

## Mapeo de Materia Orgánica a escala de lote utilizando técnicas de aprendizaje automático

García Seleme, F.A.<sup>12</sup>[0009-0003-0740-4761], Paccioretti, P.<sup>13</sup>[0000-0002-1267-6361], Balzarini, M.<sup>12</sup>[0000-0002-4858-4637] y Córdoba, M.A.<sup>12</sup>[0000-0002-4761-9623]

<sup>1</sup> Universidad Nacional de Córdoba. Facultad de Ciencias Agropecuarias. Cátedra de Estadística y Biometría. Córdoba, Argentina.

<sup>2</sup> Grupo de Estadística Vinculado. Unidad de Fitopatología y Modelización Agrícola. Consejo Nacional de Investigaciones Científicas y Técnicas (UFyMA-CONICET). Córdoba, Argentina.

<sup>3</sup> Comisión Nacional de Actividades Espaciales. Universidad Nacional de Córdoba. Instituto de Altos Estudios Espaciales Mario Gulich. Córdoba, Argentina.

`fabrizio.garcia.seleme@mi.unc.edu.ar`

**Resumen.** En el marco de la agricultura digital, la disponibilidad de mapas de propiedades del suelo, como del contenido de materia orgánica (MO), facilita la toma de decisiones en el manejo agrícola. Para generar estos mapas, pueden emplearse diversas técnicas de interpolación espacial orientadas al mapeo de variables edáficas a escala fina. El objetivo del presente trabajo fue comparar la capacidad predictiva de modelos de aprendizaje automático para el mapeo de MO intralote. Se evaluaron dos modelos, árboles de regresión cuantílica (QRF), y redes neuronales artificiales (ANN), utilizando como covariables datos intensivos de conductividad eléctrica aparente y rendimiento de cultivos obtenidos mediante monitores de rendimiento. Como método de referencia se empleó Kriging regresión (RK). Las predicciones espaciales se realizaron en siete lotes ubicados en las provincias de Córdoba y Santiago del Estero, Argentina. Los ajustes de los modelos se realizaron teniendo en cuenta la información de todos los lotes (modelo global) o de manera individual para cada lote (modelo local). En todos los casos los modelos ANN presentaron el mejor desempeño, mostrando los menores valores de error cuadrático medio de predicción. La cantidad de observaciones por lote, así como la variabilidad conjunta del contenido de MO y las covariables, influyeron en el desempeño de modelos globales y locales. Los modelos ANN se consolidan como una alternativa promisorio para el mapeo de la variabilidad espacial del suelo a escala de lote agrícola.

**Palabras clave:** Interpolación Espacial, Redes Neuronales, Agricultura de Precisión.

## Field-scale Soil Organic Matter Prediction: Evaluating Machine Learning and Deep Learning Models

**Abstract.** Within the framework of agriculture, soil features maps availability, as for example those of soil organic matter (SOM) facilitates decision-making in agricultural management. Many techniques have been developed to generate these field scale maps. The objective of this study was to compare the prediction capability of machine learning models for field scale SOM mapping. Quantile Regression Forests (QRF) and Artificial Neural Networks (ANN) were evaluated using soil features and crop yield as covariates. Regression Kriging was used as reference method. Spatial interpolation was performed in seven fields located in Córdoba and Santiago del Estero provinces, Argentina. Models were fitted using information from all fields (global models) or only from the target field (local models). For all fields, ANN models presented better fit, showing lower normalized root mean squared errors. Sample size per field, as well as the joint variability of SOM and covariates affected the performance of global versus local models. ANN models stand as a promissory option for field scale soil variability mapping.

**Keywords:** Spatial Interpolation, Neural Networks, Precision Agriculture

### 1 Introducción

El contenido de materia orgánica en el suelo (MO) desempeña un papel fundamental en los sistemas agrícolas y se considera un indicador clave de la fertilidad edáfica (Tieszen et al, 1994). A escala de lote agrícola, el mapeo de MO constituye una herramienta estratégica para identificar áreas degradadas y optimizar las medidas agroambientales (Piccini et al. 2014), favoreciendo la implementación de técnicas de manejo sitio-específico que permiten mejorar la eficiencia en el uso de insumos y reducir los costos de producción (Vázquez, 2016). Tradicionalmente la obtención de mapas de MO se ha realizado mediante el uso de técnicas de interpolación espacial aplicadas a datos obtenidos mediante muestreo de suelo. Sin embargo, este enfoque requiere una gran cantidad de puntos de muestreo para generar mapas confiables y de alta calidad. El alto costo asociado a la recolección intensiva de datos ha impulsado la necesidad de desarrollar métodos para inferir la distribución espacial de propiedades del suelo como la MO (García Tomillo et al., 2017). Estos métodos pueden incorporar variables auxiliares (covariables) que mejoran la capacidad predictiva y reducen la cantidad de sitios de muestreo requeridos para la obtención de un mapa confiable (Odeh et al. 1995). La relación entre la MO y variables edáficas como la conductividad eléctrica, ha sido ampliamente documentada (Omonode & Vyn, 2006; Kweon et al., 2013). Asimismo, se ha demostrado su relación con otras variables como el rendimiento de cultivos, variables climáticas e índices de vegetación (Oldfield et al, 2022, Bashie et al., 2024). En el marco de la agricultura de precisión, la disponibilidad de grandes volúmenes de datos

para un mismo lote se ha incrementado significativamente gracias al desarrollo y difusión de equipamiento como monitores de cosecha y sensores capaces de recopilar datos en tiempo real (Adamchuk et al., 2004).

Existen diversos métodos de interpolación para predecir valores en sitios no muestreados. Entre ellos, el método geoestadístico kriging es una de las técnicas de interpolación espacial más utilizadas en la agricultura de precisión (Oliver, 2010). Sin embargo, su aplicación requiere cumplir con ciertos supuestos estadísticos poco flexibles y disponer de un número suficiente de muestras ( $>100$ ) para llevar a cabo un modelado confiable de la semivarianza (Oliver & Webster, 2015), lo cual no siempre resulta viable en la práctica. La incorporación de covariables en el modelado ha permitido flexibilizar los métodos de interpolación geoestadística mediante su integración con modelos de regresión lineal, en un procedimiento conocido como Kriging Regresión (RK) (Odeh et al., 1995). En este enfoque, primero se ajusta un modelo lineal entre la variable respuesta y las covariables; luego se modela la estructura espacial de los residuos mediante un semivariograma, y finalmente se genera la predicción como la suma de la predicción del modelo de regresión lineal y la predicción de los residuos obtenida mediante kriging. La integración de RK con técnicas de aprendizaje automático (ML) permitió expandir el alcance de esta herramienta, combinando las fortalezas de ambos enfoques (Mohammadpour et al., 2024). Entre sus ventajas, se destaca que kriging modela en forma explícita la autocorrelación espacial de los datos y permite obtener una medida cuantificable de la incertidumbre asociada a la predicción. Por su parte, los métodos de ML son más flexibles que los enfoques tradicionales, ya que permiten capturar relaciones no lineales en los datos y mejorar la precisión predictiva. Además, facilitan la integración de información proveniente de múltiples fuentes o lotes y la incorporación de múltiples covariables sitio-específicas, enriqueciendo el modelado.

Los algoritmos de ML revolucionaron el análisis de datos, y son ampliamente utilizados para la resolución de problemáticas asociadas a la producción agrícola (Bhat & Huang, 2021). Los métodos basados en bosques aleatorios (*random forest*, RF) se destacan por su capacidad predictiva y robustez. Estos algoritmos combinan los resultados de múltiples árboles de decisión o árboles de regresión, según la naturaleza de las variables (Lantz, 2015). Específicamente, los árboles de regresión realizan predicciones a partir de un promedio de las predicciones de todos los árboles, almacenando en cada nodo de cada árbol el valor observado promedio. Una variante de RF, conocida como árboles de regresión cuántica (QRF), permite estimar cuantiles de la distribución condicional de las variables, y no solo su valor medio (Meinshausen & Ridgeway, 2006), permitiendo estimar cuantiles de la predicción. La aplicación de este método para la interpolación espacial (QRFI) a escala de lote mostró una mayor capacidad predictiva y menor tiempo de ejecución respecto a kriging (Córdoba & Balzarini, 2021).

Las redes neuronales artificiales (ANN) constituyen un método más complejo dentro de ML, a veces encajado en una categoría más específica, aprendizaje profundo, según el grado de complejidad que alcanzan (Jurafsky & Martin, 2024). Estos algoritmos pre-

sentan una estructura de pequeñas unidades de cómputo, llamadas “neuronas”, interconectadas y dispuestas en varias capas que procesan la información de manera iterativa, recibiendo señales de entrada (*input*), que son modificadas por una serie de pesos  $w_i$  y un término de sesgo  $b$  (ecuación 1). En base a la información recibida, cada neurona emite señales de salida (*output*) reguladas por una función  $f(a)$  no lineal conocida como función de activación (Candel & LeDell, 2024).

$$a = \sum_{i=1}^n w_i x_i + b \quad (1)$$

Distintos aspectos como el número de iteraciones o  $f(a)$  pueden ser fijados en distintos valores con el fin de minimizar una función de pérdida de información (Goodfellow et al., 2016). Dicha función también se puede modificar mediante técnicas de regularización para evitar sobreajuste, como Lasso (L1) y Ridge (L2). Con el mismo fin es posible controlar aspectos asociados a la detención temprana del ajuste, siendo factible determinar una métrica de detención, usualmente una medida de error, para la cual se especifica un criterio numérico de tolerancia y la cantidad de iteraciones por debajo de dicho criterio necesarias para detener el ajuste. Otros aspectos de interés cuya variabilidad está descrita en la bibliografía son la tasa de aprendizaje y la proporción de datos usada para validación respecto al total (Kamilaris & Prenafeta, 2018), aunque esta última opción suele ser más flexible ante grandes volúmenes de datos. El hallazgo de los valores óptimos para todas estas características es un proceso denominado optimización de hiperparámetros. Por otro lado, la estructura de las ANN presenta dos grandes aspectos: la cantidad de capas ocultas, y la cantidad de neuronas en cada capa oculta. Relaciones más complejas entre las variables conllevan la necesidad de emplear redes más complejas, ya sea por mayor número de neuronas o mayor cantidad de capas. Excederse en la complejidad de esta arquitectura puede llevar al sobreajuste de modelos, en tanto que estructuras demasiado simples pueden restringir la no-linealidad deseable al ajustar ANN.

En los últimos años, las ANN han ganado protagonismo en la predicción de propiedades del suelo (Wang et al., 2024; Zhang et al., 2025). Honorato Fernandes et al. (2019) calibraron y validaron modelos ANN para la predicción de MO a partir de valores de covariables de suelo mostrando buena precisión. Otros estudios (Heil et al., 2022) emplearon información de imágenes satelitales y variables ambientales para este mismo fin. Sin embargo, existen pocos estudios sobre el uso de modelos predictivos ANN para la predicción de MO a escala de lote integrando información de otros lotes y utilizando covariables de sitio y rendimiento.

El objetivo del presente trabajo fue evaluar la capacidad de predicción a escala fina de modelos de aprendizaje automático para el contenido de MO a partir de covariables de suelo y rendimiento de cultivos.

## 2 Materiales y Métodos

### 2.1 Datos

Se analizaron datos provenientes de dos establecimientos productivos, uno situado en el departamento Juárez Celman, provincia de Córdoba; otro situado en el departamento Moreno, provincia de Santiago del Estero, Argentina. Se tomaron 218 muestras distribuidas en siete lotes (cinco de Córdoba y dos de Santiago del Estero; Fig. 1) de superficie variable entre 51 y 210 hectáreas, con densidades de muestreo similares ( $0.285 \pm 0.024$  muestras por hectárea).



**Fig. 1.** Distribución de los puntos muestrales en los lotes ubicados en la provincia de Córdoba (izq.) y Santiago del Estero (der.).

Con el perímetro de cada uno de los lotes se elaboraron grillas individuales con celda de 10×10 metros. Para cada lote se contó con datos de rendimiento de soja (*Glycine max* L.) o maíz (*Zea mays* L.) en la campaña 2022/23 obtenidos con monitor de cosecha, además de valores de conductividad eléctrica superficial, subsuperficial y elevación obtenidos con rastra Veris.

Cada una de las covariables fue interpolada mediante kriging universal, extrayendo posteriormente valores predichos para cada una de ellas en los puntos de muestreo. Los valores predichos fueron posteriormente estandarizados por lote.

### 2.2 Metodología de análisis

Se evaluaron para cada uno de los lotes modelos QRFI, incluyendo un modelo local—que utiliza exclusivamente los datos del lote objetivo— y uno global, que incorpora información proveniente de todos los lotes del estudio. En estos últimos se incluyó como covariable una variable categórica que identifica la pertenencia de cada observación a un lote determinado. Estos mismos enfoques también fueron implementados me-

dante RK. Además, se ajustaron modelos supervisados basados en ANN pre-alimentadas (*feedforward*) entrenadas mediante retro-propagación a través del optimizador de descenso de gradiente estocástico (H2O.ai, 2024).

Tanto en QRFI como ANN, fue necesario incorporar la correlación espacial de los datos. Para ello, se estimó para cada celda de la grilla de predicción el valor de MO interpolado mediante el método de la distancia inversa ponderada (IDW), utilizando los cinco puntos de muestreo más cercanos y excluyendo la celda objetivo. Estos valores de predicción fueron empleados como una covariable adicional en el modelo.

La optimización de hiperparámetros para ANN se realizó mediante un criterio de búsqueda aleatorio (Bergstra & Bengio, 2012) el cual permite obtener un desempeño similar al método de grilla con una mínima fracción del tiempo computacional que éste requiere, siendo un método ideal para la optimización de hiperparámetros sobre sets de datos no analizados. Se fijó un tope de 100 modelos ajustados para cada búsqueda, empleando distintas funciones de activación y pérdida, estructuras de red y parámetros de detención temprana (Tabla 1).

Entre las estructuras evaluadas se incluyeron opciones simples de una o dos capas ocultas de entre 5 y 30 neuronas, y otras complejas, ya sea por número de neuronas por capa (256) o por número de capas (3 capas ocultas de 30 neuronas cada una). Se optó por emplear una lista discreta de valores y no un intervalo continuo para facilitar la comparación y análisis posterior de los modelos ajustados.

**Tabla 1.** Hiperparámetros evaluados en modelos de redes neuronales ajustados para la predicción a escala fina del contenido de materia orgánica, y valores puestos a prueba para cada uno.

Hiperparámetro	Valores evaluados
Estructura de la red	(5)
	(10)
	(20)
	(30)
	(5 5)
	(10 10)
	(20 20)
	(30 30)
Función de activación	(256 256)
	(30 30 30)
Función de activación	Rectifier, Tanh, Maxout
Función de pérdida L1	0, 1e-5
Función de pérdida L2	0, 1e-5
Desactivación de neuronas	Sí, No
Iteraciones de detención temprana	3, 5
Tasa de aprendizaje	0.001
	0.005

0.01

Los modelos fueron evaluados mediante validación cruzada k-fold ( $k = 10$ ). Para comparar el ajuste de los modelos, se utilizó la raíz del error cuadrático medio (RMSE), la cual se obtiene calculando la media de los residuos elevados al cuadrado (error cuadrático medio, MSE) y aplicando la raíz cuadrada. También se calculó una variante normalizada de este indicador (nRMSE) (ecuación 3) la cual divide su valor por el valor promedio de las observaciones ( $\bar{Y}$ ).

$$nRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_{pred} - y_{obs})^2}}{\bar{Y}} \quad (3)$$

Adicionalmente, los valores predichos por cada modelo para cada lote fueron comparados de a pares mediante una prueba T modificada, que ajusta el estadístico para contemplar la autocorrelación espacial de los datos (Kvamme, 1994). Esta prueba permite evaluar la correlación lineal entre las predicciones generadas por los distintos métodos de interpolación espacial. Para los análisis fueron utilizadas las librerías *automap*, *caret*, *cowplot*, *h2o*, *leaflet*, *ggmap*, *meteo*, *paar*, *SpatialPack* y *sf* del software R.

### 3 Resultados y Discusión

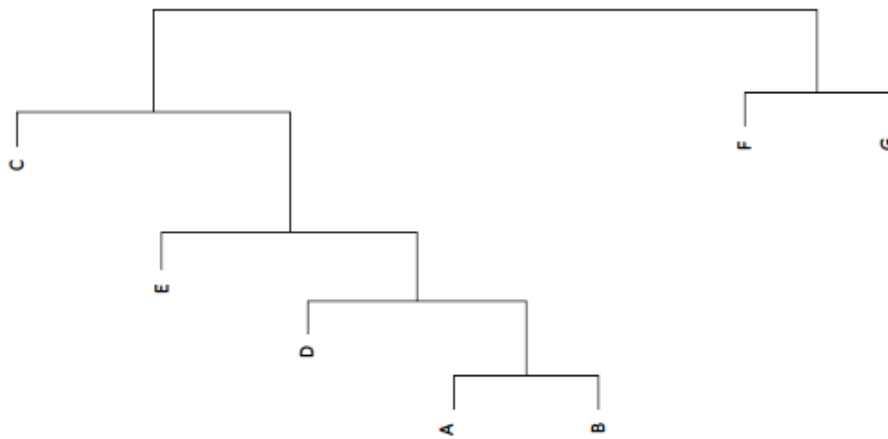
Los valores de MO observados en las muestras fueron de entre 1.45% y 2.66% (Tabla 2), con coeficientes de variación de entre 8% y 20%. Los lotes F y G, de Santiago del Estero, presentaron los valores promedio más elevados (2.23% y 2.66%, respectivamente) y menor variabilidad relativa (CV=8% y 10%, respectivamente). En el lote A se registró la mayor variabilidad (CV = 20%).

**Tabla 2.** Media, coeficiente de variación (CV), valores mínimos (Min) y máximos (Max) para el contenido de materia orgánica del suelo (%) evaluado por lote (L), y varianza multivariada (VM) por lote.

L	n	Media	CV	Min	Max	VM
A	55	1.915	20	1.15	2.82	0.142
B	57	2.07	17	1.025	2.992	0.043
C	17	1.788	14	1.357	2.339	0.361
D	17	2.085	17	1.358	2.734	2.62
E	17	1.458	18	0.774	1.84	0.663
F	26	2.272	8	1.943	2.648	0.002
G	28	2.662	10	2.287	3.371	0.015

Desde la perspectiva multivariada, considerando tanto los valores de MO como de las covariables, el lote C presentó la varianza multivariada más elevada (2.62), estimada

como el determinante de la matriz de varianzas-covarianzas; le siguieron en magnitud las varianzas de los lotes E (0.663), D (0.361) y A (0.142); para los tres lotes restantes (B, F y G) este indicador tuvo valores inferiores a 0.05. Al estudiar los lotes mediante un método de cluster jerárquico (Fig. 2) se distinguen dos grupos principales, que resultan equivalentes a los dos campos evaluados, mostrando que existe una clara diferenciación entre las características de ambos establecimientos.



**Fig. 2.** Dendrograma obtenido mediante análisis de clúster jerárquico de los lotes.

Para todos los lotes analizados en este estudio, los modelos ANN presentaron un mejor ajuste que los modelos QRFI y RK (Tabla 3). Para los lotes C, D y F el modelo que mejor ajustó fue un ANN global ( $nRMSE = 7.6, 9.4$  y  $7.2$ , respectivamente), mientras que para los demás la mejor opción fue un modelo local. En los lotes C y D, la elevada variabilidad multivariada en estos lotes podría ser un indicador de que los valores de las covariables pudieron haber comprometido la capacidad predictiva de los modelos locales.

Por su parte, los modelos QRFI mostraron en líneas generales mejor ajuste que los modelos RK. QRFI local fue el mejor modelo – excluyendo los modelos ANN – en los lotes A y F ( $nRMSE = 12.2$  y  $8.1$ , respectivamente). Para el lote G el modelo con menor grado de error fue un RK local ( $nRMSE = 10.8$ ). Los lotes F y G fueron los que presentaron menor variabilidad de MO. Esta homogeneidad relativa posiblemente haya facilitado que los modelos locales, a pesar de disponer de menos datos, hayan mostrado mejor ajuste.



**Tabla 3.** Raíz del error cuadrático medio expresada como porcentaje (nRMSE) para la predicción del contenido de materia orgánica en cada uno de los lotes (L), mediante modelos de aprendizaje automático: Quantile Regression Forest Spatial Interpolation (QRFI), Kriging Regresión (RK); y redes neuronales artificiales (ANN). En negrita se señala el mejor modelo para cada lote.

L	QRFI		RK		ANN	
	Global	Local	Global	Local	Global	Local
A	13.5	12.2	13.9	12.5	11.4	<b>11.1</b>
B	11.1	12.4	14.7	11.1	9.5	<b>9.4</b>
C	13.1	13.6	16.6	17.3	<b>7.6</b>	8.1
D	15.6	21.7	17.8	22.2	<b>9.4</b>	11.8
E	17.6	19.1	29.2	17.0	7.6	<b>7.3</b>
F	9.6	8.1	11.2	8.5	<b>7.2</b>	7.6
G	10.9	11.0	15.7	10.8	10.0	<b>6.2</b>

El número de muestras tomadas por lote resultó crucial a la hora de definir el desempeño de modelos globales o locales. Los lotes con menor número de muestras (C, D y E;  $n = 17$  muestras cada uno) exhibieron en general mejor ajuste en los modelos globales para los tres métodos de predicción, a excepción del del lote E usando ANN, donde existe una leve mejora del modelo local ( $nRMSE = 7.3$  vs  $7.6$ ). Dado que la densidad de muestreo fue relativamente homogénea en todos los lotes, los resultados sugieren que los modelos de predicción local pierden capacidad predictiva cuando la cantidad de muestras del lote es baja ( $< 30$ ) sin importar la superficie. Un muestreo más intenso en estos lotes podría reduciría los errores de predicción.

Es necesario aclarar que los modelos “globales” ajustados mediante ANN para cada lote no son iguales, dado que tanto los hiperparámetros como los pesos asignados a cada covariable han sido ajustados utilizando como set de datos de validación los datos del lote objetivo. Es decir, cada fila de la columna ANN Global representa un modelo diferente, no el ajuste de un mismo modelo sobre todos los lotes. El término Global en este trabajo hace referencia al uso de todos los datos disponibles en el set de datos de entrenamiento.

La combinación de hiperparámetros del modelo de mejor ajuste fue diferente para cada uno de los lotes (Tabla 4), aunque una tasa de aprendizaje de 0.01 se repitió en cinco lotes. En dos casos el uso de *dropout* fue necesario. En tres de los lotes fue preferible el uso de las funciones de regularización L1 y L2 en conjunto; estos tres casos presentaron una estructura neuronal de mayor complejidad. Lo contrario ocurrió para los lotes A, B, D y F, los cuales además presentaron estructuras más sencillas, con una única capa oculta. Para los lotes B, D y F no fue necesario emplear ninguna técnica de regularización.

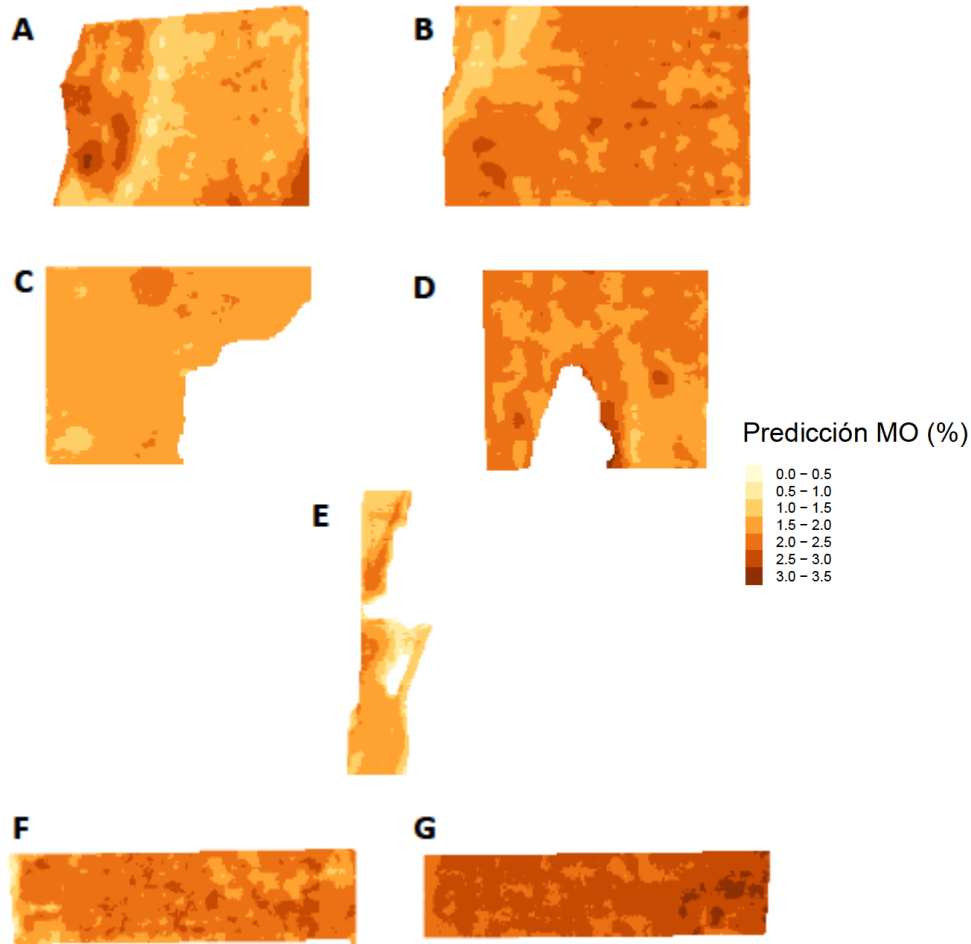
**Tabla 4.** Hiperparámetros correspondientes al mejor modelo ajustado para cada lote (L). L1 y L2: funciones de pérdida; SR: iteraciones de detención anticipada; LR: tasa de aprendizaje.

L	Método	Función de Activación	Estructura	L1	L2	SR	LR
A	Local	Maxout with Dropout	(10)	1e-5	0	3	0.01
B	Local	Rectifier	(20)	0	0	3	0.01
C	Global	Rectifier with Dropout	(256,256)	1e-5	1e-5	5	0.01
D	Global	Rectifier	(20)	0	0	3	0.01
E	Local	Tanh	(20,20)	1e-5	1e-5	5	0.001
F	Global	Maxout	(5)	0	0	3	0.005
G	Local	Tanh	(30,30,30)	1e-5	1e-5	5	0.01

Para los lotes C y G la estructura con mejor ajuste fue una estructura compleja, ya sea por número de neuronas o por número de capas. No existe a la fecha una regla o ecuación probada para la determinación de la estructura neuronal óptima. Si bien los coeficientes de error y el uso de técnicas de regularización apuntan en otro sentido, un posible sobreajuste de los modelos debe ponerse sobre la mesa al presentarse modelos de un cierto grado de complejidad frente a conjuntos pequeños de datos (Ying, 2018).

El tiempo de ajuste de los modelos locales fue notablemente menor (10.16 vs. 2.6 segundos en promedio), aunque hay que considerar el número de parámetros internos del modelo, relacionado a la cantidad de variables explicativas (13 para modelos locales y 5 para globales) y el tamaño del conjunto de datos cargado en cada uno (489 vs. 17-55). Corrigiendo el tiempo computacional requerido por los modelos por la cantidad de parámetros, los valores fueron más cercanos, aunque los modelos locales siguieron exhibiendo mayor celeridad (0.782 vs. 0.52 segundos/parámetro en promedio). El procesamiento y análisis de los datos, se realizó en un servidor remoto Ubuntu 20.04.6 LTS, con procesador Intel(R) Core(TM) i7-8700K CPU @ 3.70GHz  $\times$  12 núcleos y 62 GB de RAM.

Los mapas evaluados a partir del mejor modelo para cada lote mostraron coherencia desde el punto de vista espacial, sin observarse artefactos como líneas rectas u otros elementos no deseados (Fig. 3), los cuales son comunes en métodos que incorporan información de covariables geográficas (Xie et al., 2022). En el límite entre algunos de los lotes (ej. Sur del lote B y norte del lote C) se observan cambios a nivel del valor promedio de MO, lo cual es esperable dado que los modelos usados para la predicción fueron específicos para cada lote. No obstante, se aprecia cierta continuidad en el patrón espacial, lo que indica coherencia en la distribución de MO a mayor escala.



**Fig. 3.** Mapa de valores de materia orgánica (MO%) predichos para los lotes evaluados, utilizando redes neuronales.

Las predicciones de todos los modelos en general presentan una correlación positiva significativa ( $p < 0.05$ ) entre sí, según la prueba T modificada. Como excepción, el método RK global presentó una correlación negativa con el modelo RK local en el lote C ( $r = -0.58$  y  $-0.69$  para modelos con y sin efecto de lote,  $p < 0.05$ ) y con el modelo ANN local en el lote G ( $r = -0.54$  para ambos casos,  $p < 0.01$ ). Los modelos ANN local y global presentaron correlación significativa en todos los lotes, al igual que los modelos QRFI, con excepción del lote F, donde la correlación no fue significativa.

Los modelos propuestos en este trabajo son individuales para cada lote, es decir, su aplicabilidad se restringe únicamente a un lote, permitiendo evaluar dentro de cada uno el máximo desempeño de los distintos métodos. En vistas a una aplicación más práctica de los conocimientos generados por este trabajo, estudios futuros deberían apuntar al

ajuste de un único modelo global con un desempeño aceptable en todos los lotes a la vez.

## 4 Conclusiones

Los modelos basados en redes neuronales artificiales se consolidan como una alternativa promisorio para la predicción de MO a escala de lote. La optimización de hiperparámetros de los modelos debe apuntar a corregir potencial sobreajuste a los datos, especialmente frente a conjuntos de datos de menor dimensión. Las predicciones realizadas mediante los distintos modelos estuvieron correlacionadas entre sí, mostrando convergencia en la captura de los patrones de variabilidad espacial del contenido de MO en el suelo.

## Referencias Bibliográficas

- Adamchuk, V.I., Hummel, J.W., Morgan, M.T., Upadhyaya, S.K. (2004). On-the-go sensors for precision agriculture. *Computers and Electronics in Agriculture* 44(1): 71 – 91. <https://doi.org/10.1016/j.compag.2004.03.002>
- Bashie, A.L., Ayankukwa, A.U., Owojoku, O.D., Nicholas, O.G., Ishoro, A.P., Christiana, U.A. (2024). Predicting Environmental Covariates of Soil Organic Matter at Sub-Regional Scale for Sustainable Agricultural Development in Southeast Nigeria. *Pol. J. Environ. Stud.* 34(3): 2011 – 2021. doi: 10.15244/pjoes/186888
- Bergstra, J., Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research* 13(1): 281 – 305.
- Bhat, S.A., Huang, N. (2021). Big Data and AI Revolution in Precision Agriculture: Survey and Challenges. *Ieee Access* 9: 110209 – 110222. doi: 10.1109/ACCESS.2021.3102227
- Candel, A., LeDell, E. (2024). Deep learning with H2O. 6a ed. H2O ai, Mountain View, CA.
- Córdoba, M., Paccioretti P.A., Giannini Kurina, F., Bruno, C.I., Balzarini, M.G. (2020). Guía para el análisis de datos espaciales: Aplicaciones en agricultura. 1a ed. Editorial Brujas, Buenos Aires, Argentina. ISBN: 978-987-760-272-2
- Córdoba, M., Balzarini, M. (2020). Mapeo de materia orgánica del suelo a escala de campo. Congreso Argentino de Agroinformática, Jornadas Argentinas de Informática e Investigación Operativa 7(4): 176 – 186. ISSN: 2525-0949
- Córdoba, M., Balzarini, M. (2021). A random forest-based algorithm for data-intensive spatial interpolation in crop yield mapping. *Computers and Electronics in Agriculture* 184, 106094. <https://doi.org/10.1016/j.compag.2021.106094>
- Fryda, T., LeDell, E., Gill, N., Aiello, S., Fu, A., Candel, A., Click, C., Kraljevic, T., Nykodym, T., Ahoyoum, P., Kurka, M., Malohlava, M., Poirier, S., Wong, W., Rehak, L., Eckstrand, E., Hill, B., Vidrio, S., Jadhawani, S., ... Novotny, M. (2024). H2o: R Interface for the ‘H2O’ Scalable Machine Learning Platform. doi: 10.32614/CRAN.package.h2o
- García Tomillo, A., Mirás Avalos, J.M., Dafonte Dafonte, J., Paz González, A. (2017). Estimating soil organic matter using interpolation methods with a electromagnetic induction sensor and topographic parameters: a case study in a humid region. *Precision Agriculture* 18: 882 – 897. <https://doi.org/10.1007/s11119-016-9481-6>
- Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep Learning. MIT Press, Cambridge, MA.

- H2O.ai. (2024). Deep Learning (Neural Networks). Consultada el 22 de marzo de 2025. URL: <https://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/deep-learning.html>
- Heil, J., Jörges, C., Stumpe, B. (2022). Fine-Scale Mapping of Soil Organic Matter in Agricultural Soils Using UAVs and Machine Learning. *Remote Sens.* 14(14), 3349. <https://doi.org/10.3390/rs14143349>
- Honorato Fernandes, M.M., Prates Coelho, A., Fernandes, C., da Silva, M.F., Campos Dela Marta, C. (2019). Estimation of soil organic matter content by modelling with artificial neural networks. *Geoderma* 350: 46 – 51. <https://doi.org/10.1016/j.geoderma.2019.04.044>
- Jurafsky, D., Martin, J.H. (2024). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models* [Manuscrito sin publicar]. 3a ed. URL: [https://web.stanford.edu/~jurafsky/slp3/ed3book\\_Jan25.pdf](https://web.stanford.edu/~jurafsky/slp3/ed3book_Jan25.pdf)
- Kamilaris, A., Prenafeta-Baldú, F.X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture* 147 70 – 90. <https://doi.org/10.1016/j.compag.2018.02.016>
- Kvamme, K.L. (1994). Spatial statistics and GIS: An integrated approach. En: Andresen, J., Madsen, T., Scollar, I. (eds.). *Computing the Past: Computer Applications and Quantitative Methods in Archaeology*. Aarhus University Press, Aarhus, Dinamarca. 91 – 103.
- Kweon, G., Lund, E., Maxton, C. (2013). Soil organic matter and cation-exchange capacity sensing with on-the-go electrical conductivity and optical sensors. *Geoderma* 199: 80 – 89. <https://doi.org/10.1016/j.geoderma.2012.11.001>
- Lantz, B. (2015). *Machine learning with R*. 2a ed. Packt Publishing Ltd., Birmingham, Reino Unido. ISBN: 978-1-78439-390-8
- LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature* 521: 436 – 444.
- Mohammadpour, M., Roshan, H., Arashpour, M., Masoumi, H. (2024). Machine learning assisted Kriging to capture spatial variability in petrophysical property modelling. *Marine and Petroleum Geology* 167, 106967. <https://doi.org/10.1016/j.marpetgeo.2024.106967>
- Meinshausen, N., Ridgeway, G. (2006). Quantile Regression Forests. *Journal of Machine Learning Research* 7(6): 983 – 999.
- Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J. (1995). Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. *Geoderma* 67(3-4): 215 – 226. [https://doi.org/10.1016/0016-7061\(95\)00007-B](https://doi.org/10.1016/0016-7061(95)00007-B)
- Oldfield, E.E., Bradford, M.A., Augarten, A.J., Cooley, E.T., Radatz, A.M., Radatz, T., Ruark, M.D. (2022). Positive associations of soil organic matter and crop yields across a regional network of working farms. *Soil Science Society of America Journal* 86(2): 384 – 397. <https://doi.org/10.1002/saj2.20349>
- Oliver, M.A. (2010). *Geostatistical Applications for Precision Agriculture*. 1a ed. Springer, Dordrecht, Países Bajos. ISBN: 978-90-481-9133-8
- Oliver, M.A., Webster, R. (2015). *Basic Steps in Geostatistics: The Variogram and Kriging*. 1a ed. Springer, Cham, Suiza. ISBN: 978-3-319-15865-5
- Omonode, R.A., Vyn, T.J. (2006). Spatial dependence and relationships of electrical conductivity to soil organic matter, phosphorus, and potassium. *Soil Science* 171(3): 223 – 238. doi: 10.1097/01.ss.0000199698.94203.a4
- Piccini, C., Marchetti, A., & Francaviglia, R. (2014). Estimation of soil organic matter by geostatistical methods: Use of auxiliary information in agricultural and environmental assessment. *Ecological Indicators*, 36, 301–314.
- Tiessen, H., Cuevas, E., Chacon, P. (1994). The role of soil organic matter in sustaining soil fertility. *Nature* 371: 783 – 785.

- Vázquez, J.M. (2016). Manejo sitio específico. En: Gueçaimburu, J.M., Rojo, V., Reposo, G., Vázquez, J.M., Martínez, M. Manejo del sistema agropecuario. 1a ed. Editorial Universidad Nacional de Luján (EdUNLu), Luján, Buenos Aires, Argentina.
- Wang, Y., Shi, L., Hu, Y., Hu, X., Song, W., Wang, L. (2024). A comprehensive study of Deep learning for soil moisture prediction. *Hess* 28(4): 917 – 943.
- Xie, J., Huang, J., Zheng, C., Huang, S., Burton, G., (2022). A generic framework for geotechnical subsurface modeling with machine learning. *Journal of Rock Mechanics and Geotechnical Engineering* 14(5): 1366 – 1379. <https://doi.org/10.1016/j.jrmge.2022.08.001>
- Ying, X. (2019). An Overview of Overfitting and its Solutions. *IOP Conf. Series: Journal of Physics. Conf. Series* 1168. <https://doi.org/10.1088/1742-6596/1168/2/022022>
- Zhang, Y., Luo, C., Zhang, W., Wu, Z., Zang, D. (2025). Mapping Soil Organic Matter in Black Soil Cropland Areas Using Remote Sensing and Environmental Covariates. *Agriculture* 15(3), 339. <https://doi.org/10.3390/agriculture15030339>