



FACULTAD DE
CIENCIAS ECONÓMICAS
Y DE ADMINISTRACIÓN



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

UNIVERSIDAD DE LA REPÚBLICA
FACULTAD DE CIENCIAS ECONÓMICAS Y DE
ADMINISTRACIÓN

TRABAJO FINAL DE GRADO PARA OBTENER EL TÍTULO
DE LICENCIADO EN ESTADÍSTICA

Predicción espacial del Área Basal del Bosque Nativo
en la Cuenca Alta del Río Negro.

Cecilia Alonso y Hugo De Francisco

Tutora:
María Eugenia Riaño

Montevideo
URUGUAY
2022

UNIVERSIDAD DE LA REPÚBLICA

El tribunal docente integrado por los abajo firmantes aprueba el trabajo final de grado:

**Predicción espacial del Área Basal del Bosque Nativo
en la Cuenca Alta del Río Negro.**

Cecilia Alonso y Hugo De Francisco

Tutora:
María Eugenia Riaño

Licenciatura en Estadística

Calificación: 12.

Fecha: 29 de abril del año 2022.

Tribunal:

Profesora Natalia Berberían _____

Profesor Fernando Massa _____

Profesor Ramón Álvarez-Vaz _____

Agradecimientos

Queremos agradecer a Eugenia Riaño por su permanente ayuda, ánimo y dedicación desde su rol de tutora del presente Trabajo Final de Grado. También se agradece a Leonardo Boragno, Mariana Boscana y María Emilia Arriaga de la DGF del MGAP, por sus diversos aportes y por haber proporcionado datos para realizar este documento.

Por otra parte, agradecemos muy especialmente al Instituto de Estadística, y a la Udelar en general, por las herramientas brindadas y por habernos dado la oportunidad de conocernos y forjar una linda y larga amistad.

Hugo:

A toda mi familia, en especial a Adrián y Lautaro, por apoyarme durante este proceso. A mis amigos, los que conocí en el IESTA, y a los que yo ya traía. A nuestros seres amados, que ya no nos acompañan, pero están presentes también en estas instancias. Este proyecto no podría haberse llevado adelante sin su apoyo. Infinitas gracias.

Cecilia:

A las hermosas personas que me rodearon en este proceso y que adoro: a Martín, por apoyarme constantemente y tener siempre palabras de aliento y perspectiva; a mis amigas, a quienes aburrí con este proyecto en varias charlas de los últimos dos años; a mis padres, por la vida y apoyo incondicional; a mi abuela, por su sabiduría infinita; a mi hermano y hermana, por el cariño; a mis primos/as y tíos/as, en particular a Álvaro por sus valiosos aportes sobre el tema. Sin ellos/as probablemente nada de lo lindo que vivo y siento sería realidad. Gracias por el aliento en el proceso de la carrera y de este trabajo final.

*“Durante toda mi vida,
las nuevas visiones de la naturaleza
me hicieron regocijarme como una niña.”*

Marie Curie, 1867-1934.

Resumen

En este trabajo se realizan predicciones de área basal en una zona de la cuenca Gondwánica de Uruguay, mediante la utilización de modelos geoestadísticos, a partir de información de campo relevada en el inventario forestal nacional, imágenes satelitales Landsat 8 e información de otras fuentes vinculadas (índice coneat, precipitaciones, temperatura, evapotranspiración, índices de vegetación, distancia al curso de agua, entre otros). La metodología utilizada consiste en la aplicación de interpolación por kriging (ordinario, universal y con deriva externa).

La relevancia de los bosques nativos en Uruguay refuerza la importancia de investigar sobre métodos robustos que consideren información auxiliar en el proceso de estimación de variables de interés, como el área basal. Los elevados costos de relevamiento de campo incentivan la aplicación de herramientas para la realización de estimaciones a partir de datos disponibles como los espectrales, a fin de tener una aproximación en los lugares no muestreados.

Una de las regiones con baja cobertura de información en el inventario forestal nacional es la cuenca alta del Río Negro, la cual es particularmente relevante por ser el lugar donde se instala la segunda planta de UPM. En este contexto, en el presente trabajo se generan estimaciones de área basal para esa zona, aportando información que sirva de contralor del manejo del área.

El análisis revela que el método de kriging con deriva externa es el que resulta más preciso para estimar el área basal, presentando un *RMSE* más pequeño, de 11.87, en comparación con 12.64 (Kriging Ordinario), 12.68 (Kriging Universal) y 12.25 (Modelo Lineal de Regresión).

Las variables espectrales que más contribuyeron al modelo estimado de área basal fueron las bandas 2, 3, 5 y 6 de Landsat, esto puede deberse a que el diámetro de los árboles a la altura del pecho y por lo tanto el área basal, están directamente relacionados con el tamaño de las copas de los árboles que determina la reflectancia de la misma, la cual es captada por las bandas satelitales.

También quedaron seleccionadas en el modelo las variables elevación y pendiente, la distancia al curso de agua y las precipitaciones. Se destaca que dichas variables son determinantes del tipo de bosque que se desarrolla en un lugar particular, por lo que una posible interpretación es que la variabilidad del área basal se ve explicada por el tipo de bosque (cada tipología tiene características particulares en términos de las variables mencionadas). Es de esperarse que el comportamiento del área basal difiera según el tipo de bosque, lo que le da coherencia a los resultados de las variables seleccionadas por el modelo.

Este trabajo es el primero en realizar aplicaciones de interpolación por kriging para la predicción de área basal en Uruguay, abriendo una línea de investigación de aplicación empírica de métodos que puedan ser de utilidad para la mejora de la gestión de bosques nativos.

Palabras Clave: Área basal, inventario forestal, predicción espacial, teledetección, Landsat 8, kriging, Uruguay.

Tabla de Contenidos

Lista de Siglas	I
Índice de Figuras y Cuadros	II
1. Introducción	1
2. Antecedentes	4
3. Marco Teórico	8
3.1. Predicción Espacial	8
3.2. Variogramas y Modelización	12
3.3. Kriging	21
4. Metodología	28
4.1. Área de estudio	29
4.2. Datos utilizados	34
4.3. Unión de las capas de información a utilizar	41
5. Resultados	43
5.1. Análisis exploratorio de los datos construidos	43
5.1.1. Correlogramas	48
5.1.2. Selección de variables	49
5.2. Variogramas	52
5.2.1. Variograma Omnidireccional	52
5.2.2. Variograma Direccional	55
5.3. Kriging	56
5.3.1. Kriging Ordinario	56
5.3.2. Kriging Universal	59
5.3.3. Kriging Universal con Deriva Externa	62
5.3.4. Diagnóstico mediante validación cruzada y comparación de resultados	66
6. Conclusiones	71
7. Líneas futuras de investigación	72
Referencias	74
Bibliografía	78
8. Apéndice	80
8.1. Librerías utilizadas	80
8.2. Shiny App y GitHub	81
8.3. Análisis exploratorio de los datos (complementario)	81

Siglas

- AA** Aprendizaje Automático. 6, 73
- AB** Área Basal. 1–5, 18, 25, 26, 28–30, 34, 35, 41, 45, 46, 48–50, 53, 55, 67, 71, 73
- AIC** Criterio de Información Akaike. 50
- ANN** Redes Neuronales Artificiales. 4
- BN** Bosque Nativo. 1, 6
- CRS** Sistema de coordenadas de referencia. 41
- DAP** Diámetro a la Altura del Pecho. 1
- DGF** Dirección General de Forestación. 29–35, 71
- ETP** Evapotranspiración. 35
- EVI** Índice de Vegetación Mejorado. 34
- GI** Índice de Verdor. 34
- GPS** Sistema de posicionamiento global. 3
- IFN** Inventario Forestal Nacional. 2–6, 28–35, 38, 41–45, 58, 71, 82
- KED** Kriging con deriva externa. 24, 28, 41, 54
- LOOCV** Validación Cruzada dejando uno afuera. 66
- MAE** Error Absoluto Medio. 26, 67
- ME** Error Medio. 26, 67
- MGAP** Ministerio de Ganadería Agricultura y Pesca. 3, 40
- MLR** Modelo Lineal de Regresión. 4–6, 41
- MSE** Error Cuadrático Medio. 26, 67
- NA** Dato faltante. 45
- NDVI** Índice de Vegetación de Diferencia Normalizada. 34, 40
- OK** Kriging Ordinario. 4, 28
- RF** Random Forest. 4, 6
- RMSE** Raíz del Error cuadrático medio. 5, 26, 67
- RVI** Índice de Vegetación. 34, 40
- SIG** Sistema de Información Geográfica. 3, 6, 42
- SVM** Support Vector Machine. 4, 73
- TPI** Índice de Posición Topográfica. 34, 35
- UK** Kriging Universal. 28
- VIF** Factor de Inflación de Varianza. 48

Índice de figuras

1.	Ejemplos de los distintos tipos de Bosques.	2
2.	Interpolación estacionaria (izquierda) y no estacionaria (derecha).	10
3.	Ejemplo de semivariograma experimental y modelo ajustado (con parámetros representados).	12
4.	Comportamiento del variograma cerca del origen: (a) intercepto positivo (pepita); (b) enfoque lineal, no diferenciable; (c) continua y diferenciable.	17
5.	Principales Modelos de Variogramas	19
6.	Ecozonas por unidades morfoestructurales	30
7.	Bosque Nativo en la Cuenca Gondwánica, selección de subcuencas y puntos del IFN	31
8.	Cuencas seleccionadas por DGF (Nivel 4) para la predicción de AB	32
9.	Bosque Nativo, puntos del IFN y área de interés de predicción	33
10.	Modelo Digital de Terreno y Bosque Nativo	36
11.	Mosaicos de Landsat utilizados	38
12.	Raster Landsat Construido	39
13.	Área basal en la Cuenca Gondwánica	45
14.	Área basal en el área de predicción	46
15.	Histograma y Boxplot del Área Basal	47
16.	Clasificación de regiones a partir del AB	47
17.	Correlograma	49
18.	Nube variográfica del AB (Cuenca Gondwánica, distancia en metros).	52
19.	Variograma experimental y modelo estimado (utilizando únicamente el AB).	53
20.	Variograma experimental y modelo estimado (incorporando las coordenadas).	54
21.	Variograma experimental y modelo estimado (incorporando variables seleccionadas).	55
22.	Variograma Direccional (Cuenca Gondwánica, distancia en metros).	56
23.	Predicción de área basal con Kriging Ordinario	57
24.	Desvío estándar del Kriging Ordinario	58
25.	Predicción de AB con OK y variable relevada (círculos de tamaño proporcional al AB en campo)	59
26.	Predicción de área basal con Kriging Universal	60
27.	Desvío estándar del Kriging Universal	61
28.	Predicción de AB con UK y variable relevada (círculos de tamaño proporcional al AB en campo)	62
29.	Predicción de área basal con Kriging Universal con Deriva Externa	64
30.	Desvío estándar del Kriging Universal con Deriva Externa	65
31.	Predicción de AB con UK y variable relevada (círculos de tamaño proporcional al AB en campo)	66
32.	Comparación de las predicciones con el área basal observada	68
33.	Densidad de Kernel de los residuos de los modelos	69
34.	Comparación de los residuos de los modelos	69
35.	Variograma de los residuos del Kriging con Deriva Externa	70
36.	Boxplot del Área Basal según año de relevamiento	81
37.	Año de relevamiento de los puntos del IFN.	82
38.	Altitud, pendiente, elevación y TPI (extraído del Modelo Digital de Terreno)	83

Índice de cuadros

1.	Descripción de las bandas de las imágenes satelitales Landsat 8	37
2.	Índices de vegetación	40
3.	Medidas de resumen (Puntos del IFN en la Cuenca Gondwánica)	44
4.	MLR para predecir el AB en la Cuenca Gondwánica.	51
5.	Medidas de resumen de la predicción por Kriging Ordinario	57
6.	Medidas de resumen de la predicción por Kriging Universal	60
7.	Medidas de resumen de la predicción por KED	63
8.	Resultados de Validación Cruzada	67
9.	Comparación de predicciones en los modelos	68
10.	Comparación de Desvíos Estándar obtenidos con los modelos	68

1. Introducción

El bosque nativo se caracteriza por ser una fuente de diversidad biológica. Es un gran generador de biomasa, un sumidero importante de carbono (relevante para mitigar el aumento del efecto invernadero), mejora la calidad del agua al retener sedimentos y nutrientes arrastrados desde las áreas altas de la cuenca, regula el régimen hídrico previniendo inundaciones y mitigando sequías, sirve como amortiguador ante eventos extremos de temporales o lluvias intensas, minimizando el potencial erosivo sobre playas en territorios costeros (en zonas litorales de ríos, arroyos y lagunas, logra amortiguar durante las fases de crecidas de estos haciendo de filtro, a la vez de que disminuye el ingreso de contaminantes a los cursos de agua), combate la erosión de suelo, siendo a la vez una atracción turística (Brazeiro, 2018).

Resulta esencial contar con información referente a la presencia y magnitud de árboles en los bosques nativos, para la toma de decisiones y el diseño de estrategias de protección de dichos espacios. El Área Basal (AB) es una variable de referencia en el manejo forestal, es calculada como el área transversal de un árbol medida a la altura del pecho (aprox 1,4 mts del nivel del suelo), y generalmente se expresa en unidades cuadradas por unidad de área (en el presente trabajo se expresa en metros cuadrados por hectárea). Se parte de medir el Diámetro a la Altura del Pecho (DAP) de cada árbol (por lo general en metros), y el Área Basal (AB) se obtiene como: $AB = \pi\left(\frac{DAP}{2}\right)^2$.

La superficie de Bosque Nativo (BN) en Uruguay es de aproximadamente 835 mil hectáreas¹, de acuerdo a la Cartografía de Bosque Nativo 2016², representando aproximadamente un 5 % de la totalidad del territorio nacional. A continuación se describen los diferentes tipos de bosque nativo de Uruguay, según la Estrategia Nacional de Bosque Nativo (DGF, 2018), posteriormente en la Figura 1 se presenta un ejemplo de cada uno:

- **Bosques de galería, ribereños o fluviales:** Es el tipo de bosque nativo más abundante de Uruguay, que acompaña ríos y arroyos. Su presencia y desarrollo depende fundamentalmente del medio físico, que son el caudal y los suelos aluviales. Se caracteriza por presentar una alta cantidad de rebrotes por cepas y pocos árboles monopodiales; a causa de haber sufrido intervenciones humanas de corte para leña en diferentes grados de tala y alteración.
- **Bosque serrano:** Está asociado a las topografías más altas del país, abarcando diferentes sectores del sistema de cuchillas y serranías. Generalmente se desarrolla sobre suelos superficiales y está integrado por especies que soportan períodos de escasez hídrica.
- **Bosque de parque:** Se localiza en zonas próximas al litoral del río Uruguay, como transición entre el bosque fluvial y las comunidades herbáceas. Se trata de asociaciones xerófitas con una baja densidad de individuos.

¹El dato referente a la superficie boscosa del país es aproximado debido a que diferentes estudios de clasificación y mapeo de bosques con imágenes satelitales han generado resultados distintos. Las diferencias pueden deberse a las distintas metodologías usadas (tipo de imagen, escalas de análisis, chequeo en campo, etc.).

²Acceder a datos abiertos: Link:<https://catalogodatos.gub.uy/dataset/mgap-bosques-forestacion-medio-ambiente>

- **Bosque de quebrada:** El bosque de quebrada se desarrolla en zonas húmedas de drásticas condiciones topográficas (laderas abruptas) en el norte y noreste del país. Presenta características subtropicales, es decir, una alta densidad de especies de elevado porte.
- **Palmares:** Son reconocidos por su valor paisajístico, cultural y de biodiversidad. Existen dos comunidades, una en el este del país donde se producen las mayores concentraciones, y otra en el litoral oeste (Butia, Odorata y Yatay).
- **Bosque costero (psamófilo):** El bosque costero se caracteriza por sus formaciones vegetales leñosas ubicadas en dunas de la costa platense y atlántica a orillas del Río de la Plata y costa del océano Atlántico.

Figura 1: Ejemplos de los distintos tipos de Bosques.



Fuente: (DGF, 2018).

El Inventario Forestal Nacional (IFN) aporta información sobre las características y estado de los diferentes bosques, resultando útil para el manejo forestal sostenible de los recursos boscosos y para planificar, diseñar e implementar políticas. El último fue llevado a cabo en el año 2014, cuenta con información referente a 1500 puntos distribuidos en todo el país. No obstante, varias zonas quedaron sin cobertura debido a diversas dificultades de relevamiento de campo, por lo que resulta de interés avanzar en la identificación y aplicación de herramientas para la predicción espacial de variables como el AB en dichas zonas, a fin de tener una aproximación en los lugares no muestreados.

Una de las regiones con baja cobertura de información es la Cuenca Alta del Río Negro. El proyecto “Iniciativa del Río Negro”, llevado a cabo por el Ministerio de Ambiente y el

MGAP, pretende generar una línea base del estado de situación de la cuenca del Río Negro antes de la puesta en funcionamiento de la segunda planta de UPM. En ese sentido, contar con la información de área basal para esa zona, siendo que los datos del Inventario Forestal Nacional (IFN) no la cubrieron en su totalidad, es esencial para conocer el posible volumen de bosque nativo de esa zona, y a futuro servir como contralor del manejo de esa área.

La Teoría de Variables Regionalizadas, conocida también como Geoestadística, permite incorporar la correlación espacial en el análisis de los datos (es decir, el nivel de similitud existente entre los que se encuentran cerca en el espacio), esto permite abordar de manera más eficaz muchos problemas referidos a información regionalizada. Se originó en el campo de estudio de la minería, siguiendo el objetivo de predecir valores de variables en lugares no muestreados, teniendo como antecedentes los trabajos de Sichel (1947; 1949) y Krige (1951) en la década de 1950, y siendo posteriormente desarrollada por Georges Matheron, quien realizó una formalización en un lenguaje más riguroso en la década de 1960 (Giraldo, 2002). Desde entonces, la Geoestadística se ha aplicado a muchos otros campos relacionados con las ciencias de la tierra, por ejemplo, Hidrogeología, Hidrología, Meteorología, Oceanografía, Geoquímica, Geografía, Ecología del paisaje, Ciencias del suelo y Silvicultura.

En las últimas décadas hubo grandes avances en los SIG, GPS, y en métodos computacionales de teledetección, que han proporcionado nuevas herramientas y técnicas para generar y procesar información regionalizada. Esto ha generado grandes aportes particularmente al ámbito de la gestión forestal. Entre ellos, se encuentra la estimación de información en áreas no muestreadas en el Inventario Forestal Nacional (IFN), combinando la teledetección junto con los datos de campo (Morgenroth y Visser, 2013; Watt MS., 2016; Boisvenue C., 2016; Fayada y cols., 2016; Vicharnakorn P., 2014).

Los satélites Landsat, han adquirido continuamente imágenes espaciales de la superficie terrestre de todo el mundo en los últimos 50 años, proporcionando datos valiosos para la investigación del uso de la tierra así como también de los cambios naturales e inducidos por el hombre en el panorama mundial. No obstante, particularmente la detección remota del AB es compleja debido al tamaño, la forma y las propiedades dieléctricas de sus elementos dispersos como hojas, ramas y tallos; la diversidad espacial de esta variable hace que la relación entre el relevamiento de campo y los datos de teledetección sea un gran desafío (dos Reis y cols., 2018).

En este contexto, el **objetivo general** del presente trabajo es predecir el AB en la Cuenca Alta del Río Negro, a partir de información de campo relevada (Inventario Forestal Nacional (IFN)), imágenes satelitales (Landsat) e información de otras fuentes vinculada con el uso y características del suelo (Índice coneat, precipitaciones, temperatura, evapotranspiración, índices de vegetación, distancia al curso de agua, modelo digital de terreno, entre otros), generando así una línea de base para el AB, que luego pueda servir como contralor para el manejo del área.

Por otra parte, la presente investigación tiene como **objetivos específicos** analizar el comportamiento espacial del AB y la correlación con las variables obtenidas por teledetección y con variables ambientales.

El documento se estructura de la siguiente forma: En la sección 2 se presentan los antecedentes. En la sección 3 se describe el marco teórico en el cual se basa el presente trabajo final de grado. En la sección 4 se desarrolla la metodología y en la 5 se reportan los resultados junto con algunos ejercicios de validación cruzada. Por último, en la sección 6 se presentan las conclusiones que se derivan del presente trabajo, para posteriormente establecer posibles líneas futuras de investigación en la Sección 7. Adicionalmente se adjunta una sección de Apéndice donde, entre otras cosas, se puede acceder a una aplicación Shiny³ desarrollada en el marco del presente trabajo, así como también al código implementado en R (publicado en GitHub https://github.com/Cecilia-Alonso/Tesis_Final).

2. Antecedentes

Dado que las mediciones de campo de biomasa forestal son costosas, existe un gran interés en obtener estimaciones fiables sobre grandes áreas a partir de datos de teledetección, esto se ha visto reflejado en un amplio volumen de investigaciones académicas vinculadas con dicho objetivo. Dichas estimaciones se obtienen con una variedad de fuentes de datos, métodos estadísticos y estándares de predicción, y no hay acuerdo sobre cuáles son las mejores prácticas para esta tarea.

Múltiples investigaciones han tenido por objetivo la predicción espacial del AB en una zona en particular (dos Reis y cols., 2018; Zahriban M., 2019; X. Zhang, Lei, y Cao, 2010; Gebreslasie M., 2010; Freeman E., 2014; Meng Q., 2009), a partir de distintas metodologías. Para mejorar la comprensión de los diferentes métodos utilizados para realizar la predicción, así como también de los distintos datos existentes, primero se lleva a cabo una revisión de la literatura disponible.

Se destaca que no se encontraron antecedentes nacionales de predicción espacial de variables relacionadas a la biomasa, como el AB. Por lo que gran parte de la importancia que tiene la presente investigación radica en la escasa producción académica nacional en esta temática.

Uno de los principales antecedentes del presente trabajo fue el de Reis (2018). Este documento tiene por objetivo estimar y mapear el AB y el volumen de rodales de eucalipto mediante la integración del IFN, la teledetección (imágenes Landsat 5 con las que calculan índices de vegetación), métodos paramétricos y no paramétricos de predicción espacial, Modelo Lineal de Regresión (MLR), Random Forest (RF), Kriging Ordinario (OK), Support Vector Machine (SVM) y Redes Neuronales Artificiales (ANN), para la zona de Río Grande Brasil. Una diferencia importante con el presente trabajo es que dicho documento se centra en plantaciones forestales, en lugar de bosque nativo. Entre los resultados obtenidos se encuentra que un índice de vegetación que se basa en las bandas 4 y 5 fue la variable espectral que presentó mejores valores de correlación con el AB ($r = -0.91$) y volumen ($r = -0.52$) y también fue la variable que más contribuyó a las estimaciones de volumen y AB por el MLR y RF. El algoritmo de RF presentó errores de volumen y AB más pequeños en comparación con otros algoritmos de

³El link al visualizador es el siguiente: https://cinve.shinyapps.io/Area_Basal_Resultados/

machine learning y MLR (dos Reis y cols., 2018).

Galeana et al (2014), llevaron a cabo una investigación en Ciudad de México, en donde se exploró la capacidad de las variables de teledetección (índices de vegetación, principalmente) para modelar la distribución espacial de la biomasa aérea. Los resultados mostraron que un método geoespacial híbrido, basado en datos y un interpolador de regresión-kriging, tiene un buen potencial para estimar la biomasa de los suelos.

Fassnacht et al (2014) estiman el AB para regiones de bosque nativo de Alemania y Chile, mediante la integración del IFN, la teledetección, métodos paramétricos y no paramétricos de predicción espacial. Destacan la importancia del tamaño de muestra en los procesos de predicción espacial remota del AB y concluyen que el tipo de datos de predicción es el factor más importante para las estimaciones de este tipo de variable, siendo LIDAR (Laser de detección y rango de imágenes) preferible a otro tipo de imágenes. A su vez, remarcan que a la fecha (2014) se utilizaban diferentes métodos para la estimación de biomasa y AB, sin un acuerdo general sobre las mejores prácticas, advirtiéndose diferencias sustanciales en la precisión entre los distintos métodos.

Zhang et al (2014) estimó el AB presente en bosques de Alberta Canadá. Integró un conjunto de datos espaciales, una combinación de datos del IFN de 1968 e imágenes satelitales (LIDAR). En esta investigación se mostró que la información de datos de suelo, con los datos de inventario y las imágenes resultaron ser eficientes para estimar cantidad y distribución de la biomasa forestal en grandes regiones.

Por otro lado, Zahriban et al (2019) buscaron evaluar las mejoras de las estimaciones de volumen y AB usando datos satelitales (imágenes multiespectrales del satélite Pléyades 2013), complementados con datos auxiliares (precipitación, temperatura, pendiente y la elevación), para el área de estudio del bosque de Sari en Darabkola, Irán. Encuentran que integrando los datos espectrales con datos auxiliares mejoró los resultados del modelado de volumen (disminución del 5 % en RMSE), y del AB (disminución del 2,5 % en RMSE).

En la investigación llevada a cabo al noreste de México por López et al (2016), se utilizaron imágenes Landsat 5 para estimar el AB de bosque nativo. De las distintas bandas, la banda infrarrojo cercano es la que resultó más relevante para la predicción.

Fayada et al (2016) mapea la biomasa de bosques (principalmente tropicales) de la Guyana Francesa, utilizando métodos de kriging. Para esto combina imágenes satelitales, inventarios forestales, e información ambiental. En esta investigación se obtiene como resultados que la incorporación de una mayor variedad de datos, logra mejorar el ajuste y desempeño predictivo de los modelos, principalmente para valores altos de AB.

Por otra parte, se encuentran algunos antecedentes que buscan predecir variables forestales a partir de métodos diferentes a Kriging. En España, García et al (2015) procedieron a realizar una comparación entre distintas técnicas de predicción de variables forestales a partir de un

vuelo de teledetección LIDAR, en particular la metodología AA y el clásico MLR. El primero resultó ser la mejor de las técnicas implementadas.

Latifi et al. (2010) busca predecir y mapear el volumen total y la biomasa en un bosque templado mixto en el suroeste de Alemania, para lo que empleó métodos no paramétricos de predicción, y tres fuentes de datos de detección remota con las que generan 113 variables. En cuanto a los resultados, encuentran que el conjunto de datos es relevante en la predicción de ambas variables (volumen total y biomasa), el RF demostró ser superior a los otros métodos examinados como NN (de su acrónimo en inglés Near Neighbour, o vecino más cercano). El mapa con las predicciones finales lo realizan con RF.

Boisvenue et al (2016) busca estimar el crecimiento forestal tanto de manera espacial como temporal, describiendo para ello el cambio en las áreas de biomasa con la información derivada de la serie temporal Landsat (1984-2012) en una gran área de Saskatchewan, Canadá. Encuentran que el modelo de RF explicó el 70 % de la variación en los datos, utilizando como insumo las bandas 1-5 y 7, datos de IFN de elevación y pendiente.

Existen múltiples documentos de trabajo que han utilizado imágenes Landsat para generar predicciones espaciales de variables vinculadas a lo forestal (Boisvenue C., 2016; Vicharnakorn P., 2014; dos Reis y cols., 2018; Meng Q., 2009; Galeana, López, López, Silván, y Couturier, 2014a; López Serrano, Corral, Díaz Varela, Álvarez, y López, 2016; Cutler, Boyd, Foody, y Vetrivel, 2012; Ahmed, Franklin, Wulder, y White, 2015; Stojanova, Panov, Gjorgjioski, Kobler, y Džeroski, 2010; Gao y cols., 2016; Mutanga, 2015).

A su vez, también existen antecedentes nacionales en el uso de imágenes Landsat. Altamirino et al (1996) llevaron a cabo un estudio de la Cuenca Alta del Río Negro utilizando imágenes satelitales y SIG, con el objetivo de diagnosticar el uso de los recursos suelo y agua principalmente bajo el cultivo de arroz, así como evaluar las posibilidades de expansión del mismo. A partir de la información generada se calculó el área de posible expansión del cultivo, y se detectaron y cuantificaron áreas sensibles desde el punto de vista de la sustentabilidad.

Posteriormente, Petraglia et al (2006) trabajaron en la actualización de la carta forestal digital del Uruguay, mediante el uso de imágenes satelitales Landsat 5 del año 2004. Detectaron en el país cerca de 1.400.000 has de forestación de las cuales el 53 % es BN, mientras que el 40 % corresponde a plantaciones industriales y el resto a montes de abrigo y sombra, bosques costeros y parques. En cuanto a la distribución territorial, la mayor parte de la forestación se encuentra en Rivera y Tacuarembó.

Almeida (2019), en una investigación llevada a cabo en la Amazonia Brasil, obtuvo como resultado que los satélites Landsat 8 y Sentinel-2 mostraron el mismo rendimiento en términos de precisión (con precisiones ajustadas por área de 95,7 % y 96,7 % respectivamente). Sin embargo, Landsat 8 mapeó un 36,9 % más de área de tala selectiva en comparación con Sentinel-2. La infraestructura de tala se detectó mejor a partir de Sentinel-2 (43,2 %) que los datos de Landsat 8 (35,5 %), lo que confirma su potencial para mapear la tala a pequeña escala. También

se evaluaron el área impactada mediante la tala selectiva con una cuadrícula regular de 300m 300m sobre los resultados basados en píxeles, lo que resultó en 1143 ha de bosque alterado a partir de Sentinel-2 y 1197 ha con Landsat 8.

Por otra parte, en otro estudio llevado a cabo en Argentina por Avogadro et al (2019), que tenía por objetivo la comparación de la información aportada por imágenes satelitales de Landsat8 y Sentinel 2, obtuvieron como resultado que ambos sensores son comparables con un nivel de correlación de alrededor 0,90. El satélite Sentinel 2 permite un mayor nivel de detalle con su resolución espacial mejorando el análisis intra-urbano.

Por lo expuesto anteriormente, en línea con la selección del tipo de satélite de la mayoría de los antecedentes (Boisvenue C., 2016; dos Reis y cols., 2018; Cutler y cols., 2012; Galeana, López, López, Silván, y Couturier, 2014b), se decide trabajar con imágenes LANDSAT.

3. Marco Teórico

3.1. Predicción Espacial

La predicción espacial no solo es valiosa para los científicos que intentan modelar procesos espaciales, sino también para los responsables de la formulación de políticas que necesitan planificar, evaluar el riesgo y gestionar los resultados de los procesos espaciales en diferentes escalas (Harris, Charlton, y Fotheringham, 2010).

Cuando el objetivo de la investigación es realizar la predicción de una variable $Z(x)$ (regionalizada), en un sitio no muestreado a partir de la información incluida en puntos muestrales de sitios vecinos, puede utilizarse el predictor lineal kriging (Cressie, 1993). Este tiene en cuenta la dependencia espacial (estimada a través de una función de semivarianza) para realizar la estimación del efecto de cada variable en la predicción. Las variables $Z(x_i)$, es decir, $Z(x_1), Z(x_2), \dots, Z(x_n)$, son el valor del vector de variables observado en el sitio x_i para $i = 1, \dots, n$.

Para realizar una predicción espacial se pueden identificar dos etapas. Inicialmente se procede a hacer un análisis estructural, en el cual se lleva a cabo una descripción de la correlación entre los puntos del espacio. Posteriormente, en la segunda etapa se realiza la predicción en sitios de la región no muestreados por medio de la técnica de Kriging. Esta técnica calcula el promedio ponderado de las observaciones realizadas en la muestra, considerando la estructura de correlación espacial desarrollada en la primera etapa a la hora de asignar los pesos a los valores muestrales.

Variable regionalizada⁴.

Una variable regionalizada, que se representa como $Z(x)$, puede ser definida como un campo aleatorio cuyo dominio se halla incluido en un espacio euclídeo de dimensión “ d ”, o sea R^d . En el caso de $d = 2$, $Z(x)$ será una variable medida, asociada a un punto x del plano.

De llevarse a cabo las mediciones sobre una superficie, $Z(x)$ se asocia a una variable aleatoria, donde x representa las coordenadas del punto, las cuales pueden estar expresadas en términos planos o geográficos y Z representa la variable en cada una de estas coordenadas.

Momentos de una variable regionalizada

En la variable regionalizada definida por el campo aleatorio $Z(x)$ con $Z(x): x \in D \subset R^d$ pueden ser establecidos los siguientes momentos univariados y bivariados:

⁴Basado en Giraldo (2002), capítulos 2 y 3.

- $E(Z(x_i)) = m(x_i)$
- $V(Z(x_i)) = E[Z(x_i) - m(x_i)]^2 = \sigma_i^2$
- $C(Z(x_i), Z(x_j)) = E[Z(x_i) - m(x_i)][Z(x_j) - m(x_j)]$
- $\gamma(Z(x_i), Z(x_j)) = (\frac{1}{2})E[Z(x_i) - Z(x_j)]^2$ (función de semivarianza)

Estacionariedad

La variable regionalizada es considerada como estacionaria si su función de distribución conjunta es invariante respecto a cualquier traslación del vector h (tamaño del salto) lo que equivale a decir que la función de la distribución del vector aleatorio $\vec{Z} = [Z(x_1), Z(x_2), \dots, Z(x_n)]^t$ es idéntica a la función de distribución del vector $Z(x + \vec{h}) = [Z(x_1 + h), Z(x_2 + h), \dots, Z(x_n + h)]^t$ para cualquier valor de h .

La estacionariedad puede ser clasificada en términos de los momentos de la variable, definidos anteriormente.

Estacionariedad de segundo orden

La variable regionalizada $Z(x)$ con $Z(x): x \in D \subset R^d$ se dice que es estacionaria de segundo orden si cumple las siguientes dos condiciones:

- El valor esperado de la variable aleatoria es finito y constante para todo punto del dominio, lo que implica que $E(Z(x)) = k$, $k \in R$, $k < \infty$, $\forall x \in D \subset R^d$
- Para toda pareja de puntos $(Z(x), Z(x + h))$ separados una distancia h , la covarianza existe y es una función que depende únicamente de la separación h , lo que implica que $cov(Z(x), Z(x + h)) = c(h) < \infty$

Estacionariedad débil

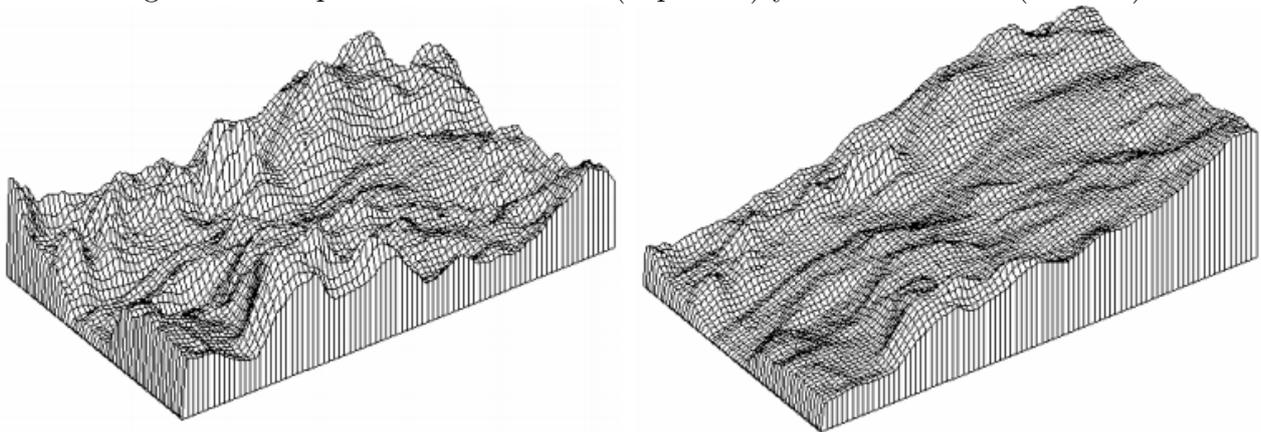
En situaciones donde no se cumpla una de las dos condiciones señaladas de la estacionariedad, más específicamente cuando se analizan fenómenos en los que la varianza no es finita, se trabaja con la hipótesis de que los incrementos $[Z(x + h) - Z(x)]$ sean estacionarios. Para ello deberá cumplirse:

- Si $Z(x)$ tiene esperanza finita y constante para todo punto del dominio, la esperanza de los incrementos debe valer cero, lo que implica que $E[Z(x + h) - Z(x)] = 0$
- Para cualquier vector h la varianza de los incrementos está definida y es una función que depende de h , lo que implica que $V[Z(x + h) - Z(x)] = E[Z(x + h) - Z(x)]^2 = 2\gamma(h)$

Es claro que si una variable regionalizada presenta estacionariedad fuerte, entonces también será estacionaria débil y en ambos casos se puede considerar, por lo que la correlación entre los datos no depende de la dirección en la que esta sea calculada y en este caso decimos que el fenómeno considerado es isotrópico.

Cuando la esperanza de la variable considerada no sea la misma en todas las direcciones, la estacionariedad no estará presente y este tipo de fenómenos son considerados como anisotropía. La isotropía queda manifiesta, si existe, en las funciones de autocovarianza en varias direcciones que de presentar formas significativamente distintas, el supuesto de que la esperanza de la variable regionalizada es constante, no es válido, correspondiendo el tratamiento de la variable como no estacionaria. En la figura que sigue, se muestra como es el comportamiento de una variable que no es estacionaria.

Figura 2: Interpolación estacionaria (izquierda) y no estacionaria (derecha).



Fuente: Giraldo, R. (2002) Introducción a la Geoestadística. Teoría y Aplicación.

Correlación Espacial

Para la determinación de la dependencia espacial entre los datos de una variable se usan las funciones de covarianza, correlograma y semivariograma, todas en función de h .

La covarianza es la función que describe la dispersión muestral entre pares de observaciones que se encuentran separadas una distancia h .

$$C(h) = COV(Z(x+h), Z(x)) = \frac{\sum_{i=1}^n (Z(x_i+h) - m)(Z(x_i) - m)}{n} \quad (1)$$

Donde:

- $C(h)$ es el estimador de la covarianza.

- m es el valor promedio de la variable de interés en la región en estudio.
- n es el número de parejas de puntos que se encuentra separadas una distancia h .

El correlograma muestral se obtiene asumiendo que el fenómeno en estudio es estacionario y estimando la varianza de la variable regionalizada a través de la varianza de la muestra. Dicho correlograma representa el cociente entre la variabilidad de parejas de puntos que están separados una distancia h , en relación a la varianza regionalizada estimada a través de la varianza de la muestra.

$$\hat{r}(h) = \frac{c(h)}{c(0)} \quad (2)$$

Asumiendo que la varianza de los incrementos de la variable regionalizada es finita, es decir, que estamos en un contexto de estacionariedad débil, se le denomina variograma a $2\gamma(h)$, función de h , por lo que a $\gamma(h)$ se le denomina semivariograma. Este último caracteriza la dependencia espacial del proceso.

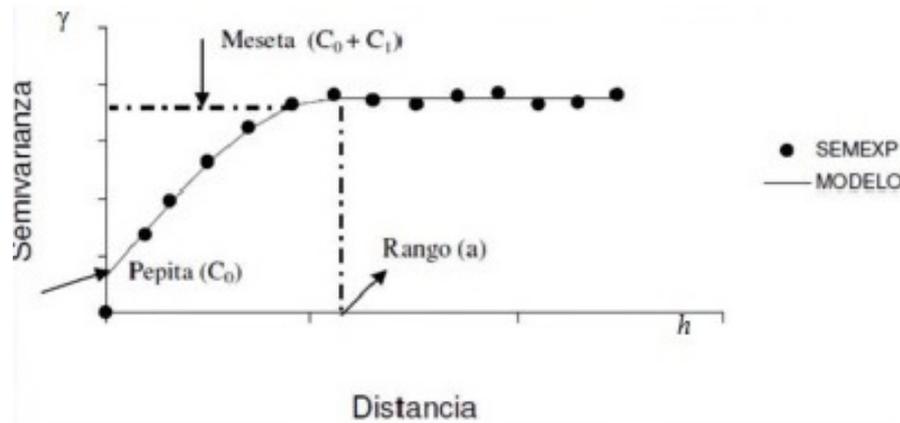
$$\hat{\gamma}(h) = \frac{\sum_{i=1}^n (Z(x_i + h) - Z(x_i))^2}{2n} \quad (3)$$

Donde:

- $Z(x_i)$ es el valor de la variable regionalizada de interés en el punto x_i .
- $Z(x_i + h)$ es otro valor de la muestra separado del punto anterior por una distancia h .
- n es el número de parejas de puntos que se hallan a una distancia de h .

Los modelos de semivariograma estarán definidos por un conjunto de parámetros que se pueden observar en la siguiente figura.

Figura 3: Ejemplo de semivariograma experimental y modelo ajustado (con parámetros representados).



Fuente: Giraldo, R. (2002) Introducción a la Geoestadística. Teoría y Aplicación.
 Nota: *SEMEXP* refiere al Semivariograma Experimental y *MODELO* es el modelo ajustado.

Donde:

- C_0 : representa la varianza a distancia 0, que se la conoce como *nugget* o *pepita* (discontinuidad en el origen).
- C_1 : representa la variabilidad espacial, que se la conoce como *partial sill* o *meseta* parcial. La varianza de los datos da una idea de por donde puede estar la meseta del variograma (no siempre coinciden).
- a : representa el *rango*, es la distancia a la cual el variograma se estabiliza, es decir, distancia a partir de la cual no hay correlación, y los datos ya no presentan relación espacial (a partir del punto donde la semivarianza se mantiene constante).
- C_0+C_1 : representa la máxima variabilidad espacial, que se la conoce como *sill* o *meseta*. Es el valor constante que toma el variograma en distancias mayores al rango.

Con la varianza de los datos (S) y el efecto pepita (C_0) se puede hallar la meseta parcial o contribución: $C_1 = S - C_0$. De esta forma, C_1/S da cuenta del aporte de la componente espacial que cuanto más elevado sea, más apropiado será el uso de metodologías que incorporen la componente espacial (como lo hace la metodología *kriging*).

3.2. Variogramas y Modelización

En geoestadística, la autocorrelación espacial ha sido tradicionalmente modelada por un variograma que describe el grado en el que las observaciones cercanas tienen valores similares (Cressie, 1993; Hengl, 2009). El variograma experimental es la estimación del variograma teórico y se construye a partir de los datos muestrales.

Una nube variográfica es un diagrama de dispersión de pares de datos, en el que se grafica la semivarianza en función de la distancia entre los puntos. La semivarianza $\gamma(h)$ se define formalmente como la diferencia al cuadrado en el AB entre las ubicaciones consideradas. Bajo el supuesto de que el proceso aleatorio $Z(x)$ tiene media constante (proceso estacionario), se obtiene por el método de los momentos el siguiente estimador del variograma:

$$2\hat{\gamma}(h) = \frac{1}{N(h)} \sum_{|x_i - x_j| = |h|} (z(x_i) - z(x_j))^2 \quad (4)$$

donde

$$N(h) = \{(x_i, x_j) : |x_i - x_j| = h; \quad i, j = 1, \dots, n\}$$

Siendo $N(h)$ el número de pares de coordenadas que distan a una distancia h .

El procedimiento para obtener el variograma experimental es:

1. Escoger una distancia h .
2. Escoger una dirección θ .
3. Calcular $\gamma(h)$ para los valores de $h, 2h, 3h, \dots, nh$.
4. Graficar $\gamma(h)$ versus los valores de $h, 2h, 3h, \dots, nh$.

El variograma se puede utilizar para la predicción espacial, ajustándole un modelo paramétrico (puede ser exponencial, lineal, esférico, gaussiano, entre otros). El análisis del variograma es un elemento clave para definir qué modelo estadístico se va a utilizar para la interpolación de los datos.

En cuanto a los métodos para estimar los parámetros del modelo, puede ser por mínimos cuadrados ordinarios (MCO), mínimos cuadrados ponderados (MCP) o mínimos cuadrados generalizados (MCG). El orden de eficiencia de los mencionados es creciente, siendo el MCG el mejor, mientras que el de simplicidad es decreciente, siendo el MCO el mejor. La estimación por MCP es la más utilizada en la práctica debido su fácil implementación y a ventajas computacionales. No obstante, presenta limitaciones, como por ejemplo, depende de las estimaciones del semivariograma, que son muy sensibles a la selección de las distancias y las regiones de tolerancia utilizadas para su cálculo (Melo Martínez, 2012).

La variación intrínseca y el variograma

Un proceso aleatorio estacionario puede ser representado con el modelo:

$$Z(\mathbf{x}) = \mu + \varepsilon(\mathbf{x}) \quad (5)$$

Esto establece simplemente que el valor de Z en x es la media del proceso, más un componente aleatorio extraído de una distribución con media cero y función covarianza:

$$C(\mathbf{h}) = E[\varepsilon(\mathbf{x})\varepsilon(\mathbf{x} + \mathbf{h})] \quad (6)$$

La mayor dificultad y la preocupación más común acerca de la estacionariedad débil, es que la media parece cambiar en una región y la varianza crece, si no se limita, a medida que aumenta el área de interés. En estas circunstancias, la covarianza no puede definirse.

Matheron (1962) reconoció el problema que esto creaba y su solución fue una importante contribución a la geoestadística práctica. Consideró que, si bien la media podría no ser constante, lo sería al menos para pequeños $|h|$, de forma que las diferencias esperadas sean cero:

$$E[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})] = 0 \quad (7)$$

Además, reemplazó las covarianzas por las varianzas de las diferencias como medidas de relación espacial, que dependen del desfase y no de la posición absoluta, al igual que la covarianza. Esto llevó a que:

$$\text{var}[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})] = E[\{Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})\}^2] = 2\gamma(\mathbf{h}). \quad (8)$$

Las dos ecuaciones anteriores constituyen la hipótesis intrínseca de Matheron. Este paso libera de las limitaciones de la estacionariedad de segundo orden, abriéndose un campo más amplio de aplicación.

La cantidad $\gamma(h)$ se conoce como la semivarianza en paso h . El concepto “semi” se refiere al hecho de que es la mitad de una varianza; en este caso, es la mitad de la varianza de una diferencia. No obstante, en la práctica el semivariograma generalmente se denomina como variograma.

Equivalencia con covarianza

Para procesos estacionarios de segundo orden, el variograma y la covarianza son equivalentes, por lo que se cumple:

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}) \quad (9)$$

Por lo tanto, un gráfico del variograma es simplemente una imagen especular de la función de covarianza sobre una recta o plano paralelo a la abscisa.

También podemos relacionar la semivarianza al coeficiente de autocorrelación:

$$\gamma(\mathbf{h}) = \sigma^2\{1 - \rho(\mathbf{h})\} \quad (10)$$

Si un proceso es intrínseco solo, no hay equivalencia porque la función de covarianza no existe.

Tipos de Variogramas

El *variograma teórico*, puede verse como el promedio de los variogramas de todas las realizaciones posibles del proceso. Es necesario diferenciarlo del *variograma regional* y el *variograma experimental*.

El *variograma regional* es el de la realización particular, en un número finito en la región de interés, no representa necesariamente todo el conjunto. Es el que se podría calcular si se tuviera información completa de la región.

Por otra parte, el *variograma experimental* se calcula a partir de una muestra de datos de la región, es decir, estima el variograma regional para esa realización, también se le llama variograma de muestra. Suele ser el único variograma que se conoce, y cualquier inferencia a partir de él requiere un modelado.

Si la estructura de correlación espacial es la misma en todas las direcciones, esto se denomina isotrópica, la función de covarianza depende solo de la magnitud del vector de retardo, h , y no de la dirección. Sin embargo, puede ocurrir que la autocorrelación muestre una estructura que es diferente para diferentes direcciones, esta propiedad de un proceso espacial se conoce como anisotropía, y se representa a partir de un *variograma direccional*. La anisotropía analiza la variación en la continuidad espacial en relación con la dirección, buscando identificar si existe una dirección predominante. Para ello se construyen variogramas direccionales, los cuales se calculan típicamente a 0, 45, 90 y 135 grados. Si la forma de los mismos no se diferencia significativamente en cada una de las trayectorias con la angulación mencionada, esto está dando cuenta de que la componente dirección no es algo que se deba considerar a posteriori para el análisis del comportamiento del AB.

Restricciones matemáticas del variograma

El modelo que se elige debe describir la variación aleatoria, y la función debe ser tal que no dé lugar a combinaciones de variables aleatorias con "varianzas negativas", como se explica a continuación, en base a Webster (2007).

Sea $Z(x_i)$ una realización de la variable aleatoria Z , con $i = 1, 2, \dots, n$, con función de covarianza $C(h)$ y variograma $\gamma(h)$. Si se considera la suma lineal:

$$Y = \sum_{i=1}^n \lambda_i z(\mathbf{x}_i) \quad (11)$$

donde los λ_i son pesos arbitrarios y la variable aleatoria Y tiene varianza:

$$\text{var}[Y] = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(\mathbf{x}_i - \mathbf{x}_j) \quad (12)$$

donde $C(x_i, x_j)$ es la covarianza de Z entre x_i y x_j . La varianza de Y puede ser positiva o cero, pero no puede ser negativa. El lado derecho de la última ecuación debe asegurar esto. La función de covarianza, $C(h)$, debe ser semidefinida positiva y la ecuación anterior se puede escribir como:

$$\text{var}[Y] = \lambda^T \mathbf{C} \lambda \geq 0 \quad (13)$$

donde λ es el vector de pesos y C es la matriz de covarianzas. Si esta última es semidefinida positiva entonces también lo es la función de covarianza. De hecho, al estar tratando con 'variables', la varianza no puede ser cero, por lo que $C(h)$ debe ser definida positiva. Si la covarianza no existe, porque la variable es solo intrínseca y no estacionaria de segundo orden, entonces se reescribe la ecuación anterior como:

$$\text{var}[Y] = C(\mathbf{0}) \sum_{i=1}^n \lambda_i \sum_{j=1}^n \lambda_j - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) \quad (14)$$

donde $\gamma(x_i, x_j)$ es la semivarianza de Z entre x_i y x_j . El primer término en el lado derecho de la ecuación contiene $C(\mathbf{0})$, la covarianza en el paso 0 ($h=0$), que no sabemos, pero podemos eliminar haciendo que los pesos sumen 0 sin pérdida de generalidad.

$$\text{var}[Y] = - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) \quad (15)$$

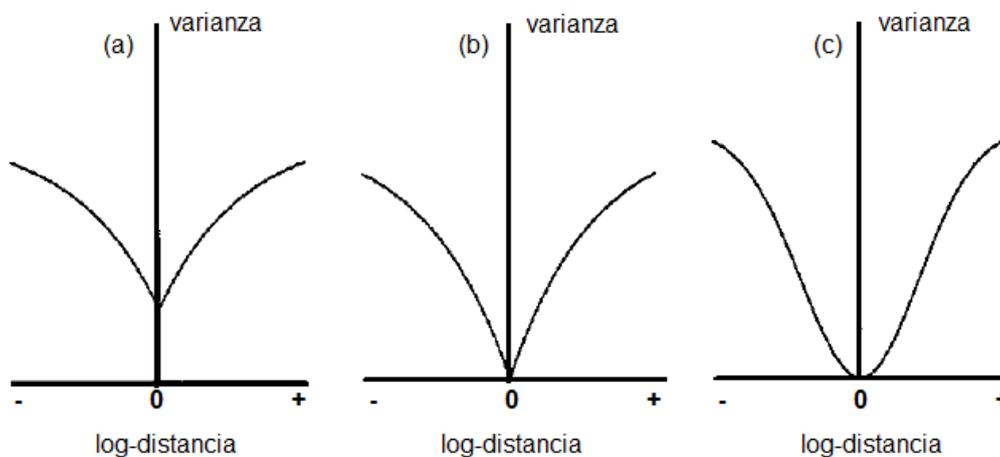
Esto tampoco puede ser negativo, pero se observa el signo de menos, por lo que el variograma debe ser semidefinido negativo condicional, siendo la condición que los pesos en la ecuación anterior sumen cero.

Solo las funciones que aseguran variaciones distintas de cero pueden usarse para variogramas. Ellos son llamados modelos autorizados en gran parte de la literatura.

Comportamiento cerca del origen

La forma en que el variograma se aproxima al origen está determinada por la continuidad o discontinuidad de la variable $Z(x)$. Se puede distinguir los siguientes casos, todos comparten la característica de que el variograma es simétrico con respecto al origen:

Figura 4: Comportamiento del variograma cerca del origen: (a) intercepto positivo (pepita); (b) enfoque lineal, no diferenciable; (c) continua y diferenciable.



Fuente: Webster (2007), p.80.

A continuación se describen los casos representados en la figura anterior:

- **(a) intercepto positivo (pepita):**

La semivarianza en $|h| = 0$ es por definición 0. Sin embargo, a menudo sucede que cualquier línea o superficie proyectada a través de los valores experimentales para la ordenada lo interseca en algún valor positivo, como en la figura (a), lo que implica un discontinuidad en $Z(x)$. Esto sucede con frecuencia en la minería, especialmente la de oro, y es atribuido a la ocurrencia independiente espacialmente de pepitas de oro en cuerpos

minerales, por lo que se le llama al fenómeno el *efecto pepita* y a la intersección *varianza pepita* (Webster, 2007).

Es posible imaginar discontinuidades derivadas de la dispersión de pequeñas pepitas de oro en un gran cuerpo de roca. Lo mismo podría ocurrir para ciertas características del suelo, como el AB, aunque las discontinuidades en las propiedades físicas y químicas del suelo son más difíciles de imaginar, y cualquier variación aparente de pepita generalmente surge de errores de medición y variación espacial dentro del intervalo de muestreo más corto.

- **(b) Enfoque lineal:**

El variograma puede aproximarse al origen linealmente con distancia de retardo decreciente:

$$\gamma(\mathbf{h}) \approx b|\mathbf{h}| \quad (16)$$

mientras $|\mathbf{h}| \rightarrow 0$ y donde b es el gradiente. El variograma pasa por el origen, como en la figura (b), pero a diferencia de la figura (a), su primera derivada es discontinua allí: su gradiente cambia abruptamente de negativo a positivo. Sin embargo, hay continuidad en $Z(x)$ porque:

$$\lim_{|\mathbf{h}| \rightarrow 0} E [\{Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})\}^2] = 0 \quad (17)$$

- **(c) Enfoque parabólico:**

La figura (c) ilustra la situación en la que el variograma pasa suavemente por el origen con un gradiente de 0 allí:

$$\gamma(\mathbf{h}) = b|\mathbf{h}|^2 \quad \text{mientras } |\mathbf{h}| \rightarrow 0 \quad (18)$$

El variograma es dos veces diferenciable en el origen y $Z(x)$ es en si mismo diferenciable, varía suavemente y no es diferenciable. El exponente 2 representa un límite estricto a las funciones de potencia para describir procesos aleatorios. Un variograma sin procesar

que parece parabólico en el origen sugiere que hay tendencia local, es decir, variación determinista de corto alcance. La esperanza de $Z(x)$ no es estacionaria sino que depende de la posición x , por lo tanto: $E[Z(\mathbf{x})] = u(\mathbf{x})$

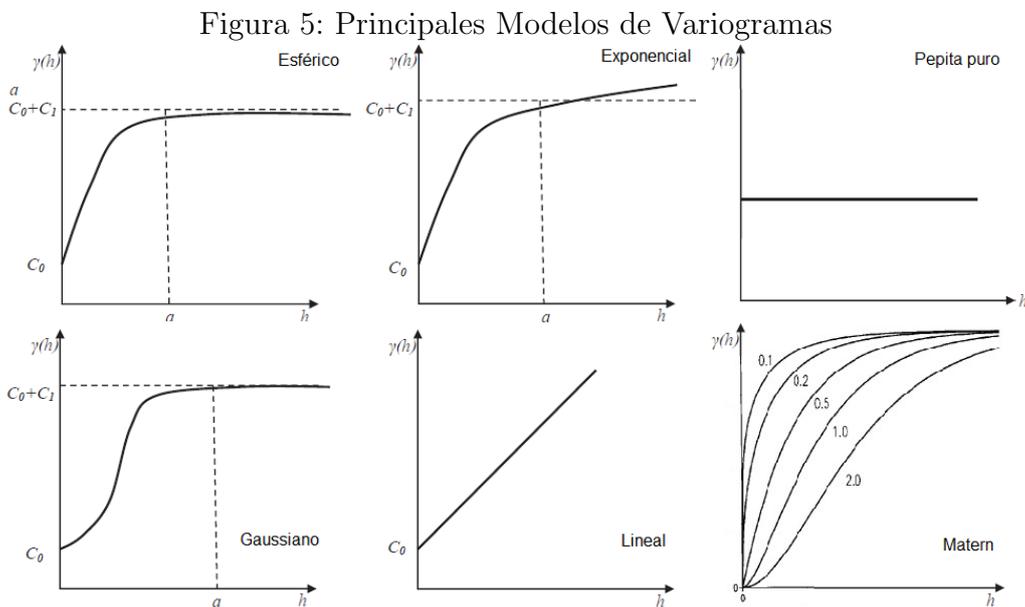
Comportamiento hacia el infinito

La forma en que se comporta un variograma al aumentar la distancia del paso h está restringida por:

$$\lim_{|h| \rightarrow \infty} \frac{\gamma(h)}{|h|^2} = 0 \tag{19}$$

El variograma debe aumentar menos que el cuadrado de la distancia del paso a medida que este último se acerca al infinito; si no lo hace, entonces el proceso no es completamente aleatorio. Cualquier función que aumenta más, no es compatible con la hipótesis intrínseca. Un variograma que aumenta más rápido que $|h|^2$ sugiere que hay una tendencia de largo alcance.

Principales Modelos de Variogramas



Fuente: En base a Webster, R. (2007) y Giraldo, R. (2002).

Modelo esférico

Este modelo se caracteriza por presentar una tendencia curvilínea hasta llegar a la meseta, a una distancia a o r (rango actual), a partir de la cual es constante. El semivariograma del modelo esférico tiene la ecuación siguiente:

$$\gamma(h) = \begin{cases} (C_0) + (C_1)(1,5(\frac{h}{a}) - 0,5(\frac{h}{a})^3) & \text{si } h \leq a \\ (C_0) + (C_1) & \text{si } a \leq h \end{cases} \quad (20)$$

Donde $h \in \mathbb{R}^+$

Modelo Exponencial

Se caracteriza por alcanzar la meseta de manera asintótica, por lo que la distancia a o r (rango actual) es el valor para el cual se ha alcanzado el 95 % de esta meseta. El semivariograma del modelo exponencial tiene la ecuación siguiente:

$$\gamma(h) = (C_0) + (C_1)(1 - e^{(\frac{-3h}{r})^3}) \quad (21)$$

Donde $h \in \mathbb{R}^+$

Modelo Gaussiano

Al igual de lo que ocurre con el modelo exponencial, el modelo gaussiano alcanza la meseta de manera asintótica, por lo que la distancia a o r (rango actual) es el valor para el cual se ha alcanzado el 95 por ciento de esta meseta. El semivariograma del modelo gaussiano tiene la ecuación siguiente:

$$\gamma(h) = C_0 + C_1(1 - e^{(\frac{-h^2}{r^2})}) \quad (22)$$

Donde $h \in \mathbb{R}^+$

Modelo Lineal

No presenta meseta, este modelo es utilizado para describir procesos donde la dispersión es creciente con el tamaño de h , es decir, procesos donde la dispersión puede alcanzar valores tan grandes que se los considera como infinito. El semivariograma del modelo lineal tiene la ecuación siguiente:

$$\gamma(h) = (C_0) + kh \quad (23)$$

Donde $h \in \mathbb{R}^+$

Modelo Pepita Puro (Pure Nugget)

Presenta una comportamiento constante, implica que no existe correlación espacial. El semivariograma del modelo pepita puro tiene la ecuación siguiente:

$$\gamma(h) = C_0 \quad (24)$$

Donde $h \in R^+$

Modelo Matern

La familia Matern es una clase más general de funciones que depende de un parámetro adicional que regula la suavidad. Son modelos con pendiente no nula en el origen. Un mayor v implica menor $\gamma(h)$, y en consecuencia mayor continuidad espacial. En particular, el caso $v = 0,5$ corresponde al modelo Exponencial, a partir de $v = 0,6$ se observa un cambio en la concavidad, cuando se trata del modelo Gaussiano.

$$\gamma(h) = C \left\{ 1 - \frac{1}{2^{v-1}\Gamma(v)} \left(\frac{h}{r}\right)^v K_v \left(\frac{h}{r}\right) \right\} \quad (25)$$

Donde:

C es la varianza del proceso.

v el parámetro de forma.

r rango.

$h \in R^+$.

K_v la función de Bessel modificada de segunda clase.

$\Gamma(v)$ la función Gamma del parámetro de forma.

3.3. Kriging

La mayoría de las variables vinculadas al medio ambiente podrían medirse en cualquiera de los infinitos puntos, pero en la práctica las mediciones se realizan solamente en una muestra reducida de puntos, principalmente por razones económicas y de disponibilidad de tiempo (Webster, 2007).

Por lo tanto, si se desea conocer el valor de las variables en otro lugar, por ejemplo sobre áreas en las que no ha sido posible medir directamente, se deben realizar estimaciones a partir de los datos que sí se pudieron obtener.

Dentro de los procedimientos de predicción espacial, son ampliamente aplicados los métodos de Kriging (Krige, 1951). Los mismos son una metodología de interpolación que utilizan un modelo de variograma. Estos métodos, parten del principio de que puntos próximos en el espacio tienden a tener valores más parecidos que los más distantes (correlación espacial). La base del método es un conjunto de algoritmos de regresión generalizada por mínimos cuadrados.

Características generales del Kriging: ⁵

- Kriging proporciona una solución al problema de la estimación basada en un modelo continuo de variación espacial estocástica.
- Hace el mejor uso de la información existente, teniendo en cuenta la forma en que una propiedad varía en el espacio a través del modelo de variograma.
- En su formulación original, una estimación krigeada en un lugar era simplemente una suma lineal o un promedio ponderado de los datos en su vecindad.
- En la actualidad kriging se utiliza para abordar problemas cada vez más complejos en minería, ingeniería petrolera, control de la contaminación, y servicios públicos de salud. El término ahora es genérico y abarca varios tipos distintos de kriging, tanto lineales como no lineales.

El valor desconocido $Z(x)$ en cualquier ubicación se descompone típicamente en un componente medio (media) $\mu(x)$ y un componente residual $\varepsilon(x)$, por lo que el modelo en el que se basa el kriging es de la forma $Z(x) = \mu(x) + \varepsilon(x)$

Existen diferentes variantes de kriging, estas se diferencian según se asuma que el componente medio $\mu(x)$ es constante o espacialmente variable (Goovaerts, 1997). Cuando el componente medio es constante y desconocido, el kriging ordinario es el procedimiento para estimar el valor desconocido en cualquier ubicación x no muestreada. Alternativamente, el componente medio puede modelarse como variable espacialmente expresándolo como una función de variables auxiliares (predictores) que varían en el espacio, por ejemplo, se puede adoptar una función lineal. Por lo tanto, la forma funcional desarrollada por μ es la que determina el valor esperado de $Z(x)$.

- $\mu(s) = \mu$ constante y conocida (Kriging Simple)
- $\mu(s) = \mu$ desconocida (Kriging Ordinario)
- $\mu(x) = f(x)^{tr} \beta$ (Kriging Universal)

La componente $\varepsilon(x)$ se supone espacialmente autocorrelacionada y es la fuente de aleatoriedad del modelo. La misma se caracteriza por las siguientes condiciones:

- $\varepsilon(x) = 0$

⁵Basado en Webster et al (2007), capítulo 8.

- $E(\epsilon(x+h) - \epsilon(x))^2 = \gamma_\epsilon(h)$
- $E[\epsilon(x+h), \epsilon(x)] = C_\epsilon(h)$

El objetivo de la técnica de kriging es minimizar la varianza del error de predicción sujeto a la restricción de insesgadez del estimador, es decir, se minimiza $\text{Var}(\hat{Z}(x) - Z(x))$ sujeto a $E(\hat{Z}(x) - Z(x)) = 0$, siendo $\hat{Z}(x)$ el valor estimado para la posición x .

Por lo tanto, el predictor empleado por el método de kriging tiene las propiedades de ser el mejor estimador lineal e insesgado (BLUP: Best Linear Unbiased Predictor), a continuación se explica esto en mayor detalle:

- Lineal: es una combinación lineal ponderada de los datos.
- Insesgado: el error de estimación tiene una media igual a 0, es decir, $E(\hat{Z}(x)) = \mu$
- El mejor: tiene un error de varianza mínima para un modelo dado de variograma.

Como se mencionó anteriormente, los métodos geoestadísticos modelan explícitamente la estructura de datos de autocorrelación espacial e incorporan esta información en la estimación de la variable de interés para ubicaciones no muestreadas (Cressie, 1993; Hengl, 2009). Es importante destacar que algunos de estos métodos, como el kriging universal y el kriging con deriva externa, hacen uso de datos auxiliares recopilados de manera más exhaustiva para el área de interés, como lo son las imágenes satelitales. Los distintos métodos difieren en la forma en que utilizan la información auxiliar para estimar la variable de interés en áreas no muestreadas (Hengl, Heuvelink, y Rossiter, 2007). Se puede encontrar una revisión de varios métodos de interpolación en otros artículos (Caruso y Quarta, 1998; Li y Heap, 2014).

Kriging Ordinario

El kriging ordinario de una sola variable, es el método más robusto y el más utilizado. Se utiliza para predecir el comportamiento de una variable predeterminada, en un escenario de los menos estrictos, supone únicamente que se verifican las 2 condiciones que definen la estacionariedad intrínseca:

1. $E(\epsilon(x)) = 0$
2. $E(\epsilon(x+h) - \epsilon(x))^2 = \gamma_\epsilon(h)$

Supone que el valor medio de la variable aleatoria de interés se mantiene constante y es desconocido a priori.

Kriging Universal

En el caso del kriging universal se deben agregar las coordenadas X e Y como explicativas al conjunto de datos.

Kriging Universal con Deriva Externa (KED)

Además de usar información de la variable de interés solamente, el kriging puede usar información secundaria para mejorar la predicción (Goovaerts, 1997)⁶. El componente medio puede modelarse como variable espacialmente expresándolo como una función de variables auxiliares (predictores) que varían en el espacio. Por ejemplo, se puede adoptar la siguiente función lineal:

$$\mu(x) = \sum_{i=0}^A b_i x_i(x) \quad (26)$$

donde $x_i(x)$ representa el valor de la variable externa i -ésima en ubicación x , y b_i denota el coeficiente de regresión asociado con esa variable externa; por convención $b_0 = 1$.

La principal diferencia del kriging universal con el KED es que en el primero la tendencia es a menudo una función de variables internas (coordenadas) mientras que en el segundo la tendencia es a través de los datos secundarios (variables externas). Matemáticamente son idénticos, solo que la tendencia puede ser una regresión lineal de variables externas.

Otros tipos de Kriging⁷

- El kriging simple se usa poco porque generalmente no se conoce la media.
- El kriging lognormal es el kriging ordinario de los logaritmos de los valores medidos. Se utiliza para datos fuertemente sesgados positivamente que se aproximan a una distribución lognormal.
- Kriging factorial es de particular valor cuando la variación está anidada, es decir, está presente más de una escala de variación. El kriging factorial estima los componentes individuales de la variación por separado, pero en un solo análisis.
- El cokriging ordinario es la extensión del kriging ordinario de una sola variable a dos o más variables. Debe haber cierta correogionalización entre las variables para que sea aplicable.
- El indicador kriging es una forma no lineal y no paramétrica de kriging en el que las variables continuas se convierten en binarias, unos (indicadores) y ceros. Se está volviendo popular porque puede manejar distribuciones de casi cualquier tipo. También puede incorporar información cualitativa para mejorar la predicción.

⁶Otro algoritmo que podría utilizarse es el de cokriging colocado ordinario (OCCK: ordinary collocated cokriging), donde los datos secundarios, que también se consideran una variable aleatoria, influyen en la predicción del kriging directamente, y no solo en la tendencia (como en el KED)

⁷Basado en Webster et al (2007), capítulo 8.

- El kriging disyuntivo también es un método no lineal de kriging, pero es estrictamente paramétrico. Es valioso para la toma de decisiones porque puede estimar las probabilidades de exceder o no exceder un umbral predefinido.
- El kriging bayesiano fue introducido para situaciones en las que existe algún conocimiento previo sobre la deriva. Es intermedio entre kriging simple, usado cuando no hay deriva, y kriging universal donde se sabe que hay deriva. Las ecuaciones son las de kriging simple, pero con covarianzas no estacionarias.

Pesos⁸

Cuando se resuelven las ecuaciones de kriging para obtener los pesos λ_i , en general, los puntos cercanos al punto o bloque a krigear son los que tienen mayor ponderación.

Los cuatro o cinco puntos más cercanos pueden aportar el 80 % del peso total, y los siguientes diez más cercanos casi todo el resto. Los pesos también dependen de la configuración del muestreo.

Podemos resumir los factores que afectan a los pesos de la siguiente manera:

1. Los puntos cercanos tienen más peso que los más distantes. Dependen de las posiciones de los puntos de muestreo y del variograma, cuanto mayor es la varianza de la pepita, menores son los pesos de los puntos más cercanos al punto objetivo.
2. Los pesos relativos de los puntos también dependen del tamaño del bloque, cuando el tamaño del bloque aumenta, los pesos de los puntos más cercanos disminuyen y los de los más lejanos aumentan, hasta que los pesos se vuelven casi iguales.
3. Los puntos agrupados tienen menos peso individualmente que los aislados en la misma distancia.

Todos estos efectos son intuitivamente deseables, a su vez, se constatan en la aplicación empírica que se lleva a cabo en el presente trabajo, como se señalará en la sección de resultados.

Validación del método de Kriging

Mediante este procedimiento se busca evaluar la capacidad predictiva de los modelos elegidos, así como también permite decidir cuál será el mejor estimador de kriging. En el presente trabajo se utiliza la técnica de Validación Cruzada *Dejando uno afuera* (leaving one out cross-validation, con sigla LOOCV), el procedimiento implica estimar sucesivamente mediante Kriging cada observación, a partir de utilizar las observaciones restantes. Es decir, dadas las n observaciones, y los valores observados de AB $Z(x_1), Z(x_2), \dots, Z(x_n)$, en cada punto i se estimó con el conjunto de las $n - 1$ observaciones restantes, es decir, excluyendo al punto i , obteniendo $\hat{Z}_{[i]}(x_i)$.

⁸Basado en Webster et al (2007), capítulo 8.3.

Si el modelo de semivarianza o covarianza elegido describe bien la estructura de autocorrelación espacial, entonces la diferencia entre el valor observado y el predicho deberá ser pequeña (Melo Martínez, 2012).

Cabe señalar que en los métodos de interpolación kriging los valores de las predicciones coinciden con los valores observados para los puntos muestreados, por lo que la validación cruzada busca dar una idea de qué tan buenas son las predicciones, para indagar cuál modelo tiene mayor exactitud (Melo Martínez, 2012).

Por lo tanto, los diferentes enfoques para la estimación del AB se evaluaron comparando los valores observados y pronosticados en los conjuntos de datos de ajuste y validación. Las medidas que se usaron para estas comparaciones son: la media del residuo de validación cruzada o Error Medio (ME), que idealmente es pequeño; el ME dividido por la media de los valores observados (*ME_mean*), para ver qué tan grande es el *ME* respecto a la media del conjunto de datos; el Error Absoluto Medio (MAE); el Error Cuadrático Medio (MSE); la correlación entre los valores observados y predichos, que idealmente es 1 (*cor_obspred*); la correlación entre los valores predichos y residuales (*cor_predres*), que idealmente es 0; la Raíz del Error cuadrático medio (RMSE), cuanto más pequeño, más exacto es el método.

$$ME = \frac{\sum_{i=1}^n (Z(x_i) - \hat{Z}_{[i]}(x_i))}{n} \quad (27)$$

$$ME_mean = \frac{ME}{\bar{Z}} \quad (28)$$

$$MAE = \frac{\sum_{i=1}^n |Z(x_i) - \hat{Z}_{[i]}(x_i)|}{n} \quad (29)$$

$$MSE = \frac{\sum_{i=1}^n (\hat{Z}_{[i]}(x_i) - Z(x_i))^2}{n} \quad (30)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Z(x_i) - \hat{Z}_{[i]}(x_i))^2}{n}} \quad (31)$$

donde n es el número de valores en el conjunto de datos; $\hat{Z}_{[i]}(x_i)$ es el valor estimado de la variable principal; $Z(x_i)$ es el valor observado en los conjuntos de predicción y validación.

Por último, en base a Reis et al (2018)⁹, se utilizará un indicador de mejora relativa (*relative improvement*), para comparar el cambio en el $RMSE$ cuando se incorporan covariables en el kriging, es decir, para comparar los resultados del KED respecto al OK . El mismo se expresa en porcentaje.

$$RI = \frac{RMSE_{ok} - RMSE_{ked}}{RMSE_{ok}} \times 100 \% \quad (32)$$

donde $RMSE_{ok}$ es el $RMSE$ de la predicción espacial del método Kriging Ordinario, y $RMSE_{ked}$ es el correspondiente a Kriging con Deriva Externa.

⁹Cabe señalar que en dicho antecedente se comparan diferentes abordajes de Kriging con el mencionado indicador.

4. Metodología

En esta sección se delimita el área de estudio y se describen los datos que se utilizaron para llevar a cabo el presente trabajo, así como también se definen las variables construidas para operativizar conceptos teóricos, se presentan algunas descriptivas de las mismas y se detallan sus restricciones.

A continuación se resumen las etapas del procedimiento realizado:

1. Definición del área de estudio donde se estimará el modelo y el área de interés en donde se realizará la predicción.
2. Obtención de las variables a incorporar al modelo: generación de rasters para las bandas Landsat para todo el país, recorte y reproyección de mosaicos, extracción de la pendiente y elevación del raster del Modelo Digital de Terreno; elaboración de otros rasters de interés a partir de polígonos (Índice Coneat, Evapotranspiración), y la obtención de rasters de otras fuentes de información como el de los Índices de vegetación, precipitaciones, temperatura, agua, entre otros.
3. Extracción de la información de los rasters anteriores en los puntos del IFN, creando un archivo de puntos para el área de estudio, que contenga información de todas las variables disponibles en los rasters.
4. Realizar un análisis descriptivo de las variables disponibles (mediante mapas, gráficos, resúmenes de variables).
5. Selección de las variables que serán utilizadas para modelar el AB, utilizando un algoritmo *stepwise* en contexto de un MLR.
6. Generación de un *brick*¹⁰ de los rasters correspondientes a las variables seleccionadas en el paso anterior, que será utilizado en la aplicación de Kriging con deriva externa (KED).
7. Estimación de variogramas omnidireccionales y direccionales del AB.
8. Estimación del modelo kriging para sus tres variantes: OK, UK y KED, generando los mapas de predicción para el área de interés.
9. Selección del mejor método de predicción utilizando validación cruzada.

Este trabajo fue realizado principalmente con el lenguaje de programación R y la interfaz de usuario R studio (R Core Team, 2017, RStudio Team, 2016), en el Apéndice se detallan las librerías utilizadas y el código desarrollado.

¹⁰Es un objeto de varias capas (o multibanda).

4.1. Área de estudio

Si bien se cuenta con datos para todo el país, abordarlo en su totalidad en este trabajo presenta varias limitaciones. En primer lugar, el AB tiene un comportamiento diferente en distintas zonas del país, que se ve influenciado por variables inobservables que no se pueden introducir en los modelos. Entre ellas se encuentran diversas dinámicas que pueden variar entre zonas del país, siendo de interés analizar zonas más pequeñas por separado. Un ejemplo de esto es la competencia entre los distintos tipos de árboles y arbustos por la luz, agua y nutrientes; así como también la presencia de arbóreas invasoras¹¹ sobre la regeneración de especies nativas, tanto por competencia como por alteración de las condiciones micro-ambientales vinculadas al reclutamiento (Brazeiro, 2018).

En línea con lo anterior se encuentran diversos antecedentes donde se selecciona un área reducida de bosques para realizar las predicciones mediante Kriging, incluso más pequeña que la seleccionada en el presente trabajo. Por ejemplo, Vicharnakorn et al (2014) consideró un área de interés de aproximadamente 100x160km para la estimación. López Serrano et al (2016) trabajaron con parcelas cuadradas de 50x50 metros (superficies de 2500m²), y el área de estudio fue de 80x100 kilómetros. Fayada et al (2016) realizó el mapeo del AB, para cuadrículas de distintos tamaños (500m, 1000m, y 2000m), en un área de estudio de 250x300km con diferentes tipos de bosque.

La selección del área de estudio, también denominada ventana, resulta una tarea compleja, siendo importante la consideración de varios aspectos, existiendo *trade-off*. Por un lado, resulta necesario contar con una cantidad suficiente de puntos para obtener estimaciones precisas, donde el AB sea la variable de interés y se utilicen diversas variables explicativas. Por otro lado, cuanto mayor es el área de estudio, más probable resulta que el comportamiento del AB responda a procesos de diferentes características, existiendo mayor influencia de variables inobservables y la superposición de efectos.

Brazeiro et al (2015) identifica varias eco-regiones con características diferenciadas en el territorio nacional, entre las que se encuentran diferencias en el AB e índices de vegetación, entre otros; las eco-regiones se representan en la Figura 6. Para el presente trabajo se decide analizar el comportamiento del AB y estimar los modelos en la Cuenca Gondwánica, para posteriormente realizar la predicción únicamente en una sub-área definida por la Dirección General Forestal (DGF). Esta última consiste en una selección de cuencas de nivel 4, de particular interés para el organismo¹².

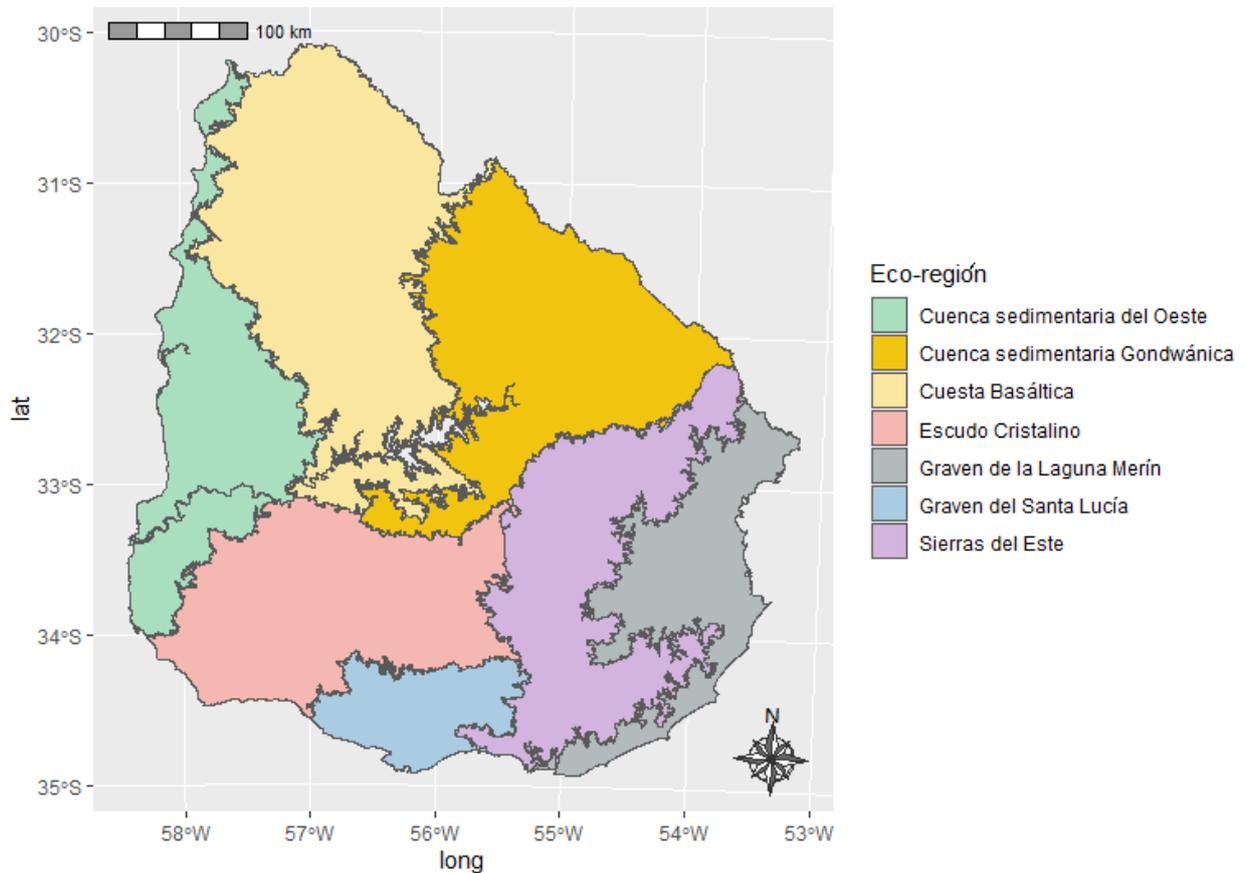
La Cuenca Gondwánica es la segunda del país en extensión, con una superficie de más de 3 millones de hectáreas. Presenta rangos altitudinales de 90 a 380 m, con formaciones geológicas dominantes compuestas por areniscas de edad gondwánica y geoformas de lomadas. A su vez,

¹¹Las especies invasoras compiten por el espacio y por la luz, crecen mucho más rápido que el bosque nativo. A su vez, no resulta sencillo diferenciar BN de especies invasoras a partir de imágenes satelitales, lo que puede impactar en el análisis del presente trabajo.

¹²El área de interés en la cual se tiene como objetivo predecir, cuenta con vacíos de información relevada en el IFN de puntos, siendo especialmente importante la predicción de AB en el lugar.

el uso de la tierra es pastoril y forestal. Se trata de una muy importante ecorregión con gran diversidad de suelos (Brussa, 2015).

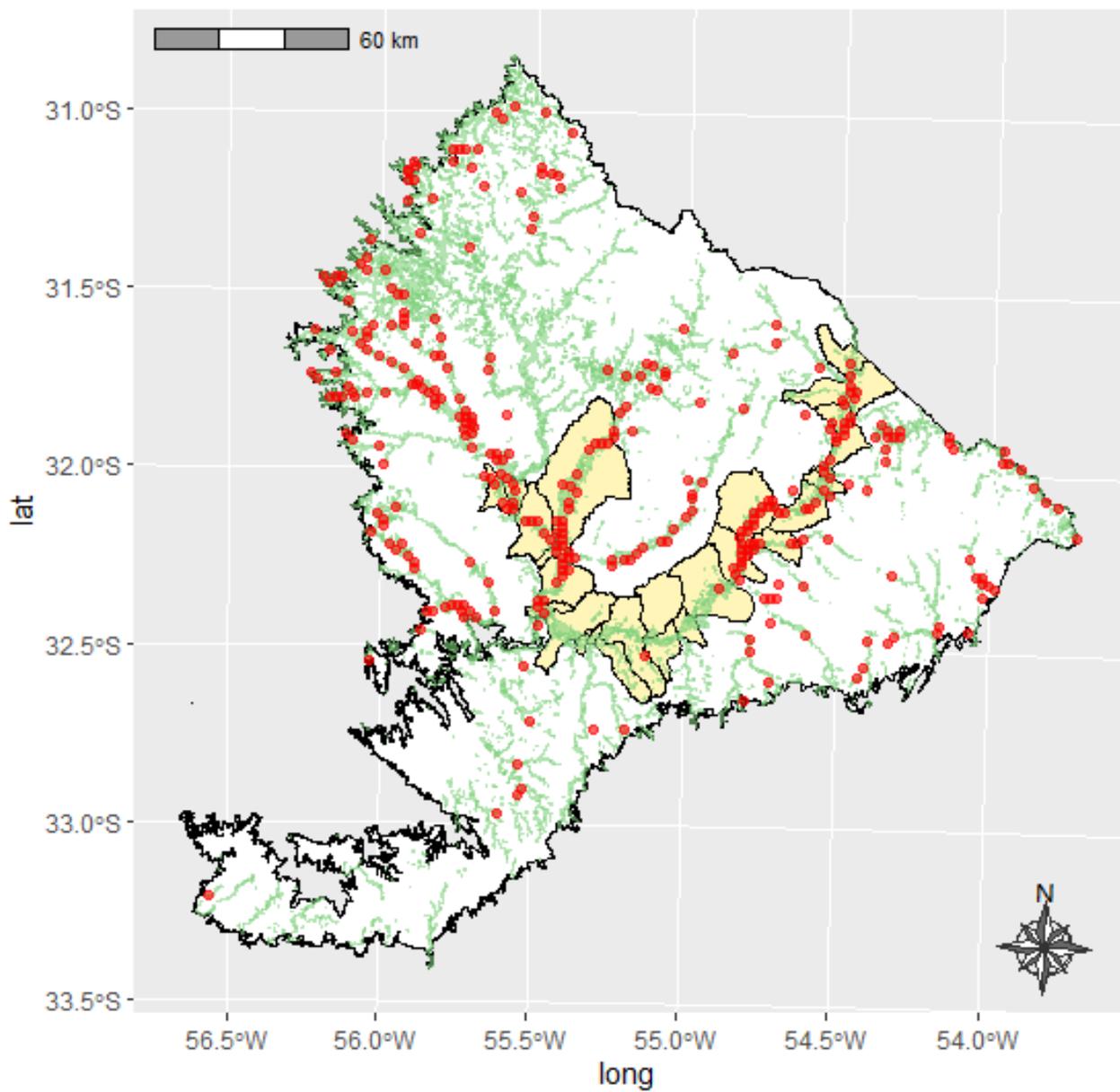
Figura 6: Ecozonas por unidades morfoestructurales



Fuente: Elaboración propia en base a Brazeiro, A. (2015).

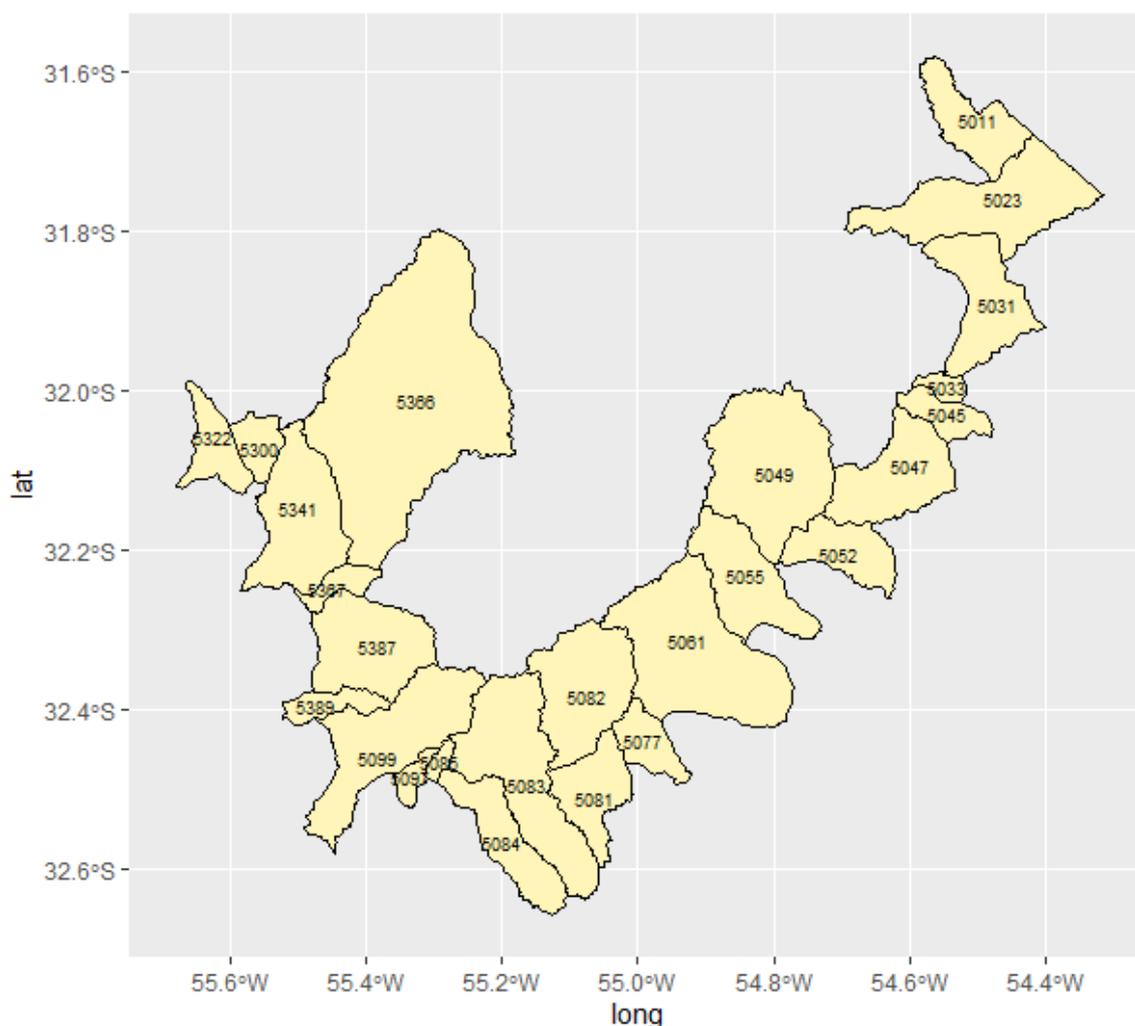
A continuación se representa la Cuenca Gondwánica, el Bosque Nativo (en verde), se superponen los puntos del IFN (en rojo), que son 331, a partir de los cuales se estimará el modelo explicativo de AB. A su vez, se representa la zona seleccionada por la Dirección General de Forestación (DGF) para la realización de las predicciones de AB (en amarillo).

Figura 7: Bosque Nativo en la Cuenca Gondwánica, selección de subcuencas y puntos del IFN



Fuente: Elaboración propia en base a shapefiles de Cuencas e información enviada por DGF.

Figura 8: Cuencas seleccionadas por DGF (Nivel 4) para la predicción de AB

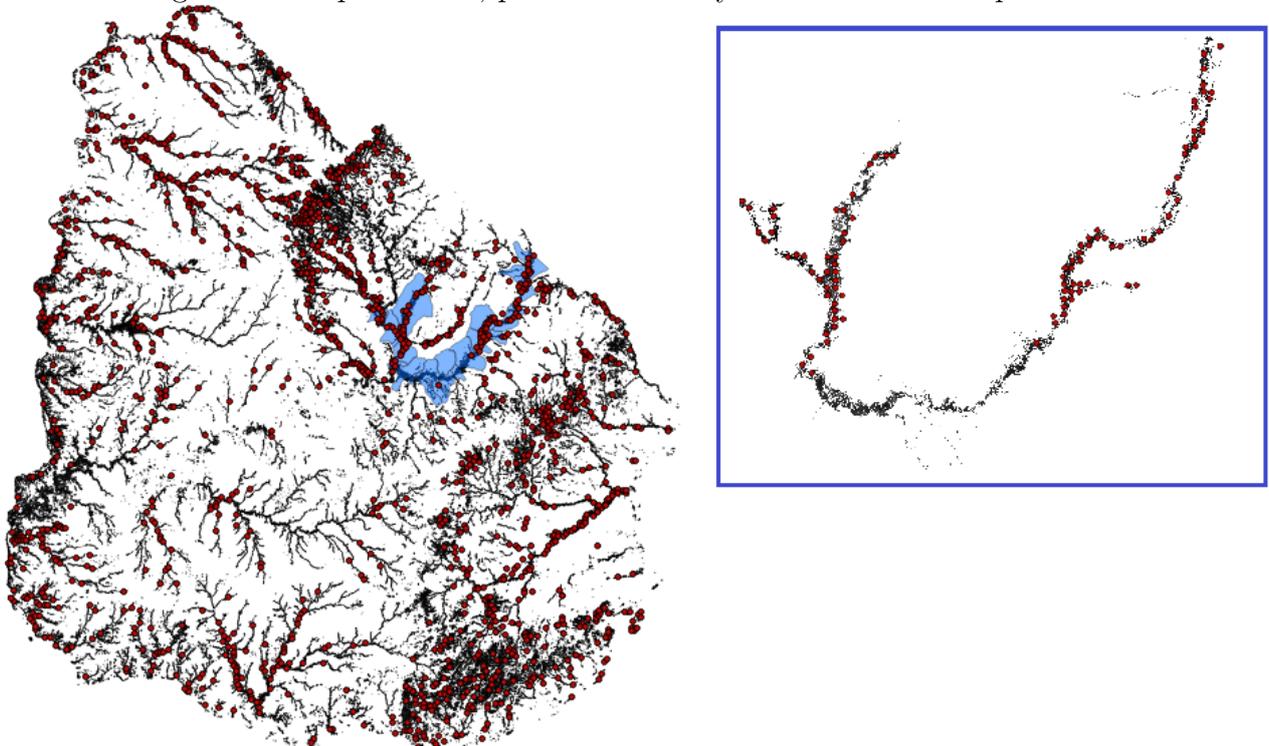


Fuente: Elaboración propia en base a shapefiles de Cuencas e información enviada por DGF.

A continuación se representa el área de Bosque Nativo (datos abiertos¹³) en todo el país, superpuesto con los puntos del IFN (en rojo). Para obtener el área de predicción del presente trabajo, se recortan dichas capas por el polígono de subcuencas seleccionadas por la DGF (representado en celeste), obteniendo el resultado del recuadro azul. Cabe señalar que los puntos del IFN que quedan por fuera de la cartografía de Bosque Nativo están sujetos a mayores errores de medida, por lo que serán descartados en los procedimientos de estimación.

¹³Los sensores utilizados para la elaboración de los mapas de cobertura de bosque nativo de Uruguay para el año 2016 fueron Landsat 8 y Sentinel 2, la metodología y datos abiertos se encuentra disponible en: [link](#)

Figura 9: Bosque Nativo, puntos del IFN y área de interés de predicción



Fuente: Elaboración propia en base a shape files de Cuencas e información enviada por DGF.

Es importante destacar que si se utilizara información únicamente del área de interés de predicción (conjunto de subcuencas), para la estimación del modelo (en lugar de la Cuenca Gondwánica), las estimaciones tendrían mayores limitaciones, principalmente porque se estaría utilizando un modelo con un conjunto numeroso de variables explicativas (inicialmente son 23), a partir de pocos puntos muestrales (95). Por ello la modelización se hace a nivel de la Eco región Gondwánica, donde se cuenta con información de 331 puntos. Por otra parte, existen otras limitaciones de los datos vinculadas al relevamiento, que al contar con una mayor cantidad de puntos para la estimación, se espera que se vean mitigadas. Por todo lo expuesto anteriormente, se decide estimar el modelo a nivel de la Cuenca Gondwánica, para posteriormente realizar la predicción en la zona de interés.

4.2. Datos utilizados

En la presente sección se introduce el conjunto de procedimientos seguidos, con la finalidad de obtener, procesar y unificar los datos utilizados en la investigación.

En la etapa inicial del abordaje empírico del presente trabajo se realizó una búsqueda de shapefiles y rasters con información de variables que pudiesen estar conceptualmente vinculadas al AB. Para esto se indagó acerca de diversas herramientas disponibles: SIGRAS¹⁴, IDE¹⁵, USGS¹⁶, Google Earth Engine (GEE)¹⁷.

GEE es una plataforma que permite a los usuarios acceder a una gran colección de archivos de datos de teledetección y ejecutar análisis geoespaciales en la infraestructura de Google. Permite desarrollar un script en la plataforma, mediante el cual, por ejemplo, se implementen filtros de fecha y nubosidad para la obtención de las imágenes. Actualmente ofrece soporte solo para Python y JavaScript, sin embargo, existe una librería para R denominada *rgee*, que permite incorporar una sesión de Python dentro de una sesión de R, permitiendo una interoperabilidad perfecta (Gorelick y cols., 2017; Aybar, Bautista, Yali, Barja, y Wu, 2020)¹⁸.

Los datos utilizados para el presente trabajo fueron los siguientes:

- Shapefile de puntos del IFN (DGF)¹⁹.
- Shapefile de polígonos del Bosque Nativo elaborado a partir de imágenes Landsat (DGF).
- Shapefile de Eco-regiones (DGF).
- Shapefile de cuencas hidrográficas (datos abiertos).
- Raster del Modelo Digital de Terreno (datos abiertos), con el que se obtiene un raster de pendiente, uno de elevación y otro de Índice de Posición Topográfica (TPI).
- Rasters de bandas (de la 1 a la 11) de las imágenes LANDSAT (construido a partir de datos abiertos).
- Rasters de índices de vegetación²⁰: EVI, GI, NDVI, RVI.
- Rasters de precipitaciones y de temperatura (datos abiertos)

¹⁴Es un Sistema de Información Geográfica web (Link al SIGRAS) desarrollado por el INIA, tiene por objetivo brindar información actual e histórica del estado de la vegetación y agua en el suelo, climatología y cartografía general. La información geográfica está en formato shapefile y puede ser descargada. No se utilizó para el presente trabajo dado que el nivel de agregación de la información es más elevado de lo que se busca.

¹⁵Infraestructura de Datos Espaciales (IDE): link

¹⁶Servicio Geológico de EE. UU. (Earth Explorer): Link

¹⁷Link a Google Earth Engine

¹⁸En la etapa inicial del presente trabajo se descargaron imágenes mediante *rgee*, pero posteriormente no fueron utilizadas en el análisis ya que por otros medios se obtuvieron mejores imágenes vinculadas al AB.

¹⁹La DGF también hizo llegar una planilla con información de los puntos, entre lo que se encontraba la variable de AB, información que se unió al shapefile.

²⁰Todos provenientes de imágenes satelitales Landsat 8.

- Raster de distancia al curso de agua más cercano (construido a partir de datos abiertos)
- Raster de Densidad de Drenaje por escorrentía (DGF)
- Raster de Índice Coneat (construido a partir de shapefile)
- Raster de Evapotranspiración (ETP), vinculado al Déficit Hídrico (construido a partir de shapefile)

El archivo **Shapefile de puntos del IFN**, proporcionado por la DGF, contiene diversas variables²¹ producto del relevamiento en campo realizado en el IFN durante el período 2009-2016, la información utilizada en el presente trabajo es la referente al AB y la ubicación de los puntos. Una aclaración relevante es que el inventario incluye el relevamiento de especies invasoras, sin embargo, en el área de estudio no hay gran presencia de estas. La cuenca Gondwánica es la ecoregión de menor porcentaje de presencia de especies invasoras (Uruguay, 2019).

La forma de relevamiento del AB en campo fue mediante la medición y suma del área en 200 metros cuadrados de cada punto. Es decir, se midió el área transversal de cada vara de árbol a la altura del pecho (aprox 1,4 mts del nivel del suelo) y se sumaron todas estas mediciones en 200 mts², luego se multiplicó esta suma por 50 para llevarlo a hectárea. Obteniendo de esta forma, una magnitud de AB en metros cuadrados, promedio por hectárea.

Los **shapefiles de Bosque Nativo, Eco-regiones y Cuencas Hidrográficas**, fueron utilizados para la construcción del área de estudio en la que se estiman los modelos, y del área de predicción. Esto fue descrito en la sección anterior.

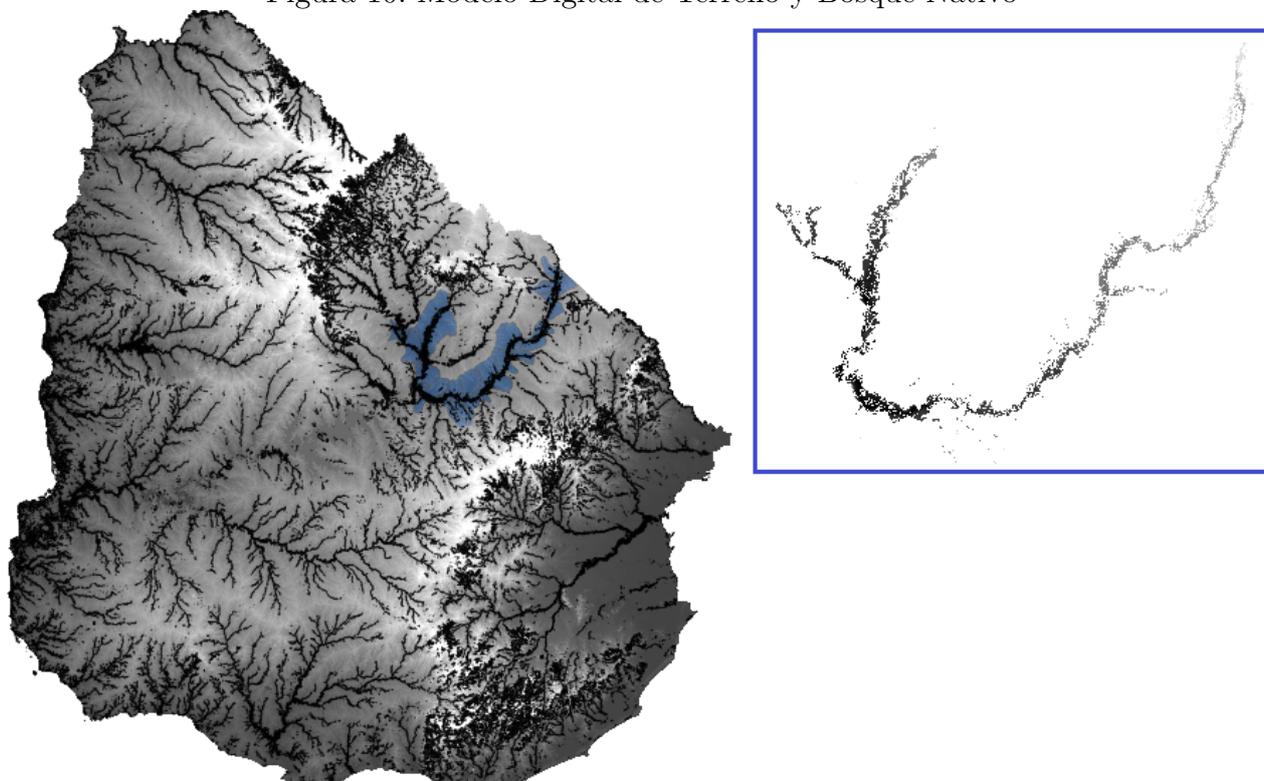
Por otra parte, un raster es una estructura empleada para datos espaciales, puede ser una fotografía aérea digital a la cuál se le asocia una matriz, organizada en filas y columnas, en la que cada celda (o píxel) tiene información referente a alguna variable. Los valores de celda o píxel pueden contener una categoría (como por ejemplo el tipo de uso del suelo), una magnitud (por ejemplo un indicador de precipitaciones o temperatura), la altura (elevación de la superficie por encima del nivel medio del mar) o valor espectral (reflectancia de la luz y color); pueden ser positivos o negativos, enteros o de punto flotante.

El **Modelo Digital de Terreno** se encuentra en formato raster, con una resolución espacial de 30 x 30 metros y una precisión de interpolación entre puntos de 2,5 metros²². Es un conjunto de capas que representan distintas características de la superficie terrestre derivadas de una capa de elevaciones. Las variables incluidas son de gran importancia en muchos procesos ambientales (precipitación, temperatura, flujos hídricos, erosión, etc) por tanto, son un elemento clave al estimar otras variables mediante procedimientos de interpolación. En el presente estudio, a partir del procesamiento del Modelo Digital de Terreno se obtienen rasters de pendiente, elevación y TPI.

²¹Algunas de las variables son: altura, tipo de bosque, ubicación, uso de la tierra, grado de erosión, drenaje, infiltración, humedad, pedregosidad y rocosidad, raíces y tipo de ganado.

²²Por más información: [Link](#)

Figura 10: Modelo Digital de Terreno y Bosque Nativo



Fuente: Elaboración propia en base a datos abiertos del Modelo Digital de Terreno.

Por otra parte, **Landsat** es un Sistema satelital de observación de la Tierra, genera información a través de una imagen multibanda, que permite evaluar y monitorear el estado de vegetación permitiendo coberturas tanto globales como regionales.

En 1972 el USGS y la NASA desarrollaron y lanzaron el primer satélite civil de observación de la Tierra, llamado Earth Resources Technology Satellite (ERTS-1), más tarde se lo pasó a llamar Landsat 1. Los lanzamientos de Landsat 2, Landsat 3 y Landsat 4 fueron en 1975, 1978 y 1982, respectivamente. Landsat 5 fue lanzado en 1984 y brindó datos globales de alta calidad de la tierra durante 28 años y 10 meses, estableciendo oficialmente un nuevo Récord mundial para la observación más prolongada de la Tierra. Landsat 6 no logró alcanzar la órbita en 1993; sin embargo, Landsat 7 se lanzó con éxito en 1999 y continúa proporcionando datos globales. Landsat 8, lanzado en 2013, continúa la misión, y Landsat 9 está tentativamente planeado para lanzarse próximamente.

Los productos del Landsat son de acceso libre, mediante servicios como GloVis y EarthExplorer del USGS. Éstos brindan un archivo que al descomprimirse nos proporciona una serie de archivos de texto (meta-datos que describen las condiciones en las que se produjeron las imágenes) e imágenes multibanda.

Tabla 1: Descripción de las bandas de las imágenes satelitales Landsat 8

Bandas de Landsat 8	Longitud de onda (micrómetros)	Resolución (metros)
Banda 1 - Aerosol Costero	0,43 – 0,45	30
Banda 2 - Azul	0,45 – 0,51	30
Banda 3 - Verde	0,53 – 0,59	30
Banda 4 - Rojo	0,64 – 0,67	30
Banda 5 - InfraRojo Cercano (NIR)	0,85 – 0,88	30
Banda 6 - SWIR 1	1,57 – 1,65	30
Banda 7 - SWIR 2	2,11 – 2,29	30
Banda 8 - Pancromática	0,50 – 0,68	15
Banda 9 - Cirrus	1,36 – 1,38	30
Banda 10 - Infraroja Térmica (TIRS 1)	10,60 – 11,19	100
Banda 11 -Infraroja Térmica (TIRS 2)	11,50 – 12,51	100

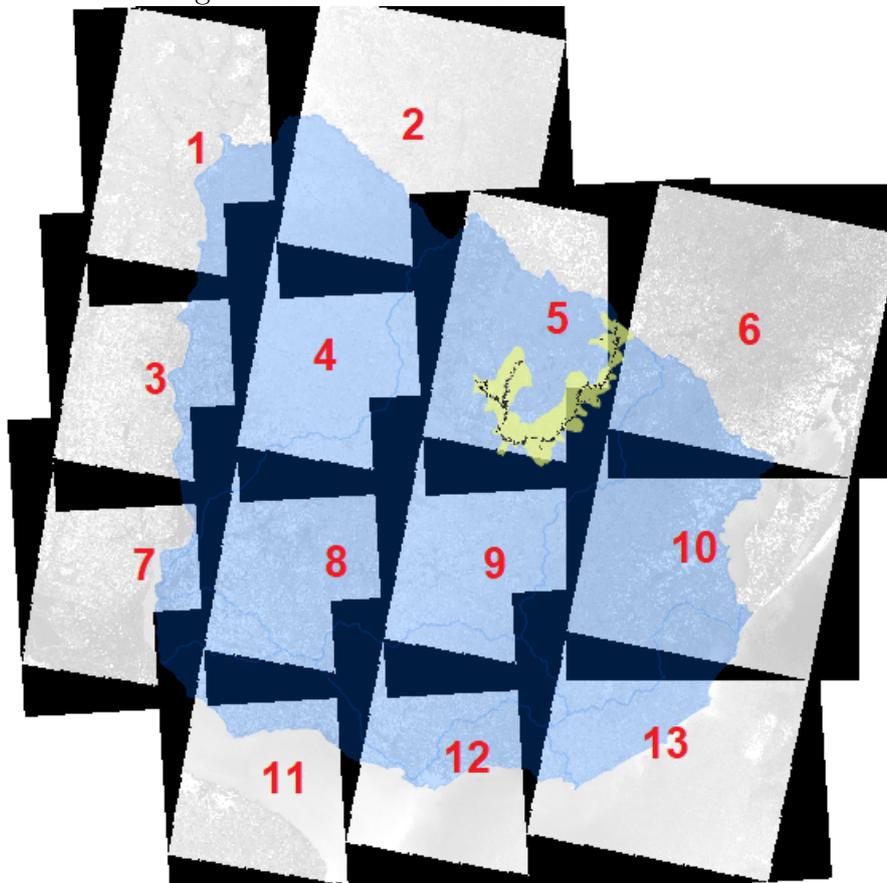
Fuente: NASA (Link) .

En el procedimiento realizado para incorporar esta información se utilizaron varias herramientas: QGIS, R, USGS²³ . Inicialmente se descargaron los mosaicos de Landsat 8 necesarios para abarcar todo el territorio uruguayo, indicados en la Figura 11, accediendo a USGS desde QGIS²⁴, luego se unieron (asegurando previamente la reproyección de cada uno a CRS 32721), y por último se recortaron por el polígono de Uruguay, obteniendo lo que se muestra en la Figura 12. También se descargó de Google Earth Engine, un único Raster de Landsat 8 para todo el territorio nacional, pero la resolución es muy inferior, siendo mucho más grande el tamaño del pixel, por lo que se descartó su uso.

²³En las siguientes páginas también se pueden visualizar y descargar imágenes satelitales: Libra, EOS, Sepal, Glovis-USGS

²⁴Para esto es necesario obtener un usuario y contraseña en la página de USGS, que luego debe ingresarse en QGIS o en el R para poder acceder a los rasters.

Figura 11: Mosaicos de Landsat utilizados



Fuente: Elaboración propia.

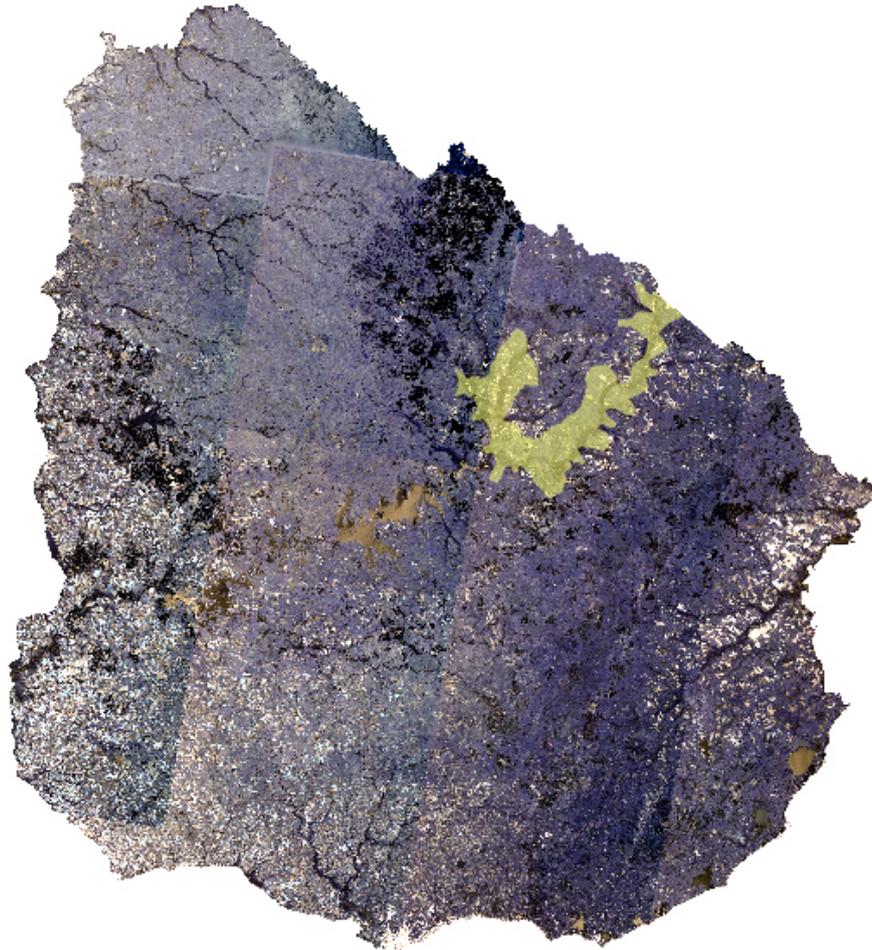
Nota: Referencias de mosaicos utilizados (código de descarga):

1. Imagen LC08_L1TP_225081_20161106_20170318_01_T1;
2. Imagen LC08_L1TP_224081_20140211_20170425_01_T1;
3. Imagen LC08_L1TP_225082_20171211_20171223_01_T1;
4. Imagen LC08_L1TP_224082_20161030_20170318_01_T1;
5. Imagen LC08_L1TP_223082_20161124_20170317_01_T1;
6. Imagen LC08_L1TP_222082_20141128_20170417_01_T1;
7. Imagen LC08_L1TP_225083_20171211_20171223_01_T1;
8. Imagen LC08_L1TP_224083_20161217_20170316_01_T1;
9. Imagen LC08_L1TP_223083_20161124_20170317_01_T1;
10. Imagen LC08_L1TP_222083_20151115_20170402_01_T1;
11. Imagen LC08_L1TP_224084_20161217_20170316_01_T1;
12. Imagen LC08_L1TP_223084_20161124_20170317_01_T1;
13. Imagen LC08_L1TP_222084_20151115_20170402_01_T1.

La selección de las escenas se realizó teniendo en cuenta factores como la presencia de nubes y la mayor cercanía posible entre las fechas de las distintas imágenes seleccionadas con el relevamiento de los puntos del IFN (2014-2016), siguiendo la metodología señalada en Bernardi et al (2019). Se eligen fechas próximas a la primavera para que la señal espectral que refleja la actividad fotosintética del bosque nativo sea clara. De esta forma, se evita el período de senescencia foliar de muchas especies que lo componen y se disminuye los errores de clasificación por confusiones con otras coberturas. La colección utilizada del USGS presenta datos que ya cumplen con los requisitos de calidad geométrica y radiométrica, por lo que no se requieren

rectificaciones adicionales (Bernardi y cols., 2019).

Figura 12: Raster Landsat Construido



Fuente: Elaboración propia.

Nota: El procedimiento para construir el raster consiste en inicialmente descargar todos los mosaicos señalados en la imagen anterior, reprojectarlos a CRS 32721, unirlos y recortarlos por el polígono de Uruguay.

A partir de los sistemas satelitales de observación de la Tierra, entre los que se encuentra Landsat, es posible evaluar y monitorear el estado de vegetación, permitiendo realizar una caracterización tanto global como regional. Para llevar a cabo lo anterior, se construyen **Índices de Vegetación** realizando operaciones con diferentes bandas de información satelital. Es decir, los índices resultan de la aplicación de una serie de operaciones algebraicas sobre los valores numéricos asociados a los píxeles de la imagen. De esta forma, un índice de vegetación es un parámetro calculado a partir del fenómeno físico de reflexión de ondas, las cuales son sensibles a la densidad vegetal.

La mayoría de los índices están correlacionados positivamente con la vegetación, de modo que valores bajos de estos se asocian a vegetación poco vigorosa, mientras que valores altos se asocian con vegetación muy vigorosa. Sin embargo, existen algunos índices de vegetación como

lo son el RVI y el NDVI, para los cuales la relación es inversa.

A los índices de vegetación se los puede agrupar en aquellos que se basan en pendiente, y los que se basan en distancia. Los primeros comparan las pendientes de las líneas que pasan por el origen y de los píxeles representados en un gráfico, con las reflectancias de las bandas en los ejes. El índice de vegetación basado en pendiente más utilizado es el NDVI, este minimiza efectos topográficos y produce escala lineal de medición. La escala va de -1 a 1 con el cero representando el valor aproximado donde empieza la ausencia de vegetación. Los valores negativos representan superficies sin vegetación. La normalización que realiza, reduce el efecto de la degradación de calibración del sensor y la influencia de los efectos atmosféricos (ñoz13, ñoz13).

Tabla 2: Índices de vegetación

Índice de Vegetación	Fórmula	Referencia
NDVI	$(BA4-BA3)/(BA4+BA3)$	Rouse et al.(1973)
SAVI	$[(BA4-BA3)/(BA4+BA3+0.5)]1.5$	Huete et al.(1988)
MSAVI	$(2BA4+1)-\sqrt{(2BA4+1)^2-8(BA4-2BA3)}/2$	QI et al.(1998)
EVI	$2.5[(BA4-BA3)/(BA4+6BA3-7.5BA1+1)]$	Justice et al.(1998)

Aclaración de la notación: BA Banda de Landsat

Fuente: Extraído de Reis et al (2018).

Múltiples investigaciones han utilizado índices de vegetación como insumo para la predicción espacial de variables forestales (Vicharnakorn P., 2014; dos Reis y cols., 2018; Gebreslasie M., 2010; Galeana y cols., 2014a; Mutanga, 2015; Dai, Zhou, Lv, Wang, y Liu, 2014).

El **Raster de Precipitaciones y el de Temperatura** se obtuvieron de *World Clim*²⁵, una plataforma con información sobre el clima mundial. A su vez, estos fueron construidos a partir del MODIS (Moderate-Resolution Imaging Spectroradiometer) a 500mts. Este es un instrumento lanzado por la NASA a la órbita terrestre en satélites, el Terra (EOS AM) en 1999 y el Aqua en 2002. Presenta 36 bandas que miden la radiación visible e infrarroja, obtiene datos que se utilizan para diversas temáticas como la vegetación, la cobertura de la superficie terrestre, la fluorescencia de la clorofila oceánica, las propiedades de las nubes, la capa de nieve en la tierra y la ocurrencia de incendios.

El archivo raster con información de **CONEAT** utilizado en el presente trabajo se obtuvo a partir de convertir un shapefile de polígonos con dicha información, proporcionado por la DGF. Los grupos **CONEAT** constituyen áreas homogéneas definidas por su capacidad productiva en términos de carne bovina, ovina y lana en pie (Art. 65 de la Ley 13.695). Esta capacidad se expresa por un índice relativo a la capacidad productiva media del país, a la que corresponde el índice 100²⁶.

²⁵Por más información: link

²⁶Extraído de la página del MGAP: link

El archivo raster utilizado con información de **Déficit Hídrico** se obtuvo a partir de convertir un shapefile de polígonos con dicha información. El **Déficit Hídrico** es el estado en el que el caudal de agua disponible (ríos, arroyos o lagunas) es insuficiente para satisfacer la demanda de la vegetación presente en la zona. Al concepto de Déficit Hídrico se le asocia el de escasez hídrica, la cual, para los caudales de agua mencionada anteriormente, la cantidad de agua esperada por registros históricos en un momento dado es inferior a la presente en dicho instante de tiempo. Este déficit es medido a través de la variable evapotranspiración (ETP), cuando es significativa, se denomina sequía y produce distintos daños en la vegetación.

El raster de **Distancia al curso de Agua más cercano** se construye en base al archivo de líneas de la hidrografía del Uruguay²⁷. Se calcula para cada píxel la distancia a la que se encuentra al curso de agua más cercano, y a partir de estos valores se construye el raster. Este procedimiento fue realizado con el software QGIS²⁸.

El raster de **Densidad de Drenaje por escorrentía** fue proporcionado por la DGF, presenta resolución MODIS. La densidad de drenaje se define como la longitud total del canal en una cuenca de drenaje dividida por el área total, en este caso, se calcula como el promedio del tamaño de cuenca de drenaje de los píxeles del Modelo Digital de Terreno para cada km².

4.3. Unión de las capas de información a utilizar

Una vez que se obtuvieron todos los rasters señalados anteriormente, se realiza una extracción de la información de cada uno, en los puntos del IFN de la Cuenca Gondwánica, unificando todo en un mismo elemento. De esta forma, se construye un archivo de 331 puntos en la Cuenca Gondwánica, con la variable AB y variables referentes a todos los rasters descritos anteriormente, estos son los datos a utilizar en la estimación del MLR, que permitirá seleccionar las variables que sean significativas para explicar el AB. Las capas (rasters) de las variables resultantes de la selección, serán posteriormente superpuestas en un *Raster Brick*²⁹ para la estimación del KED.³⁰

Para hacer la extracción de la información de cada raster en los puntos del IFN, es necesario que todos se encuentren en el mismo CRS, el que corresponde es el *epsg:32721*. A su vez, se debe decidir los criterios con los que se realiza esta extracción. Se optó por utilizar la opción de interpolación *bilinear*³¹ dado que es recomendada para variables numéricas, como lo son todas las capas del brick (en lugar de la opción *ngb*). En el método bilineal el valor de la celda de

²⁷Se puede descargar a continuación desde el visualizador de la IDE: link

²⁸Por más información: link

²⁹Es un objeto de varias capas (o multibanda). También podría utilizarse un *Raster Stack*, pero el tiempo de procesamiento de estos suele ser más largo

³⁰Otro posible procedimiento es inicialmente unir todas las capas de los rasters (hacer un *brick*), para posteriormente extraer la información en los puntos. No se procede de esta forma dado que implica mayor trabajo, la construcción del raster *brick* es uno de los procedimientos más costosos computacionalmente, y cuanto más capas de información contiene es más difícil hacer la superposición.

³¹*method='bilinear, sp=TRUE*

salida es el promedio lineal ponderado por la distancia de los cuatro valores de celda de entrada más cercanos.

A modo de comentario, además del procedimiento de extracción anteriormente señalado se realizó otro con el fin de comparar si existían grandes diferencias entre ambas estrategias. Este fue mediante un buffer ³² de 50 y la utilización de la media como medida de resumen³³, mediante la cual se extrae la información en los puntos de interés (utilizando los píxeles vecinos). Los dos métodos de extracción de información de *rasters* en los puntos generaron resultados similares, tanto en el valor de las variables, como consecuentemente en el resultado del análisis descriptivo y la estimación de modelos, por lo tanto, se optó por presentar únicamente los resultados del primero.

Limitaciones de los datos utilizados:

Las limitaciones de los datos se deben en la mayoría de los casos a errores de medición en el relevamiento del IFN. Se pueden tener:

- Sustitución de puntos en lugares de mayor accesibilidad.
- Sesgos en las mediciones de AB asociado a que lo hicieron diferentes personas según la zona.
- Sesgos en las mediciones de AB producidos por diferentes momentos de relevamiento de campo. Existe un intervalo intertemporal de 7 años entre que se tomaron los primeros datos (2009) y los últimos (2016). Esto puede llevar a pensar que en ese tiempo podría haber rotatividad en los equipos y metodologías de trabajo (aspecto que desconocemos), o incluso si se mantienen los equipos, las personas podrían mejorar su método de relevamiento con la experiencia, introduciendo también un sesgo.

Por último, se destaca que el ciclo de crecimiento del Bosque Nativo es muy largo, mucho más que el de otras plantaciones, motivo por el cual también resulta extremadamente importante su preservación (si se tala, demora mucho en crecer). Dicha característica tiene como consecuencia que de un año a otro no se espera que hayan grandes cambios en el BN, por lo que no resulta ser una limitación importante el hecho de que el relevamiento de información del IFN se dio entre el 2009 y 2016.

³²En el ámbito de los SIG, un buffer, o también conocido como zona de influencia, es el polígono que encierra el área de dominio resultante de dar una determinada distancia entorno a un objeto geográfico (una circunferencia generalmente).

³³*buffer=50, fun=mean, sp=TRUE*

5. Resultados

5.1. Análisis exploratorio de los datos construidos

La primer aproximación a la caracterización de las variables en el espacio puede realizarse mediante el análisis exploratorio espacial visualizando en diferentes gráficos y mapas el comportamiento de las variables bajo estudio. En la presente sección se analiza la información disponible en los puntos de interés, que surge del relevamiento de campo reflejado en el IFN y de la extracción de información de los rasters, tal como se describió en secciones anteriores.

A continuación se presenta una Tabla con medidas de resumen de las variables utilizadas, obtenidas para un área de 900 m² (correspondiente al área de cada píxel de los rasters). Luego se introducen mapas, con la superposición de los puntos a la Cuenca Gondwánica, Bosque Nativo (representado en verde claro) y las cuencas seleccionadas (de color amarillo).

Tal como se observa en la Tabla 3, hay algunas variables que presentan datos faltantes, es decir, un N inferior a 331 que es el total de puntos con información de área basal en la Cuenca. En cuanto al AB, presenta un mínimo de 1.3 metros cuadrados por hectárea y un máximo de 82.5, siendo el promedio 28.4.

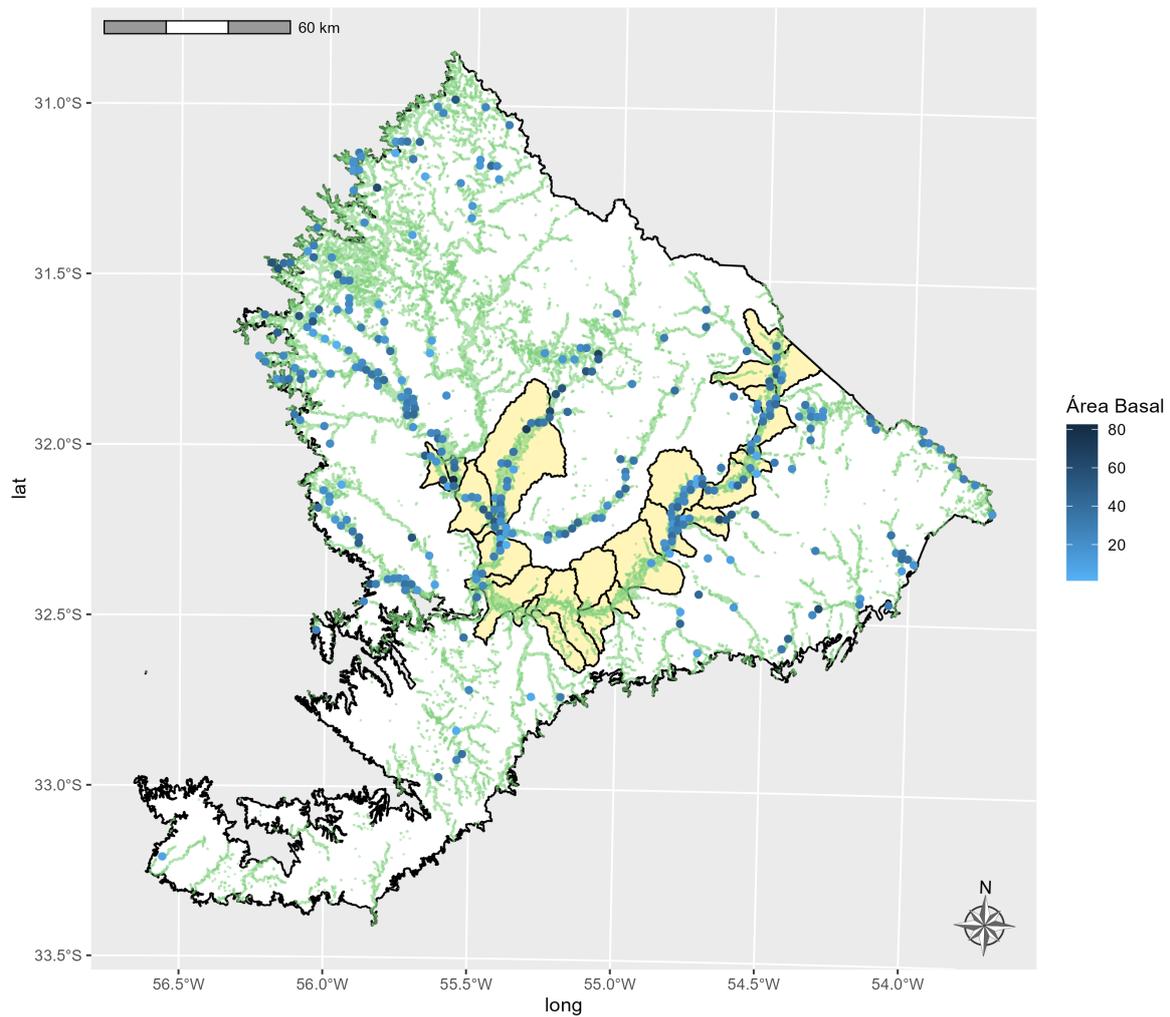
Tabla 3: Medidas de resumen (Puntos del IFN en la Cuenca Gondwánica)

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
AB	331	28.4	13.2	1.3	20.1	36.8	82.5
ETP	273	13.8	3.8	10.5	12.4	12.7	42.8
drenaje	272	20,886.8	69,105.6	4.2	46.5	7,846.7	789,323.8
elevacion	272	122.4	39.1	77.3	94.8	137.0	276.9
pendiente	280	1.6	2.8	0.0	0.2	1.6	21.0
NDVI	277	0.7	0.05	0.3	0.7	0.8	0.8
EVI	277	0.7	0.1	0.3	0.6	0.7	0.8
GI	277	1.6	0.1	1.2	1.5	1.6	1.7
RVI	279	0.2	0.1	0.1	0.1	0.2	1.1
TPI	272	-0.1	0.3	-1.4	-0.1	0.01	1.0
coneat	272	67.3	18.4	18.8	56.9	76.7	126.3
agua	331	239.2	366.4	0	14.3	333.4	2,128
temperatura	272	178.3	2.0	171.4	177.5	179.0	184.0
precipitaciones	272	1,242.7	53.7	1,141.5	1,211.2	1,263.1	1,434.0
B1	331	9,658.9	174.9	9,110.9	9,516.2	9,766.5	10,693.4
B2	331	8,667.8	195.7	8,161.5	8,525.6	8,759.4	10,106.0
B3	331	8,239.3	275.4	7,375.8	8,068.0	8,360.1	10,158.1
B4	331	7,098.5	365.6	6,483.7	6,875.6	7,210.1	10,589.5
B5	331	18,754.6	1,838.3	12,686.5	17,526.9	20,141.1	23,184.1
B6	331	11,603.1	1,179.4	8,603.3	10,966.2	12,009.3	18,079.8
B7	331	7,894.1	733.3	6,480.5	7,445.2	8,146.6	12,420.1
B9	331	5,050.9	9.0	5,022.5	5,045.1	5,057.0	5,075.9
B10	331	27,852.5	973.2	26,051.1	27,245.4	28,364.4	31,332.6
B11	331	25,405.8	684.1	24,091.5	24,933.4	25,747.8	27,813.3

Fuente: Elaboración propia.

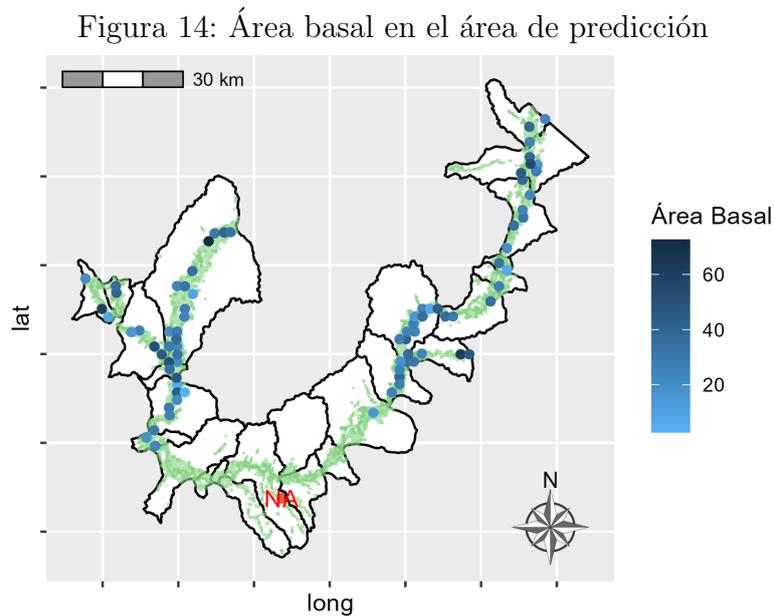
Nota: Tabla construida con el paquete *stargazer*.

Figura 13: Área basal en la Cuenca Gondwánica



Fuente: Elaboración propia.

En la Figura 14 se representa el AB en los puntos del área seleccionada para predicción, se destaca la existencia de un dato faltante en el punto sur (señalado con NA en rojo), que resulta ser el más alejado del resto, dejando una extensa zona sobre la cual no hay información en el IFN.



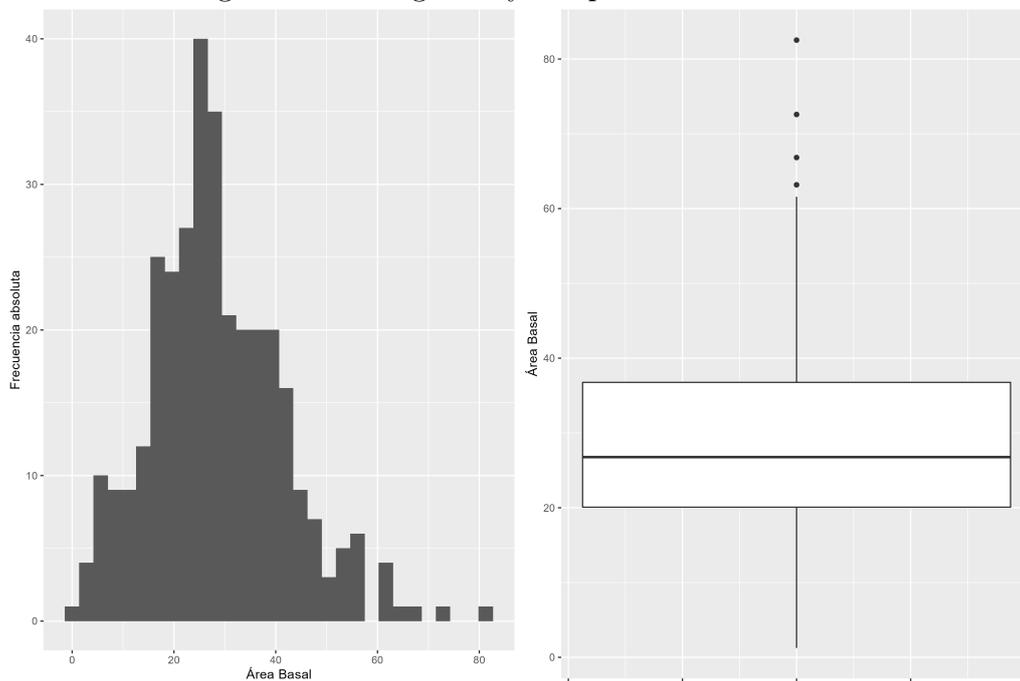
Fuente: Elaboración propia.

Si bien el AB tiene una distribución caracterizada por la presencia de varios valores atípicos cuando se considera todo el territorio nacional, por el contrario, a la hora de analizar la Eco-región se observa que la distribución es más simétrica y existen únicamente 4 valores atípicos, como puede apreciarse en el diagrama de caja de la Figura 15.

A continuación se introduce la Figura 16 extraída de Hernandez, J. (2002), a modo de guía para la interpretación de las magnitudes del área basal en el Bosque Nativo. El promedio de AB en la Cuenca Gondwánica (28.4), se corresponde a una región *Completa*, mientras que el mínimo (1.3) resulta ser *Deficiente*, entre el primer cuartil (20.1) y el tercero (36.8), las regiones se clasifican mayoritariamente como *Completa*, aunque esto depende también del número de árboles por hectárea y el Diámetro Cuadrático Promedio. Por último, el máximo de AB (82.5) corresponde a una región *Sobrepoblada*.

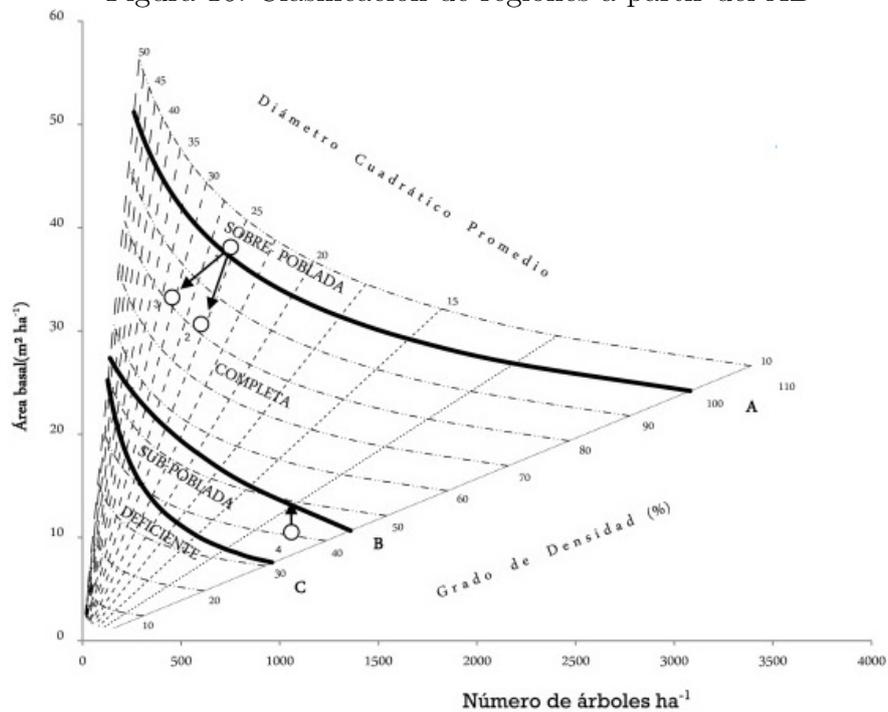
En el Apéndice se encuentran los mapas con la representación de todas las variables disponibles.

Figura 15: Histograma y Boxplot del Área Basal



Fuente: Elaboración propia.

Figura 16: Clasificación de regiones a partir del AB



Fuente: Hernandez, J. (2002) Guía de densidad para manejo de bosques nativos.

5.1.1. Correlogramas

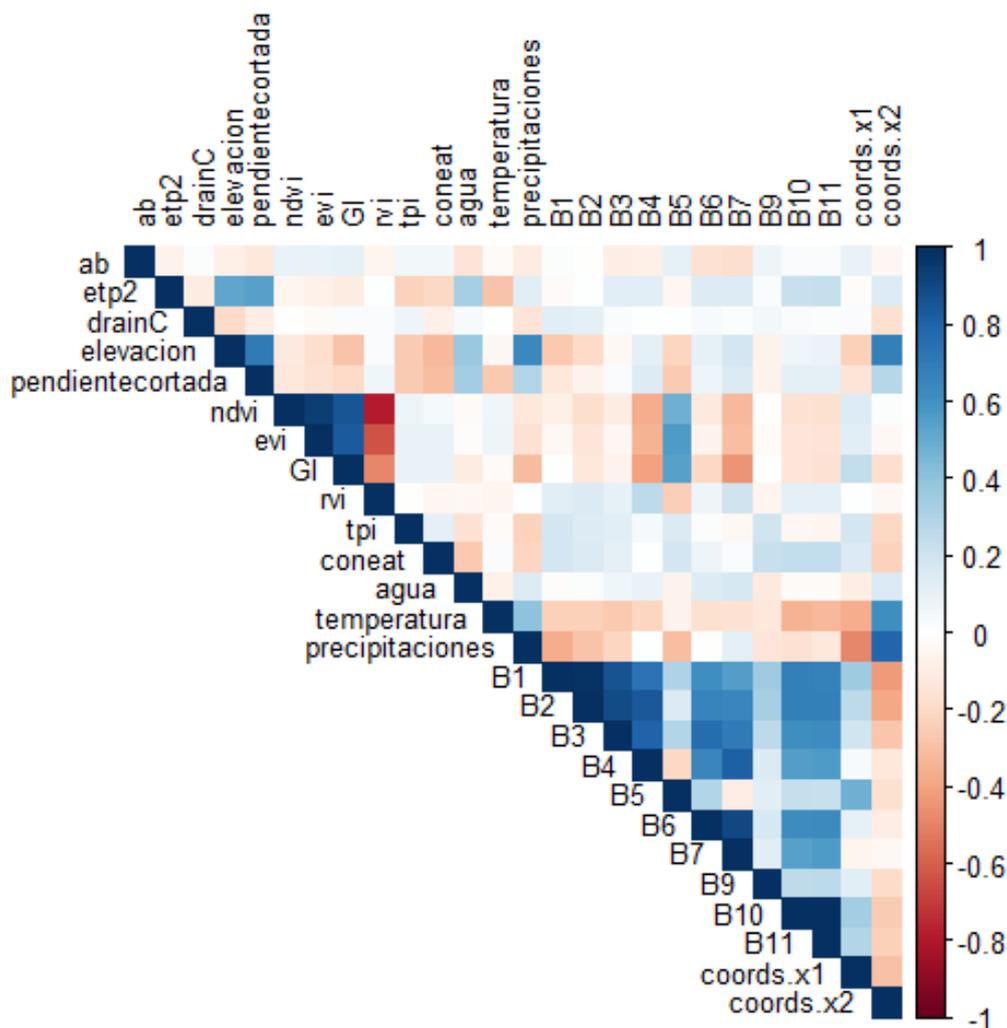
Una de las etapas del análisis geoestadístico es la determinación del vínculo entre la variable de interés (en este caso el AB) y el resto de las variables disponibles, así como también su dependencia espacial. Para ello se emplean los correlogramas y variogramas.

En la Figura 17 se presenta el correlograma de las variables utilizadas en este trabajo. La correlación positiva, implica misma direccionalidad de las variables, es identificada con colores celeste y azul. El celeste corresponde, según la escala colorimétrica que se presenta en el gráfico, a valores bajos de correlación, mientras que azul se corresponde con correlaciones altas. Por otra parte, la correlación negativa, que implica opuesta direccionalidad de las variables, es identificada con colores rosado y rojo. Rosado corresponde a valores bajos de correlación y rojo con valores altos.

Se observa que las correlaciones que el AB tiene con el resto de las variables (primera fila del correlograma de la Figura 17) son bajas, las que presentan mayor correlación son algunas bandas Landsat y la distancia al curso de agua, aunque todas muy cercanas a cero.

Para el resto de las variables es posible determinar la correlación entre ellas en la celda que resulta de la intersección. En particular se observa la alta correlación existente entre los índices de vegetación y las bandas satelitales, dado que se construyen a partir de estas. A su vez, existe alta correlación de las bandas entre sí, también de la ETP con la elevación, la pendiente, la distancia al curso de agua y la temperatura, entre otros. La elevada correlación entre las variables que serán utilizadas como explicativas en los modelos, implica la existencia de multicolinealidad en los mismos, por lo que esto deberá ser tenido en cuenta, analizando el Factor de Inflación de Varianza (VIF) luego de cada estimación.

Figura 17: Correlograma



Fuente: Elaboración propia.

5.1.2. Selección de variables

Se ajusta un modelo de regresión lineal para identificar las variables que explican el AB, inicialmente se utilizan todas las covariables provenientes de los rasters utilizados, para posteriormente proceder a seleccionar las variables de utilidad.

Sea ab_i el valor de la variable dependiente (AB) en el punto i , a continuación se plantea el modelo empírico clásico a estimar inicialmente:

$$\begin{aligned}
ab_i = & \beta_0 + \beta_1 B1_i + \beta_2 B2_i + \beta_3 B3_i + \beta_4 B4_i + \beta_5 B5_i + \beta_6 B6_i \\
& + \beta_7 B7_i + \beta_8 B8_i + \beta_9 B9_i + \beta_{10} B10_i + \beta_{11} B11_i \\
& + \beta_{12} evi_i + \beta_{13} ndvi_i + \beta_{14} gi_i + \beta_{15} rvi_i \\
& + \beta_{16} pendiente_i + \beta_{17} elev_i + \beta_{18} tpi_i + \beta_{19} coneat_i + \beta_{20} drain_i \\
& + \beta_{21} agua_i + \beta_{22} precipitaciones_i + \beta_{23} temperatura_i + \epsilon_i
\end{aligned} \tag{33}$$

Se utilizó un método de eliminación de variables por pasos (*stepwise*) junto con el AIC ($AIC = n \ln(RMSE) + 2p$) para seleccionar solo aquellas variables que mejor explicaban el AB. A su vez, luego de obtener las variables seleccionadas por el algoritmo *stepwise*, se utilizó el factor de inflación de la varianza (VIF) para detectar posibles correlaciones entre variables explicativas (multicolinealidad). El valor de corte VIF adoptado fue 10 (dos Reis y cols., 2018), lo que implicó el descarte de 2 variables (individualmente, repitiendo el proceso, ambas tenían un VIF de 61 y 57, implicando alta multicolinealidad). En la Tabla 4 se presentan los resultados del modelo seleccionado, el mayor VIF en dicho modelo es de 7.

Tal como se puede observar en la Tabla 4, las variables espectrales que más contribuyeron al modelo estimado fueron las bandas 2, 3, 5 y 6 de Landsat, a su vez, en el correlograma se observó que estas presentaron mejores valores de correlación con el AB, esto puede deberse a que la variable de interés se deriva solo del diámetro de los árboles (DAP), que está directamente relacionado con el tamaño de las copas de los árboles que determina la reflectancia de la copa, la cual es captada por las bandas (dos Reis y cols., 2018).

Por otra parte, también quedaron seleccionadas en el modelo las variables elevación y pendiente (extraídas del Modelo Digital de Terreno), la distancia al curso de agua y las precipitaciones. Se destaca que dichas variables son determinantes del tipo de bosque que se desarrolla en un lugar particular, por lo que una posible interpretación es que lo que en realidad explica la variabilidad del área basal es el tipo de bosque. Por ejemplo, a 50 mts del curso de agua, con determinada elevación y pendiente, se tiene seguramente determinado desarrollo de monte ribereño, también ocurre que las quebradas tienen unas características muy distintas al resto de los tipos de bosque en términos de las variables seleccionadas. Tal como se identifica claramente en la Figura 1, resulta coherente pensar que el comportamiento del área basal en un Bosque Ribereño resulta diferente a la de uno de Parque, Quebrada o Palmar, lo que le da coherencia a los resultados del modelo seleccionado.

Tabla 4: MLR para predecir el AB en la Cuenca Gondwánica.

	<i>Variable dependiente:</i>
	AB
elevacion	0.1** (0.04)
pendiente	-0.6 (0.4)
agua	-0.00* (0.00)
precipitaciones	-0.03 (0.02)
B2	0.04*** (0.01)
B3	-0.03*** (0.01)
B5	0.00*** (0.00)
B6	-0.00** (0.00)
Constant	-48.4 (63.0)
Observations	253
R ²	0.1
Adjusted R ²	0.1
Residual Std. Error	12.1 (df = 244)
F Statistic	4.6*** (df = 8; 244)

Note: *p<0.1; **p<0.05; ***p<0.01

Nota: Tabla construida con el paquete *stargazer*.

La prueba de significación conjunta tiene como resultado que se rechaza la igualdad de todos los coeficientes a cero, por lo que conjuntamente las variables son significativas para explicar el AB. A su vez, se observa que varias variables son significativas individualmente. Sin embargo, el R^2 ajustado resulta bajo, solo un 10 % de la variabilidad del AB es explicada por las variables independientes, esto impactará sobre las predicciones realizadas a partir del modelo KED, donde se utilizarán las variables seleccionadas mediante este procedimiento.

La principal diferencia entre modelos clásicos como la regresión lineal y los modelos espaciales se basa en el tratamiento e incorporación a estos últimos del efecto del espacio, los modelos tradicionales tienden a ignorarlos y en caso de que exista un vínculo con el espacio se violan los supuestos de Gauss-Markov. Por un lado, la dependencia espacial genera el incumplimiento del supuesto de que en un muestreo repetitivo las variables explicativas son fijas e independientes

y, por tanto, ni estas ni sus residuos seguirán una distribución normal e independiente. Por el otro, la heterogeneidad espacial viola el supuesto de que existe una sola relación lineal a lo largo de observaciones de una muestra de datos, lo cual sugiere problemas de heterocedasticidad (Pérez-Pineda, 2006).

Los residuos de los modelos de regresión fueron analizados para evaluar el cumplimiento de los supuestos del modelo y la existencia de tendencias en los errores. Se realiza el contraste de normalidad Shapiro y Jaque Bera aplicado a los residuos, el pvalor de los contrastes es respectivamente 0.02 y 0.03, por lo que a un 1% de significación no se rechaza la normalidad de los residuos, mientras que considerando un 5% se rechaza.

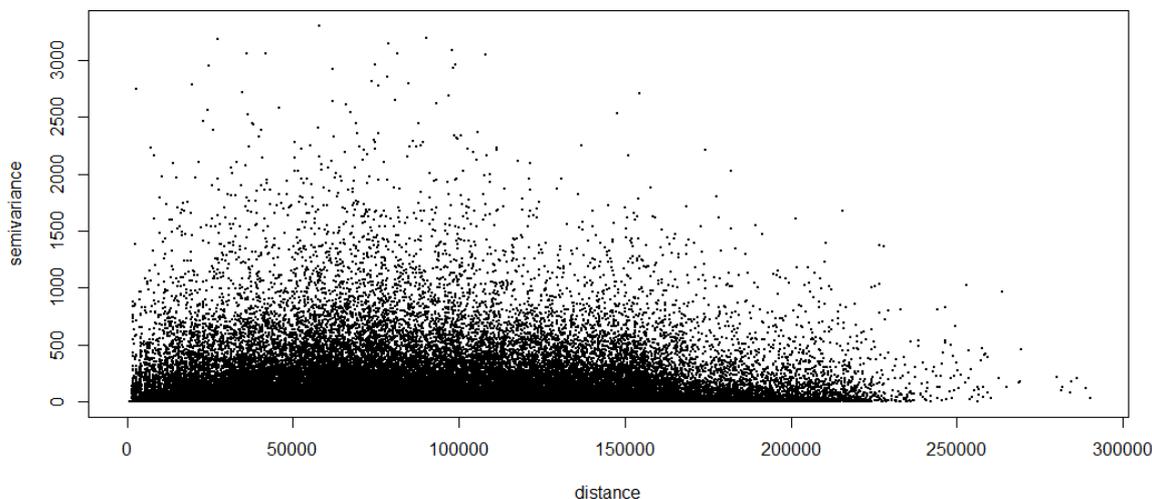
5.2. Variogramas

La presente etapa del análisis consiste en calcular un variograma experimental a partir de los datos disponibles y posteriormente ajustarle un modelo. Dicha herramienta se utiliza en geoestadística para representar la autocorrelación espacial y describir el grado en que las observaciones cercanas tienen valores similares.

5.2.1. Variograma Omnidireccional

Al calcular el cuadrado de las diferencias entre cada pareja de puntos (dispersión de pares de datos), y representarlo frente a la distancia entre puntos, se obtiene el siguiente gráfico de nube variográfica. Para obtener una estimación mas clara se suelen promediar estos valores en grupos, obteniendo el variograma experimental que se muestra en la Figura 19.

Figura 18: Nube variográfica del AB (Cuenca Gondwánica, distancia en metros).



Fuente: Elaboración propia.

Nota: Elaborado en base a los 331 puntos de la Cuenca Gondwánica.

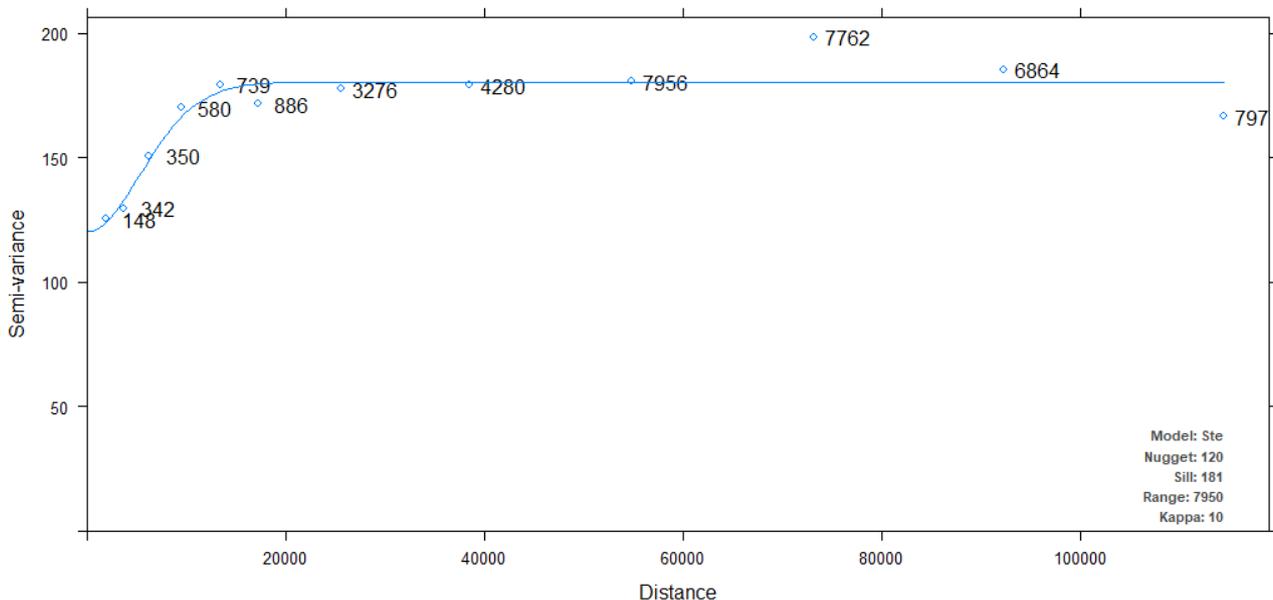
En cuanto a la interpretación de una nube variográfica, en la esquina superior izquierda se encuentran los pares de puntos que están muy juntos, pero tienen valores muy diferentes. En la esquina inferior izquierda están los puntos muy cercanos entre sí, tanto en valor como en distancia. En la superior derecha, están los puntos muy separados, tanto en valor como en distancia y en la inferior derecha los valores similares pero muy separados en distancia.

En la Figura 19 se presenta el variograma experimental y el modelo ajustado utilizando únicamente el AB, es el ajuste que será utilizado posteriormente en la aplicación de Kriging Ordinario. La evaluación del desempeño de cada modelo de semivariograma y la selección de los mejores modelos se basó en validación cruzada. Los puntos del gráfico muestran el variograma empírico, los números indican el número de pares con el que se calculó el promedio y la línea continua es el modelo de variograma ajustado. En este caso, el modelo seleccionado fue Matern, la parametrización de M. Stein (nombre en clave "Ste"), con los cuatro valores de los parámetros indicados en la esquina inferior derecha de la imagen.

En cuanto a los resultados del ajuste al variograma, la *nugget* es de 120 y la varianza de la variable es de 174, por lo que el cociente entre ambos es de 69 %, siendo este porcentaje de varianza no explicada, esto significa que el componente espacial no es muy fuerte, aunque no resulta despreciable (31 %).

Se observa que el variograma es estacionario, es decir, se estabilizan los puntos, se vuelven horizontales, y el rango no es muy elevado (7950). A su vez, existe efecto *pepita*, dado que el variograma no comienza en el origen, por lo que el proceso es muy irregular, esto puede deberse al error de medida de la variable. La presencia de atípicos afectan los valores de los parámetros.

Figura 19: Variograma experimental y modelo estimado (utilizando únicamente el AB).

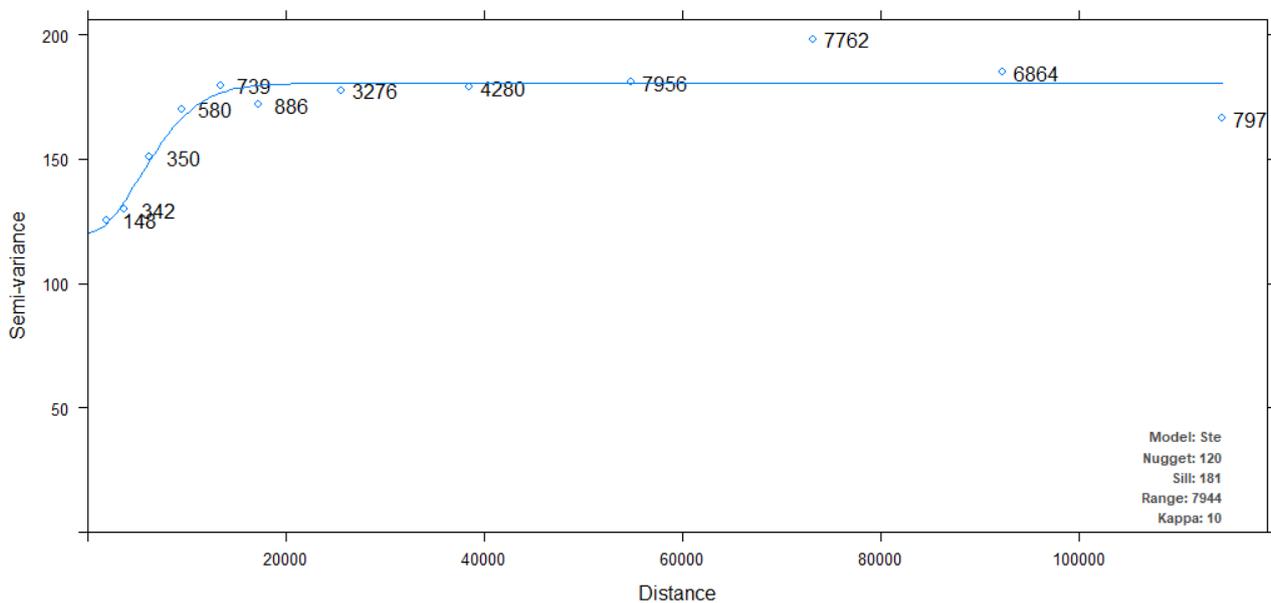


Fuente: Elaboración propia.

Nota: Elaborado en base a los 331 puntos de la Cuenca Gondwánica.

En la Figura 20 se presenta el variograma experimental y el modelo ajustado incorporando las coordenadas, es el ajuste que será utilizado posteriormente en la aplicación de Kriging Universal. En cuanto a los resultados del ajuste, el modelo seleccionado es nuevamente el *Ste*, con *kappa* igual a 10. La *nugget* es de 120, por lo que el porcentaje de varianza no explicada es de 69%, al igual que en el OK. El 31% de la varianza es explicada, siendo un valor no despreciable y que valida la implementación de interpolación espacial a partir de UK.

Figura 20: Variograma experimental y modelo estimado (incorporando las coordenadas).

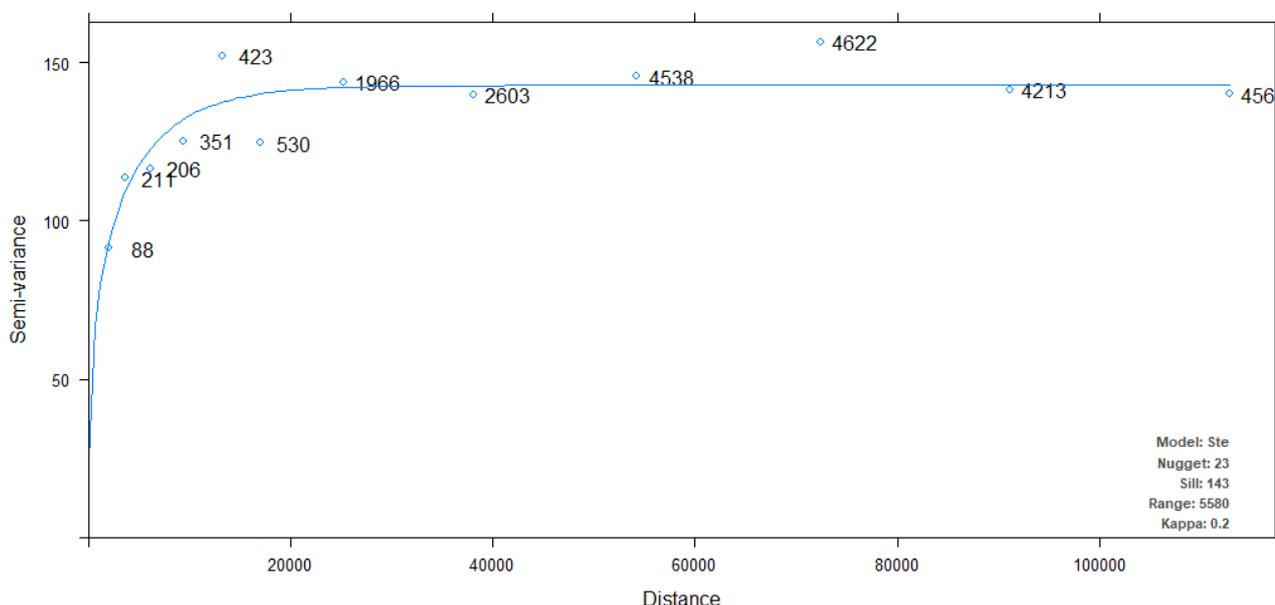


Fuente: Elaboración propia.

Nota: Elaborado en base a los 331 puntos de la Cuenca Gondwánica.

En la Figura 21 se presenta el variograma experimental y el modelo ajustado incorporando las variables seleccionadas en la sección anterior, es el ajuste que será utilizado posteriormente en la aplicación del KED. El modelo de ajuste al variograma es nuevamente el *Ste*, con *kappa* igual a 0.2. La *nugget* es de 23, por lo que el porcentaje de varianza no explicada es de 13.2%, implicando una mejora respecto a los métodos anteriores (OK y UK). A su vez, el rango es de 5580.

Figura 21: Variograma experimental y modelo estimado (incorporando variables seleccionadas).



Fuente: Elaboración propia.

Nota: Elaborado en base a los 253 puntos de la Cuenca Gondwánica (excluidos los datos faltantes de las variables involucradas).

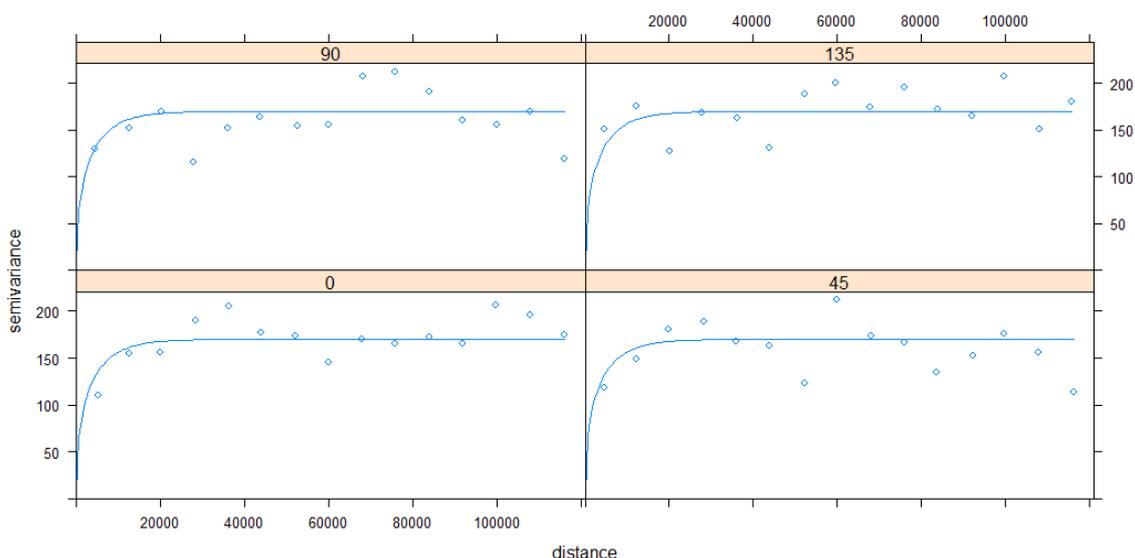
5.2.2. Variograma Direccional

A continuación se analiza la anisotropía, es decir, la variación en la continuidad espacial del AB en relación con la dirección³⁴, buscando identificar si existe una dirección predominante. Se presentan semivariogramas empíricos direccionales para el AB, que incluyen solo los valores de los pares de datos que se encuentran dentro de ciertas bandas direccionales, con el objetivo de identificar la dirección principal de anisotropía o de mayor continuidad espacial en la variable.

En la Figura 22, cada gráfico corresponde a una determinada banda direccional, es importante tener en cuenta que los semivariogramas direccionales generalmente son más ruidosos debido al número reducido de pares de datos utilizados para la estimación. Los resultados no muestran evidencia de anisotropía. Se agrega el modelo de variograma ajustado anteriormente, para analizar si se ajusta a los datos en las cuatro bandas direccionales.

³⁴Con el paquete *gstat* se puede dar cuenta de la anisotropía especificando bandas direccionales, la función `variograma()` acepta el parámetro `alfa` para indicar la dirección en el plano, en grados positivos en el sentido de las agujas del reloj.

Figura 22: Variograma Direccional (Cuenca Gondwánica, distancia en metros).



Fuente: Elaboración propia

Nota: Elaborado en base a los 331 puntos de la Cuenca Gondwánica.

5.3. Kriging

Una vez escogido el modelo de variograma, será utilizado para la generación del mapa de predicción del AB y de los desvíos estándar de los errores predichos. De esta manera se obtiene un mapa con la estimación de la variable en el área de interés. Para ello, previamente se debe generar una grilla a partir de un raster³⁵ que abarque la superficie en la que se desea predecir, en este caso, la selección de cuencas. Una vez construida la misma, y utilizando la información proveniente del modelo de variograma, mediante la interpolación por Kriging se logra estimar cada punto de la grilla.

5.3.1. Kriging Ordinario

En la presente sección, a partir de la aplicación de Kriging Ordinario se obtiene la predicción de área basal y el desvío estándar de la misma, el modelo de variograma utilizado fue presentado en la Figura 19. A continuación se presenta un resumen de los resultados (Tabla 5) y se mapean las predicciones (Figura 23) y el desvío estándar de las estimaciones (Figura 24).

³⁵La forma de hacer esto es utilizando un raster recortado por el polígono donde se desee predecir, en este caso la selección de cuencas, y aplicarle una transformación del tipo: *as(RasterRecortado,"SpatialGridDataFrame")*. De esta forma se obtiene una grilla con celdas del mismo tamaño que el raster, en este caso 30x30.

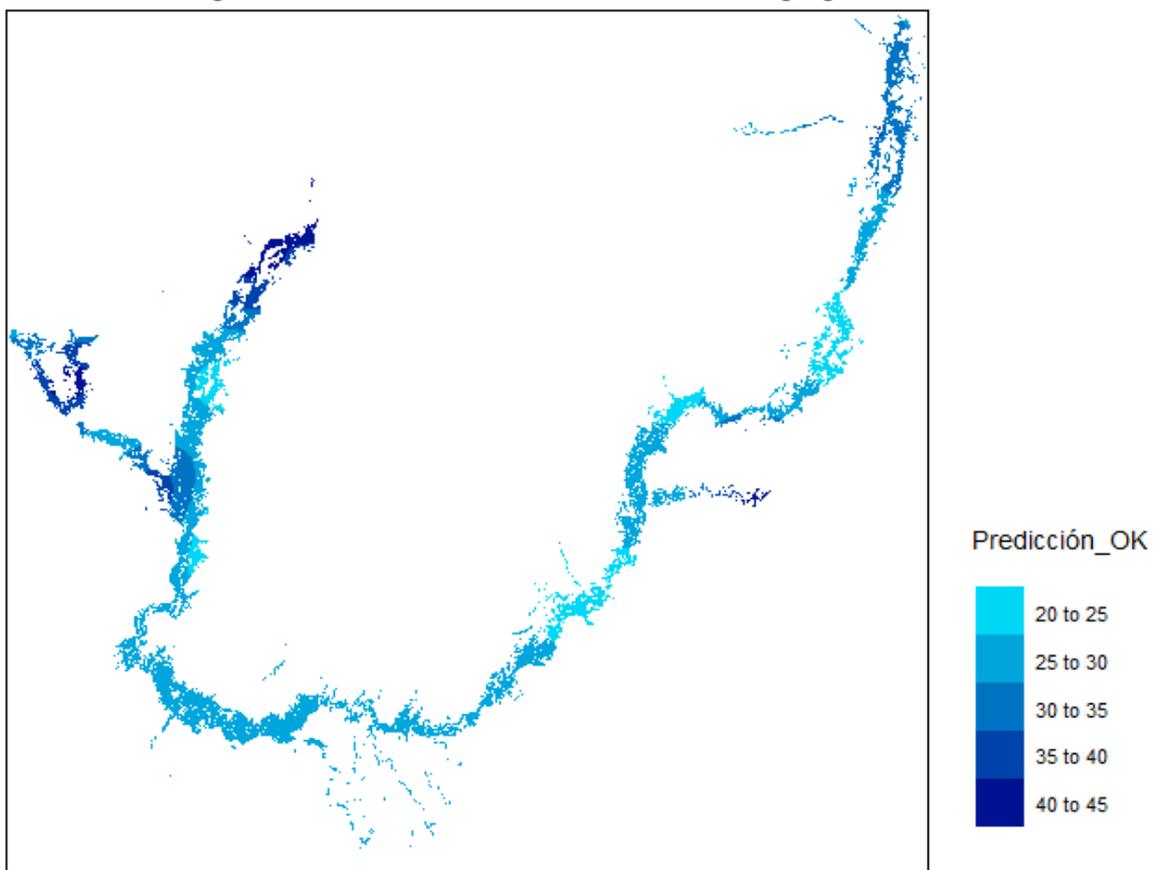
Tabla 5: Medidas de resumen de la predicción por Kriging Ordinario

	Predicción OK	Desvío Estándar OK
Min.	20,20	11,48
1st Qu.	26,35	11,86
Median	27,99	12,06
3rd Qu.	30,40	12,93
Max.	43,71	13,48

Nota: Elaborado en base a los 331 puntos de la Cuenca Gondwánica.

El mínimo de área basal predicha es de 20.2 y el máximo es de 43.7, el desvío estándar se encuentra entre 11.5 y 13.5. En la Figura 23 y 24 se observan las predicciones y desvíos en el mapa, las zonas al norte son las que se estima presentan un promedio mayor de área basal y un menor desvío estándar, mientras que las del sur presentan promedios inferiores y desvíos estándar elevados.

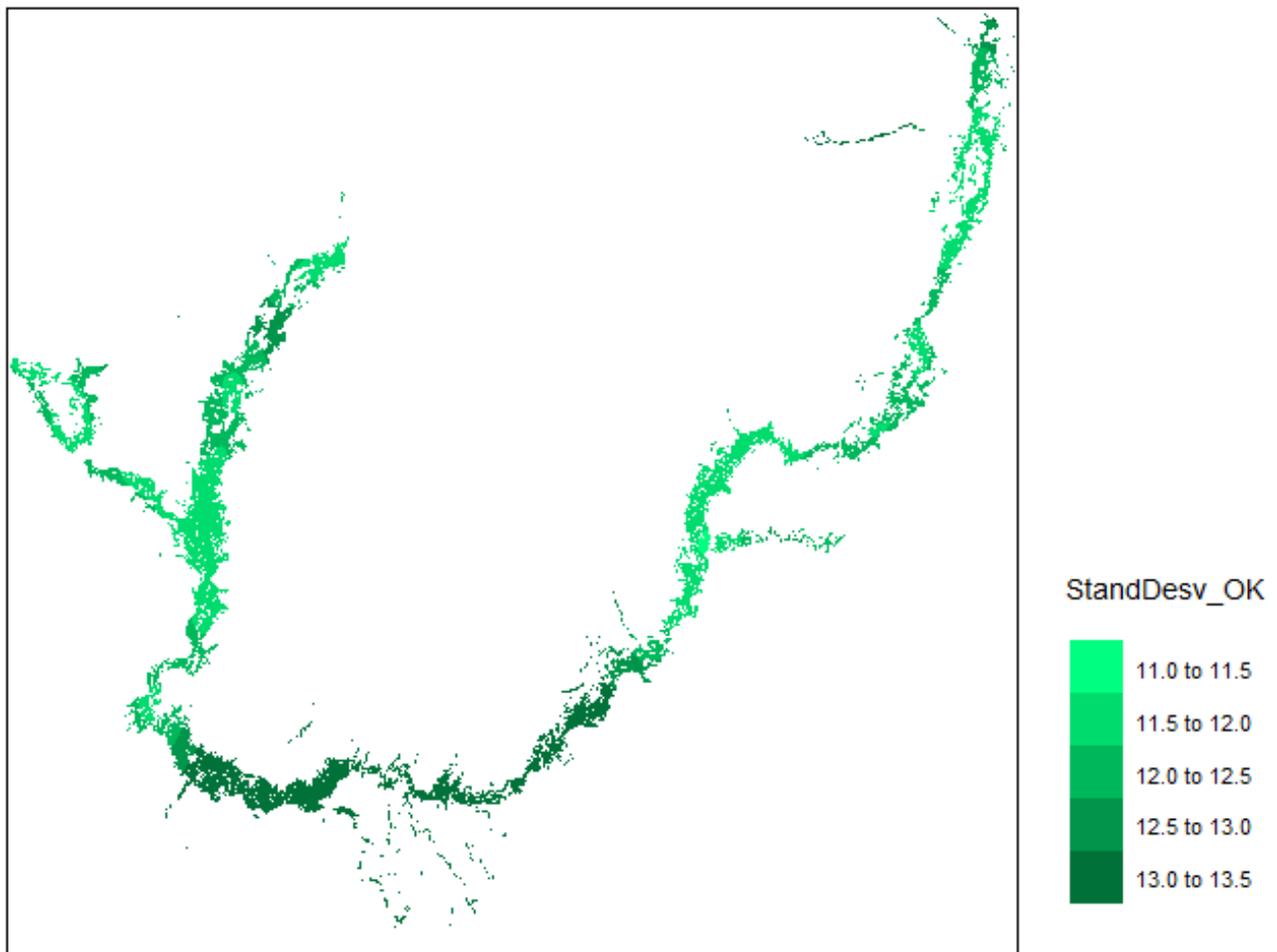
Figura 23: Predicción de área basal con Kriging Ordinario



Fuente: Elaboración propia.

Nota: Elaborado en base a los 331 puntos de la Cuenca Gondwánica.

Figura 24: Desvío estándar del Kriging Ordinario

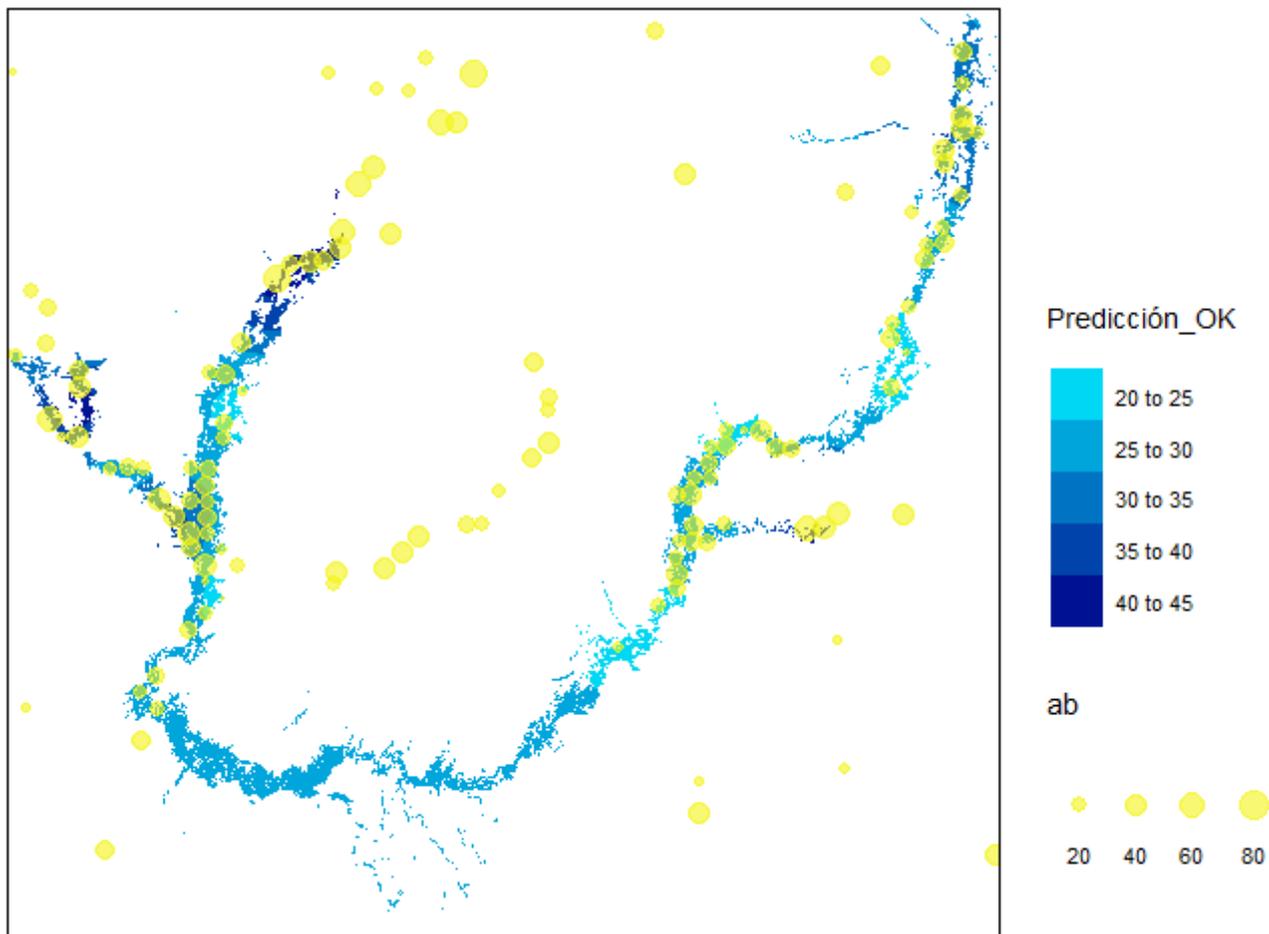


Fuente: Elaboración propia.

Nota: Elaborado en base a los 331 puntos de la Cuenca Gondwánica.

En la Figura 25 se le superpone los datos del área basal relevado en el IFN a la Figura 23 vinculada a las predicciones, estos fueron utilizados para la realización de las estimaciones. Se observa que la zona sur que presenta las predicciones más pequeñas de área basal, no cuenta con información en el IFN de puntos, motivo por el cuál presenta mayor desvío estándar. A su vez, la zona norte es la que presenta mayor área basal observada (representada con círculos de mayor tamaño), por lo que coincide con ser la de mayores predicciones.

Figura 25: Predicción de AB con OK y variable relevada (círculos de tamaño proporcional al AB en campo)



Fuente: Elaboración propia.

Nota: Elaborado en base a los 331 puntos de la Cuenca Gondwánica.

5.3.2. Kriging Universal

En la presente sección, a partir de la aplicación de Kriging Universal se obtiene la predicción de área basal y el desvío estándar de la misma, el modelo de variograma utilizado fue presentado en la Figura 20. En la Tabla 6 se presenta un resumen de los resultados y en la figura 26 se mapean las predicciones y en la 27 el desvío estándar de las estimaciones.

Para poder realizar la predicción mediante UK se deben agregar las coordenadas X e Y al raster, dado que el modelo es $z = \mu + x + y + \varepsilon$. Para ello primero se aplica la función $xyFromCell()$ a la grilla generada, obteniendo las coordenadas de los centroides de cada píxel, estando en condiciones de crear un nuevo *rasterBrick* que contenga las coordenadas para poder aplicar UK.

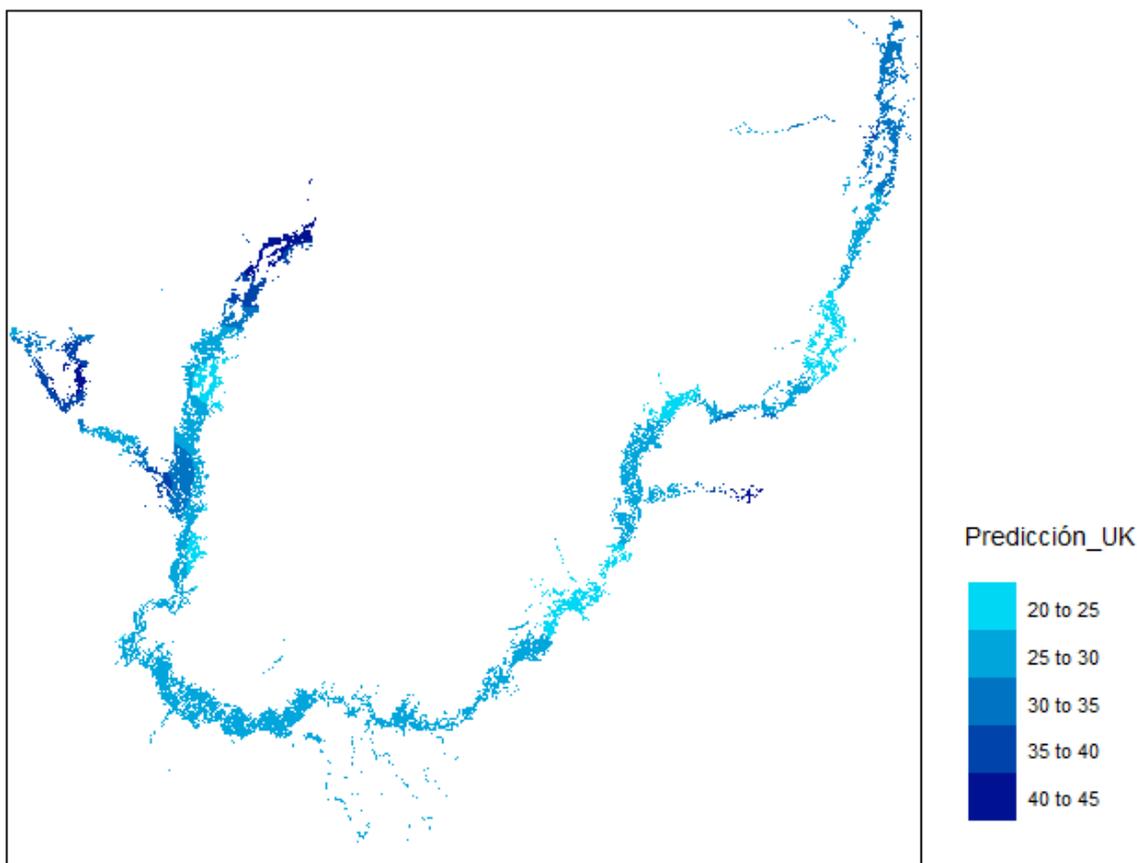
Tabla 6: Medidas de resumen de la predicción por Kriging Universal

	Predicción UK	Desvío Estándar UK
Min.	20,22	11,48
1st Qu.	26,34	11,85
Median	27,69	12,05
3rd Qu.	30,61	12,94
Max.	43,73	13,58

Nota: Elaborado en base a los 331 puntos de la Cuenca Gondwánica.

El mínimo de área basal predicha es de 20.2 y el máximo es de 43.7, ambas estimaciones coinciden con las de OK considerando un dígito después de la coma, el desvío estándar se encuentra entre 11.5 y 13.6. En la Figura 26 se observan las predicciones y en la 27 los desvíos, al igual que se observó para el caso de OK, las zonas al norte son las que se estima presentan un promedio mayor de área basal y un menor desvío estándar, mientras que las del sur presentan promedios inferiores y desvíos estándar elevados.

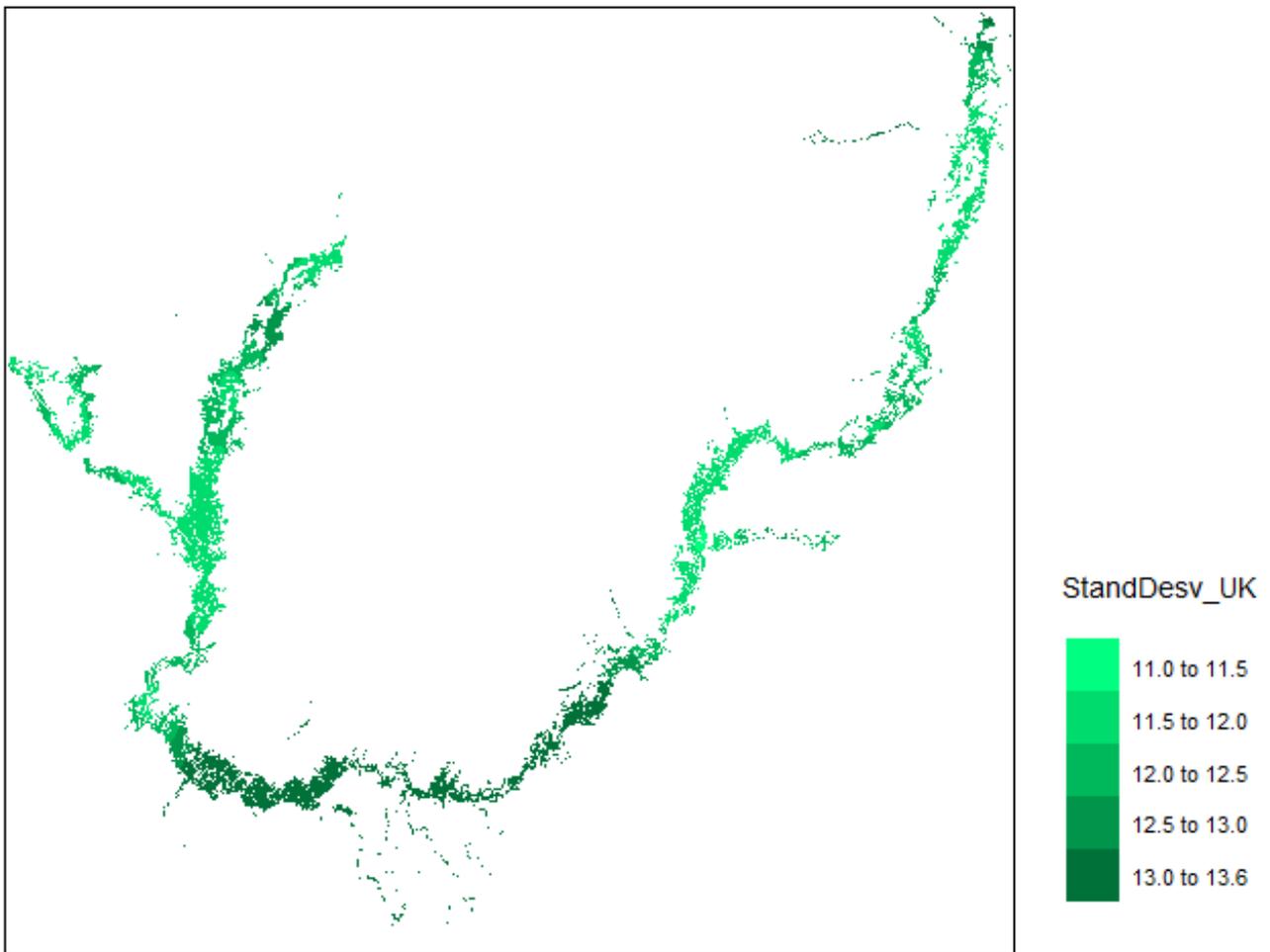
Figura 26: Predicción de área basal con Kriging Universal



Fuente: Elaboración propia.

Nota: Elaborado en base a los 331 puntos de la Cuenca Gondwánica.

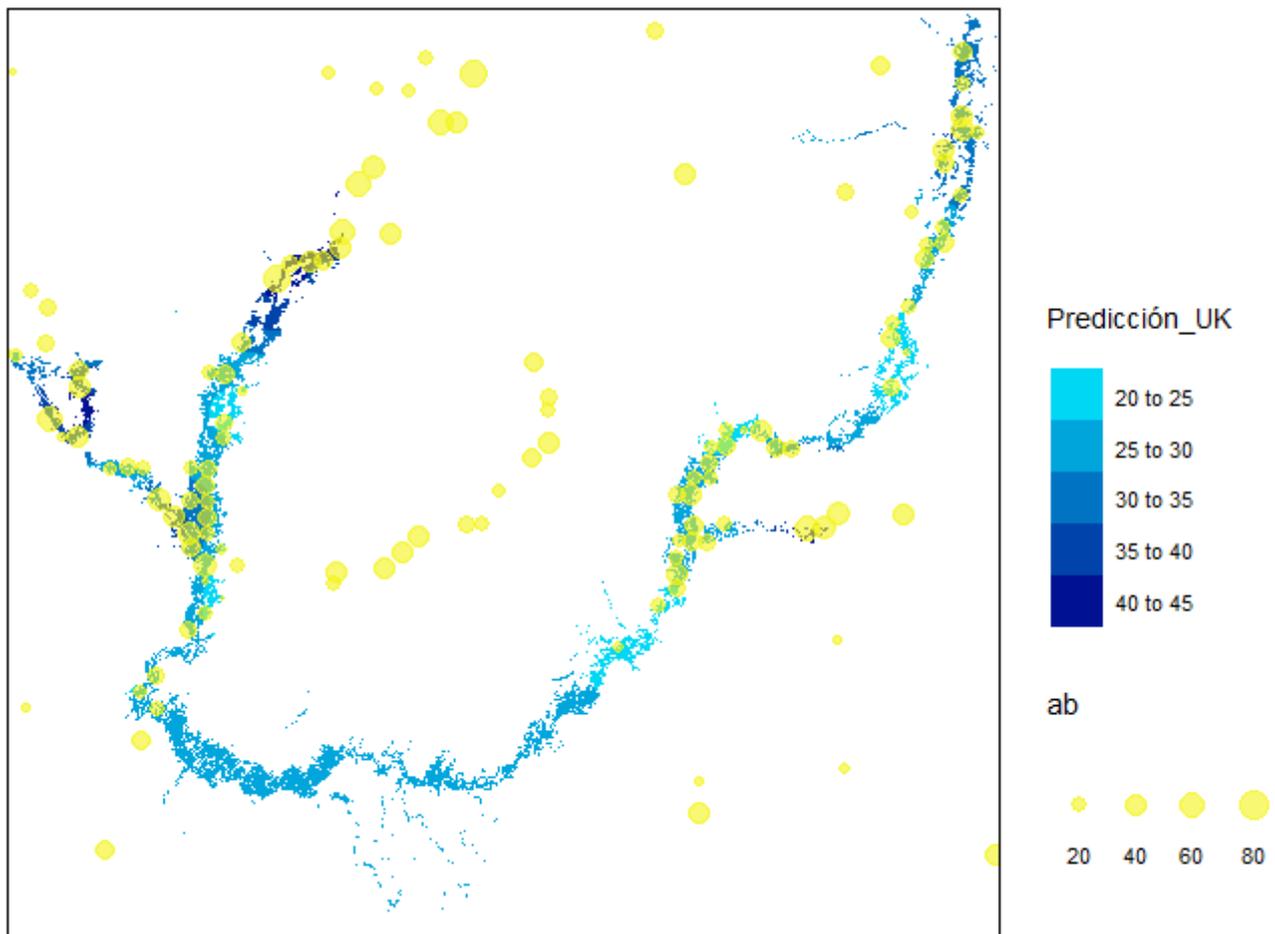
Figura 27: Desvío estándar del Kriging Universal



Fuente: Elaboración propia.

Nota: Elaborado en base a los 331 puntos de la Cuenca Gondwánica.

Figura 28: Predicción de AB con UK y variable relevada (círculos de tamaño proporcional al AB en campo)



Fuente: Elaboración propia.

Nota: Elaborado en base a los 331 puntos de la Cuenca Gondwánica.

5.3.3. Kriging Universal con Deriva Externa

En la presente sección, a partir de la aplicación de *KED* se obtiene la predicción de área basal y el desvío estándar de la misma, el modelo de variograma utilizado fue presentado en la Figura 21. A continuación se introduce un resumen de los resultados (Tabla 7) y se mapean las predicciones (Figura 29) y el desvío estándar de las estimaciones (Figura 30).

Tabla 7: Medidas de resumen de la predicción por KED

	Predicción KED	Desvío Estándar KED
Min.	-12,70	6,72
1st Qu.	24,96	10,10
Median	29,25	10,81
3rd Qu.	33,33	11,87
Max.	132,39	43,54

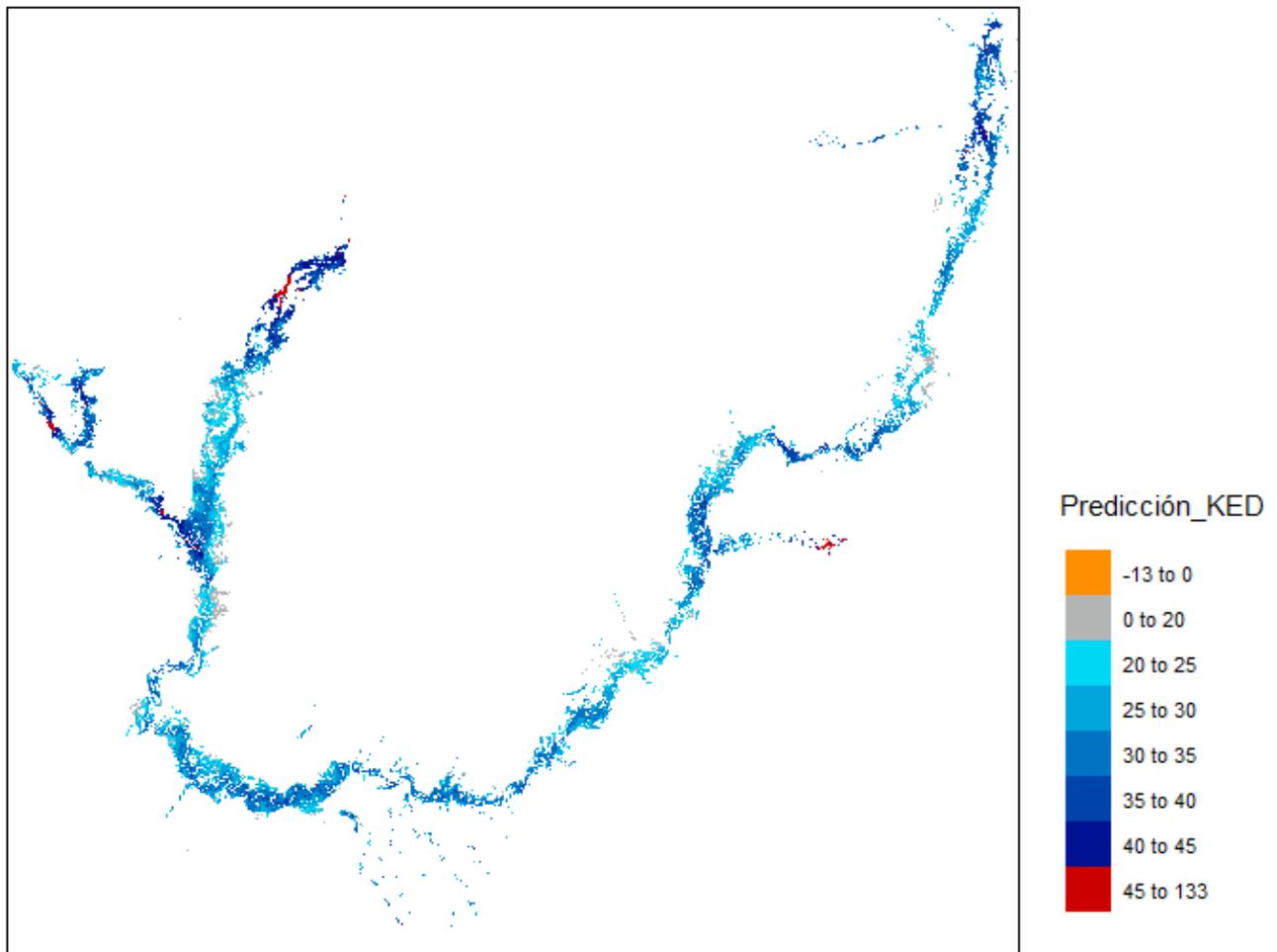
Nota: Elaborado en base a los 253 puntos de la Cuenca Gondwánica (excluidos los datos faltantes de las variables involucradas).

En las Figuras 29 y 30 se observan las predicciones y desvíos en el mapa, al igual que se observó para el caso de OK y UK, las zonas al norte son las que se estima presentan un promedio mayor de área basal y un menor desvío estándar, mientras que las del sur presentan promedios inferiores y desvíos estándar elevados. No obstante, las estimaciones a partir de *KED* en el sur tienden a ser más elevadas que las obtenidas con *OK* y *UK*, lo que se asocia a la mayor incorporación de información en las estimaciones (las variables utilizadas son elevación, pendiente, distancia al curso de agua, precipitaciones, bandas Landsat 2, 3, 5 y 6). A su vez, el desvío estándar en la zona norte es más pequeño en los resultados del *KED* respecto a los otros kriging, lo cual también puede asociarse al mismo motivo.

Por otra parte, se observa que el *KED* genera algunas predicciones atípicas, incluso inferiores a cero (solamente en 9 píxeles³⁶, siendo imperceptible en la visualización del mapa), también superiores al máximo observado en la información de campo relevada. En las figuras presentadas se utilizan los mismos intervalos y colores para representar los resultados de los 3 modelos de kriging, con el fin de que sean comparables.

³⁶Las predicciones negativas se dan en solo 9 píxeles y toman los siguientes valores: -12.7, -4.8, -3.4, -2.1, -1.6, -1.5, -1.2, -0.33, -0.30.

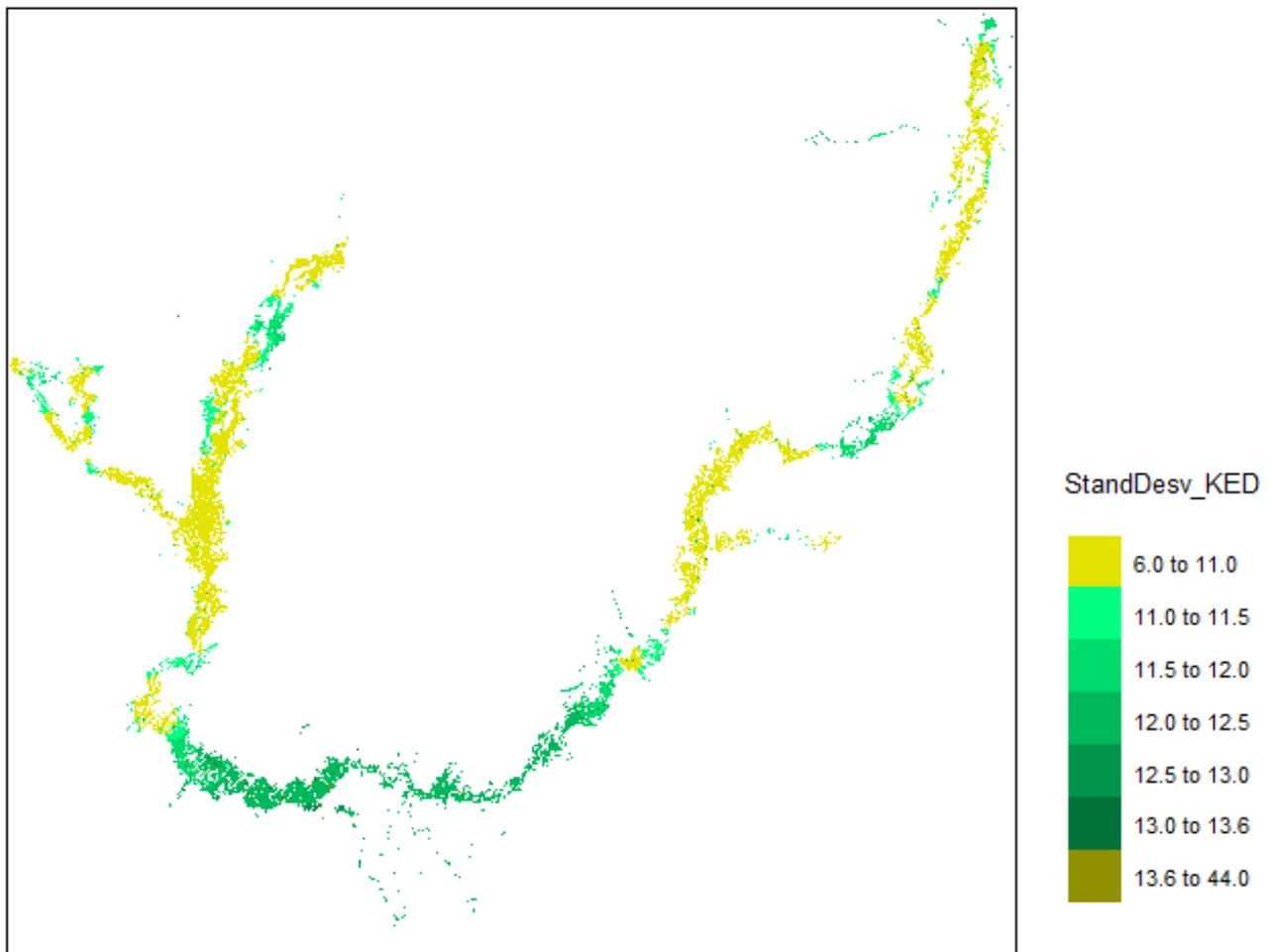
Figura 29: Predicción de área basal con Kriging Universal con Deriva Externa



Fuente: Elaboración propia.

Nota: Elaborado en base a los 253 puntos de la Cuenca Gondwánica (excluidos los datos faltantes de las variables involucradas).

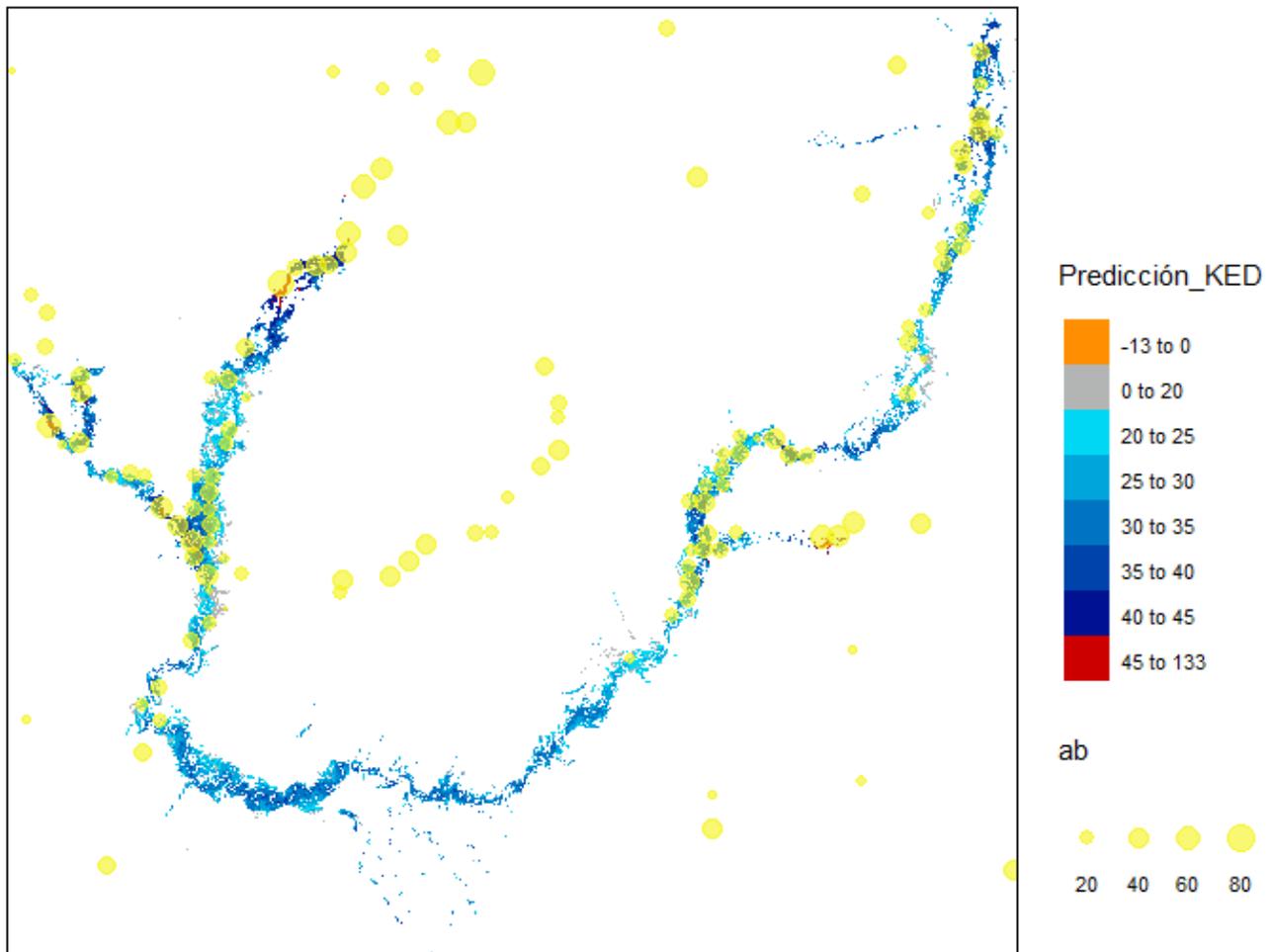
Figura 30: Desvío estándar del Kriging Universal con Deriva Externa



Fuente: Elaboración propia.

Nota: Elaborado en base a los 253 puntos de la Cuenca Gondwánica (excluidos los datos faltantes de las variables involucradas).

Figura 31: Predicción de AB con UK y variable relevada (círculos de tamaño proporcional al AB en campo)



Fuente: Elaboración propia.

Nota: Elaborado en base a los 253 puntos de la Cuenca Gondwánica (excluidos los datos faltantes de las variables involucradas).

5.3.4. Diagnóstico mediante validación cruzada y comparación de resultados

Si la variable de interés presenta autocorrelación espacial y el modelo de variograma elegido la describe bien, el comportamiento de la misma presentará un buen ajuste al modelo propuesto y esto quedará manifiesto en el hecho de que las diferencias entre los valores observados y predichos por el modelo deberán ser pequeñas en relación a lo observado. Resulta necesario aplicar un proceso de validación cruzada para detectar qué tan buenas son las predicciones, así como también seleccionar el modelo que provee estimaciones más exactas.

Se utiliza la técnica de validación cruzada “leave-one-out” (LOOCV), tal como fue explicado en el marco teórico, consiste en remover una observación de los n puntos muestrales, y luego estimar con los $n - 1$ valores restantes y el modelo de variograma seleccionado. Este procedimiento se realiza en forma secuencial con cada uno de los puntos muestreados y así se obtiene

un conjunto de n errores de predicción.

En este marco, los diferentes modelos para la estimación del AB se evaluaron comparando los valores observados y predichos en los conjuntos de datos de ajuste y validación. Las medidas que se usaron para estas comparaciones son: la media del residuo de validación cruzada o Error Medio (ME), que idealmente es pequeño; el ME dividido por la media de los valores observados (ME_mean), para ver qué tan grande es el ME respecto a la media del conjunto de datos; el Error Absoluto Medio (MAE); el Error Cuadrático Medio (MSE); la correlación entre los valores observados y predichos ($cor_obspred$), que cuanto más cerca de 1 mejor; la correlación entre los valores predichos y residuales ($cor_predres$), que idealmente es 0; la Raíz del Error cuadrático medio (RMSE), que se busca sea lo más pequeña posible.

Tabla 8: Resultados de Validación Cruzada

	OK	UK	KED	MLR
mean_error	-0,060	-0,069	-0,039	0,002
me_mean	-0,002	-0,002	-0,001	0,000
MAE	9,786	9,798	9,362	9,841
MSE	159,600	160,700	140,900	150,097
cor_obspred	0,282	0,273	0,369	0,280
cor_predres	-0,040	-0,051	-0,098	0,088
RMSE	12,640	12,680	11,870	12,251

Fuente: Elaboración propia.

Nota: Tabla construida con el paquete *stargazer*.

Según los resultados de la Tabla 8, entre los métodos de Kriging el *KED* es el que presenta los mejores indicadores de validación cruzada (excepto por $cor_predres$ que es levemente más alejado de cero que para los otros métodos). La principal medida elegida para la comparación de los métodos es el *RMSE*, que se busca sea lo más pequeño posible, por lo que se selecciona el *KED*, las estimaciones del AB con dicho método resultan ser más precisas respecto a los otros. No obstante, se destaca que no existen grandes diferencias en el nivel de precisión de las estimaciones entre los diferentes métodos, particularmente entre el *MLR* y el *KED*. La introducción de covariables (en el *KED*) no representó una gran mejoría en el método de Kriging (en comparación a *OK* y *UK*).

Al aplicar el indicador de mejora relativa (*relative improvement*) descrito en el marco teórico, para comparar el cambio en el *RMSE* cuando se incorporan covariables en el kriging, es decir, para comparar los resultados del *KED* respecto al *OK*, el resultado obtenido es bajo, indica una mejora de solamente un 6%.

En la Tabla 9 se presentan comparaciones de las predicciones realizadas a partir de los kriging, la mayor diferencia se presenta entre los valores estimados máximo y mínimo entre el *KED* y los otros métodos de predicción espacial, esto se alinea a los resultados obtenidos por

Reis et al (2018). Seguidamente, se introducen figuras con visualizaciones utilizando la variable observada de área basal y las predicciones para evaluar los resultados. Se identifica que el KED resultó en estimaciones levemente más alineadas a las observaciones de campo.

Tabla 9: Comparación de predicciones en los modelos

	OK	UK	KED
Min.	20,20	20,22	-12,70
1st Qu.	26,35	26,34	24,96
Median	27,99	27,69	29,25
3rd Qu.	30,40	30,61	33,33
Max.	43,71	43,73	132,39

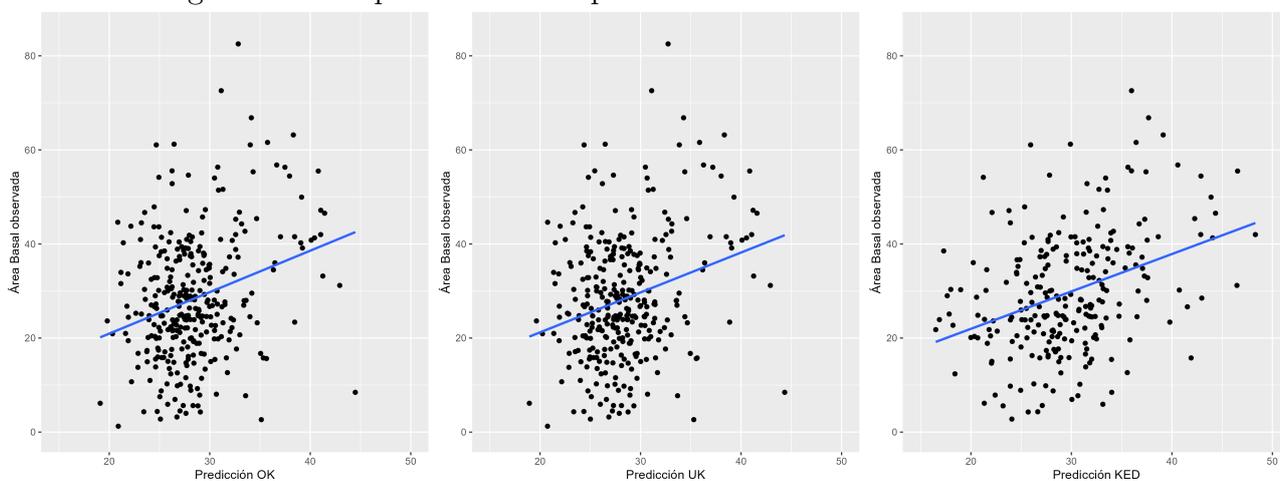
Fuente: Elaboración propia.

Tabla 10: Comparación de Desvíos Estándar obtenidos con los modelos

	OK	UK	KED
Min.	11,48	11,48	6,72
1st Qu.	11,86	11,85	10,10
Median	12,06	12,05	10,81
3rd Qu.	12,93	12,94	11,87
Max.	13,48	13,58	43,54

Fuente: Elaboración propia.

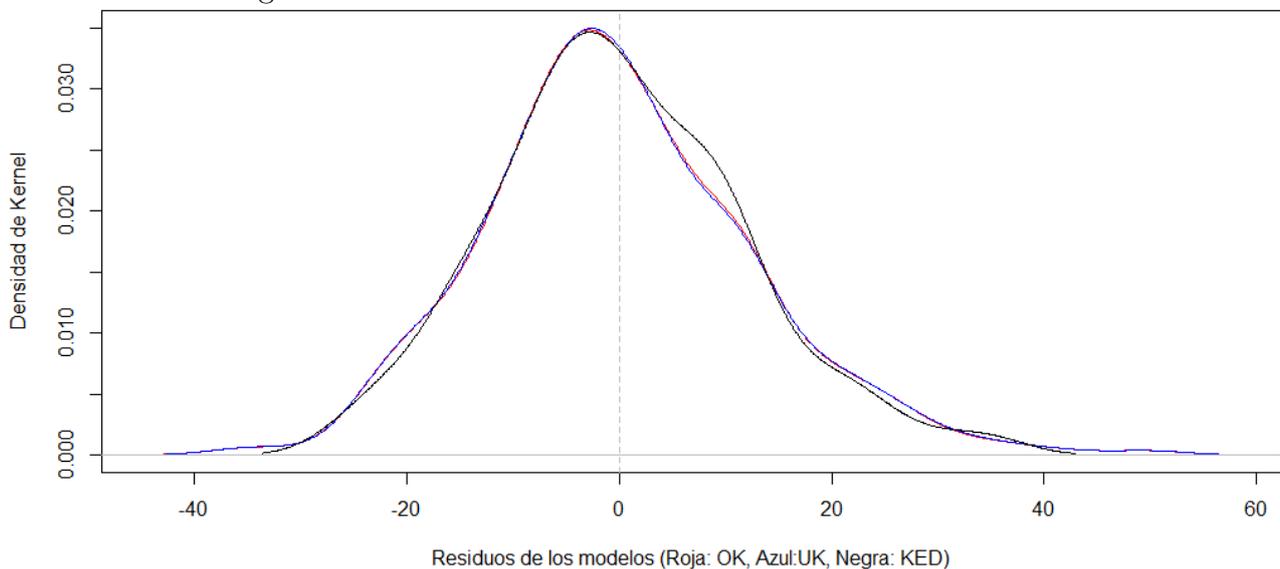
Figura 32: Comparación de las predicciones con el área basal observada



Fuente: Elaboración propia.

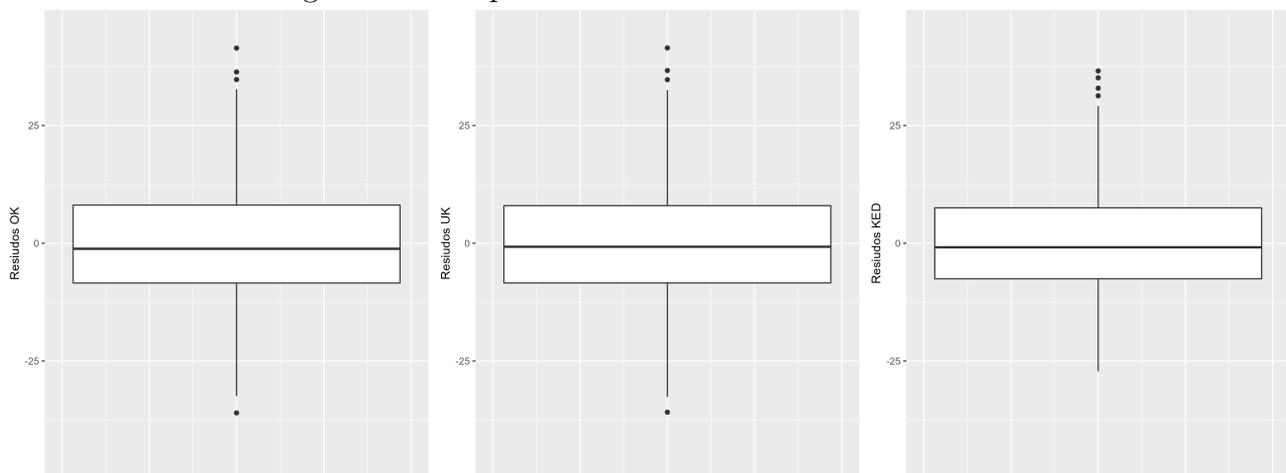
En la Figura 33 se presentan algunas visualizaciones comparativas de los residuos de los kriging. Se observa que los residuos con mejor comportamiento en términos de simetría son los resultantes del *KED*, el modelo seleccionado como la mejor alternativa en el diagnóstico de validación cruzada. Se realizan contrastes de normalidad de dichos residuos, el de Shapiro arroja un *pvalor* de 0.08 por lo que tanto a un 1% como a un 5% no se rechaza la hipótesis nula de normalidad de los residuos, lo mismo ocurre con el test Jarque Bera con un *pvalor* de 0.06. Se observa una mejora en el comportamiento de los residuos al comparar los resultados del *KED* respecto a los de la *RLM*, esto es producto de la incorporación del componente espacial a las estimaciones mediante kriging.

Figura 33: Densidad de Kernel de los residuos de los modelos



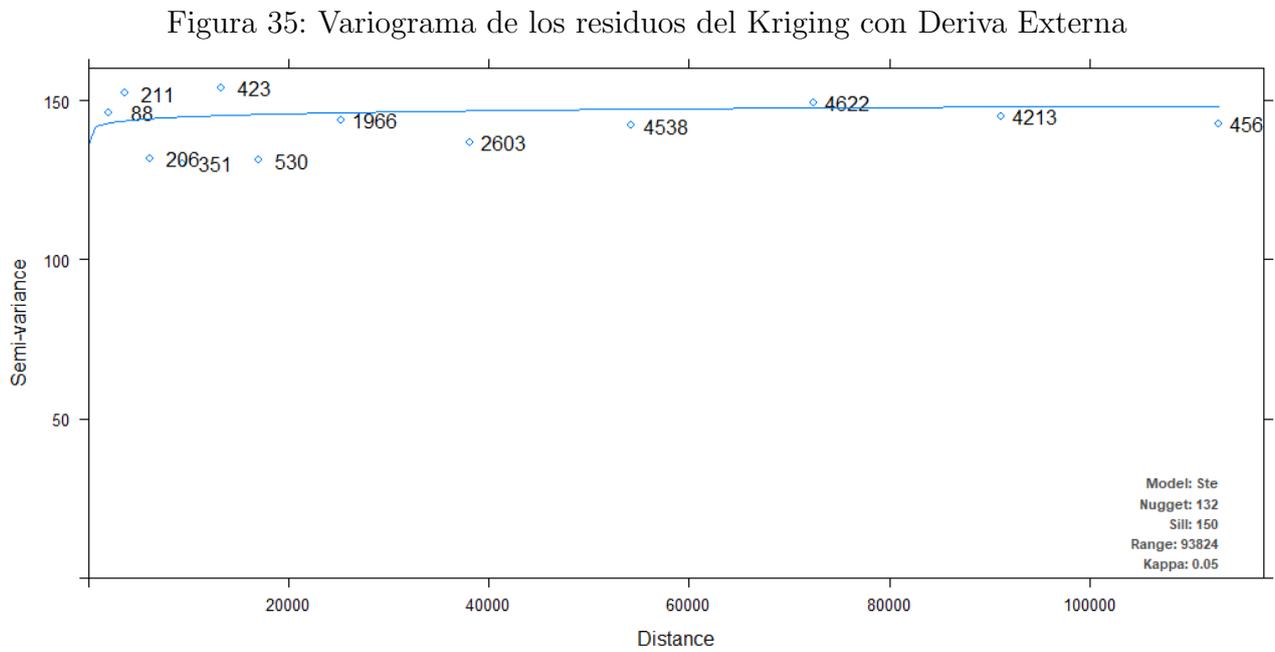
Fuente: Elaboración propia.

Figura 34: Comparación de los residuos de los modelos



Fuente: Elaboración propia.

En la Figura 35 se presenta el variograma de los residuos del método de predicción espacial *KED*, el seleccionado como la mejor herramienta en el presente abordaje empírico. Tal como se busca, no presentan estructuras de dependencia espacial.



Fuente: Elaboración propia.

6. Conclusiones

Este trabajo tuvo como principal objetivo predecir el área basal mediante la utilización de modelos geoestadísticos, a partir de información de campo relevada (IFN), imágenes satelitales (Landsat) e información de otras fuentes vinculadas con el uso y características del suelo (Índice Coneat, precipitaciones, temperatura, Evapotranspiración, Índices de vegetación, distancia al curso de agua, modelo digital de terreno, entre otros). Los modelos utilizados en este estudio proporcionan un marco para la integración de datos de campo y multiespectrales, con el fin de obtener predicciones espaciales de AB.

Los costos de relevamiento de campo son uno de los motivos por el cuál es importante el abordaje que se realiza en el presente documento, de predicción de área basal mediante modelos geoestadísticos.

La relevancia de los bosques nativos en Uruguay refuerza la importancia de investigar sobre métodos robustos que consideren información auxiliar en el proceso de estimación de variables de interés, como el AB. Las metodologías presentadas aquí son algunas de las herramientas para la realización de estimaciones a partir de los datos espectrales obtenidos de Landsat 8 y de otros rasters.

La selección del área de estudio (*ventana*), resulta una tarea compleja con inevitables *trade-offs*, que a su vez presenta importantes impactos sobre los resultados obtenidos. Por un lado, resulta necesario contar con una cantidad suficiente de puntos para obtener estimaciones precisas. Por otro lado, cuanto mayor es el área de estudio, más probable resulta que el comportamiento del AB responda a procesos de diferentes características, existiendo mayor influencia de variables inobservables y la superposición de efectos. En línea con esto último, Brazeiro et al (2015) identifica varias eco-regiones con características diferenciadas en el territorio nacional, para el presente trabajo se decide seleccionar la Cuenca Gondwánica para analizar el comportamiento del AB y estimar los modelos, para posteriormente realizar la predicción únicamente en una sub-área de interés definida por la Dirección General Forestal (DGF).

Se utilizan herramientas de interpolación por Kriging, en particular de tipo Ordinario, Universal y con Deriva Externa. Este último es el que resulta ser más preciso para estimar el área basal, presentando un *RMSE* más pequeño, de 11.87, en comparación con 12.64 (OK), 12.68 (UK) y 12.25 (MLR). No existen grandes diferencias en el nivel de precisión de las estimaciones entre los diferentes métodos, particularmente entre el *MLR* y el *KED*.

Las variables espectrales que más contribuyeron al modelo estimado fueron las bandas 2, 3, 5 y 6 de Landsat, estas presentaron mejores valores de correlación con el AB, esto puede deberse a que la variable de interés se deriva solo del diámetro de los árboles (DAP), que está directamente relacionado con el tamaño de las copas de los árboles que determina la reflectancia de la copa, la cual es captada por las bandas (dos Reis y cols., 2018). No obstante, la introducción de covariables representó una muy leve mejoría en el método de kriging.

Por otra parte, también quedaron seleccionadas en el modelo las variables elevación y pendiente (extraídas del Modelo Digital de Terreno), la distancia al curso de agua y las precipitaciones. Se destaca que dichas variables son determinantes del tipo de bosque que se desarrolla en un lugar particular, por lo que una posible interpretación es que la variabilidad del área basal se ve explicada por el tipo de bosque (cada tipología tiene características particulares en términos de las variables mencionadas). Es de esperarse que el comportamiento del área basal en un Bosque Ribereño sea diferente al de uno de Parque, Quebrada o Palmar, lo que le da coherencia a los resultados de las variables seleccionadas por el modelo.

Es importante señalar que la detección remota del AB en Bosque Nativo es compleja debido al tamaño, la forma y las propiedades dieléctricas de sus elementos dispersos como hojas, ramas y tallos. La diversidad espacial del área basal hace que la relación entre el relevamiento de campo y los datos de teledetección sea un gran desafío (dos Reis y cols., 2018). En el presente abordaje empírico, lo anterior se vio reflejado en el bajo poder predictivo de los modelos estimados.

Los resultados del estudio actual necesitan ser interpretados con cautela, ya que presentan diversas limitaciones. En primer lugar se encuentran los posibles errores de medición de las variables y los sesgos en las mediciones de área basal producidos por diferentes momentos de relevamiento de campo. Existe un intervalo intertemporal de 7 años entre que se tomaron los primeros datos (2009) y los últimos (2016). Esto puede llevar a pensar que en ese tiempo podría haber rotatividad en los equipos y metodologías de trabajo (aspecto que desconocemos), o incluso si se mantienen los equipos, las personas podrían mejorar su método de relevamiento con la experiencia, introduciendo también un sesgo.

A pesar de las limitaciones, este trabajo es el primero en realizar aplicaciones de interpolación por kriging para la predicción de área basal en Uruguay, abriendo una línea de investigación a futuro para desarrollar otros métodos que puedan ser de utilidad para la mejora de la gestión de Bosques Nativos. En la siguiente sección se presentan posibles líneas futuras de investigación en función de los resultados arrojados por el presente trabajo y la revisión bibliográfica en la que se enmarcó.

7. Líneas futuras de investigación

La resolución espacial Landsat (30 m) puede ser considerada demasiado gruesa para mapear de manera confiable variables vinculadas a los árboles a pequeña escala. Existe otro conjunto de imágenes satelitales denominadas Sentinel-2, con una resolución espacial remuestreada de 10 m, que podría mejorar la estimación de área basal (Almeida Lima y cols., 2019; Avogadro y Padró Garcia, 2019). En el presente trabajo no se utilizaran imágenes Sentinel, debido a que en la revisión de antecedentes realizada, se encuentra que no proporciona resultados muy diferentes a la utilización de Landsat. No obstante, queda como línea futura de investigación la realización de un análisis comparativo entre ambos grupos de imágenes para el presente estudio. También

podría utilizarse el vuelo de la IDE, de un tamaño de píxel aún menor³⁷.

Por otra parte, podrían incorporarse nuevos rasters con información sobre otras variables relevantes en la determinación del comportamiento del área basal, que no se hayan utilizado en el presente documento.

Otra línea futura de investigación es el abordaje de estrategias empíricas de predicción de AB a partir de otros modelos de predicción espacial (Cokriging, kriging de ventana móvil, Random Forest, SVM, Artificial Neural Networks, etc).

Varios estudios han demostrado la superioridad de la incorporación de algoritmos de aprendizaje en relación con la estadística clásica en varias áreas, como en el manejo forestal. Por ejemplo, Ahmed et al. (2015) modeló la estructura de datos en una serie temporal de imágenes Landsat en conjunción con datos de tipo LiDAR. Por lo tanto, el análisis temporal de la evolución del AB, es otra posible línea de investigación futura (Boisvenue C., 2016; Gao y cols., 2016; Melo Martínez, 2012).

Por otra parte, también podrían realizarse aplicaciones de algoritmos de AA, estos pueden aprender continuamente de nuevos datos y mantener todo el conocimiento acumulado de conjuntos de datos anteriores. El uso de todas las variables auxiliares en el proceso de estimación es otra ventaja sobre los métodos de regresión tradicionales, ya que los algoritmos de AA no están restringidos por la correlación entre las variables de entrada, evitando así la pérdida de información importante en el proceso de estimación de la variable de interés. Sin embargo, estos métodos tienen como desventaja que la relación causal entre las entradas y los resultados del proceso de estimación no está claro, lo cual implica una interpretación biológica limitada (dos Reis y cols., 2018; García, Martínez, Troncoso, y Riquelme, 2015).

Otra metodología que podría ser abordada es el Kriging de ventana móvil (Haas, 1990), resulta particularmente apropiado para el presente caso de estudio, debido a la influencia que tiene la selección de la ventana en los resultados, que se refleja en variogramas locales diferentes entre sí. Este método implica calcular los parámetros del semivariograma rango, nugget y meseta parcial varias veces a partir de ventanas vecinas pequeñas. Es decir, se utiliza una ventana móvil centrada en la ubicación que se va a predecir y se crea un semivariograma para cada contigüidad. De esta forma, la predicción de cada punto se realiza secuencialmente a medida que la ventana se mueve a través del área de estudio (Harris y cols., 2010).

³⁷Este abordaje exige mayores recursos tecnológicos de los que se dispone para el presente trabajo, los procedimientos de kriging a partir de píxeles pequeños en los rasters utilizados, implican altos tiempos de máquina

Referencias

- Ahmed, O., Franklin, S., Wulder, M., y White, J. (2015, 03). Characterizing stand-level forest canopy cover and height using landsat time series, samples of airborne lidar, and the random forest algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101. doi: 10.1016/j.isprsjprs.2014.11.007
- Almeida Lima, T., Beuchle, R., Langner, A., Grecchi, R., Griess, V., y Frédéric, A. (2019, 04). Comparing sentinel-2 msi and landsat 8 oli imagery for monitoring selective logging in the brazilian amazon. *Remote Sensing*, 11. Descargado de https://www.researchgate.net/publication/332571528_Comparing_Sentinel-2_MSI_and_Landsat_8_OLI_Imagery_for_Monitoring_Selective_Logging_in_the_Brazilian_Amazon doi: 10.3390/rs11080961
- Avogadro, E. G., y Padró Garcia, J. C. (2019, dic.). Diferenciación de plantaciones forestales en entrerríos (argentina): Comparación de métodos de clasificación aplicados a imágenes sentinel-2 y landsat-8. *GeoFocus. Revista Internacional de Ciencia y Tecnología de la Información Geográfica*(24), 117–139. Descargado de <https://www.geofocus.org/index.php/geofocus/article/view/652> doi: 10.21138/GF.652
- Aybar, C., Bautista, L., Yali, R., Barja, A., y Wu, Q. (2020, 07). rgee: An r package for interacting with google earth engine. *The Journal of Open Source Software*, 5, 2272. doi: 10.21105/joss.02272
- Bernardi, L., Bocco, A., Miguel, C., Olivera, J., Penengo, C., y Rama, G. (2019). Metodología aplicada en la elaboración del mapa de cobertura de bosque nativo de uruguay para el año 2016 con imágenes sentinel 2. *Proyecto REDD+ Uruguay..*
- Boisvenue C., W. J. K. W. W. M., Smiley B. (2016). Integration of landsat time series and field plots for forest productivity estimates in decision support models. *Forest Ecology and Management*(376:284-297). Descargado de https://www.researchgate.net/publication/304628852_Integration_of_Landsat_time_series_and_field_plots_for_forest_productivity_estimates_in_decision_support_models
- Brazeiro, A. (2015). Eco-regiones de uruguay: biodiversidad, presiones y conservación. Descargado de <https://ppduruguay.undp.org.uy/wp-content/uploads/2020/06/Ecoregiones.pdf>
- Brazeiro, A. (2018). *Recientes avances en investigación para la gestión y conservación del bosque nativo de uruguay*. Montevideo: Facultad de Ciencias, MGAP, BMEL. Descargado de <https://chm.cbd.int/api/v2013/documents/8401E92D-C23D-4CB8-1C59-5CA71AD404A8/attachments/205743/Seminario%20Investigaci%C3%B3n%20Bosque%20Nativo.pdf>
- Brussa, C. (2015). Programa de relevamientos y caracterización de ambientes y flora. *Bosques del Uruguay. Etapa 15. Revisión y preselección de predios..* Descargado de <https://docplayer.es/68710629-Programa-de-relevamientos-y-caracterizacion-de-ambientes-y-flora-bosques-del-uruguay-etapa-i-revision-y-preseleccion-de-predios.html>
- Caruso, C., y Quarta, F. (1998). Interpolation methods comparison. *Computers Mathematics with Applications*, 35(12), 109-126. Descargado de <https://www.sciencedirect.com/science/article/pii/S0898122198001011> doi: [https://doi.org/10.1016/S0898-1221\(98\)00101-1](https://doi.org/10.1016/S0898-1221(98)00101-1)
- Cressie, N. (1993). *Statistics for spatial data*. New York: Wiley Series in Probability and

Statistics.

- Cutler, M., Boyd, D., Foody, G., y Vetrivel, A. (2012, 06). Estimating tropical forest biomass with a combination of sar image texture and landsat tm data: An assessment of predictions between regions. *ISPRS Journal of Photogrammetry and Remote Sensing, Volume 70, p. 66-77.*, 70, 66-77. doi: 10.1016/j.isprsjprs.2012.03.011
- Dai, F., Zhou, Q., Lv, Z., Wang, X., y Liu, G. (2014, 10). Spatial prediction of soil organic matter content integrating artificial neural network and ordinary kriging in tibetan plateau. *Ecological Indicators*, 45, 184–194. doi: 10.1016/j.ecolind.2014.04.003
- DGF. (2018). *Estrategia nacional de bosque nativo*. Montevideo: MGAP. Descargado de https://www.gub.uy/ministerio-ganaderia-agricultura-pesca/sites/ministerio-ganaderia-agricultura-pesca/files/documentos/publicaciones/estrategia_nacional_de_bosque_nativo.pdf
- dos Reis, A. A., Carvalho, M. C., de Mello, J. M., Gomide, L. R., Ferraz Filho, A. C., y Acerbi Junior, F. W. (2018). Spatial prediction of basal area and volume in eucalyptus stands using landsat tm data: an assessment of prediction methods. *New Zealand Journal of Forestry Science*, 48. Descargado de <https://doi.org/10.1186/s40490-017-0108-0> doi: 10.1186/s40490-017-0108-0
- Fassnacht, F., Hartig, F., Latifi, H., Berger, C., Hernandez, J., Corvalan, P., y Koch, B. (2014, 07). Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. *Remote Sensing of Environment*, 154, 102-114. doi: 10.1016/j.rse.2014.07.028
- Fayada, I., Baghdadi, N., Guitet, S., Baill, J., Hérault, B., Gonde, V., ... Minha, D. (2016). Aboveground biomass mapping in french guiana by combining remote sensing, forest inventories and environmental data. *International Journal of Applied Earth Observation and Geoinformation*(52). Descargado de <https://hal.archives-ouvertes.fr/hal-01381116/document>
- Freeman E., C. J. W. B., Moisen G. (2014). *Random forests and stochastic gradient boosting for predicting tree canopy cover: Comparing tuning processes and model performance* (Vol. 16) (n.º 5).
- Galeana, J., López, A., López, P., Silván, J., y Couturier, S. (2014a, 08). Modeling the spatial distribution of above-ground carbon in mexican coniferous forests using remote sensing and a geostatistical approach. *International Journal of Applied Earth Observation and Geoinformation*, 30, 179–189. doi: 10.1016/j.jag.2014.02.005
- Galeana, J., López, A., López, P., Silván, J., y Couturier, S. (2014b, 08). Modeling the spatial distribution of above-ground carbon in mexican coniferous forests using remote sensing and a geostatistical approach. *International Journal of Applied Earth Observation and Geoinformation*, 30, 179–189. Descargado de <https://www.sciencedirect.com/science/article/abs/pii/S0303243414000440> doi: 10.1016/j.jag.2014.02.005
- Gao, T., Zhu, J., Deng, S., Zheng, X., Zhang, J., Shang, G., y Huang, L. (2016). Timber production assessment of a plantation forest: An integrated framework with field-based inventory, multi-source remote sensing data and forest management history. *International Journal of Applied Earth Observation and Geoinformation*, 52(C), 155-165. doi: 10.1016/j.jag.2016.06.004
- García, J., Martínez, F., Troncoso, A., y Riquelme, J. (2015, 05). A comparison of machine learning regression techniques for lidar-derived estimation of forest variables. *Neurocomputing*, 167, 24-31. doi: 10.1016/j.neucom.2014.09.091

- Gebreslasie M., A. V., Ahmed F. (2010, 02). Predicting forest structural attributes using ancillary data and aster satellite data. *International Journal of Applied Earth Observation and Geoinformation*, 12S, S23-S26. doi: 10.1016/j.jag.2009.11.006
- Giraldo, R. (2002). *Introducción a la geoestadística. teoría y aplicación*. Universidad Nacional de Colombia.
- Goovaerts, P. (1997, 01). Geostatistics for natural resource evaluation. En (Vol. 42).
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., y Moore, R. (2017). Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18-27. Descargado de <https://www.sciencedirect.com/science/article/pii/S0034425717302900>
- Harris, P., Charlton, M., y Fotheringham, A. (2010, 12). Moving window kriging with geographically weighted variograms. *Stochastic Environmental Research and Risk Assessment*, 24, 1193-1209. Descargado de https://www.researchgate.net/publication/225458502_Moving_window_kriging_with_geographically_weighted_variograms doi: 10.1007/s00477-010-0391-2
- Hengl, T. (2009). *A Practical Guide to Geostatistical Mapping* (Inf. Téc.). Luxembourg: Office for Official Publications of the European Communities. Descargado de http://spatial-analyst.net/book/system/files/Hengl_2009_GEOSTATE2c1w.pdf
- Hengl, T., Heuvelink, G., y Rossiter, D. (2007, 10). About regression-kriging: From equations to case studies. *Computers Geosciences*, 33, 1301-1315. Descargado de https://www.researchgate.net/publication/223155790_About_regression-kriging_From_equations_to_case_studies doi: 10.1016/j.cageo.2007.05.001
- Krige, D. (1951). A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, 52(6), 119-139. Descargado de https://journals.co.za/doi/abs/10.10520/AJA0038223X_4792 doi: 10.10520/AJA0038223X\ _4792
- Latifi, H., Nothdurft, A., y Koch, B. (2010, 09). Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: Application of multiple optical/lidar-derived predictors. *Forestry*, 83. doi: 10.1093/forestry/cpq022
- Li, J., y Heap, A. D. (2014, mar). Spatial interpolation methods applied in the environmental sciences. *Environ. Model. Softw.*, 53(C), 173-189.
- López Serrano, P., Corral, J. J., Díaz Varela, R., Álvarez, J., y López, C. (2016, 04). Evaluation of radiometric and atmospheric correction algorithms for aboveground forest biomass estimation using landsat 5 tm data. *Remote Sensing*, 2016, 369. doi: 10.3390/rs8050369
- Matheron, G. (1962). *Traite de geostatistique apliquee, tome i*. Memoires bureau de Recherches Geologiques et Minieres, N 24. Editions Bureau de Recherche et Minieres, Paris.
- Melo Martínez, C. E. (2012). *Análisis geoestadístico espacio tiempo basado en distancias y splines con aplicaciones*. Universitat de Barcelona. Departament d'Estadística. Descargado de <http://hdl.handle.net/10803/101202>
- Meng Q., M. M., Cieszewski C.J. (2009, 01). Large area forest inventory using landsat etm+: A geostatistical approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64, 27-36. doi: 10.1016/j.isprsjprs.2008.06.006
- Morgenroth, J., y Visser, R. (2013). Uptake and barriers to the use of geospatial technologies in forest management. *New Zealand Journal of Forestry Science*(43:16). Descargado de <https://nzjforestryscience.springeropen.com/track/pdf/10.1186/1179-5395-43-16.pdf>

- Mutanga, O. (2015, 10). Investigating the robustness of the new landsat-8 operational land imager derived texture metrics in estimating plantation forest aboveground biomass in resource constrained areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108. doi: 10.1016/j.isprsjprs.2015.06.002
- Petraglia, C., y Dell'Acqua, M. (2006). Actualización de la carta forestal del Uruguay con imágenes del año 2004. *MGAP*. Descargado de http://web.renare.gub.uy/media/cobertura/informes/CARTA_FORESTAL2004.pdf
- Pérez-Pineda, J. (2006, 10). Econometría espacial y ciencia regional. *Investigación Económica*, 65, 129-160.
- Sichel, H. (1947). *An experimental and theoretical investigation of bias error in mine sampling with special reference to narrow gold reefs*. Trans. Inst. Min. Metall.
- Sichel, H. (1949). *Mine valuation and maximum likelihood*. University Witwatersrand.
- Stojanova, D., Panov, P., Gjorgjioski, V., Kobler, A., y Džeroski, S. (2010, 07). 1 2 estimating vegetation height and canopy cover from remotely sensed data with machine learning 3 4 5. *Ecological Informatics*, 5, 256-266. doi: 10.1016/j.ecoinf.2010.03.004
- Uruguay, P. R. (2019). Análisis de las especies exóticas invasoras (eei) en bosques nativos del Uruguay, en base a parcelas del inventario forestal nacional (ifn). *Ministerio de Ganadería, Agricultura y Pesca - Ministerio de Vivienda, Ordenamiento Territorial y Medio Ambiente. Montevideo*.
- Vicharnakorn P., N. M. S. A. K. S., Shrestha R. (2014). Carbon stock assessment using remote sensing and forest inventory data in savannakhet, lao pdr. *Remote Sensing*, 6(6), 5452-5479. Descargado de <https://www.mdpi.com/2072-4292/6/6/5452> doi: 10.3390/rs6065452
- Watt MS., W. P. B. S., Dash JP. (2016). Multi-sensor modelling of a forest productivity index for radiata pine plantations. *New Zealand Journal of Forestry Science*(46:9). Descargado de <https://nzjforestryscience.springeropen.com/track/pdf/10.1186/s40490-016-0065-z.pdf>
- Webster, O. M., R. (2007). *Geostatistics for environmental scientists. 2nd edition*. John Wiley Sons, Ltd. Descargado de https://www.academia.edu/6871734/Geostatistics_for_Environmental_Scientists_Second_Edition
- Zahriban Hesari, M., Fallah, A., Shataee, S., Kalbi, S., y Persson, H. (2019, 10). Estimating the forest stand volume and basal area using pleiades spectral and auxiliary data.
- Zahriban M., S. S. K. S. P. H., Fallah A. (2019). Estimating the forest stand volume and basal area using pleiades spectral and auxiliary data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Descargado de <https://pdfs.semanticscholar.org/3d83/7814d573cb88471d8ebf6a64497856e3dc77.pdf>
- Zhang, J., Huang, S., Hogg, E., Lieffers, V., Qin, Y., y He, F. (2013, 12). Estimating spatial variation in alberta forest biomass from a combination of forest inventory and remote sensing data. *Biogeosciences Discussions*, 10, 19005-19044. doi: 10.5194/bgd-10-19005-2013
- Zhang, X., Lei, Y., y Cao, Q. (2010, 12). Compatibility of stand basal area predictions based on forecast combination. *Forest Science*, 56, 552-557.

Bibliografía

- Altamirano A., Califra A., Dell'Acqua M., Molfino J., Petraglia C, (1996) Estudio de la Cuenca Alta del Río Negro utilizando imágenes satelitales y sistemas de información geográfica
- Arriaga M. (2018). Resultados de la Cartografía Nacional Forestal. Dirección General Forestal Ministerio de Ganadería Agricultura y Pesca.
- Bivand, R., Lewin-Koh, N. (2021). *maptools: Tools for Handling Spatial Objects*. R package version 1.1-1.
- Bivand, R., Keitt, T., Rowlingson, B. (2021). *rgdal: Bindings for the 'Geospatial' Data Abstraction Library*. R package version 1.5-23.
- Cabrero, Y., García, A. (2015). Análisis estadístico de datos espaciales con QGIS y R.
- Chang, W., Cheng, J., Allaire, J., Xie, Y., y McPherson, J. (2020). *shiny: Web Application Framework for R*.
- Dell Aqua (2006) Actualización de la carta forestal de Uruguay con imagenes Landsat 2004.
- Demaria, M (2014) Información Geográfica Digitalizada de Uruguay
- Diamantopoulou M. (2012) Assessing a reliable modeling approach of features of trees through neural network models for sustainable forests.
- Diggle, P. J., Ribeiro, P. J. (2007). *Model-based Geostatistics*. (Springer Series in Statistics). Springer.
- Dunnington, D. (2021). *ggspatial: Spatial Data Framework for ggplot2*. R package version 1.1.5.
- Evans JS (2021). *spatialEco*. R package version 1.3-6
- Ferraco H., de Castro F., Soares J., Burkhart H., McTague J., Regis M., Araújo R., da Fonseca S., Cardoso R. (2016) Modeling dominant height growth of eucalyptus plantations with parameters conditioned to climatic variations
- Haas TC (1990) Lognormal and moving window methods of estimating acid deposition. *J Am Stat Assoc* 85:950–963
- Harrell Jr, with contributions from Charles Dupont and many others. (2021). *Hmisc: Harrell Miscellaneous*. R package version 4.5-0.
- Hernandez, J. (2002) Guía de densidad para manejo de bosques nativos. *Rev. Mex. Cien. For.* Vol. 4 Núm. 19.
- Hiemstra, P.H., Pebesma, E.J., Twenhofel, C.J.W. and G.B.M. Heuvelink (2008). Real-time automatic interpolation of ambient gamma dose rates from the Dutch Radioactivity Monitoring Network. *Computers Geosciences*, accepted for publication.
- Huete, A, Didan, K, Miura, T, Rodriguez, EP, Gao, X, Ferreira, LG. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83, 195–213.
- Huete, AR. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25, 295–309.
- Immitzer M., Atzberger C., Koukal T. (2012) Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data.
- Kennedy W., Dieu T. , Øystein B. , Bal R. (2014) A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afrotropical landscape
- Kuhn, M. (2021). *caret: Classification and Regression Training*. R package version 6.0-88.
- Marek, H. (2018). *stargazer: Well-Formatted Regression and Summary Statistics Tables*. R package version 5.2.1.

- Moss, J., Tveten, M. (2020). *kdensity*: Kernel Density Estimation with Parametric Starts and Asymmetric Kernels. R package version 1.1.0.
- Nothdurft A, Saborowski J., Breidenbach J. (2009) Spatial prediction of forest stand variables
- Olaya, V. (2011) *Sistemas de Información Geográfica*.
- Özçelik R., Diamantopoulou M., Crecente F., Eler U. (2013) Estimating Crimean juniper tree height using nonlinear regression and artificial neural network models.
- Pebesma, E.J., 2004. Multivariable geostatistics in S: the *gstat* package. *Computers Geosciences*, 30: 683-691.
- Pebesma, E.J., R.S. Bivand, 2005. Classes and methods for spatial data in R. *R News* 5 (2)
- Pebesma, E., 2018. Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal* 10 (1), 439-446,
- QGIS Development Team (2020). QGIS Geographic Information System. Open Source Geospatial Foundation Project.
- Riaño M. (2019) Metodología y resultados de la validación del mapa de cobertura de bosque nativo de Uruguay para el año 2016 con imágenes Sentinel 2. Proyecto REDD+ Uruguay.
- Ribeiro Jr, Paulo J., Diggle, J., Schlather, M., Bivand, R., Ripley, B. (2020). *geoR*: Analysis of Geostatistical Data. R package version 1.8-1.
- Robert J. Hijmans (2021). *raster*: Geographic Data Analysis and Modeling. R package version 3.4-13.
- Rijo, L., Santiñaque (2011) Investigación de metodologías para el análisis espacio temporal del evento de granizo en Uruguay. Tesis de Licenciatura en Estadística. UDELAR.
- Rouse, J, Haas, R, Schell, J, Deering, D, Harlan, J (1973). Monitoring the vernal advancements and retrogradation (greenwave effect) of nature vegetation. NASA/GSFC final report. Greenbelt: NASA.
- Tennekes M (2018). *tmap*: Thematic Maps in R. *Journal of Statistical Software*, *84*(6), 1-39. doi: 10.18637/jss.v084.i06.
- van de Kasstelee, J., Stein, A., Dekkers, A. Velders, G. (2009), 'External drift kriging of NO_x concentrations with dispersion model output in a reduced air quality monitoring network', *Environmental and Ecological Statistics* 16, 321-339.
- Viana H., Aranha J., Lopes D., Cohen B. (2011) Estimation of crown biomass of *Pinus pinaster* stands and shrubland above-ground biomass using forest inventory data, remotely sensed imagery and spatial prediction models
- Wei, T.; Simko, V. (2021). R package '*corrplot*': Visualization of a Correlation Matrix (Version 0.90).
- Wikle, C. K., Zammit-Mangion, A., Cressie, N. (2019), *Spatio-Temporal Statistics with R*.
- Wickham, H. (2016). *ggplot2*: Elegant Graphics for Data Analysis. Springer-Verlag New York.
- Wickham, H. (2018). *stringr*: Simple, Consistent Wrappers for Common String Operations. R package version 1.3.1.
- Wickham, H., Francois, R., Henry, L., and Muller, K. (2018). *dplyr*: A Grammar of Data Manipulation. R package version 0.7.5.
- Wickham, H., Hester, J., and Francois, R. (2017). *readr*: Read Rectangular Text Data. R package version 1.1.1.
- Wickham, H. and Miller, E. (2017). *haven*: Import and Export 'SPSS', 'Stata' and 'SAS' Files. R package version 1.1.0.

8. Apéndice

8.1. Librerías utilizadas

Este trabajo fue realizado utilizando el lenguaje de programación R y la interfaz de usuario R studio (R Core Team, 2017, RStudio Team, 2016). Se utilizaron las siguientes librerías:

Utilitarios y manipulación de datos

- dplyr (Wickham et al., 2018)
- ggplot2 (Wickham, 2016)
- lubridate (Grolemund and Wickham, 2011)
- readr (Wickham et al., 2017)
- Hmisc (Harrell et al., 2021)
- corrplot (Wei and Simko, 2021)
- stargazer (Marek, 2018)
- kdensity (Moss and Tven, 2020)

Manipulación de información geográfica y estimaciones geoestadísticas

- geoR (Ribeiro et al, 2020).
 - rgdal (Bivand et al, 2021).
 - raster (Robert J. Hijmans, 2021)
 - automap (Paul Hiemstra, 2008).
 - sp (Edzer Pebesma, Roger Bivand 2005).
 - sf (Edzer Pebesma, 2018).
 - tmap (Tennekes, 2018)
 - maptools (Roger Bivand, Nicholas Lewin-Koh, 2021)
 - rgee (Aybar et al, 2020)
 - ggspatial (Dunnington, 2021)
 - caret (Kuhn, 2021)
 - spatialEco (Evans, 2021)
 - gstat (Pebesma, 2004)
-

8.2. Shiny App y GitHub

En el marco del presente trabajo se elaboró un visualizador, desarrollado como una aplicación web usando la librería *shiny*³⁸ en R (Chang et al., 2020), donde se puede encontrar la publicación de los resultados del documento, a partir de visualizaciones interactivas.

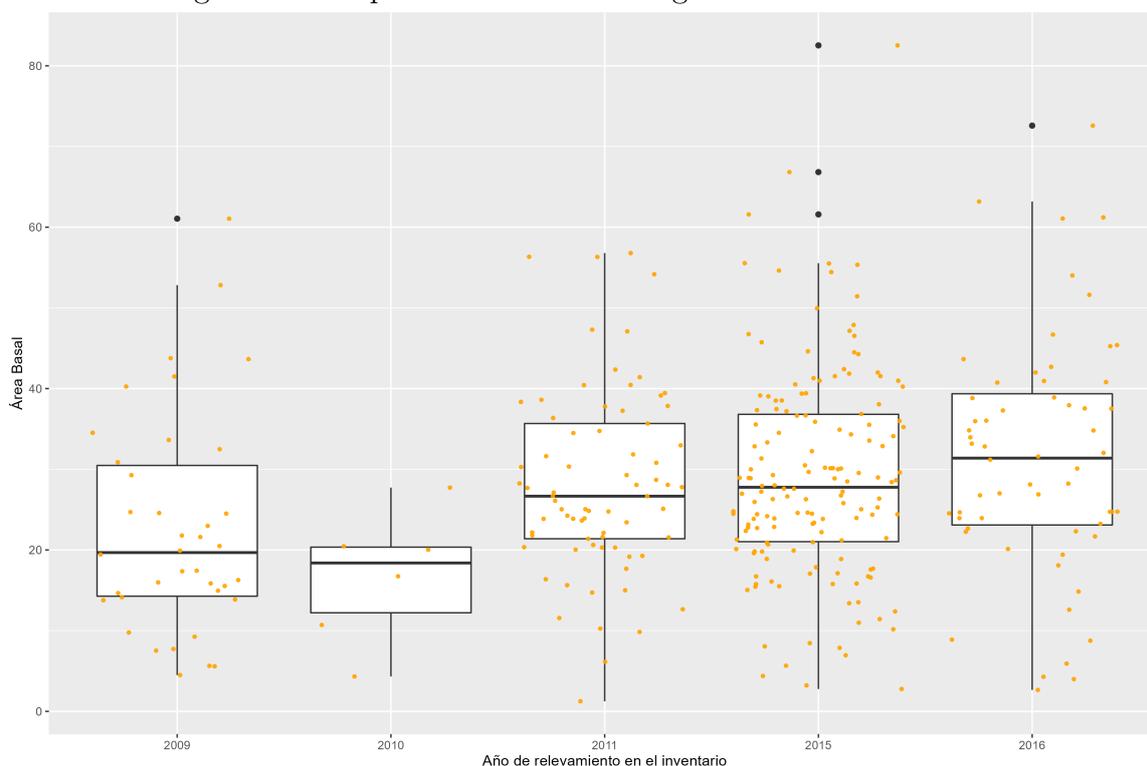
El link al visualizador es el siguiente: https://cinve.shinyapps.io/Area_Basal_Resultados/

Por otra parte, los códigos desarrollados en el presente trabajo (incluso el vinculado a Shiny) pueden encontrarse en GitHub:

https://github.com/Cecilia-Alonso/Tesis_Final

8.3. Análisis exploratorio de los datos (complementario)

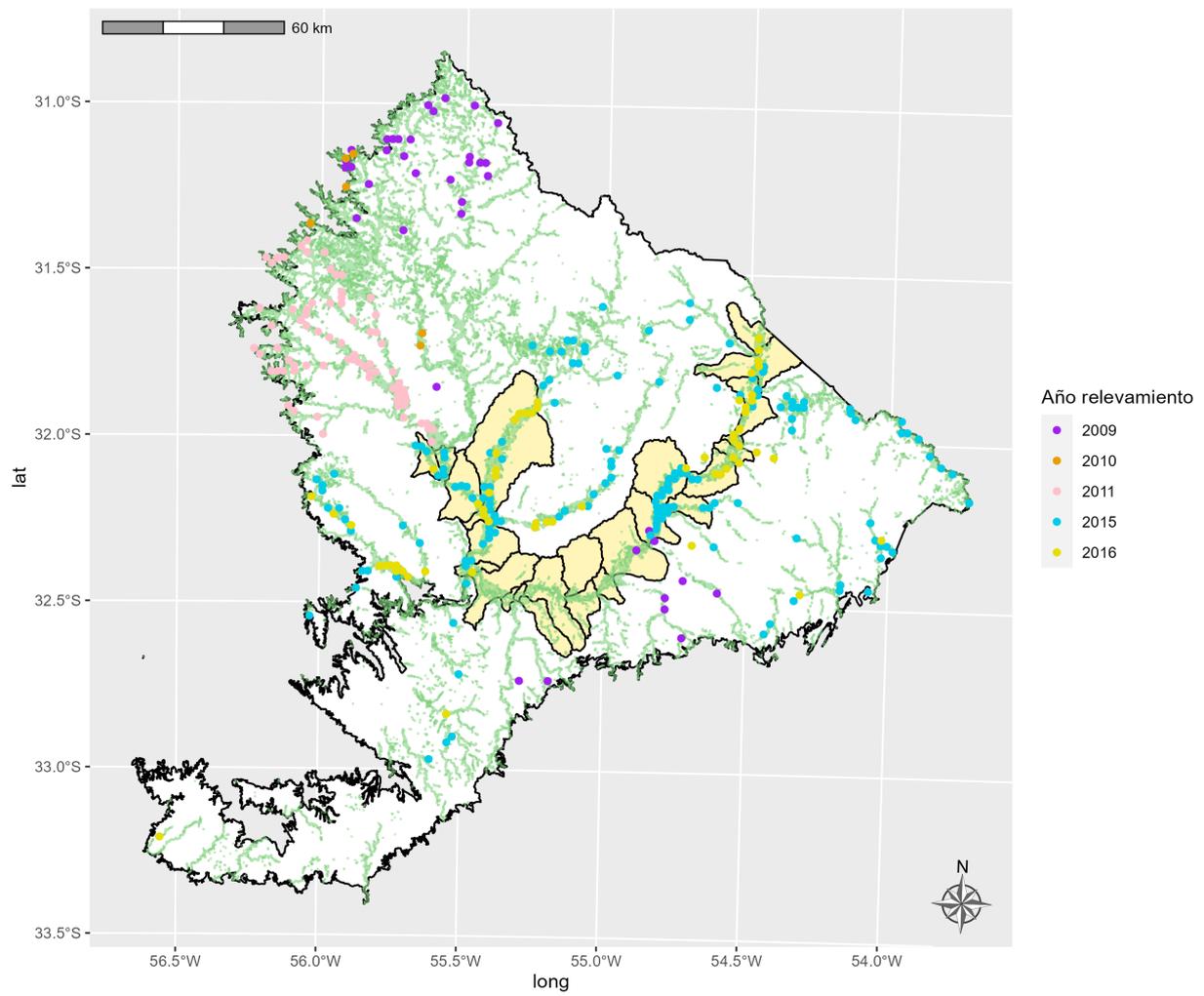
Figura 36: Boxplot del Área Basal según año de relevamiento



Fuente: Elaboración propia.

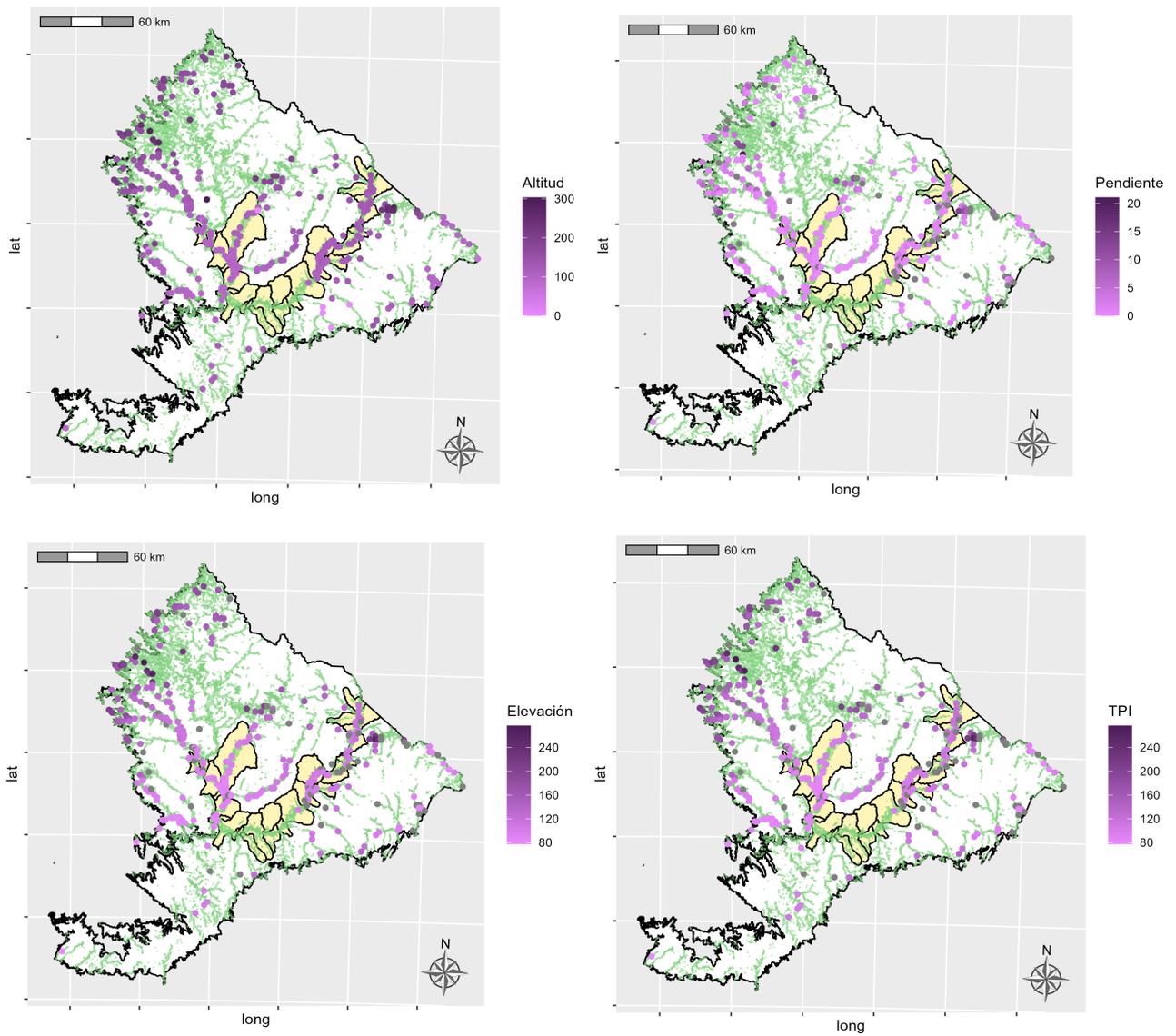
³⁸La estructura de todas las aplicaciones de Shiny es un archivo `app.R` que contiene la definición de la UI y el Servidor o conjunto de instrucciones para hacer el build, estas aplicaciones terminan siendo un documento HTML que puede ser alojado en un servidor. Por más información acceder a: <https://shiny.rstudio.com/>

Figura 37: Año de relevamiento de los puntos del IFN.



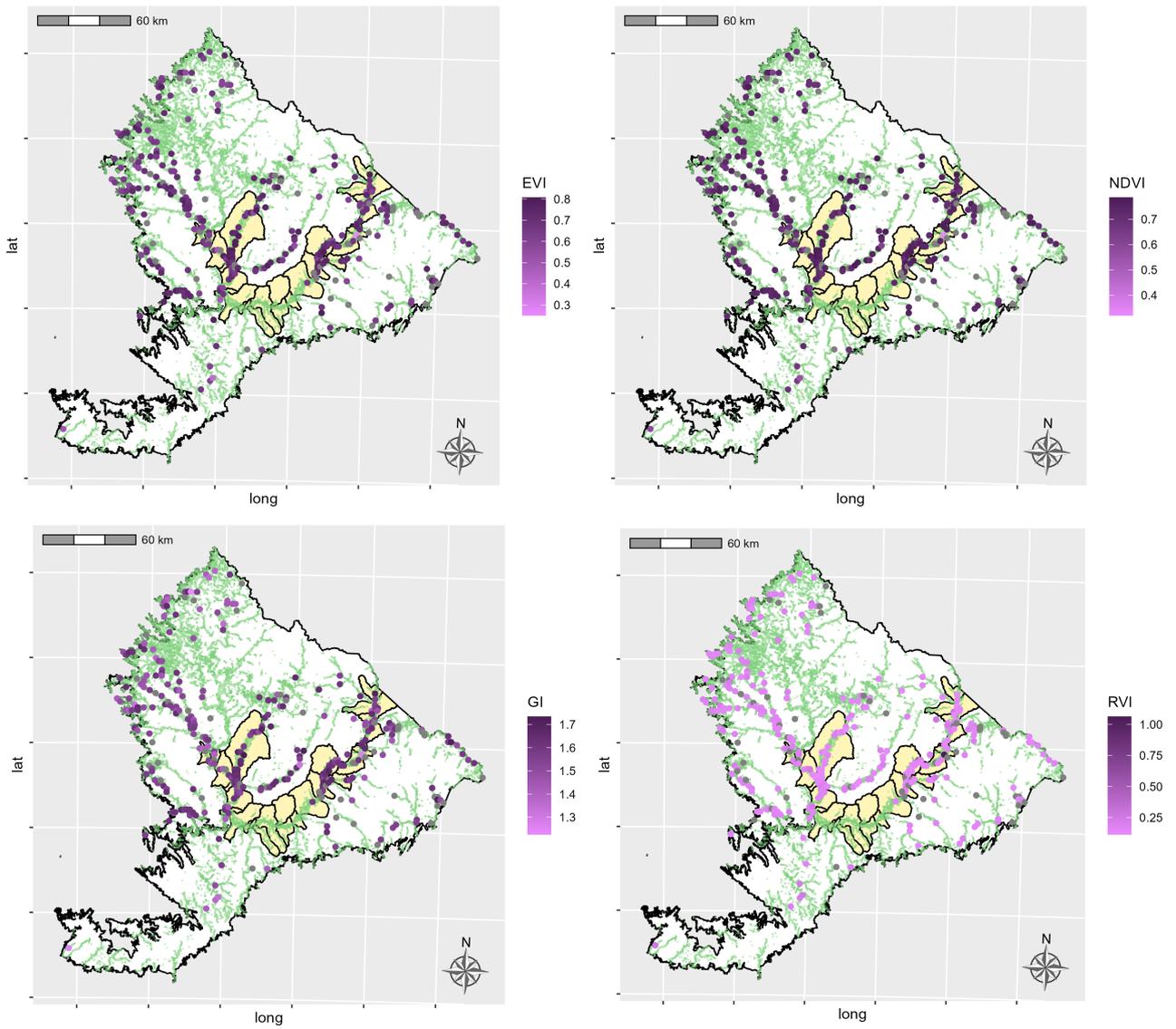
Fuente: Elaboración propia.

Figura 38: Altitud, pendiente, elevación y TPI (extraído del Modelo Digital de Terreno)



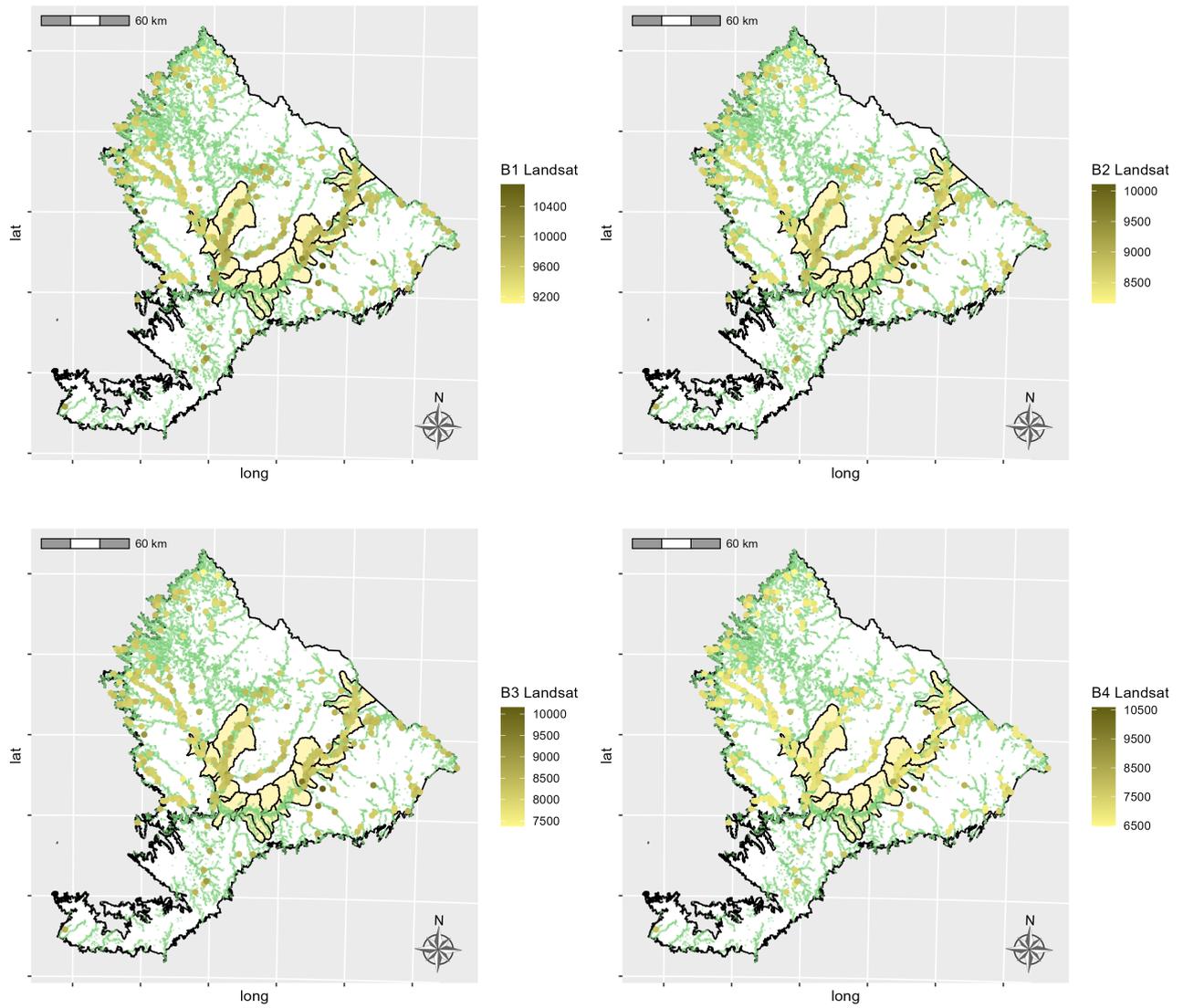
Fuente: Elaboración propia.

Índices de Vegetación



Fuente: Elaboración propia.

Bandas Landsat



Fuente: Elaboración propia.

Bandas Landsat

