

**UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS AGRONÓMICAS
ESCUELA DE PREGRADO**

Memoria de Título

**MODELACIÓN GEOESTADÍSTICA DEL CONTENIDO DE CARBONO ORGÁNICO DEL
SUELO ENTRE LAS REGIONES DE VALPARAÍSO Y DE LOS RÍOS, CHILE**

José Sergei Padarian Campusano

Santiago, Chile

2011

**UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS AGRONÓMICAS
ESCUELA DE PREGRADO**

Memoria de Título

**MODELACIÓN GEOESTADÍSTICA DEL CONTENIDO DE CARBONO ORGÁNICO DEL
SUELO ENTRE LAS REGIONES DE VALPARAÍSO Y DE LOS RÍOS, CHILE**

**GEOSTATISTIC MODELLING OF SOIL ORGANIC CARBON CONTENT BETWEEN
VALPARAÍSO AND LOS RÍOS REGIONS (32-40° S), CHILE**

José Sergei Padarian Campusano

Santiago, Chile

2011

**UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS AGRONÓMICAS
ESCUELA DE PREGRADO**

**MODELACIÓN GEOESTADÍSTICA DEL CONTENIDO DE CARBONO ORGÁNICO DEL
SUELO ENTRE LAS REGIONES DE VALPARAÍSO Y DE LOS RÍOS, CHILE**

Memoria para optar al Título Profesional de:
Ingeniero Agrónomo

Mención:
Manejo de Suelos y Aguas

José Sergei Padarian Campusano

Profesores Guías	Calificaciones
Jorge Pérez Q. Ingeniero Agrónomo, PhD.	7,0
Oscar Seguel S. Ingeniero Agrónomo, Dr.	7,0
Profesores Evaluadores	
Gerardo Soto M. Ingeniero Forestal, Dr.	6,8
Tomislav Curkovic S. Ingeniero Agrónomo, PhD.	7,0
Colaborador	
Luis Morales S. Prof. de Cs. Nat. y Física, Dr.	

Santiago, Chile

2011

ÍNDICE

RESUMEN	1
Palabras claves	1
ABSTRACT	2
Keywords	2
INTRODUCCIÓN	3
MATERIALES Y MÉTODOS	5
Área de estudio	5
Fuente de datos	5
Análisis exploratorio de datos	6
Autocorrelación y asociación espacial	6
Modelación del semivariograma	7
Predicción espacial	8
Kriging Ordinario (OK)	8
Co-Kriging (CK)	9
Kriging Ordinario con semivariogramas locales (LV)	10
Kriging de residuales de red neuronal (ANNRK)	10
Validación	12
Estimación de CO	12
RESULTADOS Y DISCUSIÓN	14
Análisis exploratorio	14
Modelación del semivariograma	16

Predicción espacial	18
Evaluación numérica	18
Evaluación visual	20
Estimación de CO	25
CONCLUSIONES	27
BIBLIOGRAFÍA	28
APÉNDICES	

RESUMEN

El suelo es uno de los grandes reservorios de carbono del planeta. El cambio que se genera al pasar de ecosistemas naturales a agroecosistemas genera pérdidas importantes del carbono orgánico (CO) acumulado, por lo que es necesario disponer de métodos adecuados que permitan evaluar su estado. Para esto se utilizaron 440 series de suelos con el objetivo de modelar la distribución espacial del CO en los suelos de Chile entre las Regiones de Valparaíso y de Los Ríos (32° 9' 2" – 40° 40' 52" lat. S y 0 – 600 m.s.n.m.) mediante el uso de técnicas geoestadísticas con el fin de estimar la cantidad de CO almacenado en los primeros 25 cm de suelo. Se utilizaron dos *líneas de modelación*, una con todos los datos (0,3 – 20% CO) y otra que excluyó 19 outliers (0,3 – 11,8% CO) y se utilizaron cuatro métodos: Kriging Ordinario, Co-Kriging, Kriging Ordinario con semivariogramas locales y Kriging de residuales de red neuronal. Todos los métodos mostraron buenos resultados ($R^2 > 0,67$) luego de comparar los valores reales de CO con los predichos por los modelos mediante validación cruzada. Sus desempeños mostraron ser dependientes de la *línea de modelación*, reforzando la influencia de los outliers en el proceso de modelación. La cantidad total y promedio por hectárea de los suelos del área de estudio fue estimada en $873,58 \pm 43 \text{ Tg}$ (10^{12} g) y $75,12 \pm 3,18 \text{ Mg ha}^{-1}$ respectivamente.

Palabras claves: Kriging, Redes neuronales, semivariogramas locales, stock de carbono orgánico.

ABSTRACT

Soil is one of the major reservoirs of carbon on Earth. Changes generated when moving from natural ecosystems to agroecosystems cause significant losses of the organic carbon (OC) accumulated on it, making necessary to have appropriate methods to evaluate its status. To achieve this, data of 440 soil series were used with the objective of modelling the spatial distribution of soil organic carbon (SOC) between Valparaíso and Los Ríos Regions (32° 9' 2" – 40 ° 40' 52" lat. S and 0 – 600 m.a.s.l) using geostatistical techniques, in order to estimate the amount of OC stored in the first 25 cm. Two modelling lines were used, one with all the data (0,3 – 20% OC) and another that excluded 19 outliers (0,3 – 11,8% OC), using four methods: Ordinary Kriging, Co-Kriging, Ordinary Kriging with local semivariograms and Artificial Neural Network Residual Kriging. All methods showed good results ($R^2 > 0,67$) after comparing the real values of SOC with those predicted by the models using cross-validation. Performance proved to be dependent on the modelling line, reinforcing the influence of outliers in the modelling process. Total SOC and average SOC per hectare in the study area was estimated at $873,58 \pm 43 \text{ Tg}$ (10^{12} g) and $75,12 \pm 3,18 \text{ Mg ha}^{-1}$ respectively.

Keywords: Kriging, Neural networks, local semivariogram, organic carbon stock.

INTRODUCCIÓN

El recurso suelo es un cuerpo dinámico que se encuentra en constante evolución y que ha acompañado al Hombre desde su aparición en la Tierra. Ya en el imperio chino, 4000 años atrás, clasificaban los suelos según color y estructura con fines fiscales. También los romanos usaban sistemas básicos de clasificación según la adaptabilidad de los cultivos de vid, frutales, etc. (Nuñez, 2000).

Para lograr una correcta caracterización y clasificación de suelos es necesario tener en cuenta que sus propiedades no son constantes. Su variabilidad es resultado de muchos procesos actuando e interactuando a través de un continuo espacio-temporal y su inherente dependencia espacial (Parkin, 1993); cada uno de estos procesos puede ser intrínseco, proveniente de factores geológicos, hidrológicos o biológicos que afectan la pedogénesis, o bien extrínseco, como el tipo de uso del suelo y las prácticas de manejo.

Hengl (2007) menciona que uno de los atributos del suelo que presenta alta variabilidad espacial es el carbono orgánico (CO). En esta variabilidad tiene importancia la zona geográfica y la combinación de atributos topográficos como pendiente, exposición, curvatura del terreno, entre otros (Mueller y Pierce, 2003; Thompson *et al.*, 2006). Además, como gran reservorio de carbono, el suelo es un actor importante a nivel ecosistémico, siendo el destino final de la vasta mayoría del carbono fotosintético fijado en los ecosistemas del planeta (Sparling *et al.*, 2006). Las pérdidas de CO a nivel mundial, debido al cambio histórico de su uso, principalmente al convertir ecosistemas naturales de alta concentración de carbono (bosques y praderas) a agroecosistemas de menor concentración de carbono, han sido estimadas en alrededor de 200 Pg¹ (DeFries *et al.*, 1999) a tasas de $1,6 \pm 1,0$ Pg año⁻¹ (Dixon *et al.*, 1994).

Los atributos del suelo, además de presentar la variabilidad ya mencionada, frecuentemente presentan dependencia espacial, por lo cual, muestras colectadas cercanas entre sí, tienden a estar más correlacionadas que las que se colectan distanciadas (Van Es, 2002). Esta característica es representada fielmente por la Primera Ley de la Geografía de Tobler (1970) que enuncia: “*Todo está relacionado entre sí, pero las cosas más cercanas tienen mayor relación que las lejanas*”.

La estadística paramétrica es inadecuada para el análisis de variables que presenten dependencia espacial, ya que ésta asume que las observaciones son independientes a pesar de su distribución en el espacio (Hamlett *et al.*, 1986). La caracterización de la dependencia espacial, o autocorrelación, es posible de realizar con análisis geoestadísticos como el variograma. Sólo si existe esta autocorrelación, es posible desarrollar un mapa de interpolación, dado que en este caso, tendremos en dicho mapa zonas en las cuales la variable

¹1 Pg = 10¹⁵ g

poseerá niveles más altos o más bajos (Best y León, 2006).

El uso de técnicas geoestadísticas se vuelve útil a la hora de predecir atributos en zonas no muestreadas y puede complementar las técnicas de generación de mapas convencionales, que por lo general están basadas en la experiencia empírica del especialista. Técnicas de estadística espacial como Kriging y sus derivados, inicialmente desarrollado en el área de la minería, han sido ampliamente usados, mejorando considerablemente las predicciones de atributos como el CO del suelo (Terra *et al.*, 2004).

En Chile, el estudio del CO del suelo se acota principalmente a estudios de su balance en diversos sistemas de manejo agroforestales (Aguilera *et al.*, 1999; Apezteguía y Sereno, 2002; Sandoval *et al.*, 2008) o enfocado a cuantificar su contenido en ecosistemas específicos (Muñoz *et al.*, 2007) o en áreas de condiciones edafoclimáticas particulares y acotadas (Matus *et al.*, 2008; Perez-Quezada *et al.*, 2011), pero en lo referente a modelar su contenido a escala regional, aún no hay intentos.

El objetivo general de esta memoria fue estudiar la distribución espacial del contenido de CO superficial de los suelos, de áreas reconocidas por estudios agrológicos, entre las regiones de Valparaíso y Los Ríos de Chile.

Además, se definieron como objetivos específicos: (i) generar un modelo de predicción espacial del contenido de CO en las áreas reconocidas de suelos, entre las regiones de Valparaíso y de Los Ríos, mediante técnicas usadas en geoestadísticas y, (ii) estimar la cantidad de CO en suelos del área de estudio, hasta 25 cm de profundidad sobre la base de la cartografía raster generada.

MATERIALES Y MÉTODOS

Área de estudio

El área de estudio se extiende desde la Región de Valparaíso hasta la Región de los Ríos de Chile, desde 32° 9' 2" lat. S, 73° 43' 35" long. O hasta 40° 40' 52" lat. S, 70° 27' 31" long. O. Se consideró esta área debido a que es donde se concentra la mayor producción agrícola, por lo que hay mayor disponibilidad de información; así mismo se excluyeron zonas altas de la Cordillera de los Andes (sobre 600 m.s.n.m, aproximadamente) dada la escasa información disponible de suelo, evitando así extrapolaciones. El área de estudio quedó definida en aproximadamente 110.000 km².

Las temperaturas medias anuales varían desde 15 °C en la zona norte hasta 11 °C en la zona más austral. Al igual que la temperatura, el volumen de las precipitaciones también varía, desde 200 – 400 mm año⁻¹ en la zona norte, hasta 1300 – 2000 mm año⁻¹ en la zona sur (INIA, 1989). Las temperaturas y precipitaciones son fuertemente afectadas por la influencia del relieve, principalmente por una cordillera que bordea la costa, limitando el área de influencia marina. Dichas condiciones, en conjunto con la diversidad geológica del país (Charrier *et al.*, 2007) generan una gran gama de condiciones pedogénicas.

Fuente de datos

Los datos utilizados provinieron de dos fuentes: a) la información georeferenciada de 440 series de suelos de los estudios realizados por el Centro de Información de Recursos Naturales de Chile (CIREN, 1996a, 1996b, 1997a, 1997b, 1999, 2002, 2003) y b) los modelos digitales de elevación (DEM, sigla en inglés de Digital Elevation Model) realizados por el programa Shuttle Radar Topography Mission (SRTM) de la NASA (USGS, 2008).

La información específica utilizada fue el contenido de CO, determinado por el método de Walkley y Black (1934), profundidad de los horizontes, coordenadas de la observación de los estudios del CIREN, la altitud, gradiente de pendiente y forma de la pendiente, entendiéndose esta última como el grado de concavidad (valores negativos) o convexidad (valores positivos) de la superficie (Cuadro 1). El gradiente de la pendiente y la forma de ésta fueron obtenidas a partir del DEM, el cual tiene una resolución de 3 arco-segundos (aproximadamente 90 m).

Cuadro 1. Estadísticas generales de los atributos topográficos utilizados derivados del DEM.

	Altitud (m)	Gradiente de pendiente (%)	Forma de pendiente (x10)
Media	243	3,35	0,02
Min	2	0,20	-5,71
Cuartil 1	107	0,89	-0,50
Mediana	178	1,70	-0,13
Cuartil 3	319	3,27	0,38
Max	1183	43,57	8,50
±DS	190	5,20	1,22

Análisis exploratorio de datos

El análisis exploratorio de datos (EDA, Exploratory Data Analysis) es un grupo de técnicas que resumen las propiedades de los datos y además son capaces de detectar patrones, identificar “outliers” (datos inusuales), errores en los datos, etc. Son técnicas basadas en los datos, por lo que ideas preconcebidas sobre la población son excluidas. Además, tienen la característica de ser resistentes, lo que implica que no son afectadas de sobremanera por outliers (Good, 1983).

Autocorrelación y asociación espacial

La variante espacial de EDA, ESDA (Exploratory Spatial Data Analysis), considera además los aspectos espaciales de los datos: su dependencia y heterogeneidad espacial (Haining, 2003).

Se determinó si existe algún grado de autocorrelación espacial global de los datos mediante la obtención del índice global I de Moran (1950). Las relaciones espaciales se conceptualizaron mediante el inverso de las distancias Euclidianas, de forma que los puntos más alejados tienen una menor influencia que los más cercanos.

Además, se realizó un análisis mediante el uso de indicadores locales de asociación espacial (LISA), específicamente Moran local (Anselin, 1995), el cual permite determinar asociaciones entre una observación y su vecindad, detectar áreas de heterogeneidad espacial e indicar bolsones de no estacionaridad de una forma más específica que el índice de Moran.

Una vez corroborado, mediante el índice I de Moran, que los datos no presentan una distribución aleatoria, se procedió con la modelación de los semivariogramas.

Modelación del semivariograma

En términos generales, técnicas geoestadísticas como Kriging y sus derivados intentan usar de forma óptima los datos, tomando en cuenta cómo varían las propiedades en el espacio, a través del modelo del semivariograma. Su objetivo es predecir los valores en puntos sin muestrear basado en el modelo de un proceso estacionario estocástico:

$$Z(x) = \mu + \varepsilon(x) \quad (1)$$

lo que implica que el valor de Z en el punto x es la media (μ) del proceso más un componente aleatorio (ε) de una distribución con media cero y una función de covarianza (C):

$$C(h) = E[\varepsilon(x)\varepsilon(x+h)] \quad (2)$$

la cual es sólo función de la distancia h entre dos puntos.

La estacionaridad del modelo implica que, en su primer orden, su media $\mu = E[Z(x)]$ es constante para todo valor de x y el problema que se presenta cuando esta estacionaridad de primer orden tiende a no cumplirse, es que la media varía dentro del área de interés y, mientras el área se incrementa, la varianza también lo hace.

En este caso la media no es constante, pero podría serlo, al menos para magnitudes pequeñas de $|h|$ y su diferencia ser cero (Matheron, 1965):

$$E[Z(x) - Z(x+h)] = 0 \quad (3)$$

Luego, al ser reemplazada la covarianza por las varianzas de las diferencias como medida de la relación espacial, la cual, al igual que la Ecuación 2, sólo depende de la distancia h y no de la posición absoluta del punto x , resulta:

$$\begin{aligned} \text{var}[Z(x) - Z(x+h)] &= E[\{Z(x) - Z(x+h)\}^2] \\ &= 2\gamma(h). \end{aligned} \quad (4)$$

Entonces $\gamma(h)$ es la semivarianza (la mitad de la varianza) a una distancia h , considerando los puntos como pares, correspondiendo esta función al *semivariograma* del proceso. La Ecuación 5 muestra uno de los modelos más ampliamente usados, el semivariograma

exponencial.

$$\gamma(h) = \begin{cases} c_0 + c_1 \{1 - \exp(-\|h\|/R)\} & h > 0 \\ 0 & h = 0 \end{cases} \quad (5)$$

donde c_0 , c_1 y R son “nugget”, umbral y rango respectivamente. Nugget corresponde a la varianza relacionada con errores de medición más variaciones que ocurren a distancias menores al intervalo de muestreo, umbral corresponde a la varianza máxima del proceso y rango a la distancia h a la cual se alcanza.

Predicción espacial

Luego del análisis completo de las observaciones mediante ESDA y de estudiar la naturaleza de éstas mediante el semivariograma, se realizó una predicción espacial con el fin de determinar el contenido de CO en sitios no muestreados. En el desarrollo de este trabajo se exploraron diversas herramientas usadas en geoestadística, seleccionándose finalmente cuatro de ellas. A continuación se describen las principales características de cada una de ellas.

Kriging Ordinario (OK)

Es un método de interpolación espacial que encuentra la mejor estimación lineal de una variable aleatoria con estacionaridad de segundo orden, con una media desconocida constante, mediante:

$$\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i), \quad \sum_{i=1}^n \lambda_i = 1 \quad (6)$$

donde $\hat{Z}(x_0)$ es el valor estimado por Kriging en un punto sin muestrear x_0 ; $Z(x_i)$ un valor real en un punto vecino x_i ; y λ_i el peso del factor $Z(x_i)$.

Su error de estimación $R(x_0)$, está dado por:

$$R(x_0) = \hat{Z}(x_0) - Z(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) - Z(x_0) \quad (7)$$

donde $Z(x_0)$ es el valor verdadero en el punto x_0 .

En este método, un único semivariograma es modelado a partir de todos los datos, obteniendo una representación de la estructura de la covarianza espacial, dentro de toda el área de estudio. Asumiendo la estacionaridad del proceso, la predicción está limitada a las observaciones en el vecindario de cualquier punto desconocido y sus ponderaciones son obtenidas al minimizar su error de predicción bajo una condición sin sesgo (Cressie, 1991). La varianza de la predicción $\sigma^2(x_0)$ (el error cuadrado medio minimizado) también es calculada en cada locación mediante:

$$\sigma^2(x_0) = \psi + \sum_{i=1}^n \lambda_i \gamma(x_i, x_0) \quad (8)$$

donde ψ es el multiplicador de Lagrange y $\gamma()$ es el modelo del semivariograma.

Co-Kriging (CK)

Es una variación multivariable de Kriging (Ecuación 6), donde la predicción se realiza tomando en cuenta, además de los datos de la propiedad de interés, los datos de propiedades secundarias correlacionadas. Su cálculo se basa en la siguiente ecuación:

$$\hat{Z}_1(x_0) - \mu_1 = \sum_{i_1=1}^{n_1} \lambda_{i_1} [Z_1(x_{i_1}) - \mu_1(x_{i_1})] + \sum_{j=2}^{n_v} \sum_{i_j=1}^{n_j} \lambda_{i_j} [Z_j(x_{i_j}) - \mu_j(x_{i_j})] \quad (9)$$

donde μ_1 es la media estacionaria de la variable principal; $Z_1(x_{i_1})$ es el valor de la variable principal en el punto i_1 ; $\mu_1(x_{i_1})$ es la media de los datos dentro del área de búsqueda; n_1 es el número de puntos muestreados dentro del área de búsqueda, para el punto x_0 , usados para realizar la estimación; λ_{i_1} es el peso seleccionado para minimizar la varianza de la estimación de la variable principal; n_v es el número de variables secundarias, n_j es el número de puntos de la j -ésima variable secundaria dentro del área de búsqueda, λ_{i_j} es el peso asignado al i_j -ésimo punto de la j -ésima variable secundaria, $Z_j(x_{i_j})$ es el valor en el i_j -ésimo punto de la j -ésima variable secundaria y $\mu_j(x_{i_j})$ es la media de las muestras de la j -ésima variable secundaria dentro del área de búsqueda.

El semivariograma cruzado puede ser estimado con la ecuación:

$$\hat{\gamma}_{12}(h) = \frac{1}{2n} \sum_{i=1}^n [z_1(x_i) - z_1(x_i + h)] [z_2(x_i) - z_2(x_i + h)] \quad (10)$$

donde n es el número de pares de puntos de la variable z_1 y z_2 en el punto x_i y $x_i + h$ (Burrough y McDonnell, 1998).

Para la realización de este método, además de las observaciones con la variable principal (CO), se distribuyeron 1750 puntos al azar dentro de área de estudio con información de las variables secundarias: altitud, gradiente de pendiente y forma de la pendiente.

Kriging Ordinario con semivariogramas locales (LV)

Este método, inicialmente desarrollado por Haas (1990), también se basa en la Ecuación 6. La principal diferencia de este método, es que el semivariograma es estimado y modelado en la vecindad de cada punto x_0 . Para la modelación se requiere establecer áreas de muestreo variable.

Se utilizó el programa VESPER (Minasny *et al.*, 2006), desarrollado por la Universidad de Sydney, para predecir puntos distanciados cada 500 metros. Para la modelación de los semivariogramas se consideraron vecindarios de área variable que contienen entre 20 y 40 observaciones.

Un modelo exponencial, como el que se muestra en la Ecuación 5, fue ajustado de forma automática usando una optimización no lineal de cuadrados mínimos, ponderados por el número de pares usados para determinar la semivarianza $\gamma(h)$ en el intervalo $|h|$. Se eligió este modelo por ser el que presenta mayor estabilidad cuando se realizan ajustes automáticos, si se compara con modelos más complejos como el Gaussiano (ACPA, 2006).

Kriging de residuales de red neuronal (ANNRK)

Las redes neuronales son herramientas de análisis basadas en modelos de estructura neuronal y funciones de procesamiento del cerebro (Anderson, 1995).

Como muestra la Figura 1, una red neuronal típica consiste en elementos de proceso interconectados que están organizados en capas de dependencia secuencial: una capa de entrada, una o más capas ocultas, y una capa de salida. La capa de entrada contiene las variables a analizar, la de salida los resultados obtenidos por el sistema y la capa oculta consiste en una serie de nodos asociados con funciones de transferencia. Cada capa de la red neuronal está conectada por ponderaciones que deben ser determinadas mediante un algoritmo de aprendizaje (Somaratne *et al.*, 2005).

Uno de los puntos críticos al momento de diseñar una red neuronal, es determinar el número de nodos de la o las capas ocultas. Usar pocos nodos implica que la red no sería capaz de captar con plenitud las interacciones entre las variables; usar muchos puede incrementar los tiempos de aprendizaje y, lo más importante, generar un problema llamado “overtraining” (sobre-entrenamiento), el cual aumenta la capacidad de proceso de la red, lo

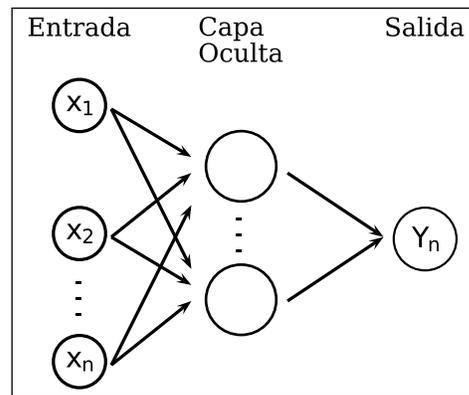


Figura 1. Esquema de red neuronal con una capa oculta.

que conlleva al aprendizaje de aspectos irrelevantes del fenómeno, haciendo que el modelo pierda generalidad (Duin, 2000).

Varios autores sugieren relaciones para determinar el número óptimo de nodos. Hecht-Nielsen (1987) sugiere que, para que una red neuronal sea capaz de aproximar cualquier función continua, es necesario que se cumpla:

$$N_h \leq 2N_i + 1 \quad (11)$$

donde N_h es el número de nodos ocultos y N_i el número de variables de entrada. Rogers y Dowla (1994) aconsejan que, para evitar el “overtraining” de la red, debe cumplirse que:

$$N_h \leq N_{tr}/(N_i + 1) \quad (12)$$

donde N_{tr} es el número de muestras usadas en el entrenamiento.

Wanas *et al.* (1998) demuestran de forma empírica que el mejor rendimiento de una red neuronal se obtiene cuando:

$$N_h = \log(N_{tr}) \quad (13)$$

Ya que sólo se pretende que la red neuronal capte la estructura no lineal de gran escala (tendencias) de los datos, en el desarrollo de esta memoria se utilizó una capa oculta de dos nodos, satisfaciendo las Ecuaciones 11, 12 y 13.

Se utilizó el programa JMP (versión 5.1; SAS Institute Inc., Cary, NC) y se generó la red con el método de Gauss-Newton, suministrando coordenadas, altitud, gradiente de pendiente y forma de pendiente como entradas, el contenido de CO como salida, realizando 2000 iteraciones.

Si bien los residuales son un indicador de la calidad del entrenamiento de la red, pudiendo incluso usarse por sí misma como modelo de mapeo, en este método es necesario realizar un análisis de la estructura espacial de éstos (Demianov *et al.*, 1998). Si no existe correlación espacial entre los residuales significa que toda la información ha sido extraída por la red.

Luego de interpolar los residuales mediante el método LV, ya descrito, obteniendo una modelación de las variaciones de pequeña escala presentes en éstos, se procedió a realizar una suma algebraica de los dos modelos (LV de residuales + red neuronal).

Validación

La validación de los diferentes modelos se realizó mediante validación cruzada “leave one out” (Geisser, 1975; Browne, 2000), donde cada dato es sucesivamente descartado de la modelación y usado en la validación. Esta validación es realizada por defecto al realizar la modelación de los métodos basados en Kriging y fue replicada en la red neuronal. Se escogió este tipo de validación debido a que los datos disponibles no eran muchos y resultaba improductivo generar un set de validación excluyendo un porcentaje de ellos.

Luego de las estimaciones, los valores predichos fueron comparados con los valores reales y conocidos de cada locación para determinar la calidad de predicción de los modelos, utilizando como criterio de calidad: (i) el coeficiente de determinación (R^2) y (ii) la raíz de la media de los errores al cuadrado ($RMSE$), la cual indica la desviación estándar del error.

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m [\hat{Z}(x_i) - Z(x_i)]^2} \quad (14)$$

siendo m el número de puntos totales.

Estimación de CO

Una vez definidos los modelos se generó cartografía raster con una resolución de 500 m para cada uno de los escenarios. Se procedió a realizar una comparación visual de éstos y posteriormente se cuantificó la cantidad total y promedio de CO por hectárea, para lo cual se subdividió el área de estudio en siete subsectores que corresponden a divisiones administrativas del país (Figura 3). Para efectos prácticos se asumió una densidad aparente de los suelos constante e igual a $1,0 \text{ Mg m}^{-3}$ (Ellert *et al.*, 2002).

Si bien suelos de texturas gruesas podrían presentar valores de densidad aparente del orden de $1,4 \text{ Mg m}^{-3}$ y los Andisoles del orden de $0,7 \text{ Mg m}^{-3}$, también existe una distribución diferenciada del contenido de CO en profundidad: más superficial en la zona norte del área de estudio (con una fuerte disminución en el horizonte B) y a mayor profundidad para los suelos de la zona sur. El valor de densidad aparente de $1,0 \text{ Mg m}^{-3}$ se reduce a un concepto de profundidad, ponderando una zona de almacenamiento menor en los suelos con densidades aparentes mayores a $1,0 \text{ Mg m}^{-3}$ y una profundidad de almacenamiento mayor en suelos con densidades aparentes menores a $1,0 \text{ Mg m}^{-3}$, siendo esta aproximación posible de mejorar en trabajos futuros con información más detallada y confiable.

RESULTADOS Y DISCUSIÓN

Análisis exploratorio

El histograma y gráfico de caja de las 440 locaciones (Figura 2) presenta una clara asimetría (factor de asimetría de 2,49), con valores de CO entre 0,3 y 20%, una media de 2,68% y mediana de 1,8%. También es posible observar, en el gráfico de caja, la presencia de posibles outliers con altos contenidos de CO.

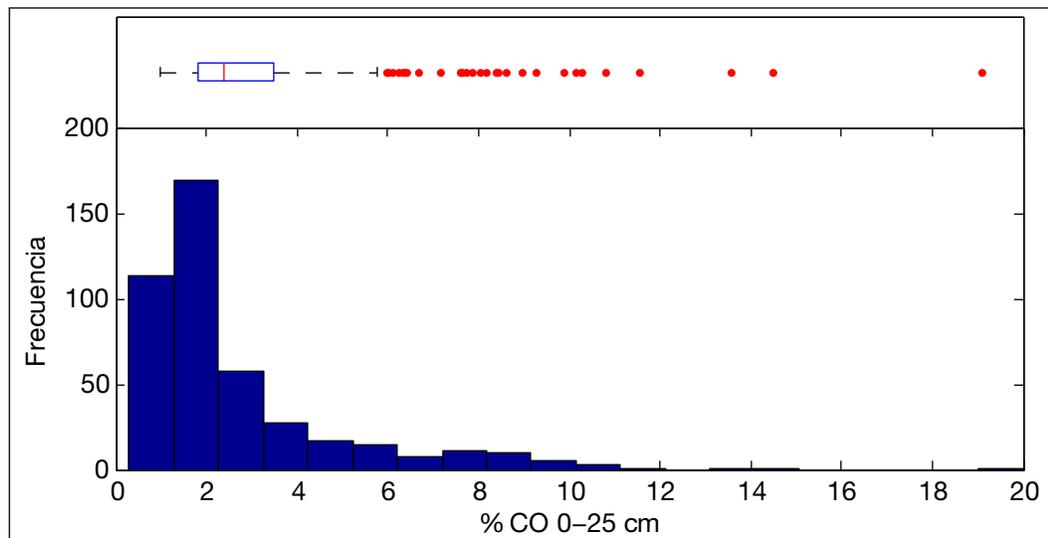


Figura 2. Histograma y gráfico de caja de los valores del contenido de CO.

El indicador Moran local (Apéndice I) indicó zonas con la presencia de outliers, que presentaron niveles de CO muy altos si se comparan con zonas circundantes. Además, se observaron agrupaciones (“clusters”) de datos con contenidos bajos o altos de CO, especialmente en la zona sur del área de estudio.

Los outliers detectados, tanto por el histograma como mediante Moran local (LISA), corresponden a bolsones de no estacionaridad debidos a características particulares de formación de suelo, como áreas de depositación lacustrinas (Luzio, 2010).

Debido a la existencia de outliers, se decidió proseguir con dos líneas de modelación, una utilizando un set con todos los datos y otra luego de excluir los outliers que se encontrasen fuera del rango $\text{cuartil} \pm 1,5 * \text{intercuartil}$ de su vecindad. Los outliers se ubican según muestra la Figura 3.

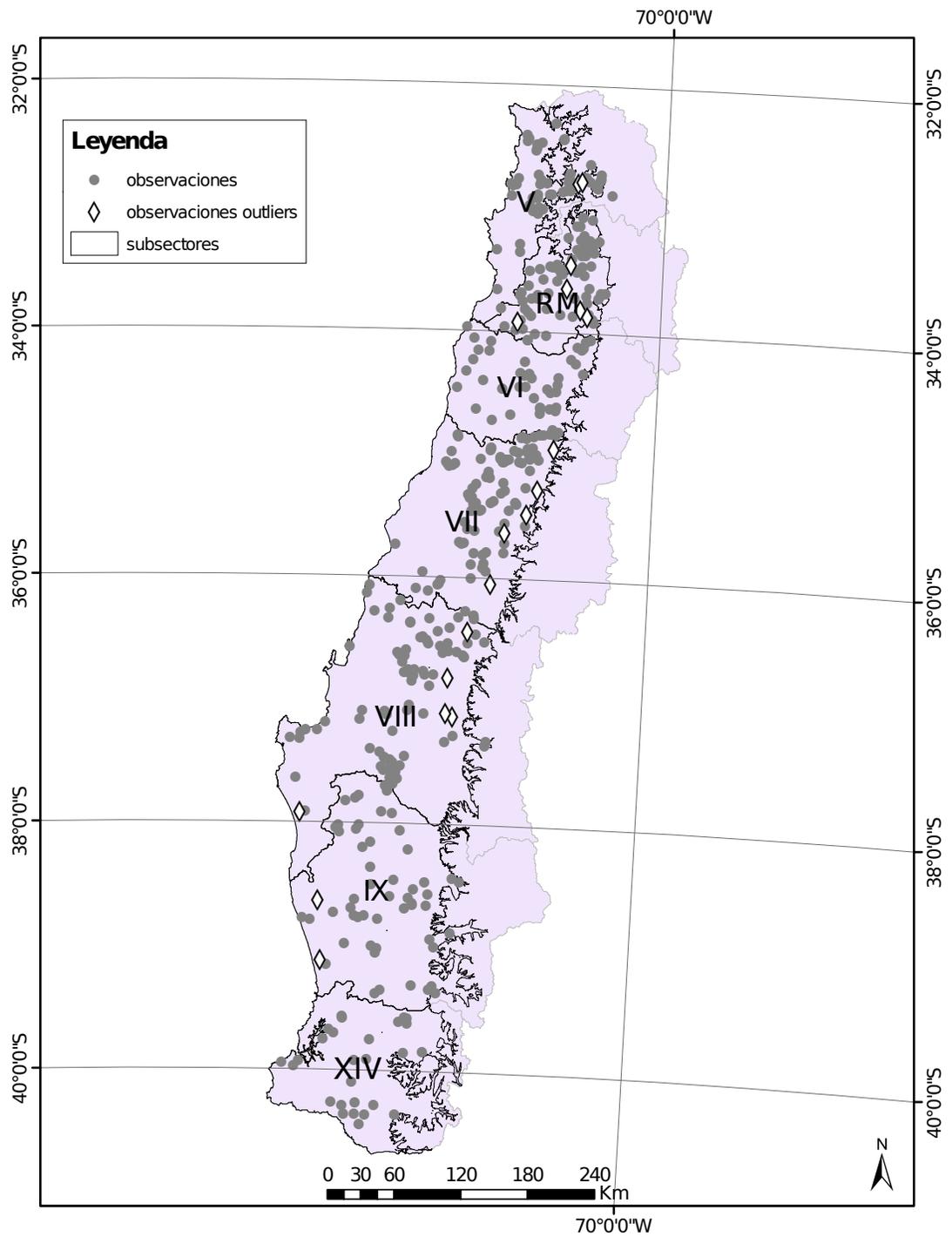


Figura 3. Subsectores (correspondientes a divisiones administrativas) en los que se dividió el área de estudio para realizar la cuantificación de CO, ubicación de las observaciones y de los outliers.

Luego de descartar los outliers, el rango de los valores de CO fue 0,3 – 11,8%, con una media de 2,45%, un factor de asimetría de 2,15 y mediana de 1,8%. Esta última no se vio afectada debido a que es una medida resistente a la presencia de los outliers a diferencia de la media, la cual es muy sensible, especialmente cuando el set de datos es pequeño (Haining, 2003).

El índice I de Moran global obtenido para la línea de modelación con todos los datos, fue de 0,30 (Z Score=6,91) y para la sin outliers de 0,33 (Z Score=7,20). Esto indica que, con un grado de significancia del 1% en ambos casos, los datos presentan un grado de agrupación no atribuible a un proceso aleatorio, por lo que fue posible realizar la modelación de los semivariogramas.

Modelación del semivariograma

Los semivariogramas para toda el área de estudio de ambas *líneas de modelación* presentaron una estructura similar (Figura 4). Se observó que la varianza aumenta sin llegar a un umbral (“unbounded”), lo que podría ser debido a que la distancia abarcada por el semivariograma es menor al rango de correlación (Webster y Oliver, 2007), por lo que es imposible estimar una media global. Además es posible diferenciar dos estructuras de variación, una en el rango de los 400 km y otra de mayor escala, con un rango no determinado y superior a 800 km, siendo posible realizar un análisis diferenciado de éstas en futuras investigaciones.

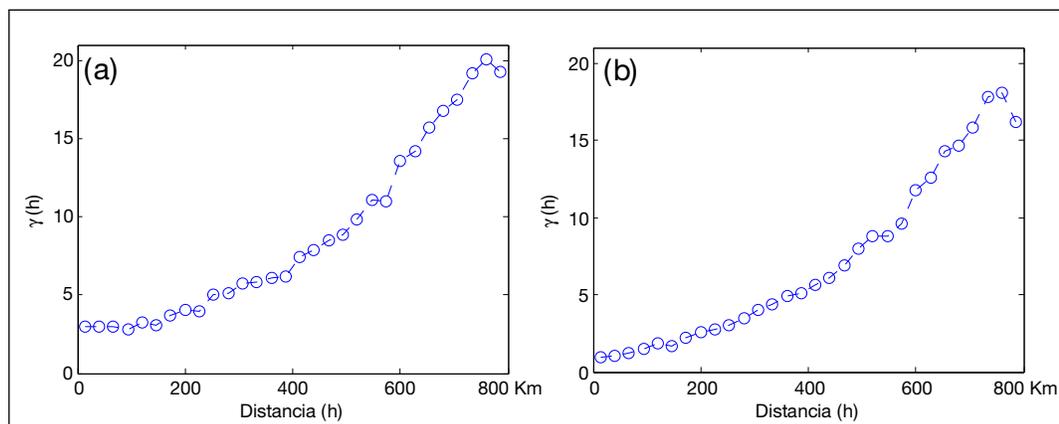


Figura 4. Semivariogramas generales: (a) Set con todos los datos; (b) Set sin outliers.

Para el método OK se ajustó un modelo exponencial con valores nugget, umbral y rango de 2,26; 2,08 y 241.701,30 respectivamente para la línea de modelación con todos los datos y de 0,63; 0,56 y 54.190,15 para la línea de modelación que excluye los outliers.

Un modelo exponencial también fue utilizado para modelar los semivariogramas cruzados en el método CK, con un rango de 264.871,76 m y nugget de: 2,27; 133.032,95; 361,16 y 0,20 para el contenido de CO, altitud, gradiente de pendiente y forma de la pendiente,

respectivamente. Los umbrales para cada uno de los covariogramas se presentan en el Cuadro 2.

Cuadro 2. Umbrales parciales utilizados en los semivariogramas cruzados en el método CK.

	%CO	Altitud	Gradiente de pendiente	Forma de pendiente
%CO	2,26	-349,00	-4,74	0,0091
Altitud		974.866,09	7.560,31	-9,77
Gradiente de pendiente			126,23	-0,10
Forma de pendiente				0,00011

Para el caso del método ANNRK, luego del entrenamiento de la red neuronal, el índice de Moran global de los residuales obtenidos fue de 0,12 (Z Score=2,93), siendo menor al presentado en los datos originales, pero aún así mostrando un grado de agrupación (significancia del 1%) de los contenidos de CO en el espacio. Esto indica que la red neuronal sólo explica una parte de la variación espacial del fenómeno, por lo que es posible continuar con la modelación de sus residuales mediante Kriging.

En la Figura 5 es posible apreciar que la red neuronal entrenada en la línea de modelación en que se excluyeron los outliers (Figura 5b) presenta un mejor resultado en comparación a la que incluye todos los datos (Figura 5a), debido principalmente a que sin los outliers resulta más simple realizar generalizaciones acerca de los datos, lo que es reflejado en una menor varianza en los residuales.

En la Figura 5a se observa que el semivariograma de los residuales de la línea de modelación que incluye todos los datos presenta una estructura no monotónica, un fenómeno denominado “hole effect” (Journel y Huijbregts, 1978). Estas estructuras generalmente indican ciclicidad, pero no necesariamente la presencia de estratas regularmente distribuidas; en este caso particular, se debería entender como agrupaciones de lentes irregulares de tipos de suelo. Este fenómeno es común y generalmente enmascarado por la escala de estudio. En el semivariograma general, si se limita la distancia máxima de alcance (rango) a 400 km, comparado con los 800 km en la Figura 4, es posible observar un comportamiento similar (Figura 6).

Modelar este tipo de estructuras es complejo, siendo necesario ajustar el proceso a semivariogramas de naturaleza sinusoidal o cosinusoidal, lo que escapa de los alcances de este trabajo. En vista que se presentó este tipo de fenómeno, se decidió utilizar la técnica que utiliza semivariogramas locales.

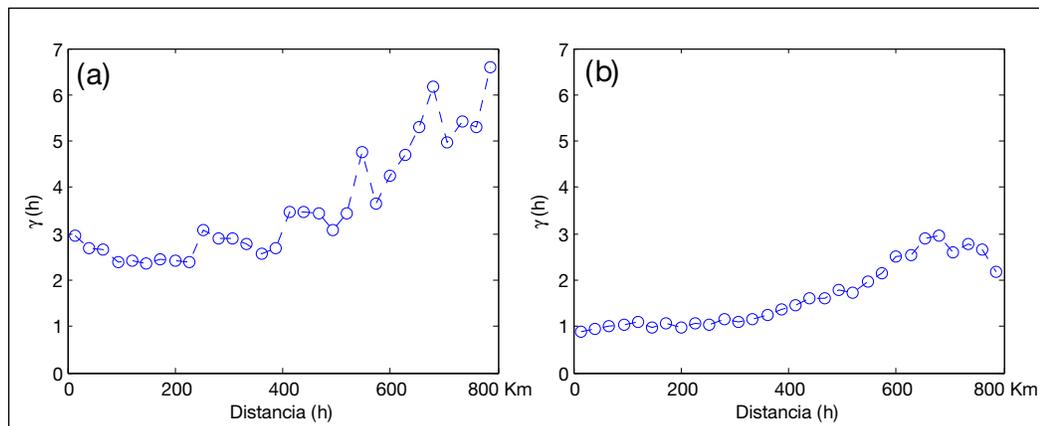


Figura 5. Semivariogramas de residuales de red neuronal: (a) Set con todos los datos; (b) Set sin outliers.

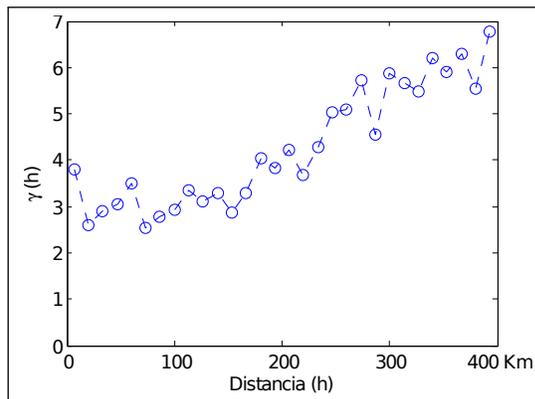


Figura 6. Semivariograma general de la Figura 4a con un rango limitado a 400 km, mostrando el fenómeno “hole effect”.

Predicción espacial

Evaluación numérica

En el Cuadro 3 se muestran los criterios de calidad de las predicciones en los diversos escenarios. La línea de modelación en la cual se descartaron los outliers presenta mejor rendimiento en todos los métodos de modelación, en comparación con la que se utilizaron todos los datos. Esto era de esperar, ya que, según Kerry y Oliver (2007), los outliers en la variable a predecir afectan de forma particular la modelación, debido a que los puntos son considerados como pares y el outlier estaría presente en varios de éstos, además del hecho que la modelación del semivariograma está basado en el cuadrado de sus diferencias (Ecuación 4), acentuando su efecto.

Cuadro 3. Estadísticas de validación de la calidad de predicción de cada método, por línea de modelación.

<i>Línea de modelación</i>		OK	CK	LV	ANNRK
Todos	R^2	0,67	0,67	0,69	0,69
	RMSE (%CO)	1,43	1,43	1,38	1,38
Sin outliers	R^2	0,88	0,90	0,84	0,82
	RMSE (%CO)	0,74	0,68	0,85	0,90

OK: Kriging Ordinario; CK: Co-Kriging;

LV: Kriging Ordinario con semivariogramas locales.

ANNRK: Kriging de residuales de red neuronal.

RMSE: Raíz media de los errores al cuadrado.

El similar rendimiento de los métodos OK y CK, también descrito por Goovaerts (1994), en ambas *líneas de modelación*, se debería a que la correlación entre el contenido de CO y la altitud, gradiente de pendiente y forma de la pendiente del set de datos no es significativa (coeficiente de correlación de Pearson ($p \geq 0,05$): 0,09; 0,05 y 0,08 respectivamente) como para que el método CK realice predicciones más precisas que OK (Odeha *et al.*, 1994; Goovaerts, 1999; Triantafilis *et al.*, 2001). Esto se debe a que la zona de estudio es muy extensa, la densidad de observaciones es baja y se encuentran en condiciones pedogénicas muy heterogéneas, presentándose contenidos de CO muy disímiles en rangos similares de los atributos topográficos utilizados.

Los métodos que obtuvieron mejores resultados en la línea de modelación con todos los datos fueron LV y ANN RK, debido a que ambos métodos son más resistentes que OK y CK a la presencia de datos anómalos, ya que la presencia de outliers afecta los semivariogramas generales de estos últimos.

En la línea de modelación sin outliers, CK y OK tienen mejor desempeño. Esto se debe a que sin outliers el proceso es mucho más homogéneo y métodos simples suelen presentar mejores resultados (Terra *et al.*, 2004). El menor rendimiento de ANN RK en esta línea de modelación, al contrario de lo descrito por Chowdhury *et al.* (2010), quienes usaron redes neuronales y obtuvieron mejoras de un 15% al mapear contenidos de arsénico en aguas subterráneas en comparación a Kriging Ordinario, se podría deber a que la disminución en la cantidad de datos y en mayor medida la disminución de la diversidad de escenarios que estos cubren, sea insuficiente para asignar eficientemente los pesos de las funciones de transferencia (Brown, 2004), incurriendo en generalizaciones deficientes.

Ninguno de los métodos debiera ser descartado completamente, ya que, dependiendo de los objetivos del investigador, la naturaleza de los datos y la disponibilidad de variables secundarias correlacionadas, podrían variar los resultados. Si se tiene la intención de generar un modelo que no busque representar zonas con valores extremos, removiendo outliers, o de enfrentarse a un proceso relativamente homogéneo, la utilización de un método simple como OK podría ser una buena alternativa. Si se cuenta con información adicional de buena calidad,

es decir, que las variables secundarias a utilizar presenten alta correlación con la variable primaria, métodos multivariados como CK o ANNRK podrían mejorar las predicciones.

Evaluación visual

Los mapas generados por los distintos métodos y que muestran la distribución espacial del contenido de CO en el área de estudio se presentan en las Figuras 7 a la 10. Estos fueron evaluados bajo los criterios de la caracterización de la variación a escala regional y a escala local. Los contenidos de CO presentan una clara tendencia regional, aumentando de norte a sur, debido principalmente al amplio gradiente de precipitación y temperatura que se genera entre los extremos del área de estudio, el cual representa un factor dominante en los ciclos del CO de los suelos en escalas regionales a globales (Post *et al.*, 1982). Este gradiente determina que en la zona norte del área de estudio se encuentren valores inferiores a 3%, mientras que en la zona sur los niveles son mayores a 6%. A partir de los 35° lat. S. se aprecia un aumento del contenido de CO hacia el extremo Este del área de estudio, dado por la aparición de los primeros Andisoles y luego un aumento importante aproximadamente a los 38° lat. S. debido principalmente al cambio de un régimen de humedad xérico a údico (Luzio, 2010), que favorece la mayor generación de biomasa y acumulación de CO en el suelo.

La representación de las variaciones locales del contenido de CO difiere en los distintos escenarios. En la línea de modelación con todos los datos es posible apreciar el habitualmente descrito efecto de suavizado de los métodos OK y CK (Watson, 1984; Isaaks y Srivastava, 1989), por lo que no se evidencian las zonas con contenidos extremos de CO, especialmente las con contenidos bajos, en comparación a LV y ANNRK.

En la línea de modelación que excluye los outliers, los mapas generados son similares, con algunas diferencias en los métodos CK y ANNRK, los cuales presentan mayor cantidad de zonas con contenidos bajos de CO.

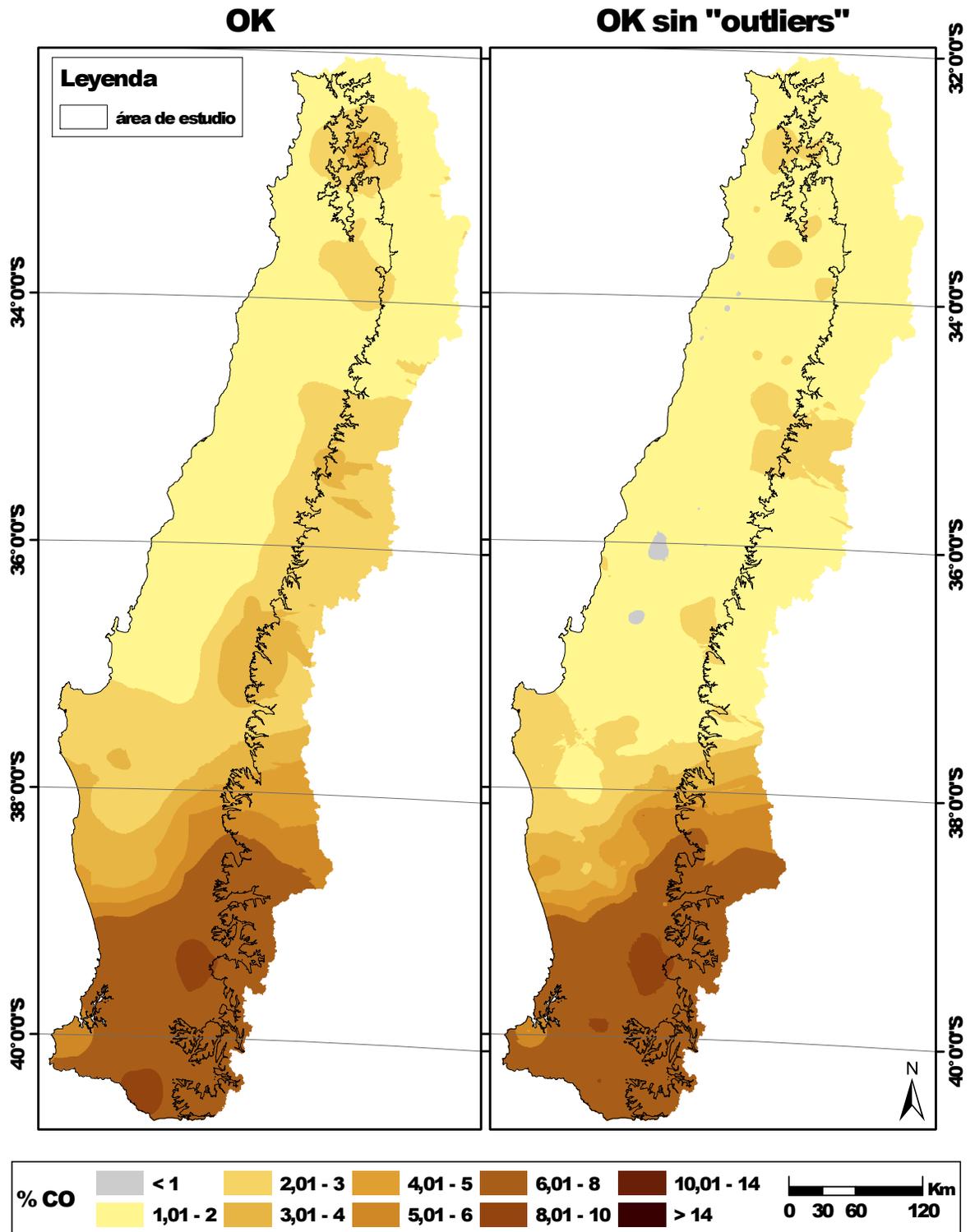


Figura 7. Mapas generados con el método Kriging Ordinario (OK) en ambas líneas de modelación.

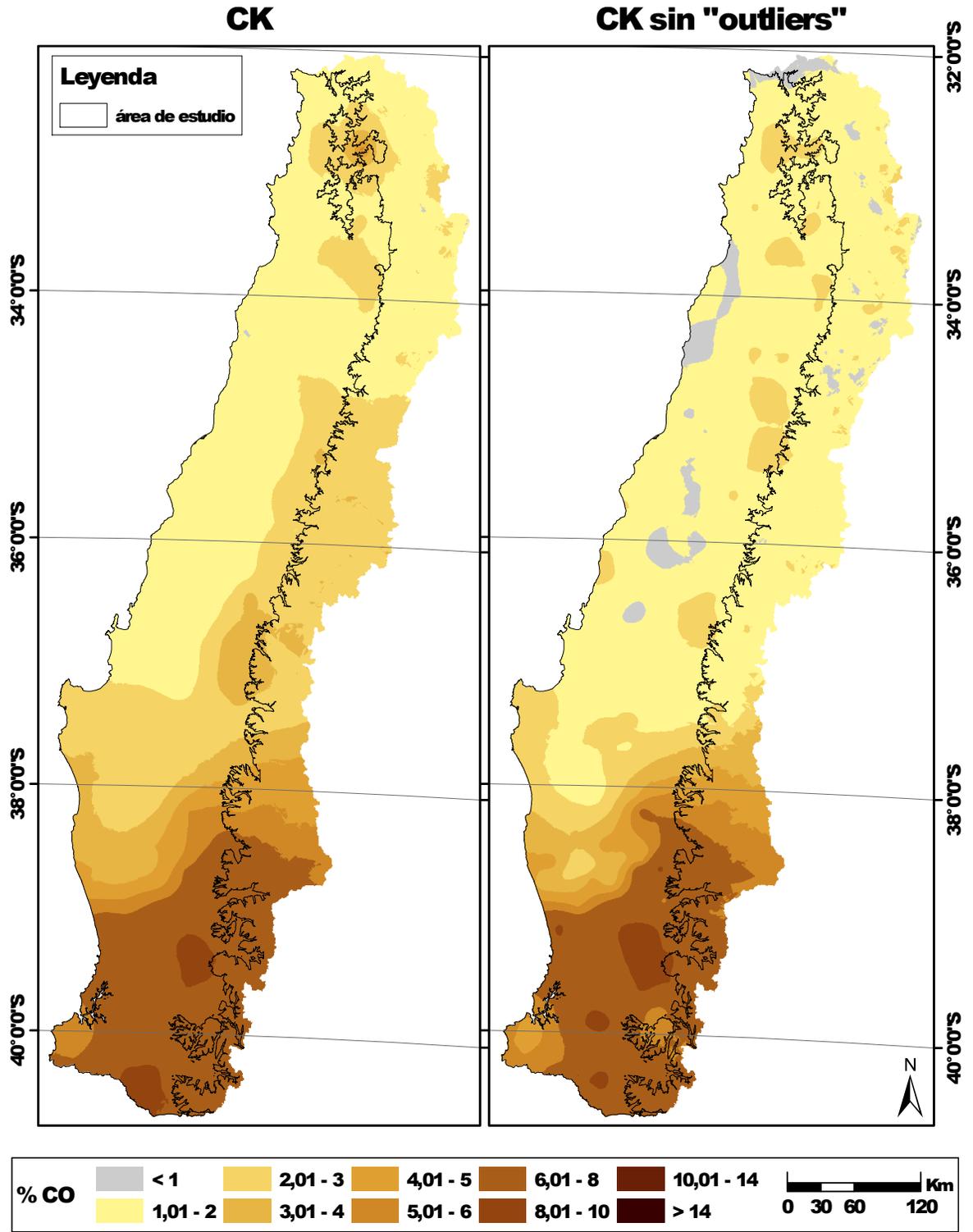


Figura 8. Mapas generados con el método Co-Kriging (CK) en ambas líneas de modelación.

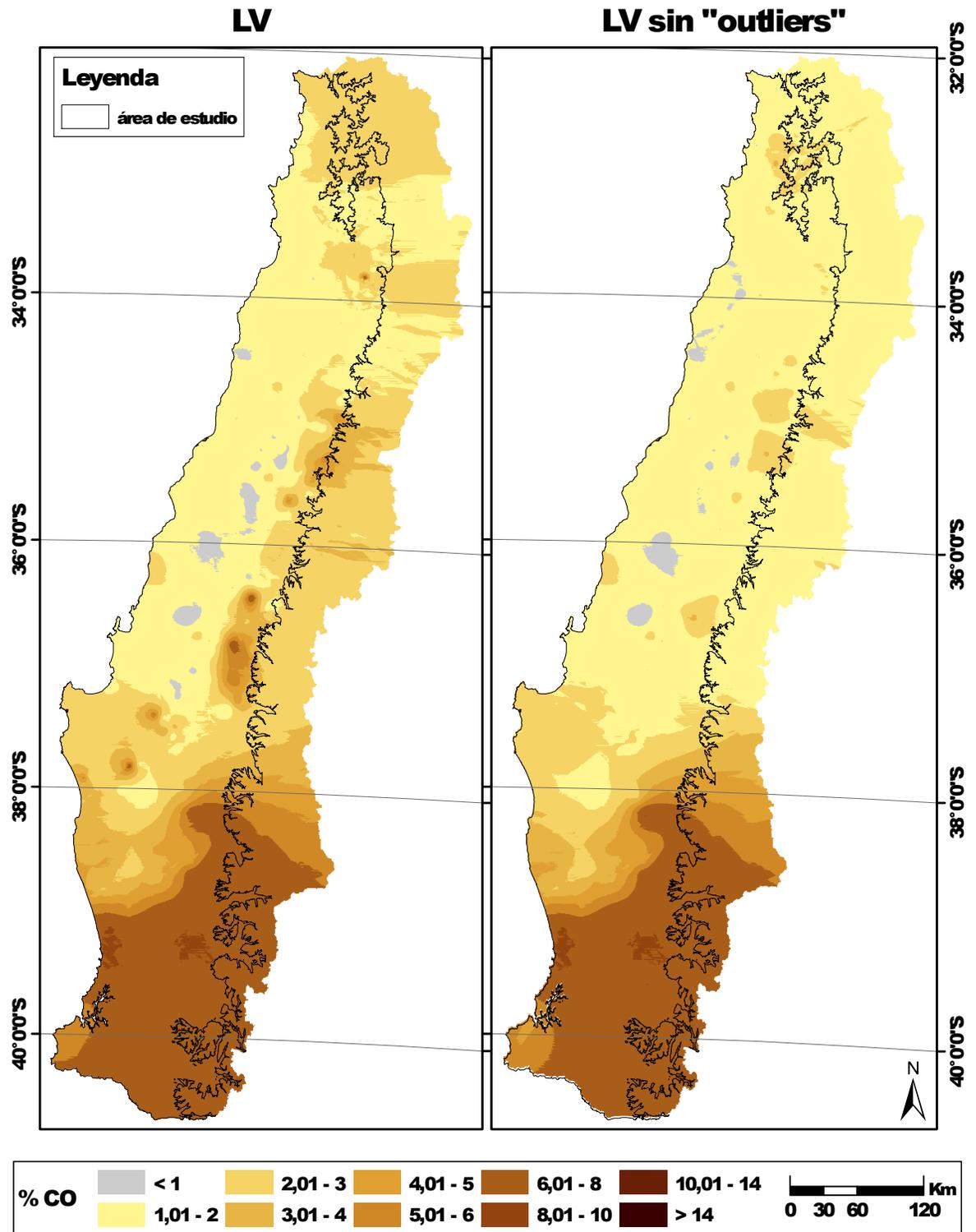


Figura 9. Mapas generados con el método Kriging Ordinario con semivariogramas locales (LV) en ambas líneas de modelación.

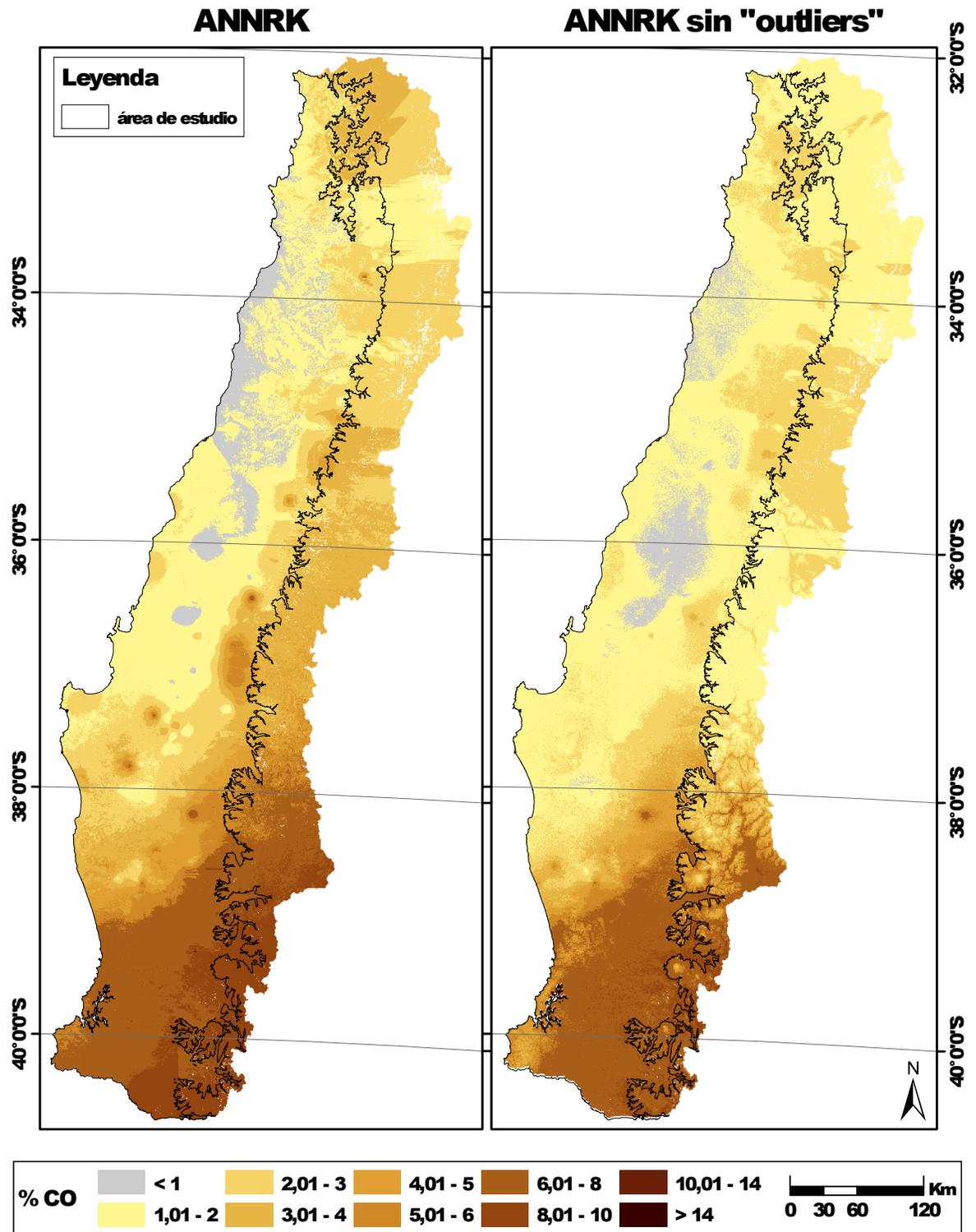


Figura 10. Mapas generados con el método Kriging de residuales de red neuronal (ANNRK) en ambas líneas de modelación.

Estimación de CO

El Cuadro 4 muestra la estimación del contenido total y promedio por hectárea de CO en cada uno de los escenarios modelados.

Cuadro 4. Estimación del contenido total y promedio por hectárea de CO en los primeros 25 cm de suelo, por método, línea de modelación y subsector (zonas administrativas).

		RM	V	VI	VII	VIII	IX	XIV	Total / \bar{x}^*
OK	Total (Tg)	36,29	37,38	41,07	80,39	162,10	288,19	263,81	909,23
	\bar{x}^*	47,83	45,90	39,18	41,98	59,49	127,15	181,14	77,52
OK sin outliers	Total (Tg)	32,44	33,84	40,03	68,42	124,92	282,41	259,71	841,78
	\bar{x}	42,76	41,55	38,19	35,73	45,85	124,60	178,32	72,43
CK	Total (Tg)	35,95	37,24	40,91	79,25	161,23	287,53	263,42	905,53
	\bar{x}	47,38	45,72	39,03	41,38	59,17	126,86	180,87	77,20
CK sin outliers	Total (Tg)	31,56	31,65	39,03	68,12	127,54	288,30	251,54	837,74
	\bar{x}	41,60	38,86	37,24	35,57	46,81	127,19	172,70	71,42
LV	Total (Tg)	36,49	44,72	42,93	80,70	161,65	291,80	264,58	922,86
	\bar{x}	48,11	54,90	40,95	42,14	59,32	128,73	181,58	79,39
LV sin outliers	Total (Tg)	32,43	35,79	40,18	68,33	131,41	282,87	253,51	844,52
	\bar{x}	42,92	44,09	38,34	35,91	48,48	125,47	175,99	73,03
ANNRK	Total (Tg)	33,85	37,40	37,75	78,80	162,82	286,06	276,31	913,00
	\bar{x}	44,66	46,36	36,10	41,25	59,91	126,28	190,70	77,89
ANNRK sin outliers	Total (Tg)	34,04	35,01	41,81	70,93	122,35	257,54	252,29	813,97
	\bar{x}	45,08	44,13	40,19	37,52	45,43	114,29	177,69	72,05
\bar{x} (Tg)		34,13	36,63	40,46	74,37	144,25	283,09	260,65	873,58
\bar{x}^*		45,04	45,19	38,65	38,94	53,06	125,07	179,87	75,12
$\pm DS$ (Tg)		1,93	3,83	1,61	5,88	19,09	10,77	8,31	
$\pm DS^*$		2,52	4,66	1,55	3,01	6,92	4,53	5,30	

*Mg ha⁻¹

1Tg = 10⁶Mg = 10¹²g

OK: Kriging Ordinario; CK: Co-Kriging; LV: Kriging Ordinario con semivariogramas locales;

ANNRK: Kriging de residuales de red neuronal.

DS: Desviación estándar.

Tanto en el contenido total de CO del área, como en el promedio por hectárea, se observó una diferencia significativa ($P < 0,0001$) entre las *líneas de modelación*, considerando los subsectores (divisiones administrativas) como bloques (Apéndice II). Además no existieron diferencias significativas entre los métodos en la cuantificación de ambas magnitudes considerando las *líneas de modelación* como bloques (datos no mostrados).

En los distintos subsectores, los promedios por hectárea fluctúan en el rango de 35,57 a 178,32 Mg ha⁻¹ en la línea de modelación sin outliers y de 36,10 a 190,70 Mg ha⁻¹ en la línea

de modelación con todos los datos. La diferencia promedio entre las *líneas de modelación* es de $5,77 \pm 1,2 \text{ Mg ha}^{-1}$, lo que equivale a $78,16 \pm 15,8 \text{ Tg}$ de CO en toda el área de estudio, lo que refuerza la influencia de los outliers en el proceso de modelación, ya que al incluirlos, la estimación del contenido total de CO en el área de estudio aumenta un 8,95%.

Las mayores diferencias considerando métodos y *líneas de modelación* se presenta en el subsector VIII, debido a la mayor superficie que posee y a la mayor heterogeneidad de la naturaleza de los suelos (Luzio, 2010). Además, en la línea de modelación sin excluir outliers, se observa que los valores máximos del contenido total se presentan en los métodos LV y ANNRK debido al ya comentado efecto de suavizado que presentan los métodos OK y CK.

Las estimaciones realizadas son consistentes con otros estudios. Muñoz *et al.* (2007) presentan valores de CO para el ecosistema mediterráneo árido (Alfisol del subsector VII), el cual abarca los subsectores RM, VI y VII, del orden de $30 - 40 \text{ Mg ha}^{-1}$. Dixon *et al.* (1994) señalan valores de 120 Mg ha^{-1} para ecosistemas de bosques de latitudes bajas (escala global) comparables con los obtenidos para el subsector IX, al igual que Matus *et al.* (2008), quienes realizaron su estudio en Andisoles del subsector IX, presentando contenidos de CO de $122,5 \text{ Mg ha}^{-1}$.

En definitiva, considerando los antecedentes discutidos, entre los cuatro métodos planteados, no es posible recomendar uno particular, ya que su comportamiento depende de la naturaleza de los datos. Sin embargo, nuevas herramientas surgen día a día, y métodos basados en geostatística multivariada como CK comienzan a fusionarse con métodos locales como LV², abriendo nuevas posibilidades para estudiar fenómenos de esta clase. Por otra parte, la problemática del calentamiento global y la emisión y captura de carbono está presente en el pensamiento colectivo, habiéndose generado un incremento notable en el número de investigaciones del carácter de ésta.

²Dr. Budiman Minasny, The University of Sydney, Australian Centre for Precision Agriculture, 2010, Australia (Comunicación personal).

CONCLUSIONES

Basándose en cuatro diferentes métodos, es posible ajustar modelos de predicción espacial del contenido de CO en el área de estudio, presentando todos ellos buenos resultados en la caracterización de los procesos involucrados, con valores de R^2 entre 0,67 y 0,90.

Debido a la detección de outliers, se generaron dos *líneas de modelación*, una con todos los datos y otra que los excluye.

El método Co-Kriging (CK) es el que obtuvo mejores resultados ($R^2 = 0,90$), excluyendo los outliers. Métodos con enfoque local como Kriging Ordinario con semivariogramas locales (LV) y Kriging de residuales de red neuronal (ANNRK), se comportaron de mejor manera ($R^2 = 0,69$), sin excluir los outliers.

La exclusión de los outliers generaron una disminución de 8,95% en las predicciones del contenido total de CO en el área de estudio, reforzando la importancia que implica el análisis correcto de éstos.

Con respecto al contenido de CO almacenado en los primeros 25 cm de suelo de los 110.000 km² que abarca el área de estudio, este trabajo presenta metodologías simples para su estimación a escala regional y que pueden ser utilizados en distintos escenarios, dependiendo de calidad de la información y de los objetivos específicos del investigador.

BIBLIOGRAFÍA

- ACPA. 2006. Vesper User Manual. Australian Centre for Precision Agriculture. The University of Sydney, Sydney, Australia. 25 p.
- Aguilera, S. M., Borie, G. and Peirano, P. 1999. Dinámica del carbono en suelos con distintos sistemas de labranza. *Frontera Agrícola* 5: 33–38.
- Anderson, J. A. 1995. *An Introduction To Neural Networks*. The MIT Press, Cambridge, USA. 672 p.
- Anselin, L. 1995. Local indicators of spatial association-LISA. *Geographical Analysis* 27 (2): 93–115.
- Apezteguía, H. and Sereno, R. 2002. Influencia de los sistemas de labranza sobre la cantidad y calidad del carbono orgánico del suelo. *Agricultura Técnica* 62 (3): 418–426.
- Best, S. and León, L. 2006. pp. : 145–170. *In*: Bongiovanni, R., Chartuni, E., Best, S. and Roel, Á. (eds.). *Agricultura de precisión: Integrando conocimientos para una agricultura moderna y sustentable*. Instituto Interamericano de Cooperación para la Agricultura. Montevideo, Uruguay. 244 p.
- Brown, G. 2004. Diversity in neural network ensembles. Unpublished Doctoral Thesis, University of Birmingham.
- Browne, M. 2000. Cross-validation methods. *Journal of Mathematical Psychology* 44 (1): 108–132.
- Burrough, P. and McDonnell, R. 1998. *Principles of GIS*. Oxford University Press, London, UK. 299 p.
- Charrier, R., Pinto, L., Rodríguez, M., Moreno, T. and Gibbons, W. 2007. *The Geology of Chile*. The Geological Society, London, UK. 440 p.
- Chowdhury, M., Alouani, A. and Hossain, F. 2010. Comparison of ordinary kriging and artificial neural network for spatial mapping of arsenic contamination of groundwater. *Stochastic Environmental Research and Risk Assessment* 24: 1–7.
- CIREN. 1996a. Estudio Agrológico VI Región. Descripciones de suelos, materiales y símbolos. Publicación N° 114. Centro de Información de Recursos Naturales (CIREN), Santiago, Chile. 479 p.
- CIREN. 1996b. Estudio Agrológico Región Metropolitana. Descripciones de suelos, materiales y símbolos. Publicación N° 115. Centro de Información de Recursos Naturales (CIREN), Santiago, Chile. 425 p.

- CIREN. 1997a. Estudio Agrológico V Región. Descripciones de suelos, materiales y símbolos. Publicación N° 116. Centro de Información de Recursos Naturales (CIREN), Santiago, Chile. 366 p.
- CIREN. 1997b. Estudio Agrológico VII Región. Descripciones de suelos, materiales y símbolos. Publicación N° 117. Centro de Información de Recursos Naturales (CIREN), Santiago, Chile. 659 p.
- CIREN. 1999. Estudio Agrológico VIII Región. Descripciones de suelos, materiales y símbolos. Publicación N° 121. Centro de Información de Recursos Naturales (CIREN), Santiago, Chile. 550 p.
- CIREN. 2002. Estudio Agrológico IX Región. Descripciones de suelos, materiales y símbolos. Publicación N° 122. Centro de Información de Recursos Naturales (CIREN), Santiago, Chile. 338 p.
- CIREN. 2003. Estudio Agrológico X Región. Descripciones de suelos, materiales y símbolos. Publicación N° 123. Centro de Información de Recursos Naturales (CIREN), Santiago, Chile. 374 p.
- Cressie, N. 1991. Statistics for spatial data. John Wiley & Sons, New York, USA. 920 p.
- DeFries, R., Field, C., Fung, I., Collatz, G. and Bounoua, L. 1999. Combining satellite data and biogeochemical models to estimate global effects of human-induced land cover change on carbon emissions and primary productivity. *Global Biogeochemical Cycles* 13 (3): 803–815.
- Demyanov, V., Kanevsky, M., Chernov, S., Savelieva, E. and Timonin, V. 1998. Neural network residual kriging application for climatic data. *Journal of Geographic Information and Decision Analysis* 2 (2): 215–232.
- Dixon, R., Brown, S., Houghton, R., Solomon, A., Trexler, M. and Wisniewski, J. 1994. Carbon pools and flux of global forest ecosystems. *Science* 263 (5144): 185–189.
- Duin, R. 2000. pp. : 9–13. *In*: van Vliet, L., Heijnsdijk, J., Kielmann, T. and Knijnenburg, P. (eds.). *ASCI 2000, Sixth Annual Conference of the Advanced School for Computing and Imaging*. Lommel, Belgium.
- Ellert, B., Janzen, H. and Entz, T. 2002. Assessment of a method to measure temporal change in soil carbon storage. *Soil Science Society of America Journal* 66 (5): 1687–1695.
- Geisser, S. 1975. The predictive sample reuse method with applications. *Journal of the American Statistical Association* 70 (350): 320–328.
- Good, I. 1983. The philosophy of exploratory data analysis. *Philosophy of Science* 50 (2): 283–295.

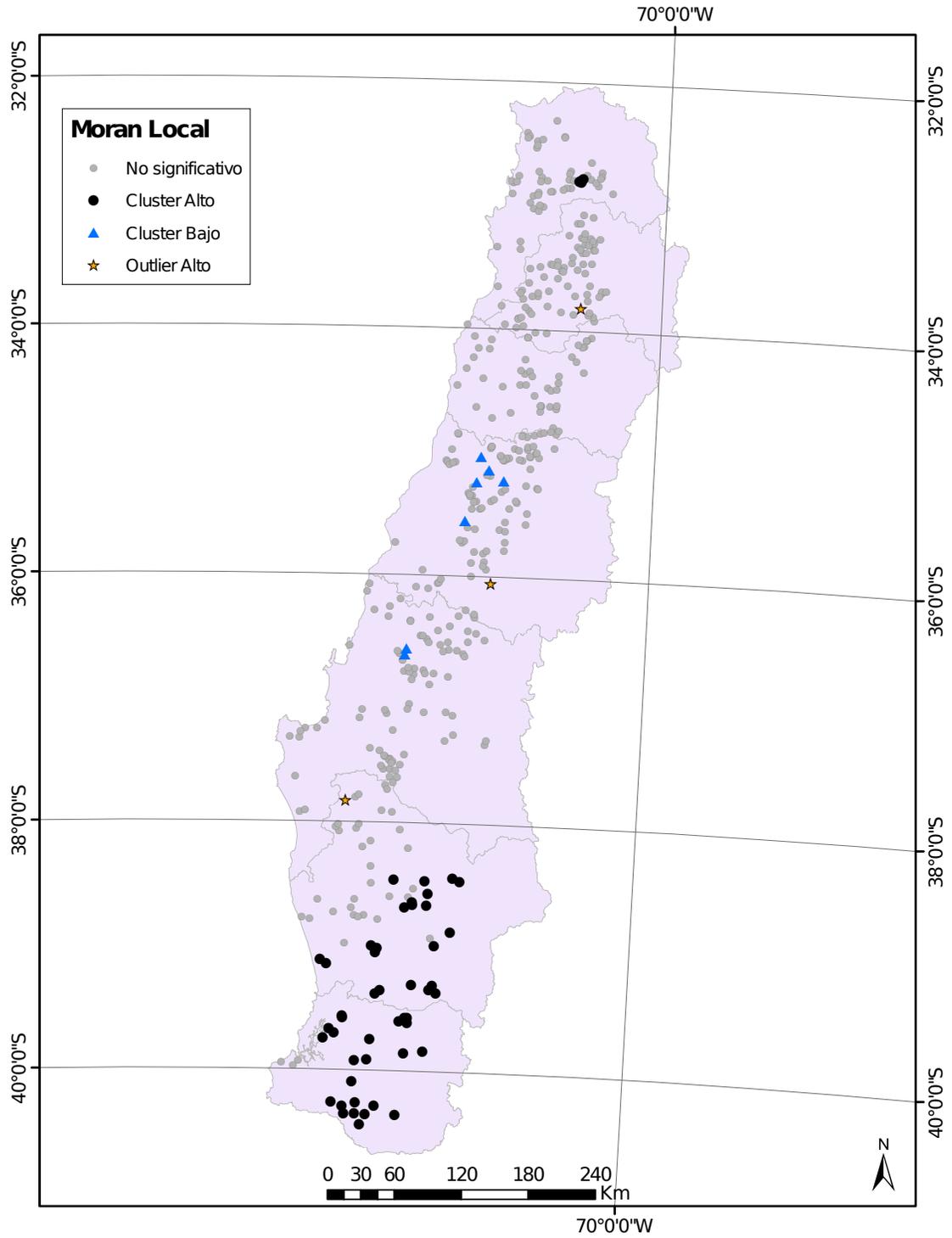
- Goovaerts, P. 1994. Comparative performance of indicator algorithms for modeling conditional probability distribution functions. *Mathematical Geology* 26 (3): 389–411.
- Goovaerts, P. 1999. Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma* 89 (1-2): 1–45.
- Haas, T. 1990. Kriging and automated variogram modeling within a moving window. *Atmospheric Environment. Part A. General Topics* 24 (7): 1759–1769.
- Haining, R. 2003. *Spatial data analysis: theory and practice*. Cambridge Univ. Press, Cambridge, UK. 452 p.
- Hamlett, J., Horton, R. and Cressie, N. 1986. Resistant and exploratory techniques for use in semivariogram analyses. *Soil Science Society of America Journal* 50: 868–875.
- Hecht-Nielsen, R. 1987. pp. : 11–14. *In: Proceedings of the international conference on Neural Networks*. Vol. 3. New York: IEEE Press.
- Hengl, T. 2007. *A practical guide to geostatistical mapping of environmental variables*. European Commission, Ispra, Italia. 143 p.
- INIA. 1989. *Mapa agroclimático de Chile*. Instituto de Investigación Agropecuaria. Ministerio de Agricultura, Chile. 221 p.
- Isaaks, E. and Srivastava, R. 1989. *An introduction to applied geostatistics*. Oxford Univ. Press, New York, USA. 561 p.
- Journel, A. and Huijbregts, C. 1978. *Mining geostatistics*. Academic Press, New York, USA. 600 p.
- Kerry, R. and Oliver, M. 2007. Determining the effect of asymmetric data on the variogram. II. Outliers. *Computers & Geosciences* 33 (10): 1233–1260.
- Luzio, W. (ed.) 2010. *Suelos de Chile*. Universidad de Chile, Santiago, Chile. 364 p.
- Matheron, G. 1965. *Les variables régionalisées et leur estimation*. Masson et Cie., Paris, France. 306 p.
- Matus, F., Garrido, E., Sepúlveda, N., Cárcamo, I., Panichini, M. and Zagal, E. 2008. Relationship between extractable Al and organic C in volcanic soils of Chile. *Geoderma* 148 (2): 180–188.
- Minasny, B., McBratney, A. and Whelan, B. 2006. *VESPER, Version 1.6*. Australian Centre for Precision Agriculture, McMillan Building A05, The University of Sydney. NSW.
- Moran, P. 1950. Notes on continuous stochastic phenomena. *Biometrika* 37 (1-2): 17–23.

- Mueller, T. and Pierce, F. 2003. Soil carbon maps: Enhancing spatial estimates with simple terrain attributes at multiple scales. *Soil Science Society of America Journal* 67 (1): 258–267.
- Muñoz, C., Ovalle, C. and Zagal, E. 2007. Distribución del carbono orgánico del suelo almacenado en el perfil de un Alfisol en ecosistemas Mediterráneos de Chile. *Revista de la Ciencia del Suelo y Nutrición Vegetal* 7 (1): 15–27.
- Núñez, J. 2000. *Fundamentos de edafología*. Editorial Universidad Estatal a Distancia, San José, Costa Rica. 188 p.
- Odeha, I., McBratney, A. and Chittleborough, D. 1994. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. *Geoderma* 63 (3-4): 197–214.
- Parkin, T. B. 1993. Spatial variability of microbial processes in soil. *Journal of Environmental Quality* 22 (3): 409–417.
- Perez-Quezada, J., Delpiano, C., Snyder, K., Johnson, D. and Franck, N. 2011. Carbon pools in an arid shrubland in Chile under natural and afforested conditions. *Journal of Arid Environments* 75: 29–37.
- Post, W., Emanuel, W., Zinke, P. and Stangenberger, A. 1982. Soil carbon pools and world life zones. *Nature* 298: 156–159.
- Rogers, L. and Dowla, F. 1994. Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling. *Water Resources Research* 30 (2): 457–481.
- Sandoval, M., Stolpe, N., Zagal, E., Mardones, M. and Celis, J. 2008. Aporte de carbono orgánico de la labranza cero y su impacto en la estructura de un Andisol de la precordillera andina chilena. *Agrociencia* 42 (2): 139–149.
- Somaratne, S., Seneviratne, G. and Coomaraswamy, U. 2005. Prediction of soil organic carbon across different land-use patterns: a neural network approach. *Soil Science Society of America Journal* 69 (5): 1580–1589.
- Sparling, G., Vesely, D. and Schipper, E. 2006. What is soil organic matter worth? *Journal of Environmental Quality* 35: 548–557.
- Terra, J., Shaw, J., Reeves, D., Raper, R., Van Santen, E. and Mask, P. 2004. Soil carbon relationships with terrain attributes, electrical conductivity, and a soil survey in a Coastal Plain landscape. *Soil Science* 169 (12): 819–831.
- Thompson, J., Pena-Yewtukhiw, E. and Grove, J. 2006. Soil-landscape modeling across a physiographic region: Topographic patterns and model transportability. *Geoderma* 133 (1-2): 57–70.

- Tobler, W. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46: 234–240.
- Triantafilis, J., Huckel, A. and Odeh, I. 2001. Comparison of statistical prediction methods for estimating field-scale clay content using different combinations of ancillary variables. *Soil Science* 166 (6): 415–427.
- USGS. 2008. The National Map Seamless Server, Earth Resources Observation and Science (EROS). U.S. Department of the Interior & U.S. Geological Survey.
Disponible en: <http://www2.jp1.nasa.gov/srtm/index.html>. Leído el 20 de Noviembre de 2010.
- Van Es, H. 2002. pp. : 1–13. *In*: Dane, J. y Topp, G. (eds.). *Methods of soil analysis, part 4, Physical Methods*. Soil Science Society of America (SSSA). Book Series N°5. Madison , Wisconsin, USA. 1692 p.
- Walkley, A. and Black, I. 1934. Estimation of soil organic carbon by the chromic acid titration method. *Soil Science* 37: 29–38.
- Wanas, N., Auda, G., Kamel, M. and Karray, F. 1998. pp. : 918–921. *In*: IEEE Canadian Conference on Electrical and Computer Engineering. Vol. 2. Waterloo, Ontario, Canada. 944 p.
- Watson, G. 1984. Smoothing and interpolation by kriging and with splines. *Mathematical Geology* 16 (6): 601–615.
- Webster, R. and Oliver, M. 2007. *Geostatistics for environmental scientists*. 2th Edition. John Wiley & Sons Inc, Chichester, UK. 330 p.

APÉNDICES

Apéndice I. Índice Moran local de Anselin.



Apéndice II. Test de Wilcoxon para determinar diferencias entre las *líneas de modelación* utilizando subsectores como bloques.

