Metodología para la Estimación Areal de Propiedades Petrofísicas Mediante el Uso de Técnicas Geoestadísticas

José Joaquín Domínguez Muñoz

Trabajo de Grado para Optar el título de Especialista en Ingeniería de Yacimientos

Directora
Maika Gambús Ordaz
PhD en Ingeniería de Petróleos

Universidad Industrial de Santander
Facultad de Ingenierías Fisicoquímicos
Escuela de Ingeniería de Petróleos
Especialización en Ingeniería de Yacimientos
Bucaramanga
2020

Dedicatoria

Dedico de manera muy especial este logro y avance en mi vida profesional a mi padre, a mi esposa y a mis dos hijos, por siempre estar a mi lado apoyándome y motivándome, por sus palabras y abrazos, por toda la paciencia que me tuvieron durante el desarrollo de este proyecto, por todos sus actos y palabras de afecto que en los momentos difíciles me dieron impulso a seguir y culminar con éxito.

Agradecimientos

Quiero agradecer a Dios quien es el soporte y el eje de mi vida, a mi directora la Dra Maika Gambús por todo su apoyo, recomendaciones y sugerencias, quien siempre estuvo dispuesta a ayudarme para avanzar en cada una de las fases de este proyecto, A Juan Soto quien fue como mi Codirector, gracias por todo el tiempo que me dedicó apoyándome, siempre buscando hacer eficiente y de una mejor manera todos los pasos y procesos en este proyecto.

Contenido

Introucción 14

1. Objetivos 17

1.1. Objetivo General 17

1.2. Objetivos Específicos 17

2. Antecedentes 18

3. Marco de referencia 20

3.1. Definición y clasificación de la Estadística 20

3.1.1. Estadística descriptiva. 20

3.1.2. Estadística inferencial. 20

3.2. Análisis Estadístico o Exploratorio de Datos 21

3.2.1. Histograma. 23

3.2.2. Boxplot 24

3.2.3. Gráfico Q-Q plot. 25

3.3. Prueba Gamma 26

3.3.1. Fundamentación teórica de la prueba Gamma. 26

3.4. Definición de Geoestadistica. 31

3.4.1. Historia de la Geoestadística.....32

3.4.2. Análisis estructural.....34

3.4.3. Interpolación y métodos de interpolación espacial.....34

3.4.4. Técnicas Geoestadísticas.....35

3.5. Porosidad.....36

3.5.1. Distribución de los minerales de arcilla.....36

3.6. Permeabilidad.....37

3.6.1. Método de Tixier.....38

3.6.2. Método de Timur.....38

3.6.3. Método de Coates.....38

4. Marco Metodológico.....40

4.1. Tipo de investigación.....40

4.2. Método.....40

4.3. Fuente de la información.....41

4.4. Análisis de los datos y tratamiento de la información.....41

4.4.1. Inventario, selección de los pozos e intervalo de interés.....43

4.4.2. Análisis Exploratorio de los Datos.....46

4.4.3. Aplicación de la prueba Gamma a los registros geofísicos RHOB, NPHI y DT.....47

4.4.4. Cálculo de porosidad, Volumen de arcilla, porosidad efectiva y porosidad promedio 47

4.4.5. Cálculo de permeabilidad por los métodos empíricos de Timur, Tixier y Coates.....49

4.4.6. Mapa de distribución de propiedades de porosidad efectiva y permeabilidad por los métodos empíricos de Timur, Tixier y Coates.....50

5. Resultados 51

5.1. Pozos seleccionados.....51

5.2. Análisis Exploratorio de los Datos de los registros RHOB, NPHI y DT 52

5.2.1. Resumen del AED para los tres registros RHOB NPHI y DT.....55

5.3. Aplicación de prueba Gamma a los registros geofísicos RHOB, NPHI y DT 59

5.4. Cálculo de porosidad, volumen de arcilla, porosidad efectiva y porosidad promedio 60

5.5. Resumen del AED para la porosidad efectiva de los 19 pozos.....62

5.6. Cálculo de permeabilidad por los métodos empíricos de Timur, Tixier y Coates.....64

5.7. Mapas de distribución de propiedades de porosidad efectiva y permeabilidad.....66

5.7.1. Semivariograma nivel medio. 67

5.7.2. Semivariograma nivel superior. 69

5.7.3. Semivariograma nivel inferior. 71

5.8. Mapas estimación de la distribución de permeabilidad 73

5.9. Ajuste del modelo de porosidad usando los estimadores los OLS, WLS, ML y RML 75

5.10. Mapa de distribución de porosidad efectiva usando el semivariograma ajustado 79

6. Análisis de resultados 81

7. Conclusiones 83

8. Recomendaciones 84

Referencias Bibliográficas 85

Lista de Tablas

Tabla 1 Coordenadas espaciales de los 52 pozos....43

Tabla 2 Combinaciones de los registros para la prueba Gamma....47

Tabla 3 Coordenadas de los pozos con registros RHOB, NPHI y DT....51

Tabla 4 Resumen del AED para el pozo GT05_07 del registro RHOB....53

Tabla 5 Resumen del AED para el registro RHOB....55

Tabla 6 Resumen del ED para el registro NPHI....57

Tabla 7 Resumen del AED para el registro DT....58

Tabla 8 Resumen de la prueba Gamma aplicada a los 19 pozos....59

Tabla 9 Resumen valores calculados de porosidad, volumen arcilla y porosidad efectiva para el pozo GT02_07....61

Tabla 10 Resumen de los valores de porosidad efectiva calculada para los 19 pozos....61

Tabla 11 Resumen del AED para la porosidad efectiva promedio de los 19 pozos....62

Tabla 12 Resumen permeabilidades calculados por métodos de Timur, Tixier y Coates....65

Tabla 13 Resumen de los valores de sill, rango y nugget para los semivariogramas generados para la porosidad efectiva....67

Tabla 14 Resumen del valor del error aplicando diferentes estimadores en cada nivel....76

Tabla 15 Resumen de los valores usando el estimador WLS-npairs....78

Lista de Figuras

Figura1 Fases del proyecto. 42

Figura 2 Mapa de distribución espacial de los 52 pozos. 45

Figura 3. Registros geofísicos GR, ILN, ILD, RHOB, NPHI y DT del pozo 02_07. 46

Figura 4. Mapa de distribución espacial de los 19 pozos seleccionados. 52

Figura 5. Histograma de frecuencia para el registro RHOB en el pozo GT05_07. 53

Figura 6. Boxplot para el registro RHOB para el pozo GT05_07 54

Figura 7. Q-Q plot para el registro RHOB en el pozo GT05_07 54

Figura 8. Grafica de regresión lineal YM (k) vs δ_M (k) para el pozo GT11_07. 59

Figura 9. Histograma para la porosidad efectiva de los 19 pozos. 63

Figura 10. Box plot para porosidad efectiva de los 19 pozos. 63

Figura 11. Q-Q plot para porosidad efectiva de los 19 pozos. 64

Figura 12. Representación del cálculo de la permeabilidad por diferentes métodos. 66

Figura 13. Semivariograma ajustado automáticamente (nivel medio). 68

Figura 14. Mapa distribución de la porosidad efectiva (nivel medio). 68

Figura 15. Mapa varianza de la porosidad efectiva (nivel medio). 69

Figura 16. Semivariograma ajustando a los datos experimentales (nivel superior). 70

Figura 17. Mapa distribución de la porosidad efectiva (nivel superior). 70

Figura 18. Mapa varianza de la porosidad efectiva (nivel superior). 71

Figura 19. Semivariograma ajustando a los datos experimentales (nivel inferior). 72

Figura 20. Mapa distribución de la porosidad efectiva (nivel inferior). 72

Figura 21. Mapa varianza de la porosidad efectiva (nivel inferior).....73

Figura 22. Mapa de estimación de permeabilidad por el método de Timur. 74

Figura 23. Mapa de estimación de permeabilidad por el método de Tixier. 74

Figura 24. Mapa de estimación de permeabilidad por el método de Coates. 75

Figura 25. Ajuste del semivariograma nivel medio usando los estimadores.....77

Figura 26. Ajuste del semivariograma nivel superior usando los estimadores.....77

Figura 27. Ajuste del semivariograma nivel inferior usando los estimadores.....78

Figura 28. Semivariograma nivel superior construido con el estimador WLS-spairs. 79

Figura 29. Mapa distribución de la porosidad efectiva (nivel superior) ajustado con el estimador.

.....80

Figura 30. Mapa varianza de la porosidad efectiva (nivel inferior) ajustado con el estimador. 80

Lista de Apéndices

Nota. Estos apéndices podrán ser consultados en la base de datos de la Biblioteca

A. Datos de los 52 pozos.
B. Código Análisis Exploratorio de Datos.
C. Código prueba Gamma.
D. Código Cálculo de porosidad, Volumen de arcilla, porosidad efectiva y porosidad promedio.
E. Código Cálculo de permeabilidad por los métodos empíricos de Timur, Tixier y Coates.
F. Código Generación del mapa con krige de la propiedad DPHIe_mean y permeabilidades.

.

Resumen

Título: Metodología para la Estimación Areal de Propiedades Petrofísicas Mediante el Uso de Técnicas Geoestadísticas[*]

Autores: José Joaquín Domínguez Muñoz[**]

Palabras Clave: Caracterización de yacimientos, Geoestadística, Propiedades petrofísicas, Prueba Gamma

Descripción: Se presenta una metodología para estimar la distribución areal de propiedades petrofísicas, la cual se basa en la prueba Gamma, un algoritmo que estima el Error Cuadrático Medio (ECM) de una variable dependiente, para este caso la prueba Gamma se aplica a registros geofísicos de porosidad: densidad (RHOB), neutrón (NPHI) y sónico (DT) de pozos perforados en la provincia de Alberta en Canadá. La prueba Gamma muestra el registro de densidad (RHOB) con el menor ECM en 9 de los 19 pozos seleccionados. A partir de este registro se calcula la porosidad, se incluyen las correcciones ambientales de diámetro de pozo, densidad de matriz y densidad de fluido de perforación, se realiza el cálculo de porosidad efectiva corrigiendo el efecto por volumen de arcillas (Vsh) y distribución de estas en el yacimiento. Luego mediante la implementación de técnicas geoestadísticas se construye el semiovariograma experimental para la porosidad efectiva, el cual se ajusta a un modelo esférico, con este modelo es poblada la malla en los sectores donde no se tiene información. Finalmente, al calcular la permeabilidad empleando los métodos empíricos de Timur, Tixier y Cotes se encuentra que, de los tres métodos, el de Timur presenta los mayores valores de permeabilidad, por el contrario, el de Coates los menores. Se emplea el software libre y de código abierto Rstudio para todos los cálculos, incluido el Análisis exploratorio de los Datos con el cual se logra tener una primera aproximación de las densidades litológicas y tipo de fluido presente en el yacimiento.

---

[*] Trabajo de Grado

[**] Facultad de Ingenierías Fisicoquímicos. Escuela de Ingeniería de Petróleos. Directora Maika Gambús Ordaz PhD en Ingeniería de Petróleos

Abstract

Títle: Methodology for Estimating Areal of Petrophysical Properties Using Geostatistics Techniques*

Authors: José Joaquín Domínguez Muñoz**

Keywords: Gamma test, Characterization reservoir, Geostatistics, Petrophysical properties
Description: This methodology presents a way to estimate the areal distribution of petrophysical properties, which is on based the Gamma Test, it is an algorithm that estimates the Mean Error square (MES) of the independent variable, in this case, the Gamma, is applicated to geophysics logs of porosity: density (RHOB), neutron (NPHI) and Sonic (DT) in wells drilled in Alberta province in Canada. The Gamma Test finds that the density log (RHOB) presents the minimum value of the Mean Error Square in 9 of 19 selected wells. With this log is calculated the porosity, this calculate includes environmental corrections of the diameter of the borehole, density of matrix, and density of fluid of perforation. Is calculated the effective porosity including corrections for the presence and distribution of clay in the reservoir. With applying of geostatic techniques is to build an experimental semivariogram of effective porosity, this model is fitted to a spherical model, using this model is populated the grid in sectors where do not exist data. Finally, the permeability is calculated using empirical methods of Timur, Tixier, and Coates, this result shows that the empirical method of Timur presents high values and on the other hand the Coates low values of permeability. To perform the calculations, it used Rstudio, free and open source software that includes Exploratory Analysis with which it is possible to have a first approximation of lithology densities and fluid class present in the reservoir.

---

* Bachelor Thesis

** Physiochemical Enginnering Faculty. School of Petroleum Engineering. Manager

Maika Gambús Ordaz PhD Petroleum Engineering.

1. Introducción

Antes de realizar la perforación de nuevos pozos de desarrollo o de avance, es necesario caracterizar el yacimiento, es decir, estar en la capacidad de entender y poder predecir algunas propiedades en áreas en donde no se ha adquirido información. Esta caracterización se realiza mediante la creación de modelos donde se cuantifican características que pueden ser estables a lo largo del tiempo y considerarse estáticas, como el modelamiento estructural, espesores de formación, litologías y distribución de porosidad y permeabilidad, sin embargo, estos modelos presentan un alto grado de incertidumbre. Para poder entender y predecir la distribución de propiedades del yacimiento y reducir el valor de la incertidumbre, se necesita disponer de información geológica adquirida directamente del yacimiento como núcleos de pozo o la adquisición de registros geofísicos se hace obligatorio.

La adquisición de núcleos de pozo económicamente presenta un alto costo, razón por la cual no se realiza en todos los pozos que se perforan, por otro lado, las adquisiciones de registros geofísicos presentan un menor costo, por tanto, algunos de estos son adquiridos durante la perforación o una vez se finaliza esta.

La falta de información o el uso de información sin un tratamiento adecuado pueden generar modelos que no representan de manera adecuada el yacimiento. De continuar usando información sin un tratamiento adecuado, se seguirán generando modelos que no representan ni predicen las características del yacimiento, por tanto, el riesgo de seguir perforando pozos en una

ubicación equivocada, con baja rentabilidad económica representa un alto costo para las empresas operadoras.

La estimación y los modelos de distribución areal de propiedades petrofísicas son útiles al momento de planificar nuevos pozos para un óptimo desarrollo del campo o delimitación del mismo, sin embargo, se pueden presentar situaciones donde el modelo propuesto para una propiedad no representa de manera adecuada el yacimiento o presenta un alto grado de incertidumbre, esto se debe a que la información geológica obtenida directamente en campo es escasa, o también a que la información que se tiene no ha tenido un tratamiento adecuado previamente a la generación de modelos.

De acuerdo con lo anterior se hace necesario mejorar los modelos de distribución areal de propiedades petrofísicas en un yacimiento, por tal razón surge el presente trabajo donde se presenta una metodología para estimar la distribución areal de propiedades petrofísicas haciendo uso de la prueba Gamma y técnicas geoestadísticas mediante el uso de software libre como Rstudio. Esta metodología servirá como guía a todas aquellas personas que están involucradas en la caracterización de yacimientos, quienes constantemente generan y/o actualizan estos modelos y en el área de desarrollo de campos o para la delimitación de estos.

El alcance de este trabajo de monografía es presentar una metodología para la estimación areal de propiedades petrofísicas mediante uso Rstudio (software libre y de código abierto), dando previamente un tratamiento a los datos de entrada (registros geofísicos) mediante la prueba Gamma y posteriormente mediante el uso de técnicas geoestadísticas poblar propiedades

petrofísicas (ejemplo porosidad, permeabilidad) en zonas donde no se tiene información y de esta manera mejorar los modelos de distribución de propiedades.

La metodología en este caso fue desarrollada con datos de registros geofísicos de un yacimiento ubicado en la provincia de Alberta en Canadá, esta se dividió en seis (6) fases, en la fase cero (0) se realizó el inventario y se seleccionaron los pozos con los tres registros geofísicos de porosidad: densidad (RHOB), neutrón (NPHI), sónico (DT) y selección del intervalo de interés, en la fase un (1) se realizó el Análisis Exploratorio de los datos, en la fase (2) selección del mejor registro para el cálculo de porosidad aplicando la prueba Gamma, en la fase tres(3) cálculo de porosidad y correcciones ambientales, en la fase cuatro (4) cálculo de permeabilidad por los métodos empíricos de Timur, Tixier y Coates, en la fase cinco (5) construcción de semivariograma y modelos de estimación de las propiedades de porosidad y permeabilidad.

Finalmente se presentan los resultados, conclusiones y recomendaciones de la metodología propuesta.

2. Objetivos

2.1. Objetivo General

Proponer una metodología para la estimación areal de propiedades petrofísicas usando un software libre para el tratamiento de los datos y la interpolación geoestadistica.

2.2. Objetivos Específicos

- Realizar un análisis exploratorio de los registros geofísicos usados como dato de entrada mediante estadística descriptiva.
- Validar los datos de entrada considerando el modelamiento no lineal usando la función Prueba Gamma para el cálculo de porosidad.
- Estimar la porosidad efectiva considerando las correcciones ambientales propias de una evaluación petrofísica y la selección del mejor dato de entrada de acuerdo con los resultados obtenidos con la prueba Gamma.
- Calcular la permeabilidad del yacimiento objeto de estudio, mediante métodos empíricos.
- Generar distribuciones de porosidades y permeabilidades usando técnicas geoestadísticas como método de interpolación.

3. Antecedentes

Para poder estimar cuanto puede llegar a producir un yacimiento es importante conocerlo, es decir, tenerlo caracterizado. Schlumberger (1989) sugiere que:

> Los principales parámetros petrofísicos necesarios para evaluar un reservorio son la porosidad, saturación de hidrocarburos, espesor, área y permeabilidad. Adicionalmente, la geometría del reservorio, la temperatura y presión de formación, y litología son importantes en la evaluación, completamiento y producción de un reservorio. (p.2-1)

Desde hace años los ingenieros de yacimientos han encontrado en la geoestadistica, más precisamente en el modelamiento estocástico, una herramienta para generar modelos más precisos de yacimientos con heterogeneidades, la incertidumbre de la distribución espacial de los parámetros es modelada siendo conscientes de la existencia del riesgo e incertidumbre del modelo generado Journel (1990). Entre los métodos de interpolación en geoestadistica más conocidos para poblar espacios donde no se tiene información se tiene el Kriging y co-kriging, estos métodos se han usado para poblar con parámetros petrofísicos, pero sin antes considerar el error inherente al dato, los datos deben estar libres de error, porque si el dato es anormal o anómalo podría no ser representativo, independiente de la precisión del modelo. En un caso de estudio Ochie & Rotimi (2018) seleccionan nueve pozos con registros de pozo caliper, gamma ray, densidad, densidad neutrón, densidad porosidad, SP, resistividad y acústicos y proponen cinco flujos de trabajo para estimar la varianza, entre ellos la prueba Gamma, la cual es una técnica no paramétrica que sirve para evaluar la calidad de los datos Koncar (1997), en las dos

primeras se calcula la permeabilidad por los métodos de Kriging y cokriging, en la tercera se calculó la porosidad después de aplicar la prueba Gamma seleccionando las combinaciones de registros con valor de Vratio más pequeño en cada pozo y en las dos últimas, sin aplicar la prueba Gamma, como resultado se tuvo que la prueba Gamma evaluó la mejor combinación de registros y con esto se construyó el modelo del yacimiento.

Un par de años antes a Ochie & Rotimi (2018) la prueba Gamma fue calificada como un revolucionario estimador de ruido por Iturrarán et al. (2014), y fue usada como una herramienta para seleccionar la mejor combinación de atributos sísmicos para estimar la permeabilidad, porosidad, y la mejor combinación de registros de pozo para estimar la atenuación intrínseca 1/Q.

La prueba Gamma también fue usada para modelar un componente hidrogeológico clave como lo es la radiación solar en Brue Catchment (UK) por Remesan et al. (2008), en el estudio se logró resolver la importancia de los datos de entrada y optimizar el número de puntos requeridos para construir un modelo confiable. La relación entre los datos meteorológicos de entrada y salida fue lograda estimando el error de la varianza antes del modelado usando la prueba Gamma.

4. Marco de referencia

4.1. Definición y clasificación de la Estadística

En la literatura se encuentra una gran variedad de definiciones de lo que es la estadística, Urías & Salvador (2014) la definen como un conjunto de métodos para coleccionar, resumir, organizar, presentar y analizar la información de un conjunto de datos y puede clasificarse en descriptiva e inferencial:

4.1.1. Estadística descriptiva.

Conjunto de métodos que se utilizan para organizar, clasificar y presentar la información contenida en los datos por medio de graficas o tablas, así como el conjunto de medidas para indicar su dispersión y su localización. Esta rama de la estadística presenta una descripción de la información contenida en los datos.

4.1.2. Estadística inferencial.

Conjunto de métodos que se utilizan para deducir alguna característica de la población con únicamente información parcial. Con esta rama de la estadística se infiere como es la población de donde se obtuvo un subconjunto de datos. (p.2)

Otra definición la presenta Martin & de Paz Santana (2007), definiendo la estadística descriptiva como:

> La estadística descriptiva tiene como objetivo la descripción del colectivo estudiado, este puede ser de toda la población (Censo) o de una parte de esta (muestra). Cuando en una

> Población dada se selecciona adecuadamente un grupo de individuos: Muestra, la descripción de estos es el objetivo de la Estadística descriptiva. (p.33)

De acuerdo con Solano & Alvarez (2005) "La Estadística Descriptiva se compone de aquellos métodos que incluyen técnicas para recolectar, presentar, analizar e interpretar datos" (p. 1.1.3). De esta manera es darles un manejo a los datos recopilados en cuanto a la forma de ordenarlos y presentarlos, y así se puede lograr poner en evidencia características en una forma objetiva y útil.

4.2. Análisis Estadístico o Exploratorio de Datos

Antes de aplicar cualquier técnica estadística se debe realizar un Análisis Exploratorio de Datos (AED). De esta manera el analista logra un entendimiento básico de los datos y las relaciones que existen entre las variables que se analizan.

Con el AED se pueden organizar y preparar los datos, también detectar fallas en el diseño y en la recolección de estos, dar un tratamiento y evaluación de los datos ausentes, identificar casos atípicos y comprobar los supuestos subyacentes en la mayor parte de las técnicas multivalentes. Antes de cualquier análisis de datos es necesario un examen visual de estos. El AED comparte los mismos objetivos que la Estadística Descriptiva, pero también busca detectar anomalías o errores en las distribuciones univariantes de los datos, descubrir en los datos patrones o modelos. Para esto ha incorporado nuevas técnicas gráficas y busca estadísticos resistentes y robustos basados en el orden y centrado en la mediana Martin & de Paz Santana (2007).

Según Martin & de Paz Santana (2007) Si las distribuciones son univariantes, el AED nos informara sobre:

- La localización, la desviación y la forma de distribución de los datos.
- La simetría o asimetría de estos.
- El número y localización de agujeros (vacíos) y puntas en la distribución.
- Presencia y número de valores alejados.

Los componentes principales del AED son:

- Las representaciones gráficas nos revelan visualmente el comportamiento de los datos y la estructura del conjunto.
- Pone mucha atención en el análisis, es decir, en las diferencias que hay entre los datos reales y el resultado de su ajuste a un modelo previamente determinado o subyacente.
- Utiliza la transformación de los datos, que consiste en encontrar la escala que más simplifique y clarifique el análisis, como, ejemplo, con el uso de funciones matemáticas simples, como raíz cuadrada, logaritmos, etc.
- Valora la resistencia que representan algunos estadísticos, que les hace poco sensibles a la influencia de uno o varios valores sensiblemente distantes de la mayoría de la distribución. la mediana resistente, mientras que la media no.
- Busca estadísticos (robustos) que presentan algunos estadísticos, que les hace poco sensibles a desviaciones de los supuestos básicos. Un estadístico robusto es válido sin que

la forma, el tipo el modelo de la distribución tengan un peso importante en su cuantía. (p.33-34)

En Rstudio existe la función AED (Exploratory Data Analysis) con la cual se obtiene las gráficas de histograma, boxplot y Q-Q plot.

4.2.1. Histograma.

Indican que una manera de presentar los datos de manera clara y sucintamente es en forma de histogramas o gráficas de barras. En un histograma se pueden resumir los datos en una escala continua y mostrar la distribución de frecuencia de alguna característica como la tendencia central y la dispersión de los datos. En una gráfica de barras se representa la frecuencia con la cual se presenta las características de los datos, y su altura indica el número de veces que una característica fue observada Krajewski & Ritzman (2000).

Dependiendo del número de picos que se presenten en el histograma, este se clasificar en unimodal, cuando la curva sube, se presenta un solo pico y después baja, bimodal cuando presenta dos picos distintos, esta bimodalidad resulta cuando el conjunto de datos consiste en observaciones de dos clases bastante distintas de individuos u objetos, y por último multimodal cuando presenta más de dos picos. Por otro lado, puede ser simétrico o presentar un sesgo, es simétrico si la mitad izquierda es imagen especular de la mitad derecha, sesgo positivo si el extremo derecho o superior se extiende comparado con el extremo izquierdo o inferior, sesgo negativo si se extiende a la izquierda Devore (2008)

4.2.2. Boxplot.

El boxplot o gráfica de caja y bigote, es una manera de presentar o reflejar propiedades de una muestra Walpole et al. (1992). En este tipo de gráfico los datos se encuentran distribuidos en cuatro grupos con igual cantidad de datos, con un 25% c/u Rojas (2013). Para construir un boxplot se deben organizar los datos en orden ascendente, luego se identifican cinco (5) números: valor mínimo, valor máximo, mediana o segundo cuartil (Q2), primer cuartil (Q1) y tercer cuartil (Q3), una vez hecho esto se pueden identificar dos rangos, un rango de los datos (valor máximo – valor mínimo) y otro rango llamado rango intercuartílico (RIC), este rango se calcula realizando la resta entre Q3-Q1, en este rango se agrupan el 50% de los datos. En el boxplot también se pueden visualizar datos con valores atípicos leves ecuación(1) y (2) y atípicos extremos ecuación (3) y (4).

Los valores de los datos atípicos son clasificados en leves y extremos como se muestra a continuación:

Valores atípicos leves. Para calcular los valores de los datos atípicos leves se debe conocer el rango intercuartílico (RIC), los valores atípicos leves corresponden a los valores que se encuentran 1.5 veces esa distancia de uno de esos cuartiles

$$Min = valor\ minimo - 1.5 * (RIC) \quad (1)$$

$$Max = valor\ maximo + 1.5 * (RIC) \quad (2)$$

Valores atípicos extremos. Para calcular los valores de los datos atípicos extremos se debe conocer el rango intercuartílico (RIC), los valores atípicos extremos corresponden a los valores que se encuentran 3 veces esa distancia de uno de esos cuartiles Uribe (2010)

$$Min = valor\ minimo - 3 * (RIC) \quad (3)$$

$$Max = valor\ maximo + 3 * (RIC) \quad (4)$$

Debe saberse que las pruebas estadísticas se utilizan para identificar valores atípicos, no para retirarlos del grupo de datos. De acuerdo con Murphy & Lau (2008) “Técnicamente, una observación no debe retirarse a menos que una investigación halle una causa probable para justificar esta acción.”

4.2.3. Gráfico Q-Q plot.

El Gráfico Q-Q plot, también conocido como gráfico cuantil-cuantil o gráfico de normalidad, es un tipo de gráfico usado para estudiar la hipótesis nula (muestra que proviene de una población con distribución normal), para no rechazar la hipótesis nula la gráfica debe tener una apariencia donde los puntos de este se encuentren alineados en una recta, de lo contrario se rechaza la hipótesis nula. Útil (2013). En el eje X se colocan los cuantiles teóricos de una normal, los cuales van de -3 a 3, en el eje Y las observaciones ordenadas de la muestra y se tiene una línea de referencia. Se considera que cuando la media y la mediana coinciden existe una hipótesis de normalidad, Shapiro Wilds y Kolmogorov Smirnov Rojas (2013)

4.3. Prueba Gamma

La prueba Gamma es un algoritmo para analizar datos, fue desarrollado por Koncar, N. junto con Stefánson y Antonia J. Jones en el año de 1997, este algoritmo estima el mejor Error Cuadrático Medio (MSE) de una variable dependiente, la cual se puede emplear en modelos de datos continuos o uniformes sin que estos sean sobre ajustados Koncar (1997). La prueba Gamma se utilizó para seleccionar una combinación de variables de entrada que se pueden emplear para generar un modelo más suave de salida Tsui & De Oliveira (2002). Una justificación matemática formal del método y una demostración de la prueba Gamma fue presentada por Evans et al. (2002) y Evans & Jones (2002)

4.3.1. Fundamentación teórica de la prueba Gamma.

De acuerdo con Everitt & Dunn (2001) “La prueba Gamma es una herramienta de análisis y modelamiento no lineal, la cual permite examinar la relación entre los datos de entrada y salida” también ayuda a seleccionar los mejores datos de entrada antes de calibrar y probar modelos, reduciendo la incertidumbre encontrada durante la selección de datos de entrada Iturrarán et al. (2014). La prueba Gamma puede ser considerada una equivalente no lineal del error de la suma de los cuadrados usada en la regresión Kemp et al. (2004), la cual remueve errores de los datos por la estimación de la varianza de los datos de salida y que no puede explicarse con un modelo suavizado, esto responde al ruido en el conjunto de datos. Un modelo suave en el cual la transformación de entrada a salida es continua.

Según Kemp et al. (2004) “la prueba Gamma trabaja bajo la suposición que si dos puntos ***x´ y x*** están cercanos en el espacio de entrada entonces sus correspondientes salidas ***y’ e y*** debe ser cercanos en el espacio de salida. Si la salida no es cercana entonces se considera que esta diferencia se debe al ruido.” (p.67-68)

Supongamos que se tienen unos datos de la forma como se muestra en la ecuación (5)

$$\{(xi, yi), 1 \leq i \leq M\} \quad (5)$$

En la cual $\boldsymbol{x \in R^m}$ en la entrada y el correspondiente $\boldsymbol{y \in R}$ en la salida, donde se supone que $\boldsymbol{x}$ contiene factores predictivamente útiles para influir en la salida $\boldsymbol{y}$. La única suposición es que el sistema es de la forma como se muestra en la ecuación (6)

$$y = f(x1 \ldots xm) + r \quad (6)$$

Donde $\boldsymbol{f}$ es una función suavizada, y $\boldsymbol{r}$ es una variable aleatoria que representa el ruido. Se puede asumir que la media de la distribución de $\boldsymbol{r}$ es cero, ya que cualquier sesgo constante puede ser incluido en la función desconocida **(*f*)** y que la varianza del ruido ***Var (r)*** es restringida. El dominio de posibles modelos es ahora restringido a la clase de modelos suaves que tienen derivadas parciales limitadas. El estadístico Gamma $\boldsymbol{\Gamma}$ es el estimado de esa parte de la varianza de la salida que no puede ser explicado por un modelo de datos suavizado.

$x_N[i, k]$ Denota el k-esimo vecino más cercano, en términos de distancia Euclidiana $x_i(1 \leq i \leq M), (1 \leq k \leq p)$, ecuaciones (7) y (8)son las principales ecuaciones para el cálculo de $\boldsymbol{\Gamma}$**.**

$$\delta_M(k) = \frac{1}{M}\sum_{i=1}^{M} \lceil x_N[i,k] - \ x_i \rceil^2 \ (1 \le k \le p) \tag{7}$$

Donde $\lceil . \rceil$ la distancia Euclidiana,

$$\Upsilon_M(k) = \frac{1}{2M}\sum_{i=1}^{M} \lceil y_N[i,k] - \ y_i \rceil^2 \ (1 \le k \le p) \tag{8}$$

La manera de calcular **Γ** es por una regresión lineal de mínimos cuadrados construido con los puntos $\boldsymbol{p}\ \delta_M(k), \Upsilon_M(k)$ .El intercepto en la vertical es el valor de **Γ** ver ecuación (9)

$$\Upsilon_M(k) \to Var(r)\ en\ probabilidad\ como\ \delta_M(k)\ \to 0 \tag{9}$$

La justificación formal del método puede verse en detalle en Evans & Jones (2002) y Evans et al. (2002)

Vratio. puede calcularse a partir de Γ y σ2 (y) como se muestra en ecuación (10)

$$\boldsymbol{\nu} = \frac{\boldsymbol{\Gamma}}{\boldsymbol{\sigma^2(y)}} \tag{10}$$

Donde σ^2 ***(y)*** es la varianza de la salida ***y***. Un valor de ***Vratio*** cercano a cero indica que existe un alto grado de predictibilidad dado un valor de salida ***y***.

Según Monte (1999) Si el valor de V ratio es cercano a uno el valor de salida es equivalente a un camino aleatorio. (p.143-148)

Algoritmo KNN. De acuerdo con Lantz (2013):

El Algoritmo KNN en Rstudio utiliza la aproximación de los vecinos más cercanos para realizar su clasificación, presenta algunas fortalezas como que es simple y efectivo, no hace suposiciones sobre la distribución de datos subyacentes y presenta una fase de entrenamiento rápida. Por otro lado, se encuentran debilidades como: no produce un modelo, lo que limita la capacidad de encontrar relaciones entre las características, presenta una fase de clasificación lenta, requiere una gran cantidad de memoria y las características nominales y datos perdidos requieren un procesamiento adicional. (p.67-74)

Cálculo de la distancia. Según Lantz (2013):

Existen diferentes maneras para calcular la distancia, pero el algoritmo KNN tradicionalmente usa la distancia Euclidiana. La distancia Euclidiana es especificada a continuación, donde **p y q** son los ejemplos para comparar y cada uno tiene **n** características. El término $\mathbf{p_1}$ se refiere al valor de la primera característica del ejemplo *p* mientras que q_1 se refiere al valor de la primera característica del ejemplo **q** ver ecuación (11) (p.70)

$$\boldsymbol{dist}\,(\boldsymbol{p},\boldsymbol{q}) = \sqrt{(\boldsymbol{p_1}+\boldsymbol{q_1})^2 + (\boldsymbol{p_2}+\boldsymbol{q_2})^2 + \ldots + (\boldsymbol{p_n}+\boldsymbol{q_n})^2} \qquad (11)$$

Selección apropiada del valor de K. Como lo indica Lantz (2013)

> Poder determinar cuántos vecinos **k** usar para KNN determina que también el modelo se generalizará a datos futuros, pues debe existir un balance entre sobre ajustar o sub ajustar los datos de entrenamiento y de esta manera evitar un problema conocido como compensación de sesgo o equilibrio sesgo varianza. En el caso de dar un valor alto a k se reduce el impacto o la variación causada por datos ruidosos, pero se puede sesgar y correr el riesgo de ignorar patrones pequeños e importantes. (p.71)

"En la práctica el valor de ***k*** depende de la dificultad del concepto que se va a aprender y del número de registros en los datos de entrenamiento" Lantz (2013), usualmente el valor dado a ***k*** esta entre 3 y 10. Por otro lado, una práctica común para elegir el valor de ***k*** es calcular la raíz cuadrada del número de datos de entrenamiento, realizando algunas aproximaciones cuando el número obtenido no es entero.

Estas reglas no siempre dan el mejor valor de ***k***, entonces surge una alternativa de probar varios valores de ***k*** en varios conjuntos de prueba y elegir el que ofrece el mejor rendimiento de clasificación. "Por otro lado, a menos que los datos sean muy ruidosos, los conjuntos de datos de entrenamiento más grandes y representativos pueden hacer que la elección de ***k*** sea menos importante. Esto se debe a que incluso los términos más sutiles tendrán un conjunto suficientemente grande de ejemplos para votar como vecinos más cercanos" Lantz (2013).

Preparación de los datos para el uso de KNN. De acuerdo con Lantz (2013)

Antes de aplicar el KNN los datos generalmente son transformados a un rango estándar. La razón de este paso es que la fórmula de la distancia depende de cómo se miden las

características. Pueden existir características que tienen valores mucho más grandes que otros, entonces las mediciones de distancia estarán fuertemente dominadas o influenciadas por los mayores valores. Es aconsejable que los valores de las características que se quieren medir se encuentren en una misma escala. Lo que se busca es una forma de "reducir o re escalar" las diversas características de tal manera que cada una contribuya relativamente por igual en la fórmula de la distancia. (p.72)

El método tradicional de re escalar características o propiedades de estudio para el KNN es mediante normalizaciones mínimo - máximo como se muestra en la ecuación (12)

$$x_{new} = \frac{x - min\ (x)}{max(x) - min\ (x)} \tag{12 }$$

Otra manera es mediante variable estandarizada o normalizada como se muestra en la ecuación (13) Journel & Huijbregts (1978)

$$x_{new} = \frac{X - \mu}{\sigma} = \frac{X - \text{Mean}\ (X)}{StdDev(X)} \tag{13 }$$

4.4. Definición de Geoestadistica.

El término geoestadística presenta varias definiciones, Giraldo (2002) presenta una recopilación de varias de estas dadas entre el año 1978 a más recientes "La geoestadística es una rama de la estadística que trata fenómenos espaciales Journel & Huijbregts (1978). Su interés primordial es la estimación, predicción y simulación de dichos fenómenos Warrick & Myers (1987). Esta herramienta ofrece una manera de describir la continuidad

espacial, que es un rasgo distintivo esencial de muchos fenómenos naturales, y proporciona adaptaciones de las técnicas clásicas de regresión para tomar ventajas de esta continuidad Isaaks & Srivastaca (1989). Por otro lado Petitgas (1996), la define como "una aplicación de la teoría de probabilidades a la estimación estadística de variables espaciales". (p.17), sin embargo, la geoestadística es solo una de las áreas del análisis de datos espaciales.

4.4.1. Historia de la Geoestadística.

Según Giraldo (2002) los fenómenos con correlación espacial con métodos geoestadísticos comenzaron a estudiarse en los años 60´s con el propósito de predecir valores de las variables en sitios no muestreados. Como antecedentes se citan los trabajos realizados por Sichel, Krige, Materon y Samper y Carrera

Sichel (1947) Observó una distribución asimétrica en el contenido de oro en unas minas sudafricanas, la equiparó a una distribución de probabilidad Log-normal y desarrolló fórmulas para esta distribución. De esta manera propone una primera estimación de las reservas, esto bajo el supuesto que las mediciones eran independientes, en una clara contradicción puesto que la experiencia muestra que existen algunas zonas con más contenido que otras.

Krige (1951) Tratando de darle solución al problema de Sichel propone una variante del método de medias móviles, el cual es considerado como el equivalente al krigeado simple.

Matheron (1962) Realiza una formulación más rigurosa al problema de predicción.

Posteriormente Calvete & Ramirez (1990) Indican que en los años sucesivos la teoría se ha depurado, ampliando su campo de validez y reduciendo las hipótesis necesarias.

De la minería las técnicas geoestadísticas se han "exportado" a muchos otros campos como hidrología, física del suelo, ciencias de la tierra y más recientemente al monitoreo ambiental y al procesamiento de imágenes de satélite.

Robertson (1987) afirma que "Aunque la aplicación de herramientas geoestadísticas es bastante resiente, son innumerables los ejemplos en los que se ha utilizado esta técnica en estudios ambientales con el ánimo de predecir fenómenos espaciales" y resalta la importancia de la geoestadística en la siguiente frase "La columna vertebral del análisis geoestadístico es la determinación de la estructura de autocorrelación entre los datos y su uso en la predicción a través de las técnicas conocidas como Kriging y co-Kriging". (p.6-7)

De acuerdo con Viera & González (2002) el análisis geoestadístico está compuesto por tres (3) etapas: (a) análisis exploratorio de los datos, (b) análisis estructural y (c) predicciones (Kriging o simulaciones).

La primera de estas, el Análisis exploratorio de los datos ya fue tratado anteriormente, ahora se tratará el Análisis estructural y las predicciones (kriging o simulaciones).

4.4.2. Análisis estructural.

Según Viera & González (2002) “El análisis estructural se encarga de la caracterización de la estructura espacial de una propiedad o fenómeno regionalizado”. (p.3) y partir de este se obtiene un modelo geoestadístico para la función aleatoria que se estudia.

Es decir, el análisis estructural estima y modela una función que refleja la correlación espacial de la variable regionalizada a partir de la hipótesis más adecuada acerca de su variabilidad. Es decir, en dependencia de las características de estacionariedad del fenómeno se modelará la función de covarianzas o de la semivarianzas.

En el análisis estructural se calcula el variograma experimental utilizando una función de correlación espacial, bien sea la semivarianza o covarianza. De acuerdo con Garcia & Guevara (2004):

> A la nube de puntos generada en el semiovariograma experimental se busca el modelo teórico (Esférico, Exponencial, Gaussiano o Potencial) que mejor se ajuste y se obtiene el modelo del semiovariograma, se define el número y tamaño (se recomienda que el tamaño del Lag sea similar a la distancia promedio que existe entre la ubicación espacial de los datos) de los Lag que se van a utilizar en el modelo. (p.15)

4.4.3. Interpolación y métodos de interpolación espacial.

Según la página web de Geoinnova:

> La interpolación espacial de datos es una parte de la geoestadística que se basa en el cálculo de los valores desconocidos de una variable espacial a partir de otros valores cuyo

valor es conocido. Este cálculo es uno de los principales análisis geoestadísticos que se llevan a cabo cuando se trabaja con datos espaciales y modelos predictivos. El resultado obtenido mediante el proceso de interpolación se conoce habitualmente por el nombre de superficie estadística, una superficie continua con valores interpolados a partir de otros conocidos.

Los métodos de interpolación espacial permiten generar superficies continuas a partir de medidas en localizaciones puntuales (muestras o sitios muestrales), y los métodos se pueden clasificar en Deterministas o Geoestadísticos.

Deterministas. Generan superficies continuas mediante el grado de similitud o suavizado. Dentro de esta categoría encontramos los métodos globales, locales, IDW y Spline.

Geoestadísticos. Generan superficies continuas a partir de las propiedades estadísticas de los datos de partida. Dentro de esta categoría encontramos Kriging y Co-Kriging Quesada (2020)

4.4.4. Técnicas Geoestadísticas.

Kriging. De acuerdo con Giraldo (2002):

El Kriging está compuesto por un conjunto de métodos de predicción espacial, los cuales tienen su fundamentación en la minimización del error cuadrático medio. Estos métodos se aplican con frecuencia con el propósito de predecir, pero también pueden ser aplicados en la simulación y en el diseño de redes óptimas de muestreo. (p.32)

Según Viera & González (2002) las definiciones de kriging simple y ordinario presentadas se muestran a continuación, sin embargo, para cada una de estas se presenta un fundamento matemático un poco más profundo, el cual se puede consultar en Giraldo (2002).

> *Kriging simple.* Asume que las medias locales son relativamente constantes e iguales a la media de la población, la que es bien conocida. La media de la población es usada como un factor en cada estimado, con todas las muestras en la misma vecindad. No es un método muy usado.
>
> *Kriging Ordinario.* Asume que la media local no está necesariamente cercana a la media de la población, y que usa solamente para el estimado la muestra para la vecindad local. Es el método usado más comúnmente por su robustez. (p.134)

4.5. Porosidad

La porosidad se puede calcular a partir del registro de densidad como se muestra en la ecuación (14)

$$\emptyset = \frac{\rho_{matriz} - \rho_{registro}}{\rho_{matriz} - \rho_{fluido}} \qquad (14\ )$$

4.5.1. Distribución de los minerales de arcilla.

La presencia de minerales de arcilla afecta las propiedades como porosidad efectiva, saturación de agua efectiva y permeabilidad y por ende la producción del yacimiento. Las arcillas se clasifican según su distribución en laminadas, dispersas y estructurales Ruhovets & Fertl (1982). En las ecuaciones (18), (19) y (20) se expresa la manera para el cálculo de la porosidad efectiva de acuerdo a su distribución en el yacimiento.

Arcillas laminadas.

$$\emptyset_e = \emptyset_{max} - (1 - V_{cl}) \quad (15)$$

Estas arcillas se depositan en forma de láminas dentro de la arena, las propiedades de la porosidad y permeabilidad no son afectadas y su composición debe ser similar a las arcillas cercanas.

Arcillas dispersas.

$$\emptyset_e = \emptyset_{max} - V_{cl} \quad (16)$$

Estas arcillas se encuentran dispersas llenando espacios intergranulares, pueden generar disminución de la porosidad, y pueden diferir de las cercanas.

Arcillas estructurales.

$$\emptyset_e = \emptyset_{max} \quad (17)$$

Los granos de cuarzo han sido sustituidos por partículas de arcilla, la porosidad de la arena es poco afectada, la composición de las arcillas puede diferir de las arcillas cercanas.

4.6. Permeabilidad

De acuerdo con Guamialamá et al. (2020) el cálculo de la permeabilidad puede ser estimado usando los registros de pozos y empleando métodos empíricos, métodos estadísticos y medidas virtuales. Los métodos empíricos están basados en una correlación entre la permeabilidad, la porosidad y la saturación de agua irreducible. Kozeny (1927) fue el primero en proponer una relación entre las propiedades de la roca y la permeabilidad, luego Tixier (1949) y

los trabajos de Wyllie & Rose (1950) extiende la relación empírica propuesta en 1949, Sheffield (1956) basado en la ecuación de Kozeny (1927) propone una nueva ecuación para estimar la permeabilidad, Timur (1968) basado en los trabajos de Kozeny (1927) y Wyllie & Rose (1950) encuentra una nueva relación, Coates & Denoo (1981) proponen una nueva ecuación, la cual satisface permeabilidad cero cuando la porosidad es cero y cuando la saturación de agua es 100% la formación debe estar en saturación de agua irreducible.

4.6.1. Método de Tixier.

Usando relaciones empíricas entre la resistividad y la saturación de agua, entre la saturación de agua y la presión capilar, y la presión capilar y la permeabilidad y continuando con el trabajo Tixier (1949) logra establecer una relación, ecuación (18)

$$K_{Tixier} = \left(\frac{250 * DPHIe^3}{swirr}\right)^2 \quad (18)$$

4.6.2. Método de Timur.

Es un método empírico para estimar la permeabilidad a partir de la porosidad y la saturación de agua irreducible Timur (1968), como se muestra en la ecuación **¡Error! No se encuentra el origen de la referencia.** existe un factor que depende de la densidad del hidrocarburo, 250 para hidrocarburos con densidad de 0.8 g/cm^3 y 79 para gas.

$$K_{Timur} = \left(\frac{100 * DPHIe^{2.25}}{swirr}\right)^2 \quad (19)$$

4.6.3. Método de Coates.

Al igual que otros métodos empíricos Coates & Denoo (1981) y Coates & Dumanoir (1973) presentan una ecuación (20) para estimar la permeabilidad a partir de la porosidad y la saturación de agua irreducible.

$$\boldsymbol{K_{Coates} = \ (70 * DPHIe^2 * \frac{1 - swirr}{swirr})^2} \qquad (20\ )$$

Donde:

DPHIe: es la porosidad efectiva

Swirr: es la saturación de agua irreducible

5. Marco Metodológico

5.1. Tipo de investigación

Este proyecto se desarrolló como una investigación práctica donde se utilizó el algoritmo de la prueba Gamma para seleccionar entre los registros geofísicos de porosidad RHOB, NPHI y DT el más apropiado para calcular la propiedad petrofísica de porosidad. La prueba Gamma es un algoritmo utilizado para el análisis de datos con el cual se estima el error cuadrático medio de la variable dependiente, puede ser utilizado en modelos de datos continuos o uniformes buscando que el modelo que se genere no sea sobre ajustado. La prueba Gamma no debe ser confundida con la distribución de probabilidad Gamma usada en estadística.

5.2. Método

El proyecto se desarrolló en seis (6) fases (Figura1). En la fase cero (0) se realizó el inventario, selección de pozos de acuerdo a los registros geofísicos y definición del intervalo de interés, en la fase uno (1) usando el software libre R versión 3.6.3 (2020-02-29) con el paquete PASWR se realizó el Análisis Exploratorio de los Datos, en la fase dos (2) con el paquete RANN la prueba Gamma para los registros geofísicos de porosidad RHOB, NPHI y DT, en la fase tres (3) se calculó la porosidad realizando algunas correcciones ambientales, se estimó porosidad efectiva del intervalo y se calculó la permeabilidad por los métodos de Timur, Tixier y Coates,

finalmente, en la fase cinco (5) se estimó la distribución de las propiedades petrofísicas de porosidad y permeabilidad empleando técnicas geoestadísticas.

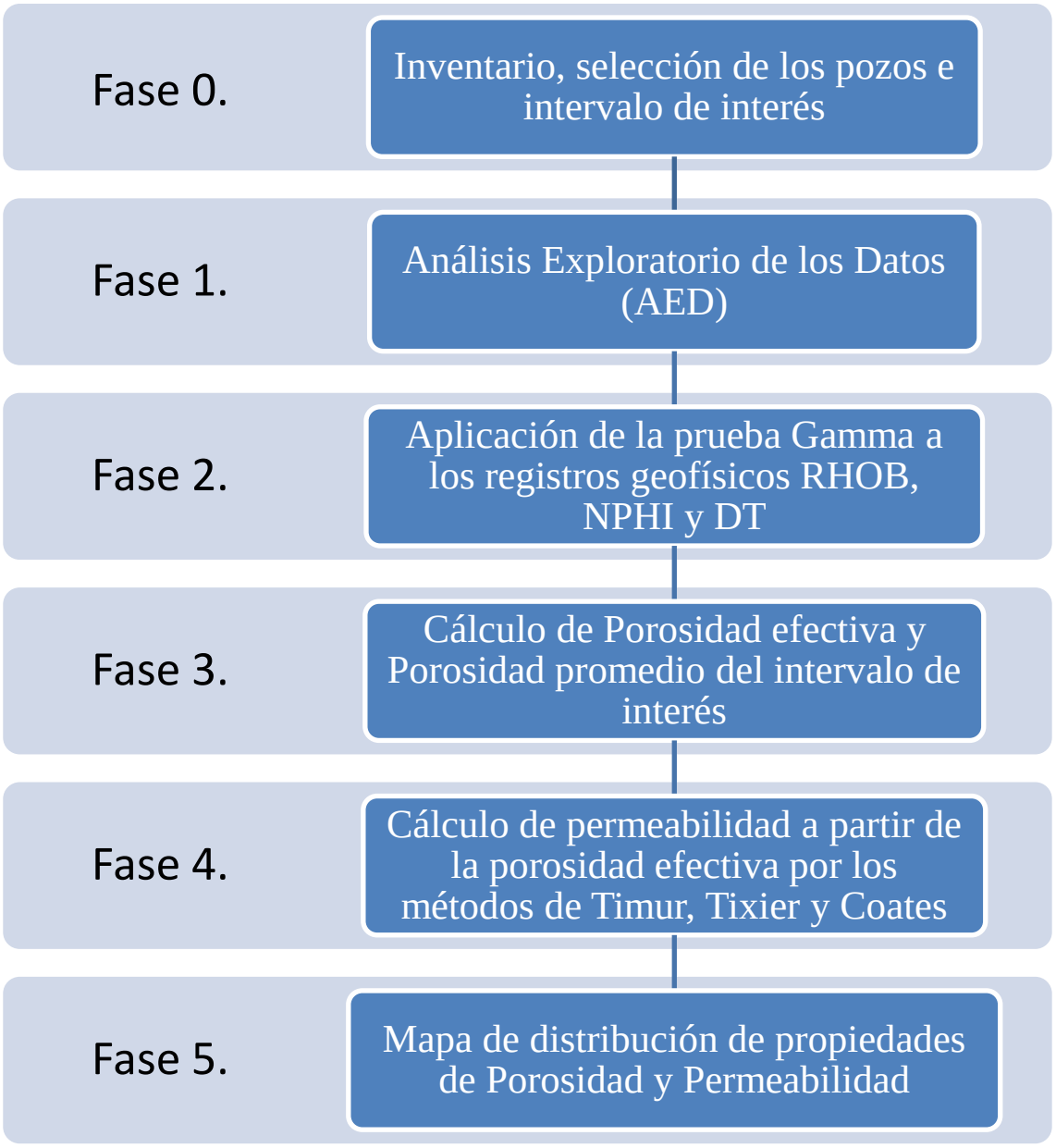
5.3. Fuente de la información

La información de los pozos y sus registros geofísicos corresponden a un yacimiento en la provincia de Alberta en Canadá, los archivos en formato (.LAS) muestran las profundidades y los registros adquiridos en cada pozo, pero no reporta información en el encabezado, esta información es necesaria para realizar algunos cálculos, como no se cuenta con esta información, se asumió un valor acorde a yacimientos de la misma provincia.

5.4. Análisis de los datos y tratamiento de la información

La información en su mayor parte fue tratada y procesada con el software libre R versión 3.6.3 (2020-02-29) con el uso de algunas paquetes y librerías de acuerdo con cada fase del proyecto.

Figura1 *Fases del proyecto.*



5.4.1. Inventario, selección de los pozos e intervalo de interés.

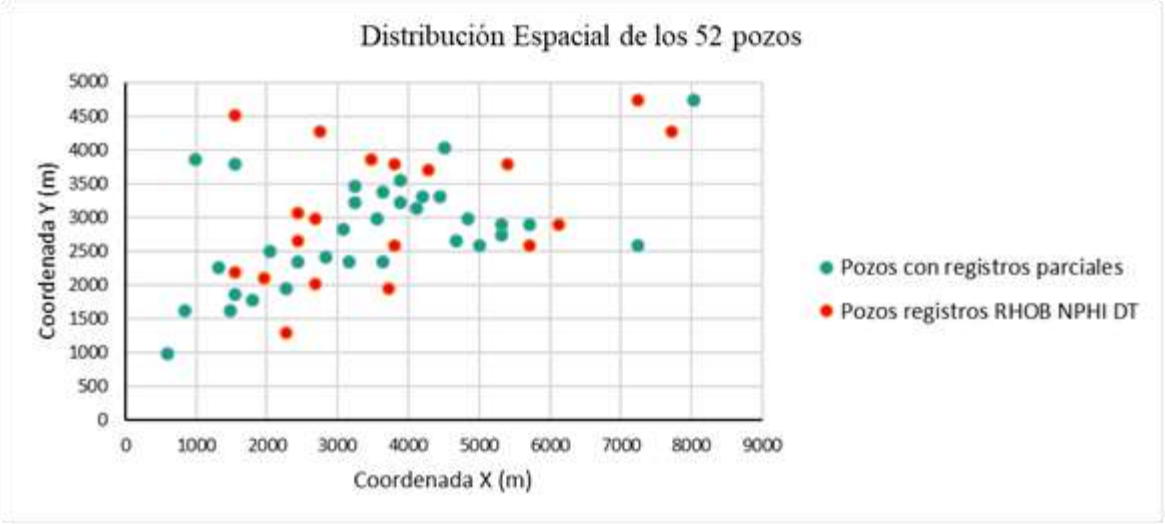
Inventario y selección de los pozos. Se muestran las coordenadas de los 52 pozos, en color rojo aparecen los pozos que tienen información de los tres (3) registros geofísicos de porosidad RHOB, NPHI y DT en formato (.LAS) y en color verde los que tienen información parcial, es decir uno (1) o dos (2) de los tres (3) de los registros (Tabla *1*). Se muestra la distribución espacial de los 52 pozos (Figura 2).

Tabla 1
Coordenadas espaciales de los 52 pozos

Pozo	Coordenadas X (m)	Coordenadas Y (m)
01_12	1556.74	1862.6
01_17	4436.74	3302.6
01_18	3236.74	3462.6
02_07	2676.74	2022.6
02_17	4196.74	3302.6
03_07	2276.74	1942.6
03_08	3716.74	1942.6
03_17	3876.74	3542.6
04_07	1796.74	1782.6
04_17	3636.74	3382.6
05_07	1956.74	2102.6
05_17	3476.74	3862.6
06_07	2436.74	2342.6
06_08	3636.74	2342.6
06_09	5316.74	2742.6
06_13	996.74	3862.6
06_15	7716.74	4262.6
06_16	5396.74	3782.6
06_17	3796.74	3782.6
07_07	2836.74	2422.6

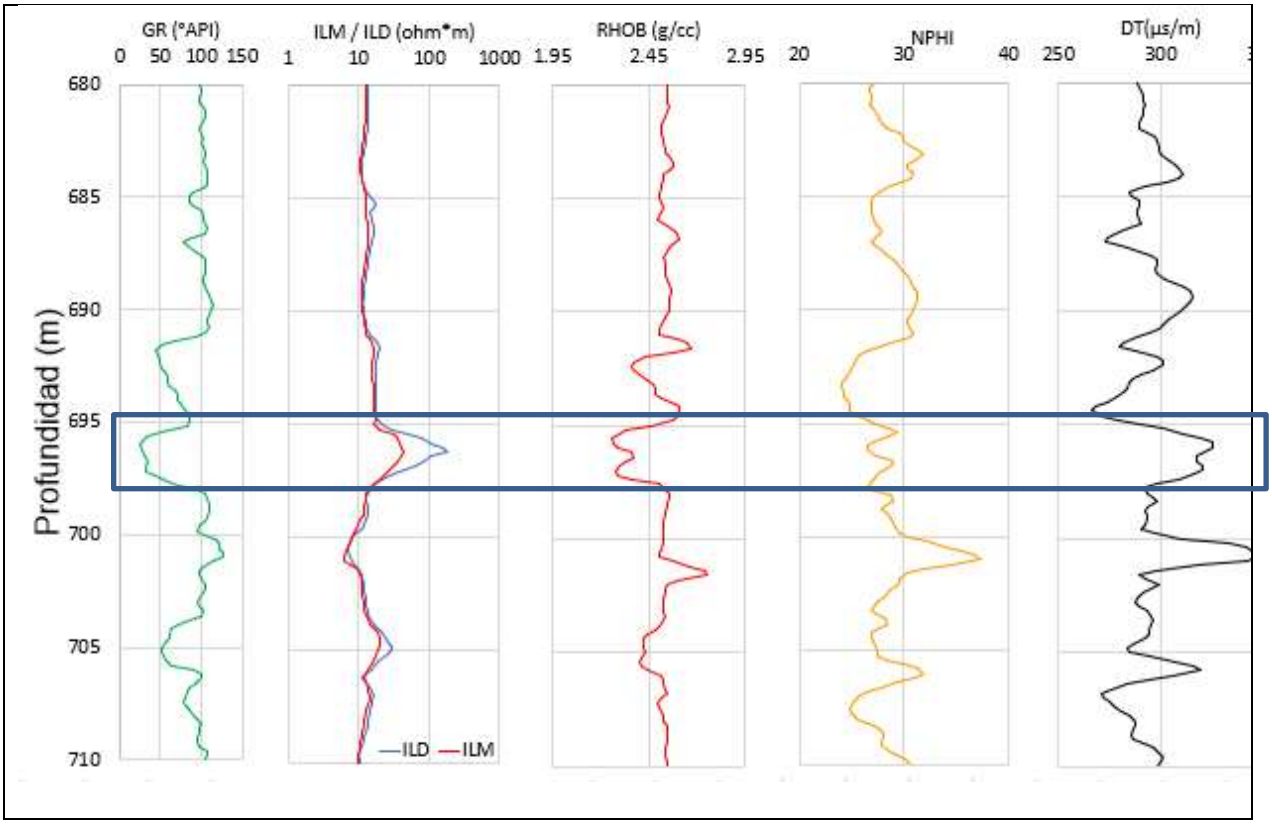
07_17	4276.74	3702.6
08_07	3156.74	2342.6
08_12	1556.74	2182.6
08_13	1556.74	3782.6
08_17	4516.74	4022.6
09_07	3076.74	2822.6
09_08	4676.74	2662.6
09_15	8036.74	4742.6
10_09	5716.74	2582.6
10_10	7236.74	2582.6
10_18	2756.74	4262.6
11_07	2436.74	2662.6
11_08	3796.74	2582.6
11_15	7236.74	4742.6
12_01	596.74	982.6
12_07	2036.74	2502.6
12_09	4996.74	2582.6
13_08	3556.74	2982.6
14_01	836.74	1622.6
14_06	2276.74	1302.6
14_07	2436.74	3062.6
14_08	3876.74	3222.6
14_09	5316.74	2902.6
15_07	2676.74	2982.6
15_08	4116.74	3142.6
15_09	5716.74	2902.6
16_01	1476.74	1622.6
16_07	3236.74	3222.6
16_08	4836.74	2982.6
16_09	6116.74	2902.6
16_13	1556.74	4502.6
8a_12	1316.74	2262.6

Figura 2 *Mapa de distribución espacial de los 52 pozos.*



Definición del intervalo de interés. El intervalo de interés se definió teniendo en cuenta la respuesta de los registros Gamma ray (GR), resistivos ILM e ILD, densidad (RHOB), neutrón (NPHI) y sónico (DT). Se encontró un intervalo de interés con valores de Gamma ray bajos característico de litologías arenosas, un aumento en la resistividad asociado a la presencia de un fluido (aceite o gas), disminución en la densidad por el aumento de porosidad, disminución de neutrón respondiendo a la cantidad de hidrógeno presente en la formación y aumento en el registro sónico, a manera de ejemplo se resalta el intervalo de interés en el pozo 02_07 entre 695m a 700m (Figura 3).

Figura 3. *Registros geofísicos GR, ILN, ILD, RHOB, NPHI y DT del pozo 02_07.*



5.4.2. Análisis Exploratorio de los Datos.

El Análisis Exploratorio de los Datos para los registros geofísicos RHOB, NPHI y DT para los 19 pozos se realizó con el paquete PASWR y con la función EDA, con el cual se obtiene una tabla resumen y al mismo tiempo las gráficas del histograma, boxplot y Q-Q plot

Datos atípicos. Para conocer los valores de los datos atípicos en cada uno de los registros y de cada pozo, se empleó la función boxplot.stats.

Eliminación de datos atípicos. En el caso que se quieran eliminar los datos que son atípicos, se debe tener instalado las librerías tidyverse y datasets.

5.4.3. Aplicación de la prueba Gamma a los registros geofísicos RHOB, NPHI y DT.

La prueba Gamma se aplicó usando combinaciones entre el registro de profundidad y cada uno de los registros RHOB, NPHI y DT (Tabla 2), para que el algoritmo funcione correctamente se debe tener instalado el paquete RANN.

Tabla 2
Combinaciones de los registros para la prueba Gamma

Depth	RHOB	NPHI	DT
1	1	0	0
1	0	1	0
1	0	0	1

5.4.4. Cálculo de porosidad, Volumen de arcilla, porosidad efectiva y porosidad promedio

Cálculo de porosidad. Tomando el registro de densidad (RHOB) se calculó porosidad DPHI teniendo en cuenta las correcciones ambientales por diámetro de pozo, se debe conocer la densidad de matriz y densidad de fluido como se muestra en la ecuación (21)

Para la corrección por diámetro de pozo se asumió un diámetro de 12 in, con este diámetro la gráfica de FDC corrección de pozo, indica que el valor de densidad de matriz se debe sumar 0.02, se asumió un valor de densidad de matriz (dm=2.65 g/cm^3) y un valor de densidad de fluido (df=1.03 g/cm^3), estos valores son basados en Schlumberger (1991) donde 2.65 g/cm^3

es un valor típico para una matriz siliciclastica y el valor de densidad de fluido se encuentra entre 1-1.1 dependiendo de la salinidad, temperatura y presión.

$$\emptyset = \frac{\rho_{matriz-corregida} - \rho_{registro}}{\rho_{matriz-corregida} - \rho_{fluido}} \quad (21)$$

En donde

$$\rho_{\text{registro}_corregido} = (\rho_b)_{clean} * (1 - Vsh) + (\rho_b)_{shale} * (Vsh)$$

Cálculo del volumen de arcilla (Vsh). Para diferenciar arenas de arcillas se empleó el método de rayos gamma, en el cual se asume una respuesta lineal entre el Vsh y el índice de rayos gamma (IRG) como se muestra en la ecuación (22), las respuestas a los métodos no lineales como Larionov (1969), Steiber (1970), Clavier (1971) son aplicados cuando se conoce la edad de las rocas reservorio.

$$V_{sh} = IGR = \frac{GR_{registro} - GR_{min}}{GR_{max} - GR_{min}} \quad (22)$$

De acuerdo con las observaciones del registro GR en los diferentes pozos, se estableció un valor de corte de 80 °API., si Vsh < 80 °API arenas, si Vsh > °80 API arcillas, esta estimación está basada en las observaciones del registro GR, ya que no se cuenta con información de descripción litológica (ripios) de la perforación. Usando la gráfica de Comparación de los métodos de cálculo de volumen de arcilla Crain (2002) para un valor de 80 °API en el modelo lineal se tiene un valor de Vsh de 0.65.

Cálculo de porosidad efectiva. La porosidad efectiva es el resultado de restar el valor del volumen de arcilla al valor de la porosidad total como se muestra en la ecuación (23), para este caso se asume que el yacimiento tiene una distribución de arcilla dispersa.

$$\emptyset_e = \emptyset - V_{sh} \tag{23}$$

El cálculo de la porosidad efectiva va a depender de la distribución de la arcilla (laminar, dispersa o estructural)

Cálculo de porosidad promedio. La porosidad efectiva promedio (DPHIe_mean), corresponde al valor estadístico de la media de las porosidades efectivas calculadas.

Análisis exploratorio de los datos de la porosidad efectiva. Se realiza el AED de la porosidad efectiva de los 19 pozos identificando tendencias, si la distribución encontrada es normal, se procede directamente a la construcción del diagrama experimental y el modelo que mejor se ajuste, de lo contrario se emplean estimadores o transformaciones que permitan un mejor ajuste con el semivariograma experimental.

5.4.5. Cálculo de permeabilidad por los métodos empíricos de Timur, Tixier y Coates.

La permeabilidad se calculó empleando los métodos empíricos de Timur, Tixier y Coates, considerando que Canadá cuenta principalmente con reservorios no convencionales (Canada, https://www.nrcan.gc.ca/, s.f.) y que estos yacimientos presentan un alto valor de saturación de agua irreducible Martin (1991) se estimó un valor de saturación de agua irreducible Swirr de 50% Moore et al. (2016)

5.4.6. Mapa de distribución de propiedades de porosidad efectiva y permeabilidad por los métodos empíricos de Timur, Tixier y Coates.

Para que los archivos en formato de Excel o csv puedan ser leídos de manera correcta y poder elaborar el variograma, se debe tener instalado el paquete de Rstudio Readxl,y/o read.csv y las librerías readxl, read.csv, gstat y sp.

Generación del variograma experimental y ajuste del modelo. Se debe conocer la ruta en donde está ubicada la tabla con la información de coordenadas y propiedad a mapear, sean las porosidades o las permeabilidades. Para el ajuste del variograma experimental se probaron diferentes modelos en Rstudio.

Creación de la grilla. Se creó el nombre de la grilla que incluye los valores de las coordenadas en el eje X e Y

X=1500 a 8000

Y= 1200:5000

Generación del mapa con krige de la propiedad DPHIe_mean y permeabilidades. De acuerdo con el modelo ajustado con el semivariograma con los valores de sill, rango y nugget, se utilizó el krige para poblar la malla creada anteriormente en los sectores donde no se tiene información.

6. Resultados

El presente capítulo tiene como finalidad mostrar los resultados obtenidos al aplicar una metodología para la estimación areal de propiedades petrofísicas haciendo uso de técnicas geoestadísticas. Los pozos seleccionados fueron 19, los cuales contaban con los tres (3) registros geofísicos: RHOB, NPHI y DT. Al aplicar la prueba Gamma a cada uno de los registros en cada pozo, se encontró que el registro RHOB estadísticamente fue el que mayormente presentó valores de Vratio cercanos a cero (0). Con el registro RHOB se calculó la porosidad DPHI, luego la porosidad efectiva DPHIe y luego mediante los métodos empíricos de Timur, Tixier y Coates la permeabilidad, finalmente se generaron mapas de distribución de porosidad y permeabilidad.

6.1. Pozos seleccionados

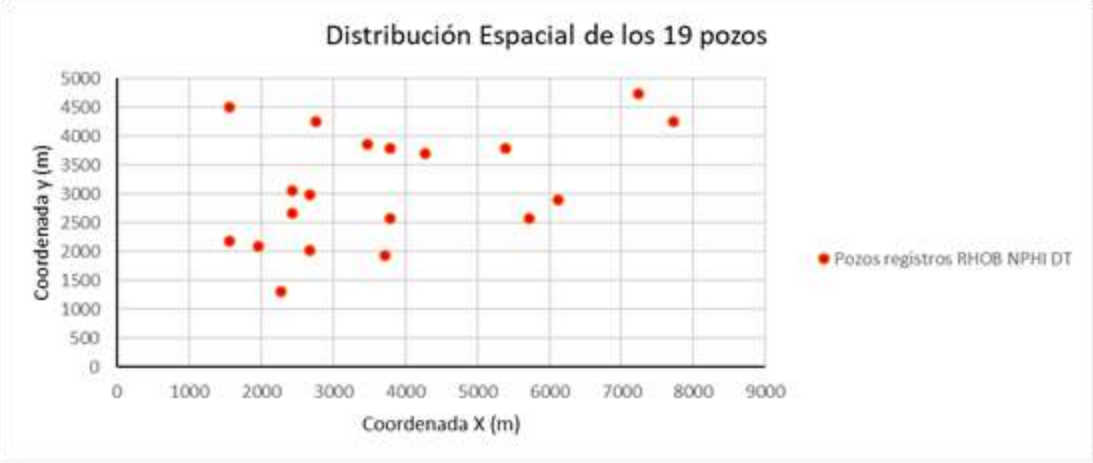
Se muestran las coordenadas de los pozos seleccionados que presentan los tres (3) registros geofísicos de porosidad para el área de estudio (Tabla 3) y el mapa de distribución espacial de los 19 pozos (Figura 4).

Tabla 3
Coordenadas de los pozos con registros RHOB, NPHI y DT

Pozo	Coordenada X(m)	Coordenada Y m)
02_07	2676.74	2022.6
03_08	3716.74	1942.6
05_07	1956.74	2102.6
05_17	3476.74	3862.6
06_15	7716.74	4262.6
06_16	5396.74	3782.6
06_17	3796.74	3782.6
07_17	4276.74	3702.6

08_12	1556.74	2182.6
10_09	5716.74	2582.6
10_18	2756.74	4262.6
11_07	2436.74	2662.6
11_08	3796.74	2582.6
11_15	7236.74	4742.6
14_06	2276.74	1302.6
14_07	2436.74	3062.6
15_07	2676.74	2982.6
16_09	6116.74	2902.6
16_13	1556.74	4502.6

Figura 4. *Mapa de distribución espacial de los 19 pozos seleccionados.*



6.2. Análisis Exploratorio de los Datos de los registros RHOB, NPHI y DT

El Análisis exploratorio para cada pozo y para cada uno de los registros arroja una tabla resumen (Tabla 4), y una representación gráfica del histograma de frecuencia (Figura 5), boxplot (

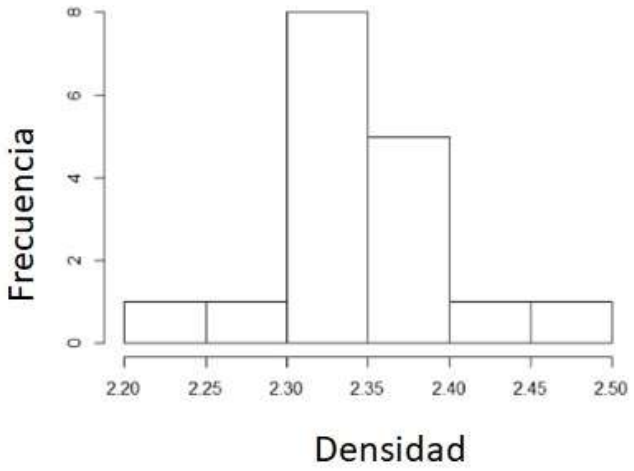
Figura 6) y Q-Q plot (

Figura 7)

Tabla 4
Resumen del AED para el pozo GT05_07 del registro RHOB

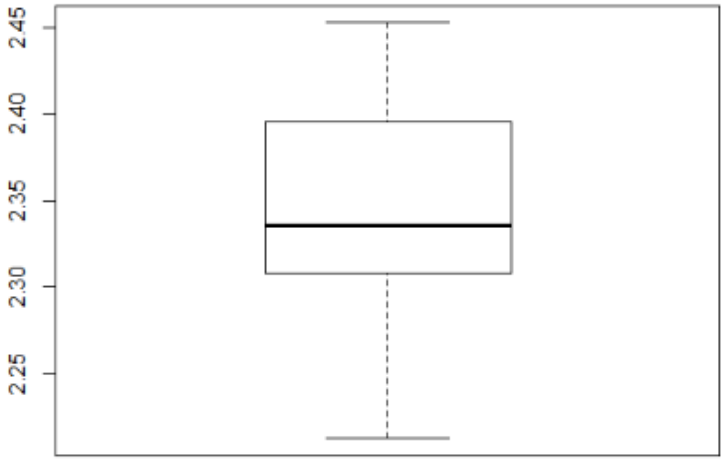
(n)	Missing	Minimum	1st Qu	Mean	Median	TrMean
17	0	2.213	2.307	2.347	2.336	2.347
3rd Qu	Max.	Stdev.	Var.	SE Mean	I.Q.R.	Range
2.396	2.453	0.062	0.004	0.015	0.089	0.24
Kurtosis	Skewness	SW p-val				
-0.612	-0.106	0.552				

Figura 5. *Histograma de frecuencia para el registro RHOB en el pozo GT05_07.*



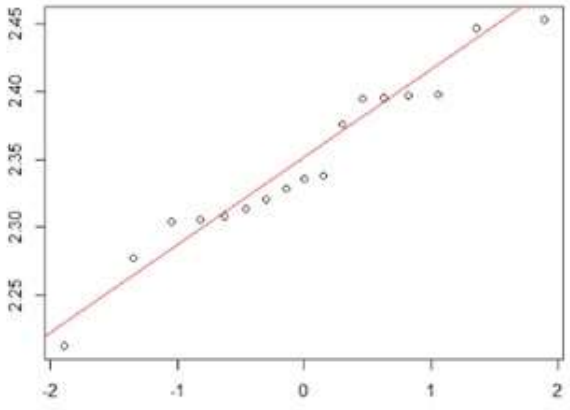
El histograma muestra ser unimodal, simétrico y con mayor frecuencia entre 2.30-2.40 g/cm^3

Figura 6. *Boxplot para el registro RHOB para el pozo GT05_07*



No presenta datos atípicos.

Figura 7. *Q-Q plot para el registro RHOB en el pozo GT05_07*



Q-Q plot indica que presenta una distribución normal.

6.2.1. Resumen del AED para los tres registros RHOB NPHI y DT.

Se muestra el resumen del Análisis exploratorio para los 3 (tres) registros de los 19 pozos, para el registro RHOB (Tabla 5), **¡Error! No se encuentra el origen de la referencia.**para el registro NPHI (Tabla 6) y para el registro DT (Tabla 7).

Tabla 5
Resumen del AED para el registro RHOB

Pozo	N° datos	Histograma	Sesgo	Datos atípicos	Normalidad
02_07	15	Unimodal	Der	Si	No
03_08	20	Unimodal	Der	Si	No
05_07	17	Unimodal	Sime	No	Si
05_17	25	Unimodal	Der	No	No
06_15	8	Bimodal	Izq	No	No
06_16	16	Bimodal	Der	No	No
06_17	12	Bimodal	Izq	No	Si
07_17	12	Unimodal	Der	Si	No
08_12	17	Multimodal	Izq	No	No
10_09	15	Bimodal	Der	No	Si
10_18	18	Bimodal	Izq	No	Si
11_07	15	Unimodal	Der	Si	No
11_08	11	Bimodal	Izq	No	Si
11_15	11	Bimodal	Izq	Si	No
14_06	16	Bimodal	Der	No	No
14_07	16	Unimodal	Sime	No	No
15_07	21	Unimodal	Der	No	No
16_09	20	Bimodal	Sime	No	No
16_13	21	Bimodal	Der	No	No

La distribución de un histograma unimodal, bimodal y multimodal del registro de densidad es la manifestación por un lado de la litología y por otro del tipo de fluido que se encuentra presente.

Si la distribución del histograma es unimodal, este sugiere que existe solo un tipo de litología y un solo tipo de fluido, el cual se encuentra distribuido en todo el espacio poral del yacimiento.

Por otro lado, si el histograma es bimodal, este puede sugerir dos cosas, la primera que se está en un yacimiento con dos litologías diferentes, la segunda que el yacimiento presenta una sola litología, pero dentro de este se encuentran dos fluidos diferentes.

Finalmente, si el histograma es multimodal, se puede estar en un yacimiento el cual presenta intercalaciones de litológicas (arenas-arcilla-arena), o un solo yacimiento con tres fluidos diferentes (agua-aceite-gas).

También es importante observar si el histograma presenta sesgo, si el sesgo es hacia la izquierda, indica altos valores de densidad, por tanto, baja porosidad y el fluido puede ser agua, si por el contrario el sesgo es a la derecha indica bajos valores de densidad, por tanto, alto porosidad y posible presencia de hidrocarburo, si es simétrico se debe prestar una mayor atención a los valores de densidad para inferir el tipo de fluido.

Tabla 6
Resumen del EAD para el registro NPHI

Pozo	N° datos	Histograma	Sesgo	Datos atípicos	Normalidad
02_07	15	Bimodal	Der	Si	Si
03_08	20	Bimodal	Der	No	No
05_07	17	Bimodal	Der	No	No
05_17	25	Unimodal	Izq	No	No
06_15	8	Unimodal	Izq	No	No
06_16	16	Unimodal	Der	No	No
06_17	12	Bimodal	Izq	No	No
07_17	12	Bimodal	Izq	No	No
08_12	17	Bimodal	Der	Si	Si
10_09	15	Unimodal	Izq	No	No
10_18	18	Multimodal	Izq	No	No
11_07	15	Unimodal	Der	Si	Si
11_08	11	Bimodal	Der	No	No
11_15	11	Bimodal	Izq	No	No
14_06	16	Bimodal	Izq	No	No
14_07	16	Unimodal	Sime	No	No
15_07	21	Unimodal	Der	No	No
16_09	20	Bimodal	Izq	No	No
16_13	21	Unimodal	Izq	No	No

Si la distribución del histograma del registro de neutrón es unimodal, sugiere que en el yacimiento solo se encuentra un tipo de fluido, pero si es bimodal o multimodal se sugiere que existe más de un fluido presente en el yacimiento, el sesgo indica también el tipo de fluido si es agua, aceite o gas, en el caso que el sesgo sea hacia la derecha, sugiere presencia de hidrocarburos, por el contrario, si es a la izquierda sugiere que el fluido es agua.

Tabla 7
Resumen del AED para el registro DT

Pozo	N° datos	Histograma	Sesgo	Datos atípicos	Normalidad
02_07	15	Unimodal	Izq	No	No
03_08	20	Unimodal	Sime	Si	Si
05_07	17	Bimodal	Izq	No	No
05_17	25	Unimodal	Der	No	No
06_15	8	Unimodal	Der	Si	Si
06_16	16	Unimodal	Izq	No	No
06_17	12	Bimodal	Izq	Si	Si
07_17	12	Unimodal	Sime	Si	Si
08_12	17	Bimodal	Der	No	No
10_09	15	Bimodal	Izq	No	No
10_18	18	Unimodal	Izq	No	No
11_07	15	Unimodal	Izq	No	No
11_08	11	Bimodal	Sime	No	No
11_15	11	Unimodal	Sime	Si	Si
14_06	16	Bimodal	Sime	No	No
14_07	16	Unimodal	Izq	No	No
15_07	21	Unimodal	Izq	No	No
16_09	20	Bimodal	Sime	No	No
16_13	21	Unimodal	Izq	No	No

Si la distribución del histograma del registro sónico es unimodal, sugiere que en el yacimiento solo se encuentra un tipo de fluido, el cual no produce cambio en el tiempo de tránsito, pero si bimodal o multimodal se sugiere que existe más de un fluido presente en el yacimiento, por lo cual se presentan variaciones en el tiempo de tránsito.

6.3. Aplicación de prueba Gamma a los registros geofísicos RHOB, NPHI y DT

A continuación, se presenta la regresión lineal de la prueba Gamma para el pozo GT11_07 con sus cuatro vecinos más cercanos (Figura 8), y también los resultados obtenidos de la prueba Gamma para cada uno de los pozos, donde se muestra con cuál de las tres combinaciones se obtuvo el valor de V ratio más cercano a cero (Tabla 8).

Figura 8. *Grafica de regresión lineal YM (k) vs δ_M (k) para el pozo GT11_07.*

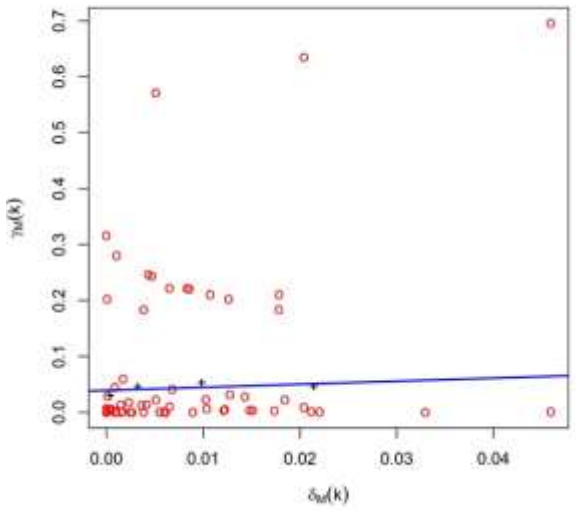


Tabla 8
Resumen de la prueba Gamma aplicada a los 19 pozos

Pozo	Depth	RHOB	NPHI	DT	gamma.val	varianza.val	Vratio.val
02_07	1			1	0.0452	0.0865	0.5224
03_08	1		1		0.0132	0.0855	0.1549
05_07	1	1			0.0347	0.0672	0.5167

05_17	1		1		0.0055	0.0695	0.0784
06_15	1			1	0.0017	0.0957	0.0180
06_16	1	1			0.0672	0.0967	0.6945
06_17	1		1		0.0038	0.1206	0.0314
07_17	1		1		0.0514	0.1140	0.4507
08_12	1	1			0.0953	0.1370	0.6958
10_09	1	1			0.0715	0.0822	0.8697
10_18	1		1		0.0813	0.1360	0.5979
11_07	1	1			0.0393	0.0660	0.5944
11_08	1	1			0.0417	0.1198	0.3483
11_15	1		1		0.0033	0.1648	0.0199
14_06	1	1			0.0347	0.1288	0.2696
14_07	1	1			0.0612	0.1445	0.4235
15_07	1			1	0.0570	0.1081	0.5277
16_09	1	1			0.0224	0.1419	0.1577
16_13	1			1	0.0236	0.0712	0.3321

De acuerdo con los resultados de la prueba Gamma, se encuentra que, de los 19 pozos, nueve (9) combinaciones presentan el valor de Vratio más cercano a cero (0) con el registro de densidad (RHOB), seis (6) con el registro neutrón (NPHI) y cuatro (4) con el de sónico (DT), esto indica que de los tres registros el que presenta menor varianza en sus datos es el registro de densidad.

Con base en estos resultados el registro que se seleccionó para la fase 3 es el registro de densidad, con este registro se realizarán los cálculos de porosidad.

6.4. Cálculo de porosidad, volumen de arcilla, porosidad efectiva y porosidad promedio

En cada pozo fue calculado el valor de la porosidad DPHI a partir del registro RHOB teniendo en cuenta las correcciones ambientales (diámetro de pozo, densidad de matriz y

densidad de fluido), el volumen de arcilla y la porosidad efectiva, se muestran los resultados de los cálculos realizados para el pozo GT02_07 (Tabla 9), y el resumen de los valores de porosidad efectiva en cada uno de los pozos (Tabla 10).

Tabla 9
Resumen valores calculados de porosidad, volumen arcilla y porosidad efectiva para el pozo GT02_07

Depth	RHOB	NPHI	DT	GR	ILD	DPHI	Vsh	DPHIe
695.4	2.303	29.307	310.998	51.772	31.566	0.224	0.436	0.126
695.6	2.263	28.568	319.743	34.496	35.107	0.248	0.150	0.211
695.8	2.265	27.094	325.481	29.251	38.189	0.247	0.063	0.231
696	2.288	26.606	325.091	25.486	41.444	0.233	0.000	0.233
696.2	2.360	26.424	320.904	27.455	44.589	0.189	0.033	0.183
696.4	2.372	27.185	317.750	30.489	41.668	0.182	0.083	0.166
696.6	2.335	28.723	317.700	33.018	36.335	0.204	0.125	0.179
696.8	2.305	28.949	320.184	33.576	31.394	0.223	0.134	0.193
697	2.280	28.291	320.110	32.808	27.114	0.238	0.122	0.209
697.2	2.295	27.533	316.765	31.456	23.704	0.229	0.099	0.206
697.4	2.359	27.212	308.601	43.633	20.367	0.190	0.301	0.133
697.6	2.509	26.805	298.340	58.944	17.126	0.098	0.556	0.044

Tabla 10
Resumen de los valores de porosidad efectiva calculada para los 19 pozos

X	Y	Pozo	DPHIe_mean
2676.74	2022.6	02_07	0.181
3716.74	1942.6	03_08	0.110
1956.74	2102.6	05_07	0.168
3476.74	3862.6	05_17	0.049
7716.74	4262.6	06_15	0.074
5396.74	3782.6	06_16	0.194
3796.74	3782.6	06_17	0.081
4276.74	3702.6	07_17	0.187

1556.74	2182.6	08_12	0.177
5716.74	2582.6	10_09	0.151
2756.74	4262.6	10_18	0.066
2436.74	2662.6	11_07	0.225
3796.74	2582.6	11_08	0.127
7236.74	4742.6	11_15	0.036
2276.74	1302.6	14_06	0.118
2436.74	3062.6	14_07	0.134
2676.74	2982.6	15_07	0.209
6116.74	2902.6	16_09	0.177
1556.74	4502.6	16_13	0.027

Hasta este punto se ha logrado estimar el valor de la porosidad efectiva promedio del intervalo de interés seleccionado en cada uno de los pozos, este dato será el nuevo insumo para los siguientes cálculos como el cálculo de permeabilidad.

6.5. Resumen del AED para la porosidad efectiva de los 19 pozos

Se muestra el resumen del AED de la porosidad efectiva para los 19 pozos (Tabla 11), el AED indica que el histograma es unimodal, presenta un sesgo a la derecha, no presentan datos atípicos ni tampoco una distribución normal, y se presentan las gráficas del histograma de frecuencias (Figura 9), box plot (Figura 10) y Q-Q plot (Figura 11).

Tabla 11
Resumen del AED para la porosidad efectiva promedio de los 19 pozos

N° datos	**Histograma**	**Sesgo**	**Datos Atípicos**	**Normalidad**
19	Unimodal	Der	No	No

Figura 9. *Histograma para la porosidad efectiva de los 19 pozos.*

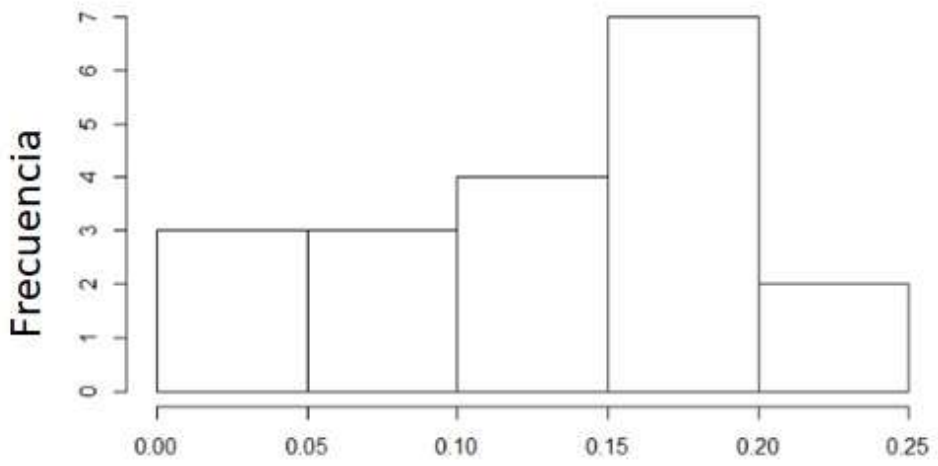


Figura 10. *Box plot para porosidad efectiva de los 19 pozos.*

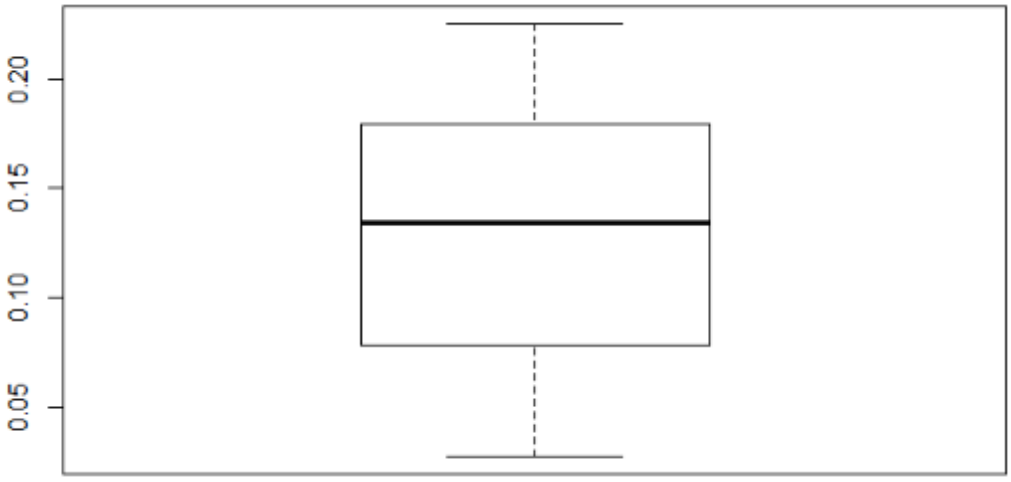
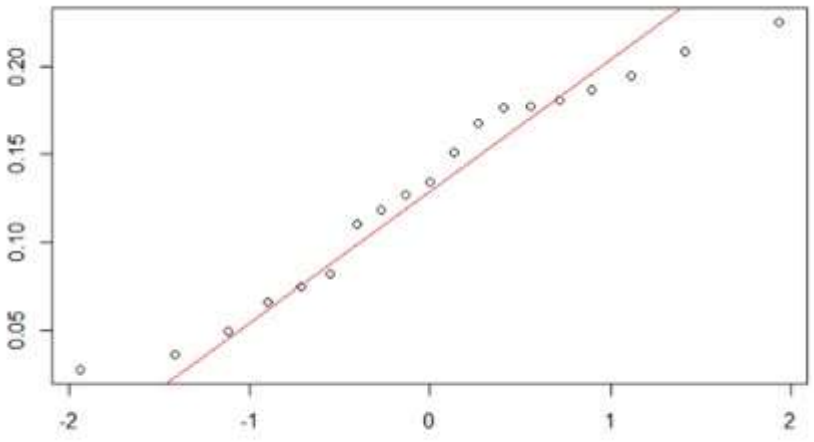


Figura 11. *Q-Q plot para porosidad efectiva de los 19 pozos.*



El AED muestra que los valores de porosidad efectiva promedio se encuentran en un rango entre 0.027 y 0.225, la mayor frecuencia es de siete para valores entre 0.15 y 0.20 y la menor es de dos para valores entre 0.20 y 0.25, el diagrama de box plot no muestra que existan datos atípicos, y la gráfica Q-Q plot sugiere que los datos no presentan una distribución normal. Como los datos no presentan una distribución normal, se emplearon técnicas geoestadísticas que usan estimadores para encontrar un mejor modelo.

6.6. Cálculo de permeabilidad por los métodos empíricos de Timur, Tixier y Coates

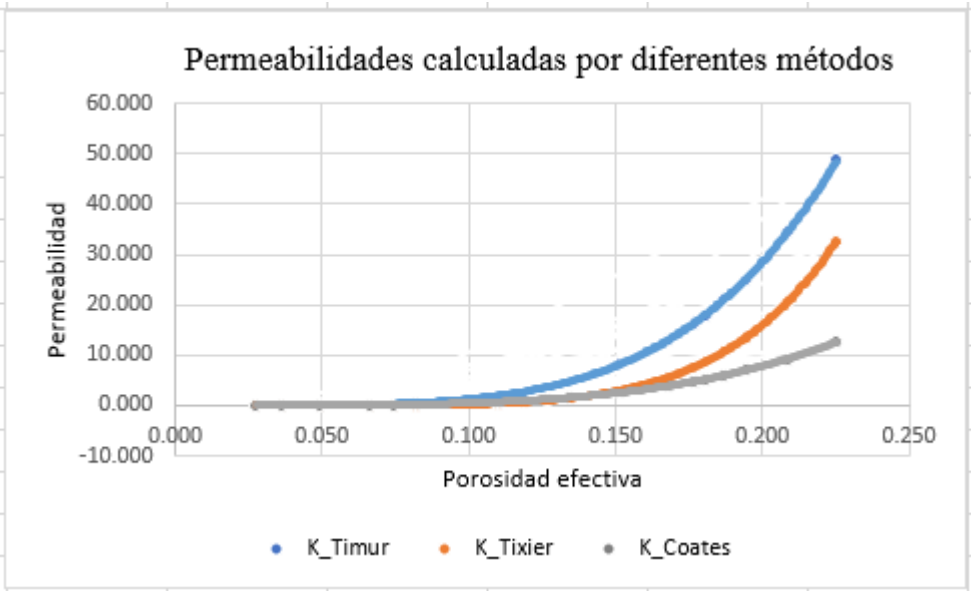
Se presenta el resumen de las permeabilidades calculadas por los métodos de Timur, Tixier y Coates (Tabla 12) a partir de la porosidad efectiva promedio del intervalo, al graficar el comportamiento de cada uno de los métodos en una sola gráfica, se observa que el método de Timur es el más optimista en el sentido que presenta los valores de permeabilidad más altos,

también se observa que para valores de porosidades aproximadamente menores a 0.12 en los tres métodos arrojan resultados similares (Figura 12).

Tabla 12
Resumen permeabilidades calculados por métodos de Timur, Tixier y Coates

Pozo	DPHIe_mean	K_Timur	K_Tixier	K_Coates
02_07	0.181	18.048	8.651	5.203
03_08	0.110	1.944	0.443	0.718
05_07	0.168	12.947	5.556	3.873
05_17	0.049	0.052	0.004	0.028
06_15	0.074	0.335	0.042	0.15
06_16	0.194	25.178	13.486	6.996
06_17	0.081	0.504	0.073	0.216
07_17	0.187	20.998	10.586	5.953
08_12	0.177	16.722	7.814	4.862
10_09	0.151	8.128	2.986	2.561
10_18	0.066	0.194	0.021	0.093
11_07	0.225	48.729	32.527	12.582
11_08	0.127	3.718	1.052	1.278
11_15	0.036	0.013	0.001	0.008
14_06	0.118	2.710	0.691	0.965
14_07	0.134	4.739	1.455	1.585
15_07	0.209	34.530	20.549	9.263
16_09	0.177	16.346	7.581	4.765
16_13	0.027	0.004	0	0.003

Figura 12. *Representación del cálculo de la permeabilidad por diferentes métodos.*



Se observa que a valores de porosidad cercanos a 0.12 los valores de permeabilidad son similares, y a medida que el valor de porosidad aumenta la diferencia en el valor permeabilidad presenta mayor distanciamiento.

6.7. Mapas de distribución de propiedades de porosidad efectiva y permeabilidad

Se presenta el resumen de los valores de sill, rango y negget para los semivariogramas generados en los tres niveles, todos fueron ajustados a un modelo esférico, el primero de estos (nivel medio) es autoajustado por el software Rstudio, los dos siguientes (nivel superior e inferior) son ajustados manualmente cambiando los valores de los parámetros de sill, rango y nugget procurando que pase por la mayor cantidad de puntos de la parte experimental (Tabla 13).

Tabla 13
Resumen de los valores de sill, rango y nugget para los semivariogramas generados para la porosidad efectiva

Nivel	Modelo	sill	rango	nugget
Parte media	Spherical	0.00220	2001.53200	0.00107
Parte superior	Spherical	0.00290	1200.00000	0.00150
Parte inferior	Spherical	0.00250	1800.00000	0.00005

6.7.1. Semivariograma nivel medio.

Se encontró que el semivariograma experimental se ajusta un modelo esférico (Figura 13), con el cual se estimó el mapa de distribución de la porosidad (Figura 14) y el mapa distribución de la varianza (Figura 15).

Figura 13. *Semivariograma ajustado automáticamente (nivel medio).*

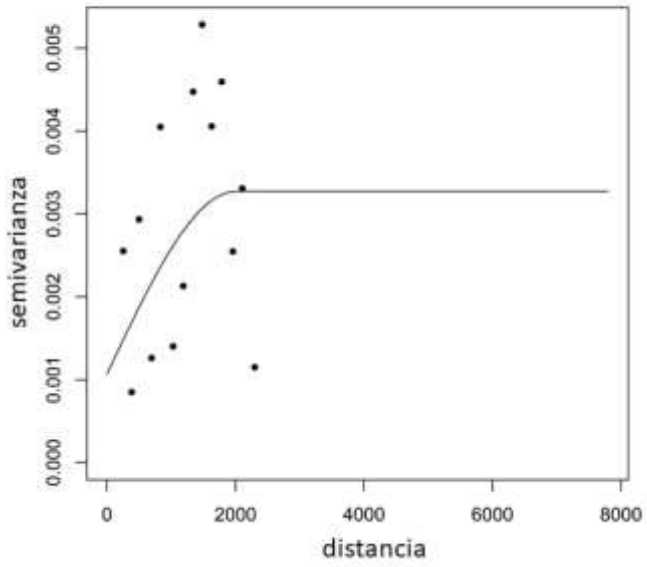


Figura 14. *Mapa distribución de la porosidad efectiva (nivel medio).*

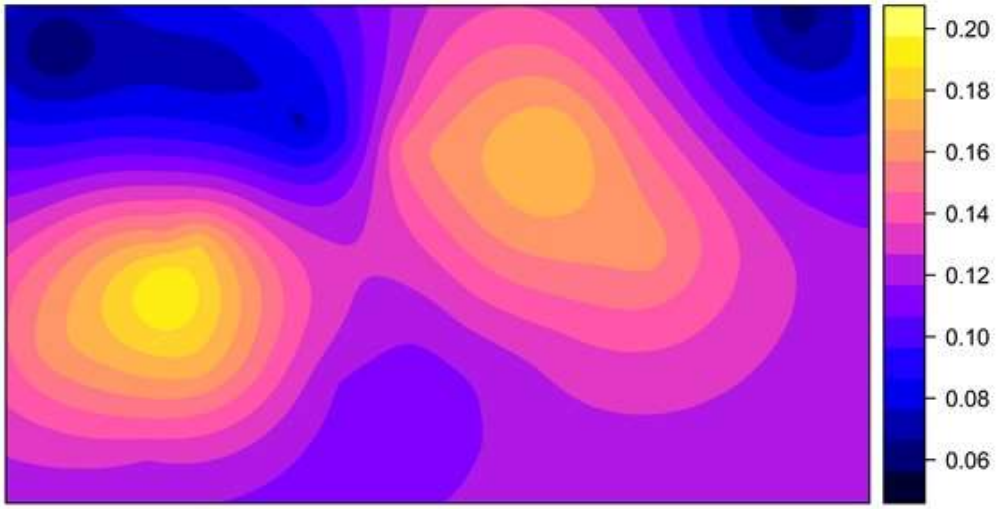
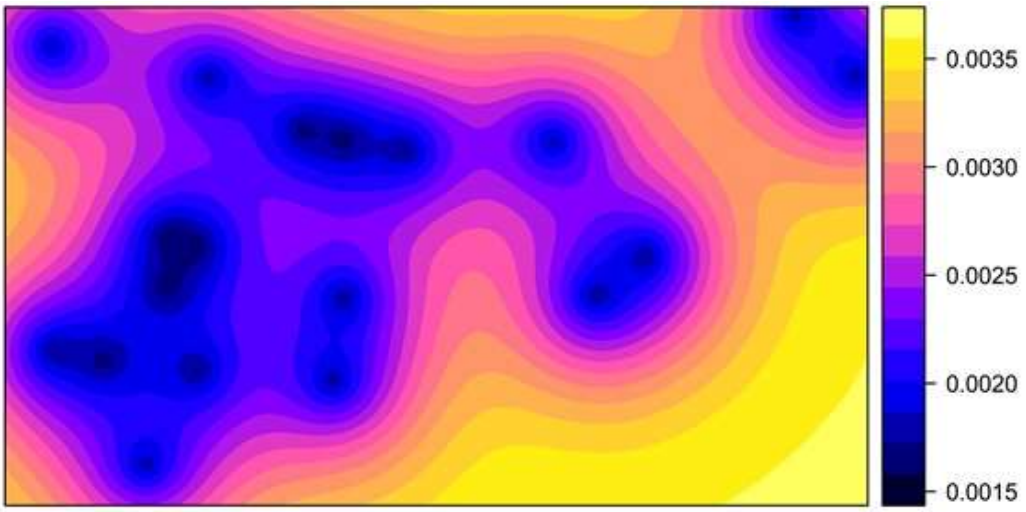


Figura 15. *Mapa varianza de la porosidad efectiva (nivel medio).*



El mapa de distribución de la porosidad efectiva muestra una tendencia preferencial en sentido suroeste-noreste, luego presenta un giro en sentido sureste (este giro puede ser interpretado como meandros). Los valores de la varianza se encuentran entre 0.0015 y 0.0035, como era de esperarse el mapa muestra los valores más bajos en los sectores en donde se cuenta con información.

Se propusieron otros dos modelos, uno por encima (nivel superior) y otro por debajo (nivel inferior) del generado automáticamente.

6.7.2. Semivariograma nivel superior.

Se encontró que el semivariograma experimental se ajusta un modelo esférico (Figura 16), con el cual se estimó el mapa de distribución de la porosidad (Figura 17) y el mapa distribución de la varianza (

Figura 18).

Figura 16. *Semivariograma ajustando a los datos experimentales (nivel superior).*

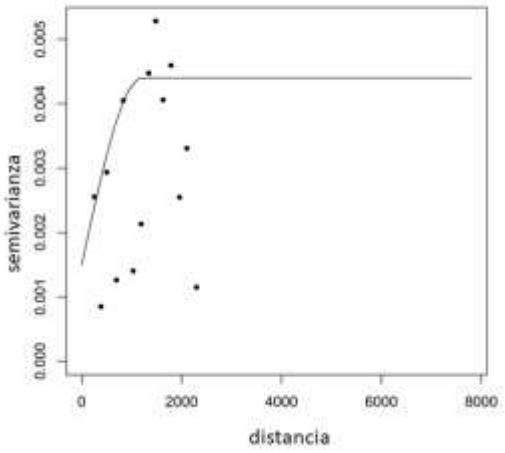


Figura 17. *Mapa distribución de la porosidad efectiva (nivel superior).*

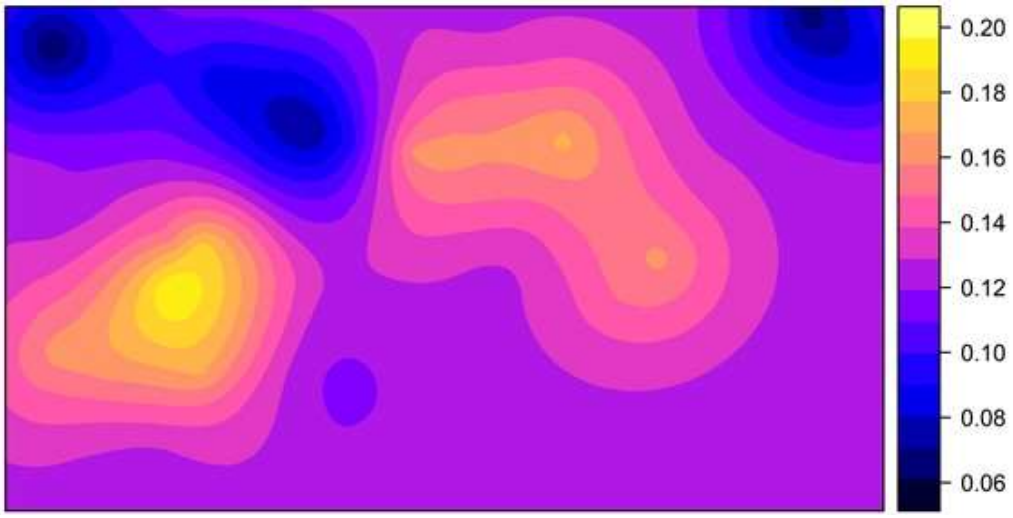
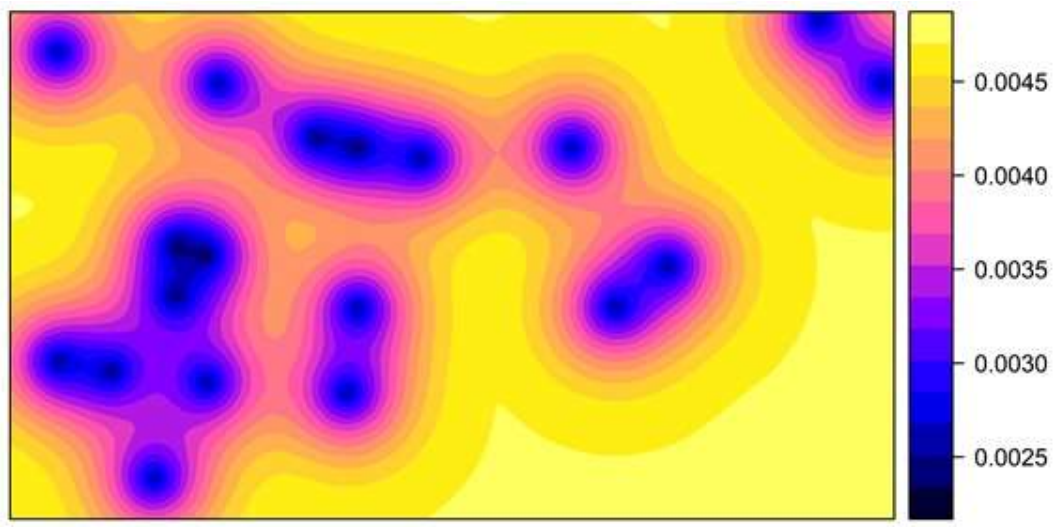


Figura 18. *Mapa varianza de la porosidad efectiva (nivel superior).*



Al ajustar el semivariograma a un modelo con datos experimentales del nivel superior, el mapa de distribución de porosidad muestra una mayor extensión con valores de mayor porosidad con respecto al mapa autogenerado del nivel medio, es decir, se puede sobre estimar sectores con mayor porosidad.

6.7.3. Semivariograma nivel inferior.

Se encontró que el semivariograma experimental se ajusta un modelo esférico (Figura 19), con el cual se estimó el mapa de distribución de la porosidad (

Figura 20) y el mapa distribución de la varianza (Figura 21).

Figura 19. *Semivariograma ajustando a los datos experimentales (nivel inferior).*

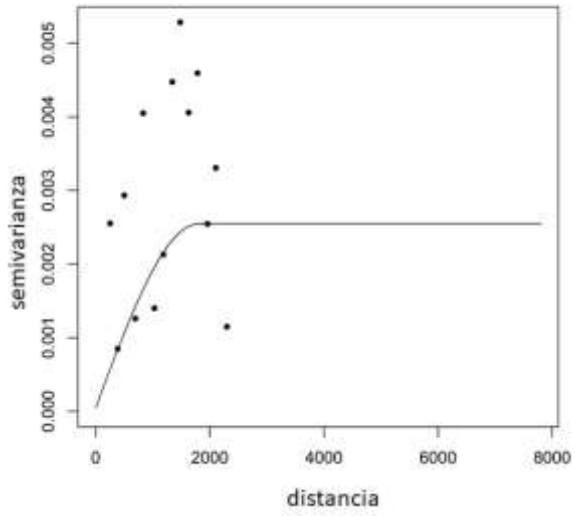


Figura 20. *Mapa distribución de la porosidad efectiva (nivel inferior).*

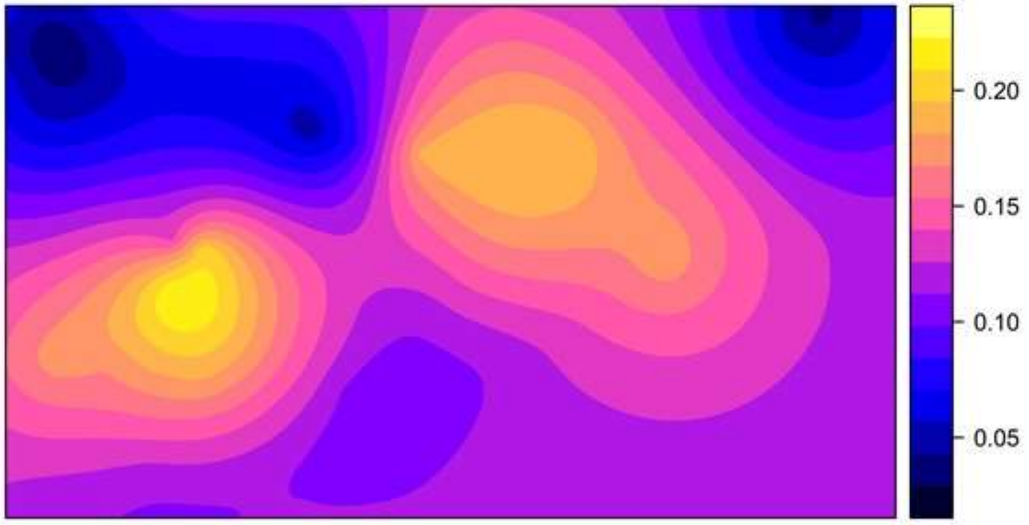
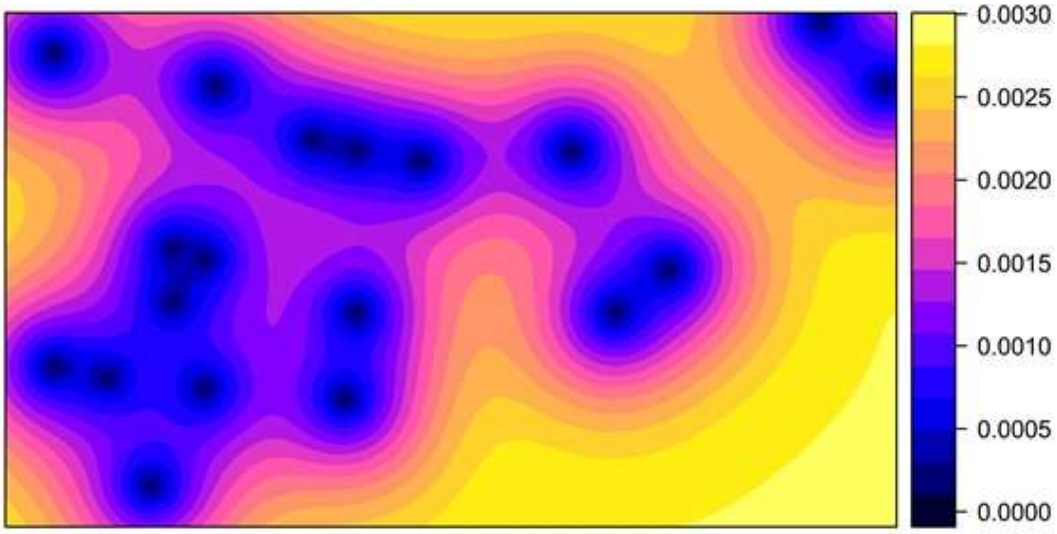


Figura 21. *Mapa varianza de la porosidad efectiva (nivel inferior).*



6.8. Mapas estimación de la distribución de permeabilidad

Se muestran los mapas de distribución de permeabilidad calculados por los métodos de Timur (Figura 22), Tixier (Figura 23) y Coates (Figura 24).

Figura 22. *Mapa de estimación de permeabilidad por el método de Timur.*

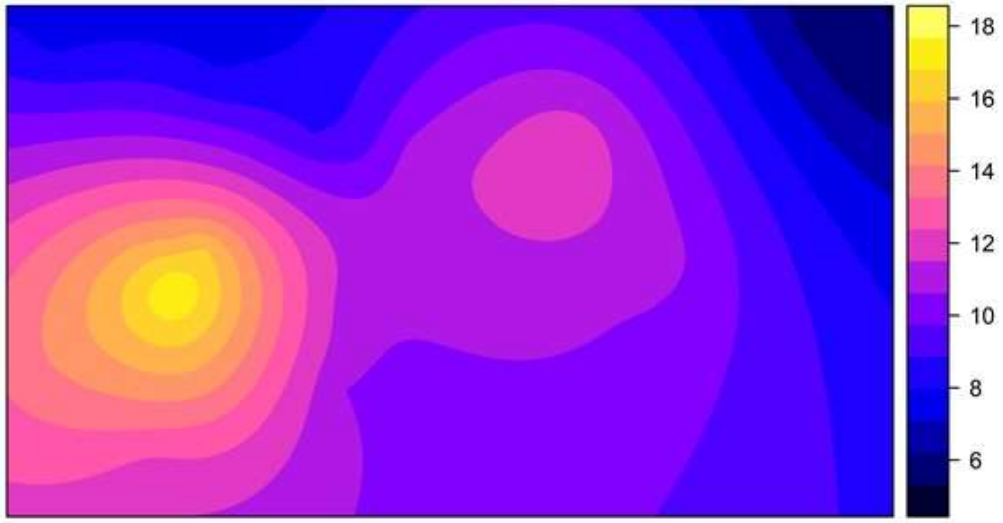


Figura 23. *Mapa de estimación de permeabilidad por el método de Tixier.*

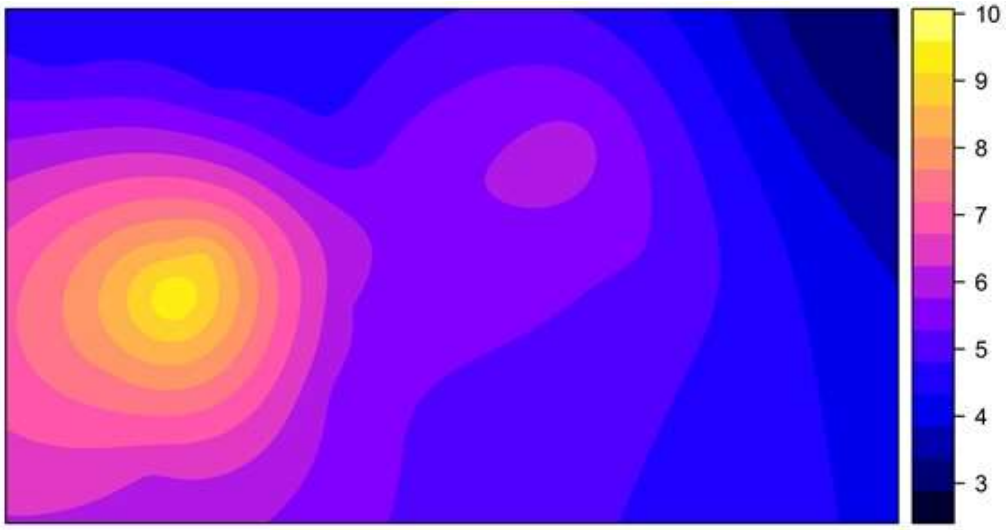
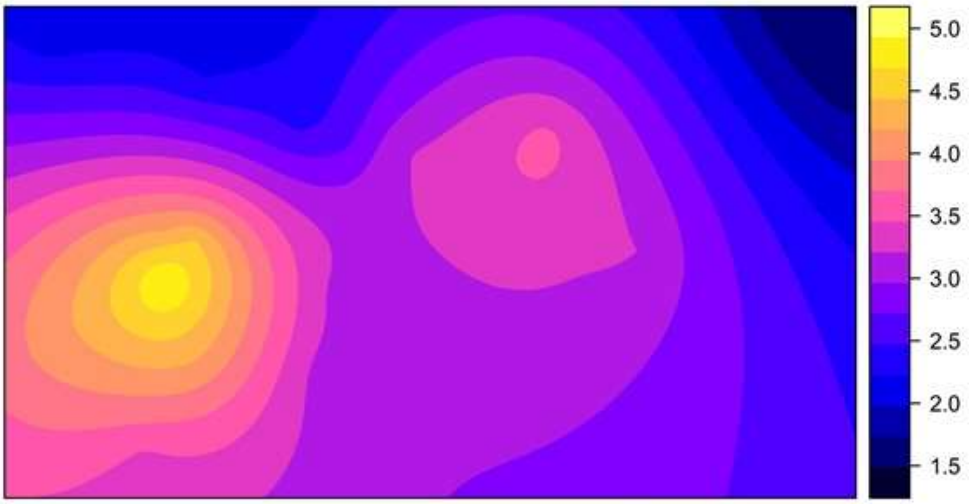


Figura 24. *Mapa de estimación de permeabilidad por el método de Coates.*



Los mapas de distribución de permeabilidades muestran que el método de Tixier estima los valores más altos entre los tres métodos, sin embargo, los tres mantienen una misma tendencia en cuanto a la distribución de esta propiedad.

6.9. Ajuste del modelo de porosidad usando los estimadores los OLS, WLS, ML y RML

Para seleccionar cuál de los tres semivariogramas (nivel medio, superior o inferior) presenta un mejor ajuste a los datos experimentales, se aplicaron métodos de estimación de parámetros y la técnica de validación cruzada, los estimadores usados fueron OLS (mínimos cuadrados ordinarios), WLS (mínimos cuadrados ponderados), ML (máxima verosimilitud) y RML (máxima verosimilitud restringida). Se muestran los resultados de los estimadores y el valor del error de cada uno (Tabla 14).

Tabla 14
Resumen del valor del error aplicando diferentes estimadores en cada nivel

Nivel	Modelo	vEstimador-peso	vError
Inferior	Spherical	ML	0.0523
Inferior	Spherical	REML	0.0940
Inferior	Spherical	WLS-npairs	0.0975
Inferior	Spherical	WLS-Cressie	0.2941
Inferior	Spherical	OLS	0.3042
Medio	Spherical	ML	0.0487
Medio	Spherical	WLS-npairs	0.0659
Medio	Spherical	REML	0.0834
Medio	Spherical	OLS	0.3128
Medio	Spherical	WLS-Cressie	0.4119
Superior	Spherical	WLS-npairs	0.0285
Superior	Spherical	ML	0.0523
Superior	Spherical	REML	0.0940
Superior	Spherical	OLS	0.2343
Superior	Spherical	WLS-Cressie	0.2967

Se muestra el ajuste de los tres semivariogramas con los cinco estimadores usados OLS, WLS, ML y REML para cada uno de los niveles, y el ajuste del semivariograma para cada uno de los niveles, nivel medio (Figura 25), nivel superior (**¡Error! No se encuentra el origen de la referencia.**) y nivel inferior (Figura 27).

Figura 25. *Ajuste del semivariograma nivel medio usando los estimadores.*

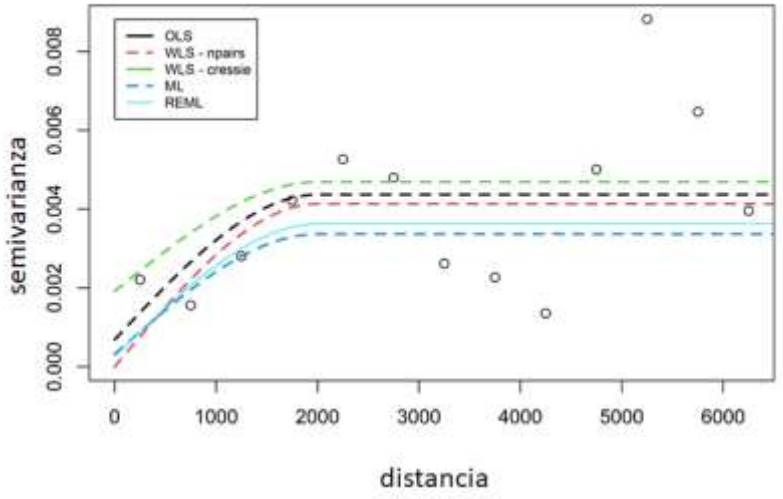


Figura 26. *Ajuste del semivariograma nivel superior usando los estimadores*

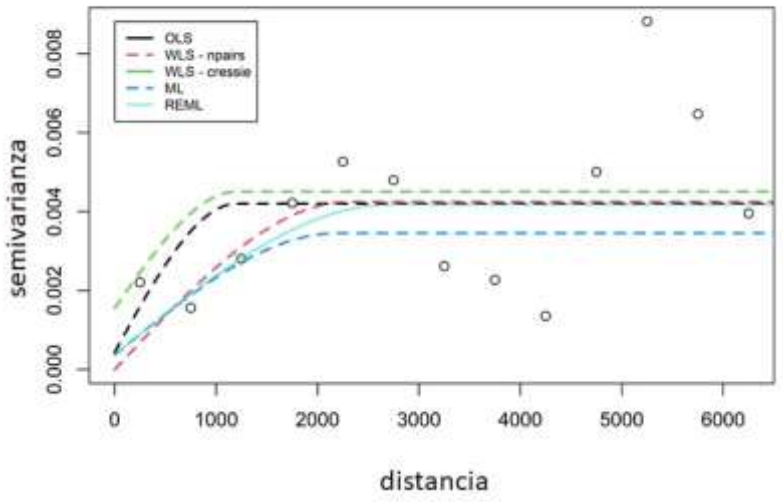
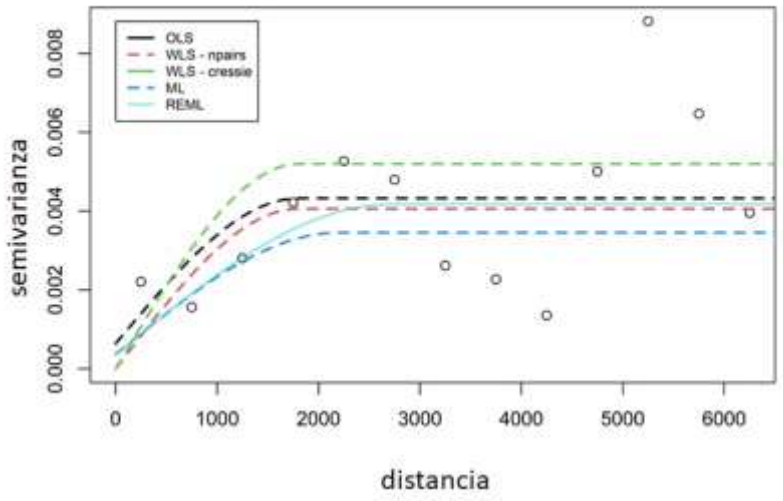


Figura 27. Ajuste del semivariograma nivel inferior usando los estimadores.



De acuerdo con los tres ajustes usando los estimadores propuestos, se encontró que el estimador con el menor error es para el semivariograma propuesto del nivel superior con el estimador WLS-npairs, en donde el valor de error de 0.0285, de esta manera se pudo encontrar los valores sill, rango y nugget (Tabla 15).

Tabla 15
Resumen de los valores usando el estimador WLS-npairs

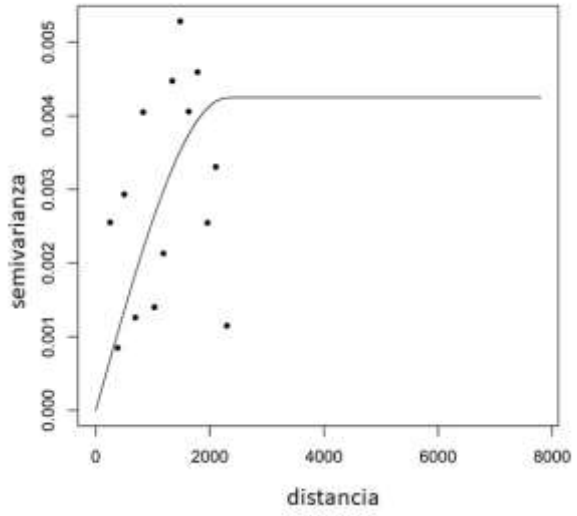
Nivel	Modelo	sill	rango	nugget
Parte superior	Spherical	0.00425	2310.114	0

Con los valores encontrados de ajuste se procedió a construir un modelo para el semivariograma experimental usando los valores de sill, rango y nugget del estimador WLS-npairs, y de esta manera poder generar un nuevo mapa de estimación para la distribución de porosidad.

6.10. Mapa de distribución de porosidad efectiva usando el semivariograma ajustado

Como se mencionó anteriormente, el semivariograma ajustado con los parámetros obtenidos con el estimador WLS-apairs (Figura 28) correspondiente al nivel superior es el que presenta el menor error de los tres propuestos.

Figura 28. *Semivariograma nivel superior construido con el estimador WLS-spairs.*



Usando los valores de sill, rango y nugget encontrados para el nivel superior con el estimador WLS-apairs, se construyó un nuevo modelo de distribución de porosidad (Figura 29) y uno de varianza (Figura 30).

Figura 29. *Mapa distribución de la porosidad efectiva (nivel superior) ajustado con el estimador.*

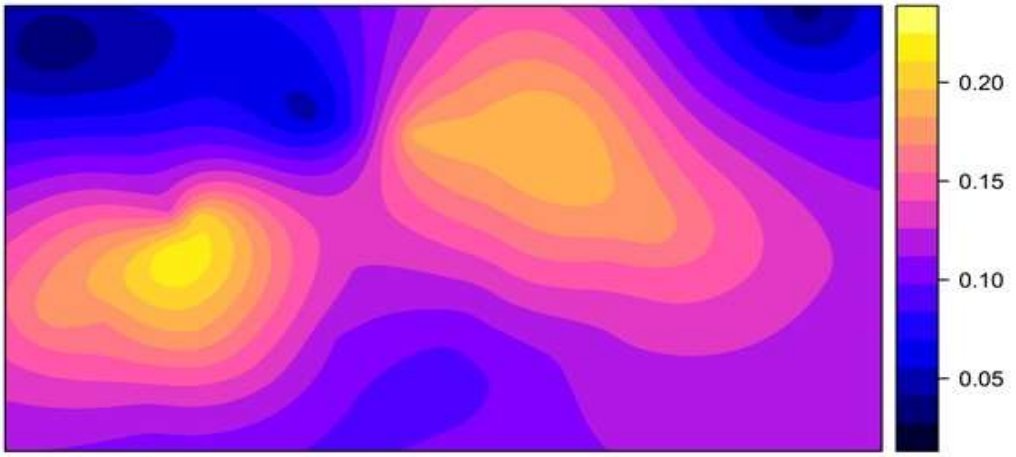
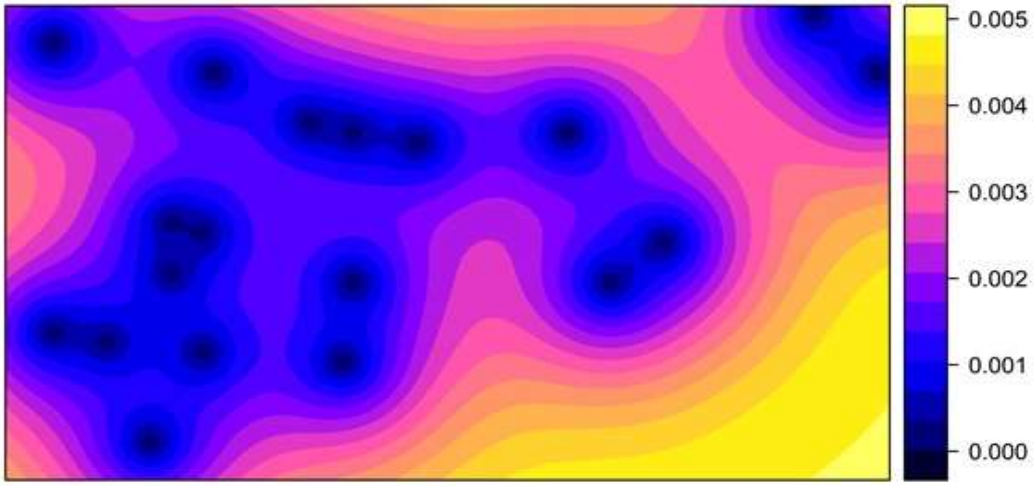


Figura 30. *Mapa varianza de la porosidad efectiva (nivel inferior) ajustado con el estimador.*



La metodología permitió estimar la distribución areal de las propiedades petrofísicas empleando técnicas geoestadísticas.

7. Análisis de resultados

De acuerdo con el registro GR el intervalo de interés corresponde a un paquete arenoso en el cual se observa un aumento en el valor del registro de resistividad asociado a la presencia de fluidos, hacia el sector noroeste en los pozos 05_17, 06_17, 10_18 y 16_13 y sector noreste en los pozos 06_15 y 11_15 se evidencia un cambio lateral de facies donde predominan principalmente litologías arcillosas lo cual no lo hacen interesante como yacimiento. Es decir, 13 de los 19 pozos presentan un intervalo arenoso que podría almacenar hidrocarburos.

Con el análisis del histograma y sesgo del registro RHOB se puede tener una primera aproximación de la calidad del yacimiento en cuanto a su litología y fluidos que se encuentran presentes, de igual manera el histograma y sesgo del análisis exploratorio de los datos para el registro de NPHI muestra una aproximación al número de fluidos presentes en el yacimiento e infiere si es un hidrocarburo o agua, lo dicho anteriormente puede ser correlacionable con el registro DT, el cual muestra variaciones en el tiempo de transito dependiendo del número de fluidos presentes en el yacimiento.

La prueba Gamma mostró un valor de Vratio cercano a cero para el registro RHOB en 9 de los 18 pozos seleccionados, se debe notar que estos 9 pozos se encuentran dentro del grupo de los 13 que presentan el intervalo arenoso y que pueden almacenar hidrocarburos, lo cual indica

que puede existir una relación directa con la litología, sin embargo, de los 4 pozos restantes con características para almacenar hidrocarburos, 2 de ellos 02_07 y 15_07 muestran valor de Vratio cercano a cero con el registro DT y los pozos 03_08 y 07_17 con el registro NPHI.

Se podría tener un mejor entendimiento del área de estudio al contar con información adicional como la descripción de muestras durante la perforación de cada uno de los pozos, adquisición y análisis de núcleos, información de petrografía, registro gráficos compuesto, registros de medición de gases durante la perforación, entre otros. De esta manera se podría ajustar el modelo petrofísico del área.

La ecuación para el cálculo de porosidad involucra diferentes variables como el diámetro de hueco, la densidad de matriz del yacimiento y densidad de matriz del fluido utilizado durante la perforación, de no conocer con algún grado de precisión los valores adecuados se incurre en estimaciones equivocadas en el valor de porosidad.

Los valores más altos de porosidad efectiva promedio DPHIe_mean se encuentran en una dirección preferencial en el eje suroeste-noreste en la parte central del campo, variando entre 0.027 y 0.225

El valor de corte de Vshale para diferenciar arenas y arcillas puede llegar a ser algo subjetivo si no existe un buen amarre entre los registros litológicos (GR y SP) y la descripción de ripios de la perforación.

La porosidad efectiva es afectada por el volumen de arcilla y su distribución en el yacimiento, es importante tener información de petrografía para realizar los ajustes a los modelos el petrofísicos.

Es primordial tener en cuenta un buen ajuste entre el semivariograma experimental y el modelo matemático para evitar subestimar o sobrestimar los valores en el yacimiento, para este caso de estudio se logró proponer diferentes modelos.

Entre los métodos empíricos para calcular la permeabilidad, se encontró que el método de Timur es el más optimista mostrando los valores de permeabilidad más altos, caso contrario por el método de Coates que muestra los valores más bajos, los tres modelos no muestran diferencias significativas en el valor de permeabilidad cuando los valores de porosidad son aproximadamente menores a 0.12.

8. 7 . Conclusiones

La metodología mostró que con el análisis exploratorio de los datos (AED) se logra identificar de manera rápida y sencilla características y comportamientos de los registros geofísicos, e inferir litologías y fluidos presentes en el yacimiento.

La prueba Gamma para este caso de estudio mostró que el mejor registro geofísico para realizar los cálculos de porosidad es el registro de densidad (RHOB) en litologías arenosas, este registro presentó el valor más bajo de Vratio en 9 de los 19 pozos seleccionados.

Mediante el uso de estimadores y correlación cruzada fue posible cuantificar el valor del error en los tres modelos propuestos para la porosidad, seleccionando el modelo con el menor error.

Una desventaja de los tres métodos para el cálculo de permeabilidad es que se debe tener el valor de la saturación de agua irreducible, lo cual obligaría a tener información de núcleo y análisis de laboratorio, sin embargo, estos cálculos se pueden realizar con datos de campos cercanos.

El software libre y de código abierto Rstudio mostró tener todas las características para desarrollar la metodología desde el análisis exploratorio hasta la generación de mapas.

8. Recomendaciones

Para obtener un mejor amarre entre los registros y la litología es importante tener la descripción de las muestras de zanja de cada uno de los pozos, de igual manera conocer las características físicas de cada uno de los fluidos usados durante la perforación, y en lo posible data obtenida en laboratorio de los núcleos, ya que con esta se puede verificar y ajustar los modelos petrofísicos generados, finalmente revisar en la literatura las opciones que existen para modelamiento geoespacial, como uso de estimadores propios de cada tipo de modelo, y transformaciones como Boxcox, normal score entre otras.

Referencias Bibliográficas

Calvete, F., & Ramirez, J. (1990). *Geoestadística: Aplicaciones a la hidrogeología subterránea.* Centro Internacional de Métodos Numéricos en Ingeniería.

canada, G. o. (s.f.). *https://www.nrcan.gc.ca/*. Obtenido de https://www.nrcan.gc.ca/

canada, G. o. (s.f.). *https://www.nrcan.gc.ca/our-natural-resources/energy-sources-distribution/clean-fossil-fuels/natural-gas/shale-tight-resources-canada/geology-shale-and-tight-resources/17675*. Obtenido de https://www.nrcan.gc.ca

canada, G. o. (s.f.). *https://www.nrcan.gc.ca/science-data/data-analysis/energy-data-analysis/energy-facts/crude-oil-facts/20064*. Obtenido de https://www.nrcan.gc.ca/

Clarkson , C., & Pedersen , P. (2011). Production Analysis of Western Canadian Unconventional Light Oil Plays. *OnePetro, Canadian Unconventional Resources Conference*.

Coates, G., & Denoo, S. (1981). The producibility answer product. *The Technical Review*, 29(2), 55-63.

Coates, G., & Dumanoir, J. (January de 1973). A new approach to improved log-derived permeability. *In SPWLA 14th Annual Logging Symposium*.

Crain, E. (2002). *Crain's petrophysical handbook.* Spectrum 2000 Mindware Limited.

Devore, J. (2008). *Probabilidad y estadística para ingenierías y ciencias.* Learning Editores.

Evans, D., & Jones, A. J. (2002). A proof of the gamma test. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 2759-2799.

Evans, D., Jones, A. J., & Schmidt, W. M. (2002). Asymptotic moments of near–neighbour distance distributions. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 2839-2849.

Everitt, B. S., & Dunn, G. (2001). *Applied Multivariate Data Analysis.* London: Arnold.

Garcia, C. C., & Guevara, J. B. (2004). *Territorio y medio ambiente: métodos cuantitativos y técnicas de información geográfica.* Murcia: EDITUM.

Giraldo, R. (2002). *Introducción a la geoestadística: Teoría y aplicación.* Bogotá: Universidad Nacional de Colombia.

Guamialamá, C., Freire, J., & Valencia , R. (Mayo-Julio de 2020). mpacto de los Métodos Aplicados para la Determinación de la. *Revista Politécnica, Vol. 45, No. 02*, 41-50.

Isaaks, E., & Srivastaca, R. (1989). *Applied Geostatistics.* Oxford University Press.

Iturrarán-Viveros, U., & Parra, J. O. (2014). Artificial neural networks applied to estimate permeability, porosity and intrinsic attenuation using seismic attributes and well-log data. *Journal of Applied Geophysics*, 107, 45-54.

Journel, A. (1990). Geostatistics for Reservoir Characterization. *Society of Petroleum Engineers*.

Journel, A., & Huijbregts, C. (1978). Mining geostatistics (Vol. 600). London.

Kemp, S. E., Wilson, I. D., & Ware, J. A. (2004). A tutorial on the gamma test. *International Journal of Simulation: Systems, Science and Technology*, 6(1-2), 67-75.

Koncar, N. (1997). *Optimisation methodologies for direct inverse neurocontrol* . Doctoral dissertation, University of London.

Kozeny, J. (1927). Uber kapillare leitung der wasser in boden. *Royal Academy of Science, Vienna*, 136, 271-306.

Krajewski, L., & Ritzman, L. (2000). *Administración de operaciones: estrategia y análisis.* Pearson education.

Krige, D. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, 52(6), 119-139.

Lantz, B. (2013). *Machine learning with R.* Packt Publishing Ltd.

Martin, I. (1991). The Medicine Hat gas field: A large Upper Cretaceous reservoir in southeastern Alberta. *American Association of Petroleum Geologists, 75*.

Martin,, Q. M., & de Paz Santana, Y. (2007). *Tratamiento estadístico de datos con SPSS.* Paraninfo.

Matheron, G. (1962). *Traité de géostatistique appliquée.* Editions Technip.

Monte, R. (1999). The random walk for dummies. *MIT Undergraduate J. Math*, 1(1), 143-148.

Moore, W., Ma, Y., Pirie, I., & Zhang, Y. (2016). *In Unconventional Oil and Gas Resources Handbook.* Gulf Professional Publishing.

Murphy, T., & Lau, A. (Noviembre-Diciembre de 2008). *Standardization News*. Obtenido de https://www.astm.org/

Ochie, K., & Rotimi, O. (August de 2018). Geostatistics–Kriging and Co-Kriging Methods in Reservoir Characterization of Hydrocarbon Rock Deposits. (I. S. Exhibition, Ed.) *Society of Petroleum Engineers*.

Petitgas, P. (1996). Geostatistics and their applications to fisheries survey data. *Computers in fisheries research* , 113-142.

Quesada, L. (Mayo de 2020). *https://geoinnova.org*. Obtenido de https://bit.ly/2PXngPy

Remesan, R., Shamim, M., & Han, D. (2008). Model data selection using gamma test for daily solar radiation estimation. *Hydrological processes*, 22(21), 4301-4309.

Robertson, G. (1987). Geostatistics in Ecology: Interpolating With Known Variance. *Ecological Archives E068-003*, Ecology, 68(3), 744-748.

Rojas, Y. M. (2 de abril de 2013). Obtenido de Yovany Martin Quijano Rojas: https://www.youtube.com/watch?v=vdvO6SJyVmA

Ruhovets, N., & Fertl, W. (1982). Volumes, types, and distribution of clay minerals in reservoir rocks based on well logs. En *In SPE Unconventional Gas Recovery Symposium.* Society of Petroleum Engineers.

Schlumberger. (1989). *Log interpretation principles/applications.* Sugarland, TX: Schlumberger Wireline & Testing,.

Schlumberger Oilfield Glossary. (s.f.). Obtenido de (https://www.glossary.oilfield.slb.com/es/Terms/r/reservoir_characterization_model.aspx)

Schlumberger Oilfield Glossary. (s.f.). Obtenido de (https://www.glossary.oilfield.slb.com/es/Terms/r/reservoir_characterization_model.aspx)

Schumberger. (1991). *Log interpretation Principles/Applications.* Sugar Land, Texas: Schlumberger.

Sheffield. (1956).

Sichel. (1947). The lognormal frequency distribution.

Solano, H. L., & Alvarez, C. R. (2005). *Estadística descriptiva y distribuciones de probabilidad.* Univesidad del Norte.

Timur, A. (January de 1968). An investigation of permeability, porosity, and residual water saturation relationships. *In SPWLA 9th annual logging symposium*.

Tixier, M. (1949). Evaluation of permeability from electric-log resistivity gradients. *The Oil and Gas Journal*, 16, 113-133.

Tsui, A. P., & De Oliveira, A. G. (2002). Tsui, A. P., Jones, A. J., & De Oliveira, A. G. (2002). The Construction of smooth models using irregular embeddings determined by a gamma test analysis. *Neural Computing & Applications*, 318-329.

Uniphyton. (s.f.). Obtenido de https://unipython.com/que-es-el-analisis-exploratorio-de-datos/

Urias, H. Q., & Salvador, B. (2014). *Estadística para ingeniería y ciencias.* Grupo Editorial Patria.

Uribe, I. (2010). *Guía metodológica para la selección de técnicas de depuración de datos.* Medellin: Universidad Nacional de Colombia.

Útil, E. (27 de Octubre de 2013). *https://www.youtube.com/watch?v=kx_o9rnI4DE*. Obtenido de https://www.youtube.com/watch?v=kx_o9rnI4DE

Viera, M., & González, R. C. (2002). *Geoestadística aplicada. Instituto de Geofísica, Universidad Nacional Autónoma de México.* Mexico DF: Ministerio de Ciencia, Tecnología y Medio Ambiente de Cuba.

Walpole, R., Myers, R., Myers, S., & Cruz, R. (1992). *Probabilidad y estadística* . México: McGraw-Hill.

Warrick, A., & Myers, D. (1987). Optimization of sampling locations for variogram calculations. *Water Resources Research*, 23(3), 496-500.

www.researchgate.net. (s.f.). Obtenido de www.researchgate.net

Wyllie, M., & Rose, W. (1950). Some Theoretical Considerations Related to the Quantitative Evaluation of the Physical. *Vol. 189*.