

INFORME

MAPEO DIGITAL DE FRACCIONES GRANULOMÉTRICAS DEL SUELO

DE LA PROVINCIA DE CÓRDOBA 2024-2025

JULIO 2026

IDECOR - INFRAESTRUCTURA DE DATOS ESPACIALES DE LA PROVINCIA DE CÓRDOBA

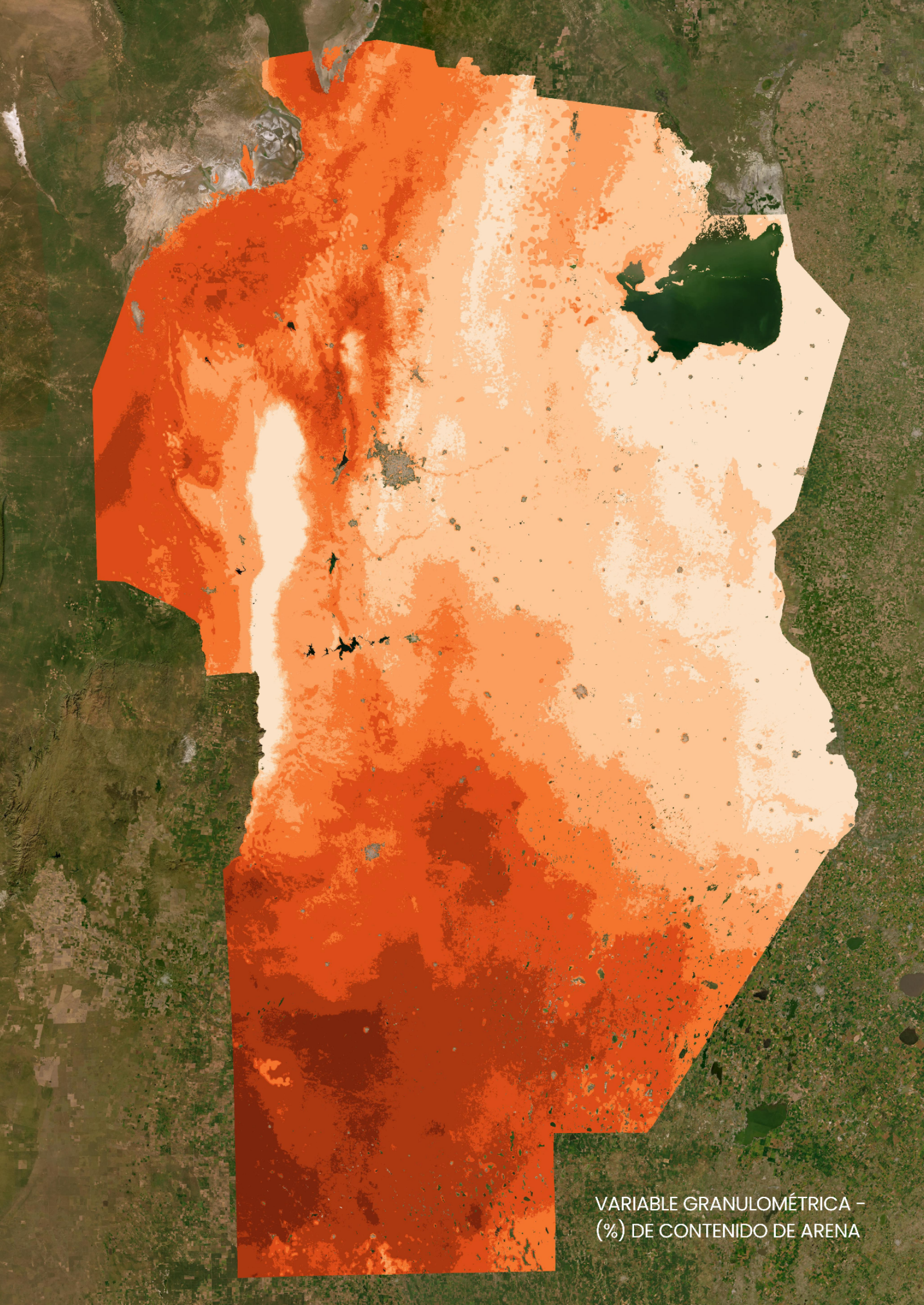
GOBIERNO DE LA PROVINCIA DE CÓRDOBA

Colabora:

MINISTERIO DE
BIOAGROINDUSTRIA

Coordina:





VARIABLE GRANULOMÉTRICA -
(%) DE CONTENIDO DE ARENA

Organismos participantes y equipo de trabajo

Desarrollo, covariables, apoyo científico e infraestructura tecnológica

IDECOR (Infraestructura de Datos Espaciales de la Provincia de Córdoba)

Unidad de Fitopatología y Modelización Agrícola (UFyMA), unidad de doble dependencia INTA-CONICET.

Secretaría de Agricultura de la Provincia de Córdoba

INTA Centro Regional Córdoba

Participantes

Mariano Córdoba

Santiago Dellarossa

Matías Godino

Ezequiel Pozzi

Federico Monzani

Hernán Morales

Colaboraciones

Instituto Nacional de Tecnología Agropecuaria (Juan Cruz Molina, Carolina Álvarez, Lautaro Faule, Mauro Lanfranco).

Secretaría de Agricultura/Ministerio de Bioagroindustria de la Provincia de Córdoba (Marcos Blanda, Santiago Dellarossa, Matías Godino).

Facultad de Ciencias Agropecuarias (FCA), Universidad Nacional de Córdoba (UNC) (Susana Hang).

Facultad de Agronomía y Veterinaria, Universidad Nacional de Río Cuarto (UNRC) (Catalina Bozzer).

Instituto Multidisciplinario de Biología Vegetal - IMBIV CONICET (Esteban Kowaljow).

Servicios Tecnológicos AGD S.A. (Sebastián Storti, Mariano Granatelli).

RAVIT (Esteban Tronfi).

Orbely (Franco Vizzio).

Versión y licencia

Versión 2, año 2025. Documento de trabajo para revisión técnica.

MAPEO DIGITAL DE FRACCIONES GRANULOMÉTRICAS DEL SUELO EN LA PROVINCIA DE CÓRDOBA, año 2025. Está distribuido bajo una Licencia Creative Commons Atribución 4.0 Internacional. Libre para compartir, distribuir, copiar y adaptar, siempre y cuando se otorguen los créditos correspondientes.



Citar como: Córdoba M., Bozzer C., Faule L., Álvarez C., Lanfranco M., Hang S., Monzani F., Morales H. Mapeo Digital de Fracciones Granulométricas del Suelo en la Provincia de Córdoba, 2025. Infraestructura de Datos Espaciales de Córdoba (IDECOR), julio de 2026.

Aviso de uso

Los datos desarrollados en el marco del presente trabajo son parte del trabajo de mapeo de variables de suelo, desarrollado por IDECOR con el apoyo de CONICET, para la realización de estudios territoriales e inmobiliarios en particular. No obstante, las características técnicas de los resultados obtenidos posibilitan otros usos científicos y técnicos, quedando éstos a criterio y responsabilidad de los usuarios.

Los mapas se diseñaron para servir como herramienta cartográfica a escala regional y provincial; no fueron creados con objetivos específicos y/o para trabajos que demanden escalas y detalles mayores.

El equipo técnico se desliga de cualquier uso indebido que pueda realizarse de los mapas y datos desarrollados fuera del ámbito para el cual fueron diseñados y/o sus características técnicas posibilitan, quedando estos otros usos bajo responsabilidad de los usuarios

Contenido

Resumen.....	6
1. Antecedentes.....	7
2. Alcance de los productos.....	7
3. Datos.....	7
3.1. Estandarización y procesamiento.....	8
3.2. Análisis descriptivo de la muestra y de la predicción.....	8
4. Metodología.....	10
4.1. Covariables.....	10
a. Clima.....	10
b. Vegetación y cobertura del suelo.....	11
c. Topografía.....	11
4.2. Selección de covariables.....	12
4.3. Algoritmos de aprendizaje automático, calibración y validación.....	13
4.4. Métricas de desempeño.....	14
4.5. Predicción e incertidumbre.....	14
5. Resultados.....	16
5.1. Desempeño predictivo de los modelos.....	16
5.2. Mapas de predicción e incertidumbre.....	17
6. Consideraciones finales.....	20
7. Referencias bibliográficas.....	21

Resumen

El presente informe documenta la actualización de los mapas de fracciones granulométricas del suelo —arcilla, limo y arena— para la Provincia de Córdoba, desarrollados en el marco conceptual del Mapeo Digital de Suelos (MDS). La actualización retoma el antecedente de 2021, cuando se publicaron mapas provinciales de granulometría con resolución espacial de 500 m, e incorpora la mejora metodológica iniciada en 2023-2024 para generar productos a 90 m de resolución, equivalente a 0,81 ha por celda.

La base de datos utilizada contiene 2.892 observaciones georreferenciadas armonizadas a una profundidad equivalente de 0–20 cm para las tres fracciones texturales. Las predicciones se realizaron sobre aproximadamente 19,3 millones de celdas válidas. Las covariables ambientales y fuentes de información siguen la lógica de los trabajos recientes sobre MDS llevados a cabo por IDECOR: clima, vegetación, cobertura y uso del suelo, topografía y otras variables derivadas de sensores remotos y productos geospaciales. Para cada fracción granulométrica se evaluaron modelos de aprendizaje automático, incluyendo Quantile Regression Forest (QRF), Cubist, XGBoost, Support Vector Machine (SVM) y redes neuronales multicapa (MLP).

Los mejores resultados fueron obtenidos por QRF, con valores de RMSE de 3,53 puntos porcentuales para arcilla, 6,39 para limo y 8,19 para arena. Cubist y XGBoost presentaron desempeños cercanos, mientras que SVM mostró menor precisión y MLP resultó inferior para estas bases de datos. En consecuencia, QRF fue utilizado para la generación de los mapas finales y de la incertidumbre condicional de predicción.

Los mapas muestran un patrón espacial consistente con la distribución regional de materiales parentales, relieve y ambientes geomorfológicos de Córdoba: mayor proporción de arena hacia sectores del oeste y sudoeste, y mayores contenidos de limo y arcilla en ambientes de llanura del centro, este y noreste. Esta organización espacial es coherente con la transición desde ambientes serranos y de piedemontes hacia llanuras loésicas y sectores con diferente historia geomorfológica, que condicionan la distribución de tamaños de partículas y, por lo tanto, propiedades físicas relevantes como retención de agua, infiltración, aireación, susceptibilidad a compactación y capacidad de intercambio catiónico asociada a los tipos de arcillas y contenidos de materia orgánica.

1. Antecedentes

La generación de información espacial de suelos en Córdoba tiene antecedentes recientes vinculados a la integración de datos públicos y privados, sensores remotos, covariables ambientales y modelación estadística. En 2021 se amplió el alcance de los mapas provinciales de propiedades del suelo para incluir, además de materia orgánica (MO), fósforo disponible (P) y pH, las fracciones granulométricas arena, arcilla y limo. Ese producto informó las variables en la capa superficial del suelo de 0-20 cm, acompañado por mapas de incertidumbre de predicción a escala provincial y con resolución de 500 m, siguiendo protocolos de Mapeo Digital de Suelos (MDS).

La actualización 2024 consolidó el cambio de escala espacial iniciado en 2023, pasando de 500 m a 90 m para los mapas de MO, P, pH y COS. Esa mejora permitió capturar con mayor detalle la variabilidad espacial de propiedades edáficas y utilizar covariables satelitales y de cobertura del suelo de alta resolución. En continuidad con ese avance, el presente informe actualiza los mapas de textura del suelo —arcilla, limo y arena— a 90 m.

La textura del suelo es una propiedad intrínseca determinada por la génesis del suelo, de gran relevancia agronómica, hidrológica y ambiental. La proporción relativa de arena, limo y arcilla condiciona la capacidad de retención de agua, la infiltración, la aireación, la susceptibilidad a la erosión —tanto hídrica como eólica—, y la dinámica de nutrientes, a través de su efecto sobre la capacidad de intercambio catiónico y la protección física de la materia orgánica. En sistemas agropecuarios, constituye además un factor explicativo clave de las diferencias en productividad, respuesta a insumos, disponibilidad de agua y capacidad de almacenamiento de secuestro de carbono orgánico.

2. Alcance de los productos

Los productos generados corresponden a mapas continuos de contenido de arcilla, limo y arena en el horizonte superficial del suelo, expresados en porcentaje (%), para la Provincia de Córdoba y con resolución espacial de 90 m. Las predicciones se acompañan con mapas de incertidumbre expresados como desvío estándar (DE) condicional de la predicción, en las mismas unidades de las variables texturales.

- Resolución espacial: 90 m (0,81 ha por celda).
- Variables mapeadas: arcilla (%), limo (%) y arena (%).
- Profundidad de referencia: capa superficial, 0-20 cm.
- Producto principal: mapas de predicción media de arcilla, limo y arena a 90 m de resolución espacial.
- Producto complementario: mapas de incertidumbre de predicción expresados como DE.

3. Datos

La base de datos de granulometría integrada para esta actualización contiene 2.892 observaciones georreferenciadas con información simultánea de arcilla, limo y arena. Las fuentes de esta información provinieron del Plan Mapas de Suelo (Ministerio de

Agricultura y Ganadería de Córdoba-INTA), la Facultad de Ciencias Agropecuarias de la Universidad Nacional de Córdoba y la Facultad de Agronomía y Veterinaria de la Universidad Nacional de Río Cuarto. Por su parte, las empresas privadas aportantes fueron Seiker (en 2020 y 2021), Orbely (en 2020 y 2021), RAVIT (2021) y AGD (2020). La información fue procesada para garantizar consistencia espacial, compatibilidad de unidades y adecuada vinculación con las covariables ambientales utilizadas por los modelos de predicción.

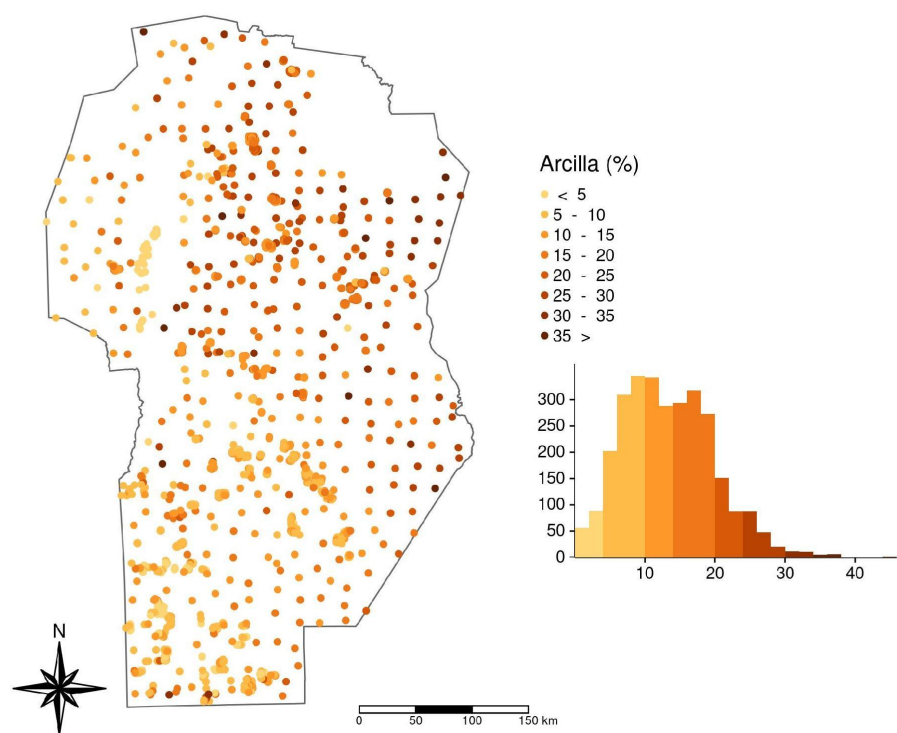
3.1. Estandarización y procesamiento

La base de datos fue sometida a un proceso de estandarización y depuración previo al modelado. En primer lugar, se revisó la consistencia de las unidades de medida y se conservaron únicamente registros completos, es decir, aquellos que presentaban información simultánea de arcilla, limo y arena. Esto permitió trabajar con una base homogénea para las tres fracciones granulométricas y evitar diferencias en el soporte muestral entre variables. Los datos provenientes de perfiles de suelo o aquellos con mediciones en múltiples profundidades fueron estandarizados a una profundidad de 0–20 cm mediante funciones de suavizado basadas en áreas equivalentes, conocidas como splines (Malone et al., 2009). Este procedimiento permite armonizar observaciones tomadas a distintas profundidades y obtener valores comparables para el horizonte superficial, que constituye la profundidad de referencia de los mapas generados.

3.2. Análisis descriptivo de la muestra y de la predicción

La distribución espacial de las muestras se presenta en la Figura N°1. Las muestras cubren una amplia porción del territorio provincial, aunque con diferencias en densidad espacial entre regiones.

Figura N°1. Distribución espacial de puntos de muestreo e histograma de frecuencias absolutas de valores de Arcilla en la Provincia de Córdoba.



Fuente: Elaboración propia, 2026.

Las medidas resumen (Tabla N°1) muestran que la muestra presentó valores promedio de 13,4% para arcilla, 45,1% para limo y 40,4% para arena. La mayor variabilidad relativa se observó en arena (CV = 54%) y arcilla (CV = 49%), mientras que limo presentó menor variabilidad relativa (CV = 38%). Estos resultados indican una marcada heterogeneidad textural en la provincia, consistente con la diversidad geomorfológica, litológica y edáfica del territorio cordobés.

Tabla N°1. Estadísticos descriptivos de la muestra para valores de arcilla, limo y arena. n: número de observaciones; CV: coeficiente de variación (%); Mín: valor mínimo; Max: valor máximo; P(05): percentil 5; P(95): percentil 95.

Variable	n	Media	CV	Min.	Max.	P05	P95
Arcilla	2.892	13,4	49	<0,00	44,6	4,0	24,7
Limo		45,1	38	<0,00	89,1	14,3	64,6
Arena		40,2	54	0,05	96,6	13,6	77,5

Fuente: Elaboración propia, 2026.

4. Metodología

El estudio se basó en técnicas de Mapeo Digital de Suelos (MDS), que integran observaciones de campo, covariables ambientales y modelos estadísticos o de aprendizaje automático para predecir propiedades edáficas en el espacio. El marco conceptual corresponde al modelo SCORPAN (McBratney et al., 2003), que representa una propiedad del suelo como función de factores relacionados con el suelo, clima, organismos, relieve, material parental, edad/tiempo y posición espacial.

4.1. Covariables

La selección adecuada de covariables es un elemento crítico en el MDS, ya que determina la capacidad de los modelos para capturar y explicar la variabilidad espacial de las propiedades edáficas (Yigini et al., 2018). Estas covariables actúan como *proxies* de los factores que influyen en los procesos de formación y distribución del suelo, integrando información climática, biológica y geomorfológica en los modelos predictivos.

a. Clima

Las variables climáticas empleadas en este estudio fueron derivadas de la base de datos CHELSA v2.1 (*Climatologies at High Resolution for the Earth's Land Surface Areas*), que proporciona estimaciones de temperatura y precipitación mediante técnicas de reescalado (*downscaling*) del modelo de reanálisis climático ERA-Interim a una resolución espacial de 1 km (Karger et al., 2017). CHELSA integra información atmosférica y orográfica, lo que permite capturar con mayor detalle los gradientes climáticos locales, especialmente en áreas con variaciones topográficas significativas. A partir de los datos mensuales de temperatura y precipitación, se calcularon diversas variables bioclimáticas para representar condiciones promedio, extremos térmicos e hídricos, y factores climáticos potencialmente limitantes. Estas variables incluyeron promedios anuales, como temperatura media anual y precipitación anual acumulada; indicadores de extremos, como temperatura de los meses y trimestres más cálidos y fríos, y precipitación de los meses y trimestres más húmedos y secos; e índices climáticos derivados que permiten resumir contrastes térmicos e hídricos relevantes para la diferenciación ambiental. Adicionalmente, se incorporaron variables de evapotranspiración calculadas mediante la ecuación de Penman-Monteith (Allen et al., 1998), que estima la pérdida de agua desde el suelo y la vegetación bajo diferentes condiciones climáticas. Estas incluyeron evapotranspiración media mensual, máxima y mínima por mes, representando distintos niveles de demanda atmosférica.

Aunque la textura del suelo es una propiedad relativamente estable y no responde directamente a variaciones climáticas recientes, las covariables climáticas son útiles para representar ambientes pedogenéticos de largo plazo. Gradientes de temperatura, precipitación y evapotranspiración pueden influir en la intensidad de meteorización, el desarrollo del perfil, la estabilidad de agregados, los procesos de erosión y la redistribución de partículas finas. Por este motivo, estas variables fueron incorporadas como predictores indirectos de las condiciones ambientales que contribuyen a explicar la distribución espacial de arcilla, limo y arena. Para integrar estas variables en los modelos predictivos, se realizó un reescalado de la resolución espacial original de 1 km a 90 m, utilizando interpolación bicúbica (Keys, 1981). Este procedimiento permitió

compatibilizar las covariables climáticas con la resolución espacial de los productos finales y con otras capas predictoras, como las variables de topografía y vegetación.

b. Vegetación y cobertura del suelo

Las variables relacionadas con la vegetación se obtuvieron principalmente de datos satelitales generados por los sensores Landsat 8 y MODIS, que ofrecen una cobertura espacial y temporal adecuada para caracterizar la dinámica de la cobertura terrestre. A partir de estas fuentes, se calcularon métricas estadísticas como la mediana y el desvío estándar del Índice de Vegetación de Diferencia Normalizada (NDVI), indicador de la densidad y vigor de la vegetación. Estos cálculos se realizaron por trimestre para la serie histórica 2013–2023, proporcionando una caracterización temporal de los patrones vegetativos. Además, se aplicaron análisis similares para otros indicadores, como la temperatura de superficie terrestre (LST), útil para evaluar interacciones entre vegetación, cobertura superficial y balance energético; y la fracción de radiación fotosintéticamente activa absorbida por la vegetación (fPAR), que representa un proxy de la productividad primaria. También se evaluaron propiedades específicas de reflectancia, como la mediana y el desvío estándar de la banda 6 de Landsat 8, correspondiente al infrarrojo de onda corta, durante el trimestre junio-agosto.

El Mapa de Cobertura y Uso del Suelo 2022–2023 de la provincia de Córdoba (IDECOR, 2023) fue otra fuente importante de información. Este mapa clasifica 22 tipos de cobertura terrestre, incluidos bosque, arbustales, matorrales, pastizales, cultivos anuales, pasturas, plantaciones forestales, suelo desnudo y rocas. Cada clase fue utilizada para calcular la proporción de cobertura dentro de celdas de 90 m, permitiendo integrar estas proporciones como variables en los modelos predictivos. Este producto, con una resolución espacial original de 10 m y un área mínima mapeable de 1 ha, facilitó una caracterización detallada de la cobertura y uso del territorio.

En el caso específico de la granulometría, las covariables de vegetación y cobertura no deben interpretarse necesariamente como causas directas de la textura, sino como indicadores indirectos de ambientes edáficos, geomorfológicos y de uso del suelo. Por ejemplo, ciertas coberturas pueden estar asociadas a suelos someros, afloramientos rocosos, sectores de acumulación, ambientes arenosos, áreas agrícolas de llanura o zonas con distinto grado de desarrollo del perfil. De esta manera, la información satelital y de cobertura aporta señales espaciales complementarias que pueden mejorar la predicción de las variables granulométricas.

c. Topografía

Las variables geomorfométricas son esenciales para comprender cómo la morfología del terreno influye en la formación, distribución y redistribución espacial del suelo. Estas variables se obtuvieron de la base de datos global Geomorpho90m (Amatulli et al., 2020), que utiliza el Modelo Digital de Elevación MERIT a una resolución aproximada de 90 m. Este producto constituye una fuente estandarizada de información topográfica y geomorfométrica a escala global, adecuada para aplicaciones de MDS. Se incluyeron 26 variables geomorfométricas estandarizadas que describen diferentes aspectos del relieve, entre ellas derivadas de primer y segundo orden, como pendiente, curvatura y exposición; métricas de rugosidad y variabilidad superficial; y variables asociadas a formas geomorfológicas, tales como

valles, crestas, mesetas y otras unidades del paisaje definidas a partir de propiedades geométricas del terreno.

Estas variables permiten capturar procesos como drenaje, retención de humedad, erosión, transporte y deposición de materiales, todos ellos directamente relacionados con la variación espacial de las fracciones granulométricas. En ambientes serranos y de piedemonte, el relieve puede estar asociado a suelos someros, mayor presencia de fragmentos gruesos, afloramientos rocosos, erosión diferencial y cambios abruptos de material parental. En ambientes de llanura, pequeñas variaciones topográficas pueden reflejar diferencias en acumulación de sedimentos finos, presencia de paleocauces, bajos, áreas de escurrimiento o sectores con distinto grado de desarrollo del perfil. Por ello, las covariables topográficas resultan particularmente relevantes para la predicción de las fracciones granulométricas en una región con marcada diversidad fisiográfica.

Finalmente, todas las covariables fueron integradas en una grilla común de 90 m de resolución. A cada punto de muestreo se le extrajeron los valores de las covariables correspondientes a su localización, generando la matriz de datos utilizada para la selección de predictores, calibración de modelos y posterior predicción espacial continua. El procesamiento de la información satelital y topográfica se realizó utilizando Google Earth Engine (Gorelick et al., 2017).

4.2. Selección de covariables

Se realizó una selección de predictores con el objetivo de identificar, para cada variable granulométrica y cada algoritmo de modelación, el subconjunto de covariables con mayor capacidad predictiva. La selección de covariables se realizó mediante un procedimiento de eliminación recursiva de predictores (Recursive Feature Elimination, RFE) (Kuhn & Johnson, 2013). Este método ajusta modelos sucesivos utilizando distintos subconjuntos de variables y evalúa su desempeño predictivo mediante validación cruzada. A partir de la importancia relativa de los predictores y de la capacidad predictiva obtenida para diferentes tamaños de subconjuntos, el procedimiento identifica el conjunto de covariables que ofrece el mejor compromiso entre precisión y parsimonia. El procedimiento fue aplicado de manera independiente para arcilla, limo y arena, debido a que cada fracción granulométrica puede responder de forma diferente a los gradientes ambientales representados por las covariables disponibles.

Asimismo, la selección de covariables se realizó en forma separada para cada algoritmo evaluado. Esta decisión se debe a que los modelos difieren en su forma de representar relaciones entre predictores y variable respuesta. En consecuencia, un subconjunto de covariables adecuado para un algoritmo no necesariamente resulta óptimo para otro. Para la implementación de RFE se evaluaron subconjuntos crecientes de covariables, desde combinaciones reducidas hasta conjuntos más amplios, utilizando validación cruzada de cinco particiones. En cada iteración, los modelos fueron entrenados y evaluados con el objetivo de identificar el tamaño de subconjunto que maximizaba el desempeño predictivo. Posteriormente, las covariables seleccionadas para cada combinación de variable respuesta y algoritmo fueron utilizadas en la etapa de calibración final de los modelos.

4.3. Algoritmos de aprendizaje automático, calibración y validación

Se evaluaron cinco algoritmos de aprendizaje automático con características complementarias:

- Quantile Regression Forest (QRF) (Meinshausen, 2006): algoritmo basado en árboles de decisión que extiende el enfoque de Random Forest. Permite obtener predicciones puntuales y, adicionalmente, estimar la distribución condicional de la variable respuesta. Esta característica permite derivar medidas de incertidumbre, como el desvío estándar de la predicción.
- Cubist (Kuhn et al., 2016): modelo híbrido que combina árboles de decisión, reglas de partición y regresiones lineales locales. El algoritmo divide el espacio de covariables en regiones más homogéneas y ajusta modelos lineales específicos dentro de ellas.
- Extreme Gradient Boosting (XGBoost) (Chen & Guestrin, 2016): algoritmo de boosting basado en árboles de decisión. Construye modelos en forma secuencial, donde cada nuevo árbol intenta corregir los errores residuales de los árboles anteriores. Es un método ampliamente utilizado por su capacidad para capturar relaciones no lineales e interacciones entre predictores.
- Support Vector Machine (SVM) (Smola & Schölkopf., 2004): método basado en funciones kernel que permite modelar relaciones no lineales entre covariables y variable respuesta. En este estudio se utilizó una formulación para regresión con kernel radial.
- Multilayer Perceptron (MLP) (Rumelhart et al., 1986): red neuronal artificial compuesta por capas de neuronas interconectadas, capaz de aproximar relaciones no lineales complejas entre las covariables ambientales y las variables respuesta.

Se realizó la calibración final de los modelos mediante validación cruzada de 10 particiones. En cada iteración, una parte de los datos fue utilizada para entrenar el modelo y otra parte fue retenida para evaluar su capacidad predictiva. Este procedimiento permitió obtener predicciones independientes para todas las observaciones y calcular métricas comparables entre modelos y variables respuestas. Durante la calibración se evaluaron distintas combinaciones de hiperparámetros para cada algoritmo. En QRF se probaron combinaciones del número de variables candidatas en cada división de los árboles (*mtry*), el tamaño mínimo de nodo y la regla de partición, incluyendo los criterios *variance* y *extratrees*. En Cubist se evaluó el número de comités y el número de vecinos utilizados para ajustar las predicciones. En XGBoost se exploraron combinaciones del número de iteraciones, profundidad máxima de los árboles, tasa de aprendizaje, regularización, proporción de covariables y proporción de observaciones utilizadas en cada iteración. Para SVM se ajustaron los parámetros asociados al kernel radial y a la penalización del modelo. En el caso de MLP, se evaluaron diferentes configuraciones de capas ocultas, número de neuronas por capa y penalización por decaimiento de pesos.

La comparación de modelos se realizó a partir de las predicciones obtenidas en validación cruzada. Para ello se utilizaron métricas complementarias orientadas a evaluar sesgo, precisión, asociación, concordancia y eficiencia predictiva. Este

conjunto de métricas permitió analizar el desempeño de los modelos desde distintas perspectivas y seleccionar el algoritmo más adecuado para la generación de los productos cartográficos finales.

4.4. Métricas de desempeño

El desempeño de los algoritmos de aprendizaje automático fue evaluado mediante el error absoluto medio (MAE), la raíz del error cuadrático medio (RMSE), el coeficiente de concordancia de Lin (CCC) y el coeficiente de eficiencia del modelo (MEC). Estas métricas describen las diferencias entre los valores observados y predichos de la siguiente manera:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

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

$$\rho_c = \frac{2 \cdot \sigma_{\hat{y}} \cdot \sigma_y}{\sigma_{\hat{y}}^2 + \sigma_y^2 + (\mu_{\hat{y}} - \mu_y)^2} \quad (3)$$

$$MEC = 1 - \frac{\sum_{i=1}^N \{y_i - \hat{y}_i\}^2}{\sum_{i=1}^N \{y_i - \mu_y\}^2} \quad (4)$$

donde n , $\hat{y}_i$, y_i , $\sigma_{\hat{y}}$, σ_y , $\mu_{\hat{y}}$, μ_y , y ρ son, respectivamente, el tamaño muestral, los valores predichos, los valores observados, el desvío estándar de los valores predichos, el desvío estándar de los valores observados, la media de los valores predichos, la media de los valores observados y el coeficiente de correlación entre valores predichos y observados.

El MAE y el RMSE evalúan la magnitud del error en las mismas unidades de la variable objetivo, en este caso puntos porcentuales de arcilla, limo o arena. El RMSE penaliza con mayor intensidad los errores grandes, mientras que el MAE expresa el error promedio absoluto de forma más directa. El CCC evalúa la capacidad del modelo para reproducir simultáneamente la asociación y la concordancia entre valores observados y predichos. El MEC cuantifica la eficiencia predictiva del modelo respecto de una predicción de referencia basada en la media de los valores observados. Un valor de MEC igual a 1 indica ajuste perfecto, mientras que un valor igual a 0 indica que el modelo no mejora la predicción obtenida al utilizar la media observada. Adicionalmente, se reportaron el error medio (ME), como indicador de sesgo promedio, y el coeficiente de determinación (R^2), calculado como el cuadrado de la correlación entre valores observados y predichos.

4.5. Predicción e incertidumbre

Una vez calibrados y evaluados los modelos, se seleccionó el algoritmo utilizado para la generación de los mapas finales de arcilla, limo y arena. La selección se realizó a partir del desempeño obtenido en validación cruzada y de la posibilidad de generar indicadores de incertidumbre espacial asociados a las predicciones. La predicción se

realizó sobre una grilla regular de 90 m de resolución, correspondiente a la unidad espacial de los productos cartográficos finales. Para cada celda de la grilla se extrajeron los valores de las covariables ambientales y se estimaron las fracciones granulométricas del horizonte superficial de 0–20 cm. Con el objetivo de facilitar el procesamiento computacional, la predicción se efectuó por teselas, que posteriormente fueron integradas en mosaicos continuos para obtener los mapas provinciales.

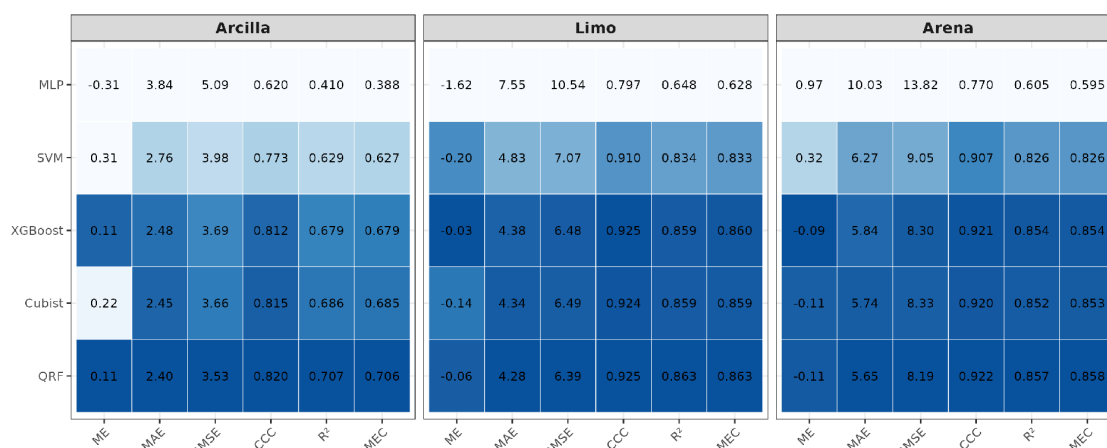
Los mapas finales fueron generados considerando las máscaras previamente definidas en el proyecto. En particular, se excluyeron del producto aquellas celdas correspondientes a cuerpos de agua, zonas urbanas y áreas salinas, donde no se reportan valores de fracciones granulométricas. De este modo, los mapas informan valores únicamente en las áreas consideradas válidas para la representación de las fracciones granulométricas. Además de los mapas de predicción media, se generaron mapas de incertidumbre espacial. La incertidumbre fue representada mediante el desvío estándar condicional de la predicción, expresado en las mismas unidades que la variable predicha, es decir, puntos porcentuales. Valores más altos indican zonas donde el modelo presenta mayor dispersión en la distribución predictiva y, por lo tanto, menor certeza relativa sobre el valor estimado. La interpretación de la incertidumbre debe considerar la densidad y distribución espacial de las muestras, la representatividad de las covariables, la complejidad ambiental del territorio y el grado de similitud entre las condiciones de cada celda y aquellas observadas durante el entrenamiento.

5. Resultados

5.1. Desempeño predictivo de los modelos

La comparación de modelos mostró diferencias entre los algoritmos evaluados, aunque con patrones relativamente consistentes entre las tres fracciones granulométricas. En general, los modelos basados en árboles y ensamblados —QRF, Cubist y XGBoost— presentaron los mejores desempeños predictivos, mientras que SVM mostró un desempeño intermedio y MLP presentó los resultados menos favorables. La Figura N°2 resume el desempeño relativo de los modelos para arcilla, limo y arena, considerando simultáneamente métricas de sesgo, error, concordancia y eficiencia predictiva. La intensidad de color representa el desempeño relativo dentro de cada variable y métrica, mientras que los valores numéricos corresponden a las métricas originales calculadas a partir de la validación cruzada.

Figura N°2. Comparación de modelos mediante métricas de capacidad predictiva. La intensidad de color representa el desempeño relativo de cada modelo dentro de cada variable respuesta, donde valores más intensos indican mejor desempeño. Los valores numéricos dentro de cada celda corresponden a las métricas originales calculadas en validación cruzada. Se presentan las métricas de sesgo medio (ME), error absoluto medio (MAE), raíz del error cuadrático medio (RMSE), coeficiente de concordancia de Lin (CCC), coeficiente de determinación (R^2) y eficiencia del modelo (MEC) para los modelos QRF, Cubist, XGBoost, SVM y MLP, en la predicción de arcilla, limo y arena.



Fuente: Elaboración propia, 2026.

La comparación de modelos mostró un patrón consistente entre las tres variables granulométricas. QRF, Cubist y XGBoost presentaron los desempeños más altos y valores muy próximos entre sí, lo que indica que los enfoques basados en árboles y ensamblados captaron adecuadamente las relaciones no lineales entre las fracciones texturales y las covariables ambientales. En arcilla, QRF alcanzó un RMSE de 3,53 puntos porcentuales, un (R^2) de 0,707 y un MEC de 0,706. En limo y arena, el desempeño predictivo fue mayor, con RMSE de 6,39 y 8,19 puntos porcentuales, (R^2) de 0,863 y 0,857, y MEC de 0,863 y 0,858, respectivamente. Estos resultados sugieren que, para el conjunto de datos y covariables utilizado, las fracciones limo y arena presentaron una estructura espacial más estable o mejor representada por los predictores ambientales disponibles.

SVM presentó un desempeño intermedio, con errores mayores pero aún con niveles aceptables de asociación, concordancia y eficiencia predictiva. En cambio, MLP presentó el menor desempeño en las tres variables, con una reducción más marcada en limo y arena. Este comportamiento podría relacionarse con la mayor sensibilidad del modelo MLP a la parametrización, al tamaño efectivo de la muestra, a la estructura de las covariables y a la necesidad de mayores volúmenes de entrenamiento para estabilizar relaciones no lineales complejas. En función de estos resultados, QRF fue seleccionado como modelo final por su combinación de precisión, estabilidad entre variables y capacidad para estimar incertidumbre condicional.

5.2. Mapas de predicción e incertidumbre

Las estadísticas descriptivas de los mapas de predicción se presentan en la Tabla N°2. En el conjunto de celdas válidas, los valores medios predichos fueron de 16,0% para arcilla, 40,6% para limo y 40,1% para arena. La variabilidad relativa fue mayor en arena (CV = 44%), seguida por limo (CV = 33%) y arcilla (CV = 33%). Los percentiles 5 y 95 indican que el 90% central de las predicciones se ubicó entre 7,4 y 23,9% para arcilla, 20,9 y 60,4% para limo, y 14,8 y 67,8% para arena, lo que resume la amplitud de valores representada en los mapas provinciales.

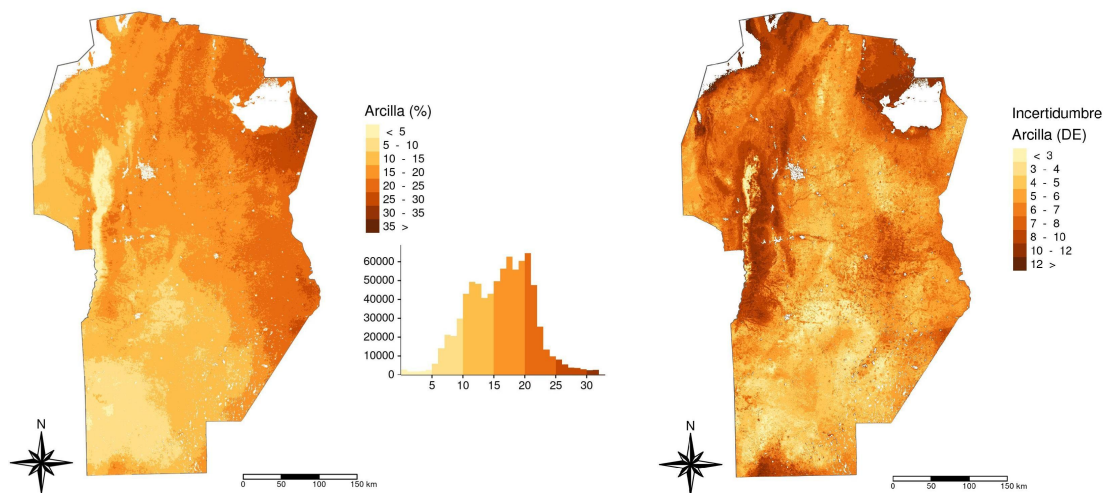
Tabla N°2. Estadísticos descriptivos de la predicción de arcilla, limo y arena. n: número de observaciones; CV: coeficiente de variación (%); Mín: valor mínimo; Max: valor máximo; P(05): percentil 5; P(95): percentil 95.

Variable	n	Media	CV	Min.	Max.	P05	P95
Arcilla	19.316.346	16,0	32	0,1	33,3	7,4	23,9
Limo		40,6	33	0,4	74,3	20,9	60,4
Arena		40,1	44	0,4	87,8	14,8	67,8

Fuente: Elaboración propia, 2026.

La Figura N°3 presenta el mapa de predicción de arcilla junto con su correspondiente mapa de incertidumbre. La arcilla presentó valores mayormente entre 10 y 20% en gran parte del área mapeada. Los contenidos más elevados, generalmente superiores a 20% y localmente cercanos a 25–30%, se observaron en sectores del centro, este y noreste provincial. En contraste, los menores contenidos, por debajo de 10% y en algunos sectores inferiores a 5%, se observaron principalmente hacia el oeste y sudoeste provincial. En este último sector, los bajos contenidos de arcilla se asocian con ambientes medanoso-arenosos correspondientes a la denominada Pampa Medanosa, ampliamente representada en el sudoeste provincial, especialmente en los departamentos Río Cuarto y General Roca. Este patrón es compatible con la diferenciación regional entre ambientes con mayor proporción de materiales finos en las llanuras y ambientes de fracción gruesa y menor desarrollo edáfico hacia el oeste provincial. La incertidumbre de arcilla fue comparativamente menor que la observada para limo y arena, en concordancia con su menor rango absoluto de variación.

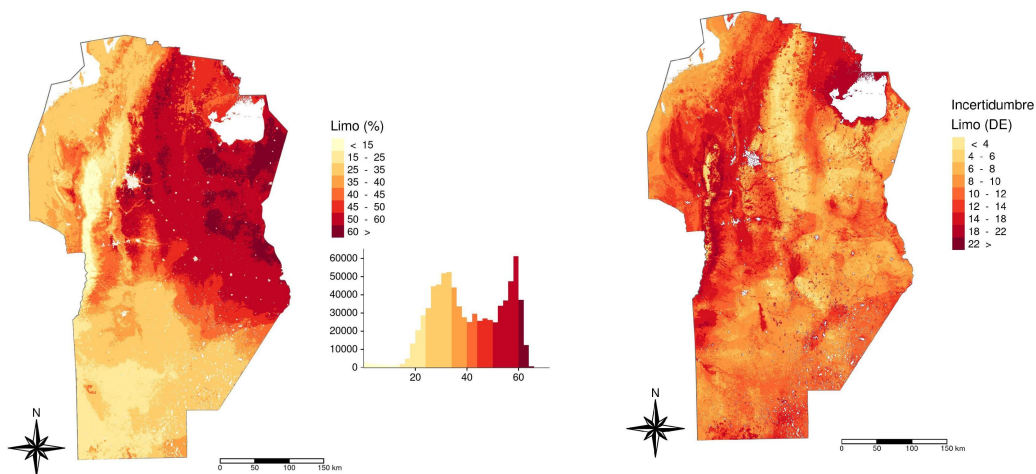
Figura N°3. Predicción del contenido de arcilla (%) e incertidumbre asociada, expresada como desvío estándar condicional de la predicción.



Fuente: Elaboración propia, 2026.

La Figura N°4 muestra la distribución espacial del limo y su incertidumbre asociada. El limo presentó valores frecuentes entre 25 y 60%, con mayores contenidos, superiores a 50% y localmente cercanos o mayores a 60%, en amplias zonas del centro, este y noreste provincial. Este comportamiento es compatible con la importancia de materiales loésicos y limo loessoides en las llanuras de Córdoba, donde la fracción limo representa un componente dominante de una gran proporción de superficie de la provincia. Los menores contenidos de limo, generalmente inferiores a 25–35%, se observaron hacia el oeste y sudoeste, coincidiendo con áreas de mayor proporción relativa de arena. En este sentido, la distribución de limo resultó, en términos generales, complementaria a la de arena.

Figura N°4. Predicción del contenido de limo (%) e incertidumbre asociada, expresada como desvío estándar condicional de la predicción.

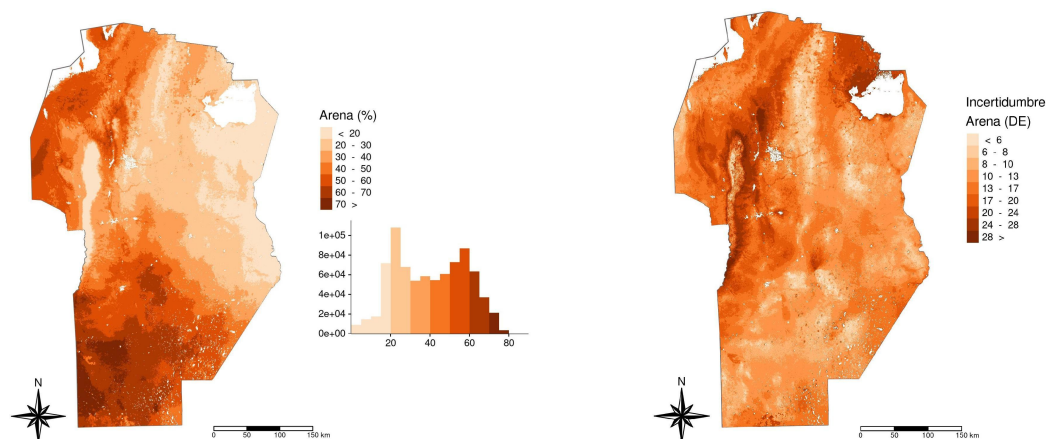


Fuente: Elaboración Propia, 2026.

La Figura N°5 presenta el mapa de arena junto con su mapa de incertidumbre. La arena mostró los mayores contenidos, superiores a 50% y localmente mayores a

60–70%, principalmente hacia sectores del oeste y sudoeste de la provincia. En cambio, los valores más bajos, generalmente inferiores a 20–30%, se ubicaron en áreas del centro, este y noreste, donde predominan fracciones más finas. Este gradiente textural es coherente con la transición entre ambientes serranos, piedemontes y llanuras, así como con diferencias regionales en materiales parentales, transporte y deposición de sedimentos.

Figura N°5. Predicción del contenido de arena (%) e incertidumbre asociada, expresada como desvío estándar condicional de la predicción.



Fuente: Elaboración propia, 2026.

El patrón textural observado tiene implicancias directas para la interpretación de limitantes y potencialidades productivas. Las áreas con mayor proporción de arena suelen presentar menor capacidad de retención de agua y de provisión de nutrientes, mayor drenaje y mayor susceptibilidad a la erosión hídrica y eólica. En contraste, las zonas con predominio de limo y arcilla presentan mayor capacidad de almacenamiento de agua, con mayor susceptibilidad a compactación o anegamiento, según el contexto topográfico y de manejo. En particular, las zonas donde el contenido de limo supera el 50 % requieren especial atención sobre la cobertura del suelo, ya que esta fracción confiere una estabilidad estructural más débil respecto a suelos con mayor contenido de arcilla en el horizonte superficial. Esta condición puede aumentar la susceptibilidad a la erosión hídrica, especialmente en suelos con altos contenidos de limo ubicados en áreas con pendiente, como sectores del piedemonte central y norte. En cuanto a la capacidad de provisión de nutrientes y secuestro de carbono, estos suelos presentan mayor potencial que los ubicados hacia el oeste y sudoeste del área provincial.

Los mapas de incertidumbre muestran que la certeza de las predicciones no es homogénea en todo el territorio. En general, los mayores valores de desvío estándar se concentraron en sectores serranos, piedemontes, áreas de transición geomorfológica y zonas con menor soporte muestral, donde la variabilidad ambiental es más compleja o menos representada en la base de entrenamiento. Por el contrario, las zonas agrícolas de llanura, con mayor densidad de observaciones y gradientes ambientales más continuos, presentaron predicciones más estables. En este sentido, los mapas de incertidumbre deben utilizarse junto con los mapas de predicción para identificar áreas donde las estimaciones requieren una interpretación más cautelosa.

6. Consideraciones finales

El presente trabajo permitió actualizar la cartografía de fracciones granulométricas del suelo de la provincia de Córdoba, generando mapas de arcilla, limo y arena para la capa superficial de 0–20 cm. Los productos fueron desarrollados a partir de una base de datos georreferenciada, covariables ambientales actualizadas y modelos de aprendizaje automático evaluados mediante validación cruzada. Entre los modelos evaluados, QRF, Cubist y XGBoost presentaron los mejores desempeños predictivos, con diferencias reducidas entre ellos. QRF fue seleccionado como modelo final por su combinación de precisión y capacidad para estimar incertidumbre condicional, lo que permitió generar mapas de predicción acompañados por mapas de desvío estándar.

Los mapas obtenidos muestran patrones espaciales coherentes con la heterogeneidad ambiental y geomorfológica de la provincia, y constituyen un insumo relevante para estudios de aptitud productiva, planificación territorial, modelación hidrológica y análisis de limitantes físicas del suelo. La información de incertidumbre permite identificar áreas donde las predicciones deben interpretarse con mayor cautela y orientar futuros esfuerzos de muestreo. Los mapas de textura generados a 90 m constituyen una mejora sustancial respecto del antecedente de 2021 a 500 m, ya que permiten representar gradientes y transiciones espaciales con mayor detalle. No obstante, la escala de uso sigue siendo regional o provincial. En este sentido, deben emplearse como información de apoyo a la caracterización física de los suelos, no como reemplazo de análisis granulométricos de detalle, diagnósticos prediales específicos o decisiones de manejo de alta precisión.

7. Referencias bibliográficas

- Allen R.G., Pereira S.L., Raes D. & Smith M. (1998) FAO Irrigation and Drainage Paper No. 56. Irrigation and Drainage, 56, No. 97. Food and Agriculture Organization of the United Nations, Rome.
- Amatulli, G., McInerney, D., Sethi, T., Strobl, P., & Domisch, S. (2020). Geomorpho90m, empirical evaluation and accuracy assessment of global high-resolution geomorphometric layers. *Scientific Data*, 7(1), 162.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Córdoba, M., Bozzer, C., Morales, H., Álvarez, C., Carranza, J., & Monzani, F. (2021). Mapeo de Propiedades de Suelo en la Provincia de Córdoba, 2021. *Infraestructura de Datos Espaciales de Córdoba (IDECOR)*.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18-27.
- IDECOR (2023). Mapa de Cobertura y Uso de Suelo de la Provincia de Córdoba 2022-2023. *Infraestructura de Datos Espaciales de Córdoba (IDECOR)*, noviembre de 2023.
- Karger, D. N., Conrad, O., Böhrner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, 4(1), 170122.
- Keys, R. (1981). Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(6), 1153-1160. <https://doi.org/10.1109/TASSP.1981.1163711>
- Kuhn, M., & Johnson, K. (2013). An Introduction to Feature Selection. En *Applied Predictive Modeling* (pp. 487-519). Springer New York. https://doi.org/10.1007/978-1-4614-6849-3_19
- Kuhn, M., Weston, S., Keefer, C., & Coulter, N. (2016). Cubist Models For Regression. *R package Vignette R package version 0.0*, 1992.
- Malone, B. P., Mcbratney, A. B., Minasny, B., & Laslett, G. M. (2009). Geoderma Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma*, 154(1-2), 138-152.
- McBratney, A. B., Santos, M. L. M., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117(1-2), 3-52.
- Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, 7(Jun), 983-999.

- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536. <https://doi.org/10.1038/323533a0>
- R Core Team. (2025). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14, 199–222.
- Yigini, Y., Olmedo, G. F., Reiter, S., Baritz, R., Viatkin, K., & Vargas, R. (2018). Soil Organic Carbon Mapping Cookbook 2nd edition (R. Olmedo, G.F., Reiter, S., Baritz, R., Viatkin, K. and Vargas, Ed.; FAO).

Organismos participantes y colaboradores:

