

SELECCIÓN DE CARACTERÍSTICAS
WAVELET Y RECONOCIMIENTO DE
PATRONES APLICANDO SVM PARA LA
IDENTIFICACIÓN DE ZONAS
POTENCIALMENTE PRODUCTORAS DE
HIDROCARBUROS

**SELECCIÓN DE CARACTERÍSTICAS WAVELET Y RECONOCIMIENTO DE
PATRONES APLICANDO SVM PARA LA IDENTIFICACIÓN DE ZONAS
POTENCIALMENTE PRODUCTORAS DE HIDROCARBUROS**

**JORGE EDUARDO OSSA CORONADO
JAVIER ERNESTO LOPEZ MORENO**

**UNIVERSIDAD SURCOLOMBIANA
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA ELECTRÓNICA
NEIVA – HUILA
2010**

**SELECCIÓN DE CARACTERÍSTICAS WAVELET Y RECONOCIMIENTO DE
PATRONES APLICANDO SVM PARA LA IDENTIFICACIÓN DE ZONAS
POTENCIALMENTE PRODUCTORAS DE HIDROCARBUROS**

**JORGE EDUARDO OSSA CORONADO
JAVIER ERNESTO LOPEZ MORENO**

**Trabajo de investigación presentado como requisito para optar el título de
Ingeniero Electrónico**

**Director:
CARLOS PEREZ
Ingeniero Electrónico**

**UNIVERSIDAD SURCOLOMBIANA
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA ELECTRÓNICA
NEIVA – HUILA
2010**

Nota de aceptación:

Firma del presidente del jurado

Firma del jurado

Firma del jurado

Neiva, 06 de Octubre de 2010

DEDICATORIA

“Agradezco a Dios por mi familia, la vida, por permitirme llegar hasta este momento, por la salud para así tener fortaleza para lograr mis metas. A mis padres María Ademir y Amín Asisclo, porque a través de su esfuerzo me entregaron el mejor regalo: la educación, porque hicieron posible mi gran sueño, haciéndolo suyo también. A mis hermanas Lorena y María Camila a mi hijo Tomas Santiago quien es mi mayor motivación para salir adelante. A los maestros de la Facultad de Ingeniería Electrónica por permitirme ser parte de una generación que busca ser útil a la sociedad; finalmente a todos mis amigos compañeros de estudio por su tiempo y constante apoyo.”

Jorge Eduardo Ossa Coronado

“Dedico este gran logro a Dios por todo lo que me ha entregado y es la fortaleza que me ayuda a seguir luchando, a mis padres, Filemón y Elsa por su constante y desinteresado apoyo, gracias por sus consejos y comprensión, por estar presentes cada momento, sin ellos este sueño no sería una realidad, a mis hermanos Juan Carlos, Julian, Jorge por su cariño y afecto por ser incondicionales, hacen falta palabras para expresarle toda mi gratitud y aprecio, son a ellos a quienes les debo todo, gracias.”

Javier Ernesto López Moreno

AGRADECIMIENTOS

Los autores expresan sus agradecimientos a:

La UNIVERSIDAD SURCOLOMBIANA, por su valiosa colaboración en la realización de este proyecto, su receptividad y apoyo a la investigación científica.

CARLOS PEREZ CAMACHO, Ingeniero Electrónico, director del proyecto, por la oportunidad brindada y por su apoyo durante el desarrollo de este proyecto.

HUGO ALBERTO BERNAL, Ingeniero Electrónico y asesor del proyecto, por su valiosa orientación y constante apoyo en el desarrollo de este proyecto.

A todas aquellas personas quienes de una u otra forma colaboraron y apoyaron en la terminación de éste proyecto.

CONTENIDO

	pág.
INTRODUCCIÓN	18
1. PERFILAJE DE POZOS	20
1.1 TIPOS DE PERFILES DE POZO	20
1.1.1 Potencial Espontáneo (SP)	20
1.1.2 Resistividad	21
1.1.3 Rayos Gamma (GR)	22
1.1.4 Registros de Porosidad	22
1.1.4.1 Registro Neutrónico	22
1.1.4.2 Perfil de Densidad de Formación	22
2. TRANSFORMADA WAVELET	24
2.1 ALGUNAS WAVELETS	25
2.1.1 La Wavelet Daubechies	25
2.1.2 Wavelet Coiflet	26
2.1.3 Wavelets Biortogonales	27
2.1.4 La Wavelet Mexican Hat	28
2.2 TRANSFORMADA WAVELET DISCRETA (DWT)	28
2.2.1 Funciones de Escala y Funciones Wavelets	29
2.3 TRANSFORMADA RAPIDA FWT	30
2.4 TRANSFORMADA WAVELET PACKET	31

2.5	FUNCIONES DE COSTO DE INFORMACION	33
2.5.1	Entropía de Shannon	33
2.5.2	Logaritmo de la Energía	33
2.6	ANALISIS DE COMPONENTES PRINCIPALES PCA	34
3.	MAQUINA DE SOPORTE VECTORIAL (MSV)	35
3.1	CLASIFICACION POR VECTORES DE SOPORTE	35
4.	DESARROLLO METODOLÓGICO	40
4.1	OBTENCION DE DATOS	41
4.2	ANALISIS GRAFICO DE LAS CARACTERISTICAS FISICAS DE LOS PERFILES	41
4.2.1	Estimación de la línea base de lutita para el SP	42
4.2.2	Estimación del perfil de Rayos Gamma GR	42
4.2.3	Estimación de la resistividad verdadera R_t	43
4.2.4	Estimación de la resistividad verdadera R_{xo}	43
4.2.5	Estimación del perfil neutrónico PHIN	43
4.2.6	Estimación del perfil de densidad de formación RHOB	43
4.3	SELECCION DE LA MEJOR ONDA MADRE	44
4.4	NORMALIZACION	46
4.5	IMPLEMENTACIÓN DEL ALGORITMO WAVEPACKET (WP)	46
4.6	REDUCCION DE DIMENSIONALIDAD	48
4.7	SELECCION DE LA POBLACION DE ENTRENAMIENTO	50
4.8	IMPLEMENTACION MAQUINA DE SOPORTE VECTORIAL (SVM)	52
4.8.1	Entrenamiento SVM por parejas y por segmentos ft	53

4.8.2 Entrenamiento de todas las zonas con todos los segmentos ft	53
5. RESULTADOS	55
5.1 OBTENCION ONDA MADRE PARA TODOS LOS POZOS	55
5.2 SELECCION MEJOR ONDA MADRE POR SEGMENTOS	63
5.3 REDUCCION DATOS DEL PCA	64
5.3 RESULTADOS ENTRENAMIENTO	64
5.3.1 Entrenamiento de las Máquinas por parejas de perfiles y segmentos	64
5.3.2 Entrenamiento de todas las zonas	66
5.3.3 Entrenamiento sin el PCA	67
5.3.4 Comparación resultados SVM respecto al método aplicado en el trabajo anterior con RNA.	69
6. CONCLUSIONES	71
BIBLIOGRAFIA	73
ANEXOS	75

LISTA DE FIGURAS

	pág.
Figura 1. Análisis Wavelet	24
Figura 2. La Wavelet de Haar	25
Figura 3. Las Wavelets de Daubechies	26
Figura 4. Symmlets (symN)	26
Figura 5. Coiflets (coifN)	27
Figura 6. Wavelets Biortogonales (biorNr.Nd)	27
Figura 7. Mexican Hat	28
Figura 8. Primera etapa de la DWT	30
Figura 9. DWT Unidimensional	31
Figura 10. Diagrama multiresolución de la señal	31
Figura 11. Árbol de paquetes de descomposición Wavelet	32
Figura 12. Maximización del margen m	36
Figura 13. Transformación de dos conjuntos no separables linealmente	38
Figura 14. Diagrama esquemático general del sistema implementado	40
Figura 15. Representación grafica de los perfiles	41
Figura 16. Representación de la ventana de desplazamiento de cada onda Wavelet	45
Figura 17. Esquema árbol de descomposición	47
Figura 18. Árbol con la menor distribución de información	47
Figura 19. Ordenamiento vector característico como vector fila	48
Figura 20. Ordenamiento vector característico como vector columna	48
Figura 21. Agrupación de las componentes de los perfiles para la reducción	49
Figura 22. Selección de las componentes de los perfiles por el PCA	49
Figura 23. Matriz con los vectores de entrenamiento	50
Figura 24. Generación vectores de soporte	53
Figura 25. Población de entrenamiento con los datos de la matriz del PCA en dos vectores	67
Figura 26. Clasificación población con los datos de la matriz del PCA en dos vectores	67
Figura 27. Población de entrenamiento sin utilización del PCA	68
Figura 28. Población clasificada sin la utilización del PCA	62
Figura 29. Visualización entorno del programa	75
Figura 30. Barra de menú	76
Figura 31. Dirección datos de perfil	76
Figura 32. Menú métodos aplicados	77
Figura 33. Opción método por pares	77
Figura 34. Campo para escoger valor de ventana para la máquina dual	77
Figura 35. Opción por espectro ordenado	77

Figura 36. Campo para escoger valor de ventana para el método por espectro ordenado del vector columna	78
Figura 37. Opción por espectro no ordenado	78
Figura 38. Campo para escoger valor de ventana para el método por espectro no ordenado del vector fila	78
Figura 39. Valores indicando las zonas de interés	79
Figura 40. Inserción zona de interés	79

LISTA DE CUADROS

	pág.
Cuadro 1. Valores típicos de Gamma Ray (<i>GR</i>)	42
Cuadro 2. Valores de ρ_e y ρ_b	44
Cuadro 3. Población de entrenamiento pozo 72 zonas reservorias y no reservorias	51
Cuadro 4. Población de entrenamiento pozo 77 zonas reservorias y no reservorias	51
Cuadro 5. Selección onda madre para el perfil potencial espontáneo con el método de Logaritmo de Energía	55
Cuadro 6. Selección onda madre para el perfil Potencial Espontáneo con el método de Entropía de Shannon	56
Cuadro 7. Selección onda madre para el perfil Gamma Ray con el método de Logaritmo de Energía	56
Cuadro 8. Selección onda madre para el perfil RT con el método de logaritmo de Energía	57
Cuadro 9. Selección onda madre para el perfil RT con el método de entropía de Shannon	58
Cuadro 10. Selección onda madre para el perfil Neutrónico con el método de logaritmo de Energía	58
Cuadro 11. Selección onda madre para el perfil Neutrónico con el entropía de Shannon	59
Cuadro 12. Selección onda madre para el perfil Bulk Density con el método de logaritmo de Energía	59
Cuadro 13. Selección onda madre para el perfil Bulk Density con el método de entropía de Shannon	60
Cuadro 14. Selección onda madre para el perfil Rxo con el método de logaritmo de Energía	61
Cuadro 15. Selección onda madre para el perfil Rxo con el método de entropía de Shannon	61
Cuadro 16. Entrenamiento zonas de diferente longitud	62
Cuadro 17. Vectores característicos a diferentes segmentos	63
Cuadro 18. Vectores característicos a diferentes segmentos	63
Cuadro 19. Eficiencia entrenamiento para zonas de diferente segmentación pozo LL-72	64
Cuadro 20. Eficiencia entrenamiento para zonas de diferente segmentación pozo 77	65
Cuadro 21 Resultados validación de zonas con clasificador neuronal.	69

LISTA DE ANEXOS

	pág.
Anexo 1. TUTORIAL DEL PROGRAMA	75

GLOSARIO

ANALISIS DE COMPONENTE PRINCIPAL (PCA): Método por el cual un determinado grupo de variables es reducido sin perder características fundamentales del grupo original.

BASE ORTONORMAL: Es un conjunto de elementos cuyo espacio vectorial generado es denso, en el que los elementos son mutuamente ortogonales y normales, es decir, de magnitud unitaria.

BASES WAVELET: Funciones generadas a partir de una función prototipo.

CARACTERÍSTICA: Medida escalar de los atributos de un objeto.

CLASES: Agrupación de datos con características similares reconocibles mediante técnicas de reconocimiento de patrones.

CLASIFICADOR: Sistema de reconocimiento de patrones.

COEFICIENTES DE APROXIMACION: Secuencia producida mediante un filtro pasa bajas asociado a cierta wavelet.

COEFICIENTES DE DETALLE: Secuencia producida mediante un filtro pasa altas asociado a cierta wavelet.

CONJUNTO DE ENTRENAMIENTO: Conjunto de pares de entrenamiento consistente de un vector de valores de entrada junto con un vector 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.

DESCOMPOSICIÓN: Proceso de dividir una señal en dos o más componentes aditivas.

ENTROPÍA: Se interpreta como la medida del desorden de un sistema, es la magnitud que mide la información contenida en un grupo de datos.

FUNCIONES KERNEL: Medida de similitud originada a partir de una representación particular de los patrones.

LITOLOGIA: Parte de la geología que trata sobre el estudio de las rocas, su tamaño de grano, del tamaño de las partículas y de sus características físicas y químicas.

LUTITA O SHALE: Es una roca detrítica, es decir, formada por detritos, y está integrada por partículas del tamaño de la arcilla y del limo, cuyo componente mayor es la arcillas. Las arcillas son esencialmente silicatos hidratados de aluminio formados por hidrólisis natural o acción atmosférica (meteorización) de los feldespatos y varios aluminios silicatos alcalinos originarios de rocas ígneas.

MAQUINA DE SOPORTE VECTORIAL: algoritmo de aprendizaje para clasificación de patrones, el principio básico es encontrar un hiperplano lineal (o no lineal) óptimo tal que sea minimizado el error de clasificación esperado para un conjunto de validación.

ORTOGONALIDAD: Es una generalización de la noción geométrica de perpendicularidad.

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

PERFILAJE: Actividad relevante en la exploración y producción de hidrocarburos la cual consiste en la toma y monitoreo de los perfiles o registros del pozo.

PERMEABILIDAD: Es la propiedad que posee la roca para permitir que los fluidos se muevan a través de sus poros.

POROSIDAD: La porosidad es una propiedad de la roca y es definida como el porcentaje del volumen total de la misma. Esta es expresada en porcentaje y de acuerdo a que tanto por ciento se tenga de porosidad se puede saber que tanto fluido puede almacenar dicha roca.

REGISTRO: Es una representación grafica en función de la profundidad de alguna de las características de las formaciones rocosas atravesadas, son respuestas de los diferentes herramientas de medición mientras desciende este instrumento.

REPRESENTACIÓN: Selección básica y fundamental de una señal, sin pérdida de información relevante de la información original.

SEÑALES NO ESTACIONARIAS: Señales con contenido frecuencial variante en el tiempo.

ZONA RESERVORIA: Corresponde al subsuelo que contiene en sus espacios porosos hidrocarburos líquidos o gaseosos con características que permiten su explotación comercial.

RESUMEN

Este trabajo presenta un estudio centrado en el tratamiento digital de señales obtenidas a través del perfilaje de pozos para analizar y detectar zonas potencialmente productoras de hidrocarburos.

Para el análisis de los registros de los perfiles se inicia con una primera fase donde se realiza un procesamiento de la señal utilizando la Transformada Wavelet, que hace una compresión de los datos para obtener la mejor caracterización de los perfiles. En esta etapa se aplica el método del algoritmo de la mejor base Wavelet Packet, utilizando dos funciones de costo de información realizando una comparación entre ellas (Logaritmo de la Energía, Entropía de Shannon). En la segunda fase se realiza una extracción de características para reducir dimensionalidad del conjunto de datos aplicando una técnica denominada, Análisis de Componentes Principales (PCA).

En la última fase se implementó la Máquina de Soporte Vectorial (SVM) haciendo uso de sus algoritmos de entrenamiento y clasificación para obtener mejores resultados que los ya alcanzados en el trabajo anterior con Redes Neuronales Artificiales.

El resultado de este trabajo es un entorno realizado en Matlab 7.6 donde los registros derivados de los perfiles de Resistividad, Gamma Ray, Potencial Espontáneo, Porosidad (Bulk Density y neutrónico), son cargados para evaluar y clasificar las zonas potencialmente productoras de interés calificadas por la (SVM).

ABSTRACT

This work presents a study focused on the treatment digital signal obtained through the profiling of wells to analyze and identify areas potentially producing hydrocarbons.

For the analysis of the records of the profiles begins with the first stage where there is a signal processing using the Transformed Wavelet, which makes a compression of the data to get the best characterization of the profiles. At this stage applies the algorithm method of the best basis Wavelet packet, using two functions of the cost of information making a comparison between them (logarithm of Energy, Entropy Shannon). In the second phase is a feature extraction to reduce dimensionality set of data using a technique called, Analysis of Major Components (PCA).

In the last phase was implemented the Machine of Support Vector (SVM) making use of their algorithms for training and classification to get better results than those already achieved in the previous work with Neural Networks Artificial.

The result of this work is an environment done in Matlab 7.6 where records derived from the profiles of Resistivity, Gamma Ray, Potential Spontaneous, Porosity (Bulk Density and neutron), are loaded to evaluate and classify areas potentially producing interest qualified by the (SVM).

INTRODUCCIÓN

Actualmente en los estudios o análisis para la detección de hidrocarburos es fundamental la utilización de métodos para la interpretación de perfiles, donde se miden propiedades eléctricas, acústicas, radioactivas y electromagnéticas de las rocas, para obtener la mayor cantidad de información del subsuelo; cada uno de estos perfiles entrega un conjunto de datos en función de sus capacidades físicas, esto se logra por medio de diferentes herramientas que son enviadas al interior del pozo. La información obtenida por los instrumentos de registro es enviada a la superficie y posteriormente procesada para conseguir el perfil del pozo en tiempo real.

Gracias al desarrollo tecnológico se han modernizado las herramientas o instrumentos que se bajan al pozo logrando mejores resultados que permiten datos mucho más exactos, en años recientes las empresas exploradoras han hecho uso de equipos portátiles computacionales que registran el proceso. El objetivo fundamental de este trabajo es interpretar los registros de pozo abierto para poder evaluar y caracterizar la zona reservoria.

En la industria del petróleo existen diversas maneras para el análisis de estos datos, muchas empresas utilizan personas para lograr un análisis e interpretación del perfil, donde no se hace uso de técnicas computacionales para llevar a cabo soluciones a la hora de establecer zonas de interés. El reto de la industria petrolera debido a la gran demanda de energía que requiere el planeta es la búsqueda de nuevos yacimientos además de hacer un máximo esfuerzo en la recuperación de los campos petroleros ya explorados de una manera rápida y segura.

En el escenario actual, nuevas técnicas se ponen al servicio de la investigación, de ahí la importancia de utilizar algoritmos que realicen un correcto procesamiento de la información que es obtenida mediante el perfilaje de pozos, en donde se pretende buscar la manera de lograr un mejor análisis de los datos obtenidos. Aparece como opción importante la Transformada Wavelet como herramienta para la caracterización de los datos de perfilaje, en donde la señal es representada en una escala tiempo-frecuencia lo que permite identificar rasgos distintivos que de otra manera serian imposibles de identificar. En años recientes ha aparecido una herramienta computacional generando buenas expectativas en tareas de clasificación de patrones. Esta técnica denominada Maquina de Soporte Vectorial surge como instrumento para la identificación y clasificación de zonas potencialmente productoras.

El propósito de este trabajo consiste en el procesamiento de señales obtenidas mediante el perfilaje de pozo abierto, que fueron proporcionadas por ECOPETROL

para su desarrollo. El proyecto se limita a un estudio para tratar de obtener un avance o mejora en relación con el trabajo desarrollado con Redes Neuronales Artificiales (RNA), llamado IDENTIFICACION DE ZONAS PRODUCTORAS DE HIDROCARBURO APLICANDO DSP Y METODOS DE CORRELACION DE REGISTROS DE PERFILAJE. Se requiere conseguir una herramienta que sea de utilidad para la obtención de resultados más precisos y confiables para la identificación y clasificación de zonas de interés.

En este trabajo investigativo el eje de desarrollo son los datos obtenidos para su posterior procesamiento mediante la Transformada Wavelet, ya que por su gran eficacia en el área del tratamiento de señales se incluyó como herramienta de aplicación; seguidamente se hace uso de otra técnica denominada Análisis de Componente Principal (PCA) como un reductor de dimensionalidad, finalmente la Máquina de Soporte Vectorial como clasificador y así conjuntamente conseguir una buena obtención de resultados y determinar las probabilidades de existencia de hidrocarburos en el reservorio.

1. PERFILAJE DE POZOS

El petróleo se detecta mediante el estudio del subsuelo por detección remota y bajando sensores dentro del pozo. El perfilaje de pozos hace el monitoreo de los perfiles mediante dispositivos que registran los datos para más tarde transmitirlos a la superficie. Cada uno de estos métodos estudia la tierra a diferentes profundidades y a diferentes escalas desde la superficie hasta kilómetros bajo tierra, un perfil es una grabación contra profundidad de alguna de las características de las formaciones rocosas atravesadas. La exploración petrolera necesita una gran cantidad de recursos donde el costo de la perforación es compensado por el hallazgo de hidrocarburos. La importancia de esta actividad es que a través de los perfiles se registran una gran variedad de características de la formación, como propiedades geológicas y petrofísicas de las rocas que se encuentran a lo largo de la perforación. Estos datos informan acerca de los fluidos que están alojados en los poros de las rocas (agua, petróleo y gas).

1.1 TIPOS DE PERFILES DE POZO

Estos tipos de registros entregan información acerca de las propiedades eléctricas acústicas, radioactivas y electromagnéticas de las rocas, estas propiedades físicas se pueden interpretar en términos de litología, porosidad y saturación de hidrocarburos.

1.1.1 Potencial Espontáneo (SP). El *SP* es el registro de un fenómeno natural que ocurre en las formaciones atravesadas por un pozo, es el voltaje que aparece entre dos electrodos, uno que se desplaza por todo el pozo y otro que se encuentra fijo en la superficie.

Aplicaciones *SP*¹:

- Correlación.
- Identificación de cambios entre agua dulce y agua salada.
- Indicador de arcillosidad.
- Determinación de la resistividad de agua de formación, (*RW*).
- Indicio de permeabilidad.

Estas señales de voltaje eléctrico se deben a las diferentes variaciones de la salinidad entre los fluidos. En las gráficas de este registro, la curva de (*SP*) registra el potencial eléctrico producido por la interacción del agua de formación, lodo de perforación (conductivo) y lutitas².

¹ SCHLUMBERGER. Potencial Espontáneo (SP), p.5

² OSORIO, Rafael. Registro SP. UNI Perú, p.1

La corriente se origina a partir del movimiento iónico presente en los espacios vacíos de las rocas porosas y que es detectada por la herramienta que se introduce dentro del pozo. Cuando esta herramienta se encuentra en una zona de arcillas, esta zona no puede invadirse de lodo por lo que el electrodo entregará una información de cero voltios; en otro caso, cuando el electrodo móvil se encuentra en una zona permeable, habrá una diferencia de potencial no nula en las gráficas de los perfiles, las curvas se mueven hacia la izquierda o hacia la derecha en referencia de una línea base, llamada la línea base de shale. Las deflexiones hacia la derecha o hacia la izquierda de esta línea, dependen de la salinidad del agua de formación. Cuando la curva se mueve hacia la izquierda se considera agua salada en la formación y cuando se encuentra a la derecha habrá agua dulce. El SP³ se utiliza para la identificación de capas porosas, para calcular la salinidad del agua de formación y la resistividad del agua de formación (R_w).

1.1.2 Resistividad. La resistividad es la capacidad que tienen las rocas de oponerse al paso de corriente eléctrica inducida y es el inverso de la conductividad. En una formación la capacidad de conducir electricidad está relacionada con la resistividad de agua de la formación, la cantidad de agua presente y la geometría estructural y tamaño de los poros proporcionando 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. Las rocas compactas poco porosas como las calizas masivas poseen resistividades altas.

Aplicaciones Resistividad⁴:

- Cálculo de saturación de agua.
- Identificación del tipo de agua.
- Análisis de invasión.
- Correlación.
- Identificación de litología.

El parámetro resistivo de mayor interés es la resistividad verdadera R_t , debido a que está relacionado con la saturación de hidrocarburos. La resistividad de la zona lavada, R_{xo} , es también de mucho interés, ya que con la comparación de R_t y R_{xo} se puede tener una idea de la movilidad de los hidrocarburos⁵.

³ www.scribd.com. Registro o Perfilaje de Pozos, p.3

⁴ SCHLUMBERGER. Resistividad, p.4

⁵ Andrade Trujillo, Guillermo. Polanco, Juan Manuel. Evaluación de Algoritmos Basados en Lógica Difusa para el Procesamiento de Señales de Registro de Pozo abierto. Universidad Surcolombiana, 2009. P 26

1.1.3 Rayos Gamma (GR). La herramienta de GR mide las emisiones naturales de rayos gamma que poseen las rocas. Durante la meteorización de las rocas, los elementos radiactivos que estas contienen se desintegran en partículas de tamaño arcilla, por lo tanto las lutitas tienen emisiones de rayos gamma mayores que las arenas. Mientras mayor es el contenido de arcilla de las rocas mayor es la emisión de GR de las mismas. Los minerales radiactivos principales son: el potasio (K), el torio (Th) y el uranio (U) cuyos rayos emitidos espontáneamente son detectados por la herramienta de GR. 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.

Aplicaciones (GR)⁶:

- Correlación.
- Indicador de litología.
- Evaluación cuantitativa de arcillosidad.

1.1.4 Registros de porosidad. Proporcionan información acerca de la porosidad del yacimiento. Son los mejores perfiles para detectar y delimitar los yacimientos de gas.

1.1.4.1 Registro Neutrónico. Se basa en la medición de concentraciones de hidrógenos, lo que indica la presencia de agua o petróleo de la roca. Posee una fuente de neutrones, los cuales colisionan con los hidrógenos presentes en los poros de la roca. La herramienta también posee un receptor que mide los neutrones dispersos liberados en las colisiones. La herramienta se llama CNL, sirve para estimar la porosidad neutrónica de las rocas (NPHI). Si el registro neutrónico es alto indica alta índice de neutrones, y si es bajo indica bajo índice de neutrones. 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 %).

1.1.4.2 Perfil de Densidad de Formación. Esta técnica es utilizada principalmente para determinar la porosidad de la formación. Una fuente radioactiva que emite radiación gamma, es colocada en una almohadilla y aplicada contra la pared del pozo. Esta fuente emite hacia la formación rayos gamma, los cuales interactúan con los electrones de la formación según el efecto Compton, según el cual el resultado del choque de un rayo gamma con un electrón es la pérdida de energía en el rayo gamma y la salida del electrón. Los rayos son dispersados por los núcleos de la formación de donde se obtienen rayos gamma de Compton, que es una radiación secundaria producida en los átomos de la formación y que se originan porque la formación cede energía a los átomos dejándolos en estado excitados⁷.

⁶ SCHLUMBERGER. Rayos Gamma (GR), p.7

⁷ FALLA VILLEGAS, Elías Jhon. Interpretación de Registros de Pozos de Petróleo. Tesis Digitales UNMSM, p.4

Estos últimos rayos son detectados y evaluados como una medida de la densidad de la formación, ya que el número de rayos gamma compton está directamente relacionado con el número de electrones de la formación.

Requisitos para medir la densidad de formación⁸:

- La fuente debe emitir rayos gamma al nivel de energía de compton scattering.
- La distancia fuente-detector debe permitir interacción con electrones de la formación.
- La distancia fuente-detector tan grande que los rayos gamma pierdan la energía antes de llegar al detector.

⁸

SCHLUMBERGER. Densidad, p.8

2. TRANSFORMADA WAVELET

En este estudio se analizan las señales que entregan las herramientas de perfilaje las cuales son no estacionarias, discontinuas, donde su frecuencia esta variando continuamente; por este motivo es fundamental para este proceso, la Transformada Wavelet. Esta transformada se utiliza para obtener información de la señal y permitir su caracterización.

En el procesamiento de señales se pueden encontrar diferentes tipos de señales estacionarias y no estacionarias. En el caso de las señales de comportamiento no estacionario, surge la necesidad de contar con una representación localizada en el tiempo y la frecuencia, es así como se introduce el análisis Wavelet como herramienta que permite obtener dicha representación.

Figura 1. Analisis Wavelet



Fuente: Wavelet Toolbox 4 User's Guide

Una Wavelet es una función integrable y oscilatoria cuya media es 0.

$$\int \varphi(t) dt = 0 \quad (\text{Ecuación 1})$$

Las Wavelet son funciones matemáticas que separan la información en diferentes componentes frecuenciales, a través del análisis de multiresolución con ventanas de longitud variable, adaptadas al cambio de frecuencia de la señal.

El análisis Wavelet adopta una función madre, el análisis temporal lo realiza una versión contraída y de alta frecuencia de esta función, mientras el análisis frecuencial lo realiza una versión dilatada y de baja frecuencia de la misma. La Transformada Wavelet continua de una función f es:

$$C(a, b) = \int_{-\infty}^{+\infty} f(t) \varphi_{a,b}^*(t) dt \quad (\text{Ecuación 2})$$

donde

$$\varphi(t)_{a,b} = \frac{1}{\sqrt{a}} \varphi\left(\frac{t-b}{a}\right), \quad a, b \in R, a > 0 \quad (\text{Ecuación 3})$$

Los parámetros a y b pueden ser variados en forma continua (CWT) o en forma discreta (DWT).

$$a = a_0^m, \quad b = nb_0a_0^m \quad \text{con } m, n \in \mathbb{Z}$$

$$a_0 > 1, b_0 > 0$$

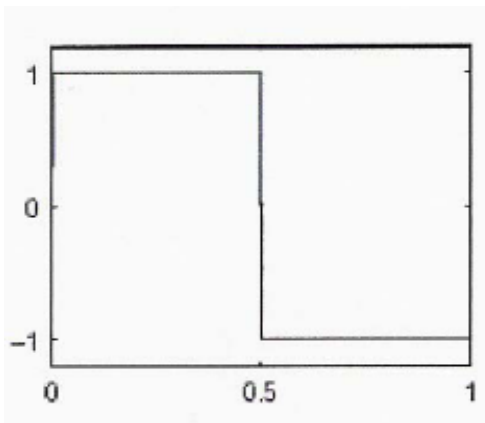
Donde a es el parámetro de escalamiento o dilatamiento y b es el desplazamiento o traslación. Entonces, esta familia de funciones son una copia de una función prototipo $\varphi(t)$, denominada Wavelet madre, trasladada y escalada mediante las variables a y b . Existen diferentes Wavelets que ya son utilizadas de forma constante y que tienen definiciones establecidas, la elección de un tipo de Wavelet depende de la aplicación específica que se le vaya a dar.

2.1 ALGUNAS WAVELETS

2.1.1 La Wavelet Daubechies. La familia Daubechies, tiene la notación 'dbN' donde N indica el orden y $N \in \mathbb{Z}$. Esta Wavelet posee soporte compacto, ortogonalidad, regularidad y continuidad. La Wavelet de Haar o Daubechies de orden 1, es la primera y mas sencilla de las Wavelet, aunque tiene soporte compacto no tiene buena localización tiempo-frecuencia, esta Wavelet no es continua, y por ende no diferenciable.

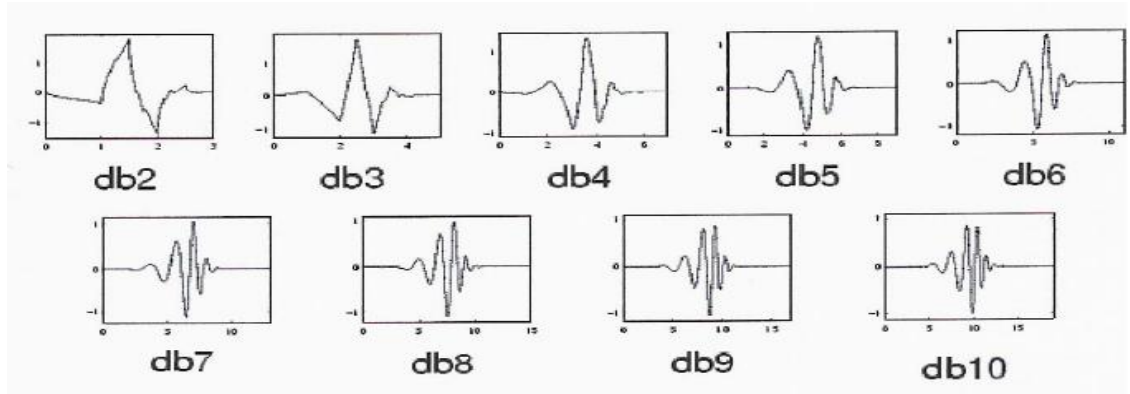
$$h(t) = \begin{cases} 1: 0 < t < 1/2 \\ -1: 1/2 < t < 1 \\ 0: \text{otherwise} \end{cases} \quad \text{(Ecuación 4)}$$

Figura 2. La Wavelet de Haar



Fuente: NIETO, Natalia y OROZCO Diana M. El uso de la Transformada Wavelet discreta en la reconstrucción de señales senosoidales.

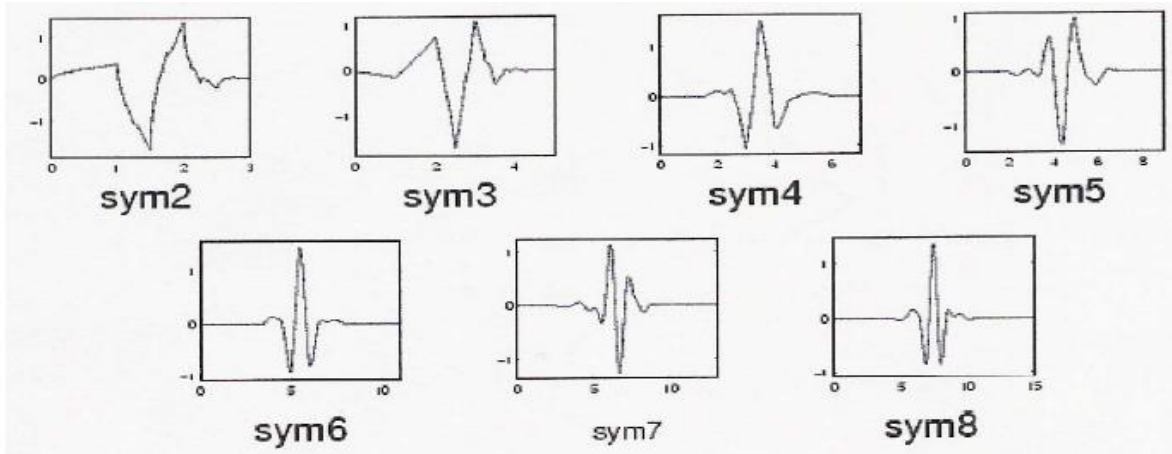
Figura 3. Las Wavelets de Daubichies (dbN)



Fuente: NIETO, Natalia y OROZCO Diana M. El uso de la Transformada Wavelet discreta en la reconstrucción de señales senosoidales.

Con el fin de agregar simetría a sus Wavelets, Daubechies creó la siguiente familia de Wavelets: Symmlets. Esta familia es de soporte compacto y puede realizar la transformada continua y discreta, las Symmlets pueden ser ortogonales, biortogonales y están cerca de ser simétricas. Su apariencia es la siguiente.

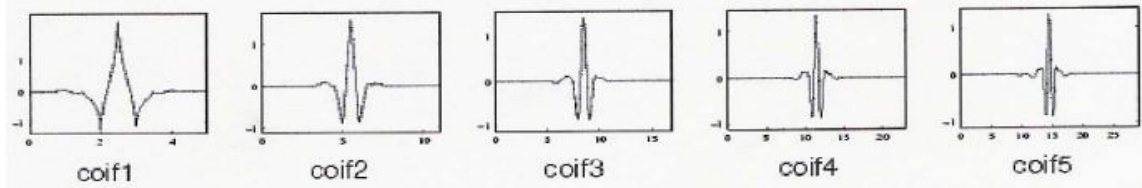
Figura 4. Symmlets (symN)



Fuente: NIETO, Natalia y OROZCO Diana M. El uso de la Transformada Wavelet discreta en la reconstrucción de señales senosoidales.

2.1.2 Wavelet Coiflet. Estas wavelets fueron igualmente creadas por Daubechies con ayuda de Coifman. La función Wavelet y la de escalamiento son mucho más simétricas que las Wavelets presentadas con anterioridad, poseen igualmente soporte compacto y el número de filtros es $6N$. Su apariencia es de la siguiente manera:

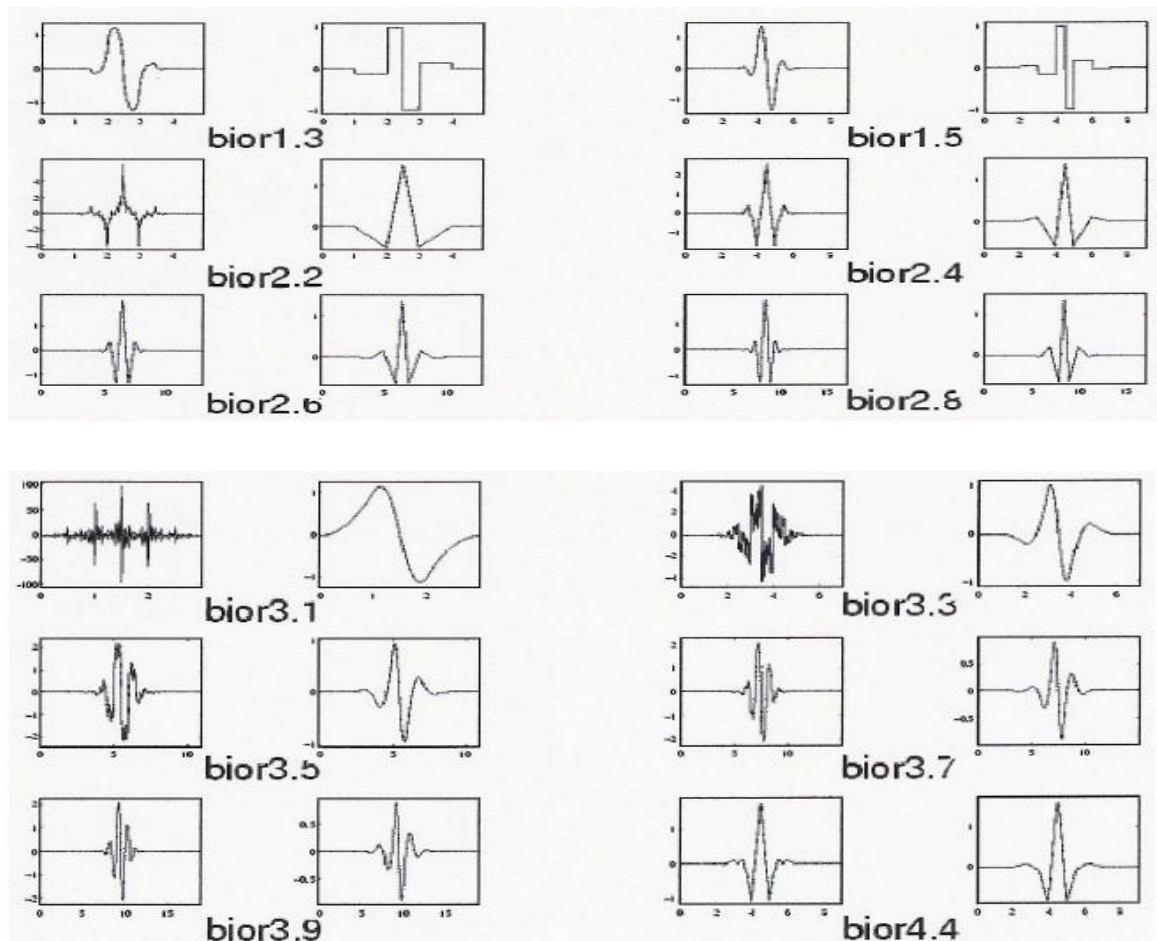
Figura 5. Coiflets (coifN)



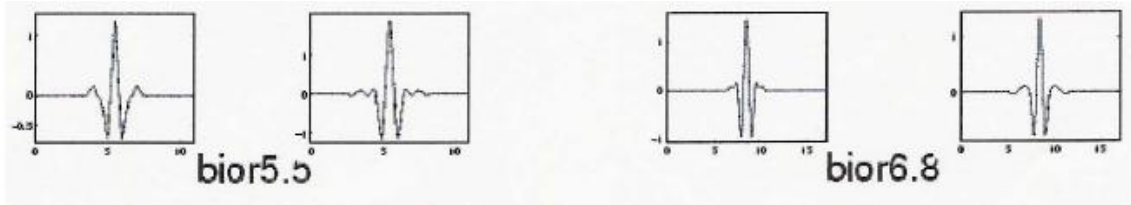
Fuente: NIETO, Natalia y OROZCO Diana M. El uso de la Transformada Wavelet discreta en la reconstrucción de señales senosoidales.

2.1.3 Wavelets Biortogonales. Poseen soporte compacto y la simetría, así como la reconstrucción exacta de la señal son posibles con filtros FIR (Finite Impulse Response), lo cual en las wavelets ortogonales es imposible excepto en la Wavelet Haar. El orden de estas Wavelets está dado por N_r y N_d , la primera para la reconstrucción y la segunda para la descomposición. Son de la siguiente manera:

Figura 6. Wavelets Biortogonales (biorNr.Nd)



Continuación Figura 6.



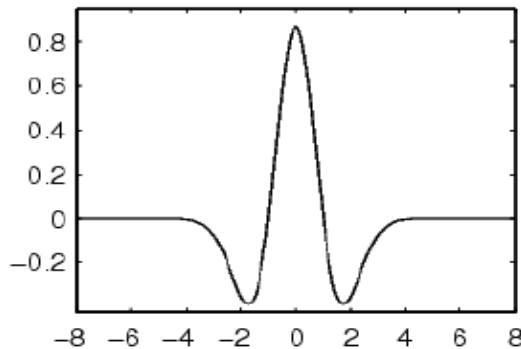
Fuente: NIETO, Natalia y OROZCO Diana M. El uso de la Transformada Wavelet discreta en la reconstrucción de señales senosoidales.

2.1.4 La Wavelet Mexican Hat. Otra Wavelet muy recurrida es para el análisis de señales es la Wavelet Mexican Hat por la forma de su gráfica y es la segunda derivada de la función de densidad de probabilidad Gaussiana⁶:

$$mex(t) = \frac{2(1-t^2)e^{-t^2/2}}{\pi^{1/4}\sqrt{2}} \quad (\text{Ecuación 5})$$

Su apariencia es la siguiente:

Figura 7. Mexican Hat



Fuente: NIETO, Natalia y OROZCO Diana M. El uso de la Transformada Wavelet discreta en la reconstrucción de señales senosoidales.

2.2 TRANSFORMADA WAVELET DISCRETA (DWT)

Debido a la variabilidad en forma continua de los parámetros de escala como de traslación, es necesario contar con una herramienta que permita discretizar la señal. Se pasará de un mapeo continuo a un conjunto de finitos valores, a través del cambio de la integral por una aproximación con sumatorias⁹.

⁹ NIETO, Natalia y OROZCO Diana M. El uso de la Transformada Wavelet discreta en la reconstrucción de señales senosoidales. Universidad Tecnológica de Pereira, p.381

La discretización permite representar una señal en términos de funciones elementales acompañadas de coeficientes.

$$f(t) = \sum_{\lambda} C_{\lambda} \varphi_{\lambda}$$

Es posible representar una señal $f(t)$ como una sumatoria de funciones Wavelet y funciones de escala:

$$f(t) = \sum_k \sum_j C_{j,k} \phi(t) + \sum_k \sum_j d_{j,k} \varphi(t) \quad (\text{Ecuación 6})$$

2.2.1 Funciones de Escala y Funciones Wavelets. Una forma de discretizar los parámetros de escala y frecuencia es mediante un muestreo exponencial, para garantizar una mejor aproximación, donde se pueden redefinir los parámetros a valores discretos de la siguiente manera:

$$a = s^{-j} \quad u = kns^{-j}$$

De esta manera y reemplazando en la Ecuación 3 se obtiene la familia de funciones discretizadas, que constituyen bases ortonormales de Wavelets en $L^2(R)$.

$$\begin{aligned} \varphi_{j,k}(t) &= \frac{1}{\sqrt{s^{-j}}} \varphi\left(\frac{t - kns^{-j}}{s^{-j}}\right) \\ &= s^{\frac{j}{2}} \varphi(s^j t - kn) \end{aligned} \quad (\text{Ecuación 7})$$

Para obtener una mejor aproximación de la señal en niveles de resolución muy finos, es necesario que las Wavelet sean dilatadas por un factor de 2^{-j} , permitiendo tener una resolución de 2^j , estas funciones son denominadas Wavelets diádicas.

$$= 2^{\frac{j}{2}} \varphi(2^j t - kn) \quad j, k \in Z \quad (\text{Ecuación 8})$$

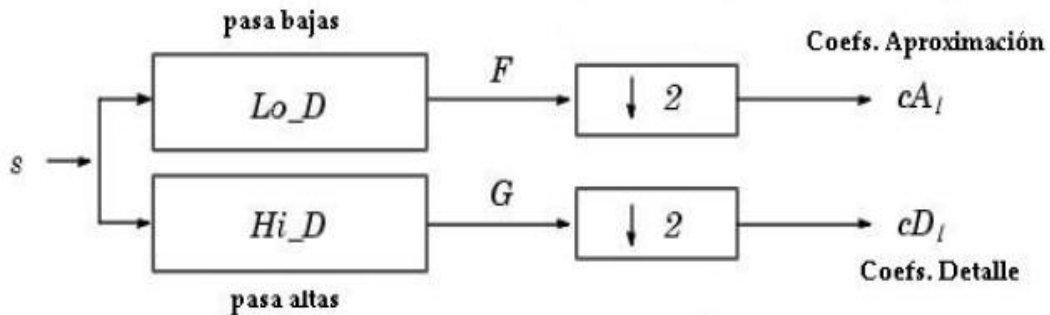
Teniendo en cuenta la Ecuación 2, la Transformada Discreta Wavelet tiene la forma:

$$DWTf(j, k) = \int_{-\infty}^{+\infty} f(t) 2^{\frac{j}{2}} \varphi(2^j t - kn) dt \quad (\text{Ecuación 9})$$

2.3 TRANSFORMADA RAPIDA FWT

Al usar la técnica Wavelet como una herramienta de extracción de características se utilizan los coeficientes de descomposición de aproximación y detalle para formar un vector de características que representen adecuadamente los datos. Dada una señal s de longitud N la Transformada Wavelet Discreta (DWT), consiste de $\log_2 N$ etapas. La primera etapa partiendo de s , produce dos conjuntos de coeficientes: coeficientes de aproximación CA_1 , y coeficientes de detalle CD_1 . Estos vectores son obtenidos mediante la convolución del filtro pasa bajas Lo_D para la aproximación.

Figura 8. Primera etapa de la DWT



Fuente: OROZCO ALZATE Mauricio. Clasificación de arritmias Cardíacas usando Transformada Wavelet y Técnicas de reconocimiento de Patrones.

Y con el filtro pasa altas Hi_D , para los detalles, seguido por una decimación diádica. La longitud de cada filtro es $2N$. si n es igual a la longitud de s las señales F y G son de longitud $n + 2N - 1$ y los coeficientes CA_1 y CD_1 son de longitud:

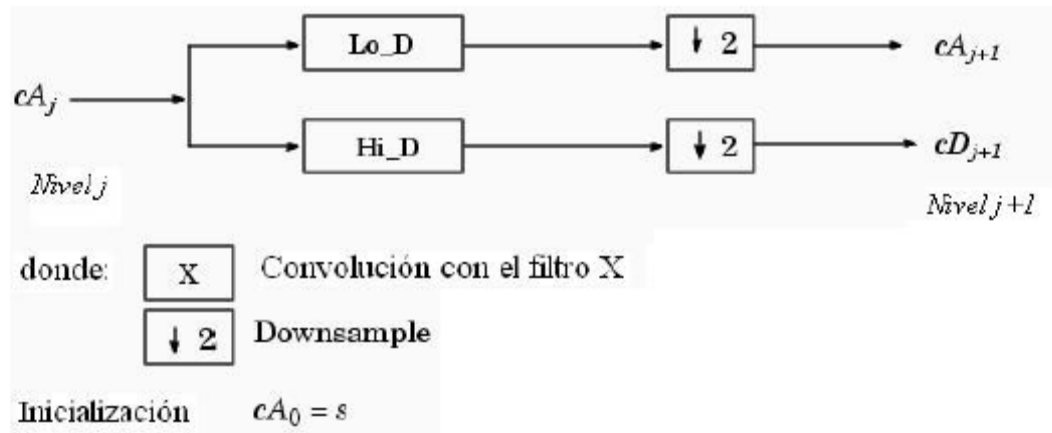
$$\text{floor}\left(\frac{n-1}{2}\right) + N$$

El siguiente paso particiona los coeficientes de aproximación CA_1 en dos partes usando el mismo esquema, reemplazando s por CA_1 , produciendo CA_2 y CD_2 , y así sucesivamente. La descomposición Wavelet de la señal s en el nivel j tiene la estructura¹⁰:

$$[CA_j, CD_j, \dots, CD_1]$$

¹⁰ OROZCO ALZATE Mauricio. Clasificación de arritmias Cardíacas usando Transformada Wavelet y Técnicas de reconocimiento de Patrones. Universidad Nacional de Colombia en Manizales, p.57

Figura 9. DWT Unidimensional



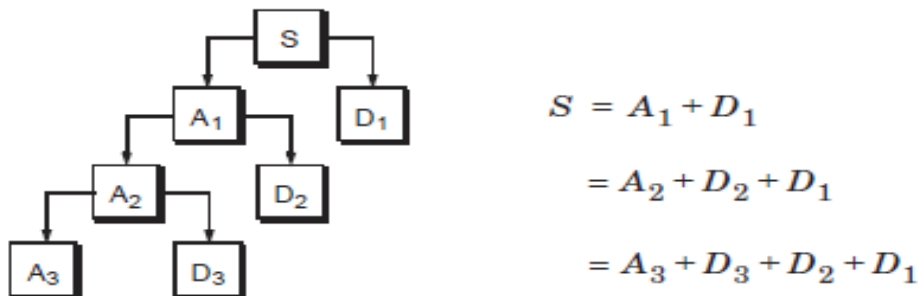
Fuente: OROZCO ALZATE Mauricio. Clasificación de arritmias Cardíacas usando Transformada Wavelet y Técnicas de reconocimiento de Patrones.

2.4 TRANSFORMADA WAVELET PACKET

El método de paquetes Wavelet es una generalización de la descomposición wavelet que ofrece una gama de posibilidades para el análisis de la señal, llamadas bases wavelet, cuya proyección sobre la señal f forma los coeficientes, y dentro de esas bases se escogen aquellas que convengan para la tarea en particular que se realice.

En el análisis Wavelet, una señal se divide en una aproximación y un detalle. La aproximación es luego dividida en una aproximación de segundo nivel y el detalle, y se repite el proceso. Para una descomposición de n-nivel, hay n + 1 posibles maneras de descomponer o codificar la señal.

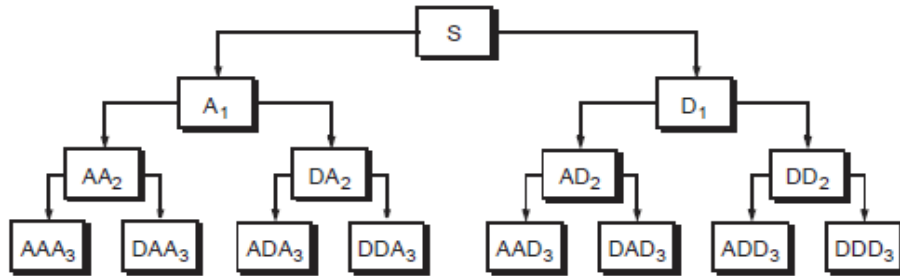
Figura 10. Diagrama multiresolución de la señal



Fuente: Wavelet Toolbox 4 User's Guide

En el análisis Wavelet Packet, los detalles, así como aproximaciones se pueden dividir. Esto produce $2^{2^{n-1}}$ maneras diferentes para codificar la señal. Este es el árbol de paquetes de descomposición Wavelet¹¹.

Figura 11. Árbol de paquetes de descomposición Wavelet.



Fuente Wavelet Toolbox 4 User's Guide

Los coeficientes de aproximación como los de detalle son descompuestos para cada nivel, procedimiento que genera una estructura de árbol¹². El árbol de descomposición wavelet es una parte del árbol binario completo.

Por ejemplo, el análisis de paquetes wavelet permite que la señal S que se representa como $A1 + AAD3 + DAD3 + DD2$.

Este es un ejemplo de una representación que no es posible con el análisis Wavelet ordinario.

La escogencia las bases Wavelet que convengan para la tarea en particular, puede hacerse dentro de un diccionario de ondas Wavelet. La transformada Wavelet Packet (WP) ha sido usada para obtener un diccionario de la cual brinda una mayor cantidad de coeficientes correspondientes a una mayor cantidad de bases, con lo cual se incrementan las opciones de selección de ondas para la tarea predefinida.

Cada nivel de descomposición entrega una partición del plano tiempo-frecuencia que al unirlo con las particiones de los otros niveles de descomposición forman una gran familia de particiones del plano tiempo-frecuencia. Así, se mejora enormemente las opciones de escogencia para los algoritmos de selección de bases de representación y clasificación.

¹¹ CASTELLANOS German. Determinación de voces disfónicas usando bases Wavelet discriminantes. Universidad Tecnológica de Pereira, p.3

¹² Wavelet toolbox 4 User's Guide. p 59

2.5 FUNCIONES DE COSTO DE INFORMACION

Este método consiste en definir una función de costo en las secuencias y buscar su valor mínimo sobre todas las bases de la librería, dando esta cifra su más eficiente representación o mejor base. Basado en la filosofía de “divide y vencerás” se puede restringir la búsqueda sobre las funciones que parten de una manera suave sobre productos cartesianos dando como resultado un algoritmo rápido.

Las funciones de costo de la información se pueden definir de la siguiente manera:

Un mapa M de secuencias $\{X_i\}$ en R es una función de costo aditiva si:

$$M(0) = 0$$

y

$$M(\{X_i\}) = \sum_i M(X_i)$$

Esta condición garantiza tener un vector con la lista más económica de coeficientes.

2.5.1 Entropía de Shannon. Se define la entropía de Shannon-Weaver de una secuencia $X = \{X_i\}$

$$H(x) = \sum_j P_j \log P_j , \quad (\text{ecuación 10})$$

10)

Donde

$$P_j = \frac{|x_j|^2}{\|x\|^2} \quad (\text{ecuación 11})$$

Se determina que $P \log p = 0$ si $p = 0$, Esta no es una función de costo aditiva, pero la función $\lambda(x) = -\sum_j |X_j|^2$ sí lo es. La relación $H(x) = \|x\|^{-2} \lambda(x) + \log \|x\|^2$ permite, mediante su minimización, minimizar la ecuación de entropía. Un aspecto interesante acerca de la entropía es que $\text{Exp}H(x)$ es proporcional al número de coeficientes necesarios para representar una señal a un error cuadrático medio fijo.

2.5.2 Logaritmo de la energía.

$$M(x) = \sum_j \log|x_j|^2 \text{ con } \log(0) = 0 \quad (\text{ecuación 12})$$

Se determina cada vez que sea necesario. Esta función de costo aditiva puede interpretarse como la entropía de un proceso Gauss-Markov compuesto por N variables aleatorias gaussianas no correlacionadas de varianzas $\vartheta_1^2 = |x_1|^2, \dots, \vartheta_N^2 = |x_N|^2$ la redefinición del logaritmo en cero es equivalente a ignorar cualquier componente que no cambia en el proceso. Al minimizar esta función se encuentra la mejor aproximación a la base del proceso, lo cual da un mínimo global de M sobre todo el grupo ortogonal¹³.

2.6 ANALISIS DE COMPONENTES PRINCIPALES PCA

En el análisis de datos un inconveniente es la reducción de la dimensionalidad; el análisis de componentes principales es una técnica estadística de simplificación o reducción que permite transformar un conjunto de variables no correlacionadas. Es decir, es posible describir los valores de p variables por un pequeño subconjunto $r < p$ de ellas, donde se habrá reducido la dimensión del problema a costa de una pequeña pérdida de información.

Existen otras dos razones importantes para mantener la dimensionalidad del espacio de características tan pequeño como sea posible: costo de medida y precisión en la clasificación. Un número limitado de características simplifica la representación tanto del patrón como del clasificador, lo que resulta en un clasificador más rápido y que usa menos memoria. La técnica de análisis de componentes principales que representa de mejor manera los datos en un sentido de mínimos cuadrados, esta técnica tiene tres efectos:

- Ortogonaliza los componentes de los vectores de entrada (de modo que estén no correlacionadas).
- Ordena los componentes ortogonales resultantes (componentes principales) así que aquellas con la menor variación van primero.
- Elimina aquellas componentes que contribuyen menos en la variación de los datos¹⁴.

¹³ ERZO GARCIA Héctor. Determinación de la mejor Ondilla para análisis de fonemas. Universidad Autónoma Metropolitana Iztapalapo, p.17

¹⁴ OROZCO ALZATE Mauricio. Clasificación de arritmias Cardíacas usando Transformada Wavelet y Técnicas de reconocimiento de Patrones. Universidad Nacional de Colombia en Manizales, p.57

3. MAQUINA DE SOPORTE VECTORIAL (SVM)

Clasificar consiste en separar los elementos de un conjunto C en diferentes subconjuntos C_i , $i = 1, \dots, P$, denominados clases, con base en la medición de las características que los elementos de C poseen. Por ejemplo cuando se tiene elementos de un conjunto C que son el resultado de un conjunto finito de variables aleatorias en R^n , denotado por:

$$C = \{X_1, X_2, \dots, X_k\},$$

Donde X_j es una variable discreta, infinita o continua. Entonces C es formado con subconjuntos de valores representativos de cada una de las variables aleatorias X_j . La Máquina de soporte vectorial (SVM) es un sistema de aprendizaje basado en el uso de un espacio de hipótesis de funciones lineales en un espacio de mayor dimensión inducido por un Kernel, donde se busca una máxima separación entre clases, donde se modela en cierta forma el comportamiento de cada una de las variables aleatorias X_j , de forma que se puede determinar, dado un vector propuesto, a cuál de ellas pertenece.

Para la clasificación de de las zonas, se considera que el conjunto C esta formado por dos variables aleatorias:

$$C = \{X_1, X_2\},$$

Se representa a cada elemento del conjunto C de la siguiente forma:

$$(x_j, y_j), j = 1, \dots, l,$$

Donde $x_j \in R^n, y_j \in \{-1, 1\}$ y l es cardinalidad de C , si se toma una muestra una muestra representativa de cada una de las variables aleatorias, la representacion dada previamente permite establecer al conjunto C de la siguiente manera:

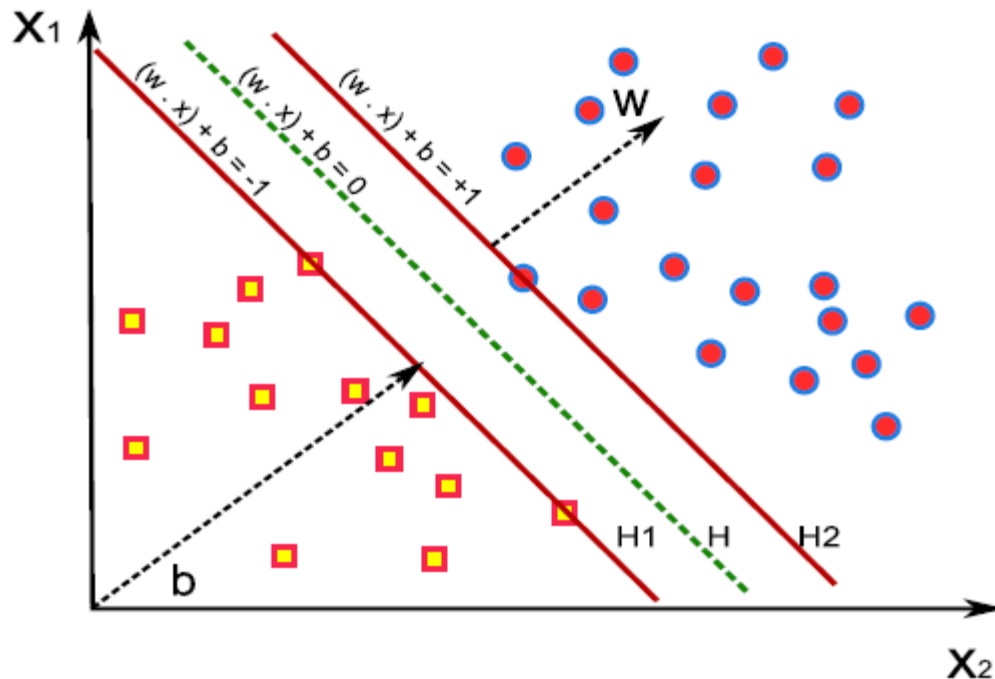
$$\begin{aligned} C &= C_1 \cup C_2, & C &= C_1 \cap C_2 = \emptyset \\ C_1 &= \{x_1, \dots, x_k\}, \\ C_2 &= \{x_{k+1}, \dots, x_l\}, \end{aligned}$$

La distribución de los puntos que conforman a C_1 y C_2 , es desconocida. Se pueden considerar dos casos; cuando C_1 y C_2 , son linealmente separables y cuando no lo son.

3.1 CLASIFICACION POR VECTORES DE SOPORTE

Se desea construir un hiperplano para resolver el problema de separar el conjunto de vectores de entrenamiento. Primero la (SVM) mapea los puntos de entrada a un espacio de características de dimensión mayor (por ejemplo si los puntos de entrada están en R^2 serán mapeados a R^3 por la (SVM), para encontrar el hiperplano que los separe y maximice el margen m entre las clases, ver Figura 12.

Figura 12. Maximización del margen m



Fuente: ANGULO J. Sanz. Máquinas de Aprendizaje para Clasificar señales Electroencefalográficas

Al maximizar el margen m , la separación de las clase es un problema de programación cuadrática y puede ser resuelto por su problema dual introduciendo multiplicadores de Lagrange. La SVM puede encontrar el hiperplano óptimo utilizando el producto punto con funciones en el espacio de características que son los llamados kernels. Los vectores de soporte son combinaciones de unos pocos puntos de entrada que permite escribir la solución del hiperplano de manera más sencilla.

- Considerando el caso de un caso de N puntos de datos de entrenamiento $\{(x_1, y_1), \dots, (x_i, y_i)\}$
- Un hiperplano

$$H_0: y = wx - b = 0 \quad (\text{Ecuación 13})$$

Donde w es la normal al hiperplano, $b/\|w\|$ es perpendicular y $\|w\|$ es la norma euclídea de w .

- Dos hiperplanos paralelos a H_0 :

$$H_1: y = wx - b = 1 \quad (\text{Ecuación 14})$$

$$H_2: y = wx - b = -1 \quad (\text{Ecuación 15})$$

Con la condición que no hay puntos de datos entre H_1 y H_2 si la distancia d_+ y d_- es la distancia más corta desde la separación del hiperplano H_0 al punto más cercano positivo o negativo, donde el hiperplano H_1 y H_2 está ubicado, entonces la distancia entre H_1 y H_2 es $d_+ + d_-$. Así $d_+ = d_- = 1/\|w\|$, entonces el margen es igual a $2/\|w\|$. El problema es encontrar el hiperplano que entregue el máximo margen, los parámetros w y b son llamados vector peso y sesgo, respectivamente. La optimización del problema se presenta con la ecuación:

$$\text{Min}_{w,b} \frac{1}{2} w^t w \text{ condicionado por } \gamma_i (wb - 1) \geq 1 \quad (\text{Ecuación 16})$$

Este es un problema de optimización de tipo cuadrático en (w, b) y cuya solución puede darse a partir del uso de la teoría de multiplicadores de Lagrange en n variables. Una función de Lagrange para la optimización del problema puede ser definida como:

$$L_p(w, b, \alpha) = \frac{1}{2} w^t w - \sum_i (\alpha_i \gamma_i (wx - b) - \alpha_i) \quad (\text{Ecuación 17})$$

Usando la formulación dual y las limitaciones del problema de optimización de Lagrange, los parámetros α_i y w, b pueden ser calculados los cuales especifican la separación del hiperplano. Se pueden calcular utilizando las siguientes ecuaciones¹⁵:

$$w = \sum \alpha_i \gamma_i x_i \quad (\text{Ecuación 18})$$

$$\alpha_i (\gamma_i (wx + b) - 1) = 0, \forall_i \quad (\text{Ecuación 19})$$

De acuerdo a la Ecuación 18 los parámetros α_i que no son iguales a cero corresponden a los datos x_i, y_i que corresponden a los vectores de soporte a, b, c y d, ver Figura 12.

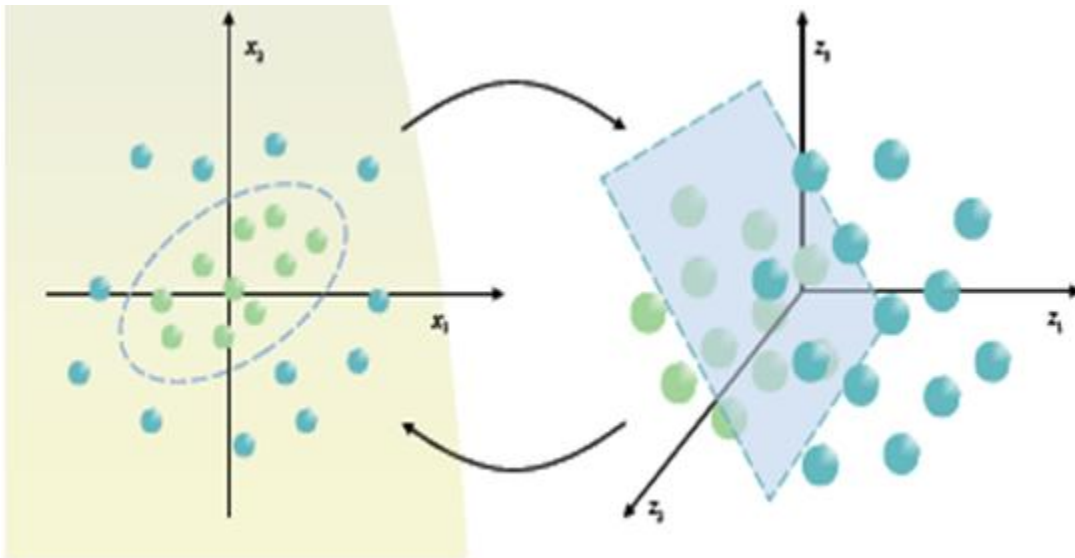
Si la superficie de separación de dos clases no es lineal, los datos pueden ser transformados en otro espacio característico de dimensión más alta donde el problema es linealmente separable.

¹⁵ ANGULO J. Sanz. Máquinas de Aprendizaje para Clasificar señales Electroencefalográficas, p.3

El tratamiento para el caso de dos conjuntos no separables linealmente consiste en utilizar una función $\phi: R^n \rightarrow R^m$ con $n, m \in N$, $m \geq n$ ó $m = \infty$. 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 se pueden denotar como $\Gamma_1 = \{\phi(x_1), \dots, \phi(x_k)\}$ y $\Gamma_2 = \{\phi(x_{k+1}), \dots, \phi(x_l)\}$ respectivamente.

El objetivo de realizar el mapeo es que los conjuntos obtenidos Γ_1 y Γ_2 sean separables linealmente o en su caso se minimice el error mediante la separación con un hiperplano, es decir que el numero de vectores clasificados incorrectamente sea mínimo. La siguiente figura muestra esta operación.

Figura 13. Transformación de dos conjuntos no separables linealmente



Fuente: GONZALEZ, Miguel. Aprendizaje estadístico, Redes Neuronales y Support Vector Machines: Un enfoque global

Si la transformación a un espacio dimensional es $\phi()$, entonces la función Lagrangiana puede ser expresada como:

$$L_D = \sum_i a_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \phi(x_i) \phi(x_j) \quad (\text{Ecuación 20})$$

El producto punto $\phi(x_i)\phi(x_j)$ en el que espacios de dimensión alta definen una función kernel

$$k(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \quad (\text{Ecuación 21})$$

Su uso es más importante que el de la función ϕ , de la cual no se requiere su conocimiento en forma explícita, ya que es suficiente, con establecer la función k para obtener una solución. Con un kernel adecuado, la MSV puede separar en el espacio característico los datos que en el espacio original de entrada no es separable. Algunos ejemplos de funciones kernel que han sido sugeridas o empleadas en problemas de clasificación son las siguientes¹⁶:

Lineal:

$$k(x_i, x_j) = x_i \cdot x_j \quad (\text{Ecuación 22})$$

Polinomial:

$$k(x_i, x_j) = (\gamma x_i \cdot x_j + r)^d, \gamma > 0, r, d \in R \quad (\text{Ecuación 23})$$

Función de base radial:

$$k(x_i, x_j) = c \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0, c \in R \quad (\text{Ecuación 24})$$

Sigmoide:

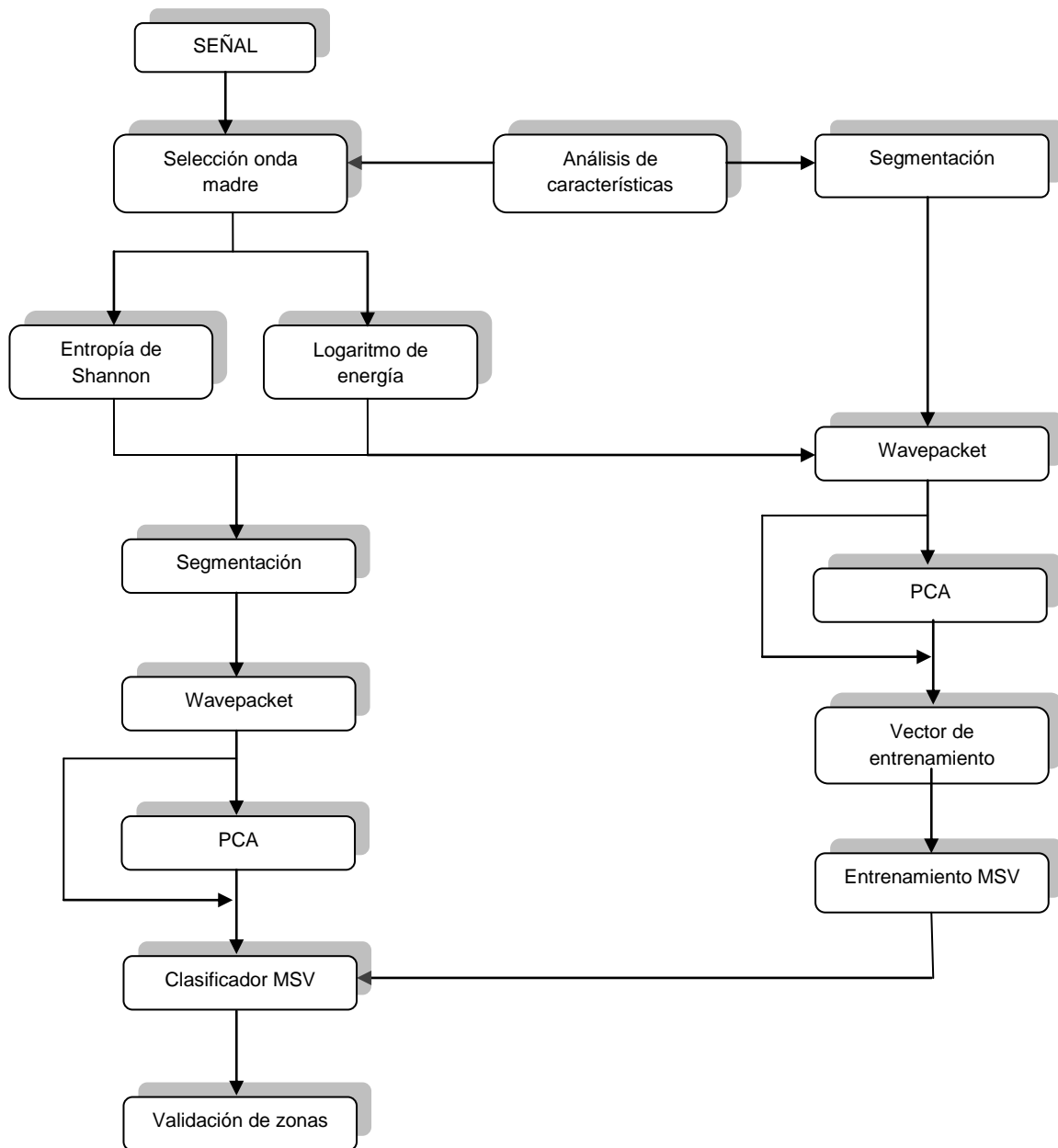
$$k(x_i, x_j) = \tanh(\gamma x_i \cdot x_j + r), \gamma, r \in R \quad (\text{Ecuación 25})$$

¹⁶ PEDROZA, Gabriel. Aplicación de las Máquinas de Soporte Vectorial al reconocimiento de hablantes. Universidad Autónoma Metropolitana, p.45

4. DESARROLLO METODOLÓGICO

En este capítulo se explica y fundamenta la metodología utilizada para el desarrollo del estudio de investigación planteado como identificación de zonas potencialmente productoras a través de registros de perfilaje. En la Figura 14 se presenta el diseño metodológico organizado en varias etapas que corresponden al desarrollo sistemático que se llevó a cabo.

Figura 14. Diagrama esquemático general del sistema implementado



La herramienta de computo para desarrollar los correspondientes algoritmos que fueron diseñados durante el desarrollo del proyecto fue Matlab 7.6 R2008a, por contar con la debida licencia en la Universidad Surcolombiana.

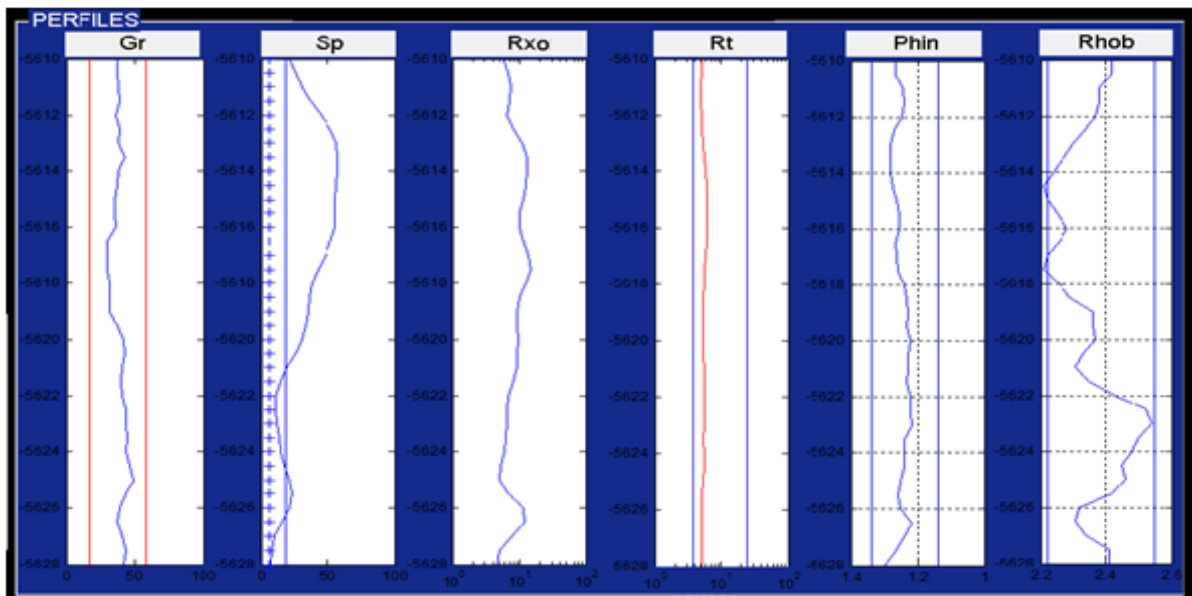
4.1 OBTENCIÓN DE DATOS

En el diagrama general de la Figura 14, se inicia con la señal correspondiente a los perfiles de los diferentes pozos suministrados por ECOPEPETROL a través del Instituto Colombiano del Petróleo ICP. Archivos organizados en formato .txt que contienen la información de los registros entregados, ubicación, datos de los perfiles y demás datos necesarios para realizar una buena identificación. Estos registros son cargados realizando una migración de los archivos de texto plano al entorno de Matlab para iniciar el procesamiento de la información contenida en ellos.

4.2 ANALISIS GRAFICO DE LAS CARACTERÍSTICAS FÍSICAS DE LOS PERFILES

En esta etapa se obtienen los parámetros necesarios para que de acuerdo a cada una de las características físicas de los perfiles utilizados (eléctricas, radioactivas y de porosidad), identificar aquellas zonas no productoras de hidrocarburos. Estos valores son necesarios debido a que las zonas consideradas como productoras fueron suministradas por ECOPEPETROL, mientras que las zonas consideradas como no productoras fueron obtenidas mediante el análisis gráfico de los perfiles como lo muestra la Figura 15.

Figura 15. Representación Gráfica de los perfiles



Las zonas no reservorias fueron obtenidas analizando cada uno de los perfiles que fueron registrados a lo largo de los pozos LL-72 y LL-77. El recorrido que se hizo a través del pozo fue realizado por segmentos de 4(ft), 5(ft), 6(ft), 7(ft), 8(ft), 10(ft), 12(ft). De esta manera se fijaron estos parámetros para que existiera una mejor identificación de las zonas no reservorias para el entrenamiento de la (SVM).

4.2.1 Estimación de la línea base de lutita para el SP. Para procesar la información correspondiente a las lecturas de potencial espontáneo, es necesario establecer algunos parámetros, uno de ellos y el más importante es la línea base de lutita o shale¹⁷. Cuando las lecturas de la curva de SP son más o menos constantes y tienden a seguir una línea recta se puede asumir esa línea recta como la línea base de shale. Se utilizó un algoritmo que calcula la línea base de shale y determinar las deflexiones hacia la derecha o hacia la izquierda de esta línea.

4.2.2 Estimación del perfil de Rayos Gamma GR. Según el Cuadro 1, de los valores típicos de GR, si este es bajo indica bajo contenido de arcilla y si es alto indica alto contenido de arcilla, para valores de GR establecidos entre 10% y 45% del mínimo y máximo de los datos respectivamente obtenidos serán considerados reservorios. Se excluyen valores muy bajos ya que estas litologías no son rocas almacenadoras de hidrocarburos; y para litologías como caolinita y esmectita los valores en unidades API son elevados representando zonas arcillosas impermeables¹⁸.

Cuadro 1. Valores típicos de Gamma Ray (GR)

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

¹⁷ Camargo Puerto Jorge Arturo, Introducción a la interpretación de perfiles de pozo abierto. Neiva: Universidad Surcolombiana, p 42

¹⁸ Bernal, Hugo. Rodríguez, Alfredo. (2009). Identificación de zonas productoras de hidrocarburos aplicando DSP Y métodos de correlación de registros de perfilaje. Universidad Surcolombiana. P 25

4.2.3 Estimación de la resistividad verdadera R_t . El parámetro resistivo de mayor interés es la resistividad verdadera R_t , esto se debe a que está relacionado con la saturación de hidrocarburo. La resistividad de la zona lavada, R_{xo} , es también de mucho interés, ya que con la comparación de R_t y R_{xo} se puede tener una idea de la movilidad de los hidrocarburos. R_t podría tomar valores de aproximadamente cuatro veces el valor R_o (R_o , resistividad de la zona virgen 100% saturada de agua de formación). Por su parte para R_o , siguiendo la misma metodología de análisis en los datos, se concluyó que su valor se encontraba entre 0 y 1, tomando 0.5 ohmios por metro como el mínimo valor de esta resistividad. $R_o < 4 R_t$.

4.2.4 Estimación de la resistividad verdadera R_{xo} . Las herramientas de resistividad miden los efectos producidos por una corriente inyectada en la formación, requieren un fluido conductivo dentro del pozo que pueda transportar la corriente inyectada. R_{xo} corresponde a la resistividad en la zona lavada, es decir la zona que fue invadida por el filtrado de lodo. R_{xo} , puede verse afectada por la invasión del lodo siendo alta cuando es afectada por el lodo; dependiendo del tipo de lodo, esta resistividad es mayor o menor que R_t .

4.2.5 Estimación del perfil neutrónico PHIN. El perfil neutrónico entrega el valor real del espacio poral lleno de fluidos. Los valores del perfil neutrónico para que fueran zonas de interés estaban comprendidos entre 20% y 34% de su capacidad para contener fluidos. Se ha demostrado que para que un reservorio arenoso tenga producción comercial, su porosidad debe ser mayor o igual a 9%. No obstante, existen areniscas con producción comercial, por debajo de este límite, debido a que están fracturadas, es decir, que poseen porosidad secundaria adicional a la porosidad primaria¹⁹. De otro lado, existen reservorios carbonatados con buena producción comercial con porosidad de tan solo 6%.

4.2.6 Estimación del perfil de densidad de formación RHOB²⁰. Este perfil tiene aplicación en la identificación de minerales en depósitos de evaporitas, descubrimiento de gas y determinación de la densidad de hidrocarburos, en donde se mide un índice de densidad electrónica, ρ_e , proporcional a la densidad total ρ_b , en gr/cm³. Partiendo de las consideraciones dadas por la literatura en la interpretación de perfiles, y por el análisis realizado a los datos se consideraron valores de Bulk Density (ρ_b) óptimos en el rango de 2.2 gr/m³ y 2.71 gr/m³. De acuerdo a lo anterior el Cuadro 2, muestra los valores ρ_e y ρ_b ideales de acuerdo al tipo de roca. Se validó lo dicho para minerales almacén como areniscas y calizas.

¹⁹ Camargo Puerto Jorge Arturo, Introducción a la interpretación de perfiles de pozo abierto. Neiva: Universidad Surcolombiana, 2007.

²⁰ Andrade Trujillo, Guillermo. Polanco, Juan Manuel. Evaluación de Algoritmos Basados en Lógica Difusa para el Procesamiento de Señales de Registro de Pozo abierto. Universidad Surcolombiana, 2009.

Cuadro 2. Valores de ρ_e y ρ_b

Mineral	ρ_e	ρ_b
Arenisca	1.81	2.65
Caliza	5.05	2.71
Dolomita	3.14	2.86
Anhidrita	5.08	2.97
Halita (NaCl)	4.65	2.03
Agua Dulce	0.35	1
Hidrocarburo	≤ 0.12	< 1
Agua Salada 200,000 ppm	1.2	1.146

4.3 SELECCIÓN DE LA MEJOR ONDA MADRE

Para la selección de la Wavelet madre, se utiliza la Transformada Wavelet Discreta gracias a su flexibilidad en la representación de señales no estacionarias. En principio, para transportar la misma información, pero empleando una menor cantidad de espacio, se hizo uso de dos funciones de costo de información (Entropía de Shannon y Logaritmo de Energía), en donde a cada registro se le calcula su función de costo y se escoge la función que tenga el menor valor de costo para obtener la onda madre que se adapte a cada uno de los perfiles. Se realiza la expansión en los bancos de filtros en todo el espectro frecuencial, y se calcula la entropía con cada uno de los dos métodos. Las siguientes fueron las funciones de costo de información utilizadas:

- Entropía de Shannon:

$$H(q) = \sum_k q(k) \log \frac{1}{q(k)} \quad (\text{Ecuación 26})$$

- Logaritmo de la Energía:

$$M(x) = \sum_j \log |x_j|^2 \text{ con } \log(0) = 0 \quad (\text{Ecuación 27})$$

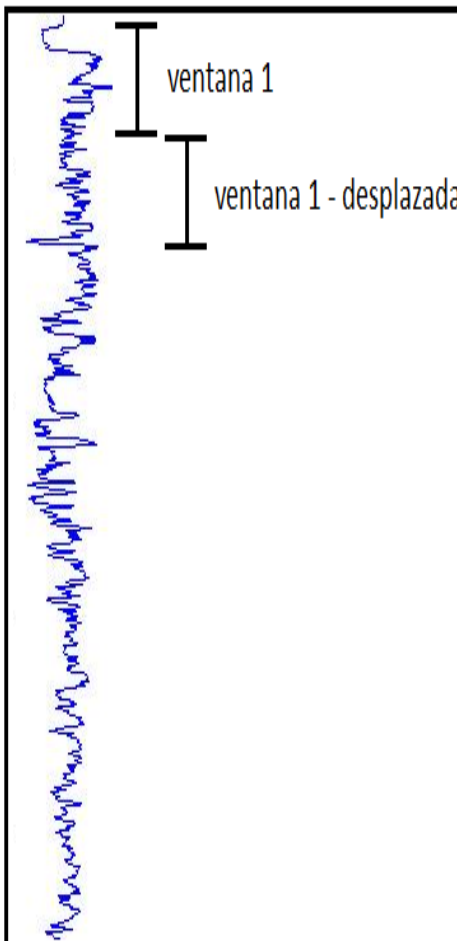
Para obtener la función de coste se utilizó la siguiente función.

$$C_{k,\lambda} = - \sum_{m=1}^N \frac{|\langle f, \psi_{m,\lambda} \rangle|^2}{\|B^\lambda\|^2} \log_e \frac{|\langle f, \psi_{m,\lambda} \rangle|^2}{\|B^\lambda\|^2} \quad (\text{Ecuación 28})$$

En el estudio para la búsqueda de la onda madre, se inicio con el análisis de todas las familias Wavelet que se pueden discretizar. Las familias que se acomodaron de una manera eficiente a cada uno de los perfiles a lo largo de los pozos, fueron las correspondientes a la db, rbio, y bior, coif. Se descartó la familia Symlet en un orden mayor a 15 por su gran costo computacional que afecta al rendimiento del sistema y no logran adaptarse de una manera eficiente a la señal.

Cada onda Wavelet recorrió cada uno de los perfiles obteniendo como resultado una onda Wavelet característica. Para obtener una mejor caracterización se desarrolló un algoritmo que evaluara los perfiles a través de diferentes ventanas a 4(ft) ,5(ft), 6(ft), 7(ft), 8(ft), 10(ft), 12(ft), como lo muestra la Figura 16, en donde se escogía la ventana dependiendo de la profundidad de la zona y se observó su respectiva representación.

Figura 16. Representación de la ventana de desplazamiento de cada onda Wavelet



4.4 NORMALIZACIÓN

Se estandarizaron los datos de los registros en la entrada a fin de que las entradas para la Máquina de Soporte Vectorial se encontraran en el rango entre cero (0) y uno (1); este proceso proporciona un rango de valores más apropiado para los procesos de reconocimiento de patrones, ya que algunos datos dentro de los registros de perfilaje, como el perfil de resistividad presentan datos atípicos y por ende su desviación estándar es demasiado alta. La siguiente ecuación fundamenta el proceso de normalización²¹:

$$X_n = \frac{X_0 - X_{\min}}{X_{\max} - X_{\min}} \quad (\text{Ecuación 29})$$

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

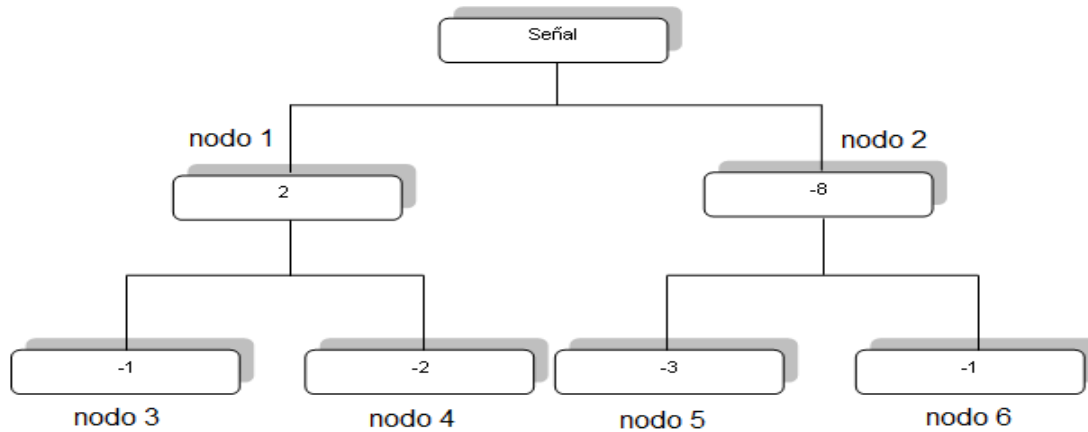
4.5 IMPLEMENTACIÓN DEL ALGORITMO WAVEPACKET (WP)

Después de haber obtenido la función con menor coste de información, para encontrar la mejor base Wavelet para esta señal se realizó a través del Wavepacket de Matlab. Para conseguir la representación se seleccionó un conjunto de nodos que se ajustó a las características de la señal; en este punto se calcula la entropía para cada una de las ramas del árbol y se hace la escogencia de cual es menor y contiene más información representativa, luego donde se encuentre la menor entropía se obtiene el vector característico de cada zona.

El proceso de selección de componentes representativos de la señal se realiza a través del concepto de entropía para cada uno de los nodos donde culmina el árbol de descomposición. El algoritmo inicia un análisis completo desde la parte inferior hasta llegar al primer nivel de descomposición ver Figura 17, analiza rama por rama de abajo hacia arriba, por todos los niveles del árbol. Estando en el último nivel analiza los dos nodos que salen del nodo anterior, y realiza una suma de sus coeficientes obteniendo un resultado que representa el costo de información en ese punto.

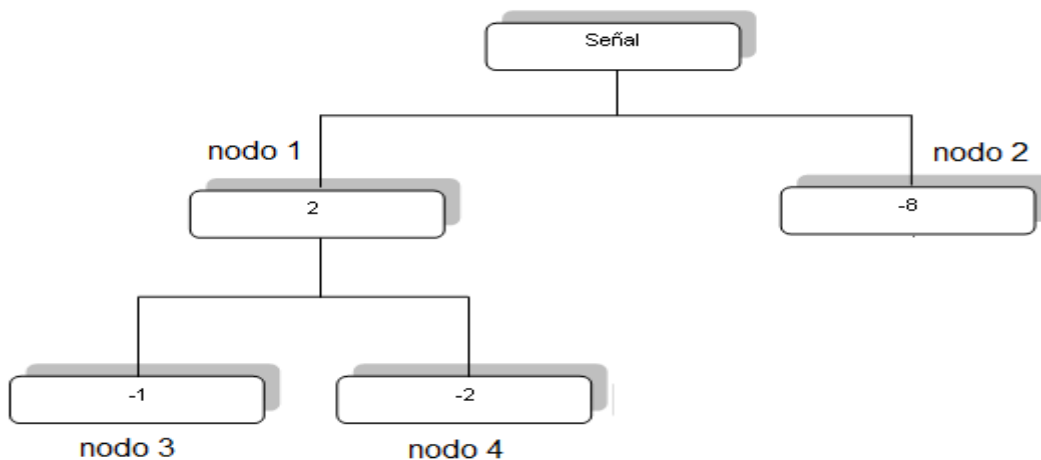
²¹ Bernal, Hugo. Rodríguez, Alfredo. (2009). Identificación de zonas productoras de hidrocarburos aplicando DSP Y métodos de correlación de registros de perfilaje. Universidad Surcolombiana. P 31

Figura 17. Esquema árbol de descomposición



Para este ejemplo el nodo 3 con el nodo 4 del nivel 2, aplica una suma entre -1 con -2 resultando -3, que indica la entropía localizada en esa parte, con este resultado el algoritmo compara con el nodo del nivel anterior, para este caso con el nodo 1, este nodo posee un costo de información mayor ya que tiene un valor de entropía de 2, mayor que el costo de información de los nodos hijos, con esto el algoritmo de la mejor base selecciona el conjunto de nodos 3 y 4, ya que posee menor costo de información que su padre, luego sigue examinando los nodos siguientes, que son el nodo 5 y el nodo 6, se obtiene una entropía de -4, comparada con el nodo 2 que tiene una distribución de información de -8, la cual el algoritmo seleccionara este nodo y omitirá sus nodos hijos (nodo 5 y nodo 6) no se tendrán en cuenta. Así queda determinado el conjunto de nodos de menor coste para la mejor representación de la señal. En la Figura 18 se ilustra el árbol seleccionado por el algoritmo.

Figura 18. Árbol con la menor distribución de información



4.6 REDUCCIÓN DE DIMENSIONALIDAD

Del análisis de la señal mediante la Transformada Wavelet y las funciones de costo de información, se obtuvo un vector característico de cada perfil, en donde sus coeficientes fueron organizados en dos vectores diferentes para ver con cuál de estos vectores conseguir una mejor extracción de características y así obtener el mejor vector de entrenamiento.

- Se agrupó en un vector fila donde se organizaron los coeficientes del vector característico de acuerdo a su espectro frecuencial. Ver Figura 19.

Figura 19. Ordenamiento vector característico como vector fila

$$\text{Vector característico como vector fila} = [f1, f2, f3 \dots fn]$$

- Se agrupó en un vector columna donde se organizaron los coeficientes del vector característico de acuerdo a su espectro frecuencial. Ver Figura 20.

Figura 20. Ordenamiento vector característico como vector columna

$$\text{Vector característico como vector columna} = \begin{bmatrix} f1 \\ f2 \\ f3 \\ \cdot \\ \cdot \\ \cdot \\ fn \end{bmatrix}$$

El Análisis de Componentes Principales (PCA) recibe el vector característico de cada perfil organizado como vector columna o como vector fila y realiza una reducción del número de variables o dimensionalidad, para eliminar aquellos componentes principales que contribuyen menos de un porcentaje especificado, perdiendo la menor cantidad de información posible. El algoritmo hace la escogencia de la medida adecuada de características que representen de buena manera y simple la señal, una medida exagerada en la selección de características dificulta el proceso analítico en toda la dimensión del espacio, reduciendo la eficiencia en el proceso; el PCA analiza la varianza entre los datos y selecciona aquellos donde su varianza es considerablemente alta y omite aquellas características donde esta varianza no influye. Después de normalizados los datos del vector característico para que tenga media cero y varianza unitaria, se aplicó un porcentaje de variación utilizado para el mapeo del PCA de 7%. Los nuevos componentes principales serán una combinación lineal de las variables originales y además independientes entre sí. El resultado es nuevamente un vector que es la entrada de la Máquina de Soporte Vectorial para la etapa del entrenamiento.

El PCA realiza un cambio de plano, donde agrupa los datos de la matriz que contiene la información de todos los perfiles, como se muestra en la Figura 21, y multiplica a estos datos por una constante lineal, para establecer una mejor agrupación de datos correlacionados, entonces el PCA analiza la agrupación de estos datos y escoge los datos o componentes que presenten un mayor comportamiento o distribución lineal en el plano, omitiendo las componentes que no se encuentran dentro de esta distribución, haciendo una reducción de datos y escogiendo las componentes correlacionadas. Las componentes que más se acerquen al margen o recta serán tenidas en cuenta por el PCA, en la Figura 22 se ilustran las componentes finales que serán las elegidas para la representación más compacta y reducida de los perfiles.

Figura 21. Agrupación de las componentes de los perfiles para la reducción

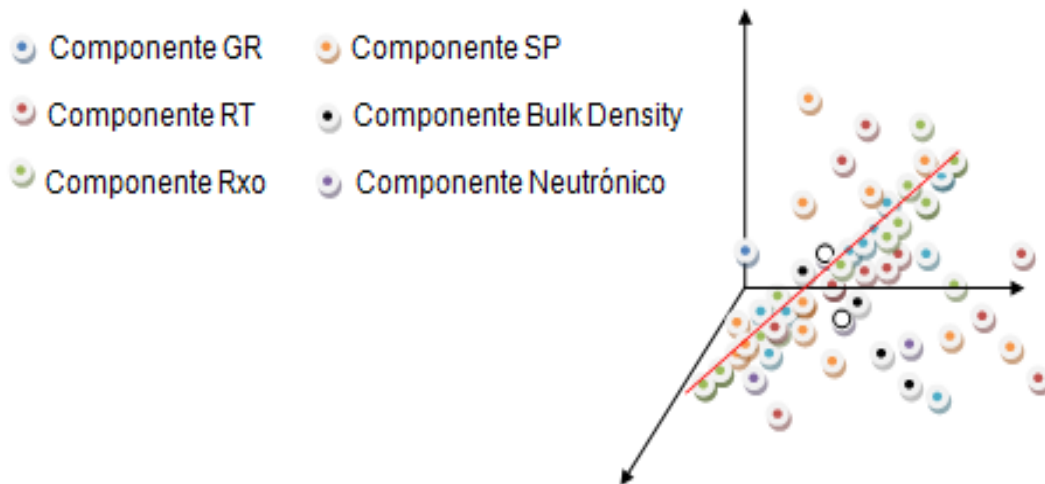
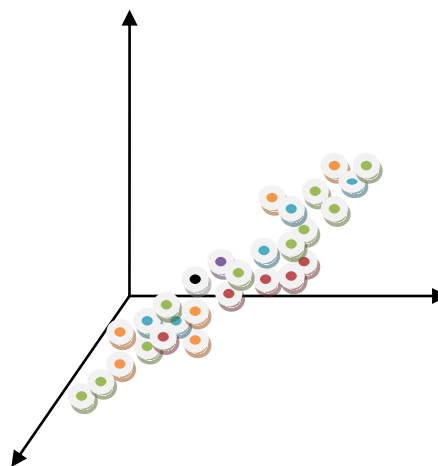


Figura 22. Selección de las componentes de los perfiles por el PCA



Se caracterizó cada uno de los perfiles y estos coeficientes se organizaron como vector fila y como vector columna donde luego se hizo la reducción de dimensionalidad en busca de mejorar la selección de características.

El resultado es nuevamente un vector de entrenamiento que es la entrada a la Máquina de Soporte Vectorial. Estos vectores llegan a una sola a matriz donde están todos los datos correspondientes a todos los perfiles de los registros, ver Figura 23.

Figura 23. Matriz con los vectores de entrenamiento

$$VCC = \begin{bmatrix} VGR \\ VSP \\ VRxo \\ VRt \\ PHIN \\ RHOB \end{bmatrix}$$

4.7 SELECCIÓN DE LA POBLACION DE ENTRENAMIENTO

Los datos escogidos para el entrenamiento son las zonas productoras que se encuentran en los datos entregados por ECOPEPETROL, y las zonas no productoras fueron obtenidas del análisis grafico de los perfiles a lo largo de los pozos LL-72 y LL-77, seleccionando zonas a lo largo de todo el perfil, como se muestra en la Figura 15, se agruparon según sus profundidades de zona, las zonas de cada pozo se clasifican según la profundidad que presentan, iniciando por las de cuatro pies hasta los valores máximos que se encuentren en las zonas productoras, donde se encontraron zonas de 4(ft), 5(ft), 6(ft), 7(ft), 8(ft), 9(ft), 10(ft), 12(ft), el tope es el inicio de la zona y la base es donde termina. En el Cuadro 3 se observa la agrupación de la población de entrenamiento que corresponde al pozo LL-72.

Para cada zona productora se adjudica una zona no productora de la misma longitud, esto para que en la etapa de entrenamiento la maquina establezca características y pueda hacer una diferenciación de clases (productoras y no productoras), La ventana mínima o zona es de un valor de cuatro pies porque su longitud esta en los limites posibles para poder hallar una reserva, en el Cuadro 4 se organiza la población de entrenamiento que corresponde al pozo LL-77. En la Figura 15, puede observarse que la zona que va desde 5610 pies hasta 5620 presenta una zona productora de 10 pies, ya que para todos los perfiles las curvas se encuentran dentro de los márgenes para zonas productoras. Después de 5620 en el perfil SP, la curva esta a la izquierda de la línea de shale, entonces para esta zona de 5620 hasta 5628 pies, se considera una zona no productora. De esta manera se evaluaron cada uno de los registros y se extrajeron zonas que son no productoras.

Cuadro 3. Población de entrenamiento para los pozos LL-72 correspondiente a las zonas reservorias y no reservorias

Pozo	Longitud (Pies)	zonas productoras		zonas no productoras	
		Tope	Base	Tope	Base
LL-72	4	6232	6236	3298	3302
	5	6400	6405	3303	3308
		6464	6469	3310	3315
	6	5892	5898	3338	3344
		5942	5948	3345	3351
		6182	6188	3352	3358
		6446	6452	3359	3365
	8	6280	6288	3366	3374
		6698	6706	3375	3383
	9	5656	5665	3384	3393
		6680	6689	3394	3403
	10	5610	5620	3404	3414
12	5534	5546	3438	3450	

Cuadro 4. Población entrenamiento pozo LL-77 zonas reservorias y no reservorias

Pozo	Longitud (Pies)	zonas productoras		zonas no productoras	
		Tope	Base	Tope	Base
LL-77	4	5636	5640	3319	3323
	5	5624	5629	3324	3329
		6461	6466	3330	3335
	6	5892	5898	3336	3342
		5942	5948	3343	3349
		6182	6188	3350	3356
		6446	6452	3357	3363
	7	5654	5661	3364	3371
		5855	5862	3372	3379
		6123	6130	3380	3387

Continuación Cuadro 4

Pozo	Longitud (Pies)	zonas productoras		zonas no productoras	
77	7	6142	6149	3388	3395
	8	5770	5708	3396	3404
		5822	5830	3406	3414
		6002	6010	3416	3424
		6224	6232	3426	3434
		6620	6628	3436	3444
		5750	5760	3445	3455
	10	5750	5760	3445	3455

4.8 IMPLEMENTACION MAQUINA DE SOPORTE VECTORIAL (SVM)

El entrenamiento del sistema de clasificación se realiza con la pareja (X_i, Y_j) , el sistema de clasificación implementado requiere de un patrón de entrenamiento el cual es el obtenido mediante la técnica de extracción de características PCA. El conjunto de vectores utilizados para conformar los de entrenamiento para la clasificación fueron las zonas productoras y las no productoras registradas en los Cuadros 6 y 7. En donde una vez obtenidos estos vectores se realizaron pruebas de validación para obtener una clasificación de:

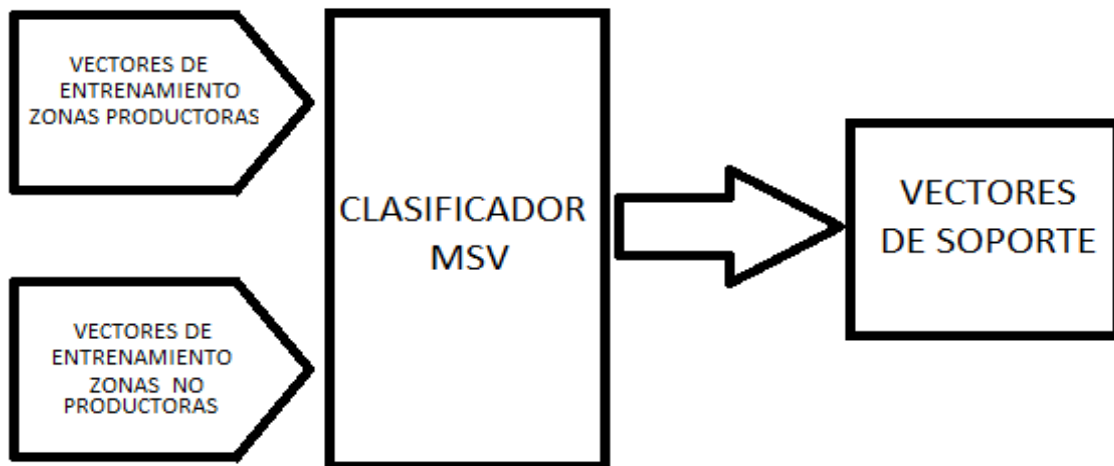
$$\text{Zonas productoras} \neq \text{Zonas no productoras}$$

La función utilizada para la clasificación es la de base radial (Ecuación 24). Las funciones que permiten especificar el método utilizado para encontrar el hiperplano de separación fueron analizadas individualmente, las opciones o métodos que fueron utilizados son: el uso de programación cuadrática QP, SMO método de optimización secuencial mínimo y LS método de mínimos cuadrados que tiene Matlab en su librería, obteniendo un hiperplano optimo con el método QP, ya que este método entrega un margen de separación mayor entre las clases lo que permite una mayor diferenciación y así le facilita a la SVM clasificarlas.

El algoritmo de clasificación implementado para la SVM determina los vectores de soporte que separan a ambas clases y de esta manera obtener un menor error en la clasificación. Para cada SVM implementada se determinó un porcentaje de eficiencia para el conjunto de entrenamiento y la clasificación. Por último se obtuvo la validación de los modelos creados mediante la realización de pruebas con los registros de los pozos LL-72 y LL-77. En la Figura 24, se muestra el esquema para que la Máquina de Soporte vectorial tomando como datos de entrenamiento las

zonas productoras y las no productoras, entrega los vectores de soporte que son los que dan el margen de separación entre las clases.

Figura 24. Generación de vectores de soporte



4.8.1 Entrenamiento SVM por parejas de perfiles y segmentos ft. Luego de obtener los vectores de entrenamiento para zonas productoras y no productoras y debido a las limitaciones de la herramienta Toolbox SVM de Matlab donde solo se puede trabajar con dos planos y ella posteriormente hace la elevación a un plano de una dimensión mayor. Como solución a este problema el entrenamiento se realiza con la comparación entre los perfiles que se encuentran en la matriz de la Figura 23, generando así una máquina por cada par, estas parejas se escogieron por similitud en relación a su aplicabilidad es decir Saturación de hidrocarburos, resistividad de la formación, porosidad de la roca etc. El entrenamiento se realizó con los vectores entre el perfil Gamma Ray GR y el perfil de Potencial Espontaneo SP donde se realizó una máquina GR-SP, el perfil Resistividad Verdadera Rt con el perfil Rxo, por similitud en sus características de resistividad creando una maquina Rxo-Rt, el perfil Neutrónico con el perfil Bulk Density por sus características de porosidad y permeabilidad, generando la maquina correspondiente.

Cada máquina contiene la información de los perfiles en un vector de entrenamiento. En totalidad se obtuvieron 21 máquinas para zonas de 4(ft), 5(ft), 6(ft), 7(ft), 8(ft), 9(ft), 10(ft), 11(ft), 12 (ft) que se hallaron en los registros de ECOPETROL, para este método se obtuvieron resultados desfavorables.

4.8.2 Entrenamiento de todas las zonas con todos los segmentos ft. Luego de haber estudiado el comportamiento del entrenamiento individualizando las zonas por segmentos se optó por hacer un entrenamiento general de los pozos para

corregir los posibles errores. Se diseñó y se implementó la máquina de soporte vectorial que entrena todas las zonas y cada uno de los perfiles entregando un solo paquete de información de soporte y no varias máquinas por separado que entregaran información y al enlazar se presentara discordancia entre ellas.

5. RESULTADOS

En este capítulo se presentan los resultados que se obtuvieron en cada uno de los códigos realizados en este trabajo, tomando como base los registros de perfilaje de pozo abierto suministrados por el ICP (Instituto Colombiano del Petróleo) – ECOPETROL.

5.1 OBTENCIÓN ONDA MADRE PARA TODOS LOS POZOS

En los siguientes Cuadros se registran los datos obtenidos en el algoritmo para seleccionar la mejor onda base Wavelet. Cada onda recorrió el pozo LL-72 con cada uno de los perfiles obteniendo como resultado una onda Wavelet característica, se tuvo en cuenta para la selección de la mejor onda madre aquella que represente la menor cantidad de concentración de datos de acuerdo a las funciones de costo de información.

Los resultados obtenidos mediante el perfil de Potencial Espontáneo (SP), en el pozo LL-72 fueron organizados en los Cuadros 5 y 6. El proceso fue realizado tanto para Logaritmo de Energía como para la Entropía de Shannon.

Cuadro 5. Selección onda madre para el perfil potencial espontáneo con el método de Logaritmo de Energía

Pozo	Madre	Costo de información	Técnica utilizada para la medida de información
72	db28	-7,46E+08	Logaritmo de Energía
77	db28	-6,37E+08	Logaritmo de Energía
81	rbio 2.6	-3,57E+08	Logaritmo de Energía
91	db28	-5,74E+08	Logaritmo de Energía
96	db28	-1,01E+08	Logaritmo de Energía
97	db28	-8,83E+08	Logaritmo de Energía
102	db28	-9,16E+08	Logaritmo de Energía
103	db28	-6,31E+08	Logaritmo de Energía
104	db28	-8,83E+08	Logaritmo de Energía
105	db28	-5,70E+08	Logaritmo de Energía
106	db28	-4,70E+08	Logaritmo de Energía
107	db28	-5,88E+08	Logaritmo de Energía

Cuadro 6. Selección onda madre para el perfil Potencial Espontáneo con el método de Entropía de Shannon

Pozo	Madre	Costo de información	Técnica utilizada para la medida de información
72	db28	-3,62E+08	Entropía de Shannon
77	db28	-4,84E+08	Entropía de Shannon
81	bior 2.4	-1,93E+08	Entropía de Shannon
91	bior 2.4	-2,37E+08	Entropía de Shannon
96	db24	-1,04E+08	Entropía de Shannon
97	db28	-8,72E+08	Entropía de Shannon
102	db28	-1,52E+08	Entropía de Shannon
103	db28	-5,54E+08	Entropía de Shannon
104	db28	-9,30E+08	Entropía de Shannon
105	db28	-1,23E+08	Entropía de Shannon
106	db28	-4,25E+08	Entropía de Shannon
107	db28	-5,71E+08	Entropía de Shannon

Para el SP en el pozo LL-72, según la información almacenada en los Cuadros 5 y 6 la onda madre que mejor se adapta a este perfil es la db28.

Los resultados obtenidos mediante el perfil de Gamma Ray, en el pozo LL-72 fueron organizados en los Cuadros 7 y 8. El proceso fue realizado tanto para Logaritmo de Energía como para la Entropía de Shannon.

Cuadro 7. Selección onda madre para el perfil Gamma Ray con el método de Logaritmo de Energía

Pozo	Madre	Costo de información	Técnica utilizada para la medida de información
72	rbio 2.6	-3,05E+08	Logaritmo de Energía
77	db28	-4,50E+08	Logaritmo de Energía
81	bior 2.4	-3,13E+08	Logaritmo de Energía
91	bior 2.4	-2,43E+08	Logaritmo de Energía
96	rbio 2.6	-1,98E+08	Logaritmo de Energía
97	rbio 2.6	-4,19E+08	Logaritmo de Energía
102	rbio 2.6	-4,38E+08	Logaritmo de Energía
103	rbio 2.6	-3,05E+08	Logaritmo de Energía

Continuación Cuadro 7

Pozo	Madre	Costo de información	Técnica utilizada para la medida de información
104	db28	-4,88E+08	Logaritmo de Energía
105	rbio 2.6	-2,49E+08	Logaritmo de Energía
106	rbio 2.6	-3,37E+08	Logaritmo de Energía
107	rbio 2.6	-2,80E+08	Logaritmo de Energía

Cuadro 8. Selección onda madre para el perfil Gamma Ray con el método de Entropía de Shannon

Pozo	Madre	Costo de información	Técnica utilizada para la medida de información
72	db28	-3,62E+13	Entropía de Shannon
77	db28	1,64E+14	Entropía de Shannon
81	sym15	-2,18E+12	Entropía de Shannon
91	bior 2.4	1.0e+012	Entropía de Shannon
96	db28	-3,70E+10	Entropía de Shannon
97	sym15	-2,50E+12	Entropía de Shannon
102	bior 2.4	-5,99E+10	Entropía de Shannon
103	db28	-4,33E+10	Entropía de Shannon
104	db28	-3,74E+13	Entropía de Shannon
105	db28	-1,23E+14	Entropía de Shannon
106	db28	-9,88E+13	Entropía de Shannon
107	db28	-4,07E+13	Entropía de Shannon

En el perfil Gamma Ray según la información almacenadas en los Cuadros 7 y 8, la onda madre que mejor se adapta es la db28 para el método de Entropía de Shannon y la rbio 2.6 por el método del Logaritmo de Energía.

Los resultados obtenidos mediante el perfil de RT, en el pozo Llanito 72 fueron organizados en los Cuadros 9 y 10. El proceso fue realizado tanto para Logaritmo de Energía como para la Entropía de Shannon.

Cuadro 9. Selección Onda madre para el perfil RT con el método de Logaritmo de Energía

Pozo	Madre	Costo de información	Técnica utilizada para la medida de información
72	db28	-2.553.900.000	Logaritmo de Energía
77	db28	-1.930.900.000	Logaritmo de Energía
81	db28	-1.904.800.000	Logaritmo de Energía
91	bior 2.4	-709.650.000	Logaritmo de Energía
96	db28	-1.964.800.000	Logaritmo de Energía
97	db28	-1.859.200.000	Logaritmo de Energía
102	db28	-1.883.600.000	Logaritmo de Energía
103	db28	-1.817.400.000	Logaritmo de Energía
104	db28	-2.401.300.000	Logaritmo de Energía
105	db28	1.254.100.000	Logaritmo de Energía
106	db28	-1.213.400.000	Logaritmo de Energía
107	db28	-3.351.000.000	Logaritmo de Energía

Para el método de obtención de la Entropía de Shannon para el perfil RT se obtuvo el siguiente Cuadro.

Cuadro 10. Selección onda madre para el perfil RT con el método de Entropía de Shannon

Pozo	Madre	Costo de información	Técnica utilizada para la medida de información
72	db28	-1,02E+10	Entropía de Shannon
77	sym15	-1,64E+12	Entropía de Shannon
81	sym15	-2,18E+12	Entropía de Shannon
91	bior 2.4	-2,27E+16	Entropía de Shannon
96	db28	-3,70E+10	Entropía de Shannon
97	sym15	-2,50E+12	Entropía de Shannon
102	bior 2.4	-5,99E+10	Entropía de Shannon
103	db28	-4,72E+10	Entropía de Shannon
104	db28	-2,95E+10	Entropía de Shannon
105	db28	-5,87E+10	Entropía de Shannon
106	db28	-4,44E+10	Entropía de Shannon

Continuación Cuadro 10.

Pozo	Madre	Costo de información	Técnica utilizada para la medida de información
107	db28	-8,60E+08	Entropía de Shannon

En el perfil RT según la información almacenadas en los Cuadros 9 y 10, la onda madre que mejor se adapta es la db28 para el método de Entropía de Shannon y Logaritmo de Energía.

Los resultados obtenidos mediante el perfil Neutrónico, en el pozo Llanito 72 fueron organizados en los Cuadros 11 y 12. El proceso fue realizado tanto para Logaritmo de Energía como para la Entropía de Shannon.

Cuadro 11. Selección onda madre para el perfil Neutrónico con el método de Logaritmo de Energía

Pozo	Madre	Costo de información	Técnica utilizada para la medida de información
72	db28	-3,29E+09	Logaritmo de Energía
77	db28	-1,64E+09	Logaritmo de Energía
81	db28	-1,90E+09	Logaritmo de Energía
91	db28	-3,59E+08	Logaritmo de Energía
96	db28	-3,85E+09	Logaritmo de Energía
97	db28	-3,58E+09	Logaritmo de Energía
102	db28	-3,76E+09	Logaritmo de Energía
103	db28	-3,54E+09	Logaritmo de Energía
104	db28	-3,95E+09	Logaritmo de Energía
105	db28	-3,07E+09	Logaritmo de Energía
106	db28	-3,16E+09	Logaritmo de Energía
107	db28	-3,49E+09	Logaritmo de Energía

Cuadro 12. Selección Onda madre para el perfil Neutrónico con el método de entropía de Shannon

Pozo	Madre	Costo de información	Técnica utilizada para la medida de información
72	db28	-8,55E+09	Entropía de Shannon
77	db28	-8,03E+09	Entropía de Shannon

Continuación Cuadro 12.

Pozo	Madre	Costo de información	Técnica utilizada para la medida de información
81	db28	-7,77E+09	Entropía de Shannon
91	db28	-1,11E+10	Entropía de Shannon
96	db28	-8,30E+09	Entropía de Shannon
97	db28	-7,34E+09	Entropía de Shannon
102	db28	-7,75E+09	Entropía de Shannon
103	db28	-7,82E+09	Entropía de Shannon
104	db28	-9,72E+09	Entropía de Shannon
105	db28	-8,74E+09	Entropía de Shannon
106	db28	-6,64E+09	Entropía de Shannon
107	db28	-6,64E+10	Entropía de Shannon

En el perfil Neutrónico según la información almacenadas en los Cuadros 11 y 12, la onda madre que mejor se adapta es la db28 para el método de Entropía de Shannon y Logaritmo de Energía.

Los resultados obtenidos mediante el perfil Bulk Density, en el pozo LLanito 72 fueron organizados en los Cuadros 13 y 14. El proceso fue realizado tanto para Logaritmo de Energía como para la Entropía de Shannon.

Cuadro 13. Selección onda madre para el perfil Bulk Density con el método de Logaritmo de Energía

Pozo	Madre	Costo de información	Técnica utilizada para la medida de información
72	db28	-2,65E+09	Logaritmo de Energía
77	db28	-2,53E+09	Logaritmo de Energía
81	db28	-2,63E+09	Logaritmo de Energía
91	db28	-2,87E+09	Logaritmo de Energía
96	db28	-2,84E+09	Logaritmo de Energía
97	db28	-2,88E+09	Logaritmo de Energía
102	db28	-3,61E+09	Logaritmo de Energía
103	db28	-2,72E+09	Logaritmo de Energía
104	db28	-3,61E+09	Logaritmo de Energía
105	db28	-2,27E+09	Logaritmo de Energía

Continuación Cuadro 13.

Pozo	Madre	Costo de información	Técnica utilizada para la medida de información
106	db28	-2,18E+09	Logaritmo de Energía
107	db28	-2,90E+09	Logaritmo de Energía

Cuadro 14. Selección Onda madre para el Bulk Density, con el método de Entropía de Shannon

Pozo	Madre	Costo de información	Técnica utilizada para la medida de información
72	db28	-3,22E+10	Entropía de Shannon
77	db28	-3,20E+10	Entropía de Shannon
81	db28	-3,21E+10	Entropía de Shannon
91	db28	-3,96E+10	Entropía de Shannon
96	db28	-2,84E+10	Entropía de Shannon
97	db28	-2,90E+10	Entropía de Shannon
102	db28	-2,80E+10	Entropía de Shannon
103	db28	-3,02E+10	Entropía de Shannon
104	db28	-2,85E+10	Entropía de Shannon
105	db28	-2,72E+10	Entropía de Shannon
106	db28	-3,03E+10	Entropía de Shannon
107	db28	-5,07E+10	Entropía de Shannon

En el perfil Bulk Density según la información almacenadas en los Cuadros 13 y 14, la onda madre que mejor se adapta es la db28 para el método de Entropía de Shannon y Logaritmo de Energía.

Los resultados obtenidos mediante el perfil Rxo, en el pozo LLanito 72 fueron organizados en los Cuadros 15 y 16. El proceso fue realizado tanto para Logaritmo de Energía como para la Entropía de Shannon.

Cuadro 15. Selección onda madre para el perfil Rxo con el método de Logaritmo de Energía

Pozo	Madre	Costo de Información	Técnica utilizada para la medida de información
72	db28	-6,22E+10	Logaritmo de Energía

Continuación Cuadro 15.

Pozo	Madre	Costo de Información	Técnica utilizada para la medida de información
77	db28	-3,62E+13	Logaritmo de Energía
81	db28	1,64E+14	Logaritmo de Energía
91	db28	-2,18E+12	Logaritmo de Energía
96	db28	1.0e+012	Logaritmo de Energía
97	db28	-3,70E+10	Logaritmo de Energía
102	db28	-2,50E+12	Logaritmo de Energía
103	db28	-5,99E+10	Logaritmo de Energía
104	db28	-4,33E+10	Logaritmo de Energía
105	db28	-3,74E+13	Logaritmo de Energía
106	db28	-1,23E+14	Logaritmo de Energía
107	db28	-9,88E+13	Logaritmo de Energía

Cuadro 16. Selección onda madre para el perfil Rxo con el método de Entropía de Shannon

Pozo	Madre	Costo de información	Técnica utilizada para la medida de información
72	db28	-2,44E+10	Entropía de Shannon
77	db28	-3,90E+10	Entropía de Shannon
81	db28	-3,57E+10	Entropía de Shannon
91	db28	-3,20E+10	Entropía de Shannon
96	db28	-2,34E+10	Entropía de Shannon
97	db28	-3,99E+10	Entropía de Shannon
102	db28	-2,12E+10	Entropía de Shannon
103	db28	-2,99E+10	Entropía de Shannon
104	db28	-3,82E+10	Entropía de Shannon
105	db28	-2,72E+10	Entropía de Shannon
106	db28	-2,05E+10	Entropía de Shannon
107	db28	-4,25E+10	Entropía de Shannon

En el perfil Rxo según la información almacenadas en los Cuadros 15 y 16, la onda madre que mejor se adapta es la db28 para el método de Entropía de Shannon y Logaritmo de Energía.

Dentro de las Wavelet utilizadas para la extracción de características aquella con las que se tuvo mejores resultados fue la db28. Este método hace la identificación de la onda madre con referencia a la adaptación en todo el perfil a lo largo de todo el pozo. La wavelet hace una caracterización con mayor cantidad de componentes, que aumenta la dimensionalidad causando el incremento en el costo de información. Para obtener una mejor caracterización se desarrolló un algoritmo que evaluara los perfiles a través de diferentes ventanas como se observa en la Figura 16, en donde se escogía la ventana dependiendo de la profundidad de la zona y la respectiva segmentación utilizada por ECOPELROL.

Cuadro 17. Vectores característicos a diferentes segmentos

Onda Madre	Vector característico
db28	4X98

5.2 SELECCIÓN MEJOR ONDA MADRE POR SEGMENTOS (ft)

Se desarrolló un algoritmo que evaluara las zonas reservorias a través de diferentes segmentos a 4(ft) ,5(ft), 6(ft), 7(ft), 8(ft), 10(ft), 12(ft), en donde se observó su respectiva representación aplicando el algoritmo para selección de la onda madre, de esta forma se seleccionó la onda que mejor se adaptó. Durante la evaluación de las diferentes zonas de todo el pozo se registraron los vectores característicos que contienen menor dimensión y mejor adaptación de la onda madre los valores de estos vectores se encuentran registrados en el Cuadro 18.

Las ondas Wavelets que entregaron los mejores resultados donde se logró la mejor caracterización por segmentos, fueron las siguientes ondas madre sym1, rbior 1.1, db2, bior 3.1. Los resultados que arrojó el algoritmo para el pozo LL-72 como en el pozo LL-77 empleando cada uno de los perfiles fueron:

Cuadro 18. Vectores característicos a diferentes segmentos

Onda Madre	Vector característico
db 2	3X 6
sym 1	3X5
rbio 1.1	3X5
bior 3.1	4 X 6

De estos resultados se seleccionó como onda madre la bior 3.1 porque se adaptó mejor con la Entropía de Shannon y logaritmo de la Energía, para entregar un vector característico por segmentos de 4x8 de buena representación.

5.3 REDUCCION DE DATOS DEL PCA

En la etapa de procesamiento para la representación de la señal, el vector característico de toda la señal con la onda madre bior 3.1 fue de 1x22, para cada perfil y 6x22 representativos para todos los perfiles, se tiene un volumen de 132 componentes. Después de organizar el vector característico como vector fila y como vector columna se logró obtener una mejor extracción de características con el vector característico organizado como vector fila. Los coeficientes fueron organizados en un vector fila de 6x22 es decir un volumen de 132 datos entran al PCA y resulta la reducción de información en un vector de entrenamiento de 4x22 o 88 datos correspondientes a los perfiles. Se deduce una reducción del 33.4% del vector característico.

5.3 RESULTADOS ENTRENAMIENTO

Las siguientes son el comportamiento del entrenamiento mediante los métodos aplicados.

5.3.1 Entrenamiento de las Máquinas por parejas de perfiles y segmentos.

Las primeras pruebas realizadas tuvieron el objetivo de determinar valores óptimos de eficiencia en la etapa de entrenamiento. La búsqueda se inició a partir de un parámetro dado para cada clase 0 para las zonas no productoras y 1 para aquellas zonas productoras, Los resultados de eficiencia en el entrenamiento utilizando diferentes maquinas en la clasificación según las zonas se muestran a continuación en los Cuadros 19 y 20.

Cuadro 19. Eficiencia entrenamiento para zonas de diferente segmentación pozo 72

Longitud de Zona (Pies)	Perfiles correlacionados en la maquina	Performance %
4	GR-SP	77.08
	RX-RT	83.33
	Neutrónico-BulkDensity	83.33
5	GR-SP	61.18
	RX-RT	62,3
	Neutrónico-BulkDensity	62.45
6	GR-SP	56.67

Continuación Cuadro 19.

Longitud de Zona (Pies)	Perfiles correlacionados en la maquina	Performance %
6	RX-RT	54.23
	Neutrónico BulkDensity	55
7	GR-SP	63.57
	RX-RT	63.6
	Neutrónico-BulkDensity	72.6
8	GR-SP	58.9
	RX-RT	62.1
	Neutrónico-BulkDensity	64.45
10	GR-SP	72.9
	RX-RT	83.57
	Neutrónico-BulkDensity	64
12	GR-SP	86.95
	RX-RT	70
	Neutrónico-BulkDensity	87.95

Cuadro 20. Eficiencia entrenamiento para zonas de diferente segmentación pozo
77

Longitud de Zona (Pies)	Perfiles correlacionados en la maquina	Performance %
4	GR-SP	71.4
	RX-RT	79.8
	Neutrónico-BulkDensity	81.4
5	GR-SP	60.7
	RX-RT	60.6
	Neutrónico-BulkDensity	61.8
6	GR-SP	56.67
	RX-RT	53.7
	Neutrónico-BulkDensity	62.7
7	GR-SP	62.1
	RX-RT	70.3

Continuación Cuadro 20.

Longitud de Zona (Pies)	Perfiles correlacionados en la maquina	Performance %
7	Neutrónico-BulkDensity	56.7
8	GR-SP	63.5
	RX-RT	61.6
	Neutrónico-BulkDensity	63.7
10	GR-SP	71.7
	RX-RT	82.4
	Neutrónico-BulkDensity	64
12	GR-SP	85.9
	RX-RT	71
	Neutrónico-BulkDensity	88.3

Debido a que Matlab carece de una herramienta para trabajar en una dimensión mayor a 2 para la SVM se aplicó el método de entrenamiento de las máquinas por parejas de perfiles y segmentos. El resultado fueron 21 Máquinas que después se conectaron entre sí. Con estas máquinas ya entrenadas, al ejecutar el programa para la evaluación y validación de las zonas de interés, esta resultó en un nivel bajo de clasificación debido a que la máquina no logró hacer una buena diferenciación de clases, estos datos entrenados confundieron al clasificador y no arrojó buenos resultados en la validación.

5.3.2 Entrenamiento de todas las zonas. Para intentar otro método de entrenamiento donde se obtuvieran mejores resultados, se tomaron los componentes de la matriz de la Figura 23, que representan a los 6 perfiles en un vector de dimensión 4x22 se organizó en dos vectores de dimensión 2x22 para ingresarlos en la SVM como una sola máquina debido a la restricción de dimensionalidad que tiene Matlab. Este vector de entrenamiento se organiza de tal manera que sea representado en un plano de dimensión en R^2 , para que la máquina lo eleve a R^3 . A continuación en la Figura 25 se muestran los componentes entrenados, y la Figura 26 se muestran los vectores de soporte obtenidos al trazar el hiperplano, de esta clasificación se obtuvo un performance o nivel de eficiencia, del 52.2655%.

Figura 25. Población de entrenamiento con los datos de la matriz del PCA en dos vectores

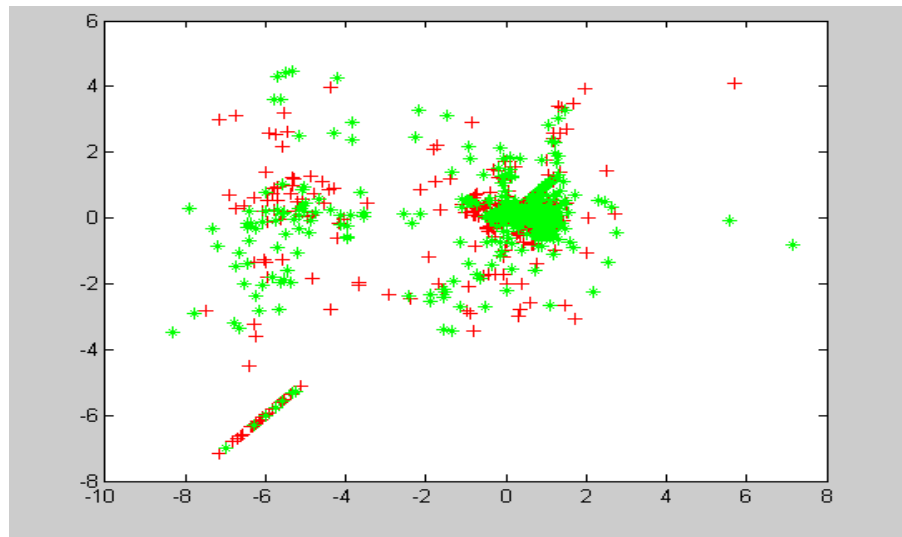
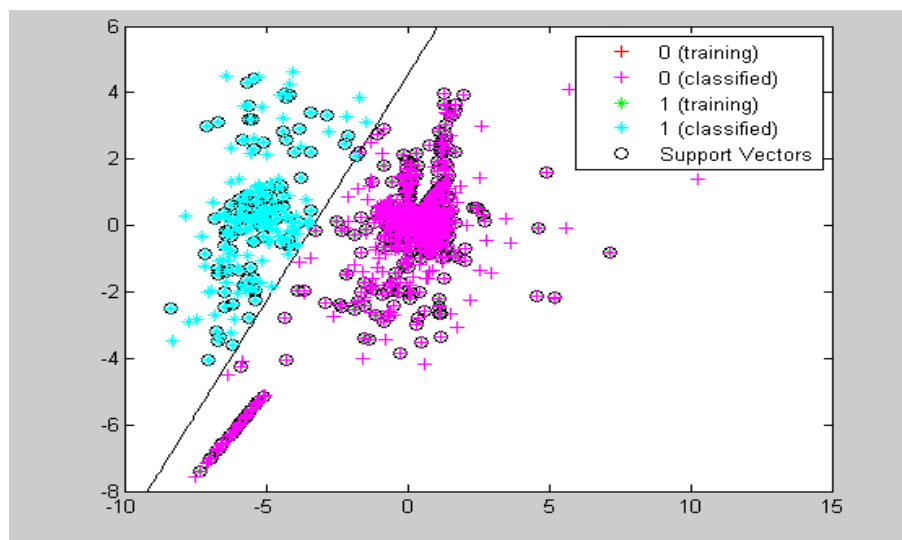


Figura 26. Clasificación población con los datos de la matriz del PCA en dos vectores.



5.3.3 Entrenamiento sin el PCA. Se realizó el mismo proceso con la ausencia de PCA, es decir con los componentes del vector característico arrojado de la Wavelet, en donde igualmente este vector se organizo en dos vectores para el entrenamiento de la SVM. A continuación en la Figura 27 se muestran los componentes entrenados, y la Figura 28 se muestran los vectores de soporte

obtenidos al trazar el hiperplano, de esta clasificación se obtuvo un performance o nivel de eficiencia, del 49.8177%.

Figura 27. Población de entrenamiento sin utilización del PCA

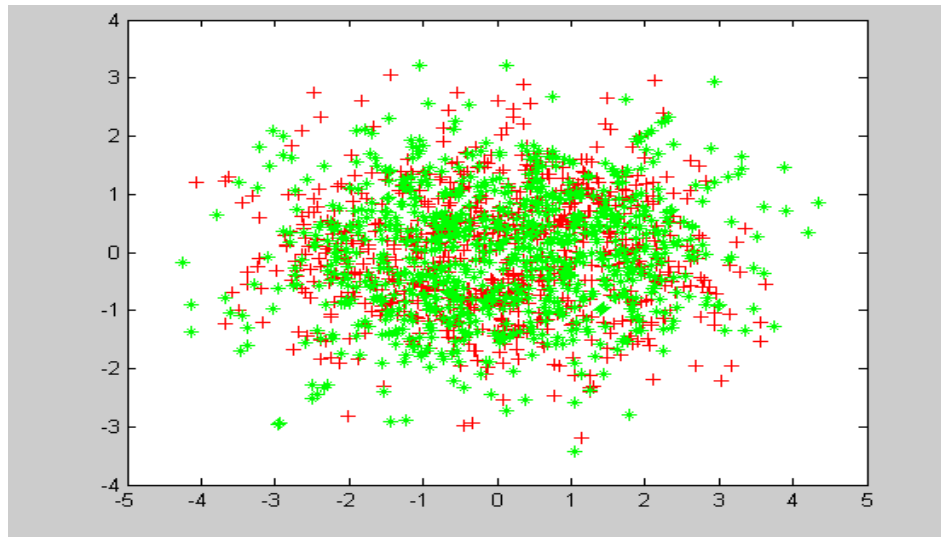
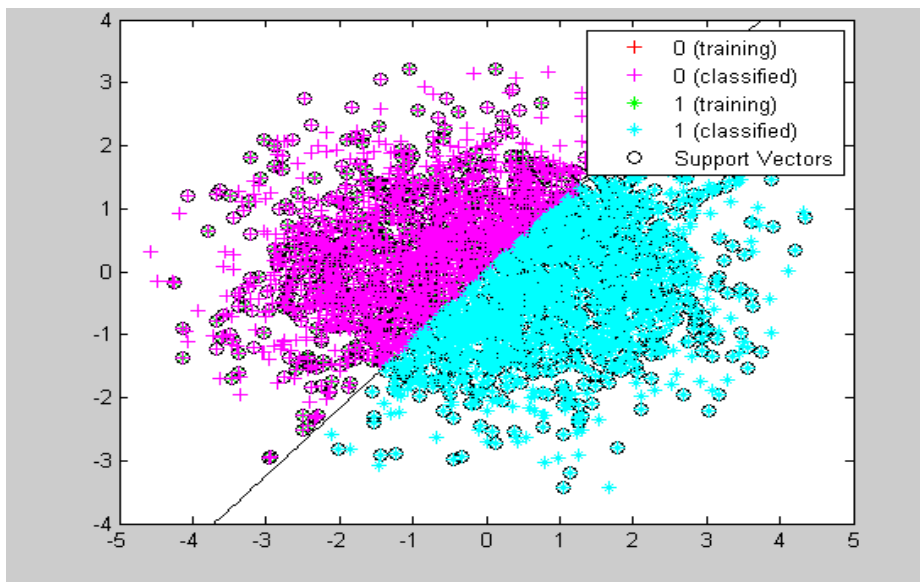


Figura 28. Población clasificada sin la utilización del PCA



Con la utilización del PCA como reductor de dimensionalidad se logró extraer las características relevantes de los perfiles, consiguiendo una mejor selección de características, disminuyendo la dimensionalidad de los componentes. La Máquina

de Soporte Vectorial no pudo hacer diferenciación de clases debido a que no pudo lograr el mayor margen de separación entre ellas.

5.3.4 Comparación resultados SVM respecto al método aplicado en el trabajo anterior con RNA. Los resultados que arroja el clasificador SVM en comparación con el clasificador neuronal fueron un gran número de zonas que tomo como reservorias sin tener características de regiones productoras, ya que su método de clasificación no es flexible al no realizar una aproximación o acercamiento como lo hace la RNA. El clasificador RNA obtuvo una eficiencia alta para identificar las zonas productoras ya que alcanzó un 94.73%, de las zonas reservorias entregadas por Ecopetrol para el pozo LL-72, como se observa en el Cuadro 21, el clasificador SVM también las identifica pero son más las zonas que identifica erróneamente que las zonas de interés, los datos que llegan a la Maquina de Soporte presentan una naturaleza o comportamiento no ideal para que el clasificador pueda entregar buenos resultados

Cuadro 21 Resultados validación de zonas con clasificador neuronal.

ZONA CAÑONEADA		CLASIFICADOR NEURONAL	
TOPE	BASE	TOPE	BASE
5442	5450	5442	5450
5534	5546	5538	5546
5594	5600	5594.5	5599.5
5610	5620	5610.5	5620
5656	5665	5656	5665
5674	5680	5674	5679
5843	5850	5843	5850
5854	5859	5854	5859
5901	5907	***	***
6133	6139	6133	6139
6232	6236	6232	6236
6280	6288	6280	6288

Continuación Cuadro 21.

ZONA CAÑONEADA		CLASIFICADOR NEURONAL	
TOPE	BASE	TOPE	BASE
6400	6405	6400	6405
6410	6416	6410	6415.5
6454	6460	6454	6459.5
6464	6469	6464	6469
6478	6484	6478	6484
6680	6689	6680	6689
6698	6706	6698	6706

6. CONCLUSIONES

- La transformada Wavelet se utilizó como herramienta para la caracterización de señales de comportamiento transitorio. Las Wavelets representan la señal dividiéndola de una manera controlada con una resolución variable a diferentes escalas, con la problemática tiempo-frecuencia que ellas presentan.
- Se destaca la importancia de la utilización del algoritmo de selección de una onda madre que mejor hiciera una representación de la señal, usar una u otra onda es de gran importancia para la representación de esta. Dentro un conjunto de familias el algoritmo escogió la que mejor adaptación presentó a la señal, se desarrolló en cada pozo la utilización de este algoritmo para determinar dentro de todas las ondas Wavelets cual se comportaba mejor.
- Gracias a las funciones de costo de información (Entropía de Shannon y Logaritmo de Energía), se logra identificar donde se encuentra la mayor cantidad de información distribuida que representa la señal, la selección final de la Wavelet madre se obtiene en base a la función que tenga el menor valor de la función de costo. Las funciones de costo de información de la entropía de Shannon encuentra la condición de concentración de información en un vector mientras que el logaritmo de la energía observa la variabilidad que tienen los datos.
- La entropía de Shannon y el logaritmo de Energía se utilizaron como técnicas de funciones de costo de información y mediante su utilización realizar una comparación entre ellas y así determinar cual es mejor para el tratamiento de las señales de perfilaje, se aplicaron para validar en donde se podía encontrar mayor información, y también para la selección de la onda madre en cada uno de los perfiles. Donde se encontró que las dos funciones de costo información dieron como resultado la onda madre bior 3.1.
- En este estudio se utilizó algoritmo de selección de la mejor base de wavelet, el cual usa el conjunto de datos de la señal, y escoge la mas eficiente descomposición tiempo-frecuencia para la representación de la señal.
- Con el PCA se logró extraer las características relevantes de los perfiles, consiguiendo una mejor selección de características, disminuyendo la dimensionalidad del dato.
- La Máquina de Soporte Vectorial como sistema de clasificación no fue eficiente, esto se debe a que los datos que llegan a la SVM presentan una naturaleza o comportamiento no ideal para que el clasificador pueda entregar

buenos resultados, por esta razón la SVM no pudo obtener el margen óptimo de separación y hacer una buena diferenciación de clases.

- Los resultados arrojados por la Máquina de Soporte Vectorial no arrojaron buenos resultados en comparación con la técnica de reconocimiento de patrones por medio de las redes neuronales debido a que la SVM no tiene un porcentaje de acercamiento en la diferenciación de clases, es decir el algoritmo implementado mediante Redes Neuronales entrega valores en un rango de 0 a 1, y la Máquina de Soporte Vectorial arroja resultados que o son 0 o son 1 en la diferenciación de clases y debido a esto no pudo separar las zonas productoras de las zonas no productoras.

BIBLIOGRAFIA

Andrade Trujillo, Guillermo. Polanco, Juan Manuel. (2009). Evaluación de Algoritmos Basados en Lógica Difusa para el Procesamiento de Señales de Registro de Pozo abierto. Universidad Surcolombiana.

Angulo, Naive. Sáenz, José. (2009). Máquinas de Aprendizaje para Clasificar señales Electroencefalográficas. Universidad de Los Andes. Venezuela.

Aljady Erazo, Jorge Héctor. (1999). determinación de la mejor ondilla para análisis de fonemas. Universidad Autónoma Metropolitana.

Barrera, Alejandro. Flores, Jesús. Ricardo, Cámara. (2006). Entrenamiento de maquinas de soporte Vectorial. Universidad autónoma del Carmen CA.

Bernal, Hugo. Rodríguez, Alfredo. (2009). Identificación de zonas productoras de hidrocarburos aplicando DSP Y métodos de correlación de registros de perfilaje. Universidad Surcolombiana.

Betancourt, Gustavo. (2005). Las máquinas de soporte vectorial (SVMs). Universidad Tecnológica de Pereira.

Camargo Puerto, Jorge Arturo. (2007). Introducción a la interpretación de perfiles de pozo abierto. Universidad Surcolombiana.

Coifman, R. Wickerhauser, M. (1992). Entropy based algorithms for best basis selection. Universidad de Yale.

Castellanos, Germán. Sepúlveda, F. (2004). Determinación de voces disfónicas usando bases wavelet discriminantes. Universidad Nacional de Manizales.

Faundez, P. Fuentes, A. (2006). Procesamiento digital de señales acústicas utilizando Wavelets. Instituto de Matemáticas UACH.

Kouro, S. Musalem, R. (2002). Tutorial introductorio a la Teoría de Wavelet. Artículo presentado como trabajo de la asignatura Técnicas Modernas en Automática. Universidad Técnica Federico Santa María.

MathWorks Inc. (2000). Wavelet Toolbox 4 User's Guide. For Use with MATLAB. Natick, MA.

Orozco Alzate, Mauricio. (2003). Clasificación de Arritmias Cardiacas usando Transformada Wavelet y Técnicas de Reconocimiento de Patrones. Universidad Nacional de Colombia.

Pedroza Bernal, Juan Gabriel. (2007). Aplicación de las Máquinas de Soporte Vectorial al reconocimiento de hablantes. Universidad Autónoma Metropolitana.

Serrano, Pedro Eduardo. (2001). Introducción A La Transformada Wavelet Y sus Aplicaciones Al Procesamiento De señales de emisión acústica. Escuela de Ciencia y Tecnología. Universidad Nacional de General San Martín.

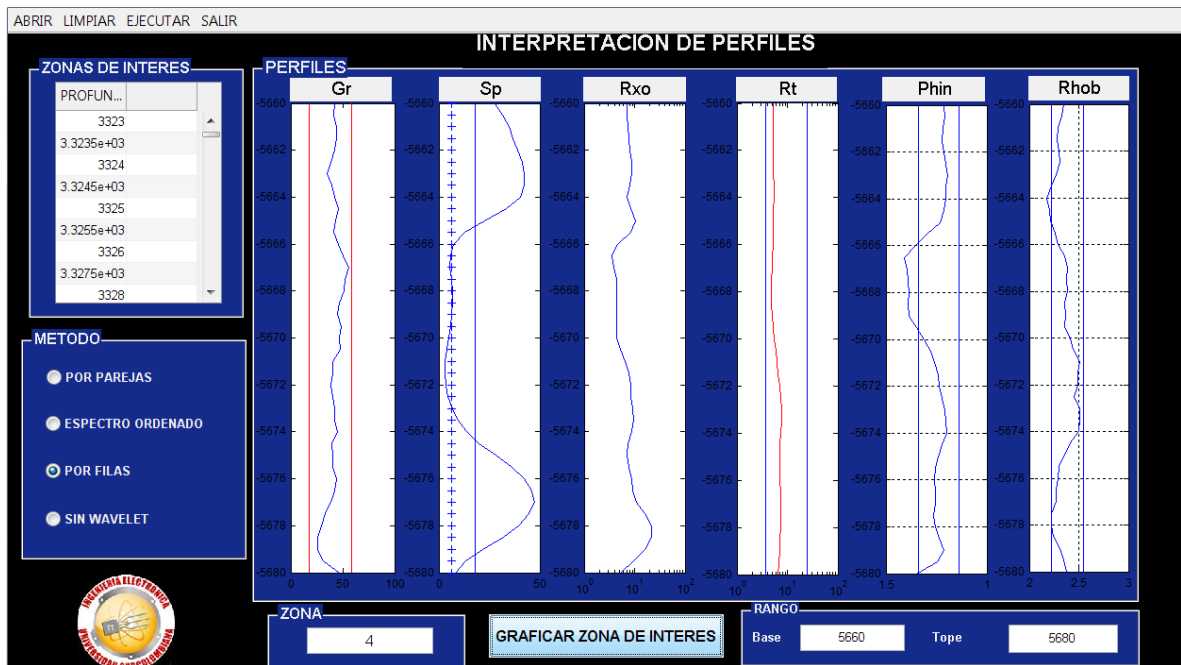
SCHLUMBERGER. (1978). Fundamentos de la Interpretación de Perfiles. Schlumberger Well Service.

Vicente Villardon, José Luis. (2005). Análisis de Componentes Principales (PCA). Departamento de Estadística, Universidad de Salamanca.

ANEXO 1. TUTORIAL DEL PROGRAMA

Este programa es desarrollado en el entorno gráfico de Matlab Guide, que es una interfaz de programación visual que ofrece Matlab para poder realizar y ejecutar programas de Simulación. El Programa de Interpretación de Registros Aplicando MSV es un programa que permite cargar registros de pozo abierto con lecturas de perfiles en formato plano .txt ver Figura 26. A continuación se ilustra el tutorial del programa.

Figura 29. Visualización entorno del programa

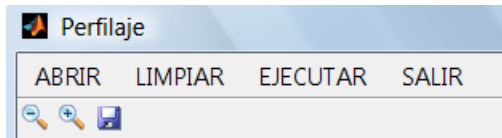


Se muestran los perfiles de pozo abierto, de acuerdo a la zona de interés.

- * Rayos Gamma (GR)
- * Potencial espontaneo (SP)
- * Resistividad zona lavada (Rxo)
- * Resistividad Profunda (RT)
- * Perfil de Densidad (RHOB)
- * Porosidad Neutrónica (PHIN)

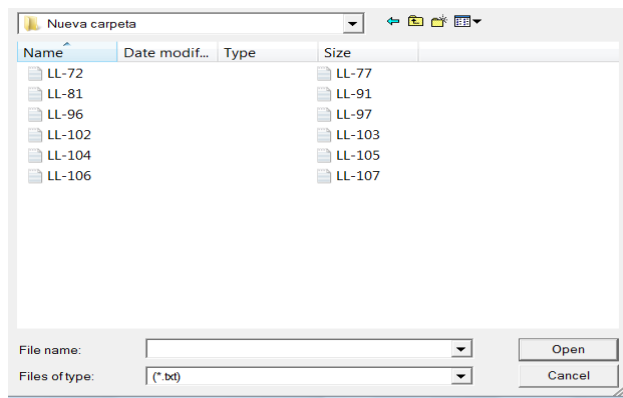
El entorno tiene una barra de menú, como se muestra en la figura 27.

Figura 30 Barra de menú



Abrir > Se examina la dirección donde se encuentra alojado el registro del pozo, luego se carga el registro de pozo abierto para esperar a que se ejecute.

Figura 31. Dirección datos de perfil



Ejecutar > Da inicio a la ejecución del programa.

Limpiar > Realiza un reset al programa blanqueando las graficas para iniciar una nueva carga de registro.

Salir > Cierra el programa

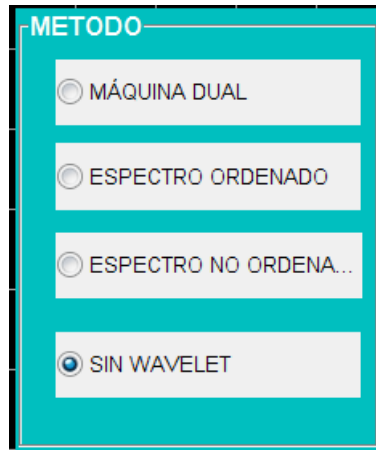
Zoom on > Amplia la zona el eje vertical de cada una de las gráficas.

Zoom off > Disminuye el eje vertical de cada una de las gráficas.

Este submenú da la opción de escoger los métodos de clasificación y entrenamiento empleados en el estudio y anteriormente explicados.

- Métodos empleados en el estudio

Figura 32. Menú métodos aplicados



- Máquina dual: al seleccionar esta opción el programa ejecuta el método de entrenamiento en el que se utiliza una máquina vectorial con la comparación de dos perfiles, entrenamiento por zonas de diferente longitud.

Figura 33. Opción máquina dual



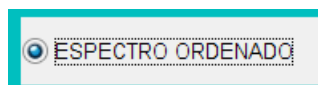
En esta parte se escoge el valor en pies (ft) de la zona o ventana, que van de valores de 4 hasta 16 pies (ft).

Figura 34. Campo para escoger valor de ventana para la máquina dual



- Espectro ordenado: al seleccionar este modo o método, se aplica el ordenamiento del vector característico que entrega el PCA, como vector columna.

Figura 35. Opción por espectro ordenado



Después de haber seleccionado la opción, se inserta la longitud de la zona que se desee analizar. Entre valores de 4 hasta 16 pies (ft).

Figura 36. Campo para escoger valor de ventana para el método por espectro ordenado del vector columna

A rectangular form with a black border. At the top left, the word "ZONA" is written in white. Below it, the text "Inserte valor de zona" is displayed in white. To the right of this text is a white rectangular input field.


- Espectro no ordenado: al seleccionar este modo o método, se aplica el ordenamiento del vector característico que entrega el PCA, como vector fila

Figura 37. Opción por espectro no ordenado

A rectangular button with a cyan border. On the left side, there is a radio button icon. To the right of the icon, the text "ESPECTRO NO ORDENA" is written in black.

Después de haber seleccionado la opción, se inserta la longitud de la zona que se desee analizar. Entre valores de 4 hasta 16 pies (ft).

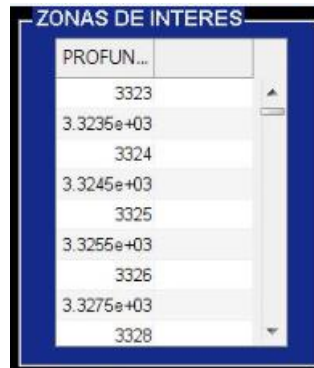
Figura 38. Campo para escoger valor de ventana para el método por espectro no ordenado del vector fila

A rectangular form with a black border. At the top left, the word "ZONA" is written in white. Below it, the text "Inserte valor de zona" is displayed in white. To the right of this text is a white rectangular input field.

- Determinar las zonas de interés:

En esta parte se muestran las zonas identificadas como productoras por el clasificador, entregando el límites superior o tope que es donde inicia la zona, a este valor se le suma el valor de Zona que se indico, para tener la longitud de la zona de interés.

Figura 39. Valores indicando las zonas de interés



PROFUN...
3323
3.3235e+03
3324
3.3245e+03
3325
3.3255e+03
3326
3.3275e+03
3328

- Graficar zonas de interés:

Este submenú, permite observar un segmento en particular de la señal que entrega el clasificador como zona de interés. Se introduce el segmento a graficar, colocando en tope el primer valor o inicio de la zona, y en el cuadro base, se inserta el valor de profundidad final o base de la zona de interés que se desee. Posteriormente se ejecuta dando clic en el botón “Zona de interés”.

Figura 40. Inserción zona de interés



GRAFICAR

Zona de Interes Tope Base