



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO  
POSGRADO EN CIENCIAS FÍSICAS**

# **APLICACIÓN DE REDES BAYESIANAS PARA LA INFERENCIA DE SATURACIÓN DE AGUA**

**TESIS**

**QUE PARA OPTAR POR EL GRADO DE:**

**MAESTRA EN CIENCIAS (FÍSICA)**

**PRESENTA:**

**ROSA MARÍA MARISCAL ROMERO**

**TUTOR:**

**DR. HÉCTOR BENÍTEZ PÉREZ**

**INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS**

**MIEMBROS DEL COMITÉ TUTOR:**

**DR. ALEJANDRO CRISTIAN RAGA RASMUSSEN**

**INSTITUTO DE CIENCIAS NUCLEARES**

**DR. PEDRO EDUARDO MIRAMONTES VIDAL**

**FACULTAD DE CIENCIAS**

**CIUDAD UNIVERSITARIA, CIUDAD DE MÉXICO**

**Mayo 2017**



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



*A mi madre, Paulina Mariscal Romero, gracias por darme la posibilidad de estudiar y por inculcarme el deseo de salir adelante.*

*A mi hermano Francisco Javier Mariscal Romero.*

*Al Dr. Ernesto Rubio Acosta.*

*A Cristina Oñate.*

*A mis amigos.*



## Agradecimientos

Al Dr. Héctor Bénitez Pérez, por permitirme trabajar con él, por su siempre amable disposición, por su motivación, por su apoyo y paciencia.

A los doctores, Pedro Eduardo Miramontes Vidal y Alejandro Cristian Raga Rasmussen por revisar mi trabajo y apoyarme como mis tutores.

A los doctores, David Philip Sanders, Alberto Contreras Cristán, Martín Alberto Díaz Viera y Arturo Iglesias Mendoza por criticar mi tesis y darme consejos para mejorar mi trabajo.

A los doctores, Denis Boyer, Octavio Miramontes Vidal y Germinal Cocho por asesorarme durante el posgrado y ser miembros de mi comité tutorial.

Al Dr. Alejandro Reyes Esqueda por su excelente asesoramiento como coordinador del posgrado.

Al Instituto de Investigación en Matemáticas Aplicadas y en Sistemas de la UNAM por darme cobijo.

Al Instituto de Física de la UNAM por su formación.

Al CONACYT por otorgarme beca de maestría durante mis estudios y posteriormente por el proyecto de Ciencias Básicas 176556.

Al Proyecto “Determinación de la Saturación de Aceite Remanente en Yacimientos Naturalmente Fracturados, a través de la integración de diferentes técnicas de Laboratorio y de Campo (Análisis de Núcleos, Registros Geofísicos y Pruebas de Trazadores) – Y.60011, HAGREPE: Herramienta de Análisis Gráfico con base en Registros SENER-CONACYT-Hidrocarburos 175936 (Consortio IIMAS-IMP).

Al Proyecto Sener-CONACYT 146640.

Al PAPIT IN104516.



## Abstract

A Descriptive Bayesian Network Model has been developed to estimate water saturation according to probabilistic dependencies observed in petrophysical logs.

Estimating water saturation ( $S_w$ ) in oil reservoirs has been one of the main tasks of Petrophysics, which has a direct impact on the risk of economic investment.

In 1942, Archie presented an equation which exhibits the principles of the relationship between resistivity, porosity, and water saturation.

Unfortunately, to estimate water saturation by conventional methods might be a complicated task if the necessary information such as certain logs, well information or core information, is missing in addition to subjective decisions that can be made in its application.

Thus the problem consists of giving a good estimation with few resources (or logs), without losing the essence of the physical properties of rock characterization and their meanings.

In the last years there have been many papers that have applied Machine Learning Techniques such as Artificial Neural Networks (ANN) for estimating water saturation with their variants. Although they have been shown to be functional, they have left a gap in physical meaning of the relationships of internal weights that are generated by ANN.

At present Bayesian Network Models have been applied, in which importance of physical dependencies in the system is weighted in some way. Thus, Bayesian networks provide a method for probabilistic inference based on the statistical dependence of a set of variables.

In this work, known dependencies between petrophysical logs, such as resistivity, neutron porosity, gamma ray, bulk density, sonic, and water saturation, are collected to apply them in the construction of a descriptive Bayesian Network.

This model is applied to a set of existing wells. The results are compared with those obtained using ANN and conventional methods.



## Resumen

Se ha desarrollado un Modelo de Red Bayesiana Descriptiva para estimar la Saturación de Agua de acuerdo a las dependencias probabilísticas observadas en los registros petrofísicos.

La estimación de la Saturación de Agua en yacimientos petroleros es una de las principales tareas de la Petrofísica, donde tiene una repercusión directa de su riesgo en la inversión económica.

En 1942, Archie presentó una ecuación que exhibe los principios de la relación resistividad-porosidad y sus efectos en la estimación de la Saturación de Agua.

Desafortunadamente, estimar la Saturación de Agua por los métodos convencionales puede resultar una labor complicada si se carece de la información necesaria como ciertos registros, información de pozo o información de núcleos<sup>1</sup>, además de las decisiones subjetivas que se pueden encontrar en sus aplicaciones.

Así el problema desemboca en dar un buen estimado con pocos recursos (o registros) sin perder la esencia de las propiedades físicas de la roca y sus significados de caracterización.

En los últimos años se han presentado múltiples trabajos donde se ha aplicado técnicas de Machine Learning tales como las Redes Neuronales Artificiales (ANN) para la estimación de la Saturación de Agua. Aunque han mostrado ser funcionales, han dejado un vacío en el significado físico de las relaciones internas que se generan dentro de la ANN. Inspirados en estos, en este trabajo se ha aplicado un modelo de Red Bayesiana para calcular la Saturación de Agua, donde se rescatan las dependencias probabilísticas en el sistema, y estas se usan como eje principal para calcular los valores por medio de una probabilidad condicional.

Las redes bayesianas proporcionan un método para obtener inferencias probabilísticas en función de la dependencia estadística de un conjunto de variables.

Se recopila las dependencias probabilísticas entre los registros petrofísicos: Resistivo, Porosidad Neutrón, Rayos Gama, Densidad, Sónicos con la Saturación de Agua para construir una red bayesiana de tipo descriptiva.

Se aplica este modelo a un conjunto de pozos reales y se comparan estos resultados con los que se obtienen usando ANN y la aplicación de la ecuación de Archie.

---

<sup>1</sup> Un núcleo consiste en una muestra de roca tomada del pozo a una profundidad específica, por medios especiales, preservando su estructura geológica y sus características fisicoquímicas de la mejor manera posible, con la finalidad de realizar análisis petrofísicos y geológicos.

# Índice general

	<b>Página</b>
Abstract.....	7
Resumen.....	8
<b>1. Introducción.....</b>	<b>11</b>
1.1. Motivación.....	11
1.2. Objetivos.....	12
1.3. Metas.....	12
1.4. Metodología.....	13
1.5. Estructura de la tesis.....	14
<b>2. Antecedentes.....</b>	<b>15</b>
2.1. Bases teóricas para la estimación de la Saturación de Agua por medio de la interpretación de registros geofísicos.....	15
2.1.1. Metodología para la estimación de Saturación de Agua.....	17
<b>3. Teoría de redes bayesianas.....</b>	<b>35</b>
<b>4. Planteamiento del problema.....</b>	<b>47</b>
4.1. Generalidades.....	47
4.2. Problema.....	49
4.3. Hipótesis.....	50
<b>5. Metodología de construcción de Modelo de Red Bayesiana.....</b>	<b>51</b>
5.1. Preparación de datos.....	52
5.2. Identificación de relaciones de probabilidad.....	59
5.3. Discretización de variables.....	64
5.4. Entrenamiento de Red Bayesiana.....	66
5.5. Implementación de Red Bayesiana.....	69
<b>6. Resultados.....</b>	<b>75</b>

<b>7. Conclusiones.....</b>	<b>85</b>
7.1. Logros.....	87
7.2. Trabajo Futuro.....	87
<b>Apéndices.....</b>	<b>89</b>
A. Descripción de registros geofísicos.....	89
B. Tablas descriptivas de los datos que conforman el CampoB.....	93
C. Metodología para la generación de registros sintéticos.....	101
Bibliografía.....	111

# Capítulo 1

## Introducción

Este trabajo se presenta como una propuesta alternativa para la estimación de la saturación de agua con fines de estimar la saturación de aceite.

La necesidad de estimar la saturación de aceite surge de la colaboración en el Proyecto: “Determinación de la Saturación de Aceite Remanente en Yacimientos Naturalmente Fracturados, a través de la integración de diferentes técnicas de Laboratorio y de Campo (Análisis de Núcleos, Registros Geofísicos y Pruebas de Trazadores) – Y.60011, HAGREPE: Herramienta de Análisis Gráfico con base en Registros SENER-CONACYT-Hidrocarburos 175936 (Consortio IIMAS-IMP).

### 1.1. Motivación

El desarrollo actual de nuestro contexto social está basado en tener acceso a fuentes de energía. En los últimos años se han realizado múltiples investigaciones sobre la obtención y uso de energía solar, eólica, geotérmica o la producida por fisión o fusión nuclear. Aunque este tipo de investigaciones e industrias han generado resultados muy optimistas en cuanto a las posibilidades de su aplicación a la vida cotidiana, la fuente principal de energía, por cuestiones de costo, seguridad y facilidad de obtención, sigue siendo el consumo de fuentes fósiles.

En México la exploración y explotación del petróleo es nuestra principal industria y una de las áreas donde se tiene mayor desarrollo científico y tecnológico. Por lo que es importante comprender las características y comportamientos de los hidrocarburos y del medio en el que son generados y atrapados.

La precisión en la determinación de la Saturación de Agua es decisiva para llevar a cabo una evaluación realista de las reservas de hidrocarburos.

La caracterización de yacimientos se refiere al proceso de asignar propiedades cuantitativas utilizando todos los datos de campo disponibles. En el proceso de interpretación las propiedades del yacimiento tales como la porosidad, la permeabilidad y las

saturaciones de líquido en formaciones altamente heterogéneas pueden predecirse utilizando la información de los registros de pozos.

Las propiedades que se suelen calcular en el proceso de caracterización de un yacimiento son: la porosidad, la permeabilidad, las saturaciones de líquidos<sup>2</sup> y los espesores de capas en formaciones. Estas se predicen a partir de los registros geofísicos: eléctricos, nucleares, acústicos y la utilización de todos los datos de campo disponibles.

Debido a que la adquisición de datos de pozos es muy costosa, se apoya la implementación de procesos físicos y matemáticos que con poca información puedan arrojar resultados confiables.

La motivación principal de este trabajo es presentar la utilidad de los modelos probabilísticos, en particular los de Redes Bayesianas, para la modelación de las propiedades petrofísicas y en particular para la predicción de Saturación de Agua.

## 1.2. Objetivos

El objetivo general es construir una metodología, aplicando modelos de red bayesiana, para inferir la Saturación de Agua a partir de las relaciones de probabilidad entre los registros geofísicos.

Los objetivos específicos son:

- Analizar y modelar las relaciones entre registros geofísicos en relación con la Saturación de Agua.
- Construir una red bayesiana que represente las relaciones identificadas.
- Construir un algoritmo para que la red bayesiana asimile la nueva información.

## 1.3. Metas

En base a los objetivos planteados, en este trabajo se desea obtener los siguientes resultados.

- i. Automatización de la metodología de Archie-Pickett para la estimación de la Saturación de Agua.
- ii. Análisis descriptivo de los registros de línea de pozo que se utilizarán para calcular la Saturación de Agua.
- iii. Análisis de correlación entre los registros geofísicos.

---

<sup>2</sup> La saturación de cualquier fluido en una formación se define como “La fracción del fluido de un espacio poral dado”. Se expresa como un porcentaje. Si el espacio poral corresponde a la porosidad total, la saturación se conoce como la saturación total del fluido. [42]

- iv. Grafo representativo de la red bayesiana del modelo.
- v. Curva de Saturación de Agua calculada por medio de la red bayesiana.

## 1.4. Metodología

Este trabajo inició con la construcción de una base adecuada de datos, se trabajó con un conjunto de pozos correspondientes a un Campo de un Yacimiento Areno Arcilloso. Los datos fueron modificados con el fin de guardar su confidencialidad, estos se presentan en la Tabla 1.1.

El campo se eligió por su diversidad de datos de una base de datos restringida del IIMAS la cual contiene tres campos, se escogió el CampoB.

Nombre de Campo <sup>3</sup>	Características Geológicas	No. Pozos	Registros Geofísicos
<b>CampoB</b>	<p>- Era Geológica: Roca Terciaria</p> <p>- Tipo de Yacimiento: Areno Arcilloso</p>	19	<p>- MSFL – Resistivo</p> <p>- GR – Rayos Gamma</p> <p>- NPFI – Porosidad</p> <p>- RHOB - Densidad</p> <p>- LLD – Resistivo</p>

**Tabla 1. 1. Campos que conforman la base de datos utilizada en este trabajo.**

Orden	Pozos de	
	Entrenamiento	Aplicación
1	38	41
2	40	44
3	42	47
4	43	58
5	50	62
6	53	63
7	55	65
8	56	67
9		68
10		70
11		72
No.Datos	112094	108129
%Datos	56.20%	43.80%

**Tabla 1. 2. Conjunto de pozos del CampoB utilizados para Entrenamiento y Aplicación**

<sup>3</sup> Campo de petróleo: Una acumulación o grupo de acumulaciones de petróleo en el subsuelo. Un campo de petróleo está formado por un yacimiento con una forma adecuada para el entrapamiento de hidrocarburos y que se encuentra cubierto por una roca impermeable que actúa como sello. [42]

Como se indica cada Campo está compuesto por un conjunto de pozos. En cada pozo suele realizarse una serie de medidas llamadas registros petrofísicos, que son medidas físicas de propiedades de las rocas en medidas regulares cada cierto espaciamiento, que usualmente es 0.16m.

Cada pozo es nombrado por un indicador numérico que no es secuencial. En la Tabla 1.2 se presenta el conjunto de pozos con el que se realizó el procedimiento de entrenamiento de la Red Bayesiana que es el 56.2% del total de datos y también se presenta el conjunto de pozos con el cual se validó la aplicación de los algoritmos que es el 43.8% de los datos, restantes.

En la primera parte de este trabajo se obtuvo la información muestral de los registros más frecuentemente medidos en los pozos para comprender su comportamiento.

Se aplicó la metodología teórica para calcular Saturación de Agua.

Se realizaron gráficas cruzadas de los registros para identificar las relaciones existentes entre los registros y con estas se construyeron las tablas de relación entre ellos.

Se planteó un grafo para presentar las relaciones de registros en una estructura de red donde las relaciones estuvieran ligadas a la relación estadística de las variables.

Esta estructura fue posteriormente aplicada a una muestra de control.

## 1.5. Estructura de la tesis

Esta tesis está estructurada como sigue:

El Capítulo 2 presenta la **Metodología teórica** que se utiliza en el área de geofísica para estimar la Saturación de Agua utilizando los registros de pozo. También se presenta la teoría para la construcción de redes bayesianas y los fundamentos teóricos de su aplicación.

El Capítulo 3 se realiza el **Planteamiento del problema** y se presenta la hipótesis a responder en este trabajo de investigación.

El Capítulo 4 presenta la **Metodología de construcción del modelo de Red Bayesiana**, dividida en seis partes: la **Preparación de datos**, donde se presenta la base de datos que se utilizó el trabajo de investigación; la **Construcción de Red Bayesiana** que muestra cómo se realiza el trabajo de construcción de la estructura de red; la **Identificación de relaciones de probabilidad**; el proceso de **Discretización de Variables**; el **Proceso de entrenamiento de red**; y las **Pruebas de predicción**.

El capítulo 5 presenta los **Resultados** y el Capítulo 6 las **Conclusiones** de este trabajo.

## Capítulo 2

### Antecedentes

El objetivo de este capítulo es presentar los antecedentes teóricos necesarios para plantear la problemática que se desarrolla en este trabajo.

#### 2.1. Bases teóricas para la estimación de la Saturación de Agua por medio de la interpretación de registros geofísicos [5]

Los registros de pozos son técnicas que se utilizan en pozos para la exploración y explotación petrolera, mineral, geo-hidroológica, geotérmica y geotécnica, debido a su confiabilidad y versatilidad.

En la actualidad los registros son adecuados para la determinación del contenido de hidrocarburos, agua, minerales y competencia de las rocas sujetas a diferentes esfuerzos, también para localizar formaciones de alta y baja porosidad, identificar lutitas<sup>4</sup>, formaciones de litología<sup>5</sup> compleja y condiciones del agujero y tubería.

Se distinguen de otros métodos geofísicos porque dan un diagnóstico más confiable de las condiciones de las rocas en el subsuelo, de la tubería de revestimiento y de la cementación, ya que miden directamente las propiedades físicas y ayudan a verificar los parámetros medidos en la superficie.

Un **registro geofísico de pozo** es una representación digital de una propiedad física que se mide contra la profundidad in-situ en metros. La obtención de este se realiza a partir

---

<sup>4</sup> Lutitas. Roca sedimentaria detrítica, fisible, de granos finos, formada por la consolidación de partículas del tamaño de la arcilla y el limo en capas relativamente impermeables de escaso espesor. Es la roca sedimentaria más abundante. Las lutitas pueden contener cantidades relativamente grandes de material orgánico, en comparación con otros tipos de rocas y, por consiguiente, poseen el potencial para convertirse en rocas generadoras ricas en hidrocarburos, aunque una lutita típica contiene sólo un 1% de materia orgánica. El tamaño de grano fino típico y la falta de permeabilidad, que son consecuencia de la alineación de sus granos laminares o escamosos, permiten que las lutitas formen buenas rocas de cubierta para trampas de hidrocarburos. [42]

<sup>5</sup> Litología. La naturaleza macroscópica del contenido mineral, la granulometría, la textura y el color de las rocas. [42]



de una sonda que va recorriendo la trayectoria del pozo y de un equipo superficial que traduce la información enviada por la sonda a través de un cable que se registra en el subsuelo en una cinta magnética o película fotográfica.

Las sondas son cilindros de metal que tienen en su parte inferior circuitos electrónicos muy sofisticados, los cuales varían de acuerdo a la propiedad que mide, a las necesidades de investigación, condiciones y profundidad del pozo. Las maneras de medir las distintas propiedades de las rocas parte de un sistema básico común: un emisor (electrodo, bobina, emisor de radiactividad, emisor de neutrones, emisor de ondas acústicas) que envía una señal (corriente eléctrica, campo electromagnético, radiactividad, neutrones u ondas acústicas) hacia la formación, donde uno o varios receptores miden la respuesta de la formación a cierta distancia del emisor.

La estructura de los pozos y como se toman los datos por la sonda es representada en la Figura 2.1. En la figura se señala un punto  $X_n$  a lo largo de un pozo donde se realizaron las mediciones de los registros, donde

$$X_n = (y_n, x_{n1}, x_{n2}, x_{n3}, x_{n4}), \quad (2.1)$$

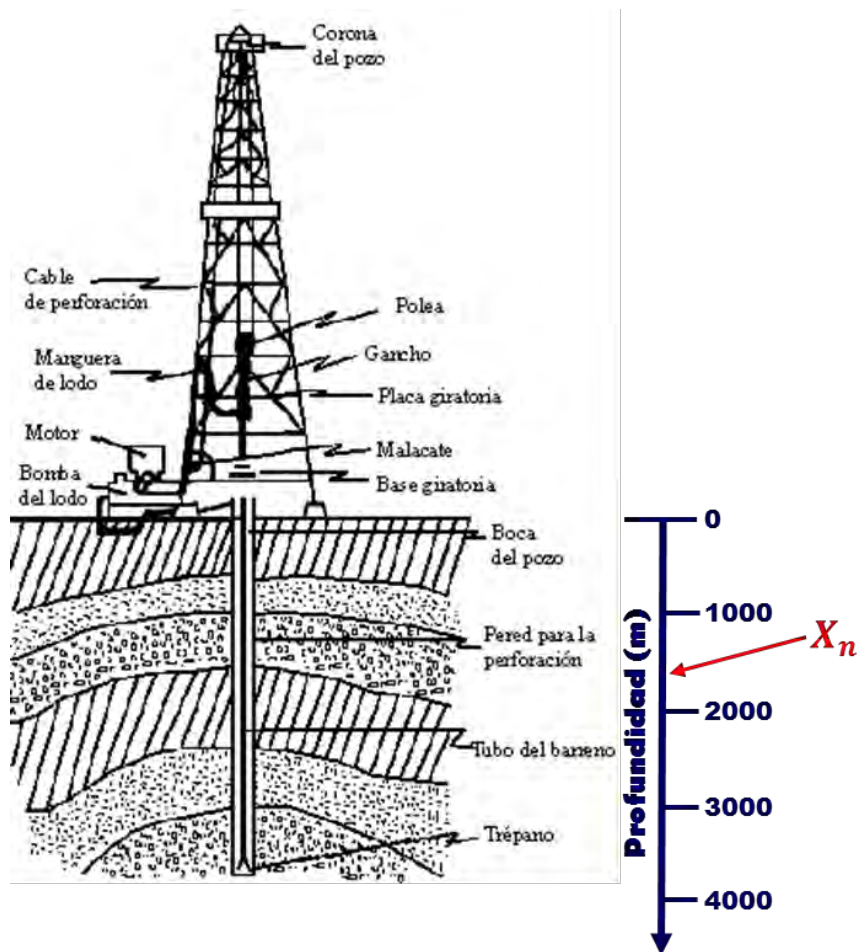


Figura 2. 1. Diagrama de un pozo petrolero. [46]

$y_n$  representa un valor de profundidad,  $x_{n1}$  es el valor del registro GR a la profundidad  $y_n$ ,  $x_{n2}$  es el valor del registro MSFL a la profundidad  $y_n$ ,  $x_{n3}$  es el valor del registro NPHI a la profundidad  $y_n$  y  $x_{n4}$  es el valor del registro RHOB a la profundidad  $y_n$ . Cada registro forma una secuencia de datos donde por ejemplo

$$GR = \{x_{i1} : i = 1, \dots, N\}, \quad (2.2)$$

donde N es el tamaño de la muestra de datos que se tomó de cada pozo.

El conjunto de registros utilizado para la secuencia de esta metodología fue {GR, MSFL, NPHI, RHOB}, ver Figura 2.2. Para conocer la composición de las herramientas que miden estos registros y sus unidades, ir a Apéndice A: Descripción de registros geofísicos.

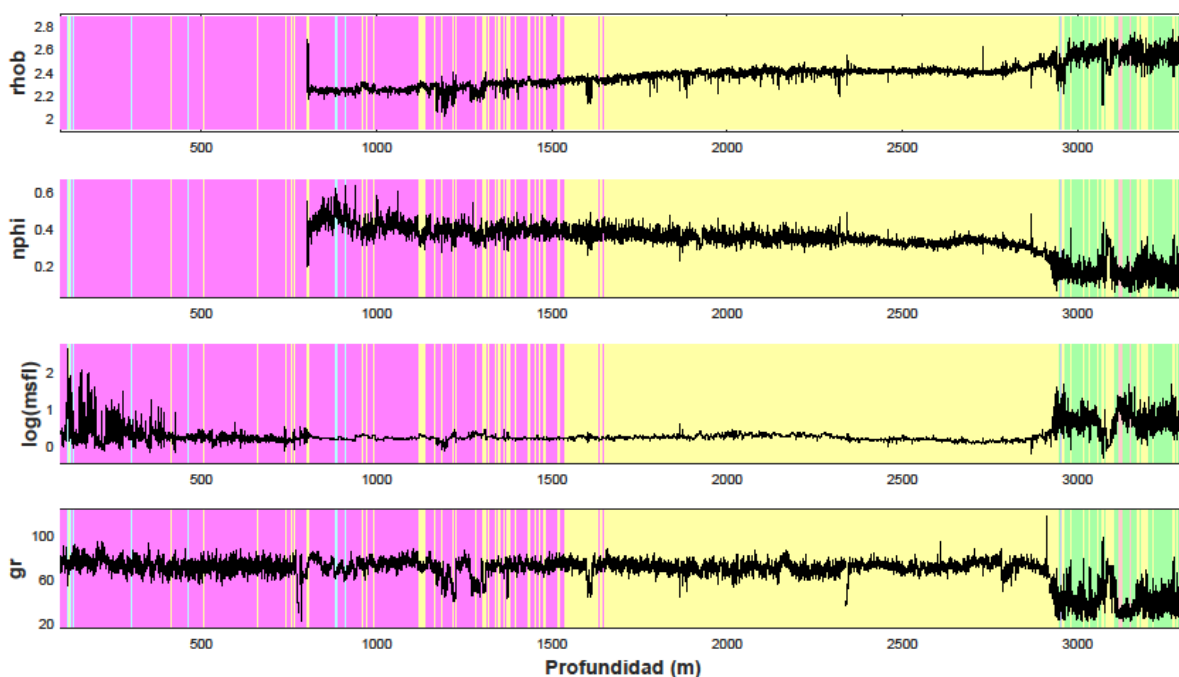


Figura 2. 2. Registros del pozo 67 de Datos Limpios – CampoB.

### 2.1.1. Metodología para la estimación de Saturación de Agua

La **saturación de un fluido** se define como la fracción del volumen del poro de una roca que está llena con este fluido;

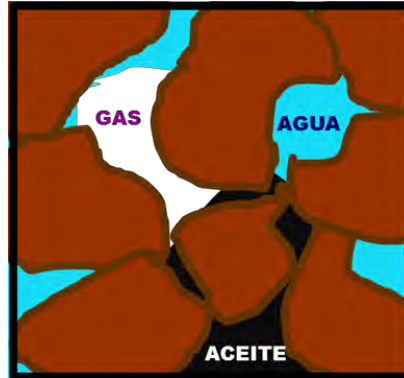
$$S = \frac{\text{Volumen del fluido}}{\text{Volumen del poro}}. \quad (2.3)$$

Al representar una fracción de volumen el rango de cualquier saturación siempre está entre 0 y 1.

En particular, una forma simple de indicar el porcentaje de fluidos que componen una roca es pensar que el 100% de volumen de poros de una roca está compuesto por agua, gas y aceite, ver Figura 2.3, de forma que la saturación del agua,  $S_w$ , más la del aceite,  $S_o$ , más la del gas  $S_g$  dan por resultado el volumen total, es decir,

$$S_w + S_o + S_g = 1. \quad (2.4)$$

Estimar el valor de  $S_w$  en la industria petrolera, permite estimar el porcentaje de aceite presente en la roca.



**Figura 2. 3. Representación del compuesto de gas, agua y aceite en un volumen de roca.**

En registros de pozos generalmente se habla de la saturación del agua, debido a que al poder inferir este valor es posible estimar el valor de los otros fluidos involucrados en una misma mezcla.

La Saturación de Agua se calcula por medio de la Ecuación de Archie [4],

$$S_w = \left( \frac{F R_w}{R_t} \right)^{1/n}, \quad (2.5)$$

donde  $R_w$  es la resistividad del agua de la formación,  $R_t$  es la resistividad verdadera<sup>6</sup> de la roca en su estado natural sin contaminación del filtrado del lodo,  $n$  es el exponente de saturación<sup>7</sup>, siendo su valor usual 2, y  $F$  es el Factor de Formación, el cual se define como la resistividad  $R_o$  de una roca 100% saturada con un fluido (generalmente agua salada) entre la resistividad  $R_w$  del agua que la satura,

<sup>6</sup> Para fines prácticos la “resistividad verdadera” es la resistividad que es medida en el registro geofísico resistivo de penetración profunda: LLD.

<sup>7</sup> El exponente de saturación expresa el efecto sobre la resistividad de la de saturación de la muestra, o del remplazo del agua con un fluido no conductor. En las rocas humedecidas con agua, petrofísicamente simples (rocas de Archie),  $n$  es constante para diferentes valores de  $S_w$ , y puede hallarse un promedio único  $n$  para un yacimiento o una formación determinada. Un valor habitual es 2. En las rocas con minerales conductores, tales como las arenas arcillosas,  $n$  se vuelve cada vez más bajo conforme  $S_w$  se reduce. [42]

$$F = \frac{R_o}{R_w}. \quad (2.6)$$

Este factor se puede relacionar con la porosidad a través de la siguiente expresión:

$$F = \frac{a}{m}, \quad (2.7)$$

donde  $a$  es el coeficiente de tortuosidad<sup>8</sup>,  $\phi$  es la porosidad y  $m$  es el exponente de cementación<sup>9</sup>. Los valores  $a$  y  $m$  dependen de la geometría interna y del tipo de porosidad de la roca.

Cuando se tienen rocas de tipo granular se utiliza,

$$F = \frac{0.62}{2.15} \approx \frac{0.81}{2}, \quad (2.8)$$

expresión que es conocida como relación de Humble [44], mientras que para rocas carbonatadas se utiliza,

$$F = \frac{1}{2}. \quad (2.9)$$

Cuando se tiene lutita<sup>3</sup>, arcilla, anhidrita<sup>10</sup>, yeso, gas, etc., la relación no se conserva, por lo que en dicha situación se requiere de un número mayor de registros para hacer correcciones que eliminen la influencia de estos elementos.

En el proceso de perforación, el agujero puede aumentar o disminuir de diámetro, el fluido de perforación<sup>11</sup> puede invadir zonas permeables y se pueden generar enjarres.

Las lutitas a menudo se dañan durante la perforación. Este daño puede ser el resultado de hidratación, causando que se hinche y aumente el diámetro del agujero por derrumbes o por solución.

<sup>8</sup> En la mecánica de fluidos de los medios porosos, la tortuosidad es la relación del largo de la corriente natural de un fluido —una línea o trayecto de flujo— entre dos puntos y la distancia en línea recta entre esos puntos. La tortuosidad se relaciona con la relación entre el coeficiente de difusión de un fluido cuando no está confinado por un medio poroso y su coeficiente de difusión efectivo cuando está confinado en un medio poroso. La tortuosidad se relaciona además con el factor de formación, que es la relación entre la resistividad eléctrica de un fluido conductivo en un medio poroso y la resistividad eléctrica del fluido propiamente dicho. [42]

<sup>9</sup> El exponente de cementación o porosidad,  $m$ , en la relación entre el factor de formación,  $F$ , y la porosidad,  $\phi$ . En la ecuación de Archie,  $F = 1 / \phi^m$ . H. Guyod designó  $m$  al exponente de cementación porque se observó que  $m$  era más alto en la roca cementada. El término más general es "exponente de porosidad." [42][43]

<sup>10</sup> Anhidrita. Un integrante del grupo de minerales evaporíticos y la roca blanda que comprende la anhidrita formada por la precipitación del sulfato de calcio a partir de la evaporación del agua de mar. [42]

<sup>11</sup> El fluido de perforación es un líquido o gas que circula a través de la sarta de perforación hasta a la barrena y regresa a la superficie por el espacio anular. Hasta la fecha un pozo de gas o aceite no se puede perforar sin este concepto básico de fluido circulante. [42]

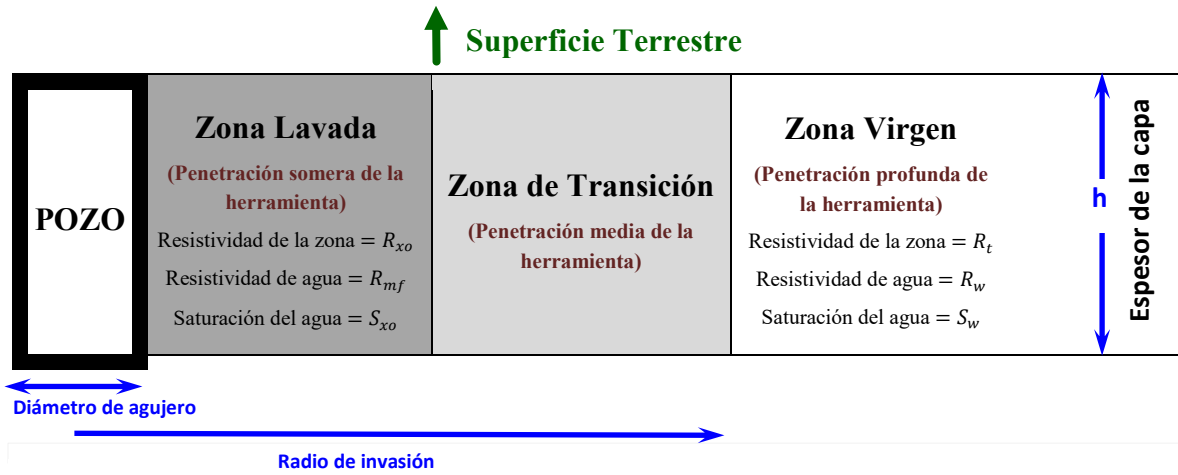


Figura 2. 4. Sección transversal esquemática de un agujero. Se muestra el esquema de una formación permeable. Los fluidos en la zona de transición son una combinación de filtrado de lodo y agua de la formación.

PARAMETRO	ZONA LAVADA	ZONA DE TRANSICIÓN	ZONA VIRGEN
Resistividad de la roca	$R_{xo}$	$R_i$	$R_t$
Saturación del fluido	$S_{xo}$	$S_i$	$S_w$
Porosidad			
Resistividad del fluido	$R_{mf}$	$R_z$	$R_w$
Ecuación de Archie	$S_{xo} = \left(\frac{FR_{mf}}{R_{xo}}\right)^{1/n}$	$S_i = \left(\frac{FR_z}{R_i}\right)^{1/n}$	$S_w = \left(\frac{FR_w}{R_t}\right)^{1/n}$
con invasión	$S_{xo} > S_i > S_w$		
sin invasión	$S_{xo} = S_i = S_w$		

Tabla 2. 1. En esta tabla se presenta la relación de nombres que reciben los parámetros para las diferentes zonas de invasión que se generan en el proceso de evaluación de las diferentes herramientas. [5]

La hidratación no sólo cambia el valor de resistividad de las lutitas también otros parámetros en la vecindad del agujero.

Las herramientas que miden la resistividad están clasificadas por la profundidad de investigación. Ver Figura 2.4 y Tabla 2.1. Si la herramienta solo mide algunas pulgadas dentro de la formación, la resistividad que mide se conoce como  $R_{xo}$ ; si la medida está en un rango intermedio se mide  $R_i$  y si la herramienta mide muy profundo  $R_t$ . Debe de recordarse que si la invasión es muy somera o no existe, todas las herramientas medirán  $R_t$ .

El uso de versión de la Ecuación de Archie correspondiente vendrá determinado por la herramienta utilizada para los registros resistivos que tengamos disponibles.

En particular en este trabajo de investigación se enfatizó el uso de la Ecuación de Archie para la Zona Lavada debido a que el conjunto de registros resistivos que teníamos fueron obtenidos por una herramienta de alcance de penetración somera.

Para poder utilizar la ecuación de saturación se requiere calcular el valor de la resistividad de formación y del exponente de cementación. En condiciones donde estos valores no son obtenidos directamente de la información conocida del pozo es necesario realizar una secuencia de pasos para su estimación.

En la Figura 2.5 se presenta el proceso para calcular el valor de Saturación de Agua como la secuencia de pasos que se utilizó para este trabajo de tesis.

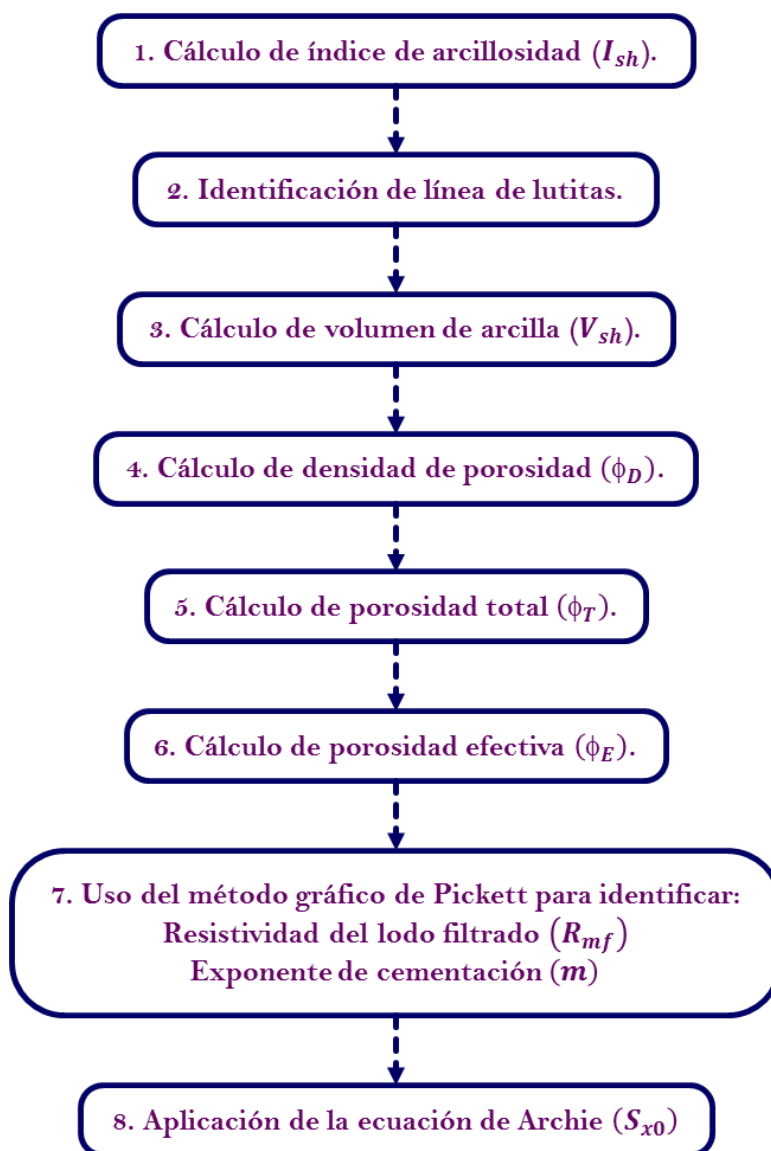


Figura 2. 5. Secuencia de pasos para la estimación de la Saturación de Agua en Zona Lavada.

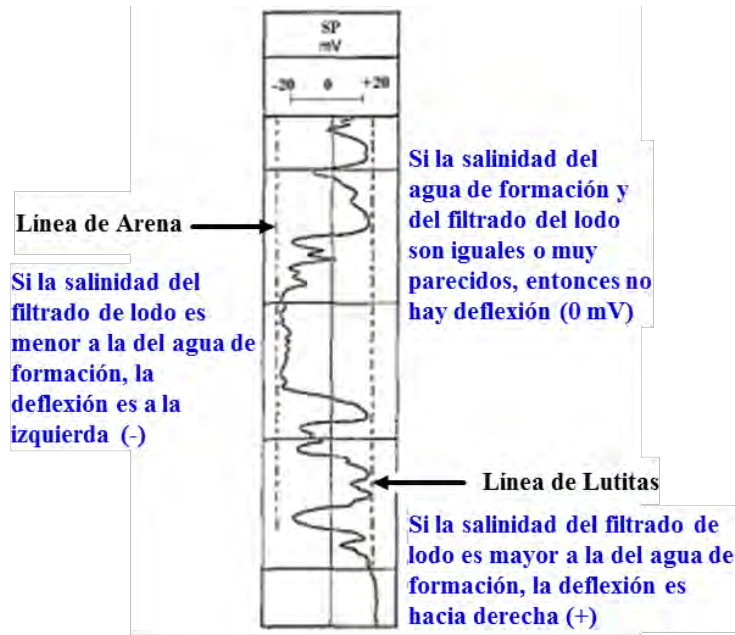


Figura 2. 6. Registro Potencial Natural o Potencial Espontaneo (SP). [45]

Antes de realizar los cálculos, hay que dividir el pozo en zonas que tengan características relativamente homogéneas. Se debe identificar cualitativamente los límites entre capas y en estos intervalos se considera que los valores que se toman son los promedios.

Para identificar regiones uniformes de aplicación suele utilizarse el registro de potencial natural<sup>12</sup> (SP), donde el espesor de las capas se obtiene de los puntos de inflexión de esta curva respecto a la línea base de lutitas.

Una versión alterna a la metodología del uso del registro SP para la identificación de capas en un pozo es la del uso del Registro de Rayos Gamma (GR), debido a que en zonas donde la radiactividad no sea muy alta se considera como zona limpia.

Ambas curvas SP y GR son indicadoras de zonas porosas y permeables y también indican la arcillosidad de la formación. La curva SP tiene un comportamiento estable en presencia de arcilla, ver Figura 2.6. La curva GR presenta valores más altos de radioactividad en arcillas y arenas arcillosas como resultado del incremento de potasio. A menudo es posible ver que las dos curvas son correlacionales cuando la salinidad de la formación es mayor que la del lodo.

<sup>12</sup> El registro potencial natural (SP) mide el potencial eléctrico que se produce debido a la interacción del agua de formación innata, el fluido de perforación conductivo y ciertas rocas selectivas de iones (lutitas). Esta diferencia de potencial se mide con ayuda de dos electrodos uno móvil en el pozo y otro electrodo fijo en la superficie. [42]

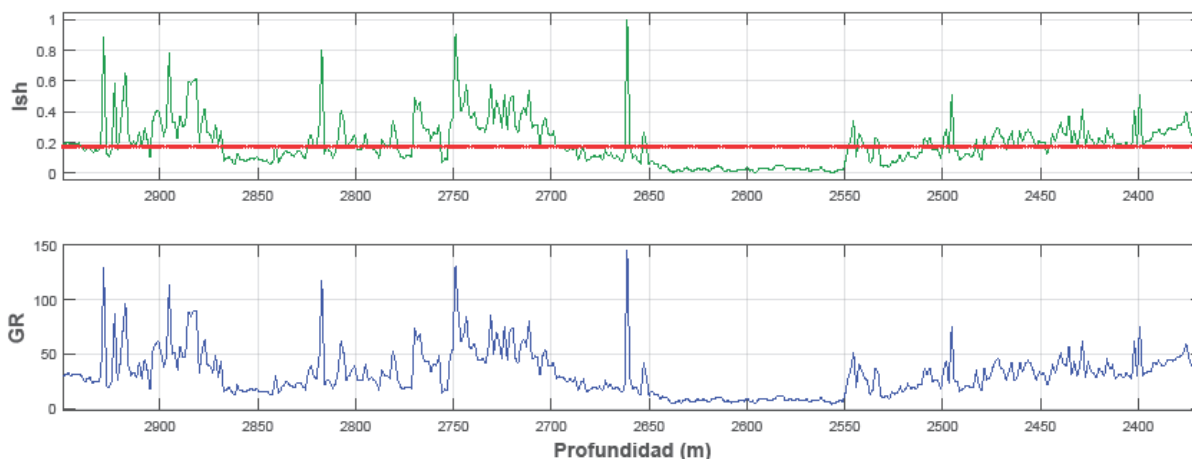
## Índice de arcillosidad

La Arcillosidad en una formación puede ser representada como el contenido de mineral arcilloso y/o lutita que se encuentra en la misma, es decir, define qué tan limpia puede estar la arena que deseamos evaluar.

Para estimar el Índice de Arcillosidad  $I_{sh}$ <sup>13</sup> se aplica la ecuación

$$I_{sh} = \frac{GR - GR_{\min}}{GR_{\max} - GR_{\min}}. \quad (2.10)$$

Dado que la ecuación de Saturación de Archie [4] se usa para determinar la Saturación de Agua en formaciones limpias, en aquellas que contienen menos del 15% de volumen de arcilla, podemos utilizar este porcentaje para definir un valor límite del Índice de arcillosidad para separar las regiones de formaciones limpias. Este límite tiene una relación directa con la línea base de lutitas del potencial natural (SP) y de igual forma puede utilizarse para zonificar un pozo.



**Figura 2. 7. Curva GR y curva Ish de un pozo. La línea verde sobre la curva de Ish indica la línea base de lutitas respecto a la cual por medio de inflexiones<sup>14</sup> identificamos las formaciones limpias. El rectángulo azul indica una capa identificada por medio de este método.**

<sup>13</sup> Índice de Arcillosidad  $I_{sh}$ , por sus siglas en ingles de shaliness index.

<sup>14</sup> Inflexión. Curvatura, desviación o torcimiento de la línea del registro de forma muy pronunciada.



## Identificación de línea base de lutitas

Utilizando la curva  $I_{sh}$ , definimos la línea base de lutitas como el 15% de esta, por lo que todos los puntos en la curva que tienen un valor inferior se consideran como puntos de una formación limpia. Ver Figura 2.7.

A partir de este punto el pozo es segmentado en regiones homogéneas de análisis donde se considera que los valores de  $R_{mf}$  y  $m$  son valores constantes.

## Volumen de arcilla

Una vez que se obtuvo el  $I_{sh}$  de la formación analizada, el siguiente paso es calcular el valor del volumen de arcilla,  $V_{sh}$ , para cada intervalo (capa).

Se podría considerar que el  $V_{sh}$  es igual al  $I_{sh}$ , pero lo ideal es considerar la era o edad geológica que corresponde al pozo que se está analizando. Existen relaciones empíricas que estiman el  $V_{sh}$  considerando la corrección de la edad geológica a la que corresponde; las más utilizadas son:

Para rocas terciarias la relación de Larionov [37]:

$$V_{sh} = 0.083(2^{3.7 \cdot I_{sh}} - 1). \quad (2.11)$$

Para rocas del Cretácico superior, Stieber [38]:

$$V_{sh} = \frac{I_{sh}}{3 - 2 \cdot I_{sh}}. \quad (2.12)$$

Para rocas del Cretácico inferior, Clavier [39]:

$$V_{sh} = 1.7 - [3.38 - (I_{sh} + 0.7)^2]^{0.5}. \quad (2.13)$$

Para rocas de Jurásico, Larionov [37]:

$$V_{sh} = 0.33(2^{2 \cdot I_{sh}} - 1). \quad (2.14)$$

En la Figura 2.8 se presentan las diferencias gráficas que tiene cada ecuación para el cálculo del Volumen de Arcilla, de acuerdo al tipo de era geológica a la que corresponde la formación.

Las variaciones que se den en los valores obtenidos del  $V_{sh}$  en los rayos gamma, fluctúa entre un 17% y un 26%, dependiendo de la ecuación que se utilice. Dependiendo del criterio del analista.

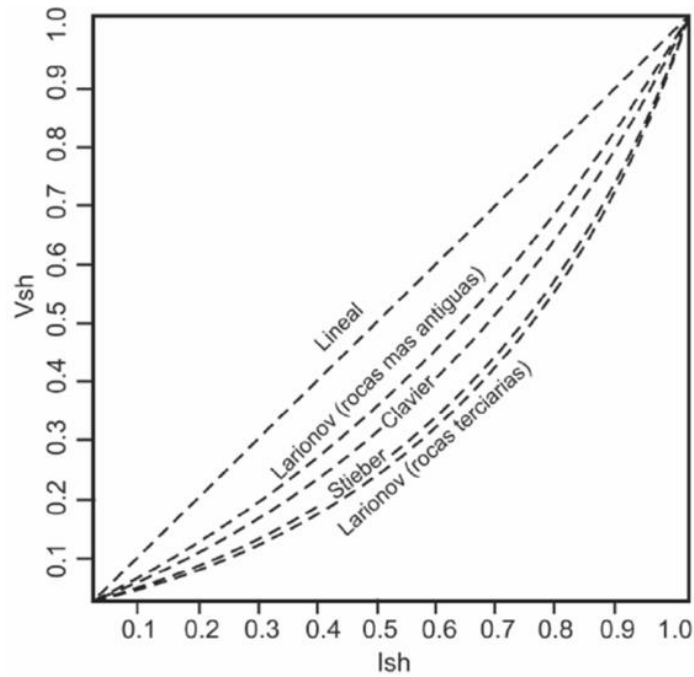


Figura 2.8. Correlación entre  $I_{sh}$  y  $V_{sh}$  en el cálculo de la arcillosidad en las formaciones. [5]

## Porosidad

La porosidad o fracción de huecos es la medida de espacios vacíos internos de una roca que pueden ser ocupados por un fluido,

$$= \frac{V_p}{V_t}, \quad (2.15)$$

donde es la porosidad,  $V_p$  el volumen de poros disponibles ocupados por el fluido y  $V_t$  el volumen total.

es un valor importante en la evaluación petrofísica, específicamente para el cálculo de saturaciones.

## Densidad de porosidad

La ecuación que relaciona la densidad con la porosidad se expresa como

$$D = \frac{\rho_{ma} - \rho_b}{\rho_{ma} - \rho_f} = \frac{\rho_{ma} - RHO B}{\rho_{ma} - \rho_f}, \quad (2.16)$$

donde  $\rho_D$  es la Densidad de Porosidad,  $\rho_{ma}$  es la densidad de la matriz<sup>15</sup> cuando la formación no tiene porosidad,  $\rho_b$  es la densidad del registro y  $\rho_f$  es la densidad del fluido; 1.0, cuando el agua del lodo es dulce; y 1.1 cuando el lodo es muy salado.

Los valores de la densidad de la matriz,  $\rho_{ma}$ , dependen del tipo de litología<sup>4</sup>. Los valores de matriz conocidos se presentan en la Tabla 2.2.

Las lutitas tienen densidades que varían de 2.4 a 2.7 gr/cc por lo que el comportamiento de la densidad de porosidad varía considerablemente entre la elección de una u otra densidad de matriz. El registro de densidad de porosidad muestra densidades muy altas cuando hay presencia de gas. El cálculo de la densidad de porosidad se vuelve complejo cuando se desconoce la litología o cuando se presentan más de dos minerales con proporciones desconocidas.

Las combinaciones más comunes de minerales en rocas sedimentarias son:

Sustancia	Densidad (g/cc)
Arenisca	2.65
Caliza	2.71
Dolomía	2.87
Anhidrita	2.98
Halita	2.03
Cal	2.35
Arcilla	2.7-2.8
Agua Potable	1.0
Agua de Mar	1.03-1.06
Aceite	0.6-0.7
Gas	0.015

**Tabla 2. 2. Tabla de valores de  $\rho_{ma}$  para diferentes tipos de roca. [18]**

- En rocas carbonatadas: Calcita/dolomía; Calcita/sílice; Dolomía/sílice y Dolomía/anhidrita

<sup>15</sup> Matriz. Las partículas intersticiales de grano más fino, que se encuentran entre partículas más grandes, o en las cuales se encuentran encastradas las partículas más grandes en las rocas sedimentarias, tales como las areniscas y los conglomerados.

- En arenas: Sílice/dolomía; Sílice/calcita; Sílice/minerales pesados (pirita o siderita).

### Identificación de zonas de gas [5]

El registro de porosidad NPHI responde al contenido de hidrogeno. Cuando hay presencia de gas la respuesta de la porosidad es baja. Las arcillas tienen agua impregnada en sus paredes, en este caso la respuesta del registro de porosidad es más alta. Ver Figura 2.9.

Una arena arcillosa con gas puede comportarse como una arena limpia. Por tal motivo se requiere un análisis para descartar intervalos con alta presencia de gas.

El registro de Densidad muestra un incremento en la porosidad en zonas con gas. Una manera de detectar la presencia de gas es a partir de la combinación Neutrón-Densidad.

La manera de realizar esta prueba es comparar simultáneamente los registros Densidad de Porosidad  $\rho_p$ , el Registro de Neutrones NPHI y el Registro Sónico DT y comparar las respuestas de los mismos para los diferentes fluidos.

Cada capa es una mezcla de Aceite-Agua-Gas en diferentes concentraciones, en cada capa se realiza una separación de puntos considerando el efecto que se observa en los registros de densidad  $\rho_p$  y porosidad  $N$  de acuerdo a la Tabla 2.3. Como se indicó la presencia de gas produce efectos de baja porosidad  $N$  y alta porosidad de densidad  $\rho_p$ , de tal forma que

$$\rho_p - N > 0.03. \tag{2.17}$$

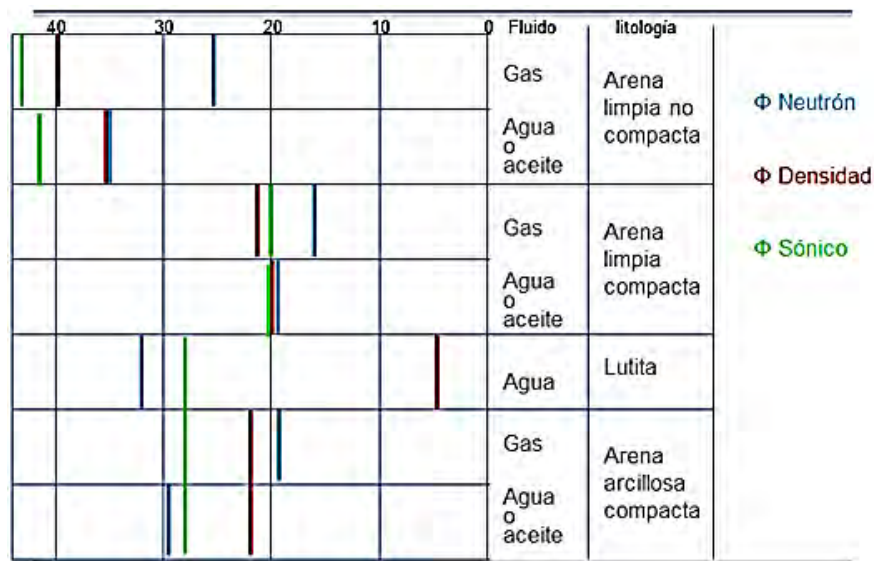


Figura 2.9. Respuesta de los registros Neutrón-Densidad-Sónico para diferentes tipos de arenas y fluidos. [5]

Litología	Fluido	Porosidad	Densidad	Sónico	Densidad-Porosidad
Arena limpia no compacta	Gas	0.26	0.395	0.43	0.135
	Agua o Aceite	0.345	0.355	0.42	0.01
Arena limpia compacta	Gas	0.17	0.22	0.2	0.05
	Agua o Aceite	0.195	0.2	0.205	0.005
Lutita	Agua	0.325	0.045	0.275	-0.28
Arena arcillosa compacta	Gas	0.19	0.225	0.275	0.035
	Agua o Aceite	0.29	0.225	0.275	-0.065

Tabla 2. 3. Relación numérica de la respuesta de los registros Neutrón-Densidad-Sónico para diferentes tipos de arenas y fluidos.

## Porosidad total

Porosidad Total ( $v_t$ ) es la relación del volumen total de espacios „vacíos“ (poros canales vesículas, etc.) existentes entre los elementos minerales de la roca, al volumen total de la roca,

$$v_t = \frac{v_t - v_s}{v_t} = \frac{v_p}{v_s}, \quad (2.18)$$

donde  $v_p$  representa el volumen de espacios „vacíos“ generalmente ocupados por fluidos: agua, gas o aceite,  $v_s$  es el volumen ocupado por los elementos sólidos y  $v_t$  es el volumen total de la roca.

Para  $D$  y  $N$  la respuesta es diferente. En combinación es posible manejar casos distintos para determinar la porosidad total de formación, en condiciones de gas

$$T = \sqrt{\frac{\frac{2}{D} + \frac{2}{N}}{2}}. \quad (2.19)$$

En condiciones sin gas

$$T = \frac{D + N}{2}. \quad (2.20)$$

## Porosidad efectiva

La porosidad Efectiva ( $e$ ) es la porosidad accesible a los fluidos, excluyendo a la porosidad no conectada y al volumen ocupado por el agua ligada a las partículas de arcilla. La determinación in-situ de este parámetro se obtiene de

$$e = \tau(1 - V_{sh}). \quad (2.21)$$

## Métodos gráficos [18]

La técnica para estimar el valor de la resistividad del agua de formación  $R_{mf}$  para cada intervalo viene de analizar la curva de Potencial Natural<sup>11</sup> (SP), aunque la metodología no es única. En casos en que no se tiene la curva Potencial Natural (SP) o no tiene buena resolución, existen otros procedimientos para estimar su valor.

Algunas técnicas para determinar el valor de  $R_{mf}$ , son conocidas como *Métodos Gráficos de Interpretación Resistividad-Porosidad*. Las gráficas de interrelación resistividad-porosidad pueden ser de dos tipos y sirven para analizar intervalos donde:

- Se presentan rocas consolidadas donde se tiene un rango muy grande de porosidades.
- No se tiene una evidencia obvia de la presencia de agua, esto es, cuando la curva SP no tiene una buena deflexión.
- Sólo se tiene un registro para el cálculo de porosidad.
- Se tienen formaciones de porosidad baja.
- Se desconoce  $R_{mf}$ , pero puede ser constante en todo el intervalo.
- Se desconoce la densidad o la velocidad de la matriz.
- La formación de interés está limpia.

Las gráficas que se utilizan son la de Hingle y la de Pickett.

## Gráfica de Hingle [10]

El *Método Gráfico de Hingle* es una solución gráfica a la ecuación de Archie, que consiste en graficar la porosidad vs  $\left(\frac{1}{\text{resistividad}}\right)^{1/m}$  para obtener los valores de  $R_{mf}$  y densidad o velocidad de matriz. Ver Figura 2.10.

De la Ecuación de Archie [4] para la zona lavada considerando  $m = 2$ ,  $n = 2$  y  $a = 1$ ,

$$S_{xo} = \left( \frac{R_{mf}}{2R_{xo}} \right)^{1/2}, \quad (2.22)$$

se puede reacomodar la expresión como:

$$\left[ \frac{1}{R_{xo}} \right]^{1/2} = \left[ \frac{S_{xo}^2}{R_{mf}} \right]^{1/2}. \quad (2.23)$$

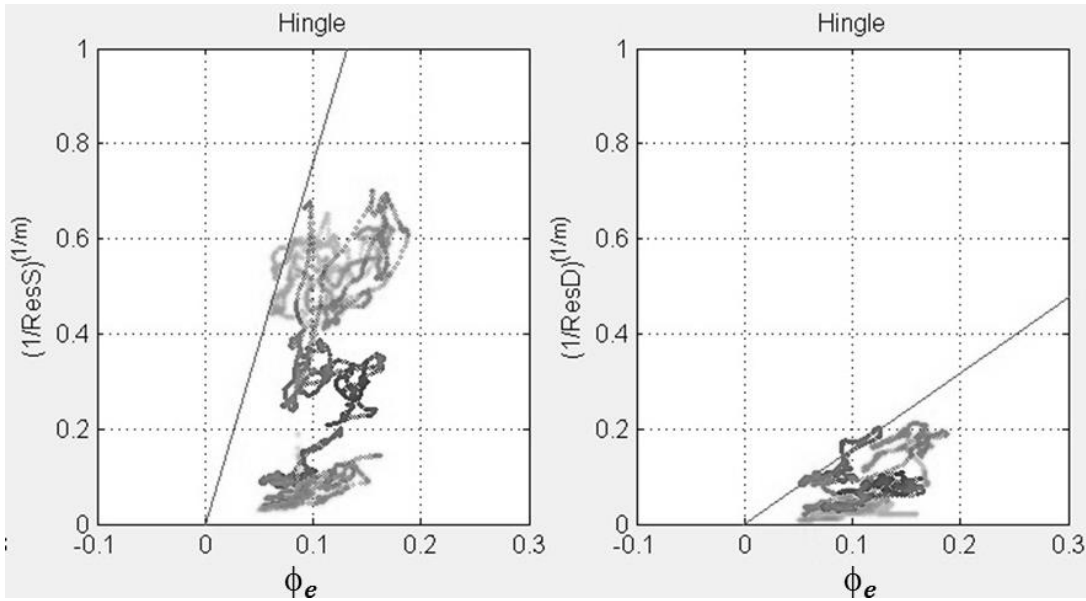


Figura 2. 10. Gráfica de Hingle.

Si se piensa la Ecuación (2.23) como la de una recta  $y = mx + b$ , donde la pendiente es igual al término  $\left[ \frac{S_{xo}^2}{R_{mf}} \right]^{1/2}$  y la ordenada al origen es 0.

La Ecuación (2.23) define un conjunto de rectas cuya pendiente depende del valor de  $S_{xo}$ , si  $R_{mf}$  se considera constante.

Se observa que la porosidad aumenta linealmente en el eje horizontal y la resistividad decrece de una manera no lineal de acuerdo a la expresión  $1/\sqrt{R_{xo}}$ .

El procedimiento es colocar en la gráfica los valores de  $\phi_e$  y  $R_{xo}$  para todos los niveles de interés en el intervalo. Se recomienda que el intervalo no sea muy grande. Aquellos puntos que caen en la zona de baja resistividad para cierta porosidad dada, representan niveles de 100% de agua. Una línea recta, que representa  $S_{xo} = 1$ , se dibuja desde el pivote ( $\phi_e = 0, R_{xo} = \infty$ ) siguiendo la tendencia de los puntos, lo más hacia el noroeste. A esta línea se le conoce también como  $R_o$ . La ecuación para la línea  $S_{xo} = 1$  de la Ecuación (2.23) es:

$$\left[\frac{1}{R_{xo}}\right]^{1/2} = \cdot \left[\frac{1}{R_{mf}}\right]^{1/2} \quad (2.24)$$

$R_{mf}$  se determina sustituyendo los valores de  $S_{xo}$  y  $R_{xo}$  para cualquier punto en la línea.

Las líneas que representan valores menores de  $S_{xo}$  se establecen como sigue. Para cierto valor de porosidad la Ecuación de Archie (2.22) puede escribirse como:

$$S_{xo} = \sqrt{R_{mf}/R_{xo}} \quad (2.25)$$

Hay que seleccionar valores convenientes de porosidad para que  $R_o$  sea fácil de identificar.<sup>16</sup>

Una vez hechas las líneas necesarias se puede estimar el valor de  $S_{xo}$  para cada intervalo.

### Gráfica de Pickett [30]

El *Método Gráfico de Pickett* es una solución gráfica a la ecuación de Archie, que consiste en una gráfica doble logarítmica de una medición de resistividad en el eje x vs una medición de porosidad en el eje y. La gráfica debe su nombre a G.R. Pickett y se basa en la obtención del logaritmo de la ecuación de Archie. Por medio de este método es posible estimar los valores de la resistividad del agua de formación,  $R_{mf}$ , y el exponente de cementación,  $m$ , de un intervalo. Ver Figura 2.11.

Sea la forma generalizada de la Ecuación de Archie [4] se tiene la expresión:

$$S_{xo}^n = \frac{a \cdot R_{mf}}{m \cdot R_{xo}} \quad (2.26)$$

Tomando logaritmos y despejando se tiene:

$$\log_{10} = -\frac{1}{m} \log_{10} R_{xo} + \frac{1}{m} \log_{10}(a \cdot R_{mf}) - \frac{n}{m} \log_{10} S_{xo} \quad (2.27)$$

Al igual que con Hingle, si se piensa la Ecuación (2.27) como la de una línea recta  $y = mx + b$ , donde:

$$y = \log_{10} \quad (2.28)$$

$$x = \log_{10} R_{xo} \quad (2.29)$$

$$\text{pendiente} = -\frac{1}{m} \quad (2.30)$$

<sup>16</sup> En esto radica lo subjetivo que es la aplicación del Método Gráfico de Hingle para calcular el valor de  $R_{mf}$

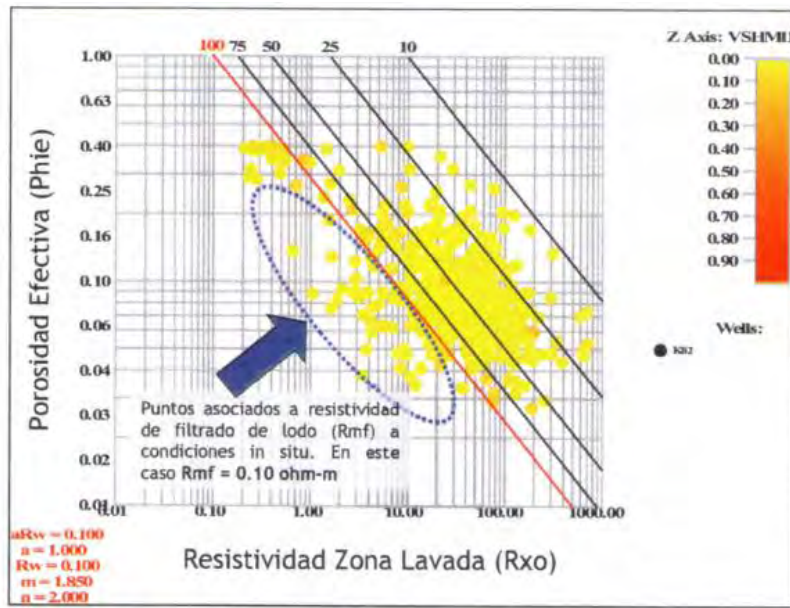


$$\text{y la ordenada al origen} = \frac{1}{m} \log_{10}(a \cdot R_{mf}) - \frac{n}{m} \log_{10} S_{xo}. \quad (2.31)$$

La Ecuación (2.27) muestra que si  $a$ ,  $R_{mf}$ ,  $n$ , y  $S_{xo}$  son constantes, generará una línea recta en una gráfica  $\log(R_{xo})$  vs  $\log(\ )$ .

En particular, se suele tomar como referencia la recta que se origina cuando  $S_{xo} = 1$ , que es el caso de 100% de Saturación de Agua, debido a que la ordenada al origen sólo queda en función de  $a$  y  $R_{mf}$ ,

$$\frac{1}{m} \log_{10}(a \cdot R_{mf}). \quad (2.32)$$



Gráfica Rxo-Phie para re-consideración de Rmf que entrara al cálculo de Sxo

Figura 2. 11. Identificación de recta, en el análisis gráfico de Pickett. [11]

Las zonas de agua en las gráficas como la de la Figura 2.11 se ubican en la parte inferior extrema de la gráfica, dirección NW-SE, donde se tienen los datos con porosidad y resistividad baja. El método consiste en realizar un gráfico log-log del registro resistivo contra la porosidad efectiva y ajustar el cúmulo de puntos a una recta con esta dirección. La recta que se ajusta es una propuesta que realiza el especialista y de preferencia se selecciona en dirección del eje mayor del cúmulo si se piensa este como una elipse alargada.

Por proponer una recta pegada al cúmulo de datos, pero en el área de menor porosidad y resistividad se considera que se está proponiendo la recta con  $S_{xo} = 1$ , por lo que una vez propuesta la recta se calcula sobre la gráfica el valor de la pendiente y la ordenada al origen y estos valores se sustituyen en las Ecuaciones (2.30) y (2.32) de los que

se puede despejar los valores del exponente de cementación  $m$  y la resistividad del agua de formación como:

$$m = -\frac{1}{\text{pendiente}}, \quad (2.33)$$

$$R_{mf} = \frac{1}{a} 10^{m \cdot \text{ordenada}}. \quad (2.34)$$

Una vez establecida la línea de agua  $S_{xo} = 1$ , pueden trazarse otras líneas paralelas para diferentes  $S_{xo}$ , asumiendo valores como  $n$  y  $a$  constantes. En la Figura 2.11 se ejemplifica la ubicación de las líneas  $S_{xo} = 0.75$ ,  $S_{xo} = 0.5$ ,  $S_{xo} = 0.25$  y  $S_{xo} = 0.1$ , en azul, con respecto a la línea,  $S_{xo} = 1$ , en rojo.

### Ecuación de Archie [4]

La *Ecuación de Archie* [4] es la relación entre el factor de formación, el registro de resistividad, la porosidad y la Saturación de Agua a un valor de profundidad constante. Esta relación parte de que se puede utilizar el registro de resistividad eléctrica para determinar las características de yacimiento, aunque su precisión depende de factores como la exactitud con la que se puede determinar la resistividad verdadera de formación, tener valores de porosidad no muy variados, el alcance de las medidas de resistividad para las características de la formación; la información disponible sobre la conductividad de las aguas innatas o de formación; el conocimiento geológico respecto a los cambios probables de facies dentro de los horizontes y los efectos resultantes sobre las propiedades eléctricas del yacimiento. La Ecuación de Archie se presenta como:

$$S_{xo} = \left( \frac{FR_{mf}}{R_{xo}} \right)^{1/n} = \left( \frac{a \cdot R_{mf}}{R_{xo} \cdot m} \right)^{1/n}, \quad (2.35)$$

donde  $S_{xo}$  representa la Saturación de Agua en Zona Lavada,  $F = \frac{a}{\phi^m}$  es el Factor de Formación,  $\phi$  es la porosidad,  $R_{mf}$  es la resistividad de lodo filtrado,  $R_{xo}$  es la resistividad en la zona lavada,  $m$  es el Exponente de Cementación y  $n$  es el Exponente de Saturación [11]. Aunque los valores de  $\phi$  y  $R_{xo}$  pueden ser obtenidos directamente de los registros, la determinación de  $a$ ,  $R_{mf}$ ,  $m$  y  $n$  requiere ser obtenida del análisis de núcleos, en casos donde no se cuenta con esta información es necesario deducir sus valores.

La Ecuación de Archie se aplica en formaciones limpias, requiere los registros de resistividad  $R_{xo}$  y porosidad  $\phi$ , se suele relacionar el valor de  $R_{xo}$  con algún registro de

resistividad de investigación somera<sup>17</sup> (MSFL). La porosidad que se utiliza es la del registro NPHI.

La saturación de hidrocarburos se calcula como

$$S_{or} = 1 - S_{xo}. \quad (2.36)$$

Existe un gran número de métodos para la interpretación de arenas arcillosas siendo actualmente los más usados el de Dos Aguas [21] y el de Waxman-Smits [40]. Se basan en Archie pero introducen un nuevo concepto, denominado CEC, Capacidad de Intercambio Catiónico.

## Resumen del Capítulo 2

En este capítulo se presentó la Metodología que se utiliza para estimar la Saturación de Agua por medio de técnicas gráficas y la aplicación de la Ecuación de Archie.

También se presentaron los conceptos necesarios para la comprensión del contexto geofísico del trabajo.

---

<sup>17</sup> Dentro de los Registros de Resistividad Laterales tenemos:

- a) MSFL = Microspheric Laterolog. Para las proximidades (1.0 y 6.0''). Lee la resistividad de la zona lavada (Rxo).
- b) MLL = LLM = Micro Laterolog. Para las proximidades (1.0 y 6.0'')
- c) SLL = LLS = Someric Laterolog. Para profundidades someras (0.5 y 1.5')
- d) DLL = LLD = Deep Laterolog. Para profundidades de más de 3.0'. Miden resistividad de la formación (Rt).

## Capítulo 3

### Teoría de redes bayesianas

El objetivo de este capítulo es definir los elementos de la Teoría de Probabilidad, Estadística y Redes Bayesianas necesarios para el desarrollo de este trabajo.

#### Conceptos Básicos de la Teoría de la Probabilidad [47]

Sea  $X = \{X_1, \dots, X_n\}$  un conjunto de variables aleatorias discretas y  $\{x_1, \dots, x_n\}$  el conjunto de sus posibles realizaciones, donde

$$p(x_1, \dots, x_n) = p(X_1 = x_1, \dots, X_n = x_n) \quad (3.1)$$

es la *función de probabilidad conjunta* de las variables de  $X$ .

La *función de probabilidad condicional* de la  $i$ -ésima variable  $X_i$  dado el subconjunto de variables  $\{X_1, \dots, X_k\}$  viene dada por

$$p(x_i | x_1, \dots, x_k) = \frac{p(x_i, x_1, \dots, x_k)}{p(x_1, \dots, x_k)}, \quad (3.2)$$

donde la suma se extiende a todos los valores posibles de  $X_i$ .

Sean  $X$  e  $Y$  dos subconjuntos disjuntos del conjunto de variables aleatorias  $\{X_1, \dots, X_n\}$ . Entonces se dice que  $X$  es independiente de  $Y$  si y solamente si

$$p(x | y) = p(x), \quad (3.3)$$

para todos los valores posibles  $x$  e  $y$  de  $X$  e  $Y$ ; en otro caso,  $X$  se dice dependiente de  $Y$ .

Las variables aleatorias  $\{X_1, \dots, X_n\}$  se dice que son independientes si y sólo si

$$\begin{aligned} p(x_1, x_2, \dots, x_n) &= p(x_1, \dots, x_k) p(x_{k+1}, \dots, x_n) = p(x_1) p(x_2) \dots p(x_n) \\ &= \prod_{i=1}^n p(x_i), \end{aligned} \quad (3.4)$$

para todos los valores posibles  $x_1, \dots, x_n$  de  $X_1, \dots, X_n$ . En otro caso, se dice que son dependientes

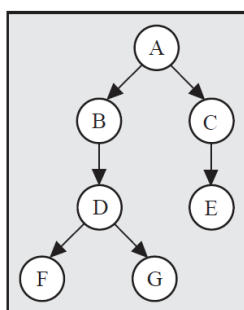
Utilizando la Ecuación (3.2), se obtiene

$$p(x_i | x_1, \dots, x_k) = \frac{p(x_i, x_1, \dots, x_k)}{\sum_{x_i} p(x_i, x_1, \dots, x_k)} = \frac{p(x_i)p(x_1, \dots, x_k | x_i)}{\sum_{x_i} p(x_i)p(x_1, \dots, x_k | x_i)} \quad (3.5)$$

La ecuación (3.5) se conoce como *Teorema de Bayes*.

## Conceptos sobre Grafos [47]

Supóngase un conjunto de variables  $X = \{X_1, X_2, \dots, X_n\}$  que pueden relacionarse entre sí. El conjunto  $X$  puede ser representado gráficamente por una colección de nodos o vértices, asociando un nodo a cada elemento de  $X$ . Estos nodos pueden conectarse por aristas, indicando las relaciones existentes entre los mismos. Por tanto, un *grafo* puede definirse de forma intuitiva mediante el conjunto de nodos,  $X$ , y las relaciones entre los mismos,  $L$ . En la Figura 3.1 se presenta un ejemplo de un grafo, donde los nodos están representados por círculos y las aristas por líneas que unen los nodos correspondientes.



**Figura 3. 1. Ejemplo de un grafo o red. Los  $X$  nodos se indican como los círculos A, B, C, D, E, F y G. Las relaciones  $L$  entre los nodos como las flechas que lo conectan.**

Un *grafo* es un par de conjuntos  $G = (X, L)$ , donde  $X = \{X_1, X_2, \dots, X_n\}$  es un conjunto finito de elementos (nodos), y  $L$  es un conjunto de aristas, es decir, un subconjunto de pares ordenados de elementos distintos de  $X$ . Los términos grafo y red se emplean como sinónimos en este libro.

Dado un grafo  $G = (X, L)$ , si  $L_{ij} \in L$  y  $L_{ji} \notin L$ , la arista  $L_{ij}$  entre los nodos  $X_i$  y  $X_j$  se denomina *dirigida* y se denota mediante  $X_i \rightarrow X_j$ .

Dado un grafo  $G = (X, L)$ , si  $L_{ij} \in L$  y  $L_{ji} \in L$ , la arista  $L_{ij}$  se denomina *no dirigida* y se denota mediante  $X_i - X_j$  o  $X_j - X_i$ .

Un grafo en el cual todas las aristas son dirigidas se denomina grafo dirigido, y un grafo en el que todas sus aristas son no dirigidas se denomina no dirigido.

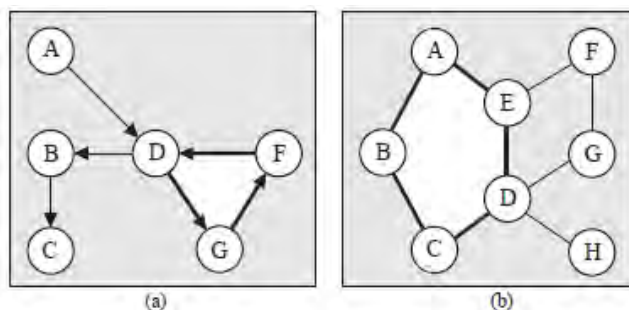


Figura 3. 2. Ejemplos de un grafo dirigido (a), y uno no dirigido (b).

Un camino del nodo  $X_i$  al nodo  $X_j$  es una sucesión de nodos  $(X_{i1}, \dots, X_{ir})$ , comenzando en  $X_{i1} = X_i$  y finalizando en  $X_{ir} = X_j$ , de forma que existe una arista del nodo  $X_{ik}$  al nodo  $X_{i_{k+1}}$ ,  $k = 1 \dots r - 1$ . La longitud del camino,  $(r - 1)$ , se define como el número de aristas que contiene.

Un camino  $(X_{i1}, \dots, X_{ir})$  se dice que es cerrado si el nodo inicial coincide con el final, es decir,  $X_{i1} = X_{ir}$ . En la Figura 3.2 se presentan algunos ejemplos de caminos cerrados.

Un grafo no dirigido se denomina completo si contiene una arista entre cada par de nodos. Por tanto, existe un único grafo completo de  $n$  nodos.

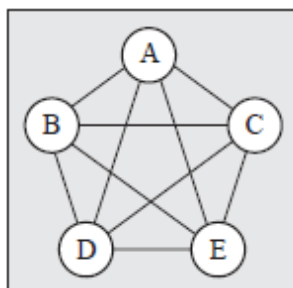


Figura 3. 3. Grafo completo de cinco nodos.

Existen grafos que representan relaciones de dependencia e independencia [28], en donde cada nodo en el grafo representa una variable. En el caso de los mapas de dependencia se asegura que, si dos nodos están conectados, existe una relación de dependencia entre las variables a las que representan. Los mapas de independencia aseguran que si dos nodos no están conectados, entonces las variables a las que representan son independientes.

Los arcos dirigidos representan una dependencia causal y los no dirigidos representan una dependencia contemporánea. Los grafos dirigidos representan comúnmente relaciones de causalidad, por lo que en ocasiones son llamados grafos causales. [12]

En los modelos de grafos que utilizan arcos dirigidos se hacen necesarias algunas definiciones:

- Un *vértice*  $A$  es un Padre de un vértice  $B$  y  $B$  es un hijo de  $A$  si y solo si existe un arco dirigido de  $A$  a  $B$ .
- Un *ancestro* de un vértice  $V$  es cualquier vértice  $W$  tal que existe un camino dirigido de  $W$  a  $V$ .
- Un *descendiente* de un vértice  $V$  es cualquier vértice  $W$  tal que existe un camino dirigido de  $V$  a  $W$ .
- $Padres(V)$  es el conjunto de padres del vértice  $V$ .
- $Hijos(V)$  es el conjunto de hijos del vértice  $V$ .
- $Ancestros(V)$  es el conjunto de ancestros del vértice  $V$ .
- $Descendientes(V)$  es el conjunto de descendientes del vértice  $V$ .

Otras definiciones importantes son aquellas que se refieren a los nodos de aristas convergentes. Para grafos dirigidos, se tienen las siguientes definiciones [33]:

- Un vértice  $V$  en un grafo  $G$  es un nodo de aristas convergentes (colíder) en un camino no dirigido  $\pi$  si y solo si existen dos distintos arcos en  $\pi$  incidentes a  $V$ .
- Un vértice  $V$  en un grafo  $G$  es un nodo de aristas convergentes sin blindaje en  $\pi$  si  $V$  es un nodo de aristas convergentes en  $\pi$ ,  $V$  es adyacente a dos vértices  $V_1$  y  $V_2$ ,  $V_1 \neq V_2$ , y  $V_1$  y  $V_2$  no son adyacentes en  $G$ .

Considerando estas definiciones:

La **Red Bayesiana o Red Probabilística** es un grafo acíclico<sup>18</sup> dirigido en el cual cada nodo representa una variable y cada arco una dependencia probabilística, en la cual se especifica la probabilidad condicional de cada variable dados sus padres. Al contar con un modelo de este tipo es posible conocer, de manera cuantitativa, el efecto que tiene sobre el entorno la modificación de una variable.

---

<sup>18</sup> Acíclico. Sin caminos cerrados.

La estructura de una Red Bayesiana se refiere al grafo acíclico dirigido que representa las relaciones de dependencia/independencia entre las variables. Los parámetros se refieren a las distribuciones de probabilidad condicional que se encuentran en los nodos de la Red Bayesiana. Es importante recordar que, dado que la distribución de probabilidad condicional para cada nodo depende de los padres del mismo, la estimación de los parámetros de una Red Bayesiana requiere una estructura definida.

Los modelos de Red Bayesiana tienen la ventaja de ser ampliamente intuitivos. Tanto la estructura como las distribuciones son fácilmente comprensibles.

Las redes bayesianas describen de manera compacta la distribución de probabilidad que gobierna a un conjunto de variables, especificando suposiciones de independencia condicional entre las mismas, así como distribuciones de probabilidad condicional que expresan de manera cuantitativa las relaciones entre ellas. Estas relaciones son probabilísticas y aunque para su identificación se pueden utilizar relaciones físicas conocidas, estas se pueden deducir de las muestra de datos a modelar con la red.

Las redes bayesianas se fundamentan en la teoría de la probabilidad y del teorema de Bayes para construir un grafo o red, que permite representar un modelo causal de las relaciones de dependencia o independencia de las variables que forman parte del dominio de la aplicación [29].

Existen condiciones que permiten caracterizar a un modelo de grafo. Una de las condiciones básicas para estos modelos es la condición de Markov. Para grafos dirigidos, la condición de Markov se puede expresar como sigue [33]:

Un grafo acíclico dirigido  $G$  sobre  $V$  con una **distribución de probabilidad**  $P(V)$  satisface la *condición de Markov* si y solo si, para todo  $W \in V$ ,  $W$  es independiente de  $V \setminus (\text{Descendientes}(W) \cup \text{Padres}(W))$  dado  $\text{Padres}(W)$ .

En base a la Condición de Markov es posible definir la Condición de Minimalidad. Esta condición identifica a aquellos grafos que tienen el mínimo número de arcos para cumplir la condición de Markov, eliminando así los casos triviales:

Sea  $G$  un grafo acíclico dirigido sobre  $V$ , y  $P$  una distribución de probabilidad sobre  $V$ , Decimos que  $\langle G, P \rangle$  satisface la *Condición de Minimalidad* si y solo sí para cada subgrafo propio  $H$  de  $G$  con vértices  $V$ ,  $\langle H, P \rangle$  no satisface la Condición de Markov.

Además de las relaciones de independencia exigidas por la condición de Markov, existen otras que se pueden expresar en un grafo. La d-separación es una de las relaciones de independencia más importantes en la representación de distribuciones de probabilidad con



grafos. Básicamente, la d-separación obedece a la independencia de dos variables dada una tercera en la distribución de probabilidad representada por el grafo.

La condición que permite saber si una distribución es representada por un grafo es la condición de fidelidad:

Sea  $G$  un grafo causal y  $P$  una distribución de probabilidad generada por  $G$ .  $\langle G, P \rangle$  satisface la *Condición de Fidelidad* si y solo si toda relación de independencia condicional presente en  $P$  es implicada por la Condición de Markov aplicada a  $G$ .

Una distribución  $P$  es fiel a  $G$  si y solo si satisface tanto la Condición de Markov como la Condición de Fidelidad. Una condición de fidelidad más estricta es la condición de fidelidad normal para grafos acíclicos dirigidos [7]:

Cuando se estudian los modelos con grafos y se desean expresar relaciones de causalidad, se utiliza la causalidad de Granger [15]. La causalidad de Granger establece que una serie de tiempo es causal de otra si la predicción de esta última se ve afectada al tomar en cuenta la información contenida en la primera serie. Formalmente se define como sigue:

*Causalidad de Granger.* Considérese una distribución condicional respecto a dos conjuntos de información disponibles en el tiempo  $t$ , digamos  $I_t$  e  $I_t^* = \{I_t, x_t, x_{t-1}, \dots\}$ , en donde  $x_t$  denota a una variable posiblemente causal. La variable  $x_t$  es definida como causal de la variable  $y_t$  en el sentido de Granger si existe un  $h \in \{1, 2, \dots\}$  tal que:

$$E(y_{t+h}|I_t) \neq E(y_{t+h}|I_t^*)$$

En la gran parte de la literatura sobre modelos de grafos [9], el símbolo  $\perp$  representa ortogonalidad (no correlación) entre variables. Asimismo,  $A \perp B | Z$  significa que  $A$  y  $B$  son ortogonales una vez que se eliminan los efectos lineales de  $Z$ .

Si  $V$  representa un conjunto de variables y  $A \subseteq V$ , entonces  $X_A = \{X_A(t)\}$  representa un proceso multivariado dado por los las variables en  $A$ , y  $\overline{X}_A(t) = \{X_A(s), s < t\}$  denota el pasado del subproceso  $X_A$ .

## Extracción de parámetros

Cuando se tienen datos completos y suficientes para todas las variables en el modelo, es posible obtener los parámetros, asumiendo que la estructura está dada. El método más común es el llamado estimador de máxima verosimilitud (MLE, por sus siglas en inglés), bajo el cual se estiman las probabilidades en base a las frecuencias de los datos. [32]

MLE asume que los parámetros desconocidos  $\theta$  son constantes o que no existe información previa sobre ellos. Este estimador presenta propiedades asintóticas, especialmente cuando el número de muestras es grande.

La estimación  $\theta_{MLE}$  de los parámetros  $\theta$  se elige de modo que maximice la función de probabilidad

$$p(x_n|\theta) = p(x(1), x(2), \dots, x(n)|\theta), \quad (3.6)$$

de las mediciones  $X_n = \{x(1), x(2), \dots, x(n)\}$ . El valor obtenido para  $\theta_{MLE}$  es aquel que maximiza la probabilidad de los valores  $x(1), x(2), \dots, x(n)$ .

La construcción de la función de estimación para MLE puede ser muy complicada si las mediciones son dependientes. Por lo tanto, es usual asumir que las mediciones  $x(j)$  son independientes entre sí, con lo que se obtiene una expresión del tipo

$$p(X_n|\theta) = \prod_{j=1}^n p(x(j)|\theta) \quad (3.7)$$

A partir de la estadística se puede obtener una estimación MLE. La densidad de probabilidad *a priori* de los parámetros  $\theta$ , denotada  $p_\theta(\theta)$ , representa una densidad de probabilidad conocida o asumida antes de aplicar el método. Es decir,  $p_\theta(\theta)$  representa conocimiento previo sobre la distribución  $\theta$ . Similarmente,  $p_{\theta|X}(\theta|X_n)$  denota la densidad posterior de los parámetros  $\theta$  dadas las mediciones  $X_n$ , es decir, la densidad de probabilidad calculada por el estimador.

Es posible aplicar el mismo principio de MLE a la estimación bayesiana, para estimar la hipótesis más probable (MAP, *máximum a posteriori hypotesis*), denotado por  $\theta_{MAP}$ . El estimador MAP se define como el valor del parámetro  $\theta$  que maximiza la densidad posterior  $p_{\theta|X}(\theta|X_n)$  de  $\theta$  dadas las mediciones  $X_n$ . El estimador MAP puede ser interpretado como el valor más probable de los parámetros  $\theta$  dados los datos  $X_n$ .

Dado un valor para  $\theta$ , la densidad posterior puede ser calculada utilizando la Regla de Bayes:

$$p_{\theta|X}(\theta|X_n) = \frac{p_{X|\theta}(X_n|\theta)p_\theta(\theta)}{p_X(X_n)} \quad (3.8)$$

El denominador  $p_X(X_n)$  de la ecuación anterior es la densidad de probabilidad de los datos  $X_n$ , la cual no depende de los parámetros  $\theta$  y solamente sirve para normalizar la densidad posterior, de modo que para encontrar el estimador MAP basta con encontrar el valor de  $\theta$  que maximiza el numerador. Dicho valor de  $\theta$  es la densidad conjunta:

$$p_{\theta, X}(\theta|X_n) = p_{X|\theta}(X_n|\theta)p_\theta(\theta) \quad (3.9)$$

Si la distribución  $p_\theta(\theta)$  es uniforme, entonces MLE y MAP se convierten en el mismo estimador.

Sea  $\theta$  la densidad de probabilidad de un proceso estocástico, y sea  $D = \{x_1, x_2, \dots, x_n\}$  un conjunto de datos generados por dicho proceso. La **función de verosimilitud** de la distribución de probabilidad  $\theta$  dado un conjunto de datos D es igual a la probabilidad de los datos dada la distribución, y se denota:

$$L(\theta: D) = p(D|\theta) \quad (3.10)$$

Una estadística suficiente es una función que sintetiza, a partir de los datos, la información relevante para el cálculo de probabilidad. Así, la función de verosimilitud se puede calcular a partir de una estadística suficiente. [20]

La obtención de reglas permite comprender mejor cualquier fenómeno en estudio. Entiéndase por regla la abstracción de una estructura, generalmente oculta, dentro de una serie de tiempo o un conjunto de series de tiempo.

Una de las formas más usuales para expresar reglas de relación entre variables heterogéneas es llevar dichas relaciones al espacio de probabilidad. En este sentido, el conjunto de variables  $X_1, X_2, \dots, X_N$ , se convierte en un conjunto de variables aleatorias, de las cuales es posible extraer la función de densidad de probabilidad conjunta:

$$f(X_1, X_2, \dots, X_N): X_1 \times X_2 \times \dots \times X_N \rightarrow [0, \dots, 1] \quad (3.11)$$

De esta forma, no importa el tipo de las variables con las que se trabaje. Cada variable toma valores en su propio espacio, independiente del espacio de otras variables. Sin embargo, esta representación involucra una restricción de tipo computacional debido a que es difícil representar funciones continuas cuyo comportamiento no es previamente conocido. Debido a esto, es común discretizar las variables continuas a fin de que la función de densidad de probabilidad conjunta se convierta en una función discreta, en donde cada una de las variables toma  $r_i$  valores  $x_{i1}, x_{i2}, \dots, x_{i(r_i)}, i = 1, \dots, N$

$$\begin{aligned} Pr(X_1 = x_{11}, X_2 = x_{21}, \dots, X_N = x_{N1}) &= p_1 \\ Pr(X_1 = x_{12}, X_2 = x_{22}, \dots, X_N = x_{N2}) &= p_2 \\ &\vdots \\ Pr(X_1 = x_{1r_1}, X_2 = x_{2r_2}, \dots, X_N = x_{Nr_N}) &= p_M \end{aligned} \quad (3.12)$$

El tamaño del dominio de cada variable es independiente del resto de las variables. El número de valores de probabilidad necesarios para especificar la densidad conjunta de probabilidad es:

$$M = \prod_{i=1}^N r_i \quad (3.13)$$

Como se observa, si la cardinalidad del espacio de probabilidad ( $r_i$ ) es la misma para todas las variables, el número de valores de probabilidad crece exponencialmente con el número de variables.

La información contenida en la densidad conjunta de probabilidad puede representarse también en términos de probabilidad condicional. Sea el conjunto  $I = \{X_{I_1}, X_{I_2}, \dots, X_{I_K}\}$ , cada entrada de la densidad conjunta de probabilidad se puede describir como:

$$\begin{aligned} \Pr(X_1, X_2, \dots, X_N, X_{I_1}, X_{I_2}, \dots, X_{I_K}) \\ = \Pr(X_1, X_2, \dots, X_N | X_{I_1}, X_{I_2}, \dots, X_{I_K}) \cdot \Pr(X_{I_1}, X_{I_2}, \dots, X_{I_K}) \end{aligned} \quad (3.14)$$

Si las variables contenidas en el conjunto  $I$  son independientes del resto de las variables, la probabilidad conjunta se puede separar en dos factores:

$$\Pr(X_1, X_2, \dots, X_N, X_{I_1}, X_{I_2}, \dots, X_{I_K}) = \Pr(X_1, X_2, \dots, X_N) \cdot \Pr(X_{I_1}, X_{I_2}, \dots, X_{I_K}) \quad (3.15)$$

De este modo, el número de valores de probabilidad necesarios para especificar la distribución de probabilidad conjunta se reduce a:

$$M = \prod_{X_i \in I} r_i + \prod_{X_i \notin I} r_i \quad (3.16)$$

Aplicando este razonamiento sobre las variables contenidas en  $I$  y en su complemento  $I^c$ , se pueden formar conjuntos cada vez más pequeños hasta que todas las variables contenidas en los conjuntos sean dependientes entre sí.

Sin embargo, en la distribución conjunta de probabilidad existe gran redundancia debido a que no se toma en cuenta la independencia entre variables. Esto ha motivado el desarrollo de modelos gráficos que expresan relaciones de dependencia e independencia entre las variables involucradas.

## Recuperación de la serie de tiempo

Para recuperar una serie de tiempo<sup>19</sup> discretizada  $\hat{X}_R$  a partir de una Red Bayesiana, es necesario muestrear el valor de dicha serie sobre la densidad marginal de probabilidad  $f_R$ , asociada a  $X_R$ . Si  $\text{Ancestros}(X_R) \neq \emptyset$  o  $\text{Descendientes}(X_R) \neq \emptyset$  en la Red Bayesiana, entonces  $f_R$  depende del valor de un conjunto de  $k$  variables  $P = \{X_1, X_2, \dots, X_k\}$  en la red.

Sea  $S \subseteq P$ , es obvio que  $f_R$  se verá alterada al asignar distintos valores a cada una de las variables contenidas en  $S$ . Sean  $\{x_{i1}, x_{i2}, \dots, x_{iw}\}$   $w$  valores asignados a la variable

---

<sup>19</sup> Serie de tiempo se refiere a datos estadísticos que se recopilan, observan o registran en intervalos de tiempo regulares.

$X_i \in S$ , la densidad de probabilidad  $f_R$  es una función de la posición  $t$  ( $1 \leq t \leq w$ ), por lo que puede escribirse como  $f_R(t)$ . Para un valor dado de  $t$   $f_R(t)$  puede expresarse como:

$$\begin{aligned} p(X_R = x_1 \mid X_i = x_{it} \dots X_j = x_{jt}) &= p_{1 \ t} \\ p(X_R = x_2 \mid X_i = x_{it} \dots X_j = x_{jt}) &= p_{2 \ t} \\ &\vdots \\ p(X_R = x_{r_R} \mid X_i = x_{it} \dots X_j = x_{jt}) &= p_{r_R \ t} \end{aligned} \quad (3.17)$$

en donde  $r_R$  es el número de valores que puede tomar la variable  $X_R$ . Al muestrear sobre  $f_R(t)$  se obtiene un valor estimado  $\hat{X}_R(t)$  para la variable  $X_R$  en la posición  $t$ . Si se muestrea para una secuencia de valores de  $t$  entre 1 y  $w$ , se obtiene una estimación  $\hat{X}_R$  de la serie de tiempo discretizada  $X_R$ :

$$\hat{X}_R = \hat{x}_R(1) \ \hat{x}_R(2) \ \dots \ \hat{x}_R(t) \quad (3.18)$$

Dado que cada valor  $\hat{X}_R(t)$  ha sido tomado como una muestra aleatoria, la estimación  $\hat{X}_R$  representa un resultado entre varios posibles. En consecuencia, para cada posición  $t$  se pueden tomar  $u$  muestras  $x_R^1(t) \ x_R^2(t) \ \dots \ x_R^u(t)$ , obteniéndose  $u$  posibles estimaciones para cada posición. Al aplicar el mismo razonamiento para cada posición  $t$  se obtienen  $u$  estimaciones para la serie de tiempo discretizada:

$$\begin{aligned} \hat{X}_R^1 &= (x_R^1(1) \ x_R^1(2) \ \dots \ x_R^1(w)) \\ \hat{X}_R^2 &= (x_R^2(1) \ x_R^2(2) \ \dots \ x_R^2(w)) \\ &\vdots \\ \hat{X}_R^u &= (x_R^u(1) \ x_R^u(2) \ \dots \ x_R^u(w)) \end{aligned} \quad (3.19)$$

Al recuperar la serie continua  $\hat{X}_R^i$  a partir de los valores discretos de la estimación  $\hat{X}_R^i$ , se obtiene una trayectoria posible para la serie de tiempo. Es decir, debido a que durante la discretización de la serie de tiempo se pudo tomar en cuenta la variación de cada punto respecto al anterior, la variación en un valor discreto  $\hat{x}_R^i(t)$  podría afectar la recuperación de los valores continuos posteriores  $\hat{x}_R^i(t) \ \hat{x}_R^i(t+1) \ \dots \ \hat{x}_R^i(w)$ .

La distribución de los valores continuos recuperados en un punto  $t$ ,  $\hat{x}_R^1(t) \ \hat{x}_R^2(t) \ \dots \ \hat{x}_R^u(t)$  no puede asumirse como normal debido a que se desconocen, entre otros, la naturaleza de la serie de tiempo y el valor de  $\sigma$  utilizado durante la discretización. Este hecho descalifica el uso de la media de los valores  $\hat{x}_R^1(t) \ \hat{x}_R^2(t) \ \dots \ \hat{x}_R^u(t)$  para la estimación de un valor final  $\hat{X}_R(t)$  para la posición  $t$ .

Inicialmente, se pensó en utilizar la mediana de los valores  $\hat{x}_R^1(t) \ \hat{x}_R^2(t) \ \dots \ \hat{x}_R^u(t)$  como el valor  $\hat{X}_R(t)$  para cada posición debido a su robustez ante la presencia de valores extremos (outliers). No obstante, se observó que es necesario tomar en cuenta la densidad de

las trayectorias para los valores dados, es decir, tomar un valor cercano a aquel valor en torno al cual se agrupan la mayoría de los valores  $\hat{x}_R^i(t)$ .

### **Resumen de Capítulo 3**

En este capítulo se presentaron los conceptos de probabilidad y Redes Bayesianas que se utilizaron para el planteamiento de la metodología propuesta en este trabajo de tesis.



## Capítulo 4

### Planteamiento del problema

El objetivo de este capítulo es mostrar la problemática que se presenta en el análisis de pozos para realizar la inferencia de Saturación de Agua. En la primera parte se presenta un análisis sobre lo que ya se ha publicado respecto al tema, para finalizar con la enunciación de la problemática y de la hipótesis que se pretende responder por medio de este trabajo.

#### 4.1. Generalidades

Tradicionalmente, el método para predecir la Saturación de Agua en función de las propiedades de las rocas y la profundidad de contacto han caído en dos grupos, los basados en la curva de presión capilar [22] y los métodos basados en registros. Los métodos basados en registros son los concernientes a resolver la Ecuación de Archie [4] o alguna de sus variantes más usadas en la industria como Dual-Water [21] o Waxman-Smits [40].

Determinar los valores del Exponente de Cementación  $m$ , el Exponente de Saturación  $n$  y la Resistividad de Lodo Filtrado  $R_{mf}$  resulta en una tarea fundamental para el contexto geológico que se presenta. En particular  $m$  da información de la consolidación de la roca y de la medida de tortuosidad de la geometría del poro de flujo de corriente y garantiza resultados fiables para los cálculos de la Saturación [19].

Existen varias metodologías que se utilizan para determinar los valores de  $m$ ,  $n$  y  $R_{mf}$ , entre las que sobresale el uso de los gráficos de Pickett [30], que sirve además en los casos donde se tienen pocos datos. Estos gráficos son el de Porosidad Aparente vs Resistividad para identificar el valor de  $m$ , el de Resistividad de Agua Aparente ( $R_{wa}$ ) vs Profundidad de pozo para estimar el valor de  $R_{mf}$  y el  $S_o$  vs  $R_o$  para determinar la Saturación de Agua. Estos métodos presentan algunas restricciones como que no es conveniente su aplicación cuando se desconoce tanto  $m$  como  $R_{mf}$  o cuando se tienen intervalos cortos del mismo tipo de roca. Los mejores resultados en términos de ajuste de datos se obtienen si el Exponente de Saturación,  $n$ , es fijado como „1“ [13]. Tanto Archie



como Pickett enfatizan la importancia de utilizar toda la información disponible y las restricciones que se pueden presentar en la aplicación de los métodos.

La justificación de utilizar estos métodos „tradicionales“ para la estimación de la Saturación de Agua radica en que han mostrado tener muy buena relación con otras metodologías más rigurosas como la monitorización de la saturación in-situ de los núcleos, que es la medida de atenuación de rayos X de 1-2 unidades de saturación en cortes continuos de 2mm debajo de la longitud del núcleo durante intervalos de tiempo de aproximadamente 30 segundos [34].

En conclusión, sobre las metodologías como Archie-Pickett, se tienen muy buenos resultados de los valores de Saturación de Agua estimados que en el mejor de los casos coinciden con los medidos directamente de la roca, pero se presentan problemas de confiabilidad en casos de no tener el análisis de núcleos o tener registros deficientes. El uso de métodos gráficos suele estar sujeto a errores de interpretación visual. [6]

## **Introducción de los Modelos de Machine Learning en la Caracterización de Yacimientos**

En trabajos posteriores se introdujeron los modelos de redes neuronales artificiales (ANN) como un nuevo paradigma que diera respuesta a los problemas de falta de información para realizar los cálculos de la metodología clásica (Archie-Pickett) [24].

Los modelos de ANN fueron propuestos por ser modelos matemáticos flexibles que podían aprender<sup>20</sup> relaciones complejas entre muchas variables e inferir algunas conclusiones, sin conocimiento explícito de las propiedades del fluido y la roca como es requerido de manera convencional [36], pero lo más importante es que este tipo de trabajos abrió un área de aplicación para las técnicas de Inteligencia Artificial en la petrofísica [24].

Los primeros trabajos que se desarrollaron en este sentido eran modelos de redes neuronales de tipo perceptrón multicapa (MLP) que predecían el valor de Saturación de Agua [25] [22] [26], la saturación de hidrocarburos [25], la macro-porosidad [31], la porosidad de fractura [29] o la Saturación de gas directamente de los registros de pozo, dando aproximaciones aceptables con exactitud del 90% [31]. Siguiendo esta línea de trabajo se han presentado muchas mejoras a los modelos de Redes Neuronales aplicados a la Estimación de la Saturación del Agua: ANN de Bases Radiales [25], ANN que integran aprendizajes con Algoritmos de Competencia en la contribución óptima (ICA) [6], ANN con aprendizaje incremental de base-árbol (LOLIMOT) [19], ANN que realiza gestión de datos para la selección del conjunto más óptimo de aprendizaje: clasificación difusa (FR) [35],

---

<sup>20</sup> Aprender en el contexto de aprendizaje automático o aprendizaje de máquina (en inglés, “Machine Learning”) se trata de crear programas capaces de generalizar comportamientos a partir de una información suministrada en forma de ejemplos. [48]

Modelo Neuro Difuso Lineal Local [3]. Además de que los modelos de Redes Neuronales han servido también para producir de forma sintética los parámetros necesarios para la resolución de la Ecuación de Archie o realizar el reconocimiento de patrones en los Gráficos de Pickett [30].

En todos los modelos de ANN utilizados las variables de entrada para el modelo son los registros de línea fija: densidad, neutrón, resistividad, photo-electrico y rayos gamma, además de la Saturación de Agua como parámetro objetivo [3, 6, 36]. Esto es debido a que estos datos son los que están disponibles de manera más abundante en la mayoría de los pozos y proporcionan valiosa, pero indirecta, información sobre las propiedades de la roca [26]. En algunos modelos se presentan algunas variantes respecto a los parámetros que se utilizan como entrada en la red neuronal, donde se cambia el conjunto mencionado por la porosidad, la permeabilidad ( $k$ ) y la profundidad por encima del nivel de agua libre [22], la litología [25].

La comprobación de los resultados obtenidos por medio de ANN suele ser la comparación entre la curva obtenida por ANN y la obtenida por alguno de los métodos clásicos, dando Errores de alrededor del 0.1% [36, 3, 17]. Adicionalmente suele calcularse el error cuadrático medio (RMSE), que físicamente para el problema de la Saturación de Agua representa una fracción de volumen de poros P.V. [25].

Estos cálculos también suelen ser usados para identificar tipos de roca [26], cambios litológicos [1] o la simulación de yacimientos [36].

Aunque los modelos de ANN han demostrado ser superiores a los métodos estadísticos convencionales, tales como la regresión lineal múltiple [25], se presentan algunas inconformidades respecto a la construcción de estos modelos tales como que la arquitectura de la red neuronal suele ser resultado de múltiples experimentos numéricos, estas arquitecturas no son replicables y dependen completamente de la base de datos que se haya utilizado para el entrenamiento, las estrategias para mejorar la precisión de los resultados están basadas en el sobre-entrenamiento, es difícil realizar una conexión entre las relaciones internas de la red con las relaciones físicas conocidas [36].

## 4.2. Problema

La estimación de la Saturación de Agua en los pozos es uno de los cálculos más importantes en la caracterización de yacimientos. Para estimar este valor se resuelve la Ecuación de Archie. Resolver esta ecuación implica conocer el valor de  $m, n$  y  $R_{mf}$  y contar con los registros de densidad, resistividad, rayos gamma y porosidad neutrón. Si no se tiene la información necesaria suele recurrirse a métodos gráficos como Pickett. Los métodos gráficos tienen restricciones como dar resultados subjetivos y no ser conveniente su uso cuando se desconoce simultáneamente  $m$  y  $R_{mf}$ .

En casos extremos donde no se puede estimar  $m, n$  y  $R_{mf}$  se ha recurrido a modelos de redes neuronales artificiales (ANN), que ya no requieren información de la roca o del fluido para estimar el valor de la Saturación de Agua.

Las inconformidades que presentan con los modelos de ANN para estimar la Saturación de Agua son que para disminuir el error cuadrático medio muchas veces suele recurrirse al sobre-entrenamiento<sup>21</sup>, la estructura de red suele ser resultado de experimentos numéricos por lo cual no es replicable y no suele existir conexión entre las relaciones internas de la estructura de la red neuronal con las relaciones físicas observadas como la relación entre porosidad y resistividad del fluido exhibidos en la Ecuación de Archie.

Es posible plantear un modelo de Inteligencia Artificial que reconcilie la estructura interna del modelo con las relaciones físicas conocidas.

### 4.3. Hipótesis

El problema a tratar es, si es posible plantear un modelo que a partir de las relaciones probabilísticas que existen entre los registros se pueda utilizar para inferir el valor de la Saturación de Agua. En particular en este trabajo la propuesta que se plantea es la del modelo de una Red Bayesiana.

---

<sup>21</sup> El sobre-entrenamiento es la tendencia que tienen la mayoría de los algoritmos de Machine Learning a ajustarse a unas características muy específicas de los datos de entrenamiento que no tienen relación causal con la función objetivo que se está buscando para generalizar. El ejemplo más extremo de un modelo sobreenentrenado es un modelo que solo memoriza las respuestas correctas; este modelo al ser utilizado con datos que nunca antes ha visto va a tener un rendimiento azaroso, ya que nunca logró generalizar un patrón para predecir. [48]

## Capítulo 5

# Metodología de construcción de Modelo de Red Bayesiana

El objetivo de este capítulo es presentar la obtención de un modelo de grafo para estimar Saturación de Agua a partir de un conjunto de registros de pozo. El modelo construido expresa cada registro como un nodo y las relaciones directas entre estos como arcos dirigidos en el grafo. Además de presentarse un esquema gráfico de la relación entre registros, el modelo mantiene información acerca del comportamiento de cada relación en forma de densidad de probabilidad condicional. Este tipo de modelos se conocen como Red Bayesiana.

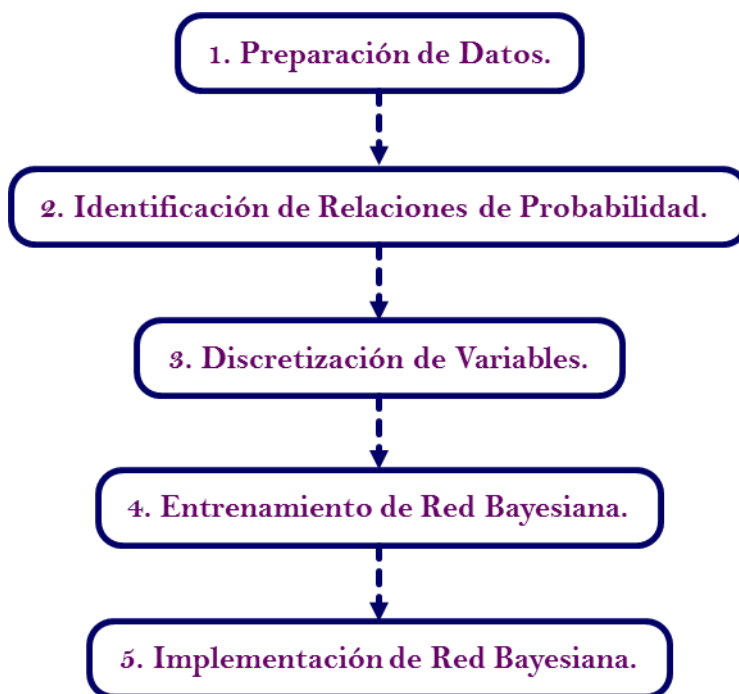


Figura 5. 1. Secuencia de pasos para la construcción del Modelo de Red Bayesiana.

Los datos de los registros son representados tomando en cuenta tanto la amplitud como las variaciones entre cada punto subsecuente, amplitud y cambio. Este método introduce una menor cantidad de ruido para la predicción de variables.

La Red Bayesiana fue entonces extraída a partir de la secuencias de los registros, utilizando algoritmos de aprendizaje.

Con este trabajo se potencia una de las características más importantes de las redes Bayesianas, su utilidad para fines de evaluación y predicción. Para realizar predicción es necesario únicamente sustituir los valores de los registros para un valor de profundidad dado, con el objetivo de obtener la densidad de probabilidad en el nodo relacionado a Saturación de Agua y muestrear sobre este. Con estas predicciones es posible estimar un error para cada punto.

El **Aprendizaje Bayesiano** provee un método para inferencia. Se basa en la consideración de que las variables están gobernadas por distribuciones de probabilidad, y de que es posible tomar decisiones óptimas en base al razonamiento sobre tales probabilidades unidas a los datos observados. [17]

En la aplicación planteada que es la construcción de una Red Bayesiana para inferir Saturación de Agua, es necesario realizar trabajo de minería de datos<sup>22</sup> sobre registros de pozo para extraer la información de las relaciones estadísticas que puedan existir.

Las relaciones estadísticas permiten identificar el conjunto de reglas que después serán aplicadas para la construcción de un grafo dirigido.

En la Figura 5.1 se presenta un esquema con los pasos seguidos en este trabajo para la construcción y aplicación del modelo de red bayesiana.

En la primera parte „Preparación de Datos“ se presenta la información estadística de los campos, pozos y registros utilizados.

## 5.1. Preparación de datos

Para todos los estudios realizados en esta investigación se utilizó una base de datos conformada por 19 pozos. Ver Tabla 1.1. Para la Metodología de Entrenamiento y construcción de las Redes Bayesianas se utilizó exclusivamente el subconjunto de pozos de entrenamiento que representa el 56.2% de toda la población de datos y es un total de 8 pozos de entrenamiento.

La información se muestra en las Tabla B.1 del Apéndice B es el número de datos o tamaño de muestra, valor mínimo, valor máximo y valor medio de los registros medidos con

---

<sup>22</sup> Minería de datos es un campo de la estadística y las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos Utiliza los métodos de la inteligencia artificial, aprendizaje automático, estadística y sistemas de bases de datos. [49]

mayor frecuencia en los pozos. Son los registros medidos con mayor frecuencia en los pozos, pero no en todos los pozos fueron medidos todos los registros que se muestran en la Tabla B.1, por ejemplo el registro SP o potencial natural jamás se mide en los pozos del CampoB.

La muestra de cada registro se conformó con la unión de los datos de cada registro de cada pozo que tenían un valor numérico asignado, es decir que, los datos que eran nulos en los registros se retiraron para realizar las medidas de muestra. Es por este motivo que se puede observar que los tamaños de las muestras para cada registro difieren considerablemente en la columna **No. Puntos** de la Tabla B.1.

En la metodología teórica que se utilizó en este trabajo para estimar Saturación de Agua sólo se utilizaron los registros GR, MSFL, NPHI y RHOB, por lo que los posteriores análisis se centran exclusivamente en estos.

Datos de Entrenamiento CampoB						
Registro	No. Datos	Min	Max	Mean	Median	Std
gr	112094	7.023	184.125	56.582	57.748	12.947
msfl	112094	0.258	1950	3.106	1.587	29.565
nphi	112094	0	0.819	0.309	0.34	0.137
rhob	112094	1.874	2.952	2.37	2.367	0.112
log(msfl)	112094	-0.588	3.29	0.284	0.201	0.254
Sx0	108207	0.008	1	0.407	0.372	0.162

**Tabla 5. 1. Información muestral de los registros medidos en los pozos del CampoB para los datos Limpios.**

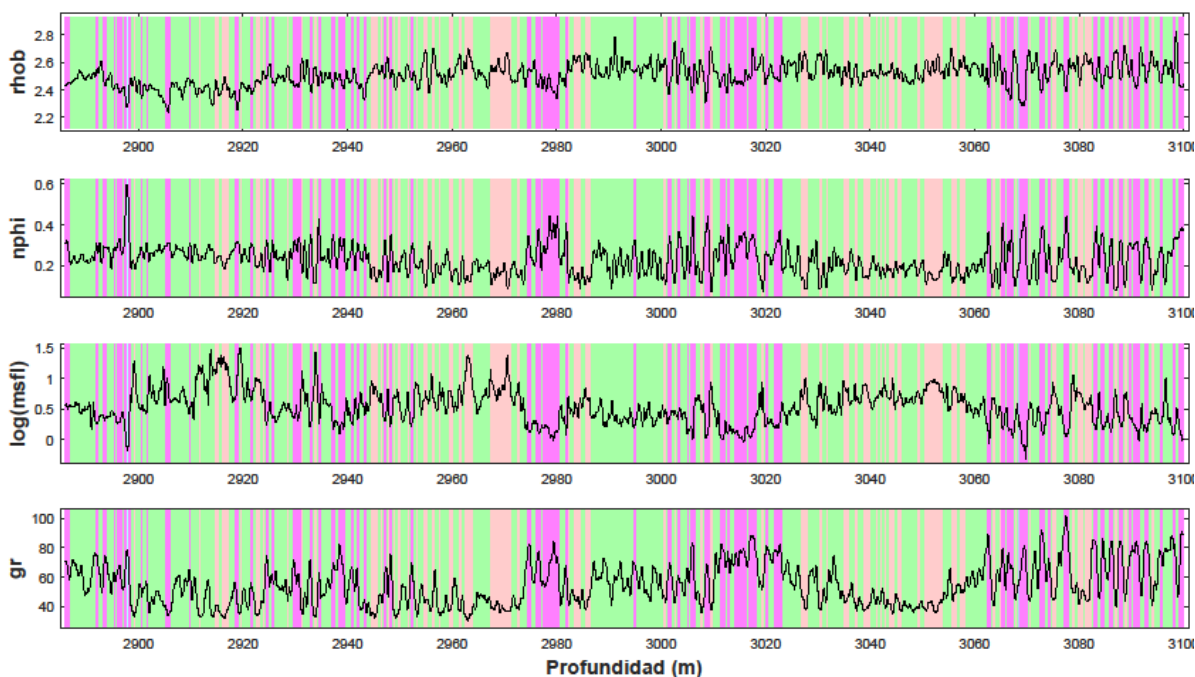
Lo más importante a resaltar en B.1 son los valores que se tienen para **Mínimo** y **Máximo** para cada registro, estos están definiendo un rango para las variables y como se observa en muchos casos indica que los registros tienen valores negativos lo cual resulta inconsistente con las medidas físicas que se realizan, las cuales indican que cualquiera de los registros sólo pueden tomar valores positivos, por este motivo los posteriores análisis se realizan sobre los Datos Limpios los cuales son los Datos Originales donde los datos negativos son sustituidos por el valor Nulo.

Para los Datos Limpios se estimó nuevamente la información muestral de forma similar a lo que se realizó para los Datos Originales. Esta información se presenta en la Tabla 5.1. Como se observara además de los registros que se seleccionaron (GR, MSFL, NPHI y RHOB) se presentan adicionalmente la información muestral de log(MSFL) y Sx0, que es la Saturación de Agua. El registro de Saturación de Agua (Sx0) muestreado en las tablas corresponde a la curva calculada para cada pozo utilizando la metodología de Archie-Pickett presentada en el Capítulo 2, esto se menciona de una forma detallada más adelante.

La Tabla 5.1 muestra los rangos, media y desviación estándar que se presentan en la unión de los registros de los pozos del campo. Se agregó la información correspondiente a

$\log(\text{MSFL})$  debido a que por ser MSFL un registro de tipo logarítmico que conviene tratar como una ley de potencias. De cualquier forma siempre se presentaran en posteriores análisis la información correspondiente tanto a MSFL como  $\log(\text{MSFL})$ .

La información presentada en la Tabla 5.1 es muy general sobre la información de los registros en cada campo por lo que en las Tablas B.3 se presenta la información muestral de los registros de forma más detallada para cada pozo que conforma el CampoB. En estas tablas es posible identificar cualquier pozo y ver los rangos que tienen los registros para ese pozo. Podemos notar que los valores medios, **Prom**, para cada pozo de RHOB son relativamente similares así como sus desviaciones estándar.



**Figura 5. 2. Registros del pozo 40 de Datos Limpios – CampoB. Ver código de color de clases Tabla 5.3.**

En las Figuras 5.2 y 2.2 se ilustran los registros GR, MSFL, NPHI y RHOB de los pozos 40 y 67 del CampoB. Estas muestran gráficamente el porqué de las diferencias de tamaños en las muestras de cada registro en las Tablas 5.1, esto es debido a que desde los Datos Originales, no todas las medidas se toman desde el principio de profundidad indicado para cada pozo, y tampoco no todos los registros terminan de ser tomados a la misma profundidad, ni en la profundidad final que en principio se tiene para cada pozo. Adicionalmente algunos registros presentan „huecos“ con datos nulos entre valores debido al proceso que se realizó para pasar de los Datos Originales a los Limpios, es decir el de retirar los valores negativos y sustituirlos por el nulo.

Los Datos Limpios fueron sustituidos por datos generados de forma sintética como se presenta en el Apéndice C. Con estos procedimientos de preparación de datos hemos obtenido los insumos necesarios para la construcción de la red bayesiana y además hemos generado dos conjuntos de datos que utilizaremos de forma diferente, los Datos Limpios como el conjunto con el cual se diseñara la estructura de la red y se realizara su entrenamiento y el de los Datos Monte Carlo sobre el cual se aplicará la estructura de red bayesiana de los Datos Limpios.

## Agrupamiento de datos en Clases Petrofísicas

Los registros geofísicos de un pozo suelen ser medido a lo largo de varios cientos o miles de metros a lo largo del pozo, por lo que geológicamente se observa que se están midiendo las características físicas de distintos tipos de sedimentos o arenas que se van apilando sobre la superficie. Se espera tener capas con distintos estratos.

Un tratamiento adecuado para la consideración de que los datos están representando distintas unidades litológicas, es realizar un análisis litológico con los registros por pozo y asociar por intervalo la etiqueta correspondiente, para posteriormente trabajar con datos de la misma litología de manera independiente.

Este trabajo utilizó un algoritmo de agrupamiento para identificar grupos en los datos con características petrofísicas similares, que buscan cierta correspondencia litológica. El algoritmo de agrupamiento de datos utilizado es **The fuzzy c-means clustering algorithm** [50] que consiste en identificar la similitud entre un posible elemento y un grupo por una función de pertenencia, que toma valores entre cero y uno. Los valores cercanos a uno indican una mayor similitud, mientras que los cercanos a cero indican una menor similitud. Además, cumplen que para cada muestra la suma de sus participaciones en cada grupo es uno. Este algoritmo define un criterio de agrupamiento en la forma de una función objetivo que depende de la partición difusa. El procedimiento consiste en minimizar iterativamente esta función hasta obtener una partición difusa óptima.

Al aplicar este algoritmo a cada vector de datos del CampoB, configurado como  $X = \{x_1 = gr \ x_2 = \log(msfl) \ x_3 = nphi \ x_4 = rhob\}$ , se generan diferentes grupos petrofísicos con diferentes valores medios de los registros que dependen de la partición difusa identificada por el algoritmo.

En la Tabla 5.2 se presentan los valores medios de los registros para las grupos identificados con Fuzzy C-Means considerando primero 2 clases independientes, después 3 clases independientes y así hasta tener 6 clases independientes. **Se realizaron las pruebas con varias posibles divisiones de grupo 2, 3, 4, 5 y 6 por falta de evidencia litológica que indicara un número adecuado de litologías diferentes en las que se tenía que clasificar los datos.**



CampoB							
No.		Valor	Valor	Valor	Valor	Valor	Tamaño
Grupos	Id	Medio	Medio	Medio	Medio	Medio	de
Totales	Grupo	GR	MSFL	NPHI	RHOB	log(MSFL)	Muestra
<b>2</b>	1	49.099	2.959	0.159	2.465	0.471	65509
	2	62.756	1.677	0.399	2.337	0.225	154714
<b>3</b>	1	47.607	3.156	0.143	2.468	0.499	56732
	2	62.48	1.748	0.453	2.282	0.243	76009
	3	62.562	1.645	0.339	2.395	0.216	87482
<b>4</b>	1	42.056	4.511	0.151	2.543	0.654	36759
	2	58.279	1.586	0.122	2.313	0.2	18974
	3	62.442	1.658	0.337	2.399	0.219	87090
	4	62.507	1.741	0.452	2.283	0.241	77400
<b>5</b>	1	37.898	6.049	0.121	2.567	0.782	22897
	2	52.286	2.379	0.229	2.479	0.376	26831
	3	58.711	1.562	0.118	2.307	0.194	17866
	4	62.363	1.774	0.458	2.277	0.249	69642
	5	63.407	1.596	0.353	2.385	0.203	82987
<b>6</b>	1	34.792	7.852	0.113	2.575	0.895	16402
	2	49.571	2.643	0.193	2.504	0.422	27033
	3	58.701	1.576	0.119	2.3	0.198	17328
	4	61.992	2.101	0.494	2.268	0.323	35215
	5	62.368	1.505	0.406	2.311	0.178	60351
	6	63.363	1.662	0.332	2.408	0.221	63894

Tabla 5. 2. Valores medios de los registros petrofísicos para los grupos petrofísicos identificados por Fuzzy C-Means para los datos del CampoB.


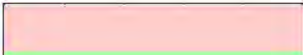

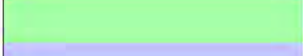








	Color Indicador de Clase para puntos y contornos	Color Indicador de Clase para fondo de intervalos
Clase Petrofísica 1		
Clase Petrofísica 2		
Clase Petrofísica 3		
Clase Petrofísica 4		
Clase Petrofísica 5		
Clase Petrofísica 6		

Tabla 5. 3. Código de color utilizado en este trabajo para indicar las distintas clases petrofísicas en las gráficas de la tesis.

Estas clases petrofísicas se utilizaron durante toda la metodología para la obtención de la Red Bayesiana. Se obtuvieron tantas Redes Bayesianas como Grupos Independientes que fueron considerados, es decir, si se clasificaron los datos en 6 grupos de características petrofísicas similares independientes, entonces se generaron 6 Redes Bayesianas independientes para indicar las relaciones de probabilidad para cada clase petrofísica. En la Tabla 5.3 se indica el código de color utilizado en este trabajo para diferenciar los distintos grupos petrofísicos.

## Aplicación de Archie – Pickett para obtener la curva teórica de Saturación de Agua

Como se observa en la Tabla 5.1 y B.3 se estimó un registro de Saturación de Agua  $S_x0$  teórico utilizando la metodología descrita en el capítulo 2.1. La metodología seguida para calcular la curva teórica de la Saturación de Agua de la Ecuación de Archie [4] consistió en identificar una serie de intervalos por debajo de la línea base de lutitas sobre el registro del índice de arcillosidad y sobre ellos se aplican los pasos de estimación de volumen de arcilla y estimación de porosidades, hasta llegar al gráfico de Pickett por medio del cual es posible calcular la resistividad del lodo filtrado ( $R_{mf}$ ) y el exponente de cementación ( $m$ ). En la Tabla B.15 se presentan el número de intervalos que se lograron identificar contra los intervalos en los que se pudo evaluar el gráfico de Pickett, las columnas de Profundidad de Inicio y Final indica el rango de profundidades en las que fueron identificados los intervalos.

El gráfico de Pickett es la gráfica del registro resistivo MSFL vs la porosidad efectiva  $e$ , ecuación (2.27). Para calcular los valores de  $m$  y  $R_{mf}$  se identifica una recta „secante“ que deje 10% de los puntos por debajo de la línea; la elección es de carácter subjetivo por lo que con fines de automatizar esta metodología para la evaluación de todos los pozos. Se generaron cinco rectas por cinco métodos diferentes que se indican a continuación.

**1. Cúmulos Extremos.** El largo del cúmulo se divide en 8 partes y se utilizan los puntos que queden entre  $\left[ x_{min}, x_{min} + \frac{x_{max} - x_{min}}{8} \right]$  con estos puntos se calcula un centroide de coordenadas  $(x_1, y_1)$ . El ancho del cúmulo se divide en 8 partes y se utilizan los puntos que queden entre  $\left[ y_{min}, y_{min} + \frac{y_{max} - y_{min}}{8} \right]$  con estos puntos se calcula un centroide de coordenadas  $(x_2, y_2)$ . Utilizando  $(x_1, y_1)$  y  $(x_2, y_2)$  se calcula la pendiente de la recta que pasa por estos puntos.

**2. Ajuste Lineal.** Se realiza el ajuste de mínimos cuadrados del cúmulo se utiliza la pendiente de este ajuste.

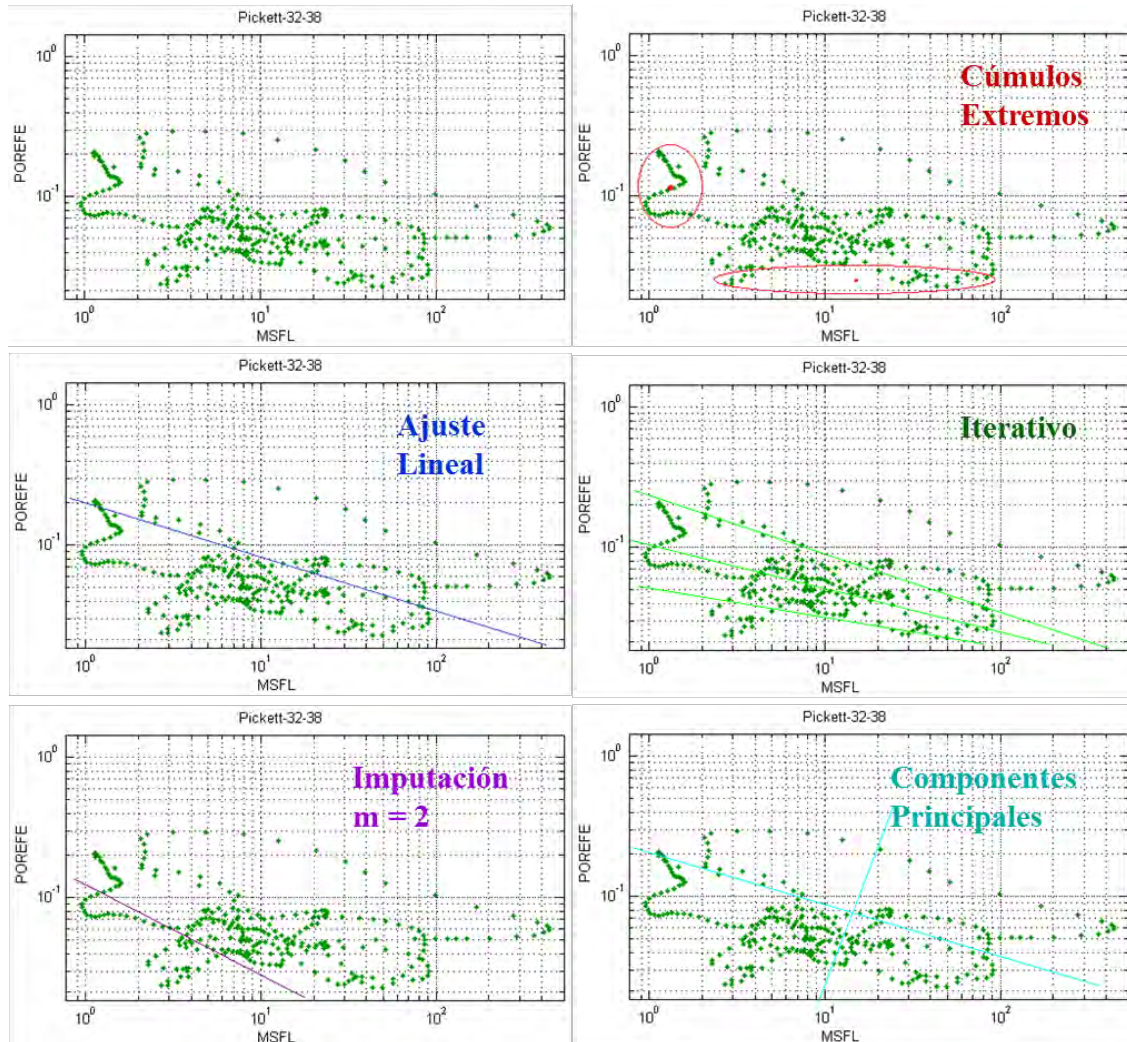
**3. Iterativo.** Se realiza un ajuste con mínimos cuadrados del cúmulo se utiliza la pendiente resultante del ajuste se calcula la recta con esta pendiente que pase por el centro geométrico

del cúmulo y se genera un nuevo cúmulo compuesto de todos los puntos del anterior que queden en la parte inferior de la recta. Con este nuevo cúmulo se vuelve a realizar el mismo proceso hasta quedarse con la octava parte de puntos. La pendiente que se utiliza es la resultante del ajuste del último cúmulo que quedo.

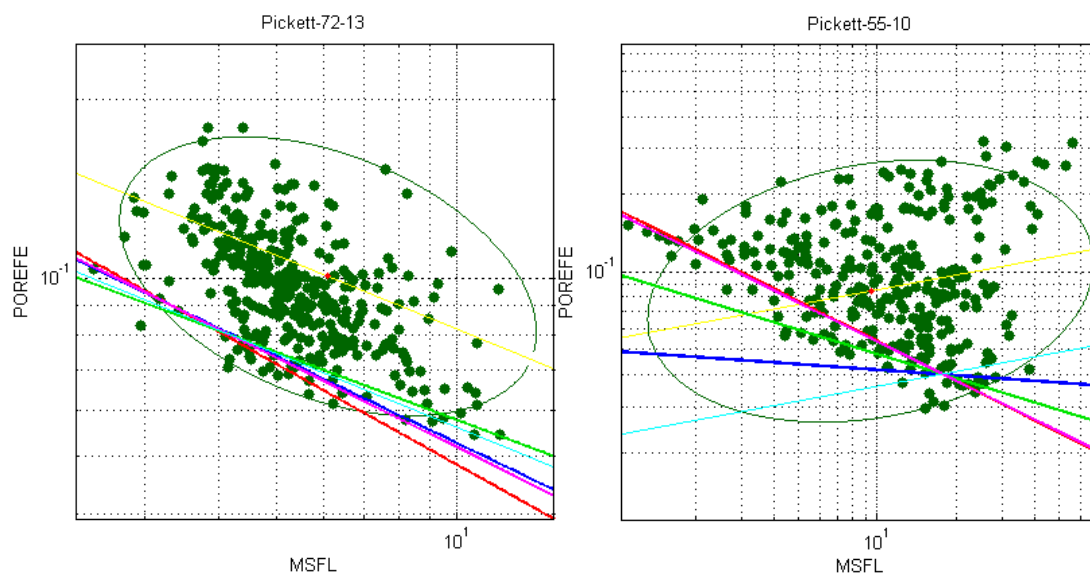
**4. Imputación  $m = 2$ .** Se utiliza pendiente igual a  $-0.5$  en la ecuación (2.25) para obtener un exponente de cementación  $m = 2$  este valor sólo es una aproximación.

**5. Componentes Principales.** Se aplica el método de componentes principales para buscar el ángulo de rotación del cúmulo. El ángulo utilizado cumple con la característica de que el lado más largo del cúmulo coincide con el eje X en el plano rotado. La pendiente que se utiliza es  $\tan \theta$  donde  $\theta$  es el ángulo de rotación.

En la Figura 5.4 se presenta la Ejemplificación de la selección de la recta de 100% de saturación de agua para el Método Gráfico de Pickett de acuerdo a los criterios e puestos previamente.



**Figura 5. 4. Ejemplificación de la selección de la recta de 100% de saturación de agua para el Método Gráfico de Pickett.**



**Figura 5. 5. Gráfico de Pickett para el pozo 72 intervalo 13 y para el pozo 55 intervalo 10 respectivamente.**

Existen muchos factores por los cuales un intervalo pudo ser descartado para ser evaluado con Pickett, como principal causa se puede indicar la falta de datos de algunos registros, otra de las causas es que cuando se aplican las diferentes técnicas de obtención de la recta „secante“ a la nube de puntos en el gráfico de Pickett por su forma arroja resultados sin significado físico con valores para  $m$  y  $R_{mf}$  que no corresponden a ningún tipo de roca.

En las Figuras 5.5 se presentan algunos ejemplos de los gráficos de Pickett obtenidos para los pozos, donde es posible observar las diferencias que se observan de la elección de una recta a otra y la influencia que tiene la forma de la nube en la pendiente de la recta.

La elección de cada recta arroja parejas de  $m$  y  $R_{mf}$  diversas y por tanto, se obtienen curvas de Saturación de Agua diferentes. Para este trabajo se seleccionó un intervalo y una recta en particular para cada pozo la cual tenía la característica de ser lo más parecida a  $m = 2$  y un valor pequeño para  $R_{mf}$ . En la Tabla B.4 se presenta la recta e intervalo seleccionado para cada pozo, el número de puntos que conformo la nube de datos y los valores de exponente de cementación ( $m$ ), resistividad de lodo filtrado ( $R_{mf}$ ) y densidad de matriz ( $\rho_{ma}$ ) obtenidos para la roca.

## 5.2. Identificación de relaciones de probabilidad

De la sección 5.1, se ha presentado un conjunto de cuatro registros {GR MSFL NPHI RHOB} para calcular un quinto de manera teórica que es la Saturación de Agua. Estas cinco variables son el conjunto que describe nuestro problema.

Para generar una estructura de red que relacione nuestras variables podemos generar como primera aproximación una en la que todas las variables estén relacionadas entre sí, como en la Figura 5.6.

La Figura 5.6 es una estructura de red que relaciona todas las variables entre sí, donde cada nodo representa una variable aleatoria la cual tiene asociada una función de probabilidad y cada enlace representa una relación estadística bidireccional no dirigida, aunque como condición de relación estadística

$$p(x | y) \neq p(y | x). \quad (5.1)$$

Para que la red de relaciones entre las variables de la Figura 5.6 sea una Red Bayesiana se tiene que dar dirección a los arcos que la conforman es decir asociar una relación de nodo padre a nodo hijo con cada una de las parejas y además es necesario asegurarse de que las relaciones que se tengan entre las variables no generen ciclos o caminos de un nodo a sí mismo. Ver Capítulo 2 sección 2.

Existen varios procedimientos para obtener la estructura de relación entre variables para una Red Bayesiana, pero en este trabajo se partió inicialmente de un grafo densamente conectado de la Figura 5.6 y se realizó un análisis de la fuerza de la conexión entre la relación de cada pareja de variables.

La identificación de las relaciones de probabilidad se realizaron utilizando dos métodos: El análisis de los gráficos cruzados entre variables y por medio del análisis de componentes principales.

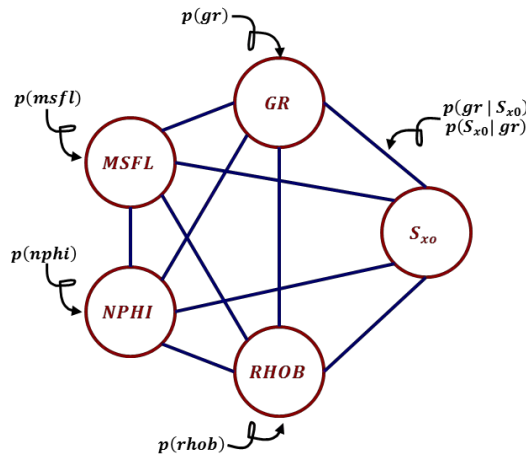


Figura 5. 6. Primera estructura de red con relación máxima entre las variables aleatorias.

## Análisis de gráficos cruzados entre variables

El análisis de gráficos cruzados es un método propuesto en este trabajo para obtener la relación de probabilidad condicional entre las variables aleatorias o nodos de nuestra red.

Para cada pareja de variables  $X$  y  $Y$  se realiza un gráfica de  $X$  vs  $Y$  y se analiza cómo queda distribuidos los puntos de  $Y$  en función de los de  $X$ , para identificar la relación que existe entre cada pareja de variables se realiza un ajuste lineal de un subconjunto del original  $[X', Y'] \subset [X, Y]$ <sup>23</sup>, donde  $(x', y') \in [X', Y']$  si  $x' \in [\bar{X} - \sigma(X), \bar{X} + \sigma(X)]$  y  $y' \in [\bar{Y} - \sigma(Y), \bar{Y} + \sigma(Y)]$ .

En las Tablas B.10 se muestran la pendiente y ordenada al origen del ajuste de la nube de puntos formada en el grafico cruzado de las diferentes combinaciones entre registros para cada campo y grupo petrofísico respectivamente, ver Figuras 5.8.

En las Tablas 5.3 se muestran las matrices de las relaciones entre los registros del CampoB para cada grupo petrofísico de 6 clases, donde “||” significa que no hay relación.

Relación gráfica entre registros CampoB Grupo 1 de 6							Relación gráfica entre registros CampoB Grupo 4 de 6						
	gr	msfl	nphi	rhob	log(msfl)	Sx0		gr	msfl	nphi	rhob	log(msfl)	Sx0
gr		p(msfl)	p(nphi)		1		gr		p(msfl)	1	p(rhob)		
msfl			p(nphi)				msfl						
nphi				1			nphi		p(msfl)		p(rhob)		1
rhob	p(gr)	p(msfl)	1			1	rhob	1	p(msfl)				
log(msfl)	p(gr)		p(nphi)	1			log(msfl)	p(gr)		p(nphi)	p(rhob)		1
Sx0	p(gr)	p(msfl)	p(nphi)	p(rhob)	p(log(msfl))		Sx0	p(gr)	p(msfl)	p(nphi)	p(rhob)	p(log(msfl))	
Relación gráfica entre registros CampoB Grupo 2 de 6							Relación gráfica entre registros CampoB Grupo 5 de 6						
	gr	msfl	nphi	rhob	log(msfl)	Sx0		gr	msfl	nphi	rhob	log(msfl)	Sx0
gr		p(msfl)		1	p(log(msfl))		gr		p(msfl)	1	p(rhob)	1	
msfl							msfl						
nphi	1	p(msfl)		1	p(log(msfl))	1	nphi		p(msfl)			1	
rhob		p(msfl)			p(log(msfl))		rhob		p(msfl)	1		p(log(msfl))	
log(msfl)	1						log(msfl)						1
Sx0	p(gr)	p(msfl)	p(nphi)	p(rhob)	p(log(msfl))		Sx0	p(gr)	p(msfl)	p(nphi)	p(rhob)	p(log(msfl))	
Relación gráfica entre registros CampoB Grupo 3 de 6							Relación gráfica entre registros CampoB Grupo 6 de 6						
	gr	msfl	nphi	rhob	log(msfl)	Sx0		gr	msfl	nphi	rhob	log(msfl)	Sx0
gr		p(msfl)	1		1		gr		p(msfl)	p(nphi)	p(rhob)	p(log(msfl))	
msfl							msfl						
nphi	p(gr)	p(msfl)			p(log(msfl))	1	nphi		p(msfl)			p(log(msfl))	1
rhob	1	p(msfl)	1		1		rhob	1	p(msfl)	p(nphi)		p(log(msfl))	1
log(msfl)			1				log(msfl)						1
Sx0	p(gr)	p(msfl)	p(nphi)	p(rhob)	p(log(msfl))		Sx0	p(gr)	p(msfl)	p(nphi)	p(rhob)	p(log(msfl))	

**Tabla 5. 3. Matrices de conexión entre los registros petrofísicos del CampoB para 6 grupos petrofísicos.**

Las relaciones en las Tablas 5.3 se obtuvieron de analizar la tendencia de las nubes de puntos en la Figura 5.8 de acuerdo al siguiente criterio. Como se desea calcular  $p(Y | X)$  al realizar el grafico cruzado de ambas variables podemos presentar tres posibles casos de ajuste al cúmulo de puntos de las variables que se presentan en la Figura 5.7, el caso A nos indicaría que la función de relación es constante, lo que significa que:

<sup>23</sup> En teoría de conjuntos  $A$  es subconjunto de  $B$  se simboliza como  $A \subset B$  significa: cada elemento de  $A$  es también elemento de  $B$ , pero  $A \neq B$ .

$$p(Y | X) \approx p(Y). \quad (5.2)$$

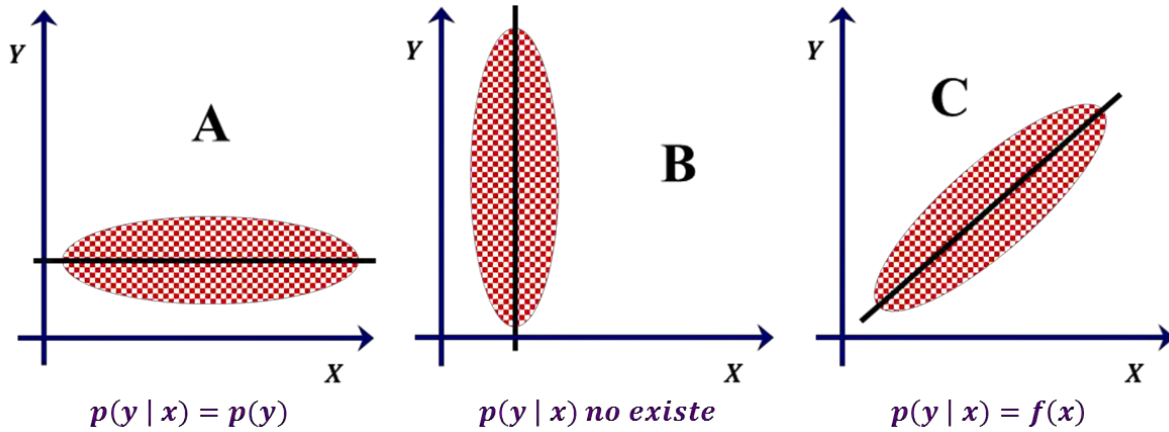


Figura 5. 7. Posibles relaciones entre las parejas de variables aleatorias.

Estos se indican en las Tablas 5.3 de manera que en el cruce de  $X$  y  $Y$  aparece  $p(Y)$ . El caso B indica la ausencia de una relación funcional para  $p(Y | X)$  lo cual se indica como que „No hay relación“ o “| |” en las tablas.

El caso C es aquel donde existe una relación funcional para  $p(Y | X)$  y aunque en la Figura 5.7 esta se muestra lineal, en realidad la relación funcional puede ser cualquiera, como un polinomio o un logaritmo; esto nos indica que para la estructura de red bayesiana existe una conexión entre  $X$  y  $Y$  este vínculo se indica en las Tablas 5.3 como „1“.

Como se observa el registro MSFL se repite dos veces en las Tablas 5.3 como el registro MSFL original y su logaritmo  $\log(\text{MSFL})$  esto es porque de la Ecuación de Archie se sabe que la relación entre  $Sx_0$  y MSFL es del tipo  $Sx_0 \propto \text{MSFL}^N$ , por lo que es conveniente analizar también los gráficos cruzados de  $\log(\text{MSFL})$  contra los demás registros.

Las Tablas 5.3 se nombraron como matriz de conexión entre las variables; puesto que si todos los espacios que no tengan 1 los sustituimos con 0 los 1's indican la conexión para construir una red. En las Figuras 5.9 se muestra el diseño de red que surgió a raíz de estas matrices para el cada grupo petrofísico de 6. En algunos casos en particular en la relación que existe entre NPHI y RHOB la relación es bidireccional, en estos casos se seleccionó una dirección que no generara ciclos en la estructura de red.

Las estructuras presentadas en las Figura 5.9 son las Redes Bayesianas identificadas para el CampoB; estas cumplen con la definición mencionada en el Capítulo 3 y con ellas es posible construir una densidad de probabilidad para cada nodo que la conforma de acuerdo al Teorema de Bayes.

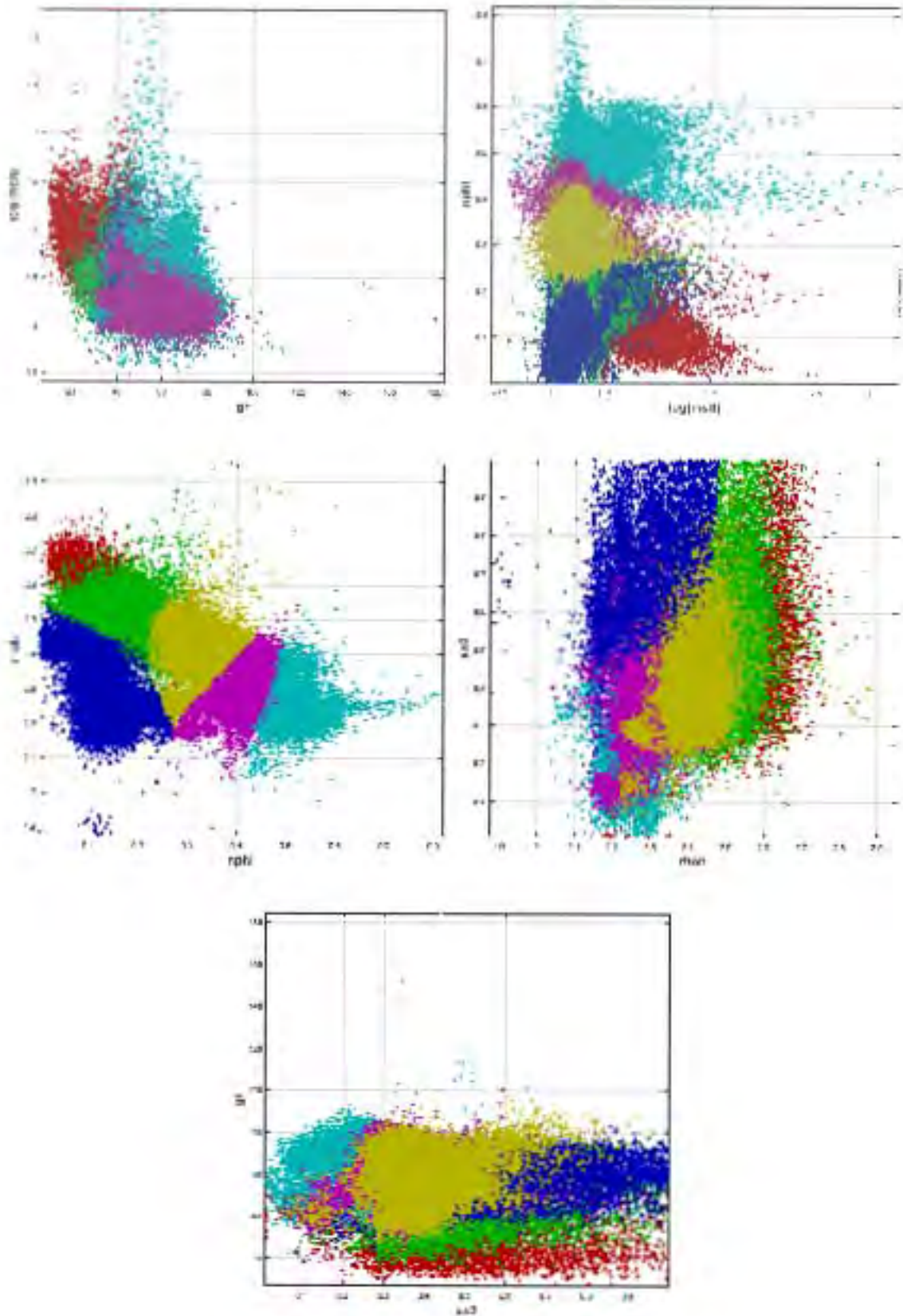


Figura 5. 8. Gráficas cruzadas de algunas parejas de registros del CampoB.



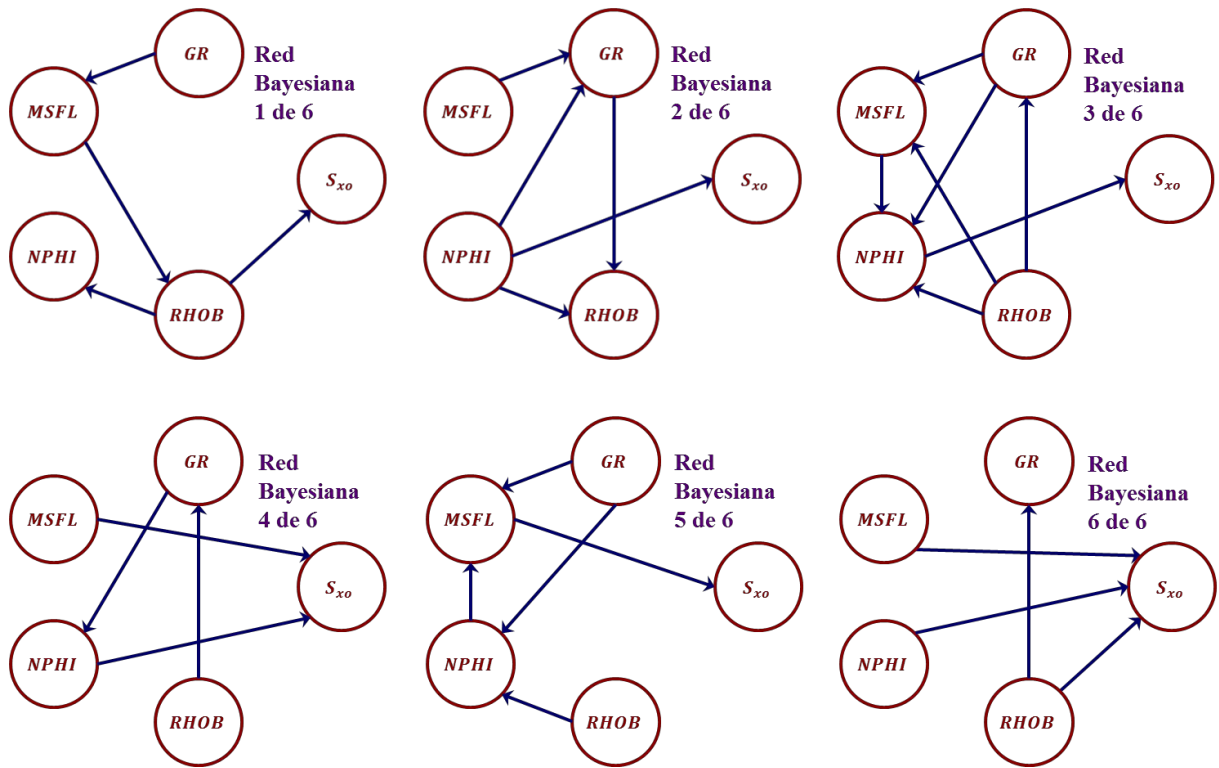


Figura 5. 9. Gráficas cruzadas de algunas parejas de registros del CampoB.

### 5.3. Discretización de variables

El área de extracción de Redes Bayesianas se divide en proponer una estructura y en el cálculo de los parámetros subyacentes a la estructura.

La discretización de las variables permite expresar las distribuciones de probabilidad de manera mucho más sencilla, al tiempo que facilita la implementación de los algoritmos.

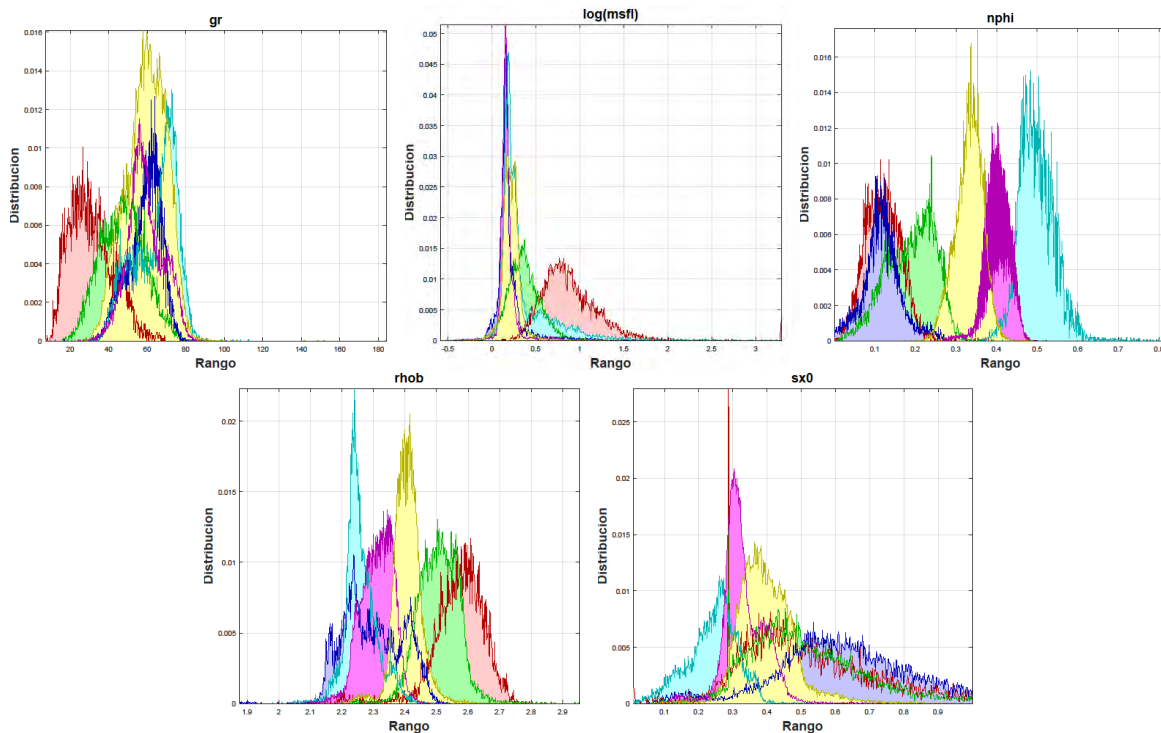
Este tipo de representación del problema suele realizarse en los casos en los que se sabe que las variables toman valores de un conjunto discreto, que aunque no es el caso de las variables de los registros geofísicos, ha sido la metodología utilizada para la estimación de las probabilidades. En las Tablas 5.1 se presenta el inicio y fin del rango de variación de cada uno de los registros medidos en ambos campos, estos representan los rangos de variable sobre los que se realizará el proceso de discretización.

El proceso de discretización para cualquiera de las señales se realiza dividiendo el rango total de fluctuación de la variable en  $M$  niveles discretos. La serie de tiempo que representa el registro  $X$  es entonces sustituida por una señal de números naturales donde cada número de muestra debe decidirse cuál de los  $M$  niveles es la mejor aproximación al valor de la señal. Eligiendo el valor más cercano, éste se mantiene hasta el siguiente tiempo de muestra. Este proceso de cuantificación introduce diferencias respecto al valor real de la señal, las cuales pueden considerarse como ruido. Si se aumenta el número de niveles de

cuantificación  $M$  tiende a reducir este ruido. La cuantificación puede ser lineal (con niveles igualmente espaciados) o no lineal (con niveles espaciados de manera no uniforme). A la diferencia que existe entre niveles contiguos se le llama tamaño del intervalo de cuantificación, y comúnmente se denota por la letra  $h$ . [51][52]

Después de cuantificar la señal se asigna un dígito a cada nivel, de manera que exista correspondencia uno a uno entre los niveles y el conjunto de dígitos. Esto se llama digitalización de la señal y reduce la señal a un conjunto de dígitos en tiempos de muestra sucesivos, originando un sistema de modulación completamente digital.

Las técnicas de discretización en este campo se concentran en la definición de tamaños de intervalo óptimos.



**Figura 5. 10. Histogramas de frecuencia de cada registro considerando un No. de Cajas = 4 3, para el CampoB con los Datos Monte Carlo.**

Al realizar este tipo de discretización se produce una malla sobre el espacio continuo  $n$ -dimensional completo de instancias, donde  $n$  corresponde al número de variables o registros independientes que estamos considerando, cada variable es discretizada en regiones independientes de las demás. La malla contiene  $\prod_{i=1}^n k_i$  regiones, en donde cada  $k_i$  es el número de particiones en la variable  $i$ -ésima

Posteriormente a realizar la discretización de los datos para cada variable es posible determinan el número de intervalos  $k$  a producir de manera independiente para cada una de ellas.

El método de discretización más simple es dividir cada variable como intervalo de ancho fijo. Simplemente se divide el intervalo de valores observados en  $k$  cajones de igual tamaño, en donde  $k$  es un parámetro constante previamente definido.

El no. de cajas con el cual se realiza la discretización en el campo fue de 403, estos números se tomaron porque representa el número de datos mínimo sin datos repetidos de los registros según cada campo. En las Figuras 5.10 se presentan los histogramas de frecuencias obtenidos con este tipo de discretización para el CampoB de los Datos modificados con Monte Carlo utilizado para entrenamiento.

## 5.4. Entrenamiento de Red Bayesiana

Existen múltiples formas de tratamiento, la aplicada en este trabajo es la consideración de la Base de Datos como un conjunto de casos, en donde cada caso es una instancia de un conjunto de los registros.

La propuesta de consideración de estados posibles para este sistema se logra utilizando la discretización mostrada en la Sección 5.3 como una forma de dividir el espacio de 5 dimensiones representadas por los 5 registros como cubos en este espacio esta discretización se puede utilizar para estimar el valor de probabilidad de los nodos y las condicionales necesarias para utilizar las estructuras diseñadas en 5.3.

Para obtener cualquier Red Bayesiana en este trabajo de tesis se partió de una base de datos donde se podía observar la relación estadística entre los registros geofísicos que utilizamos {GR MSFL NPHI RHOB  $\log(\text{MSFL})$ } con la Saturación de Agua,  $S_{x0}$ , por lo que la Red Bayesiana se puede representar como una función que recibe como variables a los registros y asocia como resultado un valor de Saturación de Agua.

El proceso de generación y entrenamiento de la Red Bayesiana es la identificación de las reglas probabilísticas que gobiernan esta función. **Para la estimación de todas las relaciones de probabilidad que se utilizaron en este trabajo se realizó el proceso de „Discretización de Variables“ que como se mencionó en la Sección 5.3 consistió en que para cada variable se identificó el rango del dominio y tal intervalo fue dividido en  $k$  segmentos iguales donde  $k$  es un parámetro que aquí se nombra como „No. de Cajas“.** Este valor en algunos trabajos se suele relacionar con la raíz del número de datos que conforma la muestra,  $N$ , aunque, en este trabajo, se propusieron algunos valores diferentes para  $k$  con los que se realizaron pruebas y comparaciones. Este parámetro es para este trabajo de suma importancia, debido a que la precisión y la cantidad de reglas estimadas para las relaciones entre registros dependen del valor de  $k$ .

Se partió de una matriz de datos de  $5 \times N$ , donde  $N$  es el número de datos totales que sí tienen un valor de  $S_{x0}$  asociado. Inicialmente se utilizó la Metodología de Archie-Pickett para estimar un valor de saturación de agua guía para cada punto contra la dimensión del

vector de variables conformado por  $\{GR \ MSFL \ NPHI \ RHOB \ \log(MSFL)\}$  que es 5. La matriz de  $5 \times N$  esta asociada a un vector  $1 \times N$  que representa la solución de Saturación de Agua estimada por Archie-Pickett.

Para cada variable se realizó el proceso de discretización, que consistió en que de acuerdo a su valor se le asignó un número de caja que correspondía, considerando los límites entre los segmentos iguales previamente estimado de la división del rango de la variable por el  $k$ , no. cajas. Este proceso también se realizó al valor de Saturación de Agua asociado al vector. En la Tabla 5.4 se presenta un ejemplo de este procedimiento. Se utilizó el vector de valores para la profundidad  $1931.001m$  del pozo 40 y para el caso de un valor de  $k = 10$ , se le asignó un vector de estado  $[2, 1, 2, 5, 5] \rightarrow [4]$  que tiene relación con su respectiva caja de  $Sx0$ , este mismo vector para  $k = 100$  tiene un vector de estado  $[11, 1, 11, 49, 42] \rightarrow [37]$ , para  $k = 1000$  presenta  $[109, 1, 101, 486, 420] \rightarrow [366]$ , etc. Nuevamente se realiza este proceso para el siguiente vector de profundidad  $1931.154m$  y así en general hasta cubrir  $N$  datos que conforman la muestra.

NoPozo	Profundidad	GR	MSFL	NPHI	RHOB	log(MSFL)	Sx0
15	1931.001	31.139	10.934	0.1	2.723	1.039	0.372
	<b>No. Cajas</b>	<b>eGR</b>	<b>eMSFL</b>	<b>eNPHI</b>	<b>eRHOB</b>	<b>eLog(MSFL)</b>	<b>eSx0</b>
	<b>10</b>	2	1	2	5	5	4
	<b>100</b>	11	1	11	49	42	37
	<b>1000</b>	109	1	101	486	420	366
NoPozo	Profundidad	GR	MSFL	NPHI	RHOB	log(MSFL)	Sx0
15	1931.154	32.978	11.121	0.098	2.718	1.046	0.38
	<b>No. Cajas</b>	<b>eGR</b>	<b>eMSFL</b>	<b>eNPHI</b>	<b>eRHOB</b>	<b>eLog(MSFL)</b>	<b>eSx0</b>
	<b>10</b>	2	1	1	5	5	4
	<b>100</b>	12	1	10	49	43	38
	<b>1000</b>	115	1	98	485	421	374

**Tabla 5. 4. Representación de cómo se relaciona el vector de datos de las variables con un estado discreto, con su valor de Saturación de Agua y con el estado discreto de la Saturación de Agua. Los estados están en función del valor de  $k$ , No. de Cajas, utilizado para discretizar las variables, en la tabla  $k = \{10, 100, 1000\}$ .**

La Red Bayesiana se obtiene entonces de la realización del conteo de la frecuencia de ocurrencia de un estado como  $[11, 1, 11, 49, 42] \rightarrow [37]$  sobre todo el conjunto de datos. Una vez identificada su frecuencia para cada uno de los datos de este tipo sobre el subconjunto de datos es posible calcular su valor medio de  $Sx0$  y su probabilidad, por medio la Ecuación (5.3) donde  $n$  representa el número de elementos que conforma el subconjunto.

$$\overline{V_{Sx0}(eGR \ eMSFL \ eNPHI \ eRHOB \ eLog(MSFL))} = \frac{1}{n} \sum_{i=1}^n v_{Sx0i}(eGR \ eMSFL \ eNPHI \ eRHOB \ eLog(MSFL)). \quad (5.3)$$

La Red Bayesiana se presentará como una tabla de relaciones de los estados de las variables con los de la Saturación de Agua, un valor medio de saturación, la probabilidad del valor de saturación y la frecuencia de ocurrencia de la relación entre el estado general y el de saturación. La probabilidad del valor de saturación se considera como la obtenida del histograma de frecuencia de  $Sx0$ . En la Tabla 5.5 se presenta un fragmento de una la Red Bayesiana *CampoB-6-10* estimada en este trabajo.

eGR	eMSFL	eNPHI	eRHOB	eLog(MSFL)	eSx0	vSx0_medio	probabilidad	frecuencia
	1	1	5	4	2	0.162	0.176	1
	1	1	5	7	4	0.355	0.157	3
	1	1	6	7	2	0.162	0.176	5
	1	2		6	2	0.162	0.176	9
1	1	1	5	4	9	0.848	0.032	62
1	1	1	5	4	7	0.651	0.066	52
1	1	1	5	4	5	0.454	0.120	45
1	1	1	5	5	5	0.454	0.120	78
1	1	1	5	6	4	0.355	0.157	423
1	1	1	5	6	3	0.257	0.188	342
1	1	1	5	9	2	0.162	0.176	118
1	1	1	5	9	6	0.553	0.092	8
1	1	1	6	3	5	0.454	0.120	25
1	1	1	6	3	6	0.553	0.092	20
1	1	1	6	5	10	0.947	0.023	41
1	1	2	5	4	3	0.257	0.188	75
1	1	2	5	5	2	0.162	0.176	72
1	1	2	5	6	5	0.454	0.120	165
1	1	2	5	6	6	0.553	0.092	69
1	1	2	5	8	2	0.162	0.176	139
2	1	1	5	6	8	0.750	0.046	341
2	1	1	5	7	3	0.257	0.188	1623
2	1	1	5	7	4	0.355	0.157	1226
2	1	1	5	7	6	0.553	0.092	279
2	1	1	5	7	7	0.651	0.066	138
2	1	1	5	8	2	0.162	0.176	1139
2	1	1	5	8	3	0.257	0.188	668
2	1	1	6	3	6	0.553	0.092	32
2	1	1	6	6	6	0.553	0.092	360
2	1	1	6	6	5	0.454	0.120	331
2	1	1	6	6	2	0.162	0.176	43
2	1	1	6	7	4	0.355	0.157	491
2	1	2	5	7	1	0.070	0.101	362
2	1	3	4	7	1	0.070	0.101	64
2	1	3	5	5	3	0.257	0.188	68
3	1	1	5	4	5	0.454	0.120	30
3	1	1	5	6	5	0.454	0.120	137
3	1	2	5	4	7	0.651	0.066	92
3	1	2	5	4	5	0.454	0.120	54
3	1	2	5	6	3	0.257	0.188	223

**Tabla 5. 5. Fragmento de Red Bayesiana obtenida a partir de los Datos del Campo con  $k = 10$  y 6 grupos petrofísicos.**

Las relaciones dependen entonces del valor de discretización  $k$  y del conjunto de Datos que se utilizó para su entrenamiento. Por lo que a cada Red Bayesiana se le asigna un nombre como  $[Nombre\ del\ campo] - [No.\ de\ Grupos\ Petrofísicos] - [No.\ de\ Cajas]$ , por ejemplo el fragmento presentado en la Tabla 5.5 corresponde a la Red *CampoB-6-10*, donde se estimó la Red Bayesiana por Grupo Petrofísico correspondiente y por tanto se tienen 6 Redes Bayesianas para 6 Grupos Petrofísicos asignados a los Datos.

## 5.5. Implementación de Red Bayesiana

Una vez construida la Red Bayesiana, es posible determinar la distribución de probabilidad para una variable determinada utilizando reglas tales como la Regla de la Cadena y la Regla de Bayes. Esta distribución puede ser calculada tomando en cuenta el conocimiento acerca del valor de las otras variables.

Si a cada variable de un conjunto  $P = \{m, g, r, n\}$  se le asignan los valores de sus respectivas secuencias, al calcular la distribución de probabilidad para algún otro nodo  $s \in P$  conectado con algún nodo  $\in P$ , ésta debe tender a producir los valores de la secuencia correspondiente a  $s$ . Así, si la estructura y parámetros obtenidos para la Red Bayesiana son correctos, se pueden recuperar valores similares a los de la serie original muestreada sobre la distribución resultante. De este modo, si se calcula una secuencia de valores, lo que se obtendrá será una de varias trayectorias que podría tomar los datos recuperados. Es posible seleccionar la trayectoria más adecuada calculando varias trayectorias posibles y obteniendo el punto óptimo para cada valor de la profundidad.

Con base en varias trayectorias calculadas para la curva  $s$ , es posible calcular una estimación del error que tendrá la reconstrucción de la serie midiendo la divergencia de las trayectorias calculadas.

Podemos pensar que la curva que caracteriza cada registro puede ser representada como un proceso estocástico<sup>24</sup>.

Sea  $\{X_j(t)\}$  un proceso estocástico definido en tiempos discretos  $t_1, t_2, \dots, t_n$ . Para caracterizar a este proceso es necesario conocer la densidad de probabilidad conjunta de todas las variables aleatorias  $X$ . La densidad de probabilidad conjunta expresa la probabilidad de ocurrencia de cada posible configuración de las variables, es decir, las probabilidades de que cada variable tome un valor determinado.

Cada nodo en la Red Bayesiana tiene asociada una distribución de probabilidad que depende solamente de sus padres. Es posible calcular la distribución conjunta a partir de la Red Bayesiana utilizando algunas leyes de probabilidad, de modo que cualquier cálculo que

---

<sup>24</sup> Un proceso estocástico es una función aleatoria del tiempo, definida sobre algún intervalo de observación, con la característica de que, antes de hacer un experimento, no es posible conocer su valor.

utilice la distribución conjunta de probabilidad puede realizarse a partir de la red. Además, una Red Bayesiana permite realizar procesos razonamiento tales como predicción y abducción utilizando reglas de probabilidad. [27]

Cuando se habla de una Red Bayesiana, se conoce como estructura al grafo dirigido que representa la relación de dependencia entre los nodos, y se conoce como parámetros a las densidades de probabilidad existentes en cada nodo. Es importante notar que las densidades de probabilidad que se obtienen para cada nodo son condicionales respecto a los valores de sus nodos “Padre” es decir sea  $\text{Padres}(X) = \{\text{Padre}_1(X), \text{Padre}_2(X), \dots, \text{Padre}_n(X)\}$  el conjunto de nodos padre de  $X$ , la densidad de probabilidad de  $X$  es dependiente de  $\text{Padre}(X)$ :

$$X \sim f(X \mid \text{Padre}_1(X), \text{Padre}_2(X), \dots, \text{Padre}_n(X)). \quad (5.4)$$

Por lo tanto, la extracción de los parámetros requiere necesariamente una estructura definida para la red.

## Aplicando la Red Bayesiana

La red bayesiana se aplicó al conjunto de datos de validación, un subconjunto de pozos del CampoB, que no fueron con los que se realizó el entrenamiento, indicados en la Tabla 1.2.

NoPozo	Profundidad	GR	MSFL	NPHI	RHOB	log(MSFL)
26	2308.479	30.845	467.546	0.1	2.774	2.670
	No. Cajas	eGR	eMSFL	eNPHI	eRHOB	eLog(MSFL)
	10	2	1	1	5	8
	100	11	3	9	50	71
	1000	108	24	85	500	710
NoPozo	Profundidad	GR	MSFL	NPHI	RHOB	log(MSFL)
26	2308.631	29.400		0.087	2.770	
	No. Cajas	eGR	eMSFL	eNPHI	eRHOB	eLog(MSFL)
	10	2		1	5	
	100	11		9	50	
	1000	103		87	499	

**Tabla 5. 6. Representación de la asignación de estado para un vector de datos para valores de  $k = \{10, 100, 1000\}$ .**

Cuando se desea aplicar una Red Bayesiana previamente entrenada, inicialmente se debe de discretizar el conjunto de datos a estimar y asociar cada dato un grupo petrofísico de características similares de acuerdo a la Red Bayesiana entrenada. Los datos a los que se desea asignar un valor de Saturación de Agua por medio de la Red Bayesiana estructuralmente también deben formar una matriz de 5  $M$ , donde el 5 representa a la

dimensión de las variables y  $M$  al número de líneas a estimar. La discretización,  $k$ -NoCajas, que se realiza a los datos a estimar debe ser la misma que tiene la Red Bayesiana que se desea aplicar.

El proceso de discretización asigna un estado numérico a un vector de valores de los parámetros geofísicos que conocemos. La matriz de valores a estimar  $5 \times M$ , se transforma en una matriz  $5 \times M$  de números enteros que constituyen los „Estados“ de los valores a estimar. En la Tabla 5.6 se presenta la asignación de estado para un vector a profundidad 2308.479m del pozo 72 de forma similar a lo expuesto en la Tabla 5.4. En la asignación de estado a un vector de parámetros puede ocurrir que como para el ejemplo de la Tabla 5.6 para la profundidad 2308.631m del pozo 26 algunas variables no tengan un valor definido, en tal caso de cualquier forma se le puede asignar un estado indicando con huecos la ausencia de valores.

La aplicación de la Red Bayesiana consistirá entonces en que para cada línea de la matriz  $5 \times M$  representada por un estado  $[eGR, eMSFL, eNPFI, eRHOB, eLog(MSFL)]$  se buscará el conjunto de reglas más afines del tipo  $[eGR, eMSFL, eNPFI, eRHOB, eLog(MSFL)] \rightarrow [eSx0]$  en la Red Bayesiana correspondiente al grupo petrofísico asignado al dato.

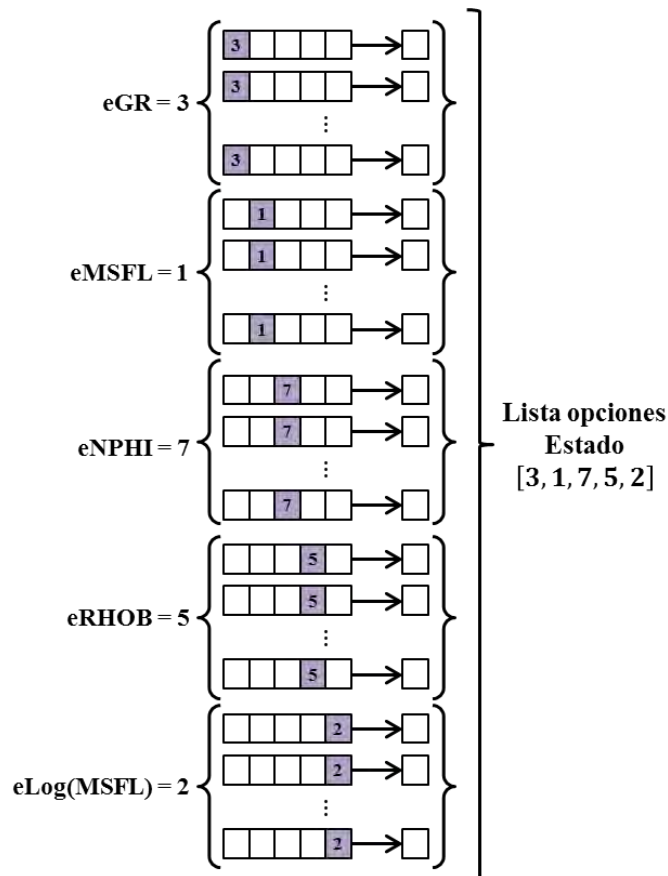


Figura 5. 11. Ejemplo de cómo se genera la lista de opciones para un estado.



El proceso de búsqueda de reglas afines de la Red Bayesiana se realiza por etapas. En la primera etapa se genera un „Listado de opciones“ bajo el criterio de unir los sub-listados de las reglas coincidentes para cada estado de cada variable de forma independiente. Por ejemplo para un vector de Estado = [3, 1, 7, 5, 2], inicialmente se realiza una primera búsqueda de estados afines considerando los casos de la Red Bayesiana que tengan  $eGR = 3$ , posteriormente se buscan los estados que tengan  $eMSFL = 1$  y así con cada variable, posteriormente estos sub-listados se unen y forman la „Lista de opciones“ para ese vector. Esta operación se ilustra en la Figura 5.11.

La segunda etapa del procedimiento de aplicación consiste en realizar un histograma de frecuencia sobre las reglas del „Listado de Opciones“ e identificar aquellas que resultan más representativas del fenómeno observado en el estado a estimar. El conjunto de reglas de mayor ocurrencia en el „Listado de Opciones“ y que mejor describen el fenómeno del estado forman el „Espectro de Opciones“ de Saturación de Agua para el Estado. Un estado que aparece como tal en la Red Bayesiana tendrá una frecuencia en el histograma del „Listado de Opciones“ de 5, correspondiente a cada vez que coincidió en la búsqueda de cada variable. Puede ocurrir también que el Estado como tal no se presente en el conjunto de reglas de la Red Bayesiana, sin embargo el „Espectro de Opciones“ quedara formado por las reglas que tuvieron mayor ocurrencia en el „Listado de Opciones“ aunque la coincidencia con el estado sea menor al de los 5 valores.

Cuando el „Espectro de Opciones“ sólo tiene un elemento este indica el valor correspondiente de „eS 0“ y asigna directamente un valor esperado para la Saturación con una desviación estándar de 0. Esto nos agrupa un conjunto de casos de las variables {GR MSFL NPHI RHOB log(MSFL)} con un comportamiento definido de  $S \times 0$  con una probabilidad de 1.

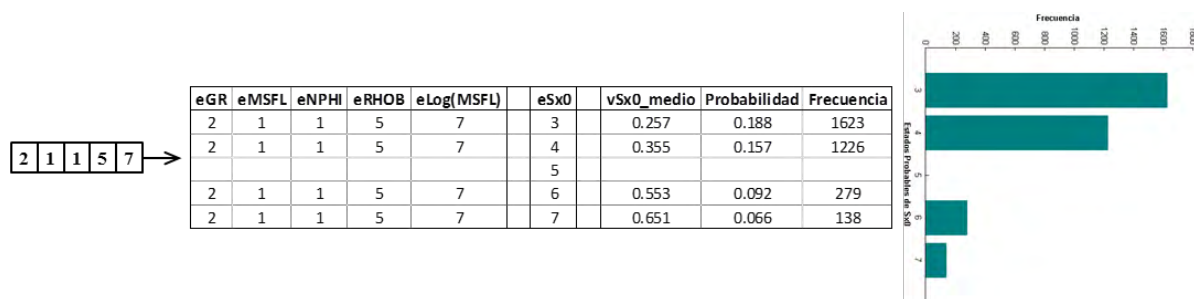


Figura 5. 12. Ejemplo de la identificación del „Espectro de Opciones“ de la Red Bayesiana *CampoB-6-10* para el Estado [2 1 1 5 2].

En la mayoría de los casos el „Espectro de Opciones“ será un conjunto de múltiples reglas que nos darán un posible comportamiento para la Saturación de Agua. Por medio de este espectro se podrá estimar el valor esperado de Saturación de Agua, así como la desviación o incertidumbre que se tiene para el valor esperado.

$$\begin{aligned} \mu_{S_{x_0}} &= \overline{S_{x_0}(eGR \ eMSFL \ eNPHI \ eRHOB \ eLog(MSFL))} \\ &= \sum_{i=1}^m S_{x_0}(eGR \ eMSFL \ eNPHI \ eRHOB \ eLog(MSFL)) \\ &\quad * p(S_{x_0} \ eGR \ eMSFL \ eNPHI \ eRHOB \ eLog(MSFL)) \end{aligned} \quad (5.5)$$

$$\begin{aligned} \sigma_{S_{x_0}}^2 &(eGR \ eMSFL \ eNPHI \ eRHOB \ eLog(MSFL)) \\ &= \sum_{i=1}^m (S_{x_0}(eGR \ eMSFL \ eNPHI \ eRHOB \ eLog(MSFL)) - \mu_{S_{x_0}})^2 \\ &\quad * p(S_{x_0} \ eGR \ eMSFL \ eNPHI \ eRHOB \ eLog(MSFL)) \end{aligned} \quad (5.6)$$

$$norma = \sum_{i=1}^m \text{Frecuencia}(S_{x_0} \ eGR \ eMSFL \ eNPHI \ eRHOB \ eLog(MSFL)) \quad (5.7)$$

En la Figura 5.12 se ejemplifica el „Espectro de Opciones“ para el Estado = [2, 1, 1, 5, 7] obtenido del fragmento de la Red Bayesiana *CampoB-6-10* mostrado en la Tabla 5.5. En el ejemplo se puede observar que el „Espectro de Opciones“ de un conjunto de posibilidades con las cuales se puede estimar un valor esperado y la desviación estándar del valor por medio de las Ecuaciones 5.5 y 5.6, donde  $m$  es el número de elementos que contiene el „Espectro de Opciones“ para el Estado  $S_{x_0}(c_i)$  es el valor de saturación medio para la caja  $c_i$ ,  $f(S_{x_0}(c_i))$  es la probabilidad de la regla para el „Espectro de Opciones“ la cual es la frecuencia indicada en la Red Bayesiana normalizada por la ecuación 5.7, que es la suma de todas las frecuencias de los elementos que conforman el „Espectro de Opciones“ para el Estado a estimar.

A partir de las Redes Bayesianas mostradas en las Figura 5.8 y 5.9 es posible recuperar la Saturación de Agua,  $S_{x_0}$ , donde se evalúan las relaciones mostradas en cada red sobre el nodo como estados. En estos casos los demás nodos se presentan como padres del de Saturación de Agua. Esta herramienta resulta entonces muy valiosa porque para casos en que sólo tengamos el valor de un solo registro es posible estimar un valor esperado posible de acuerdo a la experiencia de la Red Bayesiana y además asignar un valor de incertidumbre o desviación del dato que se está asignando.

## Resumen de Capítulo 6

En este capítulo se desarrolló de manera detallada la Metodología de Entrenamiento y Aplicación de Redes Bayesianas a un conjunto de registros petrofísicos.



## Capítulo 6

### Resultados

En este capítulo se presentan los resultados de la aplicación de las estructuras de Red Bayesiana obtenidas en el capítulo anterior Figuras 5.9.

Inicialmente se presenta la metodología de obtención de la Red Bayesiana por medio de los registros y la metodología de implementación para estimar el valor de la saturación y posteriormente se presentan los resultados de su aplicación al subconjunto de pozos de validación del CampoB indicados en la Tabla 1.2.

#### Metodología de entrenamiento de Red Bayesiana con Registros para estimar $S_x0$

1. Identificar la **Información Muestral** del conjunto de Registros Geofísicos: Rango, Valor Mínimo, Valor Máximo, Valor Medio, Mediana, Tamaño de Muestra.
2. En caso de que el conjunto de datos de entrenamiento presente valores nulos o este incompleto. Descomponer cada registro en función de su tendencia, ciclicidad, estacionalidad y función de ruido. Utilizar esta información para **Reconstruir los Registros** de tal forma que todos los datos estén completos y se tenga un mayor conjunto de casos posibles.
3. **Agrupar los Datos de Entrenamiento en Clases Petrofísicas**, es decir, en grupos de datos que compartan características petrofísicas entre ellos. Para este fin se puede considerar un análisis de litología previo o utilizar algún clasificador. En este trabajo se utilizó Fuzzy C-Means Clustering.
4. **Aplicar el Método de Archie-Pickett** para calcular la curva teórica de Saturación de Agua al nuevo conjunto de datos.
5. Identificar **Relaciones de Probabilidad** entre los registros para cada clase petrofísica en base al análisis de gráficos cruzados. Proponer el Grafo de las Relaciones de probabilidad.
6. **Discretizar las Variables**. Muestreo de Variables.
7. **Asociar un Estado** a cada combinación posible de las líneas que conforman la muestra.

8. **Generar un Conjunto de reglas** en base al conjunto de estados posibles presentes para cada clase petrofísica que conforma la muestra y al grafo de registros.
9. **Estimar la Probabilidad Condicional** de los casos para cada regla que conforma la Red Bayesiana.

## Metodología de aplicación de Red Bayesiana a Registros para estimar $Sx0$

1. Identificar la **Información Muestral** del conjunto de Registros Geofísicos: Rango, Valor Mínimo, Valor Máximo, Valor Medio, Mediana, Tamaño de Muestra.
2. Del conjunto de datos, **clasificar los datos en las clases petrofísicas correspondientes**, es decir, que compartan características petrofísicas entre ellos con las clases consideradas para los datos de entrenamiento. Se puede considerar un análisis de litología previo o utilizar algún clasificador como Fuzzy C-Means.
3. **Discretizar las variables** con el mismo No. de Cajas con el que se entrenó la Red Bayesiana.
4. **Asociar un estado** a cada combinación posible de registros.
5. **Buscar reglas afines** dentro de la estructura de Red Bayesiana para cada estado que se desee estimar.
6. Generar un „**Espectro de Opciones**“ para el estado.
7. Estimar en base al Espectro de Opciones“ el **Valor Esperado** y el **Rango de Desviación** del valor.

## Comparando resultados

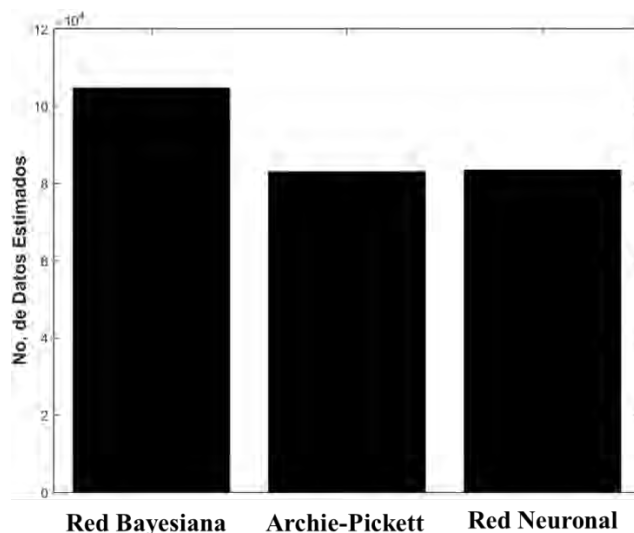
En la Sección 4.1 se planteó la problemática para inferir el valor de la Saturación de Agua en pozos, también se describieron algunas soluciones alternas que se han propuesto en algunos trabajos de investigación para estimar su valor, el paradigma comúnmente utilizado ha sido la aplicación de Redes Neuronales (ANN, por sus siglas en inglés Artificial Neural Network), el tipo más común de aplicación es el de Perceptrón Multicapa [22][24][25][26].

Para tener un algoritmo adicional contra el cual comparar los resultados se entrenó una Red Neuronal tipo Perceptrón Multicapa. Las estructuras de red utilizadas se obtuvieron de la aplicación de la función de MATLAB **fitnet**, la cual se puede estructurar y entrenar con los datos propuestos por el usuario.

La estructura que se implementó para la Red Neuronal **fitnet** de MATLAB estaba conformada por una capa oculta de 50 neuronas que se entrenó en un ciclo de 1000 épocas. Esta estructura se entrenó para cada clase petrofísica con el mismo conjunto de datos de entrenamiento que se utilizaba para las redes bayesianas, a fin de que la comparación de los resultados tuviera el mismo sentido.

Se entrenaron las Redes Bayesianas considerando los diferentes grupos petrofísicos asociados a los datos, además de diferentes  $k$ -No. Cajas o Grupos de Muestreo de discretización al listado de pozos de entrenamiento.

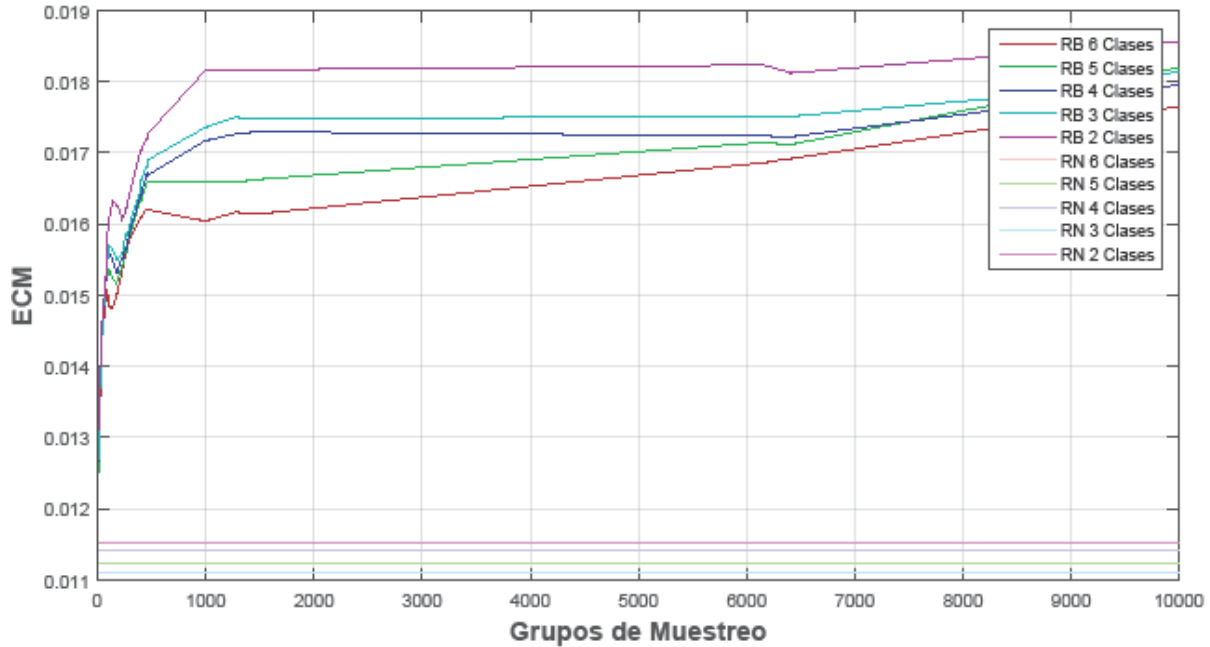
Los resultados presentados son los resultados generales de la aplicación de la Red Bayesiana y la Red Neuronal para estimar  $S_{x0}$  en todos los pozos que conforman el conjunto de datos de Aplicación del CampoB.



**Figura 6. 1. No de Datos Medio estimado con la Red Bayesiana, Archie-Pickett y Red Neuronal**

Aunque se tenga datos de los cuatro registros utilizados en la Metodología de estimación por medio de la Ecuación de Archie esto no garantiza que se tenga un valor entre 0 y 1 para la Saturación de Agua. En casos de no tener registro de porosidad o resistivo significa que no se puede evaluar la Ecuación. Por tales motivos el No. De Datos de Saturación de Agua estimados con la Ecuación de Archie no necesariamente coincide con el No. De vectores de estado a los cuales se quiso asociar un valor. Un caso similar pasa con la aplicación de la Red Neuronal que requiere tener un vector completo de valores para todos los registros que se desean asociar con un valor de Saturación de Agua.

En la Figura 6.1 se presentan No. de Datos de Saturación de Agua Medio que se pudieron estimar con cada Método utilizado en este trabajo: Red Bayesiana, la Ecuación de Archie y la Red Neuronal. Uno de los resultados más importantes es que en comparación la cantidad de datos que se pudieron estimar por medio de la „Red Bayesiana“ fue mucho mayor que los estimado por „Archie-Pickett“ o la „Red Neuronal“ lo cual es una de las virtudes de las „Redes Bayesianas“ y es poder dar un valor esperado y una incertidumbre del valor de acuerdo a la experiencia de los casos que conforman las reglas de la Red Bayesiana aún *sólo con el valor de un registro*.



**Figura 6. 2. Comparación del Error Cuadrático Medio de la Red Bayesiana y la Red Neuronal con los resultados de la Ecuación de Archie.**

En la Figura 6.2 se presenta el Error Cuadrático Medio de la aplicación de las Redes Bayesianas considerando 6, 5, 4, 3 y 2 Clases Petrofísicas, también se presenta el Error Cuadrático Medio de la Red Neuronal considerando también 6, 5, 4, 3 y 2 Clases Petrofísica. Se observa que aunque los Errores Obtenidos con la Red Bayesiana son mayores a los de la Red Neuronal estos se encuentran en el mismo rango de valores.

Las medidas del Error Cuadrático Medio entre curvas se calcularon como la media de los cuadrados de los errores

$$MSE = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2 \quad (6.1)$$

En donde  $N$  es el número de valores,  $x_1, x_2, \dots, x_N$  son los valores obtenidos con la Red Bayesiana o la Red Neuronal y  $y_1, y_2, \dots, y_N$  son los valores obtenidos con la Ecuación de Archie.

En la Figura 6.3 se presenta la comparación entre los resultados de Saturación de Agua obtenidos con la Red Bayesiana y la Red Neuronal. Estos Errores son de mucho menor rango que los observados en el Figura 6.2 de su comparación con Archie-Pickett.

En todos los casos los Errores se estabilizan después de 1500 grupos de muestreo. Esto concuerda con que el valor mínimo de datos en alguna de las clases petrofísicas es de 1520. Eso es coherente ya que para este tipo de clase petrofísica el no. de datos disponibles de entrenamiento son un límite en los grupos de muestreo posibles.

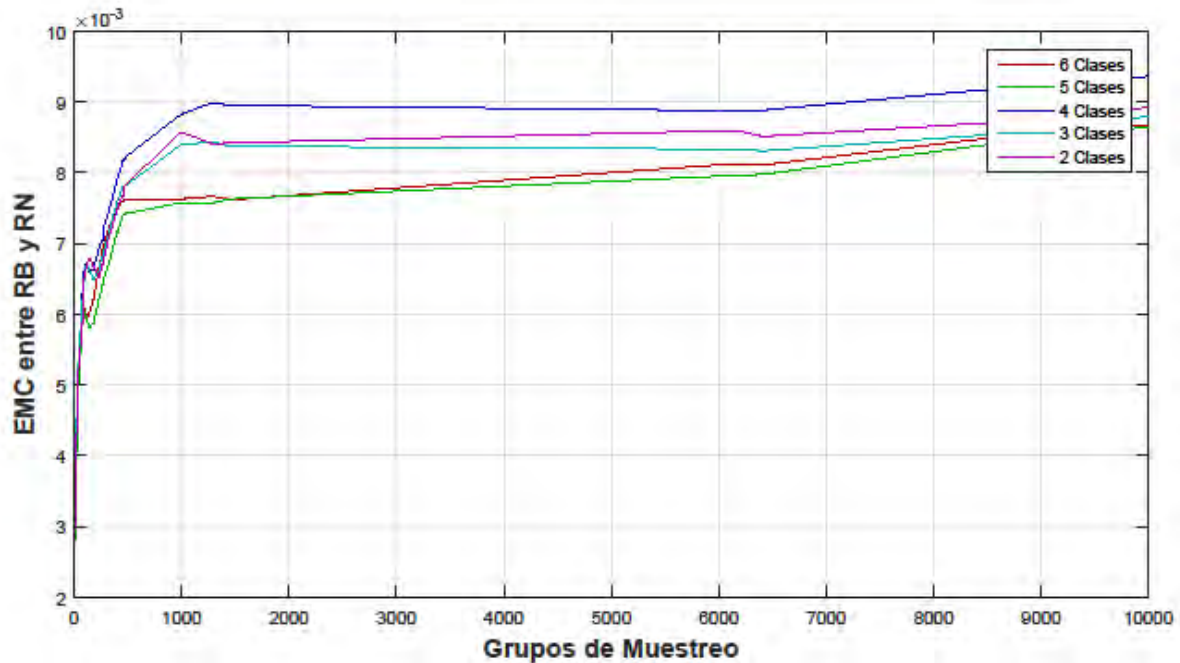


Figura 6. 3. Error Cuadrático Medio entre la curva de Saturación de Agua obtenida con las Redes Bayesianas y la obtenida con las Redes Neuronales.

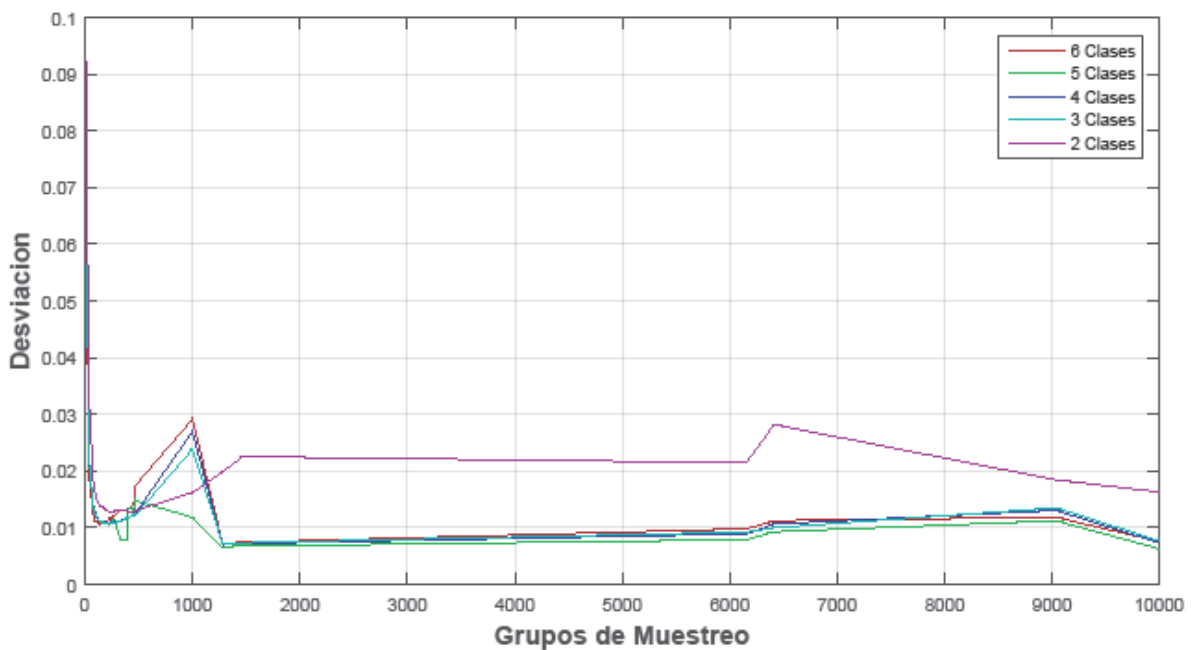


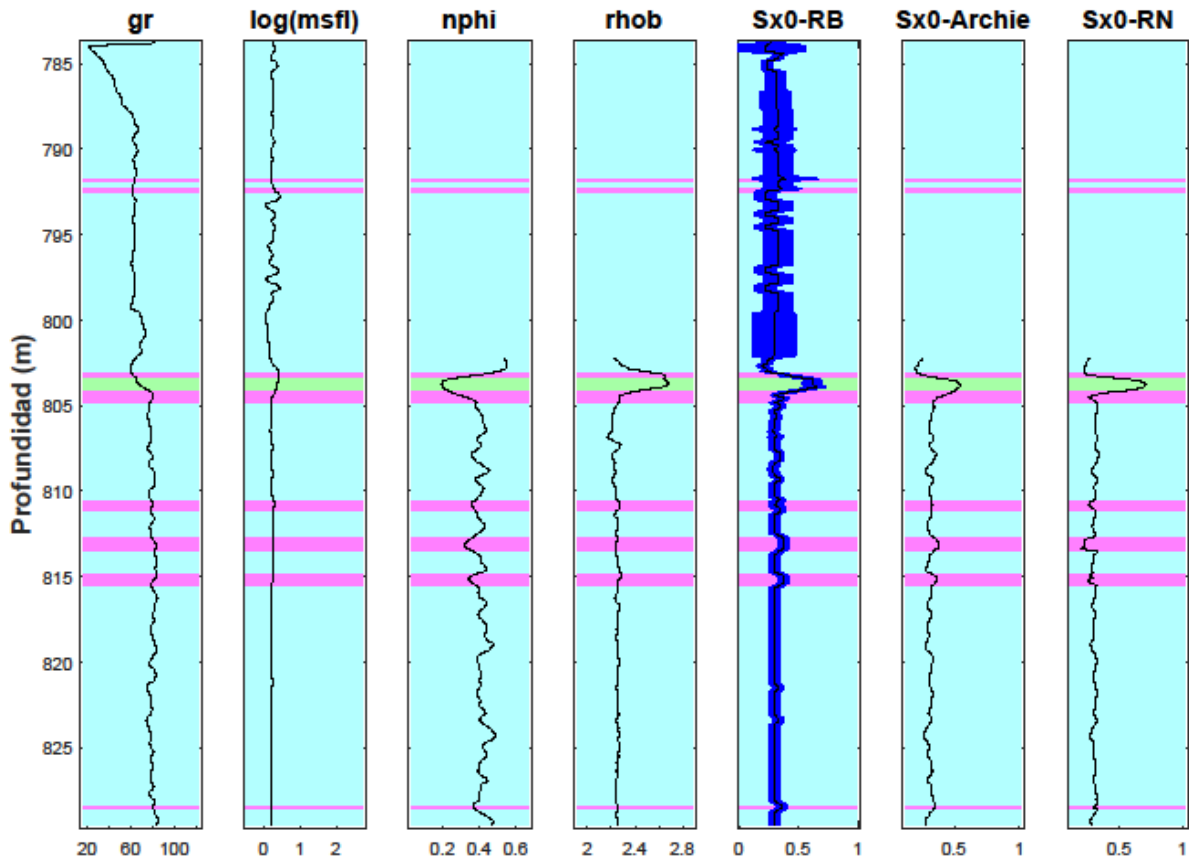
Figura 6. 4. Rango de Desviación Medio del Valor Esperado de Saturación de Agua estimado con la Red Bayesiana

Por medio de la Ecuación 5.6 se describe la forma en que se calcula la desviación del valor esperado que se calcula por medio del „espectro de opciones“ que ofrece la Red Bayesiana cuando se desea estimar el valor correspondiente a un estado. En la Figura 6.4 se



presenta el Rango de Desviación Medio que se estima para las Redes Bayesianas considerando 6, 5, 4, 3 y 2 Clases Petrofísicas y diferentes Grupos de Muestreo. Este rango muestra sus mejores valores para Grupos de Muestreo menores a 1500.

A continuación se presentan algunas gráficas de los pozos de validación a los cuales se les aplico las tres Metodologías para estimar la Saturación de Agua. Para mostrar de forma más directa los resultados hasta ahora presentados.



**Figura 6. 5. Fragmento del pozo 67 considerando 5 clases petrofísicas y 10 grupos de muestreo. Desviación = 0.07878, ECM de RB = 0.00097 y ECM de RN = 0.00188.**

## Descripción de Figuras de Resultados

Todas las figuras que presentarán los resultados en comparación tienen la misma estructura:

- Los primeros cuatro carriles corresponden a los registros utilizados para la estimación de Saturación de agua.
- El quinto carril presenta la curva de Saturación de Agua estimada con la Red Bayesiana acompañada de su desviación estándar en una sombra azul.

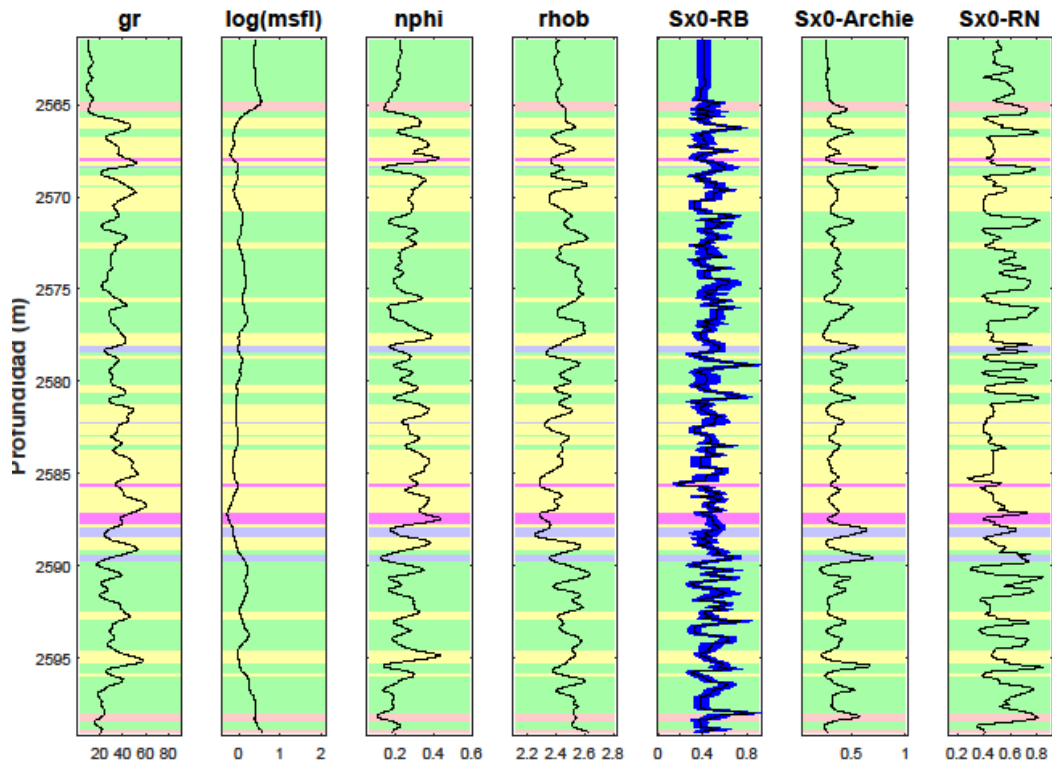


Figura 6. 6. Fragmento del pozo 41 considerando 6 clases petrofísicas y 10 grupos de muestreo. Desviación = 0.06783, ECM de RB = 0.02296 y ECM de RN = 0.04452.

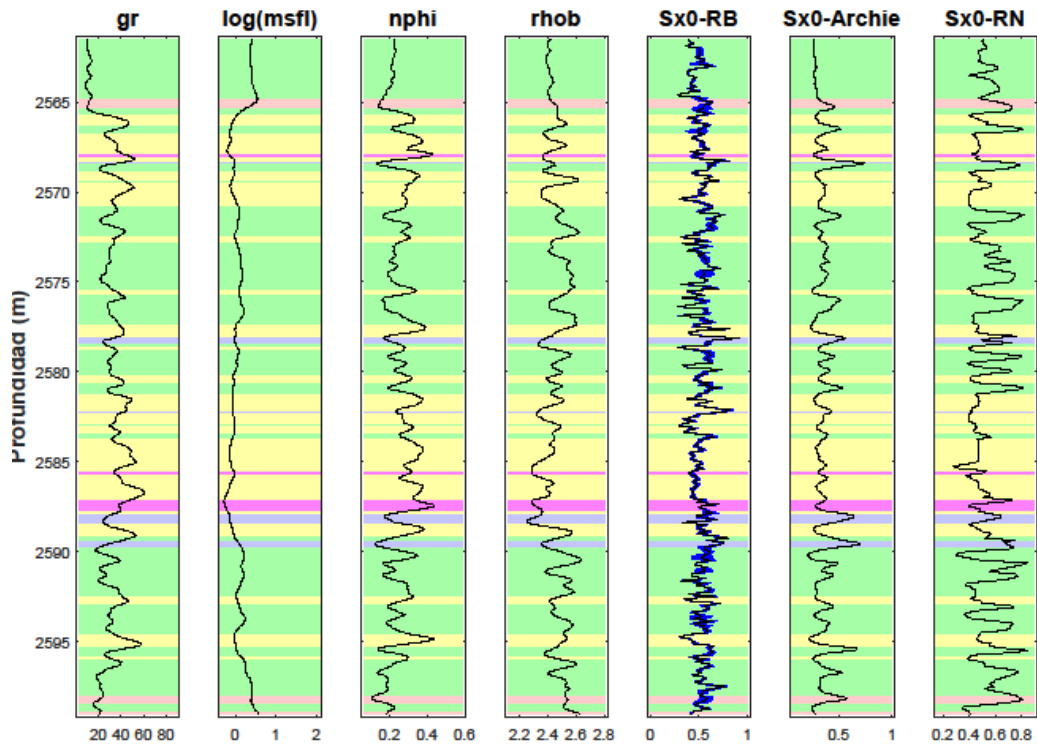


Figura 6. 7. Fragmento del pozo 41 considerando 6 clases petrofísicas y 329 grupos de muestreo. Desviación = 0.02068, ECM de RB = 0.04292 y ECM de RN = 0.04452.

- El sexto y séptimo carril presentan las curvas de Saturación de Agua estimadas con la Ecuación de Archie y la Red Neuronal respectivamente.
- Al fondo de cada carril se indica en bloques de color uniforme la correspondencia de cada intervalo con las clases petrofísicas. El código de color utilizado es el de la Tabla 5.3.

En la Figura 6.5 se presenta un ejemplo correspondiente al resultado expuesto en la Figura 6.1, que es, casos donde la Red Bayesiana es capaz de inferir un posible valor para la Saturación de Agua cuando la Ecuación de Archie o la Red Neuronal no pueden. En el caso de la Figura 6.5 se debe a la falta de información de los registros de porosidad y densidad. También se muestra que tener menos información para inferir el valor de la Saturación de agua se traduce en tener un rango de desviación más grande del resultado.

## Discusión

La identificación de relaciones de probabilidad por medio de los gráficos cruzados de variables fue el proceso más importante en este trabajo para la extracción de la estructura de red. Este tipo de análisis se ha realizado con anterioridad para encontrar las relaciones en artículos como los expuestos en el estado del arte, pero por lo revisado en este trabajo se presentan por primera vez con un enfoque probabilístico que desde este trata de resolver la problemática de la estimación de la Saturación de Agua.

El proceso de discretización de las variables para la cuantificación de estados posibles de las variables puede resultar subjetivo; lo que se buscó en este trabajo es conseguir estados representativos y al mismo tiempo variados, es decir, proponer un valor de discretización donde unos muy pocos puntos se encuentren en cada caso de miles no resulta en una buena aproximación general, tampoco lo es una en que una sola banda englobe más de la tercera parte de la población de datos, en este caso considerar la el número de cajas,  $k$ , como la raíz del número total de datos resulto bastante conveniente en este trabajo.

Al elegir determinadas variables de entrada es posible que los algoritmos de extracción de la Red Bayesiana generen estructuras adulteradas, en cuyos casos los arcos entre variables no reflejan relaciones de causalidad o dependencia Sin embargo, es posible realizar un tipo de validación asignando valores conocidos en la recuperación de cualquiera de los nodos que se muestre conectado en la red. De forma que en caso de que los resultados del registro recuperado no asemeje la información esperada esto será un indicio de que la estructura obtenida no es del todo valida, o al menos no refleja relaciones suficientemente fuerte entre las variables.

Las estructuras de red bayesianas generadas en este trabajo están basadas en valores discretos los cuales generan estados que pueden ser identificados con un valor esperado de Saturación de Agua para cada caso.

Aunque es posible utilizar la misma Red Bayesiana con diferentes parámetros de discretización, la calidad de los resultados disminuye debido a la pérdida de dependencia entre los datos discretizados.



## Capítulo 7

### Conclusiones

En este trabajo, se desarrolló una metodología de construcción de una red bayesiana como parte del análisis muestral de registros geofísicos. Aunque la Red Bayesiana se utilizó para inferir el valor de la Saturación de Agua, por el Teorema de Bayes es posible inferir una probabilidad condicional en función de la estructura de red para cualquiera de los nodos que la integran.

A continuación se enumeran las conclusiones y contribuciones realizadas en este trabajo de tesis para el cálculo de Saturación de Agua a partir de Registros Geofísicos.

1. Para el Método Gráfico de Pickett, que es el análisis del cúmulo de puntos formado por una gráfica log-log del registro resistivo MSFL, vs la porosidad efectiva  $e$ , donde se identifica una línea recta „secante“ al cúmulo; se propusieron cinco maneras diferentes para seleccionar esta recta „secante“ de forma automatizada de acuerdo a la forma geométrica del cúmulo de puntos, los métodos propuestos fueron: Cúmulos Extremos, Ajuste Lineal, Iterativo, Imputación de  $m = 2$  y Componentes Principales. (Ver página 46, Aplicación de Archie-Pickett para obtener la curva teórica de Saturación de Agua) Proponer estos métodos para la identificación de la recta de 100% de Saturación de Agua ayudo a que el análisis gráfico de Pickett no fuera subjetivo y pudiera replicarse. Estas propuestas fueron de mucha utilidad para la implementación de Archie-Pickett para la estimación de la curva de Saturación de Agua teórica que posteriormente se utilizó en este trabajo para la identificación de las reglas que constituyeron la Red Bayesiana.
2. La reconstrucción de datos para cada registro de cada pozo que conformó el CampoB por medio de la identificación de la Tendencia, Ciclicidad, Estacionalidad y Ruido que conformaba cada curva no sólo sirvió para generar un nuevo conjunto de datos sintéticos a los cuales llamamos „Datos Monte Carlo“. (Ver página 50 Metodología para la construcción de Datos con Monte Carlo) Si no que considerando cada registro tratado como una Serie de Tiempo se le pudieron asociar dichas componentes o características. Estas componentes pueden tener relación con algunos tratamientos de geofísica como la

*Tendencia* con la corrección de temperatura que se realiza a los registros, la *Cicilicidad* con la formación de las capas de diferentes tipos de roca o litología que se repiten en diferentes periodos en el subsuelo o los *Cambios Aleatorios* o abruptos en los registros con la presencia de fracturas. Por lo que la reconstrucción de datos siguiendo esta metodología pone en consideración los posibles efectos con los que se pueden topar las herramientas encargadas de realizar los registros.

3. En este trabajo se propuso un análisis gráfico para la identificar la relación de probabilidad condicional entre las variables aleatorias o nodos que conforman la Red Bayesiana. Las relaciones se obtuvieron de analizar la tendencia de la nube de puntos de acuerdo al criterio expuesto en la Página 69, Identificación de relaciones de probabilidad. Se observó que las estructuras de las Redes Bayesianas obtenidas para CampoB por medio de esta metodología resultaron ser muy parecidas entre sí, aunque las regiones de mayor densidad de los cúmulos analizados para cada crossplot de parejas de registros se presentaban en regiones diferentes. En particular las que involucraban a NPHI y MSFL. En este punto se consideró conveniente proponer una estructura de Red Bayesiana por Clase Petrofísica.
4. En este trabajo se presentó la extracción de una Red Bayesiana a raíz de un conjunto de registros geofísicos (Ver Página 82, Entrenado la Red Bayesiana) y la aplicación del conjunto de reglas condicionales obtenidas para la estimación de la Saturación de Agua. (Ver Página 85, Aplicando la Red Bayesiana) Donde cada estado considerado se presentaba como una combinación  $[x_1, x_2, x_3, x_4, x_5]$  con una probabilidad  $p(S_{x0} | GR = x_1, MSFL = x_2, NPHI = x_3, RHOB = x_4, \log(MSFL) = x_5) | x_5 = \log(x_2)$  fue estimado por medio de la Red Bayesiana originando lo que en este trabajo se llamó el „Espectro de Opciones“ y por medio del cual fue posible asignar un valor esperado para la Saturación de Agua y una desviación del valor esperado.
5. El modelo de Red Bayesiana que se presenta es capaz de capturar el patrón de relación entre  $S_{x0}$  y los registros utilizados  $\{gr, msfl, nphi, rhob\}$ .
6. La comparación con los resultados del método Archie-Pickett mostraron rendimientos consistentes para ambos campos.
7. El Modelo de Red Bayesiana puede determinar con cierta precisión el valor de  $S_{x0}$  en los casos en que el método de Archie-Pickett no pudo ser utilizado, ni la Red Neuronal, porque sólo necesita el valor de uno de los parámetros para identificar un „Espectro de Opciones“ que permite sugerir un valor estimado posible. Es importante recalcar que la precisión en el valor esperado surgido de la evaluación de la Red Bayesiana dependerá de tener la mayor cantidad de información posible, es decir, valores para cada parámetro con el que se conforma el Estado, también recalcar que entre menos información se tenga mayor será la desviación o incertidumbre del resultado arrojado por la Red Bayesiana.
8. El modelo de red bayesiana es atractivo debido a su simplicidad y facilidad de implementación. Además, es posible identificar las relaciones entre los casos

significativos resistividad, porosidad, y la Saturación de Agua. Algo que tampoco ya no se desarrolló en este trabajo es la agrupación de reglas de las Redes Bayesianas relacionándolas con características de Rocas ya conocidas o Litologías. En este caso la Red Bayesiana puede ser usada para identificar patrones y realizar trabajo de composición y estructura en la roca cortada por el pozo.

## 7.1. Logros

La aportación más importante de este trabajo es proporcionar un enfoque alternativo a la solución de algunos cálculos que se realizan en la Interpretación de Registros. En particular en este trabajo se está proponiendo para tal propósito una solución de tipo probabilístico.

Se aplica un modelo de Inteligencia Artificial muy estudiado [27], que es el de Redes Bayesianas, a un área donde previamente no se ha hecho el planteamiento y por tanto no se ha explotado la riqueza de información que se puede obtener de este tipo de modelos.

Se ha automatizado el proceso de aplicación de la Ecuación de Archie, para este fin se generó una serie de programas en MATLAB que trabajan en conjunto con una arquitectura de Base de Datos adecuada para el tratamiento de los datos de registros en Pozos.

Como parte del análisis de las relaciones entre registros se implementaron Mapas Auto Organizados para identificar los patrones en los gráficos de Pickett, lo cual permitió una mejor agrupación de los tipos de nubes de datos que se utilizan en la metodología de Archie-Pickett.

Se ha logrado presentar un análisis estadístico detallado de los registros geofísicos presentes en la inferencia del valor de la Saturación de Agua.

Como parte de la comprobación de veracidad de la curva resultante del modelo de Red Bayesiana propuesto se diseñó una Red Neuronal Artificial de tipo Perceptrón Multicapa.

Se ha propuesto un nuevo método estadístico para analizar los efectos de los registros en el valor de la Saturación de Agua, el cual hasta ahora no ha sido explorado en el área de la petrofísica, es decir, la aplicación de Redes Bayesianas.

## 7.2. Trabajo Futuro

Este trabajo sirve como fundamento para la conciliación entre las relaciones observadas en los modelos de Inteligencia artificial y las observadas en las ecuaciones físicas clásicas de la Saturación de Agua.



El grafo de red bayesiana para inferir Saturación de Agua puede ser utilizado para inferir otras propiedades de la caracterización de yacimientos como la permeabilidad, la saturación de gas o aceite, los espesores de capas y la identificación de litologías.

Como un trabajo futuro se desea proponer un modelo replicable, para ser aplicado en cualquier yacimiento y se desea abrir un nuevo paradigma en la construcción de modelos de la geofísica.

En este trabajo no se profundizó en el tema de identificar cual era el mejor valor de  $k$ -No. Cajas para discretizar las variables y obtener una mejor aproximación en los valores de Saturación de Agua estimados por la Red. Sobre esto, se puede profundizar en trabajos posteriores. Este es uno de los temas en los que más se podría profundizar en este trabajo a futuro.

Se pueden considerar en trabajos futuros que las variables sean continuas

Este trabajo puede ser extrapolado a trabajos de Sísmica y Exploración de Hidrocarburos.

## Apéndice A

### Descripción de registros geofísicos [5]

#### Registro potencial natural (SP)

Es el registro de la diferencia del potencial eléctrico entre un electrodo colocado en la superficie en la presa del lodo y un electrodo móvil en el agujero. La diferencia de potencial entre los electrodos se registra en función de la profundidad. Los potenciales naturales se desarrollan dentro del agujero como resultado de la diferencia de salinidad entre el fluido de perforación (lodo) y el agua de la formación.

Las unidades en que se mide el registro son los millivolts (mv). No existe un cero absoluto; sólo cambios en el potencial.

El principal uso de la curva SP es para seleccionar zonas permeables, obtener la resistividad del agua y estimar el grado de arcillosidad

Este registro puede utilizarse para estimar la resistividad del agua de formación, entre los pasos que suelen utilizarse con este fin está el establecer un límite conocido como la **línea base de lutitas** y a partir de estas líneas identificar las deflexiones, donde las deflexiones a la izquierda indican la presencia de un fluido más salado. Esta forma suele utilizarse para zonificar.

#### Registro rayos gamma (GR)

Mide la radiactividad natural de las rocas atravesadas por el agujero. Este registro es causado por el proceso de decaimiento radiactivo natural que ocurre en los minerales de las rocas, esto es los rayos gamma producidos por el potasio 40 y el decaimiento de los productos de las series del Uranio y Thorio. Los tres elementos contribuyen igualmente en la respuesta de registro.

Las lutitas son generalmente más radioactivas que las areniscas, calizas, dolomías, anhidritas, etc. Por este motivo suele ser un indicador efectivo de arcillosidad, para marcar zonas limpias.

La curva que se obtiene se puede correlacionar con la SP, cuando se corre en lodo dulce conductor. Se ocupa también para hacer correlaciones estratigráficas. Es efectiva donde la curva SP no se puede obtener. Se puede aplicar para determinar la cantidad de arcilla encontrada en las formaciones.

La calibración de las herramientas se hace en secuencias de calizas en unidades API. Se define una unidad API como la milésima parte de la respuesta radiactiva de una roca tipo que se localiza en un pozo de calibración en la Universidad de Houston.

### **Registros de resistividad eléctrica profunda (LLD, LLS)**

Las curvas de resistividad se obtienen por medio de una herramienta conformada por dos electrodos de corriente y dos de potencial que están colocados a diferente espaciamiento dentro de una sonda. El espaciamiento de los electrodos proporciona la profundidad de penetración de la corriente.

La **resistividad** se define como la medida de la dificultad que la corriente eléctrica encuentra al circular a través de un material. Es una de las propiedades más importantes en Registros de Pozos. Se mide en ohm-m.

Las herramientas de resistividad que se utilizan actualmente tienen más de cuatro electrodos, se les conoce como registros Laterolog o de Guarda. El propósito que tienen es el de enfocar la corriente dentro de la formación y reducir los efectos de capas laterales y de agujero. La profundidad de penetración de la corriente depende del espaciamiento de los electrodos. Tienen la ventaja de que puede correrse más de un arreglo simultáneamente. Se puede observar y cuantificar la invasión del filtrado del lodo.

Ejemplo de estos registros son el Laterolog Deep Resistivity, LLD, y el Laterolog Shallow Resistivity, LLS.

### **Registro de resistividad eléctrica somera (MSFL)**

Existen varios arreglos que dependen del tipo de lodos que se utilicen. Este tipo de registros fueron diseñados para medir  $R_{xo}$ . Este tipo de herramientas son muy semejantes a las de resistividad convencional pero difieren en que los electrodos van montados sobre una almohadilla que va pegado a la pared del pozo. Los registros de este tipo son utilizados para ayudar a corregir efectos de invasión y en otros casos para estimar la porosidad.

Las herramientas con esta configuración son Microlog, Microlaterolog, Proximidad y Microesférico Enfocado.

Tienen una profundidad de investigación muy somera con el propósito de medir la resistividad de la zona de flujo ( $R_{xo}$ ). La razón por la que se les llama dispositivos de porosidad es porque la relación  $R_{xo}/R_{mf}$  da el factor de formación.

### **Registro de neutrones (NPHI)**

El registro de neutrones mide la habilidad de las formaciones de atenuar el paso de neutrones a través de estas.

Los registros de neutrones son de dos tipos. El primero tiene una fuente y un detector, es un sistema que va pegado a la pared de pozo. El receptor en este caso mide neutrones epitermales de la fuente.

El segundo y más común tiene una fuente y dos receptores que miden neutrones termales, tiene un sistema que compensa por rugosidad y diámetro de agujero. En ambos casos el principio es el mismo.

Los detectores miden el flujo de neutrones a un nivel de energía particular, es decir, cuanta energía de los neutrones ha disminuido a partir de la fuente. La alta energía de los neutrones disminuye por la colisión con los núcleos atómicos de varios elementos en la formación. Los elementos que tienen una masa atómica muy cercana a la de los neutrones tienden a detenerlos como el hidrógeno. Es por ello que se dice que la herramienta mide el contenido de hidrógeno que es proporcional a la porosidad. La respuesta se calibra en rocas saturadas con agua. Cuando hay gas en la formación la atenuación de los neutrones es menor, en esta situación el registro da valores pesimistas.

En zonas arcillosas y con yeso la herramienta da un valor mayor de porosidad.

### **Registro de densidad (RHOB)**

Este registro mide la densidad total de la formación. El equipo consta de una fuente emisora y dos detectores de rayos gamma. El flujo de rayos gamma que se recibe en los detectores es función de la dispersión producida por el efecto Compton y que a su vez es función de la densidad electrónica de la roca.

Los rayos gamma tienen la particularidad de viajar en línea recta hasta que son desviados, moderados o absorbidos por la formación. El material de que está constituida la roca de la formación tiene una cierta densidad electrónica. La densidad total de la formación es función de la densidad del material que constituye la matriz de la roca, de su porosidad y de la densidad de los fluidos contenidos en ella. Esto es, la intensidad de rayos gamma registrada por el detector será inversamente proporcional a la densidad de la formación.

Un detector cercano a la fuente se ocupa de compensar los efectos de espesor de enjarre y rugosidad del agujero, por lo que la presentación del registro incluye una curva  $\rho_b$  y otra  $\rho_b$  que es la cantidad de corrección que fue necesaria aplicar vía el detector corto para obtener  $\rho_b$ ; sirve como control de calidad del registro.

### Registro sónico (DT)

El registro mide el menor tiempo que requiere una onda compresional para recorrer verticalmente un pie de longitud a través de la roca adyacente a la pared del pozo. La unidad de medida es el tiempo de tránsito.

Las ondas que son de interés son las compresionales y transversales. Las ondas compresionales son aquellas en que la partícula vibra en la misma dirección de la propagación, paralelas al eje del pozo. Las transversales son aquellas en las que las partículas vibran perpendiculares al eje del pozo. Las ondas compresionales viajan aproximadamente 1.7 veces más rápido que las transversales.

El tiempo total que requiere un pulso para viajar a través de la roca es proporcional a la cantidad de fluido en el espacio poroso y a la cantidad de matriz. La ecuación que relaciona la medición del tiempo de tránsito del registro con la porosidad es:

$$= \frac{\Delta t - \Delta t_{ma}}{\Delta t_f - \Delta t_{ma}} \quad (2.3)$$

Donde  $\rho$  es la porosidad,  $\Delta t$  es el tiempo de tránsito medido por el registro,  $\Delta t_{ma}$  es el tiempo de tránsito de la matriz y  $\Delta t_f$  es el tiempo de tránsito del fluido. Los dos últimos valores se obtienen de acuerdo al tipo de roca y al tipo de fluido.

La herramienta es más efectiva en formaciones consolidadas y compactas. Donde la roca no está consolidada, el tiempo de transito está influenciado por los fluidos de la formación y por la arcillosidad.

Ya que esta herramienta mide el primer arribo de los pulsos acústicos, estos normalmente se entranpan en alguna cavidad o fractura, retardando el tiempo de llegada de la onda. El rango óptimo de efectividad de la sonda esta entre 5 y 20% de porosidad.

## Apéndice B

### Tablas descriptivas de los datos que conforman el CampoB

Datos Originales - CampoB						
Némonico	Tipo de registro	Unidades	No. Puntos	Mínimo	Máximo	Valor Medio
MSFL	Resistivo Micro Esférico	ohmm	220030	0.26	1950	3.05
GR	Rayos Gamma	api	220143	-10068.09	184.13	50.45
DT	Sónico	us/ft	199360	10.74	116.03	105.15
RHOB	Densidad	gr/cm3	167834	-4018.67	3.01	2.38
NPHI	Porosidad Neutrón	vol/vol	168024	-0.01	0.82	0.31

Tabla B. 1. Información muestral de los registros medidos en los pozos del CampoB para los datos Originales.

Datos Limpio - CampoB					
Registro	No. Datos	Min	Max	Promedio	Std
gr	219497	6.03	184.13	58.72	13.04
msfl	217386	0.26	1950	2.96	21.49
nphi	167509	0.0002	0.82	0.31	0.11
rhob	167666	1.87	3.01	2.41	0.1
log(msfl)	217386	-0.59	3.29	0.3	0.26
sx0	162745	0.01	1	0.41	0.14

Tabla B. 2. Información muestral de los registros medidos en los pozos del CampoB para los datos Limpios.

Datos Limpios - CampoB															
Pozo	GR					MSFL					NPHI				
	Datos	Min	Max	Prom	Std	Datos	Min	Max	Prom	Std	Datos	Min	Max	Prom	Std
38	24359	9.85	184.13	55.66	10.38	24359	0.37	1950	3.9	56.17	9794	0.001	0.35	0.11	0.06
40	1405	30.37	101.22	54.82	14.01	1405	0.48	31.69	4.27	3.64	1405	0.07	0.59	0.23	0.07
41	20514	6.03	86.02	56.15	12.23	20491	0.52	101.23	2.39	4.82	10404	0.09	0.57	0.38	0.06
42	23260	12.36	74.67	52.02	10.11	23246	0.4	915.42	2.2	8.98	19374	0.04	0.59	0.35	0.08
43	18505	19.03	84.98	60.71	11.23	18722	0.42	86.39	2.41	3.08	14671	0.06	0.56	0.35	0.08
44	20854	32.37	165.2	60.74	8.38	20754	0.41	110.89	2.38	2.73	16836	0.02	0.56	0.34	0.09
47	1346	15.44	117.42	47.38	18.13	1346	0.37	124.96	5.22	8.35	1346	0.08	0.61	0.25	0.09
50	20078	26.84	95.43	68.89	9.04	19976	0.3	276.17	2.43	5.01	15724	0.0002	0.82	0.36	0.1
53	4502	23.55	106.44	51.66	14.14	4498	0.26	38.69	3.41	3.12	4502	0.05	0.48	0.23	0.08
55	16021	7.02	91.03	44.51	10.24	16050	0.38	1934.03	4.47	33.75	11770	0.02	0.53	0.3	0.1
56	3544	37.74	113.24	64.08	10.28	3545	0.48	83.49	4.08	6.24	3451	0.04	0.6	0.24	0.1
58	15795	33.76	126.83	64.81	8.68	15787	0.36	57.25	2.53	3.49	15808	0.03	0.6	0.37	0.11
62	11208	14.52	101.31	66.03	11.17	8883	0.54	138.7	1.99	4.1	8882	0.05	0.54	0.37	0.07
63	8335	27.73	128.74	56.75	12.31	8335	0.46	59.91	4.43	4.36	8335	0.02	0.51	0.21	0.08
65	1347	15.98	92.55	53.24	17.87	1347	0.75	70.91	4.64	5.99	1347	0.06	0.44	0.24	0.09
67	21053	21.14	117.28	67.22	12.49	21074	0.43	429.85	2.45	6.42	16404	0.05	0.64	0.34	0.09
68	1511	27.75	97.13	51.72	14.88	1707	0.31	55.79	5.14	4.17	1595	0.02	0.53	0.15	0.06
70	3741	23.5	111.08	49.17	14.8	3741	0.39	58.75	5.6	4.79	3741	0.01	0.48	0.13	0.08
72	2119	24.02	102.19	46.57	15.66	2120	0.76	28.56	4.27	2.71	2120	0.04	0.52	0.21	0.08

Datos Limpios - CampoB															
Pozo	RHOB					log(MSFL)					Sx0				
	Datos	Min	Max	Prom	Std	Datos	Min	Max	Prom	Std	Datos	Min	Max	Prom	Std
38	10181	1.87	2.75	2.48	0.08	24359	-0.43	3.29	0.24	0.24	6972	0.31	1	0.58	0.19
40	1405	2.24	2.82	2.51	0.08	1405	-0.32	1.5	0.53	0.28	1243	0.18	1	0.67	0.16
41	10450	2.23	2.69	2.38	0.05	20491	-0.28	2.01	0.25	0.23	10404	0.1	0.96	0.2	0.05
42	19368	2.01	2.8	2.37	0.09	23246	-0.4	2.96	0.24	0.21	19337	0.16	0.98	0.34	0.07
43	14726	2.03	2.69	2.37	0.09	18722	-0.37	1.94	0.29	0.22	14659	0.12	0.91	0.43	0.06
44	16826	2.04	2.72	2.38	0.09	20754	-0.39	2.04	0.29	0.21	16495	0.25	1	0.5	0.09
47	1346	2.08	2.75	2.46	0.09	1346	-0.43	2.1	0.46	0.43	1309	0.11	1	0.51	0.15
50	15405	1.89	2.86	2.39	0.1	19976	-0.53	2.44	0.28	0.22	15565	0.13	0.98	0.34	0.08
53	4494	2.17	2.95	2.49	0.08	4498	-0.59	1.59	0.41	0.3	4445	0.1	1	0.4	0.14
55	11777	2.08	2.82	2.42	0.09	16050	-0.42	3.29	0.32	0.32	11510	0.01	1	0.44	0.14
56	3451	2	2.88	2.51	0.09	3545	-0.32	1.92	0.42	0.34	3311	0.24	1	0.58	0.13
58	15804	1.99	2.73	2.38	0.11	15787	-0.44	1.76	0.29	0.24	15698	0.15	1	0.46	0.11
62	8884	2.02	2.71	2.36	0.08	8883	-0.27	2.14	0.21	0.18	8877	0.14	0.96	0.43	0.06
63	8335	2.06	2.72	2.48	0.09	8335	-0.34	1.78	0.53	0.3	8181	0.09	1	0.39	0.12
65	1344	2.27	2.77	2.45	0.1	1347	-0.12	1.85	0.5	0.33	1346	0.12	0.83	0.37	0.1
67	16414	2.02	2.77	2.38	0.1	21074	-0.37	2.63	0.27	0.23	16383	0.16	0.97	0.39	0.08
68	1595	2.09	2.77	2.53	0.07	1707	-0.5	1.75	0.61	0.3	1578	0.09	0.99	0.45	0.12
70	3741	2.14	3.01	2.56	0.1	3741	-0.41	1.77	0.62	0.34	3351	0.14	1	0.49	0.18
72	2120	2.20	2.74	2.49	0.08	2120	-0.12	1.46	0.55	0.27	2081	0.15	1	0.45	0.13

Tabla B. 3. Información muestral de los registros para cada pozo del CampoB para Datos Limpios.

Información Geofísica de Datos Limpios - CampoB						
			No.	Exponente de	Resistividad del	Densidad
		Tipo	Puntos	Cementación	Lodo Filtrado	de Matriz
Pozo	Intervalo	Línea	Intervalo	m	Rmf	Rho_ma
38	95	2	18	1.99	0.01	2.58
40	2	1	25	1.99	0.08	2.42
41	18	5	240	2.07	0.01	2.5
42	11	5	35	1.99	0.02	2.56
43	4	5	49	1.88	0.04	2.6
44	86	1	16	2	0.05	2.42
47	3	1	27	1.95	0.04	2.54
50	29	5	11	2.07	0.02	2.68
53	3	1	90	1.96	0.02	2.54
55	10	1	293	1.96	0.03	2.57
56	2	3	13	1.96	0.04	2.47
58	16	3	13	1.95	0.05	2.62
62	7	2	22	2	0.04	2.13
63	19	5	17	2.03	0.02	2.65
65	2	1	107	2.05	0.02	2.62
67	17	2	15	1.93	0.03	2.54
68	7	1	149	1.79	0.02	2.58
70	8	3	23	1.99	0.01	2.67
72	13	2	345	2.06	0.02	2.56

**Tabla B. 4. Información de roca para obtener las curvas de saturación de agua con Archie-Pickett.**

Información de Ciclo y Tendencia de los Registros por Pozo para Datos Limpios - CampoB													
		Puntos	Linea de Tendencia		No.	No.	Tipo de	Indice	Indice	Profundidad	Profundidad	Periodo	Repeti-
Pozo	Registro	Suavi-	Pendiente	Ordenada	Míni-	Máxi-	Ciclo	Inicio	Final	Inicio	Final	Estacionalidad	ciones
		zado			mos	mos		Ciclo	Ciclo	Ciclo	Ciclo	Ciclo	Ciclo
38	gr	101	0.001	54.7	243	234	maxAmax	9249	14466	1440.48	2239.67	5245	5
38	log(msfl)	101	-0.00004	0.32	242	246	maxAmax	45	639	33.53	125.12	602	41
38	nphi	101	-0.00002	0.16	101	107	maxAmax	2066	2876	2537.61	2662.73	822	13
38	rhob	101	0.0001	2.11	88	102	minAmin	2952	5008	2647.95	2961.28	2057	5
40	gr	27	0.07	-144.78	47	58	minAmin	195	1275	2915.56	3080.16	1081	2
40	log(msfl)	27	-0.001	3.97	55	51	maxAmax	195	1276	2915.56	3080.31	1082	2
40	nphi	27	-0.0002	0.79	56	48	maxAmax	68	1348	2896.21	3091.28	1281	2
40	rhob	27	0.001	0.81	50	57	maxAmax	34	697	2891.03	2992.07	664	3
41	gr	101	-0.003	60	222	218	minAmin	18945	20507	2418.88	2614.13	1563	14
41	log(msfl)	101	-0.0001	0.42	193	176	minAmin	5789	8120	776.25	1068	2335	9
41	nphi	101	-0.0001	0.56	111	108	maxAmax	7129	8664	2204.38	2396.25	1536	7
41	rhob	101	0.0001	2.24	104	106	maxAmax	5263	10406	1971.63	2614.5	5144	3
42	gr	101	-0.01	61.71	232	219	maxAmax	16554	17367	2634.08	2757.98	814	29
42	log(msfl)	101	0.00002	0.2	220	233	minAmin	17632	22063	2800.35	3475.79	4433	6
42	nphi	101	-0.00007	0.5	195	195	minAmin	15487	17827	3063.54	3420.31	2342	9
42	rhob	101	0.00008	2.19	188	181	maxAmax	15481	17886	3063.54	3430.07	2406	9
43	gr	101	-0.01	74.13	188	178	minAmin	17037	18173	2700.58	2873.71	1137	17
43	log(msfl)	101	0.00002	0.26	161	184	minAmin	6287	18124	1053.9	2858.77	11844	2
43	nphi	101	-0.00009	0.52	151	135	maxAmax	11877	12612	2513.44	2625.45	736	20
43	rhob	101	0.0001	2.18	138	158	maxAmax	6528	14331	1690.93	2883.16	7824	2

**Tabla B. 5. Información de suavizado de curva, tendencia e identificación de ciclos para cada registro de los pozos entre 38 y 43 del CampoB.**



Información de Ciclo y Tendencia de los Registros por Pozo para Datos Limpios - CampoB													
Pozo	Registro	Puntos Suavizado	Linea de Tendencia		No. Mínimos	No. Máximos	Tipo de Ciclo	Indice Inicio Ciclo	Indice Final Ciclo	Profundidad		Periodo Estacionalidad	Repeticiones
			Pendiente	Ordenada						Inicio Ciclo	Final Ciclo		
44	gr	101	0.0009	59.19	203	203	maxAmax	17839	18929	2825.45	2991.57	1091	20
44	log(msfl)	101	0.00004	0.22	183	220	minAmin	6192	18860	1053.8	2991.57	12716	2
44	nphi	101	-0.0001	0.53	171	165	minAmin	14620	15915	2927.41	3124.77	1296	13
44	rhob	101	0.00009	2.21	160	161	maxAmax	6487	15867	1690.84	3124.77	9410	2
47	gr	25	-0.05	188.67	55	56	maxAmax	151	1339	2573.88	2754.93	1189	2
47	log(msfl)	25	0.002	-3.63	54	52	maxAmax	528	1057	2631.34	2711.96	530	3
47	nphi	25	-0.0004	1.43	56	57	maxAmax	344	1307	2603.3	2750.06	964	2
47	rhob	25	0.0004	1.46	59	54	minAmin	343	1337	2603.14	2754.63	995	2
50	gr	101	-0.01	77.42	207	201	maxAmax	12273	17634	1970.25	2787.27	5362	4
50	log(msfl)	101	0.000004	0.27	185	219	minAmin	5404	16789	925.25	2679.52	11512	2
50	nphi	101	-0.0001	0.58	166	146	maxAmax	45	13232	771.17	2781.63	13193	2
50	rhob	101	0.0001	2.15	150	155	maxAmax	5695	15380	1680.39	3160.04	9710	2
53	gr	89	0.03	-39.61	46	51	minAmin	415	3734	2803.14	3308.95	3320	2
53	log(msfl)	89	-0.0001	0.87	50	51	maxAmax	443	3782	2808.01	3316.88	3340	2
53	nphi	89	-0.00004	0.36	46	49	minAmin	378	3775	2797.5	3315.2	3398	2
53	rhob	89	0.0001	2.11	49	49	maxAmax	512	3769	2819.14	3315.5	3258	2
55	gr	101	-0.01	54.19	152	158	maxAmax	12001	12879	2436.38	2570.19	879	19
55	log(msfl)	101	0.0001	0.09	149	166	minAmin	12623	15592	2526.76	2979.23	2970	6
55	nphi	101	-0.0002	0.63	115	116	minAmin	8755	11767	2589.55	3048.58	3013	4
55	rhob	101	0.0001	2.15	113	114	minAmin	7548	9495	2404.53	2701.26	1948	7
56	gr	69	0.01	22.58	47	55	minAmin	505	2754	2676.93	3019.68	2250	2
56	log(msfl)	69	0.0001	0.25	50	50	maxAmax	515	2793	2678.3	3025.47	2279	2
56	nphi	69	-0.0001	0.43	51	48	minAmin	772	2661	2731.8	3019.68	1890	2
56	rhob	69	0.0001	2.27	50	50	maxAmax	4	2659	2614.75	3019.37	2656	2
58	gr	101	-0.0001	65.1	152	150	minAmin	12364	15413	2630.12	3094.79	3050	6
58	log(msfl)	101	0.0001	-0.003	151	160	maxAmax	12315	15437	2625.85	3101.64	3123	6
58	nphi	101	-0.0001	0.62	154	153	minAmin	12339	15402	2626.31	3093.11	3064	6
58	rhob	101	0.0001	2.11	149	164	minAmin	11680	13674	2526.03	2830.37	1998	8
62	gr	101	0.002	62.29	111	113	maxAmax	4317	8892	1190.25	2337	4588	3
62	log(msfl)	101	0.0001	0.03	85	99	maxAmax	7404	8589	2546.25	2842.5	1186	8
62	nphi	101	-0.0001	0.5	81	86	maxAmax	6092	8871	2218.5	2913.25	2780	4
62	rhob	101	0.0001	2.18	89	83	maxAmax	4253	8116	1758	2724	3865	3
63	gr	101	0.01	20.17	79	92	minAmin	3359	7991	3241.71	3947.63	4633	2
63	log(msfl)	101	-0.0002	1.16	81	82	maxAmax	1753	7878	2996.95	3930.4	6126	2
63	nphi	101	0.00003	0.1	78	91	minAmin	3273	7886	3228.6	3931.62	4614	2
63	rhob	101	0.00001	2.44	80	85	maxAmax	3925	7886	3327.97	3931.62	3962	3
65	gr	25	-0.01	96.08	45	51	maxAmax	34	1098	2884.96	3047.11	1065	2
65	log(msfl)	25	0.0001	0.24	55	53	maxAmax	426	842	2944.7	3008.1	417	4
65	nphi	25	-0.0001	0.42	44	51	maxAmax	113	1098	2897	3047.11	986	2
65	rhob	25	-0.0003	3.37	56	53	maxAmax	404	1313	2941.35	3079.88	910	2
67	gr	101	-0.01	78.07	218	210	maxAmax	17059	18034	2700.69	2849.28	976	22
67	log(msfl)	101	0.0001	0.18	207	210	minAmin	7186	19612	1193.3	3087.02	12427	2
67	nphi	101	-0.0001	0.53	162	171	minAmin	14342	15417	2991.47	3155.3	1076	16
67	rhob	101	0.0001	2.13	160	165	maxAmax	14308	15426	2984.91	3155.3	1119	15
68	gr	29	-0.10	362.07	58	56	maxAmax	28	1492	2904.1	3127.21	1465	2
68	log(msfl)	33	0.001	-1.79	54	60	maxAmax	935	1231	3042.32	3087.43	297	6
68	nphi	31	-0.0003	1.13	60	59	maxAmax	69	1587	2910.34	3141.69	1519	2
68	rhob	31	0.0004	1.44	53	61	maxAmax	424	1231	2964.45	3087.43	808	2
70	gr	73	-0.005	65.26	47	50	minAmin	786	2191	3199.64	3413.76	1406	3
70	log(msfl)	73	0.001	-1.48	55	51	maxAmax	785	3321	3199.49	3585.97	2537	2
70	nphi	73	-0.0002	0.65	56	53	minAmin	86	2157	3092.96	3408.58	2072	2
70	rhob	73	0.0002	1.91	51	50	maxAmax	80	2175	3092.04	3411.32	2096	2
72	gr	41	-0.04	154.03	51	54	maxAmax	47	2109	2877.01	3191.41	2064	2
72	log(msfl)	41	0.001	-1.21	52	49	maxAmax	350	1646	2923.18	3120.69	1297	2
72	nphi	41	-0.0003	1	55	56	minAmin	269	1890	2910.84	3157.88	1622	2
72	rhob	41	0.0003	1.55	53	54	minAmin	639	1934	2967.23	3164.59	1296	2

**Tabla B. 6. Información de suavizado de curva, tendencia e identificación de ciclos para cada registro de los pozos entre 44 y 72 del CampoB.**

Datos MonteCarlo - CampoB															
Pozo	GR					MSFL					NPHI				
	Datos	Min	Max	Prom	Std	Datos	Min	Max	Prom	Std	Datos	Min	Max	Prom	Std
38	24461	9.85	184.13	55.68	10.36	24461	0.37	1950	3.9	56.06	24461	0.001	0.44	0.13	0.06
40	1405	30.37	101.22	54.82	14.01	1405	0.48	31.69	4.27	3.64	1405	0.07	0.59	0.23	0.07
41	20569	6.03	86.02	56.12	12.24	20569	0.52	101.23	2.38	4.81	20569	0.09	0.65	0.44	0.08
42	23269	12.36	74.67	52.02	10.1	23269	0.4	915.42	2.2	8.98	23269	0.04	0.62	0.37	0.08
43	18734	19.03	84.98	60.46	11.42	18734	0.42	86.39	2.41	3.08	18734	0.06	0.62	0.38	0.09
44	20854	32.37	165.2	60.74	8.38	20854	0.41	110.89	2.38	2.73	20854	0.02	0.57	0.36	0.09
47	1346	15.44	117.42	47.38	18.13	1346	0.37	124.96	5.22	8.35	1346	0.08	0.61	0.25	0.09
50	20114	26.84	95.43	68.88	9.03	20114	0.3	276.17	2.43	4.99	20114	0.0002	0.82	0.4	0.11
53	4502	23.55	106.44	51.66	14.14	4502	0.26	38.69	3.41	3.12	4502	0.05	0.48	0.23	0.08
55	16064	7.02	91.03	44.53	10.23	16064	0.38	1934.03	4.47	33.73	16064	0.02	0.61	0.34	0.11
56	3545	37.74	113.24	64.08	10.28	3545	0.48	83.49	4.08	6.24	3545	0.04	0.6	0.24	0.1
58	15808	33.76	126.83	64.82	8.68	15808	0.36	57.25	2.54	3.49	15808	0.03	0.6	0.37	0.11
62	11225	14.52	101.31	66.04	11.16	11225	0.54	138.7	1.91	3.66	11225	0.05	0.56	0.38	0.07
63	8335	27.73	128.74	56.75	12.31	8335	0.46	59.91	4.43	4.36	8335	0.02	0.51	0.21	0.08
65	1347	15.98	92.55	53.24	17.87	1347	0.75	70.91	4.64	5.99	1347	0.06	0.44	0.24	0.09
67	21077	21.14	117.28	67.23	12.48	21077	0.43	429.85	2.45	6.42	21077	0.05	0.64	0.36	0.09
68	1707	27.75	97.13	50.94	14.39	1707	0.31	55.79	5.14	4.17	1707	0.02	0.53	0.15	0.06
70	3741	23.5	111.08	49.17	14.8	3741	0.39	58.75	5.6	4.79	3741	0.01	0.48	0.13	0.08
72	2120	24.02	102.19	46.57	15.66	2120	0.76	28.56	4.27	2.71	2120	0.04	0.52	0.21	0.08

Datos MonteCarlo - CampoB															
Pozo	RHOB					log(MSFL)					Sx0				
	Datos	Min	Max	Prom	Std	Datos	Min	Max	Prom	Std	Datos	Min	Max	Prom	Std
38	24461	1.87	2.75	2.35	0.13	24461	-0.43	3.29	0.25	0.24	21232	0.01	1	0.56	0.2
40	1405	2.24	2.82	2.51	0.08	1405	-0.32	1.5	0.53	0.28	1243	0.18	1	0.67	0.16
41	20569	2.23	2.69	2.35	0.05	20569	-0.28	2.01	0.25	0.23	19529	0.07	1	0.69	0.18
42	23269	2.01	2.8	2.35	0.09	23269	-0.4	2.96	0.24	0.21	23247	0.01	0.98	0.32	0.08
43	18734	2.03	2.69	2.35	0.09	18734	-0.37	1.94	0.29	0.22	18734	0.06	0.91	0.4	0.09
44	20854	2.04	2.72	2.36	0.09	20854	-0.39	2.04	0.29	0.21	18935	0.14	1	0.74	0.14
47	1346	2.08	2.75	2.46	0.09	1346	-0.43	2.1	0.46	0.43	1309	0.11	1	0.51	0.15
50	20114	1.89	2.86	2.35	0.11	20114	-0.53	2.44	0.28	0.22	20093	0.02	0.98	0.31	0.1
53	4502	2.17	2.95	2.49	0.08	4502	-0.59	1.59	0.41	0.3	4449	0.1	1	0.4	0.14
55	16064	2.08	2.82	2.4	0.09	16064	-0.42	3.29	0.32	0.32	15805	0.01	1	0.4	0.13
56	3545	2	2.88	2.51	0.09	3545	-0.32	1.92	0.42	0.34	3404	0.24	1	0.59	0.13
58	15808	1.99	2.73	2.38	0.11	15808	-0.44	1.76	0.29	0.24	15719	0.15	1	0.46	0.11
62	11225	2.02	2.71	2.33	0.09	11225	-0.27	2.14	0.2	0.17	11213	0.17	0.99	0.49	0.08
63	8335	2.06	2.72	2.48	0.09	8335	-0.34	1.78	0.53	0.3	8181	0.09	1	0.39	0.12
65	1347	2.27	2.77	2.45	0.1	1347	-0.12	1.85	0.5	0.33	1346	0.12	0.83	0.37	0.1
67	21077	2.02	2.77	2.35	0.11	21077	-0.37	2.63	0.27	0.23	21056	0.02	0.97	0.37	0.09
68	1707	2.09	2.77	2.53	0.07	1707	-0.5	1.75	0.61	0.3	1700	0.07	0.97	0.37	0.12
70	3741	2.14	3.01	2.56	0.1	3741	-0.41	1.77	0.62	0.34	3351	0.14	1	0.49	0.18
72	2120	2.2	2.74	2.49	0.08	2120	-0.12	1.46	0.55	0.27	2081	0.15	1	0.45	0.13

Tabla B. 7. Información muestral de los registros para cada pozo del CampoB para Datos Monte Carlo.

Información Petrofísica de Datos MonteCarlo - CampoA										
Pozo	Intervalos Identificados	Intervalos Evaluados con Pickett	Profundidad Inicial	Profundidad Final	Intervalo Principal	Tipo Línea	Tamaño Intervalo	Exponente de Cementación m	Resistividad del Lodo Filtrado Rmf	Densidad de Matriz Rho_ma
38	118	112	26.21	3753.92	64	3	8	2	0.01	2.5
40	10	10	2886	3099.97	2	1	25	1.99	0.08	2.42
41	23	23	49	2620	1	2	15	2.19	0.12	2.34
42	37	36	110.03	3656.08	11	5	35	1.99	0.02	2.56
43	11	11	95	2949.91	5	5	49	1.88	0.04	2.6
44	149	147	106.94	3284.94	7	1	185	1.97	0.13	2.3
47	13	13	2551.02	2756	3	1	27	1.95	0.04	2.54
50	31	31	100	3165.22	29	5	11	2.07	0.02	2.68
53	32	32	2740.04	3425.99	3	1	90	1.96	0.02	2.54
55	42	42	601.03	3049.03	10	1	293	1.96	0.03	2.57
56	31	31	2599.97	3140.08	2	3	13	1.96	0.04	2.47
58	63	63	746	3154.98	16	3	13	1.95	0.05	2.62
62	18	18	110	2916	4	5	8	2	0.05	2.32
63	89	89	2729.95	4000.05	19	5	17	2.03	0.02	2.65
65	8	8	2879.93	3085.06	2	1	107	2.05	0.02	2.62
67	29	29	98	3309.98	17	2	15	1.93	0.03	2.54
68	13	13	2899.98	3159.98	12	3	59	1.89	0.01	2.61
70	32	32	3080	3649.98	8	3	23	1.99	0.01	2.67
72	16	16	2870	3192.93	13	2	345	2.06	0.02	2.56
Mínimo	8	8	26.21	2620						
Máximo	149	147	3080	4000.05						
Promedio	40.26	39.79	1435.74	3230.01						
Total	765	756								

**Tabla B. 8. Intervalos identificados y que pudieron ser evaluados con Pickett. Información de roca para obtener saturación de agua con Archie-Pickett para CampoB de Datos Monte Carlo.**

Datos Limpios - CampoA				Datos Limpios - CampoB				
Intervalos Identificados	Intervalos Evaluados con Pickett	Profundidad Inicial	Profundidad Final	Pozo	Intervalos Identificados	Intervalos Evaluados con Pickett	Profundidad Inicial	Profundidad Final
65	65	1930.54	2907.12	38	118	64	26.21	3749.8
8	8	2352.02	2856.15	40	10	10	2886	3099.97
48	46	2022.04	2694.13	41	23	15	49	2620
4	4	2312.04	2739.98	42	37	36	110.03	3656.08
26	26	2592.54	2964.7	43	9	9	95	2949.91
3	3	2367.09	2949.56	44	149	101	106.94	3284.94
30	30	2243.02	2788.01	47	13	13	2551.02	2756
8	8	2243.94	2678.58	50	31	31	100	3165.22
7	7	2546.04	2947.77	53	32	32	2740.04	3425.99
39	4	2229.99	2887.14	55	43	42	601.03	3049.03
30	30	2212.95	2787	56	31	31	2599.97	3140.08
24	24	2360.04	2713.6	58	63	63	746	3154.98
16	16	1949.96	2845.46	62	18	13	110	2916
8	8	2615.49	3035.5	63	89	89	2729.95	4000.05
50	50	2011.99	2905.51	65	8	8	2879.93	3085.06
58	57	1899.93	2791.47	67	29	27	98.3	3309.83
55	54	1749.98	2559.83	68	10	10	2899.98	3159.98
50	16	696.62	3378.86	70	32	32	3080	3649.98
79	49	1795.4	5020.33	72	16	16	2870	3192.93
39	36	1185.82	2700.68	Mínimo	8	8	26.21	2620
143	24	728.17	4411.92	Máximo	149	101	3080	4000.05
3	3	696.62	2559.83	Promedio	40.05	33.79	1435.76	3229.78
143	65	2615.49	5020.33	Total	761	642		
27 62	26 9	2002 17	2026 82					

**Tabla B. 9. Intervalos identificados y que pudieron ser evaluados con Pickett para CampoB.**

Ajuste lineal de crossplots de los registros del CampoA											
registro1	registro2	parejas	pendiente	ordenada	angulo	registro1	registro2	parejas	pendiente	ordenada	angulo
gr	msfl	216722	-0.04	4.67	-2.56	rhob	gr	167241	0.15	57.99	8.72
gr	nphi	167186	-1378.48	80447.5	-89.96	rhob	msfl	167385	-0.13	3.04	-7.18
gr	rhob	167241	3293.73	-192213.76	89.98	rhob	nphi	167040	-0.72	2.06	-35.62
gr	log(msfl)	216722	-0.002	0.32	-0.1	rhob	log(msfl)	167385	0.77	-1.62	37.49
gr	sx0	162437	-6621.42	386940.17	-89.99	rhob	sx0	162292	-3.25	8.22	-72.9
msfl	gr	216722	1.12	55.4	48.33	log(msfl)	gr	216722	-8.6	62.47	-83.37
msfl	nphi	167247	31.82	-87.29	88.2	log(msfl)	msfl	217386	24.88	-4.48	87.7
msfl	rhob	167385	-68.37	189.87	-89.16	log(msfl)	nphi	167247	3.25	-0.66	72.89
msfl	log(msfl)	217386	0.004	0.29	0.21	log(msfl)	rhob	167385	0.2	2.35	11.32
msfl	sx0	162745	-348.91	956.36	-89.84	log(msfl)	sx0	162745	-21.18	6.72	-87.3
nphi	gr	167186	-0.08	58.39	-4.78	sx0	gr	162437	-0.61	58.69	-31.58
nphi	msfl	167247	0.15	2.71	8.27	sx0	msfl	162745	-0.66	3.01	-33.33
nphi	rhob	167040	-0.54	2.58	-28.54	sx0	nphi	162745	-0.18	0.42	-10.35
nphi	log(msfl)	167247	-0.9	0.53	-41.97	sx0	rhob	162292	0.15	2.33	8.74
nphi	sx0	162745	-0.33	0.51	-18.02	sx0	log(msfl)	162745	0.09	0.18	5.38

**Tabla B.10. Ajuste lineal de nubes de puntos de las gráficas cruzadas entre registros del CampoB.**



## Apéndice C

# Metodología para la generación de registros sintéticos

### Metodología para la construcción de Datos con Monte Carlo

Si consideramos cada secuencia de registro como una serie de tiempo (serie de profundidad o serie espacial es una mejor interpretación) podemos aplicar mucho del tratamiento que se le da a las mismas para su reconstrucción. A continuación se presenta la metodología seguida para la modificación de los Datos Limpios para obtener los Datos de Monte Carlo. [41]

A cada serie de tiempo podemos representarla por:

**Tendencia:** El patrón que muestra de crecimiento o decrecimiento.

**Ciclicidad:** Las fluctuaciones en forma de ciclo u onda que presenta la secuencia de datos.

**Estacionalidad:** El periodo en el cual tienden a repetirse el comportamiento de los datos.

**Ruido o Fluctuaciones irregulares:** Cambios aleatorios que se dan en la sucesión de datos.

Con estas características en mente, cada registro se trató para medir estos comportamientos. En las siguientes ecuaciones  $X$  representa la profundidad y  $Y$  el registro que se desea reconstruir.

El primer paso para la reconstrucción consistió en realizar un **suavizado** con promedio móvil de  $Y$ , esto es un procedimiento con el cual se puede eliminar el ruido de la serie de tiempo, que inicialmente es lo que se desea, ya que lo recomendable es analizar el ruido de manera independiente, con esto lo que se obtuvo fue  $Y_{suave}$ . El promedio móvil consiste en tomar una vecindad de puntos en torno al que se quiere suavizar, sumarlos y

promediarlos. Por ejemplo si la vecindad es de 3, los puntos de suavizado serán 7 y el valor que tendrá el  $i$ -ésimo dato será:

$$x_{i_{\text{suavizado}}} = \frac{x_{i-3} + x_{i-2} + x_{i-1} + x_i + x_{i+1} + x_{i+2} + x_{i+3}}{7}. \quad (\text{C.1})$$

Para obtener la tendencia lineal de los datos se realizó un ajuste lineal con mínimos cuadrados a las parejas  $[X, Y_{\text{suave}}]$ , de lo cual se obtuvo una relación del tipo (C.2), donde el signo de  $m$  indica la tendencia de crecimiento o decrecimiento:

$$Y_{\text{tendencia}} = mX + b. \quad (\text{C.2})$$

Se generó un registro sin tendencia  $Y_{st}$ , restando la recta estimada en (C.2) a la secuencia suavizada  $Y_{\text{suave}}$ :

$$Y_{st} = Y_{\text{suave}} - Y_{\text{tendencia}} = Y_{\text{suave}} - (mX + b). \quad (\text{C.3})$$

Se consideró entonces que la serie de tiempo sin tendencia obtenida era una serie de tiempo que iba a presentar estacionalidad y ciclicidad sin importar qué tan irregular fuera la forma del ciclo que se repitiera y sin conocimiento de cuál era el periodo de repetición del ciclo. Para identificar el ciclo se realizó una lista de máximos locales  $A = \{a_1, a_2, \dots, a_n\}$  y otra de mínimos locales  $B = \{b_1, b_2, \dots, b_n\}$ ; para cada lista se obtuvo la distancia en profundidad entre cada uno de sus miembros:

$$\begin{aligned} \text{separación}_A &= \{d(a_1, a_2), d(a_1, a_3), \dots, d(a_2, a_3), \dots\}, \\ &\text{donde } d(a_i, a_j)_{j>i} = x(a_j) - x(a_i); \\ \text{separación}_B &= \{d(b_1, b_2), d(b_1, b_3), \dots, d(b_2, b_3), \dots\} \\ &\text{donde } d(b_i, b_j)_{j>i} = x(b_j) - x(b_i); \end{aligned} \quad (\text{C.4})$$

donde el ciclo se considera que va los puntos  $[c_i, c_j)$  que cumplen con:

$$[c_i, c_j) = \min \{ \max \{ \text{separación}_A \}, \max \{ \text{separación}_B \} \}. \quad (\text{C.5})$$

En la Figura C.1 se ilustra el procedimiento de identificación de un ciclo para el registro de densidad RHOB del pozo 15, donde la línea morada sobre la verde indica el ciclo identificado, los  $a_i$  representan algunos de los valores máximos de la secuencia y los  $b_i$  los valores mínimos de la misma.

El periodo o estacionalidad del ciclo viene dado por:

$$\text{periodo} = x(c_j) - x(c_i). \quad (\text{C.6})$$

Una vez identificado un ciclo, (inicio, fin), por tanto su periodo es posible repetir este a la izquierda y a la derecha sobre la serie de tiempo sin tendencia  $Y_{st}$  y comparar su comportamiento en diferentes repeticiones. En la Figura C.2 se repite el ciclo identificado en la Figura C.1 que va de 2100 m a 2360 m aproximadamente. La repetición a la derecha no

permite que se genere un ciclo completo, sin embargo a la izquierda se repite en dos ocasiones. En estas repeticiones a la izquierda se puede comparar la línea de los ciclos repetidos de manera punteada con la línea que siguió el registro y podemos observar sus semejanzas y diferencias.

Una manera de identificar un ciclo que represente tanto al ciclo originalmente identificado como a sus repeticiones es obtener un ciclo promedio del ciclo original y sus repeticiones, este procedimiento es el que se realiza para identificar un **ciclo representativo** de cada registro. En la Figura C.3 se presenta el **ciclo promedio** que se obtuvo para RHOB y su repetición a lo largo del registro, en esta se puede observar que la curva cíclica representa de forma más fiel el comportamiento del registro.

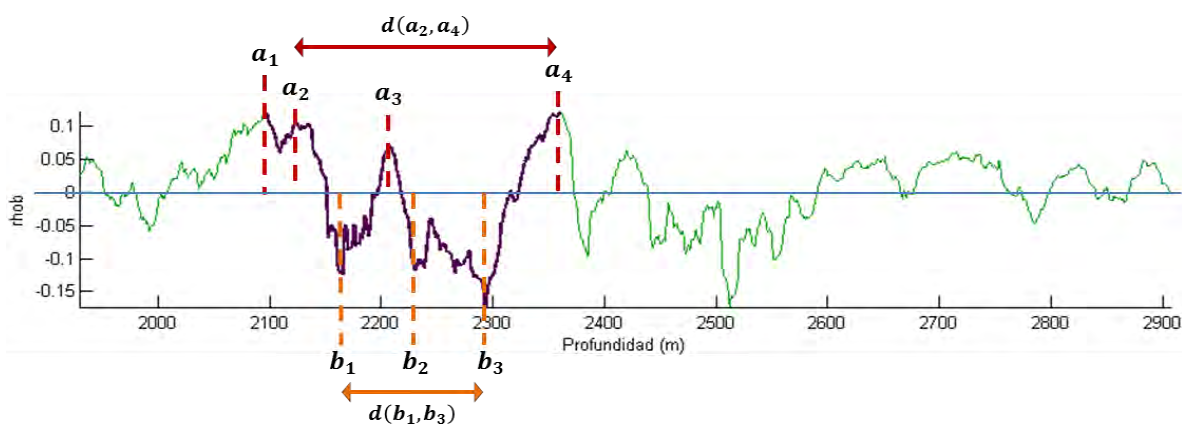


Figura C. 1. Selección de ciclo para RHOB del pozo 50, sobre la serie de tiempo suavizada y sin tendencia.

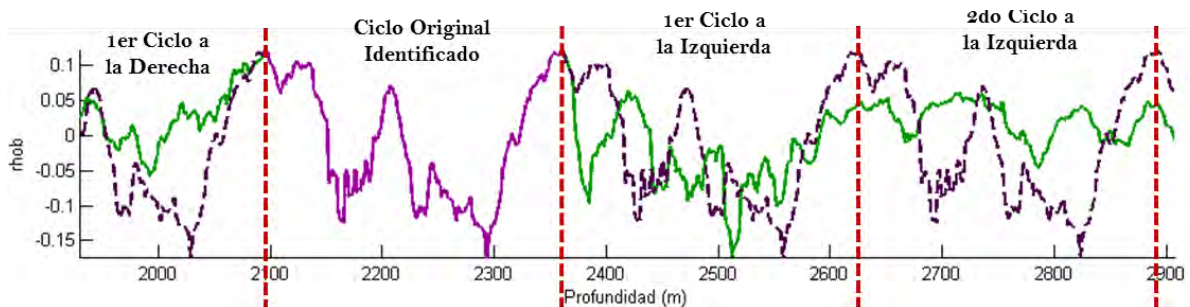


Figura C. 2. Repetición del ciclo identificado sobre el registro RHOB del pozo 50.

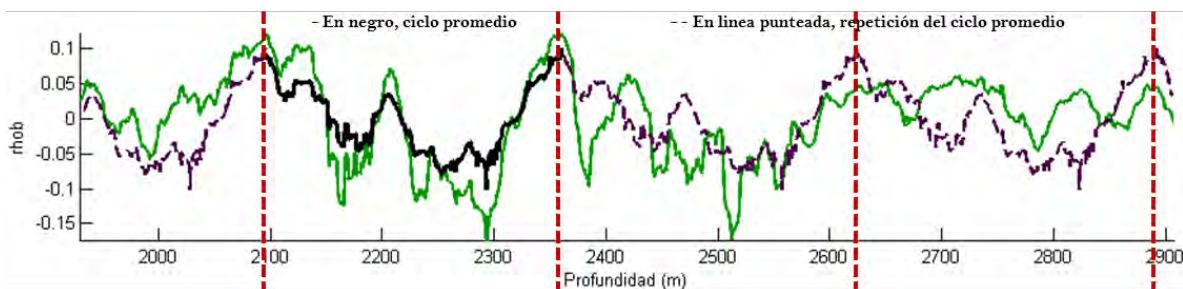


Figura C. 3. Repetición de ciclo promedio sobre el registro RHOB del pozo 50.



Para cada registro de los pozos del CampoB de los Datos Limpios se aplicó la misma metodología descrita. En las Tablas B.5 y B.6 se presentan el número de puntos, **Puntos Suavizado**, que se utilizó para suavizar cada curva, la **Pendiente y Ordenada** al origen de la línea que describe la **Tendencia** del registro, como se ha mencionado el patrón de la dirección de crecimiento de los datos es dada por el signo que tiene el valor de la pendiente. En las columnas posteriores se presenta la información correspondiente al ciclo promedio identificado para cada registro. La columna de **No. Mínimos** indica el total de mínimos locales identificados y la columna **No. Máximos** el número de máximos locales que se registraron. La columna **Tipo de ciclo** indica si el ciclo representativo elegido fue de máximo a máximo o de mínimo a mínimo de acuerdo a la ecuación (C.5) las siguientes columnas indican su **inicio** y **final** tanto en **índice** de secuencia como en su valor en metros de la **profundidad**, posteriormente se presenta la duración del **Periodo** y las **veces que se repitió el ciclo** para obtener el ciclo promedio.

Como se ha mencionado con el ciclo promedio,  $f(\tau)$ , como una secuencia de valores que se repiten de  $Y_{st}$  se puede estimar el periodo o estacionalidad del mismo y construir una serie de tiempo cíclica,  $Y_{ciclica}$  donde los elementos que la conforman son una repetición de la secuencia de valores  $f(\tau)$ , el valor  $y_i$  de  $Y_{ciclica}$  que corresponda dependerá del periodo y el inicio del ciclo identificado como en la ecuación

$$y_i = f\left(i + e \text{ tra a bajo} - \text{periodo} \left\lfloor \frac{i - 1 + e \text{ tra a bajo}}{\text{periodo}} \right\rfloor\right), \quad (\text{C. 7})$$

$$\text{dónde: } e \text{ tra a bajo} = \text{periodo} - \left( (\text{inicio} - 1) - \text{periodo} \left\lfloor \frac{\text{inicio} - 1}{\text{periodo}} \right\rfloor \right). \quad (\text{C. 8})$$

En este punto se puede plantear que la serie de tiempo que conforma cada registro se puede descomponer de la siguiente manera:

$$Y = mX + b + Y_{ciclica} + g(w | \Omega(\omega)), \quad (\text{C. 9})$$

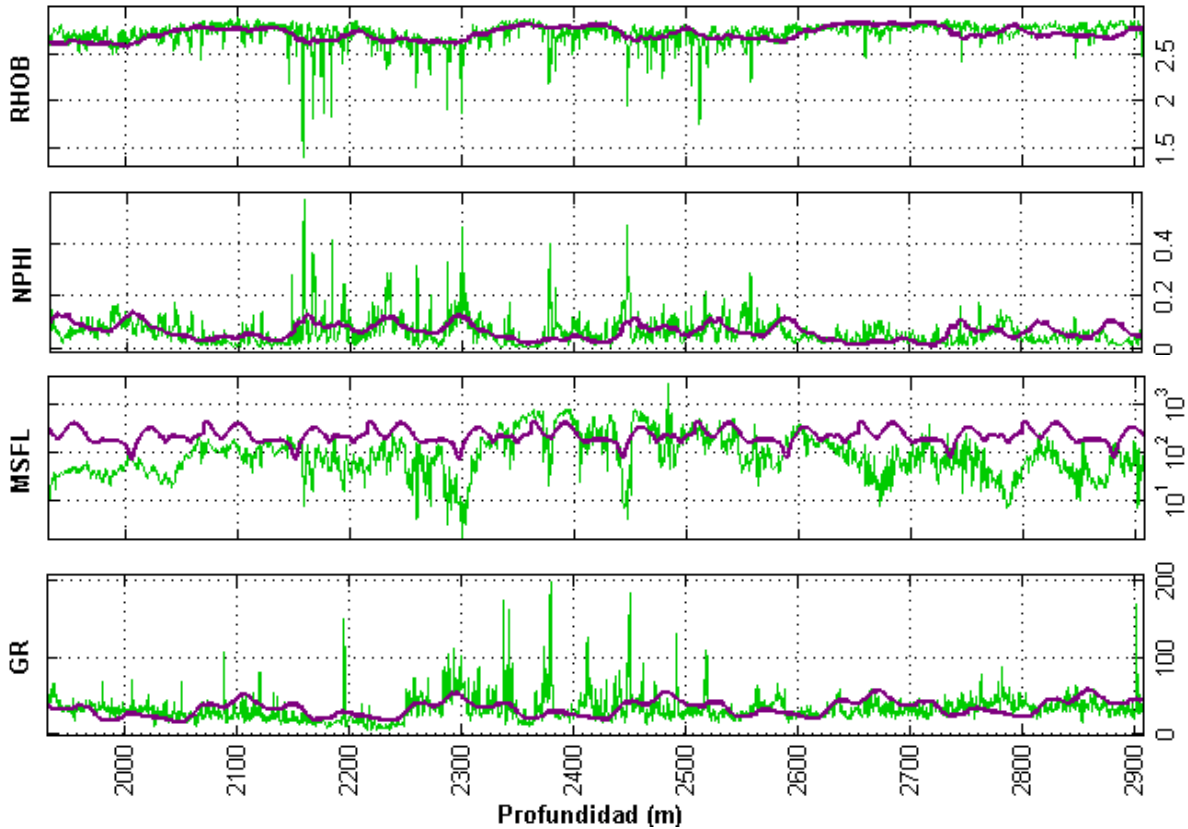
donde  $mX + b$  corresponde a la línea de tendencia que depende de la profundidad,  $Y_{ciclica}$  es una secuencia de valores periódicos y  $g(w | \Omega(\omega))$  representan las variaciones irregulares que se pueden presentar en cada punto de la serie de tiempo original  $Y$ .

$g(w | \Omega(\omega))$  en la ecuación (C.9) representa una función estocástica donde  $w \in [a, b]$ , es decir,  $w$  es una variable aleatoria definida en un rango de cambio del registro entre  $a$  y  $b$ ; y  $\Omega(\omega)$  es la distribución de probabilidad sobre la cual se muestrea  $w$ . Una manera de estimar la distribución de probabilidad de los cambios irregulares en  $g(w | \Omega(\omega))$  es construir una curva  $Y_{ciclo+tendencia}$  como

$$Y_{ciclo+tendencia} = Y_{ciclica} + mX + b. \quad (\text{C. 10})$$

Y compararla con la curva original  $Y$ . En la Figura C.4 se presentan los cuatro registros del pozo 15 y se compraran con sus respectivas curvas  $Y_{ciclo+tendencia}$  construidas.

La diferencia entre ambas curvas puede considerarse para los fines de la reconstrucción de la serie de tiempo como „ruido“ representado por  $g(w | \Omega(\omega))$  en la ecuación (C.9).



**Figura C. 4.** Comparación entre la serie de tiempo original (línea FINA),  $Y$ , y la serie construida de ciclo más tendencia (línea GRUESA),  $Y_{ciclo+tendencia}$ , para los cuatro registros del pozo 50.

Para estimar la distribución de probabilidad de  $g(w | \Omega(\omega))$ , se calcula una serie de tiempo  $Y_{diff}$  como:

$$Y_{diff} = Y - Y_{ciclo+tendencia} \quad (C. 11)$$

Esto se realizó para cada registro. Para tener una muestra considerable, para cada registro se unió en una sola serie de tiempo la  $Y_{diff}$  de cada pozo que conformaba cada campo y sobre estas cinco  $Y_{diff}$  correspondientes a  $\{gr, msfl, nphi, rhob, \log(msfl)\}$  del CampoB se obtuvo la información descriptiva que se presenta en las Tablas C.1, donde con las columnas Min y Max se puede definir el rango para la variable aleatoria  $w \in [a, b]$  de  $g(w | \Omega(\omega))$ .

Información de Distribuciones de Cambio de CampoB					
Registro	No. Datos	Min	Max	Promedio	Std
gr	219217	-91.84	93.5	0	3.09
msfl	217061	-1946.04	1940.12	0	16.2
nphi	167191	-0.24	0.21	0	0.02
rhob	167478	-0.37	0.65	0	0.03
log(msfl)	217061	-2.69	3.11	0	0.07

Tabla C. 1. Información descriptiva de  $Y_{diff}$  para Datos Limpios - CampoB.

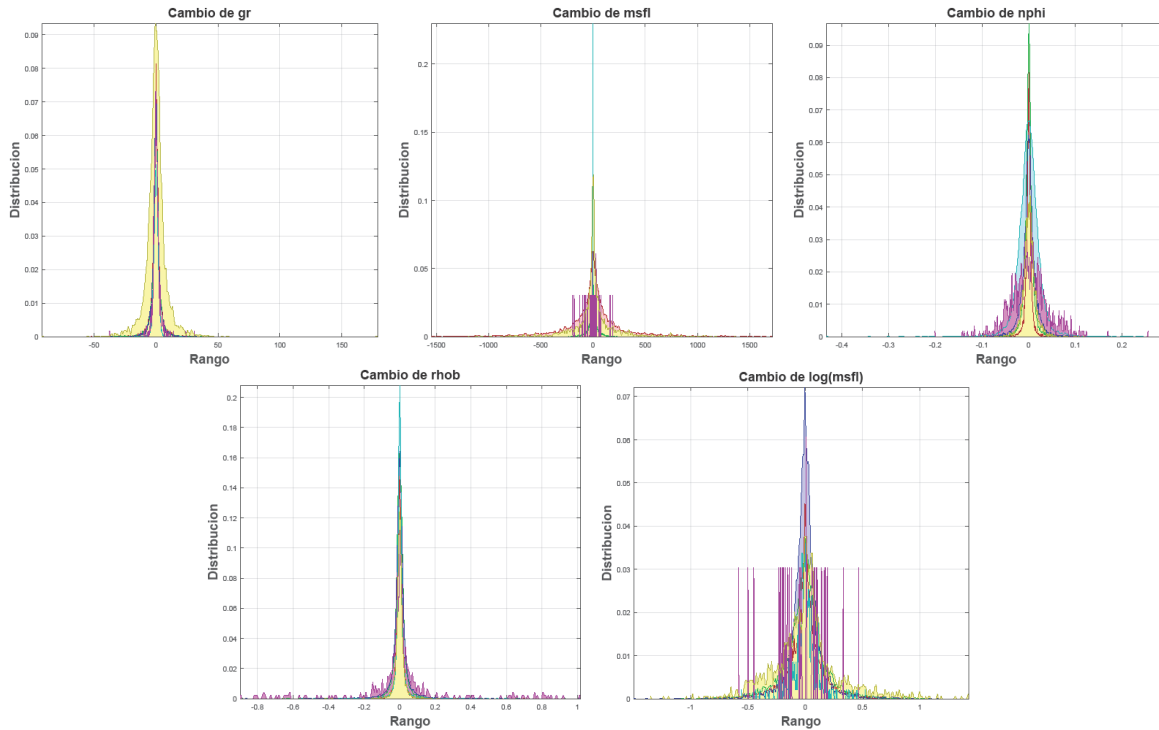


Figura C. 5. Histogramas de frecuencia con 403 cajas para  $Y_{diff} = Y_{original} - Y_{ciclo+tendencia}$  de cada registro para Datos Limpios – CampoB.

Con la serie de tiempo  $Y_{diff}$  es posible construir un histograma de frecuencia para  $w \in [a, b]$  con un número finito de divisiones o cajas y utilizar este como referencia para generar una función de probabilidad aproximada  $\Omega(\omega)$  dónde:

$$\Omega(\omega) = \frac{\text{frecuencia}(\omega)}{\text{No. Datos}}. \quad (\text{C. 12})$$

En las Figuras C.5 se presentan los histogramas de frecuencias de  $Y_{diff}$  para los registros en el CampoB, respectivamente.

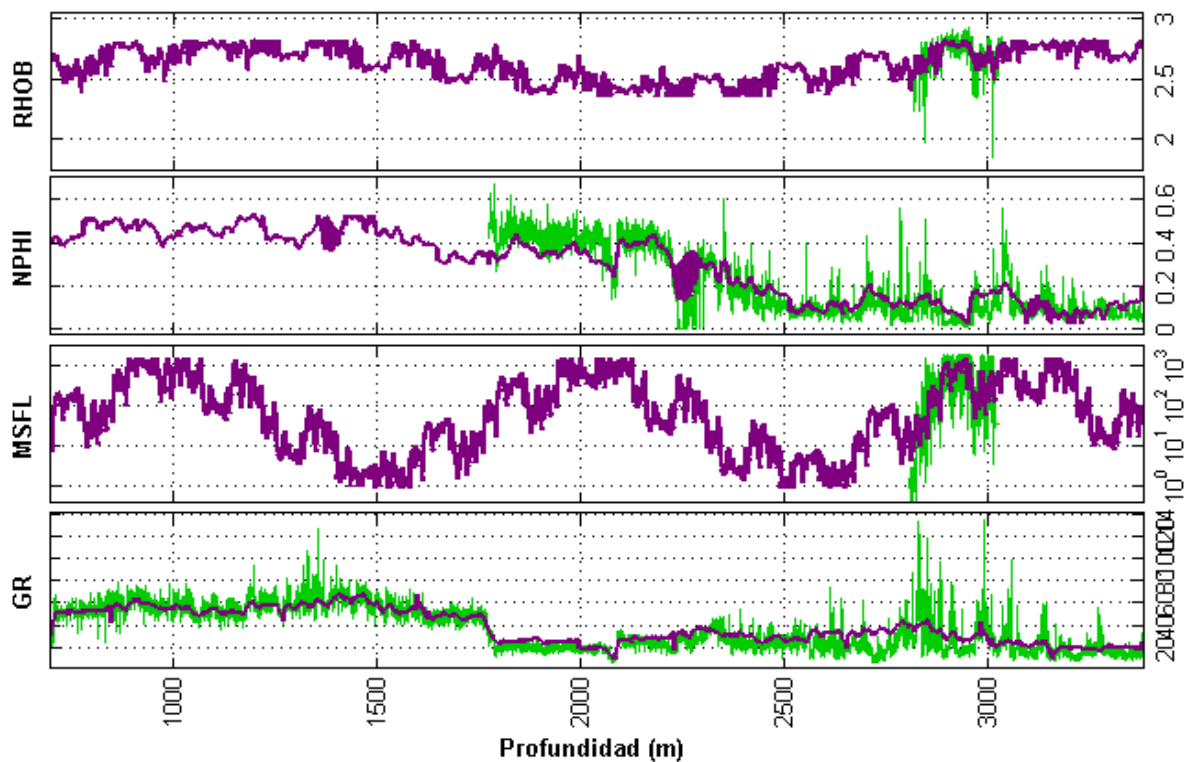


Figura C. 6. Comparación entre la serie de tiempo original,  $Y$ , y la serie construida de ciclo más tendencia,  $Y_{ciclo+tendencia}$ , para los cuatro registros del pozo 38.

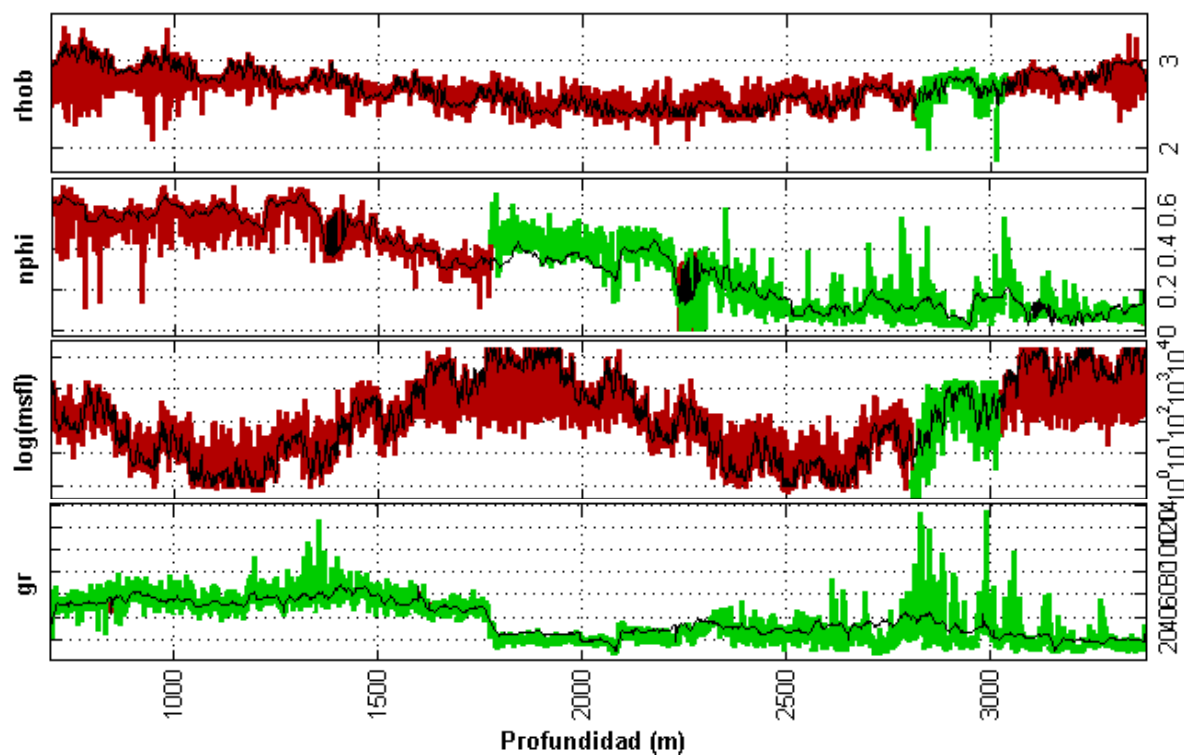


Figura C. 7. Comparación entre la serie de tiempo original,  $Y$ ,  $Y_{ciclo+tendencia}$  y  $Y_{montecarlo}$  para los cuatro registros del pozo 38.

Como se ha indicado anteriormente los **Datos Limpios** presentan muchos valores nulos y en algunos casos vacíos en tramos grandes de información. Los **Datos Monte Carlo** son una modificación de los Datos Limpios donde los datos nulos o faltantes fueron reconstruidos a partir de la información que tenemos en la ecuación (C.9).

En la Figura C.6 se presenta la comparación entre la serie de tiempo original y la construida como  $Y_{ciclo+tendencia}$  para cada registro del pozo 34, donde se puede observar la cantidad de datos desconocidos que tienen las curvas. También se observa que la curva  $Y_{ciclo+tendencia}$  proporciona un camino probable de continuidad en las partes vacías para el registro original.

Como se observa en la Figura C.6, utilizar  $Y_{ciclo+tendencia}$  como un camino continuo viable para reconstruir la serie de tiempo original es un ejercicio de especulación donde la fidelidad del mismo dependerá del tamaño de la muestra de la serie de tiempo original, en el caso de la Figura C.6 las curvas MSFL y RHOB no se tiene una muestra muy grande, sin embargo con esos datos se realiza una propuesta posible para  $Y_{ciclo+tendencia}$ .

Utilizando esta información se construyo  $Y_{montecarlo}$ , como:

$$Y_{montecarlo} = Y_{tendencia} + Y_{ciclica} + g(w | \Omega(\omega)), \quad (C.13)$$

donde las variaciones estocásticas  $g(w | \Omega(\omega))$  fueron generadas de las funciones de probabilidad aproximadas que surgieron de los histogramas de frecuencia de las Figuras C.5. Este proceso de reconstrucción de datos se realizó para los cuatro registros para todos los pozos del CampoB. En la ecuación (C.13)  $Y_{tendencia}$  y  $Y_{ciclica}$  dependen directamente de la curva a reconstruir, pero  $g(w | \Omega(\omega))$  proviene de las diferencias estocásticas globales de todo el campo.

En la Figura C.7 se presentan los resultados de la aplicación de la reconstrucción de los datos utilizando la ecuación (C.13), en la figura se compara la serie de tiempo  $Y_{montecarlo}$  reconstruida, con la serie de tiempo  $Y_{ciclo+tendencia}$  que funge como un camino o envolvente y la serie de tiempo original para los registros del pozo 34. En las Figuras C.9 y C.10 se presentan dos ejemplos adicionales de la reconstrucción para obtener los Datos Monte Carlo de los Datos limpios.

Estas técnicas de reconstrucción pueden tener una relación con lo que se realiza como tratamiento en geofísica, por ejemplo la búsqueda de la tendencia puede tener relación con la corrección de temperatura que a veces se realiza a los registros y la búsqueda de ciclos con capas de diferentes litológicas que se repiten a diferentes periodos en el manto, los cambios aleatorios repentinos con la presencia de alguna fracturas, etc.

Para los Datos Monte Carlo del CampoB se realizaron las mismas medidas de muestra que para los Datos Limpios, en la Tabla C.2 se presenta la Información de Muestra de los campos para los Datos de Monte Carlo esta tabla es el análogo de las Tablas 5.1 para los Datos Limpios.

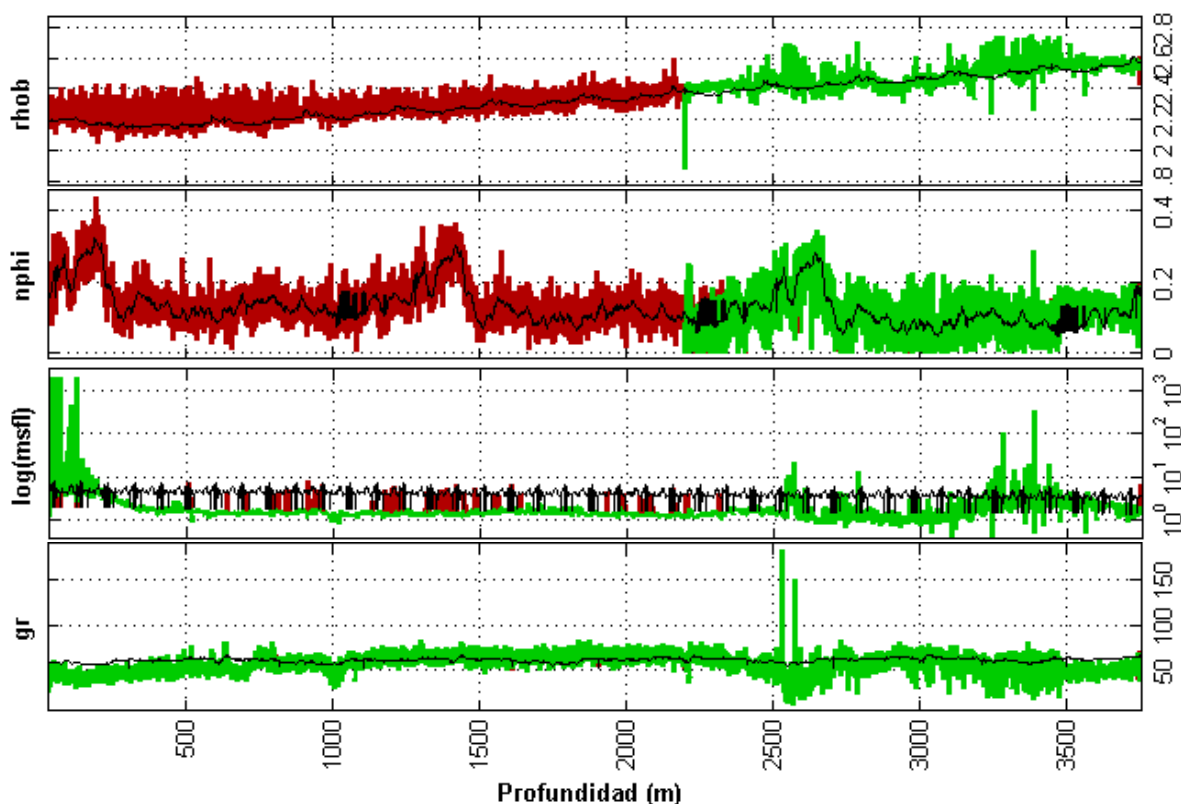


Figura C. 10. Comparación entre la serie de tiempo original,  $Y$ ,  $Y_{ciclo+tendencia}$  y  $Y_{montecarlo}$  para los cuatro registros del pozo 55 del CampoB.

Datos Montecarlo - CampoA						Datos Montecarlo - CampoB					
Registro	No. Datos	Min	Max	Promedio	Std	Registro	No. Datos	Min	Max	Promedio	Std
gr	146125	0.01	287.04	40.34	20.98	gr	220223	6.03	184.13	58.69	13.05
msfl	146125	0.05	19999.99	418.3	1597.75	msfl	220223	0.26	1950	2.95	21.35
nphi	146125	0.0001	1	0.21	0.17	nphi	220223	0.0002	0.82	0.33	0.13
rhob	146125	0.93	4.62	2.59	0.21	rhob	220223	1.87	3.01	2.37	0.11
log(msfl)	146125	-1.32	4.3	1.67	1.02	log(msfl)	220223	-0.59	3.29	0.3	0.26
sx0	133044	0.001	1	0.31	0.24	sx0	212627	0.01	1	0.47	0.19

Tabla C. 2. Información muestral de los registros en los pozos del CampoB para los Datos Monte Carlo.

Se puede observar que con los Datos de Monte Carlo se ha logrado igualar el tamaño de muestra para todos los registros y se ha mantenido el rango de valores, valor medio y desviación estándar observado en las Tablas 5.1. La Tabla C.2 presenta también la información muestral de la Saturación de Agua,  $S_{x0}$ , obtenida teóricamente de la Aplicación de Archie-Pickett expuesta previamente en este capítulo.

En las Tablas B.13 se presentan la información muestral por registro por pozo de los registros de los Datos Monte Carlo de manera más detallada. Es en estas tablas en las cuales

se puede apreciar descriptivamente la modificación realizada a los Datos Limpios con el Método de Monte Carlo descrito.

En las Tablas B.8 se presentan los intervalos de Pickett obtenidos para los Datos de Monte Carlo que aumentaron considerablemente en comparación con sus Tablas análogas B.4. En las Tablas B.8 se presenta también el Intervalo Principal seleccionado para obtener el exponente de cementación y la resistividad del lodo filtrado y el tipo de línea seleccionada en el gráfico de Pickett de dicho intervalo, la información de roca para obtener la Saturación de Agua de los Datos Monte Carlo va a diferir de la presentada en las Tablas B.4 para los Datos Limpios.

## Bibliografía

- [1] Al-Bulushi, Nabil, et al. "Development of artificial neural network models for predicting water saturation and fluid distribution." *Journal of Petroleum Science and Engineering* 68.3 (2009): 197-208.
- [2] Al-Bulushi, N. I., et al. "Artificial neural networks workflow and its application in the petroleum industry." *Neural Computing and Applications* 21.3 (2012): 409-421.
- [3] Amiri, Morteza, et al. "Improving water saturation estimation in a tight shaly sandstone reservoir using artificial neural network optimized by imperialist competitive algorithm—A case study." *Journal of Petroleum Science and Engineering* 127 (2015): 347-358.
- [4] Archie, Gustave E. "The electrical resistivity log as an aid in determining some reservoir characteristics." *Transactions of the AIME* 146.01 (1942): 54-62.
- [5] Arroyo Carrasco, F.A., *Bases Teóricas e Interpretación de Registros Geofísicos de Pozos*, México, UNAM, Facultad de Ingeniería, 1996.
- [6] Barros, Carolina, and André Andrade. "Determination of water saturation by angular competitive neural network." *Journal of Petroleum Science and Engineering* 102 (2013): 47-56.
- [7] Cheng, Jie, David A. Bell, and Weiru Liu. "Learning belief networks from data: An information theory based approach." *Proceedings of the sixth international conference on Information and knowledge management*. ACM, 1997.
- [8] Da Silva Camargo, Sandro, and Paulo Martins Engel. "Artificial neural networks for predicting petroleum quality." *Journal of Applied Computing Research* 1.1 (2011): 61-68.
- [9] Dahlhaus, Rainer, and Michael Eichler. "Causality and graphical models in time series analysis." *Oxford Statistical Science Series* (2003): 115-137.
- [10] Darwin V.E. y Julian M.S., *Well Logging for Earth Scientists*; Springer, 2do. Edition, 2008.
- [11] Del Ángel Morales, Juan Gerardo, et al. "Methodology to Determine the Remaining Oil Saturation (ROS) in the Gas Cap of a Highly Fractured Carbonate Field in the Gulf of Mexico." *SPWLA 54th Annual Logging Symposium*. Society of Petrophysicists and Well-Log Analysts, 2013.
- [12] Eichler, Michael. *Markov properties for graphical time series models*. Technical report, Dept. Statistics, Univ. Heidelberg, 2001.
- [13] El-khatib, Noaman. "A fast and accurate method for parameter estimation of Archie saturation equation." *Middle East Oil Show and Conference*. Society of Petroleum Engineers, 1997.
- [14] Ewings, Sean M., et al. "A Bayesian network for modelling blood glucose concentration and exercise in type 1 diabetes." *Statistical methods in medical research* (2014): 0962280214520732.
- [15] Granger, Clive WJ. "Investigating causal relations by econometric models and cross-spectral methods." *Econometrica: Journal of the Econometric Society* (1969): 424-438.



- [16] Green, P. L. "Bayesian system identification of a nonlinear dynamical system using a novel variant of simulated annealing." *Mechanical Systems and Signal Processing* 52 (2015): 133-146.
- [17] Green, P. L., E. J. Cross, and K. Worden. "Bayesian system identification of dynamical systems using highly informative training data." *Mechanical systems and signal processing* 56 (2015): 109-122.
- [18] Hearst, J.R., Nelson, P.H., *Well Logging for Physical Properties*, Mc Graw Hill Book Co, Nueva York, 1985.
- [19] Helle, Hans B., and Alpana Bhatt. "Fluid saturation from well logs using committee neural networks." *Petroleum Geoscience* 8.2 (2002): 109-118.
- [20] Hyvärinen, Aapo, Juha Karhunen, and Erkki Oja. *Independent component analysis*. Vol. 46. John Wiley & Sons, 2004.
- [21] Kadhim, Fadhil Sarhan, Ariffin Samsuri, and Ahmad Kamal. "A review in correlation between cementation factor and carbonate rock properties." *Life Sci. J* 10.4 (2013): 2451-2458.
- [22] Kamalyar, Kiumars, Y. Sheikhi, and M. Jamialahmadi. "Using an artificial neural network for predicting water saturation in an iranian oil reservoir." *Petroleum Science and Technology* 30.1 (2012): 35-45.
- [23] Mitchell, Tom M. "Machine learning. 1997." Burr Ridge, IL: McGraw Hill 45 (1997): 37.
- [24] Mohaghegh, Shahab, et al. "Petroleum reservoir characterization with the aid of artificial neural networks." *Journal of Petroleum Science and Engineering* 16.4 (1996): 263-274.
- [25] Mollajan, Amir, and Hossein Memarian. "Estimation of water saturation from petrophysical logs using radial basis function neural network." *J. Tethys* 1.2 (2013): 156-163.
- [26] Mollajan, Amir. "Application of local linear neuro-fuzzy model in estimating reservoir water saturation from well logs." *Arabian Journal of Geosciences* 8.7 (2015): 4863-4872.
- [27] Pearl, J., and S. Russell. "Handbook of Brain Theory and Neural Networks, chapter Bayesian Networks." Cambridge (2002).
- [28] Pearl, Judea. "Probabilistic reasoning in intelligent systems: Networks of plausible reasoning." (1988).
- [29] Pearl, Judea. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 2014. (1988)
- [30] Pickett, G. R. "A review of current techniques for determination of water saturation from logs." *Journal of Petroleum Technology* 18.11 (1966): 1-425.
- [31] Pickett, G. R., and D. S. Artus. "Prediction from Logs of Recoverable Hydrocarbon Volume, Ordovician Carbonates: Williston Basin." *Geophysics* 35.1 (1970): 113-123.
- [32] Sijbers, Jan, and A. J. Den Dekker. "Maximum likelihood estimation of signal amplitude and noise variance from MR data." *Magnetic Resonance in Medicine* 51.3 (2004): 586-594.
- [33] SPRITES, P. GLYMOUR, and C. Scheines. "R. 2001. Causation, Prediction, and Search."
- [34] Trewin, B., and S. Morrison. "Reconciliation of core and log residual oil saturation through application of in situ saturation monitoring." *International Symposium of the Society of Core Analysts*. 1992.

- [35] Wang, Baijie, Xin Wang, and Zhangxin Chen. "A hybrid framework for reservoir characterization using fuzzy ranking and an artificial neural network." *Computers & Geosciences* 57 (2013): 1-10.
- [36] Zerrouki, Ahmed Ali, Tahar Aïfa, and Kamel Baddari. "Prediction of natural fracture porosity from well log data by means of fuzzy ranking and an artificial neural network in Hassi Messaoud oil field, Algeria." *Journal of Petroleum Science and Engineering* 115 (2014): 78-89.
- [37] Larionov, V. V. (1969). *Borehole radiometry*. Nedra, Moscow.
- [38] Stieber, S. J. (1970, January). Pulsed Neutron Capture Log Evaluation-Louisiana Gulf Coast. In *Fall Meeting of the Society of Petroleum Engineers of AIME*. Society of Petroleum Engineers.
- [39] Clavier, C., Hoyle, W., & Meunier, D. (1971). Quantitative interpretation of thermal neutron decay time logs: part I. Fundamentals and techniques. *Journal of Petroleum Technology*, 23(06), 743-755.
- [40] Waxman, M. H., & Smits, L. J. M. (1968). Electrical conductivities in oil-bearing shaly sands. *Society of Petroleum Engineers Journal*, 8(02), 107-122.
- [41] Brockwell, P. J., & Davis, R. A. (2013). *Time series: theory and methods*. Springer Science & Business Media.
- [42] Glossary, S. O. (2010). Schlumberger. Retrieved from the internet < URL: <http://glossary.connect.slb.com> > Terms: Electrical Stability Test, Electrical Resistivity, Invert Emulsion.
- [43] Guyod H: Fundamental Data for the Interpretation of Electric Logs, *The Oil Weekly* 115, no. 38 (30 de octubre de 1944): 21-27.
- [44] Winsauer WO, Shearin HM, Masson PH y Williams M: Resistivity of Brine-Saturated Sands in Relation to Pore Geometry, *AAPG Bulletin* 36 (1952): 253-277.
- [45] <http://www.ingenieriadepetroleo.com/registros-sp-y-gamma-ray.html/> Consulta: 10 de Abril de 2017.
- [46] Earl Cook W. H. Freeman "Man Energy and Society" 1976.
- [47] Castillo, E.; Gutiérrez, M.; Hadi, A. *Sistemas expertos y modelos de redes probabilísticas*. Universidad de Cantabria.[En red]. 2008.
- [48] Anzai, Y. (2012). *Pattern recognition and machine learning*. Elsevier.
- [49] Oded Maimon and Lior Rokach (2010). *Data Mining and Knowledge Discovery Handbook*. Springer, New York.
- [50] Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2-3), 191-203.
- [51] Sturges, H. A. (1926). The choice of a class interval. *Journal of the American Statistical Association*, 21(153), 65-66.
- [52] Venables, W. N., & Ripley, B. D. (2013). *Modern applied statistics with S-PLUS, "Density Estimation"*. Springer Science & Business Media.