

# **MÉTODO DE DOWNSCALING ESTADÍSTICO CON MODELO DE REGRESIÓN PONDERADA ESPACIO-TEMPORAL PARA CAMPOS DE VIENTO EN LA REGIÓN DE VALPARAÍSO**

**Trabajo de Título para optar al título profesional de  
Ingeniero Civil Matemático**

Alonso Nicolás Lagos Orellana

**Profesora Guía: Dra. Karine Bertin**

Universidad de Valparaíso

**Profesor Co-Guía: Dr. Lisandro Fermín**

Universidad de Valparaíso

**Profesora Co-Guía: Dra. Andrea Jiménez**

Universidad de Valparaíso

**Miembros de la comisión:**

**Andrea Jiménez, Karine Bertin y Lisandro Fermín**

Universidad de Valparaíso

Instituto de Ingeniería Matemática  
Valparaíso, 14 de Agosto de 2025.

# Resumen

El creciente aumento en las últimas décadas de fenómenos climáticos extremos, de la tasa de incendios y de demanda energética, han llevado a la necesidad de estudiar variables climáticas en zonas espaciales locales para generar soluciones a estas problemáticas. Este trabajo busca implementar un esquema de downscaling estadístico para campos de intensidad de viento en la región de Valparaíso, el cual integra un modelo de regresión ponderada espacio-temporal (GTWR) para estimar mallas anidadas de modelos Weather Research and Forecasting (WRF) ( $3\text{ km} \rightarrow 1\text{ km} \rightarrow 1/3\text{ km}$ ). El método se apoya en hipótesis locales de regularidad espacial, homogeneidad espacial, autosimilaridad espacial y estacionariedad de los residuos. Para esto, se calibran factores de escala no-paramétricos mediante mínimos cuadrados ponderados (WLS) con un kernel espacio-distribucional. Luego, se consideran los campos de intensidad del viento reales a resolución 3 km y 1 km obtenidos por modelos WRF (73 mediciones temporales por hora entre 02/09/2023 y el 05/09/2023) con dominio espacial en la región de Valparaíso (coordenadas latitud Este-Oeste  $-33,43985$  y  $-32,88078$ , y longitud Norte-Sur  $-71,89282$  y  $-70,93353$ ). Para los datos reales, se clasificaron las zonas topográficas de mar, valle y montaña, por un algoritmo del vecino más lejano (FPC) sobre las distancias distribucionales de la diferencia absoluta de las variaciones. Después, se aplica el modelo GTWR entre campos de 3 km a 1 km para recuperar los factores de escala, obteniendo una excelente capacidad predictiva en las zonas de mar y valle (MSE bajo y coeficientes de determinación altos, para espacio y tiempo). No obstante, el GTWR muestra dificultades para adaptarse a la topografía local en zonas de montaña (error cuadrático medio alto y coeficiente de determinación bajo-intermedio, para espacio y tiempo). Se finaliza el esquema aplicando los factores de escala para obtener el campo de viento objetivo a  $1/3\text{ km}$ , en donde la predicción mejora la resolución espacial a  $1/3\text{ km}$  capturando cambios en la intensidad del viento, desde aumentos locales leves en zonas de mayor homogeneidad (mar y valle), hasta grandes aumento en zonas de mayor variabilidad topográfica (montaña). Se concluye que el modelo GTWR es eficiente en costos computacionales y muestra una alta capacidad predictiva para zonas espaciales en donde la topografía es de variación leve o mediana, y en zonas de alta variabilidad topográfica el modelo sobreajusta el cambio de escala en las series de tiempo. Este comportamiento se deriva posiblemente del aumento de la heterogeneidad espacial al aplicar el downscaling espacial en zonas de montaña. Para mejorar la efectividad del modelo GTWR se recomienda: implementar un modelo por zonas topográficas, calibrar los anchos de ventana por cluster y/o por un método espacio-temporal, y estudiar la incorporación de nuevas covariables al modelo (dirección del viento, temperatura, variable de altura, entre otras).

# Agradecimientos

No es complejo agradecer por el apoyo incondicional de tu núcleo cercano y amistades, lo difícil es retribuirlo con el mismo amor, apoyo, aliento y comprensión indescriptible que recibí en el camino. El amor de mi vida, Javiera, que siempre estuvo a mi lado, con palabras de ánimo, aliento y preocupación en momentos donde todo se daba para bien y cuando las circunstancias eran adversas, ahora te alcanzaré en nuestra etapa de adultos independientes. Mi Padre del cielo y a mi familia: Cecilia, Héctor, Nélon, Bruno, Abdón, Arnaldo, Delia, Matías, Verónica y muchos más, siempre atentos y preocupados, sin si quiera solicitárselos una vez, son mi inspiración y un privilegio tenerlos de familia.

Vienen a mi mente muchos rostros de amistades con los que compartí experiencias alegres y tristes, en comidas, recreos, salidas y deporte. A todos los que pasamos horas de dedicación en lograr desafíos y metas, con éxitos y fracasos, aquellos con los que compartí la experiencia del deporte físico (en el basket UV) y electrónico (en LoL-ICM), en especial a quienes me acogieron en su hogar cuando lo necesité. Les quiero y estimo por igual.

Agradecer a Andrea, Karine y Lisandro, quienes me guiaron en este largo proceso, me animaron y enseñaron muchas características importantes para mi futuro laboral y académico que llevaré siempre conmigo. Además, me voy con la mejor impresión de la amena calidad humana y excelente docencia, la comprensión, el apoyo y aliento del equipo que conforma el IngeMat-UV. Siempre me animaron a sacar adelante mis estudios, sobre todo en momentos de dificultad.

Agradecer al Fondo Nacional de Desarrollo Científico y Tecnológico, 1220071 FONDECYT REGULAR CROSSINGS, DECOMPOSITIONS AND COLORINGS FOR DISCRETE STRUCTURES, por apoyar esta investigación.

# Índice general

<b>Resumen</b>	<b>2</b>
<b>Agradecimientos</b>	<b>3</b>
<b>1. Introducción</b>	<b>6</b>
<b>2. Fundamentos</b>	<b>10</b>
2.1. Procesos espacio-temporales	10
2.2. Clasificación de patrones variacionales del viento	15
2.3. Métodos de downscaling	18
2.3.1. Downscaling Dinámico	19
2.3.2. Downscaling Estadístico	21
2.4. Modelos de regresión ponderados espacio-temporal (GTWR)	22
2.5. Estimación no-paramétrica del factor de escala $\beta_i$	24
2.6. Métricas de evaluación del modelo GTWR	25
2.6.1. Métricas espaciales	25
2.6.2. Métricas temporales	26
2.6.3. Métricas globales	26
<b>3. Metodología de downscaling estadístico</b>	<b>27</b>
3.1. Hipótesis del downscaling estadístico	29
3.1.1. Regularidad y homogeneidad espacial local	29
3.1.2. Autosimilaridad espacial local	29
3.1.3. Estacionariedad local de los residuos	31
3.2. Metodología de Downscaling Estadístico	31
<b>4. Resultados</b>	<b>33</b>
4.1. Caso de estudio	33
4.2. Estudio de simulación	37
4.3. Resultados del esquema de downscaling estadístico	38
4.3.1. Análisis descriptivo WRF 3 y 1 km	38
4.3.2. Modelo GTWR: estimación del campo a 1 km	44
4.3.3. Predicción del campo 1/3 km	58
<b>5. Conclusiones</b>	<b>64</b>

<b>A. Resultados Simulación</b>	<b>66</b>
A.1. $\sigma_{res}^2 = 0,005$ . . . . .	66
A.2. $\sigma_{res}^2 = 0,01$ . . . . .	71
A.3. $\sigma_{res}^2 = 0,03$ . . . . .	76
A.4. $\sigma_{res}^2 = 0,05$ . . . . .	81
A.5. $\sigma_{res}^2 = 0,1$ . . . . .	86
<b>B. Documentación y Respaldo</b>	<b>90</b>

# Capítulo 1

## Introducción

El sistema climático terrestre es una red compleja e interconectada de componentes físicos, químicos y biológicos que interactúan a través de escalas espaciales y temporales. Este sistema incluye la atmósfera, la hidrosfera, la criosfera, la litosfera y la biosfera, cuyas dinámicas conjuntas determinan el estado del clima global y regional. Dentro de este entramado, el viento, una variable meteorológica clave, se origina por diferencias de presión atmosférica inducidas por la radiación solar, la rotación terrestre y la interacción con la superficie terrestre y oceánica. Su comportamiento está profundamente influenciado por procesos tanto globales como locales, lo que lo convierte en un indicador sensible de la variabilidad climática.

Comprender la biosfera ha sido un desafío en materia climática. Esta representa el entramado vital que sostiene la vida en el planeta, y su relevancia radica en la interacción dinámica entre sus componentes: organismos vivos, suelos, cuerpos de agua, atmósfera y energía solar. Cada uno de estos elementos cumple funciones esenciales para el equilibrio climático y ecológico. Por ejemplo, los bosques actúan como sumideros de carbono regulando la concentración de gases de efecto invernadero, los océanos absorben calor y dióxido de carbono moderando la temperatura global. Además, es importante entender cómo se generan los fenómenos meteorológicos y cuáles circunstancias se producen con mayor frecuencia, cómo el avance de la humanidad ha influido en la producción de estos fenómenos y de qué forma podemos prevenir eventos climáticos de gran calibre en el futuro.

El cambio climático se ha consolidado como una de las principales problemáticas globales en las últimas décadas, debido a su creciente impacto en los sistemas naturales y sociales. Entre sus repercusiones más evidentes se encuentran el aumento de la temperatura media global en la superficie terrestre, la elevación del nivel medio del mar, y la intensificación de fenómenos meteorológicos extremos. Estos incluyen ráfagas de viento inusualmente intensas en regiones atípicas, huracanes de alta categoría que se desarrollan en períodos de tiempo cada vez más breves, y un incremento sostenido en la ocurrencia e intensidad de incendios forestales a nivel mundial. Además, la interacción entre el cambio climático y fenómenos naturales como El Niño podría estar modificando su frecuencia e intensidad, lo que añade un nivel adicional de complejidad a la dinámica climática planetaria.

El Fenómeno del Niño, parte del sistema ENSO (El Niño - Oscilación del Sur), corresponde a una fase de calentamiento anómalo de las aguas del Pacífico ecuatorial central y oriental. Este evento oceánico-atmosférico tiene repercusiones globales, alterando patrones de circulación atmosférica y afectando el clima en distintas regiones del mundo. En Chile, su influencia se manifiesta principalmente a través de una modificación en la posición y la intensidad del anticiclón subtropical del Pacífico Sur, ver figura 1.1,

lo que debilita los vientos alisios y altera la surgencia costera. Esta modificación atmosférica impacta directamente en la variable viento, reduciendo su intensidad en la zona costera central y generando una mayor variabilidad en el comportamiento del viento en regiones como Valparaíso. Durante los episodios de El Niño, se observa una tendencia a vientos más cálidos y menos persistentes, con una disminución en los flujos provenientes del sur-suroeste, lo que puede favorecer condiciones más estables en superficie pero también incrementar la ocurrencia de eventos extremos, como incendios forestales, al combinarse con temperaturas elevadas y baja humedad relativa. Desde [8] y [25] se observa que la complejidad topográfica y las condiciones oceánicas de Chile central, en particular la región de Valparaíso, influyen fuertemente en la respuesta regional del viento, lo que demanda modelos de alta resolución para capturar los patrones locales con precisión.

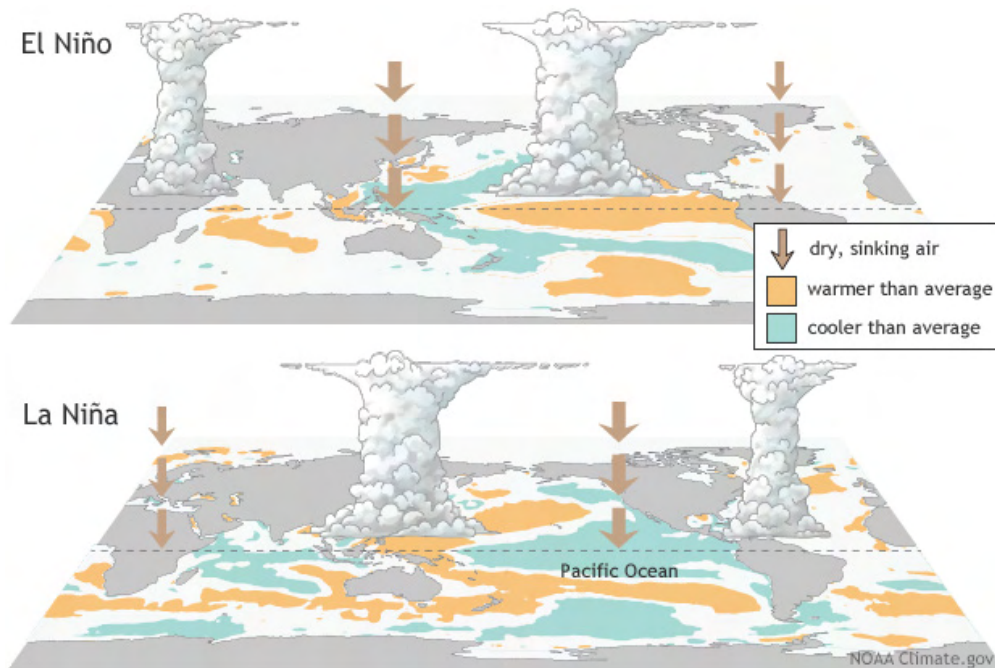


Figura 1.1: Descripción de los fenómenos climáticos asociados al ENSO. Elaborada por [18].

En las últimas décadas, se han desarrollado tecnologías avanzadas gracias a significativos esfuerzos interdisciplinarios orientados a la observación, modelación y comprensión del sistema climático global. Herramientas como satélites meteorológicos, redes de observación terrestre y modelos climáticos de circulación general y regional han permitido notables avances en la predicción y análisis del clima tanto a nivel planetario como continental [3, 16]. Sin embargo, a pesar de estos avances, persisten limitaciones sustanciales cuando se intenta representar fenómenos a escalas locales. Factores como la topografía compleja, la cobertura y uso del suelo, así como las interacciones microclimáticas, introducen una alta heterogeneidad espacial que los modelos climáticos globales no logran resolver de forma adecuada. Esto resalta la necesidad de aplicar enfoques de alta resolución espacial, como el downscaling, para mejorar la representación de procesos atmosféricos en contextos geográficos específicos [34].

En este contexto, la comunidad científica y energética se ha visto en la necesidad de innovar en el desarrollo de energías renovables, tanto para satisfacer la creciente demanda energética como para mitigar los impactos ambientales asociados a las fuentes convencionales. Dentro de estas alternativas, destacan la energía fotovoltaica y la energía eólica, las cuales requieren un análisis detallado de la

viabilidad geográfica, topográfica y ambiental para su adecuada implementación. En particular, la energía eólica se ha posicionado como una opción relevante en regiones con alta heterogeneidad espacial, donde las condiciones climáticas, la topografía accidentada o los riesgos de sobrecalentamiento limitan el rendimiento de plantas fotovoltaicas. A diferencia de estas últimas, las granjas eólicas pueden adaptarse mejor a terrenos complejos y a climas más variables [36]. No obstante, para optimizar su instalación, resulta fundamental contar con herramientas que permitan modelar el comportamiento del viento a escalas espaciales y temporales finas. En este contexto, el uso de modelos de downscaling espacial cobra especial relevancia, ya que permite proyectar la intensidad del viento en mallas de alta resolución, facilitando así la identificación de sitios potencialmente óptimos para la generación eólica [28].

Este estudio toma como fundamento el contexto multifactorial que ha sido desarrollado a lo largo de esta introducción. La creciente incidencia de fenómenos meteorológicos extremos vinculados al cambio climático, así como la influencia regional del fenómeno El Niño sobre la intensidad del viento en zonas costeras de Chile central, evidencian la necesidad de estudiar esta variable con mayor nivel de detalle espacial y temporal. En particular, la alta heterogeneidad topográfica de la región de Valparaíso dificulta la captura precisa de patrones locales mediante modelos climáticos tradicionales de baja resolución. En este escenario, surge la necesidad de emplear metodologías que permitan una desagregación espacial de información climática, lo que se conoce como downscaling. Esta técnica permite transformar proyecciones de modelos globales o regionales en estimaciones más detalladas, ajustadas a las características geográficas específicas del área de estudio. Aplicar este tipo de herramientas para caracterizar la intensidad del viento no solo representa un avance en la comprensión de su variabilidad a escala local, sino que también contribuye a la identificación de territorios con potencial para el desarrollo de energías renovables, especialmente en zonas donde la energía eólica se perfila como una alternativa estratégica y sustentable para enfrentar los desafíos energéticos y ambientales del país.

En [2] se presenta un estudio detallado de campos de viento en escalas finas, proponiendo una metodología mixta de downscaling geométrico y estadístico que, partiendo de simulaciones WRF a 3 km y 1 km de resolución para la región de Valparaíso con un tiempo de ejecución computacional de un día para un horizonte temporal de un mes, genera con gran eficiencia un campo de viento a 1/2 km. La estrategia se traduce en tres pasos: i) un upscaling determinista que interpola de 1 km a una malla intermedia de 1,5 km; ii) el ajuste de un modelo de regresión lineal local en vecindades de cuatro puntos, fundamentado en hipótesis de regularidad local y autosimilaridad espacial, que establece empíricamente la relación entre los campos de 3 km y 1,5 km; y iii) un downscaling geométrico que aplica dicha relación para obtener el campo a 0,5 km preservando la anidación de las mallas y evitando problemas de colinealidad e identificabilidad propios de métodos de alta dimensionalidad. Al combinar interpolación determinista y modelado empírico, este enfoque reduce significativamente el costo computacional, el WRF tarda hasta un mes en procesar resoluciones de 0,5 km, sin sacrificar la precisión, aportando a la comunidad científica una herramienta práctica y escalable para estudios de viento a pequeña escala. No obstante, la regresión lineal local tiene sus limitaciones como downscaling estadístico para capturar la influencia de la topografía, los micro-climas y el uso de suelo a micro-escala.

En este estudio, se propone un metodología de downscaling estadístico siguiendo un esquema directo en resolución espacial  $3 \text{ km} \rightarrow 1 \text{ km} \rightarrow 1/3 \text{ km}$  por medio de la implementación de un modelo de regresión ponderada espacio-temporal (GTWR). El esquema, a diferencia de [2], es directo en el sentido de la metodología que se emplea y no requiere pasos intermedios para llegar a la resolución objetivo, con  $3 \text{ km} \rightarrow 1 \text{ km}$  se estiman los mejores factores de escala para llegar de  $1 \text{ km} \rightarrow 1/3 \text{ km}$ . El modelo GTWR brinda robustez a fenómenos orográficos latentes, gracias a la incorporación de ponderaciones

determinadas por una distancia espacial-temporal (distribucional) y una función de kernel, en el aumento de resolución espacial a micro-escala. Por lo tanto, permite obtener campos de intensidad de viento a resolución  $1/3$  km, los cuales pueden ser de gran utilidad en los contextos de medio ambiente y energías renovables, tales como: colocación de molinos, simulación y predicción de fenómenos de turbulencia en granjas eólicas, estudio de zonas de alto riesgo de incendio, entre otros.

En el Capítulo 2, contiene los fundamentos y notaciones necesarias para comprender y estudiar los campos de vientos, clasificar series de tiempo del viento e implementar métodos de downscaling dinámicos y estadísticos. En el Capítulo 3, se fundamenta el método de downscaling seleccionado para el problema del estudio del viento, el que consiste en la construcción de un modelo de regresión ponderada espacio-temporal para series de tiempo sobre el esquema de downscaling estadístico a relación de escala  $1 : 3$ . En el Capítulo 4, se divide en: un estudio de simulación del campo a  $1$  km con estructura movimiento Browniano fraccionario simple para estudiar la capacidad de adaptación del modelo GTWR a relaciones localmente lineales, el caso de estudio que justifica el dominio espacial en la región de Valparaíso, una clasificación de las series de tiempo para un análisis descriptivo de los campos de intensidad del viento del caso de estudio por clusters (mar, valle y montaña), se presentan los resultados del modelo GTWR mediante un análisis espacial y temporal, para luego finalizar con la predicción del campo de viento a resolución  $1/3$  km. Por último, el Capítulo 5 consta de las conclusiones del esquema de downscaling y comentarios en torno al impacto del modelo, recomendaciones y desafíos para el futuro.

## Capítulo 2

# Fundamentos

En este capítulo se introducirán las nociones básicas para construir y estudiar los campos de viento sobre una zona geográfica, el esquema de downscaling estadístico con énfasis en modelos lineales, ampliamente utilizados en la investigación climática moderna, para aumentar la resolución espacial de la variable climática en un dominio espacial fijo, y mediante conceptos para clasificar las zonas geográficas en base a las estructuras espaciales y fenómenos climáticos que emergen del aumento de la resolución espacial.

### 2.1. Procesos espacio-temporales

A continuación, se presentan los conceptos teóricos fundamentales para el estudio de campos espacio-temporales, incluyendo procesos estocásticos, estructuras de dependencia y propiedades estadísticas. Estos son necesarios para explicar factores meteorológicos del sistema climático, permitiendo analizar sus características principales que representan propiedades asociadas al tiempo y espacio en que se estudia la variable climática. Estas características suelen ser: dependencia temporal, dependencia espacial, colinealidad, entre otros.

Consideraremos los siguientes objetos básicos de la teoría de la medida y probabilidad: Sea  $(\Omega, \mathcal{F}, P)$  un espacio de probabilidad, donde  $\Omega$  es el espacio de estados,  $\mathcal{F}$  una  $\sigma$ -álgebra del espacio  $\Omega$  y  $P$  una medida de probabilidad. Sea  $\mathcal{B}(\mathbb{R}^n)$  el conjunto de los borelianos de  $\mathbb{R}^n$ . Los procesos estocásticos comúnmente se utilizan para representar la interacción de una variable dentro de un sistema en el tiempo. Uno de estos es el sistema climático, en donde sus variables climáticas son representadas por procesos estocásticos indexados en el tiempo. Estos procesos están completamente definidos por sus familia de funciones de distribución finito dimensional, que permiten caracterizar su comportamiento, comparar procesos entre si y poder aplicar modelos sobre estos. A continuación, solo vamos a considerar procesos estocásticos a valores en  $\mathbb{R}$ .

**Definición 1** Sea  $(\Omega, \mathcal{F}, \mathbb{P})$  un espacio de probabilidad, y sea  $\mathcal{T}$  un conjunto no vacío llamado conjunto de índices (por ejemplo, el tiempo o el espacio). Sea  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  el espacio de estados equipado con su  $\sigma$ -álgebra de Borel. Una **variable aleatoria** es una función medible  $X : \Omega \rightarrow \mathbb{R}$ , es decir, tal que  $X^{-1}(B) \in \mathcal{F}$  para todo  $B \in \mathcal{B}(\mathbb{R})$ . Un **proceso estocástico** es una colección de variables aleatorias  $X = \{X_t : t \in \mathcal{T}\}$  definidas en  $(\Omega, \mathcal{F}, \mathbb{P})$  y con valores en  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . Es decir, para cada  $t \in \mathcal{T}$ , la función  $X_t : \Omega \rightarrow \mathbb{R}$  es una variable aleatoria.

**Definición 2** Sea un proceso estocástico  $X = (X_t, t \in \mathcal{T})$  y  $t_1, t_2, \dots, t_n \in \mathcal{T}$ . La función de distribución finito dimensional del proceso se define como:

$$\mathbf{F}_{t_1, \dots, t_n}(x_1, \dots, x_n) := \mathbf{P}(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n), \quad x_1, \dots, x_n \in \mathbb{R}.$$

**Definición 3** El valor esperado de la variable aleatoria  $X_t$ , con función de distribución  $\mathbf{F}_t(\cdot)$ , se define como  $E[X_t] := \int_{-\infty}^{\infty} x \mathbf{F}_t(dx)$ , y el segundo momento de  $X_t$  es  $E[X_t^2] := \int_{-\infty}^{\infty} x^2 \mathbf{F}_t(dx)$ . Llamamos a un proceso  $(X_t, t \in \mathcal{T})$  de **segundo orden** cuando  $E[X_t^2] < \infty$  para todo  $t \in \mathcal{T}$ .

**Definición 4** Dado el proceso  $X = (X_t, t \in \mathcal{T})$ . Se define su función de media, varianza y covarianza para  $t, t' \in \mathcal{T}$ , respectivamente:

- $\mu_t = E[X_t]$
- $\sigma_t^2 = \text{Var}(X_t) = E[(X_t - E[X_t])^2]$
- $\gamma(t, t') := \text{Cov}(X_t, X_{t'}) = E[(X_t - E[X_t])^T (X_{t'} - E[X_{t'}])]$

Estudiar las estructuras de dependencia de los procesos que representan variables climatológicas, es necesario para poder entender sus características a través del tiempo y/o distintas zonas espaciales. Dentro de unos de los supuestos principales para aplicar modelos sobre datos que representen comportamientos lineales, es necesario saber si son estacionarios. La estacionariedad, por medio de las características estadísticas del proceso, nos permite entender si estos varían en el tiempo en sus funciones de media, varianza y covarianza.

**Definición 5** Un proceso es fuertemente estacionario si la distribución finito dimensional es invariante ante traslaciones, es decir,  $\forall t_1, \dots, t_n \in \mathcal{T}$  y  $\forall x_1, \dots, x_n \in \mathbb{R}$ :

$$\mathbf{F}_{t_1, \dots, t_n}(x_1, \dots, x_n) = \mathbf{F}_{t_1+h, \dots, t_n+h}(x_1, \dots, x_n), \quad \forall h \in \mathcal{T}.$$

**Definición 6** Un proceso  $X = (X_t, t \in \mathcal{T})$  es débilmente estacionario si su función de covarianza es invariante ante traslaciones, es decir,  $\gamma(t+h, t'+h) = \gamma(t, t')$ , para todo  $h \in \mathcal{T}$ .

Desde una proceso estocástico  $X = (X_t, t \in \mathcal{T})$ , con  $\mathcal{T} = \mathbb{Z}$ , podemos extender la definición a un proceso espacio-temporal. Un proceso espacio-temporal es un proceso estocástico indexado en  $\mathcal{T} = \mathbb{R}^n \times \mathbb{Z}$ , y lo denotaremos de la forma  $X = \{X_{s,t} : s \in \mathcal{D} \subset \mathbb{R}^n, t \in \mathbb{Z}\}$ , con  $\mathcal{D}$  el dominio espacial del proceso, para  $n = 2$ , y  $s \in \mathcal{D}$  son los puntos espaciales del dominio.

Para comparar dos procesos espacio-temporales, estudiaremos las siguientes propiedades en los dominios espaciales y relaciones estadísticas entre las series del proceso. Primero, estudiar estructuras de dependencia espacial entre los dominios espaciales (si existen puntos espaciales que coinciden en los dominios) e identificar relaciones de escala espacial ( $\lambda > 0$ ) entre estos dominios. Segundo, relacionar estadísticamente las series temporales de los dominios espaciales, comparando funciones de distribución finito-dimensional, función de media y función de covarianza. Esto se resume en las siguientes definiciones que se estudiarán en el siguiente capítulo: Denotamos  $\stackrel{d}{=}$  la igualdad en términos de la función de distribución finito-dimensional entre dos procesos.

**Definición 7 (Autosimilaridad espacial)** Sea  $X = \{X_{s,t} : s \in \mathcal{D}, t \in \mathbb{Z}\}$  un proceso espacio-temporal definido sobre dominio espacial  $\mathcal{D} \subset \mathbb{R}^2$  y discreto en el tiempo. Se dice que el proceso presenta **autosimilaridad espacial** [19] si existe  $H > 0$ , llamado parámetro de Hurst, tal que, para todo  $\lambda > 0$  y un tiempo fijo  $t \in \mathbb{Z}$ , se cumple la siguiente relación:

$$\{X_{\lambda s,t}\}_{s \in \mathcal{D}} \stackrel{d}{=} \{\lambda^H X_{s,t}\}_{s \in \mathcal{D}}.$$

**Definición 8 (Regularidad espacial)** Sea  $X = \{X_{s,t} : s \in \mathcal{D}, t \in \mathbb{Z}\}$  un proceso espacio-temporal indexado sobre una malla espacial  $\mathcal{D} \subset \mathbb{R}^2$ . Se dice que el proceso presenta **regularidad espacial** si la función  $\gamma_{s,t}$  es continua para todo  $s, t \in \mathbb{Z}$ , donde  $\gamma_{s,t}(h) := \text{Cov}(X_{s,t}, X_{s+h,t})$  con  $h \in \mathcal{D}$ , y además se cumple:

$$\lim_{\|h\| \rightarrow 0} \gamma_{s,t}(h) = \text{Var}(X_{s,t}) < \infty,$$

donde  $\|\cdot\|$  es la norma euclidiana de  $\mathbb{R}^n$ .

**Definición 9 (Homogeneidad espacial)** Sea  $X = \{X_{s,t} : s \in \mathcal{D}, t \in \mathbb{Z}\}$  un proceso espacio-temporal. Se dice que el proceso es **espacialmente homogéneo** si su distribución finito-dimensional es invariante bajo traslaciones espaciales, es decir, para todo desplazamiento entre puntos espaciales  $h \in \mathcal{D}$  se cumple:

$$\{X_{s,t}\}_{s \in \mathcal{D}} \stackrel{d}{=} \{X_{s+h,t}\}_{s \in \mathcal{D}}, \quad \forall t \in \mathbb{Z}.$$

La autosimilaridad espacial, se traduce en que el proceso sigue un patrón de variabilidad espacial determinado por el exponente  $H$  al hacer un aumento de resolución. La regularidad espacial explica la presencia de formas suaves (continuas) en el dominio espacial regular, lo que explica la calidad de un modelo de downscaling estadístico en reconstruir detalles a una escala más fina. Luego, la homogeneidad espacial permite establecer un único modelo estadístico en el dominio espacial, mediante alguna caracterización de la variabilidad sistemática de este.

**Ejemplo 1** Una de las distribuciones más conocidas y utilizadas para simular procesos es la distribución Gaussiana. Esta satisface propiedades sobre su función de media y función de autocovarianza, lo que permite tener control sobre el entorno de simulación y estructuras de dependencia temporal. Sea una variable aleatoria  $X$  con distribución Gaussiana,  $X \sim \mathcal{N}(\mu, \sigma^2)$ , es decir, su función de densidad es:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Llamaremos al proceso estocástico  $X = (X_t, t \in \mathbb{Z})$  Gaussiano multivariado. Se define un proceso espacio-temporal Gaussiano  $X = \{X_{s,t} : s \in \mathcal{D}, t \in \mathbb{Z}\}$ , si  $X_{s,t}$  posee una función de distribución Gaussiana para todo  $s \in \mathcal{D} \in \mathbb{R}^d$  y  $t \in \mathbb{Z}$ . Llamaremos solo proceso Gaussiano al proceso anterior. Sea un proceso Gaussiano  $X = \{X_{s,t} : s \in \mathcal{D}, t \in \mathbb{Z}\}$ . El proceso Gaussiano  $X$  es débilmente estacionario si y solamente si es fuertemente estacionario. Esto se deduce directamente de las funciones de media, varianza y covarianza sobre distribución de un proceso Gaussiano.

En la introducción del problema, se destaca la necesidad de desarrollar e implementar modelos a resoluciones de escala suficientemente pequeñas. Para esto, introduciremos las notaciones esenciales para entender los objetos de campos de intensidad del viento, mallas espaciales y resolución espacial. Denotamos por  $T$  al número de observaciones temporales,  $b \in \{1/3, 1, 3\}$  al nivel de resolución espacial en kilómetros y  $\lambda = 1/3$  al factor de escala espacial. Para representar espacialmente un proceso espacio-temporal, consideramos una malla espacial  $M^{(b)}$  regular (puntos equidistantes en dirección horizontal y vertical) y de resolución espacial  $b$ , caracterizada por un conjunto numerable de puntos espaciales  $s_1, \dots, s_{n_b} \in \mathcal{D}$ , con  $\mathcal{D} \subset \mathbb{R}^2$  un dominio espacial compacto y convexo. Denotamos la malla espacial regular a resolución espacial  $b$ .

$$M^{(b)} := \{s_i \in \mathcal{D} \subseteq \mathbb{R}^2 : i = 1, \dots, n_b\},$$

donde  $n_b$  denota la cantidad de puntos que contiene la malla espacial. Sobre esta, cada punto espacial representa una medición de una variable atmosférica, en particular para este estudio se mide la intensidad del viento, así cada punto espacial de la malla espacial tendrá asociado la medición de la intensidad del viento en la zona geográfica respectiva. El campo de intensidad de viento observado en la malla espacial  $M^{(b)}$  se denota

$$X^{(b)} = \{x_{s_i,t}^{(b)} : i = 1, \dots, n_b \text{ y } t = 1, \dots, T\} \in \mathbb{R}^{n_b \times T}.$$

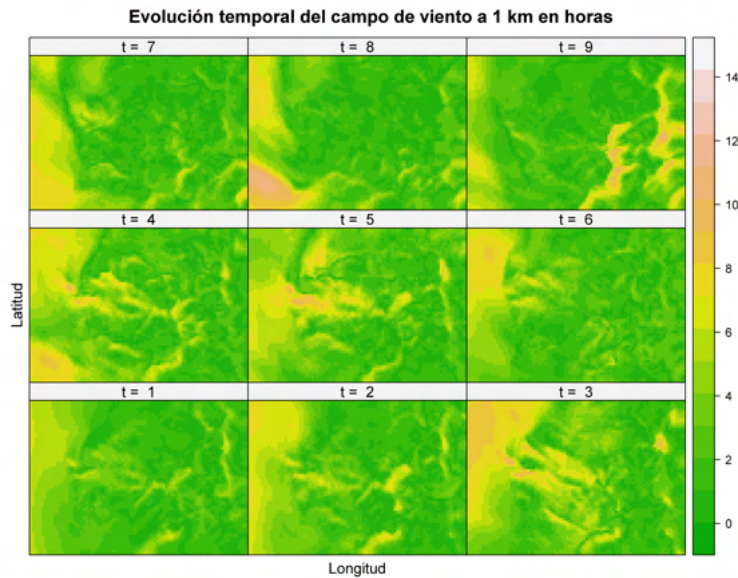


Figura 2.1: Evolución temporal del campo de intensidad de viento en  $m/s$   $X^{(1)}$  y el tiempo en horas. Se observan las primeras 9 horas de medición,  $t = 1$  hasta  $t = 9$ .

Los campos de intensidad del viento, cómo se ven en las figuras 2.1 y 2.2, se componen de observaciones espacio-temporales. Fijando un punto espacial  $s_i \in M^{(b)}$  en el campo de viento e indexando en el tiempo, observamos series de tiempo para cualquier  $i$

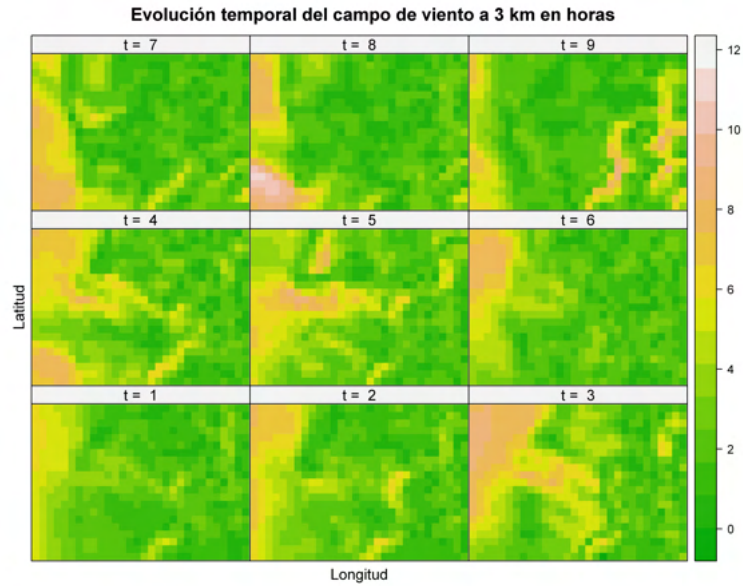


Figura 2.2: Evolución temporal del campo de intensidad de viento en  $m/s$   $X^{(3)}$  y el tiempo en horas. Se observan las primeras 9 horas de medición,  $t = 1$  hasta  $t = 9$ .

$$X_i^{(b)} := \{x_{s_i,t}^{(b)} : t = 1, \dots, T\}, \quad \text{para } i = 1, \dots, n_b,$$

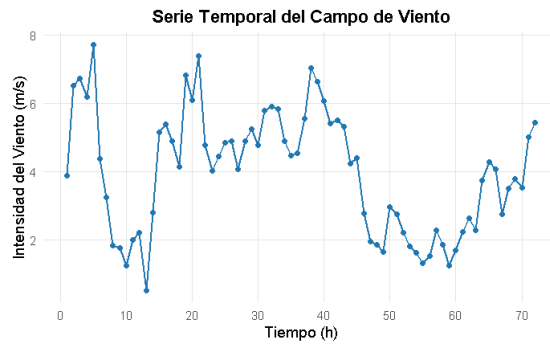


Figura 2.3: Serie de tiempo  $X_i^{(b)}$  del campo de viento en  $m/s$  sobre el punto espacial  $s_i$ .

y si queremos observar el campo de intensidad del viento en una unidad temporal fija  $t \in \mathbb{Z}$  lo denotaremos como:

$$X_{.,t}^{(b)} := \{x_{s_i,t}^{(b)} : i = 1, \dots, n_b\}, \quad \text{para } t = 1, \dots, T.$$

## 2.2. Clasificación de patrones variacionales del viento

Sobre la malla espacial se observarán fenómenos micro-climáticos con mayor relevancia a medida que aumentamos la resolución espacial y además, se evidencia la influencia de la topografía sobre al terreno y la orografía del punto espacial respectivo, siendo este una variable relevante para distinguir zonas urbanas, rurales, forestales, cuerpos de aguas, pies de montaña, entre otros. La intensidad del viento se observa a una altura de aproximadamente 10 metros sobre el nivel del mar, mostrando un comportamiento más homogéneo en distinta regiones, lo que dificulta delimitar con precisión la topografía de las zonas geográficas. La orografía y los fenómenos determinados por micro-climas suele ocurrir en zonas de cambios de variables meteorológicas y altitud con respecto al nivel del mar, cómo la transición entre una ladera y una montaña, mar y límite costero, entre otros.

Una manera de identificar las zonas de mayor variación es primero caracterizando la variación de la intensidad del viento centrado las vecindades con un promedio de sus vecinos más cercanos de primer orden y calculando su desviación estándar. A este procedimiento le llamaremos incrementos espaciales y se caracterizan por su capacidad de identificar zonas de alta variabilidad, representando fenómenos micro-climáticos y asociados a la topografía del terreno. Sea  $s_i \in \mathcal{D}$ . Llamamos a  $V_i^{(b)} \subset M^{(b)}$  la vecindad del punto espacial  $s_i$ , con  $V_i^{(b)} = \{s_{1,1}, s_{1,2}, s_{1,3}, s_{2,1}, s_{2,2}, s_{2,3}, s_{3,1}, s_{3,2}, s_{3,3}\}$  un conjunto finito de puntos de  $M^{(b)}$ , talque para  $s_{j,k}$  se tiene

$$\begin{cases} |s_{j+1,k} - s_{j,k}| = b, \\ |s_{j,k+1} - s_{j,k}| = b, \\ s_{2,2} = s_i, \end{cases}$$

ver figura 2.4.

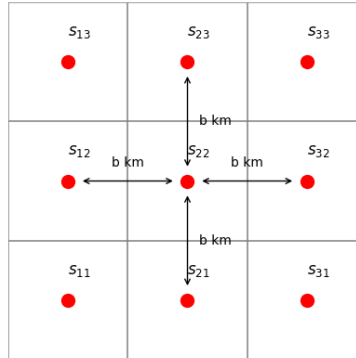


Figura 2.4: Vecindad  $V_i^{(b)} \subset M^{(b)}$  y visualización de la regularidad de la distancia espacial entre los puntos de la malla  $M^{(b)}$ .

Sea la vecindad  $V_i^{(b)}$ , con  $s_i$  el punto central de la vecindad. Llamamos a  $\tilde{\mu}_i$  el promedio de las series de tiempo de los puntos espaciales  $s_{j,k} \in V_i^{(b)}$ , talque  $s_{j,k} \neq s_i$ . Sea  $|V_i^{(b)}| = N$  la cantidad de puntos espaciales de la vecindad  $V_i^{(b)}$ . El promedio de los vecinos espaciales de  $s_i$  es

$$\tilde{\mu}_i := \left\{ \tilde{\mu}_{i,t} = \frac{1}{N-1} \sum_{\substack{s_{j,k} \in V_i^{(b)} \\ s_{j,k} \neq s_i}} x_{s_{j,k},t}^{(b)} : t = 1, \dots, T \right\}.$$

Se denotan las series de tiempo de los incrementos espaciales asociada al punto  $s_i$  como:

$$\tilde{X}_i^{(b)} := \left\{ x_{i,t}^{(b)} - \tilde{\mu}_{i,t} : t = 1, \dots, T \right\}.$$

En particular, las series de tiempo poseen estructuras de dependencia temporal que generan artefactos en modelos si no las tratamos. Al tomar un conjunto de mediciones temporales en horas y un total de 72 horas, existe estacionalidad en la intensidad del viento generalmente entre el día y la noche, solo en ciertas zonas espaciales. Una forma de tratar con estos efectos es generar ponderaciones entre las series para promediar el efecto local en todas las series de la malla espacial. Los incrementos espaciales, al estar centrados localmente, eliminan la relevancia de los posibles efectos locales en las vecindades locales.

En el contexto de los campos de la intensidad del viento la topografía está relacionada con patrones distribucionales en el dominio espacial. En particular, los incrementos espaciales  $\tilde{X}^{(b)}$  permiten analizar transiciones en la malla espacial con mayor precisión, ver figura 2.5. Recientemente, el agrupamiento jerárquico divisivo ha sido empleado en meteorología para identificar regiones con comportamientos similares [2, 9] para poder comparar zonas geográficas con distinta topografía efectivamente. Este método es particularmente útil en problemas espaciales dónde se busca identificar valores en zonas geográficas a una escala de resolución muy pequeña, ya que divide iterativamente los datos en clusters basados en medidas de disimilaridad.

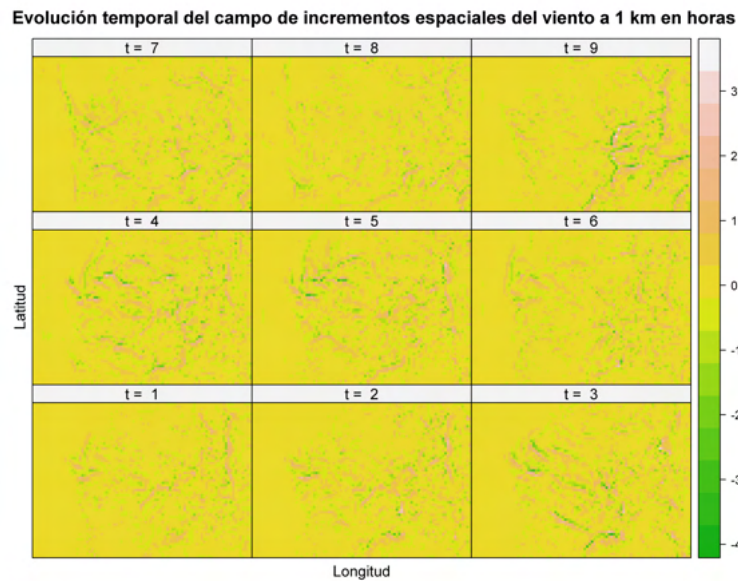


Figura 2.5: Evolución temporal del campo de incrementos espaciales de la intensidad de viento en  $m/s$   $\tilde{X}^{(1)}$ . Se observan las primeras 9 horas de medición,  $t = 1$  hasta  $t = 9$ .

Siguiendo el marco abordado en [33], podemos definir las siguiente distancias asociada a las funciones de distribuciones finito dimensional entre las series de tiempo espacio-temporal asociadas a los puntos

espaciales  $s$  y  $s'$ , respectivamente, con la siguiente definición:

**Definición 10** Sea  $\sigma_i := \sqrt{\text{Var}(X_i^{(b)})}$ , la desviación estándar de la serie de tiempo asociada al punto espacial  $s_i$ . Se define la distancia distribucional entre las series de tiempo en los puntos espaciales  $s_i$  y  $s_j$ :

$$d(i, j) := |\sigma_i - \sigma_j|, \quad (2.1)$$

para todo  $i, j = 1, \dots, n_b$ .

Usaremos esta distancia distribucional para clasificar zonas de variación topográfica, utilizando los incrementos espaciales en la malla  $M^{(1)}$ . Medir la variación temporal entre las series de tiempo permite diferenciar dónde se producen cambios abruptos y significativos de la intensidad del viento. Particularmente, fenómenos climáticos como los de turbulencia, están principalmente determinados por la variación del viento. Con esta distancia se puede estudiar la rugosidad de la malla espacial e identificar patrones topográficos. Así, al momento de identificar las zonas que caracterizan los tipos de variaciones, se pueden identificar representantes de máxima variación. Estos representantes definen un comportamiento en común variacional de las series de tiempo y se denominan clusters.

En nuestro contexto de clasificación, no poseemos etiquetas existentes sobre nuestras series de tiempo, lo que impide poder utilizar métodos de clasificación supervisada como el K-means. Por ende, debemos clasificar las distancias distribucionales por las variaciones de la intensidad del viento de las series de tiempo, para poder representar de manera eficiente las medidas de disimilaridad entre cada serie de tiempo del dominio espacial.

Dado un conjunto discreto de observaciones  $O = \{o_1, \dots, o_N\}$ , estas se agrupan en  $m$  clases disjuntas entre ellas y juntas completan el conjunto de observaciones, es decir, dadas las clases  $\{C_1, \dots, C_m\}$ , con  $1 \leq m \leq N$ , se cumple que

1.  $C_i \cap C_j = \emptyset, \quad \forall i \neq j$
2.  $\bigcup_{i=1}^m C_i = O$ .

El algoritmo de clasificación del vecino más lejano (FPC, por sus siglas en inglés) plantea en términos generales que para cada clase  $C_k$  se escoge un elemento representante  $c_k$  y los elementos restantes se asignan a la clase del representante más cercano según la medida de disimilaridad (distancia distribucional) de estos. El siguiente algoritmo muestra el método de clasificación FPC:

---

**Algoritmo 1:** Clasificación con el método del vecino más cercano (FPC)

---

**Entrada:** Sea el conjunto  $O = \{o_1, \dots, o_N\}$ . Sea  $m > 1$  el número de clusters objetivo.

**Salida:** Conjunto  $\{C_1, \dots, C_m\}$  de  $m$  clusters.

1.  $c_1 \leftarrow o_1$

2. **Para**  $i \leftarrow 2$  **hasta**  $m$  **hacer**

2.1 Calcular:

$$j^* = \operatorname{argmax}_{j=1, \dots, N} \left\{ \min_{k=1, \dots, i-1} d(o_j, c_k) \right\}$$

2.2  $c_i := o_{j^*}$

3.  $C_k := \{c_k\}$  para todo  $k = 1, \dots, m$

4. **Para**  $j \leftarrow 1$  **hasta**  $N$  **hacer**

4.1 Calcular:

$$k^* = \operatorname{argmin}_{k=1, \dots, m} \{d(o_j, c_k)\}$$

4.2  $C_{k^*} = C_{k^*} \cup \{o_j\}$

5. **Retorna**  $\{C_1, \dots, C_m\}$

---

Este tipo de clasificación nos permite identificar zonas representativas de la topografía del terreno, esto mediante las medidas de disimilaridad, o distancia distribucional. Las medidas de disimilaridad están determinadas por el máximo de las diferencias absolutas entre las variaciones de las series de tiempo, lo que nos permitirá identificar patrones de transición entre distintos tipos de uso de suelo, tales como litorales, quebradas, pies de montaña y montañas con mayor precisión, pues a medida que aumentamos la resolución de nuestras mallas espaciales, debe aumentar la variabilidad intra-cluster dada la orografía de la zona.

### 2.3. Métodos de downscaling

Estudiar los cambios del sistema climático en distintas escalas ha sido posible gracias a los avances e innovación en el desarrollo de técnicas y modelos sofisticados para su modelación, mediante el uso de herramientas como satélites, sensores y aparatos de medición, entre otros. Estas técnicas reúnen el esfuerzo de distintas disciplinas para poder estimar el comportamiento de la dinámica climática en toda la biosfera, de manera que seamos capaces de medir procesos físicos tales como: la temperatura global de la corteza terrestre, identificar cambios en las corrientes de convección del viento cómo los vientos alisios del Sur-Este y vientos árticos, junto a la circulación de Walker, ver figura 2.6, entre otros. La dinámica del clima se puede representar por principios de conservación de masa, momentum y energía, mediante un conjunto de ecuaciones diferenciales parciales con condiciones de borde (escenarios climáticos), que definen la base del estudio de la dinámica del clima [20, 6]. Existen tres tipos de escalas para modelar la circulación atmosférica, que se aceptan a nivel internacional, la escala global o macro-escala que tiene generalmente resoluciones entre 50 a 300 km, pues busca modelar toda la superficie del planeta; la meso-escala va de 1 a 50 km, generalmente se centra en regiones limitadas y con comportamientos inestables; y la micro-escala que tiene resolución menor a 1 km, que busca representar interacciones topográficas y el efecto de los micro-climas.

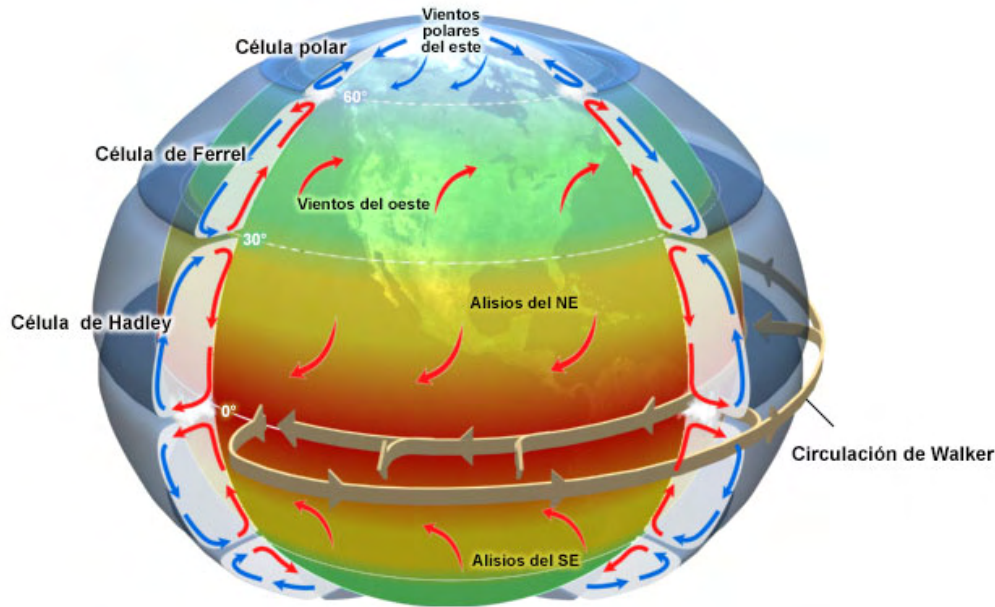


Figura 2.6: Dinámica de circulación y convección del viento.

Los Modelos de Circulación Global (GCM) constituyen la piedra angular de la modelación climática moderna, ya que simulan el sistema climático terrestre mediante ecuaciones físico-matemáticas que representan procesos atmosféricos, oceánicos, terrestres y de hielo marino a escala global, precisamente ecuaciones diferenciales parciales del tipo Navier-Stokes con condiciones de borde determinadas por funciones simples [20] que se resuelven con métodos de integración numérica. Estos modelos se desarrollan a partir de principios de dinámica de fluidos geofísicos tridimensionales, integrando interacciones complejas como la radiación, la convección, la turbulencia, la formación de nubes y los intercambios de energía entre la superficie terrestre y la atmósfera. Aunque su resolución espacial es típicamente del orden de 100 a 250 km, lo que limita su capacidad para representar fenómenos locales, su valor radica en proporcionar proyecciones coherentes del clima futuro bajo distintos escenarios de emisiones de gases de efecto invernadero, ver figura 2.7.

Los GCM son una primera instancia para comprender fenómenos del cambio climático a gran escala. Sin embargo, según [5, 31], estos no son una herramienta bien adaptada para captar fenómenos climáticos dado que abarcan una zona geográfica muy amplia y a baja resolución, obteniendo mayor homogeneidad espacial en las estimaciones de variables climáticas que dependen de la topografía del terreno y procesos micro-climáticos. Desde [1], dada la falta de una mirada realista de la topografía asociada a los fenómenos climáticos, se realizaron avances significativos para desarrollar modelos que capten tendencias temporales asociadas a la topografía y micro-climas del terreno a meso-escala y micro-escala. En general a estos modelos se les conocen como downscaling dinámico y downscaling estadístico.

### 2.3.1. Downscaling Dinámico

El downscaling dinámico es un método que preserva la dinámica climática de los modelos GCM pero buscan reducir la zonas de interés a meso-escala para identificar patrones topográficos y micro-climáticos. Los modelos climatológicos regionales (RCM) o modelos anillados derivan de la aplicación del downscaling dinámico sobre los modelos GCM. El modelo anillado proviene de una relación jerárquica

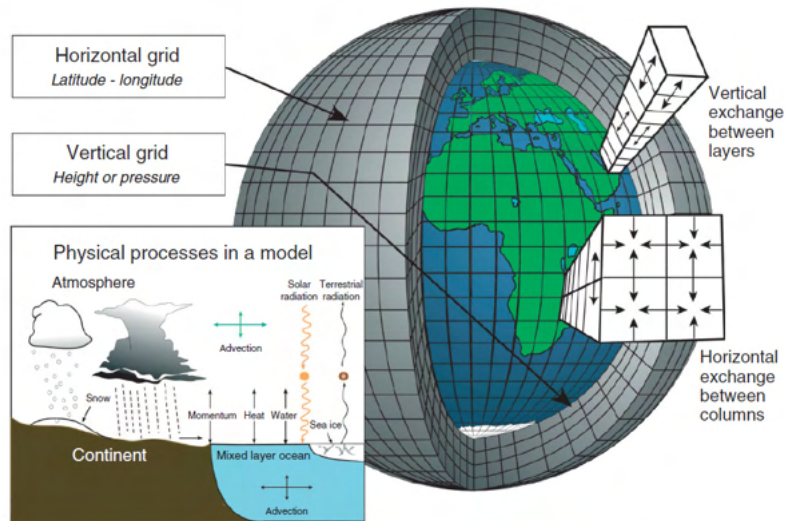


Figura 2.7: Esquema de la malla espacial 3D de los modelos GCM y una descripción de los procesos físicos descritos por este. Elaborado por [6].

entre las mallas a gran resolución horizontal, lo que se debe al tomar una malla del GCM y sobre esta, mediante condiciones de borde y transformaciones polares, se obtiene la malla anillada de mayor resolución, figura 2.8. Los RCM poseen una mayor capacidad predictiva para modelar mallas espaciales a resolución de meso-escala, dado que consideran factores topográficos y micro-climáticos propios del aumento de resolución espacial [5], pues estos suelen parametrizar y aplicar condiciones de frontera para las soluciones de un modelo GCM para así estimar factores a una mayor resolución. También, los RCM penalizan el costo computacional de reparametrizar las soluciones a una escala menor con acotar el dominio espacial de modelación. Sin embargo, estos se vuelve potencialmente costoso en términos computacionales a medida que la resolución espacial aumenta, mostrando limitaciones a su capacidad predictiva del sistema climático. Desde aquí surge la necesidad de adoptar técnica de downscaling dinámico que tengan mayor precisión y permitan predecir a resoluciones desde meso-escala hasta incluso micro-escala.

Los modelos **Weather Research and Forecasting (WRF)**, son un esfuerzo interdisciplinario entre diversas instituciones para generar un salto en la forma de predecir datos climáticos [29]. Estos modelos climáticos regionales del tipo numérico predictivo, predicen principalmente a niveles de meso-escala o escala regional. Según el centro nacional de investigación atmosférica (National Center of Atmospheric Research, NCAR), este modelo usa ecuaciones diferenciales parciales del tipo Navier-Stokes para representar la dinámica del clima, procesos físicos y además datos reales para generar predicciones de mallas, además de ofrecer flujos ideales para pruebas de datos, pues añaden historia del clima mediante forecasting. Con esta información como parámetros, definen escenarios climáticos bajo condiciones atmosféricas para obtener predicciones del sistema climático en una zona ideal entre 1 a 10 km de resolución horizontal, entre meso-escala y micro-escala [29, 30, 22]. También, mediante la dinámica de los modelos RCM, las mallas espaciales del modelo WRF tienen una proporción en su resolución con el fin de preservar el comportamiento de mallas anidadas.

El tiempo de ejecución del modelo WRF se condiciona a los varios factores de inicialización, principalmente la resolución horizontal y el horizonte temporal de la malla a simular, en el caso de mallas anidadas

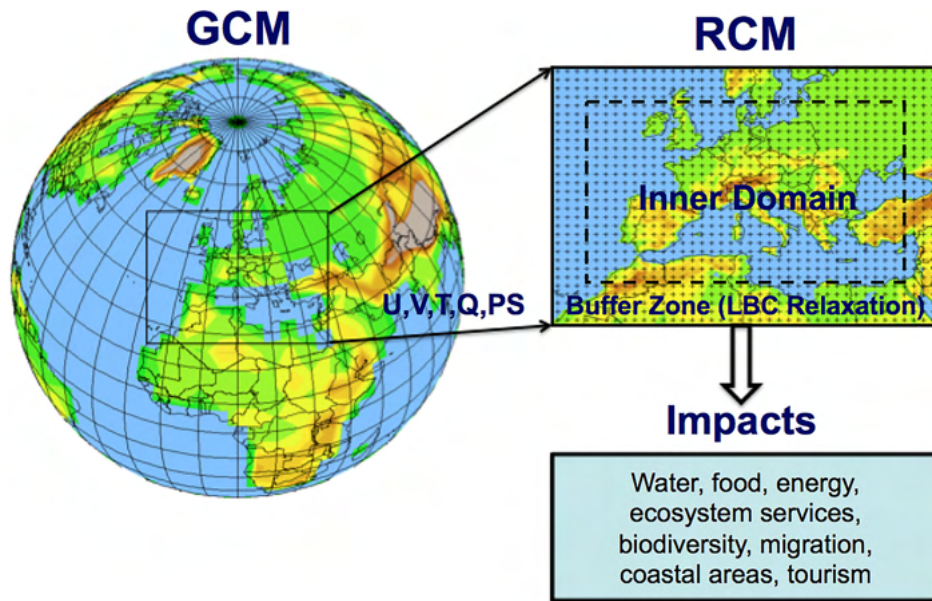


Figura 2.8: Esquema de ejemplo de mallas anilladas entre un GCM y un modelo RCM. Se observan las condiciones de borde y el dominio interno del RCM. Elaborado por [10].

el tiempo se reduce considerablemente. Este tiempo de ejecución tiene un orden de costo computacional máximo de  $\mathcal{O}(\lambda^2)$ , siendo  $\lambda$  el cambio de resolución espacial u horizontal de la malla regular que en nuestro estudio es de 1/3 de kilómetro, preservamos el horizonte temporal y vertical [26]. Por ende, este modelo dinámico debe usarse con precaución para la simulación del clima en dominios espaciales a alta resolución (menor a 1 km).

El estudio del viento en la modelación del clima se ha visto fuertemente beneficiada de los modelos WRF, dado que permite tener un escenario de la evolución del viento fijando condiciones iniciales, condiciones de borde y al forecasting del modelo, además de los factores del sistema climático. En [12] se menciona el potencial de estos modelos para simular todo el espectro de la circulación atmosférica necesario para describir el comportamiento del viento a meso-escala. No obstante, desde [37], se evidencia la dificultad de los modelos WRF para realizar predicción temporal de escenarios lluviosos en zonas de alta variación topográficas como la cordillera de las costas y cordillera de los andes, necesitando técnicas para mejorar la capacidad predictiva local de este modelo.

### 2.3.2. Downscaling Estadístico

El downscaling estadístico es una técnica climatológica para estimar información climática de alta resolución espacial (meso o micro escala) mediante información a baja resolución (escala global o regional), usando modelos estadísticos entre variables climáticas a gran escala y variables locales. Usualmente, la entrada de estos modelos son proyecciones de otros modelos ya determinados a gran escala (GCM o RCM) y necesitan recolectar suficiente información histórica del fenómeno para captar comportamientos y patrones temporales. Una de las principales formas de representar los factores de escala son mediante análisis de regresión [24]. Estos sirven para modelar el impacto climático a nivel local, alimentar modelos de hidrología, agronomía o energía para la planificación de recursos, evaluación de riesgos climáticos y diseño de infraestructura a largo plazo.

Dado que los modelos WRF usan proyecciones de modelos GCM, estos comparten la representación espacial y temporal de la variable climática seleccionada. Al elegir dos modelos WRF de este tipo con sus mallas anidadas a un factor de cambio de escala  $\lambda$ , los fenómenos climáticos se preservan, pudiendo variar levemente según la resolución espacial del modelo. Luego, al aplicar un downscaling estadístico que explique una relación  $\lambda = 1/3$  entre campos de intensidad del viento es necesario plantear las siguientes relaciones fundamentales del método [32]:

- La relación estadística entre variables predictoras y predictando no cambia en el tiempo
- Las variables a gran escala representan componentes climáticas y son sensibles a cambios en el tiempo.
- Las relaciones entre variables predictoras y predictando son altamente colineales.
- El modelo predictor interpreta el proceso físico-climático del predictando.

Existen diversas formas de aplicar downscaling estadísticos sobre mallas espaciales. En [2], se plantea un modelo de downscaling estadístico que construye una relación de escala 1/2 km sobre las mallas de dos modelos WRF a resoluciones 3 km y 1,5 km (este obtenido con la de 1 km con upscaling) por medio de una regresión lineal local con una noción de localidad de vecinos más cercanos de orden uno. También, en [35], se presenta un modelo de downscaling estadístico ampliamente conocido SDSM-DC, por sus siglas en inglés, que plantea una regresión lineal directa para determinar días lluviosos y aplica un downscaling para predecir la precipitación total y temperatura del aire con una regresión lineal local considerando el error del modelo un ruido blanco, entre algunas de sus aplicaciones. Por ende, como se menciona en el aporte del estudio en la Introducción 1, buscamos encontrar un modelo de downscaling estadístico que capture el comportamiento no solo local, sino que también temporal en términos de la variación de las series de la intensidad del viento. A continuación, se presenta el modelo a aplicar en el método de downscaling estadístico mejor conocido como regresión ponderada espacio-temporal.

## 2.4. Modelos de regresión ponderados espacio-temporal (GTWR)

Para implementar un modelo de downscaling estadístico sobre campos de intensidad de viento a una resolución fina, es necesario poder establecer una relación estadística que interprete con robustez la topografía local del terreno, el uso del suelo y los micro-climas que se producen, dado que son causantes del comportamiento irregular de la intensidad del viento cuando hacemos una estimación a micro-escala. Uno de los principales modelos de downscaling estadístico son los modelos de regresión lineal simple [14], que permiten establecer una relación estadística directa entre dos mallas espaciales  $M^{(b)}$  y  $M^{(\lambda b)}$ , sobre una única variable predictor. En los capítulos posteriores consideramos el caso  $\lambda = 1/3$  para  $b = 3$  y  $b = 1$ .

Llamaremos a punto espacial  $v_b(i) \in M^{(\lambda b)}$ , al pixel de la malla  $M^{(b)}$  más cercano al punto espacial  $s_i \in M^{(\lambda b)}$ . Sean las series de tiempo  $X_{v_b(i)}^{(b)} := \{x_{v_b(i),t}^{(b)} : t = 1, \dots, T\}$  que representan a la series de tiempo de la malla espacial  $M^{(b)}$  más cercana al punto espacial  $s_i$  de la malla espacial  $M^{(\lambda b)}$ , con  $\lambda \in (0, 1)$ . Llamamos al modelo de regresión localmente lineal espacio-temporal sobre una malla espacial  $M^{(\lambda b)}$ :

$$X_i^{(\lambda b)} = \beta_i X_{v_b(i)}^{(b)} + \varepsilon_i \quad (2.2)$$

para  $i = 1, \dots, n_{\lambda b}$ , con  $\beta_i \in \mathbb{R}$  los factores de escala,  $\varepsilon_i \in \mathbb{R}^T$  los residuos del modelo, con  $E[\varepsilon_i] = 0$  y  $E[\varepsilon_i^2] < +\infty$ , y  $v_b(i)$  como el vecino de  $M^{(b)}$  más cercano al punto espacial  $s_i$  en la malla  $M^{(\lambda b)}$ .

Este modelo de regresión lineal espacio-temporal solo interpreta características espaciales sobre los campos de viento observados para generar la relación estadística, es decir, si los campos de viento son de origen de un modelo WRF, el modelo adapta con leves variaciones localizadas los mismos patrones topográficos y micro-climáticos de la malla de menor resolución. Sin embargo, esto no es relevante para nuestro análisis, por lo que se propone utilizar los modelos de regresión ponderada espacio-temporal que se adaptan la rugosidad espacial y la variación temporal. En particular, se proponen pesos espacio-temporales  $w_i$  que penalizan la relación topográfica y espacial para cada serie respectiva al punto  $s_i$ , mediante una distancia espacial y una distancia distribucional. La distancia espacial trivialmente mide distancia geodésica (Euclidiana) en metros y la distancia distribucional, ver Definición 10, mide la diferencia de la variación absoluta en el tiempo de las series entre cada punto espacial  $s_i$  y  $s_j$ , penalizando los cambios topográfico y micro-climático locales.

Usamos los resultados de [7, 15] para definir un modelo de regresión ponderada espacio-temporal (GTWR) que se expresa de forma matricial definiendo la matriz de pesos  $W_i$ . Sea

$$X^{(\lambda b)} = \left( X_1^{(\lambda b)}, X_2^{(\lambda b)}, \dots, X_{n_{\lambda b}}^{(\lambda b)} \right)^T,$$

el campo de viento a resolución  $\lambda b$  en forma vectorial y

$$X_{v_b}^{(b)} = \left( X_{v_b(1)}^{(b)}, X_{v_b(2)}^{(b)}, \dots, X_{v_b(n_{\lambda b})}^{(b)} \right)^T,$$

el vector de series del vecino mas cercano en la malla a resolución  $b$  respecto a cada punto espacial en el indexado. Sea el vector de errores

$$\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{n_{\lambda b}})^T.$$

Las ponderaciones se consideran mediante un matriz diagonal  $W_i$ , de coordenadas  $W_{ij}$  definidas mas adelante, denotada de la forma:

$$W_i = \begin{pmatrix} W_{i1} & 0 & \dots & 0 \\ 0 & W_{i2} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & W_{in_{\lambda b}} \end{pmatrix},$$

para  $i = 1, \dots, n_{\lambda b}$ , un escalar espacial  $\beta_i$ , y  $W_{ij}$  los pesos entre los puntos espaciales  $s_i, s_j \in M^{(b)}$  que se definirán a continuación. Para definir los pesos de la matriz de ponderaciones, usamos la distancia espacial  $d_1$  y la distancia distribucional  $d_2$ , vista en la Definición 10, como:

$$d_1(i, j) = |\sigma_i - \sigma_j| \quad \text{y} \quad d_2(i, j) = \|s_i - s_j\|^2,$$

respectivamente, donde  $\|\cdot\|^2$  hace referencia a la distancia euclidiana tradicional,  $|\cdot|$  al valor absoluto y  $\sigma_i$  a la desviación estándar de la serie de tiempo  $X_{v_b(i)}^{(b)}$ . Se resume estas distancias por medio de una re-parametrización:

$$d(i, j) := d_1(i, j)/h_1 + d_2(i, j)/h_2,$$

con  $h_1, h_2 > 0$  anchos de banda arbitrarios. Sea  $h > 0$  el ancho de banda del kernel exponencial. Se denotan los pesos de la matriz de ponderación como:

$$W_{ij} = \exp \left\{ - \left( \frac{d(X_{v_b(i)}^{(b)}, X_{v_b(j)}^{(b)})}{h} \right) \right\}, \quad \forall i, j = 1, \dots, n_{\lambda b}. \quad (2.3)$$

Este tipo de kernel es fundamental porque permite al modelo capturar la variabilidad local del clima, reconociendo que los procesos atmosféricos relevantes para una región pueden ser muy distintos a los de otra, incluso a corta distancia, y que esas relaciones pueden evolucionar en el tiempo. Al asignar mayor peso a observaciones cercanas (en el espacio y tiempo) y penalizar las más distantes, el GTWR mejora la representación de fenómenos climáticos y topográficos locales, que los modelos regionales como el WRF no pueden obtener en un tiempo de ejecución razonable [26]. Desde el modelo de la ecuación (2.2), multiplicando a cada ecuación indexada en  $i$  por el respectivo peso  $W_{ij}$ , ecuación (2.3), se obtiene el sistema de ecuaciones:

$$\begin{cases} W_{i1}X_1^{(\lambda b)} &= \beta_i W_{i1}X_{v_b(1)}^{(b)} + W_{i1}\varepsilon_1 \\ W_{i2}X_2^{(\lambda b)} &= \beta_i W_{i2}X_{v_b(2)}^{(b)} + W_{i2}\varepsilon_2 \\ &\vdots \\ W_{in_{\lambda b}}X_{n_{\lambda b}}^{(\lambda b)} &= \beta_i W_{in_{\lambda b}}X_{v_b(n_{\lambda b})}^{(b)} + W_{in_{\lambda b}}\varepsilon_{n_{\lambda b}} \end{cases} \quad (2.4)$$

Escribiendo (2.4) en forma matricial, sea el modelo de regresión ponderada espacio-temporal (GTWR) para datos dependientes y matriz de pesos  $W_i$ :

$$\boxed{W_i X^{(\lambda b)} = \beta_i W_i X_{v_b}^{(b)} + W_i \varepsilon}, \quad (2.5)$$

para cada  $i = 1, \dots, n_{\lambda b}$ . El factor de escala  $\beta_i$ , que representa la influencia del uso de suelo, topografía y micro-climas para el punto  $s_i$ , se estima minimizando la norma cuadrática ponderada

$$\|W_i(X^{(\lambda b)} - \beta_i X_{v_b}^{(b)})\|_{L^2}^2.$$

La técnica empleada se explica en la Sección 2.5 por métodos de estimación no-paramétrica por núcleo. El objetivo es predecir el campo de viento  $X_i^{(\lambda b)}$  de forma que este adapte las variaciones topográficas complejas y la orografía espacial a alta resolución con una precisión aceptable. Entonces, el campo de viento a resolución  $\lambda b$  se estima como:

$$\hat{X}_i^{(\lambda b)} = \hat{\beta}_i X_{v_b(i)}^{(b)}, \quad \forall i = 1, 2, \dots, n_{\lambda b}. \quad (2.6)$$

Con esto se define la predicción del campo de viento a resolución  $\lambda b$  km como:

$$\hat{X}^{(\lambda b)} := \left\{ \hat{\beta}_i x_{v_b(i),t}^{(b)} : i = 1, \dots, n_{\lambda b}, \text{ y } t = 1, \dots, T \right\} \in \mathbb{R}^{n_{\lambda b} \times T}.$$

## 2.5. Estimación no-paramétrica del factor de escala $\beta_i$

Para el modelo (2.5), los factores de escala  $\beta_i$  en la práctica se estiman de forma no-paramétrica. Uno de los métodos convencionales es por minimización del error cuadrático medio, que en los modelos de downscaling estadístico de regresión ponderada se conoce como método de cuadrados ordinarios ponderados (WLS, por sus siglas en inglés). En nuestro contexto, este método se obtiene desde el modelo (2.5) aplicando la siguiente relación:

$$\hat{\beta}_i := \operatorname{argmin}_{\beta \in \mathbb{R}} \left\{ \sum_{j=1}^n \frac{1}{T} \sum_{k=1}^T (x_{j,k}^{(\lambda b)} - \beta x_{v_b(j),k}^{(b)})^2 W_{ij} \right\}$$

El parámetro  $\beta$  que minimiza la expresión anterior se obtiene diferenciando e igualando a cero:

$$\frac{\partial}{\partial \beta} \left( \sum_{j=1}^n \frac{1}{T} \sum_{k=1}^T (x_{j,k}^{(\lambda b)} - \beta x_{v_b(j),k}^{(b)})^2 W_{ij} \right) = 0$$

Desarrollando la derivada en el término cuadrático y distribuyendo el factor dado por regla de la cadena:

$$2 \sum_{j=1}^n \frac{1}{T} \sum_{k=1}^T \left( \beta(x_{v_b(j),k}^{(b)})^2 W_{ij} - x_{j,k}^{(\lambda b)} W_{ij} x_{v_b(j),k}^{(b)} \right) = 0$$

Separando la sumatoria y despejando el factor  $\beta_i$  se muestra la forma del estimado WLS

$$\hat{\beta}_i = \frac{\sum_{j=1}^n \frac{1}{T} \sum_{k=1}^T x_{j,k}^{(\lambda b)} W_{ij} x_{v_b(j),k}^{(b)}}{\sum_{j=1}^n \frac{1}{T} \sum_{k=1}^T x_{v_b(j),k}^{(b)} W_{ij} x_{v_b(j),k}^{(b)}}.$$

Expresado en forma matricial, para  $T$  el operador traspuesta matricial:

$$\hat{\beta}_i = \left( (X_{v_b}^{(b)})^T W_i X_{v_b}^{(b)} \right)^{-1} (X_{v_b}^{(b)})^T W_i X^{(\lambda b)}. \quad (2.7)$$

En general, para condiciones de independencia espacio-temporal, el estimador  $\hat{\beta}$  es consistente e insesgado [14]. No obstante, estas condiciones son difíciles de obtener en datos realísticos, por lo que se introducirá en el capítulo 4 un estudio para datos simulados con estructura de dependencia espacial y temporal simple, es decir, condiciones de sesgos ideales.

## 2.6. Métricas de evaluación del modelo GTWR

Para medir la eficiencia del modelo de regresión ponderada a la adaptación de patrones topográficos complejos y fenómenos micro-climáticos, usaremos las métricas del coeficiente de estimación de Pearson ( $R^2$ ), el error cuadrático medio (MSE) y la raíz del error cuadrático medio (RMSE). El coeficiente de determinación, explica la capacidad predictiva de un modelo lineal con la proporción de la variabilidad explicada por el modelo entre los datos observados y los predichos, respecto al promedio de los datos observados. El MSE toma el promedio de los errores y el RMSE toma el la raíz cuadrada del promedio de los errores al cuadrado, ambas evaluando la diferencia entre los valores predichos y los observados en promedio.

### 2.6.1. Métricas espaciales

Para el análisis espacial, se toman los campos de viento espaciales  $X_{\cdot,t}^{(b)}$  y  $X_{\cdot,t}^{(\lambda b)}$  fijando un tiempo en  $t \in \{1, \dots, T\}$  e indexando las sumatorias en los puntos espaciales  $i = 1, \dots, n_{\lambda b}$ . Sea  $\bar{X}_{\cdot,t}^{(\lambda b)} = \frac{1}{n_{\lambda b}} \sum_{i=1}^{n_{\lambda b}} x_{i,t}^{(\lambda b)}$  el promedio espacial del campo de viento  $X^{(\lambda b)}$  en el tiempo  $t$ . Denotamos a las métricas de evaluación espacial como:

$$R_t^2 = 1 - \frac{\sum_{i=1}^{n_{\lambda b}} \left( x_{i,t}^{(\lambda b)} - \hat{x}_{i,t}^{(\lambda b)} \right)^2}{\sum_{i=1}^{n_{\lambda b}} \left( x_{i,t}^{(\lambda b)} - \bar{X}_{\cdot,t}^{(\lambda b)} \right)^2}, \quad t \in \{1, \dots, T\},$$

$$\text{MSE}_t = \frac{1}{n_{\lambda b}} \sum_{i=1}^{n_{\lambda b}} \left( x_{i,t}^{(\lambda b)} - \hat{x}_{i,t}^{(\lambda b)} \right)^2, \quad t \in \{1, \dots, T\},$$

$$\text{RMSE}_t = \sqrt{\frac{1}{n_{\lambda b}} \sum_{i=1}^{n_{\lambda b}} \left( x_{i,t}^{(\lambda b)} - \hat{x}_{i,t}^{(\lambda b)} \right)^2}, \quad t \in \{1, \dots, T\}.$$

### 2.6.2. Métricas temporales

Para el análisis espacial, se toma una serie de tiempo  $X_i^{(\lambda b)}$  para un punto espacial fijo  $i \in \{1, \dots, n_{\lambda b}\}$ . Sea

$$\bar{X}_i^{(\lambda b)} = \frac{1}{T} \sum_{t=1}^T x_{i,t}^{(\lambda b)}$$

el promedio de la intensidad de viento de la serie en el punto  $s_i$ , y denotamos a las métricas de evaluación temporal como:

$$R_{s_i}^2 = 1 - \frac{\sum_{t=1}^T \left( x_{i,t}^{(\lambda b)} - \hat{x}_{i,t}^{(\lambda b)} \right)^2}{\sum_{t=1}^T \left( x_{i,t}^{(\lambda b)} - \bar{X}_i^{(\lambda b)} \right)^2}, \quad i \in \{1, \dots, n_{\lambda b}\},$$

$$\text{MSE}_{s_i} = \frac{1}{T} \sum_{t=1}^T \left( x_{i,t}^{(\lambda b)} - \hat{x}_{i,t}^{(\lambda b)} \right)^2, \quad i \in \{1, \dots, n_{\lambda b}\},$$

$$\text{RMSE}_{s_i} = \sqrt{\frac{1}{T} \sum_{t=1}^T \left( x_{i,t}^{(\lambda b)} - \hat{x}_{i,t}^{(\lambda b)} \right)^2}, \quad i \in \{1, \dots, n_{\lambda b}\}.$$

### 2.6.3. Métricas globales

Para el análisis global, se toma todas las observaciones del campo de viento  $X^{(\lambda b)}$ . Sea

$$\bar{X}^{(\lambda b)} = \frac{1}{T \times n_{\lambda b}} \sum_{t=1}^T \sum_{i=1}^{n_{\lambda b}} x_{i,t}^{(\lambda b)}$$

el promedio de la intensidad de viento espacio-temporal. Denotamos a las métricas de evaluación global como:

$$R^2 = 1 - \frac{\sum_{t=1}^T \sum_{i=1}^{n_{\lambda b}} \left( x_{i,t}^{(\lambda b)} - \hat{x}_{i,t}^{(\lambda b)} \right)^2}{\sum_{t=1}^T \sum_{i=1}^{n_{\lambda b}} \left( x_{i,t}^{(\lambda b)} - \bar{X}^{(\lambda b)} \right)^2},$$

$$\text{MSE} = \frac{1}{T \times n_{\lambda b}} \sum_{t=1}^T \sum_{i=1}^{n_{\lambda b}} \left( x_{i,t}^{(\lambda b)} - \hat{x}_{i,t}^{(\lambda b)} \right)^2,$$

$$\text{RMSE} = \sqrt{\frac{1}{T \times n_{\lambda b}} \sum_{t=1}^T \sum_{i=1}^{n_{\lambda b}} \left( x_{i,t}^{(\lambda b)} - \hat{x}_{i,t}^{(\lambda b)} \right)^2}.$$

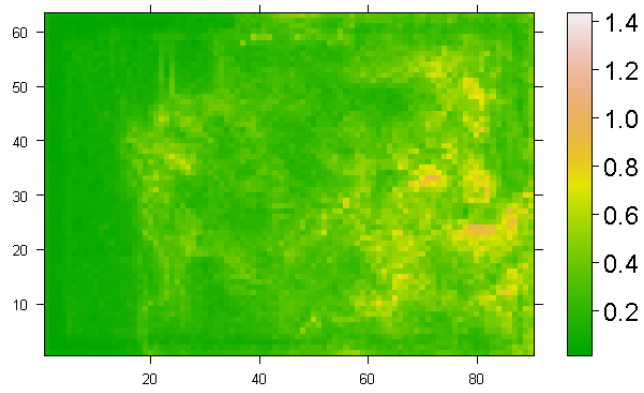
## Capítulo 3

# Metodología de downscaling estadístico

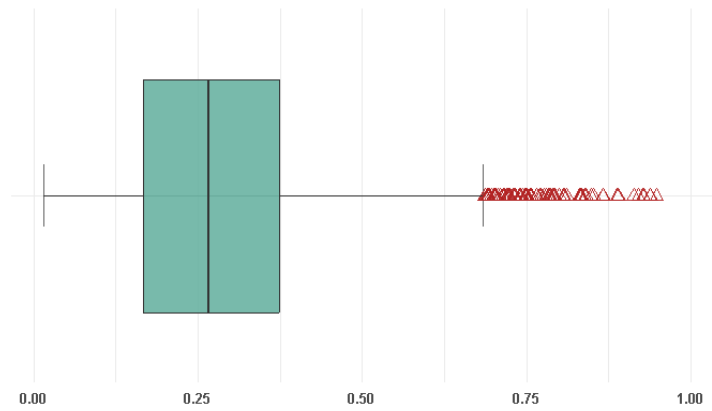
Los campos de viento obtenidos por los modelos WRF dependen de la topografía espacial y la altura con respecto al nivel del mar en que se simulan las variables climatológicas, puesto que a mayor altura las mediciones se vuelven en su totalidad homogéneas en mallas de resolución más baja (mesoescala o escala global). De esta forma, la metodología a aplicar sobre los datos necesita hipótesis que logren plantear una relación estadística suficientemente fuerte para predecir la intensidad del viento en mallas espaciales a una resolución que nos permita identificar diferentes condiciones micro-climáticas y topográficas, por ejemplo, para simular un campo de viento en un estudio de turbulencia en granjas eólicas se necesitan resoluciones espaciales del orden de la micro-escala (al menos el diámetro de un molino eólico de 300 km aproximadamente). Para esto, tomaremos campos de la intensidad de viento de 3 km y 1 km, siendo esta última resolución que admite el modelo WRF para generar predicciones en un tiempo razonable (menor a un día) para una ventana temporal de tres días, dado que, a micro-escala, el costo computacional de los modelos WRF se extiende a un orden exponencial.

Plantear relaciones estadísticas entre los campos de vientos obtenidos por modelos WRF de resolución  $b$  km y otro de resolución  $\lambda b$  es el principal motor que usamos. Se observa que en [2] se emplean métodos de downscaling estadístico con regresiones lineales para vecindades de puntos espaciales establecidas y se contrastan los resultados con un métodos de clasificación para las series de espacio-temporales, concluyendo que las distribuciones estimadas de las clases, determinadas por funciones de kernel, se diferencian por su variabilidad. Por lo tanto, para este estudio, enfocaremos la clasificación presentada en la Sección 2.2 sobre las variabilidad de las series de tiempo, con un fin descriptivo sobre las clases obtenidas por el método sobre la representación topográfica en el dominio espacial de estas, y dar evidencia al uso de ponderaciones espacio-distribucionales al modelo GTWR 2.5.

Para esto, se consideran los incrementos espaciales de la intensidad del viento, ver 2.2, centrando las series en media espacial local y se conserva solo la variación asociada a la topografía en la malla espacial de las series. Sobre los incrementos espaciales se pueden ver los segmentos espaciales representativos de la topografía del dominio espacial evidenciando que la variabilidad de las series caracterizan el terreno asociado, ver figura 3.1. Aplicando el método de clasificación sobre los incrementos espaciales, tres clusters altamente representativos de zonas geográficas: zonas de mar, zonas de valle y zona de montaña, ver figura 4.4, estos son altamente identificables geomorfológicamente en la región de Valparaíso. De esta manera, podemos caracterizar cada zona por una vecindad espacial local con el fin de estudiar el comportamiento espacio-temporal local del viento. A continuación se presentan las hipótesis sobre los campos de intensidad del viento que permiten establecer un downscaling estadístico.



(a) Campo espacial de las variaciones de los incrementos espaciales.



(b) Boxplot del campo espacial de las variaciones de los incrementos espaciales.

Figura 3.1: El campo espacial de las variaciones de los incrementos espaciales (a) y su boxplot (b). En cada pixel  $s_i$  se observa  $\sigma_i$ .

## 3.1. Hipótesis del downscaling estadístico

La naturaleza de los datos dependientes y la topografía asociada a las mallas espaciales llevan a la necesidad de plantear hipótesis que sustenten realizar un modelo de regresión ponderada en la aplicación de una metodología de downscaling estadístico. Como se mostró en la Sección 2.3.2, para proponer una relación lineal entre las mallas  $M^{(b)}$  y  $M^{(\lambda b)}$ , estas mallas al menos deben compartir una serie de tiempo asociada a un punto espacial en común y que pertenezcan a una única vecindad de la malla objetivo, esto se traduce en un relación autosimilar entre las mallas espaciales, pues el fenómeno físico a predecir está explicado por el campo predictor. Además, es necesario que en vecindades espaciales localmente pequeñas las series sean regulares, homogéneas y, al plantear una relación lineal, los errores de estimación sean estacionarios en el tiempo. A continuación, se define la metodología para verificar los supuestos del downscaling estadístico, aplicando las definiciones 6, 7, 8 y 9, sobre dominios espaciales locales, es decir, vecindades suficientemente pequeñas.

### 3.1.1. Regularidad y homogeneidad espacial local

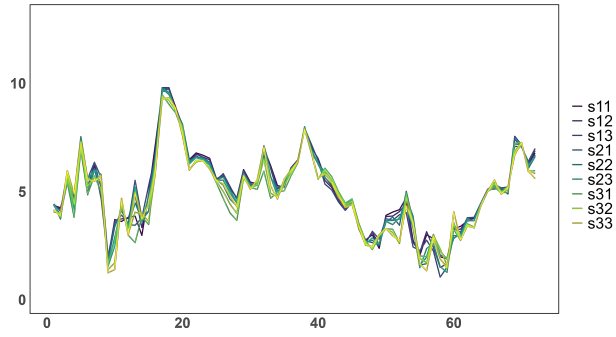
Cuando tratamos campos de intensidad del viento, es casi imposible tener los supuestos derivados de regularidad y homogeneidad espacial, Definición 8 y 9, respectivamente. De aquí se propone trabajar en vecindades localizadas, que llamamos vecindades suficientemente pequeñas o vecindades locales,  $V^{(b)} \subset M^{(b)}$ , definidas en la Sección 2.2, en dónde se espera que sea posible evidenciar comportamientos estadísticos similares entre las series de tiempo que componen estas. Estas vecindades están compuestas de puntos espaciales que serán la principal fuente para predecir la malla espacial y su campo de viento refinado. Además, nos permite identificar los pesos  $W_{ij}$  definidos en la Sección 2.4 para zonas geográficas en dónde al hacer un aumento de resolución espacial, se pueden identificar patrones topográficos localizados.

La noción de regularidad espacial local busca definir un comportamiento de alta dependencia estadística sobre las distribuciones finito-dimensionales de las series de tiempo en una vecindad local (ver Figura 2.4) la que decrece según la distancia espacial de estas, es decir, el comportamiento de viento está inducido por la topografía local del terreno, para así identificar transiciones espaciales suaves entre las series de tiempo de esta vecindad suficientemente pequeña. Por medio de esta suavidad espacial de las distribuciones, podemos definir kernels espacio-temporales para métodos de estimación local suavizados útiles para el contexto de downscaling espacio-temporal.

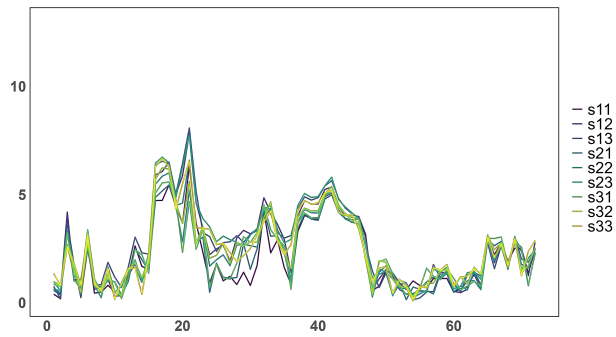
La homogeneidad espacial local implica que el comportamiento estadístico del proceso no cambia significativamente entre las series de tiempo de vecindades suficientemente pequeñas del dominio espacial, permitiendo el uso de modelos lineales con parámetros constantes o suavemente variables en el espacio. En la Figura 3.2, se observan las series de tiempo de los puntos de las vecindades que solo contienen puntos espaciales ubicados en zonas de mar, valle y montaña, determinadas según la clasificación espacial mencionada en la sección anterior. Por lo que la homogeneidad espacial local evidencia comportamientos estadísticos similares entre puntos espaciales contenidos en una vecindad a los puntos del de la malla objetivo en el downscaling.

### 3.1.2. Autosimilaridad espacial local

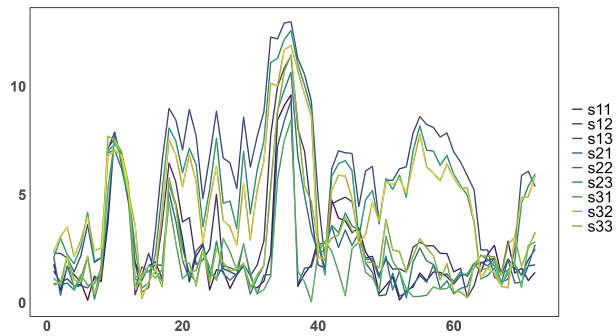
Dado que la malla espacial de los modelos WRF se obtiene desde proyecciones de modelos GCM fijas por condiciones de bordes, las propiedades geométricas sobre las mallas espaciales se preservan si están anilladas. Esta es una característica fundamental de los modelo regionales como el WRF, por



(a) Series de tiempo en zona de mar.



(b) Series de tiempo en zona de valle.



(c) Series de tiempo en zona de montaña.

Figura 3.2: Gráficas de las series de tiempo en vecindades de clusters. Se tiene la zona de mar (a), la zona de valle (b) y la zona de montaña (c).

lo que el factor de escala, ver Definición 7, está presente en la simulación del modelo WRF de mayor resolución. Seleccionamos dos modelos WRF a resolución de malla espacial  $b$  y  $\lambda b$ , por características de los modelos regionales (anillados), la malla de mayor resolución contiene totalmente a la malla menor resolución. Es más, estas mallas por su relación  $\lambda = 1/3$ , comparten un punto central en cada vecindad local, ver Figura 3.3, es decir, existe un único punto espacial  $s_i \in M^{(b)}$ , para cada vecindad de  $M^{(\lambda b)}$ , que permitirá construir una relación lineal directa con los demás puntos espaciales de esta vecindad, que a diferencia de [2], evita problemas de identificabilidad entre vecindades al aplicar una metodología de downscaling estadístico. Por ende, las distribuciones de las series en los puntos espaciales entre las mallas  $M^{(b)}$  y  $M^{(\lambda b)}$  son iguales salvo un factor de escala a encontrar.

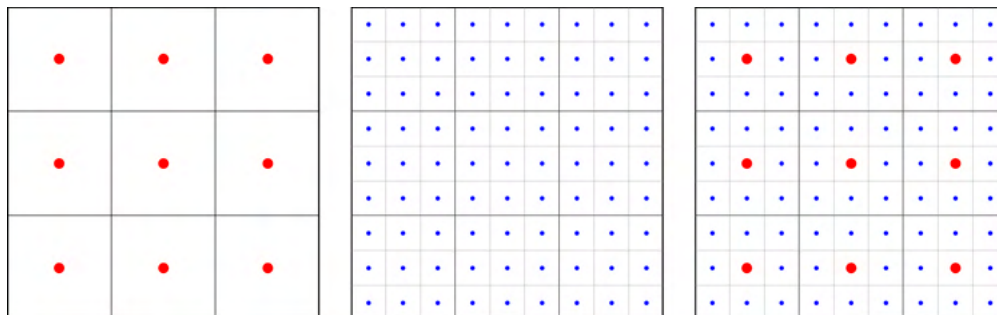


Figura 3.3: Malla espacial a resolución  $b$  km,  $\lambda b$  km y la superposición de estas mallas, respectivamente. Elaboración propia.

### 3.1.3. Estacionariedad local de los residuos

La relación estadística temporal del modelo GTWR se debe preservar en el tiempo para mayor estabilidad del modelo. La estacionariedad sobre los residuos es difícil de sostener de forma global, pero basta encontrar vecindades suficientemente locales en donde los residuos de nuestro modelo sean estadísticamente constantes o con variaciones no significativas. Desde la definición 6, los residuos de la estimación de la intensidad de viento a la malla de 1 km, muestran indicios de estacionariedad local sobre un intervalo de confianza del 99 %, por lo que se asume estacionariedad local para todas las clases representativas del dominio espacial, ver figura 3.4.

Para validar el comportamiento visual de los residuos consideramos el *PP-test* de la raíz unitaria para las series de tiempo de los residuos, en donde para que existan indicios de estacionariedad, se debe encontrar un valor *PP-estadístico* bajo el umbral del 5 % del valor crítico [21]. Esto nos da el paso a incorporar un kernel espacial y distribucional, el que puede eliminar dependencias o efectos no aleatorios en los residuos [13] y se obtienen predicciones con menor sesgo, para así identificar patrones en los factores de escala espacial de manera determinista asegurando efectividad predictiva sobre el modelo.

## 3.2. Metodología de Downscaling Estadístico

La metodología de downscaling estadístico se basa en establecer una relación de  $\lambda = 1/3$  entre las mallas espaciales  $M^{(3)}$  y  $M^{(1)}$ , por medio de una modelo GTWR directamente entre los campos de viento respectivos, con el fin de predecir los valores de los factores de escala  $\beta$ , con el procedimiento mostrados en la Sección 2.4 y 2.5, buscando un ancho de ventana  $h$  óptimo para el kernel exponencial. Una vez detectado los mejor estimador  $\beta_i$ , se aplica al modelo GTWR, Sección 2.4, entre las mallas

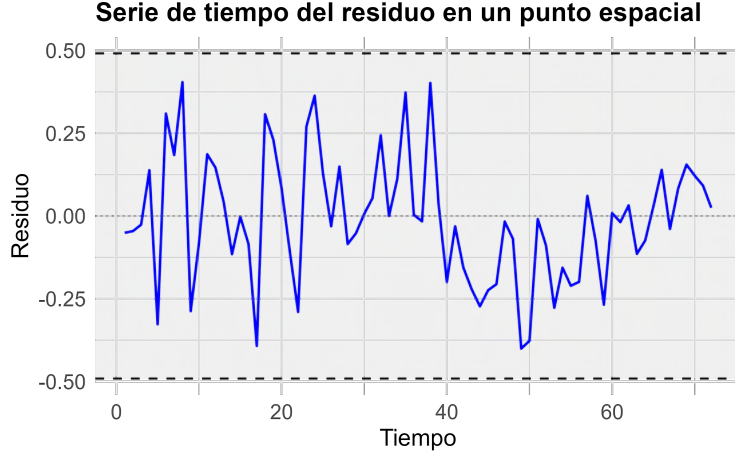


Figura 3.4: Estacionariedad del residuo local  $\varepsilon_{s_{287}}$  en posición  $s_{287} \in M^{(1)}$ , con intervalos de confianza al 99%.

$M^{(1)}$  y  $M^{(1/3)}$ . Considerar que  $\beta_i$  se debe adaptar a la escala de la malla objetivo  $M^{(1/3)}$ , en nuestro caso, con el método del vecino más cercano de primer orden. Por último, se predice el campo de viento sobre la malla  $M^{(1/3)}$  con el mismo procedimiento anterior y usando los factores de escala re-escalados a esta nueva malla a predecir. El procedimiento a continuación resume esta metodología de downscaling estadístico.

En primera instancia, se define la variable predictora del modelo GTWR, el campo de viento del vecino más cercano de primer orden a 3 km,  $X_{v_3(i)}^{(3)}$  sus series de tiempo, junto a la matriz diagonal de ponderaciones espacio-temporal determinada por los pesos  $W_{ij}$ , caracterizadas por el kernel exponencial y la distancia espacio-temporales, Definición 2.4 y siguiendo con esta línea se obtiene el modelo GTWR:

$$\boxed{W_i X_i^{(1)} = \beta_i W_i X_{v_3(i)}^{(3)} + W_i \varepsilon_i}, \quad (3.1)$$

para cada  $i = 1, \dots, n_1$ .

Implementar el modelo GTWR requiere de estimar los factores de escala, como se planteó en la Sección 2.5, se obtienen los estimadores de  $\beta_i$ :

$$\hat{\beta}_i = ((X_{v_3}^{(3)})^T W_i X_{v_3}^{(3)})^{-1} (X_{v_3}^{(3)})^T W_i X_i^{(1)}.$$

El esquema de downscaling debe ser suficientemente robusto para implementar la relación  $\lambda = 1/3$  entre los campos de viento  $X^{(3)}$  y  $X^{(1)}$ , por medio de la asignación de un ancho de banda óptimo para el kernel exponencial de los pesos espacio-temporales. Luego de establecer un ancho de banda que reduzca el sobre ajuste del modelo, se utiliza el método del vecino más cercano sobre los factores de escala  $\beta_i$  estimados en el campo a 1 km, así obteniendo los factores de escala estimados para el campo a resolución 1/3 km, para aplicar el modelo lineal con el campo de viento  $X_{v_1}^{(1)}$  y obtener la predicción del campo de viento  $X^{(1/3)}$ , es decir:

$$\boxed{\hat{X}_i^{(1/3)} = \hat{\beta}_i X_{v_1(i)}^{(1)}}, \quad \forall i = 1, \dots, n_{1/3}. \quad (3.2)$$

## Capítulo 4

# Resultados

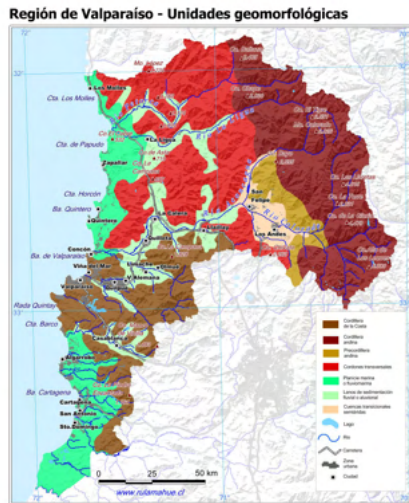
### 4.1. Caso de estudio

El presente estudio se centra en Chile, país que se emplaza en el extremo sur del continente americano y cuya principal característica es una vasta diversidad geográfica y geomorfológica. Su configuración longitudinal abarca, de este a oeste, un litoral costero, la cordillera de la Costa, el valle central y la cordillera de los Andes. Esta heterogeneidad se manifiesta también en su eje norte-sur, donde los ecosistemas transitan desde el desierto de Atacama hasta la Patagonia chilena, atravesando estepas, tundras y bosques templados. Dicha diversidad implica que las distintas regiones del país presentan variaciones significativas en su topografía, biomas y ecosistemas, lo cual es un factor determinante para el estudio de variables climáticas locales como el viento.

La Región de Valparaíso, seleccionada como zona de estudio, presenta una notable riqueza geográfica y una alta relevancia estratégica para el país. Su importancia económica se sustenta en la actividad de sus puertos, como Valparaíso y San Antonio, que canalizan una parte fundamental del comercio nacional. Sin embargo, esta dinámica se ve contrapuesta por una creciente vulnerabilidad a desafíos ambientales, como el aumento en la tasa de incendios forestales que han afectado gravemente zonas urbanas e infraestructuras críticas. En este contexto, la región también se ha posicionado como un polo de desarrollo e innovación en energías renovables, evidenciado por la implementación de proyectos como el Parque Eólico Cerro Piedra [27] y la concentración de parques fotovoltaicos en comunas como Petorca, San Felipe y Los Andes [17]. Esta dualidad entre potencial económico y desafíos climáticos justifica la necesidad de modelar con precisión sus recursos eólicos.

Recordando, el objetivo central de este trabajo es modelar campos de intensidad de viento a una resolución final de 1/3 km, aplicando un esquema de downscaling estadístico. La metodología se desarrolla en dos etapas. Primero, se implementa un modelo GTWR para establecer la relación entre los campos de viento de 3 km y 1 km, obtenidos del modelo climático regional WRF. Este paso permite recuperar los factores de escala espacial a 1 km, considerando las características topográficas y micro-climáticas de vecindades suficientemente locales. En una segunda etapa, se aplica nuevamente el esquema de downscaling, utilizando el campo de viento de 1 km como predictor y los factores de escala previamente estimados, adecuadamente ajustados a la nueva malla, para generar la predicción sobre el campo objetivo de 1/3 km.

Los datos base para este estudio consisten en campos de intensidad de viento obtenidos del modelo WRF V3.8.1, con resoluciones espaciales de 3 km y 1 km. Las mediciones corresponden a una altura



(a) Zonas geomorfológicas.



(b) Zonas con sombras de altura.

Figura 4.1: Mapas temáticos de la región de Valparaíso. Elaborado por [23].

de 10 metros sobre el nivel del mar y comprenden 73 observaciones temporales en horas, abarcando el período desde el 02/09/2023 00:00:00 hasta el 05/09/2023 00:00:00. Es relevante destacar que durante este intervalo, Chile se encontraba bajo los efectos del fenómeno de El Niño, condición que se manifiesta en una reducción de la intensidad de los vientos alisios y un aumento en los vientos costeros. Este contexto climático particular proporciona un escenario idóneo para evaluar la capacidad del modelo de downscaling estadístico (con una relación de escala  $\lambda = 1/3$ ) para capturar dinámicas atmosféricas complejas.

Dada la naturaleza anidada de los modelos WRF, el dominio espacial del estudio queda definido por la malla de mayor resolución (1 km), cuyas coordenadas para la Región de Valparaíso abarcan las latitudes Este-Oeste entre -33.43985 y -32.88078, y las longitudes Norte-Sur entre -71.89282 y -70.93353, ver Figura 4.2. Dentro de este dominio compartido, la malla de baja resolución  $M^{(3)}$  se compone de 630 puntos espaciales, mientras que la malla de alta resolución  $M^{(1)}$  contiene 5670 puntos. Cabe señalar que, para el análisis, se omite la primera observación temporal de la simulación ( $t = 0$ ), ya que esta representa las condiciones iniciales del modelo y exhibe una homogeneidad espacial que no es representativa de la dinámica climática posterior.

El dominio espacial definido abarca una compleja geomorfología que incluye zonas de mar, planicies marinas, llanos de sedimentación y la cordillera de la Costa. Dada esta heterogeneidad topográfica, y como se fundamentó en el capítulo anterior, es indispensable aplicar un método de clasificación para identificar patrones espaciales coherentes. Utilizando el método de clasificación basado en las variaciones de los incrementos espaciales (descrito en la Sección 2.1), se logró caracterizar la topografía del área de estudio. Este proceso resultó en la identificación de tres clústeres altamente interpretables que representan zonas topográficas distintas: 1) Mar, caracterizado por variaciones mínimas; 2) Valle, con zonas de variación intermedia; y 3) Montaña, que agrupa áreas de alta variabilidad (ver Figura 4.4). A partir de esta clasificación, se seleccionaron los centros representativos de cada clúster para definir las vecindades de análisis sobre las cuales se evaluará la capacidad predictiva del modelo GTWR.

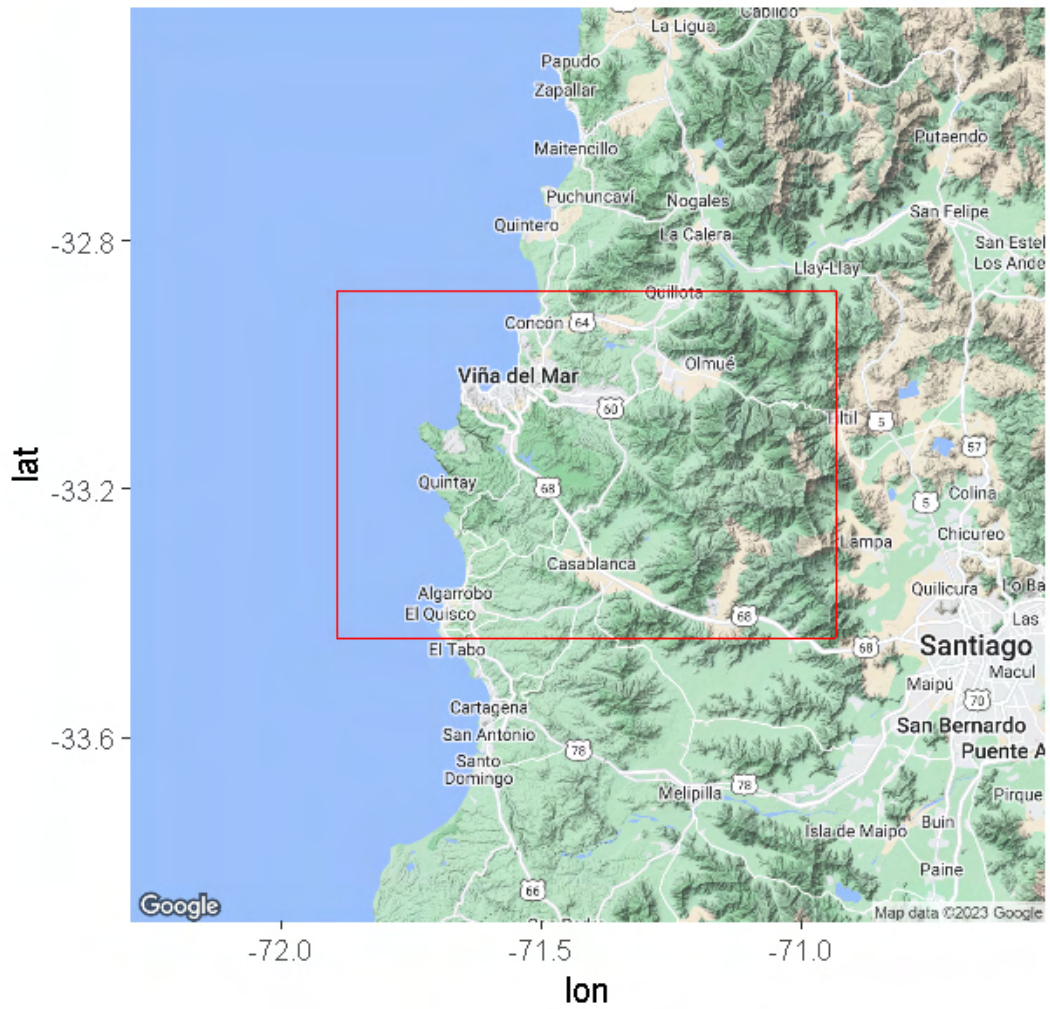


Figura 4.2: Zona geográfica del caso de estudio, donde la zona en rojo el campo de viento a estudiar, con coordenadas latitud Este-Oeste  $-33,43985$  y  $-32,88078$ , y longitud Norte-Sur  $-71,89282$  y  $-70,93353$ .

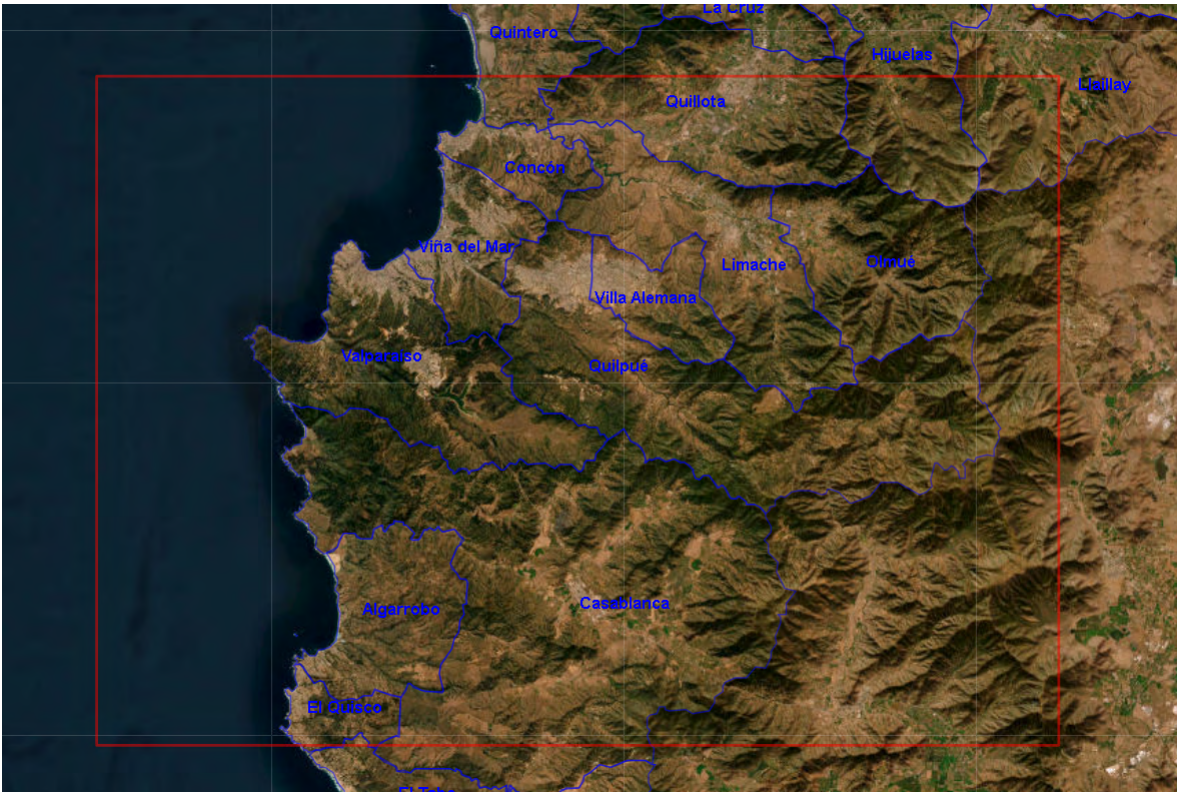


Figura 4.3: Campo de viento a estudiar delimitado en rojo y subdivisión comunal de la región de Valparaíso, con coordenadas latitud Este-Oeste  $-33,43985$  y  $-32,88078$ , y longitud Norte-Sur  $-71,89282$  y  $-70,93353$ .

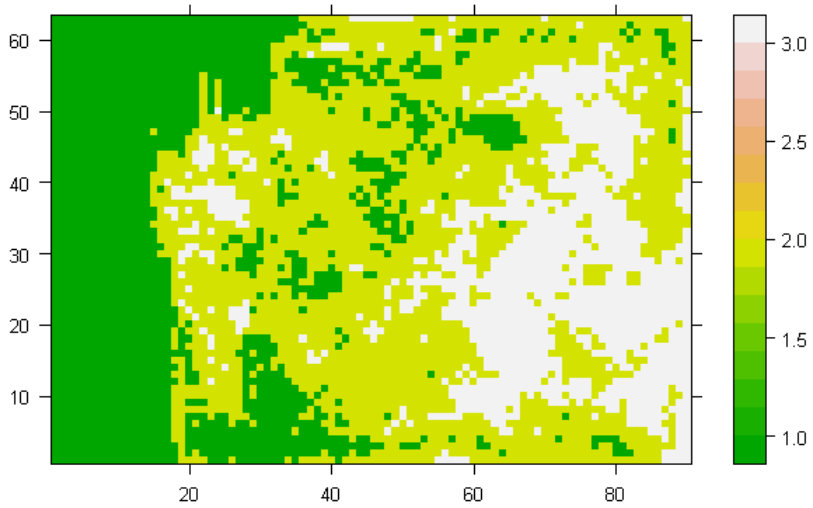


Figura 4.4: Zonas de clusters aplicando la clasificación FPC para  $k = 3$ .

## 4.2. Estudio de simulación

Para evaluar la capacidad predictiva intrínseca del modelo GTWR en un entorno controlado, se diseñó un estudio de simulación. El objetivo es verificar la habilidad del modelo para estimar correctamente los parámetros de un proceso localmente lineal en presencia de ruido estocástico, que simula la aleatoriedad inherente al clima. El procedimiento se desarrolló en tres fases:

1. Simulación de factores de escala reales ( $\beta_{\text{real}_i}$ ): Primero, se generó un campo de factores de escala de referencia a 1 km. Para ello, se utilizó el índice de uso de suelo proporcionado por el modelo WRF re-parametrizados a un valor continuo  $\beta_{\text{lu\_index}_i} \in (1, 2)$ , y se le adicionó un ruido blanco gaussiano  $\varepsilon_{\text{lu\_index}_i} \sim \mathcal{N}(0, \sigma_{\text{lu\_index}}^2)$  para introducir una variabilidad realista resultando en  $\beta_{\text{real}_i} = \beta_{\text{lu\_index}_i} + \varepsilon_{\text{lu\_index}_i}$ , para todo  $i = 1, \dots, n_1$ .
2. Construcción del campo de viento sintético ( $X_{\text{sim}}^{(1)}$ ): A continuación, se construyó un campo de viento objetivo a 1 km aplicando el modelo de regresión localmente lineal 2.2. Este campo sintético se calculó como  $X_{\text{sim}}^{(1)} = \beta_{\text{real}_i} X_{v_3(i)}^{(3)} + \varepsilon_i$ , con  $\varepsilon_i \sim \mathcal{N}(0, \sigma_{\text{error}}^2)$ .
3. Aplicación y evaluación del modelo: Finalmente, se aplicó el modelo GTWR al esquema  $X^{(3)} \rightarrow X_{\text{sim}}^{(1)}$  para estimar los factores de escala  $\hat{\beta}_{\text{real}_i}$  y predecir el campo de viento  $\hat{X}_{\text{sim}}^{(1)}$ . Para analizar la robustez del modelo, este proceso se repitió para un conjunto de distintas varianzas del error residual  $\varepsilon_i$ :  $\sigma_{\text{error}}^2 \in \{0,005; 0,01; 0,03; 0,05; 0,1\}$ .

El campo simulado  $X_{\text{sim}}^{(1)}$  fue diseñado para poseer características estadísticas específicas. El ruido añadido tanto en el factor de escala como en el residuo es, por construcción, idénticamente e independientemente distribuido (i.i.d.). Esta propiedad asegura que el proceso resultante preserve la estacionariedad local y la autosimilaridad espacial, condiciones teóricas fundamentales para el correcto desempeño del modelo. Bajo estos supuestos ideales, y en línea de [11], se espera que los estimadores de los factores de escala obtenidos mediante el método de Mínimos Cuadrados Ponderados (WLS) sean consistentes y converjan en probabilidad a los valores reales ( $\beta_{\text{real}_i}$ ) utilizados en la simulación. Este diseño nos permite establecer una base para evaluar el rendimiento del modelo.

El núcleo del experimento consistió en cuantificar la sensibilidad del modelo GTWR ante perturbaciones de ruido de distinta amplitud. Para ello, se aplicó el modelo de forma no paramétrica para cada uno de los niveles de varianza residual  $\sigma_{\text{error}}^2$  definidos, utilizando el campo real  $X^{(3)}$  como predictor. Un parámetro crucial en esta etapa es el ancho de ventana del kernel,  $h$ , que se fijó en 0.02. Este valor fue seleccionado tras un análisis exploratorio, buscando un equilibrio óptimo que prioriza la cercanía espacial y la similitud variacional entre puntos, pero evitando el sobreajuste que podría derivarse de un ancho de ventana excesivamente pequeño. La capacidad predictiva y el grado de sobreajuste del modelo se evaluaron cuantitativamente utilizando las métricas definidas en la Sección 2.4 ( $R^2$ ,  $MSE$  y  $RMSE$ ). El análisis se realizó descriptivamente en el contexto de campos espaciales, examinando la precisión del modelo en instantes de tiempo específicos ( $t \in \{2, 20, 38, 56\}$ ) para observar el rendimiento en diferentes momentos de la simulación, observar las Figuras A.2, A.4, A.6, A.8 y A.10.

Los resultados de la simulación demuestran una alta eficacia del método WLS (ver Sección 2.5) para recuperar los factores de escala. Como se observa en las Figuras A.1, A.3, A.5, A.7 y A.9, los factores de escala estimados son notablemente similares a los valores reales simulados en la mayoría de los escenarios. Sin embargo, en el caso de la varianza residual más alta ( $\sigma_{\text{error}}^2 = 0,1$ ), se observa un leve sesgo en la estimación: el modelo tiende a subestimar ligeramente el factor de escala en zonas de

alta variabilidad topográfica simulada, como las cimas de montaña. Este ligero sobreajuste se manifiesta en un valor máximo estimado de 2.2, frente al máximo simulado de 2.0. Para los niveles de ruido más bajos, las diferencias entre los campos de factores de escala simulados y estimados son visualmente imperceptibles, confirmando la robustez del estimador.

Observamos que en la variación del error del campo simulado, la capacidad predictiva del modelo en base al nivel de las variación temporales de las series de tiempo es alta y estable, observando las métricas globales (Sección 2.6.3) del modelo GTWR, ver Cuadro 4.1, se obtuvieron errores de un orden menor a  $\mathcal{O}(10^{-3})$  para el  $MSE$ , un error del orden menor a  $\mathcal{O}(10^{-2})$  para el  $RMSE$  y un  $R^2$  muy cercano a la unidad, explica una excelente capacidad predictiva para explicar el campo de viento simulado. Se concluye que para variaciones leves, moderadas y altas del error del campo simulado  $X_{sim}^{(1)}$ , el modelo es robusto en captar patrones y estructuras espacio-temporales cuando los ruidos son (i.i.d.) y estacionarios.

$\sigma_{res}^2$	MSE	RMSE	$R^2$
0,005	0,00016	0,01269	1,0000
0,010	0,00018	0,01348	1,0000
0,030	0,00035	0,01875	1,0000
0,050	0,00071	0,02661	0,9999
0,100	0,00230	0,04791	0,9997

Cuadro 4.1: Métricas globales (sección 2.6.3) obtenidas para distintos variaciones  $\sigma_{error}^2$ .

En conclusión, la simulación de un campo de intensidad del viento a 1 km mediante una relación localmente lineal con un comportamiento de ruido Browniano simple, muestra que el modelo GTWR entiende relaciones de dependencia espacial y temporal local. Sin embargo, en el contexto de los datos reales, estas estructuras pasan a ser de mayor complejidad en resoluciones espaciales finas y las estructuras en los datos reales, describen los fenómenos micro-climáticos, cambios de topografía y orografía asociada sobre el dominio espacial.

### 4.3. Resultados del esquema de downscaling estadístico

Dado que la topografía y los fenómenos micro-climáticos son los principales determinantes del comportamiento del viento, es fundamental evaluar la capacidad del esquema de downscaling para representar estas características a alta resolución. En esta sección se abordan los resultados de dicho esquema, comenzando con la selección de tres vecindades espaciales que encapsulan la diversidad del terreno: mar, valle y montaña. A través de un análisis descriptivo de estas zonas, se buscará evidenciar las relaciones estadísticas que sustentan las hipótesis del modelo. Posteriormente, se discutirá el desempeño del modelo GTWR, analizando la precisión de sus estimaciones tanto en el dominio espacial como en el temporal. Por último, presentamos el campo de viento estimado a una resolución de 1/3 de km, realizando una comparación descriptiva con las resoluciones de 3 y 1 km para contrastar el aumento de resolución y la calidad del ajuste del método propuesto.

#### 4.3.1. Análisis descriptivo WRF 3 y 1 km

Esta sub-sección se dedica al análisis descriptivo de los datos WRF de 3 km y 1 km, sentando las bases para la aplicación del modelo GTWR. Como se ha establecido, la compleja topografía de la región de Valparaíso induce un comportamiento globalmente irregular y heterogéneo en la intensidad del viento.

Sin embargo, la hipótesis fundamental del enfoque de downscaling local es que, al reducir el análisis a vecindades espaciales suficientemente pequeñas, es posible identificar dominios donde el comportamiento estadístico de las series de tiempo es más homogéneo y predecible. Por lo tanto, el objetivo de este análisis es caracterizar estas vecindades locales para confirmar que representan comportamientos estadísticos únicos y consistentes, validando así la pertinencia del enfoque local.

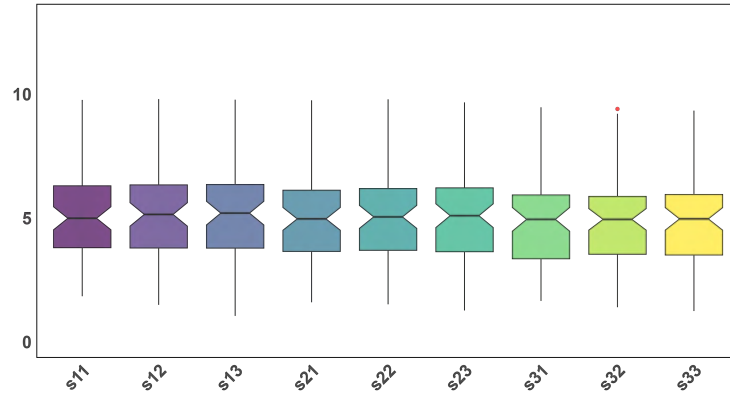
Para vincular cuantitativamente la topografía con la variabilidad del viento, se utilizó una métrica basada en los incrementos espaciales. Al centrar las series de tiempo respecto a sus vecinos locales, se aísla la variabilidad puramente espacial y se mitigan los efectos de dependencias temporales a gran escala, como la estacionalidad periódica. La desviación estándar de estas series de incrementos se convierte entonces en un indicador robusto de la rugosidad topográfica local. Posteriormente, se aplicó el método de clasificación del vecino más lejano (FPC, descrito en la Sección 2.1) sobre esta métrica de variabilidad. El resultado es una clasificación que identifica eficazmente zonas de homogeneidad local, delineando comportamientos comunes que se corresponden directamente con la geomorfología típica del dominio: mar, valle y montaña.

Tras la clasificación, se seleccionaron las vecindades representativas de cada clúster para un análisis estadístico detallado, donde cada vecindad se centra en el punto representativo de su clase. Las características de cada zona son las siguientes:

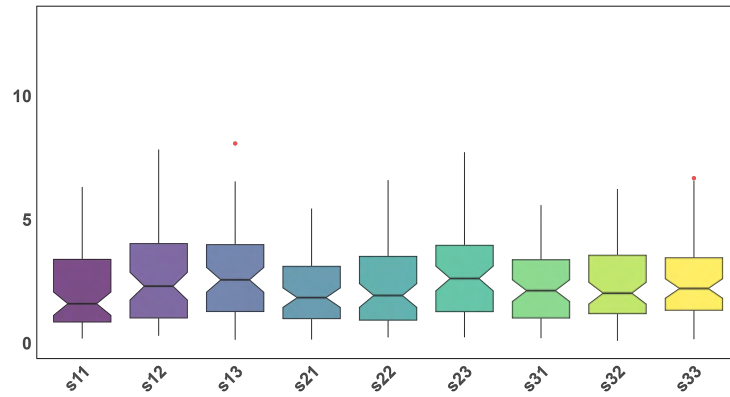
- Clúster de Mar (Figuras 4.5a y 4.6a): Presenta la menor variabilidad. Las series de tiempo en los puntos de su vecindad mar ( $s_{ij} \in V_{\text{mar}}$ ) muestran un rango de valores entre  $[1,05; 9,8]$   $m/s$ , con medianas entre  $[1,53; 1,84]$   $m/s$  y desviaciones estándar entre  $[1,75; 1,83]$   $m/s$ .
- Clúster de Valle (Figuras 4.5b y 4.6b): Muestra un comportamiento intermedio. El rango de valores se sitúa entre  $[0,08; 8,08]$   $m/s$ , con medianas entre  $[1,59; 2,61]$   $m/s$  y desviaciones estándar entre  $[1,46; 1,84]$   $m/s$ .
- Clúster de Montaña (Figuras 4.5c y 4.6c): Evidencia mayor dispersión y valores más extremos. El rango de intensidad del viento es el más amplio, entre  $[0,05; 12,98]$   $m/s$ , con medianas que varían significativamente entre  $[1,42; 6,28]$   $m/s$  y las desviaciones estándar más altas, entre  $[1,77; 3,05]$   $m/s$ , lo que es consistente con su compleja topografía.

Detectar correlación y existencia de relaciones lineales entre series de una misma vecindad local es fundamental para construir nuestro modelo GTWR. Si los puntos de una vecindad local están altamente correlacionados, al aplicar un ponderación por los vecinos más cercanos para determinar el factor de escala, encontraremos que el factor de escala es muy cercano a la unidad, es decir, la mejor aproximación es la misma serie predictora. Por el contrario, si el coeficiente de correlación no es alto, se espera estimar un factor de escala que llega incluso a duplicar o dividir a la mitad el valor de la intensidad del viento. En las Figuras 4.7, 4.8 y 4.9, la correlación entre las series espaciales de una vecindad está determinada por su variación, siendo las vecindades de montaña las candidatas a tener un ajuste determinado por cambios abruptos en la intensidad del viento por el modelo GTWR. Pese al aumento variacional en la vecindad de montaña, es posible establecer relaciones lineales.

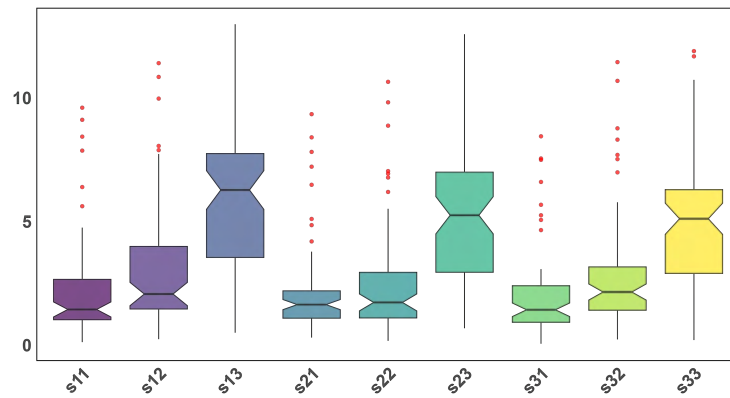
Un paso fundamental en este análisis descriptivo es la evaluación de la correlación y la linealidad entre las series de tiempo dentro de cada vecindad, ya que estas propiedades sustentan la construcción del modelo GTWR. Teóricamente, si las series en una vecindad están perfectamente correlacionadas, el factor de escala estimado tenderá a la unidad. Por el contrario, una menor correlación sugiere que



(a)

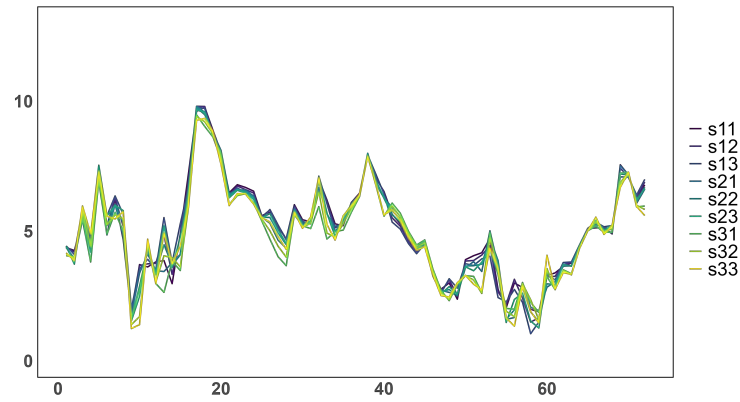


(b)

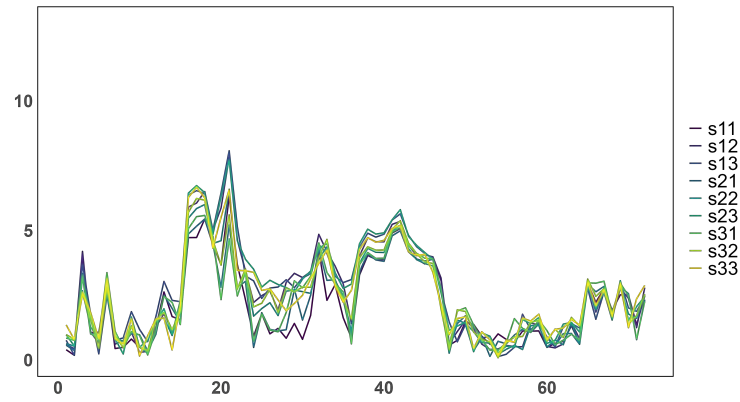


(c)

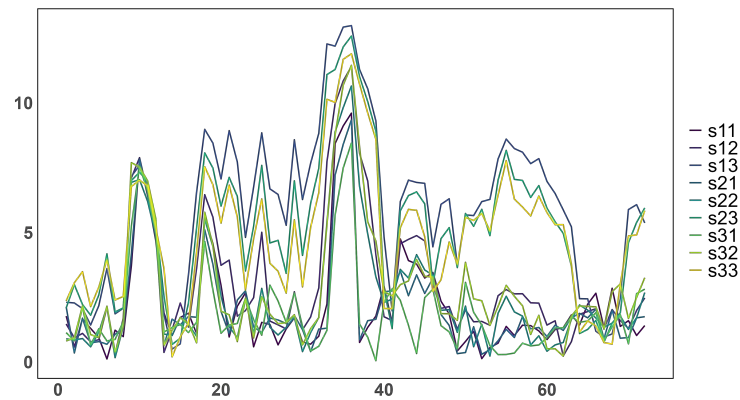
Figura 4.5: Gráficas boxplot de vecindades del campo de intensidad de viento. (a-f) Se observan los boxplots de las vecindades de mar, valle, montaña, mar-valle, valle-montaña y mar-valla-montaña, respectivamente.



(a) Vecindad de Mar



(b) Vecindad de Valle



(c) Vecindad de montaña

Figura 4.6: Gráficas series de tiempo en vecindades del campo de intensidad de viento. Se observan las series de tiempo de las vecindades de mar (a), valle (b) y montaña (c), respectivamente.

el modelo necesitará un factor de escala significativamente distinto de uno para capturar los cambios de intensidad. El análisis de las matrices de dispersión (Figuras 4.7, 4.8 y 4.9) revela que, a pesar del aumento progresivo de la variabilidad desde el mar hacia la montaña, en todas las vecindades se mantienen relaciones predominantemente lineales. Esto es muy importante, ya que confirma que incluso en las zonas de montaña, donde los cambios de viento son más abruptos, es factible establecer las relaciones lineales que el modelo GTWR requiere para su ajuste.

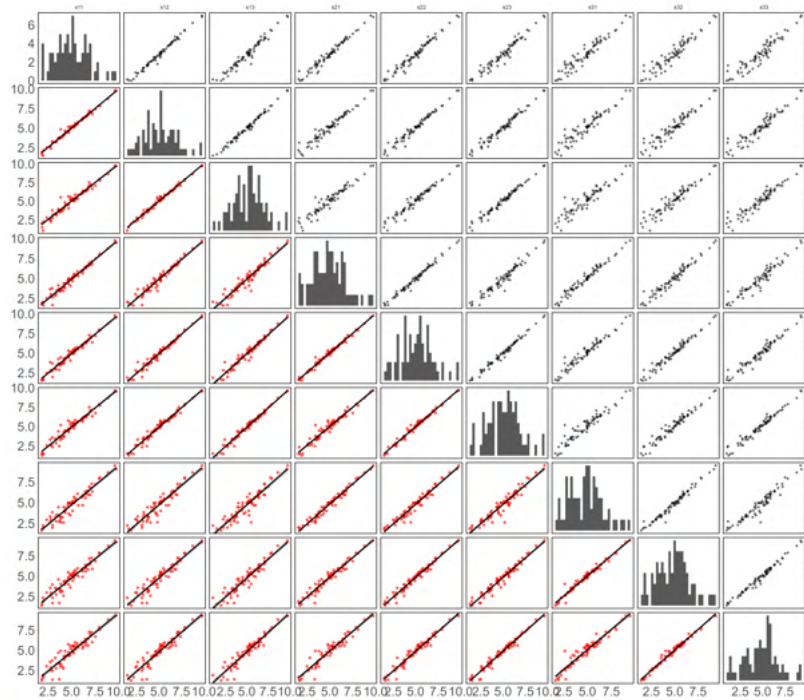


Figura 4.7: Matriz de dispersión, regresión lineal en rojo, entre las series de la vecindad en el mar. En la diagonal se observan los histogramas de las series.

Habiendo establecido un núcleo de evidencia estadística a partir de este análisis descriptivo, que confirma la existencia de homogeneidad local, relaciones lineales y la coherencia de la clasificación topográfica, se procede en la siguiente sección a presentar y discutir los resultados de la aplicación del modelo GTWR. Este análisis previo, respaldado además por los hallazgos del estudio de simulación, proporciona la justificación empírica necesaria para la implementación del método de downscaling estadístico sobre los datos reales.

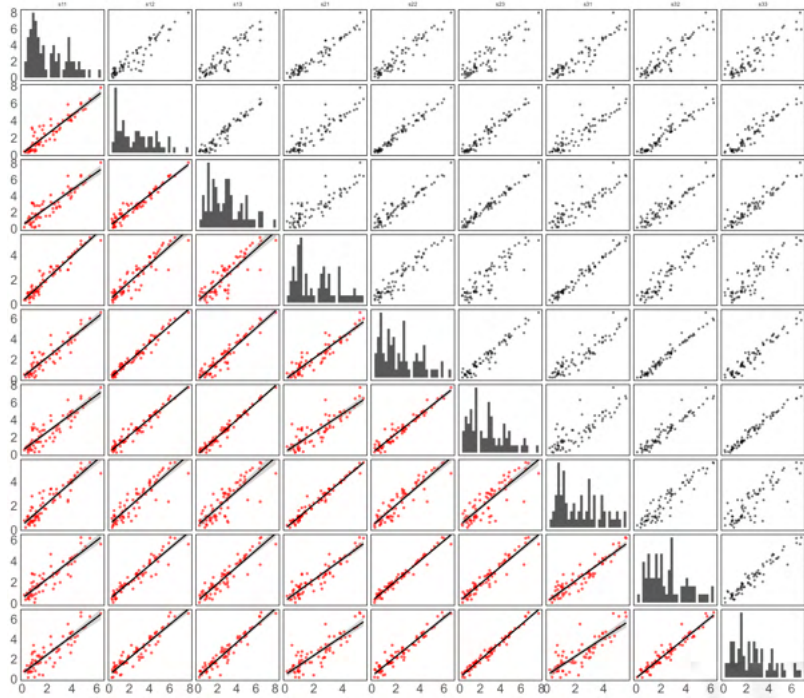


Figura 4.8: Matriz de dispersión, regresión lineal en rojo, entre las series de la vecindad en la valle. En la diagonal se observan los histogramas de las series.

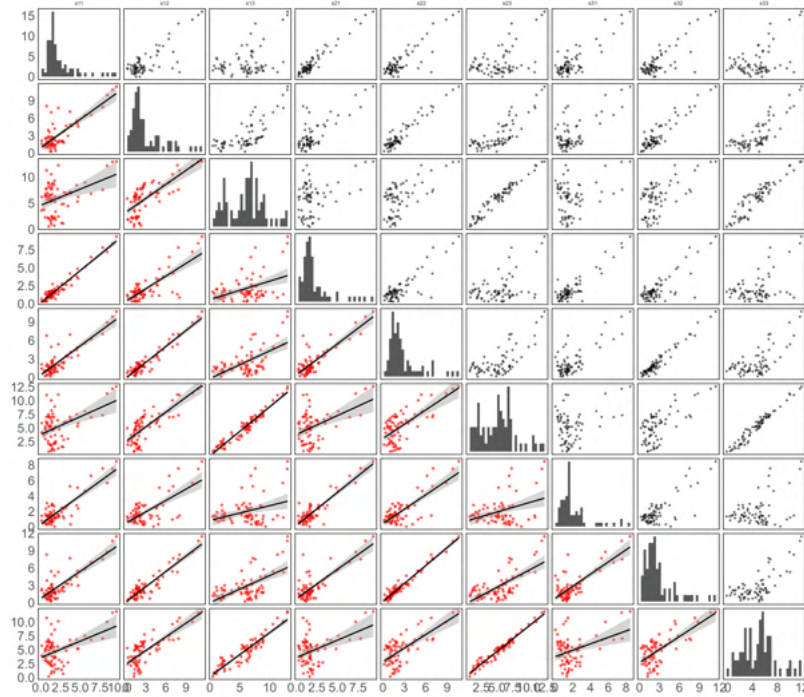


Figura 4.9: Matriz de dispersión, regresión lineal en rojo, entre las series de la vecindad de montaña. En la diagonal se observan los histogramas de las series.

### 4.3.2. Modelo GTWR: estimación del campo a 1 km

Con las hipótesis del modelo validadas empíricamente en la sección anterior, se procede a la implementación del esquema GTWR para estimar el campo de viento de 1 km a partir del campo de 3 km. El primer paso consiste en la construcción de la matriz de ponderaciones, que se basa en una distancia combinada: una distancia espacial (con un ancho de banda  $h_1 = 10000$  m) y una distancia distribucional (con  $h_2 = 0,01$ ). El parámetro más sensible en este esquema es el ancho de banda global del kernel exponencial,  $h$ , que modula la influencia del vecindario en cada estimación. Para su calibración, se empleó un método de validación cruzada temporal (Temporal Leave-One-Out, TLOO, por sus siglas en inglés). El análisis TLOO reveló un comportamiento clave: a medida que  $h$  tiende a cero, el error cuadrático medio se minimiza, pero el modelo colapsa a una regresión localmente lineal con el vecino más cercano, lo que provoca un sobreajuste e ignora la información del resto de la vecindad espacial. Por lo tanto, se seleccionó un valor de  $h = 0,02$  como un equilibrio pragmático. Este valor es suficientemente pequeño para dar prioridad a los puntos más cercanos y con variaciones similares, pero lo suficientemente grande para evitar el sobreajuste y capturar una estructura de dependencia espacial más robusta.

El desafío central del modelo GTWR reside en alcanzar un balance óptimo: debe ser lo suficientemente flexible para capturar las variaciones locales de la orografía del terreno, pero evitando sobreajuste al ruido o a las particularidades de un único punto predictor. La fortaleza del método radica precisamente en ponderar un vecindario de puntos con distribuciones estadísticas similares, lo que le brinda robustez. Sin embargo, es previsible que este balance sea más difícil de lograr en vecindades de alta complejidad topográfica. Para evaluar sistemáticamente este desempeño, el análisis de los resultados se divide en dos ramas: un análisis espacial, para evaluar la capacidad del modelo de ajustar patrones geográficos en instantes de tiempo fijos, y un análisis temporal, para estudiar su habilidad de adaptar las trayectorias temporales en puntos específicos. Ambos análisis se realizarán sobre las zonas representativas de cada clúster (vecindad del representante de clase).

#### Análisis espacial

La capacidad de predecir eventos climáticos en la intensidad del viento del modelo GTWR está determinada por los procesos micro-climáticos y la orografía presente en el dominio espacial. En particular, si observamos los tiempos  $t \in \{20, 38\}$  alcanza un MSE mínimo de 0,42 y un  $R^2 = 0,97$ , aproximadamente, lo que indica que el modelo ajusta espacialmente con mayor precisión cuando se presentan brisas costeras intensas y vientos alisios. En las ventanas temporales cuando existen mayores zonas de homogeneidad, en  $t \in \{2, 56\}$ , el modelo muestra una leve baja en el rendimiento, con un MSE más alto de 0,66 y 0,54, respectivamente, y coeficientes de determinación de 0,84 y 0,76, respectivamente. Esto se debe a las características de las ponderaciones, que toman valores cercanos a la unidad y generan valores promedios muy cercanos al valor original de los vecinos más cercanos.

El análisis espacial evalúa la capacidad del modelo para reconstruir la estructura geográfica del campo de viento en cuatro instantes temporales representativos de un ciclo diurno, permitiendo observar su respuesta a diferentes cambios microclimáticos.

- $t = 2$  (Madrugada, ver figura 4.10): En condiciones de calma relativa, el modelo reconstruye correctamente las brisas costeras. Sin embargo, muestra sus primeras limitaciones: subestima la intensidad en el farellón costero y sobreestima el viento en valles interiores, donde la orografía compleja puede generar circulaciones locales que el modelo tiende a suavizar.

- $t = 20$  (Ocaso, ver figura 4.11): Durante la transición vespertina, el modelo captura con gran precisión el gradiente de intensidad entre la costa y el interior, así como el aumento del viento en la bahía de Cartagena, un fenómeno de convección característico del ocaso.
- $t = 38$  (Mediodía, ver figura 4.12): Cerca del máximo de radiación solar y con patrones de viento más definidos, el modelo exhibe su mejor rendimiento. Se adapta correctamente al patrón de viento en el farellón costero (sector de curaumilla) y reproduce el aumento significativo de la intensidad en las cimas de la cordillera de la Costa, demostrando su capacidad para modelar efectos orográficos a gran escala.
- $t = 56$  (Amanecer, ver figura 4.13): En la transición matutina, con menor heterogeneidad espacial general, el comportamiento del modelo evidencia una tendencia sistemática al suavizado: tiende a subestimar la intensidad en las cimas de montaña y sobreestimar en zonas bajas. Este comportamiento es una consecuencia directa del kernel de ponderación, que promedia las influencias del vecindario, atenuando los valores extremos.

El análisis cuantitativo, se calculan las métricas espaciales (Sección 2.6.1) para los instantes de tiempo mencionados anteriormente, ver Cuadro 4.2, corrobora estas observaciones. El modelo alcanza su máxima precisión ( $MSE \approx 0.42$ ,  $R^2 \approx 0.97$ ) durante los períodos con patrones de viento bien establecidos ( $t = 20$  y  $t = 38$ ), como las brisas costeras intensas y los vientos alisios. En los períodos de transición ( $t = 2$  y  $t = 56$ ), donde dominan circulaciones más débiles y complejas, el rendimiento disminuye ligeramente ( $MSE \approx 0.66$  y  $0.54$ , respectivamente), aunque se mantiene en niveles altos.

t	$MSE_t$	$RMSE_t$	$R_t^2$
2	0,66459	0,81522	0,8380
20	0,42510	0,65200	0,9717
38	0,42208	0,64968	0,9567
56	0,53996	0,73482	0,7607

Cuadro 4.2: Métricas de validación espacial para el  $t = \{2, 20, 38, 56\}$  fijo, de la estimación del modelo GTWR.

Finalizando con el análisis de residuos, ver Figuras 4.21 y 4.22. Por un lado, indican que en los tiempos  $t \in \{2, 20\}$  (02/09/2023 02:00:00 hrs y 02/09/2023 20:00:00 hrs) las ponderaciones del modelo son más suaves en las zonas de máxima variabilidad, con una diferencia de hasta  $5 \text{ m/s}$  por debajo y  $4 \text{ m/s}$  sobre los datos originales para estos tiempos. Por otro lado, cuando el rango intercuartil está negativamente desplazado, en los tiempos  $t \in \{38, 56\}$  (03/09/2023 14:00:00 hrs y 04/09/2023 08:00:00 hrs), se explica que el modelo tiende a sobre-ajustar más en zonas con una topografía variable, pues hay una diferencia de hasta  $5 \text{ m/s}$  por debajo y sobre los datos originales, para  $t = 38$ ; y una diferencia de hasta  $4,5 \text{ m/s}$  por debajo y  $3 \text{ m/s}$  por sobre los datos originales, para  $t = 56$ .

En resumen, el análisis espacial evidencia que el modelo GTWR posee una alta capacidad predictiva para reconstruir la estructura espacial del campo de viento, adaptándose correctamente a los patrones de viento bien definidos, como las brisas costeras y los efectos orográficos. No obstante, se observa un comportamiento sistemático de sub-ajuste en zonas de alta complejidad, el cual se deriva directamente de las ponderaciones (y el kernel empleado). Este efecto suaviza la intensidad del viento, lo que se traduce en una sub-estimación consistente en las cimas de montaña y una sobre-estimación en los valles y pies de montaña. Por lo tanto, se concluye que la capacidad predictiva espacial del modelo es robusta

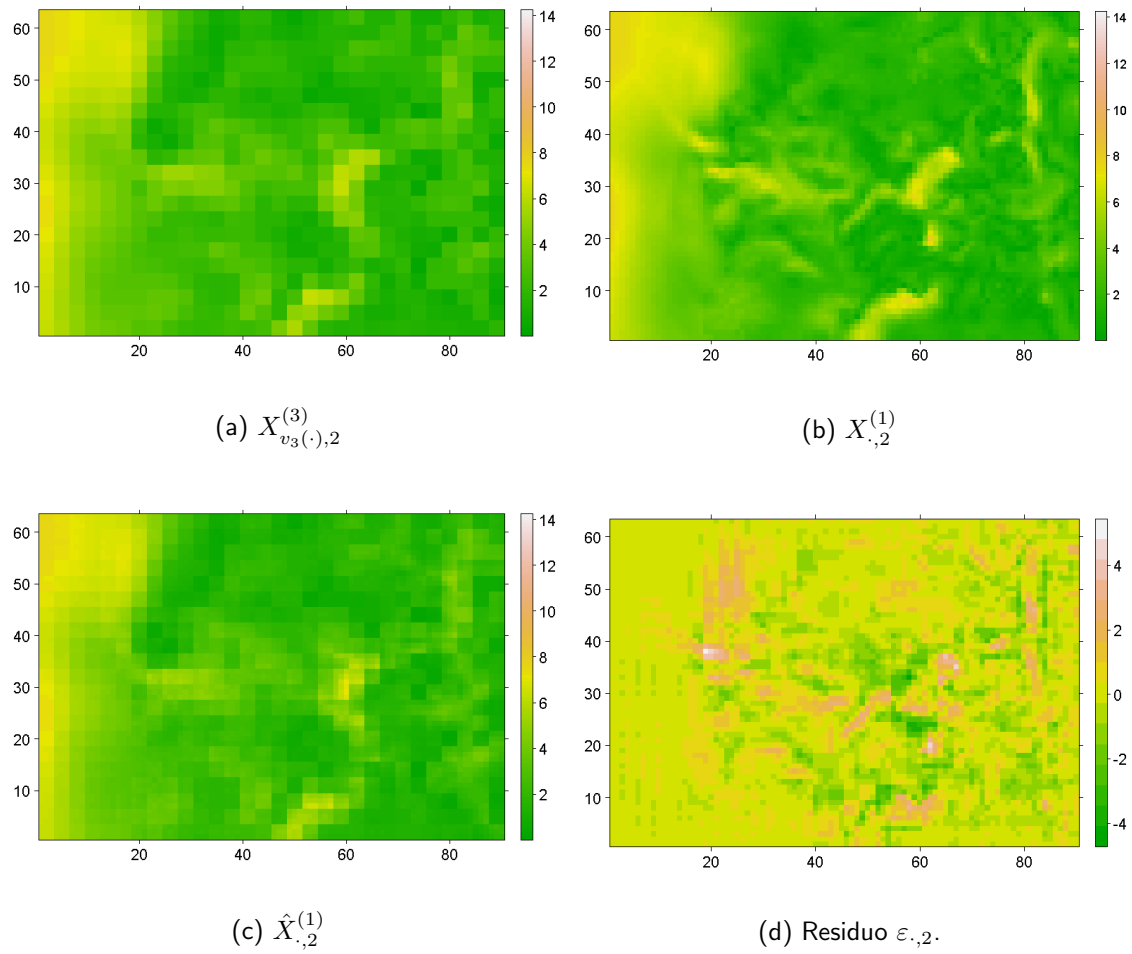


Figura 4.10: Resultados del modelo GTWR sobre las mallas espaciales  $M^{(3)}$  a  $M^{(1)}$  en el corte temporal  $t = 2$ .

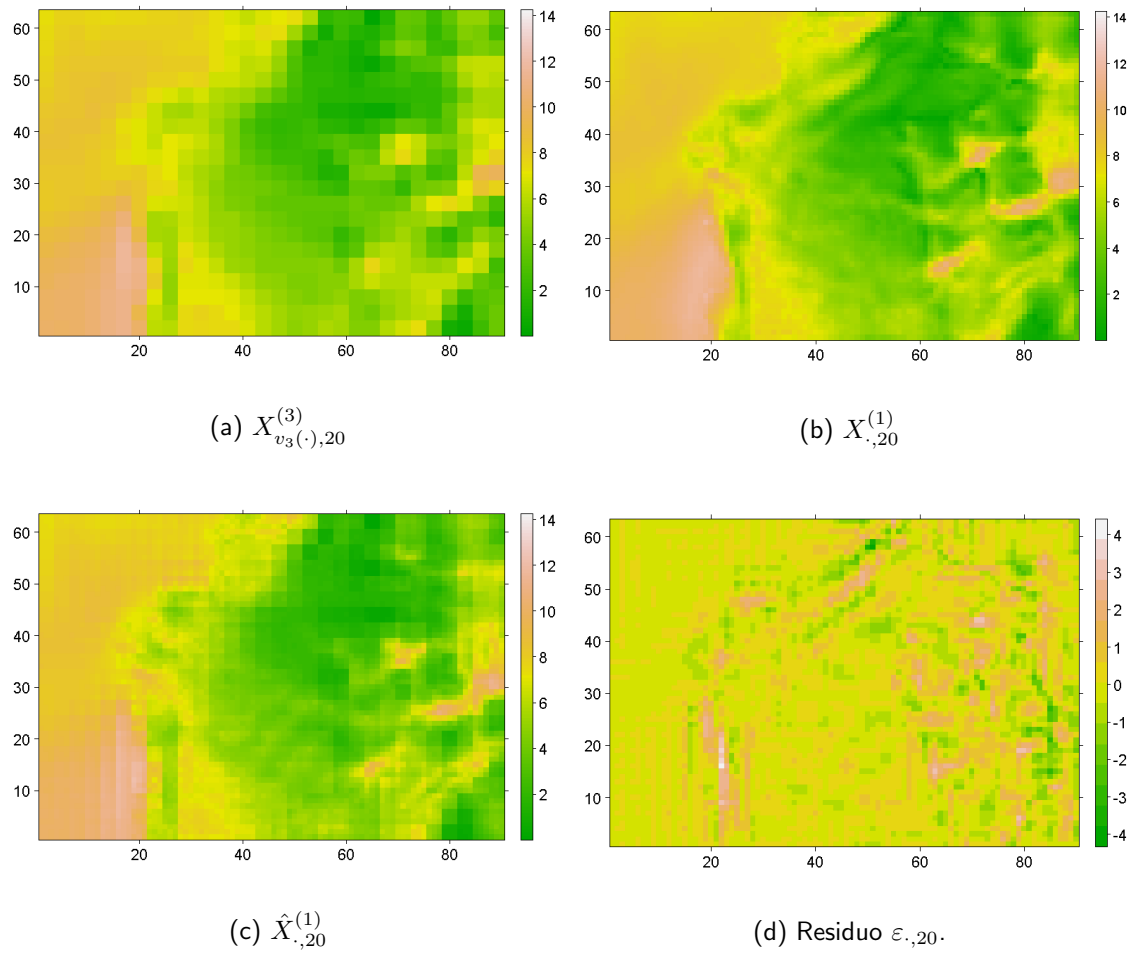


Figura 4.11: Resultados del modelo GTWR sobre las mallas espaciales  $M^{(3)}$  a  $M^{(1)}$  en el corte temporal  $t = 20$ .

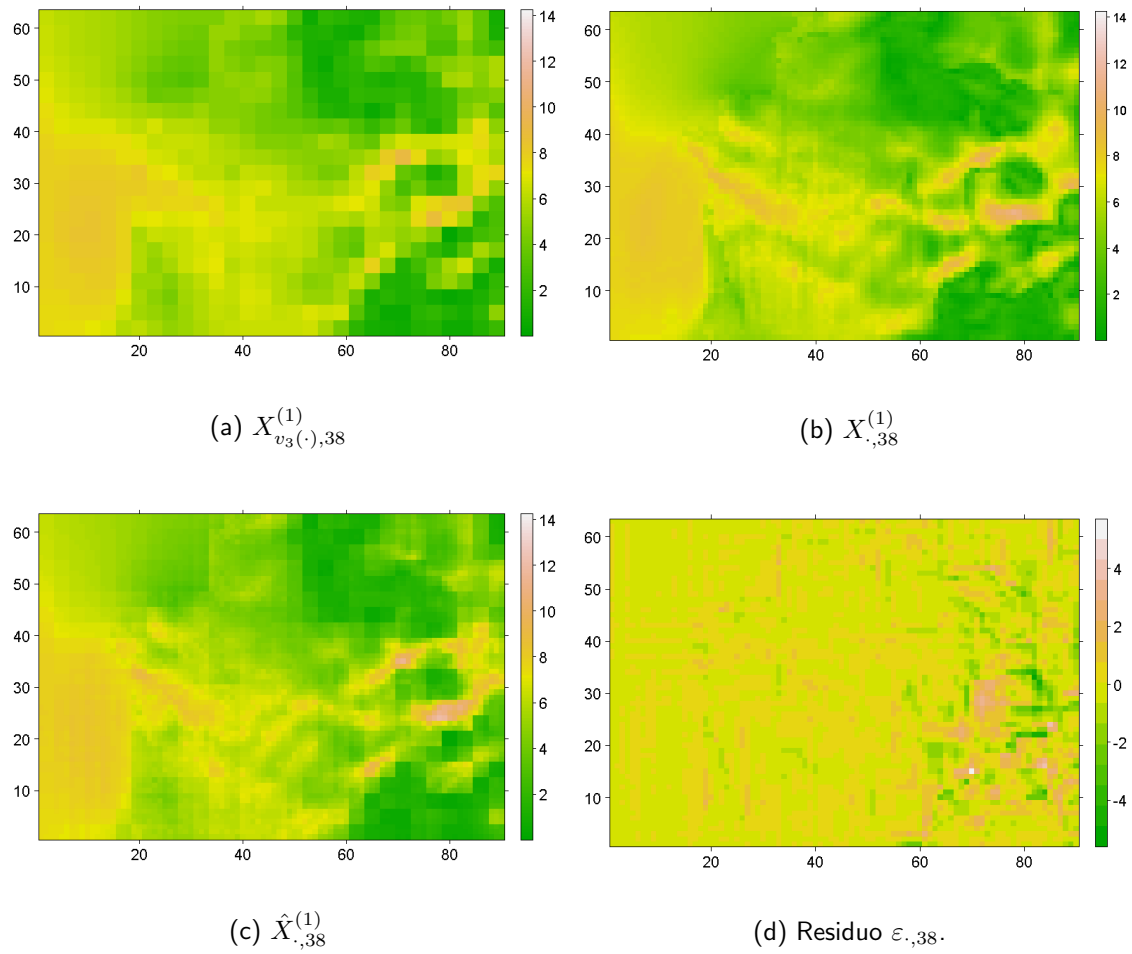


Figura 4.12: Resultados del modelo GTWR sobre las mallas espaciales  $M^{(3)}$  a  $M^{(1)}$  en el corte temporal  $t = 38$ .

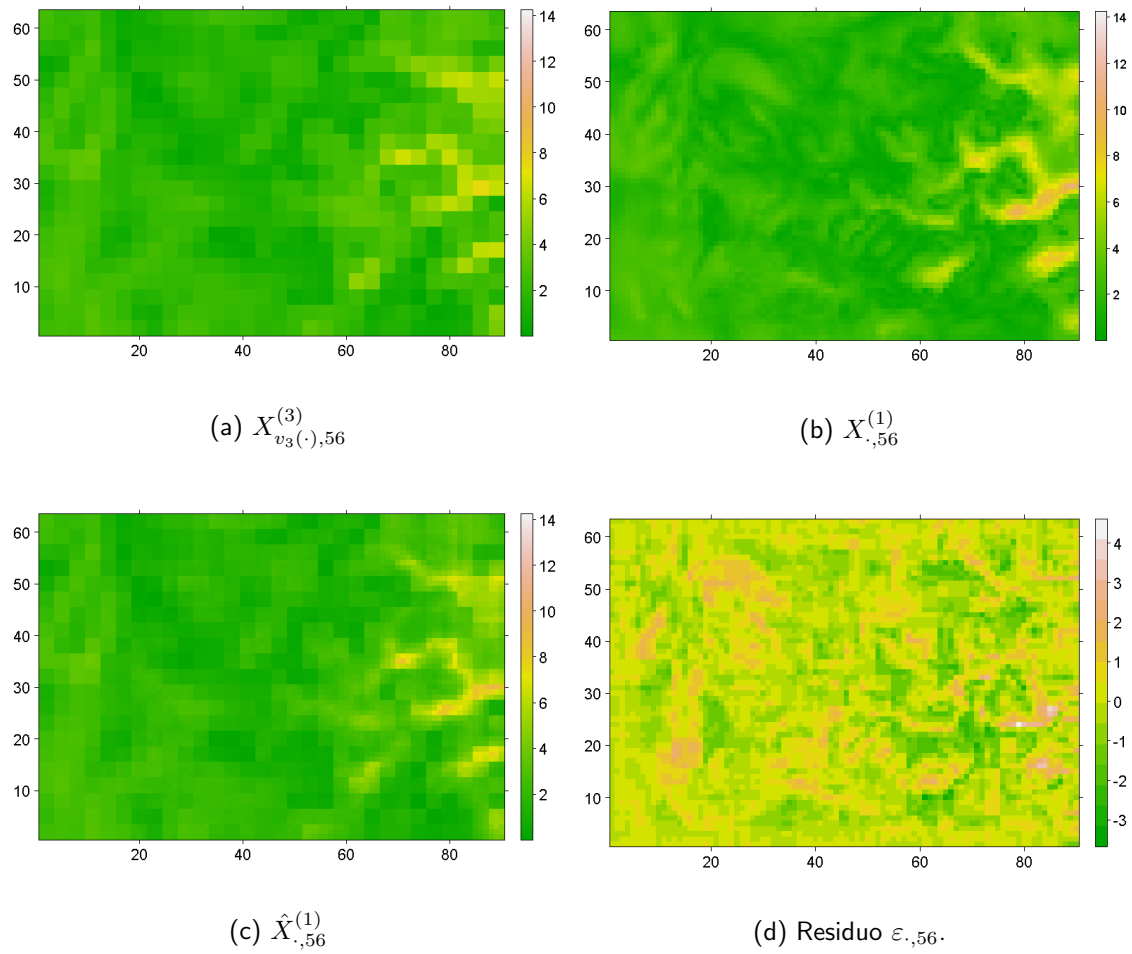
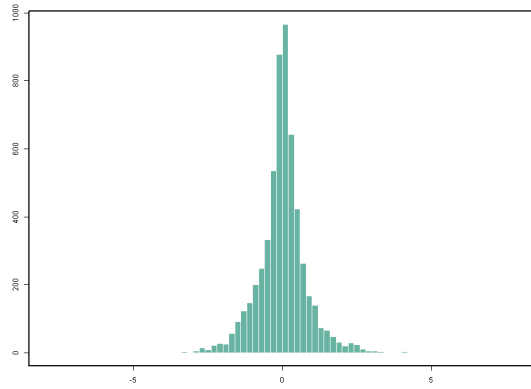
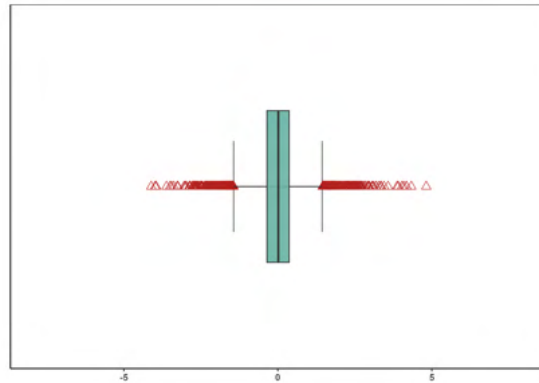


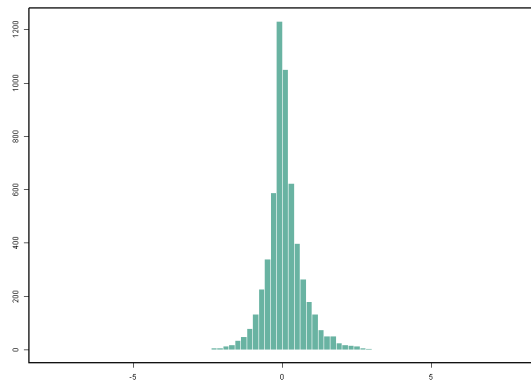
Figura 4.13: Resultados del modelo GTWR sobre las mallas espaciales  $M^{(3)}$  a  $M^{(1)}$  en el corte temporal  $t = 56$ .



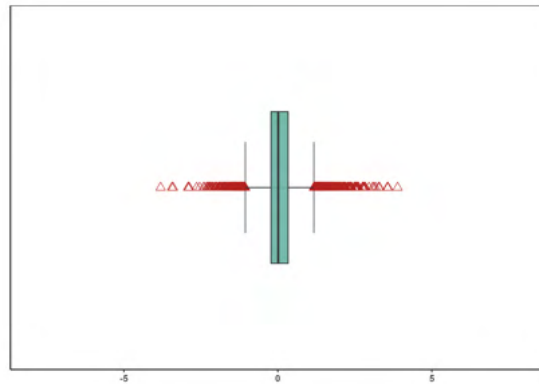
(a) Histograma de los residuos  $\varepsilon_{\cdot,2}$ .



(b) Boxplot de los residuos  $\varepsilon_{\cdot,2}$ .



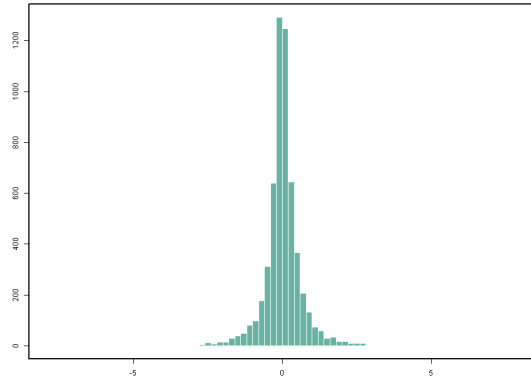
(c) Histograma de los residuos  $\varepsilon_{\cdot,20}$ .



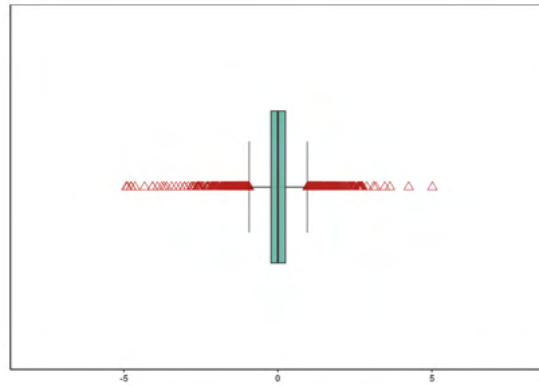
(d) Boxplot de los residuos  $\varepsilon_{\cdot,20}$ .

Figura 4.14: Visualización de la distribución de los residuos espaciales con su histograma y boxplot en los tiempos  $t = 2$ , (a,b) y  $t = 20$ , (c,d), respectivamente.

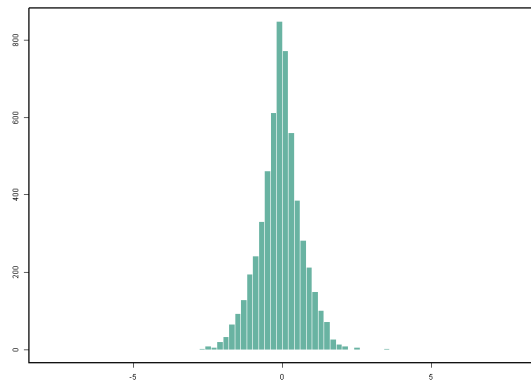
y su aplicación es adecuada para gran parte del dominio, pero su efectividad se ve condicionada por la heterogeneidad topográfica global, concentrando los mayores errores en las zonas de alta variabilidad que componen el clúster de montaña.



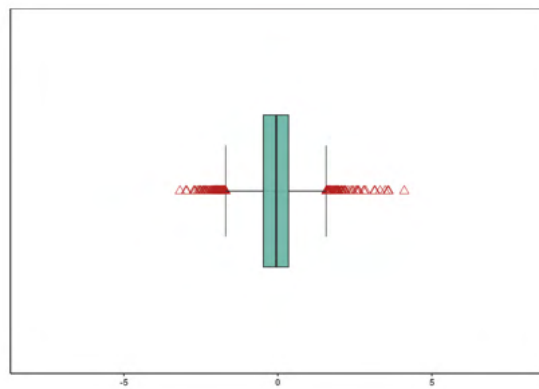
(a) Histograma de los residuos  $\varepsilon_{.,38}$ .



(b) Boxplot de los residuos  $\varepsilon_{.,38}$ .



(c) Histograma de los residuos  $\varepsilon_{.,56}$ .



(d) Boxplot de los residuos  $\varepsilon_{.,56}$ .

Figura 4.15: Visualización de la distribución de los residuos espaciales con su histograma y boxplot en los tiempos  $t = 38$ , (a,b) y  $t = 56$ , (c,d), respectivamente.

## Análisis temporal

El análisis temporal se enfoca en estudiar el comportamiento de las series de tiempo observadas versus las estimadas, con el fin de identificar la capacidad predictiva del modelo ante los patrones temporales que caracterizan a cada zona topográfica. Para esto, se compara el desempeño del modelo en las vecindades representativas de mar, valle y montaña.

- Vecindad de Mar (ver Figura 4.16): El comportamiento de las series estimadas es muy similar al de las trayectorias predictoras, generando leves sobre-estimaciones y sub-estimaciones en momentos puntuales. Este resultado es consistente con la teoría del modelo: dado que la variación espacial es mínima en esta zona, los pesos del GTWR tienden a la unidad, resultando en una predicción que suaviza la serie original.
- Vecindad de Valle (ver Figura 4.17): Se evidencia una mayor capacidad de adaptación. Mientras algunos puntos de la vecindad siguen un comportamiento similar al predictor, otros muestran un ajuste significativamente mejorado a las series originales a 1 km, particularmente entre las 25 y 35 horas. Esto indica que el modelo comienza a capturar fenómenos micro-climáticos que diferencian el comportamiento del viento dentro de la misma vecindad.
- Vecindad de Montaña (ver Figura 4.18): Como se esperaba, esta zona presenta el mayor desafío. El modelo muestra dificultades para interpretar las subidas abruptas de intensidad del viento, especialmente en las primeras 15 horas. No obstante, en otros tramos evidencia un ajuste importante a las variaciones, aunque hacia el final de la serie tiende a un comportamiento promedio. Esto sugiere que el modelo puede adaptarse a la alta variabilidad, pero no de forma consistente a lo largo de toda la trayectoria temporal.

Para cuantificar estas observaciones, se aplicaron las métricas de evaluación temporal (Sección 2.6.2) a las series predichas y observadas en cada vecindad de cluster.

- Vecindad de Mar (ver Cuadro 4.3): El ajuste del modelo es excelente, presentando errores bajos (MSE y RMSE) y una alta precisión ( $R^2 > 0,9$ , en general). Este resultado se traduce en una alta estabilidad temporal y es consistente con la homogeneidad espacial de la zona marítima.
- Vecindad de Valle (ver Cuadro 4.4): Se presentan errores moderados y una precisión que se mantiene alta ( $R^2 \in (0,8; 0,9)$ ). Esto da cuenta de que la capacidad predictiva del modelo sigue siendo robusta, incluso ante la aparición de fenómenos micro-climáticos que determinan un cambio en la evolución temporal.
- Vecindad de Montaña (ver Cuadro 4.5): Las métricas confirman las dificultades visualizadas. Se observan errores altos ( $MSE \geq 1$ ) y una precisión predictiva que decae a un nivel de moderada a baja ( $R^2 \in (0,2; 0,7)$ ). Este comportamiento se atribuye directamente al aumento de la heterogeneidad espacial, asociada a los fenómenos micro-climáticos y a los abruptos cambios de topografía local.

Un supuesto clave para la validez del modelo es la estacionariedad de los residuos. El análisis visual de sus trayectorias (Figura 4.19) muestra que en las zonas de mar y valle, las series de los residuos son centradas y presentan un comportamiento en torno a un valor fijo, lo que da indicios de estacionariedad. No obstante, el comportamiento en la vecindad de montaña es más errático, lo que hace necesario aplicar un test robusto para encontrar evidencias estadísticas concluyentes.

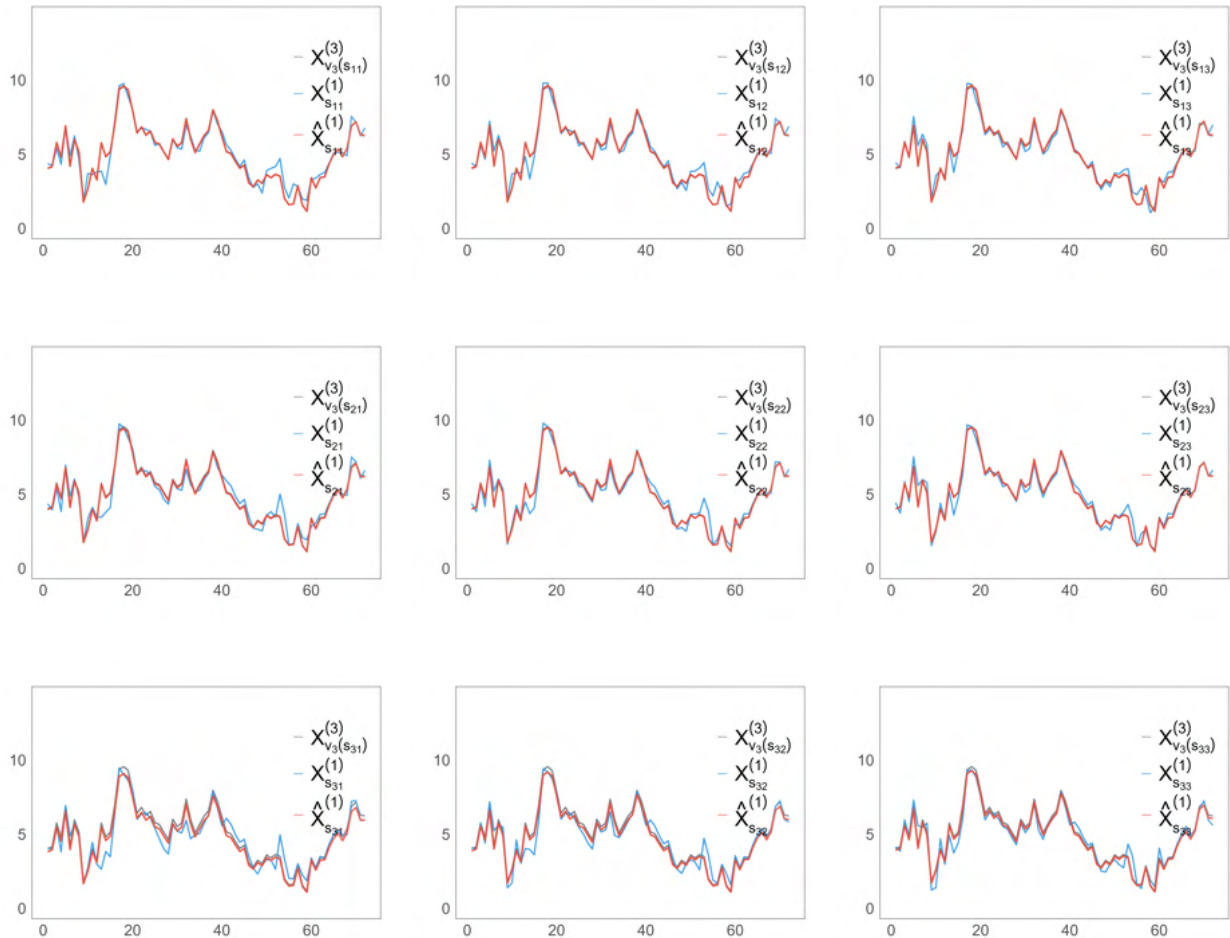


Figura 4.16: Comparación temporal del series de intensidad del viento reales contra las predicciones del modelo GTWR en la vecindad local representante de mar  $V_{s_{22}} \subset M^{(1)}$ .

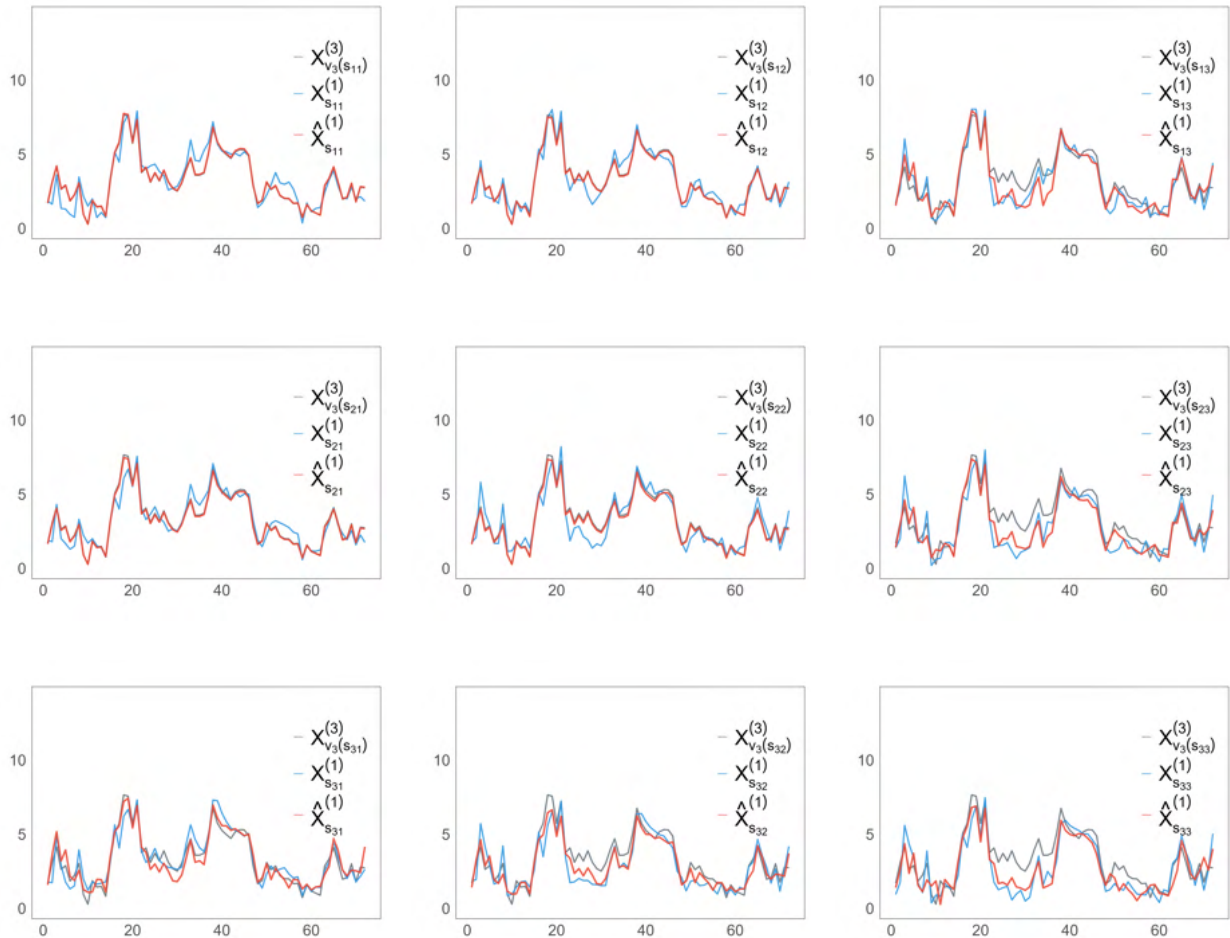


Figura 4.17: Comparación temporal del series de intensidad del viento reales contra las predicciones del modelo GTWR en la vecindad local representante de valle  $V_{s_{22}} \subset M^{(1)}$ .

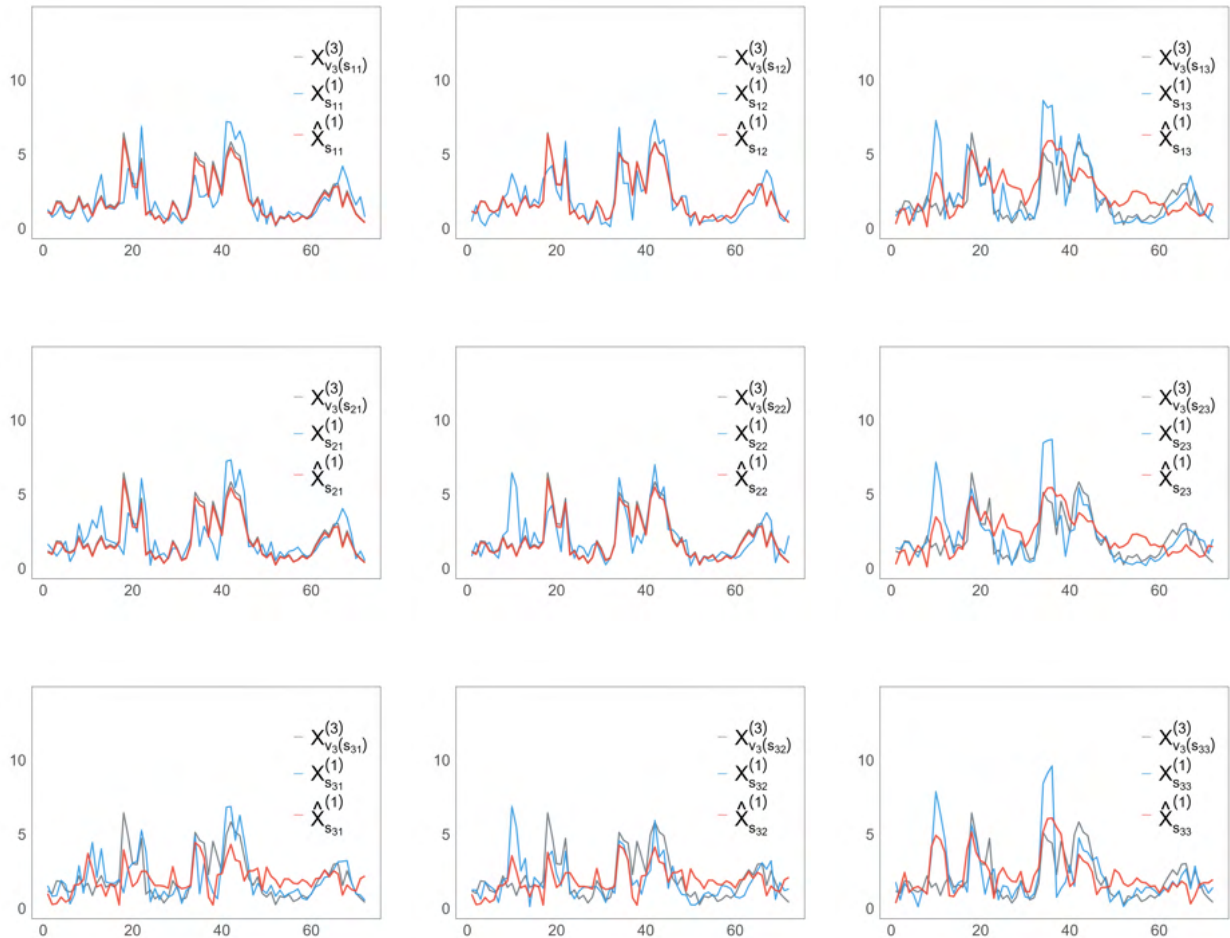
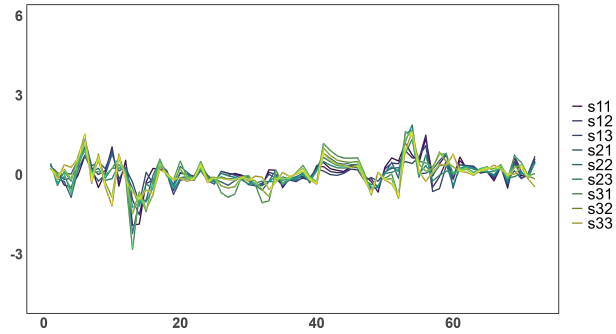
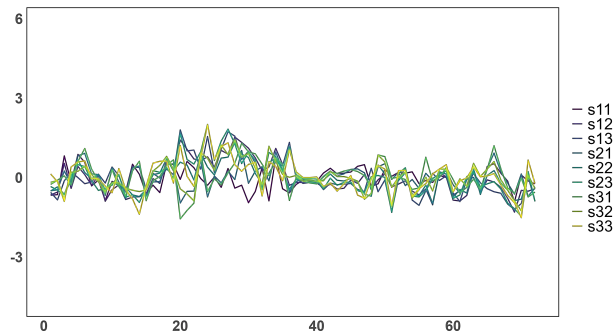


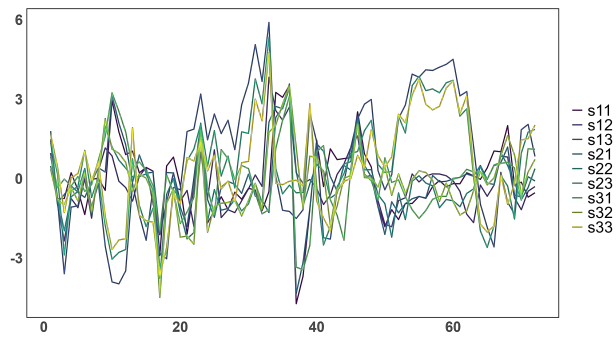
Figura 4.18: Comparación temporal del series de intensidad del viento reales contra las predicciones del modelo GTWR en la vecindad local representante de montaña  $V_{s_{22}} \subset M^{(1)}$ .



(a) Series de residuos en vecindad de mar



(b) Series de residuos en vecindad de valle



(c) Series de residuos en vecindad de montaña

Figura 4.19: Gráficas de las series de tiempo de los residuos en vecindades locales del campo de intensidad de viento de mar (a), de valle (b) y de montaña (c), respectivamente.

$s$	$MSE_s$	$RMSE_s$	$R_s^2$
$s_{11}$	0,27314	0,52263	0,9129
$s_{12}$	0,19796	0,44493	0,9373
$s_{13}$	0,15366	0,39199	0,9525
$s_{21}$	0,29767	0,54559	0,9066
$s_{22}$	0,21661	0,46541	0,9327
$s_{23}$	0,16228	0,40284	0,9503
$s_{31}$	0,42447	0,65151	0,8606
$s_{32}$	0,29730	0,54525	0,9046
$s_{33}$	0,23524	0,48502	0,9284

Cuadro 4.3: Métricas temporales por punto espacial  $s$  en la vecindad de mar.

$s$	$MSE_s$	$RMSE_s$	$R_s^2$
$s_{11}$	0,46645	0,68297	0,8525
$s_{12}$	0,30657	0,55369	0,8956
$s_{13}$	0,39090	0,62522	0,8838
$s_{21}$	0,38098	0,61723	0,8512
$s_{22}$	0,52403	0,72390	0,8244
$s_{23}$	0,43165	0,65700	0,8681
$s_{31}$	0,51710	0,71910	0,8147
$s_{32}$	0,39396	0,62766	0,8555
$s_{33}$	0,54335	0,73712	0,8366

Cuadro 4.4: Métricas temporales por punto espacial  $s$  en la vecindad de valle.

$s$	$MSE_s$	$RMSE_s$	$R_s^2$
$s_{11}$	1,37643	1,17321	0,5082
$s_{12}$	0,85631	0,92537	0,7009
$s_{13}$	2,13879	1,46246	0,5192
$s_{21}$	1,57930	1,25670	0,3824
$s_{22}$	1,30455	1,14217	0,4721
$s_{23}$	2,18046	1,47664	0,4203
$s_{31}$	1,83351	1,35407	0,2042
$s_{32}$	1,15216	1,07339	0,3681
$s_{33}$	2,04020	1,42836	0,5009

Cuadro 4.5: Métricas temporales por punto espacial  $s$  en la vecindad de montaña.

Para este fin, se utilizó el test de Phillips-Perron (PP-test), una prueba estadística que evalúa la hipótesis nula de existencia de una raíz unitaria frente a la hipótesis alternativa de estacionariedad [21]. Se eligió este test por su robustez ante la heterocedasticidad presente en los datos [4]. Los resultados, presentados en los Cuadros 4.6, 4.7 y 4.8, muestran que para todas las series de residuos en las tres vecindades, el valor del *PP-estadístico* es menor al valor crítico con un nivel de significancia del 5 %.

Por lo tanto, se rechaza la hipótesis nula, lo que da un fuerte refuerzo estadístico a la suposición de estacionariedad de los residuos temporales. Este hallazgo es fundamental, pues confirma que, a pesar de las dificultades predictivas en la zona de montaña, los residuos del modelo mantienen un comportamiento

estadísticamente estable en todo el dominio de estudio.

	$s_{11}$	$s_{12}$	$s_{13}$	$s_{21}$	$s_{22}$	$s_{23}$	$s_{31}$	$s_{32}$	$s_{33}$
<i>PP-estadístico</i>	-6,260	-6,471	-6,061	-5,832	-5,197	-6,050	-5,500	-5,656	-6,481
<i>Valor crítico al 5 %</i>	-2,902	-2,902	-2,902	-2,902	-2,902	-2,902	-2,902	-2,902	-2,902

Cuadro 4.6: Resultados del *PP-test* para vecindad Mar.

	$s_{11}$	$s_{12}$	$s_{13}$	$s_{21}$	$s_{22}$	$s_{23}$	$s_{31}$	$s_{32}$	$s_{33}$
<i>PP-estadístico</i>	-8,117	-3,486	-4,271	-8,203	-4,790	-4,095	-6,059	-5,170	-5,866
<i>Valor crítico al 5 %</i>	-2,902	-2,902	-2,902	-2,902	-2,902	-2,902	-2,902	-2,902	-2,902

Cuadro 4.7: Resultados del *PP-test* para vecindad Valle.

	$s_{11}$	$s_{12}$	$s_{13}$	$s_{21}$	$s_{22}$	$s_{23}$	$s_{31}$	$s_{32}$	$s_{33}$
<i>PP-estadístico</i>	-5,800	-5,092	-3,205	-5,983	-5,127	-3,612	-5,033	-5,243	-3,893
<i>Valor crítico al 5 %</i>	-2,902	-2,902	-2,902	-2,902	-2,902	-2,902	-2,902	-2,902	-2,902

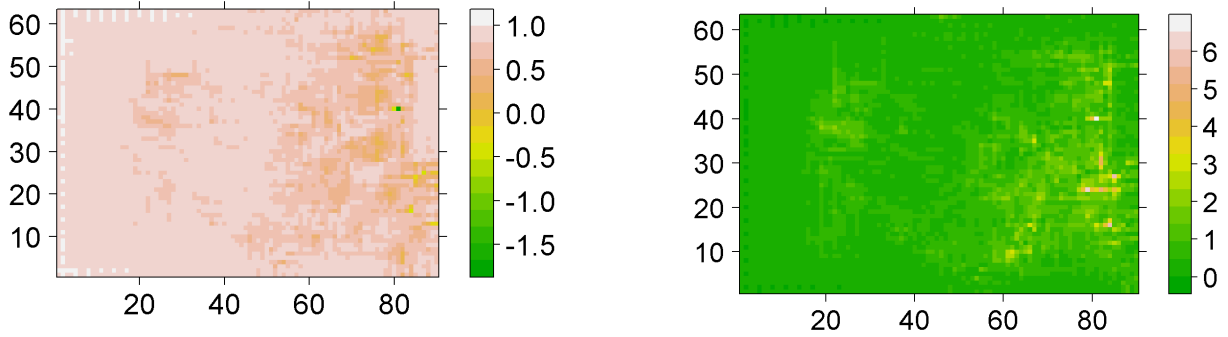
Cuadro 4.8: Resultados del *PP-test* para vecindad Montaña.

El comportamiento temporal de las estimaciones locales está determinado por la construcción de los factores de escala que dependen de la distancia espacial y distancia distribucional variacional entre las series respecto al punto de interés. Esta posee la característica de distinguir estadísticamente entre series que son similares, pero presenta dificultades para ajustar patrones temporales caracterizados por los cambios del ciclo diario de calentamiento de la superficie terrestre y fenómenos meteorológicos entrantes. También, las series de tiempo en zonas que presentan una mayor complejidad topográfica, zonas de montaña, concentran los MSE más altos y los coeficientes de determinación más bajos (incluso negativos en puntos específicos), ver Figura 4.20.

En conclusión, el análisis temporal revela que la capacidad predictiva del modelo está directamente determinada por la heterogeneidad topográfica de cada zona. En las vecindades de mar y valle, donde la homogeneidad espacial es mayor, el modelo GTWR muestra un ajuste robusto y métricas de alta precisión. Sin embargo, en la vecindad de montaña, el modelo presenta dificultades para adaptar su comportamiento a las trayectorias temporales, especialmente ante los cambios abruptos en la intensidad del viento. Este comportamiento se deriva de la construcción de los factores de escala, que si bien distinguen estadísticamente entre series, tienden a suavizar los patrones temporales más complejos. Como consecuencia, las zonas de mayor complejidad topográfica concentran los errores cuadráticos medios más altos y los coeficientes de determinación más bajos. A pesar de esto, el hallazgo fundamental de la estacionariedad de los residuos en todas las zonas otorga validez estadística al modelo, sugiriendo que sus limitaciones son de precisión predictiva y no de estabilidad estructural.

### 4.3.3. Predicción del campo 1/3 km

Concluida la etapa de estimación y validación del modelo GTWR, continúa el paso final de predicción del campo de viento a resolución 1/3 de km. Para esto, tomaremos el campo de viento a 1 km y los factores de escala estimados en la sección anterior relacionándolos con cada punto espacial de la malla a



(a) Campo espacial de los  $R_i^2$ , para todo  $i = 1, \dots, n_1$ . (b) Campo espacial del  $MSE_i$ , para todo  $i = 1, \dots, n_1$ .

Figura 4.20: Campos de métricas del coeficiente de determinación espacial (a) y del error cuadrático medio espacial (b), entre series de tiempo observadas y estimadas en la malla espacial  $M^{(1)}$ .

1/3 km con el vecino más cercano, es decir, tomamos  $X_{v_1(i)}^{(1)}$  y  $\beta_i^{(1)}$  como los factores de escala ajustados a la dimensión de la malla 1/3 km (por el vecino más cercano), para cada  $i = 1, \dots, n_{1/3}$ , tomando la ecuación (3.2), se tiene la relación  $\hat{X}_i^{(1/3)} = \hat{\beta}_i^{(1)} X_{v_1(i)}^{(1)}$ .

Los resultados de este esquema de predicción se presentan en las Figuras 4.21 y 4.22 para los instantes de tiempo  $t = \{2, 20, 38, 56\}$ . Al comparar el campo predictor de 1 km con el campo predicho de 1/3 km, se observa un efecto principal: un aumento considerable en la intensidad del viento. Sin embargo, este efecto no es uniforme, sino que se concentra selectivamente en las zonas de alta complejidad topográfica. Se evidencia así que el factor de escala recuperado por el modelo GTWR se adapta a la orografía, generando un cambio significativo en la resolución espacial de la malla objetivo. No obstante, en las zonas de mayor homogeneidad espacial (mar y valle), el modelo realiza solo leves ajustes, un comportamiento que se deriva de las características de las ponderaciones consideradas y de las distancias espacio-distribucionales seleccionada.

Dada la estructura de las mallas  $M^{(3)}$ ,  $M^{(1)}$  y  $M^{(1/3)}$ , los clusters comparten puntos entre estas considerando la proporción 3:1. Consideramos un punto por cluster compartido entre las tres mallas, para comparar el comportamiento de la intensidad del viento en el tiempo. En la Figura 4.23, se observa el comportamiento de las series predichas por el esquema de downscaling estadístico, respecto a las series originales, en donde al representante de la vecindad de mar y valle mantiene un comportamiento similar de aumentos y disminuciones en la intensidad de viento, salvo instantes temporales, sin embargo, para el representante de montaña, el esquema aumenta significativamente la intensidad de viento respecto a la estimaciones del modelo WRF, alcanzando intensidades de 17 m/s en cerca del tiempo  $t = 35$  (máximo global de 20 m/s aproximadamente).

Por ende, se concluye que el impacto del esquema de downscaling se ajusta de forma más acentuada en las zonas de mayor complejidad topográfica. Este efecto es resultante de la capacidad del modelo para interpretar la orografía local y traducir los fenómenos micro-climáticos asociados en un aumento

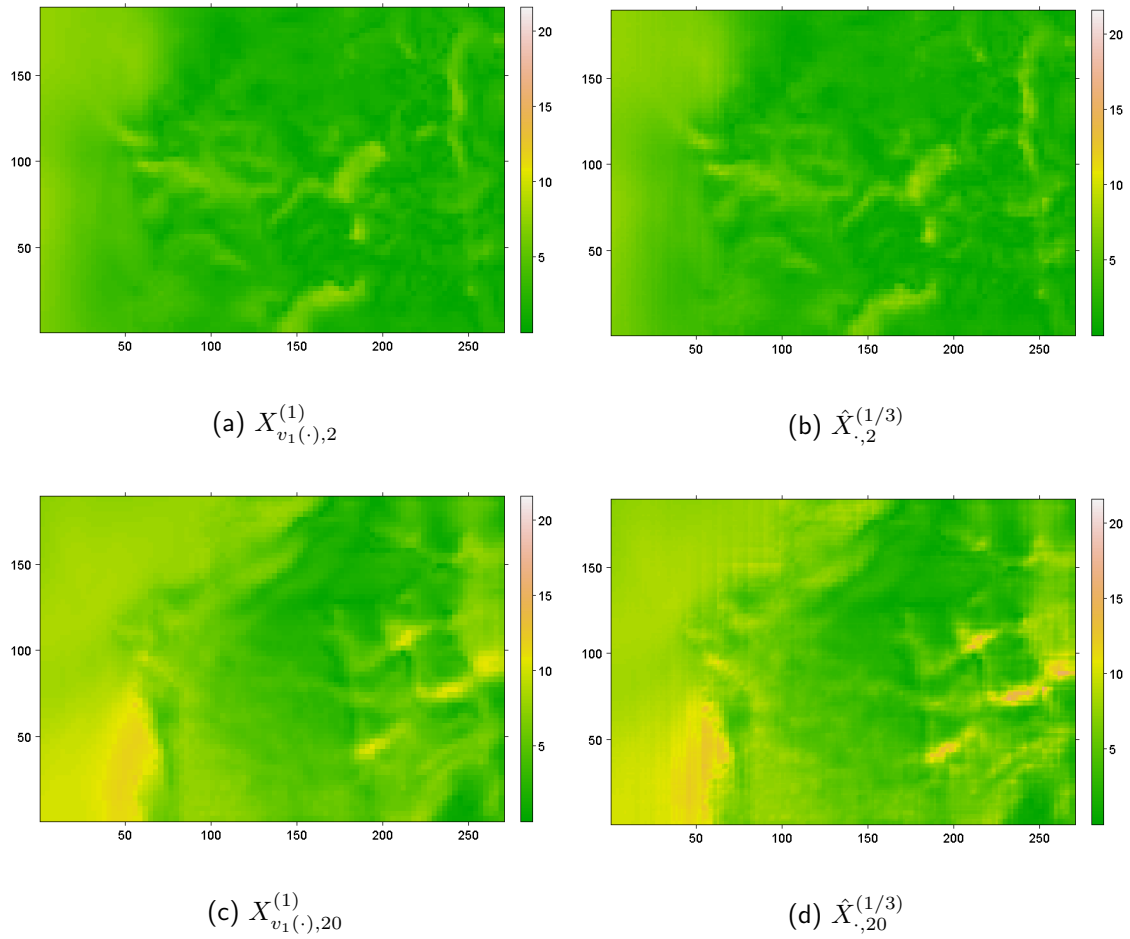


Figura 4.21: Resultado del método de downscaling a 1/3 km. Se muestra el campo  $X_{v_1(\cdot),t}^{(1)}$  y el campo predicho  $\hat{X}_{\cdot,t}^{(1/3)}$ , en los cortes temporales  $t = 2$  (a,b) y  $t = 20$  (c,d), respectivamente.

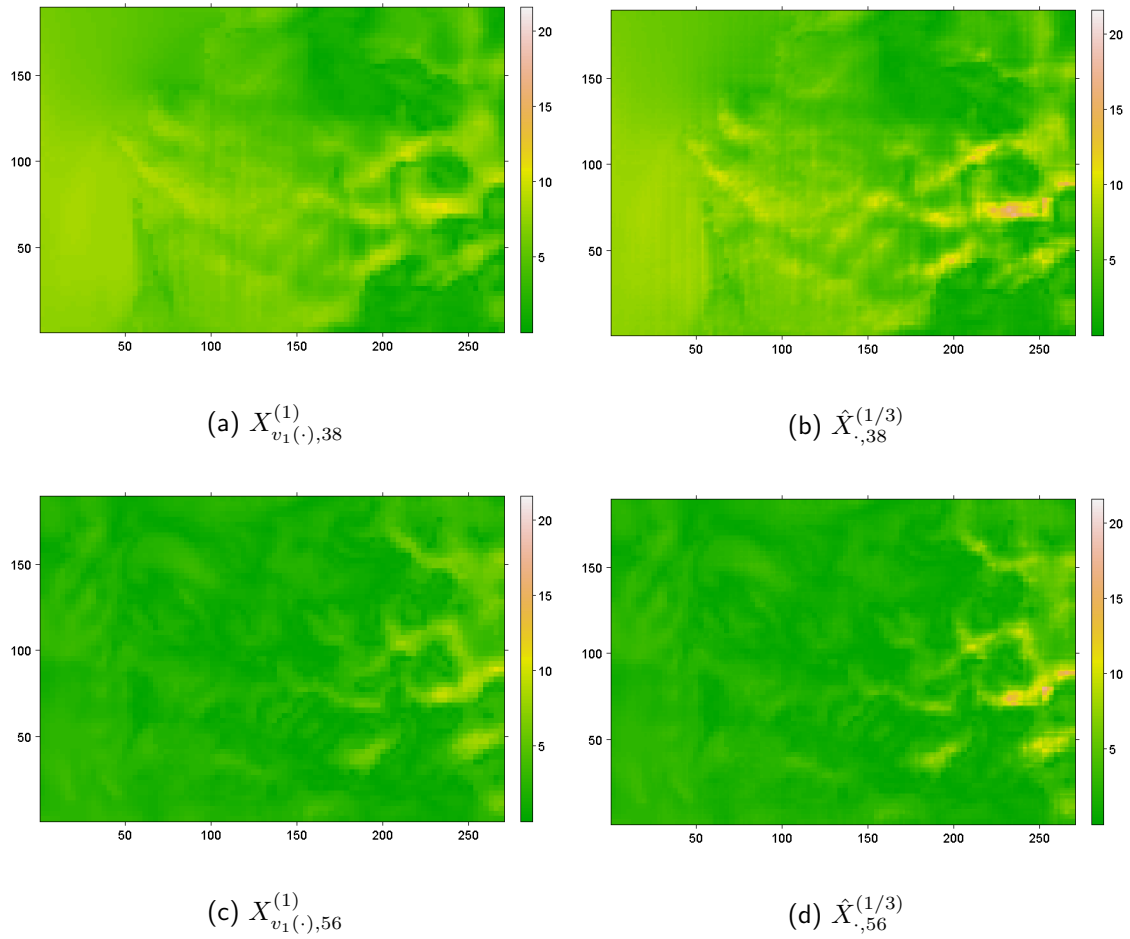
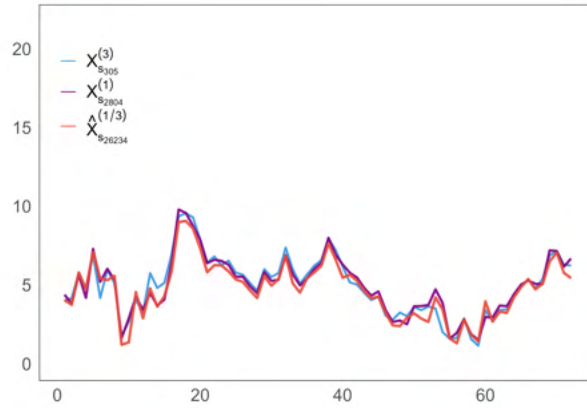
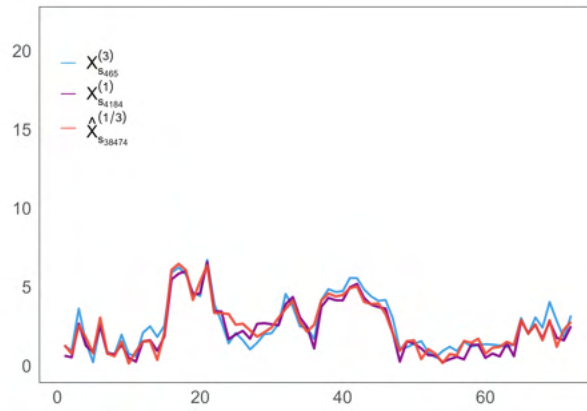


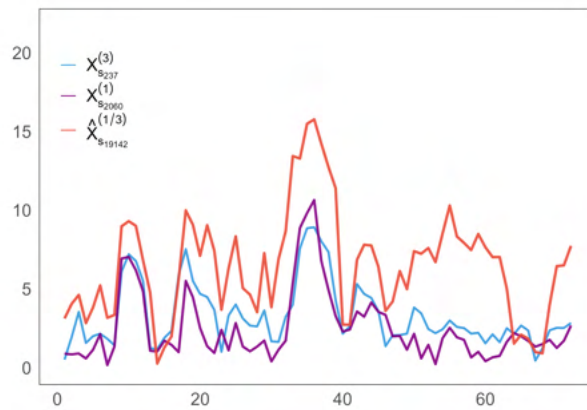
Figura 4.22: Resultado del método de downscaling a 1/3 km. Se muestra el campo original  $X_{v_1(\cdot),t}^{(1)}$  y el campo predicho  $\hat{X}_{\cdot,t}^{(1/3)}$ , en los cortes temporales  $t = 38$ , (a,b) y  $t = 56$ , (c,d), respectivamente.



(a) Serie del representante de mar.



(b) Serie del representante de valle.



(c) Serie del representante de montaña.

Figura 4.23: Comparación temporal del series de intensidad del viento en las vecindades de mar (a), valle (b) y montaña (c).

de la intensidad del viento al incrementar la resolución. No obstante, en las zonas de mayor homogeneidad local, como el mar y el valle, el modelo identifica leves cambios. Estos ajustes sutiles son igualmente importantes, pues representan la captura de fenómenos micro-climáticos de menor escala (como variaciones en las brisas costeras o vientos de valle) que los modelos de menor resolución no logran discernir.

## Capítulo 5

# Conclusiones

El desafío central de este trabajo fue construir e implementar un método de downscaling estadístico capaz de identificar y modelar las estructuras de dependencia espacio-temporal en campos de viento. Para este fin, se seleccionó y aplicó un esquema de downscaling estadístico a relación  $\lambda = 1/3$  sobre un modelo de Regresión Ponderada Espacio-Temporal (GTWR) para mallas anidadas de 3 km y 1 km obtenidas desde un modelo dinámico Weather Research and Forecasting (WRF). Los resultados confirman que el modelo GTWR es una herramienta robusta ante fenómenos orográficos para estimar campos de intensidad del viento a resolución fina, cuyo desempeño está fuertemente condicionado por la heterogeneidad topográfica del dominio.

El estudio de simulación validó la robustez del modelo para interpretar relaciones localmente lineales. Al introducir ruido blanco con variabilidad creciente, se observó que las diferencias entre los factores de escala reales y los estimados por el modelo eran prácticamente nulas. Las leves discrepancias se presentaron únicamente en el escenario de máxima variabilidad y se concentraron en las zonas de extrema heterogeneidad espacial, confirmando la alta capacidad del modelo para recuperar los parámetros subyacentes en condiciones controladas.

En zonas de homogeneidad espacial (mar y valle), el modelo exhibió una alta capacidad predictiva. Sin embargo, en zonas de alta complejidad topográfica (montaña), su desempeño fue moderado, presentando dificultades para seguir con precisión las trayectorias temporales más abruptas. En síntesis, el esquema final demuestra que el modelo no solo refina la resolución, sino que interpreta la orografía al aplicar ajustes leves en zonas homogéneas y significativos en zonas complejas. Esta capacidad de adaptar el factor de escala según la topografía local constituye la principal contribución de la metodología propuesta.

### Impacto y limitaciones del estudio

El método de downscaling estadístico propuesto tiene un impacto directo en múltiples áreas. Contribuye al desarrollo de metodologías de downscaling a micro-escala, aporta evidencia sobre la aplicación de modelos localmente lineales a datos con dependencia espacio-temporal y promueve el uso de herramientas de modelamiento estadístico para generar mallas de alta resolución. Estas mallas son la fuente principal de aplicaciones en medio ambiente, la planificación de energías renovables y la prevención de incendios forestales, sobre todo en zonas de alta densidad poblacional y estructural.

La principal limitación del estudio radica en la precisión predictiva del modelo en zonas de alta variabilidad topográfica y en el alcance limitado del dominio espacio-temporal analizado, lo que abre

vías para trabajos futuros.

## Recomendaciones y pasos futuros

Para extender y mejorar la robustez del esquema propuesto, se plantean las siguientes líneas de trabajo futuro, que abordan tanto el modelo como el diseño experimental:

1. Incorporar nuevas variables predictoras: Incluir Covariables como la dirección del viento, la temperatura o la humedad relativa, disponibles en el modelo WRF, podría mejorar significativamente la precisión de la estimación del GTWR. En particular, la dirección del viento permitiría determinar con mayor exactitud cómo la intensidad es afectada por fenómenos de convección y turbulencia, mejorando el ajuste temporal en las zonas topográficas complejas.
2. Refinar el modelo GTWR: Para mejorar el desempeño en zonas de alta variabilidad, se recomienda:
  - Implementar un enfoque por clúster: Aplicar y calibrar modelos GTWR independientes para cada zona topográfica (mar, valle, montaña).
  - Extender la distancia distribucional: Incorporar métricas de variación a diferentes escalas temporales (diaria, semanal) para capturar patrones estacionales.
  - Métodos de validación: Implementar métodos de validación cruzada más robustos que incorporen el factores espacio-temporal o trabajarlo por separado, para minimizar sesgos en la optimización del ancho de ventana del kernel (junto a la distancia) seleccionado.
3. Ampliar el diseño experimental:
  - Dominio espacio-temporal: Es crucial ampliar el dominio geográfico para incluir una mayor diversidad topográfica y extender el horizonte temporal de las simulaciones para identificar componentes de estacionalidad y tendencia.
  - Estudio de simulación: Realizar simulaciones más robustas, implementando campos gaussianos con estructuras de alta dependencia espacio-temporal, para evaluar el modelo en escenarios más realistas y desafiantes.

## Apéndice A

# Resultados Simulación

A.1.  $\sigma_{res}^2 = 0,005$

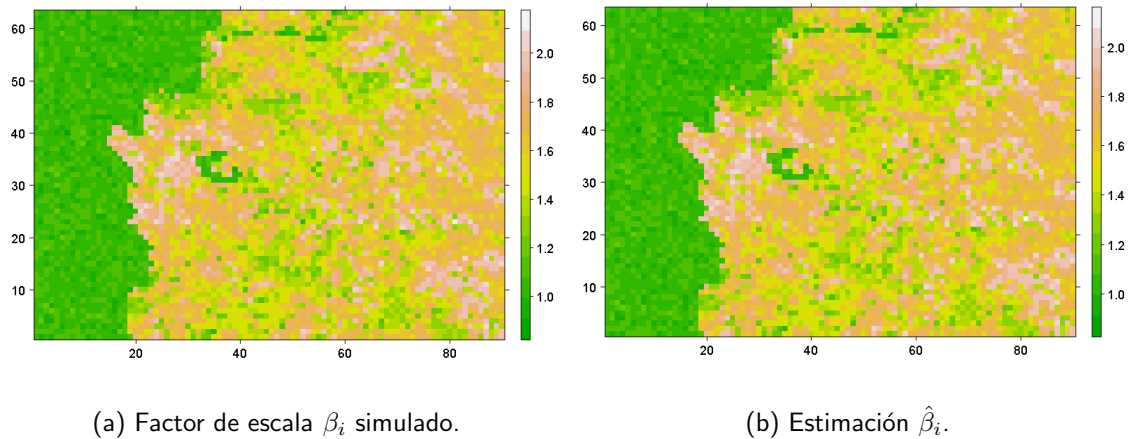


Figura A.1: Estimación WLS de los factores de escala de  $\beta_i$  para el estudio de simulación con  $\sigma_{error}^2 = 0,005$ .

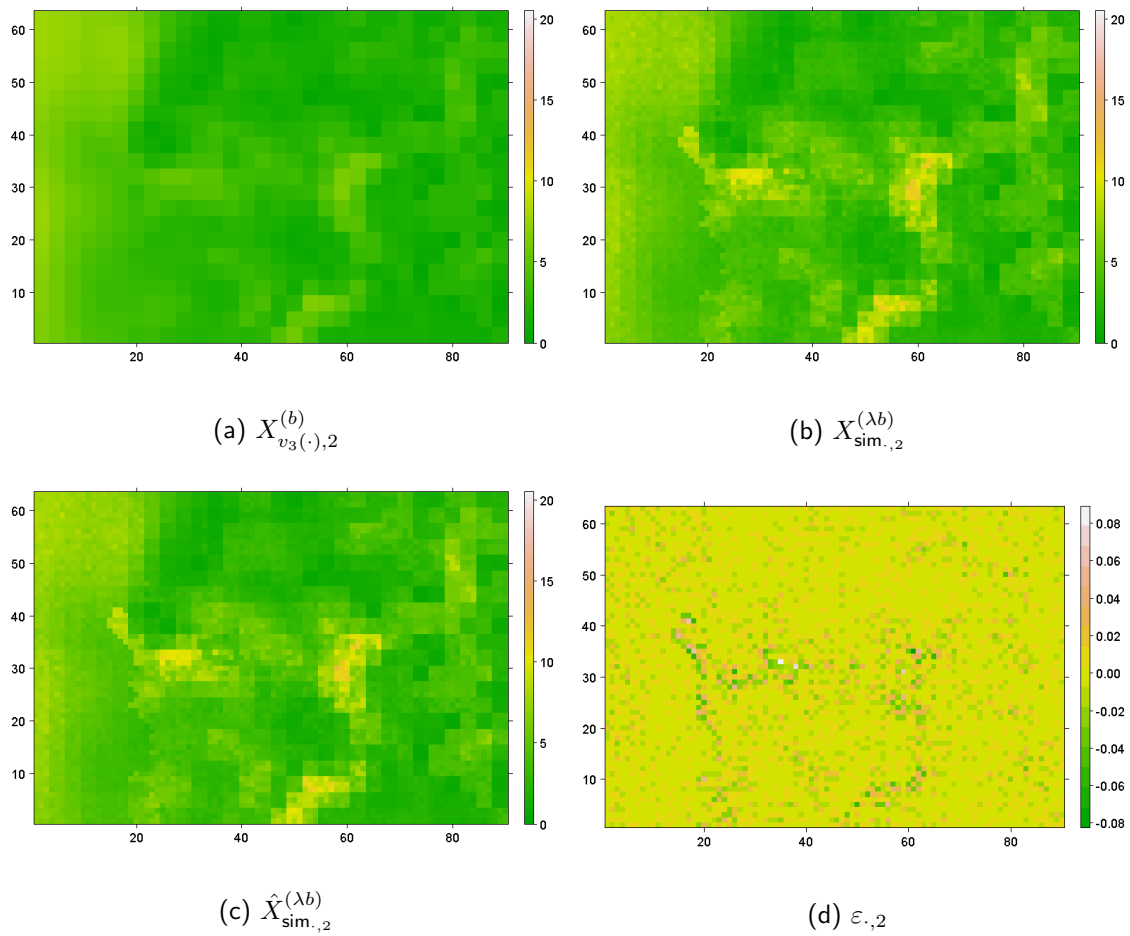


Figura A.2: Resultados de la simulación para  $\sigma_{\text{error}}^2 = 0,005$  con  $t = 2$ .

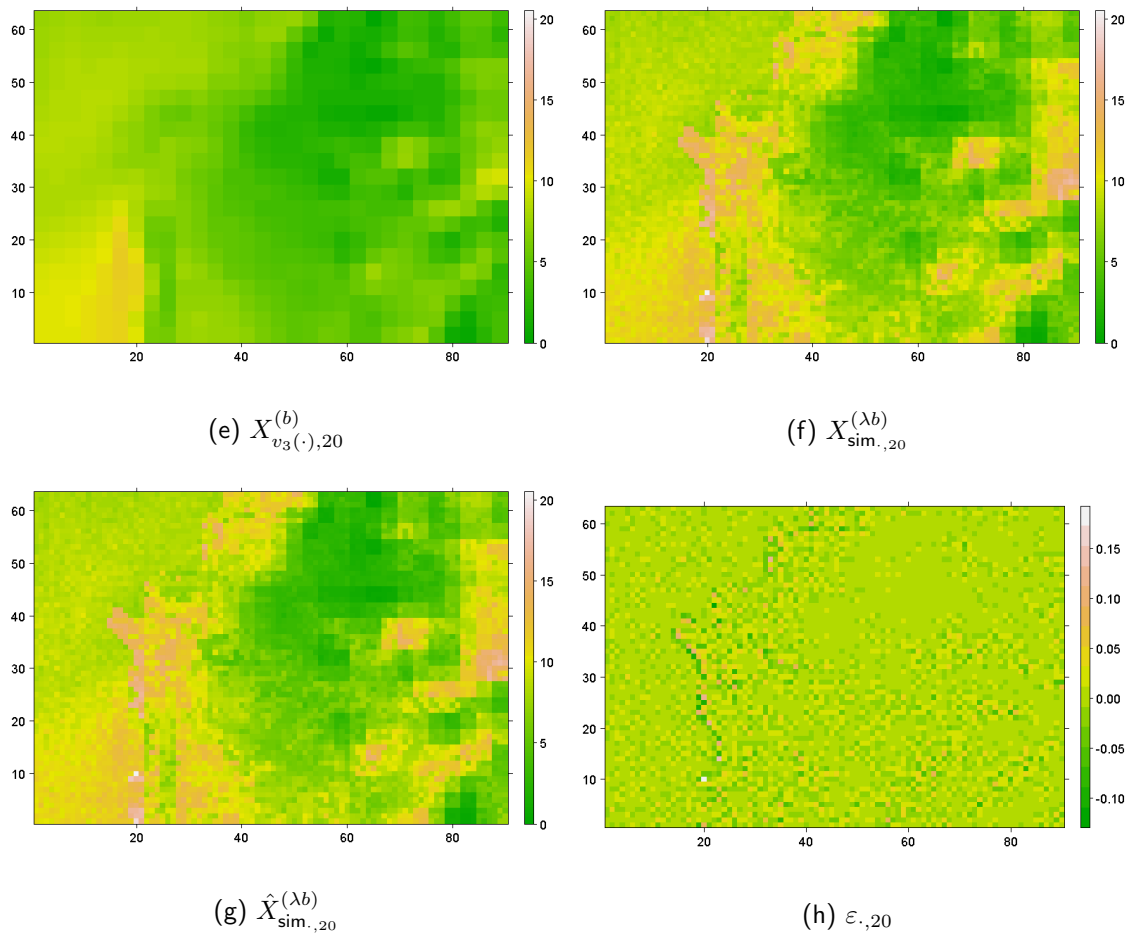


Figura A.2: Resultados de la simulación para  $\sigma_{error}^2 = 0,005$  con  $t = 20$ .

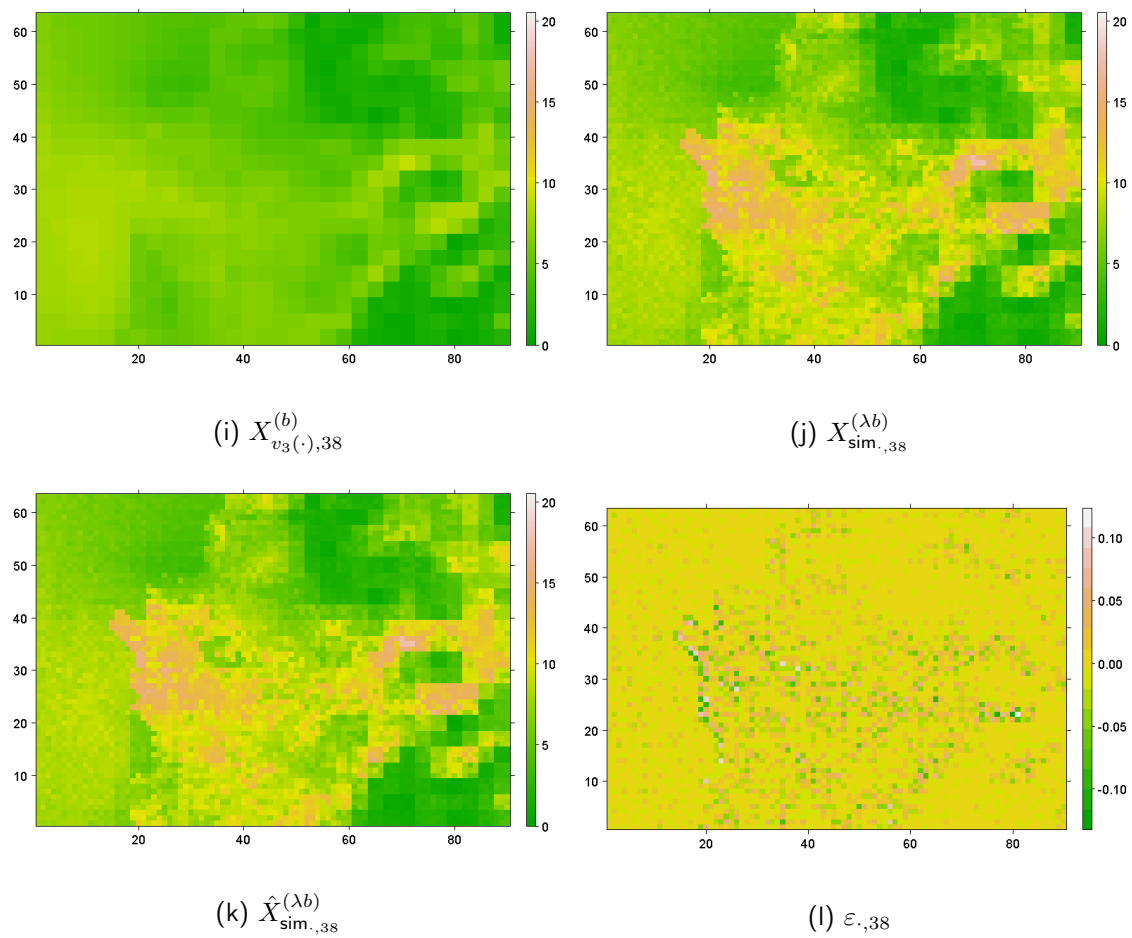


Figura A.2: Resultados de la simulación para  $\sigma_{error}^2 = 0,005$  con  $t = 38$ .

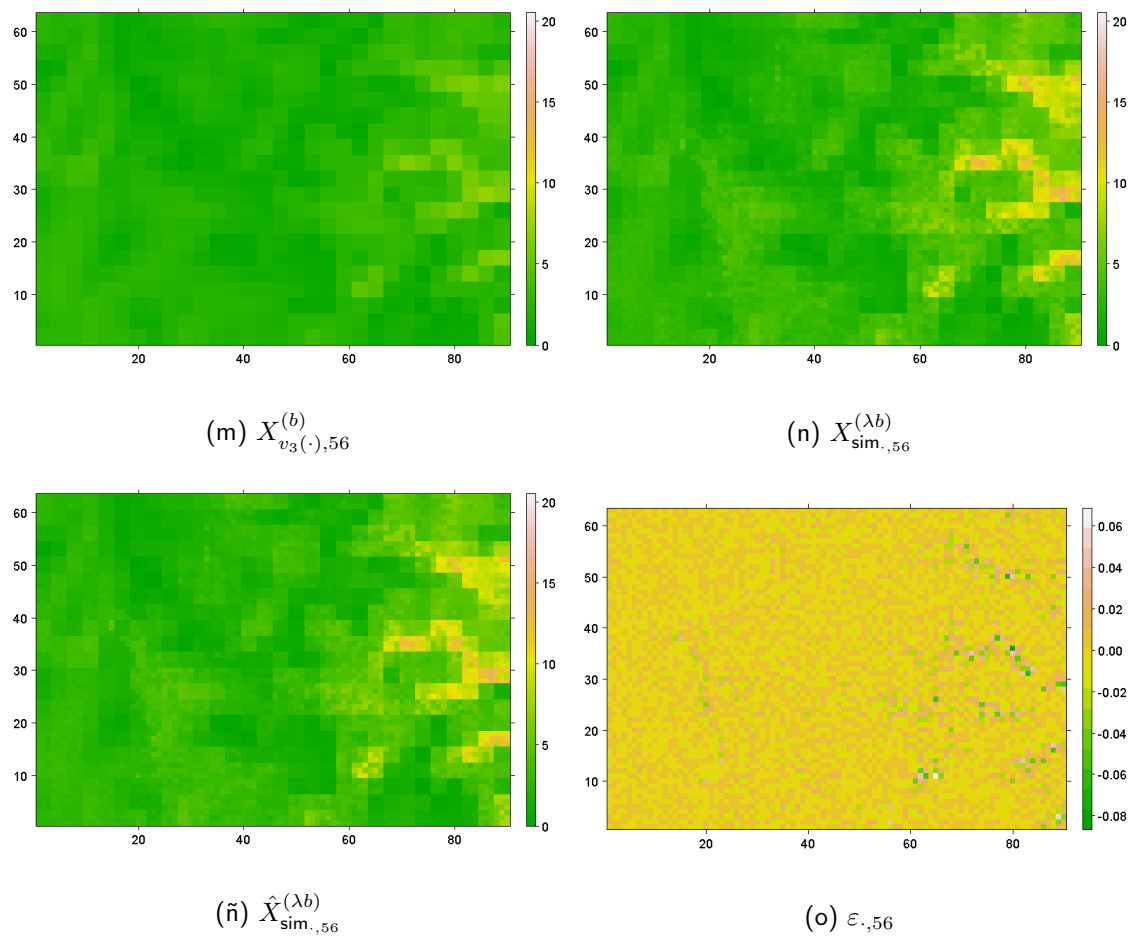
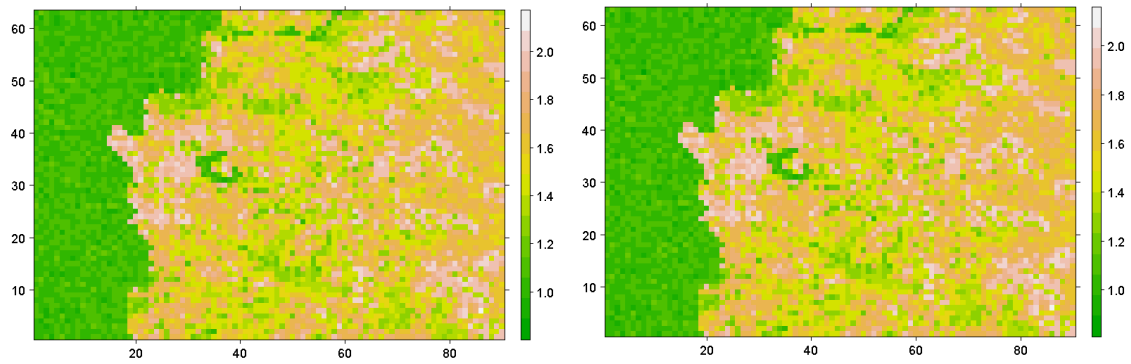


Figura A.2: Resultados de la simulación para  $\sigma_{\text{error}}^2 = 0,005$  con  $t = 56$ .

**A.2.**  $\sigma_{res}^2 = 0,01$



(a) Factor de escala  $\beta_i$  simulado.

(b) Estimación  $\hat{\beta}_i$ .

Figura A.3: Estimación WLS de los factores de escala de  $\beta_i$  para el estudio de simulación con  $\sigma_{error}^2 = 0,01$ .

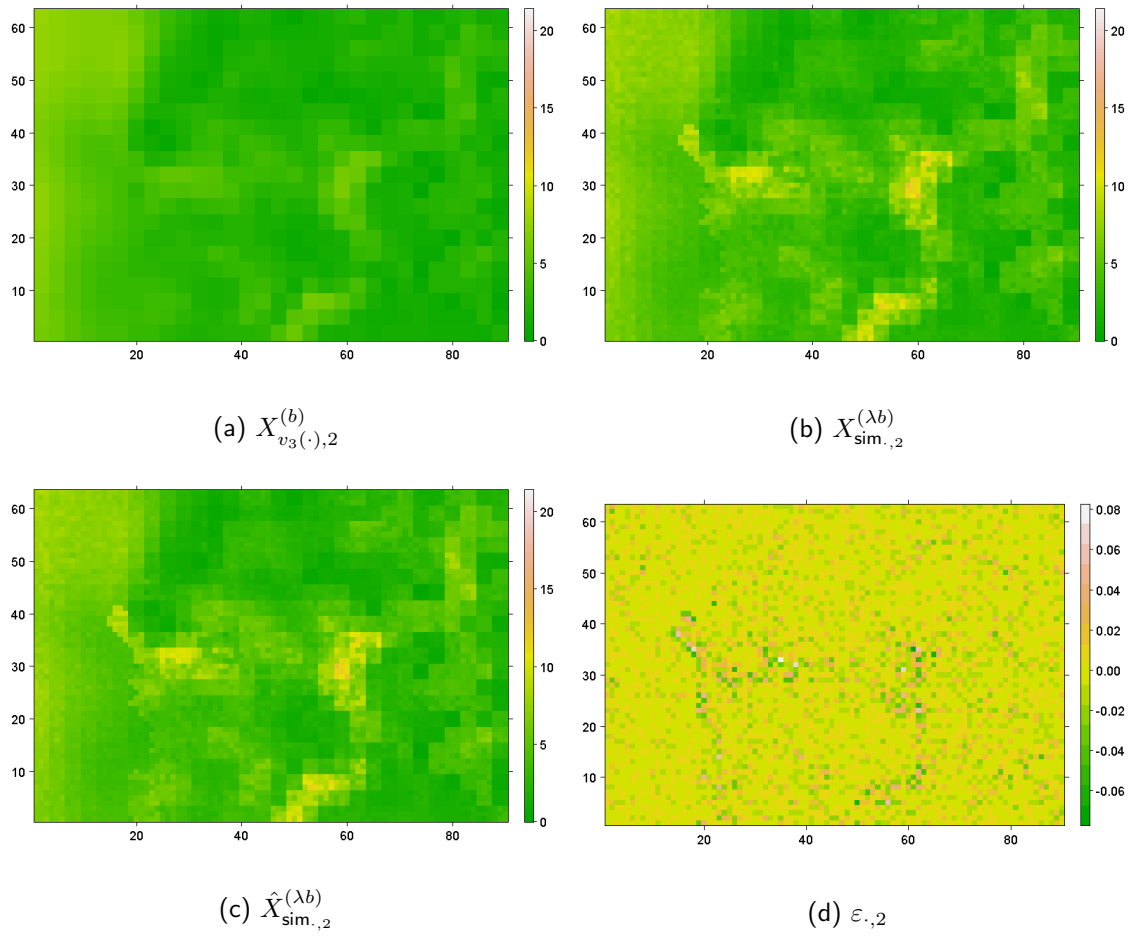


Figura A.4: Resultados de la simulación para  $\sigma_{\text{error}}^2 = 0,01$  con  $t = 2$ .

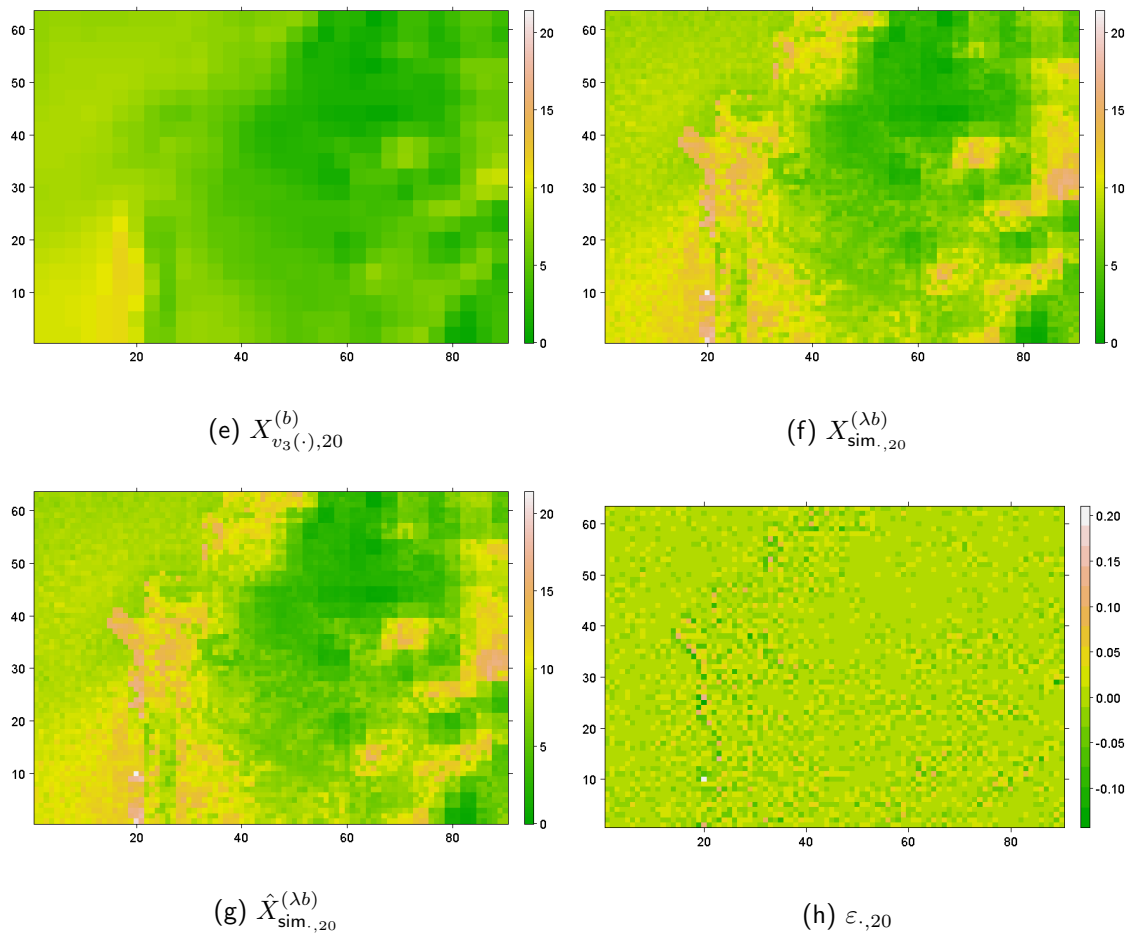


Figura A.4: Resultados de la simulación para  $\sigma_{\text{error}}^2 = 0,01$  con  $t = 20$ .

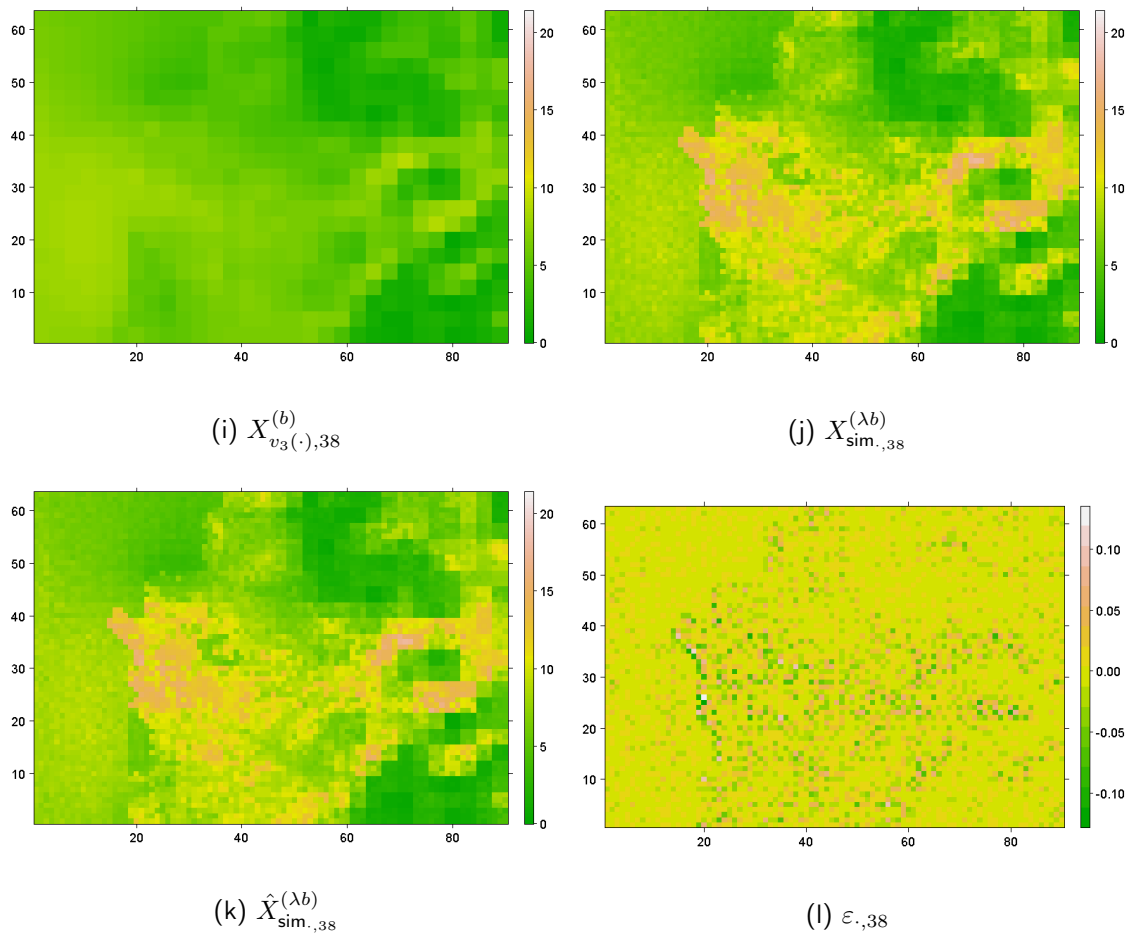


Figura A.4: Resultados de la simulación para  $\sigma_{\text{error}}^2 = 0,01$  con  $t = 38$ .

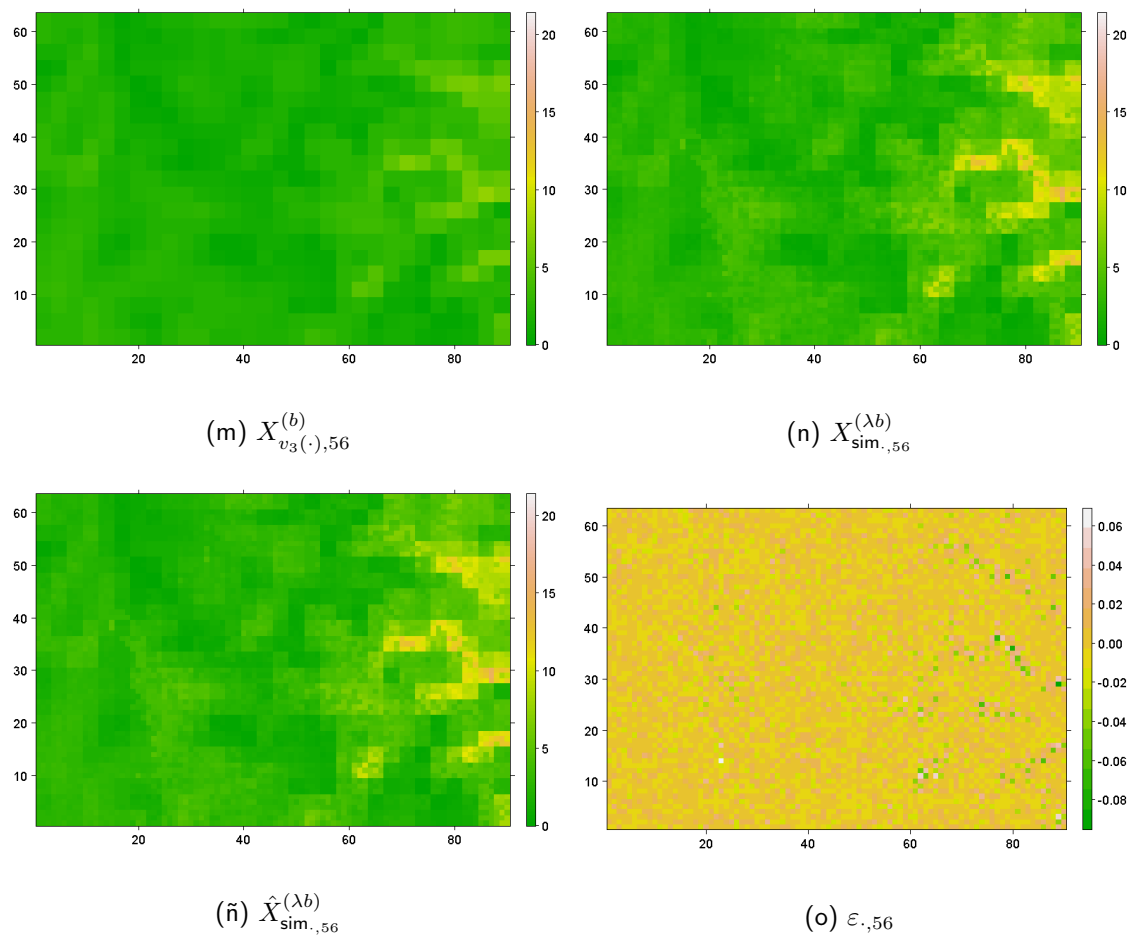
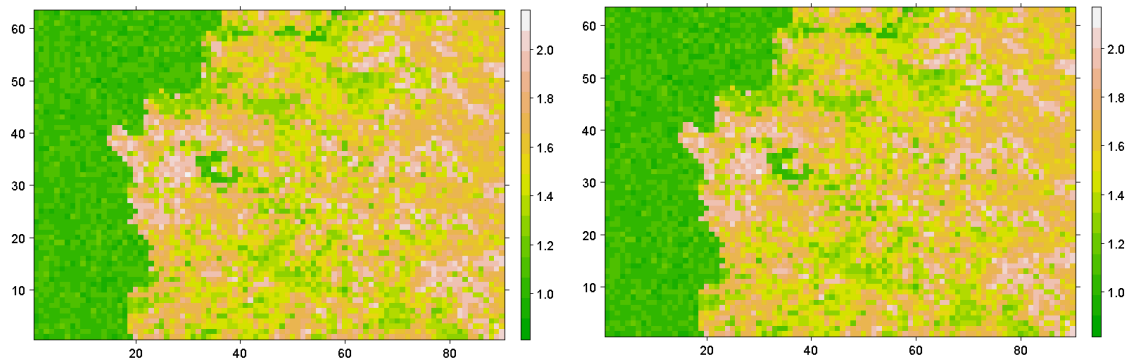


Figura A.4: Resultados de la simulación para  $\sigma_{\text{error}}^2 = 0,01$  con  $t = 56$ .

**A.3.**  $\sigma_{res}^2 = 0,03$



(a) Factor de escala  $\beta_i$  simulado.

(b) Estimación  $\hat{\beta}_i$ .

Figura A.5: Estimación WLS de los factores de escala de  $\beta_i$  para el estudio de simulación con  $\sigma_{error}^2 = 0,03$ .

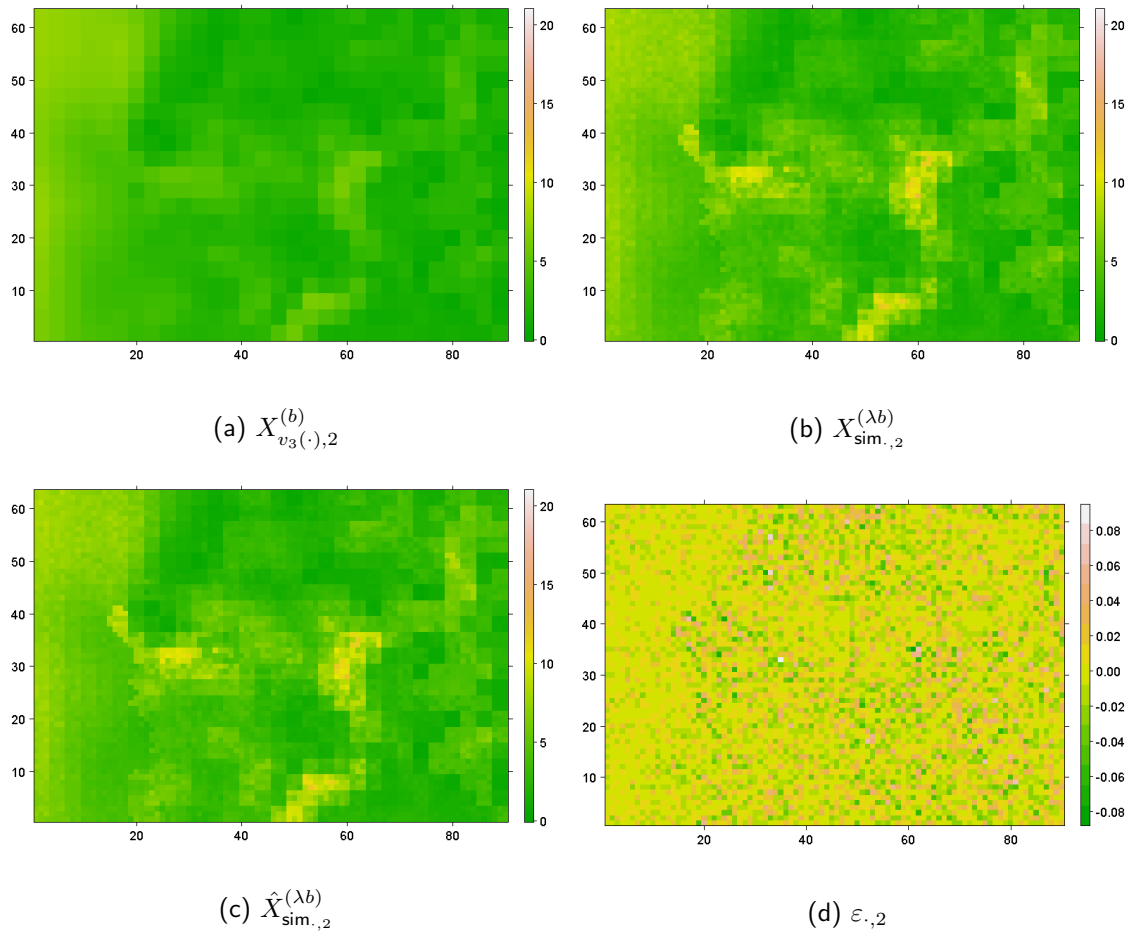


Figura A.6: Resultados de la simulación para  $\sigma_{\text{error}}^2 = 0,03$  con  $t = 2$ .

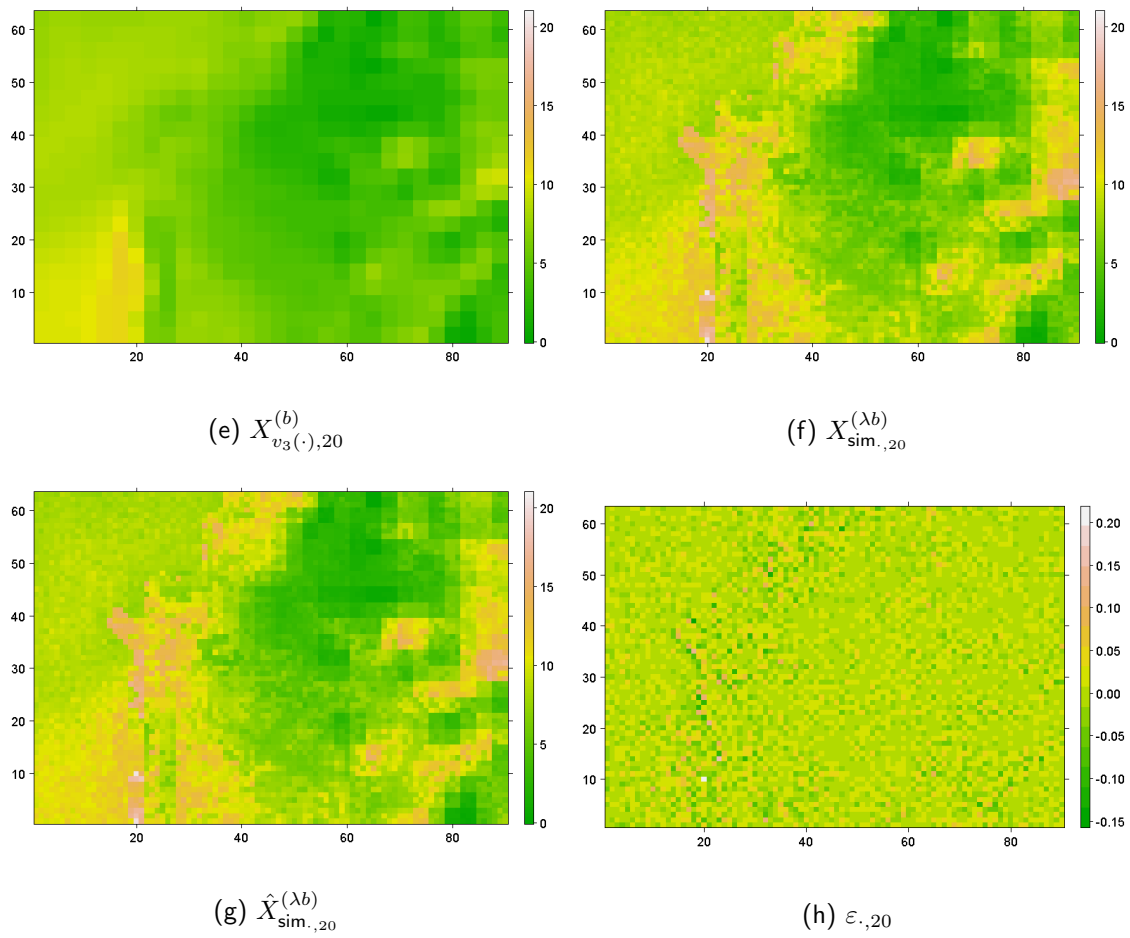


Figura A.6: Resultados de la simulación para  $\sigma_{\text{error}}^2 = 0,03$  con  $t = 20$ .

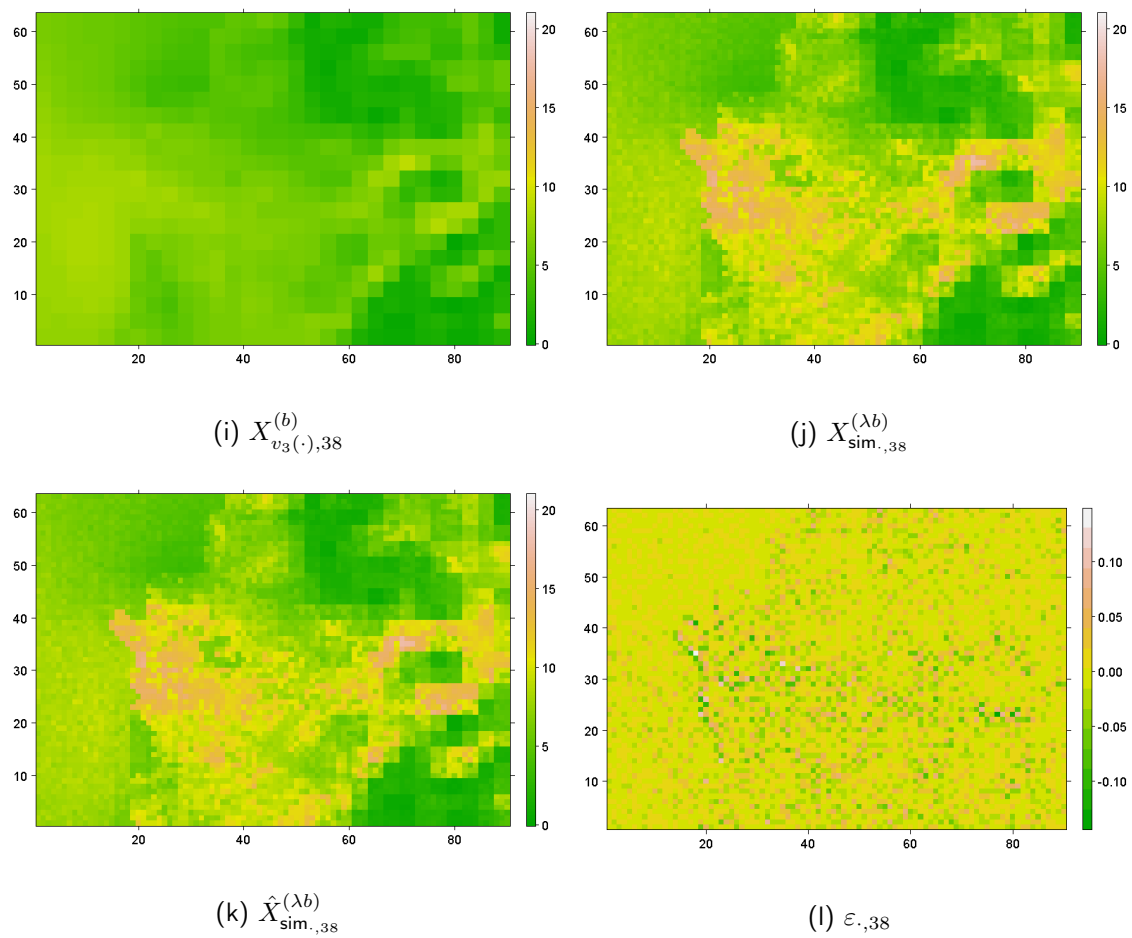


Figura A.6: Resultados de la simulación para  $\sigma_{\text{error}}^2 = 0,03$  con  $t = 38$ .

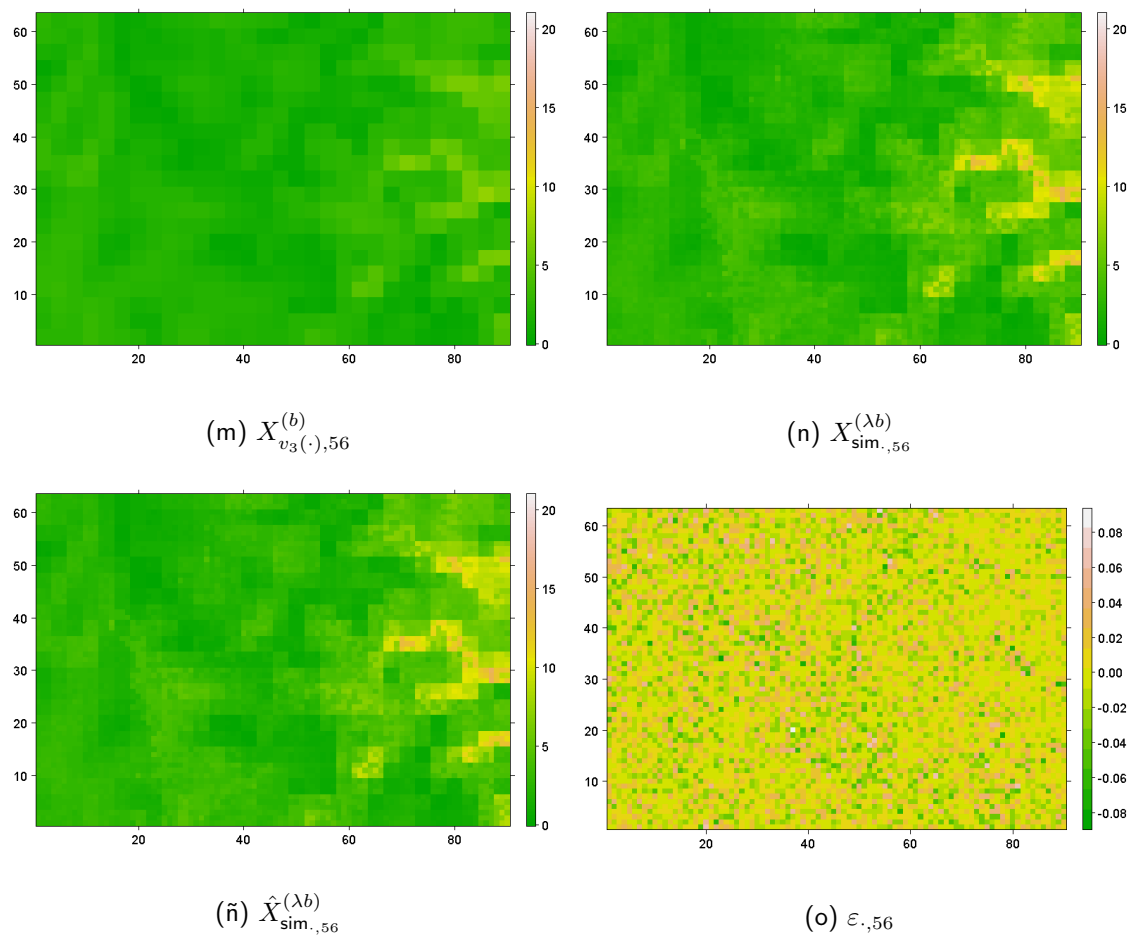
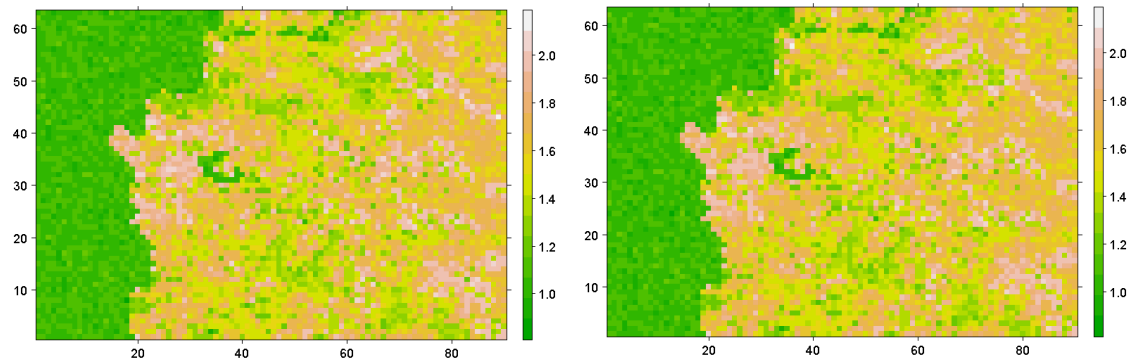


Figura A.6: Resultados de la simulación para  $\sigma_{\text{error}}^2 = 0,03$  con  $t = 56$ .

**A.4.**  $\sigma_{res}^2 = 0,05$



(a) Factor de escala  $\beta_i$  simulado.

(b) Estimación  $\hat{\beta}_i$ .

Figura A.7: Estimación WLS de los factores de escala de  $\beta_i$  para el estudio de simulación con  $\sigma_{error}^2 = 0,05$ .

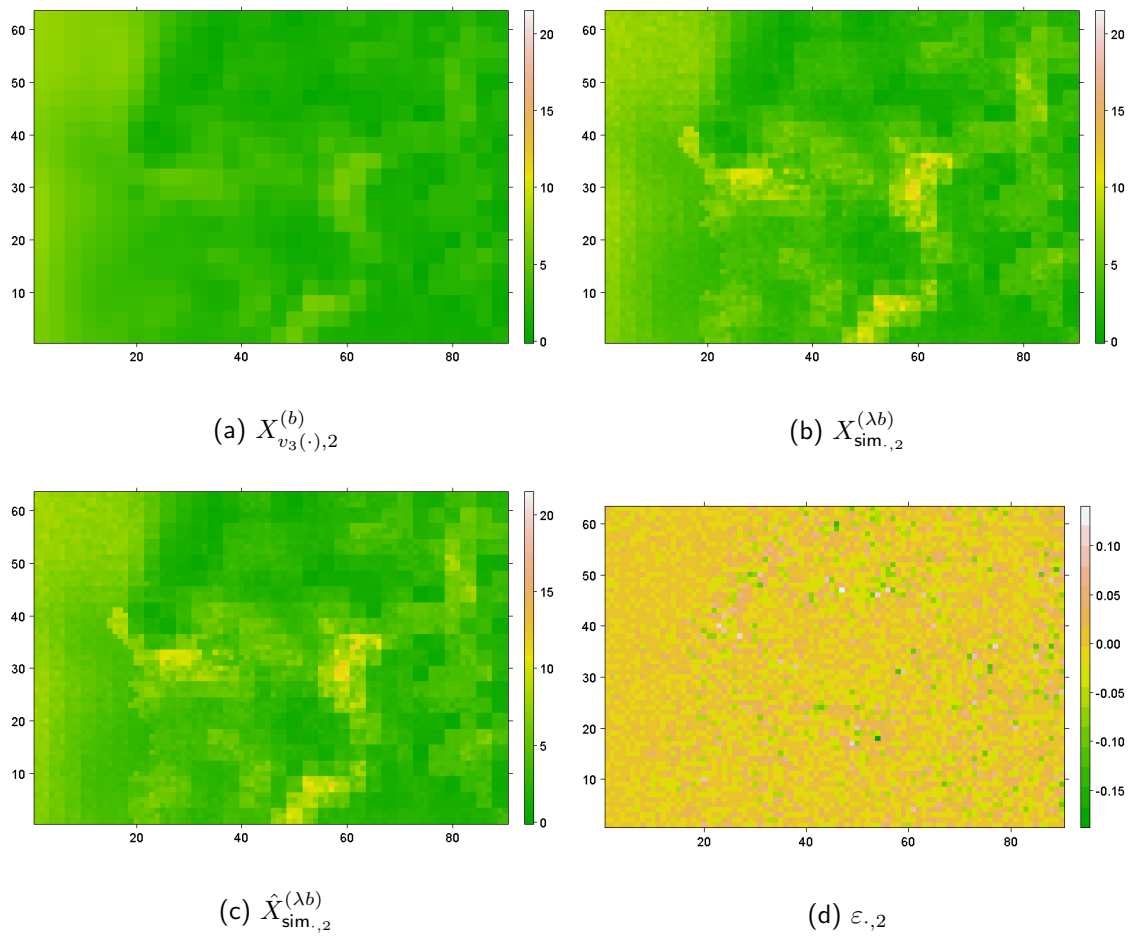


Figura A.8: Resultados de la simulación para  $\sigma_{\text{error}}^2 = 0,05$  con  $t = 2$ .

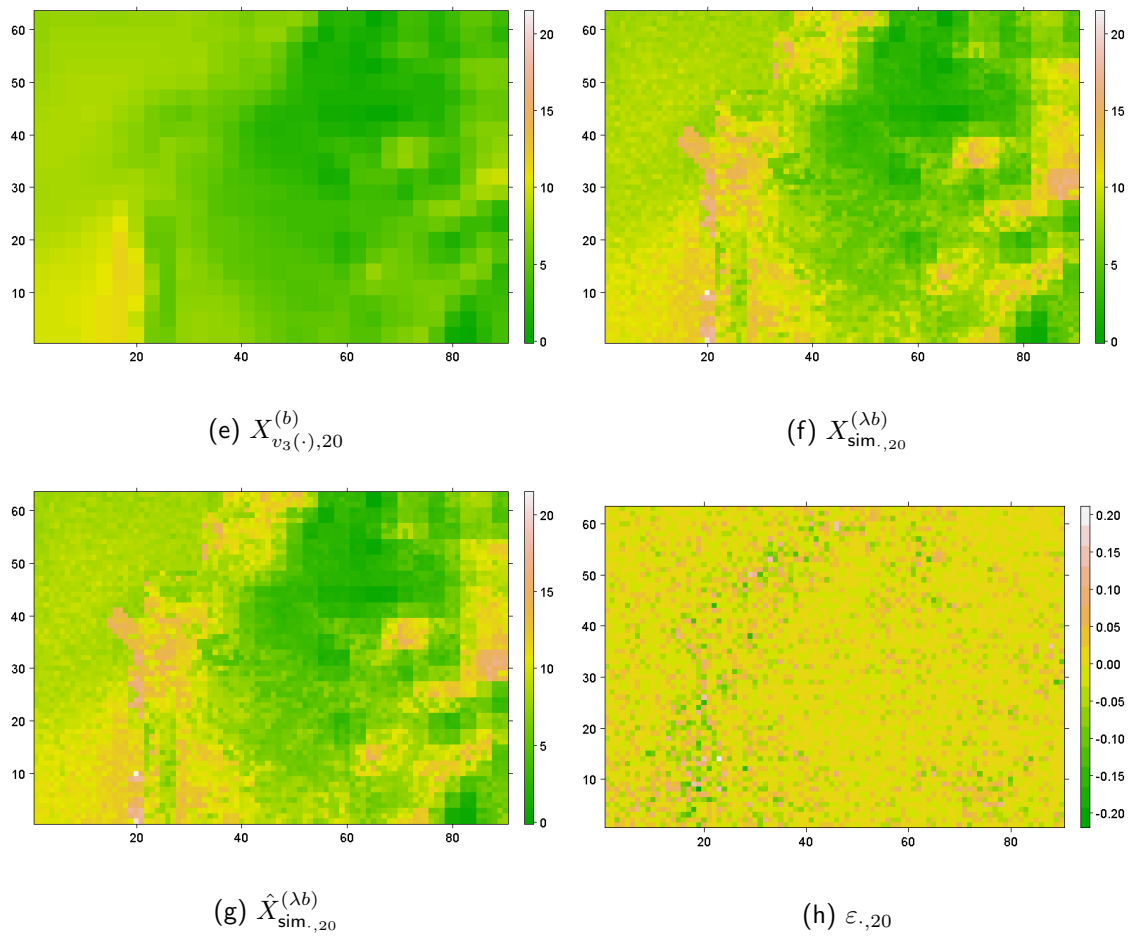


Figura A.8: Resultados de la simulación para  $\sigma_{\text{error}}^2 = 0,05$  con  $t = 20$ .

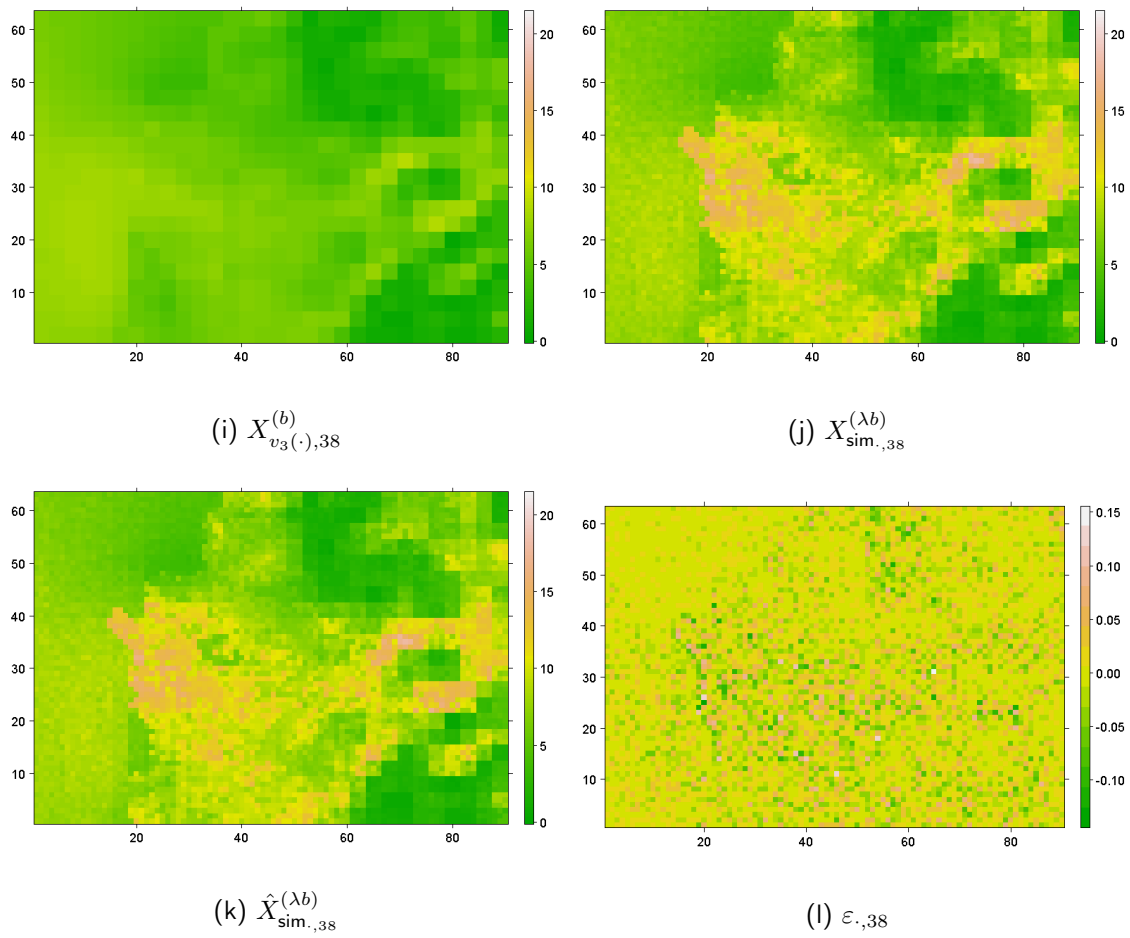


Figura A.8: Resultados de la simulación para  $\sigma_{\text{error}}^2 = 0,05$  con  $t = 38$ .

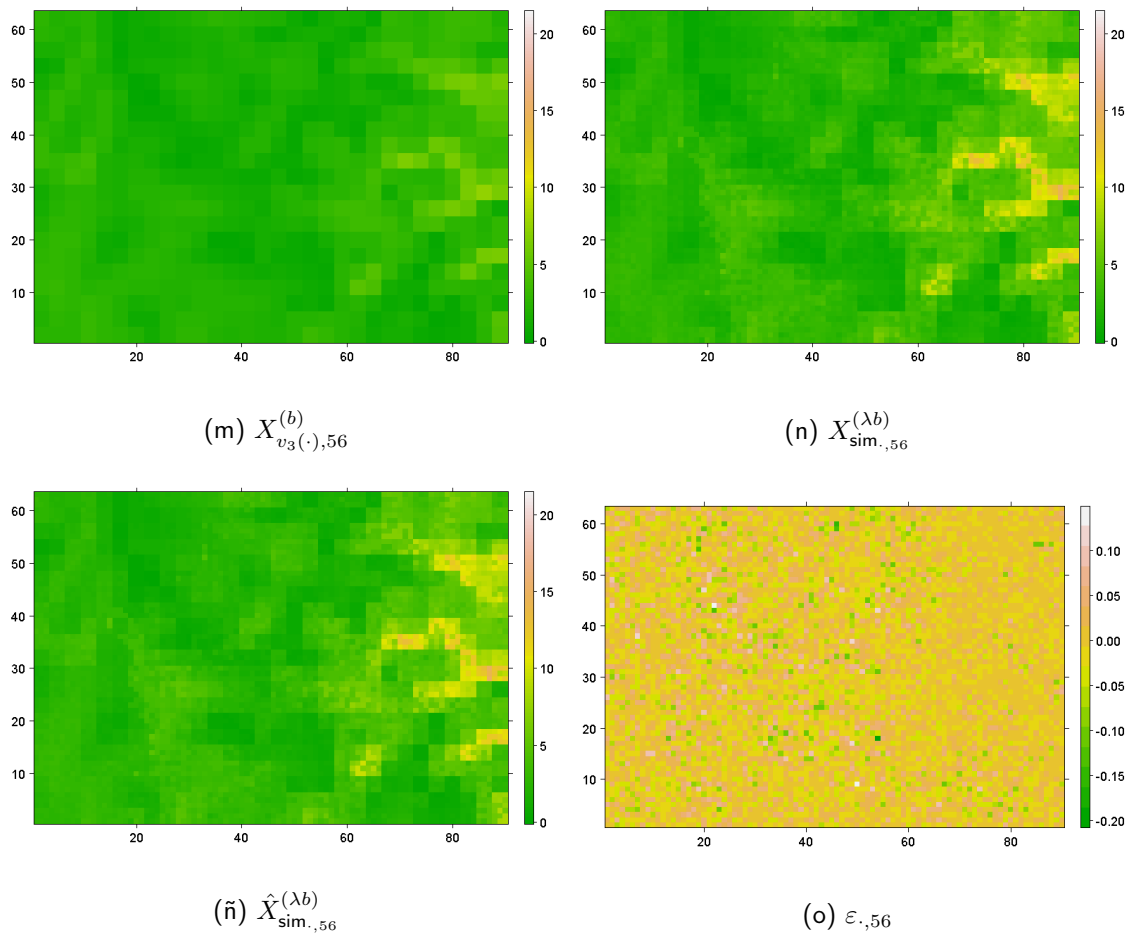


Figura A.8: Resultados de la simulación para  $\sigma_{\text{error}}^2 = 0,05$  con  $t = 56$ .

**A.5.**  $\sigma_{res}^2 = 0,1$

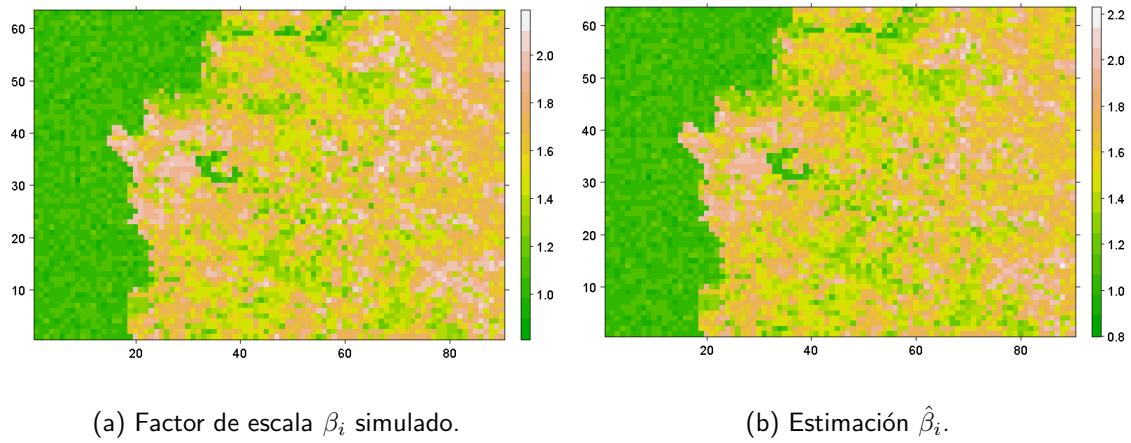


Figura A.9: Estimación WLS de los factores de escala de  $\beta_i$  para el estudio de simulación con  $\sigma_{error}^2 = 0,1$ .

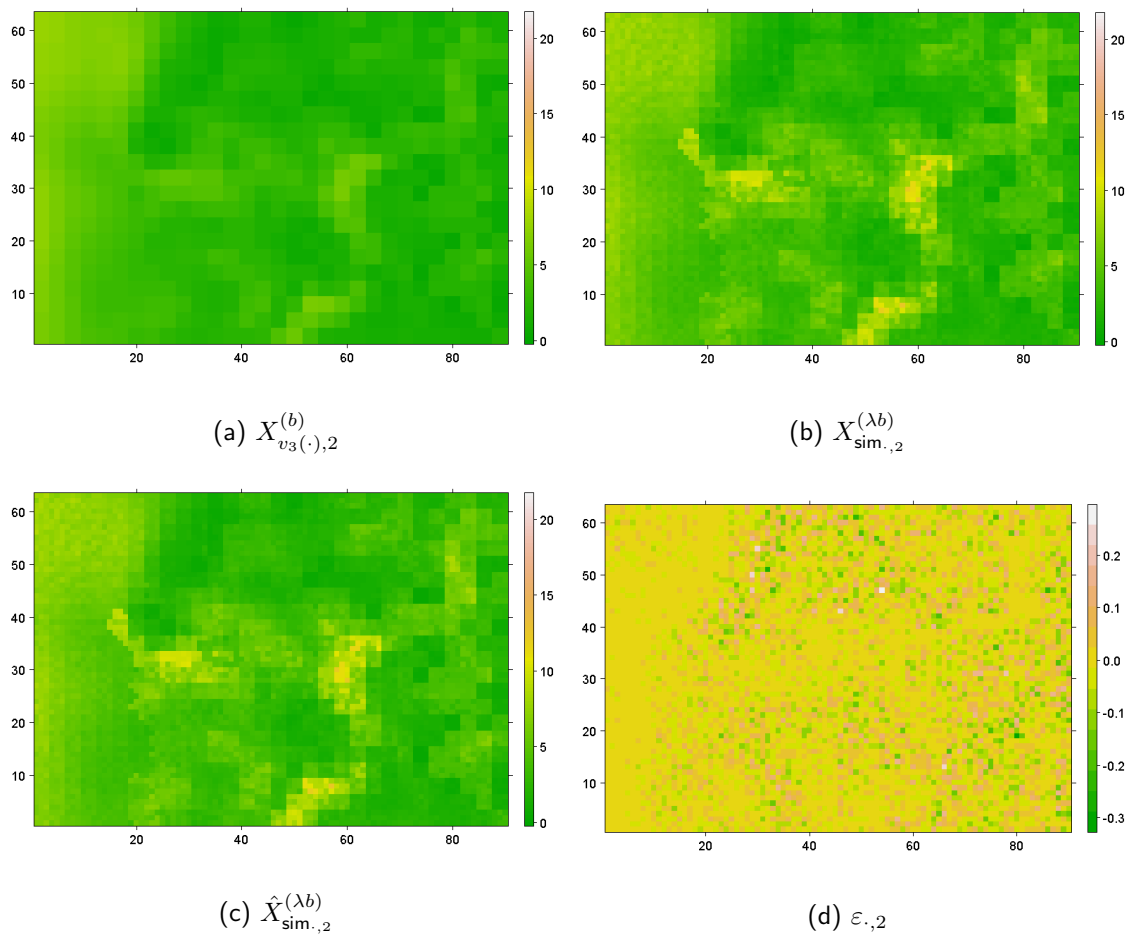


Figura A.10: Resultados de la simulación para  $\sigma_{error}^2 = 0,1$  con  $t = 2$ .

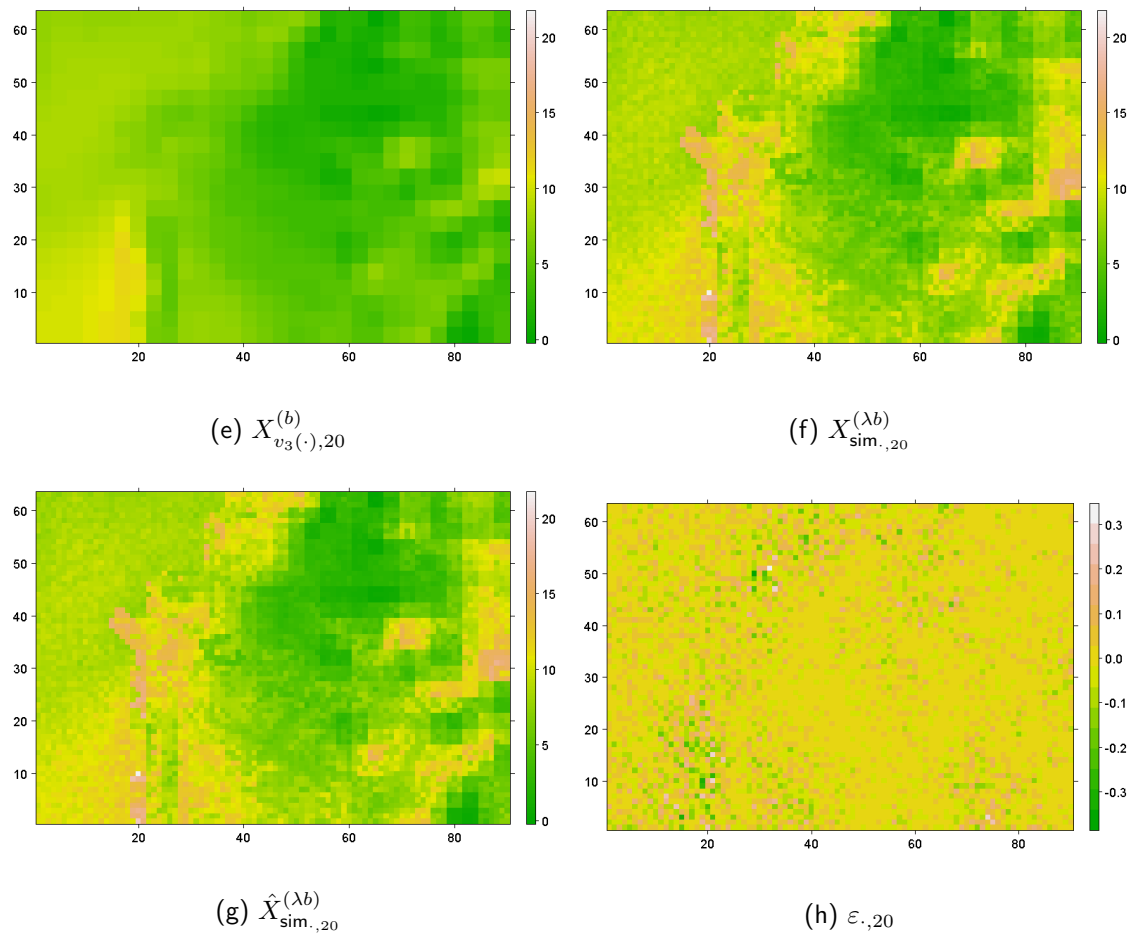


Figura A.10: Resultados de la simulación para  $\sigma_{\text{error}}^2 = 0,1$  con  $t = 20$ .

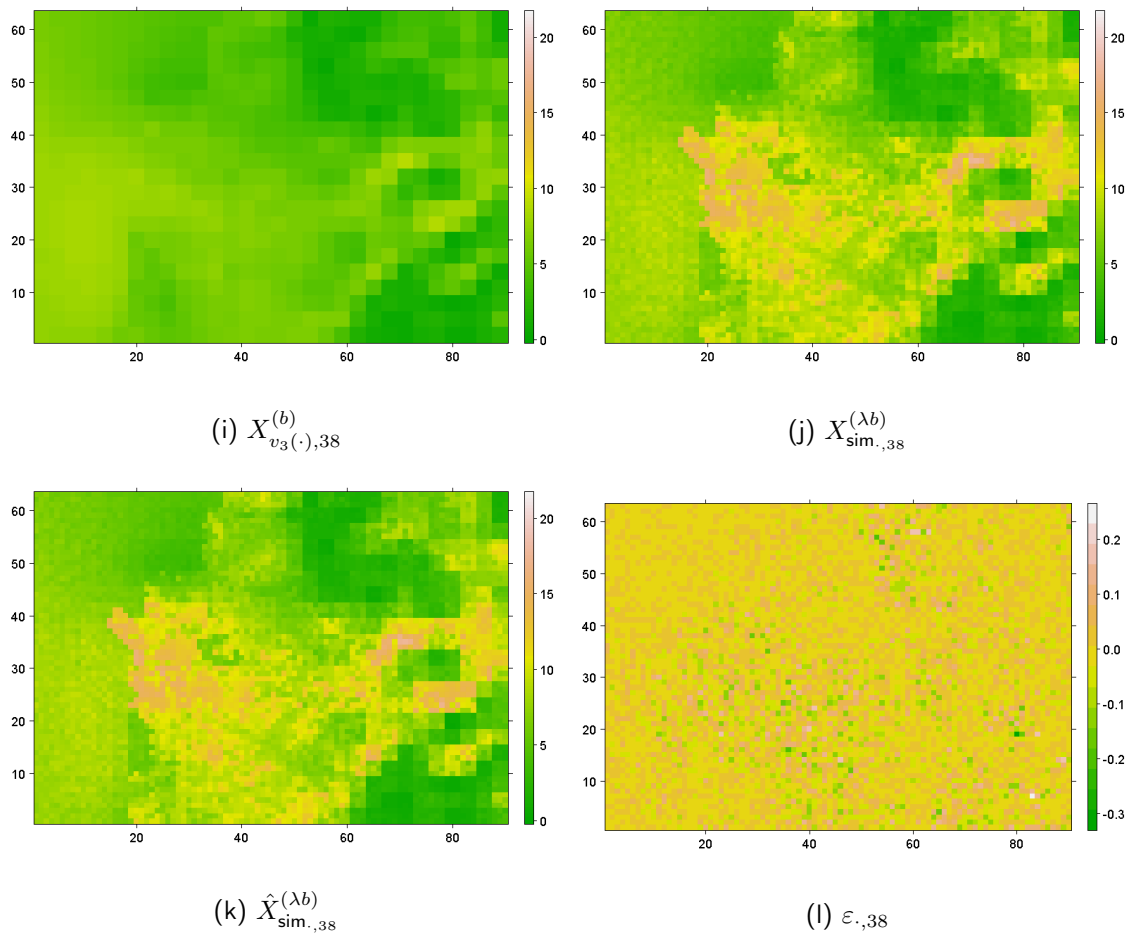


Figura A.10: Resultados de la simulación para  $\sigma_{error}^2 = 0,1$  con  $t = 38$ .

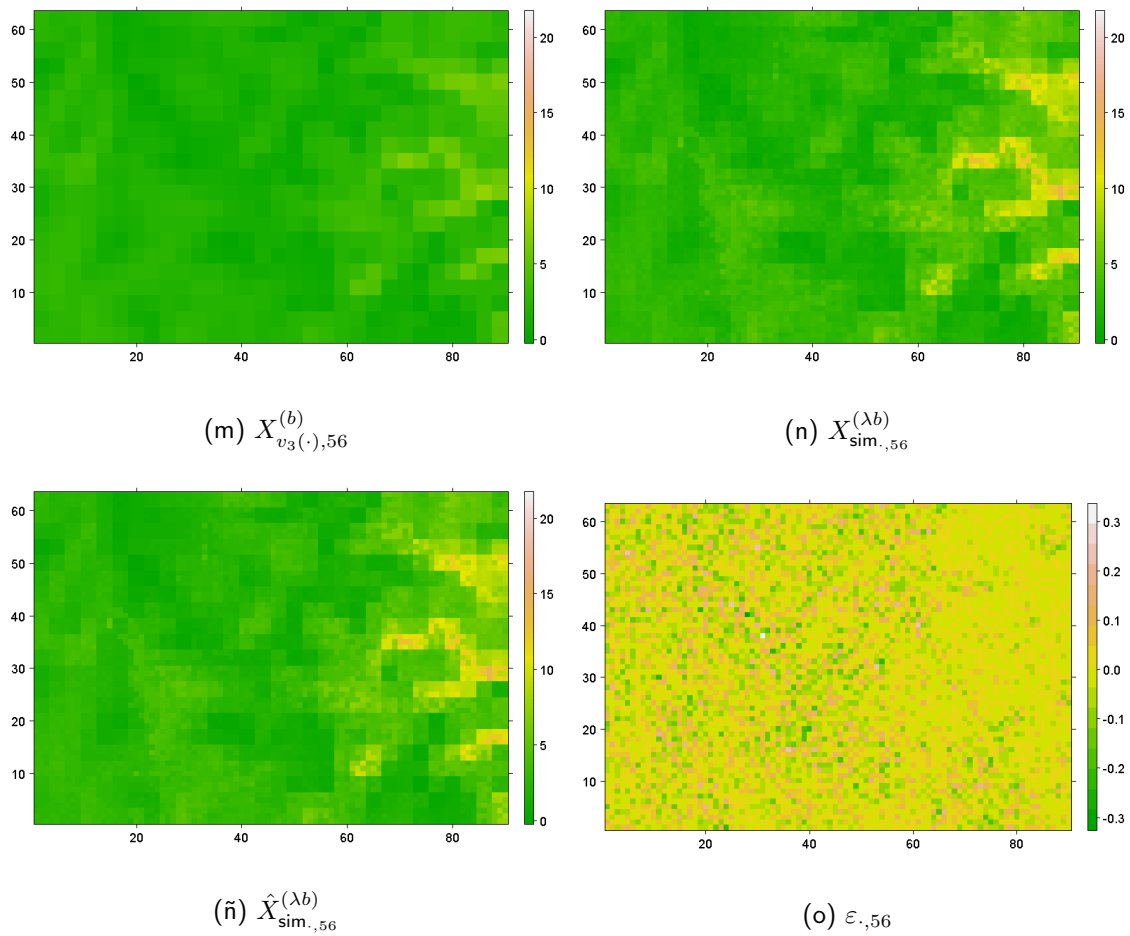


Figura A.10: Resultados de la simulación para  $\sigma_{\text{error}}^2 = 0,1$  con  $t = 56$ .

## Apéndice B

# Documentación y Respaldo

La documentación y respaldo de los códigos ejecutados en el proyecto están contenidos en el repositorio de github " *Proyecto Downscaling GTWR* ", en el siguiente [link](#), el cual contiene los análisis descriptivos, clasificación con el método FPC, el estudio de simulación para estructuras de ruido browniano simple, implementación del modelo GTWR en clases de R y sugerencias de código. El repositorio esta compuesto por los archivos:

- *data* (Carpeta): contiene los datos de los modelos WRF a 3km y 1km.
- *TallerdeTítulo.Rmd*: Exploración de las mallas, Análisis descriptivo, Clasificación FPC, Simulación, Implementación GTWR y Aplicación a malla objetivo.
- *README.md*: Archivo markdown que describe el contenido y cómo ejecutar los archivos del repositorio.

Como sugerencias de código se incorporan: chunks con métodos de validación cruzadas espacial y temporal en la clase *GTWRbase*, y métodos para aplicar interpoladores en el campo predictor  $X_{v_b}^{(b)}$ .

# Bibliografía

- [1] Tércio Ambrizzi, Michelle Simões Reboita, Rosmeri Porfírio da Rocha, and Marta Llopart. The state of the art and fundamental aspects of regional climate modeling in south america. *Annals of the New York Academy of Sciences*, 1436(1):98–120, 2019. [19](#)
- [2] F. Benavides. Metodología de downscaling aplicado a un modelo de mesoescala en la región de valparaíso, 2017. [8](#), [16](#), [22](#), [27](#), [31](#)
- [3] R.E. Benestad, I. Hanssen-Bauer, and D. Chen. Empirical-statistical downscaling. *World Scientific Series on Meteorology of East Asia*, 2008. [7](#)
- [4] David A. Dickey and Wayne A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a):427–431, 1979. [57](#)
- [5] R. E. Dickinson, R. M. Errico, F. Giorgi, and G. T. Bates. A regional climate model for the western united states. *Climatic Change*, 15(3-4):383–422, 1989. [19](#), [20](#)
- [6] Paul N. Edwards. History of climate modeling. *WIREs Climate Change*, 2(1):128–139, 2011. [18](#), [20](#)
- [7] Stewart Fotheringham, Chris Brunsdon, and Martin Charlton. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley, 2002. [23](#)
- [8] René D. Garreaud. The andes climate and weather. *Advances in Geosciences*, 22:3–11, 2009. [7](#)
- [9] Michael Ghil, Myles R Allen, Michael D Dettinger, Kayo Ide, Dmitri Kondrashov, Michael E Mann, Andrew W Robertson, Adam Saunders, Yimin Tian, Ferenc Varadi, et al. Advanced spectral methods for climatic time series. *Reviews of Geophysics*, 40(1):1–41, 2002. [16](#)
- [10] Filippo GIORGI and Xue-Jie GAO. Regional earth system modeling: review and future directions. *Atmospheric and Oceanic Science Letters*, 11(2):189–197, 2018. [21](#)
- [11] Silfrido Gómez Polanco. Un modelo de regresión con ruido fraccionario coloreado. 2023. [37](#)
- [12] Andrea N. Hahmann, Claire L. Vincent, Alfredo Peña, Jacob Lange, and Charlotte B. Hasager. Wind climate estimation using WRF model output: method and model sensitivities over the sea. *International Journal of Climatology*, 35(12):3422–3439, 2014. [21](#)
- [13] Bo Huang, Bo Wu, and Michael Barry. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, 24(3):383–401, 2010. [31](#)
- [14] Michael H. Kutner, Christopher J. Nachtsheim, John Neter, and William Li. *Applied Linear Statistical Models*. McGraw-Hill/Irwin, 5 edition, 2004. [22](#), [25](#)

- [15] Xiaolei Ma, Jiyu Zhang, Chuan Ding, and Yunpeng Wang. A geographically and temporally weighted regression model to explore the spatiotemporal influence of built environment on transit ridership. *Computers, Environment and Urban Systems*, 70:113–124, 2018. 23
- [16] L. et al. Mearns. The north american regional climate change assessment program: Overview of phase i results. *Bulletin of the American Meteorological Society*, 93(9):1337–1362, 2012. 7
- [17] Ministerio de Energía de Chile. Aumentan los parques solares y la generación distribuida en la región de valparaíso, 2025. Consultado el 27 de junio de 2025. 33
- [18] National Oceanic and Atmospheric Administration. El niño and la niña (enso). Online, 2023. Accessed: November 15, 2023. 7
- [19] Kihong Park and Walter Willinger, editors. *Self-Similar Network Traffic and Performance Evaluation*. John Wiley & Sons, Inc., New York, 2000. 12
- [20] J P Peixoto and A H Oort. *Physics of climate*. New York, NY (United States); American Institute of Physics, 01 1992. 18, 19
- [21] Peter C.B. Phillips and Pierre Perron. Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346, 1988. 31, 57
- [22] J. G. Powers, J. B. Klemp, W. C. Skamarock, C. A. Davis, J. Dudhia, D. O. Gill, and Coauthors. The weather research and forecasting model: Overview, system efforts, and future directions. *Bulletin of the American Meteorological Society*, 98(8):1717–1737, 2017. 20
- [23] Rulamahue Consultores Forestales. Ficha técnica CL05: Caracterización de propiedades de la madera, 2023. 34
- [24] Markku Rummukainen. Methods for statistical downscaling of gcm simulations. Technical report, Swedish Meteorological and Hydrological Institute (SMHI), 1997. Retrieved from SMHI website. 21
- [25] José A. Rutllant and Humberto A. Fuenzalida. Synoptic aspects of the central chile rainfall variability associated with the southern oscillation. *International Journal of Climatology*, 11(1):63–76, 1991. 7
- [26] Christoph Schär, Oliver Fuhrer, Andrea Arteaga, Nikolina Ban, Christophe Charpiloz, Salvatore Di Girolamo, Laureline Hentgen, Torsten Hoefler, Xavier Lapillonne, David Leutwyler, Katherine Osterried, Davide Panosetti, Stefan Rüdüsühli, Linda Schlemmer, Thomas C. Schulthess, Michael Sprenger, Stefano Ubbiali, and Heini Wernli. Kilometer-scale climate models: Prospects and challenges. *Bulletin of the American Meteorological Society*, 101(5):E567–E587, 2020. 21, 24
- [27] Servicio de Evaluación Ambiental. Estudio de impacto ambiental “parque eólico cerro piedra”. <https://infofirma.sea.gob.cl/DocumentosSEA/MostrarDocumento?docId=13/24/09bf96cd48597b7336a2d9b4d21d5baa2a2e>, 2020. Titular: Energía Eólica Curauma SpA. Consultado el 27 de junio de 2025. 33
- [28] Vahid Khalilpour Shorabeh, Shahram Shokri, and Naser Pariz. Optimal site selection for wind farms using gis-based multi criteria decision analysis (case study: Iran). *Energy Sources, Part B: Economics, Planning, and Policy*, 6(4):400–409, 2011. 8

- [29] William C. Skamarock, Joseph B. Klemp, Jimy Dudhia, David O. Gill, Dale Barker, Michael G. Duda, Xiang-Yu Huang, Wei Wang, and Jordan G. Powers. A description of the advanced research WRF version 3. Technical Report NCAR/TN-475+STR, University Corporation for Atmospheric Research, Boulder, CO, USA, jun 2008. Mesoscale and Microscale Meteorology Division. [20](#)
- [30] William C Skamarock, Joseph B Klemp, Jimy Dudhia, David O Gill, Zhiquan Liu, Judith Berner, Wei Wang, Jordan G Powers, Michael G Duda, Dale M Barker, et al. A description of the advanced research wrf model version 4. *National Center for Atmospheric Research: Boulder, CO, USA*, 145(145):550, 2019. [20](#)
- [31] Bjorn Stevens and Sandrine Bony. What are climate models missing? *Science*, 340(6136):1053–1054, 2013. [19](#)
- [32] Sylwia Trzaska and Emilie Schnarr. A review of downscaling methods for climate change projections, 09 2014. [22](#)
- [33] A. W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer, 1996. [16](#)
- [34] R.L. Wilby and T.M.L. Wigley. Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography*, 30:651–688, 2009. [7](#)
- [35] Robert L. Wilby, Charles W. Dawson, Conor Murphy, Paul O'Connor, and Ed Hawkins. The statistical downscaling model - decision centric (sdsdc): conceptual basis and applications. *Climate Research*, 61:259–276, 2014. [22](#)
- [36] J. J. Wimhurst, C. C. Nsude, and J. S. Greene. Standardizing the factors used in wind farm site suitability models: A review. *Heliyon*, 9(5):e15903, 2023. [8](#)
- [37] Gonzalo Yáñez-Morróni, Jorge Gironás, Marta Caneo, Rodrigo Delgado, and René Garreaud. Using the weather research and forecasting (wrf) model for precipitation forecasting in an andean region with complex topography. *Atmosphere*, 9(8):304, aug 2018. Article Number: 304. [21](#)