

# Una clase de campos aleatorios no Gaussianos con distribuciones marginales *two-piece*

Camilo Esteban Gómez Narváez  
22 de enero de 2020

## **Profesor Guía**

Moreno Bevilacqua. Ph.D.  
Instituto de Estadística, Universidad de Valparaíso

## **Profesor Co-Guía**

Christian Caamaño. Ph.D.  
Departamento de Estadística, Universidad del Bío-Bío

## **Proyecto de titulación para optar al:**

grado académico de: *Licenciado en Estadística*  
título profesional de: *Ingeniero en Estadística*  
menor en: *Modelización Estadística*

# Resumen

En este trabajo se propone una nueva clase de modelos para el análisis de regresión y dependencia en el contexto de observaciones espaciales o temporales continuas que incorporen asimetría o colas pesadas. En particular, se estudian campos aleatorios débilmente estacionarios con distribuciones marginales del tipo *two-piece*.

Se presentan estos modelos de manera general y el estudio de sus propiedades de segundo orden asociadas a la función de covarianza obtenida a partir la construcción presentada. Así mismo, se propone una alternativa computacionalmente eficiente para la estimación de parámetros basado en el método de la verosimilitud compuesta por parejas ponderada junto con un estudio de simulación.

Respecto a la aplicación de los modelos, se propone un predictor óptimo lineal para realizar predicciones sobre nuevas localizaciones bajo estos modelos. Además, toda la implementación computacional para la simulación, estimación y predicción para estos modelos está disponible para la librería (en desarrollo) de R, `GeoModels`.

Finalmente, se estudia el desempeño de los modelos con una aplicación a las observaciones temporales del mercado de divisas (*Forex*) del dólar Australiano sobre el Yen Japonés para el mes de febrero del año 2019.

*Lo intuitivo siempre aparece en el campo de lo estudiado.*  
**S. Prado Ríos.**

# Algunas Palabras

Me daré el pequeño gusto de escribir un poco acerca de lo que ha sido este periplo de estudiar una carrera de pregrado y agradecer a quienes han sido parte, tanto de manera indirecta y por supuesto a los que constantemente han estado ahí para brindarme de sus experiencias y enseñarme un poco más.

Quisiera partir por dedicar todo el trabajo y esfuerzo de estos 5 años continuos a mi núcleo familiar principal, quienes siempre han sido un aporte a mi formación y jamás han perdido la esperanza de que logre las metas que he trazado a lo largo de este viaje. A ellos con mucho cariño todo mi trabajo ya que jamás han dejado de estar ahí.

En segundo lugar y jamás menos importante, quisiera dedicar también este trabajo a mis contertulios del grupo *Interactuando con la Estadística* y a todos los miembros que han ido rotando a través de los años, sin duda una de las experiencias que más alegrías y satisfacciones me ha traído en nivel personal y académico por tener un cálido grupo humano con quienes compartir nuestras locas ideas sobre educación, didáctica, difusión, academia y mucha risa.

Sobre mi paso de estos 5 años por las enseñanzas de mucha gente, quisiera agradecer también al grupo de personas (compañeros de carrera, magíster, doctorado y de otros programas también, académicos, funcionarios, etc.) que han compartido conmigo, por los bellos lazos que hice y espero se mantengan, así como las innumerables oportunidades que me han dado para investigar, trabajar y reflexionar. También por quienes han dejado sus huellas en mi camino y que por distintas circunstancias no pude continuar los vínculos.

También quiero dedicar mis palabras a aquellas personas que fueron un ejemplo para mí ya que sin duda me han enseñado que el esfuerzo y la dedicación siempre tendrán una recompensa y que la mediocridad o el mínimo esfuerzo jamás debe formar parte de mi conjunto de virtudes. Quiero también incorporar a aquellos individuos que me mostraron cómo definitivamente no debo hacer las cosas ya que me permitieron establecer mis propios parámetros de la persona que definitivamente no quiero ser en un futuro.

De manera muy especial quisiera agradecer a mis profesores tutores. Al profesor Moreno Bevilacqua por ofrecerme la oportunidad de trabajar con él en una de nuestras conversaciones en un curso de segundo año que me dictó ya que además de ser una oferta prematura que sin duda me alegra hasta el día de hoy, fue y sigue siendo un privilegio trabajar con él; también por involucrarme siempre en su investigación y proyectos que sin duda derivó en muchas alegrías y logros que han fortalecido mi carrera académica, por su excelente criterio en cada situación que fuese necesario, sus consejos, paciencia y apoyo cada vez que lo requerí. Al profesor Christian Caamaño por siempre compartir sus ideas, enseñanzas y consejo; además, le agradezco también su gentil trato y luz cuando fue necesario.

Especialmente quiero agradecer la amistad incondicional del grupo de madrugadores ya que han sido parte fundamental de mi experiencia universitaria y personal, sin duda el mejor grupo humano del que pude formar parte estos años ya que he aprendido un montón. Por sus consejos, guía, luz, café y retos cuando fue necesario; además de mostrarme la forma crítica de afrontar cada desafío y establecer un estándar de trabajo que sin duda me ha marcado y seguirá siendo el piso de rigurosidad siempre. También por el soporte y perfección de cada locura que se nos ocurrió desarrollar en estos años, junto con otras más descabelladas que algún día terminaremos.

Para finalizar, quisiera agradecer enormemente a quienes he conocido en distintos viajes que he podido realizar este último año, en especial, a todas las increíbles personas de *La Serena School of Data Science 2019* ya que la experiencia fue inolvidable, pero también han sido una fuente enorme de inspiración de cómo un grupo de personas excepcionalmente *top* en sus áreas pueden comprometerse tanto con la formación de las futuras generaciones, investigar, trabajar muy duro y jamás dejar de ser personas amables, cercanas y guías en el camino.

## INTERACTUANDO CON LA ESTADÍSTICA



De manera especial, quiero agradecer a los miembros de la comisión revisora, el Dr. Héctor Araya y a la Dra. Daira Velandia por su impecable trabajo y muy buenas sugerencias respecto al trabajo aquí desarrollado. Por último, agradezco mucho haber conocido al grupo de las cazuelas; por sus consejos y muy buenos momentos.

# Índice general

<b>Resumen</b>	<b>2</b>
<b>Algunas Palabras</b>	<b>6</b>
<b>1. Preliminares</b>	<b>11</b>
1.1. Modelización de observaciones espacialmente dependientes . . . . .	12
1.1.1. Procesos espaciales y sus propiedades . . . . .	13
1.1.2. Campos aleatorios Gaussianos . . . . .	14
1.1.3. Funciones de covarianza isotrópicas . . . . .	15
1.1.4. Estimación de parámetros . . . . .	21
1.1.5. Predicción . . . . .	23
1.1.6. Campos aleatorios no Gaussianos . . . . .	24
1.1.7. Propuesta de investigación . . . . .	25
1.2. Notas del capítulo . . . . .	26
<b>2. Una clase general de campos aleatorios</b>	<b>27</b>
2.1. Aspectos generales asociados al modelo . . . . .	28
2.1.1. Modelo probabilístico general . . . . .	30
2.2. Extensión a campos aleatorios . . . . .	32
2.2.1. Momentos de primer y segundo orden . . . . .	33
2.2.2. Funciones de distribución . . . . .	33
2.2.3. Propiedades geométricas . . . . .	36
2.3. Casos particulares . . . . .	37
2.3.1. Caso $t$ de student . . . . .	38
2.3.2. Caso Gaussiano . . . . .	40
2.3.3. Caso Tukey- $h$ . . . . .	43
2.4. Notas del capítulo . . . . .	48
<b>3. Implementación computacional</b>	<b>49</b>
3.1. Sobre software . . . . .	50
3.2. Estudio por simulación . . . . .	51
3.3. Aplicación a datos reales . . . . .	54
3.4. Notas del capítulo . . . . .	60
<b>4. Comentarios Finales</b>	<b>61</b>
4.1. Conclusiones . . . . .	62
4.2. Preguntas abiertas . . . . .	63

<b>A. Demostraciones</b>	<b>65</b>
A.1. Densidad univariada <i>two-piece</i> . . . . .	66
A.2. Densidad bivariada <i>two-piece</i> . . . . .	68
A.3. Función de covarianza y correlación <i>two-piece</i> . . . . .	70
<b>B. Rutinas Computacionales</b>	<b>72</b>
<b>Referencias</b>	<b>79</b>

# Índice de figuras

1.1.	Funciones de correlación Matérn parametrizadas por: (a) $\mathcal{M}_{v,(2),(0)}(\mathbf{h})$ , (b) $\mathcal{M}_{(2,5),\alpha,(0)}(\mathbf{h})$ y (c) $\mathcal{M}_{(2,5),(2),\tau}(\mathbf{h})$ . . . . .	17
1.2.	Funciones de correlación Wendland Generalizada parametrizadas por: (a) $\mathcal{W}\mathcal{G}_{\gamma,(1),(6),(0)}(\mathbf{h})$ , (b) $\mathcal{W}\mathcal{G}_{(2),\varsigma,(6),(0)}(\mathbf{h})$ , (c) $\mathcal{W}\mathcal{G}_{(2),(1),\delta,(0)}(\mathbf{h})$ y (d) $\mathcal{W}\mathcal{G}_{(2),(1),(6),\tau}(\mathbf{h})$ . . . . .	18
1.3.	Realizaciones de un campo aleatorio Gaussiano estándar con función de correlación Matérn parametrizada por: (a) $\mathcal{M}_{(0,15),(2),(0)}(\mathbf{h})$ , (b) $\mathcal{M}_{(0,15),(0,5),(0)}(\mathbf{h})$ , (c) $\mathcal{M}_{(0,15),(2),(0,5)}(\mathbf{h})$ y (d) $\mathcal{M}_{(0,15),(0,5),(0,5)}(\mathbf{h})$ . . . . .	20
1.4.	Realizaciones de un campo aleatorio Gaussiano estándar con función de correlación Wendland Generalizada parametrizada por: (a) $\mathcal{W}\mathcal{G}_{(0,15),(3),(4),(0)}(\mathbf{h})$ , (b) $\mathcal{W}\mathcal{G}_{(0,15),(1),(4),(0)}(\mathbf{h})$ , (c) $\mathcal{W}\mathcal{G}_{(0,15),(3),(4),(0,5)}(\mathbf{h})$ y (d) $\mathcal{W}\mathcal{G}_{(0,15),(1),(4),(0,5)}(\mathbf{h})$ . . . . .	21
1.5.	Ejemplo de dominio creciente asintótico en una grilla regula de dimensión 10x10 con distancia mínima de 0,25 unidades para: (a) 25, (b) 441 y (c) 1681 localizaciones espaciales. . . . .	23
1.6.	Ejemplo de dominio fijo asintótico en una grilla regula de dimensión 1x1 con distancia mínima de 0,25, 0,025 y 0,0125 unidades para: (d) 25, (e) 1681 y (f) 6661 localizaciones espaciales. . . . .	23
2.1.	Comparación funciones de distribución de probabilidad Gaussianas y <i>two-piece</i> resultante. . . . .	28
2.2.	Comparación funciones de distribución de probabilidad <i>two-piece</i> : (a) Gaussiana estándar, (b) exponencial potencia ( $\alpha = 0,3$ ) y (c) logística estándar. . . . .	31
2.3.	Ilustración sobre características de un campo aleatorio <i>two-piece</i> t de student: (a) distribuciones marginales ( $v = 4$ ), (b) realización del proceso ( $\eta = 0,8$ y $v = 6$ ) con función de correlación $\mathcal{M}_{(0,5),(0,1),(0)}(\mathbf{h})$ , (c) función de correlación y (d) histograma de una realización. . . . .	40
2.4.	Ilustración sobre características de un campo aleatorio <i>two-piece</i> Gaussiano: (a) distribuciones marginales, (b) realización del proceso ( $\eta = 0,8$ ) con función de correlación $\mathcal{M}_{(0,5),(0,1),(0)}(\mathbf{h})$ , (c) función de correlación y (d) histograma de una realización. . . . .	41
2.5.	Comparación graficos de contorno para las funciones de distribución bivariadas estándar: (a) Gaussiana [ $\rho = 0$ ] (b) Gaussiana [ $\rho = -0,5$ ] (c) Gaussiana [ $\rho = 0,9$ ] (d) <i>two-piece</i> Gaussiana [ $\rho = 0$ ][ $\eta = 0$ ] (e) <i>two-piece</i> Gaussiana [ $\rho = -0,5$ ][ $\eta = 0$ ] (f) <i>two-piece</i> Gaussiana [ $\rho = 0,9$ ][ $\eta = 0$ ] (g) <i>two-piece</i> Gaussiana [ $\rho = 0$ ][ $\eta = 0,3$ ] (e) <i>two-piece</i> Gaussiana [ $\rho = -0,5$ ][ $\eta = 0,7$ ] (f) <i>two-piece</i> Gaussiana [ $\rho = 0,9$ ][ $\eta = -0,70$ ]. . . . .	42
2.6.	Ilustración sobre características de un campo aleatorio Tukey- <i>h</i> : (a) distribuciones marginales, (b) realización del proceso ( $h = 0,4$ ) con función de correlación $\mathcal{M}_{(2,5),(0,033),(0)}(\mathbf{h})$ , (c) función de correlación y (d) histograma de una realización. . . . .	45

2.7.	Ilustración sobre características de un campo aleatorio <i>two-piece</i> Tukey- <i>h</i> : (a) distribuciones marginales ( $h = 0,45$ ), (b) realización del proceso ( $\eta = 0,8$ y $h = 0,3$ ) con función de correlación $\mathcal{M}_{(0,5),(0,1),(0)}(\mathbf{h})$ , (c) función de correlación y (d) histograma de una realización. . . . .	46
2.8.	Comparación gráficos de contorno para las funciones de distribución bivariadas estándar: (a) Tukey- <i>h</i> [ $\rho = 0$ ] (b) Tukey- <i>h</i> [ $\rho = -0,5$ ] (c) Tukey- <i>h</i> [ $\rho = 0,9$ ] (d) <i>two-piece</i> Tukey- <i>h</i> [ $\rho = 0$ ][ $\eta = 0$ ] (e) <i>two-piece</i> Tukey- <i>h</i> [ $\rho = -0,5$ ][ $\eta = 0$ ] (f) <i>two-piece</i> Tukey- <i>h</i> [ $\rho = 0,9$ ][ $\eta = 0$ ] (g) <i>two-piece</i> Tukey- <i>h</i> [ $\rho = 0$ ][ $\eta = 0,3$ ] (e) <i>two-piece</i> Tukey- <i>h</i> [ $\rho = -0,5$ ][ $\eta = 0,7$ ] (f) <i>two-piece</i> Tukey- <i>h</i> [ $\rho = 0,9$ ][ $\eta = -0,70$ ]. . . . .	47
3.1.	Escala de color <i>viridis</i> (imagen capturada desde R Pubs). . . . .	51
3.2.	Gráficos de caja para la estimación de parámetros del campo aleatorio <i>two-piece</i> Tukey- <i>h</i> para: (a) $\beta_0 = 0,5$ , (b) $\beta_1 = -0,25$ , (c) $\delta = 0,2$ , (d) $\sigma^2 = 1$ , (e) $\eta = 0,3$ y (g) $h = 0,1$ . . . . .	52
3.3.	Gráficos de caja para la estimación de parámetros del campo aleatorio <i>two-piece</i> <i>t</i> de student para: (a) $\beta_0 = 0,5$ , (b) $\beta_1 = -0,25$ , (c) $\delta = 0,2$ , (d) $\sigma^2 = 1$ y (e) $\eta = 0,3$ . . . . .	53
3.4.	Gráficos de caja para la estimación de parámetros del campo aleatorio <i>two-piece</i> Gaussiano para: (a) $\beta_0 = 0,5$ , (b) $\beta_1 = -0,25$ , (c) $\delta = 0,2$ , (d) $\sigma^2 = 1$ y (e) $\eta = 0,3$ . . . . .	53
3.5.	Descripción gráfica del conjunto de observaciones <i>Forex</i> de apertura para febrero del 2019 con la distribución marginal (histograma a la izquierda) y el comportamiento de la realización en el tiempo (a la derecha). . . . .	54
3.6.	Dispersograma de las observaciones <i>Forex</i> AUDJPY a distancia temporal: 1, 7, 15, 25, 40, 50 y 70 unidades. . . . .	55
3.7.	Semi-variograma empírico para las observaciones <i>Forex</i> AUDJPY. . . . .	55
3.8.	Gráficas para el diagnóstico distribucional (izquierda) y espacial (derecha) de los modelos: (a) Gaussiano, (b) Gaussiano asimétrico y (c) <i>two-piece</i> Gaussiano. . . . .	59
4.1.	Ilustración sobre características de la distribución marginal de un campo aleatorio ArcSinh Gaussiano estándar. . . . .	63
4.2.	Realizaciones de campos aleatorios <i>two-piece</i> Weibull parametrizados por: (a) $P_{\eta=0;\varphi=2}^*$ , (b) $P_{\eta=-0,4;\varphi=5}^*$ y (c) $P_{\eta=0,8;\varphi=10}^*$ ( $\mathbf{h}$ ) con función de correlación Matén $\mathcal{M}_{(0,5),(0,015)}(\mathbf{h})$ . . . . .	64

# Índice de tablas

3.1.	Lista de modelos de campos aleatorios disponibles en <i>GeoModels</i> . . . . .	50
3.2.	Sesgo y error cuadrático medio para la estimación de parámetros de un campo aleatorio <i>two-piece</i> Gaussiano con función de correlación Wendland Generalizado bajo verosimilitud compuesta por parejas ponderada. . . . .	51
3.3.	Sesgo y error cuadrático medio para la estimación de parámetros de un campo aleatorio <i>two-piece</i> t de student con función de correlación Wendland Generalizado bajo verosimilitud compuesta por parejas ponderada. . . . .	52
3.4.	Sesgo y error cuadrático medio para la estimación de parámetros de un campo aleatorio <i>two-piece</i> Tukey- <i>h</i> con función de correlación Wendland Generalizado bajo verosimilitud compuesta por parejas ponderada. . . . .	52
3.5.	Estadística descriptiva para las observaciones de <i>Forex</i> AUDJPY de apertura. . . . .	54
3.6.	Lista de modelos de campos aleatorios para la aplicación en las observaciones de <i>Forex</i> . . . . .	56
3.7.	Estimación de parámetros de los modelos propuestos para los datos de <i>Forex</i> y valor que alcanza máximo verosimilitud compuesta por parejas ponderada (Valor máx.). . . . .	57
3.8.	Estimación de parámetros de los modelos propuestos para los datos de <i>Forex</i> con sus errores estándar de estimación (en paréntesis), valor que alcanza máximo verosimilitud compuesta por parejas ponderada (Valor máx.) y la evaluación de los criterios de selección de modelos (PLIC y BLIC). . . . .	58
3.9.	Desempeño de predicción bajo un predictor óptimo lineal de los modelos propuestos en las observaciones de <i>Forex</i> para distintos porcentajes de observaciones en la estimación (% de Obs.), RMSE y MAE. . . . .	58

# Capítulo 1

## Preliminares

En este capítulo se presentan las nociones preliminares que sustentan este trabajo. En primer lugar se entregan los conceptos teóricos básicos de la modelización de observaciones espacialmente dependientes. Seguido de esto se estudia el campo aleatorio Gaussiano como principal herramienta para la modelización de observaciones espacialmente dependientes, sus características y algunas propiedades relacionadas a los momentos de segundo orden.

Se finaliza este capítulo con la motivación de este trabajo respecto a modelos de campos aleatorios continuos con características distintas a la Gaussianidad en las observaciones, concluyendo así con las preguntas de investigación para este trabajo y un resumen del resto de los capítulos.

## 1.1. Modelización de observaciones espacialmente dependientes

La resolución de problemas modernos es una constante demanda de nuevas herramientas que permitan la abstracción matemática o estadística para una parcial comprensión de diversos fenómenos y que logren un aporte a la toma de decisiones. En consideración de lo anterior, generar nuevos modelos es importante ya que permiten mejorar el proceso de tomar decisiones bajo incertidumbre.

En los últimos años se ha podido observar que se han complejizado los modelos estadísticos en cuanto a los diversos factores que se consideran al modelizar algún fenómeno específico. Particularmente, uno de estos factores es la dependencia entre las observaciones. Este factor se ha abordado desde distintas propuestas de modelos que se utilizan para el análisis de datos bajo particulares estructuras de dependencia como la dependencia temporal a tiempos discretos (modelos de series temporales) (Palma, 2016) o de manera más general procesos estocásticos (Rao, 2011).

Otro tipo de dependencia que ha tenido un auge de investigación y aplicación es la denominada dependencia espacial. Este tipo de dependencia se sustenta en que las observaciones siguen el principio de la Geología, es decir, que a medida que las observaciones o mediciones de algún fenómeno en particular tienen una menor distancia entre ellas (dentro de algún espacio), estas observaciones tienden a ser parecidas (Porcu, Montero, y Schlather, 2012). Bajo este concepto se introduce el área estadística que se ha encargado de estudiar este tipo de dependencia que es la Estadística Espacial o Geoestadística.

Para el estudio del como varían las observaciones en un espacio se han introducido distintas funciones denominadas funciones de correlación (Stein, 1999), funciones que están definidas sobre la distancia entre dos puntos. En general, se estudian modelos para observaciones espacialmente dependiente sobre un plano, pero también se han investigado funciones de correlación en la esfera y contextos espacio temporales (White y Porcu, 2019). Más detalles sobre funciones de correlación y sus propiedades se mostrarán en las siguientes subsecciones.

Junto al análisis de dependencia (funciones de correlación) también se considera en Estadística Espacial la naturaleza de las observaciones (por ejemplo, observaciones continuas o discretas) y bajo este punto en que el modelo más conocido es el campo aleatorio Gaussiano (Gelfand y Schliep, 2016) ha tomado un rol fundamental en distintos contextos de aplicación por la facilidad que entrega en su implementación computacional; aunque en la práctica, no todo se puede ajustar a un modelo Gaussiano.

Respecto a la naturaleza de las observaciones, este trabajo pretende extender los modelos de campos aleatorios disponibles para la aplicación en escenarios de observaciones continuas. Una consideración es que este trabajo se basa en campos aleatorios aunque existen otras propuestas como variaciones a los modelos de regresión como modelos mixtos o modelos mixtos geoaditivos (Schmidt, Cruse-Sanders, Chamberlain, Ferreira, y Young, 2019) y modelos basados en cópulas (Han y De Oliveira, 2016).

A continuación se presentan la definiciones generales y propiedades de procesos espaciales, particularizando al campo aleatorio Gaussiano como herramienta principal de modelización de observaciones espacialmente dependientes. Luego, se realiza una revisión de algunas funciones de correlación espacial que se utilizarán en este trabajo y algunos aspectos de la estimación de parámetros para el campo aleatorio Gaussiano. Para finalizar la sección se hace mención de nuevas propuestas de

modelización bajo campos aleatorios no Gaussianos.

### 1.1.1. Procesos espaciales y sus propiedades

La noción general de un proceso espacial es considerar por  $X = \{X(\mathbf{s}), \mathbf{s} \in S\}$  una colección o familia de variables aleatorias indexadas en  $\mathbf{s}$ , para  $\mathbf{s} \in S \subseteq \mathbb{R}^d$  un conjunto de índices o localizaciones pertenecientes al subconjunto  $S$  dentro de un espacio  $d$ -dimensional en  $\mathbb{R}^d$ .

Desde el punto de vista práctico, las observaciones o datos corresponden a una realización o muestra del proceso  $X$ , es decir, un vector de observaciones de la forma  $X(\mathbf{s}_1), X(\mathbf{s}_2), \dots, X(\mathbf{s}_n)$ , para  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n \in S$ . Por ejemplo, el conjunto de localizaciones en una dimensión ( $d = 1$ ) podrían corresponder a instantes específicos en el tiempo y en dos dimensiones ( $d = 2$ ) corresponder a la longitud y latitud en un sistema de coordenadas dentro de un mapa.

Para el estudio de procesos espaciales la atención principal está en estudiar  $X$  a través de la función de distribución conjunta asociada al proceso y las distintas características que se puedan atribuir a esta función como la media, varianza o covarianza (Abrahamsen, 1997). Entonces, para un proceso espacial o campo aleatorio  $X$ , la función de distribución conjunta o finito dimensional  $F_{X(\mathbf{s}_1), \dots, X(\mathbf{s}_n)}(x_1, \dots, x_n)$  queda definida por:

$$F_{X(\mathbf{s}_1), \dots, X(\mathbf{s}_n)}(x_1, \dots, x_n) = \mathbb{P}[X(\mathbf{s}_1) \leq x_1, \dots, X(\mathbf{s}_n) \leq x_n] . \quad (1.1)$$

Una aclaración de notación importante es señalar a que a partir de esta sección se denominará por campo aleatorio a cualquier proceso espacial  $X = \{X(\mathbf{s}), \mathbf{s} \in S\}$  cuya dimensión de índices  $\mathbf{s} \in S$  sea mayor o igual a dos, es decir,  $\mathbf{s} \in S \subseteq \mathbb{R}^d$  y  $d \geq 2$ .

### Estacionaridad

Tal como se mencionó, el estudio de procesos espaciales está basado en las características que se atribuyen a la función de distribución conjunta (1.1) de un proceso  $X$ , específicamente estas características o supuestos se realizan sobre los momentos del proceso.

Uno de los supuestos frecuente es asumir la estacionaridad estricta del proceso, es decir, que independiente del conjunto de localizaciones siempre se cumple que:

$$\mathbb{P}[X(\mathbf{s}_1) \leq x_1, \dots, X(\mathbf{s}_n) \leq x_n] = \mathbb{P}[X(\mathbf{s}_1 + \mathbf{h}) \leq x_1, \dots, X(\mathbf{s}_n + \mathbf{h}) \leq x_n] , \quad (1.2)$$

para  $\mathbf{h} \in \mathbb{R}^d$  una traslación dentro del espacio.

Dado que el supuesto de estacionaridad estricta puede ser muy restrictivo para la construcción de nuevos procesos, un segundo supuesto puede también ser considerado. Para un proceso  $X$ , se puede definir que es estacionario débil si es que se cumple que:

$$\begin{aligned} \mathbb{E}[X(\mathbf{s})] &= \mu , \\ \text{Cov}[X(\mathbf{s}_i), X(\mathbf{s}_j)] &= \mathbb{C}(\mathbf{h}) , \end{aligned} \quad (1.3)$$

para  $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$  la separación entre dos localizaciones. Este supuesto señala que el proceso  $X$  tendrá media constante independientemente de su localización ( $\mathbf{s}$ ) y que la covarianza es una función ( $\mathbb{C}$ ) que dependerá exclusivamente la separación ( $\mathbf{h}$ ) entre dos localizaciones ( $\mathbf{s}_i, \mathbf{s}_j \in S$ ). Se denomina  $\mathbb{C}(\mathbf{h})$  por función de covarianza y por función de correlación a:

$$\rho(\mathbf{h}) = \frac{\mathbb{C}(\mathbf{h})}{\mathbb{C}(0)} . \quad (1.4)$$

Otro supuesto de la estacionaridad débil es que el proceso debe tener varianza constante, por lo que cualquier función de covarianza se puede expresar de la siguiente manera:

$$\mathbb{C}(\mathbf{h}) = \sigma^2 \rho(\mathbf{h}) , \quad (1.5)$$

esto implica que el estudio sobre características de funciones de covarianza puede ser equivalente al estudio sobre funciones de correlación bajo estacionaridad débil sin pérdida de generalidad (Abrahamsen, 1997).

Es importante señalar que la condición de estacionaridad estricta (1.2), cuando los momentos de segundo orden son finitos, implica estacionaridad débil (1.3), pero esta implicancia no siempre se cumple en el sentido contrario.

### 1.1.2. Campos aleatorios Gaussianos

La herramienta de modelización Geoestadística más utilizada es el campo aleatorio Gaussiano (Gelfand y Schliep, 2016). Este modelo es bastante estudiado debido a una gran cantidad de propiedades de modelización de observaciones espacialmente dependientes que se pueden derivar a partir de la función de correlación asociada al proceso. Además, para la inferencia de parámetros del proceso, las expresiones algebraicas tienden a ser no tan complejas por lo que utilizar campos aleatorios Gaussianos se vuelve muy conveniente.

Se define por un campo aleatorio Gaussiano  $G = \{G(\mathbf{s}), \mathbf{s} \in S\}$  si es que su función de distribución de probabilidad conjunta es una función de distribución de probabilidad Gaussiana multivariada. Dado que la función de distribución conjunta Gaussiana multivariada se caracteriza por la estructura de media y covarianza, fácilmente se pueden atribuir propiedades al campo aleatorio Gaussiano.

Un campo aleatorio Gaussiano es intrínicamente estacionario si se cumple:

$$\mathbb{E}[G(\mathbf{s})] = \mu , \quad (1.6)$$

$$\begin{aligned} \text{Var}[X(\mathbf{s}_i) - X(\mathbf{s}_j)] &= 2\gamma(\mathbf{s}_j - \mathbf{s}_i) , \\ &= 2\gamma(\mathbf{h}) , \end{aligned} \quad (1.7)$$

con  $2\gamma(\mathbf{h})$  la función llamada variograma. Intrínicamente estacionario es una propiedad más débil que estacionariedad débil (1.3), pero si se asume que el campo aleatorio  $G$  es débilmente estacionario se puede demostrar que:

$$\gamma(\mathbf{h}) = \mathbb{C}(0) - \mathbb{C}(\mathbf{h}) . \quad (1.8)$$

Una manera más general de caracterizar un campo aleatorio Gaussiano es considerando la siguiente representación:

$$G(\mathbf{s}) = \mu(\mathbf{s}) + \sigma G^*(\mathbf{s}) + \tau \epsilon(\mathbf{s}) , \quad (1.9)$$

entonces, definiendo  $G$  como un campo aleatorio Gaussiano no estacionario cuyas componentes son:  $\mu(\mathbf{s}) \in \mathbb{R}$  como una función espacialmente dependiente que puede ser definida, en una versión más tradicional, como un arreglo de covariables  $M(\mathbf{s})^T \boldsymbol{\beta}$  para un análisis de regresión espacial con vector de parámetros  $\boldsymbol{\beta} \in \mathbb{R}^p$ .

A (1.9) también se incorpora un parámetro de escala al modelo,  $\sigma \in \mathbb{R}^+$  y otro campo aleatorio estrictamente estacionario  $\epsilon = \{\epsilon(\mathbf{s}), \mathbf{s} \in S\}$  de media 0, varianza 1 y con correlación espacial nula, además de ser independiente de  $G^*$  un campo aleatorio Gaussiano débilmente estacionario con  $\mathbb{E}[G^*(\mathbf{s})] = 0$ ,  $\text{Var}[G^*(\mathbf{s})] = 1$  y  $\text{Corr}[G^*(\mathbf{s}_i), G^*(\mathbf{s}_j)] = \rho(\mathbf{h})$ .

La ventaja de considerar una representación del tipo (1.9) para un campo aleatorio Gaussiano es que permite realizar el análisis de regresión y de dependencia espacial dado que los momentos del campo aleatorio (1.9) quedan establecidos por:

$$\mathbb{E}[G(\mathbf{s})] = \mu(\mathbf{s}) , \quad (1.10)$$

$$\text{Var}[G(\mathbf{s})] = \sigma^2 + \tau^2 , \quad (1.11)$$

$$\text{Cov}[G(\mathbf{s}_i), G(\mathbf{s}_j)] = \sigma^2 \rho(\mathbf{h}) + \tau^2 I(\mathbf{h} = 0) , \quad (1.12)$$

$$\text{Corr}[G(\mathbf{s}_i), G(\mathbf{s}_j)] = \frac{\sigma^2 \rho(\mathbf{h}) + \tau^2 I(\mathbf{h} = 0)}{\sigma^2 + \tau^2} , \quad (1.13)$$

para  $I$  una función indicatriz y  $\tau \in \mathbb{R}_0^+$  el llamado efecto *nugget* el cual se traduce en una variación a pequeña escala que no se puede diferenciar de un posible error de medición cuyos efectos en la función de correlación se estudiarán en la siguiente subsección.

Bajo la representación (1.9) de un campo aleatorio Gaussiano, la función de distribución conjunta está dada por:

$$f_G(\mathbf{g}; \boldsymbol{\mu}, \Omega) = \frac{1}{(2\pi)^{n/2} |\Omega|^{1/2}} \exp \left\{ \frac{-(\mathbf{g} - \boldsymbol{\mu})^T \Omega^{-1/2} (\mathbf{g} - \boldsymbol{\mu})}{2} \right\} , \quad (1.14)$$

donde  $\mathbf{g}$  es un vector de observaciones  $n$ -dimensional (conocido también como la realización del campo aleatorio  $G$ ),  $\boldsymbol{\mu} = [\mu(\mathbf{s}_i)]_{i=1}^n$  un vector  $n$ -dimensional correspondiente al arreglo de covariables,  $\Omega \in \mathbb{R}^n \times \mathbb{R}^n$  la matriz de covarianzas cuyos elementos se definen por (1.12) y  $\rho(\mathbf{h})$  la función de correlación asociada al campo aleatorio  $G^*$ .

### 1.1.3. Funciones de covarianza isotrópicas

Para el análisis de dependencia espacial asociado a campos aleatorios, la función de covarianza ( $\mathbb{C}(\mathbf{h})$ ) asociada al proceso tiene un rol fundamental para la definición de los modelos. Formalmente, una función de covarianza debe ser una función definida positiva (Abrahamsen, 1997), es decir, debe cumplir para  $\mathbf{s}_i, \mathbf{s}_j \in S$  y  $c_i \in \mathbb{R}$  con  $i = (1, \dots, n)$ :

$$\sum_{i=1}^n \sum_{j=1}^n c_i c_j \mathbb{C}(\mathbf{s}_i, \mathbf{s}_j) \geq 0 . \quad (1.15)$$

Respecto a la propiedad de isotropía, se refiere a que la función de correlación solo dependerá de la distancia Euclidiana entre dos localizaciones espaciales, es decir:

$$\mathbb{C}(\mathbf{s}_i, \mathbf{s}_j) = \mathbb{C}(\mathbf{h}) \quad , \quad (1.16)$$

para  $\mathbf{h} = \|\mathbf{s}_i - \mathbf{s}_j\|$  la distancia Euclidiana entre dos localizaciones  $\mathbf{s}_i, \mathbf{s}_j \in S$ .

Bajo campos aleatorios estacionarios, la función de covarianza estudiada tiene la siguiente estructura:

$$\mathbb{C}(\mathbf{h}) = \sigma^2 \rho(\mathbf{h}) + \tau^2 I(\mathbf{h} = 0) \quad . \quad (1.17)$$

Las propiedades de las funciones de covarianza para campos aleatorios estacionarios de segundo orden se pueden resumir en:

1.  $\mathbb{C}(0) \geq 0$ .
2.  $\mathbb{C}(\mathbf{h}) = \mathbb{C}(-\mathbf{h})$  (función par).
3.  $\mathbb{C}(0) \geq \mathbb{C}(\mathbf{h})$ .
4.  $\sum_{i=1}^k \mathbb{C}_i(\mathbf{h})$  es una función de covarianza si solo si  $\mathbb{C}_i(\mathbf{h})$  corresponden a funciones de covarianza.
5.  $\prod_{i=1}^k \mathbb{C}_i(\mathbf{h})$  es una función de covarianza si solo si  $\mathbb{C}_i(\mathbf{h})$  corresponden a funciones de covarianza.
6.  $\mathbb{C}(\mathbf{h})$  es una función de covarianza para  $\mathbb{R}^r$  si lo es para  $\mathbb{R}^d$ , con  $r \geq d$ .

La caracterización de funciones de covarianza o de correlación, si bien no es una tarea sencilla, se presenta de manera unificada en [Abrahamsen \(1997\)](#). El autor señala que cualquier función de correlación isotrópica para campos aleatorios debilmente estacionarios se pueden representar por:

$$\rho(\mathbf{h}) = 2^{\frac{d-2}{2}} \Gamma\left(\frac{d}{2}\right) \int_0^\infty (\mathbf{w}\mathbf{h})^{-\frac{d-2}{2}} \mathbf{J}_{\frac{d-2}{2}}(\mathbf{w}\mathbf{h}) dG(\mathbf{w}) \quad , \quad (1.18)$$

para  $\mathbf{J}_k$  es una función de Bessel de primera especie de orden  $k$  y  $G$  una medida no decreciente acotada en  $\mathbb{R}^+$  y con  $G(0) = 0$ .

Para este trabajo se utilizarán dos funciones de correlación paramétricas, funciones estudiadas en [Bevilacqua, Faouzi, Furrer, y Porcu \(2019c\)](#):

1. Función de correlación Matérn:

$$\mathcal{M}_{v,\alpha}(\mathbf{h}) = \frac{2^{1-v}}{\Gamma(v)} \left(\frac{\mathbf{h}}{\alpha}\right)^v \mathcal{K}_v\left(\frac{\mathbf{h}}{\alpha}\right) \quad . \quad (1.19)$$

Este modelo es uno de los más utilizados ya que permite ajustar de manera flexible dos características. La primera característica, asociada al parámetro  $v \in \mathbb{R}_{[0,5,\infty)}^+$ , es la capacidad de la función de controlar la suavidad en el origen. La segunda característica, asociada al parámetro  $\alpha \in \mathbb{R}^+$ , es controlar el decaimiento de la función a medida de que la distancia entre dos localizaciones espaciales aumenta.

Sobre las componentes de la función de correlación Matérn (1.19),  $\mathcal{K}_v$  es una función de Bessel modificada de segundo tipo de orden  $v$  y  $\Gamma$  una función Gamma. De manera ilustrativa, los efectos de los parámetros en (1.19) pueden ser observados en la Figura 1.1 para la parametrización de la función de correlación (1.19) que considera un efecto *nugget*, es decir,  $\mathcal{M}_{v,\alpha,\tau}(\mathbf{h}) = \frac{(\tau^2 \mathcal{M}_{v,\alpha}(\mathbf{h}) + I(\mathbf{h}=0))}{(1+\tau^2)}$ .

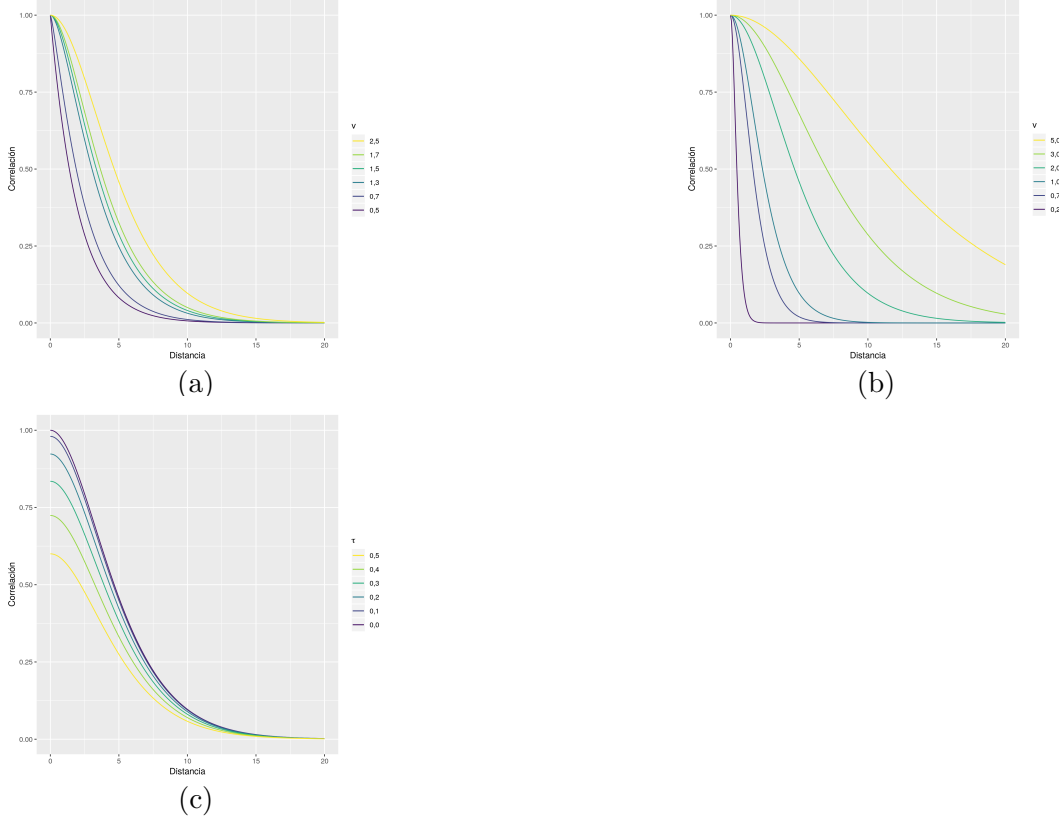


Figura 1.1: Funciones de correlación Matérn parametrizadas por: (a)  $\mathcal{M}_{v,(2),(0)}(\mathbf{h})$ , (b)  $\mathcal{M}_{(2,5),\alpha,(0)}(\mathbf{h})$  y (c)  $\mathcal{M}_{(2,5),(2),\tau}(\mathbf{h})$ .

## 2. Función de correlación Wendland Generalizada:

$$\mathcal{WG}_{\gamma,\varsigma,\delta}(\mathbf{h}) = \begin{cases} \frac{\int_{\mathbf{h}/\delta}^1 u \left( u^2 - (\mathbf{h}/\delta)^2 \right)^{\varsigma-1} (1-u)^\gamma du}{B(2\varsigma,\gamma+1)} & , \quad \mathbf{h} \leq \delta \quad , \\ 0 & , \quad e.o.c. \quad . \end{cases} \quad (1.20)$$

Si bien el modelo Wendland Generalizado ha sido menos utilizado como función de correlación en la literatura; los parámetros del modelo permiten controlar, al igual que en el modelo Matérn, el decaimiento de la función a medida que la distancia aumenta ( $\gamma \in \mathbb{R}^+$ ) y la suavidad de la función en el origen ( $\varsigma \in \mathbb{R}^+$ ). Pero la ventaja de utilizar este modelo es que se incorpora el parámetro de soporte compacto ( $\delta \in \mathbb{R}^+$ ) el cual controla el punto dentro del recorrido de la distancia en que la correlación es cero.

Desde el punto de vista computacional, una ventaja de las funciones de correlación con soporte compacto es que generan matrices *sparse* o con muchos ceros, lo cual disminuye el tiempo de cómputo para simular o estimar con estos modelos (Bevilacqua y cols., 2019c).

De la misma manera que para la función de correlación Matérn (1.19), se puede ilustrar el efecto de cada parámetro en la función de correlación Wendland Generalizada tal como se aprecia en la Figura 1.2 para la parametrización de la función de correlación (1.20) que considera un efecto *nugget*, es decir,  $\mathcal{W}\mathcal{G}_{\gamma,\varsigma,\delta,\tau}(\mathbf{h}) = \frac{(\tau^2 \mathcal{W}\mathcal{G}_{\gamma,\varsigma,\delta}(\mathbf{h}) + I(\mathbf{h}=0))}{(1+\tau^2)}$ .

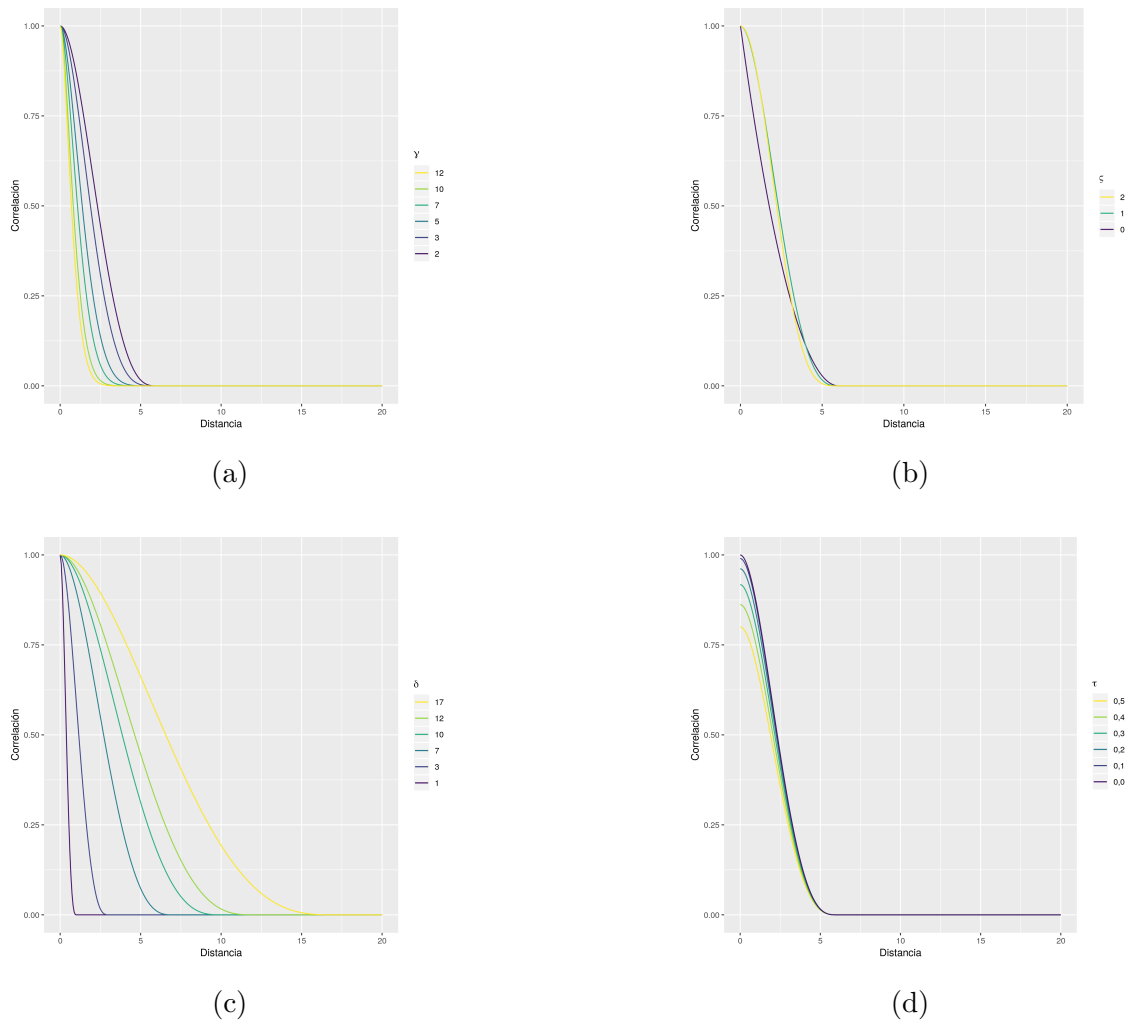


Figura 1.2: Funciones de correlación Wendland Generalizada parametrizadas por: (a)  $\mathcal{W}\mathcal{G}_{\gamma,(1),(6),(0)}(\mathbf{h})$ , (b)  $\mathcal{W}\mathcal{G}_{(2),\varsigma,(6),(0)}(\mathbf{h})$ , (c)  $\mathcal{W}\mathcal{G}_{(2),(1),\delta,(0)}(\mathbf{h})$  y (d)  $\mathcal{W}\mathcal{G}_{(2),(1),(6),\tau}(\mathbf{h})$ .

## Propiedades geométricas de un campo aleatorio

Tal como se puede apreciar en las Figuras 1.1 y 1.2, la elección de los parámetros que controlan la suavidad de las funciones de correlación en el origen tienen un efecto distinto en la diferenciabilidad de la función de correlación en el origen al igual que la elección de un efecto *nugget*. Estas características están asociadas a dos propiedades geométricas relacionadas a los campos aleatorios, estas son:

1. Continuidad en media cuadrática del campo aleatorio:

Un campo aleatorio es continuo en media cuadrática si para una secuencia de localizaciones  $\{\mathbf{s}_n\}$  dentro de un espacio  $B$  para la cual  $\|\mathbf{s}_n - \mathbf{s}\| \rightarrow 0$  cuando  $n \rightarrow \infty$ , entonces:

$$\mathbb{E}[|X(\mathbf{s}_n) - X(\mathbf{s})|^2] \rightarrow 0$$

cuando  $n \rightarrow \infty$  para todo  $\mathbf{s} \in B$ .

2. Diferenciabilidad en media cuadrática del campo aleatorio:

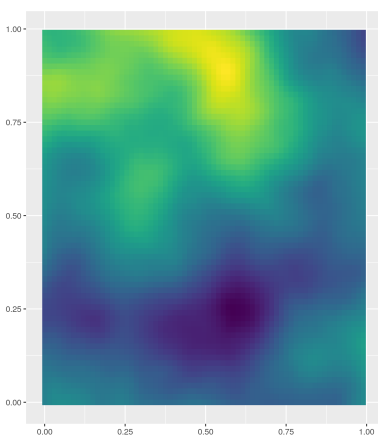
Un campo aleatorio es diferenciable en media cuadrática para cada secuencia de localizaciones  $\{\mathbf{s}_n\}$  dentro de una sección  $B$  para la cual  $\|\mathbf{s}_n - \mathbf{s}\| \rightarrow 0$  mientras  $n \rightarrow \infty$ , entonces:

$$\mathbb{E}[|X_i(\mathbf{s}_n) - X_i(\mathbf{s})|^2] \rightarrow 0$$

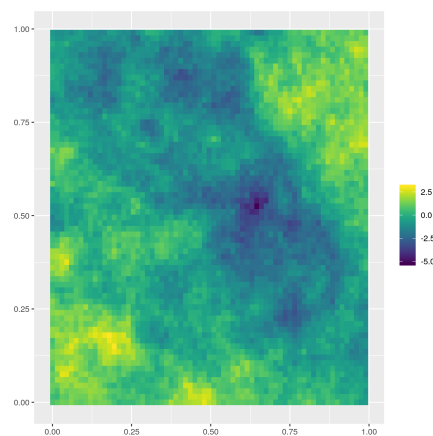
mientras  $n \rightarrow \infty$ ,  $\forall i = 1, \dots, n$ , para todo  $\mathbf{s} \in B$ .

Tal como se explicita en [Abrahamsen \(1997\)](#) ambas propiedades se pueden estudiar para el campo aleatorio equivalentemente al estudio del comportamiento de la función de correlación asociada al campo aleatorio en el origen respecto a su diferenciabilidad y continuidad.

A modo de observar el efecto de las funciones de correlación en las realizaciones de campos aleatorios, se presentan simulaciones de campos aleatorios Gaussianos estándar para apreciar el comportamiento de las realizaciones respecto a las propiedades geométricas asociadas a la función de correlación Matérn (1.19) y Wendland Generalizada (1.20).



(a)



(b)

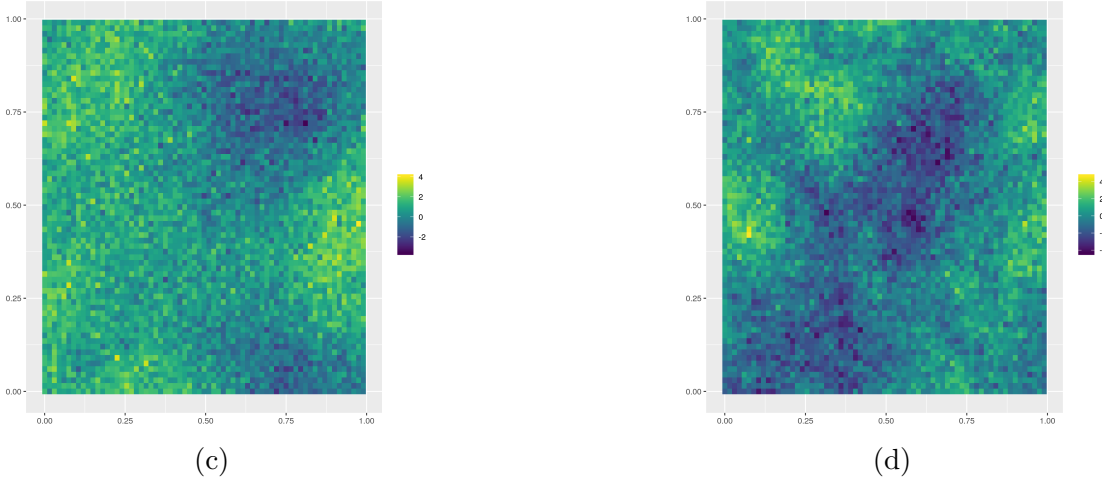
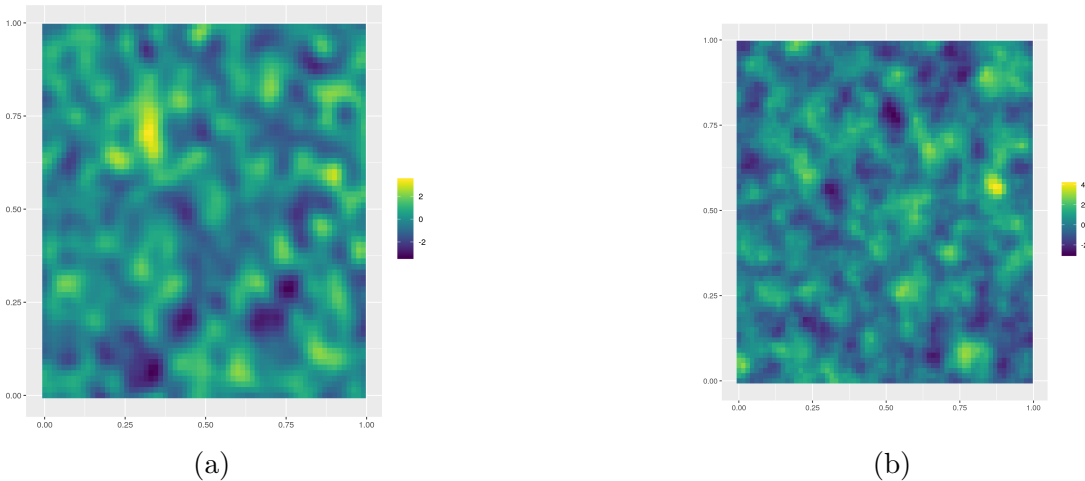


Figura 1.3: Realizaciones de un campo aleatorio Gaussiano estándar con función de correlación Matérn parametrizada por: (a)  $\mathcal{M}_{(0,15),(2),(0)}(\mathbf{h})$ , (b)  $\mathcal{M}_{(0,15),(0,5),(0)}(\mathbf{h})$ , (c)  $\mathcal{M}_{(0,15),(2),(0,5)}(\mathbf{h})$  y (d)  $\mathcal{M}_{(0,15),(0,5),(0,5)}(\mathbf{h})$ .

El comportamiento que corresponde a las realizaciones en los distintos escenarios se puede diferenciar claramente con el uso del parámetro de soporte compacto en la función de correlación Wendland Generalizada ya que para el caso (a) la realización es suave, en la Figura 1.4 se puede apreciar que se forman sectores suaves de distancia similar al valor del soporte compacto.

Además, también se puede notar la diferencia de utilizar funciones de correlación continuas o diferenciables en media cuadrática ya que en los ejemplos se aprecia claramente que considerar o no estas características se refleja en la suavidad de la realización y con efecto *nugget* una imagen no suave o pixelada.



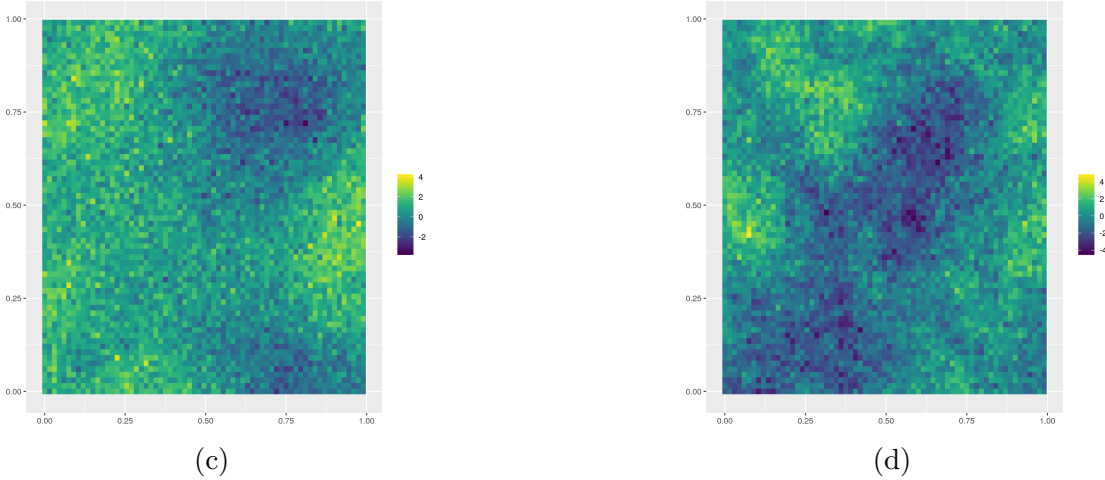


Figura 1.4: Realizaciones de un campo aleatorio Gaussiano estándar con función de correlación Wendland Generalizada parametrizada por: (a)  $\mathcal{W}\mathcal{G}_{(0,15),(3),(4),(0)}(\mathbf{h})$ , (b)  $\mathcal{W}\mathcal{G}_{(0,15),(1),(4),(0)}(\mathbf{h})$ , (c)  $\mathcal{W}\mathcal{G}_{(0,15),(3),(4),(0,5)}(\mathbf{h})$  y (d)  $\mathcal{W}\mathcal{G}_{(0,15),(1),(4),(0,5)}(\mathbf{h})$ .

Para finalizar esta sección, cabe señalar que existen funciones de correlación definidas tanto en el plano (distancia Euclidiana) como en otro tipo de espacios como una esfera (distancias geodésicas), pero aquí se estudiarán campos aleatorios en el plano sin pérdida de generalidad.

#### 1.1.4. Estimación de parámetros

En Estadística Espacial diversos autores han estudiados métodos para la estimación de parámetros asociados a los modelos tales como: el métodos de máxima verosimilitud, validación cruzada, inferencia bayesiana y aproximaciones de funciones de verosimilitud (Bachoc, 2014; Bevilacqua y cols., 2019c). En inferencia estadística el método más utilizado y cuyas propiedades han sido muy estudiadas es el de máxima verosimilitud, sin embargo en el caso de Estadística Espacial no siempre se puede llevar a cabo. El primer trabajo teórico es el de Mardia y Marshall (1984), quienes demuestran la consistencia de las estimaciones bajo campos aleatorios Gaussianos.

Una de las principales razones por las cuales no siempre es posible utilizar el método de máxima verosimilitud se debe a que para campos aleatorios no Gaussianos, es muy difícil obtener expresiones cerradas para las funciones finito dimensionales asociadas a los campos aleatorios. De manera alternativa, una aproximación es el método de la verosimilitud compuesta por parejas ponderada (Varin, Reid, y Firth, 2011) que solo requiere la función de distribución de probabilidad bivariada asociada al modelo.

Dentro de las ventajas de utilizar esta aproximación se encuentran: (a) solo requiere la función bivariada asociada al modelo, (b) puede lidiar de manera eficiente respecto al costo computacional y (c) las estimaciones de los parámetros convergen a las obtenidas por máximo verosimilitud. Además, las estimaciones de los parámetros se obtienen al maximizar una función de (log) verosi-

miltud (1.21) que se define por:

$$pl(\boldsymbol{\theta}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \log(f_{\mathbf{X}_{ij}}(x_i, x_j, \boldsymbol{\theta})) c_{ij} , \quad (1.21)$$

cuyas componentes son:  $\mathbf{x} = [x_i]_{i=1}^n$  un vector de  $n$  observaciones,  $f_{\mathbf{X}_{ij}}$  la función de distribución bivariada asociada a un modelo  $X$ , un vector de parámetros  $\boldsymbol{\theta}$  asociados al modelo y  $c_{ij} \in \mathbb{R}_0^+$ . Entonces, los estimadores se obtienen de  $\hat{\boldsymbol{\theta}} = Arg. max. pl(\boldsymbol{\theta})$ .

En el contexto de campos aleatorios, en específico campos aleatorios Gaussianos, [Bevilacqua, Gaetan, Mateu, y Porcu \(2012\)](#) proponen el estudio de las estimaciones de parámetros bajo el método de verosimilitud compuesta por parejas ponderada definiendo los pesos de la función de verosimilitud ( $c_{ij}$ ) de la siguiente manera:

$$c_{ij} = \begin{cases} 1 & ; \|s_i - s_j\| < dist_{max} , \\ 0 & ; \|s_i - s_j\| \geq dist_{max} , \end{cases} \quad (1.22)$$

donde  $dist_{max}$  es un valor de distancia entre las localizaciones espaciales a criterio de quien utilice el método.

Es necesario señalar que hasta el momento no existe una forma no arbitraria o numérica de escoger los pesos de la función de verosimilitud, pero los resultados teóricos han mostrado que por (1.22) las estimaciones convergen a las obtenidas por máximo verosimilitud y que los tiempos de cómputo se reducen ya que solo se consideran las parejas de localizaciones que cumplen (1.22) ([Bevilacqua y cols., 2012](#)).

Un segundo aspecto importante para la estimación de parámetros en Estadística Espacial es el enfoque asintótico con el que se trabaja ya que para este esquema en la literatura existen dos tipos de dominios y los resultados que se obtienen respecto a las estimaciones son distintos ([Bevilacqua y cols., 2019c](#)). El primer dominio es el enfoque asintótico de dominio creciente (*increasing domain*), el que se caracteriza por mantener una distancia mínima entre las localizaciones espaciales y que para obtener más datos es necesario cubrir una mayor área dentro del espacio (Figura 1.5). Por otro lado, el segundo enfoque, denominado dominio fijo (*infill asymptotic*) es aquel que solo mantiene un área en el espacio fija y que a medida que se deseen más observaciones, es necesario cubrir con más localizaciones dentro del una determinada área (Figura 1.6).

Si bien en este trabajo no se demostrarán distribuciones asintóticas de las estimaciones obtenidas bajo los dos tipos de dominios, se puede acudir a los resultados de [Mardia y Marshall \(1984\)](#); [Bevilacqua y cols. \(2019c\)](#) para el enfoque asintótico de dominio creciente y [Bachoc, Bevilacqua, y Velandia \(2019\)](#) para el enfoque asintótico de dominio fijo y estimación por máxima verosimilitud o verosimilitud compuesta por parejas ponderada junto con todas las referencias dentro de aquellos trabajos para campos aleatorios Gaussianos.

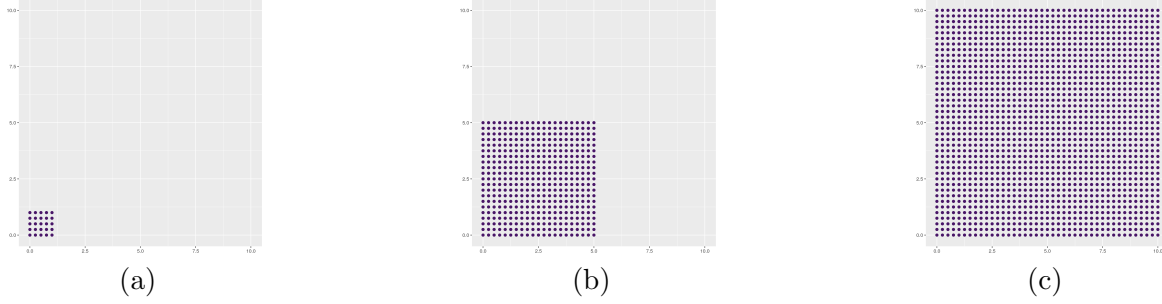


Figura 1.5: Ejemplo de dominio creciente asintótico en una grilla regular de dimensión 10x10 con distancia mínima de 0,25 unidades para: (a) 25, (b) 441 y (c) 1681 localizaciones espaciales.

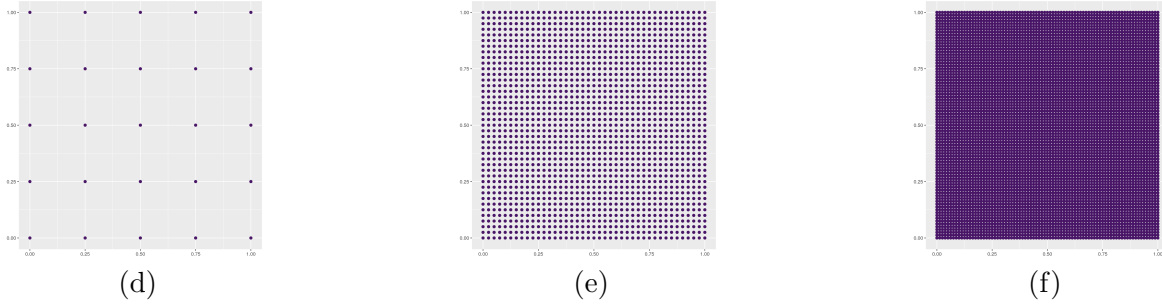


Figura 1.6: Ejemplo de dominio fijo asintótico en una grilla regular de dimensión 1x1 con distancia mínima de 0,25, 0,025 y 0,0125 unidades para: (d) 25, (e) 1681 y (f) 6661 localizaciones espaciales.

### 1.1.5. Predicción

Hasta el momento se han mostrados algunos resultados presentes en la literatura respecto a la construcción y estimación de modelos en campos aleatorios, esta aproximación considera que uno de los aspectos más importantes para la modelización de observaciones espacialmente dependientes es la elección de la función de correlación. Un segundo aspecto importante en la aplicación de estos modelos es la predicción de nuevas observaciones en una localización espacial no muestreada denotada por  $\mathbf{s}_0 \in S$ .

Según [Cressie \(1993\)](#), en Estadística Espacial, el predictor más utilizado hasta el momento es el Kriging, denominado así por [Matheron \(1960\)](#) debido al autor (D. G. Krigde) del primer trabajo sobre el que se tenga registro sobre la modelización de observaciones espacialmente dependientes en minería ([Krigde, 1951](#)). Este método consiste en que la predicción en una nueva localización espacial está dada por:

$$y(\mathbf{s}_0) = \sum_{i=1}^n w_i x(\mathbf{s}_i) + k, \quad (1.23)$$

cuyas componentes son:  $\mathbf{X} = [x(\mathbf{s}_i)]_{i=1}^n$  un vector de  $n$  observaciones del campo aleatorio  $X$ ,  $w_i$  ( $i = 1, \dots, n$ ) pesos sujetos a  $\sum_{i=1}^n w_i = 1$ . Entonces el predictor estimado para la nueva localización espacial ( $\widehat{y(\mathbf{s}_0)}$ ) que como criterio de optimización minimiza el error cuadrático medio ( $\mathbb{E}[(X(\mathbf{s}_0) - \widehat{y(\mathbf{s}_0)})^2]$ )

sobre  $(w_1, \dots, w_n, k)$  está establecido por:

$$\widehat{y(\mathbf{s}_0)} = c^T \Omega^{-1} (\mathbf{X} - \boldsymbol{\mu}) + \mu(\mathbf{s}_0) , \quad (1.24)$$

donde  $c = [\text{Cov}[X(\mathbf{s}_i), X(\mathbf{s}_0)]]_{i=1}^n$ ,  $\Omega = [\text{Cov}[X(\mathbf{s}_i), X(\mathbf{s}_j)]]_{i,j=1}^n$  la matriz de covarianzas,  $\mathbf{X}$  el vector de observaciones y  $\boldsymbol{\mu} = [\mu(\mathbf{s}_i)]_{i=1}^n$  un vector de medias. Además, se puede derivar una medida del error de predicción asociada al predictor (1.23), cuya expresión está dada por:

$$\text{Var}[\widehat{y(\mathbf{s}_0)}] = \text{Cov}[\mathbf{s}_0, \mathbf{s}_0] - c^T \Omega^{-1} c . \quad (1.25)$$

Distintas versiones de (1.23) dependerán de los supuestos sobre  $\mu(\mathbf{s})$ . Si la media es conocida y constante se denomina Kriging simple, si es desconocido el valor se denomina Kriging ordinario y si se asume desconocido y se puede representar por un arreglo de covariables, es decir,  $\mu(\mathbf{s}) = M(\mathbf{s})^T \boldsymbol{\beta}$  es denominado Kriging universal.

La derivación de las expresiones algebraicas se pueden desarrollar para el caso de campos aleatorios Gaussianos considerando la media y varianza condicional de la nueva observación dado el resto de las observaciones. Bajo este escenario el predictor óptimo lineal y el error de estimación coinciden con (1.23) y (1.25) respectivamente.

Para campos aleatorios no Gaussianos estas expresiones algebraicas se vuelven más complicadas ya que no siempre se puede obtener una expresión cerrada para la distribución finito dimensional y los momentos condicionales. Para dar una solución a este problema se puede proponer un predictor lineal óptimo (denotado por  $\widehat{y_{lo}(\mathbf{s}_0)}$ ) (Bevilacqua, Caamaño, Arellano-Valle, y Morales-Oñate, 2019a) de la siguiente manera:

$$\widehat{y_{lo}(\mathbf{s}_0)} = \mu(\mathbf{s}_0) + c_X^T R^{-1} (\mathbf{X} - \boldsymbol{\mu}) , \quad (1.26)$$

cuyas componentes son  $c_X = [\rho_X(\mathbf{s}_i - \mathbf{s}_0)]_{i=1}^n$ ,  $R = [\rho_X(\mathbf{s}_i - \mathbf{s}_j)]_{i,j=1}^n$ . Además, el error de predicción asociado a (1.26) está dado por:

$$\text{Var}[\widehat{y_{lo}(\mathbf{s}_0)}] = \sigma^2 (1 - c_X^T R^{-1} c_X) . \quad (1.27)$$

Considerando entonces esta una alternativa práctica ya que solo depende del conocimiento de la correlación asociada al campo aleatorio  $X$ . Otras soluciones para resolver (1.23) podrían ser como se plantea en Zhang y El-Shaarawi (2010) con métodos de simulación Monte Carlo, pero esta solución es computacionalmente muy costosa.

### 1.1.6. Campos aleatorios no Gaussianos

La introducción de campos aleatorios no Gaussianos (en inglés *non-Gaussian random fields*) se debe a la tarea de modelización de fenómenos u observaciones espacialmente dependientes que presenten diversas características distintas a la distribución Gaussiana como asimetría o colas pesadas, recorridos con soporte en  $\mathbb{R}_0^+$  o compactos e incluso observaciones discretas.

Diversas propuestas de modelización se han estudiado en los últimos años para campos aleatorios con distribuciones marginales continuas, algunos ejemplos de estos modelos pueden ser campos aleatorios con distribuciones marginales con soporte positivo como los campos aleatorios log-Gaussianos

(De Oliveira, 2006), Chi-cuadrado (Ma, 2010a), Birnbaum-Saunders (Garcia-Papani, Leiva, Uribe-Opazo, y Aykroyd, 2018), Gaussiano log-Gaussiano (Zareifard, Khaledi, Rivaz, y Vahidi-Asl, 2018) o Gamma y Weibull (Bevilacqua, Caamaño, y Gaetan, 2019b).

También se ha ampliado el rango de modelos a distribuciones continuas con soporte en los reales, pero que presenten asimetría o colas más o menos pesadas que la distribución Gaussiana tales como campos aleatorios con distribuciones marginales elípticas (Ma, 2010b; Genton y Zhang, 2012), de valor extremo (Sang y Genton, 2014), Gaussiano-envuelto para datos direccionales (Alegría, Bevilacqua, y Porcu, 2016), Gaussiana asimétrica (Zhang y El-Shaarawi, 2010; Alegría, Caro, Bevilacqua, Porcu, y Clarke, 2017), Tukey  $g$ -and- $h$  (Xu y Genton, 2017) o  $t$  de student y  $t$  de student asimétrica (Bevilacqua y cols., 2019a).

Todos estos modelos tienen la particularidad de ser transformaciones de campos aleatorios Gaussianos, permitiendo caracterizar teóricamente los modelos además de poder estudiar las características asociadas a los momentos. Otras aproximaciones a modelos no Gaussianos son representaciones jerárquicas y modelos generalizados mixtos (Diggle y Ribeiro, 2007), pero estas metodologías tienen la desventaja de que introducen efecto *nugget* en la función de correlación asociada por lo que se pierde flexibilidad respecto a escoger o no modelos continuos en media cuadrática cuando esto se puede observar en los conjuntos de datos; además, no existe una clara fundamentación teórica al respecto como demostraciones de existencia de estos procesos (Abrahamsen, 1997).

Con respecto a campos aleatorios con distribuciones marginales discretas, existen muy pocos trabajos. Para la modelización de observaciones dicotómicas espacialmente dependientes se ha estudiado una versión espacial para el modelo probit (Heagerty y Lele, 1998). Otros modelos de conteo con distribuciones tipo Poisson o binomial se han estudiado como construcciones jerárquicas a través de modelos lineales generalizados mixtos (Diggle y Ribeiro, 2007), pero estos modelos presentan algunas características no tan deseadas desde el punto de vista de la flexibilidad que entregan (inducir efecto *nugget*), por lo que nuevas metodologías de construcción de campos aleatorios con distribuciones marginales discretas son un área de investigación aún abierta.

### 1.1.7. Propuesta de investigación

Dentro de la investigación para campos aleatorios con distribuciones marginales continuas, colas pesadas y posible asimetría que se propone en la tesis doctoral de Caamaño (2018) se introducen dos modelos contruidos a partir de transformaciones de campos aleatorios  $t$  de student. Estos nuevos modelos son: el campos aleatorios  $t$  de student asimétricos y *two-piece*  $t$  de student.

Si bien el modelo  $t$  de student asimétrico generaliza al modelo  $t$  de student y permite el análisis de observaciones espacialmente dependientes de colas pesadas y asimetría, la falta de expresiones cerradas para la función de distribución lo convierten en un modelos que existe hasta el momento solo en la teoría.

A diferencia del modelo antes mencionado, el modelo *two-piece*  $t$  de student permite la estimación de parámetros y además extenderlo a una clase general de campos aleatorios asimétricos dado que su construcción está dada por:

$$P_{\eta}(\mathbf{s}) = \mu(\mathbf{s}) + \sigma|X(\mathbf{s})|K_{\eta}(\mathbf{s}) ,$$

cuyas componentes se definirán en detalle en el Capítulo (2), pero lo importante es que bastaría reemplazar en  $X$  cualquier proceso con distribuciones marginales simétricas como un campo aleatorio Gaussiano para obtener modelos con colas pesadas o livianas y posibles grados de asimetría en las distribuciones marginales.

Dado que hasta el momento no existe en la literatura campos aleatorios con distribuciones marginales *two-piece*, en este trabajo se propone entonces estudiar algunos casos particulares para establecerlo como una clase de campos aleatorios para el estudio de observaciones espacialmente dependientes con asimetría o colas pesadas.

## 1.2. Notas del capítulo

A continuación se dará un resumen de la investigación que se desarrolla en este trabajo junto con una breve descripción del contenido de cada capítulo.

A manera resumen, en este trabajo se estudia la metodología de construcción, estimación y predicción para campos aleatorios con distribuciones marginales continuas que permitan asimetría y posible presencia de colas pesadas a partir de la clase general de modelos *two-piece*. Además, se estudian las propiedades derivadas de la función de correlación asociada a los campos aleatorios junto con el estudio de las funciones de distribución de probabilidad bivariadas asociadas a casos particulares.

Lo que continua del documento se dividirá en cuatro capítulos que abordarán lo siguiente:

1. En el capítulo dos se estudiará la metodología para la construcción general de la clase general de campos aleatorios *two-piece*, dando paso al estudio de su distribución uni y bivariada. Además, se estudiarán las propiedades geométricas a partir de la función de correlación. Luego de definido el campo aleatorio *two-piece*, se detallarán tres casos particulares.
2. Una vez que se estudian los modelos descritos en el capítulo dos, en el capítulo tres se realizarán simulaciones para estudiar el comportamiento de las estimaciones de los parámetros de los modelos. Se finalizará este capítulo con una aplicación a los datos de cambios de divisas en el tiempo, disponibles en *Forex*.
3. Este documento terminará con algunas conclusiones de la investigación y la discusión respecto a posibles problemas abiertos que se generan a partir de la clase general *two-piece*.

Adicionalmente, se proporcionarán dos apéndices que contienen las demostraciones de algunas expresiones algebraicas y parte de las rutinas computacionales para la reproducción de resultados.

## Capítulo 2

# Una clase general de campos aleatorios

El capítulo dos comprende la presentación del modelo probabilístico *two-piece*, una revisión de algunos trabajos en la literatura que contemplan el origen y algunas características asociadas a este. Seguido de la presentación del modelo para variables aleatorias, se desarrollará la metodología para la extensión del modelo a campos aleatorios junto con el estudio de algunas características de los procesos como los momentos y las funciones de distribución de probabilidad. Se finalizará el capítulo con el estudio de tres casos particulares para la clase de modelos propuesta, comparaciones y algunas conclusiones respecto a los modelos aquí presentados.

## 2.1. Aspectos generales asociados al modelo

El modelo probabilístico *two-piece* fue introducido en una primera versión en 1860 para el análisis de datos astronómicos, botánicos y meteorológicos, entre otros; debido a que el supuesto de Gaussianidad no era tan válido dada la presencia de asimetría en la distribución de frecuencia de las observaciones (Wallis, 2014). Este modelo fue propuesto como el resultado de unir y reescalar dos distribuciones Gaussianas con igual parámetro de localización y distintos parámetros de escala.

Su construcción se define para dos variables aleatorias Gaussianas con parámetro de localización  $\mu \in \mathbb{R}$  y parámetros de escala  $\sigma_1^2$  y  $\sigma_2^2 \in \mathbb{R}^+$ , respectivamente. Entonces, la función de distribución de probabilidad para el nuevo modelo queda establecida por:

$$f(g) = \begin{cases} \frac{2}{(\sigma_1 + \sigma_2)\sqrt{2\pi}} \exp\left\{-\frac{(g-\mu)^2}{2\sigma_1^2}\right\} & ; \quad g < \mu \quad , \\ \frac{2}{(\sigma_1 + \sigma_2)\sqrt{2\pi}} \exp\left\{-\frac{(g-\mu)^2}{2\sigma_2^2}\right\} & ; \quad g > \mu \quad , \\ \mu & ; \quad g = \mu \quad . \end{cases}$$

Dada su definición, el modelo *two-piece* Gaussiano permite generar una distribución continua y asimétrica dependiendo de la elección de los parámetros de escala de las distribuciones Gaussianas que lo conforman. Un ejemplo gráfico se puede ver en la Figura 2.1 en la cual se obtiene un modelo con asimetría a la izquierda (cuando  $\sigma_1 < \sigma_2$ ).

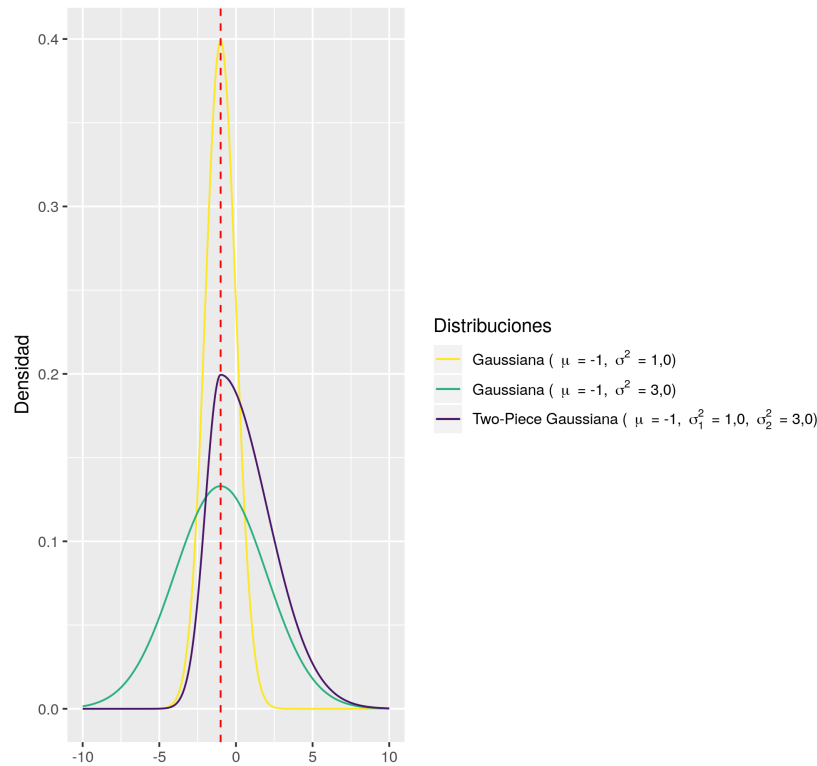


Figura 2.1: Comparación funciones de distribución de probabilidad Gaussianas y *two-piece* resultante.

Tal como lo explicita [Wallis \(2014\)](#), el modelo ha adoptado distintas versiones como en [Fernandez y Steel \(1998\)](#) para modelos asimétricos de colas pesadas. Particularmente, uno de los trabajos sobre la unificación del modelo es el de [Arellano-Valle, Gómez, y Quintana \(2005\)](#), quienes lo proponen como una clase general de modelos asimétricos con colas más o menos pesadas que la distribución Gaussiana y que además se deriva de una sencilla representación estocástica para su simulación.

Esta clase general posee la capacidad de modelización flexible respecto a distintos grados de asimetría presente en las observaciones, una característica bastante estudiada en la literatura como el modelo Gaussiano asimétrico ([Azzalini y Capitanio, 2014](#)) y posteriores derivaciones de este ya que en aplicaciones, la Gaussianidad no siempre está presente. Además de ser un modelo flexible respecto a la asimetría de la distribución, la clase general *two-piece* recibe su adjetivo “general” debido a que puede adoptar distintas formas o casos particulares dada su construcción, permitiendo establecer modelos probabilísticos con asimetría además de colas más o menos pesadas que el modelo Gaussiano y considerando volver a modelos simétricos también como casos especiales dentro del conjunto paramétrico del modelo .

La incorporación de características como asimetría, curtosis o soportes compactos a los modelos de probabilidad ha sido bastante abordado en los modelos para variables aleatorias tanto para el caso univariado como multivariado, pero obtener estas características deseables o flexibles en los modelos no siempre es una tarea sencilla cuando los modelos incorporan estructuras de dependencia entre las observaciones, esta dificultad se debe principalmente a que no siempre es posible conseguir procesos teóricamente fundamentados o existentes ([Abrahamsen, 1997](#)).

Respecto al modelo probabilístico *two-piece*, existe un primer acercamiento en el contexto de procesos espaciales o campos aleatorios en la tesis doctoral de [Caamaño \(2018\)](#), quien propone la metodología para la construcción, estimación y predicción para campos aleatorios *two-piece* t de student en un completo estudio de procesos espaciales con distribuciones marginales asimétricas y de colas pesadas. Hasta el momento, no existe registro de algún trabajo publicado que proponga otros casos particulares de campos aleatorios *two-piece*.

En esta sección se mostrará la clase general de modelos probabilísticos *two-piece* presentado por [Arellano-Valle y cols. \(2005\)](#), continuando con la entrega y detalle de las expresiones para las funciones de distribución de probabilidad uni y bivariada, junto con el desarrollo metodológico para la extensión del modelo a campos aleatorios y las respectivas expresiones algebraicas (generales) para la función de covarianza entre dos localizaciones espaciales según el trabajo de [Caamaño \(2018\)](#).

Un segundo análisis de la clase general de campos aleatorios *two-piece* se orientará al estudio de algunas características deseables en modelos para procesos espaciales que evalúan la flexibilidad de adaptación de estos cuando son utilizados en contextos de aplicaciones, las dos características principales que se evaluarán aquí serán la continuidad y diferenciabilidad en media cuadrática, ambas evaluadas al estudiar la función de covarianza del proceso.

Finalizando el capítulo, se estudiarán tres casos particulares del modelo. En primer lugar el caso t de student presentado por [Caamaño \(2018\)](#), seguido de la versión Gaussiana del modelo *two-piece* y un tercer caso a partir de campos aleatorios con distribuciones marginales del tipo Tukey-*h* ([Goerg, 2015](#)) junto con algunas conclusiones y posibilidades que se pueden generar a partir de la formulación general del campo aleatorio.

### 2.1.1. Modelo probabilístico general

En el trabajo de [Arellano-Valle y cols. \(2005\)](#) se estableció una clase general del modelo probabilístico *two-piece*, trabajo en el cual proponen una manera de describirlo completamente con la siguiente función de distribución de probabilidad:

$$f(t) = \frac{2}{a(\eta) + b(\eta)} \left\{ f_X \left( \frac{t}{a(\eta)} \right) I(t \geq 0) + f_X \left( \frac{t}{b(\eta)} \right) I(t < 0) \right\} , \quad (2.1)$$

donde  $\eta$  es un parámetro de asimetría,  $a(\eta)$  y  $b(\eta)$  dos funciones conocidas y  $X$  una variable aleatoria simétrica con soporte en  $\mathbb{R}$ , aunque los autores declaran que la construcción (2.1) puede ser utilizada para variables aleatorias con soporte en  $\mathbb{R}^+$ .

Una de las ventajas de esta construcción general es que el modelo contiene a la clase completa de modelos simétricos si  $a(\eta) = b(\eta)$ , además de contener distintas parametrizaciones del modelo *two-piece* dependiendo de las especificaciones de  $a(\eta)$  y  $b(\eta)$ .

Dentro del mismo estudio, [Arellano-Valle y cols. \(2005\)](#) presentan una manera sencilla de simular cualquier modelo perteneciente a la clase *two-piece* a partir de la siguiente representación estocástica:

$$P_\eta = \mu + \sigma |X| K_\eta , \quad (2.2)$$

cuyas componentes son: un parámetro de localización  $\mu \in \mathbb{R}$ , un parámetro de escala  $\sigma \in \mathbb{R}^+$ , un parámetro de asimetría  $\eta$ ,  $X$  una variable aleatoria perteneciente a la familia de modelos simétricos como el modelo Gaussiano,  $t$  de student, logístico, exponencial potencia, entre otros; y una variable aleatoria discreta  $K_\eta$  que permite generar la asimetría en la variable aleatoria  $P_\eta$ , esta variable aleatoria discreta se describe de la siguiente manera:

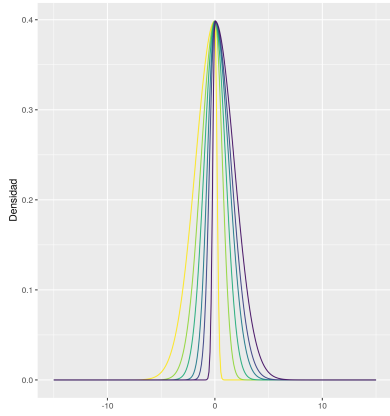
$$K_\eta = \begin{cases} a(\eta) & ; & x < \mu , \\ -b(\eta) & ; & x \geq \mu , \end{cases}$$

entonces, si  $P_\eta$  es una variable aleatoria *two-piece* de localización y escala, su función de distribución de probabilidad está dada por:

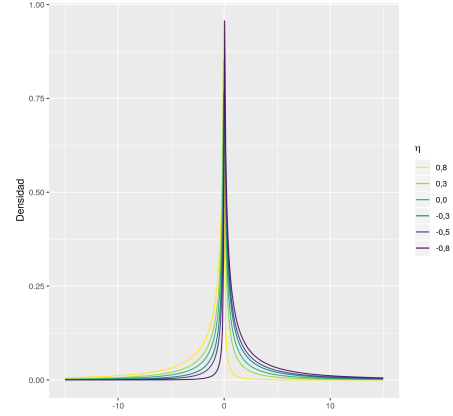
$$f_{P_\eta}(t) = \frac{2}{\sigma(a(\eta) + b(\eta))} \left\{ f_X \left( \frac{t - \mu}{\sigma a(\eta)} \right) I(t \geq \mu) + f_X \left( \frac{t - \mu}{\sigma b(\eta)} \right) I(t < \mu) \right\} , \quad (2.3)$$

En el artículo de [Arellano-Valle y cols. \(2005\)](#), si bien los autores mencionan distintas parametrizaciones para  $a(\eta)$  y  $b(\eta)$  junto con las conexiones entre las alternativas de modelos *two-piece*, ellos estudian el modelo (2.2) para  $a(\eta) = 1 - \eta$  y  $b(\eta) = 1 + \eta$ , con  $\eta \in \mathbb{R}_{(-1,1)}$  debido al comportamiento de la función de distribución de probabilidad (2.3) en los valores límites del espacio paramétrico del parámetro de asimetría.

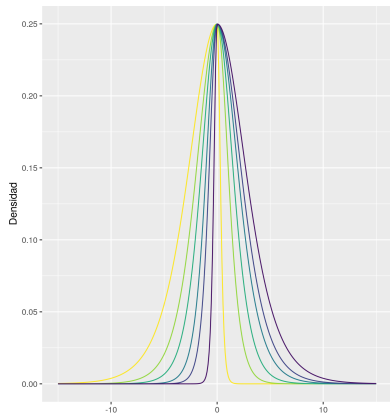
Casos particulares del modelo se pueden obtener al considerar distintos modelos simétricos, permitiendo así generar modelos flexibles respecto al grado de asimetría y además de ser flexible respecto a la pesadez de las colas del modelo, algunas ilustraciones se pueden apreciar en la Figura 2.2.



(a)



(b)



(c)

Figura 2.2: Comparación funciones de distribución de probabilidad *two-piece*: (a) Gaussiana estándar, (b) exponencial potencia ( $\alpha = 0,3$ ) y (c) logística estándar.

Extensiones del modelo *two-piece* para el caso de variables aleatorias se pueden encontrar en [Maleki y Mahmoudi \(2017\)](#) al considerar mixtura en localización y escala de distribuciones Gaussianas. También en contextos de regresión, en el trabajo de [Maleki, Contreras-Reyes, y Mahmoudi \(2019\)](#) considerando errores del tipo *two-piece* y en [Arellano-Valle, Azzalini, Ferreira, y Santoro \(2020\)](#) para modelos con errores en las variables, así como [Moravveji, Khodadadi, y Maleki \(2019\)](#) para la estimación bayesiana del modelo. Para una revisión completa de otras extensiones del modelo, se pueden estudiar las referencias dentro de los artículos anteriormente mencionados.

## 2.2. Extensión a campos aleatorios

Tal como lo señala [Caamaño \(2018\)](#), el modelo probabilístico *two-piece* (2.2) puede ser extendido a un proceso espacial o campo aleatorio al modificar su representación estocástica, definiendo así un campo aleatorio  $P_\eta = \{P_\eta(\mathbf{s}), \mathbf{s} \in S\}$  con distribuciones marginales *two-piece*. El proceso *two-piece* se construye entonces, tal como se explicitó en el Capítulo 1, de la siguiente manera

$$P_\eta(\mathbf{s}) = \mu(\mathbf{s}) + \sigma|X(\mathbf{s})|K_\eta(\mathbf{s}) . \quad (2.4)$$

De la representación (2.4), se puede notar que las componentes del modelo, a diferencia de (2.2), varían por cada localización espacial  $\mathbf{s} \in S$ . Esta nueva representación estocástica está compuesta por un parámetro de asimetría  $\eta \in \mathbb{R}_{(-1,1)}$ , una función de localización,  $\mu(\mathbf{s}) \in \mathbb{R}$ , un parámetro de escala  $\sigma \in \mathbb{R}^+$ . Además,  $X = \{X(\mathbf{s}), \mathbf{s} \in S\}$  es un campo aleatorio débilmente estacionario con distribuciones marginales simétricas de localización 0 y escala 1, y  $K_\eta = \{K_\eta(\mathbf{s}), \mathbf{s} \in S\}$  un campo aleatorio discreto definido por:

$$K_\eta(\mathbf{s}) = \begin{cases} 1 - \eta & ; & H(\mathbf{s}) < q_{g(\eta)} , \\ -1 - \eta & ; & H(\mathbf{s}) \geq q_{g(\eta)} . \end{cases} \quad (2.5)$$

A diferencia del modelo (2.2), [Caamaño \(2018\)](#) sugiere un cambio en la función de perturbaciones  $K_\eta$  al realizar la versión espacial del modelo. Este cambio está dado por considerar que  $K_\eta$  es un campo aleatorio y no una variable aleatoria, aludiendo así al argumento de [Genton y Zhang \(2012\)](#) acerca de la identificabilidad de parámetros de procesos espaciales, por lo que es necesario que algunas componentes de la representación estocástica del modelo sean espacialmente dependientes ya que de manera contraria, los parámetros no son consistentemente estimables dado que en las observaciones que se obtienen de una realización del proceso (lo equivalente a decir que el tamaño de muestra es 1), no se encuentra suficiente información para la estimación de los parámetros.

El proceso discreto (2.5) a su vez se compone de un segundo campo aleatorio  $H = \{H(\mathbf{s}), \mathbf{s} \in S\}$ , independiente de  $X$ . Por simplicidad de las expresiones para las funciones de distribución y covarianza de los campos aleatorios *two-piece*,  $H$  tiene la misma función de covarianza que  $X$  y además,  $H$  se define como un campo aleatorio débilmente estacionario con distribuciones marginales Gaussianas estándar. La segunda componente de (2.5) es  $q_\alpha$ , que es el cuantil  $\alpha$  de la distribución marginal de  $H$  y  $g_\eta = (1 - \eta)/2$ .

Con esta nueva representación estocástica (2.4), el modelo *two-piece* ya no es una variable aleatoria si no que un campo aleatorio o proceso espacial, permitiendo así la modelización de observaciones espacialmente dependientes continuas y con distintos grados de asimetría.

La definición general del campo aleatorio *two-piece* (2.4) aquí utilizada no es restrictiva a proponer modificaciones como cambiar el índice ( $\mathbf{s}$ ) en algunas de sus componentes. Por ejemplo, se podría considerar que el parámetro de escala del modelo ( $\sigma$ ) también puede variar en el espacio ( $\sigma(\mathbf{s})$ ). Respecto a la definición del proceso discreto (2.5), también se podría considerar que  $H$  es un campo aleatorio con otro tipo de distribuciones marginales o también que posee otra función de covarianza.

### 2.2.1. Momentos de primer y segundo orden

Una vez definida la clase general de campos aleatorios *two-piece* se pueden calcular expresiones para la media, varianza y función de covarianza del proceso, todas a partir de la representación estocástica (2.4). Parte del cálculo de las expresiones involucra conocer los momentos del campo aleatorio discreto (2.5) así como también es necesario estudiar algunos procesos auxiliares como  $K_\eta(\mathbf{s})^2$  y  $K_\eta(\mathbf{s}_i)K_\eta(\mathbf{s}_j)$  para obtener las expresiones algebraicas relacionadas a los momentos de los campos aleatorios *two-piece*. Para la consulta de las demostraciones sobre las expresiones aquí presentadas, acudir al apéndice A.

De la definición del campo aleatorio discreto  $K_\eta$  presentada en (2.5), se puede obtener el valor esperado y la varianza dando origen a las siguientes expresiones:

$$\mathbb{E}[K_\eta(\mathbf{s})] = -2\eta \quad ,$$

$$\text{Var}[K_\eta(\mathbf{s})] = 1 - \eta^2 \quad .$$

Así entonces, las expresiones generales para el valor esperado y la varianza de un campo aleatorio *two-piece* están dadas por:

$$\mathbb{E}[P_\eta(\mathbf{s})] = \mu(\mathbf{s}) - 2\eta\sigma\mathbb{E}[|X(\mathbf{s})|] \quad ,$$

$$\text{Var}[P_\eta(\mathbf{s})] = \sigma^2 \{ (1 + 3\eta^2)[\text{Var}[|X(\mathbf{s})|] + \mathbb{E}[|X(\mathbf{s})|^2]] - 4\eta^2\mathbb{E}[|X(\mathbf{s})|]^2 \} \quad .$$

De la misma manera, considerando un par de localizaciones espaciales  $\mathbf{s}_i$  y  $\mathbf{s}_j \in S$ , se obtiene una manera de computar la covarianza entre estas dos localizaciones.

$$\text{Cov}[P_\eta(\mathbf{s}_i), P_\eta(\mathbf{s}_j)] = \sigma^2 \{ (4\pi_{11}(\mathbf{h}) - 1 + 3\eta^2 + 2\eta)\mathbb{E}[|X(\mathbf{s}_i)||X(\mathbf{s}_j)|] - 4\eta^2\mathbb{E}[|X(\mathbf{s})|]^2 \} \quad ,$$

donde  $\pi_{11}(\mathbf{h}) = F_{\mathbf{H}_{i,j}}(q_{g_\eta}, q_{g_\eta}, \rho(\mathbf{h}))$  y es  $F_{\mathbf{H}_{i,j}}$  la función de distribución acumulada bivariada Gaussiana estándar asociada al campo aleatorio latente  $H$ .

Finalmente, si se desea obtener una expresión para la correlación entre dos localizaciones espaciales para el proceso *two-piece*, basta con calcular la siguiente expresión:

$$\text{Corr}[P_\eta(\mathbf{s}_i), P_\eta(\mathbf{s}_j)] = \frac{\text{Cov}[P_\eta(\mathbf{s}_i), P_\eta(\mathbf{s}_j)]}{\sqrt{\text{Var}[P_\eta(\mathbf{s}_i)]}\sqrt{\text{Var}[P_\eta(\mathbf{s}_j)]}} \quad .$$

Un aspecto importante de la función de covarianza asociada a los campos aleatorios *two-piece* es que no depende del signo del parámetro de asimetría, problema que también es mencionado y solucionado por Zhang y El-Shaarawi (2010) para los campos aleatorios Gaussianos asimétricos.

### 2.2.2. Funciones de distribución

#### Funciones de distribución univariada

A partir de la formulación general del modelo (2.4) se pueden derivar las expresiones para las funciones de distribución de probabilidad y funciones de distribución acumulada para el modelo

propuesto tanto para el caso univariado como bivariado.

Mediante el método de la función de distribución acumulada de transformaciones de variables aleatorias (Dobrow, 2014), se puede computar primero la función de distribución acumulada correspondiente al proceso espacial *two-piece* definido en (2.4), la cual está dada por:

$$F_{P_\eta}(t) = \begin{cases} (1 + \eta)F_X\left(\frac{t-\mu}{\sigma(1+\eta)}\right) & ; \quad t < \mu, \\ \eta + (1 - \eta)F_X\left(\frac{t-\mu}{\sigma(1-\eta)}\right) & ; \quad t \geq \mu. \end{cases} \quad (2.6)$$

Para obtener la función de distribución de probabilidad del proceso *two-piece*, basta con derivar cada componente de la construcción (2.6), obteniéndose:

$$f_{P_\eta}(t) = \begin{cases} \frac{1}{\sigma}f_X\left(\frac{t-\mu}{\sigma(1+\eta)}\right) & ; \quad t < \mu, \\ \frac{1}{\sigma}f_X\left(\frac{t-\mu}{\sigma(1-\eta)}\right) & ; \quad t \geq \mu. \end{cases} \quad (2.7)$$

Tanto para (2.6) como para (2.7) las funciones  $F_X$  y  $f_X$  corresponden a las funciones de distribución acumulada y función de distribución de probabilidad asociadas a la distribución marginal del campo aleatorio simétrico  $X$  que se escoja como caso particular de (2.4).

## Funciones de distribución bivariada

Al igual que para el caso univariado, se puede definir la representación estocástica del proceso *two-piece* para cada componente de un vector bivariado:

$$\mathbf{P}_{\eta;ij}(\mathbf{s}) = [\mu(\mathbf{s}_i) + \sigma|X(\mathbf{s}_i)|k_\eta(\mathbf{s}_i), \mu(\mathbf{s}_j) + \sigma|X(\mathbf{s}_j)|k_\eta(\mathbf{s}_j)]^T. \quad (2.8)$$

Entonces, para  $\rho$  la correlación asociada al vector aleatorio  $\mathbf{X}_{ij} = [X_i, X_j]^T$ , con  $F_H$  y  $F_{\mathbf{H}_{ij}}$  las funciones de distribución acumulada uni y bivariada Gaussiana estándar asociadas al campo aleatorio  $H$  y al vector bivariado  $[H_i, H_j]^T$  respectivamente. Además de  $F_{|X|}$  y  $F_{|\mathbf{X}_{ij}|}$  las funciones de distribución acumulada uni y bivariada estándar asociadas al campo aleatorio  $|X| = \{|X(\mathbf{s})|, \mathbf{s} \in S\}$  y vector bivariado  $[|X_i|, |X_j|]^T$  respectivamente, cuyas expresiones son:

$$F_{|X|}(x) = 2F_X(x) - 1,$$

$$F_{|\mathbf{X}_{ij}|}(x_i, x_j, \rho) = F_{\mathbf{X}_{ij}}(x_i, x_j, \rho) + F_{\mathbf{X}_{ij}}(-x_i, -x_j, \rho) - F_{\mathbf{X}_{ij}}(-x_i, x_j, \rho) - F_{\mathbf{X}_{ij}}(x_i, -x_j, \rho),$$

la función de distribución acumulada bivariada está dada por:

$$F_{\mathbf{P}_{\eta;ij}}(t_i, t_j, \rho) = \begin{cases} \left\{ 1 + F_{|\mathbf{X}_{ij}|} \left( \frac{-t_i + \mu_i}{\sigma(1+\eta)}, \frac{-t_j + \mu_j}{\sigma(1+\eta)}, \rho \right) - F_{|X|} \left( \frac{-t_i + \mu_i}{\sigma(1+\eta)} \right) - F_{|X|} \left( \frac{-t_j + \mu_j}{\sigma(1+\eta)} \right) \right\} \\ \left\{ 1 + F_{\mathbf{H}_{ij}}(q_g(\eta), q_g(\eta), \rho) - 2F_H(q_g(\eta)) \right\}, \\ t_i \leq \mu_i, t_j \leq \mu_j. \\ \\ \left\{ 1 + F_{|\mathbf{X}_{ij}|} \left( 0, \frac{-t_j + \mu_j}{\sigma(1+\eta)}, \rho \right) - F_{|X|} \left( \frac{-t_j + \mu_j}{\sigma(1+\eta)} \right) \right\} \left\{ 1 + F_{\mathbf{H}_{ij}}(q_g(\eta), q_g(\eta), \rho) - 2F_H(q_g(\eta)) \right\} \\ + \left\{ F_H(q_g(\eta)) - F_{\mathbf{H}_{ij}}(q_g(\eta), q_g(\eta), \rho) \right\} \left\{ F_{|X|} \left( \frac{t_i - \mu_i}{\sigma(1-\eta)} \right) - F_{|\mathbf{X}_{ij}|} \left( \frac{t_i - \mu_i}{\sigma(1-\eta)}, \frac{-t_j + \mu_j}{\sigma(1+\eta)}, \rho \right) \right\}, \\ t_i > \mu_i, t_j \leq \mu_j. \\ \\ \left\{ 1 + F_{|\mathbf{X}_{ij}|} \left( \frac{-t_i + \mu_i}{\sigma(1+\eta)}, 0, \rho \right) - F_{|X|} \left( \frac{-t_i + \mu_i}{\sigma(1+\eta)} \right) \right\} \left\{ 1 + F_{\mathbf{H}_{ij}}(q_g(\eta), q_g(\eta), \rho) - 2F_H(q_g(\eta)) \right\} \\ + \left\{ F_H(q_g(\eta)) - F_{\mathbf{H}_{ij}}(q_g(\eta), q_g(\eta), \rho) \right\} \left\{ F_{|X|} \left( \frac{t_j - \mu_j}{\sigma(1-\eta)} \right) - F_{|\mathbf{X}_{ij}|} \left( \frac{-t_i + \mu_i}{\sigma(1+\eta)}, \frac{t_j - \mu_j}{\sigma(1-\eta)}, \rho \right) \right\}, \\ t_i \leq \mu_i, t_j > \mu_j. \\ \\ \left\{ 1 + F_{\mathbf{H}_{ij}}(q_g(\eta), q_g(\eta), \rho) - 2F_H(q_g(\eta)) \right\} + \\ \left\{ F_H(q_g(\eta)) - F_{\mathbf{H}_{ij}}(q_g(\eta), q_g(\eta), \rho) \right\} \left\{ F_{|X|} \left( \frac{t_j - \mu_j}{\sigma(1-\eta)} \right) - F_{|\mathbf{X}_{ij}|} \left( \frac{t_i - \mu_i}{\sigma(1-\eta)}, 0, \rho \right) \right\} + \\ \left\{ F_H(q_g(\eta)) - F_{\mathbf{H}_{ij}}(q_g(\eta), q_g(\eta), \rho) \right\} \left\{ F_{|X|} \left( \frac{t_j - \mu_j}{\sigma(1-\eta)} \right) - F_{|\mathbf{X}_{ij}|} \left( 0, \frac{t_j - \mu_j}{\sigma(1-\eta)}, \rho \right) \right\} + \\ \left\{ F_{|\mathbf{X}_{ij}|} \left( \frac{t_i - \mu_i}{\sigma(1-\eta)}, \frac{t_j - \mu_j}{\sigma(1-\eta)}, \rho \right) F_{\mathbf{H}_{ij}}(q_g(\eta), q_g(\eta), \rho) \right\}, \\ t_i > \mu_i, t_j > \mu_j. \end{cases} \quad (2.9)$$

Finalmente, la función de distribución de probabilidad *two-piece* bivariada se puede obtener de manera similar al caso univariado pero realizando las derivadas parciales de la función de distribución acumulada bivariada *two-piece* (2.9) resultando:

$$f_{\mathbf{P}_{\eta;ij}}(t_i, t_j, \rho) = \begin{cases} \frac{(F_{\mathbf{H}_{ij}}(q_g(\eta), q_g(\eta), \rho) + \eta)}{\sigma^2(1+\eta)^2} f_{|\mathbf{X}_{ij}|} \left( \frac{-t_i + \mu_i}{\sigma(1+\eta)}, \frac{-t_j + \mu_j}{\sigma(1+\eta)}, \rho \right) & , t_i \leq \mu_i, t_j \leq \mu_j. \\ \\ \frac{(1-\eta-2F_{\mathbf{H}_{ij}}(q_g(\eta), q_g(\eta), \rho))}{2\sigma^2(1-\eta)^2} f_{|\mathbf{X}_{ij}|} \left( \frac{t_i - \mu_i}{\sigma(1-\eta)}, \frac{-t_j + \mu_j}{\sigma(1+\eta)}, \rho \right) & , t_i > \mu_i, t_j \leq \mu_j. \\ \\ \frac{(1-\eta-2F_{\mathbf{H}_{ij}}(q_g(\eta), q_g(\eta), \rho))}{2\sigma^2(1-\eta)^2} f_{|\mathbf{X}_{ij}|} \left( \frac{-t_i + \mu_i}{\sigma(1+\eta)}, \frac{t_j - \mu_j}{\sigma(1-\eta)}, \rho \right) & , t_i \leq \mu_i, t_j > \mu_j. \\ \\ \frac{F_{\mathbf{H}_{ij}}(q_g(\eta), q_g(\eta), \rho)}{\sigma^2(1-\eta)^2} f_{|\mathbf{X}_{ij}|} \left( \frac{t_i - \mu_i}{\sigma(1-\eta)}, \frac{t_j - \mu_j}{\sigma(1-\eta)}, \rho \right) & , t_i > \mu_i, t_j > \mu_j. \end{cases}$$

### 2.2.3. Propiedades geométricas

Respecto a las propiedades geométricas de los campos aleatorios *two-piece*, a continuación se estudiará la continuidad y diferenciabilidad en media cuadrática del proceso, ambas a partir de la función de correlación asociada. La expresión algebraica para la función de correlación está dada por:

$$\rho_{P_\eta}(\mathbf{h}) = \frac{(4\pi_{11}(\mathbf{h}) - 1 + 3\eta^2 + 2\eta)\mathbb{E}[|X(\mathbf{s}_i)||X(\mathbf{s}_j)|] - 4\eta^2\mathbb{E}[|X(\mathbf{s})|]^2}{(1 + 3\eta^2)[\text{Var}[|X(\mathbf{s})|] + \mathbb{E}[|X(\mathbf{s})|^2] - 4\eta^2\mathbb{E}[|X(\mathbf{s})|]^2}} . \quad (2.10)$$

Para evaluar la continuidad en media cuadrática del proceso, se debe analizar el valor de la función de correlación asociada en el origen, es decir, cuando  $\mathbf{h} = 0$ . El resultado de esto está dado por:

$$\rho_{P_\eta}(0) = \frac{(4\pi_{11}(0) - 1 + 3\eta^2 + 2\eta)\mathbb{E}[|X(\mathbf{s})|^2] - 4\eta^2\mathbb{E}[|X(\mathbf{s})|]^2}{(1 + 3\eta^2)[\text{Var}[|X(\mathbf{s})|] + \mathbb{E}[|X(\mathbf{s})|^2] - 4\eta^2\mathbb{E}[|X(\mathbf{s})|]^2}} . \quad (2.11)$$

El análisis de la expresión anterior debe considerar la siguiente equivalencia  $\text{Var}[|X(\mathbf{s})|] + (\mathbb{E}[|X(\mathbf{s})|])^2 = \mathbb{E}[|X(\mathbf{s})|^2]$ , por lo que se simplifica la expresión algebraica a:

$$\rho_{P_\eta}(0) = \frac{(4\pi_{11}(0) - 1 + 3\eta^2 + 2\eta)\mathbb{E}[|X(\mathbf{s})|^2] - 4\eta^2\mathbb{E}[|X(\mathbf{s})|]^2}{(1 + 3\eta^2)\mathbb{E}[|X(\mathbf{s})|^2] - 4\eta^2\mathbb{E}[|X(\mathbf{s})|]^2}} . \quad (2.12)$$

Para la finalización del análisis de continuidad en media cuadrática del proceso, si se define que la función de correlación ( $\rho(\mathbf{s})$ ) asociada al proceso latente  $H$  es continua en media cuadrática, entonces  $\rho(0) = 1$  y para  $F_{\mathbf{H}_{ij}}(q_{g_\eta}, q_{g_\eta}, \rho(0))$  se tiene que cuando  $\rho(0) = 1$ , entonces  $F_{\mathbf{H}_{ij}}(q_{g_\eta}, q_{g_\eta}, 1) = g_\eta$ , es decir, que al reemplazar en (2.12) el resultado es 1.

Este resultado hace alusión a que la clase general de campos aleatorios *two-piece* definida por (2.4) será continua en media cuadrática siempre que la función de covarianza asociada a los procesos latentes  $X$  y  $H$  lo sea, lo que permite una propuesta flexible de considerar o no un efecto *nugget* según se pueda detectar en un análisis preliminar de las observaciones que se deseen modelizar.

Respecto a la diferenciabilidad en media cuadrática de los campos aleatorios *two-piece*, para probar que los procesos *two-piece* son  $k$  veces diferenciables, es necesario demostrar que la función de correlación en  $\mathbf{h} = 0$  es  $2k$  veces diferenciable. El primer paso entonces será evaluar el límite de la pendiente de la función de correlación del proceso cuando  $\mathbf{h} = 0$ , análisis que a continuación se desarrollará a partir del siguiente límite:

$$\lim_{\mathbf{h} \rightarrow 0} \frac{\rho_{P_\eta}(\mathbf{h}) - 1}{\mathbf{h}} . \quad (2.13)$$

El límite anterior evaluado directamente tiene por resultado 0/0, por lo que una alternativa es aplicar la regla de L'Hopital quedando la primera derivada de la expresión anterior por:

$$\rho'_{P_\eta}(\mathbf{h}) = \frac{\pi'_{11}(\mathbf{h})\mathbb{E}[|X(\mathbf{s}_i)||X(\mathbf{s}_j)|]}{(1 + 3\eta^2)[\text{Var}[|X(\mathbf{s})|] + \mathbb{E}[|X(\mathbf{s})|^2] - 4\eta^2\mathbb{E}[|X(\mathbf{s})|]^2}} . \quad (2.14)$$

Entonces, evaluar (2.13), sería equivalente a evaluar  $\lim_{\mathbf{h} \rightarrow 0} \pi'_{11}(\mathbf{h})$ , expresión dada por:

$$\begin{aligned}
\lim_{\mathbf{h} \rightarrow 0} \pi'_{11}(\mathbf{h}) &= \lim_{\mathbf{h} \rightarrow 0} \left\{ \int_{\infty^-}^{q_{g\eta}} \int_{\infty^-}^{q_{g\eta}} \frac{1}{2\pi\sqrt{1-(\rho(\mathbf{h}))^2}} \exp\left(-\frac{x_i^2 - 2\rho(\mathbf{h})x_i x_j + x_j^2}{2(1-(\rho(\mathbf{h}))^2)}\right) dx_i dx_j \right\}' , \\
&= \lim_{\mathbf{h} \rightarrow 0} \int_{\infty^-}^{q_{g\eta}} \int_{\infty^-}^{q_{g\eta}} \left( \frac{\rho(\mathbf{h})}{2\pi(1-(\rho(\mathbf{h}))^2)^{3/2}} + \frac{\frac{x_i x_j}{1-(\rho(\mathbf{h}))^2} + \rho(\mathbf{h}) \frac{x_i^2 - 2\rho(\mathbf{h})x_i x_j + x_j^2}{(1-(\rho(\mathbf{h}))^2)^2}}{2\pi\sqrt{1-(\rho(\mathbf{h}))^2}} \right) \\
&\quad \exp\left(-\frac{x_i^2 - 2\rho(\mathbf{h})x_i x_j + x_j^2}{2(1-(\rho(\mathbf{h}))^2)}\right) dx_i dx_j , \\
&= \lim_{\mathbf{h} \rightarrow 0} \int_{\infty^-}^{q_{g\eta}} \int_{\infty^-}^{q_{g\eta}} \left( \frac{\rho(\mathbf{h})}{2\pi(1-(\rho(\mathbf{h}))^2)^{3/2}} + \frac{x_i x_j}{(1-(\rho(\mathbf{h}))^2)(2\pi\sqrt{1-(\rho(\mathbf{h}))^2})} + \right. \\
&\quad \left. \rho(\mathbf{h}) \frac{x_i^2 - 2\rho(\mathbf{h})x_i x_j + x_j^2}{(2\pi\sqrt{1-(\rho(\mathbf{h}))^2})(1-(\rho(\mathbf{h}))^2)^2} \right) \\
&\quad \exp\left(-\frac{x_i^2 - 2\rho(\mathbf{h})x_i x_j + x_j^2}{2(1-(\rho(\mathbf{h}))^2)}\right) dx_i dx_j , \\
&= \infty^+ . \tag{2.15}
\end{aligned}$$

Si bien en este punto no se ha terminado de demostrar completamente que el proceso no es diferenciable en media cuadrática, se puede argumentar de que si se continua iterando respecto a la  $k$ -ésima derivada de  $\pi_{11}(\mathbf{h})$  se obtendrá que el límite es  $\infty^+$  dado que los términos involucrarán la expresión exponencial, función que produce la divergencia del límite anterior.

Así se concluye entonces que los campos aleatorios *two-piece* no son diferenciables en media cuadrática, incluso cuando la función de covarianza asociada a los procesos latentes lo sean, por lo que en la práctica, la elección de funciones de covarianza para la modelización de procesos *two-piece* debe considerar que estas no sean diferenciables en media cuadrática.

### 2.3. Casos particulares

Si bien los campos aleatorios *two-piece* representan una clase general de modelos para observaciones espacialmente dependientes que muestren asimetría, caracterizar casos particulares no es tan simple como en el caso de variables aleatorias (Arellano-Valle y cols., 2005) al considerar distintos modelos simétricos en la representación estocástica (2.2) ya que en el contexto de campos aleatorios, la representación estocástica (2.4) debe necesariamente considerar procesos existentes (Abrahamsen, 1997).

A continuación, se presentarán tres casos particulares de los campos aleatorios *two-piece*, el primer caso será una revisión al proceso *two-piece*  $t$  de student presentado en Caamaño (2018), seguido de la introducción de campo aleatorio *two-piece* Gaussiano y *two-piece* Tukey- $h$ ; modelos que hasta el momento no existen en la literatura.

### 2.3.1. Caso t de student

#### Campos aleatorios t de student

En el extenso trabajo de construcción y estudio de campos aleatorios con distribuciones marginales asimétricas y de colas pesadas, [Caamaño \(2018\)](#) introduce procesos espaciales con distribuciones marginales t de student y una posterior construcción de procesos *two-piece* t de student al escoger como proceso simétrico en (2.4) el campo aleatorio t de student estándar  $Y_v^* = \{Y_v^*(s), s \in S\}$  definido por:

$$Y_v^*(s) = W(s)_v^{-1/2} G^*(s) , \quad (2.16)$$

con  $v \in \mathbb{N}$  los grados de libertad,  $W_v = \{W(s)_v, s \in S\}$  un proceso con distribuciones marginales Gamma  $(v/2, v/2)$  ([Bevilacqua y cols., 2019b](#)) y  $G^* = \{G(s), s \in S\}$  un proceso Gaussiano estándar con función de correlación  $\rho(\mathbf{h})$ .

El proceso (2.16) tiene distribuciones marginales t de student con  $v$  grados de libertad y función de distribución de probabilidad dada por:

$$f_{Y_v^*}(y) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi v}} \left(1 + \frac{y^2}{2}\right)^{-\frac{(v+1)}{2}} , \quad (2.17)$$

cuyos momentos son  $\mathbb{E}[Y_v^*(s)] = 0$  y  $\text{Var}[Y_v^*(s)] = \frac{v}{v-2}$ , para  $v > 2$ . Además, [Caamaño \(2018\)](#) obtiene las expresiones para la función de correlación asociada al proceso y la función de distribución de probabilidad bivariada asociada, expresiones dadas por:

$$\rho_{Y_v^*}(\mathbf{h}) = \frac{(v-2)\Gamma^2\left(\frac{v-1}{2}\right)}{2\Gamma\left(\frac{v}{2}\right)} \left\{ {}_2F_1\left(\frac{1}{2}, \frac{1}{2}, \frac{v}{2}, \rho^2(\mathbf{h})\right) \right\} , \quad (2.18)$$

$$\begin{aligned} f_{\mathbf{Y}_{v;ij}^*}(y_i, y_j) &= \frac{v^v l_{ij} \Gamma^2\left(\frac{v+1}{2}\right)}{\pi \Gamma^2\left(\frac{v}{2}\right) (1 - (\rho_{Y_v^*}(\mathbf{h}))^2)^{-\frac{(v+1)}{2}}} F_4\left(\frac{v+1}{2}, \frac{v+1}{2}, \frac{1}{2}, \frac{v}{2}, \frac{(\rho_{Y_v^*}(\mathbf{h}))^2}{l_{ij}}, \frac{v^2(\rho_{Y_v^*}(\mathbf{h}))^2}{l_{ij}}\right) + \\ &\frac{(\rho_{Y_v^*}(\mathbf{h}))^2 y_i y_j v^{v+2} l_{ij}^{-\frac{v}{2}-1}}{2\pi (1 - (\rho_{Y_v^*}(\mathbf{h}))^2)^{-\frac{(v+1)}{2}}} F_4\left(\frac{v}{2} + 1, \frac{v}{2} + 1, \frac{3}{2}, \frac{v}{2}, \frac{(\rho_{Y_v^*}(\mathbf{h}))^2}{l_{ij}}, \frac{v^2(\rho_{Y_v^*}(\mathbf{h}))^2}{l_{ij}}\right) , \end{aligned} \quad (2.19)$$

con  $l_{ij} = (y_i^2 + v)(y_j^2 + v)$  y:

$$F_4(a, b; c, c'; w, z) = \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} \frac{(a)_{k+m} (b)_{k+m} w^k z^m}{k! m! (c)_k (c')_k} , \quad |\sqrt{w}| + |\sqrt{z}| < 1 ,$$

una función hipergeométrica de Appell de cuarto tipo.

Para una revisión más profunda de los procesos t de student, incluyendo las demostraciones, revisar [Bevilacqua y cols. \(2019a\)](#).

#### Campos aleatorios *two-piece* t de student

Al considerar entonces un campo aleatorio débilmente estacionario t de student estándar (2.16) como la elección del proceso simétrico en la formulación general *two-piece* (2.4), bastaría entonces

definir los momentos y función de distribución de probabilidad bivariada de un proceso  $|Y_v^*| = \{|Y_v^*(\mathbf{s})|, \mathbf{s} \in S\}$ , expresiones dadas por (Caamaño, 2018):

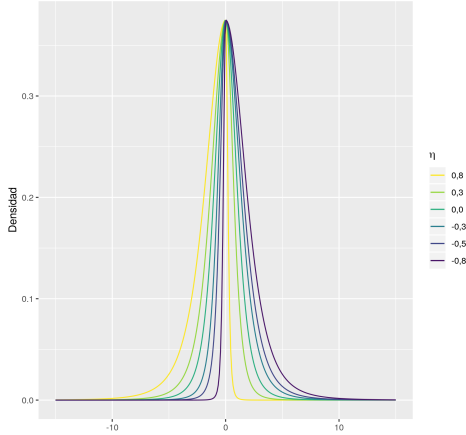
$$\mathbb{E}[|Y_v^*(\mathbf{s})|] = \frac{\Gamma\left(\frac{v-1}{2}\right)\sqrt{v}}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi}}, \quad (2.20)$$

$$\text{Var}[|Y_v^*(\mathbf{s})|] = \frac{v}{v-2} - \frac{\Gamma^2\left(\frac{v-1}{2}\right)v}{\Gamma^2\left(\frac{v}{2}\right)\pi}, \quad (2.21)$$

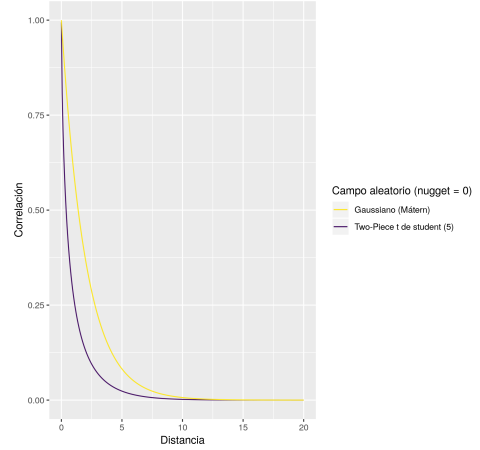
$$f_{|Y_{v;i_j}^*|}(y_i, y_j) = \frac{v^v l_{ij}^{\frac{-(v+1)}{2}} \Gamma^2\left(\frac{v+1}{2}\right)}{\pi \Gamma^2\left(\frac{v}{2}\right) (1 - (\rho_{Y_v^*}(\mathbf{h}))^2)^{\frac{-(v+1)}{2}}} F_4\left(\frac{v+1}{2}, \frac{v+1}{2}, \frac{1}{2}, \frac{v}{2}, \frac{(\rho_{Y_v^*}(\mathbf{h}))^2 y_i^2 y_j^2}{l_{ij}}, \frac{v^2 (\rho_{Y_v^*}(\mathbf{h}))^2}{l_{ij}}\right). \quad (2.22)$$

Finalmente, solo falta definir una expresión para  $\mathbb{E}[|Y_v^*(\mathbf{s}_i)||Y_v^*(\mathbf{s}_j)||]$  para dejar completamente definido un campo aleatorio *two-piece*  $t$  de student estándar, es decir,  $F_{\eta,v}^*(\mathbf{s}) = |Y_v^*(\mathbf{s})|K_\eta(\mathbf{s})$ :

$$\mathbb{E}[|Y_v^*(\mathbf{s}_i)||Y_v^*(\mathbf{s}_j)||] = \frac{v\Gamma^2\left(\frac{v-1}{2}\right)}{\pi\Gamma^2\left(\frac{v-1}{2}\right)} {}_2F_1\left(\frac{1}{2}, \frac{1}{2}, \frac{v}{2}, (\rho_{Y_v^*}(\mathbf{h}))^2\right) {}_2F_1\left(-\frac{1}{2}, -\frac{1}{2}, \frac{v}{2}, (\rho_{Y_v^*}(\mathbf{h}))^2\right) \quad (2.23)$$



(a)



(b)

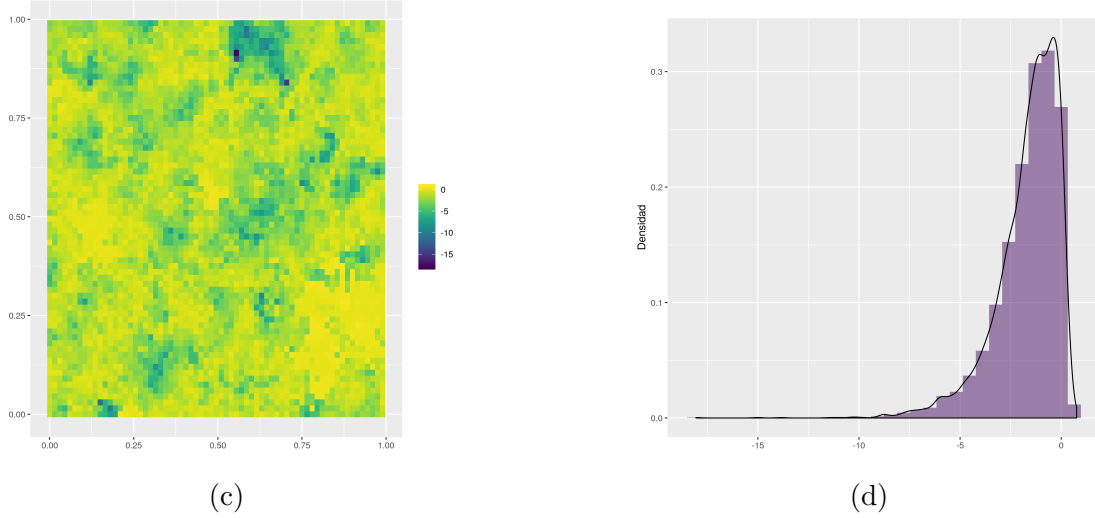


Figura 2.3: Ilustración sobre características de un campo aleatorio *two-piece*  $t$  de student: (a) distribuciones marginales ( $v = 4$ ), (b) realización del proceso ( $\eta = 0,8$  y  $v = 6$ ) con función de correlación  $\mathcal{M}_{(0,5),(0,1),(0)}(\mathbf{h})$ , (c) función de correlación y (d) histograma de una realización.

### 2.3.2. Caso Gaussiano

Para la construcción del modelo *two-piece* Gaussiano estándar es necesario considerar un campo aleatorio Half-Gaussiano  $|G^*| = \{|G^*(\mathbf{s})|, \mathbf{s} \in S\}$ , definido como el valor absoluto de un campo aleatorio Gaussiano estándar con función de correlación asociada  $\rho_{G^*}(\mathbf{h})$ . En este caso, las expresiones para el valor esperado y la varianza del proceso  $|G^*|$  están dadas por:

$$\mathbb{E}[|G^*(\mathbf{s})|] = \sqrt{\frac{\pi}{2}} , \quad (2.24)$$

$$\text{Var}[|G^*(\mathbf{s})|] = 1 - \frac{2}{\pi} . \quad (2.25)$$

El nuevo proceso *two-piece* Gaussiano, según la construcción general (2.4), queda completamente definido al conocer las expresiones para los momentos, la función de correlación y la función de distribución de probabilidad bivariada. En el trabajo de [Zhang y El-Shaarawi \(2010\)](#), los autores calculan la expresión para:

$$\mathbb{E}[|G^*(\mathbf{s}_i)||G^*(\mathbf{s}_j)|] = \frac{2}{\pi} \left\{ \sqrt{(1 - (\rho_{G^*}(\mathbf{h}))^2)} + \rho_{G^*}(\mathbf{h}) \arcsin(\rho_{G^*}(\mathbf{h})) \right\} . \quad (2.26)$$

Finalmente, la función de distribución de probabilidad bivariada asociada al vector bivariado  $|G_{ij}^*| = [|G_i^*|, |G_j^*|]^T$  puede ser fácilmente calculada a través del método de transformaciones a través de la función de distribución acumulada, quedando así la expresión:

$$f_{|G_{ij}^*|}(g_i, g_j, \rho) = f_{G_{ij}^*}(g_i, g_j, \rho) + f_{G_{ij}^*}(-g_i, -g_j, \rho) + f_{G_{ij}^*}(-g_i, g_j, \rho) + f_{G_{ij}^*}(g_i, -g_j, \rho) . \quad (2.27)$$

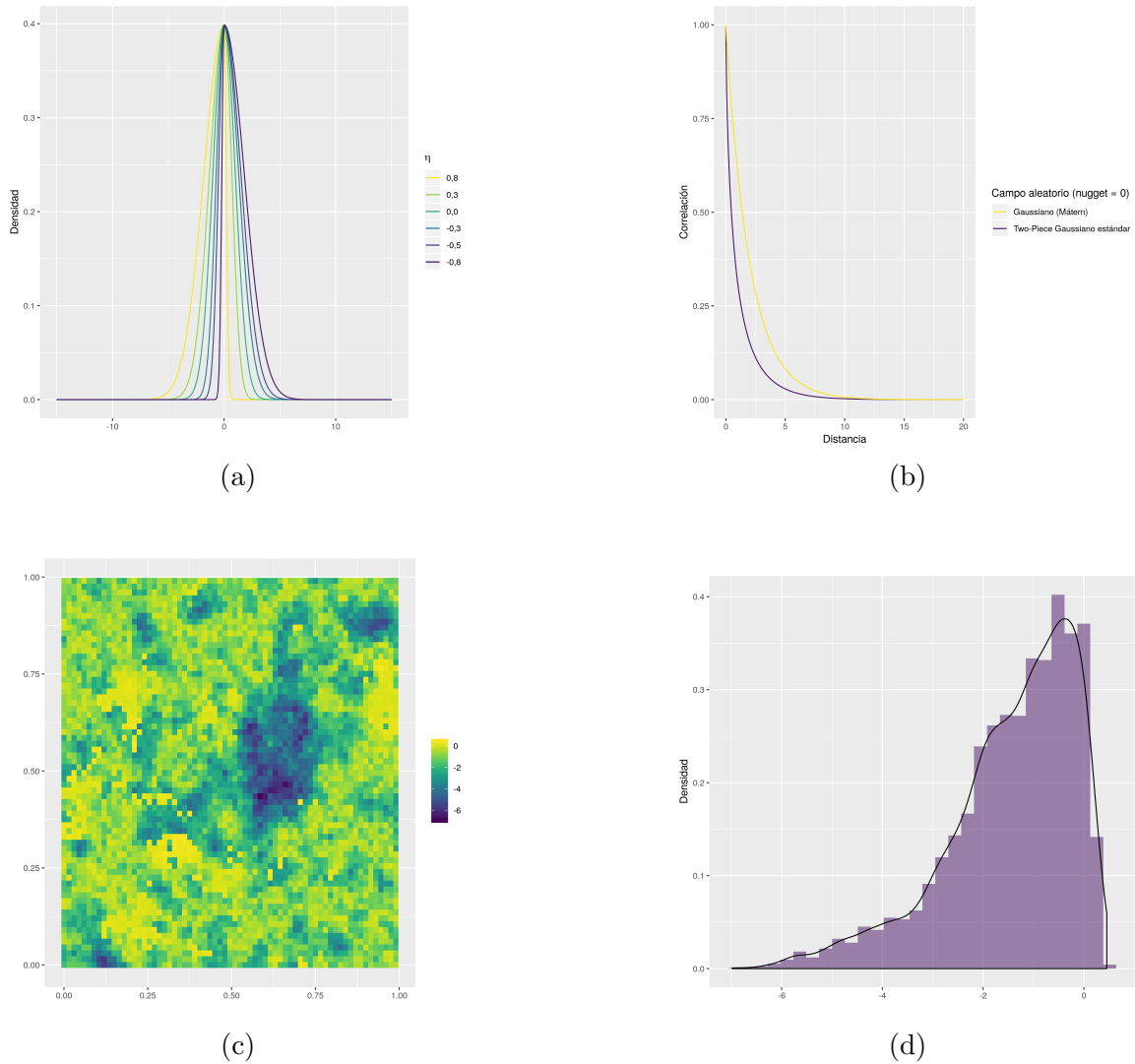


Figura 2.4: Ilustración sobre características de un campo aleatorio *two-piece* Gaussiano: (a) distribuciones marginales, (b) realización del proceso ( $\eta = 0,8$ ) con función de correlación  $\mathcal{M}_{(0,5),(0,1),(0)}(\mathbf{h})$ , (c) función de correlación y (d) histograma de una realización.

Una característica especial de la función de distribución bivariada asociada al proceso *two-piece* Gaussiano es que genera un modelo con dependencia del tipo no elíptica (Babić, Ley, y Veredas, 2019). Esta característica en la dependencia ha sido ligeramente estudiada en la literatura dado que para lograr una mayor flexibilidad en modelos multivariados se han incorporado características como colas más o menos pesadas que el modelo Gaussiano multivariado (distribuciones elípticas) y asimetría (modelos elípticos asimétricos), pero estas propuestas conservan el elipticidad de las distribuciones.

Los efectos de la correlación ( $\rho$ ) y el parámetro de asimetría ( $\eta$ ) se muestran en la comparación gráfica de los gráficos de contorno para las distribuciones bivariadas de media nula y varianza

unitaria (estándar) Gaussiana y *two-piece* Gaussiana en la Figura 2.5 para distintos valores dentro de los espacios paramétricos.

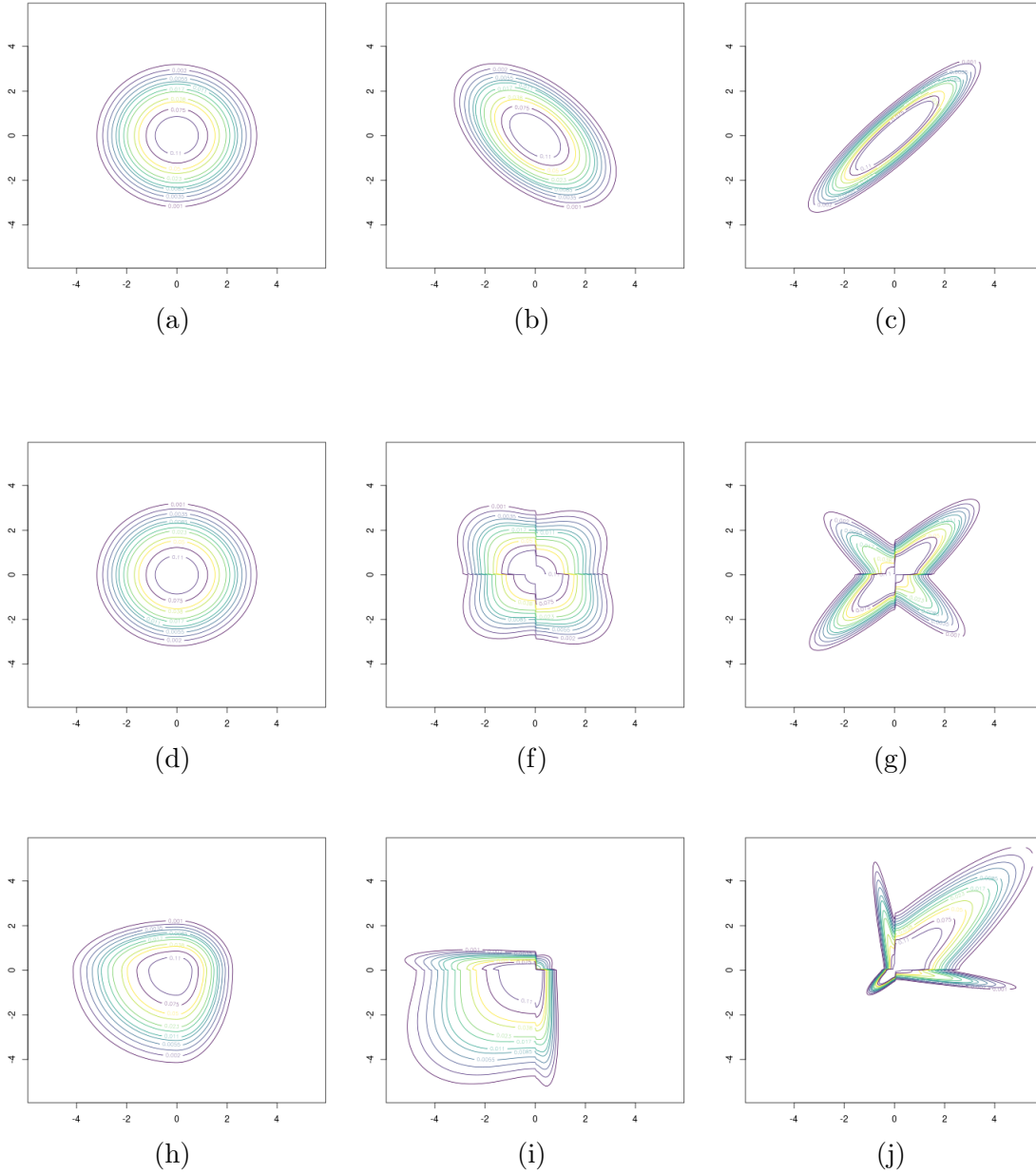


Figura 2.5: Comparación gráficos de contorno para las funciones de distribución bivariadas estándar: (a) Gaussiana  $[\rho = 0]$  (b) Gaussiana  $[\rho = -0,5]$  (c) Gaussiana  $[\rho = 0,9]$  (d) *two-piece* Gaussiana  $[\rho = 0][\eta = 0]$  (e) *two-piece* Gaussiana  $[\rho = -0,5][\eta = 0]$  (f) *two-piece* Gaussiana  $[\rho = 0,9][\eta = 0]$  (g) *two-piece* Gaussiana  $[\rho = 0][\eta = 0,3]$  (e) *two-piece* Gaussiana  $[\rho = -0,5][\eta = 0,7]$  (f) *two-piece* Gaussiana  $[\rho = 0,9][\eta = -0,70]$ .

Un aspecto interesante es destacar que la distribución bivariada no contiene al caso Gaussiano estándar como caso particular dado que el contorno de la distribución no es el mismo y a medida

que la correlación cambia esto se vuelve más evidente.

### 2.3.3. Caso Tukey- $h$

La modelización de observaciones espacialmente dependientes ha demandado que los métodos de ajuste de modelos también consideren implementaciones computacionales eficientes (Bevilacqua y cols., 2019c), argumento por el cual campos aleatorios *two-piece*  $t$  de student no serían tan recomendados dado que dependen de funciones especiales.

Alternativamente al proceso *two-piece*  $t$  de student, se podría escoger cualquier otro modelo simétrico estándar de colas pesadas para generar un nuevo proceso *two-piece* con el fin de modelizar observaciones espacialmente dependientes, pero que la estimación de parámetros no se vuelva una tarea computacionalmente costosa. Particularmente, se escogerá el campo aleatorio Tukey- $h$  a partir del modelo asimétrico y de colas pesadas Tukey  $g$ -and- $h$  presentado por Xu y Genton (2017).

A continuación se presenta una revisión al modelo probabilístico Tukey- $h$ , la caracterización de este modelo como campo aleatorio de colas pesadas y el proceso espacial *two-piece* Tukey- $h$ .

#### Modelo probabilístico Tukey- $h$

Algunos modelos probabilísticos se pueden generar a partir de transformaciones de variables aleatorias con funciones de distribución de probabilidad conocidas, Goerg (2015) muestra una manera general de construir modelos de colas pesadas y de manera especial el modelo Tukey- $h$  definido por la siguiente transformación:

$$\tau_h(G^*) = G^* \exp\left(\frac{h}{2}G^{*2}\right) , \quad (2.28)$$

particularmente  $G^*$  es una variable aleatoria Gaussiana estándar,  $h \in \mathbb{R}_{[0, \infty+)}$  el parámetro que permite controlar la pesadez de las colas del modelo, permitiendo así que sea una transformación biyectiva y con una expresión cerrada para la transformación inversa (Goerg, 2015). Esta transformación inversa, denotada por  $\tau_h^{-1}(T_h)$ , se define como:

$$\tau_h^{-1}(T_h) = \text{sign}(T_h) \left( \frac{W(hT_h^2)}{h} \right)^{1/2} = G^* ,$$

para  $W$  la función de Lambert (Goerg, 2011).

Las funciones de distribución acumulada y de distribución de probabilidad univariadas de esta nueva variable aleatoria Tukey- $h$  estándar ( $T_h^*$ ) están dadas por:

$$F_{T_h^*}(t) = F_{G^*}(\tau_h^{-1}(t)) , \quad (2.29)$$

$$f_{T_h^*}(t) = \frac{\tau_h^{-1}(t)}{t(1+W(ht^2))} f_{G^*}(\tau_h^{-1}(t)) , \quad (2.30)$$

donde  $F_{G^*}$  y  $f_{G^*}$  son las funciones de distribución acumulada y de distribución de probabilidad Gaussiana estándar, respectivamente.

La transformación se puede extender a cada componente de un vector aleatorio  $n$ -dimensional Gaussiano  $\mathbf{G}^* = [G_i^*]_{i=1}^n$  con vector de medias  $\boldsymbol{\mu} = \mathbf{0}$  y matriz de correlación  $\Omega$ , es decir,  $\mathbf{G}^* \sim \text{Gauss}_{n\text{-dim}}(\mathbf{0}, \Omega)$ . Entonces, al aplicar la transformación se obtiene  $\mathbf{T}_h^* = [t_i^*]_{i=1}^n$ , para  $t_i^* = \tau_h(G_i^*)$ .

La función de densidad de probabilidad multivariada de  $\mathbf{T}_h^*$  está dada por:

$$f_{\mathbf{T}_h^*}(\mathbf{t}^*) = \prod_{i=1}^n \left( \frac{\tau_h^{-1}(t_i^*)}{t_i^*(1 + W(ht_i^{*2}))} \right) f_{\mathbf{G}^*}(\tau_h^{-1}(\mathbf{t}^*), \mathbf{0}, \Omega) , \quad (2.31)$$

donde  $\tau_h^{-1}(\mathbf{t}^*) = [\tau_h^{-1}(t_i^*)]_{i=1}^n$  y  $f_{\mathbf{G}^*}$  es la función de densidad de probabilidad multivariada  $n$ -dimensional Gaussiana con media  $\mathbf{0}$  y matriz de covarizancias  $\Omega$ .

### Campos aleatorios Tukey- $h$

Al igual que todas las construcciones de campos aleatorios aquí presentadas, se pueden construir modelos a partir de representaciones estocásticas, en este caso se puede extender la transformación Tukey- $h$  (2.28) si se define  $G^* = \{G^*(\mathbf{s}), \mathbf{s} \in S\}$  como un campo aleatorio Gaussiano de media 0 y varianza unitaria con función de correlación  $\rho(\mathbf{h})$ . Entonces se puede construir el campo aleatorio Tukey- $h$  estándar  $T_h^* = \{T_h^*(\mathbf{s}), \mathbf{s} \in S\}$  de la siguiente manera:

$$T_h^*(\mathbf{s}) = G^*(\mathbf{s}) \exp\left(\frac{h}{2} G^*(\mathbf{s})^2\right) , \quad (2.32)$$

cuyos momentos de primer y segundo orden están dados por:

$$\mathbb{E}[T_h^*(\mathbf{s})] = 0 , \quad (2.33)$$

$$\text{Var}[T_h^*(\mathbf{s})] = (1 - 2h)^{-3/2} , \quad (2.34)$$

$$\text{Cov}[T_h^*(\mathbf{s}_i), T_h^*(\mathbf{s}_i)] = \frac{\rho(\mathbf{h})}{(1 + h(\rho(\mathbf{h}) - 1))(1 - h - h\rho(\mathbf{h}))\sqrt{1 + h(-2 + h - h(\rho(\mathbf{h}))^2)}} \quad (2.35)$$

donde  $\rho(\mathbf{h})$  es la función de correlación asociada a  $G^*(\mathbf{s})$  y  $h \in [0, 1/2)$  con el fin de que  $\text{Var}[T_h^*(\mathbf{s})]$  sea positiva.

El modelo (2.32) tiene como funciones de distribución de probabilidad y de distribución acumulada marginales las expresiones (2.30) y (2.29), además la función de distribución de probabilidad finito dimensional del proceso (2.31), a diferencia de los campos aleatorios Tukey  $g$ -and- $h$  (Xu y Genton, 2017), tiene una expresión cerrada permitiendo así realizar la estimación de parámetros por máxima verosimilitud.

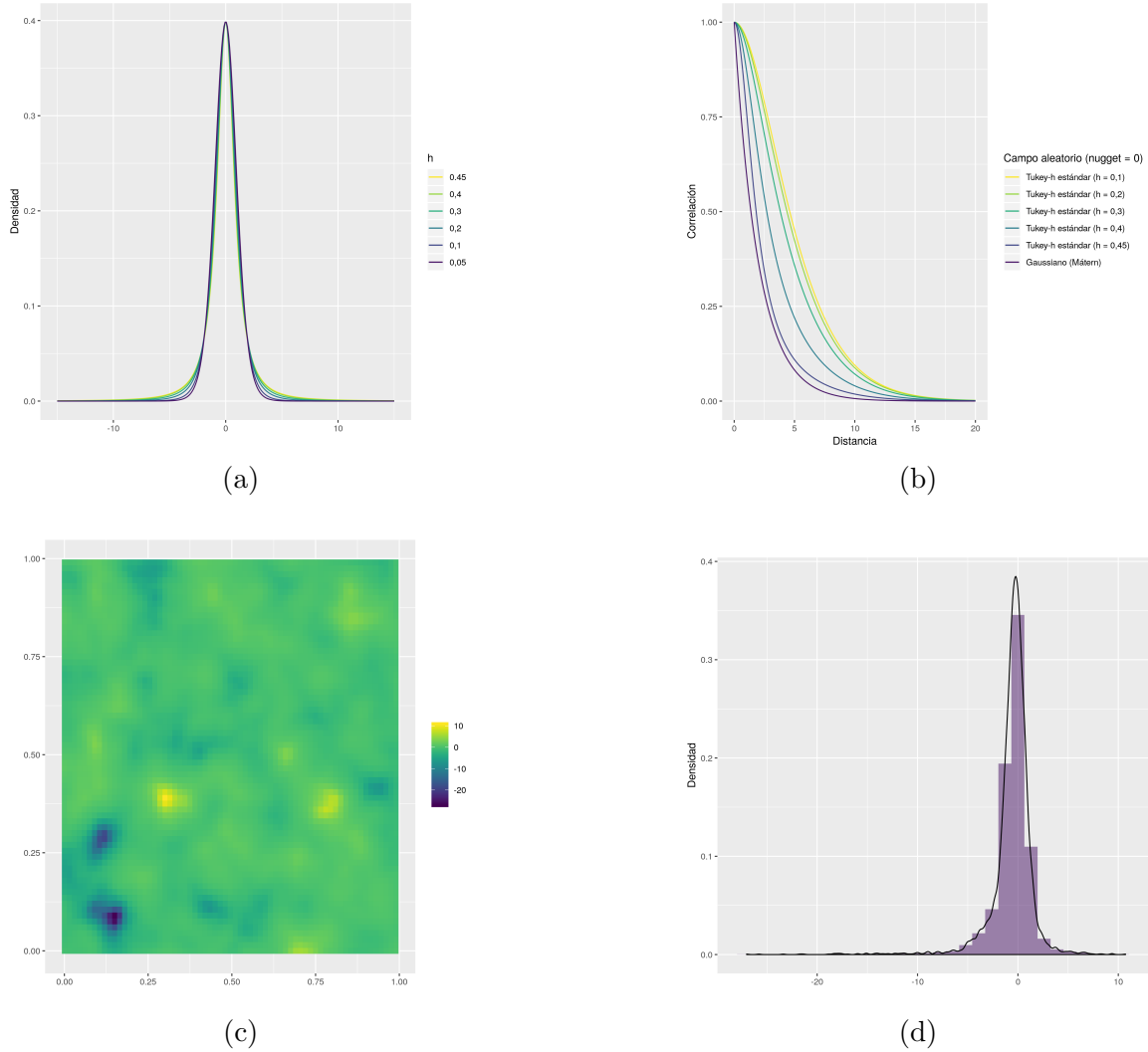


Figura 2.6: Ilustración sobre características de un campo aleatorio Tukey- $h$ : (a) distribuciones marginales, (b) realización del proceso ( $h = 0,4$ ) con función de correlación  $\mathcal{M}_{(2,5),(0,033),(0)}(\mathbf{h})$ , (c) función de correlación y (d) histograma de una realización.

### Campos aleatorios *two-piece* Tukey- $h$

Si  $T_h^*(\mathbf{s})$  es un campo aleatorio Tukey- $h$  estándar definido en (2.32), entonces se puede definir el siguiente proceso:

$$P_{\eta,h}^*(\mathbf{s}) = |T_h^*(\mathbf{s})|K_{\eta}(\mathbf{s}) , \quad (2.36)$$

como un campo aleatorio *two-piece* Tukey- $h$  estándar con  $\eta \in \mathbb{R}_{(-1,1)}$  que controla el grado de asimetría y  $h \in \mathbb{R}_{[0,1/2)}$  la pesadez de las colas. Además, un proceso discreto  $K_{\eta}$  definido por (2.5). Una vez definido  $T_{\eta,h}^*$  las expresiones para los momentos de primer y segundo orden están dadas por:

$$\mathbb{E}[|T_{\eta,h}^*(\mathbf{s})|] = \left( \frac{1}{1-h} \right) \left( \frac{2}{\pi} \right)^{1/2} , \quad (2.37)$$

$$\text{Var}[|T_{\eta,h}^*(\mathbf{s})|] = (1 - 2h)^{-3/2} - \left(\frac{1}{1-h}\right) \left(\frac{2}{\pi}\right)^{1/2}, \quad (2.38)$$

$$\mathbb{E}[|T_h^*(\mathbf{s}_i)||T_h^*(\mathbf{s}_j)|] = \frac{2(1 - \rho^2(\mathbf{h}))^{3/2}}{\pi g^2(\mathbf{h}, h)} {}_2F_1\left(1, 1, \frac{1}{2}, \frac{\rho^2(\mathbf{h})}{g^2(\mathbf{h}, h)}\right), \quad (2.39)$$

donde  $g(\mathbf{h}, h) = 1 - (1 - \rho^2(\mathbf{h}))h$ .

Finalmente, la función de distribución de probabilidad bivariada asociada al vector bivariado  $[|T_{\eta;i}^*|, |T_{\eta;j}^*|]^T$  se puede obtener igual que el caso Gaussiano, quedando como expresión:

$$\begin{aligned} f_{|T_{\eta;i;j}^*|}(t_i, t_j) &= f_{T_{\eta;i;j}^*}(t_i, t_j, \rho_{T_{\eta}^*}(\mathbf{h})) + f_{T_{\eta;i;j}^*}(-t_i, -t_j, \rho_{T_{\eta}^*}(\mathbf{h})) \\ &\quad + f_{T_{\eta;i;j}^*}(-t_i, t_j, \rho_{T_{\eta}^*}(\mathbf{h})) + f_{T_{\eta;i;j}^*}(t_i, -t_j, \rho_{T_{\eta}^*}(\mathbf{h})). \end{aligned} \quad (2.40)$$

Donde  $f_{T_{\eta;i;j}^*}$  es la función de distribución de probabilidad bivariada Tukey- $h$  estándar como caso particular de (2.31).

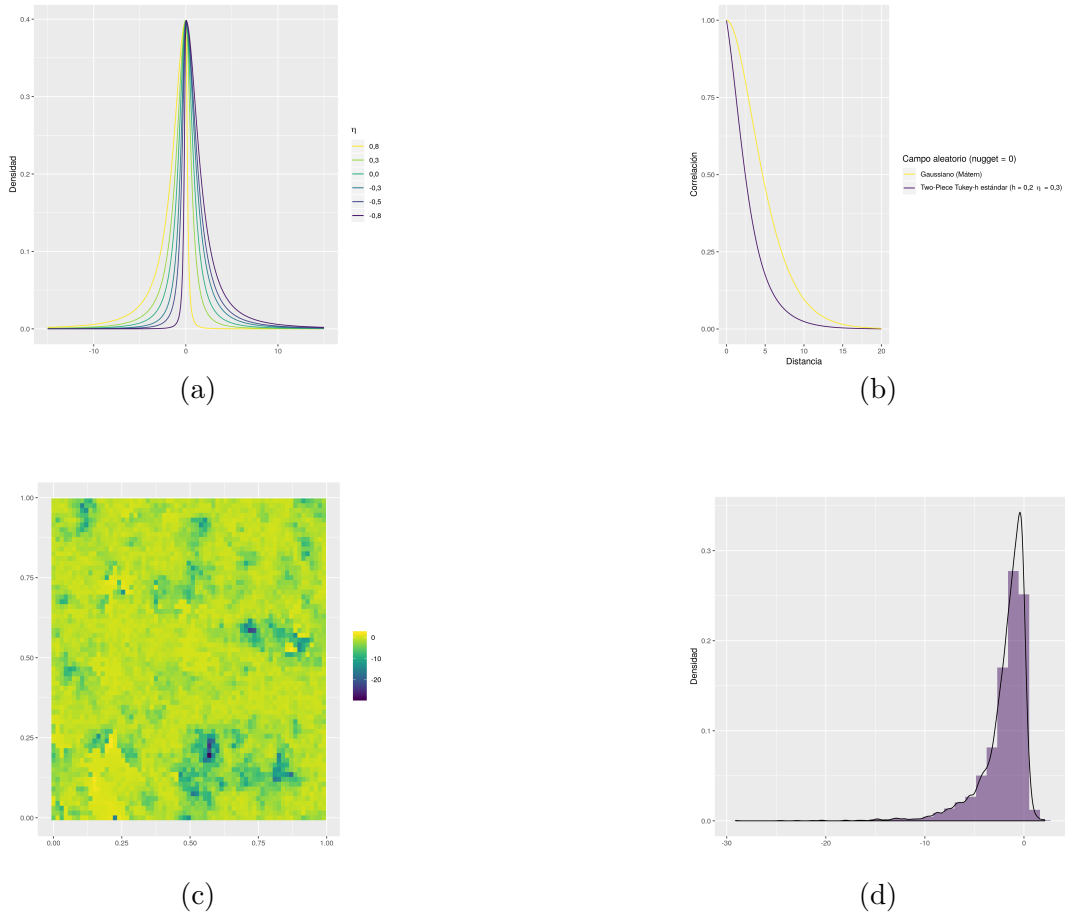


Figura 2.7: Ilustración sobre características de un campo aleatorio *two-piece* Tukey- $h$ : (a) distribuciones marginales ( $h = 0,45$ ), (b) realización del proceso ( $\eta = 0,8$  y  $h = 0,3$ ) con función de correlación  $\mathcal{M}_{(0,5),(0,1),(0)}(\mathbf{h})$ , (c) función de correlación y (d) histograma de una realización.

De la misma manera que para la función de distribución bivariata *two-piece* Gaussiana y las comparaciones realizadas en la Figura 2.5, se ilustra a continuación las comparaciones obtenidas para los gráficos de contorno de las distribuciones bivariadas estándar Tukey- $h$  y *two-piece* Tukey- $h$  y el efecto de los parámetros de correlación y asimetría para un valor del parámetro de pesadez de cola  $h = 0,2$ .

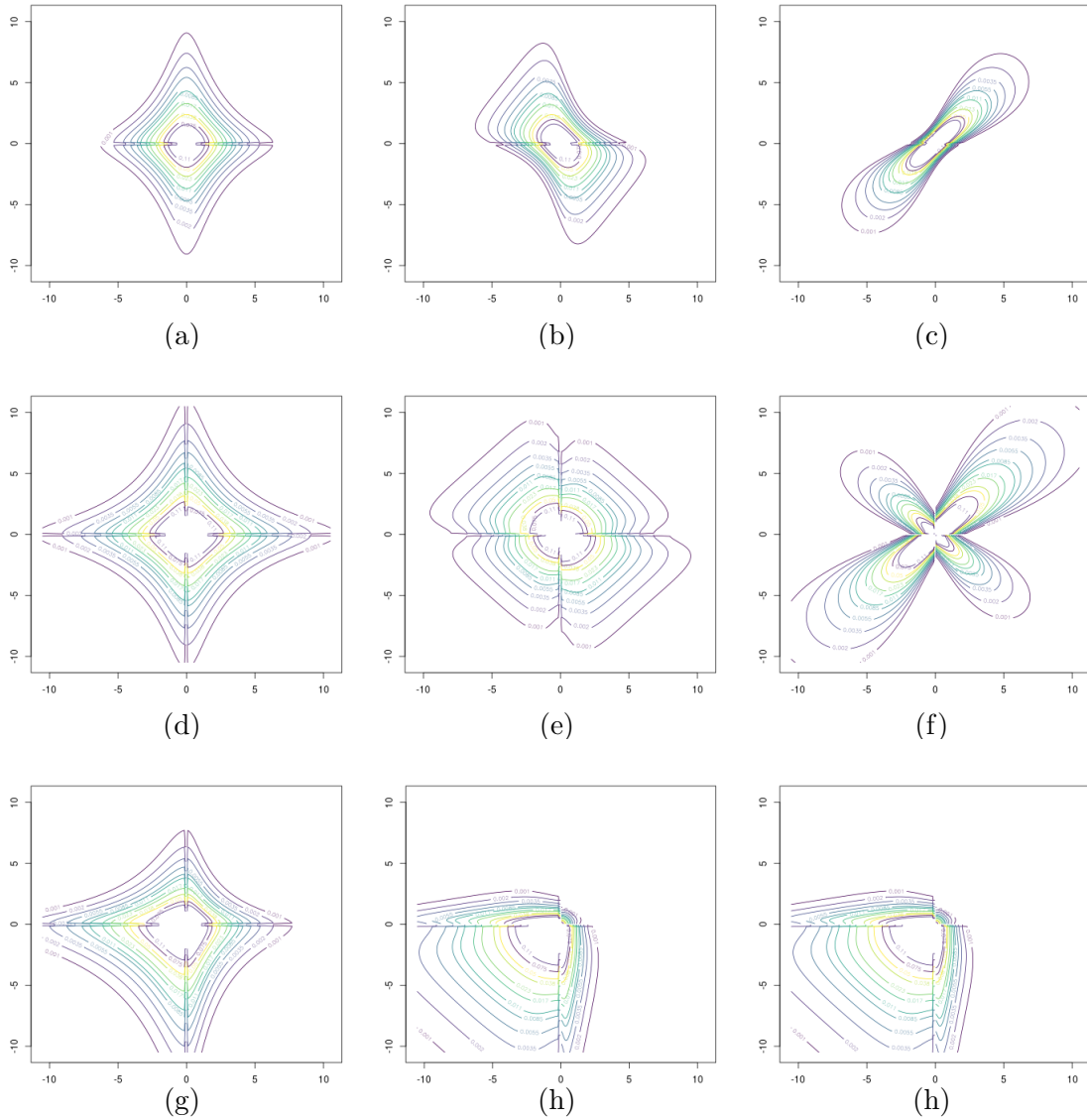


Figura 2.8: Comparación gráficos de contorno para las funciones de distribución bivariadas estándar: (a) Tukey- $h$  [ $\rho = 0$ ] (b) Tukey- $h$  [ $\rho = -0,5$ ] (c) Tukey- $h$  [ $\rho = 0,9$ ] (d) *two-piece* Tukey- $h$  [ $\rho = 0$ ][ $\eta = 0$ ] (e) *two-piece* Tukey- $h$  [ $\rho = -0,5$ ][ $\eta = 0$ ] (f) *two-piece* Tukey- $h$  [ $\rho = 0,9$ ][ $\eta = 0$ ] (g) *two-piece* Tukey- $h$  [ $\rho = 0$ ][ $\eta = 0,3$ ] (e) *two-piece* Tukey- $h$  [ $\rho = -0,5$ ][ $\eta = 0,7$ ] (f) *two-piece* Tukey- $h$  [ $\rho = 0,9$ ][ $\eta = -0,70$ ].

Una característica de la función de distribución bivariada *two-piece* Tukey- $h$  (Figura 2.8) es que a diferencia del modelo *two-piece* Gaussiano (Figura 2.5), el incorporar un parámetro que controle la pesadez de las colas de la distribución permite un modelo bivariado más flexible.

## 2.4. Notas del capítulo

Una de las ventajas de presentar una construcción general de campos aleatorios es que las expresiones algebraicas para los momentos y funciones asociadas permiten estudiar el modelo de manera global. En primer lugar, para particularizar (2.4), solo es necesario conocer los momentos asociados al proceso simétrico  $X$  además de la función de distribución de probabilidad bivariada del proceso.

Respecto a las propiedades geométricas de los campos aleatorios *two-piece*, se concluye que estos modelos serán flexibles respecto a la elección de funciones de covarianza que consideren o no un efecto *nugget* según la elección de los usuarios, pero el modelo pierde flexibilidad al no ser diferenciable en media cuadrática ya que si las observaciones en estudio presentan suavidad en la trayectoria de la realización de los procesos, aunque se escojan funciones de covarianza diferenciables, estos modelos no pueden capturar esa información debido a la manera en que son construidos.

Desde la perspectiva de la aplicación a la modelización de observaciones espacialmente dependientes y continuas en  $\mathbb{R}$ , los campos aleatorios presentados aquí componen una alternativa a otros modelos. En primer lugar, el modelo *two-piece* t de student se presentó en Caamaño (2018) como alternativa a la modelización de fenómenos con colas pesadas y asimetría bajo los campos aleatorios t de student asimétricos (Bevilacqua y cols., 2019a) ya que no se ha obtenido hasta el momento una expresión algebraica para la distribución bivariada de los procesos, por lo que no se puede realizar inferencia de parámetros bajo este modelo.

Otra desventaja que presentan los campos aleatorios derivados de procesos t de student, es que estos modelos están limitados por sus grados de libertad, esto se debe a que estos valores se encuentran restringidos a la mitad de valores enteros debido a la construcción de los procesos Gamma (Bevilacqua y cols., 2019a). Una manera de solucionar esta restricción sería poder estudiar una familia de funciones de covarianza que permitan la flexibilidad de obtener cualquier valor real para los grados de libertad, pero hasta el momento ese resultado solo ha sido estudiado para la función de covarianza exponencial.

Respecto al tercer caso particular, considerar campos aleatorios con distribuciones marginales simétricas como la Tukey- $h$  (Goerg, 2015) que también presenta colas pesadas y funciones de distribución que no dependen de funciones especiales como los campos aleatorios t de student genera la enorme ventaja de que computacionalmente se reducen los tiempos de ajuste de los modelos.

Así mismo, los campos aleatorios *two-piece* Gaussianos también se podrían considerar como una alternativa a los campos aleatorios Gaussianos asimétricos (Zhang y El-Shaarawi, 2010) siempre que las trayectorias de la realización de los procesos presenten indicios de no diferenciability en media cuadrática.

Una consideración importante respecto a los campos aleatorios *two-piece* es que son los únicos modelos hasta el momento presentes en la literatura de Estadística Espacial para procesos que pueden lidiar con dependencias del tipo no elípticas en el tiempo o en el espacio, esto se debe a la característica particular que presentan las funciones de distribución bivariada respecto a los modelos multivariados elípticos asimétricos (Azzalini y Capitanio, 2014).

## Capítulo 3

# Implementación computacional

En este capítulo se abordará todo lo que hace referencia a pruebas numéricas de los casos particulares de la clase de campos aleatorios estudiadas en el capítulo anterior. Estas pruebas se dividen en dos secciones: en una primera parte se estudiará el comportamiento de la estimación de parámetros a partir de la verosimilitud compuesta por parejas ponderada, seguido de esto se dará paso a la aplicación de los modelos con los datos del cambio de divisas para el dólar Australiano sobre el Yen Japonés (AUDJPY) junto con el análisis de capacidad de predicción de los modelos bajo un predictor óptimo lineal.

### 3.1. Sobre software

Mucha de la teoría propuesta en la literatura ha sido sin duda un gran aporte respecto a la construcción de modelos con características particulares en la modelización de dependencia. Uno de los problemas es que muchas de las implementaciones computacionales requieren un alto esfuerzo y tiempo para la reproducción de resultados y la aplicación de modelos cuando no existe código computacional disponible o paquetes computacionales de libre acceso.

Para este trabajo se propone que todas las implementaciones computacionales estén a la disposición de futuros usuarios de los modelos presentados, por esta razón es que todas las rutinas computacionales estarán disponibles en el paquete en desarrollo de R (R Core Team, 2017), `GeoModels` (Bevilacqua y Morales-Oñate, 2018).

`GeoModels` es un paquete de R en el cual se pueden realizar procedimientos de simulación, estimación (verosimilitud completa y compuesta por parejas dependiendo del modelo) y predicción para el análisis de campos aleatorios Gaussianos y no Gaussianos en contextos espaciales, espacio-temporales y bivariados. A diferencia de otros paquetes para Estadística Espacial, `GeoModels` permite escoger dentro de una amplia familia de funciones de covarianza para contextos Euclidianos y en la esfera.

Modelo	Características
Gaussiano	soporte en $\mathbb{R}$
Gaussiano asimétrico	soporte en $\mathbb{R}$ y posible asimetría
<i>two-piece</i> Gaussiano	soporte en $\mathbb{R}$ y posible asimetría
t de student	soporte en $\mathbb{R}$ y colas pesadas
<i>two-piece</i> t de student	soporte en $\mathbb{R}$ , colas pesadas y posible asimetría
Tukey- <i>h</i>	soporte en $\mathbb{R}$ y colas pesadas
<i>two-piece</i> Tukey- <i>h</i>	soporte en $\mathbb{R}$ , colas pesadas y posible asimetría
Gamma	soporte en $\mathbb{R}^+$
Weibull	soporte en $\mathbb{R}^+$
Log Gaussiano	soporte en $\mathbb{R}^+$
Bernulli	discreto
Binomial	discreto
Poisson	discreto

Cuadro 3.1: Lista de modelos de campos aleatorios disponibles en *GeoModels*.

Respecto a la visualización de datos, gran parte de los gráficos y figuras se han construido bajo el paquete de visualización de R, `ggplot2` (Wickham, 2016) y para la paleta de colores (Figura 3.1) se ha utilizado `viridis` (Garnier, Ross, Rudis, Sciaini, y Scherer, 2018) ya que esa permite una apreciación uniforme sobre los cambios de colores respecto a cambios de escala numérica en comparación a otras paletas de colores que vienen por defecto en los paquetes computacionales de visualización, permitiendo así una visualización clara incluso para personas con problemas en la visión.



Figura 3.1: Escala de color *viridis* (imagen capturada desde [R Pubs](#)).

## 3.2. Estudio por simulación

Para ilustrar el comportamiento de la estimación de parámetros bajo el método de verosimilitud compuesta por parejas ponderada, se muestran los resultados de realizar el proceso de estimación 1000 veces en una grilla regular en el cuadrado unitario con 800 localizaciones espaciales para los modelos presentados en el Capítulo 2. A continuación se entrega el escenario implementado y los resultados junto con las medidas de sesgo y error cuadrático medio de estimación.

Sobre el valor de los parámetros a estimar, se considerará  $\sigma^2 = 1$  junto con dos parámetros de regresión, es decir,  $\mu(\mathbf{s}_i) = \beta_0 + \beta_1 u(\mathbf{s}_i)$  para  $\beta_0 = 0,5$ ,  $\beta_1 = -0,25$  en el cual  $u(\mathbf{s}_i)$  es una realización desde una distribución uniforme  $(0, 1)$ . Respecto a cada modelo, se fija  $\nu = 1/6$  and  $h = 0,1$  para el modelo *two-piece* t de student y *two-piece* Tukey-h respectivamente. Además, de fijar el parámetro de asimetría por  $\eta = 0,3$  en los tres casos.

Para el modelo de correlación, se escoge el modelo Wendland Generalizado (1.20) no diferenciable en media cuadrática ( $\varsigma = 0$ ), parámetro de soporte compacto  $\delta = 0,2$  y parámetro de decaimiento  $\gamma = 4$ , considerando  $\gamma$  como fijo y conocido. Para  $\varsigma = 0$ , el modelo Wendland Generalizado adopta la siguiente estructura según Bevilacqua y cols. (2019c):

$$\rho_{\gamma,\delta}(\mathbf{h}) := \begin{cases} (1 - \mathbf{h}/\delta)^\gamma & \mathbf{h} < \delta \text{ ,} \\ 0 & \text{e. o. c. .} \end{cases} \quad (3.1)$$

Finalmente, para la función de verosimilitud compuesta por parejas ponderada (1.21), se considera para  $dist_{max} = 0,03$  además de asumir que el valor de los grados de libertad del modelo *two-piece* t de student es conocido.

En general, las estimaciones obtenidas para estos modelos sugieren una estimación simétrica y centrada en los valores reales de los parámetros para los tres modelos tal como se puede observar tanto en las Cuadros 3.2, 3.3, 3.4 y Figuras 3.2, 3.3 y 3.4 que a continuación se muestran.

Parámetro	Sesgo	Error cuadrático medio
$\hat{\beta}_0$	0,00197	0,00018
$\hat{\beta}_1$	0,00310	0,00035
$\hat{\delta}$	0,00102	0,00047
$\hat{\sigma}^2$	-0,00309	0,00820
$\hat{\eta}$	0,00386	0,00075

Cuadro 3.2: Sesgo y error cuadrático medio para la estimación de parámetros de un campo aleatorio *two-piece* Gaussiano con función de correlación Wendland Generalizado bajo verosimilitud compuesta por parejas ponderada.

Si bien el escenario de simulación puede ser simple, los resultados aquí expuestos son interpretables en función de que estimar bajo verosimilitud compuesta por parejas ponderada para los modelos *two-piece* es razonable frente al hecho de no tener una expresión cerrada para la función de distribución finito dimensional y no poder estimar por máximo verosimilitud.

Parámetro	Sesgo	Error cuadrático medio
$\hat{\beta}_0$	0,00190	0,00014
$\hat{\beta}_1$	0,00291	0,00034
$\hat{\delta}$	0,00133	0,00048
$\hat{\sigma}^2$	0,00821	0,01271
$\hat{\eta}$	0,00304	0,00122

Cuadro 3.3: Sesgo y error cuadrático medio para la estimación de parámetros de un campo aleatorio *two-piece* t de student con función de correlación Wendland Generalizado bajo verosimilitud compuesta por parejas ponderada.

Parámetro	Sesgo	Error cuadrático medio
$\hat{\beta}_0$	0,00201	0,00014
$\hat{\beta}_1$	0,00397	0,00036
$\hat{\delta}$	0,00083	0,00051
$\hat{\sigma}^2$	0,00461	0,01346
$\hat{\eta}$	0,00570	0,00117
$\hat{h}$	-0,00144	0,00089

Cuadro 3.4: Sesgo y error cuadrático medio para la estimación de parámetros de un campo aleatorio *two-piece* Tukey- $h$  con función de correlación Wendland Generalizado bajo verosimilitud compuesta por parejas ponderada.

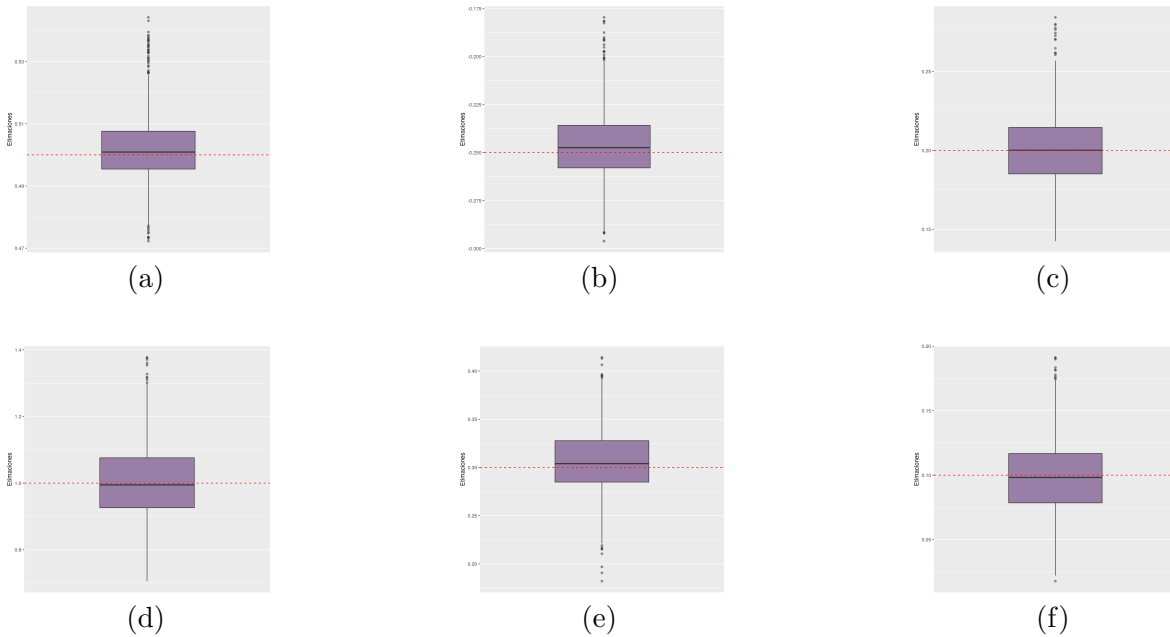


Figura 3.2: Gráficos de caja para la estimación de parámetros del campo aleatorio *two-piece* Tukey- $h$  para: (a)  $\beta_0 = 0,5$  , (b)  $\beta_1 = -0,25$ , (c)  $\delta = 0,2$  , (d)  $\sigma^2 = 1$ , (e)  $\eta = 0,3$  y (g)  $h = 0,1$ .

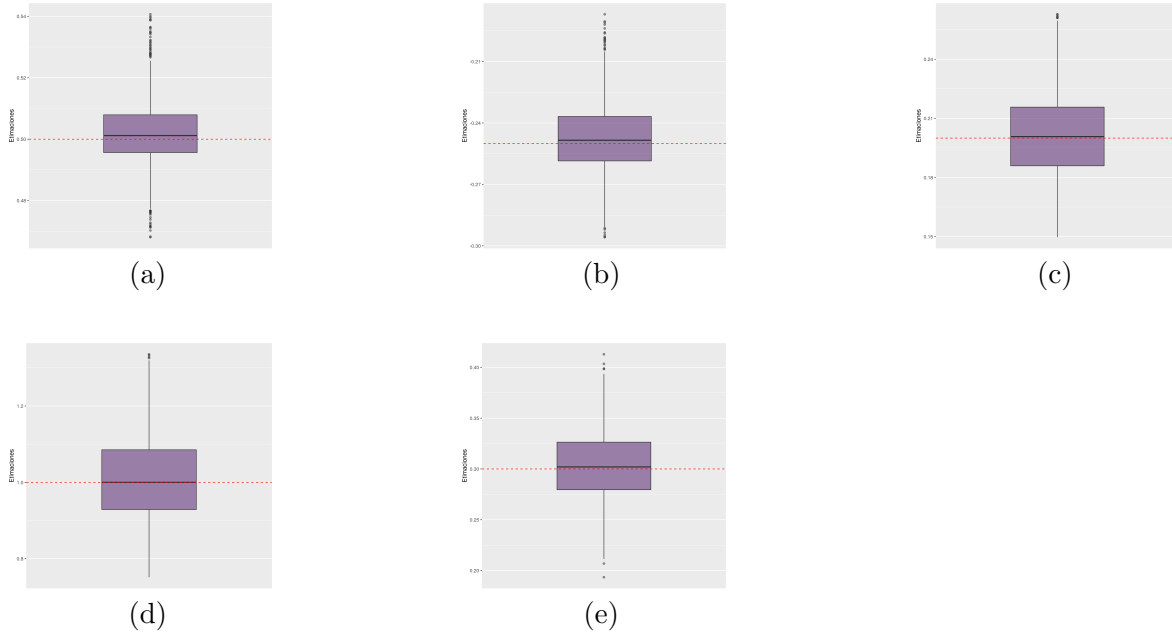


Figura 3.3: Gráficos de caja para la estimación de parámetros del campo aleatorio *two-piece t* de student para: (a)  $\beta_0 = 0,5$  , (b)  $\beta_1 = -0,25$ , (c)  $\delta = 0,2$  , (d)  $\sigma^2 = 1$  y (e)  $\eta = 0,3$ .

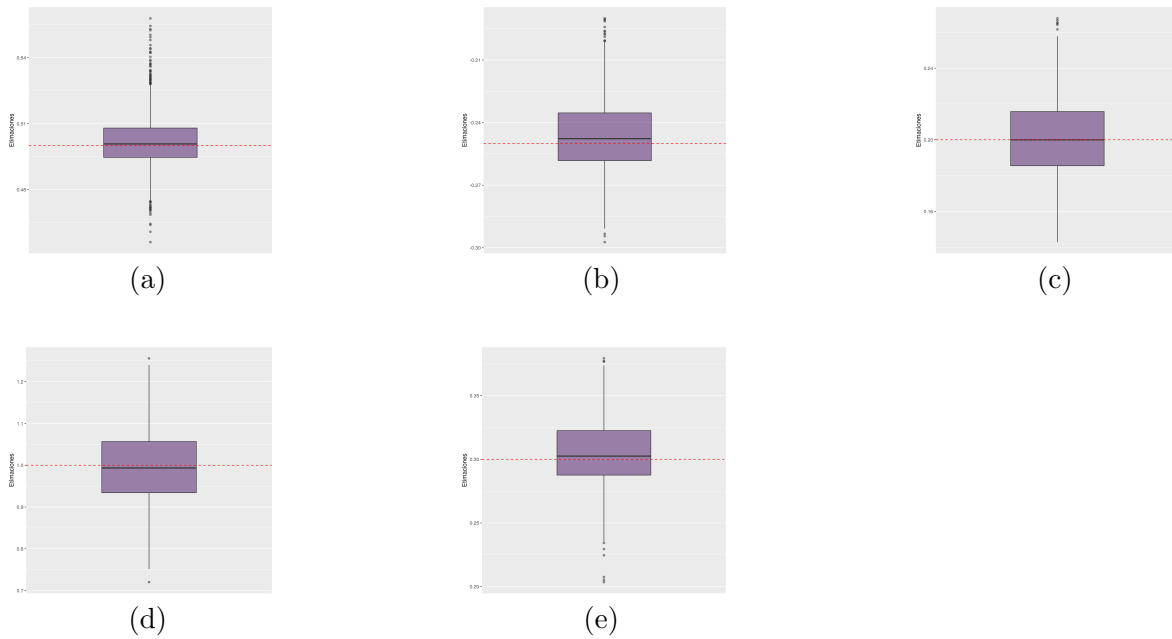


Figura 3.4: Gráficos de caja para la estimación de parámetros del campo aleatorio *two-piece* Gaussiano para: (a)  $\beta_0 = 0,5$  , (b)  $\beta_1 = -0,25$ , (c)  $\delta = 0,2$  , (d)  $\sigma^2 = 1$  y (e)  $\eta = 0,3$ .

### 3.3. Aplicación a datos reales

A modo de ilustración se estudiará la capacidad de predicción y ajuste de los campos aleatorios *two-piece* para el subconjunto de datos de *Forex* correspondiente al cambio del dólar Australiano sobre el Yen Japonés (AUDJPY) para las observaciones correspondientes al mes de febrero del año 2019. El acceso al conjunto de datos se obtuvo de manera gratuita en <https://www.truefx.com>.

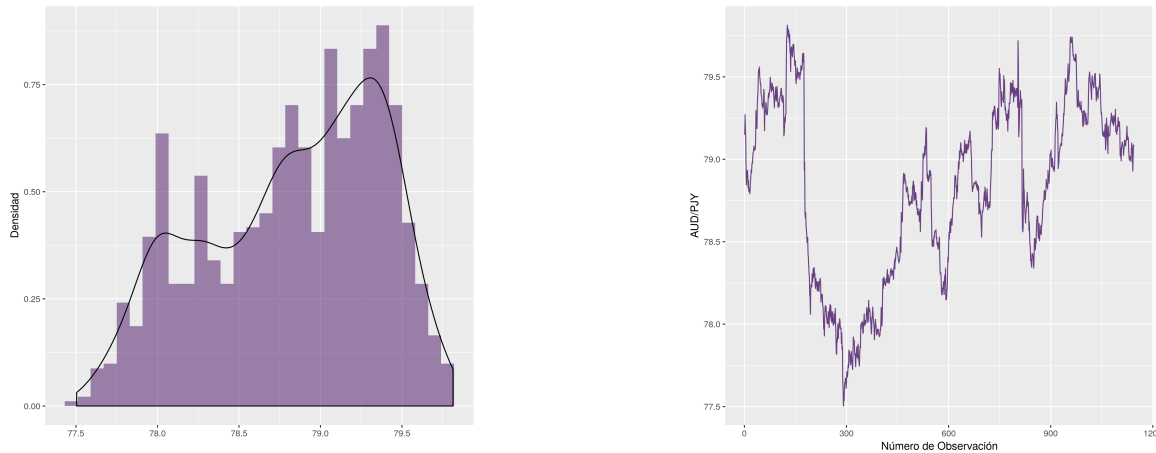


Figura 3.5: Descripción gráfica del conjunto de observaciones *Forex* de apertura para febrero del 2019 con la distribución marginal (histograma a la izquierda) y el comportamiento de la realización en el tiempo (a la derecha).

Una particularidad de estas observaciones es que los registros originales para el mes de febrero corresponden aproximadamente a 1 millón 700 mil observaciones dado que en tiempo real se registran todos los cambios producidos sobre estas monedas (registro en *ticks*). Debido a las limitaciones de capacidad computacional se ha reducido el conjunto de datos al resumen cada 25 minutos (apertura, cierre, mínimo y máximo) generando entonces una serie temporal de 1145 observaciones (Figura 3.5). El análisis que se presenta solo considerará la serie de tiempo de apertura de los valores.

Estadístico	Valor
Media	78,828
Mediana	78,9
Desviación estándar	0,541
Mínimo	77,504
Máximo	79,813
Coefficiente de asimetría	-0,387
Coefficiente de Curtosis	2,078

Cuadro 3.5: Estadística descriptiva para las observaciones de *Forex* AUDJPY de apertura.

El primer paso en este análisis es poder determinar la presencia o no de dependencia del tipo elíptica, para esto se utilizó un dispersograma disponible en R (Gräler, Pebesma, y Heuvelink, 2016) que básicamente consiste en un gráfico de dispersión de cada observación sobre todas las otras observaciones que se encuentren a determinadas distancias (temporales). Tal como se puede

observar en la Figura 3.6, la dispersión señala que existe una presencia de no elipicidad, este tipo de observaciones en otros casos se les catalogaría de observaciones atípicas dado que no conservan un comportamiento elíptico.

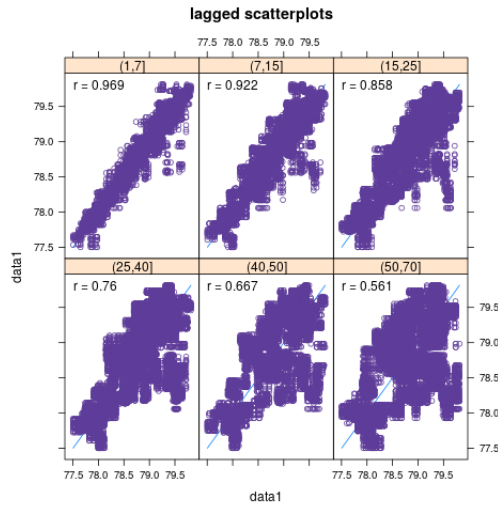


Figura 3.6: Dispersograma de las observaciones *Forex* AUDJPY a distancia temporal: 1, 7, 15, 25, 40, 50 y 70 unidades.

De la misma manera, también se presenta el semi-variograma empírico de las observaciones en la Figura 3.7. Se puede apreciar que no existe un efecto *nugget* en la realización del proceso por lo que las parametrizaciones de los modelos que a continuación se propondrán no considerarán la estimación de este parámetro.

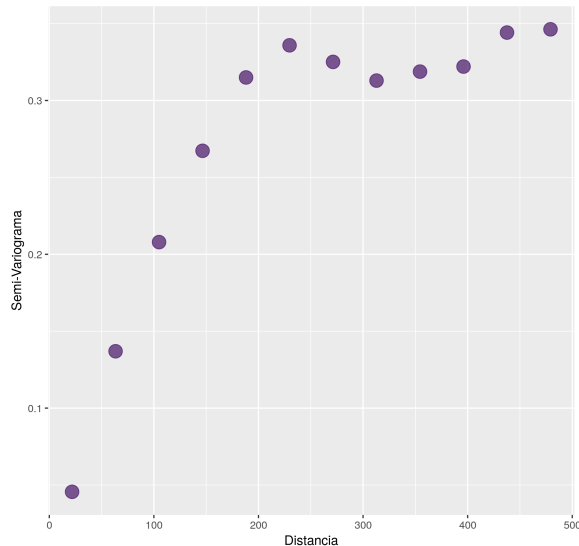


Figura 3.7: Semi-variograma empírico para las observaciones *Forex* AUDJPY.

Respecto a los modelos que se utilizarán para esta aplicación, se dispondrá de ajustar los modelos

del Cuadro 3.6 para la siguiente parametrización general:

$$Y(\mathbf{s}_i) = \beta_0 + \sigma X_{\boldsymbol{\theta}}^*(\mathbf{s}_i) \ , i = 1, \dots, 1145. \quad (3.2)$$

Considerando que el proceso  $Y$  serán las mediciones del cambio de divisas AUDJPY,  $\beta$  la media constante,  $\sigma$  el parámetro de dispersión y  $\boldsymbol{\theta}$  el conjunto de parámetros asociado al proceso estándar  $X^*$  según corresponda el modelo (Cuadro 3.6). Finalmente, para la función de correlación se escogerá el modelo Wendland Generalizado (3.1) asumiendo el parámetro de suavidad fijo y conocido ( $\gamma = 4$ ).

Modelo
Gaussiano
Tukey- $h$
Gaussiano asimétrico
<i>two-piece</i> Gaussiano
<i>two-piece</i> Tukey- $h$

Cuadro 3.6: Lista de modelos de campos aleatorios para la aplicación en las observaciones de *Forex*.

De manera similar al método de máxima verosimilitud, se estimarán los errores de estimación a través de una equivalencia a la matriz de información de Fisher, en este caso es la matriz de información de Godambe ( $\mathcal{G}_n(\theta)$ ). Esta matriz se obtiene de la misma manera que para el método de máxima verosimilitud, pero a través de la función de verosimilitud compuesta por parejas ponderada (1.21) obteniendo la siguiente expresión:

$$\mathcal{G}_n(\theta) = \mathcal{H}_n(\theta) \mathcal{J}_n(\theta)^{-1} \mathcal{H}_n(\theta)^T \ , \quad (3.3)$$

con  $\mathcal{H}_n(\theta) = \mathbb{E}[-\nabla pl(\theta)]$  y  $\mathcal{J}_n(\theta) = \text{Var}[\nabla pl(\theta)]$  (Bevilacqua y cols., 2019a), con  $pl(\theta)$  la función de verosimilitud compuesta por parejas ponderada (1.21) . Según la consideración de tiempos de cómputo establecida en Bevilacqua y cols. (2012),  $\mathcal{J}_n(\theta)$  se puede estimar a través de técnicas de remuestreo (método implementado en `GeoModels`).

Para establecer un mejor modelo para la modelización de las observaciones, se propone aplicar dos criterios de selección de modelos basados en la verosimilitud compuesta por parejas ponderada tal como se propone en Bevilacqua y cols. (2019a) como alternativa equivalentes al criterio de información de Akaike (PLIC) y el criterio de información bayesiana (BLIC), expresiones que se presentan a continuación:

$$PLIC = -2pl(\hat{\boldsymbol{\theta}}) + 2tr(\mathcal{H}_n(\hat{\boldsymbol{\theta}})\mathcal{G}^{-1}(\hat{\boldsymbol{\theta}})) \ , \quad (3.4)$$

$$BLIC = -2pl(\hat{\boldsymbol{\theta}}) + \log(n)tr(\mathcal{H}_n(\hat{\boldsymbol{\theta}})\mathcal{G}^{-1}(\hat{\boldsymbol{\theta}})) \ . \quad (3.5)$$

Luego del ajuste de modelos y concluyendo con la aplicación, se medirá en un segundo análisis el desempeño en cuanto a la predicción se utilizará el error absoluto medio (MAE por su descripción en inglés, *mean absolute error*) y la raíz cuadrada del error cuadrático medio (RMSE por su descripción en inglés, *root mean squared error*). Ambos valores se calculan al realizar la estimación de parámetros con una fracción de los datos y realizar predicción sobre las observaciones restantes, en

específico se realizará esto para el 50, 60, 75, 90 y 95 por ciento de las observaciones y se computará:

$$MAE = \frac{1}{m} \sum_i^m |y_i - \widehat{Y}_j(\mathbf{s})| , \quad (3.6)$$

$$RMSE = \frac{1}{m} \sqrt{\sum_i^m (y_i - \widehat{Y}_j(\mathbf{s}))^2} , \quad (3.7)$$

donde  $m$  es la cantidad de observaciones sobre las que se estimarán los parámetros,  $\widehat{Y}(\mathbf{s}_i)$  el valor predicho sobre la  $i$ -ésima localización  $\mathbf{s}$  y  $y_i$  el verdadero valor de la  $i$ -ésima observación.

A modo de diagnóstico de los modelos propuestos, se calculan los residuos estandarizados de la forma:

$$\widehat{Y}^*(\mathbf{s}_i) = \frac{Y^*(\mathbf{s}_i) - \widehat{\beta}_0}{\widehat{\sigma}} , \quad (3.8)$$

para estudiarlos como una realización de los procesos mencionados (Cuadro 3.5) en sus versiones estándar de localización y escala para el análisis de regresión y también para estudiar el comportamiento de los semi-variogramas empíricos con respecto al ajuste del modelo de covarianza en el análisis de dependencia espacial.

Comenzando con la presentación de los resultados de esta aplicación se muestra a continuación el resumen del ajuste de los modelos propuestos para todas las observaciones y la evaluación de la estimación de los parámetros junto con el valor que alcanza la función de verosimilitud compuesta.

Resumen Estimación de Parámetros						
Modelo	$\hat{\eta}$	$\hat{h}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\sigma}^2$	Valor máx.
Gaussiano			78,828	554,859	0,293	587,08
Tukey- $h$		$6^{-15}$	78,830	550,200	0,292	587,07
Gaussiano asimétrico	-0,895		79,555	1512,189	0,022	657,01
<i>two-piece</i> Gaussiano	0,045		78,934	477,371	0,261	559,60
<i>two-piece</i> Tukey- $h$	0,132	$6^{-17}$	78,990	494,900	0,276	569,92

Cuadro 3.7: Estimación de parámetros de los modelos propuestos para los datos de *Forex* y valor que alcanza máximo verosimilitud compuesta por parejas ponderada (Valor máx.).

Una primera observación respecto a la estimación de parámetros es que para los modelo Tukey- $h$  y *two-piece* Tukey- $h$  el parámetro de pesadez de las colas ( $h$ ) de la distribución es casi nulo, por lo que es equivalente a solamente considerar los modelos Gaussianos en ambos casos para continuar con este análisis.

Una vez descartado los dos modelos Tukey- $h$ , se continua con la estimación de parámetros de los tres modelos restantes (Gaussiano, Gaussiano asimétrico y *two-piece* Gaussiano) acompañado de los errores de estimación de los parámetros y el cómputo de los criterios de selección de modelos ya que requieren mayor tiempo de cómputo debido al método en que son calculados (remuestreo). Al observar los valores obtenidos en el Cuadro 3.8 se puede apreciar de que aunque los valores estimados del parámetro de media constante ( $\hat{\beta}$ ) son similares, el error de estimación es más bajo para el modelo *two-piece* Gaussiano. Además, para ambos criterios de selección de modelos el modelo

Modelo	Resumen Ajuste de Modelos						
	$\hat{\eta}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\sigma}^2$	Valor máx.	PLIC	BLIC
Gaussiano		78,828 (0,221)	554,859 (0,094)	0,293 (179,989)	587,08	-668,161	607,759
Gaussiano asimétrico	-0,895 (0,4285)	79,555 (0,465)	1512,189 (708,980)	0,0216 (0,098)	657,01	2623,972	12553,954
<i>two-piece</i> Gaussiano	0,045 (0,041)	78,934 (0,006)	477,371 (0,0797)	0,261 (125,817)	559,60	-719,193	289,439

Cuadro 3.8: Estimación de parámetros de los modelos propuestos para los datos de *Forex* con sus errores estándar de estimación (en paréntesis), valor que alcanza máximo verosimilitud compuesta por parejas ponderada (Valor máx.) y la evaluación de los criterios de selección de modelos (PLIC y BLIC).

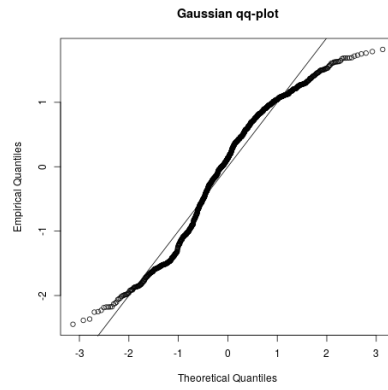
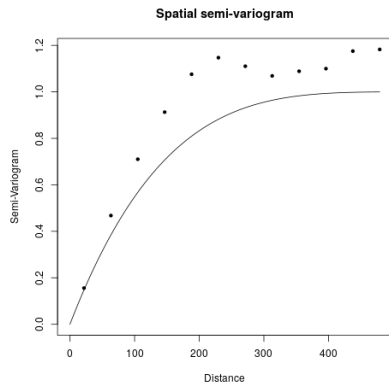
*two-piece* Gaussiano tiene los valores más bajos lo que indicaría de que el modelo con dependencia no elíptica está capturando la información que se pudo apreciar en la Figura 3.6.

La siguiente mejora que presenta el modelo *two-piece* Gaussiano es que el desempeño de predicción también se beneficia al considerar estos modelos con respecto a los otros propuestos ya que disminuyen el error al evaluar el RMSE y MAE (Cuadro 3.9).

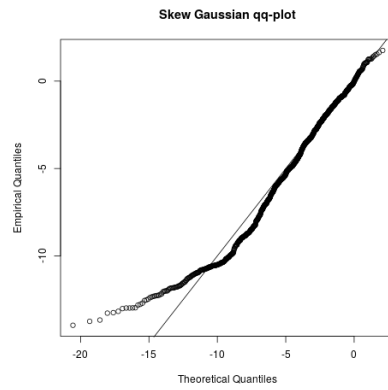
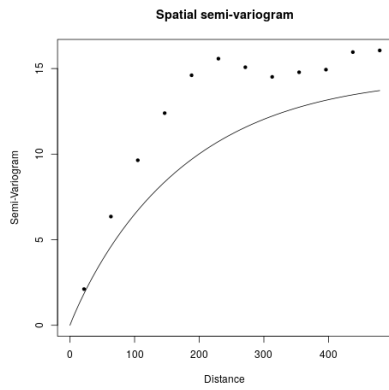
% de Obs.	Modelo	RMSE	MAE
50	Gaussiano	0,5709	0,4988
	Gaussiano asimétrico	0,5907	0,5179
	<i>two-piece</i> Gaussiano	0,4673	0,4013
60	Gaussiano	0,4455	0,3193
	Gaussiano asimétrico	0,4361	0,3130
	<i>two-piece</i> Gaussiano	0,4211	0,3004
75	Gaussiano	0,3394	0,1779
	Gaussiano asimétrico	0,3378	0,1772
	<i>two-piece</i> Gaussiano	0,2946	0,1522
90	Gaussiano	0,0410	0,0107
	Gaussiano asimétrico	0,0567	0,0169
	<i>two-piece</i> Gaussiano	0,0257	0,0071
95	Gaussiano	0,0174	0,0031
	Gaussiano asimétrico	0,0174	0,0031
	<i>two-piece</i> Gaussiano	0,0129	0,0025

Cuadro 3.9: Desempeño de predicción bajo un predictor óptimo lineal de los modelos propuestos en las observaciones de *Forex* para distintos porcentajes de observaciones en la estimación (% de Obs.), RMSE y MAE.

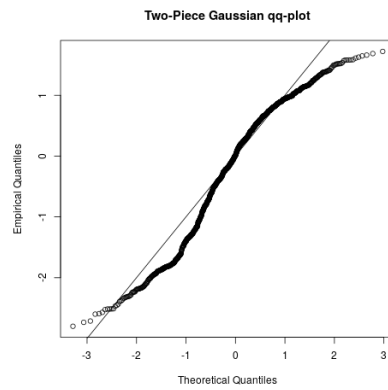
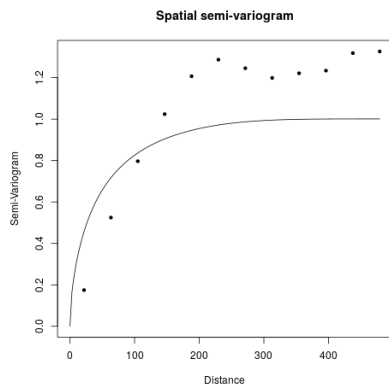
Finalmente, se puede apreciar también que el modelo *two-piece* presenta una mejora en el análisis espacial de residuos con respecto a los otros modelos tal como se aprecia en la Figura 3.8 dado que logra ajustar mejor en modelo de covarianza a las observaciones.



(a)



(b)



(c)

Figura 3.8: Gráficas para el diagnóstico distribucional (izquierda) y espacial (derecha) de los modelos: (a) Gaussiano, (b) Gaussiano asimétrico y (c) *two-piece* Gaussiano.

Respecto al análisis de regresión y los residuos, la mejora no es demasiado ya que la distribución marginal de los datos presenta multimodalidad.

### 3.4. Notas del capítulo

En este capítulo se ha estudiado el desempeño de estimación para los modelos *two-piece* y también una manera de hacer predicciones para nuevas observaciones en un contexto de aplicación de los modelos. Sobre la simulación y estimación de parámetros, el método de verosimilitud compuesta por parejas ponderada entrega una solución viable en el sentido de poder estimar los parámetros del modelo debido a que no se puede por máxima verosimilitud.

En la aplicación se pudo constatar de que el modelo de dependencia no elíptica si es útil para mejorar la precisión de modelización de observaciones dependientes en el tiempo dado que se reduce el error de predicción y también mejora el ajuste de un modelo de covarianza que alcanza a capturar dependencia en un rango más amplio de lo capturado por los modelos elípticos.

Otro aspecto novedoso de estas pruebas numéricas es que las funciones están disponibles en el paquete en desarrollo de R, `GeoModels`, por lo que se encuentra disponible para futuros usuarios los modelos *two-piece* para el análisis de observaciones temporales, espaciales o espacio temporales en contextos Euclidianos o en la esfera.

## Capítulo 4

# Comentarios Finales

A modo de capítulo de cierre para este trabajo, se terminará con algunas conclusiones respecto a los modelos y metodologías aquí presentadas. También se proporcionará una lista de preguntas abiertas para futura investigación en el sentido de extensión de casos particulares y otro tipos de campos aleatorios no Gaussianos.

## 4.1. Conclusiones

En este trabajo se han propuesto casos particulares a partir de una clase de campos aleatorios no Gaussianos para la modelización de observaciones espacialmente dependientes continuas en  $\mathbb{R}$  con posible asimetría y colas pesadas, modelos que hasta el momento no existen en la literatura y que también se pueden llevar a contextos espacio-temporales o en la esfera. Si bien solo se estudiaron tres casos particulares del modelo, se podrían considerar más casos particulares a modo de extensión de la clase siempre que se puedan obtener expresiones algebraicas cerradas; algunas posibilidades son propuestas en la siguiente sección como preguntas abiertas de investigación.

Como un resultado novedoso, las distribuciones bivariadas asociadas a los campos aleatorios podrían ser estudiadas en el contexto de modelos multivariados dado que hasta el momento han sido muy poco estudiadas en la literatura la dependencia del tipo no elíptica. Esta característica en el contexto de campos aleatorios significa ampliar el rango de aplicaciones de estos modelos en contextos de observaciones con dependencia no elíptica comprobable (dispersograma), asimetría o colas pesadas.

Posibles limitaciones en los modelos de campos aleatorios es que si bien estos son continuos en media cuadrática, no son diferenciables en media cuadrática por lo que es una limitante de flexibilidad en los modelos. Esta característica puede ser interesante dado que algunos modelos de correlación con soporte compacto en la esfera tienen la particularidad de no permitir la diferenciable en media cuadrática como condición de funciones definidas positivas.

La propuesta de utilizar el método de verosimilitud compuesta por parejas ponderada ([Bevilacqua y cols., 2012](#)) es una alternativa que resuelve de manera simple la estimación de parámetros cuando no se puede generar una expresión cerrada para las distribuciones finito dimensionales asociadas a los campos aleatorios. Además, el método estudiado es eficiente desde el punto de vista computacional respecto a los tiempos de cómputo por lo que sigue siendo una alternativa atractiva para realizar la estimación de parámetros para campos aleatorios no Gaussianos en ausencia de utilizar la máxima verosimilitud.

A modo de consideración importante sobre la estimación de parámetros con el método de verosimilitud compuesta por parejas ponderada es que el proceso de selección de los valores iniciales para el proceso de optimización debe ser iteradamente revisado con estimaciones y con los resultados a través de herramientas como el variograma ya que dependiendo de los valores iniciales (por ejemplo los parámetros de la función de correlación) se obtienen ajustes óptimos o resultados pobres en la modelización. Así mismo, se puede recomendar que para los parámetros de regresión se pueden utilizar como valores iniciales la media y varianza empírica de la variable de interés.

Finalmente, este trabajo cuenta con implementaciones computacionales de libre acceso de los modelos aquí estudiados permitiendo así que futuros usuarios puedan acceder a modelizar observaciones espacialmente dependientes con los modelos *two-piece*, todo compactado en `GeoModels` ([Bevilacqua y Morales-Oñate, 2018](#)).

## 4.2. Preguntas abiertas

Tal como se mencionó en el Capítulo 2 la clase de campos aleatorios *two-piece* permite realizar modificaciones en las componentes de su construcción. A continuación se presentan algunas preguntas abiertas de investigación que se pueden derivar como otros casos particulares de la clase *two-piece* así como también derivaciones de otros modelos que se pueden generar a partir de los campos aleatorios *two-piece*.

1. Introducir campos aleatorios *two-piece* con distribuciones marginales con colas livianas en comparación a los campos aleatorios Gaussianos. Una alternativa es considerar una transformación ArcSinh a un campo aleatorio Gaussiano estándar ( $G^*$ ) con la siguiente representación estocástica:

$$S_{\alpha}^*(\mathbf{s}) = \text{ArcSinh} \left( \frac{\alpha G^*(\mathbf{s})}{2} \right), \quad (4.1)$$

produciendo así un modelo con colas menos pesadas ( $\alpha \in \mathbb{R}_{(0,2)}$ ) y bimodalidad ( $\alpha \in \mathbb{R}_{(2,\infty+)}$ ) tal como se puede apreciar en la siguiente figura:

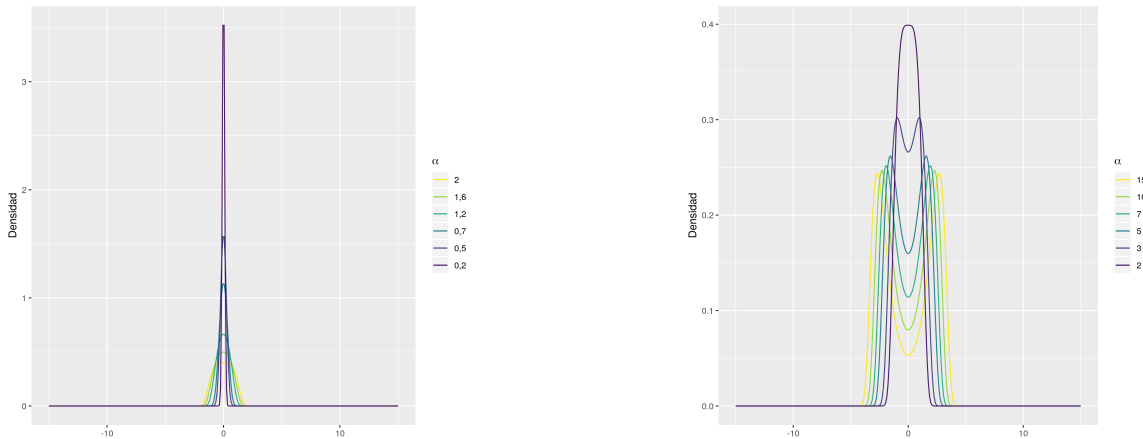


Figura 4.1: Ilustración sobre características de la distribución marginal de un campo aleatorio ArcSinh Gaussiano estándar.

Una de las dificultades que tiene estudiar este modelo es la obtención de expresiones cerradas para la función de covarianza del modelo.

2. Considerar dentro de la construcción *two-piece* (2.4) campos aleatorios con distribuciones marginales en  $\mathbb{R}^+$ . Esta propuesta tiene la particularidad de generar campos aleatorios continuos en  $\mathbb{R}$  y con opciones de estudiar bimodalidad en las observaciones debido al aumento en sus parámetros de forma (ver ejemplo en Figura 4.2), por ejemplo se podría introducir en la construcción Two-Piece (2.4) un campo aleatorio Weibull con parámetro de forma  $\varphi \in \mathbb{R}^+$  (Bevilacqua y cols., 2019b).

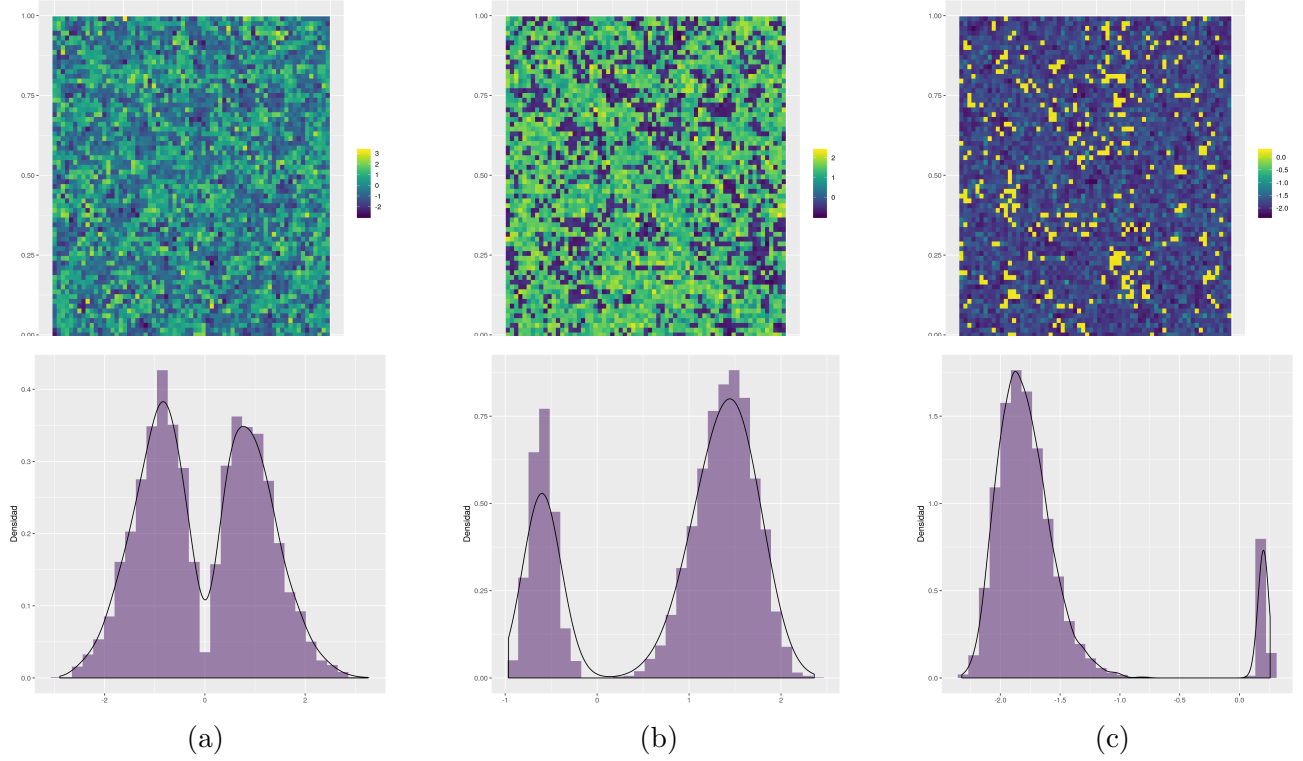


Figura 4.2: Realizaciones de campos aleatorios *two-piece* Weibull parametrizados por: (a)  $P_{\eta=0; \varphi=2}^*$ , (b)  $P_{\eta=-0,4; \varphi=5}^*$  y (c)  $P_{\eta=0,8; \varphi=10}^*(\mathbf{h})$  con función de correlación Matén  $\mathcal{M}_{(0,5), (0,015)}(\mathbf{h})$ .

3. Otra línea de campos aleatorios no Gaussianos que puede abrirse con los modelos *two-piece* es introducir casos específicos para generar campos aleatorios discretos, en especial, campos aleatorios binarios que puedan lidiar con cantidades desbalanceadas de eventos de interés y de no interés al igual que en el caso de regresión, modelos llamados *skew-probit* (Bazán, Bolfarine, y Branco, 2010).

Una manera de generar estos modelos es considerar la representación estocástica propuesta por Heagerty y Lele (1998) para modelos probit espacial y en vez de utilizar un campo aleatorio Gaussiano se reemplaza algún caso particular de los procesos *two-piece* tal como se muestra a continuación:

$$B_{\eta}(\mathbf{s}) = \begin{cases} 0 & ; \quad P_{\eta}(\mathbf{s}) < 0 \\ 1 & ; \quad P_{\eta}(\mathbf{s}) \geq 0 \end{cases} \quad (4.2)$$

El problema no trivial que tienen estos campos aleatorios binarios es la falta de identificabilidad de parámetros lo que impide poder estimar por separado el parámetro de asimetría y el de media constante para la regresión dado que algunas reparametrizaciones frecuentes en la regresión *skew-probit* perjudican la flexibilidad del modelo binario respecto a las cantidades de eventos de interés y de no interés.

# Apéndice A

## Demostraciones

Para el desarrollo de las demostraciones asociadas a la construcción del proceso espacial *two-piece* y las funciones de probabilidad asociadas se debe considerar la representación estocástica y la definición general de estos procesos.

**Definición :** (*Proceso Espacial two-piece*)

Se define  $P_\eta = \{P_\eta(\mathbf{s}), \mathbf{s} \in S\}$  como un proceso espacial o campo aleatorio del tipo *two-piece* si su representación estocástica está dada por:

$$P_\eta(\mathbf{s}) = \mu(\mathbf{s}) + \sigma|X(\mathbf{s})|K_\eta(\mathbf{s}) , \quad (\text{A.1})$$

con un conjunto de localizaciones espaciales  $\mathbf{s} \in S \subseteq \mathbb{R}^d$  y  $d \in \mathbb{N}$ ,  $X = \{X(\mathbf{s}), \mathbf{s} \in S\}$  un proceso espacial estándar con distribuciones marginales del tipo simétricas, un parámetro de asimetría  $\eta \in \mathbb{R}_{(-1,1)}$ , una función o parámetro de localización  $\mu(\mathbf{s}) \in \mathbb{R}$ , un parámetro de escala  $\sigma \in \mathbb{R}^+$  y  $K = \{K_\eta(\mathbf{s}), \mathbf{s} \in S\}$  un campo aleatorio discreto definido por:

$$K_\eta(\mathbf{s}) = \begin{cases} 1 - \eta & ; & H(\mathbf{s}) < q_{g(\eta)} , \\ -1 - \eta & ; & H(\mathbf{s}) \geq q_{g(\eta)} . \end{cases} \quad (\text{A.2})$$

considerando  $\mathbb{P}(K_\eta(\mathbf{s}) = 1 - \eta) = g(\eta) = \frac{1-\eta}{2}$ ,  $H = \{H(\mathbf{s}), \mathbf{s} \in S\}$  es un segundo campo aleatorio, independiente de  $X$ , con distribuciones marginales Gaussiana estándar y  $q_\alpha$  el cuantil  $\alpha$  de la distribución marginal de  $H$ .

Una segunda consideración es una equivalencia probabilística de la función de distribución acumulada bivariada (Dobrow, 2014).

**Propiedad:** (*Función de Distribución Acumulada Bivariada*)

$$\mathbb{P}(g_{11} \leq G_1 \leq g_{12} ; g_{21} \leq G_2 \leq g_{22}) = \mathbb{P}(G_1 \leq g_{11} ; G_1 \leq g_{21}) + \mathbb{P}(G_1 \leq g_{12} ; G_1 \leq g_{22}) - \mathbb{P}(G_1 \leq g_{11} ; G_1 \leq g_{22}) - \mathbb{P}(G_1 \leq g_{12} ; G_1 \leq g_{21}) . \quad (\text{A.3})$$

## A.1. Densidad univariada *two-piece*

Desde la representación estocástica de los procesos espaciales *two-piece* (A.1) y su versión estandarizada en localización y escala ( $P_\eta^* = \frac{P_\eta - \mu}{\sigma}$ ), se calcula la función de densidad de probabilidad de la distribución *two-piece* estándar a través del método de transformación de variables aleatorias de la siguiente manera:

$$\mathbb{P}[P_\eta^* \leq t^*] = \mathbb{P}[|X|K_\eta \leq t^*] . \quad (\text{A.4})$$

Dado que no se tiene una expresión para la distribución conjunta de  $|X|K_\eta$ , se reemplaza por la equivalencia del teorema de Bayes ( $\mathbb{P}[A \cap B] = \mathbb{P}[B]\mathbb{P}[A/B]$ ) y se escribe A.4 en los dos segmentos que componen la construcción del proceso espacial *two-piece* (A.1).

$$\begin{aligned} \mathbb{P}[P_\eta^* \leq t^*] &= \mathbb{P}[|X|(1-\eta) \leq t^*/H < q_{g(\eta)}] \mathbb{P}[H < q_{g(\eta)}] + \\ &\quad \mathbb{P}[|X|(-1-\eta) \leq t^*/H \geq q_{g(\eta)}] \mathbb{P}[H \geq q_{g(\eta)}] , \\ &= \mathbb{P}[|X|(1-\eta) \leq t^*] \mathbb{P}[H < q_{g(\eta)}] + \\ &\quad \mathbb{P}[-t^* \leq |X|(1+\eta)] \mathbb{P}[H \geq q_{g(\eta)}] , \\ &= \mathbb{P}[-t^* \leq X(1-\eta) \leq t^*] \mathbb{P}[H < q_{g(\eta)}] + \\ &\quad \{\mathbb{P}[X(1+\eta) \leq t^*] + \mathbb{P}[X(1+\eta) \geq -t^*]\} \mathbb{P}[H \geq q_{g(\eta)}] . \end{aligned} \quad (\text{A.5})$$

Utilizando la misma propiedad de equivalencia de la función de distribución acumulada (A.3) para el caso univariado, es decir,  $\mathbb{P}(g_{11} \leq G_1 \leq g_{12}) = \mathbb{P}(G_1 \leq g_{12}) - \mathbb{P}(G_1 \leq g_{11})$ , además de suponer que la función de distribución acumulada asociada a  $X$  es  $F$ , la expresión A.5 se puede reescribir como:

$$\begin{aligned} \mathbb{P}[P_\eta^* \leq t^*] &= \left\{ \frac{1-\eta}{2} F\left(\frac{t^*}{1-\eta}\right) - \frac{1-\eta}{2} F\left(\frac{-t^*}{1-\eta}\right) \right\} I(H < q_{g(\eta)}) + \\ &\quad \left\{ \frac{1+\eta}{2} F\left(\frac{t^*}{1+\eta}\right) + \frac{1+\eta}{2} \left(1 - \Phi\left(\frac{-t^*}{1+\eta}\right)\right) \right\} I(H \geq q_{g(\eta)}) . \end{aligned} \quad (\text{A.6})$$

En este punto se tiene una manera de escribir la función de distribución acumulada de la distribución *two-piece* estándar, pero la expresión depende de la función indicatriz ( $I$ ) evaluada en la desigualdad evaluada en el cuantil  $q_{g(\eta)}$  de la distribución marginal asociada a  $H$  (A.2), por lo que se debe recurrir a cambiar esos límites para que dependan de  $t^*$ . Para cambiar los límites solo se deben analizar los signos de las expresiones que componen la construcción del campo aleatorio *two-piece* estándar tal como se describe a continuación.

$$P_\eta^* = \begin{cases} (1-\eta)|X| & ; H < q_{g(\eta)} , \\ (-1-\eta)|X| & ; H \geq q_{g(\eta)} . \end{cases} \quad (\text{A.7})$$

Desde la construcción anterior (A.7), de debe considerar que  $|X|$  es siempre positivo. Entonces evaluar la primera parte de la construcción ( $H < q_{g(\eta)}$ ) implica que  $(1-\eta)|X|$  es siempre positivo ya que  $1-\eta$  toma valores positivos para todos los valores posibles de  $\eta \in \mathbb{R}_{(-1,1)}$ .

El mismo análisis se realiza para la segunda parte de la construcción ( $H \geq q_{g(\eta)}$ ) y finalmente la función de distribución acumulada, considerando el cambio dentro de las funciones indicatrices, queda escrito de la siguiente manera:

$$F_\eta(t^*) = \left\{ \frac{1-\eta}{2} F\left(\frac{t^*}{1-\eta}\right) - \frac{1-\eta}{2} F\left(\frac{-t^*}{1-\eta}\right) \right\} I(t^* > 0) + \left\{ \frac{1+\eta}{2} F\left(\frac{t^*}{1+\eta}\right) + \frac{1+\eta}{2} \left(1 - F\left(\frac{-t^*}{1+\eta}\right)\right) \right\} I(t^* \leq 0) .$$

Considerando el cambio  $F(t^*) = 1 - F(-t^*)$  y realizando la simplificación de las expresiones, la función de distribución acumulada de la distribución *two-piece* estándar es:

$$F_{\eta}(t^*) = \begin{cases} (1+\eta)F\left(\frac{t^*}{1+\eta}\right) & ; t^* \leq 0 , \\ \eta + (1-\eta)F\left(\frac{t^*}{1-\eta}\right) & ; t^*(s) > 0 . \end{cases} \quad (\text{A.8})$$

Un detalle a considerar es que la segunda sección de [A.8](#) incluye la expresión  $\eta$  dado que es lo correspondiente a incluir la primera sección de [A.8](#) evaluada en el límite superior, es decir, en cero.

Finalmente, para obtener la función de distribución de probabilidad de la distribución *two-piece* estándar basta con derivar cada parte de la construcción ([A.8](#)), dando como resultado:

$$f_{\eta}(t^*) = \begin{cases} f\left(\frac{t^*}{1+\eta}\right) & ; t^* \leq 0 , \\ f\left(\frac{t^*}{1-\eta}\right) & ; t^* > 0 . \end{cases} \quad (\text{A.9})$$

## A.2. Densidad bivariada *two-piece*

Para la construcción de la densidad bivariada *two-piece* se utiliza la misma representación estocástica (A.1), pero ahora considerando un vector aleatorio de la siguiente manera:

$$\begin{bmatrix} P_\eta(\mathbf{s}_i) \\ P_\eta(\mathbf{s}_j) \end{bmatrix} = \begin{bmatrix} \mu(\mathbf{s}_i) \\ \mu(\mathbf{s}_j) \end{bmatrix} + \sigma \begin{bmatrix} |X(\mathbf{s}_i)| K_\eta(\mathbf{s}_i) \\ |X(\mathbf{s}_j)| K_\eta(\mathbf{s}_j) \end{bmatrix} . \quad (\text{A.10})$$

El campo aleatorio discreto  $K_\eta$  cumple las mismas condiciones marginalmente para cada localización espacial al igual que en (A.2). Entonces la construcción de la densidad bivariada estándar, siguiendo la metodología para el caso univariado, está dada por:

$$\mathbb{P}[P_{\eta;i}^* \leq t_i^* ; P_{\eta;j}^* \leq t_j^*] = \mathbb{P}[|X_i|K_{\eta;i} \leq t_i^* ; |X_j|K_{\eta;j} \leq t_j^*] .$$

Dado que no se conoce la distribución conjunta de  $[|X_i|K_{\eta;i} ; |X_j|K_{\eta;j}]$  se procede a descomponer la expresión en secciones de la construcción de la siguiente manera:

$$\begin{aligned} \mathbb{P}[P_{\eta;i}^* \leq t_i^* ; P_{\eta;j}^* \leq t_j^*] &= \mathbb{P}[(1-\eta)|X_i| \leq t_i^* ; (1-\eta)|X_j| \leq t_j^*/H_i < q_{g(\eta)} ; H_j < q_{g(\eta)}] \\ &\quad \mathbb{P}[H_i < q_{g(\eta)} ; H_j < q_{g(\eta)}] + \\ &\quad \mathbb{P}[(1-\eta)|X_i| \leq t_i^* ; (-1-\eta)|X_j| \geq t_j^*/H_i < q_{g(\eta)} ; H_j \geq q_{g(\eta)}] \\ &\quad \mathbb{P}[H_i < q_{g(\eta)} ; H_j \geq q_{g(\eta)}] + \\ &\quad \mathbb{P}[(1-\eta)|X_i| \geq t_i^* ; (1-\eta)|X_j| \leq t_j^*/H_i \geq q_{g(\eta)} ; H_j < q_{g(\eta)}] \\ &\quad \mathbb{P}[H_i \geq q_{g(\eta)} ; H_j < q_{g(\eta)}] + \\ &\quad \mathbb{P}[(1-\eta)|X_i| \geq t_i^* ; (-1-\eta)|X_j| \geq t_j^*/H_i \geq q_{g(\eta)} ; H_j \geq q_{g(\eta)}] \\ &\quad \mathbb{P}[H_i \geq q_{g(\eta)} ; H_j \geq q_{g(\eta)}] , \\ &= \mathbb{P}[|X_i| \leq \frac{t_i^*}{1-\eta} ; |X_j| \leq \frac{t_j^*}{1-\eta}] \mathbb{P}[H_i < q_{g(\eta)} ; H_j < q_{g(\eta)}] + \\ &\quad \mathbb{P}[|X_i| \leq \frac{t_i^*}{1-\eta} ; |X_j| > \frac{-t_j^*}{1+\eta}] \mathbb{P}[H_i < q_{g(\eta)} ; H_j \geq q_{g(\eta)}] + \\ &\quad \mathbb{P}[|X_i| \geq \frac{-t_i^*}{1+\eta} ; |X_j| < \frac{t_j^*}{1-\eta}] \mathbb{P}[H_i \geq q_{g(\eta)} ; H_j < q_{g(\eta)}] + \\ &\quad \mathbb{P}[|X_i| \geq \frac{-t_i^*}{1+\eta} ; |X_j| \geq \frac{-t_j^*}{1+\eta}] \mathbb{P}[H_i \geq q_{g(\eta)} ; H_j \geq q_{g(\eta)}] . \end{aligned}$$

Se puede extender la notación dentro de las probabilidades para utilizar la propiedad (A.3) de la siguiente manera:

$$\begin{aligned} \mathbb{P}[P_{\eta;i}^* \leq t_i^* ; P_{\eta;j}^* \leq t_j^*] &= \mathbb{P}[|X_i| \leq \frac{t_i^*}{1-\eta} ; |X_j| \leq \frac{t_j^*}{1-\eta}] \mathbb{P}[H_i < q_{g(\eta)} ; H_j < q_{g(\eta)}] + \\ &\quad \mathbb{P}[\infty^- \leq |X_i| \leq \frac{t_i^*}{1-\eta} ; \frac{-t_j^*}{1+\eta} \leq |X_j| \leq \infty^+] \mathbb{P}[\infty^- < H_i < q_{g(\eta)} ; q_{g(\eta)} < H_j < \infty^+] + \\ &\quad \mathbb{P}[\frac{-t_i^*}{1+\eta} \leq |X_i| \leq \infty^+ ; \infty^- \leq |X_j| \leq \frac{t_j^*}{1-\eta}] \mathbb{P}[q_{g(\eta)} < H_i < \infty^+ ; \infty^- < H_j < q_{g(\eta)}] + \\ &\quad \mathbb{P}[\frac{-t_i^*}{1+\eta} \leq |X_i| \leq \infty^+ ; \frac{-t_j^*}{1+\eta} \leq |X_j| \leq \infty^+] \mathbb{P}[q_{g(\eta)} < H_i < \infty^+ ; q_{g(\eta)} < H_j < \infty^+] . \end{aligned}$$

Utilizando la propiedad (A.3) y simplificando las expresiones se obtiene la siguiente expresión:

$$\begin{aligned}
\mathbb{P}[P_{\eta;i}^* \leq t_i^* ; P_{\eta;j}^* \leq t_j^*] &= F_{\mathbf{H}_{ij}}\left(\frac{t_i^*}{1-\eta}; \frac{t_j^*}{1-\eta}\right) F_{\mathbf{X}_{ij}}(q_{g(\eta)}; q_{g(\eta)}) I_{(t_i^* \geq 0 \wedge t_j^* \geq 0)} + \\
&\quad (F_H\left(\frac{t_i^*}{1-\eta}\right) - F_{\mathbf{H}_{ij}}\left(\frac{t_i^*}{1-\eta}; \frac{-t_j^*}{1+\eta}\right)) \left(F_X(q_{g(\eta)}) - F_{\mathbf{X}_{ij}}(q_{g(\eta)}; q_{g(\eta)})\right) \\
&\quad I_{(t_i^* \geq 0 \wedge t_j^* < 0)} + \\
&\quad (F_H\left(\frac{t_j^*}{1-\eta}\right) - F_{\mathbf{H}_{ij}}\left(\frac{-t_i^*}{1+\eta}; \frac{t_j^*}{1-\eta}\right)) \left(F_X(q_{g(\eta)}) - F_{\mathbf{X}_{ij}}(q_{g(\eta)}; q_{g(\eta)})\right) \\
&\quad I_{(t_i^* < 0 \wedge t_j^* \geq 0)} + \\
&\quad (1 + F_{\mathbf{H}_{ij}}\left(\frac{-t_i^*}{1+\eta}; \frac{-t_j^*}{1+\eta}\right) - F_H\left(\frac{-t_i^*}{1+\eta}\right) - F_H\left(\frac{-t_j^*}{1+\eta}\right)) \\
&\quad \left(1 + F_{\mathbf{X}_{ij}}(q_{g(\eta)}; q_{g(\eta)}) - 2F_X(q_{g(\eta)})\right) I_{(t_i^* < 0 \wedge t_j^* < 0)}
\end{aligned}$$

Con  $F_{\mathbf{H}_{ij}}$  la función de distribución acumulada bivariada asociada a  $[H_i; H_j]^T$ . Finalmente, basta con construir la función de distribución acumulada bivariada *two-piece* por los segmentos. En primer lugar se comienza por el segmento que corresponde a evaluar la función de distribución acumulada en valores menores a cero. A medida que se continúan con los segmentos, se deben incorporar los anteriores y sus respectivas evaluaciones (en cero) según corresponda, dando como resultado final la siguiente expresión.

$$F_{\mathbf{P}_{\eta;ij}}(t_i^*, t_j^*, \rho) = \left\{ \begin{array}{l} \left\{ 1 + F_{|\mathbf{X}_{ij}|}\left(\frac{-t_i^*}{(1+\eta)}, \frac{-t_j^*}{(1+\eta)}, \rho\right) - F_{|X|}\left(\frac{-t_i^*}{(1+\eta)}\right) - F_{|X|}\left(\frac{-t_j^*}{(1+\eta)}\right) \right\} \\ \left\{ 1 + F_{\mathbf{H}_{ij}}(q_{g(\eta)}, q_{g(\eta)}, \rho) - 2F_H(q_{g(\eta)}) \right\}, \\ t_i^* \leq 0, t_j^* \leq 0. \\ \\ \left\{ 1 + F_{|\mathbf{X}_{ij}|}\left(0, \frac{-t_j^*}{(1+\eta)}, \rho\right) - F_{|X|}\left(\frac{-t_j^*}{(1+\eta)}\right) \right\} \left\{ 1 + F_{\mathbf{H}_{ij}}(q_{g(\eta)}, q_{g(\eta)}, \rho) - 2F_H(q_{g(\eta)}) \right\} \\ + \left\{ F_H(q_{g(\eta)}) - F_{\mathbf{H}_{ij}}(q_{g(\eta)}, q_{g(\eta)}, \rho) \right\} \left\{ F_{|X|}\left(\frac{t_i^*}{(1-\eta)}\right) - F_{|\mathbf{X}_{ij}|}\left(\frac{t_i^*}{(1-\eta)}, \frac{-t_j^*}{(1+\eta)}, \rho\right) \right\}, \\ t_i^* > 0, t_j^* \leq 0. \\ \\ \left\{ 1 + F_{|\mathbf{X}_{ij}|}\left(\frac{-t_i^*}{(1+\eta)}, 0, \rho\right) - F_{|X|}\left(\frac{-t_i^*}{(1+\eta)}\right) \right\} \left\{ 1 + F_{\mathbf{H}_{ij}}(q_{g(\eta)}, q_{g(\eta)}, \rho) - 2F_H(q_{g(\eta)}) \right\} \\ + \left\{ F_H(q_{g(\eta)}) - F_{\mathbf{H}_{ij}}(q_{g(\eta)}, q_{g(\eta)}, \rho) \right\} \left\{ F_{|X|}\left(\frac{t_j^*}{(1-\eta)}\right) - F_{|\mathbf{X}_{ij}|}\left(\frac{-t_i^*}{(1+\eta)}, \frac{t_j^*}{(1-\eta)}, \rho\right) \right\}, \\ t_i^* \leq 0, t_j^* > 0. \\ \\ \left\{ 1 + F_{\mathbf{H}_{ij}}(q_{g(\eta)}, q_{g(\eta)}, \rho) - 2F_H(q_{g(\eta)}) \right\} + \\ \left\{ F_H(q_{g(\eta)}) - F_{\mathbf{H}_{ij}}(q_{g(\eta)}, q_{g(\eta)}, \rho) \right\} \left\{ F_{|X|}\left(\frac{t_j^*}{(1-\eta)}\right) - F_{|\mathbf{X}_{ij}|}\left(\frac{t_i^*}{(1-\eta)}, 0, \rho\right) \right\} + \\ \left\{ F_H(q_{g(\eta)}) - F_{\mathbf{H}_{ij}}(q_{g(\eta)}, q_{g(\eta)}, \rho) \right\} \left\{ F_{|X|}\left(\frac{t_j^*}{(1-\eta)}\right) - F_{|\mathbf{X}_{ij}|}\left(0, \frac{t_j^*}{(1-\eta)}, \rho\right) \right\} + \\ \left\{ F_{|\mathbf{X}_{ij}|}\left(\frac{t_i^*}{(1-\eta)}, \frac{t_j^*}{(1-\eta)}, \rho\right) F_{\mathbf{H}_{ij}}(q_{g(\eta)}, q_{g(\eta)}, \rho) \right\}, \\ t_i^* > 0, t_j^* > 0. \end{array} \right.$$

### A.3. Función de covarianza y correlación *two-piece*

Adicionalmente, se debe considerar el cómputo de la covarianza o correlación para hacer predicciones bajo este modelo. Establecer una expresión para la covarianza y correlación del modelo involucra también describir el proceso discreto  $K_\eta$ , tal como se detalla a continuación.

Sea  $K_\eta$  un proceso espacial discreto definido por (A.2), entonces el momento de primer orden está dado por:

$$\begin{aligned}
\mathbb{E}[K_\eta] &= (1-\eta)\mathbb{P}[K_\eta=1-\eta] + (-1-\eta)\mathbb{P}[K_\eta=-1-\eta] , \\
&= (1-\eta)\frac{(1-\eta)}{2} + (1+\eta)\frac{(1+\eta)}{2} \\
&= \frac{(1-2\eta+\eta^2) - (1+2\eta+\eta^2)}{2} , \\
&= -\frac{4\eta}{2} , \\
&= -2\eta .
\end{aligned} \tag{A.11}$$

Continuando con la descripción de los momentos de  $K_\eta$  puede obtener una expresión para la varianza de la siguiente manera:

$$\mathbb{V}ar[K_\eta] = \mathbb{E}[K_\eta^2] - \mathbb{E}[K_\eta]^2 .$$

Para continuar con el cálculo de la varianza, es necesario definir un nuevo proceso, el cual está dado por:

$$K_\eta^2(\mathbf{s}) = \begin{cases} (1-\eta)^2 & ; H(\mathbf{s}) < q_g(\eta) , \\ (1+\eta)^2 & ; H(\mathbf{s}) \geq q_g(\eta) . \end{cases}$$

Entonces, el valor esperado del nuevo proceso  $K_\eta^2$  está dado por:

$$\begin{aligned}
\mathbb{E}[K_\eta^2] &= (1-\eta)^2\mathbb{P}[K_\eta=(1-\eta)^2] + (1+\eta)^2\mathbb{P}[K_\eta=(1+\eta)^2] , \\
&= \frac{(1-\eta)(1-\eta)}{(1-\eta)^2+(1+\eta)^2} + \frac{(1\eta)(1+\eta)}{(1-\eta)^2+(1+\eta)^2} , \\
&= \frac{1+6\eta^2+\eta^4}{1+\eta^2} .
\end{aligned} \tag{A.12}$$

Finalmente, con (A.12), la expresión para la varianza de  $K_\eta$  queda dada por:

$$\begin{aligned}
\mathbb{V}ar[K_\eta] &= \frac{1+6\eta^2+\eta^4}{1+\eta^2} - 4\eta^2 , \\
&= 1 - \eta^2 .
\end{aligned} \tag{A.13}$$

Así mismo se describe la covarianza para el proceso  $K_\eta$  de la siguiente manera:

$$\mathbb{C}ov[K_{\eta;i} ; K_{\eta;j}] = \mathbb{E}[K_{\eta;i}K_{\eta;j}] - \mathbb{E}[K_{\eta;i}]\mathbb{E}[K_{\eta;j}] .$$

Nuevamente es necesario definir un nuevo proceso auxiliar  $K_{\eta;i}K_{\eta;j}$ :

$$K_{\eta;i}K_{\eta;j} = \begin{cases} (1-\eta)^2 & ; H_i \wedge H_j < q_g(\eta) , \\ (1-\eta)(-1-\eta) & ; H_i < q_g(\eta) \wedge H_j \geq q_g(\eta) \text{ o} \\ (1+\eta)^2 & ; H_i \geq q_g(\eta) \wedge H_j < q_g(\eta) \\ (1+\eta)^2 & ; H_i \wedge H_j \geq q_g(\eta) . \end{cases} ,$$

Por lo que se puede obtener una expresión para  $\mathbb{E}[K_{\eta;i}K_{\eta;j}]$  de la siguiente manera:

$$\begin{aligned}
\mathbb{E}[K_{\eta;i}K_{\eta;j}] &= (1 - \eta)^2 \mathbb{P}[K_{\eta;i}K_{\eta;j} = 1 - \eta^2] + 2(1 - \eta^2)(-1 - \eta^2) \\
&\quad \mathbb{P}[K_{\eta;i}K_{\eta;j} = (1 - \eta^2)(-1 - \eta^2)] + (1 + \eta)^2 \mathbb{P}[K_{\eta;i}K_{\eta;j} = (1 + \eta)^2] , \\
&= \pi_{11}(1 - 2\eta + \eta^2 + 2 - 2\eta^2 + 1 + 2\eta + \eta^2) + (1 - \eta)(-1 - \eta)(1 - \eta) + \\
&\quad 1 + 2\eta + \eta^2 + (1 + 2\eta + \eta^2)(\eta - 1) , \\
&= 4\pi_{11} - (1 - \eta^2)(1 + \eta) + 1 + 2\eta + \eta^2 + \eta - 1 + 2\eta^2 - 2\eta + \eta^3 - \eta^2 , \\
&= 4\pi_{11} - 1 - \eta + \eta^2 - \eta^3 + 1 + 2\eta + \eta^2 + \eta - 1 + 2\eta^2 - 2\eta + \eta^3 - \eta^2 \\
&= 4\pi_{11} - 1 + 2\eta + 3\eta^2 . \tag{A.14}
\end{aligned}$$

Dado que ya está definido  $\mathbb{E}[K_{\eta;i}K_{\eta;j}]$ , se puede determinar una expresión para la covarianza del proceso espacial *two-piece* tal como se expone a continuación:

$$\begin{aligned}
\mathbb{C}ov[P_{\eta;i}^* ; P_{\eta;j}^*] &= \mathbb{E}[P_{\eta;i}^*P_{\eta;j}^*] - \mathbb{E}[P_{\eta;i}^*]\mathbb{E}[P_{\eta;j}^*] , \\
&= \mathbb{E}[|X_i||X_j|]\mathbb{E}[K_{\eta;i}K_{\eta;j}] - \\
&\quad \mathbb{E}[|X_i|]\mathbb{E}[|X_j|]\mathbb{E}[K_{\eta;i}]\mathbb{E}[K_{\eta;j}] . \tag{A.15}
\end{aligned}$$

Se puede observar que algunas de las componentes se pueden reemplazar por resultados previos, quedando la expresión anterior de la siguiente manera:

$$\mathbb{C}ov[P_{\eta;i}^* ; P_{\eta;j}^*] = (4\pi_{11} - 1 + 3\eta^2 + 2\eta)\mathbb{E}[|X_i||X_j|] - (-2\eta)^2\mathbb{E}[|X|]^2 . \tag{A.16}$$

Si se desea obtener una expresión para la correlación entre dos localizaciones espaciales para el proceso *two-piece*, basta con hacer la siguiente operación:

$$\mathbb{C}orr[P_{\eta;i}^* ; P_{\eta;j}^*] = \frac{\mathbb{C}ov[P_{\eta;i}^* ; P_{\eta;j}^*]}{\sqrt{\mathbb{V}ar[P_{\eta;i}^*]}\sqrt{\mathbb{V}ar[P_{\eta;j}^*]}} .$$

Considerando la propiedad:

$$\mathbb{V}ar[XY] = \mathbb{V}ar[X]\mathbb{V}ar[Y] + \mathbb{V}ar[X]\mathbb{E}[Y]^2 + \mathbb{V}ar[Y]\mathbb{E}[X]^2 ,$$

la expresión final para la correlación de un proceso espacial *two-piece* para dos localizaciones espaciales, está dada por:

$$\mathbb{C}orr[P_{\eta;i}^* ; P_{\eta;j}^*] = \frac{\mathbb{C}ov[P_{\eta;i}^* ; P_{\eta;j}^*]}{(1 + 3\eta^2)[\mathbb{V}ar[|X|] + \mathbb{E}[|X|]^2] - 4\eta^2\mathbb{E}[|X|]^2} .$$

## Apéndice B

# Rutinas Computacionales

En este apéndice se encuentra la rutina computacional completa para la reproducción del análisis completo de las observaciones AUDJPY de *Forex* presentada en el capítulo 3.

```
require(GeoModels)
require(gstat)
require(fields)
require(GoFKernel)
require(sn)
require(sp)
require(data.table)
require(moments)

df1 = fread("data set/bid_by_sec.csv",header = T,sep = ",")
df1 = data.frame(df1)
df1 = df1[complete.cases(df1),]

head(df1)
dim(df1)

data1 = (as.numeric(df1$open))
coords = cbind(seq(1,length(data1)),rep(0,length(data1)))

par(mfrow = c(1,2))
hist(data1)
plot(seq(1,length(data1)),data1,type = "l")
par(mfrow = c(1,1))

#####
#
# 1. Semi-variograma y dispersograma
#
#####

maxdist = max(dist(coords))
fit      = GeoVariogram(coordx = coords,data = data1,maxdist = 500)
```

```
plot(fit$centers,fit$variograms,xlab = 'h',ylim = c(0,max(fit$variograms)),
     pch = 20,main = "Semi-variogram")
```

```
DDS = as.data.frame(cbind(coords,data1))
attach(DDS)
coordinates(DDS) = ~V1+V2
hscat(data1~1, DDS, c(1,7,15,25,40,50,70),col = "#5E3C99")
```

```
#####
#
# 2. Ajuste de modelos y diagnóstico de residuos
#
#####
```

```
#####
# Selección del porcentaje del conjunto de datos para
# estimar y predecir
#####
```

```
percentage_of_train = 0.95
sel_data             = seq(1,round(length(data1)*percentage_of_train),1)
```

```
### conjunto para la estimación
datanew             = data1[sel_data]
coordsnew           = as.matrix(coords[sel_data,])
NN                  = length(datanew)
probabilities       = (1:NN)/(NN+1)
```

```
### conjunto para la predicción
data_to_pred       = data1[-sel_data]
loc_to_pred        = coords[-sel_data,]
```

```
#####
# A. Gaussiano
#####
```

```
fixed<-list(nugget=0,power2=4)
start<-list(scale=600,mean=mean(datanew),sill=var(datanew))
```

```
fit1 <- GeoFit(data=datanew,coordx=coordsnew,corrmodel="Wend0",
               maxdist=1,sensitivity=TRUE,
               start=start,fixed=fixed,model="Gaussian")
```

```
fit1
```

```

res = GeoResiduals(fit1)
vario = GeoVariogram(data = res$data, coordx = coordsnew, maxdist = 500)
GeoCovariogram(res, show.cov = FALSE, show.vario = TRUE, vario = vario, pch = 20)

gaussian.quantiles = qnorm(probabilities)
plot(sort(gaussian.quantiles), sort(c(res$data)), xlab = "", ylab = "",
      main = "Gaussian qq-plot")
abline(0,1)

#####
# B. Tukey-h
#####
tail = exp(-9)

fixed<-list(nugget=0, power2=4, tail=tail)
start<-list(scale=500, mean=mean(datanew), sill=var(datanew))#, tail = 0.2

fit2 <- GeoFit(data=datanew, coordx=coordsnew, corrmodel="Wend0",
              maxdist=1, sensitivity= TRUE,
              start=start, fixed=fixed, model="Tukeyh")
fit2

res = GeoResiduals(fit2)
vario = GeoVariogram(data = res$data, coordx = coordsnew, maxdist = 500)
GeoCovariogram(res, show.cov = FALSE, show.vario = TRUE, vario = vario, pch = 20)

tail = tail#as.numeric(fit2$param["tail"])
pTukeyh = function(x,tail){
  t = sqrt(VGAM::lambertW(tail*x*x)/tail)*sign(x)
  pnorm(t)
}

f = function(x) pTukeyh(x,tail = tail)
f.inv = inverse(f, lower = -Inf, upper = Inf)
tukeyh.quantiles = sort(as.numeric(lapply(probabilities,f.inv)))
plot(tukeyh.quantiles, sort(c(res$data)), xlab="", ylab="", main="Tukey-h qq-plot")
abline(0,1)

#####
# C. Gaussiano asimétrico
#####
start<-list(scale=400, mean=mean(datanew), sill=var(datanew), skew = -2)
fixed<-list(nugget=0, power2=4)#

fit3 = GeoFit(data= datanew, coordx= coordsnew,
              corrmodel= "Wend0", model= "SkewGaussian",
              sensitivity= TRUE, maxdist= 1,

```

```

        start=start,fixed=fixed)
fit3

res  = GeoResiduals(fit3)
vario = GeoVariogram(data = res$data,coordx = coordsnew,maxdist = 500)
GeoCovariogram(res,show.cov = FALSE,show.vario = TRUE,vario = vario,pch = 20)

omega = sqrt((fit3$param["skew"]^2 + fit3$param["sill"])/fit3$param["sill"])
alpha = fit3$param["skew"]/fit3$param["sill"]^0.5

skgauss.quantiles = qsn(probabilities,xi = 0,omega = as.numeric(omega),
                        alpha = as.numeric(alpha),solver = "RFB")
plot(sort(skgauss.quantiles),sort(c(res$data)),xlab="",ylab="",
     main="Skew Gaussian qq-plot")

#####
# D. Two-Piece Gaussiano
#####
fixed<-list(nugget=0,power2=4)
start<-list(scale=400,mean=mean(datanew),sill=var(datanew),skew=0.1)

fit4 <- GeoFit(data=datanew,coordx=coordsnew,corrmodel="Wend0",
              maxdist=1,sensitivity=TRUE,
              start=start,fixed=fixed,model="TwoPieceGaussian")
fit4

res  = GeoResiduals(fit4)
vario = GeoVariogram(data = res$data,coordx = coordsnew,maxdist = 500)
GeoCovariogram(res,show.cov = FALSE,show.vario = TRUE,vario = vario,pch = 20)

skew      = as.numeric(fit4$param["skew"])

ptpGaussian = function(x,skew){
  (1+skew)*pnorm(x/(1+skew))*I(x<0) + (skew + (1-skew)*pnorm(x/(1-skew)))*I(x>=0)
}

f = function(x) ptpGaussian(x,skew = skew)
f.inv = inverse(f,lower = -Inf,upper = Inf)

twopieceGaussian.quantiles = sort(as.numeric(lapply(probabilities,f.inv)))

plot(twopieceGaussian.quantiles,sort(c(res$data)),xlab="",ylab="",
     main="Two-Piece Gaussian qq-plot")
abline(0,1)

#####
# E. Two-Piece Tukey-h
#####

```

```

fixed<-list(nugget=0,power2=4)
start<-list(scale=450,mean=mean(data1),sill=var(data1),skew=0.1,tail=0.1)

fit5 <- GeoFit(data=datanew,coordx=coordsnew,corrmodel="Wend0",
              maxdist=1, sensitivity= TRUE,sparse = TRUE,
              start=start,fixed=fixed, model="TwoPieceTukeyh")
fit5

res  = GeoResiduals(fit5)
vario = GeoVariogram(data = res$data,coordx = coordsnew,maxdist = 500)
GeoCovariogram(res,show.cov = FALSE,show.vario = TRUE,vario = vario,pch = 20)

skew      = as.numeric(fit5$param["skew"])
tail      = as.numeric(fit5$param["tail"])

ptpTukeyh = function(x,skew,tail){
  t = sqrt(VGAM::lambertW(tail*x*x)/tail)*sign(x)
  (1+skew)*pnorm(t/(1+skew))*I(t<0) + (skew + (1-skew)*pnorm(t/(1-skew)))*I(t>=0)
}

f = function(x) ptpTukeyh(x,skew = skew,tail = tail)
f.inv = inverse(f,lower = -Inf,upper = Inf)

twopieceTukeyh.quantiles = sort(as.numeric(lapply(probabilities,f.inv)))
plot(twopieceTukeyh.quantiles,sort(c(res$data)),xlab="",ylab="",
     main="Two-Piece Tukey-h qq-plot")
abline(0,1)

#####
#
# 3. Cálculo de los criterios de selección de modelos y errores estándar
#
#####

model = fit4

Stud_m = GeoVarestbootstrap(model,K = 200,sparse = TRUE)
Stud_m$claic;
Stud_m$clbic;
Stud_m$stderr;
Stud_m$varcov;

#####
#
# 4. Desempeño de predicción
#
#####

```

```

# A. Gaussian
pr_gaussian = err_gauss_m = NULL
rmse_gauss_m = mae_gauss_m = NULL

# B. Skew Gaussian
pr_skewgaussian = err_skewgaussian_m = NULL
rmse_skewgaussian_m = mae_skewgaussian_m = NULL

# C. Two-Piece Gaussian
pr_twopiecegaussian = err_twopiecegaussian_m = NULL
rmse_twopiecegaussian_m = mae_twopiecegaussian_m = NULL

corrmodel = "Wend0"
N2          = length(datanew)

## A. Gaussiano
fixed       = list(nugget=0,power2=4)
pr_gaussian = GeoKrig(data=datanew, coordx=coordsnew,loc=loc_to_pred,
                      corrmodel=corrmodel,
                      model="Gaussian",param= as.list(c(fit1$param,fixed)))

err_gauss_m = pr_gaussian$pred-data_to_pred
rmse_gauss_m = c(rmse_gauss_m,sqrt(sum(err_gauss_m^2)/N2))
mae_gauss_m  = c(mae_gauss_m,sum(abs(err_gauss_m))/N2)

# B. Gaussiano asimétrico
fixed       = list(nugget=0,power2=4)
pr_skewgaussian = GeoKrig(data=datanew, coordx=coordsnew,
                          loc=loc_to_pred,corrmodel=corrmodel,
                          model="SkewGaussian",
                          param= as.list(c(fit3$param,fixed)))

err_skewgaussian_m = pr_skewgaussian$pred-data_to_pred
rmse_skewgaussian_m = c(rmse_skewgaussian_m,sqrt(sum(err_skewgaussian_m^2)/N2))
mae_skewgaussian_m  = c(mae_skewgaussian_m,sum(abs(err_skewgaussian_m))/N2)

## c. Two-Piece Gaussiano
fixed       = list(nugget=0,power2=4)
pr_twopiecegaussian = GeoKrig(data=datanew, coordx=coordsnew,loc=loc_to_pred,
                              corrmodel=corrmodel,
                              model="TwoPieceGaussian",
                              param= as.list(c(fit4$param,fixed)))

err_twopiecegaussian_m = pr_twopiecegaussian$pred-data_to_pred
rmse_twopiecegaussian_m = c(rmse_twopiecegaussian_m,sqrt(sum(err_twopiecegaussian_m^2)/N2))
mae_twopiecegaussian_m  = c(mae_twopiecegaussian_m,sum(abs(err_twopiecegaussian_m))/N2)

```

```
# A. Gaussiano
print("Gaussian")
mean(rmse_gauss_m);mean(mae_gauss_m);

# B. Gaussian asimétrico
print("Skew Gaussian")
mean(rmse_skewgaussian_m);mean(mae_skewgaussian_m);

# C. Two-Piece Gaussiano
print("Two-Piece Gaussian")
mean(rmse_twopiecegaussian_m);mean(mae_twopiecegaussian_m);
```

# Referencias

- Abrahamsen, P. (1997). *A review of gaussian random fields and correlation functions*. Oslo, Noruega: Norwegian Computing Center.
- Alegría, A., Bevilacqua, M., y Porcu, E. (2016). Likelihood-based inference for multivariate space-time wrapped-gaussian fields. *Journal of Statistical Computation and Simulation*, 86(13), 2583-2597. Descargado de <https://www.tandfonline.com/doi/abs/10.1080/00949655.2016.1162309> doi: <https://doi.org/10.1080/00949655.2016.1162309>
- Alegría, A., Caro, S., Bevilacqua, M., Porcu, E., y Clarke, J. (2017). Estimating covariance functions of multivariate skew-gaussian random fields on the sphere. *Spatial Statistics*, 22, 388 - 402. Descargado de <http://www.sciencedirect.com/science/article/pii/S2211675316301762> doi: <https://doi.org/10.1016/j.spasta.2017.07.009>
- Arellano-Valle, R. B., Azzalini, A., Ferreira, C. S., y Santoro, K. (2020). A two-piece normal measurement error model. *Computational Statistics and Data Analysis*, 144, 106863. Descargado de <http://www.sciencedirect.com/science/article/pii/S016794731930218X> doi: <https://doi.org/10.1016/j.csda.2019.106863>
- Arellano-Valle, R. B., Gómez, H. W., y Quintana, F. A. (2005). Statistical inference for a general class of asymmetric distributions. *Journal of Statistical Planning and Inference*, 128(2), 427 - 443. Descargado de <http://www.sciencedirect.com/science/article/pii/S0378375804000254> doi: <https://doi.org/10.1016/j.jspi.2003.11.014>
- Azzalini, A., y Capitanio, A. (2014). *The skew-normal and related families*. Cambridge, Inglaterra: Cambridge University Press.
- Babić, S., Ley, C., y Veredas, D. (2019). Comparison and classification of flexible distributions for multivariate skew and heavy-tailed data. *Symmetry*, 11(10). Descargado de <https://www.mdpi.com/2073-8994/11/10/1216> doi: <https://doi.org/10.3390/sym11101216>
- Bachoc, F. (2014). Asymptotic analysis of the role of spatial sampling for covariance parameter estimation of gaussian processes. *Journal of Multivariate Analysis*, 125, 1 - 35. Descargado de <http://www.sciencedirect.com/science/article/pii/S0047259X13002571> doi: <https://doi.org/10.1016/j.jmva.2013.11.015>
- Bachoc, F., Bevilacqua, M., y Velandia, D. (2019). Composite likelihood estimation for a gaussian process under fixed domain asymptotics. *Journal of Multivariate Analysis*, 104534. Descargado de <http://www.sciencedirect.com/science/article/pii/S0047259X18303841> doi: <https://doi.org/10.1016/j.jmva.2019.104534>
- Bazán, J. L., Bolfarine, H., y Branco, M. D. (2010). A framework for skew-probit links in binary regression. *Communications in Statistics - Theory and Methods*, 39(4), 678-697. Descargado de <https://www.tandfonline.com/doi/citedby/10.1080/03610920902783849?scroll=top&needAccess=true> doi: <https://doi.org/10.1080/03610920902783849>
- Bevilacqua, M., Caamaño, C., Arellano-Valle, R. B., y Morales-Oñate, V. (2019a). Non-gaussian geostatistical modeling using (skew) t processes. *ArXiv e-prints*. Descargado de <https://>

[arxiv.org/abs/1812.06310](https://arxiv.org/abs/1812.06310)

- Bevilacqua, M., Caamaño, C., y Gaetan, C. (2019b). On modelling positive continuous data with spatio-temporal dependence. *ArXiv e-prints*. Descargado de <https://arxiv.org/abs/1808.03829>
- Bevilacqua, M., Faouzi, T., Furrer, R., y Porcu, E. (2019c). Estimation and prediction using generalized wendland covariance functions under fixed domain asymptotics. *The Annals of Statistics*, 47(2), 828–856. Descargado de <https://projecteuclid.org/euclid.aos/1547197240#abstract> doi: <https://doi.org/10.1214/17-AOS1652>
- Bevilacqua, M., Gaetan, C., Mateu, J., y Porcu, E. (2012). 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. Descargado de <https://amstat.tandfonline.com/doi/abs/10.1080/01621459.2011.646928#.XHLPCtnPyV5> doi: <https://doi.org/10.1080/01621459.2011.646928>
- Bevilacqua, M., y Morales-Oñate, V. (2018). Geomodels: A package for geostatistical gaussian and non gaussian data analysis [Manual de software informático]. Descargado de <https://vmoprojs.github.io/GeoModels-page/> (R package version 1.0.3-4)
- Caamaño, C. (2018). *On modelling and estimation of some non gaussian random fields* (Tesis Doctoral no publicada). Universidad de Valparaíso, Valparaíso.
- Cressie, N. A. C. (1993). *Statistics for spatial data revised edition*. Nueva York, Estados Unidos: Wiley.
- De Oliveira, V. (2006). On optimal point and block prediction in log-gaussian random fields. *Scandinavian Journal of Statistics*, 33(3), 523–540. Descargado de <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1467-9469.2006.00494.x> doi: <https://doi.org/10.1111/j.1467-9469.2006.00494.x>
- Diggle, P. J., y Ribeiro, P. J. (2007). *Model-based geostatistics*. Nueva York, Estados Unidos: Springer.
- Dobrow, R. (2014). *Probability with applications and r*. Nueva Jersey, Estados Unidos: John Wiley and Sons.
- Fernandez, C., y Steel, M. F. J. (1998). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441), 359–371. Descargado de <https://www.tandfonline.com/doi/abs/10.1080/01621459.1998.10474117> doi: <https://doi.org/10.1080/01621459.1998.10474117>
- Garcia-Papani, F., Leiva, V., Uribe-Opazo, M. A., y Aykroyd, R. G. (2018). Birnbaum-saunders spatial regression models: Diagnostics and application to chemical data. *Chemometrics and Intelligent Laboratory Systems*, 177, 114 - 128. Descargado de <http://www.sciencedirect.com/science/article/pii/S0169743918300686> doi: <https://doi.org/10.1016/j.chemolab.2018.03.012>
- Garnier, S., Ross, N., Rudis, B., Sciaini, M., y Scherer, C. (2018). viridis: Default color maps from ‘matplotlib’ [Manual de software informático]. Descargado de <https://cran.r-project.org/web/packages/viridis/index.html> (R package version 0.5.1)
- Gelfand, A. E., y Schliep, E. M. (2016). Spatial statistics and gaussian processes: A beautiful marriage. *Spatial Statistics*, 18, 86 - 104. Descargado de <http://www.sciencedirect.com/science/article/pii/S2211675316300033> doi: <https://doi.org/10.1016/j.spasta.2016.03.006>
- Genton, M., y Zhang, H. (2012). Identifiability problems in some non-gaussian spatial random fields. *Chilean Journal of Statistics*, 3(2), 171–179. Descargado de <http://chjs.mat.utfsm.cl/issues.html#vol3-number2-2012>

- Goerg, G. M. (2011). Lambert w random variables—a new family of generalized skewed distributions with applications to risk estimation. *The Annals of Applied Statistics*, 5(3), 2197-2230. Descargado de <https://projecteuclid.org/euclid.aos/1318514301> doi: <https://doi.org/10.1214/11-AOAS457>
- Goerg, G. M. (2015). The lambert way to gaussianize heavy-tailed data with the inverse of tukey's h transformation as a special case. *The Scientific World Journal*, 2015, 1–16. Descargado de <https://www.hindawi.com/journals/tswj/2015/909231/> doi: <https://doi.org/10.1155/2015/909231>
- Gräler, B., Pebesma, E., y Heuvelink, G. (2016). Spatio-temporal interpolation using gstat. *The R Journal*, 8, 204-218. Descargado de <https://journal.r-project.org/archive/2016/RJ-2016-014/index.html>
- Han, Z., y De Oliveira, V. (2016). On the correlation structure of gaussian copula models for geostatistical count data. *Australian and New Zealand Journal of Statistics*, 58(1), 47-69. Descargado de <https://onlinelibrary.wiley.com/doi/abs/10.1111/anzs.12140> doi: <https://doi.org/10.1111/anzs.12140>
- Heagerty, P. J., y Lele, S. R. (1998). A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association*, 93(443), 1099-1111. Descargado de <https://www.tandfonline.com/doi/abs/10.1080/01621459.1998.10473771> doi: <https://doi.org/10.1080/01621459.1998.10473771>
- Kridge, D. G. (1951). A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the Chemical Metallurgical and Mining Society of South Africa*(6), 119-139. Descargado de [https://journals.co.za/content/saimm/52/6/AJA0038223X\\_4792](https://journals.co.za/content/saimm/52/6/AJA0038223X_4792)
- Ma, C. (2010a).  $\chi^2$  random fields in space and time. *IEEE Transactions on Signal Processing*, 58(1), 378-383. Descargado de <https://ieeexplore.ieee.org/abstract/document/5204175/citations#citations> doi: <https://doi.org/10.1109/TSP.2009.2029722>
- Ma, C. (2010b). Elliptically contoured random fields in space and time. *Journal of Physics A: Mathematical and Theoretical*, 43(16). Descargado de <https://iopscience.iop.org/article/10.1088/1751-8113/43/16/165209/meta> doi: <https://doi.org/10.1088/1751-8113/43/16/165209>
- Maleki, M., Contreras-Reyes, J. E., y Mahmoudi, M. R. (2019). Robust mixture modeling based on two-piece scale mixtures of normal family. *Axioms*, 8(2). Descargado de <https://www.mdpi.com/2075-1680/8/2/38> doi: <https://doi.org/10.3390/axioms8020038>
- Maleki, M., y Mahmoudi, M. R. (2017). Two-piece location-scale distributions based on scale mixtures of normal family. *Communications in Statistics - Theory and Methods*, 46(24), 12356-12369. Descargado de <https://www.tandfonline.com/doi/full/10.1080/03610926.2017.1295160?scroll=top&needAccess=true> doi: <https://doi.org/10.1080/03610926.2017.1295160>
- Mardia, K. V., y Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71(1), 135-146. Descargado de <http://oup.prod.sis.lan/biomet/article-pdf/71/1/135/1001019/71-1-135.pdf> doi: <https://doi.org/10.1093/biomet/71.1.135>
- Matheron, G. (1960). Principles of geostatistics. *Economic Geology*, 1246-1266. Descargado de <https://pubs.geoscienceworld.org/segweb/economicgeology/article-abstract/58/8/1246/17275/Principles-of-geostatistics?redirectedFrom=fulltext> doi: <https://doi.org/10.2113/gsecongeo.58.8.1246>
- Moravveji, B., Khodadadi, Z., y Maleki, M. (2019). A bayesian analysis of two-piece distributions based on the scale mixtures of normal family. *Iranian Journal of Science and Technology, Transactions A: Science*, 43(3), 991–1001. Descargado de <https://link.springer.com/>

- [article/10.1007/s40995-018-0541-9](https://doi.org/10.1007/s40995-018-0541-9) doi: <https://doi.org/10.1007/s40995-018-0541-9>
- Palma, W. (2016). *Time series analysis*. Nueva Jersey, Estados Unidos: John Wiley and Sons.
- Porcu, E., Montero, J., y Schlather, M. (2012). *Advances and challenges in space-time modelling of natural events*. Berlín, Alemania: Springer.
- R Core Team. (2017). R: A language and environment for statistical computing [Manual de software informático]. Vienna, Austria. Descargado de <http://www.R-project.org/>
- Rao, B. (2011). *Statistical inference for fractional diffusion processes*. Nueva Jersey, Estados Unidos: John Wiley and Sons.
- Sang, H., y Genton, M. G. (2014). Tapered composite likelihood for spatial max-stable models. *Spatial Statistics*, 8, 86 - 103. Descargado de <http://www.sciencedirect.com/science/article/pii/S2211675313000420> doi: <https://doi.org/10.1016/j.spasta.2013.07.003>
- Schmidt, J. P., Cruse-Sanders, J., Chamberlain, J. L., Ferreira, S., y Young, J. A. (2019). Explaining harvests of wild-harvested herbaceous plants: American ginseng as a case study. *Biological Conservation*, 231, 139 - 149. Descargado de <http://www.sciencedirect.com/science/article/pii/S0006320718313612> doi: <https://doi.org/10.1016/j.biocon.2019.01.006>
- Stein, M. (1999). *Interpolation of spatial data. some theory of kriging*. Nueva York, Estados Unidos: Springer.
- Varin, C., Reid, N., y Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21(1), 5–42. Descargado de <http://www.jstor.org/stable/24309261>
- Wallis, K. F. (2014). The two-piece normal, binormal, or double gaussian distribution: Its origin and rediscoveries. *Statistical Science*, 29(1), 106–112. Descargado de <https://projecteuclid.org/euclid.ss/1399645739> doi: <https://doi.org/10.1214/13-STS417>
- White, P., y Porcu, E. (2019). Towards a complete picture of stationary covariance functions on spheres cross time. *Electron. J. Statist.*, 13(2), 2566–2594. Descargado de <https://projecteuclid.org/euclid.ejs/1564732820> doi: <https://doi.org/10.1214/19-EJS1593>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Descargado de <https://ggplot2.tidyverse.org>
- Xu, G., y Genton, M. G. (2017). Tukey g-and-h random fields. *Journal of the American Statistical Association*, 112(519), 1236-1249. Descargado de <https://www.tandfonline.com/doi/full/10.1080/01621459.2016.1205501> doi: <https://doi.org/10.1080/01621459.2016.1205501>
- Zareifard, H., Khaledi, M. J., Rivaz, F., y Vahidi-Asl, M. Q. (2018). Modeling skewed spatial data using a convolution of gaussian and log-gaussian processes. *Bayesian Analysis*, 13(2), 531–557. Descargado de <https://projecteuclid.org/euclid.ba/1502762659> doi: <https://doi.org/10.1214/17-BA1064>
- Zhang, H., y El-Shaarawi, A. (2010). On spatial skew-gaussian processes and applications. *Environmetrics*, 21(1), 33-47. Descargado de <https://onlinelibrary.wiley.com/doi/abs/10.1002/env.982> doi: <https://doi.org/10.1002/env.982>