

UNIVERSIDAD DE CONCEPCIÓN  
DIRECCIÓN DE POSTGRADO  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
PROGRAMA MAGÍSTER EN ESTADÍSTICA



**LA GEOESTADÍSTICA COMO HERRAMIENTA DE ANÁLISIS  
ESPACIAL PARA INVENTARIOS FORESTALES**

Tesis presentada a la Facultad de Ciencias Físicas y Matemáticas de la  
Universidad de Concepción para optar al grado académico de Magíster  
en Estadística

**POR: CAROLINA ISABEL VILLALOBOS ROJAS**

Profesor Guía: Dra. Daisy Arroyo Fernandez

Abril 2021

CONCEPCIÓN - CHILE



© 2021, La Geoestadística como Herramienta de Análisis Espacial para Inventarios Forestales

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento



UNIVERSIDAD DE CONCEPCIÓN  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

**LA GEOESTADÍSTICA COMO HERRAMIENTA DE ANÁLISIS ESPACIAL  
PARA INVENTARIOS FORESTALES**

Tesis presentada a la Facultad de Ciencias Físicas y Matemáticas de la  
Universidad de Concepción para optar al grado académico de Magíster en



**Profesor Guía:**

Dra. Daisy Arroyo Fernandez<sup>1</sup>

Firma:.....

**Comisión:**

Dra. María Paz Casanova Laudien<sup>1</sup>

Firma:.....

MSc. Sebastián Niklitschek Soto<sup>1</sup>

Firma:.....

Dr. Simón Sandoval Rocha<sup>2</sup>

Firma:.....

**Memorante:**

Carolina Villalobos Rojas

Firma:.....

Abril 2021

Concepción, Chile

<sup>1</sup>Departamento de Estadística, Universidad de Concepción

<sup>2</sup>Departamento Manejo de Bosques y Medio Ambiente, Universidad de Concepción

## AGRADECIMIENTOS

En primer lugar quiero agradecer a mi profesora guía Dra. Daisy Arroyo Fernandez por su guía y consejos durante todo el desarrollo de la tesis.

Además, agradezco a mi familia por su apoyo y paciencia.

Finalmente, al financiamiento de la Agencia Nacional de Investigación y Desarrollo (ANID), vía el proyecto ANID/CONICYT FONDECYT INICIACIÓN 11170529, titulado “Multivariate geostatistical simulation and its application to mineral resource evaluation”.



# Índice general

<b>AGRADECIMIENTOS</b>	<b>IV</b>
<b>RESUMEN</b>	<b>XII</b>
<b>ABSTRACT</b>	<b>XIII</b>
<b>1 Introducción</b>	<b>1</b>
<b>2 Marco Teórico</b>	<b>5</b>
2.1 Geoestadística . . . . .	5
2.1.1 Variable Regionalizada . . . . .	6
2.1.2 Función aleatoria . . . . .	8
2.2 Covarianza Espacial . . . . .	9
2.2.1 Estacionariedad . . . . .	9
2.2.1.1 Estacionariedad estricta . . . . .	9
2.2.1.2 Estacionariedad de orden 2 . . . . .	10
2.2.1.3 Estacionariedad intrínseca . . . . .	11
2.3 Variograma . . . . .	12
2.3.1 Comportamiento en el origen . . . . .	15
2.3.2 Gráfico del Variograma . . . . .	16
2.3.3 Anisotropía . . . . .	17
2.3.4 Modelos de Variograma . . . . .	18
2.4 Desagrupamiento . . . . .	20
2.5 Kriging . . . . .	21
2.5.1 Tipos de Kriging . . . . .	22
2.5.2 Validación Cruzada . . . . .	24
2.6 Simulación geoestadística . . . . .	24
2.6.1 Simulación No Condicional . . . . .	26
2.6.2 Simulación Condicional . . . . .	26
2.6.3 Algoritmos de simulación . . . . .	26
2.6.3.1 Algoritmo Espectral de Bandas Rotantes . . . . .	27
2.6.4 Implementación del algoritmo de simulación . . . . .	28

---

2.6.5	Validación de las simulaciones . . . . .	29
2.6.6	Validación cruzada . . . . .	30
2.7	Marco Forestal . . . . .	30
2.7.1	Antecedentes generales . . . . .	30
2.7.2	Mensura Forestal . . . . .	32
2.7.3	Geoestadística aplicada a recursos forestales . . . . .	33
<b>3</b>	<b>Análisis Estadístico</b>	<b>36</b>
3.1	Análisis exploratorio de datos . . . . .	36
3.2	Valores Atípicos . . . . .	41
<b>4</b>	<b>Análisis</b>	<b>47</b>
4.1	Descripción Espacial . . . . .	47
4.2	Desagrupamiento de los datos . . . . .	51
4.3	Variograma Experimental Omnidireccional . . . . .	55
4.4	Ajuste Variográfico . . . . .	58
4.4.1	Ajuste considerando datos originales . . . . .	58
4.4.2	Ajuste considerando datos originales y desagrupados . . . . .	60
4.4.3	Ajuste considerando datos truncados . . . . .	61
4.4.4	Ajuste considerando datos truncados desagrupados . . . . .	62
4.5	Predicción de la variable con Kriging . . . . .	63
4.5.1	Visualización de las predicciones . . . . .	65
4.5.2	Varianza de las predicciones . . . . .	66
4.5.3	Validación de los resultados de kriging . . . . .	68
4.6	Predicción con un algoritmo de simulación . . . . .	72
4.6.1	Anamorfosis Gaussiana . . . . .	72
4.6.2	Ajuste Variográfico . . . . .	76
4.6.3	Visualización de realizaciones . . . . .	78
4.6.4	Validación de la simulación . . . . .	79
4.6.5	Medición de la exactitud . . . . .	87
4.6.6	Cuantificación de la precisión . . . . .	88
<b>5</b>	<b>Conclusión</b>	<b>90</b>
<b>6</b>	<b>Glosario</b>	<b>93</b>
	<b>Bibliografía</b>	<b>96</b>

# Índice de cuadros

2.7.1	Superficie Forestal por uso, Chile . . . . .	30
2.7.2	Superficie Forestal por Región, Chile . . . . .	31
3.1.1	Rodales de Inventario Forestal . . . . .	37
3.1.2	Estadísticos Inventario Forestal, por rodal. . . . .	39
3.2.1	Estadísticos Inventario Forestal . . . . .	44
3.2.2	Resultados Test de Hipótesis . . . . .	45
3.2.3	Estadísticos Inventario Forestal . . . . .	46
4.1.1	Distancia (metros) entre las Parcelas de Inventario . . . . .	47
4.1.2	Matriz de correlación, Rodal 29 . . . . .	50
4.1.3	Matriz de correlación, Rodal 30 . . . . .	50
4.1.4	Matriz de correlación, Rodales 29 y 30 . . . . .	51
4.2.1	Resultados desagrupamiento, datos originales y datos truncados . . . . .	53
4.2.2	Estadísticos: Desagrupamiento de los datos originales con tamaño de celda con media mínima . . . . .	54
4.2.3	Estadísticos: Desagrupamiento de los datos truncados con tamaño de celda con media mínima . . . . .	55
4.3.1	Parámetros del variograma experimental omnidireccional . . . . .	55
4.4.1	Parámetros del variograma ajustado, considerando los datos originales . . . . .	59
4.4.2	Parámetros del variograma ajustado considerando datos originales desagrupados . . . . .	60
4.4.3	Parámetros del variograma ajustado considerando los datos truncados . . . . .	61
4.4.4	Parámetros del variograma ajustado considerando los datos truncados y desagrupados . . . . .	63
4.5.1	Errores de predicción obtenidos con Kriging condicional a los diferentes set de datos . . . . .	69
4.6.1	Estadísticos de la variable transformada para los diferentes set de datos. . . . .	76
4.6.2	Parámetros del variograma ajustado con datos Gaussianos . . . . .	77
4.6.3	Estadísticos datos originales, volumen de inventario y predicho por kriging y simulación . . . . .	82

---

4.6.4	Estadísticos datos originales desagrupados, volumen de inventario y predicho por kriging y simulación . . . . .	83
4.6.5	Estadísticos datos truncados, volumen de inventario y predicho por kriging y simulación . . . . .	84
4.6.6	Estadísticos datos truncados desagrupados, volumen de inventario y predicho por kriging y simulación . . . . .	85
4.6.7	Errores de predicción obtenidos con simulación por el método Espectral Bandas Rotantes condicional a los diferentes set de datos . . . . .	87



# Índice de figuras

2.3.1	Región de tolerancia $T(\mathbf{h})$ alrededor del vector $\mathbf{h}$ , caso bidimensional. (Emery, 2013, p. 31) . . . . .	14
2.3.2	Comportamiento del variograma en el origen (Sarma, 2010, p. 82): (A) parabólico, (B) lineal, (C) con discontinuidad, (D) puramente aleatorio	15
2.3.3	Gráfico de variograma (Sarma, 2010, p. 81): (A) acotado, (B) no acotado	16
2.3.4	Anisotropía (Sarma, 2010, p. 83): (A) geométrica, (B) zonal . . . . .	17
2.3.5	Modelos de Variograma (Sarma, 2010, p. 87): (A) esférico, (B) exponencial, (C) gaussiano, (D) cúbico, (E) potencia . . . . .	20
2.7.1	Ciclo Forestal . . . . .	32
3.1.1	Mapa Ubicación Parcelas Inventario Forestal ( $n =$ número de parcelas)	38
3.1.2	Histograma a) rodal 29, b) rodal 30. $n =$ número de parcelas . . . . .	39
3.1.3	Gráficos cuantil-cuantil a) rodal 29, b) rodal 30 . . . . .	40
3.2.1	Gráfico de caja, Rodales 29 y 30. . . . .	42
3.2.2	Mapa Ubicación Parcelas Inventario Forestal, con parcelas atípicas . .	43
3.2.3	(a) Histograma Rodales 29 y 30, juntos. (b) Distribución acumulada Rodales 29 y 30, juntos. . . . .	44
3.2.4	a) Diagrama de caja Rodales 29 y 30, juntos. (b) Gráfico cuantil-cuantil Rodales 29 y 30, juntos. . . . .	45
4.1.1	Ubicación parcelas de inventario forestal: Rodales 29 y 30, juntos . . .	48
4.1.2	(a) Nube direccional Norte, Rodal 29. (b) Nube direccional Este, Rodal 29 . . . . .	49
4.1.3	(a) Nube direccional Norte, Rodal 30. (b) Nube direccional Este, Rodal 30 . . . . .	50
4.1.4	(a) Nube Direccional Norte, Rodales 29 y 30. (b) Nube Direccional Este, Rodales 29 y 30 . . . . .	51
4.2.1	Desagrupamiento (a) de los datos originales, (b) de los datos truncados	54
4.3.1	Variograma experimental omnidireccional, datos originales con y sin desagrupamiento . . . . .	57
4.3.2	Variograma experimental omnidireccional, datos truncados con y sin desagrupamiento . . . . .	58

4.4.1	Variograma ajustado omnidireccional, considerando los datos originales	59
4.4.2	Variograma omnidireccional considerando los datos originales con desagrupamiento . . . . .	61
4.4.3	Variograma omnidireccional considerando los datos truncados . . . . .	62
4.4.4	Variograma omnidireccional de los datos truncados y desagrupados . .	63
4.5.1	Grilla utilizada para el Kriging . . . . .	64
4.5.2	Mapas de la predicción con kriging: (a) condicional a los datos originales, (b) condicional a los datos originales y desagrupados . . . . .	66
4.5.3	Mapas de la predicción con kriging: (a) condicional a los datos truncados, (b) condicional a los datos truncados y desagrupados . . . . .	66
4.5.4	Mapas de la varianza de kriging: (a) condicional a los datos originales (b) condicional a los datos originales desagrupados. . . . .	67
4.5.5	Mapas de la varianza de kriging: (a) condicional a los datos truncados, (b) condicional a los datos truncados desagrupados. . . . .	67
4.5.6	Histograma de errores de estimación: (a) datos originales, (b) datos originales desagrupados. . . . .	69
4.5.7	Histograma de errores de estimación: (a) datos truncados, (b) datos truncados desagrupados. . . . .	70
4.5.8	Nube de dispersión Kriging: (a) datos originales, (b) datos originales desagrupados. . . . .	71
4.5.9	Nube de dispersión Kriging: (a) datos truncados, (b) datos truncados desagrupados. . . . .	71
4.6.1	Anamorfosis Gaussiana: (a) datos originales, (b) datos originales desagrupados. . . . .	72
4.6.2	Anamorfosis Gaussiana: (a) datos truncados, (b) datos truncados desagrupados. . . . .	73
4.6.3	Histogramas: (a) valores originales, (b) valores Gaussianos. . . . .	74
4.6.4	Histogramas: (a) valores originales desagrupados. (b) valores Gaussianos. . . . .	74
4.6.5	Histogramas: (a) valores truncados (b) valores Gaussianos. . . . .	75
4.6.6	Histogramas: (a) valores truncados desagrupados, (b) valores Gaussianos. . . . .	75
4.6.7	Variograma Experimental y ajustado: (a) valores Gaussianos originales, (b) valores Gaussianos originales desagrupados. . . . .	77
4.6.8	Variograma Experimental y ajustado: (a) valores Gaussianos truncados, (b) valores Gaussianos truncados desagrupados. . . . .	78
4.6.9	Mapa de una realización: (a) simulación condicional a los datos originales, (b) simulación condicional a los datos originales desagrupados. . . . .	79
4.6.10	Mapa de una realización: (a) simulación condicional a los datos truncados, (b) simulación condicional a los datos truncados desagrupados. . . . .	79
4.6.11	Nube de dispersión de una realización versus valores reales: (a) simulación condicional a los datos originales, (b) simulación condicional a los datos originales desagrupados. . . . .	80

4.6.12	Nube de dispersión de una realización versus valores reales: (a) simulación condicional a los datos truncados, (b) simulación condicional a los datos truncados y desagrupados. . . . .	81
4.6.13	Histogramas de datos originales provenientes de: (a) inventario forestal, (b) kriging universal, (c) simulación con algoritmo Espectral Bandas Rotantes. . . . .	83
4.6.14	Histogramas de datos originales desagrupados provenientes de: (a) inventario forestal, (b) kriging universal, (c) simulación con algoritmo Espectral Bandas Rotantes. . . . .	84
4.6.15	Histogramas de datos truncados provenientes de: (a) inventario forestal, (b) kriging universal, (c) simulación con algoritmo Espectral Bandas Rotantes. . . . .	85
4.6.16	Histogramas de datos truncados desagrupados provenientes de: (a) inventario forestal, (b) kriging universal, (c) simulación con algoritmo Espectral Bandas Rotantes. . . . .	86
4.6.17	Nubes de dispersión de valores simulados promedio sobre 100 realizaciones versus valores reales, considerando el condicionamiento a los (a) datos originales, (b) datos originales desagrupados. . . . .	87
4.6.18	Nubes de dispersión de valores simulados promedio sobre 100 realizaciones versus valores reales, considerando el condicionamiento a los (a) datos truncados, (b) datos truncados desagrupados . . . . .	88
4.6.19	Diagramas de Incertidumbre (a) datos originales, (b) datos originales desagrupados. . . . .	89
4.6.20	Diagramas de Incertidumbre (a) datos truncados, (b) datos truncados desagrupados. . . . .	89

## RESUMEN

La aplicación de la metodología geoestadística para el análisis de inventarios forestales ha sido poco explorado en el sector forestal chileno. La presente tesis expone la estimación de volumen de cosecha de *Pinus radiata* ( $\text{m}^3/\text{ha}$ ) a partir de parcelas de inventario forestal georreferenciadas, localizadas en la comuna de Santa Juana, Región del Biobío. A partir del conjunto de datos disponibles, se consideran cuatro sets de datos: datos originales, datos originales desagrupados, datos truncados y datos truncados desagrupados.

La metodología empleada es kriging universal y simulación condicional con el algoritmo Espectral Bandas Rotantes. Se propone un modelo de variograma ajustado anidado compuesto por efecto pepita (*nugget effect*), un modelo esférico y un modelo potencia, el cual fue considerado tanto para la predicción con kriging como para obtener las realizaciones con el algoritmo de simulación. Se obtuvieron estimadores insesgados y de mínima varianza en los cuatro sets de datos analizados, con valores de correlación de Pearson, entre valores estimados y originales, desde 0.35 a 0.36. Por otro lado, los mapas de las realizaciones obtenidas con el algoritmo de simulación entregan información más detallada de la continuidad y variabilidad espacial, permitiendo obtener varios escenarios de la variable volumen, cuyo valor medio presenta una distribución normal. La correlación de Pearson obtenida por éste método fluctúa entre 0.23 y 0.28.

Finalmente se concluye que ambas metodologías son una alternativa para predecir sectores no muestreados por inventarios forestales.

**Keywords** – Kriging, Geostatistics, Simulation, Forestry

---

## ABSTRACT

The application of the geostatistical methodology for the analysis of forest inventories has been little explored in the Chilean forestry sector. This article exposes the estimate of *Pinus radiata* harvest volume ( $\text{m}^3/\text{ha}$ ) from forest inventory plots georeferenced, located in the Biobío Region, Chile. The methodology used is universal kriging and conditional simulation with the algorithm Spectral Turning Bands. A nested fitted variogram model is proposed composed of a nugget effect, a spherical model, and a power model, which was considered for both the prediction with kriging and to obtain realizations with the simulation algorithm. Unbiased and minimum variance estimators were obtained for the different data sets analyzed extracted from the original data set. On the other hand, the maps of the realizations obtained with the simulation algorithm provide more detailed information on continuity and spatial variability, allowing to obtain various scenarios of the volume variable, whose mean value has a normal distribution. Finally, it is concluded that both methodologies are an alternative to predict sectors not sampled by forest inventories.

**Keywords** – Kriging, Geostatistics, Simulation, Forestry

---

# Capítulo 1

## Introducción

La geoestadística tuvo sus inicios en la industria minera en 1945 y se consolidó con la tesis doctoral de G. Matheron, en 1965 (Matheron and Kleingeld, 1987, p.11). En Chile la metodología geoestadística ha estado relacionada principalmente con la minería, desde hace más de 30 años, teniendo un desarrollo incipiente en otros sectores económicos asociados a recursos naturales.

La motivación de la presente tesis de Magíster en Estadística de la Universidad de Concepción es aplicar la predicción geoestadística a recursos forestales. En Chile se han realizado escasas investigaciones en esta área (Hernández and Emery, 2009), mientras que en otros países se han planteado varias aplicaciones para apoyar la operación forestal (Chica-Olmo, 2005) y disminuir los costos de inventarios forestales (Raimundo et al., 2017; Zhang and Yao, 2008).

Chile posee una superficie de 14.7 millones de hectáreas de bosque nativo y 2.3 millones de hectáreas de plantaciones forestales. Éstas últimas son de manejo intensivo, están conformadas mayoritariamente por *Pinus radiata* y *Eucalyptus spp*

---

y tienen como objetivo principal generar productos forestales, tales como, madera aserrada, celulosa y tableros para consumo interno y externo. Una plantación forestal requiere varios años para su desarrollo, en Chile las plantaciones de *Pinus radiata* se cosechan entre los 18 y 25 años, mientras que las de *Eucalyptus spp* entre 10 y 15 años. Durante el crecimiento de la plantación es necesario realizar varias intervenciones para favorecer el desarrollo de ésta, como controles de maleza, fertilizaciones, podas y raleos (extracción de árboles). Además, periódicamente se realizan mediciones o inventarios forestales que permiten conocer el crecimiento de la plantación y estimar el volumen a obtener al momento de la cosecha. El diseño del inventario forestal determina la cantidad de parcelas por hectárea que es necesario realizar para obtener un bajo error en el muestreo, sin embargo no es factible muestrear toda la superficie, por lo cual es necesario implementar metodologías que permitan predecir el volumen en las zonas no muestreadas.

La geoestadística es una disciplina que estudia fenómenos regionalizados (es decir, fenómenos que presentan una estructura espacial), dado que el variograma o la función de covarianza entrega información acerca de la continuidad y variabilidad espacial de la variable de interés. La geoestadística provee herramientas fundamentales para la predicción de variables regionalizadas, como ley de un mineral, contenido de contaminantes, precipitación caída en una zona, etc.

Debido a la gran cantidad de superficie forestal presente en Chile y la información que provee el análisis geoestadístico acerca de la continuidad espacial de una variable, el tema de la presente tesis tiene relevancia debido a la aplicabilidad de la metodología desarrollada tanto en el sector público como privado.

El objetivo de la presente tesis es utilizar diferentes metodologías geoestadísticas para estimar volumen de cosecha de *Pinus radiata* ( $m^3/ha$ ), a partir de parcelas de

inventario forestal georreferenciadas. Se trabajará con una base de datos obtenida de inventarios forestales de una plantación de *Pinus radiata*, establecida el año 2004 en un predio localizado en la comuna de Santa Juana, Región del Biobío. Se disponen de 244 parcelas de inventario.

Como se señaló anteriormente, la variable con la cual se trabajará durante el estudio es el volumen total de cosecha en  $\text{m}^3/\text{ha}$ . Con el fin de predecir la variable de interés en toda la superficie, se hará uso de la técnica conocida como kriging, el cual predice el valor de una variable regionalizada en sitios donde no hay información, a partir de los datos ya existentes y la continuidad espacial que existe entre ellos. La predicciones realizadas con kriging, interpolan el valor en un sitio no muestreado entregando una predicción insesgada y con mínima varianza del error. Estos procesos tienen ciertas limitaciones, siendo una de éstas el suavizamiento que genera en las variables predichas. Para hacer frente a esta problemática se utilizan simulaciones, que permiten generar diversos escenarios equiprobables de la variable de interés, capturando la variabilidad espacial e incertidumbre que mediante el kriging no son considerados. Es así, que en este trabajo también se consideran simulaciones con el algoritmo Espectral Bandas Rotantes presentado por [Arroyo and Emery \(2018\)](#). Finalmente se entregan los resultados y validaciones para ambos métodos: kriging y simulación.

Los resultados de la predicción de volumen forestal aplicado con herramientas geoestadísticas se presentan en los siguientes capítulos:

**Capítulo 1:** Marco Teórico - Se presentan los fundamentos teóricos de geoestadística: variable regionalizada, variograma, kriging, simulación geoestadística condicional y no condicional. En kriging y simulación el enfoque se centra en las metodologías aplicadas en la presente tesis, kriging Universal y simulación con el algoritmo

Espectral Bandas Rotantes. Además, se exponen antecedentes de la temática forestal relacionada con el alcance de la tesis.

**Capítulo 2:** Análisis Estadístico - Se desarrolla el análisis exploratorio de datos de las parcelas de inventario forestal.

**Capítulo 3:** Análisis Geoestadístico - Contiene la descripción espacial de los datos y la aplicación de la metodología geoestadística para realizar la predicción en sitios no muestreados. Se muestran los resultados obtenidos con kriging y con el algoritmo de simulación Espectral Bandas Rotantes, y la validación de éstos.

**Capítulo 4:** Conclusiones - Se presentan las conclusiones y análisis desprendidos de la realización del trabajo de la tesis.



# Capítulo 2

## Marco Teórico

### 2.1. Geoestadística



La geoestadística tiene como objetivo proporcionar descripciones cuantitativas de variables naturales, distribuidos en el espacio o, en el tiempo y espacio (Chilès and Delfiner, 2012, p. 1). Ejemplos de tales variables son:

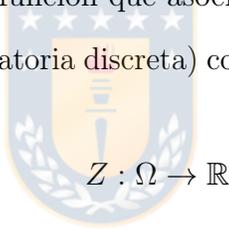
- Leyes de mineral en un depósito mineral.
- Profundidad y espesor de una capa geológica.
- Porosidad y permeabilidad en un medio poroso.
- Densidad de árboles de una determinada especie en un bosque.
- Propiedades del suelo en una región.
- Precipitación sobre una zona de captación.
- Presión, temperatura y velocidad del viento en la atmósfera.

- Concentraciones de contaminantes en un sitio contaminado.

La Geoestadística tuvo sus inicios en Sudáfrica con el trabajo de H.S Sichel, con una aplicación de la distribución log-normal a una mina de oro. Este fue seguido por la famosa contribución de D.G. Krige y se consolidó con la tesis doctoral de G. Matheron en 1965, con el desarrollo de la “Teoría de variables regionalizadas”.

### 2.1.1. Variable Regionalizada

En teoría de probabilidades se define el espacio muestral, que se denota  $\Omega$ , como el conjunto de resultados posibles de un experimento, cada elemento se denota como  $\omega$ . Una variable aleatoria es una función que asocia un número real (variable aleatoria continua) o entero (variable aleatoria discreta) con cada elemento del espacio muestral:


$$Z : \Omega \rightarrow \mathbb{R}$$

Para variables medioambientales, como ph del suelo, precipitación caída sobre una zona o cantidad de árboles por hectárea, existen múltiples factores conocidos y/o desconocidos que determinan el valor de la variable, por lo que es factible considerar dicha variable como aleatoria. La notación considera  $Z$  como variable aleatoria y  $z$  la realización de dicha variable. Además, en estadística espacial la variable aleatoria tiene una localización que puede ser uni, bi o tridimensional, que se denotará por  $\mathbf{x}$ . Los datos considerados en la presente tesis son bidimensionales, es decir  $\mathbf{x} = (x_1, x_2)$ . Por lo tanto,  $Z(\mathbf{x})$  corresponde a la variable aleatoria en el lugar  $\mathbf{x}$  y  $z(\mathbf{x})$  el valor de  $Z$  en la ubicación  $\mathbf{x}$ .

El término variable regionalizada fue acuñado por [Matheron \(1963\)](#) para enfatizar dos aspectos ([Sarma, 2010](#), p. 78):

- Un aspecto aleatorio que cuenta con variaciones locales.
- Un aspecto estructurado que refleja una tendencia a gran escala del fenómeno.

Algunas características importantes de las variables regionalizadas se detallan a continuación (Matheron, 1963, p. 1249):

- Es localizada.
- Puede mostrar mayor o menor continuidad en su variación espacial. En algunas circunstancias, esta continuidad no será confirmada y se habla de “efecto pepita”.
- Puede mostrar diferentes tipos de anisotropías, es decir, en una dirección no muestra una variación significativa, mientras que en otra dirección varía rápidamente.

En un dominio dado  $D$ , al haber muchos puntos  $\mathbf{x}$ , se obtiene un conjunto de variables aleatorias  $\{Z(\mathbf{x}), \mathbf{x} \in D\}$  que forman una función aleatoria o proceso estocástico.

La geoestadística asocia la aleatoriedad con la variable regionalizada en sí, mediante el uso de un modelo estocástico en el que la variable regionalizada se considera una de las muchas posibles realizaciones de una función aleatoria (Chilès and Delfiner, 2012, p. 3). En la literatura una función aleatoria es también llamada proceso estocástico cuando  $\mathbf{x}$  varía en un espacio 1D (tiempo) y es definido como un campo aleatorio cuando  $\mathbf{x}$  varía en un espacio de más de una dimensión (Chilès and Delfiner, 2012, p. 12).

### 2.1.2. Función aleatoria

Una función aleatoria es descrita por sus distribuciones finito dimensionales. El conjunto de todas las distribuciones multidimensionales de  $k$ -tuplas  $(Z(x_1), Z(x_2), \dots, Z(x_k))$  para todos los valores finitos de  $k$  y todas las configuraciones de los puntos  $x_1, x_2, \dots, x_k$ . Para abreviar, se denominará a esto, distribución espacial (Chilès and Delfiner, 2012, p. 12).

“Como los datos disponibles de la variable regionalizada son finitos, no es posible inferir la distribución de probabilidad completa de la función aleatoria, que contiene un número infinito de distribuciones de probabilidad” (Emery, 2013, p.23), por lo tanto se calculan los momentos de la distribución, para resumir la información relevante (Ecuaciones 2.1.1 a 2.1.5).



**Esperanza matemática (momento de orden 1):**

$$E[Z(\mathbf{x})] = m(\mathbf{x}). \quad (2.1.1)$$

**Varianza (momento de orden 2):**

$$Var[Z(\mathbf{x})] = E[Z(\mathbf{x}) - m(\mathbf{x})]^2. \quad (2.1.2)$$

**Covarianza (momento de orden 2)**, centrada entre dos variables aleatorias

$Z(\mathbf{x}_i)$  y  $Z(\mathbf{x}_j)$ :

$$\begin{aligned} Cov[Z(\mathbf{x}_i), Z(\mathbf{x}_j)] &= E[Z(\mathbf{x}_i) - m(\mathbf{x}_i)] * [Z(\mathbf{x}_j) - m(\mathbf{x}_j)] \\ &= E[Z(\mathbf{x}_i)Z(\mathbf{x}_j)] - m(\mathbf{x}_i) * m(\mathbf{x}_j). \end{aligned} \quad (2.1.3)$$

**Correlograma** o coeficiente de correlación lineal entre dos variables aleatorias  $Z(\mathbf{x}_i)$  y  $Z(\mathbf{x}_j)$ :

$$\rho(\mathbf{x}_i, \mathbf{x}_j) = \text{corr}[Z(\mathbf{x}_i), Z(\mathbf{x}_j)] = \frac{\text{cov}[Z(\mathbf{x}_i), Z(\mathbf{x}_j)]}{\sqrt{\text{var}[Z(\mathbf{x}_i)]\text{var}[Z(\mathbf{x}_j)]}}. \quad (2.1.4)$$

**Semi-variograma (momento de orden 2):**

$$\gamma(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{2}\text{Var}[Z(\mathbf{x}_i) - Z(\mathbf{x}_j)]. \quad (2.1.5)$$

## 2.2. Covarianza Espacial

La función covarianza permite medir el grado de similitud entre dos variables  $z_i$  y  $z_j$  ubicadas en  $\mathbf{x}_i$  y  $\mathbf{x}_j$ , respectivamente. Para determinar la relación espacial entre las variables aleatorias, de acuerdo a la covarianza (ver Ecuación 2.1.3) se requiere, el valor de la variable en dos sitios y las medias en ambos sitios. Sin embargo solo es posible obtener una medición en ambos sitios, por lo cual es necesario establecer los supuestos que soportan el modelo geoestadístico.

### 2.2.1. Estacionariedad

En un proceso estocástico se consideran como estacionarios aquellos procesos que tienen un comportamiento constante a lo largo del tiempo y/o espacio. En el modelo geoestadístico se consideran las siguientes definiciones de estacionariedad:

- Estacionariedad estricta
- Estacionariedad de orden 2
- Estacionariedad intrínseca

#### 2.2.1.1. Estacionariedad estricta

Establece que la distribución espacial de una función aleatoria es invariante bajo una traslación arbitraria de los puntos por un vector  $\mathbf{h}$ :

$$Pr\{Z(\mathbf{x}_1) < z_1, \dots, Z(\mathbf{x}_k) < z_k\} = Pr\{Z(\mathbf{x}_1 + \mathbf{h}) < z_1, \dots, Z(\mathbf{x}_k + \mathbf{h}) < z_k\}. \quad (2.2.1)$$

Físicamente, esto significa que el fenómeno es homogéneo en el espacio, es decir, se repite en todo el espacio (Chilès and Delfiner, 2012, p. 16).

### 2.2.1.2. Estacionariedad de orden 2

Cuando una función aleatoria es estacionaria, sus momentos, si existen, son invariantes bajo traslación (Chilès and Delfiner, 2012, p. 17). Si se consideran sus primeros dos momentos únicamente, se tienen para los puntos  $\mathbf{x}$  y  $\mathbf{x} + \mathbf{h}$  de  $\mathbb{R}^d$ :

**Esperanza matemática:**

$$E[Z(\mathbf{x})] = m, \quad \forall \mathbf{x} \in \mathbb{R}^d \quad d = 1, 2 \text{ ó } 3 \quad (2.2.2)$$

**Covarianza:**

$$Cov[Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h})] = E[Z(\mathbf{x}) - m][Z(\mathbf{x} + \mathbf{h}) - m] = C(\mathbf{h}). \quad (2.2.3)$$

La media es constante y la función de covarianza sólo depende de la distancia de separación o lag,  $\mathbf{h}$ .

Por definición, una función aleatoria que satisface las condiciones anteriores (Ecuaciones 2.2.2 y 2.2.3) es *estacionaria de segundo orden* o también llamada débilmente estacionaria, o estacionaria en sentido amplio. Una función aleatoria estacionaria es *isotrópica* si su función de covarianza sólo depende de la longitud  $|\mathbf{h}|$  del vector  $\mathbf{h}$  y no de su orientación (Chilès and Delfiner, 2012, p. 17).

Además, se tiene que la varianza es constante e igual a la covarianza evaluada en el origen y el semi-variograma solo depende del vector de separación:

**Varianza:**

$$\text{Var}[Z(\mathbf{x})] = E[Z(\mathbf{x}) - m]^2 = C(0) \quad \forall \mathbf{x} \in \mathbb{R}^d \quad d = 1, 2 \text{ ó } 3 \quad (2.2.4)$$

**Semi-variograma:**

$$\gamma(\mathbf{x}, \mathbf{x} + \mathbf{h}) = \frac{1}{2} E \{ [Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})]^2 \} = C(0) - C(h) \quad \forall \mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^d \quad d = 1, 2 \text{ ó } 3 \quad (2.2.5)$$

### 2.2.1.3. Estacionariedad intrínseca

Se asume que para cada vector  $\mathbf{h}$ , el incremento  $Y_{\mathbf{h}}(\mathbf{x}) = Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})$  es una función aleatoria estacionaria de segundo orden en  $\mathbf{x}$  (Chilès and Delfiner, 2012, p. 17). Entonces  $Z(\mathbf{x})$  se denomina función aleatoria intrínseca (*Intrinsic Random Field*, abreviado como IRF) y se caracteriza por las siguientes relaciones (Ecuaciones 2.2.6, 2.2.7):

$$m(\mathbf{h}) = E[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] = \langle a, \mathbf{h} \rangle, \quad (2.2.6)$$

$$\text{Var}[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] = 2\gamma(\mathbf{h}), \quad (2.2.7)$$

donde  $\langle a, \mathbf{h} \rangle$  es la deriva lineal de la función aleatoria intrínseca (IRF) y  $\gamma(\mathbf{h})$  es la función variograma.

En este caso, el semi-variograma, también denominado variograma, ya no depende solo del vector de separación:

**Semi-variograma:**

$$\text{Var}[Z(\mathbf{x}_i) - Z(\mathbf{x}_j)] = \gamma(\mathbf{x}_i, \mathbf{x}_j) \quad \forall \mathbf{x} \in \mathbb{R}^d \quad d = 1, 2 \text{ ó } 3. \quad (2.2.8)$$

En la Sección 2.3 se explica en mayor detalle el semi-variograma, que por simplicidad de ahora en adelante se mencionará solo como variograma.

## 2.3. Variograma



El variograma, describe estadísticamente cómo los valores en dos puntos se vuelven diferentes a medida que aumenta la separación entre estos puntos. “El variograma es la forma más sencilla de relacionar la incertidumbre con la distancia de una observación” (Chilès and Delfiner, 2012, p. 4).

Las propiedades matemáticas del variograma son (Emery, 2013, p. 43):

- Paridad:

$$\gamma(\mathbf{h}) = \gamma(-\mathbf{h})$$

- Nulidad en el origen:

$$\gamma(\mathbf{0}) = 0$$

- Positividad:

$$\mathbf{h} \geq 0$$

- Comportamiento en el infinito:

$$\lim_{|\mathbf{h}| \rightarrow +\infty} \gamma(\mathbf{h})/|\mathbf{h}|^2 = 0$$

- Función de tipo *negativo condicional*:

$$\forall k \in \mathbf{N}^*, \forall \lambda_1, \dots, \lambda_k \in \mathbb{R}, \text{ tal que } \sum_{i=1}^k \lambda_i = 0, \quad \forall \mathbf{x}_1, \dots, \mathbf{x}_k \in D,$$

$$\sum_{i=1}^k \sum_{j=1}^k \lambda_i \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) \leq 0.$$

La última propiedad es una condición necesaria y suficiente para que  $\gamma$  sea el variograma de una función aleatoria. El adjetivo *condicional* se refiere al hecho de que la desigualdad es válida solo para una clase restringida de ponderadores  $\{\lambda_1, \dots, \lambda_k\}$ , los de suma total nula.

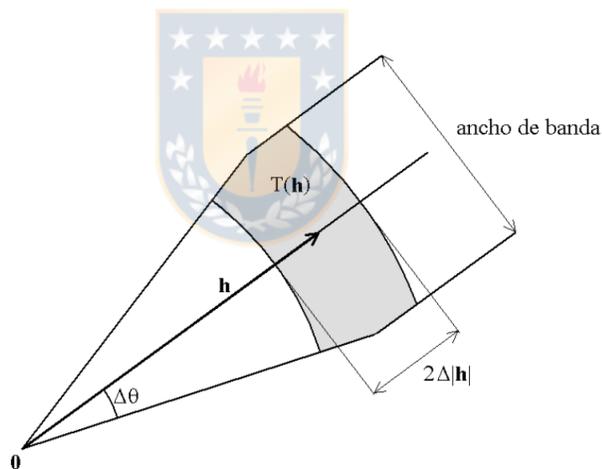
La Ecuación 2.3.1 es el estimador tradicional del variograma, el cual se calcula a partir de los datos disponibles y es más conocido como **variograma experimental**.

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2p(\mathbf{h})} \sum_{i=1}^{p(\mathbf{h})} \{z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})\}^2, \quad (2.3.1)$$

donde  $z(\mathbf{x}_i)$  y  $z(\mathbf{x}_i + \mathbf{h})$  son los valores de  $z$  observados en las ubicaciones  $\mathbf{x}_i$  y  $(\mathbf{x}_i + \mathbf{h})$ , y  $p(\mathbf{h})$  es el número de datos pareados para la distancia (lag o paso)  $\mathbf{h}$ .

El variograma experimental se puede calcular para diferentes distancias  $\mathbf{h}$  y direcciones, además como en general los puntos no están distribuidos de manera regular en el espacio, se establecen tolerancias para distancias y direcciones, como se representa en la Figura 2.3.1 (Emery, 2013, p. 31; Ortiz, 2004, p. 46), donde:

- $T(\mathbf{h})$  tolerancia del paso o distancia, de manera que los puntos puedan encontrarse a una distancia mayor o menor al paso. Por defecto, este valor es 0.5 veces el paso  $\mathbf{h}$ .
- $\theta$  corresponde al acimut o dirección de interés, medida en un plano horizontal respecto al Norte, en el sentido de los punteros del reloj.
- $\Delta\theta$  es la tolerancia del acimut.
- $\Delta\mathbf{h}_H$  Ancho de banda en el acimut, corresponde a una banda dentro de la cual se consideran válidos los datos para el cálculo del variograma. Se mide perpendicular a la dirección del acimut.



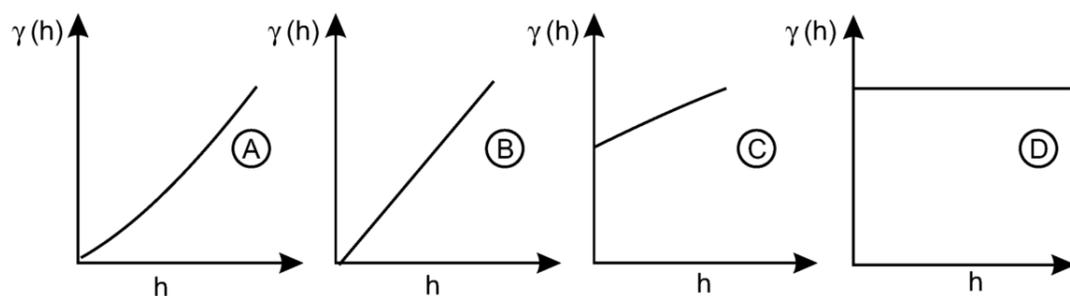
**Figura 2.3.1:** Región de tolerancia  $T(\mathbf{h})$  alrededor del vector  $\mathbf{h}$ , caso bidimensional. (Emery, 2013, p. 31)

Puesto que el variograma experimental es un estimador del variograma, se requiere ajustarlo a un variograma teórico o modelo de variograma, esto se denomina **variograma ajustado o modelado**. La acuciosidad en el cálculo del variograma experimental y en el posterior ajuste de éste son fundamentales para obtener predicciones insesgadas y de varianza mínima con el kriging.

### 2.3.1. Comportamiento en el origen

“La manera cómo el variograma se aproxima al origen está determinada por la continuidad (o carencia de continuidad) de la variable  $Z(\mathbf{x})$ ” (Webster and Oliver, 2007, p. 80). La Figura 2.3.2, extraída de Sarma (2010, p. 82) presenta los distintos comportamientos del variograma en el origen:

- A) Forma parabólica: la variable regionalizada es muy regular en el espacio.
- B) Forma lineal: la variable es continua, pero menos que la de forma parabólica.
- C) Discontinuidad en el origen: la variable es altamente irregular a cortas distancias. El salto en el origen se denomina efecto pepita y se utiliza para describir variabilidad de corto alcance, incluso cuando puede deberse a otros factores como errores de medición, errores en la ubicación, etc. (Sarma, 2010, p. 82).
- D) Curva plana: “Una curva paralela al eje horizontal, representa aleatoriedad pura o ruido blanco. Las variables regionalizadas no están correlacionadas para cualquier  $h$ . Este es el caso límite de una total carencia de estructura”(Sarma, 2010, p. 82).



**Figura 2.3.2:** Comportamiento del variograma en el origen (Sarma, 2010, p. 82): (A) parabólico, (B) lineal, (C) con discontinuidad, (D) puramente aleatorio

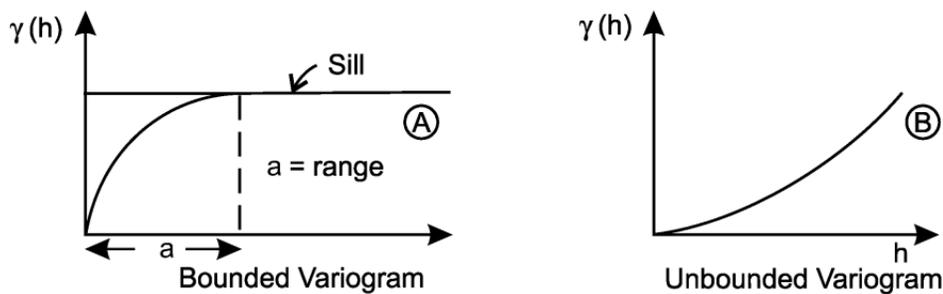
### 2.3.2. Gráfico del Variograma

El gráfico del variograma es de gran interés ya que resume la continuidad espacial de la variable regionalizada. La Figura 2.3.3 (Sarma, 2010, p. 81) muestra los dos tipos de gráficos que es factible encontrar, acotado y no acotado, se describen a continuación:

- A) Variograma acotado: representa un fenómeno estacionario, en el cual el variograma crece hasta un límite denominado meseta (*sill*) y la distancia a la cual se alcanza la meseta se denomina alcance o rango (*range*). “Además se demuestra que la meseta es igual a la varianza *a priori*” (Emery2013, p. 44).

Se considera que hasta la distancia “*a*” denominada alcance, existe correlación entre  $Z(\mathbf{x})$  y  $Z(\mathbf{x} + \mathbf{h})$ , el alcance corresponde a la zona de influencia.

- B) Variograma no acotado, representa fenómenos no estacionarios, en este caso el gráfico no presenta meseta ni alcance. “En este caso, la varianza *a priori* es infinita y no existe la función de covarianza ni el correlograma” (Emery2013, p. 45).



**Figura 2.3.3:** Gráfico de variograma (Sarma, 2010, p. 81): (A) acotado, (B) no acotado

### 2.3.3. Anisotropía

Cuando un variograma no cambia con la dirección, se denomina isotrópico. Es entonces una función del módulo del vector ( $\mathbf{h}$ ). Si lo anterior no ocurre, el variograma es anisotrópico (Chilès and Delfiner, 2012, p. 55). Dos típicos casos de anisotropía se describen a continuación:

- A) Anisotropía geométrica: la meseta es constante, pero el rango o alcance varía con la dirección.
- B) Anisotropía zonal: el variograma despliega una menor meseta en una dirección específica.

La Figura 2.3.4 describe ambos tipos de anisotropías (Sarma, 2010, p. 83). En la práctica para determinar si existen anisotropías se calculan variogramas en diferentes direcciones y se representan en un mismo gráfico.

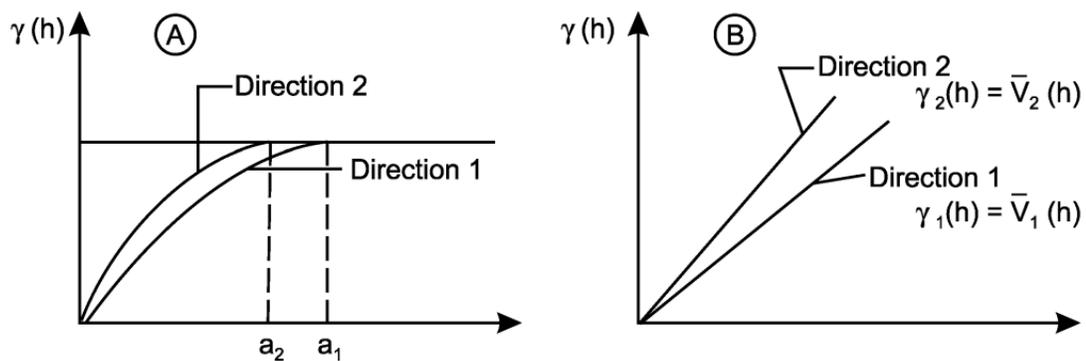


Figura 2.3.4: Anisotropía (Sarma, 2010, p. 83): (A) geométrica, (B) zonal

### 2.3.4. Modelos de Variograma

“Una función es un variograma si y sólo si es de tipo negativo condicional. Es una propiedad muy restrictiva por lo que en general, se elige un modelo de variograma entre las funciones que son de tipo negativo condicional” (Emery, 2013, p. 47). A continuación, se presentan los modelos de variogramas comúnmente utilizados (Sarma, 2010, p. 87).

A) Modelo Esférico: Es el mas comúnmente usado, describe una variable regionalizada continua. Este modelo exhibe un comportamiento lineal en el origen y alcanza su meseta en la distancia conocida como alcance (Figura 2.3.5A). El variograma esférico de alcance  $a$  y meseta  $C$  se define como (Emery, 2013, p. 48):

$$\gamma(\mathbf{h}) = \begin{cases} C \left[ \frac{3}{2} \frac{|\mathbf{h}|}{a} - \frac{1}{2} \left( \frac{|\mathbf{h}|}{a} \right)^3 \right] & \text{si } |\mathbf{h}| \leq a \\ C & \text{en caso contrario} \end{cases} \quad (2.3.2)$$

B) Modelo Exponencial: Solo alcanza la meseta asintóticamente cuando  $\mathbf{h} \rightarrow \infty$ , el alcance práctico es  $3a$ , cuando el variograma llega al 95 % de su meseta (Chilès and Delfiner, 2012, p. 88). En la Figura 2.3.5B se aprecia visualmente este modelo.

$$\gamma(\mathbf{h}) = C \left\{ 1 - \exp \left( -\frac{|\mathbf{h}|}{a} \right) \right\} \quad (2.3.3)$$

C) Modelo Gaussiano: “Representa un fenómeno extremadamente continuo”(Sarma, 2010, p. 86), tiene comportamiento parabólico en el origen, la meseta se alcanza

asintóticamente y el alcance práctico es  $1.73a$  (Figura 2.3.5C).

$$\gamma(\mathbf{h}) = C \left\{ 1 - \exp \left( -\frac{|\mathbf{h}|^2}{a^2} \right) \right\} \quad (2.3.4)$$

D) Modelo Cúbico: Modelo parabólico en el origen, se aproxima con mayor gradualidad a la meseta que el variograma esférico, como se aprecia en la Figura 2.3.5D.

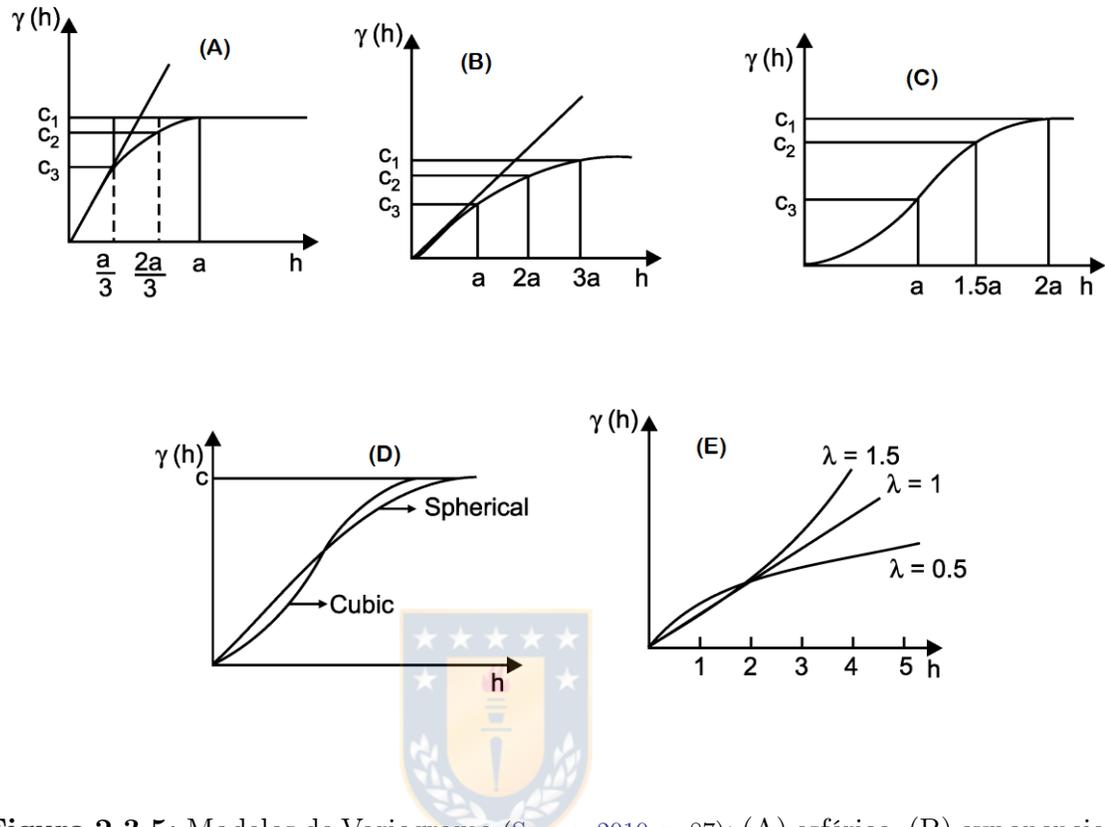
$$\gamma(\mathbf{h}) = \begin{cases} C \left[ 7\frac{\mathbf{h}^2}{a^2} - \frac{35\mathbf{h}^3}{4a^3} + \frac{7\mathbf{h}^5}{2a^5} - \frac{3\mathbf{h}^7}{4a^7} \right] & \text{si } |\mathbf{h}| \leq a \\ C & \text{en caso contrario} \end{cases} \quad (2.3.5)$$

E) Modelo Potencia: Representa una variable regionalizada no estacionaria, por lo tanto no posee meseta ni alcance. Se define de la siguiente forma:

$$\gamma(\mathbf{h}) = \omega |\mathbf{h}|^\lambda, \quad \text{con } 0 < \lambda < 2, \quad (2.3.6)$$

donde  $\omega$  = pendiente y  $\lambda$  = exponente.

La Figura 2.3.5E muestra el modelo potencia con diferentes valores de  $\lambda$ . Cuando  $\lambda = 1$  el variograma es lineal y para valores mayores la curva del variograma es parabólica.



**Figura 2.3.5:** Modelos de Variograma (Sarma, 2010, p. 87): (A) esférico, (B) exponencial, (C) gaussiano, (D) cúbico, (E) potencia

## 2.4. Desagrupamiento

Cuando los datos muestreados están agrupados puede afectar la utilidad del variograma. Chilès and Delfiner (2012, p. 45) señala que cuando no es posible una subdivisión clara en dominios homogéneos, es necesario realizar desagrupamiento de los datos. Consiste en asignar un peso específico  $w_\alpha$  a cada dato  $z_\alpha$  y aquellos datos más agrupados tendrán un peso  $w_\alpha$  menor que los datos regularmente localizados. Existen varias técnicas de desagrupamiento que permiten calcular los pesos o ponderadores. En el presente estudio se considera el método de desagrupamiento de las celdas, el cual se ejecutó con el software GSLIB (*Geostatistical Software Library*). La superficie

bajo análisis se divide en celdas idénticas y se asigna a cada celda un ponderador inversamente proporcional al número de muestras contenidas en la celda.

La Sección 4.2.1 detalla los resultados obtenidos con los datos de la tesis.

## 2.5. Kriging

Kriging es el método geoestadístico de predicción. La correlación espacial permite predecir valores para sitios no muestreados desde los datos muestreados. Si las muestras no están correlacionadas en el espacio, el mejor estimador de  $Z(\mathbf{x})$  sería la media global  $E[Z(\mathbf{x})] = m$ . El kriging usa el variograma (Ecuación 2.2.5) o la función de covarianza (Ecuación 2.2.3) como modelo de la continuidad geoespacial y estima los puntos no muestreados sobre esta base (Zhang and Yao, 2008, p. 183).

La predicción por kriging es una suma lineal ponderada de los datos. Las predicciones se pueden realizar para puntos o bloques, usando kriging puntual o kriging de bloques, respectivamente.

### Kriging puntual

Supongamos que los valores de una variable aleatoria  $Z$  han sido registrados y muestreados en los puntos  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ , obteniéndose  $N$  datos,  $z(\mathbf{x}_i)$ ,  $i = 1, 2, \dots, N$ . Para kriging puntual interesa predecir la variable aleatoria  $Z$  para cualquier punto nuevo  $x_0$  mediante la siguiente ecuación:

$$\hat{Z}(\mathbf{x}_0) = a + \sum_{i=1}^N \lambda_i Z(\mathbf{x}_i), \quad (2.5.1)$$

donde  $\lambda_i$  y  $a$  son las incógnitas, y  $Z(\mathbf{x}_i)$  son los datos conocidos o muestreados.

Además, se debe cumplir con la restricción de insesgo, es decir el error de estimación tiene esperanza nula:

$$E \left[ \hat{Z}(\mathbf{x}_0) - z(\mathbf{x}_0) \right] = 0. \quad (2.5.2)$$

La varianza de predicción debe ser mínima (Restricción de optimalidad):

$$Var \left[ \hat{Z}(\mathbf{x}_0) \right] = E \left[ \{ \hat{Z}(\mathbf{x}_0) - z(\mathbf{x}_0) \}^2 \right]. \quad (2.5.3)$$

## Kriging de bloques



“El kriging en bloque se utiliza cuando es necesario predecir valores en áreas más grandes que el soporte de los datos” (Oliver and Webster, 2015, p. 45).

El kriging de bloques es de interés para varios dominios de aplicación, como la evaluación de recursos minerales (estimar leyes de mineral sobre unidades selectivas de explotación) o las ciencias medio-ambientales (estimar la concentración de contaminantes sobre unidades de remediación). Para que los cálculos tengan un sentido físico, es necesario que la variable estudiada sea aditiva (Emery, 2013, p. 73).

### 2.5.1. Tipos de Kriging

De acuerdo a los supuestos, los principales tipos de kriging son:

- **Kriging simple** - se asume que la media de la variable regionalizada es conocida y el variograma es estacionario.

- **Kriging ordinario** o de media desconocida - se basa en el supuesto que la variación es aleatoria y espacialmente dependiente, y que el proceso aleatorio subyacente es intrínsecamente estacionario con media constante y varianza que depende solo de la distancia de separación y dirección entre los puntos y no de las posiciones absolutas (Oliver and Webster, 2015, p. 44).
- **Kriging con deriva** - se sustenta en la hipótesis que la función aleatoria ya no es estacionaria y su valor esperado o deriva, que se denota como  $m(\mathbf{x})$  varía en el espacio, reflejando una tendencia sistemática en la distribución espacial de los valores (Emery, 2013, p. 72). Según la forma de la deriva, se dividen en: kriging universal, kriging trigonométrico y kriging con deriva externa. Se explicará en mayor detalle el kriging universal, dado que es el tipo de kriging que se considera en la presente tesis.
  - **Kriging Universal** - es un método de estimación de una variable regionalizada no estacionaria, es decir cuando hay tendencia o deriva. Supone que la deriva es un polinomio de las coordenadas.

Previo a la realización del kriging o predicción de la variable de interés es necesario realizar las siguientes actividades:

- Análisis exploratorio de datos.
- Análisis de outliers.
- Cálculo del variograma experimental en una o varias direcciones.
- Ajuste del variograma experimental a un modelo teórico para obtener el variograma ajustado.
- Generación de la grilla (puntos no muestreados) para definición del sector a

predecir.

### 2.5.2. Validación Cruzada

Isaaks and Srivastava (1989, p. 351) define validación cruzada como una “técnica que permite comparar valores verdaderos y estimados usando solo la información disponible en el set de datos”. El método consiste en estimar, mediante kriging un dato a partir de todos los restantes, por lo tanto permite calcular el error de estimación en cada sitio y además realizar un análisis estadístico de los errores cometidos en todos los sitios con datos (Emery, 2013, p. 78).

El método de validación cruzada permite evaluar distintos tipos de kriging y/o diferentes modelos de variograma. En la presente tesis se consideran distintos modelos de variogramas con 4 set de datos: datos originales, datos originales desagrupados, datos truncados, datos truncados desagrupados, los cuales se evaluarán mediante validación cruzada.

## 2.6. Simulación geoestadística

Una simulación consiste en construir una variable regionalizada parecida a la variable de estudio, es decir, que reproduce su continuidad espacial y también coincide con los valores medidos en los sitios con datos. Se recurre a las técnicas de simulación porque el conocimiento exhaustivo de la variable real es imposible, por falta de tiempo y presupuesto o a causa de obstáculos naturales. Estas técnicas se apoyan en la interpretación de la variable aleatoria como una realización de una función aleatoria y en el modelamiento de su distribución espacial (Emery, 2013, p. 95).

Desde el punto de vista de las aplicaciones, existen 3 categorías que se pueden

simular (Chilès and Delfiner, 2012, p. 486):

1. Variables Continuas: incluye mediciones de propiedades físicas, tales como grado de un mineral, espesor de una capa geológica, o la velocidad del sonido.
2. Variables categóricas, son variables discretas que representan elementos de una clasificación.
3. Objetos, representan entidades del dominio de la aplicación, cuya morfología se captura como un todo y se simula en una sola toma, ejemplo una fractura.

Las propiedades de la simulación son (Emery, 2013, p. 96):

- **Interpolación exacta:** cuando la simulación es condicional.
- **Insesgo:** la variable regionalizada o, mas generalmente, una función de esta variable puede ser estimada sin sesgo por la misma función aplicada a los valores simulados.
- **No suaviza:** la dispersión de los valores simulados es la misma que la dispersión de los verdaderos.
- **No es precisa:** el error entre valor real y valor simulado no tiene varianza mínima.

Las simulaciones permiten cuantificar la incertidumbre asociada al desconocimiento de los valores reales de una variable regionalizada, mediante la construcción de diversos escenarios equiprobables de dicha variable. Según se consideren o no los datos originales al momento de simular se hablará de simulación condicional o no condicional, respectivamente.

### 2.6.1. Simulación No Condicional

Una simulación no condicional de una función aleatoria  $\{Z(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^d\}$  es simplemente una realización de  $Z(\mathbf{x})$ , aleatoriamente seleccionada de un conjunto de realizaciones posibles. Esta construcción requiere el conocimiento de la distribución de la función aleatoria  $Z(\mathbf{x})$ , al menos implícitamente (Chilès and Delfiner, 2012, p. 479).

### 2.6.2. Simulación Condicional

En la simulación geoestadística condicional se construye la función aleatoria con la misma distribución espacial que la variable regionalizada y conociendo los valores que debe tomar en los sitios muestreados. Las simulaciones condicionales son útiles cualitativamente, para obtener imágenes realistas de la variabilidad espacial, el objetivo es imitar la variación espacial tan similarmente como sea posible. Es una realización aleatoriamente seleccionada desde el subconjunto de realizaciones que coinciden con los puntos de muestra (Chilès and Delfiner, 2012, p. 482). La cantidad de simulaciones a realizar depende del objetivo y de la estructura del fenómeno.

### 2.6.3. Algoritmos de simulación

Existen varios algoritmos que permiten simular funciones aleatorias. Dentro de éstos es posible distinguir los que condicionan de forma directa (por ejemplo, el método secuencial (Chilès and Delfiner, 2012, p. 489) y el de descomposición matricial (Chilès and Delfiner, 2012, p. 493), y los que no lo hacen y requieren una etapa adicional de condicionamiento, por ejemplo el método Espectral Bandas Rotantes (Emery et al., 2016, p. 1590). A continuación se describirá el algoritmo de simulación

que se utilizará en la presente tesis. La simulación de una función aleatoria intrínseca de orden  $k$  (IRF- $k$ ) se puede realizar de manera eficiente con el algoritmo Espectral Bandas Rotantes (Arroyo and Emery, 2018) y luego aplicar un procesamiento de condicionalidad mediante kriging. La mayoría de los otros algoritmos de simulación multigaussianos fallan o son aproximados cuando se aplican a campos aleatorios no estacionarios.

### 2.6.3.1. Algoritmo Espectral de Bandas Rotantes

Este método es un caso particular del algoritmo Bandas Rotantes ((Lantuéjoul, 2002; Emery and Lantuéjoul, 2006)), donde las simulaciones uno-dimensionales son obtenidas utilizando cosinusoides basado en la representación espectral de la función de covarianza (Arroyo and Emery, 2018). Si denotamos  $Y_s(\mathbf{x})$  como el campo aleatorio simulado, tenemos:

$$\forall \mathbf{x} \in \mathbb{R}^d, Y_s(\mathbf{x}) = \frac{1}{\sqrt{L}} \sum_{l=1}^L \sqrt{\frac{2f(\mathbf{U}_l)}{g(\mathbf{U}_l)}} \cos(2\pi\langle \mathbf{x}, \mathbf{U}_l \rangle + \phi_l), \quad (2.6.1)$$

donde  $L$  es un número entero grande,  $\{\mathbf{U}_l : l = 1, \dots, L\}$  son vectores aleatorios mutuamente independientes con densidad de probabilidad  $g$ ,  $\{\phi_l : l = 1, \dots, L\}$  son variables aleatorias mutuamente independientes distribuidas uniformemente en  $[0, 2\pi]$ , independientes de  $\{\mathbf{U}_l : l = 1, \dots, L\}$ , y  $\langle \cdot, \cdot \rangle$  representa el producto interno en  $\mathbb{R}^d$ .

El condicionamiento de las realizaciones a un conjunto de datos requiere un paso adicional basado en kriging o cokriging (Chilès and Delfiner, 2012, p. 486), dependiendo de si hay una o varias variables de interés, respectivamente.

Los pasos del algoritmo Espectral Bandas Rotantes son:

1. Generar las líneas considerando un número entero  $L$  grande.
2. Simular un campo aleatorio a lo largo de cada línea con el método espectral.

Este algoritmo es bastante rápido y permite el cómputo en paralelo, requiriendo muy poco espacio de memoria. Además, es muy versátil puesto que permite simular cualquier campo aleatorio multivariable que tenga funciones de covarianza continuas y absolutamente integrables.

#### 2.6.4. Implementación del algoritmo de simulación

Antes de llevar a cabo las simulaciones condicionales, es necesario realizar pasos previos sobre los datos, que consideran:

- Estudio exploratorio de los datos.
- Transformación de los datos originales en datos Gaussianos (anamorfosis de los datos originales).
- Análisis variográfico de los datos transformados y ajuste del modelo de variograma.
- Simulación no condicional.
- Condicionamiento de las simulaciones mediante kriging.
- Transformación de los valores simulados a su distribución original.
- Procesamiento de las realizaciones.

Para el condicionamiento de las realizaciones, se utiliza kriging universal con una vecindad móvil dividida en octantes. Se simulan cien campos aleatorios intrínsecos

independientes con incrementos Gaussianos mediante el algoritmo Espectral Bandas Rotantes (Arroyo and Emery, 2018), con las siguientes especificidades:

- Dominio: grilla regular con 25632 nodos cada 10 metros, que cubre el área de interés.
- Modelo de variograma: según el caso considerado, se considera un variograma potencia con exponente  $\lambda=1.79$  ó  $\lambda=1.99$  y pendiente 1.
- Número de realizaciones: 100.
- Número de bandas rotantes: 500 (secuencia equidistribuida).
- Condicionamiento con kriging universal.

### 2.6.5. Validación de las simulaciones

Según Leuangthong et al. (2004, p. 132), un modelo simulado debe reproducir:

1. Los valores en los sitios muestreados
2. La distribución multivariable de los atributos de interés:
  - a) Forma del histograma.
  - b) Rango de valores simulados.
  - c) Resumen de estadísticas: medias, mediana y varianza.
  - d) Gráfico cuantil-cuantil.
3. La continuidad espacial de las variables: variogramas directos y cruzados.

### 2.6.6. Validación cruzada

Esta herramienta permite validar las simulaciones realizadas sobre un conjunto de datos (en algunos casos la totalidad de éstos) y analizar la capacidad de predicción de la incertidumbre del modelo empleado, en términos de precisión y exactitud. Consiste en simular sucesivamente cada dato considerando para esto solo los datos restantes. Es importante destacar que errores en la validación no indican un modelo deficiente, sino que pueden existir parámetros mal empleados para su realización los cuales deben ser verificados.

## 2.7. Marco Forestal



### 2.7.1. Antecedentes generales

En Chile, de acuerdo a información del Instituto Forestal al 2018 (Soto et al., 2020, p. 28), existen 3.1 millones de hectáreas de plantaciones forestales, principalmente *Pino Insigne* y *Eucalyptus globulus* y 14.7 millones de hectáreas de bosque nativo (ver Tabla 2.7.1). La región del Biobío posee la mayor superficie de plantaciones forestales (39.2%), mientras que en la región de la Araucanía se localiza la mayor cantidad de bosque nativo (Tabla 2.7.2).

**Cuadro 2.7.1:** Superficie Forestal, Chile

Uso Forestal	Superficie (ha)
Plantación Forestal	2303886
Bosque Nativo	14633779

**Cuadro 2.7.2:** Superficie Forestal por Región, Chile

<b>Región</b>	<b>Plantación (ha)</b>	<b>Bosque nativo (ha)</b>
Arica y Parinacota		47151
Tarapacá		33246
Coquimbo	74973	48475
Valparaíso	44606	484716
Metropolitana	5986	363955
O'Higgins	115341	459309
Maule	384690	581515
Biobío	902259	845552
Araucanía	488152	964153
Los Ríos	183574	908531
Los Lagos	74329	2827436
Aysén	29976	4398745
Magallanes		2671594

El ciclo forestal (Figura 2.7.1) consta de las fases de silvicultura, que se inicia con la producción de semillas para el establecimiento del bosque y/o plantación, luego vienen la cosecha de la plantación y la generación de productos en la industria, y culmina con la venta de los productos tales como madera aserrada, tableros y celulosa, ya sea en el mercado nacional como internacional.

El año 2019, el consumo de trozos fue de 45.307 miles de m<sup>3</sup> y las exportaciones del sector forestal totalizan US\$ 5.662 millones.



Figura 2.7.1: Ciclo Forestal

### 2.7.2. Mensura Forestal

Mensura forestal es la ciencia que se ocupa de la medición del bosque y sus productos (Prodan et al., 1997, p. 1), se sustenta principalmente en matemática y estadística, su principal objetivo es describir plantaciones o bosques a través del valor total de algún atributo, por ejemplo, volumen, área basal. Como los bosques son muy extensos y en ocasiones de difícil acceso, su descripción se basa en una pequeña muestra de árboles o unidad muestral, seleccionados de modo que representen a toda la población. Además, para que la unidad muestral sea probabilística, debe cumplirse

lo siguiente (Prodan et al., 1997, p. 153):

- Se asigna a cada árbol de la población una determinada probabilidad de ser seleccionado en una unidad muestral o parcela.
- Se localizan aleatoriamente puntos de muestreo, de modo que cualquier lugar de la superficie donde se distribuye la población de árboles, tenga la misma probabilidad de ser elegido.
- Todos los árboles de la población son susceptibles de ser seleccionados en cada punto de muestreo instalado, con la probabilidad preasignada.

Luego, a partir de las unidades muestrales probabilísticas, es factible generar estimaciones puntuales.

La ejecución del muestreo forestal, es una actividad que se lleva a cabo en terreno, en ocasiones con dificultades tales como, difícil acceso, terrenos en pendiente y/o con gran cantidad de sotobosque (vegetación que crece bajo los árboles de un bosque) que dificulta la ubicación de la unidad de muestreo y la medición de los árboles. Por lo anterior, el muestreo forestal es una actividad de alto costo en la actividad forestal y que se realiza varias veces para tener información de cómo se está desarrollando la plantación y qué rendimientos y/o productos se obtendrán en el momento de la cosecha (*Eucaliptus spp*: 10 a 15 años, *Pinus radiata* 18 a 25 años).

### 2.7.3. Geoestadística aplicada a recursos forestales

Aplicaciones de estadística espacial o geoestadística en el área forestal que se relacionan con los objetivos de la presente tesis son:

- [Matérn \(1960\)](#) en su tesis “Spatial Variation” utiliza los conceptos de estadística espacial en muestreos forestales.
- [Chica-Olmo \(2005\)](#) concluye que la metodología geoestadística se adapta bien al estudio de las variables del inventario forestal, resolviendo problemas de índole práctica que surgen en la gestión de estos recursos naturales, por medio de los métodos de estimación y simulación espacial. En su análisis utiliza kriging de indicadores y simulación condicional, para generar mapas temáticos de volumen y área basal a partir de datos provenientes de inventarios forestales.
- En la provincia de Jiangxi (China), [Daisheng and Bian \(2009\)](#) utilizan 525 puntos muestrales para estimar el índice de sitio de plantaciones de Abeto Chino (Chinese fir), combinando el uso de sistemas de información geográfico (SIG) y kriging. Los resultados obtenidos entregan error cuadrático medio (MSE) cercano a 0 y raíz del error cuadrático medio (RMSE) cercano a 1, por lo que el kriging reprodujo precisamente la distribución espacial y variabilidad del índice de sitio de Abeto Chino.
- En plantaciones de la especie *Eucalyptus spp.*, con datos de 89 parcelas de inventario permanentes, medidos desde el establecimiento de la plantación (2009) hasta el 2014, [Raimundo et al. \(2017\)](#) determinaron que utilizando kriging ordinario es factible disminuir la intensidad de muestreo un 40 %, sin afectar la precisión de las estimaciones de volumen al final de la rotación.
- [Hernández and Emery \(2009\)](#) utilizan simulación geoestadística y simulación annealing, para el diseño de un inventario forestal de la especie *Prosopis sp.*, con un error máximo de 10 % (95 % de confianza). Midieron las plantas vivas y muertas en 738 parcelas seleccionadas mediante un muestreo sistemático

con una superficie de 0.24 ha, utilizando simulación geoestadística y annealing obtuvieron que es factible reducir en un 50% los costos de muestreo sin afectar el error máximo.



## Capítulo 3

# Análisis Estadístico

### 3.1. Análisis exploratorio de datos

Los datos a analizar en la presente tesis, corresponden a información obtenida a partir de inventarios o mediciones forestales de Pino Insigne (*Pinus radiata*), cuyo objetivo principal es estimar el volumen de madera a extraer a la edad de cosecha, en  $\text{m}^3/\text{ha}$ . El predio se ubica en la región del Biobío, cercano a la ciudad de Concepción. La plantación fue establecida el año 2004 y se analizarán los rodales 29 y 30, cuya superficie e intensidad de muestreo se presentan en la Tabla 3.1.1.

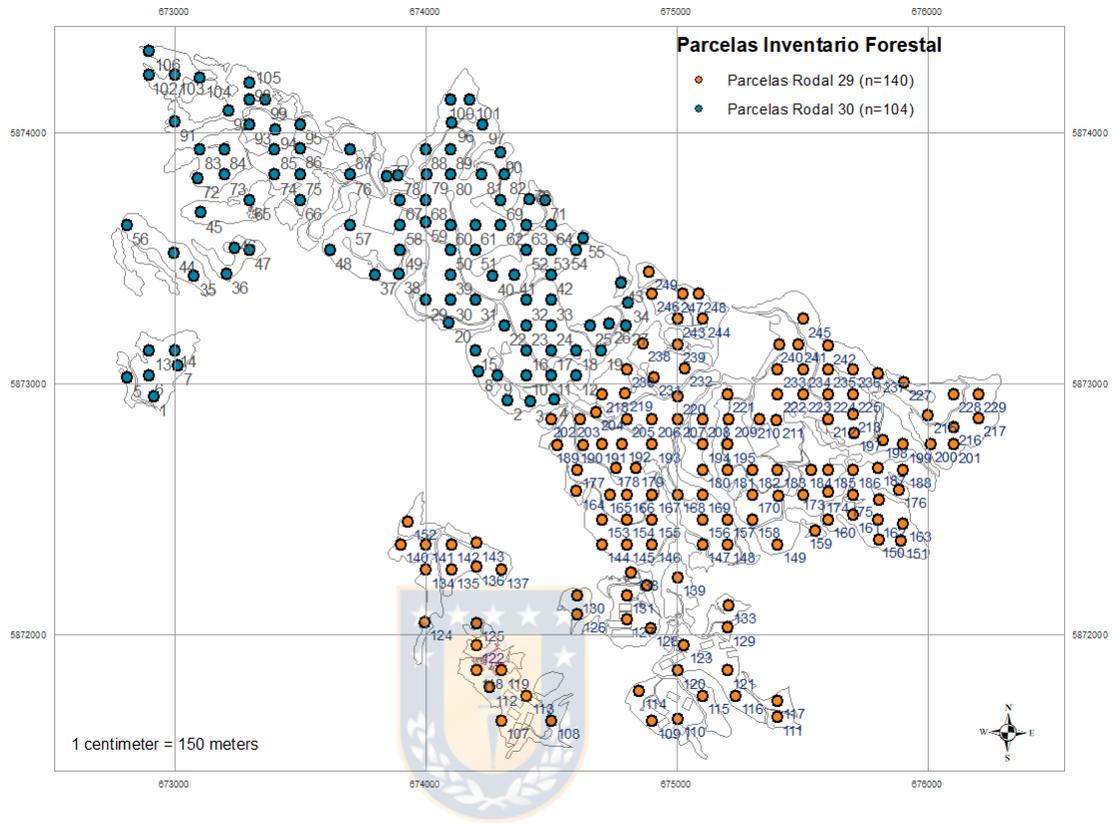
Un rodal (Cancino, 2012, p.22), es un conjunto espacialmente continuo de árboles y otros vegetales asociados, suficientemente uniforme para constituir una unidad de manejo, la que puede diferenciarse claramente de los rodales adyacentes por su edad, composición de especies, estructura, o por las características fisiográficas o climáticas en que crece. Por otra parte, la intensidad de muestreo corresponde a la cantidad de parcelas o unidades muestrales por unidad de superficie (hectáreas). La intensidad

de muestreo se define en la etapa de diseño del inventario o diseño muestral, cuyo objetivo es estimar adecuadamente los parámetros de estado de la población (bosque o plantación) y el error que afecta a la estimación (Prodan et al., 1997, p. 207).

**Cuadro 3.1.1:** Rodales de Inventario Forestal

Rodal	Superficie (ha)	Nº Parcelas	Intensidad de Muestreo (Parcelas/ha)
29	148.03	140	1.06
30	108.46	104	1.04

La Figura 3.1.1, elaborada con el software ArcGIS 10.4.1 (Aplicación de Sistema de Información Geográfica SIG), expone la ubicación de las parcelas de inventario de los rodales 29 y 30, con puntos de color naranja y azul, respectivamente. En ambos rodales se observa que las parcelas se distribuyen regularmente. Los polígonos irregulares de color gris corresponde a la superficie de cada rodal. Respecto al relieve del terreno es importante comentar que los terrenos en que se establecen plantaciones forestales son en general de pendientes superiores a 20 % y es común encontrar quebradas y/o bosque nativo que no permiten establecer una plantación continua, como ocurre por ejemplo con plantaciones agrícolas. La discontinuidad se observa en ambos rodales, a partir de los polígonos pequeños localizados al oeste de los polígonos principales.

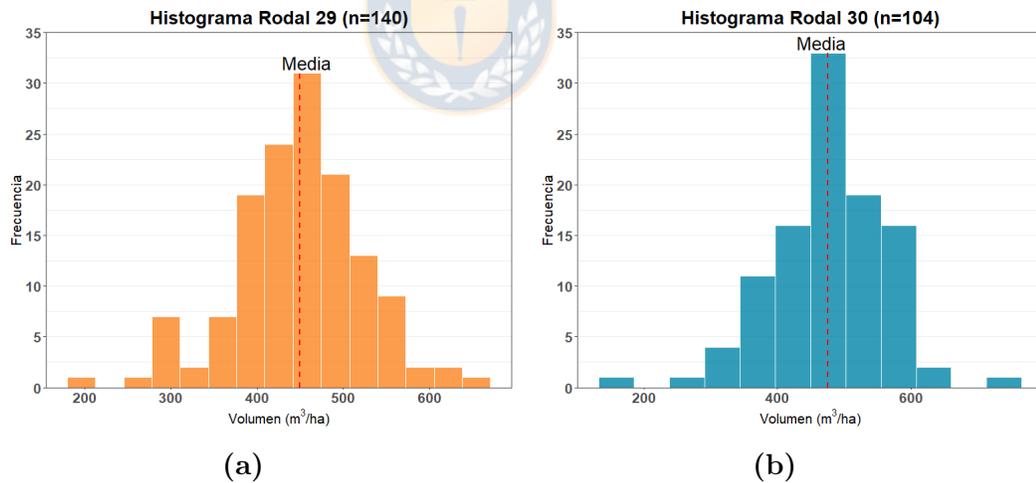


**Figura 3.1.1:** Mapa Ubicación Parcelas Inventario Forestal ( $n$  = número de parcelas)

La Tabla 3.1.2 resume los principales estadísticos del inventario forestal, por rodal y la Figura 3.1.2 muestra los histogramas, la línea punteada de color rojo corresponde al volumen medio por rodal.

**Cuadro 3.1.2:** Estadísticos Inventario Forestal, por rodal

Dato	Rodal 29	Rodal 30
Nº Parcelas	140	104
Superficie (ha)	148.03	108.46
Media (m <sup>3</sup> /ha)	449.16	475.13
Mediana (m <sup>3</sup> /ha)	452.40	478.70
Moda (m <sup>3</sup> /ha)	Multimodal	526.3
Mínimo (m <sup>3</sup> /ha)	198.70	165.50
Máximo (m <sup>3</sup> /ha)	656.70	746.70
Varianza (m <sup>3</sup> /ha) <sup>2</sup>	5519.18	7401.57
Desviación Estándar (m <sup>3</sup> /ha)	74.29	86.03
Rango Intercuartil (m <sup>3</sup> /ha)	85.65	100.9
Coefficiente de Variación	0.17	0.18
Sesgo	-0.20	-0.38
Kurtosis	0.79	1.29

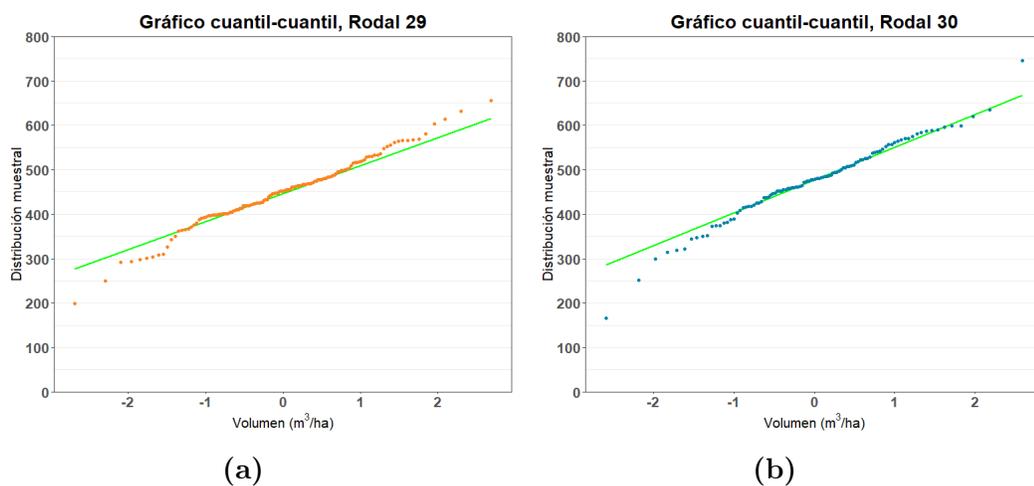
**Figura 3.1.2:** Histograma a) rodal 29, b) rodal 30.  $n$  = número de parcelas

De la información anterior podemos concluir lo siguiente:

- Los datos presentan distribución leptocúrtica ( $kurtosis > 0$ ) y asimetría negativa ( $sesgo < 0$ ).

- El sesgo es superior en el rodal 30.
- La variabilidad de los datos es alta, las medidas de dispersión observadas como varianza, desviación estándar (SD) y rango intercuartil (IQR) son superiores en el rodal 30.
- El coeficiente de variación es de 0.17 en el rodal 29 y 0.18 en el rodal 30, es decir la variabilidad de los datos respecto al promedio es levemente mayor en el rodal 30 (18%).

La Figura 3.1.3 contiene gráficos cuantil-cuantil, que comparan la distribución de los datos con la distribución normal teórica. Se aprecia que en el rodal 29 aquellas parcelas con volumen entre 350 y 500 están cercanos a la distribución normal y en el rodal 30 esto ocurre para los valores 400 a 600. Además, se pueden distinguir claramente algunos puntos atípicos.



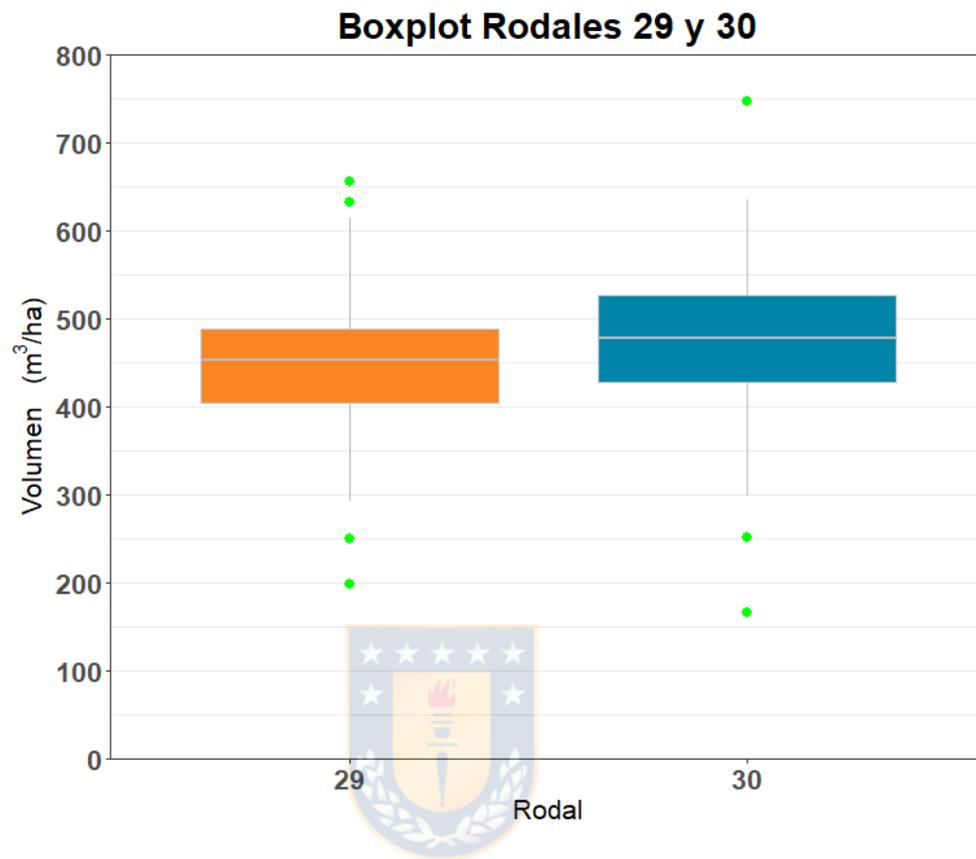
**Figura 3.1.3:** Gráficos cuantil-cuantil a) rodal 29, b) rodal 30

## 3.2. Valores Atípicos

En la Figura 3.2.1 se presentan gráficos boxplot por rodal, que nos permiten detectar datos atípicos u outliers (puntos verdes) a través del método de Tukey (1977, pp. 43–44), el cual considera outliers a todos los valores que se encuentran fuera del siguiente rango  $[Q_1 - (1.5 * IQR), Q_3 + (1.5 * IQR)]$ , donde  $IQR$  significa rango intercuartil,  $Q_1$  y  $Q_3$  corresponden al cuartil 1 (25 % de los datos es menor o igual a este valor) y cuartil 3 (75 % de los datos es menor o igual a este valor), respectivamente.

Para los datos de estudio se aprecian 4 datos atípicos en el rodal 29 y 3 en el rodal 30. Se observa también que la mediana del rodal 30 es superior al rodal 29, y que el rango intercuartil ( $IQR$ ) del rodal 30 también es mayor, lo que implica una mayor dispersión de los datos (ver Tabla 3.1.2).

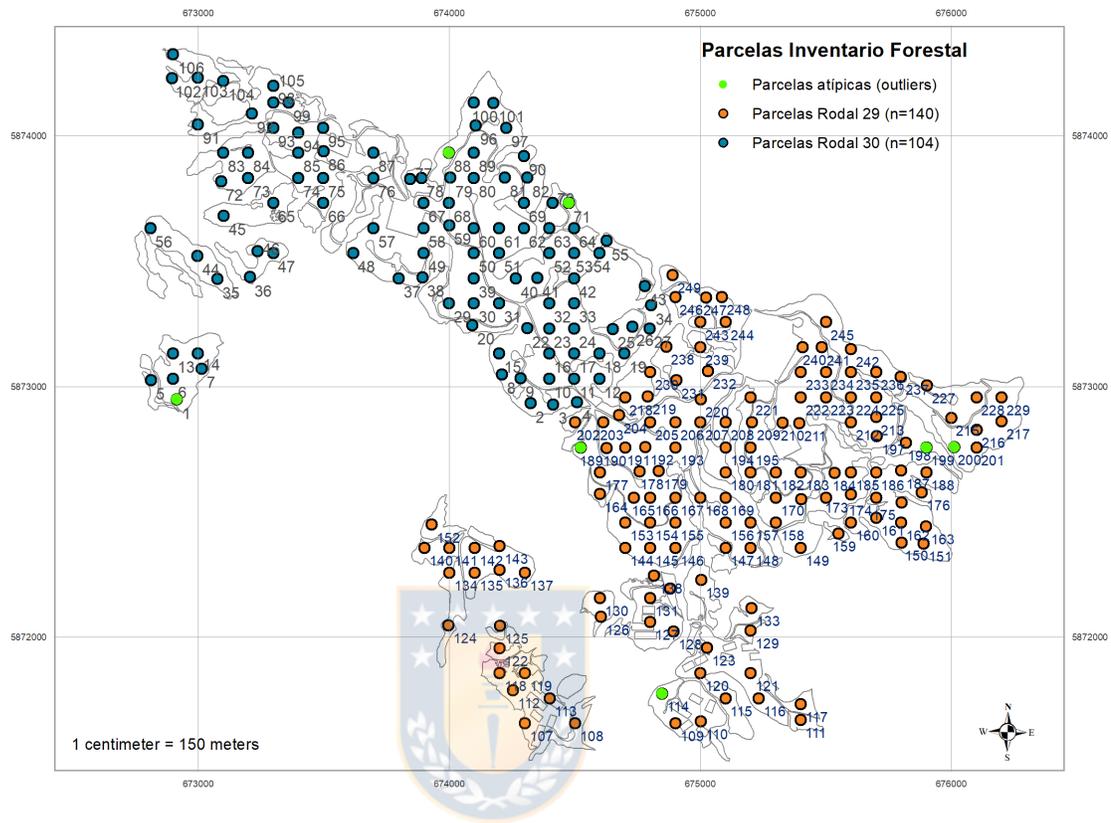




**Figura 3.2.1:** Gráfico de caja, Rodales 29 y 30.

En la Figura 3.2.2 se destaca con color verde la ubicación de las parcelas atípicas, identificadas en el gráfico boxplot (Figura 3.2.1). Se observa que los datos atípicos no están agrupados, por lo que se plantean dos hipótesis para explicar los outliers:

- Error de medición.
- Condición particular del terreno donde se ubica la parcela.



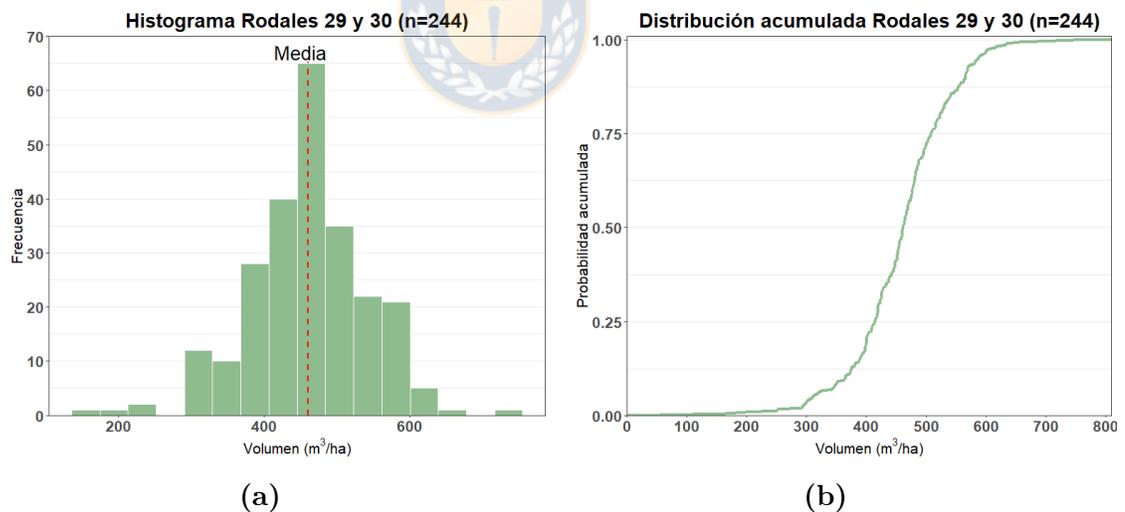
**Figura 3.2.2:** Mapa Ubicación Parcelas Inventario Forestal, con parcelas atípicas

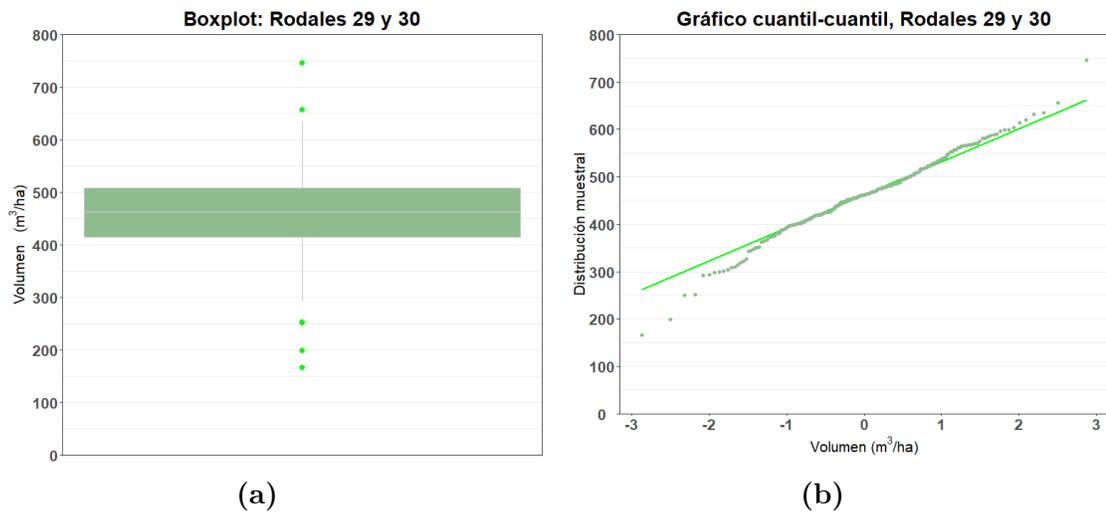
A partir del análisis exploratorio de datos se concluye que ambos rodales tienen distribuciones similares, por lo cual se analizarán ambos rodales en conjunto. En la Tabla 3.2.1 se presentan los estadísticos, considerando ambos rodales conjuntamente. Las Figuras 3.2.3 y 3.2.4, contienen histograma, distribución acumulada, gráfico de caja y cuantil-cuantil, considerando ambos rodales en conjunto.

De la información en conjunto se infiere que la distribución de los datos es levemente asimétrica (-0.22) y leptocúrtica (ver Figura 3.2.3), además en el gráfico de caja se visualizan los datos atípicos y en el gráfico cuantil-cuantil se observa que los datos se ajustan a la línea normal, excepto en los extremos (ver Figura 3.2.4).

**Cuadro 3.2.1:** Estadísticos Inventario Forestal rodales 29 y 30, juntos

Rodales 29 y 30, juntos	Datos originales
Nº Parcelas	244
Superficie (ha)	256.5
Media (m <sup>3</sup> /ha)	460.23
Mediana (m <sup>3</sup> /ha)	461.50
Moda (m <sup>3</sup> /ha)	475.1
Mínimo (m <sup>3</sup> /ha)	165.5
Máximo (m <sup>3</sup> /ha)	746.70
Varianza (m <sup>3</sup> /ha) <sup>2</sup>	6486.39
Desviación Estándar (m <sup>3</sup> /ha)	80.54
Rango Intercuartil (m <sup>3</sup> /ha)	93.98
Coefficiente de Variación	0.18
Sesgo	-0.22
Kurtosis	1.05

**Figura 3.2.3:** (a) Histograma Rodales 29 y 30, juntos. (b) Distribución acumulada Rodales 29 y 30, juntos.



**Figura 3.2.4:** a) Diagrama de caja Rodales 29 y 30, juntos. (b) Gráfico cuantil-cuantil Rodales 29 y 30, juntos.

Para complementar el análisis, se realizan en el software R tests de hipótesis de comparación de medias y varianzas, con el objetivo de determinar si existen diferencias significativas entre los rodales 29 y 30. El nivel de significancia considerado es de 5% y los resultados (Tabla 3.2.2) nos indican que no hay evidencia significativa para rechazar la hipótesis nula de igualdad, por lo tanto es factible analizar ambos rodales conjuntamente.

**Cuadro 3.2.2:** Test de Hipótesis

Test de Hipótesis	Valor $-p$
$H_0$ : Igualdad de varianzas	0.10
$H_0$ : Igualdad de medias	0.99

Finalmente, para mitigar el efecto de los outliers en la distribución sin pérdida de información, se realiza un “capping” de los datos (truncación de datos), es decir, a

los datos sobre el percentil 75 se asigna el valor del cuartil 3 ( $Q_3 = 649.4$  ( $\text{m}^3/\text{ha}$ )), mientras que a los valores muy bajos se imputa el valor del cuartil 1 ( $Q_1 = 273.5$  ( $\text{m}^3/\text{ha}$ )).

La Tabla 3.2.3 compara los estadísticos de los datos originales y datos con capping (datos truncados).

**Cuadro 3.2.3:** Estadísticos Inventario Forestal, datos originales y truncados

Rodales 29 y 30	Datos Originales	Datos Truncados	Diferencia en valor absoluto
Nº Parcelas	244	244	0
Superficie (ha)	256.5	256.5	0
Media ( $\text{m}^3/\text{ha}$ )	460.23	460.73	0.50
Mediana ( $\text{m}^3/\text{ha}$ )	461.50	461.50	0
Moda ( $\text{m}^3/\text{ha}$ )	475.1	273.54	201.56
Mínimo ( $\text{m}^3/\text{ha}$ )	165.5	273.54	108.04
Máximo ( $\text{m}^3/\text{ha}$ )	746.70	649.48	97.22
Varianza ( $\text{m}^3/\text{ha}$ ) <sup>2</sup>	6846.39	5861.72	624.67
Desviación Estándar ( $\text{m}^3/\text{ha}$ )	80.54	76.56	3.98
Rango Intercuartil ( $\text{m}^3/\text{ha}$ )	93.98	93.98	0
Coefficiente de Variación	0.18	0.17	0.01
Sesgo	-0.22	-0.11	0.11
Kurtosis	1.05	-0.03	1.08

# Capítulo 4

## Análisis

### 4.1. Descripción Espacial

La Figura 4.1.1 nos muestra la ubicación de las parcelas de inventario donde los ejes corresponden a las coordenadas UTM (*Universal Tranverse Mercator*). Es factible apreciar que no existen sectores con agrupamientos significativos de parcelas, lo que se corrobora con la información de la Tabla 4.1.1, donde se observa que para ambos rodales la distancia promedio y moda son casi idénticas, sin embargo el rodal 30 presenta un rango mas amplio de distancias entre parcelas.

**Cuadro 4.1.1:** Distancia (metros) entre las Parcelas de Inventario

Rodal	Mínimo	Media	Máximo	Moda
29	63.13	93.48	205.67	100
30	46.02	93.39	217.81	100

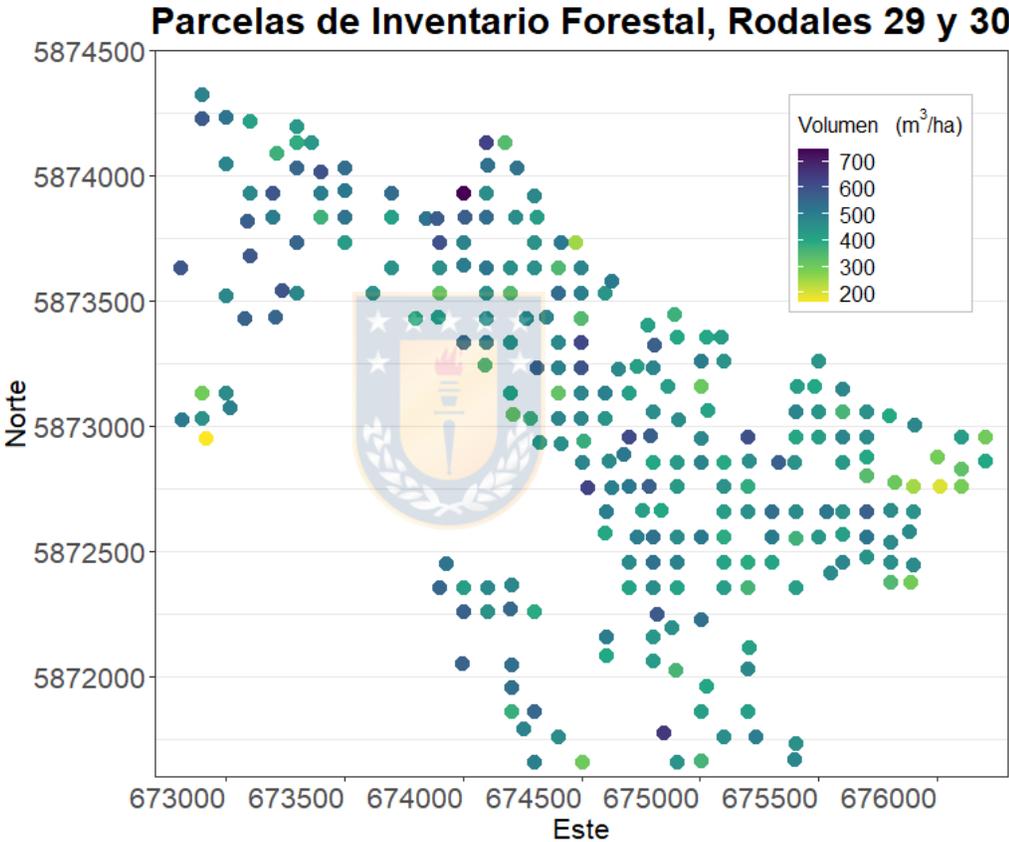
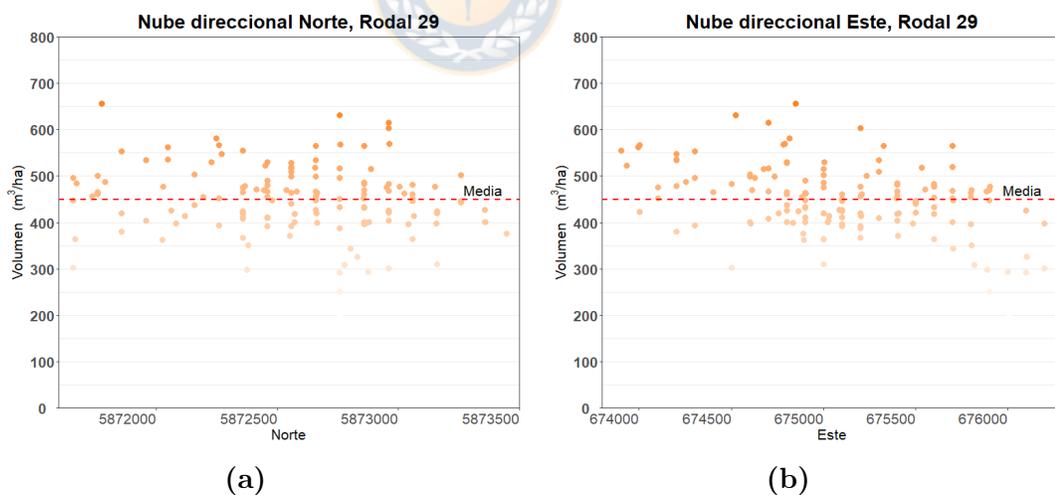


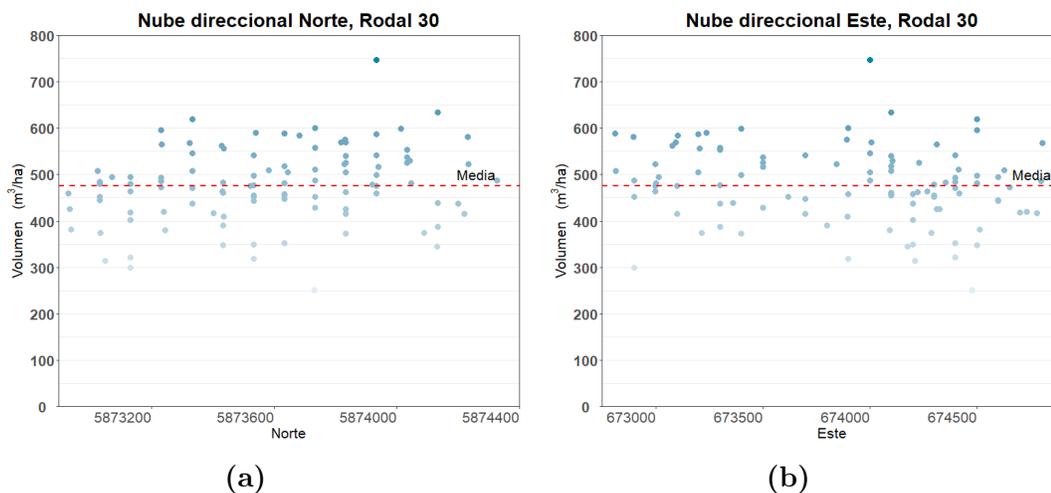
Figura 4.1.1: Ubicación parcelas de inventario forestal: Rodales 29 y 30, juntos

Una de las características del análisis geoestadístico es que nos permite determinar la continuidad espacial de la variable de interés. Para explorar tendencias se presentan en las Figuras 4.1.2 y 4.1.3 las nubes direccionales para ambos rodales. En el Rodal 29 se observa en dirección Este una disminución del volumen, teniendo las parcelas un volumen menor al promedio del rodal. Lo anterior se corrobora con la Tabla 4.1.2, en la cual se muestra el coeficiente de correlación de Pearson, el cual mide la relación lineal entre el volumen y las coordenadas UTM, para la dirección Este el valor es de  $-0.42$ .

En el Rodal 30 no es factible inferir a partir del gráfico alguna tendencia y las correlaciones en la Tabla 4.1.3 en dirección Norte y Este son  $0.28$  y  $-0.11$ , respectivamente.



**Figura 4.1.2:** (a) Nube direccional Norte, Rodal 29. (b) Nube direccional Este, Rodal 29



**Figura 4.1.3:** (a) Nube direccional Norte, Rodal 30. (b) Nube direccional Este, Rodal 30

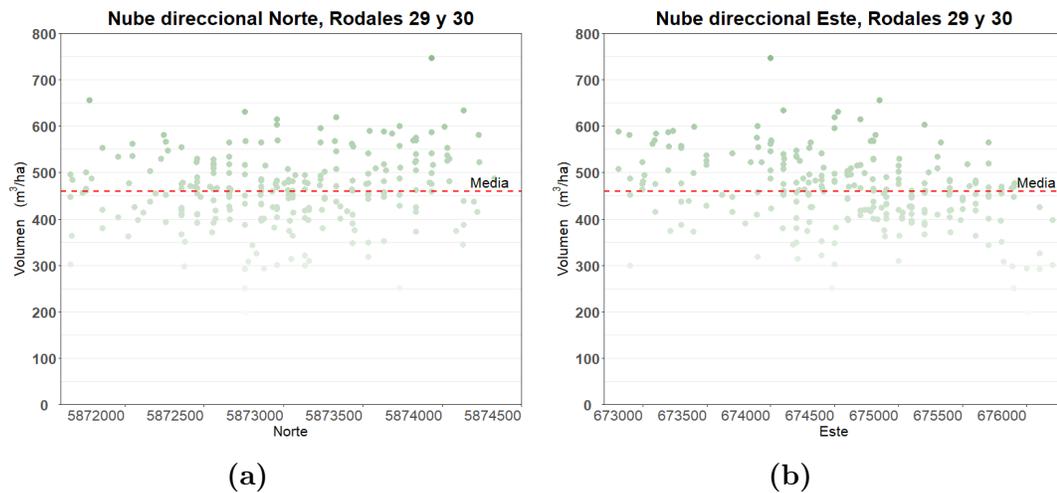
**Cuadro 4.1.2:** Matriz de correlación, Rodal 29

Rodal 29	Volumen ( $\text{m}^3/\text{ha}$ )	Norte	Este
Volumen	1.00	-0.13	-0.42
Norte	-0.13	1.00	0.37
Este	-0.42	0.37	1.00

**Cuadro 4.1.3:** Matriz de correlación, Rodal 30

Rodal 30	Volumen ( $\text{m}^3/\text{ha}$ )	Norte	Este
Volumen	1.00	0.28	-0.11
Norte	0.28	1.00	-0.38
Este	-0.11	-0.38	1.00

La Figura 4.1.4 contiene los gráficos de nube direccional considerando los datos de ambos rodales (en conjunto), se aprecia que en dirección Este el volumen de cosecha tiende a disminuir, con un valor de correlación de Pearson de -0.3 (Tabla 4.1.4).



**Figura 4.1.4:** (a) Nube Direccional Norte, Rodales 29 y 30. (b) Nube Direccional Este, Rodales 29 y 30

**Cuadro 4.1.4:** Matriz de correlación, Rodales 29 y 30

Rodales 29 y 30	Volumen ( $\text{m}^3/\text{ha}$ )	Norte	Este
Volumen	1.00	0.15	-0.30
Norte	0.15	1.00	-0.54
Este	-0.30	-0.5	1.00

## 4.2. Desagrupamiento de los datos

Dado que las parcelas de inventario no están regularmente espaciadas, se realizó desagrupamiento de los datos utilizando el software GSLIB (*Geostatistical Software Library*). Se consideran tamaños de celdas entre 0 y 300 m, con el objetivo de determinar el tamaño de celda con menor media.

En la Tabla 4.2.1 se presentan los valores de las medias obtenidas con GSLIB para cada tamaño de celda y los escenarios considerados: datos originales y datos truncados. Se destaca de color verde la menor media, que se produce para ambos escenarios

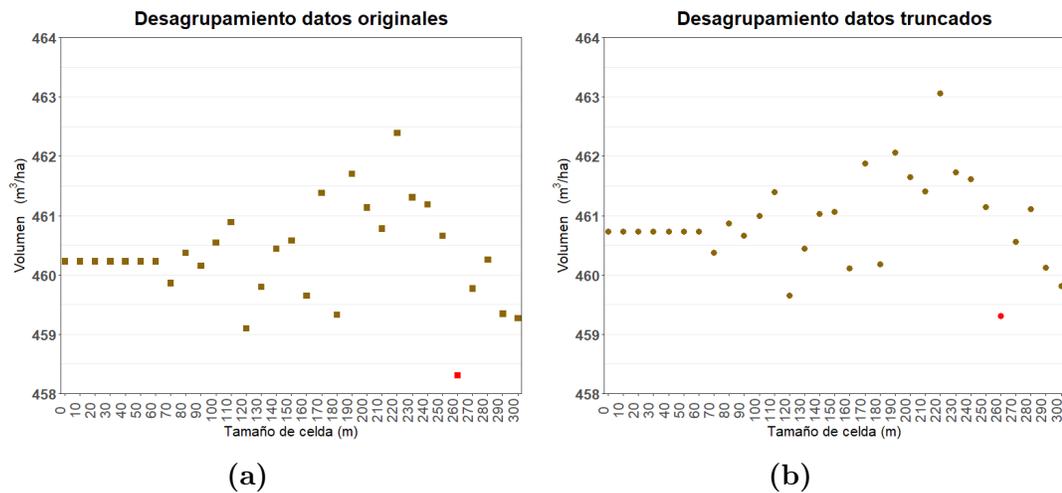
(datos originales y datos truncados) en el tamaño de celda 260 m.

Las Figuras 4.2.1a y 4.2.1b permiten observar la relación tamaño de celda versus media del desagrupamiento de los datos originales (representados con cuadrado) y los datos truncados (representados con círculo), respectivamente. Se aprecia que ambos escenarios presentan su menor media para el tamaño de celda 260 metros (punto rojo en las Figuras 4.2.1a y 4.2.1b).



**Cuadro 4.2.1:** Resultados desagrupamiento, datos originales y datos truncados

Tamaño Celda (m)	Media (m <sup>3</sup> /ha)	
	Datos originales	Datos truncados
0	460.23	460.74
10	460.23	460.74
20	460.23	460.74
30	460.23	460.74
40	460.23	460.74
50	460.23	460.74
60	460.23	460.74
70	459.86	460.37
80	460.37	460.87
90	460.16	460.67
100	460.55	460.99
110	460.89	461.40
120	459.09	459.65
130	459.80	460.44
140	460.45	461.03
150	460.58	461.07
160	459.66	460.11
170	461.39	461.88
180	459.33	460.18
190	461.71	462.06
200	461.14	461.64
210	460.78	461.41
220	462.39	463.06
230	461.31	461.73
240	461.19	461.61
250	460.66	461.14
260	458.31	459.31
270	459.77	460.56
280	460.26	461.11
290	459.35	460.12
300	459.28	459.82



**Figura 4.2.1:** Desagrupamiento (a) de los datos originales, (b) de los datos truncados

Las Tablas 4.2.2 y 4.2.3 muestran los estadísticos de los valores destacados en la Tabla 4.2.1, para los datos originales y para los datos truncados. Es factible observar en la Tabla 4.2.2, que al realizar el desagrupamiento con un tamaño de celda de 260 metros disminuye el volumen medio de 460.23 a 458.31 ( $\text{m}^3/\text{ha}$ ), sin embargo la variabilidad aumenta de 6486.39 a 7243.92 ( $\text{m}^3/\text{ha}$ )<sup>2</sup>. Para los datos truncados (Tabla 4.2.3) se observa igualmente que con desagrupamiento el volumen promedio disminuye y la varianza aumenta, sin embargo la diferencia es de una menor magnitud.

**Cuadro 4.2.2:** Estadísticos: Desagrupamiento de los datos originales con tamaño de celda con media mínima

Estadísticos	Datos originales	Desagrupamiento (260 m)
Media ( $\text{m}^3/\text{ha}$ )	460.23	458.31
Varianza ( $\text{m}^3/\text{ha}$ ) <sup>2</sup>	6486.39	7243.92

**Cuadro 4.2.3:** Estadísticos: Desagrupamiento de los datos truncados con tamaño de celda con media mínima

Estadísticos	Datos truncados	Desagrupamiento (260 m)
Media ( $\text{m}^3/\text{ha}$ )	460.74	459.31
Varianza ( $\text{m}^3/\text{ha}$ ) <sup>2</sup>	5861.72	6467.28

### 4.3. Variograma Experimental Omnidireccional

El cálculo del variograma experimental se realizó utilizando el software MATLAB (MATrix LABoratory). A continuación se presenta el variograma omnidireccional con y sin desagrupamiento de los datos originales y de los datos truncados. Los parámetros considerados en el cálculo de ambos variogramas son:

**Cuadro 4.3.1:** Parámetros del variograma experimental omnidireccional

azm	atol	dip	dtol	lag	nlag	lagtol
0	90	0	22.5	300	11	100

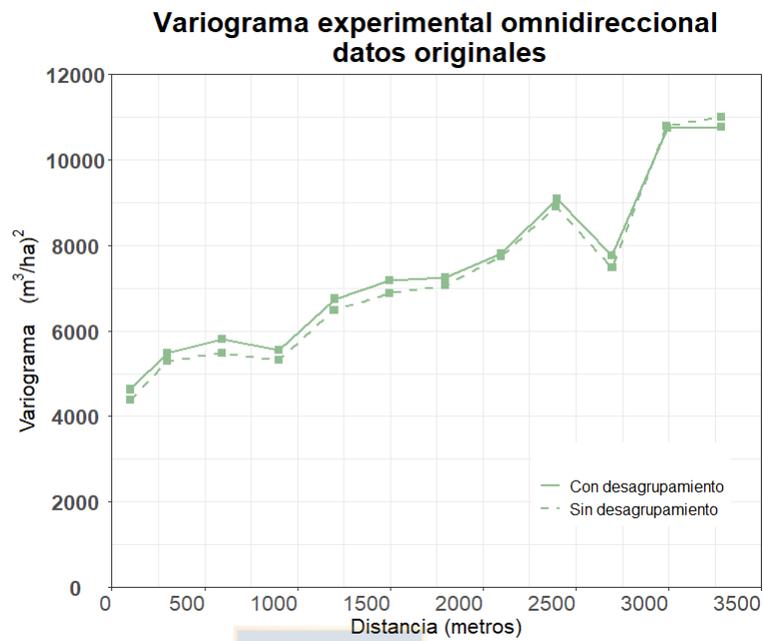
donde:

- azm: acimut o ángulo horizontal
- atol: tolerancia del acimut
- dip: ángulo vertical
- dtol: tolerancia del dip
- lag: distancia o paso

- nlag: número de pasos
- lagsol: tolerancia del lag.

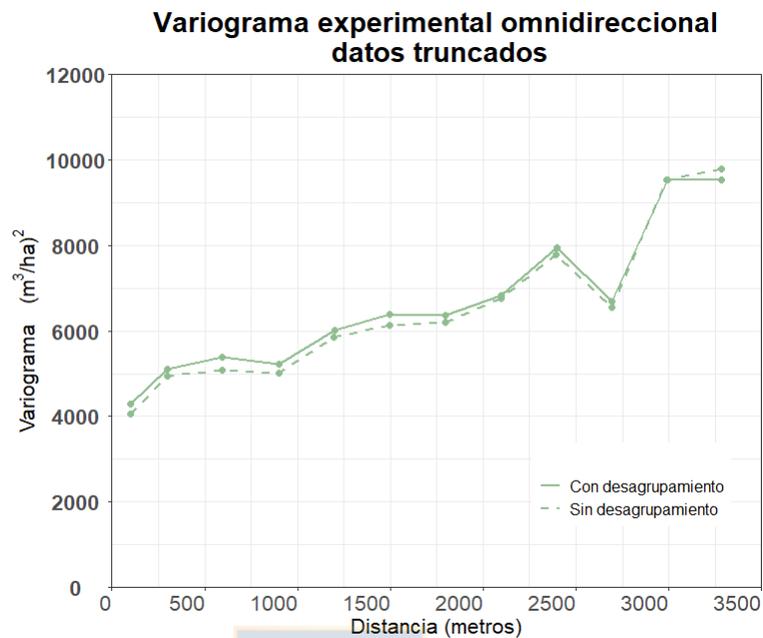
La Figura 4.3.1 muestra el variograma omnidireccional calculado para los datos originales con desagrupamiento (línea verde continua) y sin desagrupamiento (línea verde discontinua). Al comparar las curvas se observa que ambas son bastante similares, a los 1000 metros de distancia, la función variograma alcanza un valor de aproximadamente  $6000 \text{ (m}^3/\text{ha)}^2$ , a los 2400 metros llega a los  $9000 \text{ (m}^3/\text{ha)}^2$  y a los 3250 metros a  $11000 \text{ (m}^3/\text{ha)}^2$ . La variabilidad de los datos es levemente mayor en la curva con desagrupamiento.

La Figura 4.3.2 contiene el variograma experimental omnidireccional calculado con los datos truncados, con y sin desagrupamiento, línea verde continua y línea verde discontinua, respectivamente. Se observa, al igual que en la Figura 4.3.1, que el variograma con desagrupamiento tiene una variabilidad levemente mayor, sin embargo para las distancias de 1000, 2400 y 3250 metros, los valores de la función variograma son menores que en la Figura 4.3.1, alcanzando 5500, 7000 y 10000  $\text{(m}^3/\text{ha)}^2$ , respectivamente.



**Figura 4.3.1:** Variograma experimental omnidireccional, datos originales con y sin desagrupamiento





**Figura 4.3.2:** Variograma experimental omnidireccional, datos truncados con y sin desagrupamiento

## 4.4. Ajuste Variográfico

El ajuste del variograma experimental omnidireccional se realizó utilizando el software MATLAB. En los variogramas experimentales calculados a partir de los datos (Figura 4.3.1 y 4.3.2) se observa que el variograma es no acotado, ya que no alcanza una meseta, sino que crece indefinidamente (no estacionario). Dado lo anterior se consideró ajustar un modelo de variograma potencia. Los detalles para cada set de datos se presentan a continuación.

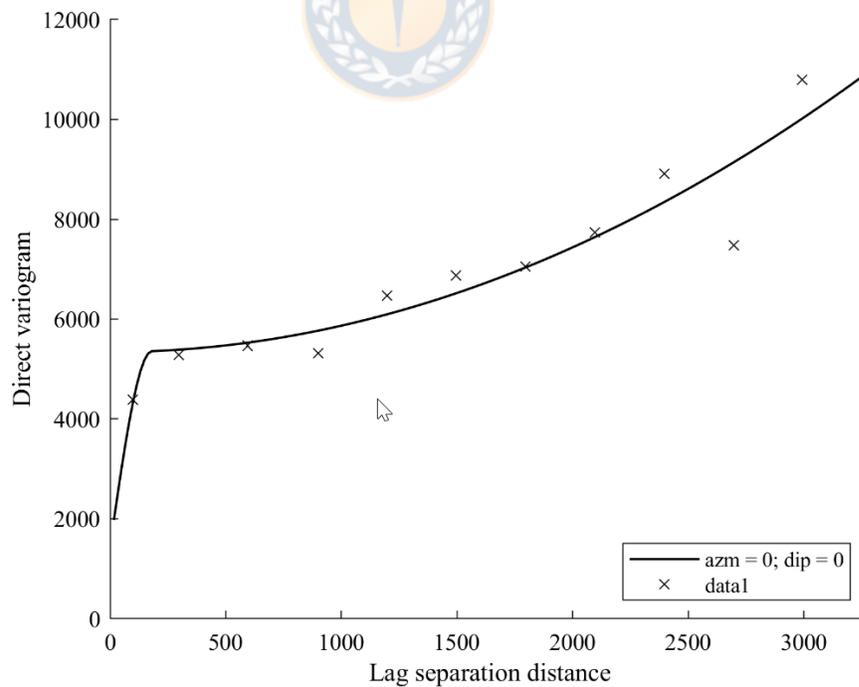
### 4.4.1. Ajuste considerando datos originales

El ajuste variográfico de los datos originales se presenta en la Figura 4.4.1, se observa que el volumen de cosecha de la plantación es no estacionario, ya que el

modelo ajustado no tiene meseta, la variabilidad crece al aumentar la distancia entre los puntos. El comportamiento en el origen tiene un efecto pepita de 1446.68, de esta forma, el ajuste es un modelo de variograma anidado compuesto por tres modelos teóricos. Los parámetros del variograma anidado ajustado se presentan en la Tabla 4.4.1, en la cual se señala con – si el valor no corresponde para el modelo.

**Cuadro 4.4.1:** Parámetros del variograma ajustado, considerando los datos originales

Modelo	Alcance (m)	Meseta ( $\text{m}^3/\text{ha}$ ) <sup>2</sup>	Exponente	Pendiente
Pepita (Nugget)	–	1446.68	–	–
Esférico	180	3894.31	–	–
Potencia (Power)	600	191.12	1.99	1



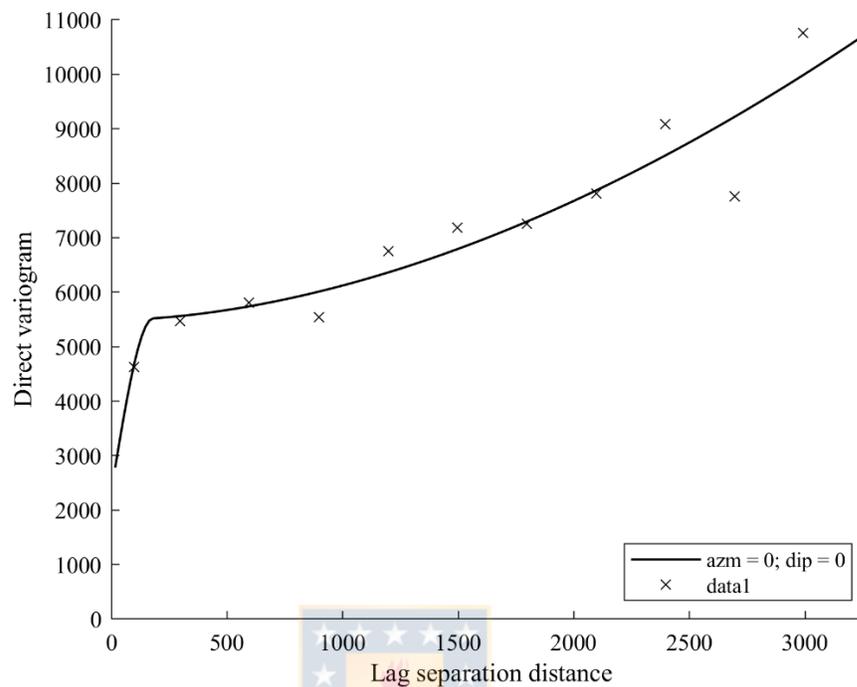
**Figura 4.4.1:** Variograma ajustado omnidireccional, considerando los datos originales

#### 4.4.2. Ajuste considerando datos originales y desagrupados

En el ajuste del variograma experimental considerando los datos originales y desagrupados, se obtiene al igual que para los datos originales un modelo anidado, sin embargo el exponente del modelo potencia es levemente menor (1.79) y la pepita mayor. Los detalles del modelo ajustado se presentan en la Tabla 4.4.2 (“-” indica que no corresponde un valor para el modelo), mientras que en la Figura 4.4.2 se visualiza gráficamente la coincidencia entre las cruces que representan el variograma experimental y la línea continua que corresponde al variograma ajustado.

**Cuadro 4.4.2:** Parámetros del variograma ajustado considerando datos originales desagrupados

Modelo	Alcance (m)	Meseta (m <sup>3</sup> /ha) <sup>2</sup>	Exponente	Pendiente
Pepita (Nugget)	-	2353.48	-	-
Esférico	180	3136.61	-	-
Potencia (Power)	600	253.38	1.79	1



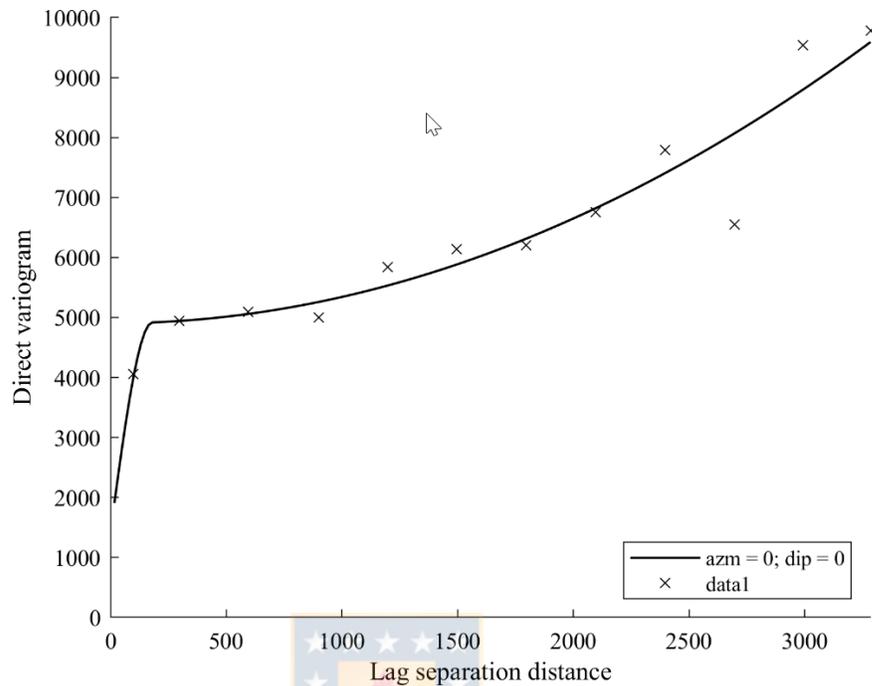
**Figura 4.4.2:** Variograma omnidireccional considerando los datos originales con desagrupamiento

#### 4.4.3. Ajuste considerando datos truncados

El ajuste variográfico considerando los datos truncados, produce los parámetros detallados en la Tabla 4.4.3. Estos son similares al variograma ajustado con los datos originales, sólo se observa diferencia en la meseta del modelo esférico. La Figura 4.4.3 presenta gráficamente el variograma ajustado.

**Cuadro 4.4.3:** Parámetros del variograma ajustado considerando los datos truncados

Modelo	Alcance (m)	Meseta ( $\text{m}^3/\text{ha}$ ) <sup>2</sup>	Exponente	Pendiente
Pepita (Nugget)	–	1434.47	–	–
Esférico	180	3468.44	–	–
Potencia (Power)	600	159.23	1.99	1



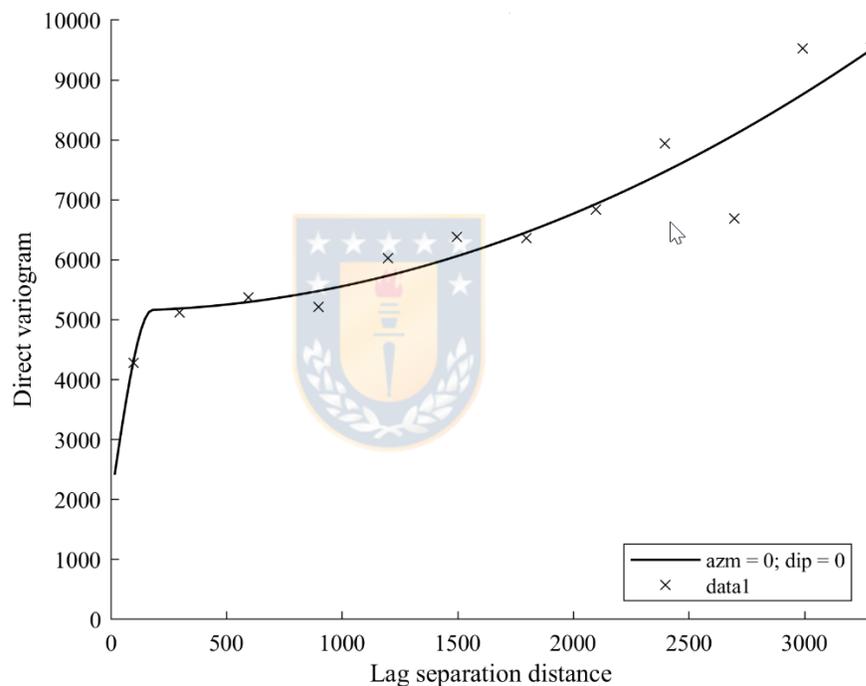
**Figura 4.4.3:** Variograma omnidireccional considerando los datos truncados

#### 4.4.4. Ajuste considerando datos truncados desagrupados

En el caso de ajustar el variograma experimental considerando los datos truncados y desagrupados, se observa un ajuste muy similar a los casos anteriores, el modelo ajustado es anidado y está compuesto por pepita, modelo esférico y potencia. Si se compara el ajuste para datos truncados y datos truncados desagrupados se observa que este último tiene mayor pepita. Los resultados se detallan en la Tabla 4.4.4 y Figura 4.4.4.

**Cuadro 4.4.4:** Parámetros del variograma ajustado considerando los datos truncados y desagrupados

Modelo	Alcance (m)	Meseta ( $\text{m}^3/\text{ha}$ ) <sup>2</sup>	Exponente	Pendiente
Pepita (Nugget)	–	1980.14	–	–
Esférico	180	3171.02	–	–
Potencia (Power)	600	147.62	1.99	1

**Figura 4.4.4:** Variograma omnidireccional de los datos truncados y desagrupados

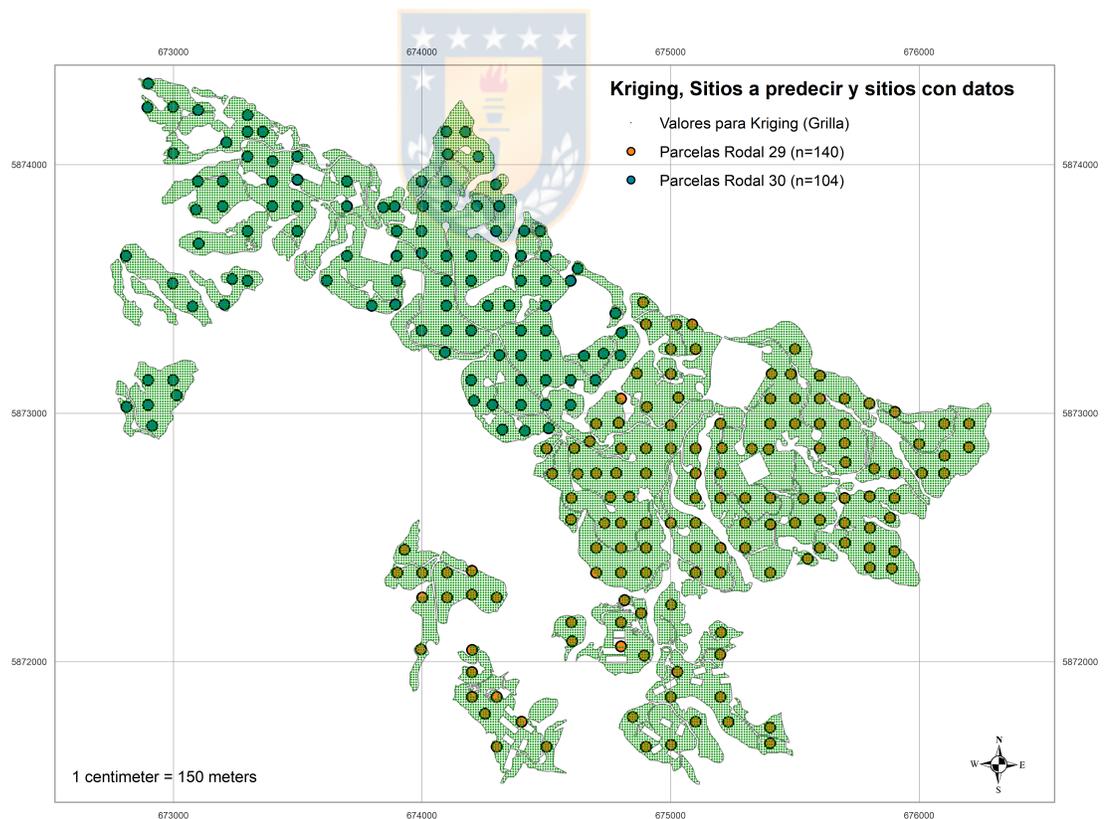
## 4.5. Predicción de la variable con Kriging

El kriging, como se comentó anteriormente, es el método de interpolación geoestadístico. Para la presente tesis, se requiere conocer el volumen de madera a extraer en los dos rodales (29 y 30), por lo que se elaboró una grilla con el software

ArcGIS, ajustada a la forma de los polígonos de los rodales bajo análisis. La grilla contiene 25332 puntos cada 10 metros y se presenta en la Figura 4.5.1.

La información necesaria para la ejecución del kriging es el variograma ajustado, y considerando que el variograma ajustado es un modelo no acotado, se utiliza kriging universal, para estimar el volumen de cosecha ( $m^3/ha$ ) en todos los puntos de la grilla.

En las siguientes subsecciones se presentan los resultados obtenidos y la validación de los resultados.



**Figura 4.5.1:** Grilla utilizada para el Kriging

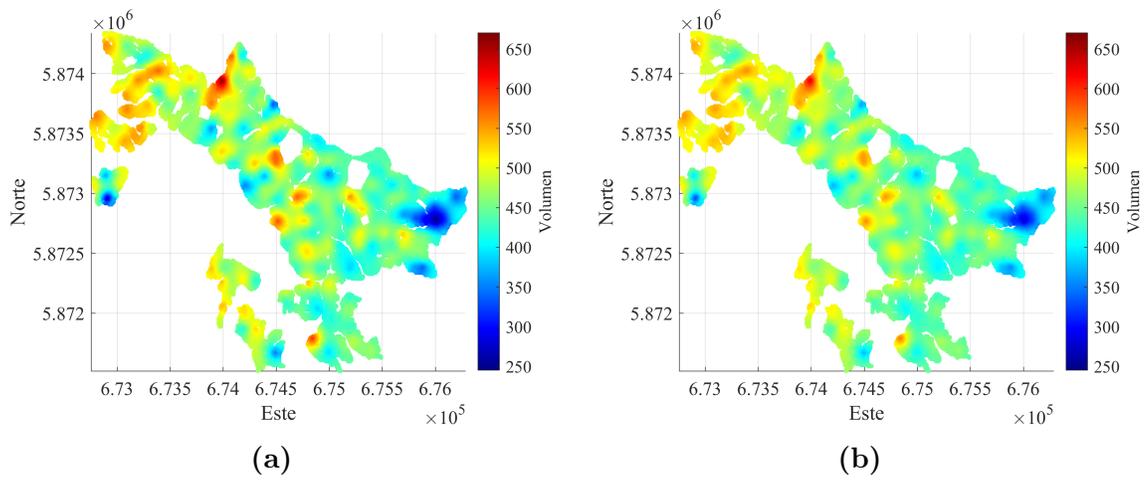
### 4.5.1. Visualización de las predicciones

Las Figuras 4.5.2 y 4.5.3 presentan los resultados del kriging universal con cada uno de los modelos ajustados en la Sección 4.4, para cada uno de los set de datos considerados (Figuras 4.3.1 a 4.4.4).

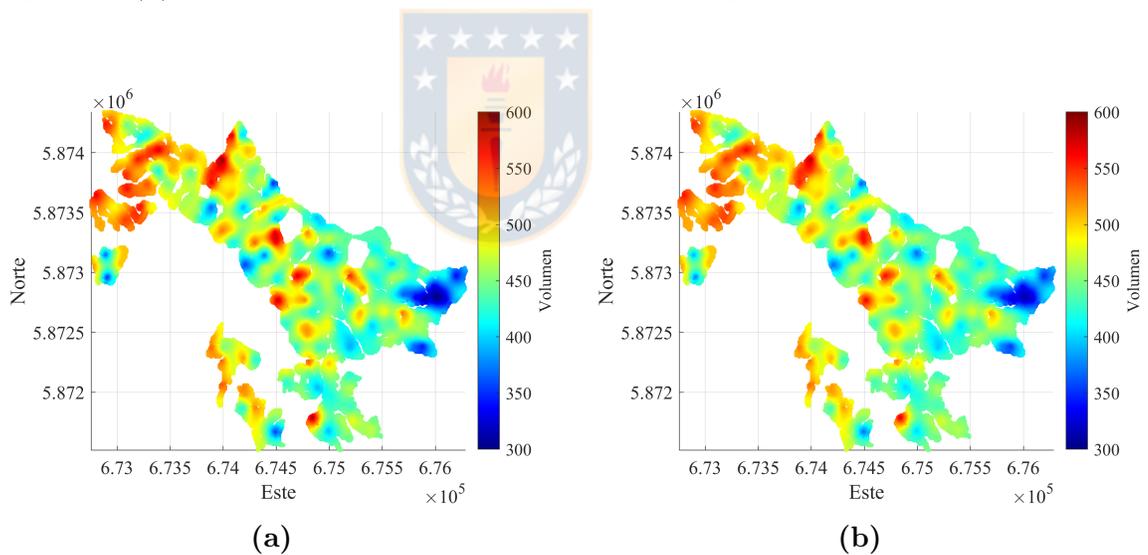
La Figura 4.5.2 presenta el mapa de predicción con kriging condicional a los datos originales (Figura 4.5.2a) y datos originales desagrupados (Figura 4.5.2b). Se observa que los valores predichos oscilan entre 250 y 650 ( $\text{m}^3/\text{ha}$ ), en ambos mapas. Sin embargo, en la Figura 4.5.2a, se aprecian colores levemente más intensos que en la Figura 4.5.2b, es decir, con los datos originales la varianza del kriging es levemente mayor (Tabla 4.6.3 y 4.6.4).

El mapa de predicción con datos truncados (Figura 4.5.3), muestra una zona más amplia de volúmenes de color rojo (500 a 600 ( $\text{m}^3/\text{ha}$ )), al compararlo con la Figura 4.5.2, es decir para el mismo sector esta prediciendo volúmenes superiores que el mapa del kriging condicional a los datos originales (Figura 4.5.2).

Respecto a la diferencia del mapa de predicción con kriging condicional a los datos truncados Figura 4.5.3a y datos truncados desagrupados Figura 4.5.3b, se observa que la predicción de volumen es más suavizada cuando se consideran los datos truncados desagrupados.



**Figura 4.5.2:** Mapas de la predicción con kriging: (a) condicional a los datos originales, (b) condicional a los datos originales y desagrupados

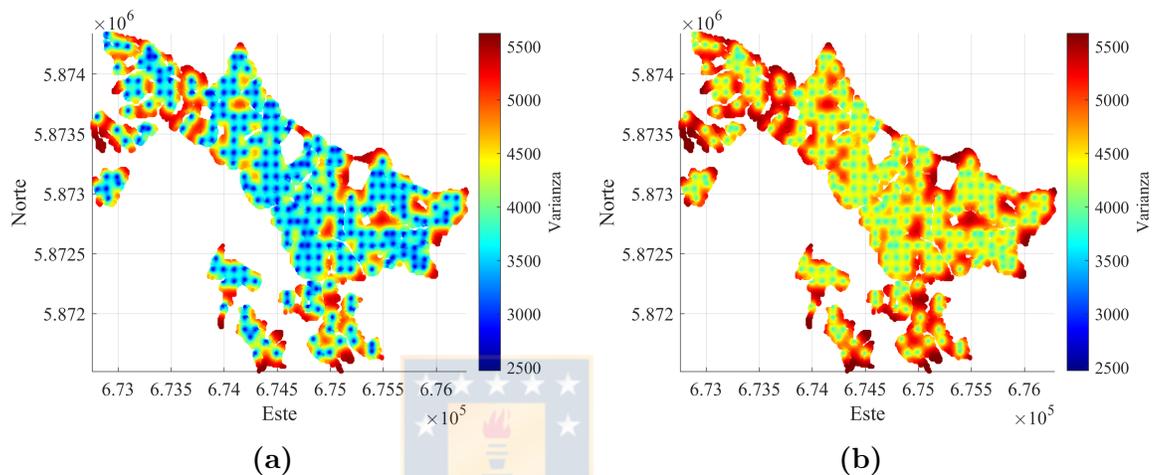


**Figura 4.5.3:** Mapas de la predicción con kriging: (a) condicional a los datos truncados, (b) condicional a los datos truncados y desagrupados

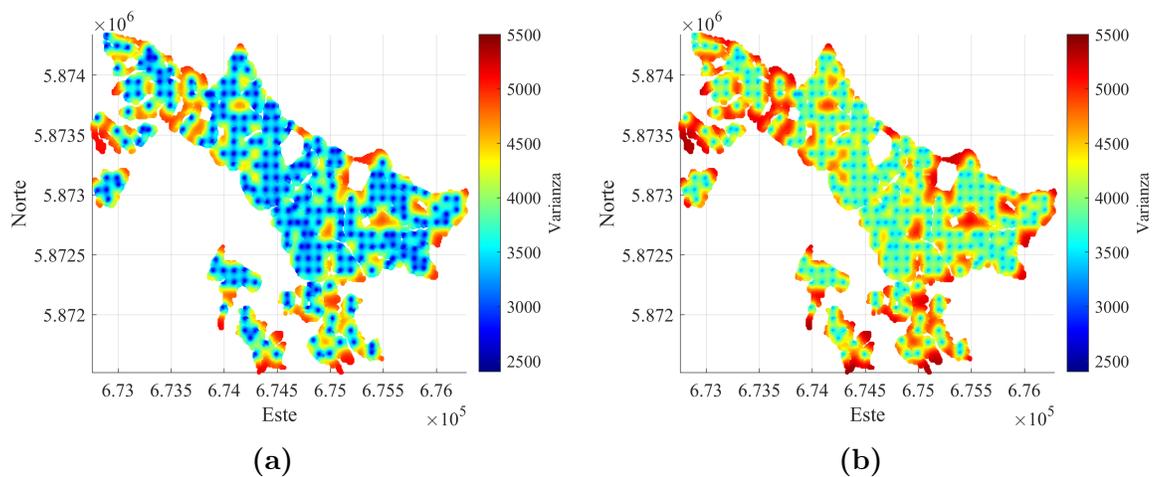
#### 4.5.2. Varianza de las predicciones

Una de las cualidades del kriging es que, como resultado, entrega también el error de predicción; las Figuras 4.5.4 y 4.5.5 muestran los mapas de varianza de predicción

condicional a los distintos set de datos. Se observa que la varianza de predicción varía entre 2500 y 5500  $(\text{m}^3/\text{ha})^2$  y que la varianza de predicción es menor para los datos sin desagrupamiento (Figuras 4.5.4a y 4.5.5a).



**Figura 4.5.4:** Mapas de la varianza de kriging: (a) condicional a los datos originales (b) condicional a los datos originales desagrupados.



**Figura 4.5.5:** Mapas de la varianza de kriging: (a) condicional a los datos truncados, (b) condicional a los datos truncados desagrupados.

### 4.5.3. Validación de los resultados de kriging

A continuación, se presenta en la Tabla 4.5.1 los datos obtenidos a partir de validación cruzada. El error medio y error medio estandarizado de predicción se calculó para los diversos casos donde se aplicó kriging, es decir, kriging con los datos originales, kriging con los datos originales y desagrupados, kriging con los datos truncados y kriging con los datos truncados y desagrupados.

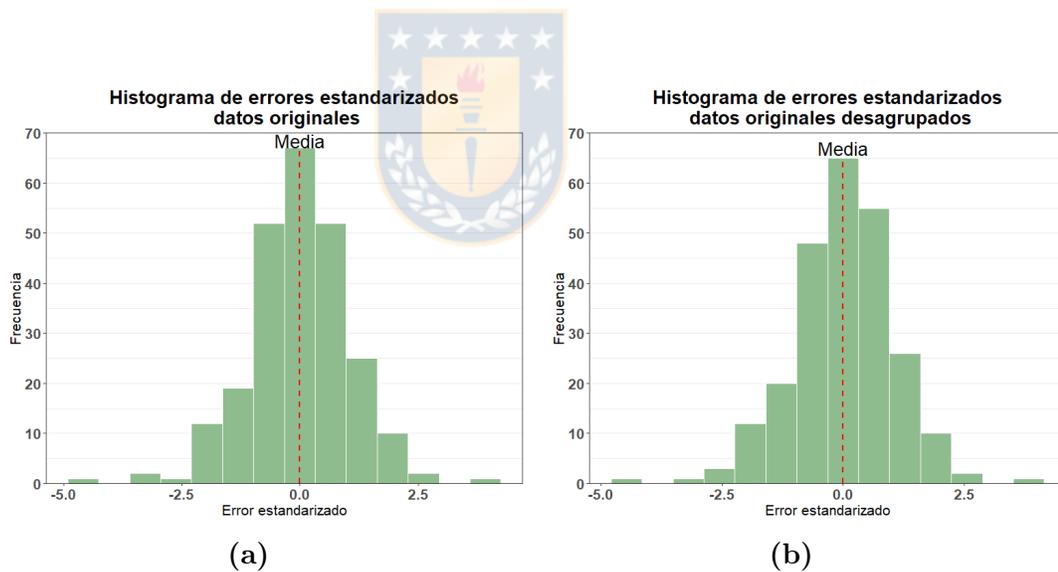
Dada la propiedad de insesgamiento se considera que la mejor predicción es aquella que presenta errores cercanos a 0. Por lo tanto, la predicción con kriging condicional a los datos originales y desagrupados, y la predicción con kriging condicional a los datos truncados y desagrupados son la mejor opción.

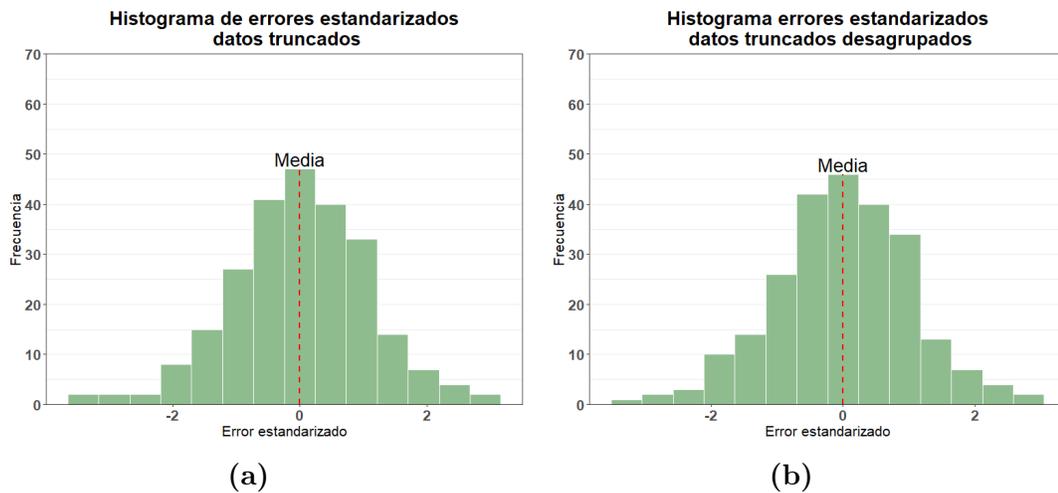
Para evaluar la precisión de la predicción con kriging se considera la desviación estándar (SD) de los errores de predicción, cuyos valores más bajos se observan en los resultados con kriging condicional a los datos truncados.

Además, se utiliza el valor de la desviación estándar de los errores estandarizados, para determinar si el variograma cuantifica adecuadamente la incertidumbre, en este caso el valor más cercano a 1 se da para los resultados con kriging condicional a los datos truncados y desagrupados (1.0260). Respecto a la media de los errores estandarizados, se observa en las Figuras 4.5.6 a 4.5.7, que para todos los set de datos el valor es 0.

**Cuadro 4.5.1:** Errores de predicción obtenidos con Kriging condicional a los diferentes set de datos

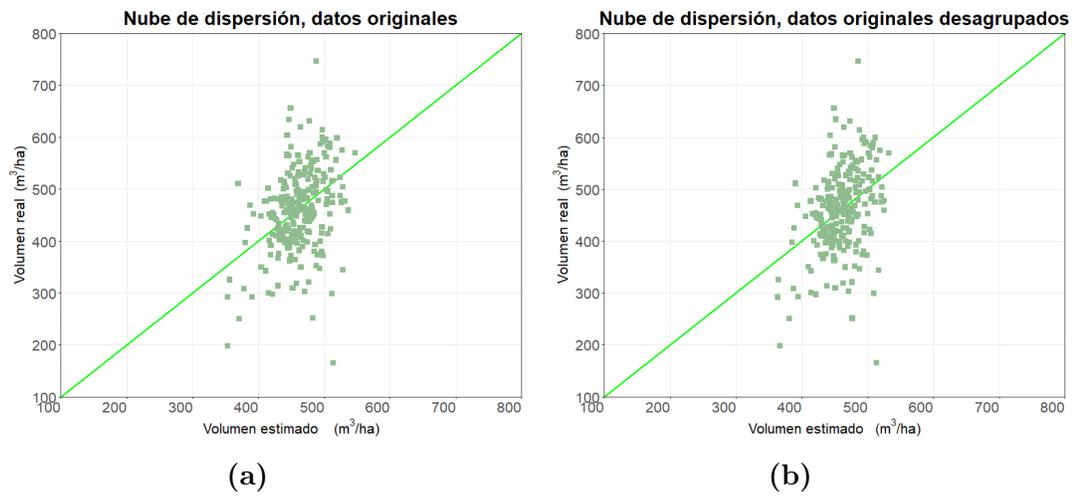
Kriging	Error		Error estandarizado	
	Media	SD	Media	SD
Condicional a datos originales	-0.1348	75.7903	-0.0009	1.0802
Condicional a datos originales y desagrupados	-0.1021	75.5788	-0.0007	1.0425
Condicional a datos truncados	-0.1318	71.7650	-0.0009	1.0656
Condicional a datos truncados y desagrupados	-0.1112	71.6345	-0.00071	1.0260

**Figura 4.5.6:** Histograma de errores de estimación: (a) datos originales, (b) datos originales desagrupados.

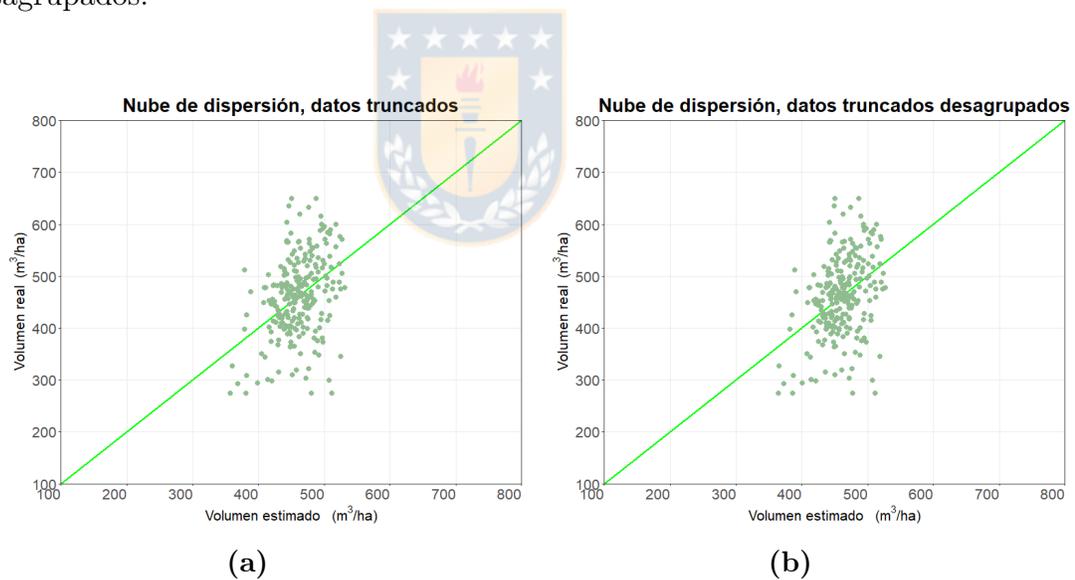


**Figura 4.5.7:** Histograma de errores de estimación: (a) datos truncados, (b) datos truncados desagrupados.

En las Figuras 4.5.8 y 4.5.9, se muestran los gráficos de dispersión para cada uno de los set de datos que se utilizaron para aplicar el método de predicción kriging. En el eje horizontal se presenta el volumen de la muestra y en el eje vertical el volumen estimado, se espera que en un ajuste ideal los datos se encuentren sobre la diagonal (línea verde). Se observa que los gráficos de los resultados, considerando datos truncados, se encuentran más cercanos a la línea verde, lo que coincide con el menor error de predicción de la Tabla 4.5.1.



**Figura 4.5.8:** Nube de dispersión Kriging: (a) datos originales, (b) datos originales desagrupados.



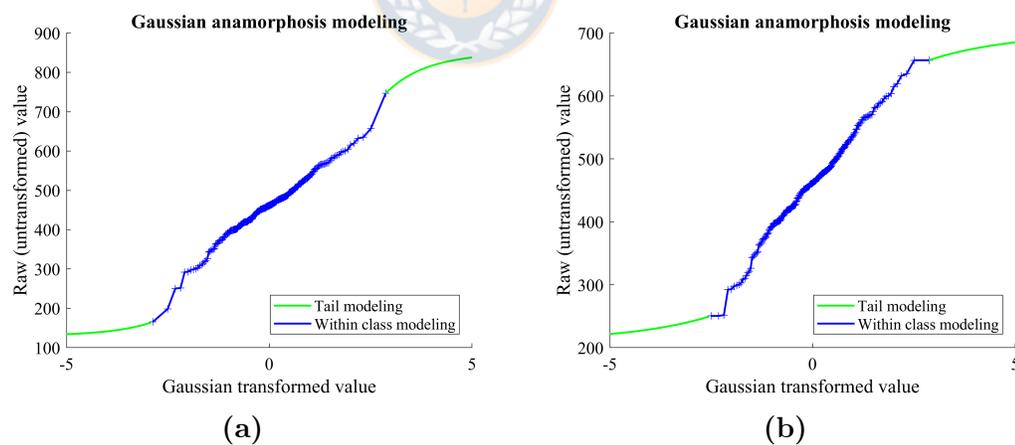
**Figura 4.5.9:** Nube de dispersión Kriging: (a) datos truncados, (b) datos truncados desagrupados.

## 4.6. Predicción con un algoritmo de simulación

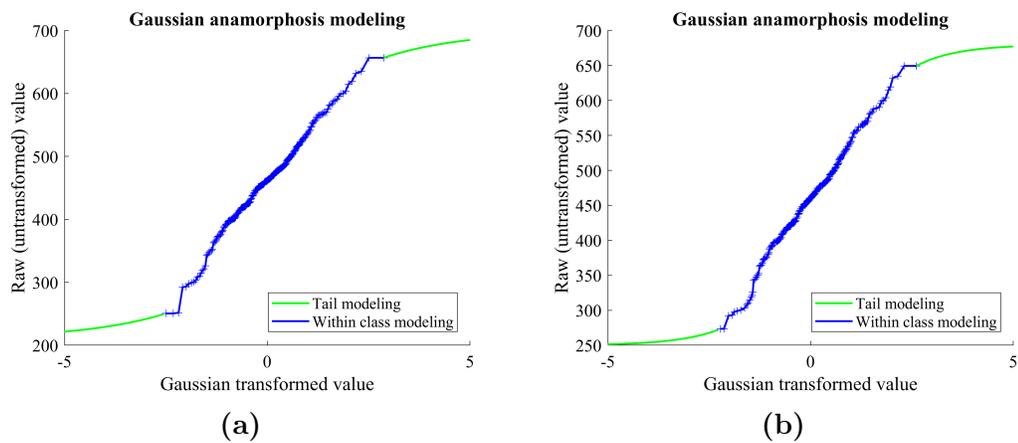
En esta sección se describe, en primer lugar la transformación de datos o anamorfosis Gaussiana, luego los resultados de la simulación con el algoritmo Espectral Bandas Rotantes y, finalmente, se presenta la validación de los resultados.

### 4.6.1. Anamorfosis Gaussiana

El modelo de simulación multigaussiano, al cual pertenece el algoritmo Espectral Bandas Rotantes, requiere que las variables posean distribución Gaussiana. Se requiere realizar un proceso de transformación denominada anamorfosis. En las Figuras 4.6.1 y 4.6.2, se presenta la transformación realizada a la variable volumen ( $\text{m}^3/\text{ha}$ ).



**Figura 4.6.1:** Anamorfosis Gaussiana: (a) datos originales, (b) datos originales desagrupados.



**Figura 4.6.2:** Anamorfosis Gaussiana: (a) datos truncados, (b) datos truncados desagrupados.

En las Figuras 4.6.3 a 4.6.6 se observa cómo cambia el histograma de los valores originales al realizar la transformación Gaussiana, para los distintos set de datos. Se aprecia que la curva de los valores transformados es similar a la distribución normal estándar.

La Tabla 4.6.1 muestra los estadísticos principales de la variable transformada. Se observa que se obtuvieron variable de media cero y varianza uno, como era esperado.

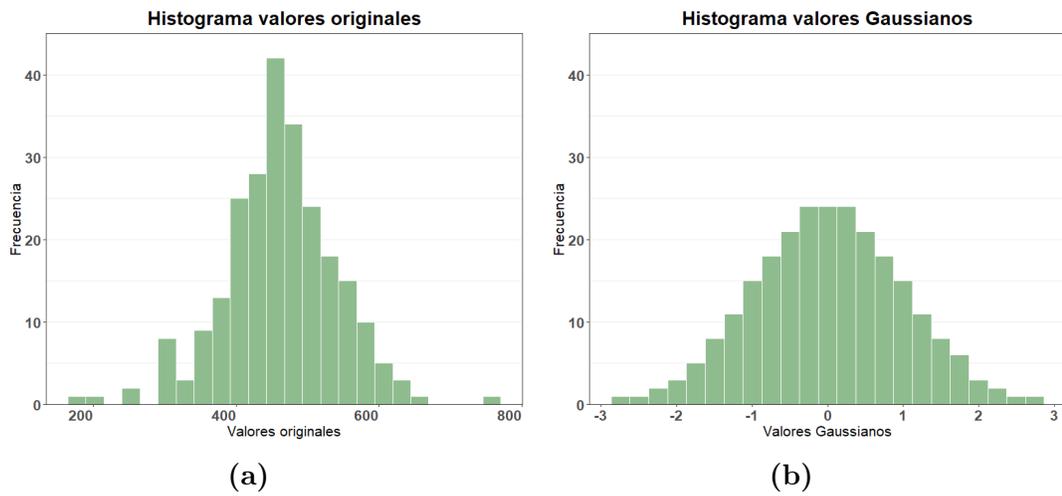


Figura 4.6.3: Histogramas: (a) valores originales, (b) valores Gaussianos.

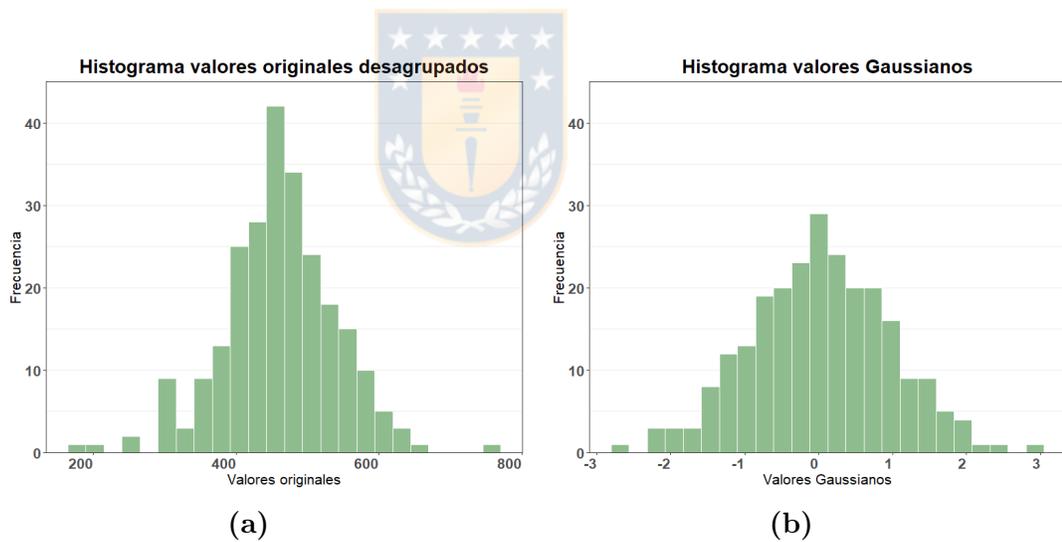
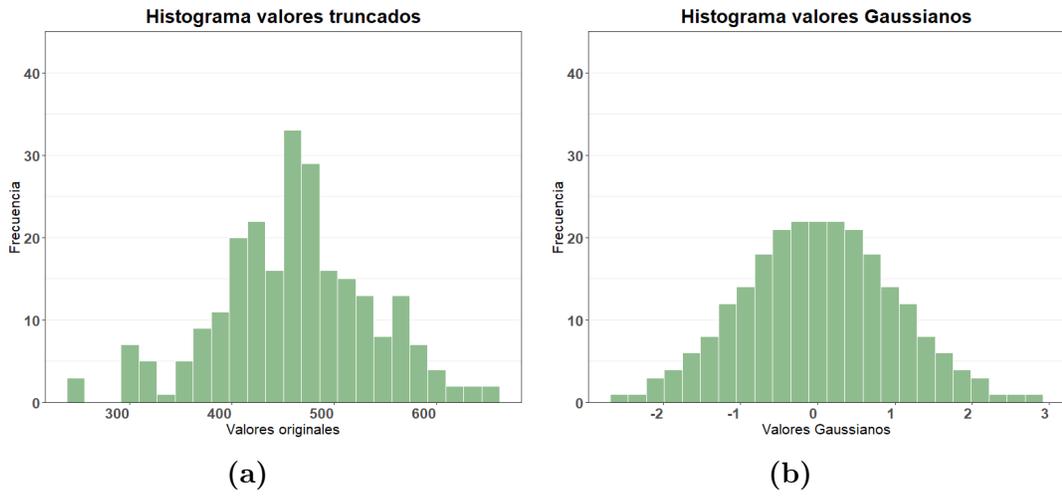
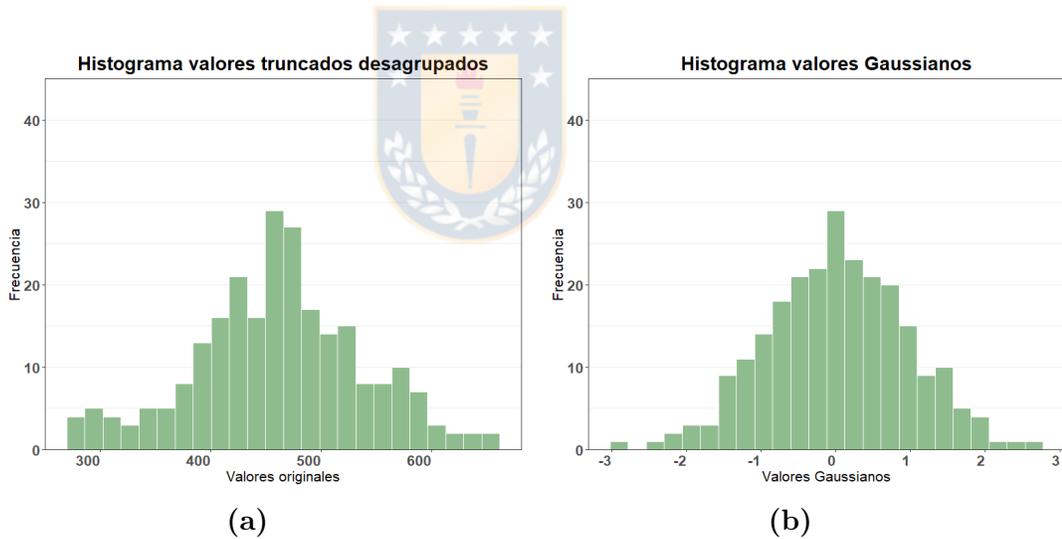


Figura 4.6.4: Histogramas: (a) valores originales desagrupados. (b) valores Gaussianos.



**Figura 4.6.5:** Histogramas: (a) valores truncados (b) valores Gaussianos.



**Figura 4.6.6:** Histogramas: (a) valores truncados desagrupados, (b) valores Gaussianos.

**Cuadro 4.6.1:** Estadísticos de la variable transformada para los diferentes set de datos.

Estadísticos	Set de Datos			
	Originales	Originales Desagrupados	Truncados	Truncados Desagrupados
Media (m <sup>3</sup> /ha)	0.00769	0.01885	0.01181	0.01483
Varianza (m <sup>3</sup> /ha) <sup>2</sup>	0.98845	0.90747	0.96877	0.90639
SD (m <sup>3</sup> /ha)	0.99421	0.95261	0.98427	0.95204
Mínimo (m <sup>3</sup> /ha)	-2.87049	-2.62594	-2.50356	-2.94852
Máximo (m <sup>3</sup> /ha)	2.87049	2.98958	2.87049	2.60788

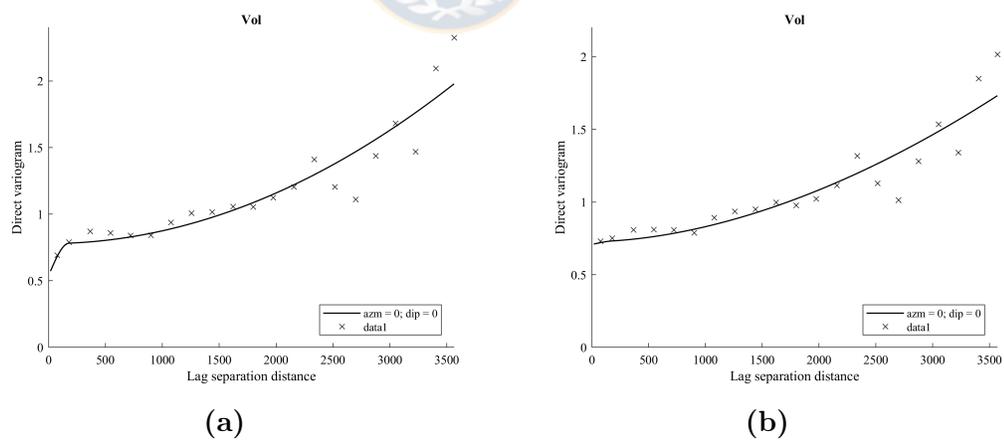
#### 4.6.2. Ajuste Variográfico

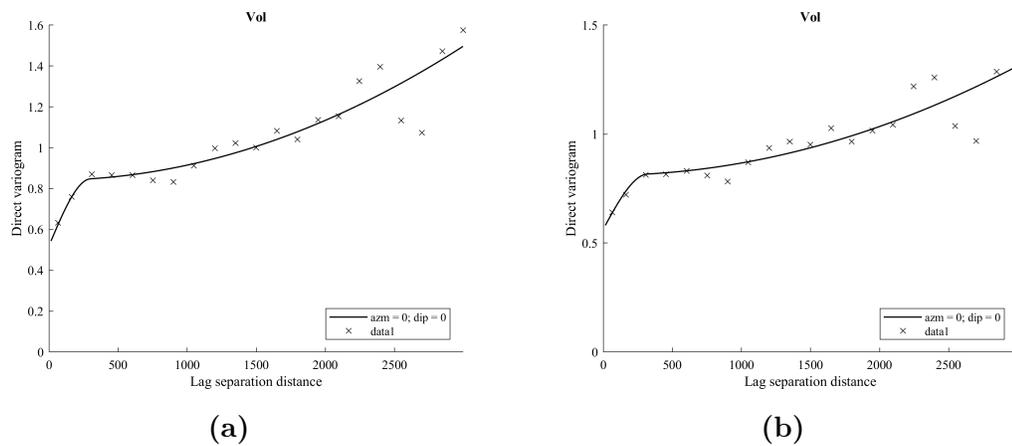
Para la simulación, el ajuste variográfico es ejecutado por un algoritmo semiautomático el cual asegura que la meseta del variograma sea igual a 1 (Emery, 2010), dado que para los campos aleatorios Gaussianos se asume variancia unitaria. Si se observa las Figuras 4.6.7 y 4.6.8, el ajuste parece satisfactorio para los campos aleatorios Gaussianos, dado que el variograma ajustado es cercano a los puntos del variograma experimental en dirección vertical y horizontal.

La Tabla 4.6.2 muestra los parámetros de los variogramas ajustados para cada set de datos.

**Cuadro 4.6.2:** Parámetros del variograma ajustado con datos Gaussianos

Set de datos	Modelo	Alcance (m)	Meseta ( $\text{m}^3/\text{ha}$ ) <sup>2</sup>
Original	Pepita (Nugget)		0.5364
	Esférico	180	0.2428
	Potencia (Power)	600	0.0346
Original Desagrupado	Pepita (Nugget)		0.7081
	Esférico	180	0.0195
	Potencia (Power)	600	0.0413
Truncado	Pepita (Nugget)		0.5193
	Esférico	180	0.3218
	Potencia (Power)	800	0.0474
Truncado Desagrupado	Pepita (Nugget)		0.5639
	Esférico	320	0.2471
	Potencia (Power)	800	0.0362

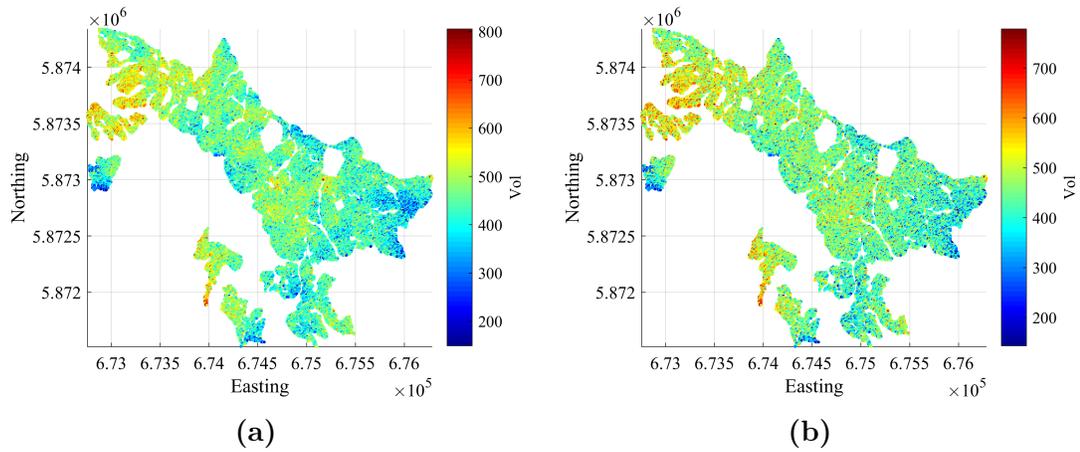
**Figura 4.6.7:** Variograma Experimental y ajustado: (a) valores Gaussianos originales, (b) valores Gaussianos originales desagrupados.



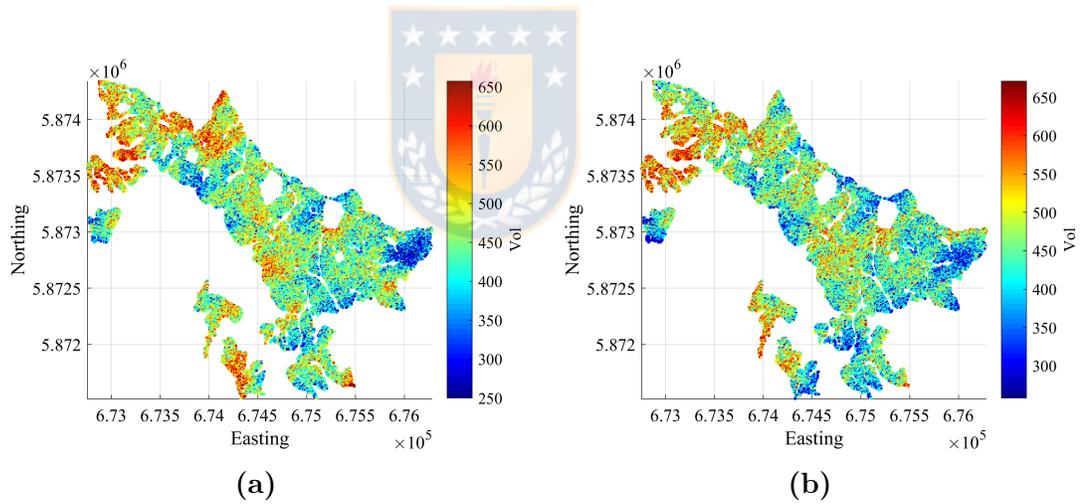
**Figura 4.6.8:** Variograma Experimental y ajustado: (a) valores Gaussianos truncados, (b) valores Gaussianos truncados desagrupados.

### 4.6.3. Visualización de realizaciones

Para visualizar el resultado de las simulaciones con el algoritmo Espectral Bandas Rotantes (Arroyo and Emery, 2018) condicional a los distintos set de datos, se presenta una realización (de las 100 realizaciones), el cual se observa en las Figuras 4.6.9 y 4.6.10. Si se comparan los mapas de las realizaciones con los mapas de predicción del kriging (Figuras 4.5.2 y 4.5.3), se aprecia una mayor irregularidad en los mapas provenientes de las simulaciones, demostrando que la simulación reproduce la variabilidad de la variable de interés, mientras que el kriging produce suavizamiento.



**Figura 4.6.9:** Mapa de una realización: (a) simulación condicional a los datos originales, (b) simulación condicional a los datos originales desagrupados.



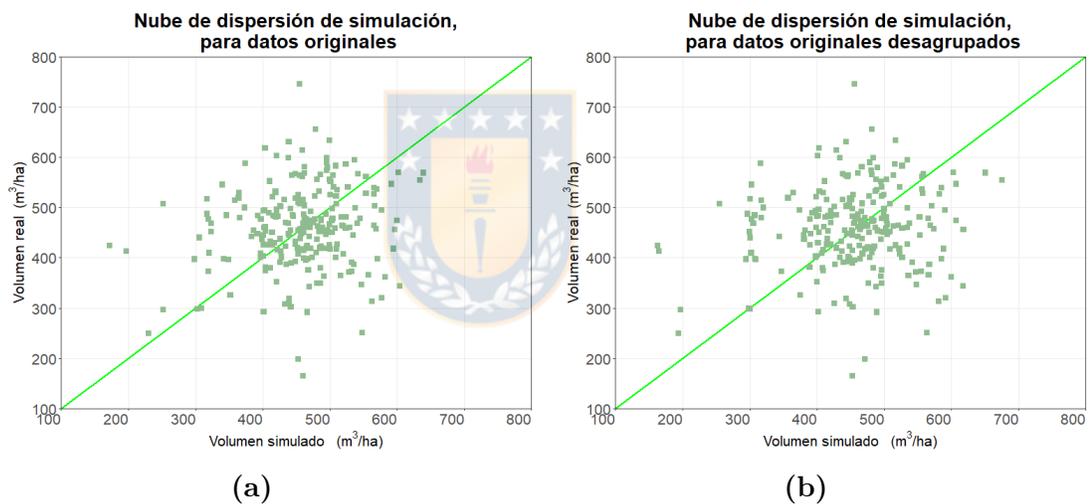
**Figura 4.6.10:** Mapa de una realización: (a) simulación condicional a los datos truncados, (b) simulación condicional a los datos truncados desagrupados.

#### 4.6.4. Validación de la simulación

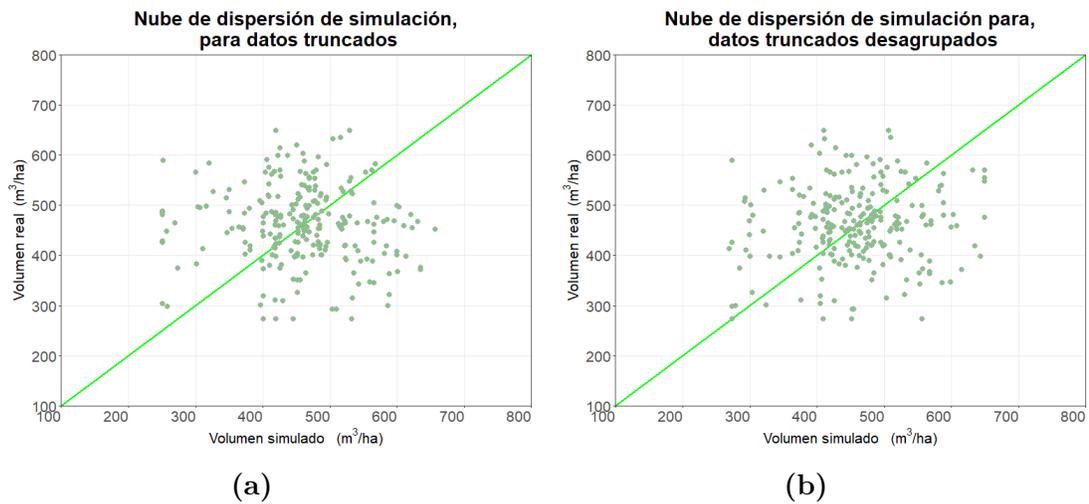
Con el objetivo de determinar la eficiencia del modelo empleado en la simulación y además verificar los parámetros empleados, se efectuó el proceso de validación cruzada. Esto consiste en la simulación de cada dato en función de los datos restantes,

permitiendo analizar la diferencia que se produce entre los valores simulados y el dato real, lo que se denomina **error de la simulación**. En este caso se consideran 100 realizaciones.

A continuación, se presentan nubes de dispersión de una de las realizaciones simulada mediante validación cruzada. Se observa en las Figuras 4.6.11b y 4.6.12b, que los datos originales desagrupados y truncados desagrupados se ubican más cercanos a la línea diagonal, es decir, el valor predicho es cercano al valor real.



**Figura 4.6.11:** Nube de dispersión de una realización versus valores reales: (a) simulación condicional a los datos originales, (b) simulación condicional a los datos originales desagrupados.



**Figura 4.6.12:** Nube de dispersión de una realización versus valores reales: (a) simulación condicional a los datos truncados, (b) simulación condicional a los datos truncados y desagrupados.

En las Figuras 4.6.13 a 4.6.16, se presentan histogramas para cada set de datos: originales, originales desagrupados, truncados y truncados desagrupados. Cada figura está compuesta por 3 histogramas, construidos con datos provenientes del inventario forestal, datos predichos por kriging universal y datos simulados con el algoritmo Espectral Bandas Rotantes.

En la Figura 4.6.13 se observa que el histograma de los datos simulados es muy similar al histograma elaborado con los datos del inventario, es decir, la simulación reproduce la variabilidad espacial de los datos. Para complementar lo anterior, en los estadísticos presentados en la Tabla 4.6.3 se observa que los valores mínimos y máximos de la simulación son más cercanos a los datos de inventario que los generados por kriging, sin embargo el kriging es un estimador insesgado y de varianza mínima, ya que se obtiene una media de  $460.4 \text{ (m}^3/\text{ha)}$  y varianza de  $1151.21 \text{ (m}^3/\text{ha)}^2$ .

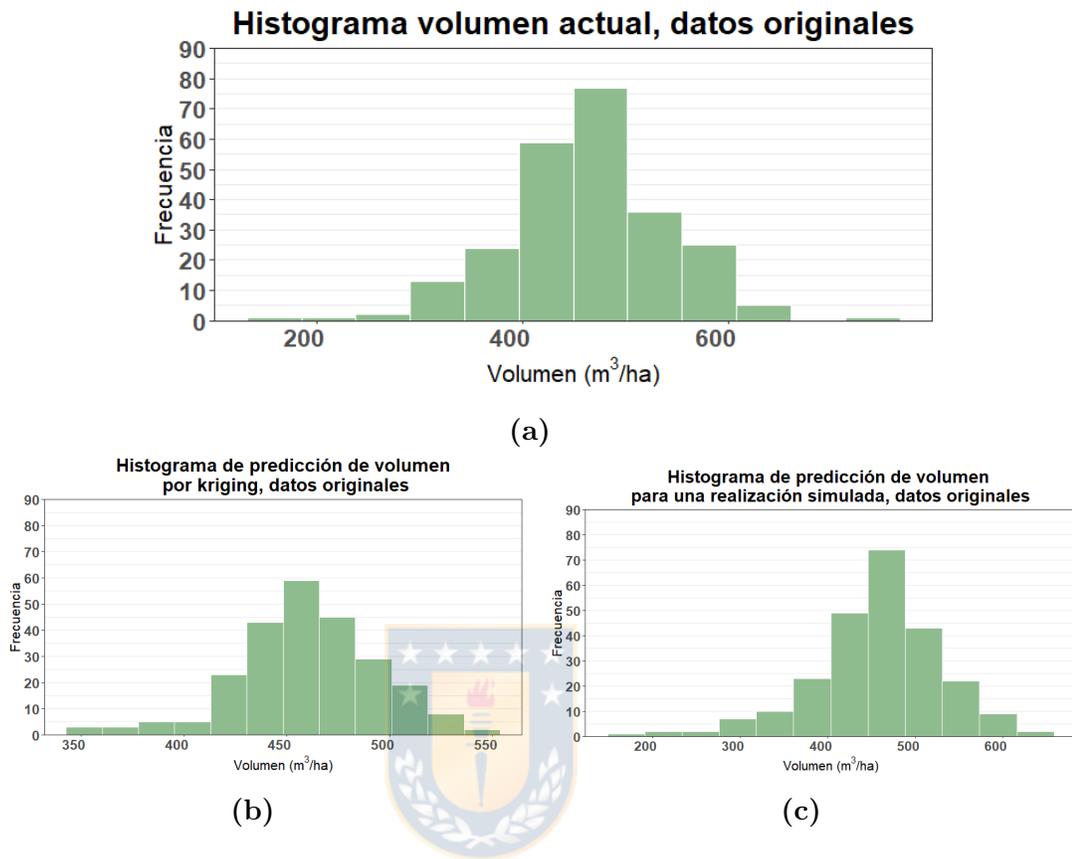
En las Figuras 4.6.14, 4.6.15 y 4.6.16 se reitera que el histograma elaborado con

los datos de simulación es similar al histograma que considera los datos muestreados, y los estadísticos de las Tablas 4.6.4, 4.6.5, 4.6.6 muestran que los valores mínimos y máximos de ambos son cercanos.

Respecto al kriging, las mínimas varianzas se observan en los datos originales desagrupados (Tabla 4.6.4) y datos truncados desagrupados (Tabla 4.6.6).

**Cuadro 4.6.3:** Estadísticos datos originales, volumen de inventario y predicho por kriging y simulación

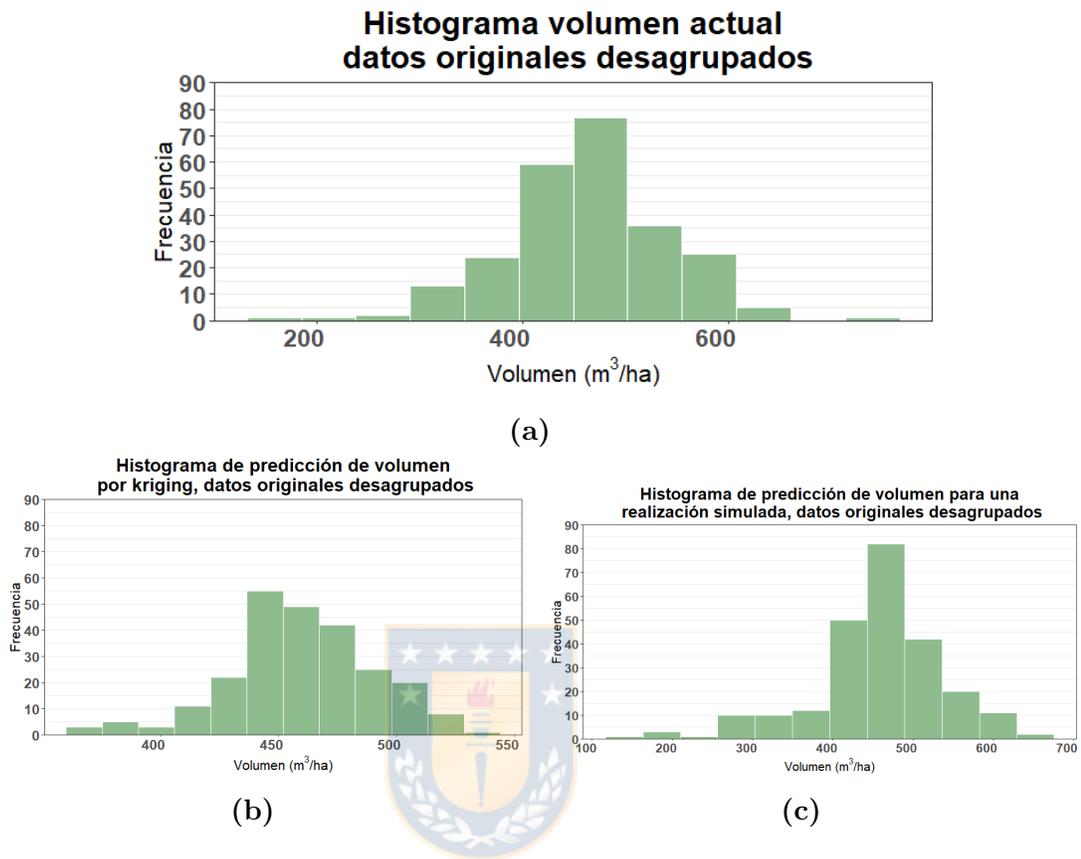
Estadísticos	Volumen (m <sup>3</sup> /ha)		
	Inventario	Kriging	Simulación
Media	460.23	460.4	456.6
Mediana	461.50	459.9	460.5
Mínimo	165.5	352.9	171.3
Máximo	746.7	546.3	638.7
Varianza	6486.4	1151.2	5310.8



**Figura 4.6.13:** Histogramas de datos originales provenientes de: (a) inventario forestal, (b) kriging universal, (c) simulación con algoritmo Espectral Bandas Rotantes.

**Cuadro 4.6.4:** Estadísticos datos originales desagrupados, volumen de inventario y predicho por kriging y simulación

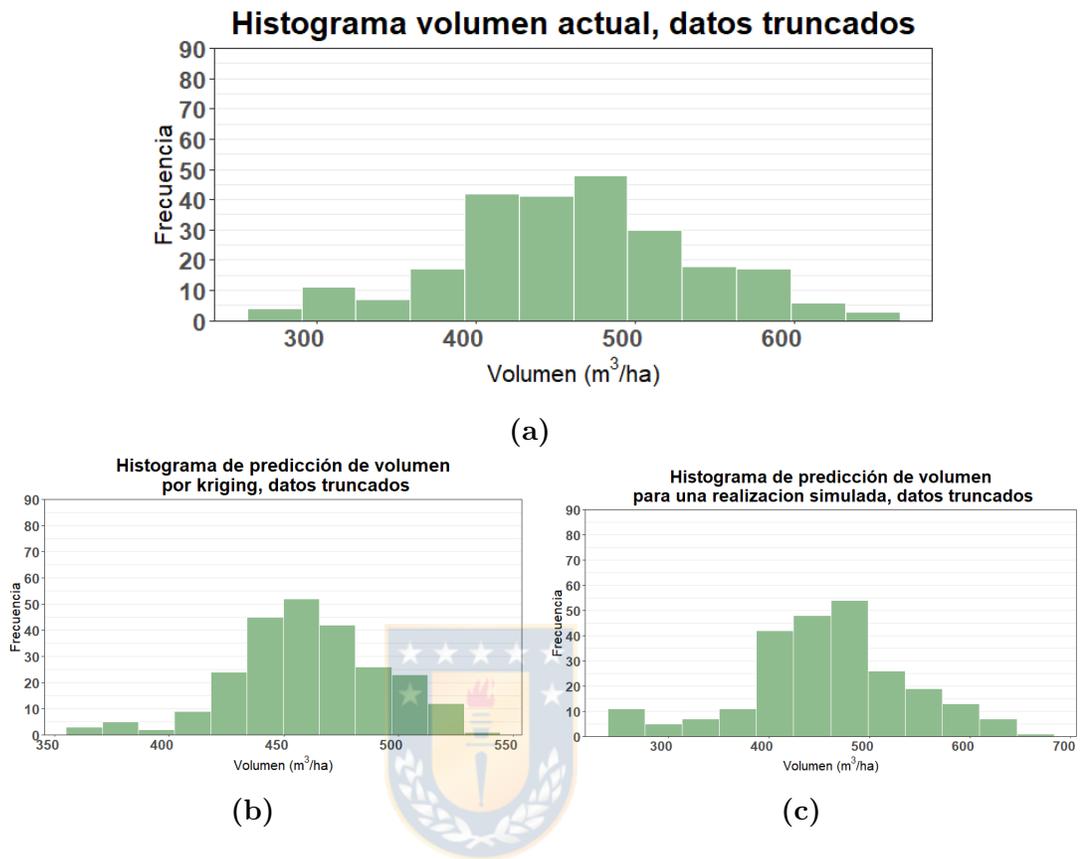
Estadísticos	Volumen ( $\text{m}^3/\text{ha}$ )		
	Inventario	Kriging	Simulación
Media	460.23	460.3	455.4
Mediana	461.5	460.9	460.5
Mínimo	165.5	363.6	162.5
Máximo	746.7	532.1	675.2
Varianza	6486.4	948.3	6680.4



**Figura 4.6.14:** Histogramas de datos originales desagrupados provenientes de: (a) inventario forestal, (b) kriging universal, (c) simulación con algoritmo Espectral Bandas Rotantes.

**Cuadro 4.6.5:** Estadísticos datos truncados, volumen de inventario y predicho por kriging y simulación

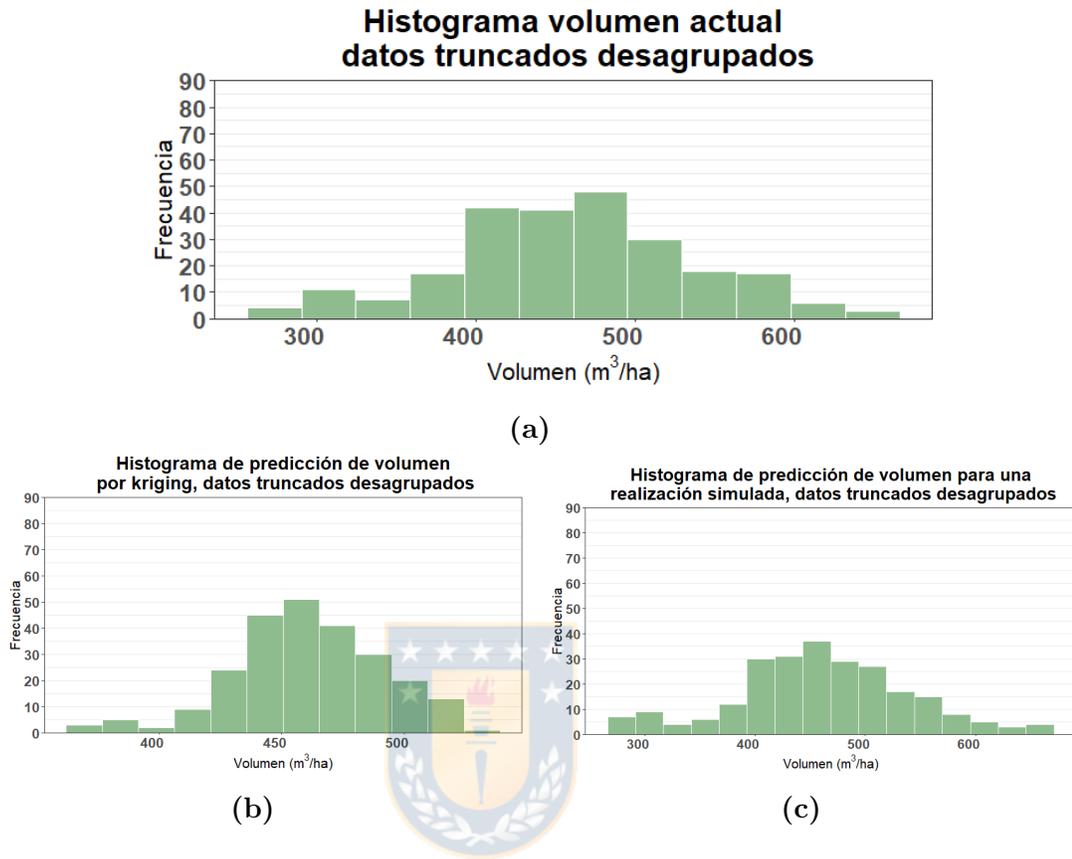
Estadísticos	Volumen ( $\text{m}^3/\text{ha}$ )		
	Inventario	Kriging	Simulación
Media	460.7	460.9	459.3
Mediana	461.5	460.6	460.5
Mínimo	273.5	357.5	250.5
Máximo	649.4	531.0	656.7
Varianza	5861.7	1060.2	6746.5



**Figura 4.6.15:** Histogramas de datos truncados provenientes de: (a) inventario forestal, (b) kriging universal, (c) simulación con algoritmo Espectral Bandas Rotantes.

**Cuadro 4.6.6:** Estadísticos datos truncados desagrupados, volumen de inventario y predicho por kriging y simulación

Estadísticos	Volumen ( $\text{m}^3/\text{ha}$ )		
	Inventario	Kriging	Simulación
Media	460.7	460.8	459.1
Mediana	461.5	460.6	458.1
Mínimo	273.5	364.6	269.4
Máximo	649.4	527.1	649.4
Varianza	5861.7	952.2	6428.4



**Figura 4.6.16:** Histogramas de datos truncados desagrupados provenientes de: (a) inventario forestal, (b) kriging universal, (c) simulación con algoritmo Espectral Bandas Rotantes.

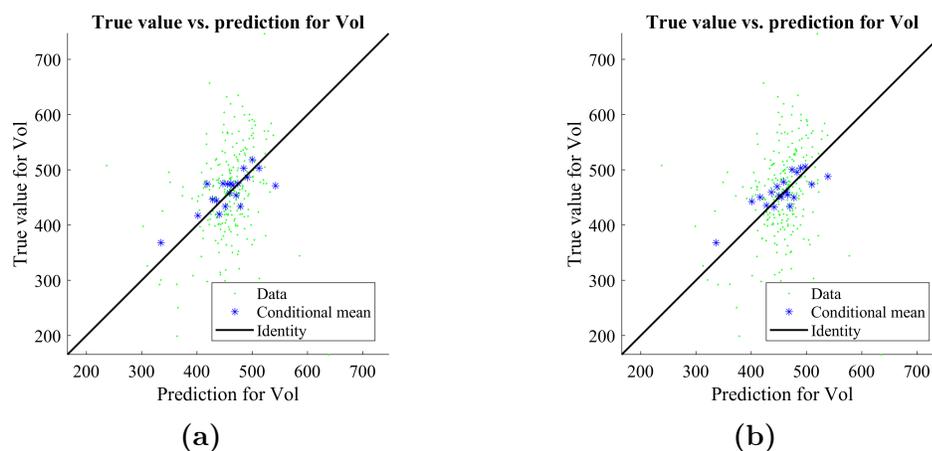
Por último, en la Tabla 4.6.7 se presentan los estadísticos de los errores de predicción de la simulación con algoritmo Espectral Bandas Rotantes condicional a los distintos set de datos. Se observa que el error de predicción con menor media, en valor absoluto, es para la simulación condicional a los datos truncados mientras que la menor desviación estándar ( $SD=99.813$ ) se genera en la simulación condicional a los datos originales.

**Cuadro 4.6.7:** Errores de predicción obtenidos con simulación por el método Espectral Bandas Rotantes condicional a los diferentes set de datos

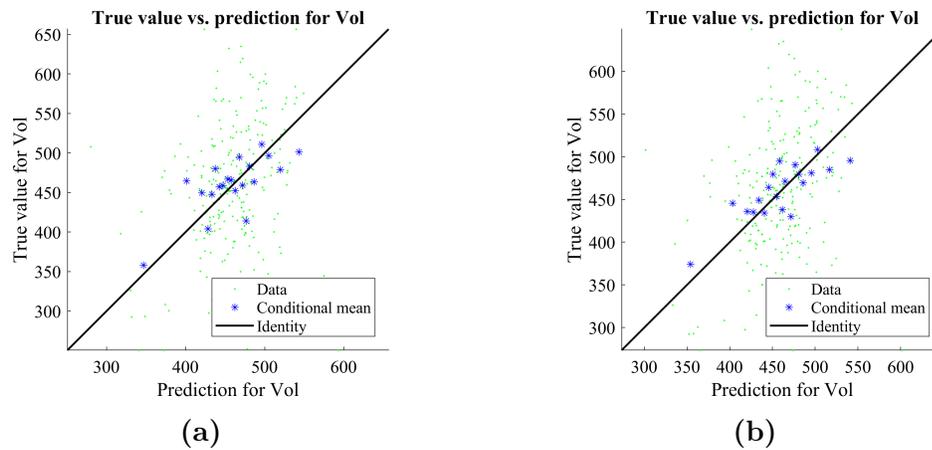
Simulación	Error de Predicción ( $\text{m}^3/\text{ha}$ )	
	Media	SD
Condicional a datos originales	-3.63	99.813
Cond. a datos originales y desagrupados	-4.85	108.108
Condicional a datos truncados	-1.48	116.011
Cond. a los datos truncados y desagrupados	-1.63	104.993

#### 4.6.5. Medición de la exactitud

Las Figuras 4.6.17 y 4.6.18 presentan las nubes de dispersión entre los valores reales y el promedio de las realizaciones para cada variable de interés. Es posible apreciar que en todos los casos los puntos azules se encuentran bastante cercanos a la diagonal, sin embargo en los datos desagrupados, originales y truncados (Figuras 4.6.17b y 4.6.18b), se visualiza mayor exactitud.



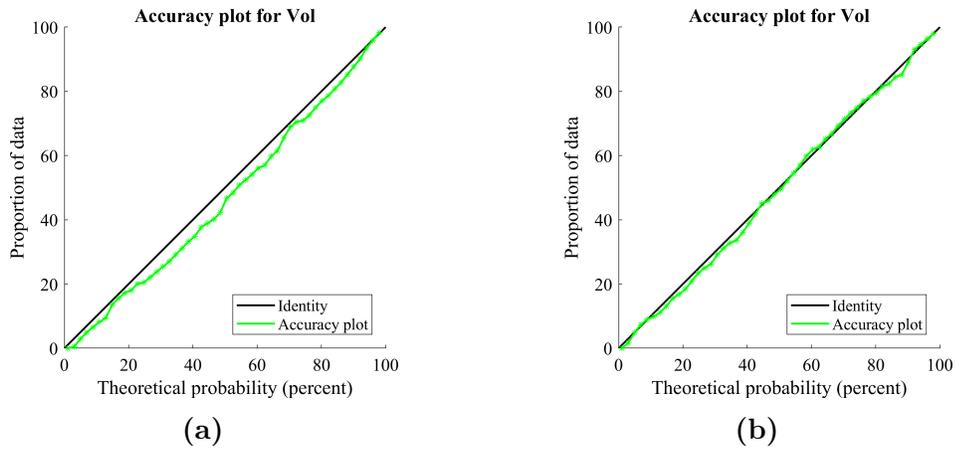
**Figura 4.6.17:** Nubes de dispersión de valores simulados promedio sobre 100 realizaciones versus valores reales, considerando el condicionamiento a los (a) datos originales, (b) datos originales desagrupados.



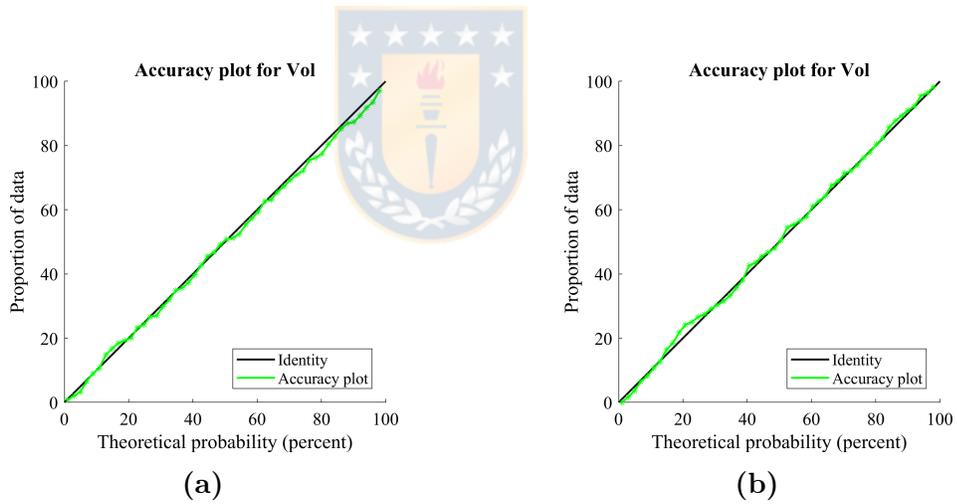
**Figura 4.6.18:** Nubes de dispersión de valores simulados promedio sobre 100 realizaciones versus valores reales, considerando el condicionamiento a los (a) datos truncados, (b) datos truncados desagrupados

#### 4.6.6. Cuantificación de la precisión

Los diagramas de incertidumbre permiten validar las simulaciones en términos de precisión, esto es, dado un intervalo de probabilidad de la simulación, la fracción de valores reales pertenecientes a dicho intervalo es igual o superior al valor del intervalo. Las Figuras 4.6.19 y 4.6.20 muestran los diagramas de incertidumbre para los 4 set de datos, se observa mayor precisión en la simulación condicional a los datos originales desagrupados (Figura 4.6.19b) y simulación condicional a los datos truncados desagrupados (Figura 4.6.20b), ya que en ambos casos la línea verde está muy cercana a la diagonal.



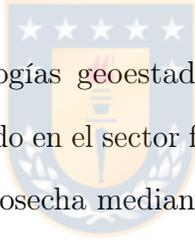
**Figura 4.6.19:** Diagramas de Incertidumbre (a) datos originales, (b) datos originales desagrupados.



**Figura 4.6.20:** Diagramas de Incertidumbre (a) datos truncados, (b) datos truncados desagrupados.

## Capítulo 5

### Conclusión



La aplicación de metodologías geoestadísticas para el análisis de inventarios forestales ha sido poco explorado en el sector forestal chileno. La presente tesis expone la estimación de volumen de cosecha mediante kriging y simulación condicional con el método Espectral de Bandas Rotantes.

Las metodologías geoestadísticas requieren previamente un detallado análisis de los datos con el objetivo de calcular y ajustar acuciosamente el variograma experimental. El variograma es la base del análisis espacial y, si hay errores en esta etapa, la predicción por kriging o simulación no será de buena calidad.

Respecto a la utilización de kriging universal (tipo de kriging aplicado en la tesis) se infiere lo siguiente:

- El cálculo del variograma experimental y posterior ajuste variográfico fue dificultoso.
  - El uso de kriging permitió obtener predicciones de la variable de interés a partir del cálculo del variograma y entregó información acerca de la continuidad
-

espacial de la variable.

- Se obtuvieron estimadores insesgados y de mínima varianza en todos los set de datos analizados.
- Las menores desviaciones estándar de los errores de predicción se generaron con el kriging condicional a los datos truncados y desagrupados ( $71.63 \text{ m}^3/\text{ha}$ ) y condicional a los datos truncados ( $71.76 \text{ m}^3/\text{ha}$ ).
- La media de los errores de predicción estandarizados para los distintos set de datos es 0 y la desviación estándar es cercana a 1, encontrándose la más cercana a 1 con el kriging condicional a los datos truncados y desagrupados (1.02).
- La correlación de los valores estimados con los originales se obtuvo mediante el coeficiente de correlación lineal de Pearson, fluctuando entre 0.35 y 0.36.
- Desde el punto de vista de la aplicabilidad de los resultados, se considera que el mapa de predicción con kriging de la variable de interés exhibe gran utilidad, ya que permitiría, por ejemplo, planificar la cosecha operativa priorizando aquellas zonas en que se obtendrá mayor volumen.
- El mapa de varianza del kriging muestra la variabilidad del error del kriging, información valiosa, desde el punto de vista operativo, ya que permite concentrar los recursos disponibles para la realización de inventarios en estos sectores y/o evaluar la necesidad de aumentar los puntos de muestreo.

Respecto a la metodología con simulación, los mapas de las realizaciones entregan información más detallada de la continuidad y variabilidad espacial de la variable, de donde se concluye lo siguiente:

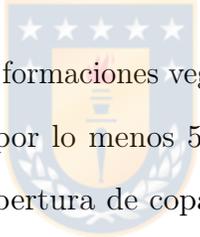
- El ajuste variográfico, en general, fue complejo.

- El procesamiento de las realizaciones permitió obtener escenarios para la variable volumen. En relación al valor medio de la variable simulada, se observó que, en general, su distribución es de tipo normal.
- Para validar el modelo propuesto, se compararon los valores predichos con kriging o simulados con el algoritmo Espectral Bandas Rotantes, con los valores condicionantes.
- La reproducción de las estadísticas de las muestras presentó errores promedio, cercanos a cero, mientras que la correlación entre los valores reales y el promedio de las 100 realizaciones simuladas tomó valores entre 0.23 y 0.28.

En relación al objetivo principal planteado durante el desarrollo de esta tesis de Magíster, se logró la predicción del volumen con métodos geoestadísticos, como el kriging y la simulación. En definitiva, se concluye que la predicción con kriging y con el algoritmo de simulación Espectral Bandas Rotantes de la variable de interés responde a los alcances planteados y se presenta como alternativa en la determinación de valores no conocidos en toda la zona y determinación de escenarios posibles.

## Capítulo 6

### Glosario



**Bosque:** Sitio poblado con formaciones vegetales en las que predominan árboles y que ocupa una superficie de por lo menos 5.000 metros cuadrados, con un ancho mínimo de 40 metros, con cobertura de copa arbórea que supere el 10 % de dicha superficie total en condiciones áridas y semiáridas y el 25 % en circunstancias más favorables (MINISTERIO DE AGRICULTURA, GOBIERNO DE CHILE, 2008, p. 1).

**Bosque nativo:** Bosque formado por especies autóctonas, provenientes de generación natural, regeneración natural, o plantación bajo dosel con las mismas especies existentes en el área de distribución original, que pueden tener presencia accidental de especies exóticas distribuidas al azar (MINISTERIO DE AGRICULTURA, GOBIERNO DE CHILE, 2008, p. 1).

**Especies:** Un grupo de organismos (plantas o animales) que son muy similares en apariencia y pueden cruzarse libremente entre sí, pero no con otros grupos (Prodan

---

et al., 1997, p. 29).

**Índice de sitio:** Una expresión de la calidad del sitio forestal, basada en la altura media de los árboles dominantes y codominantes de una especie, a una edad especificada (Prodan et al., 1997, p. 27).

**Inventario Forestal:** Proceso de muestreo de un bosque o rodal forestal, utilizado para obtener una estimación del volumen de madera y valor (Mercker, 2017, p. 13).



**Hectárea:** Unidad de superficie, una hectárea equivale a 10000 m<sup>2</sup> (Prodan et al., 1997, p. 7).

**Metros<sup>3</sup> sólidos** o m<sup>3</sup>: Unidad de medida de volumen de madera arrumada o apilada. La ruma o pila de madera es de 1 m × 1 m × 1 m (Prodan et al., 1997, p. 11).

**Parcela:** Área donde se colectan datos para proveer información acerca del bosque. Varias parcelas constituyen un inventario forestal (Mercker, 2017, p. 23).

**Pendiente:** La inclinación del terreno, usualmente expresado como el cambio en elevación (en metros) sobre 100 metros de distancia horizontal (Mercker, 2017, p. 28).

**Plantación:** Un rodal establecido por árboles plantados (Mercker, 2017, p. 23).

**Rodal:** Un rodal es un conjunto espacialmente continuo de árboles y otros vegetales asociados, suficientemente uniforme para constituir una unidad de manejo, la que puede diferenciarse claramente de los rodales adyacentes por su edad, composición de especies, estructura, o por las características fisiográficas o climáticas en que crece. Los rodales son la unidad de manejo básica de un bosque (Cancino, 2012, p. 22).

**Silvicultura:** Conjunto de técnicas y conocimientos relativos al cultivo de bosques. Se practica en 4 etapas: Establecimiento, operaciones intermedias, cosecha y protección del rodal (Mercker, 2017, p. 27).

**Volumen total:** Volumen correspondiente al fuste o tronco principal de un árbol (Prodan et al., 1997, p. 94).



## Bibliografía

- Arroyo, D. and Emery, X. (2018). Simulation of intrinsic random fields of order  $k$  with a continuous spectral algorithm. *Stochastic Environmental Research and Risk Assessment*, 32:3245–3255.
- Cancino, J. (2012). *Dendrometría Básica*. Facultad de Ciencias Forestales, Universidad de Concepción.
- Chica-Olmo, M. (2005). La geoestadística como herramienta de análisis espacial de datos de inventario forestal. *Cuadernos de la Sociedad Española de Ciencias Forestales*, 19:47–55.
- Chilès, J. P. and Delfiner, P. (2012). *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons, New York, 2nd edition.
- Daisheng, T. and Bian, F. (2009). Forest site evaluation based on GIS and Kriging. In *2009 First International Conference on Information Science and Engineering*, pages 2063–2067.
- Emery, X. (2010). Iterative algorithms for fitting a linear model of coregionalization. *Computers & Geosciences*, 36:1150–1160.
- Emery, X. (2013). *Geoestadística*. Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile.
- Emery, X., Arroyo, D., and Porcu, E. (2016). An improved spectral turning-bands algorithm for simulating stationary vector Gaussian random fields. *Stochastic Environmental Research and Risk Assessment*, 30(7):1863–1873.
- Emery, X. and Lantuéjoul, C. (2006). Tbsim: A computer program for conditional simulation of three-dimensional Gaussian random fields via the turning bands method. *Computers & Geosciences*, 32:1615–1628.
- Hernández, J. and Emery, X. (2009). A geostatistical approach to optimize sampling designs for local forest inventories. *Canadian Journal of Forest Research*, 39(8):1465–1474.
-

- Isaaks, E. H. and Srivastava, M. R. (1989). *Applied geostatistics*. Oxford University Press, Inc.
- Lantuéjoul, C. (2002). *Geostatistical Simulation: Models and Algorithms*. Springer Berlin Heidelberg.
- Leuangthong, O., McLennan, J. A., and Deutsch, C. V. (2004). Minimum acceptance criteria for geostatistical realizations. *Natural Resources Research*, 13(3):131–141.
- Matérn, B. (1960). *Spatial Variation*. Lecture Notes in Statistics. Springer-Verlag.
- Matheron, G. (1963). Principles of geostatistics. *Economic geology*, 58(8):1246–1266.
- Matheron, G. and Kleingeld, W. (1987). The evolution of geostatistics. In *APCOM 87. Proceedings of the Twentieth International Symposium on the Application of Computers and Mathematics in the Mineral Industries*, volume 3, pages 9–12.
- Mercker, D. (2017). A Glossary of Common Forestry Terms. Forestry terms.
- MINISTERIO DE AGRICULTURA, GOBIERNO DE CHILE (2008). Ley 20283: Ley sobre recuperación del bosque nativo y fomento forestal. Ley forestal.
- Oliver, M. A. and Webster, R. (2015). *Basic steps in geostatistics: the variogram and kriging*. Springer.
- Ortiz, J. (2004). *Apuntes del curso MI-54A Evaluación de yacimientos*. Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile.
- Prodan, M., Peters, R., Cox, F., and Real, P. (1997). *Mensura Forestal, GTZ/IICA*. Serie de Investigación y Educación en Desarrollo Sostenible, San José, Costa Rica.
- Raimundo, M. R., Scolforo, H. F., de Mello, J. M., Scolforo, J. R. S., McTague, J. P., and dos Reis, A. A. (2017). Geostatistics applied to growth estimates in continuous forest inventories. *Forest Science*, 63(1):29–38.
- Sarma, D. D. (2010). *Geostatistics with applications in earth sciences*. Springer Science & Business Media.
- Soto, D., Gysling, J., Kahler, C., Álvarez, V., Poblete, P., and Evaristo, P. (2020). Anuario forestal 2020. *Boletín Estadístico N° 174*.
- Tukey, J. W. (1977). *Exploratory data analysis*. Addison Wesley.
- Webster, R. and Oliver, M. A. (2007). *Geostatistics for environmental scientists*. John Wiley & Sons.
- Zhang, J. and Yao, N. (2008). The geostatistical framework for spatial prediction. *Geo-Spatial Information Science*, 11:180–185.