

**IDENTIFICACIÓN DE ZONAS POTENCIALMENTE PRODUCTORAS DE  
HIDROCARBUROS UTILIZANDO MAQUINA DE SOPORTE VECTORIAL Y ANÁLISIS  
LINEAL DISCRIMINANTE**

**HENRY MAURICIO CASTRO GERARDINO**

**COD. 2002200576**

**DIEGO ALEJANDRO DIAZ RODRIGUEZ**

**COD. 2003102836**

**UNIVERSIDAD SURCOLOMBIANA**

**FACULTAD DE INGENIERÍA**

**INGENIERÍA ELECTRÓNICA**

**2010**

**IDENTIFICACIÓN DE ZONAS POTENCIALMENTE PRODUCTORAS DE  
HIDROCARBUROS UTILIZANDO MAQUINA DE SOPORTE VECTORIAL Y ANÁLISIS  
LINEAL DISCRIMINANTE**

**HENRY MAURICIO CASTRO GERARDINO  
DIEGO ALEJANDRO DIAZ RODRIGUEZ**

**Trabajo de Grado para optar al título de  
Ingeniero Electrónico**

**Director  
CARLOS ALBERTO PEREZ CAMACHO  
Ingeniero Electrónico**

**UNIVERSIDAD SURCOLOMBIANA  
FACULTAD DE INGENIERÍA  
INGENIERÍA ELECTRÓNICA**

**2010**

## CONTENIDO

INTRODUCCIÓN .....	10
1. TRANSFORMADA WAVELET.....	15
1.2 TRANSFORMADA WAVELET CONTINUA (CWT) Y TRANSFORMADA WAVELET DISCRETA (DWT) .....	16
1.3 REDUCCIÓN DE PERTURBACIONES MEDIANTE UMBRALIZACIÓN.....	17
1.3.1 Umbralización Rígida y Flexible .....	19
1.3.2 Umbralización rígida.....	19
2. ANÁLISIS DISCRIMINANTE LINEAL .....	20
2.1 LDA PARA DOS CLASES.....	21
3. MAQUINAS DE SOPORTE VECTORIAL.....	26
3.1 CONJUNTOS SEPARABLES LINEALMENTE .....	27
3.2 CONJUNTOS NO SEPARABLES LINEALMENTE.....	30
4. PERFILAJE DE POZOS .....	34
4.1 PROPIEDADES PETROFÍSICAS DE LAS ZONAS PRODUCTORAS .....	34
4.1.1 Porosidad.....	35
4.1.2 Permeabilidad.....	35
4.1.3 Resistividad.....	35
4.1.4 Saturación de Agua.....	36
4.2 PERFILES SELECCIONADOS.....	36
4.2.1 Gamma Ray (GR).....	37
4.2.2 Potencial Espontáneo (SP).....	38
4.2.3 Resistividad (RT) .....	39
4.2.4 Porosidad Neutrónica (PHIN). .....	40
4.2.5 Densidad de Formación (ROHB) .....	41
4.3 PRESENTACIÓN DE LOS PERFILES .....	42
5. DESARROLLO METODOLÓGICO .....	44
5.1 ANALISIS DE LA SEÑAL .....	44
5.2 NORMALIZACIÓN .....	48
5.3 ELECCIÓN DE LA MEJOR BASE .....	50
5.4 REDUCCIÓN DE PERTURBACIONES MEDIANTE UMBRALIZACIÓN .....	52
5.5 SELECCIÓN DE LA POBLACIÓN DE ENTRENAMIENTO.....	56
5.5.1 Entrenamiento con Datos Preprocesados.....	57
5.5.2 Entrenamiento para Datos Preprocesados y Normalizados .....	58
5.5.3 Entrenamiento para Datos Preprocesados, Normalizados y Filtrados .....	59
5.6 ANÁLISIS LINEAL DISCRIMINANTE (LDA) .....	60
5.7 MÁQUINAS DE SOPORTE VECTORIAL (SVM) .....	62
6. ANÁLISIS DE RESULTADOS .....	66
6.1 RESULTADOS ENTRENAMIENTO .....	66
6.1.1 Resultados Entrenamiento (LDA). .....	66
6.1.2 Resultados Entrenamiento (SVM). .....	67
6.2 RESULTADOS CLASIFICACIÓN .....	68
6.2.1 Resultados Clasificación (LDA).....	68

6.2.2 Resultados de Clasificación (SVM). .....	69
7. CONCLUSIONES Y RECOMENDACIONES .....	72
8. BIBLIOGRAFIA .....	74
9. ANEXOS .....	76
9.1 TOTURIAL IDENTIFICADOR DE ZONAS POTENCIALMENTE PRODUCTORAS CON LDA Y SVM.....	76

## LISTA DE FIGURAS

Figura 1: Transformada corta de Fourier Vs Análisis Wavelet .....	15
Figura 2: Señal seno – Wavelet Daubechies db5 .....	16
Figura 3: Limitadores de umbral. (a) Umbral rígido. (b) Umbral flexible.....	19
Figura 4: Gráfica de comportamiento de Efectividad .....	20
Figura 5: Gráfico de las dos categorías con los dos predictores de ejes ortogonales.....	22
Figura 6: Nube de datos con vectores de referencia .....	23
Figura 7: Máxima transformación de la nube de datos .....	23
Figura 8: Transformación de grupo de tres categorías .....	24
Figura 9: Distribución de datos.....	24
Figura 10: Transformación para la máxima separación con traslape .....	25
Figura 11: Esquema de dos conjuntos de vectores linealmente separados mediante un hiperplano que maximiza el margen m. ....	28
Figura 12: Esquema de la transformación de dos conjuntos no separables linealmente mediante la función. ....	31
Figura 13: Esquema ilustrativo del perfilaje de un pozo .....	34
Figura 14: Niveles de Rayos Gamma en areniscas (sandstone), arcillas (shale) y carbón (coal).....	37
Figura 15: Medición de un potencial natural entre un electrodo en el pozo y en la superficie de la tierra .....	39
Figura 16: Diferentes términos utilizados para designar la resistividad de las diferentes zonas de invasión en un pozo .....	40
Figura 17: Diagrama general del sistema .....	44
Figura 18: Perfil de Resistividad Verdadera (Rt) del LL-81 .....	46
Figura 19: Perfil de Gamma Ray (GR) del LL-97.....	46
Figura 20: Perfil preprocesado de Gamma Ray del LL-97 .....	48
Figura 21: Perfil preprocesado de Resistividad Verdadera del LL-81 .....	48
Figura 22: Perfil preprocesado y normalizado de Resistividad Verdadera del LL- 81 .....	50
Figura 23: Perfil preprocesado y normalizado de Gamma Ray del LL-97 .....	50
Figura 24: Diagrama de flujo para elección wavelet básica .....	51
Figura 25: Perfil preprocesado, normalizado y filtrado nivel 3 de Gamma Ray (GR) del llano-97 .....	56
Figura 26. Diagrama de bloques LDA .....	61
Figura 27: Comparativo de Normalización con Preprocesado de la LDA .....	61
Figura 28. Diagrama de bloques de LDA .....	62
Figura 29: Comparación de una compuerta AND con un entrenamiento con 2 variables y una variable .....	63
Figura 30: Entrenamiento del perfil de Densidad de Formación (RHOB) del LL- 105 .....	64
Figura 31: Entrenamiento del perfil de Resistividad Verdadera (RT) del LL-105 ..	64

Figura 32: Clasificación del perfil de Densidad de Formación (RHOB) del LL-102 .....	65
Figura 33: Clasificación del perfil de Resistividad Verdadera (RT) del LL-105 .....	65
Figura 34: Visualización entorno grafico del programa.....	76
Figura 35: Tabla de visualización de las zonas productoras.....	77
Figura 36: Axis que ilustran los registros del perfil del pozo .....	77
Figura 37: Interfaz de Carga de Datos .....	78
Figura 38: Ventana de selección del perfil .....	78

## LISTA DE CUADROS

Cuadro 1: Porosidad de reservorios detríticos.....	35
Cuadro 2: Permeabilidad en reservorios de petróleo y gas.....	35
Cuadro 3: Resistividad de algunos fluidos, rocas y minerales .....	36
Cuadro 4: Valores típicos de GR .....	37
Cuadro 5: Valores típicos de pb.....	42
Cuadro 6: Datos de máximos y mínimos de las zonas cañoneadas en los diferentes pozos analizados. ....	45
Cuadro 7: Datos de máximos y mínimos de los diferentes pozos analizados en general. ....	45
Cuadro 8: Datos de máximos y mínimos de los pozos sustrayendo los datos atípicos.....	47
Cuadro 9: Datos de los límites utilizados en el preprocesamiento.....	47
Cuadro 10: Datos de máximos y mínimos normalizados de las zonas cañoneadas en los diferentes pozos analizados.....	49
Cuadro 11: Análisis del desempeño, media del error de umbralización flexible	53
Cuadro 12: Análisis del desempeño, media del error de umbralización rígida ...	53
Cuadro 13: Análisis del desempeño, varianza del error de Umbralización flexible .....	54
Cuadro 14: Análisis del desempeño, varianza del error de Umbralización rígida	54
Cuadro 15: Datos de máximos y mínimos filtrados a nivel 3 de las zonas cañoneadas en los diferentes pozos analizados. ....	55
Cuadro 16: Comparativo de datos preprocesados.....	57
Cuadro 17: Comparativo de datos preprocesados y Normalizados.....	58
Cuadro 18: Comparativo de datos preprocesados, normalizados y Filtrados .....	59
Cuadro 19: Comportamientos de efectividad de las mejores LDA entrenadas...	66
Cuadro 20: Resumen de las LDA entrenadas de mejor efectividad.....	66
Cuadro 21: Entrenamiento con datos Preprocesados .....	67
Cuadro 22: Entrenamiento con datos Preprocesados y Normalizados .....	67
Cuadro 23: Entrenamiento con datos Preprocesados, Normalizados y Filtrado.	68
Cuadro 24: Validación del clasificador LDA. ....	68
Cuadro 25: Comparación de efectividad punto a punto de los tres tipos de clasificaciones de SVM en todos los pozos .....	69
Cuadro 26: Zonas potencialmente productoras de SVM y LDA .....	70
Cuadro 27: Zonas potencialmente productoras de SVM y LDA.....	71

## RESUMEN

Para el desarrollo del análisis de pozos petrolíferos se debe realizar el desplazamiento de instrumentos sensoriales para la detección de perfiles, obteniendo así, las señales no estacionarias de estos mismos. Estas señales no se pueden analizar con métodos tradicionales como la transformada de Fourier, lo cual hace más difícil los análisis respectivos, con el fin de verificar y cuantificar la cantidad de hidrocarburo en poros de rocas reservorio.

En anteriores estudios se trabajó técnicas de tratamiento de señales como la transformada Wavelet (WT), Lógica difusa (Fuzzy Logic) y otras técnicas de aprendizaje como redes neuronales (RNA) y técnicas de reducción dimensional como análisis de componentes principales (PCA), que se han sumado a la tarea de la interpretación de perfiles de pozo abierto, no haciendo uso de técnicas de aprendizaje automático como Análisis Lineal Discriminante (LDA) y Máquinas de Soporte Vectorial (SVM).

El objetivo primordial de esta tesis es la búsqueda de una buena caracterización basada en patrones que determinaron cuando en un pozo existía hidrocarburo de acuerdo a las señales registradas por las herramientas, para esto es necesario el análisis estadístico de las señales para así mismo hacer uso de la normalización de datos, con el fin de mejorar la caracterización de datos, además de diseñar un filtro digital de ruido blanco gaussiano, mediante la transformada Wavelet (WT), a través del concepto de costo de información se selecciona la mejor onda madre wavelet, y posteriormente haciendo uso de la técnica umbralización: rígida ó flexible. Después con Análisis Discriminante Lineal (LDA) se ejecutará una reducción dimensional de los datos que serán posteriormente procesados con Máquina de Soporte Vectorial (SVM) para separar al máximo los elementos concluyendo con una discriminación de los casos o zonas con porcentajes altos de almacenamiento del hidrocarburo. De este procedimiento se obtuvo óptimos resultados, logrando una reducción de dimensionalidad con (LDA) muy buena y una clasificación con (SVM) que arrojó una efectividad superior al 90% con respecto a la clasificación de zonas ya cañoneadas en los pozos suministrados por ECOPETROL SA.



## ABSTRACT

To develop the analysis of oil wells should be made to the displacement of instruments for detecting sensory profiles, thus obtaining non-stationary signals of the same, not being able to analyze with traditional methods such as Fourier transform, the more difficult respective analysis to verify and quantify the amount of hydrocarbon reservoir rock pores.

In previous studies involved with signal processing techniques such as wavelet transform (WT), Fuzzy logic (FL) and other learning techniques like neural networks (ANN) and dimensional reduction techniques such as principal component analysis (PCA) who have joined the task of interpretation of well logs open, not making use of machine learning techniques such as Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM).

The primary objective of this thesis is the search for a good characterization based on patterns determining when an oil well was based on the signals recorded by the tools, this requires statistical analysis of the signals so as to use it data standards in order to improve the characterization data, and design a digital filter Gaussian white noise by wavelet transform (WT), dare the concept of information cost selecting the best mother wavelet wave and then using the thresholding technique: rigid or flexible. Then linear discriminant analysis (LDA) execute a dimensional reduction of the data to be subsequently processed with Support Vector Machine (SVM) to separate the items ending up with discrimination cases or areas with high percentages of oil storage. In this procedure we obtained excellent results, achieving a dimensionality reduction with (LDA) and a very good classification (SVM) showed an effectiveness greater than 90% over the classification of areas and canyons in the wells provided by ECOPETROL SA.

## GLOSARIO

**ALGORITMOS GENÉTICOS:** Estos algoritmos hacen evolucionar una población de individuos sometiéndola a acciones aleatorias semejantes a las que actúan en la evolución biológica, así como también a una selección de acuerdo con algún criterio, en función del cual se decide cuáles son los individuos más adaptados, que sobreviven, y cuáles los menos aptos, que son descartados.

**ANÁLISIS DEL COMPONENTE INDEPENDIENTE (ICA):** Es un método computacional que sirve para separar una señal multivariante en subcomponentes aditivos suponiendo que la señal de origen tiene una independencia estadística y es no-Gausiana. Éste es un caso especial de separación ciega de las señales.

**ANÁLISIS DE COMPONENTES PRINCIPALES:** técnica para encontrar la transformación de un conjunto original de variables en un nuevo conjunto de variables no correlacionadas.

**COEFICIENTES DE APROXIMACIÓN:** secuencia producida mediante downsampling de la secuencia filtrada con un filtro pasa bajas asociado a cierta wavelet.

**COEFICIENTES DE DETALLE:** secuencia producida mediante downsampling de la secuencia filtrada con un filtro pasa altas asociado a cierta wavelet.

**COMPONENTES PRINCIPALES:** conjunto de variables no correlacionadas obtenido sin perder información a partir de un conjunto de variables originales.

**CONDICIONES KARUSH KUHN-TUCKER:** (También conocidas como las condiciones KKT o Kuhn-Tucker) son condiciones necesarias y suficientes para que la solución de una programación no lineal con restricciones de desigualdad sea óptima. Es una generalización del método de los Multiplicadores de Lagrange.

**CONJUNTO DE ENTRENAMIENTO:** colección de pares de entrenamiento consistente de un vector de valores de entrada junto con un vector de salidas deseadas.

**CONJUNTO DE VALIDACIÓN:** es usado para probar el desempeño de una red o un clasificador que fue previamente entrenado, está compuesto de pares similares de vectores de entrada y vectores de salidas deseadas.

**CONVOLUCIÓN:** operación matemática por medio de la cual se halla la salida de un sistema, conociendo su entrada y su respuesta al impulso.

**COVARIANZA:** medida estadística del grado en el cual se mueven juntas las variables aleatorias. Una covarianza positiva implica que una variable está por encima (debajo) de su valor medio mientras la otra variable está por encima (debajo) de su valor medio.

**DESVIACIÓN TÍPICA:** modo de expresar la fluctuación de una señal alrededor de su valor promedio. Está definida como la raíz cuadrada del promedio de las desviaciones al cuadrado, donde la desviación es la diferencia entre una muestra y la media.

**FILTRO PASA ALTAS:** sistema que atenúa todas las frecuencias que están por debajo de una frecuencia de corte y permite el paso de aquellas que se encuentran por encima de la misma.

**FILTRO PASA BAJAS:** sistema que atenúa todas las frecuencias que están por encima de una frecuencia de corte y permite el paso de aquellas que se encuentran por debajo de la misma.

**FUNCIONES KERNEL:** Constituyen una forma estándar de representar los datos sustituyendo las representaciones vectoriales por otra genérica aplicable a datos no vectoriales, permitiendo así construir algoritmos de aprendizaje genéricos que pueden utilizarse sobre cualquier tipo de dato (vectorial o no).

**FUNCIONES ORTOGONALES:** conjunto de formas de onda que se usan en la descomposición de señales y con producto interno entre ellas igual a cero.

**MULTIPLICADORES DE LANGRAGE:** Es un procedimiento para encontrar los máximos y mínimos de funciones de varias variables sujetas a restricciones. Este método reduce el problema restringido con  $n$  variables a uno sin restricciones de  $n + k$  variables, donde  $k$  es igual al número de restricciones, y cuyas ecuaciones pueden ser resueltas más fácilmente.

**PATRÓN:** conjunto de características o descriptores que representan un objeto o conjunto de objetos similares.

**PROGRAMACIÓN CUADRÁTICA (QP):** es un tipo especial de problema de optimización matemática. Es el problema de optimizar (reduciendo al mínimo o maximizando) una función cuadrática de varias variables conforme a apremios lineales en estas variables.

**SNR:** (Relación señal a ruido), es una medida utilizada para cuantificar la cantidad de una señal que ha sido corrompida por el ruido.

## **AGRADECIMIENTOS**

A Dios todopoderoso, que nos iluminó para la total culminación de nuestra tesis. A nuestras familias que siempre tuvieron confianza en nosotros y por su paciencia y comprensión.

Queremos agradecer a nuestro director de tesis el ingeniero Carlos Alberto Pérez Camacho docente del programa de Ingeniería Electrónica de la Universidad Surcolombiana, por sus conocimientos impartidos, sabias orientaciones y por su apoyo total para el desarrollo de este proyecto.

A nuestros compañeros de universidad y novias, cuyo apoyo incondicional sirvió de aliciente para el desarrollo de este proyecto. A nuestros profesores de la carrera, cuya entrega a su labor fue fuente de inspiración para este proyecto. Al coordinador del convenio, Ing. Freddy Escobar Macualo, por su colaboración.

Especial agradecimiento al Instituto Colombiano del Petróleo (ICP) por habernos facilitado los registros de perfilaje que fueron la base fundamental para la consecución de este logro.

## INTRODUCCIÓN

Desde principio del siglo pasado, entre 1912-1926 los hermanos Conrad y Marcel Schlumberger, inventaron un método de prospección minera, basada en mediciones eléctricas tomadas desde la superficie terrestre, desde ese entonces hasta el día de hoy, la evolución de los perfiles de pozo abierto es algo indudable, tanto su número como su eficacia, convirtiendo al perfilaje de pozo abierto en una poderosa herramienta para el análisis de los límites de las rocas reservorio y su capacidad de almacenar fluidos y diferenciar los fluidos.

A la par de la evolución de los perfiles de pozo abierto, la interpretación de los mismos también ha avanzado. Interpretados por un geólogo que trataba de analizar y buscar las zonas de interés, tarea que en muchos casos se realiza por simple inspección, en forma empírica, valiéndose más de su amplia experiencia para la determinación de estas zonas de interés, que en muchos casos pasando por alto algunos tramos que podrían llegar a ser de interés; Luego sumados al proceso de interpretación, se diseñaron programas por computadora que evaluaban la cantidad de hidrocarburo, a través del modelamiento matemático, relacionando cada medida a una ecuación que se podía integrar al factor de formación. Posteriormente se empezaron a utilizar métodos computacionales que realizaban de forma sistematizada las tareas realizadas de forma manual en los años anteriores, pero para su interpretación los registros necesitaban ser editados y establecer manualmente un conjunto de condiciones iniciales.

Con el devenir de nuevas tecnologías a finales del siglo XX, novedosas y eficaces técnicas se han sumado, a la tarea de interpretación de perfiles de pozo abierto, desde las redes neuronales artificiales (RNA), hasta sistemas basados en algoritmos computacionales, técnicas de gran desarrollo en innumerables aplicaciones en diferentes áreas.

Dadas estas condiciones, además de los trabajos desarrollados anteriormente especialmente en la Universidad Surcolombiana se ha escogido utilizar la transformada Wavelet, El análisis lineal discriminante (LDA) y las máquinas de soporte vectorial (SVM) los cuales son los métodos utilizados en este proyecto investigativo.

En la primera parte de este proyecto se presenta una corta explicación de la teoría básica de la Transformada Wavelet, luego el análisis de transformada wavelet en el uso de la reducción de las perturbaciones mediante umbralización. En la segunda parte se mostrarán las nociones principales del análisis lineal discriminante como técnica estadística de síntesis de la información o reducción de la dimensión. En la etapa siguiente se integran generalidades de las

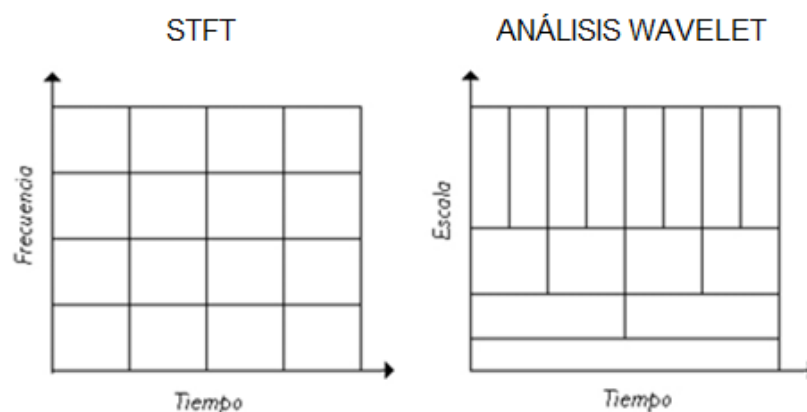
maquinas de soporte vectorial desde los conjuntos linealmente separables a los conjuntos no linealmente separables. A continuación se presentan las diferentes herramientas eléctricas de pozo abierto usados para el desarrollo del proyecto, ya que la preeminencia de los perfiles radica en que la información que se obtiene en su análisis e interpretación, da solución a la mayoría de las incógnitas que se plantean especialistas de diferentes disciplinas que participan en la prospección y desarrollo de los yacimientos de petróleo. Luego se plasma el desarrollo del sistema o desarrollo metodológico para el proyecto. Se podrá ver desde el mismo pre-procesamiento de los diferentes perfiles, normalización de los datos, procesamiento de la señal a través de la Transformada Wavelet, la reducción de perturbaciones mediante umbralización, la reducción de dimensionalidad con el análisis lineal discriminante (LDA), hasta llegar a la etapa final de clasificación de las zonas potencialmente productoras de hidrocarburos con máquinas de soporte vectorial; además se compendian los resultados y análisis de los descubrimientos, de suma importancia del proyecto. A continuación las conclusiones y anexos, donde se presenta a manera de cuadros los resultados obtenidos a través de tesis, y fundamenta nuestras conclusiones.

## 1. TRANSFORMADA WAVELET

En el procesamiento de señales se puede encontrar diferentes tipos de señales estacionarias y no estacionarias. Las señales estacionarias son localizadas en el tiempo, ya que su frecuencia no varía, este tipo de onda es estudiada por medio del análisis de Fourier, que permite su descomposición en términos de sus componentes sinusoidales, es decir, transforma la señal de la base de tiempo a la base de frecuencia y de igual manera permite el paso del dominio de la frecuencia al dominio del tiempo, sin embargo en este último se pierde información necesaria, que por ser de carácter estacionario no resulta relevante<sup>1</sup>.

Entre tanto las señales no-estacionarias, es decir, aquellas cuya frecuencia varía en el tiempo, resulta imposible efectuar el paso al dominio del tiempo con la transformada de Fourier porque no permite fijar en qué momento se presenta un cambio en la frecuencia. Como resultado de la desventaja mostrada en el análisis de Fourier, nace entonces la necesidad de contar con una representación localizada en la frecuencia y el tiempo. En la tarea por solucionar el inconveniente de la transformada de Fourier de resolución en el tiempo, Denis Gabor<sup>2</sup> adecuó la transformada haciendo uso de un procedimiento llamado ventanamiento. Este procedimiento consiste en asumir que para cada segmento la señal es estacionaria, dividiendo una señal  $x(t)$  en diminutos segmentos a través del tiempo de tal manera que se pueda así calcular la Transformada de Fourier clásica para cada porción de la señal. Al utilizar la STFT, el tamaño fijo de la ventana temporal, impide analizar pequeños detalles en señales de frecuencia variable, convirtiéndose en una deficiencia, es así como se introduce el análisis de la Transformada Wavelet WT.

Figura 1: Transformada corta de Fourier Vs Análisis Wavelet



Fuente: Wavelet Toolbox 4 User's Guide.

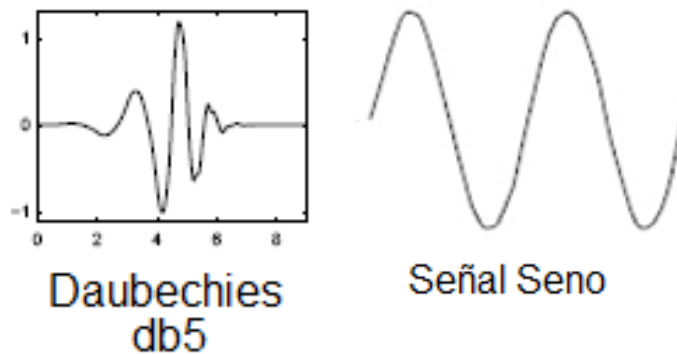
<sup>1</sup> Identificación de zonas productoras de hidrocarburos aplicando DSP y métodos de correlación de registros de perfilaje, Alfredo R. Palomino y Hugo A. Bernal, 2009

<sup>2</sup> GABOR, Denis. Theory of Communication. Journal of Institute for Electrical Engineering, 1946. p.429

De la Transformada Wavelet se pueden aprovechar ciertas propiedades: a) mediante la (WT) se puede crear un sin número de bases *wavelet*, cuya proyección sobre la señal  $f$  forma los coeficientes, y dentro de esas bases se eligen aquellas que convengan para la labor en particular que se realice, b) la capacidad de tomar información de las señales localizada en los dominios de tiempo y frecuencia. Además, la WT permite conseguir una representación, descomposición y reconstrucción de señales, que presenten cambios escabrosos en sus componentes de tiempo-frecuencia en forma instantánea a través del análisis de multirresolución con ventanas de longitud variable, adaptadas al cambio de frecuencia de la señal. Esta técnica permite el uso de intervalos grandes de tiempo en aquellos segmentos en los que se requiere mayor precisión en baja frecuencia, e intervalos más pequeños donde se requiere información en alta frecuencia<sup>3</sup>.

En oposición de Fourier, adonde las funciones base de duración infinita son senos y cosenos, en el análisis Wavelet la base son funciones localizadas en tiempo (traslación) y frecuencia (dilatación). Una Wavelet es una "pequeña onda" de permanencia delimitada, es decir, su energía está aglutinada en el tiempo alrededor de un punto, lo que suministra una adecuada herramienta para el análisis de fenómenos transitorios, no estacionarios, variables en el tiempo y aquellos que presenten discontinuidades.

Figura 2: Señal seno – Wavelet Daubechies db5



Fuente: Wavelet Toolbox 4 User's Guide.

## 1.2 TRANSFORMADA WAVELET CONTINUA (CWT) Y TRANSFORMADA WAVELET DISCRETA (DWT)

Las wavelets son familias de funciones que permiten el análisis de una señal para extraer características de espacio, tamaño y dirección. Una familia de wavelets se define como:

<sup>3</sup> S. Kouro y R. Musalem. Tutorial introductorio a la Teoría de Wavelet, Artículo presentado como trabajo de la asignatura Técnicas Modernas en Automática.



$$h_{a,b} = \frac{h\left(\frac{x-b}{\alpha}\right)}{\sqrt{|\alpha|}} \quad (\text{Ecuación 1})$$

Que es generada a partir de funciones madre  $h(x)$ , agregándole la escala a la cual permite que la onda se dilate o se contraiga, y la variable de traslación  $b$ , permitiendo que la señal se desplace en el tiempo. La transformada wavelet permite obtener información en tiempo frecuencia de señales no estacionarias. En el caso de la transformada wavelet continua CWT, se define como:

$$CWT_x^\Psi(\tau, \alpha) = \frac{1}{\sqrt{|\alpha|}} \int x(t) \Psi^*\left(\frac{t-\tau}{\alpha}\right) dt \quad (\text{Ecuación 2})$$

Donde la escala de la wavelet  $\alpha$  se puede considerar como el inverso de la frecuencia. El proceso de cálculo de la CWT se hace de manera similar al de la STFT. La CWT revela muchos detalles de la señal, debido a que realiza el proceso de cálculo en forma iterativa hasta que toda la señal es procesada cubriendo todas las escalas. No obstante, el costo computacional es mayor. La transformada wavelet discreta permite realizar el mismo procedimiento de la CWT en menor tiempo computacional, ya que calcula los coeficientes wavelet en intervalos discretos de tiempo y escala. Para el cálculo de la DWT, la wavelet  $\Psi$  es dilatada y trasladada en valores discretos.

$$\psi(2^k t + l),$$

Por lo tanto la DWT de  $f$  es una función de escala  $2^k$  y tiempo  $l$ :

$$DWT(f)(2^k, l) = \int f(t) \psi(2^k t + l) dt \quad (\text{Ecuación 3})$$

### 1.3 REDUCCIÓN DE PERTURBACIONES MEDIANTE UMBRALIZACIÓN

En los análisis basados en la WT para comprimir el ruido presente en señales no estacionarias (considerando dicho ruido como blanco gaussiano, puesto que se desconoce el nivel de ruido en la señal), la característica habitual es el manejo de un cambio en el dominio de la señal. En este dominio se realiza una serie de operaciones sobre los coeficientes, logrando normalmente una umbralización y posteriormente el cálculo de la transformada inversa para trasladar la señal al dominio del tiempo una vez ha sido modificado. Para la reducción del ruido se utiliza un procedimiento no lineal, denominado umbralización. El efecto de la

WT es filtrar la señal mediante un banco de filtros de dos tipos: pasa bajas o aproximación y pasa altas o detalles. El número de veces que es filtrada la señal viene determinado por el nivel de la descomposición. Para reducir el ruido de la señal, se eliminan los componentes obtenidos en la transformada Wavelet que estén por debajo de cierto umbral o multiplicarlos por un factor de ponderación, antes de llevar a cabo la transformada inversa. En estos elementos, el umbral o la ponderación, se encuentran las diferencias más significativas entre la mayoría de métodos propuestos en trabajos relacionados con esta aplicación. Para la reducción del ruido se utiliza un procedimiento no lineal, denominado umbralización, obtenido mediante cálculos estadísticos<sup>4</sup>.

Las condiciones exigibles a este proceso de reducción del ruido son la minimización en el error cuadrático medio (o lo mismo, que no maximice la SNR resultante) y la señal obtenida sea al menos tan suave como la señal inicial para evitar que, aunque se cumpla la primera condición, aparezcan picos en la señal resultante que se pueden confundir con información presente en la señal o puede dar lugar a que se eliminen puntos característicos de la señal<sup>5</sup>.

Para realizar la selección del umbral, se requiere minimizar el riesgo máximo relativo con respecto a la selección óptima del umbral, para esto utilizamos la siguiente ecuación.

$$\delta = \sqrt{2 \log(N)} \hat{\sigma} \quad (\text{Ecuación 4})$$

Donde la umbralización es realizada con la expresión:

$$C_{\delta}(i, j) = \begin{cases} 0, & \text{si } |C(i, j)| < \delta \\ \text{sgn}(C(i, j))(|C(i, j)| - \delta), & \text{si } |C(i, j)| \geq \delta \end{cases} \quad (\text{Ecuación 5})$$

Donde  $C(i, j)$  representa los coeficientes de los detalles obtenidos mediante la transformada Wavelet. El valor de  $\hat{\sigma}$  utilizado para este umbral que en este caso, un buen estimador de desviación típica es<sup>6 7</sup>:  $\hat{\sigma} = \text{mediana}(|C(i, j)|) / 0,6745$ . Este principio es utilizado en la estadística para diseñar estimadores porque obtiene el mínimo del máximo error cuadrático medio. Para realizar la

<sup>4</sup> D. Cuesta, "Estudio de métodos para procesamiento y agrupación de señales electrocardiográficas," Ph.D. dissertation, Valencia, 2001.

<sup>5</sup> D. Donoho, "De-noising by soft-thresholding," Tech. Rep. Dept. of Statistics, Stanford University, 1995.

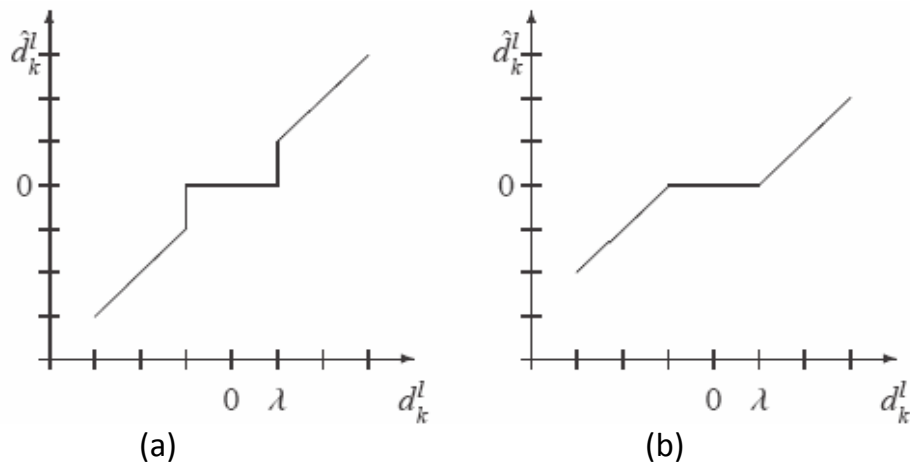
<sup>6</sup> Cristian Guarnizo, Victoria E. Montes, Julian Echeverry "Agrupamiento de impulsos neuronales en señales neurofisiológicas," Universidad Tecnológica de Pereira, 2006.

<sup>7</sup> D. Donoho, "De-noising by soft-thresholding," Tech. Rep. Dept. of Statistics, Stanford University, 1995.

selección del umbral, minimiza el riesgo máximo relativo con respecto a la selección óptima del riesgo.

**1.3.1 Umbralización Rígida y Flexible.** El método por el cual los coeficientes pequeños son movidos mientras los demás valores permanecen indelebles se denomina umbralización rígida (“keep or kill”). Una alternativa diferente es conocida como umbralización flexible: se resta el valor del umbral a la magnitud de los coeficientes por encima de éste, para que el campo entrada-salida sea continuo. Mientras a primera vista, la umbralización rígida puede parecer un acercamiento más natural, la continuidad de la operación de umbralización flexible posee importantes ventajas. Algunos algoritmos aún iguales no trabajan en combinación con umbralización rígida<sup>8</sup>.

Figura 3: Limitadores de umbral. (a) Umbral rígido. (b) Umbral flexible.



Fuente: Análisis de reducción de ruido en señales provenientes de microelectrodos de registro.

**1.3.2 Umbralización rígida.** Para la eliminación del ruido se aplica la siguiente transformada no lineal para los coeficientes wavelet:

$$F(x) = x \cdot I(|x| > \delta) \quad (\text{Ecuación 6})$$

Donde el umbral  $\delta$  es el valor a priori fijo de umbral  $I$  es la función indicadora.

**1.3.3 Umbralización flexible.** La umbralización flexible está en la elección de la transformada no-lineal sobre los coeficientes wavelet, usando:

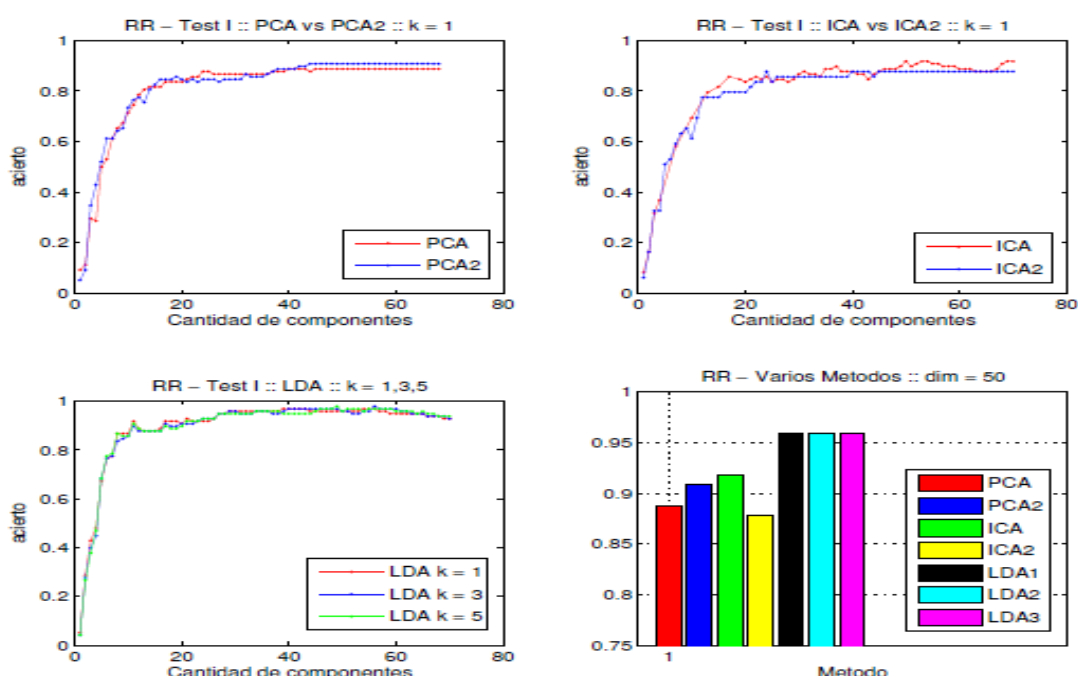
$$S(x) = \text{sgn}(x)(|x| - \delta)I(|x| > \delta) \quad (\text{Ecuación 7})$$

<sup>8</sup> M. Jansen, *Noise Reduction by Wavelet Thresholding*, Springer 2001.

## 2 ANÁLISIS DISCRIMINANTE LINEAL

Análisis Discriminante Lineal (LDA) y el correspondiente lineal discriminante de Fisher son los métodos utilizados en la estadística, para el reconocimiento de patrones y aprendizaje automático mediante la obtención de una combinación lineal de criterios que caracterizan o separan dos o más clases de objetos o acontecimientos. La combinación resultante puede ser utilizada como un clasificador lineal, o, más comúnmente, para reducción de dimensionalidad previo a la clasificación de sistemas compuestos de datos. Se puede ver en la gráfica 4, que muestra un estudio realizado de comparación de algoritmos de detección de patrones conducido para la detección de rostros, en el que se ven los porcentajes de efectividad obtenidos por diferentes métodos.

Figura 4: Gráfica de comportamiento de Efectividad



Fuente: Gráfica de Mauricio Delbracio, Matías Mateu

Como se ve en la gráfica, el algoritmo que demostró más eficiencia en el momento de la prueba fue LDA con porcentajes superiores al 95% de eficacia; para la realización de esta prueba se contó con más de 15 variables entre las cuales se encontraban: espacio entre ojos, ancho de la boca, tamaño de la nariz, junto a otros parámetros tomados, realizadas a los sujetos de pruebas a diferentes niveles de luz. LDA está estrechamente relacionado con ANOVA<sup>9</sup> (análisis de la varianza) y el análisis de regresión, que también tratan de expresar una variable dependiente como una combinación lineal de otras características o medidas<sup>10</sup>.

<sup>9</sup> Análisis Multivariante, HAIR, ANDERSON, TATHAM, BLACK. 5° Edición Editorial PEARSON Prentice Hall

<sup>10</sup> McLachlan (2004) Análisis discriminante y Reconocimiento de Patrones de Estadística en: Wiley Interscience

En los otros dos métodos sin embargo, la variable dependiente es una cantidad numérica, mientras que para LDA es una variable categórica (*es decir*, la etiqueta de clase). Regresión logística y regresión probit son más similares a LDA, como ellos también trabajan variables categóricas. Estos otros métodos son preferibles en aplicaciones donde no es razonable suponer las variables independientes y que tengan una distribución normal, que es un supuesto fundamental del método LDA.

LDA también está estrechamente relacionado con el análisis de componentes principales (PCA) y análisis de factores en que ambos buscan combinaciones lineales de las variables que mejor explican los datos<sup>11</sup>. LDA explícitamente intenta modelar la diferencia entre las clases de datos. PCA por otra parte no tiene en cuenta cualquier diferencia en la clase, y el análisis factorial construye las combinaciones de características basado en las diferencias más que las similitudes. El análisis discriminante es también diferente de análisis factorial, ya que no es una técnica de interdependencia: una distinción entre las variables independientes y variables dependientes (también llamadas variables de criterio) se debe hacer. LDA funciona cuando las mediciones efectuadas en las variables independientes para cada observación son cantidades continuas. Cuando se trata de variables independientes categóricas, la técnica equivalente es el análisis de correspondencias discriminante<sup>12</sup>.

## 2.1 LDA PARA DOS CLASES

Considere un conjunto de observaciones  $\vec{x}$  (También llamado características, atributos, variables o mediciones) para cada muestra de un objeto o evento con clase conocida  $y$ . Este conjunto de muestras se llama el conjunto de entrenamiento. El problema de clasificación consiste entonces en encontrar un buen predictor de la clase  $y$  de cualquier muestra de la misma distribución (no necesariamente de la formación conjunto) dado sólo una observación  $\vec{x}$ .

LDA aborda el problema asumiendo las funciones condicionales de densidad de probabilidad  $p(\vec{x}|y = 0)$  y  $p(\vec{x}|y = 1)$  son a la vez una distribución normal con media y covarianza de los parámetros  $(\vec{\mu}_0, \Sigma_{y=0})$  y  $(\vec{\mu}_1, \Sigma_{y=1})$ , Respectivamente. Bajo este supuesto, la solución óptima es de Bayes para predecir los puntos como de la segunda clase si la relación entre el diario de probabilidades está por debajo de un umbral  $T$ , de modo que;

$$(\vec{x} - \vec{\mu}_0)^T \Sigma_{y=0}^{-1} (\vec{x} - \vec{\mu}_0) + \ln |\Sigma_{y=0}| - (\vec{x} - \vec{\mu}_1)^T \Sigma_{y=1}^{-1} (\vec{x} - \vec{\mu}_1) - \ln |\Sigma_{y=1}| < T \quad (\text{Ecuación 8})$$

<sup>11</sup> Análisis Multivariado: Método de Componentes Principales, Secretaria General de los Estados Americanos, Programa Regional de Desarrollo Científico y Tecnológico.

<sup>12</sup> H. Abdi (2007) "análisis de correspondencias discriminantes." En: "MI Salkind (Ed.): Enciclopedia de medición y estadísticas". Thousand Oaks (CA): Sage. pp 270 a 275.

Sin ningún tipo de suposiciones, el clasificador resultante se conoce como QDA (Análisis Discriminante Cuadrática). LDA también hace que la simplificación de homoscedástica supuesto (*es decir*, que la clase covarianzas son idénticos, por lo que  $\Sigma_y = 0 = \Sigma_y = 1 = \Sigma$ ) y que las covarianzas tiene rango completo. En este caso, varios términos se solucionan y el criterio de decisión anterior se convierte en un umbral en el producto escalar

$$\vec{w} \cdot \vec{x} < c \quad (\text{Ecuación 9})$$

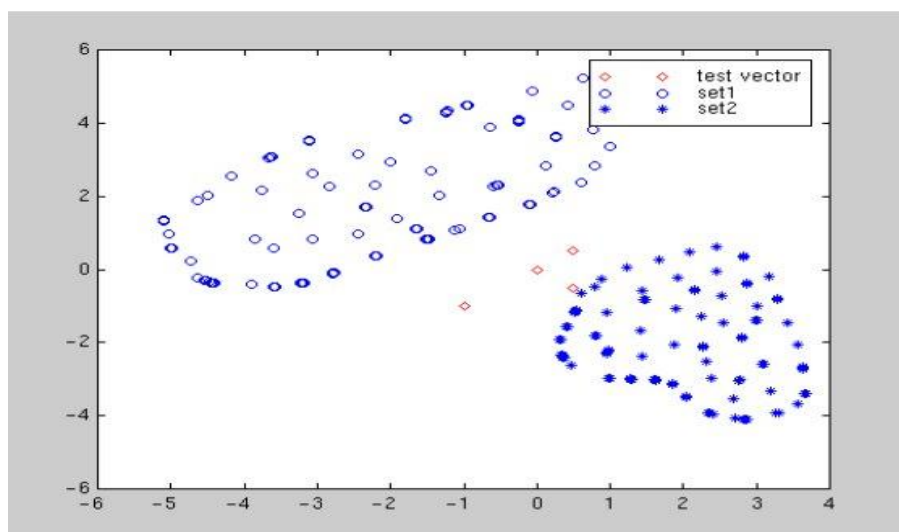
Para algunos  $c$  umbral constante, donde

$$\vec{w} = \Sigma^{-1}(\vec{\mu}_1 - \vec{\mu}_0) \quad (\text{Ecuación 10})$$

Esto significa que el criterio de una entrada  $\vec{x}$  estar en una clase  $y$  es puramente una función de esta combinación lineal de las observaciones conocidas.

A menudo es útil para ver a esta conclusión en términos geométricos: el criterio de una entrada  $\vec{x}$  estar en una clase  $y$  es puramente una función de la proyección del espacio multidimensional punto  $\vec{x}$  en la dirección  $\vec{w}$ . En otras palabras, la observación pertenece y si corresponde  $\vec{x}$  se encuentra en un cierto lado de una perpendicular al hiperplano  $\vec{w}$ . La ubicación del plano está definida por el umbral de  $c$ . Una forma más clara de comprender el funcionamiento de LDA, será a través de un ejemplo grafico, vamos a considerar una clasificación en la que participan dos categorías de objetivos y dos variables predictoras. En la Figura 5 se muestra un esquema de las dos categorías con los dos predictores de ejes ortogonales<sup>13</sup>:

Figura 5: Gráfico de las dos categorías con los dos predictores de ejes ortogonales



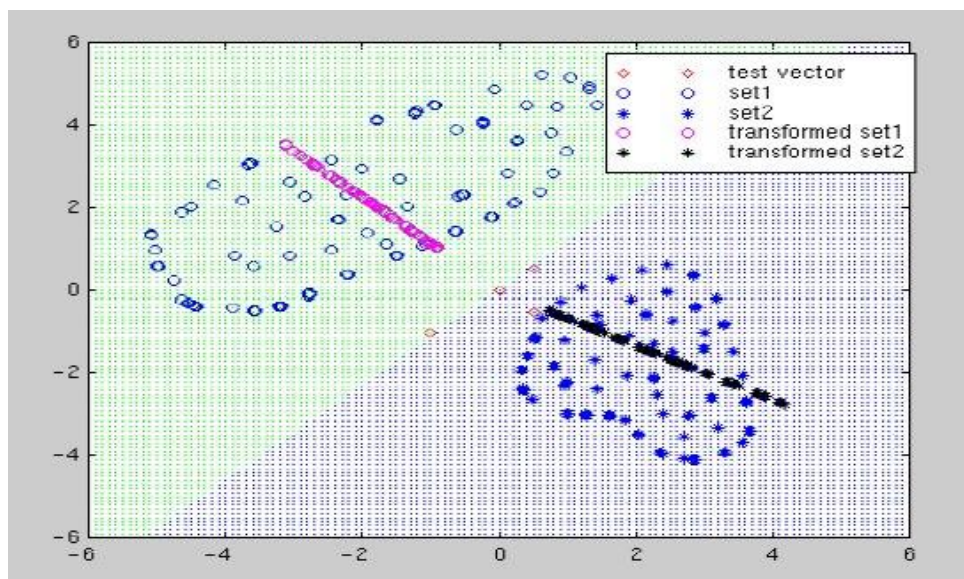
Fuente: figura por Balakrishnama y Ganapathiraju

<sup>13</sup> <http://www.dtrek.com/lda.html>

Una inspección visual muestra que los objetos de la categoría 1 (círculos abiertos) tienden a tener valores más grandes de la predicción en el eje Y y los valores más pequeños en el eje X. Sin embargo, existe una superposición entre las categorías de destino en ambos ejes, lo que se puede realizar una clasificación exacta con sólo uno de los predictores.

Análisis discriminante lineal encuentra una transformación lineal ("función discriminante") de las dos variables predictoras, X e Y, que produce un nuevo conjunto de valores transformados que proporciona una precisa discriminación, la Figura 6 muestra la nueva distribución usando la función de transformación:

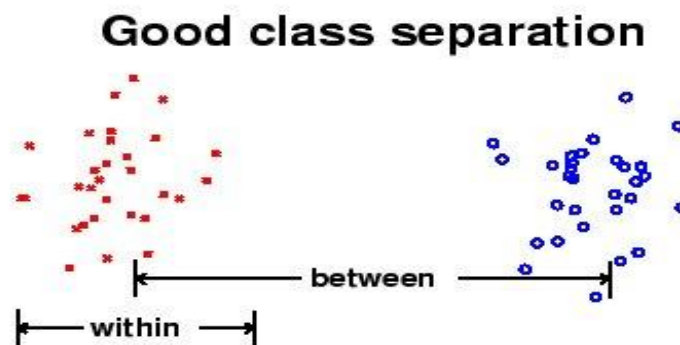
Figura 6: Nube de datos con vectores de referencia



Fuente: de Balakrishnama y Ganapathiraju

Una función de transformación se encuentra que maximiza la relación de clase entre las varianzas de las clases como se ilustra en esta figura:

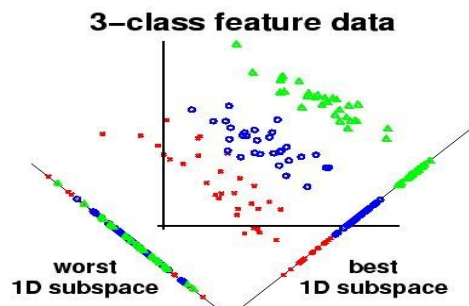
Figura 7: Máxima transformación de la nube de datos



Fuente: Ludwig Schwardt y Johan du Preez

La transformación busca girar los ejes de modo que cuando las categorías se proyectan en los nuevos ejes, las diferencias entre los grupos se maximicen. En la figura 8, se puede apreciar otro ejemplo con claridad en el cual se emplearon tres categorías, y en el que se notó la importancia de usar transformaciones de planos para encontrar la que mejor se comporte para encontrar la ortogonalidad de los grupos de datos, se muestra que a la derecha del eje menor, se alcanza la máxima separación entre las categorías mientras que la proyección de la izquierda del eje presenta los peores rendimientos para encontrar la separación de los datos.

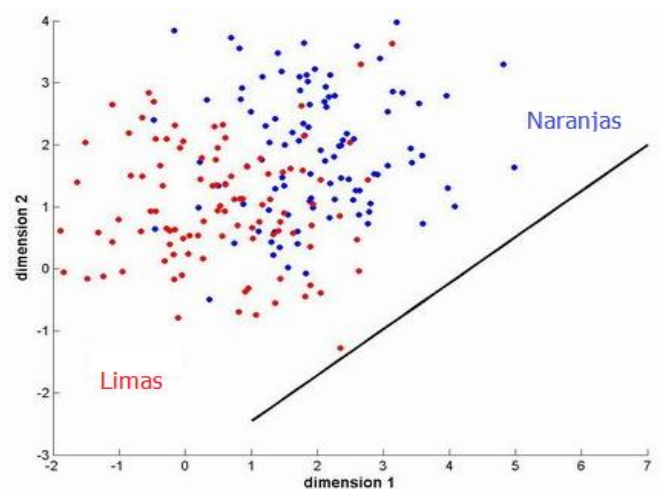
Figura 8: Transformación de grupo de tres categorías



Fuente: Ludwig Schwardt y Johan du Preez

Se aprecia que para este caso se puede generar una transformación que proporciona una ortogonalidad entre las categorías; sin embargo, en la mayoría de los casos no hay transformación que proporcione una separación completa, por lo que el objetivo es encontrar la transformación que reduce al mínimo la superposición de las distribuciones. La Figura 9 muestra una distribución de dos categorías (naranjas en azul y Limas en rojo). La línea de negro muestra el eje óptimo encontrado por LDA que maximiza la separación entre los grupos.

Figura 9: Distribución de datos

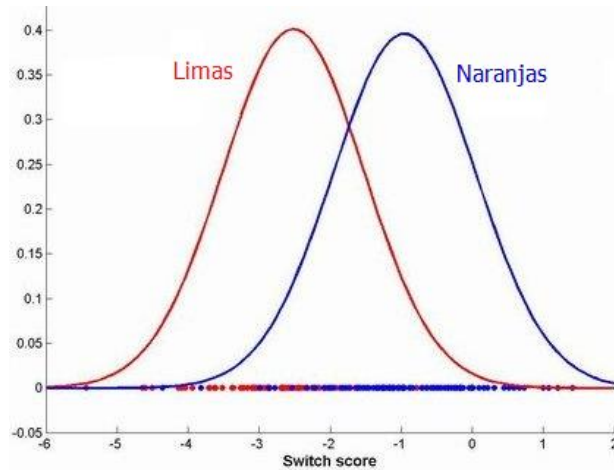


Fuente: La figura de Alex Parque y Fry Christine



En la Figura 10, se muestra la distribución las dos categorías (Naranjas y Limas) según la proyección en el eje de transformación (es decir, la línea de negro se muestra en la Figura 9):

Figura 10: Transformación para la máxima separación con traslape



Fuente: La figura de Alex Parque y Fry Christine

Tenga en cuenta que, incluso después de la transformación existe una superposición entre las categorías, es por ello que tenemos que garantizar la mayor ortogonalidad de los casos para poder conseguir elevados índices de eficiencia en este método de clasificación.

### 3. MAQUINAS DE SOPORTE VECTORIAL

Clasificar implica inevitablemente el concepto de conjunto; de modo general el proceso de clasificación radica en realizar una separación de los elementos de un conjunto  $C$  en diferentes subconjuntos  $C_i, i=1, \dots, P$ , designados clases, con base en la medición de las características que los elementos de  $C$  poseen. Una vez que se establecen las propiedades de los subconjuntos en los que se clasificará al conjunto original (modelos), los elementos de este son comparados con cada uno de los modelos, para establecer a cuál de ellos pertenecen.

Matemáticamente este proceso puede concebirse como una función que mapea el conjunto  $C$  al conjunto de clases  $\{C_i\}_{i=1}^P$ . Partimos de la hipótesis de que, sin importar la naturaleza del conjunto  $C$ , sus elementos pueden ser representados de forma numérica. Esta representación puede ser en  $\mathbb{R}^n$ , para algún  $n \in \mathbb{N}$ . Por supuesto bajo esta idea, la labor de clasificación no guarda problema alguno, esta inicia, por ejemplo, cuando se considera que los elementos del conjunto  $C$  son el resultado de un conjunto finito de variables aleatorias en  $\mathbb{R}^n$  denotado por

$$C = \{X_1, X_2, \dots, X_k\} \text{ (Ecuación 11)}$$

Donde  $X_j$  es una variable aleatoria discreta infinita o continua. Si partimos que las clases  $C_i$  son una partición de  $C$  entonces, dadas las hipótesis, el proceso de clasificación puede no ser exhaustivo, por lo que, de manera práctica,  $C$  es formado con subconjuntos de valores representativos de cada una de las variables aleatorias  $X_j$ . Así, la finalidad de una Máquina de Soporte Vectorial (Support Vector Machine, SVM)<sup>14</sup> consiste en modelar por así decirlo el comportamiento de cada una de las variables aleatorias  $X_j$ , de tal forma que se pueda establecer, dado un vector propuesto, a cuál de ellas pertenece. Para este caso en particular puede considerarse sin pérdida de generalidad, que el conjunto  $C$  está formado por dos variables aleatorias, es decir que

$$C = \{X_1, X_2\}. \text{ (Ecuación 12)}$$

Es posible representar a cada elemento del conjunto  $C$ , de la siguiente forma:

$$(x_j, y_j), j = 1, \dots, l, \text{ (Ecuación 13)}$$

---

<sup>14</sup> Vapnik incluso introduce el termino Máquinas de soporte vectorial conocido por sus siglas en ingles SVM, por el de algoritmos de aprendizaje de el tipo de soporte vectorial

Donde  $x_j \in \mathbb{R}^n$ ,  $y_j \in \{-1, 1\}$  y  $l$  es la cardinalidad de  $C$ . Se presume que se toma una muestra representativa de cada una de las variables aleatorias, la representación dada previamente permite establecer al conjunto  $C$  de la siguiente forma:

$$\begin{aligned} C &= C_1 \cup C_2, & C_1 \cap C_2 &= \emptyset, \\ C_1 &= \{x_1, \dots, x_k\}, & & \text{(Ecuación 14)} \\ C_2 &= \{x_{k+1}, \dots, x_l\}. \end{aligned}$$

Por las consideraciones hechas, la distribución de los puntos que conforman a  $C_1$  y  $C_2$  es desconocida a priori. Podemos entonces considerar, de forma general, dos casos; cuando  $C_1$  y  $C_2$  son linealmente separables y cuando no lo son<sup>15 16</sup>.

### 3.1 CONJUNTOS SEPARABLES LINEALMENTE

Se dice que  $C_1$  y  $C_2$  son linealmente separables cuando existe un hiperplano en  $\mathbb{R}^n$  determinado por un vector  $w$  perpendicular al mismo de forma que

$$w \cdot x + b = 0, \quad b \in \mathbb{R}, \quad \text{(Ecuación 15)}$$

Para cualquier punto  $x$  en el hiperplano y además

$$w \cdot x_i + b > 0, \quad \forall x_i \in C_1, \quad \text{(Ecuación 16)}$$

$$w \cdot x_j + b < 0, \quad \forall x_j \in C_2. \quad \text{(Ecuación 17)}$$

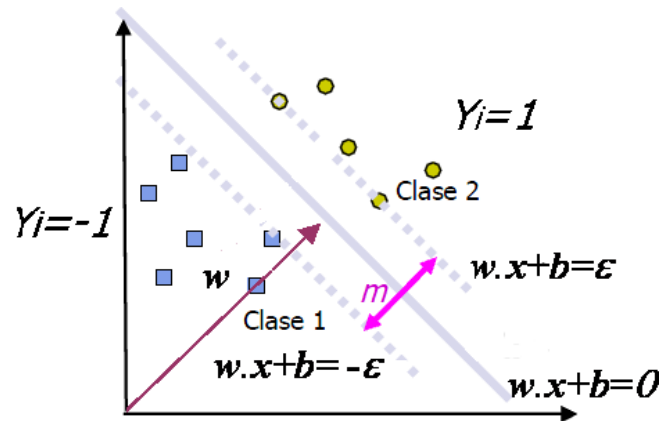
Puede conseguirse fácilmente que si  $C_1$  y  $C_2$  son linealmente separables entonces la existencia de un hiperplano tal, determinado por un vector  $w$ , no es única, de hecho existen una infinidad de tales vectores. Así, es necesario establecer un criterio que permita establecer cual de ellos se tomaría para la clasificación. También es inmediato que si ambos conjuntos son separables entonces existe una distancia mínima entre ambos conjuntos.

---

<sup>15</sup> Si se desea una mayor información sobre el tema desarrollado consultar *Learnig kernel Classifiers Theory and Algorithms*, Ralf Herbrich, 2002

<sup>16</sup> Si se desea una mayor información sobre el tema desarrollado consultar *An Introduction to Support Vector Machines and Other Kernel-Based learning methods*, Nello Cristianini y John Shawe-Taylor, 2006

Figura 11: Esquema de dos conjuntos de vectores linealmente separados mediante un hiperplano que maximiza el margen  $m$ .



Fuente: Las Máquinas de Soporte Vectorial (SVM), Gustavo A. Betancourt, 2005

De la figura anterior, se puede observar la construcción de dos hiperplanos paralelos al original, determinado este por  $w$ , los cuales delimitan un margen entre los conjuntos y cuya magnitud  $m$  se puede relacionar con dichos hiperplanos de la siguiente forma: Se puede suponer que  $w$  está contenido en el hiperplano inferior entonces sucede que

$$w \cdot w + b = -\varepsilon, \quad (\text{Ecuación 18})$$

Para alguna constante  $\varepsilon > 0$ . Entonces el vector dado por

$$w + \frac{w}{\|w\|} m, \quad (\text{Ecuación 19})$$

Está incluido en el hiperplano superior, lo cual implica que

$$w \cdot \left( w + \frac{w}{\|w\|} m \right) + b = \varepsilon. \quad (\text{Ecuación 20})$$

Restando las ecuaciones: (Ecuación 18) y (Ecuación 20), simplificando se obtiene

$$\|w\| m = 2\varepsilon \Rightarrow m = \frac{2\varepsilon}{\|w\|}. \quad (\text{Ecuación 21})$$

Dada la (Ecuación 21) se infiere que si se desea maximizar la magnitud del margen  $m$  es necesario minimizar la magnitud de  $w$ . Retomando nuevamente la notación  $(x_j, y_j)$  con  $y_j \in \{1, -1\}$ , para cada uno de los vectores  $x_j$  de  $C$ , entonces el problema de encontrar un hiperplano que separe a  $C_1$  y a  $C_2$ , maximizando el margen de separación entre dichos conjuntos, queda planteado como:

$$\min\{f(w) = \frac{1}{2\varepsilon} \|w\|^2\} \quad (\text{Ecuación 22})$$

$$\begin{aligned} y_i(w \cdot x_i + b) &\geq \varepsilon, \\ i &= 1, \dots, l, \\ \varepsilon &> 0, b \in \mathbb{R}. \end{aligned}$$

Este es un problema de optimización de tipo cuadrático sujeto a  $l$  restricciones en  $\mathbb{R}^n$  y cuya solución puede darse a partir del uso de la teoría de multiplicadores de Lagrange en  $n$  variables.

Denotemos las restricciones que aparecen en el problema de la (Ecuación 22), de la siguiente forma:

$$g_i(w) = y_i(w \cdot x_i + b), \quad i = 1, \dots, l, \quad (\text{Ecuación 23})$$

Entonces si denotamos por  $w_k$  a la  $k$ -ésima entrada de  $w$ . Sabemos que se cumple

$$\begin{aligned} \frac{\partial f}{\partial w_k}(w) &= \sum_{i=1}^l \alpha_i \frac{\partial g_i}{\partial w_k}(w), \quad k = 1, \dots, l && (\text{Ecuación 24}) \\ \alpha_i g_i(w) &= \varepsilon, \quad i = 1, \dots, l && (\text{Ecuación 25}) \end{aligned}$$

Para ciertas constantes  $\alpha_i \in \mathbb{R}$  a determinar. Sustituyendo en la (Ecuación 24) a la función  $f$  y a las funciones  $g_i$  se tiene el siguiente desarrollo:

$$\begin{aligned} \frac{\partial f}{\partial w_k} = \frac{1}{\varepsilon} w_k &= \sum_{i=1}^l \alpha_i \frac{\partial g_i}{\partial w_k} = \sum_{i=1}^l \alpha_i y_i x_i(k) \Rightarrow \\ &w = \varepsilon \sum_{i=1}^l \alpha_i y_i x_i. \end{aligned} \quad (\text{Ecuación 26})$$

Por otro lado de la (Ecuación 25) se tiene que

$$\frac{\partial \sum_{i=1}^l \alpha_i g_i(w)}{\partial b} = 0 \Rightarrow \sum_{i=1}^l \alpha_i y_i = 0. \quad (\text{Ecuación 27})$$

Por lo tanto la solución al sistema de la (Ecuación 22) está dada por

$$w = \varepsilon \sum_{i=1}^l \alpha_i y_i x_i \quad (\text{Ecuación 28})$$

$$\sum_{i=1}^l \alpha_i y_i = 0 \quad (\text{Ecuación 29})$$

Para obtener el valor de cada una de las constantes  $\alpha_i$  sustituimos la (Ecuación 28) en las restricciones iniciales en la (Ecuación 25) con lo que se obtiene la siguiente ecuación:

$$\alpha_j y_j (\varepsilon x_j \cdot \sum_{i=1}^l \alpha_i y_i x_i + b) = \varepsilon, \quad j = 1, \dots, l. \quad (\text{Ecuación 30})$$

Las igualdades descritas conforman un sistema lineal de  $l$  ecuaciones con incógnitas  $\alpha_i, i = 1, \dots, l$ , por lo que se tiene un sistema de la forma  $A\alpha = B$ , con  $A \in \mathbb{R}^{l \times l}$  y  $B \in \mathbb{R}^l$ , siendo  $\alpha$  el vector de incógnitas. Este sistema tiene solución única sólo si  $A$  tiene inversa. Es importante recordar que el valor  $l$  es la cardinalidad del conjunto  $C$ , por lo que la complejidad en la obtención de una solución mediante alguna implementación, dependerá del orden de dicho conjunto. Los valores de las constantes  $\alpha_i$  quedarían en términos de  $b$  y  $\varepsilon$ , de los cuales uno de ellos puede ser propuesto y el otro determinado, por ejemplo, con la (Ecuación 18). Así, el vector  $w$  dado en la (Ecuación 28) es conocido como el vector de soporte del hiperplano que separa a  $C_1$  y a  $C_2$ , de donde deriva el nombre de Máquinas de Soporte Vectorial.

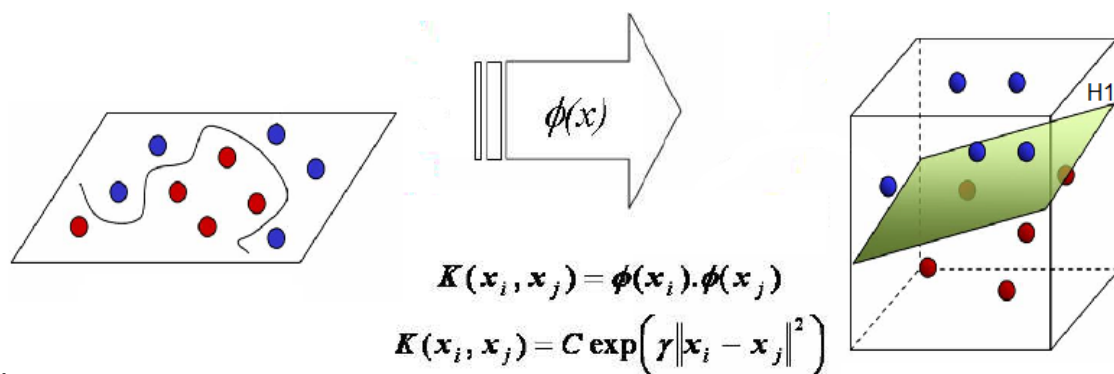
### 3.2 CONJUNTOS NO SEPARABLES LINEALMENTE

Puede considerarse que el sistema lineal de ecuaciones dado en la (Ecuación 30), es una conclusión al tratar el caso de dos conjuntos  $C_1$  y  $C_2$ , separables linealmente. Sin embargo dicho sistema puede ser obtenido también para

cualesquiera dos conjuntos en  $\mathbb{R}^n$ . Así, podemos decir que  $C_1$  y  $C_2$  no son separables linealmente cuando la matriz correspondiente al sistema de la (Ecuación 30) no tenga solución, en cuyo caso no es posible la construcción de un hiperplano que satisfaga las condiciones del problema planteado en la (Ecuación 22).

El tratamiento para el caso de dos conjuntos no separables linealmente consiste en utilizar una función  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , con  $n, m \in \mathbb{N}, m \geq n$  ó  $m > n$ . Dicha función mapea a los conjuntos  $C_1 = \{x_1, \dots, x_k\}$  y  $C_2 = \{x_{k+1}, \dots, x_l\}$ , a un espacio de mayor dimensión, donde podemos denotarlos por  $\Gamma_1 = \{\phi(x_1), \dots, \phi(x_k)\}$  y  $\Gamma_2 = \{\phi(x_{k+1}), \dots, \phi(x_l)\}$  respectivamente. El objetivo de realizar este mapeo es que los conjuntos obtenidos  $\Gamma_1$  y  $\Gamma_2$  sean separables linealmente o en su caso se minimice el error mediante la separación con un hiperplano, es decir que el número de vectores clasificados incorrectamente sea mínimo.

Figura 12: Esquema de la transformación de dos conjuntos no separables linealmente mediante la función.



Fuente: Clasificación de grandes conjuntos de datos usando Máquinas de Soporte Vectorial (SVM), Dra. Xiaou Li

Como se mencionó anteriormente, idealmente la función  $\phi$  mapea los conjuntos de vectores no separables linealmente a un espacio de mayor dimensión. Estos vectores son separables linealmente y la solución al problema original mediante el procedimiento explicado en la sección 3.1, es un hiperplano  $H_1 \subset \mathbb{R}^m$ . En este caso se considera que existen vectores de los conjuntos  $\Gamma_1$  y  $\Gamma_2$  que se encuentran contenidos en  $H_1$ . A dichos vectores se les conoce como vectores de soporte. Para obtener una solución en el espacio  $\mathbb{R}^n$  original, se realiza el mapeo inverso de los vectores de soporte, los cuales determinarán las fronteras que separarán a los conjuntos  $C_1$  y  $C_2$ . Se considera que los vectores que determinan estas fronteras conforman el modelo para  $C_1$  (o equivalentemente para  $C_2$ ).

No es posible determinar a priori, dados dos conjuntos de vectores  $C_1, C_2 \subset \mathbb{R}^n$ , una función  $\phi$  que cumpla los objetivos descritos previamente, por lo que el procedimiento para determinarla no es constructivo. Por tanto el tratamiento

para este caso está basado en la realización de ensayos con funciones  $\phi$  conocidas. Así, la función  $\phi$  es de especial importancia en la solución del problema de clasificación.

Del procedimiento dado en 3.1, puede observarse que las operaciones con vectores involucran el producto punto o producto interno canónico en  $\mathbb{R}^n$ . Este proporciona una función que determina una norma y a su vez una métrica para el espacio:

$$\|x\| = \sqrt{x \cdot x}, \quad (\text{Ecuación 31})$$

$$d(x, y) = \|x - y\|. \quad (\text{Ecuación 32})$$

Tales normas y métricas, respectivamente, son empleadas también al separar los conjuntos  $\Gamma_1$  y  $\Gamma_2$  en  $\mathbb{R}^m$ . Así que el problema análogo a la (Ecuación 22), planteado en este nuevo espacio es

$$\min\{f(w) = \frac{1}{2\varepsilon} \|w\|^2, \quad w \in \mathbb{R}^m\} \quad (\text{Ecuación 33})$$

$$\begin{aligned} y_i(w \cdot \phi(x_i) + b) &\geq \varepsilon, \\ i &= 1, \dots, l, \\ \varepsilon &> 0, \quad b \in \mathbb{R}. \end{aligned}$$

Por lo anterior la ecuación análoga a (Ecuación 31) en este nuevo espacio está dada por

$$\alpha_j \|\phi(x_j)\|^2 + y_j \sum_{i=1}^l \alpha_i y_i \phi(x_i) \cdot \phi(x_j) = -\frac{y_j b}{\varepsilon}, \quad j = 1, \dots, l. \quad (\text{Ecuación 34})$$

Se infiere de la (Ecuación 34), que la función puede ser vista como una que modifica la norma y métrica del espacio original (dadas en la (Ecuación 31) y (Ecuación 32)), por las siguientes:

$$\|x\|_\phi = \sqrt{\phi(x) \cdot \phi(x)}, \quad (\text{Ecuación 35})$$

$$d_\phi(x, y) = \|x - y\|_\phi. \quad (\text{Ecuación 36})$$



A la función definida por

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j), \quad (\text{Ecuación 37})$$

se le conoce como función núcleo y su uso es más importante que el de la propia función  $\phi$ , de la cual no se requiere su conocimiento en forma explícita, ya que es suficiente, como lo muestra la (Ecuación 34), con establecer la función núcleo K para obtener una solución.

Algunos ejemplos de funciones núcleo que han sido sugeridas o empleadas en problemas de clasificación son las siguientes <sup>17</sup> :

Lineal:

$$K(x_i, x_j) = x_i \cdot x_j. \quad (\text{Ecuación 38})$$

Polinomial:

$$K(x_i, x_j) = (\gamma x_i \cdot x_j + r)^d, \quad \gamma > 0, r, d \in \mathbb{R}. \quad (\text{Ecuación 39})$$

Función de Base Radial: (*Radial Basis Function, RBF*)

$$K(x_i, x_j) = c \exp(-\gamma \|x_i - x_j\|^2), \quad \gamma > 0, c \in \mathbb{R}. \quad (\text{Ecuación 40})$$

Sigmoide:

$$K(x_i, x_j) = \tanh(\gamma x_i \cdot x_j + r), \quad \gamma, r \in \mathbb{R}. \quad (\text{Ecuación 41})$$

Aún cuando existen diferentes funciones núcleo, es común el uso de la Función de Base Radial, por los resultados obtenidos durante la clasificación<sup>18 19</sup>. Sin embargo puede optarse por el uso de otros núcleos dependiendo de los resultados obtenidos para un caso particular.

Es importante mencionar que dado el origen y la naturaleza de los vectores obtenidos de los perfiles de pozo abierto, se espera tratar con un caso de conjuntos no separables linealmente, por lo que en este trabajo se propone utilizar inicialmente la función de base radial (dada en la (Ecuación 40)), para realizar la clasificación.

---

<sup>17</sup> H. Chi-Wei, Ch. Chih-Chung, L. Chih-Jen, A Practical Guide to Support Vector Classification

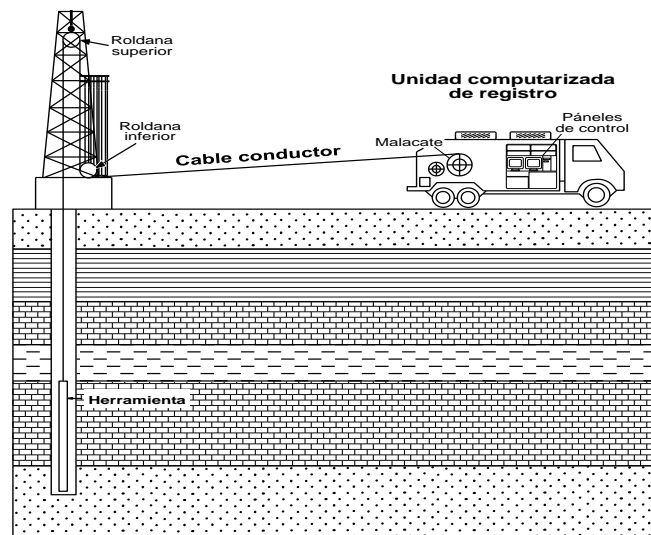
<sup>18</sup> C.J.C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition, Microsoft Research, Data Meaning and Knowledge Discovery 2, pp. 121-167, 1998.

<sup>19</sup> H. Chi-Wei, Ch. Chih-Chung, L. Chih-Jen, A Practical Guide to Support Vector Classification

## 4. PERFILAJE DE POZOS

El perfilaje de pozo consiste en la bajada de una sonda (herramienta), hasta el fondo, por medio de un cable conductor que la vincula con el equipo de superficie, el cual procesa la información enviada por la sonda y se produce el perfil del pozo en tiempo real, como se muestra en la Figura 13. La sonda en movimiento de abajo hacia arriba va recogiendo información de las paredes del pozo, que luego se envía por telemetría hacia la computadora del equipo de superficie. Un pozo en perforación es un sistema dinámico, en el sentido que el lodo de perforación utilizado en el pozo invade las rocas porosas permeables en las inmediaciones de las paredes del hueco y en consecuencia afecta, en algún grado, las mediciones de todas las herramientas que se bajan al pozo, y a su vez, las propiedades de las rocas interceptadas afectan el movimiento y las propiedades del lodo de perforación dentro del pozo<sup>20</sup>.

Figura 13: Esquema ilustrativo del perfilaje de un pozo



Fuente: Introducción a la interpretación de perfiles de pozo abierto, Ing. Jorge Arturo Camargo Puerto

### 4.1 PROPIEDADES PETROFÍSICAS DE LAS ZONAS PRODUCTORAS DE HIDROCARBUROS

Las propiedades petrofísicas que se relacionan directa o indirectamente con las mediciones realizadas con las herramientas de perfilaje son las siguientes:

- Porosidad
- Permeabilidad
- Saturación de agua
- Resistividad

<sup>20</sup> Introducción a la interpretación de perfiles de pozo abierto, Ing. Jorge Arturo Camargo Puerto

**4.1.1 Porosidad.** La porosidad,  $\Phi$ , es una propiedad de una roca reservorio de almacenar fluidos, la porosidad en las rocas sedimentarias obedece a la contextura de la roca, la cual a su vez, es función de varios factores como forma, orientación, grado de selección, forma de empaque de los granos, grado de compactación y por la forma de distribución del cemento dentro de los poros. En términos cualitativos la porosidad en reservorios detríticos se clasifica como se muestra en el Cuadro 1:

Cuadro 1: Porosidad de reservorios detríticos

$\Phi$ (%)	CLASIFICACIÓN
0 – 5	Despreciable
5 – 10	Baja
10 – 15	Regular
15 – 20	Buena
> 20	Muy buena

Fuente: Introducción a la interpretación de perfiles de pozo abierto, Ing. Jorge Arturo Camargo Puerto

**4.1.2 Permeabilidad.** La permeabilidad es la capacidad de una roca reservorio de acceder al paso de los fluidos; Cualitativamente la permeabilidad de los reservorios de petróleo y gas se clasifica, como se muestra en el Cuadro 2:

Cuadro 2: Permeabilidad en reservorios de petróleo y gas

K(mD)	CLASIFICACION
1 – 15	Mala a regular
15 – 50	Moderada
50 – 250	Buena
250 – 1.000	Muy buena
>1.000	Excelente

Fuente: Introducción a la interpretación de perfiles de pozo abierto, Ing. Jorge Arturo Camargo Puerto

Frecuentemente la permeabilidad crece con la porosidad, sin embargo, existen rocas escasamente porosas con una gran permeabilidad y viceversa rocas de alta porosidad y poca permeabilidad, como es el caso de las rocas arcillosas, que son impermeables.

**4.1.3 Resistividad.** La resistividad es la resistencia que ofrece un material al flujo de la corriente eléctrica, la cual es una propiedad inherente a todos los materiales, independientemente de su forma y tamaño. Rocas ígneas, metamórficas sedimentarias son malos conductores de la corriente eléctrica. Por lo tanto, muestran una conductividad eléctrica muy baja, y una resistividad eléctrica muy alta. Ciertos minerales, como siempre, tienden a tener resistencia anormalmente baja con respecto a las rocas circundantes, esto hace posible que la localización de ellos se haga mediante la medición de resistividad. Las rocas reservorio son rocas sedimentarias comúnmente porosa y

permeable. La mayoría de las rocas sedimentarias porosas contienen agua en sus poros, y esta suele contener algo de sal disuelta. El grado de salinidad es muy variable. Además del agua, las rocas reservorio puede contener aceite y / o gas natural. A continuación se muestra en el cuadro 3 la resistividad de algunos elementos.

Cuadro 3: Resistividad de algunos fluidos, rocas y minerales

MATERIALES	RESISTIVIDAD ( $\Omega\cdot m$ )
Petróleo, gas y aire	$\infty$
Hielo	$10^5 - 10^8$
Agua destilada	$2 \cdot 10^5$
Agua lluvia	30-1000
Agua del mar	0.2
<b>SEDIMENTOS NO CONSOLIDADOS</b>	
Arcillas	2 - 30
Arenas (varía con la composición)	100 - 5000
<b>SEDIMENTOS CONSOLIDADOS</b>	
Shale	200 - 500
Areniscas	50 - 1000
Calizas	100 - 50000
Rocas ígneas y metamórficas	10 - 10000
<b>MINERALES CONDUCTIVOS</b>	
Metales nativos: Au, Ag, Pt	10-May
Sulfuros: galenita, pirita, calcopirita, arsenopirita	10/05/2001
Óxidos: casiterita, bornita, magnetita y grafito	10/05/2001
Otros: anhídrita y bauxita	$1 - 10^5$
<b>MINERALES NO CONDUCTORES</b>	
fCuarzo, feldespato, micas, sal gema, petróleo	$10^5 - 10^{12}$

Fuente: Introducción a la interpretación de perfiles de pozo abierto, Ing. Jorge Arturo Camargo Puerto

**4.1.4 Saturación de Agua.** La saturación de agua  $S_w$  se define como el porcentaje del volumen poral ocupado por agua. Cuando exclusivamente hay agua en los poros, en otras palabras cuando  $S_w = 100\%$ , sólo existe permeabilidad absoluta del agua  $K_a$ . La  $S_w$  hace parte de los parámetros más significativos en la interpretación de registros, porque permite valorar las reservas de hidrocarburo en las rocas reservorio.

## 4.2 PERFILES SELECCIONADOS

Para este proyecto serán utilizados 5 perfiles que coinciden en todos los pozos entregados por ECOPETROL SA., para analizar y los más dicientes y necesarios a la hora de encontrar zonas potencialmente productoras. Los perfiles utilizados para este proyecto fueron 5 los cuales son:

- Gamma Ray (GR)
- Potencial Espontáneo (SP)
- Resistencia Verdadera (RT)
- Porosidad Neutrónica (PHIN)
- Densidad de formación (ROHB)

**4.2.1 Gamma Ray (GR).** El perfil de rayos gamma es un técnica de medición natural de radiación gamma para identificar la rocas o sedimentos en un pozo. Los diferentes tipos de rocas emiten diferentes cantidades y diferentes espectros de radiación gamma . En particular, las lutitas suelen emitir más rayos gamma que otras rocas sedimentarias, como arenisca, dolomita o piedra caliza porque el potasio radiactivo es un componente común en su contenido de arcilla, y porque la capacidad de intercambio catiónico de la arcilla les ayuda en la absorción de uranio y torio, lo cual permite diferenciar las litologías porosas permeables potencialmente almacenadoras de hidrocarburos de las litologías arcillosas improductivas.

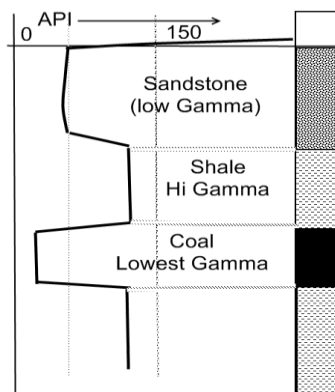
El perfil de rayos gamma mide de manera continua la radioactividad natural de las formaciones. Se lee de izquierda a derecha. Si el GR es bajo indica bajo contenido de arcilla y si es alto indica alto contenido de arcilla, como se muestra en la Figura 14. La unidad de medida es en grados API, con un rango de valores que generalmente va de 0 a 200 API. En el Cuadro 4 muestra los valores típicos de GR para algunos minerales

Cuadro 4: Valores típicos de GR

Mineral	Rayos Gamma (API)
Arenisca	25
Caliza	10
Dolomita	10
Halita	5
Anhidrita	5
Caolinita (Arcilla)	200+
Esmectita (Arcilla)	80-130

Fuente: Introducción a la interpretación de perfiles de pozo abierto, Ing. Jorge Arturo Camargo Puerto

Figura 14: Niveles de Rayos Gamma en areniscas (sandstone), arcillas (shale) y carbón (coal).



Fuente: <http://coalgeology.com>

Dentro de los usos de este perfil, se destaca:

- Revela y valora minerales radioactivos entre ellos el potasio y uranio. Los rayos gamma también se pueden utilizar para descubrir y evaluar depósitos de uranio pero no permite cuantificar la “riqueza” de los depósitos minerales. Es usado para la detección de minerales no radioactivos incluyendo estratos de carbón<sup>21</sup>.
- Establecer límites de formaciones y correlacionar litologías entre pozos<sup>22</sup> para estimar el contenido de shale o lutita en los reservorios y delimitar la calidad de la roca reservorio en los yacimientos.
- Es muy útil para casos en los cuales el registro SP no da respuesta (lodos no conductivos, pozos entubados, pozos vacíos), para la definición de estratos de lutita cuando la curva del potencial espontáneo (SP) esta redondeada en formaciones muy resistivas o poco resistivas.

**4.2.2 Potencial Espontáneo (SP).** Este registro indica las permeabilidades de las rocas en el pozo midiendo la cantidad de corriente eléctrica generada entre el fluido de perforación y el agua de la formación presente en los espacios porosos de la roca reservorio. Las Areniscas porosas con altas permeabilidades tienden a generar mayor electricidad que las Lutitas impermeables. Cuando el SP pasa frente a un nivel arcilloso, por ser éste impermeable no permite ser penetrado por lodo, por tanto no permite intercambio catiónico y no se genera diferencia de potencial. Cuando pasa frente a un nivel permeable, allí donde el lodo perdió agua que se introdujo en la formación, la curva hace una deflexión a la izquierda (negativo) cuando el agua contenida es salada y a la derecha cuando es dulce, como se muestra en la Figura 15. De este modo, el registro SP se usa a menudo para diferenciar areniscas de lutitas. Al pasar por zonas de lutita (arcilla), la respuesta de SP es algo constante, y guarda una tendencia a seguir una línea recta nombrada línea base de lutitas o Shale. Generalmente las zonas permeables están comprendidas entre capas de lutitas o shale. Además, si las resistividades del filtrado del lodo y el agua de formación son de igual valor, las variaciones del SP serán chicas y la curva será achatada sin variaciones apreciables<sup>23</sup>.

Los principales usos del perfil del potencial espontáneo SP consiste en:

- Determinar topes y bases de las rocas porosas permeables y correlacionar litologías entre pozos.

---

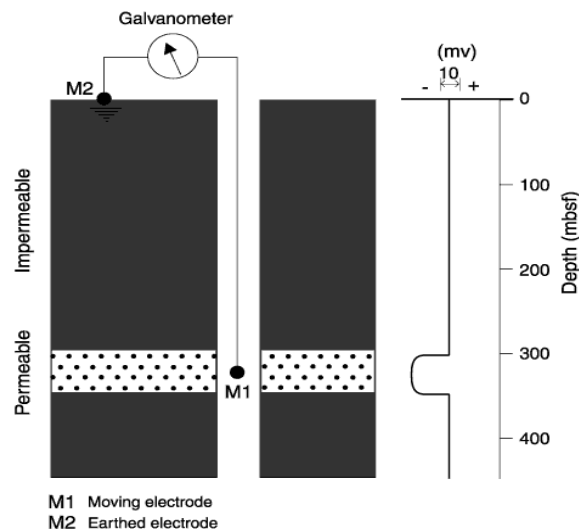
<sup>21</sup> Tixier, M. P., and Alger, R. P., Log Evaluation of Non-Metallic Mineral Deposits, 1967

<sup>22</sup> Schlumberger Production Log Interpretation, Schlumberger Limited, 1970

<sup>23</sup> Schlumberger, C and M., and Leonardon . E. G, A New Contribution to Subsurface by Means of Electrical Measurements in Drill Holes, Trans. AIME, Vol 110, 1934.

- Distinguir las litologías porosas permeables, virtualmente almacenadoras de hidrocarburos, de las litologías arcillosas improductivas.
- Apreciar la resistividad  $R_w$  de las aguas de formación. Evaluar el contenido de arcillosidad  $V_{sh}$  en las rocas reservorio.

Figura 15: Medición de un potencial natural entre un electrodo en el pozo y en la superficie de la tierra



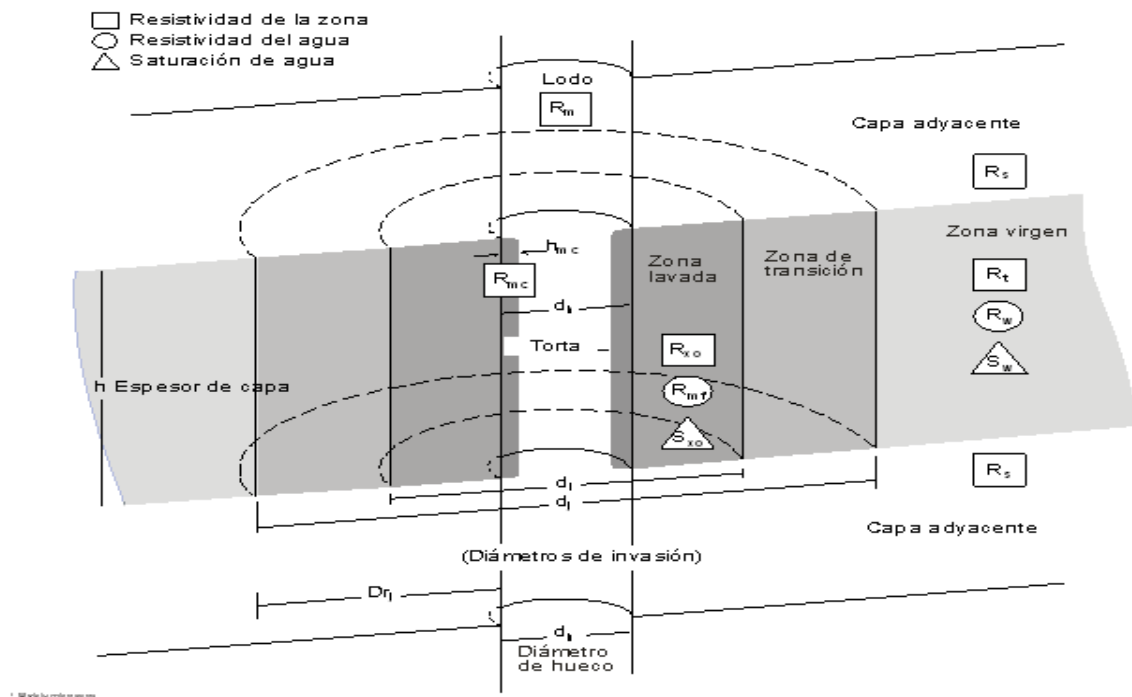
Fuente: Rider, 1996

**4.2.3 Resistividad (RT).** La resistividad de una sustancia es la capacidad de la misma para impedir el paso de corrientes eléctricas. Proporciona evidencias del contenido de fluidos en las rocas. Si los poros de una formación contienen agua salada presentará alta conductividad y por lo tanto la resistividad será baja, pero si están llenos de petróleo o gas presentará baja conductividad y por lo tanto la resistividad será alta. En una formación la capacidad de conducir electricidad está directamente relacionada a:

- La resistividad del agua en formación.
- La cantidad de agua presente
- La geometría estructural y tamaño de los poros.

En los instrumentos eléctricos, la corriente fluye radialmente hacia la formación y atraviesa zonas de diferente resistividad ( $R_m$ ,  $R_{xo}$  y  $R_t$ ) que actúan en serie y por lo tanto las superficies equipotenciales dejan de ser esféricas y se prolongan en la dirección del pozo, porque el lodo es más conductor, como se muestra en la Figura 16.

Figura 16: Diferentes términos utilizados para designar la resistividad de las diferentes zonas de invasión en un pozo



Fuente: Tomado de Schlumberger, 2000

Dentro del registro de perfil de resistividad, se halla un parámetro de gran utilidad en la estimación de hidrocarburos y tiene que ver con la resistividad verdadera ( $R_t$ ), resistividad de la zona no invadida por los fluidos de perforación, parámetro que está estrechamente unido a la saturación del hidrocarburo. El perfil de resistividad además registra otro parámetro, resistividad de la zona lavada ( $R_{xo}$ ). Si se confrontan estos dos parámetros se puede tener una clara idea acerca de la movilidad del hidrocarburo. Otra medida de gran relevancia es ( $R_s$ ), indicador de la resistividad de la zona de transición o de profundidad media, zona situada entre la zona virgen y la zona invadida. Los usos que se le dan al perfil de resistividad son:

- Diferenciar los fluidos presentes en los poros para así distinguir zonas con hidrocarburos de zonas acuíferas.
- Determinar porosidad y Medir  $R_t$  y  $R_{xo}$
- Establecer zonas permeables

**4.2.4 Porosidad Neutrónica (PHIN).** Las herramientas neutrónicas utilizadas para el perfil neutrónico se convirtieron en los primeros instrumentos en utilizar fuentes radioactivas para evaluar la porosidad de las formaciones en un pozo. Posteriormente, cuando surgió la herramienta de densidad de formación, la herramienta neutrónica fue utilizada a la par con esta para comparar y corregir



las mediciones de la porosidad, conseguidas en reservorios que contienen gas o shale. El perfil neutrónico se fundamenta en la medición de concentraciones de hidrógenos, lo que posteriormente nos indicará la presencia de petróleo o agua de la roca. Tiene una fuente de neutrones, los cuales chocan con los hidrógenos que se encuentran en los poros de la roca. La herramienta también posee un receptor que mide los neutrones dispersos liberados en las colisiones. Entre más grande sea el valor de la medida de la herramienta neutrónica mayor es la porosidad, en consecuencia se puede establecer una relación directa entre la lectura de la herramienta y la porosidad. Se lee de derecha a izquierda. La unidad de medida es en fracción o en %, con un rango de valores que va desde – 0.15 a 0.45 (–15 a 45 %). El perfil neutrónico se puede correr en cualquier tipo de pozo, abierto o revestido, lleno con aire o con lodo ya sea base agua o base aceite.

El perfil neutrónico se usa para:

- Determinar porosidad
- Identificar zonas con gas cuando se corre en combinación con el perfil de densidad
- Determinar el volumen de shale cuando se corre en combinación con el perfil de densidad
- Evaluar litologías complejas en combinación con los perfiles de densidad y sísmico.

**4.2.5 Densidad de Formación (ROHB).** El primer instrumento de densidad que utilizó el efecto de la absorción y propagación de los rayos gamma. Esta herramienta inicialmente hacia mediciones de la densidad de las rocas subterráneas, concebida para ser soporte a los geofísicos en la interpretación de la información gravimétrica. Hoy en día esta herramienta esta diseñada para determinar porosidad y adicionalmente identificar litología. El aparato de densidad de formación (RHOB, Bulk Density,  $\rho_b$ ), se basa en la medición de la densidad de la formación, por medio de la atenuación de rayos gamma entre una fuente y un receptor. Posee una fuente de rayos gamma, los cuales colisionan con los átomos presentes en la roca. La herramienta también posee un receptor que mide los rayos gamma dispersos liberados en las colisiones. Sirve para estimar la densidad del sistema roca –fluido (RHOB) que posteriormente servirá para calcular la porosidad por densidad (DPHI). Si el registro de densidad es bajo indica alta porosidad y si es alto indica baja porosidad. Se lee de izquierda a derecha. La unidad de medida es  $\text{gr}/\text{cm}^3$ , con un rango de valores que va desde 1.96 a 2.96  $\text{gr}/\text{cm}^3$ . A continuación se mostrará los valores típicos de  $\rho_b$  en el Cuadro 5.

Cuadro 5: Valores típicos de  $\rho_b$

Mineral	Perfil de Densidad $\rho_b$ (gr/m <sup>3</sup> )
Silvinita	1.98
Halita	2.16
Yeso	2.32
Arenisca	2.65
Caliza	2.71
Dolomita	2.86
Anhidrita	2.97

Fuente: Introducción a la interpretación de perfiles de pozo abierto, Ing. Jorge Arturo Camargo Puerto

El instrumento del perfil de densidad de formación es usado en pozos abiertos, sin importar con que tipo de lodo. A la par con otros Instrumentos de perfilaje de porosidad, además de estimar porosidad, sirve para:

- Valorar areniscas arcillosas y litologías más complejas y así establecer que tipo de litología se presenta.
- Determinar tipo de fluidos porales

#### 4.3 PRESENTACIÓN DE LOS PERFILES

Los registros de perfilaje, utilizados para este proyecto, y suministrados por Ecopetrol S.A., están en formato de texto plano (.txt ó .LAS). Siguiendo una estandarización del Instituto Americano del Petróleo API se establecieron unas pautas adaptables a todos los perfiles de pozo abierto. El formato de encabezado tiene un espacio para la siguiente información:

- nombre de la compañía de servicio
- identificación de pozo
- localización del pozo
- referencias de profundidad
- fecha de registro
- profundidad del pozo
- profundidad de revestimiento
- diámetro y tipo del revestimiento
- diámetro de la broca
- intervalo registrado información del lodo entro otros.

El siguiente es el prototipo general de la presentación de los perfiles:

~V

VERS. 2.00: Ver 2.00

WRAP. NO:

VEND. PRIZM:GeoGraphix

~W

STRT.F 0.0000:START DEPTH

STOP.F 7857.5000:STOP DEPTH

STEP.F 0.5000:STEP

NULL. -999.2500:NULL VALUE

WELL. LL-72 :WELL

COMP. ECOPETROL:COMPANY

UWI. LLAN0072:UNIQUE WELL ID

FLD. LLANITO:FIELD

CNTY. B/BERMEJA:COUNTY

STAT. SANTANDER:STATE

CTRY. COLOMBIA:COUNTRY

PROV. SANTANDER:PROVENCE

SRVC. GEOGRAPHIX:SERVICE

PDAT. KB:PERMANENT DATUM

LMF. KB:LOG MEASURED FROM

DMF. KB:DRILLING MEASURED FROM

LOC. :LOCATION DESCRIPTION

FTG. :LOCATION FOOTAGE DESCRIPTION

LOC1. :LOCATION LINE 1

LOC2. :LOCATION LINE 2

DATE. :LOG DATE

~C

DEPT.F :DEPTH

GR. :Gamma Ray

SP.UNKNOWN :Spontaneous Potential

CALI.IN :Caliper

RS.OHMM :

RT.OHMM :Resistivity

PHIN.V/V :Neutron Porosity

RHOB.G/C3 :Bulk Density

~P

EKB. 228.0000:ELEVATION OF KELLY BUSHING

EDF. -999.2500:ELEVATION OF DRILLING FLOOR

EGL. 206.0000:ELEVATION OF GROUND LEVEL

~A

0.0000 -999.2500 -999.2500 -999.2500 -999.2500 -999.2500 -999.2500 -999.2500

0.5000 -999.2500 -999.2500 -999.2500 -999.2500 -999.2500 -999.2500 -999.2500

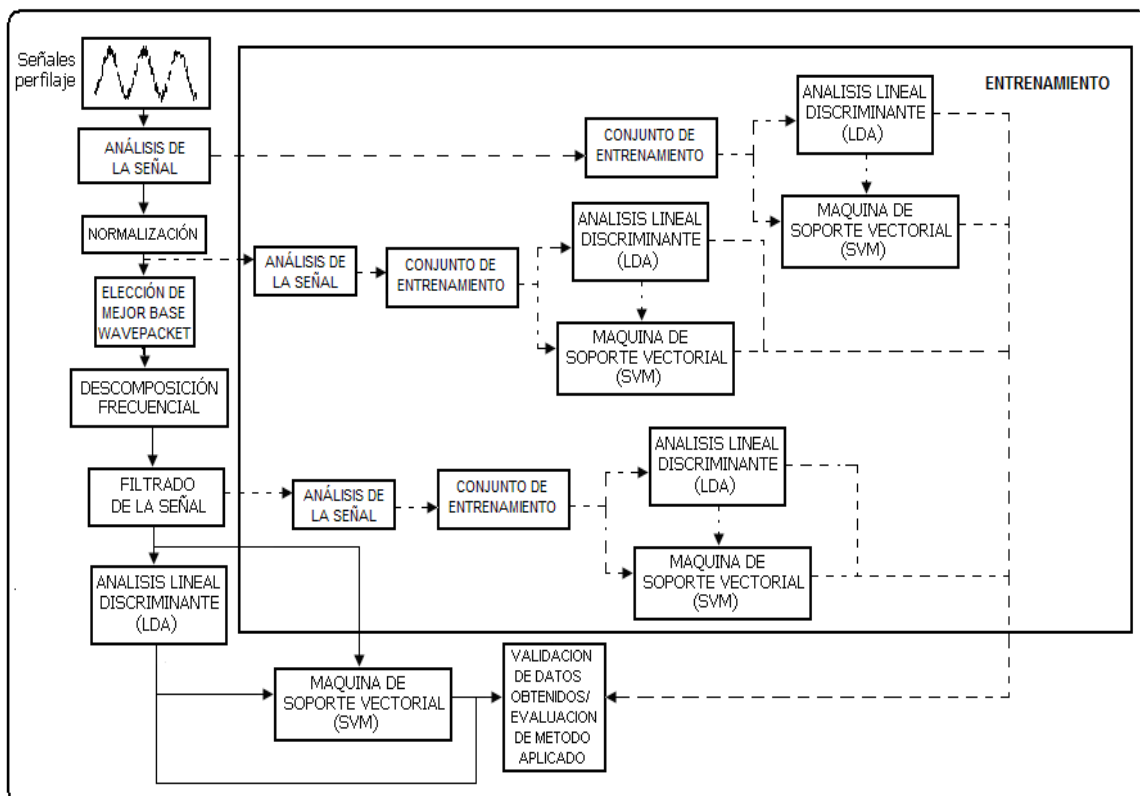
1.0000 -999.2500 -999.2500 -999.2500 -999.2500 -999.2500 -999.2500 -999.2500

## 5. DESARROLLO METODOLÓGICO

Para este proyecto, en el cual nuestro objetivo es encontrar los patrones que caracterice a cada uno de los perfiles de pozo abierto por eso es necesario un análisis estadístico de datos, para así lograr encontrar los mejores patrones que se adecuen a los diferentes perfiles de pozo abierto y lograr una buena caracterización de cada uno de ellos.

En esta sección se dan a conocer las principales pautas que se tuvieron en cuenta para la realización de este proyecto. El siguiente diagrama de bloques (Figura 17), ilustra claramente las etapas para el desarrollo del proyecto.

Figura 17: Diagrama general del sistema



Inicialmente, a las señales analizadas, correspondientes a las señales de los perfiles, se les hizo un procesamiento de la señal.

### 5.1 ANALISIS DE LA SEÑAL

Como primera medida se analizaron cada uno de los pozos y se tomó los datos de las zonas en donde se habían encontrado hidrocarburo, tomando los datos máximos y mínimos, esto con el fin de conocer los límites de las zonas potencialmente productoras. Estos datos se muestran en el cuadro 6.

Cuadro 6: Datos de máximos y mínimos de las zonas cañoneadas en los diferentes pozos analizados.

Llanos Analizados	GAMMA RAY (GR)		POTENCIAL ESPONTANEO (SP)		RESISTENCIA VERDADERA (RT)		POROSIDAD NEUTRONICA (PHIN)		BULK DENSITY (RHOB)	
	min	Max	min	max	min	max	min	max	min	max
LL-72	14.394	46.911	-7.255	-101.705	3.82	9.91	0.119	0.356	2.18	2.58
LL-77	41.879	96.761	-4.374	-87.504	3.861	13.193	0.161	0.5	2.163	2.67
LL-81	41.496	98.925	-2.566	-79.703	4.44	26.31	0.127	0.364	2.12	
LL-91	39.079	82.154	-1.6	-77.342	3.33	13.18	0.174	0.393	2.19	2.554
LL-96	28.061	132.735	-2.103	-64.437	3.754	19.338		0.38	2	2.67
LL-97	42.667	135.078	-5.016	-77.154	3.686	22.255	0.174	0.356	2.183	2.67
LL-102	42.765	96.43	-43.018	-75.253	2.857	11.891	0.22	0.415	2.169	2.467
LL-103	34.641	73.453	-4.203	-75.227	2.55	13.307	0.197	0.456	2.22	2.488
LL-104	31.913	84.319	-3.497	-87.18	2.278	13.307	0.257	0.408	2.21	2.536
LL-105	36.831	99.808	-3.658	-53.239	1.812	9.341	0.215	0.377	2.195	2.67
LL-106	39.154	116.474	0.249	-55.019	1.992	12.148	0.209	0.545	1.725	2.667
LL-107	42.97	93.17	-2.003	-66.63	5.21	27.504	0.13	0.372	2.12	2.52
Promedio	36.321	96.351	-6.587	-75.033	3.299	15.974	0.18	0.41	2.123	2.59
varianza típica	8.0148	23.4029	11.1345	13.2433	0.9839	6.0027	0.0419	0.0578	0.1325	0.0776

Analizando los datos se tomaron dos medidas estadísticas: el promedio y la varianza típica. El promedio da una idea de los valores que servirán de patrones para la matriz de entrenamiento. La varianza típica muestra una noción de que tan dispersos se encuentran esos datos, y que tan efectivos pueden ser. Como se puede apreciar en el Cuadro 6 por ser la varianza típica tan alta, los datos encontrados como promedios son alejados de algunos datos hallados en los perfiles de cada pozo, en especial el LL-72, lo cual hace casi imposible encontrar un patrón integral para todos los pozos. A continuación se mostrará los datos de los perfiles de los pozos en general, con el fin de encontrar los límites de los perfiles. Estos datos se muestran en el cuadro 7.

Cuadro 7: Datos de máximos y mínimos de los diferentes pozos analizados en general.

Llanos Analizados	GAMMA RAY (GR)		POTENCIAL ESPONTANEO (SP)		RESISTENCIA VERDADERA (RT)		POROSIDAD NEUTRONICA (PHIN)		BULK DENSITY (RHOB)	
	min	max	min	max	min	max	min	max	min	max
LL-72	12.148	109.26	2.046	-104.477	0.73	21.09	0.079	0.504	2.04	2.7
LL-77	35.093	298.597	2.889	-107.835	0.513	48.096	0.079	0.513	1.824	2.67
LL-81	27.467	355.263	6.21	-88.132	0.68	96.31	0.106	0.531	1.61	
LL-91	24.702	169.482	1.909	-100.05	0.65	19.05	0.135	0.656	1.853	2.825
LL-96	20.717	182.812	0.712	-84.448	0.885	19.338		0.583	1.698	2.67
LL-97	37.282	321.73	0.838	-87.384	0.923	42.565	0.145	0.535	2.02	2.67
LL-102	32.967	157.754	0.901	-88.208	1.068	32.724	0.179	0.553	2.111	2.672
LL-103	23.636	193.507	3.175	-87.756	1.04	24.6	0.142	0.589	1.692	2.746
LL-104	25.851	96.93	1.087	-98.934	0.899	13.37	0.199	0.568	2.146	2.582
LL-105	34.858	187.306	0.143	-57.247	0.637	25.987	0.176	0.871	1.7	2.678
LL-106	31.774	212.686	2.045	-80.325	0.694	26.463	0.172	0.863	1.536	2.709
LL-107	25.201	203.327	5.242	-95.793	0.719	118.071	0.051	0.716	1.373	2.725
Promedio	27.64133333	207.3878333	2.266416667	-90.0490833	0.7865	40.63866667	0.133	0.6235	1.80025	2.695181818
varianza típica	7.1849	79.9063	1.8602	13.3409	0.1727	32.9399	0.0483	0.1286	0.2423	0.0601

En el Cuadro 7 se encuentra una varianza típica muy alta en el perfil de Gamma

Ray (GR) y en el perfil de resistencia verdadera (RT), esto se presenta por datos anómalos al final de los pozos, en especial en estos perfiles, como se muestra en la Figura 18 y Figura 19.

Figura 18: Perfil de Resistividad Verdadera (Rt) del LL-81

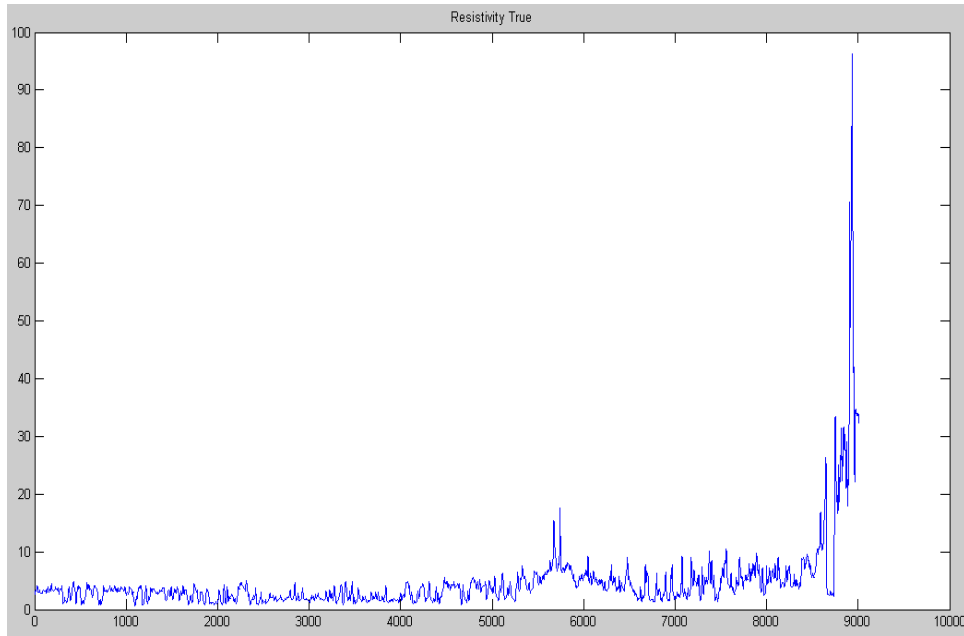
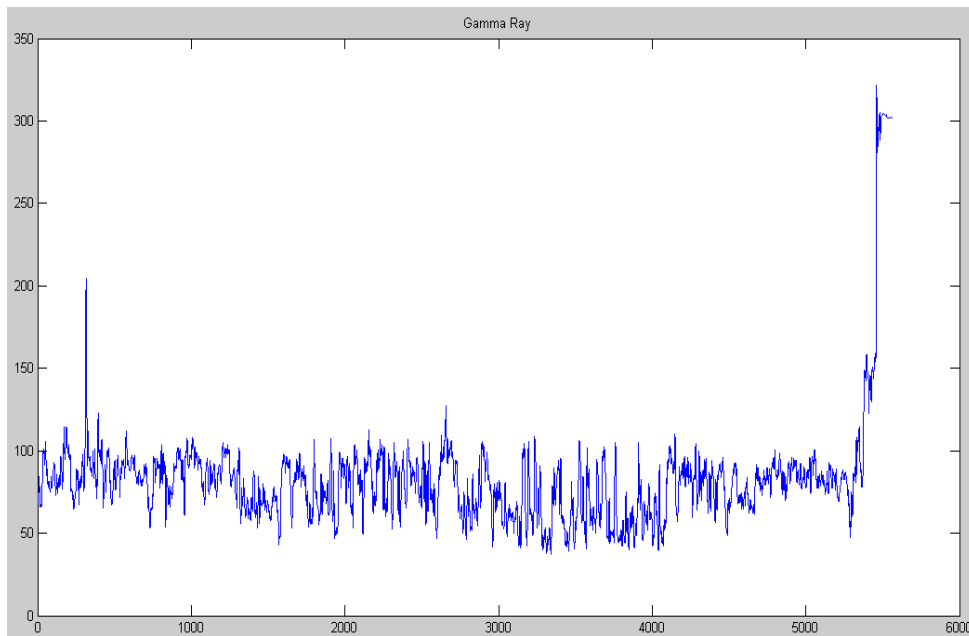


Figura 19: Perfil de Gamma Ray (GR) del LL-97



Para dar solución a este problema, de los perfiles, que se presenta en el Cuadro 7, se elimina los datos al final de cada registro o perfil, en donde empiezan aparecer estos datos, y los resultados fueron los mostrados en el cuadro 8:

Cuadro 8: Datos de máximos y mínimos de los pozos sustrayendo los datos atípicos.

Llanos Analizados	GAMMARAY (GR)		POTENCIAL ESPONTANEO (SP)		RESISTENCIA VERDADERA (RT)		POROSIDAD NEUTRONICA (PHIN)		BULK DENSITY (RHOB)	
	min	max	min	max	min	max	min	max	min	max
LL-77	12.1480	109.2600	2.8892	-107.8352	0.5130	29.7800	0.0790	0.5125	1.8240	2.6700
LL-81	35.0930	189.5000	6.2097	-88.1320	0.6800	26.3100	0.1055	0.5312	1.6100	2.7100
LL-91	27.4670	223.8000	1.9088	-100.0498	0.6500	19.0500	0.1352	0.6558	1.8526	2.8254
LL-96	24.7020	169.4820	0.7121	-84.4482	0.8850	19.3380	0.1192	0.5825	1.6980	2.6700
LL-97	20.7170	182.8120	0.8379	-87.3837	0.9230	22.2550	0.1450	0.5350	2.0200	2.6700
LL-102	37.2820	212.4000	0.9007	-88.2077	1.0680	32.7240	0.1791	0.5534	2.1110	2.6720
LL-103	32.9670	157.7540	3.1745	-87.7560	1.0400	13.3700	0.1415	0.5893	1.6920	2.7460
LL-104	23.6360	193.5070	1.0869	-98.9338	0.8990	13.3700	0.1993	0.5682	2.1460	2.5820
LL-105	25.8510	96.9300	0.1434	-57.2474	0.6372	25.9865	0.1757	0.8713	1.6997	2.6775
LL-106	34.8580	187.3060	2.0448	-80.3252	0.6937	24.0685	0.1718	0.8633	1.5356	2.7094
LL-107	31.7740	178.9000	5.2419	-95.7933	0.7186	30.2086	0.0514	0.7163	1.3726	2.7247
Promedio	25.2010	172.8774	2.2864	-88.7375	0.7916	23.3146	0.1366	0.6344	1.7783	2.6961
varianza típica	7.1849	37.2408	1.8588	12.5433	0.1718	6.2338	0.0431	0.1231	0.2304	0.0575

Como resultado se encuentra una reducción significativa en la varianza típica en los perfiles de Gamma Ray (GR) y en el de Resistividad Verdadera (RT), como consecuencia de la eliminación de los datos atípicos. Estos datos del promedio del Cuadro 8 son parecidos a los suministrados en el marco teórico, concerniente a valores límites de los perfiles; por lo cual se hace preciso una especie acción, la cual se llamará como de preprocesamiento, en la cual a las medidas de los perfiles serán limitadas a los valores encontrados aquí, así eliminaremos a futuro estos datos inusuales, como se muestra en la Figura 20 y en la Figura 21. Además se hace un cambio en el perfil de Potencial Espontáneo (SP), para así garantizar que todos los datos sean negativos y luego hacemos estos valores como valores absolutos, así todos los datos serán positivos. A continuación se mostrará los datos utilizados para este preprocesamiento en el cuadro 9.

Cuadro 9: Datos de los límites utilizados en el preprocesamiento.

Perfil	Min	Max
Gamma Ray (GR)	0	200
Potencial Espontaneo (SP)	0	120
Resistencia Verdadera (RT)	0	35
Porosidad Neutronica (Phin)	0.08	0.65
Bulk Density (ROHB)	1.5	2.96

Figura 20: Perfil preprocesado de Gamma Ray del LL-97

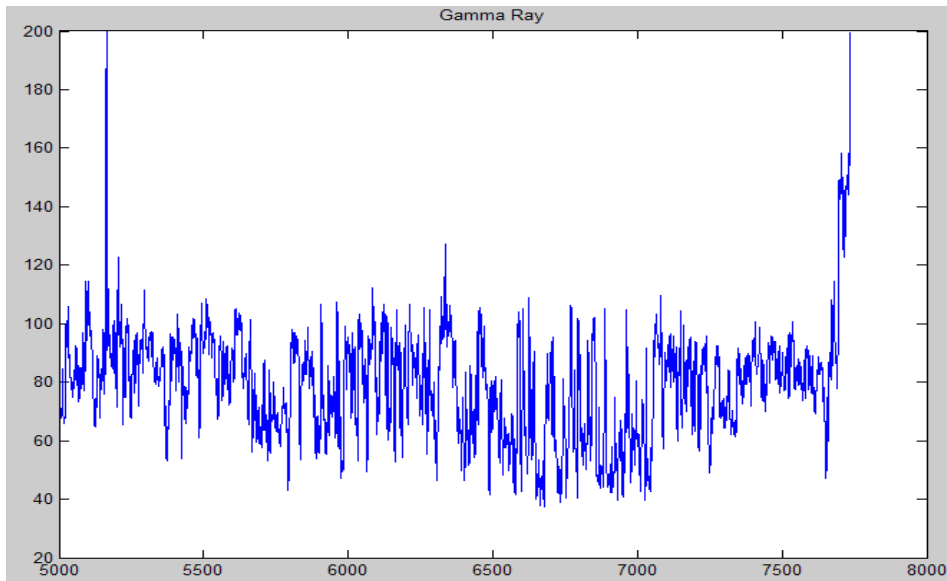
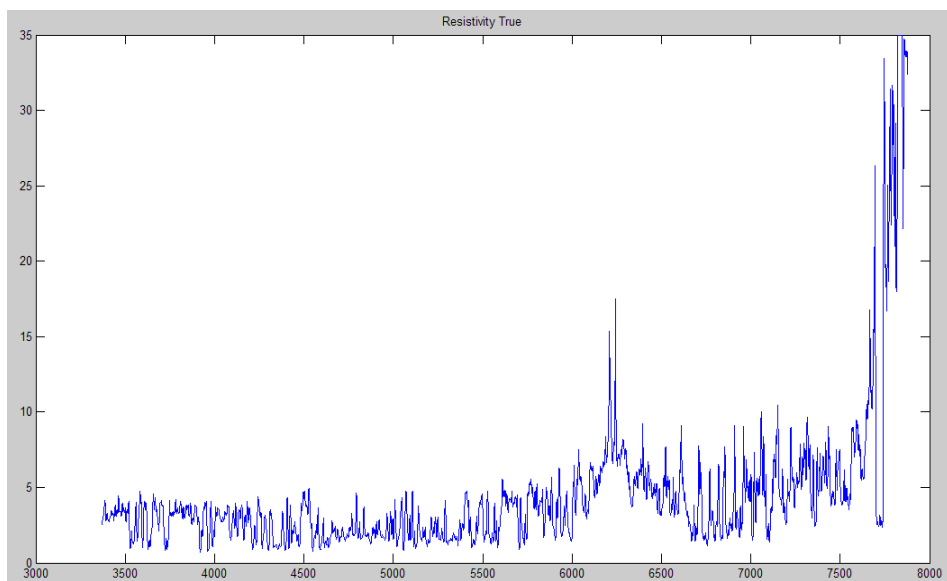


Figura 21: Perfil preprocesado de Resistividad Verdadera del LL-81



## 5.2 NORMALIZACIÓN

Se ajustaron los datos de los perfiles a fin proporcionar un rango de valores más apropiado para los procesos de reconocimiento de patrones, ya que se encontraron mediciones dentro de los registros de perfilaje, como el perfil de Gamma Ray (GR) que presenta datos inusuales y por ende su varianza típica es demasiado alta. La ecuación 42 establece el proceso de normalización así:



$$X_n = \frac{X_0 - X_{\min}}{X_{\max} - X_{\min}} \quad (\text{Ecuación 42})$$

$X_n$ , valor normalizado de  $X_0$

$X_0$ , valor del dato a normalizar

$X_{\min}$ , mínimo valor de los datos a normalizar

$X_{\max}$ , máximo valor de los datos a normalizar

Al normalizar los datos, de nuevo se analizó los datos y se tomó los datos de las zonas en donde se cañoneo, y a continuación se mostrará los resultados en el Cuadro 10:

Cuadro 10: Datos de máximos y mínimos normalizados de las zonas cañoneadas en los diferentes pozos analizados.

Llanos Analizados	GAMMA RAY (GR)		POTENCIAL ESPONTANEO (SP)		RESISTENCIA VERDADERA (RT)		POROSIDAD NEUTRONICA (PHIN)		BULK DENSITY (RHOB)	
	min	max	min	max	min	max	min	max	min	max
LL-77	1.0411	1.3740	1.1656	1.8164	1.0971	1.3677	1.1871	1.9699	1.4007	2.0000
LL-81	1.0813	1.4142	1.1930	1.9107	1.1096	1.7468	1.0505	1.6079	1.4636	2.0000
LL-91	1.0993	1.3968	1.1944	1.7773	1.1457	1.6810	1.0747	1.4948	1.3466	1.7207
LL-96	1.0453	1.6911	1.1888	1.7650	1.0904	2.0000	1.1405	1.5325	1.0000	2.0000
LL-97	1.0331	1.6010	1.1664	1.8840	1.0811	1.6260	1.0744	1.5410	1.2508	2.0000
LL-102	1.0785	1.5086	1.1584	1.8546	1.0480	1.3419	1.1098	1.6300	1.1034	1.6506
LL-103	1.0648	1.2933	1.1811	1.8622	1.0887	1.5233	1.1248	1.7025	1.5028	1.7514
LL-104	1.0853	1.8226	1.1458	1.8825	1.1111	2.0000	1.1575	1.5657	1.1468	1.8945
LL-105	1.0129	1.4261	1.1662	1.9302	1.0463	1.3433	1.0558	1.2894	1.5064	1.9923
LL-106	1.0439	1.5035	1.1175	1.6928	1.0657	1.4445	1.0539	1.5400	1.1617	1.9663
LL-107	1.0747	1.3889	1.1717	1.8168	1.0109	1.5435	1.1188	1.4825	1.5557	1.8491
Promedio	1.0600	1.4927	1.1681	1.8357	1.0813	1.6016	1.1043	1.5778	1.3126	1.8932
varianza típica	2.6447	15.6433	2.2554	7.0496	3.7090	23.8430	4.6059	16.6448	18.9767	13.0940

Como se puede ver la varianza típica disminuye con respecto a sus antecesores, lo que nos comprueba que normalizar ayudo a la estandarización de los datos, lo que nos garantiza un mejor patrón integral para todos los pozos en general.

A continuación se gráfica el perfil de Resistividad Verdadera del LL-81, preprocesado y normalizado en la Figura 22 y el perfil de Gamma Ray del LL-97 preprocesado y normalizado en la Figura 23.

Figura 22: Perfil preprocesado y normalizado de Resistividad Verdadera del LL-81

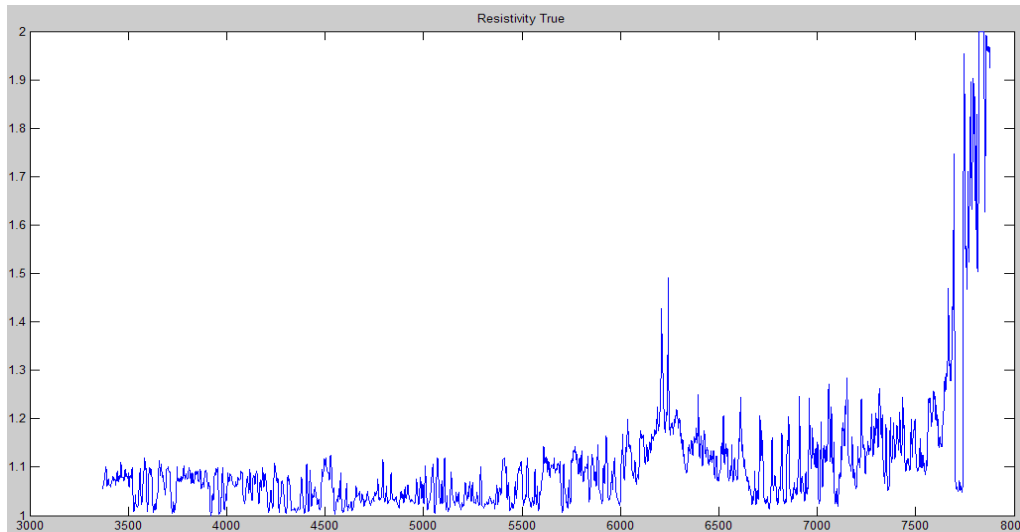
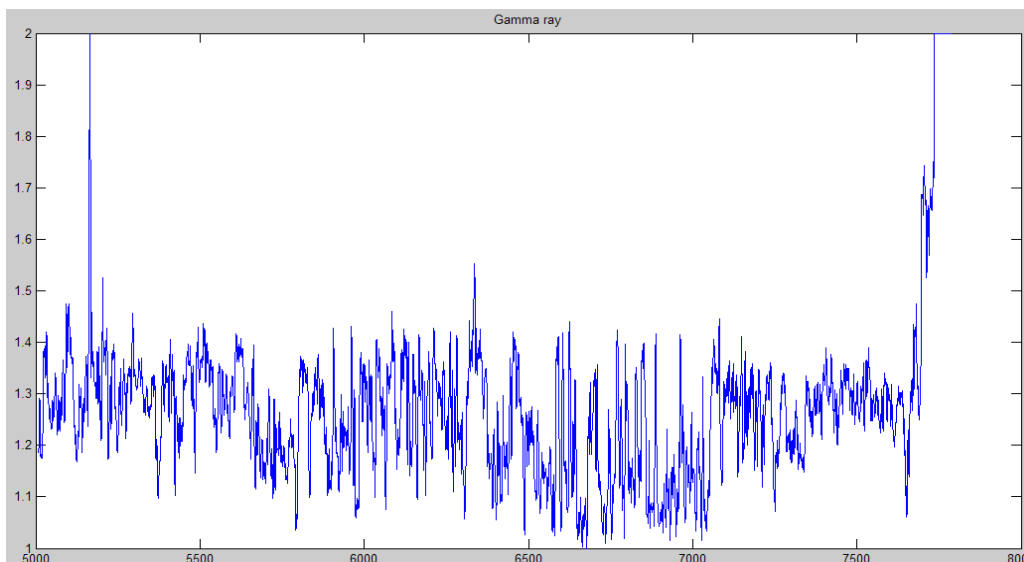


Figura 23: Perfil preprocesado y normalizado de Gamma Ray del LL-97



### 5.3 ELECCIÓN DE LA MEJOR BASE

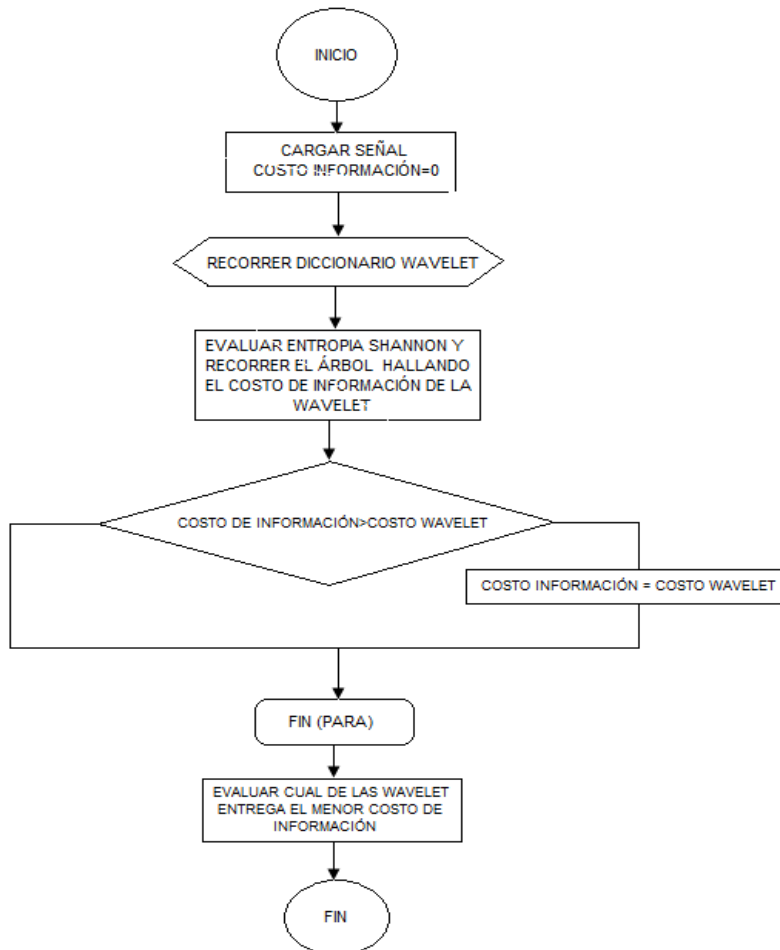
Inicialmente se trata de la exploración de la wavelet básica o madre que se adecue a cada uno de los registros de los perfiles. La elección de la *wavelet* básica o *wavelet* madre depende mucho de la naturaleza de la señal y del objetivo del procesamiento digital de esas señales<sup>24</sup>. Para la selección de la *wavelet* básica en parámetros de representación se hará uso de las medidas de información. Con este método se obtiene una mayor ventaja debido a que al

<sup>24</sup> Y. M. Stéphane Jaffard y R. D. Ryan, *Wavelets: Tools for Science and Technology*. Society for Industrial and Applied Mathematics, 2001.

momento de presentar resultados este método brinda algo más sólido sobre el qué sustentar las afirmaciones, además es menos subjetiva<sup>25</sup>.

Se valoraron todas las *wavelets* básicas, luego se eligió una enciclopedia con las *wavelets* que mejor representaban a las señales de los registros. De las 95 familias, se eligieron solo 21. Es así que se tuvo en cuenta la mejor *wavelet* básica de acuerdo al concepto de Entropía de Shannon que representara la menor cantidad de medidas de información. A cada registro se le calcula la función de costo. Aquella *wavelet* para la cual el costo fuera mínimo se establece como la más favorable para ese registro en particular. Lo mismo se hace para todos los registros de la base de datos. La cantidad de registros para los cuales la determinada *wavelet* básica de prueba diese mejor es entregada obteniéndose aquella para la cual el porcentaje es mayor. Como se muestra en la Figura 24.

Figura 24: Diagrama de flujo para elección wavelet básica



Fuente: Identificación de zonas productoras de hidrocarburos aplicando DSP y métodos de correlación de registros de perfilaje

<sup>25</sup> Identificación de zonas productoras de hidrocarburos aplicando DSP y métodos de correlación de registros de perfilaje, Alfredo R. Palomino y Hugo A. Bernal, 2009

## 5.4 REDUCCIÓN DE PERTURBACIONES MEDIANTE UMBRALIZACIÓN

Sea  $y[n]$ , la señal adquirida,  $y[n]=x[n]+\mu e[n]$  donde  $n= 1,\dots,N$ ,  $x[n]$ , es la señal sin ruido,  $e[n]$ , representa ruido blanco gaussiano con media nula y varianza 1, mientras  $\mu$  representa su nivel. El proceso genérico para reducir el ruido de una señal utilizando la WT se muestra a continuación:

- 1) Escoger una wavelet, un nivel y calcular la descomposición Wavelet de la señal  $s$  en el nivel  $J$ .
- 2) Aplicar un umbral a los coeficientes de detalle, del nivel 1 al  $J$ . El cálculo de este umbral y la forma de aplicarlo, determinará el resultado del proceso. El umbral será de la forma  $\delta = [\delta_1, \delta_2, \dots, \delta_j]$ .
- 3) Reconstruir la señal, basándose en los coeficientes de aproximación y los coeficientes modificados de detalle.

Los dos métodos a experimentar son la Umbralización rígida y Umbralización flexible, para una mejor elección del método a escoger. Se realizará un análisis de desempeño y así saber que método de umbralización escoger. Para esto se calculará el valor absoluto, la media del error y la varianza del error. Estos datos son resultados de las siguientes ecuaciones<sup>26</sup>:

$$e = | \text{Señal Filtrada} - \text{Señal Original} | \quad (\text{Ecuación 43})$$

$$\text{media del error} = \mu = \frac{1}{n} \sum_{i=1}^n e_i \quad (\text{Ecuación 44})$$

$$\text{varianza del error} = \frac{1}{n} \sum_{i=1}^n (e_i - \mu)^2 \quad (\text{Ecuación 45})$$

Para un mejor entendimiento y comparación de los datos obtenidos, se organizará la información en Cuadros, en el Cuadro 11 y Cuadro 12 se escribirán los datos de la media del error para la umbralización flexible y rígida, además en el Cuadro 13 y Cuadro 14 la varianza del error de ambos métodos, con el fin de determinar cuál método es mejor.

---

<sup>26</sup> Pulgarín, Juan., Acosta, Carlos., Castellanos, Germán., 2007. Reducción de Perturbaciones Mediante Molificación Discreta y Umbralización por Wavelets.

Cuadro 11: Análisis del desempeño, media del error de umbralización flexible

Niveles de Filtrado	Umbralización Flexible					
	LL-72	LL-77	LL-81	LL-91	LL-96	LL-97
Nivel 1	0.006	0.003	0.004	0.003	0.002	0.002
Nivel 2	0.014	0.010	0.011	0.010	0.007	0.009
Nivel 3	0.028	0.023	0.023	0.021	0.017	0.023
Nivel 4	0.046	0.039	0.037	0.036	0.035	0.042
Nivel 5	0.063	0.053	0.052	0.051	0.056	0.061
Nivel 6	0.076	0.065	0.064	0.061	0.073	0.077
Nivel 7	0.085	0.076	0.074	0.071	0.085	0.089
Nivel 8	0.090	0.080	0.080	0.077	0.091	0.094
Nivel 9	0.095	0.091	0.097	0.080	0.098	0.100
Nivel 10	0.096	0.111	0.102	0.083	0.102	0.104
	LL-102	LL-103	LL-104	LL-105	LL-106	LL-107
Nivel 1	0.002	0.002	0.002	0.006	0.007	0.004
Nivel 2	0.005	0.007	0.007	0.014	0.016	0.012
Nivel 3	0.016	0.019	0.020	0.027	0.028	0.023
Nivel 4	0.037	0.036	0.043	0.043	0.043	0.037
Nivel 5	0.060	0.051	0.072	0.060	0.058	0.051
Nivel 6	0.078	0.065	0.091	0.074	0.069	0.062
Nivel 7	0.086	0.075	0.100	0.083	0.076	0.070
Nivel 8	0.093	0.079	0.107	0.087	0.080	0.076
Nivel 9	0.097	0.086	0.110	0.091	0.084	0.082
Nivel 10	0.099	0.102	0.114	0.114	0.101	0.086

Cuadro 12: Análisis del desempeño, media del error de umbralización rígida

Niveles de Filtrado	Umbralización Rígida					
	LL-72	LL-77	LL-81	LL-91	LL-96	LL-97
Nivel 1	0.006	0.003	0.004	0.003	0.002	0.002
Nivel 2	0.014	0.010	0.011	0.010	0.007	0.009
Nivel 3	0.028	0.023	0.023	0.021	0.017	0.023
Nivel 4	0.046	0.039	0.037	0.036	0.035	0.042
Nivel 5	0.063	0.053	0.052	0.051	0.056	0.061
Nivel 6	0.077	0.065	0.065	0.062	0.073	0.077
Nivel 7	0.086	0.076	0.075	0.072	0.085	0.089
Nivel 8	0.091	0.080	0.080	0.078	0.092	0.094
Nivel 9	0.096	0.094	0.096	0.081	0.099	0.100
Nivel 10	0.098	0.103	0.099	0.085	0.104	0.106
	LL-102	LL-103	LL-104	LL-105	LL-106	LL-107
Nivel 1	0.002	0.002	0.002	0.006	0.007	0.004
Nivel 2	0.005	0.007	0.007	0.014	0.016	0.012
Nivel 3	0.016	0.019	0.020	0.027	0.028	0.023
Nivel 4	0.037	0.036	0.043	0.043	0.043	0.037
Nivel 5	0.060	0.051	0.072	0.060	0.058	0.051
Nivel 6	0.078	0.065	0.092	0.074	0.069	0.062
Nivel 7	0.086	0.075	0.101	0.084	0.076	0.070
Nivel 8	0.093	0.079	0.108	0.088	0.081	0.076
Nivel 9	0.097	0.087	0.112	0.092	0.085	0.082
Nivel 10	0.101	0.101	0.117	0.123	0.103	0.087

Cuadro 13: Análisis del desempeño, varianza del error de Umbralización flexible

Niveles de Filtrado	Umbralización Flexible					
	LL-72	LL-77	LL-81	LL-91	LL-96	LL-97
Nivel 1	1.324	1.346	1.355	1.316	1.377	1.321
Nivel 2	1.315	1.339	1.348	1.310	1.373	1.314
Nivel 3	1.301	1.326	1.337	1.298	1.362	1.299
Nivel 4	1.283	1.310	1.322	1.284	1.345	1.281
Nivel 5	1.267	1.296	1.307	1.269	1.323	1.262
Nivel 6	1.253	1.284	1.295	1.258	1.306	1.246
Nivel 7	1.244	1.273	1.285	1.248	1.295	1.234
Nivel 8	1.239	1.269	1.280	1.243	1.289	1.229
Nivel 9	1.235	1.258	1.262	1.240	1.282	1.223
Nivel 10	1.233	1.248	1.257	1.237	1.278	1.219
	LL-102	LL-103	LL-104	LL-105	LL-106	LL-107
Nivel 1	1.321	1.350	1.364	1.315	1.320	1.419
Nivel 2	1.317	1.345	1.359	1.307	1.311	1.411
Nivel 3	1.307	1.333	1.346	1.294	1.300	1.399
Nivel 4	1.286	1.316	1.322	1.278	1.284	1.386
Nivel 5	1.262	1.301	1.293	1.261	1.269	1.372
Nivel 6	1.245	1.286	1.274	1.247	1.259	1.360
Nivel 7	1.236	1.277	1.265	1.238	1.252	1.352
Nivel 8	1.230	1.272	1.259	1.234	1.247	1.346
Nivel 9	1.226	1.266	1.255	1.230	1.243	1.341
Nivel 10	1.223	1.250	1.252	1.207	1.226	1.337

Cuadro 14: Análisis del desempeño, varianza del error de Umbralización rígida

Niveles de Filtrado	Umbralización Rígida					
	LL-72	LL-77	LL-81	LL-91	LL-96	LL-97
Nivel 1	1.324	1.346	1.355	1.316	1.377	1.321
Nivel 2	1.315	1.339	1.348	1.310	1.373	1.314
Nivel 3	1.301	1.326	1.337	1.298	1.362	1.299
Nivel 4	1.283	1.310	1.322	1.284	1.345	1.281
Nivel 5	1.266	1.296	1.307	1.269	1.323	1.262
Nivel 6	1.252	1.284	1.295	1.258	1.306	1.246
Nivel 7	1.244	1.273	1.285	1.248	1.294	1.234
Nivel 8	1.238	1.269	1.279	1.242	1.288	1.228
Nivel 9	1.233	1.255	1.264	1.239	1.280	1.223
Nivel 10	1.231	1.246	1.260	1.235	1.276	1.217
	LL-102	LL-103	LL-104	LL-105	LL-106	LL-107
Nivel 1	1.321	1.350	1.364	1.315	1.320	1.419
Nivel 2	1.317	1.345	1.359	1.307	1.311	1.411
Nivel 3	1.307	1.333	1.346	1.294	1.300	1.399
Nivel 4	1.286	1.316	1.322	1.278	1.284	1.386
Nivel 5	1.262	1.301	1.293	1.261	1.269	1.372
Nivel 6	1.245	1.286	1.274	1.247	1.259	1.360
Nivel 7	1.236	1.277	1.264	1.237	1.252	1.352
Nivel 8	1.230	1.272	1.258	1.233	1.246	1.346
Nivel 9	1.225	1.265	1.254	1.228	1.242	1.341
Nivel 10	1.222	1.250	1.248	1.198	1.224	1.336

Se Puede observar en los anteriores Cuadros que los dos métodos tienen desempeño de errores muy parecidos, lo cual hace que en la práctica sea difícil determinar que método es mejor. Para esto se la decisión se fundamentará en la teoría, que indica que para este tipo de ondas, es mejor por continuidad, de la onda resultante el método de umbralización flexible<sup>27</sup>. Habiendo determinado el método falta determinar el nivel de filtrado; como los datos son normalizados, estos datos son porcentajes, los resultados revelan que es conveniente un filtrado de nivel 3, lo cual mantiene el error por debajo de 3%.

Al filtrar los datos, de nuevo se analizó los datos y se tomó los datos de las zonas en donde se cañoneo, y a continuación se mostrará los resultados en el Cuadro 15:

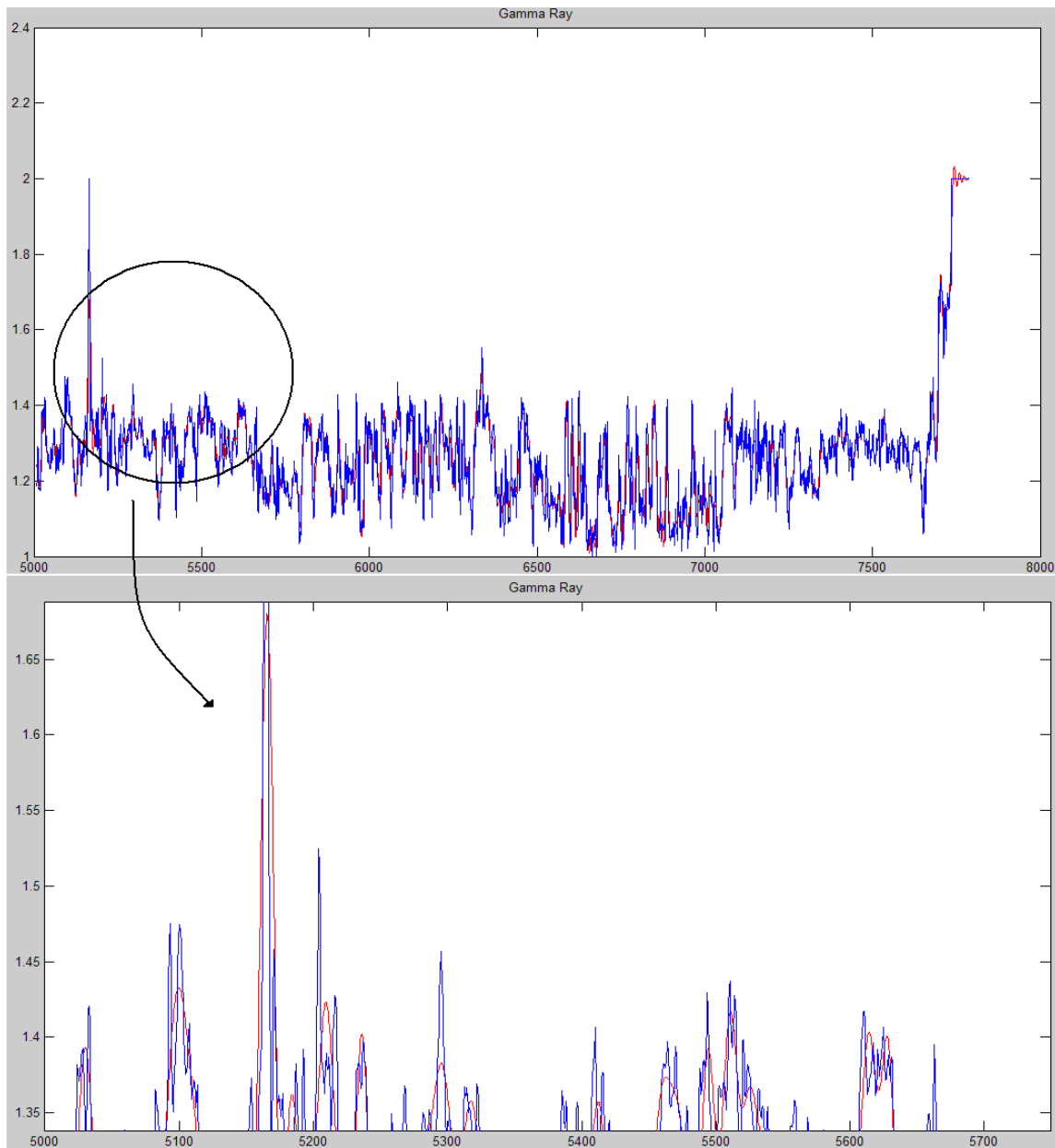
Cuadro 15: Datos de máximos y mínimos filtrados a nivel 3 de las zonas cañoneadas en los diferentes pozos analizados.

Llanos	GR		SP		RT		POROSIDAD		BULK DENSITY	
	min	max	min	max	min	max	min	max	min	max
LL-72	1.0260	1.3729	1.2063	2.0000	1.1548	1.3905	1.1336	1.6225	1.2611	1.6880
LL-77	1.0336	1.4000	1.2397	1.8252	1.1019	1.3504	1.2411	1.8737	1.4947	2.0000
LL-81	1.1013	1.3738	1.1948	1.9115	1.1158	1.7105	1.1038	1.6216	1.5749	1.9031
LL-91	1.0993	1.3968	1.1944	1.7773	1.1457	1.6810	1.0747	1.4948	1.3466	1.7207
LL-96	1.0471	1.5425	1.1993	1.7160	1.0934	1.9300	1.1265	1.5208	1.4677	2.0000
LL-97	1.0146	1.5507	1.2204	1.8865	1.0802	1.5828	1.0646	1.5304	1.2397	1.9001
LL-102	1.1124	1.4845	1.1755	1.8606	1.0443	1.3401	1.1272	1.5873	1.1036	1.6182
LL-103	1.0775	1.2838	1.1757	1.8516	1.0880	1.4947	1.1080	1.6067	1.4932	1.7404
LL-104	1.1489	1.7374	1.2142	1.8784	1.1142	1.9953	1.1473	1.4974	1.1646	1.8198
LL-105	1.0338	1.5372	1.2085	1.9395	1.0365	1.3234	1.1010	1.2599	1.4947	1.7980
LL-106	1.1010	1.4103	1.1710	1.7178	1.0477	1.3682	1.0859	1.4298	1.5691	1.8920
LL-107	1.1018	1.3354	1.1976	1.7275	1.0661	1.4578	1.1070	1.5120	1.5952	1.8495
Promedio	1.0748	1.4521	1.1998	1.8410	1.0907	1.5521	1.1184	1.5464	1.4004	1.8275
varianza típica	4.2406	12.4115	2.0004	9.1409	3.8271	23.1975	4.5606	14.3564	17.0454	11.9974

A continuación se mostrará una gráfica del perfil de Gamma Ray (GR) del LL-97, filtrado a nivel tres y sin filtrar, en la Figura 25.

<sup>27</sup> M. Jansen, *Noise Reduction by Wavelet Thresholding*, Springer 2001.

Figura 25: Perfil preprocesado, normalizado y filtrado nivel 3 de Gamma Ray (GR) del LL-97



### 5.5 SELECCIÓN DE LA POBLACIÓN DE ENTRENAMIENTO

Para la selección de la población de entrenamiento tanto para Análisis lineal Discriminante (LDA) y Máquina de Soporte Vectorial (SVM), se tomó los datos de los diferentes análisis que se han hecho a los perfiles a lo largo de todo este proceso desarrollado en esta tesis. Por esto último se hizo tres tipos de entrenamientos, siendo el último el escogido como ideal por los resultados obtenidos en cuanto a porcentaje de efectividad y porcentaje de error. Los tres tipos de entrenamientos enunciados anteriormente son los siguientes:



- Entrenamiento con Datos Preprocesados.
- Entrenamiento con Datos Preprocesados y Normalizados.
- Entrenamiento con Datos Preprocesados, Normalizados y Filtrados a nivel 3.

**5.5.1 Entrenamiento con Datos Preprocesados.** Para crear una matriz de entrenamiento de datos preprocesados se tomará los datos de los promedios del Cuadro 6 y los límites de los diferentes perfiles del Cuadro 8. A continuación se mostrará los datos de la matriz de entrenamiento en el Cuadro 16.

Cuadro 16: Comparativo de datos preprocesados

Datos preprocesados		GR		SP		RT		PHIN		RHOB	
		Valor discreto	Valor en %	Valor discreto	Valor en %	Valor discreto	Valor en %	Valor discreto	Valor en %	Valor discreto	Valor en %
Zona inferior improductora	min	0	0	0	0	0	0	0.08	0	1.5	0
	max	36.2707	5.9145	16.5869	9.7112	3.2991	11.5649	0.1802	10.4457	2.1228	32.9716
Zona productora	min	36.2708	5.9146	16.5870	9.7113	3.2992	11.5650	0.1803	10.4458	2.1229	32.9717
	max	96.3516	48.7014	85.0328	83.8414	15.9737	72.0292	0.4102	98.4844	2.5902	88.5691
Zona superior improductora	min	96.3517	48.7015	85.0329	83.8415	15.9738	72.0293	0.4103	98.4843	2.5903	88.5692
	max	100	100	120	100	35	100	0.65	100	2.96	100

El Cuadro anterior proporciona una visión de cómo se encuentran distribuidos los datos tanto de zonas productoras como de no productoras, y además ayuda hacer una estimación de los perfiles.

En el caso del perfil de Gamma Ray (GR), los datos indican que las zonas productoras son de litologías entre un 6% y un 48%, ósea areniscas y valores más bajos de Esmeclita (Arcilla), según el Cuadro 3. Excluyendo valores muy bajos pertenecientes a Calizas y Dolomitas entre otros; y para litologías como la Caolinita (Arcilla), con unidades API elevados representando zonas impermeables.

Para el perfil de Potencial Espontáneo (SP), los datos muestran que deflexiones por debajo de un 10% aproximadamente, son zonas permeables por lo cual son potencialmente productoras.

Para el perfil de Resistividad Verdadera (RT), los datos revelan que valores superiores a 3.3 Ohms., pertenecen a zonas potencialmente productoras y además que valores superiores a 16ohms., pueden demostrar presencia de otro fluido presente, al mismo tiempo que el petróleo ó diferente al petróleo ó presencia de gas.

El siguiente perfil es Porosidad Neutrónica (PHIN), los resultados dejan ver que valores superiores a 10% corresponden a zonas potencialmente productoras de hidrocarburos, pues su porosidad tiene la capacidad de almacenar fluidos.

El último perfil es Densidad de Formación (ROHB), los resultados enseñan que valores superiores de 2.13 gr/m<sup>3</sup> a 2.59 gr/m<sup>3</sup>, pertenecientes a litologías con una densidad de formación entre el yeso y la arenisca.

**5.5.2 Entrenamiento para datos preprocesados y normalizados.** Para crear una matriz de entrenamiento, con solo datos preprocesados y normalizados, se tomará los datos de los promedios del Cuadro 10. A continuación mostramos los datos de la matriz de entrenamiento en el Cuadro 17.

Cuadro 17: Comparativo de datos preprocesados y Normalizados

Datos preprocesados y normalizados		GR	SP	RT	PHIN	RHOB
		Porcentaje %	Porcentaje %	Porcentaje %	Porcentaje %	Porcentaje %
Zona inferior improductora	min	0	0	0	0	0
	max	5.99	6.00	8.12	10.42	31.25
Zona productora	min	6.00	6.10	8.13	10.43	31.26
	max	49.27	93.00	60.16	57.78	89.32
Zona superior improductora	min	49.28	93.10	60.17	57.79	89.33
	max	100	100	100	100	100

En el caso del perfil de Gamma Ray (GR), los datos indican que las zonas potencialmente productoras son de litologías entre un 6% y un 49.27%, ósea areniscas y valores más bajos de Esmectita (Arcilla), según el Cuadro 3. Excluyendo valores muy bajos pertenecientes a Calizas y Dolomitas entre otros; y para litologías como la Caolinita (Arcilla), con unidades API elevados representando zonas impermeables. Así acercándose a la teoría<sup>28</sup>.

Para el perfil de Potencial Espontáneo (SP), los datos muestran que deflexiones por debajo de un 17% aproximadamente, son zonas permeables por lo cual son potencialmente productoras.

Para el perfil de Resistividad Verdadera (RT), los datos revelan que valores superiores a 8%, equivalente a 2.5 ohms., pertenecen a zonas potencialmente productoras y además que valores superiores a 60% equivalente a 15 ohms., pueden demostrar presencia de otro fluido presente, al mismo tiempo que el petróleo o diferente al petróleo ó presencia de gas.

<sup>28</sup> Introducción a la interpretación de perfiles de pozo abierto, Ing. Jorge Arturo Camargo Puerto

El siguiente perfil es Porosidad Neutrónica (PHIN), los resultados dejan ver que valores superiores a 10.43% corresponden a zonas potencialmente productoras de hidrocarburos, pues su porosidad tiene la capacidad de almacenar fluidos.

El último perfil es Densidad de Formación (ROHB), los resultados enseñan que valores superiores a un 31% a 89%, pertenecientes a litologías con una densidad de formación parecida entre el yeso y la arenisca.

**5.5.3 Entrenamiento para datos preprocesados, normalizados y filtrados.** Para crear nuestra matriz de entrenamiento, con solo datos preprocesados, normalizados y filtrados, se tomará los datos del promedio del Cuadro 15. A continuación se mostrará los datos de la matriz de entrenamiento en el Cuadro 18:

Cuadro 18: Comparativo de datos preprocesados, normalizados y Filtrados

Datos preprocesados, normalizados y filtrados		GR	SP	RT	PHIN	RHOB
		Porcentaje %	Porcentaje %	Porcentaje %	Porcentaje %	Porcentaje %
Zona inferior improductora	min	0	0	0	0	0
	max	7.30	19.99	8.90	11.83	40.03
Zona productora	min	7.40	20.00	9.00	11.84	40.04
	max	45.21	97.00	55.21	54.64	82.75
Zona superior improductora	min	45.22	97.10	55.22	54.65	82.76
	max	100	100	100	100	100

En el caso del perfil de Gamma Ray (GR), los datos indican que las zonas potencialmente productoras son de litologías entre un 7.5% y un 45% aproximadamente, ósea areniscas y valores más bajos de Esmectita (Arcilla), según el Cuadro 3, Excluyendo valores muy bajos pertenecientes a Calizas y Dolomitas entre otros; y para litologías como la Caolinita (Arcilla), con unidades API elevados representando zonas impermeables. Así ajustándose a la teoría<sup>29</sup>. En anteriores proyectos de tesis los análisis arrojaron que las litologías potencialmente productoras estaban entre 10% y un 45%, el valor obtenido en esta tesis es un valor muy cercano, corroborando estos análisis.

Para el perfil de Potencial Espontáneo (SP), los datos muestran que deflexiones por debajo de un 20% aproximadamente, son zonas permeables por lo cual son potencialmente productoras. En teoría cualquier deflexión superior al 30%, corresponden a litologías potencialmente productoras<sup>30</sup>, en los datos analizados en el entrenamiento de datos preprocesados, arrojó que en la práctica, este valor esta solo por encima de 10%, lo que demuestra que valores por debajo de

<sup>29</sup> Guillermo A, Trujillo. Juan M. Polanco. Evaluación de Algoritmos Basados en Lógica Difusa para el Procedimiento de Señales de Registro de Pozo Abierto. Universidad Surcolombiana, 2009.

<sup>30</sup> Introducción a la interpretación de perfiles de pozo abierto, Ing. Jorge Arturo Camargo Puerto

30% pueden ser considerados como potencialmente productores, por lo tanto un 20% es un dato de litologías aparentemente poco arcillosas.

Para el perfil de Resistividad Verdadera (RT), los datos revelan que valores superiores a 9%, equivalente a 3 ohms., aproximadamente, pertenecen a zonas potencialmente productoras y además que valores superiores a 55% equivalente a 14 ohms., pueden demostrar presencia de otro fluido presente al mismo tiempo que el petróleo o diferente al petróleo ó presencia de gas. Este dato es acorde a la teoría<sup>31</sup> donde explica que una Resistividad Verdadera debe ser cuatro veces mayor que la  $R_o$ , y que  $R_o$  varía entre 0 ohms., y 1 ohms., lo que hace que cualquier (RT) superior, ó alrededor de 3 ohms., sea considerada como potencialmente productora.

El siguiente perfil es Porosidad Neutrónica (PHIN), los resultados dejan ver que valores superiores a 11.83% corresponden a zonas potencialmente productoras de hidrocarburos, pues su porosidad tiene la capacidad de almacenar fluidos. Se ha demostrado que para que un reservorio arenoso tenga producción comercial, su porosidad debe ser mayor o igual a 9%<sup>32</sup>.

El último perfil es Densidad de Formación (ROHB), los resultados enseñan que valores superiores a un 40% a 82%, pertenecientes a litologías con una densidad de formación superiores que el yeso, pasando por las areniscas y la caliza.

## **5.6 ANÁLISIS LINEAL DISCRIMINANTE (LDA)**

El análisis lineal discriminante puede ser empleado para cumplir dos funciones, la primera es la reducción dimensional, partiendo de un sistema de n-dimensional y reduciéndolo a un sistema bidimensional, con el fin de sintetizar de la mejor forma la información tratando de perder la menor cantidad de información, al igual que lo hacen el Análisis de Componentes Principales (PCA), pero con la diferencia que PCA siendo un reductor de información, el no reduce en ninguna medida la dimensionalidad de la matriz ordenada de datos de entrada, sino que éste realiza una reducción de filas completas de datos, manteniendo el porcentaje de error a niveles permisibles. La otra forma de emplear LDA es como método estadístico, para el reconocimiento de patrones y aprendizaje automático mediante la obtención de una combinación lineal de criterios que caracterizan o separan dos o más clases de objetos.

Para nuestro trabajo emplearemos a LDA en sus dos formas esenciales y de formas separadas, partiendo de una matriz de datos de n-filas dependiendo el

---

<sup>31</sup> Introducción a la interpretación de perfiles de pozo abierto, Ing. Jorge Arturo Camargo Puerto

<sup>32</sup> Introducción a la interpretación de perfiles de pozo abierto, Ing. Jorge Arturo Camargo Puerto

caso y de 5-columnas, es decir, para el método uno, lo emplearemos solo como clasificador básico y para el segundo, caso como reductor dimensional realizando una transformación lineal de la matriz de entrada, en la cual se realiza un realce a los datos de preponderancia para terminar sintetizándolo todo en una matriz de n-filas, 2-columnas, siendo entregado a SVM para que este sistema funcione como un clasificador y se estructure con este último un sistema clasificador más complejo, para tratar de elevar la efectividad del sistema. Para el desarrollo inicialmente se trabajaría tres caminos: el primero generaría la LDA básica con la información preprocesada, por el cual se obtendría la LDA con la información preprocesada y normalizada; y un último método con la información Preprocesada, normalizada y filtrada, como se muestra en la Figura 26.

Figura 26. Diagrama de bloques LDA



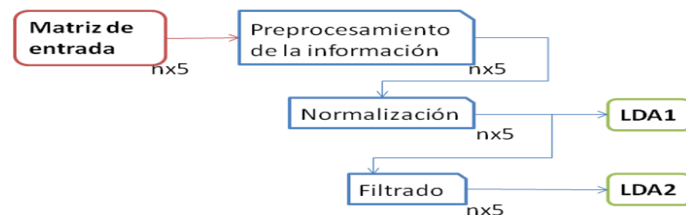
Durante la elaboración del proyecto se pudo notar que dos de los caminos propuestos inicialmente producían las mismas respuestas: LDA1 y LDA2, a la hora de obtener los resultados del clasificador, como se muestra en la Figura 27.

Figura 27: Comparativo de Normalización con Preprocesado de la LDA

590,8395236	591,9307368	1	677,3001965	678,3914097	1
598,0452267	599,1863267	1	682,4565296	683,5976296	1
609,1061326	610,3186969	1	692,0003392	693,2129035	1
637,2958611	638,2061756	1	714,8646957	715,7750102	1
655,0408201	655,8527923	1	733,6709389	734,4829111	1
656,6672698	657,2395449	1	735,3936677	735,9659428	1
648,935254	649,0900206	1	723,492458	723,6472245	1
631,3255028	631,4806141	1	702,441489	702,5966002	1
614,3585055	614,8863232	1	685,8974406	686,4252582	1
609,4538658	610,1967706	1	680,4458702	681,1887749	1
609,1485903	610,153059	1	681,3884058	682,3928745	1
612,2469675	613,3466854	1	684,8843162	685,984034	1
615,5830441	616,3414758	1	683,8223421	684,5807738	1
616,7326783	616,8389951	1	677,723271	677,8295878	1
620,3117401	619,5072327	0	674,0297339	673,2252266	0
622,0995695	620,5736765	0	672,5960979	671,070205	0
629,7989137	627,3304453	0	676,4532867	673,9848183	0
636,2690308	632,8806055	0	678,2531852	674,86476	0
Error no explorado	32		Error no explorado	32	
Error explorado	321		Error explorado	321	
Error Total	353		Error Total	353	
Porcentajes	0,82%		Porcentajes	0,82%	
	86,99%			86,99%	
	8,28%			8,28%	
LDA con Preprocesamiento			LDA		
LDA con Preprocesamiento y Normalizacion			LDA		
<b>LDA con Preprocesamiento</b>					
Error no explorado					0,82%
Error explorado					86,99%
Error Total					8,28%
<b>LDA con Preprocesamiento y Normalizacion</b>					
Error no explorado					0,82%
Error explorado					86,99%
Error Total					8,28%
<b>LDA 1</b>					
Error no explorado					29,11%
Error explorado					26,56%
Error Total					28,83%
<b>LDA 2</b>					
Error no explorado					15,85%
Error explorado					30,35%
Error Total					17,11%
<b>LDA Idealizada</b>					
Error no explorado					38,08%
Error explorado					49,32%
Error Total					39,05%

Por lo tanto el procedimiento ahora es el mostrado en la Figura 28.

Figura 28. Diagrama de bloques de LDA



Ahora, para trabajar la primera parte se obtuvieron dos clasificadores LDA, para cada pozo la razón del segundo es que mediante el estudio se encontró una variante del método original que generó una gran ventaja, proporcionando una disminución considerable en el error. Además de que a través del estudio del comportamiento de los registros de petróleos se generó una última LDA entrenada, que se basaba en un comportamiento idealizado de los registros para lo que llevó a calcular un total de 25 LDA entrenadas y se procedió a verificar el funcionamiento de tales sobre los pozos otorgados por ECOPEPETROL S.A., esto para nuestra primera etapa con el trabajo de LDA; siguiendo la segunda etapa se le realizó a los datos un filtrado adicional por lo que se tuvieron que generar nuevamente 25 LDA entrenadas para luego probar su funcionalidad sobre todos los pozos anteriormente mencionados.

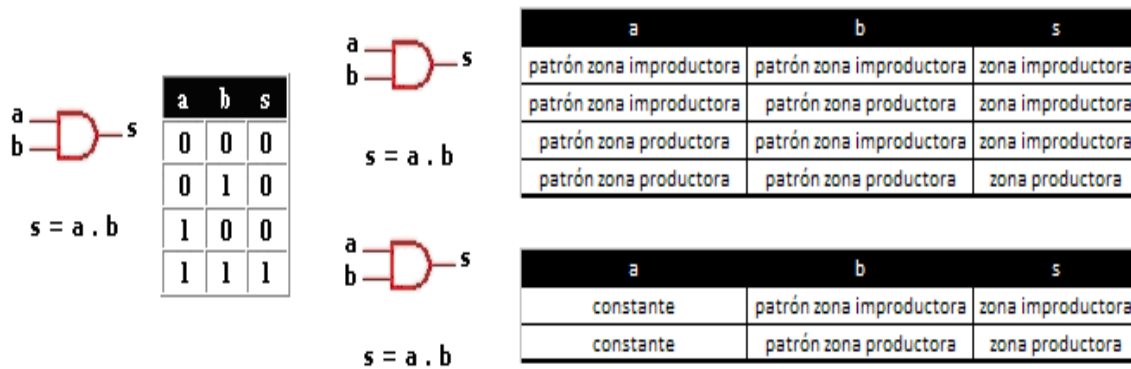
## 5.7 MÁQUINAS DE SOPORTE VECTORIAL (SVM)

El entrenamiento de una Máquina de Soporte Vectorial (SVM) se logra con un vector de entrenamiento de dos columnas y n-filas. Este método de clasificación requiere de una matriz de entrenamiento que consta de una pareja de valores, que describirán un patrón, caracterizando por aparte cada uno de los perfiles, los cuales pueden provenir de los datos filtrados a nivel 3 o del Análisis Lineal Discriminante. Los conjuntos de matrices de entrenamiento para la clasificación, fueron los descritos anteriormente en la sección 5.5, y de las matrices resultantes del Análisis Lineal Discriminante. Cada una de estas matrices cuenta con un tercer vector adicional, el cual tiene la función de dar membresía a las parejas de filas. La membresía será de 1 para parejas de puntos que describan un patrón de zona productora, y 0 cuando describa un patrón de zona no productora.

Sumado a esto cabe aclarar que las parejas de vectores, de dos columnas y n-filas, consta del registro extraído de un perfil de pozo abierto y un vector constante, así se evaluará por máquina solo un perfil, y no parejas de perfiles. Esto con el fin de simplificar tanto el entrenamiento como la clasificación.

Aunque el gasto computacional sea el doble, el margen de efectividad se elevará considerablemente. La razón por la cual se efectúa esto último es porque, cuando se tiene parejas de puntos, se deben considerar 4 situaciones, como en una compuerta AND. Si las dos entradas son cero, la respuesta es cero, si una de las entradas es cero, su respuesta es cero, pero sí las dos entradas son uno la salida es uno. Algo muy parecido sucede con las parejas de perfiles, lo cual hace que los entrenamientos sean más complejos. Por el contrario cuando se utiliza una pareja donde una de ellas es una constante, solo existe dos situaciones, si el dato del perfil describe un patrón de zona productora la respuesta es positiva, por el contrario si el dato del perfil describe un patrón de zona no productora la respuesta es negativa. Esto se asemeja a una compuerta AND donde una de sus entradas permanece en uno, lo cual hace que la respuesta de la compuerta solo dependerá del nivel que tome la otra entrada. Como se muestra en la Figura 29.

Figura 29: Comparación de una compuerta AND con un entrenamiento con 2 variables y una variable



Por ende son cinco Máquinas de Soporte Vectorial, una por cada perfil escogido, las estrenadas para todos los pozos suministrados por ECOPETROL SA., el resultado final es producto de la comparación de los resultados de todas las cinco máquinas corridas por cada llano. Así cuando se corre las cinco máquinas en un pozo, para que una zona sea declarada potencialmente productora, el resultado de las cinco máquinas tiene que ser positivo.

La función de Núcleo o función Kernel utilizada en la Máquina de Soporte Vectorial (SVM), es la función de base radial gaussiana (Gaussian Radial Basis Function, RBF), la cual es responsable de mapear los datos de entrenamiento, dentro del espacio de la función de núcleo. La función encargada de encontrar el hiperplano de separación de clases es la de Cuadrados Mínimos (Least-Squares, LS). Además el valor que especifica la tolerancia con las condiciones Karush-Kun-Tucker (KKT) es de  $1.0000e-003$  y el valor que especifica la fracción de variables a las cuales se les permite violar las condiciones KKT, es de 0.05, esto ayuda al algoritmo a converger si está fluctuando al borde de una buena

solución y por último se programo para que el número máximo de las repeticiones del bucle principal, sea el máximo (infinito), porque si este límite es sobrepasado antes de que el algoritmo converja, entonces el algoritmo se detiene y devuelve un error. A continuación se mostrará en la Figuras 30 y Figura 31, gráficas de entrenamiento de una SVM. En la Figura 32 y en la Figura 33 gráficas de clasificación de una SVM

Figura 30: Entrenamiento del perfil de Densidad de Formación (RHOB) del LL-105

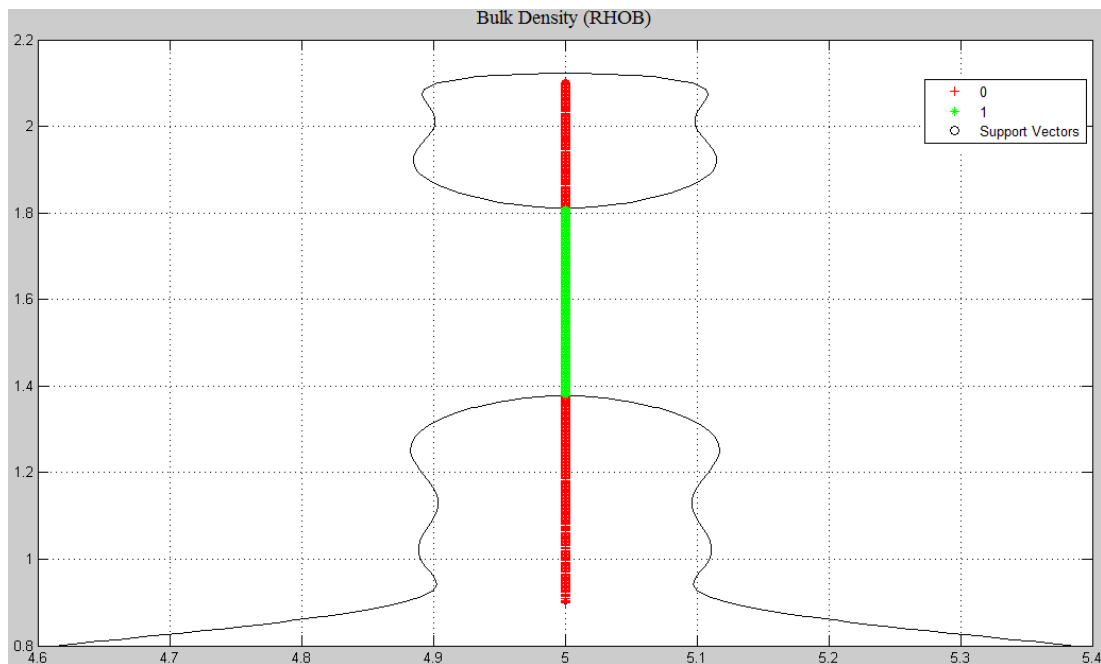


Figura 31: Entrenamiento del perfil de Resistividad Verdadera (RT) del LL-105

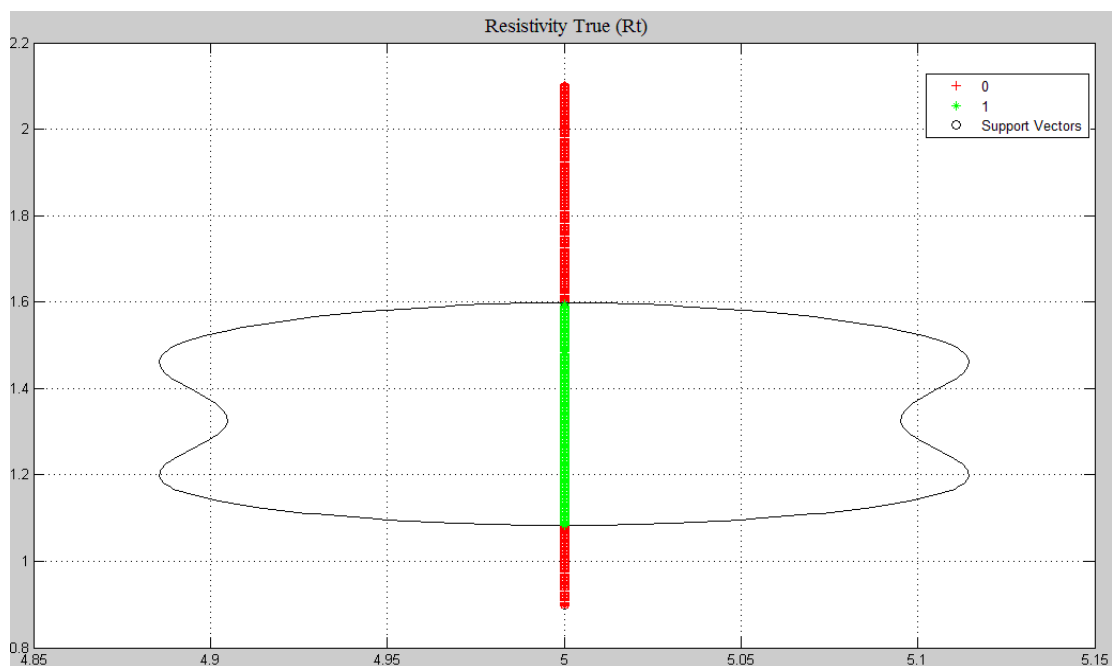




Figura 32: Clasificación del perfil de Densidad de Formación (RHOB) del LL-102

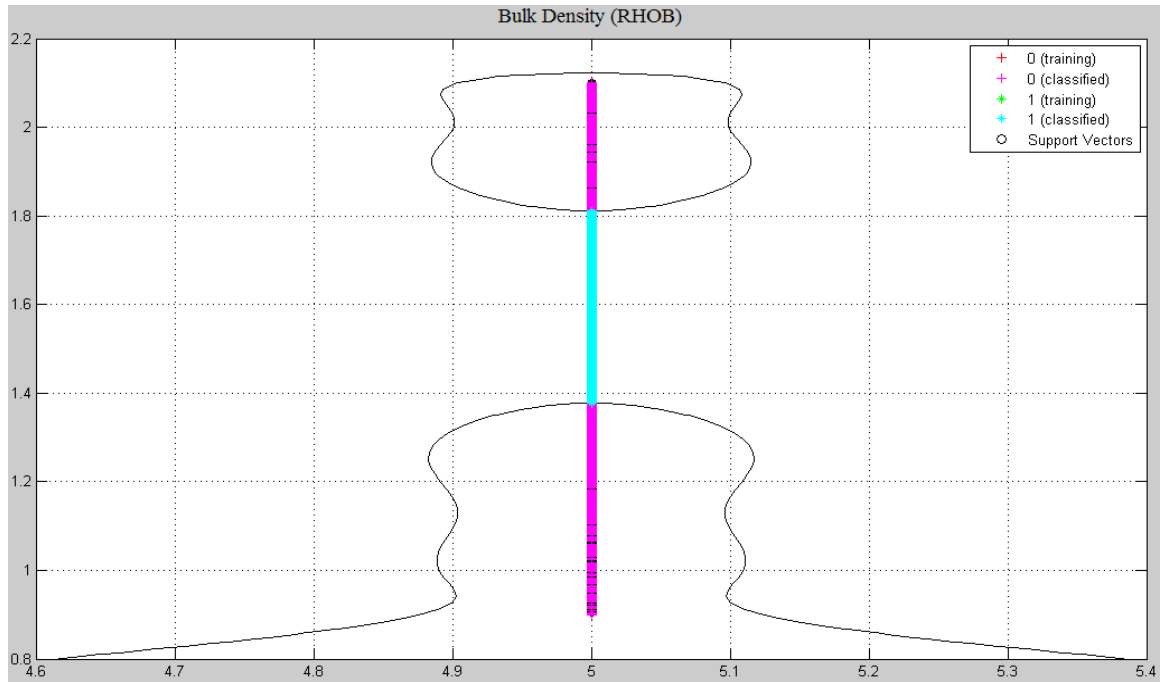
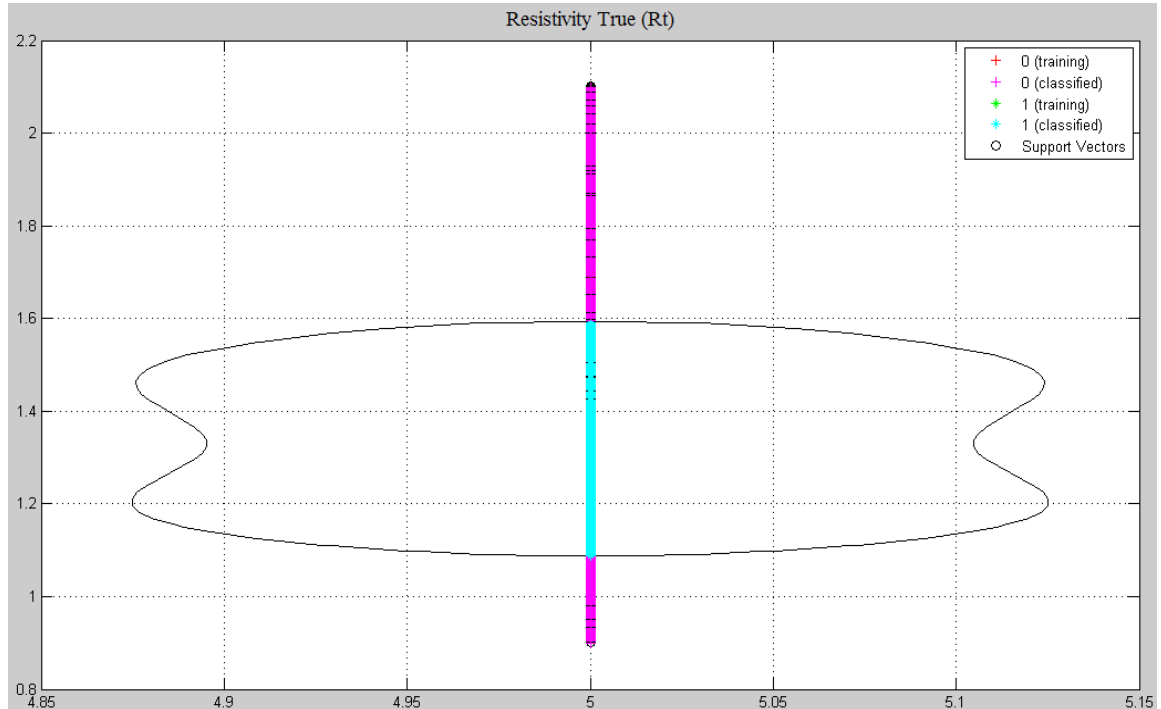


Figura 33: Clasificación del perfil de Resistividad Verdadera (RT) del LL-105



## 6. ANÁLISIS DE RESULTADOS

### 6.1 RESULTADOS ENTRENAMIENTO

**6.1.1 Resultados Entrenamiento (LDA).** Para el trabajo con análisis lineal discriminante por el primer método (Pre-procesamiento y Normalización) se obtuvieron los siguientes datos de efectividad expuestos en el siguiente Cuadro:

Cuadro 19: Comportamientos de efectividad de las mejores LDA entrenadas

Método 1										
	3	4	5	7	10	11	12	17	21	23
	T_72F2	T_77F1	T_77F2	T_81F2	T_96F1	T_96F2	T_97F1	T_103F2	T_105F2	T_106F2
Zonas no Cañoneadas	70,54%	72,37%	63,41%	70,54%	64,99%	71,59%	74,19%	77,86%	68,32%	79,22%
Zonas Cañoneadas	81,03%	82,02%	87,10%	81,03%	78,09%	82,52%	67,44%	74,16%	81,22%	76,61%

A Raíz de los anteriores resultados se puede concluir que la LDA entrenada de mejor comportamiento de efectividad sobre todos los pozos es la número 4 como se puede ver claramente en el Cuadro 19, que resume el comportamiento de efectividad de las mejores LDA del método número 1. Ahora para el método 2 se obtuvieron los datos de efectividad mostrados en el cuadro 20:

Cuadro 20: Resumen de las LDA entrenadas de mejor efectividad.

Método 2										
	5	7	8	10	11	12	14	16	20	23
	T_77F2	T_81F2	T_91F1	T_96F1	T_96F2	T_97F1	T_102F1	T_103F1	T_105F1	T_106F2
Zonas no Cañoneadas	83,58%	55,20%	52,94%	60,95%	85,12%	70,33%	76,18%	74,36%	65,84%	84,11%
Zonas Cañoneadas	68,22%	85,70%	84,85%	82,74%	69,60%	70,40%	64,90%	64,54%	77,14%	66,42%

A Raíz de los anteriores resultados se puede concluir que la LDA entrenada de mejor comportamiento de efectividad sobre todos los pozos es la número 11, porque aunque no sea clara mediante el Cuadro 40. En la cual muestra que el mejor comportamiento lo posee la LDA 12, si se observa el comportamiento completo de la LDA 11 y LDA12 la que conserva más estabilidad durante todo el estudio es la LDA 11 para el método 2.

A causa de los resultados anteriores se puede concluir que la LDA entrenada que me mejor se comporta para nuestro caso es la LDA del método 1, porque aunque sea la de datos sin filtrar para cuestiones prácticas es necesario garantizar la detección por parte de nuestro algoritmo de las zonas que Ecopetrol S.A. considero como cañoneables, pero se es permisible un error en las zonas que no fueron cañoneadas puesto que pueden haber zonas de 4 pies o

un poco mas que no fueron tomadas en cuenta, que cumplen a cabalidad con los criterios respectivos para ser tomadas como zonas de interés.

**6.1.2 Resultados Entrenamiento (SVM).** La Máquina de Soporte Vectorial es un sofisticado algoritmo de aprendizaje automático que logra por medio de una población de entrenamiento, aprender un patrón que caracterice la población de entrenamiento y así mismo lograr una clasificación. Los diferentes entrenamientos se lograron con los tres tipos de selección de población de entrenamiento. A continuación se expondrá los porcentajes de efectividad de los tres entrenamientos. En el Cuadro 21 se mostrará las efectividades del entrenamiento con datos preprocesados. En el Cuadro 22 se enseñará los porcentajes de efectividad de los datos preprocesados y normalizados. En el Cuadro 23 se mostrará las efectividades del entrenamiento con datos preprocesados, normalizados y filtrados.

Cuadro 21: Entrenamiento con datos Preprocesados

	Preprocesado				
	(GR)	(SP)	(RT)	(PHIN)	(ROHB)
LL-72	97.9	98.7	99	99.2	99.1
LL-77	97.9	98.7	99.4	97.5	98.5
LL-81	98.5	99.4	99.2	98.6	98
LL-91	98.8	98.7	98.3	97.6	98.7
LL-96	98.8	99.3	97.6	99.9	98.3
LL-97	99	97.9	98.8	98.3	99
LL-102	99.2	99.1	98.8	97.6	99.4
LL-103	98.3	98.9	97.9	98.9	99
LL-104	98.3	98.4	98.4	99.7	98.8
LL-105	98.8	98.7	98.9	98.4	99.5
LL-106	99.5	97.9	99.6	98.9	98.5
LL-107	99.8	99	99.7	99.4	98.7
Promedio	98.73333333	98.725	98.8	98.66666667	98.79166667

Cuadro 22: Entrenamiento con datos Preprocesados y Normalizados

	Preprocesado y Normalización				
	(GR)	(SP)	(RT)	(PHIN)	(ROHB)
LL-72	98.6	99.4	99.3	99.7	99.5
LL-77	97.9	98.5	99.4	97.5	98.5
LL-81	98.8	98.5	98.3	97.6	98.7
LL-91	98.5	99.4	99.2	98.6	98
LL-96	98.8	99	99.6	99.9	98.3
LL-97	99	97.9	97.6	98.3	99
LL-102	99.2	99.5	98.8	97.6	99.4
LL-103	99.5	97.9	98.4	98.9	98.5
LL-104	98.3	98.4	98.8	99.7	98.8
LL-105	97.8	98.1	98.9	98.9	99
LL-106	98.2	98.7	97.9	98.4	99.5
LL-107	99.8	99.3	99.7	99.4	98.7
Promedio	98.7	98.71666667	98.825	98.70833333	98.825

Cuadro 23: Entrenamiento con datos Preprocesados, Normalizados y Filtrado

	Preprocesado, Normalización y Filtro nivel 3				
	(GR)	(SP)	(RT)	(PHIN)	(ROHB)
LL-72	98.2	99.6	99.5	99.6	99
LL-77	98.8	99	98.3	97.6	98.7
LL-81	97.9	98.5	97.6	98.7	97.5
LL-91	99	97.9	99.1	99.6	99.4
LL-96	99.2	99.5	98.8	97.6	99.8
LL-97	98.5	99.4	99.2	98.6	97.7
LL-102	98.8	99	99.6	99.9	98.3
LL-103	99.5	97.9	98.4	98.9	98.5
LL-104	98.3	99.3	98.8	99.7	98.8
LL-105	97.8	98.1	97.8	98.3	99
LL-106	98.2	98.7	97.9	98.4	99.5
LL-107	99.8	99	99.7	99.6	98.7
Promedio	98.66666667	98.825	98.725	98.875	98.74166667

## 6.2 RESULTADOS CLASIFICACIÓN

**6.2.1 Resultados Clasificación (LDA).** Para poder ver con claridad la efectividad de la LDA escogida, se realizará un cuadro de resumen de las efectividades de los dos métodos.

Cuadro 24: Validación del clasificador LDA.

	LDA Entrenada	Método 1	Método 2
		T_77F1	T_96F2
LL-72	Zonas no Cañoneadas	85,13%	97,16%
	Zonas Cañoneadas	91,47%	64,45%
LL-77	Zonas no Cañoneadas	88,94%	91,16%
	Zonas Cañoneadas	82,95%	73,13%
LL-81	Zonas no Cañoneadas	85,13%	96,87%
	Zonas Cañoneadas	91,47%	64,93%
LL-91	Zonas no Cañoneadas	62,32%	92,97%
	Zonas Cañoneadas	98,91%	91,30%
LL-96	Zonas no Cañoneadas	84,17%	84,86%
	Zonas Cañoneadas	51,98%	55,94%
LL-97	Zonas no Cañoneadas	45,18%	85,44%
	Zonas Cañoneadas	97,86%	75,53%
LL-102	Zonas no Cañoneadas	56,33%	77,03%
	Zonas Cañoneadas	83,51%	68,04%
LL-103	Zonas no Cañoneadas	71,76%	76,44%
	Zonas Cañoneadas	57,99%	71,27%
LL-104	Zonas no Cañoneadas	62,88%	78,42%
	Zonas Cañoneadas	73,58%	77,26%
LL-105	Zonas no Cañoneadas	64,00%	80,43%
	Zonas Cañoneadas	77,46%	91,55%
LL-106	Zonas no Cañoneadas	74,19%	60,81%
	Zonas Cañoneadas	63,98%	93,79%
LL-107	Zonas no Cañoneadas	98,49%	99,85%
	Zonas Cañoneadas	16,80%	8,00%
Promedios			
Zonas no Cañoneadas		72,37%	85,12%
Zonas Cañoneadas		82,02%	69,60%

**6.2.2 Resultados de Clasificación (SVM).** A continuación se mostrará los porcentajes de efectividad de la clasificación con Máquina de Soporte Vectorial, de los tres tipos de entrenamientos realizados en el desarrollo del proyecto, analizando el porcentaje de efectividad de encontrar las zonas cañoneadas por ECOPETROL SA., las cuales son de referente para este proyecto, y el porcentaje de efectividad por así llamarlo de zonas que ECOPETROL SA., no cañoneo y que la máquina no propone que se deben cañonear. Cabe recordar que este es el objetivo de este proyecto, encontrar nuevas zonas mayores de cuatro pies potencialmente productoras de Hidrocarburo que ECOPETROL SA., no haya cañoneado, así que lo ideal es que el porcentaje de efectividad de zonas no cañoneadas sea diferente de 100%, para que existan nuevas zonas descubiertas por el clasificador de Máquina de Soporte Vectorial.

Cuadro 25: Comparación de efectividad punto a punto de los tres tipos de clasificaciones de SVM en todos los pozos

	Preprocesado		Preprocesado+Normalizado		Preprocesado+Normalizado+Filtrado	
	zonas cañoneadas	zonas no cañoneadas	zonas cañoneadas	zonas no cañoneadas	zonas cañoneadas	zonas no cañoneadas
LL-72	56.8904	77.2536	85.8657	72.9459	93.9929	67.3277
LL-77	78.6632	81.9521	82.5192	81.0016	88.4318	67.8336
LL-81	66.8246	84.2381	92.4170	80.3316	94.7867	76.5956
LL-91	90.7608	88.8799	92.9347	76.5571	97.2826	74.3152
LL-96	65.0990	77.9112	82.9207	70.3682	85.1485	67.8872
LL-97	90.0970	66.9900	92.0388	66.7129	91.8446	62.1809
LL-102	90.1808	81.9964	50.3875	81.5385	68.2170	62.6924
LL-103	91.1290	76.7345	96.2466	66.7427	97.0509	62.9060
LL-104	30.7692	78.8255	63.2107	72.9325	91.3043	53.1563
LL-105	72.4381	95.6739	79.2253	79.7133	83.0388	74.1585
LL-106	70.5872	93.5324	81.6770	84.8445	90.5923	79.7062
LL-107	83.2000	94.8800	97.6000	91.6204	100.0000	91.1950
Promedio	73.8866	83.2390	83.0869	77.1091	90.1409	69.9962

Como podemos observar el porcentaje de efectividad de los datos preprocesados es bueno, pero cuando los datos se normalizan el porcentaje de efectividad sube, y el porcentaje de efectividad de las zonas no cañoneadas baja, esto sucede porque al normalizar la maquina se hace menos excluyente, el patrón trata de abarcar más datos. Al filtrar la efectividad de zonas cañoneadas aumenta otro tanto haciéndose optimo, y se observa que el porcentaje de efectividad de zonas no cañoneadas disminuye, pues al filtrar la onda resultante es más lisa lo cual elimina espigas pequeñas (ruido), que en otro momento eran consideradas no productoras de hidrocarburo.

### 6.3 RESULTADOS DE ZONAS POTENCIALMENTE PRODUCTORAS

El objetivo final de este proyecto es proponer zonas potencialmente productoras, no cañoneadas, con una Máquina de Soporte Vectorial y un Análisis Lineal Discriminante, validados. Estos son los resultados mostrados por Máquina de Soporte Vectorial y Análisis Lineal Discriminante de zonas potencialmente productoras no cañoneadas por ECOPETROL SA.

Cuadro 26: Zonas potencialmente productoras de SVM y LDA

Nº de Zonas	LL-72		LL-77		LL-81		LL-91		LL-96		LL-97	
	base	tope	base	tope	base	tope	base	tope	base	tope	base	tope
1	5241.5	5248	5348	5358.5	5525.5	5531	5224.5	5241	5470	5475	5345.5	5355
2	5254	5261.5	5563.5	5573	5671	5682.5	5245.5	5250	5483.5	5496	5401.5	5407.5
3	5339.5	5348	5720.5	5724.5	5855.5	5866.5	5263	5271	5514.5	5524.5	5421	5427.5
4	5363	5404.5	5772.5	5791	5878.5	5890.5	5358.5	5372	5535.5	5572	5441	5452
5	5410.5	5417.5	6026	6033	5920.5	5927	5389	5400	5589.5	5595	5739.5	5750
6	5464	5468.5	6089	6093	5965	5973.5	5410	5419	5833	5849	5837.5	5851
7	5572	5578	6336.5	6377.5	6011	6018	5435	5454	5855	5859.5	6098.5	6120.5
8	5688	5741.5	6409.5	6417	6051	6061	5465	5476	5901.5	5913	6233	6237.5
9	5770.5	5778	6427	6436	6069.5	6079	5487	5499	5952.5	5962.5	6244	6257.5
10	5786.5	5800	6524	6533.5	6091.5	6098.5	5507	5519	5968.5	5988	6314.5	6318.5
11	5817.5	5940.5	6540.5	6576	6258	6268	5562	5571	6000	6011	6360.5	6382.5
12	5955	6042	6639	6652.5	6271.5	6278.5	5632	5638	6020.5	6027	6405	6410
13	6049.5	6061	6660	6666.5	6295.5	6332.5	5646	5658	6050.5	6058.5	6416.5	6422.5
14	6087	6094	6715	6722	6337.5	6374	5677	5704	6062	6068	6481.5	6486.5
15	6099.5	6123.5	6729.5	6735.5	6486	6493	5723	5740	6073	6084	6493	6505.5
16	6311	6342	6839.5	6855.5	6504	6548	5748	5799	6093	6112	6515.5	6521
17	6508.5	6521	6927.5	6936	6558	6569.5	5805	5862	6155	6161	6524.5	6544.5
18	6530	6537.5	6954	6963	6574.5	6586.5	5867	5872	6213.5	6221	6560	6566.5
19	6543	6553	6985.5	6993.5	6596	6602.5	5880	5886	6236.5	6255	6687.5	6695.5
20	6554	6587	6998.5	7006.5	6627.5	6636.5	5910	5940	6261.5	6266.5	6703	6707
21	6614	6634	7013.5	7025.5	6703	6707	5968	5991	6281	6289	6965.5	6970
22	6640.5	6657	7030.5	7037.5	6712.5	6726	5997	6018	6305.5	6324.5	7074.5	7083
23	6742.5	6747.5	7040	7045.5	6766.5	6774	6032	6039	6335	6339	7092	7098
24	6776	6790.5	7093.5	7119.5	6813.5	6824	6048	6053	6347	6352.5	7124	7133
25	6840	6874.5	7127	7136	6845	6849.5	6081	6087	6414	6427	7150.5	7160.5
26	6898	6906.5	7146.5	7191.5	6853.5	6859.5	6120	6128	6432	6437.5	7164	7173.5
27	6945	6949	7202	7213	6910	6915.5	6133	6160	6450.5	6464	7179.5	7184.5
28	6955	6965.5	7217.5	7223	6954.5	6975.5	6173	6190	6471.5	6484	7190.5	7195.5
29	6971	6992.5	7235.5	7247.5	6981.5	7004	6288	6295	6496	6502	7289.5	7295
30	7038.5	7108	7298	7304	7015	7025	6321	6342	6509.5	6520.5	7307.5	7318.5
31	7128.5	7140	7305.5	7333	7031.5	7042	6355	6369	6549	6563.5	7341	7346
32	7155.5	7170.5	7338.5	7346	7054.5	7083	6377	6383	6564	6568.5	7391.5	7396
33	7176	7187	7347	7352	7116	7132	6440	6469	6663.5	6675.5	7423.5	7429.5
34	7196.5	7217.5	7401.5	7407	7179	7186	6479	6502	6680.5	6707.5	7565	7579
35	7249	7259	7469.5	7481	7202	7258	6511	6532	6713.5	6718.5	7589	7596
36	7268	7294	7553	7590.5	7279	7294.5	6578	6606	6727	6736.5		
37	7301.5	7325.5	7637	7671.5	7349.5	7355.5	6616	6673	6739.5	6751.5		
38	7335.5	7341.5	7681.5	7691	7390	7401	6680	6685	6757.5	6763		
39	7351	7356			7476.5	7485.5	6694	6710	6771.5	6777		
40	7379	7395.5			7556.5	7562	6854	6864	6808.5	6815.5		
41	7401.5	7411			7575.5	7586	6868	6872	6820	6826		
42	7423.5	7440.5			7599	7607	6882	6935	6829.5	6833.5		
43	7462	7475					6939	6973	6876	6896		
44	7585	7593					6980	6986	6906	6915		
45	7666.5	7681.5					7007	7016	6922.5	6930		
46	7753	7767					7022	7028	6934.5	6939.5		
47							7059	7087	6948	6954		
48							7106	7116	6962	6967		
49							7189	7197	6975	6991.5		
50							7203	7211	7020	7025		
51							7248	7253	7158	7169		
52							7263	7279	7172.5	7187.5		
53							7319	7332	7206.5	7218		
54							7406	7418	7370	7376		
55							7480	7488	7454	7460		
56							7500	7513				
57							7522	7537				

Cuadro 27: Zonas potencialmente productoras de SVM y LDA

Nº de Zonas	LL-102		LL-103		LL-104		LL-105		LL-106		LL-107	
	base	tope	base	tope	base	tope	base	tope	base	tope	base	tope
1	5752	5763	5143.5	5148	5103	5107	6307	6311.5	5406	5411	5441	5447
2	5859.5	5863.5	5358.5	5365	5206	5218	6324.5	6329.5	5617.5	5625	5507	5514.5
3	6015.5	6024	5478.5	5487.5	5714	5718.5	6357.5	6362	5639	5698.5	5786.5	5803
4	6039.5	6046.5	5490.5	5498.5	5786.5	5791.5	6364	6371.5	5760	5770.5	6029.5	6034
5	6110	6123.5	5508.5	5516.5	5801	5806.5	6475	6485.5	5961.5	5964.5	6039	6052.5
6	6141.5	6153	5520	5524.5	5937	5943	6487.5	6496	6034.5	6039.5	6072	6084
7	6336.5	6344.5	5532	5541.5	5949	5954.5	6643	6649	6092.5	6101.5	6091.5	6096
8	6460	6471	5694.5	5745.5	6013	6018	7129.5	7135.5	6478.5	6483	6154	6161
9	6531	6536	5862.5	5866.5	6039.5	6063	7298	7303	6488	6492	6207	6213
10	6623	6628	5872.5	5877	6119	6131.5			6666	6678.5	6336	6341
11	6670	6679.5	5886	5899	6141.5	6145.5			6697.5	6702.5	6511.5	6515.5
12	6702	6707.5	5913	5919	6163.5	6167.5			6714.5	6725	6664.5	6672
13	6745.5	6757	5968	5982.5	6172.5	6178			6929	6934.5	7222.5	7228
14	6761.5	6766.5	6033.5	6071.5	6325.5	6333					7270.5	7280.5
15	6774.5	6783.5	6209	6218	6342.5	6350.5					7372	7376.5
16	7010.5	7016.5	6260	6272.5	6377.5	6387					7387	7393
17	7183.5	7189.5	6301.5	6308	6416.5	6421.5					7414	7419.5
18	7223	7233	6523.5	6568	6439.5	6475.5					7420.5	7443.5
19	7342.5	7347	6574.5	6587.5	6482	6499					7449	7456.5
20	7374	7380	6604.5	6617	6522.5	6528.5					7472.5	7486
21	7386.5	7401.5	6657.5	6662	6538.5	6544.5					7506.5	7512.5
22	7403.5	7408.5	6684.5	6689.5	6564.5	6574					7514.5	7524
23	7475.5	7493.5	6721	6727	6589	6612					7681	7686
24	7559	7564.5	6772	6794.5	6624	6630					7691	7696.5
25	7568.5	7574	6803.5	6810	6637	6643.5					7970	7974
26	7634	7641	6861.5	6878.5	6674.5	6680					7989.5	7994
27	7645.5	7651.5	6882	6895	6683	6695.5						
28	7762.5	7778.5	6905	6919.5	6704.5	6714.5						
29			6922.5	6929.5	6735	6739.5						
30			6946	6951	6782.5	6788.5						
31			6957	6967.5	6867	6872.5						
32			6968.5	6980	6877.5	6884						
33			6995	7001	6900.5	6904.5						
34			7011	7025	6921.5	6944						
35			7031.5	7058	6970.5	6979.5						
36			7059	7066.5	7020.5	7028.5						
37			7092	7105	7033	7037.5						
38			7334	7347.5	7044.5	7060.5						
39					7090.5	7100.5						
40					7140.5	7150						
41					7158	7165.5						
42					7190	7206						
43					7209	7234.5						
44					7265.5	7275						
45					7353.5	7371						
46					7444	7450.5						

## 7. CONCLUSIONES Y RECOMENDACIONES

- Los estudios a los perfiles demostraron que la varianza típica de cada perfil en los pozos era muy alta, haciendo que la caracterización por medio de un patrón fuera difícil, los datos confirman que es necesario procesar la información para disminuir la varianza típica y así hacer más pequeñas las diferencias.
- Los resultados demostraron que los datos atípicos al final de los perfiles eran datos prescindibles, sin la presencia de ellos los análisis estadísticos mostraron una mejoría notable, convirtiéndose en datos no representativos para cada perfil.
- Se destaca la importancia de la utilización de la normalización, pues logró excelentes resultados disminuyendo en gran parte la varianza típica, y así se consiguió un patrón mucho más robusto y eficaz.
- La elección final de la Wavelet madre se logra en base a la función que posea un mínimo valor de la función de costo, tomando como resultado la función de entropía de Shannon, en donde se consiguió identificar en qué lugar se encuentra la mayor cantidad de datos repartidos que representan la señal.
- La reducción de ruido gaussiano por medio de la umbralización fue un éxito, haciendo uso del principio del mínimax y la umbralización flexible, los resultados demostraron una mejoría y se logró una mejor caracterización de la onda.
- Aunque los dos métodos de umbralización demostraron un desempeño sumamente parecido, fue escogido el método de umbralización flexible, por su gran capacidad de adaptabilidad a cualquier onda.
- El método de selección de población de entrenamiento fue muy bueno, pues logró una población de datos representativos que hizo que los métodos de aprendizajes automáticos como Máquina de Soporte Vectorial y Análisis Lineal Discriminante, técnicas de clasificación utilizadas en este proyecto, fueran efectivas.
- Para el estudio con LDA pudimos notar que la herramienta aunque no fue totalmente acertada debido a que las categorías no conseguían ser totalmente ortogonales, además de algunos factores agravantes que fueron minimizados notablemente mediante algunos mecanismos como lo son preprocesamiento de la señales y de Procesamiento Digital de



Señales (DSP), como lo son la Normalización y el filtrado, su resultado fue muy bueno, logrando una buena eficacia.

- Al ser un método estadístico LDA tiene un efecto directo sobre la cantidad de datos con la cual fue entrenada, entonces valiéndonos de esto generamos múltiples LDA entrenadas para así llegar a obtener una LDA con una efectividad alta, mostrando entonces mediante los resultados solo las que mejor se acomodaron a nuestras necesidades.
- Máquina de Soporte Vectorial demostró una gran versatilidad y eficacia a la hora de clasificar, en especial cuando los datos eran preprocesados, normalizados y filtrados, dando una margen de eficacia superior al 90%, esto gracias al procesamiento de la señal que ayuda enormemente a facilitar la caracterización de las señales.
- El entrenamiento por individual de cada perfil provoco un gasto computacional mayor, pero ayudó considerablemente no solo a la simplicidad del entrenamiento sino a la efectividad de la misma logrando márgenes excelentes de eficacia.
- El acople LDA-SVM, es infructuoso, pues los datos proporcionados por LDA no proporcionan un patrón a SVM, por que no se halló forma de adaptar los resultados de LDA a SVM.
- Se recomienda para futuros estudios trabajar con un método estadístico de orden superior como el análisis de componentes independientes (ICA), la cual proporciona una independencia de órdenes superiores, no presume la ortogonalidad de las fuentes (característica en las señales de perfiles), y permite encontrar la línea de ruido como fuente independiente, pudiendo sustraerlas de las señales sin pérdida de información.
- Además se recomienda el uso de algoritmos genéticos el cual es un método adaptativo que pueden usarse para resolver problemas de optimización, de gran aplicación en la optimización de sistemas de compresión de datos usando wavelets y procesamiento de señales con lógica difusa.
- Por último se recomienda, aprovechar las características de las Wavelet madres para crear una nueva y más amplia base wavelet para una aplicación específica o para un determinado tipo de señal, pues al ser bastantes y muy distintas, tienen la naturaleza de adaptarse bien a muchas aplicaciones y tipos de señales.

## 8. BIBLIOGRAFÍA

- An Introduction to Support Vector Machines and Other Kernel-Based learning methods, Nello Cristianini y John Shawe-Taylor, 2006
- Learnig kernel Classifiers Theory and Algorithms, Ralf Herbrich, 2002
- Análisis Multivariado: Método de Componentes Principales. Secretaria General de la Organización de los Estados Americanos, Programa Regional del desarrollo científico y Tecnológico.
- Bernal, Hugo. Rodríguez, Alfredo. (2009). Identificación de zonas productoras de hidrocarburos Aplicando DSP y métodos de correlación de registros de perfilaje. Universidad Surcolombiana
- Guillermo A, Trujillo. Juan M. Polanco. Evaluación de Algoritmos Basados en Lógica Difusa para el Procedimiento de Señales de Registro de Pozo Abierto. Universidad Surcolombiana, 2009.
- Betancourt, Gustavo. (2005). Las Maquinas de Soporte Vectorial (SVMs). Universidad tecnológica de Pereira.
- Burges, C. (1998). A tutorial on support vector machines for pattern recognition. Knowledge Discovery and Data Mining, 2:22.
- Díaz M., Luis G. Estadística Multivariada: Inferencia y Métodos. Facultad de Ciencias, Universidad Nacional de Colombia.
- Duda, R. O., Hart, P. E., and Stork, D. G. (2001). Pattern Classification. Wiley- Interscience, New York, second edition.
- Hair, Anderson, Tathan, Black. Análisis Multivariante, Editorial Pearson, Prentice Hall. Edición 5.
- Mauricio Delbracio, Matías Mateu. (2006). Trabajo Final de Reconocimiento de Patrones: Identificación utilizando PCA, ICA y LDA.
- Mauricio Orozco Alzate. (2003). Clasificación de Arritmias Cardíacas usando Transformada Wavelet y Técnicas de Reconocimiento de Patrones Departamento de Electricidad, Electrónica y Computación.

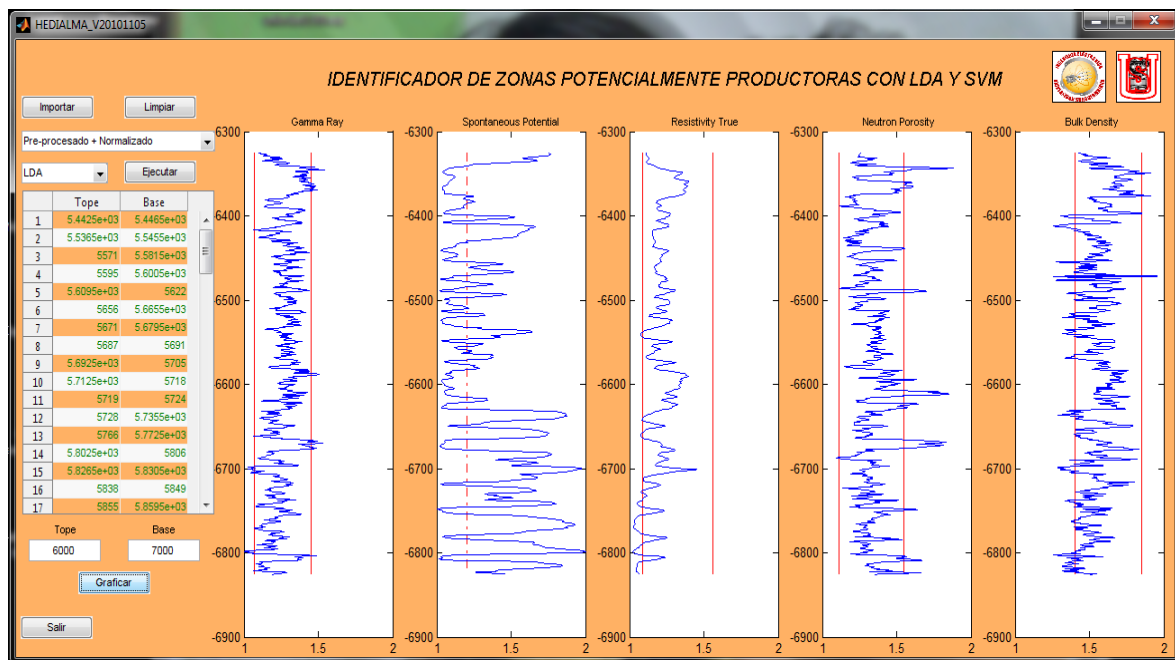
- Pedroza Bernal, Juan Gabriel. (2007). Aplicaciones de las maquinas de soporte Vectorial al reconocimiento de hablantes. Universidad Autónoma Metropolitana.
- Introducción a la interpretación de perfiles de pozo abierto, Ing. Jorge Arturo Camargo Puerto
- SCHLUMBERGER. (1978). Fundamentos de la interpretación de perfiles. Schlumberger Well Service.
- GABOR, Denis. Theory of Communication. Journal of Institute for Electrical Engineering, 1946.
- S. Kouro y R. Musalem. Tutorial introductorio a la Teoría de Wavelet, Artículo presentado como trabajo de la asignatura Técnicas Modernas en Automática.
- D. Cuesta, “Estudio de métodos para procesamiento y agrupación de señales electrocardiográficas,” Ph.D. dissertation, Valencia, 2001.
- D. Donoho, “De-noising by soft-thresholding,” Tech. Rep. Dept. of Statistics, Stanford University, 1995.
- Cristian Guarnizo, Victoria E. Montes, Julian Echeverry “Agrupamiento de impulsos neuronales en señales neurofisiológicas,” Universidad Tecnológica de Pereira, 2006.
- H. Chi-Wei, Ch. Chih-Chung, L. Chih-Jen, A Practical Guide to Support Vector Classification
- Tixier, M. P., and Alger, R. P., Log Evaluation of Non-Metallic Mineral Deposits, 1967
- Y. M. Stéphane Jaffard y R. D. Ryan, Wavelets: Tools for Science and Technology. Society for Industrial and Applied Mathematics, 2001.
- <http://people.revoledu.com/kardi/tutorial/LDA/>
- <http://www.dtreg.com/lda.htm>

## 9. ANEXOS

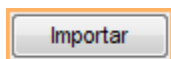
### 9.1 TUTORIAL DEL IDENTIFICADOR DE ZONAS POTENCIALMENTE PRODUCTORAS CON LDA Y SVM

Para la elaboración del programa se ha utilizado como plataforma de operación la herramienta de entorno gráfica GUIDE que ofrece el MatLab, la cual emplea una programación orientada a objetos mediante el cual podemos generar hábiles estructuras que facilitarían el trabajo e implementación de algoritmos. El programa IDENTIFICADOR DE ZONAS POTENCIALMENTE PRODUCTORAS CON LDA Y SVM, nos permite de forma fácil cargar los perfiles en formato plano \*.LAS de pozos abiertos para ser analizados de forma rápida. A continuación se puede ver en la gráfica el entorno grafico del programa.

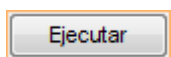
Figura 34: Visualización entorno grafico del programa



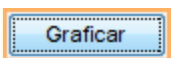
En la pantalla principal del programa se puede apreciar los siguientes accesos:



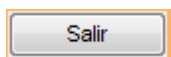
Mediante este botón podremos importar los perfiles planos el cual nos comunica con nos lleva a la interfaz de Carga de Datos.



Este botón Ejecuta el algoritmo método y el nivel de pre-procesamiento seleccionado y muestra los datos en la tabla.

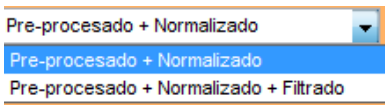


Este botón gráfica los registros desde los valores introducidos de tope y base en los axis.

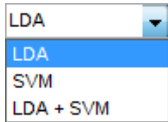


Este es el botón cierra el programa.

Se pueden ver el selector de pre-procesamiento y el selector de método:



Selector de Pre-procesamiento.



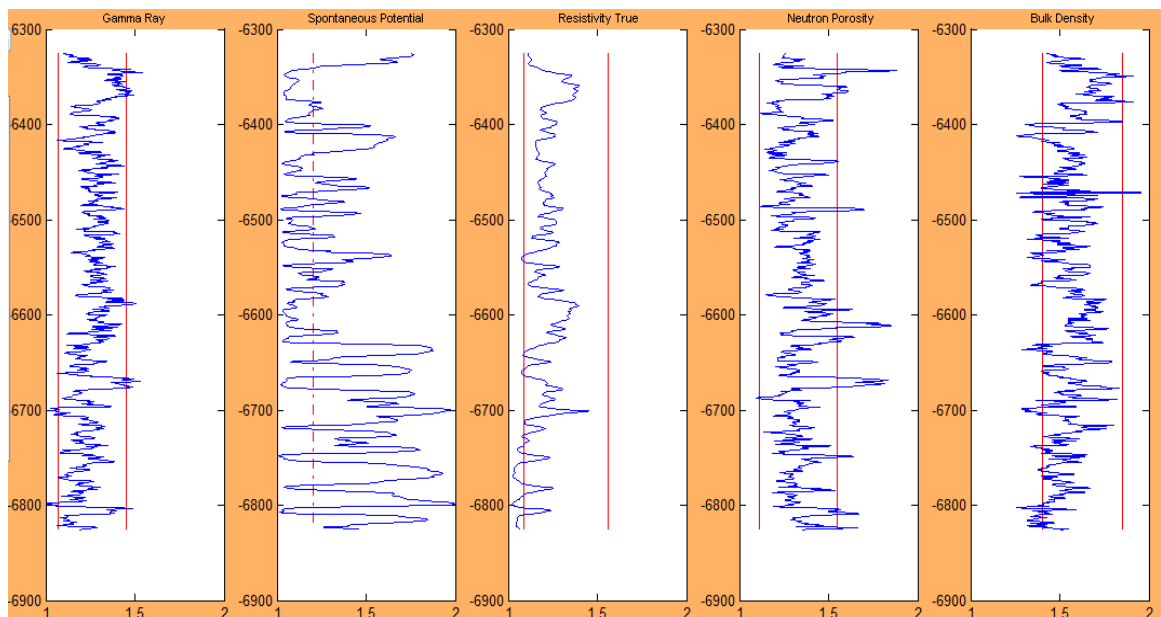
Selector de método.

Se tiene también la visualización en dos partes, que son ilustrados en las Figuras 35 y 36.

Figura 35: Tabla de visualización de las zonas productoras

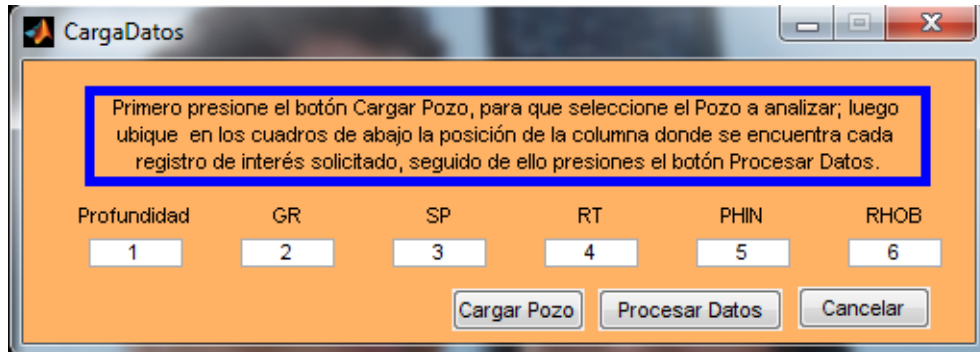
	Tope	Base
1	5.4425e+03	5.4465e+03
2	5.5365e+03	5.5455e+03
3	5571	5.5815e+03
4	5595	5.6005e+03
5	5.6095e+03	5622
6	5656	5.6655e+03
7	5671	5.6795e+03
8	5687	5691
9	5.6925e+03	5705
10	5.7125e+03	5718
11	5719	5724
12	5728	5.7355e+03
13	5766	5.7725e+03
14	5.8025e+03	5806
15	5.8265e+03	5.8305e+03
16	5838	5849
17	5855	5.8595e+03

Figura 36: Axis que ilustran los registros del perfil del pozo



A continuación mostraremos la forma en que se cargan los datos al programa, como se muestra en la Figura 37.

Figura 37: Interfaz de Carga de Datos



En cuya pantalla localizamos el botones:

Cargar Pozo	Carga los perfiles del pozo analizar
Procesar Datos	Procesar los datos del perfil de pozo abierto
Cancelar	Cancela la acción y vuelve a la ventana principal

La Figura 38 ilustra la ventana que se despliega al darle click al botón Cargar Pozo, en la cual localizamos el perfil plano, para ser procesado, por el programa.

Figura 38: Ventana de selección del perfil.

