



universidad
de león



FACULTAD DE CIENCIAS BIOLÓGICAS Y AMBIENTALES

**CARTOGRAFÍA DIGITAL DE PROPIEDADES
EDÁFICAS EN EL ÁREA DE BENAVIDES DE
ÓRBIGO (LEÓN)**

**DIGITAL SOIL MAPPING IN THE AREA OF
BENAVIDES DE ÓRBIGO (LEÓN)**

Autor: Judit Rodríguez Fernández

GRADO EN CIENCIAS AMBIENTALES

Julio, 2021

ÍNDICE

1. Introducción	1
2. Objetivos	3
3. Área de estudio	3
3.1.Geología, litología y geomorfología.....	4
3.2.Edafología.....	6
3.3.Climatología.....	7
3.4.Usos del suelo y vegetación.....	7
4. Materiales y métodos	8
4.1.Análisis de factores y obtención de covariables	10
4.1.1. Clima.....	11
4.1.2. Vegetación.....	11
4.1.3. Topografía.....	13
4.1.4. Material parental	15
4.2.Análisis estadístico	15
4.3.Análisis con SIG	17
5. Resultados	17
5.1.Análisis de linealidad y exploración de resultados. Análisis de regresión múltiple y GLM.....	17
5.2.Análisis geoestadístico.....	22
6. Discusión	25
7. Conclusiones	26
8. Referencias bibliográficas	27
ANEXO	31

Resumen

La escasez de información edafológica y cartografía ha desembocado en la búsqueda de nuevas técnicas que permitan solucionar esta problemática, de forma que se pueda obtener una cobertura completa y homogénea de la superficie. El objetivo principal de este trabajo es la elaboración de una cartografía de propiedades edáficas de interés como son materia orgánica, arena, limo y arcilla del suelo a través de la aplicación de la metodología *Digital Soil Mapping* en el área de Benavides de Órbigo (León). Para ello, se seleccionaron 75 muestreos de suelo ya existentes y distintas covariables ambientales relacionadas con los factores formadores del suelo. Con estas se llevó a cabo el análisis estadístico mediante la combinación de regresión lineal múltiple y modelos lineales generalizados para obtener el mejor modelo de predicción para cada variable estudiada, además de los residuos generados por el modelo y la estimación del error.

De todas las variables, el limo fue la única que presentó un valor de ajuste R^2 superior a 0,5, mientras que para la materia orgánica se obtuvo el error más bajo. Por todo ello, se llegó a la conclusión que, aunque el método es aplicable y útil, presenta grandes limitaciones, por lo que es importante incidir en la búsqueda de soluciones que permitan obtener mejoras en los resultados.

Palabras clave: cartografía edáfica, covariables ambientales, geoestadística, factores formadores del suelo, modelos de predicción

Abstract

The lack of edaphological information and soil maps has led into a search for new techniques to solve this problem, so that a complete and homogeneous coverage of the surface can be obtained. The main objective of this study is the mapping of edaphic properties of interest such as organic matter, sand, silt and clay through the application of the Digital Soil Mapping methodology in the area of Benavides de Órbigo (León). For this purpose, 75 existing soil sampling data and different environmental covariates related to soil-forming factors were selected. The statistical analysis was carried out through the combination of multiple linear regression and generalized linear models to obtain the best prediction model for each variable, in addition to the residuals generated by the model and the error estimation.

Among all the variables, silt was the only one that presented an R^2 value higher than 0.5, while the lowest error was obtained for organic matter. For this reason, it was concluded that, although the method is applicable and useful, it still has great limitations, so it is important to point out the search for solutions that allow to obtain improvements in the results.

Keywords: soil mapping, environmental covariates, geostatistics, soil-forming factors, prediction models

1. Introducción

El desarrollo en las últimas décadas de distintas tecnologías de la información como el Sistema de Posicionamiento Global (GPS), los Sistemas de Información Geográfica (SIG) o la Teledetección y su aplicación en diversos ámbitos del conocimiento ha permitido un gran avance en el estudio espacial de los procesos objeto de análisis, los cuales se basan en la aplicación de métodos estadísticos para elaborar una cartografía de distintas propiedades edáficas.

En el caso de los suelos y de la cartografía edafológica, se ha evolucionado desde una visión descriptiva, centrada en la clasificación, hacia una visión que intente comprender y explicar los procesos y patrones que surgen por la interacción de los suelos con los distintos factores ambientales en los ecosistemas (Grunwald, 2009). Uno de los principales ejemplos de esto se encuentra en el modelo *clorpt* (clima, litología, organismos, relieve y tiempo) (Jenny, 1941), que se centra en cuantificar las relaciones entre el suelo y los factores que controlan la formación de este (Ma *et al.*, 2019).

Además, la importancia de los suelos en la regulación de los ciclos biogeoquímicos y de los flujos de energía y agua, así como por los servicios ecosistémicos que proporcionan (Grunwald *et al.*, 2011), ha hecho que se produzca un aumento de la demanda de información relativa a los suelos (McBratney *et al.*, 2003; Lagacherie y McBratney, 2006).

Respecto a la cartografía edafológica, se observan bastantes diferencias entre los distintos países y regiones, ya que, mientras algunos como Países Bajos, cuentan con una cobertura completa de la superficie, otros como Alemania o Francia no tienen una cobertura homogénea: no todos cubren toda la superficie con la resolución espacial adecuada para permitir cierto grado de detalle ni todas las zonas están cubiertas. Esta escasez de información edafológica se debe principalmente a la problemática del muestreo convencional, el cual es lento y costoso (McBratney *et al.*, 2003; Lagacherie y McBratney, 2006).

Por todo ello, surge la necesidad de buscar soluciones que permitan la elaboración de mapas que contengan dicha información. Así se desarrolla la idea de la creación de mapas digitales de suelos que complementen a los mapas digitalizados tradicionales que se utilizaban hasta ahora. Este nuevo concepto se denomina *Digital Soil Mapping* (DMS) o cartografía digital de suelos, que se define como “la creación y alimentación de sistemas espaciales de información de suelos mediante el uso de métodos de observación de campo y laboratorio, junto con sistemas espaciales y no espaciales para la inferencia de suelos.” (Lagacherie y McBratney, 2006, p. 5).

A partir de este nuevo concepto, surgen diversos modelos basados en la inferencia estadística de las propiedades edáficas o de las clases de suelos. Un ejemplo es el modelo *scorpan*, el cual relaciona estadísticamente propiedades o clases de suelos conocidas con distintas covariables climáticas, litológicas, de vegetación, de usos del suelo, del relieve, etc. De esta forma, se extrapolan dichas relaciones a aquellas zonas que carecen de datos (Figura 1). Para ello, ha sido importante el desarrollo de la geoestadística y de la pedometría (McBratney *et al.*, 2003; Minasny y McBratney, 2016).

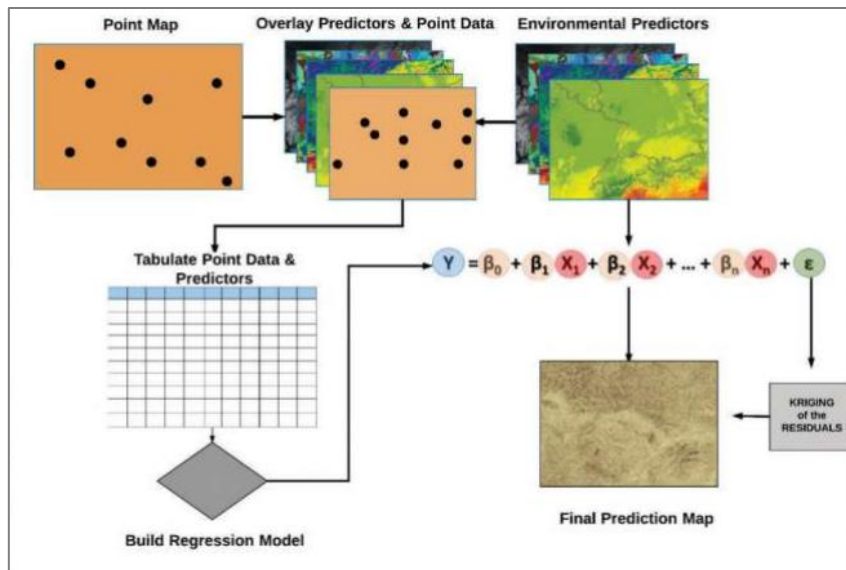


Figura 1. Esquema resumen de la metodología en la que se basa el presente trabajo (Dávila *et al.*, 2018).

Algunas ventajas que presentan estos modelos son su coste y la consistencia de la información que se utiliza, aunque las principales son la facilidad para actualizar los nuevos datos que estén disponibles y la capacidad para determinar los errores de los resultados obtenidos (Carré *et al.*, 2007b).

Cabe destacar que, aunque el coste de los muestreos tradicionales es elevado, aunque menor que el de la cartografía convencional, estos siguen siendo necesarios, pues son un elemento clave a la hora de aplicar y calibrar los distintos modelos que se utilizan para predecir las propiedades del suelo (Carré *et al.*, 2007a; Minasny y McBratney, 2016), además de que son útiles para otro tipo de estudios medioambientales (Mosleh *et al.*, 2016).

En España, la cartografía edafológica convencional existente hasta ahora es bastante escasa o no está actualizada, con mapas de suelos a escalas muy pequeñas (Instituto Geográfico Nacional, sin fecha). Sin embargo, en los últimos años se han ido elaborando mapas a escalas de 1:50.000 e incluso en algunas regiones como Asturias, a escala 1:25.000 (Gobierno del Principado de Asturias, 2018).

En el caso de Castilla y León, se han elaborado varios mapas de suelos de detalle para distintas Denominaciones de Origen y para algunas hojas del Mapa Topográfico Nacional a escala 1:50.000, como en el caso de la provincia de León (Instituto Tecnológico Agrario de Castilla y León, 2017).

En cuanto a la aplicación del DSM a nivel nacional, los ejemplos son limitados. En la mayoría de casos se ha utilizado esta metodología para la cartografía de la distribución del carbono orgánico del suelo, tanto a nivel nacional (Doblas-Miranda *et al.*, 2013) como a nivel regional, como es el caso de Andalucía o Galicia (Armas *et al.*, 2017; Rial Tubío, 2017).

2. Objetivos

El objetivo principal de este trabajo es la realización de un mapa digital de propiedades de suelos (% arena, % arcilla, % limo y % materia orgánica) en la zona de Benavides de Órbigo, en la provincia de León, a partir de la aplicación del método *Digital Soil Mapping* basado en el modelo *scorpan*.

Para su consecución, se plantean los siguientes objetivos específicos:

- Seleccionar las variables ambientales más adecuadas para cada una de las propiedades edáficas analizadas en nuestra zona de estudio.
- Analizar los errores resultantes de la aplicación del método para las cuatro propiedades estudiadas.
- Comprobar si es adecuada la utilización de un grupo de muestreos de calibración y un grupo de validación.
- Analizar si los distintos orígenes de los muestreos de suelo pueden afectar al resultado obtenido.

3. Área de estudio

El área de estudio sobre la que se centra el presente trabajo se sitúa en la hoja 160 del Mapa Topográfico Nacional 1:50.000, correspondiente a Benavides de Órbigo, en la provincia de León. Comprende altitudes entre los 830 y los 1330 metros y está bañada por los ríos Órbigo y Tuerto (Figura 2).

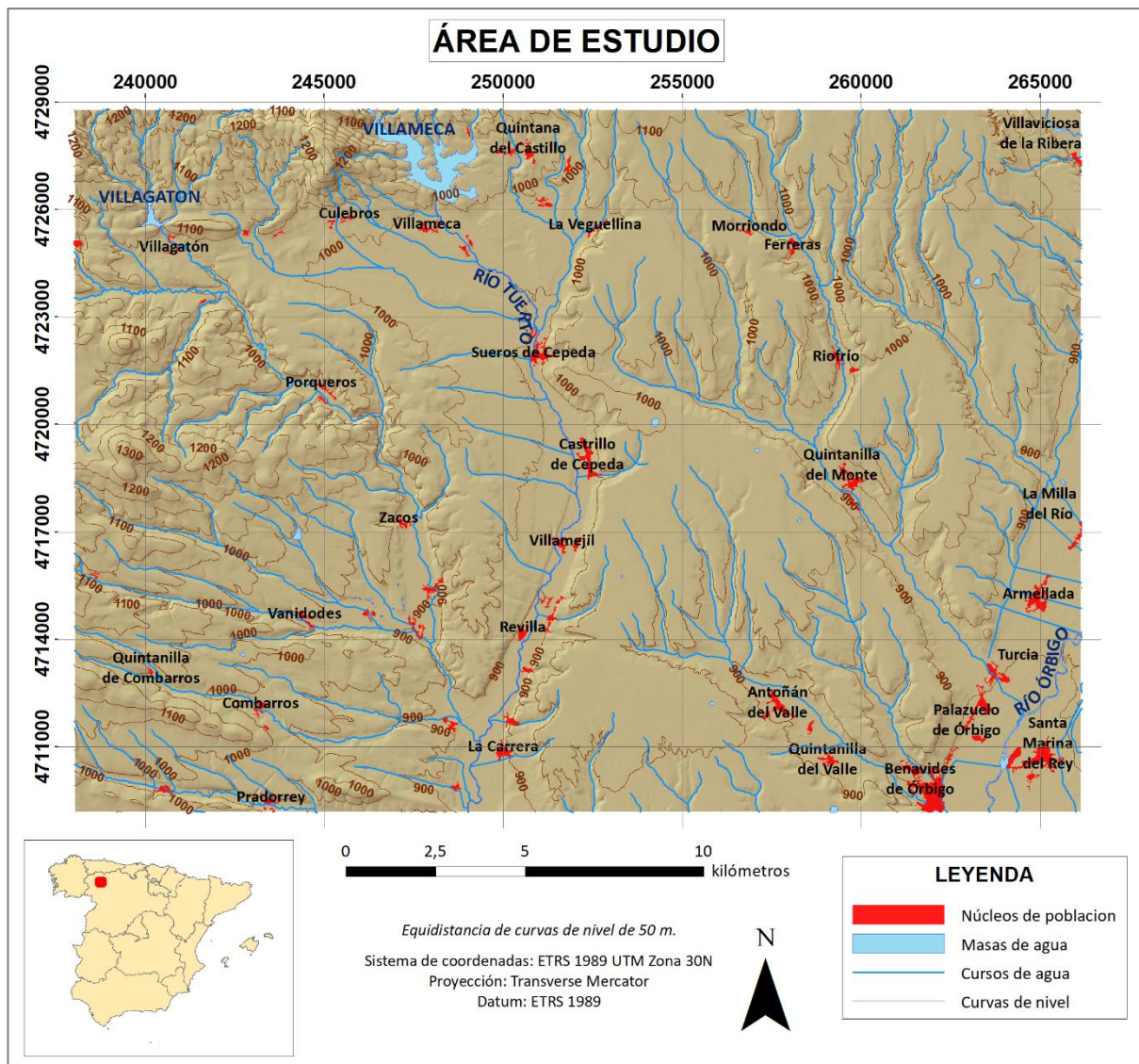


Figura 2. Localización del área de estudio.

3.1. Geología, litología y geomorfología

En el aspecto geológico (Figura 3), las litologías correspondientes al Paleozoico están formadas principalmente por pizarras, areniscas y cuarcitas de distintas formaciones de edad Cámbrico-Ordovícico (Serie de los Cabos, Pizarras de Luarca y Formación Agüeira) y se encuentran exclusivamente en la zona oeste y noroeste de la hoja, mientras que las litologías del Terciario, asociadas al Mioceno superior, y del Cuaternario, de edad Pleistoceno-Holoceno, siendo fundamentalmente materiales silíceos asociados a conglomerados, arenas, limos y arcillas, se distribuyen por el resto de la zona de estudio (Rodríguez Fernández y Pérez González, 1984). Según el grado de consolidación se pueden clasificar en tres grupos litológicos principales comprendidos por materiales del Paleozoico asociados a litologías consolidadas,

por materiales del Terciario asociados a litologías poco consolidadas, y por materiales del Cuaternario sin consolidar.

En cuanto a la geomorfología, hay un predominio de formas fluviales, sobre todo terrazas y llanuras aluviales y fondos de valle, en la zona centro y este de la hoja. Los niveles de terrazas fluviales se encuentran tapizando los relieves miocénicos, mientras que en las zonas de piedemonte se encuentran las rañas. Además, las formas poligénicas asociadas a glacis y superficies de depósito dominan en la zona centro, mientras que las superficies sin depósito correspondientes a zonas de relieves Paleozoicos son más relevantes al oeste y noroeste del área de estudio (Rodríguez Fernández y Pérez González, 1984).

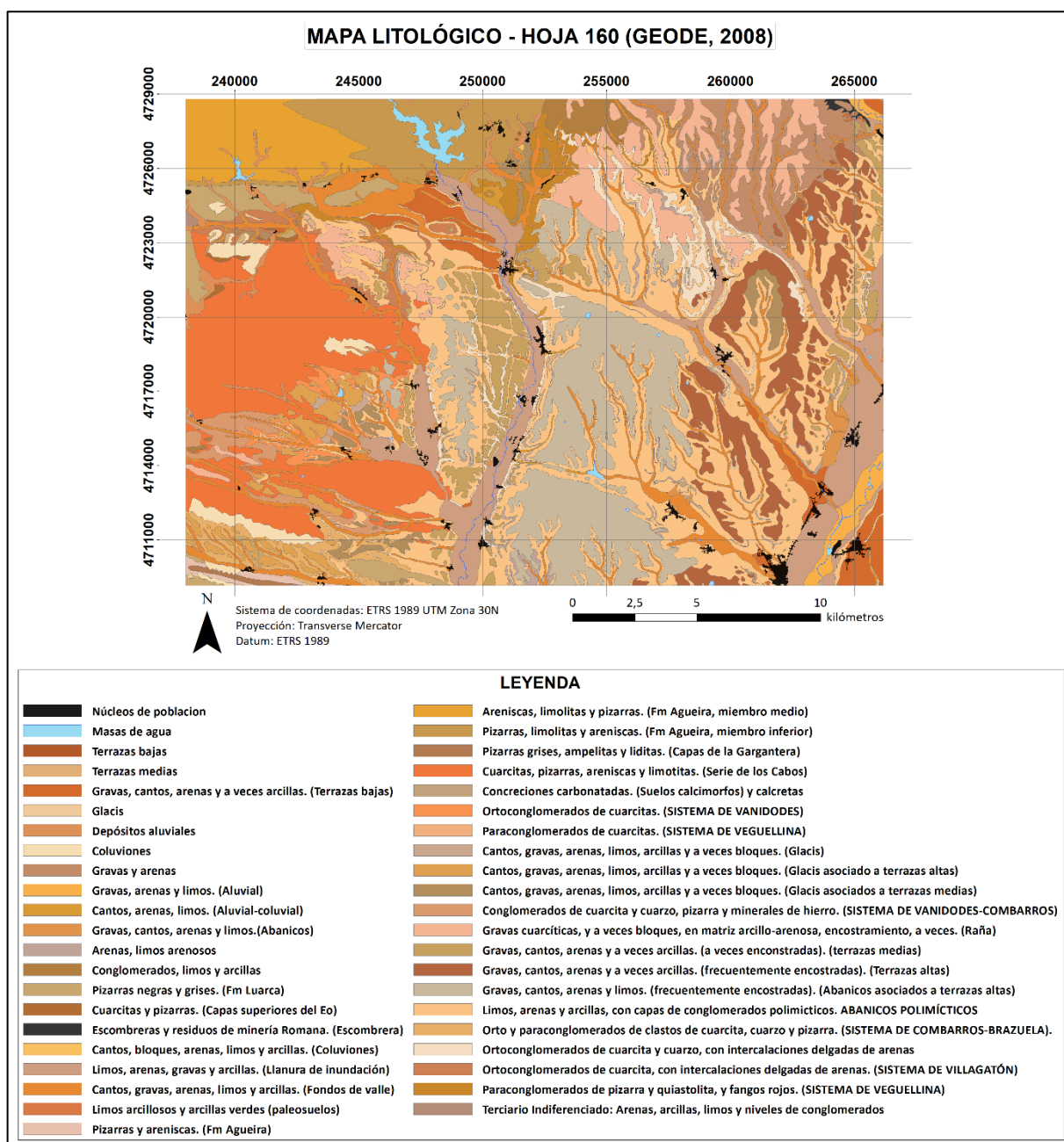


Figura 3. Mapa litológico de la hoja 160 a partir del GEODE 1:50.000 (González Menéndez *et al.*, 2008).

3.2. Edafología

Los suelos que cubren la mayor parte de la superficie de la hoja 160 (Figura 4), siguiendo la clasificación de la FAO (1998), son las asociaciones de Acrisol úmbrico / Acrisol abrupto con inclusiones de Acrisol profundo / Acrisol gléico, que dominan en la zona centro y noreste de la hoja, asociados a zonas de terrazas medias y altas. Por otro lado, las asociaciones de Regosol léptico / Umbrisol háptico y Leptosol úmbrico con inclusiones de Leptosol lítico, son las que presentan un mayor dominio en la zona norte y oeste del área de estudio (Alonso Herrero *et al.*, 2010).

Destacan, además, las asociaciones de Cambisol eútrico / Cambisol dístrico en zonas de terrazas bajas, además de Fluvisol dístrico / Regosol dístrico cercanas a los cursos de agua, y las asociaciones de Umbrisol húmico / Regosol dístrico junto con Cambisol crómico en zonas de laderas, distribuidos de forma heterogénea por toda la hoja (Alonso Herrero *et al.*, 2010).

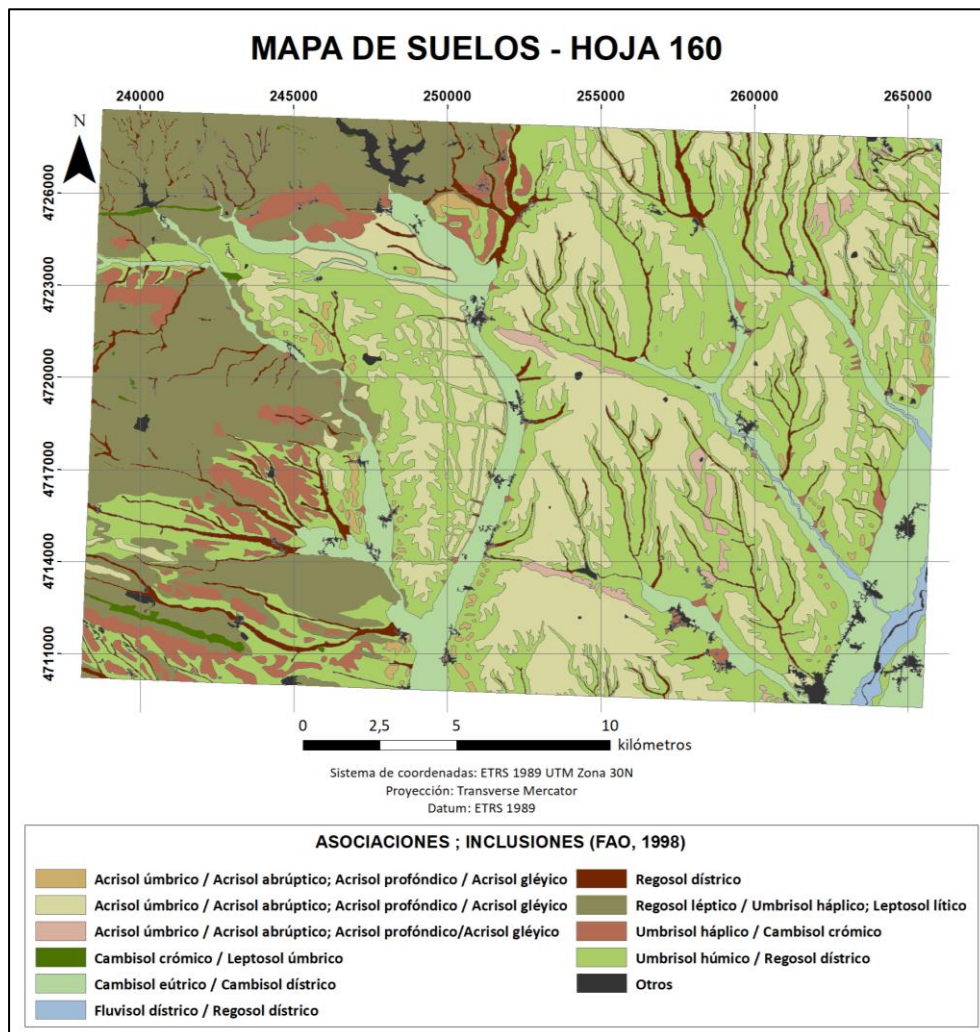


Figura 4. Mapa de suelos de la hoja 160 modificado de Alonso Herrero *et al.* (2010).

3.3. Climatología

El tipo de clima es continental templado, con veranos secos y templados (Chazarra Bernabé *et al.*, 2018). Según los datos de las estaciones termopluviométricas de Hospital de Órbigo y de Carrizo de la Ribera (De León Llamazares *et al.*, 1991), ambas situadas próximas a la zona de estudio, la temperatura media anual son 10,5°C y las precipitaciones están en torno a los 520 mm anuales, mientras que la evapotranspiración potencial anual ronda los 650 mm.

Las temperaturas más altas se alcanzan durante los meses de verano, en los cuales la ETP alcanza su máximo, en torno a 120 mm. Por otro lado, las precipitaciones se distribuyen en dos periodos máximos correspondientes a invierno y primavera, mientras que el mínimo se alcanza durante el periodo estival, siendo julio y agosto los meses más secos (Figura 5).

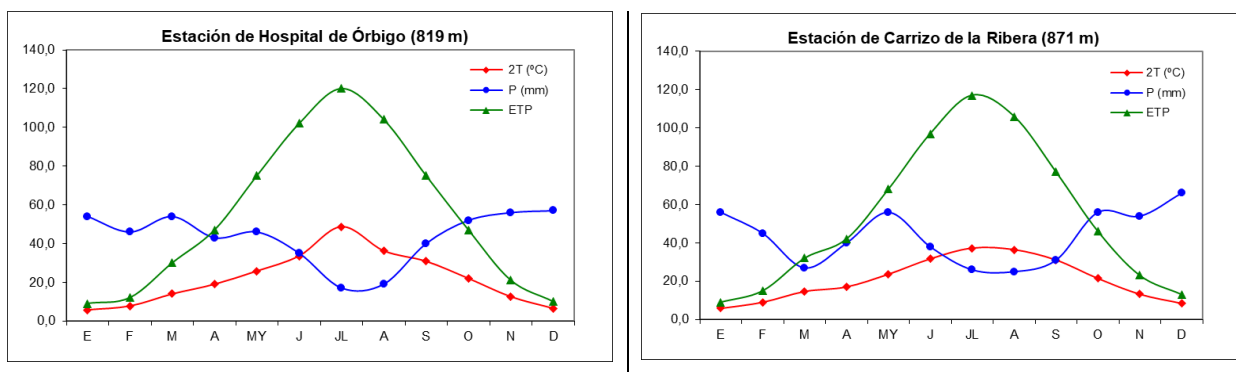


Figura 5. Climatogramas de las estaciones termopluviométricas de Hospital de Órbigo y de Carrizo de la Ribera. Fuente de datos: De León Llamazares *et al.*, 1991

3.4. Usos del suelo y vegetación

El uso de suelo predominante es el agrícola, con casi 20000 hectáreas, abarcando aproximadamente el 36% de la superficie del área de estudio. Destacan el cultivo herbáceo, en su mayoría de regadío, sobre todo en las zonas centro y sur de la hoja, y las combinaciones de cultivos con otro tipo de vegetación natural en las zonas de ribera.

Por otro lado, la vegetación asociada a matorral, pastizales o herbazales y bosques, además de distintas combinaciones de vegetación, también representa un porcentaje muy alto de la superficie del área de estudio (62%), distribuyéndose de forma heterogénea por toda la zona. En el caso de los bosques, la mayoría se corresponden a bosques de frondosas y bosques de coníferas (Figura 6).

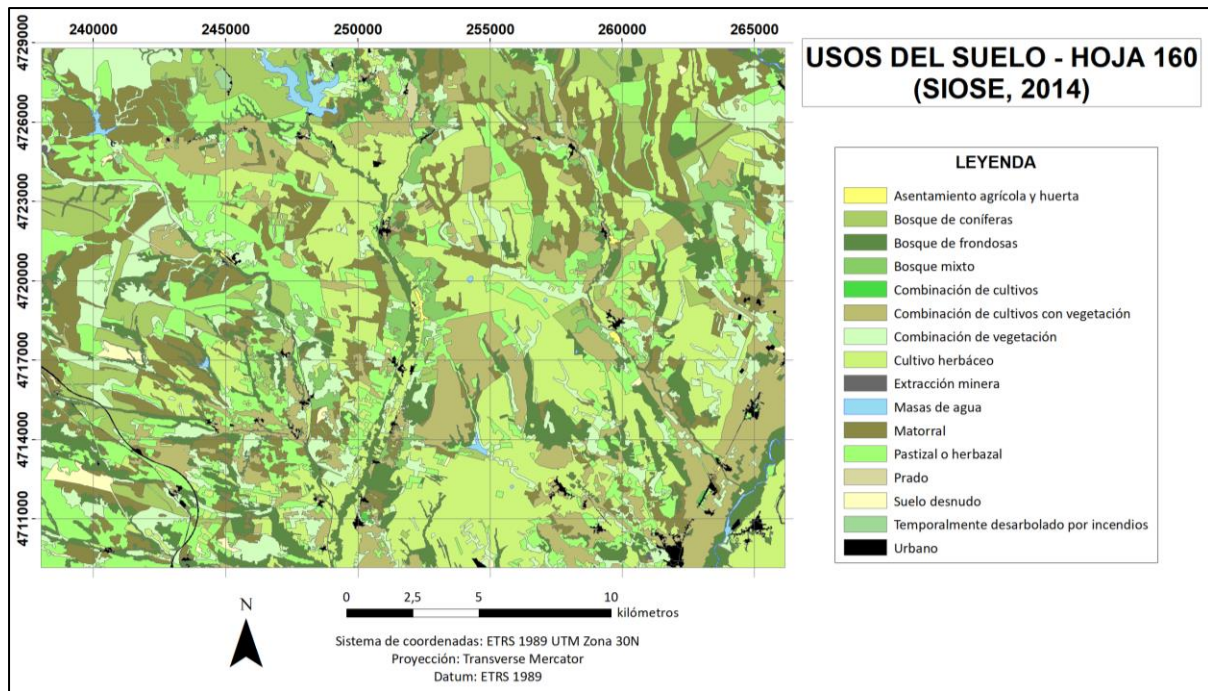


Figura 6. Usos del suelo de la hoja 160 Fuente de datos: SIOSE (Instituto Geográfico Nacional, 2014).

4. Materiales y métodos

La metodología empleada en el presente trabajo para la realización de la cartografía digital del suelo de la hoja 160 del Mapa Topográfico Nacional, se basa en el modelo *scorpan* (McBratney *et al.*, 2003; Lagacherie y McBratney, 2006; Minasny y McBratney, 2016). Este modelo considera que, a partir de la integración de distintas variables ambientales en una función, se pueden predecir las clases de suelo (S_c) o los atributos (S_a) de este (Ec. 1).

$$S_c \text{ o } S_a = f(s, c, o, r, p, a, n) \quad (\text{Ec. 1})$$

Dicha función se define a partir de puntos del territorio en los que se tienen tanto datos de variables ambientales como de suelos. De esta forma, posteriormente, es posible averiguar las propiedades del suelo en lugares que no han sido muestreados previamente (y por tanto no existen datos de suelo), pero donde sí se conocen sus variables ambientales.

Las variables ambientales utilizadas se denominan “factores formadores del suelo”, que comprenden al propio suelo (s), el clima (c), los organismos (o), el relieve o topografía (r), el material parental (p), la edad (a) y la posición espacial (n). Sin embargo, la mayoría de trabajos en este ámbito no utiliza los siete factores, debido a que no hay la misma disponibilidad de variables ambientales para todos los factores en todos los países (McBratney *et al.*, 2003).

Independientemente del número que se seleccione, cada uno de estos factores se define a partir de distintas covariables ambientales. Mediante la superposición espacial de los perfiles de suelos sobre la cartografía de las covariables, se asigna a cada perfil o punto de muestreo el valor de cada una de las covariables. De esta forma, en cada ubicación correspondiente a los perfiles de suelo, se obtiene, por un lado, los valores de las propiedades de ese suelo y, por otro, los valores de cada una de las covariables ambientales seleccionadas para el estudio (Dávila *et al.*, 2018).

Todo ello permite la realización de una regresión lineal múltiple, con el objetivo de modelizar las relaciones entre la variable dependiente y las distintas covariables ambientales en los puntos de muestreo, y usar esas relaciones para conocer los valores de la variable dependiente en lugares sin muestrear (Hengl *et al.*, 2007).

La variable dependiente en este caso se corresponde con las clases o atributos del suelo (propiedades edáficas) que se quieren conocer, mientras que las covariables ambientales o predictores son las variables independientes. Esta técnica se puede resumir mediante la siguiente ecuación (Ec. 2), donde Y es la variable dependiente, X_n son los predictores que explican la variable dependiente, β_0 es el intercepto u origen, β_n son los coeficientes que representan el peso y la relación de cada covariable ambiental con la variable dependiente, y ε son los valores residuales que no pueden ser explicados por el modelo (Dávila *et al.*, 2018).

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \dots + \beta_n \cdot X_n + \varepsilon \quad (\text{Ec. 2})$$

En el caso de estudio, la variable dependiente que se quiere conocer son los atributos del suelo, tales como los porcentajes de materia orgánica del suelo, arcilla, limo y arena.

Así pues, el suelo puede considerarse como variable dependiente (S_c o S_a) o como factor formador (s). En este trabajo, se utilizará como variable dependiente (S_a).

Los datos de la variable dependiente provienen de dos fuentes de datos. La primera es la Base de Datos de los Suelos de Castilla y León (Instituto Tecnológico Agrario de Castilla y León, 2012), mantenida por el Instituto Tecnológico Agrario de Castilla y León (ITACYL), de la que se han extraído los datos de 45 muestreos. Sus datos, a su vez, provienen de dos fuentes: el Plan 2014 para la remolacha de AIMCRA y el Inventario Nacional de Erosión de Suelos (INES) 2002-2012 realizado por el Ministerio de Agricultura, Alimentación y Medio Ambiente. Estas muestras contienen los porcentajes de materia orgánica, arcilla, limo y arena y se corresponden con los 25 o 30 primeros centímetros del suelo. Para su uso, se han seleccionado

solo aquellas muestras que se encuentran en la zona de estudio. La segunda fuente de datos son 30 perfiles realizados en la hoja 160 para el desarrollo del Plan Piloto del Mapa de suelos de la Provincia de León (Alonso Herrero *et al.*, 2010). De ellos, se han seleccionado solo los datos del horizonte A para homogeneizar la información con los datos del ITACYL. En los casos en los que hay más de un horizonte A, se realiza la media ponderada en función de la profundidad de cada horizonte (Ec. 3).

$$\%Horiz. A = \frac{(\%Horiz. Ah1 \times prof. Ah1) + (\%Horiz. Ah2 \times prof. Ah2)}{profundidad\ total} \quad (Ec. 3)$$

En cuanto a los factores formadores del suelo, para este trabajo se han seleccionado el clima (c), la vegetación (o), la topografía (r) y el material parental (p) para ser evaluados en la obtención de las variables del modelo. El factor edad (a) no se ha considerado debido a la escasez de cartografía existente y al error asociado a estos mapas (McBratney *et al.*, 2003), mientras que el factor posición espacial (n) se ha descartado ya que está incluido dentro de los perfiles del suelo al estar estos vinculados a unas coordenadas. Por otro lado, el factor suelo no se incluye como factor formador debido a que se utiliza como variable dependiente.

En la Figura 7 se muestra el método seguido en el presente estudio.

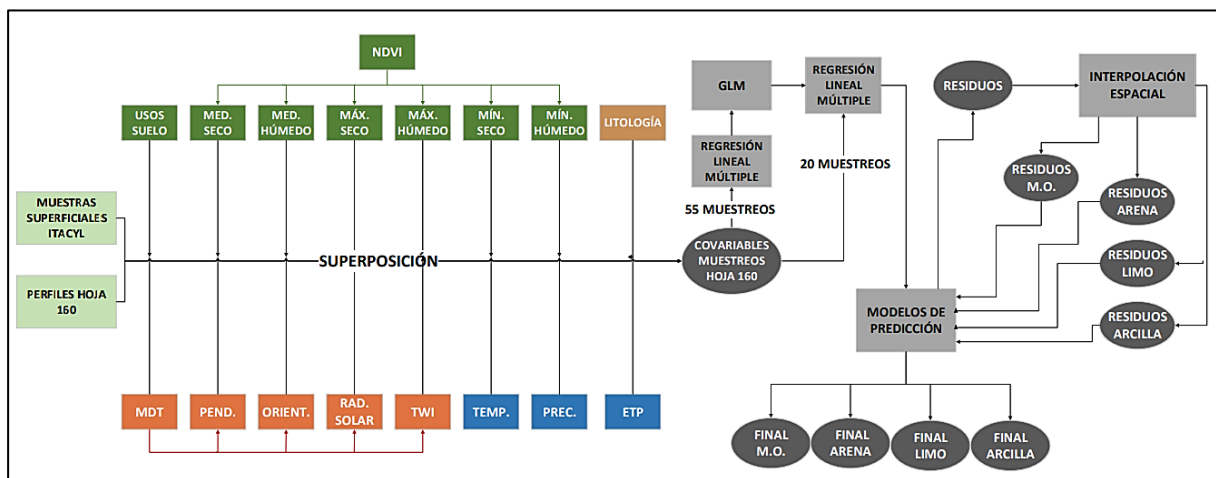


Figura 7. Diagrama de flujo resumen de la metodología empleada en el presente trabajo (en verde, las variables asociadas al factor o; en marrón, las variables asociadas al factor p; en rojo, las variables asociadas al factor r; y en azul, las variables correspondientes al factor c).

4.1. Análisis de factores y obtención de covariables

A continuación, se hace referencia a las covariables ambientales seleccionadas para el estudio de las cuatro propiedades edáficas mencionadas (% materia orgánica, % arena, % limo y % arcilla).

4.1.1. Clima

En el caso del clima, se han utilizado como covariables ambientales la precipitación media anual (mm), la temperatura media anual (°C), la evapotranspiración potencial media anual (mm) y la radiación solar (KWH/m²·año) para el año 2017. Se seleccionó un año reciente con el objetivo de tener más información disponible y poder realizar una primera aproximación al modelo.

Excepto para la radiación solar, se extrajeron los datos mensuales de dichas variables estimadas a partir de los datos derivados del modelo SIMPA 2019 (Sistema Integrado de Modelización Precipitación-Aportación). Este modelo ha sido desarrollado por el Centro de Estudios Hidrográficos del CEDEX para evaluar los recursos hídricos en régimen natural de España (Ministerio para la Transición Ecológica y el Reto Demográfico, 2019). Sus datos se encuentran en formato ráster y presentan una resolución espacial de 500 m × 500 m. Al presentar un tamaño de píxel muy elevado, es necesario transformarlo a una resolución espacial de 25 m × 25 m para que se ajuste al resto de capas que se utilizan en este trabajo. Con los datos mensuales se realizó la media anual.

Por otro lado, la radiación solar se calculó para todo el año 2017 a través de intervalos mensuales y usando el modelo de radiación difusa uniforme, lo que supone que la radiación difusa incidente es la misma desde todas las direcciones. Para su obtención se partió del Modelo Digital del Terreno (MDT). El MDT, en formato ráster, presenta una resolución espacial de 25 m × 25 m y representa las altitudes del terreno sobre el nivel del mar. Se obtuvo a través del Centro de Descargas del Centro Nacional de Información Geográfica.

4.1.2. Vegetación

Las covariables ambientales para el factor vegetación que se han analizado son los estadísticos (media, mínimo y máximo) derivados del NDVI (*Normalized Difference Vegetation Index*) para cada época extrema, período húmedo y período seco, y los usos del suelo.

El NDVI se obtuvo a partir de las imágenes de satélite Sentinel-2 L2A, las cuales ya presentan una corrección radiométrica y atmosférica previa (European Space Agency, 2021). En concreto, se obtuvieron las imágenes correspondientes a las bandas 4 (665 nm) y 8 (842

nm), que se asocian con las bandas del rojo y del infrarrojo cercano, respectivamente (Ec. 4) (Myneni *et al.*, 1995; Pettorelli *et al.*, 2005).

$$NDVI = \frac{B8 - B4}{B8 + B4} \quad (Ec. 4)$$

El cálculo del NDVI se realizó teniendo en cuenta las dos épocas extremas en la zona de estudio, para lo cual se utilizaron los datos de precipitación mensual del Modelo SIMPA 2019. Para ello, se observó la tendencia que presentaban las precipitaciones durante el periodo comprendido entre los años 1940/41 y 2017/18. De esta forma, se delimitaron dos etapas:

- Etapa seca: comprende un periodo de cuatro meses, el cual engloba los meses de junio, julio, agosto y septiembre. Los valores máximos de precipitación se encuentran en un rango de entre 27 y 59 mm; mientras que los valores mínimos son de entre 17 y 34 mm.
- Etapa lluviosa: comprende los ocho meses restantes: enero, febrero, marzo, abril, mayo, octubre, noviembre y diciembre. Presenta valores máximos de precipitación de entre 77 y 154 mm; mientras que los valores mínimos se encuentran en un rango de entre 38 y 51 mm.

Teniendo en cuenta esta tendencia, en este trabajo el periodo seco se establece desde junio de 2017 hasta septiembre de 2017, mientras que el periodo lluvioso está comprendido desde octubre de 2017 hasta mayo de 2018.

Para el cálculo, se utilizaron todas las imágenes de satélite existentes para cada periodo, excepto las correspondientes al mes de marzo, ya que estas no presentaban una calidad adecuada debido a la abundancia de nubosidad. Las imágenes descargadas se encuentran en formato ráster con una resolución espacial de 35,17 m × 35,17 m. Dado que la resolución espacial difiere respecto a la de otras capas utilizadas, esta se transformó a un tamaño de celda de 25 m × 25 m. Posteriormente, con todos los NDVI que existen para cada periodo, se realizó el cálculo de los tres estadísticos mencionados, obteniéndose seis capas derivadas del NDVI.

En cuanto a los usos del suelo, se obtuvieron a partir del Sistema de Información de Ocupación del Suelo de España (SIOSE) del año 2014. Con el objetivo de reducir el número de clases, se realizó una reclasificación de las clases presentes en la zona de estudio (Tabla 1). En el caso de la clase “Combinación de vegetación”, aunque en la ortofoto presenta varias zonas de vegetación arbórea, predomina el matorral, por lo que se clasificó dentro de la categoría “Matorral”. De cara al posterior análisis estadístico, al ser esta covariable de tipo categórico, se

realizó una transformación de esta a una variable numérica ordinal en función de la cobertura de la superficie terrestre que representan, de forma que se obtuvieron cuatro nuevos valores correspondientes a cada una de las clases en las que se dividen los usos del suelo. Además, se reclasificó la covariable con el objetivo de obtener cuatro nuevas capas correspondientes a cada valor asignado para que puedan ser utilizadas en el análisis final con SIG.

Tabla 1. Simplificación de los usos de suelo del SIOSE (Instituto Geográfico Nacional, 2014) de la zona de estudio.

Usos del suelo según el SIOSE	Simplificación	Código
Temporalmente desarbolado por incendios, Suelo desnudo	Suelo desnudo	1
Asentamiento agrícola y huerta, Cultivo herbáceo, Combinación de cultivos, Combinación de cultivos con vegetación, Pastizal, Prado	Agrícola-prado-pastizal	2
Matorral, Combinación de vegetación	Matorral	3
Bosque de frondosas, Bosque de coníferas, Bosque mixto	Bosque	4

4.1.3. Topografía

En este caso, las covariables ambientales utilizadas son la altitud, la pendiente, la orientación y el índice topográfico de humedad (TWI). La fuente de datos para su estimación fue el MDT de 25 x 25 m descrito anteriormente en el apartado 4.1.1.

Los mapas derivados de pendiente, el cual está expresado en grados, y de orientación, expresado en radianes, derivados a partir del MDT se elaboraron con el programa ArcGIS Desktop 10.8. En el caso de la orientación, esta se transformó de grados a radianes (Ec. 5).

$$\text{Orientación (radianes)} = \left(\frac{\text{orientación (grados)}}{360} \right) \times 2\pi \quad (\text{Ec. 5})$$

Por otro lado, se calculó de nuevo con SIG el índice topográfico de humedad, derivado también del MDT, el cual representa las áreas topográficamente más planas susceptibles de estar saturadas de agua. Este índice se determina a partir de la dirección del flujo y de la escorrentía acumulada (Ruhoff *et al.*, 2011).

Para su cálculo, se debe corregir el MDT inicial para eliminar aquellos píxeles que actúen como sumideros, es decir, píxeles cuyas celdas de alrededor presentan altitudes mayores

y, por tanto, no permite que el agua que llega a ellos se mueva por la superficie. Posteriormente, con el MDT corregido, se puede llevar a cabo el cálculo de la dirección de flujo y de la escorrentía acumulada o acumulación de flujo (Mattivi *et al.*, 2019).

La dirección de flujo, que representa las distintas direcciones por las que puede discurrir el agua en un territorio, se puede calcular a través de distintos algoritmos. En este caso, se utilizó el algoritmo *Deterministic Infinity* (D^∞), ya que muestra mejores resultados que otros algoritmos (Sørensen *et al.*, 2006; Ågren *et al.*, 2014). Este método asigna un número ilimitado de direcciones de flujo mediante el cálculo de la dirección con la pendiente más escarpada basado en facetas triangulares (Sørensen *et al.*, 2006; Ågren *et al.*, 2014). Además, a diferencia de otros métodos, el algoritmo D^∞ ni concentra el flujo solo en un canal principal ni lo dispersa por toda la superficie (Ruhoff *et al.*, 2011).

En cuanto a la escorrentía acumulada, esta representa las zonas de acumulación de agua. Se basa en el cálculo del número de celdas que drenan hacia una celda receptora, por lo que es necesario conocer la dirección de flujo del agua (Schmidt y Persson, 2003). Este mapa de escorrentía acumulada tiene como valores el número de celdas que drena a cada una de ellas, es decir, las zonas correspondientes a las divisorias de cuenca presentarán valores 0, mientras que la celda correspondiente a la salida de la cuenca tendrá como valor el total de celdas que componen dicha cuenca.

Para estimar el índice topográfico de humedad, además, se ha de calcular la pendiente en radianes ($pend_{rad}$) (Ec. 6) para, posteriormente, calcular su tangente ($pend_{tan}$) (Ec. 7).

$$pend_{rad} = \frac{(pendiente \times 1,570796)}{90} \quad (Ec. 6)$$

$$pend_{tan} = Con(pend_{rad} > 0, Tan(pend_{rad}), 0,001) \quad (Ec. 7)$$

A continuación, se calculó la superficie específica de contribución (SCA) teniendo en cuenta la acumulación y el tamaño del lado del píxel del modelo digital del terreno (Ec. 8).

$$SCA = (acumulación + 1) \times tamaño\ lado\ píxel \quad (Ec. 8)$$

Finalmente, teniendo en cuenta todos los aspectos mencionados anteriormente, se obtuvo el índice topográfico de humedad (Ec. 9) (McBratney *et al.*, 2003; Schmidt y Persson, 2003; Sørensen *et al.*, 2006; Ruhoff *et al.*, 2011; Ågren *et al.*, 2014; Mattivi *et al.*, 2019).

$$TWI = Ln\left(\frac{SCA}{pend_{tan}}\right) \quad (Ec. 9)$$

4.1.4. Material parental

La covariable utilizada para este factor es la litología, extraída a partir del mapa geológico GEODE (González Menéndez *et al.*, 2008), el cual se basa en una cartografía geológica digital continua a escala 1:50.000 que homogeneiza la cartografía del mapa geológico nacional MAGNA existente hasta ahora, con el objetivo de corregir los problemas surgidos por la utilización de distintos criterios a la hora de la elaboración del MAGNA (Rodríguez Fernández y Pérez González, 1984). Al igual que en los usos del suelo, se redujo el número de clases litológicas a través de una simplificación basada en la memoria MAGNA de la hoja 160 a partir de la leyenda propuesta por Alonso Herrero (1995) (ver Anexo, Tabla A1).

Además, la litología, al igual que los usos del suelo, también es una variable categórica, por lo que se transformó a una variable numérica ordinal en función de su grado de consolidación, tamaño de grano, naturaleza litológica y, finalmente su susceptibilidad a meteorizarse para poder realizar el análisis estadístico posterior, obteniéndose cuatro nuevos valores correspondientes a las cuatro clases resultantes de la simplificación de la litología. En este caso también se reclasificó la covariable, originando cuatro capas nuevas correspondientes a cada valor asignado para utilizarlas posteriormente en el análisis con SIG.

4.2. Análisis estadístico

Dado que el número total de covariables era muy elevado, se realizó un análisis de las correlaciones entre las distintas derivadas del NDVI, considerando valores del coeficiente de Pearson superiores a 0,65 como alta correlación, de forma que se eliminó una covariable de aquellas correlaciones que superasen dicho valor, suponiendo dependencia lineal entre ellas.

En cuanto a las variables dependientes, se comprobó el tipo de distribución de cada una de ellas, para tenerla en cuenta en los análisis estadísticos posteriores.

Como se comentó anteriormente, mediante análisis espaciales de superposición, se obtuvo para cada uno de los puntos de muestreo de suelo el valor de cada una de las covariables propuestas y se llevó a cabo el análisis estadístico. Para ello, primero se hizo una selección aleatoria de aproximadamente el 75% de los casos, trabajando con 55 muestreos y dejando los 20 restantes para una posterior validación del modelo.

Con los muestreos seleccionados se estimaron diversas regresiones lineales múltiples (MLR) para cada una de las cuatro variables dependientes (% arena, % limo, % arcilla y % materia orgánica). Para ello, se utilizó el método de “eliminación hacia atrás” utilizando todas las variables independientes, excluyéndolas sucesivamente en función de la menor correlación de cada una de ellas con la variable dependiente (IBM Corporation, 2021). De todos los modelos resultantes para cada una de las variables dependientes, se seleccionaron dos que presentasen un valor de R^2 adecuado teniendo en cuenta su valor de significancia y el número de covariables lo más pequeño posible. Tras la selección de los dos modelos para cada variable, se estimaron modelos lineales generalizados (GLM) utilizando las covariables que representaban cada uno de los modelos seleccionados, obteniéndose un valor del Criterio de Información de Akaike (AIC), el cual permite evaluar la idoneidad de un modelo estadístico concreto en comparación a otros modelos (Martínez *et al.*, 2009). Posteriormente, para cada variable dependiente, se seleccionó aquel modelo que menor valor de AIC presentaba o, en caso de que los modelos presentasen índices similares, se seleccionó el que menos covariables contenía. Finalmente, una vez definidas qué covariables explicaban mejor cada una de las variables dependientes, se volvió a estimar otra regresión lineal múltiple para cada variable dependiente, pero en este caso forzando a utilizar solo las covariables definitivas a fin de obtener las ecuaciones finales que explicaban cada variable dependiente y que permitieron obtener los valores estimados tanto para los 55 muestreos con los que se ha trabajado como para los 20 muestreos que permiten validar el modelo. Esos valores predichos se compararon con los valores reales medidos, de forma que se obtuvieron los residuos para cada uno de los muestreos (Ec. 10).

$$Residuo = (Valor\ real) - (Valor\ estimado) \quad (Ec. 10)$$

Además, se calculó el error cuadrático medio (RMSE) con el objetivo de analizar la validez del método habiendo utilizado 55 muestreos de calibración y 20 muestreos de validación (Ec. 11).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (medido - estimado)^2}{n}} \quad (Ec. 11)$$

4.3. Análisis con SIG

Una vez obtenidos los valores de los residuos tanto para los 55 muestreos con los que se trabajó, como para los 20 muestreos de validación del modelo, mediante SIG se realizó una interpolación para toda el área de estudio utilizando el método del inverso de la distancia (IDW), el cual consiste en una estimación de los valores de cada una de las celdas teniendo en cuenta la distancia a cada uno de los puntos de muestreo existentes, los cuales se ponderan de forma que los puntos más cercanos tendrán más influencia que los que se encuentran más lejos (Esri, 2015). Con ello, se obtuvo un ráster que contiene los residuos de cada variable dependiente para toda el área de estudio.

Finalmente, se llevó a cabo la aplicación de las distintas ecuaciones obtenidas en el análisis estadístico, incluyendo además el valor de los residuos de la predicción, de forma que se obtuvieron cuatro mapas finales correspondientes al contenido de la materia orgánica, arena, limo y arcilla de los suelos de la hoja 160.

5. Resultados

5.1. Análisis de linealidad y exploración de resultados. Análisis de regresión múltiple y GLM

El análisis de la correlación entre las distintas covariables derivadas del NDVI (Tabla 2), muestra que el NDVI medio del periodo seco (*ndvi_med_s*) y el NDVI medio del periodo húmedo (*ndvi_med_h*) son las que presentan un coeficiente de Pearson alto y una correlación elevada con un mayor número de variables (Anexo, Tabla A2), por lo que, seleccionándolas, se puede prescindir de la información que aportan el resto de covariables del NDVI relacionadas.

En cuanto al tipo de distribución de las variables dependientes, tal y como se puede observar en la Figura A1 del Anexo todas las variables siguen una distribución normal, acumulándose la mayoría de datos en el centro del histograma; excepto la materia orgánica, que presenta una distribución gamma, con la mayoría de valores en el extremo derecho del histograma.

Todo ello, además, se contrasta con los gráficos P-P, observándose que las variables siguen la distribución supuesta en cada caso.

Tabla 2. Correlaciones de las variables de NDVI con coeficiente de Pearson superior a 0,65 (*ndvi_min_s*: NDVI mínimo del período seco; *ndvi_min_h*: NDVI mínimo del período húmedo; *ndvi_med_s*: NDVI medio del período seco; *ndvi_med_h*: NDVI medio del período húmedo; *ndvi_max_s*: NDVI máximo del período seco; *ndvi_max_h*: NDVI máximo del período húmedo).

Variables		Coefficiente
<i>ndvi_min_s</i>	<i>ndvi_min_h</i>	0,717
<i>ndvi_min_s</i>	<i>ndvi_med_s</i>	0,877
<i>ndvi_min_s</i>	<i>ndvi_med_h</i>	0,772
<i>ndvi_min_s</i>	<i>ndvi_max_s</i>	0,734
<i>ndvi_min_h</i>	<i>ndvi_med_h</i>	0,870
<i>ndvi_med_s</i>	<i>ndvi_med_h</i>	0,658
<i>ndvi_med_s</i>	<i>ndvi_max_s</i>	0,942
<i>ndvi_med_h</i>	<i>ndvi_max_h</i>	0,744

Por otro lado, tras la realización de la regresión lineal múltiple MLR y del GLM (Tabla 3), en el caso de la materia orgánica, se han seleccionado los modelos *c* y *f* (Anexo, Tabla A3), con unos valores de R^2 de 0,457 y 0,449 respectivamente. Tras la realización del GLM, se ha obtenido un valor de AIC de 237,418 para ambos casos, por lo que se selecciona el modelo *f* al ser el que menos covariables contiene (Ec. 12). Se observa que las variables relacionadas con la litología y con los usos del suelo explican en gran medida el porcentaje en materia orgánica, siendo importantes los usos del suelo asociados a bosques, matorral y suelo desnudo, y las litologías consolidadas.

$$\begin{aligned}
 \text{Materia orgánica} = & 18,978 + (1,002 \times \text{usos4}) + (-0,009 \times \text{altitud}) + (-3,617 \times \text{lito1}) + \\
 & (2,113 \times \text{usos3}) + (-0,306 \times \text{orient_rad}) + (5,480 \times \text{ndvi_med_h}) + (-1,902 \times \text{lito2}) + \\
 & (7,470 \times \text{usos1}) + (0,119 \times \text{pr}_{17_160}) + (-0,096 \times \text{pend}) + (-1,049 \times \text{tem}_{17_160})
 \end{aligned}$$

(Ec. 12)

En cuanto a la arena, los modelos seleccionados son el *f* y el *h* (Anexo, Tabla A3), cuyos valores de R^2 son 0,490 y 0,483 respectivamente. Una vez realizado el GLM, los valores de AIC obtenidos han sido en ambos casos de 414,865, por lo que se selecciona el modelo *h* (Ec. 13). En este caso, las litologías consolidadas también tienen gran importancia en el modelo, además de los usos forestal y matorral.

$$\text{Arena} = 150,366 + (5,651 \times \text{usos4}) + (-12,868 \times \text{lito1}) + (3,109 \times \text{usos3}) + (-0,583 \times \text{twi}) + (-0,035 \times \text{rad_s_kw}) + (-18,272 \times \text{lito2}) + (-13,236 \times \text{ndvi_med_s}) + (-0,186 \times \text{pr_17_160}) + (-3,549 \times \text{tem_17_160}) \quad (\text{Ec. 13})$$

El limo es la variable que presenta valores de ajuste R^2 mayores. En este caso, se han seleccionado los modelos *d* e *i* (Anexo, Tabla A3), con valores de R^2 de 0,509 y de 0,501 respectivamente. Tras el GLM, el modelo seleccionado ha sido el *i*, ya que, para ambos casos, el valor de AIC es de 422,058 (Ec. 14). Se observa que ambos NDVI medios son variables que explican el porcentaje de limo en nuestra área de estudio, además de las litologías consolidadas y poco consolidadas.

$$\text{Limo} = 2,720 + (0,212 \times \text{etp_17_160}) + (0,284 \times \text{pr_17_160}) + (5,325 \times \text{ndvi_med_s}) + (17,654 \times \text{ndvi_med_h}) + (-0,740 \times \text{orient_rad}) + (8,967 \times \text{lito1}) + (12,692 \times \text{lito2}) + (-6,897 \times \text{lito3}) \quad (\text{Ec. 14})$$

Al contrario que en el caso del limo, la arcilla presenta los valores de R^2 más bajos, seleccionándose los modelos *c* y *e* (Anexo, Tabla A3), cuyos valores son 0,405 y 0,392, respectivamente. El valor del índice AIC ha sido de 361,278 para ambos modelos, por lo que se seleccionó el modelo *e* (Ec. 15). Al igual que en el caso del limo, el porcentaje de arcilla se explica gracias a los valores de NDVI medios, además de ser importantes las litologías consolidadas y poco consolidadas y los usos forestal, matorral y suelo desnudo.

$$\text{Arcilla} = 33,327 + (-4,502 \times \text{usos4}) + (6,735 \times \text{lito1}) + (5,908 \times \text{lito3}) + (-5,950 \times \text{usos3}) + (0,623 \times \text{orient_rad}) + (-17,826 \times \text{ndvi_med_h}) + (0,631 \times \text{twi}) + (-0,174 \times \text{etp_17_160}) + (5,960 \times \text{lito2}) + (-11,523 \times \text{usos1}) + (10,376 \times \text{ndvi_med_s}) + (-0,203 \times \text{pr_17_160}) \quad (\text{Ec. 15})$$

Tabla 3. Resumen de los resultados obtenidos en la regresión lineal múltiple y el GLM.

Variable dependiente	Modelo	R ²	Sig.	AIC
Materia orgánica	c	0,457	0,015	237,418
	f	0,449	0,003	237,418
Arena	f	0,490	0,001	414,865
	h	0,483	0,000	414,865
Limo	d	0,509	0,002	422,058
	i	0,501	0,000	422,058
Arcilla	c	0,405	0,051	361,278
	e	0,392	0,026	361,278

Una vez realizado el entrenamiento del modelo, se ha ejecutado la verificación con el conjunto de los 20 datos restantes para los modelos seleccionados de cada variable.

Comparando los valores medidos con los valores estimados por el modelo tanto para los 55 muestreos de calibración como para los 20 de validación para cada una de las variables dependientes (Figura 8), se observa que el limo es el que presenta mayor valor de R^2 en el caso de los muestreos de calibración. Sin embargo, la materia orgánica es la que presenta mayor R^2 para los 20 muestreos restantes, además de que presenta valores de R^2 similares para ambos grupos de muestras. Por otro lado, tanto arena como arcilla presentan mayor diferencia entre los R^2 de los muestreos de validación y los de calibración.

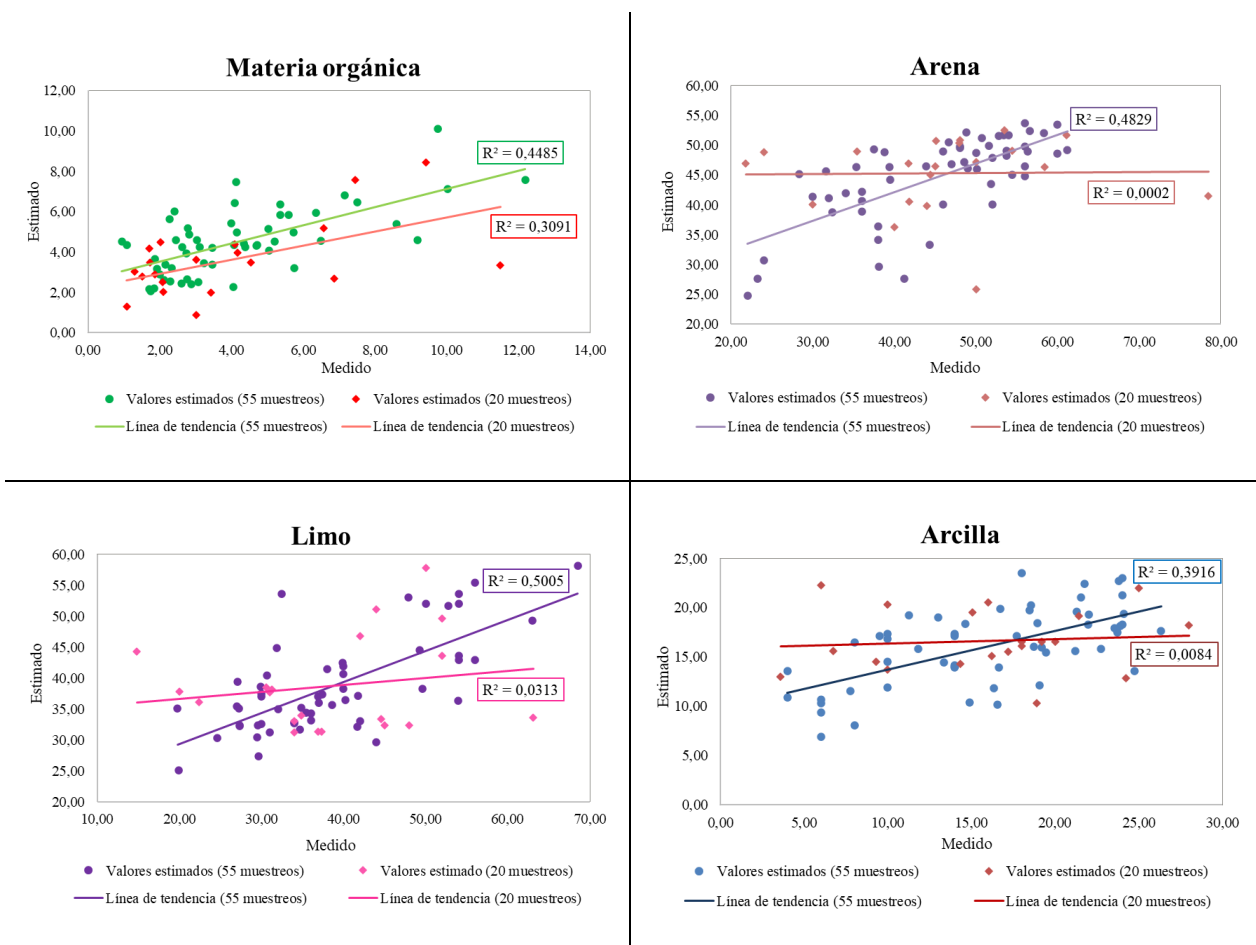


Figura 8. Comparación de los valores reales medidos respecto a los valores estimados por el modelo para cada una de las propiedades estudiadas.

Por otro lado, en la Tabla 4 se recoge el error cuadrático medio obtenido para cada una de las variables dependientes, observándose que no hay grandes diferencias entre el error de los 55 y de los 20 muestreos respecto al total de estos.

Tabla 4. Resumen del error cuadrático medio (RMSE) obtenido para los muestreos de calibración, los de validación y para los totales en las cuatro variables dependientes.

Variable	Error (RMSE)		
	55 muestreos	20 muestreos	75 muestreos
Materia orgánica	1,803	2,412	1,984
Arena	7,308	13,906	9,525
Limo	7,648	12,679	9,261
Arcilla	4,858	6,937	5,490

Además, comparando los valores reales medidos con los valores estimados por el modelo en función de los dos orígenes de datos de los muestreos (Figura 9), se observa que en todos los casos la tendencia es bastante similar, excepto en el caso del limo, en cual presenta ligeras diferencias. Cabe destacar que, de todas las propiedades analizadas, el porcentaje de arcilla es el que presenta valores de R^2 más parecidos entre ambos grupos de datos.

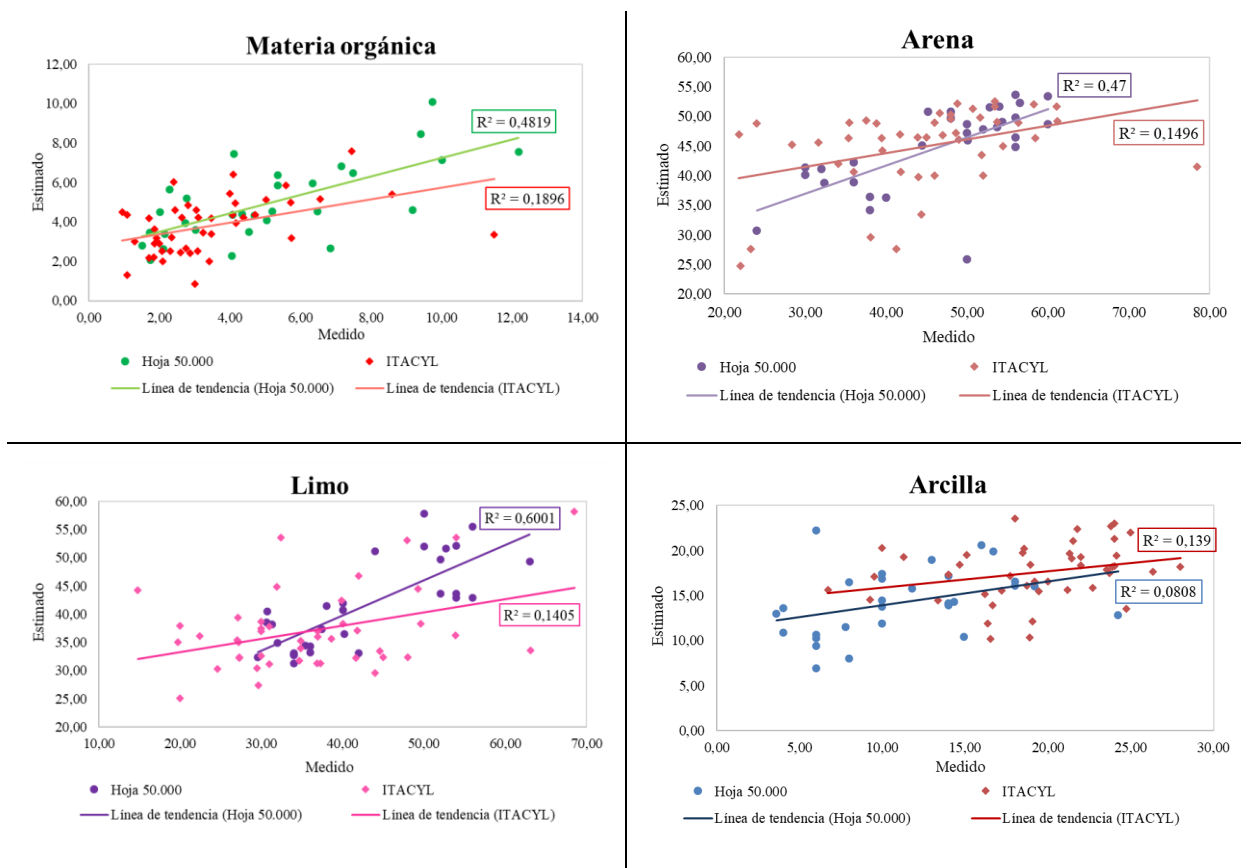


Figura 9. Comparación de los valores medidos respecto a los estimados en función de las dos bases de datos utilizadas para cada una de las propiedades estudiadas.

5.2. Análisis geoestadístico

En cuanto al análisis espacial de los residuos (Anexo, Figura A2), se observa que la arcilla, junto a la arena, son las que presentan una mayor subestimación de los resultados, mientras que en el caso del limo y la materia orgánica hay un dominio de los residuos negativos y, por lo tanto, los resultados se sobreestiman.

Además, cabe destacar que los residuos obtenidos para el total de los 75 muestreos presentan para cada una de las variables estudiadas una distribución normal, tal y como se observa en la Figura A3 del Anexo.

Como resultado final se obtuvieron cuatro mapas correspondientes a cada una de las variables dependientes estudiadas (Figuras 10 y 11). En aquellos casos en que los valores estimados fueron negativos, una vez añadidos los residuos, estos se igualaron a 0. Se observa que los mayores porcentajes de materia orgánica y de limo se encuentran en la zona noroeste de la hoja, coincidiendo con los valores más bajos de arena y arcilla. Estas últimas presentan los valores más altos en las zonas de terrazas ubicadas al este y sureste de la hoja, respectivamente.

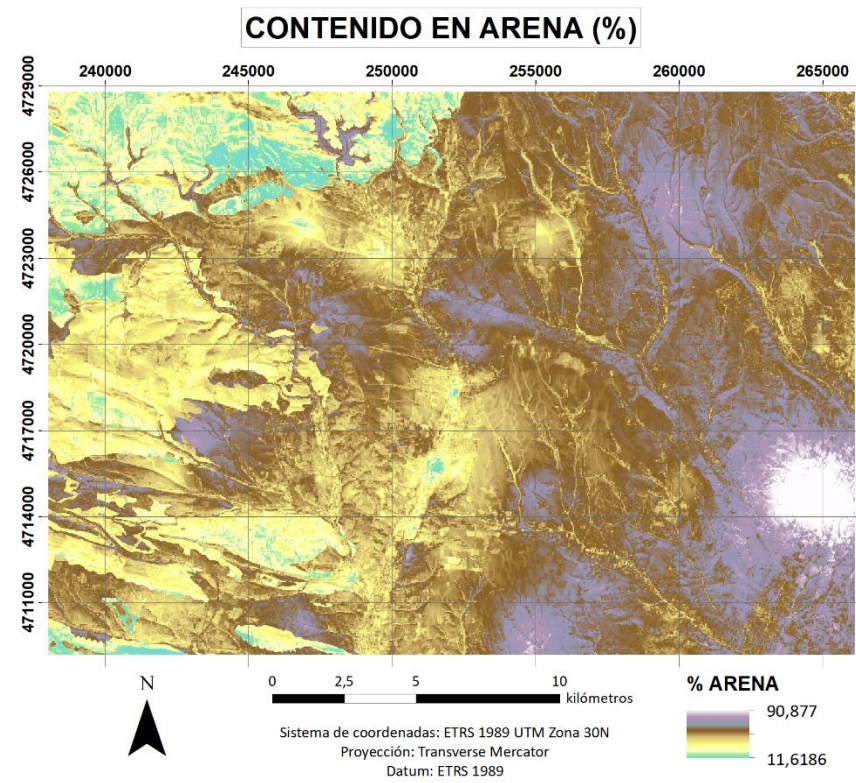
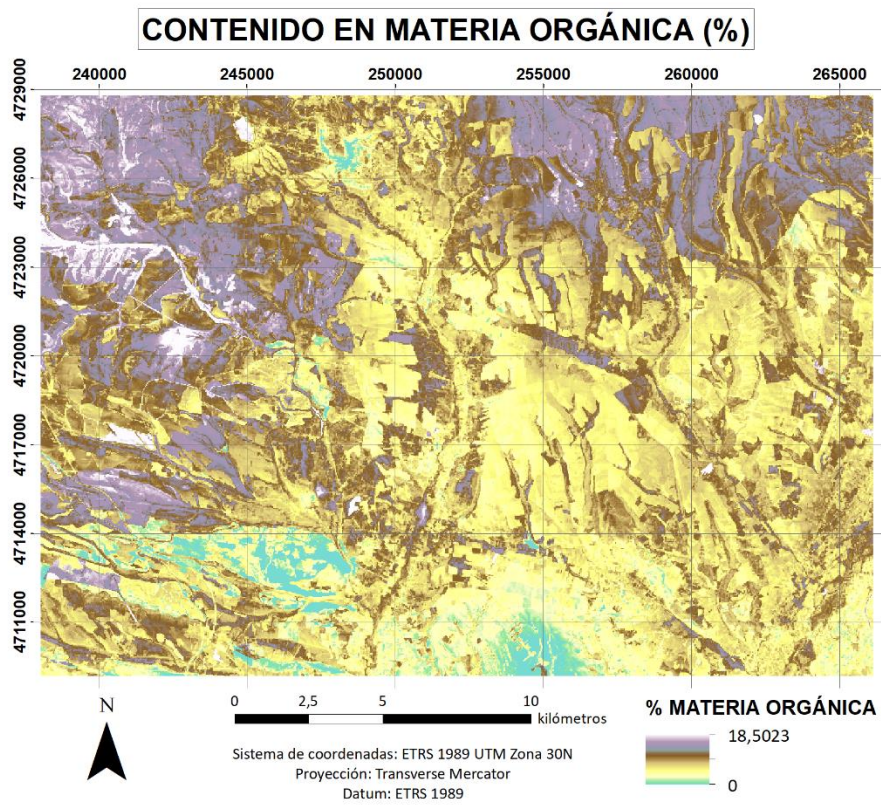


Figura 10. Mapas finales de materia orgánica y arena tras la aplicación de las respectivas ecuaciones resultantes del método.

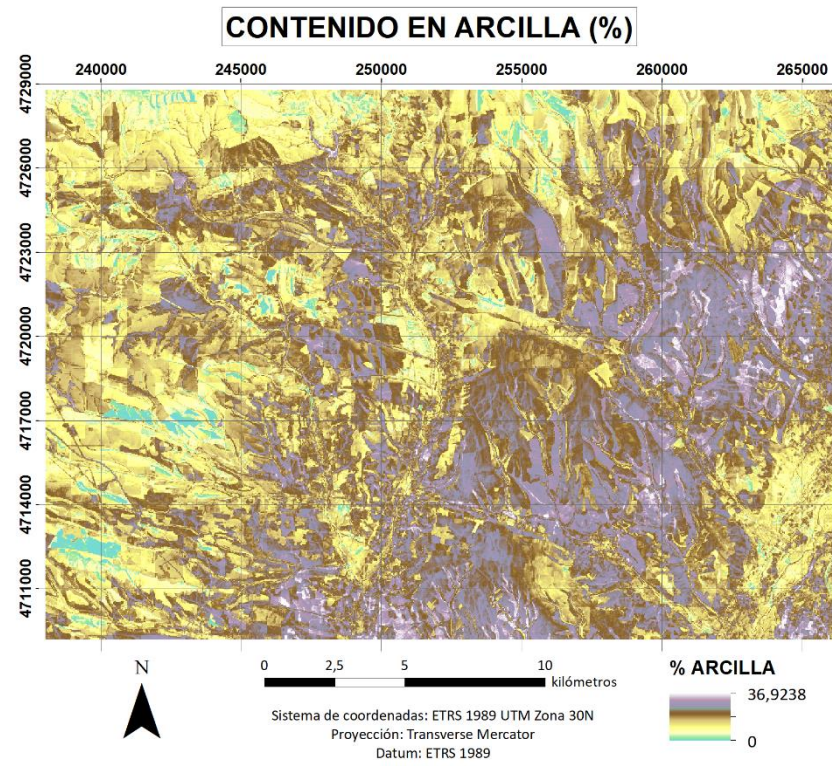
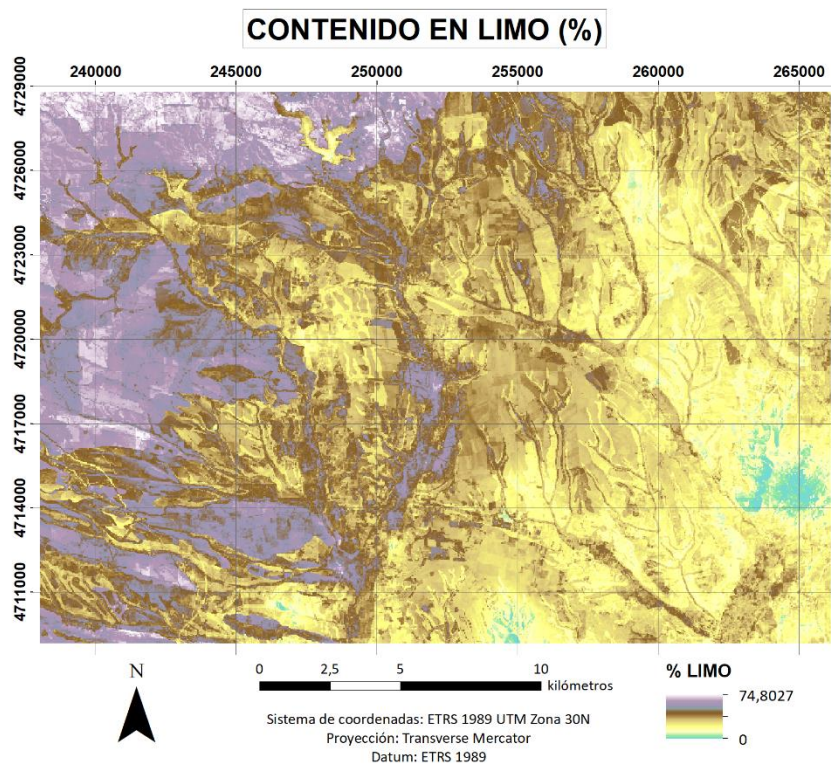


Figura 11. Mapas finales de materia orgánica y arena tras la aplicación de las respectivas ecuaciones resultantes del método.

6. Discusión

Los resultados obtenidos muestran ciertas diferencias respecto a otros trabajos publicados, tanto para los valores de R^2 como para los errores. Aunque los ajustes obtenidos en los modelos de regresión no son muy elevados, se encuentran en los rangos esperados según Hengl *et al.* (2004), Mora-Vallejo *et al.* (2008) y Mosleh *et al.* (2016). Mosleh *et al.* (2016) estudiaron, entre otras propiedades, la arena, el limo y la arcilla utilizando distintos modelos estadísticos, como el GLM y la MLR, obteniendo en esta última valores de R^2 inferiores a 0,3, siendo la arcilla la que mejor resultado presentaba con un R^2 de 0,24. Por el contrario, el GLM mostró mejores resultados, elevando los coeficientes de determinación en todos los casos, alcanzando incluso valores de 0,54 en el caso de la arena y de 0,45 en la arcilla, ligeramente superiores a los obtenidos en este trabajo. Sin embargo, el mejor resultado se obtuvo a través de la utilización de redes neuronales artificiales (ANN), obteniendo para el limo un R^2 de 0,65. En cuanto a los errores, Mosleh *et al.* (2016) obtuvieron valores mucho más elevados para todos los casos respecto a los obtenidos en el presente trabajo.

Mora-Vallejo *et al.* (2008) estudiaron a través de MLR el contenido de arcilla en los suelos del sudeste de Kenia, obteniendo un valor de R^2 de 0,84. Por el contrario, el valor del error obtenido en el trabajo mencionado es superior, tanto para los muestreos de calibración como para los de validación, siendo 6,32 en el primer caso y 9,65 en el segundo.

Para la materia orgánica, Hengl *et al.* (2004) llevaron a cabo un estudio en la región central de Croacia utilizando varios métodos estadísticos, entre ellos la regresión múltiple. En este caso, obtuvieron un valor de R^2 para la materia orgánica de 0,3, inferior al que se presenta en este trabajo. En el caso del error, presentan errores superiores en todos los casos, siendo el más bajo el obtenido en la regresión múltiple (3,4).

Los valores de R^2 obtenidos podrían deberse a las diferencias entre las distintas técnicas de muestreo empleadas para cada uno de los grupos de muestreos del suelo utilizados al proceder de dos fuentes diferentes. Sin embargo, se ha observado que el distinto origen de estos datos no ha afectado a los resultados obtenidos. A pesar de ello, es importante tener en cuenta que las covariables utilizadas y la gran variabilidad espacial del suelo pueden dar lugar a errores (McBratney *et al.*, 2003). Cabe destacar que el empleo de covariables más detalladas o covariables a distinta resolución puede mostrar ligeras mejoras en cuanto a los resultados (Taylor *et al.*, 2013; Samuel-Rosa *et al.*, 2015).

Por otro lado, la mayoría de estudios presentan el *kriging* como método de interpolación, al contrario que el presente trabajo en el que se utiliza el método IDW. Sin embargo, en algunos trabajos se han comparado ambas técnicas para la elaboración de cartografía edáfica de distintas propiedades y no se han observado grandes diferencias (Gozdowski *et al.*, 2015).

En cuanto a la división de los muestreos en dos grupos para validar y para calibrar, tanto en los resultados obtenidos (Tabla 4) como en los otros tres mencionados anteriormente, los cuales también presentan esta forma de trabajo (Hengl *et al.*, 2004; Mora-Vallejo *et al.*, 2008; Mosleh *et al.*, 2016), se observa que no hay grandes diferencias en los errores, por lo que se corrobora la utilidad de esta segmentación del conjunto de datos.

7. Conclusiones

Tras la discusión de los resultados obtenidos en el presente trabajo para la elaboración de una cartografía digital de distintas propiedades edáficas, se han obtenido las siguientes conclusiones:

- Los resultados obtenidos en los modelos seleccionados son mejorables, pues se han hallado valores de R^2 bastante bajos. Sin embargo, otros estudios, aunque han obtenido R^2 superiores, presentan errores elevados en comparación, poniendo de manifiesto menor robustez en la modelación con los datos.
- La segmentación de los datos correspondientes a las muestras de suelo originales presenta resultados óptimos, permitiendo realizar una correcta validación del método.
- El origen de los muestreos del suelo no es un aspecto que haya afectado a los resultados obtenidos, por lo que se debe estudiar qué otros aspectos pueden afectar a los resultados.
- No existe un único método estadístico exclusivo para la metodología DSM que presente resultados óptimos, de forma que, para cada caso de estudio es necesario analizar qué método estadístico es el más adecuado.
- El muestreo convencional, aunque es complejo y costoso, sigue siendo necesario, pues es un elemento clave en esta metodología y, debido a la variabilidad del suelo, es importante aumentar el número de muestreos con el objetivo de disminuir el error y obtener mejores resultados.
- El método presenta grandes limitaciones a pesar de su aplicabilidad, tales como el elevado número de variables, los errores asociados a estas variables y a la falta de homogeneidad en el análisis estadístico.

8. Referencias bibliográficas

Ågren, A. M., Lidberg, W., Strömberg, M., Ogilvie, J. y Arp, P. A. (2014). *Evaluating digital terrain indices for soil wetness mapping--a Swedish case study*. Hydrology and Earth System Sciences, 18(9), pp. 3623 - 3634.

Alonso Herrero, E. (1995). *Litología y Geomorfología*. En Gallego Valcarce, A., Alonso Herrero, E. y Penas Merino, A. (Coords.). *Atlas del Medio Natural de la provincia de León*. Instituto Tecnológico Geominero de España (ITGE) y Diputación de León, Madrid. 11-18 pp.

Alonso Herrero, E., Vidal Bardán, M., Villa Bermejo, J. E., Fuertes Gutiérrez, I., Alcalde Aparicio, S., Robles Llamazares, J. A., Rodríguez Blanco, J. y Juan González, L. A. (2010) *Plan Piloto del Mapa de Suelos de la Provincia de León. Escala 1:50.000. Memoria Hoja 160 (Benavides)*. Universidad de León: Área de publicaciones.

Armas, D., Guevara, M., Alcaraz-Segura, D., Vargas, R., Soriano-Luna, Ángeles, Durante, P. y Oyonarte, C. (2017). *Mapa digital del perfil del carbono orgánico en los suelos de Andalucía, España*. Ecosistemas, 26(3), pp. 80 - 88.

Carré, F., McBratney, A. B. y Minasny, B. (2007a). *Estimation and potential improvement of the quality of legacy soil samples for digital soil mapping*. Geoderma, 141(1-2), pp. 1-14.

Carré, F., McBratney, A. B., Mayr, T. y Montanarella, L. (2007b). *Digital soil assessments: Beyond DSM*. Geoderma, 142(1-2), pp. 69-79.

Chazarra Bernabé, A., Flórez García, E., Peraza Sánchez, B., Tohá Rebull, T., Lorenzo Mariño, B., Criado Pinto, E., Moreno García, J. V., Romero Fresneda, R. y Botey Fullat, R. (2018). *Mapas climáticos de España (1981-2010) y ETo (1996-2016)*. Madrid: Ministerio para la Transición Ecológica y Agencia Estatal de Meteorología.

Dávila, F., Pereira, G. y Fontes, F. (2018). *“Mapeo digital de carbono orgánico del suelo.”* Disponible en: <http://inia.uy/Documentos/P%C3%BABlicos/INIA%20La%20Estanzuela/Actividades%202018/2%20de%20octubre%20secuestro%20de%20carbono/taller%20nacional%20sobre%20secuestro%20de%20carbono%20%20de%20octubre%20del%202018%20F.%20Davila,%20G.%20Pereira%20y%20F.%20Fontes.pdf> (Accedido: 8 de enero de 2021).

De León Llamazares, A., Arriba Balenciaga, A. y De La Plaza, M.C. (1991). *Caracterización agroclimática de la provincia de León*. 2.^a ed. Madrid: Ministerio de Agricultura, Pesca y Alimentación.

Doblas-Miranda, E., Rovira, P., Brotons, L., Martínez-Vilalta, J., Retana, J., Pla, M. y Vayreda, J. (2013). *Soil carbon stocks and their variability across the forests, shrublands and grasslands of peninsular Spain*. Biogeosciences, 10(12), pp. 8353-8361.

Esri (2015). *How inverse distance weighted interpolation works*. Disponible en: <https://pro.arcgis.com/es/pro-app/latest/help/analysis/geostatistical-analyst/how-inverse-distance-weighted-interpolation-works.htm> (Accedido: 14 de junio de 2021).

Environmental Systems Research Institute (Esri) (2020). *ArcGIS Desktop (Versión 10.8) [Programa de ordenador]*. Redlands, CA: Environmental Systems Research Institute, Inc.

European Space Agency (2021). *Sentinel-2 MSI Introduction*. Disponible en: <https://dragon3.esa.int/web/sentinel/user-guides/sentinel-2-msi> (Accedido: 12 de enero de 2021).

Gobierno del Principado de Asturias (2018). *Cartografía de los Suelos de Asturias*. Disponible en: https://sig.asturias.es/HIPERVINCULOS/Catalogo_Suelos/mapas.htm (Accedido: 14 de junio de 2021).

González Menéndez, L., Heredia, N. y Marcos, A. (2008). Mapa Geológico Digital continuo E. 1:50.000, Zona Asturoccidental - Leonesa (Zona 1100). Disponible en: <http://info.igme.es/cartografiadigital/geologica/geodezona.aspx?Id=Z1100> (Accedido: 20 de enero de 2021).

Gozdowski, D., Stępień, M., Samborski, S., Dobers, E. S., Szatyłowicz, J. y Chormański, J. (2015). *Prediction accuracy of selected spatial interpolation methods for soil texture at farm field scale*. *Journal of soil science and plant nutrition*, 15(3), pp. 639-650.

Grunwald, S. (2009). *Multi-criteria characterization of recent digital soil mapping and modeling approaches*. *Geoderma*, 152(3-4), pp. 195-207.

Grunwald, S., Thompson, J. A. y Boettinger, J. L. (2011). *Digital soil mapping and modeling at continental scales: Finding solutions for global issues*. *Soil Science Society of America Journal*, 75(4), pp. 1201-1213.

Hengl, T., Heuvelink, G. B. y Stein, A. (2004). *A generic framework for spatial prediction of soil variables based on regression-kriging*. *Geoderma*, 120(1-2), pp. 75-93.

Hengl, T., Heuvelink, G. B. y Rossiter, D. G. (2007). *About regression-kriging: From equations to case studies*. *Computers & geosciences*, 33(10), pp. 1301-1315.

IBM Corporation (2019). *IBM SPSS Statistics (Versión 26.0.) [Programa de ordenador]*. Armonk, NY: IBM Corporation.

IBM Corporation (2021). *Métodos de selección de variables en el análisis de regresión lineal*. Disponible en: <https://www.ibm.com/docs/es/spss-statistics/version-missing?topic=regression-linear-variable-selection-methods> (Accedido: 9 de junio de 2021)

Instituto Geográfico Nacional (2014). *Centro de Descargas*. Disponible en: <http://centrodedescargas.cnig.es/CentroDescargas/> (Accedido: 12 de enero de 2021).

Instituto Geográfico Nacional (sin fecha). *Mapas edafológicos en España*. Disponible en: <https://www.ign.es/web/catalogo-cartoteca/apibadasid/cartoteca/searchAuthority/T650&18889> (Accedido: 14 de junio de 2021).

Instituto Tecnológico Agrario de Castilla y León (2017). *Zonas de Castilla y León con cartografía semidetallada de suelos*. Disponible en: http://suelos.itacyl.es/documents/426128/428516/Cartogra_DDOO_y_2HojasSA_y_2HojasLE_900x637.jpg (Accedido: 14 de junio de 2021).

Instituto Tecnológico Agrario de Castilla y León (2012). *Base de datos*. Disponible en: http://suelos.itacyl.es/base_datos (Accedido: 12 de enero de 2021).

Jenny, H. (1941). *Factors of soil formation: a System of Quantitative Pedology*. Nueva York: McGraw-Hill.

Lagacherie, P. y McBratney, A. B. (2006). *Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping*. *Developments in soil science*, 31, pp. 3-22.

Ma, Y., Minasny, B., Malone, B. P. y McBratney, A. B. (2019). *Pedology and digital soil mapping (DSM)*. *European Journal of Soil Science*, 70(2), pp. 216-235.

Martínez, D. R., Albín, J. L., Cabaleiro, J. C., Pena, T. F., Rivera, F. F. y Blanco, V. (2009). *El Criterio de Información de Akaike en la Obtención de Modelos Estadísticos de Rendimiento*. *Jornadas de Paralelismo*, (10), pp. 439-444.

Mattivi, P., Franci, F., Lambertini, A. y Bitelli, G. (2019). *TWI computation: a comparison of different open source GISs*. *Open Geospatial Data, Software and Standards*, 4(6), pp. 1-12.

McBratney, A. B., Santos, M. M. y Minasny, B. (2003). *On digital soil mapping*. *Geoderma*, 117(1-2), pp. 3-52.

Minasny, B. y McBratney, A. B. (2016). *Digital soil mapping: A brief history and some lessons*. *Geoderma*, 264, pp. 301-311.

Ministerio para la Transición Ecológica y el Reto Demográfico (2019) *Modelo SIMPA 2019*. Disponible en: <https://www.miteco.gob.es/es/agua/temas/evaluacion-de-los-recursos-hidricos/evaluacion-recursos-hidricos-regimen-natural/> (Accedido: 8 de enero de 2021).

Mora-Vallejo, A., Claessens, L., Stoorvogel, J. y Heuvelink, G. B. (2008). *Small scale digital soil mapping in Southeastern Kenya*. *Catena*, 76(1), pp. 44-53.

Mosleh, Z., Salehi, M. H., Jafari, A., Borujeni, I. E. y Mehnatkesh, A. (2016). *The effectiveness of digital soil mapping to predict soil properties over low-relief areas*. *Environmental monitoring and assessment*, 188(3), pp. 195.

Myneni, R. B., Hall, F. G., Sellers, P. J. y Marshak, A. L. (1995). *The interpretation of spectral vegetation indexes*. *IEEE Transactions on Geoscience and Remote Sensing*, 33(2), pp. 481 - 486.

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J. M., Tucker, C. J. y Stenseth, N. C. (2005). *Using the satellite-derived NDVI to assess ecological responses to environmental change*. Trends in ecology and evolution, 20(9), pp. 503-510.

Rial Tubío, M. (2017). *Modelling of the topsoil organic carbon content by analysing the potential of spectroscopic techniques for Digital Soil Mapping*. Tesis doctoral. Universidad de Santiago de Compostela.

Rodríguez Fernández, L.R. y Pérez González, A. (coord.) (1984). *Mapa Geológico de España. Escala 1:50.000. Benavides*. Segunda serie. 1.ª ed. Madrid: Servicio de Publicaciones del Ministerio de Industria y Energía.

Ruhoff, A. L., Castro, N. M. R. y Risso, A. (2011). *Numerical modelling of the topographic wetness index: an analysis at different scales*. International Journal of Geosciences, 2, pp. 476 - 483.

Samuel-Rosa, A., Heuvelink, G. B. M., Vasques, G. M. y Anjos, L. H. C. (2015). *Do more detailed environmental covariates deliver more accurate soil maps?*. Geoderma, 243, pp. 214-227

Schmidt, F. y Persson, A. (2003). *Comparison of DEM data capture and topographic wetness indices*. Precision Agriculture, 4(2), pp. 179-192.

Sørensen, R., Zinko, U. y Seibert, J. (2006). *On the calculation of the topographic wetness index: evaluation of different methods based on field observations*. Hydrology and Earth System Sciences, 10 (1), pp. 101 – 112.

Taylor, J. A., Jacob, F., Galleguillos, M., Prevot, L., Guix, N. y Lagacherie, P. (2013). *The utility of remotely-sensed vegetative and terrain covariates at different spatial resolutions in modelling soil and watertable depth (for digital soil mapping)*. Geoderma, 193, pp. 83-93.

ANEXO

Tabla A1. Simplificación de la litología de la zona de estudio.

Litología	Simplificación	Código
Cuarcitas, pizarras, areniscas y limolitas (Serie de los Cabos); Cuarcitas y pizarras (Capas superiores del Eo)	Litologías consolidadas silíceas de grano grueso con alteración con aportes finos	1
Pizarras negras y grises (Fm, Luarca); Pizarras y areniscas (Fm. Agüeira); Pizarras, limolitas y areniscas (Fm. Agüeira, miembro inferior); Areniscas, limolitas y pizarras (Fm. Agüeira, miembro medio); Pizarras grises, ampelitas y liditas (Capas de la Gargantera)	Litologías consolidadas silíceas de grano fino con alteración con aportes finos	2
Terciario indiferenciado: arenas, arcillas, limos y niveles de conglomerados; Gravas y arenas; Limos, arenas y arcillas, con capas de conglomerados polimícticos, abanicos polimícticos; Arenas, limos arenosos; Limos arcillosos y arcillas verdes (paleosuelos); Ortoconglomerados de cuarcita y cuarzo, con intercalaciones delgadas de arenas; Paraconglomerados de pizarras y quiastolita, y fangos rojos (Sistema de Veguellina); Ortoconglomerados de cuarcita, con intercalaciones delgadas de arenas (Sistema de Villagatón); Orto y paraconglomerados de clastos de cuarcita, cuarzo y pizarra (Sistema de Combarros-Brazuela); Ortoconglomerados de cuarcitas (Sistema de Vanidodes); Paraconglomerados de cuarcitas (Sistema de Veguellina); Conglomerados de cuarcita y cuarzo, pizarra y minerales de hierro (Sistema de Vanidodes-Combarros); Conglomerados, limos y arcillas	Litologías poco consolidadas de materiales miocénicos silíceos	3
Depósitos aluviales; Terrazas bajas; Terrazas medias; Glacis; Coluviones; Gravas cuarcíticas, y a veces bloques, en matriz arcillo-arenosa, encostramiento, a veces (Raña); Cantos, gravas, arenas, limos, arcillas y a veces bloques (Glacis asociados a terrazas altas); Cantos, gravas, arenas, limos, arcillas y a veces bloques (Glacis asociados a terrazas medias); Cantos, gravas, arenas, limos, arcillas y a veces bloques (Glacis); Cantos, bloques, arenas, limos y arcillas (Coluviones); Cantos, arenas, limos (Aluvial-coluvial); Gravas, cantos, arenas y limos (frecuentemente encostrados). (Abanicos asociados a terrazas altas); Gravas, cantos, arenas y limos (Abanicos); Gravas, cantos, arenas y a veces arcillas (Frecuentemente encostradas) (Terrazas altas); Gravas, cantos, arenas y a veces arcillas (A veces encostradas) (Terrazas medias); Gravas, cantos, arenas y a veces arcillas (Terrazas bajas); Cantos, gravas, arenas, limos y arcillas (Fondos de valle); Limos, arenas, gravas y arcillas (Llanura de inundación); Gravas, arenas y limos (Aluvial)	Litologías sin consolidar de depósitos cuaternarios recientes	4

Tabla A2. Correlaciones entre las distintas variables derivadas del NDVI (*: $p < 0,05$; **: $p < 0,01$).

	<i>ndvi_min_s</i>	<i>ndvi_min_h</i>	<i>ndvi_med_s</i>	<i>ndvi_med_h</i>	<i>ndvi_max_s</i>	<i>ndvi_max_h</i>
<i>ndvi_min_s</i>	1					
<i>ndvi_min_h</i>	0,717 **	1				
<i>ndvi_med_s</i>	0,877 **	0,566 **	1			
<i>ndvi_med_h</i>	0,772 **	0,870 **	0,658 **	1		
<i>ndvi_max_s</i>	0,734 **	0,415 **	0,942 **	0,543 **	1	
<i>ndvi_max_h</i>	0,420 **	0,399 **	0,410 **	0,744 **	0,409 **	1

Tabla A3. Covariables correspondientes a cada uno de los modelos seleccionados.

Variable dependiente	Modelo	Covariables													
		usos4	altitud	lito1	usos3	orient_rad	ndvi_med_h	twi	etp_17_160	lito2	usos1	ndvi_med_s	pr_17_160	pend	tem_17_160
Materia orgánica	c	usos4	altitud	lito1	usos3	orient_rad	ndvi_med_h	twi	etp_17_160	lito2	usos1	ndvi_med_s	pr_17_160	pend	tem_17_160
	f	usos4	altitud	lito1	usos3	orient_rad	ndvi_med_h	lito2	usos1	pr_17_160	pend	tem_17_160			
Arena	f	usos4	lito1	usos3	twi	rad_s_kw	lito2	usos1	ndvi_med_s	pr_17_160	pend	tem_17_160			
	h	usos4	lito1	usos3	twi	rad_s_kw	lito2	ndvi_med_s	pr_17_160	tem_17_160					
Limo	d	altitud	lito1	lito3	usos3	orient_rad	ndvi_med_h	etp_17_160	lito2	usos1	ndvi_med_s	pr_17_160	pend	tem_17_160	
	i	lito1	lito3	orient_rad	ndvi_med_h	etp_17_160	lito2	ndvi_med_s	pr_17_160						
Arcilla	c	usos4	altitud	lito1	lito3	usos3	orient_rad	ndvi_med_h	twi	rad_s_kw	etp_17_160	lito2	usos1	ndvi_med_s	pr_17_160
	e	usos4	lito1	lito3	usos3	orient_rad	ndvi_med_h	twi	etp_17_160	lito2	usos1	ndvi_med_s	pr_17_160		

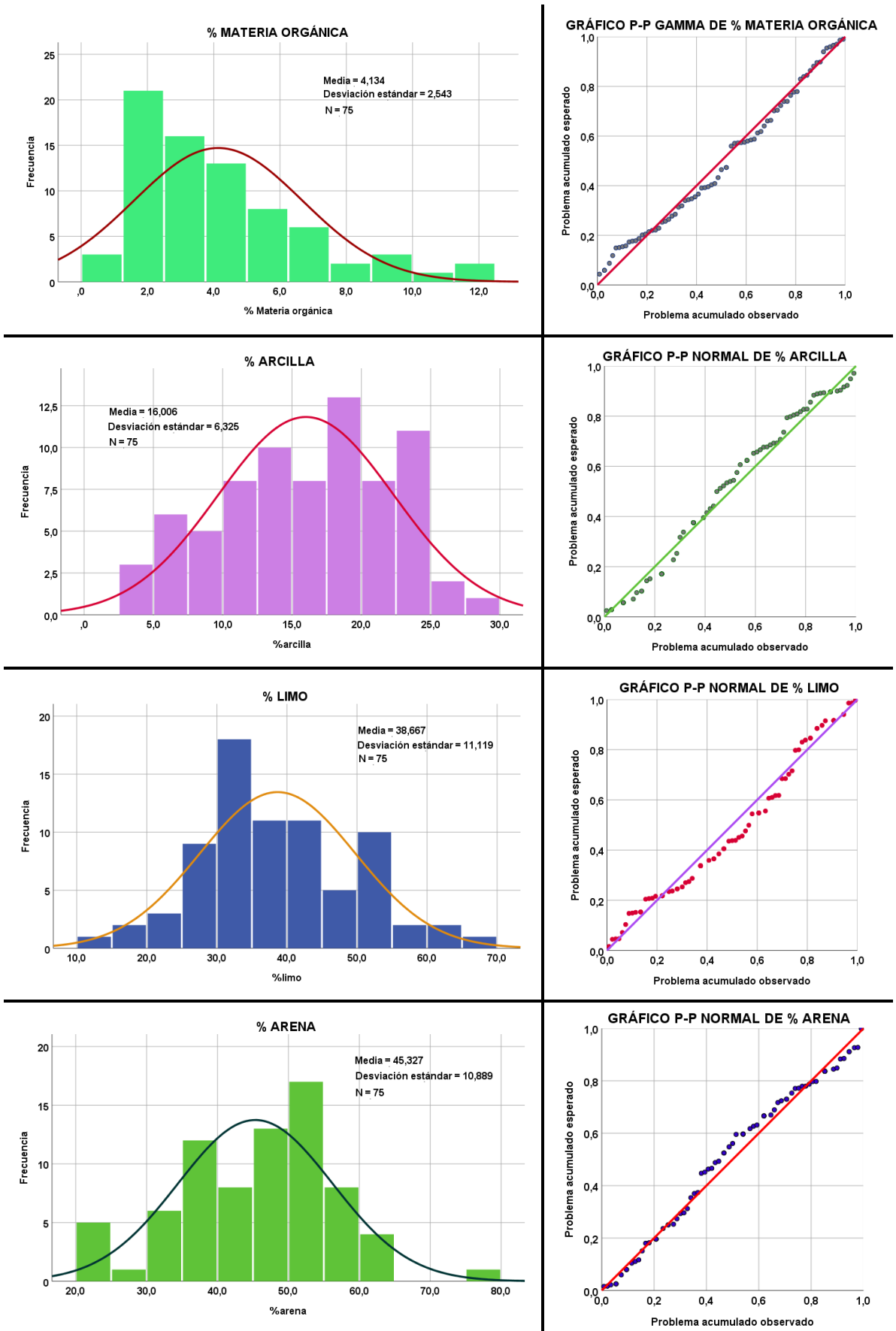


Figura A1. Tipo de distribución y gráficos P-P para cada una de las variables dependientes.

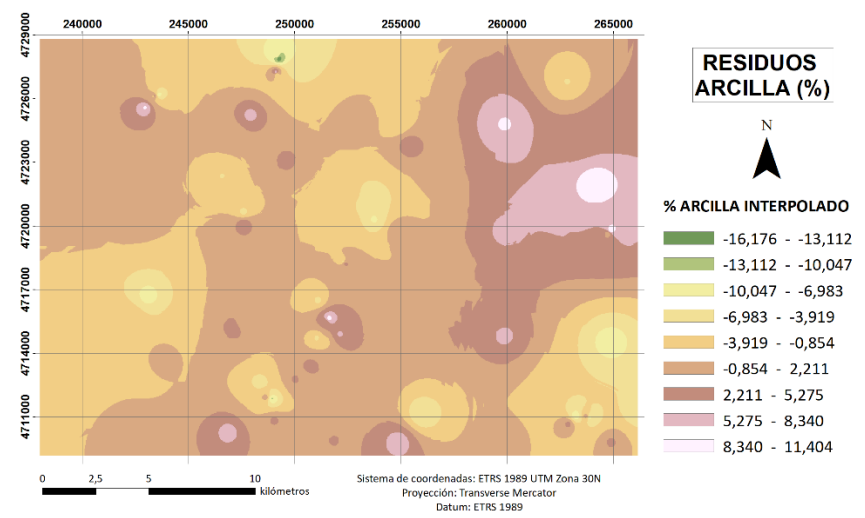
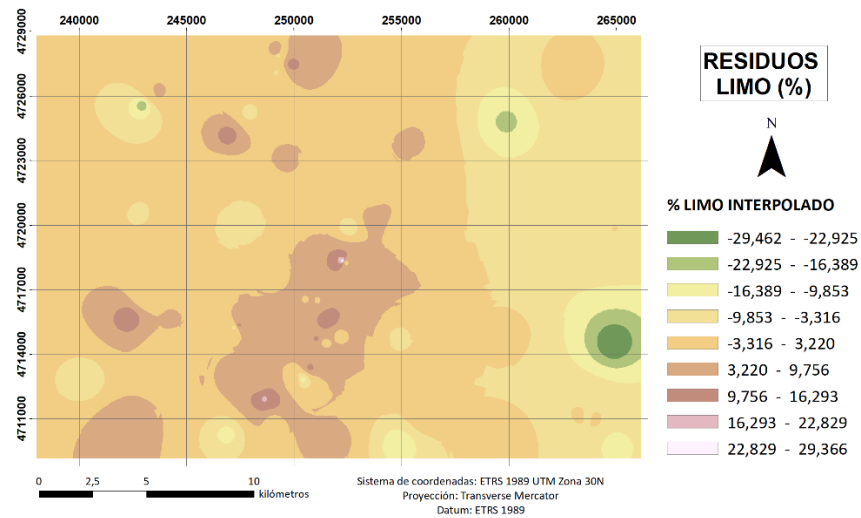
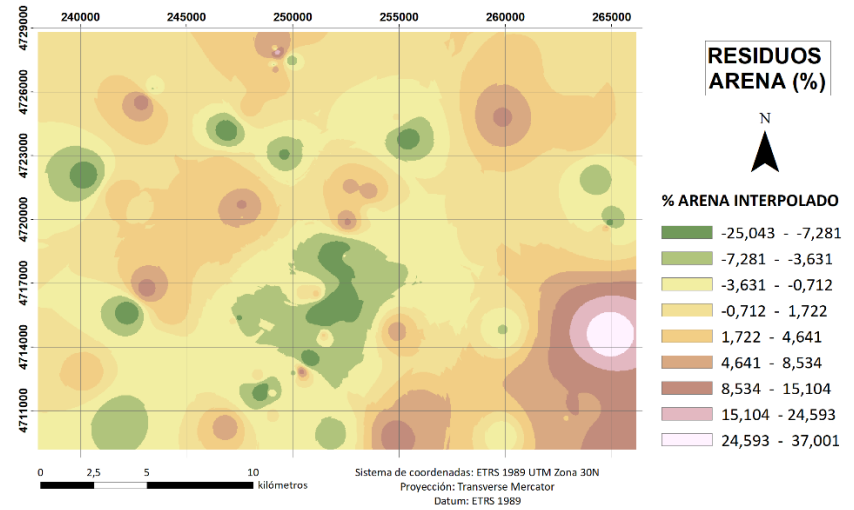
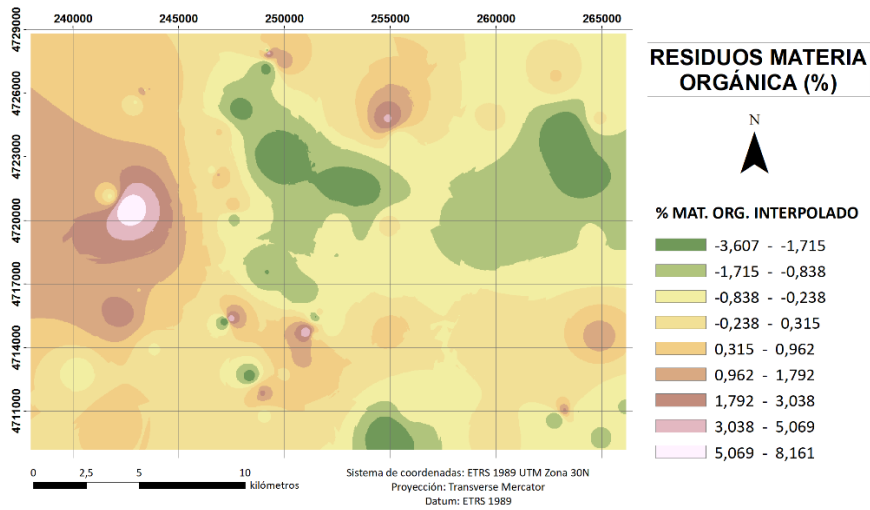


Figura A2. Mapas de los residuos interpolados para las cuatro variables dependientes estudiadas (% materia orgánica, % arena, % limo y % arcilla).

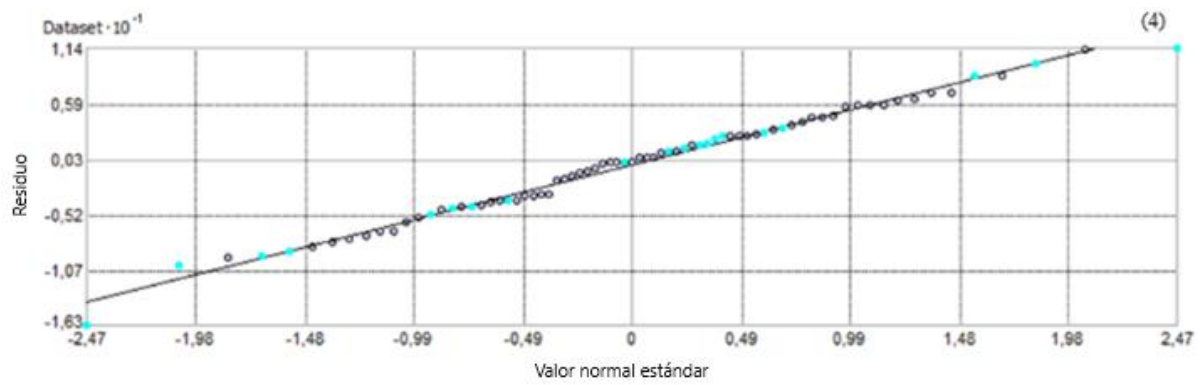
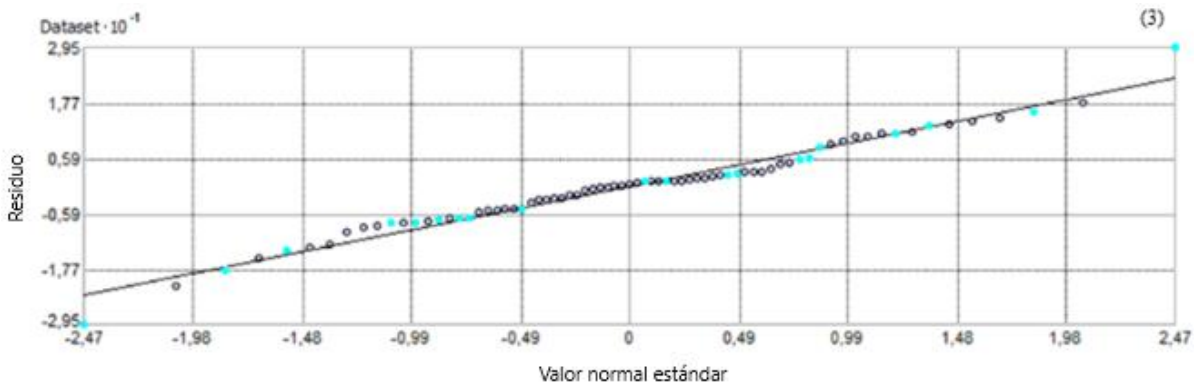
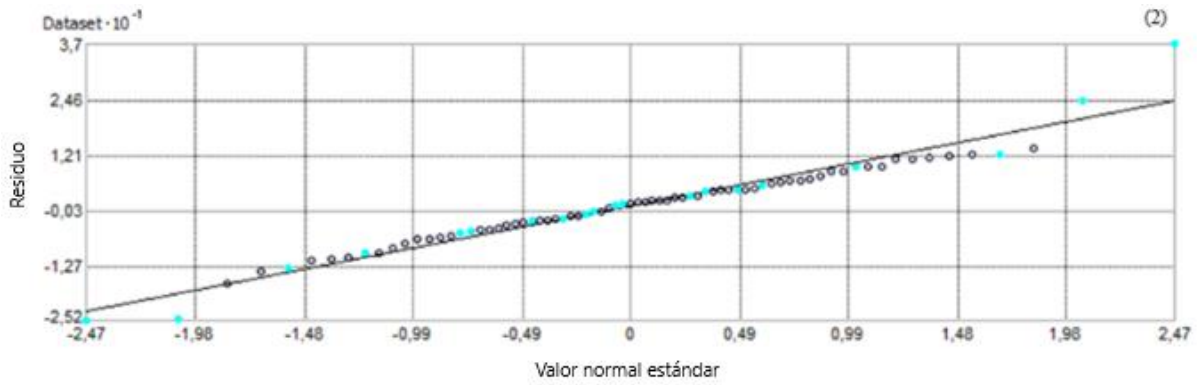
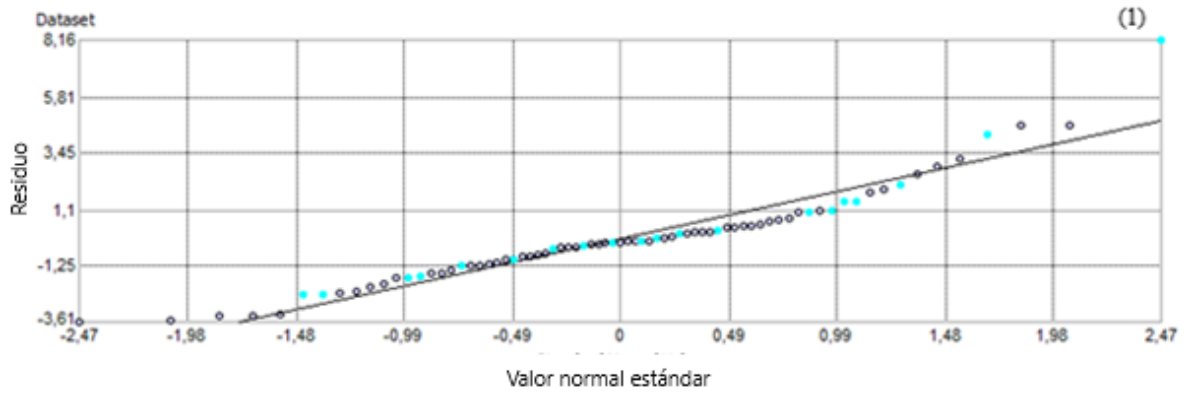


Figura A3. Gráficos de normalidad Q-Q de los residuos de los 75 muestreos (en azul, muestreos de calibración; en blanco, muestreos de validación) para la materia orgánica (1), arena (2), limo (3) y arcilla (4).