



CIATEC

**SISTEMA EXPERTO BASADO EN CONOCIMIENTO
PARA EL RECONOCIMIENTO DE PATRONES DE
DESCARGAS PARCIALES EN REDES ELÉCTRICAS
DE ALTA TENSIÓN.**

Tesis

QUE PARA OBTENER EL GRADO ACADÉMICO DE

Doctor en Ciencia y Tecnología

En la Especialidad de

Ingeniería Industrial y de Manufactura

PRESENTA:

Rubén Jaramillo Vacio

DIRECTOR DE TESIS:

DR. CARLOS ALBERTO OCHOA ZEZZATTI

CO-DIRECTOR DE TESIS:

DR. ARMANDO RÍOS LIRA



CIENCIA Y TECNOLOGIA

León, Guanajuato, septiembre de 2015

AGRADECIMIENTOS

RESUMEN

En el presente trabajo se presenta un novedoso Sistema Basado en Conocimiento (SBC) para resolver el problema de reconocimiento y clasificación de patrones de descargas parciales (DP's). La investigación fue aplicada bajo el respaldo del Laboratorio de Pruebas a Equipos y Materiales de la Comisión Federal de Electricidad (LAPEM-CFE), esto como parte del compromiso de la institución de desarrollar técnicas innovadoras de evaluación y diagnóstico para aseguramiento de la calidad de los productos que se adquieren y se instalan en la red eléctrica de la CFE.

El sistema está basado en el proceso de Descubrimiento del Conocimiento en Bases de Datos. Este proceso cuenta con tres etapas clave: el pre-procesamiento de los datos (orden, limpieza y reducción de dimensionalidad) fue realizado con un Mapa Autoorganizado (SOM), la etapa de clasificación por medio de una Red Neuronal Probabilística (RNP) y la etapa de retroalimentación del repositorio para robustecer el clasificador siempre que se presente un nuevo caso.

Una de las contribuciones más importantes de este trabajo es el estudio de las regiones de operabilidad del SOM, demostrando que altamente no lineal y la convergencia depende de los datos de entrada y el algoritmo de aprendizaje competitivo, por lo que la sintonización óptima del SOM se desarrolló utilizando el método de la función de deseabilidad.

En el contexto validar el SBC se crearon datos experimentales en condiciones controladas, con el fin de tener un marco referencial y de entrenamiento de los algoritmos implementados en el sistema. Estos datos permitieron estudiar atributos de las DP's para establecer la mejor metodología de clasificación de fuentes en la misma muestra. Se propone como alternativa de salida dos tipos de gráfico, la matriz de distancia unificada y el 3D-SOM, en los cuales se puede observar los clústeres de interés y dar la primera noción de diagnóstico. El dominio de aplicación fue en la evaluación en sitio de redes subterránea alta tensión y de equipos eléctricos que requerían conocer el nivel de DP's en el diseño, así como también se valida el desempeño del SBC ante datos que contienen error obteniendo mejores resultados que otras técnicas.

Este SBC ofrece aplicación para evaluación y diagnóstico en fábricas de producción o instalaciones operativas, ya que es robusto a ruido ambiental y muy eficiente para separación y clasificación de fuentes DP's en aislamientos eléctricos de tipo gas, líquido, sólido y mixto.

ABSTRACT

In this thesis is presented a novel Knowledge Based System (KBS) to solve the problem of pattern recognition and classification of partial discharge (PD's). The research was applied on the Laboratory Tests on Equipment and Materials of the Federal Electricity Commission (LAPEM-CFE), this as part of the commitment of the institution to develop innovative techniques of assessment and diagnosis for quality assurance of purchased and installed products in CFE.

The system is based on the process of Knowledge Discovery in Databases. This process has three key stages: the pre-processing of data (order, cleanliness and dimensionality reduction) was performed with a Self Organizing Map (SOM), the classification stage by a Probabilistic Neural Network (PNN) and Repository feedback stage to strengthen the classifier whenever there is a new case.

One of the most important contributions of this work is the study of the regions of operability of SOM, showing that highly nonlinear and convergence depends on the input data and the competitive learning algorithm, so that the optimum tuning of the SOM was developed using the desirability function method.

To validate of KBS, the experimental data was created under controlled conditions, in order to have a benchmark and training algorithms implemented in the system. These data allowed the study attributes PD's to establish the best methodology for classifying sources in the same sample. The application of KBS was in the on-site evaluation of high voltage underground networks and electrical equipment required to meet the level of PD's in the design, and performance of KBS is also validated against data containing error.

The KBS provides application on evaluation and diagnosis in production facilities or operating facilities, as it is robust to environmental noise and very efficient separation and classification of PD's sources DP's sources of gas, liquid, solid and mixed electrical insulation.

INDICE

AGRADECIMIENTOS	i
.....	ii
RESUMEN.....	iii
ABSTRACT	iv
Índice de figuras	vii
Índice de Tablas	ix
CAPÍTULO 1: INTRODUCCIÓN	1
1.1 Antecedentes	1
1.2 Descripción del problema	6
1.3 Objetivos	8
1.3.1 Objetivo General	8
1.3.2 Objetivos Particulares	9
1.4 Hipótesis	9
1.5 Motivación.....	9
CAPÍTULO 2: FUNDAMENTO TEÓRICO Y ANÁLISIS CRÍTICO DE LA LITERATURA.....	11
2.1 Descargas parciales: Fundamentos	11
.....	15
2.3 Descargas parciales y Minería de Datos	17
CAPÍTULO 3: ALGORITMOS DE APRENDIZAJE NO SUPERVISADO Y SUPERVISADO	20
3.1 Mapas Autoorganizados (SOM).	20
3.2 Calidad del SOM.....	23
3.3 Mecanismos de aprendizaje del SOM.....	24
3.3.1 Aprendizaje competitivo: El ganador toma todo (Winner take all).....	25
3.3.2 Aprendizaje competitivo: sensible a la frecuencia (Frequency sensitive).	26
3.3.3 Aprendizaje competitivo: El rival penalizado (Rival penalized).	27
3.4 Proceso de implementación del SOM.	27
3.5 Descubrimiento Conocimiento en Bases de Datos	28
3.5.1 Etapas del KDD	30
3.5.2 Minería de datos	31
3.6 Red neuronal probabilística (RNP).....	31
CAPÍTULO 4: GENERACIÓN DE DATOS EXPERIMENTALES EN LABORATORIO Y EN SITIO.....	34
4.1 Generación de datos en laboratorio.	34
4.1.1 Desarrollo de banco de pruebas experimental	38
4.2 Obtención de datos en Sitio.....	39

4.3 Pre-tratamiento de datos usando SOM	46
4.4 Aplicación de Métricas de Calidad en el entrenamiento del SOM	48
4.5 Comparación de Aprendizaje Competitivo en Entrenamiento del SOM	49
4.6 Visualización de datos utilizando SOM en 3D.....	51
CAPÍTULO 5: EVALUACIÓN DEL DESEMPEÑO DE ALGORITMOS COMPETITIVOS EN EL SOM UTILIZANDO METODOLOGÍA DE SUPERFICIE DE RESPUESTA.	53
5.1 Diseño y análisis del experimento.	53
5.2 Análisis de varianza.....	57
5.3 Optimización por Superficie de Respuesta	59
CAPÍTULO 6: DESARROLLO DE METODOLOGÍA PROPUESTA	64
6.1 Descripción y aplicación de la interfaz usuario del SBC.	66
6.2 Evaluación del sistema en casos de estudio.....	70
6.3 Validación de datos que contienen error	74
CAPÍTULO 7: CONCLUSIONES.....	76
7.1 Contribuciones de este trabajo.....	77
7.2 Trabajos futuros	77
CAPÍTULO 8: PRODUCCIÓN CIENTIFICA.....	79
8.1 Artículos en Journal Citation Report (JCR)	79
8.2 Congresos Internacionales.....	79
8.3 Congresos Internacionales con estricto arbitraje	79
8.4 Revistas en Línea	80
8.5 Capítulos de Libro.....	80
8.6 Ponencias en Congreso o Academia	81
REFERENCIAS	83

Índice de figuras

Figura 1.- Cadena de Valor de CFE.....	2
Figura 2.- Procesos Clave del LAPEM en la Calidad de Bienes de CFE	3
Figura 3.- Red Subterránea.....	5
Figura 4.- Red Aérea	5
Figura 5.- Partes que conforman un cable	5
Figura 6.- Cable de Potencia.....	5
Figura 7.- Instalación de un cable subterráneo	6
Figura 8.- Pruebas preliminares de instalación	6
Figura 9.- Terminales para cable subterráneo	6
Figura 10.- Empalmes para cables subterráneos	6
Figura 11.- Equipo de pruebas resonante.....	7
Figura 12.- Medición de DP's en cable subterráneo con diferentes fuentes de contaminación por ruido.	8
Figura 13.- Proceso de evaluación de la conformidad de acuerdo al marco normativo.....	12
Figura 14.- Diagrama esquemático del inicio de DP en una cavidad.	13
Figura 15.- Arborescencia eléctrica en material aislante consecuencia de descargas parciales internas.	13
Figura 16.- Comportamiento de los pulsos de DP en un ciclo de tensión.	14
Figura 17.- Esquema del inicio y evolución de una DP en una cavidad confinada por aislamiento.....	14
Figura 18.- Diagrama conceptual de SBC propuesto para diagnóstico de DP's.	19
Figura 19.- Arquitectura del SOM.....	20
Figura 20.- Diagrama de flujo del SOM.....	22
Figura 21.- Representación de la salida del SOM en U-matrix.....	23
Figura 22 .- Etapas del KDD	29
Figura 23 Arquitectura de una red neuronal probabilística	32
Figura 24.- Defectos artificiales, a) Descarga parcial interna con electrodos planos, b) Electrodos punta-plano para generación de Corona o DP's externas y c) Electrodo plano con material aislante para generar corriente fuga y DP's superficial	34
Figura 25.- Esquema de conexión para generación de datos en laboratorio	35
Figura 26.- Vector de datos en grafica X-Y (Magnitud-fase) de medición de Descarga Externa (Efecto Corona).....	36
Figura 27.- Vector de datos en grafica X-Y (Magnitud-fase) de medición de Descarga Interna	36
Figura 28.- Vector de datos en grafica X-Y (Magnitud-fase) de medición de Descarga Superficial.	37
Figura 29.- Plataforma experimental para medición de DP en sistemas subterráneos.....	38
Figura 30.- Muestra prototipo para medición de DP's en laboratorio.	39

Figura 31.- Sistema de pruebas móvil resonante con medición de DP's a cables subterráneos.	40
Figura 32.- Diagrama conceptual de Sistema móvil resonante (etapa de potencia).	41
Figura 33.- Ejemplo grafico de cancelación de reactancias en prueba de alta tensión a frecuencia de resonancia.	41
Figura 34.- Ejemplo de prueba con equipo resonante en un sistema de cable de 115 kV ubicado en un poste troncocónico.	42
Figura 35.- Ejemplo de prueba con equipo resonante en un sistema de cable de 115 kV ubicado en una subestación.	42
Figura 36.- Diagrama conceptual de la plataforma de trabajo experimental	43
Figura 37.- Pantalla principal de sistema de adquisición de datos.	46
Figura 38 Determinación de U-matrix para DP Externa.....	47
Figura 39.- Determinación de U-matrix para DP Interna	47
Figura 40.- Determinación de U-matrix para DP Superficial.....	47
Figura 41.- Error de cuantización y topológico por época de entrenamiento (Descarga Interna)	50
Figura 42.- Error de cuantización y topológico por época de entrenamiento (Descarga Superficial) ...	50
Figura 43.- Error de cuantización y topológico por época de entrenamiento (Descarga Externa).....	50
Figura 44.- Representación de SOM en U-matrix y en 3D (Descarga interna)	52
Figura 45.- Representación del SOM en U-matrix y en 3D (Efecto Corona)	52
Figura 46.- Representación del SOM en U-matrix y 3D (Descarga Superficial).....	52
Figura 47.- Vector de Datos y Vector Prototipo (Descarga Superficial).....	56
Figura 48.- Vector de Datos y Vector Prototipo (Descarga Interna)	56
Figura 49.- Vector de Datos y Vector Prototipo (Descarga Externa)	56
Figura 50.- Valores actuales y predichos para la respuesta <i>te</i>	58
Figura 51.- Valores actuales y predichos para la respuesta <i>time</i>	58
Figura 52.- Valores actuales y predichos para la respuesta <i>qe</i>	59
Figura 53.- Grafico de superficie de respuesta de la combinación de efectos para <i>qe</i>	60
Figura 54.- Grafico de superficie de respuesta de la combinación de efectos para <i>te</i>	60
Figura 55.- Grafico de superficie de respuesta de la combinación de efectos para <i>time</i>	60
Figura 56.- Gráfica de deseabilidad para optimización de múltiples respuestas.....	61
Figura 57.- Identificación de las etapas en el diagrama conceptual del SBC.....	64
Figura 58.- Interfaz usuario del SBC.....	67
Figura 59.- Validación del SBC con el vector de descarga interna.....	69
Figura 60.- Validación del SBC con el vector de descarga superficial	69
Figura 61.- Validación del SBC utilizando el vector de descarga externa o efecto corona	70
Figura 62.- Medición obtenida en la fase B de una línea de 115 kV en sitio.	71
Figura 63.- Disección de un empalme de 115 kV con presencia de descarga superficial	71
Figura 64.- Evidencia de deficiente instalación de accesorios dentro del empalme	72

Figura 65.- Acometida de media tensión en un tablero de potencia..... 72
Figura 66.- Mediciones obtenidas en el tablero de media tensión del centro bancario..... 73
Figura 67.- Evidencia de defecto de instalación en terminal de media tensión..... 74

Índice de Tablas

Tabla 1.- Características Técnico-económicas de las redes subterráneas.	4
Tabla 2.- Diagnostico en Descargas Parciales utilizando Inteligencia Artificial	18
Tabla 3.- Resultados de medición en muestra de fabricantes	38
Tabla 4.- Calidad del SOM	48
Tabla 5.- Parámetros de entrenamiento	49
Tabla 6.- Desempeño del algoritmos competitivos en el SOM	51
Tabla 7.- Rango experimental y niveles de variables numéricas.....	54
Tabla 8.- Rango experimental y niveles de variables categóricas	54
Tabla 9.- Matriz de tratamientos y resultados	55
Tabla 10.- Resultados de la ANOVA.....	57
Tabla 11.- Solución para diferentes combinaciones de niveles de factor numérico y categórico	62
Tabla 12.- Matriz de confusión para validación del SBC	68
Tabla 13.- Evaluación comparativa de diferentes métodos de reconocimiento y clasificación de patrones de DP's	75

CAPÍTULO 1: INTRODUCCIÓN

1.1 Antecedentes

Una de las necesidades más apremiantes es desarrollar metodologías y herramientas que ayuden a los ingenieros de todos los procesos productivos a aprender e interpretar los datos de sus procesos con el fin de tomar decisiones. En la mayoría de las empresas el análisis de datos está basado en la selección manual de variables y extracción de los datos. Esto requiere mucho tiempo y tiende al sesgo de acuerdo al nivel de experiencia que tengan los ingenieros de proceso.

Existen muchas herramientas estadísticas que apoyan al análisis en línea (recolección de datos en procesos automatizados) y fuera de línea (recolección de datos manualmente), que se convierten en datos llenos de información, que con ella pueden establecer controles estadísticos de procesos.

Es importante minimizar el error aleatorio en la toma de datos de un proceso ya que debe ser vaciado en un repositorio, de manera que los algoritmos de análisis sean lo más eficientes y confiables para ver si algo útil sale de ellos, sin embargo, siempre es importante encontrarle coherencia a la información y validar lo interpretado. Ese proceso se vuelve más complejo cuando las muestra de datos que se toman son de gran tamaño y orden, aquí es donde juega un papel importante la experiencia del analista.

Una de las alternativas más viables para interpretación de la información es a través del reconocimiento de patrones y rasgo de la información, ya que con ello se identifica, caracteriza, clasifica y reconstruye comportamientos de procesos importantes para la toma de decisiones [1] [2] [3].

Existen diferentes enfoques para el reconocimiento de patrones, entre ellos destacan:

- **Geométrico** (Clustering): Los patrones deben ser graficables, en éste enfoque se emplea el cálculo de distancias, geometría de formas, vectores numéricos, puntos de atracción, etc.
- **Estadístico**: Se basa en la teoría de la probabilidad y la estadística, utiliza análisis de varianzas, covarianzas, dispersión, distribución, etc.
- **Sintáctico-Estructural**: estudia la estructura de los objetos, es decir, usa teoría de lenguajes formales, gramáticas, teoría de autómatas, etc.
- **Neuro-Reticular**: Se utilizan redes neuronales que se ‘entrenan’ para dar una cierta respuesta ante determinados valores.
- **Lógico-Combinatorio**: se basa en la idea de que la modelación del problema debe ser lo más cercana posible a la realidad del mismo, sin hacer suposiciones que no estén fundamentadas.

Se utiliza para conjuntos difusos y utiliza lógica simbólica, circuitos combinatoriales y secuenciales, etc.

- **Sistemas híbridos:** Es actualmente donde se encuentra el estado del arte, ya que a través de la combinación de los desempeños de los algoritmos, se le establecen objetivos en etapas del análisis, obteniendo resultados mejorados que con las aplicación de las técnicas independientes

El Laboratorio de Pruebas de Equipos y Materiales (LAPEM) es el brazo tecnológico de la Comisión Federal de Electricidad (CFE), responsable de asegurar la calidad en la adquisición de sus equipos principales y la confiabilidad en la operación de sus instalaciones estratégicas, líder en México y Latinoamérica por la especialización de su personal y la capacidad de su infraestructura eléctrica en extra-alta, alta tensión y alta corriente.

El LAPEM es un área estratégica para la Comisión Federal de Electricidad, ya que a través de los servicios de pruebas, análisis, diagnósticos y estudios de comportamiento de los equipos y materiales de la CFE, contribuye a incrementar la confiabilidad, disponibilidad, continuidad y calidad del servicio que proporciona la CFE a sus clientes.

La infraestructura tecnológica del LAPEM está constituida por los siguientes laboratorios para análisis y desarrollo de pruebas a equipos y materiales utilizados en el sector eléctrico: Extra Alta Tensión, Alta Tensión y Media Tensión, Alta Potencia en Media y Baja Tensión, Metrología, Química Analítica, Electrónica, Pruebas a Relevadores de Protección, así como Laboratorios Móviles para Pruebas en Centrales Eléctricas.

En la Figura 1 muestra la cadena de valor de CFE y la Figura 2 descompone en un diagrama los procesos clave que desarrolla cada uno de los laboratorios del LAPEM de acuerdo a su especialidad.

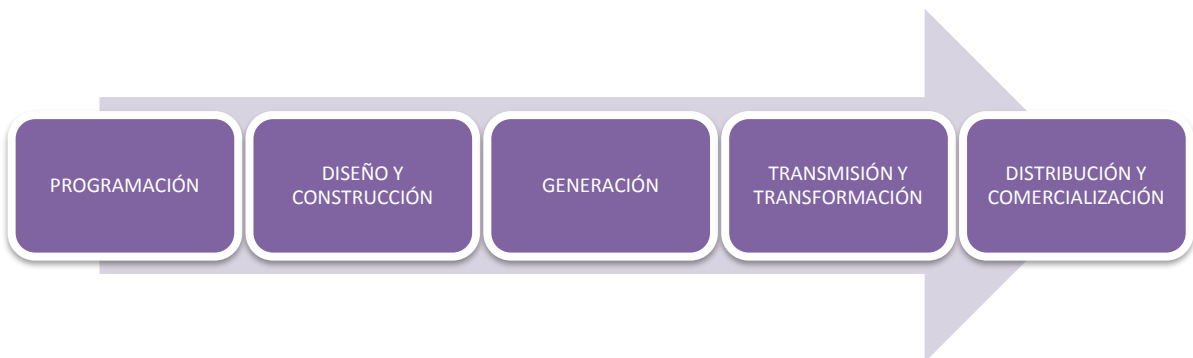


Figura 1.- Cadena de Valor de CFE.



Figura 2.- Procesos Clave del LAPEM en la Calidad de Bienes de CFE

El tamaño de información que se va generando al día con día en cada una de las funciones de LAPEM para cada uno de los procesos clave, hace imperativo el desarrollo de sistema de análisis de información basado en conocimiento.

El trabajo de investigación que actualmente se desarrolla impactará en los procesos clave de:

Normalización de productos y equipos: ya que establecerá métricas de aceptación y rechazo para productos considerando nuevas filosofías de análisis que se someterán al sistema normativo para su consideración.

Aprobación de prototipos: Se establecerán sistemas de decisión para que el proceso de evaluación del prototipo sea confiable y estándar, no basado en la experiencia del analista.

Inspección de bienes: El sistema propuesto debe ser robusto para ser utilizado tanto en laboratorio como en campo para la inspección de bienes que se entregaran a CFE.

Comportamiento de Equipos y Productos en Campo: Una de las tendencias de la CFE es la Gestión de Activos, en donde con el trabajo propuesto se obtendrán análisis de la condición operativa de los equipos y productos en campo y que formará una base de datos para seguimiento y monitoreo.

Una de las actividades más importantes de toda la cadena de valor de la CFE es la transmisión de energía desde su generación hasta su utilización, es por ello que el aislamiento es vital tanto para la seguridad de Sistema Eléctrico Nacional (SEN) como para la seguridad del personal.

En el Simposio Internacional de Aislamiento Eléctrico en Vancouver en 2008, se concluyó que el aislamiento eléctrico es la parte más débil de la mayoría de los aparatos eléctricos, tanto en media tensión como en alta tensión, los equipos susceptibles a degradación son: Transformadores, interruptores, maquinas rotatorias, líneas aéreas y desde luego cables para líneas subterráneas. Los principales factores de degradación que en su mayoría son eléctricos, térmicos, mecánicos y ambientales, provocan la pérdida de capacidad para soportar condiciones operativas.

Confiabilidad, mantenibilidad y disponibilidad de los activos del SEN implica, por lo tanto, el conocimiento de los procesos de degradación y envejecimiento del aislamiento eléctrico, el comportamiento de los aislantes a diversos factores operativos (frecuencia de operación, temperatura, sismos, etc.), de propiedades y metodologías de diagnóstico, de nuevas tecnologías de materiales (como hidrofobicidad, nano-compuestos, nano-estructurados, etc.).

Una de las actuales tendencias de la CFE es la transmisión y distribución subterránea, y como breve retrospectiva de las redes subterráneas en CFE, se tiene la visión de que la construcción de los proyectos tenga como características:

- Rentables (económicamente)
- Sencillas (técnicamente)
- Eficientes (operativamente)
- Ordenadas (con el entorno)

A continuación se presenta un análisis que demuestra la importancia de los proyectos de redes subterráneas:

Tabla 1.- Características Técnico-económicas de las redes subterráneas.

TIPO DE RED	CONTINUIDAD	MANTENIMIENTO	COSTO INVERSION INICIAL
SUBTERRANEA	+++	-	\$\$\$
HIBRIDA	++	++	\$\$
AEREA	-	+++	\$

Diferencia con respecto al entorno:

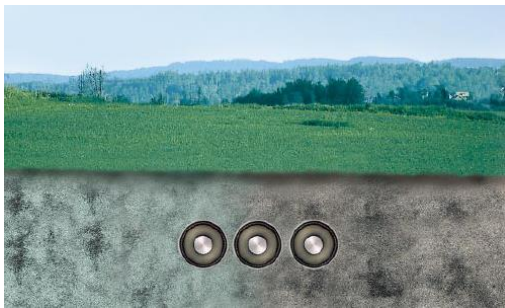


Figura 3.- Red Subterránea

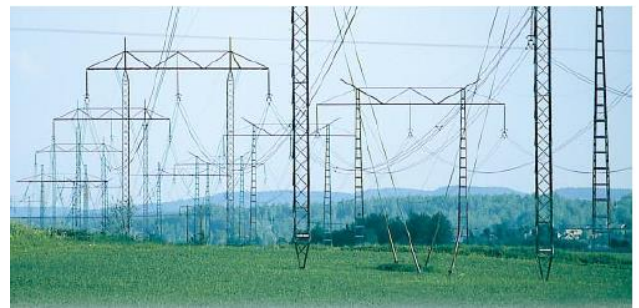


Figura 4.- Red Aérea

Los cables subterráneos son fabricados con un material denominado XLPE, que es un aislamiento polimérico de cadenas cruzadas que cumple con características dieléctricas que lo hace adecuado para aislar el conductor central. (Ver Figura 6)

Sin embargo, dentro del aislamiento, se presentan fenómenos de concentración y esfuerzos de campo eléctrico, los cuales deben ser uniformizados por capas semiconductoras que son colocadas en diferentes disposiciones del elemento aislante, así también es necesaria una capa metálica para drenar corrientes circulantes en el exterior del cable. Por estas causas, al unir un tramo de cable con otro, esa área de EMPALME, es un punto de probabilidad de falla, así como también las TERMINALES, las cuales al ser colocadas por el personal operario, se debe verificar su adecuada instalación. En general, una red subterránea como se mencionó anteriormente, consta de tres elementos básicos, que son TERMINALES, CABLES y EMPALMES.



Figura 6.- Cable de Potencia

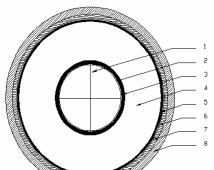
PIRELLI		Tipo de Cable		Conductor Unipolar Pantalla metálica de plomo	
Tensión	Calibre	Aislamiento	Gradiente Max		
133 / 230 kV (245 kV)	1000 mm ²	XLPE	8.5 kV/mm		
					
Item	Descripción	Espesor Nominal	Detalles		Diametro Aproximado
1	Conductor	-	4 sectores de cobre compactado		40.5
2	Cinta Separadora	-	Cinta Semiconductor		-
3	Semiconductor Interno	2.0	Polimero Semiconductor		45.0
4	Aislamiento	22.0	XLPE		90.0
5	Semiconductor Externo	1.0	Polimero Semiconductor		92.3
6	Cinta SC Anti Humedad	-	Cinta SC Anti Humedad		93.3
7	Pantalla Metálica	2.2	Aleación de Plomo		98.1
8	Cubierta Exterior	4.0	Pe Rojo		106.5
Peso Métrico (kg/m)					24.1

Figura 5.- Partes que conforman un cable



Figura 8.- Pruebas preliminares de instalación



Figura 7.- Instalación de un cable subterráneo



Figura 9.- Terminales para cable subterráneo



Figura 10.- Empalmes para cables subterráneos

1.2 Descripción del problema

Algunas las problemáticas detectadas en las instalaciones subterráneas son:

- Problemas incipientes en terminales y empalmes debido a defectos en su instalación (daños en los sistemas de alivio de esfuerzo, presencia de contaminación, competencia de los instaladores, entre otros)
- Manufactura e instalación del cable subterráneo (daños de las pantallas semiconductoras durante la fabricación, aislamiento no uniforme, excentricidad del conductor, transportación y almacenaje, entre otros)
- Envejecimiento después de algunos años de servicio (arborescencias eléctricas y arborescencias por humedad)

La utilización de las redes subterráneas se considera más segura, confiable y amigable con el entorno que los sistemas aéreos, pero para la puesta en servicio de instalaciones nuevas no se cuenta con un Sistema que nos asegure el estado del aislamiento y nos dé la “huella digital”, este mismo procedimiento se debe realizar para sistemas subterráneos que ya estén en operación, ya que cualquier falla implica un mayor tiempo de demora en el restablecimiento del servicio debido a que no se tiene acceso directo a la totalidad del sistema.

Actualmente el LAPEM como institución encargada de aseguramiento de la calidad de los suministros de la CFE, adquirió un Equipo Resonante con medición de descargas parciales (ver Figura 11) con un alcance suficiente para probar cables en 115, 230 y 400 kV, sin embargo, considerandos las técnicas de detección actuales para pruebas en campo, se requiere desarrollar una sistema de análisis de datos que permita identificar patrones del comportamiento y presencia de descargas parciales para emitir un diagnóstico integral de sistema terminales-cables-empalme, de manera que en caso de detectar alguna anomalía en el aislamiento poder ubicar el punto de falla y se pueda realizar la acción correctiva correspondiente.



Figura 11.- Equipo de pruebas resonante

La medición de descargas parciales es indudablemente la técnica más eficiente para determinación del estado del aislamiento en un laboratorio, en donde los niveles de interferencia electromagnética son controlados a través de filtros externos, sistemas de aterrizamiento y discriminadores por software, que permiten analizar cualquier indicio de ionización transitoria debida a presencia de descargas parciales, en laboratorio los valores de descargas parciales oscilan entre 0 y 2 pC. Sin embargo, para sistemas de medición de descargas parciales en sitio (subestaciones de transmisión y distribución, instalaciones de generación, redes instaladas cerca de líneas aéreas, etc.), en donde los niveles de interferencia electromagnética oscila en los 10000-50000 pC, es impráctico establecer criterios de medición de

descargas parciales en el orden de entre los 5-100 pC, ya que al aplicar filtros de señal se pierde información, así como también existe variación e inestabilidad de los datos, por lo que la identificación del fenómeno está basada en la experiencia del personal (empírico). Además el sistema de medición toma 100 muestra por ciclo, considerando que son mediciones de 1 minuto para la captura de datos, se tienen hasta 1 millón de datos agrupado en un vector de 4×10^6 . Así también uno de los grandes retos hoy es la clasificación de diferentes fuentes de DP's medidas en el mismo vector de datos.

Bajo este escenario se establece la siguiente pregunta ¿Existe un método que permita reconocer los patrones de datos de diversas fuentes en el mismo vector de DP's tomados en laboratorio y en sitio y clasificarlos para establecer un diagnóstico del sistema subterráneo?. La investigación de esta tesis se centra en **resolver el problema de clasificación de patrones de descargas parciales** proponiendo un sistema híbrido de mapas autoorganizados como pre-procesamiento y herramientas de minería de datos para la evaluación de la información.

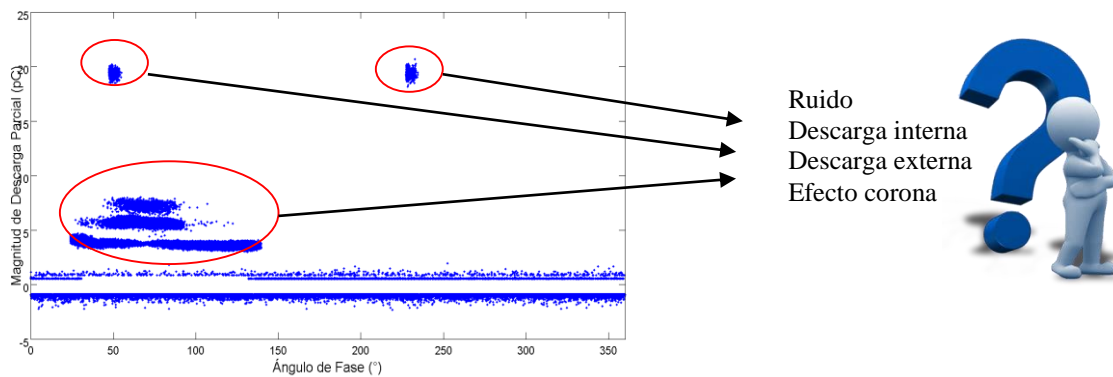


Figura 12.- Medición de DP's en cable subterráneo con diferentes fuentes de contaminación por ruido.

1.3 Objetivos

1.3.1 Objetivo General

Desarrollar una sistema de diagnóstico en redes subterráneas a través de la medición de descargas parciales (factor cuantitativo) detectando patrones comportamiento para su ubicación (factor cualitativo) utilizando técnicas heurísticas.

1.3.2 Objetivos Particulares

- Estudiar los atributos para diagnóstico de redes subterráneas utilizando la medición de descargas parciales.
- Utilizar de los mapas autoorganizado para pre-procesamiento del comportamiento de descargas parciales en cables y accesorios de las redes subterráneas durante las pruebas en sitio.
- Comparar herramientas de minería de datos para obtener la evaluación más eficiente de los patrones de datos de salida de los mapas autoorganizados.
- Estructurar los datos de salida en un formato de fácil interpretación a través de gráficos adecuados para la aplicación en los informes de pruebas del LAPEM-CFE.

1.4 Hipótesis

H1. Caracterizando patrones de DP's en laboratorio nos ayudará a extraer atributos para identificar patrones de defectos del sistema con condiciones de prueba de campo.

H2. El desarrollo de un sistema inteligente de identificación de patrones de DP's hará más eficiente la tarea de diagnóstico en lugares con condiciones adversas para la medición.

1.5 Motivación

En la actualidad, debido a las grandes ventajas que tienen las redes subterráneas, a nivel mundial, han aumentado su construcción en los proyectos de ampliación de la infraestructura eléctrica. En México existen cerca de 300 proyectos instalados en 115, 230 y 400 kV, lo equivalente a aproximadamente 10000 km lineales, 20000 empalmes y 6000 terminales tanto con aislamiento en aire como en SF6, con un crecimiento en sus instalaciones de aproximadamente el 30 % anual. Tan solo para 2015 se tiene proyectada la finalización de cerca de 7000 km de cables en diferentes proyectos con sus accesorios. Este crecimiento en México hace imperativo el contar con un sistema de diagnóstico tanto para los proyectos nuevos (pruebas de aceptación de la instalación), como para los sistemas ya instalados (pruebas de mantenimiento preventivo y predictivo).

Esta propuesta se ha tomado a nivel organizacional como un nuevo servicio que ofrece el LAPEM y que le da una ventaja competitiva en Latinoamérica, obteniendo un ingreso aproximado por servicio de

diagnóstico de \$700mil pesos. Considerando un costo de inversión de aproximadamente \$16millones de pesos que incluye el Sistema Resonante para pruebas en campo y el equipo de medición de Descargas Parciales.

Así también, de acuerdo a la norma IEC/ISO 17025, se debe desarrollar un método de validación para Unidades Métricas que no cuenten con patrones para trazabilidad de la medición, y con esta propuesta se obtendrán ahorros considerables en la calibración del sistema, ya que contando con los patrones de prueba en laboratorio, se puede establecer métodos para aseguramiento de la calidad de la medición por medio de verificaciones intermedias para mantener vigente la calibración del equipo de medición.

CAPÍTULO 2: FUNDAMENTO TEÓRICO Y ANÁLISIS CRÍTICO DE LA LITERATURA

2.1 Descargas parciales: Fundamentos.

Las características de calidad de un producto son variables aleatorias y el porcentaje de piezas defectuosas por no cumplir las especificaciones de calidad depende de la distribución de dichas variables aleatorias. Controlar el proceso de fabricación consiste en conseguir que la media μ y desviación típica σ de cada una de esas variables aleatorias sean constantes en el tiempo. La variabilidad que presenta una característica del producto puede deberse a causas comunes y de difícil eliminación, como pequeñas oscilaciones en la tensión eléctrica, heterogeneidad en la materia prima, desgaste de alguna máquina o herramienta, etc. o bien a causas de tipo puntual o accidental como una partida de materia prima defectuosa, un desajuste de una máquina, etc.

En el sector eléctrico existen productos de manufactura en serie como fusibles, interruptores, cables, transformadores de pequeña potencia, etc, o productos de fabricación especial como interruptores de potencia, tableros encapsulados, subestaciones blindadas, transformadores de potencia, etc. Los productos especiales son los que se deben de revisar e inspeccionar y probar en dos etapas: pruebas de fábrica y pruebas de aceptación.

Pruebas de Fábrica: Tienen como objeto verificar las características, condiciones de operación y la calidad de fabricación del equipo antes de ser entregados al cliente.

Pueden ser supervisadas por el cliente.

Pruebas de Aceptación: Son conocidas como las pruebas de las pruebas. Se realizan sobre equipos nuevos después de ser instalados y antes de ser energizados, también se debe efectuar en todo equipo que ha sido intervenido para realizar reparaciones mayores. Tienen como objeto lo siguiente:

- Verificar que el equipo cumple con lo especificado.
- Establecer referencias para pruebas futuras.
- Determinar si la instalación del equipo es la correcta.

Las pruebas de los equipos eléctricos tienen como fin verificar las condiciones en que estos se encuentran. Sirven para mantener la confiabilidad y continuidad del servicio eléctrico, estas pruebas son:

Pruebas de Mantenimiento Predictivo: Se ejecuta en intervalos regulares durante la vida útil del equipo. Su objetivo consiste en verificar si un equipo se encuentra en condiciones de operación adecuadas y detectar a tiempo fallas que pudieran afectar esta situación.

Pruebas Especiales de Mantenimiento: Se realizan cuando existen sospechas o certeza de que el equipo se encuentra en problemas. También se ejecutan sobre equipos sometidos a condiciones extremas.

Para que el LAPEM lleve a cabo el círculo indicado en Figura 2 se basa en normas nacionales e internacionales en el que se establecen los atributos de calidad y criterios de aceptación de cada uno de los productos. De acuerdo a la funcionalidad que se quiere validar se desarrolla procedimiento ilustrado en

Figura 13.

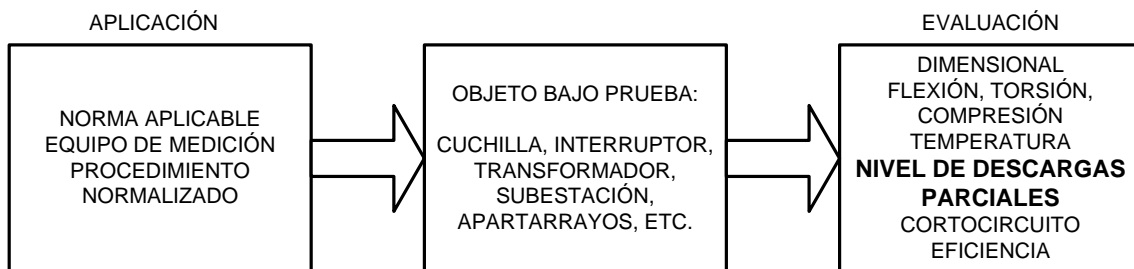


Figura 13.- Proceso de evaluación de la conformidad de acuerdo al marco normativo.

El atributo que se espera medir es el reflejo de la calidad de los materiales, procedimiento de ensamble, capacitación del personal involucrado en el proceso, diseño, capacidad de planta, etc. de esta manera el mensurando obtenido es la **calidad reflejada** del producto. Estos criterios aplican a cualquier tipo de prueba tanto en fábrica como en sitio, por lo que el personal que realice la evaluación debe contar con equipo confiable y los conocimientos necesarios para aplicar los criterios.

Uno de los retos científicos-tecnológicos está en la aplicación de los criterios de evaluación utilizando las descargas parciales [4] [5] [6].

La descarga parcial (PD) es un fenómeno que ocurre en aislamientos de alta tensión, está la definición internacional se da en la norma [7]. En general, las descargas parciales son una consecuencia del esfuerzo dieléctrico (campo eléctrico) local en el aislamiento o en la superficie del aislamiento. Este fenómeno tiene un efecto dañino en los equipos, por ejemplo, transformadores, generadores, cables de potencia, interruptores, entre otros. La medición de descargas parciales es una prueba no destructiva y puede ser utilizada para evaluar el estado del aislamiento desde la puesta en servicio (huella digital, *fingerprint* en inglés) y con ello se monitorea el sistema teniendo en cuenta la reducción de vida útil

durante el tiempo de operación, este envejecimiento depende de numerosos parámetros como los esfuerzos eléctricos, térmicos y mecánicos.

Existen dos condiciones para que ocurran descargas parciales en superficies, cavidades y defectos:

- Presencia de un electrón libre para iniciar una avalancha (ver Figura 14).
- La diferencia de potencial por campo eléctrico debe ser más alto que el potencial de inicio de ionización en el medio aislante.

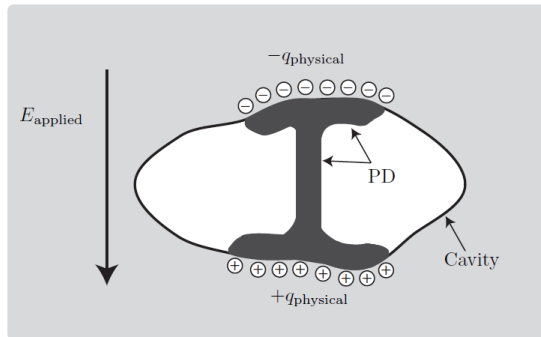


Figura 14.- Diagrama esquemático del inicio de DP en una cavidad.

Las descargas parciales son divididas en tres grupos de acuerdo a sus orígenes y características [8]:

- **Efecto Corona o Descarga Externa:** Ocurre en gases o líquidos causada por concentración de campo eléctrico en un electrodo en punta, cables delgados, aristas, etc.
- **Descarga interna:** Ocurre dentro de una cavidad que está rodeada completamente por material aislante, puede ser en forma de burbuja (ejemplo: burbujas en papel impregnado de aceite), delaminaciones (ejemplo: devanado de transformadores), fisuras (ejemplo: esfuerzos por fatiga). Arborescencias eléctricas son la consecuencia de descargas internas (Figura 15).
- **Descarga superficial:** Ocurre en la superficie de un aislamiento donde el esfuerzo tangencial es elevado (ejemplo: barras de generadores eléctricos)



Figura 15.- Arborescencia eléctrica en material aislante consecuencia de descargas parciales internas.

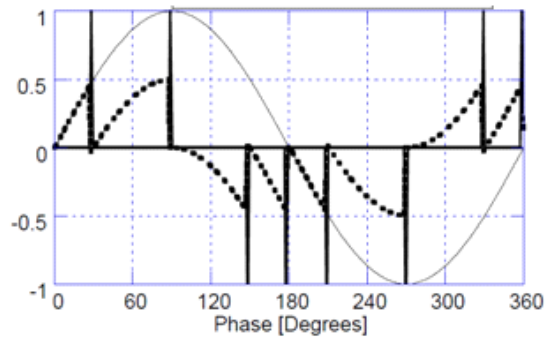


Figura 16.- Comportamiento de los pulsos de DP en un ciclo de tensión.

Para explicar cómo se presenta una DP interna durante un periodo de tensión a 60 Hz se presenta la Figura 16. Se observa que en el primero y segundo cuadrante correspondientes al semiciclo positivo, al iniciar el incremento de tensión en el ciclo también inicia el proceso de ionización en la cavidad, hasta que en determinado ángulo del ciclo ocurre la avalancha (primer cuadrante angular), esto depende del tamaño de la cavidad, como continúa incrementando el nivel de tensión comienza nuevamente la ionización hasta que ocurre otra avalancha, en el segundo cuadrante angular, nuevamente comienza la ionización hasta que se produce la descarga, el tercero y cuarto cuadrante que es el semiciclo negativo se producen los mismos fenómenos, pero en sentido inverso. Este comportamiento pulsante degrada irreversiblemente el medio que lo confina, provocando tracking y erosión en las superficies (cuando es efecto corona y descargas superficial) o el fenómeno de árboles (denominados así por el efecto de ramificaciones) en la descarga parcial interna (Figura 15).

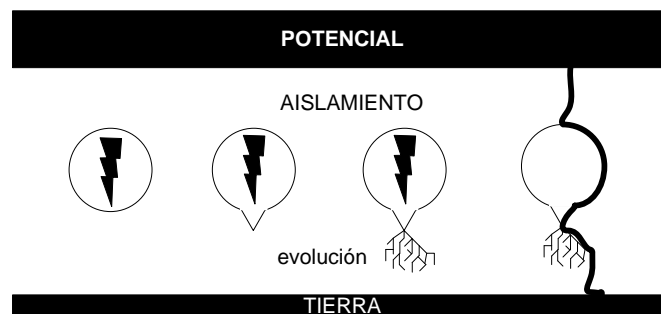


Figura 17.- Esquema del inicio y evolución de una DP en una cavidad confinada por aislamiento.

En la Figura 17 se presenta un bosquejo del fenómeno de una DP interna, en cada uno de los pulsos generados por efecto del campo eléctrico debido a la tensión, se va degradando gradualmente el medio

que confina la cavidad, es un proceso aleatorio en cuanto a su evolución, ya que depende de diversos factores físicos y operativos, uno de los más significativos es la temperatura del medio. Sin embargo, como ejemplo explicativo se describe que la evolución del daño de la DP va formando ramificaciones que disminuyen la rigidez dieléctrica del aislamiento hasta que se produce la falla, que por general es fase-tierra, aunque en el caso de algunos equipos eléctricos como los transformadores de potencia, en una cavidad confinada por aceite dentro de los devanados, el mayor esfuerzo dieléctrico es fase-fase por la misma configuración del sistema.

Como se observa, existen diferentes variables de influencia que pueden ser seleccionadas para el estudio del fenómeno de la DP's, entre las que se pueden mencionar la magnitud de la tensión, métricas estadísticas de la DP's (valores máximo-mínimo, media, mediana, desviación estándar, etc), ángulo de fase de la ocurrencia de DP, tasa de repetición de la DP, etc. Por ello, la primera aproximación a un diagnóstico es la selección de los diferentes atributos para clasificar las actividades de PD medidas y las fuentes que las generan [9].

2.3 Descargas parciales y Minería de Datos

Desde este punto de vista, la aplicación de herramientas para análisis de datos utilizando medición de Descargas Parciales han ayudado a estudiar patrones característicos del comportamiento del fenómeno, identificando potenciales daños, en este sentido en [17] se proponen los Mapas Autoorganizados para el reconocimiento de patrones de descargas parciales en transformadores de corriente encapsulados en resina, lo cual se convierte en la primera referencia de esta herramienta, potencializando su aplicación para el diagnóstico de estos fenómenos dieléctricos.

Los mapas autoorganizados (de sus siglas en inglés SOM) son una red neuronal no supervisada, se caracterizan porque en su entrenamiento no se presentan las salidas objetivo que se desean asociar con cada patrón de entrada [18] [19]. La red a partir del proceso de autoorganización, proporcionará cierto resultado, el cual será reflejo de las relaciones de similitud entre dichos patrones de entrada. Las principales aplicaciones de estos modelos son en la realización de agrupamientos de patrones, análisis exploratorio y minería de datos, además que el esquema de presentación de la información se puede adecuar a los requerimientos del usuario de la información [20] [21].

La minería de datos se refiere a la extracción del conocimiento de grandes bases de datos y en la actualidad se ha utilizado en aplicaciones de todas las ramas, desde las ciencias físicas, la informática, salud, ambiental [22] [23] [2], entre otras. La medición de descargas parciales genera información en una gran base de datos y de acuerdo de la frecuencia de muestreo del sistema, aparte de que es necesaria la toma de decisiones al momento de tener la información, sin embargo, para esa toma de decisiones ha sido necesaria la experiencia, la cual no se ha organizado ni transferido de manera sistemática, por lo que es necesario un sistema que apoye en el diagnóstico e interpretación de los datos. En [24] se presenta un estudio de consideraciones para diagnóstico en redes subterráneas a través de Clusterings Jerárquicos, enmarcando los diferentes parámetros de las señales de inyección y su análisis variando factores como magnitud, fase, etc. Este estudio fue realizado considerando condiciones de laboratorio controladas, por lo que abre la oportunidad de reproducirlo pero en condiciones de campo, en donde se deben de tomar en cuenta el filtrado de la señal para discriminación de ruido por interferencia electromagnética.

Desde el punto de vista del diagnóstico en equipo eléctrico utilizando la medición de descargas parciales, se han presentado resultados aplicando técnicas heurísticas, en [25] se presenta la aplicación de medición en línea en redes de media tensión, en [26] se presenta la aplicación el diagnóstico de subestaciones encapsuladas (GIS) utilizando momentos estadísticos y la organización de esa información con árboles de decisión. En [27] se aplica minería de datos utilizando mapas autoorganizados en transformadores de potencia, este artículo es muy importante, ya que a partir de

este análisis se agregará otro paso después de mapa autoorganizado para que el sistema ayude a tomar la decisión. En [5] se presenta un sistema de diagnóstico para redes subterráneas de alta tensión, el procesamiento de la información es utilizando redes neuronales y con un sistema computacional basado en la información del operador se emite el diagnóstico. Uno de los artículos más recientes en la materia es [25], que realiza un estudio comparativo entre SOM, SVM (Support Vector Machine) y BPN (Back Propagation Neural Network), obteniendo que SVM tiene ventajas importantes a considerar, como es el tiempo de procesamiento de los datos y la exactitud en la clasificación de los datos.

Tabla 2.- Diagnostico en Descargas Parciales utilizando Inteligencia Artificial

Autores	Herramienta y Objetivo	Restricciones
(Mazroua, 1993) (Krivda, 1995) (W. Yan, 2005) (G. C. Montanari, 2006) (K. X. Lai B. T., 2010)	Herramienta: Redes Neuronales Supervisadas. Objetivo: Reconocimiento entre diferentes fuentes formadas por cavidades cilíndricas	Reconocimiento de diferentes fuentes de DP's en la misma muestra.
(J. Kim, 2008)	Herramienta: Red Neuronal Difusa. Objetivo: Comparación entre BPN y FNN	Desempeño en caso de múltiples descargas y no se incluyen defectos externos y ruido en la medición.
(L. Ri-Cheng, 2008)	Herramienta: Optimización en Enjambre de Partículas (PSO). Objetivo: Localización de DP's en transformadores de potencia.	Mejorar la aplicación para mediciones en sitio.
(W. Y. Chang, 2008) (Babnik, 2007)	Herramienta: Mapas Autoorganizados (SOM). Objetivo: Reconocimiento de patrones y clasificación.	Calidad y optimización de la estructura del SOM.
(N. Fadilah-Ab-Aziz, 2007)	Herramienta: Maquina de Soporte Vectorial (SVM). Objetivo: Selección de Atributos y clasificación de DP's.	SVM no es eficiente para pequeños vectores de datos.
(H. Hirose, 2008)	Herramienta: Árbol de decisión Objetivo: Selección de Atributos y clasificación de DP's.	Las reglas de asignación son sensibles a pequeñas perturbaciones en el conjunto de datos (inestabilidad)

En Tabla 2 se presenta un resumen de los resultados obtenidos en el reconocimiento de patrones y extracción de atributos para diagnostico utilizando técnicas meta heurísticas o de inteligencia artificial.

El presente tema de investigación doctoral tiene por objetivo general la mejora en la clasificación de patrones considerando diferentes fuentes de DP's, al hacer esta clasificación es posible discriminar

ruido y otros atributos de las mediciones que sean de poco interés para el diagnóstico. Se realizará un estudio teórico-experimental de los atributos del SOM como son el tipo de aprendizaje competitivo y las métricas de calidad para mejorar el desempeño en el esquema general de Sistema Basado en Conocimiento propuesto (ver

Figura 18). En el capítulo 7 se desarrollará de manera extensa cada una de las etapas que lo conforma.

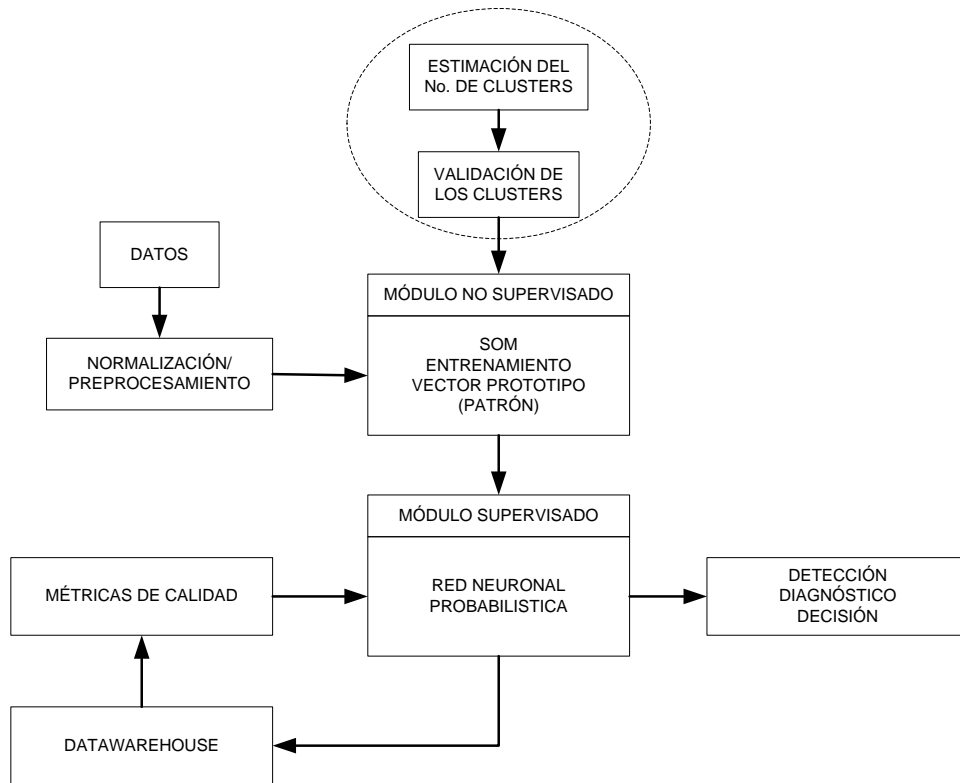


Figura 18.- Diagrama conceptual de SBC propuesto para diagnóstico de DP's.

CAPÍTULO 3: ALGORITMOS DE APRENDIZAJE NO SUPERVISADO Y SUPERVISADO

3.1 Mapas Autoorganizados (SOM).

Los SOM fueron desarrollados a lo largo de la década de los ochenta por el físico finlandés Teuvo Kohonen, como una continuación natural de la línea de desarrollo de las redes competitivas iniciada por Von der Malsburg. Aparte de su interés como una sencilla modelización de redes neuronales naturales, los SOM poseen un gran potencial de aplicabilidad práctica.

De entre las clases de problemas del mundo real en los que han demostrado su eficacia cabe citar: clasificación de patrones, cuantificación vectorial, reducción de dimensiones, extracción de rasgos y visualización. Por ejemplo, los SOM han sido empleados en reconocimiento del habla, control de robots, monitorización de procesos industriales, ayuda al diseño de circuitos integrados, reconocimiento de patrones financieros y minería de grandes bases de datos en Internet.

En este modelo, las neuronas se organizan en una arquitectura unidireccional de dos capas (ver Figura 19). La primera es la capa de entrada o sensorial, que consiste en m neuronas, una por cada variable de entrada distribuyendo la información procedente del espacio de entrada a las neuronas de la segunda capa. Las entradas son muestras estadísticas del espacio sensorial [18].

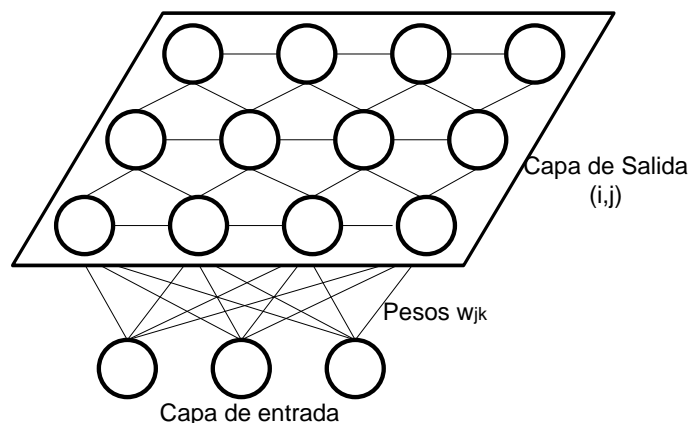


Figura 19.- Arquitectura del SOM.

El procesamiento se realiza en la segunda capa, que forma el mapa de rasgos, y consiste habitualmente en una estructura rectangular de $n \times n_y$ neuronas que operan en paralelo.

Aunque la arquitectura rectangular es la más común, a veces también se utilizan capas de una sola dimensión o de tres dimensiones.

Etiquetaremos las m neuronas de entrada con el índice k ($1 \leq k \leq m$), y las $n_x \times n_y$ neuronas del mapa con un par de índices $\mathbf{i} \equiv (i, j)$ ($1 \leq i \leq n_x, 1 \leq j \leq n_y$) que determinarán su localización espacial. Cada neurona de entrada k está conectada a todas las neuronas (i, j) del mapa mediante un peso sináptico w_{ijk} [28].

En resumen, el mapa puede describirse como una matriz de procesadores elementales (i, j) ordenados en dos dimensiones que almacenan un vector de pesos sinápticos o vector de referencia (codebook) $\mathbf{w}_{ij}(t)$, con:

$$\{w_{ij}(t) : w_{ij} \in \mathcal{R}^m, 1 \leq n_x, 1 \leq n_y\} \quad (1)$$

En la fase de ejecución (operación normal de la red), los pesos permanecen fijos. En primer lugar, cada neurona (i, j) calcula la similitud entre el vector de entradas, $\mathbf{x} \{x_k | 1 \leq k \leq m\}$ y su propio vector de pesos sinápticos \mathbf{w}_{ij} , según una cierta medida de distancia o criterio de similitud establecido.

Para nuestro estudio se considera la Distancia Euclídea entre x y w , dada por:

$$D(x, w_i) = \sqrt{\sum_k (w_{ik} - x_k)^2} \quad (2)$$

La neurona ganadora se le denomina BMU, y es cuando el valor x y w son los más cercanos, de manera que el vector de pesos en la BMU y su vecindad en la red son movidos de acuerdo a la siguiente regla:

$$\begin{aligned} \Delta w_{ij} &= \delta(c, i) \alpha(x_j - w_{ij}) \\ \Delta w_{ij} &= \alpha(x_j - w_{ij}) \text{ to } i = c \\ \Delta w_{ij} &= 0 \text{ to } i \neq c \end{aligned} \quad (3)$$

Donde c es la BMU, α es la tasa de aprendizaje y δ es la función de vecindad. El proceso se repite para numerosos patrones de entrada, de forma que al final los diferentes vectores de referencia sintonizan con dominios específicos de las variables de entrada. Dicho espacio está dividido en grupos, cada neurona se especializará en uno de ellos, y la operación esencial de la red se podrá interpretar entonces como un análisis clúster [29] [30]. Lo descrito hasta el momento responde a un esquema competitivo clásico, en el que cada neurona actúa en solitario. Sin embargo, el modelo de SOM aporta una importante novedad, pues incorpora a este esquema relaciones entre las neuronas próximas del mapa. Para ello introduce una función de vecindad δ , su efecto es que durante el aprendizaje se actualizan tanto los pesos BMU como los de las neuronas pertenecientes a su entorno [19].

Una función vecindad $V=V[n]$ (n = tiempo discreto) se puede definir de diferentes formas (cuadrada, circular, rombo, gaussiana o algún híbrido entre ellas). En realidad, esta función no tiene como único argumento el tiempo discreto "n", sino que también tiene como argumentos a la posición del nodo

ganador y la posición del nodo a actualizar. La función vecindad está centrada en el nodo ganador y su radio de influencia disminuye monótonamente a medida que avanza el entrenamiento [28].

De esta manera, en el modelo de SOM se logra que neuronas próximas sintonicen con patrones similares, quedando de esta manera reflejada sobre el mapa una cierta imagen del orden topológico presente en el espacio de entrada. En esencia, por medio del proceso descrito los SOM realizan la proyección no lineal de un espacio multidimensional de entrada \mathfrak{R}^m sobre un espacio discreto de salida, representada por la capa de neuronas. El mapa representa una imagen del espacio sensorial, pero de menor número de dimensiones, reflejando con mayor fidelidad aquellas dimensiones del espacio de entrada de mayor varianza (que suelen coincidir con los rasgos más importantes de las entradas).

La función vecindad representa matemáticamente de una forma sencilla el efecto global de las interacciones laterales existente entre las neuronas en el cerebro, pues en vez de considerar en detalle que una neurona trata de activar a sus vecinas y de inhibir a las alejadas (como sucede en el córtex), esta situación se modela mediante una sencilla función que define el tamaño de la vecindad en torno a la vencedora, dentro de la cual todas las neuronas son premiadas actualizando sus pesos, y fuera de ella son castigadas al no actualizar sus pesos o al hacerlo en sentido contrario.

La utilización de la función vecindad en el modelo de mapas autoorganizados aporta respecto del modelo competitivo sencillo dos ventajas adicionales: el ritmo efectivo de convergencia se mejora y el sistema es más robusto frente a variaciones en los valores iniciales de los pesos.

El algoritmo del SOM, en síntesis es el siguiente (ver Figura 20):

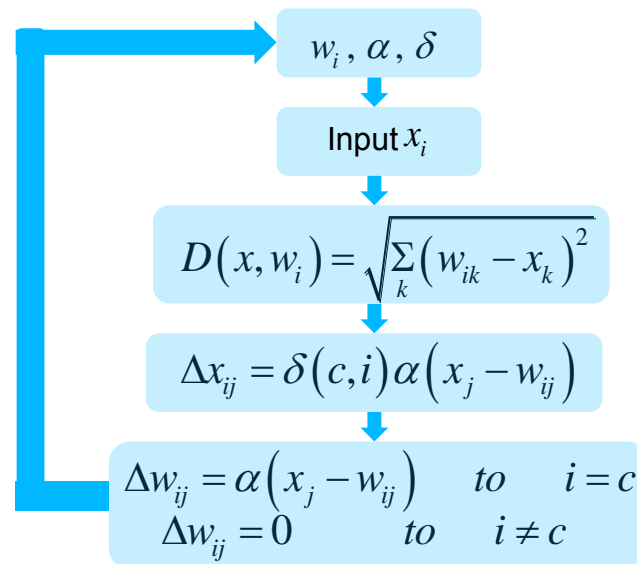


Figura 20.- Diagrama de flujo del SOM

El número de paneles de la capa de salida son acorde a $n \times ny$, esquematizado en un diagrama de salida denominado Matriz de Distancias Unificadas (U-matrix), donde su representación es en escala de color o escala de grises (ver Figura 21), la resolución se da por $(2nx-1) \times (2ny-1)$ dimensional [20].

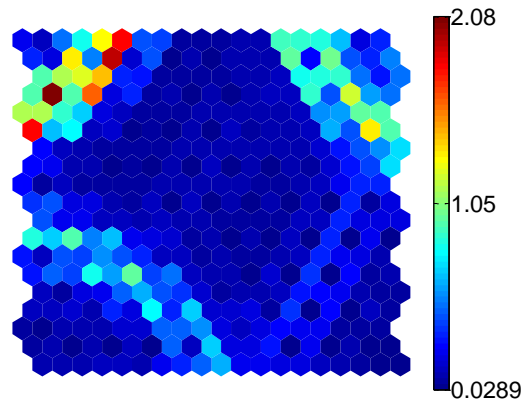


Figura 21.- Representación de la salida del SOM en U-matrix.

Como se observa en la Figura 21, cada hexágono nos da la posibilidad de establecer seis vecindades, el color de cada hexágono representa la distancia promedio entre el vector de referencia ubicado en la posición del mapa y los hexágonos vecinos. El color azul indica que el grupo de datos es cercano o similar y el rojo representa que el grupo de datos está lejano o es poco similar. De esta manera se pueden identificar los clústeres por grupo de datos.

3.2 Calidad del SOM

Para definir métricas de calidad del SOM es importante considerar el preservar la resolución y la topología, estas son las bases que establecen el desempeño adecuado del algoritmo [31]. El error de cuantización (q_e) y el error topológico (t_e) son calculados para medir la calidad del mapa definido por el vector prototipo. El error de cuantización es la distancia promedio entre cada vector de datos y su BMU, en otras palabras es la medida de resolución del mapa. La expresión está dada por:

$$q_e = \frac{1}{N} \sum \|x_i - c_{x_i}\| \tag{4}$$

Donde c_{x_i} denota la BMU de la muestra que es mapeada en el vector prototipo. Esto es, que tan exacta es la representación de los datos por el vector prototipo. Si es pequeño el error de cuantización, la media de la distancia del vector de datos al vector prototipo es pequeña, esto significa que los datos son cercanos a su prototipo [32].

Otra métrica de calidad del SOM desde un punto de vista de proyección y que considera la estructura del mapa, es el error topológico, este nos indica la comparación de la entrada con respecto a la salida, ubicando cada una de los datos en el mapa. El error topológico es la proporción de todo el vector de datos donde cada primera y segunda BMU son no adyacentes, de otra manera es considerada una violación de la topología y es penalizada incrementando el valor del error [32]. Formalmente, esto se escribe como:

$$t_e = \frac{1}{N} \sum_{k=1}^N u(x_k) \quad (5)$$

Donde N es el número de datos y:

$$u(x) = \begin{cases} 0 & \text{si la BMU de } x \text{ es adyacente a la segunda BMU de } x \\ 1 & \text{cualquier otra} \end{cases} \quad (6)$$

3.3 Mecanismos de aprendizaje del SOM

El aprendizaje de esta red neuronal consiste solo en aplicar a la red neuronal los datos de entrada, sin indicarle la salida esperada. La red neuronal es capaz de reconocer algún tipo de estructura en el conjunto de datos de entrada (normalmente redundancia de datos) y de esta forma se produce el autoaprendizaje. Durante el proceso de aprendizaje la red autoorganizada debe descubrir por sí misma rasgos comunes, regularidades, correlaciones o categorías en los datos de entrada, e incorporarlos a su estructura interna de conexiones. Se dice, por tanto, que las neuronas deben autoorganizarse en función de los estímulos (datos) procedentes del exterior. En cuanto a los algoritmos de aprendizaje no supervisado, en general se suelen considerar dos tipos, que dan lugar a los siguientes aprendizajes [19]:

- Aprendizaje hebbiano.

Esta regla de aprendizaje es la base de muchas otras, la cual pretende medir la familiaridad o extraer características de los datos de entrada. El fundamento es una suposición bastante simple: si dos neuronas n_i y n_j toman el mismo estado simultáneamente (ambas activas o ambas inactivas), el peso de la conexión entre ambas se incrementa. Las entradas y salidas permitidas a la neurona son: $\{-1, 1\}$ o $\{0, 1\}$ (neuronas binarias). Esto puede explicarse porque la regla de aprendizaje de Hebb se originó a partir de la neurona biológica clásica, que solamente puede tener dos estados: activa o inactiva.

- Aprendizaje competitivo.

En la red neuronal tiene lugar una competición entre las neuronas de forma que la neurona que gana la competición se activará, las demás neuronas no se activan. La competición la gana la neurona cuyo vector de pesos sea más parecido al vector entrada.

El aprendizaje competitivo es una eficiente herramienta para *clustering* de datos (el termino en español es clusterización, sin embargo por fines de adecuar al lenguaje internacional se utilizará el término en ingles *clustering*), universalmente utilizado en problemas de procesamiento de datos tales como compresión, clasificación, procesamiento de imágenes, reconocimiento de patrones, etc.

El resultado del aprendizaje competitivo en el caso del SOM es una partición del conjunto de datos de entrada inducida por la distribución de los datos en las neuronas. Ésta partición se realiza de manera que datos similares son agrupados por la red y representados por una sola neurona. Dicha neurona es la unidad ganadora para cada uno de los datos asociados durante la última iteración en el proceso de entrenamiento. Por lo tanto, la agrupación de los datos es realizada de manera automática, basándose en la similitud entre los datos y en la distribución de las respectivas neuronas ganadoras localizadas a lo largo y ancho de una retícula bidimensional.

Para encontrar el SOM que cumpla con la eficiencia necesaria para resolver el problema de clasificación, es importante evaluar los algoritmos de aprendizaje competitivos propuestos en la literatura, ya que cada uno tiene ventajas y desventajas, de manera que se cumpla con un adecuado tiempo de procesamiento y minimizando los errores de cuantización y topológicos propios del entrenamiento.

3.3.1 Aprendizaje competitivo: El ganador toma todo (Winner take all).

Es el aprendizaje competitivo más simple, se encuentra una neurona ganadora y se actualizan los pesos únicamente de esa neurona para hacer que tenga más posibilidades para ganar la próxima vez que un vector de entrada similar sea mostrado a la red. De esta forma se produce su especialización en el reconocimiento de este tipo de entradas al activarse únicamente ante una entrada similar.

Se puede presentar un problema en este tipo de aprendizaje: algunas neuronas pueden llegar a dominar el proceso y ganar siempre mientras otras neuronas no ganan nunca (neuronas muertas). Para evitar esto nos aseguramos de que los pesos son actualizados y normalizados en cada iteración.

Este aprendizaje se suele presentar en las neuronas de la capa de salida, a este tipo de neuronas se las llama El-Ganador-Toma-Todo de sus siglas en ingles WTA (winner-take-all).

Se produce una organización en función de la estructura de los datos de la entrada. Este método intenta asegurar que los elementos que pertenecen al mismo grupo sean lo más similares posible entre sí y lo más diferentes posibles a elementos de otros grupos.

El objetivo de este tipo de aprendizaje es categorizar conjuntos de datos estudiando su estructura y clasificándolo en grupos [33].

3.3.2 Aprendizaje competitivo: sensible a la frecuencia (Frequency sensitive).

El algoritmo de K-means tiene un problema denominado “unidades muertas” (dead units) [34], lo que significa que si un centroide es inapropiadamente elegido, nunca podrá ser actualizado y nunca podrá representar una clase o cluster. Para resolver el problema de las “unidades muertas” se ha desarrollado el algoritmo de aprendizaje competitivo “sensible a la frecuencia” de sus siglas en ingles FSCL (frequency sensitive competitive learning) [35]. Cada centro cuenta el número de ocasiones en las que ha ganado la competencia y en consecuencia reduce su velocidad de aprendizaje. Si un centro ha ganado con demasiada frecuencia "se siente culpable" y se retira de la competición. El algoritmo FSCL es una extensión del algoritmo de k-mean, que se obtiene de acuerdo con el siguiente:

$$j = \arg \min \gamma_i \|x(n) - c_i(n)\| \quad i = 1, \dots, N \quad (7)$$

Donde n es el número de entradas, N representa el número de centros, la frecuencia relativa de competiciones ganadas está dada por γ_i , definida como:

$$\gamma_i = \frac{s_i}{\sum_{i=1}^n s_i} \quad (8)$$

Donde s_i es el número de veces donde el centros c_i fue declarado ganador. Así que los centros que han ganado la competición en los últimos tienen una probabilidad reducida de ganar otra vez, proporcional a su frecuencia γ_i . Después de seleccionar el ganador, el algoritmo FSCL actualiza al ganador con la siguiente ecuación:

$$c_i(n+1) = c_i(n) - \eta [x(n) - c_i(n)] \quad (9)$$

η es la tasa de aprendizaje, de la misma manera que el algoritmo k-means, y el ajuste correspondiente es considerando la siguiente ecuación:

$$s_i(n+1) = s_i(n) + 1 \quad (10)$$

3.3.3 Aprendizaje competitivo: El rival penalizado (Rival penalized).

El algoritmo competitivo del rival penalizado de sus siglas en ingles RPCL (Rival penalized), desempeña apropiadamente el clustering sin conocimiento del número de clústers [36]. Este algoritmo no solo determina el centro ganador sino que penaliza el segundo centro ganador denominado rival, esto es descrito de la siguiente manera:

$$r = \arg \min \gamma_i \|x(n) - c_i(n)\|, \quad i = 1, \dots, N \quad i \neq j \quad (11)$$

El rival se moverá alejándose del centro de la entrada x con una taza β , llamada taza de desaprendizaje. Los demás centros no se moverán. Esta consideración es expresada como:

$$c_i(n+1) = \begin{cases} c_i(n) + \eta [x(n) - c_i(n)] & \text{if } i = j \\ c_i(n) - \beta [x(n) - c_i(n)] & \text{if } i = j \\ c_i(n) & \text{if } i \neq j \text{ and } i \neq r \end{cases} \quad (12)$$

Si la velocidad de aprendizaje η se escoge mucho mayor que β , con al menos el doble de magnitud, el número de datos de salida se encuentran automáticamente. Supongamos, el número de clases es desconocido y el número de centros N es mayor que el número de clústers, el número de centros convergerá hacia los centros de las clases de los datos de entrada. El RPCL alejará el rival en cada iteración.

3.4 Proceso de implementación del SOM.

De la revisión bibliográfica del estado del arte se encontraron trabajos que demostraban los algoritmos más eficientes del SOM, entre ellos, los de [18] [37] [28] [17] [21] [38] [39]. Tomado como base esos trabajos se desarrolló el algoritmo en MATLAB® R2010a con el fin de tener flexibilidad para su adecuación y mejora, ya que si se manipulan los "Toolbox" que contiene la actual versión de MATLAB® R2010a se puede llegar generar errores. Para el desarrollo del SOM se consideró las siguientes etapas:

- **Inicialización del SOM (número de neuronas, taza de aprendizaje, numero de vecindades por tipo de celda, función de vecindad, etc.):** Esta etapa es importante porque se establecen los atributos del SOM para su configuración inicial para
- **Lectura de datos en archivos *.CSV y *.TXT, normalización en base a la varianza de los datos para mejorar convergencia numérica:** Se considera que el sistema de medición cuenta

con la generación y denominación de datos en formato importable para su almacenamiento en el repositorio provisional y después poder leer en el espacio de trabajo de MATLAB® R2010a.

- **Grafica XY de los vectores Magnitud-Fase:** En esta etapa se grafica los datos con los atributos más significativos para el diagnóstico, esta grafica es solo para comparar el vector de datos original con la U-matrix.
- **Grafica de vector prototipo en U-matrix en escala de color o escala de grises, en 2D o 3D:** Esta es la etapa más importante de la red, ya que su representación muestra los patrones de salida basándose en la similaridad.
- **Representación de componentes planos del vector de datos:** La representación en componentes planos de cada una de las variables del vector de datos nos da la representación visual de la correlación entre ellas.

Ejemplo de código en MATLAB® R2013a para lectura, entrenamiento y visualización del SOM:

<pre> %% Inicializar Workspace clc clear all %% Cargar datos data=load ('SP_GND_4.txt'); %Corona punta a tierra %data=load ('Interna02.txt'); %Descarga interna %% Pre-tratamiento data(:,1)=[]; data(:,3)=[]; phi=data(:,1); pc=data(:,2); %% Grafico de Mediciones figure(1) plot(phi,pc, '.'); %title('Mediciones de Descargas Parciales'); xlabel('Phase (°)'); ylabel('Partial Discharge (pC)'); %% SOM sD = som_set(data); sD = som_normalize(sD,'var'); %SD = som_normalize(sD,'var'); %% Grafica SOM sM = som_make(sD,'msize', [10 10],'hexa'); %% Visualizacion figure(2) colormap(1-gray)%Para graficas in gray scale som_show(sM,'umat','all','norm','d'); %som_show_add('textSize',8,'textColor','r'); U=som_umat(sM); Um=U(1:2:size(U,1),1:2:size(U,2)); </pre>	<pre> %% Matriz de Distancias figure(3) [~,V,me,l] = pcaproj(sD,2); Pm = pcaproj(sM,V,me); % PC-proyeccion Code = som_colorcode(Pm); % color hits = som_hits(sM,sD); % hits U = som_umat(sM); % U-matrix Dm = U(1:2:size(U,1),1:2:size(U,2)); % matrix de distancia Dm = 1-Dm(:)/max(Dm(:)); Dm(hits==0) = 0; % clustering info som_grid(sM,'Label',cellstr(int2str(hits)),'Line','none', 'Marker','none','Labelcolor','k'); %som_grid(sM,'Label',cellstr(int2str(hits)),'Line','none', 'Marker','none','Labelcolor','k','Surf'); %hold on som_cplane(sM,Code,Dm); title('Codigo de Colores + Matriz de Distancias') %% Datos asociados al cluster figure(4) [~,V,me,l] = pcaproj(sD,2); Pm = pcaproj(sM,V,me); % PC-proyeccion Code = som_colorcode(Pm); % color hits = som_hits(sM,sD); % hits U = som_umat(sM); % U-matrix Dm = U(1:2:size(U,1),1:2:size(U,2)); % matrix de distancia Dm = 1-Dm(:)/max(Dm(:)); Dm(hits==0) = 0; % clustering som_cplane(sM,Code,Dm); som_grid(sM,'Label',cellstr(int2str(hits)),'Line','none', 'Marker','none','Labelcolor','k'); title('Datos asociados a los clusters') </pre>
--	---

3.5 Descubrimiento Conocimiento en Bases de Datos

Por medio del algoritmo SOM, una red neuronal aprende a representar visualmente un conjunto de datos y las relaciones entre sus variables, gracias a que la proyección preserva la topología en el conjunto de datos. El desarrollo de sistemas de software que implementan este tipo de herramientas

tiene distintos enfoques, uno de los más interesantes es el de aplicaciones computacionales que tienen como objetivo el descubrimiento de conocimiento a partir del procesamiento masivo de información contenida en grandes bases de datos. Este objetivo se alcanza por medio de la concreción de un proceso conocido como **Descubrimiento Conocimiento en Bases de Datos (KDD de las siglas en ingles de Knowledge Discovery on Databases)**. La denominada Minería de Datos (DM) es una etapa dentro del proceso KDD. En esta etapa se lleva a cabo la aplicación de distintas herramientas y métodos de análisis inteligente de datos, por medio de los cuales es posible obtener, de manera automática, información útil que una vez validada se acepta como conocimiento. El proceso de KDD está definido en términos del objetivo que su nombre indica, el cual se logra como consecuencia del cumplimiento de una serie de etapas (ver Figura 22).

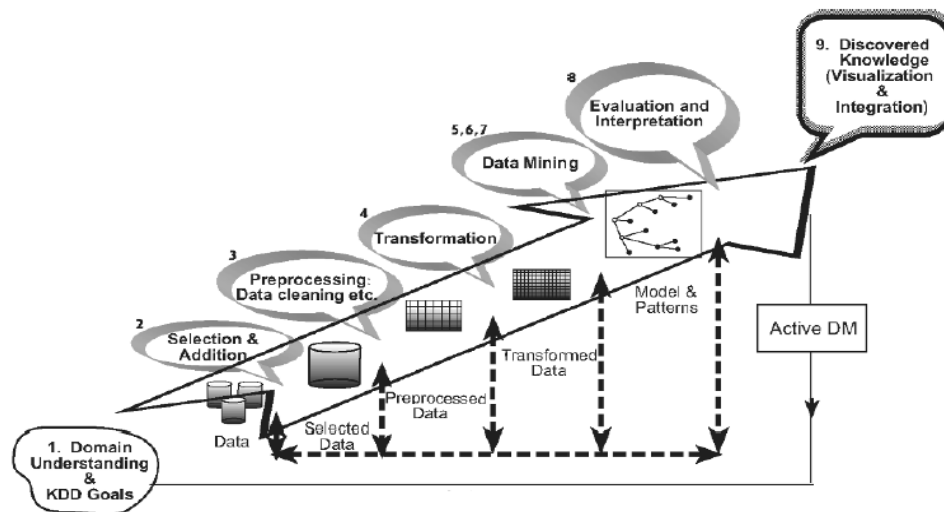


Figura 22 .- Etapas del KDD

El proceso de KDD tiene como meta principal identificar patrones o establecer modelos; válidos, nuevos, útiles y entendibles; a partir del procesamiento de grandes bases de datos. Se trata de un proceso no-trivial ya que el cómputo involucrado no es simple y por lo tanto en muchas ocasiones se requiere la aplicación de técnicas computacionales de alto desempeño.

Varios autores [16] [29] [40] [2] coinciden en la consideración de que el proceso KDD implica el cumplimiento de una serie de etapas en las cuales se involucra:

- El pre-procesamiento y la preparación de los datos.
- La búsqueda de patrones o modelos.
- La evaluación del conocimiento.

3.5.1 Etapas del KDD

Antes de iniciar el proceso es necesario desarrollar un entendimiento del dominio de la aplicación, así como contar con un conocimiento relevante previo; de manera que sea posible identificar la meta del proceso KDD desde el punto de vista del beneficiario. Entender el dominio de los datos es un prerequisite para extraer cualquier tipo de información útil. La sucesión de etapas que constituyen el proceso KDD a continuación se describen brevemente [2]:

- 1 **Crear un conjunto de datos objetivo:** Es decir, seleccionar un conjunto de datos o enfocarse en un subconjunto de variables de las muestras, para llevar a cabo el análisis y el descubrimiento.
- 2 **Limpieza de datos y pre-procesamiento:** Operaciones básicas como limpieza del ruido si es requerido; normalización de los campos que contiene cada dato; implementación del modelo de representación; decidir estrategias para el manejo de campos faltantes en los datos, entre otras tareas. Se ha considerado que esta etapa puede tomar más del 80% del tiempo total del proceso.
- 3 **Transformaciones:** En los casos en los que es posible usar métodos de reducción de la dimensión o métodos de transformación para reducir el número de variables a considerar y encontrar características útiles para representar los datos dependiendo de la meta del proceso.
- 4 **Selección del algoritmo de minería de datos:** Esto incluye decidir qué modelos y parámetros pueden ser apropiados (por ejemplo, los métodos para datos categóricos son distintos que los métodos para vectores de componentes reales) y empatar el método particular de minería de datos con el criterio global del proceso.
- 5 **Minería de datos:** Buscar patrones de interés en una forma representativa o un conjunto de estas representaciones. En algunos círculos esta etapa es considerada como todo el proceso. En la minería de datos se pueden usar muchas técnicas de la estadística y aprendizaje de máquina, tales como regla de aprendizaje, inducción por árbol de decisión, clustering, programación lógica inductiva, etc. El énfasis de la investigación de minería radica en el descubrimiento de patrones simples y entendibles.
- 6 **Interpretación de los patrones minados:** El proceso de KDD no se detiene cuando los patrones han sido descubiertos. El usuario debe entender que ha sido descubierto, para lograr esto se pueden llevar a cabo tareas tales como: la selección u ordenamiento de patrones, la visualización

de los patrones extraídos o la visualización de los datos dados los modelos extraídos, etc. El proceso de KDD es necesariamente iterativo: los resultados de la minería pueden mostrar que algunos cambios deben ser hechos en la conformación del conjunto de datos y entonces será necesario volver al principio. Muchos enfoques del proceso KDD ponen mucho énfasis en la interpretación o post-procesamiento.

- 7 **Consolidación de conocimiento descubierto:** La incorporación en otro sistema del conocimiento obtenido para futuras acciones o simplemente documentar y reportar el conocimiento obtenido. Esto también incluye buscar o resolver conflictos entre el conocimiento previo y el extraído.

3.5.2 Minería de datos

El componente de minería de datos en el proceso de KDD comúnmente involucra la aplicación repetida e iterativa de métodos del análisis inteligente de datos. En general, la mayoría de los métodos de minería de datos están basados en técnicas de calidad probada, provenientes de disciplinas como el aprendizaje de máquina, el reconocimiento de patrones y la estadística [2].

Los métodos de minería de datos pueden ser clasificados de acuerdo a las tareas específicas que desempeñan. A continuación, se listan las metas básicas y se describe brevemente en qué consisten:

- **Clasificación:** Determinar una función que asigna a cada dato una o varias clases predeterminadas.
- **Regresión:** Determinar una función que representa el comportamiento de alguna porción del conjunto de datos o el descubrimiento de relaciones funcionales entre variables.
- **Sumarización:** Encontrar una descripción compacta para un subconjunto de datos, es decir, extraer un sumario o reglas de asociación y aplicar técnicas de visualización multivariada.
- **Clustering:** identificar un conjunto finito de clases que describan los datos
- **Modelación de Dependencia:** Encontrar un modelo que describa dependencias significativas entre las variables.
- **Cambio y Detección de Desviación:** Descubrir los cambios más significativos en los datos a partir de medidas previas o valores normativos.

3.6 Red neuronal probabilística (RNP)

Las redes neuronales probabilísticas (RNP) pueden ser utilizadas para problemas de clasificación de datos. A diferencia del proceso utilizado para problemas de regresión, en el cual se realiza ajuste de

pesos y biases, de acuerdo al error presentado, en el proceso de clasificación no se realiza ningún ajuste de pesos y sólo los patrones de salida son determinados mediante la comparación y el cálculo de distancias.

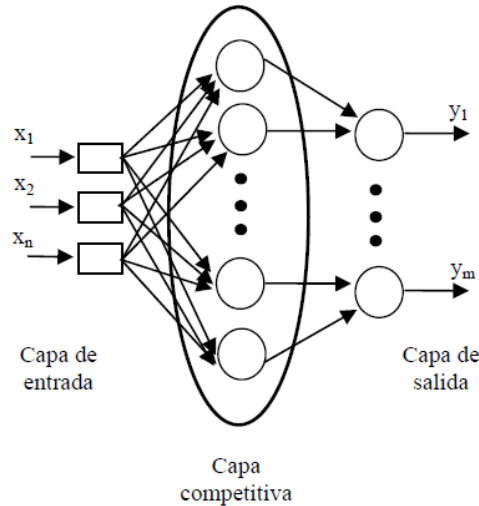


Figura 23 Arquitectura de una red neuronal probabilística

En la Figura 23 se presenta la arquitectura de la red neuronal probabilística. Cuando se le presenta a la red una entrada, la primera capa se ocupa de calcular las distancias desde el vector de entrada a los vectores de entrada entrenados y produce un vector cuyos elementos indican como está de cerca la entrada respecto a la entrada entrenada.

La función de la segunda capa es sumar las contribuciones para cada clase de entrada y producir con ellas, un vector de salida con las probabilidades. Finalmente, una función de transferencia en la salida de la segunda capa del tipo “competitiva”, selecciona la máxima de estas probabilidades y produce un 1 para esa clase y cero para las demás.

El algoritmo fue desarrollado considerando el estimador de densidad de probabilidades de Parzen, la base teórica indica que para un grupo de datos dados $X = \{x_1, x_2, \dots, x_n\}$, el estimador de densidad de Parzen es:

$$p(x) = \frac{1}{N\sigma^d} \sum_{n=1}^N G \left[\frac{\|x-x_n\|}{\sigma} \right] \tag{13}$$

Donde $x_n \in R^d$, G es la función kernel y σ es el factor de escala. El kernel es la función de tipo gaussiana:

$$G(x) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{x^2}{2}} \quad (14)$$

Basado en el condicional de la función de densidad de probabilidad (FDP) estimada, una muestra dada x puede ser clasificada como una clase i si $p(x|i) > p(x|j)$ para todas las clases $j \neq i$. En el algoritmo propuesto, la entrada de la RNP es un cluster del SOM (promedio de la FDP para n_i muestras en la i^{th} población), para esta aplicación en particular el kernel es:

$$G(x) = \frac{1}{(2\pi)^{p/2} \sigma^p n_i} \sum_{k=1}^{n_i} e^{-\frac{\|x-x_{ik}\|^2}{2\sigma^2}} \quad (15)$$

Diferentes valores de factor de escala σ conducen a diferentes actuaciones de clasificación. En primer lugar, la puntuación de desempeño del clasificador con un determinado σ se determina por el método de validación cruzada. En el proceso, cada muestra de entrenamiento se omite temporalmente del conjunto de entrenamiento y se utiliza como muestra de prueba. El conjunto de datos de entrenamiento restante se utiliza en la RNP para clasificación. Si la muestra se clasifica correctamente, la puntuación de rendimiento se incrementa en 1. Por último, una unidimensional búsqueda heurística se realiza para encontrar el óptimo σ con la puntuación más grande de desempeño.

CAPÍTULO 4: GENERACIÓN DE DATOS EXPERIMENTALES EN LABORATORIO Y EN SITIO

El procedimiento empleado para desarrollar esta investigación doctoral consta de diversas etapas estructuradas de acuerdo al cronograma de actividades, cada actividad encaminada a construir el conocimiento de manera sistémica. Algunas etapas están concluidas, sin embargo, de acuerdo a los trabajos sometidos a los congresos se han mejorado o aumentado con el fin de dar un soporte metodológico sólido.

4.1 Generación de datos en laboratorio.

La generación de datos en laboratorio es un muy importante, que gracias a que se tiene control de las variables de influencia, se pueden obtener datos para entrenar el SOM y analizar los atributos importantes para diagnóstico. Este estudio establecerá los umbrales requeridos para la identificación de los patrones de pruebas en sitio y que serán utilizados en el sistema híbrido propuesto para la evaluación. Para la generación de los datos en el laboratorio se consideraron las geometrías básicas de defectos y sus variantes (ver Figura 24). En la Figura 25 se presenta un diagrama de conexión para la medición de datos en las muestras con diferentes tipos de descarga parcial.

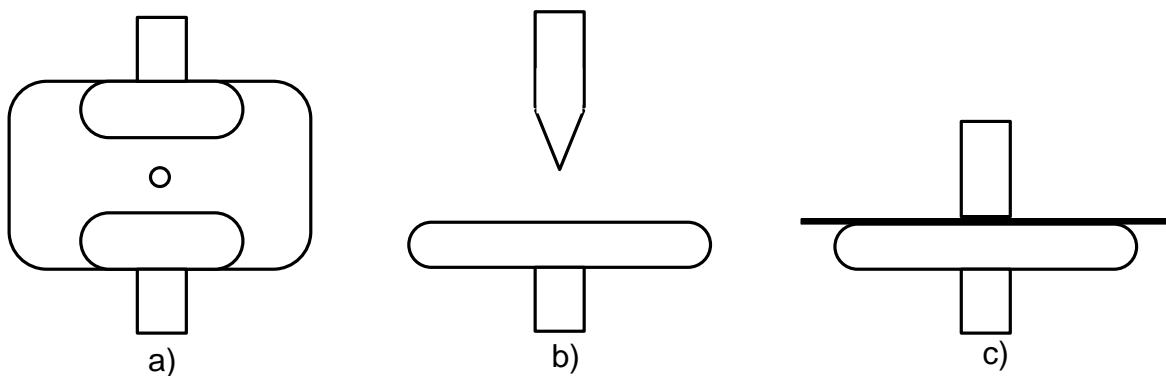


Figura 24.- Defectos artificiales, a) Descarga parcial interna con electrodos planos, b) Electrodo punta-plano para generación de Corona o DP's externas y c) Electrodo plano con material aislante para generar corriente fuga y DP's superficial

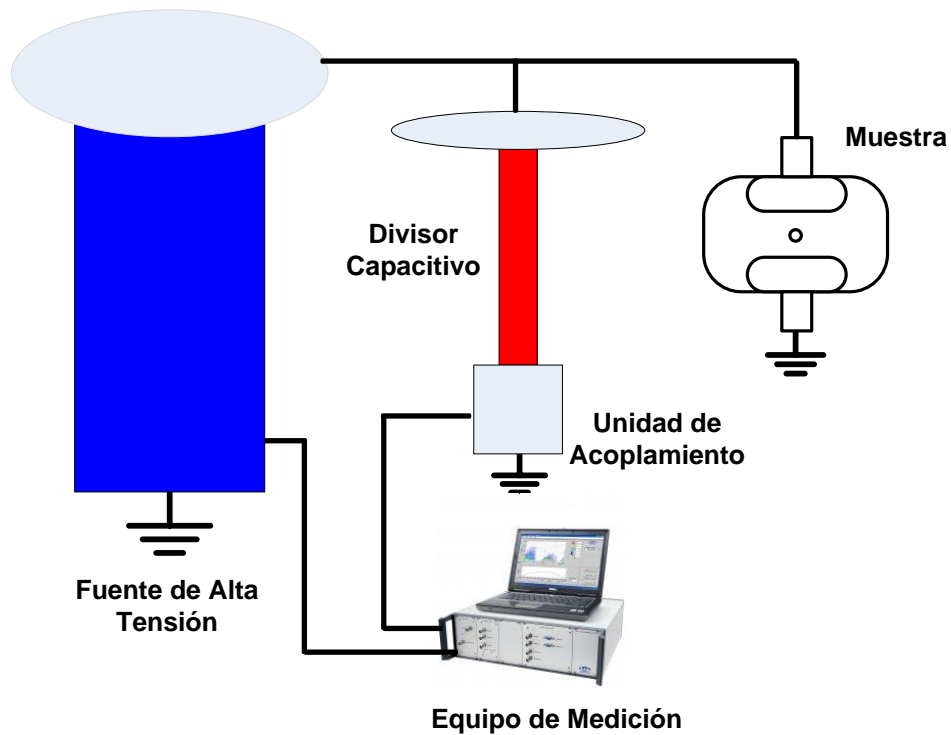


Figura 25.- Esquema de conexión para generación de datos en laboratorio

La descarga interna con electrodos planos se consideró utilizando un cable de potencia patrón en el que se tienen cuantificadas las DP's. La descarga externa o efecto corona es fácilmente generada con electrodos de tipo punta-plano y por último la descarga superficial se generó con electrodos separados por un material aislante, para efectos prácticos se utilizó madera y polietileno de alta densidad.

El efecto de ruido es fácilmente inducido al poner puntos fijos flotantes cerca del objeto bajo prueba, con falsos contactos en las conexiones y la inducción producida por los balastos electrónicos de las lámparas de iluminación (estos elementos son las fuentes generadoras de ruido más común en laboratorio y campo).

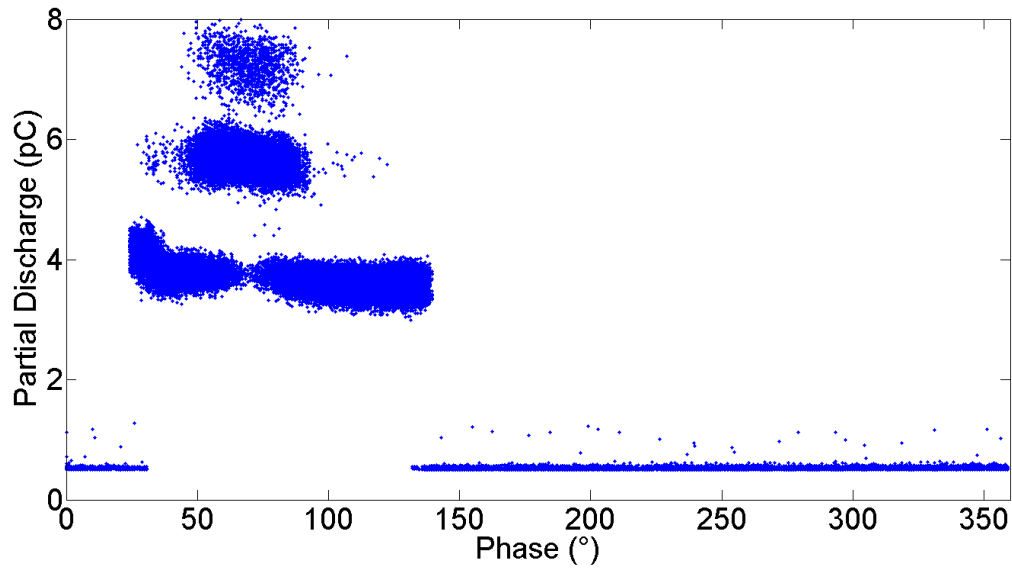


Figura 26.- Vector de datos en grafica X-Y (Magnitud-fase) de medición de Descarga Externa (Efecto Corona)

En la Figura 26 se muestra un vector de datos de medición de efecto corona considerando la configuración punta-plano, el comportamiento ha sido ampliamente reportado en la literatura, es un efecto que se presenta en el semiciclo positivo o semiciclo negativo de acuerdo a la ubicación de la tierra y el potencial. Este se considera un patrón de referencia para ese efecto. Los puntos aislados que se visualizan en la grafica se consideran ruido.

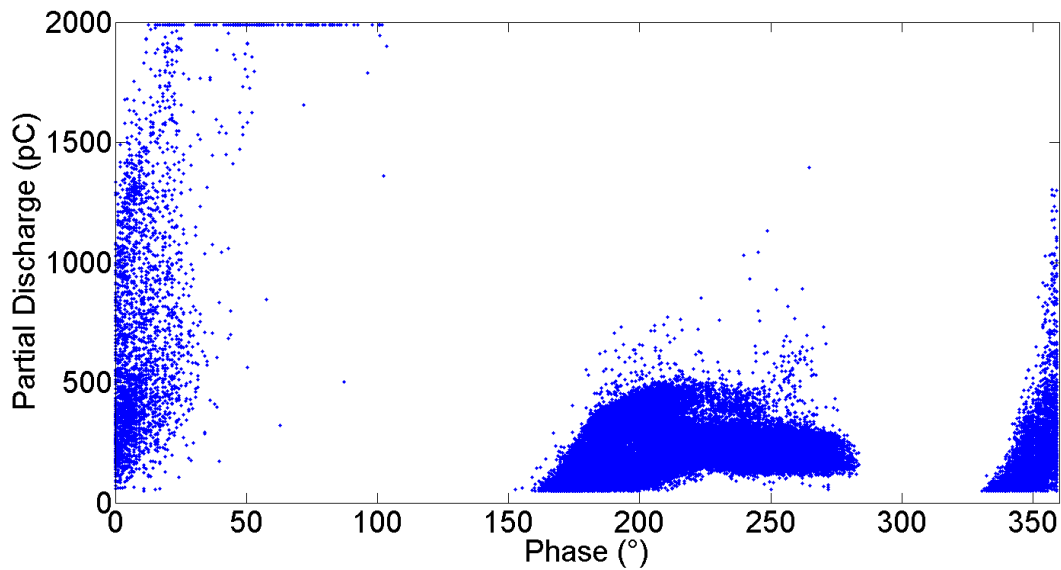


Figura 27.- Vector de datos en grafica X-Y (Magnitud-fase) de medición de Descarga Interna

En la Figura 27 se muestra el comportamiento de la Descarga superficial, en donde se puede observar que después de los cruces por cero de los semiciclos (0° , 180° y 360°) se presenta una gran concentración de datos. Los datos aislados se consideran ruido, ya que como se mencionó anteriormente, se debe presentar repetibilidad en la zona de datos para saber que existe un efecto.

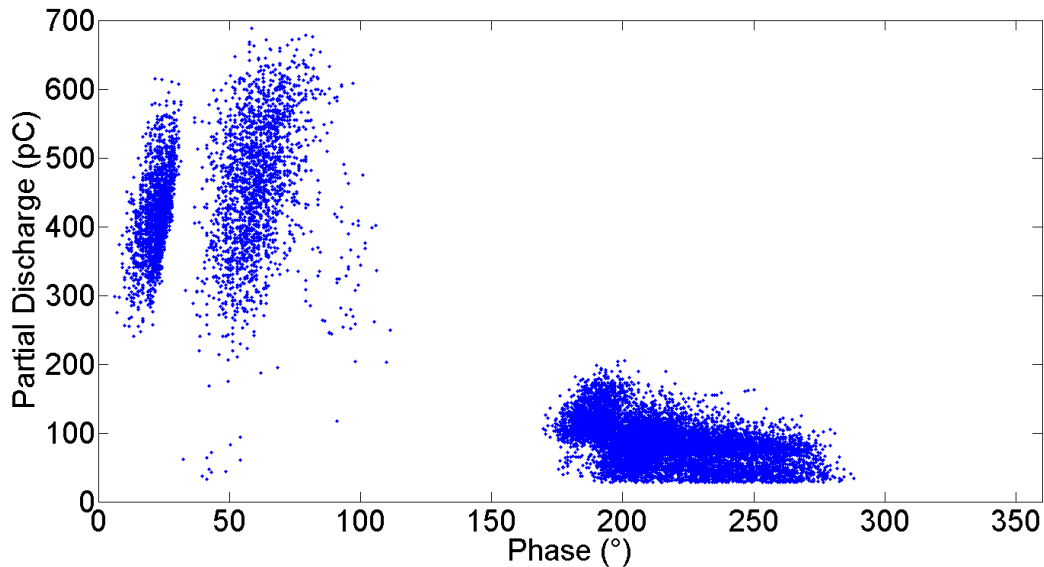


Figura 28.- Vector de datos en grafica X-Y (Magnitud-fase) de medición de Descarga Superficial.

La Figura 28 presenta el comportamiento de la descarga superficial, aquí al igual que en la Figura 27, existe la mayor concentración de datos en los cruces por cero, sin embargo, al mostrar aleatoriedad en algunos cuadrantes, el punto de mayor interés es en el cruce a 180° .

La visualización adecuada de la información es un medio para resolver problemas de toma de decisiones o la confirmación de hipótesis, es por ello que es un área de gran interés en las ciencias.

La Descarga externa (Figura 26) es visualmente interpretable, sin embargo, como se observa en la Figura 27 y Figura 28, es difícil interpretar de manera el comportamiento de los datos, y más considerando fuentes combinadas de DP's, por lo que hoy día es considerada una limitante al no caracterizar las dependencias en un nivel mas conceptual, describir cualitativamente las regularidades y determinar dependencias en factores que no hayan sido proporcionados por los datos explícitamente.

Estas graficas mostradas son interesantes, ya que demuestran el comportamiento de las descargas parciales, y al ser obtenidas en laboratorio con condiciones controladas, pueden ser caracterizadas por el SOM y por tanto se considerarlas los patrones de referencia. Cabe mencionar hay pocas referencias que describan a detalle las gráficas o patrones de comportamiento de las DP's.

4.1.1 Desarrollo de banco de pruebas experimental

Como parte del trabajo experimental se desarrolló una plataforma de pruebas en la que cinco fabricantes contribuyeron con muestras de prueba, en la cual se consideraron marcas y procesos de manufactura especializados.

El banco de pruebas consiste en un electrodo de potencial con anillos graduadores, una boquilla tipo perno y un sistema neumático para posicionar los elementos con la presión adecuada (Ver Figura 29). En esta plataforma se realizaron mediciones experimentales para evaluar la calidad de los materiales, producto terminado e instalación, quedando bajo resguardo la información de CFE y para uso con fines de investigación.

Datos obtenidos:

Tabla 3.- Resultados de medición en muestra de fabricantes

Fabricante 1	30 vectores de datos	17 resultados satisfactorios
Fabricante 2	40 vectores de datos	22 resultados satisfactorios
Fabricante 3	38 vectores de datos	12 resultados satisfactorios
Fabricante 4	42 vectores de datos	25 resultados satisfactorios
Fabricante 5	62 vectores de datos	24 resultados satisfactorios



Figura 29.- Plataforma experimental para medición de DP en sistemas subterráneos.

En la Figura 30 se muestran los elementos a evaluar aplicando la tensión de prueba de acuerdo a la NMX-J-148-ANCE-2002, en donde el criterio de evaluación indica que el nivel máximo de DP's es de 3 pC. En la Tabla 3 se muestran los resultados de evaluación de conformidad de acuerdo al criterio de norma.

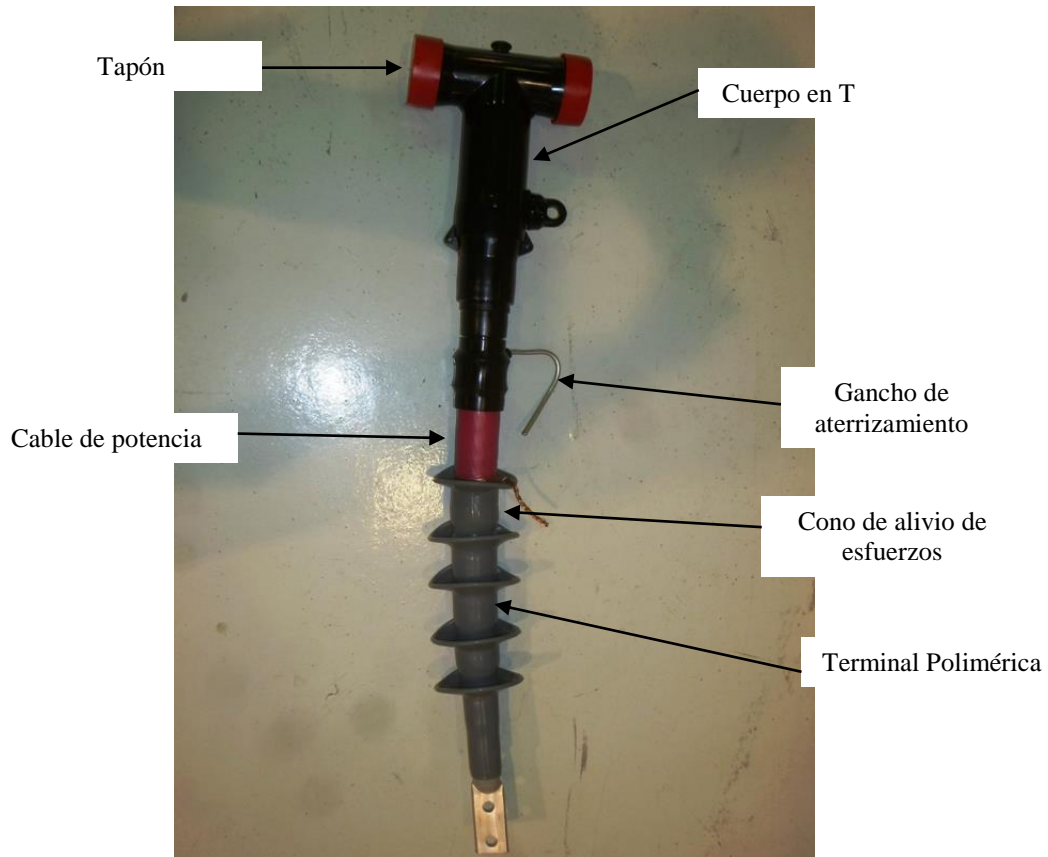


Figura 30.- Muestra prototipo para medición de DP's en laboratorio.

4.2 Obtención de datos en Sitio

Una de las principales aportaciones de este trabajo es que gracias a la infraestructura móvil que se cuenta en el LAPEM, se pueden analizar datos de pruebas en sitio realizadas por procedimientos estandarizados.

El LAPEM adquirió en 2009 una plataforma móvil para pruebas en laboratorio y en sitio para cables subterráneos de 115, 230 y 400 kV (ver Figura 31), siendo la única en América Latina, con una inversión de \$16 000 000.00 (dieciséis millones de pesos).



Figura 31.- Sistema de pruebas móvil resonante con medición de DP's a cables subterráneos.

Las características son las siguientes:

Sistema resonante de frecuencia variable 83 A, 260 kV, modelo WRT 83/260 T, marca HIGH VOLT, constituido por:

- i. Unidad de control y alimentación modelo RSE 400 serie 888658
- ii. Reactor de Alta Tensión modelo DE 30000/260-83 serie 887983
- iii. Transformador de excitación ETO 72/3-20 serie 887984
- iv. Divisor capacitivo tipo WMC serie 888996
- v. Bloque de impedancia de alta tensión serie 888997
- vi. Contenedor para sistema móvil de pruebas de alta tensión serie CV936
- vii. Computadora marca SIEMENS con software de control y medición del equipo de pruebas resonante
- viii. Equipo de medición de descargas parciales marca LDIC modelo LDS-6**
- ix. Computadora marca DELL con software para control y medición del equipo de descargas parciales
- x. Calibrador del equipo de descargas parciales para rangos de 5 – 100 pC.**

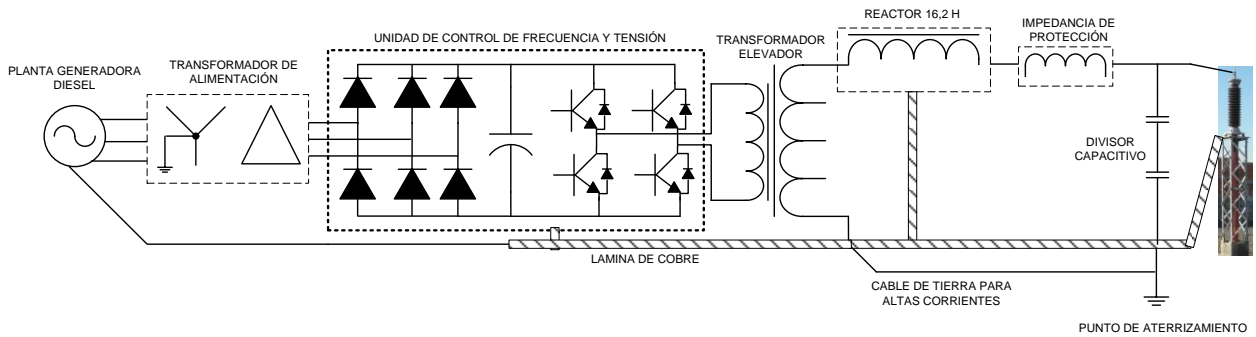


Figura 32.- Diagrama conceptual de Sistema móvil resonante (etapa de potencia).

El diagrama esquemático de la etapa de potencia del Sistema Resonante se presenta en la **Figura 32**, este equipo opera bajo el principio de la resonancia, la cual es encontrada a través de la variación de la frecuencia de la tensión del objeto bajo prueba.

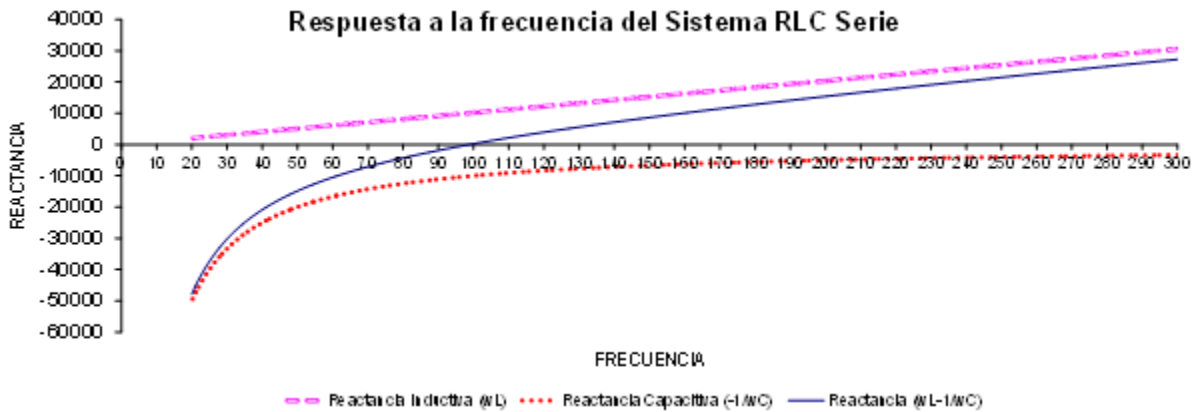


Figura 33.- Ejemplo grafico de cancelación de reactancias en prueba de alta tensión a frecuencia de resonancia.

En la **Figura 33** se presenta una gráfica de barrido de frecuencia de 20 a 300 Hz de un circuito LC serie, la línea punteada rosa es la reactancia inductiva y la línea punteada roja es la reactancia capacitiva, el punto de resonancia es cuando se igualan en magnitud ambas reactancias pero con signo contrario, lo que hace que la suma sea cero. Es considerado el punto de equilibrio energético.



Figura 34.- Ejemplo de prueba con equipo resonante en un sistema de cable de 115 kV ubicado en un poste troncocónico.



Figura 35.- Ejemplo de prueba con equipo resonante en un sistema de cable de 115 kV ubicado en una subestación.

En la **Figura 34** y **Figura 35** se presentan dos ejemplos de ubicación del equipo de pruebas, en ambos casos se realizaron medición de DP's con el fin de obtener patrones de medición con el fin de enriquecer el set de entrenamiento.

Actualmente la norma internacional IEC 60840 y la IEC 62067 que aplica a sistemas de cables de potencia subterráneos, indican en el punto 15 y 14 respectivamente, que se debe aplicar una tensión de prueba de acuerdo a la clase de aislamiento del cable durante un periodo de una hora, con el fin de evaluar la condición de instalación de sistema.

Debido a la continuidad del SEN, no es factible utilizar el sistema eléctrico (on-line) para realizar diagnóstico de sistemas de reciente construcción, ya que en caso de falla, repercutiría en la operación de las protecciones y por tanto dejaría sin energía la carga que se alimenta, Parques Industriales, Zonas Urbanas, etc., por mencionar algunas. Es por ello que la opción más viable es la utilización de equipos móviles con sistemas de adquisición. Diversos trabajos se han presentado por los Grupos de Investigación en el área de diagnóstico en equipo eléctrico utilizando la medición de DP's alimentando por fuentes independientes (off-line) [4] [16] [5] [41] [42], otros trabajos han recurrido de simulación en software como MATLAB® [43] [25].

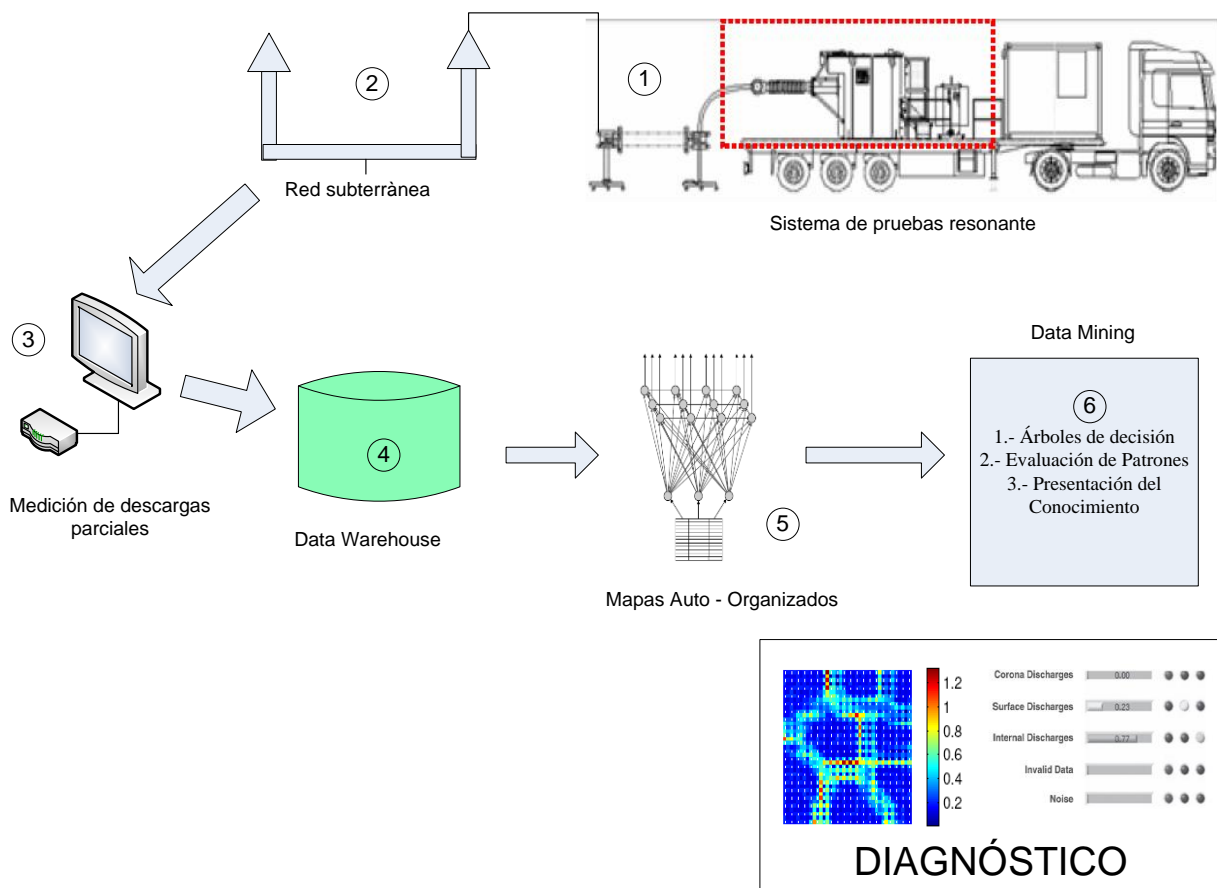


Figura 36.- Diagrama conceptual de la plataforma de trabajo experimental

Una aportación de este trabajo de investigación doctoral es la interpretación de patrones reales, medidos a condiciones de prueba normalizadas, con el fin de tener una trazabilidad de resultados a través del sistema basado en conocimiento desarrollado.

En la Figura 36 se esquematiza el sistema de trabajo para hacer la validación teórico – experimental de los algoritmos propuestos, de manera que se presenten resultados aplicables a las condiciones operativas en México, que en general por la topografía es más severa que en la mayor parte de los países. Los bloques del sistema para validación experimental son los siguientes:

1. Se utiliza un equipo resonante para realizar las pruebas en campo y obtener datos del estado de redes subterráneas nuevas y en operación.
2. Se realiza la adquisición de los datos a través de un sistema de medición de descargas parciales en sitio y se validarán los atributos principales (magnitud y fase) de la descarga parcial obtenidos en laboratorio.
3. Estos datos se concentran en el repositorio, se procesarán los datos utilizando MATLAB®.
4. Se aplica SOM para realizar el pre-procesamiento de los datos, ya que deben ser clasificados los datos de entrada identificando las características más representativas de su comportamiento. Para asegurar el desempeño del SOM se realiza un estudio comparativo del desempeño de los algoritmos competitivos de acuerdo al tamaño del vector de datos a procesar, las métricas de calidad del SOM y el tiempo de procesamiento.
5. Se aplica el minado de la información a través de herramientas heurísticas, ya que el principal objetivo es resolver el problema de Clasificación de fuentes de DP's, esta información será desplegada en un esquema fácilmente identificable, de manera que se pueda establecer un diagnóstico integral del estado de la red subterránea (fingerprint o huella digital).

Como parte del compromiso del LAPEM hacia la CFE, la propuesta y desarrollo de este trabajo de investigación será aplicada para monitoreo, evaluación y diagnóstico de líneas de transmisión (L.T.) de nueva construcción o ya estén actualmente en operación. Estos resultados son originales, ya que dentro de la revisión de estado del arte se ha podido encontrar que la mayoría de los trabajos presentados se basan en resultados a escala o en simulación de laboratorio, y después se generalizan los resultados a pruebas en sitio. Sin embargo, no se hace la evaluación de los patrones considerando el número de

fuentes de ruido, siempre presentes por las condiciones climáticas, el estado operativo de las instalaciones, la variabilidad de los transitorios por operación del sistema, etc., por ellos es muy importante minimizar el error aleatorio en la toma de datos antes de vaciar en un repositorio, para que de esta manera los algoritmos de análisis sean lo más eficientes y confiables.

Los resultados obtenidos de las primeras mediciones con condiciones en sitio (On Site) en México y aplicando la metodología normalizada fueron presentadas en [6].

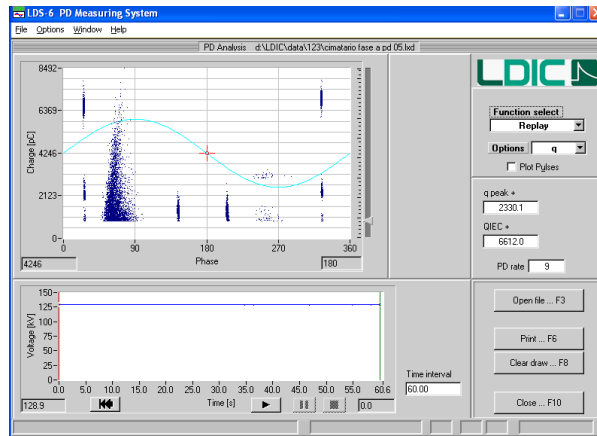


Figura 37.- Pantalla principal de sistema de adquisición de datos.

En [44] se presenta una de las únicas referencias de aplicación de diseño de experimentos para estudio de las variables de influencia en el proceso de medición. Una de las etapas importantes en la configuración del sistema de medición es la selección de los umbrales de detección, lo que nos ayuda a evitar que se sature por la medición de ruido, sin embargo, esta determinación del umbral puede hacer que se pierda información o evidencias de presencia de fuentes de DP's. Otro parámetro que se debe evaluar es el tiempo de medición, que como se mencionó anteriormente, un atributo para diagnóstico es la tasa de repetición, por lo que se deben guardar los registros necesarios para la extracción del patrón de DP's.

Una probable razón para que no se sigan aplicando las metodologías de diseño de experimentos en la selección de configuraciones óptimas de medición, es que se continúa evaluando los datos en base a la experiencia, por lo cual este proceso aún sigue siendo a prueba y error. Esto nos da la justificación para proponer umbrales de operación para determinadas condiciones de prueba en sitio.

4.3 Pre-tratamiento de datos usando SOM

Una vez teniendo los datos estructurados en archivo *.CSV o *.TXT, se realiza el tratamiento con el SOM obteniendo los siguientes resultados:

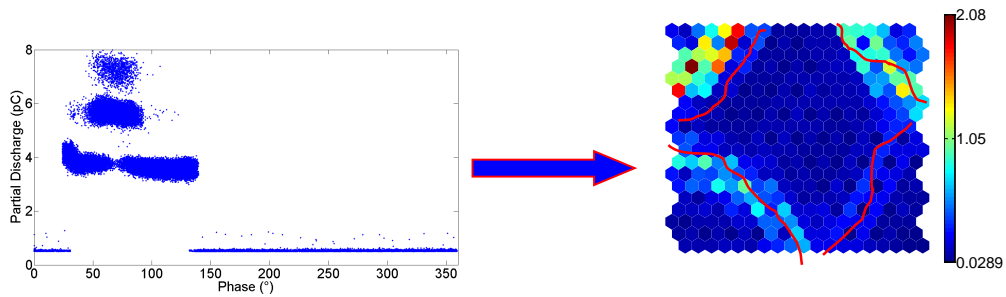


Figura 38 Determinación de U-matrix para DP Externa

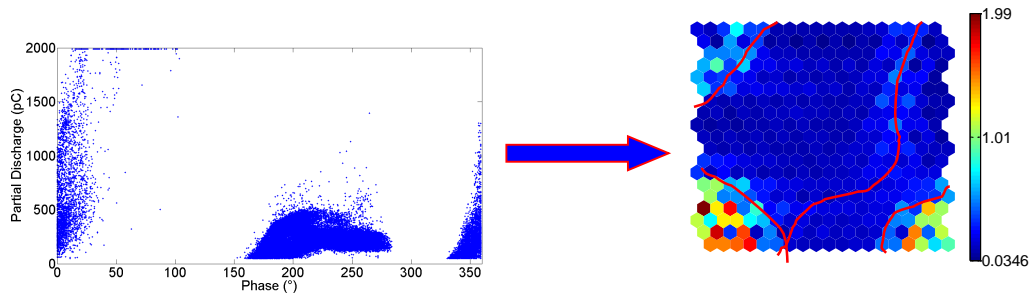


Figura 39.- Determinación de U-matrix para DP Interna

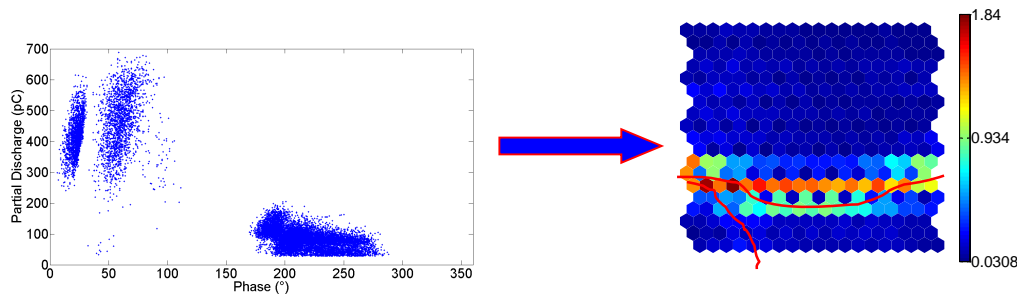


Figura 40.- Determinación de U-matrix para DP Superficial

Se observa en la Figura 38, Figura 39 y Figura 40 la salida de U-matrix para cada vector de datos, el color de cada hexágono representa la distancia promedio entre el vector de referencia ubicado en la posición del mapa y los hexágonos vecinos. El color azul indica que el grupo de datos es cercano o similar y el rojo representa que el grupo de datos está lejano o es poco similar. De esta manera se pueden identificar los clústers por grupo de datos. Aquí podemos observar puntualmente como el SOM determinó la repetibilidad de cada DP de acuerdo a su zona, agrupando en cada clúster por magnitud y fase el dato correspondiente.

Las U-matrix presentada fue considerando un arreglo de 10x10 neuronas de salida, obteniendo una gráfica bidimensional de resolución 19x19 celdas, en las que se agrupó un vector de datos de hasta 4x1000000. Los resultados obtenidos fueron presentados en [45].

Con el entrenamiento del SOM se obtuvo el nuevo vector prototipo, sin embargo, una serie de cuestionamientos deben ser respondidos, ¿se mantiene la topología de los datos? ¿Son representados adecuadamente los datos en el vector prototipo?

4.4 Aplicación de Métricas de Calidad en el entrenamiento del SOM

Como se comenta en el punto anterior, la aplicación del SOM como pre-procesamiento es un propuesta que nos ayudará a transformar los datos en un vector prototipo, sin embargo, se debe mantener la topología de los datos y algunos de los principales elementos que deben considerar son el número de neuronas con el que se hará el entrenamiento y el tipo de aprendizaje competitivo aplicado en el algoritmo.

Como se mencionó en el punto 3.2, el error de cuantización (q_e) y el error topológico (t_e) son calculados para medir la calidad del mapa definido por el vector prototipo, a continuación se muestra en la Tabla 4 los resultados considerando arreglos de neuronas de 5X5, 10X10 y 20X20.

Tabla 4.- Calidad del SOM

Vector de Datos Neuronas del SOM	Error de cuantización q_e (%)	Error Topológico t_e (%)
Descarga Superficial 5x5	0.378	0.0426
Descarga Superficial 10x10	0.181	0.0547
Descarga Superficial 20x20	0.101	0.0561
Descarga Interna 5x5	0.188	0.0865
Descarga Interna 10x10	0.097	0.0652
Descarga Interna 20x20	0.038	0.0511
Descarga Externa 5x5	0.271	0.0342
Descarga Externa 10x10	0.1326	0.0906
Descarga Externa 20x20	0.053	0.1342

En la Tabla 4 se muestra que el error de cuantización disminuye conforme el tamaño del mapa aumenta, ya que la media de la distancia del vector de datos al vector prototipo es pequeña y significa que los datos son más cercanos a su prototipo cuando el SOM crece.

En la columna del Error Topológico se observa lo contrario, conforme el mapa aumenta el error también aumenta. En el caso del vector de descarga externa se observa que el error tuvo un incremento considerable, esto es debido a que ahora los datos se están distribuyendo en un mayor número de celdas, para el caso de la DP externa, como los datos son concentrados en un cuadrante, ahora son distribuidos en una mayor área para su identificación. Lo importantes es que de acuerdo al valor obtenido por el error de cuantificación, el vector prototipo representa adecuadamente a todos los datos.

4.5 Comparación de Aprendizaje Competitivo en Entrenamiento del SOM

El SOM, como se explicó anteriormente, debe ser capaz de reconocer la redundancia de los datos, rasgos comunes, regularidades y categorías de datos, de esta forma se produce el autoaprendizaje. Para encontrar el SOM que cumpla con la eficiencia necesaria para resolver el problema de clasificación, es importante evaluar los algoritmos de aprendizaje competitivo propuestos, ya que cada uno tiene ventajas y desventajas, de manera que se cumpla con un adecuado tiempo de procesamiento y minimizando los errores de cuantización y topológicos propios del entrenamiento.

Para realizar la comparación se establece un marco referencia de competencia, en el cual se utilizarán los vectores de datos y los parámetros de entrenamiento en los algoritmos competitivos deben ser iguales. Los parámetros para cada algoritmo se presentan en la Tabla 5.

Tabla 5.- Parámetros de entrenamiento

	WTA	FSCL	RPCL
Épocas	100	100	100
η	0.1	0.1	0.05
β	0.01	0.01	0.01

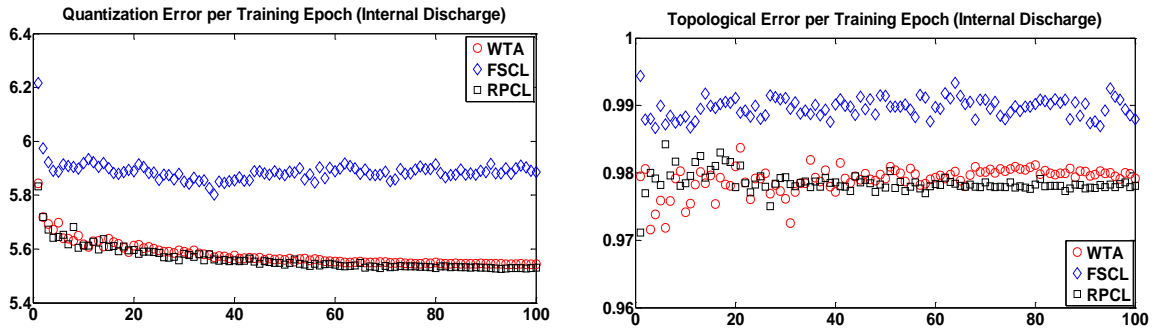


Figura 41.- Error de cuantización y topológico por época de entrenamiento (Descarga Interna)

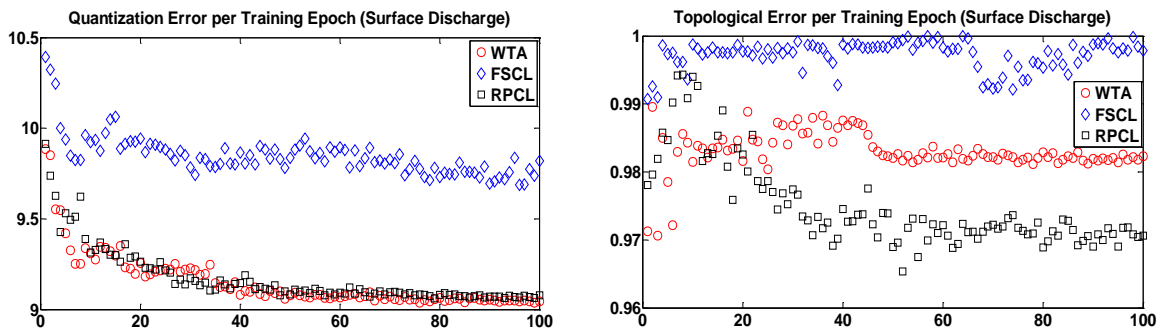


Figura 42.- Error de cuantización y topológico por época de entrenamiento (Descarga Superficial)

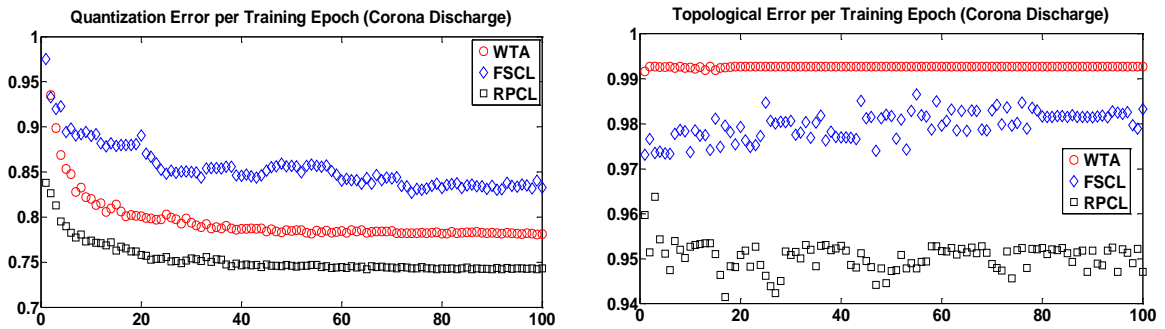


Figura 43.- Error de cuantización y topológico por época de entrenamiento (Descarga Externa)

En Figura 41, Figura 42 y Figura 43 nos demuestran claramente la convergencia de los errores en cada uno de los algoritmos y para diferentes vectores de datos. Es muy interesante ver como es el comportamiento del error topológico cuando es entrenado con el vector de datos de DP superficial, sin embargo converge adecuadamente. En la Tabla 6 se presenta el resumen del desempeño del SOM con cada algoritmo.

Tabla 6.- Desempeño del algoritmos competitivos en el SOM

	WTA	FSCL	RPCL
Descarga Superficial			
q_e	0.9	0.98	0.91
t_e	0.0985	0.1	0.097
tiempo	849 segundos	1160 segundos	1226 segundos
Descarga Interna			
q_e	0.55	0.59	0.54
t_e	0.98	0.99	0.98
Tiempo	173 segundos	222 segundos	241 segundos
Descarga Externa			
q_e	0.78	0.85	0.75
t_e	0.099	0.098	0.095
Tiempo	889 segundos	1191 segundos	1362 segundos

En el estudio comparativo presentado en la Tabla 6 se muestra que el algoritmo WTA tiene un adecuado desempeño y en algunos casos el menor error, pero siempre tiene el menor tiempo de entrenamiento. En segundo lugar de desempeño se puede considerar el algoritmo FSCL, ya que el tiempo de entrenamiento es medio y algunos valores de error son los mayores. El algoritmo RCPL es el que necesita el mayor tiempo de procesamiento, pero para vectores de datos como el de DP's externa e interna tiene los menores errores.

De acuerdo a los resultados obtenidos, el criterio de selección dependerá del tipo de DP a evaluar, ya que como se ha visto, se deben minimizar los errores topológicos y de cuantización, de manera que esta etapa de pre-procesamiento sea confiable y eficiente para que las siguientes etapas de Algoritmo Híbrido sean también eficientes.

4.6 Visualización de datos utilizando SOM en 3D

Como parte del entregable a CFE se propone trabajar en graficas de representación del vector de salida, de manera que sea interpretable y de fácil aplicación.

Una de las propuestas es representar el vector prototipo de la salida del SOM en una gráfica de superficie que muestre la distancia promedio de las neuronas vecinas y que resalte el área de mayor concentración de datos de interés.

Como parte de los resultados obtenidos en laboratorio, se cuenta con datos de referencia (ahora propuesto como benchmark para validar los algoritmos propuestos), en los que se ha trabajado en la representación gráfica del vector prototipo (Ver Figura 44, Figura 45 y Figura 46).

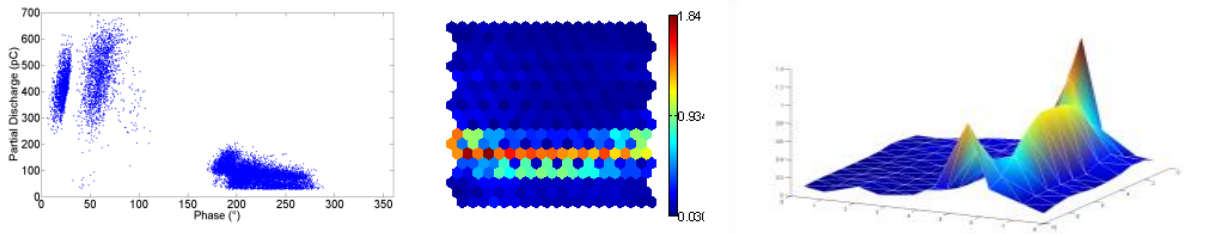


Figura 44.- Representación de SOM en U-matrix y en 3D (Descarga interna)

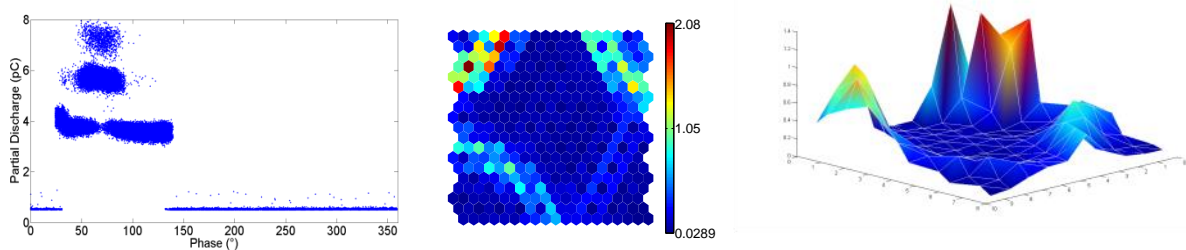


Figura 45.- Representación del SOM en U-matrix y en 3D (Efecto Corona)

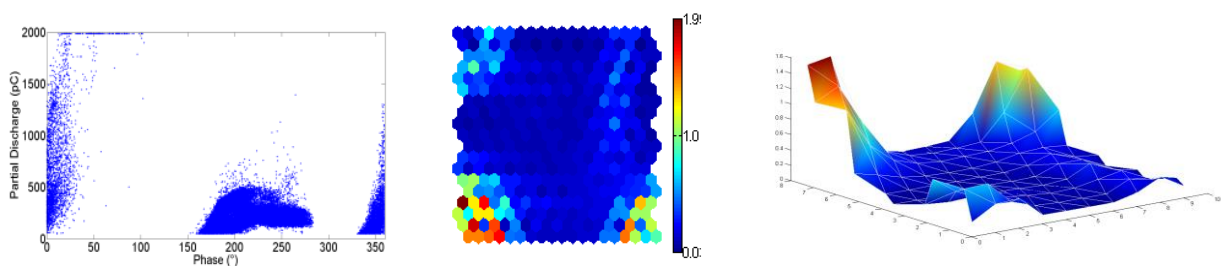


Figura 46.- Representación del SOM en U-matrix y 3D (Descarga Superficial)

CAPÍTULO 5: EVALUACIÓN DEL DESEMPEÑO DE ALGORITMOS COMPETITIVOS EN EL SOM UTILIZANDO METODOLOGÍA DE SUPERFICIE DE RESPUESTA.

En esta sección se presentan resultados originales encontrados en la comparación de los algoritmos competitivos utilizando superficie de respuesta. El principal objetivo es explorar la relación entre variables numéricas y categóricas y el cómo influyen en el desempeño del SOM. Las variables analizadas fueron el tipo de algoritmo, tasa de aprendizaje, épocas, etc, estas fueron evaluadas en base a las salidas de interés tales como las métricas de calidad del SOM y el tiempo de entrenamiento. Los gráficos de superficie de respuesta fueron utilizados para determinar los efectos de interacción de factores principales y las condiciones óptimas para mejorar el desempeño en la clasificación de fuentes de DP's.

5.1 Diseño y análisis del experimento.

Técnicas de diseños de experimentos o DOE por sus siglas en inglés (Design of Experiments), son bien conocidas en el área industrial para diseñar, mejorar u optimizar los procesos de producción, desarrollo de productos, reducción de tiempo y costo, etc.

Típicamente, un diseño de experimentos debe ser aplicado por una o más de las siguientes razones:

- a) Determinar las causas principales de la variabilidad en una respuesta medida.
- b) Encontrar las condiciones minimizar o maximizar una respuesta.
- c) Comparar la respuesta obtenida a diferentes tratamientos.
- d) Obtener un modelo matemático para predecir futuras respuestas.

La metodología de superficie de respuesta (MSR) es un conjunto de técnicas matemáticas y estadísticas para analizar los efectos de las variables independientes en la respuesta. MSR tiene amplia aplicación en los procesos de diseño y optimización de los ya existentes. Esta metodología es muy práctica comparada con los modelos teóricos que se pueden plantear en una metodología experimental porque incluye los efectos de interacción de las variables y, eventualmente, que representa todos efectos de los parámetros, como en este caso, requeridos en el algoritmo del SOM.

En un sistema multivariable, un enfoque clásico es mover una variable a la vez, con el fin de estudiar los efectos en otras variables para una respuesta en particular, lo cual es ineficiente. Una alternativa a esta estrategia es aplicar una herramienta estadística como la MSR, la cual al ser aplicada puede dar un enfoque de solución para múltiples variables en un sistema complejo. Esta metodología fue aplicada para analizar el impacto de los factores del SOM en el tiempo de entrenamiento y las variables de calidad (*qe* y *te*). La ventaja del MSR es la minimización del número de experimentos y tiempo necesario. En este proceso de optimización se estudió la respuesta de la combinación de factores estadísticamente

propuestos, estimando los coeficientes del modelo para ajustar los datos experimentales al modelo de respuesta, prediciendo la respuesta del modelo ajustado y verificando la adecuación del modelo. Después de un cribado de factores utilizando experimentos 2^k , se utilizó un diseño D-óptimo. Los diseños tradicionales como el Box-Behnken o el Diseño Central Compuesto son generales y flexibles si la región experimental es un cubo o una esfera, sin embargo para el caso de la respuesta del SOM, no existe reporte de resultados de la región de operación del algoritmo, por lo que no es trivial la selección del algoritmo de superficie de respuesta a utilizar.

La mejor alternativa es hacer referencia a los diseños en donde proponen diseños óptimos generados por computadora. Por diseño óptimo se entiende un diseño que es “mejor” respecto a un criterio. El enfoque usual es especificar un modelo, determinar la región de interés, seleccionar el número de corridas que deberán hacerse, especificar el criterio de optimalidad y después elegir los puntos del diseño de un conjunto de puntos candidatos que el experimentador consideraría utilizar. El diseño seleccionado y que fue generado por Design Expert® es el D-óptimo.

Se dice que un diseño es D-óptimo si:

$$|(X'X)^{-1}| \text{ se minimiza.}$$

Esto es que se minimiza el volumen de región de confianza conjunta para el vector de coeficientes de regresión.

Tabla 7.- Rango experimental y niveles de variables numéricas

Factor	Niveles codificados	
		-1
η (eta)	1E-5	1
β (Beta)	0.01	1

Tabla 8.- Rango experimental y niveles de variables categóricas

Factor	Niveles		
Algorithm	WTA	FSCL	RPCL
Discharge	Internal	External	Surface

El análisis de regresión fue desarrollado para determinar la superficie de respuesta como función de la ecuación polinomial de segundo orden siguiente:

$$Y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_{i \leq j}^k \beta_{ij} x_i x_j + \varepsilon \quad (17)$$

Donde Y es la respuesta predicha, β_i , β_{ii} , β_{ij} representan los efectos de interacción lineal y cuadrática. β_0 es el término de intersección y x_i, x_j, \dots, x_k son las entradas que afectan al valor de salida. El esquema de experimentos desarrollados para este estudio es presentado en la Tabla 9.

Tabla 9.- Matriz de tratamientos y resultados

Inputs				Outputs		
Eta	Beta	Alg.	Discharge	qe	te	Time
A	B	C	D			
1	0.5	WTA	External	0.86	0.99	176.50
1E-5	0.01	WTA	Surface	10.22	0.97	157.96
1	1	FSCL	External	0.96	0.98	228.97
1	0.5	FSCL	Surface	9.61	0.96	215.96
0.25	0.25	WTA	External	0.76	1	171.63
1E-5	0.01	RPCL	External	1.12E6	1	243.63
0.5	1	WTA	Surface	11.84	0.99	159.79
0.5	1	FSCL	Internal	7.84	0.98	43.97
1	0.01	FSCL	Internal	5.52	0.97	44.03
1	0.01	RPCL	Internal	7.32	0.98	47.21
0.25	0.5	FSCL	External	0.85	0.92	234.53
0.5	1	WTA	Surface	10.81	0.96	163.07
1E-5	1	RPCL	Internal	1.15E6	1	47.55
1E-5	0.5	WTA	Internal	6.39	0.97	32.91
1E-5	1	FSCL	Surface	11.41	0.96	219.15
1E-5	0.01	RPCL	External	5.82E6	1	248.45
1	1	RPCL	Surface	118.23	1	230.25
1E-5	0.01	FSCL	Internal	6.21	0.97	46.19
0.5	0.01	WTA	Internal	8.71	0.99	159.05
1E-5	0.5	RPCL	Surface	2.50E6	1	47.23
0.5	0.01	FSCL	Surface	8.68	0.98	215.00
0.5	0.5	RPCL	Internal	169.61	1	46.93
1	1	RPCL	External	50.14	1	244.14
1E-5	0.01	FSCL	Internal	6.13	0.98	44.05
1E-5	1	WTA	External	2.06	0.99	170.54
1	1	WTA	Internal	7.80	0.98	32.88
1	1	WTA	Internal	7.47	0.99	32.95
1	1	RPCL	External	57.55	1	246.15
1	0.01	RPCL	External	1.19	0.97	251.57
0.75	0.25	RPCL	Surface	119.92	1	230.02
0.75	0.25	FSCL	External	0.76	0.96	229.75
1	0.01	WTA	Surface	8.50	0.96	164.22

En la Figura 41, Figura 42 y Figura 43 se presentan ejemplos del desempeño y convergencia de los algoritmos de aprendizaje competitivo a diferentes fuentes de DP's, así también en la Figura 47, Figura 48 y Figura 49 son presentados los vectores de datos y vectores prototipo resultantes del entrenamiento del SOM.

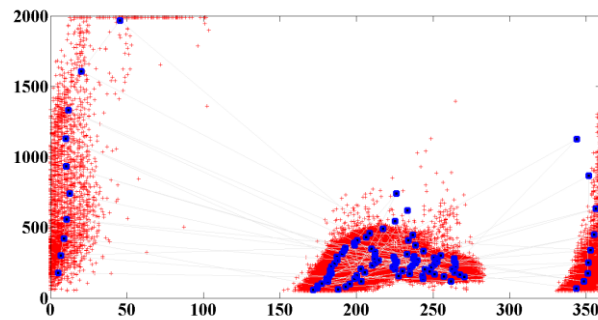


Figura 47.- Vector de Datos y Vector Prototipo (Descarga Superficial)

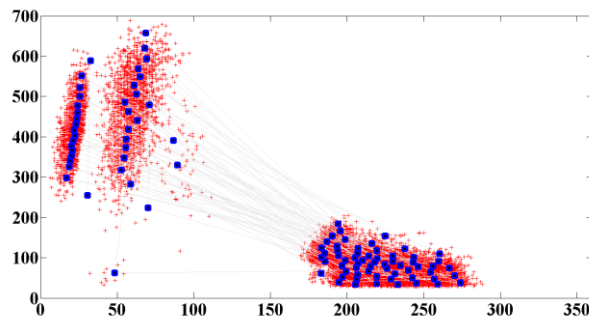


Figura 48.- Vector de Datos y Vector Prototipo (Descarga Interna)

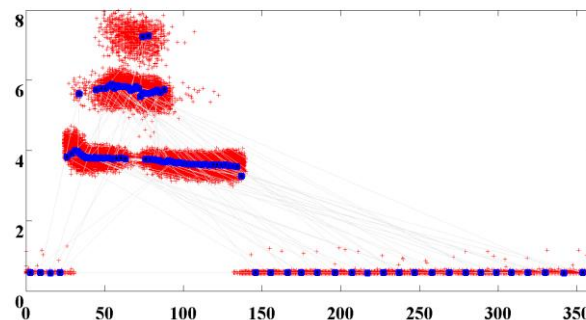


Figura 49.- Vector de Datos y Vector Prototipo (Descarga Externa)

5.2 Análisis de varianza

La significancia estadística y la prueba de bondad de ajuste de la ecuación fueron evaluados por R^2 y por F-test del análisis de varianza (ANOVA), este análisis subdivide la variación total en un grupo de datos de los componentes asociados con fuentes específicas de variación para la prueba de hipótesis en los parámetros del modelo. Un valor grande del F-test indica que la mayor variabilidad puede ser expresada por la ecuación de regresión donde un valor bajo de p-value (<0.05) indica que el modelo es considerado para ser estadísticamente significativo (Tabla 10).

Tabla 10.- Resultados de la ANOVA

	Sum of Square	DF	Mean Square	F-Value	P-value
Para <i>qe</i>					
Model	425.37	19	22.39	45.82	<0.001
Residual	5.86	12	0.49		
Lack of fit	0.2	1	0.2	0.72	0.484
Pure error	1.37	5	0.27		
Cor total	428.24	31			
$R^2 = 0.9864$					
$R^2_{adj} = 0.9649$					
Para <i>te</i>					
Model	843.7	21	40.17	143.48	<0.001
Residual	7.86	10	0.65		
Lack of fit	0.25	1	0.25	0.91	1.245
Pure error	1.4	5	0.28		
Cor total	850.2	31			
$R^2 = 0.9764$					
$R^2_{adj} = 0.9539$					
Para <i>time</i>					
Model	2134	21	101.61	23.82	<0.001
Residual	946.2	10	78.85		
Lack of fit	9.5	11	0.86	0.2024	0.742
Pure error	21.33	5	4.26		
Cor total	2182.53	31			
$R^2 = 0.9575$					
$R^2_{adj} = 0.9286$					

La relación entre la variable de respuesta y las variables independientes es expresada por las siguientes ecuaciones cuadráticas de superficie de respuesta:

$$q_e = 3.38 - 1.67A + 0.46B - 1.97C - 2.02C^2 - 4.23D + 0.81D^2 + 0.22AB + 1.67AC + 1.57AC^2 + 0.36AD + 0.21AD^2 - 0.41BC - 0.13BC^2 - 0.42BD + 0.082BD^2 + 0.5CD + 0.46C^2D + 0.11CD^2 + 0.20C^2D^2 \tag{18}$$

$$te = -0.21 + 1.575A + 7.729B + 1.551C - 0.17C^2 + 3.032D - 1.489D^2 + 8.981AB + 8.344AC + 9.763AC^2 - 4.087AD - 2.446AD^2 + 1.602BC + 7.662BC^2 - 1.931BD - 1.990BD^2 + 6.174CD + 6.882C^2D - 0.16CD^2 + 6.520C^2D^2 - 0.12A^2 + 0.17B^2 \tag{19}$$

$$Time = 143.71 + 10.73A - 8.59B - 14.10C + 10.77C^2 - 99.70D + 27.39D^2 + 1.30AB - 4.85AC - 4.85AC^2 - 15.15AD + 25.15AD^2 - 18.28BC + 12.57BC^2 - 10.13BD + 13.49BD^2 + 38.46CD - 32.99C^2D - 20.53CD^2 + 29.06C^2D^2 - 30.94A^2 + 40.27B^2 \tag{20}$$

El ANOVA para las respuestas *te*, *qe* y *time* es mostrada en la Tabla 10, el p-value para el modelo es menor de 0.05 indicando que los términos del modelo son estadísticamente significativos. Los valores actuales y predichos son mostrados en la Figura 50, Figura 51 y Figura 52.

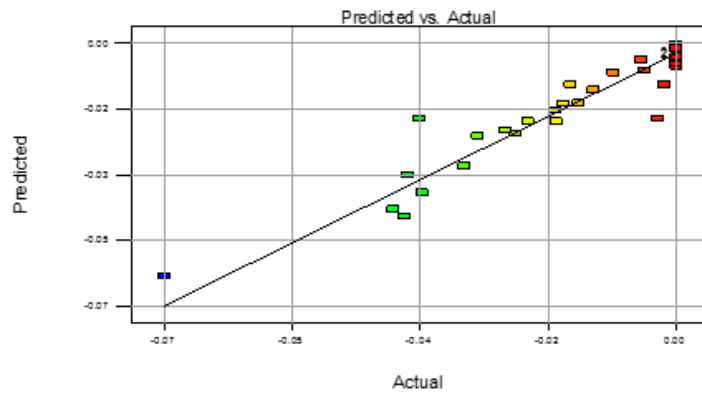


Figura 50.- Valores actuales y predichos para la respuesta *te*

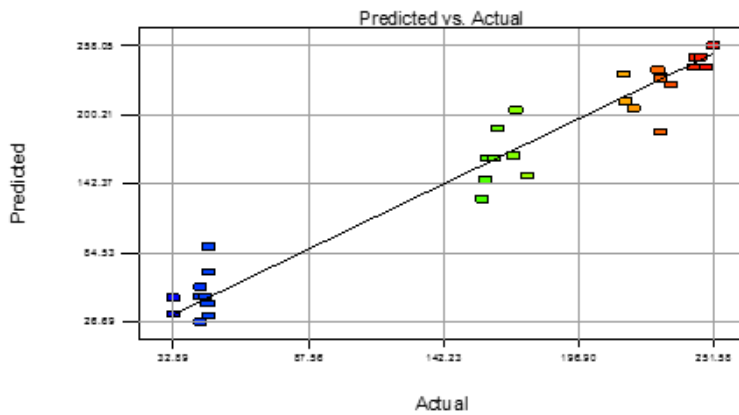


Figura 51.- Valores actuales y predichos para la respuesta *time*

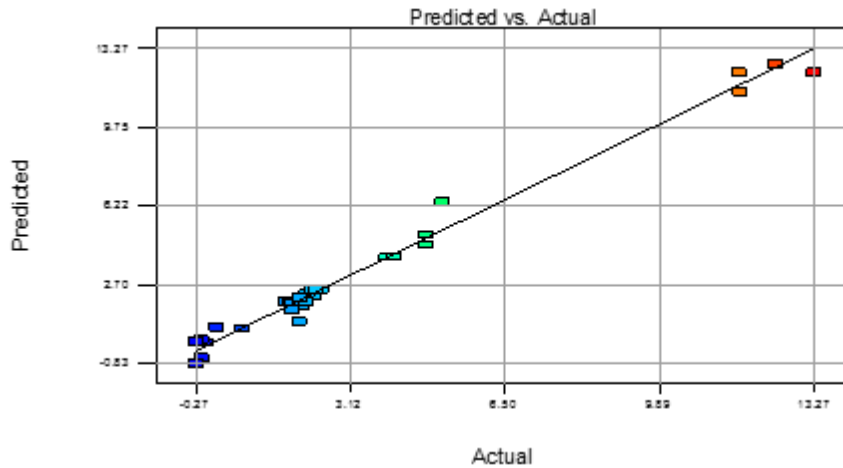


Figura 52.- Valores actuales y predichos para la respuesta *qe*

Los valores actuales son los valores medidos para un experimento en particular y los valores predichos son generados utilizando el modelo de regresión. Los valores de R^2 y $R^2_{ajustada}$ indican una gran correlación entre los valores actuales y predichos. Por ejemplo, para *qe*, el valor R^2 es de 0.9864, sugiriendo que la variación en *qe* es atribuible a las variables independientes y solo un 1.36% de la variabilidad total no puede ser representada por el modelo.

El valor F del modelo de *qe* es de 45.82, para *te* es de 143.48 y el de *time* es de 23.82, para todos ellos el $p\text{-value} < 0.0001$ indicando que es estadísticamente significativo un modelo cuadrático. Por tanto, de acuerdo a estos resultados, la relación entre las variables independientes y la respuesta puede ser representada por el modelo de regresión obtenido.

5.3 Optimización por Superficie de Respuesta

El MSR fue utilizado para estimar los efectos de los cuatro factores a diferentes niveles, en este análisis los gráficos en 3D demuestran la interacción de esos factores y sus efectos en la respuesta. Por ejemplo en la Figura 53 se muestran los tratamientos para minimizar *qe* y es claramente identificable donde se encuentra un mínimo en la figura. Sin embargo, para las respuestas *te* y *time* (Figura 54 y Figura 55) es observado un interesante comportamiento, al ser un punto silla, es una combinación de problema mini/max y función cóncavo/convexa. Por lo que para este caso es importante establecer el área factible de operación del algoritmo para obtener la respuesta óptima más exacta.

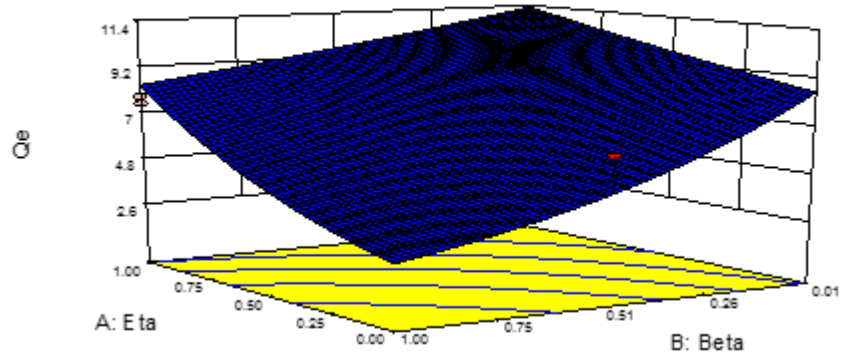


Figura 53.- Grafico de superficie de respuesta de la combinación de efectos para q_e .

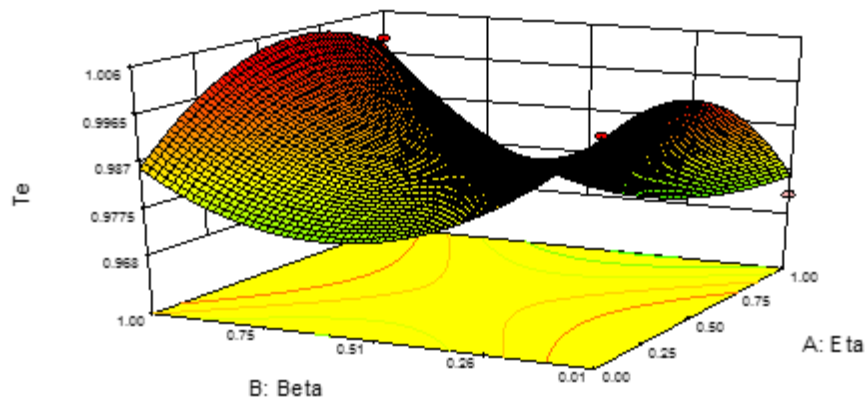


Figura 54.- Grafico de superficie de respuesta de la combinación de efectos para t_e .

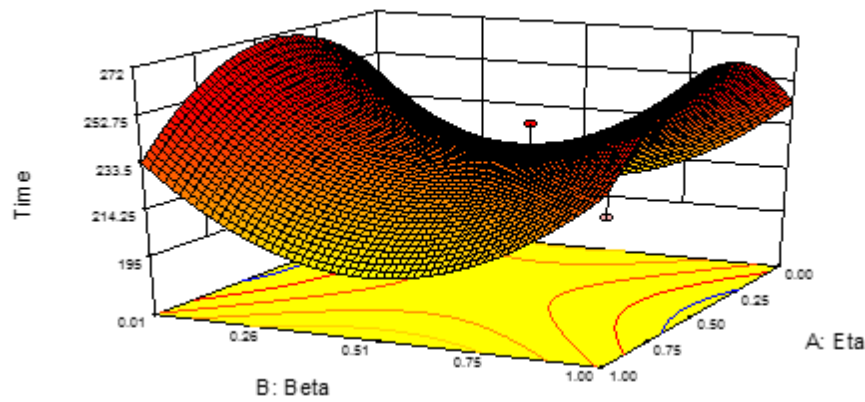


Figura 55.- Grafico de superficie de respuesta de la combinación de efectos para $time$.

En el análisis fueron encontradas las condiciones óptimas para todos los factores bajo estudio, **eta**, **beta**, tipo de algoritmo competitivo y tipo de descarga. La grafica de superficie de respuesta a condiciones óptimas es mostrada en la Figura 56, en ella se representa la función de deseabilidad, en la que un problema de múltiples respuestas es transformado en un problema de una respuesta por medio de transformaciones matemáticas.

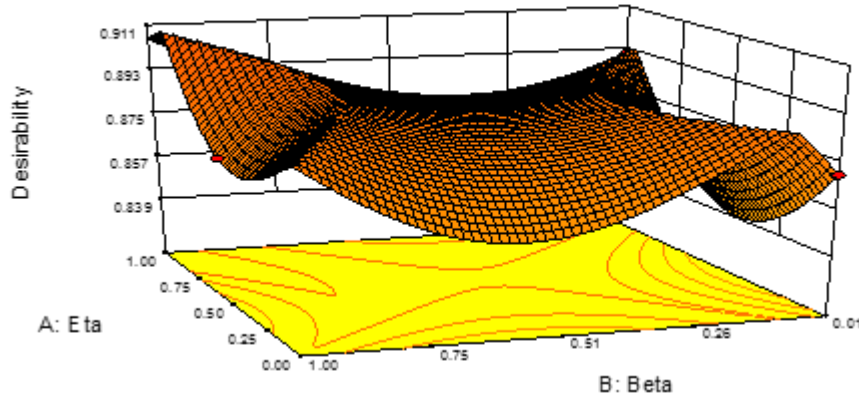


Figura 56.- Gráfica de deseabilidad para optimización de múltiples respuestas.

En la ecuación siguiente es presentada la función de deseabilidad (**dj**) con un rango de valores entre 0 y 1, donde 1 es el valor deseado (óptimo)

$$d_j = \begin{cases} \left(\frac{Y_j - Y_{\min j}}{T_j - Y_{\min j}} \right)^s & \text{if } Y_{\min j} \leq Y_j(x) \leq T_j \\ \left(\frac{Y_j - Y_{\max j}}{T_j - Y_{\max j}} \right)^s & \text{if } T_j \leq Y_j(x) \leq Y_{\max j} \\ 0 & \text{cualquier otro} \end{cases} \quad (21)$$

La mejor solución basada en los diferentes niveles de los parámetros del SOM son presentados en la siguiente Tabla 11.

Tabla 11.- Solución para diferentes combinaciones de niveles de factor numérico y categórico

INPUTS				RESPONSE			
Eta	Beta	Alg.	PD	Qe	Te	Time	Desirability
0.9	1.0	FSCL	Internal	8.0	1.0	32.9	0.9112
1.0	1.0	WTA	Internal	8.4	1.0	32.9	0.9047
1.0	1.0	WTA	Internal	8.4	1.0	32.2	0.9039
1.0	0.3	RPCL	Internal	4.2	1.0	32.9	0.9026
1.0	0.3	RPCL	Internal	5.2	1.0	32.9	0.8971
0.0	0.9	WTA	Internal	2.9	1.0	42.9	0.8951
0.0	0.9	WTA	Internal	2.9	1.0	42.0	0.8951
0.9	0.1	FSCL	Internal	7.0	1.0	32.9	0.8945
0.9	0.1	FSCL	Internal	7.0	1.0	32.9	0.8945
0.9	0.1	FSCL	Internal	6.9	1.0	32.9	0.8945
0.9	0.1	FSCL	Internal	7.0	1.0	32.9	0.8945
0.9	0.1	FSCL	Internal	7.0	1.0	32.9	0.8945
0.8	0.1	FSCL	Internal	6.9	1.0	32.9	0.8945
1.0	0.9	RPCL	Internal	12.2	1.0	32.9	0.8856
1.0	0.7	RPCL	Internal	8.4	1.0	16.8	0.8818
0.0	0.6	WTA	Surface	8.2	1.0	73.3	0.7547
0.0	0.7	WTA	Surface	8.0	1.0	74.7	0.7528
1.0	0.7	WTA	External	0.7	1.0	143.8	0.7251
1.0	0.7	WTA	External	0.7	1.0	143.8	0.7251
1.0	0.6	WTA	External	0.6	1.0	144.4	0.7232
0.0	0.7	WTA	External	1.9	1.0	151.1	0.6635
0.0	0.7	WTA	External	1.9	1.0	151.1	0.6635
1.0	0.7	WTA	Surface	14.1	1.0	136.5	0.6152
0.0	0.3	FSCL	Surface	8.7	1.0	149.8	0.5911
0.0	0.3	FSCL	Surface	8.7	1.0	150.1	0.5911
0.0	0.2	FSCL	Surface	8.4	1.0	155.0	0.5874
0.0	0.6	FSCL	Surface	10.0	0.9	155.1	0.5536
1.0	0.4	RPCL	Surface	16.7	1.0	180.3	0.5173
1.0	0.4	RPCL	Surface	17.3	1.0	180.1	0.5173
1.0	0.3	RPCL	Surface	15.3	1.0	181.2	0.5167
1.0	0.5	FSCL	External	0.5	1.0	195.9	0.4915
1.0	0.5	FSCL	External	0.5	1.0	195.9	0.4915
1.0	0.5	FSCL	External	0.5	1.0	196.2	0.4914
1.0	0.6	FSCL	External	0.5	1.0	196.2	0.4913
1.0	0.6	FSCL	External	0.5	1.0	196.5	0.4910
1.0	0.4	FSCL	Surface	10.0	1.0	208.9	0.4064
1.0	0.4	FSCL	Surface	9.9	1.0	208.9	0.4064
1.0	0.5	RPCL	External	10.8	1.0	212.8	0.3935
1.0	0.5	RPCL	External	11.0	1.0	212.7	0.3935
1.0	0.5	RPCL	External	11.2	1.0	212.7	0.3934

El análisis realizado presenta el estudio de la combinación de los efectos en el desempeño del SOM aplicando diferentes algoritmos de aprendizaje competitivo para encontrar las condiciones óptimas para clasificar fuentes de descarga parcial en vectores de datos de medición. Dos tipos de factores fueron satisfactoriamente probados: numéricos (η y β) y categóricos (Algoritmo Competitivo y tipo de DP). Un diseño D-óptimo basado en cuatro entradas y tres respuestas fue desarrollado para estimar los efectos en el tiempo de entrenamiento y en la calidad de los resultados del SOM. En el análisis del ANOVA se validó el modelo con los datos experimentales y se proponen los tratamientos para obtener una respuesta óptima, esto es validado por medio de la métrica de porcentaje de error R^2 y $R^2_{ajustada}$.

CAPÍTULO 6: DESARROLLO DE METODOLOGÍA PROPUESTA

En este capítulo se presentan los resultados originales obtenidos en la interpretación de patrones de DP's. El SBC propuesto fue aplicado en el LAPEM que es un laboratorio de pruebas y validación de prototipos de nuevos diseños para el sector eléctrico, en donde se tiene la posibilidad de realizar pruebas a equipos para diferentes aplicaciones, como transformadores de medición, cables de potencia, equipo de desconexión, etc. Se desarrolló un banco experimental para la adquisición de los primeros datos de entrenamiento y validación, donde se simularon de manera experimental los diferentes defectos de interés, como son descarga interna, descarga superficial y ruido de fondo.

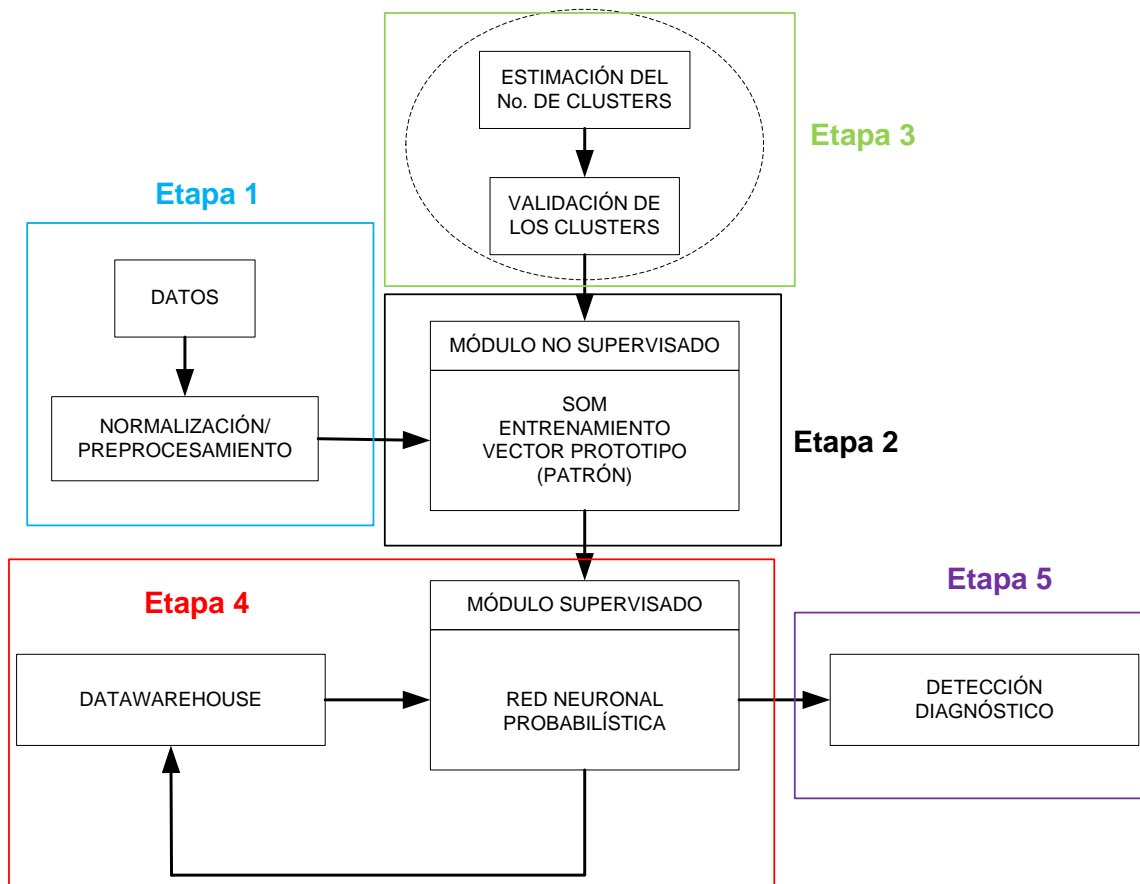


Figura 57.- Identificación de las etapas en el diagrama conceptual del SBC.

En la Figura 57 se presentan las etapas en las que está dividido el SBC, a continuación se describe a detalle el objetivo de cada una de ellas, con el fin de establecer el proceso sistémico de adquisición, pre-procesamiento, evaluación y clasificación/diagnóstico.

Etapas 1. Medición de DP's, Normalización/Pre-procesamiento.

La etapa de medición es muy importante, ya que en ello radica la certidumbre de la información. Las mediciones se realizaron en base a la norma IEC 60270 en el cual se miden los niveles de carga aparente en pC. El método convencional de interpretación es con representaciones genéricas donde se grafica el ángulo de fase a 60 Hz con respecto al valor de carga aparente, esta representación es denominada Patrones de Fase Resuelta (de sus siglas en inglés PRPD).

Los datos obtenidos en la medición son en archivos *.csv en los que se harán los pre-tratamientos correspondientes. Las métricas de distancia entre clústeres o grupos están basadas en la distancia entre datos individuales del vector, es por ello que son sensibles a la escala de las variables. El procedimiento más común de normalización o estandarización es tratar todas las variables independientemente y transformar a Z-score obteniendo la media y dividiendo entre la desviación estándar de cada variable.

$$x_{ij} = \frac{x_{ij} - \min_j(x_{ij})}{\max_j(x_{ij}) - \min_j(x_{ij})} \quad (22)$$

Este rango de escalamiento proporciona un punto de partida equitativo para cada dato y elimina el efecto escala, por lo tanto, solo variables y dependencias afectan el resultado de la cuantificación.

Etapas 2. Entrenamiento del SOM (reducción de dimensionalidad, determinación del vector prototipo).

En el trabajo presentado en [46], se determinaron los parámetros óptimos para entrenamiento del SOM. Es por ello que se determina un arreglo de 10x10 neuronas de salida, obteniendo una gráfica bidimensional de resolución 19x19 celdas, en el cual, la entrada representativa de las mediciones de vector de dimensiones 10x1000000 es transformado en un conjunto de vectores de 19x19 con las mismas propiedades topológicas, este vector es denominado vector prototipo (codebook), el cual será la entrada a la red neuronal que determinará la evaluación de la información. Una de las grandes ventajas es que procesada la información en el SOM, los vectores pueden ser utilizados para representar los datos en salidas visuales que mejoran la interpretación de acuerdo a su distribución en la última capa del SOM.

Etapas 3. Estimación del número de clústeres en el SOM y validación de los clústeres.

Algunos resultados de clusterización son usualmente abiertos a la interpretación, es por ello que es importante validar la información. Una clusterización valida cuando la distancia entre elementos del clúster es pequeña y distancia entre clústeres es grande. Para esta situación el índice de Davies-Bouldin nos indica cual es la mejor clusterización.

$$I_{DB} = \frac{1}{C} \sum_{i=1}^C \max \left\{ \frac{S_C(C_i) + S_C(C_j)}{d_{ce}(C_i, C_j)} \right\} \quad (23)$$

Etapas 4. Comparación de Vector Prototipo con el Repositorio usando la RNP.

Terminado el proceso de validación y de reducción de dimensionalidad, se realiza el proceso de evaluación de los datos utilizando una RNP, la cual previamente fue entrenada con los vectores prototipo de referencia y que son almacenados en el repositorio. Cabe mencionar que este repositorio es dinámico, al entrar un nuevo caso, la RNP se entrena nuevamente y adapta sus pesos para la identificación del nuevo caso.

Etapas 5. Detección y Diagnóstico.

En esta etapa final, de acuerdo a la calidad del entrenamiento y la eficiencia de la evaluación, el sistema separa las fuentes de DP's y emite el diagnóstico del tipo de fuente presente en la medición, ya que como se menciona anteriormente, cada fuente de DP's está asociada a un defecto. En caso de que en la medición solo se presente el ruido de fondo, significa que el objeto medido está libre de DP's, por lo tanto, cumple con los requerimientos operativos y de confiabilidad.

6.1 Descripción y aplicación de la interfaz usuario del SBC.

En la Figura 58 se presenta la interfaz gráfica desarrollada para aplicación del SBC, consta de botones funcionales de acuerdo al procesamiento. Los datos son almacenados por el equipo de medición de DP's en formato *.txt, después con la aplicación en segundo plano se abre el archivo, se le quita el ruido de fondo, se elimina la entropía, se aplica el SOM para determinar los clusters de interés, se grafica con el SOM 3D y se realiza el diagnóstico. Existen 2 ventanas adicionales para determinar el nivel de energía producida por el tipo de descarga con el cual se define el nivel de daño provocado y por último se continúa con el análisis estadístico para validarlo con otras técnicas tal y como se desarrolló en [47].

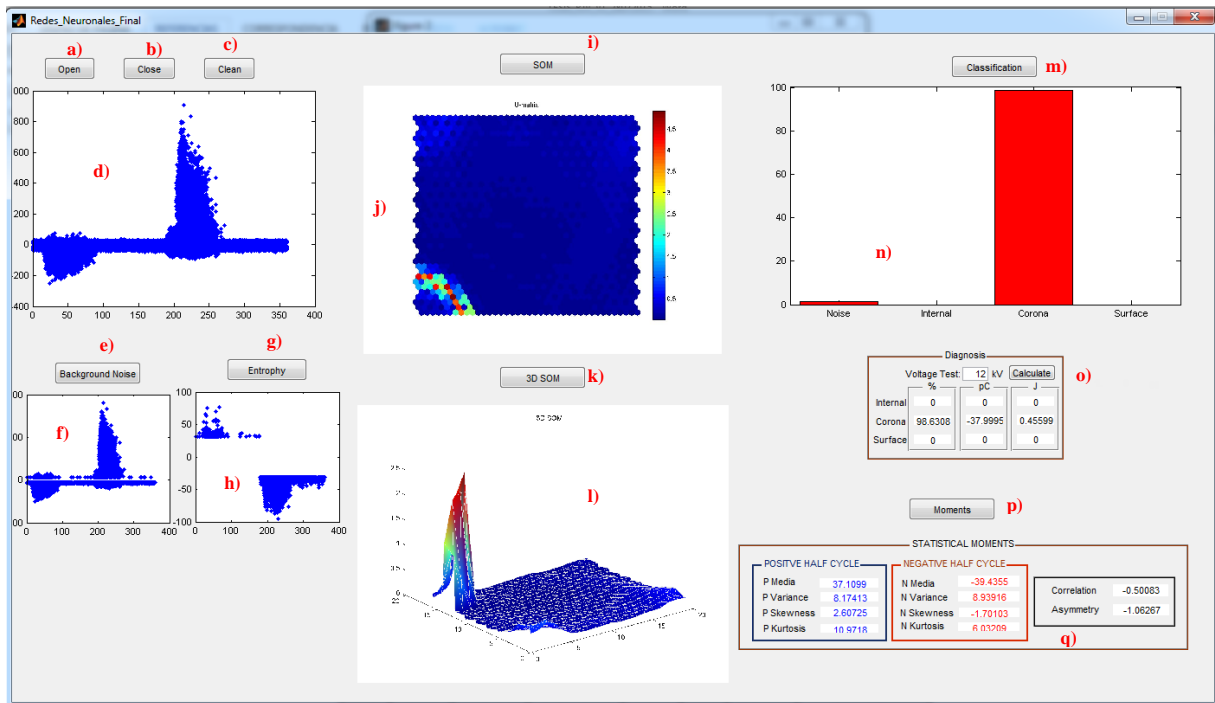


Figura 58.- Interfaz usuario del SBC.

A continuación, se describen los elementos de la aplicación del SBC.

- a) Botón para abrir archivo. Tiene la funcionalidad de ubicar el directorio donde se guardó el vector de datos.
- b) Cierra la aplicación después de realizar el análisis.
- c) Limpia la pantalla para un nuevo análisis.
- d) Gráfico del vector de datos original que mapea el de Nivel de Descargas VS Ángulo de Fase.
- e) Elimina el ruido de fondo previamente determinado en la calibración del sistema.
- f) Gráfico del vector de datos con datos filtrados del ruido de fondo.
- g) Eliminación de entropía y datos redundantes.
- h) Gráfico de vector de datos después de eliminación de entropía.
- i) Procesamiento de datos, reducción de dimensionalidad y determinación de clúster dentro del vector de datos.
- j) Gráfico de la matriz de distancias unificadas.
- k) Aplicación de algoritmo de 3D-SOM para nueva representación de los datos.
- l) Gráfico del 3D-SOM.
- m) Clasificación y diagnóstico.

- n) Gráfico de barras de determina el % de datos de acuerdo al tipo de descarga presente en la totalidad del vector de datos.
- o) Diagnóstico de acuerdo al nivel de energía de acuerdo al tipo de descarga.
- p) Aplicación del algoritmo de momentos extendidos.
- q) Tabla de valores de los momentos extendidos como herramienta adicional para el diagnóstico.

Para validar la funcionalidad del SBC se aplica en los datos de entrenamiento, con el fin de determinar si es capaz de reconocer el tipo de descarga generada con los defectos artificiales.

Como se puede observar en Figura 59, Figura 60 y Figura 61, el SBC responde adecuadamente en la evaluación y clasificación de los datos de entrenamiento. Para realizar una mejor validación se utiliza una matriz de confusión, la cual es una herramienta de visualización que se emplea para validar el aprendizaje del sistema. Cada columna de la matriz representa el número de predicciones de cada clase, mientras que cada fila representa a las instancias en la clase real. Uno de los beneficios de las matrices de confusión es que facilitan ver si el sistema está confundiendo dos clases.

Si en los datos de entrada el número de muestras de clases diferentes cambia mucho la tasa de error del clasificador no es representativa de lo bien que realiza la tarea el clasificador. Se utilizan 20 muestras para cada clase de descarga con el fin de determinar si existe un sesgo en el clasificador.

Tabla 12.- Matriz de confusión para validación del SBC

	DESCARGA INTERNA	DESCARGA SUPERFICIAL	DESCARGA EXTERNA
DESCARGA INTERNA	20	0	0
DESCARGA SUPERFICIAL	0	20	0
DESCARGA EXTERNA	0	0	20

En la Tabla 12 se observa el resultado de la validación del SBC, con lo cual se garantiza con un 0% de error que en el proceso de clasificación no se confunden las clases y por tanto se aseguran que en la interpretación del SBC no habrá evaluaciones erróneas.

Desde el punto de vista de los diagnósticos basados en experiencia, el mayor sesgo en confusión siempre se ha presentado en la clasificación de descarga interna y descarga superficial, ya que los comportamientos son simétricos en ambos semiciclos del vector de datos, el único marco de referencia para estos caso era la interpretación de los cruces por cero, sin embargo, en las pruebas en sitio, debido al gran ruido de fondo, los cruces por cero están demasiado contaminados, por lo tanto se hace muy compleja la interpretación.

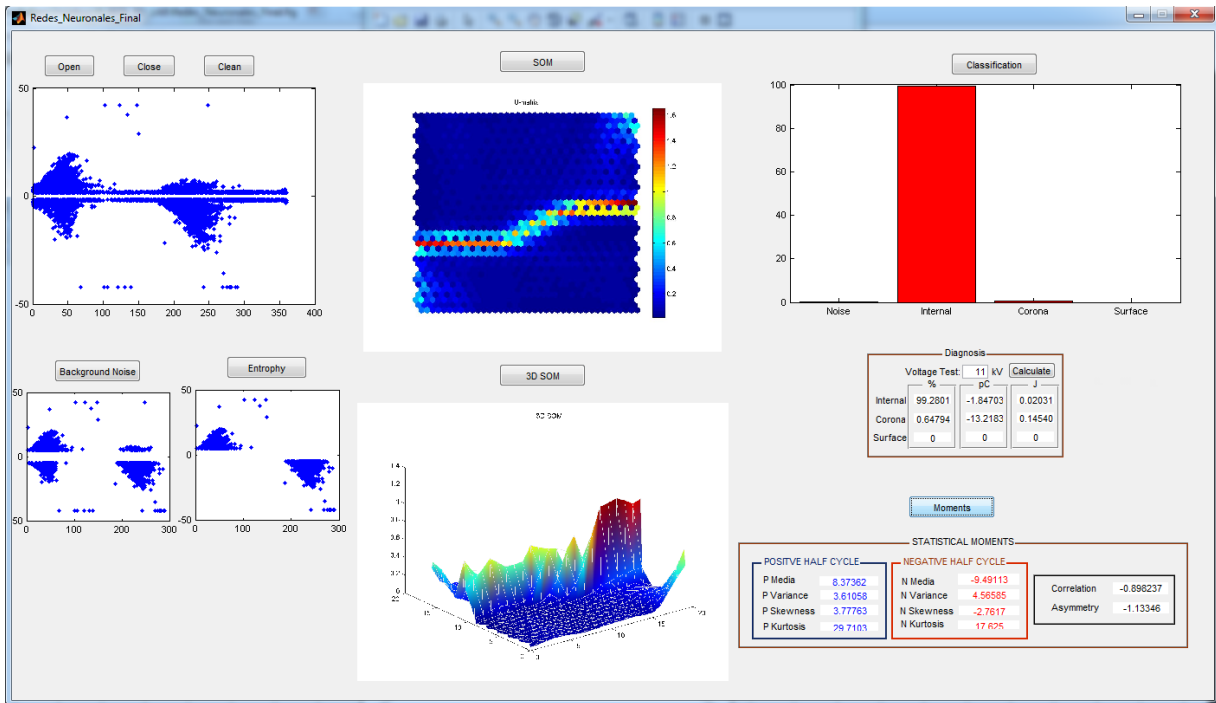


Figura 59.- Validación del SBC con el vector de descarga interna

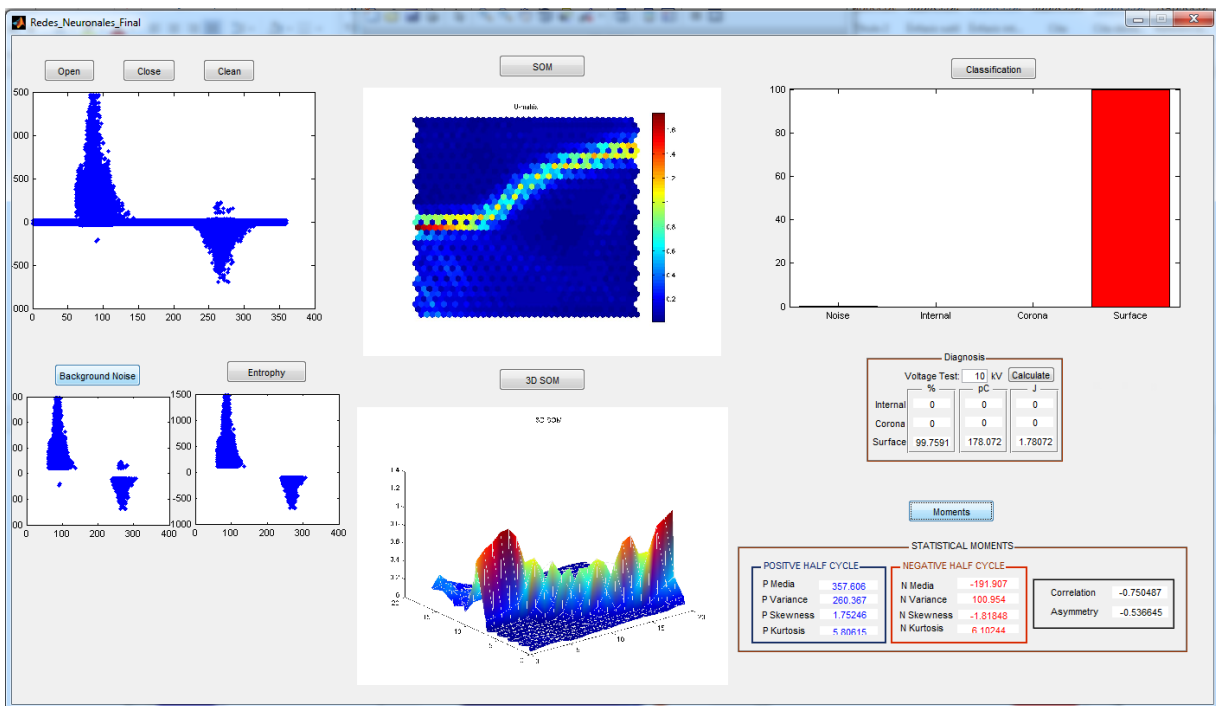


Figura 60.- Validación del SBC con el vector de descarga superficial

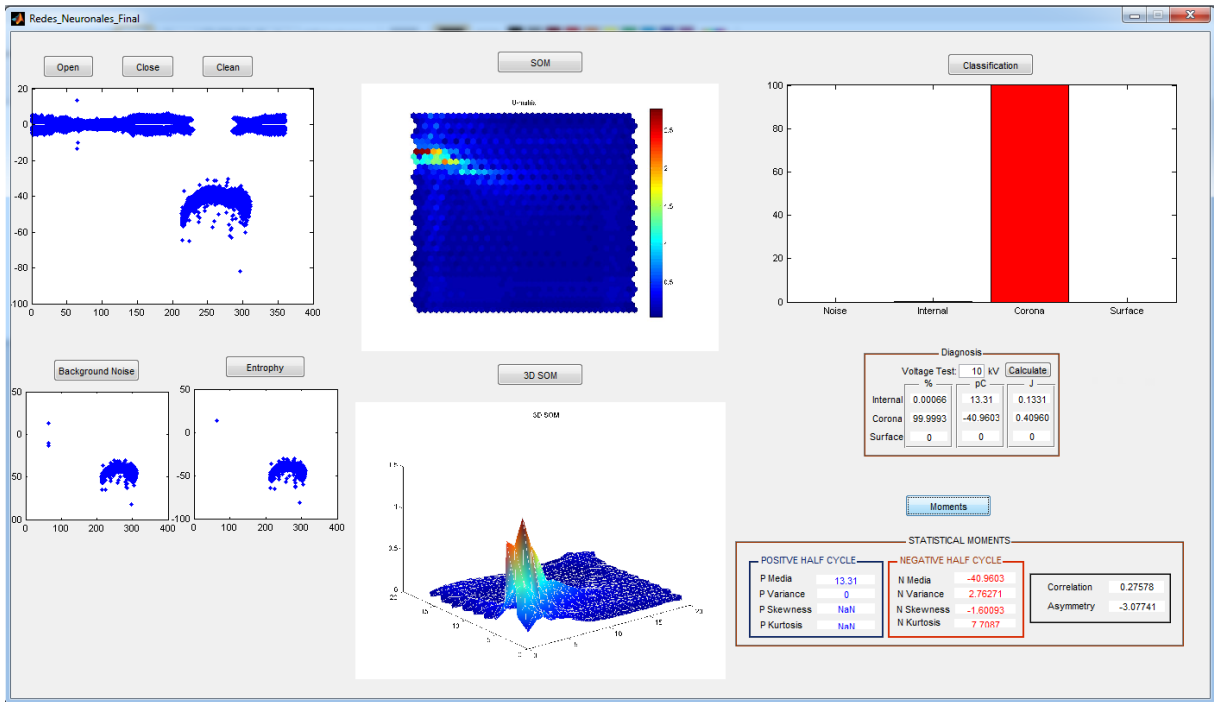


Figura 61.- Validación del SBC utilizando el vector de descarga externa o efecto corona

6.2 Evaluación del sistema en casos de estudio.

A continuación, se presenta los resultados experimentales originales sobre la aplicación del SBC en pruebas en sitio, ya que los niveles de ruido son demasiado elevados y por tanto es requerida una adecuada separación de fuentes de DP's y poder encontrar cual es la de interés en el análisis.

El primer caso de estudio presentado es relacionado con la evaluación de una red subterránea de 115 kV que presentaba un alto nivel de DP's durante una prueba de alta tensión en la Fase B. Se obtienen 15 vectores de medición de 2 minutos, con el fin de registrar el mayor número de datos. Cabe resaltar nuevamente que el sistema es robusto a vectores de gran tamaño, ya que la etapa de pre-procesamiento reduce la dimensionalidad sin pérdida de topología.

Se observan niveles de DP's en el orden de 15000 pC, siendo el criterio de aceptación 10 pC, el patrón obtenido tiene envolventes bien definidas en magnitud y fase, no presenta simetría definida, lo cual nos lleva a plantear las siguientes hipótesis:

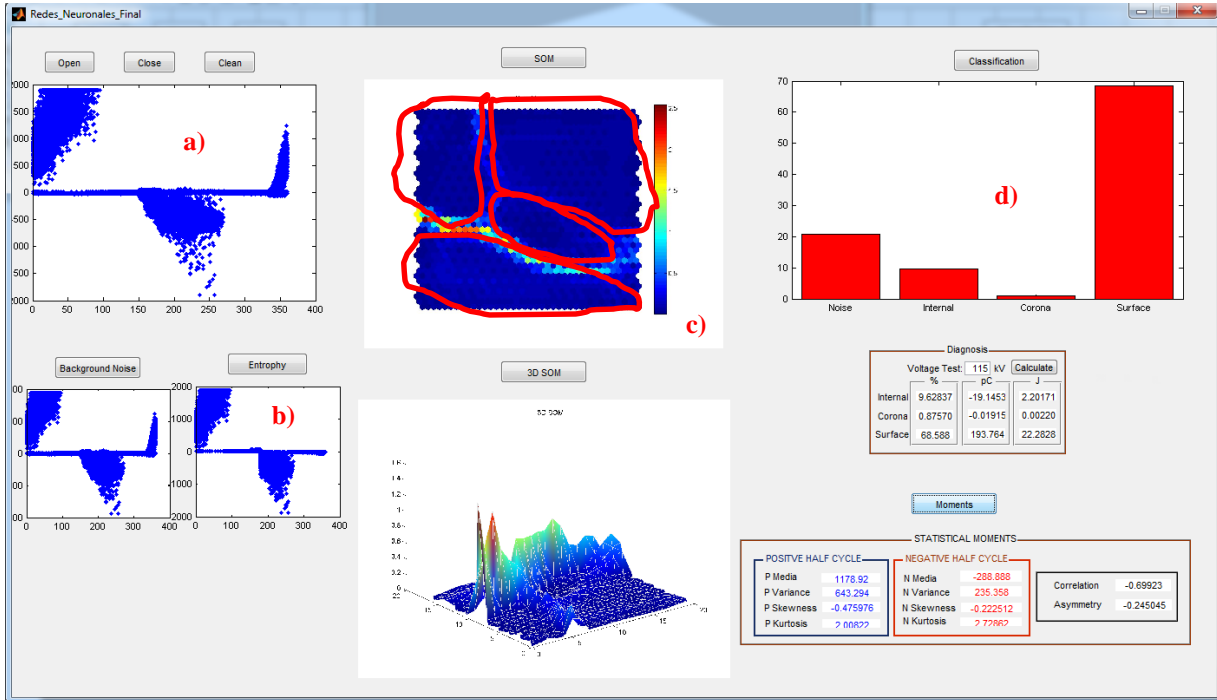


Figura 62.- Medición obtenida en la fase B de una línea de 115 kV en sitio.

En la Figura 62 se presenta el análisis de los datos utilizado el SBC en a) se presenta el vector de datos obtenido, en b) se presenta el vector ya acondicionado al eliminar ruido de fondo y entropía en el c) se presenta la matriz de distancia unificada mostrando 4 clúster representativos, en d) se observa la clasificación de los 4 clúster. Como se observa en la salida del SBC se obtiene que la clase de DP más representativa es la descarga superficial, por lo que con ello se determina que existe descarga superficial entre dos puntos a diferente potencial. Se localiza por un método acústico el punto de falla dentro de un empalme y se procede a la disección del mismo con el fin de validar el resultado obtenido.



Figura 63.- Disección de un empalme de 115 kV con presencia de descarga superficial

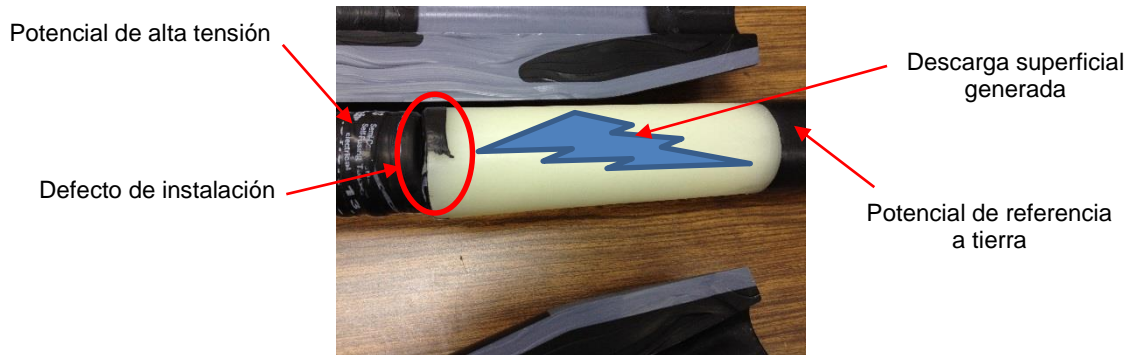


Figura 64.- Evidencia de deficiente instalación de accesorios dentro del empalme

En la Figura 63 se observa la disección del empalme para determinar el problema que fue detectado por el SBC. El caso de éxito es presentado en la Figura 64, ya que se encuentra que efectivamente fue aplicada una cinta semiconductora en el conector mecánico del empalme, sobresaliendo del aislamiento y presentando un efecto punta al campo eléctrico, este defecto al estar confinado en el empalme tendía a generar un camino de erosión, generando una falla súbita de la red subterránea.

Otro ejemplo de caso de éxito de aplicación del SBC fue en un cable de potencia que alimenta un tablero de media tensión un centro bancario muy importante en Querétaro (Figura 65), en el cual por mantenimiento se solicita a el LAPEM la evaluación del sistema de cables de energía. Se instala el equipo de medición y se capturan datos de 2 minutos en cada fase, detectando que en la fase B se presentan niveles de hasta 40 pC, siendo el criterio de calificación 5 pC.



Figura 65.- Acometida de media tensión en un tablero de potencia

Las hipótesis son validadas con las mediciones obtenidas, las cuales analizadas con el SBC obteniendo los resultados presentados en la siguiente figura.

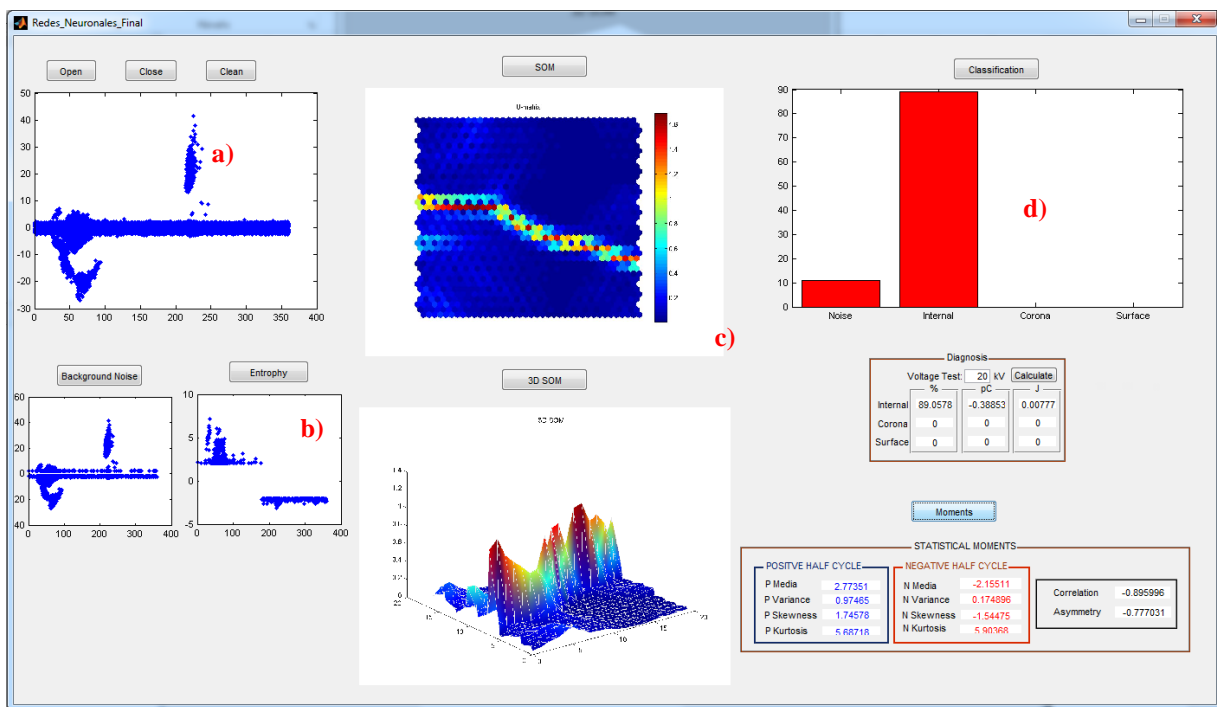


Figura 66.- Mediciones obtenidas en el tablero de media tensión del centro bancario

En la Figura 66 se presenta el análisis de los datos utilizado el SBC en a) se presenta el vector de datos obtenido, en b) se presenta el vector ya acondicionado al eliminar ruido de fondo y entropía en el c) se presenta la matriz de distancia unificada mostrando 2 clúster representativos, en d) se observa la clasificación de los 2 clúster. Como se observa en la salida del SBC se obtiene que la clase de DP más representativa es la descarga interna, por lo que con ello se determina descargas por cavidades en los accesorios. Se localiza por un método acústico el punto de falla dentro de una terminal y se procede a la disección del mismo con el fin de validar el resultado obtenido.

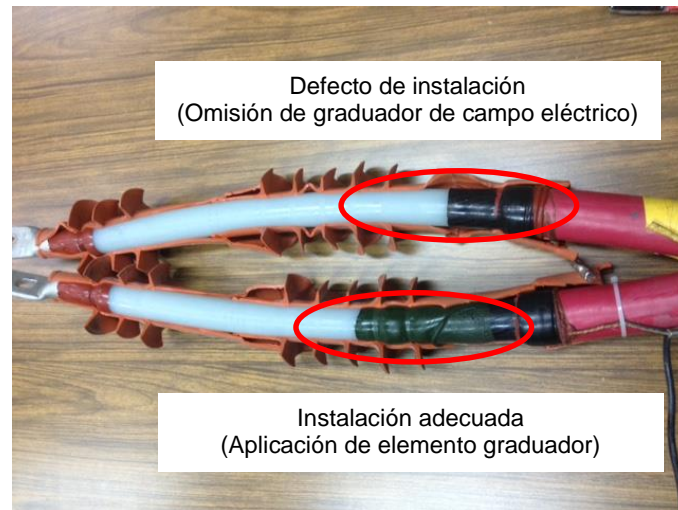


Figura 67.- Evidencia de defecto de instalación en terminal de media tensión.

En la Figura 67 se observa la disección de las terminales del sistema identificado como niveles de DP's fuera del rango de seguridad, se analiza a detalle la colocación de los elementos de la terminal, desde el conector mecánico final de la terminal hasta los traslapes entre las secciones de la terminal. Un punto importante a revisar son los cortes de la transición de la semiconductor sobre el aislamiento, ya que si no es realizado de manera adecuada, se pueden presentar fallos por concentración de campo eléctrico en el punto del corte, es por ello que es de gran importancia la aplicación de la masilla graduadora de manera que no queden huecos, puntas o aristas que provoquen esa concentración. Es por ello que se detecta una irregularidad en la instalación, ya que como se observa en el comparativo de las terminales, en la transición entre la semiconductor sobre el aislamiento y el XLPE no se aplicó la masilla graduadora (tal como lo indica el manual de instalación de la terminal), lo cual es inadecuado, ya que no permite graduar adecuadamente el campo eléctrico por efecto de cambio de permitividad y las líneas de equipotenciales de campo eléctrico no se concentran un punto específico. Por tal motivo se presentan descargas internas, lo que produce una degradación gradual y por tanto minimizando la confiabilidad del sistema.

6.3 Validación de datos que contienen error

En este experimento de validación del SBC, se obtiene un conjunto de datos de entrenamiento que contiene 150 vectores de datos, es decir, vectores de datos que contengan las cuatro instancias a evaluar (descarga, interna, externa, superficial y ruido de fondo) y los datos de prueba igual al conjunto de datos de entrenamiento, que contiene 150 vectores de datos con las instancias a evaluar. En los datos de entrada de un sistema de reconocimiento de DP's es inevitablemente que contenga de error

aleatorio como varianza en los transductores, los errores humanos, etc, lo que podría dar lugar a incertidumbres de los datos. Para incluir el ruido e incertidumbres, 150 conjuntos de datos de prueba fueron creados agregando de 5% a 30% de error aleatorio, de manera uniforme, con el fin de evaluar la capacidad de tolerancia a fallas del SBC propuesto. En la evaluación se comparó con los sistemas independientes del SOM y del BPN.

Tabla 13.- Evaluación comparativa de diferentes métodos de reconocimiento y clasificación de patrones de DP's

Porcentaje de Error (%)	Exactitud de los diferentes métodos de reconocimiento de patrones de DP's		
	SOM	BPN	SBC
± 0%	100%	100%	100%
± 5%	99%	100%	100%
± 10%	95%	100%	100%
± 15%	90%	98%	100%
± 20%	88%	95%	100%
± 25%	85%	90%	98%
± 30%	79%	90%	97%

En la Tabla 13 se presentan los resultados comparativos de diferentes métodos de reconocimiento de patrones y clasificación de patrones de DP's, demostrando que el SBC es robusto ante incertidumbres y variabilidad en la configuración del sistema de medición, error aleatorio y ruido del entorno, lo que demuestra una gran confiabilidad en el diagnóstico y clasificación de fuentes de DP's.

CAPÍTULO 7: CONCLUSIONES

Una de las necesidades más apremiantes es desarrollar metodologías y herramientas que ayuden a los ingenieros de todos los procesos productivos a aprender e interpretar los datos de sus procesos con el fin de tomar decisiones. El SBC desarrollado en este trabajo nos ayuda a automatizar el proceso de diagnóstico y evaluación de equipos eléctricos tanto en instalaciones operativas (mantenimiento y monitoreo) como en fábricas (control de calidad).

Los resultados obtenidos fueron validados con datos para entrenamiento y matriz de confusión, en la cual se presentaron resultados muy satisfactorios, se realizó la evaluación de datos obtenidos tanto en laboratorio (nivel de ruido controlado) como datos en sitio (ruido no controlado). Para demostrar la robustez y capacidad de clasificación se realiza la validación de datos con error y se compara con otras técnicas reportadas en la literatura.

En este trabajo se presenta a detalle cada una de las etapas implementadas para procesar la información, de manera que en cada una de ellas se desarrolla un proceso de validación que incrementa el nivel de confianza del resultado y provee una decisión justificada.

Considerando las hipótesis planteadas al inicio de la investigación:

H1. Caracterizando patrones de DP's en laboratorio nos ayudará a extraer atributos para identificar patrones de defectos del sistema con condiciones de prueba de campo.

H2. El desarrollo de un sistema inteligente de identificación de patrones de DP's hará más eficiente la tarea de diagnóstico en lugares con condiciones adversas para la medición.

Con los datos obtenidos en laboratorio a condiciones controladas y con defectos artificiales, se obtuvieron resultados muy satisfactorios al realizar el entrenamiento del SBC, por lo que los algoritmos lograron clasificar y reconocer fuentes de DP's cuando el objeto bajo prueba era analizado se y presentaba actividad de DP's durante la prueba, con esto la hipótesis H1 es aceptada, ya que gracias a esa información obtenida previamente se puede asegurar que el sistema realmente está "Basado en Conocimiento".

Este SBC nos da la capacidad de poder resolver el problema de reconocimiento de patrones aun cuando en la misma muestra se presenten diferentes fuentes de DP's, lo cual hace que el sistema pueda trabajar en lugares donde la condiciones de medición son adversas, ya que el nivel de ruido es filtrado y lo residual en el grupo de datos es clasificado en un clúster específico de ruido, siendo separado de las demás fuentes de interés, es por ello que la H2 es aceptada, ya que al separar las diferentes fuentes de DP's se puede dar un diagnóstico fehaciente y validado.

7.1 Contribuciones de este trabajo

Las principales contribuciones de este trabajo son:

- Evaluación exhaustiva de los algoritmos de aprendizaje competitivo para determinar el mejor algoritmo de clusterización y reducción de dimensionalidad sin perder la topología de los datos y que mejore el tiempo de respuesta del pre-procesamiento. Los resultados fueron presentados en [6] [48] [49] [50].
- Estudios de las regiones de operabilidad del SOM, demostrando que son altamente no lineal y la convergencia depende de los datos de entrada y el algoritmo de aprendizaje competitivo, por lo que la sintonización óptima del SOM se desarrolló utilizando el método de la función de deseabilidad. Los resultados fueron presentados en [46].
- Obtención de una base de datos de DP's procesada, validada, ordenada, y con gran nivel de aplicabilidad al estar estructurada en archivos *.txt, con el fin de sentar las bases de un grupo de investigación en el tema, hoy día se han desarrollado 2 tesis de maestría y está en proceso una de doctorado, en las cuales se ha seguido explotado la información.
- Desarrollo de un novedoso SBC que ofrece aplicación para evaluación y diagnóstico en fábricas de producción o instalaciones operativas, ya que es robusto a ruido ambiental y muy eficiente para separación y clasificación de fuentes DP's en aislamientos eléctricos de tipo gas, líquido, sólido y mixto. Actualmente se encuentra en trámite de registro como Modelo de Utilidad a nombre de CFE y CIATEC.

7.2 Trabajos futuros

Este trabajo abre la perspectiva de análisis de datos basadas en herramientas heurísticas, con el fin de desarrollar modelos de utilidad para CFE. Como consecuencia a esta tesis se proponen las siguientes líneas de investigación:

- Estudiar las características de las DP's como series de tiempo, esto nos da una perspectiva de vida residual, sin embargo, cabe mencionar que el atributo a analizar no es magnitud sino

vector, lo que hace que el análisis sea en base al cambio del patrón en el tiempo. En otras palabras, análisis de series de tiempo basadas en patrones.

- Desde el punto de vista de las cartas de control, estas son desarrolladas en base a magnitudes o atributos, sin embargo, de acuerdo a la naturaleza de los datos de DP's, la magnitud no es representativa, es representativo el patrón de datos, es por ello que se propone desarrollar un método de control estadístico de la calidad basada en cartas de control de patrones de datos.
- Los datos obtenidos de DP's fueron desarrollados a frecuencias entre 20 y 300 Hz, sin embargo, desde el punto de vista de los sistemas móviles de diagnóstico, existe una gran tendencia a los sistemas de impulso oscilante amortiguado, los cuales provocan distribuciones estadísticas de DP's en cada semiciclo, por lo cual hoy día no se ha podido cuantificar el nivel de daño del objeto bajo prueba. Desarrollar en base a los datos obtenidos de DP's un SBC para evaluación y diagnóstico utilizando impulso amortiguado oscilante.
- Con los datos experimentales obtenidos, validar el desempeño de otros algoritmos heurístico, metaheurísticos, mateheurísticos y bioinspirados, con el fin de encontrar alternativas de evaluación y diagnóstico de acuerdo al estado del arte de los algoritmos.

CAPÍTULO 8: PRODUCCIÓN CIENTÍFICA

8.1 Artículos en Journal Citation Report (JCR)

1. **Rubén Jaramillo-Vacio**, C. Alberto Ochoa Ortiz-Zezzatti, Armando Rios-Lira. **Knowledge-based System for partial discharge recognition on electrical insulation**. Applied Soft Computing. ISSN: 1568-4946 www.elsevier.com/locate/asoc. Aceptado. Factor de Impacto: 2.810.
2. Alberto Ochoa-Ortiz, Francisco Ornelas-Zapata, Lourdes Margain-Fuentes, Miguel Gastón Cedillo-Campos, Jons Sánchez-Aguilar, **Rubén Jaramillo-Vacio**. **Capacited vehicle routing problem for PSS uses based on ubiquitous computing: An emerging markets approach**. DYNA 82 (191), pp. 20-26, June, 2015 ISSN: **2346-2183**. Factor de Impacto: 0.217
DOI: <http://dx.doi.org/10.15446/dyna.v82n191.51141>

8.2 Congresos Internacionales

1. **Rubén Jaramillo-Vacio**, C. Alberto Ochoa Ortiz-Zezzatti, S. Jöns, Sergio Ledesma-Orozco. **Diagnosis test of power cables using a resonant test system and analysis of partial discharge on site**. 12th International Conference on Power Electronics – CIEP August 2010, San Luis Potosi, SLP. 2010.
2. **Rubén Jaramillo-Vacio**, Alberto Ochoa-Zezzatti, Julio Ponce, **Competitive Learning for Self Organizing Maps used in Classification of Partial Discharge**. Congreso Internacional de Cómputo en Optimización y Software. CICOS'2011. Cuernavaca, Mor. December 2011. **Aceptado para su publicación en la Revista Electrónica “PROGRAMACIÓN MATEMÁTICA Y SOFTWARE” Editada por la UAEM, en Volumen 5, número 1.**
3. **Rubén Jaramillo-Vacio**, Alberto Ochoa-Zezzatti, Armando Rios-Lira, Julio Ponce: **Self Organizing Maps used in classification of partial discharge**. IEEE CONIELECOMP January 2012, Tonantzintla, Puebla. 2012
4. **Rubén Jaramillo-Vacio**, Alberto Ochoa-Zezzatti, Armando Ríos-Lira: **Evaluación de Algoritmos Competitivos en Mapas Autoorganizados para la Clasificación de Patrones de Descargas Parciales**. 1er Congreso Internacional de Investigación e Innovación 2012. Cortázar, Guanajuato. Abril 2012.

Rubén Jaramillo-Vacio, Alberto Ochoa-Zezzatti, Armando Rios-Lira: **The Self Organizing Maps used in classification of partial discharge: a comparative study of algorithms**. IEEE RVP 2012, Acapulco, Guerrero. July 2012. **Aceptado para su publicación en la Revista Electrónica de CFE “TECNOLAB” Editada por la CFE-LAPEM, en Segunda edición Septiembre 2012.**

8.3 Congresos Internacionales con estricto arbitraje

1. **Rubén Jaramillo-Vacio**, Alberto Ochoa-Zezzatti, S. Jöns, Sergio Ledezma-Orozco, Camelia Chira. **Diagnosis of partial discharge using Self Organizing Maps and Hierarchical Clustering – An approach**. The 6th International Conference on Hybrid Artificial Intelligence

Systems, HAIS'2011. Wroclaw, Polonia. 2011. Publicado en Proceeding, Part I, Lecture Notes in Computer Science, Vol. 6678. ISBN 978-3-642-21218-5

2. **Rubén Jaramillo-Vacio**, Alberto Ochoa-Zezzatti, Armando Rios-Lira: **Comparison of Competitive Learning for SOM Used in Classification of Partial Discharge**. The 7th International Conference on Hybrid Artificial Intelligence Systems, HAIS'2012. Salamanca, España. 2012. Publicado en Proceeding, Part I, Lecture Notes in Computer Science, Vol. 7206. ISBN 9783642289316 3642289312
3. **Rubén Jaramillo-Vacio**, Alberto Ochoa-Zezzatti, Armando Rios-Lira: **A Comparative study of Partial Discharge by Classification's kind**. The 12th International Conference on Hybrid Intelligence Systems, HIS'2012. Pune, India. 2012. Publicado en IEEEXplorer Proceeding. ISBN 978-1-4673-5116-4

8.4 Revistas en Línea

1. **Rubén Jaramillo-Vacio**, Alberto Ochoa-Zezzatti, Armando Rios-Lira "Optimal Tuning Parameters of Self Organizing Maps for Classification of Partial Discharge Using Response Surface Methodology" / **International Journal of Engineering Research and Applications (IJERA)** ISSN: 2248-9622 www.ijera.com Vol. 3, Issue 4, Jul-Aug 2013, pp.974-984
2. **Rubén Jaramillo-Vacio**, Alberto Ochoa-Zezzatti, Julio Ponce, **Competitive Learning for Self Organizing Maps used in Classification of Partial Discharge**. **Revista Electrónica "PROGRAMACIÓN MATEMÁTICA Y SOFTWARE"** ISSN: 2007-3283, en Volumen 5, número 2, pp. 6-12. 2013.
<http://www.progmat.uaem.mx:8080/Vol5num2/vol5num2art2.pdf>

8.5 Capítulos de Libro

1. **Libro: Knowledge-Oriented Applications in Data Mining**
Publisher: InTech, Edited by: Kimito Funatsu. ISBN 978-953-307-154-1.
CAPÍTULO: New Implementations of Data Mining in a Plethora of Human Activities.
Alberto Ochoa, Julio Ponce, Francisco Ornelas, **Rubén Jaramillo**, Ramón Zatarain, María Barrón, Claudia Gómez, José Martínez, Arturo Elias.
Website: <http://www.intechopen.com/articles/show/title/new-implementations-of-data-mining-in-a-plethora-of-human-activities>
2. **Libro: Logistics Management and Optimization through Hybrid Artificial Intelligence Systems**
Publisher: IGI GLOBAL, Edited by: Carlos Alberto Ochoa Ortiz Zezzatti (Juarez City University, Mexico), Carmelia Chira (Babes-Bolyai University Cluj-Napoca, Romania), Arturo Hernandez (Mexico) and Miguel Basurto (UAEM, Mexico). ISBN13: 9781466602977, ISBN10: 146660297X, EISBN13: 9781466602984
CAPÍTULO: Data mining applications in the electrical industry
Ruben Jaramillo-Vacio, Carlos A. Ochoa-Zezzatti, Armando Rios-Lira
Website: <http://www.igi-global.com/book/logistics-management-optimization-through-hybrid/59725#author-editor-biography>

3. Libro: Data Mining Applications in Engineering and Medicine

Publisher: INTECH, Edited by Adem Karahoca, ISBN 978-953-51-0720-0, 336 pages, Publisher: InTech, Chapters published August 29, 2012 under CC BY 3.0 license DOI: 10.5772/2616

CAPÍTULO: Explaining Diverse Application Domains Analyzed from Data Mining Perspective

Alberto Ochoa, Daniel Azpeitia, Petra Salazar, Emmanuel García, Miguel Maldonado, **Rubén Jaramillo**, Jöns Sánchez, Javier González, Claudia Gómez, Julio Ponce, Sayuri Quezada, Francisco Ornelas, Arturo Elías, Edgar Conde, Víctor Cruz and Lourdes Margain
Website: <http://www.intechopen.com/books/data-mining-applications-in-engineering-and-medicine/explaining-diverse-application-domains-analyzed-from-data-mining-perspective>

8.6 Ponencias en Congreso o Academia

1. Ponente: **M. en I. Rubén Jaramillo Vacio**
Titulo: Diagnostico en Redes Eléctricas Subterráneas de Alta Tensión.
Universidad Tecnológica de Altamira - Octubre 2009
2. Ponente: **M. en I. Rubén Jaramillo Vacio**
Titulo: Propuesta de reducción de pérdidas en transformadores de distribución.
Conferencia Internacional de Grandes Redes Eléctricas (CIGRE) - Mayo 2010
3. Ponente: **M. en I. Rubén Jaramillo Vacio**
Titulo: Diagnostico en Redes Eléctricas Subterráneas de Alta Tensión.
Seminario Posgrado en Ingeniería Eléctrica UASLP - Mayo 2010
4. Ponente: **M. en I. Rubén Jaramillo Vacio**
Titulo: Diagnostico en Redes Eléctricas Subterráneas de Alta Tensión con Sistema Resonante.
3ra Semana de Tecnología del ITESI - Agosto 2010
5. Ponente: **M. en I. Rubén Jaramillo Vacio**
Titulo: Redes Neuronales No Supervisadas.
4ta Semana de Tecnología del ITESI - Agosto 2011
6. Ponente: **M. en I. Rubén Jaramillo Vacio**
Titulo: Evaluación del Aislamiento de Alto Voltaje utilizando SVM.
4ta Congreso Internacional de Ingenierías Mecánica, Eléctrica, Electrónica, Mecatrónica y Computacional - Septiembre 2011
7. Ponente: **M. en I. Rubén Jaramillo Vacio**
Titulo: Estudio del Envejecimiento en Materiales Poliméricos para Conductores Eléctricos.
XIV Jornadas Nacionales de Ingeniería de Materiales – Marzo 2012
8. Ponente: **M. en I. Rubén Jaramillo Vacio**
Titulo: Evaluación técnica de transformadores de distribución.
CONALEP – Agosto 2012
9. Ponente: **M. en I. Rubén Jaramillo Vacio**
Titulo: Análisis Técnico – Económico de Transformadores.
4ta Semana de Tecnología del ITESI - Agosto 2012

10. Ponente: **M. en I. Rubén Jaramillo Vacio**
Titulo: Aplicación de la Inteligencia Artificial en el Diagnóstico de Equipos de Alta Tensión.
4ta Semana de Tecnología del ITESI - Agosto 2012
11. Ponente: **M. en I. Rubén Jaramillo Vacio**
Titulo: Evaluación técnica de transformadores de distribución de potencia.
Semana de Ingeniería de la UASLP – Noviembre 2012
12. Ponente: **M. en I. Rubén Jaramillo Vacio**
Titulo: Aplicación de Inteligencia Artificial en la evaluación de Equipos de Alta Tensión.
ELECTROMECC'13 UDL Campus León – Noviembre 2013

REFERENCIAS

- [1] J. A. Vázquez-López, «Desarrollo de un método de Control Estadístico para variables aleatorias con distribuciones simétricas y asimétricas, basado en el reconocimiento de patrones por Redes Neuronales Artificiales,» 2009.
- [2] G. J. Myatt, *Making Sense of Data. A practical Guide to Exploratory Data Analysis and Data Mining*, Wiley-Interscience, 2007.
- [3] T. K. T. M. Ch. Kim, «Change in PD Pattern with Aging,» *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 11, nº 5, 2004.
- [4] C. G. Azcarraga-Ramos, «Medición distribuida de descargas parciales en redes subterráneas utilizando técnicas de banda ultra-ancha,» 2004.
- [5] A. C. G. C. Montanari, «A New Approach to Partial Discharge Testing of HV Cable Systems,» *IEEE Electrical Insulation Magazine*, vol. 22, nº 1, 2006.
- [6] O.-Z. C. J. S. L.-O. S. Jaramillo-Vacio R., «Diagnosis test of Power Cables using a Resonant Test System and Analysis of Partial Discharge On Site,» *12th International Conference on Power Electronics – CIEP 2010*, pp. 114-117, 2010.
- [7] IEC-60270, *High Voltage test techniques - Partial discharge measurements*, 2000.
- [8] H. Edin, «Partial discharge studies with variable frequency of the applied voltage,» 2001.
- [9] L. Wills, *Electrical Power Cable Engineering*, Ed. Marcel Dekker Inc., 1999.
- [10] B.-A. C. M. J.C. Bapt, «Corona frequency analysis in artificial cavities in epoxy resins,» *1978 Annual Report, Conference on Electrical Insulation and Dielectric Phenomena*, pp. 403-407, 1978.
- [11] G. S. S.A. Boggs, «Fundamental Limitations in the Measurement of Corona and Partial Discharge,» *IEEE Transactions EI-17*, nº 2, pp. 143-150, 1983.
- [12] R. V. Brunt, «Stochastic properties of PD phenomena,» *IEEE Transactions on Electrical Insulation*, vol. 26, pp. 902-948, 1991.
- [13] A. J. E. G. P. Cichecki, «Statistical Approach in Power Cables Diagnostic Data Analysis,» *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 15, nº 6, 2008.
- [14] F. K. W. Yan, «Features Selection for Partial Discharges Diagnosis,» *Proceedings of 12 th SPIE: Health Monitoring and Smart Nondestructive Evaluation of Structural and Biological Systems*, nº 5, 2005.
- [15] F. Western, «Condition Assessment of power cables using partial discharge diagnosis at damped AC voltage,» 2005.

- [16] F. W. E. Gulski, «PD Knowledge rules for insulation Condition Assessment of distribution power cables,» *5th Colloquium on Partial discharge Measurement*, 2007.
- [17] H. T. Y. W. Y. Chang, «Application of Self Organizing Map approach to partial discharge pattern recognition of cast resin current transformers,» *Wsea Transactions on computer research*, vol. 3, nº 3, pp. 151-160, 2008.
- [18] E. O. O. S. T. Kohonen, «Engineering Applications of the Self-Organizing Map,» *Proceedings of the IEEE*, vol. 84, nº 10, 1996.
- [19] D. M. González, «Redes Neuronales Artificiales y SOM,» 2004.
- [20] M. R. Sánchez, «Nuevos métodos para análisis visual de mapas auto-organizativos,» 2004.
- [21] J. M. Alonso, «Un Nuevo Modelo Autoorganizado aplicado a la resolución de Problemas de Geometría Computacional,» 2008.
- [22] M. Y. Márquez-Montenegro, «Determinación de perfiles de generación de RSD por tipología familiar a través de minería de datos: estudio de caso a tres comunidades de Mexicali, B. C.,» 2009.
- [23] J. A. Hernández-Aguilar, «Diseño e implementación de un sistema de evaluación remoto con seguridad avanzada para universidades utilizando minería de datos,» 2008.
- [24] J. P. R. H. J. C. Hernández-Mejía, «Determining Routes for the analysis of partial discharge signals derived from the field,» *IEEE Transactions on Dielectric and Electrical Insulation*, vol. 15, nº 6, pp. 1517-1525, 2008.
- [25] B. T. P. a. T. R. B. K. X. Lai, «Application of Data Mining on Partial Discharges Part I: Predictive Modelling Classification,» *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 17, nº 5, 2010.
- [26] M. H. S. O. H. Hirose, «Diagnosis of Electric Power Apparatus using the Decision Tree Method,» *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 15, nº 6, 2008.
- [27] T. Babnik, «Data Mining on a Transformer Partial Discharge Data Using the Self-organizing Map,» *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 15, nº 6, 2007.
- [28] J. Vesanto, «Data Exploration Process Based on the Self Organizing Map,» 2002.
- [29] M. Kantardzic, *Data Clustering, Theory, Algorithms and Methods*, Philadelphia: ASA-SIAM, 2007.
- [30] E. A. J. Vesanto, «Clustering of the Self Organizing Map,» *IEEE Transactions on Neural Networks*, vol. 11, nº 3, pp. 1082-1089, 2000.
- [31] G. Pözlbauer, «Survey and Comparison of Quality Measures for Self-Organizing Maps,» *Proceedings of the Fifth Workshop on Data Analysis*, pp. 67-82, 2004.
- [32] F. D.-M. E. Arzuaga-Iriarte, «Topology Preservation in SOM,» *International Journal of Mathematical and Computer Science*, pp. 19-22, 2005.

- [33] W. K.-L. L. K. C.-R. L. H.-C. L. J.-F. Y. Miin-Sen, «On three types of competitive learning algorithms with their comparisons and applications to MRI Segmentation,» *International Journal of Intelligent Systems*, pp. 1-22, 2010.
- [34] Y. Ming-Cheung, «k*-Means: A new generalized k-means clustering algorithm,» *Pattern Recognition Letters, Elsevier*, pp. 2883-2893, 2003.
- [35] A. K. K. P. C. D. E. M. C. Ahalt, «Competitive learning algorithms for vector quantization,» *Neural Networks*, pp. 277-290, 1990.
- [36] S. S.-Y. C. L.-H. Y. Jen-Chang, «Improving Rival Penalized Competitive Learning using Density-Evaluated Mechanism,» *Journal of Chinese Institute of Engineering*, pp. 1083-1088, 2010.
- [37] J. H. E. A. J. P. J. Vesanto, «Self Organizing Map in Matlab: The SOM Toolbox,» *Proceedings of Matlab DSP Conference*, pp. 35-40, 1999.
- [38] M. B. W. R. D. Brugger, «Automatic Cluster Detection in Kohonen's SOM,» *IEEE Transaction on Neural Networks*, pp. 442-459, 2008.
- [39] E. M. K. Tasdemir, «Exploiting Data Topology in Visualization and Clustering of Self Organizing Maps,» *IEEE Transactions on Neural Networks*, pp. 549-562, 2009.
- [40] A. J. McGrail, «Data mining techniques to assess the condition of high voltage electrical plant,» *CIGRÉ*, p. 01 – 12, 2002.
- [41] S. Markalous, «Detection and location of Partial Discharges in Power Transformers using acoustic and electromagnetic signals,» 2006.
- [42] S. Strachan, «Practical applications of data mining in plant monitoring and diagnostics,» *IEEE Power Engineering Society General Meeting*, pp. 1-7, 2007.
- [43] C. Forssén, «Modelling of cavity partial discharges at variable applied frequency,» 2008.
- [44] G. C. C. G. F. G. R. Bozzo, «Design of Experiments (DOE) procedures for the optimisation of a digital partial discharge measurement system,» *IEEE Annual Report - Conference on Electrical Insulation and Dielectric Phenomena*, pp. 64-68, 1996.
- [45] O.-Z. C. J. S. L.-O. S. C. C. Jaramillo-Vacio R., «Diagnosis of Partial Discharge using Self Organizing Maps and Hierarchical Clustering - An Approach,» *Hybrid Artificial Intelligence Systems*, pp. 91-98, 2011.
- [46] R. O. Z. A. R.-L. A. Jaramillo-Vacio, «Optimal Tuning Parameters of Self Organizing Maps for Classification of Partial Discharge Using Response Surface Methodology,» *International Journal of Engineering Research and Applications*, pp. 974-984, 2013.
- [47] F. Figueroa-Godoy, «Evaluación de aislamiento en equipos eléctricos usando mediciones de descargas parciales,» CIATEC, 2013.
- [48] R. O.-Z. A. P. J. Jaramillo-Vacio, «Competitive learning for self organizing maps used,» *Programación Matemática y Software*, vol. 5, nº 2, pp. 6-12, 2013.

- [49] R. O.-Z. A. Jaramillo-Vacio, «Evaluación de Algoritmos Competitivos en Mapas Autoorganizados para la,» de *Congreso Internacional de Innovación Tecnológica*, Cortazár, Guanajuato, 2013.
- [50] R. O.-Z. A. Jaramillo-Vacio, «Competitive Learning for Self Organizing Maps used in Classification of Partial Discharge,» de *International Conference on Electronics, Communications and Computers*, Tonantzintla, Puebla, 2012.
- [51] B. T. P. T. R. B. K. X. Lai, «Partial Discharge Analysis using PCA and SOM,» *IEEE Power Tech Laussane*, pp. 256-265, 2008.
- [52] A. Mazroua, «PD pattern recognition with neural networks using the multilayer perception technique,» *IEEE Transactions on Electrical Insulation*, pp. 1082-1089, 1993.
- [53] A. Krivda, «Automated Recognition of Partial Discharge,» *IEEE Transactions on Dielectrics and Electrical Insulation*, pp. 796-821, 1995.
- [54] W. C. S. O. K. P. S. G. J. Kim, «Partial Discharge Pattern Recognition Using Fuzzy-Neural Networks (FNNs) Algorithm,» *IEEE International Power Modulators and High Voltage Conference*, pp. 272-275, 2008.
- [55] B. K. D. C. L. S.-Y. X. G.-Z. L. Ri-Cheng, «Study on Partial Discharge Localization by Ultrasonic Measuring in Power Transformer Based on Particle Swarm Optimization,» *International Conference on High Voltage Engineering and Application*, pp. 600-603, 2008.
- [56] L. H. P. L. N. Fadilah-Ab-Aziz, «Analysis of Partial Discharge Measurement Data Using a Support Vector Machine,» *5th Student Conference on Research and Development*, pp. 1-6, 2007.
- [57] K. Lai, «Descriptive Data Mining of Partial Discharge using Decision Tree with genetic algorithms,» *AUPEC*, 2008.

