



UNIVERSIDAD NACIONAL DE COLOMBIA

# **Modelo adaptativo para la estimación de la dimensión fractal de registros de imágenes resistivas a partir de registros geofísicos de pozo**

**Juan Gabriel Agudelo Gamboa**

Universidad Nacional de Colombia

Facultad de Ingeniería, Departamento de Ingeniería de Sistemas e Industrial

Bogotá, Colombia

2016



# **Modelo adaptativo para la estimación de la dimensión fractal de registros de imágenes resistivas a partir de registros geofísicos de pozo**

**Juan Gabriel Agudelo Gamboa**

Tesis de investigación presentada como requisito parcial para optar al título de:  
**Magister en Ingeniería – Ingeniería de Sistemas y Computación**

Director:

M.Sc., Luis Hernán Ochoa Gutiérrez

Línea de Investigación:

Ciencias de la Computación - Ciencias de la Tierra, Geofísica.

Universidad Nacional de Colombia

Facultad de Ingeniería, Departamento de Ingeniería de Sistemas e Industrial

Bogotá, Colombia

2016



*The road and the tale have both been long, would you not say so? The trip has been long and the cost has been high... but no great thing was ever attained easily. A long tale, like a tall Tower, must be built a stone at a time."*

*Stephen King, The Dark Tower*



## **Agradecimientos**

En primera instancia quiero agradecer al profesor Luis Hernán Ochoa por su continuo apoyo a lo largo del proyecto. Las recomendaciones y buenas prácticas fueron fundamentales para el desarrollo del proyecto de investigación. También quiero agradecerle la oportunidad de participar en un proyecto tan interesante.

Un agradecimiento fundamental para el departamento de Ingeniería Industrial y de Sistemas, así como a la Universidad Nacional de Colombia. Su aporte fue fundamental no solo por el acompañamiento y facilidades para desarrollar el proyecto sino también por todas las personas vinculadas que de una u otra forma brindaron soporte y contribución en diálogos e ideas vitales en el transcurso de estos años.

Un agradecimiento muy especial a mi familia por su constante apoyo. A mi madre, María Liliana ya que es y ha sido el soporte fundamental para todos los proyectos emprendidos. A María Paula, mi hermana, por el apoyo, aporte y acompañamiento constante. A Jonathan Alfonso por su disponibilidad, colaboración y diálogos, los cuales fueron vitales para el desarrollo del proyecto. A mis perros, Lupe Margarita y Simón Mauricio, por todo el amor que me brindan y los buenos momentos de esparcimiento y diversión.

Finalmente un agradecimiento profundo y especial para Ángela, por su amor, apoyo, comprensión y motivación.





## Resumen

Los registros de pozo presentan gran variedad de información que permite caracterizar propiedades físicas y químicas de las rocas así como de los fluidos contenidos entre ellas. Los muestreos geofísicos en sitio caracterizan en detalle parámetros físicos de las formaciones que permiten determinar la resistividad, densidad volumétrica, litología, permeabilidad y porosidad entre otros. Dentro de los avances en predicción de fracturas naturales se han desarrollado modelos soportados por procesos dependientes de la dimensión fractal de imágenes resistivas. Sin embargo, no siempre se cuenta con la información asociada a dichas imágenes.

Esta investigación tiene como base la información de registros geofísicos de pozo y parte de la hipótesis que existen relaciones significativas entre múltiples registros con el objeto de caracterizar los valores de dimensión fractal. Se implementa un modelo de predicción utilizando técnicas de minería de datos y bajo los lineamientos de la metodología CRISP-DM. El modelo permite de una forma adaptativa el procesamiento de información proveniente de los registros de pozo con el objeto de identificar cuáles de estos registros presentan relaciones significativas para la generación de patrones que inciden en la predicción de valores de dimensión fractal.

El modelo es validado sobre un caso de estudio con información de registros de pozo realizados en la cuenca del Catatumbo, Norte de Santander, Colombia. En los capítulos asociados se presenta el análisis de información, ejecución y evaluación del modelo. El análisis de resultados de ejecución para el modelo de predicción se basa en muestreos aleatorios de valores de registros múltiples en secciones transversales y es evaluado sobre valores aleatorios de dimensión fractal. Como producto del proceso de clasificación se identificaron tres conjuntos de orígenes de datos relevantes como son registros (logs) base, registros (logs) complementarios y registros (logs) no válidos para la generación de valores de dimensión fractal. La predicción asociada al caso de estudio evidenció una respuesta altamente cercana hacia los valores esperados con una tasa de

error baja donde el coeficiente de determinación ( $R^2$ ) resultante es de 0.95, la media absoluta del error igual a 0.003 y el error de la raíz cuadrada de la media (RMSE) es igual a 0.004. La metodología propuesta ofrece una alternativa para generar e introducir valores asociados a la dimensión fractal de imágenes resistivas en escenarios donde se cuenta con información de diversos registros geofísicos de pozo pero donde la información asociada con la dimensión fractal es inexistente.

**Palabras clave:** Minería de datos, CRISP-DM, predicción, metodología de pre procesamiento, filtros digitales, dimensión fractal, registros de pozo.

## **Abstract**

Well logs have a great variety of information to characterize physical and chemical properties of rock formations and its fluid content among them. Geophysical surveys characterize detailed physical parameters of rock formations in order to determine the resistivity, bulk density, lithology, porosity and permeability, inter alia. Among the advances made on prediction of natural fractures, there have been developed different models supported by processes that depend on fractal dimension of resistive images. However, the information associated to such resistivity image logs is not always available.

The current research is based on information from geophysical well logs, and also upon the hypothesis that is possible to identify significant relationships between multiple records in order to characterize the fractal dimension values. A prediction model is implemented through a data mining technique under the guidelines of CRISP-DM methodology. The model allows an adaptively way of information processing based on well logs in order to identify which of these records show significant relationships to generate prediction patterns to estimate fractal dimension values.

The model is validated over a study case with well-log information gathered on-site at the Catatumbo basin, Norte de Santander, Colombia. In the study case chapters it is also presented data analysis, model execution and evaluation. The analysis of the execution results for the prediction model is based on random samples of multiple values in cross section and is evaluated upon random fractal dimension values. As a product of classification process three groups of relevant data sources were identified, such as: base logs, complementary logs and invalid logs for the generation of fractal dimension values. The prediction associated to the case study showed a very accurate response, in regards to the expected result values, with a low error rate where the resulting  $R^2$  is 0.95, the mean absolute error is equal to 0.003 and the root mean squared error is 0.004. The

proposed methodology is an alternative to generate and input values associated to the fractal dimension of resistive images in scenarios where well geophysical logs are available, but information regarding values of fractal dimensions is inexistent.

**Keywords:** Data mining, CRISP-DM, prediction, preprocessing methodology, digital signals, fractal dimension, well logs.

# Contenido

	Pág.
<b>Resumen</b> .....	<b>IX</b>
<b>Abstract</b> .....	<b>XI</b>
<b>Lista de figuras</b> .....	<b>XV</b>
<b>Lista de tablas</b> .....	<b>XVII</b>
<b>Lista de Símbolos y abreviaturas</b> .....	<b>XVIII</b>
<b>Introducción</b> .....	<b>1</b>
<b>1. Planteamiento del problema</b> .....	<b>3</b>
1.1 Estado del arte .....	4
1.2 Objetivos.....	8
1.2.1 Objetivo general .....	8
1.2.2 Objetivos específicos .....	8
1.3 Estructura del documento .....	8
<b>2. Marco teórico</b> .....	<b>11</b>
2.1 Generalidades geográficas y geológicas .....	11
2.1.1 Caracterización geográfica del área en estudio.....	11
2.1.2 Caracterización geológica regional.....	13
2.1.3 Caracterización geológica local.....	15
2.2 Registros de pozo.....	16
2.2.1 Registros eléctricos .....	16
2.2.2 Registros radioactivos .....	17
2.2.3 Registros acústicos .....	18
2.2.4 Otros tipos de registros .....	18
2.3 Fractales.....	20
2.3.1 Dimensión fractal .....	21
2.3.2 Dimensión fractal por conteo de cajas.....	22
2.4 Procesamiento digital de señales .....	23
2.4.1 Herramientas de procesamiento digital de señales .....	24
2.4.2 Conversión de análogo a digital .....	24
A. Muestreo .....	24
B. Cuantización .....	24
C. Análisis espectral .....	25
2.5 Filtros digitales.....	26

2.5.1	Respuesta al impulso.....	27
2.5.2	Respuesta a la frecuencia .....	28
2.5.3	Respuesta al escalón.....	28
2.5.4	Información en el dominio del tiempo.....	29
2.5.5	Información en el dominio de la frecuencia .....	30
2.5.6	Filtro pasa bajas .....	30
2.5.7	Filtro pasa altas .....	31
2.5.8	Filtro media móvil.....	31
2.6	Redes neuronales .....	32
2.6.1	Topología.....	33
2.6.2	Mecanismo de aprendizaje .....	34
2.6.3	Aprendizaje supervisado.....	34
2.6.4	Aprendizaje no supervisado.....	35
2.6.5	Aprendizaje por refuerzo.....	36
2.6.6	Redes neuronales feedforward .....	37
2.6.7	Regla de clasificación bayesiana .....	37
2.6.8	PNN – Redes neuronales probabilísticas.....	38
2.7	Minería de datos.....	40
2.8	CRISP-DM .....	41
2.8.1	Comprensión del proyecto .....	43
2.8.2	Comprensión de los datos .....	43
2.8.3	Preparación de los datos .....	43
2.8.4	Modelado.....	44
2.8.5	Evaluación .....	44
2.8.6	Despliegue, resultados e informe final .....	45
<b>3.</b>	<b>Datos y metodología.....</b>	<b>47</b>
3.1	Comprensión de datos .....	47
3.1.1	Registro de Rayos Gamma.....	48
3.1.2	Registro de imágenes resistivas .....	49
3.2	Preparación de datos .....	50
3.2.1	Recopilar información .....	51
3.2.2	Datos de registros radioactivos.....	52
3.2.3	Datos de registros generales .....	53
3.2.4	Construcción del conjunto de datos .....	54
3.2.5	Procesamiento de imágenes resistivas.....	54
3.2.6	Datos de dimensión fractal.....	56
3.2.7	Datos de filtro media móvil.....	61
3.2.8	Datos de filtro pasa bajas .....	62
3.2.9	Datos de filtro pasa altas .....	63
3.2.10	Homogeneización de información.....	64
3.3	Modelo de predicción de dimensión fractal.....	67
3.4	Evaluación experimental .....	68
<b>4.</b>	<b>Discusión de resultados.....</b>	<b>71</b>
<b>5.</b>	<b>Conclusiones y recomendaciones .....</b>	<b>79</b>
5.1	Conclusiones.....	79
5.2	Recomendaciones.....	82
	<b>Bibliografía .....</b>	<b>85</b>

# Lista de figuras

	Pág.
<b>Figura 2-1:</b> Ubicación del municipio de Tibú en Norte de Santander, Colombia.....	12
<b>Figura 2-2:</b> Cuenca del Catatumbo (Modificado de ANH) .....	14
<b>Figura 2-3:</b> Columna estratigráfica generalizada. Cuenca del Catatumbo (Modificado de ANH)	15
<b>Figura 2-4:</b> Mediciones derivadas de algunos tipos de registro de pozo utilizadas para determinar propiedades de los yacimientos.....	19
<b>Figura 2-5:</b> Escalas para el Triángulo de Sierpinski .....	21
<b>Figura 2-6:</b> Triangulo de Sierpinski .....	22
<b>Figura 2-7:</b> Muestreo de una señal. ....	25
<b>Figura 2-8:</b> Transformada rápida de Fourier de la señal de la figura anterior. ....	26
<b>Figura 2-9:</b> Respuestas de caracterización de filtro sobre una señal. ....	27
<b>Figura 2-10:</b> Respuesta al impulso .....	28
<b>Figura 2-11:</b> Equivalencia de respuesta al impulso y respuesta al escalón.....	29
<b>Figura 2-12:</b> Filtro pasa bajas .....	30
<b>Figura 2-13:</b> Filtro pasa altas .....	31
<b>Figura 2-14:</b> Estructura de una red neuronal.....	32
<b>Figura 2-15:</b> Estructura interna básica de una neurona artificial.....	33
<b>Figura 2-16:</b> Ejemplos generales de topología de ANN - Feedforward network y recurrentes	34
<b>Figura 2-17:</b> Diagrama de proceso de aprendizaje supervisado .....	35
<b>Figura 2-18:</b> Diagrama de proceso de aprendizaje no supervisado .....	36
<b>Figura 2-19:</b> Diagrama de proceso de aprendizaje por refuerzo .....	36
<b>Figura 2-20:</b> Diagrama de proceso red Feedforward.....	37
<b>Figura 2-21:</b> Arquitectura de una PNN.....	38
<b>Figura 2-22:</b> Estructura jerárquica de CRISP-DM [9] .....	41
<b>Figura 2-23:</b> Ciclo de vida de la metodología CRISP-DM [9].....	42
<b>Figura 3-1:</b> Diagrama de actividades fase de preparación de datos.....	50
<b>Figura 3-2:</b> Información disponible por pozo para el caso de estudio.....	51
<b>Figura 3-3:</b> Ejemplo de registros gamma ray y gamma ray espectral caso de estudio - pozo 10.	52
<b>Figura 3-4:</b> Ejemplo de registros generales disponibles para el caso de estudio pozo 6	53
<b>Figura 3-5:</b> Ejemplo de corrección por velocidad en una imagen resistiva .....	54

<b>Figura 3-6:</b>	Ejemplo de normalización estática y dinámica .....	55
<b>Figura 3-7:</b>	Ejemplo de registro de dimensión fractal para el caso de estudio - pozos 6 al 10	56
<b>Figura 3-8:</b>	Ejemplo conteo de pixeles alfombra Sierpinski (Aplicación <i>Imagen Fractal</i> )	57
<b>Figura 3-9:</b>	Ejemplo conteo de pixeles registro de imagen (Aplicación <i>Imagen Fractal</i> )	58
<b>Figura 3-10:</b>	Ejemplo de regresión lineal para determinar la dimensión fractal .....	59
<b>Figura 3-11:</b>	Ejemplo: filtro media móvil en dimensión fractal - pozos 6 al 10 .....	61
<b>Figura 3-12:</b>	Ejemplo: construcción de filtro pasa bajas.....	62
<b>Figura 3-13:</b>	Ejemplo: señal original y señal afectada por filtro pasa bajas .....	63
<b>Figura 3-14:</b>	Ejemplo: construcción de filtro pasa altas.....	63
<b>Figura 3-15:</b>	Ejemplo: señal original y señal afectada por filtro pasa altas .....	64
<b>Figura 3-16:</b>	Ejemplo: señal GROH con el filtro media móvil original .....	66
<b>Figura 3-17:</b>	Ejemplo: señal GROH media móvil valores desplazados .....	66
<b>Figura 3-18:</b>	Modelo de proceso.....	67
<b>Figura 3-19:</b>	Ejemplo: ventana PDPE - valor original y valores desplazados 10 al 13	68
<b>Figura 3-20:</b>	Diagrama actividades proceso de clasificación y validación de modelo..	69
<b>Figura 4-1:</b>	Diagrama de dispersión dimensión fractal y predicción de la dimensión fractal	72
<b>Figura 4-2:</b>	Error predicción de la dimensión fractal combinación seleccionada .....	72
<b>Figura 4-3:</b>	Diagrama de flujo proceso generación de registro de dimensión fractal .	74
<b>Figura 4-4:</b>	Agrupamientos para registros de entrenamiento .....	76
<b>Figura 4-5:</b>	Registros de dimensión fractal y predicción de la dimensión fractal .....	76



## Lista de tablas

	<b>Pág.</b>
<b>Tabla 2-1:</b> Características bioecológicas generales de Tibú [25].....	12
<b>Tabla 3-1:</b> Información general registros de pozo base .....	47
<b>Tabla 3-2:</b> Información general estadística del muestreo de rayos gamma .....	48
<b>Tabla 3-3:</b> Información general estadística del muestreo de imágenes resistivas .....	49
<b>Tabla 3-4:</b> Información general estadística del muestreo de dimensión fractal.....	60
<b>Tabla 4-1:</b> Información de registros de entrada para el modelo de predicción de registros de dimensión fractal.....	75

# Lista de Símbolos y abreviaturas

## Abreviaturas

Abreviatura	Término
ANH	Agencia Nacional de Hidrocarburos
ANN	Artificial Neural Network
API	American Petroleum Institute
BMP	Bit Mapped Picture
CMI	Compact Micro Imager
CRISP-DM	Cross Industry Standard Process for Data Mining
FFT	Fast Fourier Transform
IFFT	Inverse Fast Fourier Transform
KNIME	Konstanz Information Miner
Mcg	Micrograms
PSD	Power Spectral Density
PNN	Probabilistic Neural Network
RBF	Radial Basis Function
RGB	Red Green Blue
SGS	Segmented Gamma-ray Scanner

## Registros de pozo

Abreviatura	Término
BIT	Bit size
CGXT	Temperature
CLDC	Density Caliper
DCOR	Density Corrected
DDLB	Corrected Deep Laterolog
DEN	Compensated Density
DSL	Corrected Shallow Laterolog
GROH	Gamma Ray Open Hole
GCGR	SGS Corrected Gamma Ray
GRGC	Corrected Gamma Ray
GRPO	GR Potassium
GRSG	SGS Gamma Ray
GRTH	GR Thorium
GRUR	GR Uranium
HVOL	Hole Volume
MBTC	Mmr Caliper
MINV	Mmr Micro-log Inverse
MNRL	Mmr Micro-log Normal

---

<b>Abreviatura</b>	<b>Término</b>
NPRL	Limestone Neutron Porosity
PDPE	Photoelectric Factor
SMTU	DST Up-hole Tension
SPDL	Spontaneous Potential
SPGC	Spontaneous Potential
SRES	Registro de Resistividad Promedio



# Introducción

Dentro del marco de la industria de producción de hidrocarburos es de vital importancia la identificación de rocas que permitan la acumulación y explotación de dicho recurso [47]. Los modelos tradicionales se fundamentan en el análisis de rocas sedimentarias de alta porosidad, las cuales almacenan el hidrocarburo en sus poros o espacios vacíos, despreciando la posibilidad de explotar dichos hidrocarburos en rocas de porosidad muy baja debido a su grado de dificultad [34]. Sin embargo, los reservorios con rocas de alta porosidad han sido explotados en su mayoría por lo que se ha iniciado la búsqueda en rocas fracturadas [47] que a pesar de tener baja porosidad puedan almacenar hidrocarburos en dichas fracturas [34]. Para la caracterización de intervalos naturales con fracturas abiertas se han generado múltiples tipos de modelos de predicción dentro de los cuales destacamos los que tienen como base la dimensión fractal de imágenes resistivas [50] generadas a partir de otros registros geofísicos de pozo tales como rayos gamma, densidad y resistividades de diferente penetración [25].

Gracias a la evolución de los dispositivos y la capacidad de manipulación de información se encuentran múltiples oportunidades para el empleo de nuevos métodos de procesamiento para la obtención de información nueva, eficiente y relevante, lo que permite complementar el proceso de caracterización de información y toma de decisiones [54] [3] [21]. La investigación a desarrollar parte de las posibilidades actuales de procesamiento y relación de grandes volúmenes de datos con variables de interés generando nuevos métodos para los procesos involucrados en la identificación de fracturas naturales abiertas.



# 1. Planteamiento del problema

Dentro del marco de la identificación de fracturas naturales abiertas existen en la actualidad modelos que emplean registros de dimensión fractal generados a partir de registros de imágenes resistivas obtenidos mediante herramientas geofísicas en pozo [25] [50]. Leal, propone un modelo que permite de una manera automática identificar los posibles puntos de afectación por fracturas naturales abiertas donde los datos de dimensión fractal se integran con registros de rayos gamma y resistividad utilizando máquinas de soporte vectorial [25]. Sin embargo, no siempre se cuenta con la información de imágenes resistivas para obtener dicho registro fractal pero en cambio sí se cuenta con información proveniente de registros vinculados a diversas variables geofísicas relacionadas a muestreos específicos en sitio. Por lo anterior surge la pregunta ¿Es posible generar el registro de dimensión fractal de imágenes resistivas a partir de información de diferentes registros de variables geofísicas?

Para dar respuesta a esta pregunta se implementa un modelo de caracterización para la generación emulada del registro de dimensión fractal teniendo como base información proveniente de diferentes variables como son registros de rayos gamma, resistividad adquirida y porosidad entre otros. La generación del registro de dimensión fractal utilizará técnicas de minería de datos que permitan la identificación automática de patrones asociados a las diferentes características de cada uno de los componentes contenidos en el registro en mención, donde los patrones identificados se encuentran directamente relacionados con las variables disponibles en sitio. Los patrones identificados automáticamente se aplicarán puntualmente en el caso de estudio para determinar los valores asociados a la dimensión fractal objetivo.

El caso estudio se centra en el empleo de atributos asociados al muestreo de variables geofísicas sobre objetivos de referencia previamente adquiridos en la cuenca del Catatumbo, departamento de Norte de Santander - Colombia.

## 1.1 Estado del arte

El análisis fractal ha sido utilizado tradicionalmente para describir sistemas naturales. Dentro del campo de la geología algunas investigaciones se han enfocado en describir sistemas de fallas y fracturas [25]. La dimensión fractal permite describir características fundamentales de sistemas auto-similares [52]. Se resalta que este tipo de caracterización no comprende todos los rasgos geométricos significativos [25]. Sin embargo, podemos identificar claramente la incidencia de la dimensión fractal en múltiples procesos asociados a la predicción de características geológicas particulares. A continuación presentamos algunos trabajos representativos asociados con la aplicación de geometría de fractales y su relación con caracterizaciones geológicas.

Dentro de los principales aportes en este campo encontramos la aplicación de análisis fractal para modelar sistemas de fracturas en el subsuelo aplicado en el campo de extracción de energía geotérmica. El trabajo fue presentado por Watanabe y Takahashi en 1993, teniendo como base modelos de fracturas de dos y tres dimensiones sobre geometría fractal [55]. Los modelos son generados por la distribución aleatoria de la relación fractal asociada a la longitud de fractura  $r$ , el número de fracturas  $N$  y la dimensión fractal  $D$  (ver ecuación (1.1)).

$$N = C * r^{-D} \quad (1.1)$$

En la ecuación 1.1  $C$  corresponde a una constante relacionada con la densidad de las fracturas vinculadas a una masa de roca. El caso de estudio tiene como orígenes de datos información de núcleos en varios pozos. La implementación del modelo permitió obtener datos de transmisibilidad entre pozos en función de la densidad de fracturas y el sistema de rendimiento térmico.

También se resalta la relación entre dimensión fractal y coeficiente de rugosidad en la superficie de fracturas acorde a los mecanismos que dieron origen a las fracturas desde



un punto morfológico. El trabajo fue presentado por SunYan y su equipo en 2001. Se desarrolló sobre un caso de estudio en la cuenca de Tarmim, China [50]. En 2007 Roy, Perfect, Dunne y McKay presentaron un algoritmo que permite evaluar sistemas fracturados desde el punto de vista fractal teniendo como base el método por conteo de cajas (ver ecuación (1.2)).

$$N\alpha r^{-D} \quad (1.2)$$

En la ecuación 1.2  $N$  corresponde al número de cajas las cuales cuentan con presencia de fracturas,  $r$  es el tamaño de la caja,  $D$  corresponde a la dimensión fractal y  $\alpha$  representa la relación directamente proporcional. El algoritmo presentado tiene como base un sistema fractal de fracturas sintético sobre dimensiones fractales conocidas. Los resultados asociados se encuentran directamente relacionados en mejoras sobre el método de conteo de cajas para el cálculo de la dimensión fractal. Los resultados tuvieron como base un caso de estudio donde se evidencian claramente altos grados de conectividad entre las fracturas incrementado el flujo de fluidos en el sistema [41].

En 2006 Paredes y otros presentan un trabajo sobre un mapa de trazas de fallas teniendo como base la interpretación de imágenes satelitales, fotografías aéreas, modelos digitales de elevación y datos bibliográficos. El caso de estudio se desarrolló en la isla Decepción en la Antártida donde el objetivo principal fue la definición de zonas tectónicas. Se identificaron seis áreas asociadas con diferentes patrones de fracturas y procesos volcánicos. La dimensión fractal probabilística  $D$  se computa utilizando la media del modelo fractal de Levy-Lee. Los comportamientos asociados corresponden a tres categorías donde el valor es mayor, inferior o igual a 1.5 [40].

En 2013 Leal presenta un modelo que permite de una manera automática identificar los posibles puntos de afectación por fracturas naturales abiertas utilizando imágenes resistivas y datos de dimensión fractal. Los datos de dimensión fractal se integran con registros de rayos gamma y resistividad utilizando máquinas de soporte vectorial. El modelo implementado permite la clasificación de fracturas naturales a partir del contraste de resistividad producido por la roca huésped y el material contenido en la fractura. El cálculo asociado a la dimensión fractal  $D$  se realizó por el método de conteo de cajas. El

modelo presentado utiliza minería de datos y máquinas de soporte vectorial. El caso de estudio corresponde a información de muestreo de pozos en la cuenca del Catatumbo, Norte de Santander, Colombia. La validación se realizó sobre matrices de confusión por pozo donde el porcentaje de exactitud se encuentra en el rango de 72.3% a 83.8% [25].

En 2015 Li y Huang presentan la relación entre el coeficiente de rugosidad conjunta y la dimensión fractal en superficies de rocas fracturadas. Se presenta un análisis sobre ecuaciones empíricas relacionadas con la dimensión fractal y el coeficiente de rugosidad conjunta [27].

En el periodo comprendido entre los años 2015 y 2016, Liu et al presentan dos trabajos asociados a un modelo fractal basado en ecuaciones de flujo de fluidos en fracturas para caracterizar propiedades hidráulicas de redes de fracturas. Los dos trabajos son complementarios, siendo en 2016 presentada la actualización para una nueva ecuación de gobierno. El modelo adopta la dimensión fractal  $D_f$  para representar características de distribución geométrica de fracturas. Así mismo utiliza  $D_T$  para representar la tortuosidad del flujo de fluido inducido por la superficie de rugosidad de fracturas sencillas en redes discretas de fracturas. Los resultados indican que la permeabilidad de una red de fracturas puede ser significativamente influenciada por la tortuosidad del flujo de fluido y la apertura de las fracturas [28] [29].

$$e^{6-D_T} (1.3)$$

En la ecuación (1.3)  $e$  representa la apertura y se evidencia la proporcionalidad en la tasa de flujo. El estudio presentado tiene como base un número aleatorio utilizado para la generación de la distribución de distancia fractal en la red de fracturas [28] [29].

Para concluir resaltamos la importancia de la dimensión fractal en múltiples estudios en el campo de la geología. Típicamente esta se encuentra asociada al procesamiento de imágenes y se encuentra vinculada como punto de partida de trabajos específicos aplicados. Sin embargo la investigación asociada con métodos para predecir la dimensión fractal no ha sido abordada en el área. Se encuentran algunos estudios

relacionados en otros campos del conocimiento. A continuación presentamos varios trabajos representativos donde la predicción de dimensión fractal es el objetivo final.

En 2012 Sahoo y Barman presentan el modelado de una red neuronal para predecir la dimensión fractal en operaciones de máquina. El estudio considera la dimensión fractal asociada a la caracterización de la rugosidad en la superficie de la parte de la máquina involucrada en proceso respectivo. Para el caso de estudio desarrollado se utilizan 4 máquinas vinculadas a los procesos de fresado, torneado, rectificación cilíndrica y el mecanizado de piezas mediante electroerosión por hilo realizado en piezas de trabajo en acero dulce. Los datos de rugosidad en la superficie son utilizados para desarrollar modelos que permitan predecir la dimensión fractal utilizando redes neuronales artificiales. Las entradas de la red corresponden a la velocidad de las piezas de trabajo, la alimentación longitudinal y la penetración radial. La implementación corresponde a una red multicapa del tipo feed-forward [42].

Wada y Kuwana presentan en 2010 un método numérico para predecir la dimensión fractal asociada a una flama durante una explosión de gas. Para el estudio se utiliza una simulación numérica de una propagación esférica de flama. La dimensión fractal objetivo es determinada con base en la velocidad de propagación donde los resultados son evaluados sobre datos experimentales existentes. En el método propuesto la estructura fractal es inducida por inestabilidad hidrodinámica. La velocidad de la flama es calculada en diferentes tamaños de dominio computacional para un periodo de tiempo dado. El valor aumenta relativamente constante de acuerdo al aumento en el tamaño del dominio computacional debido a la naturaleza fractal de la estructura de la flama [53].

Hu y Hassanein presentan en 2014 un modelo para estimar la dimensión fractal en materiales bombardeados con iones. El trabajo es desarrollado partiendo del simulador de iones *Monte Carlo*. El objeto de estudio es la determinación del impacto de la rugosidad en la superficie en la dependencia angular del rendimiento de pulverización catódica. Superficies fractales aleatorias son generadas por un algoritmo de desplazamiento de punto medio combinando la aplicación del simulador y máquinas de soporte vectorial. El estudio se desarrolla sobre simulaciones de dependencia angular en varias combinaciones objetivo de iones donde la entrada de rugosidad en la superficie corresponde a datos experimentales [19].

## **1.2 Objetivos**

### **1.2.1 Objetivo general**

Implementar un modelo adaptativo para la predicción del registro de dimensión fractal de imágenes resistivas basado en registros de variables geofísicas tales como rayos gamma, resistividad y porosidad entre otros.

### **1.2.2 Objetivos específicos**

Generar un modelo de predicción de registro de dimensión fractal teniendo como base el cálculo de dimensión fractal generado a partir de imágenes resistivas. El modelo permite identificar topológicamente la relación entre registros de caracterización de información y realizar el emparejamiento de objetos. La validación se realiza sobre correlaciones con los resultados obtenidos de la generación manual de registros de dimensión fractal.

Caracterizar el registro de dimensión fractal generado utilizando técnicas de minería de datos teniendo como base registros de dimensión fractal y la respectiva asociación con registros de radioactividad, porosidad y muestreos geofísicos tomados en sitio. Para el caso de estudio se utilizarán datos provenientes de registros adquiridos en la cuenca del Catatumbo, departamento de Norte de Santander – Colombia.

Generalizar el modelo aplicado sobre casos de estudio particular para determinar el modelo adaptativo de generación del registro de dimensión fractal de imágenes resistivas basado en registros de variables geofísicas.

## **1.3 Estructura del documento**

En el capítulo dos se presentan los diferentes temas relacionados con el marco teórico. En la primera sección se abordan las generalidades geográficas y geológicas de la cuenca del Catatumbo y las formaciones de yacimientos de hidrocarburos donde se realiza la toma de muestras de los diferentes registros de pozo. En principio se realiza la caracterización que permite identificar factores relevantes asociados con el contexto general de la información para el caso de estudio. En la segunda sección se explican las características particulares de los registros disponibles así como los valores objetivos. Se

---

describen los diferentes tipos de registros geofísicos de pozo realizando un apartado para abordar los conceptos generales de fractales y dimensión fractal ya que la dimensión fractal corresponde al registro más significativo y valor objetivo del caso de estudio. En la tercera parte se incluye la información relacionada con las áreas involucradas en la generación del modelo. En una primera instancia con la información general de registros de pozo se hace un análisis del procesamiento de señales junto con los respectivos filtros digitales que aplican para el caso de estudio. Adicionalmente se abordan los conceptos generales de redes neuronales y minería de datos donde se desarrolla un apartado particular asociado a la metodología CRISP-DM que soporta el proceso realizado en la investigación.

En el capítulo tres se desarrolla la información del proceso realizado acorde a la metodología. Se presentan en detalle la comprensión de los datos teniendo como base las características de origen. Así mismo, la preparación de datos donde se abordan en detalle los aspectos relacionados con la selección de registros, construcción de registros, implementación de filtros media móvil, pasa baja y pasa alta para los diferentes registros disponibles, así como la homogeneización de información. Adicionalmente se encuentra en detalle el modelo desarrollado explicando las características de cada una de las etapas implementadas. El capítulo finaliza con la respectiva evaluación del modelo realizado.

En el capítulo cuatro se encuentra el informe de resultados del proceso realizado para el caso de estudio aplicado sobre el modelo implementado y la evaluación de la investigación desarrollada.

El documento culmina con las conclusiones y recomendaciones generales asociadas al proyecto de investigación. Adicionalmente se encuentra la información de los apéndices y la bibliografía de referencia.



## 2. Marco teórico

A continuación se presentan los diferentes aspectos teóricos más relevantes que intervienen en los diversos procesos ejecutados durante el transcurso de la investigación desarrollada. Se realiza la caracterización geográfica y geológica incluyendo las principales formaciones que inciden en los registros del caso de estudio. Así mismo, los diferentes aspectos generales asociados a tipos de registro geofísicos de pozo. Adicionalmente se abordan las generalidades de los procesos de teoría de fractales y dimensión fractal, procesamiento digital de señales, filtros digitales, redes neuronales y minería de datos desde la metodología CRISP-DM.

### 2.1 Generalidades geográficas y geológicas

#### 2.1.1 Caracterización geográfica del área en estudio

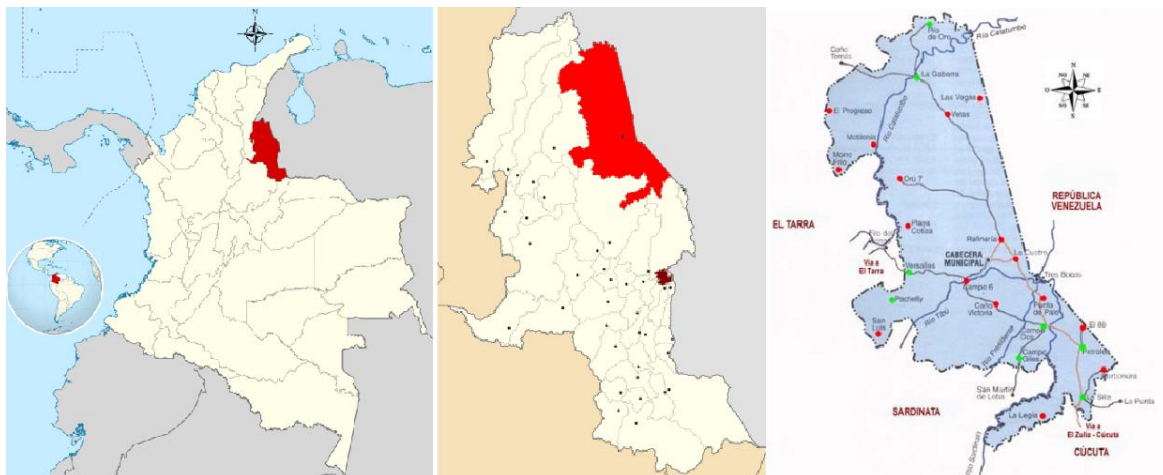
En el municipio de Tibú se encuentra el área donde se localizan los pozos base para el caso estudio. Tibú está localizado en la región nororiental del departamento de Norte de Santander en Colombia al borde del río Tibú siendo sus coordenadas geográficas longitud  $72^{\circ} 59'$ , latitud norte:  $8^{\circ} 39'$ <sup>1</sup>. Está ubicado a 15 minutos de la frontera con Venezuela, en el sitio denominado *Tres bocas* a 125 Km. de la ciudad de Cúcuta, Norte de Santander. Se encuentra rodeado por los ríos Catatumbo, San Miguel, Socuavo Norte, Chiquito, Sardinata, Nuevo Presidente, Tibú, Socuavo Sur y Río de Oro, entre otras numerosas corrientes menores. Cuenta con una extensión total de  $2.737 \text{ Km}^2$ , la altitud de la cabecera municipal es de 75 metros sobre el nivel del mar y cuenta con una temperatura promedio de  $27.3^{\circ}$  centígrados<sup>1</sup>.

---

<sup>1</sup> Latitud y longitud recuperadas de página institucional disponible en línea. Vínculo consultado el 22 de enero de 2016 en: [http://tibu-nortedesantander.gov.co/informacion\\_general.shtml](http://tibu-nortedesantander.gov.co/informacion_general.shtml)

El área de estudio donde se localizan los pozos con registros geofísicos cuenta con una extensión aproximada de  $11.69 \text{ Km}^2$  y se encuentra entre las latitudes  $8^{\circ}27'50'' \text{ N}$  y  $8^{\circ}30'7'' \text{ N}$  y las longitudes  $72^{\circ}33'19''$  y  $72^{\circ}34'50''$  [25].

**Figura 2-1:** Ubicación del municipio de Tibú en Norte de Santander, Colombia<sup>2</sup>



**Tabla 2-1:** Características bioecológicas generales de Tibú [25].

Geoférico	Las unidades litológicas presentes en superficie pertenecen a los periodos terciario en las formaciones geológicas Carbonera, León, grupo Guayabo, Necesidad y cuaternarios aluviales del río Petrólea.
Hidrología	El régimen hidrológico del área refleja en forma leve el régimen de lluvias bimodal típico de la zona. El sistema de drenaje principal del área pertenece a la cuenca del río Catatumbo. Las subcuencas de mayor importancia en el municipio son la del río Socuavo y la del río Sardinata.

<sup>2</sup> Imágenes disponibles en línea. Vínculos consultados el 22 de enero de 2016 en:

<https://es.wikipedia.org/wiki/Tib%C3%BA>

[https://es.wikipedia.org/wiki/Norte\\_de\\_Santander#/media/File:Colombia\\_-\\_Norte\\_de\\_Santander.svg](https://es.wikipedia.org/wiki/Norte_de_Santander#/media/File:Colombia_-_Norte_de_Santander.svg)

[http://tibu-nortedesantander.gov.co/mapas\\_municipio.shtml?apc=bcxx-1-&x=2336848](http://tibu-nortedesantander.gov.co/mapas_municipio.shtml?apc=bcxx-1-&x=2336848)



**Tabla 2-1:** (Continuación)

Clima	La zona posee dos clases de climas, el primero moderadamente húmedo mega termal de sur a centro y el segundo súper húmedo mega termal en el sector norte. Las precipitaciones anuales varían entre 2862.8 y 4149.7 milímetros con una temperatura promedio anual de 27.3° centígrados.
Forestal	En la zona se aprecia gran desarrollo pecuario y agrícola, con áreas cultivadas en pastos y cultivos de palma. Las zonas con vegetación arbórea se encuentran protegiendo las márgenes de los cuerpos de agua, como los ríos Petrólea, Nuevo Presidente y Socuavo. En estudios multi-temporales de la cobertura vegetal de los años 1958, 1978, 1992 y 2008, se observó una elevada tasa de deforestación.
Fauna	La vegetación en el área varía dependiendo de la alteración antrópica. Existen especies de fauna que expresan amplias tolerancias ecológicas en donde se aprovechan los fragmentos y parches de bosques así como también puede encontrar recursos para su supervivencia en lugares perturbados por el hombre.

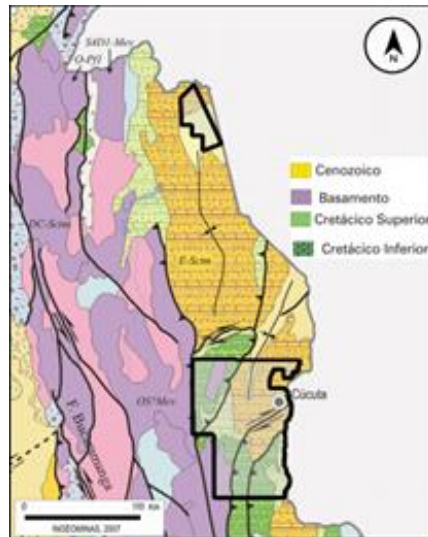
### 2.1.2 Caracterización geológica regional

La cuenca del Catatumbo es una subcuenta de la cuenca Maracaibo. Es una región compartida que se extiende desde el noreste de Colombia hasta Venezuela. Comprende una extensión de 9000 km<sup>2</sup>. Los límites geográficos son al norte con el nacimiento del río Intermedio en la Serranía del Perijá y Motilones, aguas abajo hasta su desembocadura en el río Norte y de este hasta su desembocadura en el río de Oro. Geológicamente es una depresión que se encuentra limitada por la Serranía de Perijá - Macizo de Santander (oeste) y por los Andes de Mérida (sudeste) [5].

Esta cuenca se caracteriza por la producción de petróleo de crudo liviano, medio y gas. Se han descubierto alrededor de once campos. Entre los más importantes encontramos Tibú, Río Zulia, Sardinata y Carbonera [5].

En la cuenca del Catatumbo se concentran rocas pertenecientes al periodo Cretácico del tipo areniscas marinas, lutitas y calizas de ambientes marinos someros. Adicionalmente encontramos rocas vinculadas al periodo Cenozoico del tipo lutitas fluvial-deltaicas y areniscas que se depositaron en una cuenca anti-país (foreland) [5].

**Figura 2-2:** Cuenca del Catatumbo (Modificado de ANH<sup>3</sup>)



Las características asociadas a la porosidad se observan mejor desarrolladas en areniscas del Paleógeno y las trampas se encuentran controladas por corrimientos de estructuras de anticlinales fallados, los cuales son el resultado de fallas de rumbo convergentes. El petróleo se generó en rocas del Cretácico Superior de la Formación la Luna y rocas del Cretácico Inferior pertenecientes al Grupo Uribante. Se caracterizan porque la migración comenzó en el Eoceno tardío y se sigue generando hasta el día de hoy [5] [2].

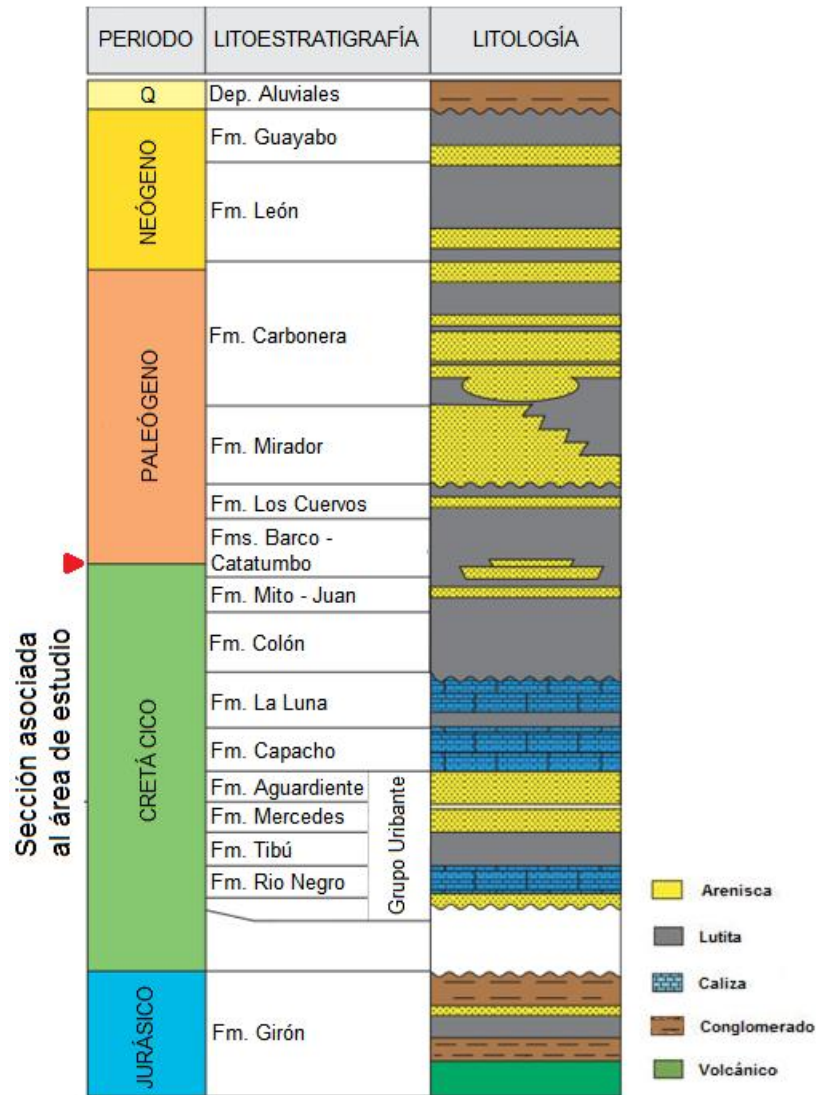
---

<sup>3</sup>ANH (Agencia Nacional de Hidrocarburos). (2010).Cuenca Catatumbo. Open Round Colombia 2010. Documento disponible en línea. Vínculo consultado el 22 de enero de 2016 en: [http://www.anh.gov.co/Informacion-Geologica-y-Geofisica/Estudios-Integrados-y-Modelamientos/Presentaciones%20y%20Poster%20Tcnicos/Catatumbo%20\(pdf\)-Ronda%20Colombia%202010.pdf](http://www.anh.gov.co/Informacion-Geologica-y-Geofisica/Estudios-Integrados-y-Modelamientos/Presentaciones%20y%20Poster%20Tcnicos/Catatumbo%20(pdf)-Ronda%20Colombia%202010.pdf)

### 2.1.3 Caracterización geológica local

Teniendo en cuenta la geología de superficie y la perforación de pozos se determinaron los domos norte y sur asociados al área de estudio.

**Figura 2-3:** Columna estratigráfica generalizada. Cuenca del Catatumbo (Modificado de ANH<sup>4</sup>)



<sup>4</sup>ANH (Agencia Nacional de Hidrocarburos). (2012).Cuenca Catatumbo. Integración Geológica de la Digitalización y Análisis de núcleos. P 106-135.Documento disponible en línea. Vínculo consultado el 22 de enero de 2016 en: <http://www.anh.gov.co/Informacion-Geologica-y-Geofisica/Tesis/4.%20Informe%20Final%20Catatumbo.pdf>

Los domos identificados corresponden a cierres estructurales. El domo norte es el campo más somero de toda la cuenca compuesto de rocas calcáreas principalmente. Se caracteriza por la presencia de un fracturamiento considerable [13] [2]. En este campo se encuentran los seis pozos asociados al caso de estudio (P-6, P-7, P-8, P-9, P-10, P-11). La secuencia estratigráfica atravesada por estos pozos corresponde a roca de edad Cretácica [25].

## 2.2 Registros de pozo

Para la caracterización litológica de cada pozo se realizan diferentes tipos de muestreos por medio del uso de sondas. Los registros resultantes corresponden a una serie de medidas en un rango de profundidad determinado. Típicamente múltiples herramientas son utilizadas simultáneamente ya que la combinación de información de resultados es significativamente más representativa en comparación con los registros aislados. El proceso de toma de registros tiene lugar después de perforar [11]. Según la característica objetivo de la medición podemos agrupar los diferentes registros en mediciones eléctricas, radioactivas o acústicas<sup>5</sup>.

### 2.2.1 Registros eléctricos

Las rocas profundas encontradas en el subsuelo típicamente se hallan saturadas con agua salada donde esta cuenta con una resistividad menor a la del petróleo y el gas. Gracias a esta característica, es posible determinar el contenido de fluido de una roca por medio de la resistencia del material al paso de la corriente eléctrica. Así, por medio de una sonda la cual puede utilizar o bobinas de inducción o electrodos, es posible generar un circuito que nos permita medir la conductividad o resistividad eléctrica. Los componentes salinos del agua de formación y del lodo de perforación permiten el movimiento de iones generando una actividad electromagnética que corresponde al registro de potencial espontáneo. Al inducir corrientes de alta penetración (Laterolog) se

---

<sup>5</sup>Oilfield Review Spring 2011 - Schlumberger. 2013. Documento disponible en línea. Vínculo consultado el 22 de enero de 2016 en: [https://www.slb.com/~media/Files/resources/oilfield\\_review/spanish11/spr11/definicion\\_del\\_concepto.pdf](https://www.slb.com/~media/Files/resources/oilfield_review/spanish11/spr11/definicion_del_concepto.pdf)

induce un potencial eléctrico entre los electrodos que genera un campo eléctrico a través de los estratos alcanzados. La variación de intensidad permite analizar el tipo de roca ya que se cuenta con un valor de conductividad según el tipo de material. De esta forma es posible promediar la resistividad generando un latero-perfil [11].

### 2.2.2 Registros radioactivos

Según las características de los elementos que se encuentran en el subsuelo es posible por medio del uso de fuentes químicas identificar la cantidad de partículas radioactivas generadas. Según el tipo de proceso los registros de *gamma ray* corresponden a los tipos GR natural o GR emitido. La radioactividad natural corresponde a la emisión espontánea de partículas  $\alpha, \beta$  ó *rayos  $\gamma$*  desde el núcleo de un átomo. Las partículas  $\alpha$  y  $\beta$  cuentan con gran energía pero bajos niveles de penetración por lo que típicamente son despreciadas ya que la aplicación en registros no es práctica. La información de este tipo de registros se encuentra vinculada con la arcillosidad presente en la roca. Los elementos radioactivos tienden a tener una mayor concentración en arcillas y lutitas.

La densidad de la formación incide directamente sobre el grado de absorción. De esta forma si tenemos la misma cantidad de material radiactivo en unidades de volumen sobre diferentes densidades, el registro resultante presenta una alta concentración de radiactividad sobre densidades bajas. Estas características son aplicadas para la detección de GR asociado al torio, potasio y uranio. Se resalta el hecho que el potasio se encuentra en formaciones sedimentarias y emite radiaciones en un solo nivel energético [11].

Los registros asociados a herramientas de litodensidad se encuentran relacionados con el gamma ray emitido y permiten caracterizar la porosidad, densidad y litología. Dentro de las curvas generadas por estos registros encontramos las relacionadas con densidad total, factor fotoeléctrico, densidad y porosidad de la densidad entre otros. Por un lado, minerales tales como los carbonatos y el cuarzo poseen radiactividad intrínseca baja o nula. Por otro lado, rocas tales como las lutitas contienen elementos radioactivos naturales. Así la radiación gamma es absorbida en proporción a la densidad de la formación [11].

Las fuentes radiactivas emiten rayos gamma hacia su entorno donde estos interactúan con los electrones donde los rayos son dispersados por los núcleos. De esta forma los átomos quedan excitados gracias a la energía que perdió la formación produciendo una radiación secundaria. Acorde a la relación del número de rayos gamma contra el número de electrones en la formación es posible determinar el registro de densidad [11].

Adicionalmente existen herramientas de porosidad neutrón las cuales permiten caracterizar la porosidad y litología basadas en neutrones emitidos [11]. Se resalta que los neutrones cuentan con una masa semejante a la del átomo de hidrógeno por lo que son absorbidos en proporción al volumen de hidrógeno. Así los neutrones son disparados a alta energía y al encontrarse con los núcleos del material circundante pierden energía donde la velocidad promedio tiende a ser la velocidad de los átomos del material. Este comportamiento se presenta en elementos como cloro, hidrógeno, y silicio entre otros. Resultado del proceso anteriormente mencionado se produce una emisión de rayos gamma tratando de generar un equilibrio energético. Este tipo de medida corresponde al registro de neutrón [11].

El promedio de las dos medidas anteriores permite determinar los valores de porosidad que corresponde a un registro de densidad-neutrón [11].

### **2.2.3 Registros acústicos**

De acuerdo con la porosidad y la composición mineral es posible transmitir un pulso acústico donde un receptor pueda capturar la transmisión original. De esta forma es posible determinar la velocidad acústica. Se hace particular referencia al hecho que las propiedades mecánicas de los sólidos pueden afectar las características de las ondas acústicas que los atraviesan. Así, es posible determinar las velocidades de las ondas para cuantificar las propiedades en mención [11].

### **2.2.4 Otros tipos de registros**

Para generar información con diferentes niveles de detalle vinculados con las principales características de las rocas se cuenta con múltiples tipos de herramientas que realizan muestreos que permiten determinar diferentes propiedades acorde a la funcionalidad

particular [11]. Dentro de estos otros tipos de registro entre los más representativos encontramos:

- Resonancia magnética o nuclear la cual permite determinar información sobre el tamaño de poros y fluidos.
- Imágenes de caracterización de la formación de las paredes de pozo por medio de mediciones de 360 grados.
- Herramientas de toma de muestras de rocas o fluidos directamente en el subsuelo.

En la figura 2-4 podemos identificar algunos tipos de registro y la forma en que estos pueden permitir identificar directa o indirectamente propiedades propias de las rocas. En la imagen se realiza una caracterización de tres colores azul, verde y ocre. Azul significa que el tipo de registro permite la caracterización de una propiedad directa. Verde que provee información parcial que al ser combinada con otro tipo de mediciones permite determinar propiedades características de las rocas. Ocre indica que el registro es sensible a una propiedad pero no permite su caracterización.

**Figura 2-4:** Mediciones derivadas de algunos tipos de registro de pozo utilizadas para determinar propiedades de los yacimientos<sup>6</sup>

Mediciones	Resistividad	Porosidad	Litología	Mineralogía	Saturación	Geometría de los poros	Permeabilidad	Propiedades de los fluidos	Propiedades geomecánicas	Estructura geológica	Estratificación geológica
Resistividad eléctrica											
Lateroperfil	■		■		■	■	■	■		■	
Inducción	■		■		■					■	■
Microlateroperfil	■		■		■					■	
Potencial espontáneo							■			■	
Propagación electromagnética	■	■	■		■	■	■	■		■	■
Nuclear											
Rayos gamma-densidad		■	■	■				■	■	■	
Porosidad-neutrón		■	■	■				■		■	
Radioactividad natural			■	■						■	
Espectrometría de rayos gamma inducidos		■	■	■			■		■	■	
Resonancia magnética nuclear		■	■	■	■	■	■	■	■	■	
Acústica		■	■	■	■	■	■	■	■	■	
Medidor de buzamiento (echado) y generación de imágenes	■	■	■	■		■		■	■	■	■
Pruebas y muestreo de formaciones											
Muestreo de rocas	■					■	■	■	■	■	■
Muestreo de fluidos					■			■		■	■
Pruebas de presión de fluidos							■	■		■	■
Sísmica		■	■		■			■	■	■	

La medición proporciona información directa sobre la propiedad del yacimiento.
  La medición es afectada por la propiedad del yacimiento o es sensible a ésta.
  La medición contribuye a comprender la propiedad del yacimiento.

<sup>6</sup>Oilfield Review Spring 2011 - Schlumberger."2013. Documento disponible en línea. Vínculo consultado el 22 de enero de 2016 en: [https://www.slb.com/~media/Files/resources/oilfield\\_review/spanish11/spr11/definicion\\_del\\_concepto.pdf](https://www.slb.com/~media/Files/resources/oilfield_review/spanish11/spr11/definicion_del_concepto.pdf)

## 2.3 Fractales

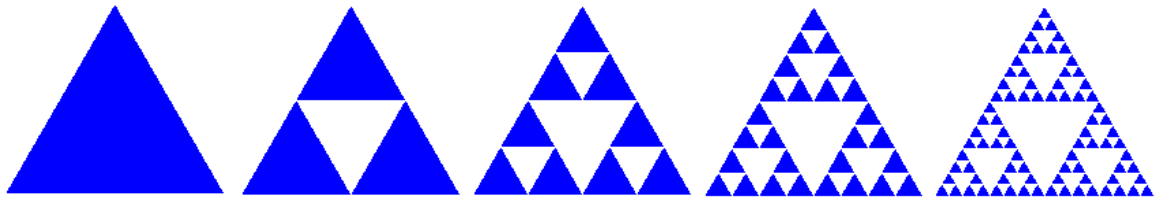
Típicamente las matemáticas se centraron en el estudio de conjuntos y funciones para aplicaciones del cálculo tradicional. Adicionalmente la geometría euclidiana descartaba el trabajar sobre conjuntos o funciones que no eran suficientemente regulares donde estos tendían a ser ignorados patológicamente. Ciertamente, se trataban como curiosidades individuales donde rara vez una teoría general era aplicable. Sin embargo, conjuntos irregulares proveen una amplia representación de fenómenos naturales [4]. Es en este punto donde Benoit Mandelbrot en 1975 introduce el término *fractal* (el cual deriva del latín *fractus* que significa quebrado o fracturado) para describir un objeto semi-geométrico cuya estructura básica, fragmentada o irregular, se repite a diferentes escalas [32].

La característica de semi-geométrico está dada ya que el objeto es más complejo que un cuerpo geométrico regular donde, a simple vista, los bordes son altamente irregulares y a distintas escalas de observación parecen conservar el mismo patrón (Auto-similitud) [14]. Acorde a la definición de Mandelbrot a un objeto geométrico fractal se le atribuyen las siguientes características [32] [4]:

- Demasiado irregular para ser descrito en términos geométricos tradicionales.
- Posee detalle a cualquier escala de observación.
- Es auto-similar (exacto, aproximado o estadísticamente).
- La dimensión de Hausdorff-Besicovitch, es decir la dimensión de un espacio topológico fraccionada no entera, es estrictamente mayor que su dimensión topológica ( $DT \leq DHB$ ).
- Se define mediante un simple algoritmo recursivo.
- No es diferenciable en ningún punto.

Estas propiedades pueden evidenciarse fácilmente en un fractal típico. Como ejemplo en la figura 2-5 visualizamos el triángulo de Sierpinski en escalas 1, 0.5, 2.25, 0.125 y 0.625 respectivamente.



**Figura 2-5:** Escalas para el Triángulo de Sierpinski

El conjunto evidenciado  $S$  tiene tres componentes donde cada uno corresponde al mismo escalado. Cada uno de sus componentes cuenta con tres subcomponentes brindando nueve imágenes en la siguiente escala y así hasta el infinito. A simple observación las funciones sucesivas  $f_1$ ,  $f_2$  y  $f_3$  son contracciones del espacio de factor  $1/n$  con puntos de corrección dados por tres vértices. En la imagen se evidencia en el inicio un triángulo correspondiente a  $T_0$  posteriormente iterando en cada nuevo nivel (ver ecuación (2.1)).

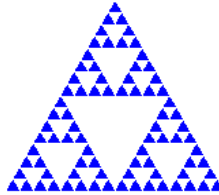
$$T_1 = f(T_0), T_2 = f(T_1), \dots \quad (2.1)$$

Se evidencia claramente que en cada nueva iteración el triángulo de Sierpinski se encuentra convergiendo a  $S$ . El proceso se identifica gráficamente en las múltiples iteraciones mostradas en la Figura 2-5 [4].

### 2.3.1 Dimensión fractal

La dimensión fractal corresponde a un número real que generaliza el concepto de dimensión ordinaria para objetos geométricos que no admiten espacio tangente [4]. Es decir, partiendo de la base que un punto es a-dimensional, una línea es bidimensional y un cubo es tridimensional se dice que la dimensión corresponde a un número que representa la cantidad de direcciones independientes o grados de libertad sobre los que el observador se puede mover en el espacio que contiene al objeto. Es decir la dimensión topológica del mismo. Mandelbrot define dimensión fractal como el número fraccionado que representa la relación de índice estadístico de complejidad que permite comparar cómo el detalle de un patrón fractal cambia con la escala en la que sea medido [32].

Un ejemplo de la dimensión fractal lo encontramos en el triángulo de Sierpinski donde:  $d=1.58496$

**Figura 2-6:** Triángulo de Sierpinski

Es importante resaltar el señalamiento de Talanquer donde afirma que los fractales que existen en la naturaleza, así como las fracturas en un yacimiento, tienden a ser irregulares y auto-similares sólo en sentido estadístico [51] [25]. En la práctica para determinar la dimensión fractal existen múltiples métodos que definen técnicas para aproximar el escalamiento y el detalle de los límites que utilizan regresión lineal, registros, logaritmos, tamaños y escalas, entre otros [24]. Dentro de los principales métodos mencionamos el conteo de cajas (Minkowski–Bouligand) [26], dimensión de información [12], dimensión de correlación [17], dimensiones de Rényi [20], dimensión de Hausdorff-Besicovitch [14], dimensión de Higuchi [18], dimensión Multifractal [35], entre otras [24]. Para nuestro caso de estudio nos centraremos en la dimensión por conteo de cajas dado que los registros de información cuentan con la dimensión fractal determinada por este método.

### 2.3.2 Dimensión fractal por conteo de cajas

Típicamente este método ha sido utilizado para determinar la dimensión fractal a sistemas fracturados [25] [27]. El método de conteo de cajas para determinar la dimensión fractal es apropiado para imágenes con o sin auto-similaridad. Esta técnica incluye el procesamiento de la imagen y la definición de los rangos de tamaño de las cajas [15].

El método consiste en enmarcar la figura objeto de estudio en un cuadrado. Posterior a ello la caja es dividida en cuatro cajas donde el lado corresponde a  $1/r$ , el número de cajas en este punto corresponde a  $N(r)$ , donde cada caja cubre alguna parte de la figura. Posteriormente cada una de las cajas es dividida en cajas de lado  $r$ , y se procede nuevamente con el conteo de cajas  $N(r)$  para las cajas que contengan alguna parte de la

figura. Se repite el proceso iterativamente. Para la representación gráfica se utiliza la ecuación  $\text{Log}(N(r))$  contra  $\text{Log}(1/r)$  [37]. Este proceso permite determinar la pendiente de la recta siendo  $Y_j = \text{Log}(N_j)$  y  $X_j = \text{Log}(1/r)$ , donde la pendiente  $m$  corresponde a la dimensión fractal obtenida (ver ecuación (2.2)) [25]:

$$m = \frac{n \sum_1^n x_j y_j - (\sum_1^n x_j)(\sum_1^n y_j)}{n \sum_1^n x_j^2 - (\sum_1^n x_j)^2} \quad (2.2)$$

## 2.4 Procesamiento digital de señales

En la actualidad los fenómenos físicos luz, sonido, movimiento, registros electromagnéticos, radioactivos, resistividad y porosidad entre otros son analizados mediante instrumentos electrónicos. Los procesos asociados producen señales que contienen la información del comportamiento particular dependiente del tipo de medida objetivo. Las señales se representan matemáticamente como funciones de una o más variables independientes. La variable independiente más común es el tiempo, para el caso de estudio el tiempo es representado mediante los valores recuperados en diferentes muestreos de profundidad [57].

Se resalta el hecho que muchas de las señales colectadas contienen datos que se salen del margen del comportamiento. Esta característica particular es generada debido a que los instrumentos electrónicos que las reciben no tienen un comportamiento totalmente ideal en su función. Adicionalmente, dependiendo del tipo propiedad asociado a la medida realizada las señales capturadas son vulnerables de ser afectadas por aspectos ambientales como temperatura, humedad y ruido entre otros. Los circuitos electrónicos permiten reemplazar los circuitos análogos por circuitos digitales que ocupan un menor volumen y están libres de los problemas de tolerancia de los componentes, calibración y deriva térmica [16].

El procesado digital de señales se ocupa de la representación, transformación y manipulación de señales discretas desde el punto de vista de la información que contienen [30]. A continuación se listan las etapas básicas que intervienen en el procesado digital de señales:

- Conversión de la señal continua en tiempo y amplitud en una señal digital.
- Procesado de la señal digital.
- Conversión de la señal digital procesada en una señal continua.

### **2.4.1 Herramientas de procesamiento digital de señales**

En el procesamiento de señales se aplican diferentes herramientas que permiten la manipulación de las mismas acorde a las necesidades objetivo. A continuación se presenta de una forma general las principales herramientas de conversión de análogo a digital, análisis espectral y filtros digitales.

### **2.4.2 Conversión de análogo a digital**

La salida de los sensores que permiten al equipo electrónico interactuar con el entorno es típicamente una señal analógica continua en el tiempo. En consecuencia, esta información debe convertirse en binaria con el fin de adaptarla a los circuitos procesadores y de presentación. Un convertidor analógico-digital (CAD) es un circuito electrónico integrado cuya salida es la palabra digital resultado de convertir la señal analógica de entrada [56]. Para realizar dicho proceso de conversión típicamente se aplican los procesos de muestreo, cuantización y análisis espectral [16].

#### **A. Muestreo**

Definimos muestreo como la cantidad de veces que medimos el valor de la señal en un periodo de tiempo definido. Según el teorema de Nyquist-Shannon la cantidad de veces que debemos medir una señal para no perder información debe de ser al menos el doble de la frecuencia máxima que alcanza dicha señal [49].

#### **B. Cuantización**

Es el proceso mediante el cual se define la amplitud de una señal en tiempo continuo y es interpretada en tiempo discreto. Normalmente es utilizada la Cuantización uniforme binaria [16].

## C. Análisis espectral

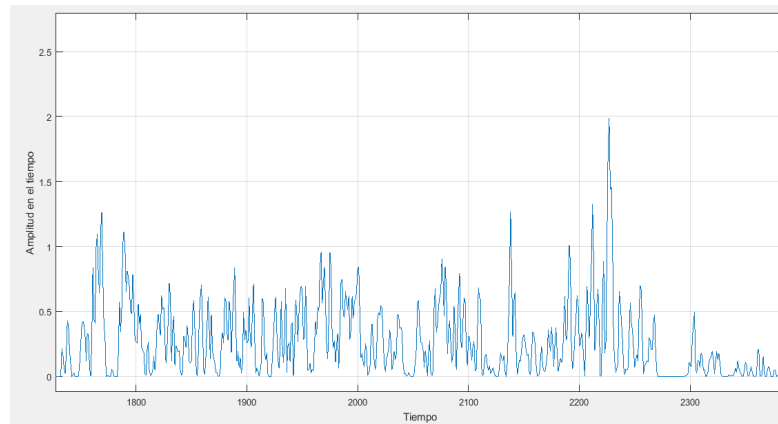
Por un lado tenemos que en muchas ocasiones la información obtenida en una serie de datos está vinculada con el dominio del tiempo existen procesos en donde dicha información no nos llega a ser de utilidad. Por otro lado tenemos el caso en que la información en una señal se encuentra en el dominio de la frecuencia de la misma señal y se requieren procesos particulares de la misma señal. Teniendo en cuenta estos escenarios la transformada de Fourier nos permite determinar el contenido de frecuencias, es decir el espectro de una señal arbitraria, sin necesidad que la señal tenga una periodicidad [39]. El paso al dominio de la frecuencia utilizando la transformada de Fourier nos da como resultado una representación infinita de señales sinusoidales y cosenoidales.

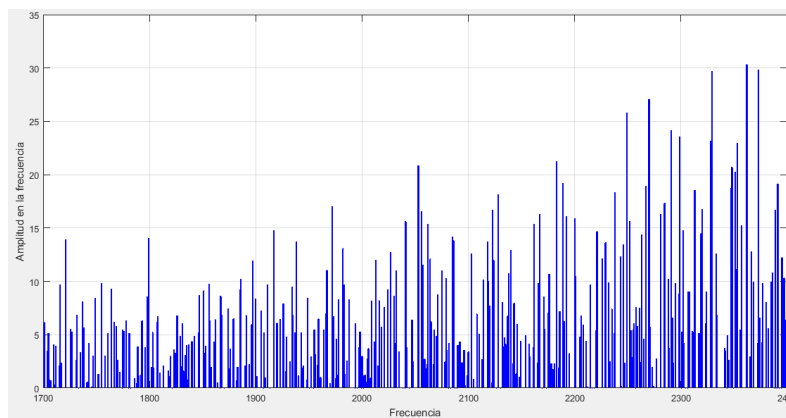
La transformada continua de Fourier se utiliza para señales en tiempo continuo donde  $F$  es una función de  $f: R \rightarrow C$ . Si  $x(n)$  tiene longitud finita se puede recuperar dicha señal a partir de muestras de su Transformada de Fourier. Para esto no es necesario conocer dicha transformada en todas las frecuencias ver ecuaciones (2.3) y (2.4)) [39].

$$F[f(t)] = F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt (2.3)$$

$$F^{-1}[F(\omega)] = f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{j\omega t} d\omega (2.4)$$

**Figura 2-7:** Muestreo de una señal.



**Figura 2-8:** Transformada rápida de Fourier de la señal de la figura anterior.

Las transformadas rápidas de Fourier (FFT) son algoritmos especializados que permiten a un procesador digital hacer el cálculo de la transformada discreta de Fourier de una forma eficiente en lo que respecta a carga computacional y tiempo de procesamiento. El objeto final es realizar la representación de una señal originalmente adquirida en el dominio del tiempo como función del dominio de la frecuencia [16] [39]. El proceso para regresar desde el dominio de la frecuencia hacia el dominio del tiempo es realizado por medio de la transformada inversa (IFFT). De esta particularidad se concluye que se puede interactuar de un dominio particular a cualquier otro. Esta característica permite realizar el análisis espectral de una señal, manipulación y regresar al dominio original para ver cómo es su comportamiento [39].

## 2.5 Filtros digitales

En su definición más general, un filtro se puede definir como toda transformación que altera la naturaleza de una señal de una forma o de otra. De esta forma un filtro digital es un proceso computacional o algoritmo mediante el cual una señal digital de entrada, secuencia de muestras numéricas, es transformada en una segunda secuencia de muestras con especificaciones determinadas o señal digital de salida [16].

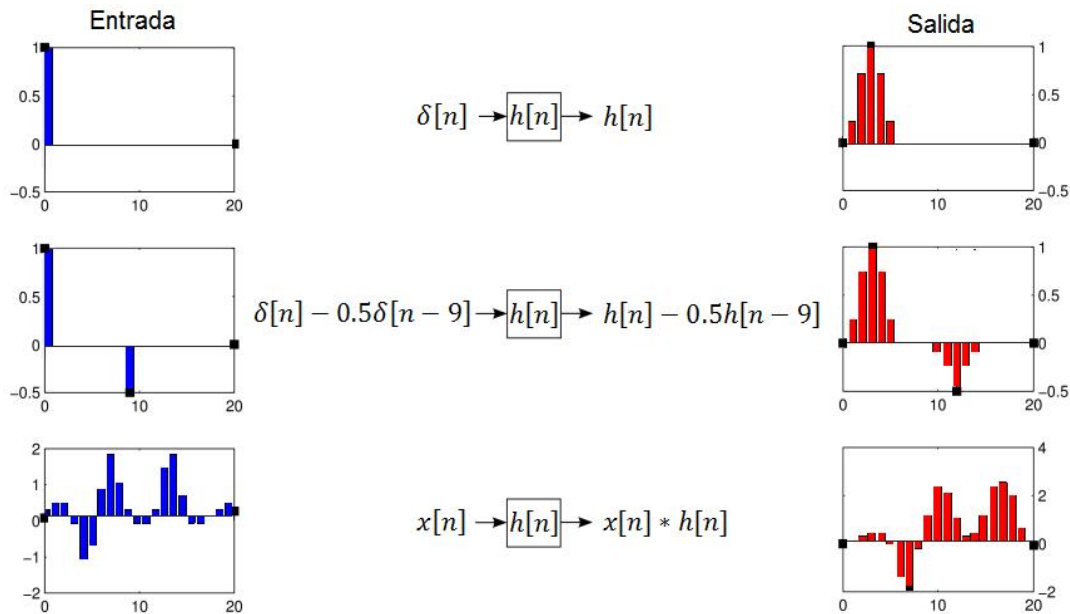
**Figura 2-9:** Respuestas de caracterización de filtro sobre una señal.

Los filtros digitales típicamente se expresan para una frecuencia en particular determinando las frecuencias que permiten pasar de la señal. Es decir, realizar un bloqueo teniendo como base una frecuencia de paso o frecuencia de corte. Adicionalmente cuentan con características como la amplitud de paso y la amplitud de rechazo [49].

Las tres caracterizaciones equivalentes de un filtro son respuesta al impulso, respuesta a la frecuencia y respuesta al escalón (ver figura 2-9) [48].

### 2.5.1 Respuesta al impulso

Conociendo la respuesta presentada ante un impulso es posible determinar la respuesta a cualquier entrada. En matemática se define el principio de superposición como la descomposición de un problema lineal en dos o más sub-problemas más sencillos donde es posible obtener el problema original como una superposición o suma de estos sub-problemas. En electrónica dicho principio se define en teoría de circuitos como que la tensión entre dos nodos de un circuito es igual a la suma de cada una de las tensiones [16] [48].

**Figura 2-10:** Respuesta al impulso

### 2.5.2 Respuesta a la frecuencia

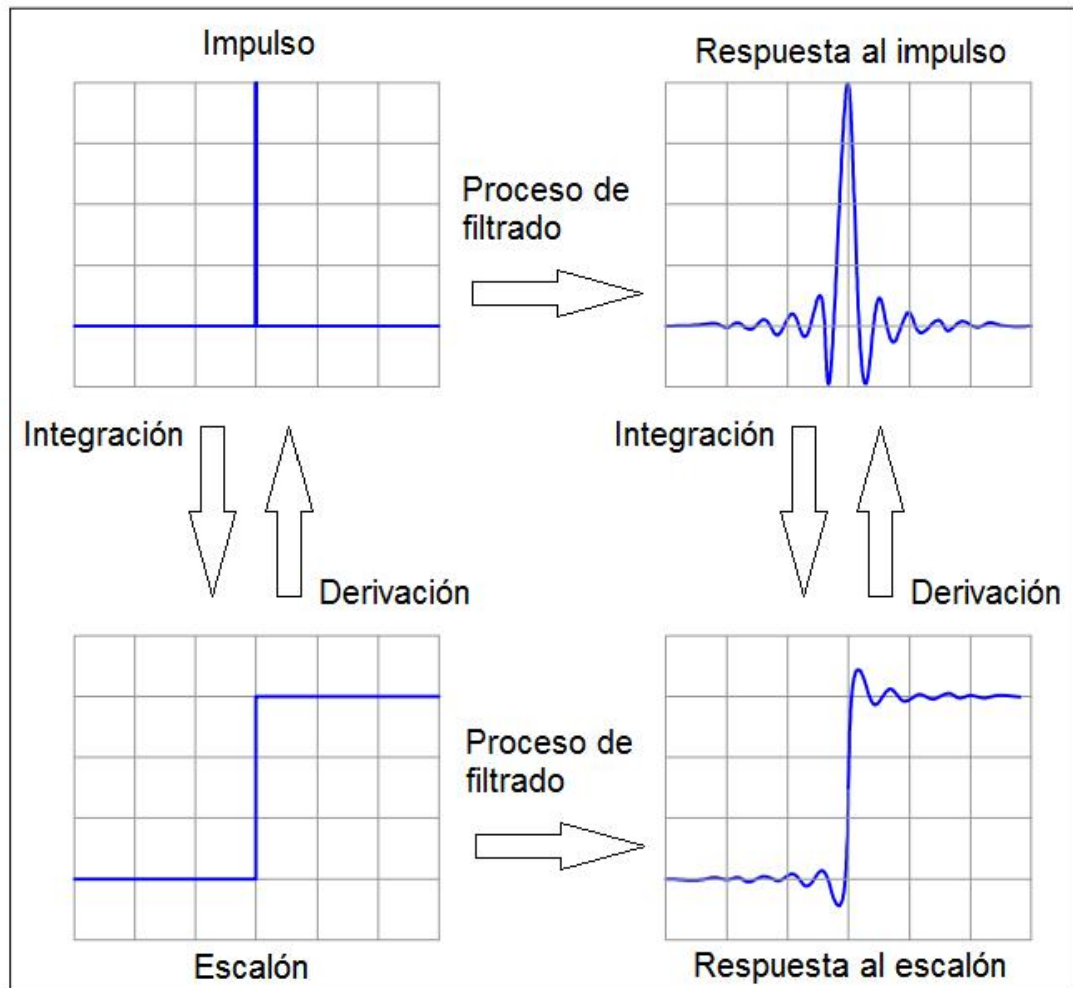
La respuesta a la frecuencia corresponde a la transformada de Fourier en tiempo discreto de la respuesta al impulso. Se caracteriza por ser una función que toma valores complejos periódicos, con periodo  $2\pi$ . Para su representación se puede presentar la parte real en notación cartesiana y la parte imaginaria en notación polar asociada a la magnitud y la fase [48].

### 2.5.3 Respuesta al escalón

La respuesta al escalón se obtiene mediante la integración discreta de un impulso (ver ecuación (2.5)).

$$y[n] = \sum_{k=-\infty}^n x[k] \quad (2.5)$$



**Figura 2-11:** Equivalencia de respuesta al impulso y respuesta al escalón

### 2.5.4 Información en el dominio del tiempo

Se encuentra asociada a la información en el momento en el que ocurren los eventos. Realiza una muestra de la magnitud del evento y se encuentra codificada sobre los intervalos de tiempo y/o longitud para el muestreo asociado. Las señales pueden ser analógicas o digitales, donde las señales analógicas tienen un número finito de valores en un rango. La data análoga generalmente se encuentra asociada a fenómenos naturales como variaciones de temperatura, presión, velocidad, distancia y sonido entre otros. Para el caso de estudio el dominio del tiempo se encuentra representado por la distancia en profundidad para la toma de muestreos. Típicamente la información asociada a este dominio responde adecuadamente a filtro de escalón [16].

### 2.5.5 Información en el dominio de la frecuencia

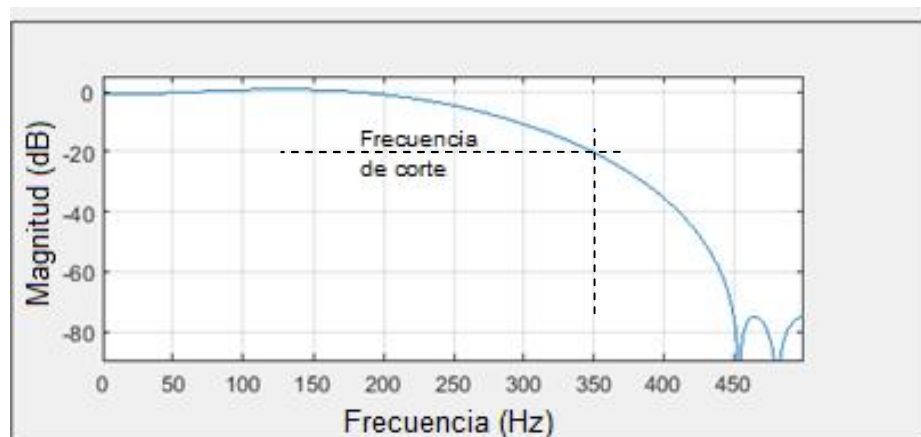
Se encuentra asociada a la información de descripción de eventos de naturaleza oscilatoria. La información no se encuentra contenida en muestras individuales sino se encuentra directamente relacionada a las muestras [16] [48].

Los filtros de respuesta en frecuencia más característicos son pasa altas, pasa bajas, pasa banda y rechaza banda.

### 2.5.6 Filtro pasa bajas

El filtro pasa bajas como su nombre lo indica permite el paso de frecuencias bajas y atenúa las frecuencias más altas donde estas tienden a cero.

**Figura 2-12:** Filtro pasa bajas

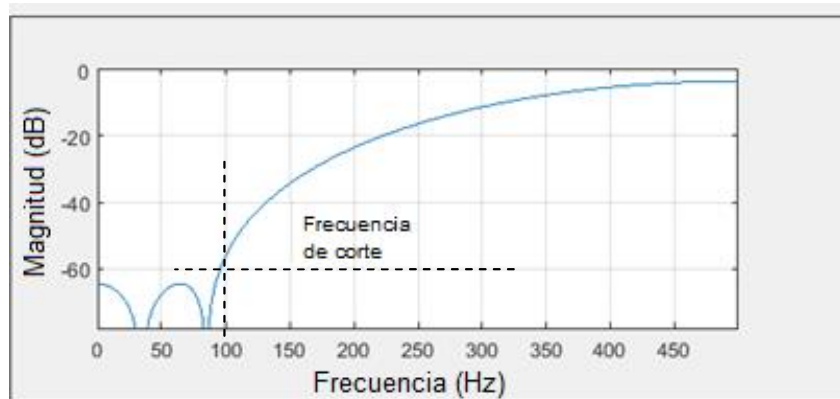


En la figura anterior se puede observar aquellas frecuencias menores a la frecuencia de corte, las cuales emiten una amplitud completa. En el caso contrario, al momento de llegar a la frecuencia de corte comienza a disminuir de forma exponencial [49] [48].

### 2.5.7 Filtro pasa altas

El filtro pasa altas como su nombre lo indica permite el paso de frecuencias altas y atenúa las frecuencias bajas.

**Figura 2-13:** Filtro pasa altas



En la figura anterior se puede observar aquellas frecuencias mayores a la frecuencia de corte, las cuales emiten una amplitud completa. En el caso contrario, cuentan con un crecimiento hasta llegar a la frecuencia de corte [49] [48].

### 2.5.8 Filtro media móvil

El origen de este filtro se encuentra en estadística y es un filtro del tipo de respuesta al impulso finito (FIR) donde la media móvil corresponde un análisis sobre un conjunto de datos generando series de promedios [49] [48]. De esta forma tenemos que los valores asociados a la media móvil son una lista de números donde cada uno corresponde al promedio de un subconjunto de los datos originales. Siendo  $M$  la media móvil (ver ecuación (2.6)).

$$y[n] = \frac{1}{M} \sum_{k=0}^{M-1} x[n-k] = \sum_{k=0}^{M-1} \frac{1}{M} x[n-k] \quad (2.6)$$

Un caso particular del filtro de media móvil es la implementación por recursión. Para este caso tenemos que la ecuación en recurrencia genérica (ver ecuación (2.7)).

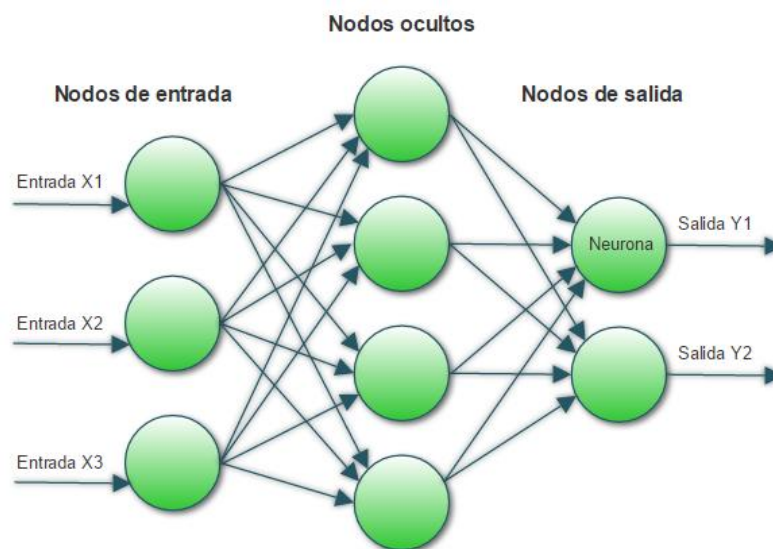
$$y[n] = y[n-1] + \frac{x[n+\frac{M-1}{2}] - x[n-\frac{M-1}{2}-1]}{M} \quad (2.7)$$

El filtro media móvil típicamente es útil en la reducción del ruido blanco donde este tipo de ruido se define como una señal aleatoria (proceso estocástico) que se caracteriza por el hecho que los valores en dos tiempos diferentes no guardan correlación estadística. Es decir, la densidad espectral de potencia (PSD) es una constante donde su gráfica es plana caracterizándose porque contiene todas las frecuencias y todas ellas muestran la misma potencia [49].

## 2.6 Redes neuronales

Desde un punto de vista biológico estructural se puede considerar al cerebro como una red de neuronas altamente conectadas a través de sinapsis o uniones intercelulares especializadas. Desde el punto de vista del aprendizaje de máquina las redes neuronales artificiales (ANN), tratan de imitar dicho comportamiento. El Dr. Robert Hecht-Nielsen definió una red neuronal como un sistema computacional conformado por un número de elementos de procesamiento simples altamente interconectados que procesan información de respuesta de estados dinámicos a partir de entradas externas [44].

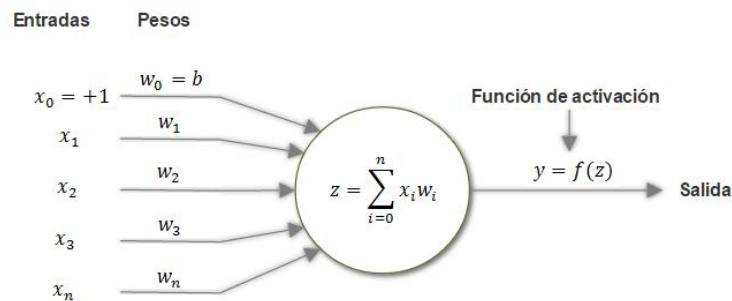
**Figura 2-14:** Estructura de una red neuronal



El objeto de las redes neuronales es generar un modelo a partir del entrenamiento con base en información conocida. Mediante el entrenamiento de una red neuronal se espera que las salidas de la red se aproximen a las salidas objetivo. Para garantizar el correcto funcionamiento de la red se incluye el ajuste de los pesos asociados a las entradas para cada neurona. Un sistema de redes neuronales típicamente está conformado por los datos asociados a cada nodo de entrada, la interconexión entre nodos, las funciones de reglas de activación de conjuntos de entradas particulares en salidas específicas y las funciones particulares de peso para la gestión de emparejamiento de entradas y salidas [43].

La red neuronal producto del proceso de entrenamiento realizado genera las diferentes salidas. Un modelo simple matemático para una red neuronal se puede definir como la función  $f: X \rightarrow Y$ . Matemáticamente una típica función de composición para redes neuronales está dada por la suma ponderada no lineal [44].

**Figura 2-15:** Estructura interna básica de una neurona artificial

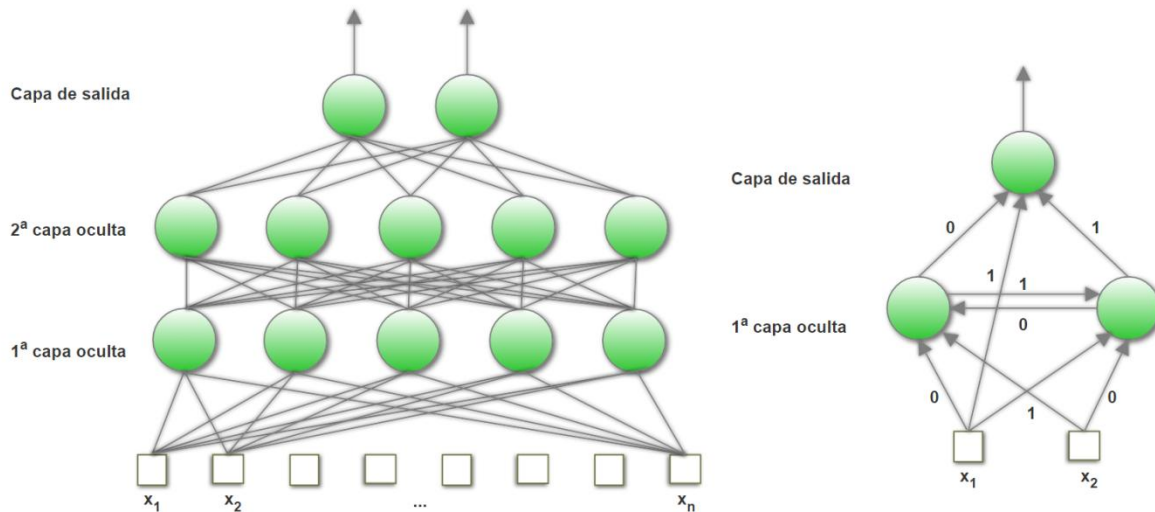


### 2.6.1 Topología

La topología en una red neuronal se encuentra directamente relacionada con la organización y distribución de las neuronas dentro de la red para formar agrupaciones de aprendizaje y funcionamiento general. La topología se caracteriza por el número de capas, el número de neuronas por capa, el grado de conectividad y tipo de conexión. Un ejemplo común de topología en aprendizaje no supervisado es el mapeo directo de entradas a una colección de unidades que representan categorías (Self-organized maps).

Las topologías más comunes en aprendizaje supervisado son totalmente conectadas, tres capas, feedforward y recurrentes. En general la topología involucra la modificación de pesos en las conexiones [36].

**Figura 2-16:** Ejemplos generales de topología de ANN - Feedforward network y recurrentes



## 2.6.2 Mecanismo de aprendizaje

El aprendizaje en una red neuronal hace referencia a la adaptación de los pesos y relaciones en la red para dar respuesta teniendo como base entradas específicas. Los cambios que se realizan son destrucción, creación y modificación de conexiones entre las neuronas. Acorde a las características del aprendizaje típicamente existen tres paradigmas los cuales exploramos a continuación [43].

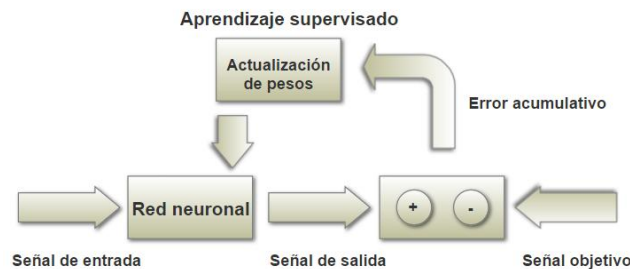
## 2.6.3 Aprendizaje supervisado

En este tipo de aprendizaje fuentes externas proveen un conjunto de entradas a la red. Estas entradas representan estímulos para las salidas con conocimiento a priori. Durante el proceso de ejecución los resultados de las salidas son continuamente comparados con los datos objetivo. Se realizan modificaciones sobre los parámetros para obtener una mejor aproximación a las salidas deseadas. El aprendizaje supervisado puede hacer referencia a los procesos de clasificación donde se cuenta con un amplio rango de

clasificadores. Los clasificadores típicos son perceptrones multicapa, máquinas de soporte vectorial, algoritmos del k-vecino más cercano, modelos Gaussianos, árboles de decisión y clasificadores de Bayes entre otros [33].

Para solucionar problemas dados se requiere considerar varios pasos en la ejecución de aprendizaje supervisado. El primer paso consiste en determinar los ejemplos de tipo de entrenamiento; en el segundo paso se reúne el conjunto de datos de entrenamiento que describen satisfactoriamente un problema dado: el tercer paso describe el conjunto de datos de entrenamiento en forma entendible para seleccionar el tipo de red a utilizar. Finalmente se ejecuta el entrenamiento y la evaluación de desempeño de aprendizaje de datos de prueba. El conjunto de datos de prueba no debe ser introducido en el entrenamiento de la red mientras se realiza el entrenamiento [44].

**Figura 2-17:** Diagrama de proceso de aprendizaje supervisado



## 2.6.4 Aprendizaje no supervisado

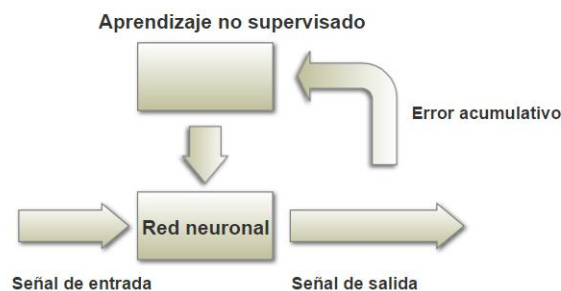
También es conocido como aprendizaje auto-organizado. Es una técnica de aprendizaje de máquina donde ninguna fuente externa alimenta la red y se basa en información local y control interno. De esta forma los datos de entrenamiento y los patrones de entrada se encuentran presentes en el sistema.

La organización del sistema de datos se encuentra determinada por categorías o conjuntos de similitud (Clusters). El conjunto de datos de entrada se encuentra presente a nivel de capa de entrada, los pesos de las conexiones de red son ajustados en el nivel orden de competencia hacia la capa de salida donde los candidatos objetivo corresponderán a los nodos con más alto valor. Es decir la red no recibe ninguna

información del entorno que le permita determinar si una determinada entrada es o no correcta [45].

En aprendizaje no supervisado se busca determinar cómo los datos son organizados y debe determinar automáticamente características, correlaciones y categorías a partir de los datos de entrada. Comúnmente se utiliza en aplicaciones del dominio de estimación como modelos estadísticos, filtrado, comprensión, separación de fuentes y búsqueda de similitudes [44].

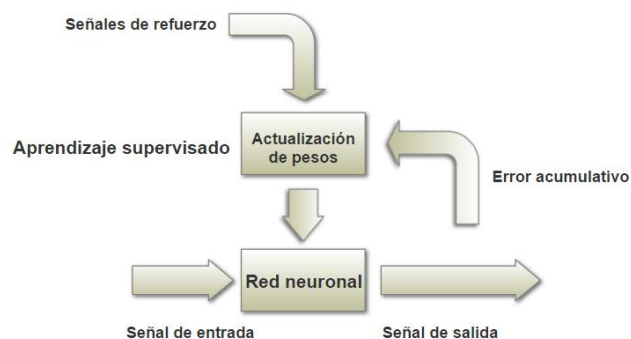
**Figura 2-18:** Diagrama de proceso de aprendizaje no supervisado



### 2.6.5 Aprendizaje por refuerzo

El aprendizaje por refuerzo se realiza cuando no es posible contar con ejemplos completos del problema a resolver. Es decir, no se indica durante el entrenamiento la salida exacta que se espera por parte de la red ante una determinada entrada.

**Figura 2-19:** Diagrama de proceso de aprendizaje por refuerzo



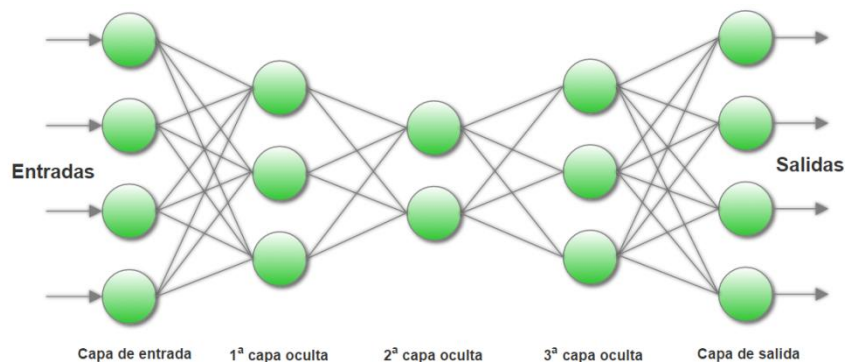


En este tipo de aprendizaje se cuenta con una función de supervisión que introduce señales de refuerzo si la salida obtenida por la red corresponde con la esperada. Así, se busca excitar o inhibir la actividad neuronal mediante el valor de refuerzo recibido un ejemplo es *éxito*=+1 o *fracaso*=-1. Los pesos son ajustados basándose en un mecanismo de probabilidades directamente vinculado con el valor de refuerzo recibido.

### 2.6.6 Redes neuronales feedforward

Las redes feedforward son redes neuronales simples donde la información se mueve únicamente en una dirección, hacia adelante. Es decir, parten de los nodos de entrada a través de los nodos ocultos y hacia los nodos de salida. No presentan ciclos en la red. Típicamente en un amplio número de unidades de procesamiento simple organizadas en capas. Cada unidad en una capa es conectada con las unidades en la capa anterior. Se resalta el hecho que las conexiones no son iguales ya que cuentan con diferentes pesos [23]. En operaciones normales de clasificación no existe retroalimentación entre las capas. Son ampliamente utilizadas en reconocimiento o clasificación de patrones.

**Figura 2-20:** Diagrama de proceso red Feedforward



### 2.6.7 Regla de clasificación bayesiana

La regla de clasificación óptima bayesiana define que dada una colección de muestras aleatorias de  $n$  poblaciones donde  $h_k$  es la probabilidad a priori que la muestra  $y_i$  pertenezca a la población  $k$ . Las clasificaciones erróneas en que una muestra no pertenezca a la población objetivo se denotan como  $l_k$ . La función de densidad de

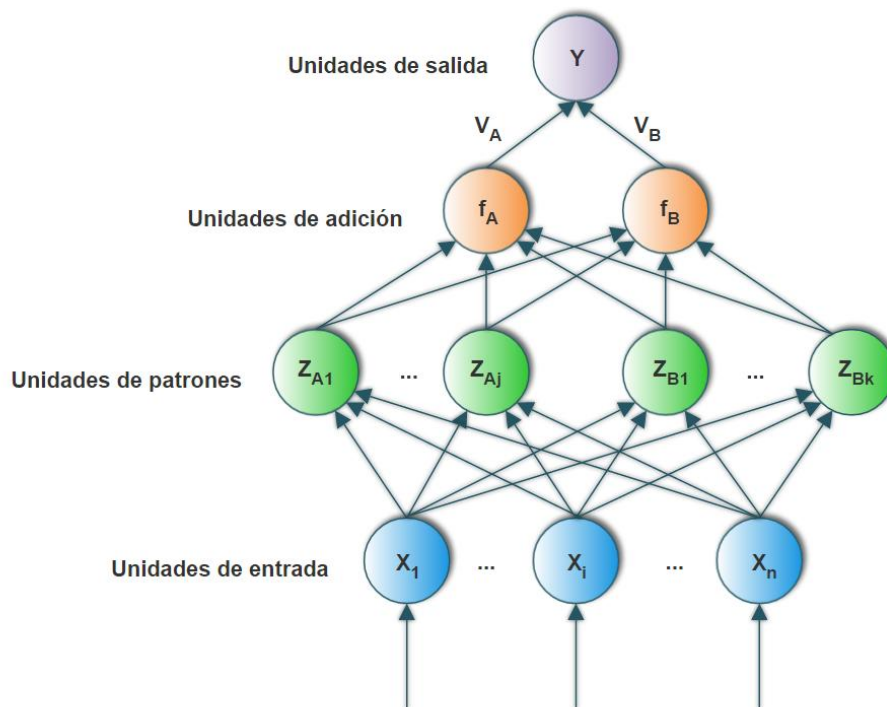
probabilidad  $f_k(y)$  determina la probabilidad condicional que una muestra específica pertenezca a la población  $k$ . Dada esta condición tenemos para dos poblaciones (ver ecuación (2.8)) [46]:

$$h_k l_k f_k(y) > h_i l_i f_i(y) \quad (2.8)$$

### 2.6.8 PNN – Redes neuronales probabilísticas

Las PNN, por sus siglas en inglés Probabilistic Neural Network, son redes del tipo feedforward con función de base radial [8] derivadas de las redes Bayesianas [31] y utilizan el algoritmo de análisis discriminante de núcleo.

**Figura 2-21:** Arquitectura de una PNN



La operación se encuentra organizada en una estructura de red de cuatro capas. Típicamente son utilizadas en problemas de clasificación cuando la entrada está presente. La primera capa calcula la distancia del vector de entrada para entrenar

múltiples vectores. Este proceso tiene como resultado un vector donde los elementos indican qué tan cerca se encuentran las entradas de los datos de entrada entrenados. La segunda capa suma la contribución de cada una de las clases y produce una red de salida como un vector de probabilidades. Finalmente una función de transferencia completa en las salidas de la segunda capa selecciona la probabilidad máxima y produce una identificación positiva para la clase objetivo así como identificación negativa para las clases no seleccionadas [1].

- Capa de entrada: en esta capa cada neurona representa una variable del predictor. En variables categóricas tenemos que  $N - 1$  neuronas son utilizadas, donde  $N$  es el número de categorías. Se estandariza el rango de valores por sustracción de la media y es dividida por el rango intercuartil. En este punto las neuronas de entrada alimentan cada una de las neuronas en la capa oculta. De esta forma tenemos una cantidad  $d$  de neuronas donde  $a$  corresponde a la dimensión de los datos de entrada [1].
- Capa de patrones: en esta capa se encuentra una neurona por cada caso en el conjunto de datos de entrenamiento. Almacena los valores de las variables del predictor para cada caso así como el valor objetivo. Una neurona oculta computa la distancia euclidiana del caso de prueba del punto central de la neurona y aplica una función de núcleo de base radial (RBF) utilizando los valores sigma [1].
- Capa de adición o sumatoria: en las PNN existe un patrón neuronal por cada categoría de las variables objetivo. La actual categoría objetivo de cada clase de entrenamiento es almacenada junto con cada neurona oculta. El peso del valor proveniente de la neurona oculta es alimentado solo si el valor del patrón corresponde con la categoría de la neurona oculta. El patrón neuronal adiciona valores por cada clase que representa [1].
- Capa de salida: en esta capa se comparan los pesos de los votos acumulados para cada categoría objetivo en la capa de patrones. Utiliza los votos más grandes para predecir la categoría objetivo [1].

Cuando a la red se presenta un vector para ser clasificado  $y = [y_1, \dots, y_d]$  la segunda capa realiza el cálculo de distancias desde los valores de entrada hasta cada uno de los vectores de ejemplo mediante la función  $\varphi_{ij}$  la cual trabaja acorde al núcleo normal estándar determinado por la función de probabilidad (ver ecuación (2.9)) [46]:

$$\varphi_{ij} = \frac{1}{(2\pi)^{\frac{d}{2}} \sigma^d} \exp \left[ -\frac{(y-x_{ij})^T (y-x_{ij})}{2\sigma^2} \right] \quad (2.9)$$

Donde  $\sigma$  es la propagación de la función Gaussiana o el parámetro de amplitud que puede tomar valores en un rango de 0 a 1. El mejor valor para la propagación es aquel que tenga los resultados de clasificación más precisos. En la capa de sumatoria el término  $P_i(y)$  indica la verosimilitud que el dato de entrada  $y$  pertenezca a la  $i$ -ésima clase. Del proceso de sumatoria tenemos (ver ecuación (2.10)):

$$P_i(y) = \frac{1}{N_i} \sum_{j=1}^{N_i} \varphi_{ij} \quad (2.10)$$

Acorde al proceso realizado en la capa de salida se asignará la clase con mayor probabilidad acorde a (ver ecuación (2.11)):

$$Clase(y) = \arg \max_i \{P_i(y)\} \quad (2.11)$$

Donde  $Clase(y)$  es la categoría a la que  $y$  pertenece [46].

## 2.7 Minería de datos

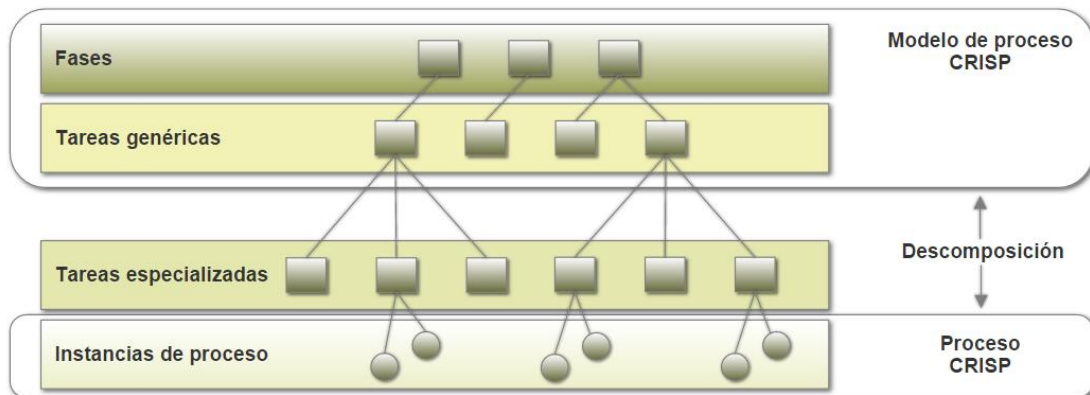
La Minería de Datos (DM) por las siglas en inglés Data Mining, es el proceso de extracción automática de conocimiento útil y comprensible que no es evidente pero que se encuentra implícitamente en los datos almacenados en distintos formatos [58]. Las herramientas de Minería de Datos involucran la participación de múltiples áreas como los sistemas de almacenamiento de datos, la estadística, el aprendizaje de máquina, la computación de alto desempeño y la visualización, donde estas interactúan con un objetivo común permitiendo la identificación de relaciones y patrones significativos presentes en los datos permitiendo la toma de decisiones [10] [38].

Sin embargo, la extracción de información derivada de los procesos de minería de datos va más allá y se encuentra enmarcado dentro de los diversos procesos no triviales de descubrir información novedosa, útil, potencialmente usable y entendible de una forma iterativa e interactiva [7] [38].

## 2.8 CRISP-DM

El proyecto se desarrolla utilizando la metodología CRISP-DM [9] (por sus siglas en inglés Cross Industry Standard Process for Data Mining). Dicha metodología surge de la necesidad de aprender nuevas técnicas para aplicar, comprender y manejar los resultados obtenidos por procesos de minería de datos de una manera jerárquica [22]. El desarrollo de la metodología tiene como base experiencias prácticas. Se resalta el hecho que gracias a su versatilidad permite interactuar sobre múltiples herramientas. El objeto final de la metodología es extraer conocimiento presente en los datos. El conjunto de tareas se describe en cuatro niveles de abstracción que van de lo general a lo particular: fase, tarea genérica, tarea especializada e instancia de procesos [9].

**Figura 2-22:** Estructura jerárquica de CRISP-DM [9]



En el nivel superior, el proceso es organizado en un número de fases donde cada una consiste en un conjunto de varias tareas genéricas de segundo nivel.

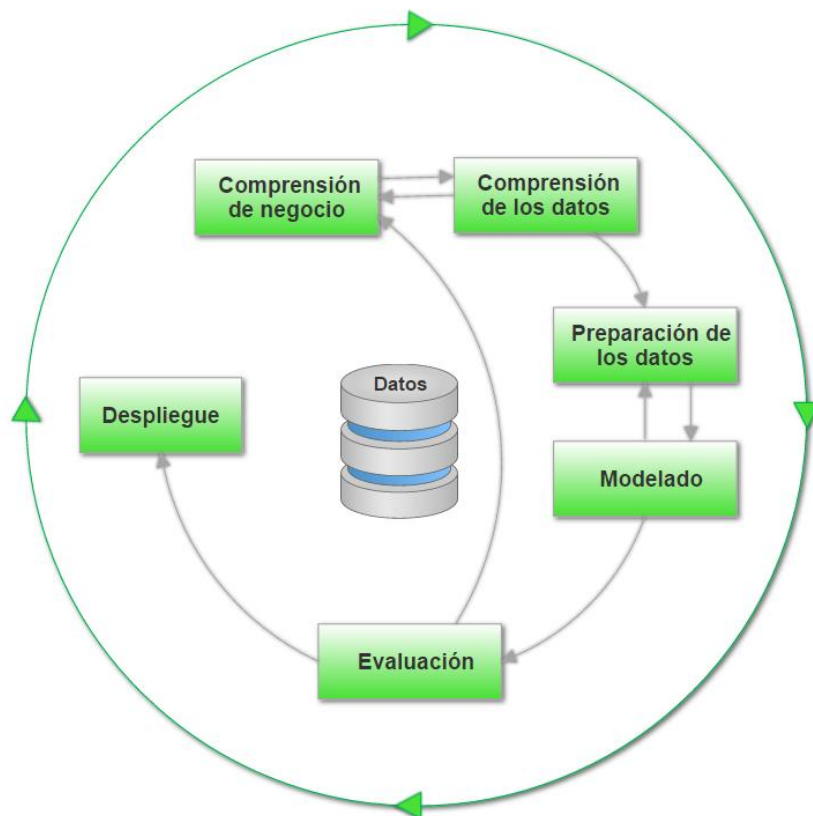
Este segundo nivel se caracteriza por ser lo bastante general para cubrir todas las situaciones posibles en los procesos de minería de datos. Las tareas genéricas están destinadas a ser tan completas y estables como sea posible. Dónde completo significa que cubre tanto al proceso entero de minería de datos como todas las posibles aplicaciones y estable significa que el modelo debería ser válido para el escenario esperado pero ser incluso extensible y adaptarse nuevos modelos [9].

El tercer nivel, el nivel de tarea especializado, se puede describir como las acciones de las tareas genéricas que deberían ser realizadas en ciertas situaciones específicas. Describe la manera en que una tarea se caracteriza en situaciones diferentes, como la limpieza de valores numéricos contra la limpieza de valores categóricos, o si el tipo de problema es agrupamiento o el modelado predictivo [9].

El cuarto nivel, la instancia de proceso, es un registro de las acciones, decisiones y los resultados de una minería de datos real aplicada.

El estándar incluye una metodología y una guía estructuradas en 6 fases [9]. La comunicación entre dichas fases permite realizar revisiones y nuevas iteraciones sobre fases anteriores dependiendo de los resultados obtenidos. Cada una de las fases es acorde a la metodología [9] y se desarrollará de una manera general ejecutando diferentes actividades asociadas a cada una de las fases respectivamente [22].

**Figura 2-23:** Ciclo de vida de la metodología CRISP-DM [9]



### **2.8.1 Comprensión del proyecto**

Esta fase inicial se enfoca en la comprensión del proyecto. Se desarrollan los objetivos y requerimientos desde una perspectiva no técnica. Posteriormente se convierte el conocimiento adquirido en una definición de problema desde la perspectiva de la minería de datos. En esta fase las principales tareas genéricas son [9]:

- Establecimiento de los objetivos del proyecto (Contexto inicial, objetivos de negocio, criterios de éxito)
- Evaluación de la situación (Inventario de recursos, requerimientos, supuestos, restricciones, riesgos y contingencias, terminologías propias del negocio, costos y beneficios)
- Establecimiento de los objetivos de la minería de datos (objetivos y criterios de éxito)
- Generación del plan del proyecto (plan, herramientas, equipo y técnicas)

### **2.8.2 Comprensión de los datos**

Esta fase de preparación permite familiarizarse con los datos teniendo presente los objetivos del negocio. Permite identificar problemas de calidad en los datos e incluso detectar subconjuntos de datos para definir hipótesis de información oculta. Algunas tareas genéricas de la fase corresponden a [9]:

- Recopilación inicial de datos
- Descripción de los datos
- Exploración de los datos
- Verificación de calidad de datos

### **2.8.3 Preparación de los datos**

Esta fase abarca todas las actividades para construir el conjunto final de datos. Las tareas de preparación de datos se realizan en múltiples ocasiones y sin un orden específico. Las principales tareas genéricas a realizar abarcan [9]:

- Selección de los datos
- Limpieza de datos
- Construcción de datos
- Integración de datos
- Normalizar / dar formato a los datos

### **2.8.4 Modelado**

En este punto múltiples técnicas son seleccionadas y aplicadas. Se realizan ejecuciones y calibraciones con el ánimo de obtener valores óptimos. Típicamente existen múltiples técnicas de minería de datos para dar solución al mismo tipo de problema. Usualmente en esta fase se requiere ir una fase atrás a la preparación de los datos. Dentro de las tareas genéricas particulares a realizar encontramos [9]:

- Selección de la técnica de modelado
- Diseño de la evaluación
- Construcción del modelo
- Evaluación del modelo

### **2.8.5 Evaluación**

En esta etapa se construye el modelo más representativo y de alta calidad desde una perspectiva de análisis de datos entre los modelos de las fases anteriores. El principal objetivo es determinar cuáles soluciones son útiles a las necesidades del problema. En este punto es de vital importancia identificar los aspectos más importantes del problema que no fueron considerados en etapas anteriores. Al final de esta fase se alcanzan los resultados de la implementación de minería de datos. Algunas de las tareas genéricas asociadas a la fase corresponden a [9]:

- Evaluación de resultados
- Revisión del proceso
- Determinar los siguientes pasos o acciones



### **2.8.6 Despliegue, resultados e informe final**

Generalmente la creación de un modelo no corresponde con el fin de un proyecto. El conocimiento adquirido debe ser organizado y presentado. Es de vital importancia explorar la utilidad de los modelos e integrarlos en las tareas de toma de decisiones. En esta etapa se pueden desarrollar tareas genéricas como [9]:

- Planificación de generación de prototipo
- Plan de monitoreo y mantenimiento
- Generación del informe final
- Revisión del proyecto



## 3. Datos y metodología

### 3.1 Comprensión de datos

Los múltiples registros disponibles son producto de muestreos en sitio desarrollados dentro del área de estudio en un periodo definido. Según el tipo de medida a recuperar se realizaron diversos procesos. En la tabla 3-1, se presenta la información general para los diferentes muestreos sobre los diversos tipos de registro utilizados en el proceso de generación de datos [25].

**Tabla 3-1:** Información general registros de pozo base

Dato	Tipo de dato	Fecha adquisición	Información general
Registro de imágenes	Registro de imágenes resistivas (ohm-m)	Periodo comprendido entre Julio de 2008 y Enero de 2009	Registro de 6 pozos con un total de 2362.2 m.
Registro de rayos gamma	Registro de radioactividad natural ( $^{\circ}$ API)		
Fotografías de núcleo	Fotografías en formato JPG.	Noviembre de 2007	Registro de 1 pozo para un total de 87.26 m.

Los registros se recuperaron en campañas de reevaluación de reservas del área definida en la cuenca del Catatumbo en el municipio de Tibú, Norte de Santander, Colombia. Para los diferentes registros posteriormente se presentarán los diversos parámetros estadísticos que describen los conjuntos de datos asociados a los registros de Rayos Gamma involucrados en el proceso y los datos de resistividad obtenidos por la herramienta de imágenes resistivas.

### 3.1.1 Registro de Rayos Gamma

Los datos de registro de Rayos Gamma se generaron con la herramienta *Compact Gamma Ray* y fueron recuperados por la compañía de servicios petroleros Weatherford Colombia [25]. La información recuperada se encuentra disponible para 6 pozos. Los registros en mención tienen como objeto realizar la medida para los diferentes niveles de radioactividad presentes en la roca sobre muestreos realizados en un intervalo definido de 0.61 metros. La medida resultante es expresada en unidades API (American Petroleum Institute). En la tabla 3.2 se presenta la información general estadística asociada al muestreo realizado por pozo (Valores redondeados en decimales [25]).

**Tabla 3-2:** Información general estadística del muestreo de rayos gamma

	Pozo 6	Pozo 7	Pozo 8	Pozo 9	Pozo 10	Pozo 11
<b>Media</b>	129.5	134.5	127.0	117.9	122.3	128.4
<b>Error típico</b>	2.05	2.02	1.90	1.92	1.88	1.82
<b>Mediana</b>	148.5	153.0	140.3	139.3	136.3	143.2
<b>Moda</b>	161.81	167.75	135.10	N/A	147.71	145.25
<b>Desviación estándar</b>	47.63	47.10	45.79	48.28	47.20	42.34
<b>Varianza</b>	2269.10	2218.63	2096.52	2331.28	2228.09	1793.13
<b>Curtosis</b>	-0.28	0.53	1.05	-0.73	-0.12	0.06
<b>Coefficiente de asimetría</b>	-0.95	-0.90	-0.62	-0.70	-0.54	-0.98
<b>Número de muestras</b>	539	542	579	633	632	539
<b>Valor mínimo</b>	23.19	23.21	4.98	22.71	18.07	23.4
<b>Valor máximo</b>	230.61	296.63	340.47	236.11	312.15	232.49

### 3.1.2 Registro de imágenes resistivas

Los datos de registro de imágenes resistivas se generaron con la herramienta compacta de imágenes resistivas CMI (*Compact Micro Imager*). Fueron realizados por la compañía de servicios petroleros Weatherford Colombia [25]. La información de registro se encuentra disponible para 6 pozos. El objetivo del registro de imágenes resistivas es identificar los valores de resistividad en las paredes del pozo. El proceso se realizó sobre muestreos en sitio donde los datos de resistividad provienen de lecturas en los patines de la herramienta la cual realiza muestreos con un intervalo definido de 0.005 metros con profundidad determinada menor a 0.07 metros hacia la pared del pozo. En la tabla 3.3 se presenta la información general estadística del muestreo realizado por pozo (Valores redondeados en decimales [25]).

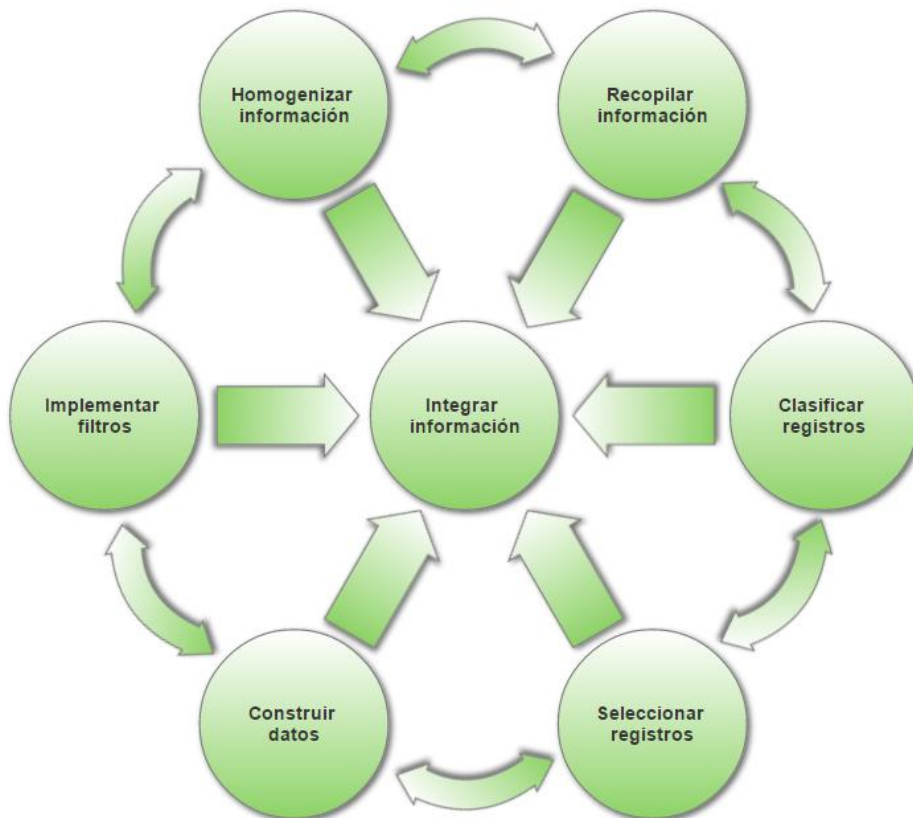
**Tabla 3-3:** Información general estadística del muestreo de imágenes resistivas

	Pozo 6	Pozo 7	Pozo 8	Pozo 9	Pozo 10	Pozo 11
<b>Media</b>	392.4	583.7	338.9	471.1	277.9	227.1
<b>Error típico</b>	68.60	119.56	48.94	93.86	36.93	33.59
<b>Mediana</b>	12.8	10.8	16.6	27.3	23.1	8.4
<b>Moda</b>	6.43	3.76	3.94	4.72	3.75	3.24
<b>Desviación estándar</b>	1592.77	2783.60	1177.52	2361.54	928.38	779.85
<b>Varianza</b>	2536927	7748423	1386555	55766868	861890	608164
<b>Curtosis</b>	91.32	113.23	100.92	101.75	114.70	62.91
<b>Coefficiente de asimetría</b>	8.71	9.76	8.38	9.45	9.17	7.06
<b>Número de muestras</b>	539	542	579	633	632	539
<b>Valor mínimo</b>	2.27	0.00	2.25	2.17	2.07	1.85
<b>Valor máximo</b>	21279.28	40662.24	18179.10	32400.00	14692.45	8852.92

## 3.2 Preparación de datos

La información base colectada presenta múltiples conjuntos de datos donde los registros asociados corresponden a diferentes fuentes de información. Es decir, existen registros con los mismos objetivos de medición pero sobre herramientas configuradas de manera particular para cada pozo. Dada esta característica, se presentan diferentes escalas de precisión, así como valores no representativos y algunos tipos de registros particulares por pozo. Dentro de la fase de preparación de información para ser procesada se realizaron diversas tareas las cuales, acorde a la metodología, se desarrollaron sobre múltiples ejecuciones y de una forma no lineal. Es decir, dependiendo del objeto de la iteración sobre el ciclo de vida se realizaron ejecuciones particulares, sin orden específico, para las diversas tareas definidas. A continuación identificamos las principales tareas asociadas a la fase así como el respectivo resultado final obtenido a través de las diversas iteraciones.

**Figura 3-1:** Diagrama de actividades fase de preparación de datos



### 3.2.1 Recopilar información

Como primera instancia se realiza la consolidación del conjunto base donde se incluye toda la información inicial disponible y donde esta corresponde a registros de pozo recuperados en sitio. Se encuentra de manera general para 6 pozos en formato de archivo tipo \*.las. Adicionalmente se cuenta con las respectivas imágenes resistivas para cada uno de los pozos impactados. En la figura 3-2 se presenta la información de pozos contra registros disponibles para cada uno de ellos. Se identifica por pozo un color donde la presencia del mismo indica que existe la respectiva información del tipo particular de registro.

**Figura 3-2:** Información disponible por pozo para el caso de estudio

	P6	P7	P8	P9	P10	P11	Descripción
INDEX	■	■	■	■	■	■	logged depth interval
CGXT	■	■	■	■	■	■	temperature
CLDC	■	■	■	■	■	■	density caliper
DCOR	■	■	■	■	■	■	density corrected
DDLB	■	■	■	■	■	■	corrected deep laterolog
DEN	■	■	■	■	■	■	compensated density
DSL	■	■	■	■	■	■	corr. shallow laterolog
GROH	■	■	■	■	■	■	gamma ray open hole
GCCR	■	■	■	■	■	■	sgs corrected gamma ray
GRPO	■	■	■	■	■	■	gr potassium
GRSG	■	■	■	■	■	■	sgs gamma ray
GRTH	■	■	■	■	■	■	gr thorium
GRUR	■	■	■	■	■	■	gr uranium
MBTC	■	■	■	■	■	■	mmr caliper
MINV	■	■	■	■	■	■	mmr microlog inverse
MNRL	■	■	■	■	■	■	mmr microlog normal
NPRL	■	■	■	■	■	■	limestone neutron porosity
PDPE	■	■	■	■	■	■	photoelectric factor
SMTU	■	■	■	■	■	■	dst uphole tension
SPDL	■	■	■	■	■	■	spontaneous potential

Posterior a la consolidación general, se realiza la selección de información homogénea por medio del cruce de pozos con los respectivos registros asociados para cada uno de ellos. Producto de este proceso se descartó la información asociada al pozo 11 ya que no cuenta con la información general para los registros de rayos gamma particulares para

torio, potasio y uranio. Así mismo, se descartó la información asociada a los registros generales relacionados con BIT, BS, DEN, GCGR, GRSG, HVOL, MBTC y SPGC.

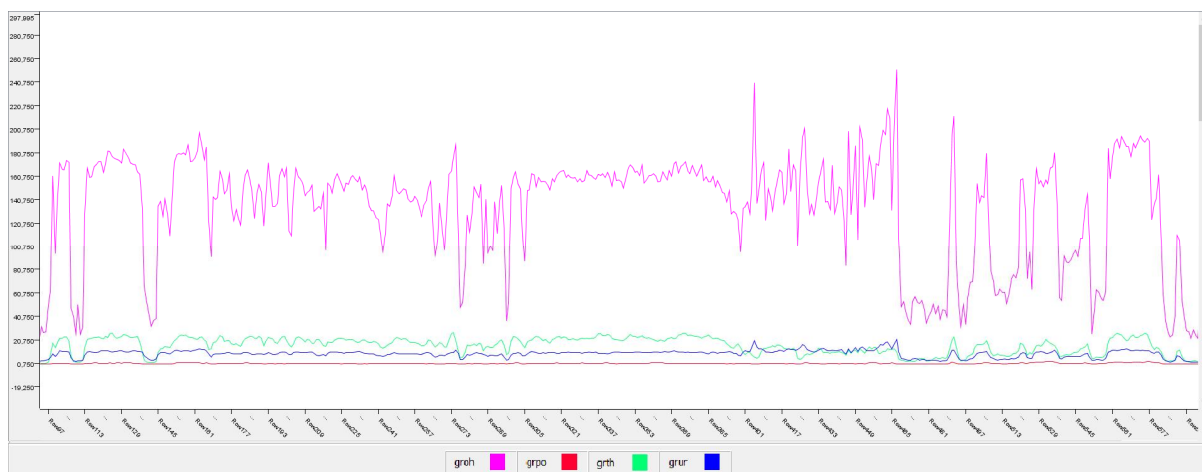
La selección resultante corresponde a la base del modelo y es presentada en la figura 3.1 resaltada en negrilla. Acorde a esta selección sobre los datos disponibles obtuvimos como resultado la base de información del caso de estudio. El conjunto de pozos resultantes abarca del 6 al 10. Así mismo, los registros base están comprendidos por CGXT, CLDC, DCOR, DDLB, DSLB, GROH, GRPO, GRTH, GRUR, MINV, MNRL, NPRL, PDPE, SMTU y SPDL.

### 3.2.2 Datos de registros radioactivos

Teniendo como base los registros que proporcionan las medidas de radioactividad natural en el caso de estudio se identifica la información base relacionada con los muestreos de:

- GROH: registro de rayos gamma (Gamma Ray Open Hole).
- GRPO: registro de rayos gamma para el Potasio (GR POTASSIUM).
- GRTH: registro de rayos gamma para el Torio (GR THORIUM).
- GRUR: registro de rayos gamma para el Uranio (GR URANIUM).

**Figura 3-3:** Ejemplo de registros gamma ray y gamma ray espectral caso de estudio - pozo 10.



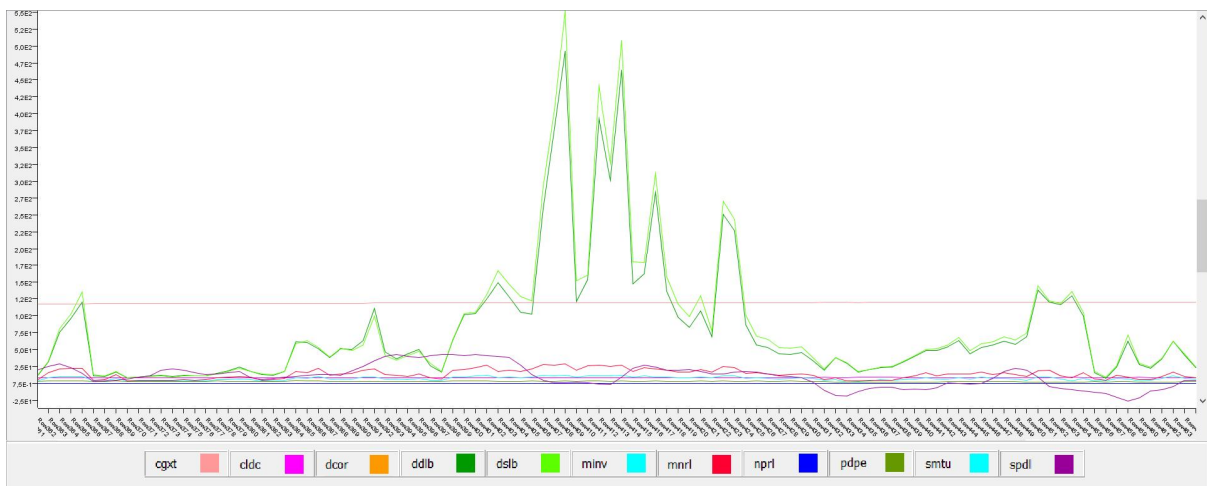


### 3.2.3 Datos de registros generales

Dentro de la información base existen registros clasificados como generales dada su diversidad. Cada muestreo realizado se encuentra asociado a un tipo de medida particular. Dentro de la selección realizada resaltamos los registros asociados a porosidad, latero-perfil y densidad. Como resultado de la clasificación realizada encontramos el conjunto compuesto por los registros:

- CGXT: temperatura.
- CLDC: calibrador del registro de densidad.
- DCOR: densidad corregida.
- DDLB: latero-perfil de resistividad profunda.
- DSLB: latero-perfil de resistividad somera.
- MINV: mmr micro-log inverso.
- MNRL: mmr micro-log normal.
- NPRL: porosidad neutrón en matriz caliza.
- PDPE: factor fotoeléctrico.
- SMTU: distancia de tensión en la boca del pozo.
- SPDL: potencial espontaneo.

**Figura 3-4:** Ejemplo de registros generales disponibles para el caso de estudio pozo 6



### 3.2.4 Construcción del conjunto de datos

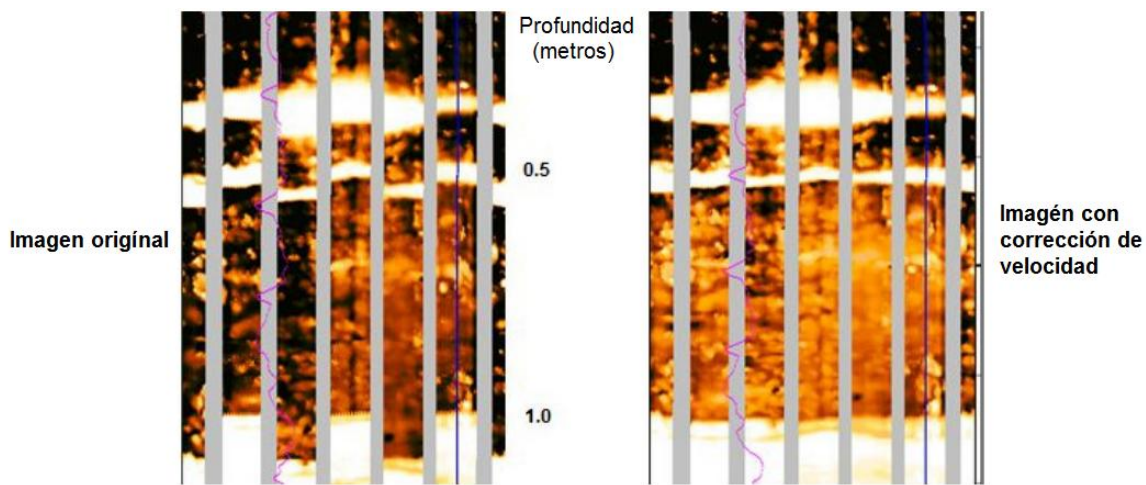
Durante el proceso de construcción del conjunto de datos se tuvieron en cuenta diferentes herramientas que permiten generar información adicional a partir de los datos base. Dentro de los elementos impactados encontramos el registro de dimensión fractal generado a partir del procesamiento de imágenes resistivas, y los respectivos filtros digitales aplicados para cada una de los registros geofísicos disponibles para el caso de estudio. Los filtros implementados corresponden a media móvil, pasa altas y pasa bajas.

Así mismo dentro del modelo de predicción de registros de dimensión fractal se contempla una sección donde se utilizan ventanas para los múltiples registros seleccionados. Las ventanas generan un desplazamiento de la señal original y permiten el recuperar información relevante de la misma señal por valores de proximidad sobre el muestreo de un punto particular y relacionándolo con los valores adyacentes definidos.

### 3.2.5 Procesamiento de imágenes resistivas

Para realizar el procesamiento de imágenes resistivas recuperadas en sitio se utilizó la aplicación *Wellimager*. Esta herramienta permite procesar e interpretar diferentes tipos de datos obtenidos de pozos. Típicamente la información predominante se encuentra relacionada con la geología, la petrofísica y la ingeniería de reservorio.

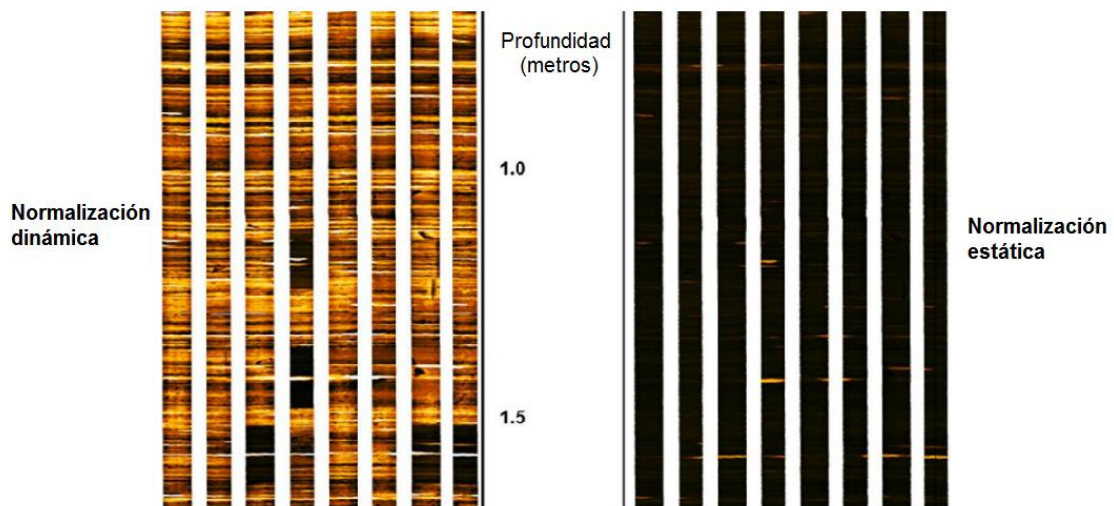
**Figura 3-5:** Ejemplo de corrección por velocidad en una imagen resistiva



El procesamiento de registros de imágenes resistivas tradicionalmente se realiza siguiendo una secuencia de proceso estándar. En la primera etapa se realizan correcciones de velocidad y aceleración asociados a la herramienta en el momento de realizar el respectivo muestreo [25]. En la figura 3-5 se presenta un ejemplo de corrección de velocidad para un intervalo definido entre 0.5 metros y 1 metro. Adicionalmente, dentro del muestreo realizado la orientación de los patines y los movimientos de la herramienta a lo largo del eje del pozo se controlan por un arreglo de magnetómetros e inclinómetros triaxiales. La información proporcionada por este monitoreo permite determinar la posición precisa de todos los elementos de la herramienta en el espacio al realizar la toma de muestras. Por lo anterior, es posible caracterizar los rasgos geológicos asociados [25].

Posteriormente en el proceso se realiza una normalización dinámica con un periodo definido de 0.305 metros. En este punto para la resistividad se define un intervalo entre 0.305 metros y 0.610 metros. El valor mínimo de resistividades asociado al color negro y el valor máximo al color blanco respectivamente. Los valores intermedios varían en una escala de 64 colores. El resultado permite resaltar pequeños detalles. Adicionalmente, se realiza una normalización estática con un periodo definido desde el tope hasta la base del registro. El producto de este proceso es resaltar grandes rasgos geológicos. Las normalizaciones realizadas tienen como objeto optimizar el funcionamiento de la herramienta bajo condiciones de resistividad y contrastes variables [25].

**Figura 3-6:** Ejemplo de normalización estática y dinámica



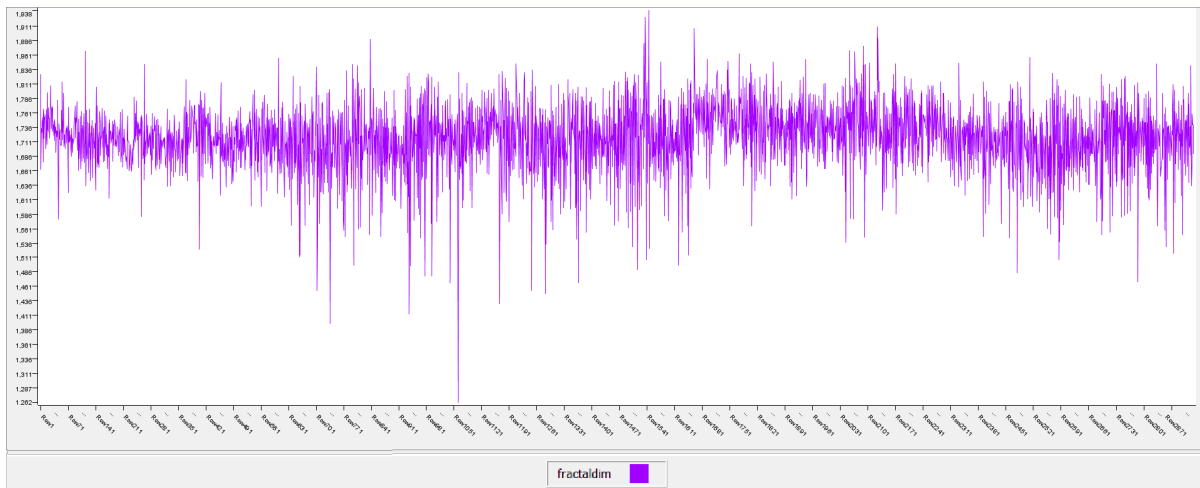
En la lectura de imágenes resistivas se utiliza un código de colores. Los tonos claros representan resistividades más altas y tonos más oscuros indican menores niveles de resistividad respectivamente. Las resistividades bajas pueden ser ocasionadas por minerales, ciertos tipos de arcillas y piritita entre otros. También pueden provenir por la presencia de filtrado de lodo no resistivo en fracturas, cavidades de disolución o en cualquier tipo de espacio poroso teniendo como resultado el oscurecimiento de la imagen [25].

El registro de resistividad promedio (SRES) se genera partiendo de la media aritmética de 8 patines para cada profundidad de registro especificada.

### 3.2.6 Datos de dimensión fractal

Los datos de dimensión fractal se calcularon a partir del registro de imágenes resistivas adquiridas en sitio. En la figura 3-7 visualizamos el comportamiento de la dimensión fractal para los diversos pozos del caso de estudio.

**Figura 3-7:** Ejemplo de registro de dimensión fractal para el caso de estudio - pozos 6 al 10

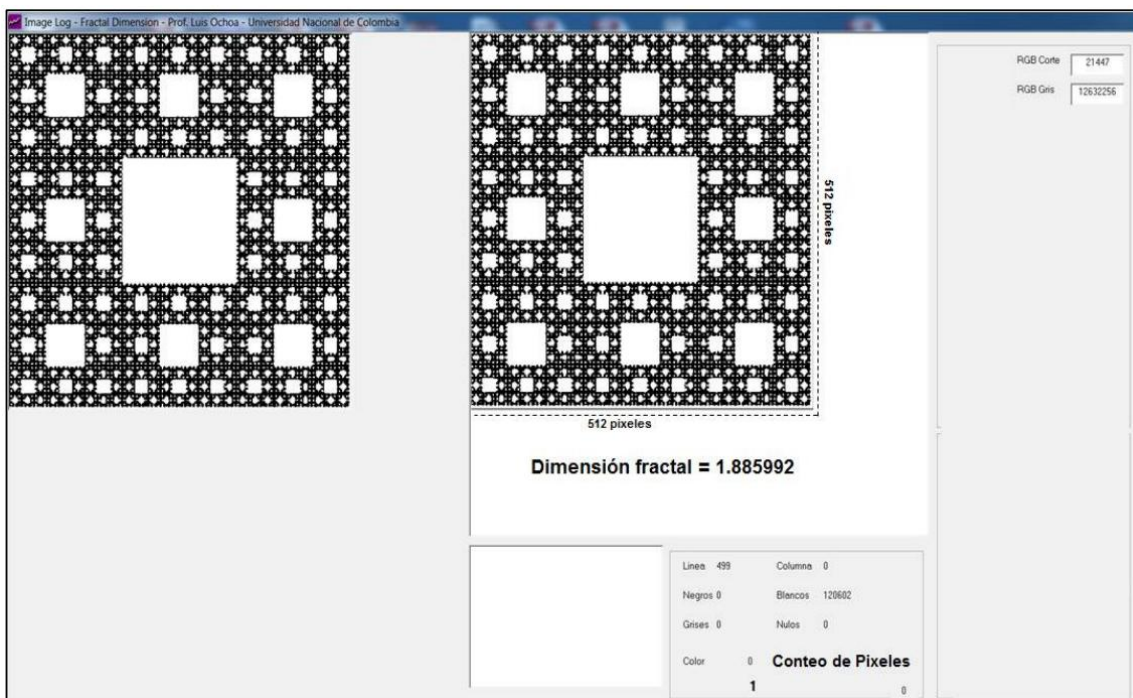


El registro de dimensión fractal se obtuvo mediante el procesamiento realizado utilizando la aplicación *Imagen Fractal* desarrollada por el profesor Luis Hernán Ochoa. Las

entradas de la aplicación corresponden a registros de imágenes resistivas en formatos de mapa de bits (BMP). Las dimensiones asociadas a los registros ingresados corresponden a 512 píxeles por 15378 píxeles. La aplicación genera un registro sobre muestreos definidos por el usuario. Las imágenes ingresadas en el sistema cuentan con una escala vertical de 1:5 donde la conversión de píxeles a profundidad es de 512 píxeles, es decir 0.663 metros [25].

El proceso realizado por la aplicación *Imagen Fractal* corresponde a la transformación de las imágenes resistivas. Se realiza una normalización binaria (*blanco y negro*), sobre las imágenes dinámicas, por medio de la aplicación de un filtro RGB (*Red, Green and Blue*). El punto de corte definido para el filtro corresponde al valor de  $RGB=21447$ , es decir, todos los valores RGB que estén por encima de este color serán transformados en negro y todos los valores RGB que se encuentren por debajo de este color serán transformados en blanco. El color gris representa a las franjas sin datos, este color gris es constante lo largo de todo el registro y corresponde al valor de  $RGB=12632256$ , dentro del procesamiento el valor en mención no es considerado para que el conjunto objetivo corresponda estrictamente a los píxeles blancos y negros [25].

**Figura 3-8:** Ejemplo conteo de píxeles alfombra Sierpinski (Aplicación *Imagen Fractal*)

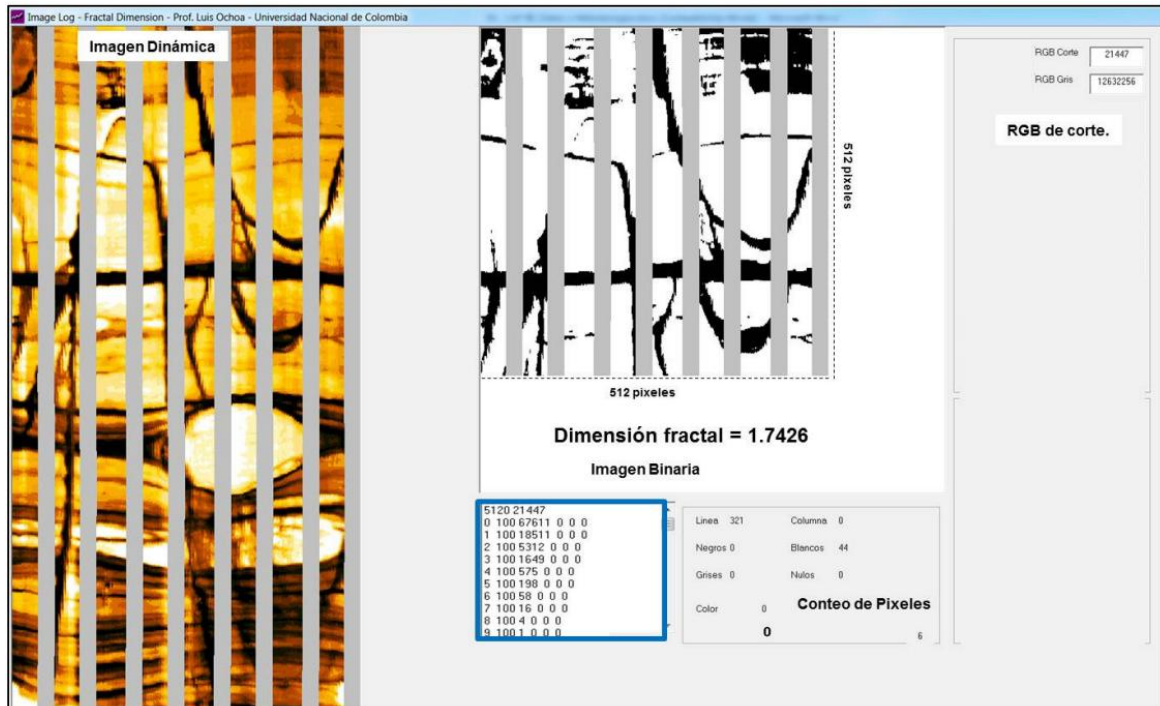




La figura 3-8, presenta como ejemplo un fractal tradicional como lo es la alfombra de Sierpinski. Para el caso particular la dimensión fractal es calculada por medio de la aplicación *Imagen Fractal*. El valor resultante del proceso por conteo de cajas anteriormente mencionado es  $m = 1.88599$ . Para la alfombra de Sierpinski el valor teórico de la dimensión fractal corresponde a  $1.89278$ . Para este caso de validación particular se puede evidenciar una certidumbre en el resultado de dimensión fractal obtenido del 99.641%.

En la figura 3-9 se presenta un ejemplo de registro de imágenes resistivas en un intervalo con fracturas naturales conductivas. Del lado izquierdo de la figura se observa la imagen de normalización dinámica. En el centro se observa la imagen binaria sobre un área definida de 512 píxeles por 512 píxeles. A la imagen binaria se aplica el método de conteo de cajas el cual arroja un valor de dimensión fractal de  $1.7426$ . A la derecha se observan los valores de corte de RGB [25].

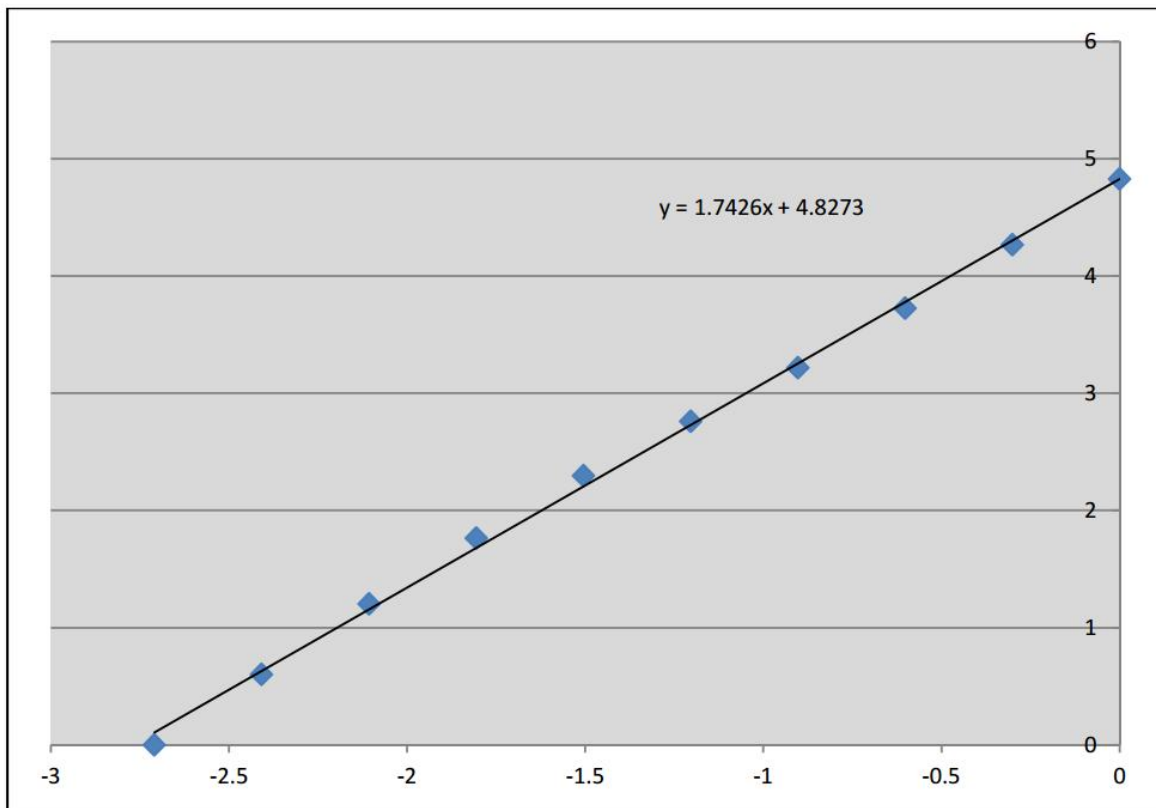
**Figura 3-9:** Ejemplo conteo de pixeles registro de imagen (Aplicación *Imagen Fractal*)



El recuadro azul de la Figura 3-9 muestra el número de datos utilizados para generar la regresión lineal. El intervalo definido de 512 píxeles corresponde a 0.063 metros. La pendiente  $m$  asociada a la regresión lineal representa la dimensión fractal resultante (ver ecuación (2.2)).

La figura 3-10, muestra la regresión lineal para el intervalo de imagen procesado en el ejemplo presentado en la figura 3-9. Donde en el eje  $x$  representa el logaritmo del inverso del tamaño de la caja que contiene píxeles negros. Respectivamente el eje  $y$  representa el logaritmo del número de cajas que contiene los píxeles impactados [25].

**Figura 3-10:** Ejemplo de regresión lineal para determinar la dimensión fractal



Resaltamos que para el ejemplo presentado la dimensión fractal para el intervalo impactado corresponde a  $1.74266$ .

El procesamiento objeto del análisis anteriormente mencionado se realizó sobre cada uno de los 6 pozos del área que contaban con la información de registros de dimensión fractal.

Como producto resultante del proceso realizado sobre la aplicación *Imagen Fractal* se generaron los registros de dimensión fractal para cada uno de los pozos. En la tabla 3-4 se presenta la descripción estadística general para los datos de dimensión fractal generados con el método de conteo de cajas. Se resalta que el valor máximo obtenido es de 1.94 en el pozo 8 y el valor mínimo es de 1.26 en el pozo 7. La media aritmética se encuentra entre 1.70 y 1.74 para todos los pozos.

**Tabla 3-4:** Información general estadística del muestreo de dimensión fractal

	Pozo 6	Pozo 7	Pozo 8	Pozo 9	Pozo 10
<b>Media</b>	1.7	1.7	1.7	1.7	1.7
<b>Error típico</b>	0.002	0.003	0.003	0.002	0.003
<b>Mediana</b>	1.7	1.7	1.7	1.8	1.7
<b>Moda</b>	1.71	1.72	1.70	1.78	N/A
<b>Desviación estándar</b>	0.04	0.07	0.07	0.05	0.06
<b>Varianza</b>	0.001	0.005	0.005	0.003	0.003
<b>Curtosis</b>	1.95	4.92	1.82	0.91	1.01
<b>Coficiente de asimetría</b>	-0.22	-1.35	-0.72	-0.31	-0.67
<b>Número de muestras</b>	523	526	563	617	616
<b>Valor mínimo</b>	1.527	1.262	1.432	1.539	1.470
<b>Valor máximo</b>	1.868	1.889	1.938	1.910	1.857

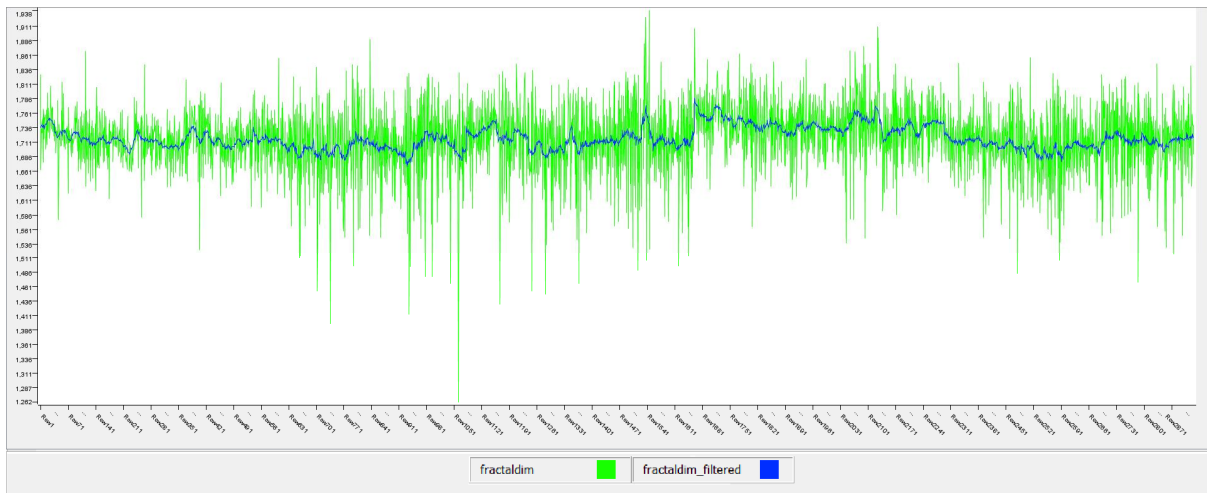


### 3.2.7 Datos de filtro media móvil

Los registros iniciales presentaron diferentes niveles de ruido. El filtro media móvil se caracteriza por permitir la atenuación del ruido blanco. Dentro del caso de estudio se implementó el filtro media móvil por el método por recursión con un orden de  $M=9$ . Es decir, los 9 valores inmediatamente disponibles tanto a derecha como a izquierda.

Dentro del proceso de construcción de datos y dado que para los valores iniciales y finales donde el conjunto de valores requeridos para el promedio no existe, se generó el valor de filtro sobre los valores disponibles. Es decir, el valor media móvil ideal se encuentra entre el intervalo comprendido por los registros  $n+9$  y  $n-9$  en cada pozo. En los casos fuera del intervalo donde por ejemplo se disponía de 8 valores a izquierda y 9 a derecha, el valor resultante se generó con los datos disponibles sobre la cantidad de registros disponibles. De esta forma fue posible generar información para todo el muestreo sin omitir registros. Sin embargo, se resalta el hecho que en los puntos asociados al origen y fin de cada pozo donde la información disponible no corresponde al intervalo típico la atenuación respectiva es menor a la esperada.

**Figura 3-11:** Ejemplo: filtro media móvil en dimensión fractal - pozos 6 al 10

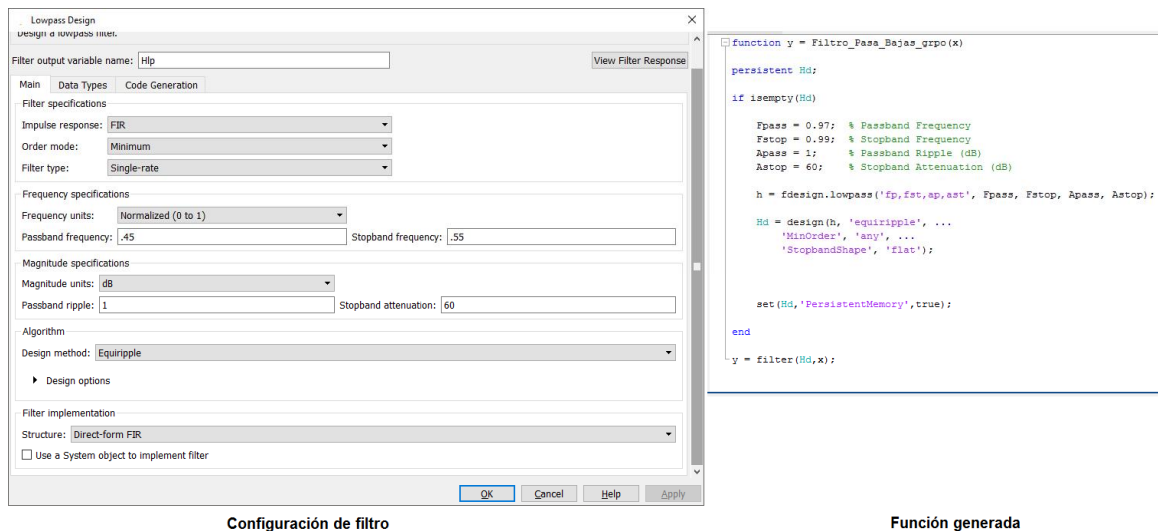


### 3.2.8 Datos de filtro pasa bajas

Las señales asociadas a los orígenes de datos son susceptibles a ruido proveniente de los equipos o factores externos que inciden en el muestreo por lo anterior se utiliza una herramienta para construir filtros. Esta permite configurar las diferentes variables requeridas para la creación del filtro. Los registros resultantes obedecen al proceso de convertir la señal del dominio del tiempo al dominio de la frecuencia, aplicar el filtro y devolver la señal al dominio del tiempo. Las señales resultantes tienen un comportamiento sinusoidal y presentan los respectivos valores de compensación producto del proceso realizado.

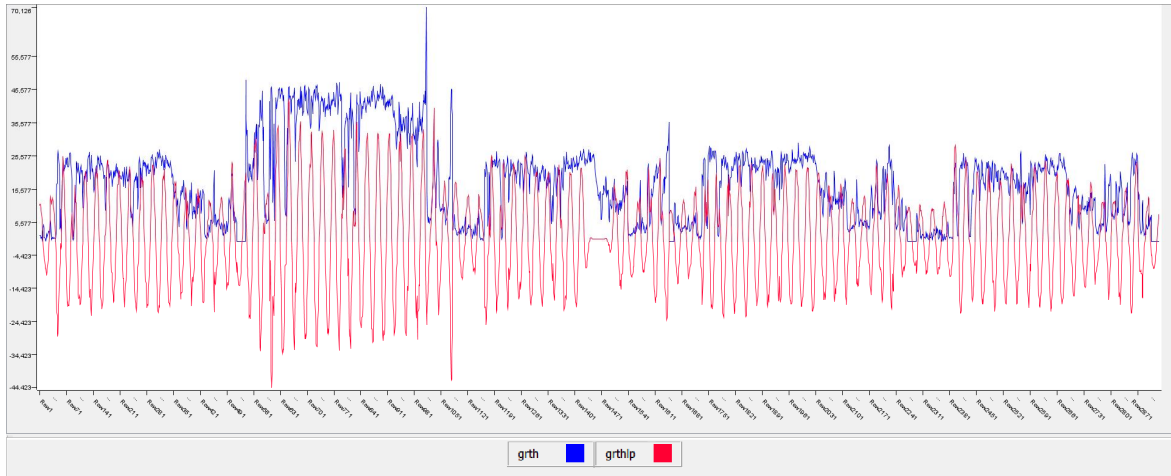
El objeto de aplicar el filtro pasa bajas es definir una frecuencia de corte que permita el paso de frecuencias bajas y atenuar el paso de frecuencias más altas, donde estas tienden a cero.

**Figura 3-12:** Ejemplo: construcción de filtro pasa bajas



Para todos los registros disponibles se generó la respectiva señal resultante del proceso de filtro de pasa bajas. La información generada fue integrada en el conjunto de datos base como registros de señales adicionales complementarias. En la figura 3.13 observamos el ejemplo de una señal original en azul con el respectivo registro resultante del proceso del filtro aplicado en rojo.

**Figura 3-13:** Ejemplo: señal original y señal afectada por filtro pasa bajas



### 3.2.9 Datos de filtro pasa altas

Bajo el mismo proceso de creación de filtros se generaron las respectivas funciones para todas las señales disponibles para el filtro pasa altas. El objeto de aplicar este filtro es definir una frecuencia de corte que permita el paso de frecuencias altas y compense el paso de frecuencias más bajas.

**Figura 3-14:** Ejemplo: construcción de filtro pasa altas

**Configuración de filtro**

```

if isempty(Hd)
    Fstop = 0.01; % Stopband Frequency
    Fpass = 0.03; % Passband Frequency
    Astop = 60; % Stopband Attenuation (dB)
    Apass = 1; % Passband Ripple (dB)

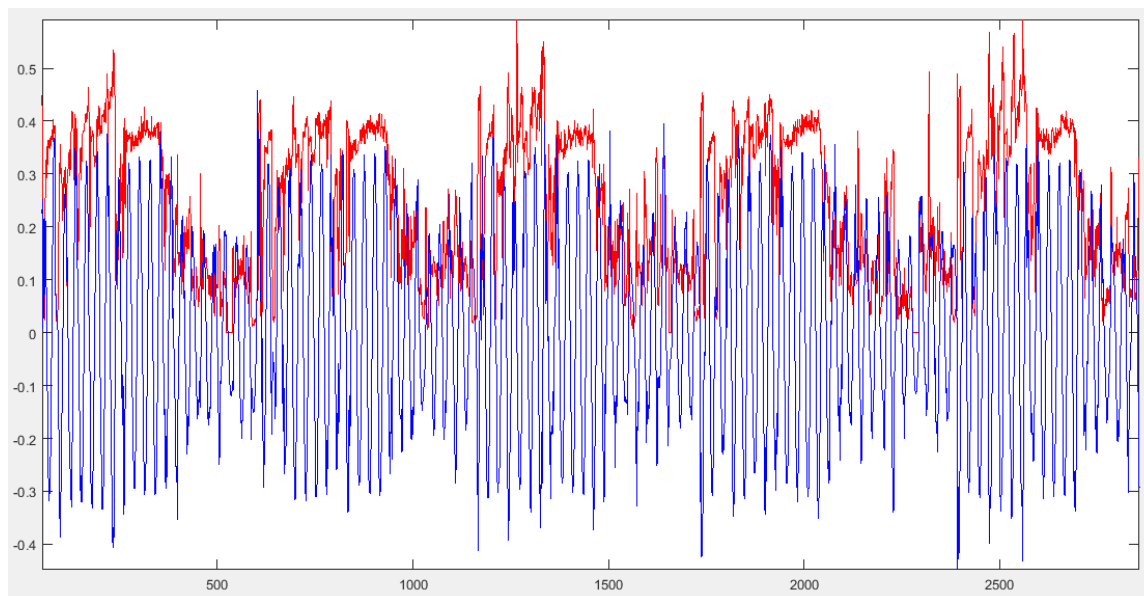
    h = fdesign.highpass('fst,fp,ast,ap', Fstop, Fpass, Astop, Apass);
    Hd = design(h, 'equiripple', ...
        'MinOrder', 'any', ...
        'StopbandShape', 'flat');

    set(Hd, 'PersistentMemory', true);
end
                    
```

**Función generada**

Para todos los registros disponibles se generaron las respectivas señales resultantes del proceso de aplicar el filtro de pasa altas. La información generada se integró en el conjunto de datos base como señales adicionales complementarias. En la figura 3.15 observamos el ejemplo de una señal original en rojo con el respectivo registro resultante del proceso del filtro aplicado en azul.

**Figura 3-15:** Ejemplo: señal original y señal afectada por filtro pasa altas



### 3.2.10 Homogeneización de información

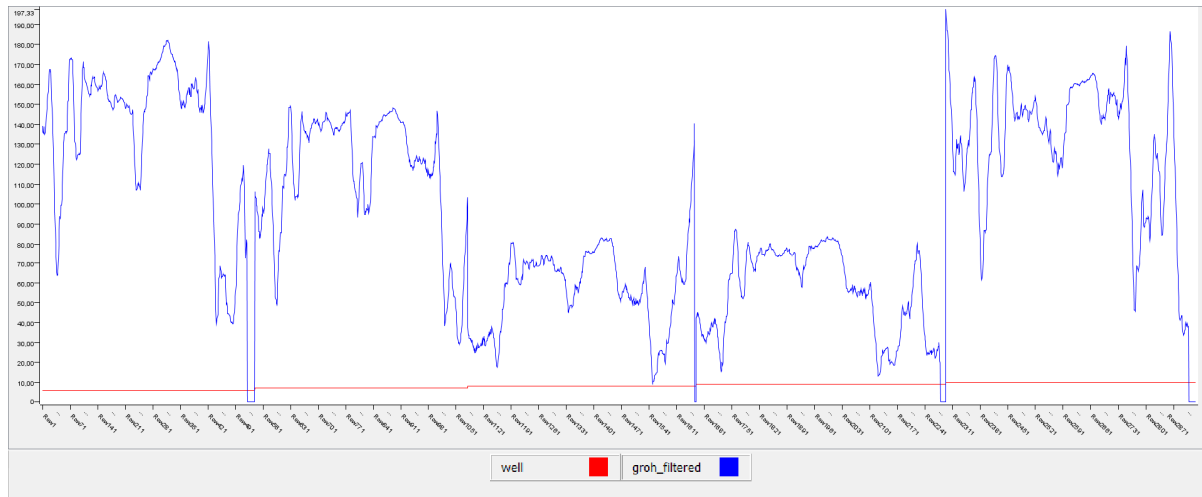
Dado que se requiere poder procesar todos los datos bajo un mismo modelo estructurado se realizaron diferentes actividades que permitieron la homogeneización de la información. De esta forma se estructuran los datos para que tengan las mismas características y puedan ser procesados uniformemente. Dentro de las actividades desarrolladas encontramos:

- Integración de datos: dado que se cuenta con múltiples orígenes de datos se realiza un volcado de información en base de datos. Por un lado se cuenta con toda la información proveniente de registros geofísicos de pozo en archivos planos del tipo \*.las. Por otro lado se cuenta con información de valores calculados en archivos \*.xls. Los valores se encuentran asociados a secciones transversales específicas para

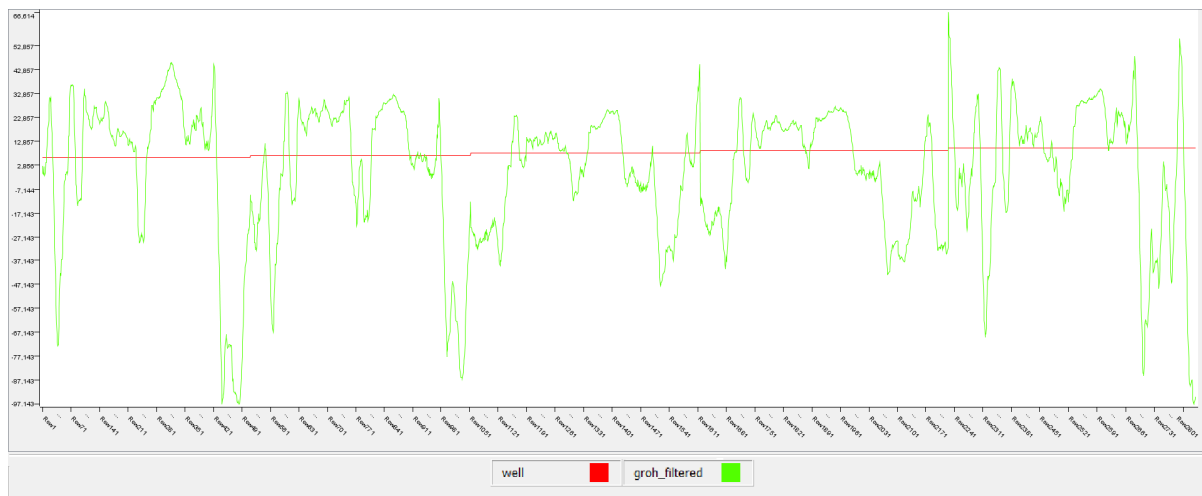
dimensión fractal por conteo de cajas, rayos gamma y valores de resistividad en promedio. Dada esta característica es necesario realizar las respectivas uniones entre el conjunto de registros originales del muestreo en sitio y el conjunto de la información vinculada a los registros calculados a partir de imágenes resistivas, así como el conjunto de los registros producto del procesamiento realizado por herramientas definidas en casos particulares. Como resultado de este proceso se obviaron los datos de los registros originales que no se encontraban en la respectiva unión con la información asociada a la dimensión fractal objetivo generada en el procesamiento de imágenes resistivas. Se resalta el hecho que el objetivo de la predicción es el valor asociado a la dimensión fractal. Las diferentes actividades realizadas en esta fase se ven reflejadas sobre los conjuntos de información de entrada para el modelo de predicción de dimensión fractal.

- Limpieza de datos: como primera instancia se eliminan valores negativos no representativos presentes en los registros geofísicos de pozo. Se identifica que en los múltiples registros recuperados en sitio se presentan valores constantes sobre las diferentes señales que corresponden a muestreos por defecto generados por la herramienta de lectura cuando no se encuentra un valor asociado a la propiedad objetivo. Típicamente las lecturas por defecto se caracterizaron en los registros originales con el valor de -999.25. En la tarea de limpieza asociada a la fase de preparación de datos como primera etapa se reemplazó dicho valor por cero. Se identificó que todos los valores representativos esperados son presentados en los registros con valores positivos estrictamente mayores a cero.
- Transformación de datos: la transformación se realiza en múltiples etapas. En una primera etapa se procede a establecer la misma precisión decimal para todos los valores asociados a los diferentes registros. La caracterización general definida corresponde a 3 decimales. En una segunda etapa se identifica que existen comportamientos esperados determinados a nivel de pozo. Sin embargo, al realizar un análisis en conjunto se evidencia que el modelo responde no al valor particular de la medida más si al comportamiento de la propiedad asociada. Por lo anterior, se realiza el desplazamiento en el eje horizontal teniendo como referencia la media de cada señal. Para esta etapa se generó la media de la señal para cada pozo y se desplazó el mismo valor en el eje vertical donde este corresponde a cero en la señal resultante.

**Figura 3-16:** Ejemplo: señal GROH con el filtro media móvil original



**Figura 3-17:** Ejemplo: señal GROH media móvil valores desplazados



Una tercera etapa se encuentra asociada a la normalización de los valores. Para el proceso de normalización se utilizó el método z-score el cual se desarrolla acorde a (ver ecuación (3.1)):

$$v' = \frac{v - \text{mean}_A}{\text{stand\_dev}_A} \quad (3.1)$$

### 3.3 Modelo de predicción de dimensión fractal

La información producto de los diferentes procesos realizados en la fase de preparación de datos fue consolidada en una única tabla en la base de datos la cual aporta la información de entrada al modelo. Posterior a la integración de información se procede con la ejecución del modelo de predicción el cual se encuentra conformado por las siguientes etapas:

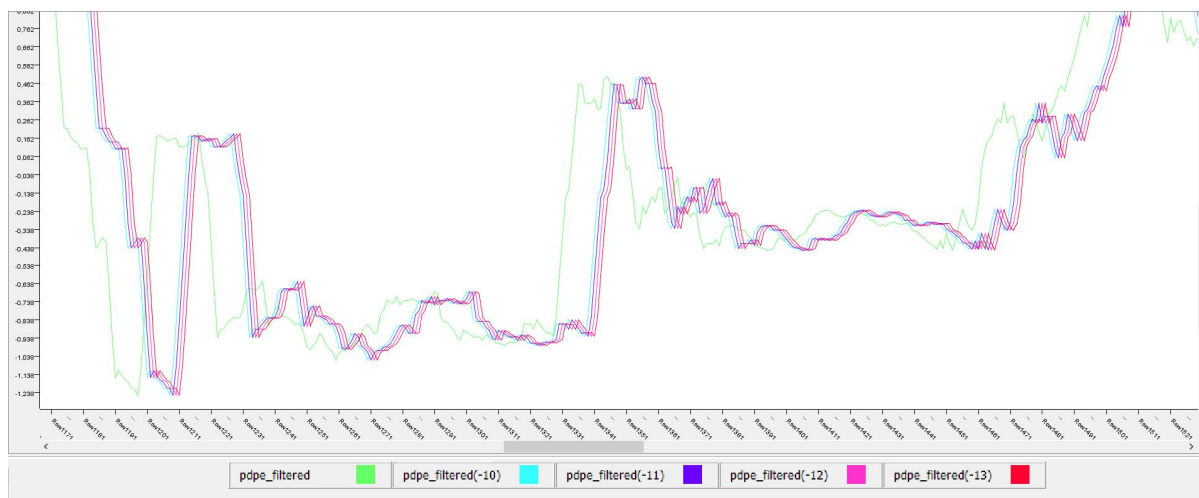
**Figura 3-18:** Modelo de proceso



- Lectura de datos: toda la información resultante de la etapa de pre-procesamiento de información es cargada en el modelo.
- Normalización de datos: para todos los datos exceptuando el filtro media móvil para la dimensión fractal el cual es el registro objetivo a predecir se realiza una normalización por el método de Z-score.
- Generación de ventanas: se generan ventanas que permiten incluir la información de los registros vecinos (la definición de tamaño de ventana pre-configurada en el modelo corresponde a 13). De esa forma se introduce la información relevante del entorno en el modelo.

- Selección de registros: se definen estrategias de combinación de registros. Acorde a las diferentes iteraciones los registros utilizados son categorizados y aquellos que no aportan información representativa para el modelo son identificados dentro del conjunto que degradan el modelo. En el transcurso de las diferentes iteraciones pueden ser redefinidos los conjuntos ya que las relaciones entre los registros afectan el comportamiento del modelo.
- Particionamiento: se realiza una partición de datos donde una parte de los datos corresponde a los registros para entrenamiento y los registros sobrantes serán objeto de la predicción.
- Red neuronal probabilística: la información es sometida a entrenamiento y se realiza la predicción de valores de dimensión fractal.

**Figura 3-19:** Ejemplo: ventana PDPE - valor original y valores desplazados 10 al 13



### 3.4 Evaluación experimental

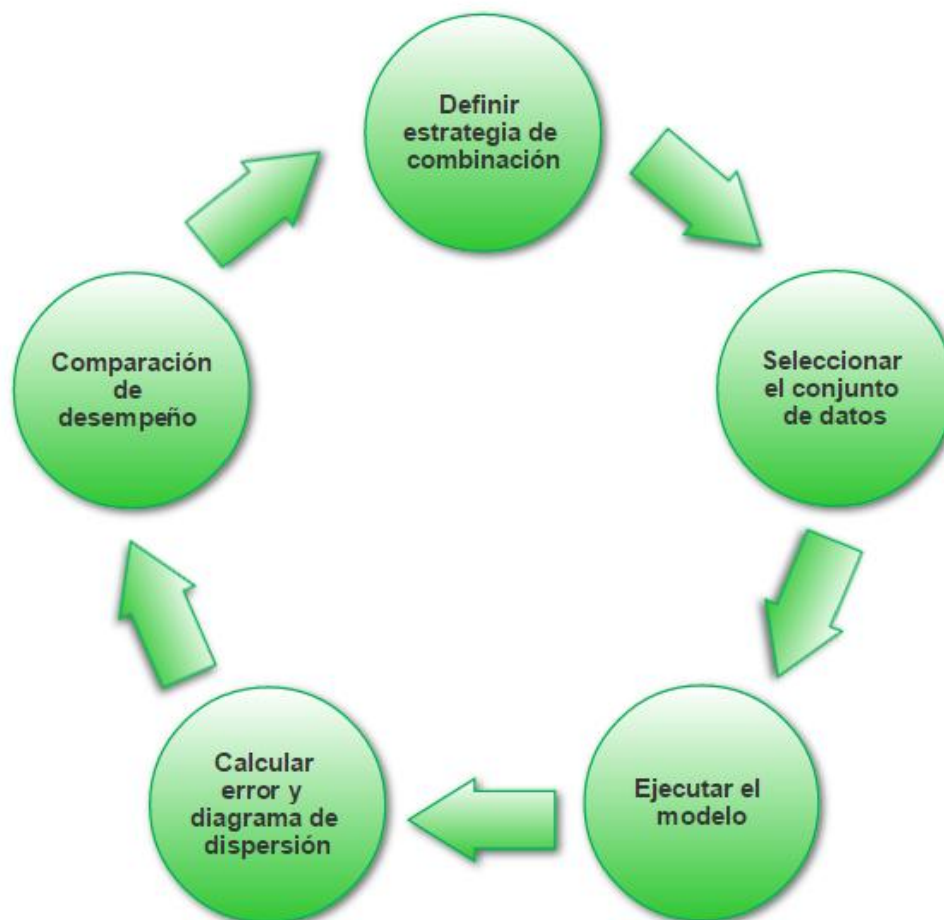
La evaluación del modelo tiene como base la caracterización de registros de dimensión fractal. Para el caso de estudio se determinó la dimensión fractal promedio filtro media móvil siendo este el valor esperado. Se realizaron múltiples iteraciones sobre el modelo determinando la validez del mismo teniendo como base el análisis gráfico de diagramas de dispersión y la validación probabilística asociada a los resultados vinculados a los datos de predicción.



El sistema fue inicializado con información relacionada con conjuntos de datos de 18 registros para 5 pozos. Los clasificadores disponibles implementados utilizan redes neuronales.

Las evaluaciones se realizaron utilizando todas las posibles combinaciones permitiendo identificar la cercanía entre los registros, el desempeño y la predicción. Para caracterizar un banco de pruebas que represente las métricas de desempeño de la selección se implementó el siguiente modelo.

**Figura 3-20:** Diagrama actividades proceso de clasificación y validación de modelo



- Definir la estrategia de combinación.
- Seleccionar el conjunto de datos.  
De acuerdo a la estrategia definida, seleccionar un grupo de diferentes registros de pozo de los disponibles en el sistema.  
Realizar la combinación de registros para caracterizar el desempeño.
- Ejecutar el modelo.  
Sobre el conjunto de datos seleccionado correr el modelo de predicción.
- Calcular el error y generar el respectivo diagrama de dispersión.
- Comparación de desempeño.  
Asignar el pivote de comparación de error y diagrama de dispersión. Para este proceso se realizan las siguientes tareas:
  - En la inicialización los datos obtenidos de la primera ejecución son asignados al pivote.
  - Computar la desviación media, el  $R^2$  e identificar el respectivo diagrama de dispersión para la combinación objetivo.
  - Comparar el desempeño obtenido entre el pivote y la combinación.
  - El mejor desempeño se asigna como nuevo pivote.

## 4. Discusión de resultados

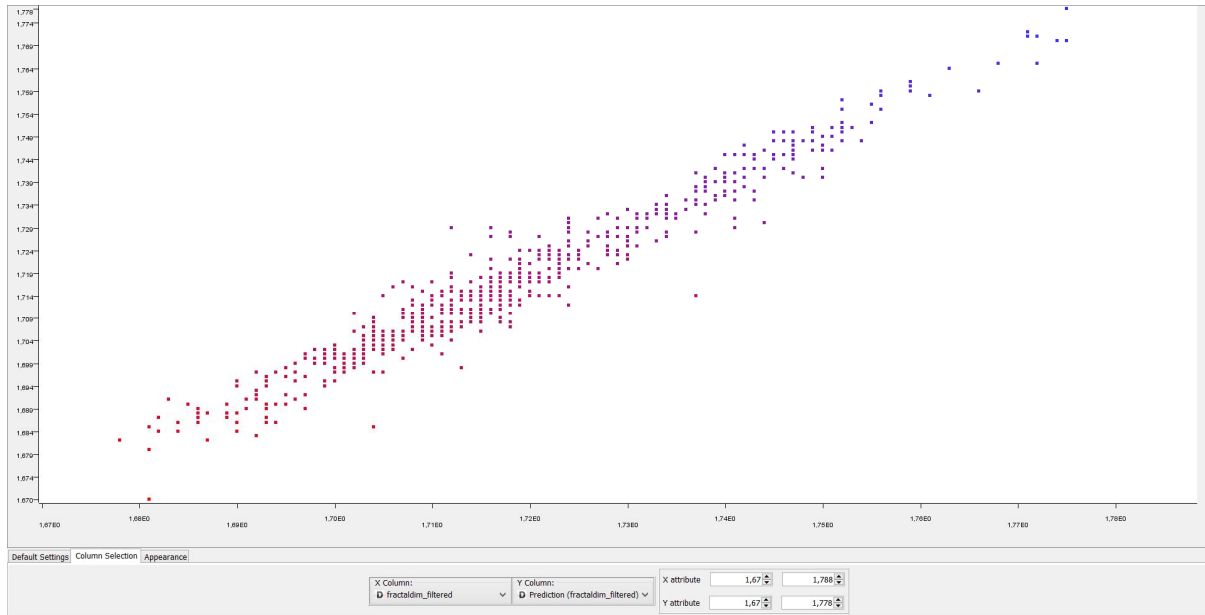
El caso de estudio se desarrolló sobre registros eléctricos de pozo adquiridos en la cuenca del Catatumbo en Norte de Santander Colombia. La información involucrada fue almacenada en una base de datos construida en POSTGRES. Los diversos procesos que tuvieron lugar asociados a la fase de preparación de datos se desarrollaron a través de la ejecución de scripts sobre la misma. Como resultado de los diferentes procesos se consolidó una única tabla la cual presenta la información de entrada para el caso de estudio requerida por el proceso. El proceso asociado al modelo de predicción de dimensión fractal fue implementado sobre la herramienta de software libre KNIME (Konstanz Information Miner) Analytics Platform.

Acorde a la metodología de CRISP-DM se desarrollaron iterativamente las diferentes fases hasta el desarrollo de la versión final del modelo. Las actividades asociadas a la fase de preparación de datos (ver figura 3.1) y las actividades de etapas identificadas en el diagrama del proceso de clasificación y validación de modelo (ver figura 3.20) se ejecutaron tantas veces como fue requerido hasta obtener la combinación que cumplía con las métricas esperadas.

En la iteración cero se realizó una combinatoria inicial con todos los registros disponibles. Teniendo como base esta iteración cero se determinó el conjunto de datos base para la ejecución del modelo. Posterior a ello se ejecutaron las diferentes iteraciones sobre múltiples estrategias de clasificación. Para cada combinación seleccionada se realizó la comparación de los datos de la combinatoria actual sobre la anterior con mayor precisión seleccionada. Es decir, el conjunto de datos seleccionados presenta niveles de error asociados a la desviación media y  $R^2$  mejores o cercanos a la anterior combinatoria seleccionada. Adicionalmente se evaluó el factor de presentar una tendencia en la dispersión que permitiera deducir que el conjunto de datos asociados a la estrategia de

clasificación seleccionados presentarán relaciones que permiten generar el valor de predicción para el valor objetivo. Para el caso de estudio el valor corresponde al filtro media móvil para la dimensión fractal.

**Figura 4-1:** Diagrama de dispersión dimensión fractal y predicción de la dimensión fractal



**Figura 4-2:** Error predicción de la dimensión fractal combinación seleccionada

R <sup>2</sup> :	0,95
Mean absolute error:	0,003
Mean squared error:	0
Root mean squared error:	0,004
Mean signed difference:	-0

Para la combinatoria seleccionada se utilizaron 2925 valores de registro donde se seleccionaron aleatoriamente el 80% de los registros para entrenamiento y el 20% fueron utilizados para la predicción y evaluación. Los registros de entrenamiento se seleccionaron aleatoriamente en sección transversal para los diferentes valores de los registros geofísicos en la fila seleccionada. La representación de dispersión nos permite identificar con claridad que existe una tendencia de los valores resultado. Si bien no son

exactos se determina que existen múltiples relaciones entre los diferentes registros de pozo que inciden sobre el valor de dimensión fractal.

El coeficiente de determinación ( $R^2$ ) resultante es de 0.95. Los registros de predicción tienen una media absoluta del error igual a 0.003. El error de la raíz cuadrada de la media (RMSE) es igual a 0.004.

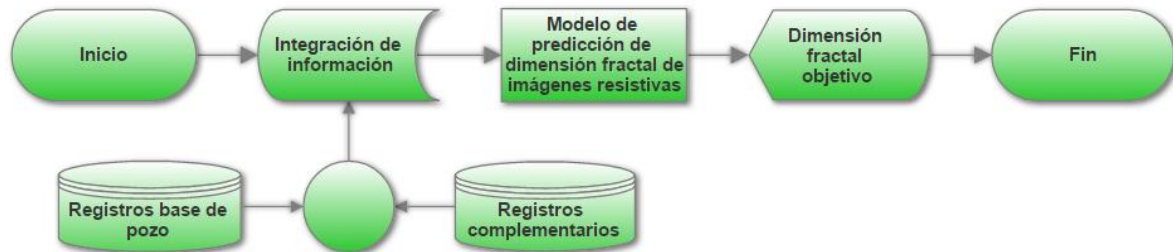
Con respecto al proyecto de investigación desarrollado tenemos que es posible caracterizar el registro de dimensión fractal a partir de registros de pozo. El modelo se adapta a las diferentes características de los múltiples registros disponibles donde por una parte encontramos el conjunto base con la información de los registros requeridos para generar la predicción de valores estimados de dimensión fractal. Por otra parte, se identifica un segundo conjunto de datos el cual contiene los registros complementarios que permiten aumentar la precisión en los valores resultantes de la predicción. Como tercer conjunto de datos encontramos los registros que no inciden en la generación del valor estimado de dimensión fractal y que adicionalmente ocasionan una degradación significativa en el resultado de la predicción.

Para el caso de estudio los registros base corresponden a los siguientes registros:

- Fractaldim: dimensión fractal
- GROH: registro de rayos gamma en pozo abierto (Gamma Ray Open Hole).
- GRPO: registro de rayos gamma para el Potasio (GR POTASSIUM).
- GRTH: registro de rayos gamma para el Torio (GR THORIUM).
- GRUR: registro de rayos gamma para el Uranio (GR URANIUM).
- GR: Gamma Ray adquirido con el registro de imágenes resistivas.

Los registros complementarios son:

- CGXT: temperatura.
- CLDC: calibrador del registro de densidad.
- DCOR: densidad corregida.
- NPRL: porosidad neutrón en matriz caliza.
- PDPE: factor fotoeléctrico.

**Figura 4-3:** Diagrama de flujo proceso generación de registro de dimensión fractal

El diagrama de flujo presenta el proceso requerido para la caracterización de registros de dimensión fractal acorde al modelo generado por el proyecto de investigación.

Producto de la ejecución iterativa del modelo de predicción de dimensión fractal se identificaron los siguientes tipos de registros los cuales no presentan relaciones significativas y que por el contrario degradan los resultados obtenidos en la generación de los valores de predicción:

- DDLB: latero-perfil de resistividad profunda.
- DSLB: latero-perfil de resistividad somera.
- MINV: mmr micro-log inverso.
- MNRL: mmr micro-log normal.
- SMTU: distancia de tensión en la boca del pozo.
- SPDL: potencial espontáneo.
- SRES: Registro de resistividad promedio calculada a partir de imágenes resistivas.

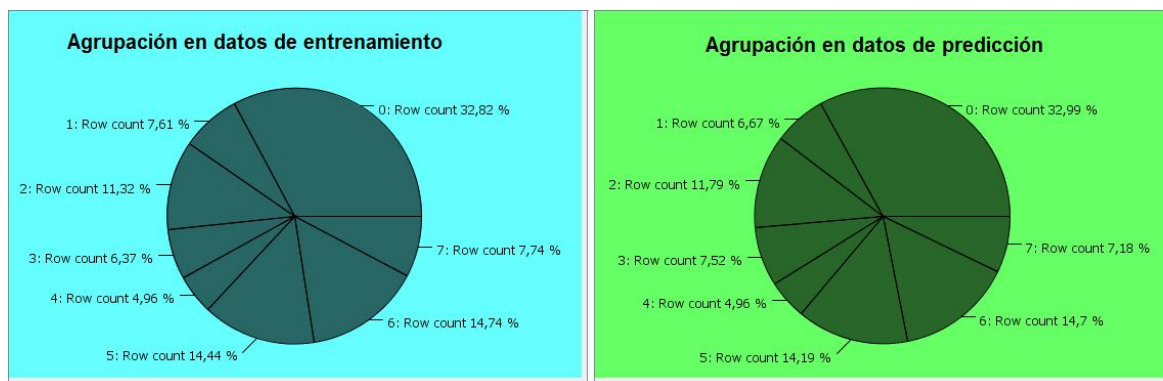
Acorde con la identificación y clasificación de los diferentes tipos de registro ejecutados sobre el modelo de predicción es posible determinar que existe un núcleo de registros que permite evidenciar una tendencia en la predicción de valores objetivo para la dimensión fractal sobre el filtro media móvil. Los registros seleccionados se complementaron con la información de los filtros generales aplicados para todos los registros disponibles. Los filtros corresponden a media móvil, pasa bajas y pasa altas. Todos los registros de entrada para el modelo se presentan en la tabla 4.1.

**Tabla 4-1:** Información de registros de entrada para el modelo de predicción de registros de dimensión fractal

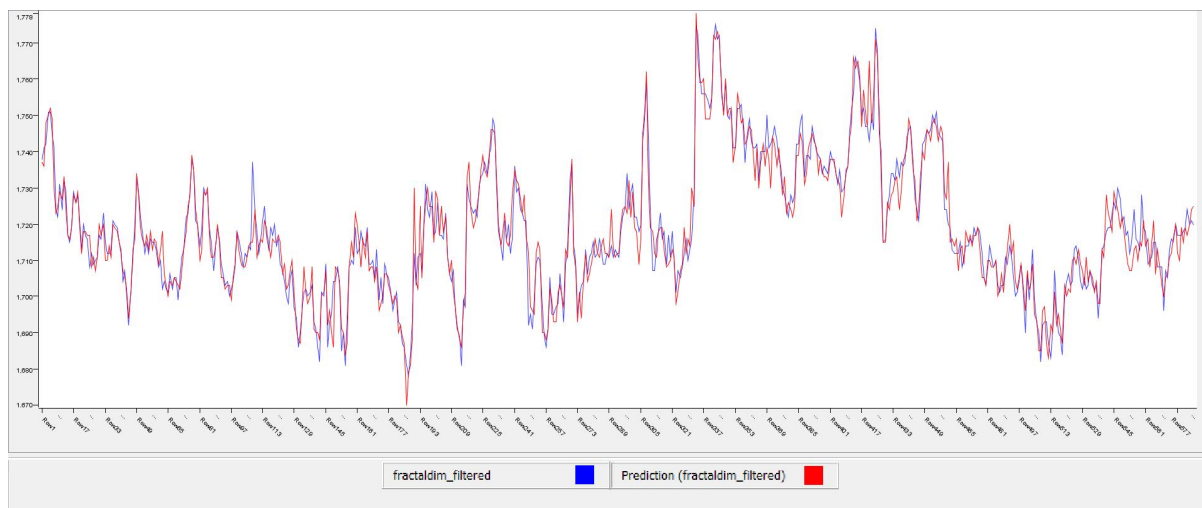
Señal original	Señal resultante	Descripción
FRACTALDIM	FRACTALDIM	Señal original
GROH	GROH_FILTERED	Filtro media móvil sobre señal original
GRPO	GRPO_FILTERED	
GRTH	GRTH_FILTERED	
GRUR	GRUR_FILTERED	
FRACTALDIM	FRACTALDIM_FILTERED	
GR	GR_FILTERED	
CGXT	CGXT_FILTERED	
CLDC	CLDC_FILTERED	
DCOR	DCOR_FILTERED	
NPRL	NPRL_FILTERED	
PDPE	PDPE_FILTERED	
GROH	GROHLP	
GRPO	GRPOLP	
GRTH	GRTHLP	
GRUR	GRURLP	
FRACTALDIM	FDIMLP	
GR	GRAYLP	
CGXT	CGXTLP	
CLDC	CLDCLP	
DCOR	DCORLP	
NPRL	NPRLLP	
PDPE	PDPELP	
GROH	GROHHP	Filtro pasa altas sobre señal original
GRPO	GRPOHP	
GRTH	GRTHHP	
GRUR	GRURHP	
FRACTALDIM	FDIMHP	
GR	GRAYHP	
CGXT	CGXTHP	
CLDC	CLDCHP	
DCOR	DCORHP	
NPRL	NPRLHP	
PDPE	PDPEHP	

Para identificar las relaciones de proximidad de los registros se realizaron 8 agrupaciones. Las secciones transversales de entrenamiento corresponden a 2340. Las secciones de validación del modelo de predicción corresponden a 585. En la figura 4.4 se presenta la variación en porcentaje de los agrupamientos de los registros de entrenamiento contra la correspondiente agrupación de los registros objeto de predicción.

**Figura 4-4:** Agrupamientos para registros de entrenamiento



**Figura 4-5:** Registros de dimensión fractal y predicción de la dimensión fractal



En la figura 4-5 se presenta la comparación gráfica del registro de dimensión fractal original en azul, contra la predicción del registro de dimensión fractal en rojo. Es posible identificar con claridad que existe una tendencia clara en la predicción hacia los valores



---

objetivo. Si bien los valores no son exactos se presenta una similitud significativa entre la señal original y la señal resultante del proceso de ejecución del modelo de predicción. Acorde con la validación de resultado de ejecución del modelo sobre los diversos registros geofísicos de pozo involucrados así como los registros generados a partir de imágenes resistivas es posible determinar que existen relaciones que permiten predecir los valores de dimensión fractal asociados para zonas con características geológicas similares.



## 5. Conclusiones y recomendaciones

### 5.1 Conclusiones

El modelo propuesto, partiendo de información proveniente de registros eléctricos de pozo, permite generar una estimación del valor asociado a la dimensión fractal generado a partir de imágenes resistivas. El modelo permite la interacción de los diferentes registros de pozo disponibles generando agrupaciones de tipos de registros acordes a la relevancia que tengan para realizar una estimación del valor de dimensión fractal. Producto de la clasificación se identificaron así tres conjuntos de tipos de registro relevantes. Dentro de estos encontramos los registros (logs) base, registros (logs) complementarios y registros (logs) no válidos para la generación de estimación de valor de dimensión fractal.

Dentro del conjunto de registros base encontramos los muestreos del tipo de radioactividad natural así como el registro de dimensión fractal. El origen de datos corresponde tanto a los registros de pozo como a los que son producto del procesamiento de imágenes resistivas del mismo tipo. Los registros (logs) identificados en este grupo base corresponden a:

- Fractaldim: dimensión fractal
- GROH: registro de rayos gamma en pozo abierto (Gamma Ray Open Hole).
- GRPO: registro de rayos gamma para el Potasio (GR POTASSIUM).
- GRTH: registro de rayos gamma para el Torio (GR THORIUM).
- GRUR: registro de rayos gamma para el Uranio (GR URANIUM).
- GR: Gamma Ray calculado a partir del registro de imágenes resistivas.

De esta agrupación se identifica claramente que estos registros se encuentran asociados directamente a características litológicas de la roca. Se resalta el hecho de que existe

una relación entre los diferentes muestreos asociados a muestreos de rayos gamma particulares y la dimensión fractal objetivo.

El conjunto de registros (logs) complementarios se caracteriza por que a partir de estos no se puede realizar una estimación de la dimensión fractal pero sí inciden de manera directa en la precisión de los valores de predicción. Este grupo de registros tiene como objeto mejorar la tendencia en los valores de predicción incidiendo directamente en la tendencia del valor objetivo. Los registros identificados en este grupo corresponden a:

- CGXT: temperatura.
- CLDC: calibrador del registro de densidad.
- DCOR: densidad corregida.
- NPRL: porosidad neutrón en matriz caliza.
- PDPE: factor fotoeléctrico.

De esta agrupación se identifica claramente la información asociada a registros de porosidad y densidad de la roca. Los valores de densidad inciden en el aumento de la precisión para la estimación del valor de dimensión fractal. Sin embargo, los diferentes factores involucrados como son la calibración y variación de la temperatura permiten caracterizar un ajuste en el valor de estimación. Adicionalmente se identifica una incidencia particular proveniente del factor fotoeléctrico. Este registro permite identificar diferentes tipos de roca o permite identificar características litológicas.

Un tercer conjunto de registros (logs) es considerado y agrupa todos aquellos que no aportan información relevante dentro del modelo de estimación de valor de dimensión fractal y que por el contrario inciden directamente de forma negativa tanto en la estimación de valores como en la tendencia de valores de predicción. Es decir, los valores asociados a estos registros (logs) no se deben considerar para la ejecución del modelo de predicción propuesto. Los registros identificados en este grupo base corresponden a:

- DDLB: latero-perfil de resistividad profunda.
- DSLB: latero-perfil de resistividad somera.
- MINV: mnr micro-log inverso.

- MNRL: mmr micro-log normal.
- SMTU: distancia de tensión en la boca del pozo.
- SPDL: potencial espontáneo.
- SRES: Registro de resistividad promedio calculada a partir de imágenes resistivas.

Los registros (logs) asociados a latero-perfil y micro-log están vinculados directamente con la resistividad de la formación. Siendo afectados principalmente por los fluidos en la roca.

Sobre el universo de registros (logs) disponibles podemos concluir que la información asociada a porosidad, propiedades de los fluidos y litología presentan relaciones que inciden en la generación de valores estimados para la dimensión fractal. Es decir, de la información proveniente de estas propiedades se encuentran relaciones que pueden incidir en la estimación de valores de dimensión fractal. Adicionalmente los registros no directos pero que se ven afectados por características litológicas particulares y propiedades de los fluidos no deben ser considerados dentro del modelo.

Acorde al trabajo de investigación realizado es posible identificar que dependiendo del tipo de registro eléctrico y las propiedades objeto de la medición se puede generar de una forma adaptativa la clasificación los registros (logs) y ejecutar el modelo de predicción de estimación de valores de dimensión fractal. El modelo implementado permite cargar toda la información y solo es requerida la selección de registros base para la ejecución del modelo generando de una forma automática no supervisada las diferentes reglas que permiten la estimación del valor de dimensión fractal. Dentro de la información requerida es necesario realizar el cálculo de dimensión fractal e incluir los filtros media móvil, pasa bajas y pasa altas para los registros involucrados en el proceso.

La aplicación *Imagen Fractal* evidenció ser confiable para el cálculo de dimensión fractal en imágenes resistivas por el método del conteo de cajas. Los resultados obtenidos durante la fase de validación tomando como ejemplo de prueba una figura fractal conocida (alfombra de Sierpinski), mostraron un error inferior al 0.007% demostrando la efectividad del método y la confiabilidad en la aplicación seleccionada.

El modelo permite caracterizar el registro de dimensión fractal utilizando técnicas de minería de datos teniendo como base registros de dimensión fractal con la respectiva asociación a registros eléctricos de pozo vinculados con la radioactividad, litología y muestreos generales. La validación del modelo se realizó sobre el caso de estudio partiendo de datos provenientes de registros adquiridos en la cuenca del Catatumbo, departamento de Norte de Santander – Colombia. Para el caso de estudio con las características particulares de la geología local y regional se evidencia que es posible estimar el valor de dimensión fractal.

Dentro de las características que se resaltan encontramos que los valores a los registros son susceptibles a diferentes tipos de ruido por lo que es necesario implementar filtros para atenuar los comportamientos asociados en las señales particulares. Acorde a lo anterior resaltamos que el ruido blanco tiene una fuerte presencia en el registro de dimensión fractal la cual es directamente el objetivo de predicción. El modelo de predicción generado permite realizar la estimación de valores sobre el registro de dimensión fractal aplicando un filtro media móvil que permita el atenuar el respectivo tipo de ruido identificado.

El modelo adaptativo de generación del registro de dimensión fractal de imágenes resistivas basado en registros de variables geofísicas tiene la característica de ser no supervisado por lo que siempre se obtendrán resultados. Sin embargo, se resalta el hecho que la tendencia en la estimación de valores de predicción se encuentra directamente relacionada con los tipos de registro seleccionados, así como los filtros aplicados y la ejecución completa de las actividades asociadas a la fase de preparación de datos.

## **5.2 Recomendaciones**

Dentro del modelo propuesto es necesaria una identificación por ejecución del modelo donde manualmente se realiza una clasificación de registros para identificar cuáles aportan información relevante en el momento de realizar la predicción de valores asociados al registro de dimensión fractal. El proceso en mención puede ser abordado por medio de la implementación de algoritmos evolutivos que realicen la selección y

clasificación de los registros (logs) que presentan afectación tanto positiva como negativa en la predicción objetivo.

El modelo se desarrolló teniendo como base los registros eléctricos de pozo disponibles para un caso de estudio asociado a un área donde se contaba con información de 6 pozos. Dada la caracterización geológica particular para definir reglas que permitan la generalización del modelo sobre regiones con características geofísicas y condiciones geológicas similares se debe alimentar el modelo con un volumen de orígenes de datos significativamente mayor.

Dentro del proceso realizado se identificó que el universo de datos iniciales contaba con 14946 secciones transversales de registros eléctricos de pozo. Fue descartada la información asociada al pozo 11 ya que esta no cuenta con la información del conjunto de registros base requeridos para el proceso los cuales correspondían a 2889 secciones. Al realizar el procesamiento de imágenes resistivas el total de secciones válidas corresponden a 2925 secciones transversales. Leal propone un modelo que permite de una manera automática identificar los posibles puntos de afectación por fracturas naturales abiertas donde los datos de dimensión fractal se integran con registros de rayos gamma y resistividad utilizando máquinas de soporte vectorial. Se recomienda actualizar dicho modelo para que utilice los valores asociados a la dimensión fractal filtrada de ruido blanco. En este punto la aplicación del modelo propuesto aporta información significativa relacionada con la adición de información para las secciones no involucradas en el proceso de generación de información partiendo de imágenes resistivas. De esta forma sería viable utilizar la información relacionada con los 12012 datos de sección transversal asociados a los pozos 6 al 10 que no fueron considerados originalmente.





## Bibliografía

- [1] Adeli, H., & Panakkat, A. (2009). A Probabilistic Neural Network for Earthquake Magnitude Prediction. *Neural Netw.*, 22(7), 1018–1024.  
<http://doi.org/10.1016/j.neunet.2009.05.003>
- [2] AIP Asesoría en Ingeniería de Petróleos. (2009). Caracterización de Yacimientos, Cuenca de Catatumbo - Colombia. *Informe Técnico*, 6–18.
- [3] Nikravesh, M., & Aminzadeh, F. (2001). Mining and fusion of petroleum data with fuzzy logic and neural network agents. *Journal of Petroleum Science and Engineering*, 29(3-4), 221–238. [http://doi.org/10.1016/S0920-4105\(01\)00092-4](http://doi.org/10.1016/S0920-4105(01)00092-4)
- [4] Barnsley, M. F., Hutchinson, J. E., & Stenflo, Ö. (2008). V-variable fractals: Fractals with partial self similarity. *Advances in Mathematics*, 218(6), 2051–2088.  
<http://doi.org/10.1016/j.aim.2008.04.011>
- [5] Pardo, A., Barrero, D., Vargas, C. A., & Martínez, J. F. (2007). *Colombian Sedimentary Basins: Nomenclature, boundaries and Petroleum Geology, a New Proposal*. Agencia Nacional de Hidrocarburos - A.N.H.-. Retrieved from [http://www.anh.gov.co/Informacion-Geologica-y-Geofisica/Cuencas-sedimentarias/Documents/colombian\\_sedimentary\\_basins.pdf](http://www.anh.gov.co/Informacion-Geologica-y-Geofisica/Cuencas-sedimentarias/Documents/colombian_sedimentary_basins.pdf)
- [6] Bouman, Q. & Gibson, G. (1964). *A Geological Review of the Barco Area*. Reporte interno Colpet. ISN 426.
- [7] Berthold, M. R., & Hand, D. (Eds.). (2003). *Intelligent Data Analysis: An Introduction* (Second edi). <http://doi.org/10.1007/978-3-540-48625-1>
- [8] Cadena, J. A., Cadena, J. M., & Pérez, S. M. (2008). Probabilistic Neural Networks Application in the Diagnosis of Transformer Incipient Faults. *Scientia et Technica*, 39(XIV), 48–53.
- [9] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R. (2000). *CRISP-DM 1.0 - Step-by-step data mining guide*.

- [10] Chen, W.-S., & Du, Y.-K. (2009). Using neural networks and data mining techniques for the financial distress prediction model. *Expert Systems with Applications*, 36(2), 4075–4086. <http://doi.org/10.1016/j.eswa.2008.03.020>
- [11] Coates, G., Xiao, L., & Prammer, M. (1999). *NMR logging: principles and applications*. Haliburton Energy Services. <http://doi.org/10.1002/cmr.1029>
- [12] Cover, T., & Thomas, J. (2006). *Elements of information theory* (2nd edition). John Wiley & Sons.
- [13] ECOPETROL. (2012). Informe Técnico Anual Contractual Año 2012. *Gerencia Catatumbo Orinoquia, Vicepresidencia de Producción*, 5–20.
- [14] Falconer, K. (2004). *Fractal Geometry: Mathematical Foundations and Applications* (2nd edition). John Wiley & Sons.
- [15] Foroutan-Pour, K., Dutilleul, P., & Smith, D. L. (1999). Advances in the Implementation of the Box-counting Method of Fractal Dimension Estimation. *Appl. Math. Comput.*, 105(2-3), 195–210. [http://doi.org/10.1016/S0096-3003\(98\)10096-6](http://doi.org/10.1016/S0096-3003(98)10096-6)
- [16] Grami, A. (2016). *Introduction to Digital Communications*. *Introduction to Digital Communications*. Elsevier. <http://doi.org/10.1016/B978-0-12-407682-2.00002-8>
- [17] Grassberger, P. (1983). Generalized dimensions of strange attractors. *Physics Letters A*, 97(6), 227–230. [http://doi.org/10.1016/0375-9601\(83\)90753-3](http://doi.org/10.1016/0375-9601(83)90753-3)
- [18] Higuchi, T. (1988). Approach to an irregular time series on the basis of the fractal theory. *Physica D: Nonlinear Phenomena*, 31(2), 277–283. [http://doi.org/10.1016/0167-2789\(88\)90081-4](http://doi.org/10.1016/0167-2789(88)90081-4)
- [19] Hu, A., & Hassanein, A. (2014). Model to estimate fractal dimension for ion-bombarded materials. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 323, 82–86. <http://doi.org/10.1016/j.nimb.2014.01.007>
- [20] Jizba, P., & Arimitsu, T. (2004). The world according to Rényi: thermodynamics of multifractal systems. *Annals of Physics*, 312(1), 17–59. <http://doi.org/10.1016/j.aop.2004.01.002>
- [21] Xia, K., Zhang, Z., Zhou, Y., & Si, S. (2013). Development on Intelligent Data Mining System in Oil Logging. *Journal of Computational Information Systems*, 5(11), 1843–1850.
- [22] Kotu, V., & Deshpande, B. (2015). *Predictive Analytics and Data Mining*. *Predictive Analytics and Data Mining*. Elsevier. <http://doi.org/10.1016/B978-0-12-801460-8.00002-1>

- [23] Krenker, A., Kos, A., & Bešter, J. (2011). *Introduction to the Artificial Neural Networks*. (P. K. Suzuki, Ed.). InTech. <http://doi.org/10.5772/15751>
- [24] Lapidus, M. (2004). *Fractal Geometry and Applications: A Jubilee of Benoit Mandelbrot*.
- [25] Leal F, J. A. (2013). Identificación y modelado de fracturas naturales en pozos de un yacimiento de hidrocarburos ubicado en la cuenca de Catatumbo, Departamento Norte de Santander - Colombia, utilizando registros de imágenes resistivas y datos de dimensión fractal. Retrieved from <http://www.bdigital.unal.edu.co/12849/>
- [26] Li, J., Du, Q., & Sun, C. (2009). An improved box-counting method for image fractal dimension estimation. *Pattern Recognition*, 42(11), 2460–2469. <http://doi.org/10.1016/j.patcog.2009.03.001>
- [27] Li, Y., & Huang, R. (2015). Relationship between joint roughness coefficient and fractal dimension of rock fracture surfaces. *International Journal of Rock Mechanics and Mining Sciences*, 75, 15–22. <http://doi.org/10.1016/j.ijrmms.2015.01.007>
- [28] Liu, R., Li, B., & Jiang, Y. (2016). A fractal model based on a new governing equation of fluid flow in fractures for characterizing hydraulic properties of rock fracture networks. *Computers and Geotechnics*, 75, 57–68. <http://doi.org/10.1016/j.compgeo.2016.01.025>
- [29] Liu, R., Jiang, Y., Li, B., & Wang, X. (2015). A fractal model for characterizing fluid flow in fractured rock masses based on randomly distributed rock fracture networks. *Computers and Geotechnics*, 65, 45–55. <http://doi.org/10.1016/j.compgeo.2014.11.004>
- [30] Pombo, H. L. (2010). Análisis y Desarrollo de Sistemas de Realidad Aumentada, 124. Retrieved from <http://eprints.ucm.es/11425/>
- [31] Maloney, P. S., & Specht, D. F. (1989). The use of probabilistic neural networks to improve solution times for hull-to-emitter correlation problems. *Neural Networks, 1989. IJCNN., International Joint Conference on 18 Jun, 1, 289–294*. <http://doi.org/10.1109/IJCNN.1989.118593>
- [32] Fowler, A. D. (1991). The fractal geometry of nature (revised edition). *Computers & Geosciences*, 17(7), 1065–1066. [http://doi.org/10.1016/0098-3004\(91\)90106-N](http://doi.org/10.1016/0098-3004(91)90106-N)
- [33] Markou, M., & Singh, S. (2003). Novelty detection: a review—part 2: *Signal Processing*, 83(12), 2499–2521. <http://doi.org/10.1016/j.sigpro.2003.07.019>
- [34] Maugeri, L. (2013). The Shale Oil Boom: A US Phenomena, (June), 64. Retrieved from [http://belfercenter.hks.harvard.edu/files/The US Shale Oil Boom Web.pdf](http://belfercenter.hks.harvard.edu/files/The_US_Shale_Oil_Boom_Web.pdf)

- [35] Meneveau, C., Sreenivasan, K. R., Kailasnath, P., & Fan, M. S. (1990). Joint multifractal measures: Theory and applications to turbulence. *Phys. Rev. A*, 41(2), 894–913. <http://doi.org/10.1103/PhysRevA.41.894>
- [36] Miikkulainen, R. (2010). Encyclopedia of Machine Learning. In C. Sammut & G. I. Webb (Eds.), (pp. 988–989). Boston, MA: Springer US. [http://doi.org/10.1007/978-0-387-30164-8\\_837](http://doi.org/10.1007/978-0-387-30164-8_837)
- [37] Moon, F. C. (2008). *Chaotic and Fractal Dynamics: Introduction for Applied Scientists and Engineers*. John Wiley & Sons.
- [38] Orallo, J. H., Ramírez, M. J., & Ferri, Q. C. (2004). *Introducción a la Minería de Datos*. Prentice Hall.
- [39] Ozaktas, H. M., Zalevsky, Z., & M. Alper Kutay. (2001). *The Fractional Fourier Transform: with Applications in Optics and Signal Processing*. John Wiley & Sons.
- [40] Paredes, C. (1999). Distribución espacial y zonificación tectónica de los morfolineamientos en la Isla Decepción ( Shetland del Sur , Antártida ), (1), 75–78.
- [41] Roy, A., Perfect, E., Dunne, W. M., & McKay, L. D. (2007). Fractal characterization of fracture networks: An improved box-counting technique. *Journal of Geophysical Research: Solid Earth*, 112(12), 1–9. <http://doi.org/10.1029/2006JB004582>
- [42] Sahoo, P., & Barman, T. K. (2012). *Mechatronics and Manufacturing Engineering. Mechatronics and Manufacturing Engineering*. Elsevier. <http://doi.org/10.1533/9780857095893.159>
- [43] Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2009). The Graph Neural Network Model. *Neural Networks, IEEE Transactions on*, 20(1), 61–80. <http://doi.org/10.1109/TNN.2008.2005605>
- [44] Shanmuganathan, S., & Samarasinghe, S. (Eds.). (2016). *Artificial Neural Network Modelling* (1st ed.). Springer International Publishing.
- [45] Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation* (2nd ed.). Upper Saddle River, NJ, USA: Prentice Hall PTR.
- [46] Sivanandam, S. N., Sumathi, S., & Deepa, S. N. (2006). *Introduction to neural networks using MATLAB 6.0*. New Delhi: McGraw Hill Education (India) Private Limited.
- [47] Schlumberger. (2006). Naturaleza de Yacimientos Naturalmente Fracturados. *OildField Review*.
- [48] Smith, J. O. (2007). *Introduction to Digital Filters: with Audio Applications*. W3K Publishing.

- [49] Smith, S. W. (1999). *The Scientist and Engineer's Guide to Digital signal processing. California Technical Publishing, 1(2)*. <http://doi.org/10.1109/79.826412>
- [50] Sun, Y., Zhu, W., Guo, J., & Kosaka, K. (2001). Comparison of Fractal Dimensions of Fracture Surfaces Originated from Different Mechanical Conditions in the Northern Tarim Basin , Northwest China, *36(36)*, 75–79.
- [51] Talanquer, V. (2002). *Fractus, fracta, fractal: fractales, de laberintos y espejos*. S.L. fondo de cultura economica de españa.
- [52] Twiss, R. J., & Moores, E. M. (2007). *Structural Geology* (2nd editio). Freeman, W. H. & Company.
- [53] Wada, Y., & Kuwana, K. (2013). A numerical method to predict flame fractal dimension during gas explosion. *Journal of Loss Prevention in the Process Industries, 26(2)*, 392–395. <http://doi.org/10.1016/j.jlp.2011.11.006>
- [54] Wang, B. (2012). UCGE Reports Reservoir Characterization and Horizontal Well Placement Guidance Acquisition by Using GIS and Data Mining Methods.
- [55] Watanabe, K., & Takahashi, H. (1993). Fractal characterization of subsurface fracture network for geothermal energy extraction system. *18th Workshop on Geothermal Reservoir Engineering Stanford University*, 119–124. Retrieved from [http://helios.osti.gov/geothermal/product.biblio.jsp?osti\\_id=888908](http://helios.osti.gov/geothermal/product.biblio.jsp?osti_id=888908)
- [56] Webster, J. G., & Eren, H. (Eds.). (2014). *Measurement, Instrumentation, and Sensors Handbook, Second Edition: Spatial, Mechanical, Thermal, and Radiation Measurement* (2nd editio). CRC Press.
- [57] Williams, D. B., & Madisetti, V. K. (Eds.). (1997). *The Digital Signal Processing Handbook* (2nd editio). CRC Press.
- [58] Witten, I. H., Frank, E., & Hall, M. a. (2011). *Data Mining: Practical Machine Learning Tools and Techniques, Third Edition. Annals of Physics* (Vol. 54). Morgan Kaufmann. <http://doi.org/10.1002/1521-3773>