

## MODELOS DE DISTRIBUCIÓN ESPACIAL DE HIDROCARBUROS EN SUELOS CONTAMINADOS Y SU IMPLEMENTACIÓN EN SOFTWARE SIG. UNA REVISIÓN

### SPATIAL DISTRIBUTION MODELS OF HYDROCARBONS IN CONTAMINATED SOILS AND THEIR IMPLEMENTATION IN GIS SOFTWARE. A REVIEW

Román Gamarra Torres<sup>1</sup>  Luis Moncada Torres<sup>1</sup> 

<sup>1</sup>Universidad Nacional de Trujillo, Perú.

**Correspondencia:** Dr. Luis Moncada Torres  
**lmoncadat@unitru.edu.pe**

**Como citar este artículo:** Gamarra, R., & Moncada, L. (2025). Modelos de distribución espacial de hidrocarburos en suelos contaminados y su implementación en software SIG. Una revisión. *Revista de Investigación Hatun Yachay Wasi*, 4(2), 106–118. DOI: 10.57107/hyw.v4i2.101

#### RESUMEN

Esta revisión analiza y compara los principales modelos empleados para representar la distribución espacial de hidrocarburos en suelos contaminados, con énfasis en su implementación en entornos de Sistemas de Información Geográfica (SIG). Se identificaron dos enfoques principales: métodos geoestadísticos como el *Kriging* ordinario y modelos de aprendizaje automático como *Random Forest* y redes neuronales 3D. El *Kriging* ordinario destaca por su equilibrio entre precisión, manejo de incertidumbre y compatibilidad con SIG, frente a la mayor complejidad de modelos avanzados. Se concluye que, el *Kriging* ordinario es el modelo más adecuado para representar este tipo de contaminación, y se recomienda su implementación con herramientas en Python, para lograr un análisis automatizado, preciso y accesible.

**Palabras clave:** Kriging ordinario; hidrocarburos; suelos contaminados; SIG; Python.

#### ABSTRACT

This review analyzes and compares the main models used to represent the spatial distribution of hydrocarbons in contaminated soils, with an emphasis on their implementation in Geographic Information Systems (GIS) environments. Two main approaches were identified: geostatistical methods such as ordinary Kriging and machine learning models such as Random Forest and 3D neural networks. Ordinary Kriging stands out for its balance between accuracy, uncertainty management, and GIS compatibility, compared to the greater complexity of advanced models. It is concluded that ordinary Kriging is the most appropriate model to represent this type of contamination, and its implementation with Python tools is recommended for achieving an automated, accurate, and accessible analysis.

**Keywords:** ordinary kriging, hydrocarbons; contaminated soils; GIS; Python.



## INTRODUCCIÓN

La contaminación del suelo por hidrocarburos representa una amenaza significativa para los ecosistemas y la salud humana, especialmente en zonas donde se desarrollan actividades extractivas, industriales o de almacenamiento de combustibles fósiles. La detección y análisis espacial de estos contaminantes no solo permite comprender su comportamiento en el subsuelo, sino también optimizar los planes de remediación y gestión ambiental. En este contexto, los modelos de distribución espacial se convierten en herramientas fundamentales para predecir y representar la heterogeneidad de los contaminantes en el terreno (Liang et al., 2021).

Los hidrocarburos del petróleo, incluyendo fracciones como F2 y F3, tienden a distribuirse de manera no uniforme debido a la influencia de variables como la textura del suelo, las condiciones topográficas, la estructura geológica y el origen del contaminante (Luca et al., 2007). Dado que las campañas de muestreo son costosas y no siempre alcanzan una cobertura densa, los modelos geoestadísticos como el kriging ordinario, kriging indicador e incluso simulaciones secuenciales han demostrado ser útiles para estimar la distribución de contaminantes en zonas no muestreadas (Atteia et al., 1994; Bengtsson & Törneman, 2009). Además, el uso de modelos basados en inteligencia artificial como los algoritmos *Random Forest*, redes neuronales y, más recientemente, redes convolucionales tridimensionales (3D-CNN), ha ganado relevancia por su capacidad para manejar datos no lineales y mejorar la precisión de las predicciones (Liu et al., 2020; Miao et al., 2025).

La combinación de estos enfoques con datos ambientales auxiliares como topografía, vegetación, uso del suelo y datos espectrales ha dado lugar a modelos integrados que alimentan plataformas SIG con información geoespacialmente precisa (Miao et al., 2025).

La implementación de estas herramientas en SIG permite generar mapas temáticos, analizar patrones espaciales de riesgo y tomar decisiones fundamentadas para la delimitación de zonas críticas y el diseño de estrategias de remediación (Stewart & Purucker, 2011).

Esta revisión bibliográfica tuvo como objetivo identificar los modelos existentes de distribución espacial aplicados al estudio de suelos contaminados por hidrocarburos; así como, sus formas de implementación en plataformas de Sistemas de Información Geográfica (SIG).

## MATERIALES Y MÉTODOS

### **Búsqueda de información**

La búsqueda se basó en los artículos publicados desde enero, 2015 hasta enero, 2025 que se encuentren en la base de datos en línea Science Direct en revistas indizadas en *Journal Citation Reports*, *Web of Science* o *Scopus*, con una clasificación mínima en el cuartil Q3 según *SCImago Journal Rank*. El idioma de los estudios seleccionados fue inglés.

Se usaron las siguientes palabras clave con los términos “hydrocarbon-contaminated soil”, “spatial distribution models”, “GIS implementation”, “kriging”, “machine learning”, y “remediation”. Se realizaron combinaciones usando los operadores booleanos “AND” y “OR”. Esta combinación permitió identificar estudios enfocados en la distribución espacial de hidrocarburos en suelos contaminados, el uso de herramientas de predicción y su integración con Sistemas de Información Geográfica (SIG), Figura 1.

### **Criterios de Elegibilidad**

#### **Criterios de inclusión**

- Los estudios seleccionados debían cumplir debían cumplir simultáneamente con los siguientes criterios:

- Analizar contaminación de suelos por hidrocarburos o contaminantes orgánicos persistentes afines (PAHs, DDT, HCH, etc).
- Aplicar algún modelo de distribución espacial (geoestadísticos, determinísticos), basado en aprendizaje automático o híbrido.
- Usar datos empíricos provenientes de muestreo ambiental real.
- Reportar métricas cuantitativas de validación del modelo (como R<sup>2</sup>, RMSE, MAE, MRE).
- Incluir alguna forma de implementación o integración en software SIG, o mencionar claramente la exportación de resultados a dicho entorno.

- No incluían validación numérica del modelo aplicado.
- Presentaban únicamente modelos conceptuales sin análisis geográfico cuantitativo.

**Extracción y análisis de datos**

La información fue extraída de los artículos seleccionados y se diseñó la tabla 1 con las siguientes características:

Primer autor y año de publicación; modelo espacial utilizado aplicado a suelos contaminados con hidrocarburos (*kriging*, IDW, 3D-CNN, RFOK, Bi-LSTM, simulaciones secuenciales); tipo (geoestadístico, determinístico, híbrido, etc); Indicadores de precisión y validación reportados; nivel de integración en SIG (directa, parcial o nula); incorporación de variables auxiliares y facilidad de implementación de SIG.

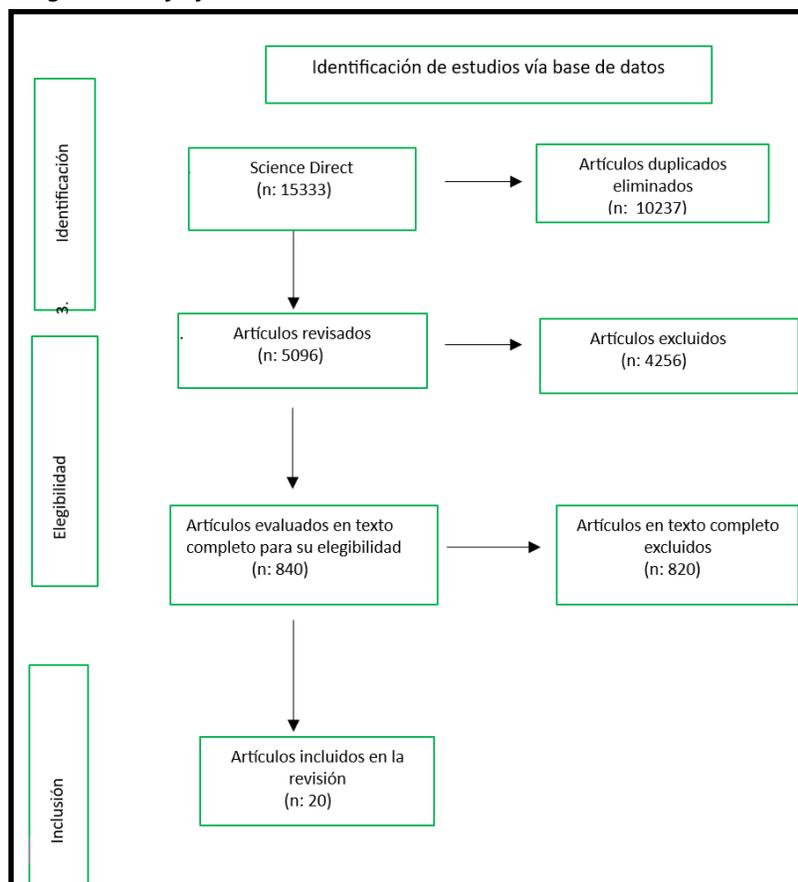
**Criterios de exclusión**

Fueron excluidos los estudios que:

- Carecían de un enfoque espacial explícito en su metodología.

**FIGURA 1**

*Diagrama de flujo de Revisión sistematica*



**TABLA 1**

*Modelos de distribución espacial de hidrocarburos en suelos contaminados*

Autor/año	Modelo	Tipo	Indicadores de Precisión	Nivel de integración en SIG	Incorporación de variables auxiliares	Facilidad de implementación SIG
Liu et al. (2017); Metahni et al. (2019) Ouabo et al. (2020)	Kriging ordinario (OK)	Geoestadístico clásico	Alta (en zonas bien muestreadas)	Moderado (usa semivariogramas)	Baja	Alta
Liu et al. (2020) ; Ouabo et al. (2020)	IDW	Determinístico	Moderada dependiente de p	Nula	Baja	Muy alta
Metahni et al. (2019)	SGS / SIS	Geoestadístico (simulación)	Alta	Alta (representación probabilística)	Baja	Media
Liu et al. (2020) ; Parvizi & Fatehi, (2025)	Random Forest (RF)	Aprendizaje automático	Muy alta	Baja (modelo de caja negra)	Alta	Media
Liu et al. (2020)	RFOK (híbrido)	Híbrido RF + OK	Muy alta (mejor en validación)	Alta (usa residuos del RF)	Muy alta	Media
Miao et al. (2025)	3D-CNN	Aprendizaje profundo	Muy alta (R <sup>2</sup> > 0.95)	Baja (difícil interpretación)	Muy alta	Baja (requiere más cómputo)
Parvizi & Fatehi, (2025)	SVM / Cubist	Aprendizaje automático	Alta (dependiente de datos)	Baja	Alta	Media
Gao et al. (2024) (JIPS); Gao et al. (2024) (Springer)	FNN	Aprendizaje profundo	Muy alta	Baja	Alta	Media
Mahanty et al. (2025)	ANN híbrido	Híbrido ANN + físico	Muy alta (R <sup>2</sup> > 0.99)	Media (por validación cruzada)	Alta	Media
Gu et al. (2020)	LIF 3D modelado	Sensor + estadístico	Alta (validación experimental)	Baja	Media	Alta
Hyde et al. (2019)	Incremental sampling (ISM)	Estadístico/ empírico	Alta (representatividad mejorada)	Media (control por replicabilidad)	Baja	Alta
Karaca, (2016); Yu et al. (2019)	Geoestadístico PAH (clásico)	Geoestadístico descriptivo	Alta (validación de campo)	Baja	Baja	Alta
Zhang et al. ( 2022)	POP mapping	Geoestadístico + diagnóstico	Alta (en casos urbanos e industriales)	Media (uso de herramientas analíticas)	Media	Alta

**Modelos de distribución espacial identificados**

En el análisis sistemático de los modelos de distribución espacial utilizados para representar la variabilidad de hidrocarburos en suelos

contaminados, se identificaron dos grandes enfoques: (i) modelos de interpolación espacial tradicionales, basados principalmente en métodos geoestadísticos y determinísticos (Fig. 2) y (ii)

modelos basados en técnicas de aprendizaje automático. A continuación, se describen los modelos más representativos en la primera categoría.

**1.- Kriging Ordinario (2D y 3D)**

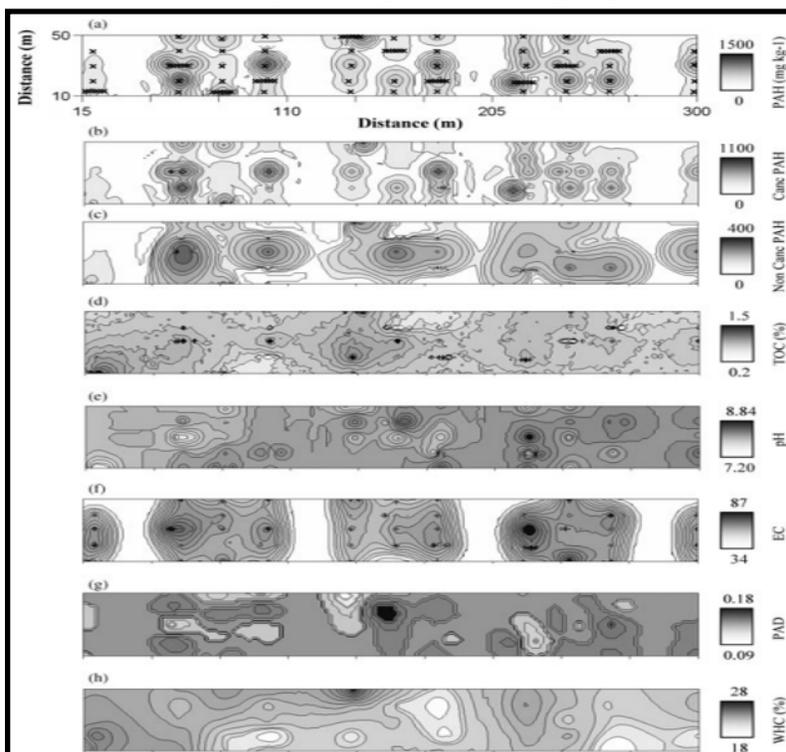
El kriging ordinario (OK, por sus siglas en inglés) ha sido uno de los métodos de interpolación más aplicados en estudios de suelos contaminados por hidrocarburos. Este modelo asume que la variable a interpolar es una realización de un proceso aleatorio estacionario con media constante, y utiliza semivariogramas para capturar la dependencia espacial (Liu et al., 2017). Estos autores utilizaron OK tanto en dos como en tres dimensiones para analizar la distribución de hidrocarburos aromáticos

policíclicos (PAHs) en suelos contaminados alrededor de una planta química en Chongqing, China, concluyendo que el modelo 3D proporcionó una estimación más precisa al incorporar la elevación como variable auxiliar significativa.

Asimismo, el estudio de Ouabo et al. (2020) comparó el desempeño del kriging ordinario frente al método de distancia inversa ponderada (IDW) para estimar la concentración de cromo (Cr) y cadmio (Cd) en suelos contaminados por residuos electrónicos. Encontraron que OK fue más preciso para Cr, con menores errores cuadráticos medios (RMSE), aunque en zonas urbanas con alta densidad de muestreo IDW resultó competitivo o incluso más representativo en mapas visuales.

**FIGURA 2**

*Patrones espaciales de las propiedades fisicoquímicas del suelo estimados mediante interpolación geoestadística (SGS).*



Nota: (a) concentración de HAP totales (mg/kg); (b) concentración de HAP cancerígenos (mg/kg); (c) concentración de HAP no cancerígenos (mg/kg); (d) carbono orgánico total (COT) (% de peso seco); (e) pH; (f) conductividad eléctrica (CE) (µS/cm); (g) distribución de agregados de partículas (DAP) (mm); (h) capacidad de retención de agua (CRA) (%). Las cruces en (a) representan las posiciones de las muestras. Se inserta una guía de escala cuantitativa a la derecha del mapa. Tomado de Bengtsson & Törneman, (2009)

En otro caso, Metahni et al. (2019) aplicaron OK junto con simulaciones gaussianas secuenciales para estimar el volumen de suelo contaminado por múltiples contaminantes (As, Cr, Cu, PCP y dioxinas/furanos). Se observó que, si bien OK ofrecía buenos estimadores de la media, la simulación gaussiana era superior para representar la incertidumbre espacial y calcular probabilidades de superar umbrales normativos.

En el trabajo de Atteia et al. (1994), se presenta el sistema clásico de ecuaciones del kriging ordinario:

**Sistema de ecuaciones de kriging**

$$\sum_{i=1}^n \lambda_i \gamma(x_i, x_j) + \mu = \gamma(x_j, B) \quad \text{para todo } j \quad (1)$$

donde:

$\lambda_i$  son los pesos para determinar

$\gamma(x_i, x_j)$  es la semivarianza entre puntos  $x_i$  y  $x_j$

$\mu$  es el multiplicador de Lagrange

$\gamma(x_i, B)$  es la semivarianza entre el punto  $x_j$  y el bloque  $B$

Varianza de estimación

$$\sigma^2(B) = \sum_{i=1}^n \lambda_i \gamma(x_i, B) + \mu - \gamma(B, B) \quad (2)$$

El semivariograma experimental se expresa como:

$$\gamma(h) = \frac{1}{2m(h)} \sum_{i=1}^{m(h)} [z(x_i) - z(x_i + h)]^2 \quad (3)$$

donde:

$\gamma(h)$  es la semivarianza para el intervalo de separación  $h$ ,

$z(x_i)$  es el valor observado en la ubicación  $x_i$ ,

$m(h)$  es el número de pares de puntos separados por la distancia  $h$ .

**2.- Interpolación por Distancia Inversa (IDW)**

El IDW se basa en la premisa de que los puntos cercanos tienen más influencia sobre un punto desconocido que los distantes. Este método es

ampliamente usado debido a su simplicidad computacional y su integración en múltiples plataformas SIG. El estudio de Ouabo et al. (2020), el IDW proporcionó resultados comparables al kriging para la interpolación de Cd, con un desempeño similar en términos de error medio absoluto (MAE) y RMSE. Además, se destacó que, al ajustar el parámetro de potencia, IDW puede generar mapas más suaves o abruptos, dependiendo del contexto. Cabe resaltar que, en sitios con baja densidad de muestreo, IDW puede subestimar las variaciones reales del contaminante, como fue observado por Liu et al. (2020), quienes indicaron que modelos más complejos como RFOK superan las limitaciones del IDW.

Según Metahni et al. (2019), el modelo IDW se expresa mediante:

$$\hat{Z}(S_0) = \frac{\sum_{i=1}^n Z(S_i) / d(S_i, S_0)^p}{\sum_{i=1}^n 1 / d(S_i, S_0)^p} \quad (4)$$

donde:

$\hat{Z}(S_0)$  es el valor interpolado en el punto  $S_0$ ,

$Z(S_i)$  son los valores observados,

$d(S_i, S_0)$  es la distancia entre el punto conocido y el estimado,

$p$  es un parámetro de potencia (habitualmente 2).

**3.- Simulación Gaussiana Secuencial (SGS) y Simulación Indicadora Secuencial (SIS)**

Las técnicas de simulación geoestadística, como SGS y SIS, son valiosas cuando se requiere evaluar la incertidumbre espacial de los contaminantes. Por ejemplo, en el estudio de Luca et al. (2007), se empleó SIS para evaluar simultáneamente la probabilidad de superar umbrales normativos en diferentes fracciones de hidrocarburos del petróleo (TPH-F2 y TPH-F3). Se concluyó que, en comparación con el kriging indicativo, SIS permitió una representación más realista de las áreas contaminadas, especialmente cuando se consideraban múltiples fracciones de contaminantes de forma conjunta (casos dependientes e independientes).

#### 4.- Modelos Polinomiales y Regresiones

Algunos estudios, como el de Liu et al. (2017), han combinado el uso de kriging con análisis de componentes principales (PCA) y regresiones polinomiales, permitiendo identificar patrones dominantes de contaminación y reducir dimensionalidad. Estos enfoques se integran bien en entornos SIG y contribuyen a delimitar áreas prioritarias de remediación.

#### 5.-LIF 3D modelado de difusión

Gu et al. (2020) emplearon fluorescencia inducida por láser (LIF) para modelar la distribución tridimensional de diésel en suelos, considerando la influencia de la humedad. El estudio permitió cuantificar volúmenes contaminados en distintas condiciones de saturación, generando modelos 3D de penetración y dispersión lateral con precisión experimental validada (RSD < 10%).

#### 6.- Incremental Sampling Methodology (ISM)

Hyde et al. (2019) aplicaron la metodología de muestreo incremental (ISM) como alternativa representativa frente al muestreo puntual tradicional en sitios contaminados con PHCs. El estudio comparó ISM y evaluación de fase II (ESA), revelando que ISM proporciona estimaciones más realistas y reproducibles de volumen contaminado, recomendándose para la planificación de remediación basada en riesgo.

#### 7.- Geoestadística clásica para PAHs

Karaca (2016) realizó un análisis detallado de la distribución espacial de PAHs en suelos urbanos de Turquía utilizando técnicas geoestadísticas convencionales. Se identificaron fuentes de emisión pirogénica y petrogénica mediante relaciones moleculares, destacando la utilidad de este enfoque para la evaluación de fuentes y la priorización de intervención.

#### 8.- Modelos de mapeo de POPs

Zhang et al. (2022) estudiaron la distribución

espacial de DDTs, HCHs y PAHs en suelos urbanos de Guiyang, China, identificando fuentes múltiples de contaminación y variación espacial significativa asociada a densidad poblacional e intensidad industrial. El estudio aplicó métodos de diagnóstico fuente y análisis geoespacial para categorizar zonas críticas de intervención ambiental.

#### *Modelos basados en aprendizaje automático y enfoques híbridos*

En la última década, el aprendizaje automático ha emergido como una herramienta poderosa en la predicción de la distribución espacial de contaminantes del suelo, particularmente en escenarios donde los patrones de dispersión son no lineales y multivariados. Estos métodos permiten incorporar múltiples fuentes de datos ambientales, como información topográfica, climática, geológica y remota, lo que ha mejorado sustancialmente la precisión en la estimación de la contaminación por hidrocarburos.

#### **1.-Random Forest (RF) y modelos híbridos RFOK**

El modelo *Random Forest* (RF) ha demostrado ser especialmente eficaz para representar la variabilidad espacial de contaminantes debido a su capacidad para manejar relaciones no lineales y altos volúmenes de datos. En el estudio de Liu et al. (2020), se propuso un modelo híbrido denominado RFOK, que combina la predicción generada por el modelo *Random Forest* con el kriging ordinario. Esta fusión permitió aprovechar la fortaleza del RF en captar relaciones complejas entre el contenido de arsénico en suelos y factores ambientales (topografía, uso del suelo, imágenes satelitales), y del OK para refinar la predicción espacial con base en la correlación espacial de residuos.

Los resultados de validación cruzada mostraron que RFOK fue superior a los métodos individuales (RF, OK, IDW, y regresión por pasos), con una reducción del error cuadrático medio (RMSE) del 62,2 % vs. RF, y del 83,6 % vs. IDW. Esta mejora sustancial

evidencia que los modelos híbridos pueden superar las limitaciones de los modelos tradicionales, en especial cuando se busca alta resolución en zonas de variabilidad espacial compleja.

**2.-Redes Neuronales Convolucionales 3D (3D-CNN)**

Otro avance significativo en este campo lo constituye el uso de redes neuronales convolucionales tridimensionales (3D-CNN), como se describe en el trabajo de Miao et al. (2025). En su investigación, se aplicó una 3D-CNN equipada con un mecanismo de atención por canal (CAM) para predecir la distribución de hidrocarburos del petróleo en suelos contaminados. Este enfoque permitió conservar y explotar la estructura tridimensional de los datos de perforación, evitando la pérdida de información espacial que ocurre al transformar los datos en vectores planos (una limitación frecuente en modelos tradicionales).

En comparación con métodos como Kriging 3D ( $R^2 = 0.318$ ) y Support Vector Regression ( $R^2 = 0.582$ ), el modelo 3D-CNN logró un coeficiente de determinación de  $R^2 = 0.954$ , reflejando una capacidad sobresaliente para capturar la complejidad espacial de los contaminantes y su distribución vertical y horizontal. Este resultado posiciona a las redes neuronales como herramientas clave para el desarrollo de software SIG de nueva generación orientado a la predicción y gestión de suelos contaminados.

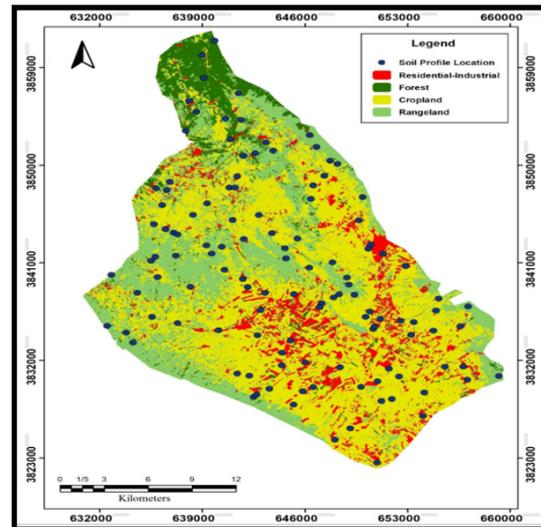
**3.- Support Vector Machines (SVM), Cubist y otros**

Parvizi & Fatehi, (2025) aplicaron diferentes algoritmos de aprendizaje supervisado, incluyendo SVM, Cubist, GLM, y LAM, para predecir el contenido de carbono orgánico en suelos en función del uso del suelo y otras variables ambientales. Aunque su aplicación no fue específicamente para hidrocarburos, sus resultados permiten extrapolar su potencial para estos contaminantes. En particular, el modelo *Random Forest* presentó el mejor desempeño ( $R^2 = 0.64$ ), seguido por SVM y

Cubist, lo que confirma su robustez, para tareas de predicción espacial multivariable, Figura 3.

**FIGURA 3**

*Distribución espacial de los perfiles de suelo en el mapa de uso del suelo del área de estudio*



Nota: Tomado de Parvizi & Fatehi, (2025)

**4.- Modelos FNN (Red Neuronal Feedforward)**

Modelos basados en redes neuronales *Feedforward* (FNN) han sido aplicados para predecir la distribución espacial y la migración vertical de hidrocarburos del petróleo en suelos industriales. Gao, Chen, et al. (2024) y Gao, Cheng, et al. (2024) demostraron que FNN supera en precisión a métodos tradicionales como kriging, al modelar patrones no lineales influenciados por el flujo subterráneo y la profundidad del acuífero, con visualización tridimensional de contaminantes en zonas de almacenamiento y producción.

**5.-ANN híbrido aplicado a bioprocesos**

Mahanty et al. (2025) propusieron un modelo híbrido con redes neuronales artificiales (ANN) para predecir la dinámica de bioremediación in situ de suelos contaminados con diésel, integrando condiciones iniciales (C/N, humedad, concentración) y tiempo. El modelo alcanzó un  $R^2 > 0.99$  en la simulación de tasas de remoción, destacando su utilidad para procesos con datos

limitados y alta variabilidad.

### **Análisis comparativo de modelos de distribución espacial**

La diversidad de enfoques metodológicos para modelar la distribución espacial de hidrocarburos en suelos contaminados refleja la complejidad inherente del fenómeno que se pretende representar (Tabla 1). Cada método posee ventajas particulares en términos de precisión, capacidad de incorporar variables auxiliares, manejo de incertidumbre y facilidad de implementación en entornos SIG. Para facilitar una comparación sistemática, esta sección presenta un análisis comparativo de los principales modelos identificados, agrupados en dos grandes categorías: métodos tradicionales de interpolación y modelos basados en aprendizaje automático.

Los métodos tradicionales como el Kriging ordinario y la Interpolación por distancia inversa (IDW) han sido ampliamente utilizados debido a su disponibilidad en software SIG y su fundamentación estadística o determinística. Por otro lado, los modelos modernos como *Random Forest*, 3D-CNN y modelos híbridos como RFOK integran información multifuente y capturan relaciones no lineales, lo que ha resultado en mejoras sustanciales en la precisión y capacidad predictiva.

El Kriging ordinario sigue siendo una herramienta confiable en escenarios donde se cuenta con datos densos y distribuidos homogéneamente, especialmente en estudios exploratorios o preliminares (García et al., 2021; Karaca, 2016). Sin embargo, sus limitaciones emergen en contextos donde la relación entre los contaminantes y las variables ambientales es no lineal o cuando los datos de muestreo son escasos o dispersos.

Por su parte, el modelo IDW, a pesar de su simplicidad, presenta un rendimiento aceptable en zonas urbanas de alta densidad (Ouabo et al., 2020; Zhang et al., 2018), pero tiende a suavizar

los resultados y no representa adecuadamente la variabilidad espacial ni la incertidumbre.

En contraste, los modelos basados en aprendizaje automático como RF, RFOK y 3D-CNN ofrecen una capacidad superior, para manejar complejidad espacial, grandes volúmenes de datos y relaciones no lineales. El modelo RFOK, al combinar la potencia de predicción del *Random Forest* con la capacidad de interpolación local del Kriging, se destaca como el más robusto para representar la distribución espacial en áreas contaminadas por hidrocarburos (Kebonye et al., 2021; Liu et al., 2020). Por su parte, el 3D-CNN ofrece un salto cualitativo al modelar en tres dimensiones, capturando la distribución vertical con alta precisión (Miao et al., 2025; Rakotonirina et al., 2025); aunque su implementación en entornos SIG convencionales aún requiere adaptación técnica.

### **DISCUSIÓN**

Los resultados de esta revisión evidencian la existencia de múltiples enfoques para modelar la distribución espacial de hidrocarburos en suelos contaminados, cada uno con fortalezas particulares según el contexto ambiental, la calidad del muestreo y los objetivos del análisis. En general, se observa que los modelos geoestadísticos como el Kriging ordinario (OK) siguen siendo ampliamente utilizados por su fundamento teórico sólido y su integración en plataformas SIG como ArcGIS. Este método es especialmente eficaz cuando se cuenta con una red de muestreo densa y homogénea, ya que permite estimar valores desconocidos en función de la autocorrelación espacial de los datos. Sin embargo, su rendimiento disminuye en presencia de alta heterogeneidad espacial o datos dispersos, donde la estructura del semivariograma se vuelve difícil de modelar (Ouabo et al., 2020; Liu et al., 2017).

Por su parte, la interpolación por distancia inversa (IDW) se presenta como un método de

fácil implementación y bajo costo computacional. Aunque no incorpora relaciones espaciales complejas, su desempeño puede ser aceptable en áreas urbanas densamente muestreadas o en etapas exploratorias del análisis. No obstante, estudios como los de Ouabo et al. (2020) y Metahni et al. (2019), han mostrado que IDW tiende a suavizar las variaciones del contaminante y no representa adecuadamente la incertidumbre espacial, lo cual limita su aplicabilidad en estudios de riesgo ambiental.

Frente a estas limitaciones, los modelos basados en aprendizaje automático han ganado terreno por su capacidad para manejar relaciones no lineales, integrar múltiples variables ambientales y mejorar la precisión de las predicciones. El modelo *Random Forest* (RF), por ejemplo, ha demostrado ser eficaz en la predicción de contaminantes como arsénico en suelos, especialmente cuando se dispone de datos auxiliares como topografía, uso del suelo o índices espectrales. Aún más prometedor es el modelo híbrido RFOK, que combina la predicción robusta de RF con la capacidad del kriging, para capturar la estructura espacial de los residuos, logrando una reducción significativa de los errores de predicción (Liu et al., 2020).

En cuanto a los modelos más avanzados, el uso de redes neuronales convolucionales tridimensionales (3D-CNN) ha marcado un hito en la representación de la distribución vertical de hidrocarburos en el suelo. A diferencia de los métodos convencionales, que suelen perder información al vectorizar datos espaciales, el 3D-CNN conserva la estructura tridimensional del terreno y logra altos niveles de precisión ( $R^2 > 0.95$ ), como lo demuestra el estudio de Miao et al. (2025). No obstante, su implementación práctica aún enfrenta desafíos, como la necesidad de recursos computacionales elevados y la dificultad para integrarse de forma directa en SIG convencionales.

Un aspecto clave que diferencia a los modelos geoestadísticos de los de aprendizaje automático es la capacidad de estimar la incertidumbre. Mientras que métodos como la simulación gaussiana secuencial (SGS) o la simulación indicadora secuencial (SIS) permiten generar múltiples realizaciones y evaluar probabilidades de excedencia, los modelos de aprendizaje automático tienden a comportarse como “cajas negras” y no brindan medidas explícitas de confianza. Esta limitación puede ser crítica en contextos regulatorios o cuando se requiere una evaluación detallada del riesgo (Buzzi et al., 2017; Ouabo et al., 2020; Stewart & Purucker, 2011).

Cabe destacar que la elección del modelo también depende de su facilidad de integración con sistemas SIG. En este sentido, modelos como OK e IDW tienen la ventaja de estar plenamente implementados en entornos como ArcGIS o QGIS, lo que facilita su uso por profesionales sin formación avanzada en programación (Ouabo et al., 2020; Stewart & Purucker, 2011). Por el contrario, modelos como RFOK o 3D-CNN requieren procesos adicionales de integración, lo cual podría ser una barrera en aplicaciones operativas. Sin embargo, esta limitación técnica se atenúa progresivamente con la expansión de bibliotecas como *scikit-learn* o *TensorFlow*, y el desarrollo de APIs compatibles con SIG (Miao et al., 2025; Rakotonirina et al., 2025).

Finalmente, enfoques innovadores como el modelado 3D mediante fluorescencia inducida por láser (Gu et al., 2020), o la aplicación de metodologías de muestreo incremental (Hyde et al., 2019), aportan nuevas herramientas para caracterizar de forma más realista la distribución espacial de hidrocarburos, especialmente en contextos donde la representatividad del muestreo es crítica. Además, estudios como los de Zhang et al. (2022) sobre mapeo espacial de contaminantes orgánicos persistentes (DDT, HCH, PAHs) y los de Karaca (2016) mediante geoestadística aplicada a

PAHs, refuerzan la utilidad de estos modelos en la toma de decisiones ambientales.

Los hallazgos de esta revisión proporcionan una base sólida, para orientar el desarrollo de un software especializado para la evaluación de la variabilidad espacial de suelos contaminados por hidrocarburos. La incorporación de modelos híbridos, capacidades de simulación de incertidumbre, y herramientas de visualización 3D, se perfila como una estrategia clave para dotar a dicho software de precisión técnica, valor predictivo y aplicabilidad práctica en escenarios reales de gestión ambiental.

### CONCLUSIONES

- La presente revisión ha permitido identificar, clasificar y analizar comparativamente los principales modelos utilizados para la distribución espacial de hidrocarburos en suelos contaminados; así como, su integración en entornos de Sistemas de Información Geográfica (SIG).
- El kriging ordinario se considera un método principal de interpolación, dadas sus ventajas en términos de precisión, fundamentación teórica y facilidad de implementación en SIG, para representar la distribución de contaminantes en contextos ambientales.
- El Kriging también incorpora una estructura de incertidumbre a través del análisis del semivariograma, lo que lo convierte en una herramienta valiosa, para la planificación de intervenciones y toma de decisiones ambientales. u integración directa en entornos SIG (como ArcGIS y QGIS); lo que permite una visualización y análisis espacial intuitivos, constituyéndose en una opción ideal para estudios aplicados de gestión ambiental.
- Los modelos como *Random Forest* o redes neuronales tridimensionales (3D-CNN) pueden

superar al kriging en términos de precisión predictiva; no obstante, presentan una mayor complejidad computacional, menor interpretabilidad estadística y una integración más limitada con plataformas SIG tradicionales.

- Para enriquecer esta implementación, se propone el uso de herramientas programáticas en Python, mediante bibliotecas como PyKrige, Geopandas, Rasterio, Scikit-learn y ArcPy. Estas permitirán automatizar el flujo de trabajo, desarrollar visualizaciones dinámicas, evaluar la estructura espacial de los datos y eventualmente integrar algoritmos más complejos en etapas futuras.
- Esta revisión contribuye así al conocimiento técnico necesario, para sustentar el diseño de un software especializado que combine capacidades analíticas, predictivas y visuales para la gestión de sitios contaminados por hidrocarburos.

### REFERENCIAS BIBLIOGRÁFICAS

- Atteia, O., Dubois, J., & Webster, R. (1994). Geostatistical analysis of soil contamination in the Swiss Jura. *Environmental Pollution*, 86(3), 315–327. [https://doi.org/10.1016/0269-7491\(94\)90172-4](https://doi.org/10.1016/0269-7491(94)90172-4).
- Bengtsson, G., & Törneman, N. (2009). A spatial approach to environmental risk assessment of PAH contamination. *Risk Analysis*, 29(1), 48–61. <https://doi.org/10.1111/j.1539-6924.2008.01128.x>.
- Gao, A., Cheng, L., Wei, X. (2024). Study on the Spatial Distribution and Migration Characteristics of Petroleum Hydrocarbons in Soil Based on Artificial Neural Networks. In: Park, J.S., Yang, L.T., Pan, Y., Park, J.J. (eds) *Advances in Computer Science and Ubiquitous Computing*. CUTECSA 2023. Lecture Notes in Electrical Engineering, vol 1190. Springer.

- Gao, A., Chen, L., Wei, X., Liu, C., & Cheng, L. (2024). Spatial Distribution Prediction and Migration Characteristic of Petroleum Hydrocarbons in Soil Based on Artificial Neural Networks. *Journal of Information Processing Systems*, 20(6). <https://doi.org/10.3745/JIPS.04.0330>.
- Gu, Y., Zuo, Z., Shi, C., & Hu, X. (2020). Feasibility study for spatial distribution of diesel oil in contaminated soils by laser induced fluorescence. *Applied Sciences*, 10(3), 1103. <https://doi.org/10.3390/app10031103>.
- Hyde, K., Ma, W., Obal, T., Bradshaw, K., Carlson, T., Mamet, S., & Siciliano, S. (2019). Incremental sampling methodology for petroleum hydrocarbon contaminated soils: volume estimates and remediation strategies. *Soil and Sediment Contamination: An International Journal*, 28(1), 51-64. <https://doi.org/10.1080/15320383.2018.1529736>.
- Karaca, G. (2016). Spatial distribution of polycyclic aromatic hydrocarbon (PAH) concentrations in soils from Bursa, Turkey. *Archives of Environmental Contamination and Toxicology*, 70, 406-417. <https://doi.org/10.1007/s00244-015-0248-2>.
- Liang, Y., Zhang, X., Wang, J., & Li, G. (2012). Spatial variations of hydrocarbon contamination and soil properties in oil exploring fields across China. *Journal of Hazardous Materials*, 241, 371-378. <https://doi.org/10.1016/j.jhazmat.2012.09.055>.
- Liu, G., Niu, J., Guo, W., Zhao, L., Zhang, C., Wang, M., Zhang, Z., & Guo, G. (2017). Assessment of terrain factors on the pattern and extent of soil contamination surrounding a chemical industry in Chongqing, Southwest China. *Catena*, 156, 237-243. <https://doi.org/10.1016/j.catena.2017.04.005>.
- Liu, G., Zhou, X., Li, Q., Shi, Y., Guo, G., Zhao, L., Wang, J., Su, Y., & Zhang, C. (2020). Spatial distribution prediction of soil As in a large-scale arsenic slag contaminated site based on an integrated model and multi-source environmental data. *Environmental Pollution*, 267, 115631. <https://doi.org/10.1016/j.envpol.2020.115631>.
- Luca, C., Si, B., & Farrell, R. (2007). Assessing spatial distribution and joint uncertainty of TPH-fractions: Indicator kriging and sequential indicator simulation. *Canadian Journal of Soil Science*, 87(4), 551-563. DOI:10.4141/CJSS07003.
- Mahanty, B., Behera, S., Godio, A., & Chiampo, F. (2025). Hybrid Modeling with Artificial Neural Networks for Predicting In-Situ Bioremediation Dynamics of Diesel Fuel-Spiked Soil. *Water Air & Soil Pollution*, 236, 285 (2025). <https://doi.org/10.1007/s11270-025-07940-0>.
- Metahni, S., Coudert, L., Gloaguen, E., Guemiza, K., Mercier, G., & Blais, J. (2019). Comparison of different interpolation methods and sequential Gaussian simulation to estimate volumes of soil contaminated by As, Cr, Cu, PCP and dioxins/furans. *Environmental Pollution*, 252, 409-419. <https://doi.org/10.1016/j.envpol.2019.05.122>.
- Miao, S., Ni, G., Kong, G., Yuan, X., Liu, C., Shen, X., & Gao, W. (2025). A spatial interpolation method based on 3D-CNN for soil petroleum hydrocarbon pollution. *PLOS ONE*, 20(1), e0316940. <https://doi.org/10.1371/journal.pone.0316940>.
- Ouabo, R., Sangodoyin, A., & Ogundiran, M. (2020). Assessment of ordinary kriging and inverse distance weighting methods for modeling chromium and cadmium soil pollution in e-waste sites in Douala, Cameroon. *Journal of Health & Pollution*, 10(26), 1-20. <https://doi.org/10.1016/j.jhp.2020.10.005>.

org/10.5696/2156-9614-10.26.200605.

Parvizi, Y., & Fatehi, S. (2025). Geospatial digital mapping of soil organic carbon using machine learning and geostatistical methods in different land uses. *Scientific Reports*, *15*, 4449. <https://doi.org/10.1038/s41598-025-88062-9>.

Rakotonirina, H., Honeine, P., Atteia, O., & Van Exem, A. (2025). Estimating Contaminated Soil Volumes Using a Generative Neural Network: A Hydrocarbon Case in France. *Mathematical Geosciences*, <https://doi.org/10.1007/s11004-025-10193-6>.

Stewart, R., & Purucker, S. (2011). An environmental decision support system for spatial assessment and selective remediation. *Environmental Modelling & Software*, *26*(6), 751-760. <https://doi.org/10.1016/j.envsoft.2010.12.010>.

Yu, H., Li, T., Liu, Y., & Ma, L. (2019). Spatial distribution of polycyclic aromatic hydrocarbon contamination in urban soil of China. *Chemosphere*, *230*, 498-509. <https://doi.org/10.1016/j.chemosphere.2019.05.006>.

Zhang, G., Lan, T., Yang, G., Li, J., & Zhang, K. (2022). Contamination, spatial distribution, and source contribution of persistent organic pollutants in the soil of Guiyang city, China: a case study. *Environmental Geochemistry and Health*, *44*, 3265–3278. <https://doi.org/10.1007/s10653-021-01089-5>.