



UNIVERSIDAD DE MURCIA

FACULTAD DE BIOLOGÍA

Estudio de la Distribución Espacial y Cartografía Digital de Algunas Propiedades Físicas, Químicas e Hidrodinámicas de Suelos de la Cuenca del Segura

Dña. Arantzazu Blanco Bernardeau
2015

UNIVERSIDAD DE
MURCIA



ESTUDIO DE LA DISTRIBUCIÓN ESPACIAL Y
CARTOGRAFÍA DIGITAL DE ALGUNAS
PROPIEDADES FÍSICAS, QUÍMICAS E
HIDRODINÁMICAS DE SUELOS DE LA CUENCA
DEL SEGURA.

Tesis Doctoral

ARANTZAZU BLANCO BERNARDEAU

DIRECTORES

Francisco Alonso Sarría

Roque Ortiz Silla

Antonio Sánchez Navarro

Esta Tesis Doctoral ha sido realizada en el Departamento de Química Agrícola, Geología y Edafología de la Universidad de Murcia y el Instituto Universitario del Agua y el Medio Ambiente (INUAMA). El trabajo ha sido posible gracias a la financiación del Instituto Euromediterráneo del Agua a través de los Proyectos "Caracterización de las propiedades físicas e hidrodinámicas de los suelos de la cuenca del Segura", y "Elaboración de la Cartografía de propiedades físico-químicas e hidráulicas de los suelos de la cuenca del Segura", en el seno del cual se concedió a la doctoranda una beca de colaboración.

Hay mucha, mucha gente a la que me gustaría dar las gracias...y darían para un capítulo entero. Compañeros de departamento, con los que compartí también años de carrera y máster; además del apoyo y la compañía, sus ideas y sus proyectos también sirven para enriquecer el trabajo. Profesores e investigadores que aunque no formaban parte de este proyecto, siempre estaban dispuestos a dar un consejo y solucionar un problema. Técnicos de laboratorio de trabajo silencioso, pero que salvan experimentos entre el maremágnum de trabajo y llenos de palabras de ánimo...

Un agradecimiento especial merecen mis directores, en primer lugar, por haber confiado en mí. Sobre todo, en aquellos momentos en que yo no lo hacía.

Me gustaría también agradecer especialmente a mi familia, por el apoyo constante a lo largo de estos años que a todos se nos han hecho largos.

Y finalmente pero no en último lugar a mis amigos, por estar ahí incluso cuando yo no estaba, por su paciencia cuando la tesis estaba por delante de cualquier cosa.

"La vida es aquello que te sucede mientras escribes la tesis"

Dedicado:

A mis padres, los primeros, y a mi hermano. Por todo, absolutamente TODO.

A Fede. Sin su aliento constante, es seguro que habría abandonado hace tiempo.

Índice

Pág.

1	Introducción y objetivos	1
1.1	Modelos predictivos	3
1.1.1	Métodos locales	4
1.1.2	Métodos globales	5
1.1.3	Funciones de edafotransferencia	10
1.2	Propiedades edáficas	10
1.2.1	Carbono orgánico	10
1.2.2	Carbonato cálcico total o equivalente	12
1.2.3	Granulometría	13
1.2.4	Densidad aparente	14
1.2.5	Capacidad de retención de humedad	15
1.3	Objetivos	17
2	Material y métodos	19
2.1	Zonas de estudio	19
2.1.1	Región de Murcia	20
2.1.2	Sierra de Mojantes	25
2.1.3	Sierra de las Moreras	26
2.2	Fuentes de información	28
2.2.1	Datos de suelos	28
2.2.1.1	Base de datos LUCDEME	28
2.2.1.2	Datos de suelos de la sierra de Mojantes	31
2.2.1.3	Datos de suelos de la sierra de las Moreras	32
2.2.2	Variables ambientales	33
2.2.2.1	Variables geomorfométricas	33
2.2.2.2	Variables climáticas	38
2.2.3	Variables cualitativas	39
2.3	Software	46
2.3.1	Sistemas de información geográfica: GRASS GIS y SAGA GIS	46
2.3.2	Software estadístico: R	46
2.4	Técnicas estadísticas	46
2.4.1	Transformación de variables	46
2.4.2	Selección de variables	47

2.4.3	Evaluación y validación de los modelos	48
2.4.4	Interpolación de los residuos	51
2.4.5	Funciones de edafotransferencia	54
3	Resultados y discusión	57
3.1	Selección de variables	57
3.2	Modelos a escala regional.	59
3.2.1	Carbono orgánico.	59
3.2.1.1	Análisis descriptivo.	59
3.2.1.2	Modelos.	63
3.2.1.3	Residuos del modelo.	71
3.2.2	Carbonato cálcico total o equivalente.	73
3.2.2.1	Análisis descriptivo.	73
3.2.2.2	Modelos.	76
3.2.2.3	Residuos del modelo.	82
3.2.3	Arcilla.	84
3.2.3.1	Análisis descriptivo.	84
3.2.3.2	Modelos.	86
3.2.3.3	Residuos del modelo.	93
3.2.4	Limo.	96
3.2.4.1	Análisis descriptivo.	96
3.2.4.2	Modelos.	99
3.2.4.3	Residuos del modelo.	105
3.2.5	Arena.	107
3.2.5.1	Análisis descriptivo.	107
3.2.5.2	Modelos.	111
3.2.5.3	Residuos del modelo.	117
3.3	Modelos a escala local: sierra de Mojantes.	119
3.3.1	Carbono orgánico	119
3.3.1.1	Análisis descriptivo	119
3.3.1.2	Modelos	124
3.3.2	Carbonato cálcico total o equivalente	132
3.3.2.1	Análisis descriptivo	132
3.3.2.2	Modelos	134
3.3.3	Arcilla	140
3.3.3.1	Análisis descriptivo	140
3.3.3.2	Modelos	142
3.3.4	Limo	147
3.3.4.1	Análisis descriptivo	147
3.3.4.2	Modelos	150
3.3.5	Arena	155
3.3.5.1	Análisis descriptivo	155
3.3.5.2	Modelos	158

4	Funciones de edafotransferencia	163
4.1	Densidad aparente	163
4.2	Contenido de humedad	168
4.2.1	Capacidad de campo	169
4.2.2	Punto de marchitamiento	178
5	Conclusiones	187
	Lista de tablas	189
	Lista de figuras	193
	Referencias bibliográficas	199
	Cartografía	A.0

Acrónimos utilizados

AIC Criterio de Información de Akaike

ANN Redes neuronales artificiales

Da Densidad aparente

BT Boosted Trees

CART Árbol de clasificación y regresión

CE Conductividad eléctrica

CC Capacidad de campo

CO Carbono orgánico

ETP Evapotranspiración potencial

GAM Modelo aditivo generalizado

GAM-CV Modelo aditivo generalizado con validación cruzada en bloques

GLM Modelo lineal generalizado

GLM-CV Modelo lineal generalizado con validación cruzada en bloques

KO Krigado ordinario

KU Krigado universal

MARS Multivariable Adaptive Regression Splines

MARS-CV MARS con validación cruzada en bloques

MRVBF Multiresolution Valley Bottom Flatness

NT Nitrógeno total

PMP Punto de marchitamiento permanente

PTF Función de edafotransferencia

RECM Raíz del error cuadrático medio

RF Random Forest

RF-CV Random Forest con validación cruzada en bloques

RF-CVIMP Random Forest con validación cruzada en bloques por importancia de variables

RK Regresión-krigeado

RSS Suma de los cuadrados de los residuos (Residual Sum of Squares)

SVM Máquina de vector soporte (Support Vector Machine)

SVM-CV Máquina de vector soporte con validación cruzada en bloques

Capítulo 1

Introducción y objetivos

El suelo es un sistema abierto que representa la interfaz a través de la cual se relacionan otros sistemas como la atmósfera, la geosfera y la hidrosfera, denominados en conjunto como ambiente.

En el sistema suelo tienen lugar una gran diversidad de procesos físicos, químicos y biológicos que dependen de sus propiedades y atributos. Estos procesos dan como resultado que el suelo cumpla importantes funciones en los ecosistemas, definidas por [Karlen *et al.* \(1997\)](#). Entre estas funciones destacan la de almacenamiento de recursos hídricos, y la de regulación de los flujos de agua y solutos. Estas funciones son de una importancia fundamental, ya que el agua almacenada en el suelo influye en la tasa de evaporación, la recarga de los acuíferos, y la generación de escurrimiento.

La gestión de los recursos hídricos, tanto en lo concerniente a la cantidad como a la calidad, constituye uno de los problemas más importantes para el desarrollo sostenible, especialmente en aquellos territorios donde los recursos son escasos, como las regiones áridas y semiáridas entre las que se encuentra la Cuenca del Río Segura ([Maestre-Valero *et al.*, 2013](#)). Este problema se vuelve más acuciante dado el actual escenario de cambio climático, cuyos efectos en la humedad del suelo dependerán no solo de la intensidad de tal cambio, sino también de las características de cada suelo, que también pueden verse afectadas por el mismo ([IPCC, 2001](#)). El cuarto informe de evaluación del IPCC sitúa además las áreas mediterráneas como zonas especialmente afectadas por el cambio climático, donde se prevee una reducción en la frecuencia de las precipitaciones con condiciones de sequía extrema, reduciéndose los flujos de carbono y agua en los ecosistemas ([IPCC, 2007](#)). El conocimiento de las propiedades del suelo es por tanto necesario para una adecuada toma de decisiones respecto a la gestión de los recursos, la ordenación y planificación del territorio, así como la monitorización de los impactos que tienen lugar sobre los mismos.

El comportamiento hídrico de los suelos viene determinado por sus propiedades físicas y sus propiedades químicas, que influyen a su vez en las primeras, de modo que es posible establecer

relaciones entre ellas. Estas propiedades edáficas están a su vez relacionadas con propiedades morfométricas, climáticas, litológicas y relacionadas con los usos del suelo, a través de distintos procesos suelo-paisaje. Estas relaciones ya han sido establecidas por un gran número de científicos, incluyendo a [Jenny \(1941\)](#), quien propuso el modelo empírico-determinístico **CLORPT**, que describe la influencia de los diferentes factores (**CL**ima, **O**rganismos, **R**elieve, material **P**arental, **T**iempo) en el desarrollo del suelo.

Sin embargo, la mayoría de la información publicada sobre las relaciones suelo-ambiente son de tipo cualitativo, y a menudo no están claramente enunciadas ([Gray et al., 2009](#)). Por otro lado, las propiedades del suelo presentan escalas de variación diversas y complejas ([Burrough, 1993](#)), resultado del amplio rango espacio-temporal de los procesos edafogénicos ([Allen y Starr, 1982](#)), incluyendo la impronta de climas pasados, el solapamiento y las sinergias entre los mismos. Esta complejidad resulta en una gran diversidad de opiniones sobre la influencia de las condiciones de formación en la distribución espacial de las propiedades del suelo, sobre todo a escala regional y local ([Gray et al., 2009](#)). Como resultado, los estudios de suelos han cambiado su orientación subjetiva por una visión más cuantitativa, a través de modelos que incluyen medidas de precisión y evaluación de la incertidumbre.

Los estudios de suelos, incluyendo las descripciones de campo, los análisis de laboratorio, la clasificación y cartografía consiguientes, así como informes e interpretaciones de las unidades o series de suelos, pueden considerarse como inventarios y se han utilizado de forma tradicional para la ordenación territorial y la gestión del suelo. Debido al alto coste, tanto en términos económicos como de tiempo, que requiere la obtención de muestras en campo y su posterior análisis, existe una carencia generalizada de datos apropiados que reflejen estos patrones espaciales, especialmente de datos referentes a las propiedades del suelo que determinan su comportamiento hídrico ([Sumfleth y Duttmann, 2008](#)). El desarrollo de los GPS, técnicas de detección remota como los LIDAR o imágenes de satélite, así como de diversas técnicas geofísicas, como la reflectometría de dominio de tiempo (TDR) o los radares terrestres (GPR) han permitido obtener medidas indirectas. Sin embargo, incluso con todas estas fuentes de datos, el suelo solamente puede ser medido de forma puntual en el espacio y en el tiempo, y las propiedades del suelo en las zonas no muestreadas han de ser estimadas. Por este motivo, en las últimas décadas se vienen desarrollando diferentes técnicas de modelización de las propiedades de los suelos, que permiten predecir espacialmente las propiedades de los mismos de forma rápida y fiable ([Dobos et al., 2006](#); [Grimm y Behrens, 2010](#)).

Estos modelos estadísticos se desarrollan a partir de atributos del suelo, medidos en el campo, en localizaciones donde se dispone de otros atributos ambientales (variables explicativas), tales como variables topográficas, climáticas, material original, tipo de vegetación y uso del suelo ([Gessler et al., 1995](#)), minimizando el número de muestras necesarias para generar las predicciones espaciales. [McBratney et al., 2003](#) propusieron como base para la cartografía digital de suelos el modelo **SCORPAN**, basado en el anterior, en el cual el Suelo (S), o cualquiera de sus propiedades, es una función de factores ambientales tales como el clima (C), organismos (O), relieve(R), material parental (P), el tiempo expresado como la edad del suelo (A) y la posición en el espacio (N), determinada por las coordenadas, de modo que la variabilidad espacial del suelo se consideraría como el resultado de las complejas combinaciones de estos factores a través de los procesos de edafogénesis. Este último modelo forma la base de la cartografía digital de

suelos (Lagacherie y McBratney, 2006; IUSS, 2006).

1.1 Modelos predictivos

Los avances de las últimas décadas en cuanto a la toma de datos y su almacenamiento digital, el aumento en la capacidad de memoria y cálculo de las computadoras, y el desarrollo de los SIG han permitido utilizar diversas aproximaciones a la hora de predecir las propiedades de los suelos y modelizar los mismos en mapas puramente digitales. Estos se denominan así para diferenciarlos de los mapas en papel digitalizados para crear mapas vectoriales de polígonos que, sin embargo, poseen las limitaciones de los mapas de coropletas que los originaron. Dado que los muestreos de suelos proporcionan datos sobre observaciones puntuales (discontinuas), la realización de la cartografía digital precisa convertir esos datos en polígonos o áreas continuas, usando diferentes métodos de interpolación para estimar el valor de las variables en los puntos en los que no se tienen observaciones (IUSS, 2006; Lagacherie y McBratney, 2006; Roecker *et al.*, 2010).

La interpolación espacial consiste en estimar el valor de una variable o propiedad del suelo $f(x)$ en un punto x , del que no se tienen observaciones, a partir de los valores conocidos de f en un cierto número de puntos situados alrededor.

Los métodos utilizados pueden clasificarse en paramétricos y no paramétricos. Los métodos paramétricos asumen premisas sobre el tipo de función, o forma, de la variable, al cual posteriormente se ajustan los datos, utilizando una serie de parámetros, reduciéndose el problema a la obtención de tales parámetros, denominados *grados de libertad* del modelo. Por el contrario, los métodos no-paramétricos no asumen tales premisas, pero precisan de mayor cantidad de datos para obtener predicciones precisas; además, pueden producir problemas de sobreajuste, en los que el modelo no genere resultados precisos para datos que no estén en el conjunto que lo generó (Hastie *et al.*, 2009).

Por otro lado, se puede diferenciar también entre métodos locales y métodos globales. Los métodos locales utilizan solamente los datos situados en la vecindad del punto a estimar, mientras que los métodos globales utilizan todos los datos disponibles para dar lugar a una estimación en los puntos cuyo valor se desconoce. A menudo, los métodos globales son utilizados para eliminar los efectos de tendencia, de forma previa al uso de métodos locales para analizar los residuos.

La selección del método adecuado de interpolación influye en la precisión del mapa de forma significativa, así como su interpretabilidad (Kravchenko y Bullock, 1999; Robinson y Metternicht, 2006; Omran, 2012). Los modelos más restrictivos, tales como los modelos lineales, son sencillos de interpretar y son más adecuados cuando el objetivo es crear modelos explicativos; los modelos más flexibles, como los basados en aprendizaje automático, ofrecen predicciones más exactas, pero son más difíciles de interpretar. La diferencia entre modelos explicativos y predictivos, es que los primeros están diseñados para comprobar teorías y dar una explicación causal a los fenómenos que describen los modelos, mientras que los segundos están pensados para predecir observaciones nuevas o futuras (Geisser, 1993; Breiman, 2001b; Shmueli, 2010). Según Shmueli (2010), ambos objetivos pueden confluir, aunque la distinción entre predicción y explicación tiene un importante efecto en los diferentes pasos del proceso de modelización y en sus consecuencias.

1.1.1 Métodos locales

Los métodos más sencillos de interpolación son los que no tratan de obtener un modelo de correlación estadística entre las muestras, sino que asumen uno a priori.

La **interpolación por ponderación inverso de la distancia** (IDW) es un método no paramétrico que asume que cada punto tiene una influencia local que disminuye de forma no lineal con la distancia, de modo que los puntos utilizados como entrada que se encuentren más cerca de los puntos cuyo valor queremos predecir reciben un peso mayor en la ponderación que los puntos más lejanos. Los pesos son proporcionales al inverso de la distancia elevado a un valor de potencia p , de modo que al aumentar la distancia, los pesos disminuyen rápidamente, dependiendo el grado de disminución del valor de p ; si $p = 0$, no hay disminución con la distancia, con lo que los pesos se mantienen constantes y la predicción es la media de todos los valores medidos; si $p = 1$, el método será lineal; al aumentar p , los pesos de los puntos distantes disminuyen, de forma no lineal. Generalmente, se utiliza un valor de $p = 2$, denominándose el método *Interpolación con ponderación cuadrática inversa a la distancia* (Mitas y Mitasova, 1999; DeMers, 2005).

La interpolación usando **superficies de tendencia** (*splines*) es un método no paramétrico consistente en definir polinomios que generan superficies cuya curvatura se minimiza; la ventaja sobre la interpolación con IDW es que pueden predecir valores por debajo o por encima de los medidos. Los *splines* proporcionan resultados satisfactorios en variables con cambios suaves, pero generan errores considerables cuando existen grandes variaciones en distancias cortas (Webster y Oliver, 2001).

Las **técnicas geoestadísticas** son un conjunto de métodos paramétricos, basados en la teoría de las variables regionalizadas desarrollada por Matheron (1971) e introducida en ciencias del suelo por Burgess y Webster (1980a), Burgess y Webster (1980b), y Webster y Burgess (1980), más conocidas como *Krigeado* por el ingeniero de minas D.G. Krige. Estas técnicas se basan en la existencia de autocorrelación espacial entre las muestras, que establece que los factores que determinan la variación de una función o atributo son la distancia y la dirección entre los puntos de muestreo. La herramienta básica de las técnicas geoestadísticas es el variograma experimental, calculado a partir de los pares de datos con diferentes distancias de separación o lag (h), y el ajuste del mismo a un modelo de variograma teórico, donde los más frecuentes son el lineal, el esférico, y el gaussiano (Regalado y Ritter, 2005). Los parámetros del variograma, que determina la estructura de correlación espacial de las variables, proporcionan el peso de los puntos a utilizar en la interpolación. Estos parámetros son la pepita (*nugget*), que describe la microvarianza (variación a distancias más cortas que el espaciado de muestreo) y los errores de muestreo, y la zona espacialmente correlacionada o estructurada, modelada en base a dos parámetros: la meseta (*sill*), que es el valor de la varianza o autocorrelación donde esta se vuelve estacionaria; y el alcance (*range*), que es la distancia o lag a la que se alcanza la meseta (Matheron, 1971).

Dentro de las técnicas de krigeado existen diferentes métodos: El **krigeado ordinario (KO)**, se basa en asumir que la variación se produce de forma homogénea entre los puntos vecinos y depende solamente de la distancia entre ellos, no de su posición; esta técnica ha sido ampliamente utilizada en la predicción de propiedades del suelo tales como el carbono orgánico (McGrath y

Zhang, 2003; Sigua y Hundall, 2008). El **krigeado universal (KU)**, tal y como lo formuló Matheron (1969) es un caso generalizado del krigeado ordinario donde además de la distancia, influye la posición de los puntos, de modo que la tendencia se modeliza como una función de las coordenadas. En el caso del **cokrigeado**, se utiliza otra variable cuya determinación es menos costosa que la variable objetivo de estudio, estando ambas variables fuertemente correlacionadas entre sí; en este caso, es preciso calcular además los variogramas cruzados. El **krigeado con deriva externa** se utiliza cuando las variables auxiliares utilizadas para mejorar la predicción están muestreadas en la misma red de puntos que la variable objetivo de estudio.

Uno de los requisitos para el uso de la geoestadística es la normalidad de los datos, ya que los sesgos en la distribución de la población pueden enmascarar las posibles estructuras espaciales (Webster y Oliver, 1990). Por ello, se precisa una transformación de los datos, previa al krigeado; en este caso, la dificultad radica en seleccionar la transformación adecuada y en la generación de errores al revertir dicha transformación, aunque se han obtenido resultados satisfactorios con transformaciones logarítmicas (Kravchenko y Bullock, 1999; Regalado y Ritter, 2005; Robinson y Metternicht, 2006).

El método de **regresión-krigeado** es una variante del krigeado con deriva externa, en el cual la tendencia o deriva se ajusta mediante regresión y los residuos a través de técnicas de krigeado (Goovaerts, 1997); la ventaja de esta técnica de regresión-krigeado es que puede combinar diferentes técnicas de regresión, tanto paramétricas como no paramétricas.

1.1.2 Métodos globales

La **Regresión Lineal Múltiple (MLR, *Multiple Linear Regression*)** es un método paramétrico ampliamente utilizado en la modelización de propiedades del suelo (Moore *et al.*, 1993; Gessler *et al.*, 1995; Hengl *et al.*, 2004; Thompson y Kolka, 2005; Sumfleth y Duttmann, 2008) y la elaboración de funciones de edafotransferencia debido a su simplicidad y sencillez de uso, y su interpretación directa. La **regresión lineal por pasos** es una variante que consiste en añadir o eliminar variables predictoras, utilizando criterios que indiquen el impacto de cada adición o sustracción y por tanto de cada variable sobre la precisión del modelo. También pueden introducirse variables ordinales o cualitativas, que deben ser previamente transformadas en variables *dummy* para un análisis adecuado. Para ello, si tenemos una variable cualitativa con n niveles o categorías, se crean n nuevas variables, que tendrán nivel 0 si la muestra no tiene esa categoría, o nivel 1 si la muestra pertenece a esa categoría.

No obstante, a pesar de la estrecha relación entre las propiedades de los suelos y las variables ambientales, existen diversos problemas, tales como que las variables no tengan una distribución normal, que su varianza no sea homogénea, o que la relación entre las variables no sea lineal. Por ello, se hace necesario utilizar modelos más robustos que permitan trabajar con modelos no lineales.

Los **Modelos Lineales Generalizados (GLM)** son métodos paramétricos que funcionan como una extensión de los modelos lineales. Los GLM no fuerzan a los datos a adoptar escalas no naturales, y por tanto permiten las estructuras no lineales y cuya varianza no es constante (Hastie y Tibshirani, 1990). Se basan en asumir una relación, denominada "función de enlace",

entre la media de la variable respuesta y una función lineal de las variables explicativas, que pueden adoptar diversas distribuciones entre las que se encuentran la normal, pero también la de Poisson, binomial, binomial relativa, o gamma, las cuales suelen ajustarse mejor a la estructura de las variables ecológicas que la distribución normal. El uso de un GLM asumiendo una relación o función de enlace normal, sería equivalente a utilizar un modelo lineal (MLR). Además, los GLM pueden adaptarse para la modelización de variables cualitativas o semicuantitativas, a través de modelos logísticos (Guisan y Harrell, 2000), asegurando la linealidad, y restringen la variable respuesta a un rango de valores posibles, por medio de la función de enlace. Sin embargo, la desventaja de los GLM es la dificultad para encontrar las funciones polinomiales y las transformaciones adecuadas para mejorar el ajuste del modelo.

Los **Modelos Aditivos Generalizados (GAMs)** son métodos semi-paramétricos derivados de los GLM que incluyen además estimaciones de incertidumbre (Hastie y Tibshirani, 1990; Poggio *et al.*, 2010). Los GAM se introdujeron como una mejora de los GLM, ya que son capaces de identificar de forma automática la mejor transformación o función de enlace, de forma independiente, para cada variable predictora que las precise, de forma que estas pueden modelizarse de forma no paramétrica junto con otras variables lineales o polinomiales, agrupándose todas ellas para calcular la variable dependiente, de lo cual resulta el calificativo de "aditivo". Esto permite explorar distintas curvas de respuesta a gradientes ambientales y ajustar modelos que responden mejor a la teoría edafológica. La dificultad de los GAM radica en la selección de los grados de libertad efectivos, en forma de grados de suavizado, para garantizar la máxima flexibilidad evitando al mismo tiempo el sobreajuste.

Los **árboles de clasificación y regresión (CART)**, engloban a un conjunto de métodos no paramétricos con los que, a partir de un conjunto de datos, se generan divisiones binarias de forma recursiva dando lugar a grupos de homogeneidad creciente (Breiman *et al.*, 1984). Al acabar, los subconjuntos de datos obtenidos, denominados "*nodos terminales*", o bien son homogéneos o bien contienen una cierta heterogeneidad definida por un umbral anteriormente establecido. En el caso de los árboles de regresión, a cada elemento de los nodos terminales se le asigna la media de la variable respuesta en ese nodo. Las reglas para realizar las divisiones se obtienen utilizando funciones estadísticas consiguen disminuir la heterogeneidad en los subconjuntos resultantes. La posibilidad de sobreajuste, común en los métodos no paramétricos, se resuelve mediante la "poda" de los árboles de regresión, eliminando el exceso de "ramas", que son enviadas a nodos de orden inferior, más cercanos al "tronco" (Torgo, 1999). Este método ha sido utilizado como herramienta de clasificación de cubiertas de vegetación (Friedl y Brodley, 1997), clases de suelo (Scull *et al.*, 2005), o en generar aplicaciones para predecir procesos relacionados con el mismo tales como erosión (Geissen *et al.*, 2007).

Cuando la variable respuesta es continua, se habla de árboles de regresión y los nodos terminales están ocupados por un modelo de regresión o por la media de los casos incluidos, dependiendo del tipo de árbol, de modo que si las variables predictoras asociadas a la observación satisfacen las reglas establecidas, se aplica un modelo de regresión para predecir la respuesta (Pachepsky y Rawls, 2004; Henderson *et al.*, 2005; Prasad *et al.*, 2006; Wiesmeier *et al.*, 2011). Las ventajas de los árboles de regresión es que no son sensibles a la existencia de huecos en los datos y valores extremos en los mismos; otra ventaja de este método es la sencillez de interpretación y la posibilidad de incorporar variables continuas y categóricas; además de todo esto, en caso

de bases de datos de gran tamaño, los árboles de regresión y clasificación permiten descubrir estructuras complejas que quedarían ocultas con otros métodos. Sin embargo, los árboles de regresión también tienen desventajas: entre ellas destacan el uso ineficiente de los datos, y que no tienen opción de marcha atrás, por lo que una vez realizada una división no se consideran alternativas; esto significa que se produce una optimización del error en cada una de las divisiones pero no considerando la totalidad del árbol. Además, el comportamiento predictivo de los árboles de regresión es bastante pobre.

Los **árboles aumentados** (*Boosted Trees*, **BT**) son una variante de los árboles de regresión que permiten mejorar el comportamiento predictivo de los mismos. Se trata de una combinación de métodos no paramétricos de regresión y aprendizaje automático, consistente en formar gran cantidad de árboles relativamente simples, que se ajustan de forma iterativa a los datos de entrenamiento ([Freund y Schapire, 1996](#); [Elith et al., 2008](#); [Lemercier et al., 2012](#)). El método se basa en la premisa de que es más sencillo encontrar muchas reglas aproximadas que una sola regla precisa ([Schapire, 2003](#)).

La técnica no paramétrica denominada **Random Forest** (**RF**), es una ampliación de los **CART** para mejorar la capacidad de predicción del modelo ([Breiman, 2001a](#)); la diferencia entre ambos radica en que en RF se construyen gran cantidad de árboles relativamente simples (del orden de cientos o miles) resultando en un "bosque de modelos". Para cada árbol se utiliza solamente un subconjunto de las variables predictoras, dependiendo el número de estas variables y la cantidad de árboles que se construyen del conjunto total de datos. Cada árbol se construye con un remuestreo mediante *bootstrapping* (con reemplazamiento) del conjunto de datos original, excluyéndose muestras que se utilizarán para la estimación de errores ([Stum, 2010](#)). El resultado de **RF** es una sola predicción, que es el promedio de todas las predicciones agregadas ([Moran y Bui, 2002](#)).

Para obtener una estimación de la importancia de las variables en los árboles de regresión, se procede del mismo modo que en los árboles aumentados. En cada división, la mejora del criterio de separación (expresada en forma de error cuadrático medio) es la medida de la importancia atribuida a la variable que se divide. Sin embargo, a diferencia de los árboles aumentados, en que algunas variables son ignoradas, en random Forest esto no es así. Random Forest también usa las muestras que no han sido usadas en la construcción de cada árbol (llamadas "out-of-bag", *oob*) para elaborar una medida de la importancia de las variables diferente a la anterior, cuya utilidad aparente es medir la fortaleza de la predicción de cada variable. Esta fortaleza se mide a través de dos medidas. La primera medida es la disminución media de la precisión (denominada %IncMSE), que registra el error de predicción (ECM en el caso de regresión) de la parte "out-of-bag"; esta operación se repite permutando cada variable predictora j . De esa forma, si la variable predictora está fuertemente relacionada con la variable respuesta, el MSE obtenido al eliminarla de la regresión aumentará.

La segunda es la disminución en la impureza del nodo (IncNodePurity), expresada como suma de los cuadrados de los residuos (en inglés, **RSS**, "Residual Sum of Squares") antes y después de realizar cada una de las divisiones ([Hastie et al., 2009](#)), promediada entre todos los árboles. Para árboles de regresión, la impureza del nodo se mide con la suma de los cuadrados de los residuos antes y después de realizar la división ([Hastie et al., 2009](#)). La suma de los cuadrados

de los residuos se determina como indica la ecuación 1.1, donde y_i es el valor n -ésimo observado de la variable a predecir, y \hat{y}_i es el valor predicho por el modelo.

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1.1)$$

Este procedimiento tiene varias ventajas, además de conseguir una mejora de la capacidad predictiva de una forma más rápida que mediante el aumento de número de árboles de los árboles descrita anteriormente; se elimina el exceso de ajuste a los datos; la correlación entre los árboles individuales es baja dado que la diversidad de árboles se consigue usando un número limitado de variables predictoras; disminuyen los niveles de sesgo y varianza; y las estimaciones de error son robustas debido a que estas se calculan con los datos no usados en el remuestreo. Por el contrario, el gran inconveniente es que es un método de "caja negra" que impide una interpretación sencilla de las relaciones entre las variables predictoras y las variables respuesta, pues es imposible investigar la estructura de cada uno de los árboles que conforman el bosque (Prasad *et al.*, 2006). Random Forest es un procedimiento ampliamente utilizado en ecología, a pesar de lo cual existen relativamente pocos ejemplos en el campo de la cartografía digital de suelos (Grimm *et al.*, 2008; Wiesmeier *et al.*, 2011; Ließ *et al.*, 2012).

Las **Redes Neuronales Artificiales (ANN)** son métodos no paramétricos de aprendizaje automático, que permiten detectar relaciones no lineales, y que han sido utilizados tanto en modelización de propiedades del suelo, cuando se dispone de bases de datos de gran extensión y con gran cantidad de variables (Elshorbagy y Parasuraman, 2008), como para la elaboración de funciones de edafotransferencia. Estas redes neuronales pretenden imitar el funcionamiento del cerebro humano; poseen un sistema de muchos elementos, comparables a neuronas, interconectados por conectores codificados de diversos modos y organizados en capas: una capa de entrada, una capa de salida, y una o más capas ocultas que extraen información útil de la capa de entrada y la utilizan para predecir los resultados en la capa de salida. Se trataría de un método de aprendizaje automático, ya que no hay una estructura asumida del modelo, sino que las redes son ajustadas o "entrenadas" para que unos datos de entrada determinados den un resultado específico (McBratney *et al.*, 2003); este modelo matemático consiste en una serie de funciones simples unidas a través de pesos. Existen estudios en los que se describen mejores resultados en la predicción que en otros métodos convencionales de regresión (Schaap y Leij, 1998b); por ejemplo, no requieren simplificar el sistema o asumir premisas respecto a los datos. Sin embargo, dado que se trata de un modelo de caja negra no se puede conocer la relación entre las capas de entrada y salida.

Las redes neuronales han sido aplicadas por lo general a la predicción de funciones de edafotransferencia, pero también a la determinación de otras propiedades del suelo (Behrens *et al.*, 2005; Anagu *et al.*, 2009; Hattab *et al.*, 2013).

Una técnica más reciente de modelización se denomina **Multivariate Adaptive Regression Splines (MARS; Friedman, 1991)**; se trata de un método no paramétrico que combina la regresión lineal, la construcción de splines, la partición binaria recursiva del conjunto de datos y algoritmos inteligentes para generar un modelo capaz de predecir el valor de una variable objetivo, que puede ser continua o categórica, a partir de variables independientes. El funcionamiento

de MARS consiste en obtener un modelo de regresión general, sobre el cual se buscan "nudos". Los nudos son puntos donde el comportamiento de la función cambia, ajustándose a las denominadas "funciones básicas". Estos nudos se establecen por pasos, tanto hacia delante como hacia atrás. En primer lugar, se genera un modelo sobreajustado, y posteriormente se eliminan "hacia atrás" los nudos que menos contribuyen a la eficiencia del modelo. El mejor modelo se selecciona mediante una validación cruzada generalizada, que penaliza los términos o nudos añadidos, favoreciendo los valores que indiquen menor complejidad.

Existen estudios que utilizan MARS para modelizar propiedades del suelo utilizando variables relacionadas con la reflectancia (Shepherd y Walsh, 2002; Bilgili *et al.*, 2010; Viscarra Rossel y Behrens, 2010), pero también fenómenos tales como el transporte de solutos (Yang *et al.*, 2003) o la ocurrencia de erosión en surcos (Gómez *et al.*, 2011).

La técnica de **Support Vector Machine** (SVM; Vapnik, 1995) es un método inicialmente diseñado para clasificación, basado en la formulación original del concepto de hiperplano óptimo que separa los datos en dos categorías. Entre los hiperplanos posibles con una tasa de error determinada, el hiperplano óptimo es el que se encuentra en medio del espacio más grande entre los casos de ambas clases, proporcionando el mayor margen de separación entre las muestras con el mínimo error de generalización. Las muestras más cercanas al hiperplano de separación $f(x)$ son conocidas como vectores soporte y sobre ellas pasan hiperplanos paralelos a $f(x)$, de modo que el margen de separación (m) geoméricamente es la distancia perpendicular entre los hiperplanos de los vectores soporte.

Este método puede extenderse a problemas de regresión (Smola y Schölkopf, 2004), en cuyo caso el objetivo es encontrar una función $f(x)$ que tiene como máximo una desviación ϵ de los valores objetivo y_i para todos los datos de entrenamiento, siendo a la misma vez lo más plana posible. El resultado es una función similar a una función lineal, con una importante diferencia: el uso de una función de pérdida con una zona de insensibilidad, donde ϵ define el radio de un tubo en el cual se encuentra la función de regresión (figura 1.1). La pérdida es igual a 0 si la diferencia entre el valor predicho ($f(x_i)$) y el valor medido (y_i) es menor a ϵ .

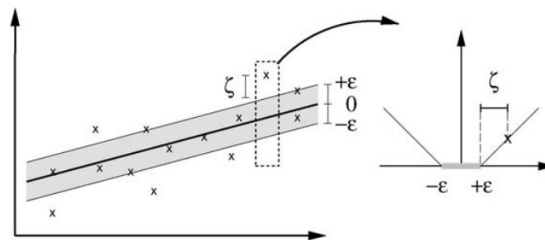


FIGURA 1.1: Concepto del margen suavizado para una SVM lineal. Fuente: Smola y Schölkopf (2004).

Como no siempre se puede encontrar esa función f que aproxima todos los valores de muestras (x_i, y_i) con precisión ϵ , es necesario permitir otros errores, por lo que se introducen las funciones de ajuste (ζ_i, ζ_i^*) , que miden la distancia entre la observación y la superficie del tubo ϵ . También se introduce una constante $C > 0$ que determina el equilibrio entre la planicidad de la función,

y la cantidad de pares de muestras que se permiten fuera del tubo determinado por ϵ .

Esta formulación puede extenderse para llevar a cabo regresión no-lineal, proyectando los datos originales a un espacio de variables de mayor dimensionalidad, donde pueda trabajarse con una función lineal; el problema se traslada a buscar la función más plana posible en este espacio de mayor dimensionalidad, en vez de en el espacio de muestras. Esto se lleva a cabo mediante determinadas transformaciones, que en realidad también se utilizan en otras técnicas, tales como las ANN, aunque en el caso de SVM con la ventaja de ser un procedimiento con menor coste de computación, las funciones *kernel* ("núcleo"). Una función *kernel* $K(x_i, x_j)$ es una función en el espacio de datos de entrada \mathcal{X} que permite evitar una proyección completa, calculando en su lugar los productos escalares en el espacio hiperdimensional \mathcal{F} de forma directa, evitando los problemas ligados a una dimensionalidad extrema. Esta función puede ser de varios tipos: lineal, polinomial, o una función de base radial Gaussiana.

Las Máquinas de Vectores Soporte han sido utilizadas en un amplio rango de problemas, tanto para clasificación como para regresión (Meyer *et al.*, 2003). En el caso de las ciencias ambientales han resultado ser un gran apoyo para tecnologías como la teledetección, caracterizadas por su gran número de observaciones. También han sido utilizadas en el campo de la edafología, tanto para la predicción de propiedades como para la clasificación de tipos de suelo o la detección de casos de salinización o contaminación del mismo (Bhattacharya y Solomatine, 2006; Ballabio, 2009; Kovacevic *et al.*, 2010; Cai *et al.*, 2010; Brungard *et al.*, 2015).

1.1.3 Funciones de edafotransferencia

Las funciones de edafotransferencia (PTFs; Bouma, 1989), que inicialmente se denominaron "funciones de transferencia" (Bouma y van Lanen, 1987) también se basan en el uso de métodos de regresión para predecir ciertas propiedades del suelo cuya determinación es costosa, a partir de otras propiedades del suelo más fáciles de medir, con las cuales están relacionadas, generalmente CO, granulometría y densidad aparente (Da).

Las primeras PTFs tenían como objetivo estimar el contenido de agua del suelo en determinados puntos de la curva de humedad considerados de interés agronómico, que solían ser la capacidad de campo (-33 KPa) y el punto de marchitamiento (-1500 KPa)(Salter y Williams, 1965); sin embargo, el avance en la capacidad de computación permitió posteriormente el desarrollo de las PTF paramétricas, basadas en modelos más complejos (Gupta y Larson, 1979; Rawls, 1983). Las funciones de edafotransferencia también pueden ser "de clase", cuando se utiliza un factor de agrupamiento previo a la aplicación de la PTF, como puede ser la clase textural o el tipo de suelo (Pachepsky y Rawls, 2004) Los métodos más comunes en los que se basan las funciones de edafotransferencia son los regresiones lineales (Wösten *et al.*, 1995; Batjes, 1996; Tomasella *et al.*, 2000) y las redes neuronales(Schaap y Boultens, 1996; Minasny y McBratney, 2002). Existen asimismo trabajos que comparan la precisión de distintas funciones de edafotransferencia, obtenidas con procedimientos diversos o en unas condiciones edafoclimáticas determinadas (Merdun *et al.*, 2006; Abbasi *et al.*, 2011; Botula *et al.*, 2012; Brahim *et al.*, 2012).

A pesar de los avances, la aplicación de las PTFs publicadas está limitada por diversas causas: estar basadas en conjuntos de datos pequeños; estar orientadas a una determinada cobertura

o tipo de suelo (Huntington *et al.*, 1989; de Vos *et al.*, 2005); incluir variables de entrada no disponibles en las bases de datos existentes, lo cual suele ocurrir con datos de estructura y la densidad aparente (Heuscher *et al.*, 2005); y no haber sido validadas con conjuntos de datos independientes a los utilizados para su elaboración.

1.2 Propiedades edáficas

1.2.1 Carbono orgánico

La importancia del CO está ampliamente reconocida en la ciencia del suelo debido a su papel en la fertilidad y los ciclos de nutrientes al constituir un almacén para el resto de organismos y en la formación de la estructura del suelo al permitir la formación de agregados, que a su vez influye en la porosidad, la aireación del suelo, en las propiedades hídricas y en la resistencia ante la erosión. Por otro lado, el CO del suelo tiene también una gran importancia debido a su papel en los distintos procesos que se dan en el ecosistema y su influencia en la composición de la atmósfera y la tasa de cambio climático (Raich y Potter, 1995; Trumbore *et al.*, 1996; Jobbagy y Jackson, 2000); esta importancia se acentúa al encontrarse la zona de estudio dentro del área mediterránea, que según el informe del IPCC (2007) será una de las zonas que sufrirán mayores aumentos en las temperaturas, los periodos de sequía y los episodios de precipitación intensas o torrenciales. Por tanto, el conocimiento y desarrollo de modelos de distribución del CO en el suelo es de gran interés para la comprensión y predicción de las propiedades hídricas del suelo tales como infiltración y capacidad de retención de humedad, relacionados a su vez con procesos de erosión y desertificación.

El contenido y tipo de CO que existe en el suelo depende de los aportes de biomasa producidos por la vegetación, y la tasa de descomposición de la misma y, por lo tanto, de la actividad microbiana existente en el suelo, que a su vez responde a los cambios de humedad, temperatura y acidez del suelo (Alexander, 1977). En relación al pH, existen estudios que relacionan pequeños cambios en la acidez del suelo con grandes variaciones en los contenidos de CO en estudios a gran escala (Brady y Weil, 2002); el pH puede afectar la disponibilidad de nutrientes y con ello la producción de biomasa, además de influir en las poblaciones microbianas que intervienen en la descomposición de materia orgánica, que a su vez puede liberar ácidos orgánicos que afecten al pH.

Las variables climáticas tienen importancia a escalas grandes, en función de la latitud (Liu *et al.*, 2012b), condicionando el régimen de precipitaciones, la evapotranspiración potencial (ETP) o la amplitud térmica; estas variables climáticas condicionan a su vez la existencia de diferentes tipos de vegetación, de la cual depende la generación de mayor o menor cantidad de biomasa y su naturaleza, pero también con los procesos de humificación que condicionan su adición al suelo, y de mineralización, que provocan su pérdida. El cambio climático que se está experimentando actualmente es una amenaza potencial para el CO en ambientes semiáridos, que tendría como consecuencia una considerable disminución de este constituyente en el suelo y el aumento de la concentración de los gases de efecto invernadero en la atmósfera. En un estudio sobre el impacto del cambio climático en el sureste ibérico Albaladejo *et al.* (2013) predijeron un descenso en el

contenido de CO con el aumento de las temperaturas y la disminución de las precipitaciones que se esperan en las zonas semiáridas.

Sin embargo, a escala regional o local, las variables morfométricas, incluyendo la altitud, pendiente, orientación, así como el uso del suelo o tipo de vegetación podrían ser los factores dominantes (Rezaei y Gilkes, 2005). Estos atributos del paisaje afectarían a la actividad de los organismos, los procesos de erosión y sedimentación, las condiciones de drenaje y la exposición del suelo a la radiación solar, el viento y las precipitaciones (Buol *et al.*, 1989). De ellos, la variable que mayor correlación muestra con el contenido en CO es la altitud, excepto en las laderas orientadas al sureste (Thompson y Kolka, 2005); los mismos autores encontraron diferencias en el contenido de CO atendiendo a la forma de la pendiente, indicando que las zonas convexas tenían mayores contenidos en este componente que las posiciones cóncavas. En paisajes con relieve poco acusado las laderas cóncavas presentaban mayores contenidos de CO (Gessler *et al.*, 1995); además, en relieves más accidentados o pendientes mayores, las pendientes cóncavas parecen presentar mayor convergencia del flujo y mayores tasas de erosión, lo que produciría suelos poco espesos y bastante rocosos con escasez de CO. Nadeu Puig-Pey (2013) encontró relaciones entre los contenidos de CO y los factores que influyen en la movilización de sedimentos a escala de cuenca, entre los que se citan factores topográficos como la pendiente, o el área de la cuenca.

El uso y tipo de cobertura vegetal, y los cambios que se producen en las mismas, influyen asimismo en el contenido de CO de los suelos, además de en la erodibilidad y por tanto la movilización y exportación de sedimentos que lo contienen. En las zonas cultivadas, el contenido en CO muestra una alta dependencia de el tipo de cultivos y las prácticas de manejo del suelo, tales como el riego, la aplicación de enmiendas, o las técnicas de laboreo (Bergstrom *et al.*, 2001; Lal, 2002), que controlan la entrada de materia orgánica al suelo; en las zonas con vegetación natural, la cantidad y tipo de materia orgánica depende de estructura y distribución de la vegetación y las especies existentes en cada comunidad, que a su vez están relacionadas con factores topográficos, como la altitud, la posición en la ladera y la orientación (Rodríguez-Murillo, 2001; Yimer *et al.*, 2007). La conversión de zonas agrícolas de secano en bosques se refleja en un aumento del CO del suelo, cuya cantidad varía dependiendo de cada región. Boix-Fayos *et al.* (2009) encontraron una ganancia de 71% en la cuenca de la Rogativa, situada dentro de la cuenca del Segura. En sentido contrario, según los mismos autores, el cambio de una vegetación de bosque a matorral, de matorral a agricultura de secano, y de pastos a agricultura de secano representaron una disminución del 46% de CO en el suelo. En zonas semiáridas, se ha observado que el uso y la cobertura vegetal modulan el efecto de la precipitación en la redistribución del CO, al determinar el estado de agregación de las partículas del suelo (Martínez Mena *et al.* (2011). Sin embargo, a pesar de las medidas para secuestrar CO en los suelos a través de la reforestación, existen estudios que afirman que éstas no son lo suficientemente eficientes (Laganière *et al.*, 2010). Además, existen escasos datos sobre los mecanismos de estabilización del CO en el suelo (Six *et al.*, 2002; De Gryze *et al.*, 2004), sobre todo en zonas semiáridas.

Otros estudios han encontrado relaciones entre la distribución del tamaño de partículas y el contenido de CO independientemente del clima. Burke *et al.* (1995) describió un aumento en el contenido de CO al aumentar la proporción de arcillas, debido al efecto protector de las partículas de arcilla en los compuestos orgánicos; la correlación entre la arcilla y el CO ha resultado ser en otros estudios más alta que entre la precipitación y el CO (Nichols, 1984). En otros lugares, la

localización, la precipitación o la temperatura (Arrouays *et al.*, 2001) estaban más relacionados con el CO del suelo que la textura.

1.2.2 Carbonato cálcico total o equivalente

Los carbonatos pueden tener un origen tanto primario o litogénico como secundario o edafogénico. Los primarios proceden de calizas, dolomías, margas u otros sedimentos detríticos que son materiales parentales frecuentes en la zona de estudio. Los carbonatos secundarios proceden de la disolución de los carbonatos existentes en el sedimento, que son lavados y translocados vertical y horizontalmente en mayor o menor medida, y finalmente vuelven a precipitar, formando nódulos, concreciones pulverulentas o capas endurecidas continuas o discontinuas generalmente paralelas al suelo (horizontes cálcicos y petrocálcicos, respectivamente). Todos estos procesos reflejan los movimientos del agua dentro de la zona vadosa del suelo, y su ocurrencia depende de diversos factores: la disponibilidad de carbonatos en el medio; la presencia de vegetación y organismos, que liberan ácidos capaces de disolver estos carbonatos; la presencia de precipitaciones en cantidad suficiente para producir lavado lateral y vertical; y la estabilidad temporal suficiente para que se produzcan estos procesos (Yaalon, 1997; Candy y Black, 2009; Wilding *et al.*, 2012).

La importancia de la presencia y cantidad de los carbonatos en el suelo radica en las múltiples relaciones que muestran con otras propiedades físico-químicas del suelo.

En relación con la estructura del suelo, los carbonatos están correlacionados con la cantidad de agregados estables, restando importancia a la materia orgánica, con la cual interaccionan, en la formación de los mismos en los suelos de zonas semiárida (Boix-Fayos *et al.*, 2001; Bouajila y Gallali, 2008; Fernández-Ugalde *et al.*, 2011). Esta interacción se produce a diferentes niveles; en primer lugar, los iones de calcio favorecen las interacciones intermoleculares (Baldock y Skjempstad, 2000), siendo más efectivo que el magnesio en este aspecto; además, el Ca^{2+} puede inhibir la dispersión de las arcillas, mientras que el Mg^{2+} tendría el efecto contrario (Zhang y Norton, 2002). Por otro lado, la precipitación de los carbonatos secundarios sobre los residuos orgánicos previene la degradación microbiana (Clough y Skjemstad, 2000). A su vez, una adecuada estructura del suelo influye favorablemente al reducir el riesgo de erosión, compactación, retención y circulación del agua en el suelo, y penetración por las raíces (Bronick y Lal, 2005).

En sentido contrario, la acumulación y precipitación de carbonatos en forma de horizontes cálcicos o petrocálcicos puede conducir a la disminución de la porosidad, a una mayor resistencia a la penetración de las raíces, la disminución de la infiltración y de los movimientos del agua dentro de la matriz del suelo (conductividad hidráulica). Los suelos calcáreos también han sido descritos como más susceptibles a la compactación y la obstrucción de los poros, a pesar de que las capas endurecidas de carbonato no son completamente impermeables, pues permiten un cierto nivel de humectación y, aunque lentamente, el agua puede circular hacia capas más profundas, donde sí queda retenida (Hennessy *et al.*, 1983). Sin embargo, otros estudios consideran que la precipitación del carbonato cálcico en el espacio poroso no afecta ni a la retención de agua ni la conductividad de la misma (Frenkel *et al.*, 1978), aunque sí se considera que los suelos carbonatados tienen un comportamiento hidráulico diferente, desde el punto de vista de la energía de adsorción y retención de la humedad (Soil Survey Staff, 2010).

Debido a estas relaciones, la variabilidad espacial de los carbonatos es de particular interés, tanto para las ciencias agronómicas como las ambientales, a la hora de conocer las propiedades hídricas de los suelos; existen estudios que consideran el contenido en carbonatos como criterio al agrupar los suelos en función de su comportamiento hídrico (Tóth *et al.*, 2008). El contenido en carbonatos ha sido también incluido en funciones de edafotransferencia como variable de entrada, tal como se describe en la revisión realizada por Wösten *et al.* (2001), y más recientemente para el caso concreto de suelos en zonas áridas (Khodaverdiloo *et al.*, 2011; Parchami-Araghi *et al.*, 2013).

Por otro lado, los carbonatos representan una vía para el secuestro de carbono en las zonas áridas y semiáridas (Lal, 2004), de forma adicional a las técnicas de manejo de suelos y la vegetación. Este secuestro se produce al suministrar cationes (calcio y magnesio) desde fuentes internas, incluyendo entre ellas las sales marinas, fertilizantes y enmiendas; la dinámica de la formación de carbonatos secundarios depende de diversos factores, existiendo una fuerte sinergia entre el CO y la formación de carbonatos secundarios.

La calidad del agua de riego también es importante, tanto como fuente de sales como agente de lavado (Nieder y Benbi, 2008). Aunque la tasa de secuestro de C atmosférico a través de la formación de carbonatos y, por tanto, su significación, varía entre los diversos estudios realizados (Schlesinger, 1997; Monger y Gallegos, 2000), en función de las características del suelo, los regímenes de humedad y temperatura y los usos del suelo, ésta presenta un potencial medio estimado de 1 Pg de C al año, calculado para 50 años, con una tasa de devolución a la atmósfera de 30000 a 90000 años; es, por tanto, una posibilidad a considerar, en combinación con otras técnicas de control de la desertificación tales como la reforestación, y la reducción de emisiones.

1.2.3 Granulometría

La distribución en los tamaños de las partículas constituye la textura del suelo. Ésta tiene una influencia significativa en otras propiedades tales como la temperatura del suelo, la estructura y la formación de agregados, la cohesión del suelo, la capacidad de retención de humedad y la conductividad hidráulica. Por todo ello, un conocimiento adecuado de la distribución de partículas es importante para la modelización hidrológica (Arya y Paris, 1981), la gestión de riesgos geológicos tales como los corrimientos de laderas (Ließ *et al.*, 2012) o la vulnerabilidad a la erosión (Le Bissonnais, 1996). Las fracciones granulométricas son también utilizadas como variables de entrada en las funciones de edafotransferencia, o incluso como factores de agrupación en las denominadas "PTF de clase". Particularmente la arcilla es un componente muy activo, que influye en la retención y los ciclos de nutrientes al determinar la capacidad de cambio catiónico, además de proteger a la materia orgánica de la degradación y determinar la formación de agregados, aunque en este caso es tan importante su proporción como su tipología.

La proporción de las distintas fracciones granulométricas en la superficie del suelo depende de diversos factores (Jenny, 1941). En primer lugar, el material original influye en la distribución de las partículas, debido a su composición mineralógica, que determina a su vez su mayor o menor resistencia a los agentes de meteorización que puedan fragmentarla, diferenciándose también los materiales en función de su estado de consolidación. En suelos profundos la relación entre el

material original y la distribución de las partículas en la superficie puede ser menos directa, por la coexistencia de otros procesos pedogenéticos superpuestos.

Las variables climáticas intervendrían en la meteorización del material original, especialmente la temperatura y la humedad, factores que determinan la velocidad e intensidad de la meteorización.

El relieve es un importante factor a la hora de modelizar las fracciones granulométricas, ya que influyen en los procesos de erosión, transporte y sedimentación. Numerosos autores (entre otros, [Gessler *et al.*, 2000](#); [Pachepsky *et al.*, 2001](#); [Wilcke *et al.*, 2008](#)) han utilizado los factores morfométricos como variables de entrada, encontrando por ejemplo que la altitud mostraba una correlación positiva con el contenido en arena, mientras que el contenido en arcilla estaba negativamente correlacionado con la altitud y la pendiente. La relación entre la curvatura y la distribución de partículas varía sin embargo en función de la escala ([Brown *et al.*, 2004](#)). Por otro lado, los organismos pueden influir en la distribución de las partículas, pudiendo colaborar en la meteorización a través de la acción física de las raíces o la acción química de los exudados y productos del metabolismo, o reteniendo los sedimentos frente la erosión hídrica o eólica. Sin embargo, en zonas de relieves bajos o suavemente ondulados, los factores ambientales tales como el relieve o la vegetación no son buenos indicadores de la variabilidad espacial de la distribución de partículas ([McKenzie y Ryan, 1999](#); [Zhu *et al.*, 2010](#)), por lo que en los últimos años se están utilizando técnicas de análisis de imágenes obtenidas por sensores remotos para la modelización de la distribución de las arcillas en superficie ([Lagacherie *et al.*, 2008](#); [Liu *et al.*, 2012a](#); [Ließ *et al.*, 2012](#)).

1.2.4 Densidad aparente

La densidad aparente (D_a) del suelo describe el estado de compactación del suelo y la estructura del suelo, relacionada con la organización de las partículas y el espacio de poros dentro del mismo. De este modo, la densidad aparente depende del tamaño y la densidad de las partículas minerales y orgánicas, y cómo estas están agrupadas.

La densidad aparente está determinada principalmente, por lo tanto, por el contenido de CO del suelo, y la proporción de las fracciones granulométricas (arena, limo y arcilla) o la clase textural ([Rawls, 1983](#); [Nemes y Rawls, 2004](#)), además de la profundidad. En zonas áridas, la densidad aparente es a la vez causa y efecto de la distribución en mosaico de la vegetación: por un lado, bajo la vegetación, el aporte de materia orgánica y la bioturbación debida a las raíces disminuye la densidad aparente y favorece la infiltración respecto las zonas desnudas de alrededor; esto convierte a estas partes del suelo en islas de recursos donde se acumulan nutrientes y agua, que a su vez promueven el crecimiento de las plantas ([Mora y Lázaro, 2014](#); [Stavi *et al.*, 2008](#)).

Sin embargo, la densidad aparente en la capa arable no es una propiedad constante, sino que varía en función de las condiciones a las que está sometida el suelo, como por ejemplo la cobertura vegetal, las actividades de manejo y laboreo ([Logsdon y Cambardella, 2000](#)), o el uso de maquinaria pesada ([Franzluebbers, 2005](#)). También puede sufrir variaciones significativas durante un mismo año, debido por ejemplo al pastoreo o la estacionalidad de las precipitaciones ([Timm *et al.*, 2006](#)), los ciclos de congelación y deshielo y los de humectación y secado en suelos ricos en esmectitas ([Pires *et al.*, 2008](#)).

Este parámetro se considera un indicador de la porosidad y el grado de compactación del suelo, de modo que el aumento de la densidad aparente se relaciona con la aireación del suelo, la resistencia a la germinación y la penetración de las raíces y por tanto la reducción del sistema radicular (Boone y Veen, 1994), que en el caso de los cultivos conduce a menores rendimientos de los mismos (Bauder *et al.*, 1981; Veihmeyer y Hendrickson, 1948). Por el contrario, los suelos demasiado sueltos se asocian con una reducción del transporte de agua y nutrientes bajo condiciones no saturadas. La estimación de la densidad aparente es también necesaria para calcular la capacidad de retención hídrica del suelo, así como la conductividad hidráulica.

Además, la determinación de la densidad aparente permite la conversión de medidas en unidades de masa, en otras referidas al volumen, como por ejemplo la capacidad de secuestro de carbono en los ecosistemas, o el almacenamiento de nutrientes en el suelo (Lark *et al.*, 2014).

A pesar de todo ello, la densidad aparente es un dato a menudo ausente en las bases de datos de suelos, debido al alto coste que tiene su determinación, y el esfuerzo que requiere. Debido a este problema, la densidad aparente se estima frecuentemente por medio de funciones de edafotransferencia.

Desde los años noventa, han sido elaboradas muchas PTFs para la estimación de la D_a , que han sido a su vez recogidas en diversos estudios comparativos, entre los que se encuentran los realizados por de Vos *et al.* (2005) o Kaur *et al.* (2002). En la mayoría, se utiliza el contenido de CO y la proporción de las fracciones granulométricas para la estimación de la densidad aparente, aunque también pueden incluirse otras propiedades físico-químicas, como el pH y el contenido de carbonato cálcico (Brahim *et al.*, 2012). Estas últimas, aunque no tienen una relación directa con la densidad aparente (Bernoux *et al.*, 1998) podrían estar relacionadas con el contenido en el catión Ca y su acción estabilizadora sobre los agregados del suelo, resultando en una mayor porosidad y por tanto, menor densidad aparente (Dupouey *et al.*, 1997).

1.2.5 Capacidad de retención de humedad

La determinación de la capacidad de retención de humedad del suelo es necesaria para describir la disponibilidad de agua para las plantas, por lo que ha sido de interés para diversos fines, entre los que se encuentran la agricultura, la ecología y la hidrología, ya que es útil para modelizar el comportamiento del agua y los solutos en la zona no saturada. La capacidad de retención se caracteriza a través de la curva de retención de humedad, que relaciona el contenido de agua en el suelo, expresado en unidades de peso o volumen, con valores determinados de potencial matricial. Esta curva adquiere diversas formas dependiendo de características del suelo tales como la textura, la temperatura, o el contenido de humedad previo a la determinación, ya que la curva es diferente en el proceso de desecación y el de humectación, fenómeno que se denomina histéresis. Sin embargo, debido al elevado coste y el tiempo que requiere su determinación, a menudo solamente se dispone de los valores correspondientes a la capacidad de campo (CC) y el punto de marchitamiento permanente (PMP); la diferencia entre ambos se denomina agua útil (AW), y es considerada la cantidad de agua disponible para las plantas.

Se define como **Capacidad de Campo (CC)** la cantidad de agua retenida en el suelo tras haber sido saturado y dejado drenar libremente en ausencia de evapotranspiración; a partir de un cierto

periodo de tiempo, la tasa de drenaje disminuye y la cantidad de agua en el suelo permanece en un aparente equilibrio, que depende de diversos factores. Debido a esta dependencia y la dificultad de realizar el experimento, la capacidad de campo es determinada en el laboratorio succionando el agua a un potencial específico, que varía en un amplio rango, desde -2.5 kPa a -50 kPa, siendo el valor de -33 kPa (Richards y Weaver, 1944) una de las aproximaciones más usuales. El contenido de humedad a capacidad de campo está condicionado por diversos factores, destacando entre ellos la macroporosidad y la estructura, lo cual hace recomendable su determinación utilizando muestras inalteradas de suelo (Minasny, 2009). Según Rawls *et al.* (2003), el contenido en CO afecta tanto a la estructura como a la propiedades de absorción del suelo y, por tanto, la retención de humedad estaría influida por los cambios en la materia orgánica del suelo, aunque los resultados de diversos estudios sobre tal relación son contradictorios. Pachepsky *et al.* (2001) describieron una correlación relativamente alta entre el agua retenida por el suelo a potenciales comprendidos entre de -10 y -33 kPa y el contenido de arena y limo; además, relacionaron el contenido de humedad a estos valores de potencial con la pendiente y la curvatura tangencial.

Al disminuir el contenido de humedad, el potencial mátrico disminuye, y el agua del suelo es retenida con más fuerza en los poros del suelo debido a las fuerzas de cohesión entre las moléculas de agua y las partículas minerales. El agua puede ser extraída del suelo debido a la evapotranspiración, hasta alcanzarse el Punto de marchitamiento permanente (PMP), que se define como el contenido máximo de humedad al cual la mayoría de las plantas se marchitan y no se recuperan al volver a rehumedecer. Richards y Weaver (1944) consideraron que esto ocurría a tensiones de aproximadamente -1500 kPa. El agua a este potencial está asociada con microporos parcialmente llenos y las superficies hidratadas de las partículas. Al contrario que el contenido de humedad a capacidad de campo, el contenido de humedad en el punto de marchitamiento muestra mayor correlación con la composición granulométrica del suelo, particularmente las arcillas, que con la estructura; por ello, puede ser determinado a partir de muestras alteradas (Aina y Perisawamy, 1985). No obstante, Pachepsky *et al.* (2001) encontraron menor relación entre la retención de humedad y el contenido en arcilla que otros estudios en los que esta fracción tiene un papel determinante, como por ejemplo, Rawls *et al.* (1991); los autores anteriores tampoco encontraron relaciones significativas entre las variables morfológicas y la retención de humedad a -1500 kPa. En relación a la materia orgánica, a pesar de la capacidad de este constituyente para retener humedad, su influencia es menor a -1500 kPa que a -33 kPa, aunque puede modificar la disponibilidad para el agua de los puestos de adsorción en los minerales de la arcilla (Rawls *et al.*, 2003). Por otro lado, McBride y MacIntosh (1984) encontraron que la materia orgánica sólo influía en la retención de humedad a -1500 kPa cuando su contenido era superior al 5%. Por el contrario, el carbonato cálcico tendría una influencia negativa en la retención de humedad, debido a la baja energía de adsorción de este constituyente (Khodaverdiloo *et al.*, 2011).

1.3 Objetivos

El objetivo de la presente tesis doctoral, es combinar técnicas estadísticas y el uso de sistemas de información geográfica para modelizar y representar algunas propiedades edáficas de la capa

arable de los suelos (30 cm superficiales) antes mencionadas, al considerarse de interés para conocer y evaluar la calidad del suelo y, de forma particular, su capacidad de almacenamiento de humedad. Para ello, se utilizan diversas técnicas estadísticas, correlacionando esas propiedades edáficas con diversas variables ambientales, y se representan los resultados en forma de cartografía temática, discutiendo los resultados y la precisión de los modelos, teniendo en cuenta la aplicabilidad de los mismos en investigaciones futuras sobre gestión de recursos hídricos y edáficos y la prevención de la degradación de los mismos. En el caso del carbono orgánico y carbonato cálcico, existe además un interés añadido, relacionado con la importancia de los suelos como sumidero de carbono, tanto en forma orgánica como inorgánica. En relación a los diferentes capítulos de la tesis, los objetivos generales se exponen a continuación:

1. Elaborar modelos a escala regional que permitan predecir el contenido de diversos constituyentes del suelo: carbono orgánico, carbonato cálcico total o equivalente, y fracciones granulométricas (arena, limo y arcilla) y representar cartográficamente el resultado de los modelos obtenidos a escala regional.
2. Obtener modelos que permitan predecir a escala local las mismas variables estudiadas en el capítulo anterior, con el fin de compararlos con los modelos obtenidos a escala regional y analizar las diferencias que puedan encontrarse.
3. Analizar la validez de diferentes funciones de edafotransferencia procedentes de la literatura existente para la estimación de propiedades hídricas de los suelos (densidad aparente, contenido de agua a capacidad de campo y en el punto de marchitamiento), y obtener en su caso nuevas funciones de edafotransferencia que se adecúen a los datos disponibles y las características de los suelos de la región de Murcia.

Capítulo 2

Material y métodos

2.1 Zonas de estudio

El área de estudio para la realización de esta tesis es la región de Murcia (Figura 2.1). Aunque tendría más sentido hidrológico trabajar a escala de cuenca, existe una limitación en la disponibilidad de datos, ya que la capa del mapa de suelos sólo se encuentra disponible para la región de Murcia. Del mismo modo, existe una limitación en la disponibilidad de datos analíticos de suelos para determinadas zonas fuera de la región, por lo que la extrapolación de los modelos a la totalidad de la cuenca en trabajos futuros, haría necesarios nuevos muestreos en esas zonas para poder realizar la correspondiente validación. De este modo, se han elaborado los modelos elaborados a escala regional para el ámbito de la región.

Dentro de la región de Murcia, se encuentran las zonas seleccionadas para la obtención de modelos a escala local (sierra de Mojantes) y la validación y obtención de funciones de edafotransferencia (sierra de las Moreras).

La selección de la sierra de Mojantes para la elaboración de los modelos a escala local se justifica por ser una zona incluida dentro del ámbito de los modelos a escala regional, donde se disponía de una base de datos adicional para las propiedades físicas estudiadas con una malla de muestreo más detallada. Además, la zona puede considerarse representativa de un área montañosa, con cotas elevadas y pendientes pronunciadas, cuya orientación determina unas características climáticas más frescas y húmedas que en el resto de la región.

Respecto a la sierra de las Moreras, se ha seleccionado esta zona para la validación y elaboración de funciones de edafotransferencia por disponerse de datos de propiedades hídricas de suelos no existentes o escasos para las zonas descritas anteriormente. Además, la zona también está incluida dentro del ámbito de aplicación de los modelos a escala regional.

2.1.1 Región de Murcia

La región de Murcia se sitúa en el sureste de la Península Ibérica, entre las coordenadas UTM $X_{min} = 509000, X_{max} = 740000, Y_{min} = 4120000$, e $Y_{max} = 4320000$, correspondientes al huso 30N, considerando el ETRS89 como sistema de referencia. La superficie de la región es de 11.854 km² englobando la práctica totalidad de la Cuenca del Segura (CHS, 2013), mientras que del resto de su superficie, 4.341 km² pertenecen a la provincia de Albacete (Castilla-La Mancha), 1.529 km² pertenecen a la provincias andaluzas de Almería, Jaén y Granada; y 1.154 km² a la provincia de Alicante (Comunidad Valenciana).

La región se caracteriza por una elevada variabilidad ambiental. Hacia el noroeste se encuentran las zonas más elevadas, alcanzándose los 2000 m s.n.m., y con mayor pendiente. Al norte de la región, se encuentra el Altiplano, con altitudes de entre 500 y 1000 m y topografía más suave, limitando al sur con pendientes más acusadas. La mayor parte de la región se encuentra a altitudes inferiores a los 500 m s.n.m, en la cual pueden encontrarse valles y depresiones que coinciden con cursos fluviales (Segura y Guadalentín). Al este y sureste de la región están situadas las cotas más bajas, inferiores a los 200 m y las pendientes más suaves, exceptuando las cadenas montañosas que bordean el litoral sur.

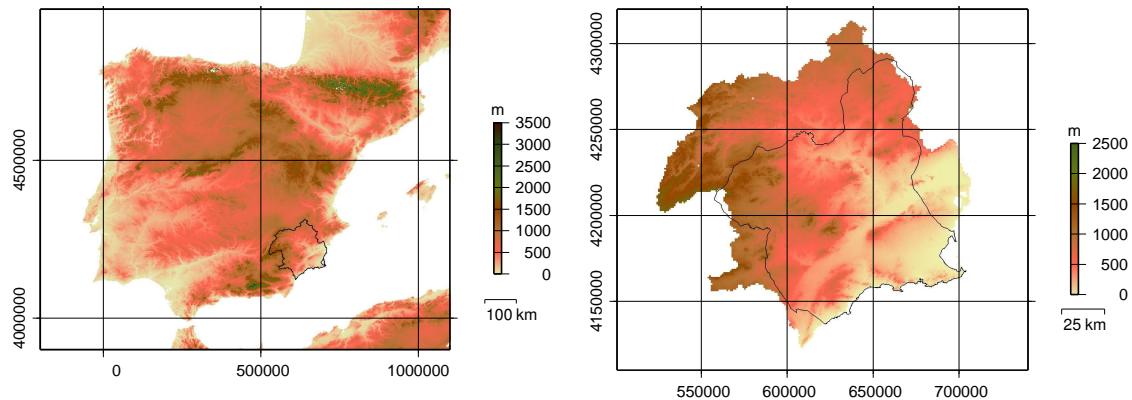


FIGURA 2.1: **Izquierda:** Situación de la cuenca del Segura en la Península Ibérica. **Derecha:** Límite de la Región de Murcia dentro de la cuenca.

Esta diversidad orográfica se corresponde con una diversidad climática. En el noroeste, las zonas situadas a cotas topográficas más altas y con relieve más accidentado son las que presentan a su vez temperaturas medias más bajas, con entradas de aire polar que producen fuertes heladas. También es en la cabecera de la cuenca donde se produce la mayor cantidad de precipitaciones, favorecidas por la disposición SO-NW de los altos arcos montañosos, que impiden el avance de las borrascas atlánticas hacia el sureste; en consecuencia, las tasas de evapotranspiración son las menores de la región. Por el contrario, al sureste de la región, y en el valle del Guadalentín-Segura, se observan las temperaturas medias más altas, y también las máximas absolutas coincidiendo con la entrada de vientos procedentes de África. Al sureste se registran también las menores tasas de precipitación, siendo la depresión de Águilas, situada en el extremo sur de la

cuenca, la que menor cantidad de lluvias recibe a lo largo del año. Estos datos se corresponden con altas tasas de evapotranspiración estival y anual, que además presenta valores máximos en las laderas de las montañas orientadas al sur, con la excepción de Sierra Espuña. Hay que tener en cuenta, sin embargo, las formas ocultas de precipitación y la alta humedad ambiental de las zonas costeras, que disminuyen la tasa de evapotranspiración y aumentan por tanto la cantidad de agua disponible para las plantas. Además de esta distribución anual de precipitaciones, es importante destacar el carácter torrencial de las mismas, principalmente en las zonas medias y bajas de la cuenca, debido a la aparición de fenómenos convectivos mediterráneos. Por todo ello, el clima en la región se identifica como Mediterráneo (Papadakis, 1966), presentándose los subtipos Mediterráneo templado, Mediterráneo continental, Mediterráneo subtropical, y Mediterráneo semiárido subtropical.

Desde un punto de vista geológico, se pueden encontrar elementos de la zona Bética, Prebética y Subbética, así como materiales terciarios post-orogénicos y cuaternarios. De la zona Bética está representados los tres dominios: Nevado-Filábride, Alpujárride y Málagaide. El complejo Nevado-Filábride, formado por rocas metamórficas (micasquistos, cuarcitas) y mármoles, se ve rodeado por el complejo Alpujárride, que litológicamente consta de micaesquistos, conglomerados, areniscas y yesos sobre los cuales se asientan calizas, dolomías y calcoesquistos. El complejo Málagaide, formado por calizas, dolomías y conglomerados, puede encontrarse en Sierra Espuña, además de en Cabo Cope y el sur de Carrascoy.

El Prebético externo muestra una litología formada predominantemente por dolomías, acompañadas de materiales detríticos y evaporitas (yesos y sal), característica de medios de depósito diversos, desde continentales a costeros. El Prebético interno, representado en el Altiplano, al norte de la cuenca, está formado por calizas y dolomías, con margas y margocalizas, depósitos propios de plataforma continental, con frecuentes y diversos residuos fósiles. El Prebético meridional, aflorante en Moratalla y Cieza, presenta características intermedias entre los dos anteriores.

El dominio Subbético está formado tres sectores: externo, medio e interno. El Subbético externo se compone de potentes tramos de margas, intercaladas con calizas y areniscas; el Subbético interno, aflorante en la sierra de Ricote, lo forman materiales de depósito, margas, calizas y dolomías, viéndose afectado por fenómenos de cabalgamiento; entre ellos se encuentra el Subbético medio, de mayor extensión que los anteriores, donde destacan las calizas, algunas de ellas con sílex, y rocas volcánicas intercaladas.

Los materiales terciarios postorogénicos engloban los depósitos alojados en las depresiones, destacando entre ellas el Campo de Cartagena, de relleno margoso con intercalaciones de conglomerados y calizas, y la depresión de Mula, con sedimentos margosos de gran espesor; a diferencia de estos sedimentos de origen marino, la depresión de Lorca presenta sedimentos continentales, donde son frecuentes los depósitos de evaporitas. Los afloramientos volcánicos son abundantes en la cuenca, particularmente de rocas lamproíticas y ultrapotásicas. Finalmente, los materiales cuaternarios son fundamentalmente de origen continental, destacando el valle fluvial del Guadalentín-Segura, y los depósitos del Campo de Cartagena, cubriendo los materiales terciarios, excepto el Cabezo Gordo, que constituye un monte-isla de materiales nevado-filábrides.

Como consecuencia de las características climáticas, fisiográficas, geológicas y litológicas descritas, así como de factores ecológicos (vegetación y actividad antropogénica), en la región se

han desarrollado diferentes tipos de suelo, descritos en los mapas y memorias elaboradas para la región (LUCDEME, 1986-2004).

Las zonas de topografía más abrupta suelen estar ocupadas por Leptosoles. Son suelos limitados en profundidad por una roca dura o material muy calcáreo (con un contenido en carbonato cálcico superior al 40%) dentro de los 25 cm a partir de la superficie o que contienen menos del 10% de tierra fina, caracterizados por un perfil A-R, sin otro horizonte diagnóstico que no sea A ócrico o móllico. Principalmente, se encuentran Leptosoles Líticos (IUSS-WRB, 2007, antes denominados Litosoles) y Leptosoles Réndzicos (IUSS-WRB, 2007, antes considerados Rendzinas). Los Leptosoles Líticos representan la tercera unidad por extensión en la región, y están caracterizados por una limitación en profundidad (menos de 10 cm), por lo que aunque estén cubiertos por abundante vegetación y posean un horizonte superficial rico en materia orgánica, no cumplen las condiciones de espesor para ser un horizonte móllico. Descansan sobre rocas de diferentes tipos, que pueden ser calizas, dolomías, areniscas, conglomerados, o rocas volcánicas. Los Leptosoles Réndzicos presentan un horizonte A móllico situado inmediatamente encima de material calcáreo, con un contenido de CO_3Ca superior al 40%, por lo que se encuentran representados en lugares donde hay calizas, margas y margocalizas con una densa cobertura vegetal, generalmente en la zona septentrional de la región y zonas de umbría, donde la incorporación de materia orgánica a la fracción mineral, siendo por tanto suelos con buenas condiciones para la vegetación, cuya limitación más importante es el espesor. Suelen estar asociados a Leptosoles Líticos y, en ocasiones, a Kastanozems Cálcidos, aunque esta última unidad está escasamente representada en la región.

Los Regosoles son suelos formados por materiales no consolidados que se encuentran muy escasamente evolucionados, siendo el horizonte A ócrico su único horizonte de diagnóstico. Esta escasa evolución se debe al continuo rejuvenecimiento del perfil motivado por los procesos de erosión y aporte, que no permiten las transformaciones que dan lugar a su desarrollo. La falta de evolución produce suelos muy uniformes, con escasa variabilidad de sus características a lo largo del perfil. Sus características y propiedades están muy influenciadas por la naturaleza del material litológico sobre el que se forman. Son suelos ampliamente representados en la región, cubriendo aproximadamente una quinta parte de su superficie, y se encuentran bajo diferentes usos, principalmente agricultura de secano, vegetación natural y terrenos marginales. De ellos, los más ampliamente extendidos en la región son los Regosoles Calcáricos, con un elevado contenido en carbonato cálcico, que suelen formarse a partir de margas neógenas y cretácicas. Son suelos de textura fina y de escasa permeabilidad, lo cual produce una elevada escorrentía superficial de las escasas pero en ocasiones intensas lluvias, provocando importantes procesos erosivos. Por otro lado, esta baja permeabilidad provoca también que una escasa proporción del agua quede disponible para las plantas, por lo que la vegetación suele ser también poco abundante, así como la incorporación de materia orgánica. Estos suelos dan lugar a paisajes de badlands, muy característicos, que pueden encontrarse en depresiones como las de Fortuna o Mula. Una parte importante de estos suelos se encuentran asociados a Calcisoles, Gipsisoles y Litosoles. En la región también pueden encontrarse otro tipo de Regosoles, aunque menos representativos por ocupar una superficie mucho menor, que son los Regosoles Eútricos y los Regosoles Lépticos.

Los Fluvisoles son suelos característicos de llanuras aluviales y fondos de valle, relacionándose con los ríos, arroyos y ramblas, barrancos y cañadas existentes en la región, y ocupan alrededor del

10% de su superficie, destacando por su extensión los que forman los valles del Guadalentín y del Segura. Son suelos caracterizados por una estratificación a lo largo del perfil, y una disminución irregular del contenido de materia orgánica en el mismo, que suele ser de tipo A-C1-C2-C3-...-Cn, donde los horizontes subyacentes (C1,...,Cn) tienen texturas muy variables, dependiendo de la energía que llevara el agua en los diferentes episodios de depósito. Los materiales depositados proceden de la erosión y posterior acumulación de suelos y productos de la meteorización de las rocas situadas en posiciones topográficamente más altas, los cuales les transmiten muchas de sus características y propiedades. Los Fluvisoles de la región son suelos utilizados tradicionalmente para cultivos y huertas, por lo que frecuentemente el horizonte superficial es un Ap (antrópico), ya que tienen un gran potencial agrícola y que la instalación de las poblaciones agrícolas precisa de aportes regulares de agua. Los Fluvisoles más abundantes y representativos por la naturaleza calcárea de la mayoría de las rocas y los suelos circundantes son los Fluvisoles Calcáricos, caracterizados por un contenido variable de carbonato cálcico (entre el 20% y el 50%), que pueden presentarse asociados a Calcisoles, Regosoles y Solonchaks.

Cuando se dan las condiciones para un mayor desarrollo edáfico, se encuentran Calcisoles (Xerosoles Cálcidos y Petrocálcidos y Cambisoles Cálcidos de [LUCDEME, 1986-2004](#)), que tienen como característica la acumulación de carbonato cálcico y un horizonte superficial ócrico aunque puede tener abundante materia orgánica. Son la unidad más ampliamente representada en la región, formándose a partir de materiales detríticos suficientemente permeables para que se produzca el lavado y posterior acumulación de carbonato, tanto en superficies llanas, como en laderas coluviales, superficies de pie de monte, conos de deyección, etc. El carbonato cálcico puede proceder tanto del material sobre el que se forma el suelo, como de zonas próximas por procesos de lavado lateral. Existe una gran variedad de Calcisoles, siendo los Calcisoles Háplidos los más representativos, junto con los Calcisoles Pétricos (antiguos Xerosoles Petrocálcidos), que conforman la segunda unidad por orden de extensión. Ambos aparecen frecuentemente asociados entre sí o con otras unidades. Los Calcisoles Háplidos muestran acumulaciones de carbonato cálcico en profundidad en forma de manchas pulverulentas o nódulos, y ocupan grandes superficies, como la práctica totalidad del Campo de Cartagena, los términos municipales de Lorca, Totana y Mazarrón y en el sector nororiental de la región. Son suelos ampliamente utilizados con fines agrícolas. Cuando la acumulación de carbonatos es muy intensa, se forman Calcisoles Pétricos, caracterizados por presentar capas continuas y endurecidas de este material (horizontes petrocálcidos Ckm) dentro de los primeros 100 cm desde la superficie, y que se observan frecuentemente sobre depósitos coluviales muy antiguos, zonas de laderas suaves y superficies de glacis donde son frecuentes los procesos de lavado lateral. Los procesos de disolución y recristalización de carbonato son el origen de las costras calizas fuertemente cementadas, que constituyen un factor limitante de la profundidad del perfil, por lo que la roturación requiere de maquinaria capaz de fragmentarlas. Por otro lado, los restos fragmentados de estas costras protegen contra la erosión y la evaporación del agua del suelo, mejorando las propiedades hídricas del mismo. Se encuentran generalmente bajo uso agrícola de secano, sobre todo de leñosas (almendros y viñedos), en las zonas del Altiplano y el Noroeste.

Los Gipsisoles (Xerosoles gípsicos, [LUCDEME, 1986-2004](#)) se caracterizan también por presentar acumulación de sales, en este caso de yeso, ya sea en forma de horizonte gípsico o petrogípsico que aparece dentro los 100 cm desde la superficie; el horizonte superficial es siempre ócrico y el

perfil es de tipo A-Cy. Se desarrollan a partir de afloramientos de arcillas y margas yesíferas del Triás, principalmente en el Altiplano, o de margas neógenas con intercalaciones de yesos, y están asociados a un régimen de humedad arídico. Existen varios tipos de Gipsisoles en función del contenido de yeso y su estado de cementación, agrupándose los que se encuentran en la región dentro de los Gipsisoles Háplicos, aunque solo representan un 1% de la superficie total. También aparecen asociados a otras unidades, como Calcisoles, Regosoles, Leptosoles, Fluvisoles o incluso Solonchaks.

Otros suelos minoritarios dentro de la región son los Solonchaks, caracterizados por presentar un horizonte sálico en los primeros 50 cm desde la superficie del suelo, sin poseer materiales flúvicos. Aparte de esto son suelos muy heterogéneos, que presentan una gran variabilidad en cuanto a la naturaleza de las sales, y la salubridad y profundidad de la capa freática que además puede oscilar a lo largo del año, por lo que pueden presentar una gran diversidad de horizontes y propiedades tanto en superficie como en profundidad, aunque los que se encuentran en la región se consideran Solonchaks Háplicos. Aparecen en saladares y marjales, tanto en el litoral (Calblanque, San Pedro del Pinatar, Carmolí) como en el interior en zonas semiendorreicas y ramblas (Saladares del Guadalentín, Rambla Salada, El Ajauque), debido a la existencia de materiales geológicos formados por margas ricas en evaporitas y otras sales más solubles que el yeso. En periodos de gran evaporación, las sales pueden ascender a la superficie y acumularse formando eflorescencias salinas o incluso costras. Se distinguen por albergar una vegetación y fauna particular de gran interés ambiental.

Cuando el desarrollo del perfil es algo mayor, en zonas con abundante materia orgánica, se forman Kastanozems, aunque esta unidad es escasamente representativa, pues actualmente ocupan solamente un 0.01% de la superficie regional. Se caracterizan por poseer un horizonte superficial móllico y concentraciones de carbonatos secundarios dentro de los primeros 100 cm del perfil, además de no tener horizontes diagnósticos que no sean árgico, cálcico, cámbico, gípsico o vértico. Se encuentran distribuidos por toda la región, clasificados como Kastanozems Cálcicos en zonas forestales de montaña situadas en la umbría, sobre materiales desarrollados a partir de coluvios calizos, heterométricos y sin consolidar. Los procesos edafogénicos que los caracterizan son por un lado la humificación, debido a la aportación de materia orgánica, la escasa mineralización y la protección ante la erosión que proporcionan los restos vegetales en superficie; y por otro lado el lavado de carbonatos que da lugar a un perfil A-Ck y en ocasiones incluso A-Bw-Ck. En la actualidad están escasamente representados, debido a la puesta en cultivo de grandes superficies bajo vegetación natural, que ha afectado al proceso de humificación, acelerándose al tiempo la mineralización de la materia orgánica; en consecuencia, el horizonte superficial pierde la consideración de móllico y pasa a ser ócrico. De esta forma, muchos de los actuales Calcisoles, Gipsisoles y Regosoles proceden de la degradación de Kastanozems Cálcicos, Gípsicos y Háplicos.

En lo relativo a los usos y coberturas del suelo en el momento de la elaboración de la base de datos ([European Environment Agency, 1995](#)), la mayor parte del territorio de la región estaba influenciada por la acción humana, principalmente por diferentes tipos de cultivo. La mayor parte de esta superficie agrícola era de cultivo en mosaico (20%), seguida por cultivos arbóreos (16%) que se distribuían por toda la región, incluyendo frutales de regadío (albaricoque, melocotón, cítricos) y leñosas de secano (almendro, viñedo y olivo), estas últimas de gran importancia en las comarcas situadas al norte y noroeste. Los cultivos herbáceos de secano eran de cereales,

tales como cebada y avena, y ocupaban alrededor de un 12% de la superficie de la región, principalmente en el noroeste y en algunos términos municipales como Lorca, mientras que los cultivos de regadío, principalmente hortícolas, representaban una superficie mucho menor, solamente el 7% de la superficie regional, pero más localizados en zonas determinadas, principalmente el Campo de Cartagena y las vegas del Segura y el Guadalentín. Las superficies artificiales, haciendo referencia a los usos urbanos y las infraestructuras derivadas, representaban un 1% de la superficie.

De los usos naturales, la vegetación de matorral era la mayoritaria, ocupando un 21% de la superficie. Esta categoría incluye el matorral esclerófilo característico formado por coscojales, lentiscales, palmitares y espinales típicos de vaguadas y ramblas, junto con tomillares y espartales, mientras que en las zonas litorales son típicos los cornicales; también se incluyen en este uso las comunidades de matorrales almohadillados de las zonas más frías. A este uso lo seguían en extensión los bosques (10% de la superficie), principalmente de pináceas, pero entre los que se podían encontrar especies perennifolias, quejigos, sabinares y enebrales, junto con especies caducifolias como robles (*Quercus faginea*) o arces (*Acer granatense*). Una pequeña proporción de la superficie la conformaban afloramientos rocosos y canchales (6%) donde el suelo se presentaba mayoritariamente desnudo. Finalmente hay que mencionar, a pesar de que representaban una mínima proporción de la superficie regional (inferior al 1%), la vegetación típica de humedales, tanto los que tienen una lámina libre de agua (charcas, fuentes) como los criptohumedales, dependientes de las aguas freáticas (saladares).

2.1.2 Sierra de Mojantes

La sierra de Mojantes se sitúa en el término municipal de Caravaca de la Cruz (Figura 2.2), entre las coordenadas UTM del Huso 30 N $X_{min} = 570000, X_{max} = 600000, Y_{min} = 4200000$, e $Y_{max} = 4220000$ (Figura 2.2). Se corresponde con el Mapa Topográfico Nacional, Hoja 889 (Moratalla) a escala 1:50.000. La elevación media de la zona es de 1000 m, alcanzándose los 1800 m s.n.m. Las pendientes varían entre el 1% de las llanuras planas que ocupan el centro y sur de la hoja, y el 30% en las sierras con laderas más empinadas, siendo el valor medio aproximado del 5%.

Las rocas que afloran en la zona pertenecen a las zonas externas de las Cordilleras Bética, concretamente al dominio subbético. Junto a estos también se han descrito materiales neógeno-cuaternarios que ocupan las posiciones topográficas más deprimidas. Los materiales subbéticos se componen en su mayor parte de calizas y dolomías, alternadas con materiales margosos, margocalizos y arenisca, por lo que la diversidad de su composición es mucho menor que en la zona de la sierra de las Moreras. Los materiales cuaternarios se clasifican mayormente como cuaternario indiferenciado, formado por conglomerados y brechas, cementados por una matriz caliza; existen además arenas y bloques poco cementados formando conos de deyección y materiales aluviales, y algunos travertinos donde se producen surgencias de aguas carbonatadas. Los materiales prebéticos apenas tienen representación.

En esta zona se han cartografiado distintos tipos de suelos en función de diferentes factores, como el uso, el material geológico, la topografía, etc. En las zonas más altas y con mayores

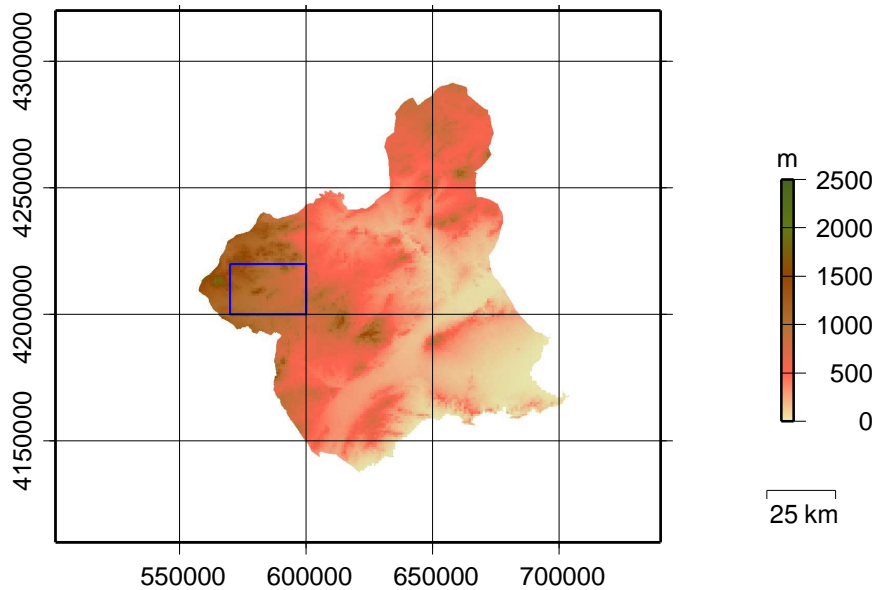


FIGURA 2.2: Zona de estudio en la sierra de Mojantes, delimitada en azul.

pendientes se encuentran Leptosoles. En las laderas se encuentran Calcisoles, y Kastanozems, en función de la orientación y del uso que se hace de los mismos, en los que el horizonte de acumulación de CO_3Ca está más o menos cementado dependiendo de la inclinación o pendiente de estas. En algunas zonas se encuentran Regosoles Calcáricos, en los que no se diferencia un perfil desarrollado. En el cauce del río y las zonas inundables se pueden observar Fluvisoles Calcáricos (WRB, 2007). Según la clasificación edafoclimática de Soil Survey Staff (2010) estos suelos se corresponden con un régimen de humedad xérico y un régimen de temperaturas méxico. Los valores de la precipitación media anual y la temperatura media anual son, respectivamente, 389 mm y 16°C , y la evapotranspiración anual es de 1312 mm, considerándose el clima de la zona como mediterráneo templado.

Los usos del suelo (European Environment Agency, 1995) son predominantemente cultivos de secano en las zonas llanas, estando relegados los cultivos arbóreos y de regadío a las zonas más cercanas a los cursos de agua. En algunos puntos existen mosaicos de cultivos. Respecto a los usos naturales, los bosques están relegados a las zonas más altas de las alineaciones montañosas, predominando los pinares sobre el resto de especies forestales, y mayoritariamente se encuentran formaciones de matorral esclerófilo característico del piso supramediterráneo, tales como coscojares y sabinas.

2.1.3 Sierra de las Moreras

La sierra de las Moreras se sitúa al SO de la región de Murcia, entre las coordenadas UTM del Huso 30 N $X_{min} = 570000$, $X_{max} = 660000$, $Y_{min} = 4150000$, e $Y_{max} = 4170000$ (Figura 2.3). Se corresponde con el Mapa Topográfico Nacional, Hoja 976 (Mazarrón) a escala 1:50.000. Se trata

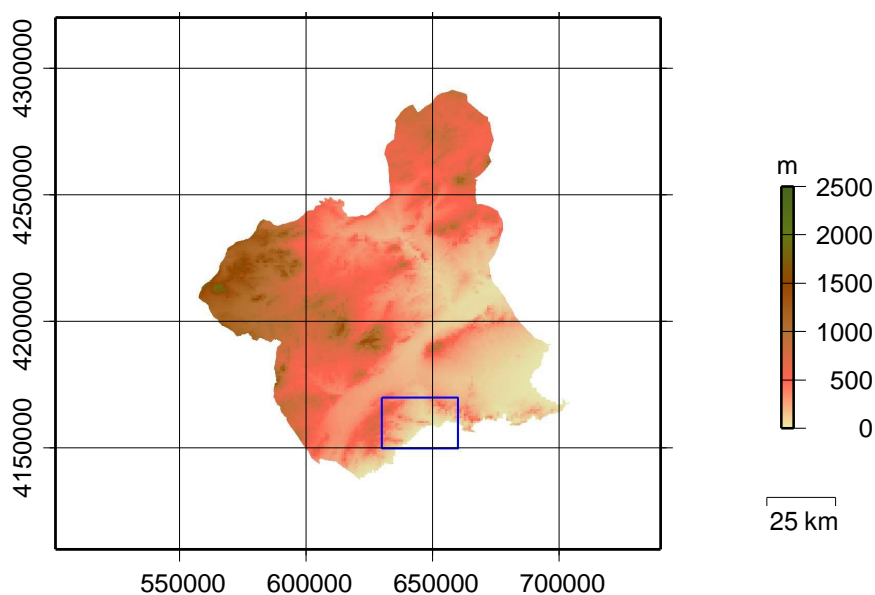


FIGURA 2.3: Zona de estudio en Mazarrón, delimitada en azul.

de una sierra costera, con una elevación media de 213 m, donde la altitud máxima alcanzada es de 886 m. La pendiente media es del 9%, con un amplio rango de variación entre las zonas bajas correspondientes a los fondos de las ramblas, y el 49% de las laderas más abruptas de las sierras que las rodean.

En la zona puede encontrarse una gran diversidad de materiales geológicos (IGME, 1993), incluyendo materiales metamórficos que forman parte de los complejos Nevado-Filábride, Maláguide y Alpujarride de las cordilleras Béticas, representados por esquistos, micasquistos, calcoesquistos, mármoles, filitas, cuarcitas y areniscas, que se alternan con materiales calizos. La zona central de la Hoja 976 está ocupada por una depresión, en cuyos bordes afloran materiales volcánicos básicos, principalmente de dacitas, riodacitas y andesitas, observándose fenómenos de hidrotermalismo que dan lugar a la formación de limonita y jarosita; junto a ellos aparecen además rocas lamproíticas ultrapotásicas al final de la serie. La depresión central está ocupada por rocas sedimentarias, formando una serie detrítica margoarenosa en la que se alternan margas, areniscas y arenas de origen predominantemente continental, coronada por sedimentos cuaternarios procedentes de los depósitos de piedemonte en forma de abanicos aluviales, en los que encontramos conglomerados de cantos unidos por una matriz calcárea. La depresión está surcada por los sistema fluvial de las ramblas de las Moreras y de Ramonete, que forman una red de drenaje encajada que recorta los abanicos mencionados anteriormente.

Los suelos formados sobre estos materiales pertenecen a cuatro de los principales grupos descritos por WRB (2007), encontrándose en orden de porcentaje de superficie ocupada Calcisoles, Leptosoles, Fluvisoles y Regosoles. Según la clasificación edafoclimática de Soil Survey Staff (2010), los suelos de la zona se corresponden con un régimen de humedad arídico y un régimen de temperaturas térmico. Los valores de la precipitación media anual y la temperatura media anual

son, respectivamente, 223.5 mm y 18.4 °C, y la evapotranspiración anual es de 912 mm, que se corresponden con un clima mediterráneo semiárido subtropical.

Los usos del suelo observados en la zona ([European Environment Agency, 1995](#)) son predominantemente agrícolas (68% de la superficie), dentro de los cuales la mayoría (54%) se corresponde a cultivos de regadío y el resto mayormente a cultivos en mosaico, con muy poca superficie ocupada por cultivo de secano. Por otro lado, existe una fuerte ocupación del suelo por usos urbanos e industriales (27%), y solo el 5% de la superficie es vegetación natural, principalmente de matorral.

2.2 Fuentes de información

Con el fin de elaborar los modelos de variabilidad espacial de las propiedades edáficas, se han utilizado diferentes fuentes de información que permiten correlacionar las observaciones empíricas de suelos con variables ambientales:

- **Datos de suelos.** Se han utilizado tres bases de datos con observaciones empíricas de suelos: la base de datos de suelos de LUCDEME ([Alias y Ortiz, 1986–2004](#)), la obtenida para el estudio de los suelos de la sierra de Mojantes y sus alrededores ([Sánchez Navarro, 1995](#)), y una tercera base de datos con observaciones de la sierra de las Moreras, que incluye los datos de [Gil Vázquez \(2010\)](#).
- **Modelo digital de elevaciones (MDE).** Se ha obtenido de la página web del CNIG ([Instituto Geográfico Nacional de España, 2013](#)), con una resolución de 25 m.
- **Datos climáticos.** Se han utilizado los datos de 625 estaciones meteorológicas situadas en la Cuenca del Segura durante el periodo 1975-2002 ([Gomariz Castillo y Alonso Sarría, 2013](#)).
- **Cartografía temática.** Mapas de usos del suelo de CORINE Land Cover 1990 ([European Environment Agency, 1995](#)), y tipos de suelo del mapa digitalizado de suelos elaborado por la CARM ([Ramírez-Santiagosa et al., 1999](#)).

Los datos obtenidos de las diferentes fuentes se han analizado utilizando el software estadístico R y el sistema de información geográfica GRASS.

2.2.1 Datos de suelos

2.2.1.1 Base de datos LUCDEME

Las propiedades del suelo se han obtenido a partir de la base de datos del proyecto LUCDEME (Lucha Contra la Desertificación en el Mediterráneo), que fue elaborada entre 1984 y 2005 y promovida por el antiguo Instituto de Conservación de la Naturaleza (ICONA). El proyecto se enmarca en las actividades recomendadas por el Plan de Acción contra la Desertificación

de las Naciones Unidas (DESCON) establecidas por la conferencia de Nairobi (1977) al ser España el único país occidental con áreas sometidas a graves procesos de desertificación (sureste peninsular), junto con otras en riesgo moderado. Desde su origen a la actualidad, el proyecto ha dado lugar a gran cantidad de información, entre la cual se encuentra la elaboración de 132 hojas de Mapas de Suelos a escala 1:100.000 que abarcan todo el sureste de la península. Originalmente en formato papel, algunos de ellos se han digitalizado posteriormente. Los Mapas incluyen las correspondientes Memorias, en las que consta la descripción de los suelos conforme al sistema de clasificación vigente en esos años (FAO-UNESCO, 1974)

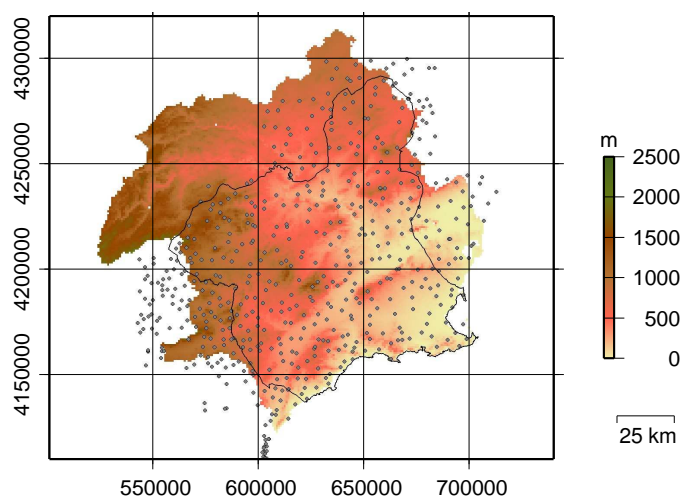
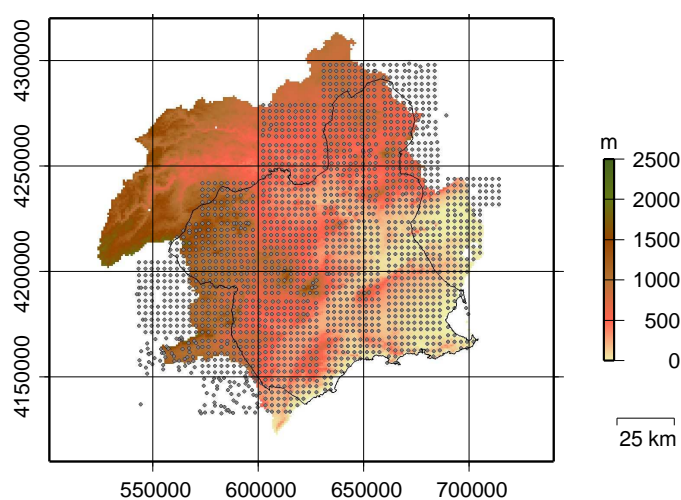


FIGURA 2.4: Base de datos LUCDEME.

La base de datos derivada de este trabajo, recopilada en el Departamento de Química Agrícola, Geología y Edafología de la Universidad de Murcia, incluye los datos empíricos de suelos de la

Cuenca del Segura, correspondientes a 547 muestras de perfiles y 1922 muestras de capa arable (figura 2.4), y que se describen en la tabla 2.1. Las muestras de perfiles (figura 2.4B) incluyen datos de cada uno de los horizontes del perfil, de las cuales en este trabajo solamente se ha tenido en cuenta el horizonte superficial, y no siguen una distribución espacial determinada, pues fueron tomadas con el fin de describir las unidades de suelos que aparecían en los Mapas. Las muestras de capa arable (figura 2.4A), por el contrario, se corresponden con los primeros 30 centímetros superficiales, y están tomadas siguiendo una malla de 3x3 kilómetros.

Variable	Unidades	Descripción
Carbono orgánico (CO)	$g \cdot kg^{-1}$	Anne (1945)
*Nitrógeno total (NT)	$g \cdot kg^{-1}$	Método de Kjeldahl (Duchafour, 1970)
*Relación C:N		
Carbonato cálcico (CO_3Ca) total o equivalente	$g \cdot kg^{-1}$	Método volumétrico del calcímetro de Bernard
*Acidez total (pH_w) y potencial (pH_{ClK})		Peech (1965)
*Capacidad de cambio catiónico (CCC)	$cmol^+ \cdot kg^{-1}$	Chapman (1965)
Conductividad eléctrica (CE)	$dS \cdot m^{-1}$	En extracto de saturación (Rhoades <i>et al.</i> , 1989)
*Capacidad de campo (CC)	% (peso)	Richards y Weaver (1944)
*Punto de marchitamiento permanente (PMP)	% (peso)	Richards y Weaver (1944)
Granulometría		8 fracciones de tierra fina
Arcilla ($0-2 \mu \varnothing$)	%	Pipeta de Robinson
Limo ($2-20$ y $20-50 \mu \varnothing$)	%	Pipeta de Robinson
Arena ($50-100\mu$, $100-250 \mu$, $250-500 \mu$, $500\mu-1 mm$ y $1-2 mm \varnothing$)	%	Tamizado

TABLA 2.1: Datos analíticos incluidos en la base de datos de LUCDEME. Las variables indicadas con asterisco solo están disponibles en las muestras de perfiles.

La base de datos ha sido revisada, para eliminar errores analíticos o tipográficos, en el Departamento de Química Agrícola, Geología y Edafología.

2.2.1.2 Datos de suelos de la sierra de Mojantes

La base de datos de la sierra de Mojantes fue obtenida para el estudio de los suelos situados en la misma y sus alrededores en la tesis de [Sánchez Navarro \(1995\)](#).

En la figura 2.5 se han representado las muestras de la base de datos de LUCDEME (círculos negros) y las muestras utilizadas en los modelos locales para las muestras de Mojantes (triángulos amarillos). La distancia entre estas últimas es mucho menor, ya que las muestras de capa arable forman una retícula de 1000 x 1000 m; existen además otras muestras correspondientes a perfiles, que no se distribuyen en forma de retícula sino que fueron tomados en función del tipo de suelo para representar diferentes unidades.

Las propiedades del suelo disponibles en esta base de datos son el CO ($\text{g} \cdot \text{kg}^{-1}$; [Anne, 1945](#)), el CO_3Ca total o equivalente ($\text{g} \cdot \text{kg}^{-1}$; Método volumétrico del calcímetro de Bernard), arcilla (%; pipeta de Robinson), limo (%; pipeta de Robinson), y la arena (%; tamizado). Se debe comentar que para evitar errores en la validación cruzada, dos de las muestras se reclasificaron como uso de matorral (7) al ser las dos únicas que aparecían sobre suelo desnudo (uso 8). Otras dos muestras de Calcisoles Háplicos con Fluvisoles Calcáricos (2014) fueron reclasificadas como Calcisoles Háplicos (201) por el mismo motivo.

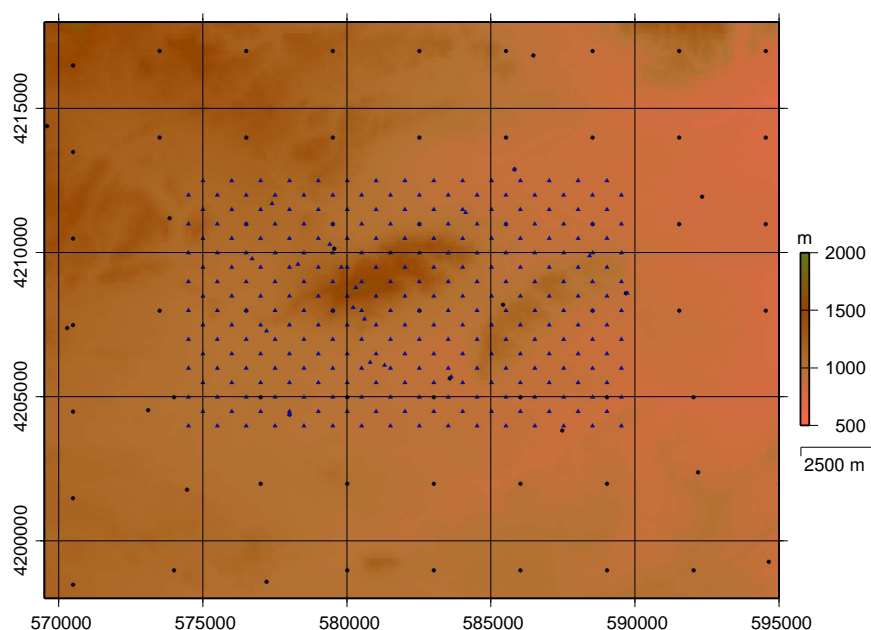


FIGURA 2.5: Mapa de elevaciones (MDE25) con muestras de base de datos LUCDEME (círculos negros) y de la base de datos de Mojantes (triángulos azules).

2.2.1.3 Datos de suelos de la sierra de las Moreras

Los datos de la sierra de las Moreras utilizados en esta tesis fueron obtenidos en un muestreo realizado en 2008, en el que se recogieron 41 muestras inalteradas de capa arable distribuidas siguiendo una malla regular de 3 x 3 km (Gil Vázquez, 2010). Además de las variables CO ($\text{g} \cdot \text{kg}^{-1}$; Anne, 1945), CO_3Ca total o equivalente (%; Método volumétrico del calcímetro de Bernard), CE ($\text{dS} \cdot \text{m}^{-1}$; Medida en extracto de saturación), arcilla (%; pipeta de Robinson), limo (%; pipeta de Robinson), arena (%; tamizado), Capacidad de Campo y Punto de Marchitamiento Permanente (% (peso/peso), Richards y Weaver, 1944), se dispone de datos de Densidad aparente (Da , $\text{g} \cdot \text{cm}^{-3}$; Hénin *et al.* (1969)) que fueron utilizados para la validación y recalibración de las Funciones de Edafotransferencia estudiadas en esta tesis.

Para evitar errores en la validación cruzada, se eliminaron dos muestras situadas sobre suelo desnudo (uso 8); además, las muestras pertenecientes a Unidades de suelos poco frecuentes fueron reclasificadas en grupos más amplios.

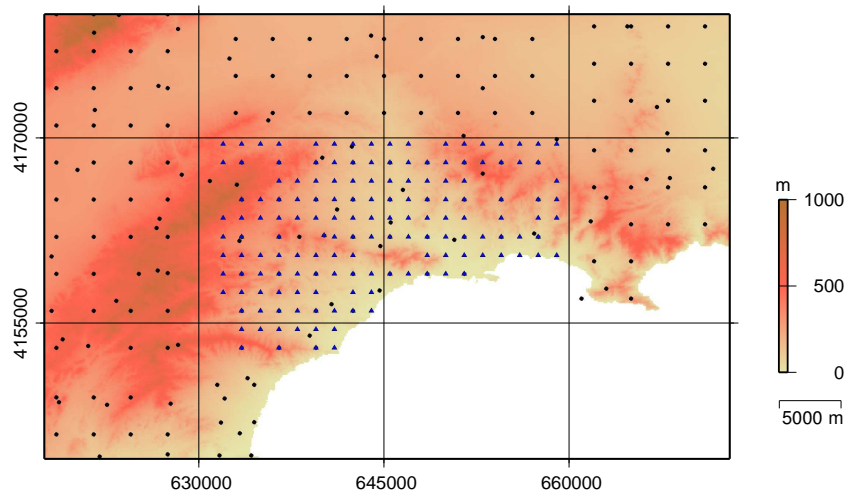


FIGURA 2.6: Mapa de elevaciones (MDE25) de la sierra de las Moreras, muestras de base de datos LUCDEME (círculos negros) y de la base de datos de la sierra de las Moreras (triángulos azules).

2.2.2 Variables ambientales

2.2.2.1 Variables geomorfológicas

Los parámetros morfométricos seleccionados son aquellos que mejor describen los procesos formadores del suelo según diferentes autores (Deng, 2005, Burrough *et al.*, 2000). Dado que el valor puntual de los diferentes parámetros no tiene por qué ser el más adecuado, y que por otra parte existe una gran incertidumbre asociada a la posición de las muestras en los datos del proyecto LUCDEME, se hace necesario incluir las diferentes propiedades geomorfológicas a diferentes escalas.

En este trabajo se ha utilizado el Modelo Digital de Elevaciones (MDE) con una resolución espacial de 25 m, descargado de la página web del CNIG, que a través de herramientas de análisis del terreno permite derivar otros parámetros cuantitativos asociados a los procesos formadores del suelo. Muchos de estos parámetros son operadores de vecindad, es decir, adjudican a cada celdilla un valor que es función de los valores de un conjunto de celdillas contiguas, denominado ventana. La ventana es generalmente de forma cuadrada y su tamaño es variable, definido por el número de celdillas que hay en el lado del cuadrado, que siempre es impar. Se han seleccionado 10 tamaños de ventana (3, 7, 11, 15, 19, 23, 31, 35, y 39 celdillas) para el cálculo de los operadores de vecindad, con el fin de tener en cuenta las diferentes escalas (desde aproximadamente 0.1 a 95 ha).

Variables relacionadas con la altitud

1. **Pendiente (SLO)**. Refleja el mayor gradiente en los valores de elevación, calculada a través de la ecuación 2.2. La pendiente indica el ángulo entre el plano horizontal y el plano tangencial a la superficie, definido por el gradiente $|\nabla Z|$ (ecuación 2.1).

$$\nabla Z = \left(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y} \right) \quad (2.1)$$

La pendiente viene expresada en grados o radianes. En esta tesis, se ha trabajado por motivos prácticos con la pendiente expresada en porcentaje, usando la ecuación 2.3.

$$\text{SLOPE} = \arctan(|\nabla Z|) \quad (2.2)$$

$$\text{SLOPE} = \tan(\text{SLOPE}) \cdot 100 \quad (2.3)$$

2. **Seno de la orientación (SIN)**. La orientación es un parámetro circular que se define como el ángulo existente entre el vector que señala el Norte y la proyección sobre el plano horizontal del vector normal a la superficie en el punto considerado, calculado mediante la ecuación 2.4.

$$\text{ASPECT}(\%) = 180 - \arctan\left(\frac{q}{p}\right) + 90 \frac{p}{|p|} \quad (2.4)$$

donde la orientación se expresa en grados, medidos en el sentido de las agujas del reloj desde el norte; p y q son derivadas de primer orden del gradiente:

$$p = \frac{\partial z}{\partial x}; q = \frac{\partial z}{\partial y} \quad (2.5)$$

Para realizar comparaciones estadísticas se considera en este caso el seno de la orientación, $\sin(\text{ASPECT})$, que alcanza un máximo (1) en orientaciones hacia el este y un mínimo (-1) en orientaciones hacia el oeste.

3. **Coseno de la orientación (COS).** De forma análoga a la variable anterior, el coseno de la orientación, $\cos(\text{ASPECT})$, equivale al efecto solana (coseno=-1) en orientaciones hacia el sur, y umbría (coseno=-1) en orientaciones hacia el norte.
4. **Curvatura del perfil (PROF).** La curvatura en un punto se define como la tasa de cambio de la pendiente y depende por tanto de las derivadas de segundo grado de la altitud. Considerando p y q tal y como se definen en la ecuación 2.5, junto con r , s y t definidas en la ecuación 2.6, la curvatura del perfil se calcula usando la ecuación 2.7.

$$r = \frac{\partial^2 z}{\partial x^2}; s = \frac{\partial^2 z}{\partial x \partial y}; t = \frac{\partial^2 z}{\partial y^2} \quad (2.6)$$

$$\text{PROF} = \frac{p^2 \cdot r + 2 \cdot p \cdot q \cdot r \cdot s + q^2 \cdot t}{(p^2 + q^2) \cdot \sqrt{(1 + p^2 + q^2)^3}} \quad (2.7)$$

La curvatura del perfil refleja la aceleración o deceleración del flujo de materiales a lo largo de una pendiente, de forma que una curvatura del perfil negativa es cóncava y el flujo sufre una deceleración relativa, mientras que una curvatura del perfil positiva refleja una pendiente convexa y en ella se observa una aceleración relativa. La curvatura del perfil tiene una relación significativa con la humedad del suelo, indicando si una celdilla tiende a acumular agua o no.

5. **Curvatura plana (PLAN).** La curvatura plana refleja la curvatura de la línea de contorno de la superficie, cuya fórmula viene expresada por la ecuación 2.8, y describe la convergencia o divergencia de las líneas de flujo.

$$\text{PLAN} = \frac{q^2 \cdot r - 2 \cdot p \cdot q \cdot s + p^2 \cdot t}{\sqrt{(1 + p^2 + q^2)^3}} \quad (2.8)$$

estando definidas p y q por la ecuación 2.5, y r , s y t por la ecuación 2.6, respectivamente. Una curvatura plana negativa indica concentración de flujo, se trata de celdillas con una curvatura plana cóncava, donde las líneas de flujo convergen, mientras que en aquellas donde la curvatura plana es positiva existe una convexidad y las líneas de flujo divergen. Combinando esta con la curvatura del perfil, puede obtenerse información sobre la erosión potencial que puede generarse a través de ese flujo, que será mayor en celdillas con una curvatura plana cóncava y una curvatura del perfil convexa que en celdillas con otras configuraciones (Figura 2.7).

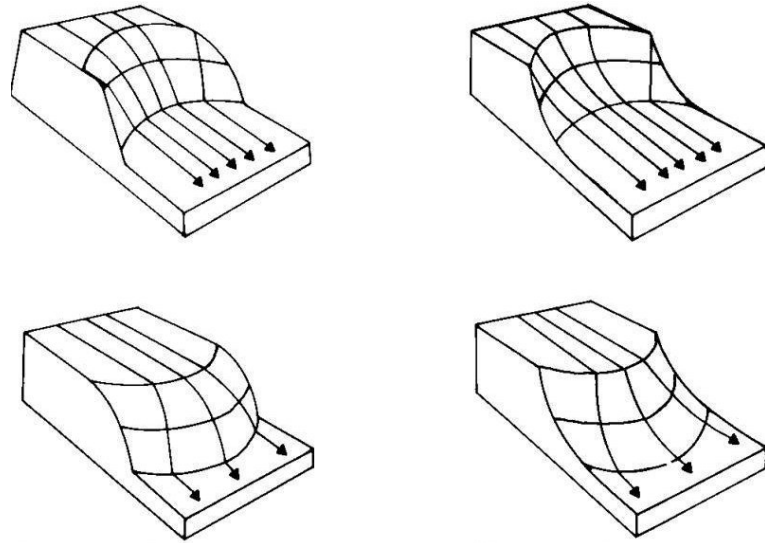


FIGURA 2.7: Distintas configuraciones de la pendiente, las flechas indican la dirección del flujo. Arriba izquierda: curvatura del perfil convexa, curvatura plana cóncava; arriba derecha: curvatura del perfil cóncava, curvatura plana cóncava; abajo izquierda: curvatura del perfil convexa, curvatura plana convexa; abajo derecha: curvatura del perfil cóncava curvatura plana convexa.

VARIABLES RELACIONADAS CON LA RUGOSIDAD.

6. **Índice de Rugosidad del Terreno (TRI)**. Es un índice de la heterogeneidad del terreno definido por [Riley et al. \(1999\)](#), que computa la suma de las diferencias de elevación entre una celda y las que le rodean, tal como se describe en la ecuación 2.9 para una ventana de 3 x 3 celdillas (figura 2.8). El método es análogo para tamaños de ventana superiores.

-1,-1	0,-1	-1,1
-1,0	0,0	1,0
1,1	0,1	1,1

FIGURA 2.8: Denominación de las celdas en una ventana de 3x3

$$TRI = \sqrt{\sum (x_{ij} - x_{00})^2} \quad (2.9)$$

7. **Medida vectorial de la Rugosidad (VRM)**. Descrito por [Sappington et al. \(2007\)](#), su cálculo consta de tres pasos. En primer lugar, para cada celda se obtiene un vector unitario ortogonal a la misma, que se descompone en las tres direcciones x , y , z usando funciones trigonométricas, la pendiente (α) y la orientación (β) de la celda (figura 2.9; ecuación 2.10). Estas tres componentes de las celdillas que componen la ventana se combinan en un vector resultante para la celda central (ecuación 2.11). Restando a uno esta magnitud ($|r|$) estandarizada (obtenida dividiendo entre el número de celdillas de la ven-

tana), según indica la ecuación 2.12 se obtiene una medida de la rugosidad del paisaje a la escala seleccionada, que varía entre 0(plano) y 1 (rugoso).

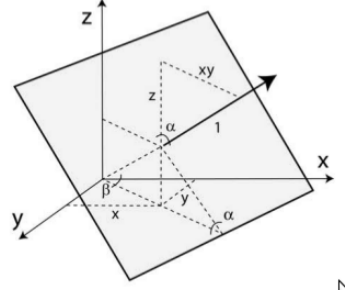


FIGURA 2.9: Obtención de las componentes del vector unitario normal para calcular VRM.

$$\begin{aligned} z &= 1 \cdot \cos(\alpha); xy = 1 \cdot \sin(\alpha) \\ x &= xy \cdot \sin(\beta); y = xy \cdot \cos(\beta) \end{aligned} \quad (2.10)$$

$$|r| = \sqrt{(\Sigma x)^2 + (\Sigma y)^2 + (\Sigma z)^2} \quad (2.11)$$

$$VRM = 1 - \frac{|r|}{n} \quad (2.12)$$

8. **Índice de Iwahasi y Kamiya (IWA).** Iwahashi y Kamiya (1995) calcularon un índice de rugosidad basado en el filtrado del DEM utilizando como operador de filtrado la mediana para extraer los valles y las crestas, lo cual implica un suavizado del mismo. Al sustraer el DEM suavizado por el filtro al DEM original, los valores positivos indican las zonas elevadas, mientras que los valles toman valores negativos, correspondiendo los ceros a las zonas planas.
9. **Índice de Melton (MEL).** Melton (1965) propuso un índice de forma para describir la rugosidad, que en principio estaba destinado a describir cuencas. Este índice se obtiene dividiendo el desnivel z entre la raíz cuadrada del área (m^2) cubierta por la ventana (ecuación 2.13).

$$\Delta z = Z_{max} - Z_{min}$$

$$MEL = \frac{\Delta z}{\sqrt{A}} \quad (2.13)$$

Estas cuatro variables, que describen la rugosidad del terreno, se calcularon utilizando los 10 tamaños de ventana descritos, por lo que en total son 40 capas.

VARIABLES RELACIONADAS CON LA POSICIÓN.

10. **Índice de Posición Topográfica (TPI)**. Este índice, desarrollado por Jenness (2005), es un operador de vecindad que compara la elevación de una celdilla con la elevación media de la ventana. Valores positivos de TPI representan celdillas con altitud mayor al promedio de la ventana en estudio, representando las crestas, mientras que los valores negativos representan celdillas con una altitud menor (valles). Los valores cercanos a cero pueden representar celdillas donde la pendiente es cero o donde la pendiente es constante. El valor del TPI puede ser discretizado en clases, teniendo en cuenta la escala de trabajo y las características del paisaje específico donde se ha calculado.

El TPI es un operador de vecindad y fue obtenido utilizando los 10 tamaños de ventana

ÍNDICES SECUNDARIOS

11. **Topographic Wetness Index (TWI)**. Basándose en la idea de que la topografía controla el movimiento del agua en una pendiente, el TWI indica la distribución espacial de humedad del suelo y la saturación de la superficie del mismo. Este índice forma parte del modelo distribuido TOPMODEL (Beven y Kirkby, 1979; Quinn *et al.*, 1991), utilizado para modelizar procesos hidrológicos relacionados con la topografía a escala de ladera y cuenca. El cálculo de este índice se realiza a través de la fórmula 2.14.

$$TWI = \ln\left(\frac{a}{\tan \beta}\right) \quad (2.14)$$

donde a es la superficie específica de acumulación de flujo, que refleja la tendencia de esa celdilla a acumular agua. El cálculo de a requiere conocer los flujos que se producen en la ladera para determinar el área total de acumulación A que vierte a través de la celdilla en estudio, x_{00} , junto con el longitud de contorno efectiva L , ortogonal al flujo, de modo que $a = A/L$, donde L se pondera multiplicando por el ángulo de la pendiente local $\tan \beta$, que representa el gradiente hidráulico, es decir, la tendencia de las fuerzas gravitacionales de mover el agua hacia abajo de la pendiente. En este trabajo, se ha utilizado el método D8 descrito anteriormente (con una sola dirección de flujo), en el que se asume que el área efectiva de contorno es igual al lado de la celdilla y el ángulo de la pendiente es el mayor en cualquier dirección de la pendiente.

12. **Parámetro LS de la ecuación USLE**. La ecuación USLE (*Universal Soil Loss Equation*), es una ecuación empírica utilizada para fines agrícolas, en la que el factor LS representa la influencia de la pendiente en la pérdida de suelo, que aumenta con la inclinación y la longitud de la misma, tal como se indica en la ecuación 2.15 (Moore y Burch, 1986).

$$LS = \left(\frac{\lambda}{22}\right)^m \left(\frac{0.43 + 0.3S + 0.043S^2}{6.574}\right) \quad (2.15)$$

donde λ es la longitud de la pendiente (m), S es la inclinación de la pendiente (%), y m recibe distintos valores dependiendo de la inclinación: si $S \leq 1$, $m = 0.2$; si $1 < S \leq 3.5$, $m = 0.3$; si $3.5 < S \leq 4.5$, $m = 0.4$ y si $S > 4.5$, $m = 0.5$.

13. **Índice Multirresolución de Fondo de Valle (Multiresolution Index of Valley Bottom Flatness, MRVBF)**. Se trata de un índice que permite describir la morfología del fondo del valle. La diferenciación de las laderas y los fondos de valle es un paso importante para identificar y caracterizar los procesos de erosión y sedimentación. Gallant y Dowling (2003) elaboraron un algoritmo que utiliza el modelo digital de elevaciones para identificar los fondos de valle basándose en la altitud y la pendiente, que opera en un amplio rango de escalas combinando los resultados en un solo índice multirresolución, y clasifica las diferentes gradaciones en la morfología del fondo del valle. De este modo, valores de MRVBF inferiores a 0.5 no son considerados como fondos de valle. Los valores entre 0.5 y 1.5 corresponderían a los valles más pequeños y abruptos, mientras que los valles más amplios y planos estarían representados por valores de 1.5, 2.5, 3.5, y así sucesivamente.

Aunque estos tres índices son dependientes de la escala, por razones prácticas solo se obtuvieron a la escala más detallada, dando lugar cada uno a una capa correspondiente al menor tamaño de ventana (3 x3). Junto con el MDE, se obtienen un total de cuatro capas.

Todas las variables geomorfológicas se calcularon utilizando GRASS (Neteler y Mitasova, 2007) excepto MRVBF, LS, TWI y TPI que se calcularon con SAGA (Olaya y Conrad, 2009).

2.2.2.2 Variables climáticas

Las series mensuales de precipitación y temperatura permiten caracterizar climáticamente una determinada región. Las variables utilizadas en este trabajo (Gomariz Castillo y Alonso Sarría, 2013) se han obtenido utilizando series temporales de 625 estaciones de la Cuenca del Segura, analizando datos que abarcan un periodo de 30 años (1975-2002). En primer lugar, se realizó una homogeneización de las series temporales utilizando el paquete **climatol**. A continuación, se aplicaron modelos lineales generalizados utilizando como variables predictoras la localización espacial, la altitud, la pendiente, la curvatura del perfil y tangencial, las componentes N-S y E-W de la pendiente, así como las segundas derivadas parciales, la distancia al mar y la radiación global. El mejor modelo se seleccionó a través de algoritmos de búsqueda y el Criterio de Información de Akaike (AIC) que proporcionan los paquetes **bestglm**, **lmtest** e **ipred**. Finalmente, los residuos del GLM se interpolaron usando diferentes métodos (interpolación ponderada con el inverso de la distancia, krigeado ordinario y splines) mediante los paquetes **MBA**, **gstat**, **maps** y **automap**, entre los cuales se seleccionó el mejor utilizando un sistema de validación cruzada dejando-uno-fuera (LOO-CV).

De esta forma, se obtuvieron mapas ráster de cinco variables climáticas, a escala mensual, además de la radiación mensual, que resultan en un total de 72 variables climáticas:

14. **Precipitación mensual (prectot)**.
15. **Temperatura media mensual (temed)**.
16. **Temperatura mínima mensual (teminabs)**.
17. **Temperatura máxima mensual (temaxabs)**.

18. Humedad mensual (eTH).

19. **Radiación mensual (RAD).** Se ha calculado la radiación mensual con la función **r.sun** de GRASS ([GRASS Development Team, 2013](#)), utilizando como variables de entrada el Modelo Digital de Elevaciones, la pendiente (SLOPE), el mapa de orientaciones (ASPECT), y el día del año (1-365). Esta función permite obtener la radiación directa, difusa y reflejada, cuya suma resulta en la radiación global, todas ellas calculadas considerando el cielo despejado. No obstante, el programa ofrece también la opción de incluir como variable de entrada la proporción de radiación disminuida por factores ambientales como las nubes. El índice de turbidez atmosférica de Linke, es considerado por defecto igual a 3 (equivalente al valor medio en zonas rurales y urbanas, considerando que el valor para una atmósfera absolutamente limpia es de 1). La fórmula incluye la posibilidad de incorporar al modelo el efecto sombra producido por las elevaciones circundantes. Las unidades en que se expresa la radiación son $\text{Wh} \cdot \text{m}^2 \cdot \text{day}^{-1}$. Se han calculado 12 capas, correspondientes a cada uno de los meses del año.

2.2.3 Variables cualitativas

20. **Mapa de usos del suelo (CORINE Land Cover).** Se ha utilizado el mapa de usos del suelo del año 1990, descargado de la página web del CNIG, simplificando los usos a 8 clases para hacerlos coincidir con los descritos en la base de datos, tal como se indica en la tabla 2.2. La figura 2.10 muestra la distribución de los usos del suelo en la región.

Uso	Código LUCDEME	CORINE Land Cover 1990
Urbano y superficies artificiales	1	111-142
Cultivo de secano	2	211
Regadío	3	212, 213
Cultivos arbóreos	4	221,222,223
Cultivos en mosaico	5	241,242,243
Bosque	6	311, 312, 313
Matorral	7	321,323,324,333,411,421,422
Suelo desnudo	8	331,332

TABLA 2.2: Reclasificación de usos CORINE Land Cover 1990

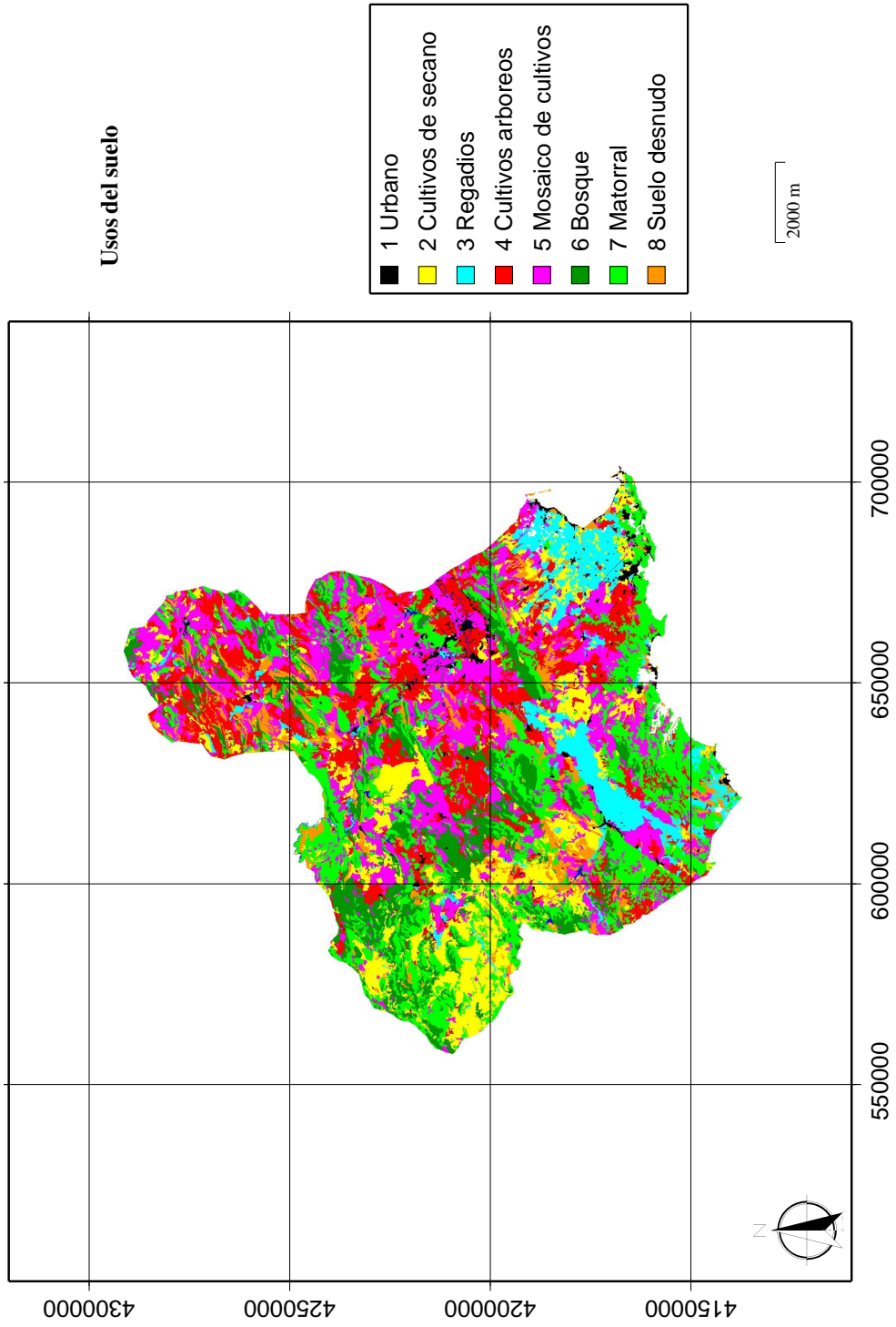


FIGURA 2.10: Mapa de usos del suelo de CORINE Land Cover 1990 reclasificado

21. **Mapa de suelos de la región de Murcia.** El mapa de suelos utilizado en este trabajo es el elaborado por [Ramírez-Santigosa *et al.* \(1999\)](#) a partir de la digitalización de los mapas de suelos de [LUCDEME \(1986-2004\)](#), cuya clasificación se ha revisado para adaptarla a la vigente en la actualidad ([WRB, 2007](#)), incluyendo las asociaciones de suelos pero no las inclusiones (que representan menos del 20% de la superficie de la unidad correspondiente). Con el fin de eliminar los grupos poco representativos en relación al número de muestras, que podrían generar problemas en la obtención y validación en bloques, se han reagrupado los tipos con menos de 7 muestras en otros tipos de suelos considerados similares, o se han eliminado otros cuando no podían ser reagrupados (etiqueta "null" en la tabla [2.3](#)). La distribución de los suelos de la región se muestra en la figura [2.11](#).

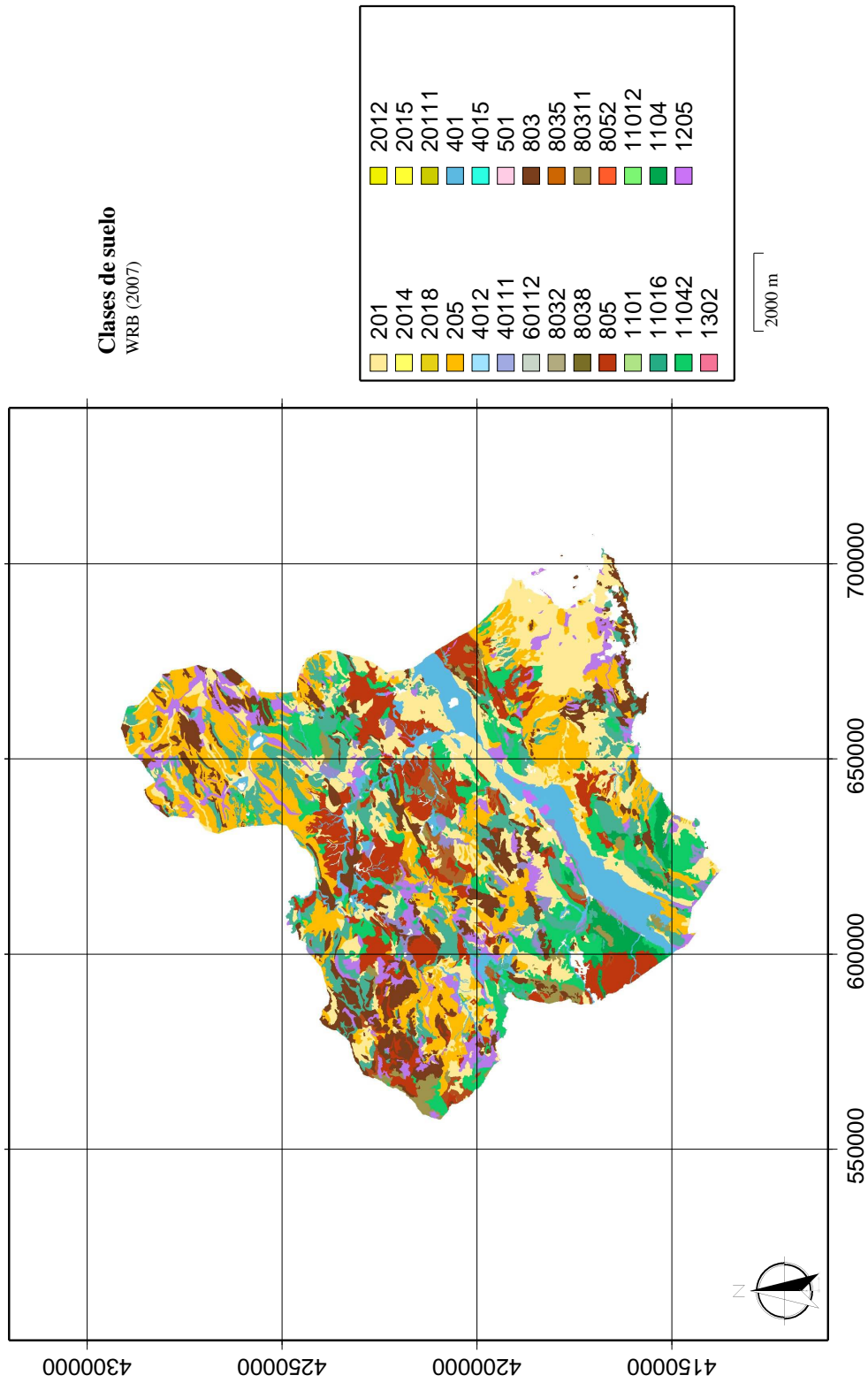


FIGURA 2.11: Mapa de suelos de LUCDEME (1986-2004) reclasificados según WRB (2007). Los números de la leyenda se refieren a las clases de suelo descritas en la tabla 2.3.

Mapa de suelos LUCDEME		Mapa reclasificado	
código	FAO-UNESCO (1974)	código	WRB (2007)
1	Arenosoles álbicos, Solonchaks gléicos	null	
2	Cambisoles cálcicos	201	Calcisoles Hápticos
3	Cambisoles cálcicos, Cambisoles petrocálcicos	2012	Calcisoles Hápticos Calcisoles Pétricos
4	Cambisoles cálcicos, Fluvisoles calcáricos	2014	Calcisoles Hápticos, Fluvisoles Hápticos
5	Cambisoles cálcicos, Regosoles calcáricos	20111	Calcisoles Hápticos, Regosoles Hápticos
6	Cambisoles cálcicos, Rendsinas órticas	2018	Calcisoles Hápticos, Leptosoles Réndzicos
7	Cambisoles eútricos, Regosoles eútricos	null	
8	Cambisoles petrocálcicos	205	Calcisoles Pétricos
9	Cambisoles petrocálcicos, Rendsinas órticas	205	Calcisoles Pétricos
10	Fluvisoles calcáricos	401	Fluvisoles Hápticos
11	Fluvisoles calcáricos, Regosoles calcáricos	40111	Fluvisoles Hápticos, Regosoles Hápticos
12	Fluvisoles calcáricos, Solonchaks órticos, Solonchaks gléicos	null	
13	Fluvisoles calcáricos, Xerosoles cálcicos	4012	Fluvisoles Hápticos, Calcisoles Hápticos
14	Fluvisoles calcáricos, Xerosoles gípsicos	4015	Fluvisoles Hápticos, Gipsisoles Hápticos
15	Fluvisoles calcáricos, Xerosoles petrocálcicos	4012	Fluvisoles Hápticos, Calcisoles Pétricos
16	Gleysoles calcáricos, Solonchaks gléicos	60112	Gleysoles Hápticos, Solonchaks Gléicos
17	Kastanosems cálcicos	null	
18	Litosoles	803	Leptosoles Líticos
19	Litosoles, Cambisoles cálcicos	8032	Leptosoles Líticos, Calcisoles Hápticos
20	Litosoles, Kastanosems cálcicos	803	Leptosoles Líticos
21	Litosoles, Phaeosems Hápticos	803	Leptosoles Líticos
22	Litosoles, Regosoles calcáricos	80311	Leptosoles Líticos, Regosoles Hápticos
23	Litosoles, Regosoles calcáricos, Regosoles litosólicos	80311	Leptosoles Líticos, Regosoles Hápticos
24	Litosoles, Regosoles litosólicos	80311	Leptosoles Líticos, Regosoles Lépticos
25	Litosoles, Rendsinas arídicas	8038	Leptosoles Líticos, Leptosoles Réndzicos
26	Litosoles, Rendsinas órticas	8038	Leptosoles Líticos, Leptosoles Réndzicos
27	Litosoles, Solonchaks gléicos	null	
28	Litosoles, Solonchaks litosólicos	null	
29	Litosoles, Xerosoles cálcicos	8032	Leptosoles Líticos, Calcisoles Hápticos

TABLA 2.3: (continúa)

Mapa de suelos LUCDEME		Mapa reclasificado	
código	FAO-UNESCO (1974)	código	WRB (2007)
30	Litosoles, Xerosoles gípsicos	8035	Leptosoles líticos, Gipsisoles Hápticos
31	Litosoles, Xerosoles petrocálcicos	8032	Leptosoles Líticos, Calcisoles Pétricos
32	Regosoles calcáricos	1101	Regosoles Hápticos
33	Regosoles calcáricos, Cambisoles cálcicos	11012	Regosoles Hápticos, Calcisoles Hápticos
34	Regosoles calcáricos, Regosoles eútricos	1101	Regosoles Hápticos
35	Regosoles calcáricos, Regosoles litosólicos	1101	Regosoles Hápticos
36	Regosoles calcáricos, Regosoles litosólicos, Regosoles eútricos	1101	Regosoles Hápticos
37	Regosoles calcáricos, Xerosoles cálcicos	11012	Regosoles Hápticos, Calcisoles Hápticos
38	Regosoles calcáricos, Xerosoles gípsicos	11016	Regosoles Hápticos, Gipsisoles Hápticos
39	Regosoles calcáricos, Xerosoles petrocálcicos	11012	Regosoles Hápticos, Calcisoles Pétricos
40	Regosoles eútricos	1101	Regosoles Hápticos
41	Regosoles eútricos, Regosoles litosólicos	1101	Regosoles Hápticos
42	Regosoles litorródicos	1104	Regosoles Lépticos
43	Regosoles litosólicos	1104	Regosoles Lépticos
44	Regosoles litosólicos, Regosoles calcáricos	1104	Regosoles Lépticos
45	Regosoles litosólicos, Xerosoles cálcicos	11042	Regosoles Lépticos, Calcisoles Hápticos
46	Rendsinas arídicas	805	Leptosoles Réndzicos
47	Rendsinas arídicas, Xerosoles cálcicos	8052	Leptosoles Réndzicos, Calcisoles Hápticos
48	Rendsinas órnicas	805	Leptosoles Réndzicos
49	Solonchaks gléicos	1205	Solonchaks Hápticos
50	Solonchaks gléicos, Solonchaks órnicos	1205	Solonchaks Hápticos
51	Solonchaks órnicos	1205	Solonchaks Hápticos
52	Vertisoles crómicos	1302	Vertisoles Hápticos
53	Xerosoles cálcicos	201	Calcisoles Hápticos
54	Xerosoles cálcicos, Fluvisoles calcáricos	2014	Calcisoles Hápticos, Fluvisoles Hápticos
55	Xerosoles cálcicos, Rendsinas arídicas	8032	Leptosoles Réndzicos, Calcisoles Hápticos
56	Xerosoles cálcicos, Xerosoles gípsicos	2015	Calcisoles Hápticos, Gipsisoles Hápticos
57	Xerosoles cálcicos, Xerosoles lúvicos	201	Calcisoles Hápticos
58	Xerosoles cálcicos, Xerosoles petrocálcicos	2012	Calcisoles Hápticos, Calcisoles Pétricos

TABLA 2.3: (continúa)

Mapa de suelos LUCDEME		Mapa reclasificado	
código	FAO-UNESCO (1974)	código	WRB (2007)
59	Xerosoles gípsicos	501	Gipsisoles Háplicos
60	Xerosoles lúvicos	mull	
61	Xerosoles petrocálcicos	205	Calcisoles Pétricos
62	Miscelánea	99	Miscelánea
63	Embalses	99	Miscelánea
64	Zona militar	99	Miscelánea
64	Zona Minera	99	Miscelánea
64	Zona urbana	99	Miscelánea

TABLA 2.3: Reclasificación de unidades de suelos

2.3 Software

2.3.1 Sistemas de información geográfica: GRASS GIS y SAGA GIS

GRASS (“Geographic Resources Analysis Support System”) es un Sistema de Información Geográfica (SIG) usado para gestión de datos, procesamiento de imágenes, producción de gráficos, modelización espacial y visualización de datos espaciales (Neteler y Mitasova, 2007). GRASS permite trabajar en conexión con el paquete estadístico R (Neteler y Mitasova, 2007) y con PostgreSQL (gestión de base de datos). En este trabajo, se ha utilizado para la visualización de datos espaciales, el cálculo de variables geomorfológicas y climáticas, y la interpolación de los residuos de los modelos cuando la estructura espacial no era lo suficientemente fuerte.

SAGA (System for Automated Geoscientific Analyses) GIS (Olaya y Conrad, 2009) es un sistema de información geográfica que soporta datos raster y vectoriales, y que se ha utilizado para análisis geográfico. Dispone de diversos módulos de análisis. En esta tesis se han utilizado los de morfometría e hidrología (`ta_morphometry` y `ta_hidrology`).

2.3.2 Software estadístico: R

R fue el software utilizado para la selección de variables, la elaboración y selección de los modelos, la validación cruzada, y la elaboración de los gráficos mostrados en la presente tesis. Para cargar las distintas capas raster como objetos de R, se ha utilizado la librería `spGRASS6` (Bivand, 2013). Los modelos han sido calculados utilizando diversos paquetes:

- `bestglm` (McLeod y Xu, 2011). Para Modelos Lineales Generalizados (GLM).
- `mgcv` (Wood, 2006). Para Modelos Aditivos Generalizados (GAM).
- `earth` (Milborrow, 2011). Para Splines de Regresión Adaptativa Multivariante (MARS).
- `randomForest` (Liaw y Wiener, 2002) Para Random Forest (RF).
- `kernlab` (Karatzoglou *et al.*, 2004). Para Máquinas de Vectores Soporte (SVM).

Todos estos modelos aparecen descritos en Hastie *et al.* (2009).

Para la interpolación de los residuos, se han utilizado los paquetes de R `automap` (Hiemstra *et al.*, 2008), con el fin de obtener el semivariograma experimental y ajustarlo a los diversos modelos, así como el Krigeado Ordinario.

2.4 Técnicas estadísticas

2.4.1 Transformación de variables

Debido a que algunas de las variables predictoras no se ajustaban a una distribución normal, estas se transformaron tal como se indica en las ecuaciones 2.16 a 2.21.

$$\text{SLO25_39}_t = \log(0.5 + \text{SLO25_39}) \quad (2.16)$$

$$\text{VRM11}_t = \log(0.1 + \text{VRM11}) \quad (2.17)$$

$$\text{VRM39}_t = \log(0.1 + \text{VRM39}) \quad (2.18)$$

$$\text{IWA25_27}_t = \log(0.01 + \text{IWA25_27}) \quad (2.19)$$

$$\text{IWA25_39}_t = \log(0.01 + \text{IWA25_39}) \quad (2.20)$$

$$\text{AGprectotPREDFIN4}_t = \log(\text{AGprectotPREDFIN4}) \quad (2.21)$$

Además, se eliminaron los valores de TPI25_3 superiores a 20.

A continuación, todas las variables cuantitativas fueron estandarizadas, utilizando la ecuación 2.22, que resta a la valor de la variable X en cada punto el valor medio de la misma (\hat{X}), dividiendo el resultado entre la desviación estándar (σ_X).

$$X_{norm} = \frac{X - \hat{X}}{\sigma_X} \quad (2.22)$$

2.4.2 Selección de variables

De las 181 variables disponibles, resumidas en la tabla 2.4, muchas de ellas presentaban un alto grado de colinealidad. El problema de tener variables colineales en el modelo es de mayor importancia cuando las señales ecológicas son débiles, y su eliminación puede tener mayor efecto en los p-valores que la eliminación de las variables no significativas. En este trabajo no se ha utilizado la aproximación habitual del análisis de componentes principales, ya que presenta el problema de que resulta difícil entender el significado de los componentes. Por ello, se ha utilizado el método del Factor de Inflación de la Varianza (Zuur *et al.*, 2010). Este método calcula para cada variable un estadístico, el Factor de Inflación de la Varianza (VIF; ecuación 2.23) que resume el vector correspondiente de la matriz de correlaciones, donde R_j^2 es el R^2 de un modelo de regresión lineal en el cual la variable X_j se utiliza como variable respuesta, y el resto como variables explicativas. De este modo cuanto mayor sea el valor del estadístico mayor será el grado de colinealidad de la variable.

$$\text{VIF} = \frac{1}{1 - R_j^2} \quad (2.23)$$

Grupo	Variable	Número de capas
Altitud	MDE25	1
Derivadas de la altitud	SLO	10
	COS25	10
	SIN25	10
	PROF25	10
	PLAN25	10
Rugosidad	TRI25	10
	VRM25	10
	IWA25	10
	MEL25	10
Derivadas de la posición	TPI25	10
Índices secundarios	TWI	1
	LS	1
	MRVBF	1
Variables climáticas	AGprectotPREDFIN	12
	AGtemaxabsPREDFIN	12
	AGteminabsPREDFIN	12
	AGtemedPREDFIN	12
	AGeTHPREDFIN	12
	RAD25	12
Total		181

TABLA 2.4: Resumen de las variables predictoras utilizadas

La aproximación habitual consiste en determinar un valor umbral por encima del cual se considera que la variable es excesivamente colineal con las demás y se elimina del análisis. [Montgomery y Peck \(1992\)](#) recomendaba un valor umbral de 10, aunque [Zuur *et al.* \(2010\)](#) utilizan un valor umbral más estricto, de 5. Para evitar la eliminación de variables correlacionadas entre sí pero no con las demás, que podría ser conveniente conservar, se ha desarrollado un procedimiento iterativo que elimina la variable con mayor VIF, recalcula los estadísticos y vuelve a eliminar la variable con mayor VIF. Este procedimiento se repite hasta que ningún VIF supera el valor umbral de 5 considerándose de este modo que las variables restantes son no colineales.

2.4.3 Evaluación y validación de los modelos

Con el fin de obtener una medida más fiable de la precisión de los modelos, es necesario calcular estos parámetros a partir de datos de validación, no utilizados en la calibración del modelo. Por esta razón, las 1436 muestras de capa arable disponibles se dividieron de forma aleatoria en cinco bloques para cada una de las variables estudiadas. Los cuatro primeros bloques, que suman en total 1150 muestras, se utilizaron como conjunto de entrenamiento para calibrar los distintos modelos mediante validación cruzada. El quinto bloque, de 284 muestras de capa arable, se utilizó como conjunto de validación.

A la hora de seleccionar los modelos, en primer lugar se realizó un procedimiento de validación cruzada por bloques para seleccionar el número de variables en el modelo minimizando el valor de RMSE en el set de calibración ($RECM_c$). Dado que los modelos más complejos presentan menor sesgo, pero una mayor varianza (figura 2.12), se hace necesario realizar un balance entre ambos parámetros, evitando del mismo modo un sobreajuste de los modelos a los datos.

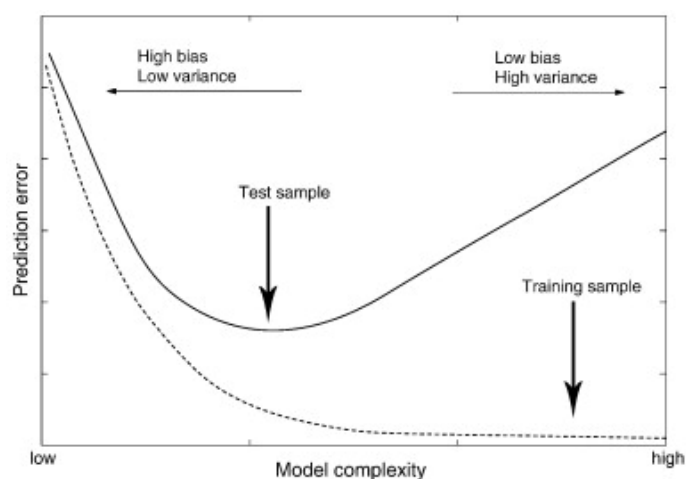


FIGURA 2.12: Equilibrio entre sesgo (bias) y varianza (variance) en la selección del modelo. Fuente: [James et al. \(2013\)](#).

La validación cruzada en bloques consiste en la división del set completo de datos (n muestras) en K bloques de la misma longitud. En la primera iteración, se utilizan las muestras de $(K-1)$ bloques como datos de entrenamiento para calibrar el modelo, y las muestras del bloque restante se utilizan como datos de validación del mismo; este proceso se reitera tantas veces como bloques (K) se han establecido, de manera que en cada iteración, uno de los bloques sirve como datos de validación, y el resto como datos de entrenamiento. Este método es similar a la validación cruzada dejando un valor fuera (Leave One Out Cross Validation, **LOOCV**), que en realidad equivaldría a una validación cruzada en bloques donde $K = n$, pero con dos ventajas frente a ésta:

- (i) por un lado, la ventaja de que su implementación requiere menor esfuerzo computacional, sobre todo en el caso de conjuntos de muestras grandes, y con métodos estadísticos complejos.
- (ii) Por otro lado, cuando se utiliza LOOCV, los conjuntos de datos de entrenamiento son casi idénticos cada vez, y por tanto, estamos calculando la media de resultados que están altamente correlacionados entre sí, obteniendo modelos con un gran ajuste y un bajo sesgo, pero a costa de una gran varianza, por lo que el modelo es altamente sensible a los cambios en los datos de entrada utilizados; sin embargo, usando la validación cruzada en bloques, los conjuntos de datos de entrenamiento presentan menor superposición, y la media de los resultados esperados presenta mayor sesgo que al utilizar LOOCV, a cambio de obtener una menor varianza, es decir, menores fluctuaciones en el resultado al introducir pequeñas variaciones en los datos de entrada.

Por lo general, se recomienda utilizar validación cruzada en bloques con $K=5$ o $K=10$, valores que, según se ha comprobado de forma empírica, no presentan un sesgo ni una varianza excesivamente altos ([Hastie et al., 2009](#)). En este trabajo se han utilizado 5 bloques, dividiendo el

conjunto de datos en función de los tipos de suelo, de modo que los diferentes grupos quedaran representados en todos los bloques de forma aproximadamente equitativa, con al menos una muestra de cada tipo de suelo en cada uno de ellos, con el fin de que no se generaran errores en la calibración y validación.

Tras este procedimiento, se calibró cada uno de los modelos con las muestras de los cuatro primeros bloques, utilizando tantas variables como indicara el paso anterior. El modelo obtenido se validó con las 284 muestras del quinto bloque, obteniéndose además el valor de la correlación y el $RECM_v$.

Finalmente, los modelos predictivos utilizados fueron:

- **Modelos Lineales Generalizados (GLM)**. Se obtuvieron dos modelos, el primero de ellos (denominado **GLM**) introduciendo todas las variables y realizando un proceso de selección por pasos, y el segundo (denominado **GLM-CV**) introduciendo el número de variables determinado con el procedimiento de validación cruzada por bloques con el cual se minimizaba el RECM. El número máximo de variables a introducir en el modelo fue establecido en 30.
- **Modelos Aditivos Generalizados (GAM)**. Se obtuvieron dos modelos. Debido al coste (en términos de tiempo) que requería el proceso de selección por pasos, el primero (**GAM**) fue el modelo saturado (con las 40 variables). El segundo modelo (**GAM-CV**) introdujo solamente el número de variables determinado por el procedimiento de validación cruzada en bloques, para obtener el mínimo valor de RECM, con un máximo de 30 variables. Dado que los modelos GAM requieren un elevado número de muestras para no generar errores debido al elevado número de coeficientes, en el caso de los modelos a escala local, donde la cantidad de muestras era menor, sólo se incluyeron en el primer GAM las variables que resultaban significativas en el modelo GLM, y un máximo de 15 variables en el caso de GAM-CV.
- **Splines de Regresión Adaptativa Multivariante (MARS)**. En primer lugar, se separaron las variables cualitativas (tipo y uso de suelo) para elaborar un modelo GLM con ellas; el valor obtenido en la predicción del GLM se restó del valor original de la variable, denominándose "residuo" a la diferencia. A continuación, se elaboraron dos modelos MARS en los que la variable dependiente era este residuo. En el primero de ellos (denominado **MARS**) se introdujeron todas las variables cuantitativas disponibles, de modo que el proceso de selección por pasos sucesivos "hacia atrás" (backward) comentado en la sección 1.1.2, eliminara aquellas que menos contribuían a la eficacia del modelo. El segundo de los modelos (**MARS-CV**) incluía el número de variables cuantitativas determinado previamente por el proceso de validación cruzada.
- **Random Forest (RF)**. Como parámetros de entrada para la construcción de cada árbol, el número de variables usadas en cada división (`mtry`) se fijó en valor por defecto, que para la regresión viene establecido como $p/3$, donde p es el número de variables totales usadas en el modelo, y se obtuvieron un total de 500 árboles (`mtree=500`). Como se indicó en la sección 1.1.2, la importancia de las variables en los árboles de regresión se ha expresado a

través de la impureza del nodo. Se realizaron tres modelos con Random Forest: el primero de ellos (denominado **RF**) utilizaba todas las variables disponibles (40); en el segundo de ellos (**RF-CV**) se incluyeron las variables resultantes del procedimiento de validación cruzada en bloques; para el tercero (**RF-CV_{IMP}**), se ordenaron en primer lugar las variables del modelo RF en orden decreciente de importancia, incluyéndose a continuación en el modelo las determinadas por el procedimiento de validación cruzada por bloques.

- **Máquinas de Vectores Soporte (SVM)**. Para la elaboración de estos modelos, las variables cualitativas fueron convertidas previamente en factores. La función núcleo seleccionada fue la gaussiana de base radial (RBF Gaussian). En el primero de los modelos obtenidos con este método (**SVM**) se introdujeron todas las variables, mientras que en el segundo (**SVM-CV**) se incluyó el número de variables resultante del proceso de validación cruzada por bloques.

Los parámetros utilizados para seleccionar y validar los modelos obtenidos para las diferentes variables fueron el coeficiente de correlación de Pearson (**cor**, ecuación 2.24) y la Raíz del Error Cuadrático Medio (**RECM**, ecuación 2.25, teniendo ambos en cuenta como medida del ajuste y capacidad de predicción de los mismos. El coeficiente de correlación de Pearson es la covarianza entre los valores reales (y) y los valores predichos por el modelo (\hat{y}), dividida por el producto de sus desviaciones estándar.

$$\rho_{y\hat{y}} = \frac{\text{cov}(y\hat{y})}{\sigma_y\sigma_{\hat{y}}} \quad (2.24)$$

La raíz del error cuadrático medio (RECM) representa la desviación estándar de las diferencias entre los valores predichos por el modelo (\hat{y}) y los valores reales (y).

$$\text{RECM} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (2.25)$$

Sin embargo, a la hora de evaluar la capacidad predictiva de los modelos, se ha utilizado preferentemente la disminución del RECM como criterio, ya que la correlación (coeficiente de Pearson) es una medida lineal que puede no ser adecuada cuando las relaciones no son lineales.

Hay que tener en cuenta, sin embargo, que tanto la correlación como el RECM son valores puntuales. Con el fin de obtener una distribución de dicho estadístico, los datos del conjunto de validación se remuestrearon 1000 veces con reemplazo mediante bootstrapping (James *et al.*, 2013) y en cada muestra se calcularon los estadísticos descriptivos de la correlación y la raíz cuadrada del error cuadrático medio (RMSE) para cada uno de los 11 modelos utilizados.

2.4.4 Interpolación de los residuos

En este trabajo, el residuo fue calculado restando al valor estimado por el modelo el valor observado en cada muestra. Tras comprobar la normalidad de los residuos, se representaron estos gráficamente en función de sus coordenadas, en diferentes colores según el valor de los mismos, con el fin de detectar patrones en la distribución espacial.

A continuación, se elaboró el semivariograma experimental y se ajustó a un semivariograma teórico con la función `autofitVariogram` del paquete `automap` (Hiemstra *et al.*, 2008). Esta función calcula el semivariograma experimental y a continuación ajusta, basándose en los datos proporcionados, el semivariograma teórico a partir de ciertas premisas de entrada: la meseta inicial se estima como la media entre el máximo y la mediana de la semivarianza, y la pepita inicial se define como el mínimo de la semivarianza. A la hora de seleccionar el modelo, realiza iteraciones con todos los tipos, escogiendo aquel en el que la suma de los cuadrados es menor respecto al semivariograma experimental. De este semivariograma se extrajeron los valores de pepita (nugget), meseta y rango, y la relación entre los dos primeros (NSR; Nugget to Sill Ratio) utilizando la ecuación 2.26. Según Cambardella *et al.* (1994), el valor de NSR permite clasificar la estructura espacial en fuerte (NSR < 25%), moderada (25% < NSR < 75%) o débil (NSR > 75%).

$$\text{NSR} = \frac{\text{Nugget}}{\text{Sill total}} \cdot 100 \quad (2.26)$$

En los casos en que los residuos mostraban una estructura espacial fuerte o moderada, se realizó una interpolación de los mismos utilizando krigeado ordinario, utilizando la función `autoKrige` (paquete `automap`, Hiemstra *et al.*, 2008), mientras que cuando la estructura espacial era débil, se utilizó una interpolación utilizando la función `v.surf.rst` de GRASS (Neteler y Mitasova, 2007).

La figura 2.13 muestra el procedimiento llevado a cabo para la cartografía de cada una de las variables físico-químicas a escala regional. Este procedimiento se resume en los puntos siguientes:

1. **Análisis descriptivo.** En primer lugar, se representaron tanto el histograma como el gráfico de cuantiles, con el fin de asegurar la normalidad de la variable, que fue también comprobada con el test de Shapiro-Wilk. Por otro lado, la homogeneidad de varianzas fue comprobada con el test de Bartlett. Cuando se cumplían ambas condiciones (normalidad y homocedasticidad), se realizó una transformación de la variable. Posteriormente se realizó, sobre la variable original o ya transformada, el test de comparación de varianzas ANOVA para buscar diferencias entre los usos y los tipos de suelo, seguido por el test de Tukey para localizar esas diferencias; si a pesar de la transformación no se cumplían las condiciones necesarias, se utilizaron tests no paramétricos con este mismo fin (test de Kruskal-Wallis y test de Wilcoxon entre pares de grupos). Finalmente, se comprobaron las correlaciones existentes entre la variable estudiada y las variables predictoras geomorfológicas y climáticas.
2. **Elaboración de modelos.** Inicialmente, se dividió el conjunto de datos en cinco subconjuntos, para realizar la validación cruzada en bloques. En este paso, se procuró que en todos los subconjuntos estuvieran representados cada uno de los tipos de suelos presentes. A continuación, se elaboraron los 11 modelos descritos (subsección 2.4.3). Todos los modelos se evaluaron en función de la correlación y el RECM obtenidos en un remuestreo del conjunto de validación, prestando también atención al número de variables incluido en cada modelo. Tras seleccionar el modelo que proporcionaba mejor resultado, se utilizó

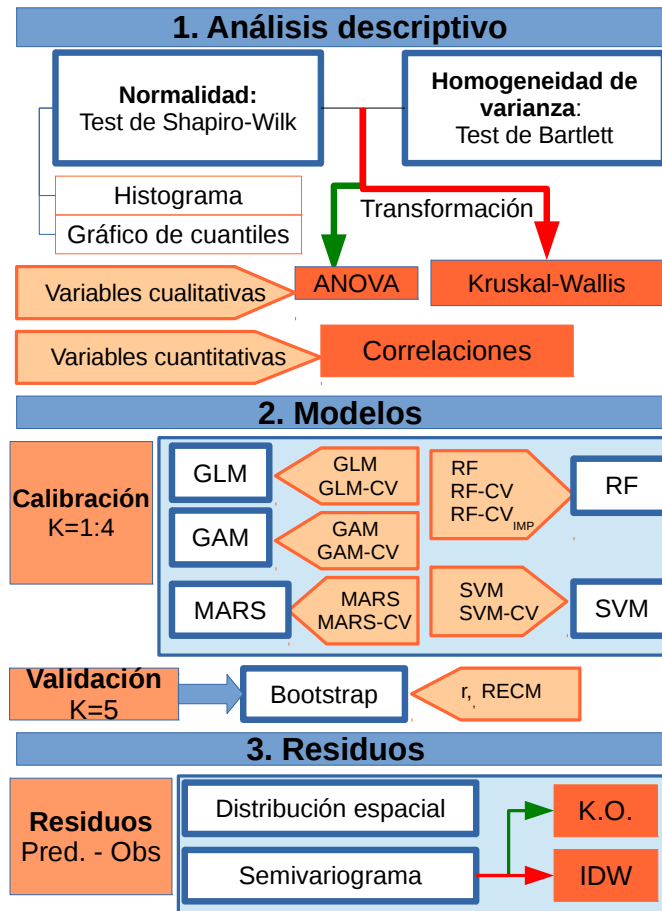


FIGURA 2.13: Esquema de flujo del trabajo realizado para obtener cada uno de los mapas de las variables físico-químicas a nivel regional.

para obtener el mapa de la predicción introduciendo como variables predictoras los mapas raster.

- 3. Interpolación de residuos.** Tras asegurar la normalidad de los residuos del modelo, se representaron los puntos utilizando una clave de colores para detectar patrones espaciales en sus valores. Además, se ha elaborado el semivariograma, ajustándolo a uno de los modelos teóricos existentes, con la función `autofitVariogram` del paquete **automap**. En función de si la estructura espacial, calculada a partir de los parámetros del semivariograma (subsección 2.4.4), se obtuvo el mapa de residuos interpolando con Krigeado Ordinario (KO; función `krige`, paquete **automap**) o con splines (función `v.surf.rst` de GRASS). Finalmente, se han sumado ambos mapas para obtener el mapa de cada variable.

En el caso de los modelos a escala local, solo se han llevado a cabo los dos primeros pasos, pero no se han elaborado los mapas ni se han interpolado los residuos. En su lugar, se ha comparado

el ajuste de los modelos a escala regional y escala local a los datos existentes en la sierra de Mojantes.

2.4.5 Funciones de edafotransferencia

A la hora de estimar las variables relacionadas con las propiedades hídras (densidad aparente, capacidad de campo y punto de marchitamiento permanente) se han utilizado funciones de edafotransferencia (PTFs).

En primer lugar, se realizó un análisis descriptivo de cada una de las variables en la base de datos de la sierra de las Moreras, que incluyó el cálculo de los estadísticos descriptivos, y la realización de pruebas de normalidad (test de Shapiro-Wilk, histograma y gráfico de cuantiles). Además, se examinó si las variables hídras estaban correlacionadas con propiedades del suelo como el contenido de carbono orgánico, el contenido de carbonato cálcico y la distribución de tamaños de partículas.

A continuación, se realizó una revisión bibliográfica para recopilar y seleccionar diferentes funciones de edafotransferencia (PTF). Estas funciones fueron validadas en la base de datos de la sierra de las Moreras. Los parámetros utilizados para la evaluación de las PTFs son, como en el caso de los modelos, la correlación (r) y la Raíz del Error Cuadrático Medio (RECM). Además, se ha calculado el Error Medio de la Predicción (EMP, ecuación 2.27). El MPE permite la detección del sesgo negativo o positivo de cada PTFs, indicando la tendencia a la infraestimación o la sobreestimación, respectivamente (de Vos *et al.*, 2005; Jalabert *et al.*, 2010; Sequeira *et al.*, 2014).

$$\text{EMP} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \quad (2.27)$$

Posteriormente, se utilizó la base de datos de la sierra de las Moreras para obtener nuevas PTFs, cuyo comportamiento se comparó con las obtenidas en la revisión bibliográfica, utilizando GLMs en un procedimiento de validación cruzada dejando una muestra fuera (LOOCV). El comportamiento de estas funciones se evaluó con los mismos parámetros que las obtenidas en la revisión. En el caso del contenido de humedad, las funciones de edafotransferencia elaboradas se validaron también utilizando las muestras de perfiles de LUCDEME, con el fin de evaluar su aplicabilidad en muestras diferentes a las utilizadas en la calibración.

Finalmente, se elaboraron dos mapas para cada variable hídrica: uno de ellos, utilizando la función de edafotransferencia recopilada en la bibliografía que mejor resultado proporcionaba, y un segundo mapa con la función de edafotransferencia elaborada en este trabajo. Ambos mapas utilizan la función `r.mapcalc` de GRASS, computando a partir de las variables físico-químicas modelizadas en pasos anteriores de esta tesis. El desempeño de cada uno de los mapas también fue evaluado con los parámetros descritos, frente los datos de la sierra de las Moreras y los de los perfiles de LUCDEME (en el caso del contenido de humedad).

En resumen, el flujo de trabajo realizado con las funciones de edafotransferencia se muestra en la figura 2.14.

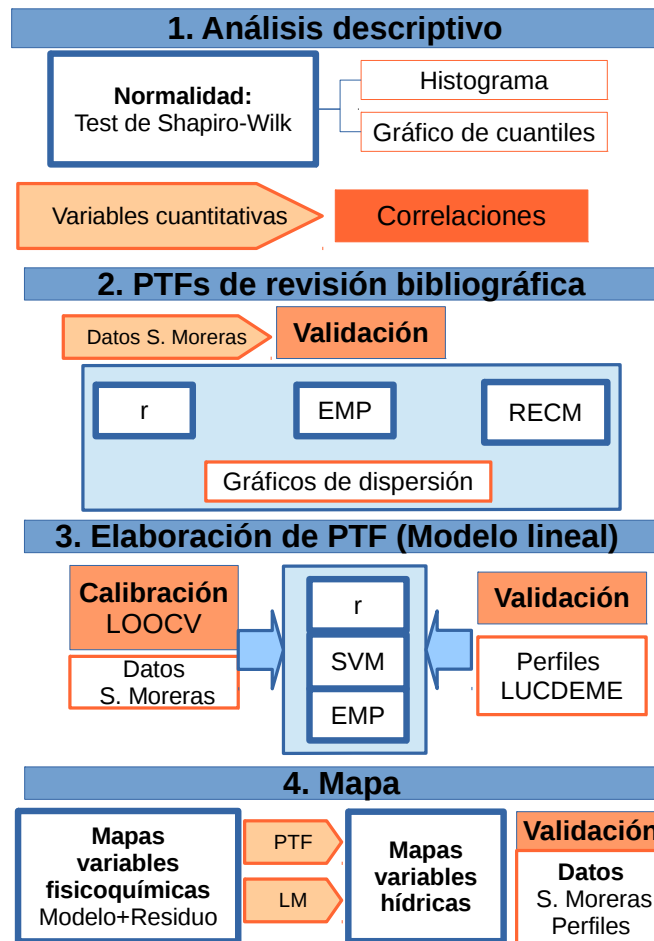


FIGURA 2.14: Esquema de flujo del trabajo realizado con las funciones de edafotransferencia para cada una de las variables hídricas analizadas.

Capítulo 3

Resultados y discusión

3.1 Selección de variables

El conjunto de variables resultante del análisis VIF es un subconjunto de las variables originales que resume la información a la vez que reduce la colinealidad de las variables. Esto significa que cada una de las variables se consideran representativas de las características morfométricas y climáticas de la zona de estudio que no han sido incluidas tras la selección. De esta forma, la radiación total de junio se considera como una aproximación de la radiación estival.

Destacan en la tabla 3.1 la curvatura del perfil (PROF) y la curvatura plana (PLAN), que se mantienen en casi todas las escalas, lo que indica que están escasamente correlacionadas con el resto de las variables. La orientación (SIN y COS) aparecen ambas a la escala menor y mayor, y además, la componente norte aparece a una escala intermedia (COS25.25). De los índices de rugosidad, se mantienen la medida vectorial de la rugosidad (VRM) y el índice de Iwahasi (IWA) a distintas escalas, mientras que el VRI y el índice de Melton desaparecen en todas ellas. También aparecen los dos índices secundarios (factor LS y MRVBF) lo que indica su independencia respecto al resto de variables. De las variables climáticas, además de la radiación anteriormente mencionada, están presentes los valores de precipitación mensual de abril, septiembre y octubre, lo cual resulta de interés al ser los meses en los que se concentran las lluvias en el área mediterránea; también aparece la temperatura estival, representada por la temperatura media mensual de julio y la máxima absoluta de agosto, y la temperatura invernal a través de la mínima absoluta de diciembre. Esta selección de variables permite la elaboración de modelos multirresolución.

A estas variables cuantitativas se añadieron las variables cualitativas (tipo y usos del suelo), sumando 40 variables en total.

<p>Derivadas del terreno</p> <p>SLO25_39 (4.828), SIN25_3 (2.821), SIN25_39 (1.665), COS25_3 (2.094), COS25_15 (4.053), COS25_39 (2.893), PROF25_3 (2.316), PROF25_7 (3.517), PROF25_11 (3.932), PROF25_23 (3.032), PROF25_39 (2.293), PLAN25_3 (1), PLAN25_7 (1.361), PLAN25_11 (1.841), PLAN_15 (2.057), PLAN_19 (2.248), PLAN_23 (2.349), PLAN25_27 (2.226), PLAN25_31 (2.408), PLAN25_35 (2.506), PLAN25_39 (1.83)</p>
<p>Relacionadas con la rugosidad</p> <p>VRM25_11 (2.314), VRM25_39 (4.448), IWA25_27 (3.165), IWA25_39 (3.36)</p>
<p>Relacionadas con la posición</p> <p>TPI25_3 (2.397)</p>
<p>Índices secundarios</p> <p>MRVBF (3.554), LS (4.689)</p>
<p>Variables climáticas</p> <p>AGprectotPREDFIN4 (2.822), AGprectotPREDFIN7 (2.826), AGprectotPREDFIN9 (2.59), AGprectotPREDFIN10 (1.446), AGtemedPREDFIN7 (3.403), AGtemaxabsPREDFIN8 (3.019), AGteminabsPREDFIN12 (1.755), RAD_6 (2.773), RAD_12 (2.48), AGetHPREDFIN9 (3.163)</p>

TABLA 3.1: Variables resultantes del proceso de selección de variables, con el valor de VIF indicado entre paréntesis.

3.2 Modelos a escala regional.

3.2.1 Carbono orgánico.

3.2.1.1 Análisis descriptivo.

El contenido de carbono orgánico (CO) en los suelos analizados muestra una gran variación (tabla 3.2), con un rango comprendido entre 0.06 y 100.6 g · kg⁻¹, pudiendo tratarse el máximo de un valor atípico, al ser el valor medio de 13.42 g · kg⁻¹, y la desviación estándar de 12.66 g · kg⁻¹. La mayor parte de los suelos tiene un contenido bajo en CO, como indica el hecho de que la mediana tenga un valor de 9.4 g · kg⁻¹.

Mín.	Mediana	Media	Máx.	DS	CV(%)
0.060	9.450	13.420	100.600	12.66	94.31

TABLA 3.2: Estadísticos descriptivos del contenido de CO (g · kg⁻¹) en los suelos de la región. **Mín.**= Mínimo; **Máx.**= Máximo; **SD**= Desviación estándar; **CV (%)**. = Coeficiente de variación.

Los datos de CO se han transformado logarítmicamente, con el fin de obtenerse un mayor ajuste a una distribución normal (Fig 3.1) y lograr la homogeneidad de varianza; en este punto, aunque no se ha conseguido una distribución normal (test de Shapiro-Wilk, $W=0.96$, p-valor $< 2.2 \cdot 10^{-16}$), al menos el histograma es razonablemente simétrico, y justifica la utilización de la distribución gaussiana en los modelos GLM, GAM, y el modelo GLM previo a MARS utilizando las variables cualitativas; por otro lado, la transformación da como resultado una variable homocedástica si se comparan los usos de suelo (test de Bartlett, $K^2 = 5.4538$, p-valor=0.4871) pero no al comparar los tipos de suelo (test de Bartlett, $K^2 = 56.6892$, p-valor= $3.995 \cdot 10^{-5}$).

Para comparar los diferentes grupos se ha utilizado un test de ANOVA seguido por un test de Tukey entre los usos de suelo, ya que se cumplía el requisito de homocedasticidad, mientras que en el caso de los tipos de suelo, donde no se cumplían los requisitos, se ha utilizado el test de Kruskal-Wallis seguido por un test de Wilcoxon entre los diferentes pares de grupos de suelos. El intervalo de confianza para todos los test ha sido fijado en $\alpha = 0.05$. La figura 3.2 muestra la distribución del CO en función del uso y el tipo de suelo.

Entre los usos del suelo (figura 3.2A), se han encontrado diferencias significativas (ANOVA, $F = 170.8$; p-valor $< 2 \cdot 10^{-16}$). Los usos agrícolas son los que menor contenido de CO tienen, especialmente los cultivos arbóreos (4) y los cultivos en mosaico (5), no existiendo diferencias significativas entre ellos respecto esta variable. Los usos naturales tienen contenidos más altos de CO, sobre todo los suelos de bosque (6), siendo significativas las diferencias entre este grupo y todos los demás. Los suelos cubiertos por matorral (7) y los suelos cubiertos por matorral o incluso suelos considerados "desnudos" (8), que representan roquedos, matorrales con escasa cobertura y arenales y dunas mostraban también diferencias significativas con los suelos agrícolas, e incluso con los suelos de bosque, siendo el contenido en CO menor que en estos últimos, pero no mostraban diferencias significativas entre sí. Además, en los suelos considerados "con cobertura natural" (6, 7, y 8) se aprecia una mayor variabilidad en el contenido en CO, mientras que en los suelos cultivados el rango de CO es mucho más estrecho.

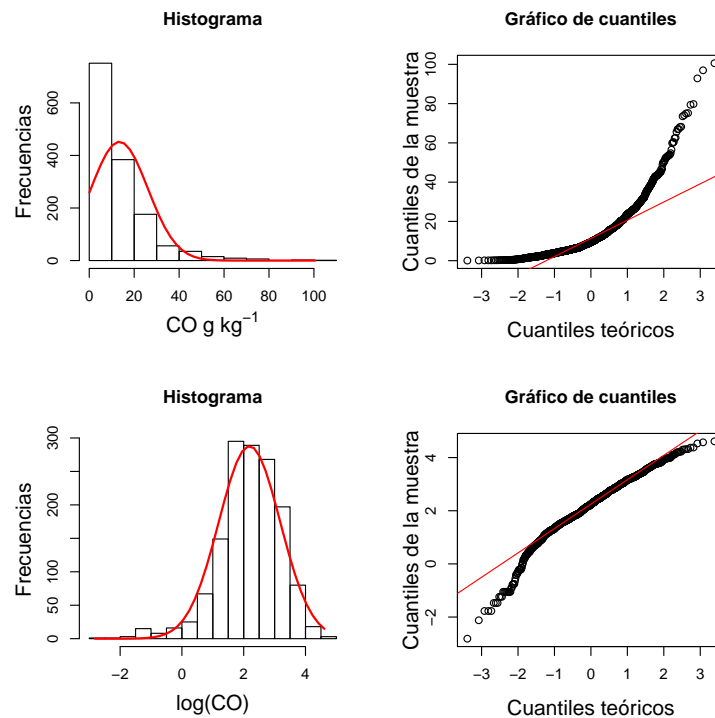
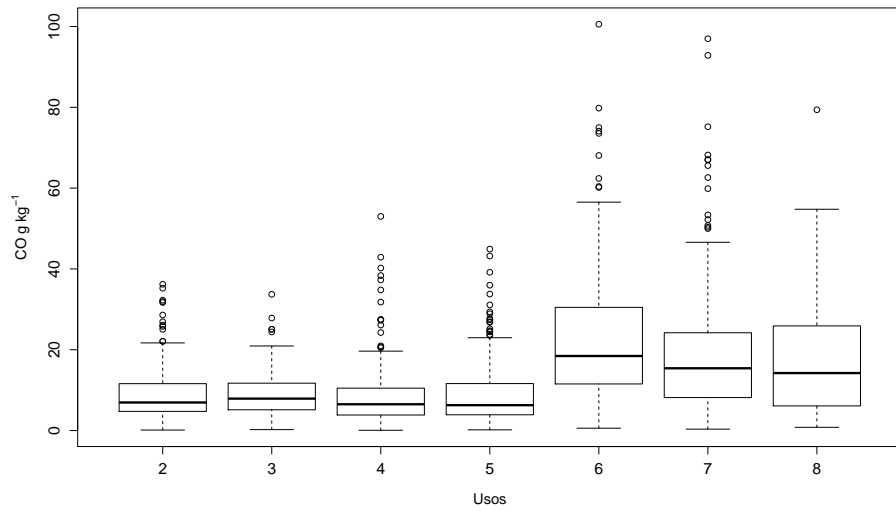


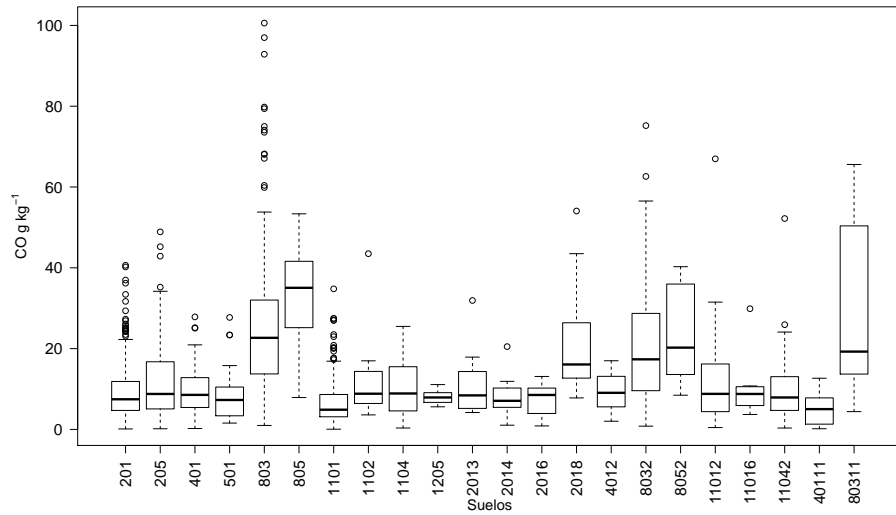
FIGURA 3.1: Histogramas y gráficos de cuantiles de CO sin transformar (arriba) y transformado logarítmicamente (abajo).

En relación a los tipos de suelo (figura 3.2B) los Leptosoles, tanto Líticos (803) como Réndzicos (805) muestran contenidos de CO similares entre sí y superiores, con diferencias significativas, a casi todos los demás grupos, excepto los Leptosoles Réndzicos con Calcisoles (8052) y los Leptosoles Líticos con Regosoles (80311). Los Calcisoles Háplicos con Leptosoles (2018), Leptosoles Líticos con Calcisoles (8032), y Leptosoles Réndzicos con Calcisoles (8052) conformarían otro grupo, con contenidos algo menores de CO pero aún mayores que el resto de suelos, pudiendo añadirse los Leptosoles Líticos con Regosoles (80311), donde se observa la mayor variabilidad en contenido de CO. Por el contrario, los Regosoles Háplicos tienen una cantidad de CO significativamente menor que casi todos los suelos restantes, con excepción de los Fluvisoles Háplicos con Regosoles (40111), los Calcisoles Háplicos asociados con Fluvisoles (2014) y con Gleysoles (2016) y los Solonchaks Háplicos (1205).

Según la tabla 3.3 las variables cuantitativas más correlacionadas de forma positiva con el contenido en CO son la rugosidad (VRM39, VRM11), la pendiente (SLO25_39) y el factor LS, y en menor proporción con la precipitación de otoño, primavera y verano. Por el otro lado, destacan por la correlación negativa la temperatura media del mes de julio, la radiación estival (RAD25_6) y el índice MRVBF, aunque los coeficientes de Pearson son menores que en el caso de las correlaciones positivas. Esto se explica porque las pendientes más abruptas tienen mayor contenido en CO por estar cubiertas de vegetación forestal o de matorral, mientras que



(A)



(B)

FIGURA 3.2: Distribución de CO (g · kg⁻¹) en función del uso y el tipo de suelo

al suavizarse la pendiente, y hacerse más regular el relieve, es común que esta vegetación sea sustituida por cultivos; además, las zonas más húmedas tendrán más cantidad de vegetación y la materia orgánica estará sometida a una menor tasa de mineralización.

Variables geomorfológicas			
Variable	r (Pearson)	Variable	r (Pearson)
SLO25_39	0.33***	SIN25_3	0.007
SIN25_39	-0.02	COS25_3	0.006
COS25_15	-0.001	COS25_39	0.01
PROF25_3	-0.02	PROF25_7	-0.03
PROF25_11	-0.036	PROF25_23	0.035
PROF25_39	0.1***	PLAN25_3	0.02
PLAN25_7	-0.04	PLAN25_11	-0.08**
PLAN25_15	-0.08**	PLAN25_19	-0.09***
PLAN25_23	-0.11***	PLAN25_27	-0.11***
PLAN25_31	-0.05*	PLAN25_35	-0.11***
PLAN25_39	-0.11***	VRM11	0.21***
VRM39	0.37***	IWA25_27	-0.04
IWA25_39	-0.10***	TPI25_3	-0.02
MRVBF	-0.17***	LS	0.30***
Variables climáticas			
Variable	r (Pearson)	Variable	r (Pearson)
AGprectotPREDFIN4	0.20***	AGprectotPREDFIN7	0.21***
AGprectotPREDFIN9	0.23***	AGprectotPREDFIN10	0.07**
AGtemedPREDFIN7	-0.26***	AGtemaxabsPREDFIN8	-0.08**
AGteminabsPREDFIN12	-0.01	AGetHPREDFIN9	-0.15***
RAD25_6	-0.17***	RAD25_12	-0.03

TABLA 3.3: Correlaciones de $\log(\text{CO})$ con las variables predictoras cuantitativas. Los valores de correlación significativos se han resaltado con fondo gris (P-valor: * < 0.05; ** < 0.01; *** < 0.001).

3.2.1.2 Modelos.

La figura 3.3 muestra el número de variables introducidas en el modelo, con el cual se minimiza el RECM en el conjunto de calibración (RECM_c) para cada uno de ellos. En ella se observa que el RECM_c en el caso de MARS-CV es considerablemente mayor que en el resto de los modelos; además la gráfica es prácticamente horizontal, es decir, el aumento del número de variables no disminuye visiblemente el RECM_c . Del resto de modelos, solamente en el caso de SVM-CV se aprecia un mínimo claro en el RECM_c utilizando 10 variables, y un aumento a partir de dicho punto; en los demás, el valor de RECM_c disminuye hasta que se introduce determinado número de variables, a partir del cual se mantiene aproximadamente constante.

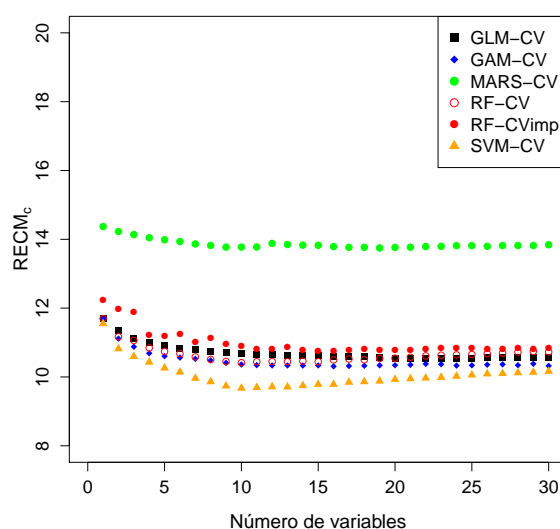


FIGURA 3.3: Raíz del error cuadrático medio (RECM_c) de los modelos a escala regional obtenidos con validación cruzada en bloques para carbono orgánico, en función del número de variables.

La tabla 3.4 y la figura 3.4 muestran, respectivamente, los estadísticos descriptivos y la distribución (histogramas) de la correlación (r_v) y la raíz del error cuadrático medio (RECM_v) para cada uno de los modelos elaborados, calculados sobre el remuestreo en el conjunto de validación. Destaca en primer lugar que al contrario que lo esperado, los valores medios de RECM_v son menores que los valores de RECM_c obtenidos en el conjunto de calibración.

El modelo RF-CV es el que produce una mayor correlación y un menor RECM_v que el resto, ocurriendo lo contrario con el modelo GAM, donde además la dispersión es mayor (alcanzando el RECM_v un valor máximo de $25.75 \text{ g} \cdot \text{kg}^{-1}$) y estos parámetros no tienen una distribución normal, sino bimodal con dos máximos de frecuencias, y en menor medida por el modelo MARS. En el resto de modelos, las distribuciones si se aproximan a una distribución normal y muestran una menor dispersión, situándose los valores de correlación entre 0.5 y 0.7 aproximadamente en todos ellos, y entre 7 y 15 en el caso de RECM_v .

Estos parámetros, junto con el número de variables incluidas en cada modelo y que también se muestra en la tabla 3.4, indican que el mejor resultado se obtiene con RF-CV, seguido por SVM-CV, incluyendo ambos 10 variables, en términos de mayor correlación y menor error cuadrático medio. Los modelos SVM y RF consiguen una alta correlación y un bajo valor de $RECM_v$, pero utilizando todas las variables (40). En el caso de GLM-CV, el uso de la validación cruzada aumenta el $RECM_v$ en comparación con el modelo GLM, utilizando además más variables, por lo que el proceso de validación cruzada da peor resultado que procedimiento que regresión por pasos. En el caso de MARS, el procedimiento de validación cruzada (GCV) interno ya reduce a 25 el número de variables, y el procedimiento de validación cruzada utilizado en MARS-CV sólo reduce ese número en 5 variables más, resultando 20 variables. En el resto de modelos (GAM, SVM, y RF) el procedimiento de validación cruzada si produce un aumento en la exactitud de la predicción a pesar del menor número de variables, menos en el caso de $RF-CV_{IMP}$ donde sólo disminuye el número de variables utilizadas pero no la exactitud.

Modelo		Mín.	Mediana	Media	Max.	SD	Nº de variables
GLM	r_v	0.49	0.66	0.66	0.80	0.05	15
	$RECM_v$	7.62	9.89	9.88	11.87	0.64	
GLM-CV	r_v	0.52	0.66	0.66	0.78	0.04	18
	$RECM_v$	8.00	10.80	10.11	12.21	0.67	
GAM	r_v	0.24	0.48	0.50	0.75	0.1	40
	$RECM_v$	8.75	14.80	14.69	25.72	3.38	
GAM-CV	r_v	0.45	0.63	0.63	0.76	0.05	15
	$RECM_v$	8.35	10.19	10.18	11.87	0.61	
MARS	r_v	0.31	0.55	0.56	0.78	0.08	24
	$RECM_v$	8.6	12.91	12.97	19.76	1.96	
MARS-CV	r_v	0.49	0.65	0.65	0.77	0.05	20
	$RECM_v$	8.03	9.6	9.8	12.02	0.60	
RF	r_v	0.55	0.71	0.72	0.82	0.04	40
	$RECM$	7.47	9.93	9.91	11.97	0.67	
RF-CV	r_v	0.56	0.73	0.73	0.83	0.04	10
	$RECM_v$	7.55	9.51	9.5	11.25	0.6	
RF-CV _{IMP}	r_v	0.52	0.69	0.68	0.81	0.05	15
	$RECM_v$	7.8	10.08	10.12	12.42	0.72	
SVM	r_v	0.51	0.65	0.66	0.79	0.05	40
	$RECM_v$	7.68	9.75	9.76	12.51	0.74	
SVM-CV	r_v	0.46	0.66	0.69	0.79	0.05	10
	$RECM_v$	7.8	9.66	9.67	11.76	0.62	

TABLA 3.4: Modelos a nivel regional obtenidos para Carbono Orgánico ($g \cdot kg^{-1}$). Estadísticos descriptivos de la correlación (r) y Raíz del Error Cuadrático Medio ($RECM_v$) obtenidos tras el remuestreo en el conjunto de validación.

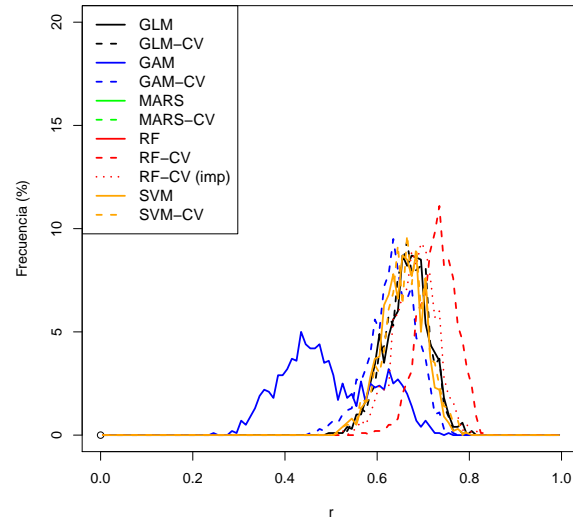
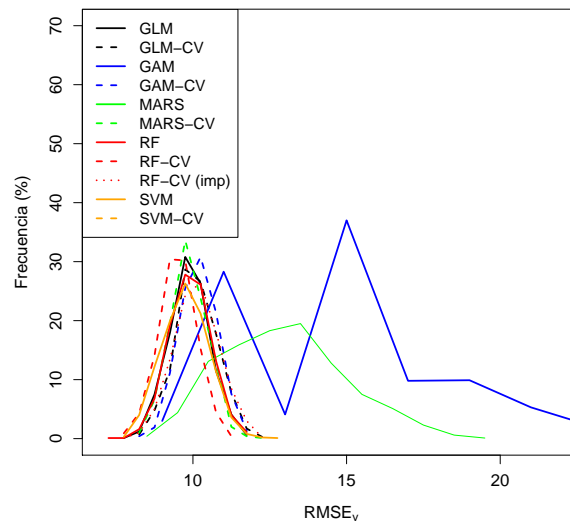
(A) Correlación (r_v)(B) Raíz del error cuadrático medio $RECM_v$

FIGURA 3.4: Modelos a escala regional obtenidos para carbono orgánico. Histogramas de la correlación (r_v) y raíz del error cuadrático medio ($RECM_v$) tras el remuestreo en el conjunto de validación

Por todo ello, se ha seleccionado el modelo RF-CV para elaborar el mapa de CO de la región, ya que es el que mejor resultado predictivo proporciona utilizando la menor cantidad de variables, explicando un 35.63% de la varianza. Sin embargo, tal y como se observa en la figura 3.5, el rango de la predicción, que oscila entre 1.57 y 40.85 $\text{g} \cdot \text{kg}^{-1}$, es más estrecho que el de los datos observados (tabla 3.2) y se aprecia una infraestimación del contenido de este constituyente, con valores de mediana (7.63 $\text{g} \cdot \text{kg}^{-1}$) y media (10.1 $\text{g} \cdot \text{kg}^{-1}$) sensiblemente menores.

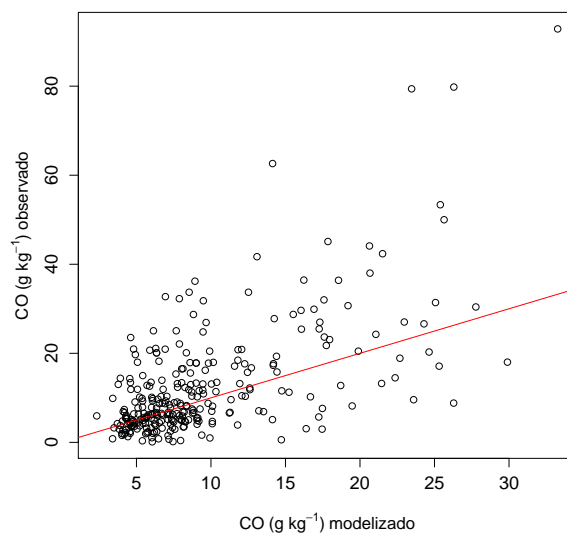


FIGURA 3.5: Diagrama de dispersión de los datos de CO ($\text{g} \cdot \text{kg}^{-1}$) observados en el conjunto de validación frente a los obtenidos con RF-CV a escala regional.

Tal como indica la figura 3.6, el tipo de suelo sería la variable de mayor importancia (es decir, cuya supresión en el modelo produce un mayor incremento en el % de RSS). Le sigue en importancia la rugosidad (VRM39), notablemente distanciada, y a continuación con escasa diferencia en importancia entre ellas, la temperatura media de julio (AGtemedPREDFIN7), los usos del suelo, y el resto de variables climáticas: la precipitación de abril (AGprectotPREDFIN4), la precipitación de septiembre (AGprectotPREDFIN9), la temperatura máxima de agosto (AGtemax-absPREDFIN8) y la precipitación del mes de julio (AGprectotPREDFIN7). Finalmente, otra medida de la rugosidad (IWA25_39) y la curvatura plana (PLAN25_39) son las variables que influyen en el contenido de CO del suelo.

La aparición de las variables relacionadas con la rugosidad, junto con los tipos de suelo, podría tener que ver con los procesos de erosión y sedimentación, ya que los suelos más irregulares suelen ser los Leptosoles, suelos de zonas más abruptas que suelen estar cubiertos de vegetación natural, mientras que en las zonas más lisas y llanas, la vegetación de matorral y de bosque suele ser reemplazada por cultivos. Las variables climáticas que aparecen se corresponden con los periodos en los que mayores precipitaciones (primavera y otoño) y temperaturas (verano) se producen. A pesar de su escasa importancia, la aparición de la curvatura plana también está

relacionada con los procesos de pérdida o acumulación de materiales, aunque su efecto se produce de forma combinada con la curvatura del perfil, que no aparece en el modelo.

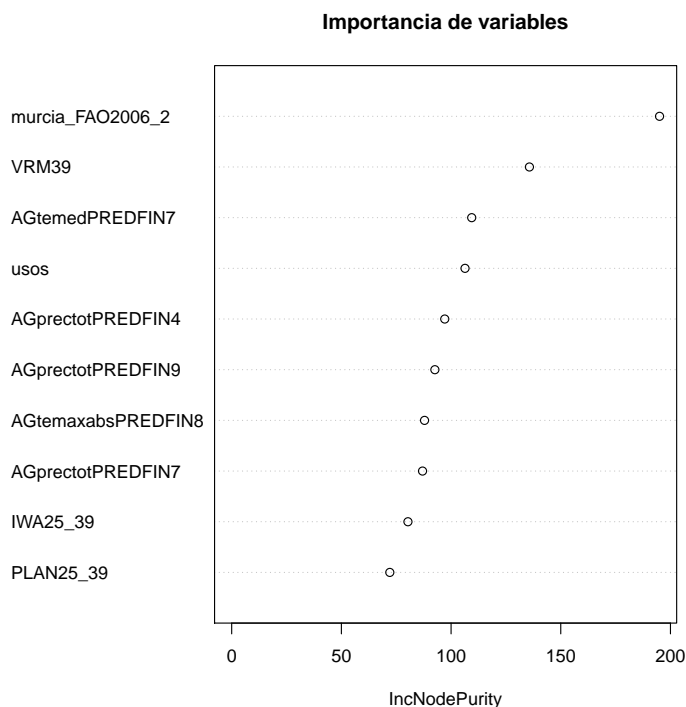


FIGURA 3.6: Importancia de variables en RF-CV expresada como criterio de impureza del nodo.

La elevada importancia del tipo y uso de suelo en el contenido de CO ha sido descrita anteriormente por [Grimm *et al.* \(2008\)](#); [Wiesmeier *et al.* \(2011, 2014\)](#), utilizando también Random Forest como método de regresión. En el estudio de [Xiong *et al.* \(2014\)](#), el uso (representado como el tipo de vegetación), junto con el gradiente de humedad del suelo, era la variable más relevante para la predicción de CO, aunque el método seleccionado fue Boosted Trees. De ellos, los dos primeros no incluyen variables climáticas entre los predictores, sino solamente variables geomorfológicas, mientras que [Wiesmeier *et al.* \(2014\)](#) incluye además la precipitación total anual y la temperatura media anual, aunque esta última no resultó significativa en el modelo que obtuvo. Por otro lado, en el trabajo de [Grimm *et al.* \(2008\)](#) esta importancia es variable, ya que otras variables geomorfológicas, relacionadas con la posición topográfica (la posición relativa y el área contribuyente) tendrían más importancia para la modelización del CO superficial (0-10 cm), mientras que a mayores profundidades (20-30 cm) el tipo de suelo tendría más importancia. En el caso de [Wiesmeier *et al.* \(2011\)](#) el uso del suelo resultaba de mayor importancia que el tipo del mismo, al contrario que en nuestro modelo.

No obstante, tal como se comentó en la sección 2, la alta capacidad predictiva de este método viene contrarrestada por la dificultad de interpretar físicamente las relaciones entre las variables dependientes y las variables predictoras, pues se trata de un promedio de un número conside-

rablemente alto de árboles de regresión. Con el fin de analizar tales relaciones, se ha obtenido un GAM con las variables utilizadas en el modelo RF-CV. Otra posibilidad es utilizar un árbol de regresión (CART) (Wiesmeier *et al.*, 2011, 2014); sin embargo, los modelos CART presentan algunas debilidades, tales como la inestabilidad, la falta de suavizado y la dificultad a la hora de modelizar estructuras aditivas (Hastie *et al.*, 2009). Por ello, se considera que el modelo GAM es más adecuado para este fin, teniendo además una gran interpretabilidad.

El GAM obtenido tiene una capacidad predictiva algo mejor que la obtenida anteriormente con GAM-CV, siendo la correlación de 0.66 y el $RECM_v$ de 9.94, aunque en este caso, este resultado predictivo se alcanza con solo 10 variables, frente las 15 que tenía el otro modelo.

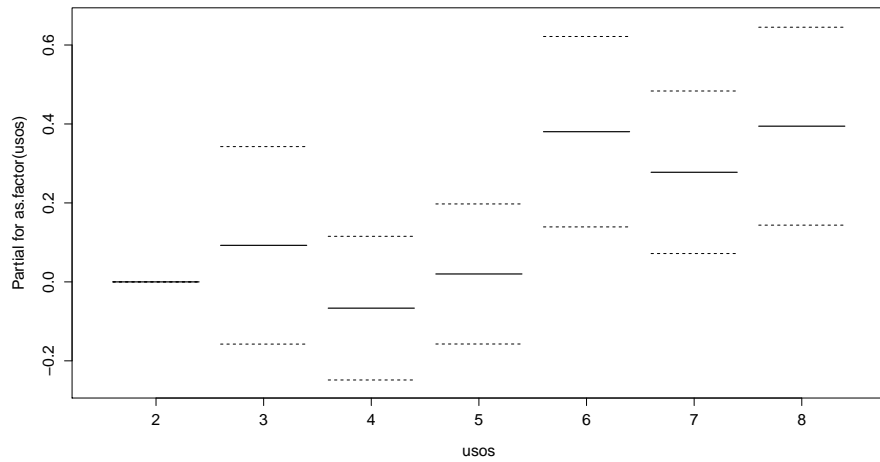
De los predictores paramétricos, resultan significativos (con p-valor < 0.05) los niveles 6,7 y 8 (bosque, matorral y suelo desnudo, respectivamente). Destaca el hecho de que el contenido en CO predicho en suelos con matorral es menor que en los suelos de roquedo y bosque, pero aún así más alto que en los suelos bajo uso agrícola, tanto herbáceas de secano (2), como regadío (3), cultivos arbóreos (4) y en mosaico (5).

Esto se ha atribuido a que las zonas consideradas como suelo desnudo son roquedales o zonas con escaso desarrollo de suelo, lo que a su vez se confirma al comprobar que coinciden con determinadas unidades de suelos que también resultan significativas: 803, 8032 y 80311 (Leptosoles Líticos, Leptosoles Líticos con Calcisoles Háplicos y Leptosoles Líticos con Regosoles Háplicos). En estas unidades se encuentran los mayores valores de CO, y se caracterizan precisamente por el escaso desarrollo en profundidad. Los Calcisoles Pétricos (205) también resultan significativos, aunque con un contenido de CO intermedio. En los Regosoles Háplicos (1101), los Regosoles Lépticos (1104), y los Regosoles Lépticos con Calcisoles (11042), los contenidos de CO predichos son menores que en el resto, lo cual puede relacionarse con las características estructurales de estos suelos, como la escasa capacidad de retención de humedad o incluso la limitación en profundidad, que resultan adversas para la vegetación.

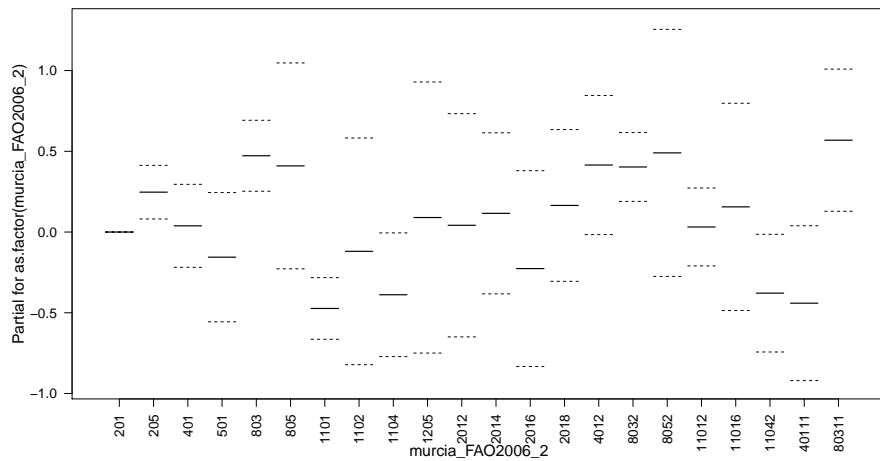
El resto de unidades no resultan significativas en el modelo, debido al amplio intervalo de confianza del error. Estos resultados coinciden con lo observado al analizar la distribución de la variable, en la figura 3.2.

De las variables suavizadas (figuras 3.8 y 3.9), la más significativa (p-valor = $1.71 \cdot 10^{-5}$) es la temperatura media del mes de julio (AGtemedPREDFIN7, transformada logarítmicamente), que muestra relación negativa en la mayor parte del rango con el contenido en CO, lo cual concuerda con la hipótesis de partida de que el aumento de la temperatura estival (época del año donde menor cantidad de precipitaciones se producen) se relaciona con una mayor mineralización de la materia orgánica.

La precipitación total del mes de abril y septiembre (AGprectotPREDFIN4, AGprectotPREDFIN9) también son significativas (p-valor = 0.04 y p-valor = 0.03 respectivamente) aunque solo en el caso de la precipitación otoñal el efecto en el contenido en CO es claramente positivo, ya que en el caso de la primavera a partir de cierto nivel de precipitaciones se observa una disminución del contenido de CO que puede estar relacionado con otras variables y con la pérdida de este constituyente debido a la erosión. La temperatura máxima del mes de agosto (AGtemaxabSPREDFIN8) y la precipitación del mes de julio (AGprectotPREDFIN7) no resultan significativas en el modelo (p-valor > 0.05). Esta relación entre los efectos (no lineales) de los factores



(A) Efecto del uso



(B) Efecto del tipo de suelo

FIGURA 3.7: Efecto de las variables cualitativas en el modelo GAM para CO obtenido con las variables de RF-CV_{IMP}.

climáticos y el contenido en CO del suelo ha sido ampliamente descrito en otros diversos estudios, recopilados por [Grumwald \(2009\)](#). La mayoría de ellos sugieren un aumento del contenido de CO al aumentar la precipitación, considerada como promotora de la producción primaria al aportar agua disponible para las plantas, mientras que las temperaturas altas favorecen la actividad microbiana necesaria para la descomposición de la materia orgánica ([Wynn *et al.*, 2006](#); [Xiong *et al.*, 2014](#)), pero también la mineralización de la misma.

De las variables geomorfométricas son significativas la rugosidad vectorial (VRM_39; p-valor = 0.008) y la curvatura plana (PLAN25_39; p-valor = 0.005); la rugosidad tiene una relación

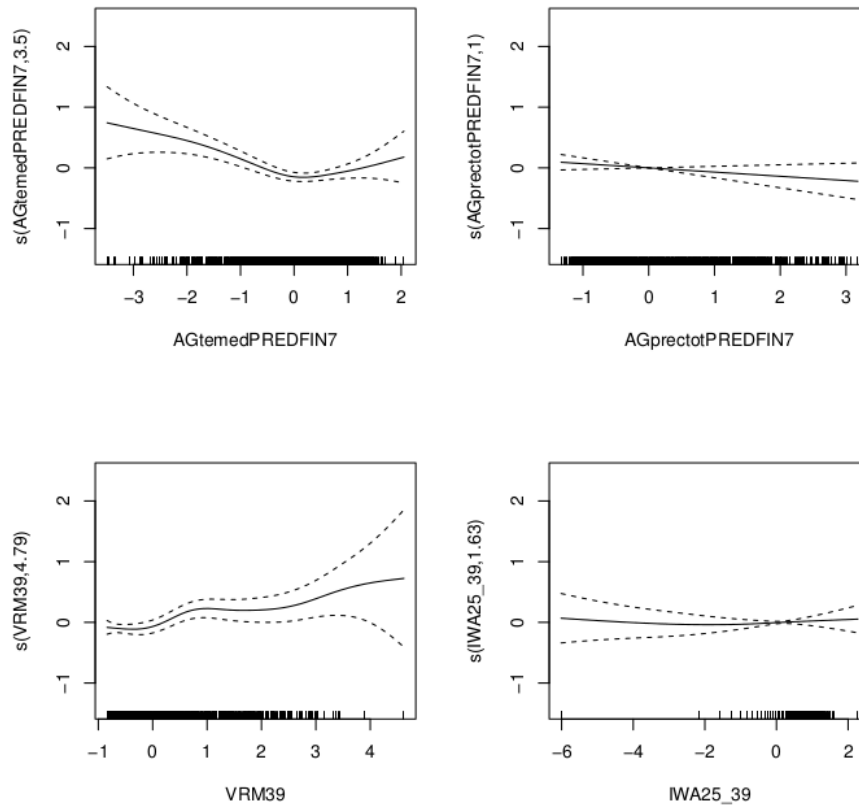


FIGURA 3.8: Modelo GAM obtenido con las variables de RF-CV_{IMP} para CO. Curvas de respuesta para las variables suavizadas (I).

positiva con el contenido en CO, que coincide con la alta correlación entre ambas variables que se observaba en la tabla 3.3, mientras la curvatura plana muestra una relación negativa y lineal con el contenido en CO, lo que sugiere que esta variable se relaciona con pendientes cóncavas en las que se produce pérdida de material. Teóricamente, una mayor rugosidad aumentaría la capacidad de almacenamiento de suelo por parte de las zonas deprimidas y, por tanto la cantidad de agua retenida por el mismo, lo que se traduciría en mejores condiciones para el crecimiento de la vegetación con la consiguiente posibilidad de que se deposite la materia orgánica (Hansen *et al.*, 1999); además, esto refleja la influencia de las diferentes escalas, pues esas zonas más rugosas representarían las áreas más montañosas e irregulares de la región, frente las zonas más lisas. La importancia de la rugosidad también viene reflejada por la presencia del índice de Iwahasi (IWA25_27, transformado logarítmicamente), aunque esta variable no ha resultado significativa en el modelo. She *et al.* (2014) obtuvo, utilizando GLM, que la pendiente y la rugosidad del terreno eran las variables topográficas que mayor influencia tenían en el contenido en CO, aunque utilizaban un método distinto para calcular la rugosidad. Sin embargo, el estudio de She *et al.*

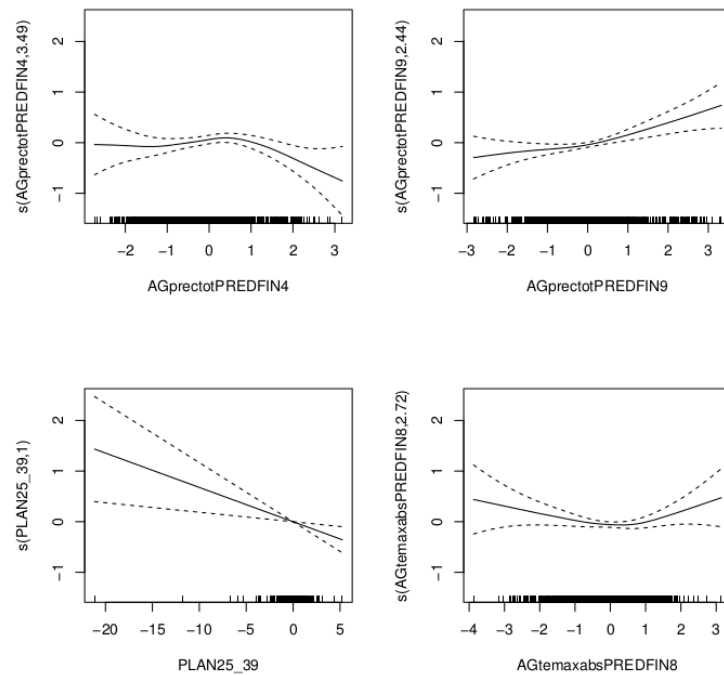


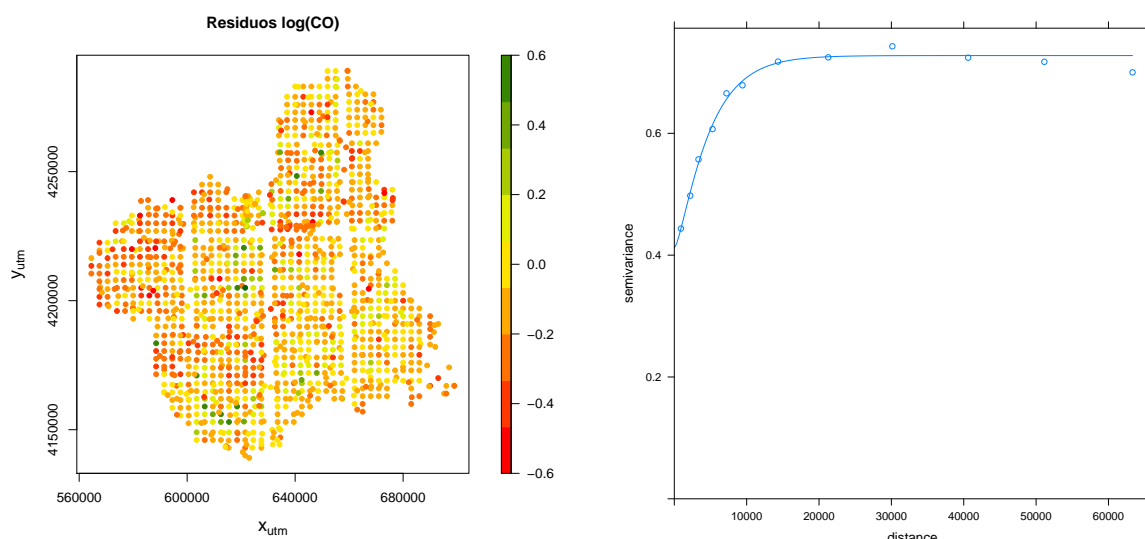
FIGURA 3.9: Modelo GAM obtenido con las variables de RF-CV_{IMP} para CO. Curvas de respuesta para las variables suavizadas (II).

(2014) se refiere a zonas con usos o cubiertas vegetales relativamente homogéneas, concretamente pastizales, mientras que en la región de estudio existen diversos usos tanto naturales como antropizados.

3.2.1.3 Residuos del modelo.

La figura 3.10A muestra la distribución de los residuos del modelo en la Región. En ella se aprecia el amplio rango en el que varían los residuos, que se han calculado restando el valor original de CO a la predicción arrojada por el modelo. Aún teniendo en cuenta que este residuo es el logaritmo del valor original, los valores oscilan entre -3 y 4. El residuo es de signo negativo en la mayor parte de la región, reflejando la subestimación del contenido en CO, sobre todo en la zona del Noroeste, y muy pocos puntos, de color verde, reflejan una sobreestimación del CO del suelo.

Respecto el semivariograma, en la figura 3.10B se observa un ajuste al modelo de Matern con parametrización de Stein, con un valor de pepita (nugget) de 0.41 frente a la meseta total de 0.72, alcanzada a los 6504 m. La relación entre nugget y meseta (NSR; Nugget to Sill Ratio) de 56% indica una estructura espacial moderada (Cambardella *et al.*, 1994) por lo que se han interpolado los residuales utilizando un Krigeado Ordinario (KO).



(A) Distribución de los residuos del modelo RF-CV en la Región

(B) Semivariograma de los residuos

FIGURA 3.10: Distribución espacial de los residuos del modelo RF-CV para carbono orgánico.

El mapa elaborado añadiendo los residuos interpolados a la predicción da como resultado una correlación de 0.76 con los datos originales en el conjunto de validación, y un error $RECM_v$ de $6.64 \text{ g} \cdot \text{kg}^{-1}$, aunque para valorar este resultado se debe tener en cuenta que no se ha realizado un procedimiento de validación cruzada en el proceso final de regresión-krigeado. La figura 3.11 muestra que los rangos de los datos observados, del mapa de la predicción (página A.1) y de la predicción con krigeado (página A.2) se superponen entre sí aunque los valores originales varíen en un rango más amplio; además, el test de Kruskal-Wallis realizado para comparar todos los mapas no refleja diferencias significativas entre los tres ($p\text{-valor} > 0.05$).

Sin embargo, este resultado del test de Kruskal puede deberse a la diferencia en las varianzas (que es mucho mayor en la variable original que en la modelizada), a los que este test sí es sensible, por lo que se ha realizado además el t-test de Welch. Este test detecta diferencias entre las medias de los valores originales de CO y los valores arrojados por el modelo ($p\text{-valor}=0.0005$) pero no entre los valores originales y el mapa de regresión-krigeado ($p\text{-valor}=0.64$). El test de Wilcoxon en cambio encuentra diferencias tanto entre los valores originales y el mapa de predicción+krigeado ($p\text{-valor}=0.0005$) como entre los valores originales y el mapa de predicción ($p\text{-valor} < 2 \cdot 10^{-16}$).

El mapa finalmente seleccionado (página A.2) muestra que el contenido en CO es mayor en las zonas más accidentadas topográficamente, destacando en este sentido la región del Noroeste y en menor medida el Altiplano. También en sierras como la de Espuña hay valores más altos de CO, así como en la sierra de Cartagena al sureste de la región. Todos ellos coincidirían con usos forestales, de matorral o canchales, mientras que por el contrario los valores más bajos se registran en las depresiones interiores y el sector suroeste, coincidiendo en su distribución con

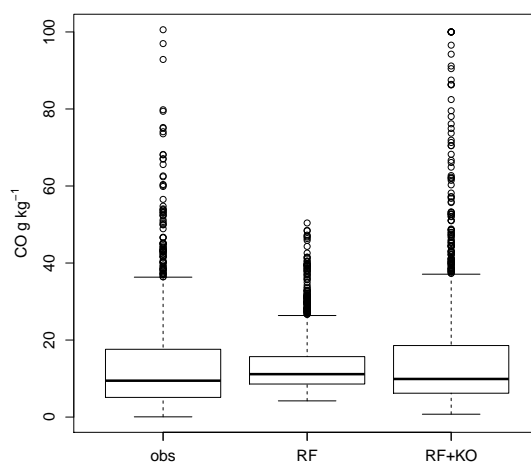


FIGURA 3.11: Gráfico de cajas con la distribución del contenido de CO a escala regional. **obs**: valores observados; **RF**: mapa de predicción del modelo RF-CV ; **RF+KO**: suma de la predicción de RF-CV y los residuos interpolados con KO.

zonas de margas y usos agrícolas de secano. Existe también un sector de la Vega del Guadalentín donde se aprecian valores medios de CO, así como en la cuenca del Mar Menor y la Vega del Segura, lo cual puede atribuirse a la presencia de cultivos de regadío. Esto coincide con las hipótesis planteadas en la introducción y el análisis descriptivo de la variable realizado anteriormente.

3.2.2 Carbonato cálcico total o equivalente.

3.2.2.1 Análisis descriptivo.

El carbonato cálcico total o equivalente ($\text{CaCO}_{3\text{total}}$), que en este estudio se ha expresado en tanto por uno y no en tanto por ciento como en la base de datos original, no mostraba una distribución normal (figura 3.12; test de Shapiro-Wilk, $W = 0.9711$, p-valor $< 2.2 \cdot 10^{-16}$), por lo que se ha elegido la distribución "Gamma" a la hora de elaborar los modelos GLM y GAM. Para ello, los valores de carbonato cálcico total iguales a 0 han debido ser convertidos en valores positivos (0.0001) con el fin de evitar errores en la modelización.

Los estadísticos descriptivos del contenido de $\text{CaCO}_{3\text{total}}$ están resumidos en la tabla 3.5. De ellos puede deducirse que los suelos de la región tienen un alto contenido de $\text{CaCO}_{3\text{total}}$ en la capa arable, siendo el valor medio de $0.40 \text{ kg} \cdot \text{kg}^{-1}$ y la mediana de $0.43 \text{ kg} \cdot \text{kg}^{-1}$ ligeramente superior a la media, estando el rango entre 0 y $0.95 \text{ kg} \cdot \text{kg}^{-1}$, y que atendiendo al valor de la desviación estándar, presentan una alta variabilidad.

A la hora de comparar el contenido de $\text{CaCO}_{3\text{total}}$ entre los diferentes usos del suelo, el test de

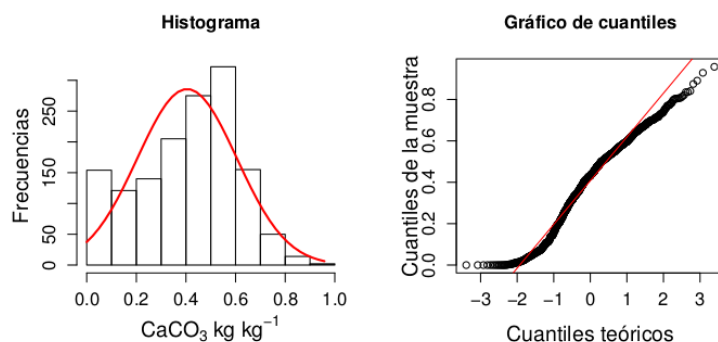


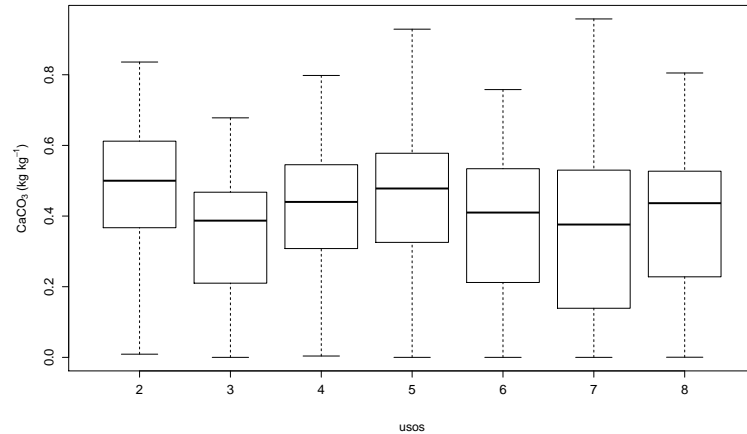
FIGURA 3.12: Histograma de frecuencias y gráfico de cuantiles para carbonato cálcico total o equivalente ($\text{kg}\cdot\text{kg}^{-1}$)

Mín.	Mediana	Media	Máx.	DS	CV(%)
0.0000	0.4320	0.4046	0.9580	0.20	49.6

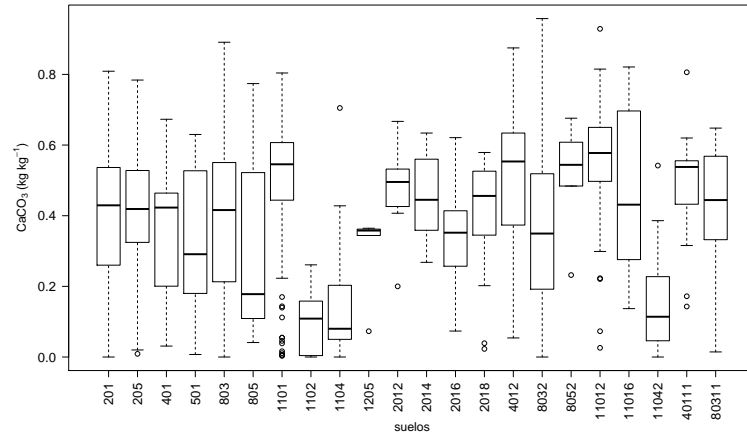
TABLA 3.5: Estadísticos descriptivos del contenido de CaCO_3 total o equivalente en los suelos de la Región de Murcia. **Mín.** = Mínimo; **Máx.** = Máximo; **SD** = Desviación estándar; **CV (%)** = Coeficiente de variación.

Bartlett indica que no hay homogeneidad de varianzas ($K^2 = 37.1936$, $p\text{-valor} = 1.614 \cdot 10^{-6}$) y por tanto se ha seleccionado el test de Kruskal Wallis, según el cual existen diferencias significativas entre ellos ($\chi^2 = 76.5852$; $p\text{-valor} = 1.81 \cdot 10^{-14}$). Posteriormente se ha utilizado el test de Wilcoxon para buscar diferencias entre pares. La figura 3.13A muestra una diferenciación aparente de los distintos usos, aunque no tan clara como en el caso del CO. Los suelos bajo cultivo de secano (2) tendrían mayor cantidad de $\text{CaCO}_{3\text{total}}$ en la capa arable, junto con los suelos bajo cultivo en mosaico (4) y con cultivos arbóreos (5), que los suelos bajo cultivo de regadío, los suelos de bosque (6), los cubiertos por matorral (7) y los suelos bajo cultivo de regadío (3), siendo estas diferencias significativas según el test de Wilcoxon, aunque menos pronunciadas entre los suelos en regadío y los suelos de roquedo. También cabe destacar que, a pesar de esa diferenciación aparente, los suelos bajo cultivos arbóreos tienen un contenido en $\text{CaCO}_{3\text{total}}$ significativamente mayor que los suelos de regadío, pero similar a los suelos desnudos o de roquedal y los suelos de bosque.

Respecto a los tipos de suelo (figura 3.13B), también existen diferencias significativas ($\chi^2 = 245.8345$; $p\text{-valor} < 2.2 \cdot 10^{-16}$). Destacan las dos unidades donde el contenido en $\text{CaCO}_{3\text{total}}$ de la capa arable es notablemente inferior al resto: Regosoles Lépticos (1104) y Regosoles Lépticos con Calcisoles (11042). Este resultado es contrario al esperado, sobre todo por el hecho de que estos suelos a menudo están formados por materiales sueltos de carácter calcáreo, como las margas, especialmente en el caso de las asociaciones con Calcisoles. Por el contrario, en los Regosoles Háplicos (1101) y otras unidades donde aparecen asociados Calcisoles (Leptosoles Réndzicos con Calcisoles, 8052; Regosoles Háplicos con Calcisoles, 11012) se observa un contenido en $\text{CaCO}_{3\text{total}}$ de la capa arable mayor que en el resto de unidades. Los Calcisoles Háplicos (201), así como los asociados con Gleysoles (2016) y Leptosoles (2018), al igual que los Calcisoles Pétricos (205) tienen contenidos menores de $\text{CaCO}_{3\text{total}}$ en la capa arable. Otras unidades muestran rangos



(A) Distribución en función del uso.



(B) Distribución en función del tipo de suelo.

FIGURA 3.13: Distribución de CaCO_3 total o equivalente ($\text{kg} \cdot \text{kg}^{-1}$) en función del uso y el tipo de suelo

bastante amplios, como es el caso de los Leptosoles Líticos (803), Leptosoles Réndzicos (805), y Regosoles Háplicos con Gleysoles (11016) que se solapan con otros grupos, por lo cual las diferencias encontradas no revisten significación estadística.

La tabla 3.21 muestra las correlaciones entre el $\text{CaCO}_{3\text{total}}$ y las variables predictoras cuantitativas. El contenido en $\text{CaCO}_{3\text{total}}$ estaba correlacionado positivamente y de forma significativa ($p < 0.05$) con todas las variables climáticas, especialmente con la humedad relativa de septiembre, la temperatura máxima absoluta de agosto y la temperatura mínima absoluta de diciembre, siendo las correlaciones más débiles en el resto de variables climáticas. Aunque este resultado coincide con la hipótesis de que en las zonas más cálidas (con mayores temperaturas máximas en agosto y mínimas en diciembre) se esperaría un mayor contenido de en la capa arable, debido a mayores tasas de evaporación, por otro lado sería de prever una correlación negativa entre la precipitación y el contenido de $\text{CaCO}_{3\text{total}}$ en la capa arable, al favorecer ésta el lavado de los carbonatos hacia los horizontes inferiores.

Variables geomorfológicas			
Variable	r (Pearson)	Variable	r (Pearson)
SLO25_39	-0.04	SIN25_3	-0.07**
SIN25_39	-0.05*	COS25_3	-0.02
COS25_15	0.019	COS25_39	-0.007
PROF25_3	-0.006	PROF25_7	-0.009
PROF25_11	0.003	PROF25_23	-0.018
PROF25_39	-0.072**	PLAN25_3	0.035
PLAN25_7	-0.03	PLAN25_11	-0.01
PLAN25_15	0.033	PLAN25_19	0.005
PLAN25_23	0.001	PLAN25_27	-0.02
PLAN25_31	-0.015	PLAN25_35	0.006
PLAN25_39	0.01	VRM11	-0.1***
VRM39	-0.14***	IWA25_27	0.014
IWA25_39	-0.006	TPI25_3	-0.03
MRVBF	-0.01***	LS	-0.06*
Variables climáticas			
Variable	r (Pearson)	Variable	r (Pearson)
AGprectotPREDFIN4	0.24***	AGprectotPREDFIN7	0.1***
AGprectotPREDFIN9	0.17***	AGprectotPREDFIN10	0.13**
AGtemedPREDFIN7	0.10***	AGtemaxabsPREDFIN8	0.28***
AGteminabsPREDFIN12	0.26***	AGetHPREDFIN9	0.30***
RAD25_6	0.13***	RAD25_12	0.067*

TABLA 3.6: Correlaciones entre el contenido en CaCO_3 total y las variables predictoras cuantitativas. Los valores de correlación significativos se han resaltado con fondo gris (P-valor: * < 0.05; ** < 0.01; *** < 0.001).

En el caso de las variables geomorfológicas, por el contrario, las correlaciones son negativas en su mayoría y mucho más débiles que en el caso de las variables climáticas, careciendo de significación estadística excepto en el caso del seno de la orientación, la curvatura del perfil (PROF25_39), la rugosidad (VRM) y el factor LS.

3.2.2.2 Modelos.

La figura 3.14 muestra la raíz del error cuadrático medio (RECM_c) obtenido en la validación cruzada en función del número de variables utilizadas, para cada uno de los modelos, medido en el conjunto de calibración. Este dato se ha utilizado para seleccionar el número de variables a incluir en los modelos con validación cruzada, seleccionando aquel con el cual el RECM_c es menor. En el gráfico destaca que el modelo RF-CV obtiene menor RECM_c que el resto de modelos, mientras que con el modelo GLM-CV el error obtenido es el mayor, seguido por RF-CV_{IMP}. Por otro lado, en los modelos GLM-CV, GAM-CV y MARS-CV, la disminución del error (RECM_c) al introducir nuevas variables es poco perceptible, al contrario de lo que ocurre con los modelos de RF-CV, RF-CV_{IMP} y SVM-CV, donde sí se observa una disminución del error al aumentar la complejidad de los modelos, hasta un determinado número de variables, a partir del cual el error aumenta al introducir más predictores.

Este mejor resultado de RF-CV se refleja al calcular los mismos parámetros en el conjunto de validación. La tabla 3.7 muestra los estadísticos descriptivos de correlación y RECM_v obtenidos en el conjunto de validación tras realizar el remuestreo (bootstrapping), cuya distribución (histogramas) puede observarse en la figura 3.15.

De los modelos elaborados, RF-CV_{IMP} proporciona el mayor promedio de la correlación (0.69) y

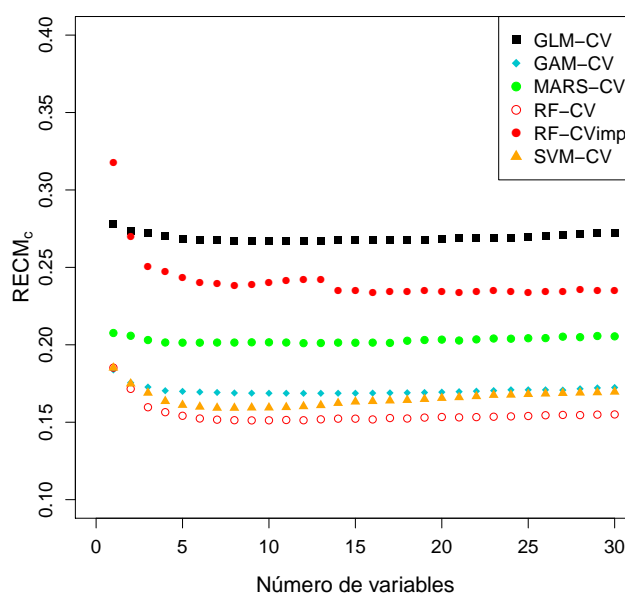


FIGURA 3.14: Raíz del error cuadrático medio ($RECM_c$) de los modelos a escala regional obtenidos con validación cruzada en bloques para carbonato cálcico total o equivalente, en función del número de variables.

al mismo tiempo, el promedio de $RECM_v$ es el más bajo (0.11), utilizando solamente 8 variables. Sin embargo, la distribución del $RECM_v$ en el remuestreo no es normal ni simétrica (figura 3.15.B). Bajo los mismos criterios, el segundo mejor método sería Random Forest con validación cruzada (RF-CV), que consigue resultados similares con una variable más. SVM-CV también da como resultado un valor de $RECM_v$ bajo utilizando solamente 7 variables, pero una menor correlación (figura 3.15A).

Por el contrario, los modelos GLM-CV, GLM, MARS-CV y RF son los que dan como resultado un mayor $RECM_v$ (figura 3.15B). Además, el gráfico de $RECM_v$ para GLM-CV tampoco sigue una distribución normal, y la correlación obtenida con este modelo es la menor, muy por debajo incluso que al utilizar regresión por pasos (GLM). El modelo MARS también resulta en una baja correlación, y el uso del procedimiento de validación cruzada resulta en una correlación similar a GLM, pero un aumento del error.

Finalmente, el modelo seleccionado es Random Forest con validación cruzada (RF-CV), por ser el que mayor correlación y menor $RECM_v$ ofrece con el menor número de variables (9), explicando el 44.41% de la varianza total. La figura 3.16 muestra que los datos observados y los modelizados se ajustan de forma aproximada a la recta; además, en los valores estimados por el modelo, la media (0.40%) y la mediana (0.43%) son similares a los valores observados. Sin embargo, el rango de los valores predichos por el modelo varía en un rango más estrecho que los datos reales (0.08-0.65%), produciéndose una infraestimación del contenido de $CaCO_{3total}$ en los

Modelo		Mín.	Mediana	Media	Max.	SD	Nº de variables
GLM	r_v	0.34	0.51	0.51	0.63	0.04	19
	RECM _v	0.22	0.25	0.25	0.28	0.01	
GLM-CV	r_v	0.11	0.16	0.26	0.57	0.20	10
	RECM _v	0.16	0.40	0.38	0.84	0.17	
GAM	r_v	0.16	0.43	0.42	0.63	0.09	40
	RECM _v	0.15	0.19	0.19	0.27	0.02	
GAM-CV	r_v	0.41	0.57	0.57	0.71	0.04	12
	RECM _v	0.14	0.16	0.16	0.19	0.007	
MARS	r_v	0.16	0.34	0.34	0.48	0.05	25
	RECM _v	0.17	0.19	0.19	0.22	0.007	
MARS-CV	r_v	0.34	0.53	0.53	0.66	0.05	12
	RECM _v	0.21	0.25	0.25	0.29	0.01	
RF	r_v	0.53	0.65	0.65	0.75	0.03	40
	RECM _v	0.19	0.22	0.22	0.25	0.009	
RF-CV	r_v	0.57	0.68	0.68	0.80	0.03	9
	RECM _v	0.12	0.14	0.15	0.17	0.007	
RF-CV _{IMP}	r_v	0.57	0.69	0.69	0.79	0.04	8
	RECM _v	0.0001	0.09	0.11	0.44	0.08	
SVM	r_v	0.35	0.51	0.51	0.66	0.04	40
	RECM _v	0.15	0.17	0.17	0.18	0.01	
SVM-CV	r_v	0.49	0.63	0.63	0.74	0.04	7
	RECM _v	0.13	0.16	0.16	0.18	0.007	

TABLA 3.7: Modelos a nivel regional obtenidos para CaCO₃ total o equivalente (kg·kg⁻¹). Estadísticos descriptivos de la correlación (r) y Raíz del Error Cuadrático Medio (RECM_v) obtenidos tras el remuestreo en el conjunto de validación.

valores más altos, si se compara con la tabla 3.5.

Las variables utilizadas en este modelo son casi todas climáticas (Figura 3.64), excepto el tipo de suelo, y dos variables geomorfométricas. Las variables que más importancia tienen son la temperatura mínima absoluta de diciembre (AGteminabsPREDFIN12) y el tipo de suelo (murcia_FAO2006_2). A continuación aparecen la precipitación total mensual de abril (AGprectotPREDFIN4), y con menor variación en el RSS las precipitaciones totales de julio (AGprectotPREDFIN7) y septiembre ((AGprectotPREDFIN9), la temperatura máxima absoluta de agosto (AGtemaxabsPREDFIN8), y la temperatura media del mes de julio (AGtemedPREDFIN7). Finalmente, aparecen las dos variables geomorfométricas, una de ellas relacionada con la orientación de la pendiente (COS25_39) y la otra con la rugosidad (IWA25_39), calculadas ambas con 39 celdillas.

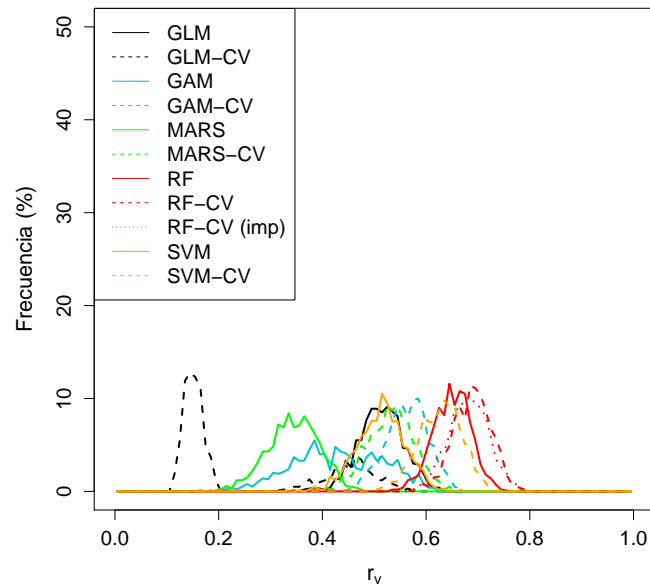
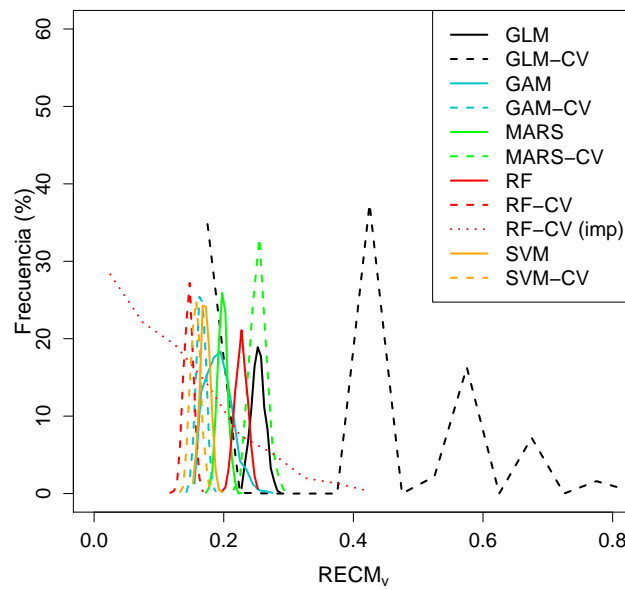
(A) Correlación (r_v)(B) Raíz del error cuadrático medio (RECM_v)

FIGURA 3.15: Modelos a escala regional obtenidos con validación cruzada en bloques para $\text{CaCO}_{3\text{total}}$. Histogramas de la correlación (r_v) y raíz del error cuadrático medio (RECM_v) tras el remuestreo en el conjunto de validación.

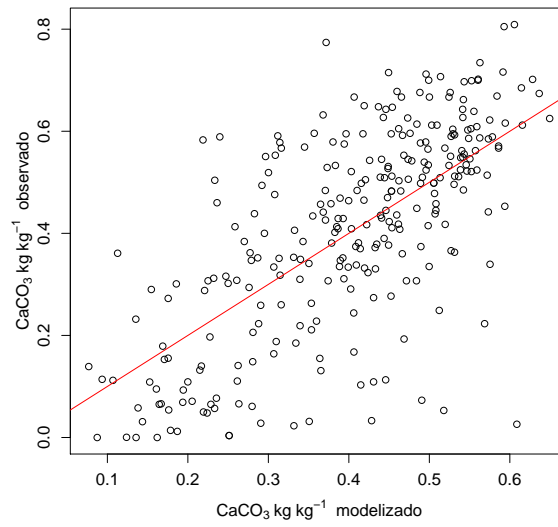


FIGURA 3.16: Diagrama de dispersión de los datos de carbonato cálcico total o equivalente ($\text{kg} \cdot \text{kg}^{-1}$) observados en el conjunto de validación frente los obtenidos con RF-CV a escala regional.

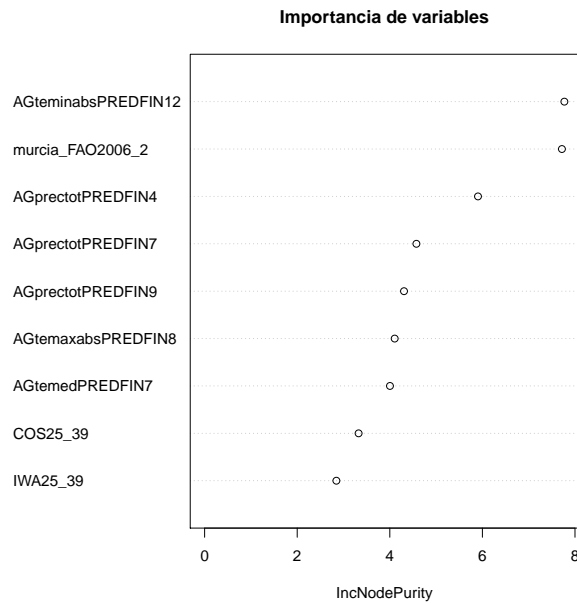


FIGURA 3.17: Importancia de variables en RF-CV para carbonato cálcico total o equivalente, expresada como criterio de impureza del nodo.

El GAM elaborado con las variables del modelo RF-CV (GAM-CV), con el fin de comprender la influencia de cada una en la concentración de $\text{CaCO}_3_{\text{total}}$, da como resultado un valor medio de correlación de 0.55 y un RECM_v de 0.17 entre los datos observados y los predichos por el modelo en el conjunto de validación, tras realizar un remuestreo, siendo similares a los obtenidos con el GAM-CV elaborado anteriormente.

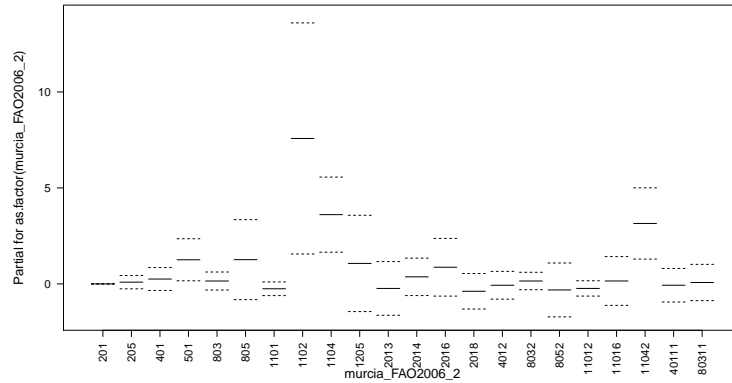


FIGURA 3.18: Modelo GAM obtenido con las variables de RF-CV para carbonato cálcico total o equivalente. Efecto de la variable suelo.

En este modelo GAM-CV resultan significativos principalmente los Regosoles Lépticos (1104) y los Regosoles Lépticos asociados con Calcisoles (11042), y en menor medida los Gipsisoles Háplicos (501) y los Regosoles Eútricos (1102).

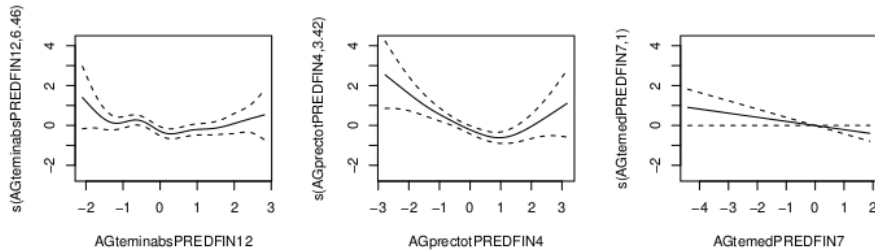


FIGURA 3.19: Modelo GAM obtenido con las variables de RF-CV para carbonato cálcico total o equivalente. Curvas de respuesta para las variables suavizadas significativas.

De las 8 variables cuantitativas, solo resultan significativas ($p < 0.05$) tres de las climáticas: la precipitación total de abril (p-valor < 0.001) y, en menor medida, la temperatura mínima absoluta de diciembre y la media de julio (p-valor < 0.05). En el caso de la precipitación total de abril la relación no es lineal y se observa una tendencia decreciente en un primer tramo, mientras que a partir de cierto valor de temperatura mínima un incremento de esta produce un aumento en el contenido de $\text{CaCO}_3_{\text{total}}$ (figura 3.19). A pesar de que se esperaría que una mayor

precipitación produjera un lavado que eliminara el $\text{CaCO}_{3\text{total}}$ de la capa arable, lo cual coincide con el tramo inicial de la gráfica de la precipitación en abril, en el cual se encuentra la mayoría del rango de las muestras, el tramo final ascendente puede deberse a interacciones no incluidas en el modelo con procesos relacionados con la erosión y transporte de sedimentos cargados de carbonatos.

El efecto de la temperatura mínima de diciembre es similar, disminuyendo el contenido de $\text{CaCO}_{3\text{total}}$ conforme aumenta la temperatura hasta un determinado valor a partir del cual, el aumento de la temperatura mínima invernal produce un incremento en el contenido de $\text{CaCO}_{3\text{total}}$, aunque la tendencia general es bastante horizontal. Por su parte, la temperatura media del mes de julio muestra una relación lineal e inversa con el contenido de carbonato cálcico total en la capa arable, lo cual se opone a la hipótesis inicial de que en áreas más cálidas el porcentaje de este constituyente sería mayor debido a la mayor tasa de evaporación.

Otros estudios han encontrado diferencias en el contenido de $\text{CaCO}_{3\text{total}}$ en la capa arable relacionadas con la posición topográfica relativa, donde las zonas más bajas muestran mayor contenido de carbonatos (Gaultier *et al.*, 2006). Por el contrario, De Alba *et al.* (2004) encontró que los perfiles de zonas topográficamente más altas presentaban este constituyente a menor profundidad, mientras que en las zonas más bajas el horizonte cálcico (o con presencia de $\text{CaCO}_{3\text{total}}$, aunque no se cumplieran los requisitos para ser considerado horizonte cálcico) aparecía a mayor profundidad. En el estudio de Stavi *et al.* (2008) se encontraron diferencias significativas en los contenidos de carbonato cálcico total, dependiendo de la orientación de la ladera, siendo estos mayores en las laderas orientadas al sur que en las laderas orientadas al norte, donde se produciría un mayor lavado y una menor evaporación.

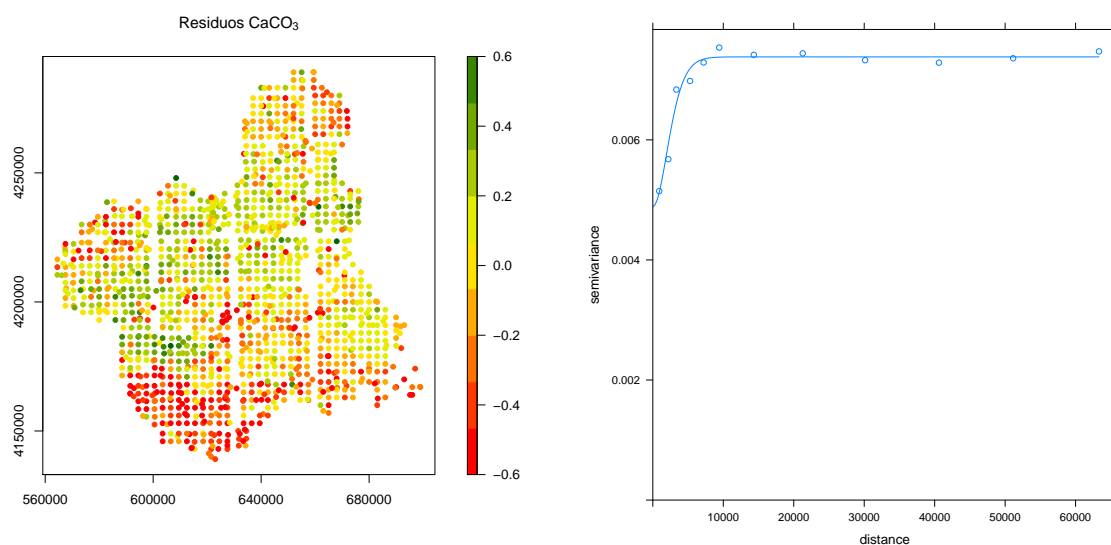
Aunque los dos últimos estudios mencionados también hacían referencia al uso del suelo, esta variable no aparece en el modelo obtenido. De Alba *et al.* (2004) estudió el efecto del laboreo en la redistribución de los constituyentes y la evolución de catenas de suelos, y Stavi *et al.* (2008) se centró en el efecto del ganado en la vegetación y algunos constituyentes del suelo entre los cuales se encontraba el $\text{CaCO}_{3\text{total}}$.

Cabe destacar también que en la literatura encontrada, casi todos los trabajos de modelización del contenido en $\text{CaCO}_{3\text{total}}$ se refiere a la profundidad a la que se encuentra el horizonte cálcico o petrocálcico, como el realizado por Bell *et al.* (1994). Otros utilizan imágenes obtenidas a partir de datos satelitales como variables predictoras de sales como los carbonatos en la capa arable (Leone *et al.*, 1995; Thwaites, 2006; Lagacherie *et al.*, 2008).

3.2.2.3 Residuos del modelo.

La figura 3.20A muestra la distribución de los residuos del modelo en la Región. En ella se aprecia que la mayor parte de los residuos tiene un valor cercano a 0, variando entre 0.6 y -0.6, y que en casi toda la región, el residuo (calculado restando los valores originales de la predicción) tiene valor negativo, es decir, se produce una infraestimación del contenido de $\text{CaCO}_{3\text{total}}$. Solo en algunas zonas puntuales (puntos verdes) el modelo predice valores mayores a los datos observados.

Respecto al semivariograma (figura 3.20B), ajustado a un modelo de Matern con parametrización de Stein, destaca el elevado valor del nugget (0.005) respecto a la meseta total (0.007), es decir, existe una elevada proporción de la variabilidad espacial de los residuos generada a una distancia inferior al intervalo de muestreo. El rango del semivariograma es de 3210 m. No obstante, la relación entre nugget y meseta (NSR; Nugget to Sill Ratio) de 66% indica una estructura espacial moderada (Cambardella *et al.*, 1994) por lo que se han interpolado los residuales utilizando un Krigeado Ordinario (KO).



(A) Distribución de los residuos del modelo RF-CV en la Región

(B) Semivariograma de los residuos

FIGURA 3.20: Distribución espacial de los residuos del modelo RF-CV para carbonato cálcico total o equivalente.

El mapa finalmente elaborado, añadiendo los residuos interpolados a la predicción (página A.4), da como resultado una correlación de 0.82 con los datos originales en el conjunto de validación, y un error $RECM_v$ de $0.12 \text{ kg} \cdot \text{kg}^{-1}$, mientras que el mapa utilizando solo la predicción (página A.3) resulta en una correlación de 0.61 y un error $RECM_v$ de $0.15 \text{ kg} \cdot \text{kg}^{-1}$. Sin embargo, la figura 3.21 muestra que las medianas en ambos mapas y en los valores originales son muy similares, y que solo existen diferencias entre los rangos de valores, más estrechos en el caso de los mapas. Esta semejanza se ha confirmado con el test de la t de Welch, que resulta en un p-valor de 0.32 entre los datos originales y el mapa con la predicción, y un p-valor de 0.60 al comparar los datos originales con el mapa final (RF+KO), por lo que no se detectan diferencias significativas.

En el mapa (página A.4) puede observarse un mayor contenido de carbonato cálcico total en el centro de la región, coincidiendo con las depresiones neógenas rellenas de margas, ricas en este constituyente, y sedimentos de las formaciones montañosas que las rodean, de naturaleza calcárea. Los valores ligeramente más bajos de esas sierras pueden atribuirse a que el carbonato forme parte de los horizontes subsuperficiales, mientras que el mapa se refiere a la capa arable, de donde el carbonato habría sido lavado. La costa sur de la región es donde el contenido en carbonatos es menor. Aunque el material original no fue incluido como variable predictora, este

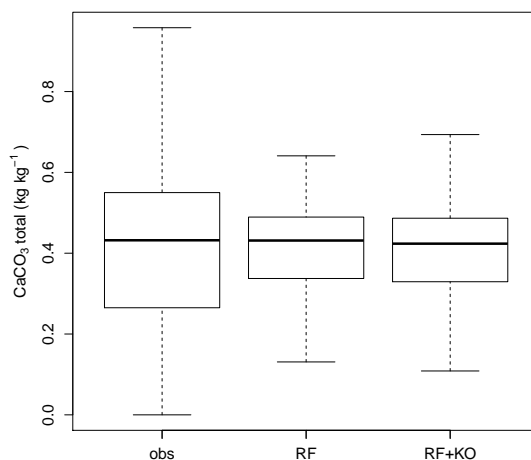


FIGURA 3.21: Gráfico de cajas con la distribución del contenido en carbonato cálcico total o equivalente a escala regional. **obs**: valores observados; **RF**: mapa de predicción del modelo RF-CV ; **RF+KO**: suma de la predicción de RF-CV y los residuos interpolados con KO.

hecho se podría atribuir a la composición de las sierras que rodean a este litoral, de naturaleza ígnea y metamórfica.

3.2.3 Arcilla.

3.2.3.1 Análisis descriptivo.

El contenido en arcilla en los suelos de la región varía entre 0 y $0.53 \text{ kg} \cdot \text{kg}^{-1}$, con un valor medio aproximado de $0.20 \text{ kg} \cdot \text{kg}^{-1}$ y una desviación estándar de $0.09 \text{ kg} \cdot \text{kg}^{-1}$ (tabla 3.8).

Mín.	Mediana	Media	Máx.	DS	CV(%)
0	0.189	0.197	0.530	0.09	44.3

TABLA 3.8: Estadísticos descriptivos del contenido de arcilla ($\text{kg} \cdot \text{kg}^{-1}$) en los suelos de la región. **Mín.**= Mínimo; **Máx.**= Máximo; **SD**= Desviación estándar; **CV (%)**. = Coeficiente de variación.

Se ha dividido la variable (expresada en porcentaje) entre 100, para expresarlo en tanto por uno (kg/kg). Tal y como se muestra en la figura 3.22, la distribución es acampanada, pero no cumple los requisitos para considerarse una distribución normal (Test de Shapiro, $W = 0.9748$, $p\text{-valor} = 1.018 \cdot 10^{-14}$), y no se ha realizado ninguna transformación. Los modelos obtenidos se han elaborado utilizando la distribución "binomial", considerando el contenido en arcilla como la probabilidad de que se encuentre arcilla en la muestra. De esta forma, se acota el resultado entre 0 (sin arcilla) y 1 (todo es arcilla), evitando la posibilidad de que salgan contenidos de arcilla negativos, o superiores a 1.

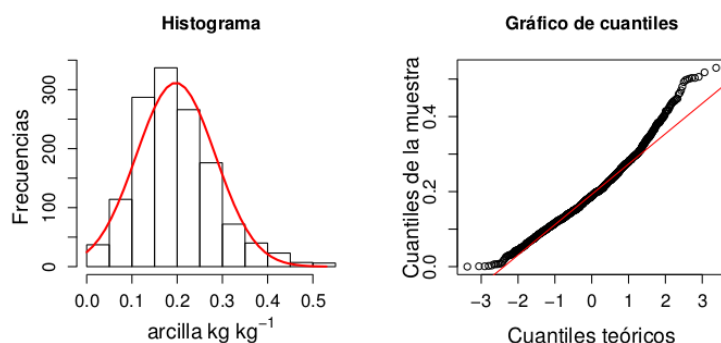


FIGURA 3.22: Histograma de frecuencias y gráfico de cuantiles de arcilla.

El test de Kruskal-Wallis indica que existen diferencias significativas en el contenido en arcilla entre los usos del suelo ($\chi^2 = 67.3663$, p-valor = $1.416 \cdot 10^{-12}$). Según la figura 3.23, se pueden diferenciar tres grupos al realizar el test de Wilcoxon por pares, entre los cuales habría diferencias significativas (p-valor < 0.05): por un lado, los suelos de cultivos de regadío (3), donde el contenido en arcilla es mayor que en el resto, lo cual se explicaría por el hecho de que este uso está asociado con los suelos de vega y los cercanos a cursos de agua, donde se reciben los aportes de los cursos fluviales; en segundo lugar, los suelos con cultivos de secano (3), con cultivos arbóreos (4), mosaicos de cultivos (5) y bosques (6), con contenidos de arcilla intermedios; finalmente, los suelos en suelos con matorral (7) y en suelos desnudos (8), el contenido de arcillas es menor, lo que se atribuiría a la menor meteorización que se ha producido en estos, asociados en a general suelos jóvenes con escaso desarrollo del perfil, pero también a sistemas dunares donde predominan partículas de mayor grosor.

En cuanto a los tipos de suelo, también existen diferencias significativas entre los grupos ($\chi^2 = 118.2878$, p-valor = $5.913 \cdot 10^{-16}$). También en este caso pueden diferenciarse tres grupos: los suelos que tienen mayor proporción de arcilla, que son los Fluvisoles Háplicos (401), los Leptosoles Réndzicos (805), los Solonchaks Háplicos (1205), los Calcisoles Háplicos asociados con Fluvisoles (2014), Gleysoles (2016) y Leptosoles (2018), los Fluvisoles Háplicos con Regosoles (40111), y los Regosoles Háplicos (1101). En el segundo grupo, estarían los suelos que tienen un contenido intermedio de arcilla, como los Calcisoles Háplicos (201) y Pétricos (205) y las asociaciones entre ellos (2012), los Gipsisoles Háplicos (501), los Leptosoles Líticos (803), los Fluvisoles asociados con Calcisoles (4012), los Leptosoles asociados con Calcisoles, tanto Lépticos (8032), como Réndzicos (8052), y los Regosoles Háplicos asociados con Calcisoles (11012). Finalmente, hay algunos grupos de suelos donde el contenido de arcilla es visiblemente menor, como los Regosoles Lépticos (1104), los Leptosoles Líticos asociados a Regosoles (80311), y los Regosoles Háplicos asociados a Gleysoles (11016). Estos últimos destacan porque, debido a las características que definen a los Gleysoles, se esperaría que tuvieran un mayor contenido de arcilla, tal y como ocurre en los Calcisoles Háplicos con Gleysoles.

Según la tabla 3.9, solamente tres variables destacan por tener una correlación significativa y notable con el contenido en arcilla. La primera y más importante, por su magnitud y significación,

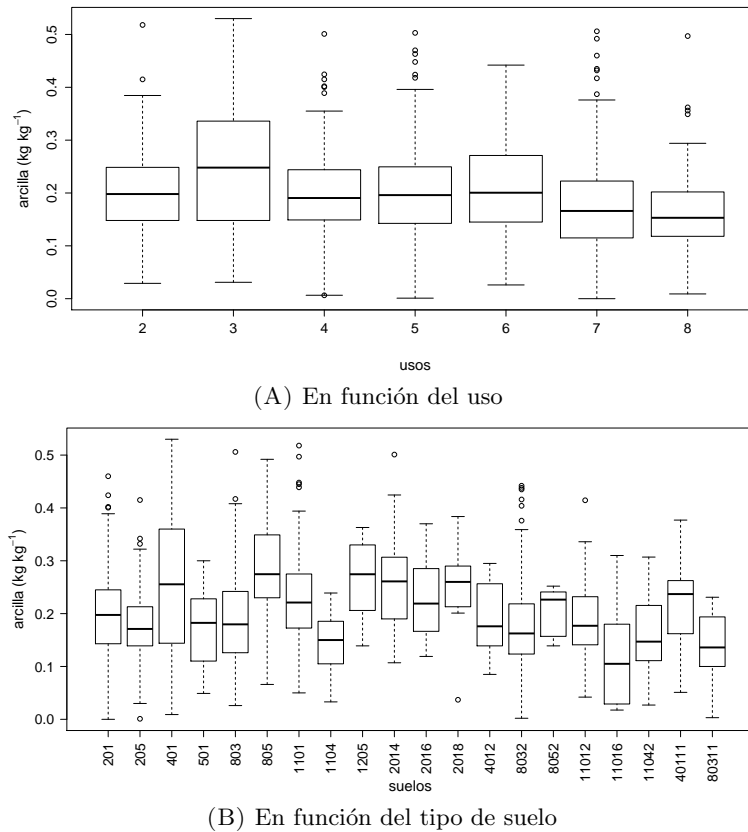


FIGURA 3.23: Análisis exploratorio: Contenido de arcillas en función del uso y tipo de suelo.

es el índice MRVBF, lo cual se explicaría por el aporte de sedimentos finos hasta mayores distancias por parte de los cursos de agua, que se acumularían en los valles más planos. También es altamente significativa la correlación con la pendiente (SLO25_39), que es además negativa, es decir, a mayor pendiente menor contenido de arcilla, lo que se fundamentaría en el hecho de que en las zonas con mayores pendientes se produciría más erosión de materiales, sobre todo los más finos. Finalmente, de las variables climáticas, hay una correlación importante con la precipitación mensual del mes de julio, donde además se producen altas temperaturas, por lo que esta variable representa las condiciones en que se favorece la meteorización más intensa, que da lugar a la formación de partículas más finas. El resto de correlaciones entre la proporción de arcilla en el suelo y las variables geomorfológicas y climáticas son de menor magnitud y significación.

3.2.3.2 Modelos.

En la figura 3.24 se muestra la evolución del $RECM_c$ al ir incrementando el número de variables, para cada uno de los modelos donde se utiliza la validación cruzada. En el caso de la arcilla, hay una gran superposición de todas las curvas, indicando que no hay una variación notable

Variables geomorfológicas			
Variable	r (Pearson)	Variable	r (Pearson)
SLO25_39	-0.2***	SIN25_3	0.031
SIN25_39	0.054*	COS25_3	-0.053
COS25_15	-0.093***	COS25_39	-0.07*
PROF25_3	0.023	PROF25_7	-0.011
PROF25_11	0.007	PROF25_23	0.013
PROF25_39	0.01	PLAN25_3	0.026
PLAN25_7	0.025	PLAN25_11	0.011
PLAN25_15	0.025	PLAN25_19	0.032
PLAN25_23	0.057*	PLAN25_27	0.069*
PLAN25_31	0.024	PLAN25_35	0.05
PLAN25_39	0.04	VRM11	-0.11***
VRM39	-0.10***	IWA25_27	0.007
IWA25_39	0.039	TPI25_3	-0.0009
MRVBF	0.29***	LS	-0.09***
Variables climáticas			
Variable	r (Pearson)	Variable	r (Pearson)
AGprectotPREDFIN4	0.028	AGprectotPREDFIN7	0.21*
AGprectotPREDFIN9	0.089***	AGprectotPREDFIN10	-0.021
AGtemedPREDFIN7	-0.011	AGtemaxabsPREDFIN8	-0.036
AGteminabsPREDFIN12	-0.055*	AGetHPREDFIN9	-0.15***
RAD25_6	0.055*	RAD25_12	-0.04

TABLA 3.9: Correlaciones del contenido en arcilla con las variables predictoras cuantitativas. Los valores de correlación significativos se han resaltado con fondo gris (P-valor: * < 0.05; ** < 0.01; *** < 0.001).

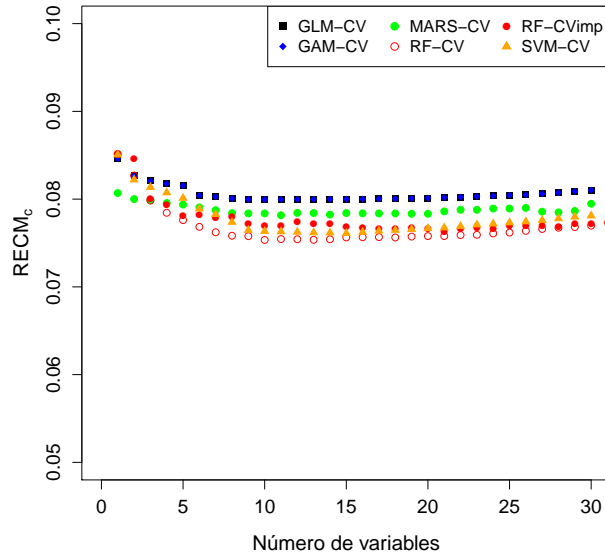


FIGURA 3.24: Raíz del error cuadrático medio ($RECM_c$) de los modelos a escala regional obtenidos con validación cruzada en bloques para arcilla, en función del número de variables.

del $RECM_c$ entre los diferentes modelos. Aunque todos ellos parten (con una sola variable) de un mismo valor de $RECM_c$, en los modelos con Random Forest (RF-CV y RF-CV_{IMP}) y

SVM-CV se aprecia una disminución del error al aumentar el número de variables predictoras en el modelo, mientras que en los modelos GLM-CV, GAM-CV y MARS-CV esta disminución es menos evidente.

Modelo		Mín.	Mediana	Media	Max.	SD	Nº de variables
GLM	r_v	0.14	0.33	0.33	0.51	0.06	9
	RECM _v	0.07	0.08	0.08	0.09	0.003	
GLM-CV	r_v	0.20	0.43	0.43	0.58	0.05	10
	RECM _v	0.07	0.08	0.08	0.09	0.003	
GAM	r_v	0.27	0.43	0.43	0.58	0.05	40
	RECM _v	0.06	0.08	0.08	0.09	0.003	
GAM-CV	r_v	0.24	0.42	0.42	0.58	0.05	9
	RECM _v	0.07	0.08	0.08	0.09	0.003	
MARS	r_v	0.20	0.39	0.39	0.57	0.06	18
	RECM _v	0.07	0.08	0.08	0.09	0.004	
MARS-CV	r_v	0.19	0.42	0.42	0.59	0.06	11
	RECM _v	0.07	0.08	0.08	0.09	0.003	
RF	r_v	0.36	0.54	0.54	0.69	0.05	40
	RECM _v	0.06	0.07	0.07	0.08	0.003	
RF-CV	r_v	0.32	0.51	0.51	0.64	0.05	10
	RECM _v	0.06	0.07	0.07	0.08	0.004	
RF-CV _{IMP}	r	0.33	0.54	0.54	0.69	0.06	21
	RECM _v	0.06	0.07	0.07	0.08	0.003	
SVM	r_v	0.30	0.47	0.47	0.6	0.05	40
	RECM _v	0.06	0.07	0.07	0.09	0.003	
SVM-CV	r_v	0.28	0.45	0.45	0.61	0.05	10
	RECM _v	0.06	0.08	0.08	0.09	0.003	

TABLA 3.10: Modelos a nivel regional obtenidos para arcilla ($\text{kg} \cdot \text{kg}^{-1}$). Estadísticos descriptivos de la correlación (r) y Raíz del Error Cuadrático Medio (RECM_v) obtenidos tras el remuestreo en el conjunto de validación

En la figura 3.24 se muestra la evolución del RECM_c al ir incrementando el número de variables, para cada uno de los modelos donde se utiliza la validación cruzada. En el caso de la arcilla, MARS-CV muestra el menor RECM_c con una y dos variables, pero al aumentar el número de predictores no se produce una disminución del error. Tampoco los modelos GLM-CV y GAM-CV experimentan una variación significativa del RECM_c y son los modelos que mayor error producen de todos los elaborados. Los modelos con Random Forest (RF-CV y RF-CV_{IMP}) y SVM-CV sí disminuyen el error RECM_c al aumentar la complejidad, siendo esta disminución más irregular en el caso de RF-CV_{IMP}, superponiéndose a SVM-CV, y quedando RF-CV por debajo del resto en la mayor parte del recorrido de la gráfica, aunque al alcanzarse las 21 variables no se diferencie de RF-CV_{IMP}) y SVM-CV.

El número de variables que finalmente se utiliza viene indicado en la tabla 3.10, donde se comparan los 11 modelos elaborados para predecir el contenido de arcilla, sirviendo de criterio de selección junto con los estadísticos descriptivos de la correlación (r_v) y el valor de RECM en el conjunto de validación (RECM_v), tras realizar el remuestreo. Las distribuciones de esos estadísticos descriptivos vienen además representadas por los correspondientes histogramas en la figura 3.25.

En el caso de la arcilla, las diferencias entre los diferentes modelos son más notables en la correlación (r_v ; figura 3.25A). Los modelos que utilizan Random Forest son los que obtienen mayores correlaciones, sobre todo RF (que utiliza las 40 variables) y RF-CV_{IMP} (que utiliza 21

variables). En cambio, el modelo RF-CV, que sólo utiliza 10 variables, resulta en una correlación ligeramente peor. Los modelos con SVM no obtienen correlaciones mejores que los modelos con Random Forest, e incluso esta es menor en el modelo con validación cruzada que en el modelo con las 40 variables. Por el contrario, con el modelo GLM, la correlación obtenida es la menor, seguido por MARS, aunque aumenta en ambos casos con el procedimiento de validación cruzada. Este aumento se produce con 10 variables en GLM-CV (frente las 9 variables que utiliza GLM), y con 11 variables en el caso de MARS-CV (frente las 18 de MARS). Los modelos con GAM obtienen resultados intermedios en cuanto a correlación, sin que se aprecien diferencias entre el modelo saturado y el modelo con validación cruzada, aunque el modelo GAM-CV solo utiliza 9 variables para llegar al mismo resultado.

Respecto al error ($RECM_v$; figura 3.25B), los valores obtenidos son muy similares en todos los modelos. Solamente cabe destacar que MARS es el modelo que mayor $RECM_v$ produce, mientras los menores valores de error se obtienen con los modelos que resultaban en una mayor correlación (RF y RF-CV_{IMP}).

Atendiendo a lo expuesto, se ha seleccionado el modelo RF-CV por considerarse el modelo que mejor resultado predictivo proporciona, tanto en relación a la exactitud como a la simplicidad. Este modelo explica un 28% de la varianza total de los datos, resultado similar al obtenido por [Ließ et al. \(2012\)](#); [Akpa et al. \(2014\)](#) y [Henderson et al. \(2005\)](#). La exactitud del modelo es también similar, en términos de $RECM$, a la obtenida por [Vaysse y Lagacherie \(2015\)](#) con el mismo método (Random Forest). Por otro lado, aunque la correlación es inferior a la conseguida por [Adhikari et al. \(2013a\)](#), se obtienen mejores resultados de $RECM_v$ (7% frente 13%).

La figura 3.26 muestra el ajuste entre los valores de arcilla predichos por el modelo, y los datos observados. El rango de valores que resulta del modelo es más estrecho que los valores originales de arcilla en el conjunto de validación, variando entre 0.11 y 0.37, aunque los valores de media y mediana son similares (0.197 y 0.186, respectivamente). Esta menor variabilidad se puede atribuir al efecto suavizador de los modelos predictivos ([Odeh et al., 1995](#)) y de forma adicional, al procedimiento que utiliza Random Forest de calcular un gran número de árboles y utilizar la predicción promediada de todos ellos ([Breiman, 2001a](#)).

La figura 3.27 muestra las variables incluidas en el modelo, en orden descendente de criterio de impureza del nodo. Las variables de mayor importancia son el índice MRVBF, el tipo del suelo y la pendiente (SLO25_39); a continuación, con menor criterio de impureza del nodo, aparecen las variables climáticas: la precipitación total del mes de julio (AGprectotPREDFIN7) y del mes de abril (AGprectotPREDFIN4), la radiación total de diciembre (RAD25_12) y la precipitación total de octubre (AGprectotPREDFIN10). Les siguen en importancia la curvatura plana (calculada con ventana de 11 celdillas, PLAN25_11) el índice de rugosidad de Iwahasi y Kamiya, calculado con ventana de 27 celdillas (IWA25_39), y finalmente los usos del suelo. Destaca por tanto en el modelo, las diferentes escalas de cálculo utilizadas para las variables geomorfológicas.

[Bui et al. \(2006\)](#) obtuvieron también modelos en los que el contenido de arcilla en la capa superficial se relacionaba con los tipos de suelo y con variables climáticas, tales como el índice de humedad mensual, la estacionalidad de las precipitaciones y la radiación mínima mensual, aunque en su caso los modelos eran árboles de regresión. Estas variables estarían relacionadas

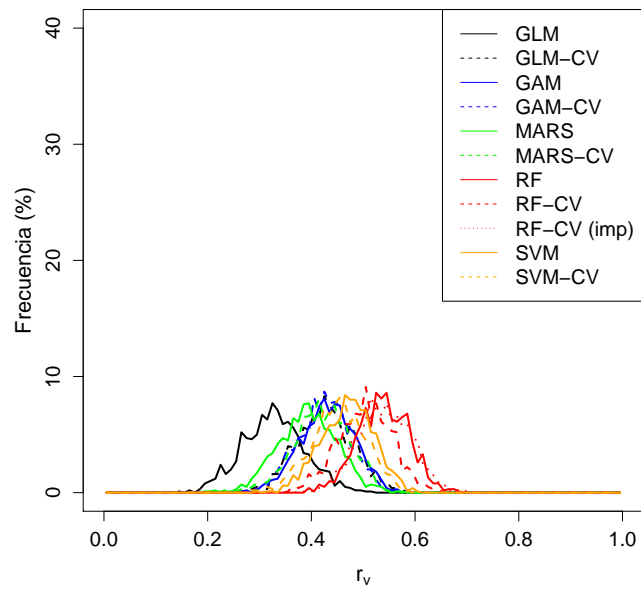
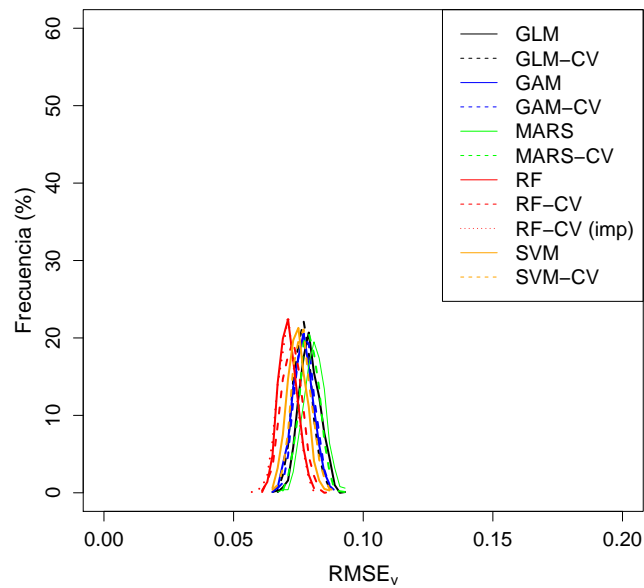
(A) Correlación (r_v)(B) Raíz del error cuadrático medio (RECM_v)

FIGURA 3.25: Modelos a escala regional obtenidos para arcilla. Histogramas de la correlación (r_v) y raíz del error cuadrático medio (RECM_v) tras el remuestreo en el conjunto de validación.

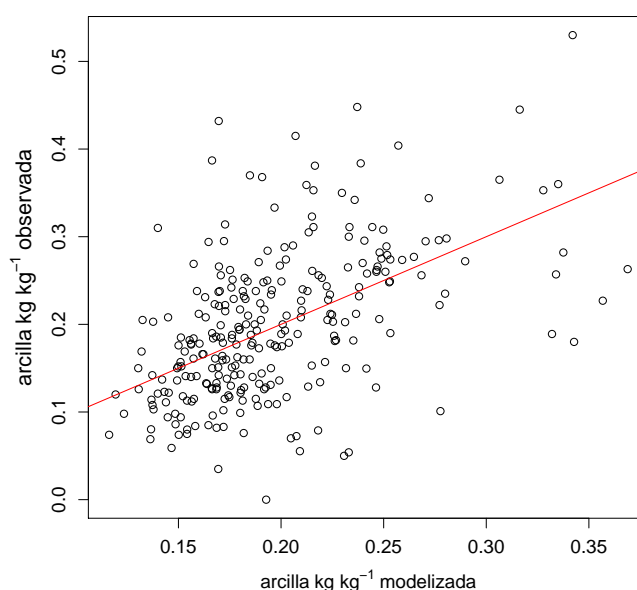


FIGURA 3.26: Diagrama de dispersión de los datos de arcilla ($\text{kg} \cdot \text{kg}^{-1}$) observados en el conjunto de validación frente los obtenidos con RF-CV a escala regional.

con la intensidad de la meteorización [Duchafour \(1988\)](#), destacando la importancia del agua en la intensidad de la misma. En nuestro caso, el índice de humedad vendría representado por la humedad relativa de septiembre, que se incluye en el modelo, a pesar de la correlación encontrada en el análisis descriptivo (tabla 3.9).

Por otro lado, [Adhikari et al. \(2013b\)](#) obtuvo reglas y árboles de regresión en los que el tipo de suelo era la variable de mayor importancia en la proporción de arcilla (considerando la importancia como el número de veces que la variable aparecía en las reglas y árboles) junto con el uso del suelo y variables geomorfológicas, que incluían el índice MRVBF, la pendiente, y el factor inclinación-longitud de la misma (LS). En nuestro modelo, sin embargo, el uso del suelo era la variable de menor importancia, y el factor LS ni siquiera está incluido, aunque también mostraba una correlación positiva con el contenido de arcilla.

La aparición del índice MRVBF como variable de mayor importancia coincide la hipótesis planteada anteriormente de que las fracciones finas se acumularían en los fondos de valles y zonas planas, tras ser erosionadas y transportadas desde zonas más altas ([Gessler et al., 2000](#)). Resultados similares fueron descritos por [Ließ et al. \(2012\)](#), que también utilizó RandomForest, donde la pendiente, la curvatura plana y la curvatura del perfil eran las variables que resultaban importantes en la predicción del contenido de arcilla.

En el caso de la curvatura plana (PLAN25_11), la proporción de arcilla disminuye en el suelo al pasar de zonas convexas a cóncavas, lo cual coincide con el resultado esperado, ya que los valores positivos de esta variable están relacionados con flujos divergentes. [Brown et al. \(2004\)](#) propuso

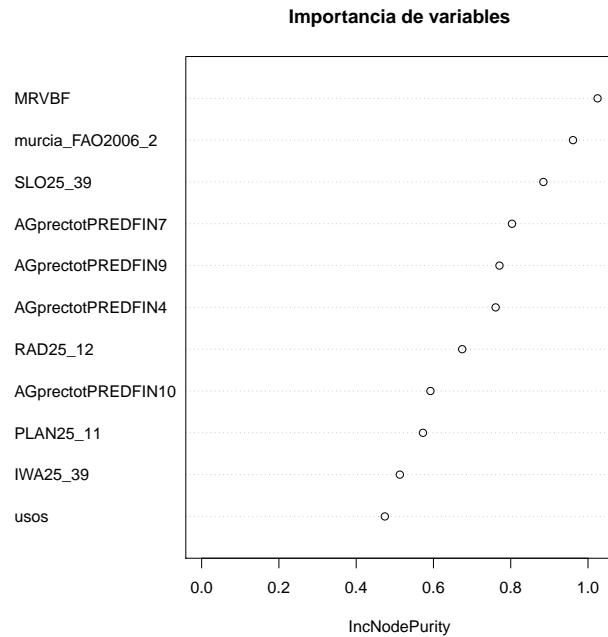


FIGURA 3.27: Importancia de variables en el modelo regional RF-CV para arcilla, expresada como criterio de impureza del nodo.

que la influencia de la curvatura sería dependiente de la escala, al observar una disminución del contenido de arcillas en zonas convergentes para suelos poco drenados a escala regional, mientras a escalas más finas este efecto sería contrario. También [Pérez Cutillas \(2013\)](#) encontró que la influencia de la curvatura en la proporción de arcilla dependía de la escala en modelos multiresolución, con un mayor efecto en las resoluciones de 200 y 400 m (que se corresponderían en nuestro trabajo con las ventanas de 9 a 17 píxeles): en la resolución de 400 m, el contenido de arcilla aumenta desde las formas convexas hasta las superficies planas, pasando a ser descendente en las formas cóncavas, mientras que en la resolución de 200 m se produce el efecto contrario.

Por su parte, el índice de Iwahasi y Kamiya (IWA25.39), aunque de escasa importancia en el modelo, reflejaría tanto la rugosidad como la posición relativa, de modo que valores altos indicarían relieves de textura gruesa, planos o poco accidentados y situados en zonas donde predominaría la deposición de materiales, tales como valles aluviales; por el contrario, los valores negativos indicarían relieves de textura fina, accidentados y con posiciones convexas donde predominaría la erosión y las partículas más gruesas (gravas y arenas).

Todas estas variables se han incluido en un nuevo GAM, con el fin de analizar las relaciones entre cada una de ellas y el contenido en arcilla.

Con el GAM elaborado utilizando las variables de RF-CV (GAM_{IMP}), se obtiene un valor medio de correlación de 0.44, ligeramente superior que el obtenido con GAM-CV, mientras el valor de RECM es ligeramente menor (0.08). Sin embargo, y tal como ocurre en el GAM-CV, ninguna de las variables incluidas resulta significativa (p -valor > 0.05). Ambos modelos tienen variables en

común, tales como el índice MRVBF, el tipo de suelo, los usos y la precipitación total del mes de julio; sin embargo, el modelo GAM-CV incluye otras variables climáticas (precipitación total de septiembre, temperatura media de julio), y geomorfométricas (curvatura plana calculada con 19 celdillas, y la orientación hacia el norte calculada con 15 y 39 celdillas).

Respecto a las variables cualitativas, los efectos de los distintos tipos de suelo y usos se solapan entre sí y no son significativos. En el caso de los usos, aún así, destaca la predicción de mayor cantidad de arcilla en el uso forestal (6), que en el análisis descriptivo aparecía con contenidos menores que los suelos de regadío (3), y el amplio rango de error en los suelos desnudos (8), aunque son los que menor cantidad de arcilla contienen. En el caso de los suelos, el rango de error de los Regosoles Háplicos con Gleysoles (11016) es el mayor de todos los tipos de suelo, aunque anteriormente se comprobó que era una de las unidades con menor proporción de arcilla; en el caso de los Leptosoles Líticos (803), también se describió una escasa proporción de arcilla en los datos de LUCDEME, mientras que la cantidad estimada de arcillas es alta en esta unidad según el modelo GAM_{IMP}, con un rango de error relativamente estrecho.

Por otro lado, aunque tampoco las variables cuantitativas tenían significación, destaca entre ellas el MRVBF que muestra una tendencia afín a la comentada, con un aumento del contenido de arcillas al aumentar el valor de esta variable. También es notable que las precipitaciones totales de septiembre, octubre y abril tienen una relación positiva con el contenido en arcilla, siendo este efecto más importante en el caso de AGprectotPREDFIN9, mientras que la precipitación de julio tiene un efecto contrario. En el caso de la curvatura plana (PLAN25_11), la relación es negativa, aunque el rango de error sea elevado; lo mismo ocurre en el modelo GAM-CV con la variable PLAN25_19. En el resto de variables la tendencia es prácticamente horizontal, o los rangos de error muy amplios.

3.2.3.3 Residuos del modelo.

En el caso de la arcilla, los residuos del modelo (figura 3.28) tienen valores que varían entre -0.26 y $0.19 \text{ kg} \cdot \text{kg}^{-1}$. Además, estos residuos son de signo positivo en la mayor parte de la Región, atendiendo a las tonalidades color verde claro de los puntos, es decir, la diferencia entre la predicción y el valor real es mayor de 0, a pesar de que como se dijo anteriormente, el modelo tendía a infraestimar el contenido de arcilla. Solo en unos pocos lugares (puntos naranjas) el modelo predice valores menores a los datos observados.

Respecto al semivariograma (figura 3.29A), ajustado automáticamente a un modelo de Matern con parametrización de Stein con la función `autofitVariogram`, destaca el elevado valor del nugget (0.006), y un escaso ajuste del semivariograma teórico con respecto al semivariograma experimental, que muestra una elevada semivarianza en los primeros pares de puntos, situados por encima de la gráfica del semivariograma teórico prácticamente horizontal, con un rango de 20147 m. Esta diferencia entre el semivariograma experimental y teórico no se daba en los residuos de CO y CaCO₃ estudiados anteriormente. A pesar de ello, la relación entre nugget y meseta (NSR; Nugget to Sill Ratio) es moderada, del 46% (Cambardella *et al.*, 1994) por lo que la interpolación se ha llevado a cabo con Krigado Ordinario.

Si, en cambio, teniendo en cuenta la gráfica del semivariograma experimental, se utiliza un mode-

lo de semivariograma esférico y se fija de entrada un rango de 7000 m (figura 3.29B), el valor de NSR disminuye al 28%, y también se reduce la suma de cuadrados de los residuos, de $1.25 \cdot 10^{-10}$ a $3.48 \cdot 10^{-12}$. La razón de que la función no seleccione este variograma automáticamente es que el modelo de Matern incluye más parámetros y por tanto, por lo general, produce un mayor ajuste a los datos.

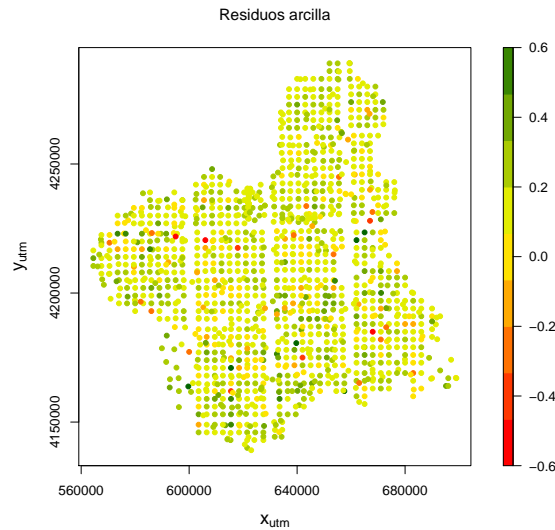
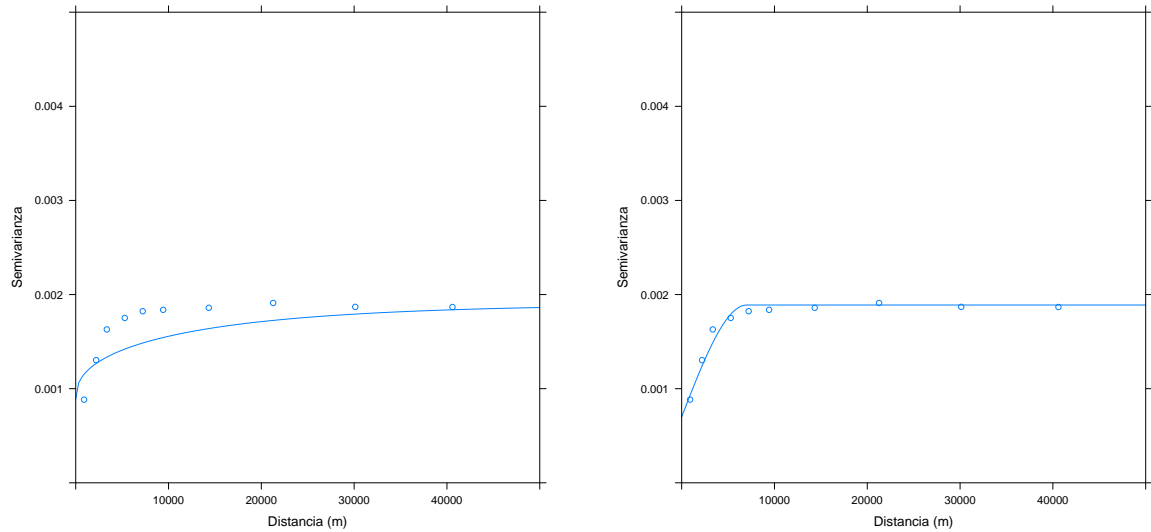


FIGURA 3.28: Distribución espacial de los residuos del modelo RF-CV_{IMP} para arcilla.

Aun teniendo en cuenta que no se ha realizado validación cruzada del proceso total de regresión-krigeado, el mapa finalmente elaborado tras añadir los residuos interpolados con KO a la predicción utilizando el semivariograma ajustado automáticamente (página A.6) da como resultado una correlación de 0.84 con los datos originales en el conjunto de validación, y un error RECM_v de $0.05 \text{ kg} \cdot \text{kg}^{-1}$. En cambio, utilizando como modelo para la interpolación el variograma esférico de la figura 3.29B (página A.7), se obtiene una correlación $r_v = 0.92$ y un RECM_v = $0.03 \text{ kg} \cdot \text{kg}^{-1}$.

Aunque estos resultados son mejores que los obtenidos sólo con el mapa de la predicción (página A.5, $r_v = 0.45$, RECM_v = $0.07 \text{ kg} \cdot \text{kg}^{-1}$), el test de Kruskal-Wallis realizado muestra que no hay diferencias significativas entre las predicciones (p-valor > 0.05), ya que en todos los casos el valor de la mediana es similar y los rangos se superponen entre sí, siendo la amplitud de estos la única diferencia. El test de la t de Welch, sin embargo, sí revela diferencias significativas entre las medias al comparar los datos observados con los predichos por el modelo RF (p-valor=0.001), siendo el valor medio en estos últimos algo menor. Lo mismo sucede al comparar las medias de los datos observados y los mapas de regresión-krigeado con semivariograma automático (RF+KOauto) y esférico (RF+KOesf), siendo los p-valores 0.003 y 0.005 respectivamente. No existe sin embargo diferencia entre los tres mapas elaborados.

El mapa finalmente seleccionado (RF+KOauto, página A.6) muestra que los mayores contenidos de arcilla se producen en las zonas bajas, como sería de esperar teniendo en cuenta el efecto de



(A) Semivariograma de Matérn generado automáticamente (B) Semivariograma esférico con rango $A=7000$ m

FIGURA 3.29: Semivariogramas de los residuos de RF-CV_{IMP} para arcilla.

la topografía, tal como se comentaba anteriormente. Destacan en este aspecto ciertas secciones de las vegas del Segura y del Guadalentín, y de la cuenca del Mar Menor, mientras que en el resto de las mismas el contenido de arcilla es intermedio. También hay contenidos intermedios en las depresiones cuaternarias que ocupan el centro de la región, y en el Altiplano se distinguen los cauces de ríos y ramblas. Los contenidos más bajos de arcilla se sitúan en las zonas más altas (Altiplano, Noroeste), así como en la zona suroeste (Mazarrón y Águilas). El mapa RF+KOesf (página A.7) muestra esta misma distribución, siendo la diferencia entre ambos la mayor presencia de "ruido" en este último.

3.2.4 Limo.

3.2.4.1 Análisis descriptivo.

El contenido de limo se ha expresado en $\text{kg} \cdot \text{kg}^{-1}$ (tanto por uno), dividiendo entre 100 el valor de la base de datos. A pesar del histograma mostrado en la figura 3.31, la distribución de esta variable no se puede considerar normal (test de Shapiro-Wilk, $W = 0.9792$, p-valor = $3.352 \cdot 10^{-13}$), a pesar de que el histograma es razonablemente simétrico y la mayor parte del gráfico de cuantiles se ajusta a la línea de la distribución, normal, por lo que no se ha realizado ninguna transformación. Los modelos GLM, GAM y el modelo GLM previo al MARS (con las variables cualitativas) se han elaborado utilizando la familia "binomial", con el fin de acotar la respuesta entre 0 (la muestra no contiene limo) y 1 (la muestra está formada enteramente por limo) sin que se generen contenidos de limo negativos o superiores a 1 en la predicción resultante. El contenido medio de limo en los suelos de la Región es de $0.43 \text{ kg} \cdot \text{kg}^{-1}$, oscilando en un rango

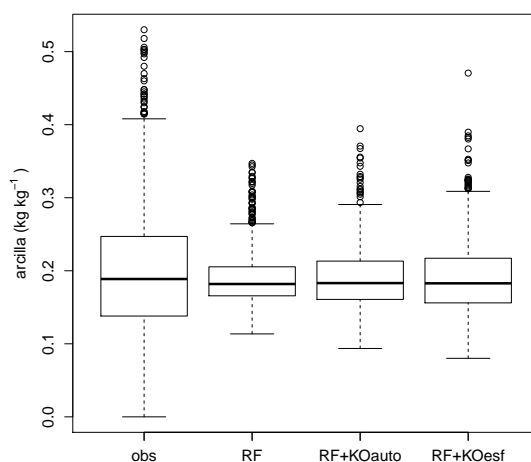


FIGURA 3.30: Gráfico de cajas con la distribución del contenido en arcilla a escala regional. **obs**: valores observados; **RF**: mapa de predicción del modelo RF-CV_{IMP}; **RF+KO**: suma de la predicción de RF-CV_{IMP} y los residuos interpolados con KO; **RF+splines**: suma de la predicción de RF-CV_{IMP} y los residuos interpolados con splines; **calc**: Valores calculados restando al total la suma de arena y arcilla.

entre 0 y 0.80 kg · kg⁻¹ (tabla 3.11), un rango considerablemente más amplio que el encontrado para las arcillas. Sin embargo, el coeficiente de variación es menor que en el caso de la arcilla, por lo que se puede decir que el contenido de limo es más homogéneo.

	Mín.	Mediana	Media	Máx.	DS	CV(%)
R	0.00	0.45	0.438	0.803	0.123	28.12

TABLA 3.11: Estadísticos descriptivos del contenido de limo (kg · kg⁻¹) en los suelos de la región. **Mín.**= Mínimo; **Máx.**= Máximo; **SD**= Desviación estándar; **CV (%)**. = Coeficiente de variación.

A pesar de que la figura 3.32A muestra contenidos de limo similares entre los usos del suelo, existen diferencias significativas entre ellos (test de Kruskal-Wallis, $\chi^2 = 23.4371$, p-valor = 0.0006). En concreto, los suelos de bosque (6) y los suelos con matorral (7) tienen menor proporción de limos que los suelos agrícolas, especialmente los suelos de regadío (2), los suelos de secano (3) y los suelos con cultivos en mosaico (5), que son los que mayor contenido de esta fracción tienen. Los suelos bajo cultivo arbóreo y los suelos desnudos (8) estarían en una posición intermedia, ya que solo muestran leves diferencias (p-valor < 0.05) con los suelos bajo mosaico de cultivo y los suelos con matorral.

Respecto a los tipos de suelo (figura 3.32B), también se observan diferencias significativas (test de Kruskal-Wallis, $\chi^2 = 143.1823$, p-valor < $2.2 \cdot 10^{-16}$). Resaltan en la figura 3.32 los Solonchaks Háplicos (1205) con mayor contenido de esta fracción, junto con los Regosoles Háplicos, como unidad (1101) o asociados con Gleysols (11016), y los Fluvisols Háplicos, solos (401) o asociados con Regosoles (40111). Salvo en el caso de los Regosoles Háplicos, cuyo contenido en limos se

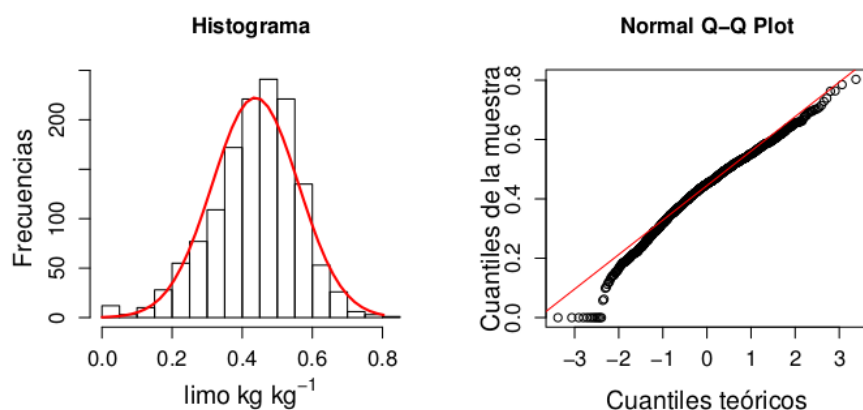
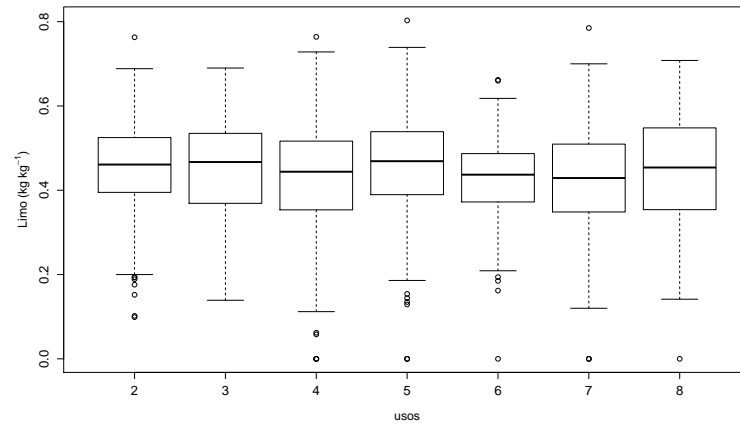
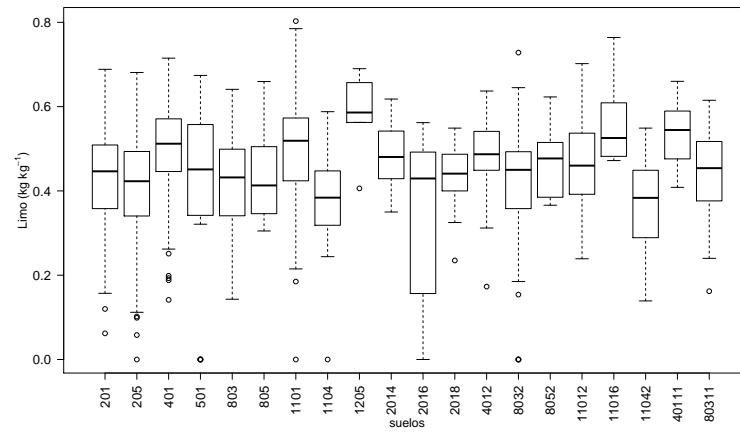


FIGURA 3.31: Histograma y gráfico de cuantiles de limo ($\text{kg} \cdot \text{kg}^{-1}$). Las líneas rojas señalan la distribución normal.

relacionaría con los materiales originales, las unidades mencionadas se relacionan con cierto nivel de hidromorfia y posiciones topográficas bajas (suelos de marismas, suelos de vega, suelos con patrones de color relacionados con periodos de inundación). Por el otro lado, los Regosoles Lépticos (1104), también cuando van asociados con Calcisoles (11042), tienen contenidos de limo menores que el resto de unidades, lo cual puede relacionarse con el escaso desarrollo del perfil, en el cual las partículas finas sean constantemente eliminadas por efecto de la erosión. Destaca también el caso de los Calcisoles Háplicos asociados con Gleysoles (2106) donde el rango de contenido en limos entre bajo y medio, es más amplio que en el resto de suelos, y pese a la presencia de los Gleysoles muestra diferencias significativas con los grupos mencionados anteriormente con contenidos altos en limo.



(A) En función del uso.



(B) En función del tipo de suelo.

FIGURA 3.32: Análisis exploratorio: Contenido de limo en función del uso y tipo de suelo.

Variables geomorfológicas			
Variable	r (Pearson)	Variable	r (Pearson)
SLO25_39	-0.1***	SIN25_3	0.07**
SIN25_39	0.08**	COS25_3	-0.02
COS25_15	-0.07**	COS25_39	-0.09***
PROF25_3	0.06*	PROF25_7	0.03
PROF25_11	0.03	PROF25_23	0.03
PROF25_39	0.003	PLAN25_3	-0.001
PLAN25_7	-0.01	PLAN25_11	-0.02
PLAN25_15	0.01	PLAN25_19	0.02
PLAN25_23	0.02	PLAN25_27	0
PLAN25_31	-0.01	PLAN25_35	0.02
PLAN25_39	0.03	VRM11	-0.02
VRM39	-0.03	IWA25_27	0.1***
IWA25_39	0.14***	TPI25_3	0.02
MRVBF	0.11***	LS	-0.01
Variables climáticas			
Variable	r (Pearson)	Variable	r (Pearson)
AGprectotPREDFIN4	0.02	AGprectotPREDFIN7	-0.22*
AGprectotPREDFIN9	0.01	AGprectotPREDFIN10	0.06*
AGtemedPREDFIN7	0.11***	AGtemaxabsPREDFIN8	0.13***
AGteminabsPREDFIN12	-0.08**	AGetHPREDFIN9	0.13***
RAD25_6	-0.03	RAD25_12	-0.02

TABLA 3.12: Correlaciones del contenido en limo ($\text{kg} \cdot \text{kg}^{-1}$) con las variables predictoras cuantitativas. Los valores de correlación significativos se han resaltado en negrita con fondo gris (p-valor: * < 0.05; ** < 0.01; *** < 0.001).

Respecto las correlaciones con las variables cuantitativas, mostradas en la tabla 3.12, éstas son generalmente de escasa magnitud, aunque tengan una alta significación (p-valor < 0.001), de forma análoga a lo que ocurría con la arcilla. Las correlaciones más fuertes se producen con variables climáticas: destaca la precipitación total del mes de julio, negativamente correlacionada con el contenido en limo, pero de menor significación que las que se producen con la temperatura máxima de agosto y la humedad relativa del mes de septiembre, y la temperatura media de julio. De las variables geomorfológicas, la mayor correlación se produce con el índice de Iwahasi y Kamiya calculado con el mayor tamaño de ventana (IWA25_39), seguido por el índice MRVBF, ambas de sentido positivo. Con el resto de variables las correlaciones son muy débiles, menores de 0.1, sean positivas o negativas.

3.2.4.2 Modelos.

La figura 3.33 muestra que, en el test de calibración, MARS-CV es el modelo en el que menos disminución de RECM_c se produce, estancándose el valor de este parámetro a partir de 4 variables y permaneciendo mayor que en el resto de los modelos a partir de ese punto. Como en el caso de la arcilla, las curvas de GLM-CV y GAM-CV se superponen y tampoco en estos modelos se produce una disminución significativa del error, aunque es menor que en el modelo MARS-CV. Los modelos RF-CV, RF-CV_{IMP} y SVM-CV sí muestran una disminución notable de RECM_c al añadir variables, siendo los modelos donde más se reduce este parámetro al aumentar la complejidad. Particularmente en RF-CV con una sola variable, el error es mayor que en el resto de modelos, seguido por RF-CV y SVM-CV con un RECM_c similar. A partir de la adición de la segunda variable, RF-CV y SVM-CV disminuyen el error de forma paralela y algo más pronunciada que RF-CV_{IMP} pero, al llegar a las 8 variables, la disminución del error se detiene. En el

caso de RF-CV, el valor de $RECM_c$ permanece aproximadamente constante, mientras que en SVM-CV aumenta hasta casi alcanzar el error obtenido con GLM-CV y GAM-CV. La evolución de $RF-CV_{IMP}$ es más irregular, pero con 30 variables su gráfica se superpone a la de RF-CV.

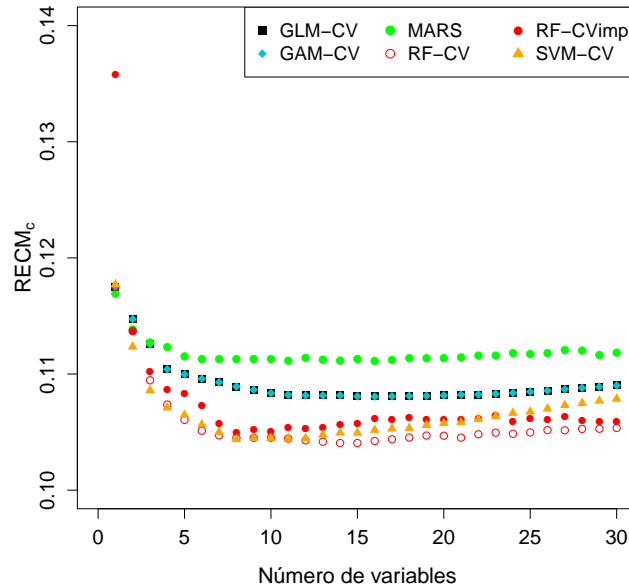


FIGURA 3.33: Raíz del error cuadrático medio ($RECM_c$) de los modelos a escala regional obtenidos con validación cruzada en bloques para limo, en función del número de variables.

Tal como se aprecia en la tabla 3.13 y la figura 3.34, la diferencia de exactitud entre los modelos obtenidos se refleja en la distribución de las correlaciones entre los datos observados y modelizados del conjunto de validación, y en el número de variables incluidas en cada modelo.

Las correlaciones obtenidas son similares en todos los modelos elaborados (figura 3.34A). Los modelos RF-CV y $RF-CV_{IMP}$ resultan en mayores correlaciones, siendo ambos prácticamente iguales en relación a este parámetro, que el resto de modelos. También se observa una mejora en la correlación respecto al modelo RF sin usar validación cruzada. En el caso de SVM-CV, la correlación obtenida es menor que con los modelos de Random Forest, a pesar de la mejora que también en este caso supone el procedimiento de validación cruzada. Esta mejora se observa también entre GLM-CV y GLM, aunque las correlaciones obtenidas sean de las más bajas de todos los modelos. Por el contrario, los modelos MARS y MARS-CV son los que menor correlación obtienen, y en ellos, al igual que entre los modelos GAM y GAM-CV, no se aprecia un incremento de la correlación al aplicar la validación cruzada.

La variación en el valor de $RECM_v$ es escasa, a excepción del modelo RF donde es considerablemente superior (figura 3.34B) a pesar de utilizar las 40 variables; además, a diferencia de lo esperado, el error en el conjunto de validación ($RECM_v$) es muy similar al obtenido en el conjunto de entrenamiento ($RECM_c$) en los modelos con validación cruzada. En este caso,

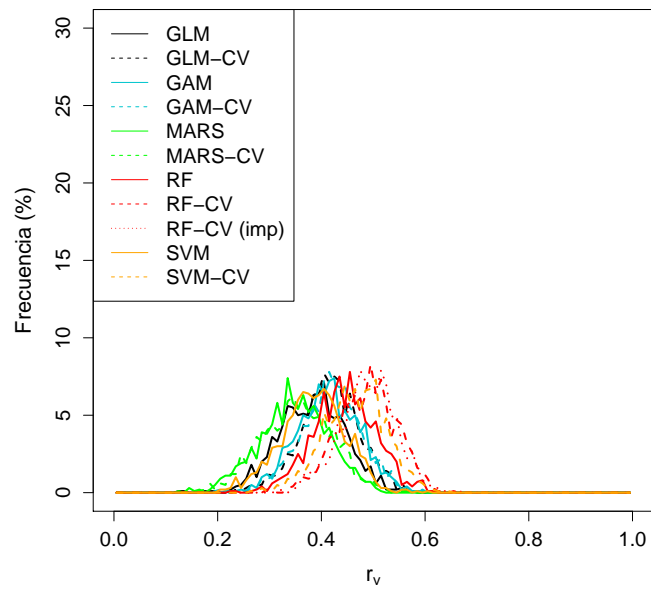
Modelo		Mín.	Mediana	Media	Max.	SD	Nº de variables
GLM	r_v	0.12	0.38	0.38	0.57	0.06	5
	RECM _v	0.09	0.12	0.12	0.14	0.007	
GLM-CV	r_v	0.23	0.42	0.42	0.59	0.06	15
	RECM _v	0.09	0.12	0.12	0.14	0.007	
GAM	r_v	0.21	0.41	0.41	0.59	0.06	40
	RECM _v	0.1	0.11	0.12	0.14	0.007	
GAM-CV	r_v	0.20	0.41	0.42	0.60	0.06	15
	RECM _v	0.09	0.12	0.02	0.14	0.005	
MARS	r_v	0.14	0.35	0.35	0.53	0.06	13
	RECM _v	0.10	0.12	0.12	0.16	0.008	
MARS-CV	r_v	0.16	0.36	0.36	0.56	0.06	13
	RECM _v	0.10	0.12	0.12	0.15	0.01	
RF	r_v	0.22	0.44	0.44	0.61	0.06	40
	RECM _v	0.09	0.11	0.11	0.14	0.007	
RF-CV	r_v	0.32	0.49	0.49	0.66	0.05	15
	RECM _v	0.09	0.11	0.11	0.13	0.006	
RF-CV _{IMP}	r_v	0.29	0.48	0.48	0.65	0.06	8
	RECM _v	0.09	0.11	0.11	0.14	0.008	
SVM	r_v	0.2	0.38	0.38	0.6	0.06	40
	RECM _v	0.1	0.12	0.12	0.14	0.007	
SVM-CV	r_v	0.27	0.47	0.47	0.64	0.05	8
	RECM _v	0.09	0.11	0.11	0.13	0.007	

TABLA 3.13: Modelos a nivel regional obtenidos para limo ($\text{kg} \cdot \text{kg}^{-1}$). Estadísticos descriptivos de la correlación (r_v) y Raíz del Error Cuadrático Medio (RECM_v) obtenidos tras el remuestreo en el conjunto de validación

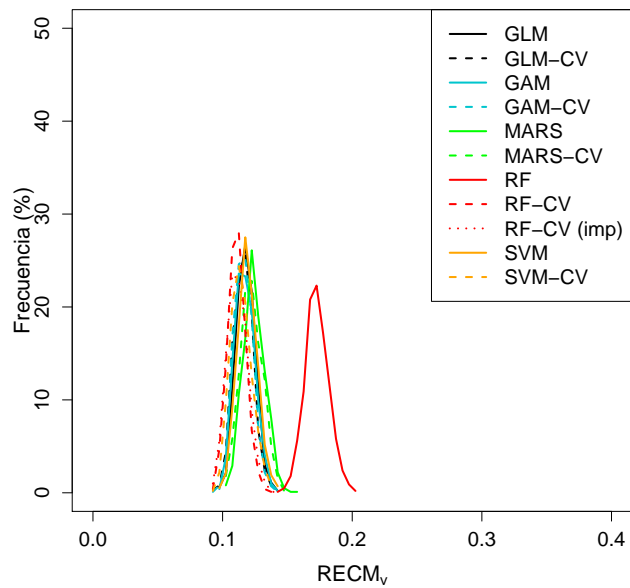
los modelos RF-CV y RF-CV_{IMP} son los que menor (RECM_v) obtienen, aunque en el caso de RF-CV_{IMP} solamente hacen falta 8 variables predictoras, frente las 15 que incluye RF-CV. SVM-CV también supone una reducción del número de variables frente SVM, pero no del RECM_v, al igual que ocurre con GAM y GAM-CV. En el caso de los modelos MARS incluyen ambos 13 variables, y también el RECM_v obtenido es similar entre ellos. Finalmente, en el caso de GLM y GLM-CV, el modelo con validación cruzada no supone una mejora frente la regresión por pasos, ya que GLM-CV precisa mayor número de variables sin reducir el RECM_v.

Por todo ello, el modelo finalmente seleccionado para la elaboración del mapa de limo es RF – CV_{IMP}, que explica un 25.89% de la varianza total en el conjunto de calibración. Este valor es similar al obtenido por Akpa *et al.* (2014) en suelos de Nigeria con el mismo método. La figura 3.35A refleja la dispersión entre los datos observados en el conjunto de validación y los obtenidos con el modelo; se observa que, aunque los valores de media y mediana son muy similares, el rango en los valores modelizados es más estrecho, variando entre 0.22 y 0.61 $\text{kg} \cdot \text{kg}^{-1}$, mientras que en los valores reales el contenido de limo varía entre 0 y 0.78 $\text{kg} \cdot \text{kg}^{-1}$.

De las variables que incluye el modelo seleccionado, solamente aparece una variable cualitativa (murcia_FAO2006_2), siendo el resto variables climáticas (figura 3.36). La variable con mayor importancia es la precipitación mensual de julio (AGprectotPREDFIN7), seguida por el tipo de suelo (murcia_FAO2006_2). Les siguen en importancia la precipitación total de septiembre (AGprectotPREDFIN9) y la oscilación térmica, representada por la temperaturas mínima absoluta de diciembre (AGteminabsPREDFIN12) y la temperatura máxima absoluta de agosto (AGtemaxabsPREDFIN8). A continuación aparecen la temperatura media de julio (AGtemedPREDFIN7), la precipitación total de abril (AGprectotPREDFIN4) y la humedad relativa del

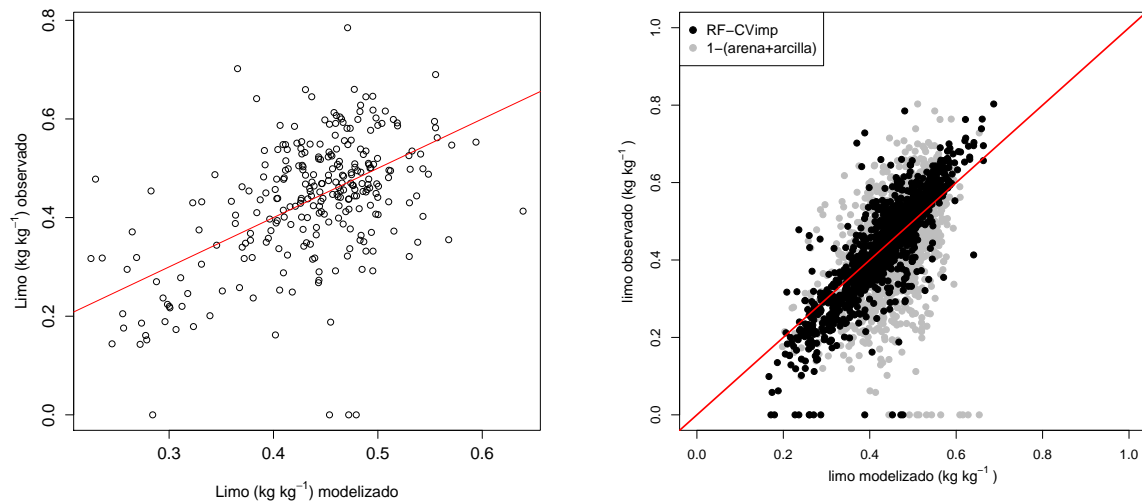


(A) Correlación



(B) Raíz del error cuadrático medio

FIGURA 3.34: Modelos a escala regional obtenidos para limo. Correlación (r_v) y raíz del error cuadrático medio ($RECM_v$) tras el remuestreo en el conjunto de validación.



(A) Ajuste del modelo en el conjunto de validación (B) Comparación entre modelo y estimación en el conjunto de datos

FIGURA 3.35: Diagrama de dispersión de los datos de limo (kg/kg) observados frente los obtenidos con $\text{RF-CV}_{\text{IMP}}$ a escala regional en el conjunto de validación y comparación con el porcentaje estimado frente las otras fracciones.

mes de septiembre (AGetHPREDFIN9).

Aunque la mayoría de ellas están correlacionadas de forma significativa con el contenido en limo, destaca la ausencia de variables geomorfométricas, como el índice MRVBF, y el índice de Iwahasi calculado con las mayores escalas (IWA25_27 e IWA25_39), que también están correlacionados con el porcentaje de limo y que en otros estudios sí aparecen como variables significativas en los modelos de distribución de este componente.

Por otro lado, no obstante, pocos de esos estudios modelizan la distribución de limos, sino que calculan el porcentaje restándolo de la proporción de arenas y arcillas (Akpa *et al.*, 2014; Ließ *et al.*, 2012). En este trabajo, hemos comparado las dos aproximaciones, realizando los mapas correspondientes. Según muestra la figura 3.35B, aunque el resultado en ambos casos es similar, la correlación en el total de los datos es mejor en el caso del limo modelizado con $\text{RF-CV}_{\text{IMP}}$ ($r = 0.85$; $\text{RECM}_v = 0.06$) que con el estimado restando el porcentaje de las otras fracciones ($r_v = 0.43$; $\text{RECM}_v = 0.11$).

El GAM elaborado con las variables de mayor importancia en $\text{RF-CV}_{\text{IMP}}$, (GAM_{IMP}) ofrece menor exactitud que el modelo GAM-CV obtenido anteriormente, ya que si bien los estadísticos significativos de la correlación tras el remuestreo son similares entre ambos, el error (RECM_v) del modelo GAM_{IMP} es mucho mayor, variando entre 0.67 y 0.71.

Por otro lado, ninguno de los tipos de suelo es estadísticamente significativo en el nuevo modelo GAM, debido a los amplios rangos de error que todos ellos muestran, sobre todo los Solonchaks Háplicos, que eran los que mayor contenido de limos tenían, tal como se comentó en el análisis descriptivo.

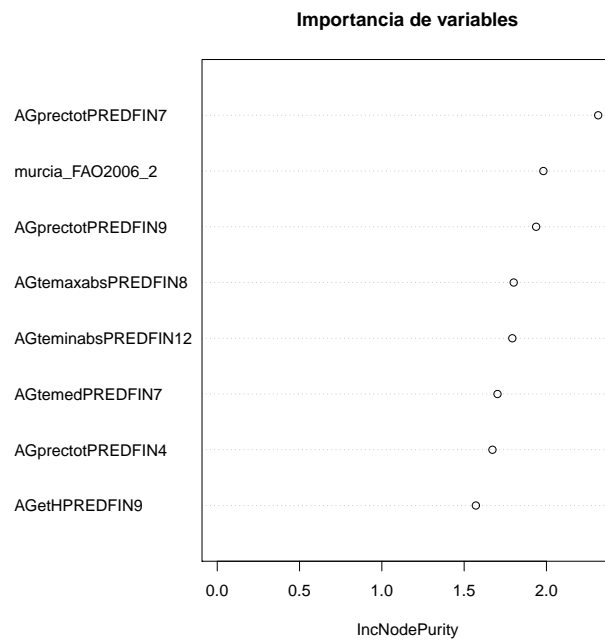


FIGURA 3.36: Importancia de variables en RF-CV_{IMP} obtenido para limo, expresada como criterio de impureza del nodo.

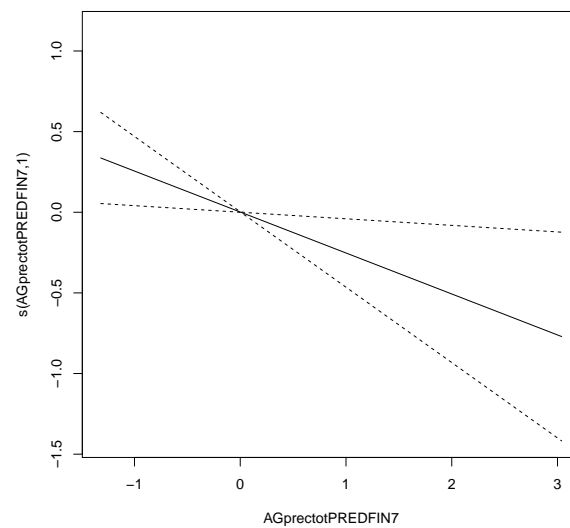


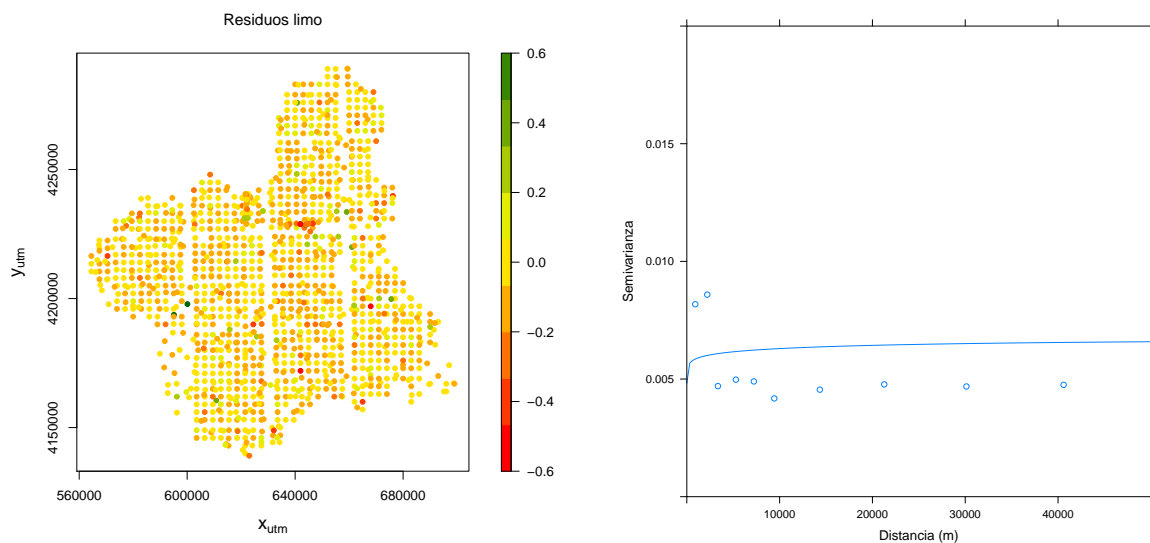
FIGURA 3.37: Modelo GAM obtenido con las variables de RF-CV_{IMP} para limo. Curva de respuesta para la variable suavizada significativa

De las variables suavizadas, solamente resulta significativa la precipitación total de julio, ya que los intervalos de error del resto de variables no permiten descartar la horizontalidad, es decir, la nula influencia de la variable en el contenido de limo. En la figura 3.37 puede observarse una relación negativa entre la precipitación de julio (AGprectotPREDFIN7). La influencia de estas variables se relacionaría con procesos de erosión y transporte de materiales, ya que en primavera y otoño es cuando suelen producirse la mayor parte de las precipitaciones en la región, de carácter generalmente torrencial. Esta hipótesis estaría apoyada con la relación también directa entre el contenido de limo y la precipitación de abril (AGprectotPREDFIN4) aunque esa variable no resulte significativa.

En el modelo GAM-CV, sin embargo, si resultan significativos algunos tipos de suelos, y determinadas variables geomorfológicas, además de variables climáticas. Estas variables (curvatura del perfil, componente norte de la orientación, y el índice MRVBF) coinciden con las variables que en otros estudios han mostrado tener influencia en la distribución de partículas.

3.2.4.3 Residuos del modelo.

Los residuos del modelo, representados en la figura 3.38A, muestran en su mayoría valores cercanos a 0, variando entre 0.42 y -0.5. Este residuo, calculado restando los valores originales de la predicción, tiene en la mayor parte de la región un valor negativo, es decir, el modelo produce una infraestimación del contenido de limo excepto en unos pocos lugares (puntos verdes) donde el modelo predice valores mayores a los datos observados.



(A) Distribución de los residuos del modelo RF-CV_{IMP} para en la Región

(B) Semivariograma de los residuos

FIGURA 3.38: Distribución espacial de los residuos del modelo RF-CV_{IMP} para limo.

Respecto el semivariograma (figura 3.38B), ajustado a un modelo de Matern con parametrización de Stein, destaca el elevado valor del nugget (0.006), y una elevada semivarianza en los pares

de puntos más cercanos, superior a la encontrada en puntos más alejados. Como consecuencia, y debido al algoritmo utilizado por la función para el cálculo del semivariograma teórico, el ajuste de las muestras es escaso en el gráfico, y el semivariograma teórico muestra una tendencia prácticamente horizontal, con un rango de 20147 m, tratándose de una mala estructura espacial que no justifica la interpolación utilizando kriging ordinario. Sin embargo, la relación entre nugget y meseta (NSR; Nugget to Sill Ratio) es moderada-débil, del 69% (Cambardella *et al.*, 1994). Debido a esto, se ha realizado la interpolación mediante Kriging Ordinario, además de con splines.

Teniendo en cuenta que al final del proceso (regresión-interpolación) no se ha realizado una validación cruzada, la comparación entre los valores observados en el conjunto de validación y el mapa obtenido sumando la predicción y el mapa interpolado con KO (página A.9) resulta en una correlación r_v de 0.57 y un valor de $RECM_v$ de $0.10 \text{ kg} \cdot \text{kg}^{-1}$, al igual que el mapa resultante de restar al total la proporción de arcilla y arena (página A.11). Ambos resultados son similares al obtenido solamente con el mapa de predicción (página A.8, $r_v = 0.57$; $RECM_v=0.11 \text{ kg} \cdot \text{kg}^{-1}$). Haciendo la interpolación con splines la correlación r_v es de 0.81 y el valor de $RECM_v$ es de $0.07 \text{ kg} \cdot \text{kg}^{-1}$, aunque el aspecto del mapa (página A.10) refleja la débil estructura espacial de los residuos.

El gráfico de cajas realizado (3.39) muestra que los rangos de los valores observados y los obtenidos con los diferentes mapas se solapan. Por otro lado, al realizar el test de Wilcoxon para comparar los rangos entre los datos observados y los diferentes mapas, las diferencias no eran significativas en el caso del mapa estimado (calc; p-valor=0.055), pero sí respecto el mapa de predicción (RF; p-valor=0.004), y los mapas de predicción y residuos interpolados con KO (RF+KO; p-valor=0.005) y con splines (RF+splines; p-valor=0.02). Estas diferencias se deben a que este test compara las medianas, que son mayores en los datos observados que en los mapas, excepto en el estimado. Por otro lado, el test de la t de Welch, que compara medias, solo encuentra diferencias significativas entre los valores medidos y los estimados (calc), que son mayores en este mapa.

Finalmente, el mapa seleccionado para limos ha sido el que resulta de restar al total el mapa de arcillas y arenas (página A.11), ya que a pesar de la diferencia en los valores medios mencionada, el mapa obtenido da lugar a que la suma de todas las fracciones sea igual a la unidad, lo cual no ocurre si se selecciona cualquiera de los otros.

El mapa finalmente seleccionado es el que estima el contenido de limo a partir del resto de fracciones. En él, se observan mayores proporciones de limo en las zonas topográficamente más bajas, como algunos puntos de las vegas del Segura y el Guadalentín, y algunas secciones del centro de la región ocupadas por cuencas neógenas, aunque de forma menos clara que en el mapa de arcillas. También en el límite norte de la cuenca del Mar Menor se aprecian zonas con elevados contenidos de limo, que coincidirían con ramblas y barrancos. En la mayor parte de la región, el contenido de limo sería intermedio, y solamente se aprecian proporciones bajas de esta fracción en un sector del Altiplano, al suroeste en la zona de Águilas y Mazarrón y otras zonas de pequeña extensión.

Además de todo lo dicho anteriormente, se debe añadir que es la fracción granulométrica en la que peor resultado predictivo se obtiene.

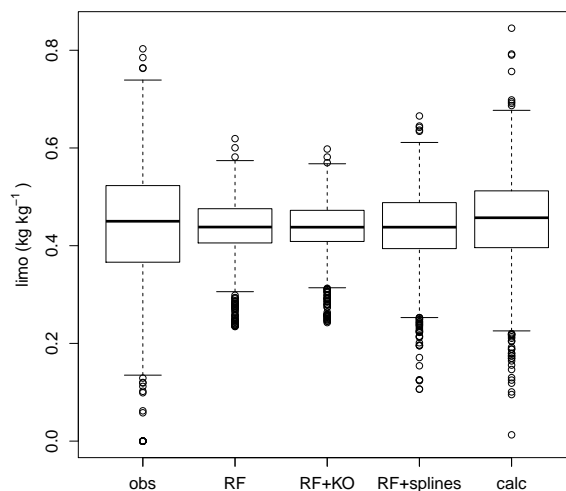


FIGURA 3.39: Gráfico de cajas con la distribución del contenido en limo a escala regional. **obs**: valores observados; **RF**: mapa de predicción del modelo RF-CV_{IMP} ; **RF+KO**: suma de la predicción de RF-CV_{IMP} y los residuos interpolados con KO; **RF+splines**: suma de la predicción de RF-CV_{IMP} y los residuos interpolados con splines; **calc**: Valores calculados restando al total la suma de arena y arcilla.

3.2.5 Arena.

3.2.5.1 Análisis descriptivo.

El contenido de arena se ha expresado en $\text{kg} \cdot \text{kg}^{-1}$ (tanto por uno), dividiendo entre 100 el valor de la base de datos. La distribución de esta variable se ajusta a una campana, pero no se puede considerar normal (test de Shapiro-Wilk, $W = 0.9932$, $p\text{-valor} = 6.286 \cdot 10^{-6}$). Como en las fracciones granulométricas estudiadas anteriormente, los modelos GLM, GAM y el modelo GLM con las variables cualitativas previo a los modelos MARS se han elaborado utilizando la distribución binomial.

La tabla 3.14 muestra los estadísticos descriptivos del contenido en arena en los suelos de la Región de Murcia. El contenido medio de arena es de $0.35 \text{ kg} \cdot \text{kg}^{-1}$, oscilando en un rango entre 0 y $0.88 \text{ kg} \cdot \text{kg}^{-1}$, similar al encontrado para el limo, aunque en este caso, el coeficiente de variación es muy superior, de lo que se deduce que la distribución de la variable es más heterogénea, como sucedía con la arcilla.

Mín.	Mediana	Media	Máx.	DS	CV(%)
0.00	0.35	0.36	0.88	0.16	47.12

TABLA 3.14: Estadísticos descriptivos del contenido de arena ($\text{kg} \cdot \text{kg}^{-1}$) en los suelos de la región. **Mín.**= Mínimo; **Máx.**= Máximo; **SD**= Desviación estándar; **CV**o (%). = Coeficiente de variación.

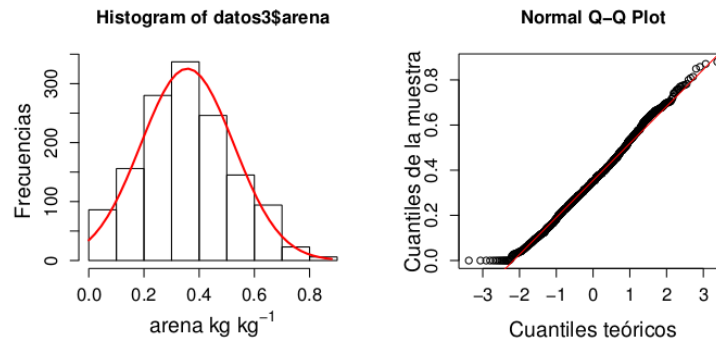
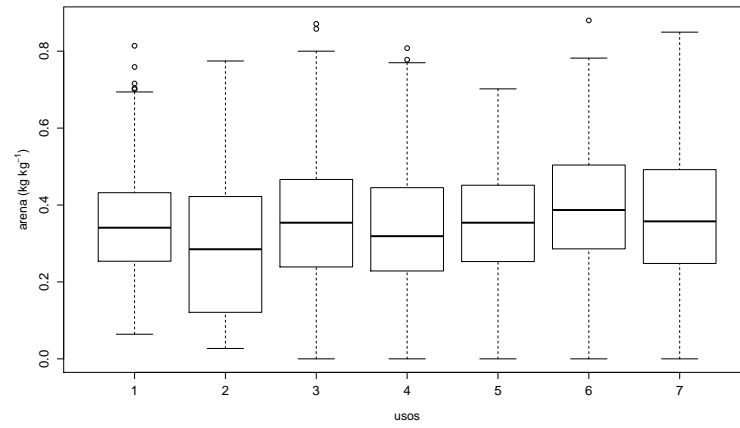


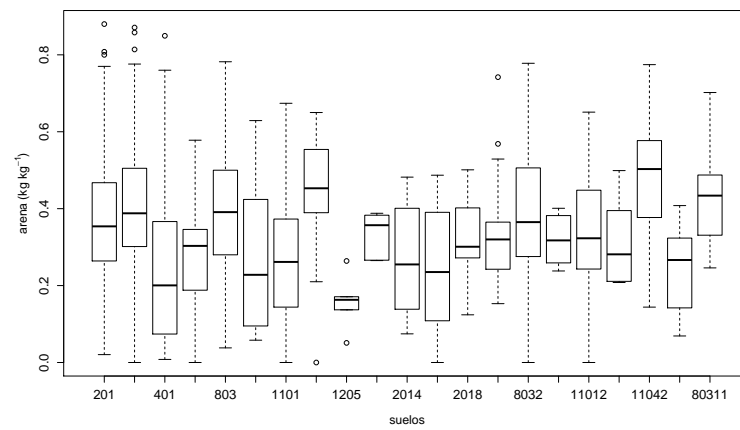
FIGURA 3.40: Histograma y gráfico de cuantiles de arena ($\text{kg} \cdot \text{kg}^{-1}$) en los suelos de la región. Las líneas rojas indican la distribución normal.

A pesar de que la figura 3.41 muestra contenidos de arena similares en todos los usos del suelo, existen diferencias significativas entre ellos (test de Kruskal-Wallis, $\chi^2 = 36.8544$, p-valor = $1.88 \cdot 10^{-6}$). El test de Wilcoxon realizado entre pares muestra que en los cultivos de regadío (3) el contenido en arena es significativamente menor (p-valor < 0.05) que en el resto de usos, mientras que en los suelos con matorral (7) el contenido en esta fracción era mayor que en los demás, excepto los suelos desnudos (8). Los demás tendrían contenidos de arena intermedios, siendo de ellos los suelos desnudos los únicos que no muestran diferencias significativas con los suelos cubiertos de matorral.

Esta distribución de las arenas en función de los usos puede atribuirse a que la mayor parte de regadío están en suelos de vega, situados en zonas de menor altitud, donde son más abundantes los sedimentos finos; por otro lado, en los suelos con matorral, la mayor proporción de arena se atribuye a su localización en las laderas y cabeceras de rambla, donde los materiales más finos serían erosionados y arrastrados a posiciones más bajas.



(A) En función del uso



(B) En función del tipo de suelo

FIGURA 3.41: Análisis exploratorio: Contenido de limo en función del uso y tipo de suelo.

En cuanto a los tipos de suelo, también existen diferencias significativas (test de Kruskal, $\chi^2 = 181.1121$, p-valor $< 2.2 \cdot 10^{-16}$), destacando por su bajo contenido de arenas los Solonchaks Háplicos (1205), en los cuales el rango es además estrecho, además de los Fluvisoles Háplicos (401) y la asociación de estos con Regosoles Háplicos (40111), aunque con rangos más amplios en el contenido de esta fracción.

Esto concuerda con el hecho de que estos suelos suelen aparecer en zonas topográficamente poco elevadas, y cursos de ríos, tal y como se comentó con los suelos de regadío en el párrafo anterior; por otro lado, la escasa cantidad de arenas en los Solonchaks se correspondería con el hecho de que en la región, pocos de estos suelos se corresponden con arenales costeros, sino con saladares de interior como los del Guadalentín y los de Abanilla, donde predomina la sedimentación de materiales finos.

En sentido contrario, los Regosoles Lépticos (1104), solos o asociados con Calcisoles (11042), y los Leptosoles Líticos con Regosoles (80311) tienen mayor proporción de arena, teniendo todos ellos, excepto los Calcisoles, la característica común de ser suelos jóvenes o con poca diferenciación de perfiles, debido al constante rejuvenecimiento por procesos de erosión. También los Calcisoles Háplicos (201) y Pétricos (205) tienen un contenido de arenas relativamente alto, que en este caso puede atribuirse asimismo a las posiciones topográficas que ocupan, mientras que el resto de suelos no muestran diferencias significativas en el contenido de esa fracción.

Variables geomorfológicas			
Variable	r (Pearson)	Variable	r (Pearson)
SLO25_39	0.16***	SIN25_3	-0.05*
SIN25_39	-0.09**	COS25_3	0.05
COS25_15	0.08**	COS25_39	0.09**
PROF25_3	-0.05*	PROF25_7	-0.01
PROF25_11	-0.009	PROF25_23	0.005
PROF25_39	0.001	PLAN25_3	-0.003
PLAN25_7	-0.008	PLAN25_11	0.007
PLAN25_15	-0.02	PLAN25_19	-0.02
PLAN25_23	-0.03	PLAN25_27	-0.02
PLAN25_31	-0.005	PLAN25_35	-0.03
PLAN25_39	-0.04	VRM11	0.07
VRM39	0.06*	IWA25_27	-0.09**
IWA25_39	-0.13***	TPI25_3	-0.01
MRVBF	-0.21***	LS	0.06*
Variables climáticas			
Variable	r (Pearson)	Variable	r (Pearson)
AGprectotPREDFIN4	-0.03	AGprectotPREDFIN7	0.21***
AGprectotPREDFIN9	-0.03	AGprectotPREDFIN10	-0.04
AGtemedPREDFIN7	-0.1***	AGtemaxabsPREDFIN8	-0.14***
AGteminabsPREDFIN12	0.05	AGetHPREDFIN9	-0.15***
RAD25_6	-0.011	RAD25_12	0.033

TABLA 3.15: Correlaciones del contenido en arena ($\text{kg} \cdot \text{kg}^{-1}$) con las variables predictoras cuantitativas. Los valores de correlación significativos se han resaltado en negra con fondo gris (p-valor: * < 0.05 ; ** < 0.01 ; *** < 0.001).

En relación a las variables cuantitativas, se han detectado correlaciones significativas entre éstas y el contenido de arena (Tabla 3.15). Destacan, de las variables geomorfológicas, la correlación negativa con el índice MRVBF, y con menores valores de correlación, la pendiente, con signo positivo. De las variables relacionadas con la rugosidad, se ha hallado una correlación positiva

con la medida vectorial VRM39, y negativa con el índice de Iwahasi a dos distintas escalas, siendo más fuerte la correlación con IWA25_39. Además, el contenido en arena estaría correlacionado con orientación de la misma respecto al norte (COS25_15 y COS25_39), de signo positivo, mientras que en el caso de la orientación respecto al este, la escala a la que se produce la correlación no coincide (SIN25_3, SIN25_39) y la correlación es negativa.

De las variables climáticas, las correlaciones más fuertes se han encontrado con la precipitación de julio, y con signo negativo, con la temperaturas estivales (temperatura media de julio y máxima absoluta de agosto), además de la humedad relativa en el mes de septiembre.

3.2.5.2 Modelos.

El número de variables incluidas en los modelos de validación cruzada se ha establecido a partir de los valores de la figura 3.42, con el fin de minimizar el valor del $RECM_c$. En el caso de los modelos MARS-CV, a este número se le han sumado las dos variables cualitativas, incluidas en el modelo lineal previo. Además, en el modelo MARS-CV no hay variación del error en función de las variables introducidas, que es siempre superior al de los demás modelos. En los modelos GLM-CV y GAM-CV tampoco hay una disminución visible del error al incrementar el número de variables, situándose su curva visiblemente por debajo de la de MARS-CV. Las gráficas de los modelos RF-CV y SVM-CV se superponen entre sí en toda la figura, partiendo de un error ($RECM_c$ igual al de GLM-CV y GAM-CV en los modelos con una y dos variables predictoras, y disminuyendo hasta el modelo con 10 variables a partir de las cuales el valor de $RECM_c$ se mantiene aproximadamente constante. El modelo RF-CV_{IMP} se superpone a la de estos últimos, a pesar de partir de un $RECM_c$ inicial similar a la obtenida con MARS-CV.

Modelo		Mín.	Mediana	Media	Max.	SD	Nº de variables
GLM	r_v	0.36	0.50	0.50	0.61	0.04	7
	$RECM_v$	0.12	0.14	0.14	0.16	0.006	
GLM-CV	r_v	0.37	0.51	0.51	0.62	0.05	12
	$RECM_v$	0.12	0.14	0.14	0.17	0.007	
GAM	r_v	0.36	0.52	0.52	0.65	0.04	40
	$RECM_v$	0.12	0.14	0.14	0.16	0.007	
GAM-CV	r_v	0.38	0.52	0.52	0.64	0.05	12
	$RECM_v$	0.12	0.14	0.14	0.16	0.007	
MARS	r_v	0.31	0.50	0.50	0.67	0.05	18
	$RECM_v$	0.12	0.15	0.15	0.18	0.007	
MARS-CV	r_v	0.32	0.48	0.48	0.62	0.05	9
	$RECM_v$	0.12	0.15	0.15	0.17	0.007	
RF	r_v	0.44	0.60	0.60	0.70	0.04	40
	$RECM_v$	0.11	0.13	0.13	0.16	0.006	
RF-CV	r_v	0.47	0.63	0.63	0.73	0.04	12
	$RECM_v$	0.11	0.12	0.13	0.15	0.006	
RF-CV _{IMP}	r_v	0.47	0.63	0.63	0.73	0.04	11
	$RECM_v$	0.11	0.13	0.13	0.15	0.006	
SVM	r_v	0.36	0.49	0.49	0.62	0.04	40
	$RECM_v$	0.13	0.15	0.15	0.17	0.006	
SVM-CV	r_v	0.47	0.60	0.60	0.70	0.04	10
	$RECM_v$	0.11	0.13	0.13	0.15	0.007	

TABLA 3.16: Modelos a nivel regional obtenidos para arena ($\text{kg} \cdot \text{kg}^{-1}$). Estadísticos descriptivos de la correlación (r) y Raíz del Error Cuadrático Medio ($RECM_v$) obtenidos tras el remuestreo en el conjunto de validación.

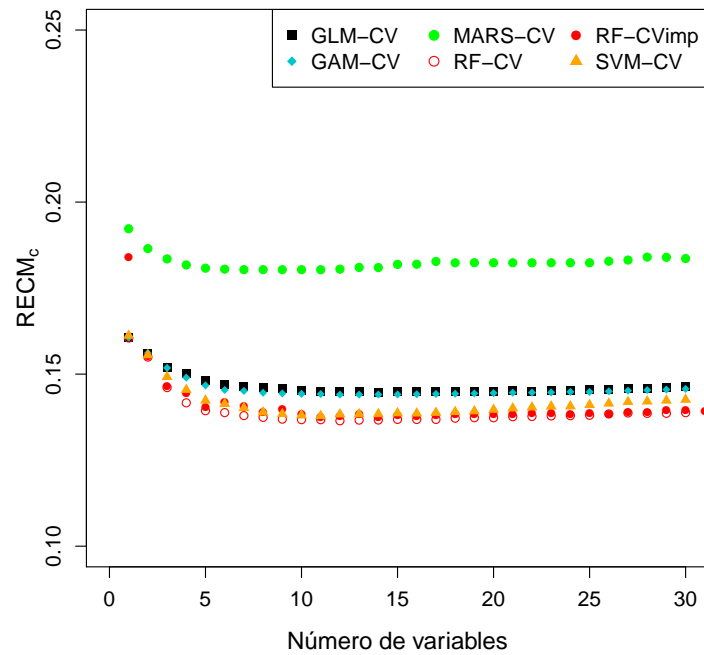
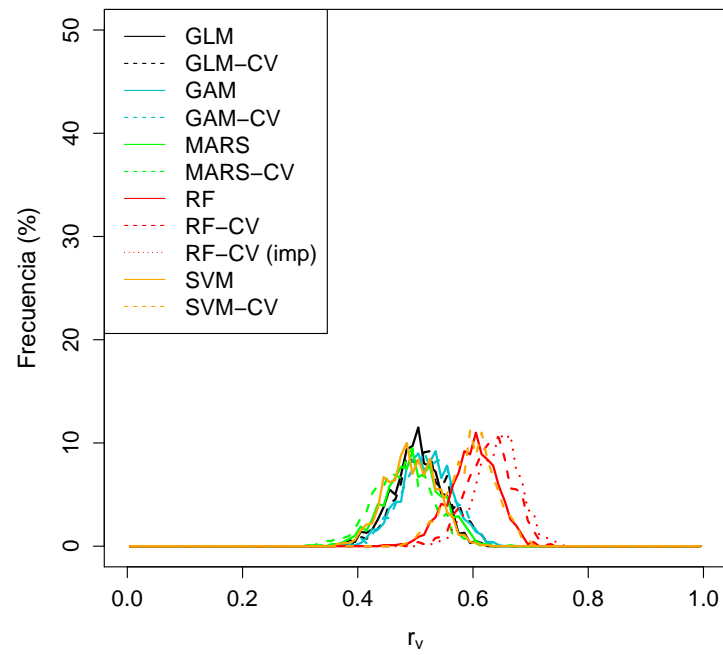
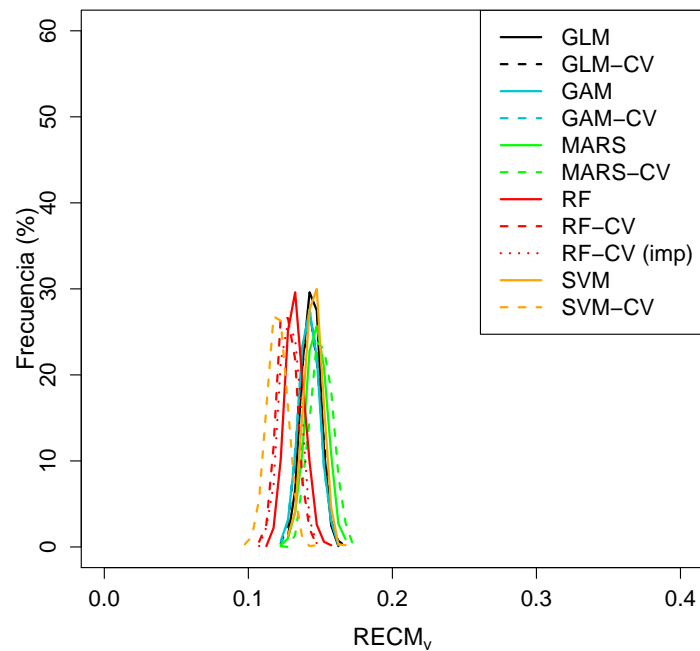


FIGURA 3.42: Raíz del error cuadrático medio ($RECM_c$) de los modelos a escala regional obtenidos con validación cruzada en bloques para arena, en función del número de variables.

En la figura 3.43 se puede observar la diferencia entre los 11 modelos elaborados, evaluados con las distribuciones de la correlación y el $RECM_v$ en el conjunto de validación. Los estadísticos descriptivos se muestran, junto con el número de variables incluidas en cada uno de ellos, en la tabla 3.16. Respecto la correlación, los modelos RF-CV_{IMP} y RF-CV son los que resultan mayores valores de correlación, que además son las más altas de las fracciones granulométricas estudiadas (figura 3.43). Les siguen los modelos RF y SVM-CV, mientras que el resto de modelos (GLM, GLM-CV, GAM, GAM-CV, MARS y MARS-CV) consiguen menores correlaciones. En este grupo se encuentra también el modelo SVM, que es el único modelo en el que la validación cruzada resulta en una mejora visible de la correlación.



(A) Correlación



(B) Raíz del error cuadrático medio

FIGURA 3.43: Modelos a escala regional obtenidos para arena ($\text{kg} \cdot \text{kg}^{-1}$). Histogramas de la correlación (r_v) y raíz del error cuadrático medio (RECM_v) tras el remuestreo en el conjunto de validación.

A diferencia de las fracciones granulométricas estudiadas anteriormente, en el caso de la arena el modelo SVM-CV es el que consigue una mayor reducción del $RECM_v$, seguido inmediatamente por RF-CV y RF-CV_{IMP} (figura 3.43.B). En cambio, el modelo SVM sin utilizar validación cruzada resulta en un mayor $RECM_v$, similar a los obtenidos con GLM, GAM y MARS. Estos últimos no producen una reducción del $RECM_v$ al utilizar el procedimiento de validación cruzada.

Respecto a la simplicidad de los modelos, expresada en función del número de variables incluidas en los mismos, la mayor exactitud en el resultado predictivo de RF-CV_{IMP} se consigue utilizando 11 predictores, mientras que RF-CV utiliza 12 variables sin conseguir el mismo resultado. En cambio, SVM-CV consigue una reducción del $RECM_v$ ligeramente mayor con solo 10 variables, aunque la correlación conseguida es menor.

Por todo ello, se ha utilizado el modelo RF-CV_{IMP} para la elaboración del mapa de arena. Este modelo explica un 35% de la varianza total de los datos, resultado similar al obtenido por Vaysse y Lagacherie (2015), mientras que Ließ *et al.* (2012) obtuvieron correlaciones menores para esta fracción. La correlación conseguida es inferior a la obtenida para los suelos de Nigeria por Akpa *et al.* (2014), aunque en nuestro caso, el $RECM$ es menor.

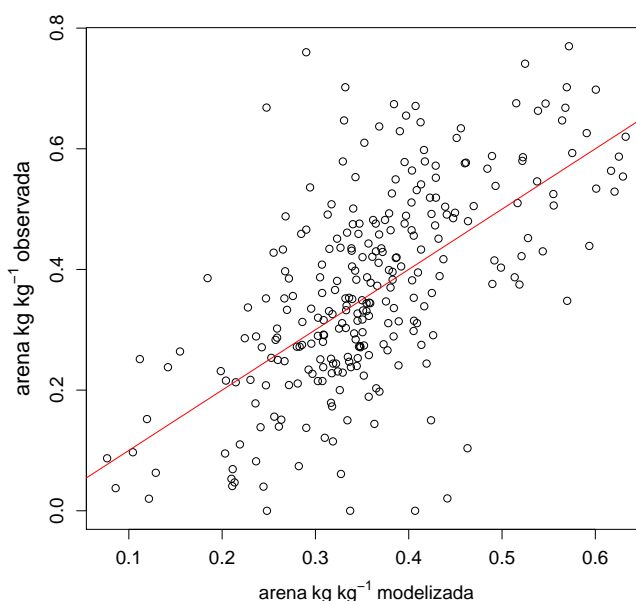


FIGURA 3.44: Diagrama de dispersión de los datos de arena ($\text{kg} \cdot \text{kg}^{-1}$) observados en el conjunto de validación frente a los obtenidos con RF-CV_{IMP} a escala regional.

La figura 3.45 muestra la dispersión entre los valores de arena observados y los obtenidos con el modelo en el conjunto de validación. En este caso, aunque el valor de la media y la mediana de los valores estimados son similares (0.36 y 0.35, respectivamente) a los de los valores observados,

el rango es menor, ya que la estimación de los valores de arena oscila entre 0.06 y 0.68. es decir, el modelo resulta en una infraestimación del contenido de arena en los suelos de la Región.

La variable de mayor importancia en el modelo seleccionado es el tipo de suelo, seguida por la precipitación total de julio (AGprectotPREDFIN7). A continuación aparecen la precipitación total de septiembre (AGprectotPREDFIN9), la temperatura mínima absoluta de diciembre (AGteminabsPREDFIN12) y la precipitación total de abril (AGprectotPREDFIN4), junto con dos de las variables geomorfológicas (índice MRVBF y medida vectorial de la rugosidad VRM39). También se incluyen, aunque con menor importancia, la temperatura máxima de agosto (AGtemaxabsPREDFIN8), la temperatura media de julio (AGtemedPREDFIN7), la pendiente (SLO25_39) y la humedad relativa de septiembre (AGetHPREDFIN9). En resumen, la proporción de arena viene determinada principalmente por variables climáticas, y algunas variables geomorfológicas calculadas con el mayor tamaño de ventana (39 celdillas).

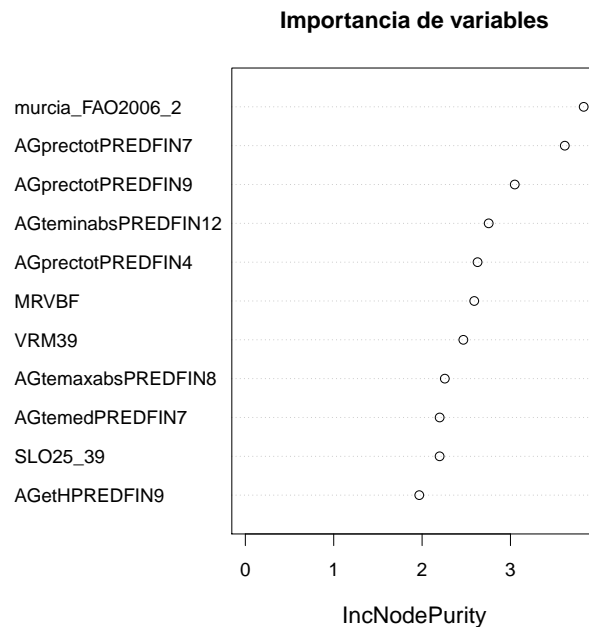


FIGURA 3.45: Importancia de variables en RF-CV_{IMP} para arena a escala regional, expresada como criterio de impureza del nodo.

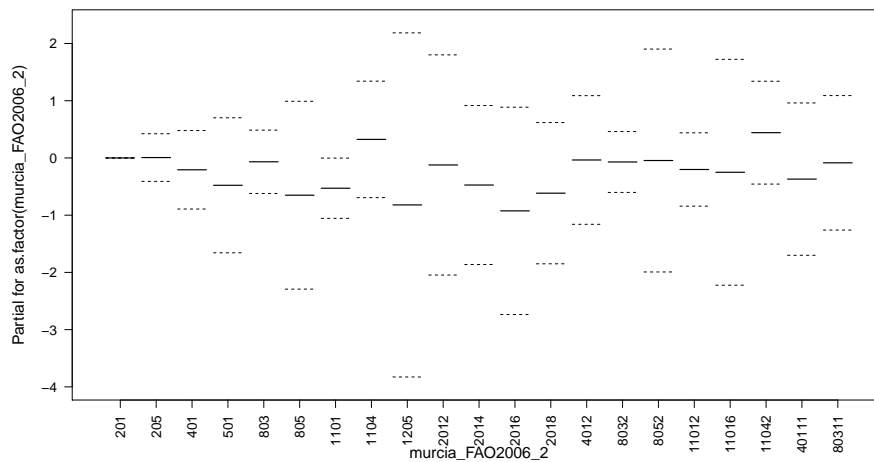


FIGURA 3.46: Modelo GAM obtenido con las variables de RF-CV_{IMP} para arena. Efecto del tipo de suelo.

Las variables climáticas (precipitación y temperatura), como se ha comentado anteriormente, influyen principalmente en la profundidad que alcanza la meteorización de los materiales, así como la intensidad de los procesos de erosión; por otro lado, el tipo de suelo recogería información diversa sobre litología y factores de edafogénesis.

De las variables geomorfométricas, hay que destacar la aparición de la medida vectorial de la rugosidad a la mayor escala. Valores altos de esta variable indicarían la presencia de grandes accidentes del relieve, mientras que los valores bajos indicarían morfologías planas y relieves lisos.

El GAM seleccionado con las variables de mayor importancia en RF-CV_{IMP} (GAM_{IMP}) da como resultado un valor medio de correlación de 0.50, ligeramente menor que GAM-CV, pero un RECM_v similar, calculado sobre el remuestreo en el conjunto de validación.

En el modelo GAM_{IMP} son significativas, de las variables paramétricas, un solo tipo de suelo, los Regosoles Háplicos (1101), mientras que el resto de unidades tiene grandes rangos de error, destacando en ese sentido los Solonchaks Háplicos, que en el gráfico mostrado en el análisis descriptivo tenía menor cantidad de arena (figura 3.46). De las variables suavizadas, solamente la precipitación de julio (AGprectotPREDFIN7) y el índice MRVBF tienen significación estadística (p -valor < 0.005) y sus gráficas se muestran en la figura 3.47. Es interesante comprobar que las gráficas reflejan las tendencias esperadas, ya que por ejemplo al aumentar la precipitación de julio disminuye la cantidad de arena, concordando con la relación positiva de esta variable con el contenido en arcilla comentado en el apartado correspondiente. También la gráfica de MRVBF muestra una tendencia acorde con lo expuesto, ya que el contenido de arena disminuye en los valles de fondo más plano. Esto coincide con los resultados descritos por [Thompson *et al.* \(2006\)](#); [Ließ *et al.* \(2012\)](#), que describieron un predominio de las fracciones más finas en las zonas más bajas y llanas, mientras que al aumentar la pendiente y la altitud aumentaría el contenido de arenas. Sin embargo, la pendiente (aunque no sea significativa) muestra una relación negativa

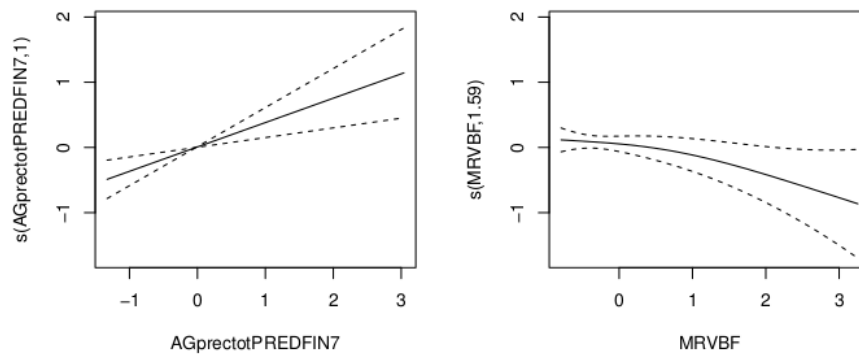


FIGURA 3.47: Modelo GAM obtenido con las variables de RF-CV_{IMP} para arena. Curvas de respuesta para las variables suavizadas significativas.

con el contenido en arena.

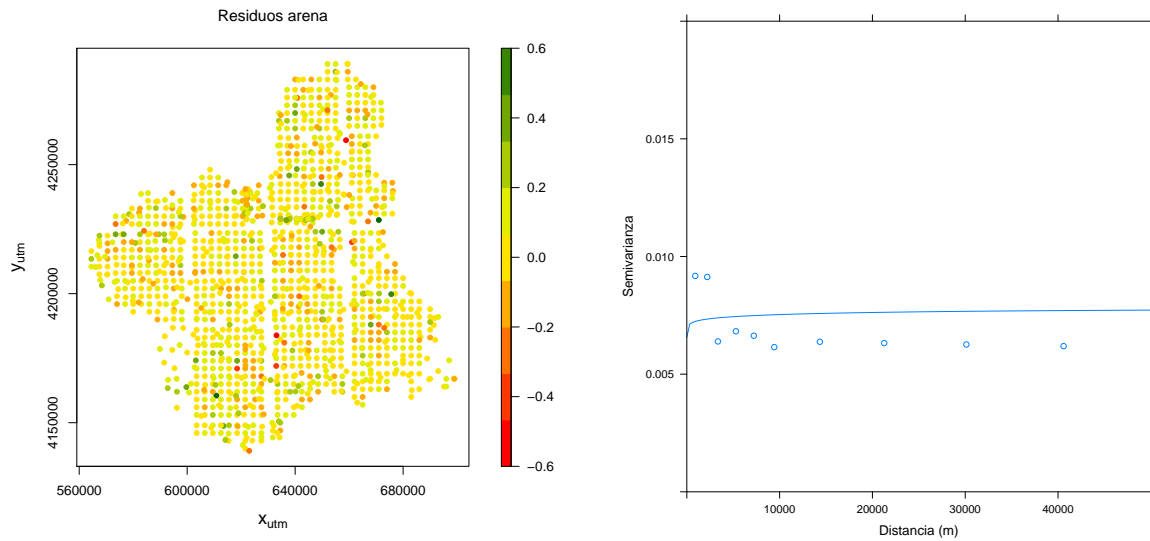
Por otro lado, las tendencias contrarias de la temperatura mínima de diciembre y la temperatura máxima de agosto (gráficas no mostradas) sugieren que el contenido de arena disminuye al aumentar la amplitud térmica. La precipitación total de abril y septiembre tienen una relación negativa con el contenido en arena, hecho que también coincide con la hipótesis inicial. Las gráficas de estas variables no se muestran al no ser significativas en el modelo, debido a los amplios márgenes de error.

3.2.5.3 Residuos del modelo.

Los residuos del modelo, representados en la figura 3.48A, muestran en su mayoría valores cercanos a 0, variando entre 0.42 y -0.5. Este residuo, calculado restando los valores originales de la predicción, tiene en la mayor parte de la región un valor negativo, es decir, el modelo produce una infraestimación del contenido de arena, excepto en unos pocos lugares (puntos verdes) donde el modelo predice valores mayores a los datos observados.

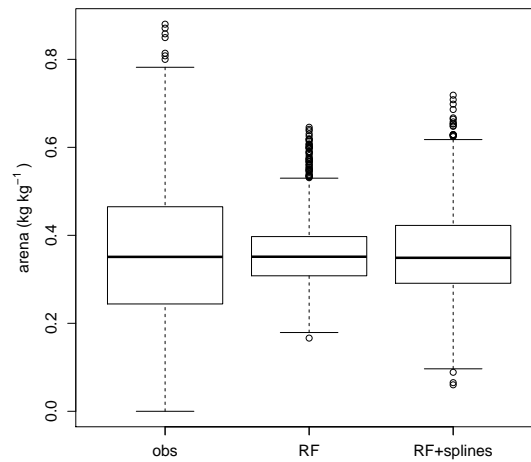
Respecto al semivariograma (figura 3.48B), ajustado a un modelo de Matern con parametrización de Stein, destaca el elevado valor del nugget (0.006) que se sitúa incluso por encima de la gráfica del semivariograma teórico, con una tendencia prácticamente horizontal, así como un rango muy amplio (20147 m), al igual que ocurría en el caso del limo. Estos datos reflejan una estructura espacial débil en los residuos, confirmada por el elevado valor de la relación entre nugget y meseta (NSR; Nugget to Sill Ratio) de 79% (Cambardella *et al.*, 1994) por lo que se ha utilizado una interpolación con splines para obtener el mapa de residuos. Sin embargo, el mapa de residuos obtenido refleja la ausencia de estructura espacial mostrada por el semivariograma.

De este modo, el mapa finalmente elaborado tras añadir los residuos interpolados con splines al mapa de predicción (página A.13), da como resultado una correlación de 0.90 con los datos originales en el conjunto de validación, y un error RECM_v de 0.07 kg · kg⁻¹. Hay que tener



(A) Distribución de los residuos del modelo RF-CV en la Región

(B) Semivariograma de los residuos

FIGURA 3.48: Distribución espacial de los residuos del modelo RF-CV_{IMP} para arena.FIGURA 3.49: Gráfico de cajas con la distribución del contenido en arena a escala regional. **obs**: valores observados; **RF**: mapa de predicción del modelo RF-CV_{IMP}; **RF+splines**: suma de la predicción de RF-CV_{IMP} y los residuos interpolados con splines.

en cuenta sin embargo que no se ha realizado una validación cruzada en el procedimiento de interpolación, y que por tanto este dato debe tomarse con precaución. Aunque el resultado supone una mejora respecto al obtenido solo con el mapa de la predicción (página A.12, $r=0.59$;

RECM_v=0.13), la adición puede considerarse como un incremento del "ruido" y no una mejora real, ya que la diferencia entre ellos no es significativa (test de Kruskal-Wallis, p-valor > 0.05). Esto se debe a que el test de Kruskal es no paramétrico y, como el de Wilcoxon, compara las medianas que son prácticamente iguales pero diferenciándose los rangos, más estrechos sobre todo en el caso de la predicción (figura 3.49). Tampoco existen diferencias entre las medias (test de la t de Welch, p-valor > 0.05).

El mapa seleccionado ha sido finalmente RF+splines, por reflejar menor el rango de arenas en la región, donde se puede observar que los mayores contenidos de arena se producen en un sector del Altiplano, y en la zona de Águilas que en apartados anteriores mostraban menores proporciones de partículas finas. Sin embargo, la mayor parte de la región muestra contenidos medios o bajos de arena. Los contenidos más bajos se encuentran en el centro de la región, que estarían ocupadas por arcillas y limos, así como en los cauces de los ríos y ramblas, la cuenca del Mar Menor y la vertiente sur de las sierras litorales del sureste, mientras que las zonas topográficamente más altas, que representan las áreas más montañosas, destacan entre ellas por tener contenidos medios de esta fracción.

3.3 Modelos a escala local: sierra de Mojantes.

En el caso de los usos del suelo, la variabilidad de usos es menor que a nivel regional, ya que la mayor parte de las 296 muestras están bajo cultivo de secano (57%) o vegetación de matorral (24.7%), siendo las muestras bajo vegetación forestal las terceras en importancia (10.8%). Por otro lado, las muestras bajo cultivo de mosaico (5) y suelo desnudo (8) fueron reclasificadas como cultivos de regadío (3) y vegetación de matorral (7), respectivamente, debido al escaso número de muestras, inferior a 5, que limitaba su representatividad además de dificultar la aparición de estas clases en los conjuntos de calibración y validación.

También en el caso de los tipos de suelo se encuentra una menor variabilidad que en el conjunto de la región, pues las muestras se corresponden con sólo 6 tipos de suelo, : la mayor parte de las muestras (74.7%) se clasifica como Calcisoles, la mayor parte de ellos (60.8%) son Calcisoles Pétricos (205) y el resto (13.8%) Calcisoles Háplicos (201); de las restantes, la mayoría se agrupan en Leptosoles (21.2%), casi todos ellos (17.5 %) Leptosoles Líticos (803), mientras que el 3.7% son Leptosoles Réndzicos (805) y Regosoles Lépticos (1101). El 4% restante se clasifican como Fluvisoles Háplicos (401) y Regosoles Háplicos (1101). Las asociaciones de suelos (Calcisoles Háplicos con Fluvisoles, 2014; Calcisoles Háplicos con Leptosoles, 2018) estaban presentes en menos de cinco muestras, y se reclasificaron previamente como Calcisoles Háplicos, por la razón expuesta en el caso de los usos minoritarios.

3.3.1 Carbono orgánico

3.3.1.1 Análisis descriptivo

El contenido de CO en la sierra de Mojantes muestra una variabilidad similar a la encontrada en el conjunto de la región, tal como se aprecia en la tabla 3.17. Casi todos los valores de la

tabla son superiores a los encontrados a nivel regional, excepto el coeficiente de variación que es ligeramente menor.

Mín.	Mediana	Media	Máx.	DS	CV(%)
0.70	14.90	20.85	112	18.14	86.99

TABLA 3.17: Estadísticos descriptivos del contenido de CO ($\text{g} \cdot \text{kg}^{-1}$) en los suelos de la sierra de Mojantes. **Mín.**= Mínimo; **Máx.**= Máximo; **SD**= Desviación estándar; **CV (%)**. = Coeficiente de variación.

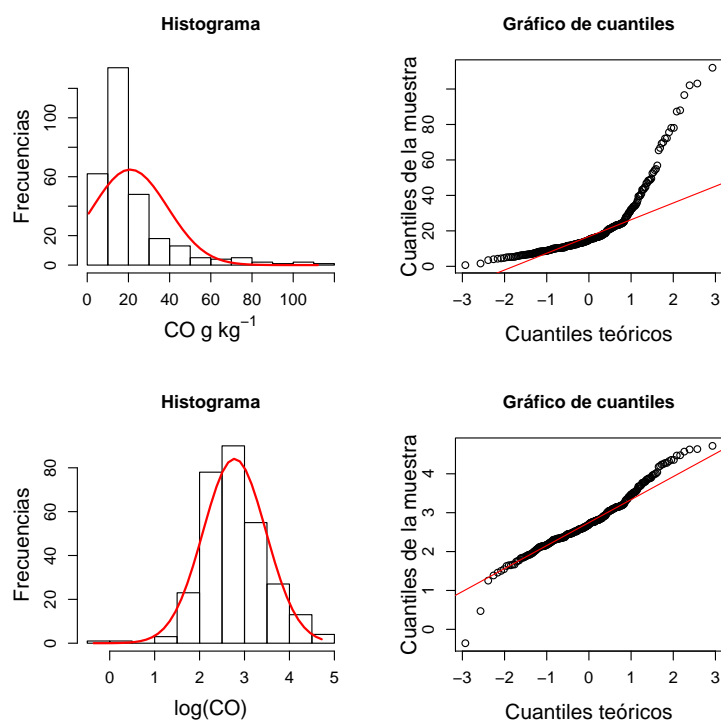
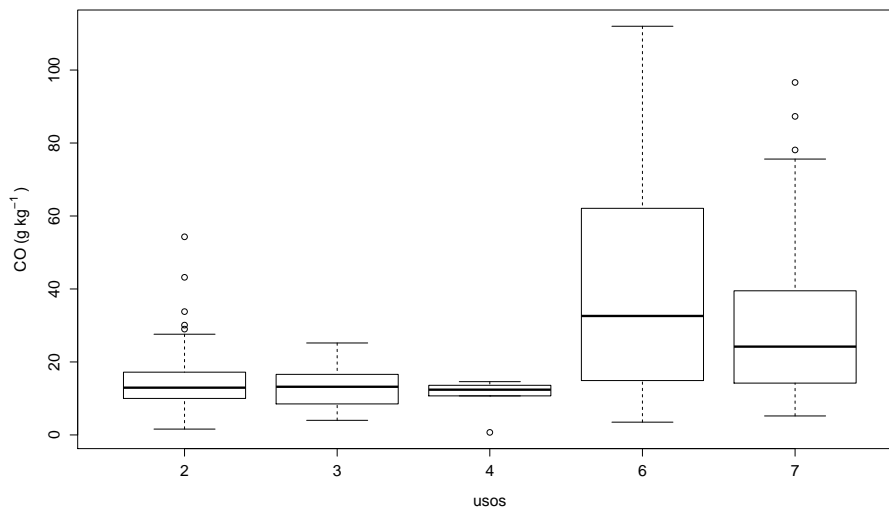


FIGURA 3.50: Histogramas y gráficos de cuantiles de CO en la sierra de Mojantes. **Arriba:** datos originales; **abajo:** datos transformados logarítmicamente.

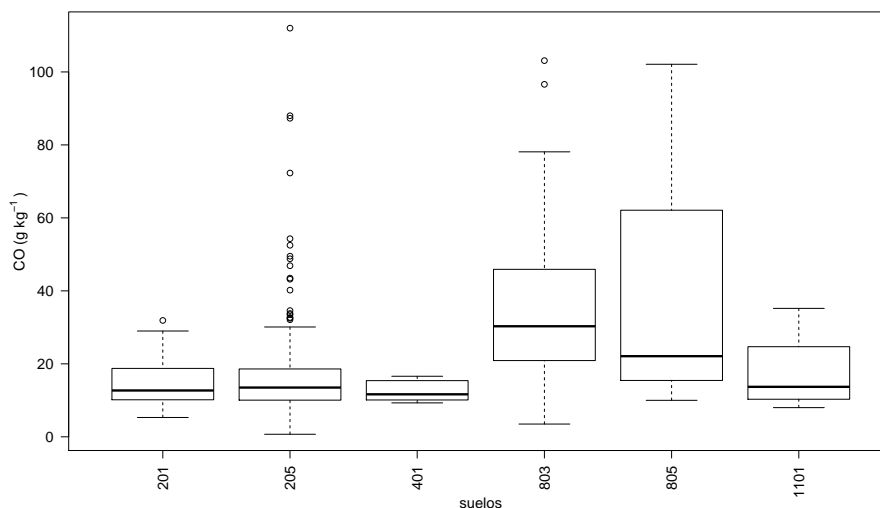
La distribución de la variable no se ajusta tampoco en este caso a una distribución normal (test de Shapiro-Wilk; $W=0.71$, $p\text{-valor} < 2.2 \cdot 10^{-16}$), ni siquiera tras la transformación logarítmica (test de Shapiro-Wilk, $W=0.96$, $p\text{-valor} < 5.95 \cdot 10^{-6}$), aunque esta sí produce un histograma razonablemente simétrico (Fig 3.50), justificando el uso de la distribución gaussiana en los modelos GLM, GAM, y el modelo GLM con las variables cualitativas previo al MARS.

La transformación tampoco resulta en una variable homocedástica si se comparan los usos de suelo (test de Bartlett, $K^2 = 45.14$, $p\text{-valor} = 4.38 \cdot 10^{-8}$) ni al comparar los tipos de suelo (test de Bartlett, $K^2 = 18.47$, $p\text{-valor} = 0.002$), por lo que se han utilizado test no paramétricos para comparar los contenidos de CO entre los distintos grupos (test de Kruskal-Wallis seguido por un test de Wilcoxon entre los diferentes pares). El intervalo de confianza para todos los test ha

lado fijado en $\alpha = 0.05$. La figura 3.51 muestra la distribución del CO en función del uso y el tipo de suelo.



(A) CO en función del uso.



(B) CO en función del tipo de suelo.

FIGURA 3.51: Distribución de CO ($\text{g} \cdot \text{kg}^{-1}$) en la sierra de Mojantes, en función del uso y tipo de suelo.

La distribución en función de los usos muestra similitudes con la anteriormente encontrada a nivel regional (figura 3.51A). Se encuentran diferencias significativas entre los grupos (test de Kruskal-Wallis, $\chi^2 = 61.3118$, $p\text{-value} = 1.537 \cdot 10^{-12}$), siendo el contenido de CO considerablemente menor en los suelos en cultivo que los suelos bajo uso natural, especialmente bajo cubierta forestal. De los suelos en cultivo, no se encuentran diferencias significativas entre los usos de regadío, secano y cultivos arbóreos; tampoco se encuentran diferencias significativas entre los usos de bosque y matorral.

En relación a los tipos de suelo, también hay diferencias significativas entre los contenidos de CO en los distintos grupos (test de Kruskal-Wallis, $\chi^2 = 51.1508$, p-value= $8.055 \cdot 10^{-10}$), que es mayor en los Leptosoles Líticos (803) y los Leptosoles Réndzicos (805) que en el resto de grupos (figura 3.51B); esto se atribuye a la cubierta de matorral y forestal de estos suelos, al no ser cultivados por su escasa profundidad. Entre ambos tipos de Leptosoles no se han encontrado diferencias significativas.

Por el contrario, los Calcisoles Háplicos (201) y Pétricos (205), que son los más abundantes en la zona de estudio, tienen contenidos de CO más bajos, inferiores por lo general a los $20 \text{ g} \cdot \text{kg}^{-1}$, aunque los Calcisoles Pétricos son también los que mayor cantidad de valores atípicos muestran. Esto puede atribuirse al calificativo Pétrico de estos suelos, indicador de una limitación en profundidad que dificultaría su puesta en cultivo, por lo que muchos de estos suelos conservarían su cubierta natural.

Los Fluvisoles Háplicos (401) y Regosoles Háplicos (1101) son los suelos menos abundantes. Ambos presentan contenidos de CO semejantes a los Calcisoles, aunque en el caso de los Fluvisoles el rango es el más estrecho, posiblemente por la puesta en cultivo; en los Regosoles el rango es más amplio, con contenidos algo mayores, lo que también se puede atribuir a la conservación de la cubierta vegetal natural, debido a la escasa aptitud agrícola de estos suelos.

Respecto a las variables cualitativas, destaca que las correlaciones del CO con las variables ambientales son diferentes en esta escala de trabajo (tabla 3.18). De este modo, igual que a nivel regional, se mantiene el nivel de significación de las correlaciones con la pendiente (SLO25_39), el índice MRVBF y el factor LS, aunque con mayor magnitud.

Lo mismo sucede con la rugosidad (VRM11 y VRM39), aunque la correlación observada con el índice de Iwahasi (IWA25_39) ya no es significativa a escala local. Otras correlaciones que pierden significación a escala local son las observadas con la curvatura plana a nivel regional; el proceso contrario ocurre con la orientación (COS25_15, COS25_39) y la curvatura del perfil (PROF25_23, PROF25_39).

De las variables climáticas, solo se mantienen las correlaciones con la temperatura media de julio (AGtemedPREDFIN7), la radiación total de junio (RAD25_6) y la humedad relativa de septiembre (AGetHPREDFIN9), que además son más intensas a esta escala, mientras que las correlaciones con las variables de precipitación desaparecen.

Variables geomorfológicas			
Variable	r (Pearson)	Variable	r (Pearson)
SLO25_39	0.44***	SIN25_3	-0.009
SIN25_39	-0.05	COS25_3	0.06
COS25_15	0.11*	COS25_39	0.13*
PROF25_3	0.12*	PROF25_7	0.09
PROF25_11	0.08	PROF25_23	0.15*
PROF25_39	0.22***	PLAN25_3	-0.06
PLAN25_7	-0.005	PLAN25_11	0.00
PLAN25_15	0.01	PLAN25_19	-0.03
PLAN25_23	-0.08	PLAN25_27	-0.11
PLAN25_31	-0.12*	PLAN25_35	-0.16**
PLAN25_39	-0.14*	VRM11	0.32***
VRM39	0.56***	IWA25_27	-0.02
IWA25_39	-0.05	TPI25_3	0.08
MRVBF	-0.27***	LS	0.13*
Variables climáticas			
Variable	r (Pearson)	Variable	r (Pearson)
AGprectotPREDFIN4	0.06	AGprectotPREDFIN7	0.04
AGprectotPREDFIN9	0.04	AGprectotPREDFIN10	-0.02
AGtemedPREDFIN7	-0.35***	AGtemaxabsPREDFIN8	0.10
AGteminabsPREDFIN12	-0.09	AGetHPREDFIN9	-0.47***
RAD25_6	-0.44***	RAD25_12	0.03

TABLA 3.18: Correlaciones de $\log(\text{CO})$ con las variables predictoras cuantitativas. Los valores de correlación significativos se han resaltado con fondo gris (P-valor: * < 0.05; ** < 0.01; *** < 0.001).

3.3.1.2 Modelos

Las distribuciones de la correlación y el $RECM_v$, así como el número de variables introducidas en cada uno de los modelos elaborados, viene reflejado en la tabla 3.19. En el caso de GLM y MARS, este número se obtiene a través del proceso de selección por pasos, mientras que en el modelo GAM se han utilizado las variables obtenidas en el modelo GLM-CV.

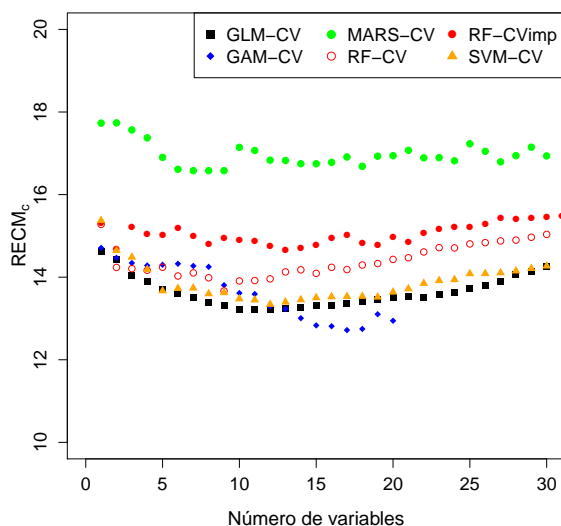


FIGURA 3.52: Raíz del error cuadrático medio ($RECM_c$) de los modelos obtenidos con validación cruzada en bloques para carbono orgánico en la sierra de Mojantes, en función del número de variables.

En los modelos con validación cruzada, el número de variables utilizado es aquel que minimiza el valor de $RECM_c$, tal como se muestra en la figura 3.52. Tal como se observa, el modelo MARS es el que mayor error produce independientemente del número de predictores incluidos, como ocurría en el modelo a escala regional; además, en este caso, las gráficas del resto de modelos muestran menor superposición. $RF-CV_{IMP}$ es el modelo que mayor error produce, con una evolución irregular al incrementar la complejidad, mientras que $RF-CV$ parte del mismo valor de error en el modelo con 1 variable, pero este disminuye hasta llegar a los 9 predictores y vuelve a aumentar a partir de ese punto, aunque sin alcanzar el valor de $RECM_c$ de $RF-CV_{IMP}$. Los modelos $GLM-CV$ y $SVM-CV$ son los que menor error obtienen en el conjunto de calibración, y puede observarse que la gráfica de $GLM-CV$ corre paralela ligeramente por debajo de $SVM-CV$. Ambas registran un descenso al aumentar el número de variables hasta aproximadamente las 10 variables, y un ascenso a continuación.

Por su parte, el modelo $GAM-CV$ parte de un error similar a $GLM-CV$ y el error se mantiene estable al aumentar la complejidad, hasta que a partir de las 8 variables disminuye y alcanza valores inferiores al del resto de modelos. Sin embargo, hay que tener en cuenta que en el $GAM-CV$ el máximo de variables a introducir es de 20, para evitar que haya mayor número de coeficientes que de datos.

Respecto al comportamiento de los modelos elaborados para el CO en el conjunto de validación, la tabla 3.19 y la figura 3.53 muestran que las correlaciones medias son similares que las obtenidas a nivel regional, aunque en este caso los rangos de la correlación en todos los modelos son más amplios (figura 3.53A); además, en este caso el modelo MARS es el que menor correlación obtiene, seguido por GAM, y el resto de histogramas de la correlación están superpuestos, sin la diferencia clara que mostraba el modelo RF-CV a nivel regional. En este caso es el modelo RF con las cuarenta variables el que mayor correlación obtiene, quedando RF-CV en segunda posición.

Modelo		Mín.	Mediana	Media	Max.	SD	Nº de variables
GLM	r_v	0.19	0.68	0.67	0.92	0.12	12
	RECM _v	9.02	18.1	18.06	27.05	3.28	
GLM-CV	r_v	0.29	0.70	0.68	0.92	0.11	11
	RECM _v	8.58	18.46	18.39	28.72	3.28	
GAM	r_v	0.13	0.55	0.55	0.83	0.12	11
	RECM _v	8.6	19.44	19.33	32.82	4.04	
GAM-CV	r_v	0.27	0.71	0.71	0.93	0.09	17
	RECM _v	8.75	17.51	17.52	28.06	3.15	
MARS	r_v	0.03	0.40	0.39	0.80	0.12	13
	RECM _v	9.25	21.49	21.49	33.76	3.89	
MARS-CV	r_v	0.24	0.72	0.71	0.89	0.08	9
	RECM _v	6.95	16.92	16.86	18.92	2.90	
RF	r_v	0.35	0.79	0.78	0.94	0.08	40
	RECM _v	8.79	17.19	17.3	26.41	2.91	
RF-CV	r_v	0.38	0.77	0.76	0.93	0.08	10
	RECM _v	7.84	16.44	16.53	25.16	2.70	
RF-CV _{IMP}	r_v	0.40	0.74	0.73	0.93	0.08	12
	RECM _v	7.27	18.06	18.01	26.99	3.14	
SVM	r_v	0.38	0.72	0.72	0.95	0.09	40
	RECM _v	9.85	18.7	18.61	29.52	3.41	
SVM-CV	r_v	0.27	0.69	0.69	0.94	0.11	12
	RECM _v	9.14	17.95	17.85	27.9	2.89	

TABLA 3.19: Modelos obtenidos para Carbono Orgánico ($\text{g} \cdot \text{kg}^{-1}$) a escala local. Estadísticos descriptivos de la correlación (r) y Raíz del Error Cuadrático Medio (RECM_v) obtenidos en el remuestreo del conjunto de validación.

En relación al RECM_v, se mantienen valores similares o algo mayores a los obtenidos en el conjunto de calibración. Las distribuciones son parecidas en todos los modelos, y los histogramas se solapan con una gran simetría entorno a la media. En la figura 3.53B se aprecia también el amplio rango de los valores de error a nivel local, cuyos mínimos son similares a los obtenidos a nivel regional, pero con un amplio rango, alcanzando los $30 \text{ g} \cdot \text{kg}^{-1}$ en algunos de los modelos. También a esta escala son los modelos MARS y GAM los que mayor error arrojan, pero sin la gran diferencia observada a escala regional. Además, el modelo RF-CV es también el que menor RECM_v obtiene, seguido de MARS-CV y RF, mientras que con el modelo RF-CV_{IMP} el error obtenido es mayor. Los modelos obtenidos con SVM no ofrecen una mejora respecto los modelos GLM-CV o GAM, arrojando errores incluso mayores que GAM-CV.

Este resultado es contrario al esperado al comienzo de la tesis, pues el objetivo de usar una malla de muestras más densa en una zona más reducida era conseguir una mayor exactitud en la predicción; sin embargo, esta menor exactitud puede atribuirse a la gran variabilidad del contenido de CO, similar a la observada a escala regional, mientras que las variables predictoras

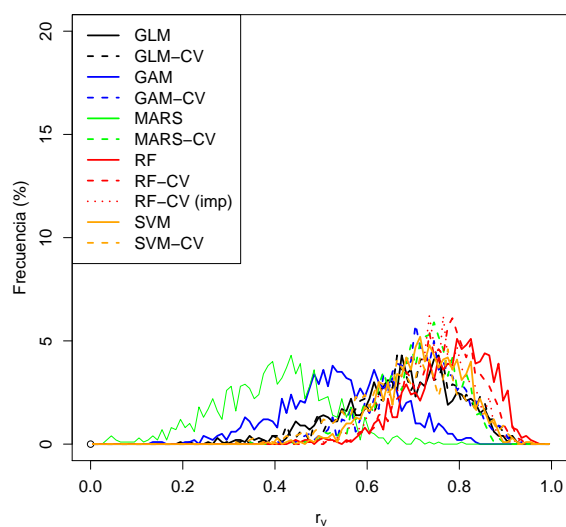
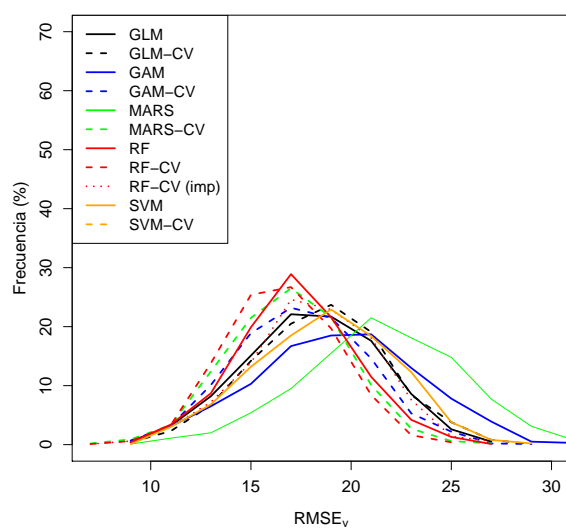
(A) Correlación (r_v)(B) Raíz del error cuadrático medio $RECM_v$

FIGURA 3.53: Modelos obtenidos para carbono orgánico en la sierra de Mojantes. Histogramas de la correlación (r_v) y la raíz del error cuadrático medio ($RECM_v$) tras el remuestreo en el conjunto de validación

utilizadas mostraban una variabilidad mucho menor.

Teniendo en cuenta el valor de $RECM_v$, la correlación y la simplicidad del modelo, se puede considerar que el modelo que mejor resultado proporciona es RF-CV, ya que además utiliza solamente 10 variables, frente otros que con el mismo número de variables resultan en mayores

valores de error, o que minimizan el error a costa de introducir mayor complejidad (número de variables) en el modelo.

Este modelo, finalmente seleccionado para generar el mapa de CO a escala local, explica un 28% de la varianza total de los datos, una proporción menor que la explicada por el modelo a escala regional. La figura 3.54 muestra la dispersión entre los datos observados y los datos predichos por el modelo en el conjunto de validación. En ella, se observa un mayor ajuste en los suelos con menores cantidades de CO, mientras que en los suelos con mayor contenido en materia orgánica se aprecia una mayor dispersión y una infraestimación general de los valores de CO. Como muestra de ello, el valor mínimo y de la mediana son mayores en los datos predichos por el modelo (10.23 y 14.17 $\text{g} \cdot \text{kg}^{-1}$) que en los datos observados (0.7 y 12.85 $\text{g} \cdot \text{kg}^{-1}$), pero el valor máximo de CO en los datos observados representa el doble del alcanzado por la predicción (102.1 frente 48.44 $\text{g} \cdot \text{kg}^{-1}$), del mismo modo que ocurría en el modelo regional.

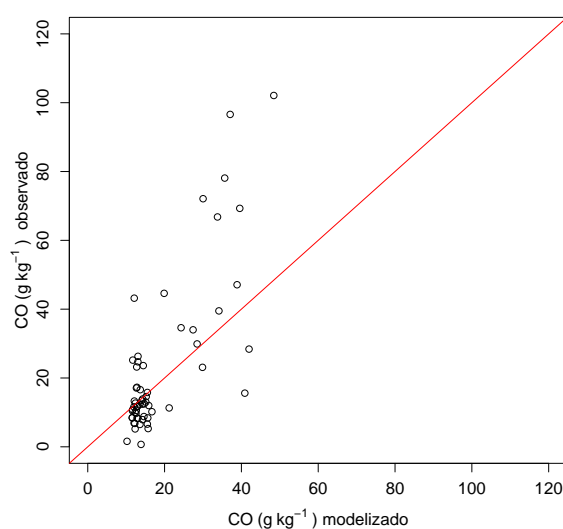


FIGURA 3.54: Diagrama de dispersión de los datos de CO observados frente los obtenidos con RF-CV a escala local en el set de validación.

Los predictores que aparecen en el modelo RF-CV a escala local son mayoritariamente variables geomorfológicas, excepto el uso del suelo (cuya importancia es mucho menor, en comparación con el modelo regional) y dos variables climáticas (la precipitación de los meses de abril, AGprectotPREDFIN4; y la de julio, AGprectotPREDFIN7), a diferencia del modelo a escala regional donde el tipo de suelo era la variable más importante, y la mayor parte de las variables eran de tipo climático. Este efecto apoya la hipótesis de que las variables que controlan la distribución espacial de CO dependen de la escala de estudio. También destaca en el modelo la aparición de variables relacionadas con la orientación: el seno de la pendiente (a la mínima y la máxima escala) y el coseno de la pendiente calculado con el menor tamaño de ventana, que estarían relacionados con las condiciones edafoclimáticas en las diferentes orientaciones. De ese

modo, las laderas orientadas al norte serían más húmedas y frescas que las orientadas al sur, favoreciendo el desarrollo de la vegetación y la incorporación de la materia orgánica al suelo.

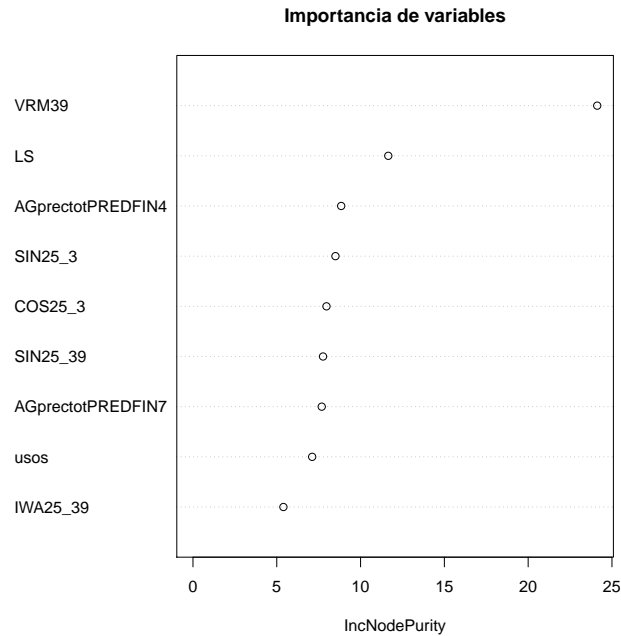


FIGURA 3.55: Importancia de variables en el modelo RF-CV para CO a escala local, expresada como criterio de impureza del nodo.

El modelo GAM obtenido con las variables de RF-CV (GAM_{IMP}) da como resultado una correlación similar en el conjunto de validación (0.75), pero superior a la que arrojaban los modelos GAM anteriormente elaborados; también el $RECM_V$ resulta en un valor parecido al que se obtenía con GAM-CV (17.57). Sin embargo, el modelo GAM_{IMP} explica mayor cantidad de varianza (41.6%) que el modelo GAM-CV (33.4%) a pesar de incluir menor cantidad de variables (10 frente a 17). Además, en comparación con GAM-CV, donde dos variables geomorfométricas y el uso forestal eran significativos, en el modelo GAM_{IMP} solamente resulta significativa la rugosidad calculada con la máxima escala (VRM39), que tiene una relación directa pero no lineal con el contenido de CO en la zona, con un aumento más acusado en la primera parte de la gráfica, y más atenuado en las zonas más rugosas (figura 3.56), lo cual concuerda con el resultado obtenido en el modelo regional.

Respecto la variable paramétrica, el modelo predice un mayor contenido de CO en las cubiertas naturales (6 y 7), aunque ninguno de los usos del suelo resulta significativo, debido a los amplios márgenes de error.

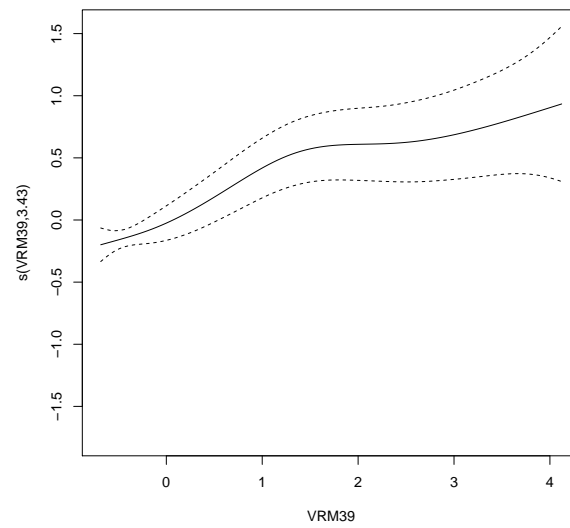


FIGURA 3.56: Modelo GAM obtenido con las variables de RF-CV para CO a escala local. Curva de respuesta de la variable suavizada significativa.

Al comparar los valores observados con el modelo a escala local y el mapa obtenido para la Región, utilizando las 296 muestras (figura 3.57), se observa que la correlación entre los valores modelizados a escala local y los datos observados es de 0.90, y el valor de RECM de $9.94 \text{ g} \cdot \text{kg}^{-1}$, a pesar de la infraestimación, reflejada en el valor medio ($17 \text{ g} \cdot \text{kg}^{-1}$ frente $22 \text{ g} \cdot \text{kg}^{-1}$ en los datos observados) y el valor máximo de la predicción ($60 \text{ g} \cdot \text{kg}^{-1}$) frente los $102 \text{ g} \cdot \text{kg}^{-1}$ de los datos en Mojantes .

Esta infraestimación es más notable en los valores derivados del modelo regional, ya que a pesar de obtener un valor medio mayor ($18.66 \text{ g} \cdot \text{kg}^{-1}$), el valor máximo es de $40 \text{ g} \cdot \text{kg}^{-1}$, un tercio del valor observado en los datos originales. A pesar de ello, la correlación entre los valores obtenidos a escala regional y los observados puede considerarse satisfactoria con un valor de 0.58, ya que los valores superiores a 40 pueden considerarse como outliers y los rangos se solapan (figura 3.58) aunque el valor de RECM sea mayor que con el modelo a escala local ($15.25 \text{ g} \cdot \text{kg}^{-1}$). En cuanto a las pruebas estadísticas realizadas, aunque el test de Wilcoxon sí muestra diferencias entre los rangos, estas pueden ser atribuidas a la gran diferencia existente en las varianzas (Fagerland y Sandvik, 2009; McDonald, 2014). Por lo tanto, hemos tomado el resultado del test de la t de Welch, que indica que no hay diferencias significativas entre la media de los datos observados y los estimados por el modelo regional.

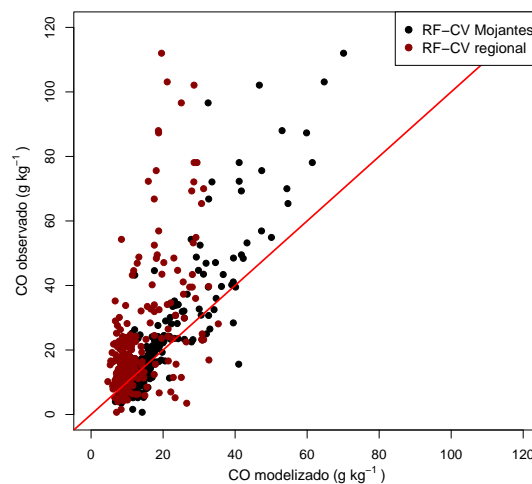


FIGURA 3.57: Comparación de los datos predichos frente los observados entre el modelo RF-CV p a escala local y la predicción de RF-CV a escala regional.

Por otro lado, la obtención de valores de correlación inferiores a 0.5 es común en la predicción de propiedades de suelos (Beckett y Webster, 1971; Adhikari *et al.*, 2013a . También Minasny *et al.* (2013) observó un aumento de la correlación al aumentar la densidad del muestreo al revisar diferentes estudios sobre cartografía de CO del suelo. Este elevado error a pesar del valor de la correlación lineal, puede atribuirse a la incertidumbre espacial de las muestras utilizadas para la obtención de los modelos a ambas escalas.

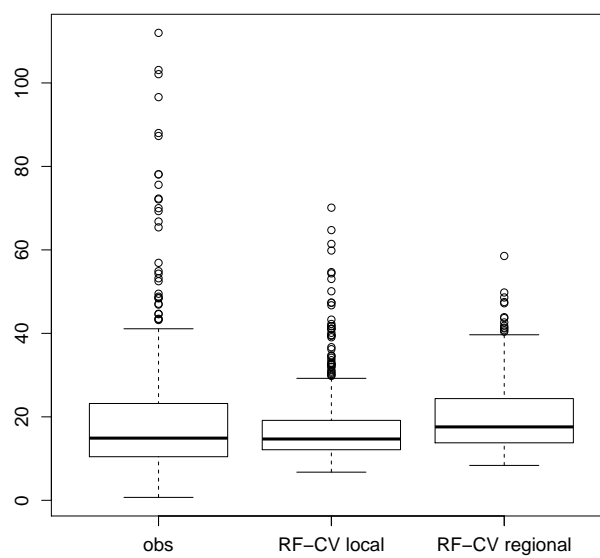


FIGURA 3.58: Gráfico de cajas del contenido de CO ($\text{g} \cdot \text{kg}^{-1}$) a escala local. **obs**: valores observados; **RF-CV local**: predicción del modelo RF-CV a escala local; **RF-CV regional**: mapa elaborado con la predicción de RF-CV a escala regional y los residuos interpolados con KO.

Mín.	Mediana	Media	Máx.	DS	CV(%)
0.026	0.5150	0.5088	0.8490	0.15	31

TABLA 3.20: Estadísticos descriptivos del contenido de CaCO_3 total o equivalente en los suelos de la sierra de Mojantes. **Mín.** = Mínimo; **Máx.** = Máximo; **SD** = Desviación estándar; **CV (%)** = Coeficiente de variación.

3.3.2 Carbonato cálcico total o equivalente

3.3.2.1 Análisis descriptivo

Los valores de carbonato cálcico total o equivalente ($\text{CaCO}_{3\text{total}}$), expresados en $\text{kg} \cdot \text{kg}^{-1}$ de suelo, varían en la sierra de Mojantes entre 0.02 y 0.85, un rango ligeramente más estrecho que el mostrado por los valores a nivel regional, aunque tanto el valor de la media como el de la mediana son más elevados (tabla 3.20). Los valores de la desviación estándar y del coeficiente de variación son más bajos que los encontrados en toda la región, lo cual indica una menor dispersión y una mayor homogeneidad de los mismos.

A pesar de no ajustarse a la distribución normal (test de Shapiro-Wilk, $W = 0.9775$, $p\text{-valor} = 0.0001385$), los valores de $\text{CaCO}_{3\text{total}}$ muestran una distribución simétrica (figura 3.59), por lo que no se ha realizado una transformación de los datos. Con el fin de evitar los valores negativos en los resultados de los modelos GLM, GAM y el modelo GLM previo al MARS, se ha utilizado la distribución gamma.

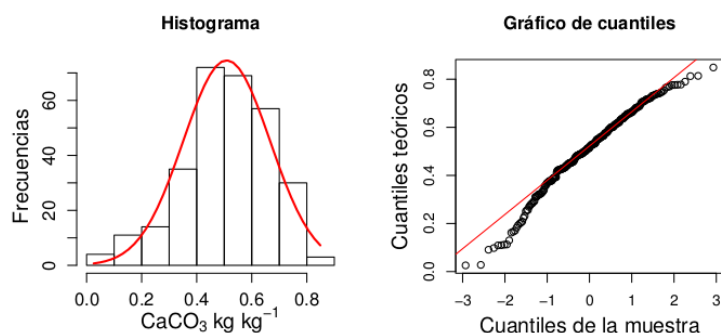
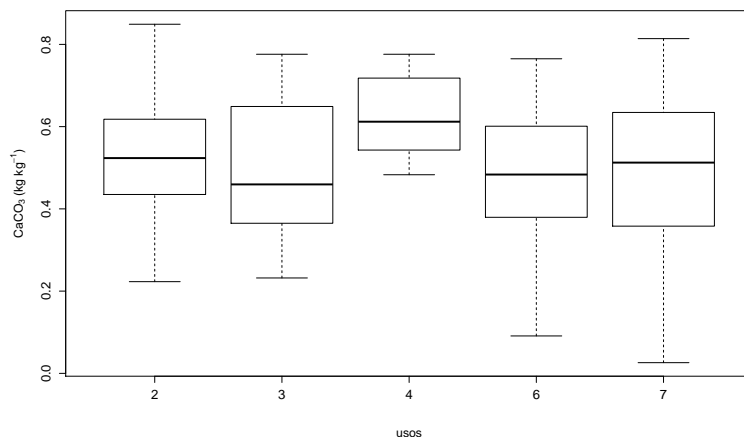


FIGURA 3.59: Histograma y gráfico de cuantiles de $\text{CaCO}_{3\text{total}}$ ($\text{kg} \cdot \text{kg}^{-1}$) en la sierra de Mojantes.

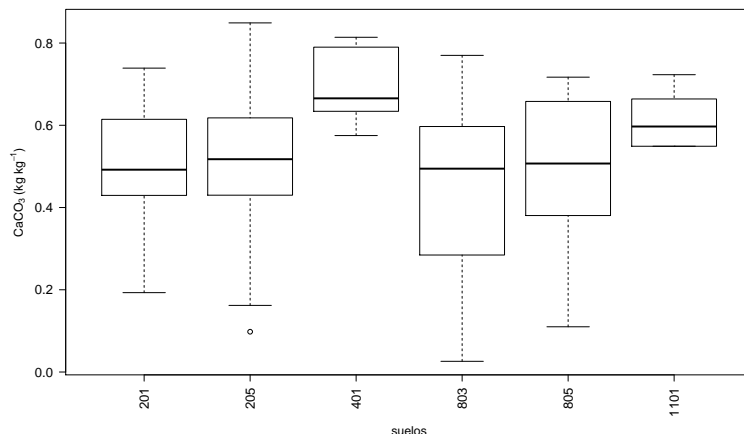
Tampoco puede asegurarse la homogeneidad de varianzas entre los diferentes usos (test de Bartlett $K^2 = 24.9731$, $p\text{-valor} = 5.094 \cdot 10^{-5}$), ni entre los tipos de suelo (test de Bartlett, $K^2 = 23.8616$, $p\text{-valor} = 0.00023$) por lo que se ha utilizado el test de Kruskal Wallis para determinar la existencia de diferencias significativas entre los grupos, seguido por el test de Wilcoxon entre pares.

En el caso de los tipos de suelo, la figura 3.60A muestra una mayor cantidad de $\text{CaCO}_{3\text{total}}$ en los cultivos arbóreos que en el resto de usos, mientras que en los suelos de bosque y bajo matorral los rangos eran más amplios que en el resto; sin embargo, las diferencias encontradas solo son

aparentes y no revisten significación estadística (p -valor > 0.05). Las diferencias encontradas entre los tipos de suelo (3.60B) sí son significativas (test de Kruskal-Wallis, $\chi^2 = 15.8499$, p -valor = 0.007286), siendo los Fluvisoles Háplicos (401) los suelos con mayores contenidos de $\text{CaCO}_{3\text{total}}$ con diferencias significativas con todos los demás grupos excepto los Regosoles Háplicos (1101), mientras que el contenido de $\text{CaCO}_{3\text{total}}$ en la capa arable de los Leptosoles Réndzicos (805) y Líticos (803) es el más variable, solapándose con los rangos en los Calcisoles Háplicos (201) y Pétricos (205).



(A) Distribución en función del uso.



(B) Distribución en función del tipo de suelo.

FIGURA 3.60: Distribución de $\text{CaCO}_{3\text{total}}$ ($\text{kg} \cdot \text{kg}^{-1}$) en función del uso y el tipo de suelo en la sierra de Mojantes.

Respecto las variables cuantitativas, destacan las relacionadas con la rugosidad VRM11 y VRM39, que también mostraban correlaciones significativas a escala regional, pero que a escala local son más intensas. Lo mismo ocurre con la curvatura del perfil, donde aumenta la fuerza de la correlación con las calculadas a 23 y 39 celdillas, que ya eran significativas a escala regional, pero además cobran significación las calculadas con 7 y 11 celdillas. En cambio, a escala local ya no resultan significativos el índice MRVBF y el factor LS. De las variables climáticas, solo resultan significativas las correlaciones con la temperatura media de julio (AGtmedPREDFIN7)

Variables geomorfológicas			
Variable	r (Pearson)	Variable	r (Pearson)
SLO25_39	-0.06	SIN25_3	-0.001
SIN25_39	-0.031	COS25_3	0.04
COS25_15	0.069	COS25_39	0.081
PROF25_3	-0.11	PROF25_7	-0.15**
PROF25_11	-0.16**	PROF25_23	-0.23***
PROF25_39	-0.26***	PLAN25_3	0.08
PLAN25_7	0.03	PLAN25_11	0.014
PLAN25_15	0.028	PLAN25_19	0.05
PLAN25_23	0.087	PLAN25_27	0.07
PLAN25_31	0.073	PLAN25_35	0.10
PLAN25_39	0.085	VRM11	-0.24***
VRM39	-0.23***	IWA25_27	0.1
IWA25_39	0.07	TPI25_3	-0.13*
MRVBF	-0.045	LS	0.0004
Variables climáticas			
Variable	r (Pearson)	Variable	r (Pearson)
AGprectotPREDFIN4	0.028	AGprectotPREDFIN7	0.017
AGprectotPREDFIN9	0.018	AGprectotPREDFIN10	0.038
AGtemedPREDFIN7	0.19***	AGtemaxabsPREDFIN8	-0.05
AGteminabsPREDFIN12	0.046	AGetHPREDFIN9	0.11*
RAD25_6	0.12*	RAD25_12	-0.04

TABLA 3.21: Correlaciones entre el contenido en $\text{CaCO}_{3\text{total}}$ y las variables predictoras cuantitativas. Los valores de correlación significativos se han resaltado con fondo gris (P-valor: * < 0.05; ** < 0.01; *** < 0.001).

y la humedad relativa de septiembre (AGetHPREDFIN9), pero con menos fuerza que a nivel regional, y en el caso de la humedad relativa, también con menor significación.

3.3.2.2 Modelos

La figura 3.61, muestra el error cuadrático medio obtenido en el conjunto de calibración (RECM_c) para los modelos que utilizan validación cruzada, en función del número de variables utilizadas en el modelos.

El número de variables para el cual se minimiza el RECM_c es el que posteriormente se incluye en cada uno de estos modelos. En el caso de la sierra de Mojantes todos los modelos obtienen un RECM_c similar, con un valor aproximado de $0.15 \text{ kg} \cdot \text{kg}^{-1}$, sin que el aumento de variables produzca una disminución significativa. Este valor es menor que el arrojado por los modelos a escala regional, donde además el valor de RECM_c variaba según el modelo elaborado. Aún así, se aprecia que los modelos MARS-CV y GLM-CV son los modelos con mayor RECM_c , y que el uso de SVM-CV no reduce este error. Por otro lado, como en el caso del modelo a escala regional, el modelo RF-CV_{IMP} es el que mayor reducción del RECM_c produce.

La tabla 3.22 muestra la distribución de la correlación y el error calculados en el conjunto de validación (RECM_v), además del número de variables finalmente incluido en cada uno de los modelos elaborados. Entre ellos destacan GLM, GLM-CV, MARS y MARS-CV por los valores mínimos negativos de la correlación. Para todos los modelos, los valores medios de la correlación eran menores que a escala regional, pero no los valores máximos, que eran similares. También eran menores los valores medios de RECM_v al compararlos con los obtenidos en los modelos a escala regional.

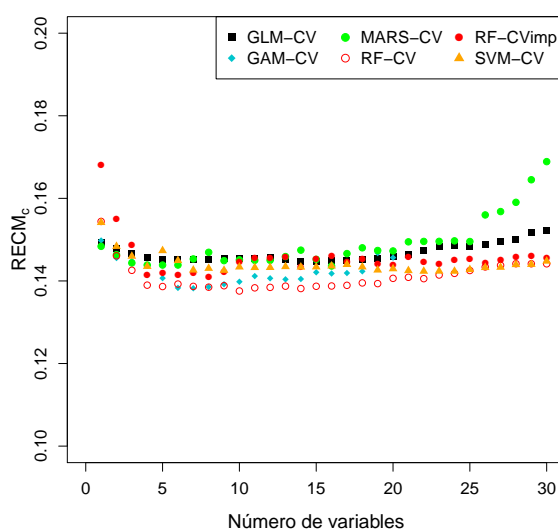


FIGURA 3.61: Raíz del error cuadrático medio ($RECM_c$) de los modelos obtenidos con validación cruzada en bloques para $CaCO_3$ total o equivalente a escala local, en función del número de variables.

En la figura 3.62A se puede observar que los rangos de la correlación eran más amplios en los modelos a escala local, indicando un comportamiento menos fiable de la predicción, y que en este caso los modelos con Random Forest no muestran una ventaja clara frente los demás. En el caso del $RECM_v$ (figura 3.62B) los histogramas están más solapados, y todos ellos, a diferencia de los observados a escala regional, tienen una distribución simétrica; RF-CV es el que mejor resultado ofrece, aunque la diferencia con el resto de modelos, especialmente SVM-CV, es menor que en el modelo a escala regional.

Finalmente, se ha seleccionado como mejor modelo RF-CV, ya que a pesar de la escasa diferencia, es el que mayor valor medio de correlación arroja, con el mismo valor medio de $RECM_v$, utilizando solo una variable más que RF-CV_{IMP}. Este modelo, sin embargo, explica solamente el 18.21% de la varianza total de los datos, un valor muy inferior al obtenido en el modelo a escala regional (44%). La figura 3.63 muestra la dispersión entre los valores observados y los valores obtenidos por el modelo, donde se observa, a semejanza del modelo para toda la región, un rango más estrecho en los valores predichos; además, aunque los valores de media y mediana son similares en los datos predichos y observados, la tendencia global es a una infraestimación del contenido de $CaCO_{3total}$.

Las variables predictoras que aparecen en el modelo RF-CV son mayoritariamente geomorfológicas, calculadas casi todas con el mayor tamaño de ventana (39 celdillas), excepto la precipitación total de los meses de septiembre y abril, y el tipo de suelo. Por orden de importancia, destacan la curvatura del perfil, en dos de las escalas de cálculo utilizadas (39 y 23 celdillas), y la medida vectorial de la rugosidad (VRM39). Les siguen las dos variables climáticas mencionadas, que tienen un índice de pureza del nodo similar a la orientación (representada por COS25_39), la

Modelo		Mín.	Mediana	Media	Max.	SD	Nº de variables
GLM	r_v	-0.11	0.19	0.26	0.63	0.11	20
	RECM _v	0.13	0.19	0.19	0.26	0.02	
GLM-CV	r_v	-0.07	0.37	0.37	0.73	0.12	14
	RECM _v	0.12	0.16	0.16	0.20	0.01	
GAM	r_v	0.007	0.38	0.38	0.70	0.10	14
	RECM _v	0.11	0.16	0.16	0.20	0.01	
GAM-CV	r_v	0.41	0.57	0.57	0.71	0.04	12
	RECM _v	0.14	0.16	0.16	0.19	0.007	
MARS	r_v	-0.02	0.39	0.38	0.78	0.11	18
	RECM _v	0.17	0.19	0.19	0.22	0.007	
MARS-CV	r_v	-0.24	0.27	0.27	0.62	0.13	5
	RECM _v	0.13	0.17	0.17	0.20	0.01	
RF	r_v	0.26	0.54	0.54	0.78	0.08	40
	RECM _v	0.10	0.15	0.15	0.18	0.01	
RF-CV	r_v	0.20	0.52	0.51	0.80	0.09	9
	RECM _v	0.11	0.14	0.15	0.18	0.01	
RF-CV _{IMP}	r_v	0.11	0.47	0.47	0.79	0.10	8
	RECM _v	0.11	0.15	0.15	0.19	0.01	
SVM	r_v	0.26	0.52	0.52	0.72	0.08	40
	RECM _v	0.10	0.14	0.14	0.19	0.01	
SVM-CV	r_v	-0.05	0.35	0.35	0.70	0.10	9
	RECM _v	0.11	0.16	0.16	0.21	0.01	

TABLA 3.22: Modelos obtenidos a escala local para CaCO₃ total o equivalente (kg·kg⁻¹). Estadísticos descriptivos de la correlación (r) y Raíz del Error Cuadrático Medio (RECM_v) tras el remuestreo en el conjunto de validación.

pendiente y la curvatura plana (PLAN25_35). Finalmente, aparecen la rugosidad expresada por el índice de Iwahasi (IWA25_39) y los tipos de suelo, con una importancia menor. Al comparar con el modelo a escala regional, destaca el hecho de que el coseno de la orientación y la medida vectorial de la rugosidad también aparecieran en el modelo, pero con menor importancia; por el contrario, el tipo de suelo es menos importante a escala local que a escala regional.

El modelo GAM elaborado con las variables de RF-CV (GAM_{RF-CV}) explica un 35.4% de la varianza total da como resultado una correlación que los modelos GAM y GAM-CV anteriores, y una disminución del RECM_v, valores similares a los obtenidos con RF-CV. De las variables introducidas, resultan significativas la medida de la rugosidad (VRM39), la curvatura del perfil (PROF25_39) y en menor medida, la pendiente (SLO25_39). Entre ellas, la figura 3.65 muestra que solamente la pendiente tiene una relación lineal con el contenido en CaCO_{3total}, y negativa, de modo que en la pendientes más escarpadas el contenido en CaCO_{3total} es menor. Respecto la rugosidad (VRM39), el aumento de esta variable se relaciona con un aumento en el contenido en CaCO_{3total}, pero poco pronunciado, y alcanzando un máximo a partir del cual la relación se invierte; esto indicaría un contenido bajo en CaCO_{3total} en las zonas llanas, que aumenta en los relieves pero no en aquellos más irregulares. La relación del contenido en CaCO_{3total} con la curvatura del perfil es poco clara, ya que en las curvaturas cóncavas (valores más bajos) se observa un contenido bajo de CaCO_{3total}, aumentando en las convexas pero no de forma uniforme, ya que disminuye de nuevo en un amplio rango de curvaturas para volver a aumentar. El resto de variables muestran gráficas horizontales. Tampoco son significativos los tipos de suelo, con rangos de error amplios para todos los grupos presentes en la zona.

La comparación de los valores observados y los modelizados a ambas escalas en el total de

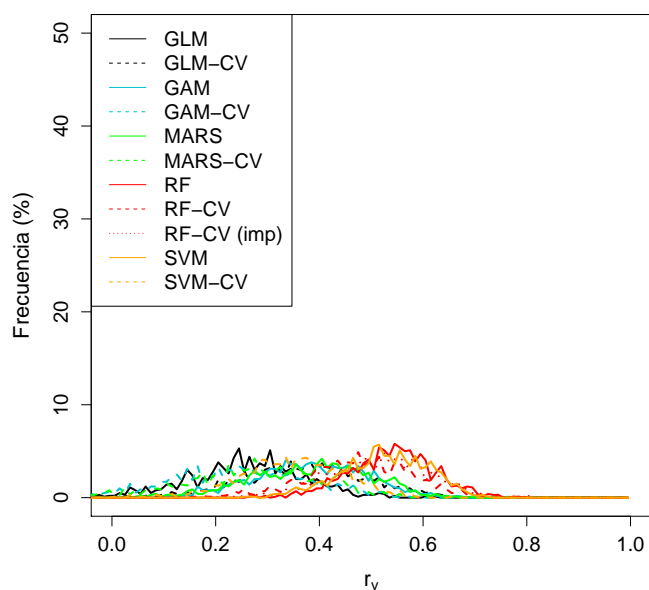
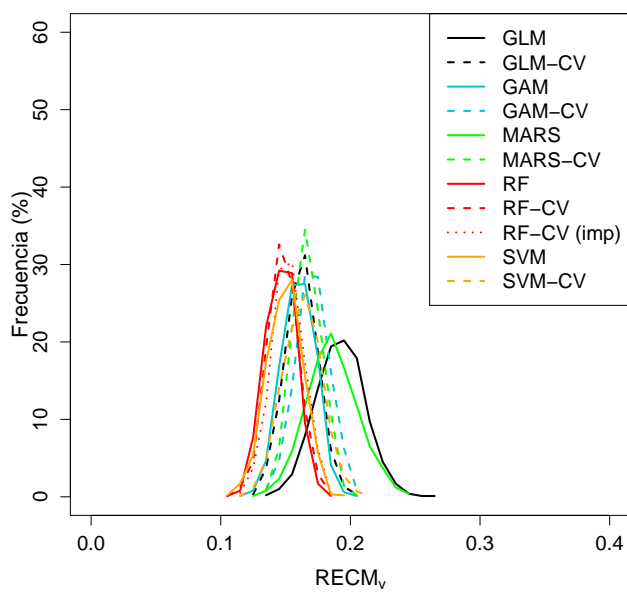
(A) Correlación (r_v)(B) Raíz del error cuadrático medio (RECM_v)

FIGURA 3.62: Modelos obtenidos a escala local para CaCO_3 total o equivalente. Histogramas de la correlación (r_v) y la raíz del error cuadrático medio (RECM_v) tras el remuestreo en el conjunto de validación.

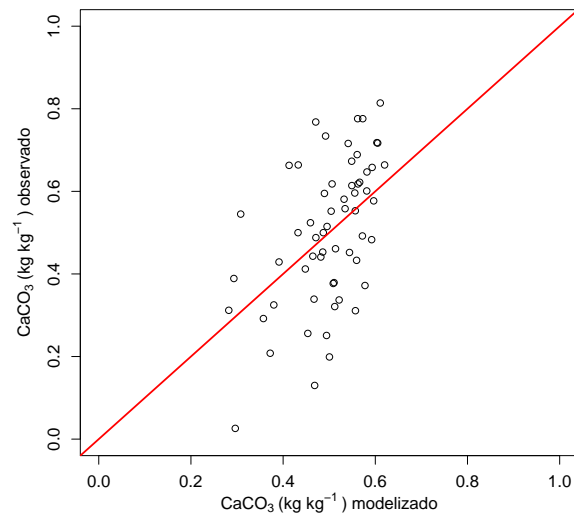


FIGURA 3.63: Diagrama de dispersión de los datos de CaCO_3 total o equivalente ($\text{kg} \cdot \text{kg}^{-1}$) observados en el conjunto de validación frente los obtenidos con RF-CV a escala local.

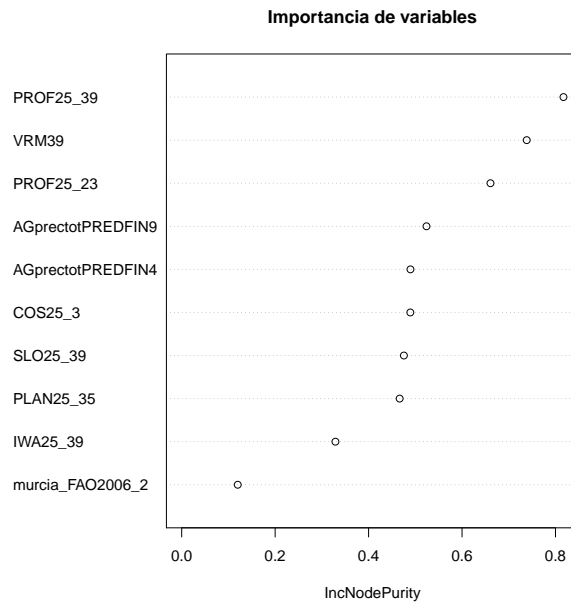


FIGURA 3.64: Importancia de variables en RF-CV para carbonato cálcico total o equivalente a escala local, expresada como criterio de impureza del nodo.

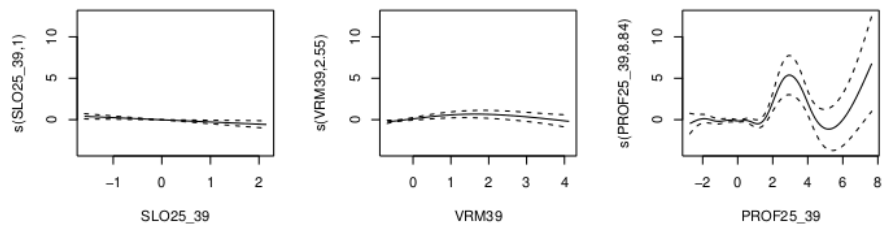
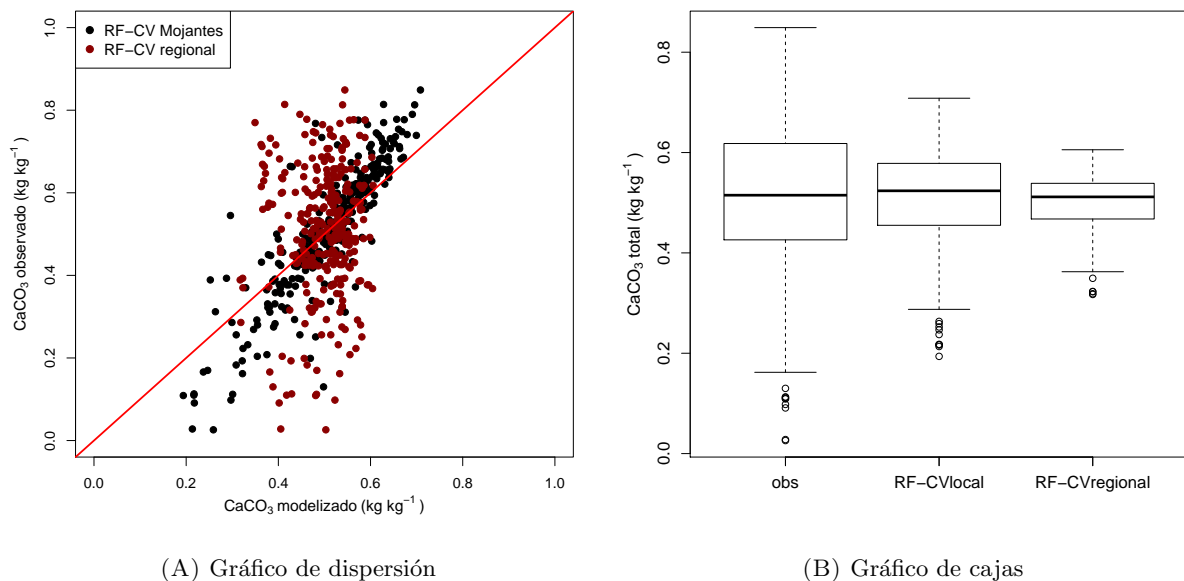


FIGURA 3.65: Modelo GAM obtenido con las variables de RF-CV para carbonato cálcico total a escala local. Curvas de respuesta para las variables suavizadas significativas.

muestras de Mojantes ($n=296$) (figura 3.66) da como resultado una correlación de 0.87 y un error RECM de $0.08 \text{ kg} \cdot \text{kg}^{-1}$ al considerar el modelo $\text{RF-CV}_{\text{Mojantes}}$, mientras que con el modelo regional la correlación es prácticamente nula y no significativa ($r=0.06$) y el error es de $0.16 \text{ kg} \cdot \text{kg}^{-1}$.



(A) Gráfico de dispersión

(B) Gráfico de cajas

FIGURA 3.66: Comparación de los datos de carbonato cálcico total observados frente los predichos por los modelos RF-CV a escala local y regional. **obs**: valores observados; **RF-CVlocal**: predicción del modelo RF-CV a escala local; **RF-CVregional**: mapa elaborado con la predicción de RF-CV a escala regional y los residuos interpolados con KO.

Esta escasa correlación en el caso del modelo regional puede atribuirse al escaso intervalo en el que están comprendidos los valores de $\text{CaCO}_3_{\text{total}}$, con un mínimo de $0.31 \text{ kg} \cdot \text{kg}^{-1}$ y un valor máximo de $0.61 \text{ kg} \cdot \text{kg}^{-1}$, mientras que con el modelo a escala local los valores varían en un rango más amplio ($0.19\text{-}0.71 \text{ kg} \cdot \text{kg}^{-1}$). A pesar de ello, los valores de media y mediana son

similares en las predicciones de ambos modelos (figura 3.66), y la comparación de estas entre sí y con los datos originales no revela diferencias significativas al utilizar el test de Wilcoxon, ni tampoco utilizando el test de la t de Welch (p -valor > 0.05).

3.3.3 Arcilla

3.3.3.1 Análisis descriptivo

Como en el caso de los modelos a nivel regional, los valores se han dividido entre 100, para expresarlos en $\text{kg} \cdot \text{kg}^{-1}$.

Los valores de arcilla en la sierra de Mojantes varían entre 0.08 y 0.42 $\text{kg} \cdot \text{kg}^{-1}$, un rango ligeramente más estrecho que en el conjunto de la región, aunque el valor de la mediana es igual. Sin embargo, en la zona de Mojantes los datos están menos dispersos y son más homogéneos que en la región, tal y como reflejan la desviación estándar y el coeficiente de variación (tabla 3.23).

Mín.	Mediana	Media	Máx.	DS	CV(%)
0.08	0.186	0.21	0.42	0.05	23.3

TABLA 3.23: Estadísticos descriptivos del contenido de arcilla ($\text{kg} \cdot \text{kg}^{-1}$) en los suelos de la sierra de Mojantes. **Mín.**= Mínimo; **Máx.**= Máximo; **SD**= Desviación estándar; **CV (%)**. = Coeficiente de variación.

La figura 3.67 indica que también en este caso los datos siguen una distribución normal, confirmada por el test de Shapiro-Wilk ($W = 0.9913$, p -valor = 0.07802), no siendo precisa su transformación. Se ha utilizado la familia "binomial" para elaborar los modelos GLM, GAM y el modelo GLM previo a MARS, con el fin de limitar el resultado arrojado por los mismos a valores entre 0 y 1.

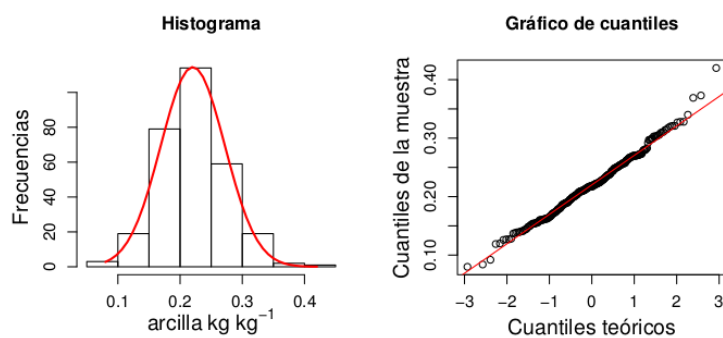
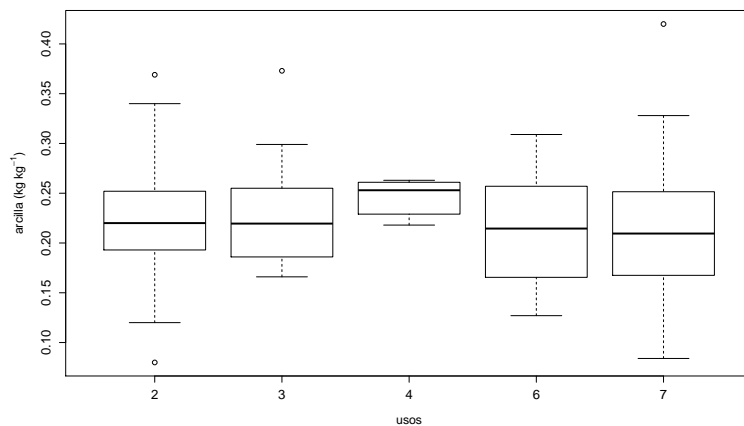


FIGURA 3.67: Histograma de frecuencias y gráfico de cuantiles de arcilla.

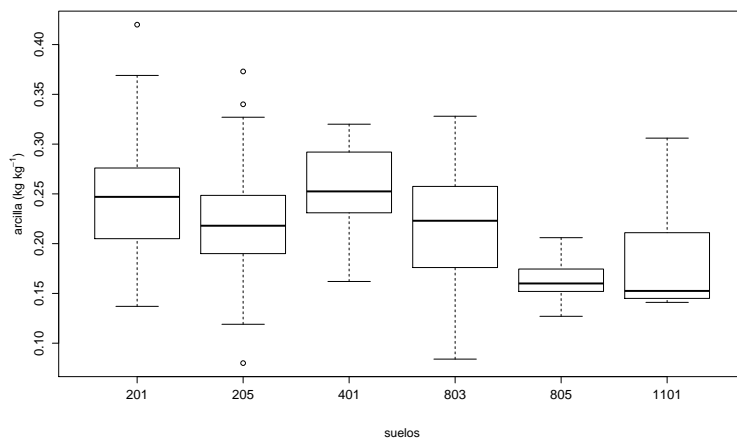
El contenido de arcilla puede considerarse homocedástico respecto a la distribución por usos (test de Bartlett, $K^2 = 7.2752$, p -valor = 0.122), donde los cultivos arbóreos (4) son los que mayor porcentaje de esta fracción tienen, frente al resto de usos donde este porcentaje es similar

(figura 3.68A; sin embargo, el test ANOVA indica que estas diferencias son solo aparentes y no tienen significación estadística ($F= 1.48$, $p\text{-valor} = 0.208$).

Respecto los tipos de suelo, el contenido de arcillas no puede considerarse homocedástico (test de Bartlett, $K^2= 12.9505$, $p\text{-value} = 0.02385$), por lo que se ha utilizado el test de Kruskal-Wallis, que sí indica la existencia de diferencias significativas ($\chi^2 = 28.09$, $p\text{-valor} = 3.495 \cdot 10^{-5}$). La figura 3.68B muestra que el contenido en arcilla es mayor en los Fluvisoles Háplicos (401) y los Calcisoles Háplicos (201). Como en el caso de la escala regional, la mayor cantidad de arcilla en los Fluvisoles puede atribuirse al aporte de sedimentos por los ríos y ramblas. Por el contrario, los suelos con menor porcentaje de arcillas son los Leptosoles Réndzicos (805) y los Regosoles Háplicos (1101). En el caso de estos últimos, se esperaría, tal como ocurre a escala regional, una mayor cantidad de partículas finas. En el caso de ambos tipos de Leptosoles la tendencia es además inversa a la observada a escala regional, donde los Leptosoles Líticos tenían una proporción mayor de arcilla que los Leptosoles Réndzicos.



(A) En función del uso



(B) En función del tipo de suelo

FIGURA 3.68: Análisis exploratorio: Contenido de arcillas en la sierra de Mojantes, en función del uso y tipo de suelo.

Variables geomorfológicas			
Variable	r (Pearson)	Variable	r (Pearson)
SLO25_39	-0.17**	SIN25_3	0.021
SIN25_39	0.047	COS25_3	0.088
COS25_15	0.053	COS25_39	0.074
PROF25_3	-0.028	PROF25_7	-0.032
PROF25_11	0.026	PROF25_23	0.022
PROF25_39	0.09	PLAN25_3	-0.028
PLAN25_7	-0.003	PLAN25_11	0.089
PLAN25_15	0.069	PLAN25_19	0.045
PLAN25_23	0.035	PLAN25_27	0.062
PLAN25_31	0.10	PLAN25_35	0.082
PLAN25_39	0.036	VRM11	-0.06
VRM39	-0.16**	IWA25_27	0.044
IWA25_39	0.075	TPI25_3	0.007
MRVBF	0.16**	LS	-0.04
Variables climáticas			
Variable	r (Pearson)	Variable	r (Pearson)
AGprectotPREDFIN4	0.084	AGprectotPREDFIN7	0.094
AGprectotPREDFIN9	0.10	AGprectotPREDFIN10	-0.071
AGtemedPREDFIN7	0.019	AGtemaxabsPREDFIN8	-0.061
AGteminabsPREDFIN12	-0.04	AGetHPREDFIN9	0.18**
RAD25_6	0.11*	RAD25_12	-0.08

TABLA 3.24: Correlaciones del contenido en arcilla con las variables predictoras cuantitativas. Los valores de correlación significativos se han resaltado con fondo gris (P-valor: * < 0.05; ** < 0.01; *** < 0.001).

Según la tabla 3.24, varias de las variables que tenían correlaciones significativas con la arcilla a escala regional, pierden esa significación en la sierra de Mojantes, quedando solamente cuatro variables correlacionadas con este constituyente. Además, la magnitud de estas correlaciones es menor que a escala regional. De las variables geomorfológicas, destaca la correlación positiva con el índice MRVBF, lo cual se explicaría por el aporte de sedimentos finos hasta mayores distancias por parte de los cursos de agua, que se acumularían en los valles más planos. También es significativa la correlación con la pendiente (SLO25_39), de igual magnitud pero negativa, indicando una disminución del contenido de arcilla al aumentar la pendiente, lo que se fundamentaría en el hecho de que en las zonas con mayores pendientes se produciría más erosión de materiales, sobre todo los más finos. La rugosidad (VRM39) también está correlacionada negativamente con el porcentaje de arcilla, lo cual se podría atribuir a que en las zonas más escarpadas el contenido de estas es menor que en las zonas más planas, en el mismo sentido que se ha comentado para MRVBF. De las variables climáticas, solamente la humedad relativa de septiembre y la radiación total de junio muestran correlaciones positivas y significativas con el contenido de arcilla.

3.3.3.2 Modelos

La figura 3.69 muestra el número de variables utilizada en cada uno de los modelos que utilizan la validación cruzada. En esta figura se observa que los modelos GLM-CV y MARS-CV son los que mayor error producen en el conjunto de calibración ($RECM_c$), mientras que GAM-CV (incluyendo hasta 19 variables para evitar el exceso de coeficientes) y SVM-CV los que menor $RECM_c$ producen, sin que en ninguno de ellos se perciba una disminución clara del error al aumentar la complejidad del modelo. El número final de variables incluido finalmente en los

modelos se muestra en la tabla 3.25.

En cuanto al resto de modelos, hay que destacar que GLM no incluye las variables del procedimiento de regresión por pasos, ya que este elimina todos los predictores, sino las variables que resultaban significativas en el análisis exploratorio; además, el modelo GAM se ha elaborado con las variables resultantes de GLM-CV.

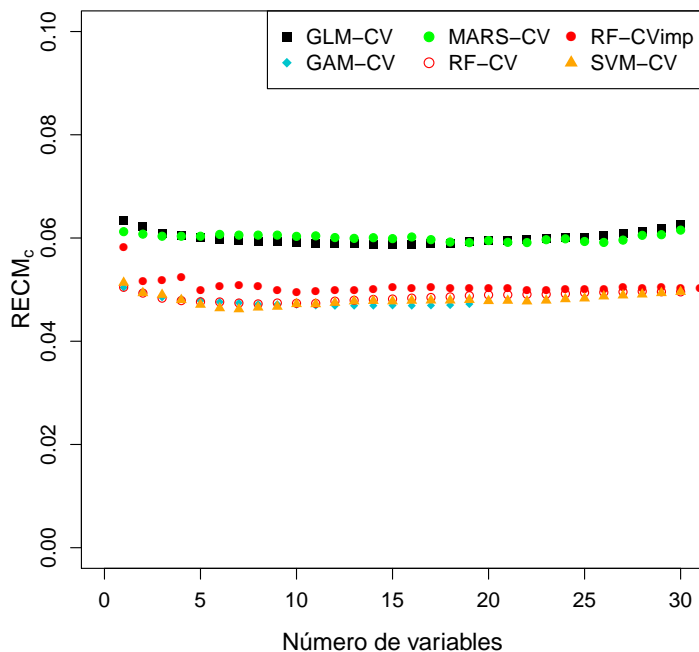


FIGURA 3.69: Raíz del error cuadrático medio ($RECM_c$) de los modelos a escala local obtenidos con validación cruzada en bloques para arcilla, en función del número de variables.

La tabla 3.25 muestra también la distribución de la correlación y el error en el conjunto de validación en cada modelo elaborado. En comparación con los modelos obtenidos a nivel regional para la misma variable, los modelos obtenidos para la sierra de Mojantes resultan en una correlación mucho menor, incluso con valores mínimos negativos, a pesar de que los valores máximos sean similares (figura 3.70A). Como resultado, el rango de la correlación es muy amplio y esto refleja una estabilidad de la predicción baja. Los modelos que dan una mayor correlación son MARS-CV, seguido por RF-CV, mientras que el resto de modelos, incluidos RF, RF-CV_{IMP}, SVM y SVM-CV, generan valores de correlación negativos más altos. Por otro lado, el procedimiento de validación cruzada no supone una mejora en la correlación de los modelos, excepto MARS-CV y en mucha menor medida, RF-CV.

En la figura 3.70B se puede ver que el rango de los errores es mucho más estrecho en torno a $0.05 \text{ kg} \cdot \text{kg}^{-1}$, con una gran superposición de los histogramas. A pesar de la superposición, se aprecia que MARS y GLM-CV son los modelos que mayor $RECM_v$ producen, mientras que

Modelo		Mín.	Mediana	Media	Max.	SD	Nº de variables
GLM	r_v	-0.22	0.17	0.17	0.48	0.11	5
	RECM _v	0.03	0.05	0.05	0.07	0.006	
GLM-CV	r_v	-0.38	0.06	0.06	0.58	0.15	14
	RECM _v	0.03	0.05	0.06	0.09	0.008	
GAM	r_v	-0.35	0.07	0.07	0.56	0.15	14
	RECM _v	0.03	0.06	0.06	0.08	0.008	
GAM-CV	r_v	-0.27	0.17	0.18	0.60	0.12	12
	RECM _v	0.03	0.05	0.05	0.08	0.008	
MARS	r_v	-0.33	0.10	0.11	0.55	0.14	11
	RECM _v	0.03	0.06	0.06	0.09	0.008	
MARS-CV	r_v	-0.07	0.31	0.31	0.65	0.10	9
	RECM _v	0.03	0.05	0.05	0.08	0.007	
RF	r_v	-0.17	0.17	0.18	0.52	0.05	40
	RECM _v	0.03	0.05	0.05	0.08	0.007	
RF-CV	r_v	-0.27	0.12	0.12	0.50	0.11	8
	RECM _v	0.03	0.05	0.05	0.08	0.007	
RF-CV _{IMP}	r_v	-0.04	0.26	0.26	0.53	0.08	10
	RECM _v	0.03	0.05	0.05	0.07	0.006	
SVM	r_v	-0.36	0.07	0.07	0.47	0.12	40
	RECM _v	0.03	0.05	0.05	0.08	0.007	
SVM-CV	r_v	-0.26	0.11	0.12	0.50	0.12	7
	RECM _v	0.06	0.08	0.08	0.09	0.003	

TABLA 3.25: Modelos obtenidos a escala local para arcilla ($\text{kg} \cdot \text{kg}^{-1}$) a escala local. Estadísticos descriptivos de la correlación (r_v) y Raíz del Error Cuadrático Medio (RECM_v) obtenidos en el remuestreo del conjunto de validación.

MARS-CV y los modelos con Random Forest son los que menos error arrojan.

Teniendo en cuenta lo expuesto, se ha seleccionado MARS-CV como mejor modelo a escala local, que además utiliza solamente 7 variables predictoras para llegar a este resultado (además de las variables cualitativas del modelo GLM), frente las 8 y 10 variables que utilizan RF-CV y RF-CV_{IMP} con peor resultado predictivo.

Las variables incluidas en el modelo son mayoritariamente geomorfométricas; por orden de importancia, estas serían el índice MRVBF, la radiación de junio RAD25_6, la precipitación total de julio (AGprectotPREDFIN7), la componente este de la orientación (SIN25_39), la precipitación total de septiembre (AGprectotPREDFIN9), la radiación de diciembre RAD25_12, y la curvatura plana (PLAN25_7). Además, el modelo utiliza 11 funciones básicas.

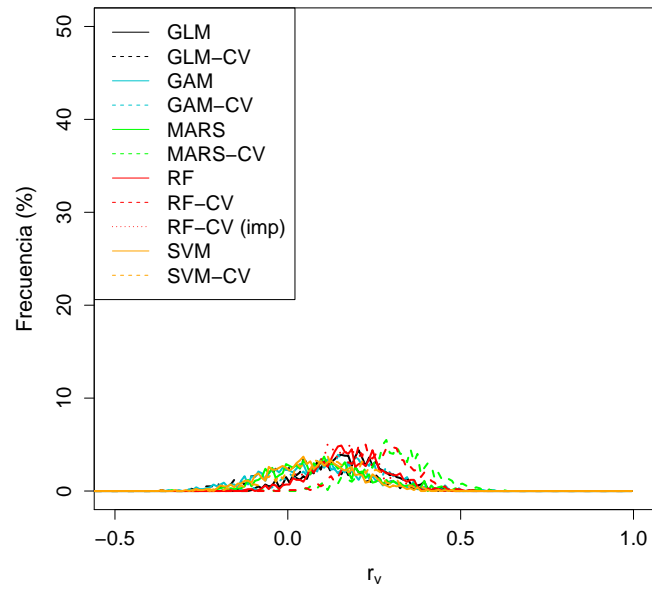
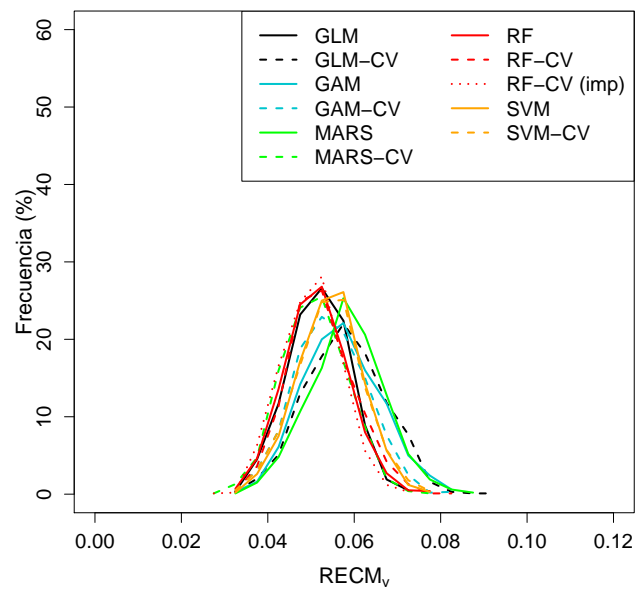
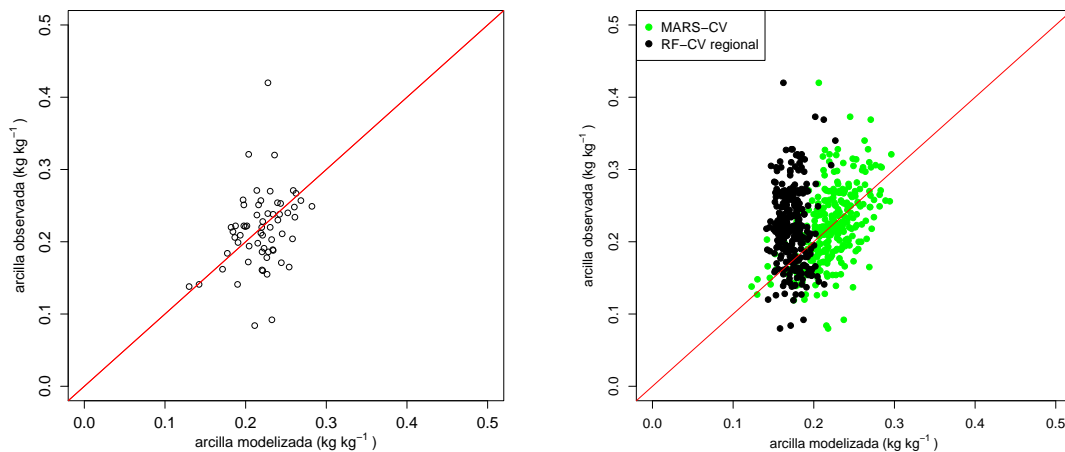
(A) Correlación (r_v)(B) Raíz del error cuadrático medio ($RECM_v$)

FIGURA 3.70: Modelos obtenidos a escala local para arcilla a escala local. Histogramas de la correlación (r_v) y la raíz del error cuadrático medio ($RECM_v$) del remuestreo en el conjunto de validación.

El ajuste del modelo se puede observar en la figura 3.71A, donde aunque el valor medio es muy similar en los datos observados y predichos por el modelo MARS-CV (0.216 y 0.22 $\text{kg} \cdot \text{kg}^{-1}$ respectivamente), el rango es más estrecho en los valores de arcilla predichos que en los datos observados en el conjunto de validación, con un valor máximo predicho que es aproximadamente la mitad del valor máximo observado. Dada la importancia del contenido en arcilla en las propiedades hídricas del suelo (sobre todo de la densidad aparente y el punto de marchitamiento permanente), esto podría reflejarse una infraestimación posterior de las mismas.

En comparación con el modelo MARS obtenido por Gómez *et al.* (2011) para arcilla, este modelo obtiene una menor correlación, pero los valores de RMSE y GCV son más bajos con una mayor simplicidad (11 funciones básicas y 7 variables, frente 67 funciones básicas y 37 variables). También el modelo MARS obtenido por Arrouays *et al.* (2014) dio como resultado un valor de correlación ligeramente mayor, en su caso para el conjunto de Europa.



(A) MARS-CV a escala local (conjunto de validación)

(B) Comparación entre el modelo MARS-CV (Mojantes) y el mapa con RF-CV a escala regional

FIGURA 3.71: Diagramas de dispersión de los datos de arcilla ($\text{kg} \cdot \text{kg}^{-1}$) observados y estimados a escala local y regional.

Finalmente, hay que señalar que la predicción del modelo a escala regional para arcilla tiene un ajuste escaso a los datos de la sierra de Mojantes, con una correlación (para las 296 muestras) de solamente 0.001, frente el valor de 0.51 que se obtiene con MARS-CV en todas las muestras. Esto podría deberse al estrecho rango de la predicción realizada por el modelo RF-CV a escala regional, que se muestra en las figuras 3.71 y 3.72 resultado de la promediación y a haber sido calibrado en un rango de valores diferente. Además, con el modelo RF-CV a escala regional se produce una infraestimación del contenido de arcilla en la zona, que según el test de Wilcoxon sería significativa (p -valor < 0.05), mientras que el modelo MARS-CV da como resultado un rango de valores más similar al observado y según este mismo test no existen diferencias significativas. El test de la t de Welch confirma el resultado del test de Wilcoxon para ambos modelos.

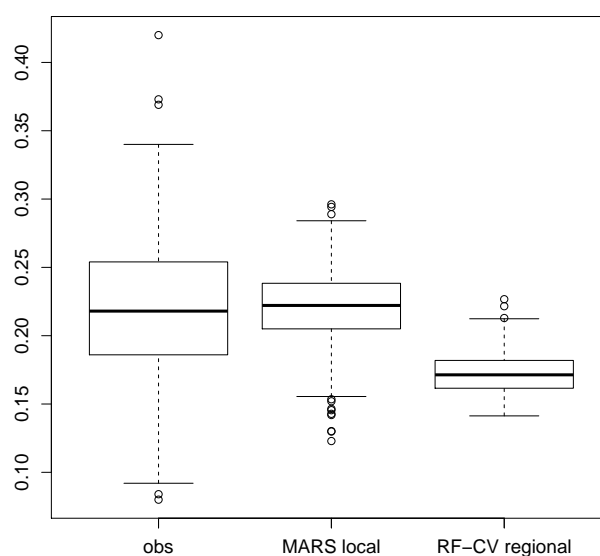


FIGURA 3.72: Gráfico de cajas con la distribución del contenido de arcilla ($\text{kg} \cdot \text{kg}^{-1}$) a escala local. **obs**: valores observados; **MARS local**: predicción del modelo MARS-CV a escala local; **RF-CV regional**: mapa elaborado con la predicción de RF-CV a escala regional y los residuos interpolados con KO.

3.3.4 Limo

3.3.4.1 Análisis descriptivo

Como en la base de datos a nivel regional, se ha dividido el valor de la base de datos entre 100, para expresarlo en tanto por uno ($\text{kg} \cdot \text{kg}^{-1}$). Los estadísticos descriptivos del limo (tabla 3.26) indican un rango más estrecho que en los datos regionales, sobre todo debido al mayor valor mínimo, aunque los valores de la mediana y la media son ligeramente superiores; por otro lado, el contenido de limo es menos disperso y más homogéneo, atendiendo a la menor desviación estándar y el menor coeficiente de variación.

Aunque la distribución no es normal (test de Shapiro-Wilk, $W = 0.9866$, $p\text{-valor} = 0.007791$), la figura 3.73 muestra un histograma simétrico, por lo que no se ha realizado ninguna transformación. Los modelos se han elaborado utilizando la familia "binomial".

Mín.	Mediana	Media	Máx.	DS	CV(%)
0.25	0.46	0.47	0.73	0.08	18.16

TABLA 3.26: Estadísticos descriptivos del contenido de limo ($\text{kg} \cdot \text{kg}^{-1}$) en los suelos de la sierra de Mojantes. **Mín.**= Mínimo; **Máx.**= Máximo; **SD**= Desviación estándar; **CV (%)**. = Coeficiente de variación.

Respecto a la distribución por usos, el limo es una variable homocedástica (test de Bartlett, $K^2 =$

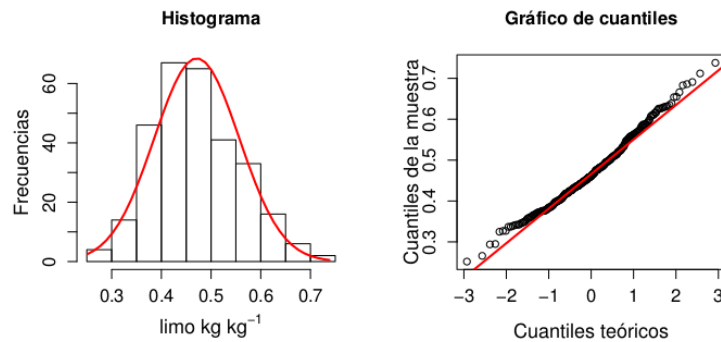


FIGURA 3.73: Histograma de frecuencias y gráfico de cuantiles de limo ($\text{kg} \cdot \text{kg}^{-1}$) en la sierra de Mojantes.

3.3428, $p\text{-value} = 0.6473$). El test ANOVA realizado indica que no hay diferencias significativas ($F = 0.071$, $p\text{-valor} = 0.79$) entre las diferentes cubiertas vegetales, aunque la figura 3.74 indica que el contenido medio de limo es mayor en los cultivos de secano (3) que en el resto; por el contrario, los valores más bajos y con un rango más estrecho se dan en los cultivos arbóreos (4).

El test de Bartlett también indica que el limo es una variable homocedástica entre los tipos de suelo ($K^2 = 9.8851$, $p\text{-valor} = 0.1952$). Sin embargo, en este caso el test ANOVA indica que las diferencias son significativas ($F = 3.668$, $p\text{-valor} = 0.0031$), aunque el test de Tukey posterior muestra que esta diferencia solo existe entre los Calcisoles Háplicos (201) y Pétricos (205). El contenido de limos en estos últimos tiene un rango más bajo que en los demás tipos de suelos, a pesar de los valores anómalos altos, pero los rangos se superponen en casi todos (figura 3.75). En el caso de los Leptosoles Réndzicos (805) el rango es especialmente amplio.

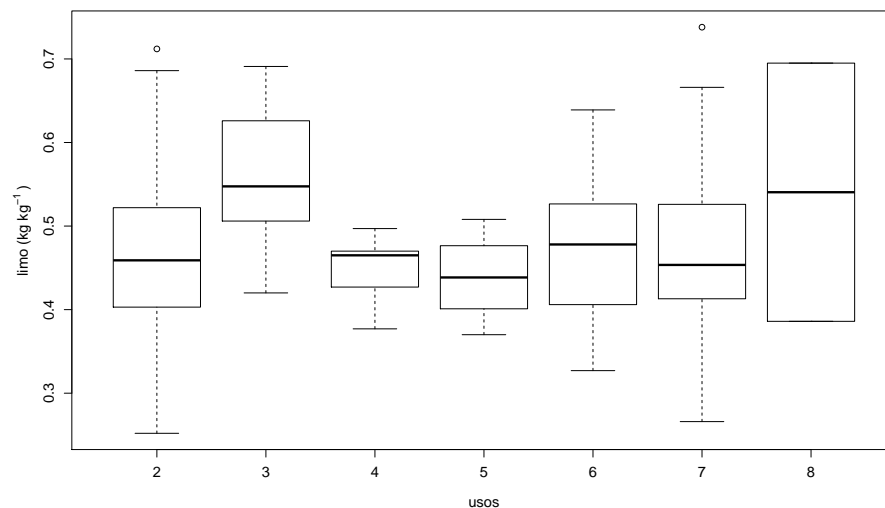


FIGURA 3.74: Análisis exploratorio: Contenido de limo ($\text{kg} \cdot \text{kg}^{-1}$) en función del uso del suelo

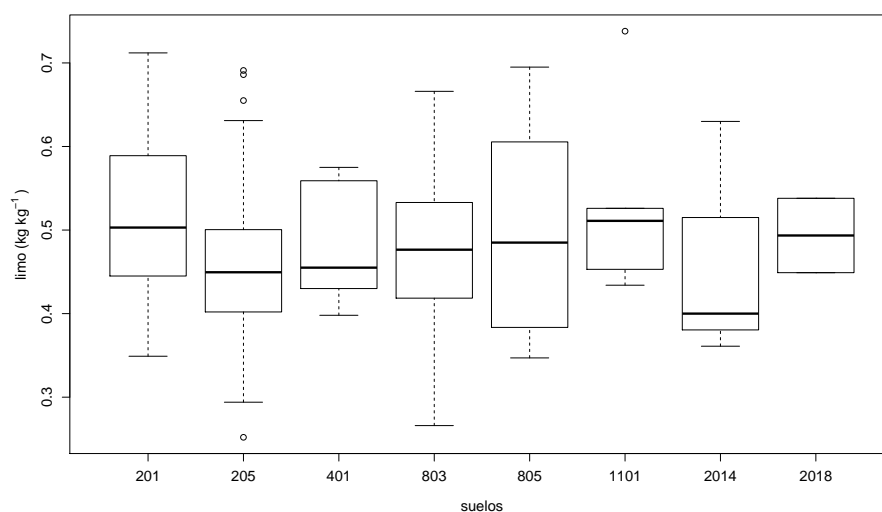


FIGURA 3.75: Análisis exploratorio: Contenido de limo ($\text{kg} \cdot \text{kg}^{-1}$) en función del tipo de suelo

En relación a las variables cualitativas, de las correlaciones detectadas a escala regional solo se mantienen las existentes con el coseno de la orientación a dos escalas de cálculo (COS25_15 y COS25_39), y con el índice MRVBF, aunque a escala local estas tienen menor significación.

Variables geomorfológicas			
Variable	r (Pearson)	Variable	r (Pearson)
SLO25_39	-0.02	SIN25_3	0.03
SIN25_39	0.01	COS25_3	-0.02
COS25_15	-0.12*	COS25_39	-0.14*
PROF25_3	0.01	PROF25_7	0.06
PROF25_11	0.03	PROF25_23	0.06
PROF25_39	0.04	PLAN25_3	0.01
PLAN25_7	-0.05	PLAN25_11	-0.004
PLAN25_15	0.005	PLAN25_19	-0.02
PLAN25_23	0.04	PLAN25_27	0.023
PLAN25_31	0.022	PLAN25_35	0.01
PLAN25_39	0.04	VRM11	-0.008
VRM39	0.009	IWA25_27	0.04
IWA25_39	0.04	TPI25_3	0.06
MRVBF	0.16**	LS	-0.022
Variables climáticas		Variables climáticas	
Variable	r (Pearson)	Variable	r (Pearson)
AGprectotPREDFIN4	-0.06	AGprectotPREDFIN7	-0.06
AGprectotPREDFIN9	-0.06	AGprectotPREDFIN10	0.05
AGtemedPREDFIN7	-0.01	AGtemaxabsPREDFIN8	-0.007
AGteminabsPREDFIN12	0.06	AGetHPREDFIN9	0.04
RAD25_6	-0.01	RAD25_12	-0.02

TABLA 3.27: Correlaciones del contenido en limo ($\text{kg} \cdot \text{kg}^{-1}$) con las variables predictoras cuantitativas en la sierra de Mojantes. Los valores de correlación significativos se han resaltado en negrita con fondo gris (p-valor: * < 0.05; ** < 0.01; *** < 0.001).

3.3.4.2 Modelos

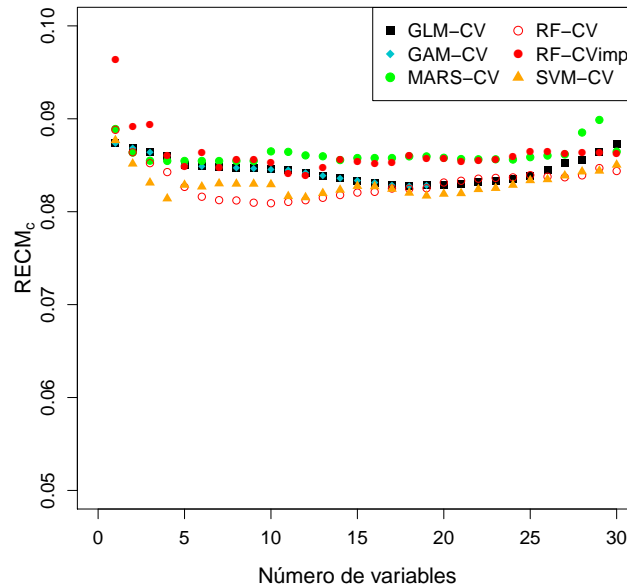


FIGURA 3.76: Raíz del error cuadrático medio $RECM_c$ de los modelos obtenidos con validación cruzada en bloques para limo a escala local, en función del número de variables.

La figura 3.76 muestra la evolución del $RECM_c$ en el conjunto de calibración, al introducir mayor cantidad de predictores en el proceso de validación cruzada. Se puede observar que los modelos GLM-CV y GAM-CV obtienen el mismo error, y sus gráficas se superponen, aunque en el caso de GAM-CV solo se han alcanzado las 20 variables. En estos modelos, el error $RECM_c$ disminuye lentamente al aumentar la complejidad, y a partir de un determinado número de predictores, la adición de más variable produce un aumento del error. El modelo MARS-CV parte del mismo valor de $RECM_c$ inicial, pero al aumentar el número de variables no se reduce sino que aumenta el error producido.

Respecto los modelos con RF y validación cruzada, el resultado es diferente al obtenido en los modelos a escala regional: en aquellos, eran los que más disminuían este error. Sin embargo, a escala local el modelo RF-CV parte de un error similar al obtenido con GLM-CV en el modelo con una sola muestra. disminuyendo rápidamente y situándose su gráfica por debajo de las restantes hasta alcanzar las 10 variables, y partir de ahí, la inclusión de variables adicionales produce un aumento del error $RECM_c$; el modelo SVM-CV no consigue superar esta disminución del error.

Por su parte, el modelo RF-CV_{IMP} tiene un comportamiento irregular: es el que mayor error produce con los modelos más sencillos, disminuyendo ligeramente al alcanzar las 5 variables, pero sin llegar a alcanzar valores inferiores a los obtenidos con GLM-CV y GAM-CV, mientras

que a partir de 12 variables el error $RECM_c$ aumenta de nuevo, hasta alcanzar el valor obtenido por MARS-CV.

El número final de variables introducidas en cada modelo se ha incluido en la tabla 3.28, que muestra además los estadísticos significativos de la correlación y el RECM de cada modelo tras el remuestreo en el conjunto de validación.

Modelo		Mín.	Mediana	Media	Max.	SD	Nº de variables
GLM	r_v	-0.23	0.24	0.24	0.58	0.13	3
	$RECM_v$	0.05	0.07	0.08	0.10	0.07	
GLM-CV	r_v	0.01	0.26	0.26	0.56	0.10	18
	$RECM_v$	0.06	0.07	0.08	0.1	0.06	
GAM	r_v	-0.06	0.27	0.27	0.60	0.1	18
	$RECM_v$	0.06	0.08	0.08	0.11	0.008	
GAM-CV	r_v	-0.23	0.21	0.21	0.55	0.12	18
	$RECM_v$	0.07	0.11	0.11	0.15	0.014	
MARS	r_v	-0.26	0.10	0.10	0.46	0.11	6
	$RECM_v$	0.06	0.09	0.09	0.12	0.008	
MARS-CV	r_v	-0.25	0.16	0.16	0.51	0.1	2
	$RECM_v$	0.07	0.09	0.09	0.11	0.006	
RF	r_v	-0.01	0.34	0.34	0.64	0.1	40
	$RECM_v$	0.05	0.0137	0.08	0.9	0.006	
RF-CV	r_v	-0.20	0.20	0.19	0.56	0.13	10
	$RECM_v$	0.09	0.13	0.13	0.17	0.011	
RF-CV _{IMP}	r_v	-0.15	0.28	0.28	0.61	0.11	12
	$RECM_v$	0.06	0.07	0.08	0.1	0.006	
SVM	r_v	-0.15	0.22	0.22	0.57	0.1	40
	$RECM_v$	0.06	0.08	0.08	0.1	0.007	
SVM-CV	r_v	-0.38	0.14	0.14	0.51	0.14	4
	$RECM_v$	0.06	0.08	0.08	0.1	0.007	

TABLA 3.28: Modelos a escala local obtenidos para limo ($\text{kg} \cdot \text{kg}^{-1}$). Estadísticos descriptivos de la correlación (r) y Raíz del Error Cuadrático Medio ($RECM_v$) obtenidos en el remuestreo del conjunto de validación.

En el conjunto de validación, las correlaciones obtenidas son inferiores que las anteriormente comentadas en el modelo regional, con mínimos de signo negativo en todos los modelos, excepto GLM-CV. Los valores máximos, sin embargo, no son muy inferiores a los obtenidos en los modelos regionales, lo que resulta en un rango amplio (figura 3.77) y una baja estabilidad de los resultados. El modelo que obtiene mayores correlaciones es RF, seguido de RF-CV_{IMP}, siendo RF-CV el que peor resultado ofrece de los obtenidos con RandomForest. Por el contrario, los modelos con MARS son los que obtienen las menores correlaciones, seguidos por los elaborados utilizando SVM.

Los valores de $RECM_v$ son algo inferiores a los obtenidos con los modelos a escala regional (figura 3.78). En este caso, no están tan solapados entre ellos, percibiéndose una diferencia clara entre RF-CV, que es el que mayores valores de error arroja, y además con un rango más amplio que el resto de modelos excepto MARS y MARS-CV, con valores de error y rangos intermedios. RF y RF-CV_{IMP} se sitúan en el extremo contrario, aunque el primero utiliza todas las variables (40) y el segundo solamente 12 variables. GLM, que solamente utiliza tres variables, arroja valores de $RECM_v$ similares a RF y RF-CV_{IMP}; también GLM-CV, GAM y GAM-CV consiguen valores de $RECM_v$ bajos, pero utilizando, respectivamente, 18, 18 y 15 variables.

Atendiendo a los valores de $RECM_v$ y correlación (r_v), y el número de variables incluidas en el

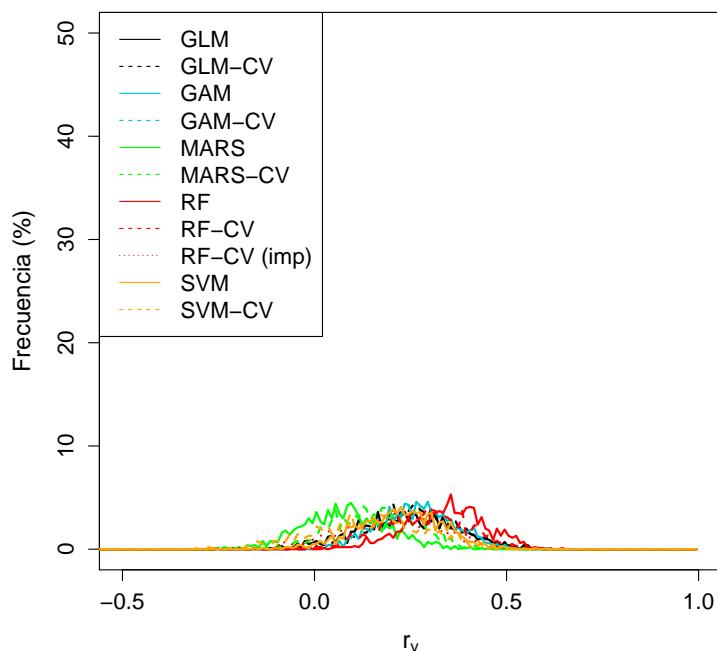


FIGURA 3.77: Modelos obtenidos con validación cruzada en bloques para limo a escala local. Correlación (r_v) tras el remuestreo en el conjunto de validación.

modelo, se ha seleccionado RF-CV_{IMP} como el que mejor resultado proporciona. Sin embargo, este modelo solo explica el 7% de la varianza total de los datos.

De las 12 variables incluidas en el modelo (figura 3.79), las de mayor importancia son variables geomorfológicas, a diferencia del modelo a escala regional, donde no aparecía ninguna de ellas. Destacan entre ellas el índice MRVBF, el factor LS, la curvatura del perfil (PROF25_39) y la rugosidad (VRM39), todas ellas relacionadas con transporte y depósito de sedimentos, junto con la pendiente y la curvatura plana. El índice de posición topográfica, aunque se incluye en el modelo, tiene mucha menor importancia, calculada con el menor tamaño de ventana (3 píxeles), por lo que reflejaría procesos geomorfológicos a una escala mucho menor que el resto. De las variables climáticas aparecen la precipitación total de abril (AGprectotPREDFIN4) y julio (AGprectotPREDFIN7), la humedad relativa de septiembre (AGetHPREDFIN9) y la radiación total de junio (RAD25_6) a diferencia del modelo a escala regional donde había una predominancia de variables climáticas. El tipo de suelo, que a escala regional era la variable más importante, pasa a ser la de menor importancia en esta zona.

La inclusión de estas variables en un nuevo modelo GAM no solo no mejora la correlación obtenida anteriormente con GAM-CV, sino que por el contrario el valor medio de la correlación es considerablemente menor (0.13), a diferencia de lo que ocurría a escala regional. Por otro lado, ninguna de las variables incluidas en el modelo son significativas.

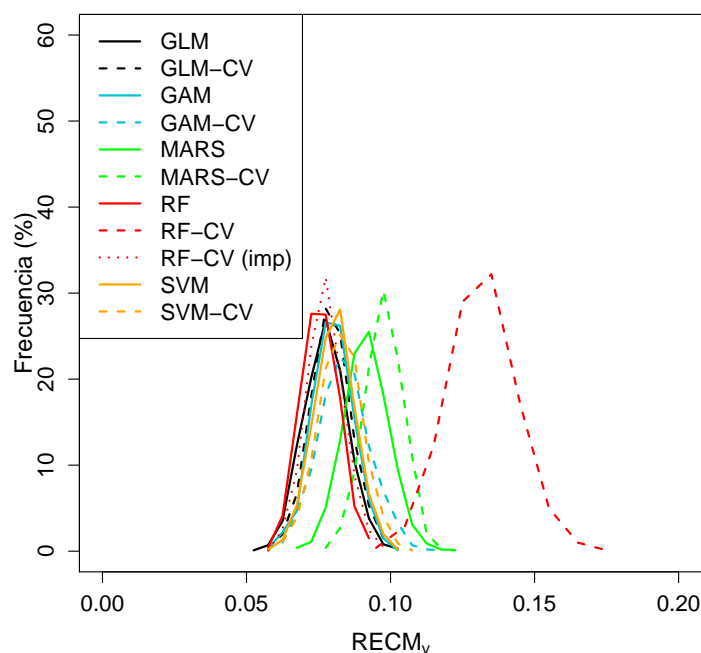


FIGURA 3.78: Modelos obtenidos con validación cruzada en bloques para limo a escala local. Raíz del error cuadrático medio ($RECM_v$) tras el remuestreo en el conjunto de validación.

Tal como se observa en la figura 3.80, el rango de limo predicho por el modelo a escala local es más estrecho ($0.34\text{-}0.62 \text{ kg} \cdot \text{kg}^{-1}$) que el observado en los datos ($0.25 - 0.74 \text{ kg} \cdot \text{kg}^{-1}$); de este modo, aunque el valor medio es en ambos casos de $0.47 \text{ kg} \cdot \text{kg}^{-1}$, la desviación estándar en los datos observados es el doble que en los datos predichos (0.08 frente 0.03). Con el mapa del modelo a escala regional (regresión+krigeado), el rango es aún más estrecho ($0.36\text{-}0.51 \text{ kg} \cdot \text{kg}^{-1}$) y la media es ligeramente menor (0.45), al igual que con el mapa resultante de restar al total la suma de arena y arcilla. Estos dos últimos muestran una correlación muy escasa, y no significativa, con los valores observados de limo a escala local.

La figura 3.81 indica que, a pesar de las diferencias en la amplitud de los rangos mencionada en el párrafo anterior, y la escasa correlación, las medianas son muy similares. El test de Wilcoxon muestra que hay diferencias significativas entre los valores observados de limo y los valores calculados tras restar el contenido de arenas y arcillas de la granulometría total ($p\text{-valor} = 0.01$), pero no con el modelo a escala local ($p\text{-valor} = 0.67$), ni tampoco con el modelo a escala regional tras interpolar los residuos ($p\text{-valor} = 0.01$), que es el que menor proporción de limos predice. En cambio, el test de la t de Welch, más sensible que el anterior, si encuentra diferencias significativas entre los datos observados y los reflejados en ambos mapas a escala regional ($p\text{-valor} = 0.01$).

Por lo tanto, los modelos obtenidos para limo a escala regional no reflejan la variabilidad de este constituyente a escala local, pero puede usarse su predicción como base para estimar el

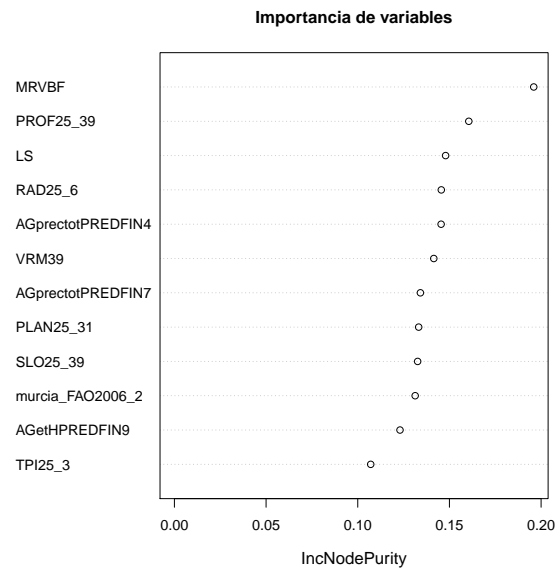
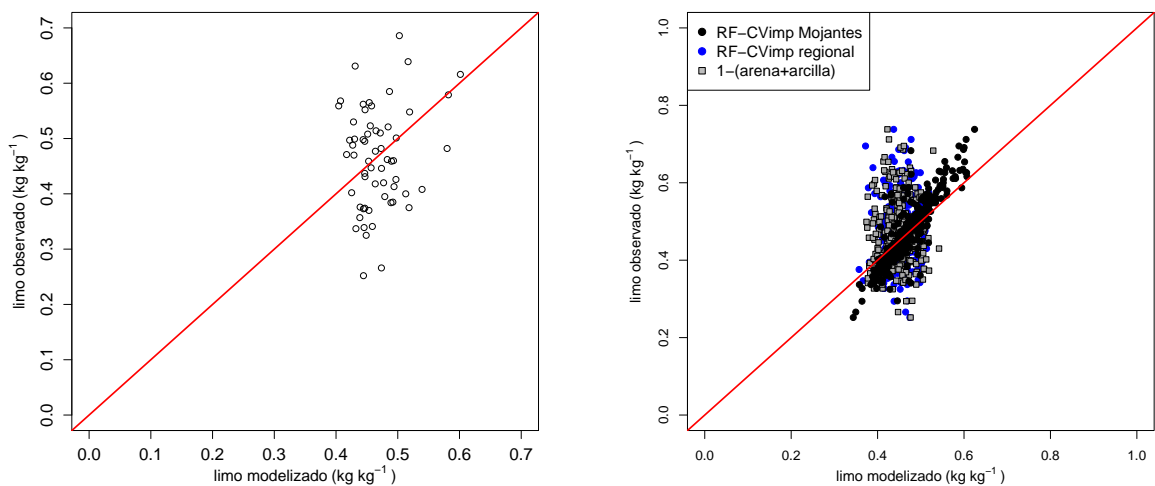


FIGURA 3.79: Importancia de variables en el modelo RF-CV_{IMP} para limo a escala local, expresada como criterio de impureza del nodo.



(A) Ajuste del modelo en el conjunto de validación

(B) Comparación entre los modelos RF-CV_{imp} a diferentes escalas y el limo estimado en el conjunto de datos

FIGURA 3.80: Diagramas de dispersión de los datos de limo $\text{kg}\cdot\text{kg}^{-1}$ observados frente los obtenidos con RF-CV_{IMP} a escala local y comparación con el resultado del mapa a escala regional.

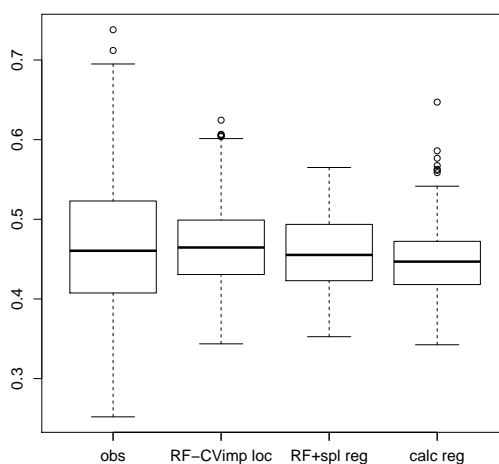


FIGURA 3.81: Gráfico de cajas con la distribución del contenido de limo ($\text{kg} \cdot \text{kg}^{-1}$) a escala local. **obs**: valores observados; **RF-CVimp loc**: predicción del modelo RF-CV_{IMP} a escala local; **RF+spl reg**: mapa elaborado con la predicción de RF-CV_{IMP} a escala regional y los residuos interpolados con splines; **calc reg**: Valores calculados restando al total la suma de arena y arcilla (a escala regional).

contenido de dicha fracción.

3.3.5 Arena

3.3.5.1 Análisis descriptivo

El contenido de arena se ha expresado en $\text{kg} \cdot \text{kg}^{-1}$ (tanto por uno), dividiendo entre 100 el valor de la base de datos. En la sierra de Mojantes, el rango de contenidos de arena es más estrecho que en el conjunto de la región, con un máximo de $0.64 \text{ kg} \cdot \text{kg}^{-1}$, y el contenido medio de arena menor (0.30 frente $0.36 \text{ kg} \cdot \text{kg}^{-1}$), tal como indica la tabla 3.14. En este caso, el coeficiente de variación es también menor, indicando una mayor homogeneidad de los datos.

Mín.	Mediana	Media	Máx.	DS	CV(%)
0.03	0.32	0.31	0.64	0.1	33.12

TABLA 3.29: Estadísticos descriptivos del contenido de arena ($\text{kg} \cdot \text{kg}^{-1}$) en los suelos de la sierra de Mojantes. **Mín.** = Mínimo; **Máx.** = Máximo; **SD** = Desviación estándar; **CV (%)** = Coeficiente de variación.

Aunque el test de Shapiro-Wilk indica que la arena no tiene una distribución normal ($W = 0.9828$, $p\text{-valor} = 0.001243$), el histograma es simétrico (figura 3.40) y no se ha realizado ninguna transformación. Los modelos se han elaborado utilizando la familia "binomial".

El contenido en arena es homocedástico respecto los distintos usos (test de Bartlett, $K^2 = 7.8349$, $p\text{-valor} = 0.09782$). por lo que se ha realizado un test ANOVA. El análisis de varianza indica que

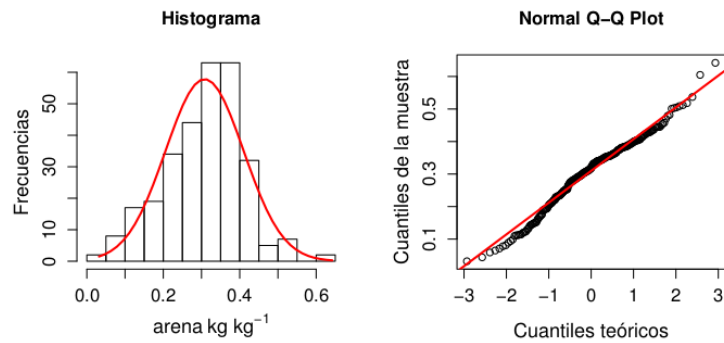


FIGURA 3.82: Histograma y gráfico de cuantiles de arena ($\text{kg} \cdot \text{kg}^{-1}$) en la sierra de Mojantes

no hay diferencias entre las medias de los grupos ($F=0.66$; $p\text{-valor}=0.41$), y de hecho en la figura 3.83 se observa que, de forma análoga a los datos de la región, el contenido en arena es similar en todos los usos del suelo, aunque con un rango más amplio y menor valor de mediana en los cultivos de regadío (3) y más estrecho en los suelos con cultivo arbóreo (4). También a escala local se puede atribuir esta diferencia a que en los suelos de regadío, normalmente situados en posiciones topográficas más bajas y cercanas a los cauces, predominan las granulometrías más finas, mientras que los Leptosoles coincidirían con zonas más altas y abruptas donde se eliminarían las partículas de menor tamaño del horizonte superficial.

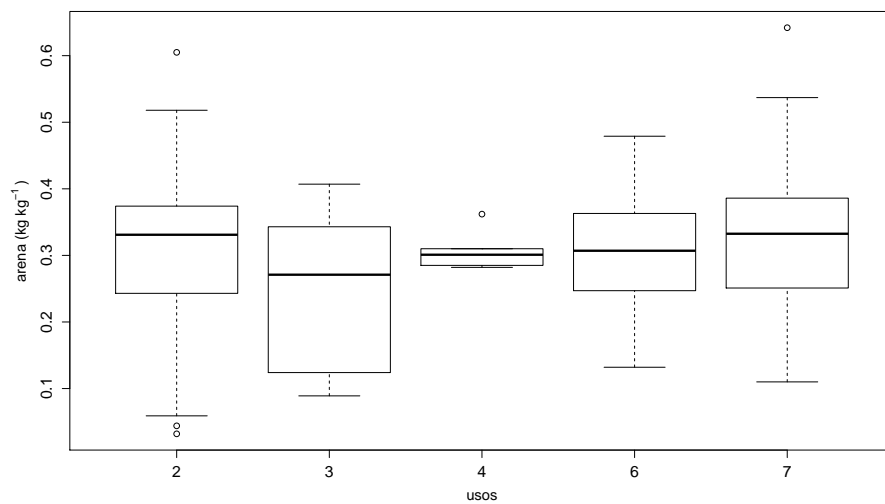


FIGURA 3.83: Análisis exploratorio: Contenido de arena en función del uso del suelo.

En cuanto a los tipos de suelo (figura 3.84), no se produce esta homogeneidad de varianzas (test de Bartlett, $K^2 = 14.34$, $p\text{-valor}= 0.01357$) por lo que se ha realizado el test no paramétrico de Kruskal-Wallis, detectándose diferencias significativas entre los diferentes grupos ($\chi^2 = 16.0953$,

p-valor = 0.006577). Tras realizar el test de Wilcoxon entre pares, los Calcisoles Háplicos (201) tienen proporciones de arena menores que los Calcisoles Pétricos (205) y los Leptosoles Háplicos (803) y, con menor nivel de significación, que los Leptosoles Réndzicos (805). Sin embargo, entre los demás grupos, estas diferencias son solo aparentes, aunque de forma similar a lo que se observaba en la región, los Fluvisoles Háplicos (401) y los Regosoles Háplicos (1101) tienen proporciones relativamente bajas de esta fracción.

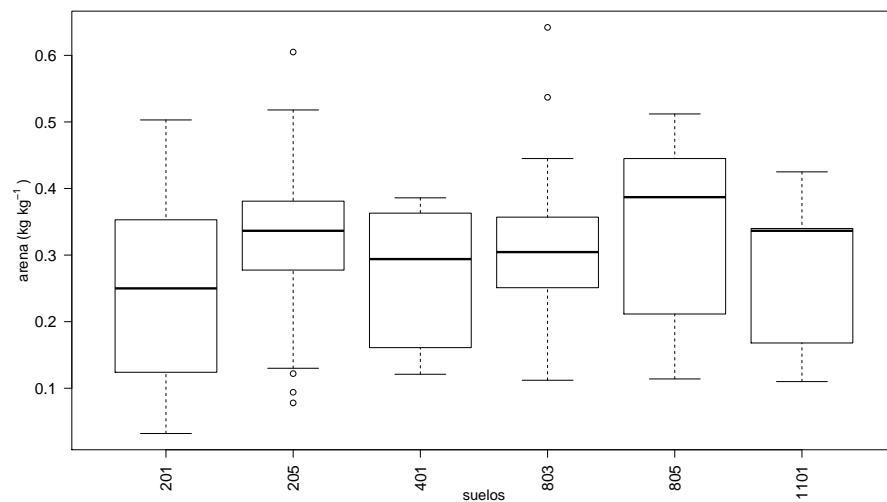


FIGURA 3.84: Análisis exploratorio: Contenido de arena en función del tipo de suelo.

Variables geomorfológicas			
Variable	r (Pearson)	Variable	r (Pearson)
SLO25_39	0.13*	SIN25_3	-0.03
SIN25_39	-0.03	COS25_3	-0.03
COS25_15	0.08	PROF25_7	-0.04
PROF25_11	-0.03	PROF25_23	-0.06
PROF25_39	-0.03	PLAN25_3	0.004
PLAN25_7	0.04	PLAN25_11	-0.04
PLAN25_15	-0.04	PLAN25_19	-0.005
PLAN25_23	-0.05	PLAN25_27	-0.05
PLAN25_31	-0.07	PLAN25_35	-0.05
PLAN25_39	-0.05	VRM11	0.03
VRM39	0.06	IWA25_27	-0.04
IWA25_39	-0.03	TPI25_3	-0.05
MRVBF	-0.22***	LS	0.03
Variables climáticas			
Variable	r (Pearson)	Variable	r (Pearson)
AGprectotPREDFIN4	0.01	AGprectotPREDFIN7	0.02
AGprectotPREDFIN9	0.003	AGprectotPREDFIN10	-0.006
AGtemedPREDFIN7	0.009	AGtemaxabsPREDFIN8	0.04
AGteminabsPREDFIN12	-0.02	AGetHPREDFIN9	-0.12*
RAD25_6	-0.05	RAD25_12	0.06

TABLA 3.30: Correlaciones del contenido en arena ($\text{kg} \cdot \text{kg}^{-1}$) con las variables cuantitativas en la sierra de Mojantes. Los valores de correlación significativos se han resaltado en negrita con fondo gris (p-valor: * < 0.05; ** < 0.01; *** < 0.001).

En cuanto a las variables cuantitativas, de las correlaciones observadas a escala regional solo se mantienen la pendiente, con signo positivo, el índice MRVBF con signo negativo, y la humedad relativa del mes de septiembre, también con signo negativo.

3.3.5.2 Modelos

El número de variables a incluir en cada modelo viene determinado por la figura 3.85, donde se aprecia la evolución del error en el conjunto de calibración ($RECM_c$). En ella se observa que el modelo MARS-CV mantiene un valor de error mayor que el resto de modelos a pesar de la inclusión de nuevas variables, dibujando una gráfica recta. En los demás modelos sí hay una disminución del error hasta llegar a un mínimo en torno a las 6 u 8 variables en todos ellos, a partir del cual la adición de variables no mejora la predicción. En el caso de GLM-CV y GAM-CV, cuyas gráficas se superponen, el aumento en la complejidad a partir del modelo de 10 variables produce un aumento del error hasta casi alcanzar el valor obtenido con MARS-CV. Los modelos con RandomForest (RF-CV y RF-CV_{IMP}) parten de un valor inicial de error similar al de MARS-CV pero disminuyen rápidamente a partir de la segunda variable introducida; estos modelos mantienen una evolución paralela, junto con SVM-CV, ya que tras alcanzar el mínimo en el valor de $RECM_c$, la adición de nuevas variables produce un ligero aumento del error, casi imperceptible.

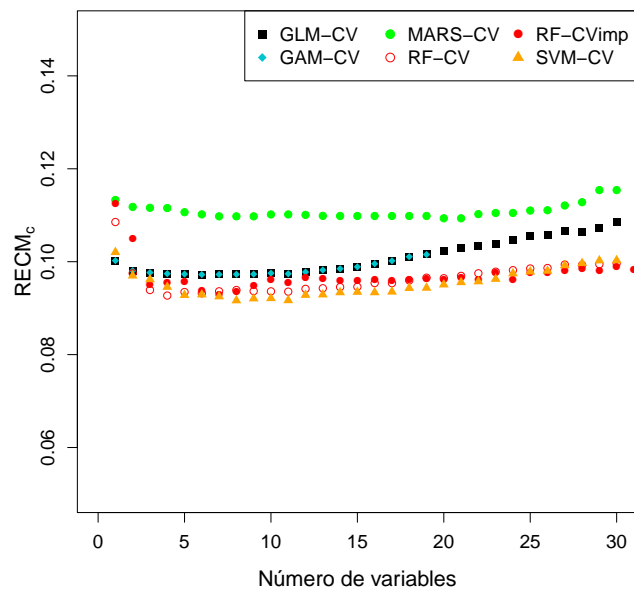


FIGURA 3.85: Raíz del error cuadrático medio ($RECM_c$) de los modelos obtenidos con validación cruzada en bloques para arena a escala local, en función del número de variables.

En el caso de GLM, el procedimiento de regresión por pasos dio lugar a la eliminación de

todas las variables, por lo que se seleccionó un modelo alternativo con las que habían resultado con correlaciones significativas en el análisis descriptivo previo. En el caso de GAM, el modelo incluye estas mismas variables, ya que el resultado era mejor que el obtenido con las variables de GLM-CV.

La tabla 3.31 muestra los estadísticos descriptivos de la correlación y el error (RECM_v), para cada uno de los modelos elaborados tras realizar el remuestreo en el conjunto de validación, cuyos histogramas se han representado en la figura 3.86. La tabla también indica el número de variables que cada modelo incluye.

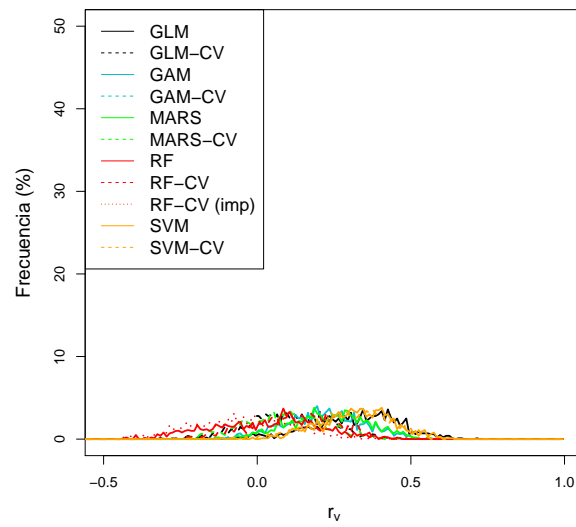
Destacan los valores mínimos negativos en la correlación de todos los modelos, especialmente en RF-CV y RF, y el valor alto de dispersión de la correlación. En el mismo sentido, estos dos modelos son los que tienen un valor medio de correlación negativo y un máximo de correlación más bajo, al contrario que en los modelos regionales donde RandomForest era el método que mejor resultado arrojaba. Esta tendencia se observa en el correspondiente histograma (figura 3.86A). El modelo RF-CV_{IMP} obtiene un resultado algo mejor, pero no supera el obtenido con MARS, que incluso utiliza una variable menos y alcanza un máximo de correlación mayor.

Modelo		Mín.	Mediana	Media	Max.	SD	Nº de variables
GLM	r_v	-0.31	0.10	0.10	0.71	0.14	4
	RECM_v	0.07	0.1	0.1	0.14	0.01	
GLM-CV	r_v	-0.38	0.10	0.10	0.51	0.13	6
	RECM_v	0.07	0.1	0.1	0.13	0.008	
GAM	r_v	-0.22	0.21	0.21	0.56	0.13	4
	RECM_v	0.07	0.1	0.1	0.12	0.007	
GAM-CV	r_v	-0.29	0.18	0.18	0.55	0.13	6
	RECM_v	0.07	0.1	0.1	0.12	0.007	
MARS	r_v	-0.27	0.23	0.23	0.62	0.13	9
	RECM_v	0.07	0.1	0.1	0.13	0.008	
MARS-CV	r_v	-0.39	0.09	0.09	0.45	0.05	8
	RECM_v	0.07	0.1	0.1	0.13	0.007	
RF	r_v	-0.43	0.05	0.05	0.46	0.17	40
	RECM_v	0.07	0.1	0.1	0.13	0.009	
RF-CV	r_v	-0.51	-0.04	-0.04	0.46	0.16	6
	RECM_v	0.07	0.11	0.11	0.14	0.009	
RF-CV _{IMP}	r_v	-0.27	0.13	0.12	0.48	0.14	8
	RECM_v	0.07	0.1	0.1	0.13	0.008	
SVM	r_v	-0.11	0.33	0.33	0.63	0.11	40
	RECM_v	0.05	0.09	0.09	0.12	0.009	
SVM-CV	r_v	-0.14	0.32	0.31	0.60	0.12	8
	RECM_v	0.06	0.09	0.09	0.12	0.007	

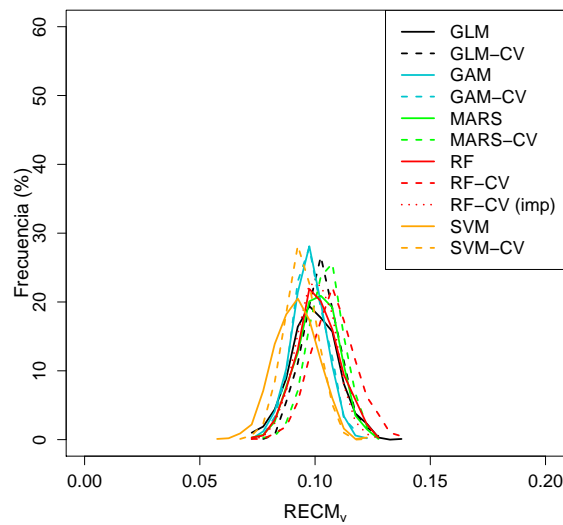
TABLA 3.31: Modelos obtenidos para arena ($\text{kg} \cdot \text{kg}^{-1}$) a escala local. Estadísticos descriptivos de la correlación (r) y Raíz del Error Cuadrático Medio (RECM_v) obtenidos en el remuestreo del conjunto de validación.

Los modelos SVM y SVM-CV son los que, a pesar de los valores mínimos aún negativos de la correlación, los valores medios son los más altos, en el caso de SVM-CV con solo 8 variables. Sin embargo, es el modelo GLM el que, con solo 4 variables, alcanza el valor máximo de correlación, aunque el valor medio sea sensiblemente inferior al obtenido con SVM-CV.

Respecto a los histogramas de RECM_v , la figura 3.86B muestra que la distribución del error es muy similar en todos los modelos obtenidos, con un valor medio aproximado de 0.1 en todos ellos. Coincidiendo con lo explicado anteriormente, el modelo RF-CV arroja los errores más



(A) Correlación



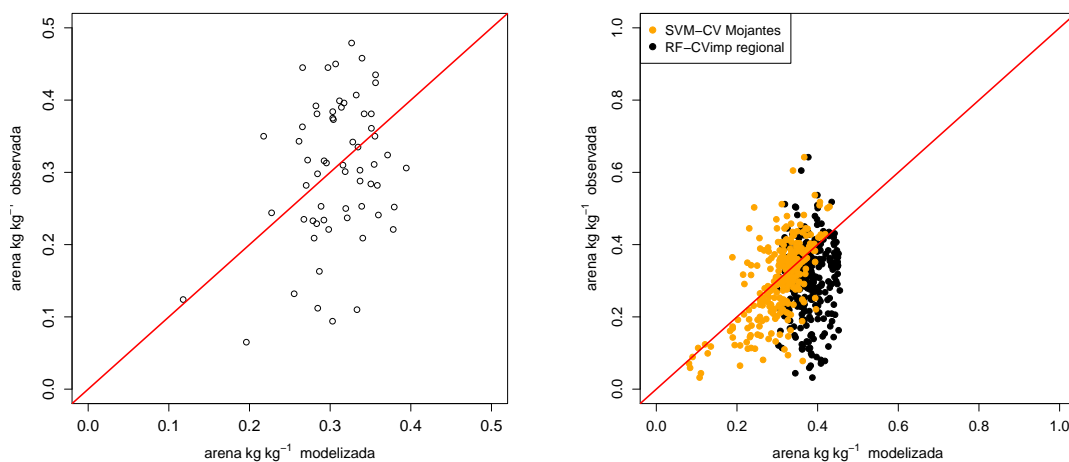
(B) Raíz del error cuadrático medio

FIGURA 3.86: Modelos obtenidos para arena a escala local. Histogramas de la correlación (r_v) y raíz del error cuadrático medio ($RECM_v$) tras el remuestreo en el conjunto de validación.

altos, mientras que los modelos utilizando Support Vector Machine son los que más reducen este parámetro. Por otro lado, MARS-CV, a pesar de obtener valores medios de correlación más altos que los elaborados con Random Forest, da como resultado valores altos de $RECM_v$ en comparación con el resto de modelos, sólo superados por RF-CV.

En resumen, atendiendo a los criterios anteriores, se ha elegido SVM-CV como el modelo que ofrece la mejor predicción del porcentaje de arena en la sierra de Mojantes.

Sin embargo, los datos predichos muestran una elevada dispersión (0.09) respecto los datos observados en el conjunto de validación (0.04), que se refleja en la figura 3.87, a pesar de que los valores medios son similares (aproximadamente $0.30 \text{ kg} \cdot \text{kg}^{-1}$). Por otro lado, aunque se produce una infraestimación de los contenidos altos de arena, la diferencia entre el máximo observado y el arrojado por el modelo es menor que en otras variables estudiadas anteriormente.



(A) Diagrama de dispersión del modelo SVM-CV en el conjunto de validación (B) Comparación entre los modelos a escala local y escala regional en el conjunto de datos

FIGURA 3.87: Diagrama de dispersión de los datos de arena ($\text{kg} \cdot \text{kg}^{-1}$) observados en el set de validación frente los obtenidos con SVM-CV a escala local y comparación con el mapa RF-CV_{IMP+KO} a escala regional.

El modelo SVM-CV incluye como variable cualitativa el tipo de suelo, siendo la mayoría de tipo geomorfométrico; estas son la curvatura plana en tres escalas de cálculo diferentes (PLAN25_39, PLAN25_31 y PLAN25_15), el índice MRVBF, y la curvatura del perfil a la menor escala de cálculo (PROF25_3). La única variable climática incluida es la radiación de junio. Es notable que de ellas, solamente el tipo de suelo y el índice MRVBF estaban relacionadas con el contenido de arena de forma significativa. Además, el tipo de suelo, que mantiene su importancia al ser la primera variable seleccionada en el modelo SVM-CV, junto con el MRVBF, aparecían también en el modelo a escala regional.

La curvatura plana y la curvatura del perfil reflejan procesos relacionados con el flujo de agua y sedimentos. Sin embargo, las curvaturas aparecen calculadas a diferentes escalas en el modelo, lo que refleja a su vez las diferentes escalas de los procesos. Por lo general, se prevé mayor porcentaje de arenas (sedimentos más gruesos) en las zonas donde las fracciones más finas hayan sido erosionadas y transportadas a otros lugares, esto es, en las curvaturas convexas que en las cóncavas, donde se se produce acumulación de sedimentos, tanto de las zonas más altas como por la convergencia de diferentes direcciones de flujo.

Al comparar el rendimiento del modelo SVM-CV a escala local con el modelo RF-CV_{IMP} obtenido anteriormente a escala regional, utilizando todos los datos (figura 3.87B), se observa que muchos de los puntos en el modelo SVM-CV se ajustan a la recta 1:1. Esto refleja un elevado ajuste del modelo a los datos usados en la calibración, confirmado por el aumento de la correlación de 0.65, mayor que el valor de 0.39 en el conjunto de validación. Los test realizados para comparar (test de la t de Welch y test de Wilcoxon) indican que no existen diferencias significativas entre las medias y las medianas de ambos rangos de valores ($p\text{-valor} > 0.05$).

En el caso de los valores del modelo elaborado a escala regional, la correlación con los valores observados es menor al calcularla con todos los datos que al utilizar solo los datos de validación (0.01 y 0.06 respectivamente). El rango de valores de la predicción es más estrecho, oscilando entre 0.24 y 0.48 $\text{kg} \cdot \text{kg}^{-1}$, y la media considerablemente mayor (0.38 $\text{kg} \cdot \text{kg}^{-1}$) que en los datos observados, produciéndose una sobreestimación del contenido de arena, reflejado en la figura 3.88. El test de la t de Welch indica también que las diferencias encontradas entre las medias son significativas ($p\text{-valor} < 2.2 \cdot 10^{-16}$), así como las encontradas entre las medianas (test de de Wilcoxon, $p\text{-valor} < 2.2 \cdot 10^{-16}$). Aún así, el contenido predicho por este modelo se encuentra dentro del rango de los valores observados.

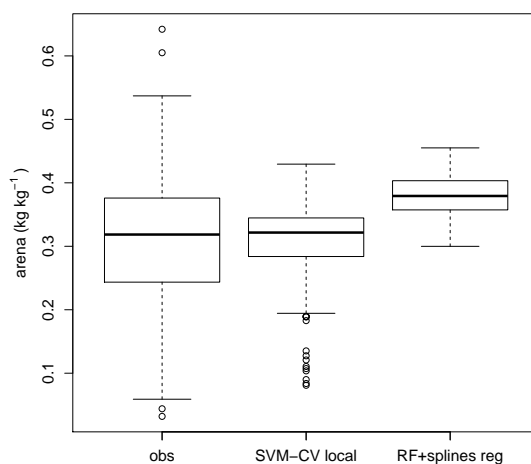


FIGURA 3.88: Gráfico de cajas con la distribución del contenido de arena ($\text{kg} \cdot \text{kg}^{-1}$) a escala local. **obs**: valores observados; **SVM-CV local**: predicción del modelo SVM-CV a escala local; **RF+splines reg**: mapa elaborado con la predicción de RF-CV_{IMP} a escala regional y los residuos interpolados con splines.

Capítulo 4

Funciones de edafotransferencia

4.1 Densidad aparente

La densidad aparente en la base de datos disponible en la sierra de las Moreras tiene un valor medio de $1.27 \text{ g} \cdot \text{cm}^{-3}$, variando este entre 1.07 y $1.45 \text{ g} \cdot \text{cm}^{-3}$ tal como se muestra en la tabla 4.1. Es, por tanto, un intervalo estrecho de valores, con una alta homogeneidad atendiendo al bajo coeficiente de variación. Además, atendiendo al histograma y el gráfico de cuantiles de la figura 4.1, la densidad aparente muestra una distribución normal (test de Shapiro-Wilk, $W = 0.9676$, p-valor = 0.2876).

Mín.	Mediana	Media	Máx.	DS	CV(%)
1.07	1.28	1.27	1.45	0.10	8.04

TABLA 4.1: Estadísticos descriptivos de densidad aparente ($\text{g} \cdot \text{cm}^{-3}$) en los suelos de la sierra de las Moreras. **Mín.**= Mínimo; **Máx.**= Máximo; **SD**= Desviación estándar; **CV (%)**. = Coeficiente de variación.

Por otro lado, la densidad aparente está significativamente correlacionada con el contenido en CO ($r=-0.34$, p-valor=0.03) y la fracción arena ($r=-0.40$, p-valor=0.009), pero no con el contenido de limo ($r=-0.18$, p-value=0.24) y arcilla ($r=-0.28$, p-value=0.06). Esto contrasta con la aparición de estas fracciones en la mayoría de las funciones de edafotransferencia. Respecto el contenido en carbonatos, la acidez y la capacidad de cambio catiónico, todas ellas relacionadas con el contenido en catión Ca^{2+} que influiría positivamente en la estructura del suelo, y de forma indirecta en la densidad aparente, tampoco se han encontrado correlaciones significativas.

La densidad aparente no refleja diferencias significativas en función del uso del suelo (test de Kruskal-Wallis, $\chi^2 = 1.6321$, p-valor=0.803), aunque en la figura 4.2 se puede ver que esta

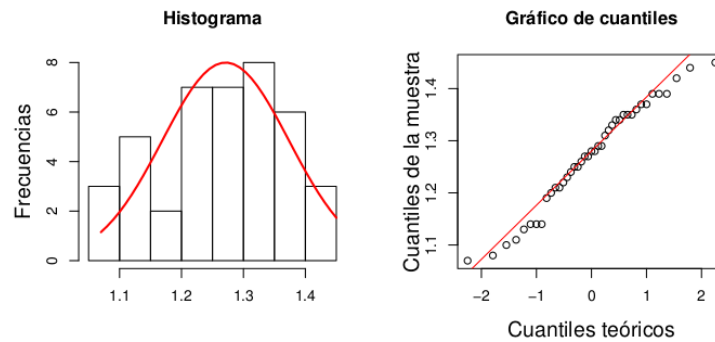


FIGURA 4.1: Histograma y gráfico de cuantiles de los datos de D_a ($\text{g} \cdot \text{cm}^{-3}$) en la sierra de las Moreras.

es menor en los suelos de regadío que en los suelos con vegetación natural de matorral. Este resultado es contrario al descrito en la introducción (apartado 1.2.4) ya que en los suelos agrícolas, con menor cantidad de materia orgánica, la densidad aparente debería ser mayor; sin embargo, este menor valor de D_a puede atribuirse al laboreo más intensivo en la capa superficial, que no explica sin embargo los valores mayores en los suelos con cultivo en mosaico. En cambio, el valor más bajo de D_a en los suelos desnudos, respecto a los suelos bajo cubierta de matorral puede atribuirse a que se corresponden con suelos ricos en arenas y a canchales con escasa fracción fina (menor de 2 mm de diámetro). El valor más bajo de densidad aparente en los suelos de bosque carece de representatividad al haber solo una muestra bajo este uso.

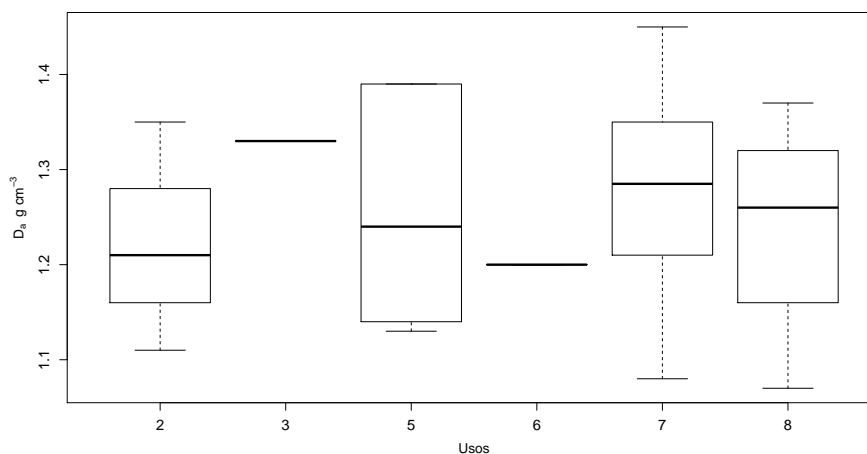


FIGURA 4.2: Distribución de la densidad aparente ($\text{g} \cdot \text{cm}^{-3}$) en función de los usos del suelo en la sierra de las Moreras

Tampoco existen diferencias respecto al tipo de suelo (test de Kruskal-Wallis, $\chi^2 = 9.28$, $p\text{-valor}=0.16$), aunque el valor medio en los Fluvisoles era menor que en el resto de suelos (1.14).

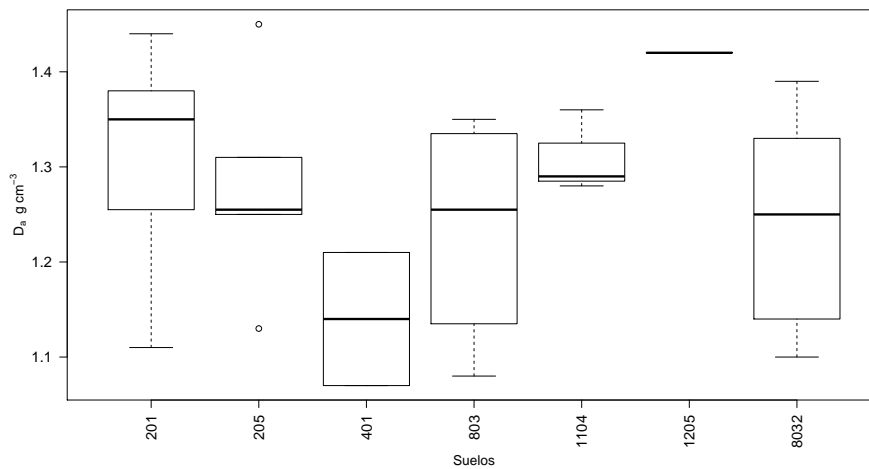


FIGURA 4.3: Distribución de la densidad aparente ($\text{g} \cdot \text{cm}^{-3}$) en función de los tipos de suelo en la sierra de las Moreras.

$n=2$), mientras que en los Solonchaks se observan los mayores valores de D_a (1.42, $n=1$). Sin embargo, debido al escaso número de muestras, estas medias son poco representativas; en el resto de grupos (Calcisoles, Leptosoles y Regosoles), donde se agrupan la mayor parte de las muestras, los valores de D_a son intermedios (figura 4.3).

Tras realizar una revisión bibliográfica, se seleccionaron 9 diferentes funciones de edafotransferencia para ser validadas en las 41 muestras de la sierra de las Moreras, cuyas fórmulas y ámbito geográfico para el que fueron elaboradas se muestran en la tabla 4.2.

Para aquellas funciones en las que aparece como variable predictora el contenido de materia orgánica, éste se ha calculado multiplicando el contenido de CO por el factor 1.72 propuesto por Broadbent (1953).

PTF	Autor	Algoritmo	Region
1	Alexander (1980)	$\rho_b = 1.66 - (0.308 \cdot \text{CO}^{0.5})$	California,EE.UU.
2	Barahona y Santos (1981)	$\rho_b = 1.5456 + (0.0015 \cdot A_p) - (0.0022 \cdot \text{Cl}_p) - (0.1219 \cdot \text{CO}_p)$	Castilla-La Mancha, España
3	Bernoux <i>et al.</i> (1998)	$\rho_b = 1.398 - (0.042 \cdot \text{CO}) - (0.0047 \cdot \text{Cl}_p)$	Amazonia, Brasil
4	Kaur <i>et al.</i> (2002)	$\ln(\rho_b) = -0.313 - (0.191 \cdot \text{CO}_p) + (0.02102 \cdot \text{Cl}_p) - (0.000476 \cdot \text{Cl}_p)^2 - (0.00432 \cdot L_p)$	Almora, India
5	Benites <i>et al.</i> (2007)	$\rho_b = 1.5688 - (0.0005 \cdot \text{Cl}) - (0.009 \cdot \text{CO})$	Brasil
6	Ruehlmann y Körschens (2009)	$\rho_b = (2.684 - 140.943 \times 0.006) \times \exp[-(0.006 \times \text{CO})]$	Diversas áreas
7	Brahim <i>et al.</i> (2010)	$\rho_b = 1.65 - (0.117 \cdot \text{CO}_p) - (0.0042 \cdot \text{Cl}_p) - (0.0036 \cdot \text{Cl}/L) + (0.031 \cdot \text{pH}_w)$	Tunisia
8a	Al-Qinna y Jaber (2013)	$\rho_b = 1.398 - 0.138 \cdot \text{CO}_p + 0.008 \cdot A_p$	Amman-Zarqa, Jordania
8b	Al-Qinna y Jaber (2013)	$\rho_b = 1.228 - (0.155 \ln \text{CO}_p) + 0.008(A_p)$	Amman-Zarqa, Jordania

TABLA 4.2: Funciones de edafotransferencia para densidad aparente. ρ_b = Densidad aparente ($\text{g} \cdot \text{cm}^{-3}$); CO= Carbono Orgánico ($\text{g} \cdot \text{kg}^{-1}$); CO_p = % Carbono orgánico; Cl = Arcilla ($\text{g} \cdot \text{kg}^{-1}$); L =Limo ($\text{g} \cdot \text{kg}^{-1}$); A = Arena ($\text{g} \cdot \text{kg}^{-1}$); Cl_p =% Arcilla; L_p =% Limo; A_p =% Arena ; Cl/L = Relación arcilla/limo; pH_w = pH en agua.

Tal como se observa en la tabla 4.3 y la figura 4.4, los valores de correlación obtenidos con estas funciones de edafotransferencia varían entre 0.34 y 0.48, siendo las elaboradas por Brahim *et al.* (2010) y Al-Qinna y Jaber (2013) aquellas que mayores correlaciones obtienen, Sin embargo,

estas mayores correlaciones no se corresponden con los menores valores de error, ya que las tres funciones tienden a una sobreestimación significativa de la densidad aparente, tal como indica el MEP, y también el valor de RECM es alto. Las PTFs obtenidas por Barahona y Santos (1981) y Benites *et al.* (2007) dan como resultado valores de correlación algo menores, aunque en este caso también se produce una sobreestimación menos acentuada de la densidad aparente.

Destaca por el lado contrario la PTF de Kaur *et al.* (2002), con un valor de correlación de 0.38, pero un error medio de predicción análogo al obtenido con la función de Barahona y Santos (1981), aunque de signo contrario, ya que esta función tiende a la infraestimación de la densidad aparente. También la elaborada por Alexander (1980) obtiene una de las correlaciones más bajas, pero consigue minimizar tanto el EMP como el RECM.

PTF	r	p-value	EMP	RECM
1	0.34	0.03	0.06	0.13
2	0.42	0.006	0.15	0.19
3	0.36	0.02	-0.48	0.59
4	0.38	0.02	-0.14	0.22
5	0.42	0.006	0.10	0.22
6	0.34	0.03	-0.43	0.45
7	0.39	0.01	0.31	0.35
8a	0.48	0.001	0.32	0.34
8b	0.48	0.001	0.32	0.35

TABLA 4.3: Resultados de las funciones de edafotransferencia para densidad aparente. **r** = correlación (Pearson); **EMP** = Error Medio de la Predicción; **RECM** = Raíz del Error Cuadrático Medio.

El mejor resultado de las PTF de Barahona y Santos (1981) y Alexander (1980) puede atribuirse a que fueron calibradas en regiones con características similares a las que se presentan en el área de estudio (Castilla-La Mancha y California); en el caso de las obtenidas por Brahim *et al.* (2010) y Al-Qinna y Jaber (2013), estas también se calibraron en zonas áridas (Túnez y Jordania), pudiendo explicarse el mayor error por la diferencia en las características edáficas, sobre todo la composición granulométrica. Por su parte, la PTF de Benites *et al.* (2007) ofrece un resultado relativamente bueno, aunque fue calibrada en suelos de Brasil, lo que podría deberse a que fue propuesta para aplicarse a un amplio rango de suelos.

La PTF seleccionada finalmente es la obtenida por Barahona, ya que además de los resultados comentados, también el propio autor describió una correlación de 0.45 y un error de 0.136. Asimismo, la validación de esta función con otras muestras de la zona han obtenido una correlación de 0.67 (Albaladejo *et al.*, 2013).

De forma paralela, se ha elaborado una función de edafotransferencia utilizando un modelo de regresión GLM y un procedimiento de validación cruzada. La regresión por pasos, con el fin de eliminar las variables no significativas del modelo, da como resultado la ecuación 4.1.

$$D_a = 0.842 + 0.0022 \cdot CO - 0.0017 \cdot A_p \quad (4.1)$$

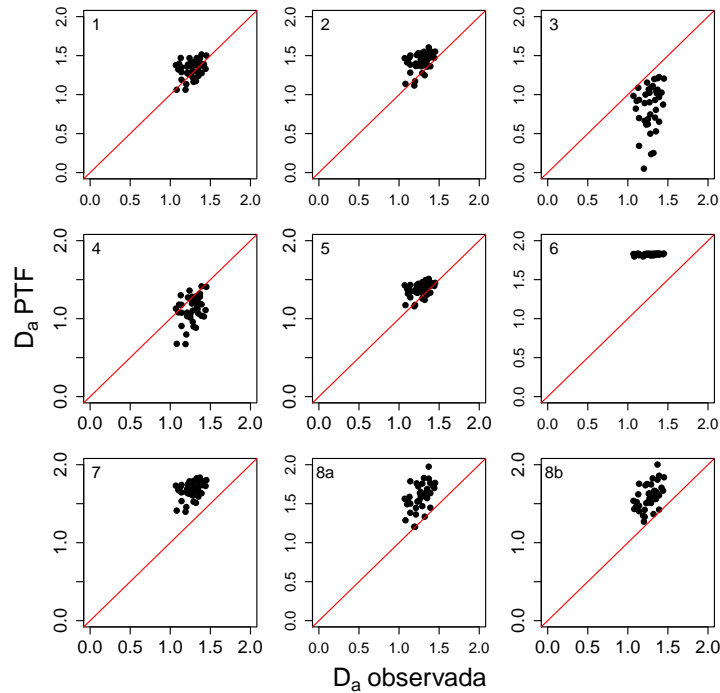


FIGURA 4.4: Diagrama de dispersión de los datos de D_a ($\text{g} \cdot \text{cm}^{-3}$) observados frente los obtenidos con las funciones de edafotransferencia recopiladas.

donde D_a es la densidad aparente ($\text{g} \cdot \text{cm}^{-3}$), CO es el contenido de carbono orgánico ($\text{g} \cdot \text{kg}^{-1}$), y A_p es la proporción de arena (%). En esta función resultan significativos los contenidos de CO y de arena, del mismo modo que ocurría con la función de [Brahim *et al.* \(2010\)](#), mientras que desaparecen en el procedimiento de regresión por pasos el contenido de arcillas que aparecía en la mayoría de las funciones anteriores, incluyendo la función de [Barahona y Santos \(1981\)](#). El resultado obtenido es mejor que el descrito con la función seleccionada anteriormente, ya que la correlación obtenida es de 0.49 (p-valor=0.001) y el RECM de $0.08 \text{ g} \cdot \text{cm}^{-3}$.

Sin embargo, al utilizar las variables modelizadas y cartografiadas anteriormente para calcular el mapa de D_a con estas PTF, se obtienen menores correlaciones y mayores valores de error. Como muestra la figura 4.5 el mapa con la función de Barahona (página A.14) produce una sobreestimación del valor de densidad aparente significativo (test de la t de Welch, $t=-9.8$, p-valor < 0.001, y una correlación mucho menor con los datos originales que la mencionada anteriormente ($r=0.19$, p-valor=0.27; RECM=0.21). El mapa obtenido con la función calculada en este trabajo (PTF-LM, no incluido en la cartografía) da como resultado una infraestimación de la densidad aparente, con una correlación negativa con los datos originales ($r=-0.15$, p-valor=0.12) y un RECM de $0.41 \text{ g} \cdot \text{cm}^{-3}$, mucho mayor que al utilizar los datos de calibración. Este resultado puede atribuirse a la propagación de errores de los modelos, y a la propia ecuación, con un valor de intercepto bajo al que se añaden un coeficiente negativo para el contenido de

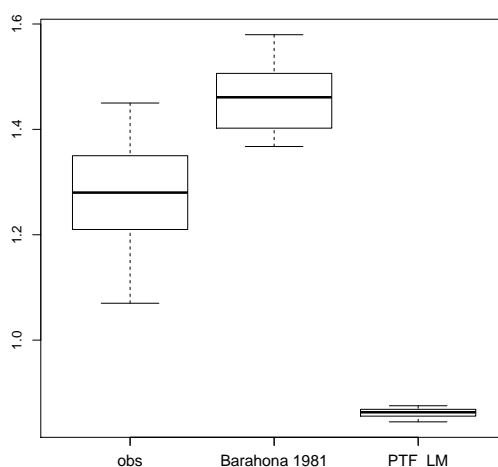


FIGURA 4.5: Gráfico de cajas con la distribución de la densidad aparente ($\text{g} \cdot \text{cm}^{-3}$) a escala local. **obs**: valores observados; **Barahona 1981**: mapa elaborado con la PTF de Barahona y Santos (1981); **PTF_LM**: mapa elaborado con la PTF elaborada con GLM-LOOCV para los datos de Mazarrón.

CO, y un coeficiente positivo en el contenido de arena. Por ello, al contrario que lo esperado y planteado en las hipótesis, las zonas con mayor cantidad de CO presentan mayores valores de Da.

4.2 Contenido de humedad

En el caso del contenido de humedad, en la base de datos de la sierra de las Moreras se dispone de valores para 149 muestras, correspondientes a los años 1991 y 2008, al eliminarse dos para las cuales no se disponía de los valores de las fracciones granulométricas.

El valor de Densidad aparente, requerido en algunas de las funciones de edafotransferencia, solo está disponible para 41 de esas muestras, por lo que ha estimado su valor para el resto utilizando la PTF de Barahona y Santos (1981). Para aquellas funciones en las que aparece como variable predictora el contenido de materia orgánica, éste se ha calculado multiplicando el contenido de CO por el factor 1.72 propuesto por Broadbent (1953).

Además, se dispone de datos de contenido de humedad en los perfiles de la base de datos de LUCDEME, con un total 495 muestras tras descartar aquellas donde faltaban datos de granulometría, carbonatos o CO, de las cuales 334 se encuentran dentro de la región de Murcia.

Para validar las funciones de edafotransferencia que estiman el contenido volumétrico de humedad, se ha tenido en cuenta la ecuación 4.2, que permite calcular $\psi_{0.33}$ multiplicando $\theta_{0.33}$ por la densidad aparente.

$$\psi_{0.33} = \theta_{0.33} \cdot \frac{\rho_b}{\rho_w} \quad (4.2)$$

donde ρ_b es la densidad aparente del suelo ($\text{g} \cdot \text{cm}^{-3}$) y ρ_w es la densidad del agua ($1 \text{ g} \cdot \text{cm}^{-3}$)

4.2.1 Capacidad de campo

El contenido de humedad a capacidad de campo en las muestras de la sierra de las Moreras varía entre el 9.31 y el 42.35 % en peso, con un valor medio de 22.21% como se observa en la tabla 4.4. Atendiendo al valor de la desviación estándar y el coeficiente de variación, se trata de una variable poco dispersa y bastante homogénea.

Mín.	Mediana	Media	Máx.	DS	CV(%)
9.31	21.46	22.21	42.35	5.65	25.55

TABLA 4.4: Estadísticos descriptivos de contenido de humedad a capacidad de cambio (% en peso) en los suelos de la sierra de las Moreras. **Mín.**= Mínimo; **Máx.**= Máximo; **SD**= Desviación estándar; **CV (%)**. = Coeficiente de variación.

Los datos de contenido de humedad a 0.33 atm no siguen una distribución normal, atendiendo al resultado del test de Shapiro-Wilk ($W = 0.9761$, $p\text{-valor} = 0.01067$). Sin embargo, la figura 4.6 muestra un histograma aproximadamente simétrico y un gráfico de cuantiles sin desviaciones sustanciales de la recta.

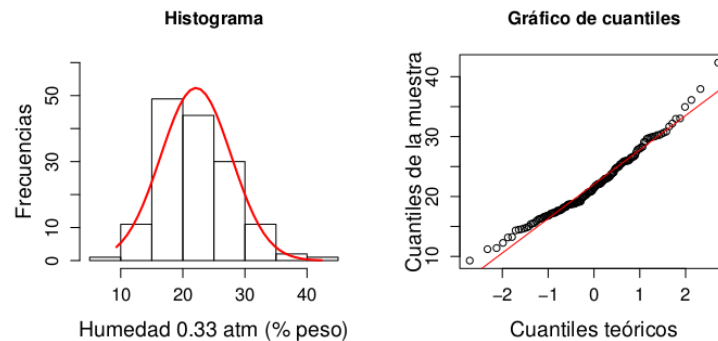
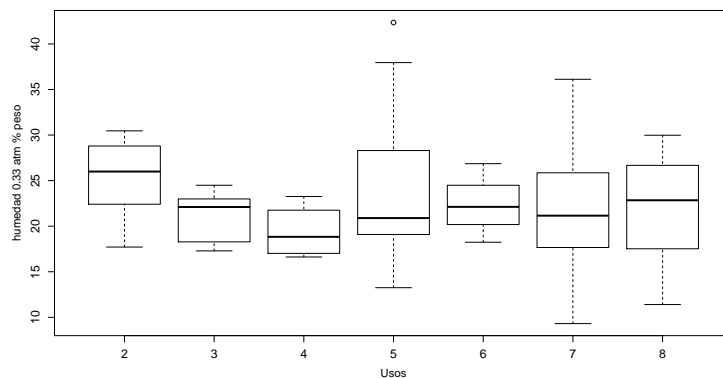
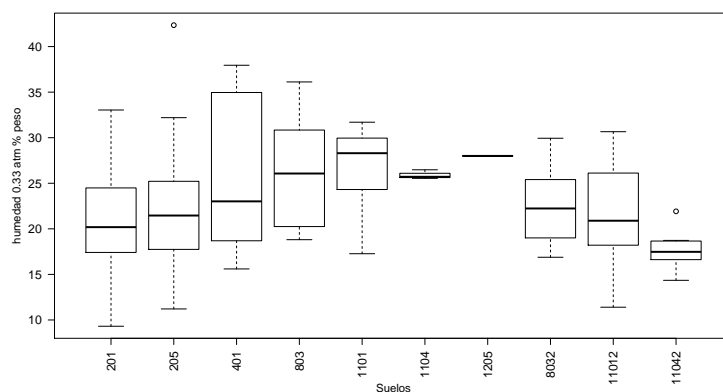


FIGURA 4.6: Histograma y gráfico de cuantiles de los datos de contenido de humedad a capacidad de campo (% en peso) en la sierra de las Moreras.

Además, el contenido de humedad está significativamente correlacionado con las fracciones granulométricas, siendo la correlación más fuerte y de signo negativo con el contenido de arena ($r = -0.62^{***}$), y de menor magnitud y de signo positivo con el contenido de limo ($r=0.52$, $p\text{-valor}=1.393 \cdot 10^{-11}$) y arcilla ($r=0.48$, $p\text{-valor}=2.68 \cdot 10^{-10}$). También se han encontrado correlaciones significativas con la Densidad aparente ($r=-0.5$; $p\text{-valor}= 0.0009385$, en las muestras en las que se disponía de ese dato) y el contenido de carbono orgánico ($r=0.31$; $p\text{-valor} = 0.0001048$).



(A) En función del uso



(B) En función del tipo de suelo

FIGURA 4.7: Distribución de los datos de contenido de humedad a capacidad de campo (% en peso) en función del uso y tipo de suelo.

El contenido de humedad a 0.33 atmósferas no muestra diferencias significativas entre los usos de suelo encontrados en la zona (test de Kruskal-Wallis, $\chi^2 = 7.6413$, p-valor = 0.2656), aunque la figura 4.7A indica que el valor de la mediana es mayor en los suelos de regadío que en el resto de usos, incluidos los naturales, mientras que los suelos con vegetación de matorral y los cultivos en mosaico muestran los rangos más amplios. Este valor más alto en los cultivos de regadío puede atribuirse a la mayor proporción de arcillas en estos suelos, que retienen el agua en el espacio de poros, ya que estos suelos suelen coincidir con zonas topográficamente bajas cercanas a cauces de agua, siendo en muchas ocasiones Fluvisoles.

En cambio, sí que se encuentran diferencias significativas entre los distintos tipos de suelo (test de Kruskal-Wallis, $\chi^2 = 22.0426$, p-valor = 0.008745). En este caso, la figura 4.7B muestra que los Solonchaks Háplicos (1205) y los Regosoles Lépticos (1104) son los que mayor cantidad de agua retienen a este potencial, aunque por el escaso número de muestras estos suelos son escasamente representativos. También los Regosoles Háplicos (1101) muestran altos contenidos de humedad

a 0.33 atm. Todos ellos tienen en común los altos contenidos de partículas finas (del tamaño de la arcilla). Les siguen los Leptosoles Líticos (803), con un valor de mediana menor pero un rango más amplio, y los Leptosoles Líticos asociados a Calcisoles. Por el contrario, los Regosoles Lépticos y Hápticos asociados a Calcisoles (11042 y 11012 respectivamente), y Calcisoles Hápticos y Pétricos (201 y 205, respectivamente) son las unidades con menor contenido de humedad a 0.33 atmósferas, lo cual podría deberse a la mayor proporción de arenas y menor contenido de CO en estos suelos. Finalmente los Fluvisoles Hápticos (401) muestran un rango amplio de valores de esta variable, que se solapa con los otros tipos de suelo, aunque el valor de la mediana es intermedio.

La revisión bibliográfica de funciones de edafotransferencia para el cálculo de esta propiedad edáfica dio como resultado 9 PTF para el contenido gravimétrico de humedad ($\theta_{0.33}$), y 4 para el contenido volumétrico ($\psi_{0.33}$) en humedad. Los algoritmos de las PTF se muestran en la tabla 4.5.

Contenido gravimétrico de humedad			
PTF	Autor	Algoritmo	Región
1	Pidgeon (1972)	$\theta_{0.33} = (0.0361 + 0.0016 \cdot L_p + (0.003 \cdot Cl_p + 0.03 \cdot CO_p) / 0.95) \cdot 100$	Suelos ferralíticos, Uganda
2	Barahona y Santos (1981)	$\theta_{0.33} = 45.996 + 0.1595 \cdot Cl_p - 0.1478 \cdot A_p + 0.2528 \cdot CO - 13.9615 \cdot D_a$	Castilla-La Mancha, España
3	Arruda <i>et al.</i> (1987)	$\theta_{0.33} = 3.07439 + 0.629329 \cdot (L_p + Cl_p) - 0.00343813 \cdot (L_p + Cl_p)^2$	São Paulo, Brasil
4	Dijkerman (1988)	$\theta_{0.33} = 36.97 - 0.35 \cdot A_p$	Sierra Leona
5	Bell (1993)	$\theta_{0.33} = 23.9 + 0.402 \cdot MO_p - 0.100 \cdot A_p$	México
6	Masutti (1997)	$\theta_{0.33} = -1.569 + 0.429 \cdot (L_p + Cl_p)$	Pernambuco, Brasil
7	Oliveira <i>et al.</i> (2002)	$\theta_{0.33} = (0.00333 \cdot L_p + 0.00387 \cdot Cl_p) \cdot 100$	Pernambuco, Brasil
8	Urach (2007)	$\theta_{0.33} = (0.643 + 0.00238 \cdot A_p - 0.26767 \cdot D_a) \cdot 100$	Rio Grande do Sul, Brasil
Contenido volumétrico de humedad			
PTF	Autor	Algoritmo	Región
1v	Gupta y Larson (1979)	$\psi_{0.33} = (0.003075 \cdot A_p + (0.005886 \cdot L_p + 0.008039 \cdot Cl_p + 0.002208 \cdot MO_p - 0.1434 \cdot D_a) \cdot 100$	EE.UU.
2v	Lal (1979)	$\psi_{0.33} = (0.334 - 0.003 \cdot A_p) \cdot 100$	Nigeria
3v	Rawls <i>et al.</i> (1982)	$\psi_{0.33} = 0.2576 \cdot -0.0020 \cdot A_p + 0.0036 \cdot Cl_p + 0.0299 \cdot MO_p$	EE.UU.
4v	Abbasi <i>et al.</i> (2011)	$\psi_{0.33} = 0.01 \cdot (15.60 - 0.323 \cdot A_p + 16.9 \cdot D_a)$	Irán

TABLA 4.5: Funciones de edafotransferencia para contenido de humedad a -0.33 atm. $\theta_{0.33}$ = Contenido de humedad gravimétrico a -0.33 atm (% peso); CO_p = % Carbono orgánico; CO = Carbono orgánico ($g \cdot Kg^{-1}$); MO_p = % Materia orgánica; MO = Materia orgánica ($g \cdot Kg^{-1}$); Cl_p = % arcilla; L_p = % limo; A_p = % arena; D_a = Densidad aparente ($g \cdot cm^{-3}$)

De las fracciones granulométricas, la más utilizada en las ecuaciones es la arena, seguida por la arcilla, mientras que el limo solamente aparece en cinco de las funciones revisadas, y en dos de ellas lo hace junto con el porcentaje de arcilla, reflejando el porcentaje de partículas finas. Esto concuerda con las correlaciones obtenidas entre el contenido de agua a capacidad de campo y las diferentes fracciones, que era más notable y de signo negativo en el caso de la arena ($r = -0.61$, $p\text{-valor} < 2.2 \cdot 10^{-16}$), mientras que con la arcilla y el limo las correlaciones eran positivas y de menor magnitud ($r = 0.48$; $p\text{-valor} = 2.68 \cdot 10^{-10}$ y $r = 0.52$; $p\text{-valor} = 1.39 \cdot 10^{-11}$, respectivamente).

De las funciones de edafotransferencia utilizadas, las obtenidas por Pidgeon (1972); Barahona y Santos (1981); Bell (1993); Gupta y Larson (1979) y Rawls *et al.* (1982) incluyen como variables predictoras la cantidad de materia orgánica, expresada como MO o CO, mientras que en el resto, solamente se utilizan variables relacionadas con la textura. El efecto de la materia orgánica en la capacidad de retención hídrica es contradictorio, ya que algunos autores han descrito una mejora de la predicción con PTFs al incluir la materia orgánica, mientras que otros no han encontrado

efecto o este depende de la textura de las muestras utilizadas (Rawls *et al.* (2003)).

En cuanto a la densidad aparente, solo aparece en cuatro de las funciones utilizadas (Barahona y Santos, 1981; Urach, 2007; Gupta y Larson, 1979 y Abbasi *et al.*, 2011). Al igual que el contenido en materia orgánica, la inclusión de la densidad aparente en las funciones de edafotransferencia mejora el resultado predictivo de estas (Rawls *et al.*, 2003), ya que refleja la estructura del suelo, su estado de agregación y por tanto la porosidad.

Contenido gravimétrico de humedad				
PTF	r	p-value	EMP	RECM
1	0.66	$< 2.2 \cdot 10^{-16}$	-4.83	6.41
2	0.66	$< 2.2 \cdot 10^{-16}$	1.93	4.81
3	0.66	$< 2.2 \cdot 10^{-16}$	3.85	5.75
4	0.62	$< 2.2 \cdot 10^{-16}$	-1.78	5.02
5	0.62	$< 2.2 \cdot 10^{-16}$	-2.90	5.68
6	0.63	$< 2.2 \cdot 10^{-16}$	-1.07	5.26
7	0.63	$< 2.2 \cdot 10^{-16}$	-3.74	5.97
8	0.67	$< 2.2 \cdot 10^{-16}$	-7.15	8.51
Contenido volumétrico de humedad				
PTF	r	p-value	EMP	RMSE
1v	0.67	$< 2.2 \cdot 10^{-16}$	-1.32	4.94
2v	0.68	$< 2.2 \cdot 10^{-16}$	-8.42	9.39
3v	0.64	$< 2.2 \cdot 10^{-16}$	-2.79	5.80
4v	0.63	$< 2.2 \cdot 10^{-16}$	-4.93	6.63

TABLA 4.6: Resultados de las funciones de edafotransferencia para contenido de humedad a Capacidad de Campo (% en peso). **r**= correlación (Pearson); **EMP** = Error Medio de la Predicción; **RECM** = Raíz del Error Cuadrático Medio.

La tabla 4.6 muestra que todas las funciones de edafotransferencia utilizadas arrojan correlaciones significativas con los valores observados de contenido de humedad a capacidad de campo, siendo los valores de esta correlación similares para todas ellas. Esta similitud puede observarse también en la figura 4.8 Sin embargo, la mayoría infraestiman este valor, tal como refleja el valor negativo de EMP, excepto la elaborada por Barahona y Santos (1981) y la obtenida por Arruda *et al.* (1987).

Por otro lado, la función que arroja la mayor correlación (Lal, 1979) resulta en los mayores valores de error, tanto en el EMP como en RECM. Lo mismo ocurre con la función de Urach (2007), que si bien consigue una correlación alta, también produce una gran infraestimación del contenido de humedad y un elevado valor de RECM. Por el contrario, la función de Bell y Van Keulen (1995) es la que menor correlación ofrece, junto con la de Dijkerman (1988). Las funciones de Masutti (1997) y Oliveira *et al.* (2002) resultan en una correlación de 0.63, diferenciándose entre sí por el error medio de la predicción. También la PTF de Abbasi *et al.* (2011) obtiene una correlación de 0.63, aunque en este caso los valores de error son considerablemente mayores. Esto confirmaría la importancia de aplicar las funciones de edafotransferencia a suelos con características similares a aquellos en los que fueron calibradas (Schaap y Leij, 1998a; Nemes

et al., 2002; Givi *et al.*, 2004), ya que excepto la función de Abbasi *et al.* (2011), elaborada para suelos de Irán, el resto fueron calibradas en suelos tropicales (Sierra Leona, Nigeria, México y Brasil), donde las condiciones son muy diferentes a las de la cuenca mediterránea.

Destaca también la función de Gupta y Larson (1979), que obtiene valores de error similares a los obtenidos por la función de Barahona y Santos (1981), y una correlación levemente mayor. En el caso de la última, fue calibrada en suelos similares a los de este trabajo, situados en Castilla-La Mancha, mientras que la de Gupta y Larson (1979) fue calibrada en un amplio rango de suelos de EE.UU, lo que explicaría su mayor aplicabilidad.

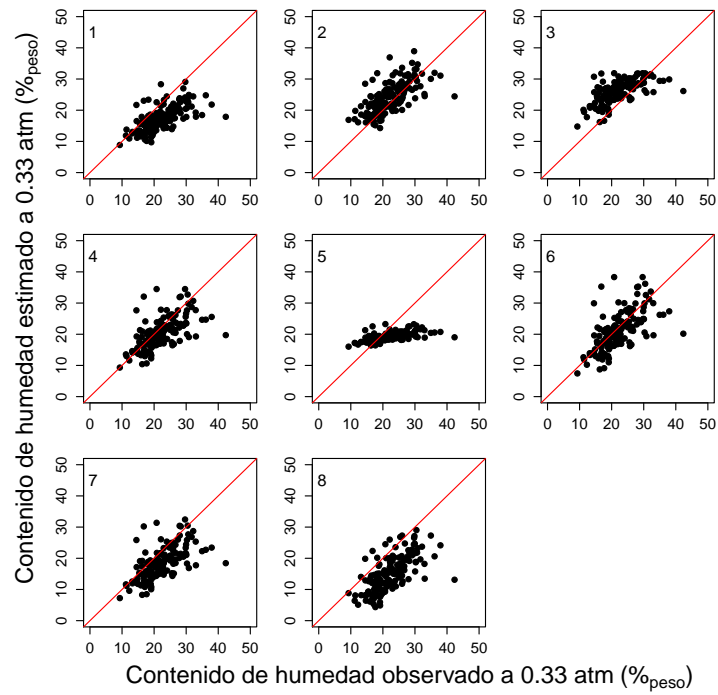
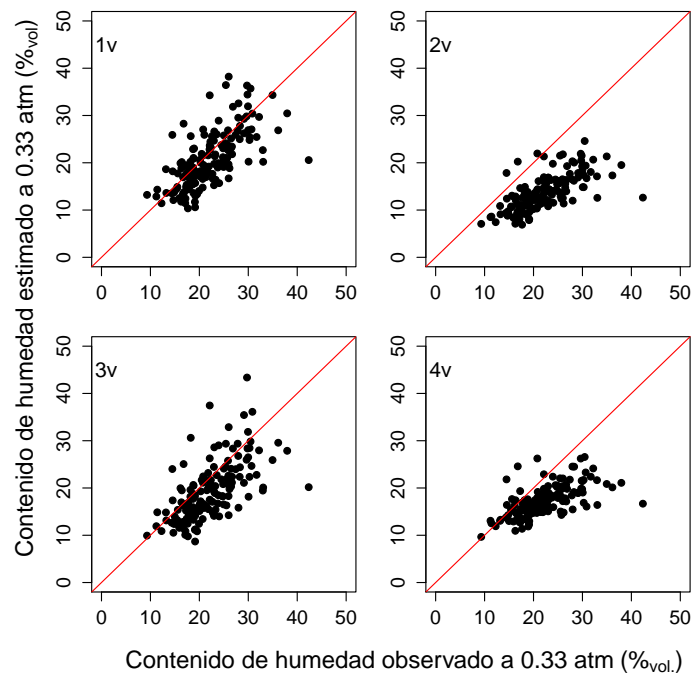
(A) Contenido gravimétrico de humedad ($\theta_{0.33}$; % en peso)(B) Contenido volumétrico de humedad ($\psi_{0.33}$; % en volumen)

FIGURA 4.8: Diagrama de dispersión del contenido de humedad a capacidad de cambio observado frente los valores obtenidos con las funciones de edafotransferencia recopiladas.

Finalmente, se ha seleccionado la función de [Barahona y Santos \(1981\)](#) para elaborar el mapa de contenido de humedad a capacidad de cambio.

Sin embargo, la función de Barahona no incluye el efecto que podría tener el carbonato cálcico en la retención de humedad, a pesar de haber sido calibrada en suelos calizos, ampliamente distribuidos en zonas áridas y semiáridas como la península ibérica y la región de Murcia.

Existen estudios que consideran los efectos de la precipitación de carbonatos en la reducción del espacio de poros, pero estos sugieren que ni la retención de humedad ni la conductividad hidráulica se ven afectados de forma significativa ([Frenkel et al., 1978](#)). Sin embargo, el efecto podría estar relacionado con la energía de adsorción y retención de las partículas carbonatadas del tamaño de las arcillas, por lo que las propiedades hídricas de los suelos calcáreos serían diferentes a las de los suelos no calcáreos.

[Khodaverdilloo et al. \(2011\)](#) comprobaron las posibles mejoras en la predicción que conllevaría la inclusión de los carbonatos en las PTF, comparando funciones en las que se incluía esta variable predictora frente funciones en las que no se incluía; sin embargo, sus PTF introducen también como variable predictora la media geométrica de las partículas, dato del que no se dispone en los datos de este trabajo, por lo que no han podido ser validadas.

En su lugar, se ha obtenido una función de regresión que utiliza las variables vistas anteriormente (materia orgánica, fracciones de arena, limo y arcilla, densidad aparente) y añadiendo el contenido de carbonatos, con un procedimiento de validación cruzada dejando una muestra fuera, y una regresión por pasos para eliminar las variables que no resultaban significativas en el modelo.

El modelo de regresión (ecuación 4.3, $R^2=0.4839$) elimina la arena de las variables significativas, a pesar de ser la fracción que mayor correlación mostraba con el contenido de humedad a este potencial; sin embargo, incluye el contenido en carbonatos, a pesar de tener menor correlación ($r=0.21$, p -valor = 0.01). El ajuste obtenido con esta función ($r=0.71$; $EMP=0.005$; $RMSE=3.99$) es mejor que el conseguido con la PTF de [Barahona y Santos \(1981\)](#). Las variables más significativas son el contenido de arcilla y limo, al contrario que la D_a y el CO , que tienen menor significación. La baja significación del contenido de CO puede ser atribuida al escaso contenido de este constituyente en los suelos estudiados ([Lal, 1979](#); [Danalatos et al., 1994](#); [Rawls et al., 2003](#)).

$$\theta_{0.33} = 22.44024 + 0.107100 \cdot CO + 0.302889 \cdot Cl_p + 0.208772 \cdot L_p - 0.005305 \cdot CO_3Ca - 8.796821 \cdot D_a \quad (4.3)$$

$$\theta_{0.33} = 22.23898 + 0.11003 \cdot CO + 0.25895 \cdot Cl_p + 0.19095 \cdot L_p - 8.65944 \cdot D_a \quad (4.4)$$

donde $\theta_{0.33}$ el contenido gravimétrico de humedad a -0.33 atm (% en peso); CO = Carbono orgánico ($g \cdot Kg^{-1}$); Cl_p = % arcilla; L_p = % limo; A_p = % arena; D_a = Densidad aparente ($g \cdot cm^{-3}$); y CO_3Ca = Carbonato cálcico ($g \cdot Kg^{-1}$)

Por otro lado, la eliminación del contenido en carbonato de la función (ecuación 4.4) no resulta en una diferencia significativa en el ajuste del modelo ($R^2=0.472$) ni en el ajuste con los datos

observados ($r=0.69$, $EMP=-0.0003$, $RECM=4.05$; figura 4.9) por lo que finalmente se ha seleccionado esta ecuación por su mayor simplicidad. Este resultado coincide con el obtenido por [Khodaverdiloo et al. \(2011\)](#), que tampoco encontraron diferencias significativas al comparar los pares de PTF incluyendo o excluyendo el contenido de CO_3Ca . Esto podría deberse a la presencia de otros predictores (sobre todo la densidad aparente o la textura), que explicarían parte de los posibles efectos del carbonato reduciendo al mismo tiempo sus efectos en el resultado de la función.

La comparación con el test de la t de Welch indica que existen diferencias significativas entre las medias del resultado de la PTF de [Barahona y Santos \(1981\)](#) y el modelo de regresión elaborado ($t = 3.3959$, $p\text{-valor} = 0.0007829$), así como entre las medias de los valores obtenidos con la función de [Barahona y Santos \(1981\)](#) y los datos observados ($t = 2.8659$, $p\text{-valor} = 0.004461$), pero no entre la predicción del modelo de regresión lineal y los datos observados en Mazarrón ($t = 7 \cdot 10^{-4}$, $p\text{-valor} = 0.9995$), cuyos valores medios son muy similares. Al realizar la misma comparación utilizando los datos de los perfiles, en cambio, se obtienen diferencias significativas entre los valores de la función de [Barahona y Santos \(1981\)](#) y el modelo de regresión ($t = -2.9768$, $p\text{-valor} = 0.002989$), y el resultado del modelo lineal y los valores observados ($t = 4.01$, $p\text{-valor} = 6.643 \cdot 10^{-5}$), al infraestimar el modelo el contenido de humedad a este potencial, pero no entre la función de [Barahona y Santos \(1981\)](#) y los datos observados ($t = 1.5086$, $p\text{-valor} = 0.1318$). Se ha seleccionado este test no paramétrico por no poder asegurar las condiciones de normalidad y homogeneidad de varianzas necesarias para la prueba de la t de Student.

Finalmente, se ha obtenido un mapa utilizando este modelo de regresión y las variables modelizadas (página A.16), para compararlo con el obtenido anteriormente con la función de [Barahona y Santos \(1981\)](#) (página A.15), cuyos parámetros de validación se pueden observar en la tabla 4.7.

La disminución en la correlación, y el aumento del error, sobre todo en la regresión lineal, se debe a las diferencias entre los valores de la base de datos, y los valores de los mapas de las variables predictoras, obtenidos con los modelos a nivel regional. Este hecho viene corroborado por la mayor correlación obtenida entre los datos de los mapas y los observados en los perfiles de LUCDEME. De este modo, los errores existentes en los mapas se propagarían al utilizar los valores en el cálculo de las variables hídricas.

Mapa	Contenido gravimétrico de humedad			
	r	p-value	EMP	RECM
PTF (Barahona y Santos (1981))	0.24	0.007	0.36	6.41
PTF _{perf}	0.32	$1.77 \cdot 10^{-9}$	3.18	9.48
LM	0.22	0.01	-1.02	5.60
LM _{perf}	0.32	$1.623 \cdot 10^{-9}$	-0.03	7.94

TABLA 4.7: Comparación de los mapas de contenido de humedad a Capacidad de Campo (% en peso) con los datos de Mazarrón y los datos de perfiles de LUCDEME (_{perf}): r = correlación (Pearson); EMP =Error Medio de la Predicción; $RECM$ = Raíz del Error Cuadrático Medio.

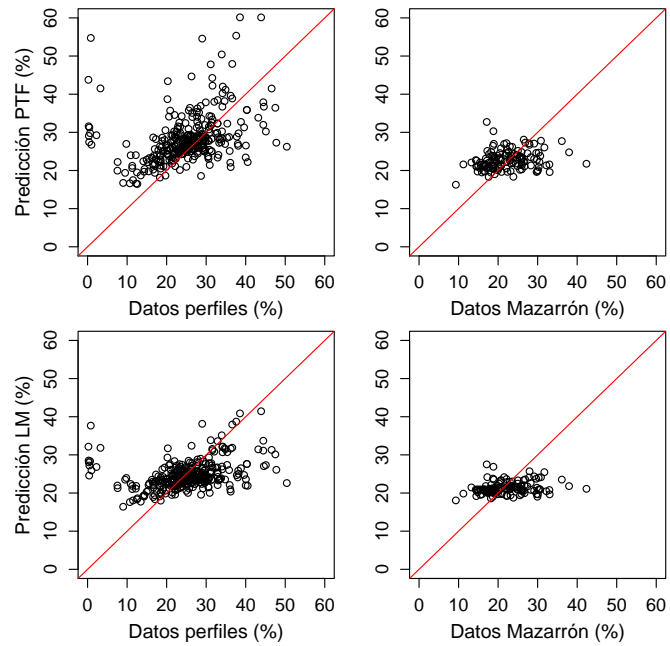


FIGURA 4.9: Diagrama de dispersión del contenido de humedad a capacidad de cambio observado frente los valores obtenidos con en los mapas con la función de edafotransferencia (PTF) y el modelo de regresión (LM)

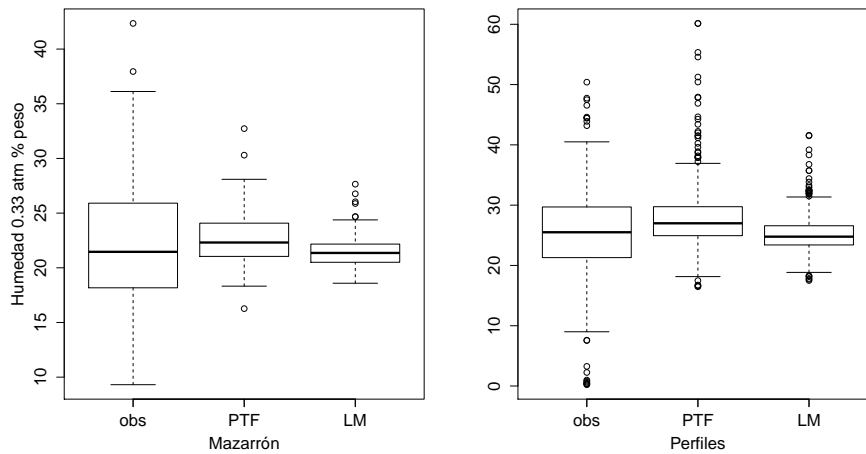


FIGURA 4.10: Diagrama de cajas del contenido de humedad a capacidad de cambio. **obs**: valores observados en las muestras de Mazarrón (izquierda) y los perfiles de LUCDEME (derecha); **PTF**: valores obtenidos en el mapa con la función de Barahona; **LM**: valores obtenidos en el mapa usando el modelo de regresión.

Sin embargo, el test de la t de Welch no muestra diferencias significativas entre el mapa de humedad a capacidad de cambio calculado con el modelo y los datos observados en Mazarrón ($t = 1.0438$, $df = 175.377$, $p\text{-value} = 0.298$), ni tampoco entre el mapa con la función de Barahona y Santos (1981) y los datos observados en esa misma zona ($t = -1.1908$, $p\text{-valor} = 0.2351$). Por el contrario, el mapa obtenido usando la PTF sobreestima el contenido de humedad respecto el observado en los perfiles, existiendo diferencias significativas entre las medias (test de la t de Welch; $t = -5.2596$, $p\text{-valor} = 1.98 \cdot 10^{-7}$), mientras que esas diferencias no existen entre los datos observados en los perfiles y los valores del mapa con el modelo (test de la t de Welch; $t = -0.0193$, $p\text{-valor} = 0.9846$). Por otro lado, los mayores valores de RECM pueden atribuirse al rango más amplio de los valores observados en los perfiles. Tanto en las muestras de Mazarrón como en las muestras de los perfiles, se observa un solapamiento de los rangos, mucho más estrechos en las estimaciones, mientras que los valores de las medianas son muy similares entre sí (figura 4.10).

4.2.2 Punto de marchitamiento

La tabla 4.8 muestra los estadísticos descriptivos del contenido de humedad a capacidad de cambio, en la que destaca el valor más alto del coeficiente de variación, y por tanto una mayor heterogeneidad en los datos.

Mín.	Mediana	Media	Máx.	DS	CV(%)
2.67	8.97	9.18	26.92	3.41	37.22

TABLA 4.8: Estadísticos descriptivos de contenido de humedad en el punto de marchitamiento (% en peso) en los suelos de la sierra de las Moreras. **Mín.**= Mínimo; **Máx.**= Máximo; **SD**= Desviación estándar; **CV (%)**. = Coeficiente de variación.

Además, los datos de contenido de humedad en el punto de marchitamiento no se ajustan a una distribución normal (test de Shapiro-Wilk, $W = 0.9459$, $p\text{-valor} = 1.623 \cdot 10^{-5}$), a pesar de que el histograma es razonablemente simétrico, si bien el gráfico de cuantiles se aleja de la recta en uno de los extremos del rango (figura 4.11).

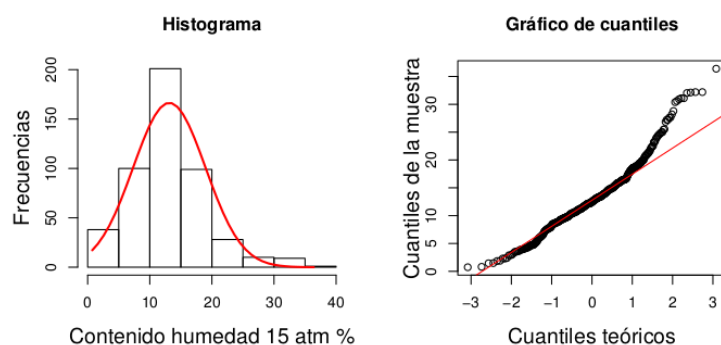


FIGURA 4.11: Histograma y gráfico de cuantiles del contenido de humedad a -15 atmósferas (% peso) en la sierra de las Moreras.

A diferencia del contenido de humedad a capacidad de cambio, el contenido de humedad en el punto de marchitamiento está más fuertemente relacionado con el contenido en arcilla ($r=0.52$, $p\text{-valor} = 1.07 \cdot 10^{-11}$). Esto se debe a que en los suelos de textura más fina, los poros son de menor tamaño y debido a la capilaridad, el agua queda retenida en los mismos con más fuerza. Sin embargo, la correlación con el contenido en limo es más débil ($r=0.18$, $p\text{-valor}= 0.03$), así como la encontrada con el contenido en arena ($r=-0.39$, $p\text{-valor}=5.722 \cdot 10^{-7}$). Por otro lado, el contenido en carbono orgánico muestra un valor de correlación similar al encontrado con $\theta_{0.33}$ ($r=0.29$, $p\text{-valor}=0.0003$). Destaca sin embargo el valor negativo en la correlación con la densidad aparente en los datos en los que originalmente se disponía de este valor ($r=-0.58$, $p\text{-valor}=6.138 \cdot 10^{-5}$), ya que sería de esperar que los suelos con mayor cantidad de arcilla, y por tanto más densos, fueran los que mayor cantidad de agua almacenaran a este potencial.

El contenido de humedad a -15 atm es diferente en función de los usos y tipos de suelo. En relación a los usos (figura 4.12), se observan mayores contenidos de humedad en los suelos de bosque (6) y los cultivos de regadío (2), mientras que los suelos con cultivos arbóreos son los que menor cantidad de humedad retienen, y el resto de usos muestran contenidos intermedios. No obstante, las diferencias halladas no revisten significación estadística (test de Kruskal-Wallis, $\chi^2 = 6.2661$, $p\text{-valor} = 0.394$). Tampoco son significativas las diferencias encontradas entre los distintos tipos de suelo ($\chi^2 = 14.8306$, $p\text{-valor} = 0.09569$), la figura 4.13 muestra que los Leptosoles Líticos (803) son los que mayor cantidad de agua almacenan a este potencial, junto con los Regosoles Háplicos (1101); por el contrario, los Regosoles Lépticos asociados a Calcisoles (11042) son los que menor cantidad de agua retienen a -15 atmósferas, junto con los Solonchaks Háplicos que no son representativos debido a la escasez de muestras. El resto de suelos almacenan poca cantidad de agua a este potencial, destacando los Fluvisoles Háplicos en los cual se esperaría mayor contenido de humedad debido a la importancia de las fracciones finas (limos y arcillas).

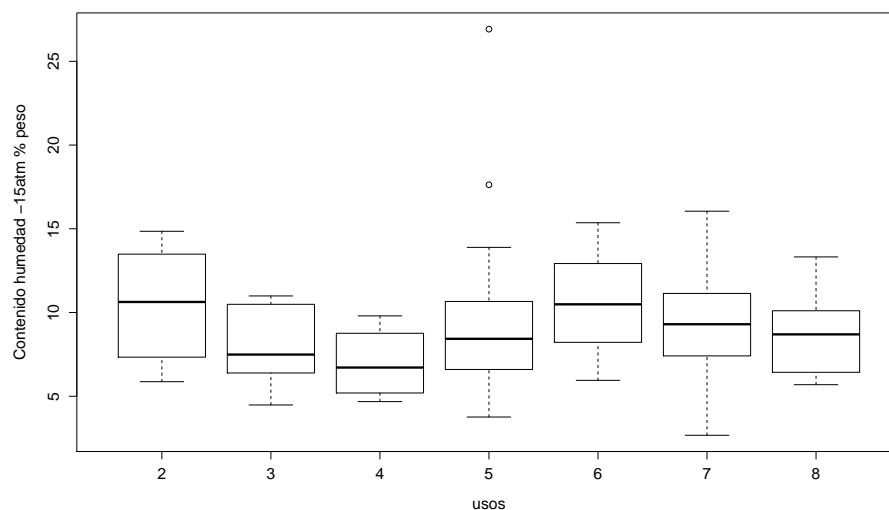


FIGURA 4.12: Distribución del contenido de humedad a -15 atm (% peso) en función de los usos del suelo en la sierra de las Moreras

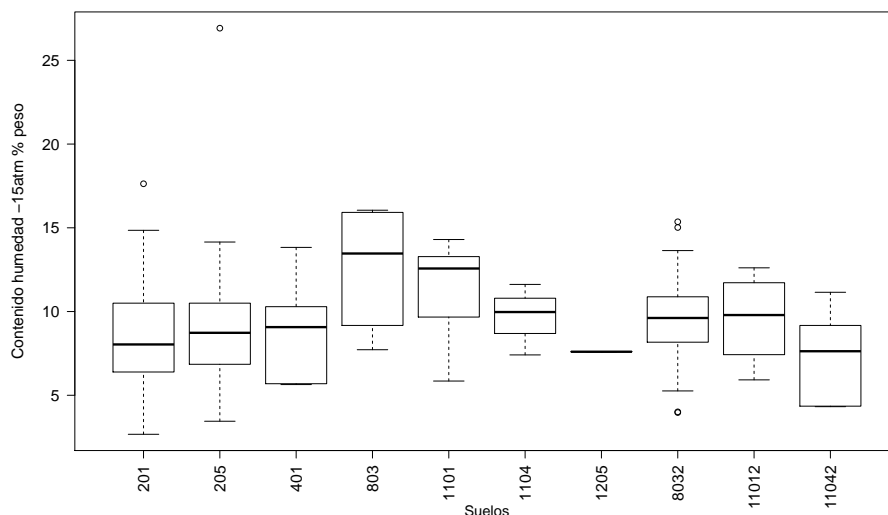


FIGURA 4.13: Distribución del contenido de humedad a -15 atm (% peso) en función de los tipos de suelo en la sierra de las Moreras.

Como en el caso de la Capacidad de Cambio, se han recopilado 8 funciones de edafotransferencia para estimar el contenido gravimétrico de humedad en el punto de marchitamiento permanente, y 4 que determinan el contenido volumétrico de humedad, cuyos algoritmos se muestran en la tabla 4.9.

Contenido gravimétrico de humedad			
PTF	Autor	Algoritmo	Región
1	Pidgeon (1972)	$\theta_{15} = (-0.0419 + 0.0019 \cdot L_p + 0.0039 \cdot Cl_p + 0.009 \cdot CO) \cdot 100$	Uganda
2	Barahona y Santos (1981)	$\theta_{15} = 8.4221 + 0.2781 \cdot Cl_p - 0.0833 \cdot A + 1.1537 \cdot CO_p$	Castilla-La Mancha (España)
3	Arruda <i>et al.</i> (1987)	$\theta_{15} = 1,074 + 0.2712(L_p + Cl_p)$	São Paulo (Brasil)
4	Dijkerman (1988)	$\theta_{15} = 0.74 + 0.39 \cdot Cl_p$	Sierra Leona
5	Bell y Van Keulen (1995)	$\theta_{15} = (-0.992 + 0.351 \cdot Cl_p + 0.047 \cdot MO_p) \cdot 100$	México
6	Masutti (1997)	$\theta_{15} = -0.530 + 0.0928 \cdot Cl_p + 0.301 \cdot L_p$	Pernambuco (Brasil)
7	Oliveira <i>et al.</i> (2002)	$\theta_{15} = [-0.00038 \cdot A_p + 0.00153 \cdot L_p + 0.0034 \cdot Cl_p - 0.0309 \cdot D_a] \cdot 100$	Pernambuco (Brasil)
	Urach (2007)	$\theta_{15} = [0.462 - 0.0022 \cdot A_p - 0.00074 \cdot L_p - 0.1838 \cdot D_a] \cdot 100$	Rio Grande do Sul (Brasil)
Contenido volumétrico de humedad			
PTF	Autor	Algoritmo	Región
1v	Gupta y Larson (1979)	$\psi_{15} = -0.000059 \cdot A_p + (0.001142 \cdot L_p) + (0.005766 \cdot Cl_p) + (0.002228 \cdot OM_p) + (0.02671 \cdot MO_p)$	EE.UU.
2v	Lal (1979)	$\psi_{15} = [0.247 - 0.003 \cdot MO_p] \cdot 100$	Nigeria
3v	Rawls <i>et al.</i> (1982)	$\psi_{15} = 0.026 \cdot Cl + 0.0158 \cdot MO_p$	EE.UU.
4v	Abbasi <i>et al.</i> (2011)	$\psi = 0.01 \cdot (6.627 + 0.315 \cdot Cl)$	Irán

TABLA 4.9: Funciones de edafotransferencia para contenido de humedad a -15 atm. θ_{15} =Contenido de humedad gravimétrico a -15 atm (% peso); CO_p = % Carbono orgánico; CO = Carbono orgánico ($g \cdot Kg^{-1}$); MO_p = % Materia orgánica; MO = Materia orgánica ($g \cdot Kg^{-1}$) Cl_p =% arcilla; L_p =% limo; A_p = % arena; D_a = Densidad aparente ($g \cdot cm^{-3}$)

Del mismo modo que ocurría con el contenido de humedad a capacidad de campo, las PTF que estiman el contenido volumétrico de humedad no arrojan resultados peores que las que estiman el contenido gravimétrico, a pesar de que los valores de densidad aparente solamente estaban disponibles para una parte de las muestras.

Las correlaciones obtenidas entre los datos observados de contenido de humedad a -15 atmósferas y los valores predichos por las funciones de edafotransferencia son menores que en el caso del contenido de humedad a capacidad de cambio, y tienen un rango de variación más amplio, entre 0.26 con la función de [Masutti \(1997\)](#) y 0.57 con la función de [Abbasi et al. \(2011\)](#). Entre estos valores, destacan la función de [Barahona y Santos \(1981\)](#) y la de [Rawls et al. \(1982\)](#), ambas con una correlación de 0.56 (tabla 4.10). El ajuste de los valores predichos por las funciones y los valores observados en Mazarrón se muestra gráficamente en la figura 4.14.

Contenido gravimétrico de humedad				
PTF	r	p-value	EMP	RECM
1	0.47	$1.51 \cdot 10^{-9}$	8.59	11.02
2	0.56	$1.09 \cdot 10^{-13}$	0.63	3.11
3	0.38	$1.39 \cdot 10^{-6}$	6.19	7.42
4	0.52	$1.07 \cdot 10^{-11}$	-2.54	3.97
5	0.53	$5.44 \cdot 10^{-12}$	-4.78	5.63
6	0.26	0.0013	3.01	5.18
7	0.48	$3.06 \cdot 10^{-10}$	-4.48	5.84
8	0.53	$3.33 \cdot 10^{-12}$	-2.32	4.31
Contenido volumétrico de humedad				
PTF	r	p-value	EMP	RMSE
1v	0.59	$3.86 \cdot 10^{-14}$	-2.41	4.12
2v	0.46	$4.43 \cdot 10^{-9}$	-1.65	3.9
3v	0.59	$2.22 \cdot 10^{-15}$	0.11	3.19
4v	0.58	$5.107 \cdot 10^{-15}$	-1.04	2.95

TABLA 4.10: Resultados de las funciones de edafotransferencia para contenido de humedad a -15 atm. r = correlación (Pearson); EMP = Error Medio de la Predicción; RECM = Raíz del error cuadrático medio.

Esta disminución en la correlación se acompaña de un aumento en los valores de error, tanto el error medio de la predicción como el RECM. El EMP es de signo negativo en la mayoría de las funciones, lo que indica una infraestimación del contenido de humedad a este potencial. Su magnitud varía entre 2.54 y 8.59 en casi todas las funciones, excepto las de [Barahona y Santos \(1981\)](#) y [Rawls et al. \(1982\)](#) donde este valor es menor de 1; sin embargo, los signos son opuestos, ya que la función de [Barahona y Santos \(1981\)](#) tiende a sobreestimar el valor de θ_{15} mientras que la de [Rawls et al. \(1982\)](#) produce una infraestimación del ψ_{15} . El valor máximo de este error no se produce sin embargo con la PTF que resultaba en una menor correlación, sino con la función de [Pidgeon \(1972\)](#), acompañado por el mayor valor de RECM. Por el contrario, el menor valor de RECM se produce con la PTF de [Barahona y Santos \(1981\)](#), por lo que se ha seleccionado esta para elaborar el mapa de contenido de humedad en el punto de marchitamiento.

Como ya se ha comentado con anterioridad, el mejor resultado de esta función puede atribuirse a la semejanza existente entre los suelos para los que fue calibrada, y los suelos utilizados para la validación, ambos situados en ambientes semiáridos. Por otro lado, la función de [Abbasi et al. \(2011\)](#) fue calibrada para suelos salinos y salino-sódicos en una zona árida (Irán) a pesar de lo cual, el valor del error medio de la predicción es mayor, lo cual podría deberse a las diferentes

características de los suelos de esa región. Por su parte, el buen rendimiento de la función de [Rawls *et al.* \(1982\)](#) puede atribuirse a que fue calibrada en un amplio rango de suelos de los Estados Unidos.

El modelo lineal elaborado con los datos de Mazarrón, utilizando validación cruzada (LOOCV) y un procedimiento de regresión por pasos da como resultado la ecuación 4.5. Este modelo de regresión lineal, en el que solamente resultan significativos el contenido de carbono orgánico y el contenido de arcilla, explica menor proporción de varianza que el elaborado anteriormente para capacidad de campo ($R_{aj}^2 = 0.33$)

$$\theta_{15} = 4.14154 + 0.12560 \cdot \text{CO} + 0.25167 \cdot \text{Cl}_p \quad (4.5)$$

donde θ_{15} es el contenido gravimétrico de humedad a -15 atm (% en peso); CO= Carbono orgánico ($\text{g} \cdot \text{kg}^{-1}$), y Cl_p =% arcilla.

En este caso, el contenido de carbonato es eliminado de la ecuación en el proceso de regresión por pasos, por lo que se deduce que este constituyente no resulta significativo en la retención de humedad a este potencial, tal como también apuntaban [Khodaverdiloo *et al.* \(2011\)](#). Tampoco aparece la densidad aparente, lo cual hace pensar en una menor importancia de la estructura. El mayor coeficiente ligado a la fracción arcilla permite deducir que las fuerzas de capilaridad ligadas al tamaño de los poros y la mayor superficie específica de estas partículas ([Hillel, 1998](#)).

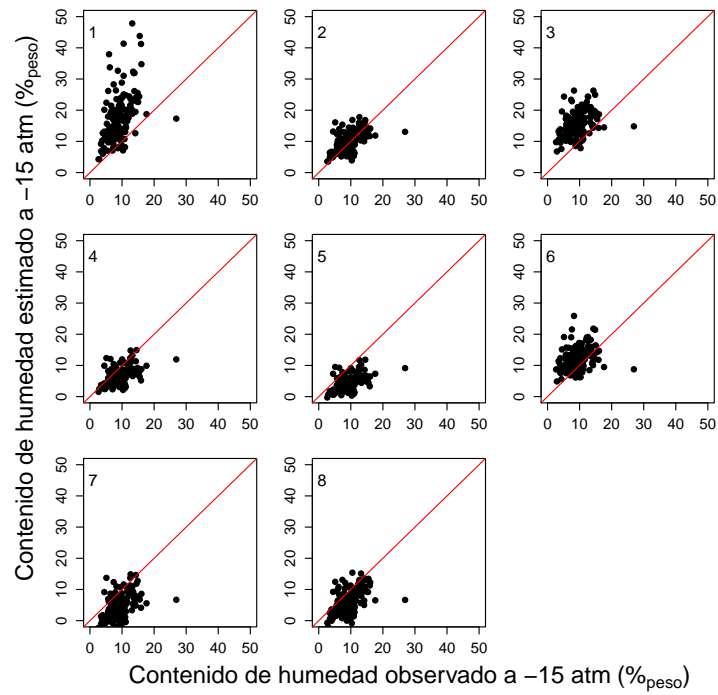
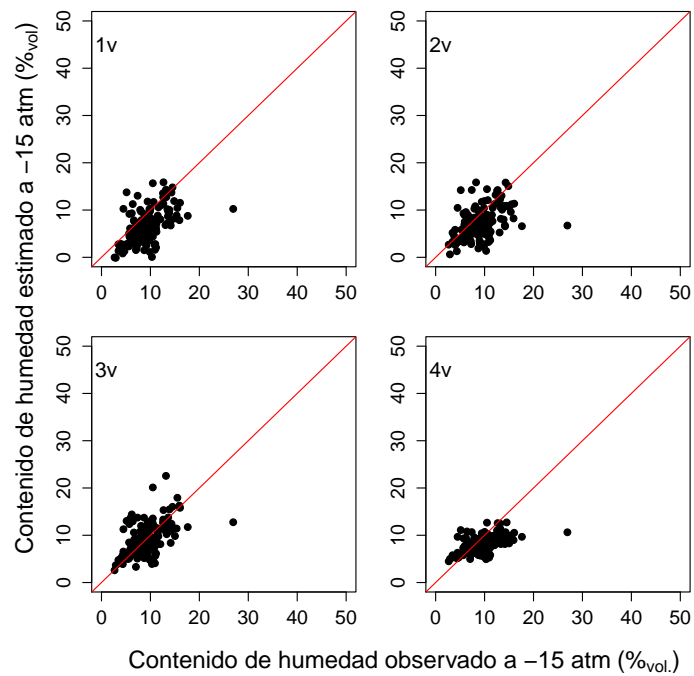
(A) Contenido gravimétrico de humedad (θ_{15} ; % en peso)(B) Contenido volumétrico de humedad (ψ_{15} ; % en volumen)

FIGURA 4.14: Diagramas de dispersión del contenido de humedad a -15 atmósferas en relación al obtenido con las funciones de edafotransferencia recopiladas.

El contenido de humedad predicho por el modelo resulta en un correlación mayor que la obtenida con la función de edafotransferencia ($cor=0.58$), y menores valores de error, sobre todo el error medio de la predicción ($EMP=0.011$), pero también la raíz del error cuadrático medio ($RECM=2.76$).

La comparación entre las tres predicciones con el test de Wilcoxon indica que no existen diferencias significativas entre los resultados de la PTF de Barahona y Santos (1981) y el modelo de regresión ($W = 12324$, $p\text{-valor} = 0.1001$), ni tampoco entre la predicción del modelo de regresión lineal y los datos observados en Mazarrón ($W = 10638$, $p\text{-valor} = 0.5345$). Entre los valores obtenidos con la función de Barahona y Santos (1981) y los datos de Mazarrón existe una diferencia mayor que las descritas, pero tampoco significativa al nivel seleccionado ($W = 9660$, $p\text{-valor} = 0.0528$). Es decir, la obtención del nuevo modelo no supone una mejora en la estimación del contenido de humedad a -15 atmósferas.

Mapa	Contenido gravimétrico de humedad			
	r	p-value	EMP	RECM
PTF (Barahona y Santos (1981))	0.18	0.04	0.48	3.60
PTF _{perf}	0.43	$2.2 \cdot 10^{-16}$	-1.43	5.34
LM	0.12	0.15	-0.15	3.53
LM _{perf}	0.57	$2.2 \cdot 10^{-16}$	-3.13	6.40

TABLA 4.11: Comparación de los mapas de contenido de humedad en el punto de marchitamiento (% en peso) con los datos de Mazarrón y los datos de perfiles de LUCDEME (_{perf}): r= correlación (Pearson); p-valor: * < 0.05; ** < 0.01; *** < 0.001; EMP=Error Medio de la Predicción; RECM=Raíz del error cuadrático medio.

No obstante, al calcular el contenido de humedad utilizando los mapas de las variables predictoras, el resultado utilizando este modelo es peor que con la función de Barahona y Santos (1981) en términos de correlación, pero no en términos de error (tabla 4.11, figura 4.15).

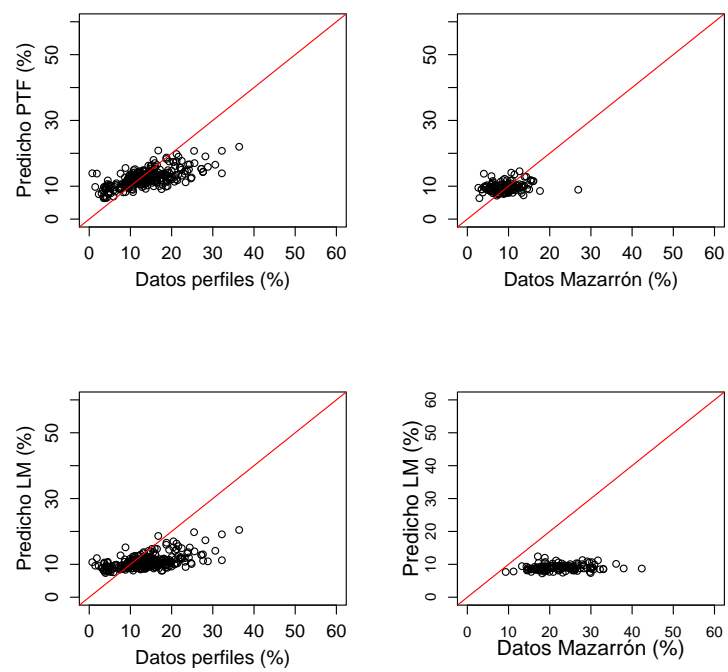


FIGURA 4.15: Diagrama de dispersión del contenido de humedad en el punto de marchitamiento observado frente los valores obtenidos con la función de edafotransferencia (PTF) y el modelo de regresión (LM)

Esta menor correlación se puede atribuir a la propagación de los errores procedentes de los modelos utilizados, y a la incertidumbre espacial de las muestras utilizadas para la calibración de estos modelos. Esta hipótesis vendría confirmada porque las correlaciones de ambos mapas con las muestras de los perfiles de LUCDEME (utilizadas en la calibración) son mayores. Los mayores valores de error, sobre todo en relación con los perfiles, también pueden atribuirse a las diferencias en los rangos de los valores observados y los valores estimados (figura 4.16).

En cuanto al mapa con la función de Barahona (página A.17), a pesar de la escasa correlación obtenida y la sobreestimación del contenido medio de humedad (indicado por el valor de EMP), no existen diferencias significativas entre las medias (test de la t de Welch, p-valor=0.12). Por el contrario, en los datos de perfiles donde la correlación era mayor, esta función produce una subestimación del contenido de humedad, con diferencias significativas entre las medias (p-valor = 0.0003). El mapa obtenido con el modelo GLM (página A.18), atendiendo al valor de EMP, produce una infraestimación del contenido de humedad en ambos conjuntos de datos, pero el valor medio es significativamente menor solo en los perfiles de LUCDEME (test de la t de Welch, p-valor = $2.385 \cdot 10^{-15}$), mientras que en las muestras de Mazarrón son muy similares (test de la t de Welch, p-valor = 0.58).

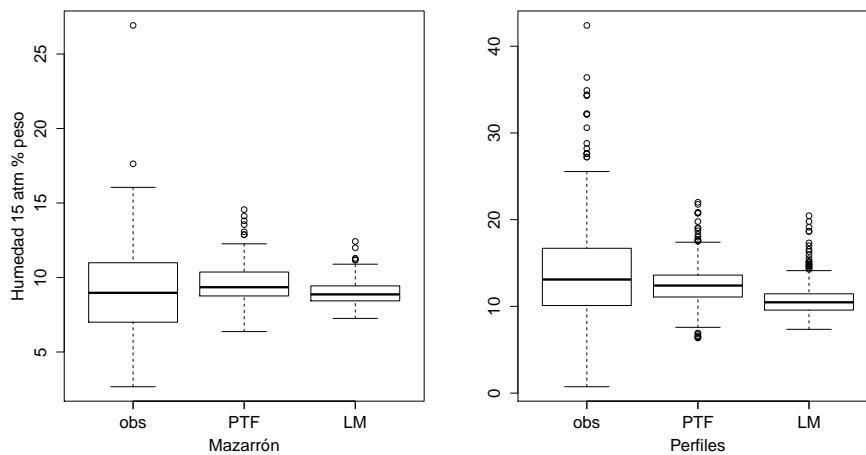


FIGURA 4.16: Diagrama de cajas del contenido de humedad a punto de marchitamiento (% peso). **obs**: valores observados en las muestras de Mazarrón (izquierda) y los perfiles de LUCDEME (derecha); **PTF**: valores obtenidos en el mapa con la función de Barahona; **LM**: valores obtenidos en el mapa usando el modelo de regresión.

Capítulo 5

Conclusiones

1. Los modelos a escala regional elaborados sirven para estimar los valores medios de las propiedades físico-químicas seleccionadas, a pesar de no ser capaces de reflejar toda la variabilidad espacial mostrada por las mismas.
2. El contenido de carbono orgánico y carbonato cálcico total o equivalente son las propiedades físico-químicas en las que se obtiene un mejor resultado de los modelos a escala regional, mientras que en el caso de las fracciones granulométricas, las predicciones obtenidas tienen una menor correlación con los valores medidos y producen mayores valores de error.
3. El método de Random Forest con validación cruzada en bloques es el que mejor resultado proporciona en la modelización de las propiedades físico-químicas a escala regional.
4. A escala regional, las variables predictoras de los modelos son mayoritariamente de tipo climático, junto con los tipos de suelos y vegetación, y algunas variables geomorfométricas que representan rasgos fisiográficos a gran escala. En los modelos a escala local, donde el clima y la vegetación son más homogéneos, las variables geomorfométricas aumentan su importancia.
5. En el caso del CO y $\text{CaCO}_3_{\text{total}}$, los residuos de los modelos mostraban una buena estructura espacial, que permitió la interpolación con Krigeado ordinario, mientras que en el caso de las fracciones granulométricas, la estructura espacial de los residuos es peor. Esto se puede atribuir a la menor dependencia de la textura respecto de las variables climáticas, usos y tipos de suelo, en comparación con el CO y el contenido en carbonatos, y una mayor dependencia de procesos relacionados con la geomorfología, que no han podido ser capturados por los modelos elaborados.

6. La adición de los residuos interpolados a la predicción del modelo, ya sea a través de krigeado ordinario, como a través de splines, aumenta la correlación de los mapas obtenidos y disminuye los valores de error. No obstante, este resultado debe tomarse con prudencia, ya que no se ha realizado una validación cruzada de estos métodos de interpolación, tarea que sería de interés llevar a cabo en el futuro.
7. La modelización de las variables físico-químicas a escala local no produce mejores resultados que los obtenidos a escala regional, expresados como un aumento de la correlación y una disminución del error de predicción. Este hecho puede atribuirse a la incertidumbre espacial de las coordenadas de las muestras a ambas escalas y a la menor variabilidad de las variables predictoras a escala local.
8. Las funciones de edafotransferencia que mejor resultado proporcionan son las obtenidas en los suelos más similares a las de la región de estudio o las elaboradas para ser aplicadas en un amplio rango de suelos.
9. De las funciones de edafotransferencia obtenidas en la revisión bibliográfica, la elaborada por [Barahona y Santos \(1981\)](#) es la que mejor resultado proporciona para las tres variables estudiadas (densidad aparente, contenido de humedad a 0.33 atm y contenido de humedad a 15 atm).
10. Las funciones de edafotransferencia obtenidas con los datos de la sierra de las Moreras no obtienen resultados mejores que los obtenidos con la función de [Barahona y Santos \(1981\)](#) al validarlas con datos externos. Este hecho puede atribuirse a la menor variabilidad en las propiedades edáficas que determinan la densidad aparente y la capacidad de almacenamiento de humedad, que dificulta la aplicación en datos distintos a los que sirvieron para su calibración.
11. El resultado de las funciones de edafotransferencia utilizando como datos de entrada los mapas de variables físicoquímicas modelizadas previamente es peor que al utilizar los datos analíticos. La causa puede atribuirse a la propagación de errores en los modelos. Sin embargo, se considera que los mapas obtenidos para densidad aparente y contenido de humedad realizan una estimación satisfactoria de estas propiedades del suelo a escala regional.
12. Tras analizar los resultados obtenidos en este trabajo, resultaría interesante en el futuro analizar la variabilidad espacial de las propiedades hídricas de los suelos con las propiedades geomorfométricas y climáticas, así como con el tipo y el uso de los suelos, a través de modelos análogos a los elaborados para las propiedades físico-químicas.

Lista de tablas

Pág.

2.1	Datos analíticos incluidos en la base de datos de LUCDEME	30
2.2	Reclasificación de usos CORINE Land Cover 1990	39
2.3	Reclasificación de unidades de suelos	45
2.4	Resumen de las variables predictoras utilizadas	48
3.1	Variables resultantes del proceso de selección de variables	58
3.2	Estadísticos descriptivos del contenido de CO ($\text{g} \cdot \text{kg}^{-1}$) en los suelos de la región	59
3.3	Correlaciones de $\log(\text{CO})$ con las variables predictoras cuantitativas	62
3.4	Modelos a nivel regional obtenidos para Carbono Orgánico ($\text{g} \cdot \text{kg}^{-1}$). Estadísticos descriptivos de la correlación (r) y Raíz del Error Cuadrático Medio (RECM_v) obtenidos tras el remuestreo en el conjunto de validación.	64
3.5	Estadísticos descriptivos del contenido de CaCO_3 total o equivalente en los suelos de la Región de Murcia.	73
3.6	Correlaciones entre el contenido en $\text{CaCO}_{3\text{total}}$ y las variables predictoras cuantitativas	76
3.7	Modelos a nivel regional obtenidos para CaCO_3 total o equivalente ($\text{kg} \cdot \text{kg}^{-1}$). Estadísticos descriptivos de la correlación (r) y Raíz del Error Cuadrático Medio (RECM_v) obtenidos tras el remuestreo en el conjunto de validación.	79
3.8	Estadísticos descriptivos del contenido de arcilla ($\text{kg} \cdot \text{kg}^{-1}$) en los suelos de la región	84
3.9	Correlaciones del contenido en arcilla con las variables predictoras cuantitativas	87
3.10	Modelos a nivel regional obtenidos para arcilla ($\text{kg} \cdot \text{kg}^{-1}$). Estadísticos descriptivos de la correlación (r) y Raíz del Error Cuadrático Medio (RECM_v) obtenidos tras el remuestreo en el conjunto de validación	88
3.11	Estadísticos descriptivos del contenido de limo ($\text{kg} \cdot \text{kg}^{-1}$) en los suelos de la región	96
3.12	Correlaciones del contenido en limo ($\text{kg} \cdot \text{kg}^{-1}$) con las variables predictoras cuantitativas	99
3.13	Modelos a nivel regional obtenidos para limo ($\text{kg} \cdot \text{kg}^{-1}$). Estadísticos descriptivos de la correlación (r_v) y Raíz del Error Cuadrático Medio (RECM_v) obtenidos tras el remuestreo en el conjunto de validación	101
3.14	Estadísticos descriptivos del contenido de arena ($\text{kg} \cdot \text{kg}^{-1}$) en los suelos de la región.	107
3.15	Correlaciones del contenido en arena ($\text{kg} \cdot \text{kg}^{-1}$) con las variables predictoras cuantitativas	110

3.16 Modelos a nivel regional obtenidos para arena ($\text{kg} \cdot \text{kg}^{-1}$). Estadísticos descriptivos de la correlación (r) y Raíz del Error Cuadrático Medio (RECM_v) obtenidos tras el remuestreo en el conjunto de validación.	111
3.17 Estadísticos descriptivos del contenido de CO ($\text{g} \cdot \text{kg}^{-1}$) en los suelos de la sierra de Mojantes	120
3.18 Correlaciones de $\log(\text{CO})$ con las variables predictoras cuantitativas en la sierra de Mojantes	123
3.19 Modelos obtenidos para Carbono Orgánico ($\text{g} \cdot \text{kg}^{-1}$) a escala local. Estadísticos descriptivos de la correlación (r) y Raíz del Error Cuadrático Medio (RECM_v) obtenidos en el remuestreo del conjunto de validación.	125
3.20 Estadísticos descriptivos del contenido de CaCO_3 total o equivalente en los suelos de la sierra de Mojantes	132
3.21 Correlaciones entre el contenido en CaCO_3 total o equivalente y las variables predictoras cuantitativas en la sierra de Mojantes	134
3.22 Modelos obtenidos a escala local para CaCO_3 total o equivalente ($\text{kg} \cdot \text{kg}^{-1}$). Estadísticos descriptivos de la correlación (r) y Raíz del Error Cuadrático Medio (RECM_v) tras el remuestreo en el conjunto de validación.	136
3.23 Estadísticos descriptivos del contenido de arcilla ($\text{kg} \cdot \text{kg}^{-1}$) en los suelos de la sierra de Mojantes. Mín. = Mínimo; Máx. = Máximo; SD = Desviación estándar; CV (%) . = Coeficiente de variación.	140
3.24 Correlaciones del contenido en arcilla con las variables predictoras cuantitativas en la sierra de Mojantes.	142
3.25 Modelos obtenidos a escala local para arcilla ($\text{kg} \cdot \text{kg}^{-1}$) a escala local. Estadísticos descriptivos de la correlación (r_v) y Raíz del Error Cuadrático Medio (RECM_v) obtenidos en el remuestreo del conjunto de validación.	144
3.26 Estadísticos descriptivos del contenido de limo ($\text{kg} \cdot \text{kg}^{-1}$) en los suelos de la sierra de Mojantes	147
3.27 Correlaciones del contenido en limo ($\text{kg} \cdot \text{kg}^{-1}$) con las variables predictoras cuantitativas en la sierra de Mojantes	149
3.28 Modelos a escala local obtenidos para limo ($\text{kg} \cdot \text{kg}^{-1}$). Estadísticos descriptivos de la correlación (r) y Raíz del Error Cuadrático Medio (RECM_v) obtenidos en el remuestreo del conjunto de validación.	151
3.29 Estadísticos descriptivos del contenido de arena ($\text{kg} \cdot \text{kg}^{-1}$) en los suelos de la sierra de Mojantes	155
3.30 Correlaciones del contenido en arena ($\text{kg} \cdot \text{kg}^{-1}$) con las variables cuantitativas en la sierra de Mojantes	157
3.31 Modelos obtenidos para arena ($\text{kg} \cdot \text{kg}^{-1}$) a escala local. Estadísticos descriptivos de la correlación (r) y Raíz del Error Cuadrático Medio (RECM_v) obtenidos en el remuestreo del conjunto de validación.	159
4.1 Estadísticos descriptivos de densidad aparente ($\text{g} \cdot \text{cm}^{-3}$) en los suelos de la sierra de las Moreras.	163
4.2 Funciones de edafotransferencia para densidad aparente	165
4.3 Resultados de las funciones de edafotransferencia para densidad aparente	166

4.4	Estadísticos descriptivos de contenido de humedad a capacidad de cambio (% en peso) en los suelos de la sierra de las Moreras	169
4.5	Funciones de edafotransferencia para contenido de humedad a -0.33 atm	171
4.6	Resultados de las funciones de edafotransferencia para contenido de humedad a Capacidad de Campo (% en peso)	172
4.7	Comparación de los mapas de contenido de humedad a Capacidad de Campo (% en peso) con los datos de Mazarrón y los datos de perfiles de LUCDEME (<i>perf</i>): r= correlación (Pearson); EMP=Error Medio de la Predicción; RECM = Raíz del Error Cuadrático Medio.	176
4.8	Estadísticos descriptivos de contenido de humedad en el punto de marchitamiento (% en peso) en los suelos de la sierra de las Moreras	178
4.9	Funciones de edafotransferencia para contenido de humedad a -15 atm	181
4.10	Resultados de las funciones de edafotransferencia para contenido de humedad a -15 atm	181
4.11	Comparación de los mapas de contenido de humedad en el punto de marchitamiento (% en peso) con los datos de Mazarrón y los datos de perfiles de LUCDEME (<i>perf</i>)	184

Lista de figuras

Pág.

1.1	Concepto del margen suavizado para una SVM lineal	9
2.1	Localización del área de estudio: Región de Murcia	20
2.2	Localización del área de estudio: sierra de Mojantes	26
2.3	Localización del área de estudio: sierra de las Moreras	27
2.4	Base de datos LUCDEME.	29
2.5	Situación de las muestras: sierra de Mojantes	31
2.6	Situación de las muestras: sierra de las Moreras	32
2.7	Configuraciones de pendiente: concavidad y convexidad.	35
2.8	Denominación de las celdas en una ventana de 3x3	35
2.9	Obtención de las componentes del vector unitario normal para calcular VRM.	36
2.10	Mapa de usos del suelo de CORINE Land Cover 1990 reclasificado	40
2.11	Mapa de suelos de LUCDEME (1986-2004) reclasificados según WRB (2007). Los números de la leyenda se refieren a las clases de suelo descritas en la tabla 2.3.	42
2.12	Equilibrio entre sesgo (bias) y varianza (variance) en la selección del modelo. Fuente: James <i>et al.</i> (2013).	49
2.13	Esquema de flujo del trabajo realizado para obtener cada uno de los mapas de las variables físico-químicas a nivel regional.	53
2.14	Esquema de flujo del trabajo realizado con las funciones de edafotransferencia para cada una de las variables hídricas analizadas.	55
3.1	Histogramas y gráficos de cuantiles de CO	60
3.2	Distribución de CO ($g \cdot kg^{-1}$) en función del uso y el tipo de suelo	61
3.3	Raíz del error cuadrático medio de los modelos a escala regional obtenidos con validación cruzada en bloques para carbono orgánico, en función del número de variables	63
3.4	Modelos a escala regional obtenidos para carbono orgánico. Histogramas de la correlación (r_v) y raíz del error cuadrático medio (RECM _v) tras el remuestreo en el conjunto de validación	65
3.5	Diagrama de dispersión de los datos de CO ($g \cdot kg^{-1}$) observados en el conjunto de validación frente los obtenidos con RF-CV a escala regional.	66
3.6	Importancia de variables en RF-CV expresada como criterio de impureza del nodo.	67

3.7	Efecto de las variables cualitativas en el modelo GAM para CO obtenido con las variables de RF-CV _{IMP}	68
3.8	Modelo GAM obtenido con las variables de RF-CV _{IMP} para CO. Curvas de respuesta para las variables suavizadas (I).	70
3.9	Modelo GAM obtenido con las variables de RF-CV _{IMP} para CO. Curvas de respuesta para las variables suavizadas (II).	71
3.10	Distribución espacial de los residuos del modelo RF-CV para carbono orgánico.	72
3.11	Gráfico de cajas con la distribución del contenido de CO a escala regional	73
3.12	Histograma de frecuencias y gráfico de cuantiles para carbonato cálcico total o equivalente ($\text{kg} \cdot \text{kg}^{-1}$)	74
3.13	Distribución de CaCO_3 total o equivalente ($\text{kg} \cdot \text{kg}^{-1}$) en función del uso y el tipo de suelo	75
3.14	Raíz del error cuadrático medio (RECM _c) de los modelos a escala regional obtenidos con validación cruzada en bloques para carbonato cálcico total o equivalente, en función del número de variables.	77
3.15	Modelos a escala regional obtenidos con validación cruzada en bloques para $\text{CaCO}_{3\text{total}}$. Histogramas de la correlación (r_v) y raíz del error cuadrático medio (RECM _v) tras el remuestreo en el conjunto de validación.	78
3.16	Diagrama de dispersión de los datos de carbonato cálcico total o equivalente ($\text{kg} \cdot \text{kg}^{-1}$) observados en el conjunto de validación frente los obtenidos con RF-CV a escala regional.	80
3.17	Importancia de variables en RF-CV para carbonato cálcico total o equivalente, expresada como criterio de impureza del nodo.	80
3.18	Modelo GAM obtenido con las variables de RF-CV para carbonato cálcico total o equivalente. Efecto de la variable suelo.	81
3.19	Modelo GAM obtenido con las variables de RF-CV para carbonato cálcico total o equivalente. Curvas de respuesta para las variables suavizadas significativas.	81
3.20	Distribución espacial de los residuos del modelo RF-CV para carbonato cálcico total o equivalente.	83
3.21	Gráfico de cajas con la distribución del contenido en carbonato cálcico total o equivalente a escala regional	84
3.22	Histograma de frecuencias y gráfico de cuantiles de arcilla.	85
3.23	Análisis exploratorio: Contenido de arcillas en función del uso y tipo de suelo.	86
3.24	Raíz del error cuadrático medio de los modelos a escala regional obtenidos con validación cruzada en bloques para arcilla, en función del número de variables	87
3.25	Modelos a escala regional obtenidos para arcilla. Histogramas de la correlación (r_v) y raíz del error cuadrático medio (RECM _v) tras el remuestreo en el conjunto de validación.	90
3.26	Diagrama de dispersión de los datos de arcilla ($\text{kg} \cdot \text{kg}^{-1}$) observados en el conjunto de validación frente los obtenidos con RF-CV a escala regional.	91
3.27	Importancia de variables en el modelo regional RF-CV para arcilla, expresada como criterio de impureza del nodo.	92
3.28	Distribución espacial de los residuos del modelo RF-CV _{IMP} para arcilla.	94
3.29	Semivariogramas de los residuos de RF-CV _{IMP} para arcilla.	94

3.30	Gráfico de cajas con la distribución del contenido en arcilla la a escala regional	95
3.31	Histograma y gráfico de cuantiles de limo ($\text{kg} \cdot \text{kg}^{-1}$)	96
3.32	Análisis exploratorio: Contenido de limo en función del uso y tipo de suelo.	98
3.33	Raíz del error cuadrático medio (RECM_c) de los modelos a escala regional obtenidos con validación cruzada en bloques para limo, en función del número de variables.	100
3.34	Modelos a escala regional obtenidos para limo. Correlación (r_v) y raíz del error cuadrático medio (RECM_v) tras el remuestreo en el conjunto de validación.	102
3.35	Diagrama de dispersión de los datos de limo (kg/kg) observados frente los obtenidos con RF-CV _{IMP} a escala regional en el conjunto de validación y comparación con el porcentaje estimado frente las otras fracciones.	103
3.36	Importancia de variables en RF-CV _{IMP} obtenido para limo, expresada como criterio de impureza del nodo.	104
3.37	Modelo GAM obtenido con las variables de RF-CV _{IMP} para limo. Curva de respuesta para la variable suavizada significativa	104
3.38	Distribución espacial de los residuos del modelo RF-CV _{IMP} para limo.	105
3.39	Gráfico de cajas con la distribución del contenido en limo a escala regional	107
3.40	Histograma y gráfico de cuantiles de arena ($\text{kg} \cdot \text{kg}^{-1}$) en los suelos de la región	108
3.41	Análisis exploratorio: Contenido de limo en función del uso y tipo de suelo.	109
3.42	Raíz del error cuadrático medio (RECM_c) de los modelos a escala regional obtenidos con validación cruzada en bloques para arena, en función del número de variables.	112
3.43	Modelos a escala regional obtenidos para arena ($\text{kg} \cdot \text{kg}^{-1}$). Histogramas de la correlación (r_v) y raíz del error cuadrático medio (RECM_v) tras el remuestreo en el conjunto de validación.	113
3.44	Diagrama de dispersión de los datos de arena ($\text{kg} \cdot \text{kg}^{-1}$) observados en el conjunto de validación frente los obtenidos con RF-CV _{IMP} a escala regional.	114
3.45	Importancia de variables en RF-CV _{IMP} para arena a escala regional, expresada como criterio de impureza del nodo.	115
3.46	Modelo GAM obtenido con las variables de RF-CV _{IMP} para arena. Efecto del tipo de suelo.	116
3.47	Modelo GAM obtenido con las variables de RF-CV _{IMP} para arena. Curvas de respuesta para las variables suavizadas significativas.	117
3.48	Distribución espacial de los residuos del modelo RF-CV _{IMP} para arena.	118
3.49	Gráfico de cajas con la distribución del contenido en arena a escala regional	118
3.50	Histogramas y gráficos de cuantiles de CO en la sierra de Mojantes	120
3.51	Distribución de CO ($\text{g} \cdot \text{kg}^{-1}$) en la sierra de Mojantes, en función del uso y tipo de suelo.	121
3.52	Raíz del error cuadrático medio (RECM_c) de los modelos obtenidos con validación cruzada en bloques para carbono orgánico en la sierra de Mojantes, en función del número de variables.	124
3.53	Modelos obtenidos para carbono orgánico en la sierra de Mojantes. Histogramas de la correlación (r_v) y la raíz del error cuadrático medio (RECM_v) tras el remuestreo en el conjunto de validación	126

3.54	Diagrama de dispersión de los datos de CO observados frente los obtenidos con RF-CV a escala local en el set de validación.	127
3.55	Importancia de variables en el modelo RF-CV para CO a escala local, expresada como criterio de impureza del nodo.	128
3.56	Modelo GAM obtenido con las variables de RF-CV para CO a escala local. Curva de respuesta de la variable suavizada significativa.	129
3.57	Comparación de los datos predichos frente los observados entre el modelo RF-CV p a escala local y la predicción de RF-CV a escala regional.	130
3.58	Gráfico de cajas del contenido de CO ($\text{g} \cdot \text{kg}^{-1}$) a escala local	131
3.59	Histograma y gráfico de cuantiles de $\text{CaCO}_{3\text{total}}$ ($\text{kg} \cdot \text{kg}^{-1}$) en la sierra de Mojantes.	132
3.60	Distribución de $\text{CaCO}_{3\text{total}}$ ($\text{kg} \cdot \text{kg}^{-1}$) en función del uso y el tipo de suelo en la sierra de Mojantes.	133
3.61	Raíz del error cuadrático medio (RECM_c) de los modelos obtenidos con validación cruzada en bloques para CaCO_3 total o equivalente a escala local, en función del número de variables.	135
3.62	Modelos obtenidos a escala local para CaCO_3 total o equivalente. Histogramas de la correlación (r_v) y la raíz del error cuadrático medio (RECM_v) tras el remuestreo en el conjunto de validación.	137
3.63	Diagrama de dispersión de los datos de CaCO_3 total o equivalente ($\text{kg} \cdot \text{kg}^{-1}$) observados en el conjunto de validación frente los obtenidos con RF-CV a escala local.	138
3.64	Importancia de variables en RF-CV para carbonato cálcico total o equivalente a escala local, expresada como criterio de impureza del nodo.	138
3.65	Modelo GAM obtenido con las variables de RF-CV para carbonato cálcico total a escala local. Curvas de respuesta para las variables suavizadas significativas.	139
3.66	Comparación de los datos de carbonato cálcico total observados frente los predichos por los modelos RF-CV a escala local y regional. obs : valores observados; RF-CVlocal : predicción del modelo RF-CV a escala local; RF-CVregional : mapa elaborado con la predicción de RF-CV a escala regional y los residuos interpolados con KO.	139
3.67	Histograma de frecuencias y gráfico de cuantiles de arcilla.	140
3.68	Análisis exploratorio: Contenido de arcillas en la sierra de Mojantes, en función del uso y tipo de suelo.	141
3.69	Raíz del error cuadrático medio (RECM_c) de los modelos a escala local obtenidos con validación cruzada en bloques para arcilla, en función del número de variables.	143
3.70	Modelos obtenidos a escala local para arcilla a escala local. Histogramas de la correlación (r_v) y la raíz del error cuadrático medio (RECM_v) del remuestreo en el conjunto de validación.	145
3.71	Diagramas de dispersión de los datos de arcilla ($\text{kg} \cdot \text{kg}^{-1}$) observados y estimados a escala local y regional.	146
3.72	Gráfico de cajas con la distribución del contenido de arcilla ($\text{kg} \cdot \text{kg}^{-1}$) a escala local	147

3.73	Histograma de frecuencias y gráfico de cuantiles de limo ($\text{kg} \cdot \text{kg}^{-1}$) en la sierra de Mojantes.	148
3.74	Análisis exploratorio: Contenido de limo ($\text{kg} \cdot \text{kg}^{-1}$) en función del uso del suelo	148
3.75	Análisis exploratorio: Contenido de limo ($\text{kg} \cdot \text{kg}^{-1}$) en función del tipo de suelo	149
3.76	Raíz del error cuadrático medio RECM_c de los modelos obtenidos con validación cruzada en bloques para limo a escala local, en función del número de variables.	150
3.77	Modelos obtenidos con validación cruzada en bloques para limo a escala local. Correlación (r_v) tras el remuestreo en el conjunto de validación.	152
3.78	Modelos obtenidos con validación cruzada en bloques para limo a escala local. Raíz del error cuadrático medio (RECM_v) tras el remuestreo en el conjunto de validación.	153
3.79	Importancia de variables en el modelo $\text{RF-CV}_{\text{IMP}}$ para limo a escala local, expresada como criterio de impureza del nodo.	154
3.80	Diagramas de dispersión de los datos de limo $\text{kg} \cdot \text{kg}^{-1}$ observados frente los obtenidos con $\text{RF-CV}_{\text{IMP}}$ a escala local y comparación con el resultado del mapa a escala regional.	154
3.81	Gráfico de cajas con la distribución del contenido de limo ($\text{kg} \cdot \text{kg}^{-1}$) a escala local	155
3.82	Histograma y gráfico de cuantiles de arena ($\text{kg} \cdot \text{kg}^{-1}$) en la sierra de Mojantes .	156
3.83	Análisis exploratorio: Contenido de arena en función del uso del suelo.	156
3.84	Análisis exploratorio: Contenido de arena en función del tipo de suelo.	157
3.85	Raíz del error cuadrático medio (RECM_c) de los modelos obtenidos con validación cruzada en bloques para arena a escala local, en función del número de variables.	158
3.86	Modelos obtenidos para arena a escala local. Histogramas de la correlación (r_v) y raíz del error cuadrático medio (RECM_v) tras el remuestreo en el conjunto de validación.	160
3.87	Diagrama de dispersión de los datos de arena ($\text{kg} \cdot \text{kg}^{-1}$) observados en el set de validación frente los obtenidos con SVM-CV a escala local y comparación con el mapa $\text{RF-CV}_{\text{IMP}}+\text{KO}$ a escala regional.	161
3.88	Gráfico de cajas con la distribución del contenido de arena ($\text{kg} \cdot \text{kg}^{-1}$) a escala local	162
4.1	Histograma y gráfico de cuantiles de los datos de D_a ($\text{g} \cdot \text{cm}^{-3}$) en la sierra de las Moreras.	164
4.2	Distribución de la densidad aparente ($\text{g} \cdot \text{cm}^{-3}$) en función de los usos del suelo en la sierra de las Moreras	164
4.3	Distribución de la densidad aparente ($\text{g} \cdot \text{cm}^{-3}$) en función de los tipos de suelo en la sierra de las Moreras.	165
4.4	Diagrama de dispersión de los datos de D_a ($\text{g} \cdot \text{cm}^{-3}$) observados frente los obtenidos con las funciones de edafotransferencia recopiladas.	167
4.5	Gráfico de cajas con la distribución de la densidad aparente ($\text{g} \cdot \text{cm}^{-3}$) a escala local	168

4.6	Histograma y gráfico de cuantiles de los datos de contenido de humedad a capacidad de campo (% en peso) en la sierra de las Moreras.	169
4.7	Distribución de los datos de contenido de humedad a capacidad de campo (% en peso) en función del uso y tipo de suelo.	170
4.8	Diagrama de dispersión del contenido de humedad a capacidad de cambio observado frente los valores obtenidos con las funciones de edafotransferencia recopiladas.	174
4.9	Diagrama de dispersión del contenido de humedad a capacidad de cambio observado frente los valores obtenidos con en los mapas con la función de edafotransferencia (PTF) y el modelo de regresión (LM)	177
4.10	Diagramas de cajas del contenido de humedad a capacidad de cambio	177
4.11	Histograma y gráfico de cuantiles del contenido de humedad a -15 atmósferas (% peso) en la sierra de las Moreras.	178
4.12	Distribución del contenido de humedad a -15 atm (% peso) en función de los usos del suelo en la sierra de las Moreras	179
4.13	Distribución del contenido de humedad a -15 atm (% peso) en función de los tipos de suelo en la sierra de las Moreras.	180
4.14	Diagramas de dispersión del contenido de humedad a -15 atmósferas en relación al obtenido con las funciones de edafotransferencia recopiladas.	183
4.15	Diagrama de dispersión del contenido de humedad en el punto de marchitamiento observado frente los valores obtenidos con la función de edafotransferencia (PTF) y el modelo de regresión (LM)	185
4.16	Diagrama de cajas del contenido de humedad a -15 atmósferas	186

Referencias bibliográficas

- ABBASI, Y.; GHANBARIAN-ALABIJEH, B.; LIAGHAT, A.M. y SHORAFI, M. (2011). «Evaluation of Pedotransfer Functions for Estimating Soil Water Retention Curve of Saline and Saline-Alkali Soils of Iran». *Pedosphere*, **21(2)**, pp. 230–237.
- ADHIKARI, K.; KHEIR, R.B.; GREVE, M.B.; BØCHER, P.K.; MALONE, B.P.; MINASNY, B.; MCBRATNEY, A.B. y GREVE, M.H. (2013a). «High-resolution 3-D mapping of soil texture in Denmark». *Soil Science Society of America Journal*, **77**, pp. 860–876.
- ADHIKARI, K.; KHEIR, R.B.; GREVE, M.B. y GREVE, M.H. (2013b). «Comparing Kriging and Regression Approaches for Mapping Soil Clay Content in a Diverse Danish Landscape». *Soil Science*, **178(9)**, pp. 00–00.
- AINA, P.O. y PERISAWAMY, S.P. (1985). «Estimating available water-holding capacity of western Nigerian soils from soil texture and bulk density, using core and sieved samples». *Soil Science*, **140**, pp. 55–58.
- AKPA, I.C.; ODEH, I.O.A.; BISHOP, T.F. y HARTEMINK, A.E. (2014). «Digital mapping of soil particle-size fractions for Nigeria». *Soil Science Society of America Journal*, **78**, pp. 1953–1966.
- AL-QINNA, M.I y JABER, S. M. (2013). «Predicting Soil Bulk density using advanced pedotransfer functions in an arid environment». *Transactions of the ASABE*, **56(3)**, pp. 963–976.
- ALBALADEJO, J.; ORTIZ, R.; GARCÍA FRANCO, N.; RUIZ NAVARRO, A.; ALMAGRO, M.; GARCÍA PINTADO, J. y MARTÍNEZ MENA, M. (2013). «Land use and climate change impacts on soil organic carbon stocks in semi-arid Spain». *Journal of Soils and Sediments*, **13**, pp. 265–277.
- ALEXANDER, E.B. (1980). «Bulk density of Californian soils in relation to other soil properties». *Soil Sc. Soc. Am. J.*, **44**, pp. 689–692.
- ALEXANDER, M. (1977). *Introduction to Soil Microbiology*. Wiley, New York, 2ª edición.
- ALIAS, L. y ORTIZ, R. (1986–2004). «Memorias y mapas de suelos de las hojas del MTN a escala 1:100.000. Proyecto LUCDEME». *Informe Técnico*, Ministerio de Medio Ambiente-ICONA.

- ALLEN, T. F. H. y STARR, T. B. (1982). *Hierarchy: Perspectives for ecological complexity*. University of Chicago Press, Chicago.
- ANAGU, I.; INGWERSEN, J.; UTERMANN, J. y STRECK, T. (2009). «Estimation of heavy metal sorption in German soils using artificial neural networks». *Geoderma*, **152**, pp. 104–112.
- ANNE, P. (1945). «Organic soil carbon by reduction of CrO₃». *Soil Science*, **40**, pp. 311–320.
- ARROUAYS, D.; DESLAIS, W. y BADEAU, V. (2001). «The carbon content of topsoil and its geographical distribution in France». *Soil Use and Management*, **17**, pp. 7–11.
- ARROUAYS, D.; MCKENZIE, N.; HEMPEL, J.; RICHER DE FORGES, A. y MCBRATNEY, A. (Eds.) (2014). *GlobalSoilMap: Basis of the global spatial soil information system*. CRC Press, Taylor and Francis.
- ARRUDA, F.B.; ZULLO, J.R. y DE OLIVEIRA, J.B. (1987). «Parâmetros de solo para o cálculo da água disponível com base na textura do solo». *Revista Brasileira de Ciência do Solo*, **11**, pp. 11–15.
- ARYA, L.M. y PARIS, J.F. (1981). «A physicoempirical model to predict soil moisture characteristics from particle-size distribution and bulk density data». *Soil Science Society of America Journal*, **45**, pp. 1023–1030.
- BALDOCK, J.A. y SKJEMPSTAD, J.O. (2000). «Role of the soil matrix and minerals in protecting natural organic materials against biological attack». *Organic Geochemistry*, **31**, pp. 697–710.
- BALLABIO, CRISTIANO (2009). «Spatial prediction of soil properties in temperate mountain regions using support vector regression». *Geoderma*, **151(3–4)**, pp. 338–350.
- BARAHONA, E. y SANTOS, F. (1981). «Estudios de correlación y regresión de diversos parámetros analíticos de 52 perfiles de suelos del sector Montiel-Alcaraz-Bienservida (Ciudad Real-Albacete)». *Anales de Edafología y Agrobiología*, **40 (5-6)**, pp. 761–773.
- BATJES, N. H. (1996). «Development of a world data set of soil water retention properties using pedotransfer rules». *Geoderma*, **71**, pp. 31–51.
- BAUDER, J.W.; RANDALL, G.W. y SWANN, J.B. (1981). «Effect of four continuous tillage systems on mechanical impedance of a clay loam soil». *Soil Science Society of America Journal*, **45**, pp. 802–806.
- BECKETT, P. y WEBSTER, R. (1971). «Soil variability: a review». *Soils and Fertilizers*, **34**, pp. 1–15.
- BEHRENS, T.; FÖRSTER, H.; SCHOLTEN, T.; STEINRÜKEN, U.; SPIES, E. y GOLDSCHMITT, M. (2005). «Digital soil mapping using artificial neural networks». *J. Plant Nutr. Soil Sci.*, **168**, pp. 21–33.
- BELL, J.C.; CUNNINGHAM, R.L. y HAVENS, M.W. (1994). «Soil drainage class probability mapping using a soil-landscape model». *Soil Science Society of America Journal*, **58**, pp. 464–470.

- BELL, M.A. (1993). «Organic matter, soil properties and wheat production in the High Valley of Mexico». *Soil Science Society of America Journal*, **59(3)**, pp. 865–871.
- BELL, M.A. y VAN KEULEN, H (1995). «Soil pedotransfer functions for four Mexican soils.» *Soil Science Society of America Journal*, **59(3)**, pp. 865–871.
- BENITES, V. M.; MACHADO, P.L.O.A.; FIDALGO, E.C.C. y COELHO, B.E., M. R. MADARI (2007). «Pedotransfer functions for estimating soil bulk density from existing soil survey reports in Brazil». *Geoderma*, **139(1–2)**, pp. 90–97.
- BERGSTROM, D.W.; MONREAL, C.M. y ST. JACQUES, E. (2001). «Influence of tillage practice on carbon sequestration is scale-dependent». *Can. J. Soil Sci.*, **81**, pp. 63–70.
- BERNOUX, M.; ARROUAYS, D.; CERRI, C.; VOLKOFF, B. y JOLIVET, C. (1998). «Bulk density of Brazilian Amazon soils related to other soil properties». *Soil Science Journal*, **162**, pp. 743–749.
- BEVEN, K.J. y KIRKBY, M.J. (1979). «Physically based, variable contributing area model of basin hydrology». *Hydrological Sciences Bulletin*, **24**, pp. 43–69.
- BHATTACHARYA y SOLOMATINE, D.P. (2006). «Machine learning in soil classification». *Neural Networks*, **19(2)**, pp. 186–195.
- BILGILI, A.V.; VAN ES, H.M.; AKBAS, F.; DURAK, A. y HIVELEY, W.D. (2010). «Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey». *Journal of Arid Environments*, **74**, pp. 229–238.
- BIVAND, R. (2013). «spgrass6: Interface between GRASS 6 and R». Último acceso: 10/10/2014. <http://cran.r-project.org/web/packages/spgrass6/>
- BOIX-FAYOS, C.; CALVO-CASES, A. y IMESO, A.C. (2001). «Influence of soil properties on the aggregation of some Mediterranean soils and the use of aggregate size and stability as land degradation indicators». *Catena*, **44**, pp. 47–67.
- BOIX-FAYOS, C.; DE VENTE, J.; ALBALADEJO, J. y MARTÍNEZ MENA, M. (2009). *Agriculture, Ecosystems and Environment*, pp. 75–85.
- BOONE, F.R. y VEEN, D.E. (1994). *Soil compaction in crop production*. Mechanisms of crop responses to soil compaction, pp. 237–264. Elsevier, New York.
- BOTULA, Y.D.; CORNELIS, W.M.; BAERT, G. y VAN RANST, E. (2012). «Evaluation of pedotransfer functions for predicting water retention of soils in Lower Congo (D.R. Congo)». *Agricultural Water Management*, **111**, pp. 1–10.
- BOUAJILA, A. y GALLALI, T. (2008). «Soil organic carbon fractions and aggregate stability in carbonated and no carbonated soils in Tunisia». *Journal of Agronomy*, **7**, pp. 127–137.
- BOUMA, J. (1989). «Using soil survey data for quantitative land evaluation». *Adv. Soil Sci.*, **9**, pp. 177–213.

- BOUMA, J. y VAN LANEN, H.A.J. (1987). «Transfer functions and threshold values: from soil characteristics to land qualities.» En: K. Beek; P.A. Barrough y D.D. McCormack (Eds.), *Proceedings of the International Workshop on Quantified Land Evaluation Procedures*, .
- BRADY, L. y WEIL, R. R. (2002). *The Nature and Properties of Soil*. Springer, Netherlands, 13^a edición.
- BRAHIM, N.; BERNOUX, M.; BLAVET, D. y GALLALI, T. (2010). «Tunisian Soil Organic Carbon Stocks». *International Journal of Soil Science*, **5(1)**, pp. 34–40.
- BRAHIM, N.; BERNOUX, M. y GALLALI, T. (2012). «Pedotransfer functions to estimate soil bulk density for Northern Africa: Tunisia case». *Journal of Arid Environments*, **81**, pp. 77–83.
- BREIMAN, L. (2001a). «Random Forests». *Machine Learning*, **45(1)**, pp. 5–32.
- BREIMAN, L. (2001b). «Statistical Modeling: The two cultures». *Statistical Science*, **16(3)**, pp. 199–215.
- BREIMAN, L.; FREIDMAN, J.; OLSHEN, R. y STONE, C. (1984). *Classification and Regression Trees*. Wadsworth and Brooks/Cole, Monterey, CA..
- BROADBENT, F.E. (1953). «The Soil Organic fraction». *Advances in Agronomy*, **5**, pp. 153–183.
- BRONICK, C.J. y LAL, R. (2005). «Soil structure and management». *Geoderma*, **124**, pp. 3–22.
- BROWN, D.J.; CLAYTON, M.K. y MCSWEENEY, K. (2004). «Potential terrain controls on soil color, texture contrast and grain-size deposition for the original catena landscape in Uganda». *Geoderma*, **122**, pp. 51–72.
- BRUNGARD, C.W.; BOETTINGER, C.L.; DUNIWAY, M.C. y WILLS, T.C., S.A. EDWARDS (JR). (2015). «Machine learning for predicting soil classes in three semi-arid landscapes». *Geoderma*, **239–240**, pp. 68–83.
- BUI, E.N.; HENDERSON, B.L. y K., VIERGEVER (2006). «Knowledge discovery from models of soil properties developed through data mining». *Ecological Modelling*, **(191)**, pp. 431–446.
- BUOL, S.W.; HOLE, F.D. y MCCRACKEN, R.J. (1989). *Soil Genesis and Classification*. Iowa State University Press.
- BURGESS, M. y WEBSTER, R. (1980a). «Optimal interpolation and isarithmic mapping of soil properties, I. The semi-variogram and punctual kriging». *European Journal of Soil Science*, **31**, pp. 315–331.
- BURGESS, M. y WEBSTER, R. (1980b). «Optimal interpolation and isarithmic mapping of soil properties, II. Block kriging». *European Journal of Soil Science*, **31**, pp. 333–341.
- BURKE, I.C.; ELLIOT, E.C. y COLE, C.V. (1995). «Influence of macroclimate, landscape position and management on soil organic matter in agroecosystems.» *Ecological Applications*, **5**, pp. 124–131.

- BURROUGH, P. A. (1993). «Soil variability: a late 20th century view.» *Soils and Fertilizers*, **56**, pp. 529–562.
- BURROUGH, P. A.; VAN GAANS, P. F. M. y MACMILLAN, R. A. (2000). «High-resolution landform classification using fuzzy k-means». *Fuzzy Sets and Systems*.
- CAI, S.; ZHANG, R.; LIU, L. y ZHOU, D. (2010). «A method of salt-affected soil information extraction based on a support vector machine with texture features.» *Mathematical and Computer Modelling*, **51(11–12)**, pp. 1319–1325.
- CAMBARDELLA, C.A.; MOORMAN, T.B.; NOVAK, J. M.; PARKIN, T. B.; KARLEN, D. L.; TURCO, R. F. y KONOPKA, A. E. (1994). «Field scale variability in central Iowa soils.» *Soil Sci. Soc. Am. J.*, **58**, pp. 1501–1511..
- CANDY, I. y BLACK, S. (2009). «The timing of Quaternary calcrete development in semiarid southeast Spain: Investigating the role of climate on calcrete genesis». *Sedimentary Geology*, **218(1–4)**, pp. 6–15.
- CHAPMAN, H.D. (1965). En: C.A. Black (Ed.), *Methods of soil analysis - Chemical and microbiological properties*, volumen 9, Cation-exchange capacity, pp. 891–901. Agronomy.
- CHS, CONFEDERACIÓN HIDROGRÁFICA DEL SEGURA (2013). «Ministerio de Agricultura, Alimentación y Medio Ambiente». Último acceso: 16/10/2014.
<http://www.chsegura.es/chs/cuenca/resumendedatosbasicos/marcoadministrativo/>
- CLOUGH, A. y SKJEMSTAD, J.O. (2000). «Physical and chemical protection of soil organic carbon in three agricultural soils with different contents of calcium carbonate». *Australian Journal of Soil Resources*, **38**, p. 1005–1016.
- DANALATOS, N.G.; KOSMAS, C.S.; DRIESSEN, P.M. y YASSOGLU, N. (1994). «Estimation of the draining soil moisture characteristics from standard data as recorded in soil surveys». *Geoderma*, **64**, pp. 155–165.
- DE ALBA, S.; LINDSTROM, M.; SCHUMACHER, T.E. y MALO, D.D. (2004). «Soil landscape evolution due to soil redistribution by tillage: a new conceptual model of soil catena evolution in agricultural landscapes». *Catena*, **58**, pp. 77–100.
- DE GRYZE, S.; SIX, J.; PAUSTIAN, K.; MORRIS, S.J.; PAUL, E. y MERCKX, R. (2004). «Soil organic carbon pool changes following land-use conversion». *Global Change Biology*, **10**, pp. 1120–1132.
- DE VOS, B.; VAN MEIRVENNE, M.; QUATAERT, P.; DECKERS, J. y MUYS, B. (2005). «Predictive quality of pedotransfer functions for estimating bulk density of forest soils.» *Soil Sc. Soc. Am. J.*, **69**, pp. 500–510.
- DEMERS, M. N. (2005). *Fundamentals of Geographic Information Systems*. Wiley, 3ª edición.
- DENG, Y. (2005). *Evaluating Input data for DEM-Based Environmental Analysis and Classification*. Dissertation, University of Southern California.

- DIJKERMAN, J.C. (1988). «An Ustult-Aquult-Tropept Catena in Sierra Leone, West Africa, II. Land Qualities and Land». *Geoderma*, **42**, pp. 29–49.
- DOBOS, E.; CARRÉ, F.; HENGL, T.; REUTER, H.I. y TÓTH, G. (2006). *Digital Soil Mapping as a support to production of functional maps*. Office for Official Publications of the European Communities, Luxemburg..
- DUCHAFOUR, PH. (1970). *Précis de Pedologie*. Masson y Cie.
- DUCHAFOUR, PH. (1988). *Pedologie*. Masson y Cie.
- DUPOUEY, J.L.; THIMONIER, A. y BEHR, P. (1997). «Variation de la densité des sols des hêtraies du nord-est de la France en relation avec leurs caractéristiques physico-chimiques». *Etude et Gestion des Sols*, **4(1)**, pp. 43–52.
- ELITH, J.; LEATHWICK, J.R. y HASTIE, T. (2008). «A working guide to boosted regression trees». *Journal of Animal Ecology*, **77(4)**, pp. 802–818.
- ELSHORBAGY, A. y PARASURAMAN, K. (2008). «On the relevance of using artificial neural networks for estimating soil moisture content». *Journal of Hydrology*, **362(1–2)**, pp. 1–18.
- EUROPEAN ENVIRONMENT AGENCY (1995). «CORINE Land Cover». *Informe Técnico*, Commission of the European Communities, Roma.
- FAGERLAND, M.W. y SANDVIK, L. (2009). «The Wilcoxon–Mann–Whitney test under scrutiny». *Statistics in Medicine*, **28**, pp. :1487–1497.
- FAO-UNESCO (1974). *Soil Map of the World*. Paris. Vol I.
- FERNÁNDEZ-UGALDE, O.; VIRTO, I.; BARRÉ, P.; GARTZÍA-BENGOETXEEA, N.; ENRIQUE, A.; IMAZ, M.J. y BESCANSÀ, P. (2011). «Effect of carbonates on the hierarchical model of aggregation in calcareous semiarid Mediterranean soils». *Geoderma*, **164(3–4)**, pp. 203–214.
- FRANZLUEBBERS, A.J. (2005). «Soil organic carbon sequestration and agricultural greenhouse gas emissions in the southeastern USA». *Soil and Tillage Research*, **83**, pp. 120–147.
- FRENKEL, H.; HADAS, A. y JURY, W.A. (1978). «The effect of salt precipitation and high sodium concentration on soil hydraulic conductivity and water retention.» *Water Resour. Res.*, **14**, pp. 217–222.
- FREUND, Y. y SCHAPIRE, R.E. (1996). «Experiments with a new boosting algorithm». En: *Machine Learning: Proceedings of the Thirteenth International Conference*, pp. 148–156.
- FRIEDL, M.A. y BRODLEY, C.E. (1997). «Decision tree classification of land cover from remotely sensed data». *Remote Sensing Environ.*, **61**, pp. 399–409.
- FRIEDMAN, J.H. (1991). «Multivariate adaptive regression splines (with discussion)». *Annals of Statistics*, **19**, pp. 1–141.

- GALLANT, J.C. y DOWLING, T.D. (2003). «A multi-resolution index of valley bottom flatness for mapping depositional areas». *Water Resour. Res.*, **39**, pp. 1347–1360.
- GAULTIER, J.; FARENHORST, A. y CROW, G. (2006). «Spatial variability of soil properties and 2,4-D sorption in a hummocky field as affected by landscape position and soil depth.» *Canadian Journal of Soil Science*, **86(1)**, pp. 89–95.
- GEISSEN, V.; KAMPICHLER, C.; LÓPEZ DE LLERGO JUÁREZ, J.J. y GALINDO ACÁNTARA, A. (2007). «Superficial and subterranean soil erosion in Tabasco, tropical Mexico: Development of a decision tree modeling approach». *Geoderma*, **139(3–4)**, pp. 277 – 287.
- GEISSER, S. (1993). *Predictive Inference: An Introduction*. Nmero 55 en Monographs on Statistics and Applied Probability. Chapman & Hall.
- GESSLER, P.E.; CHADWICK, O.A.; CHAMRON, F.; HOLMES, K. y ALTHOUSE, L. (2000). «Modeling soil-landscape and ecosystem properties using terrain attributes». *Soil Science Society of America Journal*, (**64**), pp. 2046–2056.
- GESSLER, P.E.; CHADWICK, Q.A.; CHAMRAN, F.; ALTHOUSE, L.D. y HOLMES, K.W. (1995). «Soil Landscape Modelling and Spatial prediction of soil attributes». *Soil Sci Soc. Am. J.*, **64**, pp. 2046–2056.
- GIL VÁZQUEZ, J.M. (2010). *Evaluación de indicadores edáficos para caracterizar la calidad del suelo en ecosistemas mediterráneos semiáridos (Mazarrón, Murcia)*. Tesis doctoral, Universidad de Murcia.
- GIVI, J.; PRASHER, S.O. y PATEL, R.M. (2004). «Evaluation of pedotransfer functions in predicting the soil water contents at field capacity and wilting point». *Agricultural Water Management*, **70(2)**, pp. 83–96.
- GOMARIZ CASTILLO, F. y ALONSO SARRÍA, F. (2013). «An R script to model monthly climatic variables with GLM to be used in hydrological modelling». En: *9th International R User Conference*, p. 168. Spain. Último acceso: 10/2/2015.
http://www.edii.uclm.es/~useR-2013/abstracts/files/92_gomariz_alonso2.pdf
- GÓMEZ, Á.; F.LAVADO y SCHNABEL, S. SUSANNE (2011). «Modeling soil properties at a regional scale using GIS and Multivariate Adaptive Regression Splines». (Último acceso: 12/6/2015).
<http://geomorphometry.org/system/files/GomezGutierrez2011bgeomorphometry.pdf>
- GOOVAERTS, P. (1997). *Geostatistics for Natural Resources Evaluation*. Oxford Univ. Press.
- GRASS DEVELOPMENT TEAM (2013). «GRASS GIS 6.4.4svn Reference Manual». Último acceso: 1/12/2014.
<http://grass.osgeo.org/grass64/manuals/>
- GRAY, J.M.; HUMPHREYS, G.S. y DECKERS, J.A. (2009). «Relationships in soil distribution as revealed by a global soil database». *Geoderma*, **150(3–4)**, pp. 309–323. ISSN 0016-7061.

- GRIMM, R. y BEHRENS, T. (2010). «Uncertainty analysis of sample locations within digital soil mapping approaches». *Geoderma*, **155**, pp. 154–163.
- GRIMM, R.; BEHRENS, T.; MÄRKER, M. y EISENBEER, H. (2008). «Soil organic carbon concentrations and stocks on Barro Colorado Island — Digital soil mapping using Random Forests analysis». *Geoderma*, **146**, pp. 102–113.
- GRUMWALD, S. (2009). «Multi-criteria characterization of recent digital soil mapping and modeling approaches». *Geoderma*, **152(3–4)**, pp. 195–207.
- GUISAN, A. y HARRELL, F.E. (2000). «Ordinal response regression models in ecology». *Journal of Vegetation Science*, **11**, pp. 617–626.
- GUPTA, S.C. y LARSON, W.E. (1979). «Estimating soil water retention characteristics from particle size distribution, organic matter percent, and bulk density». *Water Res. Res.*, **15**, pp. 1633–1635.
- HANSEN, B.; SCHJØNNING, P. y SIBBESEN, E. (1999). «Roughness indices for estimation of depression storage capacity of tilled soil surfaces». *Soil and Tillage Research*, **52(1–2)**, pp. 103–111.
- HASTIE, T.J.; TIBSHIRANI, R. y FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, 2ª edición.
- HASTIE, T.J. y TIBSHIRANI, R.J. (1990). *Generalized Additive Models*. Chapman & Hall.
- HATTAB, N.; HAMBALI, R.; MOTELICA HEINO, M.; BOURRAT, X. y MENCH, M. (2013). «Application of neural network model for the prediction of chromium concentration in phyto-remediated contaminated soils». *Journal of Geochemical Exploration*, **128(0)**, pp. 25–34.
- HENDERSON, B.L.; BUI, E.N.; MORAN, C.J. y SIMON, D.A.P. (2005). «Australia-wide predictions of soil properties using decision trees». *Geoderma*, **124(3–4)**, pp. 383–398.
- HENGL, T.; HEUVELINK, G. B. M. y STEIN, A. (2004). «A generic framework for spatial prediction of soil variables based on regression-kriging». *Geoderma*, **120(1–2)**, pp. 75–93. ISSN 0016-7061.
- HÉNIN, S.; GRAS, R. y MONNIER, G. (1969). *Le profil cultural. L'état Physique du sol et ses conséquences agronomiques*. Masson et Cie, Paris.
- HENNESSY, J.T.; GIBBENS, R.P.; TROMBLE, J.M. y CARDENAS, M. (1983). «Water properties of caliche». *Journal of Range Management*, **36(6)**, pp. 723–726.
- HEUSCHER, S.; BRANDT, C.C. y JARDINE, P.M. (2005). «Using Soil Physical and Chemical Properties to Estimate Bulk Density». *Soil Science Society of America Journal*, **69**, pp. 51–56.

- HIEMSTRA, P.H.; PEBESMA, E.J.; TWENHÖFEL, C.J.W. y HEUVELINK, G.B.M. (2008). «Real-time automatic interpolation of ambient gamma dose rates from the Dutch Radioactivity Monitoring Network». *Computers Geosciences*. DOI: <http://dx.doi.org/10.1016/j.cageo.2008.10.011>.
- HILLEL, D (1998). *Environmental soil physics*. Academic Press, San Diego, California.
- HUNTINGTON, T.G.; JOHNSON, A.H.; SICCAMI, T.G. y RYANS, D.F. (1989). «Carbon, organic matter and bulk density relationship in forested Spodosol.» *Soil Science*, **148**, pp. 380–386.
- IGME (1993). *Geological Map of Spain*. Geological and Mining Institute of Spain, IGME.
- INSTITUTO GEOGRÁFICO NACIONAL DE ESPAÑA (2013). <https://www.cnig.es/>. Último acceso: 20/10/2013.
- IPCC (2001). *IPCC Third Assessment Report: Climate Change 2001. Working Group II: Impacts, Adaptation and Vulnerability*. 4. Hydrology and Water Resources. Intergovernmental Panel on Climate Change.
- IPCC (2007). *IPCC Fourth Assessment Report: Climate Change 2007. Working Group II: Impacts, Adaptation and Vulnerability*. 4. Ecosystems, their Properties, Goods and Services. Cambridge University Press.
- IUSS (2006). *The future of soil science*. IUSS.
- IUSS-WRB (2007). «Base Referencial Mundial del Recurso Suelo.» *Informe Técnico*, FAO, Roma.
- IWAHASHI, J. y KAMIYA, I. (1995). «Landform classification using digital elevation model by the skills of image processing—mainly using the Digital National Land Information». *Geoinformatics*, **6(2)**, pp. 97–108.
- JALABERT, S.S.M.; MARTIN, M.P.; RENAUD, J.P.; BOULONNE, L.; JOLIVET, C.; MONTANARELLA, L. y ARROUAYS, D. (2010). «Estimating forest soil bulk density using boosted regression modeling». *Soil Use and Management*, **26**, pp. 516–528.
- JAMES, G.; WITTEN, D.; HASTIE, T. y TIBSHIRANI, R. (2013). *An Introduction to Statistical Learning*.
- JENNESS, J. (2005). «Topographic Position Index (tpi-jen.avx) Extension for ArcView 3.x.» *Informe Técnico*, Jenness Enterprises. <http://www.jennessent.com>
- JENNY, H. (1941). *Factors of soil formation*. McGraw-Hill, New York.
- JOBAGY, E. y JACKSON, R.B. (2000). «The vertical distribution of soil organic carbon and its relation to climate and vegetation». *Ecological Applications*, **10(2)**, pp. 423–436.
- KARATZOGLU, A.; SMOLA, A.; HORNIK, K. y ZEILEIS, A. (2004). «kernlab - An S4 Package for Kernel Methods in R». *Journal of Statistical Software*, **11(9)**, pp. 1–20.

- KARLEN, D.L.; MAUSBACH, M.J.; DORAN, J.W.; CLINE, R.G.; HARRIS, R.F. y SCHUMAN, G.E. (1997). «Soil quality: A concept, definition, and framework for evaluation». *Soil Sci. Soc. Am. J.*, **61**, pp. 4–10.
- KAUR, R.; KUMAR, S. y GURUNG, H.P. (2002). «Pedotransfer functions (PTFs) for estimating soil bulk density from basic soil data and its comparison with existing PTFs». *Australian Journal of Soil Research*, **40**, pp. 847–857.
- KHODAVERDILOO, H.; HOMAEE, M.; VAN GENUCHTEN, M.Th y DASHTAKI, S.G. (2011). «Deriving and validating pedotransfer functions for some calcareous soils». *Journal of Hydrology*, **399**, pp. 93–99.
- KOVACEVIC, M.; BAJAT, B. y GAJIC, B. (2010). «Soil type classification and estimation of soil properties using support vector machines». *Geoderma*, **154(3–4)**.
- KRAVCHENKO, A. y BULLOCK, D. (1999). «A comparative study of interpolation methods for mapping soil properties.» *Agron. J.*, **91**, pp. 393 – 400..
- LAGACHERIE, P.; BARET, F.; FERET, J.B.; NETTO, J.M. y ROBBEZ-MASSON, J.M. (2008). «Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyper-spectral measurements». *Remote Sensing of Environment*, **112**, pp. 825–835.
- LAGACHERIE, P. y MCBRATNEY, A.B. (2006). «Chapter 1 Spatial Soil Information Systems and Spatial Soil Inference Systems: Perspectives for Digital Soil Mapping». En: A.B. McBratney P. Lagacherie y M. Voltz (Eds.), *Digital Soil Mapping - An Introductory Perspective*, volumen 31 de *Developments in Soil Science*, pp. 3–22. Elsevier.
- LAGANIÈRE, J.; ANGERS, D.A. y PARÉ, D. (2010). «Carbon accumulation in agricultural soils after afforestation: a meta-analysis». *Global Change Biology*, **16**, pp. 439–453.
- LAL, R. (1979). «Physical properties and moisture retention characteristics of some Nigerian soils.» *Geoderma*, **21**, pp. 209–223.
- LAL, R. (2002). «Soil carbon dynamics in cropland and rangeland». *Environmental pollution*, **116**, pp. 353–362.
- LAL, R. (2004). «Carbon Sequestration in Dryland Ecosystems». *Environmental management*, **33(4)**, pp. 528–544.
- LARK, R. M.; RAWLINS, B. G.; ROBINSON, D. A.; LEBRON, I. y TYE, A. M. (2014). «Implications of short-range spatial variation of soil bulk density for adequate field-sampling protocols: methodology and results from two contrasting soils». *European Journal of Soil Science*, **65(6)**, pp. 803–814.
- LE BISSONNAIS, Y. (1996). «Aggregate stability and assessment of crustability and erodibility: 1. Theory and methodology». *European Journal of Soil Science*, **47**, pp. 425–437.

- LEMERCIER, B.; LACOSTE, M.; LOUM, M. y WALTER, C. (2012). «Extrapolation at regional scale of local soil knowledge using boosted classification trees: A two-step approach». *Geoderma*, **171–172(0)**, pp. 75 – 84.
- LEONE, A.P.; WRIGHT, G.G. y CORVES, C. (1995). «The application of satellite remote sensing for soil studies in upland areas of Southern Italy.» *International Journal of Remote Sensing* . 01/1995; 16(6):1087-1105., **16(6)**, pp. 1087–1105.
- LIAW, A. y WIENER, M. (2002). «Classification and Regression by randomForest». *R News*, **2(3)**, pp. 18–22.
- LIEß, M.; GLASER, B. y HUWE, B. (2012). «Uncertainty in the spatial prediction of soil texture: Comparison of regression tree and Random Forest models». *Geoderma*, **170(5)**, pp. 70–79.
- LIU, F.; GENG, X.; ZHU, A.; FRASER, W. y WADDELL, A. (2012a). «Soil texture mapping over low relief areas using land surface feedback dynamic patterns extracted from MODIS». *Geoderma*, **171–172**, pp. 44–52.
- LIU, ZHI-PENG; SHAO, MING-AN y WANG, YUN-QIANG (2012b). «Estimating Soil Organic Carbon Across a Large-Scale Region: A State-Space Modeling Approach». *Soil Science*, **177**, pp. 607–618.
- LOGSDON, S.D. y CAMBARDELLA, C.A. (2000). «Temporal changes in small depth incremental soil bulk density». *Soil Sci. Soc. Am. J.*, **64**, pp. 710–714.
- LUCDEME (1986-2004). «Memorias y mapas de suelos de las hojas del MTN a escala 1:50.000. Proyecto LUCDEME». *Informe Técnico*, Ministerio de Medio Ambiente.
- MAESTRE-VALERO, J. F.; MARTÍNEZ-GRANADOS, D.; MARTÍNEZ-ÁLVAREZ, V. y CALATRAVA, J. (2013). «Socio-Economic Impact of Evaporation Losses from Reservoirs Under Past, Current and Future Water Availability Scenarios in the Semi-Arid Segura Basin». *Water Resources Management*, **27(5)**, pp. 1411–1426.
- MARTÍNEZ MENA, M.; LÓPEZ, J.; ALMAGRO, M.; ALBALADEJO, J.; CASTILLO, V.; ORTIZ, R. y BOIX-FAYOS, C. (2011). «Organic carbon enrichment in sediments: Effects of Rainfall characteristics under different land uses in a Mediterranean area». *Catena*, pp. 36–42.
- MASUTTI, M.M. (1997). *Caracterização da água disponível a partir de parâmetros físico-hídricos em solos da zona da mata do Estado de Pernambuco*. Tesina o Proyecto, Universidade Federal Rural de Pernambuco.
- MATHERON, G. (1969). «Le krigeage universel». *Informe Técnico*, École de Mines de Paris.
- MATHERON, G. (1971). «The theory of regionalized variables and its applications». *Informe Técnico 5*, Cahiers du Centre de Morphologie Mathématique, Fontainebleau.
- MCBRATNEY, A. B.; MENDONÇA, M. L. y MINASNY, B. (2003). «On digital soil mapping». *Geoderma*, **117(1-2)**, pp. 3–52. ISSN 0016-7061.

- MCBRIDE, R.A. y MACINTOSH, E.E. (1984). «Soil survey interpretations from water retention data: 1. Development and validation of a water retention model». *Soil Science Society of America Journal*, **48**, p. 1338–1343.
- MCDONALD, J.H. (2014). *Handbook of Biological Statistics*. Sparky House Publishing, Baltimore, MY, 3ª edición.
- MCGRATH, D. y ZHANG, C. (2003). «Spatial distribution of soil organic carbon concentrations in grassland of Ireland». *Applied Geochemistry*, **18(10)**, pp. 1629–1639.
- MCKENZIE, N.J. y RYAN, P.J. (1999). «Spatial prediction of soil properties using environmental correlation». *Geoderma*, **89**, pp. 67–94.
- MCLEOD, A.I. y XU, CHANGJIANG (2011). *bestglm: Best Subset GLM*. R package version 0.33. Último acceso: 10/10/2013.
<http://CRAN.R-project.org/package=bestglm>
- MELTON, M.A. (1965). «The geomorphic and paleoclimatic significance of alluvial deposits in southern Arizona». *Journal of Geology*, **73**, pp. 1–38.
- MERDUN, H.; CINARB, O.; MERALA, R. y APANC, M. (2006). «Comparison of artificial neural network and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity». *Soil and Tillage Research*, **90(1-2)**, pp. 108–116.
- MEYER, D.; LEISCH, F. y HORNIK, K. (2003). «The support vector machine under test.» *Neurocomputing*, **55(1-2)**, pp. 169–186.
- MILBORROW, S. (2011). *earth: Multivariate Adaptive Regression Spline Models*. R package version 3.2-1. Derived from mda:mars by Trevor Hastie and Rob Tibshirani. Último acceso: 10/10/2014.
<http://CRAN.R-project.org/package=earth>
- MINASNY, B. (2009). «Prediction of the water content at field capacity from disturbed soil samples». Último acceso: 14/3/2015.
https://www.academia.edu/947138/Prediction_of_the_water_content_at_field_capacity_from_disturbed_soil_samples
- MINASNY, B. y MCBRATNEY, A.B. (2002). «The Neuro-m method for fitting neural network parametric pedotransfer functions». *Soil Sci. Soc. Am. J.*, **66**, pp. 352–361.
- MINASNY, B.; MCBRATNEY, A.B.; MALONE, B.P. y WHEELER, I. (2013). «Digital Mapping of Soil Carbon». *Advances in Agronomy*, **118(3,4)**, pp. 1–47.
- MITAS, L. y MITASOVA, H. (1999). *Geographical Information Systems: Principles, Techniques, Management and Applications*. Spatial Interpolation, pp. 481–492. GeoInformation International. Wiley.

- MONGER, H. C. y GALLEGOS, R. A. (2000). *Global climate change and pedogenic carbonates*. Biotic and abiotic processes and rates of pedogenic carbonate accumulation in the southwestern United States. Relationship to atmospheric CO_2 sequestration, pp. 273–290. CRC/Lewis Publishers.
- MONTGOMERY, D.C. y PECK, E.A. (1992). *Introduction to Linear Regression Analysis*. Wiley, New York.
- MOORE, I.D. y BURCH, G.J. (1986). «Modelling erosion and deposition: Topographic effects». *Transactions of the American Society of Agricultural Engineers*, **29**, pp. 1624–1630.
- MOORE, I.D.; GESSELER, P.E.; NIELSEN, G.A. y PETERSON, G.A. (1993). «Soil attribute prediction using terrain analysis.» *Journal of Soil Science of America*, **57**, pp. 443–452.
- MORA, J.L. y LÁZARO, R. (2014). «Seasonal changes in bulk density under semiarid patchy vegetation: the soil beats». *Geoderma*, **235–236**, pp. 30–38.
- MORAN, C.J. y BUI, E.N. (2002). «Spatial data mining for enhancing soil map modeling». *International Journal of Geographic Information Science*, **16**, pp. 533–549.
- NADEU PUIG-PEY, E. (2013). *Soil erosion and organic carbon mobilization at the catchment scale: factors, processes and impact on the carbon balance*. Tesis doctoral, University of Murcia.
- NEMES, A. y RAWLS, W.J. (2004). «Soil texture and particle-size distribution as input to estimate soil hydraulic properties». En: Y.A. Pachepsky y W.J. Rawls (Eds.), *Development of Pedotransfer Functions in Soil Hydrology*, pp. 47–70. Elsevier.
- NEMES, A.; SCHAAP, M. y WOSTEN, H. (2002). «Validation of international scale soil hydraulic pedotransfer functions for national scale applications». En: *Proceedings of the 17th World Congress of Soil Science*, pp. 47–70. Thailand.
- NETELER, M. y MITASOVA, H. (2007). *Open Source GIS: A GRASS GIS Approach*. Springer, New York, 3 edición.
- NICHOLS, J.D. (1984). «Relation of Organic Carbon to Soil Properties and Climate in the Southern Great Plains». *Soil Sci. Soc. Am. J.*, **48**, pp. 1382–1384.
- NIEDER, R. y BENBI, D.K. (2008). *Carbon and Nitrogen in the Terrestrial Environment*. Springer, Netherlands.
- ODEH, I.O.A.; MCBRATNEY, A.B. y CHITTLEBOROUGH, D.J. (1995). «Further results on prediction of soil properties from terrain attributes: Heterotopic cokriging and regression-kriging». *Geoderma*, **67**, pp. 215–226.
- OLAYA, V. y CONRAD, O. (2009). «Chapter 12 Geomorphometry in SAGA». En: Tomislav Hengl y Hannes I. Reuter (Eds.), *Geomorphometry - Concepts, Software, Applications*, volumen 33 de *Developments in Soil Science*, pp. 293–308. Elsevier.

- OLIVEIRA, L.B.; RIBEIRO, M.R.; JACOMINE, P.K. y RODRIGUES, J.V.V. (2002). «Funções de Pedotransferência para predição da umidade retida a potenciais específicos em solos do Estado de Pernambuco.» *Revista Brasileira de Ciência do Solo*, **26**, pp. 315–323.
- OMRAN, E.E. (2012). «Improving the Prediction Accuracy of Soil Mapping through Geostatistics». *International Journal of Geoscience*, **3**, pp. 574–599.
- PACHEPSKY, Y. y RAWLS, W.J. (2004). *Development of Pedotransfer Functions in Soil Hydrology*. Elsevier.
- PACHEPSKY, Y. A.; TIMLIN, D. J. y RAWLS, W. J. (2001). «Soil Water Retention as Related to Topographic Variables». *Soil Sc. Soc. Am. J.*, **65**, pp. 1787–1795.
- PAPADAKIS, J. (1966). *Climates of the world and their agricultural potentialities*. Ed. Albatros, Buenos Aires.
- PARCHAMI-ARAGHI, F.; MIRLATIFI, S. M.; DASHTAKI, S.G. y MAHDIAN, M. H. (2013). «Point estimation of soil water infiltration process using Artificial Neural Networks for some calcareous soils». *Journal of Hydrology*, **481**, pp. 35–47.
- PEECH, M. (1965). *Methods for Soil Analysis*. Exchange acidity, pp. 905–913. U.S. Government Printing Office.
- PIDGEON, J.D. (1972). «The measurement and prediction of available water capacity of Ferrallitic soils in Uganda». *Journal Soil Science*, **23**, pp. 431–441.
- PIRES, L.F.; COOPER, M.; CÁSSARO, F.A.M.; REICHARDT, K.; BACCHI, O.O.S. y DIAS, N.M.P. (2008). «Micromorphological analysis to characterize structure modifications of soil samples submitted to wetting and drying cycles». *Catena*, **72**, pp. 297–304.
- POGGIO, L.; GIMONA, A.; BROWN, I. y CASTELLAZZI, M. (2010). «Soil available water capacity interpolation and spatial uncertainty modelling at multiple geographical extents». *Geoderma*, **160**, pp. 175–188.
- PRASAD, A.M.; IVERSON, L.R. y LIAW, A. (2006). «Newer classification and regression tree techniques: bagging and random forests for ecological prediction». *Ecosystems*, **9**, pp. 181–199.
- PÉREZ CUTILLAS, P. (2013). *Modelización de propiedades físicas del suelo a escala regional. Casos de estudio en el Sureste Ibérico*. Tesis doctoral, CEBAS-CSIC.
- QUINN, P.; BEVEN, K.; CHEVALIER, P. y PLANCHON, O. (1991). «The prediction of hillslope flow paths for distributed hydrological modeling using digital terrain models». *Hydrological Processes*, **5(1)**, pp. 59–79.
- RAICH, J.W. y POTTER, C. S. (1995). «Global patterns of carbon dioxide emissions from soils». *Global Biogeochemical Cycles*, **9**, pp. 23–36.

- RAMÍREZ-SANTIAGOSA, I.; VICENTE-ALBADALEJO, M.; GARCÍA-BARCELÓ, J.A. y VAQUERO GÓMEZ, A. (1999). «Mapa Digital de Suelos de la Región de Murcia».
- RAWLS, W.J. (1983). «Estimating soil bulk density from particle size analysis and organic matter content». *Journal of Soil Science*, **135**, pp. 123–125.
- RAWLS, W.J.; BRAKENSIEK, D.L. y SAXTON, K.E. (1982). «Estimation of soil water properties». *Transactions of the ASAE*, **25(5)**, pp. 1316–1320.
- RAWLS, W.J.; GISH, T.J. y BRACKENSIEK, D.L. (1991). «Estimating soil water retention from soil physical properties and characteristics.» *Advances in Soil Science*, **16**, pp. 213–234.
- RAWLS, W.J.; PACHEPSKY, Y.A.; RITCHIE, J.C.; SOBECKI, T.M. y BLOODWORTH, H. (2003). «Effect of soil organic carbon on soil water retention». *Geoderma*, **116(1–2)**, pp. 61–76.
- REGALADO, C.M. y RITTER, A. (2005). «Técnicas geoestadísticas para caracterizar la variabilidad espacial de la hidrofobicidad en la zona no saturada del suelo». En: F.J. Samper Calvete y A.Paz González (Eds.), *Estudios de la Zona No Saturada del suelo*, volumen VII, pp. 95–114. Amsterdam.
- REZAEI, S.A. y GILKES, R.J. (2005). «The effects of landscape attributes and plant community on soil chemical properties in rangelands». *Geoderma*, **125**, pp. 167–176.
- RHOADES, J. D.; MANTEGHI, N. A.; SHOUSE, P. J. y ALVES, W. J. (1989). «Estimating Soil Salinity from Saturated Soil-Paste Electrical Conductivity». *Soil Science Society of America Journal*, **53**, pp. 428–433.
- RICHARDS, L.A. y WEAVER, L.R (1944). «Moisture Retention by some irrigated soils as related to soil-moisture tension». *Journal of Agricultural Research*, **69**, pp. 215–235.
- RILEY, S.J.; DE GLORIA, S.D. y ELLIOT, R. (1999). «A terrain ruggedness index that quantifies topographic heterogeneity». *Intermountain Journal of Sciences*, **5**, pp. 123–127.
- ROBINSON, P.T. y METTERNICHT, G. (2006). «Testing the performance of spatial interpolation techniques for mapping soil properties». *Computers and Electronics in Agriculture*, **50**, pp. 97–108.
- RODRÍGUEZ-MURILLO, J.C. (2001). «Organic carbon content under different types of land use and soil in peninsular Spain». *Biol. Fertil. Soils*, **33**, pp. 53–61.
- ROECKER, S.M; HOWELL, D.W.; C.A., HAYDU-HOUESHELL y C., BLINN (2010). *A Qualitative Comparison of Conventional Soil Survey and Digital Soil Mapping Approaches*. volumen 2, 29, pp. 369–384. Springer Science.
- RUEHLMANN, J. y KÖRSCHENS, M. (2009). «Calculating the effect of soil organic matter concentration on soil bulk density». *Soil Science Society of America Journal*, **73(3)**, pp. 876–885.
- SALTER, P.J. y WILLIAMS, J.B. (1965). «The influence of texture on the moisture characteristics of soils I. A critical comparison for determining the available water capacity and moisture characteristics curve of a soil». *Journal of Soil Science*, **16**, pp. 1–15.

- SÁNCHEZ NAVARRO, A. (1995). *Suelos de la Sierra de Mojantes y sus alrededores*(Caravaca, Murcia). Tesis doctoral, Universidad de Murcia.
- SAPPINGTON, J.M.; LONGSHORE, K.M. y THOMPSON, D.B (2007). «Quantifying landscape ruggedness for animal habitat analysis: A case study using bighorn sheep in the Mojave Desert». *Journal of Wildlife Management*, **71**(5), pp. 1419–1426.
- SCHAAP, M.G. y BOULTEN, W. (1996). «Modelling water retention curves of sandy soils using neural networks.» *Water Resources Research*, **32**, pp. 3033–3040.
- SCHAAP, M.G. y LEIJ, F.J. (1998a). «Database-related accuracy and uncertainty of pedotransfer functions». *Soil Science*, **163**, pp. 765–777.
- SCHAAP, M.G. y LEIJ, F.J. (1998b). «Using neural networks to predict soil water retention and soil hydraulic conductivity». *oil and Tillage Research*, **47**, pp. 37–42.
- SCHAPIRE, R. (2003). «MSRI Workshop on Nonlinear Estimation and Classification, 2002».
- SCHLESINGER, W. H (1997). *Biogeochemistry: an analysis of global change*. Academic Press, New York.
- SCULL, P.; FRANKLIN, J. y CHADWICK, O.A. (2005). «The application of classification tree analysis to soil type prediction in a desert landscape». *Ecological Modelling*, **181**(1), pp. 1–15.
- SEQUEIRA, C.H.; WILLS, S.A.; SEYBOLD, C.A. y WEST, L. T. (2014). «Predicting soil bulk density for incomplete databases». *Geoderma*, **213**, pp. 64–73.
- SHE, D.; XUEMEI, G.; JINGRU, S.; TIMM, L. C. y HU, W. (2014). «Soil organic carbon estimation with topographic properties in artificial grassland using a state-space modeling approach». *Canadian Journal of Soil Sciences*, **94**, pp. 503–514.
- SHEPHERD, K.D y WALSH, M.G. (2002). «Development of Reflectance Spectral Libraries for Characterization of Soil Properties». *Soil Science Society of America Journal*, **66**, pp. 988–998.
- SHMUELI, G. (2010). «To Explain or to Predict?» *Statistical Science*, **25**, pp. 289–310.
- SIGUA, G.C. y HUNDALL, W.H.(2008). «Kriging analysis of soil properties. Implication to landscape management and productivity improvement». *Journal of Soils and Sediments*, **8**, pp. 193–202. ISSN 0016-7061.
- SIX, J.; CALLEWAERT, P.; LENDERS, S.; DE GRYZE, S.; MORRIS, S.J.; GREGORICH, E.G.; PAUL, E.A. y PAUSTIAN, K. (2002). «Measuring and understanding carbon storage in afforested soils by physical fractionation». *Soil Science Society of America Journal*, **66**, pp. 1981–1987.
- SMOLA, A.J. y SCHÖLKOPF, B. (2004). «A tutorial on support vector regression». *Statistics and Computing*, (14), pp. 199–222.

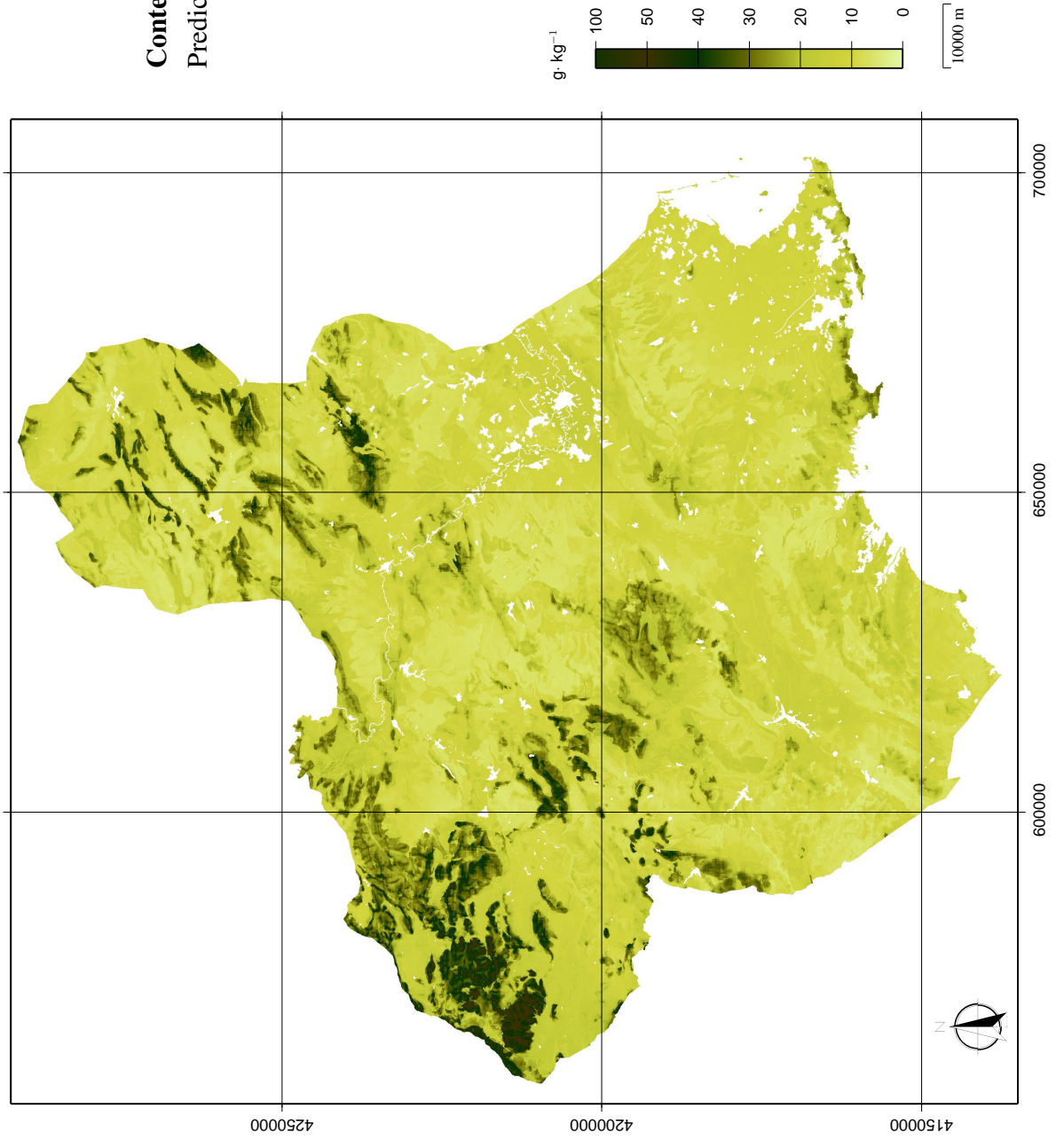
- SOIL SURVEY STAFF (2010). *Keys to Soil Taxonomy*. USDA National Resources Conservation Services, Washington, DC, 11^a edición.
- STAVI, I.; UNGAR, E.D.; LAVEE, H. y PARIENTE, S. (2008). «Grazing-induced spatial variability of soil bulk density and content of moisture, organic carbon and calcium carbonate in a semi-arid rangeland». *Catena*, **75(32)**, pp. 288–296.
- STUM, A.K. (2010). *Random Forests applied as a Soil spatial predictive model in arid Utah*. All graduate theses and dissertations. paper 736., Utah State University.
- SUMFLETH, K. y DUTTMANN, R. (2008). «Prediction of soil property distribution in paddy soil landscapes using terrain data and satellite information as indicators». *Ecological Indicators*, **8(5)**, pp. 485–501.
- THOMPSON, J. A. y KOLKA, R. K. (2005). «Soil Carbon Storage Estimation in a Forested Watershed using Quantitative Soil-Landscape Modeling». *Soil Sci Soc Am J*, **69**, pp. 1086–1093.
- THOMPSON, J.A.; PENA-YEWTUKHIW, E. M. y GROVE, J.H. (2006). «Soil-landscape modeling across a physiographic region: Topographic patterns and model transportability». *Geoderma*, **133(1–2)**, pp. 57–70.
- THWAITES, R.N. (2006). «Chapter 20 Conceptual and Digital Soil-Landscape Mapping using Regolith-Catenary Units». En: A.B. McBratney P. Lagacherie y M. Voltz (Eds.), *Digital Soil Mapping - An Introductory Perspective*, volumen 31 de *Developments in Soil Science*, pp. 257–268, 614. Elsevier. doi: DOI:10.1016/S0166-2481(06)31020-3.
- TIMM, L.C.; PIRES, L. F.; ROVERATTI, R.; ARTHUR, R. C. A.; REICHARDT, K.; DE OLIVEIRA, J.C.M. y BACCHI, O. O. S. (2006). «Field spatial and temporal patterns of soil water content and bulk density changes». *Scientia Agricola*, **63**, pp. 55 – 64.
- TOMASELLA, J.; HODNETT, M.G. y ROSSATO, L. (2000). «Pedotransfer functions for the estimation of soil water retention in Brazilian soils». *Soil Sci. Soc. Am. J.*, **64**, pp. 327–338.
- TORGO, L. (1999). *Inductive Learning of Tree-Based Regression Models*. Tesis doctoral, Faculdade de Ciencia da Universidade do Porto..
- TÓTH, B.; MAKÓ, A.; GUADAGNINI, L.; AZZELLINO, A. y GUADAGNINI, A. (2008). «Grouping of soils according to their soil water retention characteristics». En: *Calibration and Reliability in Groundwater Modelling: Credibility of Modelling. Proceedings of ModelCARE 2007*, volumen 320, pp. 154–160. IAHS Publications.
- TRUMBORE, S.E.; CHADWICK, O.A. y AMUNDSON, R. (1996). «Rapid exchange between soil carbon and atmospheric carbon dioxide driven by temperature change». *Science*, **272**, pp. 393–396.
- URACH, F.L. (2007). *Estimativa da retenção de água em solos para fins de irrigação*. Dissertação, Universidade Federal de Santa Maria.

- VAPNIK, V. (1995). *The Nature of Statistical Learning Theory*. Springer, New York.
- VAYSSE, K. y LAGACHERIE, P. (2015). «Evaluating Digital Soil Mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France)». *Geoderma Regional*, **4**, pp. 20–30.
- VEIHMEYER, F. J. y HENDRICKSON, A. H. (1948). «Soil Density and root penetration». *Soil Sci.*, **65**, pp. 487–493.
- VISCARRA ROSSEL, R.A. y BEHRENS, T. (2010). «Using data mining to model and interpret soil diffuse reflectance spectra». *Geoderma*, **158(1–2)**, pp. 46–54.
- WEBSTER, R. y BURGESS, T.M. (1980). «Optimal interpolation and isarithmic mapping, III. Changing drift and universal kriging». *Journal of Soil Science*, **31**, pp. 505–524.
- WEBSTER, R. y OLIVER, M.A. (1990). *Statistical methods in soil and land resource survey*. Oxford University Press, New York.
- WEBSTER, R. y OLIVER, M.A. (2001). *Geostatistics for Environmental Scientists*. John Wiley and Sons Ltd., London, UK.
- WIESMEIER, M.; BERTHOLD, F.; BLANK, B. y KÖGEL-KNABNER, I. (2011). «Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem». *Plant Soil*, **340**, pp. 7–24.
- WIESMEIER, M.; BERTHOLD, F.; SPÖRLEIN, P.; GEUSS, U.; HANGEN, E.; REISCHL, A.; SCHILLING, B.; ANGST, G.; VON LÜTZOW, M. y KÖGEL-KNABNER, I. (2014). «Estimation of total organic carbon storage and its driving factors in soils of Bavaria (southeast Germany)». *Geoderma Regional*, **1**, pp. 67–78.
- WILCKE, W.; YASIN, S.; SCHMITT, A.; VALAREZO, C. y ZECH, W. (2008). *Ecological Studies 198, Gradients in a Tropical Mountain Ecosystem of Ecuador*. 9. Soils along the altitudinal transect and in catchments. Springer, Berlin, Heidelberg.
- WILDING, L.P.; NOBLES, M.M.; WILCOX, B.P.; WOODRUFF, C.M y LIN, H. (2012). *Hydrogeology*. 9. Hydrogeology in Caliche Soils Weathered from Glen Rose Limestone of Lower Cretaceous Age in Texas, pp. 285–327. Elsevier.
- WOOD, S.N. (2006). *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC.
- WÖSTEN, J.H.M.; FINKE, P.A. y JANSEN, M.J.W. (1995). «Comparison of class and continuous pedotransfer functions to generate soil hydraulic characteristics». *Geoderma*, **66**, pp. 227–237.
- WÖSTEN, J.H.M.; PACHEPSKY, Y.A. y RAWLS, W.J. (2001). «Pedotransfer functions: bridging gap between available basic soil data and missing soil hydraulic characteristics.» *Journal of Hydrology*, **251**, pp. 123–150.

-
- WRB, IUSS WORKING GROUP (2007). «World reference base for soil resources 2006, first update 2007». *Informe Técnico 103*, FAO, Roma.
- WYNN, J.G.; BIRD, M.I.; VELLEN, L.; GRAND-CLEMENT, E.; CARTER, J. y BERRY, S.L. (2006). «Continental-scale measurement of the soil organic carbon pool with climatic, edaphic, and biotic controls». *Global Biogeochemical Cycles*, **20(1)**, p. GB1007. doi: 10.1029/2005GB002576. Último acceso: 24/11/2014.
http://research.eeescience.utoledo.edu/lees/papers_PDF/Wynn_2006_GBC.pdf
- XIONG, X.; S. GRUMWALD, D. B. MYERS; KIM, J.; HARRIS, W. G. y COMERFORD, N. B. (2014). «Holistic environmental soil-landscape modeling of soil organic carbon». *Environmental Modelling and Software*, **57**, pp. 202–215.
- YAALON, D.H. (1997). «Soils in the Mediterranean region: what makes them different?» *Catena*, **28**, pp. 157–199.
- YANG, C.; PRASHER, S.O.; LACROIX, R. y KIM, S.H. (2003). «A Multivariate Adaptive Regression Splines Model for Simulation of Pesticide Transport in Soils». *Biosystems Engineering*, **86(1)**, pp. 9–15.
- YIMER, F.; LEDIN, S. y ABDELKADIR, A. (2007). «Changes in soil organic carbon and total nitrogen contents in three adjacent land use types in the Bale Mountains, southeastern highlands of Ethiopia». *Forest Ecology and Management*, **242**, pp. 337–342.
- ZHANG, X.C. y NORTON, L.D (2002). «Effect of exchangeable Mg on saturated hydraulic conductivity, disaggregation and clay dispersion of disturbed soils». *J. Hydrol.*, **260**, pp. 194–205.
- ZHU, A.X.; LIU, F.; LI, B.L.; PEI, T.; QIN, C.Z.; LIU, G.H.; WANG, Y.J.; CHEN, Y.N.; MA, X.W.; QI, F. y ZHOU, C.H. (2010). «Differentiation of soil conditions over low relief areas using feedback dynamic patterns». *Soil Science Society of America Journal*, **74**, p. 861–869.
- ZUUR, A.F.; IENOL, E.N. y ELPHICK, C.S (2010). «A protocol for data exploration to avoid common statistical problems». *Methods in Ecology and Evolution*, **1**, pp. 3–14.

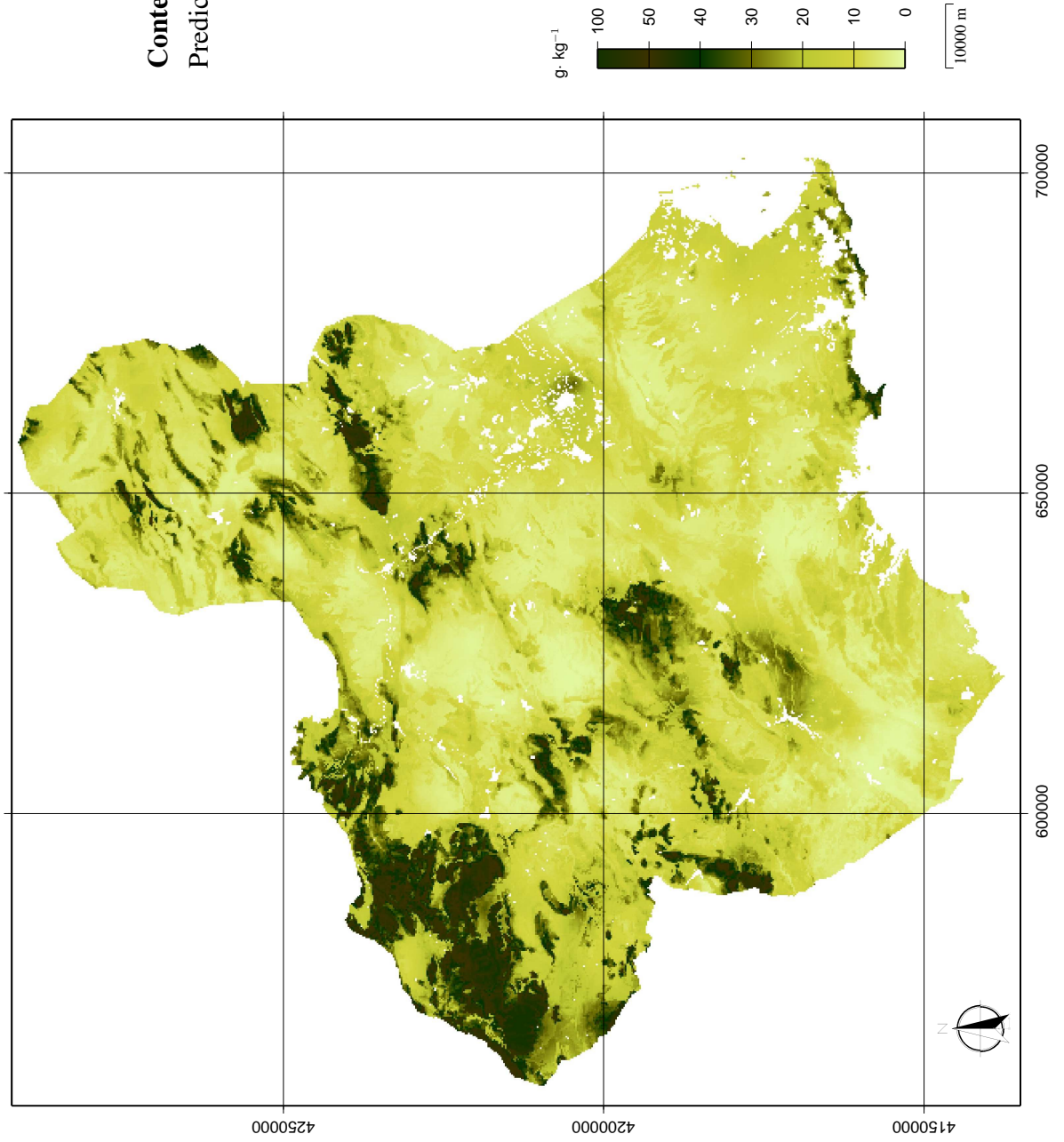
Cartografía

**Contenido de CO
Predicción**

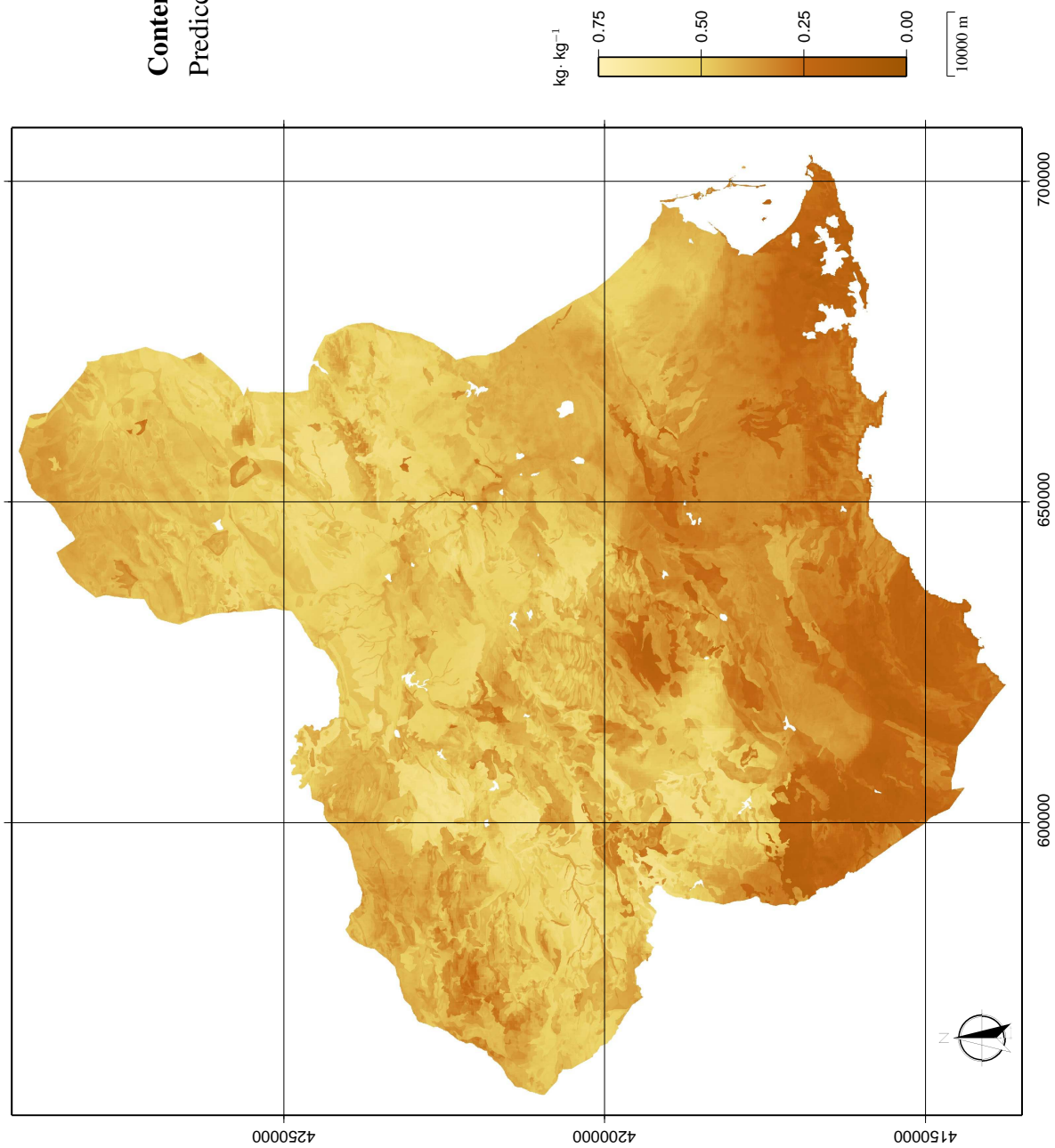


A.1

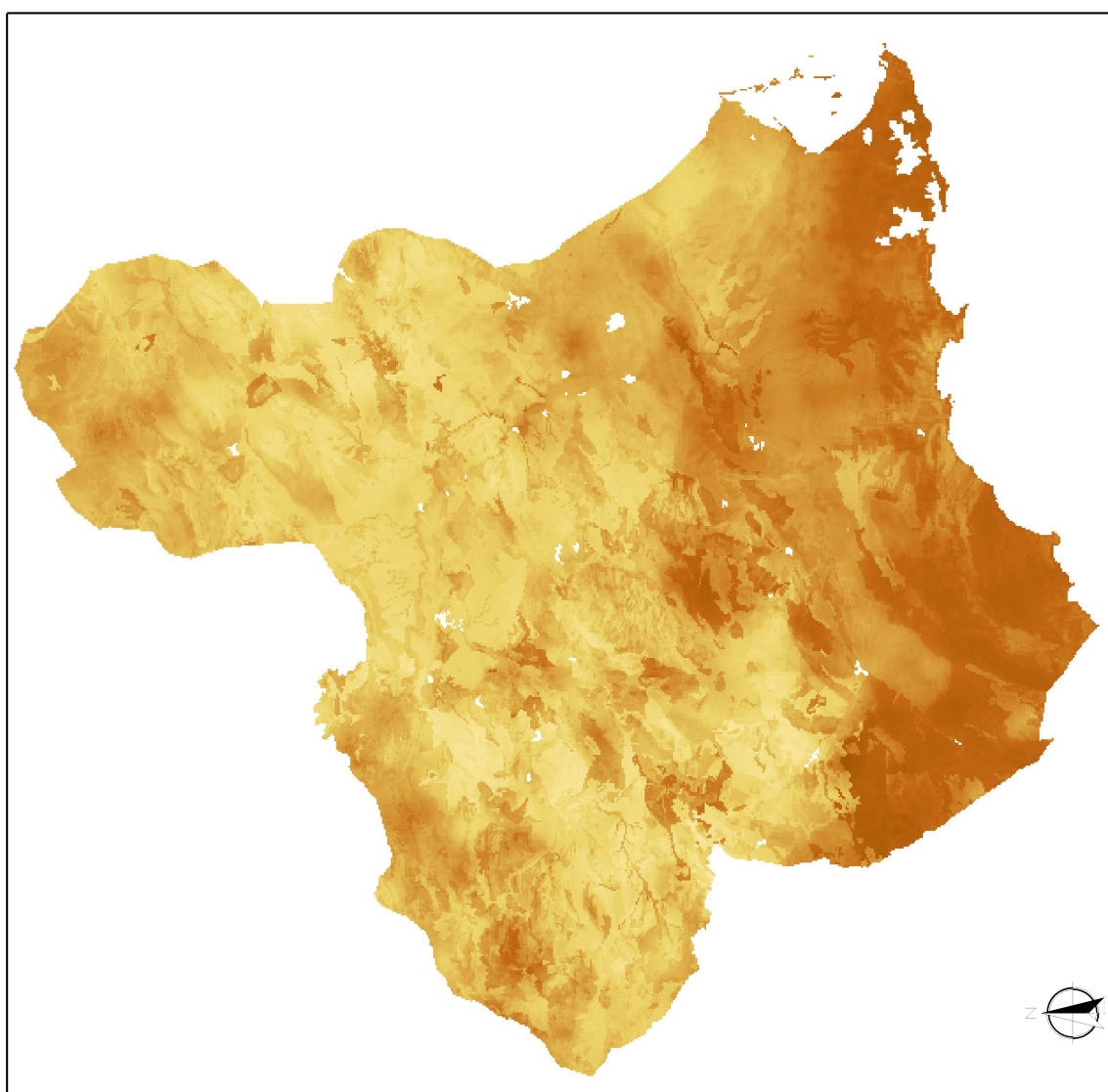
Contenido de CO
Predicción + KO



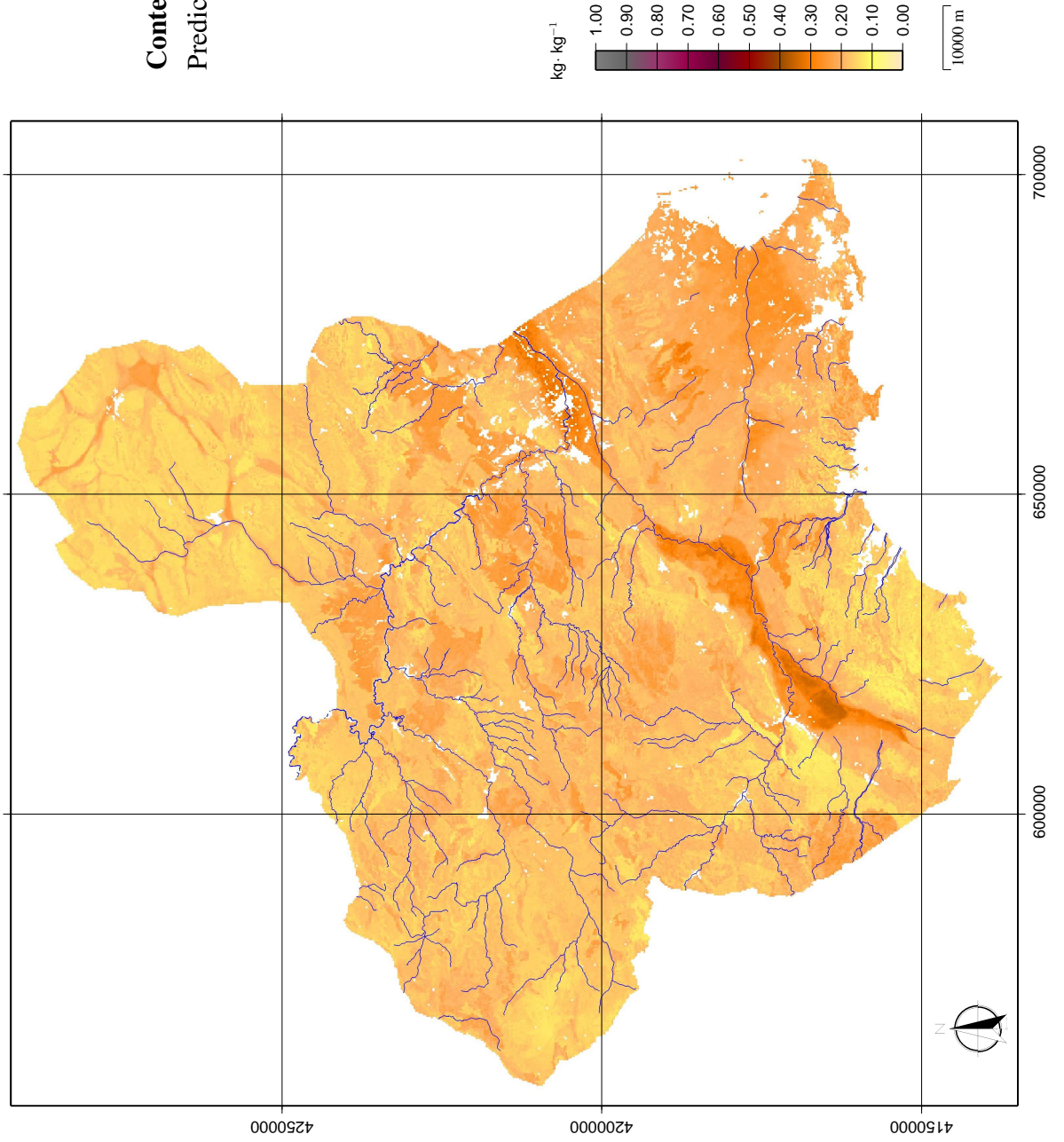
Contenido de CaCO_3 total o equivalente
Predicción



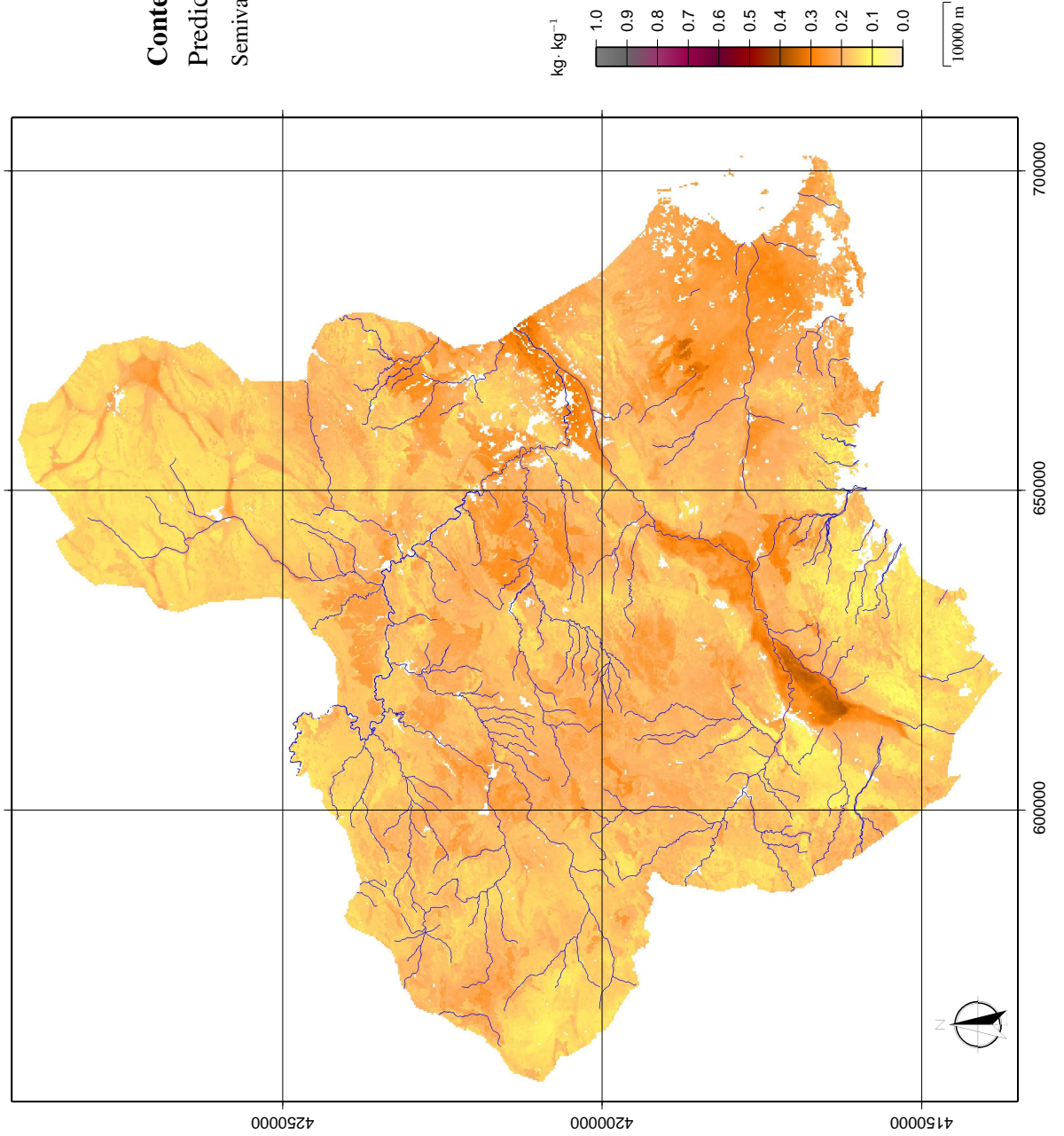
Contenido de CaCO_3 total o equivalente
Predicción + KO



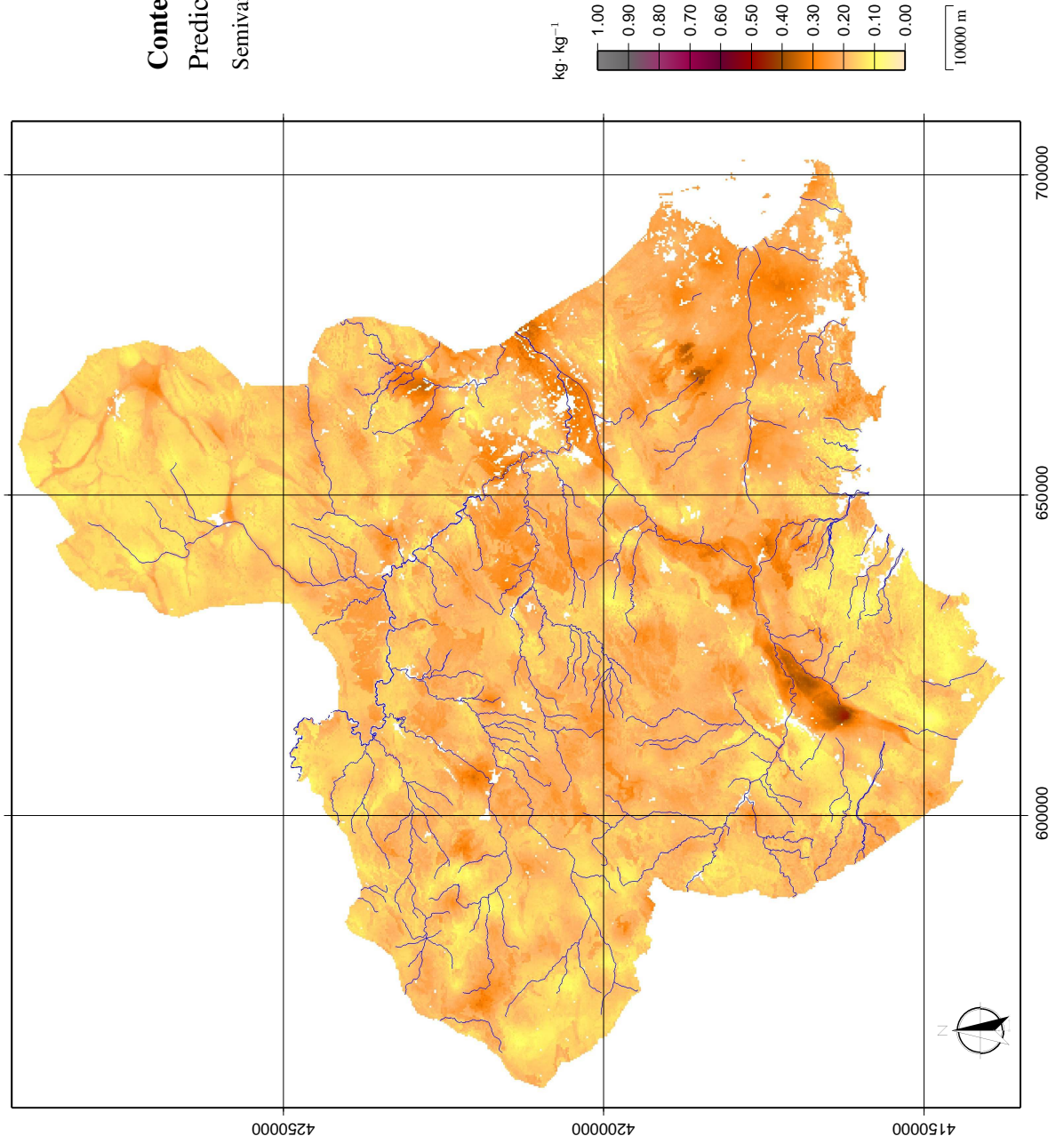
Contenido de arcilla
Predicción



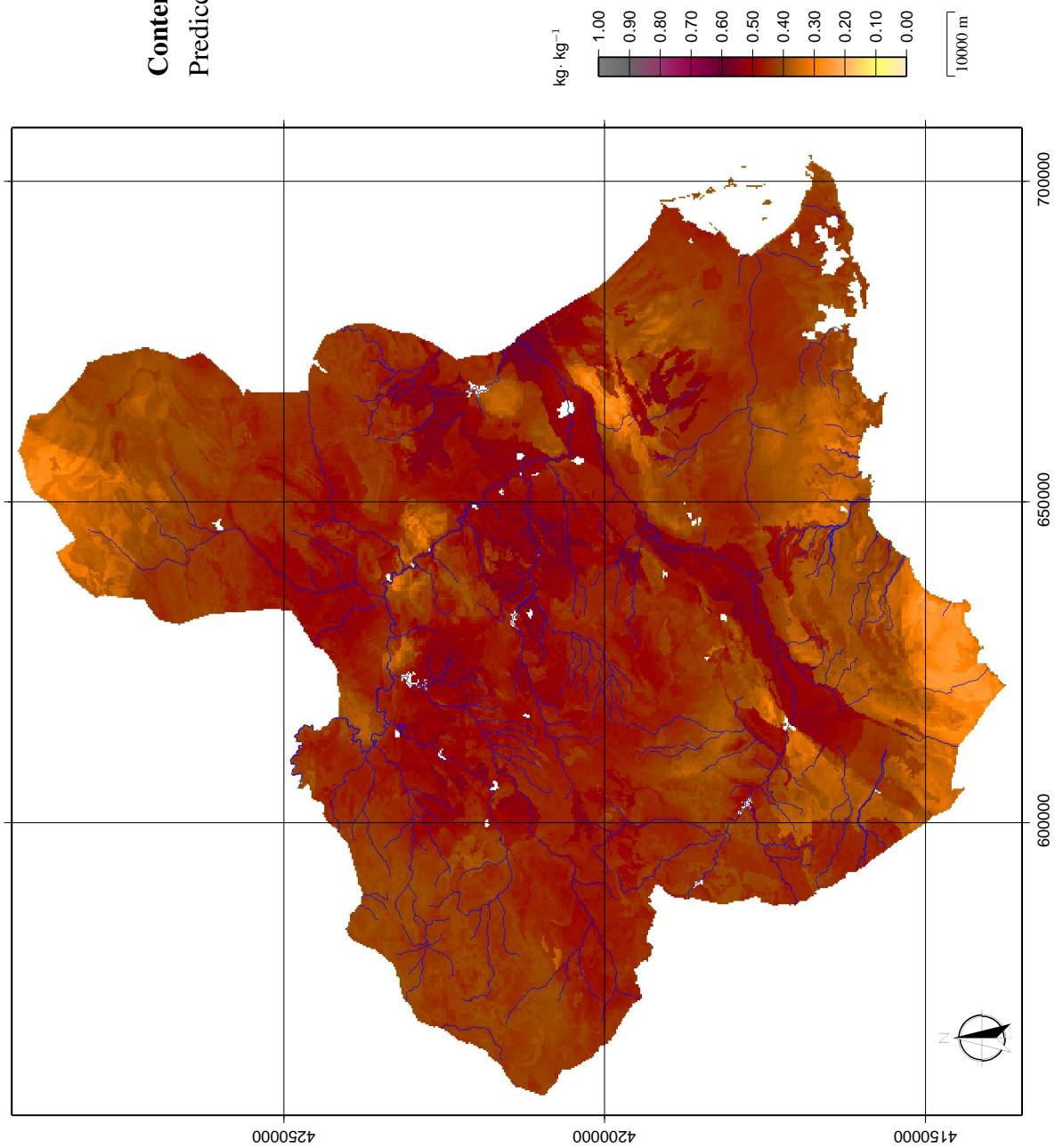
Contenido de arcilla
Predicción + KO
Semivariograma de Mättern



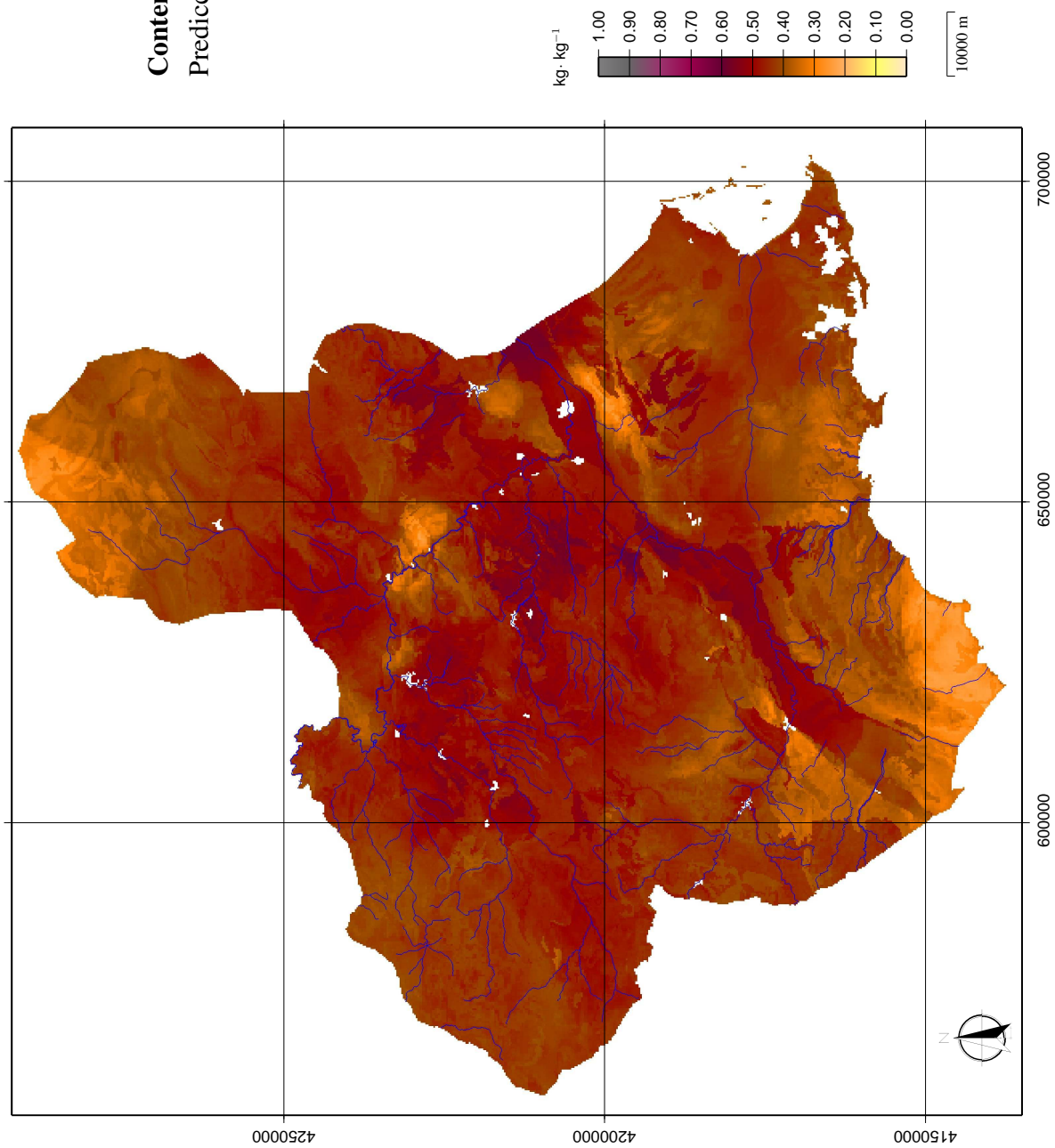
Contenido de arcilla
Predicción + KO
Semivariograma esférico



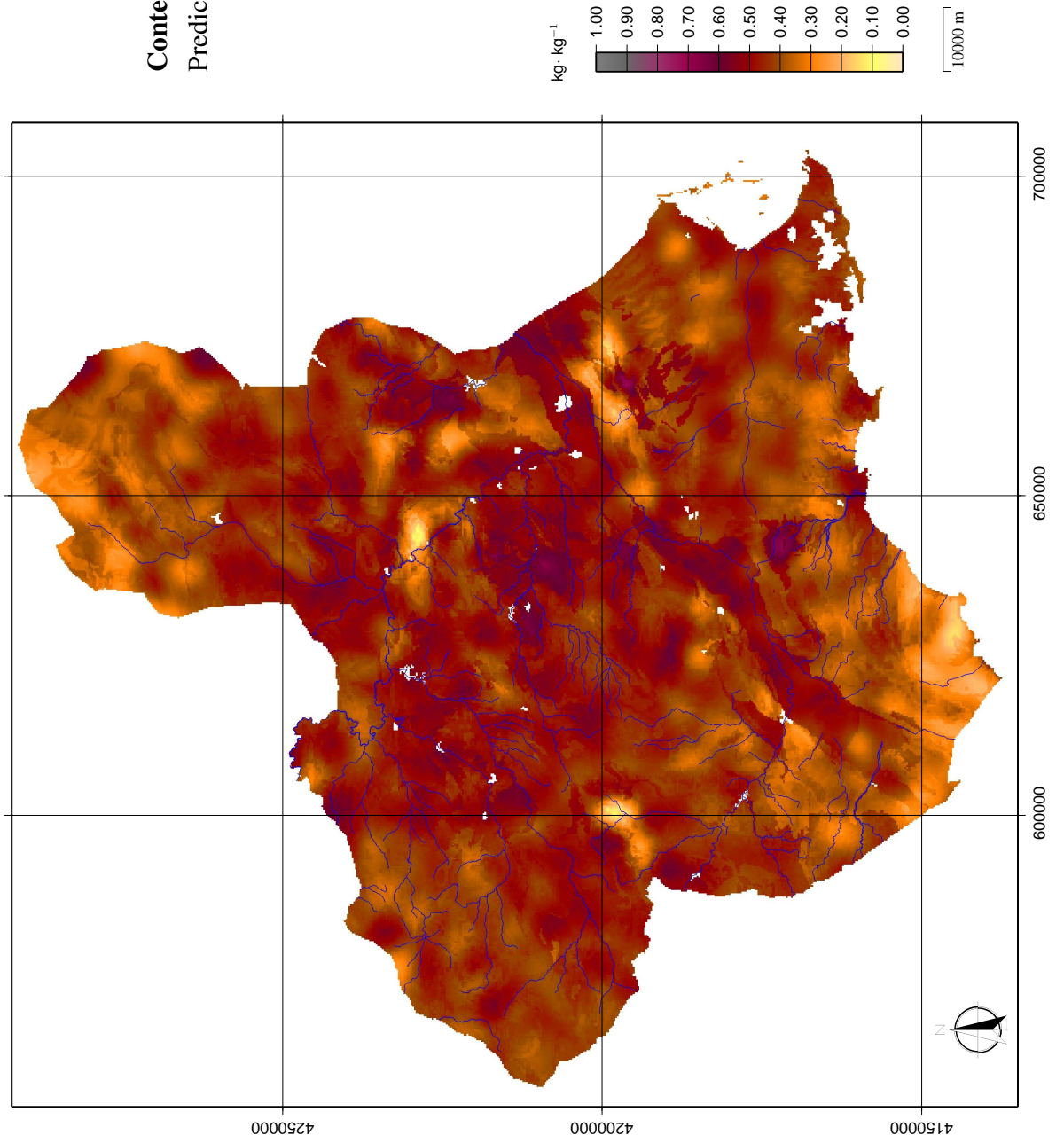
Contenido de limo
Predicción



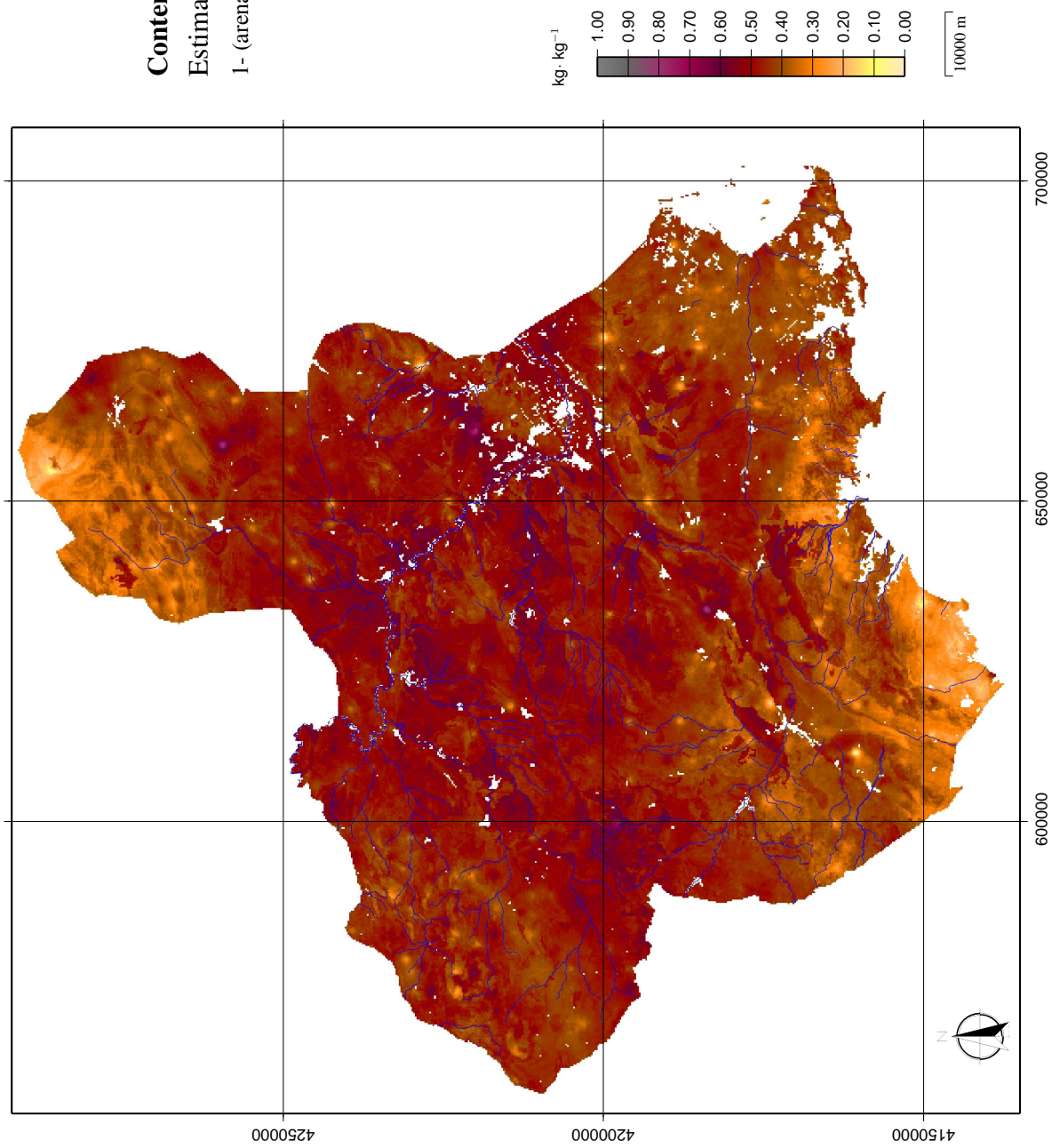
**Contenido de limo
Predicción + KO**



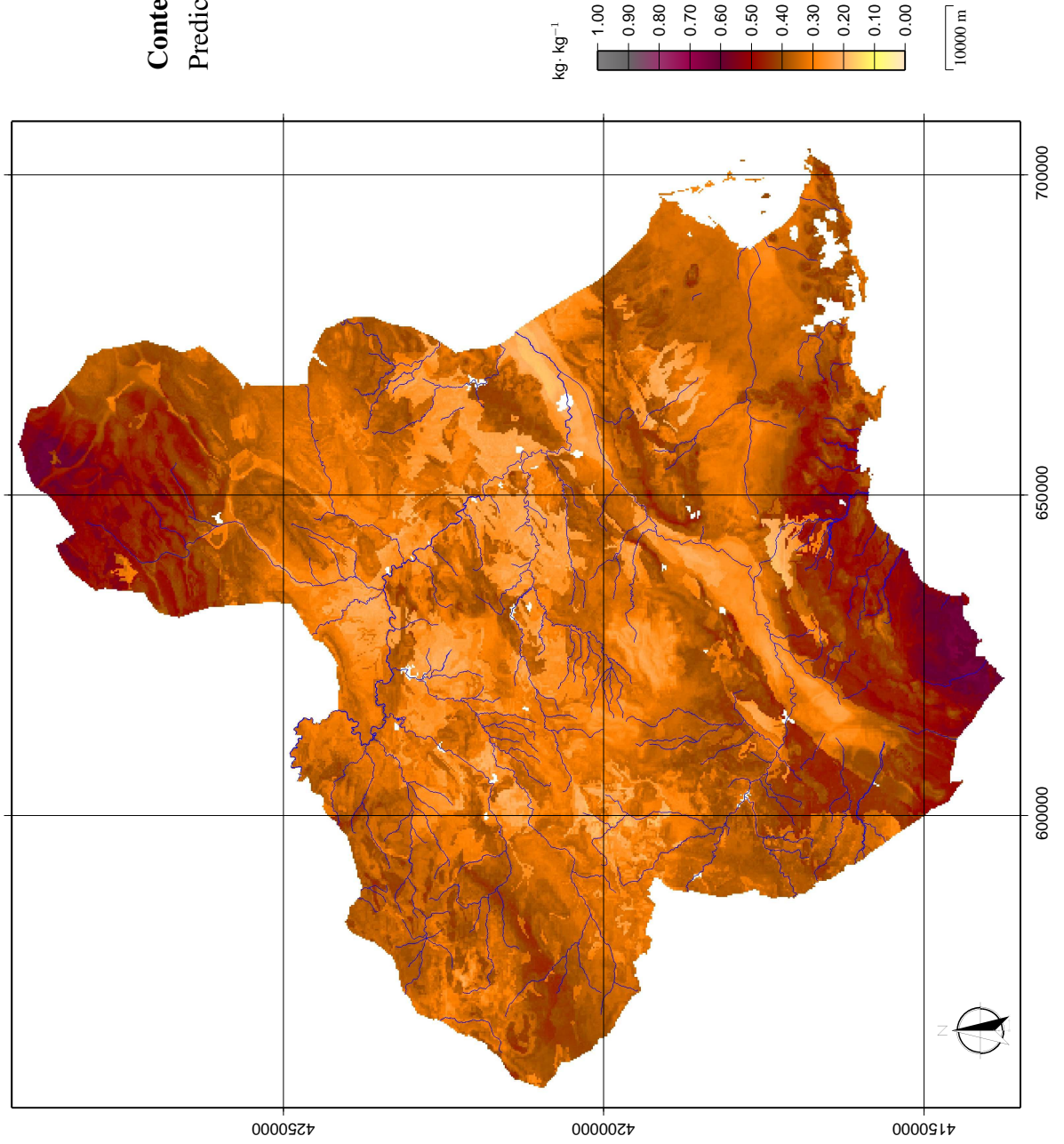
Contenido de limo
Predicción + splines



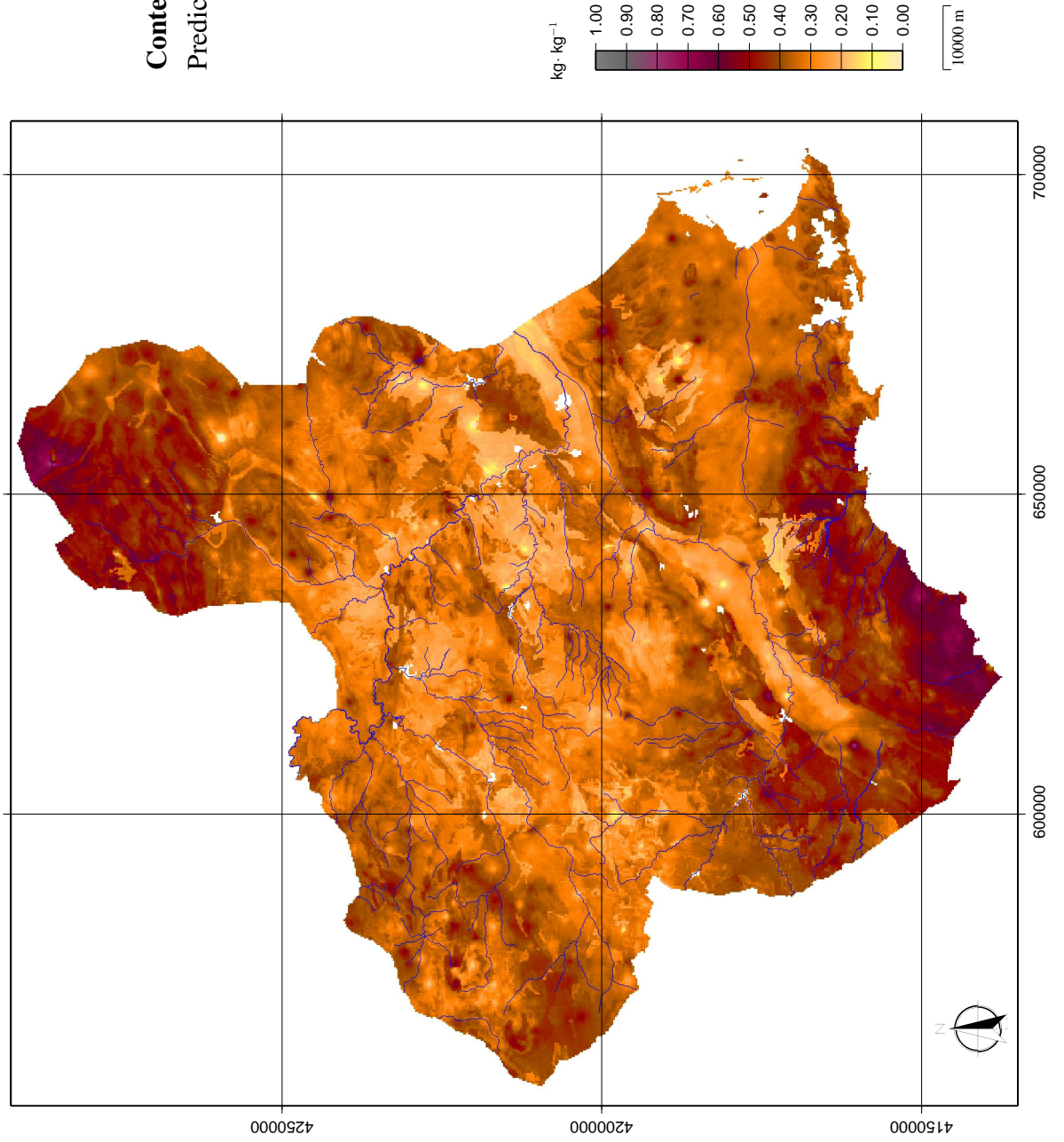
Contenido de limo
Estimado
1- (arena+arcilla)



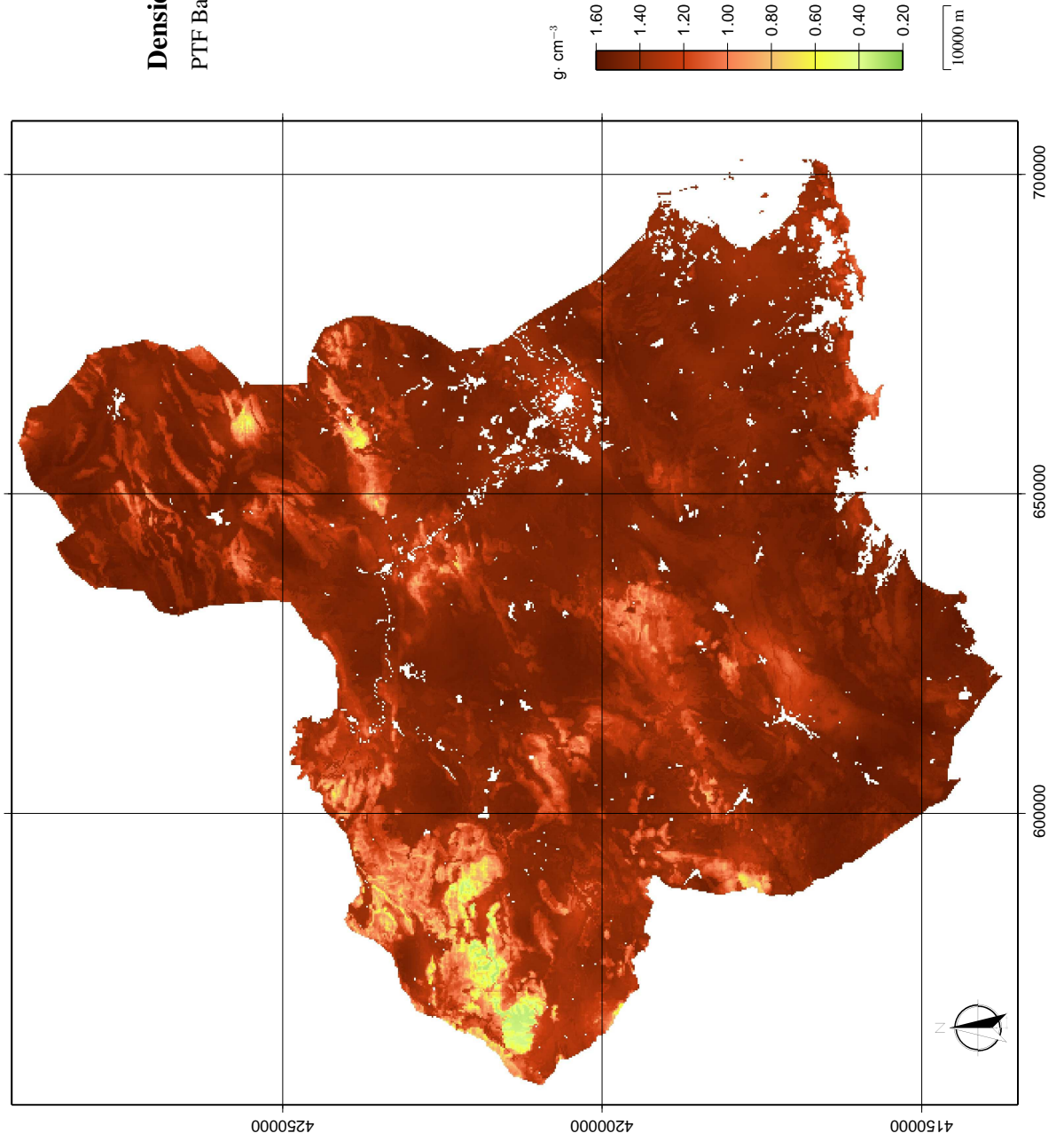
Contenido de arena
Predicción



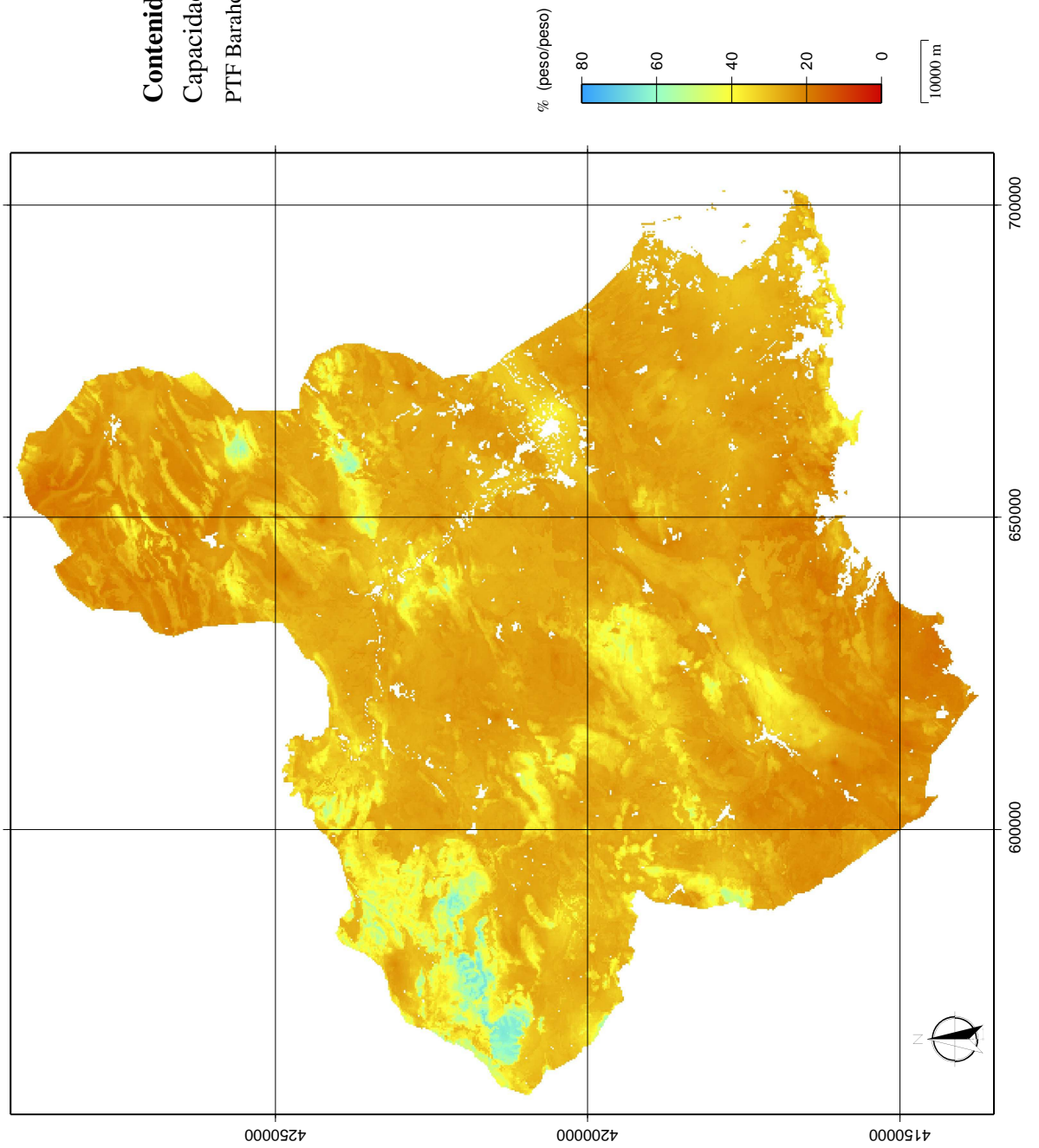
Contenido de arena
Predicción+ splines



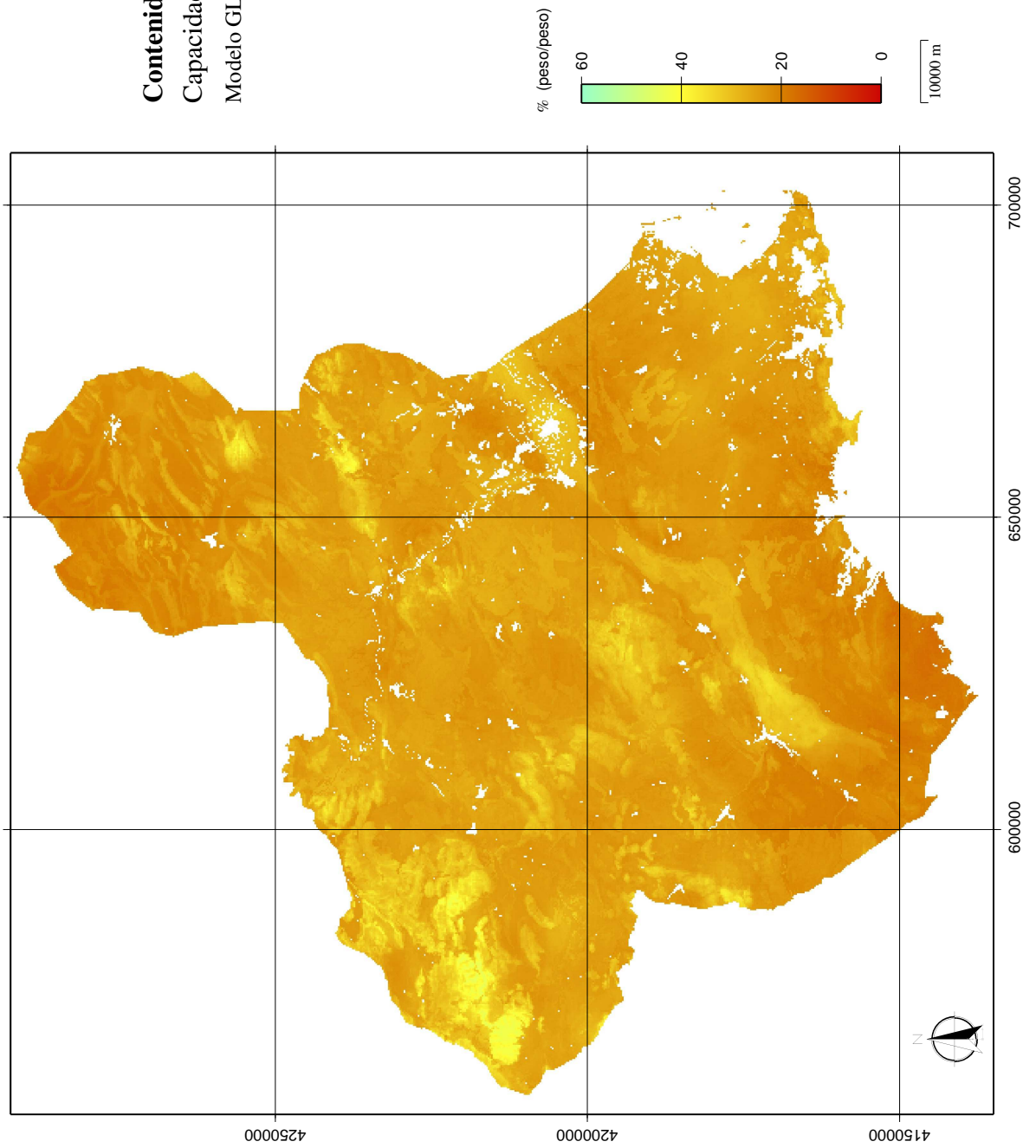
Densidad aparente
PTF Barahona y Santos (1981)



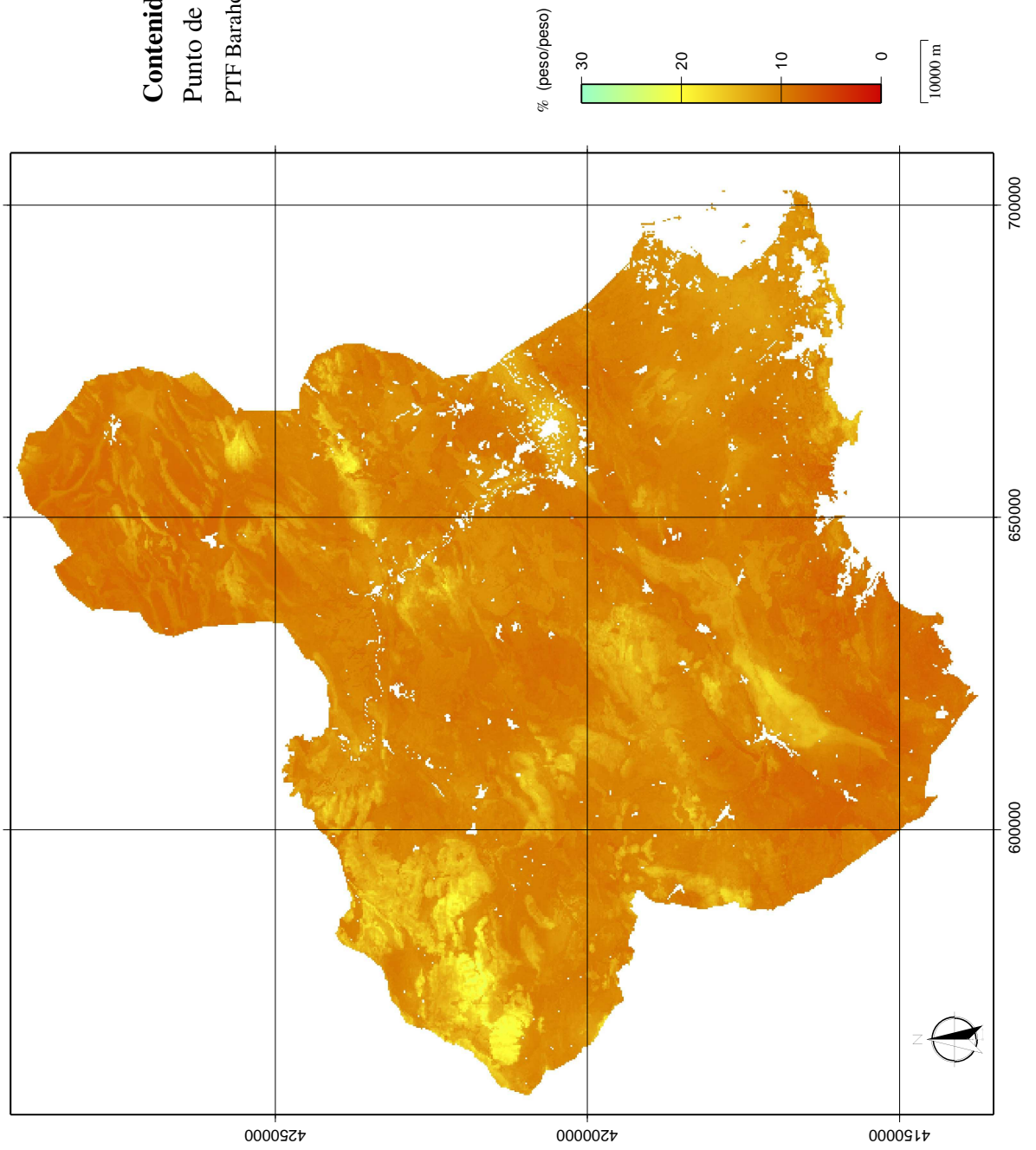
Contenido de humedad
Capacidad de campo (-0.33 atm)
PTF Barahona y Santos (1981)



Contenido de humedad
Capacidad de campo (-0.33 atm)
Modelo GLM-LOOCV



Contenido de humedad
Punto de marchitamiento (-15 atm)
PTF Barahona y Santos (1981)



Contenido de humedad
Punto de marchitamiento (-15 atm)
Modelo GLM-LOOCV

