

OPTIMIZACIÓN DE UNIDADES DE BOMBEO MECÁNICO A PARTIR DEL
RECONOCIMIENTO DE CARACTERÍSTICAS POR MEDIO DE MAPAS
AUTORGANIZADOS

LUIS CAMILO GÓMEZ TRUJILLO
MAYRA ALEJANDRA PERALTA

UNIVERSIDAD SURCOLOMBIANA
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA ELECTRÓNICA
NEIVA – HUILA
2012

OPTIMIZACIÓN DE UNIDADES DE BOMBEO MECÁNICO A PARTIR DEL
RECONOCIMIENTO DE CARACTERÍSTICAS POR MEDIO DE MAPAS
AUTORGANIZADOS

LUIS CAMILO GÓMEZ TRUJILLO
MAYRA ALEJANDRA PERALTA

Trabajo de Grado para optar al título de
Ingeniero Electrónico

Director
JOSE SALGADO PATRON
Ingeniero Electrónico

Asesores
HUGO ALBERTO BERNAL PERDOMO
Líder Inteligencia Artificial del Proyecto STAR
LEONARDO FRANCO GRAJALES
Ingeniero de desarrollo del proyecto STAR

UNIVERSIDAD SURCOLOMBIANA
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA ELECTRÓNICA
NEIVA – HUILA
2012

Nota de aceptación:

Firma del presidente del jurado

Firma del jurado

Firma del jurado

Neiva, 13 de diciembre de 2012

Mayra Alejandra Peralta Méndez

A mi Familia y a mi gran amigo.

Luis Camilo Gómez Trujillo

A cada uno de los seres que de alguna forma han influido mi existencia y la persona que soy: mis padres con su ejemplo, enseñanza, apoyo incondicional y gran amor, mi futuro hijo que me motiva a seguir adelante y Gloria por brindarme su amistad y amor incondicional en cada paso de mi vida.

AGRADECIMIENTOS

A nuestras Familias que con su apoyo incondicional nos brindaron comprensión y paciencia para culminar esta meta.

A todas las personas que han intervenido de una u otra forma en nuestro proceso de formación personal, académica y profesional.

Un especial agradecimiento a nuestros asesores, los Ingenieros Hugo Alberto Bernal Perdomo, Líder Inteligencia Artificial del Proyecto STAR y Leonardo Franco Grajales, Ingeniero de desarrollo del proyecto STAR, Pacific Rubiales.

Finalmente reconocer el trabajo en equipo, el compromiso, la disciplina académica y constancia, para terminar con éxito el trabajo.

CONTENIDO

	pág.
INTRODUCCIÓN	19
1. MARCO TEÓRICO	22
1.1 SISTEMA DE BOMBEO MECÁNICO.....	22
1.1.1 Partes del sistema de bombeo mecánico	22
1.1.2 Carta dinográfica	25
1.2 TRANSFORMADA WAVELET	26
1.2.1 Formas wavelet	27
1.2.2 Transformada discreta wavelet	28
1.2.3 Transformada wavelet Packet	30
1.3 REDES NEURONALES ARTIFICIALES	33
2. METODOLOGÍA	40
2.1 IDENTIFICAR SI EL PROBLEMA ES ADECUADO PARA RESOLVERLO POR MEDIO DE REDES NEURONALES.	42
2.2 RECOLECCIÓN DE LA INFORMACIÓN.	42
2.3 DISEÑO Y CONSTRUCCIÓN DE LA BASE DE DATOS.	43
2.3.1 Clasificación cualitativa.....	43
2.3.2 Normalización	44
2.3.3 Extracción de características	46
2.4 ENTRENAMIENTO Y EVALUACIÓN DE LA RED.	49
2.5 PRUEBA Y OPTIMIZACION DEL PROTOTIPO.....	51
3. RESULTADOS.....	55
3.1 TOPOLOGÍA DE LA RED	55
3.2 MATRIZ DE DISTANCIA DE PESOS.....	56
3.3 POSICIÓN DE PESOS	57
3.4 DATOS DE ENTRADA ASOCIADOS.....	58

3.5 TRAYECTORIAS.....	63
4. SOFTWARE DINASOM.....	74
4.1 VALIDACIÓN.....	74
4.2 INTERFAZ GRÁFICA.....	75
4.3 FUNCIONAMIENTO.....	76
4.3.1 Ingreso de datos.....	76
4.3.2 Identificación de fallas.....	77
4.3.3 Predicción de comportamiento.....	78
5. CONCLUSIONES.....	80
6. RECOMENDACIONES.....	81
BIBLIOGRAFÍA.....	82
ANEXOS.....	83

LISTA DE FIGURAS

	pág.
Figura 1. Componentes del sistema de Bombeo Mecánico.	23
Figura 2. Carga del émbolo en el fondo del pozo contra posición.....	24
Figura 3. Carta dinagráfica.....	25
Figura 4. Varias wavelets unidimensionales.....	27
Figura 5. Proceso de TDW, donde $h(n)$, corresponde al filtro pasa bajas y $g(n)$ al filtro pasa altas.	29
Figura 6. Plano Tiempo-frecuencia analizado con TW.....	30
Figura 7. Descomposición Wavelet de una señal.....	31
Figura 8. Árbol de descomposición wavelet packet.....	31
Figura 9. Comparación entre la neurona biológica (A) y neurona artificial (B).	34
Figura 10. Clasificación de redes neuronales.....	34
Figura 11. Modelo Kohonen.	36
Figura 12. Ejemplo de vecindario, donde $t1 < t2 < t3$	37
Figura 13. Arquitectura de una SOM.....	38
Figura 14. Diagrama de bloques general del proyecto.....	40
Figura 15. Algoritmo de Supervisión	41
Figura 16. Tipos de Pozos en bombeo mecánico	43
Figura 17. Información de Carga de Base de Datos.	45
Figura 18. Grafica 3D Información de carga de la Base de Datos.	45
Figura 19. Señal y Aproximaciones.....	48
Figura 20. Señal y Detalles	49
Figura 21. Topología de la red.	55
Figura 22. Matriz U.....	56
Figura 23. Posición de Pesos.....	57
Figura 24. Posición de pesos 3D.....	58
Figura 25. Muestras por neurona	58
Figura 26. Dinagrama representativo por cada neurona.	59

Figura 27. Trayectorias definidas para cada clase.	63
Figura 28. Validación del software	75
Figura 29. Pantalla principal de DinaSOM	76
Figura 30. Abrir un archivo	77
Figura 31. Identificación.	77
Figura 32. Predicción	79

LISTA DE CUADROS

	pág.
Cuadro 1. Parámetros de aprendizaje por defecto	39
Cuadro 2. Análisis de mejor nivel.	47
Cuadro 3. Resultados de estudio de mejor nivel.	47
Cuadro 4. Pruebas de eficiencia de la SOM.....	53
Cuadro 5. Validación del software	74

LISTA DE ANEXOS

pág.

ANEXO A. FORMAS DE CARTAS DINAGRAFICAS DE FONDO	84
ANEXO B. PRECLASIFICACION CUALITATIVA DE LA BASES DE DATOS PROPORCIONADAS POR ECOPETROL.	88

GLOSARIO

ANÁLISIS MULTIRESOLUCIÓN: el análisis multi-resolución consiste básicamente en aproximar una función $f(t)$ de una señal en distintos niveles de resolución, $\{f_1(t), f_2(t), f_3(t), \dots\}$ lo que entrega una descomposición multi-escala de la señal.

BOMBEO MECÁNICO: la función del sistema de bombeo mecánico por cabillas es transmitir la potencia hasta la bomba de fondo, para levantar los fluidos del yacimiento hasta la superficie.

CARTA DINAGRÁFICA: gráfico de cargas vs posición de la barra pulida durante un ciclo completo de bombeo. La medición directa se conoce como carta dinagráfica de superficie y se obtiene con un dinamómetro bien sea analógico o digital.

CARRERA ASCENDENTE Y DESCENDENTE: es el movimiento de la barra pulida o el pistón.

CARTA DINAMOMÉTRICA DE FONDO: gráfico de cargas contra posición donde el principal factor es el peso de la columna de fluido. Esto usualmente se calcula con programas que incluyan modelos de la ecuación de onda que permitan trasladar las cargas vistas en la barra pulida hacia el pistón.

COEFICIENTES DE APROXIMACIÓN: secuencia producida mediante submuestreo de un filtro pasa bajas asociado a cierta wavelet.

COEFICIENTES DE DETALLE: secuencia producida mediante submuestreo de un filtro pasa altas asociado a cierta wavelet.

CONJUNTO DE ENTRENAMIENTO: colección de pares de entrenamiento que consiste en un vector de valores de entrada junto con un vector de salidas deseadas.

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

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

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

DAUBECHIES: familia de wavelet creadas por Ingrid Daubechies, profesora del departamento de matemáticas de la Universidad de Princeton dedicada a la investigación en teoría y aplicaciones del Análisis Tiempo-Frecuencia.

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

DETALLE: secuencia filtrada mediante el filtro pasa altas asociado a cierta wavelet.

DIÁDICO: relativo a potencias de dos.

DINAGRAMA: es una herramienta gráfica en forma de diagrama, utilizada para el diagnóstico de fallas, diseño y balanceo de estructuras de bombeo mecánico. Existen expertos en el análisis e interpretación de estas cuyo criterio y experiencia permiten solucionar los diferentes problemas que conllevan estas estructuras en el levantamiento artificial de pozos petrolíferos.

DINAMÓMETRO: instrumento equipado con una impresora o computador para registrar los datos de las cargas de la barra pulida contra las posiciones de la barra pulida (carta dinagráfica) en un pozo. El dinamómetro consiste en una caja registradora, un transductor de posición y una celda de carga. La celda de carga y el transductor de posición se unen a la barra pulida para registrar cartas dinagráficas.

DISTANCIA MANHATAN: es una forma de distancia geometría donde la métrica usual de la geometría euclidiana es reemplazada por una nueva métrica donde la distancia entre dos puntos es la suma de las diferencias (absolutas) de sus coordenadas.

DISTANCIA EUCLIDIANA: es la distancia "ordinaria" (que se mediría con una regla de acero) entre dos puntos de un espacio euclídeo, la cual se deduce a partir del teorema de Pitágoras.

DOWNSAMPLING: reducir la cantidad de datos de una secuencia, mediante la disminución de la frecuencia de muestreo.

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

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

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

FUNCIÓN GAUSSIANA: es una función que es simétrica y su forma es de campana, por esta razón se denomina generalmente campana de Gauss. La característica más elemental de la curva en cuestión es que es necesaria y muy buena forma de representar la distribución de variables aleatorias. Por esta razón es muy útil para realizar cálculos de estadística. La altura de la campana se centra en el origen donde este punto en estadística corresponde al valor de la media y la anchura de la misma que representa el valor de la varianza.

LEVANTAMIENTO ARTIFICIAL: es el levantamiento desde el subsuelo a la superficie de los fluidos de un pozo por medio de bombeo mecánico, bomba sumergible, por gas, bomba jet, etc. El levantamiento artificial es necesario cuando la presión de fondo no es suficiente para que fluido del pozo fluya por sí solo.

MEYER: wavelet básica cuya función de escalamiento están definidas en el dominio de la frecuencia.

NORMALIZACIÓN: ajuste de una serie (vector) de valores (representando típicamente un conjunto de medidas de acuerdo con alguna función de

transformación para hacerlos comparables con algún punto específico de referencia.

NEUROCOMPUTACIÓN: la neurociencia computacional es una ciencia interdisciplinaria que enlaza los diversos campos de la ciencia: la ciencia cognitiva, la ingeniería eléctrica, las ciencias de la computación, la física y las matemáticas. El término fue introducido por Eric L. Schwartz en una conferencia de 1985 en Carmel, California, a petición de la Systems Development Foundation, para ofrecer un resumen del estado actual de un campo que hasta entonces era nombrado de muy diversas formas: modelado neural, teoría cerebral y redes neurales.

RED NEURONAL ARTIFICIAL: miembro de una clase de algoritmos matemáticos inspirados en el funcionamiento de las redes neuronales biológicas, este algoritmo tiene la capacidad de aprender y es entrenado mediante la presentación de ejemplos de entrada y las correspondientes salidas deseadas.

RESOLUCIÓN: la resolución viene determinada por la longitud de la palabra digital (número de bits), es decir por la agrupación de ceros y unos, con que se va componiendo (codificando) la señal.

SOM: representaciones de sistemas neuronales que se define como mapas autorganizados por sus siglas en inglés (self organizing map) y tienen la propiedad especial de crear de forma efectiva representaciones internas espacialmente organizadas de varias características de las señales de entrada y sus abstracciones.

SYMLET: wavelets casi simétricas propuestas por Ingrid Daubechies como modificaciones a la familia Db (Daubechies). Las propiedades de las dos familias wavelet son similares.

TRANSFORMADA WAVELET: es la descomposición de la función $f(x)$ en una base de funciones formada por la traslación y dilatación de una misma función, la función de escala.

TBT(The Best Tree): algoritmo que genera un árbol completo binario, el cual contiene nodos con la información de la Transformada Wavelet Packet y el objetivo de esta es obtener los nodos que se ajuste lo mejor posible a las características de la señal (mejor base).

TWM (Total Well Management): *software* de la empresa ECHOMETER, gran proveedor de equipos de medición para bombeo mecánico. Este programa permite visualizar simultáneamente la carta de superficie y fondo, la carta de superficie es capturada por *hardware* ECHOMETER y enviada al *software* TWM para calcular la carta dinográfica de fondo por medio de la ecuación de onda y realizar los análisis de inspección, verificación, diagnóstico y planes de mejora.

TOPOLOGIA: la topología, es el número de elementos procesados que forman la red y las interconexiones existentes entre ellos.

WAVELET: es una forma de onda de duración efectiva limitada que tiene un valor promedio igual a cero. La wavelet tiende a ser irregulares y asimétricas.

WAVELET PACKED: el análisis wavelet packed es una generalización de la descomposición wavelet que ofrece un amplio rango de posibilidades para realizar análisis de señales.

RESUMEN

Este trabajo presenta una alternativa de identificación y predicción, cualitativa y cuantitativa, de los problemas en el equipo de subsuelo de sistemas de bombeo mecánico, reflejados en los dinagramas de fondo para establecer una metodología en análisis de la información de 512 dinagramas de 8 pozos de la región norte del departamento del Huila, proporcionada por ECOPELROL.

Luego de hacer una recolección y clasificación manual de los dinagramas, y de formar una base de datos, se identificó si el problema era adecuado para ser resuelto por medio de redes neuronales.

Posteriormente se aclara el procedimiento para extraer las características más significativas de cada dinagrama por medio de un método de procesamiento de señales llamado Transformada Wavelet.

Se diseñó y probó distintos algoritmos de la transformada para obtener los correspondientes coeficientes de aproximaciones y detalles de las señales, para luego aplicarlos al entrenamiento de una red neuronal de mapas auto organizados (Self Organizing Maps (SOM)), capaz de realizar su interpretación de una forma más rápida y asertiva.

Los resultados de estas etapas de la investigación se adaptaron a un software final, que suministrara: Identificación y predicción de manera ágil y precisa de los problemas en los equipos de bombeo con el fin de minimizar los costos de operación y maximizar la producción en la industria.

ABSTRACT

This paper presents an alternative for identification and prediction, both qualitative and quantitative, of equipment problems in rod pumping systems, reflected in the underground dynagraph to establish a methodology of data analysis of 512 dinagramas in 8 wells, provided by ECOPETROL, in the northern region of Huila.

After making a collection and manual classification of dynagraphs, and to make a database, the problem identified was suitable to be solved by neural networks.

Later, clarify the procedure for extracting the most significant characteristics of each dynagraph by a signal processing method called Wavelet Transform.

Was designed and tested different algorithms transform to get the corresponding coefficients of approximations and details of the signals, and then apply them to the training of a neural network of self-organizing maps (SOM), capable of performing their interpretation of a faster and assertive.

The results of these research phases of a software suit final, to supply: Identification and prediction in an agile and precise problems in pumping equipment in order to minimize operating costs and maximize production in industry.

INTRODUCCIÓN

La industria del petróleo cada día busca optimizar sus métodos tanto para la exploración, explotación como para la producción de petróleo. Es así, como las empresas encargadas de este medio tratan de encontrar técnicas Neuro-computacionales que les garanticen de forma eficiente y óptima buenos resultados, contrario, a lo que las técnicas antiguas o convencionales ofrecen.

En la actualidad hay muchos estudios en distintas áreas de la Neuro-computación, acerca de la forma de aplicar las redes neuronales a las tareas diarias, como en las industrias, que han retomado ostensiblemente tales técnicas (redes neuronales), debido al aumento de la capacidad de procesamiento de las máquinas y la posibilidad de poder aplicar este método de análisis a tareas dispendiosas, realizadas por personas expertas.

En las diferentes universidades del mundo se están desarrollando estudios e investigaciones para adaptar el desarrollo de las redes neuronales a todos los campos del conocimiento, la industria petrolera, destina grandes recursos económicos para propiciar el desarrollo de técnicas que mejoren sus producciones, por esto, países como Venezuela en empresas como PDVSA apoyan el desarrollo de estas herramientas en instituciones tales como la Universidad de Oriente, en la escuela de Ingeniería de Petróleos y la Universidad de Zulia.

La producción promedio de crudo actualmente en Colombia se ubicó en 955.000 barriles por día (bpd)¹, es decir un 5,4% más frente a los 906.000 barriles que se registraron en abril de 2011². Por esta razón Colombia está ubicada en la posición 25 de los países productores de petróleo. Gran parte de esta producción proviene de pozos que utilizan el bombeo mecánico como su método de levantamiento artificial debido a que es el más económico y más fácil de mantener cuando es diseñado y operado apropiadamente. Es así como en Colombia, la Universidad Industrial de Santander (UIS), con su Grupo de Investigación en Exploración y Explotación de Hidrocarburos, poseen varios estudios de este tipo de aplicación de redes neuronales.

En el Huila, en la Universidad Surcolombiana está naciendo el interés de desarrollar proyectos basados en las técnicas de Neuro-computación (redes

¹ Disponible en Elespectador.com, 15 Mayo 2012.

² Disponible en: http://www.indexmundi.com/es/colombia/petroleo_produccion.html

neuronales), pero no existen grupos de investigación que propicien el desarrollo más acelerado de este tema. Por eso, esta investigación se ha interesado en continuar el trabajo iniciado por el ingeniero Leonardo Franco y el Ingeniero Carlos Pérez, con el fin de que sirva de inspiración para los que quieran continuar con el desarrollo en el campo de las redes neuronales y la aplicación de técnicas de procesamiento de señales tales como las de wavelet, y con ello, ampliar el conocimiento y mejorar la calidad de la investigación en la universidad.

La supervisión del bombeo mecánico produce un gran flujo de datos e información (dinagramas, perfiles de exploración, etc.) que necesita ser analizado e interpretado. En muchos casos el análisis de toda esta información se hace de forma manual y esto se torna tedioso, lento y agotador para los ingenieros expertos, por tanto es necesario desarrollar herramientas expertas para el análisis automático aplicado a la detección, predicción de problemas que permitan tomar de forma rápida decisiones acertadas y eficientes para la solución de los mismos.

Motivados por los nuevos adelantos de análisis de señales, como la Transformada de Wavelet (TW) y redes de clasificación autorganizados, SOM, que han demostrado ser una herramienta muy poderosa en Minería de Datos (Data Mining) y en metodología de Descubrimiento de Conocimiento en Base de Datos (Knowledge Discovery Database (KDD)) y con una gran variedad de aplicaciones de ingeniería tales como reconocimiento de patrones, análisis de imágenes, monitoreo de procesos y detección de fallas por nombrar algunas, el software DinaSOM, busca aplicar estos conceptos para el análisis e identificación de la información representada en cartas dinagráficas, además se desarrolló una herramienta experta que permite optimizar unidades de bombeo mecánico de forma más apropiada.

Para el trabajo propuesto se desarrolló un software experto para la optimización de unidades de bombeo mecánico, agregando descripciones de comportamientos el cual contiene: análisis de registros históricos, el estudio de trayectorias de fallas asociadas y el monitoreo de múltiples pozos presentándolos por medio de una interfaz gráfica tipo radar, con múltiples sub-pantallas para que se puedan realizar distintos tipos de observaciones sobre el comportamiento de las fallas.

Con base en los objetivos planteados, se desarrolló una explicación de la teoría básica del levantamiento artificial por medio del bombeo mecánico y las formas para realizar un registro del estado de los sistemas dinámicos que compone todo el sistema de bombeo, este sistema de registro se realiza por medio de dinagramas, que son una cantidad de datos registrados por una computadora que representan el desplazamiento de la bomba versus la fuerza que ejerce esta.

Luego, para estos datos es necesaria una pre-clasificación manual de las cartas dinagráficas de los diferentes pozos suministrados por Ecopetrol y en la clasificación se tiene en cuenta la profundidad de los mismos, ya que por inercia del fluido las mismas formas dinagráficas proporcionan fallas diferentes.

Después de realizar la clasificación manual, los datos son sometidos a un análisis de señales por medio de la TW donde se hace una comparación con el análisis de la Transformada de Fourier y se justifica la importancia y necesidad de utilizar la TW.

En la etapa siguiente, se integran generalidades de las redes neuronales artificiales, arquitectura e implementación de uno de los algoritmos más representativos y poco usados, los modelos de Mapas Auto organizativos (Self Organizing Maps (SOM)).

Por último se realiza la validación del software por medio de estadísticas aplicadas para obtener las configuraciones del análisis wavelet más eficientes y el entrenamiento más adecuado para los SOM y así obtener un mejor resultado a la hora de realizar un diagnóstico de fallas para los pozos seleccionados.

1. MARCO TEÓRICO

1.1 SISTEMA DE BOMBEO MECÁNICO

De los métodos de levantamiento artificial, el Sistema de bombeo mecánico es usualmente el más económico y fácil de mantener cuando es diseñado y operado apropiadamente.

Para optimizar el comportamiento de sistemas por bombeo mecánico es importante identificar y entender los problemas que reducen la rentabilidad. Las dos principales fuentes de reducción de rentabilidad son baja eficiencia del sistema y fallas en el equipo³. Al encontrar las posibles fallas a tiempo mejora la eficiencia en un porcentaje alto, y los métodos de predicción autónomos son la mejor forma de encontrar esas fallas a tiempo sin producir gastos innecesarios.

1.1.1 Partes del sistema de bombeo mecánico. La función del sistema de bombeo mecánico por cabillas es transmitir la potencia hasta la bomba de fondo para levantar los fluidos del yacimiento hasta la superficie, para lograr este objetivo este sistema consta de un equipo de superficie y de fondo, como se observa en la *Figura 1*.

El equipo de superficie incluye la unidad motriz, unidad de bombeo, barra pulida, prensa estopa y líneas de flujo. El equipo de fondo incluye el revestidor, tubería de producción, sarta de cabillas, bomba de fondo⁴.

1.1.1.1 Unidad motriz. Es un motor eléctrico o a gas. Su función es suministrar la potencia que el sistema de bombeo mecánico necesita.

1.1.1.2 Unidad de bombeo. Es la parte mecánica visible en la superficie y su función, es convertir el movimiento rotacional de la unidad motriz al movimiento ascendente-descendente de la barra pulida.

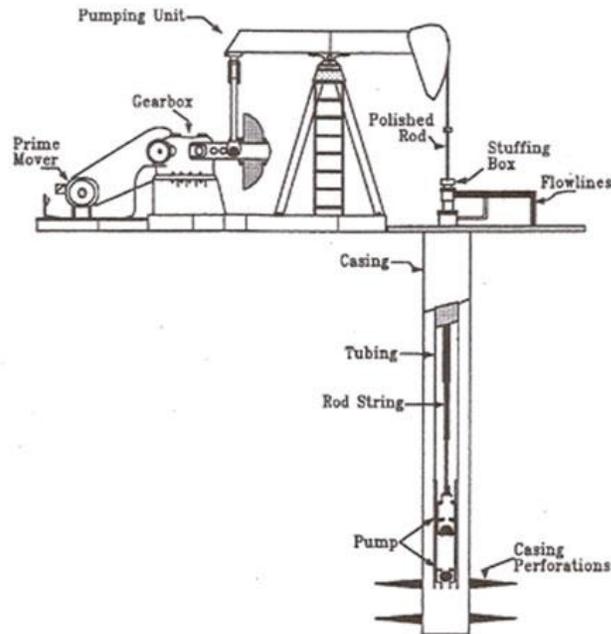
1.1.1.3 Caja de engranajes y contrapesos. La caja de engranajes consiste en una caja de velocidades (también llamada simplemente caja) es el elemento encargado de obtener en la unidad el par motor suficiente para poner en

³ THETA ENTERPRISES, Bombeo mecánico Optimización. Bera, CA 92821-USA: Theta Enterprise, Inc, 2005. p. 15.

⁴ Ibid., p. 27.

movimiento la bomba por medio de las cabillas desde parado, y una vez en marcha obtener un par suficiente sobre la bomba para vencer las resistencias debido a los fluidos al avance y una función de la caja de engranajes es convertir torques bajos y altas rpm de la unidad motriz en torques altos y bajas rpm necesarias para operar la unidad de bombeo.

Figura 1. Componentes del sistema de Bombeo Mecánico.



Fuente: Bombeo Mecánico Optimización. Theta Enterprise, Inc, 2005

1.1.1.4 Barra pulida. Conecta la unidad de bombeo a la sarta de cabillas y es la única parte de la sarta que es visible en la superficie. Su función principal es soportar el peso de la sarta de cabillas, bomba y fluido.

1.1.1.5 Prensa estopa. Las empaaduras del prensa estopa son apretadas para prevenir fugas en el cabezal. Pero, si se aprietan demasiado, podrían incrementarse las pérdidas de potencia en la barra pulida resultando en una mala interpretación de la carta dinográfica por la distorsión de las cargas sobre la barra pulida.

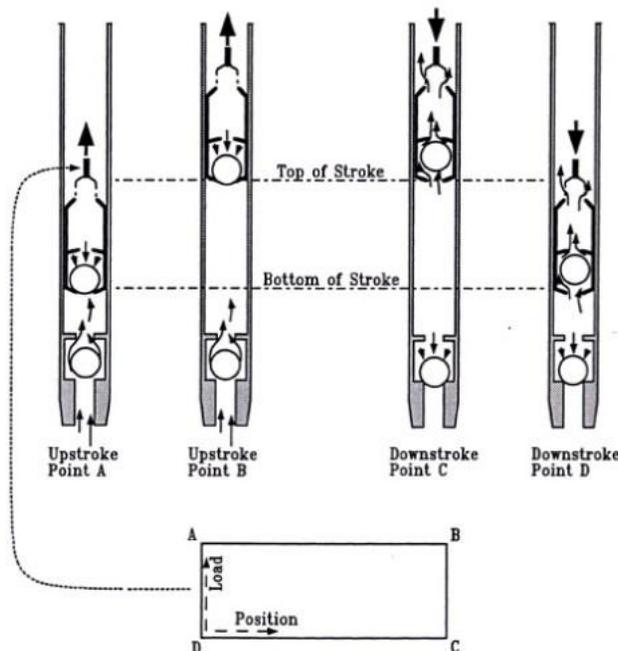
1.1.1.6 Sarta de cabillas. conecta la bomba de fondo con la barra pulida. La función principal es transmitir el movimiento oscilatorio de la barra pulida a la bomba. Esto proporciona la potencia necesaria por la bomba, para producir

hidrocarburos. La resistencia, vida útil y fuerzas friccionales de la sarta de cabillas tiene un impacto significativo en la economía de un pozo.

1.1.1.7 Bomba de subsuelo. La típica bomba de succión es un arreglo embolo-cilindro. El pistón tiene una válvula de bola y asiento llamado válvula viajera debido a que viaja hacia arriba y hacia abajo con el pistón. A la entrada del barril de la bomba existe otra válvula llamada válvula fija debido a que está fijada a la tubería y no se mueve. La operación de la bomba afecta todos los componentes del sistema.

○ **Acción de las Válvulas y Cargas de Fluido.** La acción de las válvulas ayudan a entender como las cargas del fluido ejercen diferentes fuerzas al pistón de la bomba y la sarta de cabillas formando la carta dinagráfica plasmando en estas el comportamiento de las cabillas de succión. La carta dinagráfica es una representación gráfica del estado de la bomba que muestra la carga versus posición. Si existiera la posibilidad colocarse un instrumento para medir las cargas justo arriba del pistón de la bomba, se terminaría con una carta dinagráfica de fondo. Para entender como seria la carta dinagráfica de la bomba para el caso de bomba llena, Veamos la *Figura 2*, y su posterior explicación.

Figura 2. Carga del émbolo en el fondo del pozo contra posición.



Fuente: Bombeo Mecánico Optimización. Theta Enterprise, Inc, 2005

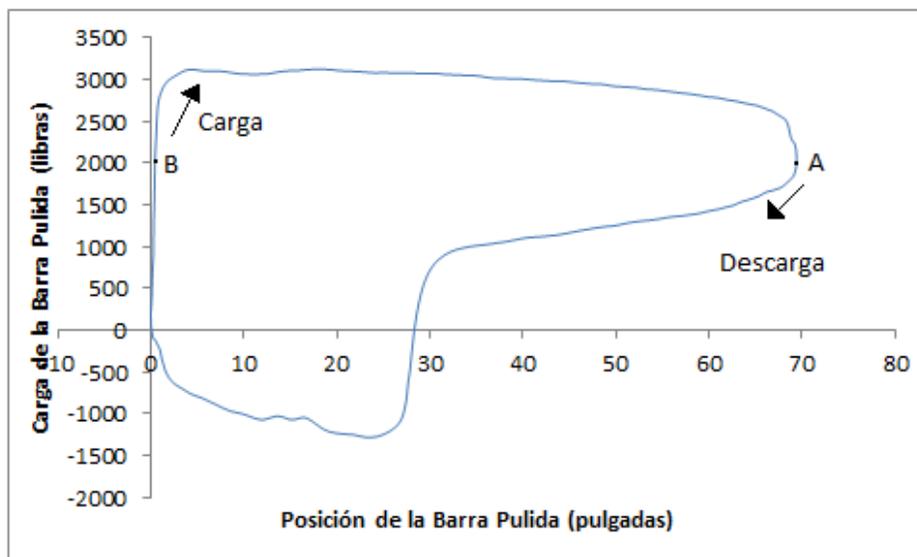
a) **Carrera ascendente.** Al inicio de la carrera ascendente, la válvula viajera realiza el cierre (*punto A de la Figura 2*) en este punto la válvula viajera comienza a levantar el fluido. Estas cargas permanecen constantes durante el recorrido ascendente (*de punto A hasta B*).

b) **Carrera descendente.** Cuando el pistón se mueve hacia abajo, la válvula viajera abre (*punto C*) en este momento la válvula viajera se libera del fluido y la presión de este se transfiere a la tubería a través de la válvula fija. Por lo tanto, la válvula viajera no lleva la carga de fluido durante la carrera ascendente (*desde el punto C al D*).

Las diferencias de carga que experimentan los puntos (*A y D*) ó (*B y C*) son las cargas del fluido sobre el pistón. Que son transferidas instantáneamente desde *D* hasta *A* y desde *B* hasta *C*. La cantidad de fluido bombeado depende de la integridad de la bomba, el tipo de fluido bombeado, espaciamento de la bomba, y si la tuberías está o no anclada, estos fenómenos se ven reflejados en los dinagramas.

1.1.2 Carta dinagráfica. Es un gráfico de cargas vs. posición registrado por un sistema dinamométrico desde la barra pulida. La *Figura 3.* muestra un ejemplo de un gráfico de carta dinagráfica.

Figura 3. Carta dinagráfica.



Existen 2 tipos de dinagramas, dinagramas de superficie y de fondo, siendo el ultimo el más adecuado y fácil de interpretar, ya que su forma depende de lo que esté pasando en la bomba. Por ejemplo, la forma de una carta dinagráfica de fondo correspondiente a bomba llena será la misma, sin importar la profundidad, unidad de bombeo, sarta de cabillas, velocidad de bombeo, etc.

1.1.2.1 Dinagrama de superficie. Registro de cargas sobre la barra pulida para determinar en superficie el comportamiento de la bomba durante cada ciclo de bombeo.

1.1.2.2 Dinagrama de fondo. Es obtenido por medio de la solución de ecuación de onda unidimensional amortiguada que modela el comportamiento de la sarta de cabillas. La técnica para la solución de la ecuación de onda depende de un modelo matemático de la sarta de cabillas numéricamente resuelto por computadora. El método se puede usar para calcular cartas dinagráficas de fondo a partir de una carta dinagráficas de superficie.

1.2 TRANSFORMADA WAVELET

El análisis wavelet es un nuevo conjunto de herramientas y técnicas capaces de revelar aspectos de datos que otras técnicas de análisis de señales pierden, aspectos como puntos de ruptura, discontinuidades en derivadas de orden superior, y auto similitudes⁵. Ya que ofrece un punto de vista diferente que los presentados por las técnicas tradicionales, el análisis wavelet puede codificar, comprimir, limpiar, reconstruir, describir, simplificar, modelar y distinguir una señal sin degradación apreciable.

La Transformada Wavelet (TW) es ideal para el análisis de señales que presenten cambios abruptos en forma instantánea a través de ventanas de longitud variable adaptadas al cambio de frecuencia de la señal. Esta técnica permite el uso de intervalos grandes de tiempo en los que queremos información de baja frecuencia más precisa y más cortos en las regiones donde se quiere información de alta frecuencia⁶.

⁵ MISITI, Michel, *et al.* Wavelet Toolbox for use with MATLAB. [s.l]: The MathWorks, 1996-1997. p. 1.2

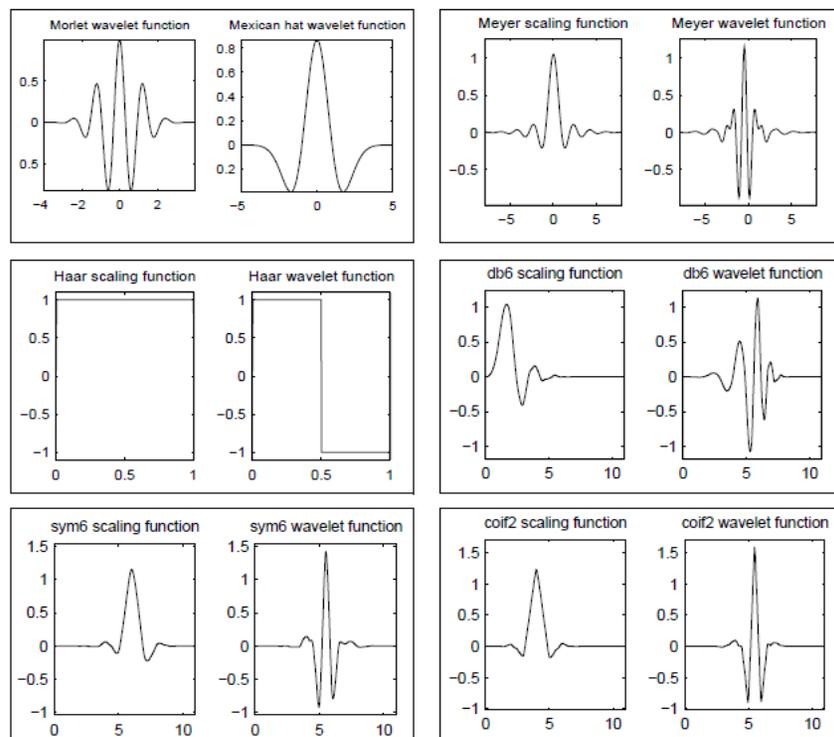
⁶ *Ibid.*, p. 1.5

Desde un punto de vista intuitivo, la descomposición wavelet consiste en calcular un “índice de semejanza” entre la señal y una wavelet localizada en una posición b y de escala a . Si el índice es grande, la semejanza es muy fuerte, de lo contrario es leve. Los índices $C(a,b)$ son llamados coeficientes; su variable de escala (a) lleva información de frecuencia y la variable de traslación (b) tiene información de tiempo.

1.2.1 Formas de Wavelet. A diferencia de Fourier, donde las funciones base son senos y cosenos de duración infinita, en el análisis wavelet la base son funciones localizadas en frecuencia y tiempo.

Una wavelet es una pequeña onda de duración limitada, es decir, su energía está concentrada en el tiempo alrededor de un punto, lo que proporciona una adecuada herramienta para el análisis de fenómenos transitorios, no estacionarios, variables en el tiempo y aquellos que presenten discontinuidades⁷. En la *Figura 4*. se muestra diferentes tipos de wavelets.

Figura 4. Varias wavelets unidimensionales.



Fuente: Wavelet toolbox User's Guide

⁷ BERNAL, Hugo y RODRIGUEZ, Alfredo. Identificación de zonas productoras de hidrocarburos aplicando DSP y métodos de correlación de registro de perfilaje. Trabajo de grado Ingeniero Electrónico. Neiva. Universidad Surcolombiana. Facultad de Ingeniería, 2009, p. 18.

Las wavelet se clasifican en familias dependiendo de propiedades como simetría, momentos de desvanecimiento, regularidad, ortogonalidad o biortogonalidad y a su vez cada familia posee órdenes. Las familias wavelets incluidas en Matlab son: Morlet, Mexican Hat, Meyer, Daubechies, Symlets, Coiflets y Biortogonales.

1.2.2 Transformada Discreta Wavelet (TDW). El cálculo de los coeficientes wavelet a todas las escalas posibles es una buena cantidad de trabajo y genera una gran cantidad de datos⁸. Para un análisis mucho más eficiente y más preciso, se han escogido escalas y posiciones basadas en potencias de dos, llamadas escalas y posiciones diádicas definidas por las *Ecuaciones (1) y (2)* respectivamente:

$$a = 2^j, j \in N \quad (1)$$

$$b = ka, k \in Z \quad (2)$$

Donde j es el nivel, $1/a$ es la resolución. Entonces, conforme la resolución incrementa la escala decrece.

Para obtener coeficientes de tipo discreto se usa una función g definida en Z y una escala igual a 2^j , para $j \in N$ y $k \in N$, entonces $g_{j,k}$ es:

$$n \in Z, \quad g_{j,k}(n) = 2^{-j/2} g(2^{-j}n - k) \quad (3)$$

De tal modo que los coeficientes para análisis discreto son:

$$C(a, b) = C(j, k) = \sum_{n \in Z} s(n) g_{j,k}(n) \quad (4)$$

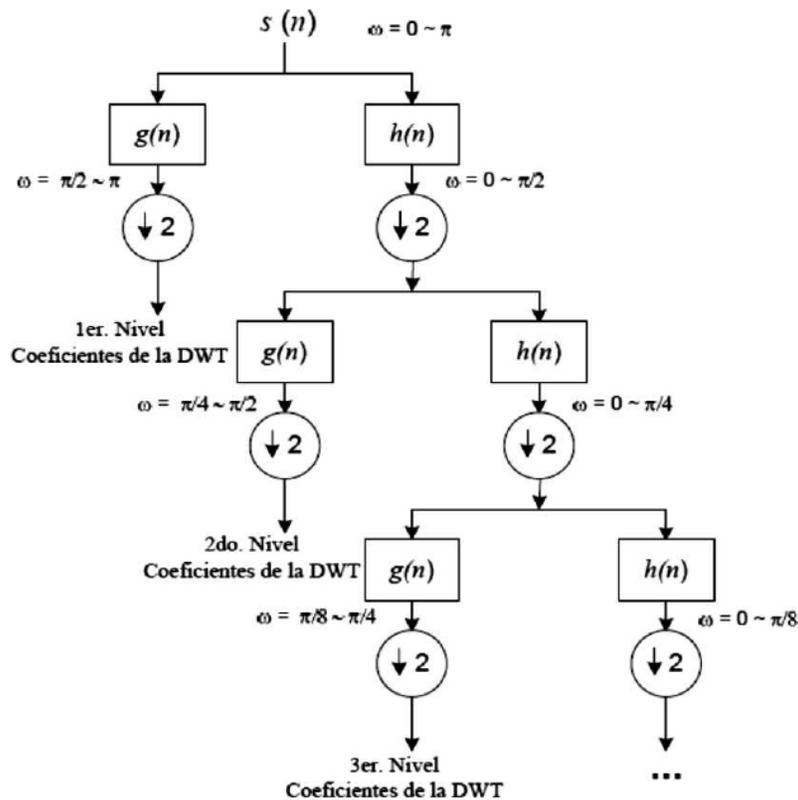
1.2.2.1 Análisis multirresolución. Se desarrolló para descomponer señales de tipo discreto, de tal modo que cada componente de frecuencia de la señal es analizada con una resolución diferente⁹, entendiéndose por resolución el número de niveles de descomposición en el dominio de las wavelet¹⁰.

⁸ MISITI. Op. cit., p.1.16.

⁹ HERNADEZ DIAZ, Marianito. Análisis Comparativo de Algoritmos para Reducción de Ruido en Señales Utilizando Wavelets. Teoría de Wavelet. Tesis profesional Licenciatura en Ingeniería en

El proceso consiste en una serie de filtrados repetitivos usando el concepto de decimado (Down sampling). Al principio se utiliza un filtro pasa baja y otro pasa alta con frecuencia de corte de la mitad de la máxima componente de frecuencia de la señal a analizar. El resultado se vuelve a filtrar bajo las mismas características, la frecuencia de corte del segundo filtro es la mitad de la máxima componente de frecuencia de la señal que va a entrar a ese filtro¹¹. La señal en la salida de cada filtro es entonces sub-muestreada por 2 (Down sampling) ya que la mitad del número de muestras son redundantes¹². Ver Figura 5.

Figura 5. Proceso de TDW, donde $h(n)$, corresponde al filtro pasa bajas y $g(n)$ al filtro pasa altas.



Fuente: Estudio de métodos para procesamiento y agrupación de señales electrocardiográficas. [Tesis doctoral]. Universidad Politécnica de Valencia; 2001.

Electrónica y comunicaciones. Cholula, Puebla, México: Universidad de las Américas Puebla. Escuela de ingeniería. Departamento de Ingeniería Electrónica, 2003. p. 29.

¹⁰ Ibid., p. 30.

¹¹ Ibid., p. 31.

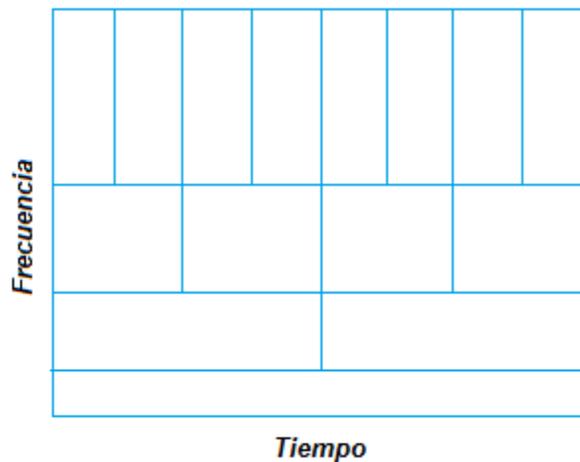
¹² POLIKAR, Robi. The Wavelet Tutorial, Citado por HERNADEZ DIAZ. Op. cit., p. 31.

De este modo la información de la localización de las componentes de frecuencia a lo largo del tiempo no se pierde. Sin embargo, la resolución de este dato depende del nivel en que se encuentre¹³.

En general, con el uso de wavelet, a altas frecuencias se tiene buena resolución en tiempo y poca resolución en frecuencia, mientras que para bajas frecuencias los resultados son buena resolución en frecuencia y poca resolución en tiempo. *Figura 6.*

Los términos producidos tras la transformación son conocidos como coeficientes de aproximación y coeficientes de detalles.

Figura 6. Plano Tiempo-frecuencia analizado con TW.



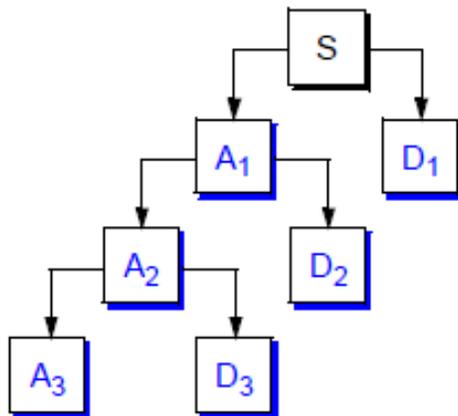
Fuente: Wavelet toolbox User's Guide.

1.2.3 Análisis Wavelet Packet. El método wavelet packet es una generalización de la descomposición wavelet que ofrece un amplio rangos de posibilidades para análisis de señales.

Se ha visto anteriormente que cuando una señal es analizada con TW, ésta se divide dentro de aproximaciones y detalles. Las aproximaciones son luego divididas dentro de aproximaciones y detalles de segundo nivel, y así sucesivamente. Ver *Figura 7.*

¹³ HERNADEZ DIAZ. Op. cit., p. 34.

Figura 7. Descomposición Wavelet de una señal.

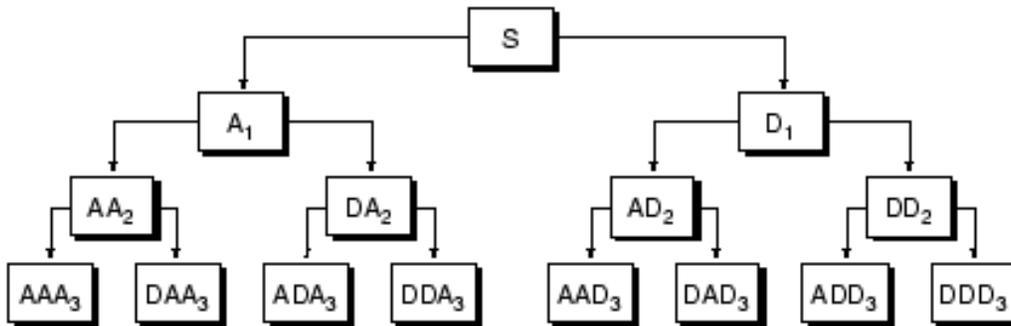


$$\begin{aligned}
 S &= A_1 + D_1 \\
 &= A_2 + D_2 + D_1 \\
 &= A_3 + D_3 + D_2 + D_1
 \end{aligned}$$

Fuente: Wavelet toolbox User's Guide.

En análisis wavelet packet tanto las aproximaciones como los detalles son subdivididos, de forma que se genera un árbol completo binario como el mostrado en la *Figura 8*.

Figura 8. Árbol de descomposición wavelet packet.



Fuente: Wavelet toolbox User's Guide.

La idea de esta descomposición es empezar desde una descomposición orientada a escala y luego se analiza las señales obtenidas en sub-bandas de frecuencia.¹⁴

¹⁴ MISITI. Op. cit., p. 6.96

Los wavelets packets son usados para obtener numerosas expansiones a partir de una señal dada. Se puede entonces seleccionar la descomposición más adecuada de la señal con respecto a un criterio de información específico. Este método de selección se conoce como el algoritmo de la mejor base¹⁵.

1.2.3.1 Algoritmo de la mejor base. Cada nodo del árbol binario formado a partir del análisis wavelet packet contiene información de la señal de entrada. El objetivo es obtener los nodos que se ajusten lo mejor posible a las características de la señal (mejor base).

Funciones de verificación con propiedades de tipo aditivo son muy adecuadas para una búsqueda eficiente de estructuras de árbol binario y divisiones fundamentales. Los criterios clásicos basados en entropía conservan estas condiciones y describen las propiedades relacionadas con la información para una representación exacta de una señal dada¹⁶.

La entropía es un concepto común en muchos campos, principalmente en el procesamiento de señales. Se puede considerar como la cantidad de información promedio que contienen los símbolos usados. Cuando todos los símbolos son igualmente probables, todos aportan información relevante y la entropía es máxima.

A continuación se presentan tres criterios de entropía:

La Entropía E debe ser una función de coste aditivo tal que,

$E(0) = 0$ y $E(s) = \sum_i E(s_i)$, donde s es la señal y (s_i) los coeficientes de s en una base ortonormal.

- Entropía de Shannon:

$$E1(s_i) = -s_i^2 \log(s_i^2), \text{ entonces} \quad (5)$$

$$E1(s) = -\sum_i -s_i^2 \log(s_i^2) \quad (6)$$

¹⁵ COIFMAN and WICKERHAUSER. "Entropy-based algorithms for best basis selection," IEEE Trans. Inform, Citado por OROZCO, Mauricio. Selección efectiva de características wavelet en la ... Trabajo de grado Magister de Ingeniería. Manizales: Universidad Nacional de Colombia. Facultad de Ingeniería y Arquitectura. Departamento de Ingeniería Eléctrica, Electrónica y Computación, 2005. 4 p.

¹⁶ MISITI. Op. cit., p. 6.105

- Concentración en l^p con $1 \leq p < 2$:

$$E2(s_i) = |s_i|^p, \text{ entonces} \quad (7)$$

$$E2(s) = \sum_i |s_i|^p = \|s\|_p^p \quad (8)$$

- Logaritmo de la energía:

$$E1(s_i) = \log(s_i^2), \text{ entonces} \quad (9)$$

$$E1(s) = \sum_i \log(s_i^2) \quad (10)$$

El algoritmo de selección de la mejor base recorre el árbol binario en sentido inverso y selecciona el conjunto de nodos de menor coste.

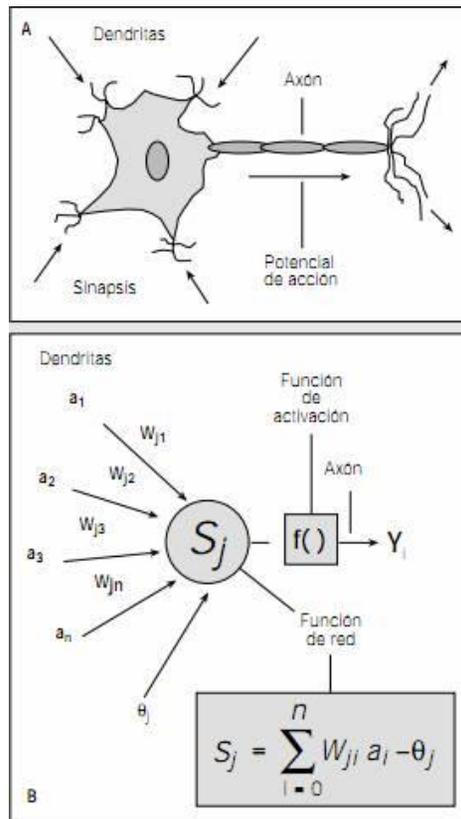
1.3 REDES NEURONALES ARTIFICIALES

Las Redes Neuronales Artificiales, RNA, están inspiradas en las redes neuronales biológicas del cerebro humano, o sea, son simuladores de sus estructuras y funcionamientos.

Están constituidas por elementos que se comportan de forma similar a la neurona biológica en sus funciones más comunes. Estos elementos están organizados de una forma parecida a la que presenta el cerebro humano¹⁷ y son conocidos como neuronas artificiales. Ver *Figura 9*.

¹⁷ BASOGAIN Olabe, Xabier. Curso: Redes neuronales artificiales y sus aplicaciones. Bilbao, España. Escuela superior de ingeniería de Bilbao. Dpto. de ingeniería de sistemas y automática. p. 2

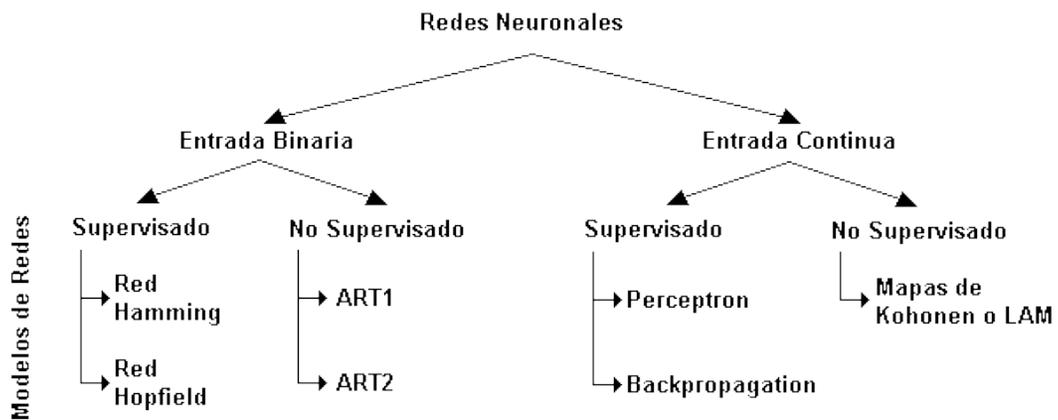
Figura 9. Comparación entre la neurona biológica (A) y neurona artificial (B).



Fuente:
http://dharma.frm.utn.edu.ar/cursos/ia/2012/material/APUNTES-FILMINAS/U4/redes_neuronales_artificiales.htm

Existen muchos tipos de redes neuronales conforme señales a estudiar. Para clasificarlas se consideran dos criterios: el tipo de entrada y el modo de aprendizaje.

Figura 10. Clasificación de redes neuronales.



Fuente: Redes neuronales artificiales

En la actualidad las redes neuronales con aprendizaje no supervisado, como los SOM, son de gran interés, pues son capaces de preservar la topología del espacio de los datos de entrada, proyectan datos altamente dimensionales a un esquema de representación de baja dimensión y tienen la habilidad de encontrar similitudes en los datos¹⁸.

MAPAS AUTORGANIZADOS. Los modelos de Mapas Auto organizativos (SOM) fueron introducidas por Teuvo Kohonen en los años 1982 y 1990, y han demostrado ser una herramienta muy poderosa en Minería de Datos (Data Mining) y en metodología de Descubrimiento de Conocimiento en Base de Datos (Knowledge Discovery Database (KDD)) con una gran variedad de aplicaciones de ingeniería tales como reconocimiento de patrones, análisis de imágenes, monitoreo de procesos y detección de fallas por nombrar algunas.

El éxito de las redes SOM se debe a su propiedad especial de crear de forma efectiva representaciones internas espacialmente organizadas de varias características de las señales de entrada y sus abstracciones¹⁹. Las SOM implementan un mapa ordenado de dimensionalidad reducida de los datos que respeta la función de densidad de probabilidad que subyace en el comportamiento de éstos.

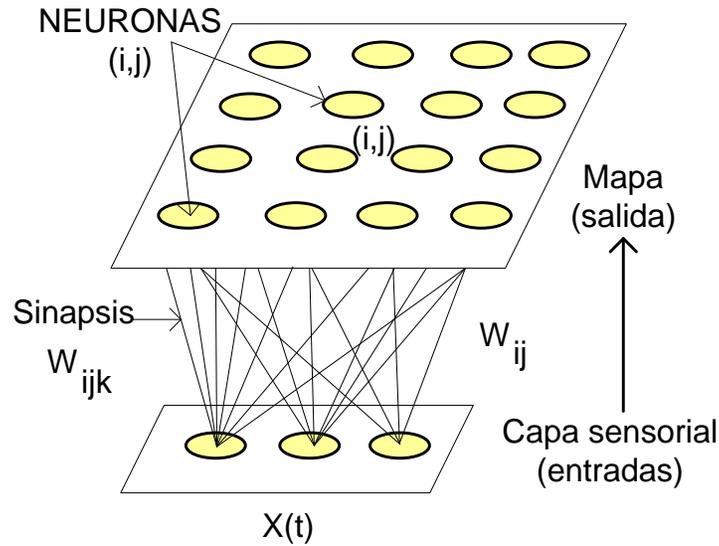
A diferencia de las redes neuronales supervisadas, La SOM no posee una capa oculta, ni utiliza una función de activación o un valor umbral. Su modelo, creado por Kohonen (ver *Figura 11*), sólo contiene una capa de entrada y un mapa de salida con una distribución especial de las neuronas. Esta distribución se conoce como topología, puede ser unidimensional, bidimensional o de orden superior; en malla, hexagonal o aleatoria.

Por no poseer una capa oculta y tener un algoritmo de aprendizaje no supervisado estas redes son relativamente fáciles de construir y pueden ser entrenadas muy rápidamente haciendo que sean muy utilizadas.

¹⁸ SALAS, Rodrigo. "Mapas Autorganizados de kohonen (SOM)". Internet: (http://www.inf.utfsm.cl/~rsalas/Pagina_Investigacion/docs/Apuntes/Redes%20SOM.pdf)

¹⁹ KOHONEN, Teuvo. The Self-Organizing Map. **En:** PROCEEDINGS OF IEEE. Septiembre, 1990. vol. 78, no. 9, p. 1464.

Figura 11. Modelo Kohonen.



Fuente: Introduction to Self Organizing Map [video].

Los procesos esenciales presentes en la formación de una SOM son:

a) **Proceso competitivo.** Considerando un vector de entrada

$$\vec{x} = [x_1, x_2, \dots, x_n]^T \in \mathbb{R}^n, \quad (11)$$

conectado en paralelo a todas las neuronas j de una red con vector de pesos

$$\vec{w}_j = [w_{j1}, w_{j2}, \dots, w_{jn}]^T \in \mathbb{R}^n; \quad (12)$$

El algoritmo competitivo consiste en una secuencia de pasos iterativos que halla un índice de relación entre el vector de entrada y los pesos de la red, para conocer cuál es la unidad que mejor se ajusta (Best match unit ó neurona ganadora) a la señal.

El criterio de relación más conveniente es basado en la distancia Euclidiana entre x y w_j , donde la distancia mínima define la “ganadora” w_c ²⁰.

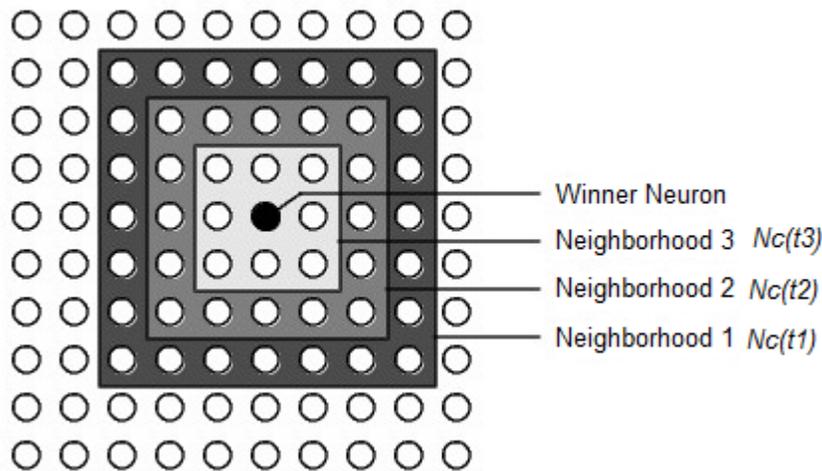
$$\|\vec{x} - \vec{w}_c\| = \min_j \{\|\vec{x} - \vec{w}_j\|\} \quad (13)$$

²⁰ Ibid., p. 1467.

b) Cooperación y Adaptación sináptica. Una vez las neuronas compiten para representar la señal de entrada es necesario introducir un concepto de interacción lateral con otras neuronas conocido como vecindario N_c . Este vecindario está centrado alrededor de la neurona ganadora c . En cada paso de aprendizaje todas las neuronas pertenecientes a N_c son actualizadas, mientras que las neuronas fuera se dejan intactas.

El radio de N_c puede ser variante en el tiempo; de hecho para un buen ordenamiento global se ha comprobado que el vecindario debe empezar con valor amplio y reducirse monótonicamente con el tiempo. *Figura 12.*

Figura 12. Ejemplo de vecindario, donde $t1 < t2 < t3$.



Fuente: Neural Network Toolbox User's Guide. Self-Organizing and Learning Vector Quantization Nets.

La unidad ganadora y sus vecinas modifican sus vectores de referencia hacia la entrada actual mediante la siguiente formulación:

$$\vec{w}_j(n+1) = \vec{w}_j(n) + \eta(n)h_{j,i}(n)[\vec{x} - \vec{w}_j] \quad (14)$$

Donde $h_{j,i}(n) = h_{j,i}(\vec{x})(n)$ es la función de vecindad "Kernel" descrita como una función Gaussiana de la forma:

$$h_{j,i}(\vec{x})(n) = h_0 \exp\left(\frac{-\|\vec{r}_j - \vec{r}_i\|}{2\sigma^2}\right) \quad (15)$$

Con:

\vec{r}_j : Vector de posición de la neurona excitada (j)

\vec{r}_i : Vector de posición de la neurona ganadora (i)

$\sigma = \sigma(n) = \sigma_0 \exp\left(\frac{n}{\tau_1}\right)$, n : número de iteraciones y τ_1 : constante de tiempo.

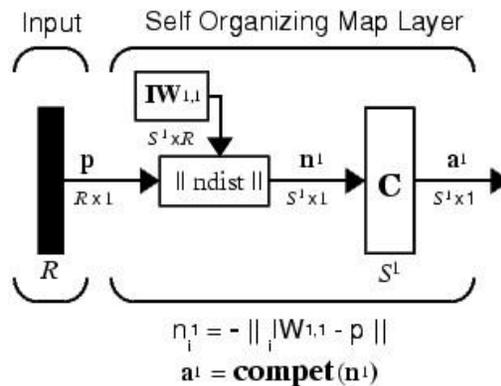
Y $\eta(n)$ es el parámetro de la tasa de aprendizaje, también decreciente en el tiempo:

$$\eta(n) = \eta_0 \exp\left(\frac{n}{\tau_2}\right) \quad (16)$$

En conclusión, en esta parte del proceso se busca aumentar la aproximación de los pesos asociados de la unidad ganadora y sus vecinas al vector de entrada (Cooperación), y disminuir gradualmente los dos parámetros que intervienen en dicho proceso, el parámetro de aprendizaje y el tamaño del vecindario (Adaptación).

c) Arquitectura de la red. La arquitectura de una red SOM es mostrada en la *Figura 13*.

Figura 13. Arquitectura de una SOM.



Fuente: Neural Network Toolbox User's Guide.

La función de transferencia competitiva C produce un 1 para el elemento de salida a_j^1 correspondiente a la neurona ganadora j . Los otros elementos en a^1 son cero²¹.

d) Entrenamiento. El entrenamiento más usual para este tipo de redes es el *trainbu* (Batch unsupervised training por sus siglas en inglés). En este tipo de entrenamiento los datos son presentados como un bloque y los nuevos vectores de peso son promedios ponderados de los vectores de datos.

Este entrenamiento posee dos fases: Fase de ordenamiento y Fase de sintonización, en las que el tamaño de la vecindad y la tasa de aprendizaje se alteran. Ver *Cuadro 1*.

Cuadro 1. Parámetros de aprendizaje por defecto

Taza de aprendizaje fase de ordenamiento	0.9
Iteraciones fase de ordenamiento	1000
Tamaño del vecindario inicial	3
Taza de aprendizaje fase de sintonización	0.02
Tamaño de vecindario fase de sintonización	1

Durante la fase de ordenamiento el algoritmo ajusta la distancia del vecindario (DV) desde el tamaño inicial, que por defecto es 3, a 1. Es durante esta fase que los pesos neuronales se ordenan en el espacio de entrada en consonancia con las posiciones de las neuronas asociadas²².

En la fase de sintonización, DV siempre se establece en 1, y se espera que los pesos se extiendan de forma relativamente uniforme sobre el espacio de entrada, manteniendo su orden topológico encontrado durante la fase de ordenamiento²³.

Las neuronas de una SOM se ordenarán con distancia aproximadamente igual entre ellas si los vectores de entrada aparecen con probabilidad uniforme a lo largo de la sección del espacio de entrada. Si los vectores de entrada se producen con frecuencia variable en todo el espacio de entrada, el mapa tiende a asignar las neuronas a un área en proporción a la frecuencia de vectores de entrada allí.

²¹ HUDSON BEALE, Mark; HAGAN, Martin and DEMUTH, Howard. Neural Network Toolbox User's Guide. Self-Organizing and Learning Vector Quantization Nets. [s.l]: The MathWorks, Inc., 1992-2012. p. 6.19.

²² *Ibid.*, p. 6.22.

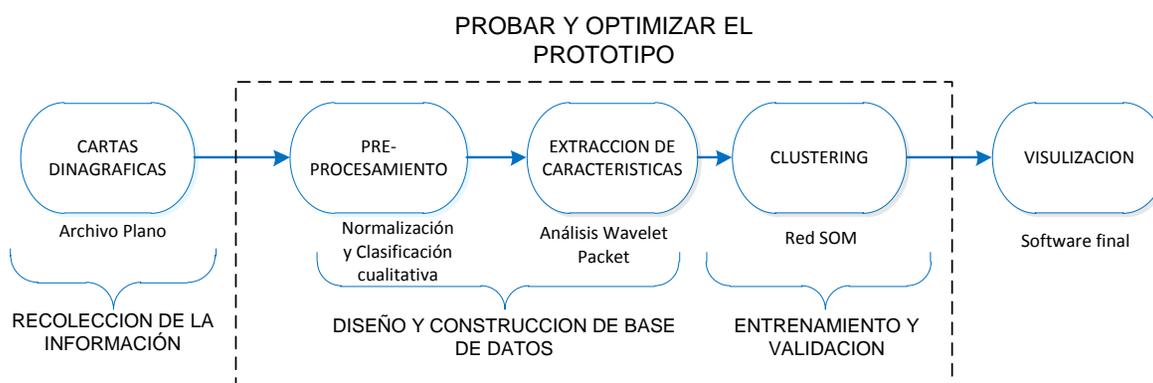
²³ *Ibid.*, p. 6.23.

2. METODOLOGÍA

De acuerdo a la información adquirida sobre la interpretación de dinagramas y con base en el comportamiento descrito en estas, el procedimiento aplicado para realizar la identificación de problemas en el equipo de subsuelo se explica en detalle en el transcurso de este capítulo donde se describe los estudios y pruebas preliminares, la metodología desarrollada y los parámetros más importantes de esta aplicación.

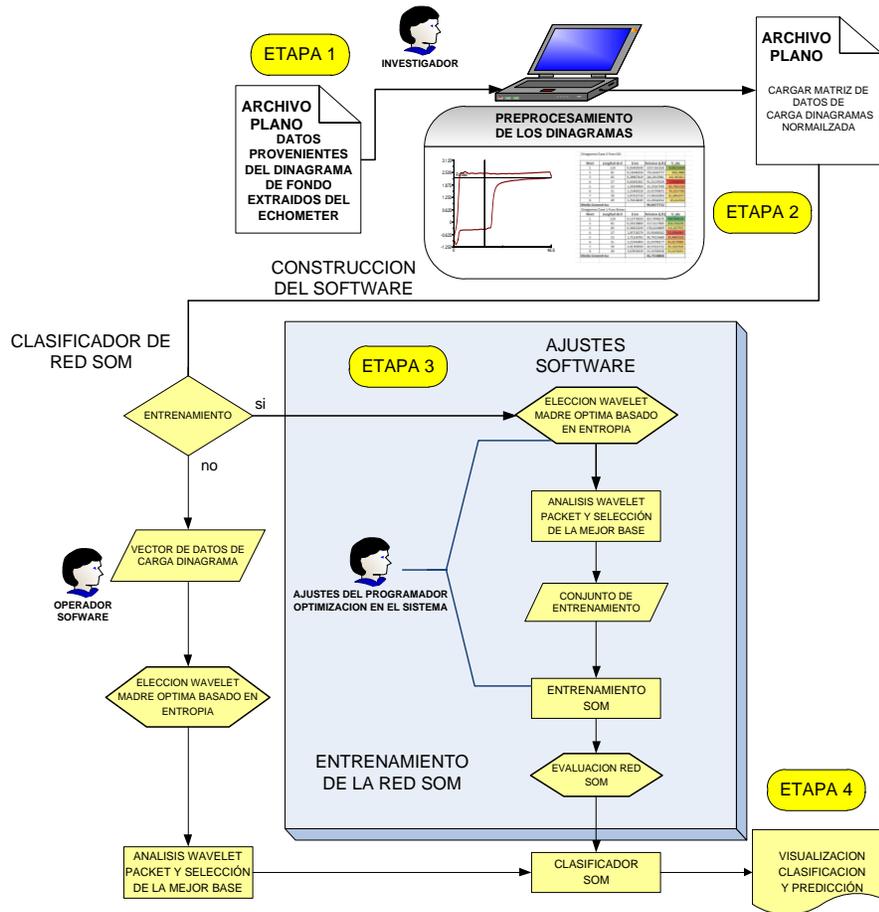
El diagrama de bloques general que se observa en la *Figura 14* inspira la producción del algoritmo aplicado al software, donde se describen cuatro etapas principales, presentes en cualquier proyecto Neurocomputacional: Recolección de la información, Diseño y construcción de la base de datos, Entrenamiento y validación de la red neuronal y visualización o Software final. Además a lo largo de las etapas centrales también se prueba y optimiza el prototipo diseñado.

Figura 14. Diagrama de bloques general del proyecto



En esta sección se dan a conocer los principales procedimientos utilizados en cada una de las etapas de la metodología para desarrollar este proyecto, el siguiente diagrama, *Figura 15*, ilustra claramente y de forma detallada cada una de las etapas descritas en el diagrama de bloques anterior. Basado en este esquema se procedió a realizar el desarrollo del trabajo.

Figura 15. Algoritmo de Supervisión



En la etapa 1 los registros de los diferentes dinagramas provenientes del ECHOMETER se han guardado en un archivo plano pasando a la etapa 2 donde se clasificaron de forma manual para saber qué clase de información se tenía del proyecto. Luego se analizaron a través de procesamiento digital de señales (DSP por sus siglas en inglés) para luego seguir con la etapa 3 en la cual se procedió con la construcción del software, donde a la base de datos de los dinagramas ya pre procesados se le aplicó la transformada wavelet, debido a que estas señales son de naturaleza no estacionarias siendo esta una poderosa herramienta para este trabajo, y finalmente en la etapa 4 se construye la red SOM y la visualización de los resultados por medio de un software adaptado a las necesidades de la industria.

A continuación se aborda el desarrollo de cada una de las etapas de manera más profunda.

2.1 IDENTIFICAR SI EL PROBLEMA ES ADECUADO PARA RESOLVERLO POR MEDIO DE REDES NEURONALES

En el caso de diagnóstico de problemas en el equipo de subsuelo de unidades de bombeo mecánico las técnicas tradicionales hablan concretamente de la interpretación y análisis de dinagramas. Esta tarea se basó en la identificación de formas o comportamientos que han sido preestablecidos para determinar los problemas y dependiendo de si estas formas patrón estaban implícitas en el dinagrama de fondo o no, el equipo de subsuelo presentaba el problema relacionado con dicha forma patrón o se encontraba funcionando normalmente.

De esta forma se describe el problema en una tarea de reconocimiento de patrones y clasificación de dinagramas con base en un conjunto de datos, ejercicio que se desarrolló por medio de la red SOM como herramienta.

El proceso de entrenamiento y ajuste de la red SOM fue posible cuando se realizó el análisis de los patrones de entrenamiento, para asegurar un reconocimiento e identificación de problemas con un error mínimo. Por lo tanto el esfuerzo más grande se debió concentrar en construir una base de datos de entrenamiento con ejemplos óptimos en cuanto a definición de los problemas y extraer cada uno de los problemas más relevantes que caracterizaban cada situación. Además es fundamental en este tipo de proyectos la utilización de una topología adecuada de las SOM y una cantidad de neuronas mayor a los patrones detectados, que en el caso del proyecto son las clases, para que no existiera ambigüedad y que permita cumplir los objetivos propuestos (reconocimiento, clasificación y predicción de los datos del dinagrama de entrada a la red).

2.2 RECOLECCIÓN DE LA INFORMACIÓN

Este aspecto fue uno de los más importantes porque la información puede ser requerida en cada fase del desarrollo del proyecto.

La información la proporcionó la empresa Ecopetrol con base en pruebas tomadas a algunos pozos cercanos a la ciudad de Neiva, estos pozos se encuentran ubicados en los campos Brisas, Cebú y Santa clara.

Los datos del dinagrama de fondo del pozo que fueron analizados son coordenadas tipo (x,y) , que corresponden a las coordenadas de posición del pistón y carga soportada por la primera varilla después de la bomba, respectivamente. Los datos fueron extraídos del software TWM (Total Well

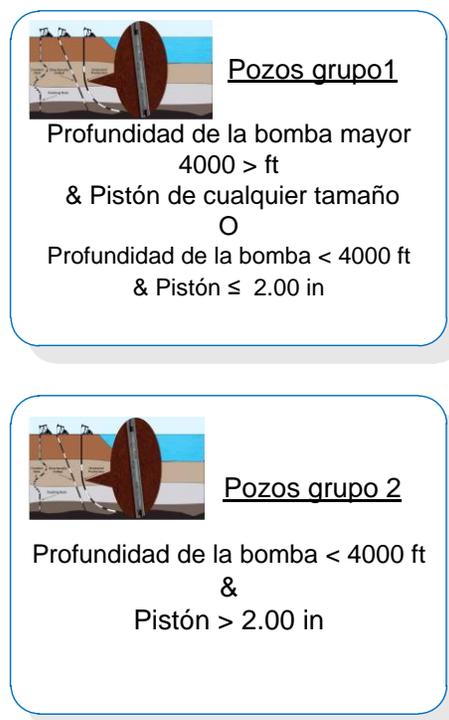
Management) Echometer y guardados separadamente en archivos planos (tipo XLS) para su posterior pre-procesamiento individual y pre-clasificación.

2.3 DISEÑO Y CONSTRUCCIÓN DE LA BASE DE DATOS

En esta etapa se estableció y organizó la información necesaria sobre cada dinagrama. De esta forma se realizó un pre-procesamiento, que consiste en una clasificación cualitativa de los pozos y una normalización, seguido de la codificación de los datos para el entrenamiento por medio de la TW.

2.3.1 Clasificación cualitativa. Ya que no todos los pozos son iguales, se requirió de diferentes herramientas para diagnosticar problemas dependiendo de las condiciones del pozo. Se dividió entonces los pozos con bombeo mecánico en dos grupos. El primer grupo recibió el nombre de “grupo 1”, incluye pozos mayores de 4000 ft de profundidad con cualquier tamaño de pistón, y pozos con menos de 4000 ft y pistones de 2 in. o menos. El segundo grupo, al cual se llamó “grupo 2”, incluye pozos con menos de 4000 ft y con pistón mayores de 2.00 in Ver *Figura 16*. Estos dos grupos de pozos poseen características únicas que se debe conocer para diagnosticar problemas con precisión.

Figura 16. Tipos de Pozos en bombeo mecánico



Fuente: Bombeo mecánico Optimización.

La razón para separar los pozos someros de alta producción del grupo 2 de los pozos profundos (grupo 1) es porque los mismos son afectados por las fuerzas de inercia de los fluidos, que con frecuencia duplican la carga sobre el pistón. Debido a que los pistones en estos pozos son grandes, éstos deben recoger el fluido y acelerarlo a la tubería en tazas mucho mayores que en pozos más profundos. Además la sarta de cabillas está rígida y no provee ningún nivel de amortiguación.

En pozos del grupo 1, para reducir la carga de la sarta de cabillas se debe usar un pistón de poco diámetro. La sarta de cabillas actúa como un amortiguador estirándose al aumentar la carga en el pistón; esto absorbe efectivamente el “golpe” de recibir la carga de fluido en la carrera ascendente y no aparecen fuerzas de dinámica de fluido en la carta dinagráfica de fondo.

Por lo tanto, sólo en pozos del Grupo 1 la condición de la bomba se puede identificar por la comparación con formas conocidas de cartas dinagráficas.

Una vez se organizaron los dinagramas en los dos tipos de pozos, se clasificaron manualmente teniendo en cuenta la forma aparente del dinagrama. Esto se hizo con el fin de reconocer a priori las fallas o patrones dominantes que podrían llegar a poseer los pozos.

2.3.2 Normalización. Como se conoce, un dinagrama está representado por un par de vectores columnas (x e y). Pero al tener datos reales se descubre que cada dinagrama posee diferentes longitudes y diferentes tasas de muestreo, por lo que es indispensable su normalización. Este proceso consiste en agrupar los valores en el rango entre cero y uno en cada eje coordenado, lo que permite estandarizar los datos para que trabajen a una misma escala y así se puedan correlacionar más fácilmente para un mejor reconocimiento en la red neuronal.

Las ecuaciones empleadas para este fin son las siguientes:

$$X_{nor} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad Y_{nor} = \frac{Y - Y_{min}}{Y_{max} - Y_{min}} \quad (17)$$

Una vez se normalizaron los datos de cada dinagrama se igualaron sus dimensiones para finalmente compilar la información en una gran matriz $M_{180 \times 1038}$. Esta matriz se obtiene de agrupar un total de 519 dinagramas, donde cada dinagrama contiene 180 datos que son obtenidos dependiendo del tiempo de muestreo de la señal tomada físicamente en el pozo.

Para la extracción de características por medio de TW se seleccionó sólo la información de carga de cada dinagrama quedando reducida la matriz a un

tamaño de 180X519. Sus respectivas gráficas en 2D y 3D se muestran en la *Figura 17* y *Figura 18*.

Figura 17. Información de Carga de Base de Datos.

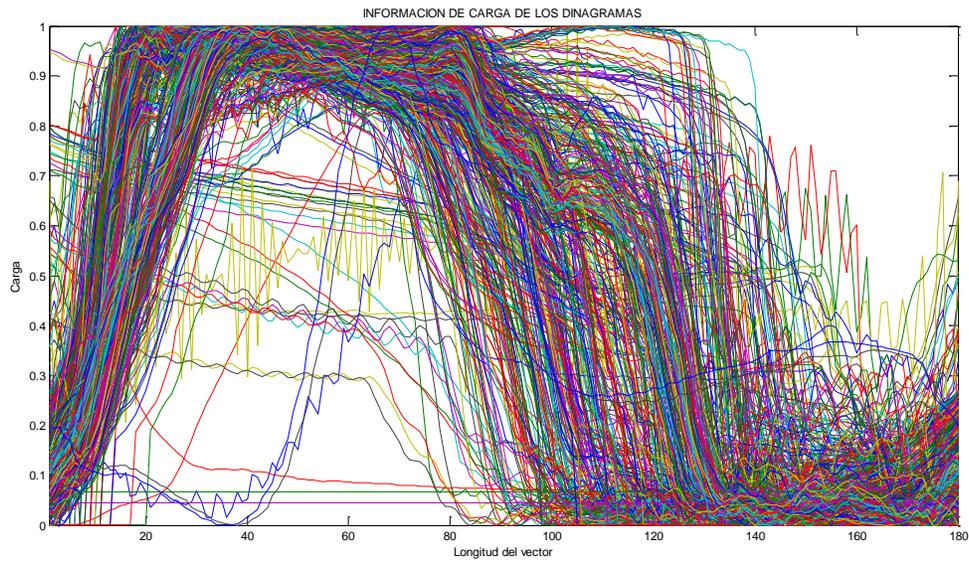
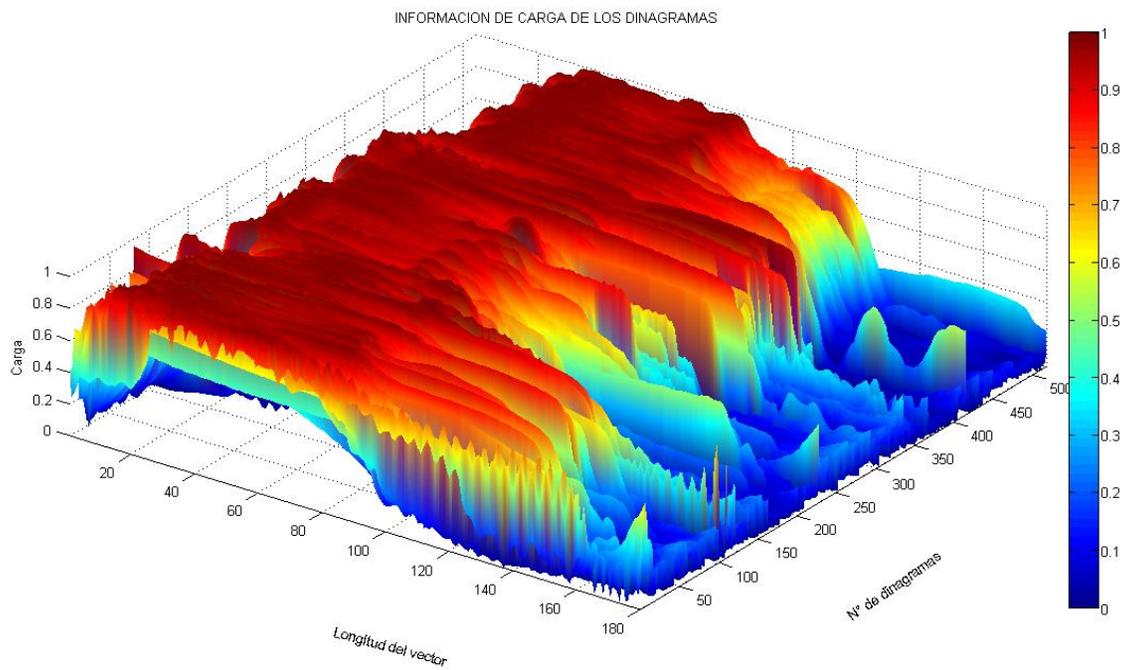


Figura 18. Grafica 3D Información de carga de la Base de Datos.



2.3.3 Extracción de características. Para efectuar un buen estudio de los Dinagramas se consultaron varios métodos de análisis y selección de patrones para que cada registro pudiera ser reconocido por la SOM.

Se optó por hacer un análisis mediante wavelet packet debido a la naturaleza no estacionaria de las señales, se seleccionó primero la wavelet madre que mejor se adapte a cada diagrama, luego un nivel óptimo de descomposición y posteriormente la mejor base (Best Basis), que contiene los nodos con la información más predominante de la señal.

2.3.3.1 Elección Wavelet madre. A partir de un conjunto de 17 wavelet básicas que comprenden las familias Daubechies, Biorthogonal, Reverse Biorthogonal y Coiflets con sus respectivas órdenes, se realizó la selección de la wavelet madre en base a un criterio de información.

Se calculó la función de costo de información basada en la Entropía de Shannon para cada señal de dinagrama con cada wavelet base del conjunto preestablecido, de tal manera, que la selección final de la wavelet madre recayó sobre la función que tuviera el menor valor.

De lo anterior se encontró que para el 98% del total de las señales, la wavelet madre seleccionada fue la Daubechies de orden 25 (db25), debido a que es la wavelet que más momentos de desvanecimiento posee, o sea, con mayor capacidad para representar el comportamiento polinomial de la señal.

2.3.3.2 Nivel óptimo de descomposición. Después de que se escogió la wavelet madre se tuvo en cuenta un número adecuado de niveles de descomposición basados en la naturaleza de la señal y se buscó una reducción significativa en la longitud del vector de coeficientes que caracterizan el dinagrama sin sacrificar la información que éste contenía.

Se encontró que a medida que se adiciona niveles el valor de Entropía en los nodos disminuye, entonces la elección del nivel óptimo por medio de un algoritmo con criterio de Entropía que daría un nivel de descomposición muy elevado y perdería todo sentido la caracterización de la señal.

De otra forma, una selección de nivel basado en el error de reconstrucción de la señal a partir de los coeficientes de aproximación de cada nivel tampoco tiene mayor sentido, pues el menor error se ubica en la aproximación más cercana de la señal, que sería en el primer nivel.

En este sentido, la idea fundamental consistió en relacionar la longitud del vector de los coeficientes del análisis wavelet packet y el error de aproximación a la señal original de manera que el nivel óptimo de descomposición conserve la relación más equitativa. Véase *Cuadro 2*.

Cuadro 2. Análisis de mejor nivel.

Nivel	Longitud de cA	Error	Relación (L/E)	V_abs
1	114	0,0350	3257,5622	3158,5144
2	81	0,1104	733,4144	634,3666
3	65	0,2697	241,0316	141,9838
4	57	0,6249	91,2113	7,8365
5	53	1,5939	33,2517	65,7961
6	51	2,2147	23,0280	76,0198
7	50	2,8733	17,4014	81,6463
8	49	3,7022	13,2354	85,8123
Media Geométrica			99,0478	

En el anterior cuadro, para cada uno de los ocho niveles, se hizo la lectura de la longitud de sus coeficientes de aproximación (cA), se realizó la Transformada Inversa Wavelet para llegar al error de reconstrucción de la señal y se halló el cociente entre la longitud y el error. Luego se buscó el valor más equitativo de estas relaciones calculando su cercanía a la media geométrica, pues ésta es menos sensible que la media aritmética a los valores extremos. La escala cromática de verde a rojo, mostrada en la última columna del cuadro, donde el verde es el valor más lejano y el rojo el más cercano a la media, muestra que el nivel que mejor conserva la proporción de la longitud del vector y contenido de información del mismo es el cuarto.

Este análisis se hizo para 60 dinagramas en general con resultados mostrados en el *Cuadro 3*. En el 80% de los casos el nivel más óptimo fue el 4 y en el 20% restante fue el nivel 5.

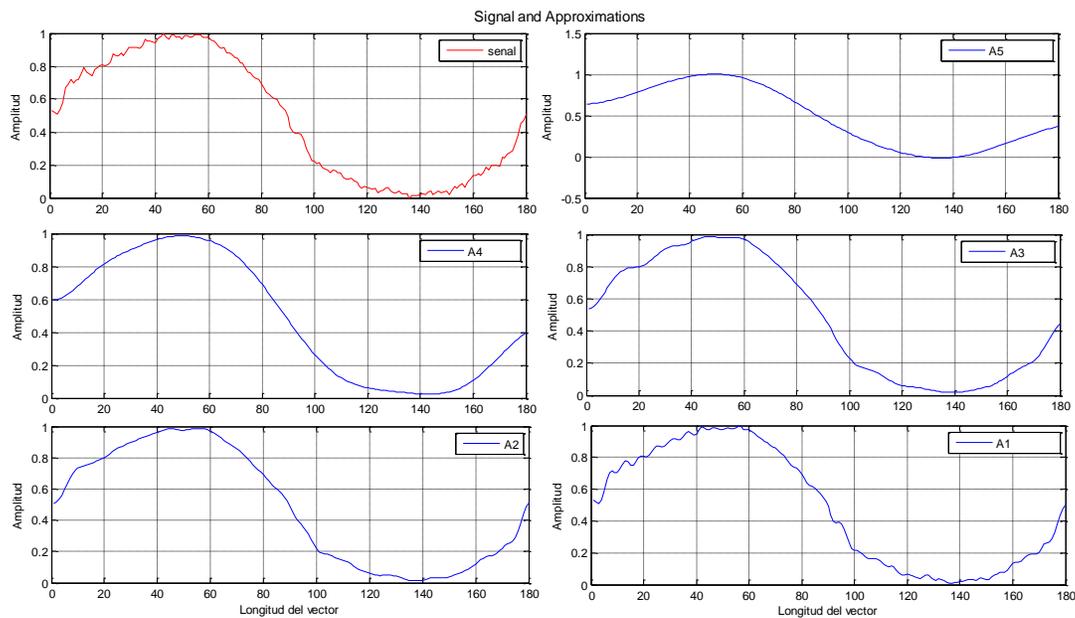
Cuadro 3. Resultados de estudio de mejor nivel.

	Cantidad de dinagramas	Porcentaje
Nivel 4	48	80%
Nivel 5	12	20%
Total	60	100%

Se corroboró, como se muestra en las siguientes dos figuras, que, para este caso de estudio, un análisis wavelet packet con cuatro (4) niveles de descomposición proporcione los mejores aspectos cualitativos (aproximaciones) y cuantitativos (detalles) de las señales.

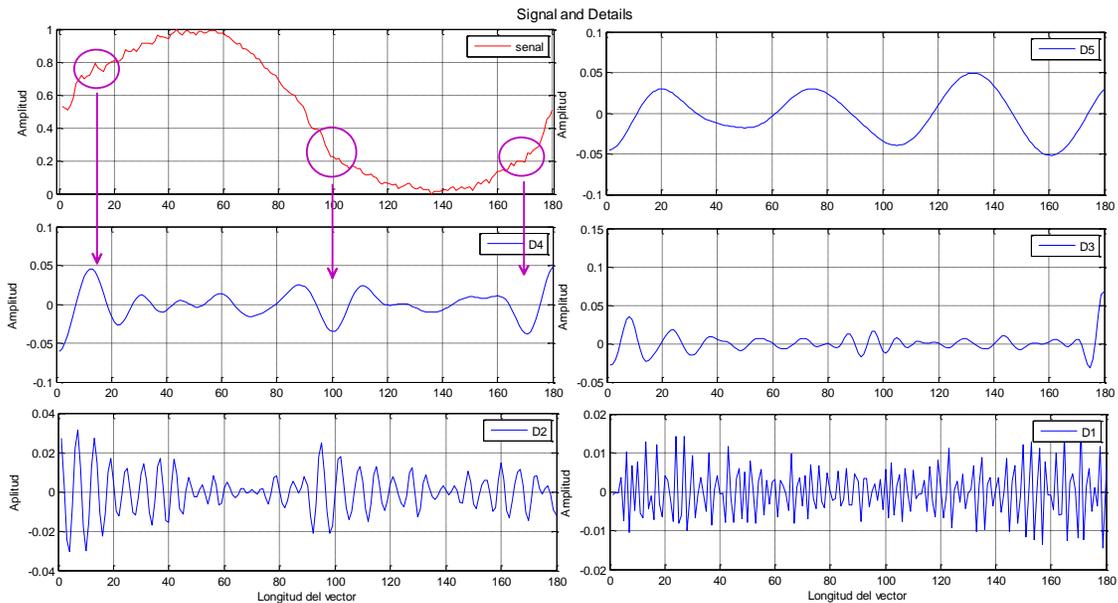
En la *Figura 19*, A1, A2 y A3, formas correspondientes a los coeficientes de aproximación del nivel 1, 2 y 3 respectivamente, presentaron una similitud muy importante de la señal original, sin embargo se ve que A4 logró una muy buena forma sin perder las pendientes principales. A5 perdió la descripción fundamental de la señal.

Figura 19. Señal y Aproximaciones



Por otro lado, en la *Figura 20*, los coeficientes de detalles de los primeros dos niveles, D1 y D2, poseen un amplitud muy pequeña y denotan un ruido de alta frecuencia a lo largo de todo el sensado. D3 es un poco más específica en la localización de los detalles, cambios de pendientes. Pero D4 presentó los valores mínimos en los cambios de pendientes más importantes de la señal que están encerrados en círculos, entre 10 y 30, entre 90 y 110, entre 160 y 180. D5 perdió casi completamente esta información.

Figura 20. Señal y Detalles



2.3.3.3 Mejor base. El árbol binario de cuatro niveles que resultó del análisis wavelet packet permite 2^4 diferentes caminos para codificar la señal. Escoger la descomposición más adecuada significa analizar cada nodo del árbol y cuantificar la información que se puede obtener de la realización de cada división mediante el algoritmo discutido en la... sección 1.2.3.1... Finalmente, se obtuvo un nuevo vector de coeficientes de aproximaciones y detalles de 114 elementos por cada dinagrama que se proporcionó a la red SOM para su posterior clasificación.

2.4 ENTRENAMIENTO Y EVALUACIÓN DE LA RED

Para asegurar un éxito en el entrenamiento, la TW se encargó de entregar los datos de cada señal, con el fin de que la información se traduzca en ideas claras que equivaldrán a los patrones o reglas a ser aprendidas por la red SOM y que luego se convirtieran en salidas concretas de cada una de las neuronas.

La SOM por ser una red neuronal de clasificación no supervisada, también conocida como técnica de clustering, no necesita conocer de antemano la pertenencia de cada dato de entrenamiento a un grupo preestablecido para realizar los agrupamientos. Sin embargo, como se disponía de esa información, se utilizó posteriormente para evaluar la calidad de los clusters y para “etiquetarlos” (poner nombre).

La clasificación manual de fallas en base a los dinagramas teóricos encontrados en el Manual de Bombeo Mecánico dio un conocimiento a priori de los siguientes 25 patrones presentes en los datos suministrados por Ecopetrol:

- 1) Golpe de fluido combinado y retraso del cierre de la válvula de viajera
- 2) Completamente sin bombeo - bomba anclada.
- 3) Desgaste moderado de la bomba.
- 4) Golpe de fluido leve - tubos sin anclar.
- 5) Bombeo completo - fondo tubería sin anclar.
- 6) Bombeo completo – golpe en la parte inferior del émbolo.
- 7) Desconocida
- 8) Interferencia Gas – alta separación en el émbolo.
- 9) Golpe de fluido grave - tubo anclado.
- 10) Desconocido
- 11) Ruido
- 12) Ruido
- 13) Fugas en la válvula de viajera o émbolo.
- 14) Golpe de fluidos - alta fricción en la carrera ascendente
- 15) Desconocido
- 16) Desconocido
- 17) Desconocido
- 18) Bombeo completo - émbolo golpeando hacia arriba. (Espaciado de la bomba demasiado alto)
- 19) Desconocido
- 20) Grave fuga en la válvula viajera.
- 21) Bombeo full - cabillas de fibra de vidrio.
- 22) Bomba desgastada.
- 23) Bomba severamente desgastada, varillas partidas o que no fluyen bien.
- 24) Golpe ligero de fluido - tubo anclado.
- 25) Golpe severo de fluido - tubo sin anclar.

Luego de razonar sobre estos patrones de entrenamiento se observó que algunos representan un mismo problema de manera general y por lo tanto se formó una base de datos bastante extensa de todos los dinagramas obtenidos.

Además de las utilidades de esta pre-clasificación para evaluar la calidad de los agrupamientos resultantes de la red SOM y su posterior etiquetado como ya se ha dicho, con este planteamiento se estableció que el tamaño del mapa de la salida de la red SOM debía contener un total de neuronas mayor al total de estas clases para disminuir la ambigüedad a la hora de hacer la identificación de las fallas.

El siguiente paso luego de que se obtuvieron los dinagramas en el formato deseado para la entrada de la red SOM, fue realizar el entrenamiento. En este punto se le proporciono además de los datos, las características de la red a entrenar como son: la topología (posiciones físicas de las neuronas), función de distancia, útil para calcular la distancia entre las neuronas, distancia inicial de vecindario y tamaño del mapa de salida.

La topología de la red puede ser rectangular, hexagonal o aleatoria, y las funciones de distancia existentes son: distancia euclidiana, distancia cuadrada (boxdist), distancia de pasos (linkdist) y distancia manhattan.

Aunque la evaluación o validación sobre esta red no tendría ningún impacto en la formación de la función de entrenamiento por ser no supervisada, si puede servir como medidas independientes de generalización de la red

La evaluación se realizó teniendo en cuenta la respuesta a los patrones de prueba constituidos por dinagramas diferentes a los del entrenamiento, si la evaluación de la red arroja resultados buenos en cuanto a el reconocimiento de uno y dos problemas, entonces se termina el proceso de evaluación de la red, si es el caso contrario es necesario modificar la estructura de la red y volver a realizar el proceso.

2.5 PRUEBA Y OPTIMIZACIÓN DEL PROTOTIPO

En esta fase se determinó si se debía modificar la topología de la red SOM, si se requerían más datos de entrenamiento, si era necesario cambiar el modelo original incluyendo el algoritmo de pre-procesamiento y el algoritmo de entrenamiento (se pone a prueba el diseño y se mejora en caso de ser necesario).

El prototipo se convirtió en el modelo óptimo cuando el sistema fue confiable, es decir que produzca las soluciones requeridas en la fase de prueba. Para cubrir este aspecto se debió ejecutar diferentes pruebas entre las cuales se evaluaron las entradas de la red, el tamaño de los mapas de salida (número de neuronas) y las diferentes topologías en cada caso.

Cuantificar la eficiencia de una SOM no es una tarea fácil, puesto que no hay un objetivo claramente especificado en el entrenamiento al cual se tenga que adaptar la red para así poder determinar un error. Sin embargo, en este trabajo se

inferieron los siguientes dos métodos que proporcionan alguna medida del comportamiento de la SOM:

a) Distancia del vector de entrada y su neurona ganadora. Con este método se intentó medir que tanto se ajusta la red a la topología del conjunto de entrada, calculando el promedio de las distancias entre los vectores de entrada (x_i) y sus correspondientes neuronas ganadoras (j). Siendo N el número de vectores o muestras de entrada, y m_j el vector de pesos de la neurona ganadora para el vector de entrada x_i , este promedio se define como:

$$Promedio = \frac{1}{N} \sum_{i=1}^N \|\vec{x}_i - \vec{m}_j\| \quad (18)$$

Luego se encontró en la literatura que este promedio es llamado error de cuantización²⁴.

b) Error Topográfico de Kiviluoto. Aparte de estudiar la precisión con que las neuronas se adaptan a la topología de entrada, se hace necesario evaluar que tanto conservan las redes sus diferentes topologías iniciales basados en el error topográfico propuesto por Kiviluoto²⁵:

$$e_t = \frac{1}{N} \sum_{i=1}^N u(x_i) \quad (19)$$

Donde,

N : Número de muestras.

x_i : La muestra i -ésima del conjunto de datos.

$u(x_i)$: Valor en función de la posición relativa entre las dos principales neuronas ganadoras de las muestras x_i en el espacio de salida.

$$u(x_i) = \begin{cases} 1 & \text{Si las dos principales neuronas ganadoras no son adyacentes.} \\ 0 & \text{Si son adyacentes.} \end{cases}$$

²⁴ E. Arsuaga Uriarte and F. Díaz Martín. Topology preservation in SOM. En: International Journal of Mathematical and Computer Sciences. 2005, p. 19.

²⁵ Curso de Sistemas inteligentes de control. [En línea]. Disponible en: <http://isa.uniovi.es/doctorado/control_avanzado/5-Curso%20doctorado.pdf>

De este modo el menor error significa que conserva su topología original.

El *cuadro 4* muestra las pruebas hechas, en base a los anteriores dos métodos, para las diferentes topologías de red neuronal, diferentes funciones de distancia y variación en la cantidad y distribución de neuronas en la salida.

Cuadro 4. Pruebas de eficiencia de la SOM

		Promedio de Distancia			Error Topográfico		
		Topología			Topología		
Función de distancia	Distribución Mapa	Cuadrada	Hexagonal	Aleatoria	Cuadrada	Hexagonal	Aleatoria
mandist	10X6	1,1350	1,1201	1,1223	0,3507	0,4855	0,4547
	8X5	1,3772	1,3718	1,3358	0,4123	0,4027	0,3642
	6X5	1,4833	1,5574	1,4744	0,3468	0,5491	0,5934
	7x4	1,5516	1,4977	1,4891	0,4027	0,3218	0,4547
	5x5	1,5570	1,5619	1,5601	0,2563	0,4143	0,5395
	5x4	1,7409	1,6878	1,6926	0,2678	0,2216	0,4971
linkdist	10X6	1,1329	1,1163	1,1521	0,4008	0,5279	0,4971
	8X5	1,3914	1,3234	1,3389	0,5202	0,5029	0,5067
	6X5	1,5631	1,5241	1,4666	0,4509	0,4412	0,5183
	7x4	1,5582	1,5470	1,5180	0,3680	0,3603	0,4451
	5x5	1,5532	1,6128	1,6315	0,2331	0,4759	0,3487
	5x4	1,7404	1,7227	1,7061	0,2871	0,4123	0,2948
boxdist	10X6	1,1360	1,1427	1,0998	0,5511	0,5260	0,4759
	8X5	1,3998	1,3806	1,3066	0,4393	0,5279	0,5376
	6X5	1,4765	1,5157	1,4660	0,3410	0,4778	0,5973
	7x4	1,5670	1,4962	1,5161	0,3931	0,5106	0,5164
	5x5	1,6073	1,6300	1,6069	0,4374	0,5723	0,5568
	5x4	1,7205	1,7244	1,7360	0,3083	0,3391	0,5337
dist	10X6	1,1309	1,1001	1,1326	0,4605	0,5087	0,6204
	8x5	1,3175	1,3725	1,3157	0,4258	0,5472	0,4143
	6X5	1,4685	1,5280	1,4595	0,2794	0,4489	0,5626
	7x4	1,5631	1,5468	1,5593	0,3911	0,3931	0,5048
	5x5	1,6137	1,5716	1,5511	0,4451	0,3719	0,5067
	5x4	1,6895	1,7131	1,7349	0,2582	0,3854	0,4875

Se observó que en la mayoría de los casos, el promedio de distancia (o error de cuantización) decreció conforme el tamaño del mapa aumenta. Esto es lógico pues como el número de neuronas aumenta, hay más neuronas para representar los datos y cada vector está más cerca de su neurona ganadora. Pero este criterio fue suficiente para determinar cuál era el tamaño del mapa y la topología más adecuada. Había que fijarse entonces que tanto conservan esas redes su topología. El error topográfico no demostraba una tendencia aparente, lo más notorio era que las redes aleatorias poseen los errores más altos, tal vez porque no tenían una distribución definida y era difícil evaluar si ésta se conserva.

Para efectos prácticos en este trabajo se optimizó la red a una con topología hexagonal de 60 unidades, pues presentó el menor error topográfico, y la distancia promedio de la neurona ganadora también es una de las más bajas.

3. RESULTADOS

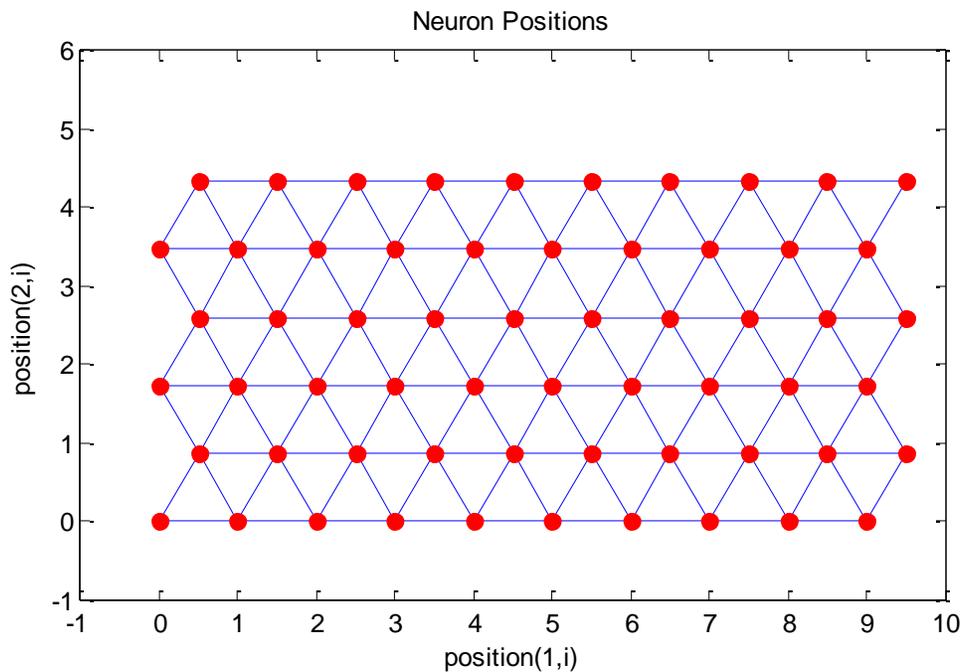
En esta sección se presenta los resultados del entrenamiento y la simulación de la red neuronal con toda la base de datos conformada y se presenta la visualización de estos resultados mediante un software final.

3.1 TOPOLOGÍA DE LA RED

En la *Figura 21* se muestra la topología inicial, donde las neuronas se distribuyen de forma equidistante en un mapa de dos dimensiones. Por lo tanto, es posible visualizar un espacio de entrada de alta dimensión en las dos dimensiones de la topología de la red.

En esta figura, cada punto rojo es una neurona y las líneas azules, la forma en cómo se interconectan. Las dimensiones son 10x6, así que hay un total de 60 neuronas, como se estableció en la...sección 2.5...

Figura 21. Topología de la red.

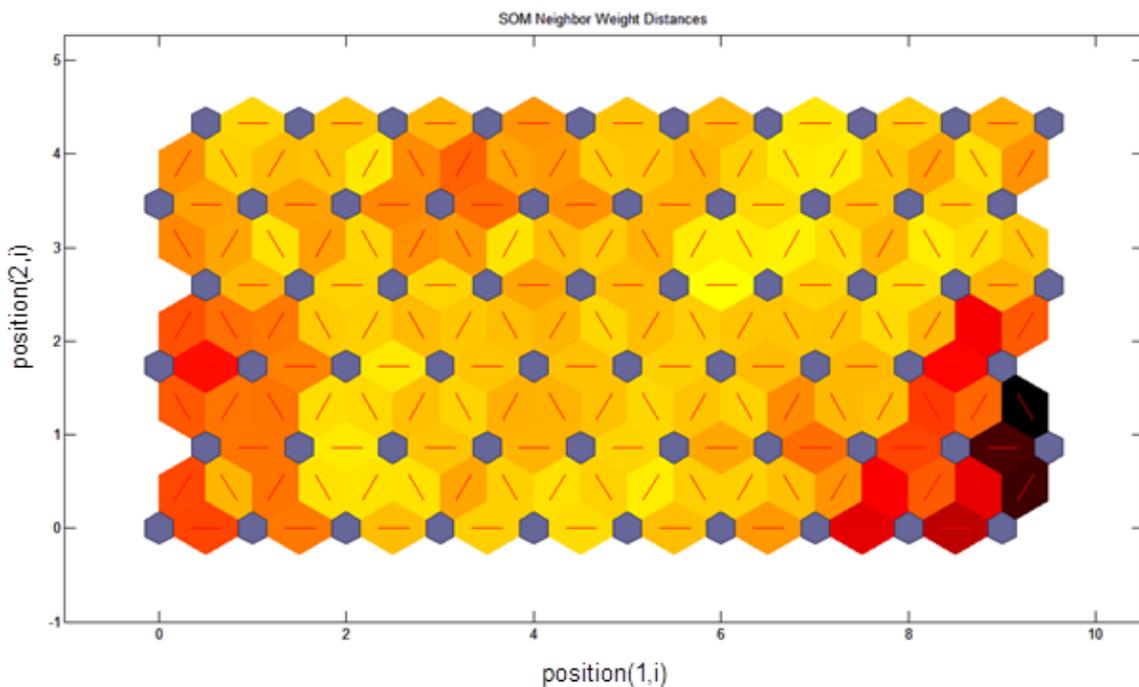


3.2 MATRIZ DE DISTANCIA DE PESOS

Finalizado el entrenamiento, las neuronas han cambiado sus distancias entre ellas de acuerdo a los clusters presentes en los datos de entrada. Esta información se conoce como matriz de distancia de los vectores de pesos de las neuronas o matriz U, y se puede observar en la *Figura 22*.

Los hexagonos azules representan las neuronas. Los colores en las regiones que contienen las líneas rojas indican las distancias entre las neuronas, mientras más oscuro es el color mayor es la distancia.

Figura 22. Matriz U



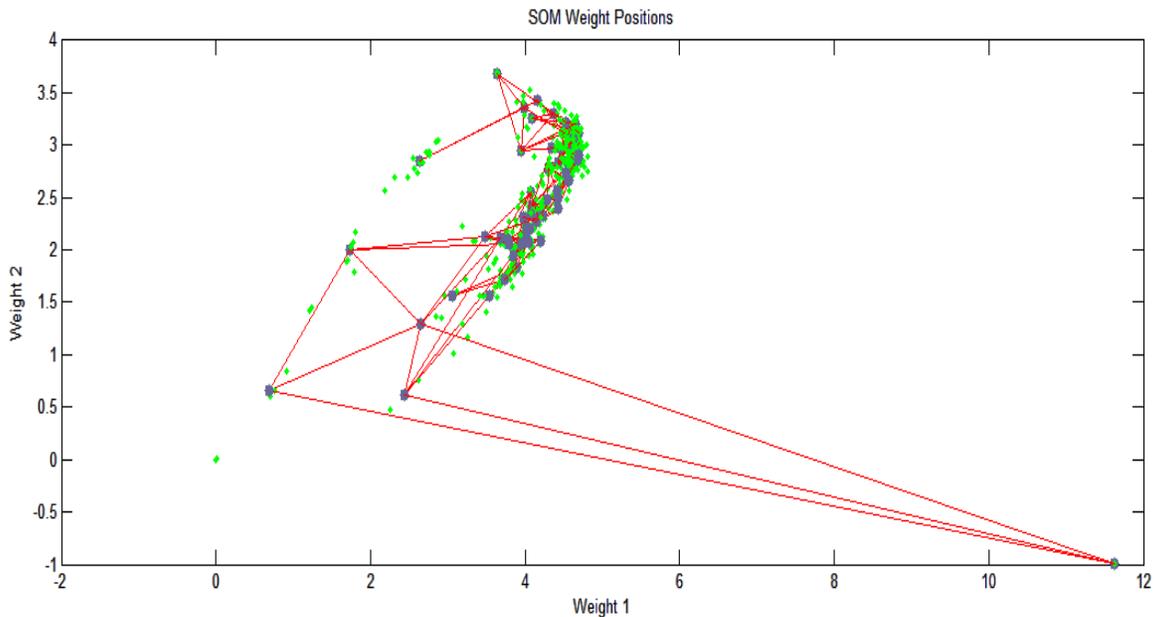
Las neuronas en la mayor parte del mapa poseen las menores distancias entre ellas, esto se puede atribuir a que haya una mayor densidad de datos en esta zona y por ende se requiera más neuronas para su clasificación. Por otra parte, hay 5 o 6 neuronas muy separadas del resto del conjunto, útiles para representar fallas atípicas o información diversa.

Para darle mayor sentido y comprensión a lo dicho anteriormente, en la siguiente sección se visualiza la posición específica de los pesos de las neuronas y los datos de entrada.

3.3 POSICIÓN DE PESOS

En la *Figura 23*, se muestra cómo la SOM clasifica el espacio de entrada, mostrando puntos de color azul grisáceo por vector de pesos de cada neurona y la conexión de las neuronas vecinas con líneas rojas. Los vectores de entrada están dibujados como puntos verdes.

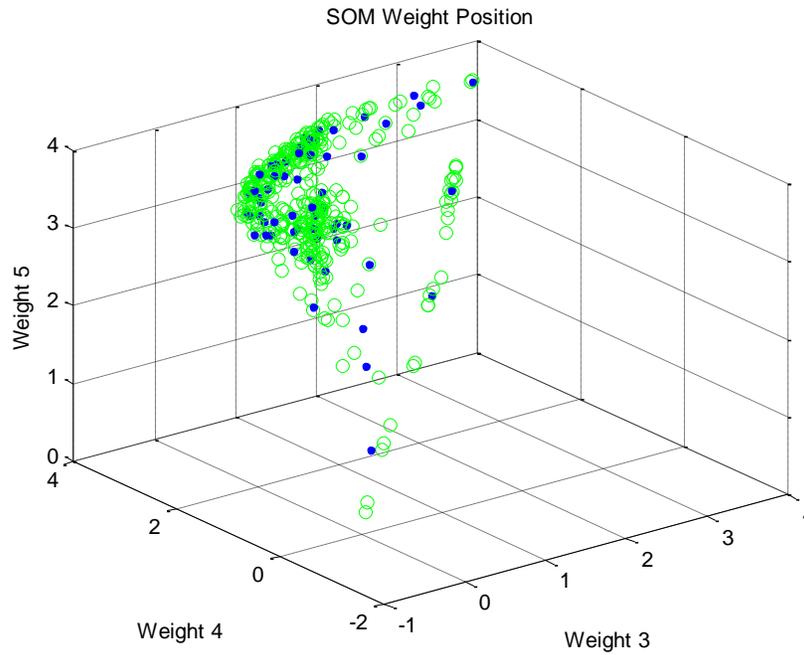
Figura 23. Posición de Pesos



Aquí se detalla más fácilmente la distancia entre cada neurona descrita en la... sección **3.2**...

Hay que advertir que la *Figura 23*, solo proyecta 2 de los 114 valores que posee cada vector de pesos. La máxima visualización que se puede obtener es en tres dimensiones, y se muestra en la *Figura 24*.

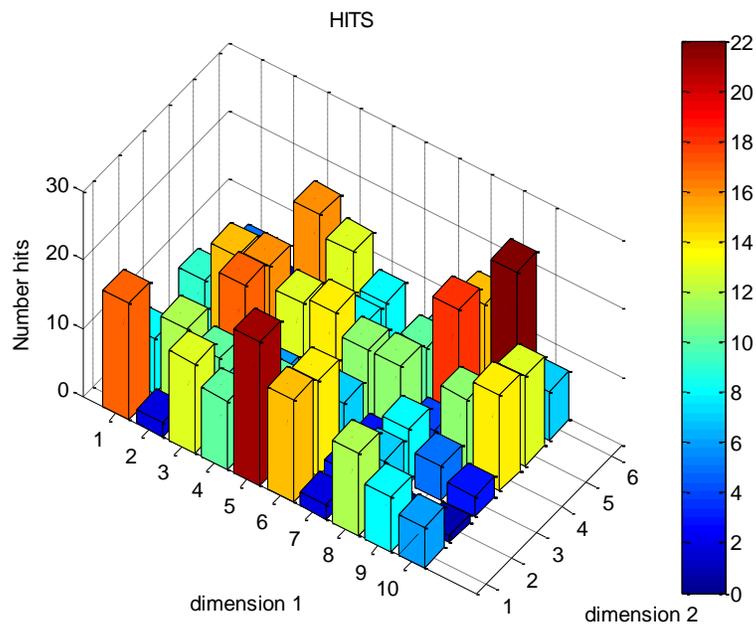
Figura 24. Posición de pesos 3D



3.4 DATOS DE ENTRADA ASOCIADOS

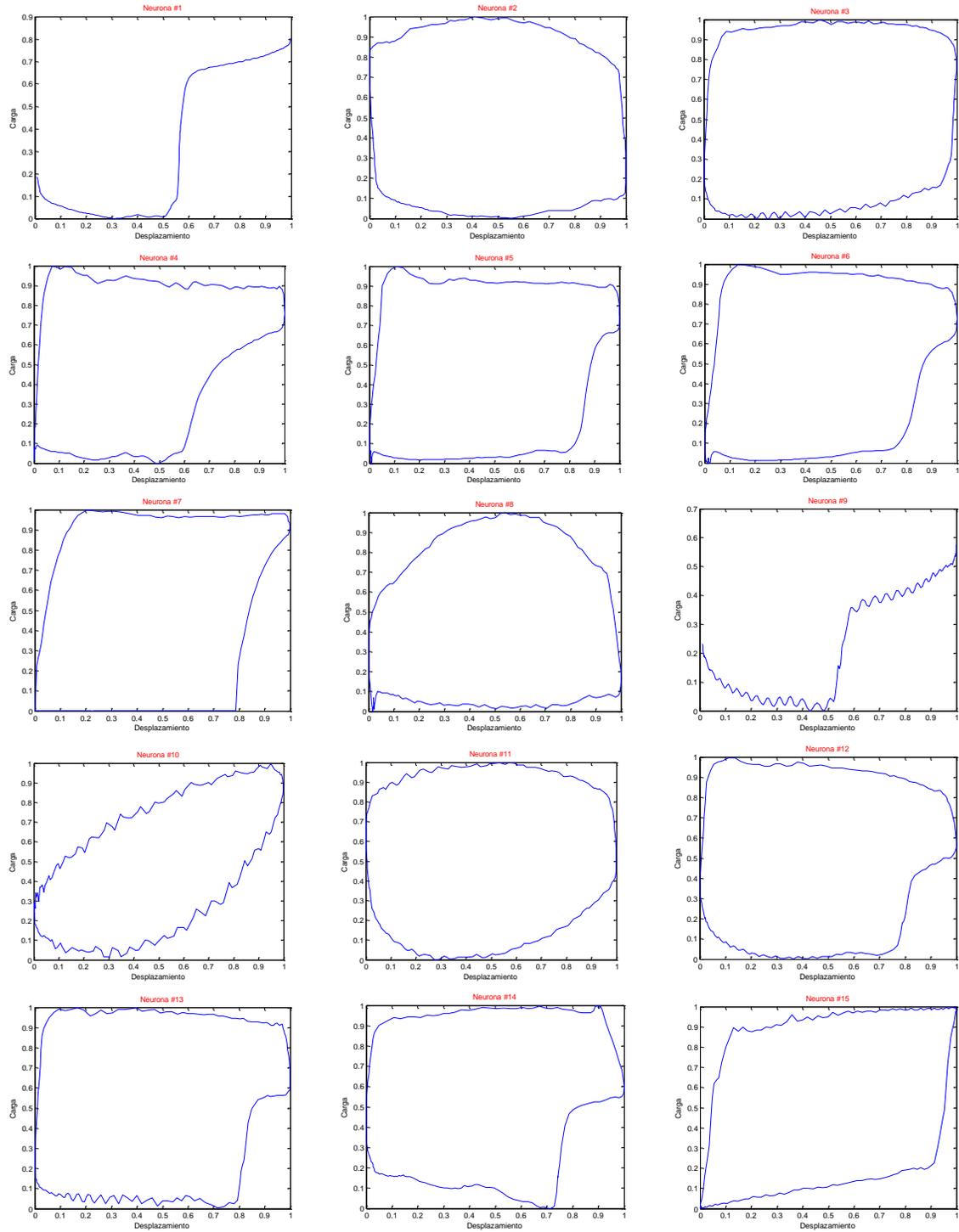
La *Figura 25* indica cuantos de los datos de entrenamiento están asociados a cada neurona. El máximo número de muestras con alguna neurona es 22.

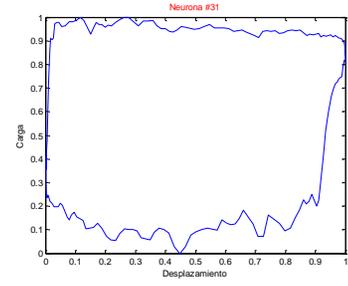
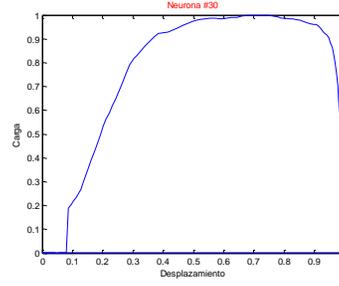
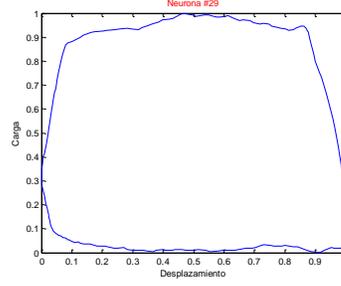
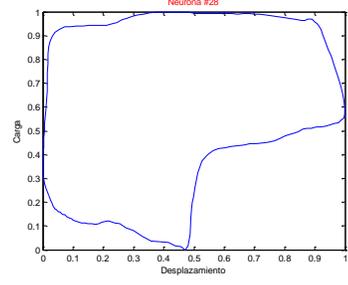
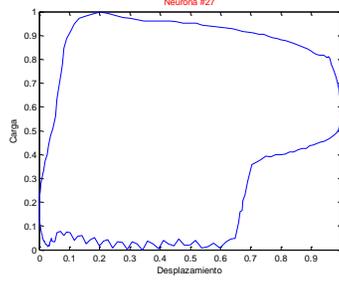
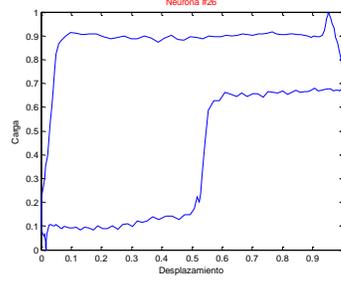
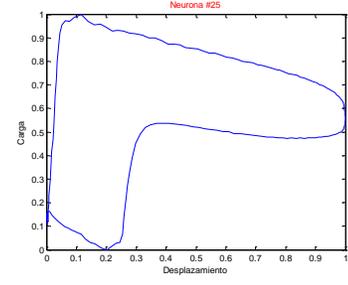
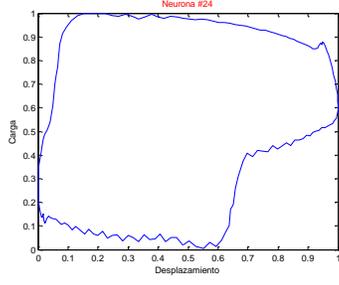
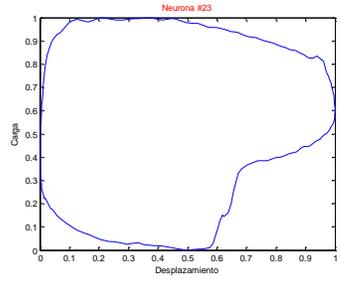
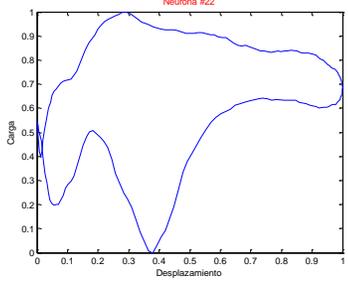
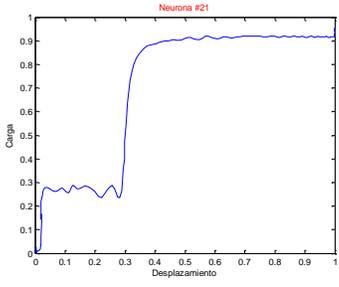
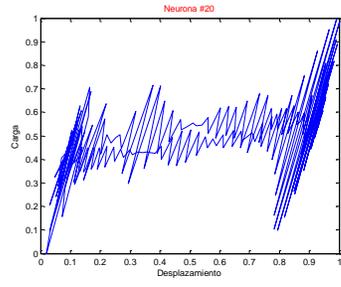
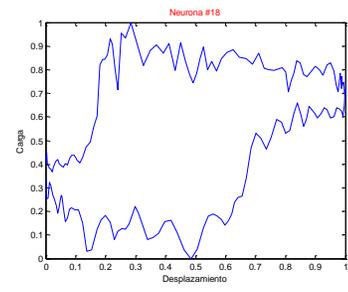
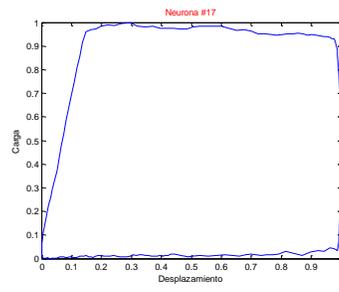
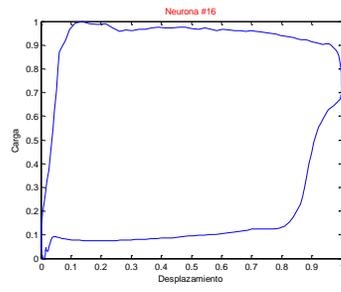
Figura 25. Muestras por neurona

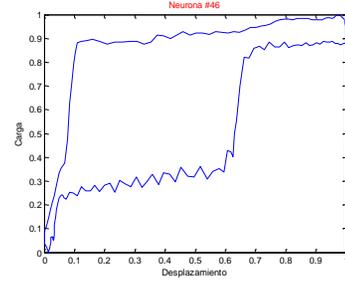
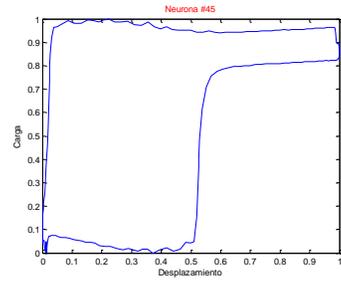
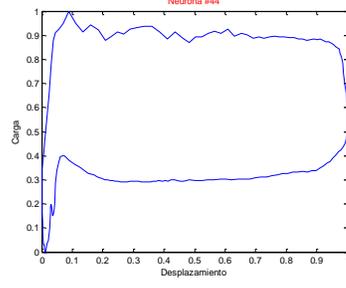
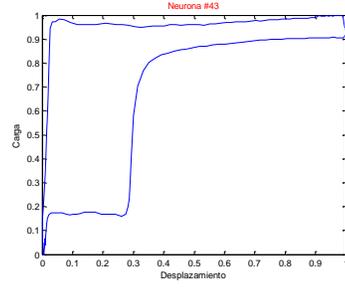
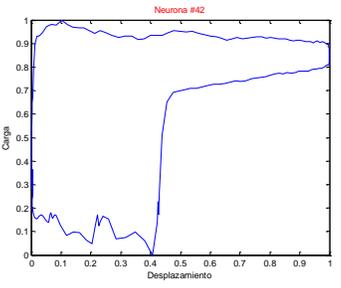
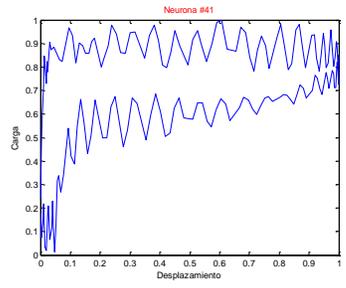
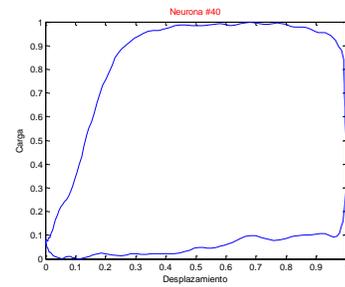
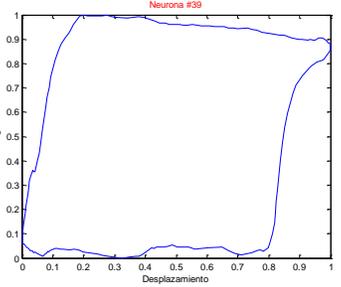
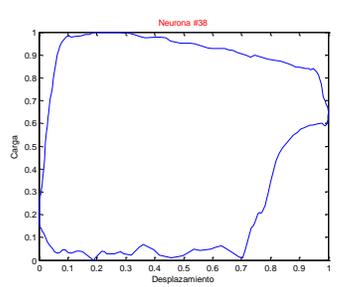
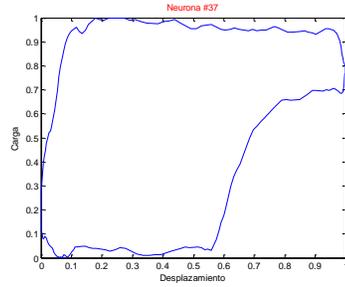
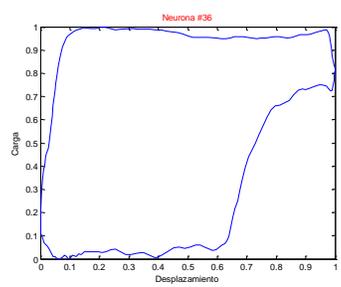
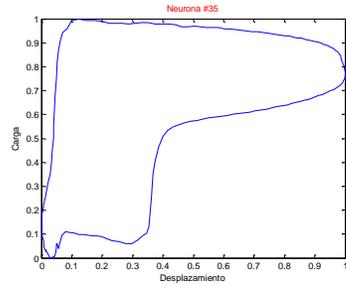
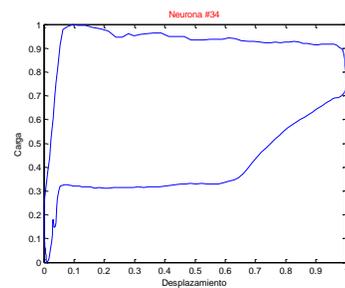
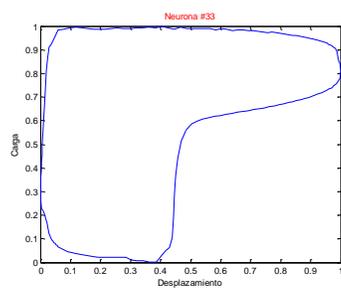
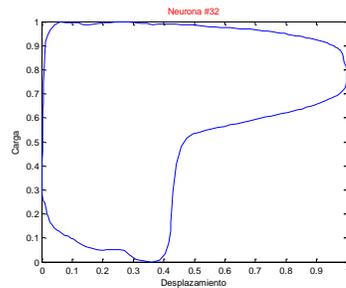


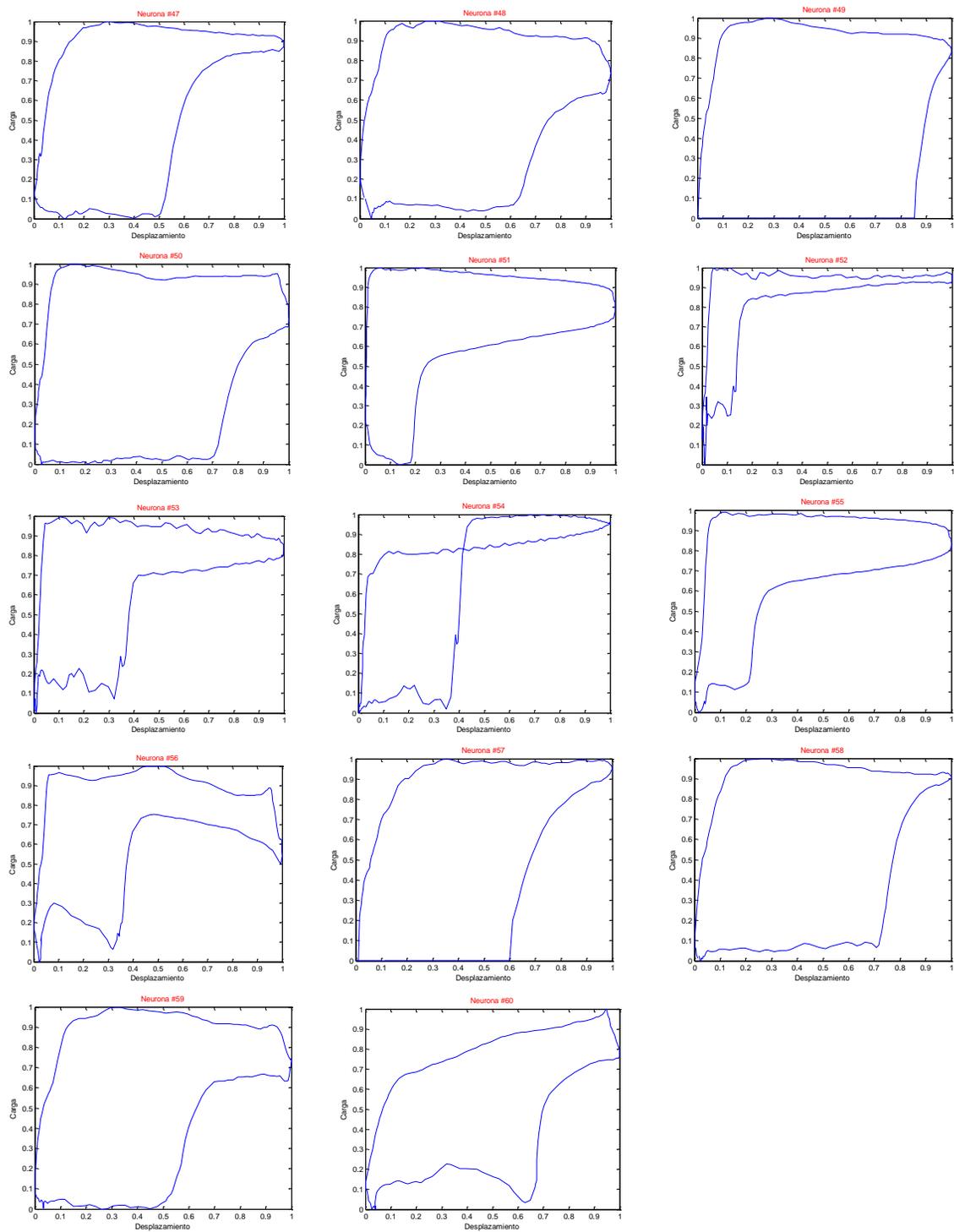
En adelante, se muestran de forma explícita los dinagramas asociados a cada neurona después de la simulación de la red con el 100% de la base de datos.

Figura 26. Dinagrama representativo por cada neurona.









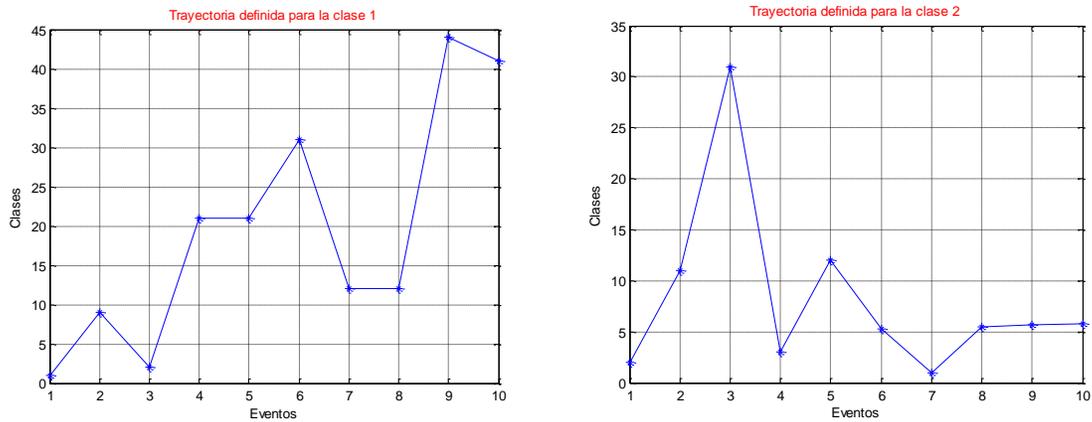
Como se pudo ver la red SOM presenta alta sensibilidad a los cambios bruscos de las señales. En adelante se hace referencia a cada neurona como cada clase obtenida.

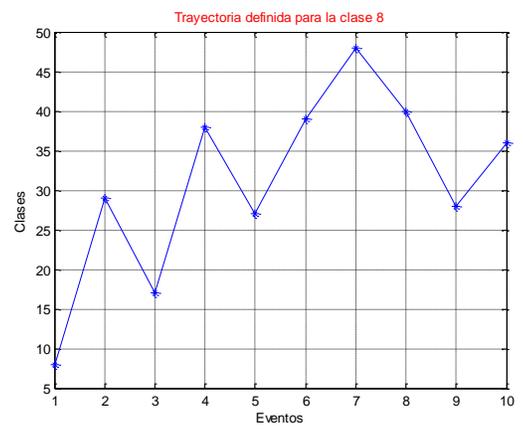
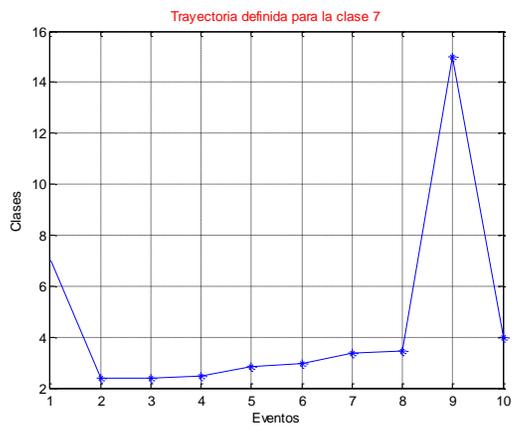
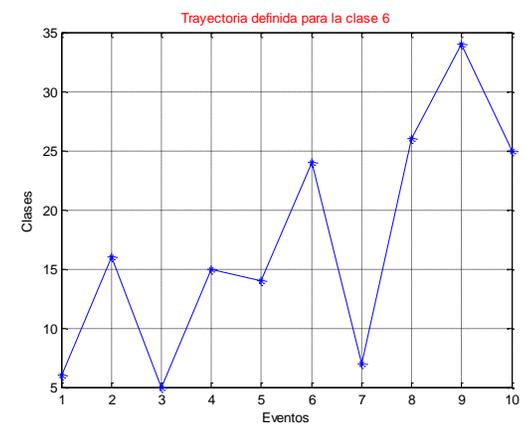
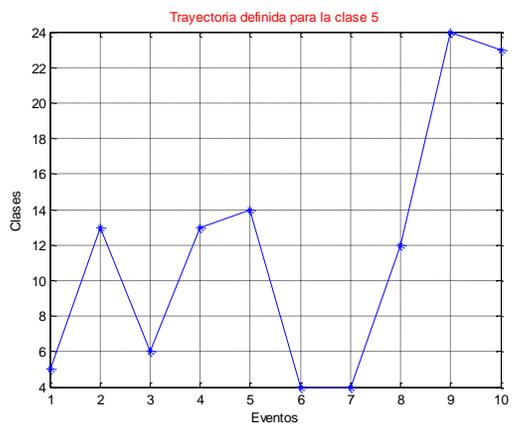
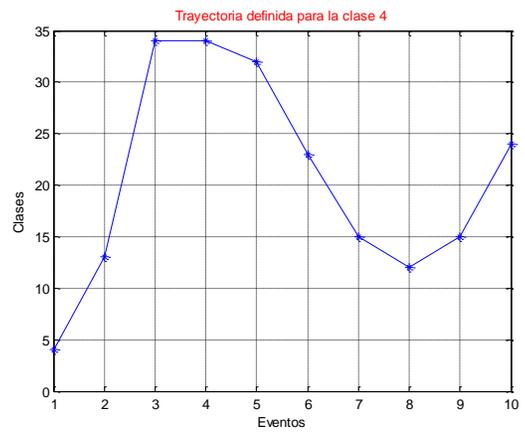
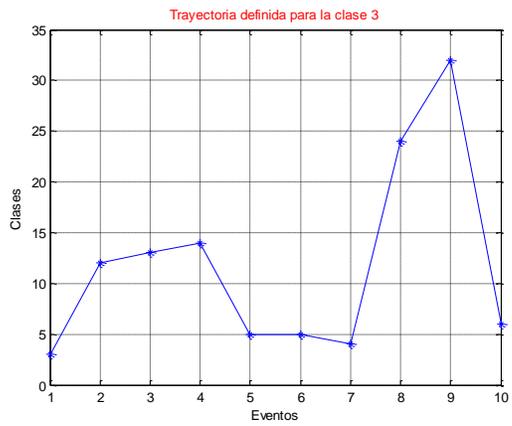
3.5 TRAYECTORIAS

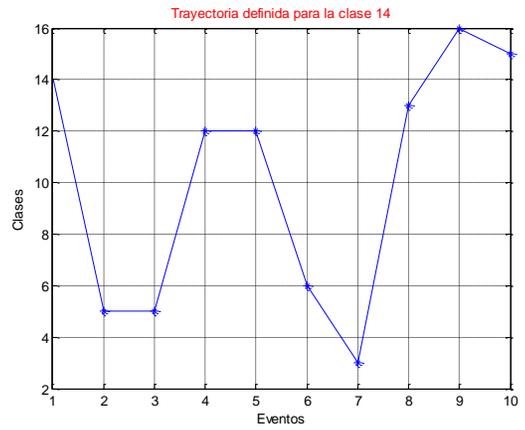
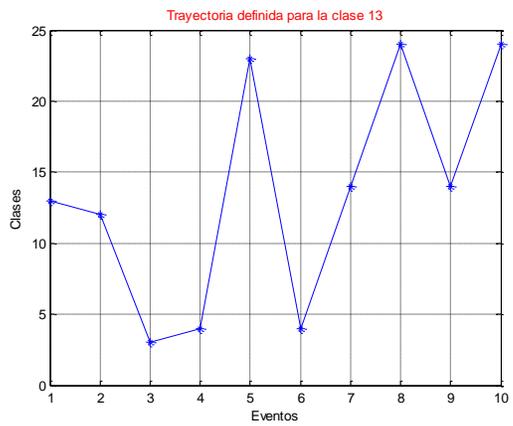
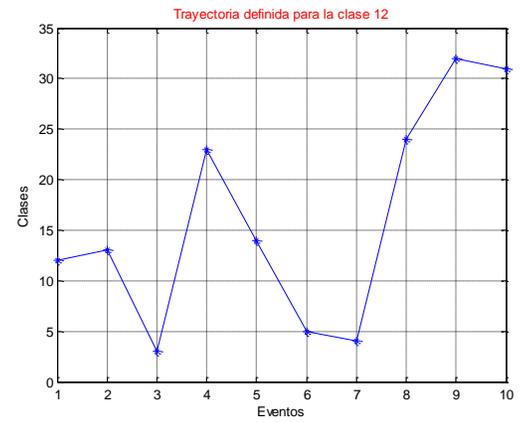
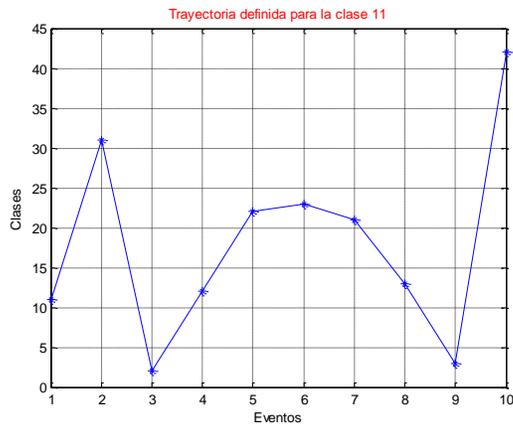
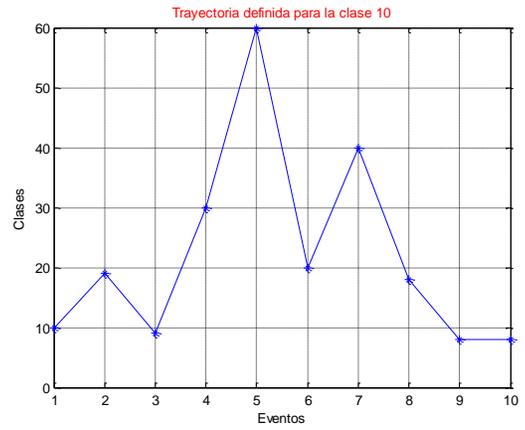
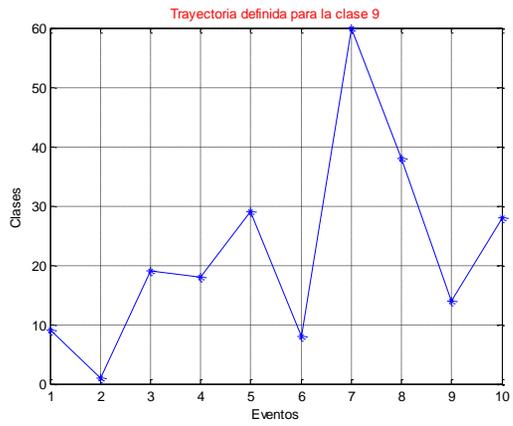
Las trayectorias se utilizan principalmente para predicción, pues muestran la tendencia de las fallas próximas a ocurrir a lo largo de los eventos. Cada trayectoria se obtiene a partir de las distancias ascendentes, desde la neurona ganadora hasta la décima neurona.

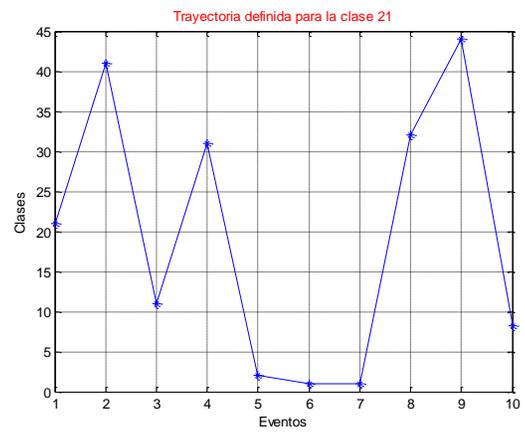
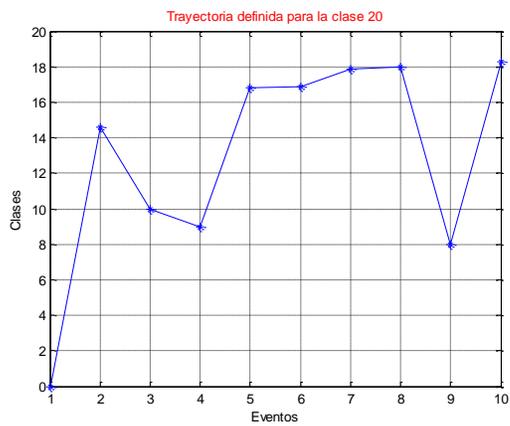
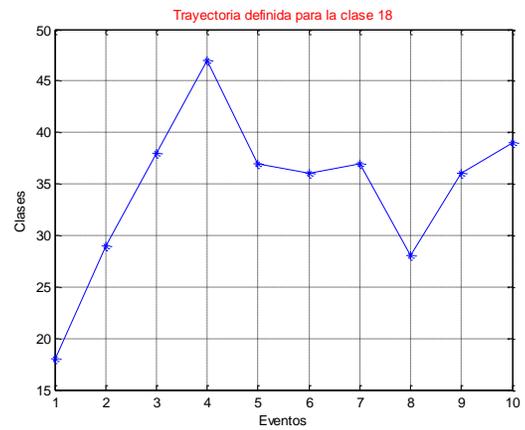
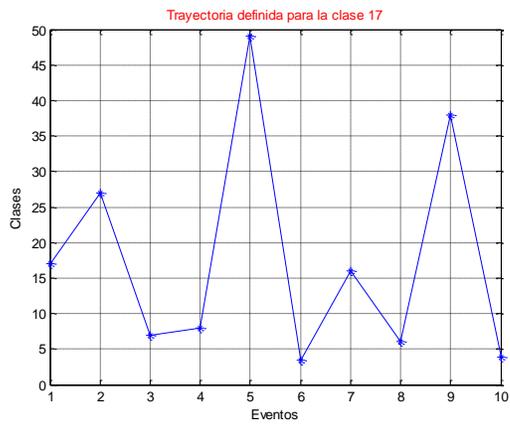
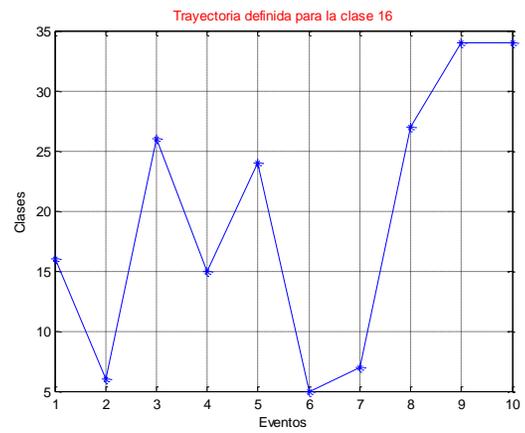
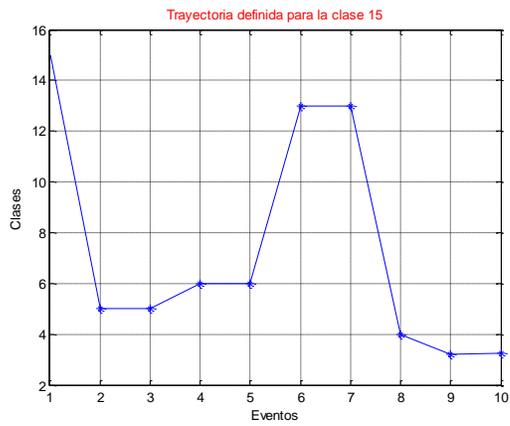
Se inició con la grafica de todas las trayectorias posibles para cada una de las clases obtenidas. Luego se estudió cual es la clase más frecuente para cada uno de los eventos calculando su porcentaje de ocurrencia, para finalmente establecer una sola trayectoria que muestre la predicción más adecuada.

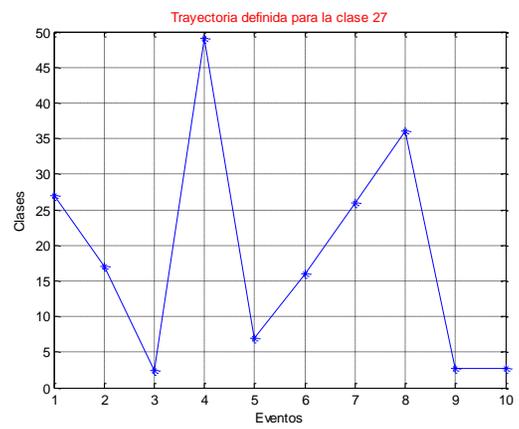
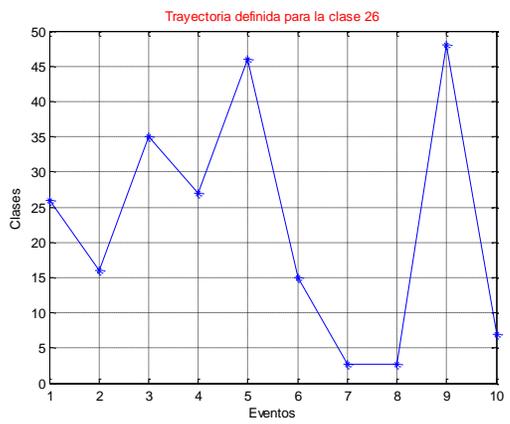
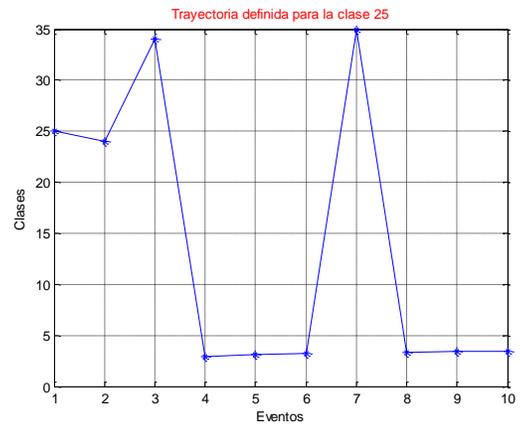
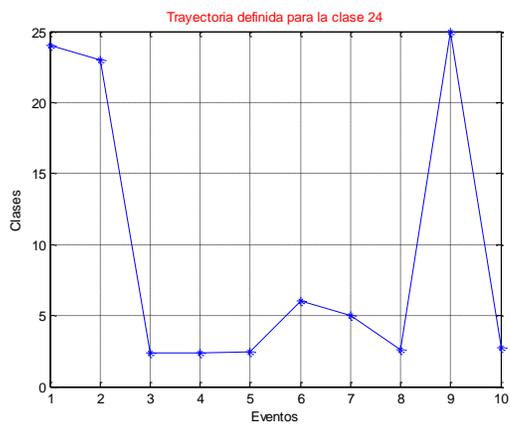
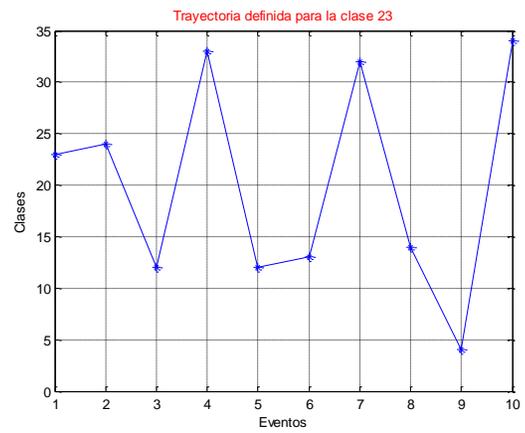
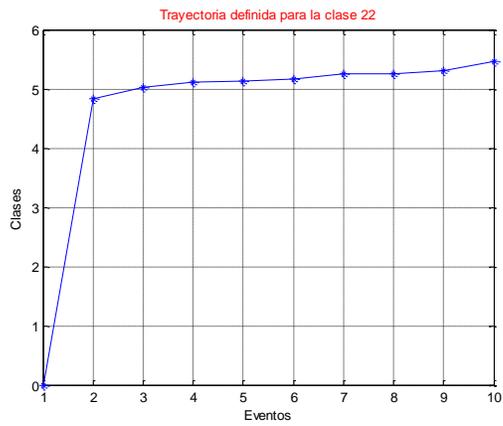
Figura 27. Trayectorias definidas para cada clase.

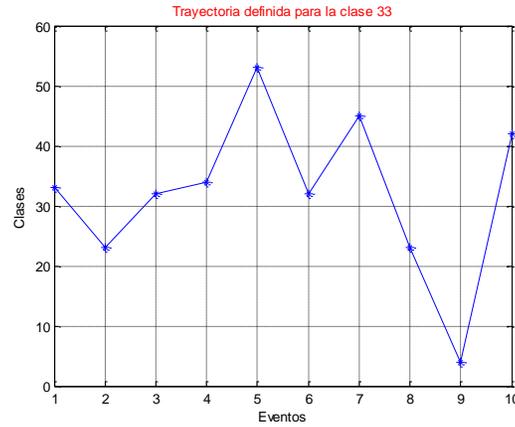
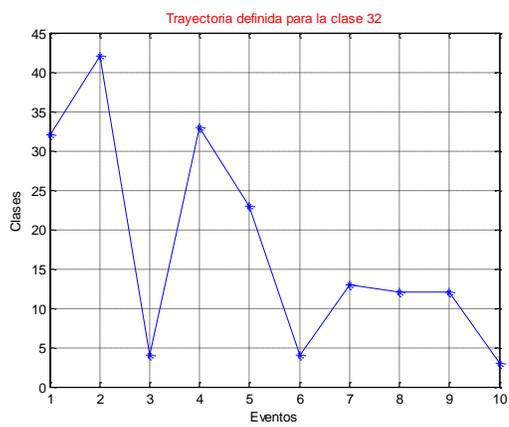
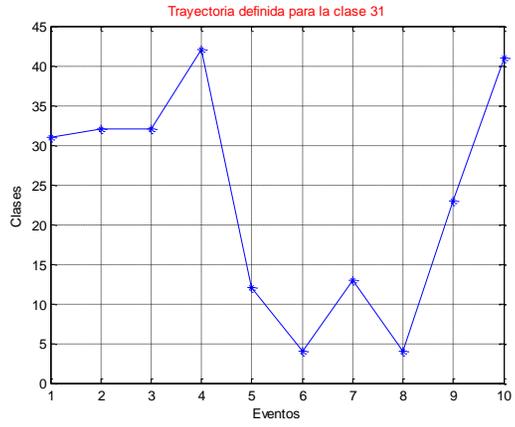
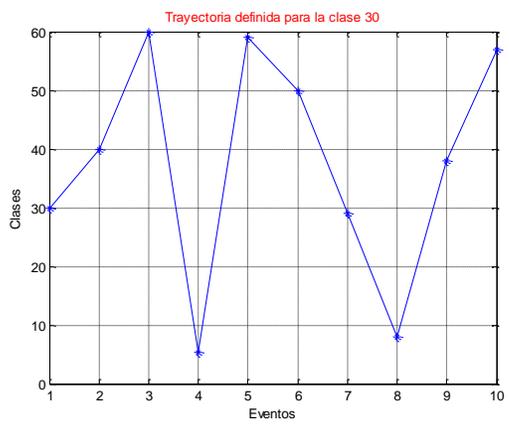
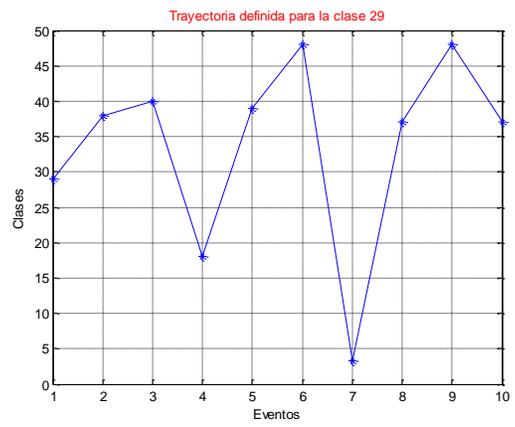
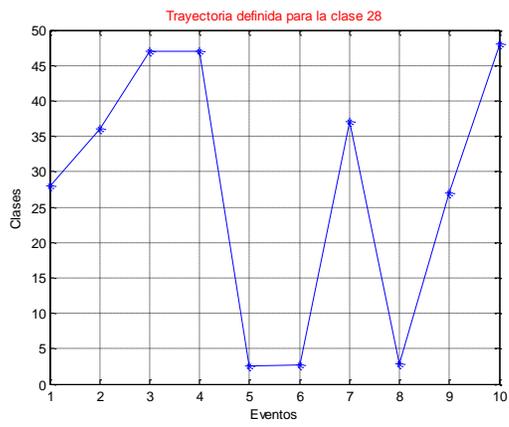


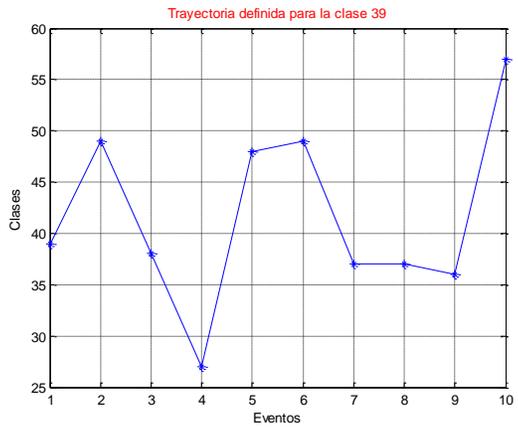
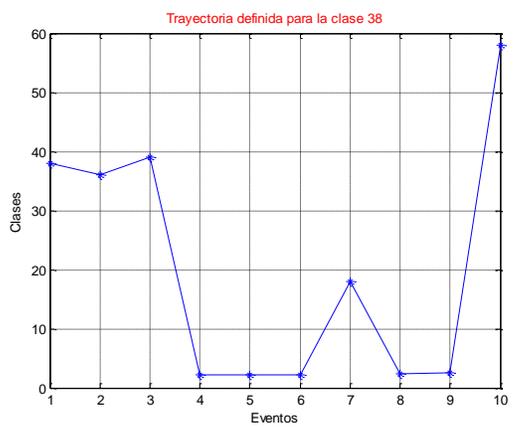
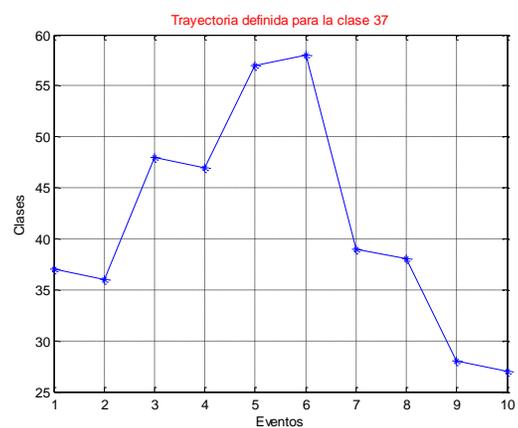
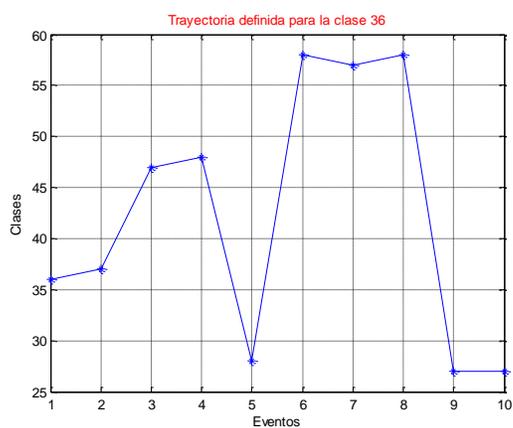
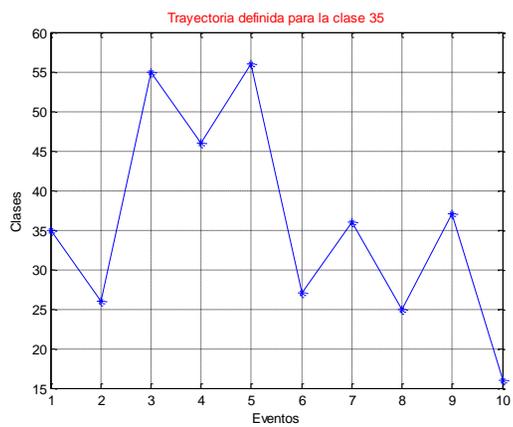
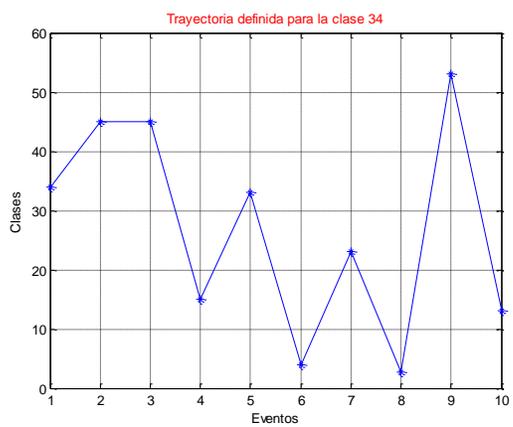


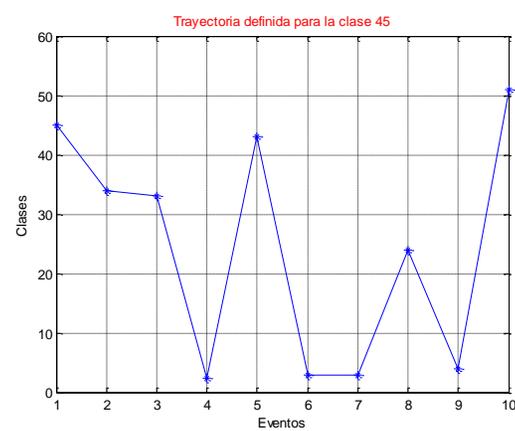
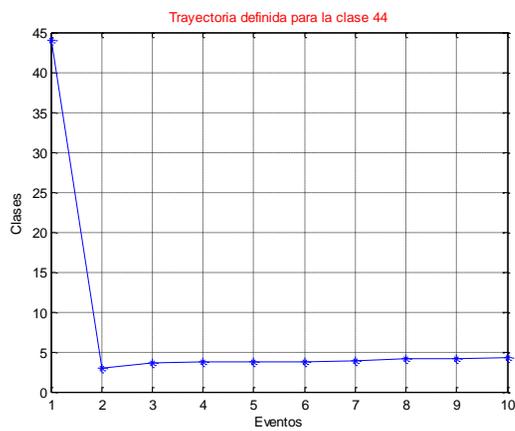
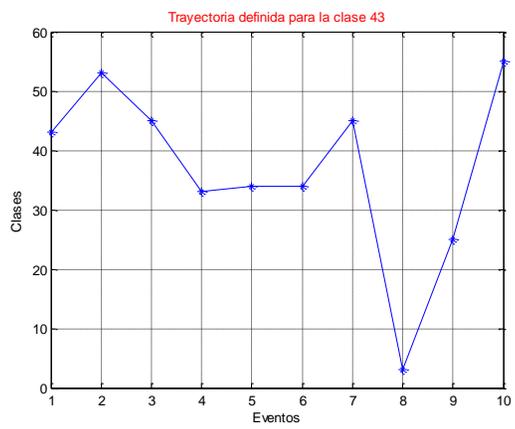
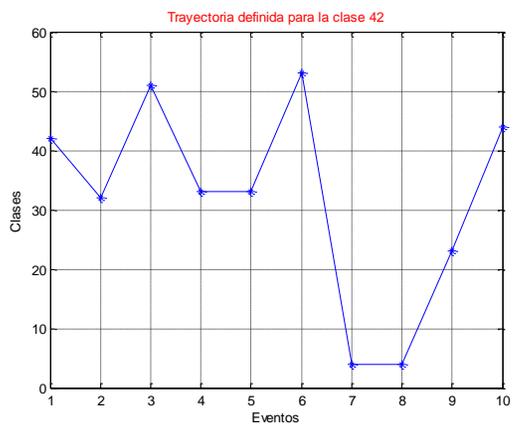
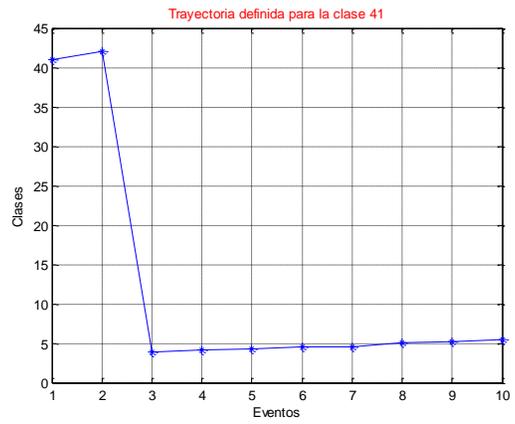
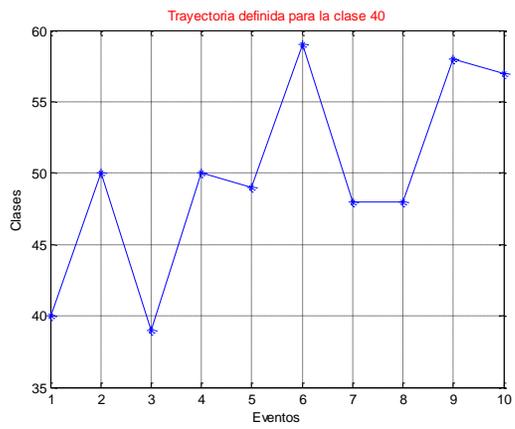


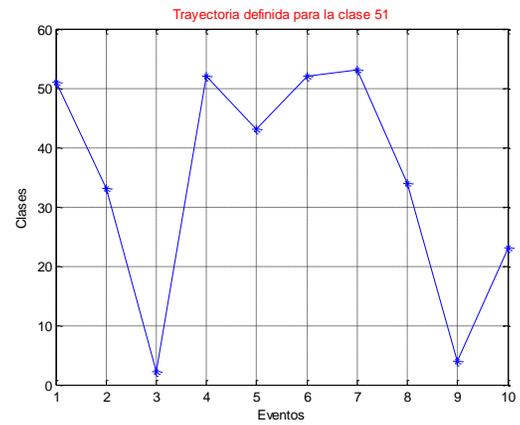
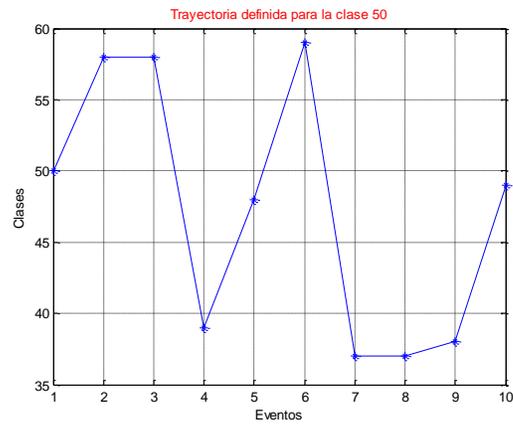
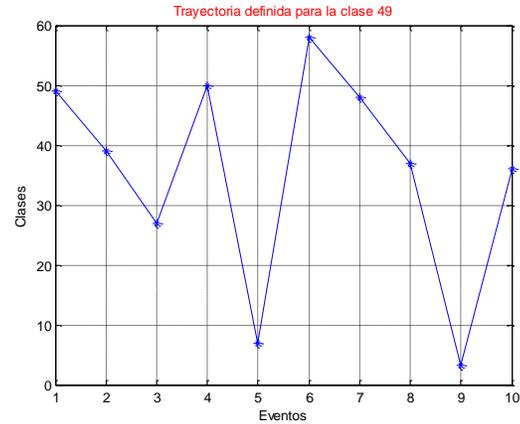
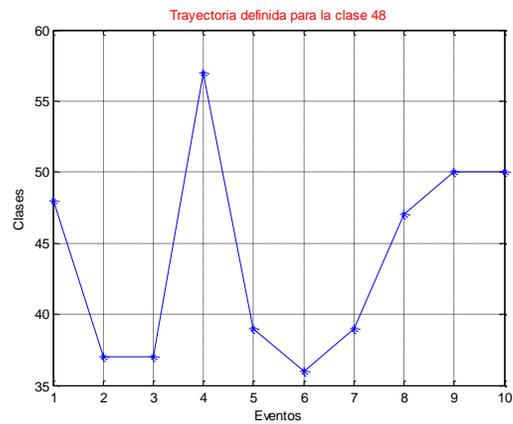
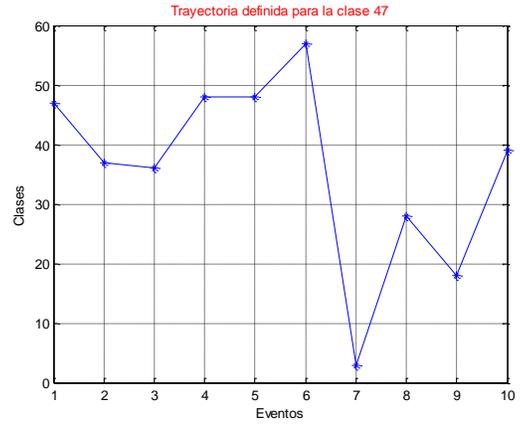
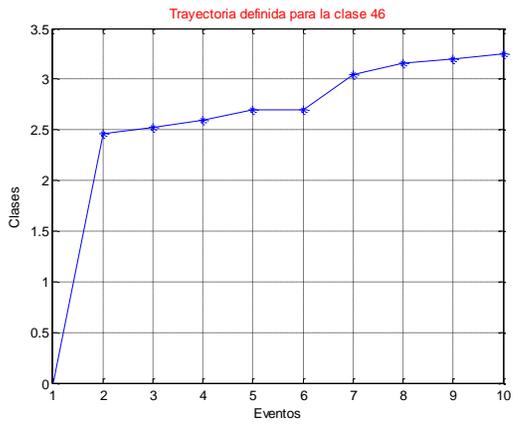


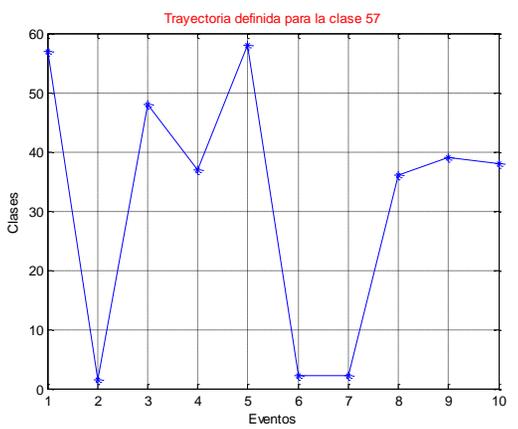
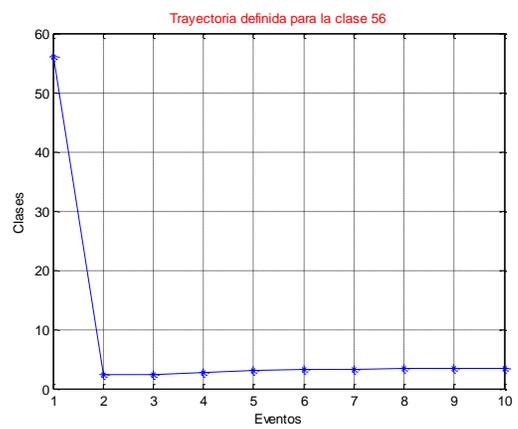
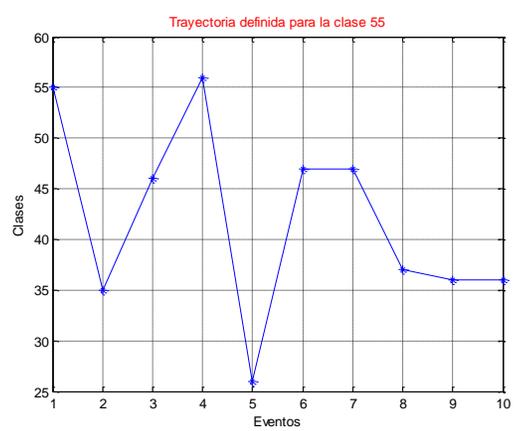
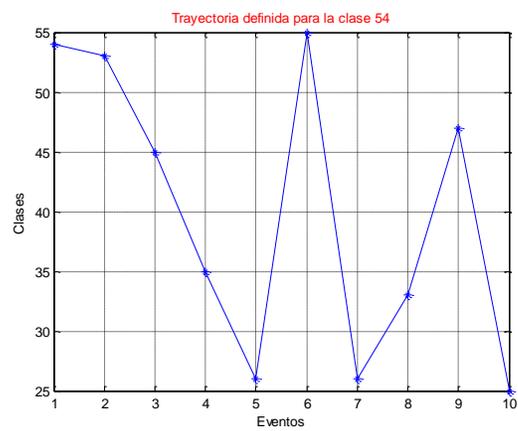
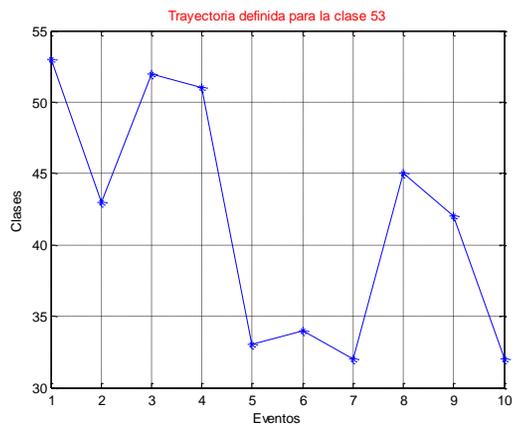
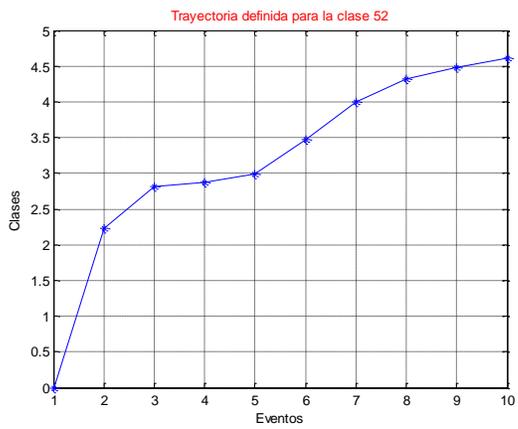


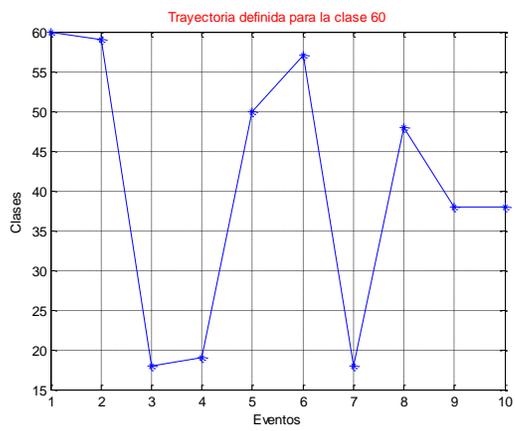
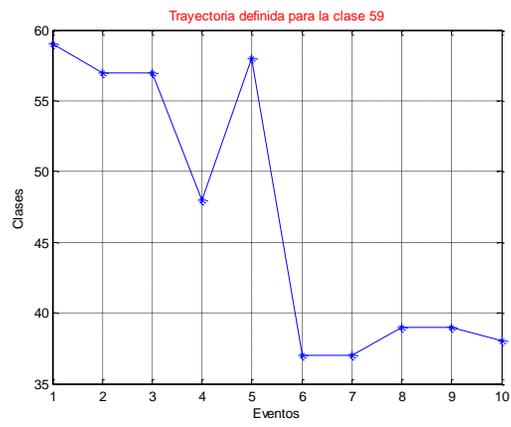
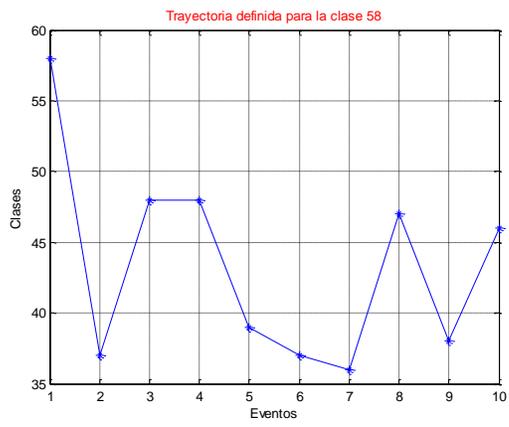












4. SOFTWARE DINASOM

El software DinaSOM usa una red neuronal SOM entrenada para la optimización de unidades de bombeo mecánico. Se diseñó una interfaz gráfica en donde el operario puede encontrar descripciones de comportamientos como: análisis de registros históricos, el estudio de trayectorias de fallas asociadas y el monitoreo por medio de una interfaz gráfica tipo radar. Esta interfaz permite relacionar formas de dinagramas con los nombres de las fallas brindándole al usuario una base de conocimiento rápida y acertada para realizar la interpretación de dinagramas, agilizando la identificación de problemas en los sistemas de bombeo mejorando el tiempo de respuesta en caso de una contingencia.

El software DinaSOM se destaca porque además de ser una herramienta de identificación y análisis muestra las tendencias de las fallas en forma de trayectorias, permitiendo tener un panorama general de la situación futura del pozo ya que no solo identifica la falla en el momento sino que muestra una serie de fallas cercanas en formas de trayectoria, con las cuales se puede prevenir un daño que saque totalmente de funcionamiento el pozo.

4.1 VALIDACIÓN

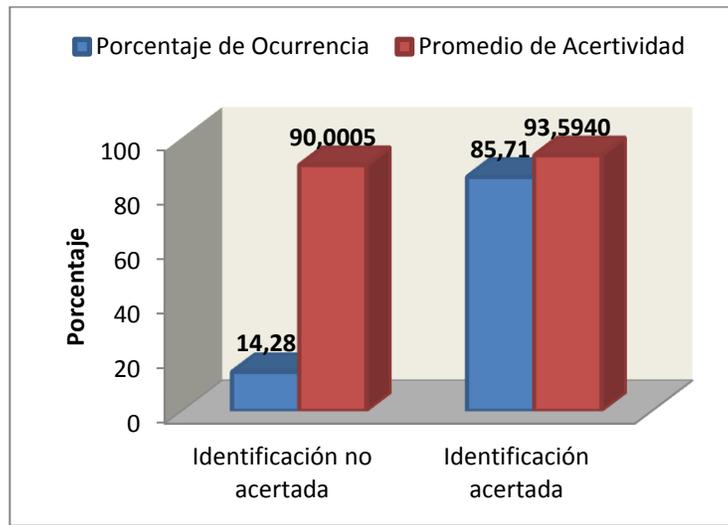
Por ser un software basado en una red de entrenamiento no supervisado su validación es más compleja, porque no existe una función específica para evaluar el performance de esta red.

Se realizaron 70 pruebas para observar el éxito del clasificador. Los resultados se muestran en el *Cuadro 5* y la *Figura 28*.

Cuadro 5. Validación del software

	Ocurrencia (%)	Porcentaje promedio de asertividad
Identificación no acertada	14,28	90,0005
Identificación acertada	85,71	93,5940

Figura 28. Validación del software

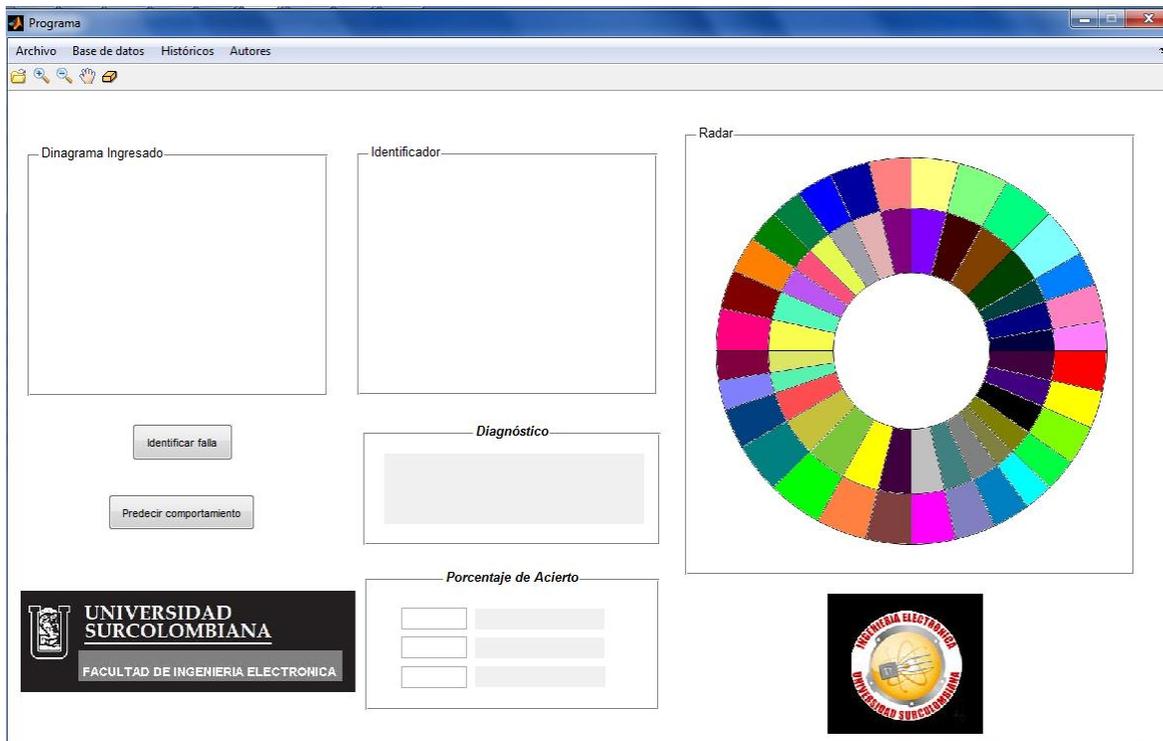


Se observa que el 85,71% del total de los casos evaluados fueron exitosos frente a un 14,28% que fueron confusos o no acertados. La ocurrencia de estos casos erróneos se debe principalmente al dinagrama escogido para la visualización de cada falla, es decir que el dinagrama representativo de cada neurona, en pocos casos, no fue el más adecuado. Se corrobora esta afirmación con los valores de los porcentajes promedios de asertividad de las neuronas ganadoras; en las dos opciones este valor fue superior al 90%, esto quiere decir que la neurona ganadora en cada caso se aproxima bastante bien a los datos de entrada. Cabe destacar que cuando la identificación es positiva la neurona ganadora tiene un porcentaje de acierto por encima del 92%.

4.2 INTERFAZ GRÁFICA

Fue diseñada con Matlab utilizando el GUIDE como se observa en la *Figura 29*. Esta interfaz es de manejo fácil y puede ser ejecutada por cualquier persona con o sin conocimientos de Matlab.

Figura 29. Pantalla principal de DinaSOM



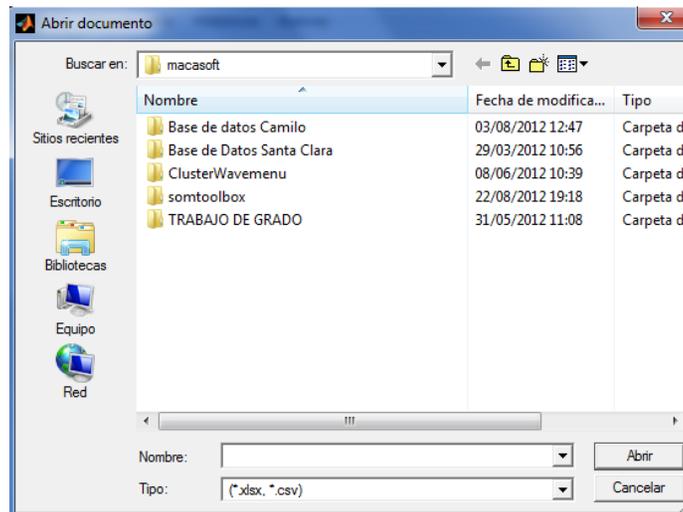
4.3 FUNCIONAMIENTO

DinaSOM cuenta con dos posibilidades para el ingreso de datos, estos datos deben estar en formato (.xls) o extensión (.csv) en forma de tabla con dos columnas con datos mayores de 180 filas.

La primera columna debe contener los datos correspondientes a la carga de los dinagramas y la segunda columna corresponde a los datos del desplazamiento tomados por el dinamómetro de la bomba y al final lo que se le ingresa al software DinaSOM en una matriz de dos columnas por 180 o más filas.

4.3.1 Ingreso de datos. En la pantalla principal, en el Toolbar, se le da clic izquierdo al icono universal de Abrir. Al seleccionar Abrir se muestra otro sub-menu como se ve la *Figura 30*.

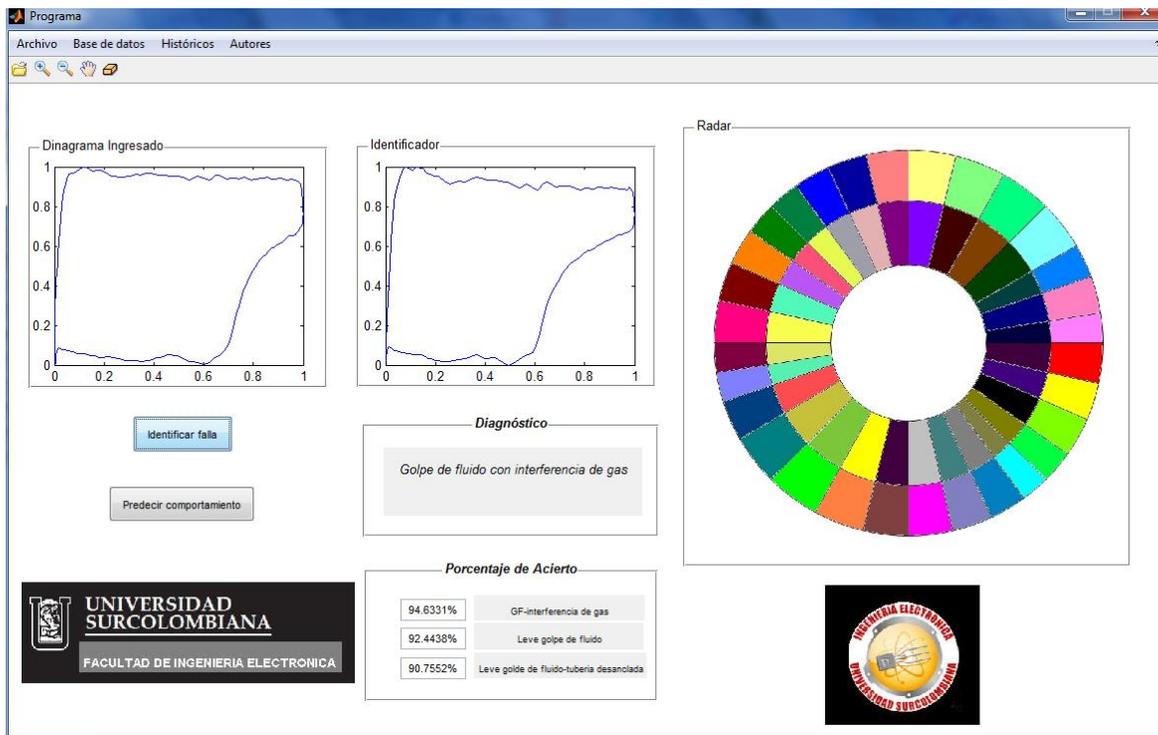
Figura 30. Abrir un archivo



Luego se busca, se selecciona y se abre el dinagrama que se desea cargar.

4.3.2 Identificación de fallas. Después de que el archivo ha sido cargado al programa, se le da clic en el botón “Identificar falla” y se tendrá un resultado como la *Figura 31*.

Figura 31. Identificación.



El dinagrama que se ha cargado previamente se muestra en el panel “Dinagrama ingresado”, y en el panel “Identificador” se muestra la falla a la cual ha sido asociado.

También posee un panel de “Diagnostico” donde se visualiza una breve descripción de la falla asignada por el diseñador para una mejor identificación del problema, esto pensando en la posibilidad de identificar de forma más rápida la falla para que el experto formule las correcciones de los daños en los equipos de forma más ágil.

Y el panel “Radar”, que apoya el proceso de identificación, puede seguir el comportamiento del pozo a medida que se van identificando las fallas y esto da la posibilidad de tener una gráfica del estado de las fallas de la bomba y cuál sería su tendencia.

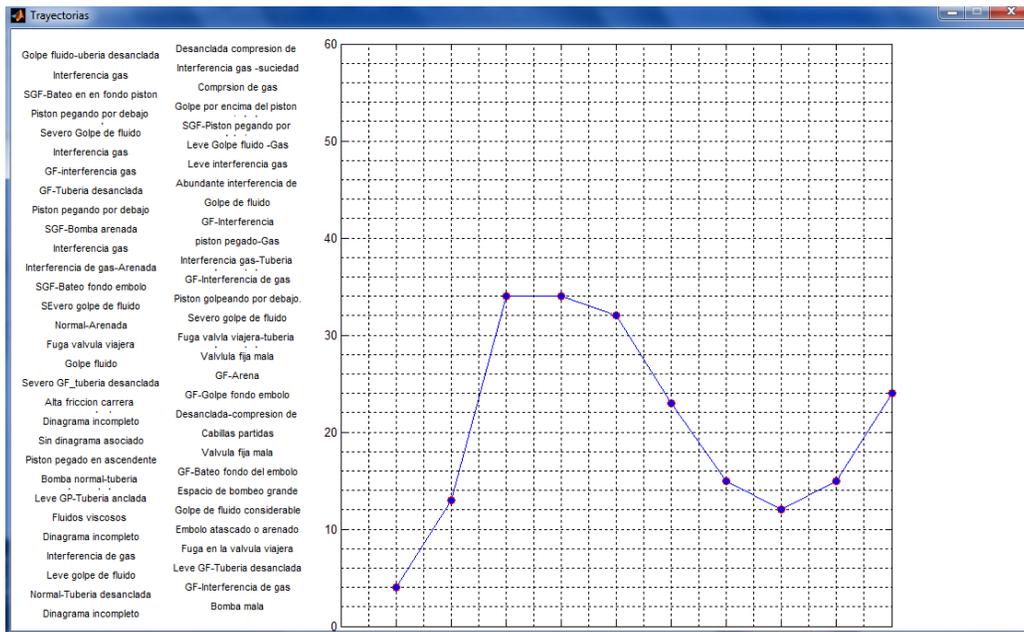
La red neuronal SOM que contiene el identificador toma las características de la carta dinográfica ingresada y selecciona la falla que más acerca a esas características. Esta buena asociación de patrones es producto del entrenamiento y la robustez con la que cuentan las redes SOM sin importar que el dinagrama de entrada nunca haya sido procesado en el entrenamiento de la red neuronal.

En la casilla con el nombre de: porcentaje de acierto, ubicada bajo el panel de diagnóstico, se visualiza porcentaje de semejanza del dinagrama ingresado y el dinagrama identificado por el *software*.

4.3.3 Predicción de comportamiento. Esta opción se implementó como herramienta para visualizar las trayectorias de las posibles futuras fallas más cercanas al dinagrama de entrada y evaluar su grado de certeza, mejorando de esta forma la confianza a la hora de tomar decisiones para ejecutar trabajos en los equipos de fondo.

Al presionar el botón “Predecir comportamiento” como se muestra en la Figura 32. El programa calcula la trayectoria de las posibles fallas más próximas a ocurrir.

Figura 32. Predicción



La gráfica muestra la trayectoria de las posibles fallas futuras sino se realiza un mantenimiento a tiempo de la bomba. Cada punto de la trayectoria hace referencia a la falla más próxima en cada evento por ocurrir.

5. CONCLUSIONES

- El desarrollo de esta aplicación inteligente y además efectiva, requiere conocimientos previos de interpretación de dinagramas, métodos para realizar procesamiento de señales y aplicación de redes neuronales artificiales SOM, debido a que si existen falencias de conocimientos en estos campos se hace muy laborioso desarrollar un algoritmo para la identificación de problemas en los equipos de subsuelo de las unidades de bombeo mecánico de manera exitosa.
- El hecho de igualar las dimensiones de todos los vectores de entrada de los dinagramas para su entrenamiento, hace que se pierda cierta información de la señal que puede ser relevante en el momento de la clasificación.
- Las redes SOM crean mapas autorganizados para clasificar muestras con tanto detalle como se desee, seleccionando el número de neuronas en cada dimensión del mapa de salida. Así proyectan datos de entradas de altas dimensiones en bajas dimensiones respetando la densidad espacial de los datos.
- Después de probar varios métodos de procesamiento para la extracción de características relevantes de los dinagramas, se encontró que el más adecuado para este tipo de señales, que tienen una frecuencia de muestreo variable o dinámica, es la Transformada Wavelet; debido a que este método utiliza un algoritmo matemático que se adapta a esa variación de la señal y proporciona diferentes resoluciones de tiempo y frecuencia.
- La información obtenida de la transformada wavelet viene discriminada en forma de niveles, donde cada nivel contiene coeficientes de aproximaciones y detalles, formando un árbol jerárquico. Esto quiere decir, que luego de analizar una señal existen diversas rutas o caminos que contienen nodos con la información de descomposición más notable de la señal.
- Las redes neuronales artificiales (SOM) presentaron un mejor reconocimiento al involucrar en su etapa de aprendizaje dinagramas con fallas combinadas, acorde a los problemas reales presentes en los pozos.
- En las redes SOM, debido a la ausencia de capas ocultas y por poseer un algoritmo no supervisado, el tiempo de entrenamiento toma de 6 a 10 segundos. Tiempo considerablemente bajo comparado con el de otras redes de carácter supervisado.

6. RECOMENDACIONES

- Es recomendable establecer nuevamente las dimensiones de la matriz para el entrenamiento en base a la media de todas las dimensiones de los vectores, es decir que en ciertos casos se tendrá que reducir la cantidad de filas con tasas de muestreo equitativa y en otros casos aumentar la dimensión de los vectores correspondientes con la media de estos.
- Para diagnosticar problemas de bombas del grupo 2 se necesita desarrollar un algoritmo predictivo de ecuación de onda que modele no solo el comportamiento de la sarta de cabillas sino también de los efectos de la mezcla del fluido.
- Se propone realizar estudios más a fondo y en conjunto con el programa de ingeniería de petróleos, acerca del análisis del significado de los diferentes dinagramas, y con este conocimiento generado tener avances más grandes en el área de la Neuro-computación aplicado a la industria petrolera.

BIBLIOGRAFÍA

BASOGAIN Olabe, Xabier. Curso: Redes neuronales artificiales y sus aplicaciones. Bilbao, España. Escuela superior de ingeniería de Bilbao. Dpto. de ingeniería de sistemas y automática. 76 p.

BERNAL, Hugo Alberto y RODRIGUEZ, Alfredo. Identificación de zonas productoras de hidrocarburos aplicando DSP y métodos de correlación de registro de perfilaje. Trabajo de grado Ingeniero Electrónico. Neiva. Universidad Surcolombiana. Facultad de Ingeniería, 2009. 55 p.

HERNADEZ DIAZ, Marianito. Análisis Comparativo de Algoritmos para Reducción de Ruido en Señales Utilizando Wavelets. Tesis profesional Licenciatura en Ingeniería en Electrónica y comunicaciones. Cholula, Puebla, México: Universidad de las Américas Puebla. Escuela de ingeniería. Departamento de Ingeniería Electrónica, 2003.

HUDSON BEALE, Mark; HAGAN, Martin and DEMUTH, Howard. Neural Network Toolbox User's Guide. Self-Organizing and Learning Vector Quantization Nets. [s.l]: The MathWorks, Inc., 1992-2012. p. 6.2 – 6.47.

INSTITUTO COLOMBIANO DE NORMAS TÉCNICAS Y CERTIFICACIÓN. Documentación. Presentación de Tesis, Trabajo de grado y Otros trabajos de investigación. Sexta actualización. NTC 1486. Bogotá DC.: ICONTEC, 2008. 41 p.

Documentación. Referencias bibliográficas. Contenido, forma y estructura. NTC 5613. Bogotá DC.: ICONTEC, 2008. 38 p.

Documentación. Referencias documentales para fuentes de información electrónica. NTC 4490. Bogotá DC.: ICONTEC, 1998. 27 p.

KOHONEN, Teuvo. The Self-Organizing Map. **En:** PROCEEDINGS OF IEEE. Septiembre, 1990. vol. 78, no. 9, p. 1464-1480.

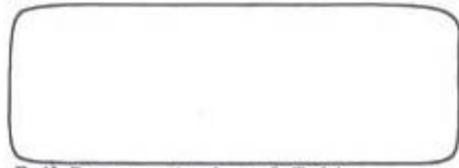
MISITI, Michel, *et al.* Wavelet Toolbox for use with MATLAB. [s.l]: The MathWorks, Inc., 1996-1997. 626 p.

OROZCO, Mauricio. Selección efectiva de características wavelet en la identificación de bioseñales 1-D y 2-D usando algoritmos genéticos. Trabajo de grado Magister de Ingeniería. Caldas: Universidad Nacional de Colombia. Facultad de Ingeniería y Arquitectura. Departamento de Ingeniería Eléctrica, Electrónica y Computación, 2005. p. 1-27

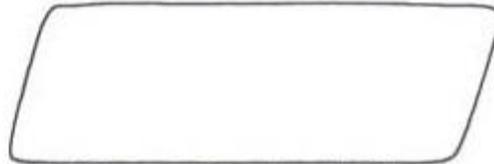
SALAS, Rodrigo. “Mapas Autorganizativos de kohonen (SOM)”. Internet: (http://www.inf.utfsm.cl/~rsalas/Pagina_Investigacion/docs/Apuntes/Redes%20SO M.pdf)

THETA ENTERPRISES. Bombeo mecánico Optimización. Traducido por Ing. Herdly Escalante y Tsu. Cherry Cardona. Bera, CA 92821-USA: Theta Enterprise, Inc, 2005. 292 p.

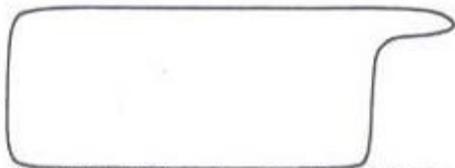
ANEXOS A. FORMAS DE CARTAS DINAGRAFICAS DE FONDO



Full Pump - Anchored Tubing



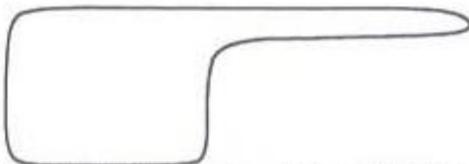
Full Pump - Unanchored Tubing
Or Tubing Anchor not Holding



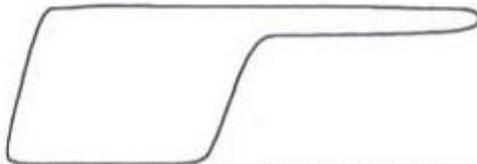
Slight Fluid Pound - Anchored Tubing



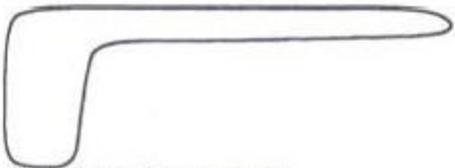
Slight Fluid Pound - Unanchored Tubing



Severe Fluid Pound - Anchored Tubing



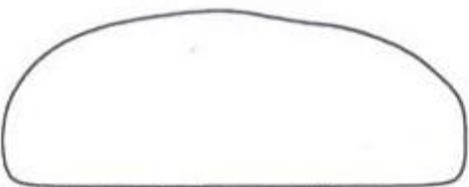
Severe Fluid Pound - Unanchored Tubing



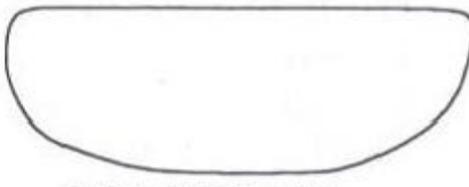
Completely Pumped off
Anchored Tubing



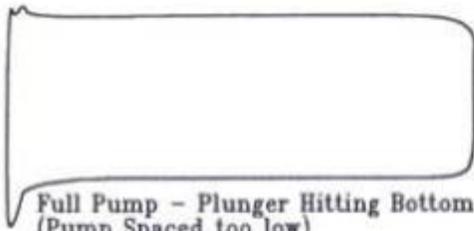
Malfunctioning Tubing Anchor



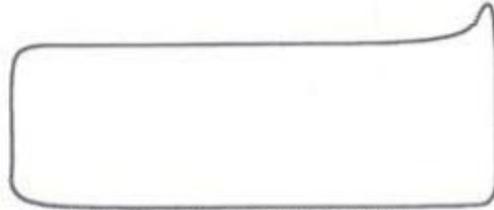
Leaking Traveling Valve or Plunger



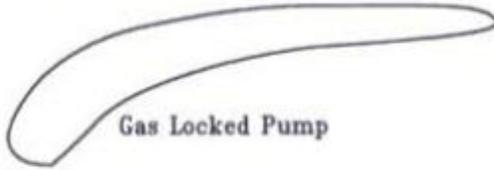
Leaking Standing Valve



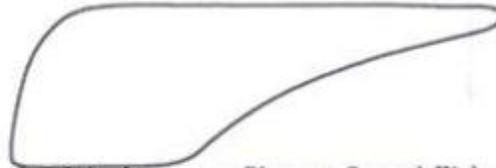
Full Pump - Plunger Hitting Bottom
(Pump Spaced too low)



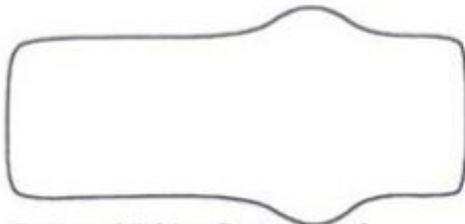
Full Pump - Plunger Hitting Up
(Pump Spaced too High)



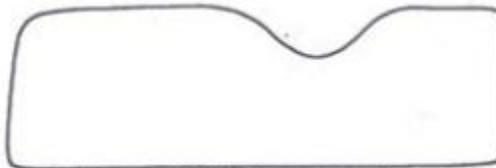
Gas Locked Pump



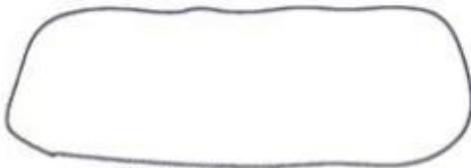
Gas Interference - Plunger Spaced High
From Bottom



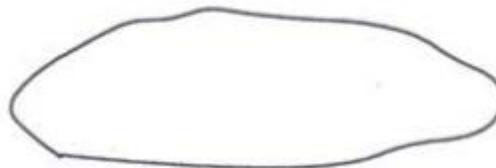
Bent or Sticking Pump Barrel



Worn or Split Pump Barrel



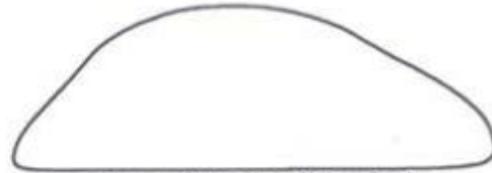
Slightly Worn Pump



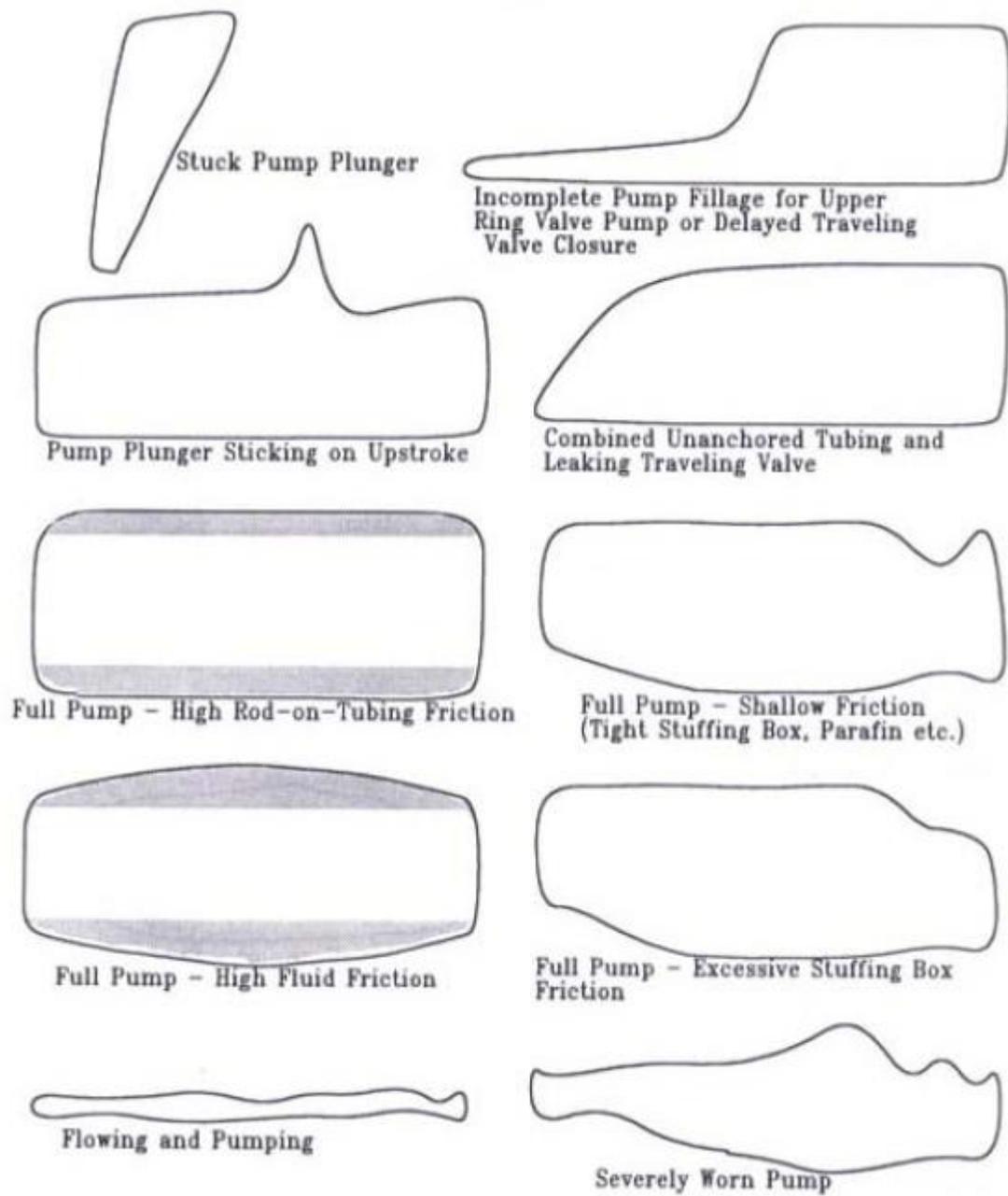
Worn Out Pump



Severly Worn Out Pump , Parted Rods
or Flowing Well



Severe Traveling Valve Leak

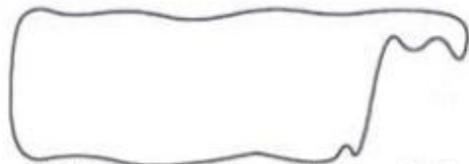




Fluid Pound - Wave Reflection from Sinker Bars



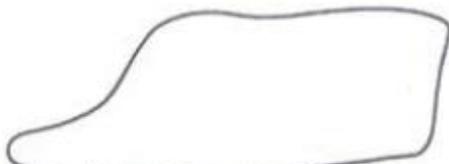
Fluid Pound - High Friction on Upstroke



Slight Fluid Pound - Fiberglass Rods



Full Pump - Fiberglass Rods

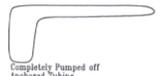
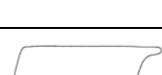
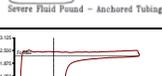


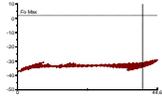
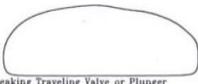
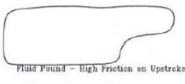
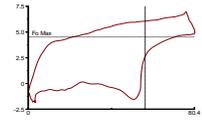
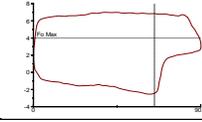
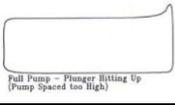
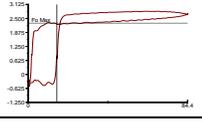
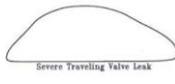
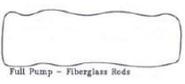
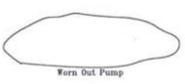
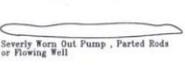
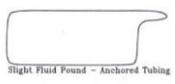
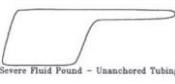
Delayed traveling valve closure caused by emulsified fluid in pump



Combined Fluid Pound and Delayed Traveling valve closure

ANEXO B. PRECLASIFICACION CUALITATIVA DE LA BASES DE DATOS PROPORCIONADAS POR ECOPETROL

CANTIDAD DE DINAGRAMAS POR POZO										
Clases	DINAGRAMA	POZOS								
		SC01	SC01S	SC02W	SC03S	SC04SD	SC05	BRI01	BRI09	Total
1	 Combined Fluid Pound and Delayed Travelling valve closure	1								1
2	 Completely Pumped off Anchored Tubing	2	19	3			7			31
3	 Slightly Worn Pump	6		1	1	1				9
4	 Slight Fluid Pound - Unanchored Tubing	3	3		41	12	5	22	63	149
5	 Full Pump - Unanchored Tubing or Tubing Anchor not Holding	4		2	9	9	17	13	4	58
6	 Full Pump - Plunger Hitting Bottom (Pump Spaced too low)	1	1			1				3
7	 Pressure vs. Time graph showing fluid pound		3	1	2					6
8	 Gas Interference - Plunger Spaced High From Bottom		4		12				13	29
9	 Severe Fluid Pound - Anchored Tubing		53		1		20	15		89
10	 Pressure vs. Time graph showing severe fluid pound		10	1	3	4	9			27
11 ruido(atipico)	 Pressure vs. Time graph showing noise		1				1			2

CANTIDAD DE DINAGRAMAS POR POZO										
Clases	DINAGRAMA	POZOS								
		SC01	SC01S	SC02W	SC03S	SC04SD	SC05	BRI01	BRI09	Total
11 ruido(atipico)			1				1			2
13	 Leaking Traveling Valve or Plunger			20						20
14	 Fluid Found - High Friction on Upstroke			6		41				47
16				8						8
17				6						6
18	 Full Pump - Plunger Hitting Up (Pump Spaced too High)					3				3
19							14			14
20	 Severe traveling Valve Leak						4			4
21	 Full Pump - Fiberglass Rods							1		1
22	 Worn Out Pump							18		18
23	 Severly Worn Out Pump , Parted Rods or Flowing Well								2	2
24	 Slight Fluid Found - Anchored Tubing							17		17
25	 Severe Fluid Found - Unanchored Tubing							15		15