiado pequeño, puede ocasionar que no existan pares de bloques que satisfagan el criterio de vecindad. En consecuencia, la función objetivo se vuelve nula y deja de ser informativa para estimar los parámetros del modelo.

En contraste con lo anterior, el método por bloques mostró una ventaja sistemática para estimar el parámetro de rango. En términos generales, la verosimilitud por bloques logra capturar con mayor precisión la escala de dependencia, lo que se reflejó en el estudio de simulación y también en el caso aplicado, donde la desviación estándar bootstrap del estimador del rango resultó menor bajo el enfoque por bloques. Esto sugiere que, cuando el objetivo principal es caracterizar la dependencia espacial, el método por bloques presenta una alternativa particularmente atractiva.

Adicionalmente, los resultados indican que la estructura de covarianza puede incidir en qué método resulta más conveniente. En particular, para el caso exponencial ($\nu = 0{,}5$)

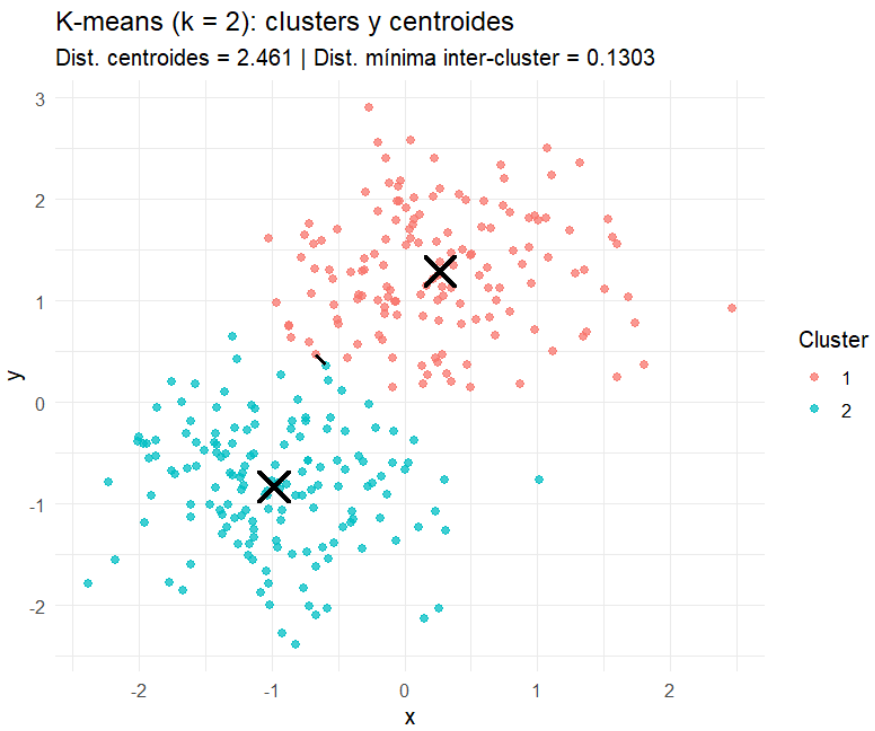


Figura 11: Ejemplo ilustrativo de efecto de borde: un umbral de vecindad excesivamente restrictivo tras el agrupamiento puede eliminar contribuciones entre bloques, llevando a una función objetivo degenerada y, por ende, inadecuada para la estimación.

la verosimilitud por parejas presentó un rendimiento comparativamente sólido incluso para umbrales relativamente grandes. En cambio, al considerar estructuras más suaves ($\nu = 1{,}5$ y $\nu = 2{,}5$), el método por bloques tendió a entregar mejores resultados cuando se emplearon umbrales de vecindad medianos o altos. Este patrón sugiere que, en covarianzas más suaves, incorporar información multivariada a través de bloques puede ser más beneficioso.

Para la estimación de la media y la varianza, no se observa una dominancia clara de un método sobre el otro al aislar efectos de umbral y suavidad. En términos generales, para umbrales bajos (o bajo la covarianza exponencial) el enfoque por parejas tiende a ser más eficiente, mientras que para covarianzas más suaves y umbrales medianos/altos el enfoque por bloques puede igualar o superar su desempeño. Esto sugiere que la diferencia en eficiencia para estos parámetros depende más de las condiciones de umbral y suavidad que de una ventaja intrínseca de un método.

En cuanto al costo computacional, el método por parejas suele ser bastante más rápido. El enfoque por bloques sólo logra superarlo en velocidad en configuraciones particulares con umbrales muy restrictivos y un número elevado de clústeres, lo que reduce el número de contribuciones en la función objetivo. Fuera de esas situaciones, el costo adicional de manejar términos de mayor dimensión hace que el enfoque por bloques tienda a ser más exigente computacionalmente.

A partir de lo anterior, se desprenden dos implicancias prácticas. Primero, para el método por bloques parece natural acoplar el tamaño de los bloques con el umbral: a medida que se considera un umbral más grande (más pares de bloques), puede ser más eficiente utilizar menos bloques, pero de mayor tamaño, con el fin de aprovechar mejor la información multivariada disponible. Segundo, para ambos métodos aumentar el umbral tiende a mejorar la eficiencia estadística, incluso en escenarios con rangos bajos.

En conclusión, la verosimilitud por bloques no pretende reemplazar la verosimilitud por parejas, pero nuestro trabajo identifica condiciones en las que su uso es recomendable. En particular, si el objetivo es inferir con mayor precisión la estructura de dependencia (rango), si se asume una covarianza más suave, o si se está dispuesto a asumir un mayor costo computacional a cambio de ganancias en eficiencia estadística, entonces la verosimilitud por bloques emerge como una alternativa competitiva y hasta en ciertos casos, superior.

10. Trabajos futuros

A partir de los resultados obtenidos, se identifican varias líneas naturales de extensión, tanto para robustecer el modelo como para ampliar el alcance metodológico y aplicado:

- **Incorporar efecto *nugget* en la estructura de covarianza.** En aplicaciones reales es habitual observar variabilidad a micro escala o errores de medición, lo que se modela clásicamente mediante un término *nugget* (ruido blanco) que se suma a la varianza en el origen [Cressie, 1993, Stein, 1999].
- **Estructuras de covarianza más generales: anisotropía (y eventualmente no estacionariedad).** Dado que fenómenos como el viento suelen exhibir direcciones preferenciales, podría resultar razonable permitir *anisotropía*, llegando a posibles conclusiones que el modelo isotrópico no logra capturar.
- **Ampliar la comparativa con otras metodologías de inferencia.** Además de las verosimilitudes compuestas por parejas y por bloques, existen aproximaciones clásicas como la pseudo-verosimilitud basada en especificaciones condicionales locales [Besag, 1974]. Por otro lado, en el caso gaussiano (y modelos relacionados) es natural comparar con aproximaciones tipo Vecchia, que factoriza la densidad conjunta en un producto de condicionales de baja dimensión [Vecchia, 1988, Katzfuss and Guinness, 2021].
- **Extensión a otros campos direccionales y a modelos multivariados direccional–lineal.** Como vimos anteriormente, existen más formas de modelar datos direccionales toroidales y por tanto se podría introducir trabajos análogos para por ejemplo campos aleatorios von Mises o normal proyectados, como posibles alternativas. Por último, muchas aplicaciones observan simultáneamente una variable angular y otra real (p. ej. dirección y magnitud), por lo que resulta natural estudiar modelos conjuntos direccional–lineal [Wang et al., 2015]. En una línea más general, también es posible abordar modelos multivariados con covarianzas Matérn cruzadas para capturar dependencia entre componentes [Gneiting et al., 2010].

11. Referencias

Referencias

Jonathan Acosta, Alfredo Alegría, Felipe Osorio, and Ronny Vallejos. Assessing the effective sample size for large spatial datasets: A block likelihood approach. *Computational Statistics and Data Analysis*, 162, 2021.

Alfredo Alegría, Moreno Bevilacqua, and Emilio Porcu. Likelihood-based inference for multivariate space-time wrapped-gaussian fields. *Journal of Statistical Computation and Simulation*, 86(13):2583–2597, 2016. doi: 10.1080/00949655.2016.1162309.

A. Alegría. Assessing the competitiveness of matrix-free block likelihood estimation in spatial models. *arXiv preprint arXiv:2401.11265*, 2024. URL https://arxiv.org/abs/2401.11265.

Julian Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–236, 1974.

Moreno Bevilacqua, Carlo Gaetan, Jorge Mateu, and Emilio Porcu. Estimating space and space-time covariance functions for large data sets: A weighted composite likelihood approach. *Journal of the American Statistical Association*, 107(497):268–280, 2012. doi: 10.1080/01621459.2011.646928.

Jens Breckling. The analysis of directional time series: Applications to wind speed and direction. *Lecture Notes in Statistics*, 61, 1989.

Petruţa C. Caragea and Richard L. Smith. Asymptotic properties of computationally efficient alternative estimators for a class of multivariate normal models. *Journal of Multivariate Analysis*, 98(7):1417–1440, 2007.

José A. Carta, Celia Bueno, and Penélope Ramírez. Statistical modelling of directional wind speeds using mixtures of von Mises distributions: Case study. *Energy Conversion and Management*, 49(5):897–907, 2008. doi: 10.1016/j.enconman.2007.10.017.

Ted Chang. Spherical regression and the statistics of tectonic plate reconstructions. *International Statistical Review*, 61(2):299–316, 1993.

Jean-Paul Chilès and Pierre Delfiner. *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons, New York, 2nd edition, 2012.

D. Collett and T. Lewis. Discriminating between the von mises and wrapped normal distributions. *Australian Journal of Statistics*, 23(1):73–79, 1981. doi: 10.1111/j.1467-842X.1981.tb00763.x.

Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*. MIT Press, Cambridge, MA, 3 edition, 2009.

Noel A. C. Cressie. *Statistics for Spatial Data*. Wiley-Interscience, New York, revised edition edition, 1993. ISBN 978-0-471-00255-0.

Peter J Diggle and Paulo J Ribeiro. *Model-based Geostatistics*. Springer, New York, 2007.

Bradley Efron and Rob Tibshirani. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1, 02 1986. doi: 10.1214/ss/1177013815.

Jo Eidsvik, Elias T. Krainski, David Bolin, Håvard Rue, and Amelia Pires. *Bayesian Spatial Modeling with R-INLA*. Chapman & Hall/CRC, Boca Raton, 2014a.

Jo Eidsvik, Benjamin A. Shaby, Brian J. Reich, Matthew Wheeler, and Jarad Niemi. Estimation and prediction in spatial models with block composite likelihoods. *Journal of Computational and Graphical Statistics*, 23(2):295–315, 2014b. doi: 10.1080/10618600.2012.760460.

Paul Elliott, Jon C Wakefield, Nicky G Best, and David J Briggs. *Spatial Epidemiology: Methods and Applications*. Oxford University Press, Oxford, 2000.

Javier Fernández-López and Klaus Schliep. rwind: download, edit and include wind data in ecological and evolutionary analysis. *Ecography*, 42(4):804–810, 2019. doi: 10.1111/ecog.03730.

Javier Fernández-López, Klaus Schliep, and Yurena Arjona. *rWind: Download, Edit and Include Wind and Sea Currents Data in Ecological and Evolutionary Analysis*, 2021a. URL `https://cran.r-project.org/package=rWind`. R package version 1.1.7.

Javier Fernández-López, Klaus Schliep, and Yurena Arjona. *wind.data: Wind data example*, 2021b. URL `https://www.rdocumentation.org/packages/rWind/versions/1.1.7/topics/wind.data`. R package documentation, `rWind` version 1.1.7.

Javier Fernández-López, Klaus Schliep, and Yurena Arjona. *wind.dl: Wind-data download*, 2021c. URL `https://www.rdocumentation.org/packages/rWind/versions/1.1.7/topics/wind.dl`. R package documentation, `rWind` version 1.1.7.

NI Fisher and AJ Lee. Analysis of a correlation coefficient for circular data. *Biometrika*, 70 (2):327–332, 1983.

Nicholas I. Fisher. *Statistical Analysis of Circular Data*. Cambridge University Press, Cambridge, UK, 1993.

Nicholas I. Fisher and A. J. Lee. Time series analysis of circular data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(2):327–339, 1994. doi: 10.1111/j.2517-6161.1994.tb01981.x.

Alan E Gelfand, Peter J Diggle, Montserrat Fuentes, and Peter Guttorp. *Handbook of Spatial Statistics*. CRC Press, Boca Raton, 2010.

Tilmann Gneiting, William Kleiber, and Martin Schlather. Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105(491): 1167–1177, 2010. doi: 10.1198/jasa.2010.tm09420.

V. P. Godambe. An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, 31(4):1208–1211, 1960. doi: 10.1214/aoms/1177705905.

Stephane Heritier, Eva Cantoni, Samuel Copt, and Maria-Pia Victoria-Feser. *Robust Methods in Biostatistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester, 2009.

Daniel Hernández-Stumpfhauser, F Jay Breidt, and Mark J van der Woerd. Recent advances in directional statistics. *Test*, 26(1):1–58, 2017.

Kurt Hornik and Bettina Grün. movmf: An r package for fitting mixtures of von mises-fisher distributions. *Journal of Statistical Software*, 58(10):1–31, 2013.

Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. Springer, New York, NY, 2 edition, 2021. doi: 10.1007/978-1-0716-1418-1.

S Rao Jammalamadaka and Y R Sarma. A correlation coefficient for angular variables. *Statistical theory and data analysis II*, pages 349–364, 1988.

S. Rao Jammalamadaka and A. SenGupta. *Topics in Circular Statistics*. World Scientific, Singapore, 2001.

Edwin T Jaynes. Information theory and statistical mechanics. *Physical Review*, 106(4): 620–630, 1957.

Harry Joe and Jianjun J. Xu. The estimation method of inference functions for margins for multivariate models. *Technical Report*, 1996. Department of Statistics, University of British Columbia.

Giovanna Jona-Lasinio, Alan E. Gelfand, and Mattia Jona-Lasinio. Spatial analysis of wave direction data using wrapped gaussian processes. *The Annals of Applied Statistics*, 6(4): 1478–1498, 2012. doi: 10.1214/12-AOAS576.

PE Jupp and KV Mardia. A general correlation coefficient for directional data and related regression problems. *Biometrika*, 67(1):163–173, 1980.

Shogo Kato, Gianluca Mastrantonio, and Masayuki Ishikawa. An interpretable family of projected normal distributions and a related copula model for Bayesian analysis of hypertoroidal data, 2025. URL https://arxiv.org/abs/2508.16432. arXiv preprint arXiv:2508.16432v2, last revised 1 Dec 2025.

Matthias Katzfuss and Joseph Guinness. A general framework for vecchia approximations of gaussian processes. *Statistical Science*, 36(1):124–141, 2021. doi: 10.1214/19-STS755.

Gerhard Kurz, Igor Gilitschenski, and Uwe D. Hanebeck. Efficient evaluation of the probability density function of a wrapped normal distribution. In *Sensor Data Fusion: Trends, Solutions, and Applications (SDF), 2014*, pages 1–6. IEEE, 2014. doi: 10.1109/SDF.2014.6954714.

Jeffrey C Lagarias, James A Reeds, Margaret H Wright, and Paul E Wright. Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on optimization*, 9(1):112–147, 1998.

Francesco Lagona, Marco Picone, and Antonello Maruotti. Analysis of time series of bivariate circular data using hidden markov models. *Environmetrics*, 26(4):252–264, 2015.

John M. Lee. *Introduction to Riemannian Manifolds*, volume 176 of *Graduate Texts in Mathematics*. Springer, Cham, 2 edition, 2018. ISBN 978-3-319-91754-2. doi: 10.1007/978-3-319-91755-9.

Christophe Ley and Thomas Verdebout. *Modern Directional Statistics*. Chapman & Hall/CRC, Boca Raton, FL, 2017.

Kanti V Mardia. Statistics of directional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 37(3):349–393, 1975.

Kanti V. Mardia and Peter E. Jupp. *Directional Statistics*. John Wiley & Sons, Chichester, UK, 2000.

Kanti V. Mardia, Charles C. Taylor, and Ganesh K. Subramaniam. Protein bioinformatics and mixtures of bivariate von Mises distributions for angular data. *Biometrics*, 63(2):505–512, 2007. doi: 10.1111/j.1541-0420.2006.00682.x.

Kanti V. Mardia, Gareth Hughes, Charles C. Taylor, and Harshinder Singh. A multivariate von Mises distribution with applications to bioinformatics. *Canadian Journal of Statistics*, 36(1):99–109, 2008. doi: 10.1002/cjs.5550360110.

Gianluca Mastrantonio. The joint projected normal and skew-normal: A distribution for poly-cylindrical data. *Journal of Multivariate Analysis*, 165:14–26, 2018.

Georges Matheron. Principles of geostatistics. *Economic Geology*, 58(8):1246–1266, 1963.

Ali Mohammadian Mosammam and Serve Mohammadi. Estimating of spatial covariance function using block differenced composite likelihood. *JSS*, 12(2):513–525, 2019.

Alexandre Navarro, Jes Frellsen, and Richard Turner. The multivariate generalised von mises: Inference and applications. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31, 02 2016. doi: 10.1609/aaai.v31i1.10943.

John A Nelder and Roger Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.

Gabriel Núñez-Antonio and Eduardo Gutiérrez-Peña. Spatial modeling of circular data: A case study analysis of forest fire orientations. *Spatial Statistics*, 23:1–16, 2018.

Margaret A Oliver and Richard Webster. *Geostatistics for Environmental Scientists*. John Wiley & Sons, Chichester, 2nd edition, 2010.

Arthur Pewsey, Terry Lewis, and M. C. Jones. Discrimination between the von Mises and wrapped normal distributions: just how big does the sample size have to be? *Statistics*, 39(2):81–89, 2005. doi: 10.1080/02331880500031597.

Arthur Pewsey, Markus Neuhäuser, and Graeme D Ruxton. *Circular Statistics in R*. Oxford University Press, Oxford, 2013.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023. URL https://www.R-project.org/.

Klaus Schmidt-Koenig. Current problems in bird orientation. In Daniel S. Lehrman, Robert A. Hinde, and Evelyn Shaw, editors, *Advances in the Study of Behavior*, volume 1, pages 217–278. Academic Press, New York, 1965.

Michael L Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York, 1999.

Waldo R Tobler. A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46(sup1):234–240, 1970.

Cristiano Varin, Nancy Reid, and David Firth. An overview of composite likelihood methods. *Statistica Sinica*, 21:5–42, 2011a.

Cristiano Varin, Nancy Reid, and David Firth. An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42, 2011b. URL http://www3.stat.sinica.edu.tw/statistica/oldpdf/A21n11.pdf.

Adrian V. Vecchia. Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(2):297–312, 1988. doi: 10.1111/j.2517-6161.1988.tb01729.x.

Fangpo Wang and Alan E. Gelfand. Directional data analysis under the general projected normal distribution. *Statistical Methodology*, 10(1):113–127, 2013. doi: 10.1016/j.stamet.2012.07.005.

Fangpo Wang and Alan E Gelfand. A general framework for circular and spherical data regression models. *Stat*, 9(1):e283, 2020a.

Fangpo Wang and Alan E Gelfand. Statistical modeling and computation for analysis of spatial circular data. *Statistics and Computing*, 30(3):653–672, 2020b.

Fangpo Wang, Alan E. Gelfand, and Giovanna Jona-Lasinio. Joint spatio-temporal analysis of a linear and a directional variable: Space-time modeling of wave heights and wave directions in the adriatic sea. *Statistica Sinica*, 25:25–39, 2015. doi: 10.5705/ss.2013.204w.