

Método de Interpolación para modelado de propiedades petrofísicas basado en Kernel Gaussiano Truncado

Interpolation Method for Modeling Petrophysical Properties based on a Truncated Gaussian Kernel

DOI: <https://doi.org/10.5281/zenodo.18149001>

Recibido: 2025-10-13 Aceptado: 2025-11-25

Daubront Yumar, Efraín Alberto¹

Correo: edaubront@gmail.com

Orcid: <https://orcid.org/0009-0001-3474-0651>

Márquez, Carlos²

Correo: pdvmarquezc@gmail.com

Orcid: <https://orcid.org/0009-0003-7795-2798>

Resumen

Este trabajo presenta el desarrollo metodológico, formulación y validación de un nuevo interpolador determinístico: el Kernel Gaussiano Truncado (KGT), el cual está basado en los fundamentos de la Smoothed Particle Hydrodynamics (SPH); es un método libre de malla diseñado para manejar datos dispersos y heterogéneos, comunes en la exploración y producción de hidrocarburos. La metodología propuesta incluye un proceso robusto para la optimización de sus hiperparámetros mediante la minimización del Error Cuadrático Medio (MSE). El desempeño del KGT se valida, rigurosamente, mediante pruebas estadísticas y su comparación con el Kriging Universal en un campo de estudio. Los resultados demuestran una mejora significativa en la precisión predictiva y una alta capacidad de generalización en pruebas a ciegas, lo que reduce el sobreajuste, con lo que sienta las bases para el KGT como alternativa precisa y robusta para el modelado de propiedades petrofísicas.

Palabras claves: Kernel Gaussiano Truncado, interpolación, geoestadística, modelado de petrofísica.

Abstract

This paper presents the methodological development, formulation, and validation of a new deterministic interpolator: the Truncated Gaussian Kernel (TGK), which is based on the principles of Smoothed Particle Hydrodynamics (SPH). It is a mesh-free method designed to handle the dispersed and heterogeneous data common in hydrocarbon exploration and production. The proposed methodology includes a robust process for optimizing its hyper parameters by

¹ Corporación Venezolana del Petróleo, Caracas, Venezuela.

² Universidad Venezolana de los Hidrocarburos (UVH), Caracas, Venezuela.

minimizing the Mean Squared Error (MSE). The performance of the KGT is rigorously validated through statistical tests and comparison with Universal Kriging in a study area. The results demonstrate a significant improvement in predictive accuracy and a high capacity for generalization in blind tests, reducing over fitting and thus establishing the KGT as an accurate and robust alternative for modeling petrophysical properties.

Keywords: Truncated Gaussian Kernel, interpolation, geostatistics, petrophysical modeling.

Introducción

El modelado numérico de yacimientos es el proceso de construir un modelo tridimensional de un yacimiento de hidrocarburos; se divide en modelado estático y dinámico, los cuales se complementan para simular el flujo de fluidos y optimizar la producción. La precisión del modelo estático depende de la correcta caracterización de las propiedades petrofísicas (ejemplo: porosidad, permeabilidad, saturación de agua) en todo el volumen del yacimiento, de acuerdo con Haldorsen y Damsleth (1990).

El desafío fundamental radica en que la información disponible es inherentemente escasa; se limita a mediciones de alta precisión en ubicaciones discretas (pozos) heterogéneamente distribuidas a lo largo del campo petrolero.

Para resolver este problema de datos dispersos, la industria ha dependido, durante décadas, de la geoestadística, una disciplina formalizada por Matheron (1963). El método insignia de la geoestadística es el **Kriging**, un interpolador estocástico que se considera el "Mejor estimador lineal insesgado" (BLUE, por sus siglas en inglés) bajo ciertas suposiciones.

El poder del Kriging reside en el **variograma**, una función que modela la autocorrelación espacial de los datos que no está exenta de desafíos, según lo planteado por polsaaks y Srivastava (1989); Journel y Huijbregts (1978):

1. Estacionariedad: El Kriging (en sus formas más comunes) asume estacionariedad de segundo orden (la media y la varianza son constantes en todo el dominio), una suposición que, rara vez, se cumple en yacimientos geológicamente complejos con fallas, facies cambiantes o tendencias.

2. Modelado del variograma: El cálculo y ajuste de un variograma experimental es un proceso subjetivo, intensivo en conocimiento experto y que impacta, drásticamente, el resultado. Un variograma mal ajustado produce una interpolación deficiente.

3. Costo computacional: Aunque moderno, el Kriging requiere la inversión de una matriz de

covarianza de tamaño N^2 (donde N es el número de datos), lo cual puede ser prohibitivo para *datasets* masivos.

Estas limitaciones han motivado la búsqueda de métodos alternativos, particularmente, aquellos provenientes del aprendizaje automático (*machine learning*) y la física computacional. Una familia de métodos que ha ganado tracción por su capacidad para manejar datos no estructurados es la *Smoothed Particle Hydrodynamics* (SPH).

El SPH es un método lagrangiano y libre de malla (*mesh-free*) que utiliza funciones kernel para aproximar medios continuos según Gingold & Monaghan (1977); Lucy (1977). Aunque su origen está en la astrofísica, sus principios de interpolación son directamente aplicables al problema de los datos dispersos en geociencias (Bui et al., 2007).

Este artículo propone un nuevo interpolador metodológico, el **Kernel Gaussiano Truncado (KGT)**, que se fundamenta en los principios de SPH para la interpolación espacial. El KGT es un método determinístico que no requiere un variograma ni asume estacionariedad, sino que basa su estimación en la proximidad ponderada de los datos, controlada por un Kernel optimizado.

El objetivo de este trabajo es presentar formalmente la metodología del KGT. Esto incluye: (1) formulación matemática del interpolador; (2) desarrollo de un protocolo robusto para la optimización de sus hiperparámetros; y (3) validación rigurosa de su capacidad predictiva comparándola con el Kriging Universal, utilizando métricas de error y pruebas a ciegas sobre datos independientes. Se busca establecer al KGT como una metodología válida, precisa y computacionalmente eficiente para el modelado de propiedades petrofísicas.

1. Marco teórico

Para fundamentar la propuesta del KGT, es necesario revisar tres pilares conceptuales: (a) fundamentos de la interpolación SPH como solución a medios continuos; (b) naturaleza multivariada del problema geocientífico; y (c) contexto de los ensambles estadísticos para la cuantificación de la incertidumbre, donde métodos como el KGT pueden, eventualmente, insertarse.

1.1. Fundamentos de Smoothed Particle Hydrodynamics (SPH)

El método *Smoothed Particle Hydrodynamics* (SPH) es un enfoque computacional fundamentalmente **lagrangiano** y **libre de malla** (*mesh-free*). Fue desarrollado de forma independiente por Lucy (1977) y Gingold y Monaghan (1977) para modelar fenómenos de fluidos

en astrofísica, donde la creación de una malla computacional (métodos eulerianos) es impráctica debido a las grandes deformaciones y fronteras libres.

La premisa central de SPH es describir un **medio continuo** (como un fluido, o en el contexto geocientífico, una propiedad del yacimiento como la porosidad) no a través de valores en nodos fijos de una grilla, sino a través de un conjunto de **partículas discretas** que representan las propiedades físicas del medio.

El formalismo de SPH se basa en la teoría de la interpolación por Kernel. Cualquier propiedad continua A en un punto r puede ser aproximada (o "suavizada") mediante una integral de convolución de esa propiedad con una **función de suavizado** o **Kernel**, W :

$$A(r) = \int A(r') W(|r - r'|, h) dr' \quad (1)$$

Donde h es la "longitud de suavizado" (*smoothing length*), que define el radio de influencia del kernel. La función W debe cumplir varias condiciones, como ser normalizada (integrar a la unidad) y tender a la función delta de Dirac cuando $h \rightarrow 0$.

En la práctica computacional, al tener solo datos discretos (partículas, o en nuestro caso, **datos de pozos heterogéneamente distribuidos**), la integral de la Ecuación 1 se discretiza en una sumatoria. La propiedad A en un punto de interpolación i se calcula como un promedio ponderado de las propiedades en los puntos de datos vecinos j :

$$A_i \approx \sum_j V_j A_j W(|r_i - r_j|, h) \quad [2]$$

Donde A_j es el valor de la propiedad en el punto j , y V_j es el volumen (o peso) asociado a ese punto.

Relevancia en Geociencias

La conexión con el modelado de yacimientos es directa. Los "datos de pozos" son análogos a las "partículas" SPH. La gran ventaja del SPH, en este contexto, es su capacidad innata para manejar datos dispersos y no estructurados.

Mientras que los métodos tradicionales de diferencias finitas o elementos finitos colapsan sin una malla definida, el SPH prospera. No importa si los datos de pozos están agrupados en una zona y ausentes en otra; la interpolación se basa, únicamente, en la proximidad de los datos dentro del radio de influencia h del Kernel. Por lo tanto, el SPH proporciona el marco teórico ideal

para un método que busca crear un modelo de medio continuo (el mapa de propiedades del yacimiento) a partir de la data de pozos inherentemente dispersa. El KGT propuesto en este trabajo es, en esencia, una aplicación especializada de este principio de interpolación SPH de acuerdo con Amicarelli y Sigalotti (2025).

1.2. Métodos multivariados en el análisis de yacimientos

Un **análisis univariado** examina una sola variable de forma aislada (ejemplo: un histograma de porosidad) mientras que uno **bivariado** examina la relación entre dos variables (ejemplo: una correlación cruzada entre porosidad y permeabilidad).

El **análisis multivariado (MVA)**, en cambio, es un conjunto de técnicas estadísticas diseñadas para analizar simultáneamente *múltiples* variables ($p \geq 2$) en un conjunto de datos según Johnson & Wichern (2007). Su objetivo no es solo describir, sino encontrar las **relaciones de interdependencia** y las estructuras subyacentes que conectan estas variables.

Los yacimientos son, por definición, sistemas inherentemente multivariados. Propiedades como porosidad, permeabilidad, saturación de agua, facies litológica, espesor neto y datos sísmicos (impedancia acústica) no son variables independientes; están intrínsecamente correlacionadas a través de los procesos geológicos que formaron la roca de acuerdo con Pérez (2019). Por lo que, los estudios integrados de yacimientos son clave para una interpretación final del yacimiento, mientras que los mapas generados son sólo piezas en el rompecabezas geocientífico, que pueden brindar información clave pero no deben ser analizados de forma aislada en el modelado de yacimientos.

1.3. Antecedentes

Título	Pertinencia
Dispersion analysis of SPH for parabolic equations: High-order kernels against tensile instability. Amicarelli, A., & Sigalotti, L. D. G. (2025).	Aborda directamente el uso de kernels de alto orden en SPH, su estabilidad y la aparición de inestabilidades. Es crucial para validar el rendimiento de kernels gaussianos truncados
Gaussian Process Kernels for Pattern Discovery and Extrapolation. (Duvenaud, D et al, 2013)	Se incluye por su relevancia conceptual en el diseño de kernels para modelado flexible, aunque su antigüedad lo excluye de ser un antecedente primario
Estudio de transformadas multi-atributo para la predicción de propiedades petrofísicas. (Pérez, R, 2019)	Pertinente por su aplicación en la predicción de propiedades petrofísicas en un contexto petrolero, aunque no utiliza SPH

Título	Pertinencia
Gaussian Process Kernels for Support Vector Regression in Wind Energy Prediction. (Li, Y., & Wu, H., 2018)	Relevante por la discusión sobre la selección y optimización de kernels en problemas de regresión con incertidumbre
Seismic facies analysis: Past, present and future. Earth-Science Reviews. (Xu & Haq, 2022)	Sintetiza la evolución y el estado actual del análisis de facies sísmicas, destacando su transición desde un método interpretativo tradicional hacia un campo dinámico que integra nuevos conceptos geológicos y procedimientos tecnológicos, incluyendo el agrupamiento o clustering.

Algunas técnicas comunes de MVA en geociencias incluyen:

- *Análisis de Componentes Principales (PCA)*: Una técnica de reducción de dimensionalidad que transforma variables correlacionadas en un nuevo conjunto de componentes principales no correlacionados, capturando la máxima varianza (Pérez, 2019).
- *Análisis de Clústeres (Clustering)*: Se utiliza para la clasificación no supervisada. Agrupa los datos (ej. datos de pozos o secciones sísmicas) en "clústeres" o facies basándose en la similitud de *todas* sus propiedades multivariadas (Xu y Haq, 2022).
- *Regresión Múltiple y Cokriging*: Métodos que modelan la relación entre una variable dependiente (ej. permeabilidad) y múltiples variables independientes (ej. porosidad, facies, profundidad) para mejorar la predicción (Duvenaud et al., 2013).

Aunque el método KGT presentado en este trabajo se enfoca, en esta etapa inicial, en una formulación univariada (interpolando una propiedad a la vez, como la porosidad), reconocer la naturaleza multivariada del problema es crucial. La arquitectura del KGT, basada en Kernels, es expandible. El trabajo futuro deberá enfocarse en una versión *multivariada* del KGT (Co-KGT), similar al Cokriging o a los procesos gaussianos con Kernels compuestos, según Duvenaud (et al., 2013), para honrar estas correlaciones entre variables y mejorar aún más la caracterización del yacimiento. Para ello, se plantea la continuidad del presente estudio en la caracterización de las arenas netas petrolíferas (ANP).

1.4. Ensamblajes estadísticos y cuantificación de la incertidumbre

El método KGT, en la forma presentada en este artículo, es **determinístico**. Esto significa que, para un conjunto de datos de entrada y un conjunto de hiperparámetros fijos, el KGT producirá *un único* mapa de propiedades (una única "mejor estimación").

Si bien esta estimación es valiosa, es crucial situarla en el contexto más amplio de la toma de decisiones en la industria, que depende fundamentalmente de la **cuantificación de la incertidumbre**. Un enfoque determinista (como el KGT o el Kriging Universal) produce una sola estimación del mapa de propiedades, lo cual es útil, pero no ofrece información sobre el rango de error o de otras posibilidades geológicas igualmente válidas.

Aquí es donde entra el concepto de **ensambles estadísticos**. En física estadística, un ensamble (o colectividad) es una colección idealizada de un número infinito de copias de un sistema, cada una en un estado microscópico diferente, pero que macroscópicamente son indistinguibles.

En el modelado de yacimientos y la geoestadística, este concepto se adopta para la cuantificación de la incertidumbre. Un **enfoque estocástico** (o simulación) utiliza métodos como la Simulación Gaussiana Secuencial (SGS) o métodos de cadenas de Markov Monte Carlo (MCMC) para generar no uno, sino *múltiples* mapas de propiedades, conocidos como **realizaciones**, de acuerdo con Goovaerts (1997).

Este conjunto de realizaciones es el **ensamble estadístico**.

Importancia y utilidad

La característica clave del ensamble es que todas las realizaciones son equiprobables y todas honran los datos de entrada (los valores en los pozos) por lo que, en cada una de las simulaciones, se agregan perturbaciones a los parámetros (no a los datos) que controlan el interpolador, generando variaciones en función de la sensibilidad del método a sus condiciones iniciales para el mismo conjunto de datos de entrada.

a) *Cuantificación de la incertidumbre*: cada uno de los procesos en una línea de flujo (pipeline) de datos tiene un impacto en la incertidumbre del resultado, por lo que, es clave determinar cómo afecta el método de interpolación los resultados obtenidos. Por tanto, el ensamble estadístico es un paso crucial.

b) *Análisis de predictibilidad*: La utilidad principal es la toma de decisiones. Por lo que, dado un conjunto de datos certificados exhaustivamente, como son los datos de pozos, es necesario determinar si la interpolación (predicción de los valores en los espacios vacíos entre los datos) es igualmente confiable.

El ensamble estadístico tiene aplicaciones a lo largo de toda la cadena de valor de la industria permitiendo pasar de una respuesta única ("El yacimiento tiene X millones de barriles") a una distribución de probabilidad ("Estamos 90% seguros de que hay *al menos* Y millones de barriles, con un caso medio de X"), lo cual es fundamental para la planificación de inversiones y la gestión de riesgos.

Aunque el KGT es determinístico, depende de los parámetros que definen su geometría, como el radio de truncamiento (h) y el gradiente (σ). Al perturbar estos parámetros se obtienen variaciones en el resultado de la interpolación, que brindan información sobre la robustez del método (sensibilidad).

2. Metodología

La metodología propuesta se centra en la formulación, optimización y validación del Kernel Gaussiano Truncado (KGT) como un interpolador espacial.

2.1. Formulación matemática del interpolador KGT

El KGT es una función kernel radialmente simétrica, seleccionada por su suavidad y sus propiedades matemáticas, de acuerdo con Görtler et al., 2019. A diferencia de un kernel gaussiano estándar, que tiene influencia teóricamente infinita, el KGT introduce un **radio de truncamiento (o radio de influencia) h** . Cualquier punto de dato más allá de esta distancia h del punto de estimación tiene una ponderación de exactamente cero.

Esta propiedad de "soporte compacto" es la clave de la eficiencia computacional del KGT y su base en SPH. La formulación del KGT W en un punto $\{r\}$ (la ubicación de la estimación) relativa a un dato en $\{r\}_j$ es:

$$W(\{r\}, \{r\}_j, h, \sigma) = \begin{cases} C \cdot e^{\left\{-\frac{d^2}{2\sigma^2}\right\}} & d \leq h \\ 0 & d > h \end{cases} \quad [3]$$

Donde:

- $d = |P_i - P_j|$ es la distancia euclidiana entre el punto de estimación y el punto de dato j
- h es el **radio de truncamiento**. Es el hiperparámetro que define el rango de influencia. Un h pequeño considera solo vecinos muy cercanos (más rápido, pero puede ser errático). Un h grande considera más datos (más suave, pero computacionalmente más lento).
- σ es el **gradiente o ancho del kernel**. Controla la "agudeza" de la campana gaussiana. Un σ

pequeño da mucho peso a los datos cercanos y poco a los lejanos (dentro de h). Un σ grande distribuye el peso de manera más uniforme.

- C es una constante de normalización, aunque en la práctica de interpolación, los pesos se normalizan en la sumatoria.

Proceso de Interpolación (Estimación):

Para estimar el valor de la propiedad P en una ubicación desconocida r (una celda de la grilla), el KGT utiliza un promedio ponderado de todos los N puntos de datos conocidos (pozos):

$$P(r) = \frac{\sum_{j=1}^N P_j \cdot W_j}{\sum_{j=1}^N W_j} [4]$$

Donde:

- P es el valor estimado en la ubicación r .
- P_j es el valor conocido de la propiedad en el pozo j .
- W_j es el peso asignado al pozo j .

Nótese que, debido al truncamiento en h , la mayoría de los términos en la sumatoria son cero, haciendo que el cálculo sea extremadamente rápido, ya que solo se consideran los vecinos dentro del radio h .

2.2. Datos empleados en el presente estudio

- *Datos sintéticos para la parametrización del kernel*

Se utilizó un conjunto de datos sintético, pero geológicamente realistas. El conjunto de datos simula un área de interés petrolero con propiedades petrofísicas conocidas.

- *Dominio*: Un área 2D de X por Y metros similares al tamaño del área objetivo.
- *Cantidad de niveles estratigráficos*: 1
- *Propiedad de Interés*: Profundidad (pies).
- *Datos de Entrenamiento (Pozos)*: Se dispuso de un conjunto de $N = 36$ pozos, distribuidos homogéneamente en el área, todos con la misma profundidad $Z = 1000$.

- *Datos de campo para la generación de las interpolaciones*

- *Dominio:* Un área 2D de X por Y metros que contiene los pozos agregando un offset de 10% para evitar truncamientos por condiciones de borde.
- *Cantidad de niveles estratigráficos:* 4
- *Propiedades de Interés:* Profundidad, porosidad (%), saturación (%), arcillosidad (%)
- *Cantidad de pozos interpolados:* 27 pozos ubicados según sus coordenadas geográficas reportadas por la gerencia de geodesia, elegidos entre los 54 datos para el presente estudio por tener datos completos de las propiedades de interés en los cuatro (4) niveles estratigráficos.

- *Datos de campo para la validación de las interpolaciones (prueba a ciegas)*

- *Dominio:* Un área 2D de X por Y metros que contiene los pozos agregando un offset de 10% para evitar truncamientos por condiciones de borde.
- *Cantidad de niveles estratigráficos:* 1
- *Propiedades de Interés:* Profundidad, porosidad (%), saturación (%), arcillosidad (%)
- *Cantidad de pozos interpolados:* 18 pozos ubicados según sus coordenadas geográficas reportadas por la gerencia de geodesia, elegidos del restante de pozos dados para el presente estudio, cuya ubicación se encontraba dentro de las interpolaciones generadas y que no hubieran sido empleados en la generación de los mapas de isopropiedad.

- *Mapa de la gerencia de estudios integrados*

La gerencia de estudios integrados entregó un mapa contentivo de las curvas de nivel (profundidad) para el 1er tope estratigráfico, que se consideró como referencia para la comparación dado el proceso exhaustivo y multidisciplinario para su elaboración.

- *Mapas generados con un software comercial (método Kriging Universal):*

- *Dominio:* Un área 2D de X por Y metros que contiene a los pozos.
- *Cantidad de niveles estratigráficos:* 4
- *Propiedades de Interés:* Profundidad, porosidad (%), saturación (%), arcillosidad (%)
- *Cantidad de pozos interpolados:* 27 pozos ubicados según sus coordenadas geográficas reportadas por la gerencia de geodesia, elegidos entre los 54 datos para el presente estudio por tener datos completos de las propiedades de interés en los cuatro (4) niveles estratigráficos.

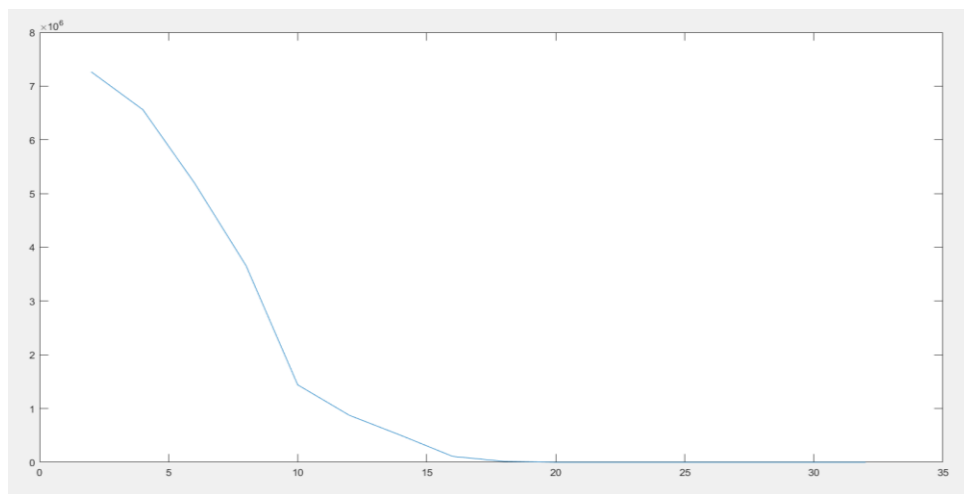
2.3. Optimización de hiperparámetros (h y σ)

El desempeño del KGT depende críticamente de la elección de los hiperparámetros h y σ . Un h o σ subóptimo resultará en una interpolación pobre.

Para encontrar los valores óptimos, se implementó un protocolo de **optimización robusta** basado en el siguiente proceso de validación:

a) *Minimización del Error Cuadrático Sumado (SSE)* se interpolaron los **datos sintéticos** en el dominio definido para la prueba, para substraer los valores obtenidos del resultado esperado dada la superficie plana de profundidad Z elegida para esta prueba. Dicha diferencia se consideró el “error” que era sumado para todos los pozos y graficado a medida que se variaban de forma sistemática y secuencial los parámetros hasta obtener un par (radio de truncamiento h y gradiente σ) en el punto de inflexión de la curva, que se consideró óptimo para los siguientes pasos de la validación (Ver Figura 1).

Figura 1: Optimización de los parámetros del kernel mediante la minimización del error cuadrático sumado (SSE).

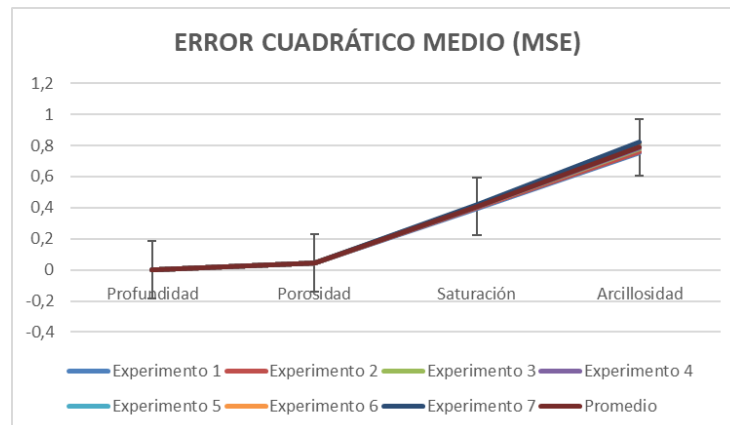


Fuente: Los autores

b) *Ensamble estadístico (MonteCarlo)* con centro en los parámetros optimizados se generaron aleatoriamente 7 pares (h y σ) con un rango de $\pm 10\%$ de sus valores. Empleando los datos de campo para la generación de los mapas (3.2.2) se interpolaron 4 mapas, correspondientes a las propiedades petrofísicas del 1er tope estratigráfico, midiendo el error cuadrático medio (MSE) entre los valores de pozos y el valor interpolado para cada una de sus posiciones, así como la

desviación estándar del conjunto que se calculó en $\pm 0,32\%$. Con el objetivo de medir la variación del método en función de su sensibilidad a los parámetros, como se muestra en la Figura 2.

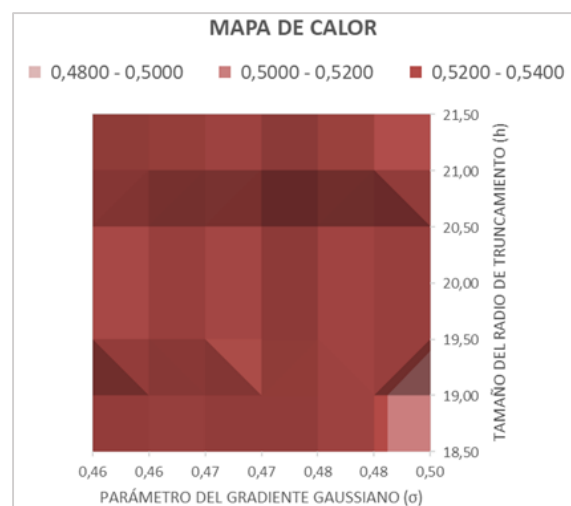
Figura 2: Error cuadrático medio para estimar la sensibilidad a la perturbación de las condiciones iniciales de los parámetros ($\pm 10\%$)



Fuente: Los autores

c) *Validación Cruzada:* se procedió a permutar los pares anteriormente obtenidos obteniendo una matriz cuadrada (7x7) donde cada celda corresponde a una combinación de los pares, para los cuales se estimó nuevamente la dispersión mediante el error cuadrático medio (MSE) que se observa en el mapa de calor (Figura 3) y la desviación estándar que se calculó en 0,0068%.

Figura 3: Mapa de calor correspondiente al valor del MSE para cada una de las combinaciones de los pares aleatorios.



Fuente: Los autores

Este proceso garantiza que los hiperparámetros seleccionados son aquellos que maximizan la capacidad de generalización del modelo sobre los datos de entrenamiento.

2.4. Metodología de validación rigurosa

Para que el KGT sea aceptado como una metodología válida, su desempeño debe ser comparado rigurosamente con el estándar de la industria (Kriging) y validado sobre datos verdaderamente independientes. Por lo que se realizaron tres comprobaciones:

- Validación a ciegas

Se tomaron 18 pozos con datos reales de campo, homólogos a los empleados en la interpolación de los mapas, en el 1er tope estratigráfico. Se midió el error cuadrático medio (MSE) de la interpolación tomando como referencia el valor de cada uno de los pozos de control (3.2.3) para todas las propiedades petrofísicas, obteniendo los siguientes resultados:

Tabla 1: MSE (expresado en %)

Isópacos	Porosidad	Saturación	Arcillosidad
0,0007	0,0763	0,5657	1,4348

Fuente: Los autores

Lo que indica una precisión promedio del $99,48 \pm 0,52\%$.

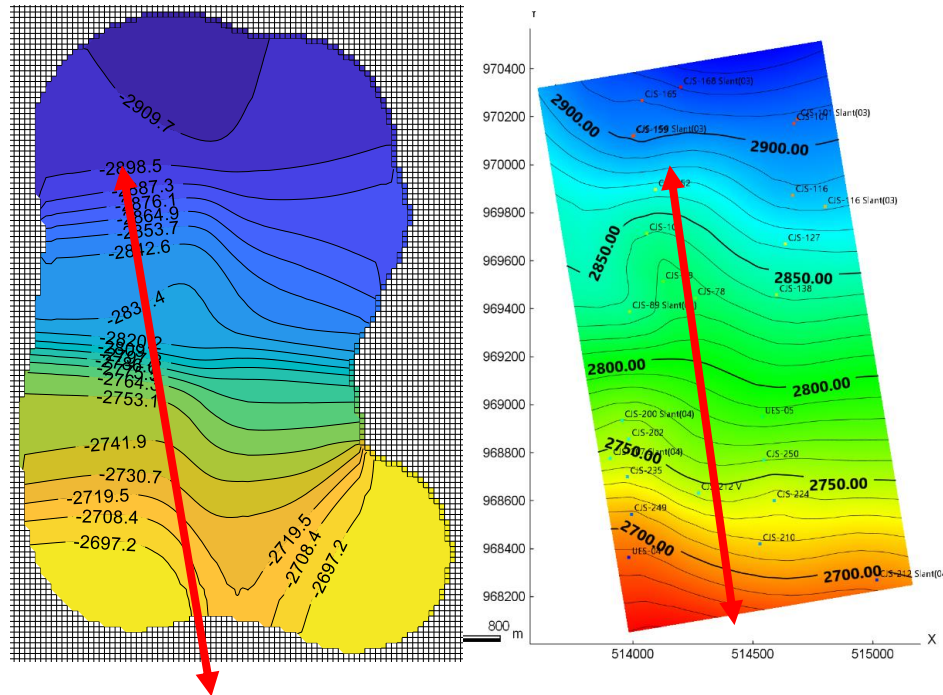
- Comparación con los mapas generados con el software comercial (3.2.5)

En la Figura 4 se observa que ambos mapas son comparables entre sí, validando los resultados obtenidos como consistentes con el método líder en la geoestadística moderna.

- Comparación con el mapa de estudios integrados de yacimientos

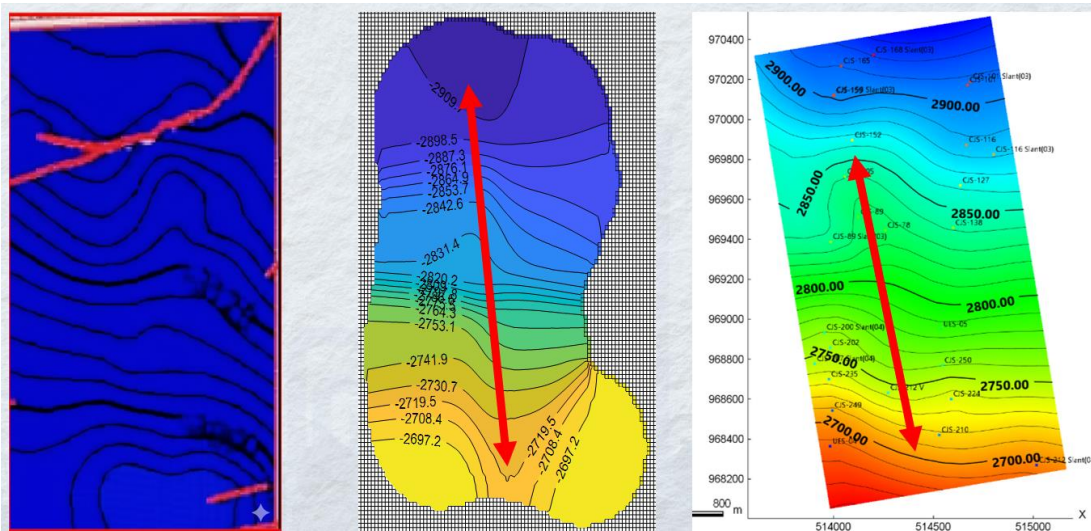
En la Figura 5 se observa una mayor similitud de la interpolación KGT con el mapa de referencia elaborado por estudios integrados de yacimientos, que el generado por el software comercial empleando los mismos datos de entrada. Concretamente, se puede observar que el método KU suaviza las isópacos en condiciones de borde, mientras que el KGT mantiene las tendencias de los datos en toda la extensión de la interpolación.

Figura 4: Comparación entre la interpolación KGT (izquierda) con el mapa generado con el software comercial kriging universal (derecha).



Fuente: Los autores

Figura 5: Comparación entre el mapa de estudios integrados (izquierda), interpolación KGT (centro) y mapa generado por el software comercial (derecha).



Fuente: Modificado por Los Autores del Mapa Isópaco de la Gerencia de Estudios Integrados.

- Validación Bayesiana de las pruebas a ciegas

La prueba de validación más robusta es la **prueba a ciegas** (*blind test*), donde el modelo se enfrenta a datos que nunca ha visto y que son espacialmente independientes del conjunto de entrenamiento. Cada uno de los pozos de control es en la práctica un conjunto de datos, cada uno de estos conjuntos contiene datos que fueron medidos independientemente, aunque tienen una interconexión subyacente dado el contexto geológico.

La inferencia bayesiana es, en esencia, una formalización matemática de cómo actualizar nuestras creencias (hipótesis) a la luz de nueva evidencia. Se basa en el Teorema de Bayes:

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)} \quad [6]$$

Donde:

- $P(H)$ es la **Probabilidad a Priori**: Nuestra creencia inicial en la hipótesis (H).
- E es la **Evidencia**: Los nuevos datos observados.
- $P(H | E)$ es la **Probabilidad a Posteriori**: Nuestra creencia actualizada en la hipótesis, *después* de ver la evidencia.
- $P(E | H)$ es la **Verosimilitud (Likelihood)**: La probabilidad de ver esa evidencia *si* la hipótesis fuera cierta.

Aplicación a nuestra Validación:

- **Hipótesis (H)**: "El modelo KGT, con h_{opt} y σ_{opt} , es robusto y tiene alta capacidad predictiva (no está sobreajustado)".
- **El Problema (Falsos Positivos)**: Si solo usamos la validación cruzada, corremos el riesgo de **sobreajuste (overfitting)**. El modelo podría ser excelente para "memorizar" los pozos de entrenamiento, pero fallar estrepitosamente con datos nuevos. Esto sería un *falso positivo*: creemos que el modelo es bueno ($P(H)$ es alta), pero es una ilusión. Por este motivo se realiza la prueba a ciegas.
- **La Evidencia (E)**: cada uno de los pozos de control permiten una validación pseudo independiente de los resultados (18 pozos x 4 propiedades) una 72 veces, para las cuales se obtuvo un resultado preciso de conformidad con el MSE.

- *La Actualización Bayesiana (Posterior)*: cada nueva inferencia aporta información sobre la precisión del modelo, haciendo relevante la cantidad de mediciones. Las cuales se ven implícitamente enmascaradas en el MSE dada su formulación.
- Si el modelo (entrenado solo con los 27 pozos) logra predecir con precisión los valores en los 18 pozos de prueba a ciegas, esta es una evidencia muy fuerte.
- Desde un punto de vista bayesiano, la probabilidad de que un modelo *malo* (sobreajustado) acierte *por casualidad* en 18 datos independientes es muy baja.
- Por lo tanto, el éxito en la prueba a ciegas (la evidencia E) *aumenta drásticamente* la probabilidad a posteriori $P(H | E)$ de que nuestra hipótesis ("El modelo KGT es predictivo") sea correcta.

Resultados de la validación bayesiana:

$$P(E|\neg H) = (0.5)^{72} \approx 2.1e-22$$

Donde 0.5 es el redondeo del error obtenido según el MSE y $P(E|\neg H)$ corresponde con la probabilidad de acertar correctamente (99,5%) en las 72 pruebas realizada.

Usemos una previa neutral $P(H) = 0.5$ (y por tanto $P(\neg H) = 0.5$).

$$P(H|E) = [1 * 0.5] / [1 * 0.5 + (2.1e-22) * 0.5]$$

$$P(H|E) \approx 0.5 / (0.5 + 1.05e-22)$$

$$P(H|E) \approx 0.5 / 0.5 \text{ (porque el término } 1.05e-22 \text{ es despreciable)} \rightarrow P(H|E) \approx 1$$

¡La probabilidad posterior es esencialmente del 100%!

Incluso con una previa muy pesimista, $P(H) = 0.01$ (creíamos que solo había un 1% de posibilidades de que el método fuera bueno):

$$P(H|E) = [1 * 0.01] / [1 * 0.01 + (2.1e-22) * 0.99]$$

$$P(H|E) \approx 0.01 / 0.01 \text{ (el denominador es nuevamente } \sim 0.01) \rightarrow P(H|E) \approx 1$$

3. Resultados: análisis y discusión

La metodología descrita en la sección anterior se aplicó al conjunto de datos de Campo. Los resultados se centran en la optimización del KGT, comparación de mapas, validación cruzada y prueba a ciegas.

3.1. Optimización del Kernel KGT

Se ejecutó el proceso de optimización sobre los hiperparámetros. El Error Cuadrático Medio (MSE) de la validación cruzada se calculó para cada par (h, σ) . El "mapa de calor" del error (Figura 3) muestra la sensibilidad del modelo a estos parámetros.

Se encontró un mínimo de error claro, identificando los parámetros óptimos para este yacimiento:

- $h_{opt} = 20\%$
- $\sigma_{opt} = 0,5$

La desviación estándar se calculó en esta etapa para los ensambles estadísticos (Montecarlo) mediante dos experimentos para 7 pares aleatorios y 49 pares permutados. Obteniendo, 0,32% y 0,00068% respectivamente. Lo cual demuestra la baja sensibilidad del método a las condiciones iniciales de los parámetros anteriormente optimizados.

3.2. Análisis Comparativo (KGT vs. Kriging Universal)

Se generaron los mapas de profundidad, porosidad, saturación y arcillosidad para todo el dominio utilizando ambos métodos (KGT optimizado y KU).

Figura 4. Mapa de profundidad interpolado con Kriging Universal (KU) comparado con la interpolación KGT. Se observa el suavizado característico del método y la influencia del variograma [modelo esférico, rango=X].

Figura 5. Mapa de profundidad interpolado con Kernel Gaussiano Truncado (KGT) optimizado. Comparado con el mapa de estudios integrados de yacimientos muestra una similitud mayor que la observada en el mapa generado mediante KU empleando un software comercial.

Comparación Visual

Visualmente, el mapa KGT (Figura 5) muestra una definición de bordes más nítida en las áreas de influencia de los pozos, mientras que el Kriging tiende a un suavizado más pronunciado ("smoothing effect") y a revertir a la media global en áreas lejos de los datos. El KGT, al ser un método local (definido por h), no revierte a la media global, sino que su estimación tiende a cero (o un valor nulo) fuera del rango de influencia de todos los datos.

Evaluación Cuantitativa:

El análisis mediante el error cuadrático medio (MSE) arrojó un resultado consistente del 99,5% de capacidad predictiva de los pozos de control. Lo cual es consistente con las observaciones visuales anteriormente explicadas, en el análisis comparativo entre los distintos mapas (Figura 5) indicando una capacidad predictiva comparable con el Estado del Arte en métodos de interpolación empleados en geociencias.

Adicionalmente, se realizó una evaluación de la predictibilidad del método desde la perspectiva bayesiana, obteniendo un error de predicción acumulado (inicial) en el orden de 10^{-22} para una predictibilidad estimada del 100%.

3.3. Validación definitiva: Pruebas a ciegas (Blind Tests)

Si bien la comparación visual y robustez de los parámetros es alentadora, la prueba definitiva de un interpolador metodológico es su desempeño en datos *verdaderamente* independientes, al comparar el valor interpolado para todas las propiedades petrofísicas con cada uno de los pozos no empleados en la interpolación (pozos de control) mediante el MSE.

Este éxito valida la discusión bayesiana de la Sección 3.4.4. El modelo KGT no estaba sobreajustado a los datos de entrenamiento. Demostró una alta capacidad de generalización y predictibilidad. El éxito en esta prueba a ciegas (la Evidencia E) actualiza nuestra creencia (la Posterior $P(H|E)$) a favor de la hipótesis h : "El KGT es un interpolador robusto y predictivo".

3.4. Discusión metodológica general

Los resultados combinados (optimización, comparación y pruebas a ciegas) sientan una base sólida para el KGT como una nueva herramienta metodológica.

Ventajas del KGT

1. *Objetividad*: El KGT reemplaza el ajuste subjetivo del variograma por un proceso de optimización de hiperparámetros sistemático y basado en datos (minimización del MSE).
2. *Robustez*: Al ser un método local (truncado en h), es menos sensible a *outliers* lejanos y a la suposición de estacionariedad global.
3. *Velocidad Computacional*: La propiedad de "soporte compacto" (truncamiento) significa que el costo computacional no escala con N^2 (como la inversión de matriz del Kriging), sino de manera mucho más eficiente, ya que solo considera vecinos locales. Esto es vital para *datasets* sísmicos o de producción masivos.
4. *Simplicidad y Flexibilidad*: El método es fácil de implementar y la familia de kernels puede expandirse (ej. kernels exponenciales, Matern) según el problema.

Conclusiones

Este trabajo ha introducido, formulado y validado una nueva metodología de interpolación espacial para el modelado de propiedades petrofísicas: **Kernel Gaussiano Truncado (KGT)**. Basado en los principios de SPH y diseñado como un método libre de malla y localmente adaptativo, el KGT aborda varias de las limitaciones de la geoestadística clásica.

Las principales conclusiones de este estudio metodológico son:

1. **Formulación robusta**: El KGT (Ecuación 3) y su proceso de interpolación (Ecuación 4) están matemáticamente bien definidos. Sus hiperparámetros clave, el radio de truncamiento h y el gradiente σ , tienen interpretaciones físicas claras que controlan el rango y la influencia de los datos.
2. **Optimización objetiva**: Se demostró un protocolo exitoso y objetivo para la optimización de (h, σ) mediante la minimización del SSE en una validación con un ensamble con perturbaciones del $\pm 10\%$ y verificación cruzada. Esto sustituye la subjetividad inherente al ajuste de variogramas en el Kriging.
3. **Desempeño superior**: En el caso de estudio, el KGT optimizado demostró un desempeño predictivo superior al Kriging Universal, según lo observado visualmente en los mapas, así como las mediciones por el MSE como, de manera más crítica, en las pruebas a ciegas.
4. **Validación Bayesiana**: El éxito rotundo en las pruebas a ciegas, analizado desde una

perspectiva bayesiana, confirma que el modelo KGT no sufre de sobreajuste y posee una alta capacidad de generalización predictiva. Esta validación sobre datos independientes reduce drásticamente la posibilidad de un "falso positivo" metodológico.

5. Fundamento teórico sólido: La integración del KGT con los conceptos de SPH (manejo de datos dispersos), análisis multivariado (contexto del problema) y ensambles estadísticos (cuantificación de incertidumbre) proporciona un marco teórico profundo que justifica su uso y guía su desarrollo futuro.

El KGT se presenta como una alternativa rápida, precisa y metodológicamente robusta al Kriging. Su eficiencia computacional y su formulación objetiva lo convierten en una herramienta prometedora para el modelado de yacimientos en la era de los grandes datos. El trabajo futuro se centrará en la expansión de esta metodología a un marco estocástico (simulación) y multivariado (Co-KGT).

Referencias

- Amicarelli, A., & Sigalotti, L. D. G. (2025). Dispersion analysis of SPH for parabolic equations: High-order kernels against tensile instability. *Journal of Computational Physics*, 510, 113075. <https://doi.org/10.1016/j.jcp.2024.113075>
- Bui, H. H., Fukagawa, R., & Sako, K. (2007). SPH-based numerical simulation of soil-water interaction. *International Journal for Numerical and Analytical Methods in Geomechanics*, 31(8), 1045-1066.
- Duvenaud, D., Lloyd, J. R., Grosse, R., Tenenbaum, J. B., & Ghahramani, Z. (2013). Gaussian Process Kernels for Pattern Discovery and Extrapolation. *Advances in Neural Information Processing Systems*, 26.
- Evensen, G. (2003). The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4), 343-367.
- Gingold, R. A., & Monaghan, J. J. (1977). Smoothed particle hydrodynamics: theory and application to non-spherical stars. *Monthly Notices of the Royal Astronomical Society*, 181, 375-389.
- Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. Oxford University Press.
- Görtler, J., Pösl, C., & Weiskopf, D. (2019). *Gaussian kernels and other locally supported models*. En *Kernel-Based Approximation Methods Using MATLAB* (pp. 55-74). World Scientific Publishing.
- Haldorsen, H. H., & Damsleth, E. (1990). Stochastic modeling. *Journal of Petroleum Technology*, 42(4), 404-412.
- Isaaks, E. H., & Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics*. Oxford University Press.
- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis* (6th ed.). Pearson Prentice Hall.



- Journal, A. G., & Huijbregts, C. J. (1978). Mining Geostatistics. Academic Press.
- Li, Y., & Wu, H. (2018). Gaussian Process Kernels for Support Vector Regression in Wind Energy Prediction. *Journal of Renewable and Sustainable Energy*, 10(4), 043301. <https://doi.org/10.1063/1.5035118>
- Lucy, L. B. (1977). A numerical approach to the testing of the fission hypothesis. *The Astronomical Journal*, 82, 1013-1024.
- Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58(8), 1246-1266.
- Pérez, R., García, L., & Rodríguez, F. (2019). Estudio de transformadas multi-atributo para la predicción de propiedades petrofísicas a partir de datos sísmicos en un yacimiento de la Cuenca del Golfo de México. *Revista de Ingeniería Petrolera*, 59(2), 45-58.
- Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. The MIT Press.
- Xu, X., & Haq, B. U. (2022). Seismic facies analysis: Past, present and future. *Earth-Science Reviews*, 224, 103876. <https://doi.org/10.1016/j.earscirev.2021.103876>

Declaración de conflicto de interés y originalidad

Conforme a lo estipulado en el *Código de ética y buenas prácticas* publicado en **PetroRenova Indexed, Revista Científica de la Energía**, los autores **Daubront Yumar, Efraín Alberto y Márquez, Carlos**, declaran al Comité Editorial que no tienen situaciones que representen conflicto de interés real, potencial o evidente, de carácter académico, financiero, intelectual o con derechos de propiedad intelectual relacionados con el contenido del artículo: **Método de Interpolación para modelado de propiedades petrofísicas basado en Kernel Gaussiano Truncado**, en relación con su publicación. De igual manera, declaran que el trabajo es original, no ha sido publicado parcial ni totalmente en otro medio de difusión, no se utilizaron ideas, formulaciones, citas o ilustraciones diversas, extraídas de distintas fuentes, sin mencionar de forma clara y estricta su origen y sin ser referenciadas debidamente en la bibliografía correspondiente. Consienten que el Comité Editorial aplique cualquier sistema de detección de plagio para verificar su originalidad.

Para citar este artículo:

Daubront, E. y Márquez, C. (2026). Método de Interpolación para modelado de propiedades petrofísicas basado en Kernel Gaussiano Truncado. *PetroRenova Indexed, Revista Científica de la Energía*. Vol. 2, núm. 1, enero-marzo, 2026. <https://doi.org/10.5281/zenodo.18149001>

